IMAGING TECHNOLOGIES AND DEEP LEARNING FOR IN-FIELD HIGH THROUGHPUT PLANT PHENOTYPING AND POSTHARVEST QUALITY ASSESSMENT

by

Yu Jiang

(Under the direction of Changying Li)

Abstract

Modern agriculture is facing tremendous challenges in its sustainability, productivity, and quality for a rapidly growing human population, changing climate, and shortfall of arable land and water resources. Improved crops and advanced farming management are essential to tackling those challenges, but both encounter the same bottleneck in the evaluation of plant performance (plant phenotyping) and postharvest quality. Thus, it is paramount to utilize sensing, automation, and data analytics technologies to develop innovative solutions for high throughput plant phenotyping and postharvest quality assessment. This dissertation focused on the development of imaging-based approaches for the accurate, rapid, and nondestructive measurements of key plant phenotypic traits and postharvest quality properties.

A ground mobile system was designed and implemented to integrate RTK-GPS, multi-view color, RGB-D, thermal, and hyperspectral imaging for field-based plant phenotyping. Computer control modules were developed for individual sensors and integrated into a LabVIEW program to control and synchronize sensors for field data collection. A number of image analysis methods were developed to extract phenotypic traits related to plant morphology, physiology, and development. Reflectance differences were studied between healthy

and bruised blueberry tissues in the spectral range from 950 nm to 1650 nm, providing a basis for non-destructive detection of blueberry bruising. A new index, the bruise ratio index, was defined and calculated using a machine learning based approach to quantify bruise severity for individual berries. Experimental results showed that the developed systems and methodologies can accurately and rapidly extract key phenotypic traits (height, width, projected leaf area, volume, photosynthetic efficiency at the canopy level, germination rate, and flowering patterns) and postharvest quality properties (bruise ratio). Such extracted traits also demonstrated their usefulness for genetics/genomics studies and in farm management. Thus, these developed systems and methodologies can be effective and efficient tools for the evaluation of plant performance (plant phenotyping) and postharvest quality.

INDEX WORDS: Multimodal Data Acquisition, Multi-threading Software Architecture, 3D Imaging, RGB-D, Portable LiDAR Instrument, Hyperspectral Imaging, Solar Induced Fluorescence, Photosynthetic Efficiency, Rapid Light Curve, Deep Convolutional Neural Network, Object Detection, Video Tracking, Plant Phenotyping, Plant Development, Flowering Pattern Characterization, and Bruising Detection

IMAGING TECHNOLOGIES AND DEEP LEARNING FOR IN-FIELD HIGH THROUGHPUT PLANT PHENOTYPING AND POSTHARVEST QUALITY ASSESSMENT

by

Yu Jiang

B.S., Ludong University, China, 2011

M.S., China Agricultural University, China, 2013

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2019

IMAGING TECHNOLOGIES AND DEEP LEARNING FOR IN-FIELD HIGH THROUGHPUT PLANT PHENOTYPING AND POSTHARVEST QUALITY ASSESSMENT

by

Yu Jiang

Approved:

Major Professor: Changying Li

Committee: Andrew H. Paterson
Kurt C. Lawrence
Mark A. Haidekker
Peter A. Kner
Tianming Liu

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2019

Dedication

To people who are struggling with hunger in the world;

To my parents;

To my wife.

Acknowledgments

I would like to express my sincere appreciation to my major advisor Dr. Changying Li for his mentoring, encouragement, and trust during my Ph.D. study. I would not have made this work and many other achievements without his immense efforts in the past six years. I would also like to thank Dr. Li's family for their kindness to me and my family. I am extremely grateful to work with a diligent, erudite, and visionary advisor.

My deepest appreciation also goes to members of my advisory committee: Dr. Andrew H. Paterson, Dr. Kurt C. Lawrence, Dr. Mark A. Haidekker, Dr. Peter A. Kner, and Dr. Tianming Liu for their constructive guidance and helpful support. I would also like to extend my deepest gratitude to many of my collaborators: Dr. Fumiomi Takeda, Dr. John L. Snider, Dr. Glen C. Rains, Dr. Peng W. Chee, Dr. Randy Beaudry, Dr. Hamid Ashrafi, and Dr. Elizabeth A. Kramer for their great assistance in collaborative studies. Furthermore, I would like to thank many of my friends and lab colleagues including Dr. Weilin Wang, Mr. Rui Xu, Mr. Shangpeng Sun, Mr. Jon Robertson, Dr. Yi Fang, Dr. Mengyun Zhang, Mr. Zikai Wei, Dr. Ruoyu Zhang, and Mr. Jessie Kuzy. Their companion and encouragement helped me to complete one of my important studying journeys. I would also like to acknowledge the assistance and support from staff and colleagues in the College of Engineering and Graduate School during my study at the University of Georgia.

Last, I am deeply indebted to my parents and wife for their unconditional love, care, and support to my study and future career.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Significance of This Study

Food security is one of the biggest challenges for the world. The global population is likely to exceed 9 billion by 2050, which will require more food, fiber, and fuel products from agricultural production systems [3]. To fulfill these increasing demands, current crop productivity needs to be almost doubled by 2050, which translates into an annual growth rate of 1.75% of total factor productivity (TFP) [4]. On average, the current TFP annual growth rate is around 1.5% globally, but the TFP annual growth rate has decreased to 0.96% in developing countries, presenting a great challenge for the improvement of crop productivity. In addition to productivity, sustainability is another crucial factor for agriculture. Crop productivity must be increased in a sustainable way because the global population will keep increasing and could be over 11 billion by 2100, causing the situation to become even more challenging [5]. Moreover, agriculture sustainability is also facing tremendous challenges from different phenomena such as a decrease in the availability of the workforce, climate change, a shortage of arable land, and limited water resources [6].

There are two potential ways to address these issues: improvement of crops and improvement in crop management. Improvement of crops aims at breeding new cultivars so that crops would naturally produce a higher yield, better quality, and improved adaptability to various environments (e.g., saline soils). Improvements in crop management through implementing advanced farming techniques such as precision agriculture would minimize the input (e.g., irrigation and chemical applications) and maximize the outcome (e.g., productivity

and quality) for an agricultural production system through technological innovations (e.g., sensing, automation, and data science techniques). Both methods face the same conundrum: how to evaluate a large amount of plants in the field. Therefore, it is paramount to develop new technologies to accurately evaluate crops in a high throughput manner.

Over the past decade, high-throughput phenotyping platforms (HTPPs) in greenhouses or growth-chambers have been developed and run by several transnational seed companies or nation-wide public institutions such as the Australian Plant Phenomics Facility (APPF), the European Plant Phenotyping Network (EPPN), and the United States Department of Agriculture (USDA) [7]. Phenotypic traits of various plants can be quickly and accurately measured by these HTPPs in a non-destructive manner under controlled environments. However, depending on experimental purposes, phenotypic traits have to be measured under field conditions, because some quantitative trait loci or candidate genes could be less effective under field conditions than in controlled environments [7, 8, 9]. This occurs primarily because some important factors for plant growth and development, such as soil conditions, cannot be simulated in the controlled environments [10, 11, 12]. For instance, plants can only access a limited volume of soil in a pot, which is a significantly different scenario from field conditions. It is very difficult to simulate natural soil conditions in a controlled environment as it is made up of a complex mixture of various nutrients, organic matter, gases, liquids, and numerous organisms [12]. Therefore, it is necessary to develop innovative technologies for field-based high throughput plant phenotyping.

In recent years, advances in proximal sensing, computing, automation, and robotic techniques have created opportunities for field-based high-throughput phenotyping [13, 14, 8, 15, 16, 17, 18]. Computing, automation and robotic technologies ensure the improvement of phenotyping throughput, and proximal sensing techniques enable the feasibility of accurately measuring phenotypic traits under field environments.

1.2 Objectives

The overall goal of the dissertation was to develop data acquisition systems and data analytical approaches for field-based high throughput plant phenotyping and laboratory postharvest quality assessment. Specific objectives were as follows:

1. Develop a ground mobile data acquisition system with multimodal imaging for field based high-throughput plant phenotyping;
2. Extract plant morphological traits (canopy size and shape) from 3D point clouds
3. Characterize canopy-level photosynthetic activities by a ground hyperspectral imaging system
4. Characterize plant development (plant seedling and flowering) by color imaging and deep learning
5. Detect and quantify blueberry bruising by near infrared (NIR) hyperspectral images

1.3 Overview of the Dissertation Chapters

This dissertation contains nine chapters. Chapter 1 introduces the significance of this study and the current status of in-field high throughput plant phenotyping, and defines the objectives of this dissertation. Chapter 2 provides a comprehensive review on deep convolutional neural networks for plant phenotyping applications.

Chapter 3 reports on the development of a ground mobile system with multimodal imaging for field data collection on plant phenotyping. Calibrations were taken to validate the accuracy of the sensor measurements. A small-scale experiment was carried out to evaluate the usefulness of the system for plant phenotyping and the management of agricultural production systems.

Chapter 4 studies the use of 3D point clouds to characterize plant canopy size and shape, which can be used to select cultivars suitable for mechanical harvesting and to guide pruning actions for farm management.

Chapter 5 introduces a new methodology for estimating photosynthetic efficiency at the canopy level through a passive sensing approach.

Chapter 6 explores the combination of a deep learning-based object detector (Faster RCNN) and a classic video tracking method (Kalman filter) for detecting and counting plant seedlings in videos collected for individual plots.

Chapter 7 investigates the use of a multi-view color imaging system and a deep learning-based object detector (Faster RCNN) to count white blooms for individual plants over a growing season, which derives flowering curves to characterize flowering patterns for angiosperms in the field.

Chapter 8 examines reflectance differences between healthy and bruised tissues in the spectral range from 950 nm to 1650 nm, which can be used to establish a classification model for bruising detection and quantification.

Chapter 9 provides general conclusions and future research directions.

Chapters 2 to 8 represent seven manuscripts written individually. Chapter 2 provides a comprehensive review on convolutional neural networks in image-based plant phenotyping, The other chapters serve to address questions related to the development of a data acquisition system (Chapter 3) and image analysis methods for extracting phenotypic traits for plant morphology (Chapter 4), physiology (Chapter 5), development (Chapters 6 and 7), and

postharvest quality (Chapter 8). As they are independent, these seven chapters remain in a general manuscript format.

Chapter 2

Convolutional Neural Networks for Image-based High Throughput Plant Phenotyping: A Review[1]

[1]Yu Jiang and Changying Li. To be submitted to *Plant Phenomics*.

2.1 Abstract

Plant phenotyping has been recognized as the bottleneck for improving the efficiency of breeding programs, understanding of plant-environment interactions, and management of agricultural systems. In the past five years, imaging approaches have shown the great potential for high-throughput plant phenotyping, resulting in more attention paid to imaging-based plant phenotyping. With an increased amount of image data, it is urgent to develop robust analytical tools that can extract phenotypic traits accurately and rapidly. The goal of this review is to provide a comprehensive overview for the latest studies using deep convolutional neural networks (CNNs) in plant phenotyping applications. We specifically review the use of various CNNs and CNN meta-models for plant stress evaluation, plant development, and postharvest quality assessment. We systematically organize the studies based on the technical development from imaging classification, object detection, and image segmentation, thereby identifying state-of-the-art solutions for certain phenotyping applications. Finally, we provide several directions for future research in the use of CNNs and CNN meta-models for plant phenotyping purposes.

2.2 Introduction

Food security is one of the biggest challenges for the world. The global population is likely to exceed 9 billion by 2050, which will necessitate more food, fiber, and fuel products from agricultural production systems [3]. To fulfill the increasing demands, the current crop productivity needs to be doubled approximately by 2050, which translates into an annual growth of 1.75% of total factor productivity (TFP) [4]. On average, the current TFP annual growth is around 1.5% globally, but the TFP annual growth decreases to 0.96% for developing countries, thus presenting a significant challenge for the improvement of crop productivity. In addition to the productivity, sustainability is another crucial factor for the agriculture. Crop productivity must be increased in a sustainable way because the global

population will continue to increase and could exceed 11 billion by 2100, making these situations even more challenging [5]. Agriculture sustainability, however, faces tremendous challenges from a decreased availability of workforce, changing climate, shortfall of arable land, and limited water resources [6]. It is thus paramount to improve simultaneously the productivity and the sustainability of agricultural production systems.

There are two potential ways to address these issues: improvement of crops and improvement of management. Improvement of crops aims to breed new crop cultivars, so that crops would naturally have a higher yield, better quality, and improved adaptability to various environments (e.g., saline soils). Improvement of management seeks to advance farming concepts such as precision agriculture, which minimizes the input (e.g., irrigation and chemical application) and maximizes the outcome (e.g., productivity and quality) for an agricultural production system through technological innovations (e.g., sensing, automation, and data science techniques). Both ways face the same bottleneck: evaluation of a large amount of plants in the field. It is therefore paramount to develop new technologies to accurately evaluate crop plants in a high throughput manner.

High throughput plant phenotyping (HTP) has been recognized as integral to overcoming this bottleneck [7, 19, 9, 20, 21, 22, 23, 18]. In the past five years, various HTP solutions were developed to dramatically improve phenotyping capability and throughput, including tower-based systems, gantry-based systems, ground mobile systems, low- and high-altitude aerial systems, and satellite-based systems. An obvious trend has been noticed in the recent development of HTP systems: imaging sensors have been used more frequently because of their ample capacity for extracting complex traits. Imaging-based solutions have been used for a wide range of phenotyping applications covering plant morphology, physiology, development, and postharvest quality. A typical pathway for imaging-based plant phenotyping is a four-step cyclic graph (Figure 2.1). The first step is to identify and define phenotypic

traits to be measured, which largely determine the use of suitable imaging modalities for plant sensing. Measurement of phenotypic traits usually demands one or more computer vision tasks (e.g., fruit counting would require object detection). This problem can be solved by developing new or improved algorithms through conventional image/signal processing, machine learning, or a combination of the two. Consequently, data processing pipelines can be designed to extract defined phenotypic traits to support and facilitate domain applications such as genetics/genomics studies, breeding programs, and farm management. Among these options, algorithm development becomes noticeably challenging because of significant disparities in image quality (e.g., illumination, sharpness, occlusions, and so forth) [24]. These image quality variations dramatically affect the performance of image/signal processing algorithms and to poor generalizations for measuring the same phenotypic traits from different datasets. Conventional machine learning (ML)-based approaches generally provide an improved generalizability, but most of them still cannot provide the acceptable generalizability needed for current phenotypic purposes. In addition, conventional ML approaches require significant efforts to design data representations (features) manually that are invariant to imaging environment changes. Feature designing is laborious and needs expertise in computing and image analysis, which prevents the use of conventional ML techniques for phenotyping applications.

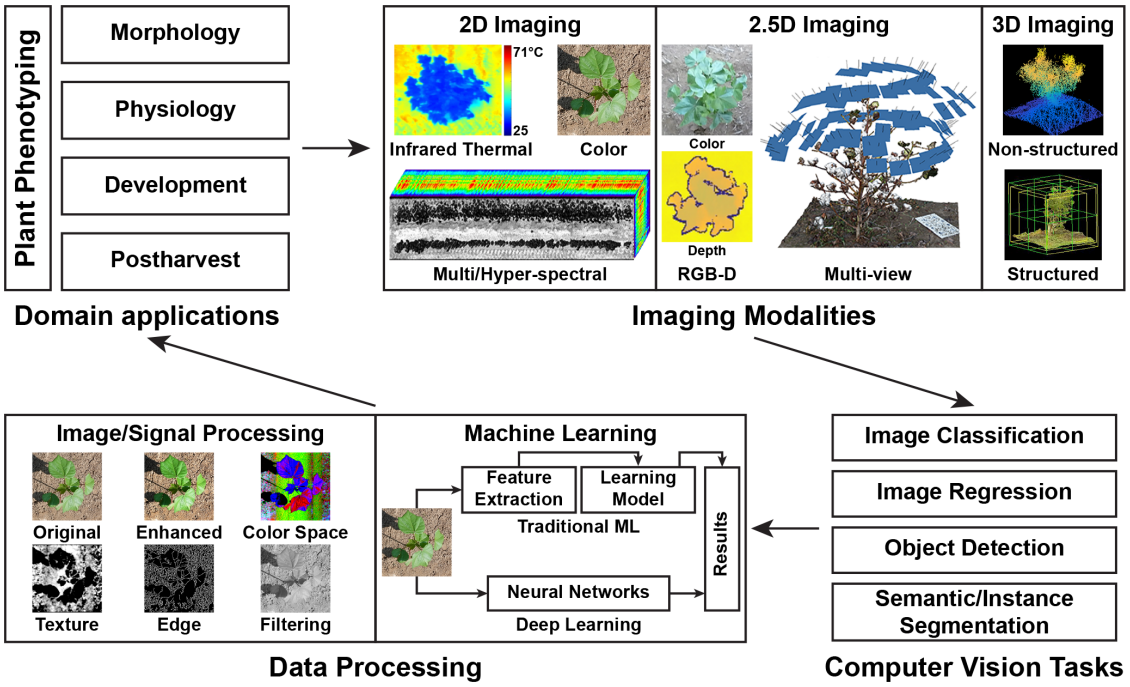

Figure 2.1: Diagram of the pathway of imaging-based plant phenotyping.

Deep learning (DL) is a subset of machine learning and it allows the learning of data representations in a hierarchical way. The key advantage of DL is that features will be learned automatically from input data, thereby breaking down barriers to developing intelligent solutions for different applications. A commonly used DL architecture is deep convolutional neural networks (CNNs) that have achieved state-of-the-art performance for important computer vision tasks, such as image classification/regression, object recognition, and image segmentation (both semantic and instance). CNNs originated in the 1980s [25] and showed their first success in the recognition of hand-written digits in the 1990s by using backpropagation-based training [26]. In 2012, a breakthrough (AlexNet) was made because of significant improvements in computational power (and therefore CNN model complexity) and availability of annotated datasets (e.g., ImageNet) [27]. Since then, various CNN models have been developed for image classification and eventually showed better performance than humans on the same dataset [28]. In addition, CNNs have been used widely as feature extractors to be integrated with a meta-model for other computer such vision

tasks as object detection and semantic and instance segmentation. CNN-based solutions have provided state-of-the-art performance compared to traditional approaches in almost all of these tasks, demonstrating a great potential to improve data analysis performance in such imaging-based applications as imaging-based plant phenotyping. In particular, the advancement of transfer learning (or domain adaptation) and the emergence of DL development framework further facilitate the use of DL techniques for domain applications, so DL approaches have been adopted rapidly for plant phenotyping in recent years, and an exponentially increasing trend is foreseen for DL-based plant phenotyping in the future. It is thus necessary to conduct a literature review to summarize the existing knowledge, good practices, limitations, and potential solutions for applying DL techniques in plant phenotyping.

Several papers have been published in the last two years that provide comprehensive reviews of DL techniques for such computer vision tasks as image classification [29], object detection [30], and semantic segmentation [31]. These reviews effectively summarize the basic principles, development history, and future trends for CNNs and CNN meta-models in computer vision, but none of them provide information related to agriculture and therefore highlight a gap between the technology theories and phenotyping applications. There have been pioneering efforts that have focused on various DL techniques for general agriculture applications [32] and plant stress phenotyping [33]. They were, however, either too broad (covering all DL techniques for all agricultural applications) or too narrow (limited to a particular phenotyping task) and lack a focused and comprehensive review on DL in image-based plant phenotyping.

The goal of this review paper is to thoroughly scrutinize the current efforts, provide insights, and identify potential research directions for the utilization of CNNs and CNN meta-models for image-based plant phenotyping. This review paper targets key phenotyping

tasks related to plant physiology, plant development, and postharvest quality assessment. With these efforts, it is expected that readers can bring CNNs and CNN meta-models into their research projects to benefit the plant phenotyping community. The rest of this paper is organized in the following way: Section 2.3 provides a concise introduction to important CNNs and CNN meta-models used in image classification, object detection, and semantic and instance segmentation. Section 2.4 provides a review of CNNs and CNN meta-models for image based plant phenotyping; Section 2.5 discusses key issues in the application of CNNs and CNN meta-models to plant phenotyping; and finally, Section 2.6 provides conclusions and future research directions.

2.3 CNNs and CNN Meta-models for Computer Vision Tasks

Since 2012, CNNs have dominated the solutions for computer vision tasks because of their superior performance. While efforts have been made to thoroughly review the development of CNNs and CNN meta-models for computer vision tasks [31, 29, 30], it will be useful to provide a brief introduction here for the completeness of this review. Most image-based phenotyping applications essentially demand solutions for one or more tasks related to image classification, object detection, and segmentation, so CNNs and CNN meta-models for those tasks are reviewed in this section. Because CNNs and CNN meta-models evolve rapidly, the following review is limited to models that provide significant performance improvements and are used widely as benchmark methods by other domain applications. For convenience, useful information is summarized for those reviewed models, including development year and group, the original reference, the key innovation concept, and the source code (or third-party implementation) if available (Table S1).

2.3.1 Image Classification

Image classification is one of the core tasks in computer vision and aims to assign images with predefined class labels. CNNs are artificial neural networks that combine a set of

mathematical operations (e.g., convolution, pooling, and activation) using various connection schemes (plain stacking, inception, and residual connection), and the operational parameters (e.g., convolutional kernels) can be learned from annotated images to predict image class labels (Image Classification in Figure 2.2). The development of modern CNNs for image classification can be grouped into three stages: 1) emergence of modern CNNs (2012 to 2014); 2) intensive development and improvement of CNN architecture (2014 to 2017); and 3) reinforcement learning for CNN architectural design (i.e., the concept of using artificial intelligence, AI, to improve AI, 2017 to present).

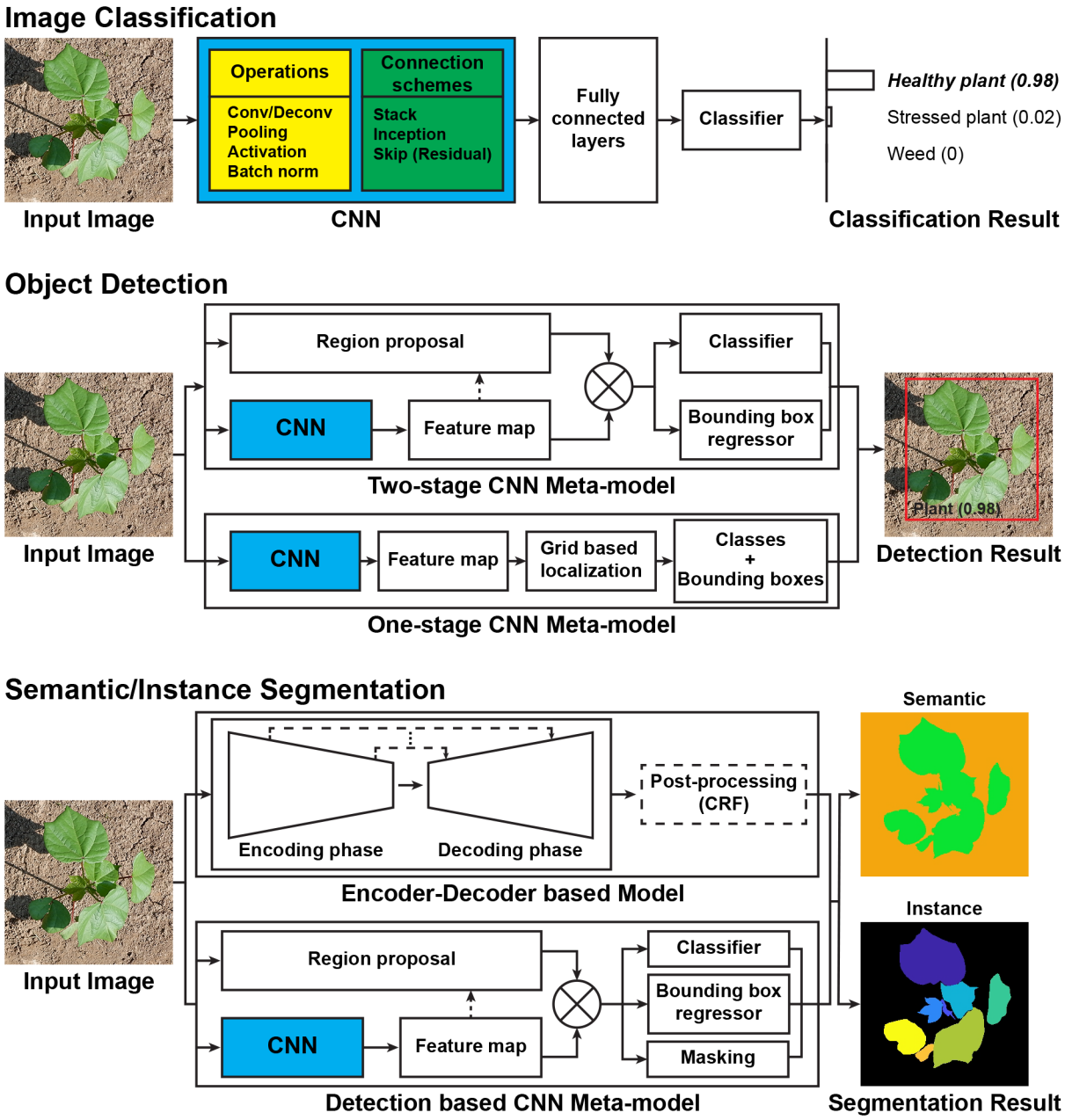

Figure 2.2: Diagrams of mechanism of CNNs and CNN meta-models for image classification, object detection, and semantic and instance segmentation.

In 2012, the first modern CNN network (also known as AlexNet) was reported by Krizhevsky et al. [27] and demonstrated breakthrough performance on image classification in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2012) competition. It showed improvements of 8.2% and 8.7% on top-1 (35.7% versus 45.7%) and top-5

(17% versus 25.7%) errors. This work led the trend of applying CNNs for image classification and other computer vision tasks. On the other hand, studies were conducted to understand the mechanism of CNNs, so CNNs could be improved in a systematic way. A pioneering work improved AlexNet to a new variant (ZFNet) using a visualization tool [34]. The visualization tool is a framework that is integrated with CNNs and can map neuron activities back to the input pixel space. Pixel-wise activations, therefore, can be visualized after each convolutional layer. This would be particularly useful for researchers to understand the CNN mechanism and improve architectural design. The study also showed that learned features could be generalized to various classifiers, suggesting CNNs could learn general representations of images rather than specific features for classification. CNN architecture was intensively studied from 2014 to 2017, and several representative CNNs were developed with improved architectural design, including VGGNet [35], Inception-based CNNs [36, 37], ResNet and its variants [28, 38], and DenseNet [39]. These improvements attempted to dramatically enhance network learning capability and reduce computational complexity by using efficient operations (e.g., 3 by 3 convolutional operation as the building block) and revised connection schemes (e.g., inception modules, residual modules, and dense blocks). With these improvements, representative CNNs can usually surpass human performance on image classification for various datasets. It should be noted that performance improvement because of CNN architectural modification was heavily dependent on human expertise and tuning efforts, which means that CNN architectural improvement could be as laborious as feature engineering in traditional ML used along with different applications. To solve this problem, a study was conducted to explore the possibility of searching optimal CNN architecture using reinforcement learning which is a learning method to reward operations yielding improved performance [40, 41]. A reinforcement learning framework was introduced to seek optimal convolutional cells on a small annotated dataset, and the resultant cells were stacked in different ways and transferred to a large unknown dataset. Experimental results showed that CNNs built by searched convolutional cells provided different degrees of perfor-

mance improvement than CNNs designed manually. This work demonstrated the capability of using AI to improve AI, which is the new direction for solving some problems in designing CNN architecture. The search process, however, is extremely expensive computationally (500 NVIDIA P100 GPUs for 4 days), which limits the potential use for other domain applications.

2.3.2 Object Detection

Object detection seeks to detect and classify all potential objects in a given image. The use of CNNs for object detection can be categorized into two groups: one-stage and two-stage CNN meta-models (Object Detection in Figure 2.2). Two-stage models were firstly studied because it was intuitive to combine region proposal methods (or ROIs) and CNNs. OverFeat framework was developed to use a single CNN to extract features for training classifiers and regressors separately [42]. The trained classifiers and regressors were used to predict class labels and bounding box coordinates, respectively, for candidate ROIs generated using a sliding window method. Although the OverFeat framework provided the best performance on the localization task of the 2013 ILSVRC competition, the high computational cost and complex framework training presented difficulties for practical applications. A region-based CNN (RCNN) family was introduced to solve those issues, including the original RCNN [43], Fast RCNN [44], and Faster RCNN [45]. Three key techniques were identified in CNN meta-models of the RCNN family. First, a region proposal network (RPN) was developed to generate candidate object ROIs using features extracted from CNNs, which simultaneously saved processing time and increased region proposal accuracy. Second, an ROI pooling operation was developed to extract a fixed number of features from ROIs with varying sizes, thereby avoiding the repeated computation of features for different ROIs. Third, a multi-task loss function was used to unify the training process, which enabled an end-to-end training for object detection. With the three improvements, Faster RCNN has been widely used as either a benchmark for performance comparison or an object detector for domain applications (e.g.,

pedestrian detection in autonomous driving) because it is easy to train and mostly provides good detection performance. Although Faster RCNN provides state-of-the-art accuracy, its efficiency is still far behind for real-time applications such as autonomous driving. This is mainly because the two-stage meta-models spend time handling different components for inference [30]. Compared with two-stage models, one-stage models can reduce time expense by global regression/classification, mapping directly from image pixels to bounding box coordinates and class probabilities. In other words, candidate object regions are generated from each pixel in feature maps, and then classified and fine-tuned to accurate object boundaries. Representative one-stage models include you-only-look-once (YOLO) family [46, 47] and single shot detector (SSD) framework [48]. A critical issue, however, has been realized for these one-stage models: an extreme imbalance between the number of object and background regions. Most image regions contain only the background information (identified as irrelevant regions), providing a limited contribution to model training process. A focal loss function was proposed to put more penalties on inaccurately detected (or classified) samples, which solved the issues resulting from sample imbalance and ultimately led to the development of an improved one-stage framework RetinaNet [49]. When using the same CNN backbone model, RetinaNet achieved comparable performance with Faster RCNN and 29% improvement of computational efficiency. Nevertheless, if detection accuracy is the most important factor to consider, two-stage models would be the option; otherwise, one-stage models provide better computational efficiencies for embedded systems.

2.3.3 Semantic and Instance Segmentation

Semantic segmentation seeks to provide masks for objects with the same semantic meaning, e.g., all plants in an image, whereas instance segmentation seeks to provide individual objects in a given image. In general, CNN meta-models for semantic/instance segmentation can be classified into two groups: encoder-decoder based framework and detection based

framework (Semantic and Instance Segmentation in Figure 2.2).

Encoder-decoder based models usually contain two phases. The encoder phase uses CNNs to extract feature maps that are semantically meaningful from input images, and the decoder phase uses transposed convolution (also known as deconvolution) for upsampling of extracted feature maps to per-pixel labels. Two techniques have been used to improve the segmentation accuracy of encoder-decoder models. First, a lateral connection scheme is used to link feature maps with the same spatial resolution between the encoder and decoder phases, which helps with the transition of semantic meaning from input images to output segmentation results [50]. Second, a conditional random field (CRF) is used as a post-processing method to improve the segmentation accuracy of object boundaries [51]. Representative encoder-decoder based models include U-Net [50], fully convolutional network (FCN) [52], and DeepLab family [53, 51]. A detection-based framework relies on CNN meta-models for object detection. Several studies have explored the use of object detection models for instance segmentation, including simultaneous detection and segmentation (SDS) based on RCNN [54] and DeepMask based on Faster RCNN [55]. They did not, however, reach an acceptable performance for the instance segmentation task [31]. Breakthrough performance was achieved by Mask RCNN that adds an FCN network on top of Faster RCNN for generating masks of individual objects [56]. Many successive studies and applications also proved that Mask RCNN could provide state-of-the-art performance for instance segmentation.

2.4 CNNs for Image based Plant Phenotyping

2.4.1 CNNs for Plant Stress Phenotyping

Plant stress phenotyping is an important aspect of plant phenotyping and can be used to evaluate resistance and tolerance among different accession lines for genetics/genomics studies and breeding programs. In addition, plant stress detection is crucial for understanding plant physiological activities. Plant stress phenotyping can be categorized into four

stages: 1) identification (presence of stress); 2) classification (type of stress); 3) quantification (severity of stress); and 4) prediction (possibility of stress occurrence) [33]. From the computer vision perspective, all the four stages can be considered as an image classification task, whereas some stages could involve other processing methods such as object detection and semantic/instance segmentation.

Studies related to stress phenotyping have focused primarily on stress identification and classification, utilizing CNNs as image classifiers to identify images that contain stressed plants or plant leaves. The development of image classification-based studies can be split into two phases. In the first phase, studies intensively investigated CNNs with representative or custom architecture because of the availability of annotated datasets and the simplicity of CNN implementation and training for image classification. There are several large annotated image sets for plant biotic and abiotic stresses that accelerate the evaluation of various CNNs for stress phenotyping. For instance, PlantVillage (https://plantvillage.psu.edu/) is a publicly available image dataset containing over 54,000 labelled images of plant leaf. These images cover 14 crop species for 26 types of stresses and healthy leaves. This can be used to either evaluate a new CNN architecture or pretrain a CNN model for transfer learning. Data annotation for image classification is also relatively easy, so a large number of images in a newly collected dataset can be annotated with reasonable time and cost, especially when a proper data collection procedure is used. As a result, studies related to plant stress detection typically have a sufficient number of annotated images (several thousand or more) for model training. In addition, a DL framework has been developed to accelerate the implementation and training of CNNs for image classification. Commonly-used DL frameworks include Caffe (University of California Berkeley), Theno (University of Montreal), TensorFlow (Google), PyTorch (Facebook), CNTK (Microsoft), and so on. Key CNNs (e.g., Inception-based CNNs, ResNet family, and DenseNet) have been implemented using a different framework, and researchers can quickly develop a simple computer program for training CNNs as long as

annotated data are available. These advancements facilitate the use of CNNs for plant stress detection. By using good training practices (e.g., data augmentation, background removal, and transfer learning), various studies showed that CNNs achieved accuracies from 87% to 99% for stress identification and classification [57, 58, 59, 60, 61, 62, 63, 64, 65, 66]. Details of these studies can be accessed in the comprehensive overview provided for those studies [33].

In the second phase, pioneering studies attempted to understand reasons leading to high performance of CNNs for stress identification and classification because the understanding would not only help to improve CNNs but also ensure biological correctness of obtained results. Although some studies adopted the deconvolution layers to visualize the activated pixels in different convolutional layers, the visualization results were not used to compare with human evaluation or correlate to biological knowledge. Until 2018, an explainable framework (xPlNet) was developed to both identify (or classify) plant stresses and generate an explainable map that showed pixels determining the identification (or classification) results [1] (Figure 2.3a). In this framework, the reference activation level (the mean pixel intensity plus three times pixel intensity variation) of healthy leaves was calculated for each of the feature maps extracted in the first convolutional layer. For a testing image, feature maps from the first convolutional layer subtracted the reference activation to calculate the feature-importance metric (weighted average of leaf pixel intensity in each of the feature maps). Feature maps were ranked based on their importance, and the top-K (K = 3 in the original study) feature maps were selected to calculate the explainable map (EM). The mean intensity of the EM can be used to quantify stress severity. A separate study also examined various techniques to understand the mechanism of CNNs for disease diagnose [2]. Explanation maps generated by xPlNet generally showed the best correlation with manual annotation and validated its efficacy for finding pixels correlated to plant diseases (Figure 2.3b). Compared with studies in the first phase, the two pioneering studies demonstrated the importance of understanding the mechanism of CNNs for stress phenotyping as

well as the potential for stress severity quantification.

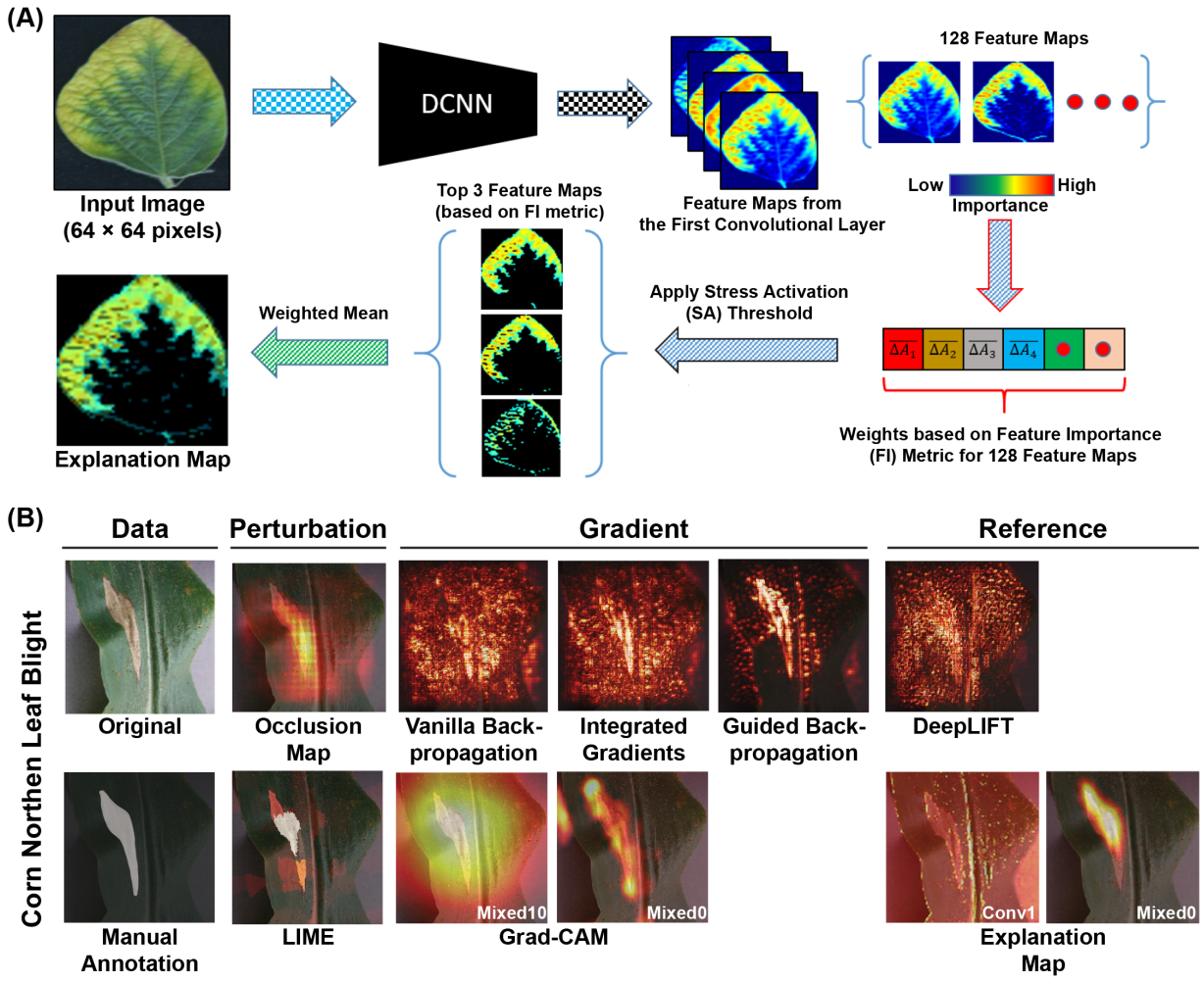

Figure 2.3: Key concept and results of xPlNet for plant stress detection: (A) diagram of the developed xPlNet for calculating explanation map for a given image, and (B) visualization results using different methods for an image containing a stressed leaf. (A) and (B) were reproduced using figures from [1] and [2], respectively.

In addition to the classification of entire images, a study was conducted to generate a heat map of stressed lesion probabilities from small patches obtained using a sliding window over a given image [67]. Generated heat maps were used as input images to a separately trained CNN for detection of stressed lesions. The developed method showed two

advantages. First, high-resolution images were directly processed without downsampling, so detailed spatial information could be utilized by CNNs. Second, generated heat maps were used as a visualization tool to explain reasons leading to classification decisions. This advantage, however, was not recognized and fully explored. As the combination of the sliding window method and CNN resembled the OverFeat framework (an early effort on using CNNs for object detection), improved CNN meta-models for object detection were used for plant stress phenotyping [68]. Three representative meta-models (Faster RCNN, SSD, and R-FCN) were trained and evaluated, and experimental results showed that the best detection accuracy was 86% at an intersection of union (IOU) level of 0.5. Trained meta-models could identify and localize the symptomatic regions. As plants could be infected by multiple diseases, the object detection based solutions could detect all possible causes, providing a more comprehensive evaluation than image classification based solutions.

Semantic/instance segmentation could provide pixel-level identification of plant stresses, so that plant stresses can be easily quantified using the ratio of stressed pixels and healthy pixels. To the best of our knowledge, however, there was no study that explored the use of segmentation-based solutions for plant stress phenotyping at the time of writing this review. This could be an important direction for research in the future.

2.4.2 CNNs for Plant Development

Plant and organ development is another important aspect for plant phenotyping. Detection and counting of plants and plant organs are the key to characterizing plant development. In this section, a comprehensive overview is provided for studies related to the detection and counting of plants and plant organs. Based on data format, these studies can be classified as two categories: detection and counting in still images and detection and counting in image sequences and videos.

Detection and Counting in Still Images

It is straightforward to apply CNN meta-models designed for the object detection task to detect plants and plant organs in still images (Detection based Methods in Figure 2.4). DeepFruits is the first study that explored the use of modern CNN meta-models (i.e., Faster RCNN) for fruit detection [69]). Several key contributions were recognized in this study. First, transfer learning was used to train a Faster RCNN model with 100 labeled images, demonstrating the potential of using limited labeled images to train CNN meta-models. Second, when using RGB images, the trained Faster RCNN model provided a 1% improvement of F1 score than the CRF model. Third, data fusion was conducted at the raw-data level and decision level for Faster RCNN models. Experimental results showed that decision-level fusion further improved the F1 score to 0.838 (additional 2% compared with Faster RCNN with no fusion). However, raw-data level fusion showed a 2% reduction of F1 score than the RGB-based Faster RCNN. Two potential reasons were given for the performance reduction. First, the decision-level fusion contained two Faster RCNN models, which had twice the parameters that a single Faster RCNN model had to model image data distribution, and ultimately resulted in the performance improvement. Second, the performance reduction of raw-data level fusion was because the weights pretrained on ImageNet dataset were more suitable for RGB color images than NIR images.

Although the two explanations are reasonable, a more possible reason might be the ineffective transfer learning of the revised Faster RCNN architecture. In order to use four-channel (RGB-NIR) images for training, the receptive field of the first layer in the backbone CNN was changed from 3 to 4, meaning each filter in the first layer had an additional dimension that had to be initialized randomly. As a consequence, the output from the revised first layer was not likely to follow the data distribution pretrained on ImageNet dataset, and this new data distribution could eventually ruin the rest of the CNN because CNNs are hierarchical and deep layers are dependent on shallow layers [29]. In other words, pretrained weights in

deep layers would not effectively model data, resulting in lower transfer learning efficiency. Even worse, if pretrained weights are somehow in a local minima or saddle, transfer learning could provide worse results than randomly initialized weights. Last, the study also applied the Faster RCNN model for other fruit such as cantaloupe, apples, avocados, mangos, and oranges, showing the generalizability of Faster RCNN for fruit detection. While the study reported much useful information, a major drawback was the limited images for training and testing. Although 100 images could let researchers train a Faster RCNN with high accuracy, the testing image sizes (from 11 to 34) were too small to confirm the achieved high performance. In particular, training and testing images were acquired in the same condition, which significantly reduced the variation of images. This may also be one major factor that CNN-based solutions showed only marginal improvements over conventional methods.

Several studies generally followed similar practices and used the region-based CNNs (e.g., RCNN and Faster RCNN) for detection and counting of mushrooms [70]. Two critical issues were not addressed by these studies. The first issue was to train CNN meta-models using high-resolution images that usually have a large size and cannot be fed into CNN meta-model for training. A new approach was developed to solve this issue by splitting one high-resolution image into multiple small patches. Each patch still had a reasonable size (e.g., 500 by 500 pixels), so that all the patches could be used to train a CNN meta-model such as Faster RCNN with high-resolution images [71]. In the testing stage, an image was still split into patches with a certain overlap (e.g., 50% between two neighboring patches) and a Faster RCNN model was used to detect maize ears in each patch. Because of the high overlap between patches, one ear could be detected in multiple patches. Overlap between each pair of detections were calculated and detections with small pixel area were removed to avoid repeated detections. This strategy substantially increased training samples and could handle images with an arbitrary resolution. The second issue was to detect small-sized objects, which is also a general challenge for CNN based object detection methods.

An intuitive solution was to use features from shallow layers for region proposals because features from shallow layers reserved more spatial information and would be able to identify small-sized objects. Based on this, features from multiple layers (shallow, middle, and deep) were used for regional proposals of Faster RCNN models [72]. Compared with standard Faster RCNN models, the modified Faster RCNN model achieved an improvement of 4.7% for the F1 score for detecting almonds in still images.

Apart from Faster RCNN, a custom two-stage framework was proposed to use superpixels generated by the simple linear iterative clustering (SLIC) algorithm as region proposals [73]. A CNN model was used to classify each superpixel as flower or non-flower objects. While this approach showed higher performance than conventional ML methods (e.g., color features and SVM classifier), it had a potential limitation in region proposal. The advantage of end-to-end CNN meta-models is that they are able to use rich features for accurate localization, especially when images varied dramatically. However, superpixels were subject to image variations and might not provide optimal region proposals. Thus, the generalizability of this approach is very likely inferior than the end-to-end methods.

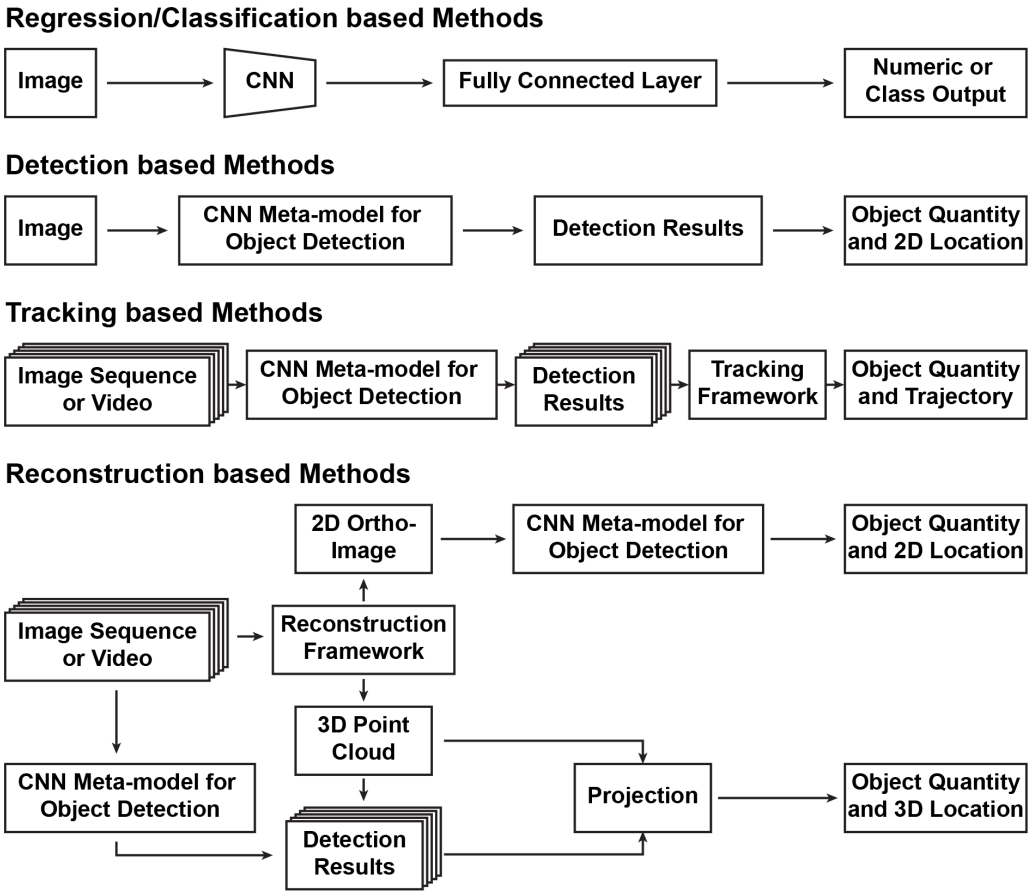

Figure 2.4: Diagrams of key concepts for using CNNs and CNN meta-models for plant/organ detection, counting, and localization.

In addition to the two-stage meta-models, one-stage meta-models were investigated for situations that required fast processing speed. YOLO-v2 was used to detect and count apples and pears in still images [74]. Compared with the original YOLO-v2 model, a modification was made to change the default grid cell from 13 by 13 to 26 by 26, so that relatively small-sized apples could be detected. The modified YOLO-v2 model achieved an F1 score of 0.9 at the IOU level of 0.5. As the study was concerned about inference speed, the authors halved the YOLO-v2 model from 24 layers to 12 layers, providing a dramatic increase of processing speed (from 4 FPS to 10 FPS) with an acceptable accuracy reduction (F1 score from 0.9 to 0.8).

For some applications, fruit/organ quantity is more important than location information, and thus problems can be simplified as an image regression or classification problem (Regression/Classification based Methods in Figure 2.4). For regression-based methods, a major modification was to replace the softmax layer of a CNN with a single neuron for regressing numeric values (e.g. fruit counts). This simple end-to-end counting solution provided high accuracies (over 90%) for counting tomatoes [75], Arabidopsis leaves [76, 77], and tassels [78].

In particular, the CNN feature extractor for tomato counting was trained using synthetic images and tested on real images, showing the potential for developing simulated learning methods to reduce the efforts on data annotation. This provided a useful direction to solve the limited availability of labeled datasets. It is should be noted that the study only tested on red tomatoes which have distinctive color features from the background. The generalizability of the developed approach should be further validated for images with challenging situations. Tassels exhibit varied poses, sizes, and colors in a given image, presenting a difficulty to use global features to characterize them. TasselNet was thus developed to solve this issue in two steps [78]. In the first step, a local CNN regression model was established to predict the number of tassels in each patch of an image. In the second step, the estimated count in each image patch was averaged on individual pixels in that patch, creating a counting map with the same spatial size as the original image. The summation of all pixel intensities in the counting map is the final tassel count in that image. Experimental results showed that TasselNet achieved counting accuracies of from 74.8% to 97.1%, which were 2 to 5 times higher than conventional methods. TasselNet used the patch-based training method which substantially increased the number of images for training. In addition, TasselNet required dot annotation rather than bounding boxes, which also reduced the difficulties in image annotation. A common finding has been identified in all these studies: a moderately complex CNN is recommended because of the potential of model overfitting. This could be a particular concern for regression based counting methods, as its learning target is much

simpler than either image classification or object detection. A major drawback that is also noteworthy is that no location information can be provided by regression based methods, limiting the potential of using these methods for other applications. For classification based methods, plant/organ counting was treated as a discrete counting (or scoring/grading) problem and thus a predefined score or grade (e.g., 10% of inflorescence) were assigned to a given image rather than an exact count.

A representative classification based method is WheatNet, which was developed to predict the percentage of flowering in wheat images [79]. Multiple images were acquired for each plot. A total of 11 classes were annotated for each plot (and thus images for that plot), corresponding to 11 visual scores of percentage heading from 0 to 100% with an interval of 10%. The average prediction of all images in a plot was the final percentage headings for that plot, which reduced counting errors because of inaccurate classification. By fitting the per-plot percentage heading into a sigmoid function, an error of 1.25 days was achieved between the manual and CNN-based measurement of heading date (50% of emerged heads), indicating the high accuracy of the developed method. It should be noted that heading dates estimated using WheatNet counts showed the comparable broad sense heritability ($H^2 = 0.987$) as those estimated using manual counts ($H^2 = 0.982$), indicating the great potential of incorporating DL for plant phenotyping and, therefore, breeding programs and genetics/genomics studies. However, the developed method has a major limitation in that it might be very difficult to generalize the method for other plants with complex canopy structures and flowering patterns such as cotton. Flowers in those plants are usually with in canopies rather than on-top of canopies, presenting more partial or full occlusions. If flowers cannot be imaged, it is not feasible to train any ML/DL model for detection and counting. Additionally, classification based counting methods have the same issue as regression based methods, which cannot provide location information that is actually important to understand plant development.

Detection and Counting in Image Sequences and Videos

Although the aforementioned studies demonstrated that the detection and counting of plants and plant organs can be fairly accurate in still images, a single image is usually not adequate to cover a plant of tree crops (e.g., an apple tree) or an entire plot of row crops. Thus, image sequences and videos should be acquired for those situations, requiring the expansion of detection and counting methods. The key challenge of object detection in image sequences or videos is to associate the same object over different images. There are currently two types of methods to solve this issue: tracking-based methods and reconstruction-based methods.

The key concept of tracking-based methods is to associate detections of the same object over consecutive image sequences or video frames, so individual objects can be tracked to avoid repeated counts (Tracking-based Methods in Figure 2.4). Two types of methods can be used for correspondence estimation.

The first type is based on trajectory information, which can be acquired using sensors such as RTK GPS and IMU devices. A framework was developed to count mangoes for yield estimation [80]. The framework firstly detected mangoes in each still image using a Faster RCNN model. Camera location and pose parameters were collected for each image, so that the geometric correspondence was calculated between pixels in two consecutive images. Thus, it was able to associate and track mango detections from one image to the next image. Experimental results showed that the developed framework achieved an accuracy of 98.6% for mango counting with inexpensive computational cost, demonstrating the efficacy and efficiency of tracking based methods. The developed framework had three limitations. First, the use of positioning devices would increase the cost of the data acquisition system, which could be an issue for small farms and research projects that lack adequate funds. Second, the accuracy of geometric correspondence was dependent on the accuracy of positioning devices, which might be problematic in applications with very tall trees that can block GPS signals.

Last, if fruit samples could be seen from both sides because of relatively open canopies, the developed framework could overestimate the number of fruit counts and thus yield load.

The second type is based on video tracking algorithms. A simple tracking algorithm was developed for sweet pepper counting [81]. Sweet peppers were detected using a Faster RCNN model in all images. In the first image, all detections were initialized as trackers. In the rest of the images, the intersection of union (IOU) and boundary measure (the ratio of the intersection between a tracker and a detection to the area of that detection) metrics were used to quantify the closeness between a detection and a tracker. For a given pair of detection and tracker, if they had an IOU value and a boundary measure that exceeded predetermined thresholds, the detection and tracker would be associated. When sweet peppers moved in or out of images, the IOU and boundary measure metrics become problematic because of the aspect ratio change of the bounding boxes. To avoid this issue, start and stop zones were configured and sweet peppers detected in these zones would not be used for tracking. A small set of image sequences were used to determine the IOU and boundary measure thresholds as well as the start and stop zones. Although the simple tracking algorithm provided an average counting accuracy of 95.9%, it might not be stable because the thresholds could be dramatically different in various datasets. If the testing image sequences and videos were acquired in slightly different conditions, the thresholds might become invalid and result in degraded performance. As a result, the developed algorithm requires a calibration step for finding proper parameters for different datasets. In addition, if fruit objects were highly occluded, the accuracy of detection-tracker association would decrease significantly. To overcome these issues, the optical flow algorithm was used as the tracking framework [82]. The optical flow provided motion information between two consecutive images, so the potential position of each bounding box in the current image can be estimated in the next image. Thus, the detection-tracker association is constrained by the image motion, which improved the association accuracy. The optical flow relies on

some assumptions such as small motion between images and brightness consistency. The first assumption can be satisfied by controlling the data collection movement speed and image (video) acquisition frame rate, whereas the second assumption is relatively easy to be broken. For instance, the optical flow provides degraded performance due to illumination changes, and the illumination change is unavoidable in field conditions [83]. Also, some plant organs (e.g., flowers) are not rigid objects and are affected by wind. When the wind blows, flower shapes can change dramatically, resulting in considerable differences in pixel intensities between images. There are other tracking algorithms such as particle filter, multiple hypotheses tracking (MHT), and probability hypothesis density (PHD), but they have not been used for plants and plant organ counting.

The key concept of reconstruction based methods is to reconstruct a global coordinate system to which objects detected in individual images can be projected (Reconstruction based Methods in Figure 2.4). For 2D reconstruction, global ortho-images were reconstructed by mosaicking image sequences or video frames, so sub-images of an entire crop tree or plot could be extracted from the ortho-images [84, 85]. Subsequently, detection based methods were used to detect and count plants and plant organs in the extracted sub-images. For 3D reconstruction, point clouds were obtained using either image sequences or video frames through photogrammetric algorithms (e.g., the structure from motion, SfM) [86, 87] or additional imaging sensors (e.g., LiDARs) [80]. A transformation relationship was established between 2D images and obtained 3D point clouds, so that objects detected in 2D images can be projected to the 3D space or vice versa. As detections of the same object would be highly overlapped in the 3D space, redundant detections could be eliminated to obtain accurate object quantity. Additionally, 3D reconstruction based methods enabled the extraction of additional objection information such as 3D location and object morphology (e.g., diameter or volume), providing great potential for comprehensive evaluation of plant/organ develop-

ment.

There were several challenges, however, for the 3D reconstruction based methods. First, highly overlapped objects were hard to be accurately detected, leading to inaccurate detection and counting. To overcome this issue, detection and classification based methods were combined [86]. Instead of detecting individual apples, a Faster RCNN model was trained to detect apple clusters, which substantially simplified the problem complexity and improved the detection accuracy. For each detected cluster, a classification based counting method was used to determine the number of apples in that cluster. Although the combination of two strategies dramatically simplified the problem complexity and improved the accuracy, the developed framework was considerably expensive computationally. Also, individual apples could not be projected in 3D space, which lost the possibility of extracting additional phenotypic traits for development characterization. Another issue was the computational cost, especially the SfM technique used to obtain 3D point clouds. The computational complexity of the SfM technique increases quadratically along with the number of used images. Certain environmental factors (e.g., wind) would also result in failure of 3D reconstruction using the SfM. Generally, these are open issues to photogrammetric 3D reconstruction in the field, which become limiting factors for 3D reconstruction based methods as well.

2.4.3 CNNs for Postharvest Quality Assessment

While the plant phenotyping community is primarily focusing on in-season plant performance, postharvest quality is also an important part for plant phenotyping because postharvest properties significantly affect the eventual crop productivity and quality. Based on the nature of the analysis, postharvest quality assessment can be classified as two categories: qualitative assessment and quantitative assessment. Qualitative assessment provides scores/grades for crop fruit, such as defect detection and freshness grading, whereas quantitative assessment provides continuous values for crop postharvest properties such as firmness

and soluble solid content (SSC).

Qualitative assessment of postharvest quality is similar to plant stress phenotyping, with a unique emphasis on fruit rather than plants. Most studies investigated the use of CNNs to detect defects for fruit such as mangosteens [88], cucumbers [89], potatoes [90], blueberries [91], and peaches [92]. These studies reported detection accuracies from 87.85% to 98.6%, which usually were 10% to 20% higher than conventional ML methods, demonstrating the advantages of using CNNs for qualitative assessment of postharvest quality. Although these efforts showed certain success in addressing problems, they had several big limitations. First, because of the availability of labeled data, most studies used very shallow CNN architecture (e.g., one convolutional layer followed by one pooling layer and one fully connected layer), meaning that the potential of CNNs for postharvest quality assessment has not been fully investigated. Even though patch-based training with data augmentation could substantially increase sample sizes, most of the image patches were highly correlated, presenting potential concerns to overfitting problems. Second, as of writing, no studies explored techniques for understanding the mechanism of CNNs, making CNN decisions and high performance unexplained. In addition to defect detection, qualitative assessment of postharvest quality includes crop grading. A CNN based system was developed to grade freshness of packed lettuce [93]. In this system, the CNN was trained to classify each pixel as lettuce, packaging, and artifacts using a small patch (3 by 3 pixels) surrounding that pixel. Experimental results showed that the trained CNN achieved an accuracy of 97.9% for pixel-level classification (equivalent to segmentation). The quality grading with segmentation was comparable with grading using the images of lettuce without packing. This demonstrated the potential of using CNN to segment lettuce for grading without removing packaging, showing the possibility for on-shelf sorting.

Quantitative assessment of postharvest quality (e.g., sugar/acid ratio and bruising) can also be processed using CNNs. A study was conducted to develop a CNN-based regression model for estimating sugar/acid ratio for citrus [94]. Images of excitation-emission matrix (EEM) were used as inputs to train a custom CNN with 8 layers for regression. Sugar/acid-ratio values were estimated with the trained CNN models for 20 testing samples and results showed that CNN-based regression model achieved the lowest prediction error of 2.48, which was 2 to 3 times less than conventional regression models. Another study investigated the use of fully convolutional network (FCN) for segmenting bruised, non-bruised, and calyx end tissues for blueberries [95]. The FCN meta-model was based on a VGG-16 network. Experimental results showed that the developed approach provided the segmentation accuracies of 73.4% to 81.2%, which was substantially higher than the SVM based segmentation method (46.6%). A partial reason was that the spectra of the calyx end were similar to bruised and non-bruised tissues, and thus conventional classifiers struggled to accurately segment them. In contrast, CNN-based approaches could learn other features such as shape and position, which significantly contributed to the improvement of the segmentation accuracy of the calyx end. This study was the only case using an end-to-end CNN meta-model for postharvest quality assessment, providing valuable experiences for future studies. However, there were several issues in the study. First, hyperspectral images have many more channels than RGB images, leading to an issue of transfer learning. In the study, an additional layer was developed to reduce the dimensionality of raw hyperspectral images from an arbitrary value to 3, so that weights pretrained on other datasets can be used for the bruising dataset. However, experimental results showed FCN models trained using transfer learning were less accurate than those trained from scratch. The authors stated that this was primarily because of the difference between the bruising dataset and the ImageNet dataset, meaning that the majority of learned filters from the ImageNet are not useful for bruising detection. This raises the critical question of how large publicly available datasets can benefit postharvest quality

assessment studies that rely more on advanced imaging modalities (e.g., multispectral and hyperspectral imaging) than RGB imaging.

2.5 Discussion

2.5.1 Data Availability

The availability of diverse annotated datasets is a key factor for all DL related studies. Adequate annotated datasets enable and ensure the fast development and evolution of DL methods. This generally holds true for domain applications such as plant phenotyping. For biotic/abiotic stress phenotyping, data annotation is relatively straightforward, resulting in several large publicly available datasets such as PlantVillage. For plant development, such as the sensing technologies are under development, not many datasets are publicly available, not to mention the lack annotated datasets. As the main concept of DL is to learn features from data, it is really hard to develop (or even use) DL techniques without sufficient annotated data. In particular, data annotation has several challenges for plant phenotyping. First, data annotation sometimes requires domain expertise. For instance, it is easy to label cars, whereas it is difficult to label particular plant diseases because of the need of domain expertise. Thus, it is not easy to crowdsource annotation tasks, which limits the efficiency and throughput of data labeling. Second, unlike common uses, plant phenotyping oftentimes relies on advanced imaging techniques such as thermal and hyperspectral imaging. Labeling of those data is dramatically more difficult than labeling color images because of fewer visual cues. Third, there are many phenotyping applications that require object detection and segmentation (semantic, instance, or panoptic), and these applications require instance-level (bounding boxes) and pixel-level (masks) annotations. Those are very time-consuming tasks and become the major limiting factor for using DL in plant phenotyping. Some of the challenges are common for general computer vision tasks, and researchers have proposed and developed some solutions. Transfer learning is one of the most important techniques for significantly reducing the requirements of labeled data. Transfer learning relies on the

assumption that a very large dataset ensures that the learned filters are common for other datasets. Thus, for some domain applications with limited labeled data, transfer learning could significantly improve the training efficiency and accuracy. However, the key challenge is whether phenotyping datasets are similar to very large common datasets (e.g., ImageNet or MS COCO), especially for some phenotyping applications using advanced imaging techniques (e.g., hyperspectral imaging). Active learning is another effort for the reduction of the cost of data labelling. Compared with conventional data annotation, active learning aims to find and label samples that maximize model performance. Thus, the majority of samples are ignored in labeling, thus saving time. Crowdsourcing is also a viable way for data annotation, which requires less investment in labor cost. Some studies have demonstrated the capability of using crowdsourcing for quickly labelling a large image datasets for machine learning applications [96]. In particular, there are some commercial services for crowdsourcing labelling such as Amazon Mechanical Turk and CrowdFlower. Through those services, a reasonable quality and throughput can be ensured for data annotation.

2.5.2 Adoption of DL Methods for Plant Phenotyping

Another important consideration is the adoption of DL methods for plant phenotyping. Technology companies have released various DL framework that accelerate the development and implementation of new DL algorithms. In particular, the DL community encourages researchers to share source codes of original studies to facilitate other research projects. These efforts considerably ease the adoption of latest DL methods for domain applications such as plant phenotyping. However, there is still a delay in the use of the latest technologies for plant phenotyping. This likely occurs for three reasons. First, some of the latest DL methods require a significant investment in computational power, which cannot easily be achieved by ordinary research labs. Second, original DL solutions might not be directly usable for plant phenotyping applications. Additional efforts are necessary to adopt those advanced DL

solutions, and sometimes these efforts are time-consuming. Thirdly, large private companies, who invest heavily in plant phenotyping, do not divulge their research efforts in this area.

2.5.3 CNNs for 3D Imaging Processing

3D imaging, an important imaging technique, has not been mentioned yet. An important plant phenotyping task is to characterize and understand plant morphology. While few studies reported the use of CNN in a scenario with 3D imaging, they have primarily focused on the detection in 2D images and projected the detections in 3D for processing, such as removal of redundant detections and determinations of detections with occlusions. None of them really utilized CNNs for plant morphology characterization and understanding. In particular, 3D point clouds can be collected using various approaches in plant phenotyping applications, and most of them need to be processed using conventional 3D processing methods. One possible reason is that even the DL community has not delivered many reliable tools for 3D point cloud processing. PointNet and PointNet++ are pioneering work for processing 3D point clouds, but they are limited to the number of points in each model (a couple of thousands points) [97, 98]. If the point cloud is too large, there is no efficient computational solution for the network training and inference. Thus, many of 3D imaging work requires the technical development from the DL community.

2.6 Conclusions

In this review, CNNs and CNN meta-models for image-based plant phenotyping were comprehensively reviewed to provide advantages and disadvantages of using them for different tasks of plant phenotyping. Through these studies, CNNs and CNN meta-models demonstrated the great potential for solving the most challenging problems encountered in various plant phenotyping applications. In particular, some end-to-end CNN meta-models significantly streamlined the process of extracting phenotypic traits from images. This would

enable the improvement of data processing and ultimately plant phenotyping applications.

Several future research directions that use CNNs and CNN meta-models for plant phenotyping are identified. First direction is to develop efficient labeling strategies or active learning methods, so that the availability of labeled datasets can be significantly increased to meet the general requirements of training CNNs and CNN meta-models. This holds true especially for postharvest quality assessment that utilizes different imaging modalities and has limited samples. The second direction is to customize a deep learning framework that can facilitate the adoption of latest DL techniques for plant phenotyping applications. Such a framework can provide common interface for algorithm integration, so that newly developed models and tools can be added for use with little or no development effort, such as visualization tools for model explanation and reinforcement learning for model improvement. The third direction is to adopt and develop CNNs and CNN meta-models for direct 3D processing, especially skeleton extraction, branch-pattern classification, and plant-development understanding.

Chapter 3

GPhenoVision: A Ground Mobile System with Multi-modal Imaging for Field-Based High Throughput Phenotyping of Cotton[1]

[1]Jiang, Y., Li, C., Robertson, J. S., Sun, S., Xu, R., and Paterson, A. H. 2018. *Scientific Reports*, 8. Reprinted here with permission of publisher.

3.1 Abstract

Imaging sensors can extend phenotyping capability, but they require a system to handle high-volume data. The overall goal of this study was to develop and evaluate a field-based high throughput phenotyping system accommodating high-resolution imagers. The system consisted of a high-clearance tractor with sensing and electrical systems. The sensing system was based on a distributed structure, integrating environmental sensors, real-time kinematic GPS, and multiple imaging sensors including RGB-D, thermal, and hyperspectral cameras. Custom software with a multilayered architecture was developed using LabVIEW for system control and data collection. The system was evaluated by scanning a cotton field with 23 genotypes for quantification of canopy growth and development. A data processing pipeline was developed to extract phenotypes at the canopy level, including height, width, projected leaf area, and volume from RGB-D data and temperature from thermal images. Growth rates of morphological traits were accordingly calculated. The traits had strong correlations (r = 0.54–0.74) with fiber yield and good broad sense heritability ($H^2 = 0.27$–0.72), suggesting the potential for conducting quantitative genetic analysis and contributing to yield prediction models. The developed system could be a useful tool for a wide range of breeding/genetic, agronomic/physiological, and economic studies.

3.2 Introduction

Agriculture is facing tremendous challenges from the rapidly growing population that demands more food, feed, fiber, and fuel, as well as from the changing climate and severe shortfall of arable land and water resources [6]. To overcome these challenges, it is necessary to select and cultivate new crop genotypes with high yield and quality while using a reduced amount of natural resources such as water. Cotton (*Gossypium*) is the most important source of natural fiber, and in recent years it has also become an important source of food and feed (e.g. cottonseed oil for humans and hulls for livestock) [99]. Consequently, improving

cotton production and quality is crucial to fulfilling the fiber and food requirement of over nine billion people by 2050 [3]. Genetic/genomic research and breeding programs hold great potential to double the current production of cotton. Two key factors have been recognized for these programs: development of diversity panels and evaluation of phenotypic traits [7]. During past decades, advances in genetic technologies paved the way for genetic analysis of large crop populations. In particular, high-throughput sequencing techniques have enabled rapid and inexpensive genotyping of crop plants to provide thousands of recombinant lines for genetic selection and genomics studies. However, current phenotyping primarily relies on manual measurements and observations, and is far behind genotyping in terms of throughput, accuracy, and repeatability. This limits the potential use of crop genotypes in genotype-phenotype mapping and characterizing genotype-environment interactions [100]. High throughput phenotyping (HTP) has been recognized as an essential part of a new 'Green Revolution' to further improve crop yield and quality as well as to better understand crop genomics [100].

In the past decade, greenhouse- and chamber-based high throughput phenotyping systems have been developed by several transnational companies, public institutions, and universities [101]. These systems are fully automated and can accurately measure phenotypic traits of individual plants. Many studies have demonstrated successful use of HTP systems to reveal relationships between genotypes and phenotypes [102, 103, 104, 105]. These studies primarily focused on crop resistance and/or tolerance to stress and nutrient-deficiency, and the experiments took place in greenhouses, where environments were artificially controlled to simultaneously provide consistent ambient condition (e.g. illumination) for data collection and minimize confounding effects (spatial heterogeneity) on experiments. In addition, automation subsystems could efficiently and precisely implement treatments (e.g. irrigation and nutrient spray) for experiments, and provide accurate environment and treatment records (irrigated water and nutrient amount) for data analyses. Although the greenhouse-

based HTP systems provide the aforementioned beneficial features, they have three major limitations. First, system development and maintenance are usually prohibitively expensive for small breeding programs. Second, most greenhouse- and chamber-based systems can only handle up to several hundred potted plants (or plant wells) in an experiment, so the maximum population size is far too small for breeding programs. Third, and most importantly, quantitative trait loci or candidate genes identified in controlled environments may be less effective in field conditions [7]. Plant growth and development can be significantly affected by environmental factors such as soil, and these factors are extremely difficult (or sometimes impossible) to simulate in controlled environments.

To address these issues, it is imperative to develop systems for field-based high throughput phenotyping (FB-HTP). In consideration of system cost, utility, and spatial and temporal resolution, FB-HTP systems based on high-clearance ground vehicles are currently preferred for breeding programs of moderate scale (up to a few thousand plots) [106, 107, 108]. Several representative systems have been developed recently [109, 110, 111, 112, 113, 114, 115]. Some of these systems [109, 110, 111] have demonstrated usefulness in breeding programs and genomics studies [116, 117, 118]. It is noteworthy that imaging techniques were frequently applied in recent systems due to their ample capacity for extracting complex traits. The imaging techniques included conventional RGB, thermal, spectral, and 3D imaging modules. Integration of multimodal imaging sensors would significantly improve the sensing capability of an FB-HTP system, because a multimodal imaging system cannot only measure traits from a single imaging module, but also provide a set of traits from different imaging modules. For instance, RGB and 3D imagers can be integrated to provide both color and position information for plants or plant organs. Generally, the data volume of imaging modules is high, thus presenting a technical challenge in data acquisition. In particular, the continuous scanning mode is preferable for high throughput data collection in the field, which further increases the demand in a data acquisition (DAQ) system dedicated for field-based phe-

notyping systems. There are three features that need to be included in the DAQ system: high-throughput, customizability, and modularity. High-throughput means that the DAQ system needs to handle high-volume data generated by imaging sensors. Customizability allows researchers to quickly develop and upgrade an integrative FB-HTP with various imaging modules for different phenotyping purposes. Modularity splits individual modules into independent working environments, which prevents entire-system malfunction and reduces the amount of effort required in development, when adding or removing imaging modules.

The overall goal of the research described in this paper was to develop and evaluate a modular and customizable ground mobile system using multiple imaging modalities for FB-HTP. Specific objectives were to: 1) develop a modular and customizable ground mobile system integrated with multiple high resolution imagery modules including RGB-D, thermal, and hyperspectral cameras, 2) calibrate and validate the sensing system, and 3) evaluate the usefulness of the FB-HTP system for breeding programs and genomics studies.

3.3 Development and Validation of the GPhenoVision System

3.3.1 Platform mechanical design and implementation

The 'GPhenoVision' system consists of five components including a platform, mechanical structures, sensing and electrical systems, and data acquisition software. A high-clearance tractor (Spider DP, LeeAgra Inc. Lubbock, TX, USA) was the platform upon which other system components were integrated (Figure 3.1a). In the current study, the wheel track and clearance of the platform were set as 1.83 m and 1.5 m, respectively, and narrow wheels were selected to minimize mechanical disturbance. An aluminum frame (width of 1.52 m) with three polyethylene tarpaulins covering the top and two sides was built as an enclosure to provide a shaded area. The platform engine and a secondary alternator were installed in the back of the platform by default, so the enclosure was attached to the front of the

tractor to reduce engine-induced variations. The enclosure covered plants for imaging to reduce ambient interferences such as strong sunlight and wind effects. An adjustable sensor frame was welded in front of the tractor and covered by the enclosure, the height of which could vary from 1.2 m to 2.4 m above the ground. Imaging sensors were installed on the sensor frame in custom camera holders, and a rubber cushion was added between metal frames to further reduce high-frequency vibrations that could be potentially transferred to sensors. This vibration reduction helped to decrease the possibility of acquiring blurry images. The installation positions of sensors were predetermined so that acquired images could be geo-referenced. On top of the enclosure and tractor cab, there were two sensor bases on which environmental sensors and positioning devices were mounted. In the middle of the platform, a metal frame was fabricated to hold instruments of the electrical system. Passenger seats were installed on each side of the tractor for testing purposes. In the testing stage, additional operators were on the tractor to test the image acquisition software during field data collection.

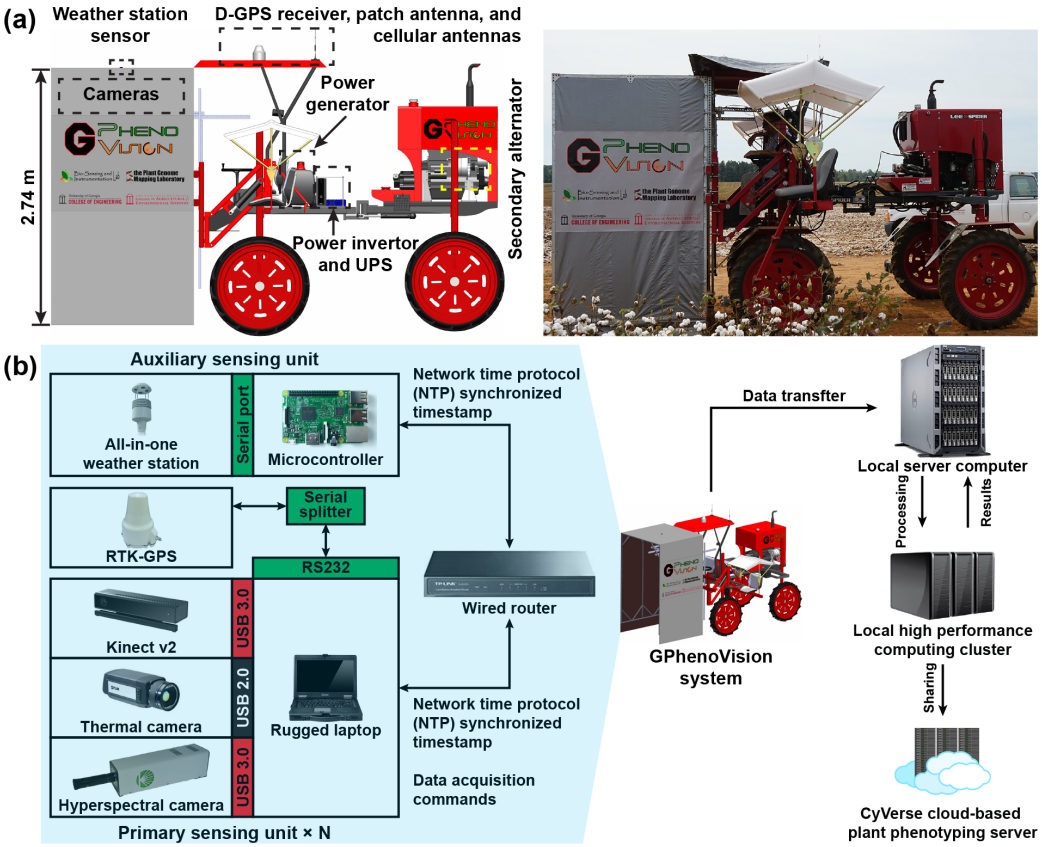

Figure 3.1: (a) Diagram and picture of the GPhenoVision system design and (b) architecture of data flow and management. N is the number of the primary sensing unit(s) integrated to the GPhenoVision system; N = 1 in the current study.

3.3.2 Sensing system

The sensing system contained two subsystems: a DAQ system and local computing resources (Figure 3.1b). The DAQ system consisted of a primary sensing unit, a positioning unit, and an auxiliary sensing unit, and was integrated into the platform for data collection in field conditions (left block in Figure 3.1b). In the primary sensing unit, an RGB-D camera (Kinect for Windows v2, Microsoft, Redmond, WA, USA), a thermal camera (A655sc, FLIR Systems Inc., Wilsonville, OR, USA), and a hyperspectral camera (MRC-923-001, Middleton Spectral Vision, Middleton, WI, USA) were connected to a rugged laptop (S400, Getac Technology Corporation, Taipei, Taiwan) that was equipped with a solid state drive (SSD; 850 Pro 1TB,

Samsung Electronics, Suwon, South Korea). The laptop used an Intel i7-4712 MQ CPU with 8 GB of RAM and Windows 10 Pro operating system. A serial port (RS232) of the laptop was utilized to receive GPS data from the positioning unit based on a real time kinematic GPS (RTK-GPS; Cruizer II, Raven Industries Inc., Sioux Falls, SD, USA). Based on each camera's field of view (FOV) and scanning mode, the sampling frequencies of the RGB-D and thermal cameras were 6 frames per second (FPS), whereas the hyperspectral cameras were 100 FPS. All cameras used global shutter to avoid the Wobble effect during data collection. Accordingly, data volume of the primary sensing unit was estimated (Table 3.1). In the auxiliary sensing unit, a single-board computer (Raspberry 3 Model B, Raspberry Pi Foundation, Cambridge, UK) was used to receive and save ambient information (air temperature, relative humidity, and pressure) by controlling a microcontroller (Arduino Uno, Arduino, Vancouver, BC, Canada) that regulated an environmental sensor (BME280, Bosch Sensortec, Gerlingen, Germany). The primary and auxiliary sensing units were wired to a router for communication of synchronized timestamps and DAQ commands.

Table 3.1: Key specification and data volume of the cameras used in the primary sensing unit

	RGB-D camera	Thermal camera	Hyperspectral camera
Manufacturer	Microsoft	FLIR	Middleton Spectral Vision
Model	Kinect for Windows v2	A655sc	MRC-923-001
Scanning mode	Area scan	Area scan	Line scan
Field of view	$70° \times 60°$	$80° \times 64.4°$	55.9°
Data type	16-bit unsigned integer (depth) 32-bit unsigned integer (color)	32-bit single precision	16-bit unsigned integer
Spectral range	Visible (color), 827–850 nm (depth)	7.5–14 μm	400–1000 nm
Image resolution	512×424 (depth) 1920×1080 (color)	640×480	$640(spatial) \times 236(spectral)$
Image size	0.42 MB (depth) 8 MB (color)	1.18 MB	0.29 MB
Designed frame rate (maximum)	6 (30) FPS (depth) 6 (30) FPS (color)	6 (30) FPS	100 (200) FPS
Data volume (maximum)	2.52 (12.6) MB/s (depth) 48 (240) MB/s (color)	7.08 (35.4) MB/s	29 (58) MB/s
Total data volume per second		86.6 (346) MB/s	

After field data collections, data could be transferred to local and remote computing resources for data processing and sharing (right block in Figure 3.1b). A server computer was equipped with a redundant array of independent disks (RAID 1) of 4 terabytes for storing data generated in one growth season. The server computer was also used to perform algorithms for extracting phenotypic traits.

3.3.3 Electrical system

The electrical system was based on the secondary alternator (200 amps) of the tractor platform, an absorbed glass mat (AGM) battery (Yellowtop D34, OPTIMA Batteries, Inc., Milwaukee, WI, USA), a power inverter (APS1000-12, Power Bright, Fort Lauderdale, FL, USA), and an uninterruptible power supply (UPS; Back-UPS Pro 1500, APC by Schneider

Electric, West Kingston, RI, USA). When the system ran, the alternator continuously generated and output power to the power inverter through the battery. The inverter boosted the input power from 12-volt direct-current (DC) to 120-volt alternating current (AC) and transferred to the UPS to provide stable AC output for the sensing system. A 2-kilowatt power generator (EU2000i, Honda Power Equipment, Alpharetta, GA, USA) was chosen as an alternative power generating source in case the secondary alternator malfunctioned. In total, the electrical system using either the secondary alternator or the power generator could provide AC output of 865 watts with the UPS protection, which was adequate to power the sensing system used in GPhenoVision (see Supplementary Table S1).

3.3.4 Data acquisition software

Custom computer software was developed using LabVIEW 2015 (National Instruments, Austin, TX, USA) and deployed on the rugged laptop to control the positioning and primary sensing units for data acquisition. The software was based on a multilayered architecture; each layer worked independently but the layers could be synchronized with user commands (Figure 3.2a). The graphical user interface (GUI) received user commands and transferred them to the sensor control layer for processing. The GUI also worked with the data display layer to present current image frames to users. The sensor control layer utilized multiple threads, with each thread controlling a single sensor through an event-driven finite state machine (EFSM) (Figure 3.2b). Due to the multithread design, the GPS and the RGB-D, thermal, and hyperspectral cameras could acquire data using different sampling frequencies. Additionally, although all sensors shared the same EFSM, sensor initialization, finalization, image preview, and acquisition relied on specific drivers or software development kits (SDKs) for each sensor. In the present study, the RGB-D camera required its driver to be provided by the manufacturer and a third-party LabVIEW library (Haro3D, HaroTek LLC, Keller, TX, USA), whereas the thermal and spectral cameras required only their official SDKs. The serial communication virtual instruments (VIs) provided by LabVIEW were used to receive

data from the GPS. As the total data volume could be up to 346 MB/s, it was necessary to design a data cache layer, which increased the performance of data transferring and storage by buffering data from sensors in the computer memory. The data transfer layer periodically inquires the data cache layer and writes buffered data back to the SSD hard drive.

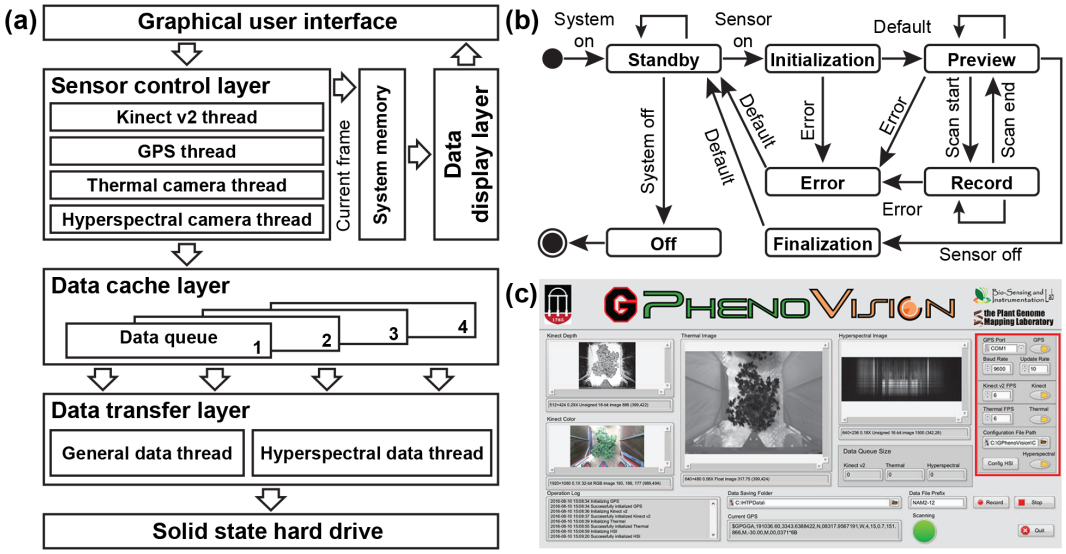

Figure 3.2: (a) Design of the data acquisition (DAQ) software for the GPhenoVision system: (a) software architecture of the DAQ program; (b) the event-driven finite state machine (EFSM) developed for sensor threads in the sensor control layer; and (c) the front panel of the custom DAQ program.

The software front panel showed data (GPS, depth, color, thermal, and hyperspectral images), memory cache status, and operation log in real time during field data collection (Figure 3.2c). This allowed operators to quickly read height or temperature of plants from raw images or investigate potential issues such as failure of certain sensors. Sensors could be individually configured and controlled through sensor control panels (indicated with red rectangle in Figure 3.2c). The collected images were saved in a folder specified by the operator. After configuring and turning on certain sensors, an operator could start and stop

data collections by using the 'Record' and 'Stop' buttons, accordingly.

In addition to the aforementioned computer software, a custom Python (Python 2.7) script was developed and deployed on the single-board computer to acquire air temperature, relative humidity, and pressure with acquisition timestamps under 1 Hz. Timestamps of the primary and auxiliary sensing units were synchronized through a network time protocol (NTP) service that was installed on the rugged laptop. Due to its small amount of data (35 bytes per second), the auxiliary sensing unit was controlled to start/stop by a physical button, and continuously recorded data during a whole data collection session without being interrupted by commands from the DAQ software controlling the primary sensing unit.

3.3.5 System calibration and validation

Calibration and validation methods

Prior to using the system, it was important to calibrate and validate sensors in the primary sensing unit because they would provide essential data for extracting phenotypic traits. The three cameras were used to acquire images for different types of traits, and thus they had various aspects to be calibrated or validated 3.2. For the Kinect v2 camera, depth accuracy is a key factor because it determines the accuracy of each point in a colored point cloud retrieved from the sensor. Although the depth accuracy of Kinect v2 camera has been proven in both laboratory and field conditions [119, 120], it was necessary to re-validate sensor measurements as the camera was integrated into a new system. To account for ambient effects such as wind, depth accuracy of the Kinect v2 camera was evaluated by measuring plant height in field conditions. The GPhenoVision system scanned 100 cotton plants on three different days while recording wind speed information with a portable anemometer (HYELEC MS6252B, Huayi Electronics Industrial Ltd., Hangzhou, Zhejiang, China). Plant heights were extracted from depth images by using the method described in a previous study [120], and also manually measured on each day for reference. The system and manual

measurements were compared using simple linear regressions, and the adjusted R^2 and root mean squared error (RMSE) were used as accuracy indicators.

Table 3.2: Summary of calibration and validation experiments for the RGB-D, thermal, and hyperspectral cameras used in the GPhenoVision system.

Sensor	Test parameter	Test location	Test objective
RGB-D camera	Depth accuracy	Field	Validate the accuracy of sensor measurements of morphological traits
Thermal camera	Measurement accuracy (compared with thermocouple measurements) for an object with various temperatures	Laboratory	Validate the measurement accuracy in a control environment to provide an accuracy baseline
Thermal camera	Measurement accuracy (compared with thermocouple measurements) for an object at different distance to the camera	Field	Validate the repeatability of measurements due to the change of object-to-camera distance
Thermal camera	Temperature differences between thermal camera measurements for plants under shaded and unshaded areas	Field	Verify the absence of the shading effect to canopy temperature measurements
Hyperspectral camera	Accuracy of the regression model between pixel locations on the spectral dimension and standard wavelengths	Laboratory	Calibrate and validate the relationship between pixel locations on the spectral dimension and wavelengths in a scan line
Hyperspectral camera	Maximum and minimum spatial resolution	Field	Calibrate and validate the spatial resolution of the camera under various object-to-camera distances

Three experiments were conducted to validate measurement accuracy of the thermal camera under various conditions. The first experiment was to provide a baseline of the thermal camera accuracy by measuring surface temperature of a blackbody (an object with emissivity of 1; 4-inch Blackbody Source, FLIR Systems Inc., Wilsonville, OR, USA) in laboratory conditions. The thermal camera and blackbody were placed on the same plane separated by a distance of 2.4 m (the maximum distance between objects and the thermal camera on the system), and a T-type thermocouple was attached to a corner of the blackbody surface to measure its temperature for reference. The camera and thermocouple measured temperatures synchronously. A total of 26 surface temperatures ranging from $24°C$ to $69°C$ were set to the blackbody by adjusting its powering voltage, and each temperature was measured by both the thermal camera and thermocouple for 200 frames; therefore, both the thermal camera and thermocouple were used to collect 200 data points. The mean value of the 200 data points collected by each sensor was calculated as the sensor measurement at each known blackbody temperature. The mean values obtained from the thermal camera and thermocouple were compared using linear regressions, and the adjusted R^2 and RMSE values were used for accuracy evaluation.

Although the first experiment provided the baseline of thermal camera accuracy, the baseline was obtained in laboratory conditions that were different from field conditions (practical data collection conditions). The second experiment was to test whether the thermal camera could repeatedly provide accurate measurements of objects at different distances from the camera in field conditions. The thermal camera was oriented towards nadir, installed on the GPhenoVision system, and raised to the highest position (2.4 m above the ground). The blackbody was set to a constant voltage (therefore a constant surface temperature in theory), and placed on a tripod that was in the center of the camera's FOV. The blackbody was raised from 0 to 1.5 m above the ground with an interval of 0.1 m, and at each height, its surface temperature was measured for 200 frames (therefore yielding 200 data points) by

both the thermal camera and thermocouple. Since blackbody temperature could vary due to fluctuations of ambient conditions, the thermal camera measurements were compensated by using Equation 3.1. Analysis of variance (ANOVA) test was conducted on the compensated measurements to identify any existence of statistical differences among temperatures measured at different heights.

$$TI_{h,i}^{compensated} = TI_{h,i}^{raw} + TC_{h,i} - TC_{h,1}, \quad h = 0, 0.1, ..., 1.5, \quad i = 1, 2, ..., 200 \tag{3.1}$$

where TI and TC represented temperatures measured by the thermal camera and thermocouple, respectively; h indicated heights of blackbody; and i was the number of frames (or data points) at a certain height.

During data collection, plants would be shaded by the enclosure and this might raise concern regarding shading effects on temperature measurements. The last experiment was to test any shading effect on measurements of canopy temperature in field conditions. In the experiment, only half of the enclosure was covered so that the thermal camera could acquire images of plants under both shaded and unshaded conditions. The GPhenoVision system with the modified enclosure was used to scan 30 plants on three days (14 July 2016, 22 July 2016, and 26 July 2016). A total of 90 pairs of images were collected, each containing the same single plant that appeared in both the shaded and unshaded areas under the modified enclosure. The plant in each image was masked by a thresholding method to calculate its mean canopy temperature. Differences in temperatures measured under the two shading conditions were computed, and the mean value and standard deviation of the temperature differences were used to evaluate shading effects. During data collection, shading would be considered to have no effect on canopy temperature measurements if a mean difference of zero was achieved with a standard deviation of less than $0.5°C$.

The hyperspectral camera was calibrated in spectral and spatial dimensions, respectively. A method proposed in a previous study [121] was used to calibrate the spectral dimension. To obtain a more accurate regression, three calibration lamps were used including a Krypton lamp (Model 6031, Oriel Instruments, Stratford, CT, USA), a Xenon lamp (Model 6033, Oriel Instruments, Stratford, CT, USA), and an Hg (Ar) lamp (Model 6035, Oriel Instruments, Stratford, CT, USA). Additionally, the spatial resolution was calculated and tested using two test targets that were printed on letter-size (215.9 mm $\times$ 279.4 mm) bright white papers. The first target contained two resolution patterns of 1/8 and 1/6 line pairs per millimeter (LP/mm) with each of 15 pairs, and the second target contained three resolution patterns of 1/20, 1/15, and 1/10 LP/mm with each of 5 pairs. The targets were placed at three heights to evaluate spatial resolutions under the worst (targets were on the ground), common (targets were 1 m above the ground), and best (targets were 1.5 m above the ground) conditions. The spatial resolution under each condition was calculated as a ratio of the target length (279.4 mm) to the number of pixels representing the target in an image. Subsequently, the calculated spatial resolutions were validated by using the resolution patterns. If a resolution pattern was correctly identified in an image, two image pixels at least should be assigned to any line pair in the pattern with one pixel for the black stripe and another for the white. As a result, the spatial resolution was confirmed by the resolution patterns using Equation 3.2.

$$\begin{cases} SR \leq \frac{1}{2\times r}, & \text{if pattern was correctly identified} \\ SR > \frac{1}{2\times r}, & \text{otherwise} \end{cases} \tag{3.2}$$

where SR is the camera spatial resolution in mm/pixel and r is the pattern resolution in LP/mm.

Calibration and validation results

Overall, the system and manual measurements of plant height were strongly correlated (*adjusted* $R^2 > 0.99$) with an RMSE of 0.034 m for data on all three days (Figure 3.3a).

Compared with results from another study [120], both $adjusted R^2$ and RMSE were slightly improved, indicating that the GPhenoVision system could accurately measure plant height in field conditions. Additionally, these results were stably obtained when using data on individual days (see Supplementary Table S2), which meant that the system had high repeatability to accurately measure plant height under various conditions.

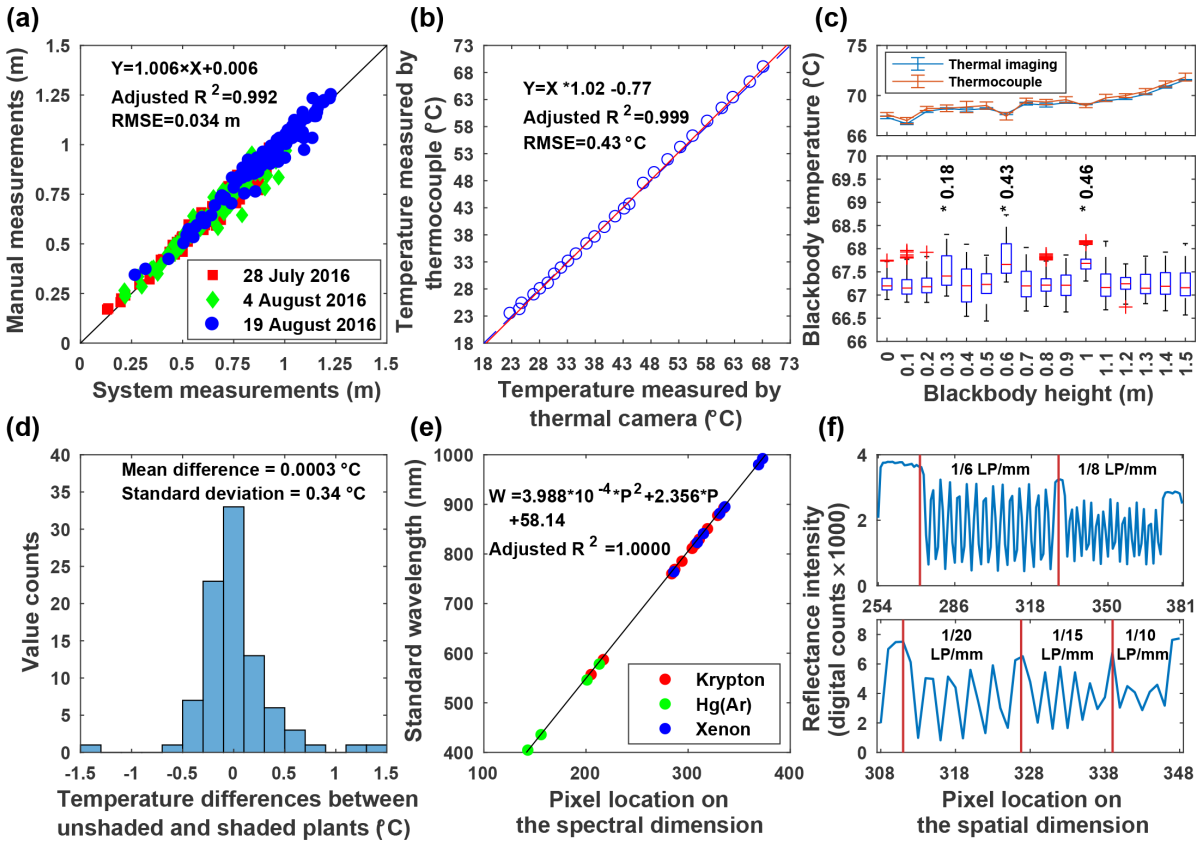

Figure 3.3: Calibration and validation results of three sensors used in the primary sensing unit. (a) Depth accuracy of the RGB-D camera. (b) Regression between blackbody temperatures measured using thermal camera and thermocouple in laboratory conditions. (c) Raw (top) and compensated (bottom) blackbody temperatures measured using thermal camera in field conditions when placing blackbody at various heights in the sensing range. (d) Canopy temperatures measured under shaded and unshaded conditions. (e) Regression between wavelengths and pixel locations on the spectral dimension in images. (f) Spatial resolution of the hyperspectral camera under the best (top) and worst (bottom) situations.

For the blackbody surface temperature, thermal camera and thermocouple measurements were highly correlated, with an RMSE of $0.43°C$, and thus thermal camera measurements were as accurate as traditional contact measurement methods (Figure 3.3b). Consequently, thermal images acquired by the camera could be directly used for extracting canopy temperature by configuring a proper emissivity. The emissivity can be set from 0.93 to 0.99

depending on crops to be studied [122]. A value of 0.96 is commonly used as cotton plant emissivity [123] and has been used for field data collection in the present study. In the second experiment, although the blackbody was set at a constant voltage, its surface temperatures measured by raw thermal images and thermocouple varied when the blackbody was placed at different heights (top chart in Figure 3.3c). This indicated that the temperature variations were primarily due to fluctuations of ambient conditions. After compensation, thermal imaging measurements showed no statistical difference with thermocouple measurements when the blackbody was placed at various heights (bottom chart in Figure 3.3b). Exceptions occurred at the heights of 0.3, 0.6, and 1 m above the ground, but the temperature differences at these three heights were less than the nominal camera accuracy ($0.5°C$). As a result, temperature measurements with the thermal camera were consistently accurate as long as plants were within the system measurement range (0 to 1.5 m). The mean difference in canopy temperatures between plants under shaded and unshaded areas was negligible, because the actual difference ($3 \times 10^{-4°}C$) was smaller than the nominal camera sensitivity ($0.05°C$). Additionally, the standard deviation was less than the camera's nominal accuracy, which meant that there was no shading effect on canopy temperature measurements (Figure 3.3d). This was probably because canopy temperature would not immediately decrease when plants were shaded for only one or two seconds during data collection. Therefore, images acquired by the thermal camera of the GPhenoVision system could be used to accurately extract physiological traits such as canopy temperature.

Following a previously developed method [121], no "keystone" or "smile" distortion was observed in the spatial and spectral dimensions (see Supplementary Figure S1), and thus correction of distortion was not needed for the hyperspectral camera. In the spectral calibration, 22 representative wavelengths from 404.66 nm to 992.3 nm were identified and used to establish the regression equation between pixel locations on the spectral dimension and wavelengths (Figure 3.3e). The regression equation was satisfactory because the standard

and calculated wavelengths were strongly correlated (*adjusted* $R^2 = 1$). The RMSE (less than 1 nm) was also acceptable because the spectrograph has a nominal spectral resolution of 2.7 nm. In the spatial calibration, the length (279.4 mm) of the two targets was recognized as 41, 74, and 128 pixels in hyperspectral images acquired at the worst (on the ground), common (1 m above the ground), and best (1.5 m above the ground) conditions. Thus, the spatial resolutions of the hyperspectral camera ranged from 2.2 to 6.8 mm/pixel and was 3.8 mm/pixel under the common condition. The worst and best resolutions were validated by the test patterns on the targets. In the worst condition, resolutions of lower than 1/15 LP/mm were successfully identified as 5 pairs, whereas resolutions of 1/10 LP/mm were not correctly recognized (bottom chart in Figure 3.3f). Based on Equation 3.2, this confirmed that the spatial resolution (6.8 mm/pixel) was better than 7.5 mm/pixel but worse than 5 mm/pixel. The hyperspectral camera successfully resolved all patterns in the best condition, which agreed with the calculated spatial resolution of 2.2 mm/pixel (better than 3 mm/pixel; top chart in Figure 3.3f). Nonetheless, even the worst spatial resolution would be 6.8 mm/pixel, which was smaller than the size of a cotton leaf. Thus, the hyperspectral camera of the GPhenoVision system has an adequate spatial resolution for many elements of phenotyping.

3.4 Methods

3.4.1 Experimental Design and Field Data Collection

The GPhenoVision system is intended to be eventually used for experiments involving thousands of lines, so it was necessary to evaluate system capability in a phenotyping scenario. As the entire project was at a fledgling stage, the system was evaluated in a small-scale experiment with a special field layout. A field (33.727458 N, -83.299273 W) contained 132 plants (11 plants per row $\times$ 12 rows), and used a single plant layout (SPL) where individual plots had in-row and across-row width of 1.52 m (see Supplementary Figure S4). The 'SPL' field was planted on 25 May 2016 with each plot comprising a single plant sampling 23

different genotypes. The genotypes belonged to three groups: elite *Gossypium hirsutum* (*G. hirsutum*), *Gossypium barbadense* (*G. barbadense*, represented by a single elite cultivar) , and exotic *G. hirsutum* genotypes, which include wild and elite cottons not adapted to the study area.

The field was scanned using the system on five days in 2016 including 14 July (51 days after planting, DAP 51), 28 July (DAP 65), 4 August (DAP 72), 19 August (DAP 87), and 26 August (DAP 94). This covered two cotton growth stages: 1) canopy development (DAP 30–80) and 2) flower and boll development (DAP 80–120). No irrigation was scheduled during the data collection period. To obtain daily maximum canopy surface temperature, data collection was strictly conducted in a period from 1200 to 1500 hours on each day (see Supplementary Table S7). Although it could collect data with all imaging modules (see Supplementary Figure S5 for representative images from all imaging modules), the GPhenoVision system in the present study, ran at a constant speed of 1 m/s, with depth, color, and thermal modules in a continuous scanning mode. To save storage space, the operator manually controlled the DAQ software to start/stop saving images at the beginning/end of each row. Cotton fiber was manually harvested and weighed for individual plants on 4 November 2016 (DAP 163). Although a precipitation sensor was not available on the farm during the data collection period, precipitation information was obtained from a public service (Weather Underground, The Weather Company, Atlanta, GA, USA) based on sensors 11 km away from the SPL field.

3.4.2 Extraction of Phenotypic Traits

After all plants germinated, positions of individual plants (defined as the position of plant main stem) were surveyed using an RTK-GPS (Cruizer II, Raven Industries Inc., Sioux Falls, SD, USA). Each of the collected images (depth, color, and thermal images) had a corresponding GPS record, and this record indicated the image acquisition position representing the image center. In the configuration (field layout and camera installation position)

of the present study, a depth, color, or thermal image captured a single plant with little or no coverage of adjacent plants if that plant was located close to the image center (image acquisition position). Because image acquisition was in a continuous mode, there were multiple images that contained one particular plant. To select only one image from each type of camera for each plant, Euclidean distances were calculated between the physical position of a plant and acquisition positions of all images. The depth, color, and thermal images with the minimum distances to the center of a plant were assigned to that plant for phenotypic trait extraction.

Subsequent image processing was performed on each plants images to extract phenotypic traits including five morphological traits and canopy temperature. Depth and color images were used to extract the morphological traits (left block in Figure 3.4). The raw depth and color images were firstly reconstructed to colored point clouds using a built-in function provided by the Kinect v2 sensor SDK, and a color filter (threshold of 0.15) based on excess green (ExG) index was applied to the colored point clouds to segment canopy points from the background. The plant points were used to calculate the five morphological traits: 1) plant height (H) was the distance between the ground surface (z equals 0) and the highest plant point; 2) width across-row (WAR) was the maximum distance along the x-axis; 3) width in-row (WIR) was the maximum distance along the y-axis; 4) projected leaf area (PLA) was the area covered by the canopy boundary; and 5) canopy volume (CV) was calculated using convex hull algorithm. Growth rates were calculated based on differences in the morphological traits between two measurement days to quantify dynamic changes of plant canopy. In total, growth rate of the morphological traits was calculated in five periods including DAP 51–65, DAP 65–72, DAP 72–87, DAP 87–94, and DAP 51–94.

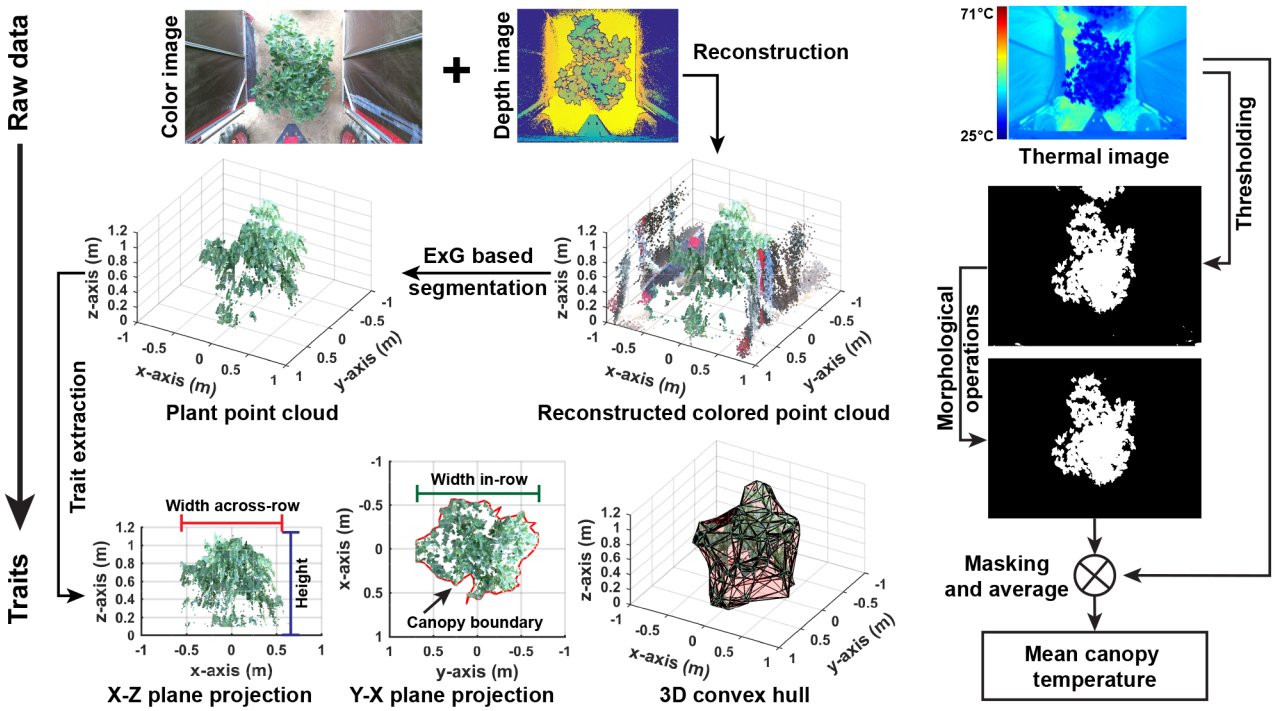

Figure 3.4: Flowchart of extracting morphological traits from depth and color images, and canopy temperature from thermal images. Excess green (ExG) index was used to segment plants from background in colored point clouds.

Canopy temperature was represented by the difference between canopy and air temperature (Tc-Ta). Air temperatures were retrieved from environmental data collected by the auxiliary sensing unit, whereas mean canopy temperatures were calculated from thermal images (right block in Figure 3.4). An adaptive thresholding algorithm (Otsu) was performed on each image to segment plants from the ground and the tractor frame. In the resultant binary images, connected components were identified and their centroid and circumcircle diameter were calculated. The connected component in the most central position was recognized as the target plant mask using Equation 3.3, and connected components representing parts of neighboring plants and/or weeds were removed using Equation 3.4. The final mask was subsequently applied to the original thermal image to calculate mean canopy temperature.

$$Index_{plant}^{CC} = argmin(\sqrt{(x_i - I_{xcenter})^2 + (y_i - I_{ycenter})^2}), \quad i = 1, 2, \ldots, n \tag{3.3}$$

where $Index_{plant}$ is the index of the connected component (CC) in the most central position, $I_{xcenter}$ and $I_{ycenter}$ are the x and y coordinates of the image center, and i is the index of a given connected component.

$$CC_i^{validated} = \begin{cases} 1, & (r_i + r_c) \leq \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \\ 0, & \text{otherwise} \end{cases}, \quad i = 1, 2, \ldots, n \tag{3.4}$$

where $CC^{validated}$ is a flag for a connected component, 1 or 0 means to include/exclude the connected component in the mask, x_i, y_i, and r_i are the x and y coordinates and the circumcircle diameter of the ith connected component , x_c, y_c, and r_c are the x and y coordinates and the circumcircle dimeter of the target plant connected component.

3.4.3 Statistical Analyses

Pearson correlation analyses were conducted between fiber yield and the extracted traits (or growth rates), and correlation coefficient (r) was used as an indicator to evaluate the potential of traits for establishing a yield prediction model useful to select high-yielding genotypes in breeding programs. Analysis of variance (ANOVA) with post-hoc Tukey-Kramer tests were conducted on the extracted traits among three genotype groups, exploring differences in plant growth and development between various cultivated and exotic species. In addition, broad sense heritability (H^2) was calculated for individual traits, and used as an indicator to evaluate the usefulness of a trait for genotype selection and/or quantitative genetic analyses such as genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping. ANOVA tests were performed on the extracted traits measured on each day between individual *G. hirsutum* genotypes. In the resultant ANOVA tables, the mean sum of squares

(MS) for the sources of 'genotype' and 'error' were the variances due to genotype and environment (including measurement error), respectively. Accordingly, H^2 values were calculated for individual traits using Equation 3.5 [124]. The *G. barbadense* species included only one genotype (Pima S6) in the present study. To avoid effects due to various species, data points of the *G. barbadense* species (Pima S6) were excluded from correlation analyses with fiber yield and H^2 calculation. All tests were performed in SAS (SAS 9.4, SAS Institute Inc., Cary, NC, USA) using a significance level of 0.05.

$$H^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E} \tag{3.5}$$

where H^2 is the broad sense heritability, V_P is the total phenotypic variance, and V_G and V_E are phenotypic variances due to genotype and environment effects.

3.5 Results

3.5.1 Representative Images Acquired by the System

The GPhenoVision system could successfully control the RGB-D and thermal cameras to acquire depth, color, and thermal images of cotton plants under field conditions (Figure 3.5). The system demonstrated the capability of simultaneously acquiring and storing high-volume images generated by multiple imaging sensors of high resolution. Therefore, the system could be a useful tool for field data collection, and the system design is reusable for FB-HTP projects aiming to utilize high-resolution imaging sensors.

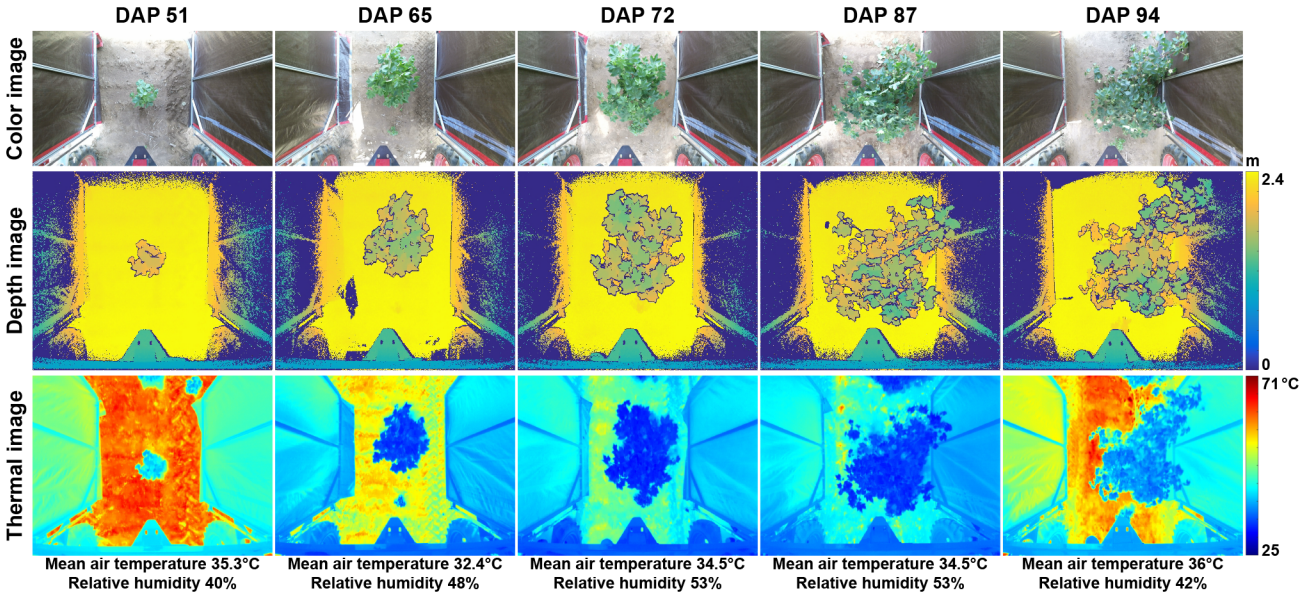

Figure 3.5: Representative color, depth, and thermal images acquired by the GPhenoVision system on five days in the field.

Additionally, the acquired images could potentially be used to extract various phenotypic traits. The depth images contained 3D information about plant canopies, which was useful to extract morphological traits such as canopy height, projection area, and volume. In combination with color images, depth images could be converted to colored point clouds in which vegetative components were easily segmented from the background. This might improve accuracies of measuring morphological traits, compared with other 3D imaging sensors such as LiDAR. Furthermore, color features were important for detecting certain plant organs such as flowers. Thermal images had an advantage in measuring not only canopy temperatures but also the temperature distribution over a canopy. Some leaves were cooler than others, probably due to their geometric distributions or physiological responses to environments.

3.5.2 Extracted Phenotypic Traits

Overall, plant canopies elongated and expanded substantially from DAP (day after planting) 51 to 87 (canopy development stage), but only slightly after DAP 87 by which time the flower and boll development stage had started (plant height, width in-row (WIR), and

width across-row (WAR) in Figure 3.6). Projected leaf area (PLA) followed the trends of plant width, whereas canopy volume (CV) followed the trends of both plant height and width (projected leaf area and canopy volume in Figure 3.6). This was because projected leaf area was primarily affected by plant expansion, but canopy volume was affected by both canopy elongation and expansion. Thus, any single dimensional development led to increases of canopy volume. Canopy temperature showed a different trend than canopy expansion (Tc-Ta in Figure 3.6). Before DAP 72, canopy temperatures decreased, as plants rapidly developed and increased leaf area available for transpiration (resulting in evaporative cooling). The decrease of canopy temperature was also due to environmental factors. The total precipitation levels were 111.76 mm and 178.82 mm during the period from the day of planting to DAP 51, and from DAP 66 to 72, respectively, resulting in an increase of average daily precipitation from 2.23 mm to 25.55 mm (see Supplementary Table S3 for detailed precipitation information). However, from DAP 73 to 94, environmental factors became dominant. Average daily precipitation was 6.21 and 3.19 mm per day from DAP 73 to DAP 84 and DAP 85 to 94, respectively. Reduced water availability may reduce transpiration, leading to higher canopy temperatures that may indicate water stress. The plant height, projected leaf area, and canopy temperature showed similar trends with previous studies [110, 111]. Moreover, GPhenoVision could measure multi-dimensional phenotypic traits such as projected leaf area and canopy volume, providing new tools for understanding canopy growth and development.

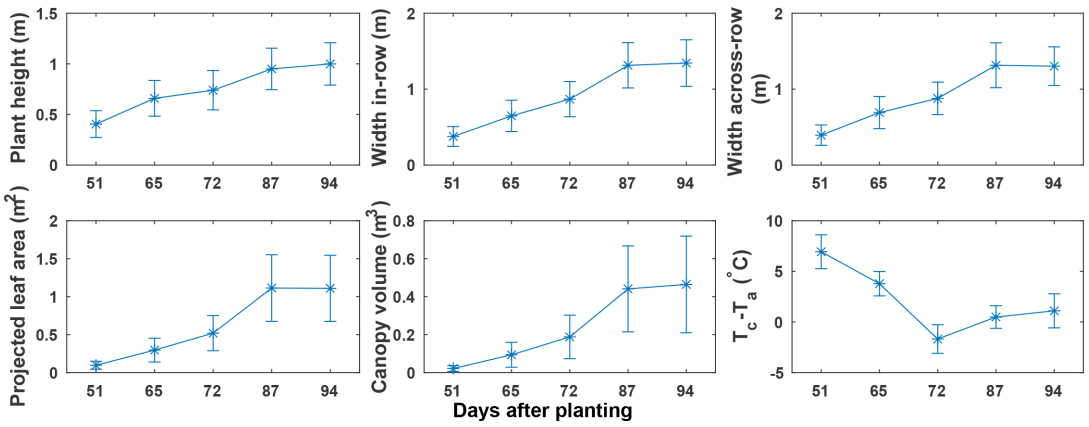

Figure 3.6: Six phenotypic traits measured using the GPhenoVision system on 14 July (51 days after plating, DAP 51), 28 July (DAP 65), 04 August (DAP 72), 19 August (DAP 87), and 26 August (DAP 94) in 2016. Asterisk and error bar indicated the mean value and standard deviation of traits for all plants in the field, respectively. DAP 51, 65, and 72 were in canopy development stage, and DAP 87 and 94 were in flower and boll development stage.

3.5.3 Correlation between Extracted Traits and Fiber Yield

Morphological traits generally had clear positive correlations ($r > 0.5$) with fiber yield (Table 3.3, see Supplementary Figure S2.1–S2.5 and Table S4.1–S4.5 for detailed results). The correlations became strong until DAP 87 and then were relatively weak on DAP 94. This is consistent with findings that canopy growth and development primarily contributes to fiber yield in early stages rather than late stages [125]. Single-dimensional traits were comparable (sometimes slightly better) with 2D and 3D traits in terms of correlation with fiber yield. This was likely due to the single plant layout where a plot contained only one plant. Compared with usual plot design, plant height in the single plant layout could be more indicative of plant growth and development, showing strong correlation with fiber yield. High canopy temperature, an indicator of plant water stress, mostly showed negative correlations with fiber yield. However, no correlation was observed between fiber yield and canopy temperatures measured on DAP 72, when ample water was available, suggesting that at that time growth was not constrained by water limitation and thus canopy temperature

was not predictive of fiber yield [111].

Table 3.3: Pearson correlation coefficients between fiber yield and phenotypic traits measured on five days after planting (DAP) in 2016. Pearson correlation tests used data from 100 plants (n=100) in the field. Asterisks (or abbreviations) indicated different statistical signifiance levels: NS for not significant, * for p-value <0.05, ** for p-value <0.01, *** for p-value <0.001.

Date (DAP)	Plant height	Width in-row	Width across-row	Projected leaf area	Canopy volume	Tc-Ta
14 July (51)	0.65***	0.67***	0.70***	0.66***	0.59***	-0.65***
28 July (65)	0.68***	0.65***	0.61***	0.60***	0.49***	-0.44***
04 August (72)	0.68***	0.64***	0.66***	0.63***	0.62***	NS
19 August (87)	0.69***	0.64***	0.62***	0.66***	0.58***	-0.45***
26 August (94)	0.56***	0.60***	0.65***	0.60***	0.53***	-0.25*

During the entire phenotyping period (DAP 51 to 94), growth rates of all morphological traits showed positive correlations with fiber yield (Table 3.4, see Supplementary Figure S3.1–S3.5 and Table S5.1–S5.5 for detailed results). Growth rates of projected leaf area and canopy volume had stronger correlations with fiber yield than plant heights and widths, suggesting that growth rates of 2D and 3D traits might be better predictors than those of 1D traits in a linear model for yield estimation. It was noteworthy that growth rates of WIR, WAR, PLA, and CV had no statistical correlation with fiber yield from DAP 87 to 94, because reproductive growth became dominant in this period and WIR, WAR, PLA, and CV did not change significantly. However, plant height increased slightly in that period, because excessive vegetative growth was from a single terminal bud located at the tallest part of branches. A moderate negative correlation was observed between fiber yield and growth rate of plant height from DAP 87 to 94, supporting the hypothesis that excessive vegetative growth may reduce reproductive growth in late stages of crop development. Therefore, a combination of growth rates of multi-dimensional morphological traits would be particularly

useful for studying plant energy use efficiency in different growth stages.

Table 3.4: Pearson correlation coefficients between fiber yield and daily phenotype growth rates calculated between different days after planting (DAP) in 2016. Pearson correlation tests used data from 100 plants (n=100) in the field. Asterisks (or abbreviations) indicated different statistical signifiance levels: NS for not significant, * for p-value <0.05, ** for p-value <0.01, *** for p-value <0.001.

Period	Plant height	Width in-row	Width across-row	Projected leaf area	Canopy volume
14 July to 28 July (DAP 51-65)	0.54***	0.39***	0.32**	0.52***	0.43***
28 July to 04 August (DAP 65-72)	NS	NS	NS	0.51***	0.58***
04 August to 19 August (DAP 72-87)	NS	0.22**	0.21**	0.52***	0.38***
19 August to 26 August (DAP 87-94)	-0.36**	NS	NS	NS	NS
14 July to 26 August (DAP 51-94)	NS	0.34***	0.29**	0.55***	0.50***

3.5.4 Differences in Extracted Traits Among Genotype Groups

Overall, *G. barbadense* had the lowest value of all traits after DAP 65, with a shorter and smaller plant canopy than *G. hirsutum* and exotic genotypes (Figure 3.7). Plant height of *G. barbadense* Pima S6 reached its maximum on DAP 65, 22 days earlier than the *G. hirsutum* and exotic genotypes, whereas other static traits generally reached the highest level on DAP 87. These data might indicate abnormal vegetative growth of *G. barbadense* genotypes, which are not well adapted to the study area and not generally grown there. Manual field assessment showed all *G. barbadense* plants to be lodged, although most were otherwise healthy and continued to grow horizontally.

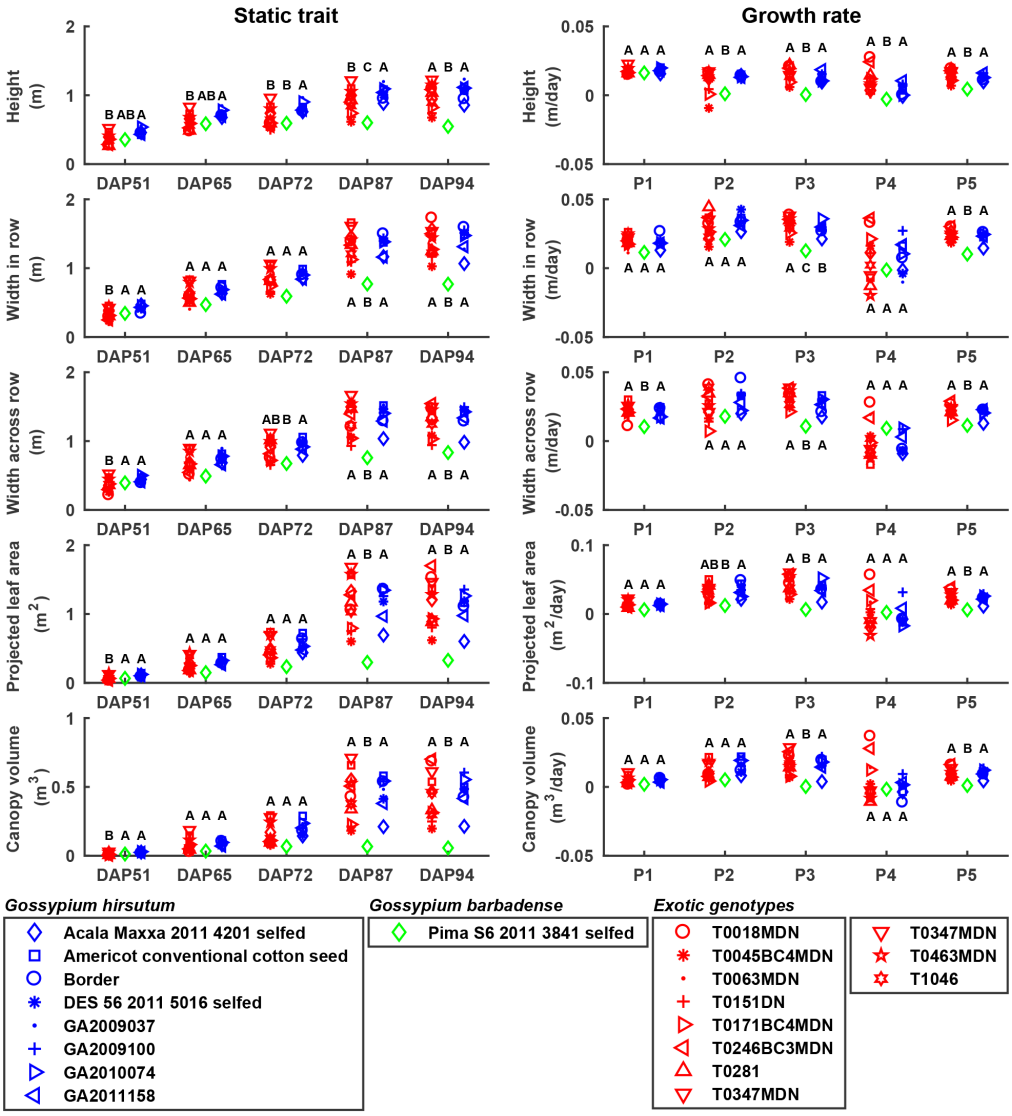

Figure 3.7: Differences in extracted traits among three cotton groups: *Gossypium hirsutum*, *Gossypium barbadense*, and 'exotic' genotypes. Growth rates were calculated in five periods including P1 (DAP 51–65), P2 (DAP 65–72), P3 (DAP 72–87), P4 (DAP 87–94), and P5 (DAP 51–97). Each marker indicated the mean value of traits for a genotype used in the study, and genotypes in the same group were rendered as the same color. Groups with different letters were statistically different from each other (see Supplementary Table S6 for detailed p-values), and group mean values were sorted alphabetically. DAP 51, 65, and 72 (accordingly P1 to P3) were in canopy development stage, and DAP 87 and 94 (P4) were in flower and boll development stage.

Plant canopy size (projected leaf area) of exotic genotypes was smaller than those of *G. hirsutum* and *G. barbadense* cultivars on DAP 51, indicating that the exotics established their maximum size slower than the two cultivated genotype groups. This was expected because the exotics tend to be late flowering and experience rapid vegetative growth in late stages. After DAP 51, the exotics did indeed experience rapid vegetative growth, consistently as high as those of *G. hirsutum* genotypes, resulting in large plant canopies that were comparable with *G. hirsutum* genotypes in late stages (after DAP 87). Some exotic genotypes (T0246BC3MDN, T0018MDN, T0347MDN, and T0151DN) had even larger canopies than *G. hirsutum* genotypes and maintained high vegetative growth rates, perhaps reflecting greater investment in vegetative growth and less in reproductive growth (seed and associated fiber).

Trait variations of the exotics were larger than those of elite *G. hirsutum* genotypes, especially static traits after DAP 87 and growth rates in the period from DAP 87 to DAP 94. This was expected, because elite *G. hirsutum* genotypes have been selected by plant breeders for many years and are generally very closely related to one another. On the other hand, the exotics harbor substantial genetic variations, some of which may be used to improve *G. hirsutum* genotypes by crossing and selection.

3.5.5 Broad Sense Heritability of Extracted Traits

Most of the measured traits had broad sense heritability (H^2) greater than 0.5, supporting the usefulness of the GPhenoVision system in genomics/genetics studies. H^2 values of morphological traits (static traits) decreased to their lowest values in the middle of the growing season, and then began to increase and ultimately reached the maximum values at the last measurement date (Figure 3.8). In particular, H^2 values for all morphological traits were larger than 0.5 after DAP 87 (H^2 of some traits was over 0.7 after DAP 94), indicating the usefulness of the traits for quantitative genetic analyses such as genome-wide association

studies (GWAS) and quantitative trait locus (QTL) mapping. On the contrary, H^2 of canopy temperature showed a random pattern. Canopy temperature is expected be useful for genotype selection on certain days but not others. This was probably due to two reasons. First, during periods when cotton plants have adequate soil moisture, there may be little or no temperature difference between drought resistant and non-resistant genotypes. Further, wide row and plant spacing provided more air movement and less competition among plants for water. Second, the experiment was small and might not have used sufficiently large samples to discern statistically significant differences in canopy temperature. Thus, it would be better to study canopy temperature in experiments with irrigation treatments and larger populations. H^2 values of growth rates (dynamic traits) showed an increasing trend until reaching maximal levels (>0.7) in the middle of the growing season (DAP 65 to 72 or DAP 72 to 87). The growth rates H^2 over the entire growing season (DAP 51 to 94) were also larger than 0.6. Low-yielding genotypes may continue to grow vegetatively while high-yielding early- and late-season genotypes are in reproductive growth stages during the period from DAP 72 to 87.

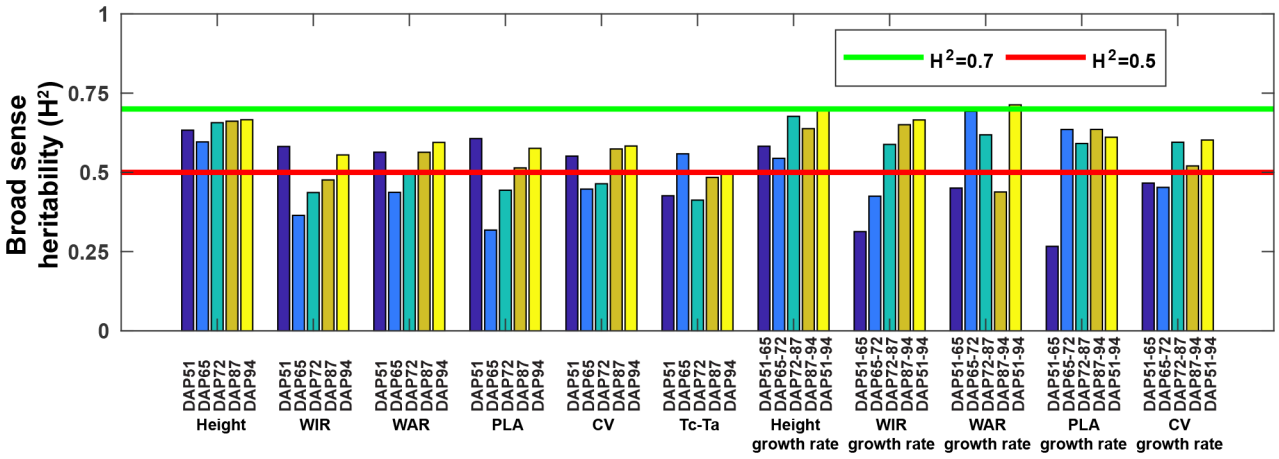

Figure 3.8: Broad sense heritability (H^2) of phenotypic traits extracted in the present study. $H^2 > 0.5$ indicated a trait would be useful for genotype selection, and $H^2 > 0.7$ indicated a trait would be useful for genotype differentiation.

3.6 Discussion

Compared with other existing integrated systems, GPhenoVision is a field-based HTP system that consists of multiple high-resolution imaging modalities. Modularity and customizability are two key features of the GPhenoVision system. The modular design is implemented and demonstrated at various levels. The GPhenoVision system is firstly decoupled into four subsystems, with each providing a particular system function: mechanical platform for system mobility, electrical system for powering, and sensing system and software for data acquisition. In particular, the sensing system and data acquisition software are separated as two parts, because the sensing system addresses hardware concerns (e.g. sensor installation position and communication interface) while the data acquisition software provides efficient solutions for user interaction and data management (e.g. data transfer, visualization, and storage). Subsystems are modules in the GPhenoVision system, and replacement/modification of any module will not affect other modules or the entire system functionality as long as the replacement/modification follows the same module interface. For instance, the electrical system is a valid module to the entire system as long as it provides stable 120V AC power output (the module interface), regardless of which power source (either the tractor secondary alternator or a separate power generator) is being used. Individual subsystems are also decoupled into modules for certain functions, and the DAQ software primarily demonstrates the modular design at this level. The DAQ software uses a multilayered architecture, with each layer being a software module. The synchronization of the sensor control, and data cache and transfer layers provides a mechanism for avoiding data acquisition latency due to the large difference between the data volume generated by imaging sensors and the input/output (I/O) speed of hard drives. Although the data transfer layer with an SSD can achieve a writing speed up to 520 MB/s, which is adequate for most imaging sensors such as LiDARs and hyperspectral cameras, they can be replaced by new hardware and system writing functions to fulfill requirements of higher resolution (thus data volume) sensors without modification of other software modules. The sensor control layer is

further decoupled, and each sensor thread is considered as a module in this layer with the module interface of the EFSM. Any EFSM-based sensor thread can be properly controlled by the sensor control layer regardless of the differences in implementation of sensor control. For instance, the RTK-GPS needs a simple LabVIEW function to initialize a serial port for communication/data transfer, whereas the hyperspectral camera needs multiple SDK functions to initialize the SDK library, open a USB3 port for communication/data transfer, and configure camera parameter. Although the initialization implementation of the two sensors is different, the sensor control layer turns them into the same state (sensor initialization) due to the modular design. In other words, sensors (sensor threads) are interchangeable to the sensor control layer because they follow the same module interface. The modularity at various levels does not only simplify the system development and maintenance, but also increases the system customizability. Various sensors can be conveniently integrated into (or removed from) the system for different phenotypes by adding (or deleting) EFSM-based sensor threads in the sensor control layer of the DAQ software. In addition, integrated sensors can be readily selected to use (or not use) in the DAQ software due to the modular design. As the priority of phenotypic traits changes between different growth stages, the system can use various combinations of imaging modules without modification of source code. For instance, canopy temperature becomes less meaningful in late stages such as cotton boll maturity, and thus the thermal camera can be turned off to reduce data storage space. The modular design ensures the stability of customization of the GPhenoVision system in various situations, because turning-off or malfunction of one sensor (module) will not affect the use of other sensors (modules).

The calibration and validation results showed that the RGB-D, thermal, and hyperspectral cameras achieved sufficient measurement performance for measuring various types of phenotypic traits. The RGB-D camera is used to measure morphological traits, and plant height as a representative morphological trait has been repeatedly measured by the

RGB-D camera with a difference of 3–4 cm to manual measurements. If considering the field conditions with a wind speed of 3–6.6 m/s, this measurement error is acceptable. The thermal camera is used to measure canopy temperature, and it achieved the nominal measurement accuracy ($0.5°C$) in both the laboratory and field conditions. In addition, there was no difference between measurements of plants under shaded and unshaded areas, suggesting no shading effect on the canopy temperature measurement during the regular data collection. The hyperspectral camera provides a high spectral sampling interval of 2.6 nm with a spectral shift less than 1 nm, and the best and worst spatial resolutions are 2.2 mm/pixel and 6.8 mm/pixel, respectively. Given this level of spectral and spatial resolutions, the hyperspectral camera may provide ample information to study plant physiological status (e.g. photosynthesis and diseases) at the organ level, but these need to be further studied by considering experimental design, plant material preparation, agronomic practices, and data analysis.

Although the present study was small-scale, it demonstrated the usefulness of measured phenotypic traits for genomics studies and breeding programs. The sensing capabilities of the GPhenoVision system created new opportunities for measuring multi-dimensional morphological traits such as projected leaf area and canopy volume. In particular, cotton plants continued vegetative growth along different dimensions during the growing season, so it would be better to characterize canopy growth and development using morphological traits in multiple dimensions. Additionally, growth rates were calculated for morphological traits in multiple dimensions, providing the possibility of studying relationships between canopy architecture at different growth stages and plant reproductive efficiency. In contrast, the usefulness of canopy temperature mostly depended on plant growth and environmental conditions. Precipitation was only 0.36 mm/day during DAP 51 to DAP 65, resulting in a large variation of canopy temperature among genotypes on DAP 65. Therefore, different levels of

irrigation are necessary to study drought-resistant genotypes by using thermal imaging data.

The GPhenoVision system is modular and customizable, and has demonstrated the potential for genetics/genomics studies and breeding programs, but it should be acknowledged that several parts of the system can be further improved. First, the current illumination configuration (just relying on solar radiation) was not optimal for all three imaging modules. The RGB-D and thermal cameras worked well in the shading condition, whereas the hyperspectral camera showed a relatively low signal intensity due to the reduced incident light. The enclosure can be further split into two sections: RGB-D and thermal section and hyperspectral section. Additional illumination sources can be configured to increase the incident light intensity (and thus reflectance intensity in images) in the section for the hyperspectral camera. Second, challenges still remain in the development of data processing algorithms to take full advantage of the sensing capabilities of the GPhenoVision system. The algorithms in the present study were developed for processing images collected in the SPL field, and need a significant modification for data collected in fields with regular plot layout. In addition, new algorithms should be able to accurately extract complex traits from 3D or hyperspectral images, which require advanced techniques in computer vision and machine learning. For instance, plant components such as flowers or cotton bolls can potentially be detected using convolutional neural networks, even in field conditions containing a complex background. Moreover, processing speed will be an important concern that affects phenotyping throughput. Therefore, it is necessary to consider new computational approaches such as cloud computing and GPU-based optimization to speed up algorithms for trait extraction. In addition, image data are usually in high volume, posing challenges in data storage, management, and sharing. If a project involves researchers from the same institution, a viable solution is to use a storage service maintained by local agencies; otherwise, cloud-based services need to be considered such as 'CyVerse', because they can provide reliable data storage, management, and sharing to users in different regions.

3.7 Conclusions

The GPhenoVision system reported in this study can control RGB-D, thermal, and hyperspectral cameras to collect images of cotton plants in field conditions. The proposed sensing system structure and DAQ software architecture would allow rapid development of a custom FB-HTP system that could handle imaging sensors that generate high-volume data. The validation and calibration results showed that the three cameras could provide accurate raw data for phenotyping purposes. Most of the measured traits had H^2 over 0.5 (some over 0.7), confirming the usefulness of using the GPhenoVision system in genomics/genetics studies. Future studies will be focused on developing image processing algorithms to extract more traits and deploying the system in a large-scale experiment for genetic analyses.

Chapter 4

3D Point Cloud Data to Quantitatively Characterize Size and Shape of Shrub Crops[1]

[1]Jiang, Y., Li, C., Takeda, F., Kramer, E. A., Ashrafi, H., and Hunter, J. 2019. *Horticulture Research*, 6(1), 43. Reprinted here with permission of publisher.

4.1 Abstract

Size and shape are important properties of shrub crops such as blueberries, and they can be particularly useful for evaluating bush architecture suited to mechanical harvesting. The overall goal of this study was to develop a 3D imaging approach to measure size-related traits and bush shape that are relevant to mechanical harvesting. 3D point clouds were acquired for 367 bushes from five genotype groups. Point cloud data were preprocessed to obtain clean bush points for characterizing bush architecture, including bush morphology (height, width, and volume), crown size, and shape descriptors (path curve λ and five shape indices). One-dimensional traits (height, width, and crown size) had high correlations (R^2 = 0.88–0.95) between proposed method and manual measurements, whereas bush volume showed relatively lower correlations (R^2 = 0.78–0.85). These correlations suggested that the present approach was accurate in measuring one-dimensional size traits and acceptable in estimating three-dimensional bush volume. Statistical results demonstrated that the five genotype groups were statistically different in crown size and bush shape. The differences matched with human evaluation regarding optimal bush architecture for mechanical harvesting. In particular, a visualization tool could be generated using crown size and path curve λ, which showed great potential of determining bush architecture suitable for mechanical harvesting quickly. Therefore, the processing pipeline of 3D point cloud data presented in this study is an effective tool for blueberry breeding programs (in particular for mechanical harvesting) and farm management.

4.2 Introduction

Blueberries are nutritious fruit, containing ample amounts of phytochemicals (e.g., antioxidants) beneficial to human health [126]. The United States (US) is the largest blueberry producer and consumer in the world, and the recognition of blueberry economic and nutritional values have prompted cultivation of blueberries in other countries (e.g., Chile, China,

Mexico, Peru, Australia, and European countries) [127]. Thus, an increasing demand for blueberries has been foreseen, requiring improvements in blueberry production technology and fruit quality in the future. These improvements require breeding programs to develop superior genotypes that are better adapted to different climates and modern agriculture production practices, including fruit harvesting with over-the-row (OTR) mechanical harvesters. Phenotyping technologies provide various traits for genotype evaluation in breeding programs [7, 100, 128]. These traits can also be used for management decision-making in commercial production fields such as the ability to use mechanical harvesting methods with limited (or even no) impacts on fruit quality.

Bush architecture is important for tree/shrub crops, because it usually can be used for growth evaluation, biomass estimation, yield prediction, harvest efficiency improvement, and utilization of plant protection products (PPPs) such as pesticides [129]. Size and shape are two important aspects of bush architecture. Size-related traits indicate the overall growth status of bushes, which are related to yield potential. Studies were conducted to use size-related traits to evaluate blueberry vegetative growth under various environments, showing a reasonable correlation between those traits (e.g., bush height) and berry yield [130, 131]. Bush shape describes the geometry of bushes, which is an important factor affecting the performance of OTR mechanical harvesters. Some blueberry growers are already using OTR mechanical harvesters. More growers expect to rely on OTR mechanical harvesters to pick blueberries for fresh market, addressing the challenges of increasing harvest labor cost and anticipated insufficient labor force. To maximize the performance of OTR mechanical harvesters, blueberry plants ideally should have a narrow and small crown (e.g., small cross-section area near the ground) and a vase-shaped canopy [132, 133, 134]. A narrow and small crown is easy to tighten with catch plates of OTR harvesters, leaving small gaps between the bush and the catch plates to prevent berries from falling to the ground [135, 136]. This reduces yield losses due to mechanical harvesting. A vase-shaped canopy positions fruit away

from the central crown of a blueberry plant, providing a relatively open area for berries to drop onto harvester catch plates. This reduces external impacts (and thus potential bruises) on machine-harvested berries for better fruit quality and longer shelf-life. To date, the determinations of bush dimension, crown size, and bush shape have largely relied on manual assessments, which are subjective and laborious.

Crop size-related traits have been widely studied using 2D and 3D imaging modalities. 2D imaging approaches were primarily used to extract unitless ratios or traits in the unit of image pixel [15]. When a reference object is provided or the imaging system is pre-calibrated, extracted traits can be converted to real world units. Conversion models are usually established for greenhouse- or chamber-based phenotyping systems due to the easy deployment of reference objects and precise configuration of a pre-calibrated imaging system [137, 138, 139, 140]. When the distance between the canopy and camera is relatively consistent, unitless ratios (e.g., canopy coverage ratio) are comparable over different data collection periods and thus they have been extracted in many 2D imaging based studies [15]. With the increased availability of 3D sensing approaches, researchers are starting to frequently use 3D imaging techniques for measuring size-related traits [141, 15]. Previous studies intensively investigated size-related traits at the plant and canopy levels for tree and shrub crops such as apples [142], pears [142], grapes [142, 143, 144], hickories [145], olives [146], almonds [147], peaches [148], and blueberries [149]. These studies showed a general trend that the accuracy of crop size measurement mostly depended on point cloud quality which is determined by sensing range and imaging approaches. Photogrammetry based 3D imaging approaches (e.g., the structure from motion (SfM)) are inexpensive and can provide detailed point cloud data, but they require considerable computational resources for 3D reconstruction. The quality of reconstructed point clouds is significantly affected by ambient conditions such as illumination changes and wind. In addition, the SfM technique requires the use of reference objects to scale reconstructed point clouds, if no metric data

(e.g., accurate metric positions of image acquisition) is provided. In such cases, reference targets need to be included in the imaging scene, introducing potential challenges for large field experiments (e.g., over several hundreds of plots) [150, 151, 152]. Active 3D imaging instruments (e.g., LiDARs) are costly but usually provide fast 3D measurements. Some active instruments have particular outdoor configurations (e.g., special emitting illumination sources) to dramatically improve the accuracy and repeatability of 3D reconstruction in the field. However, occlusions can lead to incomplete scanning of objects, presenting difficulties in trait measurement. For instance, it would be difficult to measure plant organs and branches under the canopy because they cannot be imaged by instruments using a single sensing angle. Recently, a handheld mobile laser scanner was developed so that full-view point clouds can be obtained [153]. Two studies demonstrated that the handheld laser scanner could obtain point clouds with much less missing points due to occlusion for forest structure characterization and inventory [154, 155]. Thus, it is worthwhile to explore the use of this laser scanner to obtain point cloud data for measuring size-related traits of shrub crops, especially the traits of plant parts under the canopy such as crown size of blueberry plants.

Shape analysis methods can be grouped into two categories: descriptive methods and outline-based methods [156]. Both methods have been commonly used to analyze shapes of fruits, vegetables, and plant leaves. Descriptive methods usually define landmark points that can be used to derive ratios, angles, and their combinations for quantifying object shapes. Descriptive methods have been used to study the shape of tomatoes [157, 158], eggplants [159], vineyard grape leaves [160], and peppers [161]. Outline-based methods rely on advanced mathematical tools (e.g., curve functions and elliptical Fourier analysis (EFA)) to quantitatively describe object shapes using transformed features. Studies reported the use of EFA for analyzing the shape of mistletoe berries [162], cotton leaves [163], oranges [164], ash tree fruit [165], and persimmons [166]. Shape descriptors defined in descriptive methods have

clear physical meanings, which can be easily interpreted and compared. However, defining descriptors requires a good understanding of domain needs and knowledge, involving extra efforts from domain experts. In contrast, features extracted using outline-based methods usually have no direct physical meaning, which requires visualization tools for feature interpretation and understanding. Outline-based methods use general mathematical models/framework to calculate shape features, which require almost no domain knowledge for conducting data analyses. It is also possible to use both methods for a comprehensive analysis because shape descriptors from the two methods could be complementary to each other [167]. In fact, both methods would be suitable for bush shape analysis for two reasons: 1) there is a clear physical definition of optimal bush shape for mechanical harvesting, and thus it would be straightforward to define landmark points to extract shape features; 2) previous studies [168, 162] demonstrated that a curve function (path curve) can effectively depict differences between vase-shape, cone-shape, and round shape, which would be worthwhile to explore.

To the best of our knowledge, only one study from our group has reported on the potential of using 3D imaging to extract size-related traits and shape descriptors of bush crops such as blueberries [149]. The study used an unmanned aerial system (UAS) to acquire oblique images of blueberry bushes from approximately 3 m above the ground for reconstructing point clouds using the SfM technique. It achieved a strong correlation ($R^2 = 0.92$) between imaging and manual measurements of bush dimensions (e.g., height and width) with an root mean square error (RMSE) of 0.1 m, indicating a high system measurement accuracy. However, correlations were less desirable ($R^2 = 0.380.55$) between imaging and manual measurements for crown size which is one of the most important parameters of machine harvest efficiency. The undesirable correlations occurred primarily due to the limitation of data collection system. When images were acquired on top of bushes regardless of using nadir or oblique perspectives, bush canopy occluded plant architecture close to the ground, leading

to an incompletion of bush crown reconstruction and thus inaccurate measurement of crown size. In addition, some shape descriptors defined in the study may not have been effective in identifying desired bush shapes. For instance,"blockiness" was defined as the ratio of widths at 85% and 65% canopy heights, but it had no relation to the position of the widest canopy cross-section which is the determinant among round (the widest cross-section in the middle), conical (the widest cross-section in lower canopy), and vase shapes (the widest cross-section in upper canopy). In fact, bush canopy can be round, conical, and vase-shaped for the same "blockiness" value. Therefore, it is necessary to address aforementioned issues and provide improved approaches in both data collection and analysis for measuring size-related traits and bush shape. These approaches would be particularly useful for breeding programs to select blueberry genotypes suited to machine harvesting.

The overall goal of this study was to develop a 3D imaging approach to measure blueberry bush dimensions and shape in the field. Specific objectives were to: 1) evaluate the accuracy of sensor measurements for objects with standard shapes in field conditions; 2) develop data processing algorithms to extract size-related traits (bush dimensions and crown size) and shape descriptors of blueberry bushes; 3) evaluate the accuracy of proposed method; and 4) explore the usefulness of bush shape descriptors for machine harvesting and farm management.

4.3 Materials and Methods

4.3.1 Blueberry Field and Data Collection

The study was conducted in two blueberry fields. The first field (33°53'10.7"N, 83°25'15.1"W) was located at the Horticultural Farm of the University of Georgia in Watkinsville, Georgia, USA and consisted of 7-year-old southern highbush blueberry (*Vaccinium darrowii*) bushes. The bushes had been un-pruned for two years at the time of this study. Point cloud data were collected in a sub area (23 m×15 m) containing 47 bushes (O'Neal cultivar) on 4 October

2016 with a clear sky view and an average wind speed of 2.7 m/s. A closed-loop walking path was predetermined, with at least one pass for each side of individual bushes. This walking strategy ensured that bushes would be scanned from multiple angles to improve point cloud coverage. While moving along the predefined path, a person carried the ZEB1 scanner (GeoSLAM, Ruddington, Nottinghamshire, United Kingdom) and swung the scanner node across the movement direction. The walking speed was about 1.4 m/s, and it took approximately 5 minutes to complete the scanning. A total of 20 bushes were selected for further analyses, because they were relatively small plants which in practice could be measured manually.

The second field (34°21'42"N, 77°50'11.9"W, 68 m×12.5 m) named Ideal Tract Farm was in the Horticulture Research Station in North Carolina. The study characterized 222 8-year-old bushes that have been bred for mechanical harvesting. The 222 bushes were pruned prior to data collection. Data collection was conducted on 15 March 2018 with an overcast sky and an average wind speed of 3.5 m/s. A closed-loop walking path was used with one pass on each side of individual bushes, ensuring that bushes would be scanned from multiple angles to improve point cloud coverage. While moving along the predefined path, a person carried the ZEB1 scanner and swung the scanner node across the movement direction. The walking speed was about 1.0 to 1.2 m/s due to the muddy ground condition, and it took approximately 10 minutes to complete the scanning. All 222 bushes were used for further data analyses, but no manual measurements were conducted.

4.3.2 Data Processing Pipeline of Extracting Size and Shape Traits

Point Cloud Preprocessing

A data processing pipeline was developed to extract size-related traits and bush shape, including data acquisition, preprocessing, and trait extraction (Figure 4.1 and Figure 4.2). Raw data were manually transferred from the scanner to a workstation computer and

uploaded to the manufacturer's web service (GeoSLAM Cloud, Bingham Nottingham, Notts, UK) for 3D reconstruction. The reconstructed point cloud data contained 4.7 million points (152 MB in LAS format) and 17 million points (547 MB in LAS format) for the two experimental fields.

In the preprocessing stage, clean point clouds of individual blueberry bushes were obtained (Preprocessing in Figure 4.1). The first step was to rotate the point cloud of the entire scanning area to a coordinate system in which the ground plane was paralleled to the x-y plane and the bush row direction was aligned with x-positive direction. Ground normal was calculated through plane fitting using maximum likelihood estimation sample consensus (MLESAC) [169], and rotation transform matrices (T_y and T_x) were accordingly derived for the x-z and y-z planes. Rotation matrix for the x-y plane (T_z) was derived based on the row direction identified using Hough transform. The point cloud of the scanning area was rasterized to a depth image, with each image pixel representing the maximum depth (z value) in a grid of 0.5×0.5 m^2 in the point cloud. The depth image was thresholded (the threshold was 0.3 because bushes were at least 0.3 m above the surrounding ground) to segment bush pixels. The Hough transform was performed on all bush pixels to detect lines (bush rows) in a feature space ($[\alpha, \rho]$) where α was the complementary angle to the line orientation and ρ was the distance from the line to the origin. If a line represented a bush row, the line should go through as many bush pixels as possible, and thus the high occurrence of $[\alpha, \rho]$ values (the α line with the least intersection points with curves in the Hough space graph) denoted bush rows. The best α value was accordingly selected to calculate the bush row direction. Subsequently, a rotated point cloud of the scanning area was calculated using Equation 4.1.

$$
\begin{aligned}
PtCloud_{trans} &= PtCloud_{orign}\mathbf{T_x}\mathbf{T_y}\mathbf{T_z} \\
\mathbf{T_x} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\theta_x) & sin(\theta_x) \\ 0 & -sin(\theta_x) & cos(\theta_x) \end{bmatrix} \\
\mathbf{T_y} &= \begin{bmatrix} cos(\theta_y) & 0 & sin(\theta_y) \\ 0 & 1 & 0 \\ -sin(\theta_y) & 0 & cos(\theta_y) \end{bmatrix} \\
\mathbf{T_z} &= \begin{bmatrix} cos(\theta_z) & sin(\theta_z) & 0 \\ -sin(\theta_z) & cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned} \tag{4.1}
$$

Where *PtCloud* denoted a point cloud matrix, $\mathbf{T}$ represented rotation matrices, θ_x, θ_y, and θ_z were rotation angles around the x-, y-, and z-axes, respectively.

The rotated point cloud was re-oriented to ensure that the bush row vector was towards the x-positive direction, and then the experimental area was cropped based on its dimension (23 m and 15 m along x-positive and y-positive directions). Bounding boxes were manually drawn for the selected 20 bushes, and raw point clouds of individual bushes ($PtCloud_{RawBush}$) were segregated accordingly.

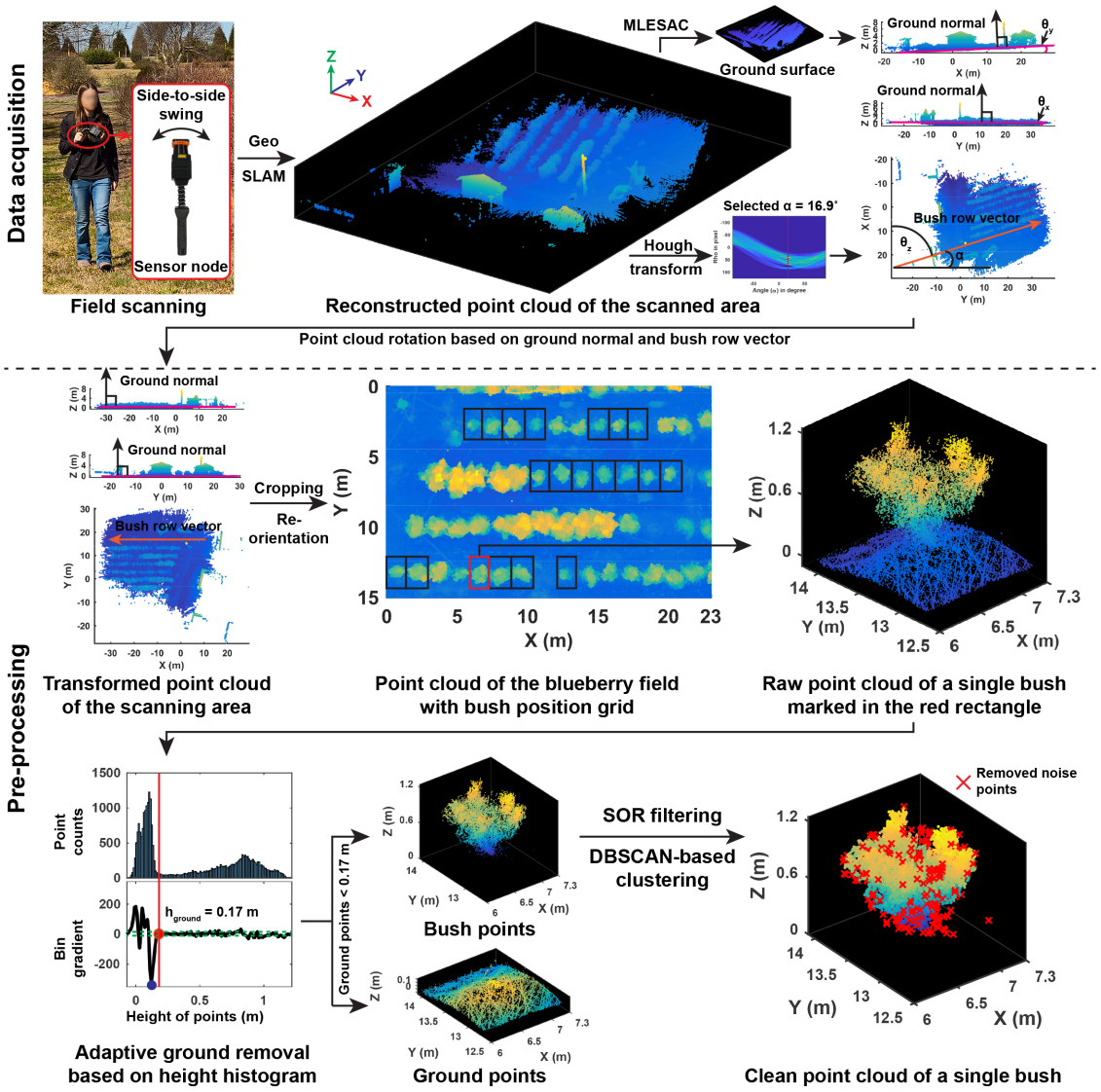

Figure 4.1: Flowchart of data acquisition and preprocessing to obtain clean point clouds of individual blueberry bushes. In the diagram, for the adaptive ground removal based on height histogram, the red and blue dot indicated the determined threshold and the height value with the least bin gradient, respectively. The green dash lines depicted the value range in which bin gradient values were close to zero.

The second step was to remove ground and noise points in raw bush point clouds, obtaining clean bush point clouds for trait extraction. The ground surface was not flat due to agronomic practices (e.g., additional woody layer), and as a result plane-fitting based

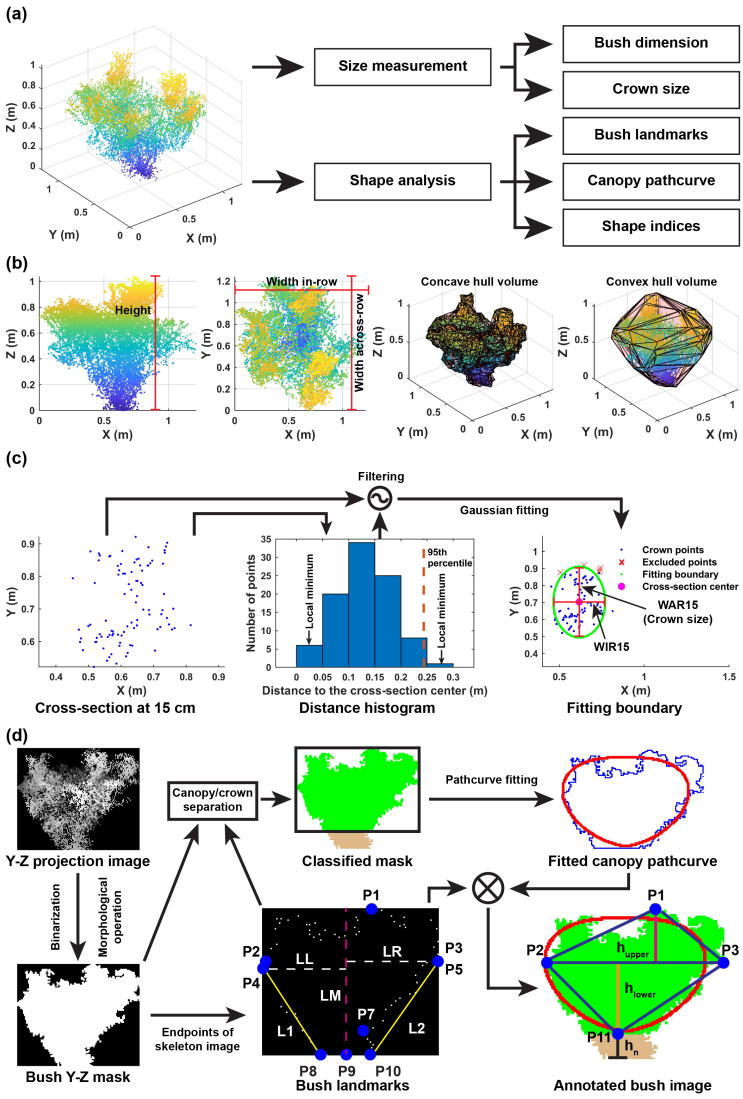

Figure 4.2: Flowchart of data processing to extract size and shape traits for blueberry bushes. (a) overall diagram of trait extraction; (b) measurement of size-related traits; (c) measurement of bush crown size; and (d) calculation of bush shape indices. In the sub panel (c), WAR15 and WIR15 were the width across-row and width in-row of the cross-section at 0.15 m above the ground, and WAR15 was used as crown size in the present study. In the sub panel (d), green and brown colors indicated the canopy and crown part of bush. Blue and red curves were the contours and the best fitted path curves of the bush canopy. P1 to P11 denoted 11 landmark points including the bush top-center point (P1), left (P2) and right (P3) endpoints of the broadest cross-section, bush leftmost (P4) and rightmost (P5) endpoints, left (P6) and right (P7) canopy-crown separation points, left (P8), center (P9), and right (P10) endpoints of the bush bottom, and the center point of the canopy bottom (P11). It should be noted that in this case, P6 did not exist and P3 and P5 overlapped. LM was the center line of the bush, and LL and LR were the left and right border lines between bush upper and lower portions. L1 and L2 were the left and right outer boundary lines of the bush bottom portion. h_n, h_{upper}, and h_{lower} denoted the height of the bush crown and canopy upper and lower triangles.

methods such as random sample consensus (RANSAC) could not remove ground points correctly. An adaptive thresholding approach was proposed for ground point removal. In each raw bush point cloud, a height histogram was generated using a bin width of 0.01 m, and bin gradients were calculated accordingly. The threshold of ground points was determined by three criteria: its bin gradient was close to zero; it was greater than the lower limit of the bin with the least gradient value; and it should be as small as possible. The threshold was calculated using Equation 4.2.

$$h_{ground} = min(\{h|0 < |G(h)| < t\} \cap \{h|h > argmin_h(G(h))\}) \tag{4.2}$$

Where h_{ground} was the determined height threshold for ground points, $G(\cdot)$ was the gradient of a bin, h was the lower limit of bins in a height histogram, and t (set to 5 in the present study) was a noise factor for selecting bins with the gradient close to zero.

If point heights were lower than the threshold, the points in raw bush point clouds ($PtCloud_{RawBush}$) were classified as ground ($PtCloud_{ground}$) or otherwise bush ($PtCloud_{bush}$). After removing ground points, noise points in bush point clouds ($PtCloud_{bush}$) were detected and excluded using statistical outlier removal (SOR) filter. For each point, Euclidean distances to its k nearest neighboring points were calculated. A point was categorized as noise, if the mean distance between that point to its k neighboring points was larger than n times of the standard deviation (Equation 4.3 and Equation 4.4).

$$D(p,k) = \frac{\sum_{i=1}^{k} \sqrt{(p_x - p_x^i)^2 + (p_y - p_y^i)^2 + (p_z - p_z^i)^2}}{k} \tag{4.3}$$

$$N(p) = \begin{cases} 1(noise), & \mu_{D(p,k)} > n \times \delta_{D(p,k)} \\ 0, & otherwise \end{cases} \tag{4.4}$$

Where *D(p, k)* denoted a range of Euclidean distances between a point p in bush point clouds to each of its k nearest neighbors. p^i indicated the ith neighboring point of p. p_x, p_y,

and p_z ($p_x^{\cdot}$, $p_y^{\cdot}$, and $p_z^{\cdot}$) were the x, y, and z coordinates of the point p (or its neighboring point p^i) in point clouds. N denoted noise flag, and μ_D and δ_D were the mean and standard deviation of *D(p, k)*, and n was the scalar of standard deviation.

Based on some preliminary tests, k and n were set as 10 and 1 in this study, respectively. The SOR filter could eliminate scattered noise points but not point clusters of relatively large objects such as weeds. Density-based spatial clustering of applications with noise (DBSCAN) algorithm was used to further filter out point clusters of non-bush objects. After SOR filtering, the points were clustered using the DBSCAN algorithm, and the largest point cluster was selected as the clean bush point cloud ($PtCloud_{CleanBush}$) (Equation 4.5).

$$PtCloud_{CleanBush} = argmax_C(|C_i|),\ i = 1, 2, ..., n \tag{4.5}$$

Where $PtCloud_{CleanBush}$ denoted the clean point cloud of a blueberry bush, C denoted a point cluster that was identified using the DBSCAN algorithm, i was the index of identified point clusters, ranging from 1 to n, and $|\cdot|$ operator calculated the number of points in a point cluster.

Size-related Trait Measurement

Size-related traits were measured from the clean bush point clouds of individual bushes (Figure 4.2a). The measurement was to calculate bush dimensions and crown size, and the shape analysis was to find the best boundary curve of canopy and derive shape indices.

Bush Dimension Bush dimension parameters included bush height, width in-row (WIR), and width across-row (WAR), which were the maximum length of a bush along the z-, x-, and y-directions, respectively (Figure 4.2b). Bush volume was estimated using concave and convex hulls.

Crown Size Crown size is an essential dimensional parameter, affecting the configuration of catch plate (also known as fish scale) and ultimately the performance (e.g., ground loss) of machine harvesters. In the horticultural community, the term "crown" refers to a cross-section at a certain height [170]. In this study, crown size was defined as the bush diameter across-row at 15 cm from the bottom of main stems where the catch plates of OTR harvesters contact with plants. Cross-section points at such height were separated using height information, and subsequently distances from individual cross-section points to the cross-section median center were calculated (Figure 4.2c). A distance histogram was generated with a bin width of 0.05 m, and local minimal bin values (the bin value is less than that of two neighboring bins) were identified to group bins into different bin clusters. The first bin cluster contained points representing the bush crown, and the $95th$ percentile distance of the first bin cluster (between the first and second local minimal bin values) was used as the threshold to exclude cross-section points that were far away from the cross-section center. The retained points ($P_{retained}$) were fitted to a two-dimensional (2D) Gaussian distribution. The distribution mean and variances were used as the center and initial values of the semi-axes for an ellipse curve. Constrained optimization was used to find the minimum values of the two axes, so that the ellipse curve could reach a predefined point coverage (Equation 4.6). The ellipse's vertical diameter was the crown size of a bush.

$$\begin{aligned} min \quad & f(d_x, d_y) = d_x + d_y \\ subject\ to \quad & \frac{|P_{covered}|}{|P_{retained}|} \geq T_{coverage},\ P_{covered} = \{p | \frac{p_x^2}{d_x^2} + \frac{p_y^2}{d_y^2} \leq 1, p \in P_{retained}\} \end{aligned} \tag{4.6}$$

Where d_x and d_y were the horizontal and vertical diameters of the ellipse curve. $T_{coverage}$ was the predefined coverage (set as 0.9 in the present study). $P_{covered}$ is a set of points covered by the fitted elliptical curve and p indicated a point in the two sets ($P_{covered}$ and $P_{retained}$). $|\cdot|$ operator calculated the number of points in a point set.

Bush Shape Analysis

Due to the importance of machine harvester configuration and performance, the across-row bush shape was analyzed in the present study (Figure 4.2d). For each bush, clean bush point cloud was projected onto the y-z plane, and rasterized to a grayscale image using a grid size of $0.01 \times 0.01\ m^2$. As the shape analysis was conducted on images, the coordinate system used the top-left corner as the origin (0, 0) and x- and y-coordinates increased along the right and downward directions. In the grayscale image, pixel intensity represented the distance from a pixel to the starting point of a bush along the x-direction. The grayscale image was thresholded to a raw bush mask, and morphological operations were used to remove noise pixels and fill holes, generating the final bush mask for successive processing.

Landmark Point Detection A total of 11 landmark points were defined in the present study, including the center point of the bush topmost row (P1), left (P2) and right (P3) endpoints of the broadest horizontal cross-section, bush leftmost (P4) and rightmost (P5) endpoints, left (P6) and right (P7) canopy-crown separation points, left (P8), center (P9), and right (P10) endpoints of the bush bottom, and the center point of the canopy bottom (P11). P1 to P5 and P8 to P10 were detected based on their definitions, and five lines (LM, LL, LR, L1, and L2) were identified accordingly. LM was the center line of bush, which is a vertical line passing through P11. LL was the left border lines, which is a horizontal line that passes through the point close to the bush bottom in P2 and P4 and intersects with LM. L1 was the left outer boundary passing through P8 and the point close to bush bottom in P2 and P4. Similarly, LR and L2 were identified on the right side of the bush. LM, LL, and LR split the skeleton endpoints into four quadrants. Distances from individual endpoints in the left-lower quadrant to L1 were calculated. P6 was an endpoint in the left-lower quadrant that satisfied four criteria: 1) it had the most considerable distance to L1, 2) its distance to L1 was larger than a threshold (the median of all left-lower endpoints' distances plus the median absolute

deviation (MAD)), 3) it was close to the bush bottom as much as possible, and 4) it located above L1 (Equation 4.7 and Equation 4.8).

$$P6 = px_k, k = argmax(px_{i,n}) \; and \; px_i \in P_{P6candidate} \tag{4.7}$$

$$\begin{aligned} P_{P6candidate} = \{px_j | & D_{ll}(px_j) = max(D_{ll}) \\ & \cap \; D_{ll}(px_j) > Med(D_{ll}) + MAD(D_{ll}) \\ & \cap \; px_{j,n} > f_{L1}(px_{j,m})\} \end{aligned} \tag{4.8}$$

Where px indicated an endpoint pixel in the lower-left quadrant, and k, i, and j were the indices of P6 pixel, P6 candidate pixels ($P_{P6candidate}$), and endpoint pixels in the lower-left quadrant, respectively. $px_{j,m}$ and $px_{j,n}$ were the horizontal and vertical coordinates of a pixel px_j in images. D_{ll} denoted the set of distances from individual left-lower endpoints to L1. Med and MAD were operators to calculate the median and MAD values of a set. f_{L1} was the function of L1.

P7 could be identified using the same criteria in the lower-right quadrant. The vertical coordinate of the one close to the bush bottom in P6 and P7 was used to separate bush crown and canopy. P11 was the center point of the separation cross-section. It should be noted that P6 and P7 are not guaranteed to be present, because bush main stems may spread at a position very close to the ground or form branches at higher positions, resulting in an unclear separation between canopy and crown. If both P6 and P7 were missing, the canopy-crown separation line would merge with the bush bottom line, and consequently P11 became the same with P9.

Canopy Contour Fitting Canopy shape is another important factor affecting the performance of machine harvesting. Vase-shaped canopy is likely to reduce the total harvesting loss and bruising damage, leading to an improved harvest yield and quality. For a blueberry bush, canopy pixels were segmented in the bush mask image using the landmarks, and

canopy contour was extracted from the segmented part. Path curve was used to quantify the canopy contour shape. A path curve is defined by a single parameter λ: the path curve is a circle when λ equals to one, and becomes conical (or vase-shaped) when λ is larger than one (or less than one). As the curve position was considered, the function for drawing a two-dimensional (2D) path curve was defined using Equation 4.9.

$$\begin{aligned} f^{ptc}_{w_{ptc},h_{ptc},\lambda} &: t \in \mathbb{R} \rightarrow S_{ptc} \in \mathbb{R} \times \mathbb{R} \\ f^{ptc}_{w_{ptc},h_{ptc},\lambda}(t) &= a_{ptc} \cdot (\mp w'_{ptc} + offset_m,\ \tfrac{e^t h^2_{ptc}}{2} + offset_n) \\ a_{ptc} &= \tfrac{1}{e^{-\lambda t} + \frac{h_{ptc}}{2} e^t} \\ w'_{ptc} &= \tfrac{w_{ptc} \times (e^{-\lambda t_{max}} + \frac{h_{ptc}}{2} \times e^{t_{max}})}{2} \\ t_{max} &= \tfrac{ln(\frac{2\lambda}{h_{ptc}})}{\lambda + 1} \\ t &\in [-20, 20] \end{aligned} \tag{4.9}$$

Where f^{ptc} denoted the function for drawing a 2D path curve, and w_{ptc}, h_{ptc}, and λ were the width, height, and shape factor of a path curve. w'_{ptc} was the base value of horizontal coordinates of a path curve given the width of w_{ptc}. a_{ptc} reached its maximum value when t was t_{max}. Based on previous study [162], the domain of definition from -20 to 20 provided adequate range for covering typical object contours.

Detailed mathematical explanations of a typical function of path curve can be found in [168, 162]. To evaluate the fitness of a path curve to a bush canopy contour, an energy function was defined as Equation 4.10.

$$e_{ptc} = \frac{\sqrt{\sum_{i=1}^{n} dist(px_i^{ptc})}}{n} \tag{4.10}$$

Where e_{ptc} was the energy function of a path curve, px_i^{ptc} denoted the ith pixel in a path curve, and n was the total number of pixels in a path curve. $dist(px_i^{ptc})$ was a function to calculate the distance from px_i^{ptc} to the nearest pixel of the canopy contour.

Gradient descent approach was used to find the optimal path curve parameters (w_{ptc}, h_{ptc}, and λ) that minimized the energy function. In the present study, the gradient descent optimization would stop, if the path curve energy reached to a minimum value with no change in the following 5 iterations or the total iteration reached 500. The path curve with the minimum energy value was selected as the best fitting curve using certain initial values. To avoid fitting to local optima, w_{ptc} and h_{ptc} were initiated with various values. w_{ptc} ranged from $\lceil \frac{w_{canopy}}{2} \rceil$ to w_{canopy}, and h_{ptc} ranged from $\lceil \frac{h_{canopy}}{2} \rceil$ to h_{canopy}, with an increment of 5 pixels for both. w_{canopy} and h_{canopy} were the width and height of the bounding box of canopy. The path curve with the lowest energy value among the best path curves using various initial values was selected as the final fitting path curve of canopy for a bush.

Shape Index Calculation The detected landmark points and fitted path curve were used to derive five shape indices: 1) noncanopy-bush height ratio (NBR), 2) canopy vertical ratio (VR), 3) canopy aspect ratio (AR), 4) canopy curvedness (CN), and 5) canopy irregularity (IRR) (Figure 4.2d). The five indices quantified bush shape aspects for machine harvesting and agronomic management. NBR was to evaluate the potential of catch plate configuration, with high values for a wide range of placing catch plates of machine harvesters. VR and AR represented the canopy overall shape. Low VR and AR values (< 1) would represent a vase-shaped canopy with short fruit dropping height, which is preferred for machine harvesting; whereas high VR and AR values (> 1) would represent a conical canopy with greater fruit dropping height, which is not ideal for machine harvesting. CN and IRR were more related to agronomic management. In particular, high IRR values indicated an irregular canopy shape that requires agronomic actions such as pruning. Mathematical definitions of the five indices were provided in Equation 4.11.

$$
\begin{aligned}
NBR &= \frac{h_n}{h_{bush}} = \frac{h_n}{h_n + h_{upper} + h_{lower}} \\
VR &= \frac{h_{upper}}{h_{lower}} \\
AR &= \frac{h_{ptc}^{fitted}}{w_{ptc}^{fitted}} \\
CN &= \frac{|S_{ptc}^{fitted}| - |S_{triangles}|}{|S_{ptc}^{fitted}|} = \frac{|S_{ptc}^{fitted}| - |S_{upper}| - |S_{lower}|}{|S_{ptc}^{fitted}|} \\
IRR &= \frac{|\complement S_{ptc}^{fitted}|}{|S_{canopy}|} = \frac{|S_{canopy}| - |S_{ptc}^{fitted}|}{|S_{canopy}|}
\end{aligned}
\tag{4.11}
$$

Where h_n, h_{upper}, and h_{lower} were heights of non-canopy part, and canopy upper (with vertices of P1, P2, and P3) and lower (with vertices of P11, P2, and P3) triangles. h_{ptc}^{fitted} and w_{ptc}^{fitted} were the height and width of the fitted canopy path curve. S_{upper}, S_{lower}, S_{ptc}^{fitted}, and S_{canopy} denoted sets of pixels within the canopy upper and lower triangles, fitted path curve, and the canopy contour. $|\cdot|$ operator calculated the number of pixels in a given set.

4.3.3 Performance Evaluation

It is important and necessary to evaluate the accuracy of measured size-related traits by the proposed method. Five size-related traits were manually measured for reference, including bush height, WIR, WAR, volume, and crown size. Bush height, WIR, WAR, and crown size were measured using a measuring tape based on their definitions, whereas bush volume was estimated using a cylindrical model (see Figure S7 in Supplementary Materials). A bush was visually and vertically segregated into layers with an interval of 5 cm, with each layer being assumed as a cylinder. Circumference of each layer was manually measured, and thus diameter could be estimated to calculate the layer volume. The summation of all layer volumes was used as a reference value of bush volume. Simple linear regression analyses were performed between sensor and manual measurements for size traits. R^2 and RMSE were used as indicators to evaluate the accuracy of sensor measurements. In addition, MAEs and MREs were calculated as additional parameters for accuracy evaluation. All analyses

were conducted in MATLAB (Statistics Toolbox 2017b, The MathWorks Inc. Natick, Massachusetts, USA).

Statistical analyses were conducted on the extracted crown size and shape descriptors to evaluate their usefulness of identifying optimal bush architecture suitable for mechanical harvesting. Although there were 16 genotypes in the North Carolina field, they have been selected for mechanical harvesting and are extremely similar in terms of bush architecture. Thus, it was reasonable to treat all genotypes in the North Carolina field as one genotype group (hereafter, NCSU_MH group). An additional point cloud dataset was used to increase the diversity of bush architecture, containing three highbush blueberry cultivars (Star, Meadowlark, and Farthing) with distinctive bush shapes [149]. In summary, crown size and shape descriptors were extracted using 367 bushes from five genotype groups (20 bushes in the O'Neal group, 222 in NCSU_MH, 42 in Star, 43 in Meadowlark, and 40 in Farthing). As sample sizes were dramatically different among groups, Kruskal-Wallis tests (nonparametric equivalent to analysis of variance test) were performed on extracted crown size and shape descriptors to identify statistical differences among the five groups. Kruskal-Wallis tests were conducted in R 3.4.2 (R Development Core Team, 2008) (package asbio) using a significance level of 0.05.

4.4 Results

4.4.1 Reconstructed Point Cloud Data

Reconstructed results contained a scanning trajectory and raw point clouds for scanned areas (Figure 4.3). In trajectories, the lowest position indicated the starting and ending points of each data collection session, and starting and ending points could be further differentiated based on the relative movement direction. The starting point was the origin in each scanned point cloud. The waveform of trajectories reflected the oscillation of the LiDAR node.

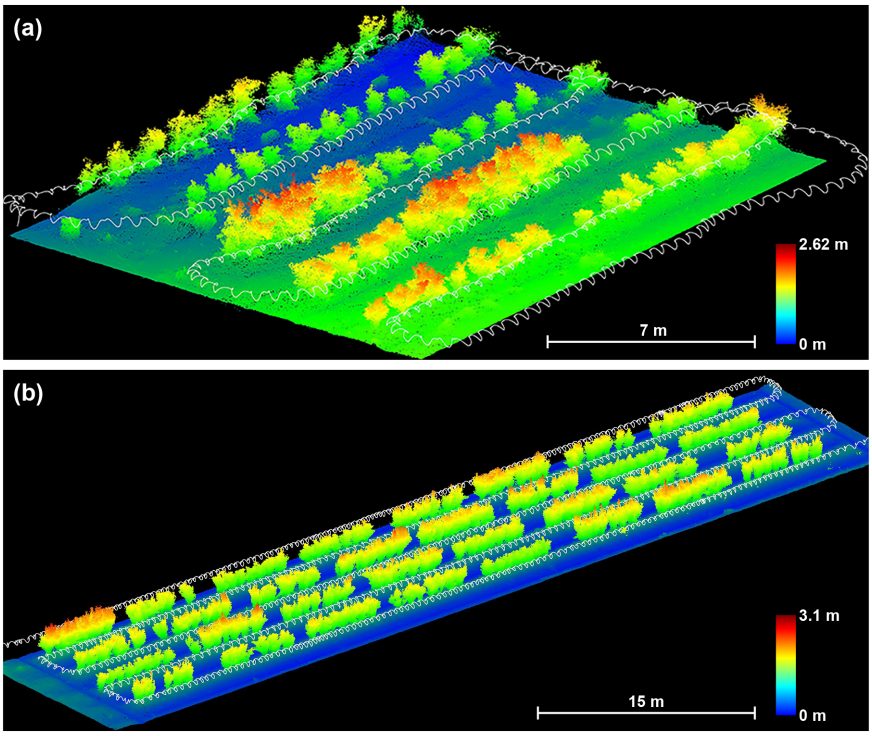

Figure 4.3: Point cloud data collected using a handheld mobile laser scanner in two fields: (a) Horticulture Farm of the University of Georgia and (b) Horticulture Research Station in North Carolina. The displayed point clouds were cropped to remove irrelevant objects. White lines indicate the walking path and sensor node oscillations of data collection sessions, and star and triangle markers show the starting and ending points of the walking path.

Raw point clouds were rendered by color using point height information, with blue to red representing low to high values. The blueberry field at the Horticulture Farm of the University of Georgia showed an obvious sloped terrain: the southern side (top part in the figure) was at a lower elevation than the northern side (bottom part in the figure), resulting in different height values (colors) of ground points (Figure 4.3a). In contrast, the terrain elevation was relatively level (ground points looked in similar blue colors) in the field at the Horticulture Research Station in North Carolina, but clear color contrasts were observed between furrows (dark blue) and plant beds (turquoise). These differences increased data variability and could be particularly challenging for point cloud preprocessing, but algorithms developed

in this study successfully removed ground and noise points, suggesting its generalizability to various field conditions (see the section of ground removal and bush point denoising in Supplementary Materials).

4.4.2 Accuracy of Size Measurements

The scanner achieved the nominal measurement accuracy (2–3 cm) when measuring objects with standard shapes (see the section of validation of measurement accuracy in Supplementary Materials), providing a performance baseline to evaluate accuracies of measuring size-related traits. Generally, correlation was high (R^2 = 0.92–0.95) between sensor and manual measurements of bush height and width (Figure 4.4a to Figure 4.4c). The RMSE and mean absolute error (MAE) of bush height were comparable with those of objects with standard shapes, whereas the RMSE and MAE of width were two times larger. This occurred primarily because blueberry bushes were non-rigid objects that could be swayed by wind during data collection. Bush movements had relatively small effects on z-direction, resulting in little or no change in height measurements. However, the movements would have substantial effects on the x- and y-directions, leading to large errors in width measurements. Nonetheless, mean relative errors (MREs) (around 5%) indicated that those errors of width measurements were acceptable.

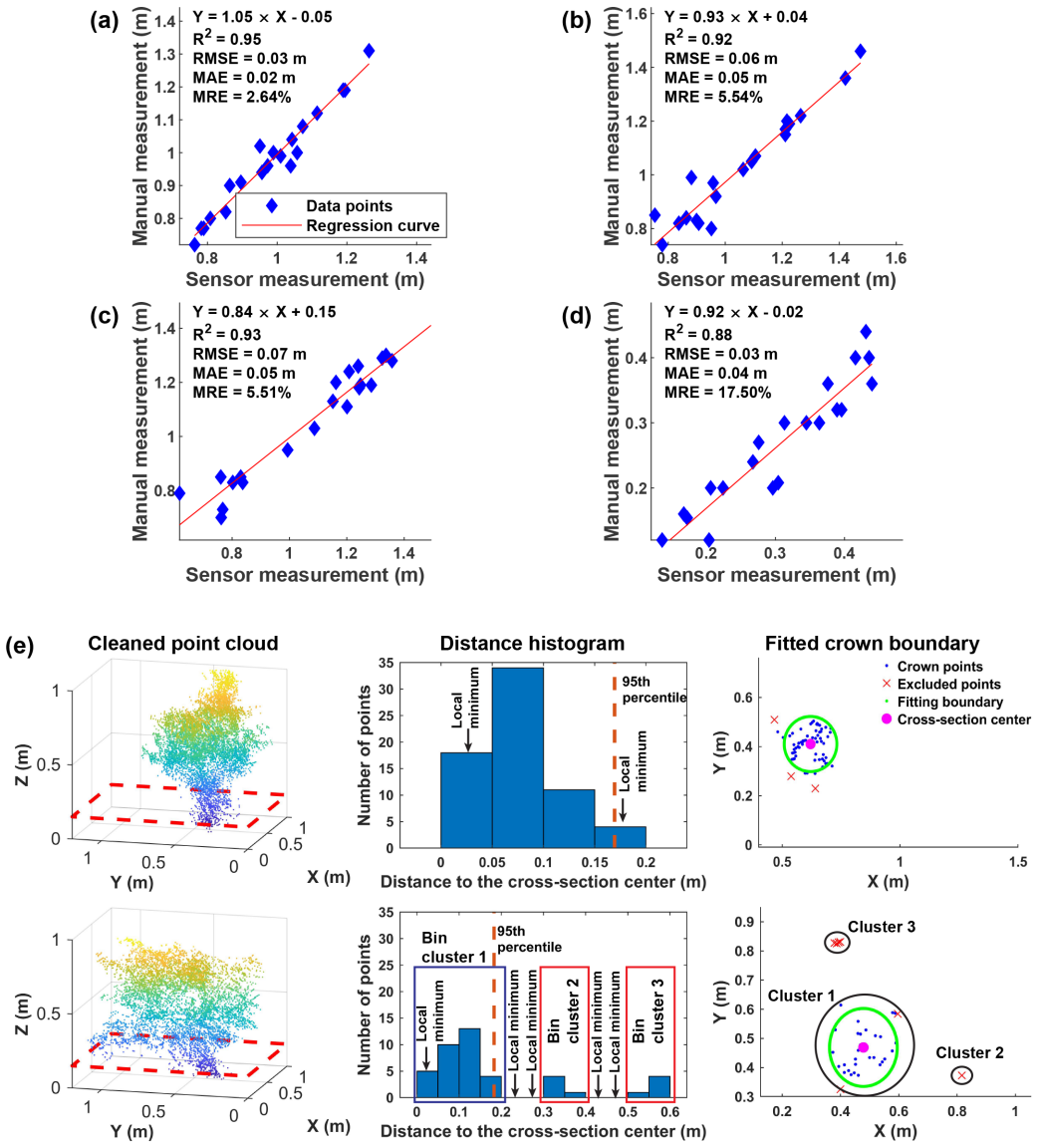

Figure 4.4: The accuracy of sensor measurements and the efficacy of crown size measurements. (a), (b), (c), and (d) are regression results between sensor and manual measurements of bush height, width in row (WIR), width across row (WAR), and crown size; and (e) is the efficacy of the present algorithm for measuring upright (top chart) and inclined (bottom chart) bushes.

For bush volume, manual measurements were less than convex hull volumes but greater than concave hull volumes (see Figure S4 (a) and (b) in Supplementary Materials). Compared with concave hull volumes, convex hull volumes were better correlated ($R^2 = 0.85$)

with manual measurements, because both convex hull and manual measurements included space that was not occupied by branches. It was noteworthy that convex hull method included considerably more of unoccupied space between branches than the manual estimation method, resulting in high MAE (0.21 m^3) and MRE (104%) of volume measurements. In particular, the convex hull method tended to substantially overestimate (around 120%) the volume of bushes with irregular architecture and tall crown (compare (c) and (d) in Figure S4 in Supplementary Materials). This occurred because irregular architecture and tall crown led to large hollow (or empty) areas among (or below) bush canopies that would be included by the convex hull method. However, the manual method used a short height interval (0.05 m in the present study) and significantly reduced the amount of hollow/void areas in volume estimation. Compared with convex hull volumes, concave hull volumes showed a lower correlation with manual measurements, but they were closer to the actual reference values (much smaller MAE (0.05 m^3) and MRE (19%)). A potential reason was that the amount of space included by the convex hull method was more related to plant size changes than that excluded by the concave hull method. When the plant size increases, the convex hull consistently includes extra space due to the expansion of plant points, but the concave hull method may or may not exclude space depending on the local surface. The concave hull method could match with manual measurements for a flat surface, while it could exclude a large space for a curved surface such as the transition section between canopy and non-canopy parts. The bushes had a large variation of the surface curvature, leading to inconsistent changes of space exclusion by the concave hull method and thus a lower correlation with the plant size changes. For bushes with a relatively regular shape, if point clouds were dense enough, the concave hull volume should be the most accurate measurements; otherwise, it represented the lower limit of bush volume. For instance, if a bush grew in a more regular shape, the concave hull volume was closer to the manual measurement (compare (c) and (d) in Figure S4 in Supplementary Materials).

A high correlation ($R^2 = 0.88$) was also achieved between sensor and manual measurements of crown size (Figure 4.4d). Both RMSE (0.03 m) and MAE (0.04 m) were close to the nominal instrument accuracy, indicating a high measurement accuracy of the present algorithm. Compared with a previous study, the correlation (R^2) increased from 0.56 to 0.88 and the RMSE decreased from 0.06 to 0.03 m, both of which were substantially improved. These improvements were achieved due to the appropriate exclusion of non-crown points (Figure 4.4e). For upright bushes, the cross section usually contained one core point cluster with several points that were somewhat away from the cluster (see first row in Figure 4.4e). Direct use of the cross-section points would result in a large error of crown size measurements, regardless of using either a fitted diameter or width across-row of the cross section as the crown size. On the contrary, the present algorithm filtered out distant points using the $95th$ percentile (an empirical value) of distances to the cross-section center, reducing the measurement error. In addition, main bush branches would not naturally distribute as a circle, so an ellipse shape was better for crown fitting and thus crown size measurement. For inclined bushes, the cross section mostly contained several point clusters (see bottom charts in Figure 4.4e). The cluster closest to the cross-section center represented the actual crown, whereas the clusters away from the cross-section center were points of branches. Thus, the measurement accuracy was improved by using only the closest cluster.

4.4.3 Efficacy of Crown Size and Shape Descriptors for Bush Identification

Crown size and shape descriptors showed statistical differences among five genotype groups including four cultivars and one research population bred in North Carolina for mechanical harvesting (NCSU_MH group hereafter) (Figure 4.5). The crown size of O'Neal cultivar was statistically more significant than that of the remaining groups, whereas crown sizes of the remaining groups were in a similar range although Meadowlark cultivar had the smallest crown size. This occurred primarily due to two reasons. Firstly, bushes in the

five groups were treated with different agronomic practices such as pruning of low-angled branches originating near the ground and large upright canes away from the core cluster. O'Neal bushes were planted in a research farm and not pruned for two years before the data collection, resulting in a larger crown size. On the contrary, other group bushes were routinely pruned (based on commercial production guideline) and regulated (only for Star, Meadowlark, and Farthing), leading to a smaller crown size with less variations. Secondly, the five groups were being evaluated for different breeding targets. The four cultivars were bred primarily for features such as high fruit quality and size, whereas the NCSU_MH group have been selected for mechanical harvesting that requires a small crown. Catch plates on OTR harvesters are pivot mounted on a rail on both sides of the harvester frame and overlap with neighboring plates (see Figure S6 (a) and (b) in Supplementary Materials). When the harvester moves to contact blueberry plants, catch plates are pushed to sides, allowing bush canes to go into the harvester, and then the plates retract to cover empty areas. When catch plates (e.g., fishscales) do not fully retract and return to a crown size area at the base of the plant, it would create an opening area where detached blueberries can potentially fall through to ground (e.g., ground loss). Smaller crown means less ground loss. Thus, the NCSU_MH group should present desired crown size even without crown regulation.

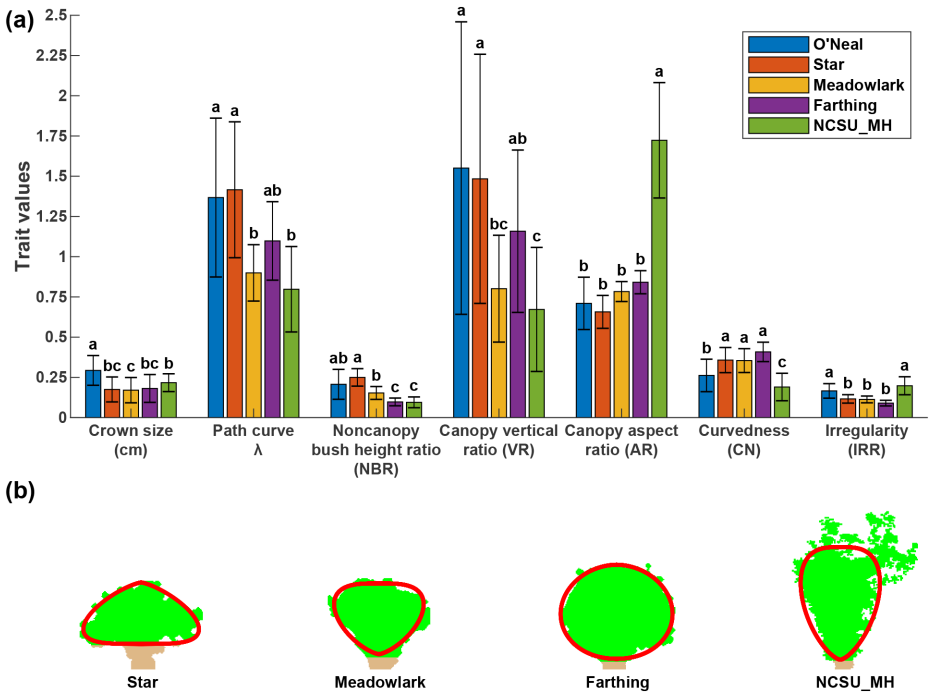

Figure 4.5: Crown size and shape analysis results of the five blueberry groups. (a) Statistical analysis results of the extracted crown size and shape descriptors and (b) fitted path curves of representative bushes. Groups with different letters are statistically significant with each other, and group mean values of each index are sorted alphabetically. In (b), green and brown colors are used to render canopy and non-canopy parts of individual blueberry plants, and red curves are the fitted path curves. No representative bush was selected for the O'Neal group due to the large variation of crown size and bush shape in the group.

All shape descriptors showed significant differences between at least two genotype groups, suggesting that the shape descriptors could be used for identifying blueberry genotypes with different bush architecture. The NCSU_MH group had the least λ values (0.8$\pm$0.26), followed by Meadowlark (0.89$\pm$0.17), Farthing (1.1$\pm$0.24), and Star (1.42$\pm$0.42). By definition, these λ value ranges indicated in general a vase-shaped canopy for NCSU_MH and Meadowlark groups, a round canopy for Farthing, and a conical canopy for Star (Figure 4.5b). Thus, NCSU_MH and Meadowlark bushes would have an optimal shape (vase shape) for mechanical harvesting, which agreed with human subjective evaluation. Meadowlark also can be

grafted on sparkleberry (*Vaccinium arboreum*) rootstock with monopodial growth habit, which creates even smaller crown diameter [171]. It should be noted that canopy vertical ratio (VR) showed the same trend and statistical results as λ. This was because VR essentially quantified the location of the widest canopy cross section where branches expanded horizontally. An ideal vase shape would have the horizontal expansion at a higher position of the canopy, leading to VR values less than 1, whereas a conical shape would have the opposite pattern. A round shape would result in VR values equal to 1. Although VR showed the same efficacy as λ in overall shape quantification, VR values had larger variations than λ, which presented a concern of using it for differentiating blueberry genotypes with a small number of replications. The capability of using extracted traits for genotype differentiation needs to be further tested when a smaller number of replications is used.

In contrast to crown size, λ, and VR, other shape descriptors (non-canopy-bush height ratio (NBR), canopy aspect ratio (AR), canopy curvedness (CN), and canopy irregularity (IRR)) could not be used for bush shape evaluation based on simple rules (e.g., small crown is preferred), requiring more domain knowledge for proper interpretation and use. Star had the highest NBR value indicating the tallest non-canopy part, which is good for mechanical harvesting due to more positions for configuration of harvester catch plates. However, an excessively tall crown is not desired because it may result in a yield reduction more substantially than the ground loss due to mechanical harvesting. In commercial field setup for mechanical harvesting, low hanging branches are pruned to eliminate their interference with the catch plates and minimize bush crown size to prevent excessive ground loss, but the pruning cannot be aggressive to impact yield. Thus, the NBR index needs to be used as a balance factor for breeding blueberry genotypes suited to mechanical harvesting. AR values of four cultivars were significantly lower than the NCSU_MH group. A low AR value indicates an oblate canopy, which is desired owing to a short dropping height (and thus reduced external impacts) for berries, but if the canopy is excessively oblate, mechanical harvesters

may damage branches as well as berries on those branches, decreasing harvest yield and berry quality. It was also noteworthy that the AR index might reflect breeding preferences due to different growing environments. The four cultivars are widely grown in southern Georgia, whereas the NCSU_MH group has been bred in a research station along the coastal area where wind would be generally strong during blueberry vegetative and reproductive growth stages (March–May). Thus, the use of AR index also requires considerations of other factors to evaluate the fitness of bush shape for mechanical harvesting. CN evaluated the curvedness of bush canopy contour and IRR indicated the likelihood of having abnormally extended portions, both of which provided useful information for agronomic management such as pruning. In particular, the IRR values indicated the management practices conducted on bushes. The ONeal and NCSU_MH groups showed statistically lower CN values and higher IRR values, suggesting a suboptimal bush architecture regarding agronomic management, but they had different reasons: the O'Neal group was due to insufficient management (no pruning for 2 years), whereas the NCSU_MH group was due to different growth periods. Data collection of the NCSU_MH group was conducted in March when bushes had little or no leaf, and consequently bushes were expected to be more irregular. Nonetheless, NBR, AR, CN, and IRR require additional considerations from various aspects for proper interpretation and cannot be used as simple criteria for bush selection and management.

4.4.4 Visualization Tool for Identification of Optimal Bush Architecture for Mechanical Harvesting

As the crown size and λ could be used for bush evaluation based on simple criteria, a visualization tool (scatter plot) was generated to identify an optimal bush architecture for mechanical harvesting (Figure 4.6). Two standard axes were made based on requirements of crown size and bush shape for mechanical harvesting. For an ideal bush architecture, the crown size needs to be less than 20.32 cm [171] and the bush needs to be vase-shaped (λ less than 1), whereas for an acceptable bush shape, the crown size can be increased to

30.48 cm [171] and the bush can be slightly conical (λ less than 1.1). The two standard axes split the space into four quadrants: 1) the upper-left is for bushes with desired crown size but undesired bush shape; 2) the upper-right is for bushes with undesired crown size and shape; 3) the lower-left is for bushes with desired crown size and bush shape, and 4) the lower-right is for bushes with desired shape but undesired crown size. When using the ideal criterion, Meadowlark was the only group having most bushes with ideal crown size and bush shape. When using the acceptable criterion, most Farthing and NCSU_MH bushes met the requirements of crown size and bush shape for mechanical harvesting, whereas the O'Neal group primarily laid in the first quadrant where both crown size and bush shape were not ideal for mechanical harvesting. Star bushes were mostly identified in the second quadrant where the bush shape was not acceptable for mechanical harvesting.

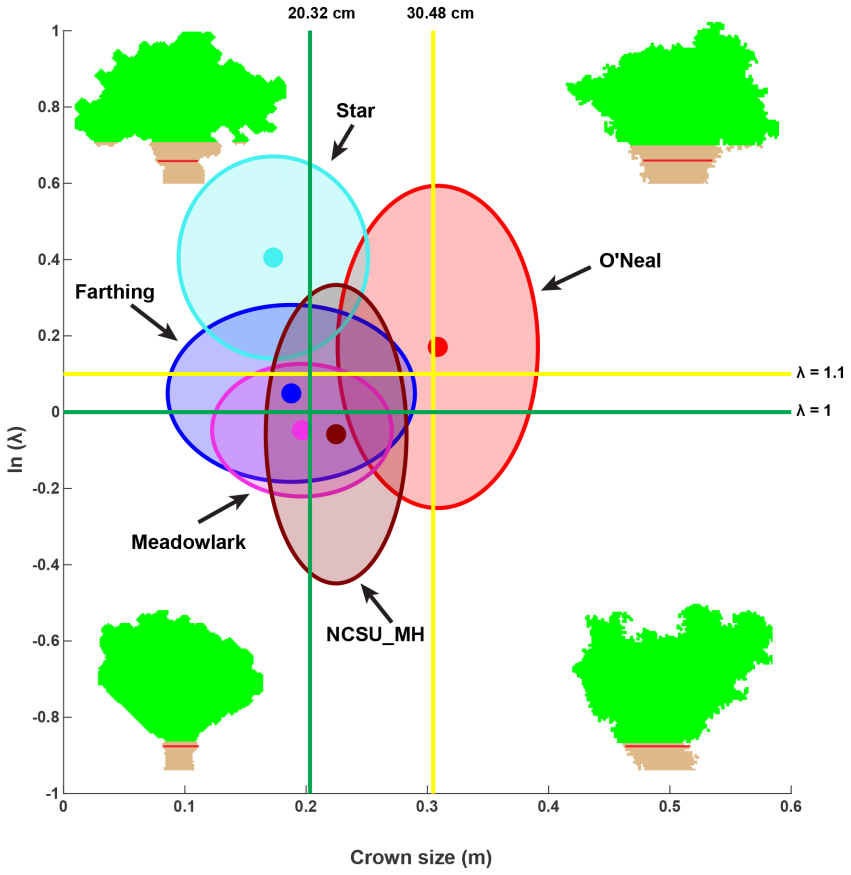

Figure 4.6: Scatter plot of crown size and the natural logarithm of λ for 145 bushes. Green (or yellow) axes indicate the value limits of crown size and path curve λ for bushes well suited to (or acceptable for) mechanical harvesting. Solid circles represent the center of individual clusters. For each representative bush, green and brown colors are used to render the canopy and non-canopy parts, and the red line in each bush silhouette indicates the height where the crown size was measured.

4.5 Discussion

The data processing pipeline has demonstrated the feasibility of using a handheld mobile laser scanner to measure size-related traits and shape descriptors of blueberry bushes in the field. For data collection, the scanner has a dynamic and expansive sensing perspective, which is a major advantage over aerial imaging systems and terrestrial LiDAR systems in

which cameras or LiDARs acquire data from an individual angle. The scanner node keeps moving along and across the operator's movement direction, so a wide range of sensing angles are used to dramatically reduce the possibility of missing points due to object occlusions. However, the scanning throughput of the handheld scanner is relatively low. If the operator keeps oscillating the scanner node and walks at 1.4 m/s (a regular walking pace), the scanning throughput is 0.42 ha/h. Considering the weight of the scanner with necessary accessories (2.5 kg in total weight), operators may become fatigued after collecting data for a period. In practice, it is also difficult for operators to continuously oscillate the sensor node, so oscillating sensor node can be problematic in a long-time data collection session, which reduces the diversity of sensing angles and thus the data quality. To increase the scanning throughput and avoid human fatigue issues, it is necessary to integrate the scanner with motorized vehicles for autonomous data collection. In fact, the scanner was originally developed for both handheld and vehicle-based applications [153], so it can be mounted on a motorized platform (e.g., a gator utility vehicle) with modifications to improve the data collection throughput.

The present data processing pipeline can accurately extract size-related traits, especially the crown size. Compared with a previous study [149], the measurement accuracy of crown size has been significantly increased due to not only the improved measurement algorithm but also a different way of collecting point cloud data. As for the aforementioned advantage, the scanner can have various sensing angles, and some angles (e.g., parallel with bush crown) can be particularly useful for acquiring points of the bush crown that is usually occluded by bush canopies from the nadir and top-to-bottom oblique views. If raw point clouds miss many points of the bush crown (or other bush parts), it is not possible to improve the measurement accuracy of algorithms. Although the processing pipeline is independent of data collection systems, the processing performance highly depends on the quality of acquired data that are affected by data collection systems to a certain extent. Due to practical reasons

(e.g., easy to measure ground truth data), small to mid-size bushes were used to evaluate the accuracy of the presented approach, which avoided a potential issue of branch entanglement between neighboring plants. The entanglement usually introduces difficulties in accurate segregation of individual plants (especially the upper canopy), which could dramatically affect the measurement of WIR and bush volume.

The extracted crown size and shape descriptors (primarily λ) provide objective evaluation and measurements for identifying bushes suitable for mechanical harvesting. In particular, the visualization tool (scatter plot of crown size and λ) is particularly useful for rapid determination of optimal bush architecture. NBR and AR indices can be used as balance factors to select bushes suited to a particular harvesting machine or growing environment and maintain other desired features such as yield. CN and IRR can be incorporated into agronomic management decision process such as pruning. However, the use of the four parameters (NBR, AR, CN, and IRR) highly depends on breeding and management purposes. Thus, thresholds or value ranges of the parameters for optimal bush architecture need to be determined with specific domain purposes and may vary dramatically among applications. In addition, all the extracted traits could be used by harvester manufacturers to improve the design of fruit catching system. The five genotype groups were selected because they had distinctive bush architecture. With a large number of replications (at least 20 reps per group), it would be relatively easy to differentiate the groups from each other using the crown size and shape descriptors. We acknowledge that it is necessary to conduct successive studies involving a wide variety of genotypes with fewer replications, so the statistical power of extracted traits can be further tested for genotype differentiation.

4.6 Conclusions

The data processing pipeline presented in this study accurately measured size-related traits and bush shape from point cloud data collected by the handheld mobile laser scanner in

the field. Shape descriptors were used to identify bushes with desired features for machine harvesting, and bushes with non-ideal shapes that required pruning actions. Thus, the present processing pipeline with the data collection instrument is particularly useful for blueberry breeding programs and farm management. Future studies will focus on the development of autonomous data collection system and experiments of using shape descriptors for genotype differentiation in a large-scale field.

Chapter 5

Ground based Hyperspectral Imaging to Characterize Canopy-Level Photosynthetic Activities[1]

[1]Jiang, Y., Snider, J. L., Li, C., Rains, G. C., and Paterson, A. H. Submitted to *ISPRS Journal of Photogrammetry and Remote Sensing*, July 19, 2019.

5.1 Abstract

Improving plant photosynthesis provides the best possibility for increasing crop yield potential, which is considered a crucial effort for global food security. Chlorophyll fluorescence is an important indicator for the study of plant photosynthesis. Previous studies have intensively examined the use of spectrometer, airborne, and spaceborne spectral data to retrieve solar induced fluorescence (SIF) for estimating gross primary productivity and carbon fixation. None of the methods, however, had spatial resolution and scanning throughput that are suitable for applications at the canopy and sub-canopy levels, thereby limiting photosynthesis analysis for breeding programs and genetics/genomics studies. The goal of this study was to develop a hyperspectral imaging approach to characterize plant photosynthesis at the canopy level. An experiment field was planted with two cotton cultivars that received two different treatments (control and herbicide treated), with each cultivar-treatment combination having 8 replicate 10-m plots. A ground mobile sensing system (GPhenoVision) was configured with a hyperspectral module consisting of a spectrometer and a hyperspectral camera that covered the spectral range from 400 nm to 1000 nm with a spectral sampling resolution of 2 nm. The system acquired downwelling irradiance spectra from the spectrometer and reflected radiance spectral images from the hyperspectral camera. On the day after 24 h of the DCMU application, the system was used to conduct six data collection trials in the experiment field from 0800 h to 1800 h with an interval of 2 hours. A data processing pipeline was developed to measure SIF using the irradiance and radiance spectral data. Diurnal SIF measurements were used to estimate the effective quantum yield and electron transport rate, deriving rapid light curves (RLCs) to characterize photosynthetic efficiency at the group and plot levels. Experimental results showed that the effective quantum yield estimated by the developed method highly correlated with those measured by a pulse amplitude modulation (PAM) fluorometer. In addition, RLC characteristics calculated using the developed method showed similar statistical trends with those derived using the PAM data. Both the RLC and PAM data agreed with destructive growth analyses. This suggests that the developed method can

be used as an effective tool for future breeding programs and genetics/genomics studies to characterize plant photosynthesis at the canopy level.

5.2 Introduction

The global population is likely to exceed 10 billion by 2050, presenting great challenges for agriculture [3]. To fulfill the needs of the rapidly growing population, the current agricultural yield must be doubled by that time, which translates into an annual increase of 1.75% total factor productivity (TFP) [4]. Currently, the global TFP growth rate is approximately 1.5%, leaving a gap of 0.25% annually. Even worse, the TFP growth rate is only approximately 0.96% in developing countries, which is far behind the required growth rate. Cotton (*Gossypium*) is one of the most important textile fibers in the world, accounting for about 25% of total world textile fiber use [172]. Thus, improvement of cotton production is vital to fulfilling the fiber requirements of over nine billion people by 2050 [100].

As with all agricultural crops that have reproductive structures of economic importance, the yield of cotton can be expressed as a function of total seasonal light interception, radiation use efficiency, and harvest index [173]. Thus, yield improvement can be achieved by increasing any one of these three variables. The "Green Revolution" introduced dwarfing genes into the most important C3 cereal crops (e.g., rice and wheat), allowing an increased biomass allocation to grain with a reduction in the total aboveground biomass (thus an increased harvest index). Breeding programs have continued to increase carbon allocation into grain [174], and in cotton specifically, genetic yield improvement historically has also been associated with an increase in biomass partitioning to reproductive units (bolls) [175, 176]. [177], however, suggested that future yield improvements in high potential environments would likely be achieved by 1) selecting cotton varieties that exhibit a more indeterminate growth habit (i.e. capitalize on the high insolation levels experienced in long growing season environments); and 2) increasing photosynthetic efficiency either through breeding or biotechnology efforts.

Photosynthesis is a process that converts radiant energy into biochemical energy, and is the basis of plant growth. In contrast with efforts to breed for desirable plant growth habits or greater harvest index, photosynthetic improvement has not yet been achieved for breeding programs, and remains a promising avenue for increasing agricultural productivity in the future. In addition, variations in photosynthetic efficiency can be used as indicators of plant stress, which can be used for selecting genotypes with high levels of stress tolerance or for making management decisions at the field scale [178]. Furthermore, there is a tremendous amount of interest in using remote sensing to model gross primary productivity of natural ecosystems according to the original framework of [173] [179]. In order to achieve this, the photosynthetic efficiency of the canopy must be estimated.

Chlorophyll fluorescence parameters are often used to evaluate photosynthetic performance and stress in plants [180]. These parameters include three measurable variables, i.e., minimal fluorescence (F_0 or $F_0^{'}$) when photosystem II (PSII) centers are open, maximal fluorescence (F_m or $F_m^{'}$) when PSII centers are closed, and steady state fluorescence (F_s or $F_s^{'}$), and several derived variables such as variable fluorescence ($F_v = F_m - F_0$ or $F_v^{'} = F_m^{'} - F_0^{'}$) and difference in fluorescence ($F_q = F_m - F_s$ or $F_q^{'} = F_m^{'} - F_s^{'}$) between F_m (or $F_m^{'}$) and F_s (or $F_s^{'}$). Variables denoted by prime are for light-adapted states, or otherwise for dark-adapted states. These parameters can be used to calculate the maximum ($\frac{F_v}{F_m}$ or $\frac{F_v^{'}}{F_m^{'}}$) and operating ($\frac{F_q^{'}}{F_m^{'}}$, a.k.a., Φ_{PSII}) efficiencies of PSII photochemistry, both of which are useful for photosynthetic efficiency evaluation and plant stress detection [181].

There are active and passive sensing approaches to measure chlorophyll fluorescence parameters. Active techniques include pulse amplitude modulation (PAM) [182] and laser induced fluorescence transient (LIFT) [183], both of which can emit predefined light to measure minimal fluorescence (F_0), steady state fluorescence (F_s), and maximal fluorescence (F_m), and can calculate variable fluorescence (F_v). Based on active techniques, portable

instruments (e.g., PAM fluorometers) have been developed and widely used for photosynthesis studies [184, 185, 186, 187, 188, 189]. Active techniques rely on artificial illumination, however, which limits the use for large scale applications [190]. Passive techniques retrieve chlorophyll fluorescence emission excited by solar illumination (natural sunlight), which is termed as solar induced fluorescence or sun induced fluorescence (SIF). Upwelling radiance from plants under solar illumination is a mixture of SIF and surface reflectance, and it is feasible to decouple SIF signals from the upwelling radiance in Fraunhofer lines of the solar spectrum in which irradiance is substantially reduced because of atmospheric absorption (e.g., hydrogen and oxygen). In the red and far-red spectral range, three Fraunhofer lines are frequently used for SIF retrieval, including $H\alpha$ at 656 nm because of hydrogen and O_2-B at 687 nm and O_2-A at 760 nm because of oxygen. Common SIF retrieval approaches include conventional Fraunhofer line discrimination (FLD), improved FLD variants, and reflectance-based ratios [191]. Many studies explored the use of ground based spectrometers to retrieve SIF of plant leaves and reported high correlations between SIF measurements and fluorescence measured using active techniques (e.g., PAM fluorometers) [191, 192, 193, 194]. Large national research institutions (e.g., European Space Agency, EPA and National Aeronautics and Space Administration, NASA) launched programs to investigate the use of hyperspectral imagery sensed remotely from planes and satellites to monitor SIF changes at the regional and global levels, to estimate carbon fixation and gross primary productivity (GPP). These data would be invaluable for research studies and policy-making to secure the food supply [195, 196, 197, 198, 199]. Spectrometer-based approaches can provide the highest accuracy of measurement location (a specific point on a leaf), but have an obvious limitation in the scanning throughput (point by point). By contrast, airborne and space-borne solutions provide a substantially faster scanning throughput, but have compromised spatial resolutions (sub-meter to meters). Neither approach, therefore, would be suitable for breeding programs and plant-science studies at the canopy or sub-canopy levels. A few studies examined the use of ground hyperspectral imaging to measure SIF, showing high

measurement accuracies and the potential of spatial variation analysis [200, 201]. However, these studies were restricted to the scope of instrumentation validation and did not utilize retrieved SIF for characterizing whole-canopy photosynthesis.

A particularly notable limitation to using SIF to estimate canopy-level photosynthetic efficiency is the method for estimation of maximal fluorescence (F'_m) at the canopy level, which has not been reported previously. If F'_m could be determined for the canopy using SIF, real-time estimates of crop-level photosynthetic efficiencies could be obtained using passive sensing approaches. Direct measurement of F'_m faces a significant challenge: the maximum intensity of solar illumination on the ground is far less than a "saturating" flash intensity used by PAM fluorometers, so maximal fluorescence at the canopy level cannot be directly measured. In fact, it has been demonstrated that leaves of plants previously acclimated to high light environments often do not close all reaction centers despite exposure to such a "saturating" flash intensity, preventing even direct measurement of F'_m by using PAM fluorometry. A multi-phase flash approach was developed, however, to estimate F'_m and electron transport rate (ETR) without the "saturating" light [202]. Using this method, a leaf sample is successively exposed to flash intensities in either ascending or descending order, and the intensity of chlorophyll fluorescence is quantified at each step. Fluorescence intensity (F) is then plotted versus the reciprocal of photosynthetically active radiation (PAR) and a function is fit to the data to obtain the y intercept and an estimate of apparent F'_m at an infinite light intensity when all reaction centers would hypothetically be closed. Since it is often necessary to expose single leaf samples to different PAR levels in succession to obtain an estimate of F'_m, it might also be possible to estimate F'_m for the canopy by taking advantage of natural, diurnal variation in PAR and SIF.

The overall goal of this study was to develop a ground-based hyperspectral imaging approach to characterize photosynthesis at the canopy level. Specific objectives were to

1) develop a ground-based hyperspectral imaging system to measure diurnal SIF changes using the FLD method; 2) calculate effective quantum yield and ETR to derive light curves for characterizing plant photosynthesis; and 3) validate the efficacy of the characterization method by comparing the hyperspectral imaging derived measurements with PAM-derived measurements and destructive crop growth analysis.

5.3 Materials and Methods

5.3.1 Plant Materials and Experiment Design

To evaluate the utility of diurnal SIF measurements for estimating whole canopy photosynthetic efficiency, a field experiment was established at the University of Georgia Lang-Rigdon research farm near Tifton, GA. Two cotton cultivars (PHY 841 RF and PHY 340 W3FE) were planted on June 19, 2018 at a 2.5 cm depth and a seeding rate of 11 seeds per m^2. Individual plots were 1 row with a length of 9 m and an inter-row spacing of 0.91 m. The two cultivars planted represent two different species of cotton adapted to different cotton production regions of the southern United States (US) and were previously shown to differ in leaf anatomical characteristics and photosynthetic response to light intensity [203]. PHY 841 RF is a Pima cotton (*Gossypium barbadense*) cultivar widely grown in the arid southwestern US, whereas PHY 340 W3FE is an upland cotton (*Gossypium hirsutum*) cultivar primarily grown in the humid southeastern US. To generate large differences in photosynthetic efficiency of the canopy, once the crop had reached the initial stages of floral bud development (referred to as "squaring"), plots received one of two possible herbicide treatments. Untreated control plots received only water applied as a foliar spray at a rate of 130 L ha-1. Diuron (3-(3,4-dichlorophenyl)-1,1-dimethylurea, also known as DCMU) treated plots had a 41% solution of DCMU (w/w; Diuron 4L) applied at a rate of 2.35 L/ha and a total application volume (water plus Diuron solution) of 130 L/ha. Diuron is a highly selective herbicide that specifically blocks the transfer of electrons from PSII to plastoquinone during

the thylakoid reactions of photosynthesis. There were two rows of buffer plants between adjacent herbicide treatments to prevent drift onto non-target plants. Fertility, pest control, and irrigation were managed according to University of Georgia Cooperative Extension recommendations [204]. The experiment was arranged as a split plot, randomized complete block design, where DCMU treatment represented the whole plot factor and cultivar represented the sub-plot factor. There were eight replicate plots for each cultivar within a particular herbicide treatment.

5.3.2 Data Acquisition

Hyperspectral Data Acquisition

The GPhenoVision system was configured with a spectral module for spectral data acquisition [205]. The spectral module consisted of one spectrometer (Flame VIS-NIR, Ocean Optics Inc., Largo, FL, USA) and one hyperspectral camera (MSV500, Middleton Spectral Vision, Middleton, WI, USA) (Figure 5.1). Both sensors were calibrated radiometrically and spectrally (also spatially for the hyperspectral camera), covering the spectral range from 400 nm to 1000 nm with a spectral sampling resolution of 2 nm. The spectrometer was equipped with a cosine corrector (field of view of 180°) facing towards the sky, whereas the hyperspectral camera was positioned nadir to the ground. The spectrometer acquired the irradiance spectra of the sunlight. Depending on the solar irradiance intensity, the sampling frequency of the spectrometer varied from 20 Hz to 50 Hz, so that signal intensities could stay in the optimal range without saturation. The hyperspectral camera was positioned 2.15 m above the ground, collecting radiance spectral images of plant canopies of two plots at a time. To ensure the spatial aspect ratio, the hyperspectral camera ran at 100 frames per second (FPS) and the platform moved at an approximate speed of 0.5 m/s. Six data collection trials were conducted on 9 August 2018 (beginning approximately 24 h after Diuron application) at 0800, 1000, 1200, 1400, 1600, and 1800 h, respectively.

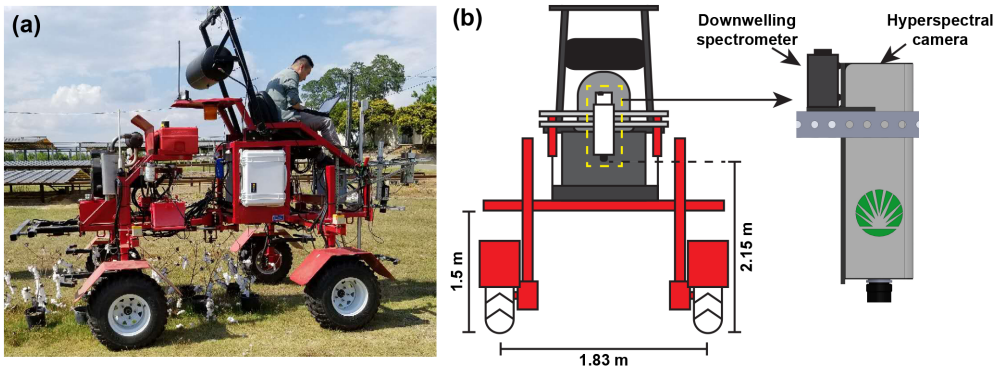

Figure 5.1: Illustration of data collection system: (a) picture of the GPhenoVision system and (b) diagram of the system configuration and sensor installation.

Fluorometry Measurement

Active chlorophyll fluorescence measurements were conducted on the uppermost fully expanded leaf at approximately the fourth mainstem node below the plant terminal using a portable pulse-amplitude-modulation (PAM) fluorometer (OS5p+, Opti-Sciences, Inc., Hudson, NH, USA). At each diurnal sampling time, the leaf blade was clipped so that the orientation of the exposed adaxial surface relative to incoming solar radiation was left unchanged, and steady state fluorescence ($F_s^{'}$) was measured under ambient light conditions. While measuring fluorescence, PAR at the leaf surface was estimated using a PAR sensor integrated into the leaf clip. Subsequently, maximal fluorescence intensity ($F_m^{'}$) was estimated using a multi-phase flash approach comparable to the methods described in [202], where relative fluorescence intensity is plotted versus the reciprocal of PAR following exposure of the leaf sample to a sequence of flashes with increasing intensity (2850, 5700, and 8550 $\mu mol/m^2/s$) for a total duration of 0.95 s. A linear function was fit to the resulting data set to estimate $F_m^{'}$ at infinite light intensity. This represents fluorescence intensity when all reaction centers are closed.

5.3.3 Characterization of Canopy-level Photosynthetic Efficiency

Retrieval of Solar Induced Fluorescence

Collected spectral data were used to retrieve SIF values at the canopy level (Figure 5.2). Irradiance spectra of the sunlight and radiance hyperspectral cubes of the plant canopy were synchronized using timestamps, resulting in meta-hyperspectral cubes of each scanning row, where individual pixels had both irradiance and radiance spectra. Based on the spatial information, a meta-hyperspectral cube was further split into two sub-cubes, with each sub-cube containing irradiance and radiance data for one plot.

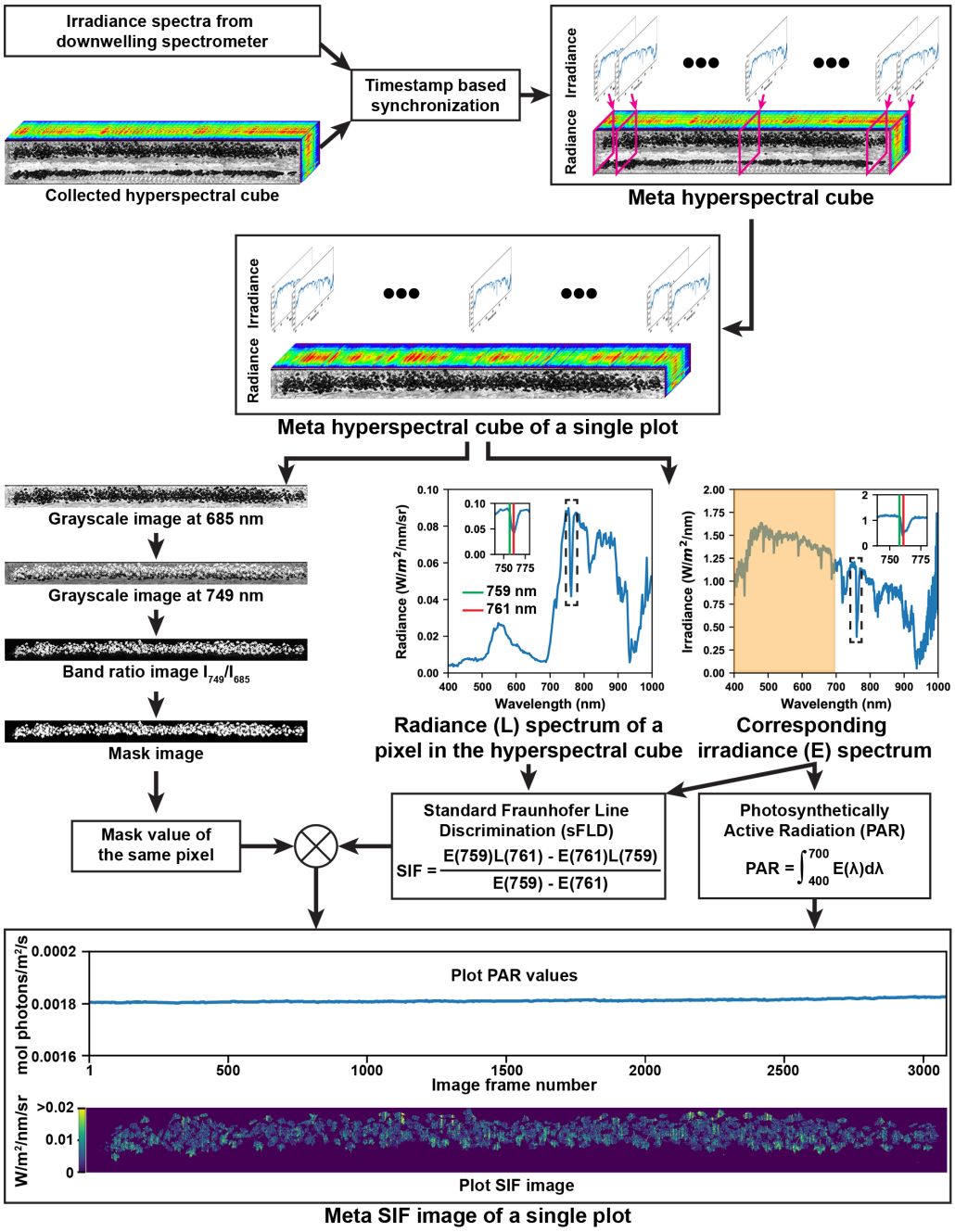

Figure 5.2: Flowchart of image processing from collected raw data to meta-SIF (solar induced fluorescence) images. In the meta-hyperspectral cube, each pixel had both irradiance and radiance spectra covering the spectral range from 400 to 1000 nm with a spectral resolution of 2 nm.

The following processes were performed for the meta hyperspectral cube of a single plot. Grayscale images at 749 nm and 685 nm were used to generate a band ratio image (I_{749}/I_{685}). A threshold was applied to the band ratio image to create the mask of plant canopies in that plot. An arbitrary value of 3 was used in the present study based on the trial-and-error method. Chlorophyll absorbs incident light (particularly blue and red light in the visible spectral range) and emits fluorescence in the red and far-red spectral range. Upwelling radiance spectra of plant canopies thus contain both reflectance and fluorescence signals in the red and far-red spectral range. Fraunhofer lines are a set of spectral absorption lines in the spectrum of sunlight related to particles in the solar and terrestrial atmosphere. $H\alpha$ at 656 nm, O_2-B at 687 nm, and O_2-A at 760 nm are three Fraunhofer lines in the red and far-red spectral range. The reduction of solar irradiance in the Fraunhofer lines results in a decrease of the canopy reflectance and an increase in the ratio of fluorescence and reflectance signals, which maximizes the suitability of decoupling chlorophyll fluorescence from the canopy reflectance. A standard Fraunhofer line discrimination (sFLD) method has been developed to use irradiance and radiance signals at two spectral bands [206]. One band is one Fraunhofer line, and the other band is a wavelength near the corresponding Fraunhofer line. In the present study, the O_2-A band (approximately at 761 nm) and its neighboring band (759 nm) were used to calculate SIF values of individual pixels using Equation 5.1.

$$SIF_p = \frac{E_p^{759} L_p^{761} - E_p^{761} L_p^{759}}{E_p^{759} - E_p^{761}} \tag{5.1}$$

Where SIF_p was the SIF value ($W/m^2/nm/sr$) of a pixel p. $E_p^{\cdot}$ and $L_p^{\cdot}$ represented the irradiance ($W/m^2/nm$) and radiance ($W/m^2/nm/sr$) intensities of p at a certain wavelength.

The plant canopy mask was multiplied with retrieved SIF values, forming the SIF image of the plot. As each image line of a radiance hyperspectral cube had a corresponding irradiance spectrum, PAR values were calculated for individual lines of SIF images using Equation 5.2.

$$PAR_p = \frac{1}{A_s t_s} \int_{400}^{700} \frac{E_p^\lambda}{e_\lambda} d\lambda = \frac{1}{A_s t_s} \int_{400}^{700} \frac{E_p^\lambda \lambda}{hc} d\lambda \tag{5.2}$$

Where PAR_p was the PAR value ($\mu mol/m^2/s$) of a pixel p. A_s was the surface area (m^2) of the cosine corrector equipped with the spectrometer, and t_s was the integration time (s) of the spectrometer. E_p^λ was the irradiance of p at the wavelength λ (nm). e_λ was the energy of a photon at the wavelength λ, h was the Planck constant, and c was the light speed in a vacuum.

Subsequently, a meta-SIF image was generated for a plot, consisting of a plot SIF image and a PAR curve (PAR values for lines along the row direction of the corresponding SIF image).

Calculation of Effective Quantum Yield and Electron Transport Rate

For data derived from hyperspectral images (HSI), the potential maximum SIF value was estimated for each of four combinations of genotype and treatment. In each genotype and treatment combination, a regression model was used to fit SIF measurements versus reciprocals of PAR, and the model y-interception (the reciprocal of PAR equaled to zero) was treated as the maximum SIF value for that combination. An exponential model was used for the two control treatment combinations, whereas three different models were used for the two Diuron combinations due to a lack of obvious data distribution patterns (likely due to damage to the photosynthetic apparatus and resulting non-photochemical quenching). The three models for the Diuron combinations included a linear regression model, quadratic regression model, and average model that used the mean value of SIF measurements as the maximum SIF. After obtaining the maximum SIF value for each approach, the effective quantum yield of PSII (ϕ_{PSII}) and photosynthetic ETR were calculated using Equation 5.3 and Equation 5.4 for data points in each combination.

$$\phi_{PSII} = \frac{SIF_m - SIF}{SIF_m} = 1 - \frac{SIF}{SIF_m} \tag{5.3}$$

$$ETR = PAR \times \phi_{PSII} \times A_{leaf} \times R_{PSII} \tag{5.4}$$

Where ϕ_{PSII} was the effective quantum yield of PSII for a SIF value (SIF). SIF_m was the estimated maximum SIF value for a genotype and treatment combination. A_{leaf} was the leaf absorbance of incident light, and a typical value (0.84) for C3 species (e.g., cotton) was used in this study. R_{PSII} was the distribution ratio of absorbed energy between photosystem I (PSI) and photosystem II (PSII), which was assumed as equal distribution (0.5) [207].

For PAM fluorometry data, the maximal fluorescence intensity and PAR were obtained, so ϕ_{PSII} was calculated using Equation 5.5 [208] while ETR was still calculated using Equation 5.4.

$$\phi_{PSII} = \frac{F_m^{'} - F_s^{'}}{F_m^{'}} = 1 - \frac{F_s^{'}}{F_m^{'}} \tag{5.5}$$

Where ϕ_{PSII} was the effective quantum yield of PSII. $F_s^{'}$ and $F_m^{'}$ were the steady state and maximal fluorescence intensities under actinic light.

Rapid Light Curve and Standardized ETR

ETR and PAR values were used to generate rapid light curves (RLCs) using the method proposed by [209]. To quantitatively analyze RLCs, RLCs were fitted using an empirical model (Equation 5.6) [210].

$$ETR = mETR \times (1 - e^{\alpha \times PAR/mETR}) \tag{5.6}$$

Where $mETR$ was the photosynthetic capacity at saturating light by which all reaction centers were hypothetically closed and α was the initial slope of the RLCs before the onset

of saturation, i.e., the slope of the light-limiting region of RLCs.

RLCs were generated and modeled at two levels: the group level and the plot level. RLCs at the group level were calculated using all data points (N=48) of individual group (a genotype and treatment combination), and RLC parameters were compared to evaluate differences in canopy-level photosynthetic efficiency among the four groups. In order to testify the statistical significance of the differences, RLCs at the plot level were calculated using data points (N=6) of individual plots. Consequently, individual group would have 8 replicates for conducting statistical analyses.

While $mETR$ can be used to compare differences of photosynthetic efficiency between groups, standardized ETR (sETR) values were calculated by setting PAR equal to 1500 $\mu mol/m^2/s$, which is a common value in the study area and is generally considered a saturating light intensity for cotton [211].

5.3.4 Growth Analysis

Crop performance was also verified by destructively harvesting all above-ground plant material in a 2 meter length of row on two sample dates to derive classical crop growth indices. Plants were sampled on 9 August 2018 (immediately after SIF measurements) and 23 August 2019 (a two week interval). On each sample date, plants were placed in plastic bags with moist paper towels to ensure that plant tissues did not desiccate between harvest and measurement. In the laboratory, plants were separated into leaves and stems, and leaf area was determined using a leaf area meter (LI-3100, LI-COR Corp., Lincoln, Nebraska USA). Total dry weight was assessed following a 48 h drying period at 80°C in a forced-air oven.

Five growth parameters were calculated including crop growth rate (CGR), net assimilation rate (NAR), relative growth rate (RGR), the difference in leaf area index (ΔLAI), and the difference in leaf mass fraction (ΔLMF) between the two sample dates. The five parameters were defined by Equation 5.7 to Equation 5.11.

$$CGR = \frac{W_{t2}^{total} - W_{t1}^{total}}{(t2 - t1) \times A^{land}} \tag{5.7}$$

$$NAR = \frac{W_{t2}^{total} - W_{t1}^{total}}{t2 - t1} \times \frac{\ln A_{t2}^{leaf} - \ln A_{t1}^{leaf}}{A_{t2}^{leaf} - A_{t1}^{leaf}} \tag{5.8}$$

$$RGR = \frac{\ln W_{t2}^{total} - \ln W_{t1}^{total}}{t2 - t1} \tag{5.9}$$

$$\Delta LAI = \frac{A_{t2}^{leaf} - A_{t1}^{leaf}}{A^{land}} \tag{5.10}$$

$$\Delta LMF = \frac{W_{t2}^{leaf}}{W_{t2}^{total}} - \frac{W_{t1}^{leaf}}{W_{t1}^{total}} \tag{5.11}$$

Where W^{leaf} and W^{total} represented the leaf and total dry weights. A^{leaf} and A^{land} represented the leaf and land areas. In the present study, the A^{land} was 1.82 m^2 (0.91 m × 2 m). $t2$ and $t1$ were the sampling date in days after planting (DAPs).

5.3.5 Statistical Analysis

To test the effectiveness of standardized ETR values, ANOVA tests were performed on the five growth parameters and standardized ETR values estimated using four approaches (PAM and three HSI-based methods). After testing the effects due to genotype, treatment, and the interaction between genotype and treatment, ANOVA tests were further performed on the traits between treatments for each genotype. ANOVA tests were performed at the significance level of 0.05 in R [212]. In addition to ANOVA tests, Pearson correlation analysis

was conducted between each pair of traits and sETR, evaluating the potential of using the sETR for growth prediction. Pearson correlation analyses were also conducted in R.

5.4 Results

5.4.1 Representative Meta-SIF Images

Meta-SIF images showed obviously different trends between the control and Diuron groups (Figure 5.3). In the control groups, SIF values had the same trend with PAR values: SIF values increased with the increase of PAR values and decreased with the reduction of PAR values. In contrast, SIF values for Diuron-treated plots exhibited no relation with PAR values: SIF values were low and relatively constant irrespective of PAR changes throughout the day. This observation agreed with the experimental design. The control groups were healthy, showing fluorescence intensity changes along with varied PAR levels, whereas an inhibitor of electron transport beyond PSII would be expected to cause damage to the photosynthetic apparatus and potentially increase non-photochemical quenching (NPQ) of the fluorescence signal.

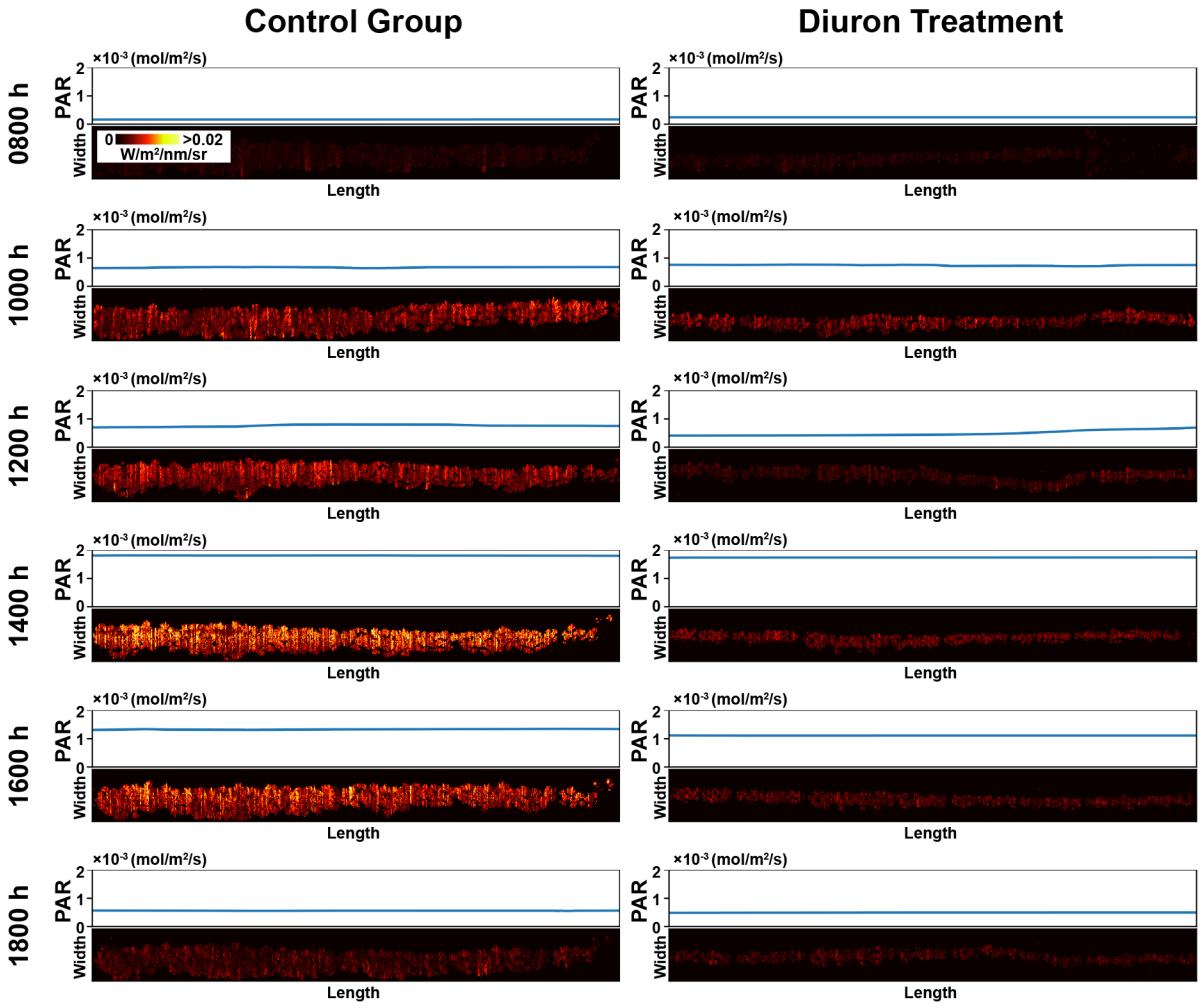

Figure 5.3: Representative meta-SIF images for control and Diuron plots. In meta-SIF images, each pixel had both SIF value and corresponding photosynthetically active radiation (PAR) values.

In addition, variations were observed within a plot (see the control group at 1400 h in Figure 5.3). Leafy regions in the plot showed higher SIF signals than less leafy regions, because the leafy regions usually had a faster vegetative growth (and thus more mature leaves), resulting in a higher capacity of conducting photosynthesis. This suggests that calculated meta-SIF images could be used for analyzing the spatial variation of photosynthesis. As the present study aimed to explore the possibility of using SIF measured by passive HSI method to characterize photosynthetic efficiencies at the canopy level, experiments were not

designed to analyze spatial variations of SIF in a plot. Thus, successive studies are needed to take the full advantage of HSI data to analyze spatial variations of SIF (and potentially other photosynthetic parameters).

5.4.2 Estimated Maximal Fluorescence

Generally, the control groups were well ($R^2 = 0.85$) fitted by the exponential model, whereas the Diuron groups showed large variations in the model that best fit the SIF response to PAR (Figure 5.4. As exponential models showed a strong relationship between SIF and PAR in control groups, it was reasonable to use the y-interception of the models as the maximal fluorescence intensity. In contrast, the three models for the Diuron groups provided different goodness-of-fit and maximal fluorescence values. Regarding goodness-of-fit, the quadratic model provided the best estimation of the maximal fluorescence, followed by the linear model and average model. This was possibly because the level of NPQ in Diuron treated plots would have been higher under the highest light intensities where the most damage would have been expected. Although a previous study showed a linear relationship between chlorophyll fluorescence and PAR under high light intensities (over 2800 $\mu mol/m^2/s$) [202], no strong linear relation was observed between SIF and PAR under solar illumination (up to 2000 $\mu mol/m^2/s$) in the present study. In particular, the linear model provided a reduced goodness-of-fit than the quadratic model, suggesting that the linear model might not be optimal for estimating the maximal fluorescence for the Diuron groups. However, there was a reasonable consideration for using the average model despite the worst goodness-of-fit. No obvious trend was identified between SIF and PAR values in the Diuron groups, and thus variations among data points could be considered as random measurement errors. If so, it was acceptable to use the mean value as the group measurement (and thus the maximal fluorescence of that group), thereby reducing the measurement error. Nonetheless, the three models provided different values for maximal fluorescence intensity and would have different

effects on the successive data processing.

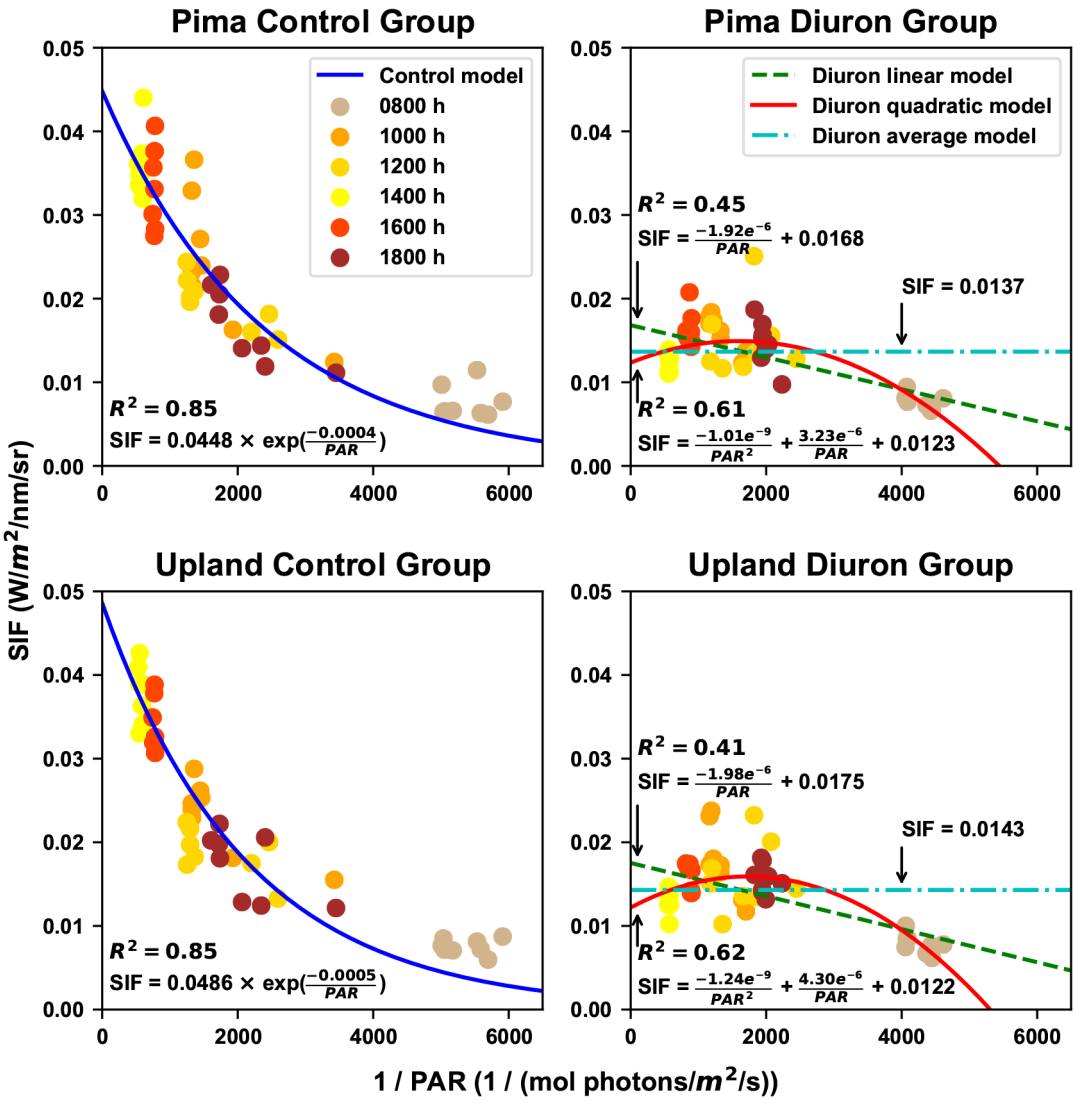

Figure 5.4: Estimation of maximal fluorescence values for control and Diuron treatment. An exponential model was used to fit the SIF and PAR reciprocal values for plots in the control group, whereas three models (linear, quadratic, and average) were used to fit the SIF and PAR reciprocal values for plots in the Diuron treatment group. The maximal fluorescence value was defined as the value when the PAR reciprocal equaled zero.

5.4.3 Calculated Effective Quantum Yield and RLCs

Overall, the Pima and Upland cultivars showed the similar trends and patterns for ϕ_{PSII} and ETR values calculated using PAM and hyperspectral data (Figure 5.5). For all four methods, ϕ_{PSII} values decreased with increasing PAR values in the range from 0 to 2000 $\mu mol/m^2/s$ for control groups, whereas ϕ_{PSII} values were relatively lower and showed no correlation with PAR values for Diuron groups. Control and Diuron groups were distinctively different from each other in ϕ_{PSII} over PAR calculated using PAM data. While control and Diuron groups were still separable, overlaps between the two groups were identified with different magnitudes in ϕ_{PSII} over PAR calculated using HSI data. The overlaps between control and Diuron groups for quantum yield were the largest when ϕ_{PSII} values were calculated using maximal fluorescence estimated by the linear model, and the overlaps became smaller and were the least when ϕ_{PSII} values were calculated using maximal fluorescence estimated by the average and quadratic models, respectively. This occurred primarily because of the difference in estimated maximal fluorescence values for the Diuron groups. The HSI quadratic model provided the lowest value of the estimated maximal fluorescence, resulting in the lowest ϕ_{PSII} values. In particular, a lower maximal fluorescence value lead to more negative ϕ_{PSII} values that were treated as zero, which increased the magnitude of the difference between the control and the Diuron groups.

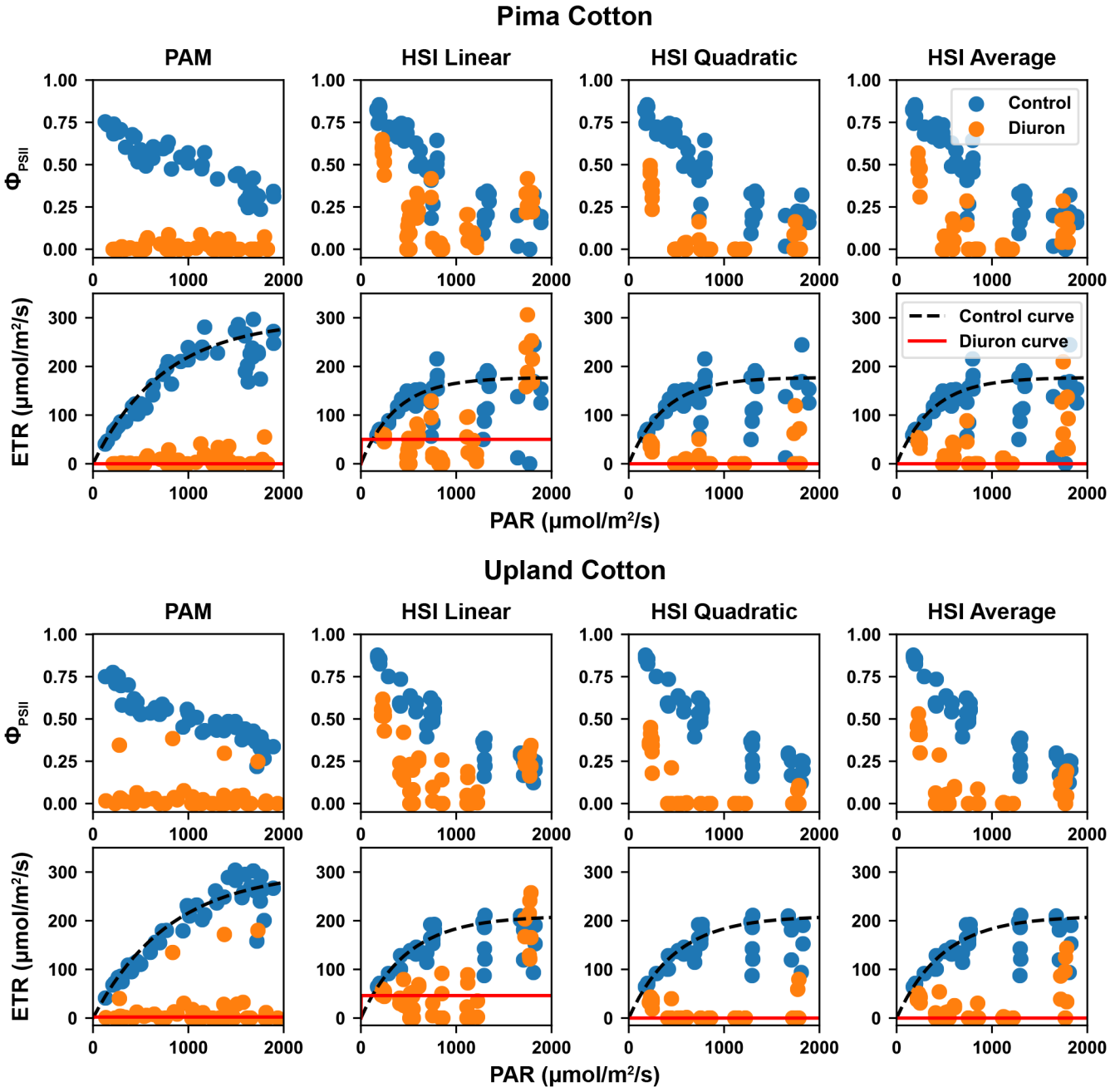

Figure 5.5: Calculated effective quantum yield of PSII (ϕ_{PSII}) and electron transport rate (ETR) for the two genotypes under control and Diuron treatments. PAR shorts for photosynthetically active radiation.

ETR values rapidly increased with increasing PAR values in the range from 0 to 600 $\mu mol/m^2/s$ and gradually reached a plateau afterwards for control groups. This trend

held true for ETR values calculated using PAM- and HSI-derived ϕ_{PSII}. ETR values calculated using PAM-derived ϕ_{PSII} remained low (near zero) in the PAR range from 0 to 2000 $\mu mol/m^2/s$ for Diuron groups, showing a clear separation from control groups. In contrast, ETR values calculated using HSI-derived ϕ_{PSII} showed different magnitudes of overlap between control and Diuron groups. In the PAR range from 0 to 1500 $\mu mol/m^2/s$, all three HSI-based methods showed a distinction between control and Diuron groups, with the largest overlap by the HSI linear method followed by average and quadratic methods. In the PAR range from 1500 to 2000 $\mu mol/m^2/s$, the HSI linear method showed no distinction between two treatment groups, whereas the HSI quadratic and average methods still showed a distinct difference. This matched with the observations of ϕ_{PSII}.

Quantitative RLC models further validated the aforementioned observations (Table 5.1). For all four methods, control groups had a substantially higher value of mETR and a lower value of initial slope (α) than Diuron groups for both cultivars. Although the HSI linear method provided ETR values that had a large overlap between control and Diuron groups, the quantitative characteristics of the fitted RLC models were dramatically different from each other, showing that even in the worst case (HSI linear method), the developed method could be used to identify the photosynthetic efficiency difference between the two groups. Although different methods showed a similar trend on the RLC characteristics between different treatments, the absolute values of those characteristics were different. mETR calculated using HSI methods was 60% to 70% lower than those calculated using the PAM method, and α were 15% to 20% higher. The reduction of absolute values for mETR was partially because the PAM and HSI methods provided measurements at different levels. The PAM method measured a single leaf near the top of a plant with little or no shading, whereas the HSI methods measured a whole canopy that had leaves at different levels of shading. So, it would be expected that the HSI methods provided lower ETR than the PAM method. In addition to herbicide treatment differences, the two cultivars showed certain

differences. The Pima cultivar had a lower minimum saturating irradiance (estimated by $\frac{mETR}{\alpha}$ [210]) than the Upland cultivar. This suggests that the Pima cultivar would enter into the stage dominated by non-photochemical quenching at lower light intensities than the Upland cultivar [213]. This finding is in agreement with a previous study showing that individual leaves of Pima cotton reached light saturation for net photosynthesis at a lower light intensity than Upland cotton [214].

Table 5.1: RLCs calculated using ETR and PAR values for each genotype and treatment combination. The RLC model was $ETR = mETR \times (1 - e^{\alpha \times PAR/mETR})$.

Method	Cultivar	Treatment	mETR	α	E_k	R^2
PAM	Pima	Control	296.70	0.3989	743.73	0.94
PAM	Pima	Diuron	0.00	0.5000	0.00	
PAM	Upland	Control	303.90	0.3771	805.98	0.99
PAM	Upland	Diuron	2.12	0.5000	4.24	
HSI_Linear	Pima	Control	177.51	0.4780	371.33	0.92
HSI_Linear	Pima	Diuron	50.13	77.0580	0.65	
HSI_Linear	Upland	Control	210.07	0.4348	483.13	0.99
HSI_Linear	Upland	Diuron	46.19	93.9418	0.49	
HSI_Quadratic	Pima	Control	177.51	0.4780	371.33	0.92
HSI_Quadratic	Pima	Diuron	0.00	0.5000	0.00	
HSI_Quadratic	Upland	Control	210.07	0.4348	483.13	0.99
HSI_Quadratic	Upland	Diuron	0.00	0.5000	0.00	
HSI_Avg	Pima	Control	177.51	0.4780	371.33	0.92
HSI_Avg	Pima	Diuron	0.00	95.0250	0.00	
HSI_Avg	Upland	Control	210.07	0.4348	483.13	0.99
HSI_Avg	Upland	Diuron	0.00	0.5000	0.00	

ϕ_{PSII} measured using different HSI methods also showed different correlations with those measured using the PAM method (Figure 5.6). In fact, a high correlation ($R^2 = 0.73$) was achieved between PAM and HSI methods for control groups. Differences in correlations mainly came from the measurements for Diuron groups. The quadratic model provided the best estimate of maximal fluorescence, exhibiting the strongest correlation between PAM and HSI-based estimates of quantum efficiency at the canopy level. This suggests that the quadratic model could be an optimal method for estimating maximal fluorescence for Diuron groups. The average model showed a reduced but comparable performance with the quadratic model. Considering its engineering rationale, the average model could also be an option for estimating maximal fluorescence for Diuron groups. Additional experiments are needed to determine which method (the quadratic or average model) would be the most optimal in terms of model generalization capability to different datasets. However, the linear model provided the lowest correlation, suggesting that the linear model would not be suitable for maximal fluorescence estimation for Diuron groups.

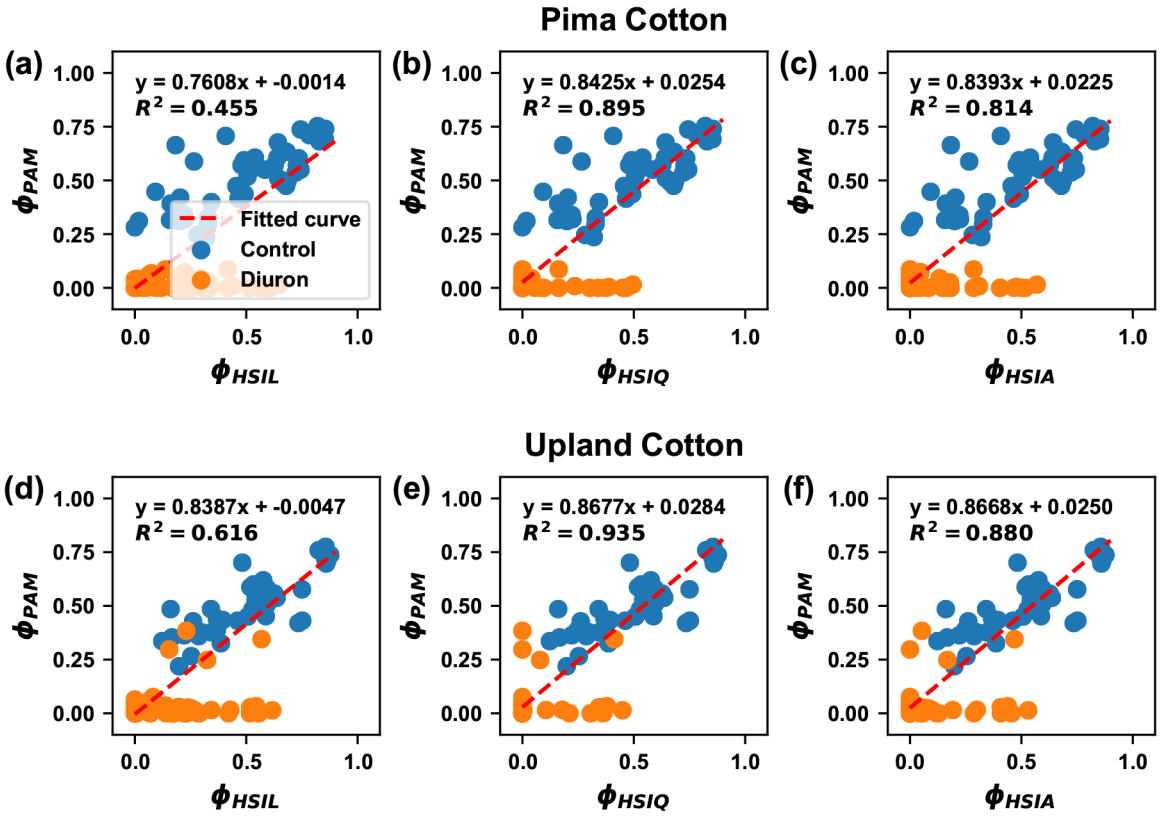

Figure 5.6: Results of correlation analysis between the effective quantum yield of PSII (ϕ_{PSII}) values derived from PAM and HSI data. ϕ_{PAM} indicated values calculated using PAM data. ϕ_{HSIL}, ϕ_{HSIQ}, and ϕ_{HSIA} indicated values derived from hyperspectral data using the maximal fluorescence value estimated by the linear, quadratic, and average models.

5.4.4 ANOVA Test Results

Generally, growth traits and sETR calculated using four methods showed significant differences between treatments but no difference between genotypes (Table 5.2). This suggests that the sETR measurements were effective in identifying the differences in plant growth between control and Diuron groups. It should be noted that although an interaction effect was identified for ΔLMF and sETR calculated using the HSI methods, control groups still showed higher values than Diuron groups irrespective of genotype, indicating the validity of main effect (treatment) significance for those traits. RGR was the only trait showing a

significant difference between genotypes.

Table 5.2: P-values of ANOVA tests on the growth traits and standardized ETR values estimated using four approaches.

Trait	Cultivar	Treatment	Interaction between cultivar and treatment
CGR	0.1568	0.0001	0.8406
NAR	0.1229	0.0002	0.3513
RGR	0.0128	0	0.2899
ΔLAI	0.1801	0.0018	0.3906
ΔLMF	0.9069	0.0008	0.006
$sETR_{PAM}$	0.0792	0	0.65
$sETR_{HSIL}$	0.2661	0	0.0237
$sETR_{HSIQ}$	0.0864	0	0.0411
$sETR_{HSIA}$	0.1663	0	0.0241

For each cultivar, growth traits and sETR calculated using four methods had statistically higher values for control groups than Diuron groups (Figure 5.7). Growth traits (e.g., CGR, NAR, RGR, and ΔLMF) were positive for control groups, indicating normal plant growth, whereas those traits were close to zero or even negative, indicating plant loss of mass and leaf area in Diuron treated plots. The same patterns were observed for sETR calculated using four methods. An exception was ΔLMF, which represents dry matter partitioning to leaf area. Both cultivars showed negative ΔLMF for control and Diuron groups, meaning both cultivars distributed a smaller fraction of total dry matter to leaves at the later sample times irrespective of treatment. The Pima cultivar exhibited a greater decline in LMF for the Diuron group than the control group, likely reflecting the Diuron damage and resulting defoliation that led to leaf mass fraction reduction. In contrast, the Upland cultivar showed no difference of ΔLMF between treatments. This is likely because the total growth was

negatively affected to a comparable extent as leaf mass.

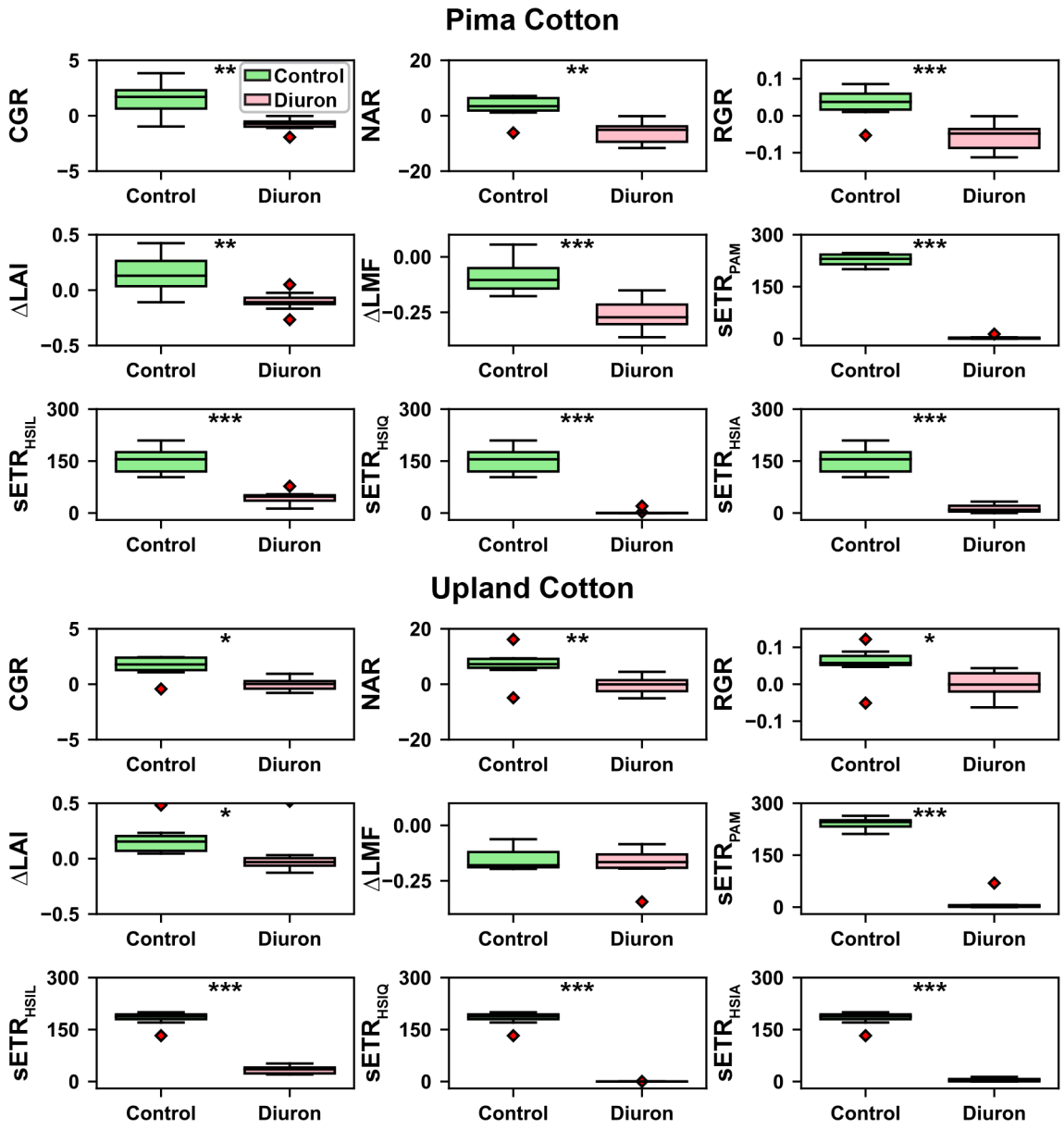

Figure 5.7: ANOVA tests on the growth traits and standardized ETR values estimated using four approaches. Asterisks indicated statistical differences between two treatments at different significance levels: * <0.05, ** <0.01, and *** <0.001. CGR, NAR, RGR, and sETR values were in the units of $g/m^2/d$, $g/m^2/d$, $g/g/d$, and $\mu mol/m^2/s$, respectively.

5.4.5 Correlation between Traits

High correlations (r = 0.92 to 0.95) were achieved for sETR calculated using the PAM method and HSI methods, further indicating the validity of using the developed method for canopy-level photosynthetic efficiency quantification (Figure 5.8). sETR, irrespective of calculation method, also showed moderate correlations (r = 0.46 to 0.67) with growth traits, indicating the potential of using calculated sETR for crop growth prediction. Among the three HSI methods, the HSI linear method provided the lowest correlation, whereas the HSI quadratic and average methods showed relatively higher correlations. This demonstrates that the estimation of maximal fluorescence affects the capability of using sETR for growth prediction as well.

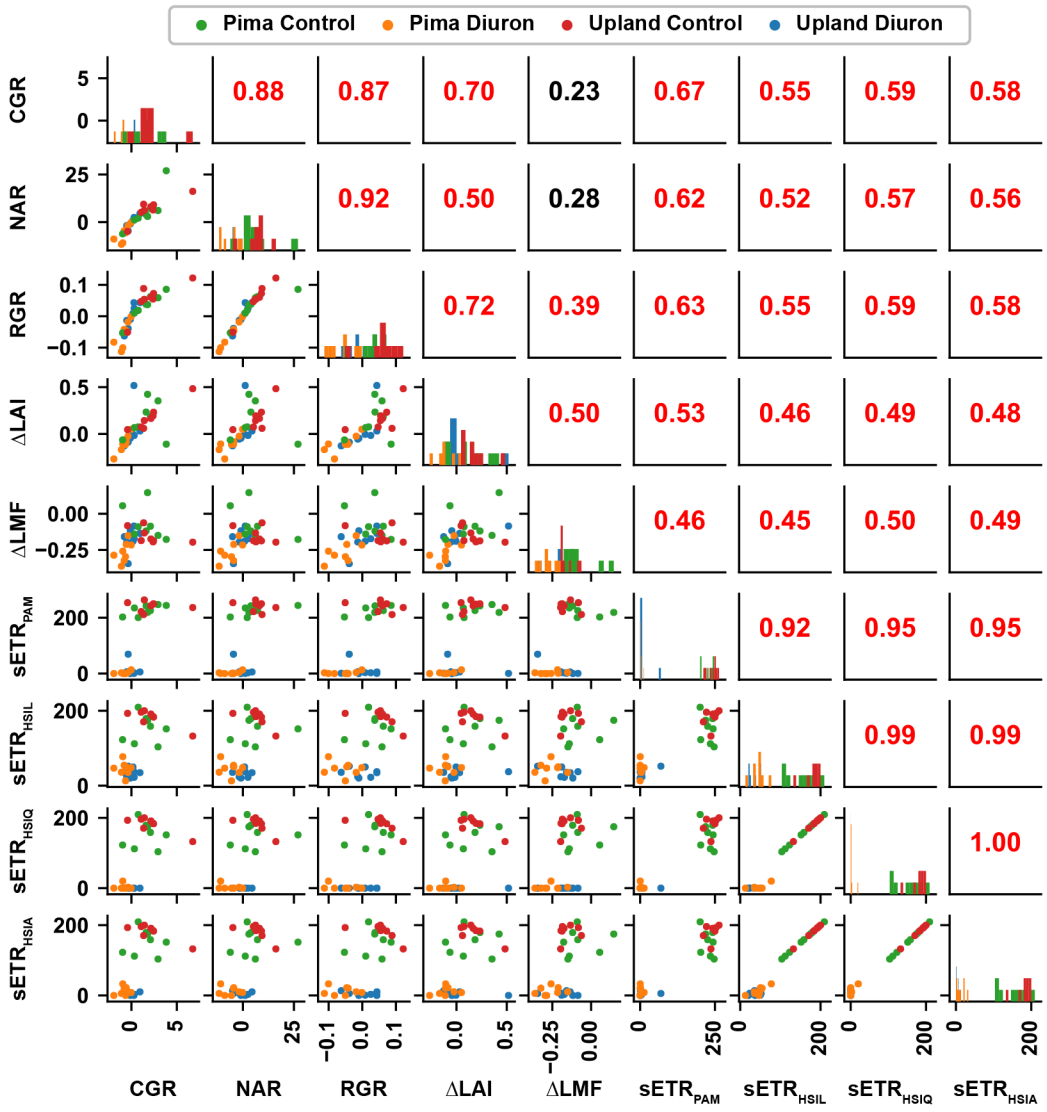

Figure 5.8: Pearson correlation values between each pair of five growth traits and standardized ETR values estimated using four approaches. Values on the upper-triangle were the correlation values for the subplots in the lower triangle. Significant (p-value <0.05) correlation values were rendered by red color or otherwise by black color.

5.5 Discussion

The developed approach used a passive hyperspectral imaging system to retrieve diurnal SIF values that were used for estimating maximal fluorescence, effective quantum yield of PSII and ETR values at the canopy level. Irrespective of the approach utilized to model

SIF to diurnal light intensity, the following conclusions can be made. Maximal fluorescence estimates for the canopy, derived from the best fit function of SIF versus the reciprocal of PAR in both the Diuron-treated plots and the control plots, were used to calculate actual quantum yield of the canopy. ϕ_{HSI} estimates were strongly correlated with ϕ_{PAM} estimates when considered across multiple diurnal sample times, in response to Diuron treatment, or for the two different species of cotton evaluated (Figure 5.6). Furthermore, ϕ_{HSI} values were used to calculate ETR at a given diurnal PAR level, and diurnal light response curves were generated for both PAM and HSI-based methods. From these data, a standardized measure of ETR at a common light intensity (1500 $\mu mol/m^2/s$) was calculated for both PAM and HSI-based methods. There was a strong correlation between HSI- and PAM-derived sETR values (r = 0.95; Figure 5.8), and canopy estimates of ETR were also predictive of whole-crop growth responses in the weeks following treatment. Thus, we suggest that is possible to estimate canopy-level photosynthetic efficiency from passive, diurnal measurements of SIF and PAR at the canopy at the time of measurement. This is a particularly notable achievement since a number previous studies have collected diurnal SIF measurements at the agro-ecosystem scale in attempts to model gross primary productivity [179, 215, 216]. The methodology reported herein would allow researchers to obtain a direct measure of ecosystem-scale photosynthetic efficiencies from existing data sets. Furthermore, the use of hyperspectral imaging to delineate plot-scale photosynthetic efficiencies, would allow for direct selection of genotypes with higher diurnal photosynthetic activities. Last, the developed method measures chlorophyll fluorescence in field conditions, which could expand fluorescence-based early detection of plant stresses (described by [217], [218], and [219]) from the laboratory to the field. This could be particularly useful for agricultural production systems.

The developed approach has two major limiting factors, however. First, ETR values measured by HSI data showed a larger absolute difference than those measured by PAM data.

This can be solved by improving the spectral sampling resolution [220]. The hyperspectral camera and spectrometer used the spectral sampling resolution of 2 nm in the present study, utilizing only around 50% depth of the Fraunhofer O_2-A line [220]. This limits the SIF retrieval accuracy and thus derivation of effective quantum yield of PSII and ETR values. On the contrary, the spectral sampling resolution of the hyperspectral camera and downwelling spectrometer can be configured to approximately 0.2 nm, which could provide greater potential for further improvement of measurement accuracy. The finest spectral resolution will result in a large increase of data volume, presenting potential challenges in data collection, management, and processing. Second, estimation of maximal fluorescence dramatically affected the calculation of effective quantum yield and ETR as well as RLC models. In the present study, a linear model provided the poorest results in which control and Diuron groups showed a large overlap. Although fitted RLC models still showed a significant difference between treatments, it could be problematic to identify subtle differences of photosynthetic efficiency among genotypes. This could limit the potential of using the developed method for genetics/genomics studies and breeding programs. Future studies, therefore, need to validate the efficacy of three models used in the present study and examine new ways to estimate maximal fluorescence using passive sensing methods. For instance, maximal fluorescence estimation models for stressed plants can be established and validated in a full PAR range from 0 to 6000 $\mu mol/m^2/s$ in a controlled environment, and transferred to field applications where data are acquired in a part of the full PAR range.

5.6 Conclusions

The developed method showed promising results in using passive hyperspectral data to estimate effective quantum yield, ETR, and RLC models for the whole canopy. Patterns observed using calculated RLC characteristics agreed closely with growth traits, indicating that the developed method can be used to differentiate plants under extreme differences in photosynthetic efficiency. Regression analysis results confirmed that the calculated values

had potential for plant growth prediction. Future studies will focus on exploring various estimation methods for maximal fluorescence of the canopy and potential applications in breeding programs and genetics/genomics studies.

Chapter 6

DeepSeedling: Deep Convolutional Network and Kalman Filter for Plant Seedling Detection and Counting in the Field[1]

[1]Jiang, Y., Li, C., and Paterson, A. H. Submitted to *Plant Methods*, April 8, 2019.

6.1 Abstract

Plant population density is an important factor for agricultural production systems due to its substantial influence on crop yield and quality. Traditionally, plant population density is estimated by using either field assessment or a germination-test-based approach. These approaches can be laborious and inaccurate. Recent advances in deep learning provide new tools to solve challenging computer vision tasks such as object detection, which can be used for detecting and counting plant seedlings in the field. The goal of this study was to develop a deep-learning-based approach to count plant seedlings in the field. Overall, the final detection model achieved F1 scores of 0.727 (at IOU_{all}) and 0.969 (at $IOU_{0.5}$) on the $Seedling_{All}$ testing set in which images had large variations, indicating the efficacy of the Faster RCNN model with the Inception ResNet v2 feature extractor for seedling detection. Ablation experiments showed that training data complexity substantially affected model generalizability, transfer learning efficiency, and detection performance improvements due to increased training sample size. Generally, the seedling counts by the developed method were highly correlated ($R^2 = 0.98$) with that found through human field assessment for 75 test videos collected in multiple locations during multiple years, indicating the accuracy of the developed approach. Further experiments showed that the counting accuracy was largely affected by the detection accuracy: the developed approach provided good counting performance for unknown datasets as long as detection models were well generalized to those datasets. The developed deep-learning-based approach can accurately count plant seedlings in the field. Seedling detection models trained in this study and the annotated images can be used by the research community and the cotton industry to further the development of solutions for seedling detection and counting.

6.2 Introduction

Plant population density is defined as the number of plant stands per unit area, which is an important factor for agricultural production systems due to its substantial influence on crop yield potential and fruit quality [221, 222, 223, 224, 225]. Plant population density is particularly important for growers right after the germination stage, providing hard date from which to evaluate the necessity of replanting the field if the density if not adequate. Thus, it is crucial to calculate the plant population density when plants are in the seedling stage.

There are two conventional ways for estimation of plant population density: field assessment and a germination-test-based approach. To estimate plant population density, a field assessment involves manually counting the number of plant seedlings/stands in randomly selected subareas of a field and using the average value to represent the plant population density. While this approach is the most straightforward, it is laborious (especially for large fields) and could be inaccurate due to improper sampling of subareas. A germination-test-based approach involves estimating the plant population density by using the total number of planted seeds and the maximum germination potential determined by laboratory germination tests. This approach is relatively simple (indeed, such data usually accompanies purchased seed), but it could provide estimation that is considerably different from the actual value due to differences of testing and field conditions.

To address these issues, several studies have been conducted to investigate the use of color images to count plant seedlings in the field [226, 227, 228]. These studies relied on conventional image processing, which primarily utilized color information to segment vegetative areas that were used for estimation of plant counts. While these approaches showed high counting accuracies (approximately 90%), they had two major disadvantages. First, color information was sensitive to ambient illumination and plant status. For instance, plants

looked darker on cloudy days than sunny days, and plants just sprouting from the soil might have different color than well-established seedlings. Second, counting models were site- and time-dependent. Typically, a calibration step was necessary: a small portion of a field would be manually counted to establish a regression model between pixel counts (or the number of segmented areas) and actual plant counts, so the regression model could be applied to the rest of the images for automatic processing. A regression model established in one experimental site (growth stage) might not transfer to another site (growth stage), requiring model validation or re-calibration. In particular, breeding programs and genetics studies involve a wide variety of genotypes with considerable variations in germination time, raising a particular concern about using these image-processing-based approaches for plant counting.

Seedling detection is an essential part of seedling counting. Recent breakthroughs in deep learning (e.g., deep convolutional neural networks, also known as, CNNs) have demonstrated strong performance for object detection [30]. In particular, faster-region-based CNN (Faster-RCNN) was developed as a CNN-based meta-architecture for object detection [45], which has been shown to provide state-of-the-art performance for various applications and competitions [30]. Researchers explored the use of Faster RCNN for *in vivo* fruit detection for peppers [69, 81], mangoes [80], apples [86, 229], almonds [72], and maize ears [71]. These studies reported promising detection accuracy (F1 score from 0.8 to 0.92) and thus per-image counting accuracy (relative counting errors from 2% to 15%). In addition, several of these studies further expanded the Faster-RCNN-based fruit detection and counting from a single image to consecutive image sequences.

The key challenge of counting in image sequences or videos is to preclude repeated counting of one fruit object. Three approaches were used to solve this challenge. The first approach reconstructed 3D point clouds of a crop row using 2D images by the structure from motion (SfM) technique, and fruit detections were projected from individual 2D images

to the reconstructed global 3D space [86, 229]. As a single fruit object occupied a unique 3D position, detections of one fruit object in different 2D images would highly overlap in the 3D space. Thus, redundant detections of one fruit object could be removed to avoid repeated counting. The second approach used the position (from RTK GPS) and pose (from IMU devices) of image acquisition to estimate the geometric correspondence between pixels in two consecutive images. With this approach, fruit detections in one image could be associated with detections in the next image, thus tracking individual detections through image sequences or videos for counting [80]. In the third approach, a tracking-via-detection strategy was developed to track and count fruit objects in image sequences [81, 229]. The key of the tracking-via-detection strategy is detection-tracker association (assign a detection to a tracker). In [81], the intersection of union (IOU) and boundary measure (the ratio of the intersection between a tracker and a detection to the area of that detection) metrics were used to quantify the closeness between a detection and a tracker. Thresholds of IOU and boundary measure were determined using a small set of image sequences. For a given pair of detection and tracker, if they had an IOU value and a boundary measure that exceeded the predetermined thresholds, the detection and tracker would be associated. In [229], optical flow was calculated to estimate object motion (center positions for trackers) between consecutive images. The estimated center positions for trackers would be compared with center positions for detections. If the center-to-center distance was the minimum, a detection and a tracker would be associated. Although all the three approaches provided fairly high counting accuracies (95.56% to 97.83% for [86], 98% for [80], and 95.9% for [81]), they had various limitations. The first approach was computationally expensive due to the SfM-based reconstruction. In addition, certain environmental factors (e.g., wind) would result in failure of 3D reconstruction using the SfM. The second approach was less computationally expensive than the first one, but the use of positioning sensors (e.g., RTK GPS and IMU) substantially increased the cost of data acquisition systems, which could be problematic for small research projects/farms. The third approach was the least expensive

in terms of computation and hardware investment, but the tracking strategy was not robust to different noises. For the method used in [81], the IOU and boundary measure thresholds were determined using only a small set of image sequences. If the testing image sequences and videos were acquired in slightly different conditions, the thresholds might become invalid and result in degraded performance. For the optical flow method [229], the calculation of optical flow could be dramatically influenced by ambient illumination changes, resulting in inaccurate motion estimation and tracking [83]. These issues could be addressed by using other tracking methods such as Kalman filter.

To the best of our knowledge, no study has reported the use of a CNN-based approach for seedling detection and counting. Based on the successes of fruit detection and counting, it is worth exploring the use of CNN-based approaches for seedling detection and counting. In particular, the combination of CNN-based detection models and sophisticated tracking framework would provide inexpensive but accurate counting solutions.

The overall goal of this study was to develop an approach based on CNN and Kalman filter to counting cotton seedlings in the field. Specific objectives were to 1) collect and annotate image datasets for detection model training and testing, 2) train Faster-RCNN models for seedling detection, 3) examine the key factors (training sample size, transfer learning efficiency, and generalizability) for detection model training, and 4) use the trained Faster-RCNN models and Kalman filter to track and count cotton seedlings in videos of individual plots or field segments.

6.3 Methods

6.3.1 Data Collection and Preparation

Videos were collected using various imaging systems in the cotton germination stage at different locations over multiple years (Table 6.1). The collected videos were split into detection and counting sets. Videos in the detection set were used to extract video frames at a frequency of 6 frames per second (FPS), forming three image datasets that were used for plant seedling detection. For convenience, the three image datasets will be referred to as TAMU2015, UGA2015, and UGA2018 hereafter. Videos in the counting set were segregated into 75 video clips (25 clips per dataset) for evaluating the developed counting algorithm. Each video clip represented an approximately 3-m long segment in the videos collected in the state of Texas or a single plot in the videos collected in the state of Georgia.

Table 6.1: Data collection summary.

Data collection	Texas A&M University	University of Georgia	University of Georgia
Location	Corpus Christi, TX, USA	Watkinsville, GA, USA	Watkinsville, GA, USA
Plot length	10.67 m	3.05 m	1.5 m
Seed spacing	0.08 m/seed (average)	0.1 m/seed (average)	0.15 m/seed
Date	12 April 2015 11 DAPs	15 June 2015 11 DAPs	13 June 2018 7 DAPs
Weather	Sunny	Cloudy	Sunny
Camera	Samsung Galaxy Note3	Panasonic DMC-G6	Fujifilm X-A10
Video configuration	1920 × 1080 @ 30 FPS	1920 × 1080 @ 60 FPS	1920 × 1080 @ 30 FPS
ISO/HDR	Auto / Auto	160 / Off	Auto / Not support
Average moving speed	0.6 m/s	0.75 m/s	0.6 m/s
Number of collected videos (detection/counting)	3 (2/1)	6 (4/2)	4 (2/2)
Number of plots per video	7	16	19

Three image datasets were preprocessed to reduce the image variability (Figure 6.1). A contrast limited adaptive histogram equalization (CLAHE) algorithm was applied to equalize the value channel of images in the HSV color space, which enhanced the image contrast and reduced the image variation due to ambient illumination changes. The preprocessed images were manually annotated with bounding boxes for objects of two classes: plant seedlings and weeds. Monocotyledon weed was the only weed type observed and labeled in the TAMU2015 dataset, whereas dicotyledons were the primary weed type in the UGA2015 dataset. Very small-sized weed objects (less than 30×30 pixels) were not labeled. As pre-emergent herbicides were applied, there was no weed identified in the UGA2018 dataset. After manual annotation, the three datasets were partitioned into training, validation, and testing sets with a ratio of 80%/10%/10% (Table 6.2). Subsequently, four comprehensive datasets were generated by combining the three datasets (Table 6.3). As the T15U15, T15U18, and U15U18 datasets were only used for model training and validation, the validation and testing sets of the original datasets (e.g., TAMU2015, UGA2015, and UGA2018) were merged into a single validation set. The $Seedling_{All}$ dataset was generated by combining all three datasets with the original partitioning.

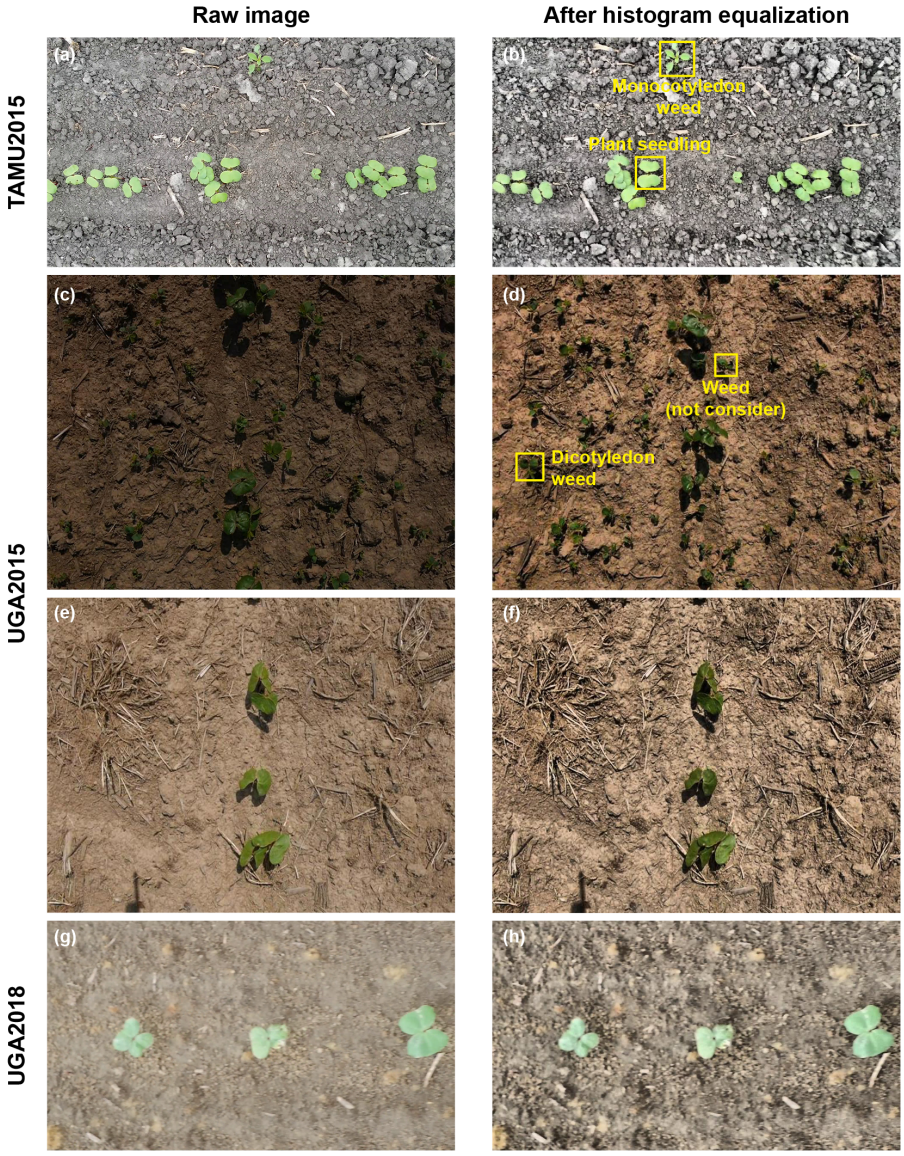

Figure 6.1: Example images in the TAMU2015, UGA2015, and UGA2018 datasets.

Table 6.2: Summary of data annotation and partitioning for the TAMU2015, UGA2015, and UGA2018 datasets

Dataset	TAMU2015	UGA2015	UGA2018
Total number of images	2204	1895	1511
Number of training images	1801	1603	1253
Number of validation images	202	146	129
Number of testing images	201	146	129
Total number of annotations (plant seedling / weed)	21915 (21133 / 782)	7802 (6849 / 953)	5880 (5880 / 0)
Number of training annotations (plant seedling / weed)	17939 (17290 / 649)	6524 (5743 / 781)	4862 (4862 / 0)
Number of validation annotations (plant seedling / weed)	1964 (1911 / 53)	643 (553 / 90)	540 (540 / 0)
Number of testing annotations (plant seedling / weed)	2012 (1932 / 80)	635 (553 / 82)	478 (478 / 0)
Number of videos for counting test	25	25	25

Table 6.3: Summary of data annotation and partitioning for the combined datasets

Dataset	T15U15	T15U18	U15U18	$Seedling_{All}$
Total number of images	4099	3715	3406	5610
Number of training images	3404	3054	2856	4657
Number of validation images	695	661	550	477
Number of testing images	N/A	N/A	N/A	476
Total number of annotations (plant seedling / weed)	29739 (27997 / 1742)	27809 (27025 / 794)	13696 (12736 / 960)	35597 (33862 / 1735)
Number of training annotations (plant seedling / weed)	24466 (23035 / 1431)	22805 (22154 / 661)	11388 (10606 / 782)	29325 (27895 / 1430)
Number of validation annotations (plant seedling / weed)	5273 (4962 / 311)	5004 (4871 / 133)	2308 (2130 / 178)	3147 (3004 / 143)
Number of testing annotations (plant seedling / weed)	N/A	N/A	N/A	3125 (2963 / 162)

6.3.2 Seedling Detection and Counting

A deep CNN based approach was developed for cotton seedling detection and counting (Figure 6.2). For a given video of cotton seedlings, video frames were extracted at the video

frame rate and enhanced using the CLAHE algorithm. The enhanced images were fed into a Faster RCNN model to detect cotton seedlings. The detected seedlings were tracked in all video frames to count the number of seedlings in the given video. Seedling detection and tracking are the key of the developed approach, so details of the two parts are provided in the following sections.

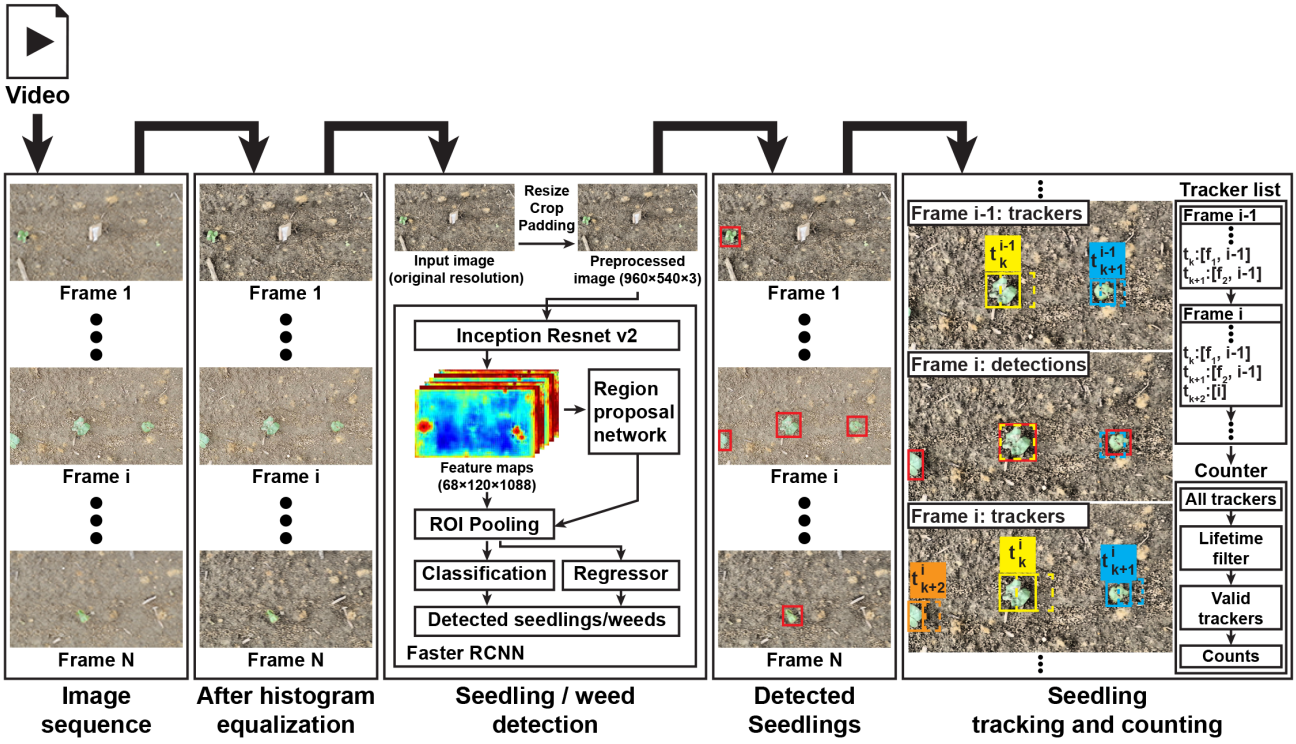

Figure 6.2: Flowchart of the deep convolutional network based approach for cotton seedling detection and counting.

Architecture, Training, and Evaluation of Detection Model

The Faster RCNN meta architecture was used due to its success in many object detection applications [30]. The architecture contains a feature extractor, a region proposal network (RPN), and a classification and regressor module. The feature extractor is usually a deep CNN network, which extracts informative feature representations from the raw input images in a hierarchical fashion. The RPN uses the extracted features to generate potential regions

of interest (ROIs), and the classification and regressor module uses the features in each ROI to identify the ROI class and refine the coordinates of ROI bounding box. In this study, the Inception ResNet v2 network [230] was used as the feature extractor due to its great potential of differentiating classes with similar appearances and shapes (e.g., dicotyledonous weed and cotton seedlings).

Transfer learning was used to improve the training efficiency and effectiveness. In the present study, the Faster RCNN model was initialized by weights pretrained on the common objects in contexts (COCO) dataset, and fine-tuned on the $Seedling_{All}$ training set. Mini-batch stochastic gradient descent (SGD) and the Adam optimizer were used for model training. While training the model, data augmentation was performed to increase the diversity of training images, including horizontal and vertical image flip and random changes of image saturation, brightness, and contrast. The Faster RCNN model and training programs were implemented using Tensorflow. Training processes were performed on two computing nodes hosted by the Georgia Advanced Computing Resource Center (GACRC), with each being configured with 14 CPU cores (2.8 GHz per core), 120 GB CPU memory, and a GPU card (Tesla V100 16GB, NVIDIA Corporation, Santa Clara, CA, USA) under the operating system of CentOS 7.5. Based on preliminary experiments, the model was trained for a total of 50,000 iterations (equivalent to 22 epochs) using an initial learning rate of 0.0001, a dropout rate of 0.5 for the RPN and classification and regressor modules, and weight decay of 0.001. Model checkpoints were saved after every 5,000 training iterations, and a total of 10 checkpoints were evaluated on the validation set to select the best model for testing and seedling counting. For the sake of brevity, the training procedure and configuration was the base training configuration.

Mean average precision (mAP), mean average recall of the top 100 detections (mAR100), and an F1 score were calculated at $IOU_{0.5}$ and IOU_{all} (from $IOU_{0.5}$ to $IOU_{0.95}$ with an

interval of 0.05) and used as metrics to evaluate the overall performance of detection models. The use of metrics at IOU_{all} was to more strictly evaluate the localization accuracy of detection models. In addition, average precision (AP), average recall of the top 100 detections (AR100), and an F1 score at $IOU_{0.5}$ were calculated for the seedling and weed classes, so per-category performance of detection could be analyzed. These evaluation metrics were used for ablation experiments as well.

Ablation Experiments

Training Sample Size While model performance benefits from a large amount of training samples, it is usually laborious to obtain a large training set for domain applications such as seedling detection. Therefore, it is important to investigate improvements of model performance due to increased training sample sizes. For each of the TAMU2015, UGA2015, and UGA2018 datasets, a total of 7 Faster RCNN models were trained using different training sample sizes, including 100, 200, 300, 500, 700, and 1000 randomly selected training images and all training images. Model training was conducted using the base training configuration, with the reduction of training iterations from 50,000 to 35,000 (based on preliminary experiments). For each training sample size, the best model checkpoint was selected based on validation performance and was used to obtain testing performance.

Transfer Learning using Different Pretrained Models In the present study, transfer learning was implemented through model initialization using pretrained weights. Thus, pretrained models may have considerable impact on the transfer learning efficiency. Model initialization using weights pretrained on large common datasets (e.g., ImageNet and COCO) would improve the model training for domain applications, especially for those with a small number of training images. However, improvements could be degraded when domain data are extremely different from the common datasets. Two model initialization methods were

used and compared to examine the transfer learning efficiency using different pretrained models: 1) model initialization using weights pretrained on the COCO dataset and 2) model initialization using weights pretrained on a domain dataset that is different from a target dataset. For example, if the UGA2018 dataset was a target dataset, Faster RCNN models would be initialized using weights pretrained on the COCO and T15U15 datasets, respectively. Subsequently, the two initialized models were trained on a subset of 100 images that were randomly selected from the UGA2018 training set. The base training configuration was used for model training and validation, with the reduction of training iterations from 50,000 to 5,000 (based on preliminary experiments). Performance on the UGA2018 testing set was obtained using the best checkpoint for each of the two models. This process was repeated 10 times, so a total of 10 testing results were calculated for each initialization method for a given target dataset, enabling statistical comparisons between the two methods. The TAMU2015 and UGA2015 were used as target datasets as well.

Model Generalizability The model generalizability was also a key factor for training deep neural networks due to the high cost of labeling a large amount of images for domain applications. To evaluate model generalizability, detection performance on a target dataset was compared between models trained using datasets acquired in the same and different data collection sessions. For instance, if the TAMU2015 testing set was a target dataset, detection performance of a Faster RCNN model trained on the TAMU2015 training set would be compared with that trained on the U15U18 dataset. The base training configuration was used for model training and validation.

Seedling Counting by Tracking

The total number of detected seedlings from all frames of a video was considerably larger than the actual number of seedlings in that video, because one cotton seedling could be repeatedly detected in consecutive video frames, resulting in recurrent counts. To address

this issue, detected seedlings were tracked in a video, so each seedling would be assigned with a single tracker and thus counted only once.

In the present study, seedling tracking was essentially to associate multiple detections (bounding boxes) of the same seedling over consecutive frames in a video. The state of a seedling tracker ($\boldsymbol{t}$) included the detection and moving speed of a seedling in a video, and was formulated using Equation 6.1.

$$\boldsymbol{t} = [u, v, s, r, \dot{u}, \dot{v}, \dot{s}]^T \tag{6.1}$$

Where u, v, s, r were the horizontal and vertical center (in pixels), area (in pixels), and aspect ratio of a bounding box. $\dot{u}, \dot{v}, \dot{s}$ were the corresponding first-order derivatives with respect to time (in the unit of video frames).

The Kalman filter [231] was used to track detected seedlings in consecutive video frames (Object tracking and counting in Figure 6.2). The seedling tracking was treated as a discrete-time filtering problem, and it was solved by two steps [232]. The first step was prediction process (also known as time update) in which states of seedling trackers in the current frame were used to predict their states in the next frame using the dynamic model of Kalman filter. The second step was update process (also known as measurement update) in which observations (seedling detections) in the next frame were associated with the seedling trackers to update the trakcer states and the dynamic model of Kalman filter. The two steps were performed alternatively over frames to track seedlings in a video.

In the first frame ($i = 1$), seedling trackers (T_1) were initialized using the seedling detections identified by a Faster RCNN model, with a value of zero for $\dot{u}, \dot{v}, \dot{s}$. Starting from the second frame ($i \geq 2$), tracker states ($\boldsymbol{t}$) and the state covariance matrix ($\boldsymbol{P}$) in the ith frame

were estimated using the information of seedling trackers in the *i-1*th frame by the prediction process (Equation6.2 and Equation6.3).

$$\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}} = \boldsymbol{F}\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i-1|i-1}},\ \boldsymbol{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.2}$$

$$\boldsymbol{P}_{\boldsymbol{i|i-1}} = \boldsymbol{F}\boldsymbol{P}_{\boldsymbol{i-1|i-1}}\boldsymbol{F}^{\boldsymbol{T}} + \boldsymbol{Q},\ \boldsymbol{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^{-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10^{-2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 10^{-4} \end{bmatrix} \tag{6.3}$$

Where $\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}}$ was the a priori estimated state of the kth seedling tracker in the ith frame, $\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i-1|i-1}}$ was the a posteriori estimated state of the kth seedling tracker in the *i-1*th frame, and $\boldsymbol{F}$ was the matrix of state transition. $\boldsymbol{P}_{\boldsymbol{i|i-1}}$ was the a priori state covariance matrix for the ith frame, and $\boldsymbol{P}_{\boldsymbol{i-1|i-1}}$ was the a posteriori state covariance matrix for the *i-1*th frame. $\boldsymbol{Q}$ was the process noise covariance matrix and determined arbitrarily in this study [232].

In the ith frame, the Kalman filter was updated by the update process using seedling trackers in the *i-1*th frame (T_{i-1}) and seedling detections in the current frame (D_i). As detections were the ground truth measurements for existing trackers, it was necessary to associate detections and trackers for updating the Kalman filter. The cost of assigning a new

detection ($\boldsymbol{d_j} \in D_i$) to an existing tracker ($\boldsymbol{t_k} \in T_{i-1}$) was the negative IOU value between the detection ($\boldsymbol{d_j}$) and the tracker's predicted detection ($\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}}$). The assignment task was optimally solved using the Hungarian algorithm [233] that minimized the assignment cost under a constraint of the minimum IOU value of 0.1 (Equation 6.4).

$$
\begin{aligned}
\min \quad & \sum_{j=1}^{N(D_i)} \sum_{k=1}^{N(T_{i-1})} c(\boldsymbol{d_j}, \hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}}) a_{j,k}, \quad a_{j,k} = \begin{cases} 1, & \boldsymbol{d_j} \; assigned \; to \; \hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}} \\ 0, & otherwise \end{cases} \\
subject \; to \quad & \forall j, \quad a_{j,k_1} = a_{j,k_2} \Rightarrow k_1 = k_2, \quad k_1, k_2 = 1, \ldots, N(T_{i-1}) \\
& \forall k, \quad a_{j_1,k} = a_{j_2,k} \Rightarrow j_1 = j_2, \quad j_1, j_2 = 1, \ldots, N(D_i) \\
& \forall j \; and \; k, \quad |c(\boldsymbol{d_j}, \hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}})| > 0.1
\end{aligned}
\tag{6.4}
$$

Where D_i was the set of seedling detections in the ith frame and T_{i-1} was the set of trackers in the *i-1*th frame. $N(\cdot)$ was the function counting the number of elements in a set. $\boldsymbol{d_j}$ was the state of the jth seedling detection in D_i and $\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}}$ was the a priori estimated state of the kth tracker ($\boldsymbol{t_k}$) in T_{i-1}. $c(\boldsymbol{d_j}, \hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}})$ was the cost for assigning $\boldsymbol{d_j}$ to $\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}}$ (i.e., the negative IOU value between the bounding boxes of $\boldsymbol{d_j}$ and $\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}}$), and $a_{j,k}$ indicated the assignment flag, with one for assigning $\boldsymbol{d_j}$ to $\hat{\boldsymbol{t}}_{\boldsymbol{k}}^{\boldsymbol{i|i-1}}$. It should be noted that one detection could be only assigned to one tracker or otherwise unassigned.

After the detection-tracker association, the detections (D_i) and trackers (T_{i-1}) were categorized into three groups: trackers associated with new detection, unassigned detections, and unassociated trackers. Trackers associated with new detection were used for the update process that calculated the a posteriori state covariance matrix in the ith frame ($\boldsymbol{P_{i|i}}$) using Equation 6.5 to Equation 6.7 and their a posteriori states of the trackers using Equation 6.8 and Equation 6.9.

$$\boldsymbol{S_i} = \boldsymbol{H P_{i|i-1} H^T} + \boldsymbol{R},\ \boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},\ \boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix} \tag{6.5}$$

$$\boldsymbol{K_i} = \boldsymbol{P_{i|i-1} H^T S_i^{-1}} \tag{6.6}$$

$$\boldsymbol{P_{i|i}} = (\boldsymbol{I} - \boldsymbol{K_i H}) \boldsymbol{P_{i|i-1}} (\boldsymbol{I} - \boldsymbol{K_i H})^T + \boldsymbol{K_i R K_i^T} \tag{6.7}$$

$$\boldsymbol{y^i} = \boldsymbol{d^t} - \boldsymbol{H \hat{t}^{i|i-1}_{associated}} \tag{6.8}$$

$$\boldsymbol{\hat{t}^{i|i}_{associated}} = \boldsymbol{\hat{t}^{i|i-1}_{associated}} + \boldsymbol{K_i y^i} \tag{6.9}$$

Where $\boldsymbol{S_i}$ was the innovation covariance matrix of the Kalman filter in the ith frame. $\boldsymbol{H}$ was the measurement matrix that mapped a tracker state to a measurement (detection) state. $\boldsymbol{R}$ was the measurement error covariance matrix and determined arbitrarily in this study. $\boldsymbol{K_i}$ and $\boldsymbol{I}$ were the optimal Kalman gain for the ith frame and identity matrix, respectively. $\boldsymbol{y^i}$ was the innovation (also known as measurement residual) between the a priori estimated state of a tracker ($\boldsymbol{\hat{t}^{i|i-1}_{associated}}$) and the state of that tracker's associated detection ($\boldsymbol{d^t}$) in the ith frame, and $\boldsymbol{\hat{t}^{i|i}_{associated}}$ was the a posteriori estimated state of that tracker.

Unassigned detections (zeros for $\dot{u}, \dot{v}, \dot{s}$) were initialized as new trackers and added into the existing tracker set (T_{i-1}). Unassociated trackers from T_{i-1} were removed, forming the new tracker set for the ith frame (T_i). All frames were processed sequentially using the prediction and update processes of the Kalman filter, which provided a list of trackers with their lifetime (number of video frames in which trackers existed).

In theory, the number of trackers would be the number of seedlings in a video as one tracker exactly corresponded to one seedling. However, in practice, seedling detection models could occasionally provide inaccurate detection results, resulting in potential initialization of noisy trackers. Thus, a lifetime filter was used to select valid trackers (trackers with a lifetime longer than a threshold), and the number of valid trackers was used as the number of seedlings in a video. In the present study, the lifetime threshold was arbitrarily set as a quarter of video frame rate, which was 7 for TAMU2015 and UGA2018 videos and 15 for the UGA2015 testing videos.

Counting Accuracy Evaluation

The developed approach with $model_{SAll}$ was used to count the number of seedlings in the 75 testing videos collected in multiple locations and years. Simple linear regression tests were performed between the video-derived counts and human field assessment, and the fitted slope, adjusted coefficient of determination (R^2), and root mean squared error (RMSE) were used as the evaluation metrics. Mean absolute error (MAE) and mean relative error (MAR) were also calculated as additional metrics for counting accuracy evaluation. In addition, the developed approach with $model_{U15U18}$, $model_{U15T15}$, and $model_{T15U18}$ was tested using 25 testing videos of TAMU2015, UGA2018, and UGA2015, respectively. The same metrics were calculated to compare with those calculated using $model_{SAll}$, which provided an evaluation of the counting accuracy of unknown datasets.

6.4 Results

6.4.1 Detection Performance on $Seedling_{All}$ Dataset

Overall, the Faster RCNN model ($model_{SAll}$) trained using the $Seedling_{All}$ training set achieved an F1 score of 0.727 (at IOU_{all}) and 0.969 (at $IOU_{0.5}$) on the $Seedling_{All}$ testing set in which images had large variations. The $model_{SAll}$ had even a better performance (F1 score of 0.998) for the seedling class, indicating the efficacy of the Faster RCNN model with

an Inception ResNet v2 feature extractor for seedling detection. The $model_{SAll}$ successfully addressed various challenges in the testing images (Figure 6.3). The primary challenge in the testing images originally collected in the TAMU2015 dataset was occlusion. Despite excessive overlap between seedling objects, the $model_{SAll}$ could detect all seedlings with corresponding bounding boxes that were tightly fitted to the seedlings ((a) and (b) in Figure 6.3). The key challenges in testing the originally collected images in the UGA2015 dataset were the background complexity and presence of dicotyledonous weeds. The background was relatively simple (no weeds) in some images but could be complex (many weeds) in other images (compare (c) and (d) in Figure 6.3). Seedings were accurately detected by the $model_{SAll}$ even under shaded conditions (the second topmost detection in Figure 6.3d). However, weeds (especially small-sized weeds) were not correctly detected when the background was very complex (dashed rectangles in Figure 6.3d). It should be noted that there was no misclassification between dicotyledonous weeds and cotton seedlings, both of which are similar in appearance (e.g., color and shape). These observations suggested that the feature extractor (Inception ResNet v2) should be powerful enough to extract features for differentiation of classes with similar appearance (cotton seedlings), whereas the RPN network was probably insufficient to provide proposals for regions of interest (ROIs) of small-sized weeds. In fact, detection of small-sized objects is a challenging issue for region based neural networks, especially for datasets with a mixture of objects of other sizes [234]. Nonetheless, the trained models would be acceptable as the primary goal of this study was to detect seedlings rather than weeds.

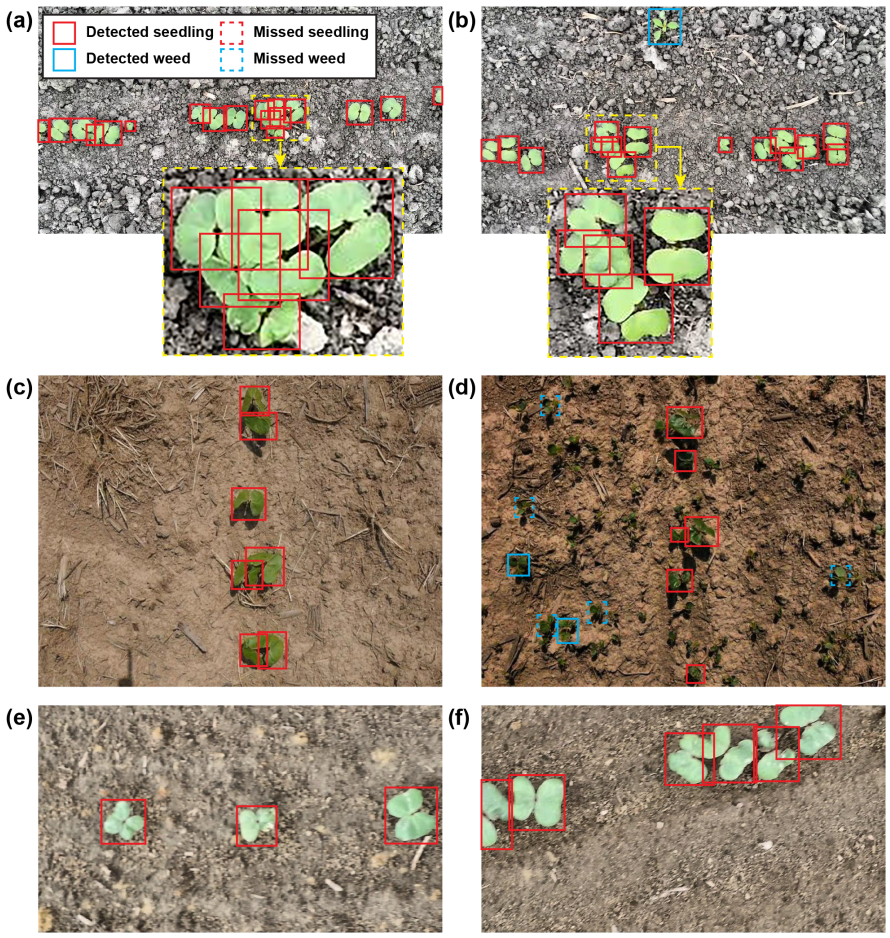

Figure 6.3: Cotton plant seedlings and weeds detected in representative images of the $Seedling_{All}$ testing set by the Faster RCNN model that was trained using the $Seedling_{All}$ training set. (a) and (b) are images originally collected in the TAMU2015 dataset, (c) and (d) in the UGA2015 dataset, and (e) and (f) in the UGA2018 dataset.

6.4.2 Ablation Experiment Results

Training Sample Size

Generally, model performance was improved with the increasing of training sample size, but the improvements heavily depended on the evaluation metric and task difficulty (Figure 6.4). IOU_{all} is more strict than $IOU_{0.5}$, forcing a higher model accuracy for object localization. Compared with $IOU_{0.5}$, it was clearer to observe performance improvements due

to increased training sample size at IOU_{all}. The model ($model_{TAMU2015}$) trained on the TAMU2015 dataset reached the maximum F1 score after 500 and 1000 training images at $IOU_{0.5}$ and IOU_{all}, respectively (Figure 6.4a). The model ($model_{UGA2018}$) trained on the UGA2018 dataset showed a slightly increasing trend at IOU_{all} (compare the curves at $IOU_{0.5}$ and IOU_{all} in Figure 6.4c). Thus, more training images would benefit the training of the bounding box regressor of Faster RCNN models. In addition, the difficulties of seedling and weed detection were different in various images. Based on the results of the $Seedling_{All}$ dataset, seedling and weed detections were the most difficult in the images from the UGA2015 dataset, followed by the TAMU2015 and UGA2018 datasets. When using the same evaluation metric, the increasing trend of model performance with higher number of images was evident for the UGA2015 dataset but less obvious for the TAMU2015 and UGA2018 datasets. This finding held true for individual classes (Figure 6.5). For the seedling class, datasets containing challenging situations (e.g., excessive occlusion in the TAMU2015 dataset and high similarities between classes in the UGA2015 dataset) required more than 300 training images to reach the best performance, whereas less-challenging datasets could use only 100 training images for the best result. For the weed class, the precision-recall (PR) curves mostly expanded to be closer to the ideal PR curve (the top-right border) when increasing the training sample size, meaning that weed detection would be relatively more complex than seedling detection for Faster RCNN models. This was because the weed class was more diverse than the seedling class in the datasets. Cotton seedling was essentially a single biological category in spite of variations due to environment changes, whereas weeds contained many subspecies that could be intrinsically different from each other in appearances. It was also noteworthy that the expansion magnitudes were different for the TAMU2015 and UGA2015 datasets. The TAMU2015 dataset contained monocotyledonous weeds that were distinct from cotton seedlings, whereas the UGA2015 dataset contained dicotyledonous weeds that were similar to cotton seedlings. Therefore, more training images

provided more improvements for the UGA2015 dataset than for the TAMU2015 dataset.

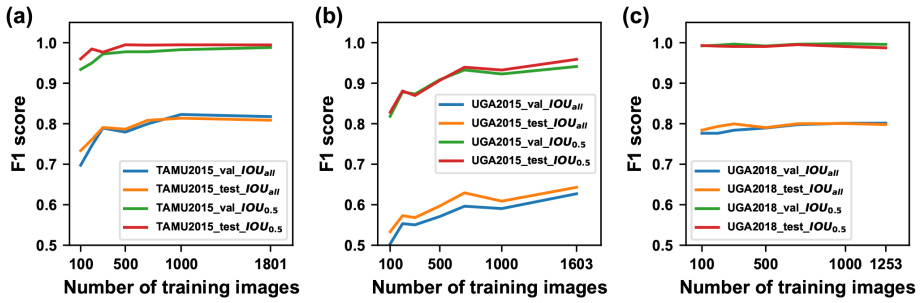

Figure 6.4: Detection performance (F1 score) calculated using different number of training images for: (a) the TAMU2015 dataset, (b) the UGA2015 dataset, and (c) the UGA2018 dataset.

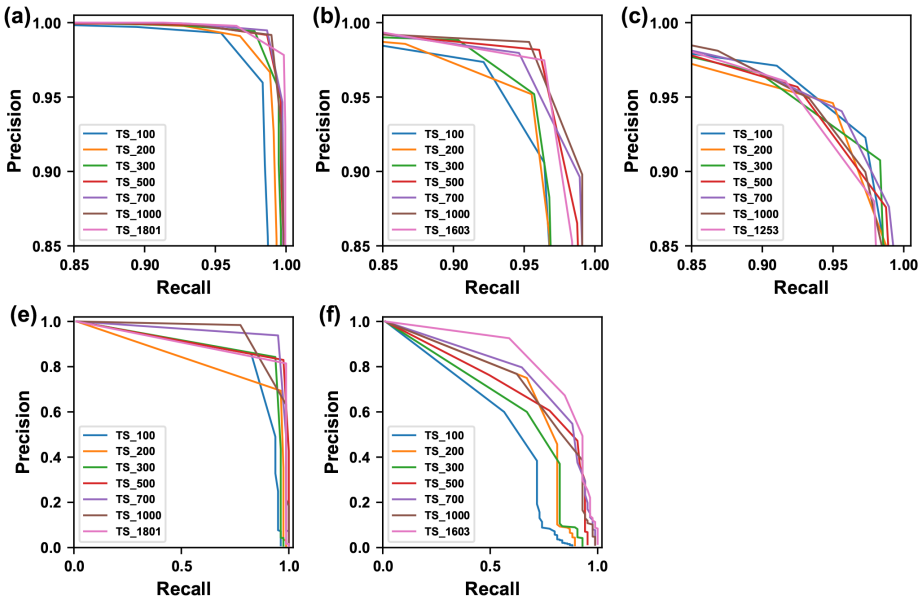

Figure 6.5: Per-category precision-recall curves generated using different number of training images. (a), (b), and (c) are for seedling detection in the TAMU2015, UGA2015, and UGA2018 datasets, and (e) and (f) are for weed detection in the TAMU2015 and UGA2015 datasets.

Transfer Learning using Different Pretrained Models

Transfer learning by model initialization using weights pretrained on a domain dataset showed varied efficiencies. These efficiency variations were dependent on evaluation metrics and datasets (Figure 6.6). When using a strict evaluation metric, model initialization using weights pretrained on a domain dataset generally yielded better performance than that using weights pretrained on a common dataset such as the common objects in context (COCO) dataset (overall performance at IOU_{all} in Figure 6.6). This occurred primarily because two domain datasets were likely similar to each other, so learned weights from one domain dataset could be more beneficial for the model training process for another domain dataset. For instance, compared with weights pretrained on a common dataset, weights (especially for the bounding box regressor) pretrained on a domain dataset would be closer to the optimal values for another domain dataset, resulting in better localization of objects and thus F1 score at IOU_{all}. An exception was identified for the experiment on the TAMU2015 dataset: the overall F1 score from the model initialized using weights pretrained on the COCO dataset was slightly higher than that on the U15U18 dataset (compare the overall performance at $IOU_{0.5}$ and IOU_{all} in Figure 6.6c). This occurred mainly because of the reduced similarity between the two domain datasets (TAMU2015 and U15U18 datasets). The TAMU2015 contained monocotyledonous weeds, whereas the U15U18 contained dicotyledonous weeds. In such a case, due to a higher object diversity, weights (features) learned on the COCO dataset would have higher possibility of representing objects in the TAMU2015 than those learned on the U15U18 dataset (see weed detection performance in Figure 6.6c). Thus, transfer learning efficiency of using weights pretrained on a common dataset was better than a domain dataset.

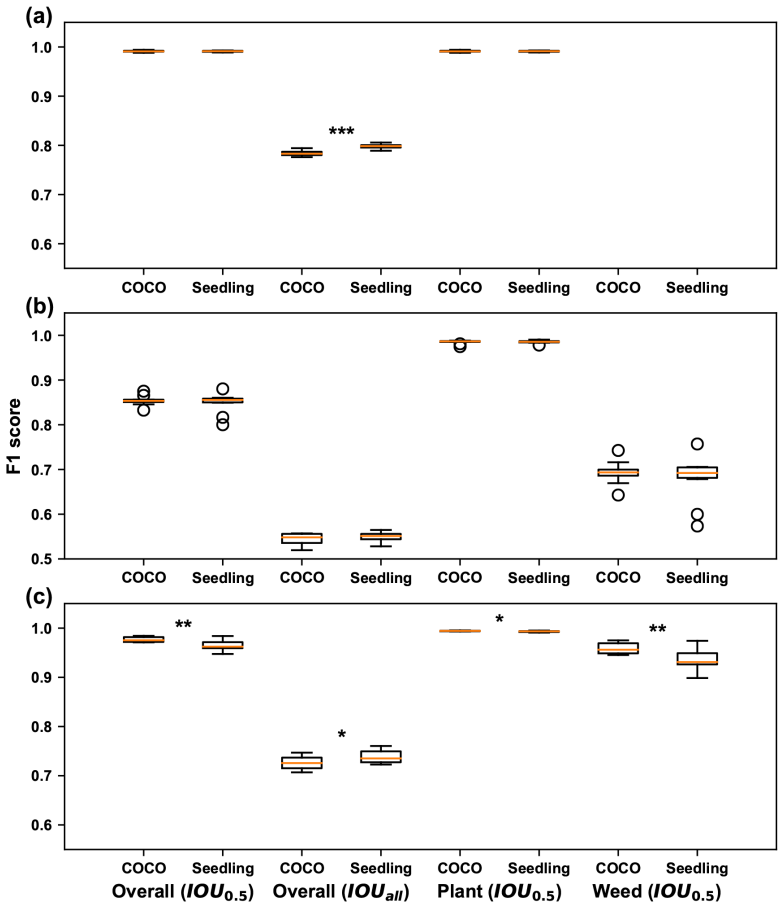

Figure 6.6: Boxplots of performance (F1 score) on the testing set for models initialized using different pretrained models. (a) are results for the UGA2018 dataset using models initialized by weights pretrained on the COCO and T15U15 datasets, respectively, (b) are results for the UGA2015 dataset using models initialized by weights pretrained on the COCO and T15U18 datasets, respectively, and (c) are results for the TAMU2015 dataset using models initialized by weights pretrained on the COCO and U15U18 datasets, respectively. Base indicates model initialization using weights pretrained on the COCO dataset, whereas DA indicates model initialization using weights pretrained on a domain dataset. For each of the TAMU2015, UGA2015, and UGA2018 datasets, a subset of 100 images were randomly selected from the training set to train a Faster RCNN model. A total of 10 models were obtained through 10 training repetitions for statistical comparisons between the models. Asterisks indicate statistical differences in model performance at the significance levels of 0.05 (*), 0.01 (**), and less than 0.001 (***).

Model Generalizability

The generalizability of trained Faster RCNN models was largely data dependent. The Faster RCNN model trained using the T15U15 dataset ($model_{T15U15}$) provided comparable performance (1% difference) to the $model_{UGA2018}$ for the UGA2018 testing set, indicating a strong model generalizability to new datasets (see results for the UGA2018 testing set in Table 6.4). This occurred primarily because the T15U15 dataset was diverse to cover possible object status (e.g., appearance variation and object occlusion) in the UGA2018 dataset. On the contrary, Faster RCNN models trained using the U15U18 ($model_{U15U18}$) and T15U18 ($model_{T15U18}$) datasets showed a substantially decreased performance for the testing set of TAMU2015 (decreasing of 36%) and UGA2015 (decreasing of 39%) respectively. This was because of large differences between training and testing datasets. Compared with the TAMU2015 dataset, the U15U18 dataset contained no monocotyledonous weed, so the $model_{U15U18}$ could not correctly detect weeds in the testing images of the TAMU2015 dataset (compare results for the TAMU2015 testing set in Table 6.5). As the U15U18 dataset contained occluded seedling objects (in spite of less occurrence than in the TAMU2015 dataset), the $model_{U15U18}$ showed an acceptable generalizability (7% reduction to the $model_{TAMU2015}$) for seedling detection. Similarly, the T15U18 dataset contained no dicotyledonous weed, so the $model_{T15U18}$ had poor performance of weed detection for the UGA2015 testing set. The missing of dicotyledonous weeds led to performance reduction to both the seedling and weed classes. Dicotyledon weeds are very similar to cotton seedlings in appearance and shape, resulting in misclassification between weeds and seedlings.

Table 6.4: Overall performance of the model generalizability

Training set	Testing set	mAP IOU_{all}	mAR100 IOU_{all}	F1 IOU_{all}	mAP $IOU_{0.5}$	mAR100 $IOU_{0.5}$	F1 $IOU_{0.5}$
T15U15 training	UGA2018 testing	0.763	0.808	0.785	0.981	1.000	0.991
UGA2018 training	UGA2018 testing	0.778	0.818	0.798	0.983	0.992	0.987
T15U18 training	UGA2015 testing	0.179	0.335	0.233	0.352	0.683	0.464
UGA2015 training	UGA2015 testing	0.599	0.695	0.643	0.923	0.997	0.959
U15U18 training	TAMU2015 testing	0.377	0.535	0.442	0.588	0.857	0.698
TAMU2015 training	TAMU2015 testing	0.791	0.827	0.809	0.989	1.000	0.995

Table 6.5: Per-category performance of the model generalizability at the $IOU_{0.5}$

Training set	Testing set	Plant AP	Plant AR	Plant F1	Weed AP	Weed AR	Weed F1
T15U15 training	UGA2018 testing	0.981	1.000	0.991	NA	NA	NA
UGA2018 training	UGA2018 testing	0.983	0.992	0.987	NA	NA	NA
T15U18 training	UGA2015 testing	0.676	0.814	0.739	0.027	0.553	0.052
UGA2015 training	UGA2015 testing	0.988	0.995	0.991	0.859	1.000	0.924
U15U18 training	TAMU2015 testing	0.911	0.940	0.925	0.266	0.775	0.396
TAMU2015 training	TAMU2015 testing	0.999	0.999	0.999	0.980	1.000	0.990

6.4.3 Counting Accuracy

Overall, seedling counts that were calculated using the developed approach with the $model_{SAll}$ model were highly correlated ($R^2 = 0.98$) with those by human field assessment (Figure 6.7a). The slope of the regression equation was one, suggesting that seedling counts calculated by the developed approach can be used directly. For the 75 testing videos, 53 (70%) videos had an absolute counting error less than or equal to 1 seeding, and 15 (20%) videos with an absolute counting error greater than 1 had a relative counting error of less than 10% (Figure 6.7b). Thus, a total of 90% of the testing videos showed acceptable counting accuracies, whereas the rest 10% of the testing videos showed large counting errors (larger than 15%). Nonetheless, the mean relative error for all the 75 videos was 7%,

indicating the efficacy of the developed counting approach.

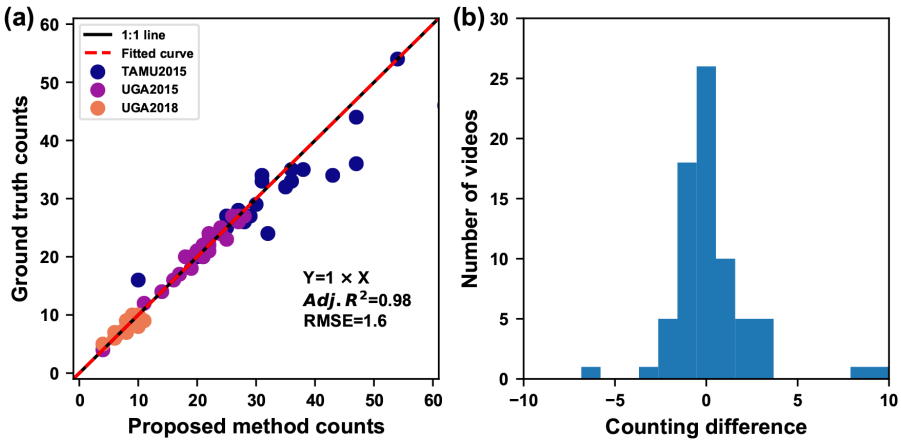

Figure 6.7: Regression results between seedlings counts calculated by the developed approach and human field assessment. (a) Results obtained using the $model_{SAll}$ detection model for all testing videos (n=75, lifetime threshold of 7 was used for the TAMU2015 and UGA2018 testing videos and 15 for the UGA2015 testing videos). (b) Counting differences between the proposed method and field assessment. A total of 53 videos (70%) had the counting differences less than 1 seedling, and 68 videos (90%) had the counting differences less than 5 seedlings.

Detection performance showed strong influences on the counting accuracy of the developed approach (Table 6.6). When detection models (e.g., $model_{T15U15}$ and $model_{U15U18}$) were well generalized to seedling detection, the developed approach with such detection models provided similar counting accuracies for videos collected in separate data collection sessions (see results for TAMU2015 and UGA2018 testing videos in Table 6.6). It was noteworthy that R^2 values decreased primarily due to the narrow range of seedling quantities. As a result, subtle counting errors could have considerable influences on the R^2 values. When detection models (e.g., $model_{T15U18}$) had poor generalizability to seedling detection, the counting performance of the developed approach with such models degraded dramatically

(see results for UGA2015 testing videos in Table 6.6).

Table 6.6: Regression results between seedling counts obtained by the proposed method and human field assessment.

Detection model	Testing videos	Regression equation	Video quantity	R^2	RMSE	MAE	MRE
$model_{SAll}$	All	Y = X	75	0.98	1.6	1.5	7%
$model_{SAll}$	TAMU2015	Y = 0.96 X	25	0.85	3.3	3.4	11%
$model_{SAll}$	UGA2018	Y = X	25	0.99	0.2	0.5	6%
$model_{SAll}$	UGA2015	Y = 1.01 X	25	0.96	1.1	0.8	4%
$model_{U15U18}$	TAMU2015	Y = 0.95X	25	0.80	3.9	3.4	11%
$model_{U15T15}$	UGA2018	Y = 1.02X	25	0.68	0.8	0.6	7%
$model_{T15U18}$	UGA2015	Y = 0.32X + 18.65	25	0.33	4.3	12.1	57%

R^2 refers to adjusted R^2

RMSE: root mean squared error

MAE: mean absolute error

MRE: mean relative error

The reduction of counting accuracy was primarily due to tracking errors caused by inaccurate seedling detection (Figure 6.8). In one of the TAMU2015 testing videos, the detection results were accurate in Frame 60 and Frame 63 but not accurate in Frame 61 and Frame 62. Consequently, some trackers lost continuity in tracking between Frame 60 and the following video frames. When all seedlings were correctly detected in Frame 63 again, although some of the detected seedlings were the same in Frame 60 and Frame 63, they were assigned to new trackers due to the tracking discontinuity. As a consequence, the number of seedling trackers would be higher than the actual number of seedlings from Frame 60 to Frame 63, which ultimately led to inaccurate counts of seedlings in that video. Depending

on the lifetime of new trackers (the number of frames new trackers could last), the seedling count in a video could be higher or lower than the actual value. This would be a major error source for the developed approach.

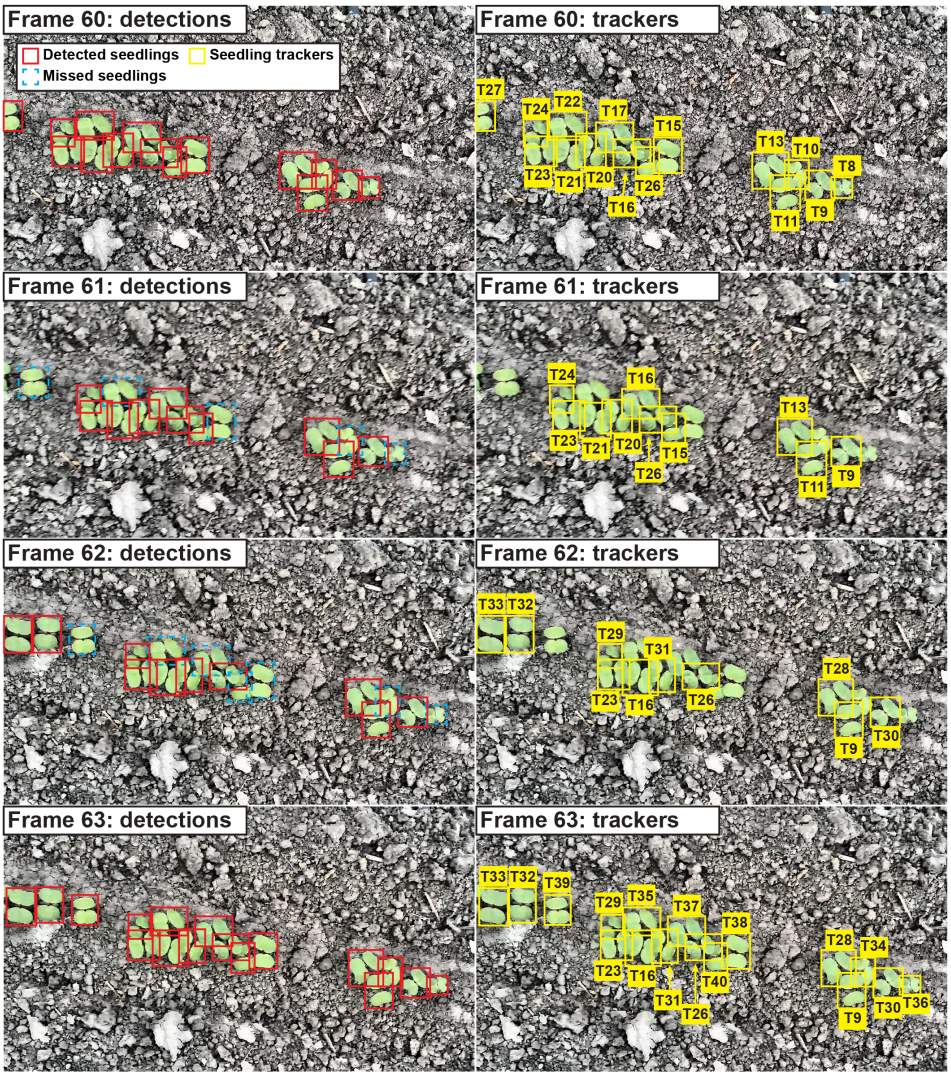

Figure 6.8: An example of seedling tracking errors due to inaccurate detection results. The detection model was $model_{SAll}$ and the testing video was from TAMU2015 dataset.

6.5 Discussion

In terms of counting performance, the developed approach provides a similar accuracy (93%) of seedling counting to other CNN-based approaches for fruit counting [81, 86, 80], showing the efficacy of using CNN-based approaches for seedling counting. Compared with the approach based on conventional image processing for seedling counting [226], the developed approach shows two advantages. First, the counting accuracy was improved from 88% to 92%. Second, and more importantly, the developed approach shows great potential to be generic for cotton seedling detection. Experiments showed trained seedling detection models, and the counting approach could be well generalized to unseen datasets, so they can potentially be used in similar applications by the cotton industry and research communities with little or no modification. To the best of our knowledge, the $Seedling_{All}$ dataset is the largest annotated dataset of cotton seedlings, and publicizing such a dataset would benefit both research communities and the cotton industry.

While showing certain advantages, the developed approach has two major limiting factors. First, detection models considerably influence the counting accuracy of the developed approach. It is not a trivial task to train an accurate and robust detection model in many applications. Three important factors have been examined in the present study, including the training sample size, transfer learning using different pretrained models, and model generalizability. Experimental results showed that all three factors were somehow data dependent. On one hand, if agronomic practices (e.g., application of pre-emergent herbicides) are implemented in a future project, seedling detection would be fairly simple due to few object categories in videos/images. Thus, seedling detection can be solved using either the $model_{SAll}$ directly, or using a new detection model that is initialized with the $model_{SAll}$ and trained on a small number of annotated images (100 to 300 images based on the present study) from the newly collected dataset. In particular, for a specific experimental site or farm, it is highly recommended to be consistent in data collection conditions (e.g., cameras

for image acquisition, camera configuration, and illumination conditions) to enhance the similarity between datasets collected over time and thus counting accuracies for long-term uses. If data in a future project are more complex (e.g., more frequent occlusions, more extreme illumination conditions, and more types of vegetation) than any of the datasets in this study, it is necessary to label a fairly large amount of data to ensure the possibility of achieving the best detection performance. If so, the value of trained models and annotated data in the present study may be relatively reduced in future studies.

Secondly, a conventional Kalman-filter-based tracking algorithm is not adequate to solve issues caused by inaccurate detection of seedlings. When a cotton seedling cannot be correctly detected in consecutive video frames, the current tracking algorithm is likely to terminate the seedling tracker in the frame where that seedling is mis-detected, and assign a new tracker in the next frame where that seedling is re-detected. Thus, that seedling could be counted repeatedly in a video, resulting in counting errors. This occurs primarily because the current strategy of detection/tracker assignment is based on the IOU metric. No detection means no intersection with any existing trackers. There are two ways to solve this issue. First, a new strategy can be used for tracker termination. If no detection can be assigned to a tracker, that tracker can be kept for extended frames (e.g., 3 video frames), which can address the tracking discontinuity due to inaccurate detection results to some extent. In the extended frames, the dynamic model for that tracker is updated using the information obtained in the last frame where that tracker has an associated detection, reducing the model accuracy. Thus, this strategy requires an additional checking procedure that ensures the correctness of detection and tracker association. Feature-based approaches are preferred to maximize the checking accuracy. Second, tracking information can be used to improve detection accuracy. The developed approach solely relies on the detection procedure to provide "ground truth measurements" for trackers. Thus, if the detection procedure is not accurate, the tracking procedure cannot be accurate. To address this issue, the tracking information needs to be

used for the detection as well. For instance, the RPN of Faster RCNN could provide inaccurate region proposals (e.g., no region proposal around an existing tracker), leading to misdetection of seedlings and therefore inaccurate tracking. In fact, the tracking procedure predicts bounding boxes of all existing trackers in the next frame. These predicted bounding boxes can be used as region proposals for detection models (e.g., a Faster RCNN model) or be evaluated by a separately trained CNN (e.g., a ResNet model) that determines the presence of plants (classified as either background or plants). With these efforts, it is expected to reduce the possibility of missing an existing tracker in the next frame and ultimately improve the tracking and counting accuracy. This detection-tracking continuum may violate some assumptions of the Kalman filter in a strict consideration. For instance, Kalman filter assumes that sensor measurement is independent of the dynamic model. To solve these potential issues, it is necessary to use other filtering approaches such as the particle filter. These solutions need to be further explored in future studies.

6.6 Conclusions

The developed approach based on deep CNNs can accurately count germinated cotton seedling in the field. Experimental results showed that the approach generalized well to unseen datasets, indicating the great potential of applying the approach for other plant or plant organ detection and tracking. Trained detection models and the annotated images can be reused by the research communities and the cotton industry. Future studies will be focused on the improvement of computation efficiency for real time online processing.

Chapter 7

DeepFlower: A Deep Learning Framework to Characterize Flowering Patterns of Angiosperms in the Field[1]

[1]Jiang, Y., Li, C., Xu, R., Sun, S., Robertson, J. S., and Paterson, A. H. To be submitted to *Plant Physiology*.

7.1 Abstract

Flowering is one of the most important processes for angiosperms (flowering plants), reflecting the transition from vegetative to reproductive growth and central importance to crop yield and adaptability. Conventionally, categorical scoring systems have been widely used to study flowering patterns, which are laborious and subjective. The goal of this study was to develop a deep learning based approach to characterize flowering patterns, as an example using cotton plants that flower progressively over several weeks, with flowers distributed across much of the plant. A ground mobile system (GPhenoVision) was modified with a multi-view color imaging module, to acquire images of a plant from four viewing angles at a time. A total of 116 plants from 23 genotypes were imaged during an approximately 2-month period with an average scanning interval of 2 to 3 days, yielding a dataset containing 8666 images. A subset (475) of the images were randomly selected and manually annotated to form training and testing datasets for training and selecting the best Faster RCNN model. With the best Faster RCNN model, a deep learning based framework (DeepFlower) was developed to detect and count individual emerging blooms for a plant on a given date. The DeepFlower framework was used to process all images to obtain bloom counts for individual plants over the flowering period, using the resulting counts to derive flowering curves (and thus flowering characteristics). Regression analysis showed that the DeepFlower framework could accurately (R^2 = 0.88 and RMSE = 0.79) detect and count emerging blooms on cotton plants, and statistical analysis showed that imaging-derived flowering characteristics had the same effectiveness as manual assessment for identifying differences among genetic categories or genotypes. The developed approach could thus be an effective and efficient tool to characterize flowering patterns for angiosperms with complex canopy architecture.

7.2 Introduction

Flowering is one of the most important processes for angiosperms (flowering plants), reflecting the transition from vegetative to reproductive growth and significantly affecting crop yield and adaptability to various environments. Therefore, characterization of flowering patterns (especially flowering time) would not only facilitate studies for understanding angiosperms genetically and physiologically, but also holding potential to contribute to breeding of new cultivars for optimal yield and environmental adaptability [235, 236, 237].

Conventionally, studies related to plant flowering patterns have required human evaluators to check experiment fields and record flowering status manually. For instance, plants and plots can be checked regularly by human evaluators to monitor characteristics such as the number of days after planting to the first bloom. In addition, human evaluators often used a categorical scoring system to assess flowering stages (e.g., estimating when 10% of plants in a plot have opened blooms), so that the time duration between particular flowering stages can be calculated. Human recorded flowering data have helped researchers to study flowering patterns for several important crops such as maize [238], rice [239], cereal [240], and sorghum [241]. Human evaluation, however, has two major disadvantages. First, the evaluation is subjective, which means that different human evaluators might give different scores for the same plant/plot. As a result, collected flowering data could contain a substantial amount of noise. Second, human evaluation is laborious, and presents great challenges for large-scale experiments and breeding programs. An automated high throughput approach to characterize flowering patterns can mitigate each of these disadvantages.

Advances in high throughput plant phenotyping and breakthroughs in deep learning enable the possibility of rapid characterization of flowering patterns for plants in the field. Several studies demonstrated the use of deep convolutional neural networks (CNNs) and meta-models (e.g., Faster RCNN meta-model [45]) to detect and count fruit in images for

crops such as mangoes [80], apples [86], and sweet peppers [81]. Although these studies achieved relatively high counting accuracies (R^2 >0.92), they were primarily used for "one time" measurements of yield estimation. A big challenge to flowering characterization is that in many plants it occurs over a long period of time, requiring one to frequently detect and count newly opened blooms on plants. Two very recent studies reported the use of deep CNNs to characterize flowering patterns. One study reported a two-stage framework to detect and count blooms in cotton plots from aerial images [85]. The framework first segmented candidate regions of blooms using a thresholding method, and subsequently classified the candidate regions as bloom or non-bloom using a custom CNN to count the number of blooms in individual cotton plots. Although the two-stage framework showed some success in counting blooms, it had a major limitation in that a considerable portion of blooms could not be captured by aerial images because of occlusions, resulting in a relatively large underestimation of bloom counts. In addition, this study only measured bloom counts for several days, lacking of the capability for flowering pattern analyses over a long flowering period. The other study reported a CNN-based approach that could classify ground images of wheat plant into 10 levels (from 0% to 100% with an interval of 10%) of inflorescence load [79]. An empirical equation was used to fit inflorescence load scores over a flowering period to estimate the peak flowering time for individual wheat plots. Experimental results showed that the imaging-derived peak flowering time had the same power to identify quantitative trait loci as the peak flowering time derived manually. As wheat forms floral heads mainly at the top of plants, however, this method also did not consider situations (e.g., cotton or many other plants) in which blooms could not always be seen from the top view. In addition, the study simplified the bloom counting problem as a classification task, which lacked the capability of detecting individual blooms. To address the aforementioned limitations, it is necessary to explore the use of side-view ground images and CNN based object detection meta-models to detect and count individual blooms over an entire flowering period.

The goal of the study was to develop a deep learning based approach to characterize flowering patterns of angiosperms in the field. Specific objectives were to 1) develop a multi-view imaging system that can acquire images of plants in a high throughput manner; 2) develop a deep learning based framework (DeepFlower) to detect and count emerging blooms on plants in images and to characterize flowering patterns for individual plants; and 3) evaluate the accuracy and efficacy of the developed approach for identification of differences in flowering patterns among genetic categories and genotypes.

7.3 Materials and Methods

7.3.1 High Throughput Imaging System and Experimental Design

A previously developed ground mobile imaging system (GPhenoVision [205]) was modified with a multi-view color imaging module for data acquisition (Figure 7.1A). The multi-view color imaging module consisted of four consumer grade mirror-less cameras (X-A10, Fujifilm Holdings Corporation, Tokyo, Japan) that faced towards the center of the system enclosure approximately 90° apart from neighboring cameras. To avoid potential issues of image quality (e.g., blurry images) because of high-frequency vibration, an inexpensive camera mount was manufactured by combining a camera ball mount and a vibration isolator, providing flexibility of the viewing angle configuration and the capability of isolating high-frequency vibrations (Figure 7.1B). A custom trigger device was developed to synchronize triggering signals to all four cameras. The trigger device and an RTK-GPS (Cruizer II, Raven Industries Inc., Sioux Falls, SD, USA) were connected to a laptop computer in which a custom LabVIEW program ran to automatically save timestamps of triggering signals and RTK-GPS records. The developed data acquisition system acquired four color images at a time with RTK-GPS information.

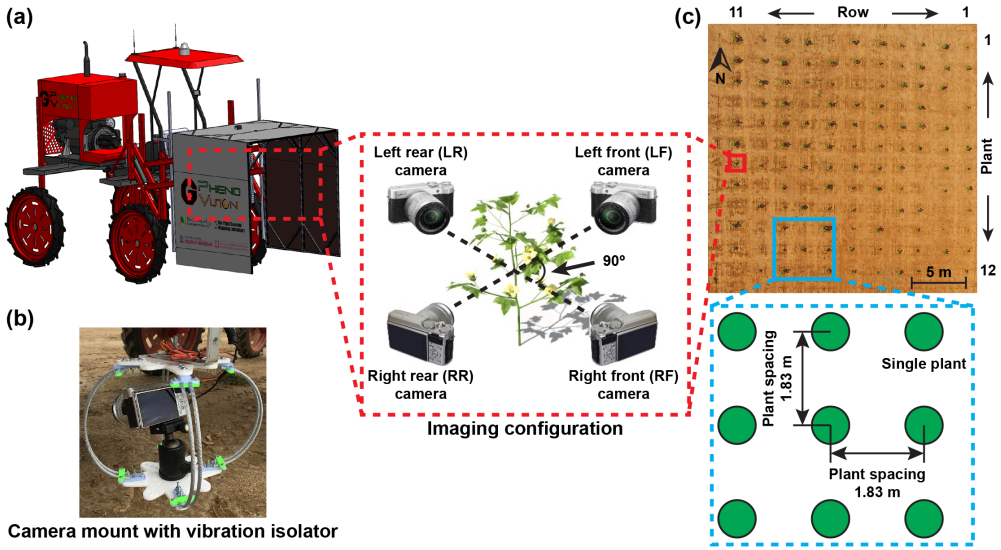

Figure 7.1: Diagram of the data acquisition system and field layout. A: GPhenoVision system with the color imaging module for acquiring four-view images of plants. B: Implementation of a specially designed camera mount for isolating high frequency vibration. C: The single plant layout (SPL) field used in the present study.

Cotton seeds of 24 genotypes (from 3 genetic categories including *Gossypium hirsutum*, *Gossypium hirsutum*, and *Gossypium barbadense*) were planted in pots in a greenhouse on 13 June 2018 to obtain cotton seedlings. An experimental field was transplanted with 132 cotton seedlings (12 plants per row $\times$ 11 rows) in a single plant layout (SPL) where individual plants (treated as one plot) had an in-row and across-row width of 1.52 m (Figure 7.1C). Two batches of transplanting were conducted. The first batch of transplanting was conducted on 26 June 2018 (13 days after planting, DAPs), yielding 75 (out of 89 survived seedlings) healthy plants over the growing season. The second batch of transplanting was conducted on 5 July 2018 (22 DAPs), yielding additional 41 (out of 43 survived seedlings) healthy plants. A total of 116 plants from 23 genotypes, therefore, were used in the present study. The modified GPhenoVision system imaged the field in a continuous scanning mode every 2 days (or 3 days if over weekends) during the flowering period from 20 August 2018 (68 DAPs) to 24 October 2018 (133 DAPs).

7.3.2 DeepFlower for Characterization of Flowering Patterns

Image Preprocessing and Labeling

Collected images were segregated to individual plants based on the collection location information, generating a dataset containing 8666 images collected for 116 plants on 26 dates. A total of 475 images were randomly selected from the dataset and manually annotated with bounding boxes of five classes (see Figure S3). The five classes included the target plant, emerging bloom, opened cotton boll, region with specular reflectance, and others (objects other than the four classes). This labeling strategy was named as the 5-class labeling strategy. The 475 images were randomly shuffled to form training (380 images) and testing (95 images) datasets for training and evaluating object detection models.

Bloom Detection

A deep learning based framework (DeepFlower) was developed to detect and count emerging blooms in the collected images (Figure 7.2). The framework consisted of three major sections including object detection, emerging bloom counting, and flowering characterization.

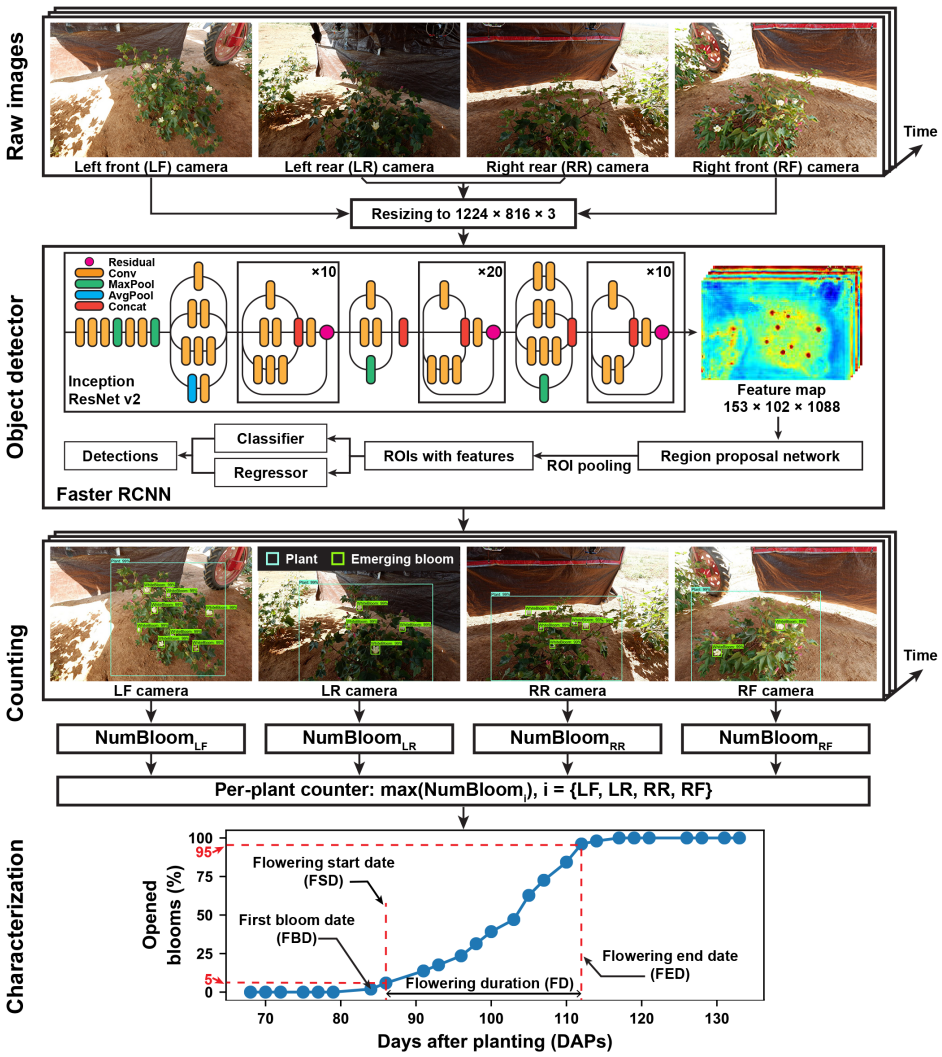

Figure 7.2: DeepFlower processing pipeline for detection, counting, and characterization of flowering patterns using deep learning method and color images.

Object detection was the key of the DeepFlower framework. Because of the success of many object detection applications [30], the Faster RCNN meta-architecture was used as the object detector in the present study (see Object detector in Figure 7.2). The architecture contains a feature extractor, a region proposal network (RPN), and a classification and regressor module. The feature extractor is usually a deep CNN network, which extracts informative feature representations from the raw input images in a hierarchical fashion. The RPN uses the extracted features to generate potential regions of interest (ROIs), and the

classification and regressor module uses the features in each ROI to identify the ROI class and refine the coordinates of ROI bounding box. As images contained diverse object classes with a similar appearance, the Inception ResNet v2 was used as a feature extractor due to its strong capability of learning adequate features to differentiate similar object classes.

As a limited number of labeled images were available, a transfer learning technique was used to facilitate model training. A Faster RCNN model was initialized using weights pretrained on the Common Objects in Context dataset (aka. COCO dataset, a large annotated image dataset open to the public) and fine-tuned on the training dataset for bloom detection. As the Faster RCNN model was trained using images labeled by the 5-class labeling strategy, the model was named as FrRCNN5-cls for conciseness. Model training was performed using a mini-batch stochastic gradient descent (SGD, batch size was 2) by the Adam optimizer with an initial learning rate of 5×10^{-5}, a dropout rate of 0.5 for the RPN and classification and regressor modules, and a weight decay of 1×10^{-3}. Based on preliminary experiments, a total 50,000 training steps (equivalent to 167 epochs) were used to ensure the model convergence for the bloom detection task. Model checkpoints were saved after every 5,000 training iterations, which were used to select the best model for bloom detection. Two computing nodes (14 2.8 GHz CPU cores, 120 GB CPU memory, and Tesla V100 16GB GPU memory) hosted by the Georgia Advanced Computing Resource Center (GACRC) were used for model training under the operating system of CentOS 7.5 with Tensorflow 1.12.0.

The trained Faster RCNN model could detect up to 100 bounding boxes of target plant and emerging blooms with classification confidence scores in a given image. If the confidence score was less than an arbitrary threshold (0.7 in the present study), a detection was removed from the detection result. Consequently, the final detection results contained only detections with a high classification confidence, which were used for bloom counting.

Bloom Counting

A counting strategy was developed to use detection results from the Faster RCNN model to count the number of emerging blooms for a plant on one day (see Counting in Figure 7.2). The strategy counted the number of emerging blooms for a plant in two steps. In the first step, emerging bloom detections were treated as blooms within the target plant if the centroids of their bounding boxes were within the bounding box of the target plant detection. Subsequently, the number of emerging blooms on the target plant was obtained for each of the four images acquired for a plant on one day. This provided an accurate bloom count for a plant from each of the four viewing angles. In the second step, we hypothesized that most (or all) emerging blooms should be seen from one of the four viewing angles, and thus the strategy selected the image (viewing angle) that provided the maximum bloom count from the four images as the number of emerging blooms for a plant on that day. As the first step only considered emerging blooms within a target plant, this counting strategy was summarized as the "plant-based counting" strategy.

Flowering Characterization

The numbers of emerging blooms per plant per day over the flowering period were used to derive flowering curves for individual plants (see Characterization in Figure 7.2). A flowering curve was defined as the cumulative percentage of total opened emerging blooms over the growing time (in DAPs). Three critical points were defined on a flowering curve, including first bloom date (FBD) when the first bloom was identified, flowering start date (FSD) when at least 5% of emerging blooms occurred on a plant, and flowering end date (FED) when at least 95% of emerging blooms occurred on a plant. Three flowering characteristics were derived from the three critical points. FBD and FSD were directly used as flowering characteristics, whereas FSD and FED were used to calculate flowering duration (FD), which was important for many applications related to improvements of environment adaptability.

7.3.3 Ablation Experiments

Labeling Strategy

While image labeling seems straightforward, it could significantly affect the performance of trained deep neural networks or deep meta-models. For the bloom detection task, a simple class definition was used to label images for training, including only three classes i.e., target plant, emerging bloom, and non-bloom. The non-bloom class contained all regions that were labeled other than plant and emerging bloom classes. This labeling strategy has been mostly used by many deep learning applications, which annotated only objects of interest. For brevity, this labeling strategy was named as the 3-class labeling strategy. Accordingly, the same training process was applied to train another Faster RCNN model ($\text{FrRCNN}_{\text{3-cls}}$) using images labeled by the 3-class labeling strategy. This model was compared with the $\text{FrRCNN}_{\text{5-cls}}$ model in terms of detection accuracy.

Counting Strategy

The "plant-based counting" strategy would provide an accurate count of emerging blooms on a target plant in an image, but it required additional efforts on labeling (e.g., annotating target plants in images) and computing (e.g., judgement of emerging bloom location within or outside of a target plant). A simplified counting strategy was to directly use the number of emerging bloom detections as the count for a plant in an image, which might save those labeling and computing efforts. This simplified strategy could be valid, because images were captured for a single plant and might not contain much information of neighboring plants. As this strategy would use all emerging bloom detections in an image, it was named as the "whole image-based countin" strategy. An ablation experiment was conducted to compare the two counting strategies in terms of counting accuracy.

7.3.4 Statistical Analysis

For detection and counting accuracies, simple linear regression analyses were performed between imaging derived and manual counts for the 116 plants on 26 dates. No interception term was used for those analyses. The slope of regression equation, coefficient of determination (R^2), and root mean squared error (RMSE) were used as indicators to evaluate performance. In addition, error analyses were conducted for the optimal counting framework (the combination of the best detection model and counting strategy) for both absolute counting and cumulative percentage calculation.

For flowering characteristics, analysis of variance (ANOVA) analyses were performed on the three flowering characteristics (FBD, FSD, and FD) among three genetic categories and genotypes, respectively, exploring differences in flowering patterns between various cultivated and exotic species. All tests were performed in R using a significance level of 0.05.

An important aspect of the present study is to guide the design of future large-scale experiments. The minimum replication number, therefore, was estimated for each flowering characteristic for experiments that are likely to include at least 200 genotypes from one population in a nested association mapping (NAM) study for cotton. Estimation was performed using the one-way ANOVA model with an effect size calculated using the present study data, a significance level of 0.05, and a statistical power of 0.95 in the G*Power software [242].

7.4 Results

7.4.1 Representative Detection Results

Generally, the Faster RCNN model (FrRCNN5-cls) could accurately detect plants and emerging blooms under different illumination, bloom load, and occlusion conditions

(Figure 7.3). If a proper viewing angle was used, the enclosure mostly provided uniform and bright illumination (e.g., Figure 7.3A). The illumination could be an issue as the enclosure did not fully cover the imaging area. When the solar zenith angle was steep or the camera was configured to face the enclosure entrance, the field of view (FOV) of cameras could include both shaded and strongly illuminated areas. Consequently, collected images could have very dark illumination for the shaded part, making it difficult to identify objects with low reflectance (e.g., the plant in Figure 7.3B). The $\text{FrRCNN}_{\text{5-cls}}$ model learned feature representations to detect the plant, showing its capability to handle object variations because of extreme illumination changes. In addition to illumination, bloom load also varied dramatically during the entire flowering process. Plants would have fewer emerging blooms in early and late stages than the peak flowering time. The $\text{FrRCNN}_{\text{5-cls}}$ model provided accurate detection results for both cases, showing the efficacy of using a single model to process images of plants in different flowering stages. Occlusion was another great challenge for detecting emerging blooms. Cotton plants were branchy and leafy during the flowering period, so blooms were frequently occluded by plant leaves and branches. Depending on the cultivar and development stages, the occlusions varied in direction and severity (see Figure 7.3E and 7.3F), which introduced issues for object detection (especially by using traditional image processing). The $\text{FrRCNN}_{\text{5-cls}}$ model learned effective features to describe and detect occluded emerging blooms, especially some heavily occluded emerging blooms (Figure 7.3F). All of these successful cases demonstrated the capability of the $\text{FrRCNN}_{\text{5-cls}}$ model to detect plants and emerging blooms in images with varied conditions.

Figure 7.3: Representative results of plants and emerging blooms detected by the trained Faster RCNN model. The top three rows demonstrate successful detections under different illumination, bloom load, and occlusion conditions. The bottom row shows two failed cases of emerging bloom detection, one of which because back-sided leaves had higher reflectance and were identified incorrectly as emerging blooms, and the other because a lower contrast between emerging blooms and the background could lead to mis-detection of emerging blooms.

The $FrRCNN_{5\text{-cls}}$ model, however, could not process certain cases. The abaxial surface of leaves has a higher reflectance than the adaxial surface [243], showing a similar contrast pattern with emerging blooms (brighter than adjacent areas). When the abaxial surface of leaves was exposed to the camera and surrounded with bracts, these leaves could not be differentiated easily from true emerging blooms by even human observation (Figure 7.3G), thereby generating false positive detections of emerging blooms. In addition, because of a high reflectance, emerging blooms under strong illumination lost the contrast with background and detailed textures, and thus became considerably more difficult to be identified in the images. In this situation, emerging bloom objects were not accurately detected by the $FrRCNN_{5\text{-cls}}$ model.

7.4.2 Results of Ablation Experiments

Labeling Strategy

Two labeling strategies were used in this study: 3-class and 5-class labeling strategies. The 3-class labeling strategy included the classes of target plant, emerging bloom, and non-bloom objects, whereas the 5-class labeling strategy further split the non-bloom class into three classes, resulting in five classes of target plant, emerging bloom, region with specular reflectance, opened boll, and others.

Overall, the model ($FrRCNN_{5\text{-cls}}$) trained using the 5-class labeling strategy had an improved performance (F1 score) than that ($FrRCNN_{3\text{-cls}}$) trained using the 3-class labeling strategy (Figure 7.4). In particular, the F1 score of emerging bloom detection increased by 5% by using the 5-class labeling strategy. Both the precision and recall contributed to improvement of the F1 score, meaning that the $FrRCNN_{5\text{-cls}}$ simultaneously reduced false positive detection and increased true positive detection (Figure S1 and Figure S2). Compared with 3-class labeling, the 5-class labeling strategy could split more efficiently classes with similar appearance. Consequently, the variation within a class would become smaller than

the differences between classes, providing benefits for training deep neural networks. For instance, there could be several types of non-bloom objects that had a distinct appearance. There were also some bright gaps between plant branches and leaves, which formed regions that had a similar appearance to emerging blooms, whereas there were some other objects (e.g., camera) that looked dissimilar from emerging blooms (see "specular reflectance" and "others" objects in Figure S3). If these regions/objects were labeled with different classes, it would be relatively easier for deep neural networks to learn features to form classification boundaries for separating classes with similar appearance. Otherwise, deep neural networks might not learn effective features, resulting in misclassification between regions/object with a similar appearance.

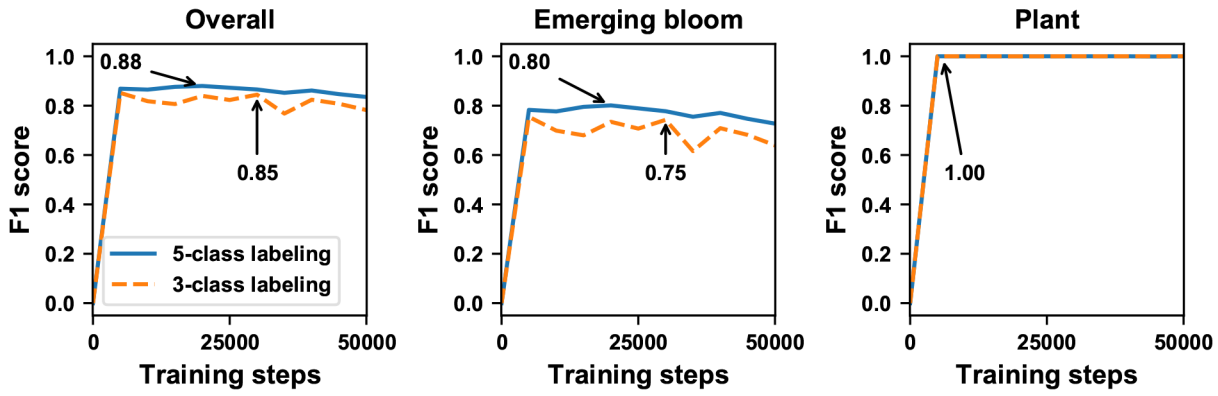

Figure 7.4: Detection accuracies (F1 score) on the validation dataset by using models trained with datasets labeled by the 5-class ($\text{FrRCNN}_{\text{5-cls}}$) and 3-class ($\text{FrRCNN}_{\text{3-cls}}$) methods, respectively. The overall accuracy was calculated using the weighted average of accuracies for emerging bloom and plant classes.

Counting Strategy

Overall, for each image, the "plant-based counting" strategy provided improved accuracy over the "whole image-based counting" strategy (Figure 7.5). Comparing imaging-derived and manual counts, although the regression slope calculated using the "plant-based counting"

strategy was slightly higher than that calculated using the "whole image-based counting" strategy, a higher correlation and lower root mean squared errors (RMSE) were achieved by using the "plant-based counting" strategy, indicating the effectiveness of the "plant-based counting" strategy for improving counting accuracy (Figure 7.5A and 7.5D). These improvements were primarily because the "plant-based counting" strategy made more samples in the counting error range within ± 1, especially a 3% increase with no counting difference (Figure 7.5B and 7.5D). As an absolute counting error of one bloom might be substantial when plants had very few emerging blooms (e.g., early and late flowering stages), relative counting errors were calculated for samples with counting errors less than one bloom (Figure 7.5C and 7.5F). Compared with the "whole image-based counting" strategy, the "plant-based counting" strategy increased the number of samples with no relative counting error by 5% and dramatically reduced the number of samples with relative counting errors over 20%. It is also noteworthy that the "plant-based counting" strategy dramatically improved the counting accuracy for samples that had a zero count with the manual method but a non-zero count with the imaging method (denoted by asterisks in Figure 7.5C and 7.5F).

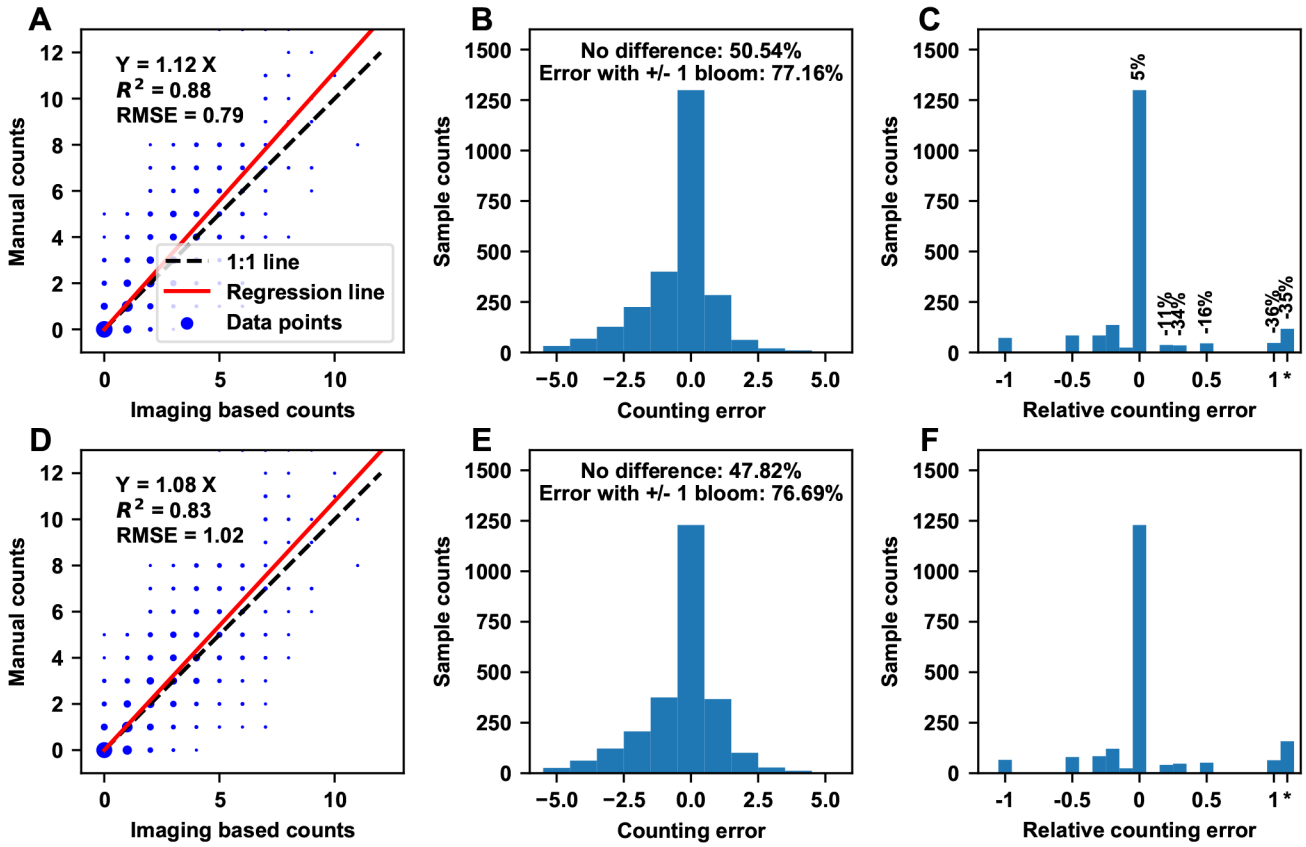

Figure 7.5: Counting accuracies calculated using the "plant-based counting" (top row) and "whole image-based counting" (bottom row) strategies, respectively, for individual plants on each of the 26 scanning dates (a total of 3016 data points). A and D are linear regression results between the imaging derived and manual counts. B and E are the histogram of counting errors. C and F are the histogram of relative counting errors for samples with an absolute counting error of less than 1. In C and F, the numbers on top of the bars indicates the relative improvement (over 5%) of using the "plant-based counting" strategy over the "whole image-based counting" strategy. The asterisk denotes samples that had a zero count with the manual method but a non-zero count with the imaging method.

Although the $\text{FrRCNN}_{\text{5-cls}}$ and the "plant-based counting" strategy demonstrated improved performance on emerging bloom detection and counting, respectively, significant counting errors were identified by jointly using the $\text{FrRCNN}_{\text{5-cls}}$ and the "plant-based counting" strategy as a counting framework (Figure 7.6). For absolute counting, the combi-

nation of the $\text{FrRCNN}_{\text{5-cls}}$ model and the "plant-based counting" strategy provided accurate measurements (less than one bloom) for plants with zero to four emerging blooms per day (approximately 79% of cases). Absolute counting errors substantially increased, however, when plants had five or more emerging blooms (approximately 21% of cases). On average, the developed counting framework also reached a plateau of 6 blooms per plant per day. Thus, when plants reached peak flowering time (over 10 emerging blooms per day), absolute counting errors were over 4 blooms per plant per day, which was equivalent to about 50% relative counting errors. This occurred primarily because of the assumption in the developed counting framework that a single image from a particular viewing angle would capture most (or even all) emerging blooms on a plant on one day, and thus the counting framework could obtain the maximum bloom count from one out of four images for a single plant. This assumption generally held true in flowering stages when plants had a small number of emerging blooms per day, so the counting framework provided accurate counts for most plants. This assumption, however, was invalid during the peak flowering time when plants had a large number of emerging blooms per day. Furthermore, emerging blooms were distributed around plant canopies, so a single image from any viewing angle would not be sufficient to capture all blooms on a plant, resulting in a significant underestimation of absolute bloom counts.

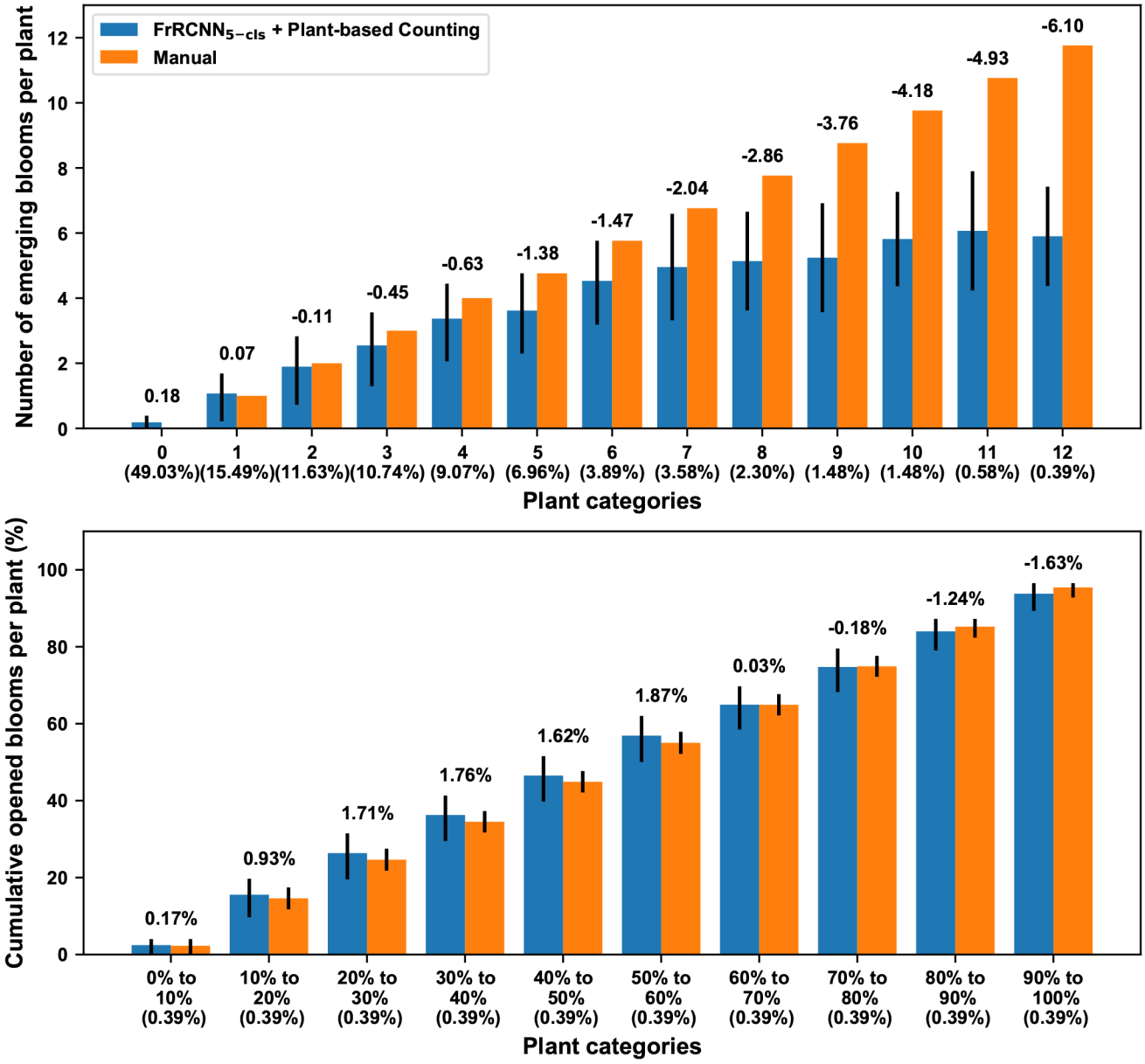

Figure 7.6: Errors of absolute counting (top chart) and cumulative percentage (bottom chart) for emerging blooms per plant by using the $\text{FrRCNN}_{5\text{-cls}}$ and the "plant-based counting" strategy. For absolute counting, plants were grouped into 13 categories based on the number of emerging blooms (0 to 12) on those plants on a particular date. For cumulative percentage, plants were grouped into 10 categories (from 0–10% to 90%–100%) of opened blooms on those plants on a particular date. The number on top of each grouped bar indicated the difference between counts (or cumulative percentage) calculated using the imaging and manual methods.

The significant underestimation of absolute bloom counts, however, showed a limited influence on the accuracy of calculating cumulative percentage of emerging blooms. Errors in the cumulative percentage of emerging blooms were less than 2% irrespective of flowering stages. A possible reason was that cumulative percentage was the ratio of total opened emerging blooms from the beginning of flowering to a specific day and total opened emerging blooms over the growing season. The underestimation of absolute counts would be included with both the numerator and denominator of the cumulative percentage, and thus the effect could be cancelled out to certain extent. Nonetheless, the high accuracy of the calculated cumulative percentage of emerging blooms could hold the great potential for flowering characterization.

7.4.3 Results of Flowering Characteristics and Statistical Analyses

Representative Flowering Curves

As the developed counting framework underestimated the number of emerging blooms on plants during the peak flowering time, the absolute bloom counting curves generated using imaging-derived counts also showed large differences from those generated using manual counts during that time (Figure S4). This suggests that the curves should not be used for quantitative analyses such as the maximal number of emerging blooms per plant over a growing season. The flowering curves derived by the imaging method, however, generally showed a similar trend as the curves derived by the manual method, suggesting that the flowering curves derived by the imaging method could be used for certain qualitative analyses. For instance, the field received precipitation (approximately 8 mm of rain) and experienced chilling temperatures (approximately 10°C cooler than the monthly-average temperature) on 16 September 2018 (95 DAPs) and 24 September 2018 (103 DAPs), respectively. After the weather changed, the plants mostly had a reduced number of emerging blooms on the next sampling day in flowering curves derived by both the manual and imaging methods. Certain genotypes (e.g. Exotic T0368BC3MDN GH196 and Elite DES 56), however, did not

show such a pattern, which means that the genotypes could be potentially more resistant to severe weather changes than other genotypes (Figure S5).

Cumulative flowering curves generated using bloom counts derived by the imaging method were very similar to those generated using manual counts (Figure 7.7). The same correspondence was also observed for individual genotypes (Figure S6). This suggests that the curves derived using the imaging method could potentially be used for both qualitative and quantitative characterization of flowering patterns. Two distinctive patterns were observed from the curves. First, exotic *G. hirsutum* presented a larger within-group variation than elite *G. hirsutum* and *G. barbadense*. This was because the exotic group contained wild genotypes that are diverse in their flowering patterns, whereas elite *G. hirsutum* were cultivated for fiber production and thus selected for flowering patterns that conferred optimal yield. There was only one cultivar in the *G. barbadense* group, which should not present large variation. Second, both exotic and elite *G. hirsutum* showed a relatively steeper slope than *G. barbadense*, indicating a potential difference in flowering duration between various species.

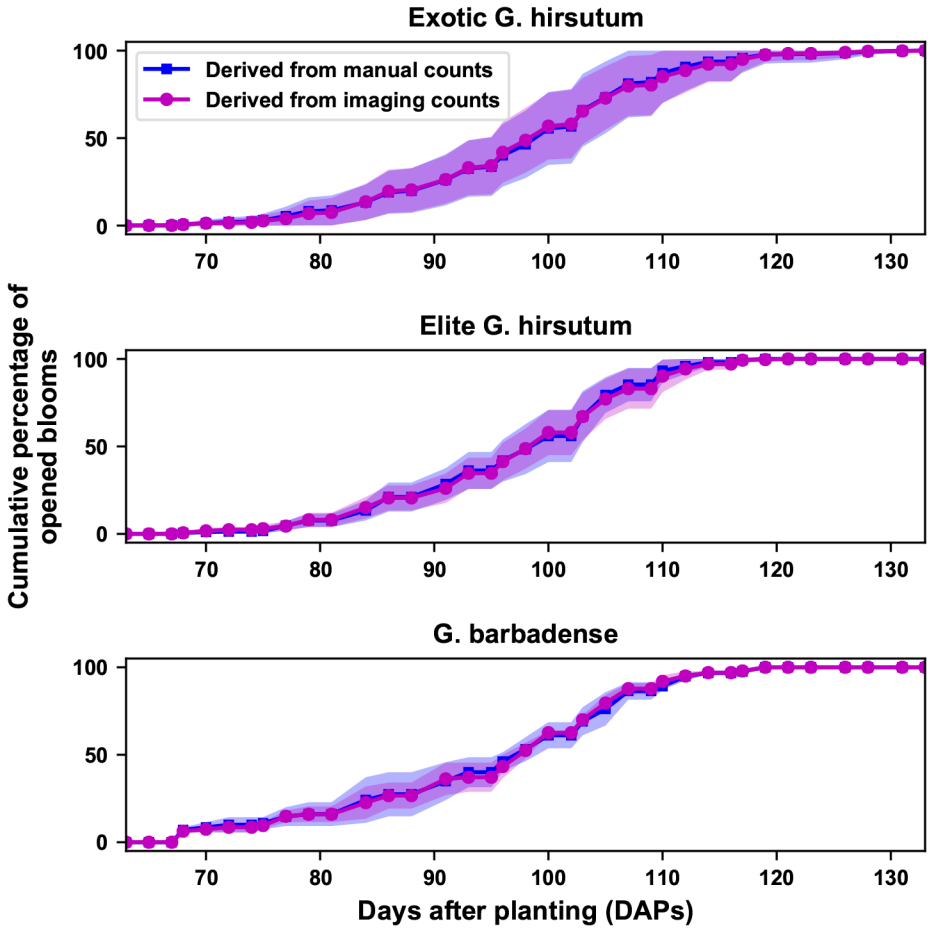

Figure 7.7: Cumulative flowering curves derived using imaging and manual counts for three genetic categories (elite *G. hirsutum*, exotic *G. hirsutum*, and *G. barbadense*) in both the first and second transplanting batches. Group mean values are drawn in lines (solid and dashed lines for results derived by the imaging and manual methods, respectively), and group standard deviations are indicated by shaded areas (magenta and blue for results derived by the imaging and manual methods, respectively).

Statistical Analysis Results

Significant interaction effects were presented on extracted flowering characteristics (first bloom date, flowering start date, and flowering duration) between the genotype and transplanting date, suggesting the necessity of analyzing flowering patterns for each transplanting batch separately (see Table S1 to Table S6). As only the much larger first transplanting batch

showed statistical significance among genetic categories or genotypes, successive sections focused on data of the first transplanting batch.

Flowering characteristics calculated using the flowering curves derived by the imaging method showed the same statistical power in differentiating the three genetic categories as those calculated using the flowering curves derived by the manual method (Figure 7.8). For the first bloom date and flowering start date, although *G. barbadense* showed the lowest values with the least standard deviation, it could not be statistically separated from the *G. hirsutum* groups for two reasons. First, exotic *G. hirsutum* contained diverse genotypes, presenting a large variation that covered the other two groups. Second, *G. barbadense* had only two replicates in the first transplanting batch, which had limited statistical power to be differentiated from other groups. For flowering duration, however, *G. barbadense* was significantly longer than the *G. hirsutum* groups, which was an expected flowering pattern for *G. barbadense* (Pima cotton) in the study area.

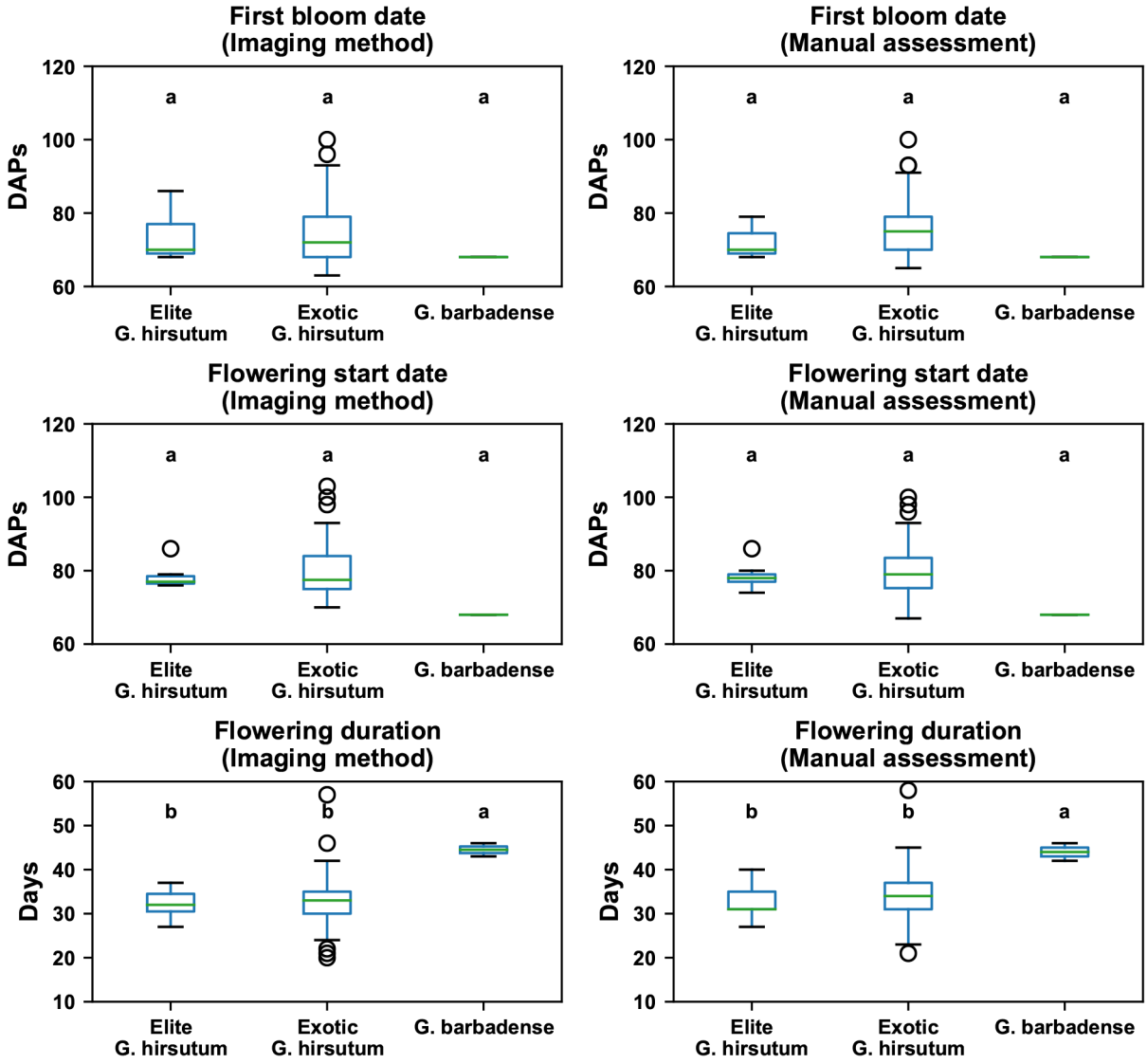

Figure 7.8: Boxplot of flowering characteristics (first bloom date, flowering start date, and flowering duration) among three genetic categories (elite *G. hirsutum*, exotic *G. hirsutum*, and *G. barbadense*) in the first transplanting batch. Groups with a statistically significant difference ($p < 0.05$) are denoted with different letters, and group mean values of each characteristic are sorted alphabetically.

While the order of individual genotypes was slightly different, flowering characteristics derived by the imaging and manual methods showed very similar statistical patterns among genotypes (Figure 7.9). Genotype T0368BC3MDN.GH196 had the first bloom (first bloom

date) and entered into the flowering period (flowering start date) significantly later than other genotypes, suggesting that T0368BC3MDN.GH196 could be used for studying genes and gene regions controlling flowering time. In addition, genotype T0368BC3MDN.GH196 had a significantly shorter flowering duration than other genotypes. This occurred likely because of environmental effects. Overall, air temperature decreased dramatically (more than 15°C) after 120 DAPs, leading to a sudden drop of emerging blooms. Although several blooms opened after 120 DAPs (see Figure S5), the freezing temperature might cause an early termination of flowering for T0368BC3MDN.GH196. Some other genotypes also presented significant differences in flowering duration, such as genotype T1046cBC1.GH212 for a longer duration and genotypes T0281aMDN.GH198 and T1046aBC1.GH210 for a shorter duration. It should be noted that genotype Pima.S6.2011.3841 had a statistically longer flowering duration using the characteristic derived from manual counts but not by that derived from imaging counts, which was the only difference in the statistical patterns between the two methods. This possibly occurred because manual counts would not miss any emerging blooms on a plant, having a relatively stronger capability to identify differences between genotypes with fewer replicates.

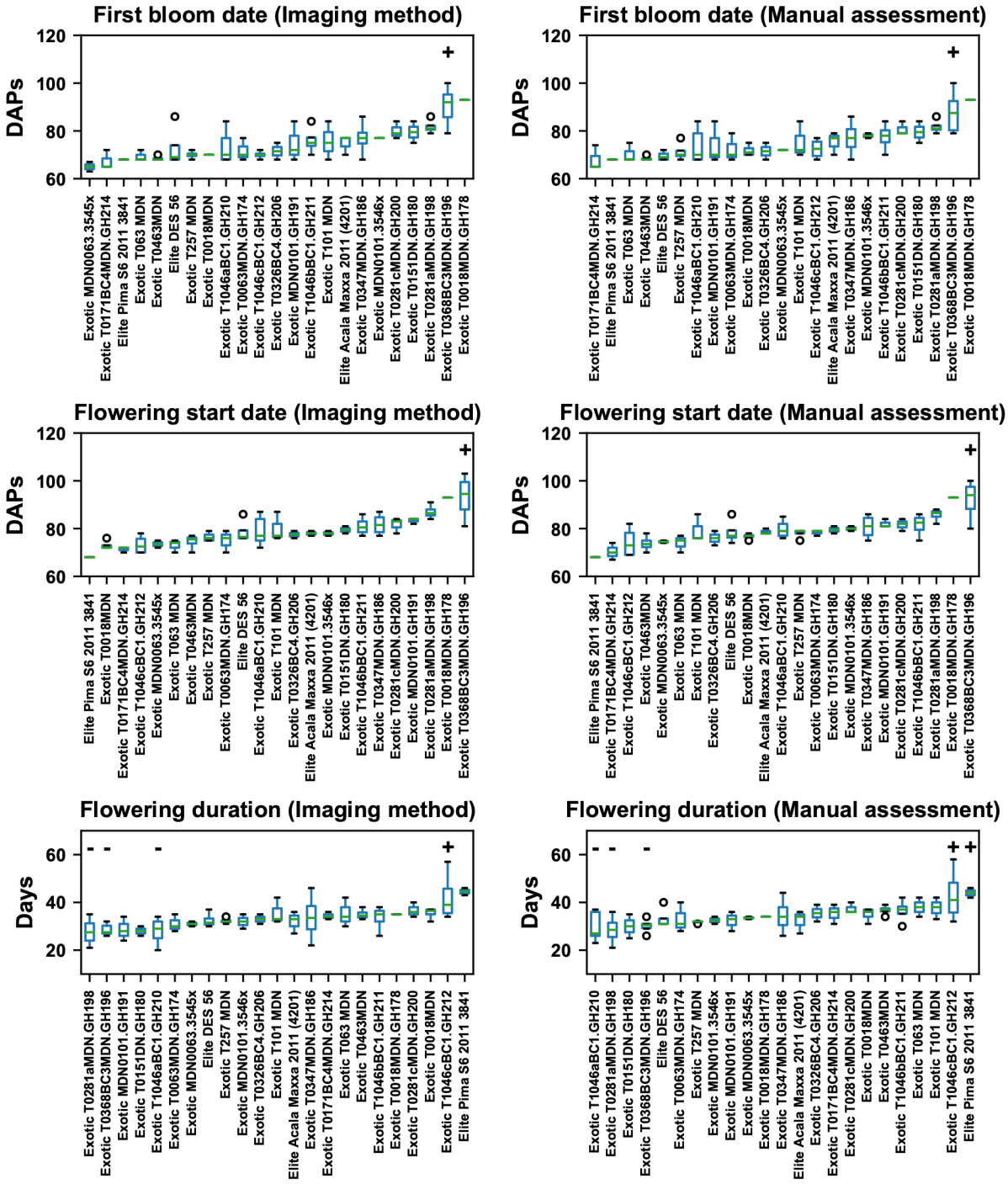

Figure 7.9: Boxplot of flowering characteristics (first bloom date, flowering start date, and flowering duration) among 23 genotypes in the first transplanting batch. Genotypes with statistically higher values are denoted by "+", whereas genotypes with statistically lower values are denoted by "-". Differences were inferred at the significance level of 0.05.

Based on the estimation, each genotype should have at least 2 replicates to ensure adequate statistical power to identify the significance of the first bloom date and flowering start date, or at least 3 replicates to ensure the statistical power to identify the significance of flowering duration among the 23 genotypes (Table 7.1). As there were only 2 replicates per genotype Pima.S6.2011.3841, no significant difference in flowering duration was identified between Pima.S6.2011.3841 and other genotypes, which agreed with the experimental result. If the variation because of genotype remains the same as that in the first transplanting batch, using more genotypes (e.g., 200 genotypes in a population) would slightly increase the statistical power for identifying significance among genotypes. Flowering curves derived by the imaging method, therefore, would remain effective for flowering pattern analyses in large-scale experiments.

Table 7.1: Estimated number of replications per genotype at the significance level of 0.05 and power of 0.95. FBD shorts for first bloom date, FSD shorts for flowering start date, and FD shorts for flowering duration. The asterisk denoted the estimation for one population in a NAM study.

Trait	Batch	Effect size F	Number of genotypes	Estimated number of replications
FBD	1	1.18	22	2
FBD*	1	1.18	200	2
FSD	1	1.52	22	2
FSD*	1	1.52	200	2
FD	1	0.88	22	3
FD*	1	0.88	200	2

7.5 Discussion

The DeepFlower framework demonstrated efficacy of detecting and counting emerging blooms in images to characterize flowering patterns for different genetic categories or geno-

types. Flower characteristics derived by the imaging method showed an almost identical capability for identifying significance among genotypes with manual counts, which further validated the effectiveness of the DeepFlower framework for studying flowering patterns. In particular, the DeepFlower framework successfully revealed flowering patterns for cotton plants that have a complex canopy architecture (and thus difficulties of emerging bloom detection and counting) and therefore should transfer well to other flowering plants that have the same or similar canopy architecture. This suggests that the combination of the image acquisition system and DeepFlower framework can be an effective and efficient tool for characterization of flowering patterns for plants in the field, holding great potential for identifying gene loci that control flowering behavior for different plant genotypes.

Although the DeepFlower framework showed promising results for extracting flowering characteristics, at least two aspects can be further improved. First, the scanning throughput is relatively low for the current configuration. The platform ran at approximately 0.25 m/s and took around 25 minutes to complete the scanning of the present experimental field (approximately 0.05 ha), resulting in a scanning throughput of 0.12 ha/h. This throughput might not be adequate for very large-scale experiments, e.g., that involve up to several thousand genotypes with at least two replicates per genotype (up to a couple of hectares). Challenges, however, would need to be identified to balance the platform cost (camera with high resolution and fast frame rate), image quality (blurry), and scanning throughput (platform moving speed). Second, the present DeepFlower framework oversimplifies the counting task by using only one single image with the maximum count among the four viewing images. The framework depends upon the assumption that most or all emerging blooms can be seen from a single one of these four viewing angles. Experimental results, however, showed that this assumption is invalid when plants enter into peak flowering time, leading to a significant underestimation of bloom counts. Consequently, absolute bloom counting curves cannot be used for quantitative analysis of flowering patterns. A viable solution is to integrate 3D

imaging so that 2D detections can be projected onto a global 3D space for counting. For instance, photogrammetric methods (e.g., structure from motion) can be used to reconstruct 3D point clouds using images from multiple viewing angles, so that for a single plant emerging bloom detections can be projected from different 2D images onto a global 3D space to remove duplicated detections (and thus counts). In the present study, preliminary tests using the collected images suggested that images from four viewing angles (approximately 90° apart from each neighboring angle) could not provide adequate image overlap to reconstruct 3D point clouds of a single plant, and thus the 2D to 3D projection. It is therefore necessary to conduct successive studies to explore the optimal image collection configuration (e.g., viewing angles and number of images) for 3D reconstruction using photogrammetric methods. Another way is to fuse 2D images with 3D point clouds acquired using separate instruments (e.g., LiDARs), which enables the 2D to 3D projection. This will also require considerable efforts to develop new sensing systems for data collection and algorithms for data fusion (especially multi-source heterogeneous data fusion).

7.6 Conclusions

The developed imaging approach (combination of the image acquisition system and DeepFlower framework) can be an efficient and effective tool for detecting and counting blooms on plants in the field, demonstrating promising results for characterization of flowering patterns. In particular, the developed approach can potentially be used for many other flowering plants that have a simpler or similar canopy architecture, providing potential for deepening understanding of the flowering process in general. Future studies will be focused on the integration of 3D imaging to further improve the counting accuracy and expand the capability of mapping bloom positions on plants.

Chapter 8

Nondestructive Detection and Quantification of Blueberry Bruising using Near-Infrared (NIR) Hyperspectral Reflectance Imaging[1]

[1]Jiang, Y., Li, C., and Takeda, F. 2016. *Scientific Reports*, 6, 35679. Reprinted here with permission of publisher.

8.1 Abstract

Currently, blueberry bruising is evaluated by either human visual/tactile inspection or firmness measurement instruments. These methods are destructive, time-consuming, and subjective. The goal of this paper was to develop a non-destructive approach for blueberry bruising detection and quantification. Experiments were conducted on 300 samples of southern highbush blueberry (Camellia, Rebel, and Star) and on 1500 samples of northern highbush blueberry (Bluecrop, Jersey, and Liberty) for hyperspectral imaging analysis, firmness measurement, and human evaluation. An algorithm was developed to automatically calculate a bruise ratio index (ratio of bruised to whole fruit area) for bruise quantification. The spectra of bruised and healthy tissues were statistically separated and the separation was independent of cultivars. Support vector machine (SVM) classification of the spectra from the regions of interest (ROIs) achieved over 94%, 92%, and 96% accuracy on the training set, independent testing set, and combined set, respectively. The statistical results showed that the bruise ratio index was equivalent to the measured firmness but better than the predicted firmness in regard to effectiveness of bruise quantification, and the bruise ratio index had a strong correlation with human assessment (R^2=0.78-0.83). Therefore, the proposed approach and the bruise ratio index are effective to non-destructively detect and quantify blueberry bruising.

8.2 Introduction

The United States (U.S.) produced 239,000 tonnes of blueberries in 2013, accounting for 57% of total world production [244]. More than half of U.S. production went to the fresh fruit market and created over 579 million dollars in revenue [245]. Much of the blueberry crop destined for fresh market is still hand-harvested. Mechanical harvesting has a major limitation in that it creates more bruises which decrease fruit quality and ultimately reduce the monetary value of the blueberry crop [246]. In fact, fruit bruising causes around 10% of total economic losses of the blueberry industry every year. Additionally, bruising may

accelerate other biological processes such as spoilage [247]. Even worse, rotten or fermented fruit could affect other healthy fruit, resulting in significant losses during long-distance transportation [248].

Bruising is a type of subcutaneous tissue damage without rupturing fruit skin [249]. Typically, the tissue damage causes the mixture of phenolic compounds and polyphenol oxidase, generating dark coloration. Bruises are indicated by the discoloration of damaged tissues and thus they can be observed and differentiated from healthy tissues [249]. Consequently, visual inspection is an intuitive way to detect and assess fruit bruising. Bruises are not externally observable for most fruits, especially fruits with dark coloration such as blueberries. Therefore, visual inspection requires slicing fruit samples. Currently, each blueberry sample is sliced along the equatorial axis and the slices are imaged by color cameras. Human graders evaluate the color images and calculate the bruising level, the ratio of the area of the discolored region to the total surface area of the slice. Typically, a blueberry with a bruising level higher than 0.2 (20%) is considered bruised, and the percent of bruised fruit in each treatment is the bruise severity for that group. A limitation of the method is that it is impossible to detect bruises that are not present on the sliced cross-section. Moreover, illumination conditions could significantly affect inspection results.

In addition to discoloration, bruised tissues are typically softer than healthy tissues [249]. A palpation method is often used in which a blueberry is held between one's thumb and index fingers and the fruit is gently squeezed and rolled. However, this method is subjective and labor- and time-consuming. Therefore, a number of automated approaches have been developed to measure fruit firmness that can be used as an indirect index for bruise quantification and assessment. Those approaches include firmness/texture analysis, acoustic impulse-response measurement [250], and resonance frequency-based method [251]. Although these methods provide objective and repeatable measurements of fruit firmness, they are

destructive to fruit samples. To overcome this issue, a laser air-puff detector was invented to measure fruit firmness in a non-contact manner [252], and was applied to blueberry firmness measurement [253]. A fruit sample is deformed by a puff of air and a laser displacement sensor is used to record the deformation to assess the fruit firmness. Although the technique does not directly contact samples, it could cause potential damage after repeated measurements because fruit deformation in each measurement could be accumulated and ultimately result in fruit damage.

In recent years, advanced imaging modalities have been explored as non-destructive approaches for bruise detection. X-ray imaging was applied to detect bruises using the difference in radiation attenuation coefficients between bruised and healthy tissues. Typically, bruised tissues have lower density and absorb less radiation than healthy tissues. The results showed that X-ray imaging could achieve over 90% accuracy in detecting bruises in apples [254]. Nonetheless, X-ray imaging techniques have potential safety issues and X-ray instruments are expensive for agricultural applications. As a non-ionizing method, magnetic resonance imaging (MRI) was also explored due to the difference in free water (water released by damaged cells) between bruised and healthy avocado tissues [255]. In early stages, more free water was released in the bruised areas, resulting in higher intensity in MRI images. The results showed that MRI detected bruises immediately after impact, but it was not practical for the food industry or large-scale food research because of the high cost of the instruments. In addition, a dynamic thermal imaging (TI) method, pulsed-phase thermography (PPT), has been studied because of the difference in thermal diffusivity between bruised and healthy fruit [256]. When heating or cooling fruit samples, the diffusivity difference caused different rates of temperature change which were used to differentiate bruised and healthy fruit. PPT is less costly than X-ray imaging and MRI without safety concerns. However, the relatively low signal-to-noise ratio (SNR) and topological effects could cause potential issues for quantitative data analysis such as bruise quantification. To date, none of the three

imaging modalities has been used for blueberry bruise detection.

Hyperspectral Imaging (HSI) has been investigated for fruit bruise detection because both physical and chemical property changes affect the spectral profile [257]. Previous studies have demonstrated the feasibility of using HSI reflectance mode to detect bruises in various fruits. In particular, the spectral range from 900 to 1700 nm was appropriate to detect bruises in apples [256, 258, 259], pears [260], and jujube [261]. In this spectral range, all the studies identified bruised tissues with over 88% accuracy, as the spectra of bruised tissues were significantly different from those of healthy tissues. The spectral difference between the bruised and healthy tissues was probably due to tissue disruption including cell wall failure and the release of free water from cells. The free water absorbed more light and affected the reflectance spectra of bruised tissues in the range from 900 to 1700 nm, especially at several key wavelengths with high water absorption such as 970 nm, 1200 nm, and 1470 nm [262]. Other than directly using them to identify bruised tissues, the spectra of the fruit could be correlated to fruit firmness by regression models such as partial least squares regression (PLSR) [263]. Two consecutive studies were conducted to predict blueberry firmness using reflectance and transmittance modes ranging from 400 to 1000 nm [264, 265]. Results showed that the predicted firmness could be used for qualitative analysis of blueberry bruising, but the root mean squared error of predictions (RMSEPs) were 20% of the mean value of the fruit firmness, which might be insufficient for quantitative assessment of blueberry bruising.

In addition to its efficacy for detecting fruit bruising, HSI is also suitable for the food industry because it is safer than X-ray imaging, more affordable than MRI, and better in image quality than thermal imaging. In particular, the HSI reflectance mode could be applied to fruit packing lines for online sorting after modifications [257]. There was no HSI study done for blueberry bruising so far, and most HSI studies primarily explored spectral differences and did not take full advantage of the technique. Therefore, the present study

utilized both the spatial and spectral information for blueberry bruising quantification by combining image processing algorithms and chemometrics.

The overall goal of this study was to develop a non-destructive approach using near-infrared (NIR) hyperspectral imaging (950 to 1650 nm) to detect and quantify blueberry bruising. Specific objectives were to (i) classify healthy and bruised tissues using the spectra extracted from the regions of interest (ROIs) of the SHB (southern highbush blueberry) and NHB (northern highbush blueberry) fruit, (ii) quantify blueberry bruising using the bruise ratio index, and (iii) compare the effectiveness of bruise quantification using the bruise ratio index, measured and predicted firmness, and human assessment.

8.3 Materials and Methods

8.3.1 Sample Collection and Preparation

Two experiments were conducted in this research (Figure 8.1). The first experiment was conducted to study the spectral differences between healthy and bruised tissues in southern highbush cultivars, and bruises were manually created at controlled positions (see Supplementary Table S13). A total of 300 blueberry samples of three southern cultivars including Camellia, Rebel, and Star were collected in May 2015 from a commercial farm in Alma, Georgia, USA. Each cultivar contained 100 samples that were divided into four groups: a control treatment of 10 samples and three bruise treatments with 30 samples each. The control treatment was kept intact, whereas the three bruise treatments were dropped onto a steel surface from heights of 15, 23, and 31 cm, respectively. In order to control the position of fruit bruising, bruises were created by a specially-designed pendulum. The wooden arm of the pendulum was connected with a sample holder made of silicon rubber through a screw eye. For the screw eye, the thread end was installed into the wooden arm, and the loop end was adhered to the back face of the sample holder. The steel screw eye enables the sample holder to be attached and released by a switchable magnet, facilitating consistent

sample dropping. For each blueberry sample, one side was stuck on the sample holder using petroleum jelly when the sample holder was attached by the magnet, and the other side would hit the contacting surface when the sample holder was released. Therefore bruises mainly occurred on the impacted face. After bruise creation, all the samples were divided equally into two groups and stored in an air-conditioned room (23°C with 30–35% relative humidity) for 24 and 48 hours, respectively, before they were imaged.

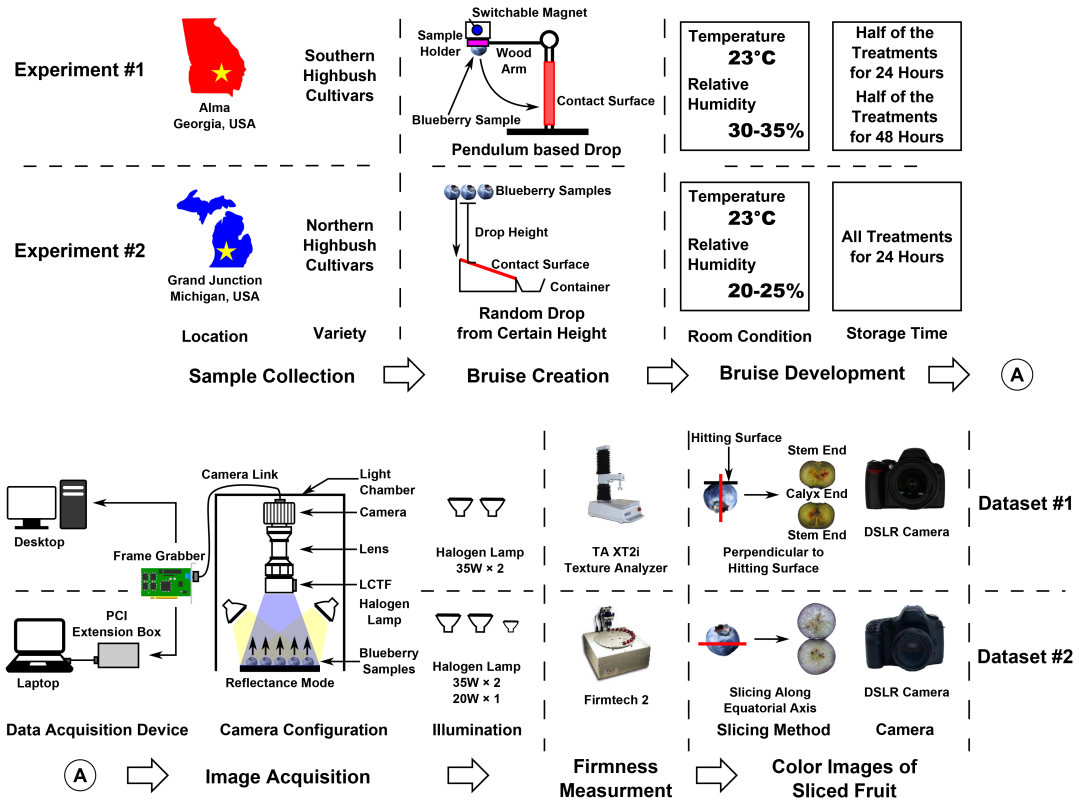

Figure 8.1: Overall flowchart of the two experiments conducted in this research.

The second experiment was conducted to study the spectral differences between healthy and bruised tissues for northern highbush cultivars, and the bruises were created by randomly dropping the fruit from certain heights (see Supplementary Table S13). A total of 1500 blueberry samples of three northern highbush cultivars including Bluecrop, Jersey, and Liberty were collected in August 2015 from a commercial farm in Grand Junction, Michigan,

USA. Each cultivar contained 500 samples that were divided equally into five treatments, with each treatment sub-divided into four replicate groups of 25 samples each. The five treatments were: control; fully-bruised (dropped from 90 cm onto a steel surface 8 times); treatments dropped onto a steel surface from 60 and 120 cm; and onto a padded surface from 120 cm. The purpose of the fully-bruised treatment was to easily and accurately extract the spectra of bruised tissues because random dropping was not able to control the bruise position and could cause difficulties in spectral extraction and classifier training. After bruise creation, all samples were stored in an air-conditioned room (23°C with 20-25% relative humidity) for 24 hours, and were then used for further image acquisition and processing.

8.3.2 Hyperspectral Image Acquisition

The samples of the first experiment were imaged using a hyperspectral imaging system previously built by the Bio-Sensing and Instrumentation Laboratory of the University of Georgia [266]. All images were acquired in a light chamber to avoid the interference of ambient light (Figure 8.1), and two 12V 35W halogen lamps (S4121, Satco Products Inc., NY, USA) were used as the illumination source. Prior to collecting images, an image of a 99% reflective panel (SRT-99-050, Labsphere Inc., North Sutton, NH, USA) was obtained as the white reference, and an image taken with the optical lens being covered was obtained as the dark reference. As the position of bruises was controlled, the samples were placed on a black cardboard holder with the bruised surfaces facing toward the camera. Control samples were imaged from three angles (stem, calyx, and equatorial axis). A total of 24 hyperspectral images were collected including 12 images for 150 samples stored for 24 hours and 12 images for 150 samples stored for 48 hours (see Supplementary Figure S7 and Table S14 for detailed image layout).

The samples of the second experiment were imaged using a hyperspectral imaging system based on a portable light chamber and data acquisition device (Figure 8.1). A frame grabber

(PCI-1426, National Instruments Corp., TX, USA) was installed in a PCIe expansion box (NA 211A-NB, Netstor Technolog, Taiwan, China) through a PCIe to PCI adapter (ST369, Sintech Electronic, Shenzhen, China). The frame grabber was used to connect the camera of the hyperspectral imaging unit and a laptop. In addition to the two 12V 35W halogen lamps, a 12V 20W halogen lamp (PC 81763, GE Lighting, OH, USA) was used to enhance the uniformity of the illumination. White and dark reference images were collected using the same procedure as used in the first experiment. Since the position of bruising was not controlled in the second experiment, it was necessary to image both the stem and calyx sides. A total of 120 (5 treatments $\times$ 4 treatment replicates $\times$ 3 cultivars $\times$ 2 sampling positions) hyperspectral images were acquired, and each image contained 25 samples from a treatment replicate. All acquired hyperspectral images covered 141 wavelengths ranging from 950 to 1650 nm with a spectral interval of 5 nm.

8.3.3 Reference Measurements

Firmness

After image acquisition, the firmness of the blueberry samples was measured for reference. In the first experiment, a texture analyzer (TA.XT2i Texture Analyzer, Texture Technologies Corp., NY, USA) was used to measure firmness by following the procedure proposed by [265]. Each blueberry fruit was compressed between two parallel plates at a constant velocity of 1 mm/s for a total deformation of 3 mm. The firmness was calculated from the slope (N/mm) of the force/deformation curve between 0.5 mm and 2.5 mm displacement, as the curve was relatively straight in this range. In the second experiment, a FirmTech machine (FirmTech 2, BioWorks, Inc., KS, USA) was used to directly measure firmness (N/mm) of individual berry samples.

Color Image of Sliced Fruit and Human Evaluation

After firmness measurement, the blueberry samples were sliced and imaged by a color camera for reference and human evaluation. In the first experiment, the samples were sliced perpendicular to the bruise position. For instance, if a bruise was created on the stem side of a fruit, the fruit was to be sliced along the stem-to-calyx axis, enabling an evaluator to observe the bruise on the slicing plane. The sliced samples were imaged by a digital single lens reflex (DSLR) camera (D40, Nikon Corp., Japan) under ambient illumination. As the slicing method was different from the approach used in a previous study [135], these color images of the sliced samples were only used for reference. However, in the second experiment, the samples were sliced by following the protocol described by Brown [135]. The sliced samples were imaged by another DSLR camera (5D Mark II, Canon Inc., Japan) under ambient illumination. These color images were used for both reference and human evaluation. Trained human graders calculated the number of pixels of the discolored (bruised) area and the total cross-sectional area of each sample, and the ratio between the two was used as the human assessment of the bruising level of a sample [135]. In total, the two experiments produced two datasets (Dataset1 for experiment #1 and Dataset2 for experiment #2), and each dataset contained hyperspectral images, firmness measurements, and color images of sliced fruit (human assessment of bruising level only for Dataset2).

8.3.4 Hyperspectral Image Processing

Automatic Blueberry Segmentation

The hyperspectral images were preprocessed by flat field correction to remove artifacts caused by non-uniform illumination or variations in the pixel-to-pixel sensitivity of the detector (Figure 8.2). The process was executed in a customized program developed in IDL (IDL 7.1, Exelis Inc., VA, USA). Subsequently, the grayscale images at 1075 nm were used to create masks for segmenting individual samples from the background. The masks were automatically generated by thresholding. The grayscale images were thresholded at intensity

levels of 737 and 942 (out of 4095) for Dataset1 and Dataset2, respectively. The resulting images were enhanced by morphological operations to fill holes and remove noise, and the refined results were used as masks. In addition, individual samples were able to be recognized by calculating connected components in the masks. The operations were implemented in MATLAB (MATLAB 2015b, The MathWorks Inc., MA, USA).

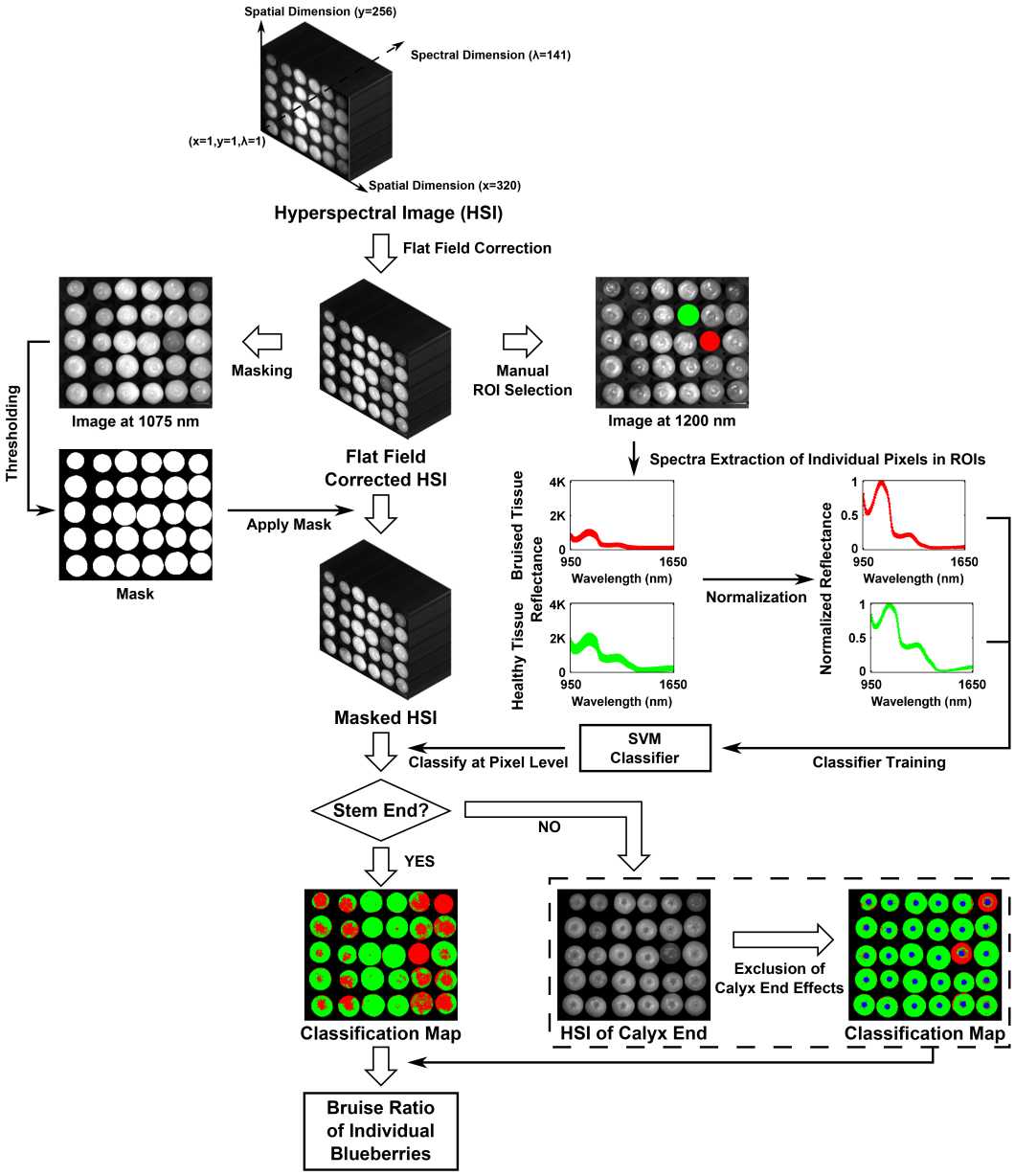

Figure 8.2: Flowchart of the hyperspectral image processing from flat field correction to calculation of the bruise ratio index.

Manual ROI Selection and Spectral Extraction

To accurately extract the spectra of healthy and bruised tissues, regions of interests (ROIs) were manually selected on the grayscale images at 1200 nm in ENVI (ENVI 4.7, Exelis Inc., VA, USA); this wavelength was chosen because the contrast between the bruised and

healthy tissues was most prominent at this wavelength. For Dataset1, to accurately extract the spectra of healthy and bruised tissues, a total of 31 samples were used for ROI selection, including 14 fully healthy and 17 fully-bruised samples. For Dataset2, a total of 600 (300 per treatment) samples from control and fully-bruised treatments were used for spectra extraction. One ROI was drawn on each selected sample, and the spectra of individual pixels in the ROIs were extracted. To balance the number of the extracted spectra, equal numbers of spectra for the two classes were selected from the extracted spectra. For the class with less spectra, all the spectra were kept, whereas for the class with more spectra, a subset of the spectra were randomly selected to keep the number of the spectra the same as the other class. Consequently, two spectral libraries (SpectraLib1 and SpectraLib2) were collected consisting of spectra of 28352 and 61580 pixels, respectively, and each contained equal numbers of spectra for healthy and bruised tissues.

Spectra and Image Classification

Support Vector Machine (SVM) was used for classification in this research. The SVM classifier was implemented by LibSVM (a software library for SVM classification) [267], and the training and evaluation were executed in MATLAB. All classifier parameters were in default values as the objective of this research was to explore the classification performance baseline of the proposed method. Three methods were used to evaluate the efficacy of the hyperspectral imaging approach: (a) two classifiers were trained and evaluated on each individual spectral library by 10-fold cross validation, respectively; (b) the classifier was trained on one spectral library but evaluated on another library; and (c) the classifier was trained and evaluated on the combined spectral library by 10-fold cross-validation. It should be noted that all three methods were for classification of the spectra extracted from ROIs.

The classifier trained by the combined spectral library was used as the final model to classify the masked hyperspectral images at the pixel level. According to a previous study

[268], the spectra of the calyx end were unrelated to bruises but often misclassified as bruised tissues, and thus the calyx end needed to be excluded in image classification. The calyx end was a 5-pixel-radius circle located at the center of each connected component in the masks, and the pixels within the circle were excluded during classification.

Based on the image classification results, bruise ratios were calculated to quantify bruises on each fruit. Since the samples in Dataset1 were only imaged on one side, the bruise ratio index of each sample was calculated by the ratio of the number of pixels classified as bruised to the total number of fruit pixels. However, in the results of Dataset2 where each fruit was imaged on two sides, the bruise ratio index of each sample was calculated by averaging the ratio of each half.

Firmness Prediction

Hyperspectral imaging has been used to predict fruit firmness by training regression models using the mean spectra of each fruit [257]. For comparison purposes, mean spectra of individual blueberry samples were extracted from the masked hyperspectral images, and the extracted spectra of each fruit with the measured firmness were used to train a PLSR model for fruit firmness prediction. According to previous studies [264, 265], to avoid effects caused by different initial conditions of blueberries, the spectra extracted from one experiment were not used to predict firmness of the samples in another experiment. Therefore, 300 and 1500 mean spectra were extracted and used for firmness prediction in the first and second experiments, respectively. The predicted firmness was used as an indirect index for blueberry bruising quantification and assessment.

8.3.5 Statistical Tests and Comparison

To rigorously prove the potential of using the hyperspectral imaging system for bruise identification, multivariate analysis of variance (MANOVA) tests were conducted to compare the extracted spectra of healthy and bruised tissues of different cultivars. All 141 wavelengths were considered as variables, and they were multivariate-normally distributed.

To explore the effectiveness of bruise quantification and assessment, bruise ratio index was compared with measured firmness, firmness predicted by PLSR, and human assessment of bruising level. Since firmness is an indirect index for bruising quantification and assessment, bruise ratio index is not directly comparable with firmness values. Multiple comparisons followed by Kruskal-Wallis test (nonparametric equivalent to analysis of variance test) were conducted to compare the differences among various treatments when using the bruise ratio index, measured and predicted firmness, respectively, as the data were not normally distributed. Two indices were considered to be equivalent if they showed the same statistical pattern among various treatments.

Since human assessment is also an area ratio of bruised tissue to cross-section of a fruit sample, the bruise ratio index and human assessment are comparable. The correlation between human assessment and the bruise ratio index was analyzed by using linear regression and the coefficient of determination (R^2). Additionally, a threshold was used to classify each fruit according to its bruise ratio index and human assessment, respectively. A fruit was classified as bruised if its bruise ratio index (or human assessment) exceeded a threshold of 0.2 (20% bruised area) in the present study. Analysis of variance (ANOVA) tests were conducted to compare the difference between the number of bruised fruit calculated using the bruise ratio index and human assessment for each treatment. The calculated number of bruised fruit in individual treatments were normally distributed.

The MANOVA and ANOVA tests were performed in SAS (package glm, SAS 9.3, SAS Institue Inc., NC, USA), whereas the multiple comparisons followed by Kruskal-Wallis test were conducted in R 3.2.4 [212] (package asbio). All tests were two-tailed and used a significance level of 0.05.

8.4 Results and Discussion

8.4.1 Grayscale Images at Representative Wavelengths

Grayscale images were observed at five representative wavelengths including three wavelengths for free water absorption (980, 1200, and 1470 nm) and two wavelengths for local peaks (1075 and 1650 nm). Generally, the blueberry samples of both southern and northern cultivars were hypo-intense (less bright) at higher wavelengths (1200, 1470, and 1650 nm) compared to lower wavelengths (980 and 1075 nm), whereas the background always absorbed most of the light at all five wavelengths (Figure 8.3). In particular, the contrast between the samples and background was the highest at 1075 nm because the samples absorbed the least light at this wavelength. Therefore, the grayscale images at 1075nm can be used to segment the blueberry samples from the background.

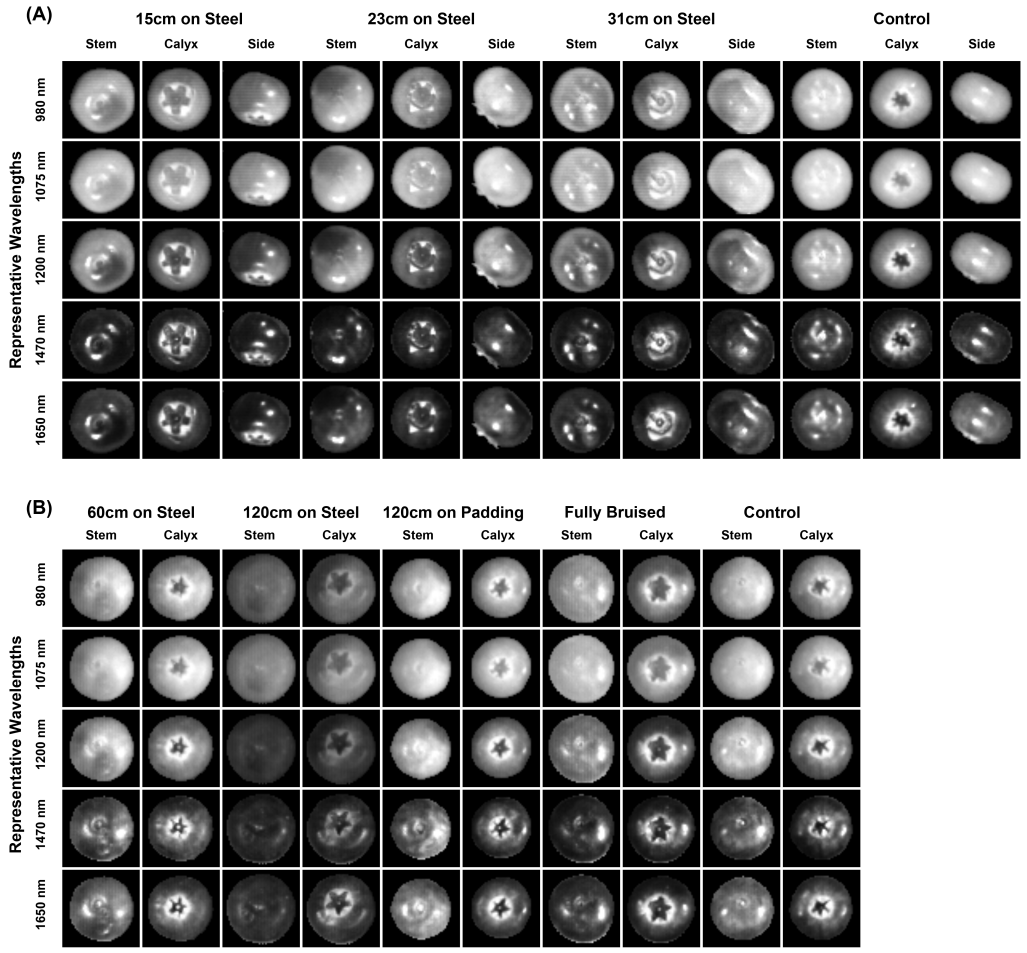

Figure 8.3: Grayscale images at representative wavelengths of blueberries: (A) for southern highbush cultivars (Dataset #1) and (B) for northern highbush cultivars (Dataset #2).

On each sample, healthy tissues were hyper-intense (brighter) while bruised tissues were hypo-intense (less bright) at 980, 1075, and 1200 nm. Both healthy and bruised tissues were hypo-intense at 1470 and 1650 nm. The contrast between the healthy tissues and bruised tissues was the highest at 1200 nm. Therefore, grayscale images at 1200 nm were used to manually select ROIs of healthy and bruised tissues for spectra extraction.

One exception was the calyx end, which appeared as dark as bruised tissues in most wavelengths. However, this phenomenon was observed from all the samples in both control and bruising treatments, and thus the low reflectance intensity was not due to bruises. In

fact, compared with other parts of the fruit, the calyx end typically is harder to bruise so the low reflectance intensity was most likely because the calyx end might have special chemical or physical properties, leading to more light absorption and less reflection. As a result, to accurately detect and quantify bruises, it was necessary to exclude the calyx end of each fruit in the following hyperspectral image classification and bruise ratio index calculation.

8.4.2 Reflectance Spectra of Healthy and Bruised Tissues

Since the present study focused on bruise detection and quantification, experimental variables were not discussed if they did not affect the separation between healthy and bruised tissues. According to statistical test results, the closest mean spectra of healthy and bruised tissues were statistically separable regardless of development time (see Supplementary Figure S1, Figure S2, and Table S1) and blueberry cultivar (see Supplementary Figure S3–S6, Table S2, and Table S3). Therefore, the spectral difference between healthy and bruised tissues was not affected by the two variables (Figure 8.4). Prior to spectral normalization, there was no obvious shape difference in the mean spectra of healthy (or bruised) tissues between southern (Dataset1) and northern (Dataset2) highbush blueberry cultivars (Figure 8.4a and 8.4b). Although the intensity of the spectra from NHB cultivars was on average higher than that of SHB cultivars, the difference was primarily caused by different illumination conditions. After removing this effect by spectral normalization, the mean spectra showed no obvious difference in both shape and intensity between SHB and NHB cultivars (Figure 8.4c and 8.4d). Therefore, the spectra extracted from one experiment can be used to train a classifier to differentiate the spectra extracted from another experiment.

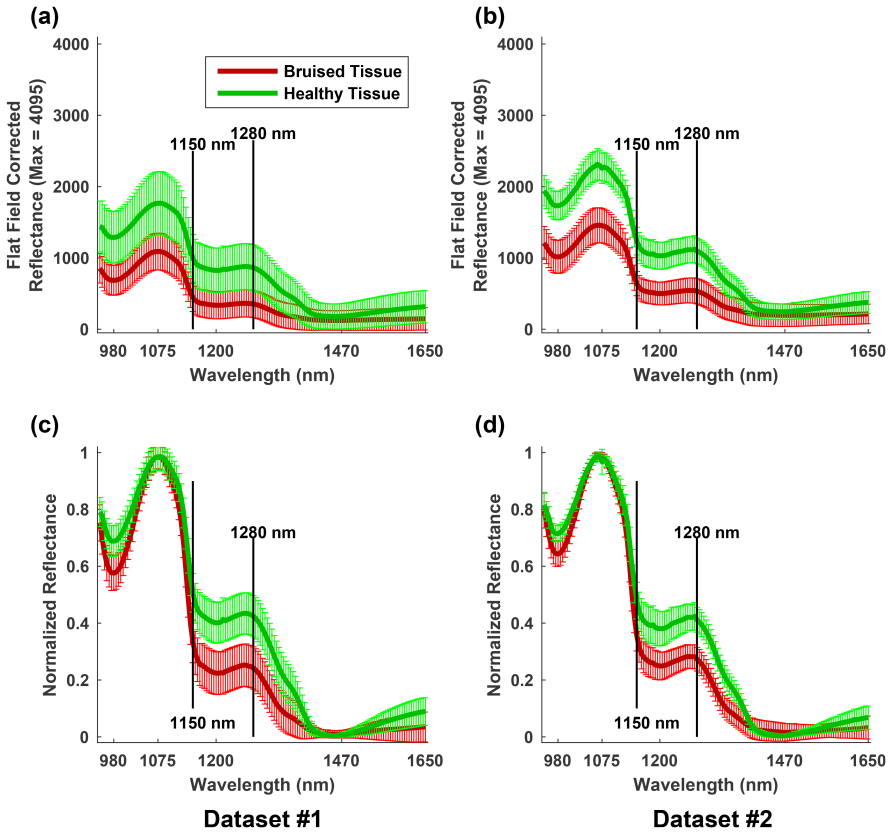

Figure 8.4: Mean spectra (solid line) and standard deviation (error bar) of healthy and bruised tissue. (a) and (b) are in flat field corrected reflectance for southern and northern cultivars, respectively; and (c) and (d) are in normalized reflectance for southern and northern cultivars, respectively.

The mean spectra between healthy and bruised tissues in both SHB and NHB cultivars were clearly different (Figure 8.4). Regarding the spectral shape, the mean spectra of healthy and bruised tissues can be grouped into three segments, i.e., 950 to 1150 nm, 1150 to 1280 nm, and 1280 to 1650 nm. The spectra of healthy and bruised tissues showed little or no difference in the first two segments but a significant difference in the third one. In the third spectral range from 1280 to 1650 nm, the intensity of the spectra decreased to the minimum at 1470 nm, and then began to increase for healthy tissues while being flat for bruised tissues between 1470 and 1650 nm. After spectral normalization, the upward trend of the spectral intensity of healthy tissues became more obvious (Figure 8.4c and 8.4d), which could be

useful to classify healthy and bruised tissues.

In addition, two local minima (at 980 and 1200 nm) and the global minimum (at 1470 nm) matched typical absorption wavelengths of liquid water [262], confirming the assumption that fruit bruising causes redistribution of free water in fruit tissues. According to a previous study [269], bruises cause damage to cell membranes and result in leaking free water. The free water stays around the bruised tissues in early stages, and then diffuses away and evaporates out of the fruit, resulting in dry cavities in late stages. Therefore, in early stages, bruised tissues contain more free water than healthy tissues and thus absorb more light at certain wavelengths with high liquid water absorption.

In terms of the spectral intensity, prior to spectral normalization, the intensity of the mean spectra of healthy tissues were always higher than that of bruised tissues (Figure 8.4a and 8.4b). The spectra of healthy and bruised tissues in southern cultivars (Dataset1) had a larger overlap than the spectra in northern cultivars (Dataset2). This occurred mainly because of the differences in manual ROI selection between the two experiments. Compared with Dataset2, Dataset1 had no fully-bruised treatment dedicated for spectra extraction of bruised tissues. Therefore, the selected ROIs for bruised tissues in Dataset1 might have higher possibility of mixing with healthy tissues than those in Dataset2, ultimately resulting in a large spectral overlap between two treatments. After spectral normalization, the differences between healthy and bruised tissues were most prominent at three ranges: 950 to 1000 nm, 1150 to 1400 nm, and 1500 to 1650 nm. Potentially, wavelengths in these three ranges could be used for feature selection and improvement of the classification performance.

8.4.3 Classification Results

Classification of Spectra Extracted from ROIs

Both the 10-fold cross-validation on the training set and the evaluation on independent test sets achieved accuracies of over 92%, indicating the proposed detecting method was accurate and robust (Table 8.1). The accuracies of the 10-fold cross-validation were over 94% with a variation up to 0.34% because normalized spectra of healthy and bruised tissues within each dataset were clearly separated with little overlap. Accuracies decreased 2.35% and 3.79% when using SpectraSet2 and SpectraSet1 as independent test set, respectively. The performance reduction was largely because testing the classifier using an independent test set is more rigorous than k-fold cross-validation. It should be noted that since all the parameters used in the SVM classifier were set to default values, they might not be optimal for all cases, and thus parameter optimization could further improve the accuracy. In addition, information at some wavelengths might be redundant features for classification, so feature selection could improve the classification performance as well.

Training set	Test set	10-fold cross validation on training set	Accuracy on test set
SpectraLib1	SpectraLib2	94.68±0.34%	92.41%
SpectraLib2	SpectraLib1	97.05±0.15%	93.29%

Table 8.1: Classification performance of using independent test set: SpectraSet1 consisting of 28352 spectra extracted from Dataset1 (southern cultivars) and SpectraSet2 consisting of 61580 spectra extracted from Dataset2 (northern cultivars).

Overall, the 10-fold cross-validation on the combined spectral data from SpectraSet1 and SpectraSet2 achieved over 96% accuracy for the test sets (Table 8.2). Although the accuracy of classifying pixels of healthy tissues (true negative) was slightly higher (2.26%) than that of bruised tissues (true positive), they were comparable classification results with satisfactory

accuracies. In addition, the classification performance on the combined spectral library was stable (performance variation was up to 0.21%). It should be noted that these performances are based on the spectral data extracted from the selected ROIs and the performance on all the images at the pixel level may vary.

	Classified as bruised	Classified as healthy	Total
Actual as bruised	42773(95.12%, true positive)	2193(4.88%, false negative)	44966
Actual as healthy	1179(2.62%, false positive)	43787(97.38%, true negative)	44966
Overall Accuracy: 96.25±0.21% for test			

Table 8.2: Confusion matrix of 10-fold cross validation on the combined spectra from Dataset1 and Dataset2

Image Classification

For Dataset1, the proposed method recognized bruises using hyperspectral images, and the results closely matched the observations of bruises in the color images of the sliced fruit (Figure 8.5). The blueberry samples were sliced perpendicular to the bruise position, which was controlled in the experiment, so the bruises were easily observed on the slicing plane. In addition, it was noteworthy that the pendulum wooden arm with the sample holder (22.5-29 g) was 15 times heavier than a typical blueberry sample (1.5-2 g), and thus bruises created by the pendulum were significantly more severe than those created by randomly dropping the fruit onto a steel surface from the same height. Therefore, the bruises created in the first experiment were clearly discolored compared to healthy tissues. Some samples in control groups were highly or fully bruised in both image classification results and color images of sliced fruit, because they may have been damaged during transportation prior to the experiments. The bruises were more severe on the stem side and equatorial axis than on the calyx side, which validated the assumption that calyx end was harder to bruise.

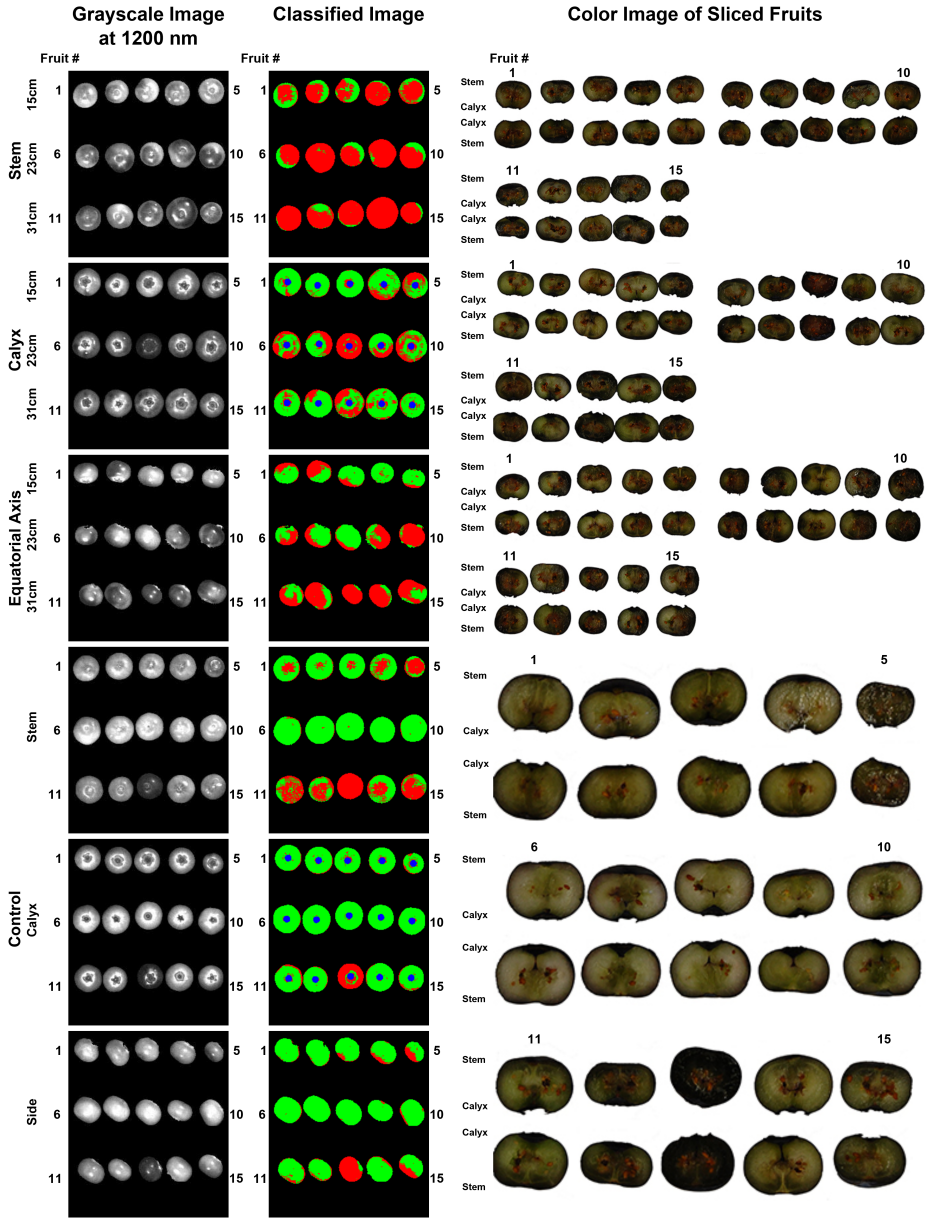

Figure 8.5: Grayscale image at 1200 nm, classified image, and color image of sliced fruit of representative results from Dataset1 (southern highbush cultivars). Red and green color represent bruised and healthy tissues, respectively.

For Dataset2, although the proposed method identified bruises, sometimes the results did not match with the observations of bruises in the color images of the sliced fruit (Figure 8.6). This occurred mainly because different bruise creation and slicing methods were used in the second experiment. Compared with those created by the pendulum, the bruises created

by randomly dropping from certain heights were less severe and tended to be shallower. Consequently, if blueberry samples were only sliced along the equatorial axis, the bruises would not be observed on the slicing plane if they occurred at shallow positions not along the equatorial axis. The bruises were consistently more severe on the stem side than on the calyx side. In addition, the amount of bruises and their severity were related to the drop height and the impact surface [268]. The classified images clearly showed this pattern. When samples were dropped from different heights onto the same steel surface, more bruises (red areas) were observed in the samples with higher drop heights. When samples were dropped onto different surfaces from the same height, less bruises (red areas) were observed in the samples dropped onto the padded surface than onto the steel surface.

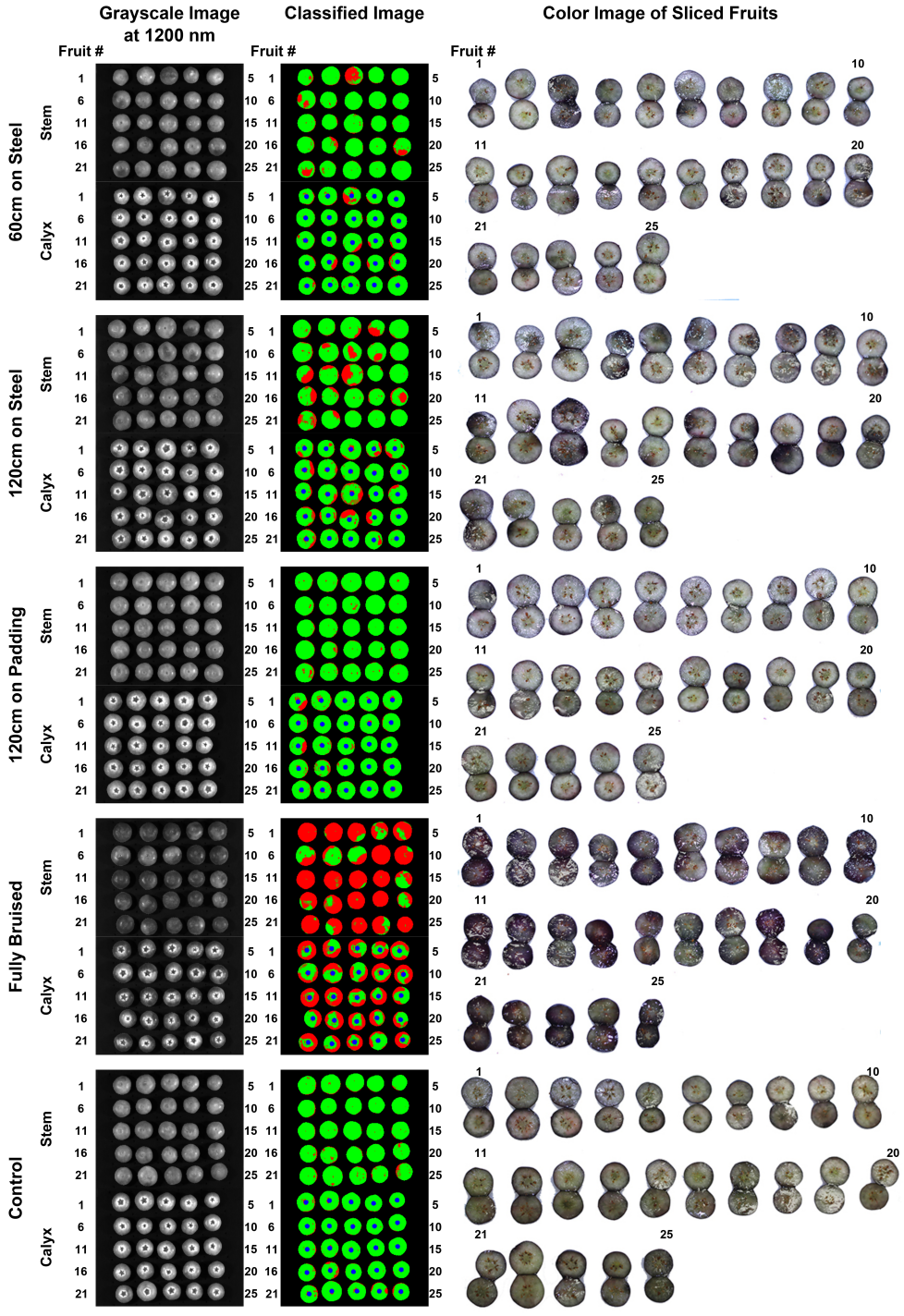

Figure 8.6: Grayscale image at 1200 nm, classified image, and color image of sliced fruit of representative results from Dataset2 (northern highbush cultivars). Red and green color represent bruised and healthy tissues, respectively.

8.4.4 Comparison between Bruise Ratio Index and Traditional Measurements

The statistical patterns obtained using bruise ratio index matched the patterns calculated using firmness measurement (see (a1), (a2) and (a4), (a5) in Figure 8.7A). When the drop height was increased, more bruises were created, and thus higher bruise ratio index and lower firmness of blueberry samples were measured. In addition, the results showed the same statistical significance among the treatments when using bruise ratio index and measured firmness, respectively (see Supplementary Table S4–S7). Therefore, the bruise ratio index calculated by the proposed non-destructive method was equivalent to measured firmness in the statistical tests.

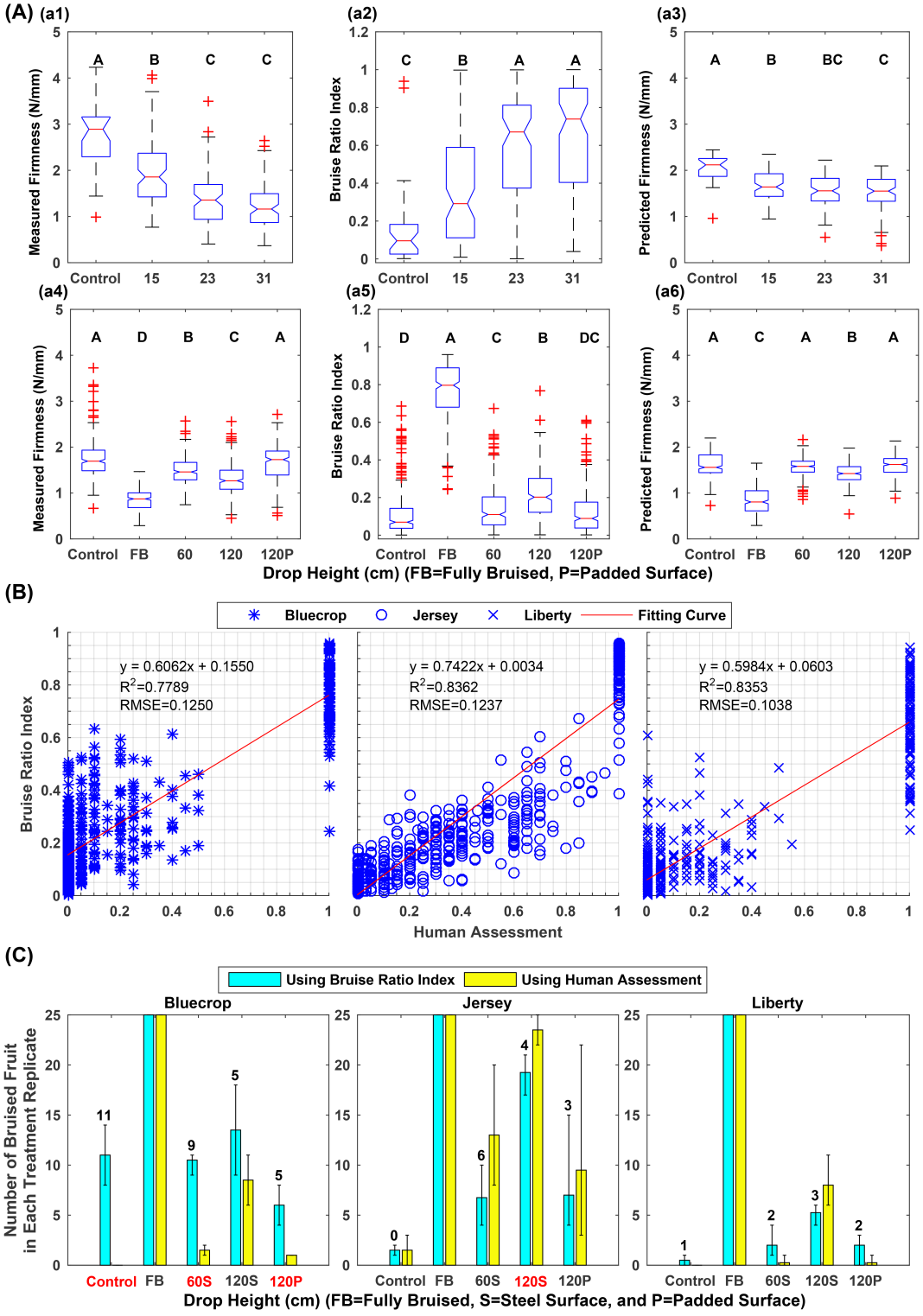

Figure 8.7: Statistical analysis results in the present study. Panel (A): boxplot of measured firmness, bruise ratio index extracted from HSI, and predicted firmness by PLSR. Treatments with different letters are statistically significant with each other (see Supplementary Table S4–S9), and treatment mean values of each index are sorted alphabetically. (a1-a3) and (a4-a6) are measured firmness, bruise ratio index, and predicted firmness for Dataset1 (southern cultivars) and Dataset2 (northern cultivars), respectively. Panel (B): linear regres-

Although it showed a certain efficacy, firmness predicted by the PLSR model did not show statistical differences for some treatments (see (a1), (a3) and (a4), (a6) in Figure 8.7A). In Dataset1, the predicted firmness of the control treatment was statistically higher than that of the bruise treatments. However, the predicted firmness of bruise treatments dropped from 15 cm was not statistically higher than that of the treatments dropped from 23 and 31 cm, and thus the predicted firmness could not accurately represent bruises caused by different drop heights (see Supplementary Table S8). In Dataset2, although the treatments dropped onto steel and padded surfaces were statistically different, there was no significant difference between the control and bruise treatments dropped from 60 cm onto a steel surface and from 120 cm onto a padded surface (see Supplementary Table S9). Thus, the predicted firmness had the effectiveness of bruising assessment in certain situations, but it could not quantify bruising as accurately as the bruise ratio index. These observations agreed with previous studies [264, 265] which showed that the predicted firmness had a relatively large RMSEP using either reflectance or transmittance spectra. Therefore, compared with the predicted firmness, bruise ratio index could be a more effective index to non-destructively quantify and assess blueberry bruising.

Overall, the R^2 value (0.78-0.84) indicated that bruise ratio index was strongly correlated with human assessment for northern cultivars (Figure 8.7B). However, the Root Mean Squared Error (RMSE) was up to 12.5% of the bruise ratio index range (0 to 1), suggesting that bruise ratio index might be considerably different from human assessment in some treatments. In fact, the correlation analyses were further validated by the results of the thresholding classification of bruised fruit (Figure 8.7C). For most cases, the differences between the number of bruised fruit calculated using bruise ratio index and human assessment were within 5 and they were not statistically different (see Supplementary Table S10–S12). This confirmed that bruise ratio index was correlated with human assessment. However, for some cases such as the control group of Bluecrop, the number of bruised fruit

calculated using bruise ratio index was significantly different from that calculated using human assessment (treatment names in red in Figure 8.7C, see Supplementary Table S10–S12). Although human visual inspection was the most intuitive approach to evaluate and quantify blueberry bruising, it can only observe and evaluate bruises that developed on the slicing plane. For those bruises that did not occur on the equatorial axis or the slicing plane, human graders could not observe and evaluate them. For instance, a sample with a firmness of 1.46 N/mm was graded as healthy fruit with no bruised tissue, whereas the proposed imaging method measured a bruise ratio index of 0.2492 (Figure 8.8a). To further explain this inconsistency between the human assessment and bruise ratio index, more advanced imaging techniques such as MRI need to be explored to characterize blueberry bruising.

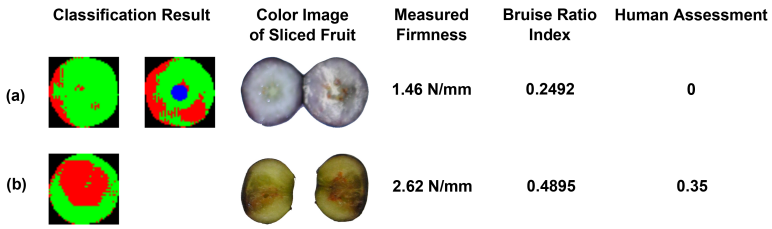

Figure 8.8: Inconsistent cases between bruise ratio index and traditional indices: (a) in consistency between bruise ratio index and human assessment (b) inconsistency between bruise ratio index and measured firmness.

Additionally, inconsistencies between measured firmness, bruise ratio index, and human assessment were observed. For instance, a blueberry with a firmness of 2.62 N/mm had a bruise ratio index of 0.4895 and a human assessment of 0.35 (Figure 8.8b). Both bruise ratio index and human assessment were high, indicating that the fruit had bruises on it. However, the firmness measurement was high as well, contrary to the fact that bruised berries tend to be softer. This likely occurred because the instrument only measures the firmness in a local area. If a bruise develops 90 degrees from the measuring axis of the instrument, the fruit firmness reading may be greater than that measured along the bruising axis. Nevertheless, the fact was that this sample had spectra similar to bruised fruit but a firmness value similar to healthy fruit. This could explain the reason that the predicted firmness had large

RMSEP, which was not good for accurately quantifying blueberry bruising. If a training set contains samples like the fruit in Figure 8.8b, the training process will force the PLSR model to fit those data, which may lead to large errors in predicting the firmness of common samples.

Bruise ratio as an index calculated by a non-destructive imaging approach showed a strong correlation with human assessment measured in a destructive manner. In addition, in the statistical tests, bruise ratio index was equivalent to firmness measured by instruments but better than firmness predicted by using hyperspectral imaging with the PLSR model. Therefore, bruise ratio index could be an effective index for quantification and assessment of blueberry bruising. Potentially, this method could also be used for assessing bruises of other berry fruits.

8.5 Conclusions

The proposed non-destructive approach based on NIR hyperspectral imaging was an accurate and stable method to detect blueberry bruising. Compared with traditional indices, bruise ratio index was an effective index for quantification of blueberry bruising. Therefore, the proposed method and index could be used to non-destructively quantify and assess blueberry bruising by both researchers and the commercial industry. Future studies will be focused on feature selection and classifier optimization to further improve the efficiency and accuracy of the method, enabling the proposed method for online sorting of bruised blueberries.

Chapter 9

Limitations and Future Work

9.1 Limitations

In this dissertation, the developed data acquisition system and analytical methods strongly enhanced the sensing and analysis capability for field-based, high-throughput plant phenotyping and for laboratory-based postharvest quality assessment. Extracted phenotypic traits also demonstrated the usefulness for genetics/genomics studies, breeding programs, and precise farm management in small- and medium-scale experiments, proving the value of technologies developed for improving agricultural productivity and sustainability. Cotton and blueberry, which represented a typical field crop and a specialty crop respectively, were used as model crops. The system developed and analytical methods applied herein could be potentially extended to plant phenotyping and postharvest quality assessment for other crops. However, several limitations were identified in this dissertation:

1. Plant structure, as an important aspect of plant morphology, needs to be further studied by using new 3D imaging modalities that can acquire high resolution, full-view 3D point clouds for plants in the field. Important parameters (e.g., internode properties and branch angles) are expected to be extracted from those high resolution, full-view 3D point clouds, supplementing the current quantification of plant morphology.

2. As the primary focus was to achieve the best detection and counting accuracy, implementations of CNN-based approaches developed for seedling and flower detection and counting were not optimized for computing efficiency, which prevents the use of those approaches in embedded systems and online applications.

3. DeepFlower framework did not take full advantage of using images taken from multiple viewing angles for bloom detection and counting, which significantly reduced the counting accuracy after certain plant growth stages.

4. Optical properties of bruised blueberry tissues have not been thoroughly studied in this dissertation, limiting the understanding of fruit bruises and the optimization of sensing and processing methods for bruising detection and quantification.

9.2 Future Work

Due to the time and scope limitations of this dissertation, future studies are expected to address several issues identified:

1. Multi-year and multi-location experiments are necessary to validate the developed methodology for estimation of canopy-level photosynthetic efficiency using a passive sensing approach. Through these experiments, it is expected to conclusively obtain the optimal method for using diurnal SIF measurements to estimate maximal fluorescence for stressed plants.

2. Nationwide collaborative efforts are expected to establish large, open, and annotated datasets and benchmarks for agricultural applications (e.g., plant phenotyping), which can significantly facilitate the adoption of DL techniques for plant phenotyping.

3. Efficient transfer learning/active learning methods will significantly improve the use of deep learning techniques for a wide range of agricultural applications.

4. It is worthwhile to explore an end-to-end CNN meta-model for fusing multimodal imaging data for the understanding of plant development.

5. Several methods reported in this dissertation are expected to be applied in large-scale experiments to leverage genetics/genomics studies.

6. In the future, robotics and information systems technologies can be used to further improve the automation of data collection, management, and processing, thereby producing a data-driven system that could support both agricultural production systems and scientific research.

References

[1] Ghosal S, Blystone D, Singh AK, Ganapathysubramanian B, Singh A, Sarkar S. An explainable deep machine vision framework for plant stress phenotyping [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(18):4613–4618.

[2] Toda Y, Okura F. How Convolutional Neural Networks Diagnose Plant Disease [Journal Article]. Plant Phenomics. 2019;2019:14.

[3] Tilman D, Balzer C, Hill J, Befort BL. Global food demand and the sustainable intensification of agriculture [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(50):20260–20264.

[4] Steensland A, Zeigler M. 2018 Global Agricultural Productivity Report [Online Multimedia]; 2018. Https://globalagriculturalproductivity.org/. April 7, 2019 Accessed.

[5] United Nations Department of Economic and Social Affairs. World Population Prospects: The 2017 Revision; 2017. Available from: `https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html`.

[6] Schmidhuber J, Tubiello FN. Global food security under climate change [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(50):19703–19708.

[7] Araus JL, Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier [Journal Article]. Trends in Plant Science. 2014;19(1):52–61.

[8] Fiorani F, Schurr U. Future Scenarios for Plant Phenotyping [Journal Article]. Annual Review of Plant Biology. 2013;64:267–291.

[9] Ghanem ME, Marrou H, Sinclair TR. Physiological phenotyping of plants for crop improvement [Journal Article]. Trends in Plant Science. 2015;20(3):139–144.

[10] Cairns JE, Hellin J, Sonder K, Araus JL, MacRobert JF, Thierfelder C, et al. Adapting maize production to climate change in sub-Saharan Africa [Journal Article]. Food Security. 2013;5(3):345–360.

[11] Cairns JE, Impa SM, O'Toole JC, Jagadish SVK, Price AH. Influence of the soil physical environment on rice (Oryza sativa L.) response to drought stress and its implications for drought research [Journal Article]. Field Crops Research. 2011;121(3):303–310.

[12] Whitmore AP, Whalley WR. Physical effects of soil drying on roots and crop growth [Journal Article]. Journal of Experimental Botany. 2009;60(10):2845–2857.

[13] Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Araus JL. High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge [Journal Article]. Journal of Integrative Plant Biology. 2012;54(5):312–320.

[14] Dhondt S, Wuyts N, Inze D. Cell to whole-plant phenotyping: the best is yet to come [Journal Article]. Trends in Plant Science. 2013;18(8):433–444.

[15] Li L, Zhang Q, Huang D. A Review of Imaging Techniques for Plant Phenotyping [Journal Article]. Sensors. 2014;14(11):20078–20111.

[16] Yang W, Duan L, Chen G, Xiong L, Liu Q. Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies [Journal Article]. Current Opinion in Plant Biology. 2013;16(2):180–187.

[17] Yeh CH, Kaplinsky NJ, Hu C, Charng YY. Some like it hot, some like it warm: Phenotyping to explore thermotolerance diversity [Journal Article]. Plant Science. 2012;195:10–23.

[18] Zhu J, Ingram PA, Benfey PN, Elich T. From lab to field, new approaches to phenotyping root system architecture [Journal Article]. Current Opinion in Plant Biology. 2011;14(3):310–317.

[19] Furbank RT, Tester M. Phenomics - technologies to relieve the phenotyping bottleneck [Journal Article]. Trends in Plant Science. 2011;16(12):635–644.

[20] Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge [Journal Article]. Nature Reviews Genetics. 2010;11(12):855–866.

[21] Ninomiya S, Baret F, Cheng ZMM. Plant phenomics: emerging transdisciplinary science [Journal Article]. Plant Phenomics. 2019;.

[22] Pieruschka R, Schurr U. Plant Phenotyping: Past, Present, and Future [Journal Article]. Plant Phenomics. 2019;2019:7507131.

[23] Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. Plant Phenomics, From Sensors to Knowledge [Journal Article]. Current Biology. 2017;27(15):R770–R783.

[24] Minervini M, Scharr H, Tsaftaris SA. Image Analysis: The New Bottleneck in Plant Phenotyping [Journal Article]. IEEE Signal Processing Magazine. 2015;32(4):126–131.

[25] Fukushima K. Neocognitron-A Self-Organizing Neural Network Model for a Mechanism of Pattern-Recognition Unaffected by Shift in Position [Journal Article]. Biological Cybernetics. 1980;36(4):193–202.

[26] LeCun Y, Boser B, Denker JS, Howard RE, Habbard W, Jackel LD, et al. In: David ST, editor. Handwritten digit recognition with a back-propagation network. Morgan Kaufmann Publishers Inc.; 1990. p. 396–404.

[27] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems; 2012. p. 1097–1105.

[28] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 770–778.

[29] Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review [Journal Article]. Neural Computation. 2017;29(9):2352–2449.

[30] Zhao Z, Zheng P, Xu S, Wu X. Object detection with deep learning: A review [Journal Article]. IEEE Transactions on Neural Networks and Learning Systems. 2019;.

[31] Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Martinez-Gonzalez P, Garcia-Rodriguez J. A survey on deep learning techniques for image and video semantic segmentation [Journal Article]. Applied Soft Computing. 2018;70:41–65.

[32] Kamilaris A, Prenafeta-Boldu FX. Deep learning in agriculture: A survey [Journal Article]. Computers and Electronics in Agriculture. 2018;147:70–90.

[33] Singh AK, Ganapathysubramanian B, Sarkar S, Singh A. Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives [Journal Article]. Trends in Plant Science. 2018;23(10):883–898.

[34] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision. Springer; 2014. p. 818–833.

[35] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [Journal Article]. arXiv preprint arXiv:14091556. 2014;.

[36] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 2818–2826.

[37] Szegedy C, Wei L, Yangqing J, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 1–9.

[38] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Springer; 2016. p. 630–645.

[39] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 4700–4708.

[40] Zoph B, Le QV. Neural Architecture Search with Reinforcement Learning [Electronic Article]; 2016. Https://ui.adsabs.harvard.edu/#abs/2016arXiv161101578Z. November 01, 2016 Accessed.

[41] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition [Journal Article]. arXiv preprint arXiv:170707012. 2017;.

[42] Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. Overfeat: Integrated recognition, localization and detection using convolutional networks [Journal Article]. arXiv preprint arXiv:13126229. 2013;.

[43] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 580–587.

[44] Girshick R. Fast R-CNN. In: IEEE International Conference on Computer Vision. IEEE; 2015. p. 1440–1448.

[45] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems; 2015. p. 91–99.

[46] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 779–788.

[47] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. In: IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 6517–6525.

[48] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: Single shot multibox detector. In: European Conference on Computer Vision. Springer; 2016. p. 21–37.

[49] Lin T, Goyal P, Girshick R, He K, Dollr P. Focal Loss for Dense Object Detection. In: IEEE International Conference on Computer Vision; 2017. p. 2999–3007.

[50] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. Springer; 2015. p. 234–241.

[51] Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [Journal Article]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018;40(4):834–848.

[52] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation [Journal Article]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(4):640–651.

[53] Chen L, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation [Electronic Article]; 2017. Https://ui.adsabs.harvard.edu/#abs/2017arXiv170605587C. June 01, 2017 Accessed.

[54] Hariharan B, Arbelez P, Girshick R, Malik J. Simultaneous detection and segmentation. In: European Conference on Computer Vision. Springer; 2014. p. 297–312.

[55] Pinheiro PO, Collobert R, Dollr P. Learning to segment object candidates. In: Advances in Neural Information Processing Systems; 2015. p. 1990–1998.

[56] He K, Gkioxari G, Dollr P, Girshick R. Mask R-CNN. In: IEEE International Conference on Computer Vision; 2017. p. 2980–2988.

[57] Barbedo JGA. Factors influencing the use of deep learning for plant disease recognition [Journal Article]. Biosystems Engineering. 2018;172:84–91.

[58] Barbedo JGA. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification [Journal Article]. Computers and Electronics in Agriculture. 2018;153:46–53.

[59] Brahimi M, Boukhalfa K, Moussaoui A. Deep Learning for Tomato Diseases: Classification and Symptoms Visualization [Journal Article]. Applied Artificial Intelligence. 2017;31(4):299–315.

[60] Cruz AC, Luvisi A, De Bellis L, Ampatzidis Y. X-FIDO: An Effective Application for Detecting Olive Quick Decline Syndrome with Deep Learning and Data Fusion [Journal Article]. Frontiers in Plant Science. 2017;8.

[61] Ferentinos KP. Deep learning models for plant disease detection and diagnosis [Journal Article]. Computers and Electronics in Agriculture. 2018;145:311–318.

[62] Liu B, Zhang Y, He D, Li Y. Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks [Journal Article]. Symmetry. 2018;10(1).

[63] Lu Y, Yi S, Zeng N, Liu Y, Zhang Y. Identification of rice diseases using deep convolutional neural networks [Journal Article]. Neurocomputing. 2017;267:378–384.

[64] Mohanty SP, Hughes DP, Salathe M. Using Deep Learning for Image-Based Plant Disease Detection [Journal Article]. Frontiers in Plant Science. 2016;7.

[65] Suh HK, Ijsselmuiden J, Hofstee JW, van Henten EJ. Transfer learning for the classification of sugar beet and volunteer potato under field conditions [Journal Article]. Biosystems Engineering. 2018;174:50–65.

[66] Wang G, Sun Y, Wang J. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning [Journal Article]. Computational Intelligence and Neuroscience. 2017;.

[67] DeChant C, Wiesner-Hanks T, Chen SY, Stewart EL, Yosinski J, Gore MA, et al. Automated Identification of Northern Leaf Blight-Infected Maize Plants from Field Imagery Using Deep Learning [Journal Article]. Phytopathology. 2017;107(11):1426–1432.

[68] Fuentes A, Yoon S, Kim SC, Park DS. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition [Journal Article]. Sensors. 2017;17(9).

[69] Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C. DeepFruits: A Fruit Detection System Using Deep Neural Networks [Journal Article]. Sensors. 2016;16(8).

[70] Kashkoush M. Region-Based Convolutional Networks for End-to-End Detection of Agricultural Mushrooms. In: 8th International Conference on Image and Signal Processing. vol. 10884. Springer; 2018. p. 319.

[71] Madec S, Jin X, Lu H, De Solan B, Liu S, Duyme F, et al. Ear density estimation from high resolution RGB imagery using deep learning technique [Journal Article]. Agricultural and Forest Meteorology. 2019;264:225–234.

[72] Mai X, Zhang H, Meng Q. Faster R-CNN with Classifier Fusion for Small Fruit Detection [Journal Article]. 2018 IEEE International Conference on Robotics and Automation. 2018;p. 7166–7172.

[73] Dias PA, Tabb A, Medeiros H. Apple flower detection using deep convolutional networks [Journal Article]. Computers in Industry. 2018;99:17–28.

[74] Bresilla K, Perulli GD, Boini A, Morandi B, Corelli Grappadelli L, Manfrini L. Single-Shot Convolution Neural Networks for Real-Time Fruit Detection Within the Tree. Frontiers in Plant Science. 2019;10:611.

[75] Rahnemoonfar M, Sheppard C. Deep Count: Fruit Counting Based on Deep Simulated Learning [Journal Article]. Sensors. 2017;17(4).

[76] Giuffrida MV, Doerner P, Tsaftaris SA. Pheno-Deep Counter: a unified and versatile deep learning architecture for leaf counting [Journal Article]. The Plant Journal. 2018;0(0).

[77] Ubbens JR, Stavness I. Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks [Journal Article]. Frontiers in Plant Science. 2017;8.

[78] Lu H, Cao Z, Xiao Y, Zhuang B, Shen C. TasselNet: counting maize tassels in the wild via local counts regression network [Journal Article]. Plant Methods. 2017;13.

[79] Wang X, Xuan H, Evers B, Shrestha S, Pless R, Poland J. High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat [Journal Article]. bioRxiv. 2019;p. 527911.

[80] Stein M, Bargoti S, Underwood J. Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry [Journal Article]. Sensors. 2016;16(11).

[81] Halstead M, McCool C, Denman S, Perez T, Fookes C. Fruit Quantity and Ripeness Estimation Using a Robotic Vision System [Journal Article]. IEEE Robotics and Automation Letters. 2018;3(4):2995–3002.

[82] Liu X, Chen SW, Aditya S, Sivakumar N, Dcunha S, Qu C, et al. Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure from Motion [Journal Article]. arXiv. 2018;.

[83] Fleet D, Weiss Y. In: Optical flow estimation. Springer; 2006. p. 237–257.

[84] Ghosal S, Zheng B, Chapman SC, Potgieter AB, Jordan DR, Wang X, et al. A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting [Journal Article]. Plant Phenomics. 2019;2019:14.

[85] Xu R, Li C, Paterson AH, Jiang Y, Sun S, Robertson JS. Aerial Images and Convolutional Neural Network for Cotton Bloom Detection. Frontiers in Plant Science. 2018;8:2235.

[86] Hni N, Roy P, Isler V. A Comparative Study of Fruit Detection and Counting Methods for Yield Mapping in Apple Orchards [Journal Article]. arxiv. 2018;.

[87] Nellithimaru AK, Kantor GA. ROLS: Robust Object-Level SLAM for Grape Counting. In: IEEE Conference on Computer Vision and Pattern Recognition; 2019. p. 0–0.

[88] Azizah LM, Umayah SF, Riyadi S, Damarjati C, Utama NA. Deep learning implementation using convolutional neural network in mangosteen surface defect detection. In: 7th IEEE International Conference on Control System, Computing and Engineering. IEEE; 2017. p. 242–246.

[89] Liu Z, He Y, Cen H, Lu R. Deep Feature Representation with Stacked Sparse Auto-Encoder and Convolutional Neural Network for Hyperspectral Imaging-Based Detection of Cucumber Defects [Journal Article]. Transactions of the ASABE. 2018;61(2):425–436.

[90] Oppenheim D, Shani G, Erlich O, Tsror L. Using Deep Learning for Image-Based Potato Tuber Disease Detection. Phytopathology. 2019;109(6):1083–1087.

[91] Wang Z, Hu M, Zhai G. Application of Deep Learning Architectures for Accurate and Rapid Detection of Internal Mechanical Damage of Blueberry Using Hyperspectral Transmittance Data [Journal Article]. Sensors. 2018;18(4).

[92] Sun Y, Lu R, Lu Y, Tu K, Pan L. Detection of early decay in peaches by structured-illumination reflectance imaging [Journal Article]. Postharvest Biology and Technology. 2019;151:68–78.

[93] Cavallo DP, Cefola M, Pace B, Logrieco AF, Attolico G. Non-destructive automatic quality evaluation of fresh-cut iceberg lettuce through packaging material [Journal Article]. Journal of Food Engineering. 2018;223:46–52.

[94] Itakura K, Saito Y, Suzuki T, Kondo N, Hosoi FJH. Estimation of Citrus Maturity with Fluorescence Spectroscopy Using Deep Learning [Journal Article]. Horticulturae. 2019;5(1):2.

[95] Zhang M, Li C. Fully convolutional networks for blueberry bruising and calyx segmentation using hyperspectral transmittance imaging. In: ASABE Annual International Meeting, Paper No: 1801489. American Society of Agricultural and Biological Engineers; 2018. p. 1.

[96] Su H, Deng J, Li F. Crowdsourcing Annotations for Visual Object Detection. 2012; 2012. Available from: `https://www.aaai.org/ocs/index.php/WS/AAAIW12/paper/view/5350`.

[97] Qi CR, Su H, Mo K, Guibas LJ. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 652–660.

[98] Qi CR, Yi L, Su H, Guibas LJ. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems; 2017. p. 5099–5108.

[99] National Cottonseed Products Association. Cottonseed and its products [Web Page]; 2017. Available from: http://www.cottonseed.org/publications/cottonseedanditsproducts.asp.

[100] Reynolds M, Langridge P. Physiological breeding [Journal Article]. Current Opinion in Plant Biology. 2016;31:162–171.

[101] Humplik JF, Lazar D, Husickova A, Spichal L. Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses - a review [Journal Article]. Plant Methods. 2015;11.

[102] Golzarian MR, Frick RA, Rajendran K, Berger B, Roy S, Tester M, et al. Accurate inference of shoot biomass from high-throughput images of cereal plants [Journal Article]. Plant Methods. 2011;7.

[103] Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, et al. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice [Journal Article]. Nature Communications. 2014;5.

[104] Neilson EH, Edwards AM, Blomstedt CK, Berger B, Mller BL, Gleadow RM. Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time [Journal Article]. Journal of Experimental Botany. 2015;66(7):1817–1832.

[105] Ge Y, Bai G, Stoerger V, Schnable JC. Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging [Journal Article]. Computers and Electronics in Agriculture. 2016;127:625–632.

[106] White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, Conley MM, et al. Field-based phenomics for plant genetics research [Journal Article]. Field Crops Research. 2012;133:101–112.

[107] Deery D, Jimenez-Berni J, Jones H, Sirault X, Furbank R. Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping [Journal Article]. Agronomy. 2014;4(3):349.

[108] Sankaran S, Khot LR, Espinoza CZ, Jarolmasjed S, Sathuvalli VR, Vandemark GJ, et al. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review [Journal Article]. European Journal of Agronomy. 2015;70:112–123.

[109] Busemeyer L, Mentrup D, Moller K, Wunder E, Alheit K, Hahn V, et al. BreedVision - A Multi-Sensor Platform for Non-Destructive Field-Based Phenotyping in Plant Breeding [Journal Article]. Sensors. 2013;13(3):2830–2847.

[110] Andrade-Sanchez P, Gore MA, Heun JT, Thorp KR, Carmo-Silva AE, French AN, et al. Development and evaluation of a field-based high-throughput phenotyping platform [Journal Article]. Functional Plant Biology. 2014;41(1):68–79.

[111] Sharma B, Ritchie GL. High-Throughput Phenotyping of Cotton in Multiple Irrigation Environments [Journal Article]. Crop Science. 2015;55(2):958–969.

[112] Barker J, Zhang N, Sharon J, Steeves R, Wang X, Wei Y, et al. Development of a field-based high-throughput mobile phenotyping platform [Journal Article]. Computers and Electronics in Agriculture. 2016;122:74–85.

[113] Bai G, Ge Y, Hussain W, Baenziger PS, Graef G. A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding [Journal Article]. Computers and Electronics in Agriculture. 2016;128:181–192.

[114] Nguyen TT, Slaughter DC, Townsley BT, Carriedo L, Maloof JN, Sinha N. In-field Plant Phenotyping using Multi-view Reconstruction: An Investigation in Eggplant. In: International Conference on Precison Agriculture; 2016. .

[115] Shafiekhani A, Kadam S, Fritschi FB, DeSouza GN. Vinobot and Vinoculer: Two Robotic Platforms for High-Throughput Field Phenotyping [Journal Article]. Sensors. 2017;17(1).

[116] Busemeyer L, Ruckelshausen A, Muller K, Melchinger AE, Alheit KV, Maurer HP, et al. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation [Journal Article]. Scientific Reports. 2013;3.

[117] Pauli D, Andrade-Sanchez P, Carmo-Silva AE, Gazave E, French AN, Heun J, et al. Field-Based High-Throughput Plant Phenotyping Reveals the Temporal Patterns of Quantitative Trait Loci Associated with Stress-Responsive Traits in Cotton [Journal Article]. G3-Genes Genomes Genetics. 2016;6(4):865–879.

[118] Sharma B, Ritchie GL, Rajan N. Near-Remote Green: Red Perpendicular Vegetation Index Ground Cover Fraction Estimation in Cotton [Journal Article]. Crop Science. 2015;55(5):2252–2261.

[119] Fursattel P, Placht S, Balda M, Schaller C, Hofmann H, Maier A, et al. A Comparative Error Analysis of Current Time-of-Flight Sensors [Journal Article]. IEEE Transactions on Computational Imaging. 2016;2(1):27–41.

[120] Jiang Y, Li C, Paterson AH. High throughput phenotyping of cotton plant height using depth images under field conditions [Journal Article]. Computers and Electronics in Agriculture. 2016;130:57–68.

[121] Lawrence KC, Park B, Windham WR, Mao C. Calibration of a pushbroom hyperspectral imaging system for agricultural inspection [Journal Article]. Transactions of the ASAE. 2003;46(2):513–521.

[122] Jones HG, Archer N, Rotenberg E, Casa R. Radiation measurement for plant ecophysiology [Journal Article]. Journal of Experimental Botany. 2003;54(384):879–889.

[123] Campbell GS, Norman JM. An introduction to environmental biophysics. Springer Science & Business Media; 2012.

[124] Burton GW, Devane EH. Estimating Heritability in Tall Fescue (Festuca-Arundinacea) from Replicated Clonal Material [Journal Article]. Agronomy Journal. 1953;45(10):478–481.

[125] National Cotton Council of America. Growth and development of a cotton plant [Web Page]; 2016. Available from: `http://www.cotton.org/tech/ace/growth-and-development.cfm`.

[126] United States Department of Agriculture, Agricultural Research Service. National nutrient database for standard reference release 28 [Online Database]; 2016. Https://ndb.nal.usda.gov/ndb/foods/show/2166. 9 August 2017 Accessed.

[127] Overview global blueberry market [Web Page]; 2017. Http://www.freshplaza.com/article/174017/OVERVIEW-GLOBAL-BLUEBERRY-MARKET. Accessed.

[128] Shakoor N, Lee S, Mockler TC. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field [Journal Article]. Current Opinion in Plant Biology. 2017;38:184–192.

[129] Rosell JR, Sanz R. A review of methods and applications of the geometric characterization of tree crops in agricultural activities [Journal Article]. Computers and Electronics in Agriculture. 2012;81:124–141.

[130] Smith ED, Harris G. Plant Nitrogen Status of Southern Highbush Blueberry (Vaccinium corymbosum L. Interspecific Hybrid) Grown in Pine Bark Amended Soils with Varying Rates of Nitrates [Journal Article]. Communications in Soil Science and Plant Analysis. 2017;48(8):878–885.

[131] Tertuliano M, Krewer G, Smith JE, Plattner K, Clark J, Jacobs J, et al. Growing Organic Rabbiteye Blueberries in Georgia, USA: Results of Two Multi-Year Field Studies [Journal Article]. International Journal of Fruit Science. 2012;12(1-3):205–215.

[132] Takeda F, Krewer G, Andrews EL, Mullinix B, Peterson DL. Assessment of the v45 blueberry harvester on rabbiteye blueberry and southern highbush blueberry pruned to V-shaped canopy [Journal Article]. HortTechnology. 2008;18(1):130–138.

[133] Takeda F, Yang W, Li C, Freivalds A, Sung K, Xu R, et al. Applying New Technologies to Transform Blueberry Harvesting [Journal Article]. Agronomy. 2017;7(2):33.

[134] Yu P, Li C, Takeda F, Krewer G, Rains G, Hamrita T. Measurement of mechanical impacts created by rotary, slapper, and sway blueberry mechanical harvesters [Journal Article]. Computers and Electronics in Agriculture. 2014;101:84–92.

[135] Brown GK, Schulte NL, Timm EJ, Beaudry RM, Peterson DL, Hancock JF, et al. Estimates of mechanization effects on fresh blueberry quality [Journal Article]. Applied Engineering in Agriculture. 1996;12(1):21–26.

[136] van Dalfsen KB, Gaye MM. Yield from hand and mechanical harvesting of highbush blueberries in British Columbia [Journal Article]. Applied Engineering in Agriculture. 1999;15(5):393–398.

[137] Gehan MA, Fahlgren N, Abbasi A, Berry JC, Callen ST, Chavez L, et al. PlantCV v2: Image analysis software for high-throughput plant phenotyping [Journal Article]. PeerJ. 2017;5.

[138] Hartmann A, Czauderna T, Hoffmann R, Stein N, Schreiber F. HTPheno: An image analysis pipeline for high-throughput plant phenotyping [Journal Article]. BMC Bioinformatics. 2011;12.

[139] Knecht AC, Campbell MT, Caprez A, Swanson DR, Walia H. Image Harvest: an open-source platform for high-throughput plant image processing and analysis [Journal Article]. Journal of Experimental Botany. 2016;67(11):3587–3599.

[140] Reuzeau C, Pen J, Frankard V, de Wolf J, Peerbolte R, Broekaert W, et al. TraitMill: a discovery engine for identifying yield-enhancement genes in cereals [Journal Article]. Plant Gene and Trait. 2010;1.

[141] Gibbs JA, Pound M, French AP, Wells DM, Murchie E, Pridmore T. Approaches to three-dimensional reconstruction of plant shoot topology and geometry [Journal Article]. Functional Plant Biology. 2017;44(1):62–75.

[142] Polo JRR, Sanz R, Llorens J, Arno J, Escola A, Ribes-Dasi P, et al. A tractor-mounted scanning LIDAR for the non-destructive measurement of vegetative volume and surface area of tree-row plantations: A comparison with conventional destructive measurements [Journal Article]. Biosystems Engineering. 2009;102(2):128–134.

[143] Rosell JR, Llorens J, Sanz R, Arno J, Ribes-Dasi M, Masip J, et al. Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LIDAR scanning [Journal Article]. Agricultural and Forest Meteorology. 2009;149(9):1505–1515.

[144] Weiss M, Baret F. Using 3D Point Clouds Derived from UAV RGB Imagery to Describe Vineyard 3D Macro- Structure [Journal Article]. Remote Sensing. 2017;9(2).

[145] Wu C, He L, Du X, Chen S, Ni K. 3D reconstruction of Chinese hickory tree for dynamics analysis [Journal Article]. Biosystems Engineering. 2014;119:69–79.

[146] Diaz-Varela RA, de la Rosa R, Leon L, Zarco-Tejada PJ. High-Resolution Airborne UAV Imagery to Assess Olive Tree Crown Parameters Using 3D Photo Reconstruction: Application in Breeding Trials [Journal Article]. Remote Sensing. 2015;7(4):4213–4232.

[147] Underwood JP, Hung C, Whelan B, Sukkarieh S. Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors [Journal Article]. Computers and Electronics in Agriculture. 2016;130:83–96.

[148] Mu Y, Fujii Y, Takata D, Zheng B, Noshita K, Honda K, et al. Characterization of peach tree crown by using high-resolution images from an unmanned aerial vehicle [Journal Article]. Horticulture Research. 2018;5.

[149] Patrick A, Li C. High Throughput Phenotyping of Blueberry Bush Morphological Traits Using Unmanned Aerial Systems [Journal Article]. Remote Sensing. 2017;9(12).

[150] Herrero-Huerta M, Gonzalez-Aguilera D, Rodriguez-Gonzalvez P, Hernandez-Lopez D. Vineyard yield estimation by automatic 3D bunch modelling in field conditions [Journal Article]. Computers and Electronics in Agriculture. 2015;110:17–26.

[151] Jay S, Rabatel G, Hadoux X, Moura D, Gorretta N. In-field crop row phenotyping from 3D modeling performed using Structure from Motion [Journal Article]. Computers and Electronics in Agriculture. 2015;110:70–77.

[152] Snavely N, Seitz SM, Szeliski R. Modeling the world from Internet photo collections [Journal Article]. International Journal of Computer Vision. 2008;80(2):189–210.

[153] Bosse M, Zlot R, Flick P. Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping [Journal Article]. IEEE Transactions on Robotics. 2012;28(5):1104–1119.

[154] Bauwens S, Bartholomeus H, Calders K, Lejeune P. Forest Inventory with Terrestrial LiDAR: A Comparison of Static and Hand-Held Mobile Laser Scanning [Journal Article]. Forests. 2016;7(6).

[155] Marselis SM, Yebra M, Jovanovic T, van Dijk AIJM. Deriving comprehensive forest structure information from mobile laser scanning observations using automated

point cloud classification [Journal Article]. Environmental Modelling & Software. 2016;82:142–151.

[156] Costa C, Antonucci F, Pallottino F, Aguzzi J, Sun DW, Menesatti P. Shape Analysis of Agricultural Products: A Review of Recent Research Advances and Potential Application to Computer Vision [Journal Article]. Food and Bioprocess Technology. 2011;4(5):673–692.

[157] Gonzalo MJ, Brewer MT, Anderson C, Sullivan D, Gray S, van der Knaap E. Tomato fruit shape analysis using morphometric and morphology attributes implemented in Tomato Analyzer software program [Journal Article]. Journal of the American Society for Horticultural Science. 2009;134(1):77–87.

[158] Rodrguez GR, Moyseenko JB, Robbins MD, Huarachi Morejn N, Francis DM, van der Knaap E. Tomato Analyzer: A Useful Software Application to Collect Accurate and Detailed Morphological and Colorimetric Data from Two-dimensional Objects [Journal Article]. Journal of Visualized Experiments. 2010;(37):1856.

[159] Hurtado M, Vilanova S, Plazas M, Gramazio P, Herraiz FJ, Andujar I, et al. Phenomics of fruit shape in eggplant (Solanum melongena L.) using Tomato Analyzer software [Journal Article]. Scientia Horticulturae. 2013;164:625–632.

[160] Klein LL, Caito M, Chapnick C, Kitchen C, O'Hanlon R, Chitwood DH, et al. Digital Morphometrics of Two North American Grapevines (Vitis: Vitaceae) Quantifies Leaf Variation between Species, within Species, and among Individuals [Journal Article]. Frontiers in Plant Science. 2017;8.

[161] Hill TA, Chunthawodtiporn J, Ashrafi H, Stoffel K, Weir A, Van Deynze A. Regions Underlying Population Structure and the Genomics of Organ Size Determination in Capsicum annuum [Journal Article]. Plant Genome. 2017;10(3).

[162] Derbidge R, Feiten L, Conradt O, Heusser P, Baumgartner S. Assessment of Shape Changes of Mistletoe Berries: A New Software Approach to Automatize the Parameterization of Path Curve Shaped Contours [Journal Article]. PLoS One. 2013;8(4).

[163] Andres RJ, Coneva V, Frank MH, Tuttle JR, Samayoa LF, Han SW, et al. Modifications to a LATE MERISTEM IDENTITY1 gene are responsible for the major leaf shapes of Upland cotton (Gossypium hirsutum L.) [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(1):E57–E66.

[164] Costa C, Menesatti P, Paglia G, Pallottino F, Aguzzi J, Rimatori V, et al. Quantitative evaluation of Tarocco sweet orange fruit shape using optoelectronic elliptic Fourier based analysis [Journal Article]. Postharvest Biology and Technology. 2009;54(1):38–47.

[165] Goto S, Iwata H, Shibano S, Ohya K, Suzuki A, Ogawa H. Fruit shape variation in Fraxinus mandshurica var. japonica characterized using elliptic Fourier descriptors and the effect on flight duration [Journal Article]. Ecological Research. 2005;20(6):733–738.

[166] Maeda H, Akagi T, Tao R. Quantitative characterization of fruit shape and its differentiation pattern in diverse persimmon (Diospyros kaki) cultivars [Journal Article]. Scientia Horticulturae. 2018;228:41–48.

[167] Liao F, Peng J, Chen R. LeafletAnalyzer, an Automated Software for Quantifying, Comparing and Classifying Blade and Serration Features of Compound Leaves during Development, and among Induced Mutants and Natural Variants in the Legume Medicago truncatula [Journal Article]. Frontiers in Plant Science. 2017;8.

[168] Almon C. Path curves, an introduction to the work of L. Edwards on bud forms [Journal Article]. Open Systems & Information Dynamics. 1994;2(3):265–277.

[169] Torr PHS, Zisserman A. MLESAC: A new robust estimator with application to estimating image geometry [Journal Article]. Computer Vision and Image Understanding. 2000;78(1):138–156.

[170] Gough RE. The highbush blueberry and its management. CRC Press; 1993.

[171] Takeda F, Krewer G, Li C, MacLean D, Olmstead JW. Techniques for Increasing Machine Harvest Efficiency in Highbush Blueberry [Journal Article]. HortTechnology. 2013;23(4):430–436.

[172] USDA-ERS. Cotton and wool: Overview [Web Page]; 2017. Https://www.ers.usda.gov/topics/crops/cotton-wool/. April 7, 2019 Accessed.

[173] Monteith JL. Solar radiation and productivity in tropical ecosystems [Journal Article]. Journal of Applied Ecology. 1972;9(3):747–766.

[174] Fischer R, Edmeades GO. Breeding and cereal yield progress [Journal Article]. Crop Science. 2010;50(Supplement1):S–85–S–98.

[175] Meredith WR, Wells R. Potential for increasing cotton yields through enhanced partitioning to reproductive structures [Journal Article]. Crop Science. 1989;29(3):636–639.

[176] Wells R. The Use of Obsolete and Modern Cultivars to Examine Advances in Yield and Dry Matter Partitioning [Journal Article]. Journal of Cotton Science. 2016;20(3):246–252.

[177] Constable GA, Bange MP. The yield potential of cotton (Gossypium hirsutum L.) [Journal Article]. Field Crops Research. 2015;182:98–106.

[178] Murchie EH, Lawson T. Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications [Journal Article]. Journal of Experimental Botany. 2013;64(13):3983–3998.

[179] Damm A, Elbers J, Erler A, Gioli B, Hamdi K, Hutjes R, et al. Remote sensing of sun-induced fluorescence to improve modeling of diurnal courses of gross primary production (GPP) [Journal Article]. Global Change Biology. 2010;16(1):171–186.

[180] Baker NR. Chlorophyll fluorescence: a probe of photosynthesis in vivo [Journal Article]. Annual Review of Plant Biology. 2008;59:89–113.

[181] Van der Tol C, Verhoef W, Rosema A. A model for chlorophyll fluorescence and photosynthesis at leaf scale [Journal Article]. Agricultural and Forest Meteorology. 2009;149(1):96–105.

[182] Bolhar-Nordenkampf H, Long S, Baker N, Oquist G, Schreiber U, Lechner E. Chlorophyll fluorescence as a probe of the photosynthetic competence of leaves in the field: a review of current instrumentation [Journal Article]. Functional Ecology. 1989;p. 497–514.

[183] Kolber Z, Klimov D, Ananyev G, Rascher U, Berry J, Osmond B. Measuring photosynthetic parameters at a distance: laser induced fluorescence transient (LIFT) method for remote measurements of photosynthesis in terrestrial vegetation [Journal Article]. Photosynthesis Research. 2005;84(1-3):121–129.

[184] Schreiber U, Khl M, Klimant I, Reising H. Measurement of chlorophyll fluorescence within leaves using a modified PAM fluorometer with a fiber-optic microprobe [Journal Article]. Photosynthesis Research. 1996;47(1):103–109.

[185] Beer S, Bjrk M. Measuring rates of photosynthesis of two tropical seagrasses by pulse amplitude modulated (PAM) fluorometry [Journal Article]. Aquatic Botany. 2000;66(1):69–76.

[186] Zarco-Tejada PJ, Miller JR, Mohammed GH, Noland TL. Chlorophyll fluorescence effects on vegetation apparent reflectance: I. Leaf-level measurements and model simulation [Journal Article]. Remote Sensing of Environment. 2000;74(3):582–595.

[187] Sakowska K, Alberti G, Genesio L, Peressotti A, Delle Vedove G, Gianelle D, et al. Leaf and canopy photosynthesis of a chlorophyll deficient soybean mutant [Journal Article]. Plant, Cell & Environment. 2018;41(6):1427–1437.

[188] Campbell PKE, Huemmrich KF, Middleton EM, Ward LA, Julitta T, Daughtry CST, et al. Diurnal and Seasonal Variations in Chlorophyll Fluorescence Associated with Photosynthesis at Leaf and Canopy Scales [Journal Article]. Remote Sensing. 2019;11(5):488.

[189] Vilfan N, van der Tol C, Verhoef W. Estimating photosynthetic capacity from leaf reflectance and chlorophyll fluorescence by coupling radiative transfer to a model for photosynthesis [Journal Article]. New Phytologist. 2019;223(1).

[190] Cendrero-Mateo MP, Moran MS, Papuga SA, Thorp K, Alonso L, Moreno J, et al. Plant chlorophyll fluorescence: active and passive measurements at canopy and leaf scales with different nitrogen treatments [Journal Article]. Journal of Experimental Botany. 2015;67(1):275–286.

[191] Meroni M, Rossini M, Guanter L, Alonso L, Rascher U, Colombo R, et al. Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications [Journal Article]. Remote Sensing of Environment. 2009;113(10):2037–2051.

[192] Ustin SL, Gitelson AA, Jacquemoud S, Schaepman M, Asner GP, Gamon JA, et al. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy [Journal Article]. Remote Sensing of Environment. 2009;113:S67–S77.

[193] S A, Hank T, Mauser W. Deriving diurnal variations in sun-induced chlorophyll-a fluorescence in winter wheat canopies and maize leaves from ground-based hyperspectral measurements [Journal Article]. International Journal of Remote Sensing. 2016;37(sup1):60–77.

[194] Xu S, Liu Z, Zhao L, Zhao H, Ren S. Diurnal Response of Sun-Induced Fluorescence and PRI to Water Stress in Maize Using a Near-Surface Remote Sensing Platform [Journal Article]. Remote Sensing. 2018;10(10):1510.

[195] Guanter L, Zhang YG, Jung M, Joiner J, Voigt M, Berry JA, et al. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(14):E1327–E1333.

[196] Sun Y, Frankenberg C, Wood JD, Schimel DS, Jung M, Guanter L, et al. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence [Journal Article]. Science. 2017;358(6360):eaam5747.

[197] Liu J, Bowman KW, Schimel DS, Parazoo NC, Jiang Z, Lee M, et al. Contrasting carbon cycle responses of the tropical continents to the 20152016 El Nio [Journal Article]. Science. 2017;358(6360):eaam5690.

[198] Guan K, Berry JA, Zhang Y, Joiner J, Guanter L, Badgley G, et al. Improving the monitoring of crop productivity using spaceborne solarinduced fluorescence [Journal Article]. Global Change Biology. 2016;22(2):716–726.

[199] Moreno J, Colombo R, Damm A, Goulas Y, Middleton E, Miglietta F, et al. Quantitative global mapping of terrestrial vegetation photosynthesis: The Fluorescence Explorer (FLEX) mission. In: IEEE International Geoscience and Remote Sensing Symposium; 2017. p. 435–438.

[200] Rossini MD, Nedbal L, Guanter L, A A, Alonso L, Burkart A, et al. Red and far red Sun-induced chlorophyll fluorescence as a measure of plant photosynthesis [Journal Article]. Geophysical Research Letters. 2015;42(6):1632–1639.

[201] Pinto F, Damm A, Schickling A, Panigada C, Cogliati S, MllerLinow M, et al. Sun-induced chlorophyll fluorescence from high-resolution imaging spectroscopy data to

quantify spatio-temporal patterns of photosynthetic function in crop canopies [Journal Article]. Plant, Cell & Environment. 2016;39(7):1500–1512.

[202] Earl HJ, Ennahli S. Estimating photosynthetic electron transport via chlorophyll fluorometry without Photosystem II light saturation [Journal Article]. Photosynthesis Research. 2004;82(2):177–186.

[203] Wise RR, Sassenrath-Cole GF, Percy RG. A Comparison of Leaf Anatomy in Field-grown Gossypium hirsutum and G. barbadense [Journal Article]. Annals of Botany. 2000;86(4):731–738.

[204] Whitaker J. Cotton irrigation, defoliation, contamination [Online Multimedia]; 2016. Accessed: March 28, 2019. Available from: `http://www.ugacotton.com/vault/file/2016-UGA-Cotton-Production-Guide.pdf`.

[205] Jiang Y, Li C, Robertson JS, Sun S, Xu R, Paterson AH. GPhenoVision: a ground mobile system with multi-modal imaging for field-based high throughput phenotyping of cotton [Journal Article]. Scientific Reports. 2018;8(1):1213.

[206] Plascyk JA. The MK II Fraunhofer line discriminator (FLD-II) for airborne and orbital remote sensing of solar-stimulated luminescence [Journal Article]. Optical Engineering. 1975;14(4):144339.

[207] Flexas J, Badger M, Chow WS, Medrano H, Osmond CB. Analysis of the relative increase in photosynthetic O2 uptake when photosynthesis in grapevine leaves is inhibited following low night temperatures and/or water stress [Journal Article]. Plant Physiology. 1999;121(2):675–684.

[208] Maxwell K, Johnson GN. Chlorophyll fluorescencea practical guide [Journal Article]. Journal of Experimental Botany. 2000;51(345):659–668.

[209] White AJ, Critchley C. Rapid light curves: a new fluorescence method to assess the state of the photosynthetic apparatus [Journal Article]. Photosynthesis Research. 1999;59(1):63–72.

[210] Ralph PJ, Gademann R. Rapid light curves: A powerful tool to assess photosynthetic activity [Journal Article]. Aquatic Botany. 2005;82(3):222–237.

[211] Ehleringer JR, Hammond SD. Solar Tracking and Photosynthesis in Cotton Leaves [Journal Article]. Agricultural and Forest Meteorology. 1987;39(1):25–35.

[212] Team RDC. R: A Language and Environment for Statistical Computing [Computer Program]; 2008. Http://www.R-project.org. Accessed.

[213] Henley WJ. Measurement and interpretation of photosynthetic lightresponse curves in algae in the context of photoinhibition and diel changes [Journal Article]. Journal of Phycology. 1993;29(6):729–739.

[214] Wise RR, Sassenrath-Cole GF, Percy RG. A comparison of leaf anatomy in field-grown Gossypium hirsutum and G-barbadense [Journal Article]. Annals of Botany. 2000;86(4):731–738.

[215] Yang X, Tang J, Mustard JF, Lee JE, Rossini M, Joiner J, et al. Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest [Journal Article]. Geophysical Research Letters. 2015;42(8):2977–2987.

[216] Liu Z, Wu C, Peng D, Wang S, Gonsamo A, Fang B, et al. Improved modeling of gross primary production from a better representation of photosynthetic components in vegetation canopy [Journal Article]. Agricultural and Forest Meteorology. 2017;233:222–234.

[217] Scholes JD, Rolfe SA. Chlorophyll fluorescence imaging as tool for understanding the impact of fungal diseases on plant performance: a phenomics perspective [Journal Article]. Functional Plant Biology. 2009;36(10-11):880–892.

[218] Atta BM, Saleem M, Ali H, Arshad HMI, Ahmed M. Chlorophyll as a biomarker for early disease diagnosis [Journal Article]. Laser Physics. 2018;28(6).

[219] Cen H, Weng H, Yao J, He M, Lv J, Hua S, et al. Chlorophyll Fluorescence Imaging Uncovers Photosynthetic Fingerprint of Citrus Huanglongbing [Journal Article]. Frontiers in Plant Science. 2017;8.

[220] Corp L, Middleton EM, Cheng YB, Campbell PE, Huemmrich KF. Impact of spectral resolution on solar induced fluorescence and reflectance indices for monitoring vegetation. In: IEEE International Geoscience and Remote Sensing Symposium. vol. 4. IEEE; 2008. p. 1387 –1390.

[221] Zhi X, Han Y, Li Y, Wang G, Du W, Li X, et al. Effects of plant density on cotton yield components and quality [Journal Article]. Journal of Integrative Agriculture. 2016;15(7):1469–1479.

[222] Adams C, Thapa S, Kimura E. Determination of a plant population density threshold for optimizing cotton lint yield: A synthesis [Journal Article]. Field Crops Research. 2019;230:11–16.

[223] Zhai L, Xie R, Ming B, Li S, Ma D. Evaluation and analysis of intraspecific competition in maize: A case study on plant density experiment [Journal Article]. Journal of Integrative Agriculture. 2018;17(10):2235–2244.

[224] Xue H, Han Y, Li Y, Wang G, Feng L, Fan Z, et al. Spatial distribution of light interception by different plant population densities and its relationship with yield [Journal Article]. Field Crops Research. 2015;184:17–27.

[225] Williams MM. Agronomics and economics of plant population density on processing sweet corn [Journal Article]. Field Crops Research. 2012;128:55–61.

[226] Chen R, Chu T, Landivar JA, Yang C, Maeda MM. Monitoring cotton (Gossypium hirsutum L.) germination using ultrahigh-resolution UAS images [Journal Article]. Precision Agriculture. 2017;.

[227] Gnadinger F, Schmidhalter U. Digital Counts of Maize Plants by Unmanned Aerial Vehicles (UAVs) [Journal Article]. Remote Sensing. 2017;9(6).

[228] Zhao B, Zhang J, Yang C, Zhou G, Ding Y, Shi Y, et al. Rapeseed Seedling Stand Counting and Seeding Performance Evaluation at Two Early Growth Stages Based on Unmanned Aerial Vehicle Imagery [Journal Article]. Frontiers in Plant Science. 2018;9.

[229] Liu X, Chen SW, Liu C, Shivakumar SS, Das J, Taylor CJ, et al.. Monocular Camera Based Fruit Counting and Mapping with Semantic Data Association [Online Multimedia]. https://arxiv.org/abs/1811.01417. Accessed 1 Feb 2019.; 2018.

[230] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI Conference on Artificial Intelligence. vol. 4. AAAI; 2017. p. 12.

[231] Kalman RE. A new approach to linear filtering and prediction problems [Journal Article]. Journal of Basic Engineering. 1960;82(1):35–45.

[232] Welch G, Bishop G. An Introduction to the Kalman Filter [Online Multimedia]. https://www.cs.unc.edu/welch/media/pdf/kalman_intro.pdf. Accessed 1 Feb 2019.; 2006.

[233] Kuhn HW. The Hungarian method for the assignment problem [Journal Article]. Naval Research Logistics. 1955;2(1-2):83–97.

[234] Chen C, Liu MY, Tuzel O, Xiao J. R-CNN for small object detection. In: Asian Conference on Computer Vision. Springer; 2016. p. 214–230.

[235] Andres F, Coupland G. The genetic basis of flowering responses to seasonal cues [Journal Article]. Nature Reviews Genetics. 2012;13(9):627–639.

[236] Bernier G, Perilleux C. A physiological overview of the genetics of flowering time control [Journal Article]. Plant Biotechnology Journal. 2005;3(1):3–16.

[237] Putterill J, Laurie R, Macknight R. It's time to flower: the genetic control of flowering time [Journal Article]. Bioessays. 2004;26(4):363–373.

[238] Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The Genetic Architecture of Maize Flowering Time [Journal Article]. Science. 2009;325(5941):714–718.

[239] Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces [Journal Article]. Nature Genetics. 2010;42(11):961–U76.

[240] Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, et al. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity [Journal Article]. Journal of Experimental Botany. 2007;58(6):1231–1244.

[241] Bhosale SU, Stich B, Rattunde HFW, Weltzien E, Haussmann BIG, Hash CT, et al. Association analysis of photoperiodic flowering time genes in west and central African sorghum [Sorghum bicolor (L.) Moench] [Journal Article]. BMC Plant Biology. 2012;12.

[242] Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses [Journal Article]. Behavior Research Methods. 2009;41(4):1149–1160.

[243] Cordon GB, Lagorio MG. Optical properties of the adaxial and abaxial faces of leaves. Chlorophyll fluorescence, absorption and scattering coefficients [Journal Article]. Photochemical & Photobiological Sciences. 2007;6(8):873–882.

[244] Statistics Division FAO of the United Nations. Crops Production [Web Page]; 2013. Available from: http://faostat3.fao.org/browse/Q/QC/E.

[245] National Agricultural Statistics Service, United States Department of Agriculture. National Statistics for Blueberries [Web Page]; 2014. Available from: http://www.nass.usda.gov/Statistics_by_Subject/result.php?B31557DC-DDD9-3C5D-9FA2-C7F176ADE896&sector=CROPS&group=FRUIT%20%26%20TREE%20NUTS&comm=BLUEBERRIES.

[246] Yu P, Li C, Takeda F, Krewer G, Rains G, Hamrita T. Quantitative evaluation of a rotary blueberry mechanical harvester using a miniature instrumented sphere [Journal Article]. Computers and Electronics in Agriculture. 2012;88:25–31.

[247] Moretti CL, Sargent SA, Huber DJ, Calbo AG, Puschmann R. Chemical composition and physical properties of pericarp, locule, and placental tissues of tomatoes with internal bruising [Journal Article]. Journal of the American Society for Horticultural Science. 1998;123(4):656–660.

[248] Opara UL, Pathare PB. Bruise damage measurement and analysis of fresh horticultural produce-A review [Journal Article]. Postharvest Biology and Technology. 2014;91:9–24.

[249] Mohsenin NN. Physical properties of plant and animal materials : structure, physical characteristics, and mechanical properties. 2nd ed. New York: Gordon and Breach; 1986.

[250] Schotte S, De Belie N, De Baerdemaeker J. Acoustic impulse-response technique for evaluation and modelling of firmness of tomato fruit [Journal Article]. Postharvest Biology and Technology. 1999;17(2):105–115.

[251] Golias J, Bejcek L, Graetz P, Klusacek S. Mechanical resonance method for evaluation of peach fruit firmness [Journal Article]. Horticultural Science. 2003;30(1):1–6.

[252] Prussia SE, Astleford JJ, Hewlett B, Hung YC. Non-destructive firmness measuring device [Generic]. Google Patents; 1994. Available from: http://www.google.com/patents/US5372030.

[253] Li C, Luo J, MacLean D. A novel instrument to delineate varietal and harvest effects on blueberry fruit texture during storage [Journal Article]. Journal of the Science of Food and Agriculture. 2011;91(9):1653–1658.

[254] Shahin MA, Tollner EW, McClendon RW, Arabnia HR. Apple classification based on surface bruises using image processing and neural networks [Journal Article]. Transactions of the ASAE. 2002;45(5):1619–1627.

[255] Mazhar M, Joyce D, Cowin G, Brereton I, Hofman P, Collins R, et al. Non-destructive 1H-MRI assessment of flesh bruising in avocado (Persea americana M.) cv. Hass [Journal Article]. Postharvest Biology and Technology. 2015;100:33–40.

[256] Baranowski P, Mazurek W, Wozniak J, Majewska U. Detection of early bruises in apples using hyperspectral data and thermal imaging [Journal Article]. Journal of Food Engineering. 2012;110(3):345–355.

[257] Wu D, Sun D. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part II: Applications [Journal Article]. Innovative Food Science & Emerging Technologies. 2013;19:15–28.

[258] Lu R. Detection of bruises on apples using near-infrared hyperspectral imaging [Journal Article]. Transactions of the ASAE. 2003;46(2):523–530.

[259] Baranowski P, Mazurek W, Pastuszka-Wozniak J. Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data [Journal Article]. Postharvest Biology and Technology. 2013;86:249–258.

[260] Lee WH, Kim MS, Lee H, Delwiche SR, Bae H, Kim DY, et al. Hyperspectral near-infrared imaging for the detection of physical damages of pear [Journal Article]. Journal of Food Engineering. 2014;130:1–7.

[261] Wu L, He J, Liu G, Wang S, He X. Detection of common defects on jujube using Vis-NIR and NIR hyperspectral imaging [Journal Article]. Postharvest Biology and Technology. 2016;112:134–142.

[262] Chaplin M. Water Absorption Spectrum [Web Page]; 2000. Available from: `http://www1.lsbu.ac.uk/water/water_vibrational_spectrum.html`.

[263] Park B, Lu R. Chapter 5: Classification and Prediction Methods. In: Hyperspectral Imaging Technology in Food and Agriculture. New York: Springer; 2015. p. 103–124.

[264] Leiva-Valenzuela GA, Lu R, Aguilera JM. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging [Journal Article]. Journal of Food Engineering. 2013;115(1):91–98.

[265] Leiva-Valenzuela GA, Lu R, Aguiler JM. Assessment of internal quality of blueberries using hyperspectral transmittance and reflectance images with whole spectra or selected wavelengths [Journal Article]. Innovative Food Science & Emerging Technologies. 2014;24:2–13.

[266] Wang W, Li C, Tollner EW, Rains GC, Gitaitis RD. A liquid crystal tunable filter based shortwave infrared spectral imaging system: Design and integration [Journal Article]. Computers and Electronics in Agriculture. 2012;80:126–134.

[267] Chang C, Lin C. LIBSVM: A Library for Support Vector Machines [Journal Article]. ACM Transactions on Intelligent Systems and Technology. 2011;2(3).

[268] Jiang Y, Li C. Non-destructive Detection of Internal Bruising of Blueberries using Hyperspectral Imaging. New Orleans, LA, USA: ASABE Annual International Meeting, Paper No: 152189892; 2015. .

[269] Lammertyn J, Dresselaers T, Van Hecke P, Jancsok P, Wevers M, Nicolai BM. Analysis of the time course of core breakdown in 'Conference' pears by means of MRI and X-ray CT [Journal Article]. Postharvest Biology and Technology. 2003;29(1):19–28.