Innovative Deep Learning Based Techniques in Structural Inspection and Biometric Authentication

by

Collins Rakowski

(Under the Direction of Thirimachos Bourlai)

Abstract

Structural health monitoring and biometric security are critical domains benefiting from AI and computer vision. This thesis tackles two challenges: generating synthetic data to improve crack detection and enhancing dorsal hand vein recognition through knuckle alignment.

For structural crack detection, data scarcity limits AI model effectiveness. Using StyleGAN3 and Brownian Bridge Diffusion Models (BBDM), we generate synthetic crack images with various blur effects to simulate real-world conditions. Integrating this data with hyperparameter-tuned DeepLabV3 achieves 65.62% MeanIoU on the Bridge Crack Library dataset, setting a new benchmark.

In dorsal hand vein biometrics, a novel knuckle alignment method enhances vein recognition accuracy, achieving 99.07% on the Jilin University dataset and 99.90% on the Wilches dataset. This study also assesses system robustness under image degradation. Results show that alignment significantly improves performance under moderate blurring but loses effectiveness with severe degradation.

Index words: [Structural Crack Detection, Synthetic Data Generation, Generative Adversarial Networks (GANs), Brownian Bridge Diffusion Model (BBDM), Dorsal Hand Vein Biometrics, Knuckle Alignment, Deep Learning, Image Degradation, Semantic Segmentation]

Innovative Deep Learning Based Techniques in Structural Inspection and Biometric Authentication

by

Collins Rakowski

B.S. in Computer Systems Engineering, University of Georgia, 2023

A Thesis Submitted to the Graduate Faculty of the

University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Master of Science

Athens, Georgia

2024

Innovative Deep Learning Based Techniques in Structural Inspection and Biometric Authentication

by

Collins Rakowski

Major Professor: Thirimachos Bourlai

Committee: Mable Fok

Kyle Johnsen

Electronic Version Approved:

Ron Walcott

Dean of the Graduate School

The University of Georgia

December 2024

Dedication

This work is dedicated to my family and closest friends, whose unwavering belief in me has been my greatest source of strength. Your patience, encouragement, and constant support have made this journey possible. I am deeply grateful to each of you for standing by me every step of the way.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to Dr. Thirimachos Bourlai. Your encouragement and steadfast belief in my abilities guided me toward discovering my passion for deep learning and artificial intelligence. Without your mentorship, I would not be where I am today.

I would also like to thank Victor Phillippe for helping me take my first steps into academic research, guiding me through the process of my first journal publication, and providing invaluable insights that laid the groundwork for this thesis.

A special thank you to Suha, whose guidance has been a beacon of inspiration as I navigated the complexities of AI and began my professional journey in this field.

Finally, I would like to express my appreciation to Dylan Meyer, a true partner in intellectual curiosity and conversation. Your willingness to entertain our wild ideas has always kept my imagination alive and has been a source of joy throughout this journey.

To all those mentioned and many others who supported me along the way, my sincere thanks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Artificial Intelligence (AI) has rapidly evolved into a cornerstone of modern technology, fundamentally transforming industries, offering innovative solutions to complex problems through advanced data processing and pattern recognition. Its rapid development has been propelled by the need to interpret vast amounts of data and make accurate predictions, leading to breakthroughs in fields such as computer vision, natural language processing, and autonomous systems.

In the domain of computer vision, AI has significantly advanced tasks such as classification, object detection, segmentation, and image generation. These advancements have led to practical applications that have reshaped various sectors. For instance, in infrastructure management, AI-powered computer vision systems enable the automatic detection and analysis of structural defects from images or video feeds. By automating the inspection process, AI not only speeds up the detection of potential issues but also reduces the risk and cost associated with manual inspections, which is critical for maintaining the safety and integrity of critical infrastructure like bridges, buildings, and pipelines.

Generative models like Generative Adversarial Networks (GANs) and diffusion models have revolutionized content creation by generating realistic images, videos, and synthetic datasets. Besides content creation for media purposes, there are many real world applications of generative AI. In applications where real-world data is scarce or sensitive, such as medical imaging or structural inspection, synthetic data generated by GANs enhances training datasets for machine learning models, improving their performance.

Advancements in AI have led to the development of better segmentation and object detection models, making it possible to tackle more complex tasks like real time tracking, security, and inspections. Segmentation and object detection models have also seen substantial progress due to AI advancements. These models enable precise identification and classification of objects within images, which is essential in applications like structural inspection. Accurate detection of defects or anomalies in infrastructure components can prevent catastrophic failures and ensure public safety. AI-driven segmentation models facilitate detailed analysis of structural elements, aiding in proactive maintenance and risk management.

The development of these advanced machine learning models, GANs, convolutional neural networks (CNNs), and segmentation architectures, has been instrumental in tackling specific challenges across various domains. By tailoring solutions to address issues like data scarcity, object identification, and biometric analysis, AI continues to push the boundaries of technological capabilities.

In this thesis, we explore the expansive capabilities of AI by delving into several advanced models, including GANs, diffusion models, Convolutional Neural Networks (CNNs), and segmentation models. Our research is centered on two key industries that are built on the foundations of these models: structural inspection and biometrics.

The structural inspection industry plays a crucial role in maintaining the safety and integrity of critical infrastructure, including bridges, buildings, pipelines, and other essential structures. Traditionally, this industry has relied heavily on manual inspections, where trained professionals visually assess structures for signs of wear, damage, or other potential issues. While effective, these methods are often time-consuming, expensive, and, in some cases, hazardous to the inspectors involved. AI has introduced a new era of efficiency and accuracy in structural inspection. By leveraging advanced computer vision techniques, AI models can automatically detect and analyze structural defects, such as cracks or deformations, from images or video feeds. This not only speeds up the inspection process but also allows for continuous monitoring of infrastructure, enabling early detection of potential problems that might otherwise go unnoticed.

Biometrics, the science of using unique physiological traits to identify individuals, has become an integral part of modern security systems. From fingerprint and facial recognition systems that unlock our smartphones to advanced iris and vein pattern recognition systems used in high-security environments, biometrics offers a level of security and convenience that traditional identification methods cannot match.

A majority of this thesis focuses on advancing the field of biometric security, particularly in the area of dorsal hand vein recognition. Although less common than other biometric modalities, dorsal hand vein recognition offers distinct advantages, such as resistance to spoofing and long-term stability. We introduce innovative alignment techniques to improve accuracy and explore how image degradation affects performance, ensuring these systems remain reliable in real-world conditions.

What follows is an in-depth exploration of four key topics, where we examine the problem definitions, relevance, and contributions of our AI-driven approaches in addressing critical challenges within the fields of structural inspection and biometrics.

1.1 Image Generation for Improved Semantic Segmentation

1.1.1 Problem Definition

In structural inspection, the accurate detection of cracks is crucial for maintaining the integrity and safety of infrastructure, particularly in bridges. Traditional methods, such as visual inspections and non-destructive testing, are often limited by subjectivity, accessibility, and the extensive resources required for data collection. Deep learning models offer a promising alternative, yet they demand large, high-quality datasets that are often difficult and time-consuming to obtain. Moreover, the annotation process is labor-intensive and prone to inconsistencies, which can affect the accuracy of the model. To circumvent the challenge of gathering and annotating new data, we design a deep learning system to generate synthetic data that accurately reflects real-world conditions thereby enhancing the segmentation performance of deep learning models for bridge inspection.

1.1.2 Relevance

Structural safety is of paramount importance in the industrial community, particularly in the context of in-service bridges. Maintaining the integrity of these structures is essential to avert accidents and ensure the safety of both people and goods that depend on them. Among the various challenges in this field, the detection of structural cracks remains a significant issue. Early detection and continuous monitoring of cracks is largely a tedious and challenging task. Traditional approaches primarily involve visual inspections, where human inspectors manually examine bridge components for cracks and anomalies. Typically an inspector conducts Non-Destructive Testing (NDT) which includes techniques such as ultrasonic testing, radiography, magnetic particle inspection, acoustic emission testing, and vibration analysis—are also employed in this domain Kashif Ur Rehman et al., 2016.

All the aforementioned approaches are limited due to various reasons. Visual inspections, which depend on the subjective judgment and experience of inspectors, can lead to inconsistencies. Non-destructive testing techniques often struggle with coverage and accessibility, making thorough inspections difficult. These traditional annotation methods can be time-intensive, require substantial manual effort, be sensitive to environmental conditions, and may not be sufficiently sensitive to detect early-stage or hidden cracks. In contrast, image processing and deep learning present notable advantages, such as automation, objectivity, continuous monitoring, and adaptability, positioning them as promising alternatives or complementary solutions to conventional crack detection methods in bridge maintenance. However, despite the appeal of these advanced approaches, the task remains complex, requiring extensive data collection and precise annotation.

The traditional method for structural crack detection using deep learning often requires extensive data collection, which can be time-consuming and resource-intensive. These deep learning models have the capability to autonomously learn to detect cracks; however, they require large datasets to fully grasp the diverse features and patterns that characterize real-world cracks Ali et al., 2022. One key research paper has introduced a groundbreaking approach Jin et al., 2023. Jin et al. Jin et al., 2023 utilized a Deep Convolutional GAN (DCGAN) to produce synthesized crack annotations and employed a Pixel2Pixel model Isola et al., 2017 to generate the corresponding synthesized crack images. They carefully documented the learning trajectory of these GANs throughout the training process, highlighting how the models evolve to create synthesized images that closely resemble real-world crack patterns. Notably, they conducted a comparative analysis to evaluate the performance of Deep Neural Networks (DNNs) trained on this synthesized dataset. Their study examines various strategies for integrating synthesized and real crack images into DNN training, with a key finding that pre-training on synthesized images followed by fine-tuning with real images yields better results than directly combining both types of data. This methodology utilizes a diffusion model to generate a synthesized crack image dataset with pixel-level annotations to address the extensive data collection process. They are the first in this domain to overcome this challenge by creating new segmentation data. In Chapter 2, we draw inspiration from this approach to attack the common problem of data scarcity and segmentation performance in the structural crack inspection domain.

As we discuss in the literature review section in Chapter 2, there is extensive research showing the capabilities of DNNs in crack segmentation, but a gap remains for us to explore the capabilities of diffusion-based synthesized datasets.

Building upon the foundation laid by the aforementioned research, our work endeavors to take diffusion-based synthetic crack data generation and semantic segmentation to new heights through newer hyper-tuned models, augmentation techniques, and ensemble modeling. While the most recent paper Jin et al., 2023 made significant strides in introducing GANs for generating synthetic crack data and their application in segmentation, there are a set of limitations that we address. These limitations include limited hyperparameter tuning, the absence of performance ensemble modeling, insufficient augmentation studies, and a lack of exploration into diffusion-based synthetic image data generation approaches.

1.1.3 Contributions

Our approach introduces several notable advancements in the domain of structural crack detection using deep learning and synthetic data generation:

1. *Novel Approach to Synthetic Data Generation*: One of the primary contributions of this research is the integration of StyleGAN3 and the BBDM to generate synthetic crack images. By utilizing StyleGAN3, we produce mask annotations that deliver precise details about the crack regions. Concurrently, BBDM is employed to generate synthetic real images that correspond to these mask annotations. This innovative methodology effectively addresses the challenge of creating segmentation data.

2. *Enhanced Performance through Hyperparameter Tuning of DeepLabV3+*: This research highlights the significant role of hyperparameter optimization in enhancing the performance of the DeepLabV3+ model. We experimented with various hyperparameters, such as altering backbones and optimizers, to boost semantic segmentation accuracy. Through careful tuning of these hyperparameters, we achieved a 9% improvement in model accuracy.

3. *Blur Data Augmentation Technique in the Context of Synthesized Dataset*: The study introduces blur data augmentation techniques, including motion blur, zoom blur, and defocus blur, to emulate real-world variations in image quality to improve segmentation performance.

4. *Ensemble Modeling on Augmented Synthesized Datasets*: A key aspect of this research is the application of ensemble modeling techniques to models trained on augmented synthetic datasets. By

combining the strengths of multiple models through ensemble approaches, particularly majority voting, the study achieves significant improvements in MeanIoU accuracy.

5. *Diffusion-Based Synthetic Image Data Generation*: The use of BBDM for synthetic data generation represents a pioneering effort in this field. This research illustrates the potential of diffusion models to generate high-quality, diverse crack images, paving the way for broader applications of diffusion-based image generation in machine learning and semantic segmentation.

We propose a comprehensive approach that addresses the challenge of generating segmentation data for structural crack detection by leveraging ensemble learning. By integrating multiple models, each trained on distinct datasets that include both real and synthetically generated images, we substantially enhance the performance of crack detection systems. This ensemble strategy not only broadens the diversity of the training data but also harnesses the unique strengths of individual models, leading to improved accuracy and robustness.

Our results exhibit the highest recorded performance on the Bridge Crack Library (BCL) dataset Jin et al., 2021, achieving a MeanIoU of 65.62%, which establishes a new benchmark in structural crack detection. This achievement highlights the effectiveness of our method in addressing the persistent issue of limited data availability, a challenge that has long hindered the application of deep learning techniques in real-world structural safety. By leveraging multiple models trained on a synergy of real and augmented synthetic data, we provide a practical solution for applying deep learning in situations where high-quality, segmented data is scarce. Our approach not only advances the field of structural crack detection but also offers a framework for tackling similar challenges in other areas where data limitations exist.

1.2 Novel Alignment Method for Dorsal Hand Vein Biometrics

1.2.1 Problem Definition

Traditional alignment techniques, such as those commonly used in face recognition, are often inadequate when applied to dorsal hand vein biometrics due to the unique anatomical structure and variability in hand positioning. While alignment is a well-studied area in other biometric modalities, there is a significant gap in developing effective alignment methods tailored specifically for dorsal hand vein recognition. We address this gap by introducing a novel knuckle alignment technique that significantly improves the consistency and accuracy of dorsal hand vein biometrics, thus reducing intra-class variability and enhancing system performance.

1.2.2 Relevance

Dorsal hand vein biometrics offers unique advantages, such as the stability of vein patterns over time and resistance to common spoofing techniques due to the internal nature of the vein structures. While biometric modalities like facial recognition, fingerprinting, and iris scanning have been widely adopted and extensively researched, with significant advancements in image enhancement techniques, feature extraction methods, and alignment strategies, dorsal hand veins have not received the same level of attention. Despite the dorsal hand vein benefits, including non-intrusive imaging, high resistance to spoofing, and the potential for high accuracy in identification and verification, the specific anatomical and positional variability of hand veins requires tailored approaches to alignment and recognition that are not adequately addressed by the existing techniques developed for other biometric modalities.

Our work intends to bridge this gap by introducing a novel knuckle alignment method that improves the consistency and accuracy of dorsal hand vein biometrics by reducing variability in hand positioning.

1.2.3 Contributions

Our proposed work marks a new benchmark in dorsal hand vein verification and identification accuracies for two public dorsal hand vein datasets (Jillian University Dorsal Hand Vein Dataset and Wilchesf Dorsal Hand Vein Dataset). The main contributions proposed in this study are:

1. *Innovative Knuckle Alignment Method*: We introduce a computer vision-based technique for dorsal hand vein alignment and scaling. This method is tailored for images captured in near-infrared (NIR) conditions where subjects are either holding a bar or presenting a closed fist, addressing variability in hand positioning and orientation.

2. *Automated Versus Manual Knuckle Alignment*: We conduct a comprehensive comparison study between automated and manual-based knuckle alignment annotation methods. This comparison assesses the effectiveness in terms of accuracy of the proposed automated versus manual alignment process.

3. *High Accuracy of the Proposed Dorsal Hand Vein Authentication System*: Our study yields the highest verification accuracies and lowest EER scores on both the JLU and DHV datasets tested, and when compared to other reported approaches in the literature.

This research marks a new benchmark in Dorsal Hand Vein verification and identification accuracies for the DHV and JLU datasets.

Overall, our study introduces a novel knuckle alignment method that significantly enhances the scaling and alignment process for dorsal hand vein identification. Unlike conventional techniques that rely on extracting regions of interest (ROIs) for detailed analysis, our method eliminates the need for such extractions by directly utilizing the entire image. Our manual alignment method yields 100% and 99.05% verification accuracy on the DHV and JLU datasets, respectively. Our automated alignment method reaches a competitive verification accuracy of 99.1% and 98.78% also on the DHV and JLU datasets, respectively.

1.3 Image Degradations for Dorsal Hand Vein Biometrcs

1.3.1 Problem Definition

Image degradation, including factors like blurriness and contrast reduction, can severely impact the performance of biometric systems. While there has been extensive research on how these degradations affect face recognition systems, similar investigations into dorsal hand vein biometrics remain sparse. As mention prior, this lack of study is particularly concerning given the benefits of dorsal hand veins as a biometric.

Dorsal hand vein recognition systems rely heavily on the clarity of vein patterns captured through near-infrared (NIR) imaging. However, the visibility of these patterns can be compromised by physiological changes (e.g., increased subcutaneous fat), environmental factors, and various forms of image degradation.

This study addresses the robustness of dorsal hand vein systems by systematically investigating the effects of global and local blurring on dorsal hand vein recognition systems. By simulating both general blurring and specific obstructions to vein visibility, this research aims to provide a comprehensive understanding of how these factors influence system performance and to establish a baseline for future investigations in this area.

1.3.2 Relevance

As biometric systems become integral to critical security applications, including border control, financial transactions, and access to sensitive information, understanding how these systems perform under various conditions becomes increasingly essential. The robustness of such systems, particularly in the face of image degradation, such as blurriness, contrast reduction, and occlusions, directly impacts their reliability and trustworthiness. Dorsal hand vein recognition, which leverages the unique vascular patterns beneath the skin, is particularly valued for its inherent resistance to surface-level spoofing techniques. However, despite its potential, the vulnerabilities of this modality under degraded imaging conditions have not been thoroughly explored, leaving a critical gap in the literature.

Given the growing reliance on biometric verification for security tasks, it is imperative to investigate how these systems respond to real-world challenges, where perfect imaging conditions cannot always be guaranteed. The findings in this study investigate two image degradation techniques. First, we provide insights into the resilience of the biometric system under general blurring techniques. Secondly, we degrade the image quality of the image by concealing vein identities and study the models dependency on vein patterns versus overall hand shape.

1.3.3 Contributions

This study sets a new benchmark in the analysis of dorsal hand vein biometric systems under blurring conditions. The main contributions of this study are as follows:

1. *Simulating Physiological Changes:* We simulate physiological changes such as increased subcutaneous fat, which obscures vein visibility, using advanced inpainting techniques from computer vision and LaMa-based methods.

2. *Comprehensive Ablation Study on Blurring Effects*: We perform an in-depth ablation study that explores the impact of both global and local blurring on dorsal hand vein biometric systems. This study provides insights into how different levels of image degradation affect system performance, highlighting the resilience of biometric systems to varying degrees of vein visibility.

3. *Comparison of Traditional and Deep Learning-based Inpainting Techniques*: We compare the effectiveness of traditional computer vision-based inpainting methods and a state-of-the-art deep learning-based inpainting model (LaMa) in concealing vein patterns and compromising biometric accuracy.

Chapter 2

Literature Review

2.1 On Enhancing Crack Semantic Segmentation using StyleGAN and Brownian Bridge Diffusion

2.1.1 Network Tuning and Architecture

Structural crack detection is an ever-advancing industry that is built on semantic segmentation models. In this section, we explore model architectures and networks that contribute to the evolution of these semantic segmentation models within the domain of crack detection.

A widely recognized model in semantic segmentation is DeepLabV3+ L.-C. Chen et al., 2018, which excels in capturing contextual information and refining object boundaries. At its core, DeepLabV3+ utilizes an Atrous Spatial Pyramid Pooling (ASPP) module to extract features at multiple scales, allowing it to handle objects of varying sizes. Additionally, the model employs depthwise separable convolutions to reduce computational complexity without sacrificing accuracy. The decoder module further enhances boundary details, making DeepLabV3+ especially effective for tasks like crack segmentation, where precise edge detection is critical. Its ability to balance performance and efficiency has established DeepLabV3+ as a leading choice for structural crack detection tasks.

Enhancements to this model include the introduction of a densely connected ASPP variant Fu et al., 2021, further improving segmentation accuracy. Its adaptability has also been demonstrated by applying it to various crack detection domains J.-J. Wang et al., 2022.

Several studies have explored the impact of different backbone architectures within DeepLabV3+ to further optimize performance. For example, comparisons have been made of the segmentation results of various backbones, including ResNet K. He et al., 2016, Xception Chollet, 2017, and MobileNetV2 Sandler et al., 2018 Atik et al., 2022. Comparative analyses have also been conducted on ResNet, DenseNet, and EfficientNet Tan and Le, 2019 in defect detection tasks Z. Nie et al., 2020. Additionally, PCR-Net (Pruned Crack Recognition Network), a real-time model designed for efficient crack detection on edge computing devices, has been developed X. Ye et al., 2022.

In addition to model architecture improvements, ensemble learning has become a highly effective strategy for boosting segmentation performance. The effectiveness of merging specialized convolutional neural networks (CNNs) designed for detecting fine crack patterns has been demonstrated Fan et al., 2020; Hirata and Takahashi, 2023. Similarly, model averaging and ensemble modeling via majority voting have been found to improve segmentation performance Kailkhura et al., 2020; Rodriguez-Lozano et al., 2020.

A similar approach was taken by performing ensemble modeling on four pretrained models: CNNs—AlexNet Krizhevsky et al., 2012, GoogleNet Szegedy et al., n.d., VGG16 Simonyan and Zisserman, 2014, and ResNet50 Maarouf and Hachouf, 2022.

These studies highlight the ongoing innovations in network architecture tuning and model improvement for semantic segmentation. The advancements in DeepLabV3+, diverse backbones, and ensemble learning techniques are crucial to the continued progress in crack detection technologies.

2.1.2 Models for Crack Segmentation

Crack segmentation has emerged as a critical task in structural integrity assessment, particularly through the application of deep learning models like Convolutional Neural Networks (CNNs). These models have been instrumental in tackling the complex nature of crack detection across various surfaces, from pavements to buildings.

A cornerstone of this evolution has been the U-Net architecture, introduced in Ronneberger et al., 2015, which has become a standard for segmentation tasks. U-Net's adaptability and performance have led to numerous enhancements for crack detection. Notable among these are Crack U-Net Tang et al., 2022, CrackW-Net Han et al., 2021, and a YOLO-based approach using YOLOv3 Redmon and Farhadi, 2018 for crack detection J. Liu et al., 2020, all of which contribute improvements in segmentation accuracy and operational efficiency across different crack detection contexts.

In addition to U-Net derivatives, other architectures have expanded the range of tools available for crack segmentation. A network incorporating a pre-trained ResNet-34 was proposed Lau et al., 2020, while a Pyramid Attention Network using DenseNet-121 G. Huang et al., 2017 was explored to enhance feature extraction W. Wang and Su, 2020. Other notable models, such as CrackU-Net Huyan et al., n.d., SDDNet Choi and Cha, 2019, and RUC-Net G. Yu et al., 2023, bring diverse improvements aimed at addressing the nuanced challenges of pavement and concrete crack detection.

These models have proven highly effective, particularly in pixel-wise pavement segmentation, as seen in studies like Cui et al., 2015; Polovnikov et al., 2021; Qu et al., 2021, 2022; S. Wang and Tang, 2012; Z. Wang et al., 2020; Yang et al., 2019; G. Yu et al., 2023. Additionally, similar methods have been applied to detect cracks in roads Shi et al., 2016; L. Zhang et al., 2016, further broadening the applicability of crack segmentation techniques.

Hierarchical learning and generative models have also made significant strides in crack segmentation. The DeepCrack model introduces a hierarchical feature learning approach Y. Liu et al., 2019, while Crack-SegAN employs a GAN-based architecture Pan et al., 2023, both contributing to higher segmentation precision. These advancements, coupled with the emphasis on balanced evaluation metrics in crack detection Ali et al., 2022, underscore the critical role of precision, recall, and F1-scores, especially when addressing imbalanced datasets.

In an increasingly CNN-dominated landscape, diffusion models have emerged as a promising alternative for image segmentation tasks. Diffusion-based data augmentation has been applied for nuclei segmentation, showcasing how these methods can enhance training datasets by generating synthetic labeled images X. Yu et al., 2023. By augmenting just 10% of a labeled dataset with synthetic samples, performance comparable to fully supervised baselines was achieved, demonstrating the efficiency of diffusion-based augmentation.

Furthermore, the work in Dhariwal and Nichol, 2021 highlights the superiority of diffusion models over traditional deep learning methods, marking a key advancement in generative modeling. These contributions indicate that diffusion models have significant potential in the domain of crack segmentation. Inspired by these successes, this study integrates diffusion models to explore their potential in generating synthetic image datasets specifically tailored for semantic segmentation tasks. This approach positions diffusion models as a cutting-edge tool for advancing image analysis and segmentation in crack detection.

2.1.3 Data Synthesis in Various Domains

Data synthesis, the process of generating synthetic images, has become an essential tool in addressing data scarcity and mitigating privacy concerns across numerous fields. The utility of data generation has been demonstrated across domains such as healthcare, environmental monitoring, and document analysis.

In the medical domain, data synthesis has played a pivotal role in overcoming the challenges of limited data availability and privacy constraints. Generative models, such as StyleGAN, have been successfully applied to create synthetic datasets for pathological and radiological images Ding et al., 2023; Fetty et al., 2020. By leveraging HoVer-Net Graham et al., 2019 for nuclei segmentation, these studies reduce the need for labor-intensive human annotations, offering a solution for the restricted access to medical data due to privacy concerns. These advances illustrate the potential of Generative Adversarial Networks (GANs) to enhance medical imaging research significantly.

Building upon these efforts, StyleGAN3, a variant designed to eliminate aliasing, was introduced Karras et al., 2021, further improving the realism and quality of generated images by generating clearer textures. This refinement marks a significant step forward in generating high-fidelity synthetic data. Similarly, the application of synthetic data to domains such as dam surface crack detection has been expanded Xu et al., 2023, highlighting the adaptability of these techniques beyond healthcare.

In the field of historical document preservation, StyleGAN models have been trained to synthesize historical document images Bartz et al., 2022. The synthetic datasets generated from these models were used to train various models for reading historical documents, including Doc-UFCN Boillet et al., 2021, EMANet X. Li et al., 2019, and TransUNet J. Chen et al., 2021.

Moreover, innovative generative techniques like Latent Diffusion Models (LDMs), introduced in Rombach et al., 2022, and the Brownian Bridge Diffusion Model (BBDM) B. Li et al., 2023, offer new methods for efficient high-resolution image synthesis. These models are particularly adept at tasks such as image-to-image translation, enabling the generation of detailed images while reducing computational demands. By establishing direct mappings between image domains without the need for conditional generation, LDMs and BBDM offer advanced solutions for high-quality data generation.

While data synthesis has seen significant exploration in domains like medical imaging and document analysis, few efforts have been directed towards generating synthetic crack datasets for structural assessment. As such, the success of generative models in other fields provides a strong foundation for their application to crack segmentation. The high-quality image synthesis capabilities of StyleGAN and the flexibility of diffusion models like BBDM offer promising avenues for generating synthetic crack datasets that could enhance segmentation tasks in civil engineering and infrastructure maintenance.

These studies collectively demonstrate the transformative potential of data synthesis across multiple disciplines.

2.1.4 Synthesis of Structural Crack Datasets for Enhanced Detection

Crack detection is critical for ensuring the safety and longevity of infrastructure such as bridges, roads, and buildings. Undetected cracks can lead to structural failures, posing serious safety risks and resulting in costly repairs. Accurate and timely detection is therefore essential for maintaining structural integrity. Synthetic datasets have emerged as a powerful solution to address these challenges in labor-intensive data collections. By generating artificial crack images, researchers can significantly reduce the time and effort required for data collection while enhancing model training with diverse, labeled images. This approach mitigates issues of data scarcity and ensures a more consistent and scalable process for training deep learning models.

Several innovative works have pioneered the generation of synthetic crack datasets for improved detection. The introduction of 3D synthetic data to enhance crack identification performance Zhai et al., 2022, and methods based on improved deep convolutional generative adversarial networks (DCGAN) to generate virtual pavement crack images Pei et al., 2021, established the viability of synthetic data in structural assessment.

One key advancement in this field was the presentation of the Bridge Crack Library (BCL) Dataset, a benchmark collection for evaluating crack detection algorithms X.-W. Ye et al., 2021. This dataset laid the groundwork for future synthetic dataset generation, offering a robust foundation for research. Building on this, generative adversarial networks (GANs) were employed to create pixel-wise annotated crack datasets Jin et al., 2023, addressing the pressing issue of data scarcity. This method, which utilizes Pix2Pix for mask image annotations and DCGAN for crack image generation, led to the development of Bridge Crack Library 2.0 (BCL 2.0) Jin and Li, 2022. This fully synthesized dataset mirrors our approach in using GANs for data synthesis, providing a valuable point of comparison.

The methodology yielded impressive results, particularly with the best-performing model, PCR-Net, which achieved a Mean Intersection over Union (MeanIoU) of 74.34%. This performance even surpassed models trained on real images. Furthermore, it was demonstrated that pre-training with synthetic images followed by fine-tuning with real data yielded better results than combining real and synthetic datasets. However, a limitation of this work lies in the lack of public availability of the 1000-image testing set, which limits the generalizability of the findings. Additionally, the relatively small size of the dataset could impact model robustness, as our research required setting aside 1000 images for testing purposes.

In our study, we build upon these earlier efforts by introducing newer generative models, specifically StyleGAN3 and the Brownian Bridge Diffusion Model (BBDM), to enhance the process of image-to-image translation for crack synthesis. By leveraging these advanced models, we generate a synthetic dataset that not only mirrors real-world crack structures but also incorporates blurring effects to simulate various degrees of degradation. This novel approach adds an extra layer of complexity to the data, testing the resilience of segmentation models under challenging conditions.

To evaluate our synthetic dataset, we selected the DeepLabV3+ architecture, a highly accessible and proven model for semantic segmentation tasks. While PCR-Net had shown strong performance in prior studies, its limited availability led us to opt for DeepLabV3+, which is widely recognized for its reliability in segmentation. Our decision was further supported by the competitive results achieved in our experiments, where DeepLabV3+ performed favorably in comparison to previous findings. This highlights the potential of accessible models like DeepLabV3+ to deliver high-precision results in crack detection, even when applied to complex, synthesized datasets.

2.2 On Designing A Near Infrared Dorsal Hand Vein Authentication System

2.2.1 Dorsal Hand Vein Biometrics

Dorsal hand vein biometrics are a relatively new and unexplored biometric modality compared to other modalities such as face, iris, and fingerprints. Bourlai et al., 2023 and Munir and Khan, 2019 study the capabilities of imaging in a variety of spectral bands, showcasing the benefits of using each spectrum. The same idea can be applied to vein biometrics. Veins are imaged using thermal and NIR imaging spectroscopies. Both thermal and NIR are non-invasive imaging methods, but they function very differently. Thermal imaging measures body heat, which can detect vein patterns, but images vary based on human physiology Lin and Fan, 2004. NIR, on the other hand, takes advantage of the transmissive and reflective properties of skin and veins. Hemoglobin, found in blood, is highly absorptive, creating a distinct outline compared to the surrounding skin when imaged in NIR Francisco et al., 2021; Jia et al., 2021.

NIR-based imaging systems are particularly advantageous for capturing dorsal hand vein patterns, which are well-suited for biometric applications due to several key attributes:

- High Recognition Performance: Dorsal hand vein patterns demonstrate strong recognition accuracy, offering distinct vein patterns for reliable identification J. Wang and Wang, 2017.
- Enhanced Anti-Counterfeiting: The inherent complexity of vein structures makes them highly resistant to forgery, increasing security against spoofing attempts Kosmala and Saeed, 2012.

- User-Friendly Acquisition: Capturing dorsal hand vein images is non-invasive, fast, and simple, increasing user comfort and acceptance of this modality J. Wang and Wang, 2017.

- Liveness Detection: While all biometric systems are vulnerable to spoofing, dorsal hand veins exhibit a stronger resistance to such attacks. Studies have demonstrated that real and fake vein patterns can be differentiated, offering an additional layer of protection against spoofing D. Huang et al., 2014; Konnova and Mizinov, 2023; Raghavendra et al., 2015; Tome et al., 2014.

Although dorsal hand vein biometrics have their advantages, they are not as popular as other biometrics like face and fingerprint recognition. This lack of research has resulted in no standardized image alignment protocols for pose. Current imaging practices vary, with different datasets capturing hand veins in open palm, closed fist, or hand-gripping positions, each introducing unique challenges and revealing different anatomical features such as knuckle structure.

In this study, we tackle two hand poses in the Jilin University Dorsal Hand Vein Dataset (JLU) F. Liu et al., 2020b and the Wilches Dorsal Hand Vein Dataset (DHV) Wilches-Bernal et al., 2020. The JLU dataset used a bar-guided approach, while the DHV dataset used a closed fist approach. These datasets will be used to evaluate our knuckle alignment method and algorithms.

2.2.2 Dorsal Hand Vein Feature Extraction Methods

Recent advancements in deep learning have significantly transformed feature extraction processes, enabling more precise and efficient methodologies in dorsal hand vein recognition systems. Convolutional Neural Networks (CNNs) have emerged as powerful tools for vein verification and identification, demonstrating their robustness across various implementations Al-johania and Elrefaei, 2019; Benaouda et al., 2021; Tazim et al., 2018; Wan et al., 2017. Several studies have explored different CNN architectures to improve recognition accuracy, with notable success reported in models like VGG-16 and GoogLeNet X. Li, Huang, and Wang, 2016; G. Wang et al., 2019a. For instance, G. Wang et al., 2019b leveraged CNN-based convolutional descriptors to enhance the extraction of vein features. Further advancements, such as attention-based CNNs, have proven effective by streamlining traditional architectures, achieving strong recognition performance without the need for extensive preprocessing or data augmentation Kocakulak et al., 2023. Additionally, Alashik and Yildirim, 2021 introduced a hybrid model combining deep learning and generative adversarial networks (GANs), which outperformed conventional approaches based on traditional characteristics such as LBP, SIFT and GABOR.

Preprocessing and feature extraction continue to be critical components in vein biometric systems. Image enhancement algorithms and statistical feature extraction has been well explored and proven Chin et al., 2020; Y. Wang et al., 2019; Zhu et al., 2013. Vein pattern visibility has largely been made possible by photometric normalization techniques such as contrast-limited adaptive histogram equalization (CLAHE) Nadiya and Gopi, 2020; Nour et al., 2024; Ruiz-Echeverri et al., 2021; Y. Wang et al., 2019; Yakno et al., 2021. CLAHE is often used alongside other contour and thesholding processes to facilitate vein segmentation Rossan and Khan, 2014; Y. Wang et al., 2022. Mathematically driven morphological methods such as graph extraction and skeletonization are preprocessing methods that has been used to simplify vein pattern representation for computationally quick comparisons Arakala et al., 2015; Cao et al., 2015; D. Huang et al., 2016; Lajevardi et al., 2014; X. Li, Huang, Zhang, et al., 2016; Luna-Benoso et al., 2021; Z. Wang et al., 2008.

In this study, we developed a dorsal hand vein recognition system that utilizes an optimized set of feature extraction techniques that we discussed (see section 3.2). Our experiments demonstrate that the proposed methods produce highly reliable vein features, leading to significant improvements in verification and identification performance.

2.2.3 Dorsal Hand Vein Alignment Methods

Alignment is a common biometric preprocessing technique used often in face recognition. It reduces the variation in pose so the model can more accuractely compare the face images. However, alignment of dorsal hand vein images do not align the same way that faces do. There are a lot of variations of dorsal hand alignment, but versions of key point alignment method stand out as ways to dynamically extract Regions of Interest (ROIs) from the back of the hand in which the models are trained on Arora et al., 2018; Damak et al., 2018; W. Nie and Zhang, 2019; Pititheeraphab et al., 2020; L. Wang et al., 2008; R. Zhang et al., 2023.

As mentioned earlier, pose and data collection has a major affect on how the carious alignment methods can be carried out. For closed fist variations, Kumar and Prathyusha, 2009 and Bhosale and Jadhav, 2017 took similar approaches by doing knuckle tip extraction methods to extract ROI boundaries using the index and right finger knuckles. Kumar et al. aligned the dorsal hand veins by identifying the index and ring knuckle tips, using either wrist-centered peaks or the distance from the knuckles to the wrist. They extracted a 400×300 pixel region of interest (ROI), placed 150 pixels below the middle finger knuckle and aligned along the line between the index and ring knuckles, then vertically aligned the ROI for vein extraction. In contrast, Bhosale et al. located the centroid and three knuckle tips (index, middle, ring), sorting peak distances from the hand contour to the centroid. They aligned the knuckles horizontally before extracting a fixed ROI. Both methods exclude the little finger or pinky knuckle.

Hsu et al., 2011 and Lee et al., 2014 aligned dorsal hand vein images by identifying two key points and rotating the image based on the angle between them. They first calculated a fixed point using the center of gravity (COG) of the hand, then identified the leftmost and rightmost contours above the COG. The area between these contours was divided into four sub-regions. The longest distances between the COG and hand contour in the first and fourth sub-regions were used to extract two key points. These points determined the axis of rotation, and the image was aligned horizontally based on the midpoint between them. A square region of interest (ROI) was then extracted from the back of the hand.

2.3 Concealing Human Vein Identity: Effect of Image Degradation on Dorsal Hand Vein Recognition

In this study, we build off the competitive results obtained in our study around designing a dorsal hand vein authentication system to study the effects that physiological changes and image degradations can have on dorsal hand vein biometric patterns.

2.3.1 Robustness to Image Degradation

As mentioned in the previous section, dorsal hand vein recognition is a relatively underexplored biometric modality, with very few studies focusing on the impact of image degradation in vein biometrics. Hand vein biometrics present unique challenges that make vein imaging difficult. Kauba et al., 2022 and Lim and Mulcahy, 2017 state that physiological changes, such as aging or increased subcutaneous fat content, can affect vein visibility, leading to more visible veins with aging and less visible veins with increased fat.

Most image degradations arise from environmental factors that affect image quality, including lighting, motion blur, and meteorological conditions. To investigate these issues, Piuri and Scotti, 2008 demonstrated that contactless biometric scanners are susceptible to defocus, motion, and image blur. Derawi et al., 2012 further emphasized the importance of illumination in acquiring clear images from biometric scanners. El-Naggar and Bourlai, 2022 found that varying levels of contrast, brightness, and blur led to performance degradation, but artifacts were responsible for a more significant decline in accuracy. Abaza et al., 2014 showed that combining various image quality metrics into a unified quality index greatly improves a recognition system's resilience to degradations such as contrast, brightness, sharpness, and focus, suggesting this approach could be applied to dorsal hand vein biometrics. Z. He et al., 2017 identified that meteorological factors, such as humidity, temperature, and pressure, are often overlooked but can significantly impact biometric performance.

Occlusions, such as tattoos or jewelry, can also hinder dorsal hand vein recognition by reducing vein visibility. F. Liu et al., 2020b noted that depending on the size of the tattoo or its transparency, vein pattern visibility may affect performance. Martin and Bourlai, 2017 found that tattoos become more or less visible under different light spectrum's, depending on skin conditions and ethnic background.

2.3.2 Research Gaps and Focus

While considerable progress has been made in feature extraction and alignment techniques, the influence of image degradation on dorsal hand vein recognition remains understudied. This paper addresses that gap by investigating how various simulated blurring effects, mimicking physiological changes, challenge the system's robustness and reliability.

Chapter 3

Methodologies

3.1 On Enhancing Crack Semantic Segmentation using StyleGAN and Brownian Bridge Diffusion

We present a novel methodology for synthetic data generation that circumvents the traditional data annotation process, which can be very timely. Our methodology comprises a multi-step process that can be found in Figure 3.3.

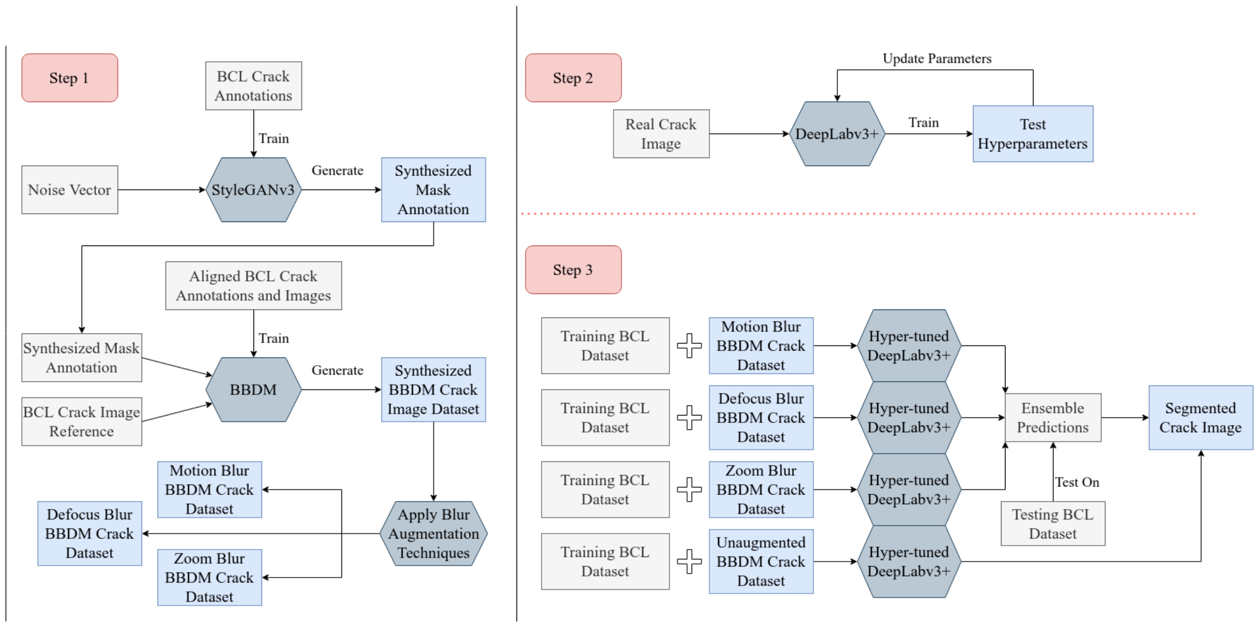

Figure 3.1: Overall Methodology Schematic: A 3-step process was implemented to produce a synthesized dataset with BBDM in the pixel and latent spaces. In step one, StyleGAN3 is trained on the ground truth mask annotation images from the BCL dataset. Once trained, a noise vector is passed into the model, and a new randomly generated mask annotation image is formed. The pixel and latent space BBDM were trained on the entire BCL dataset. The newly generated mask annotations were randomly paired with a BCL crack image for inferencing purposes. The result is a synthesized BBDM crack image dataset. Three additional blur-augmented datasets were created by copying the original synthesized BBDM crack dataset. In step two, DeepLabV3+'s hyperparameters are tuned to improve overall segmentation accuracy. In step three, 8 different datasets, 4 from the pixel space BBDM and 4 from the latent space BBDM, were trained using the new hyper-tuned DeepLabV3+. All the predictions from the models trained on the datasets with blur augmentation were ensembled together to generate a final segmented crack image.

In the first step, we employ StyleGAN3 to train on crack annotation images sourced from the BCL mask annotations. The BCL mask annotations were created by X.-W. Ye et al., 2021. Refer to Figure 3.2 and to the Bridge Crack Library Dataset section for more details.

In our initial GAN-based step, we generate annotation mask images, which lay the groundwork for the subsequent synthetic data generation. StyleGAN3 is particularly well-suited for this task, as it excels at producing fine details, making it ideal for capturing the intricate and varied patterns of cracks. StyleGAN3 can generate a diverse range of crack patterns, which is necessary for training robust semantic models. Next,

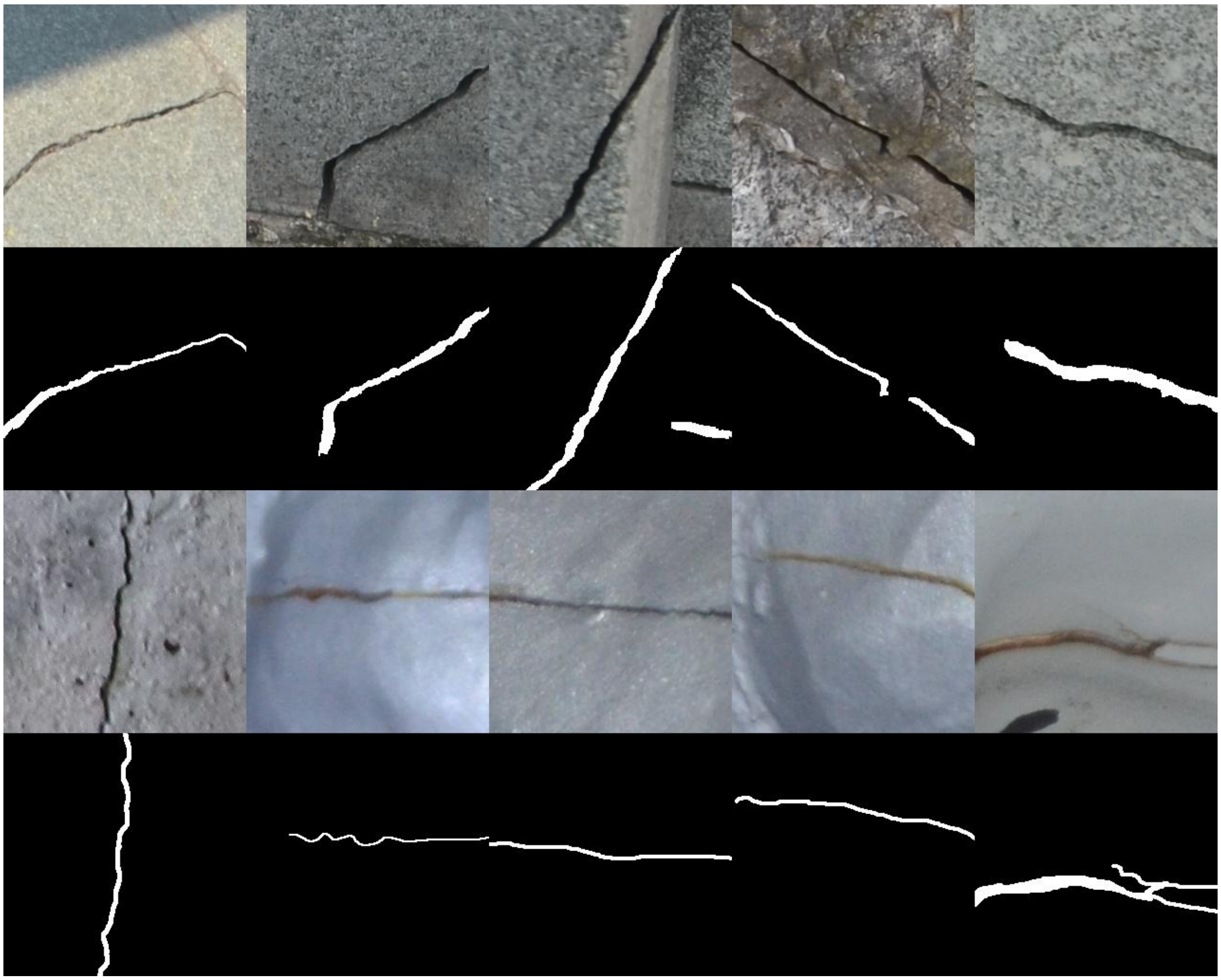

Figure 3.2: Bridge Crack Library Sample Images. The figure shows a selection of images from the Bridge Crack Library, where each picture captures different types of cracks found in materials like stone, concrete, and metal. These pictures, all with the same size of 256 by 256 pixels, are part of a big collection gathered from inspecting over 50 bridges. This collection includes 5,769 non-steel crack images (as seen in row one), 2,036 steel crack images (as seen in row three), and 3,195 images identified as noise. Each of these images also has an annotated pair as seen in rows two and four.

we incorporate BBDM into our pipeline, allowing us to translate annotation mask images into realistic synthetic structural crack images. This translation is vital for creating synthetic data that closely mimics the texture and structure of real crack images. Unlike traditional methods that depend on conditional generation processes, BBDM directly translates between two image domains through a stochastic process, ensuring that the synthesized crack images are highly realistic. This marks the first application of BBDM in crack image data generation and the first use of diffusion models for this purpose. In the second step, we fine-tune the DeepLabV3+ model using the data in Table 3.3, then train it on the blur-augmented synthesized datasets. Finally, we ensemble the trained model predictions using majority voting to achieve the best MeanIoU performance.

3.1.1 Creating Synthesized Datasets

StyleGAN3

StyleGAN3 represents an advanced iteration of GANs, renowned for its capacity to generate high-quality and diverse images Karras et al., 2021. In our methodology, StyleGAN3 is employed to learn and replicate the intricate patterns and characteristics found in crack annotation images derived from the BCL dataset. In other words, StyleGAN3 is used for generating our synthetic mask annotations of cracks. These annotation mask images delineating the outlines and boundaries of structural cracks, will serve as our ground truths for our synthetic data in subsequent steps in our pipeline.

Unlike its predecessors, StyleGAN3 incorporates several key innovations that improve both the quality and stability of the generated images Karras et al., 2021. A particularly noteworthy aspect of StyleGAN3 is its redesigned generator architecture, which includes an adaptive discriminator augmentation mechanism.

This mechanism effectively mitigates overfitting, a common issue when training on a relatively small, specialized dataset like the BCL. The network also introduces enhancements in both the mapping and synthesis networks, which contribute to the creation of more detailed and varied images, thereby bolstering the model's ability to produce diverse and realistic crack patterns.

Training StyleGAN3

The training of StyleGAN3 is conducted over multiple days using 4 NVIDIA A6000 GPUs on an Ubuntu system. All images are processed at a resolution of 256 by 256 pixels, with no preprocessing required for the BCL dataset. To optimize GPU utilization, the batch size is set to 16. The Gamma value is configured at 8 to regulate the trade-off between the fidelity and variety of the synthesized images.

The gamma parameter in StyleGAN3 is crucial for balancing the fidelity and diversity of the generated images. We selected a gamma value of 8. Over the course of approximately 10,000 epochs, we observe a gradual refinement in the quality of the generated images. Initially, the images only vaguely resembled crack patterns, but as training progressed, the images became more defined and realistic, closely resembling various crack formations.

In Figure 3.3, the sample synthesized mask image results show a wide variety of crack shapes. The black color indicates what will be the background of the image, and the white color indicates what the pixels that will be a crack in the colored synthesized image. This synthesized counterpart will be produced using BBDM.

The generated cracks exhibit a wide range of textures, resembling various types of damaged surfaces, including thin, thick, bifurcated, rugged, and smooth cracks. All the cracks appeared in the same orientation. To introduce diversity into the generated images, we adopted a pragmatic data augmentation

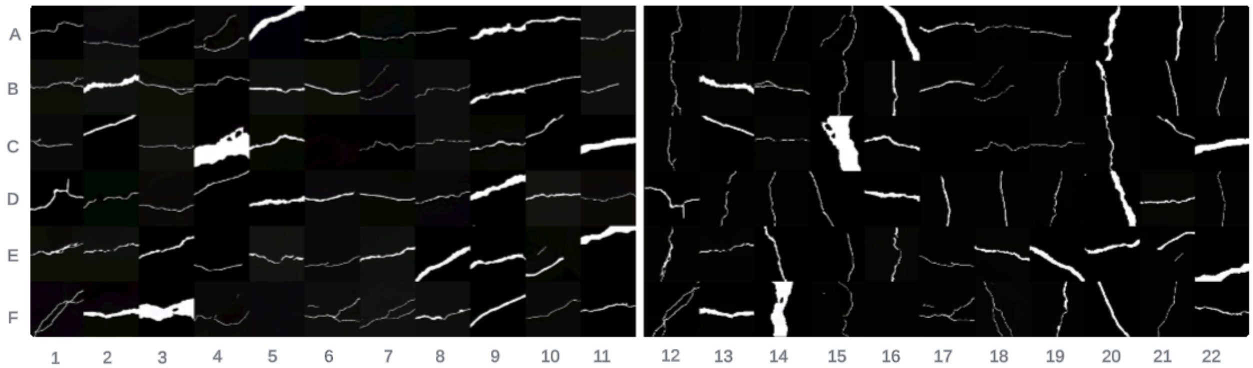

Figure 3.3: StyleGAN3 Generated and Rotated Results. Each tile displays a distinct crack formation with varying orientations and morphologies, ranging from linear to branched, from fine to wide cracks, and from smooth to rugged. The diversity in crack presentations simulates a wide array of real-world conditions. The rotation of crack images in columns 12 through 22 exemplifies the model's ability to depict varied crack orientations, enriching the dataset's diversity. Generating cracks uniformly across a single axis simplifies dataset compilation and ensures the model's impartiality to crack orientation, leading to a more reliable crack segmentation model.

strategy. We applied random flips and rotations (0, 90, 180, and 270 degrees) to the input images. This approach introduced the necessary variability, effectively ensuring a diverse representation of crack types, as illustrated in Figure 3.3. This augmentation technique was applied across all the synthesized mask annotations.

The resulting synthetic crack images displayed a remarkable array of textures and forms, closely mimicking real-world crack patterns. This outcome highlights StyleGAN3's ability to produce a comprehensive set of synthetic annotations capable of robustly supporting various semantic segmentation tasks.

BBDM

The BBDM presents a novel architecture for image-to-image translation built on the architecture of a diffusion models. This becomes particularly useful for generating synthetic crack images for our pipeline.

BBDM differentiates itself by employing a unique brownian bridge translation process which have never been used in the context of crack image synthesis.

Traditional diffusion models perform a forward process that gradually adds noise to an image until it becomes indistinguishable from random noise. The reverse process then denoises this random input to generate new images. However, BBDM distinguishes itself by employing a Brownian Bridge process, which is a stochastic process conditioned on both the initial and final states. This means that instead of starting from pure noise, the diffusion process in BBDM is guided from a specific starting image to a target image, ensuring a smooth and controlled transformation between the two states.

In our context, the starting state could be a synthetic mask image, and the ending state is an image from the BCL dataset, which contains images with specific crack patterns. This dual anchoring allows BBDM to generate images that not only transition smoothly but also closely align with the desired characteristics, such as accurate and realistic crack formations. The forward diffusion process provides detailed insights at each step by offering the marginal distributions, allowing for a better understanding of how the image evolves over time. The reverse process, starting from the conditional input, ensures that the model's output is finely tuned to produce high-quality images that maintain essential details crucial for crack detection tasks.

An important aspect of BBDM is the distinction between operating in pixel space versus latent space. Pixel space BBDM works directly on the image pixels, which can result in high-precision outputs but is computationally intensive due to the high dimensionality of image data. On the other hand, latent space BBDM operates within the compressed feature space provided by models like Vector Quantized Generative Adversarial Networks (VQGAN). By encoding images into a lower-dimensional latent space, VQGAN effectively captures the essential features of the images while reducing computational complexity.

Using latent space BBDM offers several advantages:

1. **Efficiency**: Operating in a lower-dimensional space reduces the computational resources required for training and inference, enabling faster processing times.

2. **Texture Variety**: The compressed latent space retains important structural and textural information, allowing the model to generate images with diverse and realistic textures, which is particularly useful for simulating various crack patterns.

3. **Scalability**: The reduced dimensionality makes it feasible to train on larger datasets or more complex models without incurring prohibitive computational costs.

In contrast, while pixel space BBDM may provide higher precision due to its direct manipulation of image pixels, it is less efficient and may not capture the broad variability of textures as effectively as the latent space approach.

By utilizing the latent space BBDM with VQGAN, we achieve a balance between computational efficiency and the ability to generate high-quality, diverse synthetic crack images. This approach enhances our model's capability to learn intricate crack patterns and contributes to more accurate crack detection in real-world applications.

Training BBDM

In our study, we examined two approaches to generating a synthesized dataset using the Brownian Bridge Diffusion Model (BBDM): one operating in pixel space and the other in latent space. Both models were trained on the complete set of images and annotations from the BCL dataset. To generate the synthetic dataset, we first created StyleGAN3 mask images, which were then paired with randomly selected real images from the BCL dataset to perform domain inference.

The Latent Space Brownian Bridge Diffusion Model (LBBDM) was specifically designed to prioritize efficiency and generalization. The training process consisted of 400,000 steps, with a batch size of 8. LBBDM utilized the Adam optimizer, maintaining the same learning rate and beta1 parameters as the pixel space variant. VQGAN, known for its capability to compress high-dimensional data into a latent space while preserving essential features, was employed with a UNet architecture. This setup was particularly suited to the 64x64 resolution, enabling faster training and inference. The VQGAN model was pre-trained on the VQ-f4 dataset[1].

In contrast, the Pixel Space BBDM focused on directly learning and translating crack patterns at the pixel level. This model underwent a more extensive training regime, spanning over 600,000 steps on a single GPU. The training process employed a batch size of 8 and the Adam optimizer with a learning rate of 1.e-4 and a beta1 value of 0.9. A learning rate scheduler was incorporated to adjust the learning rate dynamically, featuring a decay factor of 0.5 and a patience threshold of 3,000 steps. To further refine the model, an Exponential Moving Average (EMA) with a decay rate of 0.995 was introduced starting at step 30,000. The model architecture was optimized for handling 256x256 resolution images, with a particular emphasis on capturing the intricate details of crack patterns.

Generating Multiple Blur-Augmented Synthetic Datasets

Following the training phase of the BBDM, we generated several synthetic datasets using 200 sampling steps during the inference stage. Both real crack images from the BCL dataset and synthesized annota-

[1]VQ-F4 can be accessed from the official repository https://github.com/CompVis/latent-diffusion

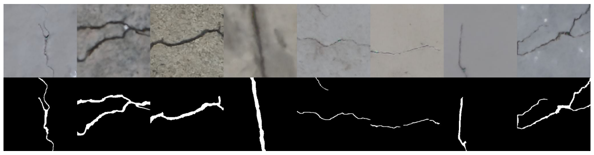

Figure 3.4: Original Pixel Space BBDM Sample Images: After 600,000 steps of training, the images exhibit the diversity of domains when reproducing crack annotations. The pixel space images very accurately represent the annotation images when compared to the latent space.

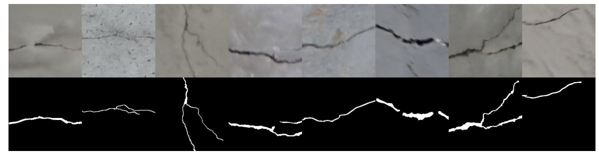

Figure 3.5: Motion Blur Pixel Space BBDM Sample Images: Motion blur was applied to all real synthesized datasets with a random severity of 3 to 12.

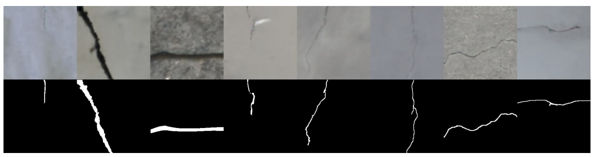

Figure 3.6: Defocus Blur Pixel Space BBDM Sample Images: Defocus emulates a camera that is not focused when the image is captured. A severity magnitude of one was applied to all images.

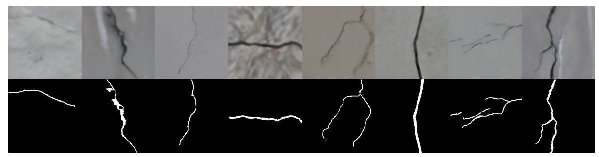

Figure 3.7: Zoom Blur Pixel Space BBDM Sample Images: A severity magnitude of one was applied to all synthesized images.

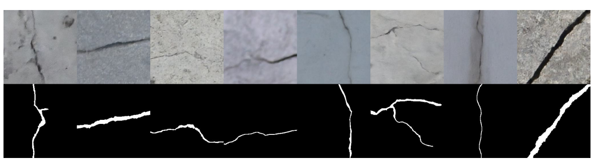

Figure 3.8: Original Latent Space BBDM Sample Images: The LBBDM was trained for efficiency and generalization, producing images with diverse and realistic backgrounds when compared to the pixel space. The latent space synthesized real images do not adhere as closely to the annotation images as the pixel space.

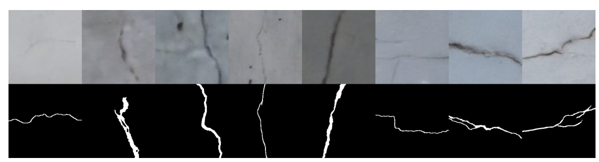

Figure 3.9: Motion Blur Latent Space BBDM Sample Images: A random severity in the range of 3 to 12 was applied to the latent space images.

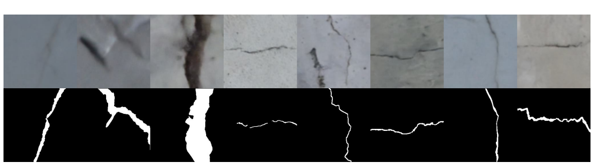

Figure 3.10: Defocus Blur Latent Space BBDM Sample Images: A strength of 1 blur was applied to the synthesized images.

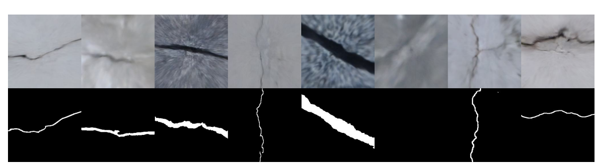

Figure 3.11: Zoom Blur Latent Space BBDM Sample Images: A strength of 1 was applied to the synthesized images.

Figure 3.12: Sample Images from Synthesized BBDM Datasets

tion images produced by StyleGAN3 served as domain reference images. Each dataset consists of 7,800 synthesized images, supplemented by a selection of real images from the BCL dataset. For both the pixel space and latent space models, we created four distinct datasets. Three of these datasets were augmented with different types of blur to introduce variety and address some of the inherent challenges associated with the BCL dataset. The fourth dataset remained unaltered and served as a baseline for comparison.

Incorporating these various blur types as data augmentation techniques plays a crucial role in enhancing the adaptability and robustness of our models, particularly in the context of the BCL dataset. We employed Imgaug, a publicly available data augmentation tool developed by Michaelis et al., 2019, to apply different magnitudes and types of blurring, as shown in Figure 3.12. This Figure contains several sample pairs of the synthetic real and mask images with various blurs applied to them.

The blurring techniques applied to the augmented datasets are as follows:

1. **Motion Blur**: This augmentation replicates the effect of rapid movement, preparing the model for scenarios involving moving subjects or capturing devices. A random severity ranging from 3 to 12 was applied to the pixel and latent BBDM datasets. Motion blur is particularly beneficial for analyzing images of cracks captured in motion, such as those taken from moving vehicles or dynamic monitoring systems, ensuring that the model maintains consistent crack detection performance even under motion.

2. **Zoom Blur**: Zoom blur mimics the effect of rapid changes in focal length, challenging the model to maintain accuracy despite variations in image focus. This augmentation is crucial for images captured from varying distances or during swift zooming actions, common in field surveys. A magnitude of one was applied to the zoom blur BBDM synthetic datasets.

3. **Defocus Blur**: Defocus blur simulates scenarios where the image is not perfectly focused. A severity of one was applied to the defocus blur BBDM synthetic datasets. This type of blur presents a challenge for the model to accurately recognize and segment cracks, even when the overall sharpness of the image is compromised.

The pixel space and latent space models generated images with distinct characteristics. The pixel space model demonstrated high fidelity in reproducing crack annotations, but it often struggled with generating realistic textured backgrounds, resulting in backgrounds that occasionally appeared washed out, as seen in Figure 3.4. Despite the accurate rendering of cracks, the less convincing textural quality of the backgrounds somewhat limits the overall realism of the images.

Conversely, the latent space BBDM excelled at producing more diverse and realistic backgrounds, as illustrated in Figure 3.8. This improvement in background texture significantly enhanced the overall realism of the images. However, this advantage was accompanied by a trade-off: the latent space model generated cracks that did not always adhere closely to the annotated patterns. The discrepancies in crack representation suggest that the model focuses more on the broader image context rather than on precise detail replication.

For each synthesized dataset, the set was also combined with the real training data specified in Table 3.3, with no synthesized data added to the validation set. Table 3.1 describes all the datasets that were generated and their contents.

Table 3.1: Description of Training Datasets: This table outlines various training datasets created to evaluate how their unique properties enhance the crack detection system's effectiveness.

Dataset	**Description**
BCL	Training and Validation dataset split described in 3.3 (BCL Training and Val).
BCL + BCL 2.0	Combines BCL Training and Val with data from BCL 2.0 images. BCL 2.0 was produced by Jin et al.Jin et al., 2023 which was trained on the original BCL dataset and contains fully synthesized images for segmentation.
M-BBDM-P	7,800 BBDM pixel space synthesized images were combined with BCL Training and Val.
M-BBDM-P-ZB	M-BBDM-P with zoom blur applied to synthesized images.
M-BBDM-P-MB	M-BBDM-P with motion blur at severity levels 3 to 12 applied only to synthesized images.
M-BBDM-P-DB	M-BBDM-P with defocus blur applied only to synthesized images.
M-BBDM-L	7,800 BBDM latent space synthesized images were combined with BCL training and validation.
M-BBDM-L-ZB	M-BBDM-L with zoom blur applied only to synthesized images.
M-BBDM-L-MB	M-BBDM-L with motion blur at severity levels 3 to 12 applied only to synthesized images.
M-BBDM-L-DB	M-BBDM-L with defocus blur applied only to synthesized images.

By training models with datasets augmented with motion, zoom, and defocus blurs, we significantly bolster their ability to interpret and analyze crack images under various real-world conditions. This approach not only enhances the model's generalization capabilities but also ensures its practical applicability in diverse structural health monitoring scenarios. The enhanced model is better equipped to handle variations in image clarity and texture, which are common in real-world structural assessments, thereby improving its reliability and effectiveness in practical applications. Incorporating zoom, defocus, and motion blur augmentations into our training datasets prepares the DeepLabV3+ model for adverse camera conditions often encountered in practical settings. Later, we introduce ensemble modeling to leverage the different synthesized augmented datasets to increase MeanIoU performance.

3.1.2 Hyperparameter Tuning DeepLabV3+

DeepLabV3+

DeepLabV3+[2] is a state-of-the-art model for semantic segmentation, renowned for its effectiveness in handling complex image segmentation tasks. The architecture of DeepLabV3+ is an evolution of its predecessors, designed to provide high precision in object boundary delineation. It employs atrous (dilated) convolutions, which allow the model to control the resolution at which feature responses are computed within Deep Convolutional Neural Networks. This feature is particularly advantageous for segmenting fine details in images, such as small or narrow cracks. At the core of DeepLabV3+ is the Atrous Spatial Pyramid Pooling (ASPP) module, which applies atrous convolution at multiple rates to capture multi-scale information, enabling the detection of cracks that vary in scale.

In addition, the model adopts an encoder-decoder structure. The encoder, enhanced by the ASPP module, captures rich contextual information, while the decoder refines the segmentation results, particularly along object boundaries. This refinement is crucial for the precise delineation of crack contours.

Within the framework of our study, the deployment of DeepLabV3+ extends beyond its role as an advanced tool for semantic segmentation. We undertook extensive hyperparameter tuning to adapt the DeepLabV3+ architecture to meet the specific demands of our dataset and the complexities of our segmentation objectives. Moreover, DeepLabV3+ served as a benchmark model to evaluate the efficacy of our innovative data generation methodologies, which include the synthesis of training data using StyleGAN and BBDM. Additionally, we explored ensemble modeling with DeepLabV3+ to enhance segmentation performance. By training multiple DeepLabV3+ models on a range of synthesized datasets and com-

[2] DeepLabV3+ implementation and pretrained model can be accessed from this GitHub repository: https://github.com/VainF/DeepLabV3Plus-Pytorch/tree/master

bining their outputs, we achieved a significant improvement in segmentation precision. DeepLabV3+ serves as the backbone for evaluating our segmentation performance, underscoring the importance of hyperparameter tuning in our project.

DeepLabv3+ Hyperparameter Tuning

The training of DeepLabV3+ was conducted on the Bridge Crack Library Dataset which exhibits a wide variety of real-world conditions. The composition of the datasets used for training, validation, and testing is detailed in Table 3.3.

Table 3.3: Composition of the Training, Validation, and Test Datasets for Hyperparameter Tuning: To emulate the training conditions of Jin et al., 2023, the BCL and BCL 2.0 datasets were combined to perform hyperparameter tuning. The BCL is made up of three classes of images: cracks, noise, and steel. Unclassified synthetic data from BCL 2.0 was added to both the validation and training data. No synthetic data was included in the testing dataset. No noise data was used in the validation set because no noise data was used in the testing set.

Dataset Group	Image Types	Number of Images
Training	Cracks	4,036
	Noise	3,195
	Steel	1,424
	Synthetic - BCL 2.0	10,919
Validation	Cracks	866
	Noise	0
	Steel	305
	Synthetic - BCL 2.0	2,340
Test	Cracks	866
	Steel	306

The primary goal of our hyperparameter tuning was to optimize the DeepLabV3+ model for accurate crack segmentation. We commenced our experiments with a basic configuration of the DeepLabV3+ model, which involved training the model for 200 epochs. In this initial setup, the learning rate was set at 0.001, with a batch size of 128. The model's output stride was fixed at 8, and we utilized the Adam optimizer with a weight decay of 0.0001 and beta values of 0.9 and 0.999. The cross-entropy loss function was employed, and the backbone architecture used was MobileNet. This initial configuration achieved an accuracy of 59.57%, as summarized in Table 4.1.

Through extensive experimentation, we identified the optimal hyperparameters that significantly enhanced the performance of the DeepLabV3+ model. The optimal configuration, also summarized in Table 4.1, involved training the model for 200 epochs with a fixed learning rate of 0.001 and a batch size of 128. We maintained the output stride at 8 and continued to use the Adam optimizer with a weight decay of 0.0001 and beta values of 0.9 and 0.999. The cross-entropy loss function was retained for consistency. To further enhance the model's performance, we switched the backbone to a pre-trained 32-layer High-Resolution Net (HRNet). Additionally, we implemented on-the-fly data augmentation during training to increase the model's robustness and improve convergence.

This fine-tuned setup yielded the best results, demonstrating a substantial improvement over the initial configuration.

3.1.3 Ensemble Modeling DeepLabV3+ Models Trained on Various Synthesized Datasets

To improve the performance of structural crack detection, we implemented an ensemble modeling approach by combining multiple DeepLabV3+ models, each trained on the distinct datasets described in Table 3.1. Specifically, a separate DeepLabV3+ model was trained on each of the six augmented datasets, allowing each model to become adept at handling images with particular types of blur.

After the training phase, the outputs of these models were combined through an ensembling technique called majority voting. Majority voting is a robust method for aggregating the predictions of multiple models. In this process, each model in the ensemble independently classifies each pixel in an image as a crack or not. For every pixel, the classification that receives the most votes across all models determines the final prediction.

This ensemble approach enhances prediction accuracy by reducing the impact of biases or errors from any single model. By leveraging the strengths of each model trained on different dataset variations, the ensemble effectively handles the diverse image qualities and conditions encountered in crack detection tasks. For illustrative results of the majority voting ensemble modeling, refer to Figure 3.13.

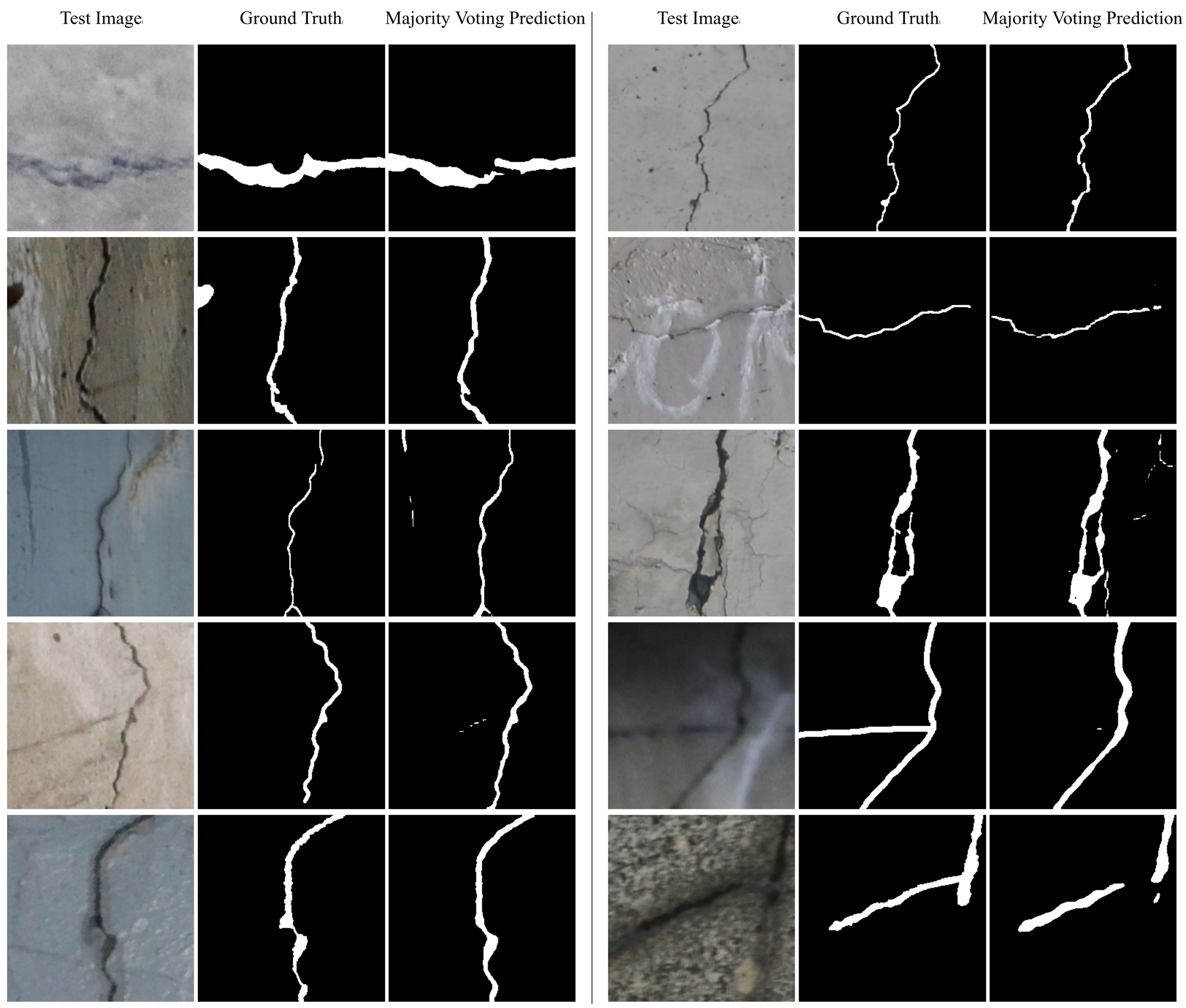

Figure 3.13: Majority Voting Ensemble Predictions. This majority voting ensemble method effectively combines individual model predictions of three DeepLabV3+ models that were trained separately on M-BBDM-P-ZB, M-BBDM-P-DB, and M-BBDM-P-MB.

3.2 On Designing A Near Infrared Dorsal Hand Vein Authentication System

We present a novel NIR-based dorsal hand vein authentication system capable of working efficiently when using two hand positions, namely the bar-guided position and the closed fist back of the hand presentation position.

Our methodology, depicted in Figure 3.20, is based on a novel knuckle alignment technique that integrates a set of image processing methods, such as extracting the largest inscribed circle on the back of the hand, with our unique proposed technique to key point knuckle alignment. These techniques ensure that high-quality human vein patterns of each individual are captured consistently, thus reducing intra-class variability and enhancing the system's accuracy and performance.

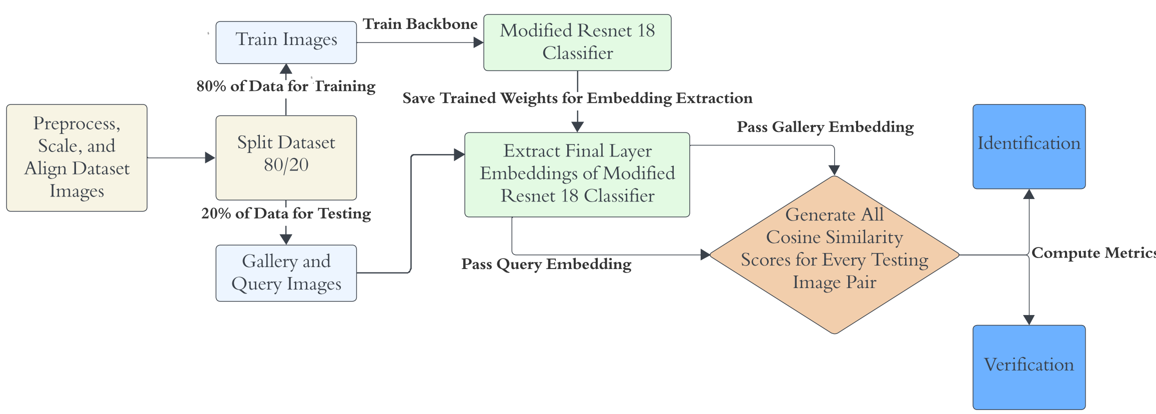

Figure 3.14: Biometric System Overview - The first step is to use our knuckle alignment-based methodology. From this, we use our training data to train a deep learning classifier for its feature extraction backbone. The trained backbone generates feature embeddings of the Gallery and Query images to calculate all the cosine similarity scores between every unique pair in the testing set. Using these cosine similarity scores, we compute the identification and verification scores such as Rank-1, EER, AUC, and accuracy.

Our proposed approach is made up of two main sections below.

1. Alignment: Unlike traditional methods that rely on knuckle-tip extraction, our approach leverages the largest inscribed circle for centering and uses two knuckle points alignment. This technique involves scaling and aligning the images to maintain uniformity across the dataset. We identify the largest inscribed circle within the hand contour, translate the image to center this circle, and use the furthest contour points in the top two quadrants to align the knuckles around this center.

2. Model Development: We designed a modified ResNet-18 classifier and performed hyperparameter tuning, i.e., we optimized the learning rate, batch size, and loss function through an exhaustive grid search. After training the optimized classifier, we procure the output embedding vector that represents the dorsal hand vein images. In the context of biometric matching using dorsal vein images for identification or verification experiments, we compute cosine similarity scores between embedding vectors associated with different or the same individuals.

Next, we will be discussing the knuckle alignment method and the matching approach we used for our studies.

3.2.1 Knuckle Alignment

Knuckle alignment is a preprocessing step in dorsal hand vein biometrics, designed to standardize the orientation and position of hand images before feature extraction. The goal is to minimize intra-class variability due to hand positioning and pose, thereby enhancing the accuracy and robustness of the feature extraction and matching process. This process involves identifying three key points: the center of the hand and two knuckle points located at the outer edges of the pinky and pointer fingers. The objective is to minimize intra-class variability attributable to hand positioning and pose, thereby enhancing the robustness of the feature extraction and matching process.

Knuckle alignment can be performed either manually or automatically, with each approach offering unique advantages and challenges, especially in the context of deep learning. Manual alignment relies on human operators to visually inspect and adjust the identified knuckle and center points in each image. This method can offer high precision, particularly in challenging scenarios where the hand positioning or image quality is inconsistent. Skilled operators can make nuanced decisions that automated systems might miss, particularly in complex images where the subject's pose differs significantly from others in the dataset.

However, manual alignment is inherently labor-intensive and becomes increasingly impractical as the size of the dataset grows. The scalability issue is a significant drawback, as larger datasets require exponentially more time and resources to process manually. Additionally, the manual approach is susceptible to human error and variability such as differences in alignment accuracy due to operator fatigue or varying levels of skill among operators.

Automated alignment leverages deep learning and computer vision algorithms to detect the key alignment points and adjust images accordingly, and it is highly scalable, enabling the rapid processing of large datasets with minimal human intervention. These systems apply consistent alignment criteria across all images, ensuring uniformity in the preprocessing stage and reducing the potential for human-induced errors.

Automated systems are particularly advantageous when dealing with vast datasets typical of modern biometric projects. They can handle repetitive tasks efficiently and at a lower cost compared to manual methods. However, the effectiveness of automated alignment depends heavily on the quality of the data and the robustness of the underlying algorithms. Poor-quality images or significant variations in pose and lighting can still pose challenges, where human judgment might surpass the current capabilities of automated systems.

The choice between manual and automated alignment depends on the specific requirements of the biometric system, including the scale of the dataset, accuracy needs, and the availability of skilled operators. Interestingly, the difference in effectiveness between manual and automated methods may be narrower than traditionally assumed, with each offering distinct benefits based on the context of use.

The automatic and manual knuckle alignment process are derived from the same core concept of identifying one center point and two knuckle alignment points for alignment. In the two subsections, we provide our methodologies for extracting these center points.

Proposed Manual Alignment

In our manual alignment process, operators select three key points: the perceived center of the back of the hand and one point on each of the leftmost and rightmost knuckles. Once these points are identified, the image is translated so that the hand's center point aligns with the geometric center of the image, without applying any scaling adjustments. To ensure a consistent orientation across the dataset, the image is then rotated to align the line connecting the outermost knuckle points with the horizontal axis.

The rotation process involves calculating the angle between the line formed by the two outermost knuckle points and the horizontal axis. Specifically, this angle is determined by measuring the deviation of the line from the horizontal plane. Once the angle is calculated, the image is rotated around its new center by this angle, effectively positioning the knuckle line to be perfectly horizontal. This ensures that all images have a uniform orientation, with the line connecting the outermost knuckles consistently aligned along the horizontal axis.

To align an image so that the selected center point is translated to the geometric center of the image, and the line formed by the outermost knuckle points is horizontal, follow these steps:

1. Translation:

   Calculate the translation required to move the selected center point (x_c, y_c) to the image's geometric center (x_g, y_g):

$$t_x = x_g - x_c \tag{EQ 1}$$

$$t_y = y_g - y_c \tag{EQ 2}$$

   The translation matrix $M_{\text{translate}}$ is given by:

$$M_{\text{translate}} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix} \tag{EQ 3}$$

   Apply the translation to the image using the affine transformation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{EQ 4}$$

2. Updating Points After Translation:

Update the coordinates of the furthest points (x_1, y_1) and (x_2, y_2) after translation:

$$x_1' = x_1 + t_x \tag{EQ 5}$$

$$y_1' = y_1 + t_y \tag{EQ 6}$$

$$x_2' = x_2 + t_x \tag{EQ 7}$$

$$y_2' = y_2 + t_y \tag{EQ 8}$$

3. Rotation:

Calculate the angle to rotate the line formed by the translated furthest points to be horizontal:

$$\Delta x = x_2' - x_1' \tag{EQ 9}$$

$$\Delta y = y_2' - y_1' \tag{EQ 10}$$

$$\theta_{\text{rad}} = \arctan\left(\frac{\Delta y}{\Delta x}\right) \tag{EQ 11}$$

$$\theta_{\text{deg}} = \theta_{\text{rad}} \times \left(\frac{180}{\pi}\right) \tag{EQ 12}$$

The rotation matrix M_{rotate} for rotating around the new center (x_g, y_g) by the calculated angle is given by:

$$M_{\text{rotate}} = \begin{bmatrix} \cos(\theta_{\text{rad}}) & -\sin(\theta_{\text{rad}}) & 0 \\ \sin(\theta_{\text{rad}}) & \cos(\theta_{\text{rad}}) & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & x_g \\ 0 & 1 & y_g \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & -x_g \\ 0 & 1 & -y_g \\ 0 & 0 & 1 \end{bmatrix} \tag{EQ 13}$$

Apply the rotation to the translated image using the affine transformation:

$$\begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} \cos(\theta_{\text{rad}}) & -\sin(\theta_{\text{rad}}) & 0 \\ \sin(\theta_{\text{rad}}) & \cos(\theta_{\text{rad}}) & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & x_g \\ 0 & 1 & y_g \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & -x_g \\ 0 & 1 & -y_g \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \tag{EQ 14}$$

Figures 3.15 and 3.16 visually demonstrate this manual knuckle alignment technique on samples from the DHV and JLU datasets, respectively. To garner a consistent result, one annotator was chosen to align all 3,680 JLU and 1,104 DHV images.

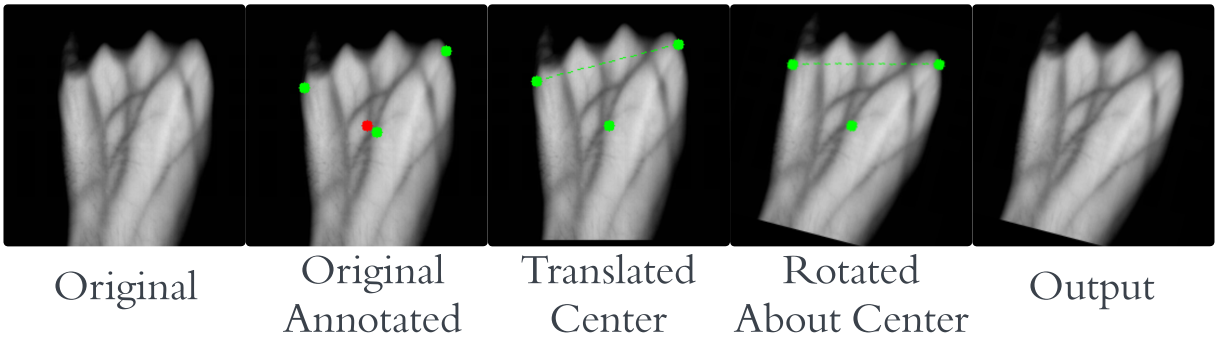

Figure 3.15: DHV Dataset Manual Alignment

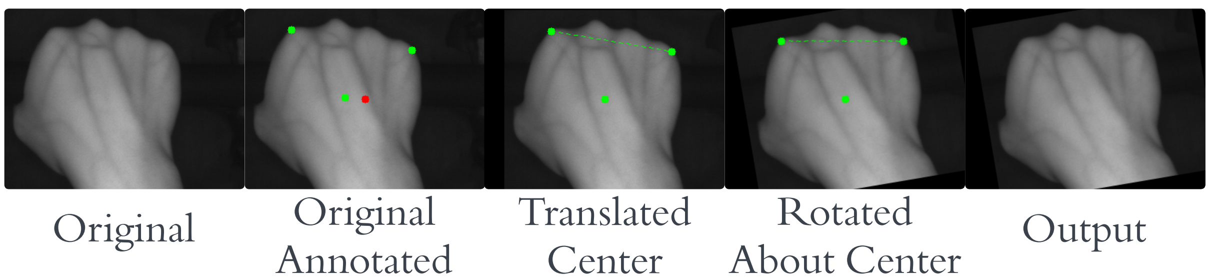

Figure 3.16: JLU Dataset Manual Alignment

Figure 3.17: Illustration of the manual knuckle alignment process for the DHV and JLU Datasets. The three green points indicate the perceived center of the hand and the left-most and right-most knuckles. The right images are the results of rotating and centering the images.

Proposed Automated Alignment

To eliminate the necessity for manual intervention in the alignment process, we introduce an automated technique that leverages sophisticated computer vision methodologies. This approach not only enhances scalability but also ensures consistency by removing operator dependency. Central to our method is the autonomous detection of knuckle points through advanced image processing algorithms, followed by image transformations that standardize their positions across all images. A key innovation in our approach is the utilization of the largest inscribed circle within hand images for alignment; a technique previously applied in vein pattern verification but primarily focused on extracted veins rather than the entire hand F. Liu et al., 2020b; Ruiz-Echeverri et al., 2021; Y. Wang et al., 2019.

Our method distinctively integrates this circle-based alignment with knuckle point alignment, a combination not extensively explored in the context of dorsal hand vein imaging Bhosale and Jadhav, 2017; Kumar and Prathyusha, 2009; Lee et al., 2014. By employing a two-step process, first centering and scaling images based on their largest inscribed circles, and then rotating them to achieve uniform knuckle orientation, we produce a consistently aligned dataset suitable for rigorous analysis.

Initially, we augment the images to accentuate the contrast between the hand and the background. This is achieved by implementing Contrast-Limited Adaptive Histogram Equalization (CLAHE) with a clip limit of 2.5 and a tile grid size of 21×21, thereby substantially enhancing image contrast. The outcome of this enhancement is depicted as the 'Enhanced Image' in Figures 3.18 and 3.19.

Subsequently, to mitigate noise and smooth the images, we apply a Gaussian blur with a kernel size of 27×27. This results in an intermediary image optimized for further processing stages (see the 'Blurred Image' in Figures 3.19 and 3.18).

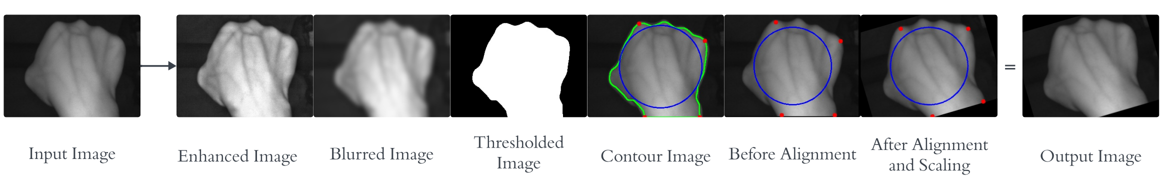

Figure 3.18: Illustration of the automated knuckle alignment process for the JLU Dataset. The input image is passed through a series of computer vision techniques such as CLAHE, Gaussian blur, and morphological operations. These techniques together automatically generate key points to properly align the images. Centering, scaling, and rotation are applied systematically to generate the aligned output image.

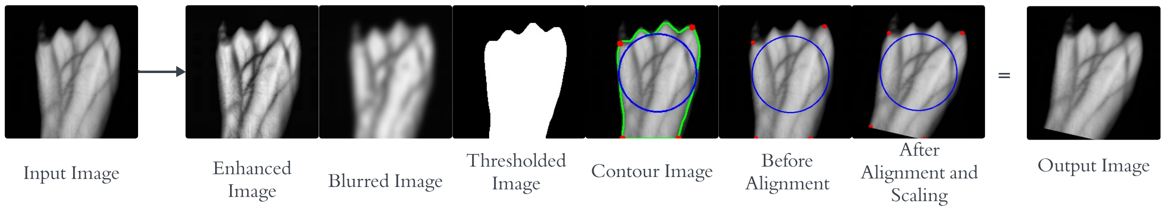

Figure 3.19: Illustration of the automated knuckle alignment process for the DHV Dataset. The input image undergoes several computer vision techniques including CLAHE, Gaussian Blur, and morphological operations to produce a binary or segmented image. From the segmented, binarized images, we find the center point of the hand and the two knuckle points. These points properly align the images by centering, scaling, and rotating.

After blurring, we employ Otsu's thresholding method Otsu et al., 1975 to segment the hand contour from the background. Morphological operations with a 5×5 kernel are then applied. Specifically, the opening operation is performed iteratively to remove small artifacts and enhance the integrity of the hand contour. The result is the 'Threshold Image' shown in Figures 3.18 and 3.19).

Following segmentation, we identify the largest contour by area, corresponding to the hand's outline. This contour is visualized as a green line in the 'Contour Image' in Figures 3.18 and 3.19. From this prominent contour, we determine the largest inscribed circle within the hand silhouette. This is achieved by applying a distance transform to the contour mask and calculating the Euclidean distance of each foreground pixel to the nearest background pixel. The pixel with the maximum distance signifies the center of the largest inscribed circle, and its distance value gives the radius. The circle is illustrated as a blue line in the 'Contour Image'.

Before applying alignment transformations, we ascertain the most distant points along the hand contour from the center of the largest inscribed circle in each of the four quadrants. The points in the top two quadrants correspond to the knuckle alignment points. These critical points are marked as red dots in the 'Before Alignment' and 'After Alignment' images in Figures 3.18 and 3.19.

Once the center of the largest inscribed circle is identified, we translate the image so that this center aligns with the center of the image frame. A translation matrix is defined to shift the circle's center to the image's center. After centering, we calculate the angle between the knuckle points using the arctangent function and convert it to degrees. A rotation matrix is then defined to rotate the image around its center by the calculated angle, effectively aligning the knuckles horizontally. The systematic application of the translation and rotation matrices results in the final aligned image (refer to Equations 9–14).

We establish a dataset where each image is uniformly centered, scaled, and rotated, facilitating reliable and consistent analysis in subsequent processing stages.

3.2.2 Model Architecture and Hyperparameters

We developed our dorsal hand vein biometric identification and verification system using the ResNet-18 architecture, known for its efficiency in image processing K. He et al., 2016. This section details the model architecture, training strategies, and key hyperparameters.

The ResNet-18 model was initialized with pre-trained ImageNet weights, allowing us to leverage the features learned from a broad and diverse dataset. Input images were resized to 224x224 pixels to conform to the model's input size requirements. To adapt the architecture for our specific task, we appended a fully connected linear layer with 512 units at the end of the network.

Later, this layer is vital for extracting the final embedding to compare the embedding of the query image and gallery image using cosine similarity (EQ 15). Given two 1x512 embedding vectors, one from the gallery G and one from the query Q, the cosine similarity is calculated as:

$$\text{Cosine Similarity}(G, Q) = \cos(\theta) = \frac{G \cdot Q}{\|G\|\|Q\|} = \frac{\sum_{i=1}^{512} G_i Q_i}{\sqrt{\sum_{i=1}^{512} G_i^2}\sqrt{\sum_{i=1}^{512} Q_i^2}} \quad \text{(EQ 15)}$$

ArcFace loss was selected for its ability to introduce an angular margin between classes, which significantly enhances the discriminative power of feature embeddings Deng et al., 2019. For both the JLU and DHV datasets, the batch size was set to 64. To optimize model performance, different learning rates and optimizers were tailored to each dataset. Specifically, the optimal learning rate was determined to be 0.003 for the JLU dataset and 0.0003 for the DHV dataset. Stochastic Gradient Descent (SGD) proved to be the most effective optimizer for the JLU dataset, while Adam was the best choice for the DHV dataset.

All models were trained for up to 500 epochs, with an early stopping mechanism implemented to prevent overfitting. The early stopping callback had a patience of 10 epochs, ensuring that training would cease if the validation loss failed to improve within this window.

3.3 Concealing Human Vein Identity: Effect of Image Degradation on Dorsal Hand Vein Recognition

To advance dorsal hand vein biometric systems, we conducted an ablation study to examine the system's robustness when vein visibility decreases. This study simulates physiological changes, such as increased subcutaneous fat, which can reduce vein visibility, and investigates the general effects of blurring degradation.

Before conducting the ablation study, we first established a dorsal hand vein authentication system for NIR-based images. The methodology for both the authentication system and the ablation study is outlined in Figure 3.20.

Please refer to the previous section, 3.2, for the discussion of the dorsal hand vein authentication system where we talk about model development and knuckle alignment.

What follows is a discussion of the ablation study.

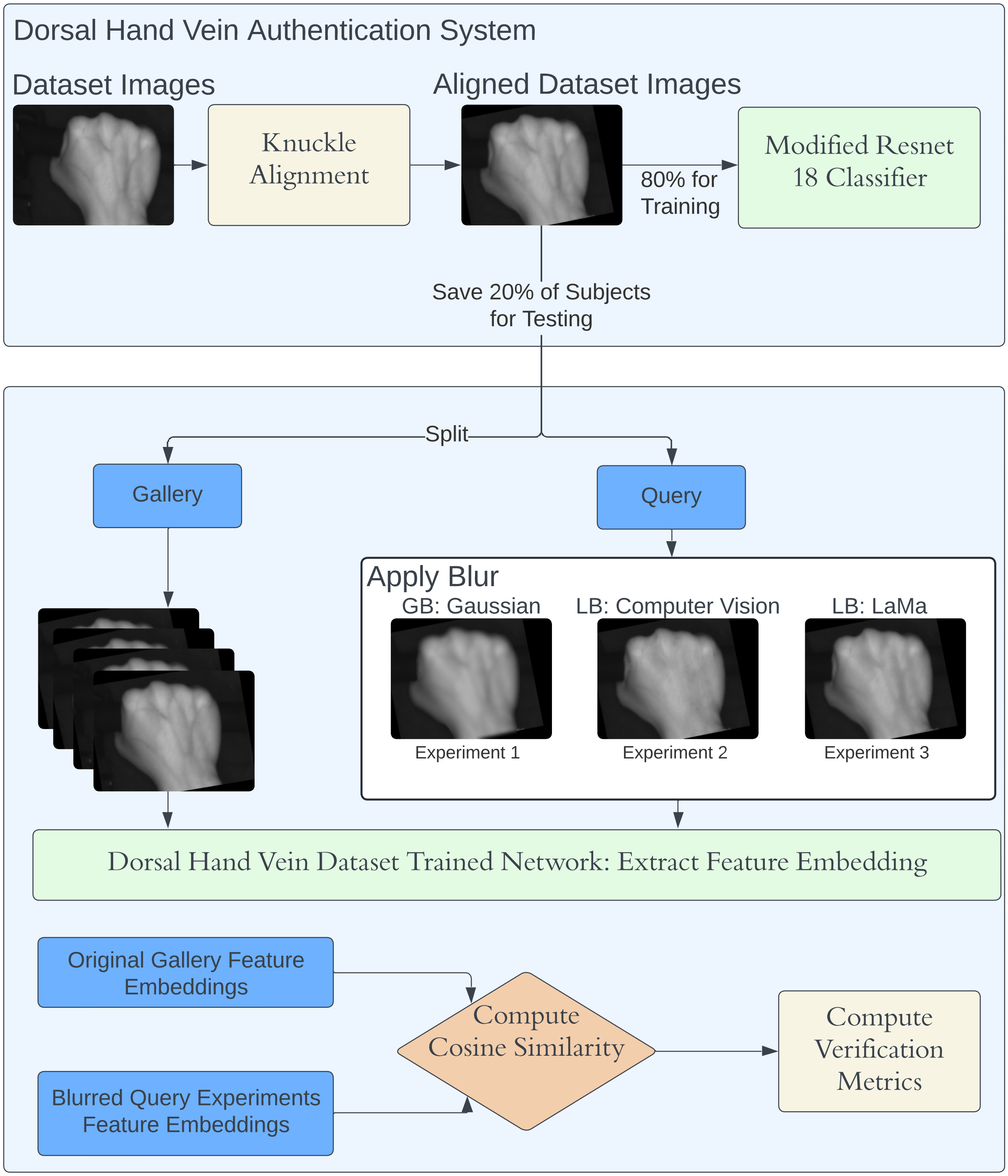

Figure 3.20: Methodology Diagram - Our methodology is split into two sections: our proposed dorsal hand vein authentication system from the previous chapter and our ablation experiments around types of blur. First, we train ResNet-18 as our feature extractor, and then we use it in our ablation study to extract feature embeddings for our gallery and query images. We apply our various ablation experiments to our query images.

3.3.1 Evaluating Global and Local Blurring in Dorsal Hand Vein Biometrics

This section outlines our investigative foray into two distinctive blurring approaches, global and local, to understand their impact on the system's performance.

Global blur simulates a scenario where the hand might be out of focus due to variations in camera focus or environmental conditions during image capture. This approach evaluates the system's ability to maintain identification performance as the hand image progressively loses sharpness. The objective of the global blur study is to assess the system's resilience to general blurring degradation and its impact on biometric accuracy.

The local blur methodology simulates scenarios where the subject experiences physiological changes in which increased subcutaneous fat could naturally obscure the visibility of the vein. By concealing the vein identities, the system challenges the system to extract other features relevant to their identities like general hand shape or skin textures. We target the veins using computer vision techniques to generate a vein mask from the black and white NIR. The vein mask is paired with the original image to virtually erase the veins via inpainting methods.

These methodologies aim to answer a pivotal question: If the vein patterns were concealed, obscured, or removed could the biometric system still discern and authenticate individuals based on other features?

This study reveals major implications for the security and robustness of the biometric system. If the system can still discern and authenticate individuals when vein patterns are blurred, obscured, or removed, it suggests that it relies on other features such as the shape of the hand or the texture of the skin. Although this enhances the system's resilience to variations in image quality or physiological changes, it also raises security concerns. If non-vein features are sufficient for authentication, a bad actor might exploit this by replicating those features through artificial means, potentially compromising the system. Therefore, understanding to what extent the system depends on these additional features is crucial to assessing vulnerabilities and strengthening biometric security measures.

Global Blurring

The clarity of vein patterns is very important in biometrics for several reasons:

- Uniqueness: Each person's vein pattern is unique, essential for distinguishing individuals.
- Accuracy of Recognition Systems: High clarity enables more accurate matches, reducing false acceptance and rejection rates.
- Robustness Against Forgery: Clear vein images are difficult to forge, enhancing security against spoofing attacks.
- Reduced Ambiguity: Clear images allow for effective application of feature extraction techniques, minimizing errors.
- Enhanced Consistency and Reliability: Clear patterns maintain system reliability despite external changes such as aging or scars.

Generally, biometric studies apply blurring techniques to the entire image. In this study, we differentiate our blurring methods by defining global blur as the blurring applied to the entire segmented hand and local blur as the blurring applied only to the veins. To simulate scenarios where both vein clarity and skin clarity might be reduced, we introduce our global blurring technique, which utilizes the Gaussian blur algorithm.

To simulate conditions where the clarity of vein and skin patterns is compromised, our global blurring technique is applied after the hand contour is extracted using the same method as in our alignment process. This technique uses the Gaussian blur algorithm to introduce ambiguity by diffusing the details within the entire hand region. Specifically, the blurring is only applied to the hand and not the background. The level of blurring is controlled by adjusting the kernel size, the larger the kernel, the greater the blurring effect. By varying the kernel size, we systematically analyze how different levels of blur impact the system's ability to accurately identify individuals based on their dorsal hand vein patterns.

This blurring approach simulates real-world conditions where factors such as motion blur, defocusing, or suboptimal image capture conditions may degrade image quality. The purpose of this investigation is to understand how susceptible the biometric system is to general blurring degradation and to evaluate its robustness in maintaining identification performance even when the image clarity is significantly reduced.

Local Blurring using Computer Vision Approach

Local blurring, which specifically targets veins within biometric images, is a sophisticated technique designed to assess a model's reliance on vein patterns for accurate verification. This approach simulates the effects of physiological changes, such as increased subcutaneous fat, which naturally diminish the visibility of these vein patterns. By focusing on the selective obscuration of vein structures through advanced computer vision methods, this technique offers insights into the model's robustness in the face of such challenges.

Central to our methodology is the development of an adaptive masking technique, which serves as the foundation for our subsequent inpainting process. This computer vision based approach precisely delineates the areas corresponding to vein patterns, enabling their targeted blurring or erasure. The process initiates with the extraction of the largest contour, following the contrast enhancement procedures detailed in the 'Knuckle Alignment' section. This contour serves as the primary region of interest, where the adaptive vein masking technique is applied to selectively obscure the vein structures.

The second key step in the vein isolation process is adaptive thresholding. Using the contour area, we apply Gaussian Adaptive Thresholding to extract veins. By calculating threshold values over localized pixel neighborhoods, we found that a block size of 21 and a constant parameter of 4 provided the best results.

After thresholding, the resulting binary mask represents vein structures but may include noise and small gaps. To refine this mask, we apply one iteration of morphological opening with a 1x1 kernel, followed by connected components analysis to remove insignificant regions, using an eight-connectivity parameter. Components smaller than 50 pixels are discarded to retain only significant vein patterns. Finally, two iterations of morphological closing with a 1x1 kernel are applied, followed by a Gaussian blur with a 3x3 kernel, resulting in a high-quality vein mask.

With the high-quality vein mask prepared, the local blurring method proceeds to the inpainting stage. We utilize two specific inpainting techniques: Navier-Stokes and Telea.

1. Navier-Stokes: This method, based on fluid dynamics, smoothly interpolates the masked areas using surrounding pixel information, effectively preserving natural image textures Temam, 2001. We applied it with an inpainting radius of 3.

2. Telea: Utilizing a fast marching technique, the Telea method propagates information from the nearest known pixels into the masked areas, resulting in sharper and more defined reconstructions

Telea, 2004. We applied the Telea method after Navier-Stokes with an inpainting radius of 6 to ensure smooth blending with surrounding pixels.

The final result is demonstrated in Figures 4.8 and 4.7 in the 'Local Blur' images. This computer vision-based approach leverages the mask to selectively target and conceal the vein patterns to simulate the images devoid of biometric vein features.

Local Blurring using LaMa

In our study of local blurring techniques, we utilize LaMa (Large Mask Inpainting), a state-of-the-art deep learning method designed to accurately reconstruct large missing areas in images. LaMa, as outlined by Suvorov et al., 2022, employs Fourier convolutions to effectively manage significant data loss while maintaining visual coherence and understanding complex image structures.

Our process begins similarly to the traditional computer vision approach, using hand contour extraction and adaptive masking to ensure precise identification of vein patterns. However, LaMa introduces a more advanced strategy for reconstructing these masked regions. Unlike conventional methods, LaMa benefits from pre-training on a diverse image dataset, which enables it to replicate intricate textures and structures in masked areas. [3]

During the inpainting process, LaMa processes the input image along with the associated vein mask as generated using the Gaussian Adaptive Thresholding method. The goal is to have pretrained LaMa model fill in the masked vein regions using the context of the unmasked parts of the hand. LaMa relies on the context provided by the surrounding pixels, there may be instances where veins are still faintly visible, particularly in low-contrast areas. The effectiveness of LaMa in obscuring veins while maintaining

[3] The pre-trained model utilized in this research can be accessed at: https://github.com/enesmsahin/simple-lama-inpainting/releases/download/v0.1.0/big-lama.pt

the natural appearance of hand textures is illustrated in Figures 4.8 and 4.7, showcasing the model's capability to smoothly integrate these modifications. This technique is valuable for assessing the resilience of biometric systems against deliberate obfuscation of critical features.

Chapter 4

Experiments

4.1 On Enhancing Crack Semantic Segmentation using StyleGAN and Brownian Bridge Diffusion

The following experiments were designed to evaluate the effectiveness of various deep learning models and techniques in the task of structural crack detection. Using the Bridge Crack Library (BCL) dataset as a foundation, we conducted a series of tests to optimize the DeepLabV3+ model through iterative hyperparameter tuning. The experiments systematically assessed the impact of different loss functions, learning rates, data augmentation techniques, and backbone architectures on the model's MeanIoU performance.

Additionally, we explored the use of synthetic datasets generated through the Brownian Bridge Diffusion Model (BBDM) and applied advanced ensemble modeling techniques to further enhance model accuracy and robustness. This section provides detailed insights into our experimental setup, the resulting performance metrics, and the effectiveness of the proposed methods in addressing the challenges of crack detection in structural health monitoring.

4.1.1 Bridge Crack Library Dataset

The Bridge Crack Library (BCL) dataset, developed by X.-W. Ye et al., 2021 in their seminal work on structural crack detection using pruned fully convolutional networks, represents a significant advancement in the domain of automated crack detection for in-service bridges. This comprehensive dataset comprises 11,000 pixel-wise labeled images of 256 by 256 resolution that were curated to include a diverse range of crack forms across various structural materials, including masonry, concrete, and steel. Notably, the dataset encapsulates 5,769 nonsteel crack images, 2,036 steel crack images, and 3,195 images categorized as noise, which were derived from the examination of over 50 bridges by experienced inspection teams over two years. Refer to back to Figure 3.2.

The BCL dataset was curated, with a comprehensive data collection process involving multiple cameras and covering a wide array of in-service bridges. During the annotations process, they performed pixel-wise annotations of the images using a specialized software and digital pens, which allowed annotators to accurately trace the contours of cracks.

The precisely annotated images in the BCL dataset provide a strong foundation for generating synthetic datasets. These synthetic datasets are vital for augmenting the diversity of training samples available for deep neural networks (DNNs).

4.1.2 Performance Metrics

We use several segmentation metrics to assess the performance of the model's accuracy and reliability.

- **Mean Intersection over Union (MeanIoU):** MeanIoU is especially relevant for segmentation tasks because it calculates the intersection over union for each class and then averages these values.

- **Precision:** Precision is defined as the ratio of correctly predicted positive observations to the total predicted positives. In the context of crack segmentation, it reflects the accuracy of the model in identifying true cracks as opposed to false positives. A higher precision indicates that the model is more effective in correctly labeling crack pixels, minimizing the instances where non-crack pixels are incorrectly classified as cracks.

- **Recall:** Recall measures the ratio of correctly predicted positive observations to the total actual positives. Recall score quantifies the model's ability to identify all actual crack instances in the images. High recall is crucial in applications like crack detection, as it ensures that the model captures as many true crack occurrences as possible, reducing the risk of missing critical defects. However, higher recall typically comes with the cost of an increased number of false positives.

- **Accuracy:** This metric is the ratio of correctly predicted observations (both true positives and true negatives) to the total observations in the dataset. While accuracy is a widely used metric, it does not provide a comprehensive assessment of performance, especially in the cases of imbalanced datasets where one class (e.g., cracks) is significantly underrepresented when compared to another (e.g., non-cracks).

- **F1-Score:** The F1-Score is the harmonic mean of precision and recall. It provides a balanced measure between these two metrics, making it particularly suitable for situations where there is an uneven class distribution, as is often the case in crack detection datasets.

4.1.3 DeepLabV3+ Hyperparameter Tuning Results

In our experiments, we iteratively introduced improvements to the DeepLabV3+ model and applied these enhancements cumulatively in subsequent tests. This approach allowed us to systematically assess the impact of each modification on the MeanIoU performance. Table 4.1 below provides a summary of these changes and their respective impacts.

Table 4.1: Summary of Hyperparameter Tuning for DeepLabV3+

Improvements	**Loss Function**	**Learning Rate**	**Augmentation**	**Backbone**	**Batch Size**	**MeanIoU (%)**	**Delta (%)**
Jin et al.: BCL + BCL 2.0	Cross-Entropy Loss	0.001	No	-	128	54.23	-
			MILAB Improvements				
Baseline	Cross-Entropy Loss	0.001	No	MobileNet	128	59.57	+5.34
Focal Loss	Focal Loss	0.001	No	MobileNet	128	59.93	+5.70
New Learning Rate	Cross-Entropy Loss	0.0003	No	MobileNet	128	58.86	+4.63
On the Fly Data Aug	Cross-Entropy Loss	0.001	Yes	MobileNet	128	62.44	+8.21
Aug + Dice Loss Variants	Dice Loss	0.001	Yes	MobileNet	128	59.98	+5.75
Aug + Focal Loss Variants	Focal Loss	0.001	Yes	MobileNet	128	61.57	+7.34
ResNet-50 Backbone	Cross-Entropy Loss	0.001	Yes	ResNet-50	128	62.12	+7.89
HRNetv2-48	Cross-Entropy Loss	0.001	Yes	HRNetv2-48	64	60.52	+6.29
HRNetv2-32, Pretrained	**Cross-Entropy Loss**	**0.001**	**Yes**	**HRNetv2-32**	**128**	**63.43**	**+9.20**

Our experiments began with a baseline configuration using MobileNet as the backbone, cross-entropy loss, a learning rate of 0.001, and no data augmentation. This setup resulted in a MeanIoU of 59.57%.

We first experimented with alternative loss functions, specifically focal loss and dice loss. The focal loss variant provided a slight improvement, achieving a MeanIoU of 59.93%, while the dice loss variant yielded 59.98%. However, these gains were marginal, indicating that the cross-entropy loss function remained a robust choice for this task. Even with data augmentation, the focal and dice loss functions did not significantly outperform cross-entropy, suggesting that the latter was the most consistent and effective for our purposes.

Next, we tested the impact of learning rate adjustments by reducing it from 0.001 to 0.0003. Contrary to expectations, this change led to a slight decrease in performance, with the MeanIoU dropping to 58.86%. Introducing on-the-fly data augmentation marked a significant turning point in our experiments.

By incorporating random scaling, cropping, flipping, rotation, and color jitter, the MeanIoU increased to 62.44%, representing an 8.21% improvement from the baseline.

After establishing the benefits of data augmentation, we explored the impact of different backbone architectures. The transition from MobileNet to ResNet-50 led to a MeanIoU of 62.12%, slightly below the augmented MobileNet configuration but still indicative of improvement. The most significant gain came with the introduction of HRNetv2 backbones. While HRNetv2-48 achieved a MeanIoU of 60.52%, the HRNetv2-32, when pretrained and combined with data augmentation, reached the highest MeanIoU of 63.43%. This result underscores the superior capability of HRNetv2-32 in capturing detailed features necessary for crack segmentation.

Comparing our results with those of Jin et al., 2023, who reported a MeanIoU of 54.23% using the BCL and BCL 2.0 datasets, our tuned model showed substantial improvements. The final configuration with HRNetv2-32, cross-entropy loss, and data augmentation outperformed the baseline by 9.20%, highlighting the effectiveness of our hyperparameter tuning strategy.

4.1.4 Using Hypertuned DeepLabV3+ for Semantic Segmentation of Generated Datasets

These results, in Table 4.2, highlight the capabilities of the BBDM in synthesizing realistic crack images and also demonstrate the effectiveness of various data augmentation and ensemble modeling techniques in improving MeanIoU.

The M-BBDM-P datasets augmented with zoom blur and defocus blur yielded impressive MeanIoU scores, with the zoom blur dataset achieving a leading score of 64.75%, closely followed by the defocus blur dataset at 64.14%. The images generated pixel space model resulted in better performance than the

Table 4.2: Testing Results for Synthesized Datasets
Model/Dataset descriptions include training datasets (excluding ensemble methods E) and model architecture. "P" and "L" in names indicate pixel and latent space models, respectively. Ensembles combine motion blur (MB), defocus blur (DB), and zoom blur (ZB), using voting, mean, or Or techniques. Red, pink, and yellow signify best, second-best, and third-best results. M-BBDM-P-ENSEMBLE-Voting showed the highest MeanIoU (65.62%) and F1 Score (71.04%), while M-BBDM-P-ENSEMBLE-Or led in Recall (83.20%) and F1 Score (71.79%). M-BBDM-L-DB excelled in Precision (73.51%). We propose M-BBDM-P-E-Or as the best model for its high recall and F1 score. More details about this proposed model can be found in the Proposed Method Discussion section. The proposed M-BBDM-L-DB model shows the second-highest accuracy in terms of Precision (73.51%).

Model/Dataset	**MeanIoU (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**F1 Score (%)**
Ye et al.	81.03	53.53	98.64	64.47	64.47
Jin et al.	54.23	62.35	77.10	98.72	68.91
BCL	63.30	68.46	78.33	98.32	70.60
BCL + BCL 2.0	63.43	73.16	70.94	98.46	69.38
M-BBDM-P	61.36	69.11	74.37	98.29	68.88
M-BBDM-P-ZB	64.75	70.59	75.44	98.44	70.59
M-BBDM-P-MB	63.35	72.95	70.37	98.46	69.02
M-BBDM-P-DB	64.14	67.28	79.88	98.36	70.82
M-BBDM-P-E-Voting	65.62	71.20	75.26	98.51	71.04
M-BBDM-P-E-Mean	65.33	71.39	75.48	98.50	71.14
M-BBDM-P-E-Or	63.89	66.01	83.20	98.27	71.79
M-BBDM-L	62.22	66.33	75.47	98.26	67.79
M-BBDM-L-ZB	63.82	72.79	70.59	98.47	69.17
M-BBDM-L-MB	63.43	72.61	70.46	98.48	68.85
M-BBDM-L-DB	64.01	73.51	69.64	98.48	68.77
M-BBDM-L-E-Voting	64.40	73.42	70.31	98.50	69.22
M-BBDM-L-E-Mean	64.38	73.28	70.23	98.51	69.16
M-BBDM-L-E-Or	64.61	70.90	75.95	98.43	71.01

images generated from the latent space model. Similarly, the augmented datasets showed enhancements in MeanIoU when compared to their counterparts.

Recognizing the comparable accuracies among the augmented models within each space, we investigated three ensemble modeling approaches to further elevate performance: majority voting, mean averaging, and logical 'Or' fusion. In majority voting, the ensemble prediction for each pixel is determined by the class most frequently predicted across the models. Mean averaging refines predictions by averaging the outputs from all models, thereby smoothing out potential anomalies. The logical 'Or' method enhances detection sensitivity by labeling a pixel as a crack if any model predicts it as such.

Employing ensemble modeling consistently resulted in improved MeanIoU and F1-Score metrics. Specifically, the majority voting ensembles achieved the highest MeanIoU in both the pixel and latent spaces, recording scores of 65.62% and 64.40%, respectively. Mean averaging and majority voting ensembles exhibited comparable performance across the evaluated metrics. Notably, the logical 'Or' ensemble attained the highest Recall and F1-Score. The final predictions resulting from these ensemble methods are illustrated in Figure 4.1.

The M-BBDM-L-DB model achieved a commendable Precision score of 73.51%, reflecting its effectiveness in minimizing false positives. The M-BBDM-P-E-Or model demonstrated the highest Recall (83.20%) and F1-Score (71.79%). The M-BBDM-P-E-Voting model exhibited the highest MeanIoU of 65.62%.

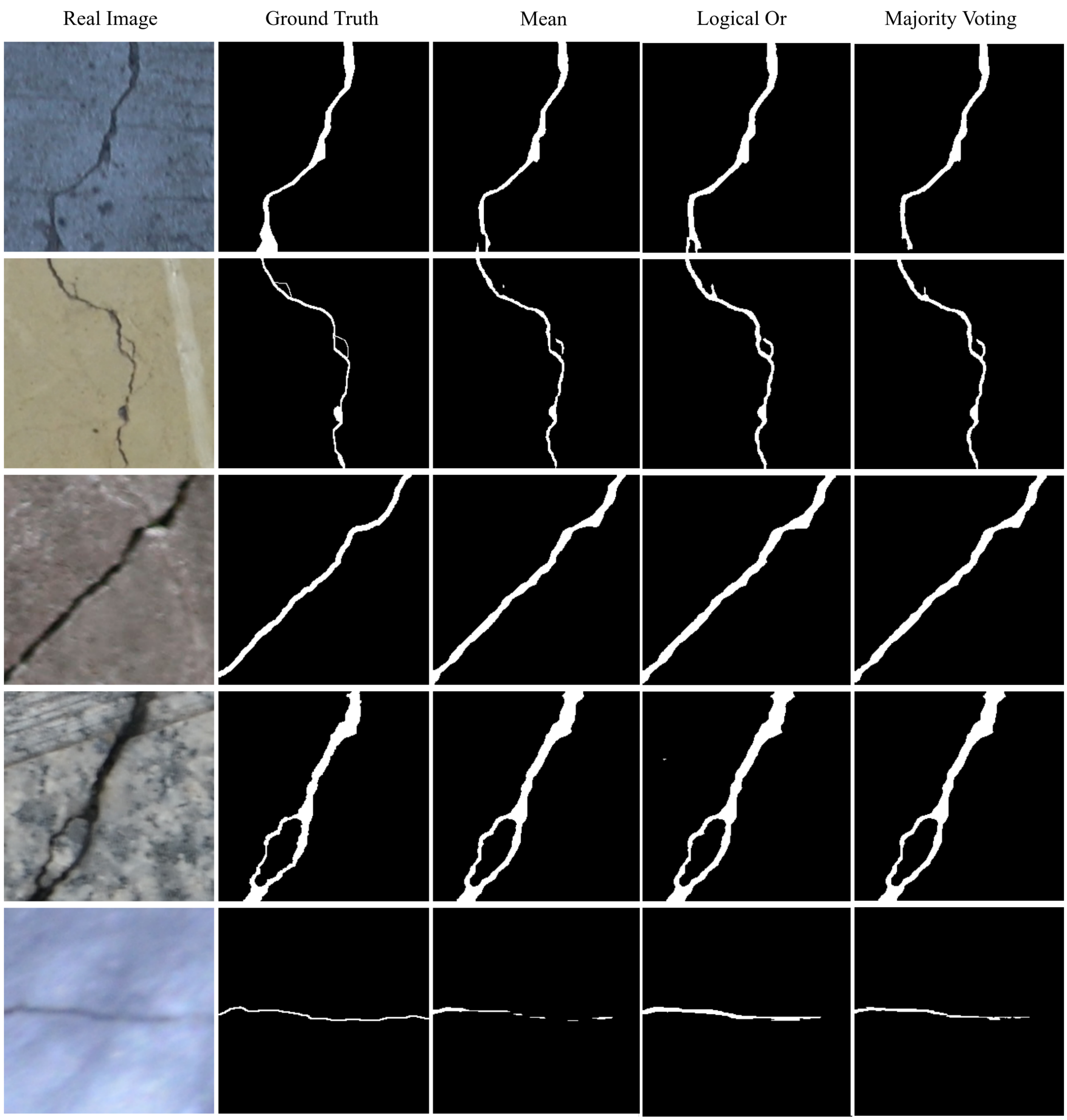

Figure 4.1: Final Prediction Results from Ensemble Modeling Synthesized BBDM Datasets. The mean, logical or, and majority voting column are the ensemble modeling predictions from the output of the DeepLabV3+ models. The ground truth image is the hand annotated image from the BCL dataset.

Proposed Method Discussion

Determining the most appropriate model for practical deployment necessitates a judicious balance of performance metrics tailored to specific operational requirements. For example, in critical infrastructure monitoring, such as assessing the integrity of bridges, dams, or nuclear facilities, the consequences of undetected cracks can be catastrophic. In this context, the M-BBDM-P-E-Or model is particularly suitable, achieving the highest Recall rate of 83.20%, thereby maximizing crack detection even if this results in an increased number of false positives.

Conversely, in situations where false positives carry significant costs or consequences, such as in automated inspection systems where each detected crack triggers a costly manual verification or halts production, minimizing false alarms becomes paramount. Here, models with higher Precision are preferable. Models like M-BBDM-P-E-Mean and M-BBDM-P-E-Voting, which deliver elevated Precision, MeanIoU, and robust F1-Scores, are advantageous choices.

If the goal is to balance precision and recall, the F1 Score becomes a key metric. The M-BBDM-P-E-Or model also excels here, with a leading F1 Score of 71.79%, indicating an effective balance between detecting cracks and minimizing false positives. If the goal is to achieve a balance between Precision and Recall, the F1-Score serves as a pivotal metric. The M-BBDM-P-E-Or model excels here as well, with a leading F1-Score of 71.79%, indicating an effective trade-off between detecting true positives and limiting false positives.

The M-BBDM-P-E-Or model stands out as the optimal choice for applications requiring high detection rates coupled with a balanced performance profile, owing to its exceptional Recall and F1-Score.

4.2 On Designing A Near Infrared Dorsal Hand Vein Authentication System

In our experiments, we assess our proposed alignment and matching approaches using the JLU and DHV datasets by performing a set of verification and identification experiments. Specifically, we conducted a comprehensive evaluation of our state-of-the-art knuckle alignment method, benchmarking our results against existing approaches within the domain of biometric recognition using dorsal hand veins.

Below, we first discuss our experimental configuration, including a detailed breakdown of the datasets, the performance metrics chosen, and the similarity score selected. Then, we discuss our two key experiments:

1. Evaluate the impact of optimization studies and hyperparameter tuning.
2. Assess the effectiveness of our proposed automated and manual knuckle alignment method on our optimized model.

What follows is a discussion of the performance metrics, datasets, and the results of these two experiments.

4.2.1 Experimental Configuration

Datasets

Jillian University Dorsal Hand Vein Dataset The Dorsal Hand Vein Database from Jilin University (JLU) F. Liu et al., 2020b provides an extensive collection of dorsal hand vein images captured using Near-Infrared (NIR) imaging technology. This dataset encompasses 736 unique participants, each contributing five images of their right-hand dorsal veins under normal conditions. For our study, we allocated data from 588 individuals (totaling 2,940 images) for training purposes, while the remaining 148 individuals (740 images) were reserved for testing and validation. Each image possesses a resolution of 320×240 pixels, offering sufficient detail for vein pattern analysis. During the acquisition process, subjects were instructed to grasp a cylindrical bar positioned beneath an overhead NIR camera. To introduce variability and simulate real-world conditions, participants adjusted their grip or altered their hand positioning between captures, ensuring that the dataset reflects natural variations in hand placement and orientation.

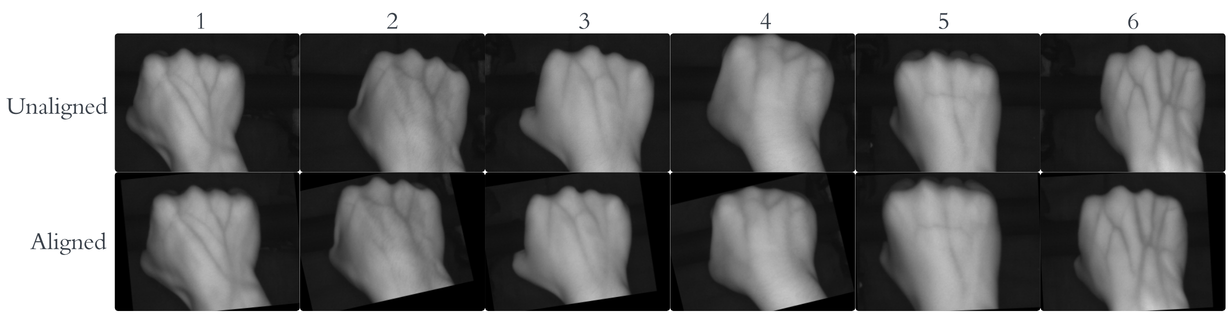

Figure 4.2: Knuckle alignment samples from the JLU dataset. Note the variety in hand position and rotation of the unaligned images. The unaligned images are in the top row and the aligned image counterpart is in the bottom row. The aligned images are generally the same size after scaling and centering, and the knuckles are generally in the same position.

Figure 4.2 illustrates a side-by-side comparison between the original unaligned images and those processed through our alignment algorithm. The upper row showcases the raw images as captured, revealing substantial variation in hand placement and rotational orientation among different samples. The lower row presents the images post-alignment, where the hands exhibit consistent orientation and positioning. This comparison underscores the efficacy of our automated alignment technique in standardizing the dataset for subsequent analysis.

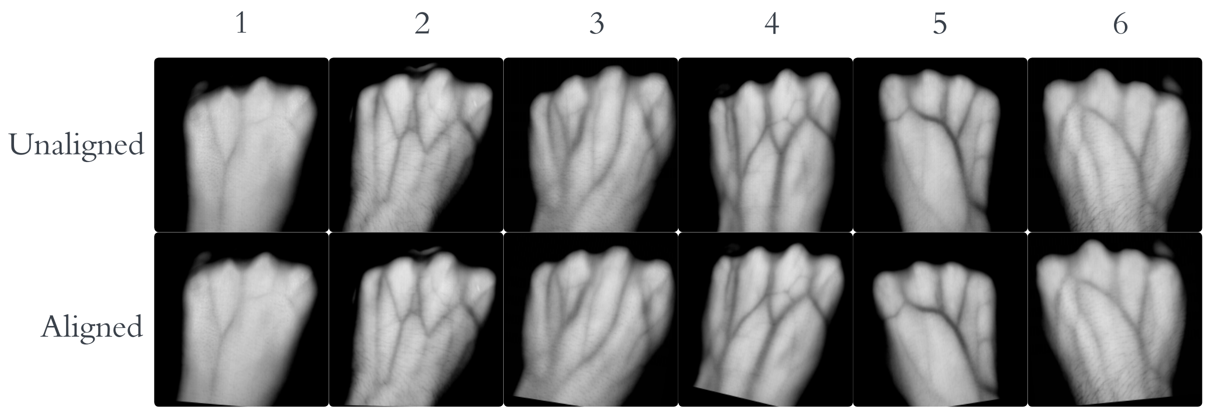

Figure 4.3: Knuckle alignment samples from the DHV dataset. The DHV dataset is already pre-centered so the unaligned data is very uniform. The unaligned images are in the top row and the aligned image counterpart is in the bottom row. After alignment, the knuckles are in the same position in all photos, but the hand is rotated slightly. Notice that the left hands, samples one to four, are rotated opposite of the right-hand samples, five and six.

Dorsal Hand Vein Dataset The Dorsal Hand Vein (DHV) Database Wilches-Bernal et al., 2020 offers an open-access repository of dorsal hand vein images captured using a custom-built infrared imaging system with specialized LED illumination optimized for vein visibility. The original images, with a high resolution of 752×560 pixels, were downscaled to 224×224 pixels to ensure compatibility with our ResNet-based neural network architecture.

The dataset comprises two separate acquisition sessions; however, to maintain balance and consistency, we focused exclusively on session one, as not all subjects participated in both sessions. Session one includes 138 individuals, each providing four images per hand, resulting in a total of 1,104 images. To expand the dataset's biometric diversity, we treated each hand (left and right) as a distinct subject, yielding 276 unique biometric identifiers. Of these, 220 identifiers (comprising 880 images) were allocated for training, while the remaining 56 identifiers (224 images) were reserved for testing and validation.

In contrast to the JLU dataset, which exhibits considerable variability in hand positioning and orientation, the DHV dataset features images that are relatively pre-aligned and centered. This inherent consistency allowed us to assess the generalizability of our alignment techniques, initially developed for the more variable JLU dataset when applied to datasets with different characteristics.

Figure 4.3 demonstrates the difference between unprocessed and aligned images from the DHV dataset. The top row displays the unaligned images, which already exhibit a high degree of uniformity due to the standardized acquisition process. The bottom row showcases the images after alignment, highlighting the enhanced consistency achieved through our preprocessing steps.

Our alignment methodology assumes that the wrist is located at the bottom of the image frame, with the knuckles positioned at the top. This assumption accommodates variations in hand pose and orientation without necessitating additional geometric normalization procedures.

Similarity Measurement

During the comparison phase, we extracted feature embeddings from the final layer of our modified ResNet-18 neural network for each image, thereby capturing high-level, discriminative representations of the input data. We selected cosine similarity as our metric of choice due to its demonstrated effectiveness across various dorsal hand vein datasets Kocakulak et al., 2023. Cosine similarity, as defined in Equation EQ 15, quantifies the cosine of the angle between two non-zero vectors, effectively measuring their directional alignment independent of vector magnitude.

Using the feature embeddings from the testing datasets, we computed all possible pairwise cosine similarity scores among embeddings belonging to the same individual (genuine comparisons) and between embeddings from different individuals (impostor comparisons). For the JLU database, this resulted in

1,480 genuine comparisons and 271,950 impostor comparisons. For the DHV database, we obtained 336 genuine comparisons and 24,640 impostor comparisons.

Performance Metrics

Our analysis evaluates both verification and identification performance using several key metrics:

1. **Verification Performance**:
   - Cosine Similarity Analysis: We assessed accuracy by analyzing the distribution of cosine similarities between genuine and imposter pairs. This metric measures the similarity between two embedding vectors, with higher values indicating greater similarity.
   - Equal Error Rate (EER): EER is calculated as the point where the false acceptance rate (FAR) equals the false rejection rate (FRR). It provides a single value that summarizes the trade-off between these two types of errors. A lower EER indicates better overall performance.
   - Verification Accuracy: This metric represents the proportion of correctly verified instances at a specific threshold, typically the threshold corresponding to the EER. Verification accuracy provides a direct measure of the model's ability to correctly verify identities.
2. **Identification Performance**:
   - Cumulative Match Characteristic (CMC) Curve: We used the CMC curve to evaluate identification performance. The curve shows the probability that the correct identity appears within the top t matches, with **Rank-1** being the probability that the correct identity is the top match. Higher Rank-1 scores indicate better identification accuracy.

3. **Overall Model Performance**:

- Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC): We calculated the AUC of the ROC curve to provide an overall assessment of the model's performance. The ROC curve illustrates the trade-off between the true positive rate (TPR) and the false positive rate (FPR). The AUC represents the model's ability to distinguish between classes, with values closer to 1.0 indicating superior performance.

4.2.2 Experiment 1: Classifier Hyperparameter Tuning

We performed extensive hyperparameter tuning on our Modified ResNet-18 model to optimize its performance as a feature extractor. This involved systematically testing various configurations of learning rates, batch sizes, optimizers, and loss functions using a grid search approach. The goal was to identify the optimal settings that maximize EER and verification accuracy for the JLU and DHV Datasets.

We have summarized the hyperparameter tuning outcomes in Table 4.3 and 4.4.

Table 4.3: Summary of Hyperparameter Tuning Results for Auto Alignment Method on JLU Dataset. The last row displays the hyperparameters that garnered the best results after an exhaustive grid search hyperparameter tuning of the learning rate, batch size, optimizer, and loss function.

Learning Rate	Batch Size	Optimizer	Loss Function	EER (% ↓)	Accuracy (% ↑)	EER Diff. (% ↓)	Acc. Diff. (% ↑)
0.03	128	Adam	Arcface	11.03	87.56	-	-
0.03	128	Adam	Softmax	11.16	91.99	-0.13	4.43
0.003	64	Adam	Softmax	6.35	94.99	-4.81	3.00
0.003	**64**	**SGD**	**Arcface**	**1.14**	**99.07**	**-5.21**	**4.08**

Table 4.4: Summary of Hyperparameter Tuning Results for Auto Alignment Method on DHV Database. The last row displays the best hyperparameters that give the best model performance for EER and verification accuracy on the DHV database.

Learning Rate	Batch Size	Optimizer	Loss Function	EER (% ↓)	Accuracy (% ↑)	EER Diff. (% ↓)	Acc. Diff. (% ↑)
0.03	128	Adam	Softmax	40.77	59.95	-	-
0.03	128	SGD	Softmax	1.26	98.74	-39.51	39.79
0.0003	128	Adam	Arcface	0.49	99.51	-0.77	0.77
0.0003	**64**	**Adam**	**Arcface**	**0.10**	**99.90**	**-0.39**	**0.39**

The hyperparameter tuning results demonstrate that specific configurations significantly enhance biometric verification performance across the JLU and DHV datasets.

For the JLU dataset, the optimal setup combined a learning rate of 0.003, a batch size of 64, the SGD optimizer, and the Arcface loss function. This configuration achieved a low EER of 1.14% and a high accuracy of 99.07%.

For the DHV dataset, the optimal settings included a learning rate of 0.0003, a batch size of 64, the Adam optimizer, and the Arcface loss function. This configuration achieved an exceptional EER of 0.10% and a verification accuracy of 99.90%. The low learning rate and the use of the Adam optimizer likely contributed to the model's ability to capture the intricate vein patterns present in the DHV dataset.

Across both datasets, a batch size of 64 emerged as the optimal choice, suggesting that this size provides a good balance between computational efficiency and the stability of gradient estimates. The Arcface loss function was consistently used in the best-performing configurations, reinforcing its capability to enhance feature discrimination significantly. SGD and Adam proved to be effective optimizers in their respective contexts.

4.2.3 Experiment 2: Knuckle Alignment

To further enhance the accuracy of our dorsal hand vein biometric system, we conducted a comprehensive examination of the efficacy of automated and manual knuckle alignment methodologies.

We applied the best-performing models from our hyperparameter tuning experiments, detailed in Tables 4.4 and 4.3, to assess the impact of alignment on system performance. Each model was trained using its optimal settings on its respective dataset, ensuring a fair and effective evaluation of the alignment methods.

The alignment procedure is designed to homogenize the presentation of hand images within the biometric system by precisely aligning the knuckles and standardizing the orientation of the hands. To provide a baseline for comparison, we also assessed the performance of the hyperparameter-tuned models on datasets without any alignment adjustments.

The results of our knuckle alignment tests are encapsulated in Table 4.5 and further validated by the ROC curves shown in Figure 4.6. These curves provide a visual representation of the trade-off between true positive rates and false positive rates giving a better understanding of total model performance.

Table 4.5: Knuckle Alignment Results. Results demonstrate the effectiveness of different alignment methods. The 'No Tuning, No Alignment' method refers to no hyperparameter tuning or alignment applied to get a baseline performance. KA means 'Knuckle Alignment.' Lower EER and higher AUC, Rank-1, and Verification Accuracy values indicate improved performance, with 'Proposed Manual' showing the best overall performance across all metrics.

Method	**JLU Dataset**				**DHV Dataset**			
	EER (%)	**AUC (%)**	**Rank-1 (%)**	**Verif. Acc. (%)**	**EER (%)**	**AUC (%)**	**Rank-1 (%)**	**Verif. Acc. (%)**
No Tuning, No Alignment	18.70	89.76	30.41	74.97	5.66	98.53	91.07	93.66
Tuning, No Alignment	1.42	99.87	100	99.09	0.29	99.99	100	99.62
Proposed Automated KA	1.14	99.91	100	99.07	**0.1**	99.99	100	99.9
Proposed Manual KA	**0.74**	**99.97**	**100**	**99.5**	0.29	99.99	100	99.71

For the JLU dataset, the manual alignment method achieved the lowest EER of 0.74% and the highest verification accuracy of 99.50%. The automated alignment method also performed exceptionally well, with an EER of 1.14% and a verification accuracy of 99.07%. Both alignment methods significantly improved performance compared to the hyperparameter-tuned model without alignment, which had an EER of 1.42% and a verification accuracy of 99.09%.

For the DHV dataset, the automated alignment method achieved the lowest EER of 0.10% and the highest verification accuracy of 99.90%. The manual alignment method also maintained strong performance with an EER of 0.29% and a verification accuracy of 99.71%. Compared to the hyperparameter-tuned model without alignment (EER of 0.29% and verification accuracy of 99.62%), the automated alignment method provided a marginal improvement. This suggests that the DHV dataset already exhibits low variability due to the controlled acquisition conditions and preprocessing techniques employed by Wilches-Bernal et al., 2020, such as the contrast-enhancing control system.

To investigate deeper into this concept, we conducted an extensive visual examination of both manual and automated alignment methods to assess the variability in hand positioning across the datasets. Figures 4.4 and 4.5 offer valuable insights into the distribution of hand center locations after annotation using our proposed methods.

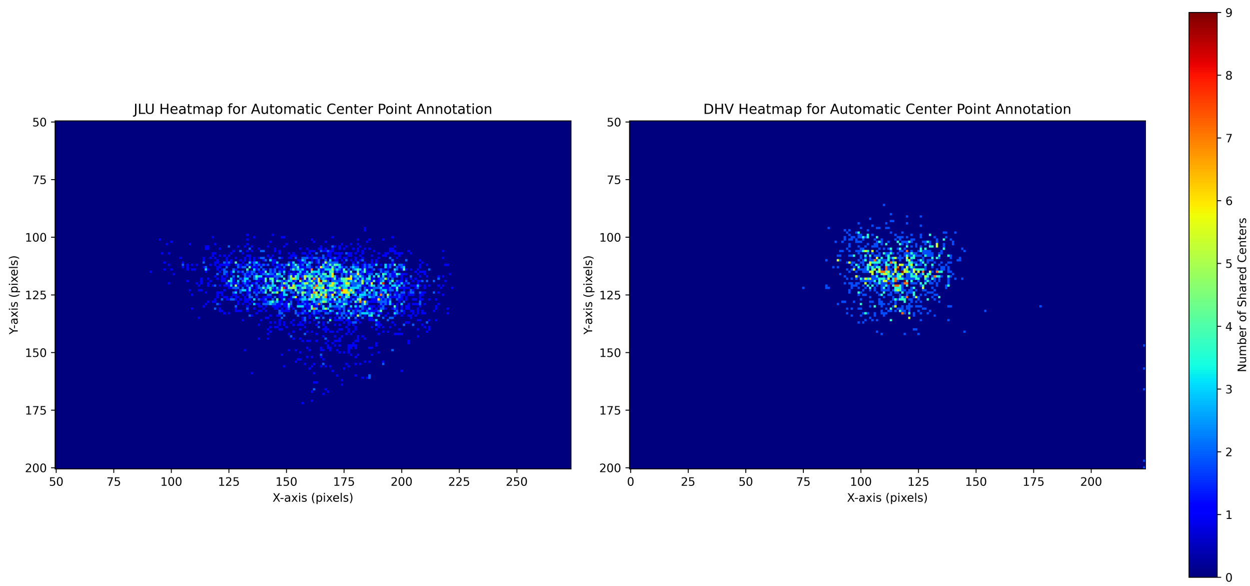

Figure 4.4: Automatic Center Point Acquisition for the JLU and DHV Datasets. These heatmaps depict the pixel locations of hand centers derived using our inscribed circle technique. The DHV heatmap shows the frequency and consistency of the circle centers, while the JLU heatmap indicates a more horizontal spread of hand positions around the y-axis at 120 pixels.

Following the annotation process, we extracted the pixel coordinates of the hand centers from both alignment methods. In the automated approach, the centers were determined using our inscribed circle

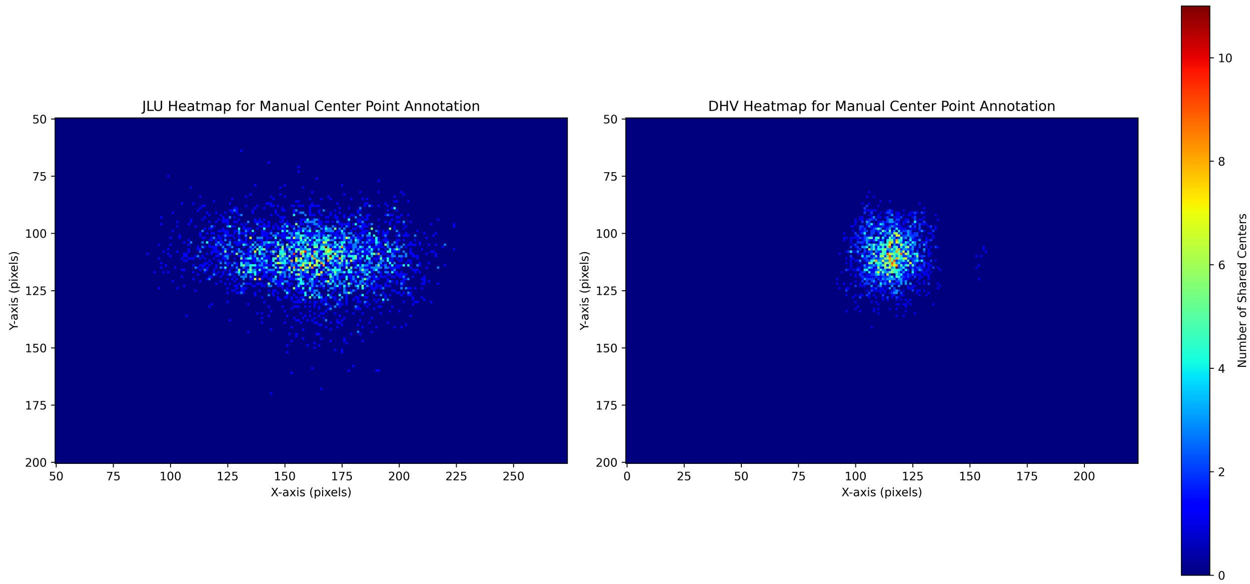

Figure 4.5: Manual Center Point Annotation for the JLU and DHV Datasets. These heatmaps display the locations of perceived hand centers as annotated manually. The DHV dataset shows a more localized distribution compared to the JLU dataset, which exhibits greater variability in hand positions.

technique, whereas in the manual method, the centers were selected based on the annotator's perception of the hand's central point.

The distributions of the hand center pixel locations exhibit similar patterns between the two methods, suggesting that our automated technique effectively identifies hand centers. Notably, in the DHV dataset, the manually extracted centers are more tightly clustered than those obtained automatically. Conversely, for the JLU dataset, the centers derived from the automated alignment are slightly more concentrated along the y-axis compared to the manual method, with fewer outliers below the y-axis value of 100. We attribute these discrepancies to the influence of knuckle contours when utilizing the largest inscribed circle; specifically, the process of detecting knuckle contours can cause the calculated hand centers to shift downward relative to the manual annotations. The DHV has a much more tight concentration when compared to the JLU dataset. This characteristic likely explains the exceptionally high initial performance metrics achieved after hyperparameter tuning with the DHV data, as the dataset's images are predominantly pre-centered, leading to superior baseline results. The JLU dataset exhibits greater variability in hand positioning, and consequently, benefits more substantially from the alignment process, which yields a marked improvement in accuracy.

The Receiver Operating Characteristic (ROC) curves presented in Figure 4.6 corroborate the high AUC values detailed in Table 4.5, affirming the robust identification capabilities of both alignment approaches. Collectively, the hyperparameter-tuned models achieve remarkably high AUC scores, with the manual method demonstrating superior performance as evidenced by its AUC values approaching 100%.

While the automated alignment method yielded only a marginal enhancement in EER on the low-variance DHV dataset, it produced a substantial positive effect on the JLU dataset, which exhibits higher intra-class variability. This outcome underscores the automated method's proficiency in achieving high accuracy, positioning it as a practical and efficient substitute for manual alignment.

We conducted a brief time comparison study to evaluate the trade-off between the time to align for the manual and automated methods. The results are summarized in Table 4.6.

Table 4.6: Time Comparison for Image Alignment of JLU Dataset. The JLU dataset contains 3,680 images. KA means 'Knuckle Alignment.'

Method	**Time Taken (seconds)**
Proposed Manual KA (Extrapolated)	20,240
Proposed Automated KA	15

Manually aligning 10 images required approximately 55 seconds. Extrapolating this rate to encompass the entire dataset of 3,680 images suggests a total processing time of roughly 5 hours and 38 minutes. Our proposed automated alignment method processed all 3,680 images in merely 15 seconds.

Having established the efficacy of our automated alignment method in terms of time efficiency, scalability, and performance, we proceeded to benchmark our optimal automated alignment strategy against

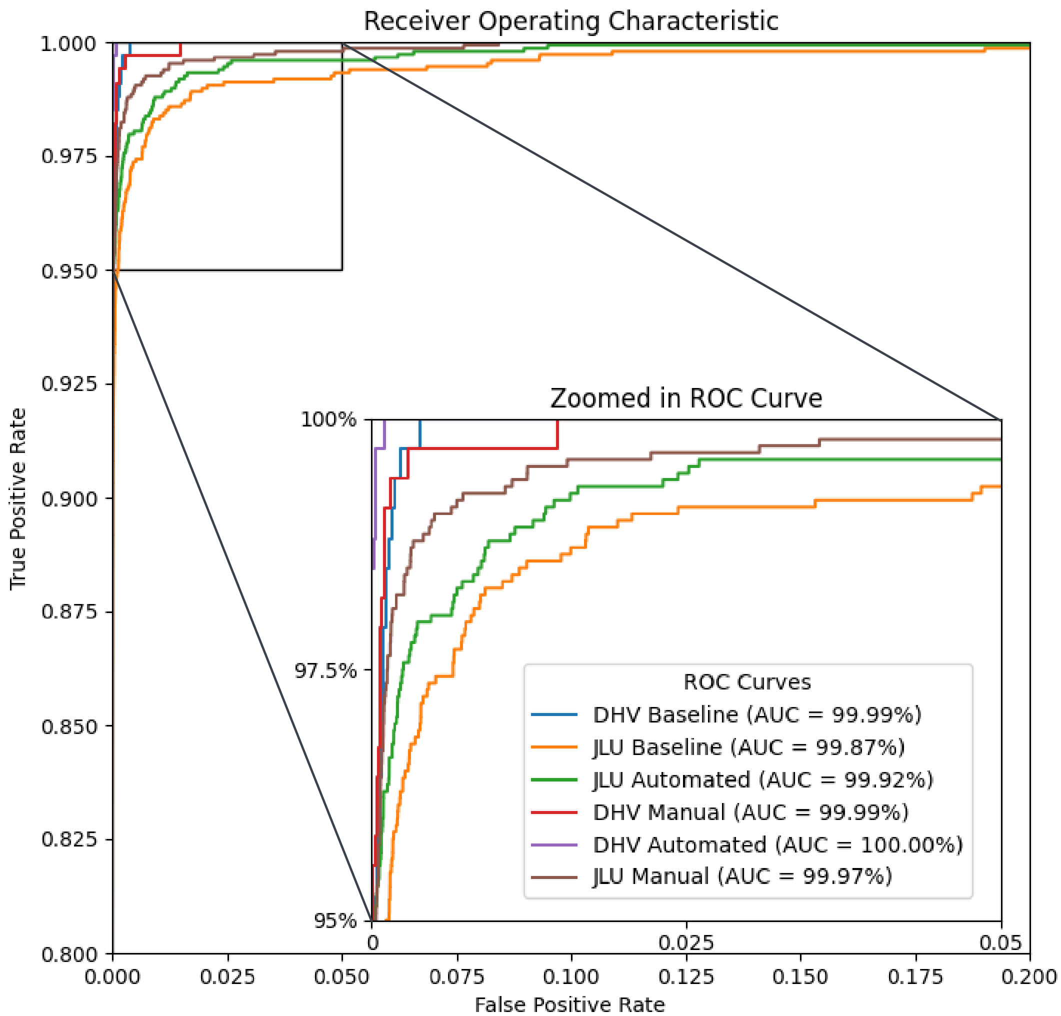

Figure 4.6: Knuckle Alignment ROC Curve Results. This ROC curve compares the performance of different classifiers, showing that the DHV Automated model achieves the highest AUC with 100.00%, indicating a perfect classification. The DHV Baseline also performs exceptionally well with 99.99% AUC. The JLU models show slightly lower, yet high performance with the Automated and Manual versions scoring 99.92% and 99.97% AUC, respectively, and the JLU Baseline at 99.87% AUC. The zoomed-in inset highlights the high true positive rate at low false positive rates for all models. The term 'Baseline' refers to the same results from 'Tuning, No Alignment' in Table 4.5.

other dorsal hand vein methodologies. Table 4.7 presents our competitors performances in terms of verification. To our knowledge, all competitors that used these public datasets for their studies did not report their identification scores.

Table 4.7: Comparison of automated alignment method with competitors on JLU and DHV datasets. Verification accuracies are reported unless noted otherwise (* denotes Classification Accuracy). Our automated knuckle alignment (KA) method significantly outperforms other automated methodologies in terms of EER across both datasets. Our verification accuracy ranks amongst the highest against all competitors except for those that chose to do dorsal hand vein classification. It is important to note that none of the competitors have reported identification scores on these datasets.

Dataset	**Author**	**Method**	**Metrics (%)**	
			EER	**Verif. Acc.**
JLU	**Our Work**	Automated KA	**1.14**	**99.07**
	Alashik and Yildirim, 2021	DL-GAN	2.47	98.36
	F. Liu et al., 2020b	IBGM	2.02	-
	R. Zhang et al., 2023	Soft-Argmax	5.81	99.5*
	F. Liu et al., 2020a	MOLPQ	-	98.07
	Nour et al., 2024	CNN & RBM	-	99.75*
DHV	**Our Work**	Automated KA	**0.1**	**99.9**
	Kocakulak et al., 2023	SA-CNN	2.17	-
	Nour et al., 2024	CNN & RBM	-	99*

Overall, our method obtains higher verification accuracies and lower EER except when compared to those that reporting classification accuracy's instead of verification metrics.

4.3 Concealing Human Vein Identity: Effect of Image Degradation on Dorsal Hand Vein Recognition

In this section, we conducted a series of experiments to assess the impact of these global and local blurring methods on the performance of dorsal hand vein biometric systems.

For details regarding experimental configuration such as model development, performance metrics, and dataset information refer to section 4.2.1. For quantitative baseline results, refer to section 4.2.3.

4.3.1 Ablation Study: Global and Local Vein Blurring

This section presents the results of our ablation study, demonstrating the effects of global and local blurring on the DHV and JLU datasets.

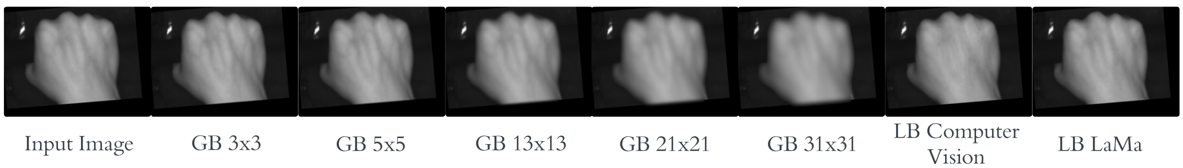

Figure 4.7: Blurring Image of JLU. 'GB' means Gaussian blur. 'LB' means Local Blur. This series demonstrates the effect of increasing Gaussian blur kernel sizes from 3x3 to 31x31 pixels applied to the whole image. The local blurring images represent blurring only the veins and leaving the rest of the hand unobstructed. Computer vision techniques such as Navier Stokes and Telea were used to inpaint the computer vision local blur image, while a pre-trained deep learning inpainting model was used to erase the veins in the 'LaMa' image.

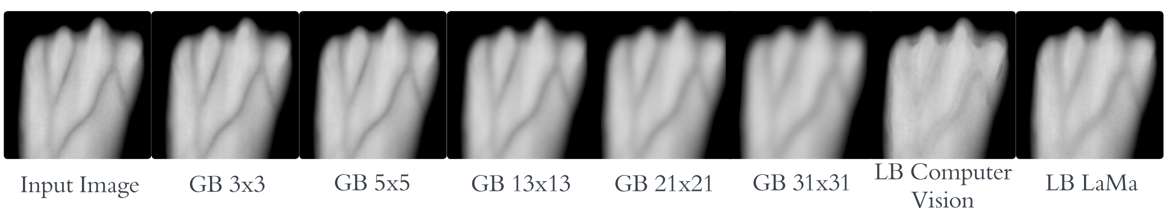

Figure 4.8: Blurring Image of DHV. 'GB' means Gaussian blur. 'LB' means Local Blur. This image sequence from the DHV dataset showcases the application of Gaussian blur across a spectrum of kernel sizes, from 3x3 to 31x31 pixels, to demonstrate the effects of varying degrees of blurriness on biometric analysis. Additionally, localized blurring techniques are employed to selectively obscure vein patterns, illustrating their impact on the accuracy of feature extraction processes critical in biometric verification systems.

Our study examines the impact of vein pattern visibility on the accuracy of biometric verification systems, specifically investigating how altering vein visibility in query images affects identity verification. To assess this, we applied global and local blurring techniques to the query images and evaluated the effects on system performance metrics. In Figure 4.9, we compare our locally blurred CV and LaMa images with the original aligned images. We also included Figure 4.10 to show off natural hands with low vein visibility.

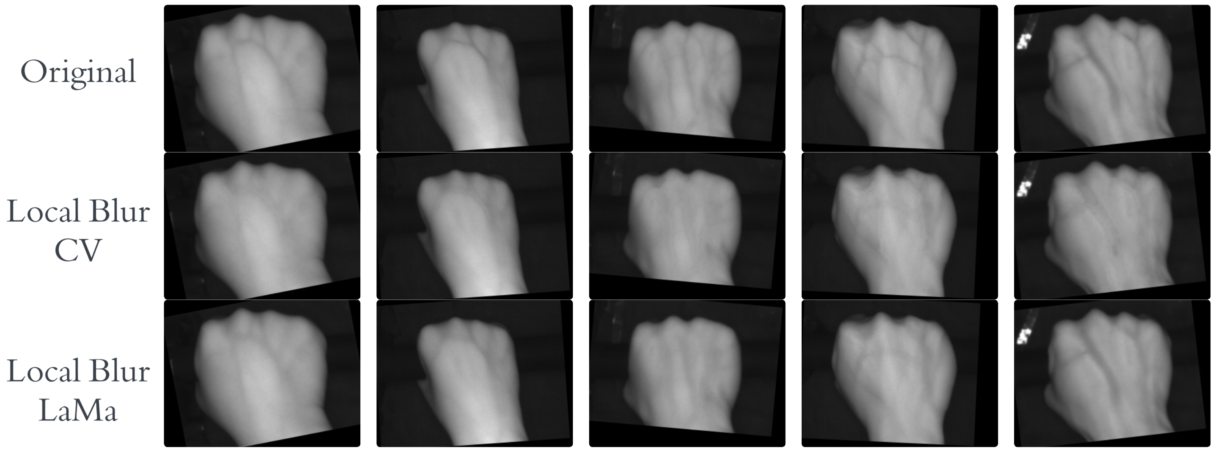

Figure 4.9: Comparison of locally blurred CV and LaMa images with the original aligned images. The first row shows the original aligned hand images. In the second row, the application of the computer vision blurring method significantly reduces vein visibility, with some veins becoming almost completely obscured. In contrast, the third row presents results from LaMa, which attempted to inpaint the hands but retained much of the vein outlines, especially around larger veins, though with additional blurring in the surrounding areas. Notably, larger veins were more resistant to blurring, as seen in the leftmost column.

Figure 4.10: Natural hand images with low vein visibility. In some cases, the visibility of smaller veins, which typically run horizontally, is significantly reduced or entirely obstructed. Larger veins tend to be the only discernible features, while the finer, more intricate vein patterns are completely obscured.

During our experiments, we applied blurring effects to a single query image per subject within the testing dataset, leaving gallery images unaltered. We also extended our analysis to include ablation experiments on the proposed knuckle alignment method and the 'Tuning, No Alignment' method, as outlined in Table 4.5. Refer to Table 4.8 for ablation study results.

General Observations The table above shows a clear trend for the global blurring tests: as kernel size and blurring intensity increase, performance consistently declines. Despite higher kernel sizes, visual comparisons indicate less severe degradation than expected. However, kernel sizes like 21x21 and 31x31 introduce significant artifacts, leading to a notable drop in verification accuracy. Both datasets confirm that larger kernels impact key metrics like EER, AUC, and Verification Accuracy.

JLU Dataset Analysis The performance of the biometric system on the JLU dataset shows clear differences between the aligned and unaligned methods, particularly as the intensity of global blurring increases. In the majority of cases, alignment improves performance, with the aligned method typically achieving better results.

Table 4.8: Performance metrics for DHV and JLU datasets under aligned and unaligned blur degradation conditions. In most cases, EER is lower and Verification Accuracy is higher when the dataset is aligned. Verif. Acc. stands for 'Verification Accuracy.'

Dataset	Condition	Aligned Metrics (%)			Unaligned Metrics (%)		
		EER	AUC	Verif. Acc.	EER	AUC	Verif. Acc.
DHV	No Blur	0.10	99.99	99.89	0.29	99.99	99.62
	GB 3x3 Kernel	0.29	99.99	99.92	1.85	99.85	97.93
	GB 5x5 Kernel	0.30	99.98	99.70	4.67	98.80	94.97
	GB 13x13 Kernel	5.08	99.00	94.45	12.80	93.72	86.98
	GB 21x21 Kernel	10.12	96.27	88.54	16.67	90.84	85.63
	GB 31x31 Kernel	16.37	92.44	80.70	20.54	87.17	81.89
	Local Blur CV	5.36	98.71	95.25	5.95	98.78	92.24
	Local Blur LaMa	2.26	99.76	97.74	2.68	99.65	97.21
JLU	No Blur	1.14	99.92	99.07	1.42	99.87	99.09
	GB 3x3 Kernel	1.22	99.91	99.27	1.42	99.87	98.72
	GB 5x5 Kernel	1.22	99.91	99.23	1.49	99.86	99.04
	GB 13x13 Kernel	1.76	99.83	98.66	2.16	99.78	98.79
	GB 21x21 Kernel	4.03	99.22	95.85	3.53	99.47	97.83
	GB 31x31 Kernel	9.90	96.66	89.56	6.49	98.27	94.74
	Local Blur CV	3.85	99.19	96.93	3.93	99.28	97.98
	Local Blur LaMa	1.68	99.82	98.31	1.76	99.81	98.17

Under the no blur condition, the aligned method achieves strong results with an EER of 1.14%, AUC of 99.92%, and Verification Accuracy of 99.07%. The unaligned method, while slightly less accurate, still performs well with an EER of 1.42% and Verification Accuracy of 99.09%.

As the Gaussian blur kernel size increases (e.g., from 3x3 to 31x31), both methods experience performance degradation, with the aligned method typically showing better resilience. For instance, at a 31x31 kernel size, the aligned method's EER rises to 9.90% with a Verification Accuracy of 89.56%, while the unaligned method has an EER of 6.49% and a Verification Accuracy of 94.74%.

When examining the local blurring methods, the LaMa-based approach continues to outperform the traditional computer vision method, particularly in maintaining higher verification accuracy and lower error rates. For the aligned condition, LaMa achieves an EER of 1.68% and a Verification Accuracy of 98.31%, compared to the Local Blur CV's EER of 3.85% and Verification Accuracy of 96.93%. In the unaligned condition, LaMa again shows better performance with an EER of 1.76% and Verification Accuracy of 98.17%, while the Local Blur CV method records an EER of 3.93% and Verification Accuracy of 97.98%. These results suggest that LaMa's advanced inpainting capabilities allow it to better preserve the underlying biometric features, even under conditions where vein patterns are partially obscured, making it a superior method for handling local blurring effects. This also shows us that the computer vision method effectively conceals the vein identity of the subjects.

DHV Dataset Analysis The DHV dataset follows a similar trend, with alignment providing a noticeable performance boost, particularly under mild blurring conditions. Under no blur conditions, the system shows near-perfect performance with an EER of 0.10% and a Verification Accuracy of 99.89%. As blurring intensifies, the aligned method generally outperforms the unaligned one, although both methods experience significant degradation with larger kernel sizes. For example, at a 31x31 kernel size, the aligned method's EER increases to 16.37%, with a Verification Accuracy of 80.70%, whereas the unaligned method reaches an EER of 20.54% and a Verification Accuracy of 81.89%.

When evaluating the local blurring methods, the LaMa-based approach consistently outperforms the traditional computer vision method, particularly in maintaining higher verification accuracy and lower error rates. In the aligned condition, LaMa achieves an EER of 2.26% and a Verification Accuracy of 97.74%, significantly better than the Local Blur CV's EER of 5.36% and Verification Accuracy of 95.25%. This trend continues in the unaligned condition, where LaMa maintains more stable performance with an EER of 2.68% and Verification Accuracy of 97.21%, compared to the Local Blur CV method, which records an EER of 5.95% and Verification Accuracy of 92.24%.

Chapter 5

Conclusion

In this thesis, we have explored the expansive capabilities of Artificial Intelligence in addressing critical challenges within structural health monitoring and biometric security. By focusing on advanced models such as Generative Adversarial Networks (GANs), diffusion models, Convolutional Neural Networks (CNNs), and segmentation architectures, as well as computer vision techniques, we have developed novel methodologies that offer practical solutions to pressing real-world problems.

5.1 Enhancing Structural Integrity through Improved Crack Detection

Structural health monitoring is essential for ensure the safety and longevity of critical infrastructure like bridges, buildings, and pipelines. Early detection of structural cracks can prevent catastrophic failures, safeguard lives, and reduce maintenance costs. Furthermore, traditional inspection methods are often time-consuming, subjective, and resource-intensive, leading to delays in identifying potential issues. We introduced an innovative approach to enhance the performance of semantic segmentation in structural crack detection by generating synthetic data using diffusion models. This novel methodology effectively addresses the challenge of generating segmentation data without the extensive effort typically required for data collection and annotation.

Through comprehensive hyperparameter optimization, the application of data augmentation techniques, and ensemble modeling, we achieved the highest recorded performance on the Bridge Crack Library (BCL) dataset, attaining a Mean Intersection over Union (MeanIoU) of 65.62%. This result sets a new benchmark in structural crack detection, demonstrating that integrating advanced data synthesis methods with model optimization can effectively mitigate data scarcity and enhance model performance. Our work contributes to more efficient and accurate structural inspections, ultimately leading to safer infrastructure and reduced maintenance costs.

5.2 Advancing Biometric Security with Dorsal Hand Vein Recognition

Biometric security systems are increasingly important in safeguarding sensitive information and controlling access in various sectors, including finance, healthcare, and national security. Dorsal hand vein recognition is an underutilized biometric modality that offers significant benefits, such as non-intrusiveness, high resistance to spoofing, and stability over time. Despite these advantages, it has not been widely adopted due to challenges in achieving high performance and consistency.

We proposed a novel knuckle alignment technique tailored for dorsal hand vein biometrics, addressing the unique challenges posed by hand positioning variability in near-infrared (NIR) imaging. Our method effectively scales, centers, and rotates hands according to their knuckle orientation, improving recognition accuracy. This iterative advancement enhances the performance of dorsal hand vein recognition systems, making them a more viable option for secure biometric authentication.

Our automated alignment method achieved exceptional verification accuracies of 99.07% on the Jillian University Dorsal Hand Vein Dataset (JLU) and 99.90% on the Wilchesf Dorsal Hand Vein Dataset (DHV), with perfect Rank-1 identification rates of 100% on both datasets. Additionally, the method offers considerable efficiency gains, being approximately 1,349 times faster than manual alignment and capable of processing thousands of images in mere seconds. These results mark a new benchmark in dorsal hand vein verification and identification accuracies, highlighting the method's potential for widespread adoption in security and identification applications.

5.3 Assessing the Robustness of Dorsal Hand Vein Biometrics under Image Degradation

Understanding the limitations and robustness of biometric systems under varying conditions is crucial for their reliable deployment in real-world applications. Image degradations, such as blurriness and obscured vein visibility, can significantly impact the performance of dorsal hand vein recognition systems, potentially compromising security and user experience.

We conducted a comprehensive study on the robustness of dorsal hand vein biometric systems under various image degradation conditions, including global and local blurring, as well as simulated physiological changes that obscure vein visibility. Our findings indicate that while alignment plays a crucial role in maintaining high recognition accuracy under moderate blurring, its effectiveness diminishes under severe degradation.

For example, under moderate blurring with a 5x5 Gaussian blur kernel on the DHV dataset, the aligned system achieved a verification accuracy of 99.70% and an Equal Error Rate (EER) of 0.30%, significantly outperforming the unaligned system. However, under severe blurring conditions, the unaligned method exhibited greater resilience, suggesting that alignment is less effective when image quality is heavily compromised.

We simulated the effects of increased subcutaneous tissue on vein visibility using a LaMa and CV-based inpainting approach. We found that the LaMa-based deep learning approach demonstrated superior performance over traditional computer vision methods, preserving underlying biometric information and resulting in less performance degradation. However, the CV-based approach was better at concealing the underlying biometric features. In addition, we learn that the models are still fairly dependent on other key biometric features besides the dorsal hand veins.

5.4 Closing Remarks

The research presented in this thesis addresses current challenges in structural health monitoring and biometric security and paves the way for future advancements. Using innovative AI and computer vision techniques, we have set new benchmarks in crack detection and dorsal hand vein recognition, contributing to safer infrastructure and more secure authentication systems.

Looking ahead, the methodologies developed in this work can be extended to other domains facing similar challenges of data scarcity, alignment complexities, and robustness under varying conditions. The integration of our alignment method to dorsal hand vein biometric systems holds great potential. As technology continues to evolve, our approaches can be adapted and expanded, fostering innovation and enhancing the effectiveness of systems across various industries.

In conclusion, this thesis lays the foundation for future research, inspiring further exploration into the applications of AI and computer vision in critical fields. By continuing to push the boundaries of what is possible, we can develop more reliable, efficient, and secure systems that have a lasting positive impact on society.

Bibliography

Abaza, A., Harrison, M. A., Bourlai, T., & Ross, A. (2014). Design and evaluation of photometric image quality measures for effective face recognition. *IET Biometrics*, *3*(4), 314–324.

Alashik, K. M., & Yildirim, R. (2021). Human identity verification from biometric dorsal hand vein images using the dl-gan method. *IEEE Access*, *9*, 74194–74208.

Ali, R., Chuah, J. H., Talip, M. S. A., Mokhtar, N., & Shoaib, M. A. (2022). Structural crack detection using deep convolutional neural networks. *Automation in Construction*, *133*, 103989.

Al-johania, N. A., & Elrefaei, L. A. (2019). Dorsal hand vein recognition by convolutional neural networks: Feature learning and transfer learning approaches. *International Journal of Intelligent Engineering & Systems*, *12*(3).

Arakala, A., Hao, H., Davis, S., & Horadam, K. J. (2015). The palm vein graph for biometric authentication. *Information Systems Security and Privacy: First International Conference, ICISSP 2015, Angers, France, February 9-11, 2015, Revised Selected Papers 1*, 199–218.

Arora, P., Srivastava, S., Hanmandlu, M., & Bhargava, S. (2018). Robust authentication using dorsal hand vein images. *IEEE Intelligent Systems*, *34*(2), 25–35.

Atik, S. O., Atik, M. E., & Ipbuker, C. (2022). Comparative research on different backbone architectures of deeplabv3+ for building segmentation. *Journal of Applied Remote Sensing*, *16*(2), 024510–024510.

Bartz, C., Raetz, H., Otholt, J., Meinel, C., & Yang, H. (2022). Synthesis in style: Semantic segmentation of historical documents using synthetic data. *2022 26th International Conference on Pattern Recognition (ICPR)*, 3878–3884.

Benaouda, A., Mustapha, A. H., & Benziane, S. (2021). A cnn approach for the identification of dorsal veins of the hand. *International Conference on Artificial Intelligence and its Applications*, 574–587.

Bhosale, A. S., & Jadhav, M. R. (2017). Dorsal hand vein pattern recognition system based on neural network. *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, *2*, 52–55.

Boillet, M., Kermorvant, C., & Paquet, T. (2021). Multiple document datasets pre-training improves text line detection with deep neural networks. *2020 25th International Conference on Pattern Recognition (ICPR)*, 2134–2141.

Bourlai, T., Rose, J., Mokalla, S. R., Zabin, A., Hornak, L., Nalty, C. B., Peri, N., Gleason, J., Castillo, C. D., Patel, V. M., et al. (2023). Data and algorithms for end-to-end thermal spectrum face verification. *IEEE Transactions on Biometrics, Behavior, and Identity Science*.

Cao, J., Xu, M., Shi, W., Yu, Z., Salim, A., & Kilgore, P. (2015). Mypalmvein: A palm vein-based low-cost mobile identification system for wide age range. *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, 292–297.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Chin, S. W., Tay, K. G., Huong, A., & Chew, C. C. (2020). Dorsal hand vein pattern recognition using statistical features and artificial neural networks. *2020 IEEE Student Conference on Research and Development (SCOReD)*, 217–221.

Choi, W., & Cha, Y.-J. (2019). Sddnet: Real-time crack segmentation. *IEEE Transactions on Industrial Electronics*, *67*(9), 8016–8025.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258.

Cui, L., Qi, Z., Chen, Z., Meng, F., & Shi, Y. (2015). Pavement distress detection using random decision forests. *International Conference on Data Science*, 95–102.

Damak, W., Boukhris Trabelsi, R., Alima Damak, M., & Sellami, D. (2018). Dynamic roi extraction method for hand vein images. *IET Computer Vision*, *12*(5), 586–595.

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4690–4699.

Derawi, M. O., Yang, B., & Busch, C. (2012). Fingerprint recognition with embedded cameras on mobile phones. *Security and Privacy in Mobile Information and Communication Systems: Third*

International ICST Conference, MobiSec 2011, Aalborg, Denmark, May 17-19, 2011, Revised Selected Papers 3, 136–147.

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, *34*, 8780–8794.

Ding, K., Zhou, M., Wang, H., Gevaert, O., Metaxas, D., & Zhang, S. (2023). A large-scale synthetic pathological dataset for deep learning-enabled segmentation of breast cancer. *Scientific Data*, *10*(1), 231.

El-Naggar, S., & Bourlai, T. (2022). Image quality assessment for effective ear recognition. *IEEE Access*, *10*, 98153–98164. https://doi.org/10.1109/ACCESS.2022.3206024

Fan, Z., Li, C., Chen, Y., Di Mascio, P., Chen, X., Zhu, G., & Loprencipe, G. (2020). Ensemble of deep convolutional neural networks for automatic pavement crack detection and measurement. *Coatings*, *10*(2), 152.

Fetty, L., Bylund, M., Kuess, P., Heilemann, G., Nyholm, T., Georg, D., & Löfstedt, T. (2020). Latent space manipulation for high-resolution medical image synthesis via the stylegan. *Zeitschrift für Medizinische Physik*, *30*(4), 305–314.

Francisco, M. D., Chen, W.-F., Pan, C.-T., Lin, M.-C., Wen, Z.-H., Liao, C.-F., & Shiue, Y.-L. (2021). Competitive real-time near infrared (nir) vein finder imaging device to improve peripheral subcutaneous vein selection in venipuncture for clinical laboratory testing. *Micromachines*, *12*(4), 373.

Fu, H., Meng, D., Li, W., & Wang, Y. (2021). Bridge crack semantic segmentation based on improved deeplabv3+. *Journal of Marine Science and Engineering*, *9*(6), 671.

Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2019). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, *58*, 101563.

Han, C., Ma, T., Huyan, J., Huang, X., & Zhang, Y. (2021). Crackw-net: A novel pavement crack image segmentation convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 22135–22144.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

He, Z., Xu, Q., Ye, Y., & Li, W. (2017). Effects of meteorological factors on finger vein recognition. *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 1–8.

Hirata, D., & Takahashi, N. (2023). Ensemble learning in cnn augmented with fully connected subnetworks. *IEICE TRANSACTIONS on Information and Systems*, *106*(7), 1258–1261.

Hsu, C.-B., Hao, S.-S., & Lee, J.-C. (2011). Personal authentication through dorsal hand vein patterns. *Optical Engineering*, *50*(8), 087201–087201.

Huang, D., Tang, Y., Wang, Y., Chen, L., & Wang, Y. (2014). Hand-dorsa vein recognition by matching local features of multisource keypoints. *IEEE transactions on cybernetics*, *45*(9), 1823–1837.

Huang, D., Zhu, X., Wang, Y., & Zhang, D. (2016). Dorsal hand vein recognition via hierarchical combination of texture and shape clues. *Neurocomputing*, *214*, 815–828.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Huyan, J., Li, W., Tighe, S., Xu, Z., & Zhai, J. (n.d.). Cracku-net: A novel deep convolutional neural network for pixelwise pavement crack detection. *Structural Control and Health Monitoring*, *27*(8), e2551.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Jia, W., Xia, W., Zhang, B., Zhao, Y., Fei, L., Kang, W., Huang, D., & Guo, G. (2021). A survey on dorsal hand vein biometrics. *Pattern Recognition*, *120*, 108122.

Jin, T., Ye, X., & Li, Z. (2023). Establishment and evaluation of conditional gan-based image dataset for semantic segmentation of structural cracks. *Engineering Structures*, *285*, 116058.

Jin, T., & Li, Z. (2022). Bridge Crack Library 2.0. https://doi.org/10.7910/DVN/TUFAJT

Jin, T., Li, Z., Ding, Y., Ma, S., & Ou, Y. (2021). Bridge Crack Library. https://doi.org/10.7910/DVN/RURXSH

Kailkhura, V., Aravindh, S., Jha, S. S., & Jayanthi, N. (2020). Ensemble learning-based approach for crack detection using cnn. *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 808–815.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., & Aila, T. (2021). Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, *34*, 852–863.

Kashif Ur Rehman, S., Ibrahim, Z., Memon, S. A., & Jameel, M. (2016). Nondestructive test methods for concrete bridges: A review. *Construction and Building Materials*, *107*, 58–86. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2015.12.011

Kauba, C., Drahanskỳ, M., Nováková, M., Uhl, A., & Rydlo, Š. (2022). Three-dimensional finger vein recognition: A novel mirror-based imaging device. *Journal of Imaging*, *8*(5), 148.

Kocakulak, M., Avcı, A., & Acır, N. (2023). Automated vein verification using self-attention-based convolutional neural networks. *Expert Systems with Applications*, *230*, 120550.

Konnova, N. S., & Mizinov, P. V. (2023). Comparative analysis of dorsal palm vein pattern images in biometric identification problems. *AIP Conference Proceedings*, *2819*(1).

Kosmala, J., & Saeed, K. (2012). Human identification by vascular patterns. In *Biometrics and kansei engineering* (pp. 67–87). Springer.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 84–90.

Kumar, A., & Prathyusha, K. V. (2009). Personal authentication using hand vein triangulation and knuckle shape. *IEEE Transactions on Image processing*, *18*(9), 2127–2136.

Lajevardi, S. M., Arakala, A., Davis, S., & Horadam, K. J. (2014). Hand vein authentication using biometric graph matching. *IET Biometrics*, *3*(4), 302–313.

Lau, S. L., Chong, E. K., Yang, X., & Wang, X. (2020). Automated pavement crack segmentation using u-net-based convolutional neural network. *Ieee Access*, *8*, 114892–114899.

Lee, J.-C., Lee, C.-H., Hsu, C.-B., Kuei, P.-Y., & Chang, K.-C. (2014). Dorsal hand vein recognition based on 2d gabor filters. *The Imaging Science Journal*, *62*(3), 127–138.

Li, B., Xue, K., Liu, B., & Lai, Y.-K. (2023). Bbdm: Image-to-image translation with brownian bridge diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1952–1961.

Li, X., Zhong, Z., Wu, J., Yang, Y., Lin, Z., & Liu, H. (2019). Expectation-maximization attention networks for semantic segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9167–9176.

Li, X., Huang, D., & Wang, Y. (2016). Comparative study of deep learning methods on dorsal hand vein recognition. *Biometric Recognition: 11th Chinese Conference, CCBR 2016, Chengdu, China, October 14-16, 2016, Proceedings 11*, 296–306.

Li, X., Huang, D., Zhang, R., Wang, Y., & Xie, X. (2016). Hand dorsal vein recognition by matching width skeleton models. *2016 IEEE International Conference on Image Processing (ICIP)*, 3146–3150.

Lim, A., & Mulcahy, A. (2017). Hand rejuvenation: Combining dorsal veins foam sclerotherapy and calcium hydroxylapatite filler injections. *Phlebology*, *32*(6), 397–402.

Lin, C.-L., & Fan, K.-C. (2004). Biometric verification using thermal images of palm-dorsa vein patterns. *IEEE Transactions on Circuits and systems for Video Technology*, *14*(2), 199–213.

Liu, F., Jiang, S., Kang, B., & Hou, T. (2020a). Improved multiscale local phase quantisation histogram for the recognition of blurred dorsal hand vein images. *Electronics Letters*, *56*(23), 1232–1235.

Liu, F., Jiang, S., Kang, B., & Hou, T. (2020b). A recognition system for partially occluded dorsal hand vein using improved biometric graph matching. *IEEE Access*, *8*, 74525–74534.

Liu, J., Yang, X., Lau, S., Wang, X., Luo, S., Lee, V. C.-S., & Ding, L. (2020). Automated pavement crack detection and segmentation based on two-step convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, *35*(11), 1291–1305.

Liu, Y., Yao, J., Lu, X., Xie, R., & Li, L. (2019). Deepcrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing*, *338*, 139–153.

Luna-Benoso, B., Martınez-Perales, J., & Cortés-Galicia, J. (2021). A methodology for biometric identification using vein geometry in infrared images of the back of the hand. *International Journal of Contemporary Mathematical Sciences*, *16*(4), 135–148.

Maarouf, A. A., & Hachouf, F. (2022). Transfer learning-based ensemble deep learning for road cracks detection. *2022 International Conference on Advanced Aspects of Software Engineering (ICAASE)*, 1–6.

Martin, M., & Bourlai, T. (2017). Enhanced tattoo image quality assessment through multispectral sensing. *IEEE sensors letters*, *1*(6), 1–4.

Michaelis, C., Mitzkus, B., Geirhos, R., Rusak, E., Bringmann, O., Ecker, A. S., Bethge, M., & Brendel, W. (2019). Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*.

Munir, R., & Khan, R. A. (2019). An extensive review on spectral imaging in biometric systems: Challenges & advancements. *Journal of Visual Communication and Image Representation*, *65*, 102660.

Nadiya, K., & Gopi, V. P. (2020). Dorsal hand vein biometric recognition based on orientation of local binary pattern. *2020 IEEE-HYDCON*, 1–6.

Nie, W., & Zhang, B. (2019). Robust and adaptive roi extraction for hyperspectral dorsal hand vein images. *IET Computer Vision*, *13*(6), 595–604.

Nie, Z., Xu, J., & Zhang, S. (2020). Analysis on deeplabv3+ performance for automatic steel defects detection. *arXiv preprint arXiv:2004.04822*.

Nour, R., Moustafa, H. E.-D., AbdelHay, E. H., & Ata, M. M. (2024). Improved unsupervised deep boltzmann learning approach for accurate hand vein recognition. *IEEE Access*.

Otsu, N., et al. (1975). A threshold selection method from gray-level histograms. *Automatica*, *11*(285-296), 23–27.

Pan, Z., Lau, S. L., Yang, X., Guo, N., & Wang, X. (2023). Automatic pavement crack segmentation using a generative adversarial network (gan)-based convolutional neural network. *Results in Engineering*, *19*, 101267.

Pei, L., Sun, Z., Xiao, L., Li, W., Sun, J., & Zhang, H. (2021). Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network. *Engineering Applications of Artificial Intelligence*, *104*, 104376.

Pititheeraphab, Y., Thongpance, N., Aoyama, H., & Pintavirooj, C. (2020). Vein pattern verification and identification based on local geometric invariants constructed from minutia points and augmented with barcoded local feature. *Applied Sciences*, *10*(9), 3192.

Piuri, V., & Scotti, F. (2008). Fingerprint biometrics via low-cost sensors and webcams. *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, 1–6.

Polovnikov, V., Alekseev, D., Vinogradov, I., & Lashkia, G. V. (2021). Daunet: Deep augmented neural network for pavement crack segmentation. *IEEE Access*, *9*, 125714–125723.

Qu, Z., Cao, C., Liu, L., & Zhou, D.-Y. (2021). A deeply supervised convolutional neural network for pavement crack detection with multiscale feature fusion. *IEEE transactions on neural networks and learning systems*, *33*(9), 4890–4899.

Qu, Z., Li, Y., & Zhou, Q. (2022). Crackt-net: A method of convolutional neural network and transformer for crack segmentation. *Journal of Electronic Imaging*, *31*(2), 023040–023040.

Raghavendra, R., Surbiryala, J., & Busch, C. (2015). Hand dorsal vein recognition: Sensor, algorithms and evaluation. *2015 IEEE international conference on imaging systems and techniques (IST)*, 1–6.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Rodriguez-Lozano, F. J., León-García, F., Gámez-Granados, J. C., Palomares, J. M., & Olivares, J. (2020). Benefits of ensemble models in road pavement cracking classification. *Computer-Aided Civil and Infrastructure Engineering*, *35*(11), 1194–1208.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *arXiv*. https://github.com/CompVis/latent-diffusion

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241.

Rossan, I., & Khan, M. H.-M. (2014). Impact of changing parameters when preprocessing dorsal hand vein pattern. *Procedia Computer Science*, *32*, 513–520.

Ruiz-Echeverri, J. M., Bernal-Romero, J. C., Ramirez-Cortes, J. M., Gomez-Gil, P., Rangel-Magdaleno, J., & Peregrina-Barreto, H. (2021). Dorsal hand veins biometrics using nir images with fusion of classifiers at score level. *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 1–6.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520.

Shi, Y., Cui, L., Qi, Z., Meng, F., & Chen, Z. (2016). Automatic road crack detection using random structured forests. *IEEE Transactions on Intelligent Transportation Systems*, *17*(12), 3434–3445.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., & Lempitsky, V. (2022). Resolution-robust large mask inpainting with fourier convolutions. *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2149–2159.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (n.d.). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *International conference on machine learning*, 6105–6114.

Tang, J., Chen, C., Huang, Z., Zhang, X., Li, W., Huang, M., & Deng, L. (2022). Crack unet: Crack recognition algorithm based on three-dimensional ground penetrating radar images. *Sensors*, *22*(23), 9366.

Tazim, R. J., Miah, M. M. M., Surma, S. S., Islam, M. T., Shahnaz, C., & Fattah, S. A. (2018). Biometric authentication using cnn features of dorsal vein pattern extracted from nir image. *TENCON 2018-2018 IEEE Region 10 Conference*, 1923–1927.

Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of graphics tools*, *9*(1), 23–34.

Temam, R. (2001). *Navier-stokes equations: Theory and numerical analysis* (Vol. 343). American Mathematical Soc.

Tome, P., Vanoni, M., & Marcel, S. (2014). On the vulnerability of finger vein recognition to spoofing. *2014 international conference of the biometrics special interest group (BIOSIG)*, 1–10.

Wan, H., Chen, L., Song, H., & Yang, J. (2017). Dorsal hand vein recognition based on convolutional neural networks. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1215–1221.

Wang, G., Sun, C., & Sowmya, A. (2019a). Learning a compact vein discrimination model with ganerated samples. *IEEE Transactions on Information Forensics and Security*, *15*, 635–650.

Wang, G., Sun, C., & Sowmya, A. (2019b). Multi-weighted co-occurrence descriptor encoding for vein recognition. *IEEE Transactions on Information Forensics and Security*, *15*, 375–390.

Wang, J.-J., Liu, Y.-F., Nie, X., & Mo, Y. (2022). Deep convolutional neural networks for semantic segmentation of cracks. *Structural Control and Health Monitoring*, *29*(1), e2850.

Wang, J., & Wang, G. (2017). Quality-specific hand vein recognition system. *IEEE Transactions on Information Forensics and Security*, *12*(11), 2599–2610.

Wang, L., Leedham, G., & Cho, D. S.-Y. (2008). Minutiae feature analysis for infrared hand vein pattern biometrics. *Pattern recognition*, *41*(3), 920–929.

Wang, S., & Tang, W. (2012). Pavement crack segmentation algorithm based on local optimal threshold of cracks density distribution. *Advanced Intelligent Computing: 7th International Conference, ICIC 2011, Zhengzhou, China, August 11-14, 2011. Revised Selected Papers 7*, 298–302.

Wang, W., & Su, C. (2020). Convolutional neural network-based pavement crack segmentation using pyramid attention network. *IEEE Access*, *8*, 206548–206558.

Wang, Y., Cao, H., Jiang, X., & Tang, Y. (2019). Recognition of dorsal hand vein based bit planes and block mutual information. *Sensors*, *19*(17), 3718.

Wang, Y., Cao, X., & Miao, X. (2022). Cross-device recognition of dorsal hand vein images by two-stage coarse-to-fine matching. *The Visual Computer*, *38*(11), 3595–3610.

Wang, Z., Zhang, B., Chen, W., & Gao, Y. (2008). A performance evaluation of shape and texture based methods for vein recognition. *2008 Congress on Image and Signal Processing*, *2*, 659–661.

Wang, Z., Yang, J., Jiang, H., & Fan, X. (2020). Cnn training with twenty samples for crack detection via data augmentation. *Sensors*, *20*(17), 4849.

Wilches-Bernal, F., Núñez-Álvares, B., & Vizcaya, P. (2020). A database of dorsal hand vein images. *arXiv preprint arXiv:2012.05383*.

Xu, J., Yuan, C., Gu, J., Liu, J., An, J., & Kong, Q. (2023). Innovative synthetic data augmentation for dam crack detection, segmentation, and quantification. *Structural Health Monitoring*, *22*(4), 2402–2426.

Yakno, M., Mohamad-Saleh, J., & Ibrahim, M. Z. (2021). Dorsal hand vein image enhancement using fusion of clahe and fuzzy adaptive gamma. *Sensors*, *21*(19), 6445.

Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., & Ling, H. (2019). Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Transactions on Intelligent Transportation Systems*.

Ye, X.-W., Jin, T., Li, Z., Ma, S., Ding, Y., & Ou, Y. (2021). Structural crack detection from benchmark data sets using pruned fully convolutional networks. *Journal of Structural Engineering*, *147*(11), 04721008.

Ye, X., Li, Z., & Jin, T. (2022). Smartphone-based structural crack detection using pruned fully convolutional networks and edge computing. *Smart Structures and Systems*, *29*(1), 141–151.

Yu, G., Dong, J., Wang, Y., & Zhou, X. (2023). Ruc-net: A residual-unet-based convolutional neural network for pixel-level pavement crack segmentation. *Sensors*, *23*(1), 53.

Yu, X., Li, G., Lou, W., Liu, S., Wan, X., Chen, Y., & Li, H. (2023). Diffusion-based data augmentation for nuclei image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 592–602.

Zhai, G., Narazaki, Y., Wang, S., Shajihan, S. A. V., & Spencer Jr, B. F. (2022). Synthetic data augmentation for pixel-wise steel fatigue crack identification using fully convolutional networks. *Smart Struct Syst*, *29*(1), 237–250.

Zhang, L., Yang, F., Zhang, Y. D., & Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. *Image Processing (ICIP), 2016 IEEE International Conference on*, 3708–3712.

Zhang, R., Zou, X., Deng, X., Wang, Z., Chen, Y., Lin, C., Xing, H., & Dai, F. (2023). Fast and accurate roi extraction for non-contact dorsal hand vein detection in complex backgrounds based on improved u-net. *Sensors*, *23*(10), 4625.

Zhu, X., Huang, D., & Wang, Y. (2013). Hand dorsal vein recognition based on shape representation of the venous network. *Advances in Multimedia Information Processing–PCM 2013: 14th Pacific-Rim Conference on Multimedia, Nanjing, China, December 13-16, 2013. Proceedings 14*, 158–169.