Complementary learning for Deep
Medical Image Segmentation and Anomaly Detection

by

Raunak Dey

(Under the Direction of Yi Hong)

Abstract

Image segmentation is a fundamental step in many biomedical image analysis applications to accurately and coherently separate out a region of interest for further analysis. However often times additional information and features which are not directly related to the region of interest is discarded. This dissertation explores ways to leverage those complementary information in order to enhance the quality of segmentation of an region of interest.

There are three major contributions; Firstly, the idea of complementary learning is formulated as a simple to use loss function for use with deep learning frameworks. Secondly, we contribute a model, Comp-Net, for supervised learning which enhances an existing segmentation algorithm to be more tolerant to unseen deviation during test set uncharacteristic to the training set. We show the robustness on two applications, skull stripping and liver lesion segmentation. Finally, we propose a new way to perform Unsupervised Anomaly Segmentation via selective two cuts with one cut chosen by the user as a reference distribution, outperforming existing state of the art methods on MS-SEG2015 dataset and show promising performance on BraTS 2019, LiTS Liver Lesion and a privately curated brain tumor dataset. The two cut allows to move away from reconstruction dependent anomaly/novelty segmentation and

moves the focus back to the anomaly itself.

Index words: Unsupervised Anomaly Detection, Unsupervised Segmentation, Unsupervised Learning, Robust Segmentation, Complementary Learning, Multi-Output Networks.

Complementary learning for Deep Medical Image Segmentation and Anomaly Detection

by

Raunak Dey

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2021

Complementary learning for Deep Medical Image Segmentation and Anomaly Detection

by

Raunak Dey

Major Professor: Yi Hong

Committee: Shannon Quinn
Suchendra Bhandarkar

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2021

Acknowledgments

Like Dr Quinn mentions in his blog, "If you can handle grad school, I believe you can handle anything", I strongly believe in that too. However, "No man is an island" and the journey that a PhD student undertakes involves a lot of people from all walks of life who make the journey possible. I will try my best to remember everyone who has made it possible for me to complete this journey and hopefully do them some justice. This section may end up being usually longer than expected, however this is the most important thing I have accomplished so far and I want to remember everyone important. Any opinions mentioned in this acknowledgement section are my own, and my own only and not the responsibility of any other party.

0.1 PhD Committee

Dr Yi Hong: I would like to begin by thanking you, my advisor, Dr Yi Hong. The kind of researcher I am today is all thanks to you. It is quite difficult to navigate a completely new field on your own and it was no different for me. When I came to UGA and when I started to work with you, I knew nothing of Deep Learning and AI, nor did I know Python. It was through the different projects you assigned me, I was able to improve myself. I would love to continue working with you after this chapter of my life is done and I believe that many of the places where I am weak, or ill equipped to handle you shine and as such we were always

a good complement to each other (like the algorithms we developed!). I have always noted that if I am able to convince you about my work no reviewer will ever reject my paper. For the most part that has been the case. You are quite the critic and for a student like me who has no focus for sticking to one thing, this was a blessing. I remember I had to ask you to be more critical of me in order for me to succeed and while you were reluctant at first you did agree to do just that. Also I thank you for replying to most of the hundreds of emails I spammed you with. I get excited too easily and every time I had a breakthrough or a pitfall I recall sending an email. Your letter of recommendation also helped me secure my future job and I am really ecstatic about the area of research I will go into and I feel you have helped me develop the skills required to handle my new work. I wish you the best in life and would like to thank you for everything. Also I hope to one day meet you again maybe at a conference or just a social visit. You have been an advisor, an elder sister, a mentor and colleague and quite frankly the single most important person in successfully completing my PhD. We have had our differences at times, but I have always admired that you forgave me even when I stepped out of line at times.

Dr Shannon Quinn: Dr Quinn, you have been an inspiration in many ways. Firstly I love your blogs, I like reading them and often times direct students to it when they have questions regarding PhD etc. On a personal level, I enjoy your teaching style and I took a few courses simply to observe and incorporate a few of your teaching styles to my own. I did drop them midway since I was already too busy with coursework and other things. I admire your pursuit of open science. That I feel is severely lacking in the deep learning community. When I was starting out, I felt quit challenged to replicate many of the paper's algorithms. However I did persevere on and I still remember spending one whole year just learning how to connect different components together. Today I can proudly say that I can replicate any neural network architecture and I can code up any kind of structure I can imagine. However if it wasnt for your Data Science Practicum class and your stress on open science I would have never replied to people who emailed me asking questions of my code simply because I was salty that I had to spend an entire year learning all of those myself. Your outlook on life and research has mellowed me out quite a bit. You also helped me a lot in organizing my dissertation and I thank you for that. I wish you a safe travels on your upcoming trip and wish you the best in your future endeavours. Finally congratulations on having a baby girl. I wish you and your family the best in life.

Dr. Suchendra Bhandarkar: I first met you when I was signing my form to register for classes as you were the graduate co-ordinator back then. I was quite taken aback hearing you speak in Bengali, which is the language of the region in India I come from. You have travelled a lot and have seen a lot which becomes quite evident when talking to you. I appreciate all the wisdom and life lessons you gave me during the times we spoke, including the different ways I could improve myself. One of the regrets I have is the fact that I never got to work on any shape based deep learning frameworks, however I plan to pursue that when I find time in the future. At the same time, I thank you for introducing me to the area of shape-based learning and I feel it will be very important in my future area of research mainly for the protein folding problem. You have also helped me tackle one of my glaring weaknesses, which is the inability to formulate my thoughts perfectly well in writing. You gave me so many guidelines on effective academic writing such as writing papers, prospectus, and how to include complex background information for my dissertation. Deep Learning research is vast and is expanding every second even as I write these lines. However, your advice was short and simple, i.e., to build on the immediately related works closest to mine, explain them and then expand on how my work enhances them. I appreciate that feedback and I will be editing parts of the dissertation with that in mind. I remember you talking about how your children were doing around 3 or 4 years ago and I hope they are continuing being successful. I wish you and your family the best in life. One other thing I never mentioned to you, is that your name is quite similar to my mother's name. Her name is Such<u>a</u>ndra. I always found it amusing that someone with a highly similar name to my mother ended up playing an important role in my PhD journey.

I am very grateful to all of you for teaching me a lot of important life lessons, I wish there were more times where we all interacted together just like during my defense and discussed more ideas. Maybe we can continue to collaborate in future and discuss many potential ways of improving the current state of research in our respective fields. I also really hope that going forward if I need help I can call on you guys to discuss things. Similarly if there is anything I can help with in your areas of work please reach out and let me know. I would love to continue working with all of you guys going forward.

0.2 University of Georgia

I would like to thank the University of Georgia for taking a chance on me and giving me the opportunity for pursuing a PhD. I spent my high school days fooling around and playing gigs across my home city and wasnt really focused on studies that much. As such I didnt have a strong under grad performance as well. However I have always been naturally curious about how world works and I loved algorithms. I appreciate the department of Computer Science at UGA for realizing my potential and taking a chance on me. Since I have joined UGA, I have published four first author research papers at top conferences and almost done with the fifth, obtaining a citation count of 138 for them at the time of writing this. One of my works have been adopted by people at Harvard Medical School for skull stripping. I have filed for a patent while interning at Ancestry.com and been part of two other publication one in security and the other in education. I hope that I have made the University proud. Going forward I will be working at Illumina, and researching into AI and Genomics. This place has been my home for the last 6 years and I have experienced a multitude of emotions and have grown a lot during this time. "Gooooooooo Dawgs! Sic ‘em! Woof! Woof! Woof! Woof! Woof!".

0.3 Labmates

I would like to thank each and every one of my labmates. I would like to thank **Zahra/Parya**, it was your presentation on W-Net way back which pretty much gave me the idea of how I can implement a network satisfying complementary constraint, and that is basically my whole dissertation. Not just that but your recent publication on graph neural network gave me a quick review about how to work with them and the scope of its usage. Meeting you has been instrumental to my research since I would be using graph neural networks intensively in the future. Also not to mention you and Omid fell in love and got married so we will all be neighbors in future. **Sharmin**, you were one of my oldest labmate and I appreciate we still keep in touch. While I will not be moving to Seattle, California is close by we should definitely catch up. I absolutely loved our late night conversations while we would work. You are an astute researcher and I wish you the best. **Dhara**, your base U-Net implementation is what started all of the segmentation journey that I have been through and I thank you for that. **Rutu** We still always end up talking for hours and hours whenever we converse. I had fun learning about attention network and discussing various ways of implementing it and talking about life in general. Since we will be around 30 mins away from each other soon lets not be strangers! Enjoy your trip through Europe and wish you and Jay the best of luck. **Jiaojiao** I hope you and your husband are doing well. You too have taught me a lot of meaningful life lessons alongside the ideas of tracing features that neural networks use for making decisions. I also want to thank you for recommending me for a job at your company. It is always a pleasure talking to you and hopefully once travel restrictions are more relaxed I can come visit you in Montreal. **Ankita** you are the only remaining PhD student at this lab. While I have not had the chance to interact with you much since you work from home mostly, I really enjoyed the time we spend here. I have learnt a lot about Image registration from you and I would be using that in my future research. I see so much potential of using a traceable step by step registration algorithm in analysing small genetic changes across the evolution timeline as well as comparing the existing tips of the tree of life. Wish you and your fiance the best in life and hopefully when you are graduating from your PhD I will be able to come visit and watch your graduation.

0.4 CS Department

Dr Liming Cai: Dr Cai, algorithms has been my initial interest when I joined UGA and I enjoyed taking your advanced algorithm and basic algorithm classes. I wish I had the opportunity to do a project with you but sadly like always time constraints did not allow me to do so. I sometimes feel like sitting in your class for fun and just revisiting some of my favourite algorithms techniques. Currently I will be working on genomics and I have been learning a lot about protein folding as its one of the subtopics I will be working on. Thank you for being an inspiration and it has always been fun talking to you.

Dr Hamid Arabnia: I felt sad seeing you retire as a full professor. Its been a pleasure to talk to you. Every interaction I learnt something new. You gave me a lot of feedback and discussed all the different responsibilities academicians hold. You also taught me the importance of being unbiased in selecting individuals for positions via one of the many stories you have shared. I will not go into details but maybe you can recall. It feels sad to see the door shut when I go into the mailroom and I really miss the days where I would just drop by and say hi and chit chat on random topics with you. Furthermore it was in your class I learnt all about the different filters and their use and realized how neural networks are simply searching for a big group of filters to solve a problem. I build upto your lessons in chapter 2. Thank You for everything.

Dr Lakshmish Ramaswamy: I will miss your enthusiasm dearly. You have participated in organizing every event we have held and you have always been proactive in getting everyone together faculties and graduate students alike. You have helped our GSA with invaluable advice. You have also advised the Indian students here at UGA a lot and for that I am grateful.

Dr Thiab Taha: I thank you for helping me make my mind up between academia and industry. Thank you for sharing your experiences after your PhD and the time after that. It was really helpful. I would have stayed in academia, first as a lecturer and then later apply to be a professor. However I got an offer in genomics, which is something I cherish and am curious about and thus I have decided to join industry for a bit. I would at one point simultaneously join as a faculty at some university alongside industry. I also thank you for helping me make the transition once Dr Hong became an adjunct professor at UGA.

Dr In Kee Kim: I would like to thank you for actively participating in our research day organization which was one of the events which brought everyone together. We discussed many games, gifts and trivia which ended up making the event successful and fun as well. Dr Ashok Goel even came to me to say he enjoyed the trivia. It was really fun and you also showed me the game you play sometimes with your kid after I mentioned I play league of legends and a bunch of Korean mmo's. Also just today, you gave me so many good advises for the future and I am grateful for that. Thank you and wish you and your family the best in life.

Dr Khaled Rasheed: I loved your Evolutionary Computation class. It basically gave me an important tool in my arsenal, the genetic algorithm. I cant thank you enough for that course. While in my current research I have not been able to use GA yet, I have a lot of ideas for the future involving a harmony between using GA's and Neural Nets alike. I feel heuristics are an amazing tool and thank you for teaching me them.

Dr Don Potter: After Dr Rasheed's class I took your class and learnt more about PSO, Fuzzy Logic and basics of neural networks amongst other things. I loved how the class was discussion based and it was probably one of my most memorable classes at UGA. Before Dr

Hong joined, I was planning to do my Masters thesis with you, however you retired that year and I never signed the form. Another reason I will never forget your class is because I met a dear friend of mine Omid Setayashfar there. All in all I thank you for teaching Computational Intelligence. I believe we were the last batch of students who got to take that course with you and I hope you are keeping well and wish you the best in your future endeavours.

Dr Soheyla Amirian: We have gotten to know each other better in the past few months. I always see you put in so much efforts in preparing for the classes, and the students reciprocate it accordingly. You kind of replace Dr Arabnia here for me. I can literally knock on your door for a quick chit chat any time you are free. Thank you for your good advises and your help when I was thinking about a lecturer track. You are an amazing person and I wish you the best. Hope to stay in touch with you. **Dr Yiheng Liang**: Thanks for helping me simplify a few of the dilemmas I had regarding students and the way to teach them as well as showing me ways to be diplomatic. I am happy to report that a few of the students who struggle initially had come to score maximum points in quizzes labs and projects after I spend additional hours with them. Wish you the best in the oncoming years. This 1301 was the last time I will be a TA and it was nice working with you.

Dr Michael E Cotterell: I have been your TA multiple times and have had the pleasure of learning a lot about handing students and answering questions. Thank You

Samantha Varghese: I would like to give thanks to Mrs. Samantha Varghese, M.Ed., in the Department of Computer Science. She was instrumental in helping me the last four years of my Phd program in Computer Science. She is focused on helping every student, transition and have a successsful graduate student experience. During peak times of stress in preparing for my defense, she encouraged and prepared me on administrative processes that gave me confidence and allowed me to give a successful doctoral defense. She was kind to keep me organized, and just for being there as student support to ease the stress of graduate studies. I appreciate her assistance over the years and would like to acknowledge the importance of university staff who invest in student success.

Anne Steward: Dear Anne. It has been amazing to have been there throughout your journey. You have also been helpful with every small and big technical thing I have ever needed. More than anything you have been an amazing friend. I will miss all the morning chats. I was so happy you visited Cuba after I told you nice stories about Cuba. Also if you ever visit California, you must let me know! I know it takes a lot of strength to go through everything that you have been through and I wish you all the happiness and hope you accomplish all you can think of!

Evette Dunbar: Any dull day can be brightened by seeing your smiling face. Thank you for everything, from organizing our research days to coffee hours to also now handling travelling related work and course related updatation on Athena amongst other things. I hope you can enjoy your break and sleep without needing melatonin! It has been great getting to know you and thank you once again.

Kimberly Buffington: I missed saying goodbye when you left the CS department. However I wish you the best and hope you are having a good time at your new workplace.

Jill Kort: I have enjoyed all the interactions we have had and off course thank you for helping set up all the funding related to TA/RA.

Brenda: I have always enjoyed our random morning conversations after working long nights and before going back home to park my car so it doesn't get towed. You wake up so early! Thank you for the fun conversations and for saying I will be a valuable dedicated addition to wherever I go. It made my day. Wish you the best.

Taerow: Wish you the best ! And happy holidays.

0.5 Other Important Mentors

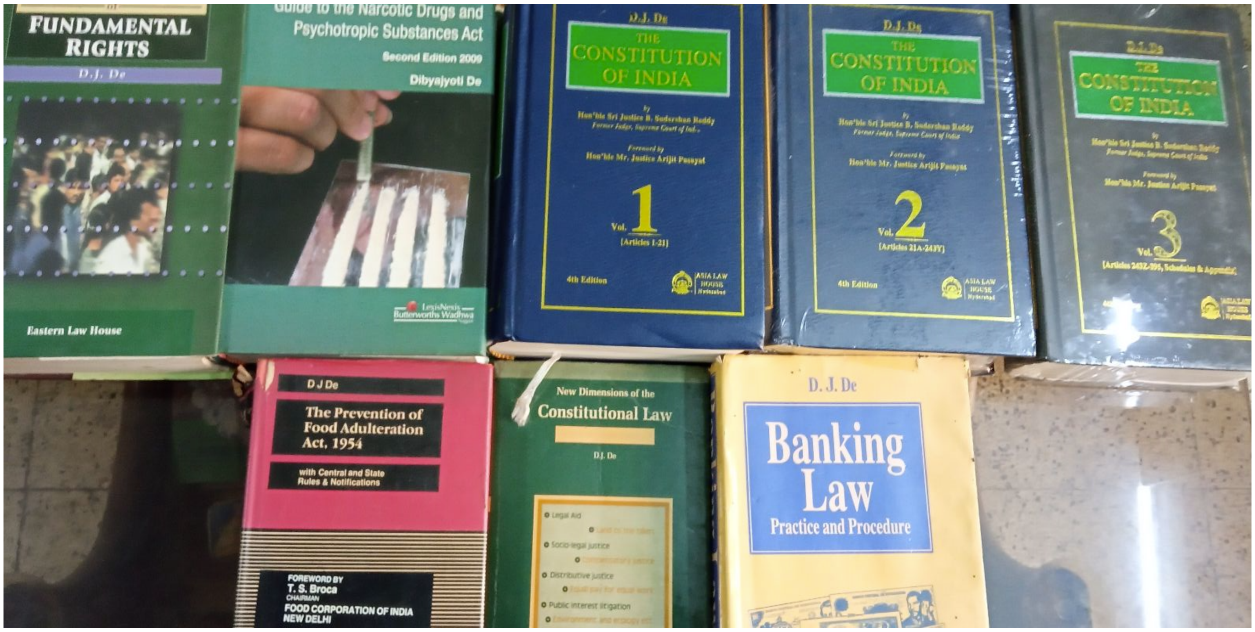
Dr R.C. Haselby, DO: The first time I met you was during my first presentation at the Marshfield Clinic Health Center where I was interning at the department of Precision Medicine. You caught my attention via some of your critical remarks regarding one of my slides. At first I was a bit irked, at least while presenting. However, once I went back to my cubical and thought about the things you said, I realized you were completely right. Then I made an appointment with you to discuss my work. That was one of my favourite days at Marshfield, you taught me so many important things. One of those lessons I still remember is about how people at middle age wants to impart more knowledge to the younger generation and they are at their peak to teach and motivate the younger generation. You also hooked me up with a full rotation with radiology department and there I learnt so much more about how exactly radiologists work and what might actually be useful to them and that has helped me a lot in my career. Without your help, I would have never have obtained the ability to shadow that department. You also took me to your favourite pizza place at Marshfield and then showed me your workshop and the multitude of trees you planted yourself years ago. It was a really great thing to behold. While I am too broke to afford good tools, I actually started getting into wood crafting myself and can make ornate and small functional objects. Furthermore I plan to buy a farm as soon as I save up enough at my new job. You also gave me important lessons regarding whistle blowing and helped me out when I had a moral dilemma. Your talk on nature vs nurture and your recommendation regarding picking up a medical degree along the way has actually shaped my future goals. Initially at Illumina, I am going to work more on data sci-ency stuff related to computational genomics and finding patterns etc in the genome. However, I am at the same time searching for resources and mentors to teach me molecular genetics. Well worst case I just hack at it from random points and learn it. Later on I do plan to get a medical degree too. But genetics and the ability to understand and work with the human body is something which has come out of all the talks we have had. So thank you Dr Haselby for helping me find my direction in life I would like to explore. Finally I want to thank you for the amazing great courses you have given me on behavior as well as on the history of the world. I never listened to it while going on cross country trips like you intended, thought I have driven from GA to Utah in 1 day once, I mostly listened to it while I did my work at my lab. All in all, you have helped shape up the direction where I will be headed in my future and for that am very grateful. I wish you and you family good health!

0.6 Family

Grandpa , I still somewhat remember a fraction of the stories you shared and it had made me curious about the world which in turn made me seek out more than just a job after my under graduate. At this point on my own time, I explore physics, chemistry, medicine and in future I will be working on genomics. The world is an amazing place and your stories sparked off that curiosity. Thank You for raising my mom on your own.

Grandma , I dont have that much memories of the time we shared. Just random fragments of this and that. I remember crying when you were leaving I think that was the last time I saw you. Thank you for being a rock and raising my Dad and his siblings.

Dibyajyoti De: Father, you have been a source of power for me at every point in life. Any time I have fallen down I recalled the different advises you gave me. The one advise of yours I always recall and live by is that "If you see something unjust, step in and do something". I appreciate this more and more these days with how people talk about the world needing changes but never really do anything to contribute substantially to bring about the change. I thank you for all the sacrifices you made and the things you never got to experience so that your children would make it well in life. You were an amazing judge and professor of law when you were alive and have always been celebrated as being efficient and thorough. However I really wish I was able to inherit your skills of writing. I cant imagine how you wrote all those big law books. I never have the patience for that maybe with age I will get better.



I am glad that I have pictures of you, because I sometimes seem to forget how you used to look. Reminds me of how fleeting life is and is it also why I make the effort to document everything that has been near and dear to me at some point of time. I feel nostalgic writing this particular paragraph, since we both share the lack of interest in accolades, yet I am giving you one by adding your memory in my dissertation. It also begs me to research into the human genome and figure out how and when we developed this sense of remembering the past and why it weights down on us so much. Because from a genetic point of view, you having fathered a child have successfully fulfil-ed one of the primary requirement of life that is to pass on your genetic code. The bond between parents and children is so strong. Is it because of the will of the genome which is passed on? I wonder. The genetic legacy that we possess is old and wise and I cannot wait to investigate into it even more in the near future.

Suchandra De: Mother, you are the strongest person I have known in my life. You have held the fort down so to speak and taken care of me and my brother the best you could have had. Even now living by yourself you are able to explore new things in life and cheering me and my brother on to succeed. I appreciate you very much and hopefully once I am able to renew my visa, get my EAD and maybe a couple of month into my new job I will come see you. Your outlook on life is quite different from mine in many ways but as I grow older I am able to understand where you are coming from even if I do not agree with them. Thank You for always being there for me no matter what. It may have taken me many years to appreciate everything you have done but at-least I have matured enough to realize it. This is a big turning point and finishing my PhD is quite possibly the biggest thing I have accomplished in my life upto this point with the help of many many others. You should get a passport so maybe sometime in the future, I can at some point introduce you to the amazing people who have shaped me as I am now, in the past six years. Thinking about dad reminds me everyday to appreciate the living while they are living. So expect more visits from me now that I will have enough money and time to move about frequently.
Saunak De: Well you grew taller than me though I am older. I dont know what to say really since I am aware you are going to call me names if I say anything too sentimental. Yet let me just say I admire the bond we share and the fact that we have always had each other's back even if we were not around each other. I am well aware of many stories where you got in fights because someone said something bad about me. I hope that we get to work on something in common in the near future. Thanks for taking care of mom while I am away in a different country. Take care of yourself and your health.
Woof: You were the best-est girl. Pat pat little sister.

My trip back to India in 2019 helped me prepare for the final stretch of my PhD and it also helped strengthen my mental. I wish to thank **Dr Bishwajyoti Dey and Dr Durga Nandini**. Ranga and Durga kakima, you guys have been inspirational to me and convined me to follow through and finish my PhD as well as stay in the US after graduation since the best research is conducted here. I wish you both good health and cant wait to see you again soon. Also Durga kakima I somewhat agree with your political opinion. But saying that, I feel politics will always be dirty and all we can even hope to do is be responsible for our own actions and simply be good to the ones around us. I feel there are three things in life we need to be happy ; food shelter clothing, a partner, and a passion for something which we can do. Lacking one of these probably makes people bitter and drive them to find relief through aggression or identifying with groups and forgetting the uniqueness of themselves as individuals. Also the first of the three is the most important and which is why we see most of the violence amongst the poorest parts of society. I do have good plans to address at least the food part of the trifecta and hopefully in the next 10 years or so I will be able to alleviate some of the troubles. I wish to thank **Dr Monojyoti Dey and Ruma Roy**. Chotokaku and Ruma kakima, thanks for encouraging me to finish my PhD as well. Chotokaku I appreciate you explaining to me about quantum mechanics while we were taking a trip in Shillong. It also helped me realize how quantum computers work as well as help me formulate a few levels of unsupervised learning one of which I implemented immediately after returning from the trip. I wish you both good health and cant wait to see you again soon. My little sister **Sukanya Dey**, you have always proactively kept in touch with your aloof brother. I really have a bad case of "out of sight out of mind" for day to day activities. You are an amazing daughter and Choto kaku and kakima are really lucky you take such good care of them. Also thank you for keeping in touch with mom and making sure shes ok. I also want to congratulate you on your new job. Thank you for keeping me motivated, sharing stories and discussing so many things. It is always a pleasure talking to you and your positive energy is infectious. See you soon! **Debjani Duttachoudhury and Ranjit Duttachoudhury**, Assam is beautiful and more so because you guys live there. I cant wait to visit you guys once again. **Rajarshi Duttachoudhury and Moutushi Chaudhury**, Rupom dada and Muku didi I was happy to see you both have neighboring apartments. It also was exciting to visit and see your kids, now me being the young adult. Just a brief while ago I was the kid and you guys were the young adults. How beautiful is the cycle. Wish you all the best in life! I would like to thank **K.R. Deb, Kharna Deb, Pinki pishi, Partha Deb** and I remember all of my visits to your place fondly. Mom and dad would be talking to you guys while me and my brother would watch cartoons. It does feel sad to not see Fulmama dadu but he will forever be there in my memory. It was really nice to see the new generation on my last visit as well.

I owe a great deal of my success to **Indu Madhav Ghose**. You were the oldest person I have known, and one of the most knowledgeable person I have met. What you have imparted to me is unique and am privileged to have met you. Also I thank you for encouraging me to become a scholar and go abroad, you were a great influence to me. I would also like to thank **Mamoni Dida**, your kind nurturing nature is something that has rubbed off on me. You also helped in raising my mother for which I am eternally grateful. You were always so full of energy and I hope I am able to carry myself like you and mejdadu when I get older. May you both rest in peace. I want to thank **Keya Das**. Keya mashi, one of the reasons I can peacefully conduct my studies here is knowing that you are there with my mother at all times. I appreciate that you never let her feel alone. Similarly **Santa Mashi, Rini Mashi** I am eternally grateful to you guys for the same reason. **Raja mama, Bappa dada, Mampi didi, Tutun dada and Tukui**, growing up with you guys was a lot of fun. And you all supported us so much after dad passed away. I am thankful to you all for also making sure mom does not feel too alone. I can only do what I do because I can have that peace of mind. Finally, I also want to thank **Bijendra Nath Mazumdar** for teaching me a lot about business and explaining to me the benefits of someday owning my own business. I do plan to have something of my own one day and would love to seek you out for more advice.

0.7 Friends

I am grateful for having met a lot of great people here at UGA. I would start with my first friend here, **Schun Uechi**. Schun san, I miss all the parties we used to have during the first two years. You have taught me a lot about astro physics and got me interested into neutron stars. Starting there I have kept exploring and I plan to keep doing so. Many of the algorithms I have developed can actually be helpful in separating meaningful signals from astral events from already known ones and that is one research I wish to pursue in future. Hopefully now that I am finishing up my degree here, I will finally have time to get back to the collaboration we were working on. You have always taken good care of me and been like a brother to me. I wish you Eri chan, Luke, and your next baby good health and hope to come see you in Japan sometime. I would also like to thank **Yang**, I will make sure to meet up with you before I leave for California. Also thanks **Burair** on all the fun times. It was really great meeting you and your family and the time we spend discussing different events and things while grabbing some food or drinking. Furthermore most recently you gave me so much to think about in regards to my new job. I wish you and your family the best and maybe we can collaborate on future projects sometime. I also would like to thank **Dr Kübra Benlİ** . You provided me this latex template and a lot of joy and happiness during the times we shared. We have had a lot of fun adventures and I fondly remember all of them, even the not so good ones. You always provided good counsel and I really miss just hanging out at your place from time to time watching movies, talking about things etc. Also currently almost all of my cutlery, and other stuff are the ones you left behind haha. I hope you are doing well in France. I wish you well and see you sometime in Paris or Istanbul next year. I would love to work with you, Schun, and Nalina in the future once I lead my own team, hopefully in a couple of years. Speaking of which, I am glad to be reuniting with **Nalina** in California in the future. Your work with monkeys I feel will be really useful and maybe we can collaborate sometime in the future as well. I also like to thank **Meltem Safak**, we have been through so much. You basically made a traveller out of me. I love all the places we visited. Never thought I would visit Cuba really. But it was a refreshing trip, and reminded me of India. Thank you for always being a friend and I cant wait to attend your and Daniel's wedding this December. You research has always been interesting to me and I wish you all the best in your future endeavours. **Dr Ceren Ocak** am glad our collaborative paper on education and AI got accepted and hope you are doing well at your new job. Wish you the best and hope you get everything you want in life. **Ilkiz**, it was good to reconnect after the long break due to covid. Its always nice to see your smiling happy face and your positive energy to most things. You always remind me to have a free spirit and I wish you the best during your time at UGA and beyond. Also let me pet Sunshine!. I would also like to thank **Aditya** for taking over GSA. That is a big responsibility and I am glad to see its in good hands! I hope to grab a beer with you before I leave. **Jane**, I have gotten to know you since I was assigned to be your mentor. I wish you all the best in your future endeavors. **Jayant**, thanks for sharing the concept of personality. It really helped me a lot by fixing some of the issues associated with my ENFP personality type. **Daniel** Its been good to get to know you. Your story about how you quit smoking will be something I shall never forget. I cant wait to attend your wedding. Congratulations and have an amazing life you and Meltem. Last and least. I jest, Last but not least I would like to thank **Dr Omid Setayashfar**. You have gotten me into wood crafting, introduced me to amazing Persian cuisine and what not. The list is endless. I am glad to have been your friend and I am even happier that we will continue to be friends past our time here at UGA. Not just that but you also fell in love with my lab mate and am so glad you guys got married during our PhD timeline. Sadly I couldnt attend as I was away on my internship but I wish I could have been there. Wishing you the best at Meta, if possible can you tell me what keywords are associated with me on facebook. I jest. Dont do that. But keep being amazing.

I would also like to give a shout out to my friends back home, **Sagnik and Aishwarya**. Am glad we somewhat keep in touch. Well mostly with Sagnik but would love to catch up to everyone else once I get back to India. I also want to wish another childhood friend **Shamik** the best of luck in his upcoming nuptials. I hope I can attend it on the 12th in NYC but even if I cant I wish you guys the best. I also hope I get to reconnect with you **Sarthak**. Its been a while since we last spoke.

For the past one and a half years, I would have suffered from crippling depression if it wasn't for the group of friends I have met through playing games. Thanks for making sure I never felt lonely or even if I did it was only temporary and harmless. You guys are the ones who helped me through the final stretch of my PhD and I am really lucky to have met you all. As we all keep advancing in life lets hope we dont get too busy to still be in touch with each other. I want to start of by thanking **Cheldon Thomas**. We have been friends for over 10 years now playing games since during our high school days. I guess that makes you my oldest friend with whom I speak with on a regular basis. I clearly remember back in highschool when my character got killed by someone and you came straight up and fought alongside me since then you have always been a co leader in the old game we played and well now you lead this one. You have also taught me a lot about crypto currency, aside from letting me play your characters for raids. I cant wait to use your advise in investing my money once I start my new job. Same goes for **Dara Leang**. I am thankful for your advise in investing and doubling the small amount I invested in a short time. You have a lot of good intuition about life in general, and I absolutely love the stories about your life in all the various scenes you worked at in the past. I wish you the best and hopefully we continue playing a bunch of games together. I am really grateful to **Ngoc Lieu** for helping me fix my CV. Ever since I took your advice on updating my CV, I have gotten a lot of calls for jobs. I cant thank you enough on your invaluable insight and all your information on corporate life as well as helping lead our BM raid. Next I would like to thank **Mike**. Thanks for the free

dental healthcare advise. It is much appreciated. You have an amazing knack for spotting details and organizing things. I also want to thank you for organizing multiple raids and in general keeping things going and never ditching us even though you have a lot going on in life right now. Also I appreciate you listening to my thesis presentation talk, the night before my defence and giving me some nice pointers. I wish you the best and who knows, our joke about making a gaming house might be a reality someday. Next I would like to thank **Judy**. When I think of Judy, the first thing that comes to mind is a nicely cooked crispy skinned peking duck. I jest of course since you love ducks so much. Get it?. You are the glue which has held our clan and in turn my sanity together for the past year. I applaud you for dealing with a crazy amount of people during your day time job, taking classes and yet finding time to play games and watching movies with us. You are a very good hearted human being and am lucky to have you as my friend. Next I would thank **Alex/Amarans** our go to tank for everything. With you joining our group we have been able to do so much more content. Not just that you have made our clan very active and in turn I have enjoyed doing more content alongside you in my free time. You are a very skilled gamer and a positive influence on everyone around. Next I would like to thank **Luna**. Alongside Amarans, you have helped our clan become active and helping everyone push on when we were just tired from daily life activities. You have always been flexible with helping us out by learning difficult roles and always being there to help out with everything with a cheerful attitude. I wish you the best in your school work as well. **Alyssa**, you have quickly become a really close friend to all of us and specially mine. You are a mature individual and your advise on life cannot be understated. You also helped me understand the different levels of corporate hierarchy and gave me ball park figures of what to expect. Aside from that, you have helped me when I have felt depressed by simply stating the things I couldn't see at the moment. I wish you and Alice the best in life and see you in SF and hope you are ready to tank our next raid. Also thank you for hosting multiple viewing streams to make us all feel included while we did our daily tasks. **Michi**, your eyebrow gif has indeed become my favourite gif to spam everywhere and besides that you are an important member of our Dawn family and hope you get back to playing with us again. **Roberto Ortiz**, you helped me learn to play my character since you were pro at it already and made my life getting into the game easier. Aside from that you have given me a lot of pointers on being a good teacher, and organizing priorities and for that I am very grateful. **Teresa**, you have helped create a fun atmosphere for all of us and joining us for raids and other events. Though you have a job and two kids, you have taught me that its still possible to do things for yourself even when there is so much responsibility. You are a great role model and hopefully I will be able to take a few pages out of your book and learn to manage my time and energy fruitfully. **Lamp**, you left your old group and decided to roll with us. I love all the shows you stream and also enjoy your sense of humor. I want to thank you for sticking around even though you are quite bored of the game . It definitely has helped me a lot having you and the others around in feeling more connected to people. **Gabriel/Snorlax**, you are one of the smartest students of computer science I have seen. And I have been a TA for multiple courses. Your attention to detail and the ability to help fellow students already shows signs that you will make you a great team leader. You could also become a researcher since you are naturally curious about things around you. Your coding skills are unparalleled for someone just starting out in computer science. In game too you are amazing at adapting to all kinds of different and unique scenarios at any given notice. It takes a lot of skill to be that accurate and that flexible. By the time you graduate I should be leading my own team. And you sir would be someone I would like to be on it. I also love your dry humor, never change thank you. **Eania**, I enjoy our long conversations about so many different things and I especially respect and love your outlook towards life in general. Thank you for always lending an ear whenever I go on random rants and sharing your insights. Hopefully you come visit the states since I dont see myself renewing my visa due to covid. Please bring over a moose when you do come. Also I want to thank you for discussing shortage of food with me and lo and behold, the idea of building an auto detector/separator bot for restaurants to re-purpose some of the wasted food pops up. **Erick Martinez**, you have had the last and final influence on my dissertation by ensuring I wouldn't fall asleep and finish my work. "What we do in the Dark" was an amazing show and having it play in the background is what helped me complete my presentations as well as sections of my dissertation on time. I think next Halloween I shall be "Nandor the Relentless". Thank you for that and other amazing suggestion. You have also been a good friend and have helped me from going down the ugly slope of depression and alcoholism. I really appreciate your comments and the discussions we had about how to handle things when life gets difficult so thank you for that too. I am looking forward to meeting you in Vegas. Hopefully its not that far driving from Foster City. I would also like to thank **Alex, Mokano, Sandra, Jinu, Steam, Phantom, Neruno, Twi, and Hux** for being a part of our group and making the games fun and enjoyable. I would have never gotten a good score on my GRE language section if it wasnt for the book you suggested, "English words from Latin and Greek elements", **Cody**. I hope you someday finish your own book. It would be great to finally get to read it. You are an amazing philosopher and you better share your wisdom with us. Wish you the best and hopefully soon I will get to stop by Texas and finally meet you. **Adam**, my love for gaming started with the guild we created and watched it grow to be the top guild on the NA server. I respect your leadership and political skills and learnt a lot from you which had an influence on the type of person I am now. Wish you, your wife and kids an amazing life ahead. Finally this dissertation wouldn't be complete without mentioning **Zeyu Zhang**, since I might not even be here at this institution if you didn't proof read my statement of purpose and help rewrite some sections. You are not with us anymore, and I hate the fact that I will never get to meet you since one of my goals is to meet each and every person who has been influential to me in game, in real life, and perhaps collaborate on something big, eventually. Thank you for your kindness and for helping a

friend out. I have yet to visit Maryland but hopefully one of these days I will get the chance to place some flowers in person, at your final resting place.

I am aware this acknowledgement is a bit long. That is obviously an understatement. However every person I have included have played an important role in me finishing my PhD.

Contents

Chapter 1

Introduction

1.1 Motivation

Pixel wise classification of a Region Of Interest (ROI) or image segmentation is an important concept in the field of medical image computing [79] and computer vision. It is the preliminary step for isolating and conducting non-invasive diagnoses of cells, tissues and organs. Examples of applications include skull stripping [42], brain and liver tumor segmentation from different modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), X-rays and Positron Emission Tomography (PET). Often time networks trained to identify the ROI may fail to segment it in the presence of uncharacteristic features in or around the ROI during testing. Leveraging features and information, complementary to the ROI provides an invaluable resource to enhance the performance of a ROI segmentation algorithm. In this dissertation, I explore two ways of leveraging information complementary to the ROI, in order to improve performance: 1) Under a supervised setting, where a ground truth of the ROI or strong guidance is provided e.g., skull stripping where the complement is the non brain regions, brain and liver tumor segmentation where the complements are the non tumor regions in the brain and liver respectively; 2) Under an unsupervised setting, where a collection of images is provided by the user to form a reference distribution and any deviation from that collection is flagged as anomaly or novelty. The two complementary sections in this case are the in distribution portion and the out of distribution portion of the input image with respect to the user defined reference distribution.

1.1.1 Supervised Segmentation

The advent of deep learning has given rise to a multitude of supervised segmentation methods for medical image analysis. The primary architecture that serves as a strong baseline for many supervised segmentation algorithms is the autoencoder with skip connections, the U-Net [80]. Given an input image the autoencoder structures aim to map it to the ground truth mask associated with the ROI. However, the restrictive nature of training a deep neural network [30] via such processes makes the neural network focus only on the features relevant for the extraction of the ROI. As such, presence of very large deviations within or around the ROI in the test set can cause the networks to ignore the afflicted sections of the ROI since those features were not seen during training. In contrast, if additional information pertaining to the remaining sections of the image space is also known, the network is able to infer the ROI region via it. One of the contributions of this dissertation is the formulation of a framework which leverages the idea of complements in the image space to train a network to learns both ROI and the complementary non ROI features without additional guidance and thus successfully operates even if the ROI or its neighborhood is warped, producing a robust segmentation. This framework is tested on the OASIS dataset [60] focusing on the objective of skull stripping and obtaining a state of the art performance on two fold cross validation as well as demonstrating the framework's robustness by simulating tumors in one of the folds used for testing. The framework also obtains the state of the art result on LiTS [11] test set without any pretraining or post-processing.

1.1.2 Unsupervised Segmentation

Classical unsupervised image segmentation [89, 72, 28, 77, 53] has been an active research topic for decades due to its potential of applying to many applications without requiring the data to be manually labelled. Existing anomaly segmentation frameworks are motivated by the research in [87, 86], where the framework learns to reconstruct manifold of healthy images only and in the presence of input with novel/anomalous sections, fails to produce a complete reconstruction of the anomalous region. The difference of the reconstruction and the input image i.e. the residual, is flagged as the anomaly. However, this framework is susceptible to the common instability issues of the Generative Adversarial Networks (GANs)[31]. Furthermore, since the expectation of learning the underlying manifold is dependent on faithful reconstruction, the framework has an additional burden to learn a good reconstruction. To alleviate these issues, in this dissertation I introduce an Unsupervised Anomaly/Novelty Segmentation framework based on complementary learning which produces a constrained two cut guided by a set of images with the anomaly being complementary to the in-distribution sections of an input image. This bypasses the residual method and instead focuses on the anomaly itself. Given images which constitute a loose definition of the set called "normalcy", the framework aims to figure out all possible deviation from it via a selective two cut and adversarial learning. The framework obtains promising results on the BraTS 2019 [4, 5, 63] and LiTs [11] datasets as well as obtain state of the art result on the public dataset MS-SEG2015 [13] in comparison to all existing anomaly segmentation methods, obtaining 20.4% increase in the mean dice score. I also verify the efficacy of the method on a private dataset where the reference distribution has verified healthy scans unlike the non tumor slices of the public datasets where there may be co-morbidities outside the scope of measurement without any medical professional's help.

1.2 Complementary Learning

In each of the two contributions above the primary idea is the notion of a complement. In essence a complement to something is the remainder which makes it a whole. In the case of image segmentation this can be achieved via two non overlapping feature space.

One way of obtaining a good segmentation of the ROI is via the use of the dice loss. Given two sets A, B the dice coefficient $Dice(A, B) = 2|A \cap B|/(|A| + |B|)$ [25] is a measure of the similarity of the two samples with the domain value ranging from $[0, 1]$. This loss measures the intersection of the two sets over their union thus making it a good function to optimize for segmentation. Instead of maximizing the dice coefficient as in the case of segmentation, I minimize it, resulting in sets which are disjoint. A dice loss equal to 0 means no overlap. The sets A, B are now disjoint and complements of each other.

Definition 1 (Complementary Constraint)**.** Images which cannot simultaneously contribute a positive value at the same coordinate in the image space formed by their union.

Definition 2 (Disjoincy Loss)**.** The loss function implemented to enforce the complementary constraint for a set of images. In this dissertation we only consider pairs of images.

In Chapters 2, 3 and 4, I incorporate the idea of complements into neural network frameworks for the purposes of supervised segmentation and unsupervised anomaly/novelty segmentation.

1.3 Thesis Statement

Thesis: *Complementary learning helps to utilize the entire information content in the image space (ROI and non ROI). Complementary learning augments can effectively enhance the robustness of a supervised framework by learning non Region of Interest features. In an* *unsupervised setting it can identify outliers from a reference distribution thus aiding in anomaly/novelty detection.*

The contributions of this dissertation are:

(i) Formulating an easy way to leverage the complementary information in the image space by minimizing the intersection over union of two regions. The resulting regions are consistently disjoint.

(ii) A model, Comp-Net, for supervised learning which represents the first case of the disjoincy loss formulation, where one cut is defined by the ROI and the other cut is obtained via the disjoincy loss. It can perform segmentation faithfully even in the presence of synthetic anomalies absent during training. It is tested on two applications, skull stripping and liver lesion segmentation.

(iii) A constrained two cut formulation for unsupervised anomaly segmentation where one of the cuts is learnt in an unsupervised fashion using a neural network training in an adversarial fashion. It outperforms existing state of the art methods on MS-SEG2015 dataset and show promising performance on BraTS2019, LiTS Liver Lesion and a privately curated brain tumor dataset. The two cut allows us to move away from reconstruction dependent anomaly/novelty segmentation and moves the focus back to the anomaly itself.

1.4 Overview of Chapters

The remainder of the dissertation is organized into the following chapters:

(i) In Chapter 2, I present the background information relevant to understanding this dissertation.

(ii) Chapter 3 introduces the first of the two algorithms based on complementary learning for the purpose of *skull stripping* using the Oasis [60] dataset. It augments an U-Net [80] to take advantage of the complementary information, making it more robust and tolerant to changes in the data distribution of a modified Oasis test set. A concept of multi-outputs is also introduced.

(iii) Chapter 4 further demonstrates the efficacy of the supervised algorithm in Chapter 3 in relations to segmenting small objects, using the public LiTS [11] liver lesion dataset and obtaining the State of the Art results on the challenge test set without any pretraining or post processing.

(iv) Chapter 5 introduces a novel way to incorporate complementary constraint to perform anomaly/novelty segmentation based on a reference distribution specified by the user.

(v) Chapter 6 concludes with a summary of the presented work and provides future directions this idea of complementary learning may be progressed in.

Chapter 2

Background

2.1 Introduction

This chapter introduces some of the historic algorithms which were relevant to the area of image segmentation. Following this, the focus is shifted towards some of the basic operations of image processing and how neural networks encapsulates them and make the tasks easier thus becoming the algorithm of choice for image segmentation in the recent times. Next, I discuss the progression of neural network architectures and ideas leading up to the U-Net for medical image segmentation which is improved upon in Chapter 3 to embody the complementary learning concept and discuss the core idea of Anomaly/Novelty Segmentation as described in the pioneering paper [87] in this field, which is redefined in Chapter 5. Next, I explain the advantages of using the Complementary constraint and the disjoincy loss and the simplicity of its use with neural network frameworks. Finally, I conclude this chapter describing the different imaging techniques involved in medical image analysis relevant to this dissertation.

2.2 Classical Techniques

In this section, I discuss some of the prevalent ways the image segmentation task was handled prior to the deep learning paradigm shift.

Thresholding

One way to obtain a segmentation is via thresholding. Thresholding attempts to cut an image histogram at different peaks or valley points to produce a subset of the input image, aka global thresholding. Other approaches subdivide the region into subsets and take thresholds based on the histograms of the sub region. Another popular thresholding method is the Otsu thresholding [73] which aims to iteratively find splits in the image histogram such that the variance within the subsets is minimal while the variance between different subsets is maximal. One of the glaring disadvantages of threshold based methods is the lack of spatial awareness which is overcome by neural network based approaches.

Clustering

One of the most popular means of segmentation via clustering involves fuzzy c-means clustering [10] which is simply a superset of k-means clustering and allows for different pixels to belong to multiple clusters unlike k-means where the membership function is set to 1 or 0 at termination i.e. a pixel belongs to a single cluster. Multiple application of fuzzy c-means clustering has been used for medical image segmentation [94, 96, 62].

Another popular method which has influenced this dissertation considerably is the Normalized N-Cut segmentation [89] wherein the image is represented as a graph with all the pixels forming the vertices/nodes and whose edges are weighted based on similarity of the two pixel/nodes in question. The primary advantage of this method over a traditional min cut lies in the penalty for cutting off isolated nodes due to the normalization factor in the denominator which is dependent on the two node's connectivity to the rest of the nodes in the graph. This in turn enables us to use this algorithm to obtain substantial segmentation clusters. In this dissertation, we model the anomaly segmentation problem as a constrained two cut problem via complementary learning described later.

Edge Detection

Edge detection involves detecting changes in the intensity across boundary of structures in images. Filters such as Canny [12], Sobel [90], Prewitt [76] are popular operators for this purpose. However, many customization's have been suggested over the years based on applications. An example of edge detection technique for segmentation of coronary vessels is presented in [91].

Region Growing

The primary idea of region growing is to start with some seeds which can consist of one or more clustered pixel groups and keep adding new neighborhood pixels to the groups if they satisfy a similarity criterion [67]. This process is continued till no more pixels can be added to the group. A popular method [36], the split and merge compare pixels with neighbors and if they satisfy a criterion of homogeneity, the pixel is associated with the neighbors. The choice of the homogeneity criterion is extremely critical for this purpose. Another flavor of region growing is obtained via morphological watershed segmentation [9, 64] which produces closed boundary segmentation. Multiple variations have been proposed since including an hybrid between split merge and watershed [2].

2.3 Basic Image Processing Techniques

In this section, I will explain the traditional core image processing techniques which are then adopted by neural networks and automated.

2.3.1 Image Filters

Image filters are tools to convolve an input image and produce a new image with some intended characteristics i.e. the shape of a function is modified by another. The two common ways to interact with any signal involves smoothing or sharpening it. The two different operations associated with it are the low pass filters or noise reduction filters and the high pass filters or sharpening filters. In the following subsections, I am going to demonstrate the capacity of using image filters using a simple and highly contrasting brain tumor image.

2.3.2 Convolution Operation

A convolution operation involves two functions or images. The first function or input image whose form is to be changed and the second function or the operator which acts on the former. The operator/kernel/filter moves across the image and performs a dot product between itself and a subset of the input image and stores the summation of the final matrix's value in the appropriate cell. Figure 2.1 demonstrates the use of the identity filter on an input image. It is to be noted that the edges of the input image/matrix is not preserved. It can however be padded to retain the same shape as the input image.

2.3.3 Low Pass Filters

Low pass filters are for example, averaging filter, gaussian filter and median filters, which convolve an image and denoises it or makes it smoother. The histogram analysis of a denoised

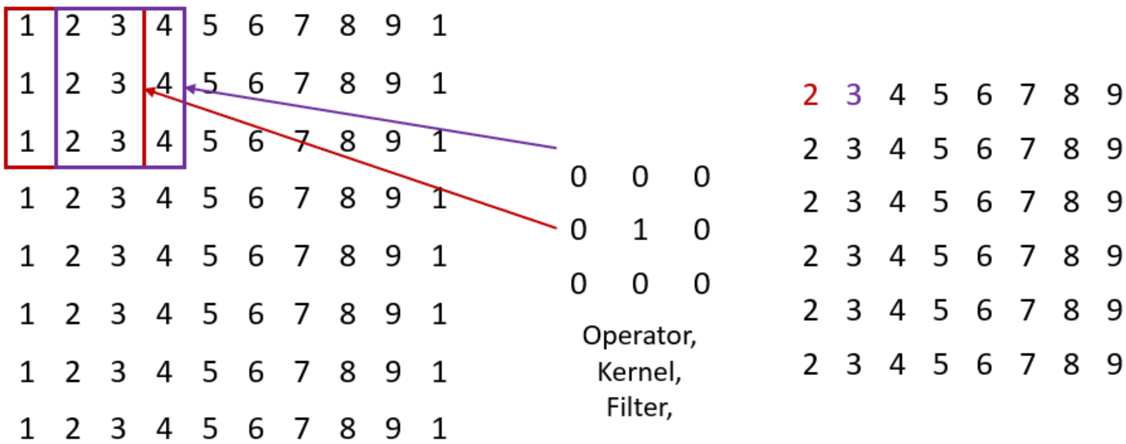


Figure 2.1: The Kernel/filter moves over the input image and produces a new image.

image in Figure 2.2 demonstrates the improvement in the interpretability in the image. Peaks in image histogram signifies areas of interest.

2.3.4 High Pass Filters

High pass filters are used for edge detection or to sharpen an input image. Examples of such filters include, Canny [12], Sobel [90] and Prewitt [76]. Figure 2.3 represents Prewitt, Kirsh and Sobel operators on an input image and their respective histogram. The peaks in the histogram tells us points of interest and using the edge detectors, we can see clear peaks around the area of the tumor pixels. Furthermore, if we overlay the created edge maps on to the input image, we get to see the tumor highlighted in more details as shown in row 3 of Figure 2.3.4. Some examples of both low and high pass filters are presented in Figure 2.4.

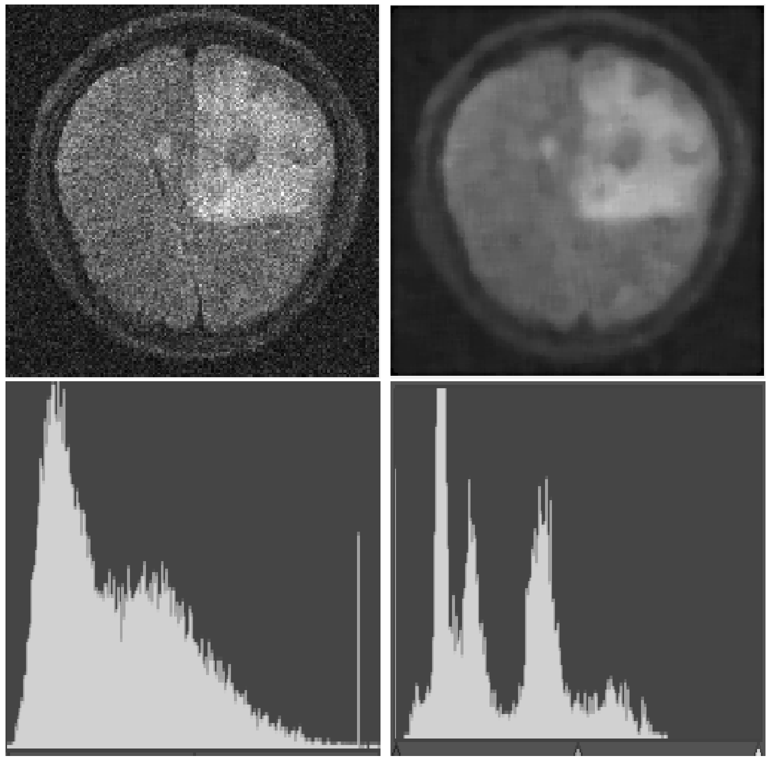


Figure 2.2: Left to right: Noisy Image, Median-Filtered image using filter size 9x9 and their histogram distribution.

2.3.5 Pyramid Operation

Often times there is a need to hierarchically apply operators and shrink the image. In the traditional image processing terms, this is called building the image pyramid as shown in Figure 2.5.

2.3.6 Deconvolution Operator

Like convolution operators, an inverse convolution operator takes in the input image and generates local duplicates of input pixels based on the spatial information encoded in the kernel.

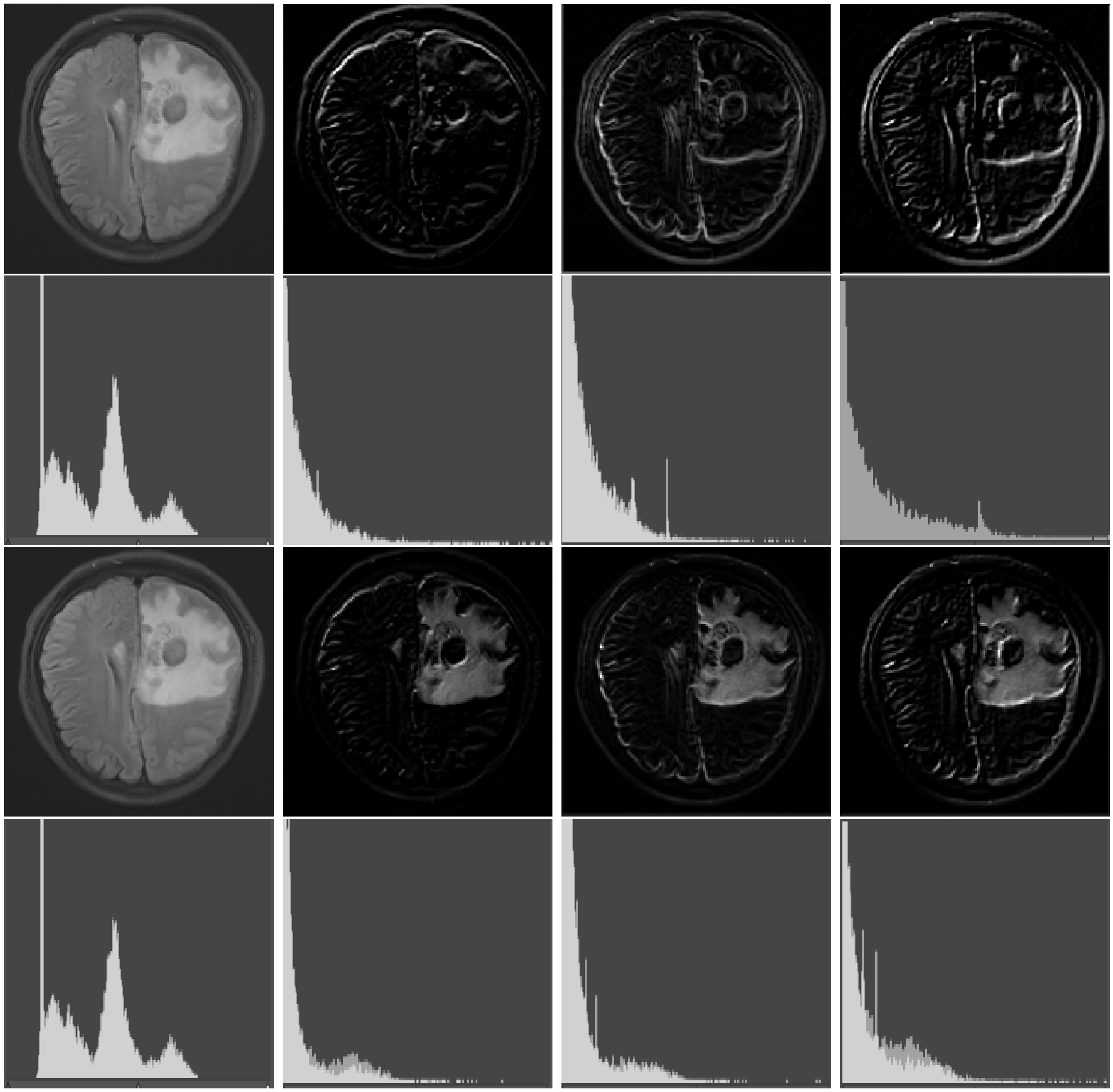


Figure 2.3: Left to right: Input, Prewitt, Kirsh, Sobel operators, their histogram, overlay on input image and their histograms.

0	0	0	1	0	-1	0	-1	0	
0	1	0	0	1	0	-1	4	-1	
0	0	0	-1	0	1	0	-1	0	
0	-1/4	0	0	-1	0	1	2	1	
-1/4	0	-1/4	-1	5	-1	2	4	2	$\frac{1}{16}$
0	-1/4	0	0	-1	0	1	2	1	

Figure 2.4: Left to right: Top Row - Identity filter, two edge detection filters; Bottom Row - An edge detection filter, Sharpening filter, and a 3x3 Gaussian filter.

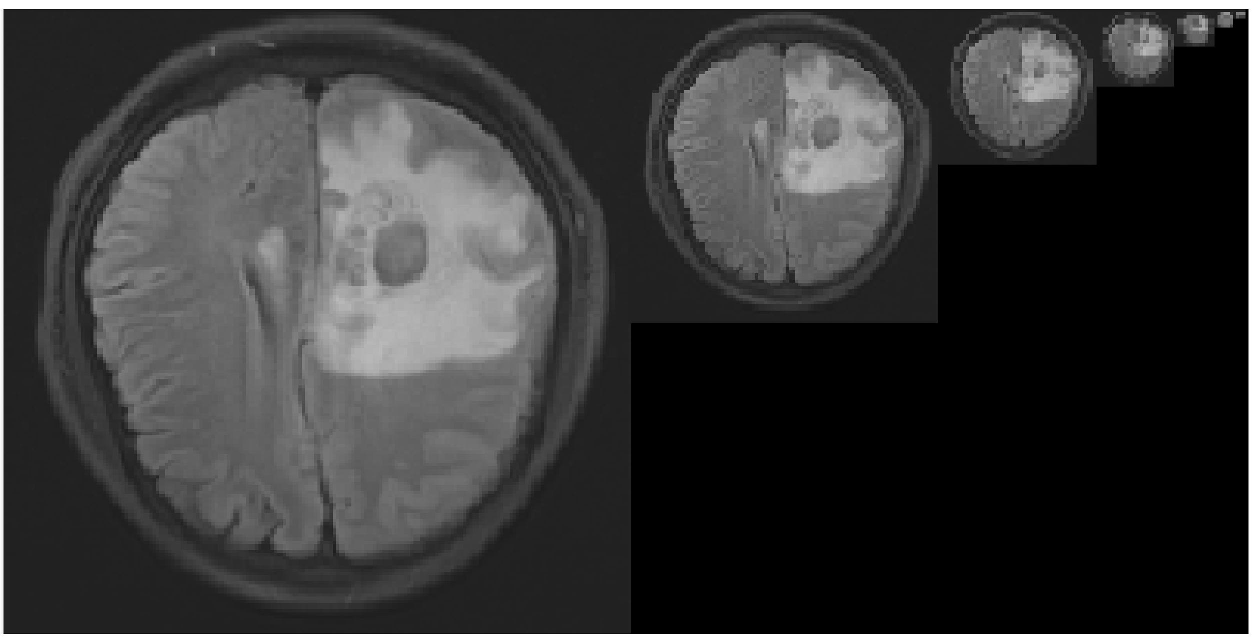


Figure 2.5: Left to right: A down sampled version of the input image in successive iterations.

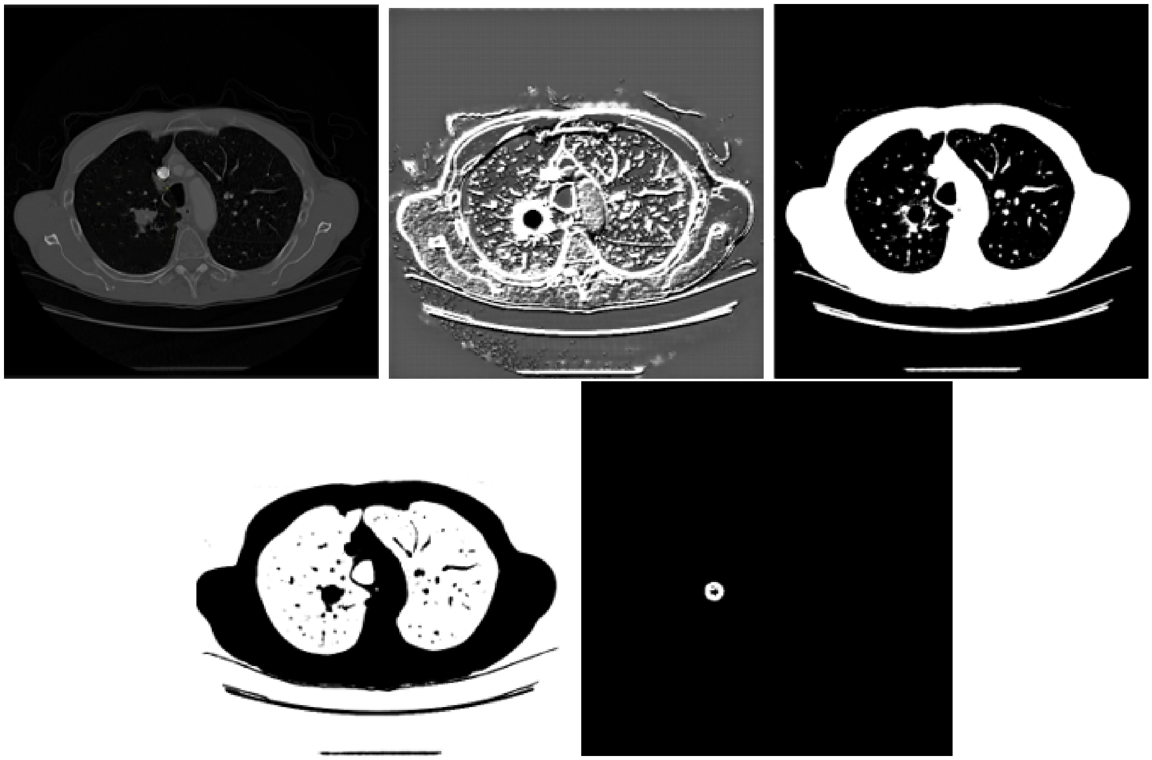


Figure 2.6: Left to right: Original Image and its different representation as it passes through successive layers for a tumor classifier network. The nodes demonstrated were chosen at random and are obtained from different layers.

2.4 Neural Networks

2.4.1 Introduction

In the above example the combination of different operators could enhance the tumor and if I took thresholds based on peaks of image histograms, I could extract the tumor. However, this is a non trivial task and the threshold point as well as the selection of the optimal operators required to achieve the problem is non trivial. Furthermore the image was carefully selected to have easy to find peaks in the first place. In Figure 5.3 we can see the different types of input histograms for different images. For smaller tumors it is even more blurred. Furthermore, edge detection operators are often times incremental operators, and the optimal choice of the correct increment is also a difficult problem. Thus, the problem of finding an optimal filters and applying them in the correct order becomes intractable for a large dataset. This is where neural networks shine. Figure 2.6 one can observe the neural network apply a variety of filters or operators which are learnt via back propagation to isolate markers for a tumor. At one of the layers, it produces a circle at the co-ordinate of the tumor. Following which it can correctly produce a label of true for tumor detection. Unlike hand crafted operators, neural network learns the operators, which can be non standardized and specific to the dataset. Furthermore the idea of image pyramid shown in Figure 2.5 is also used to further reduce the image space retaining only features relevant to the task. Aside from that, the operators, and sequences in which to apply them, are learnt to generalize across the entire dataset which is why neural networks have replaced hand crafted operator and operator sequence engineering. Neural Networks are composition of piecewise linear operators or functions with added non linearity which enables it to approximate any function imaginable and thus they are also called Universal Approximators. In the following sections, I will explain the working of the neural network including the back propagation algorithm which has caused the popularity of neural networks in recent times.

2.4.2 Tensors

Neural networks operate on tensors. A 1D tensor is normally used with a perceptron or a 1D convolutional network. A 2D tensor or an image is processed by a 2D convolutional neural network. A 3D tensor or a volume is processed by a 3D convolutional neural network. In each of the cases the biases as well as the edge weights represents their corresponding input tensor dimension. In the following sections, I will go over a basic perceptron or 1D convolution operator and then generalize it to a 2D framework which includes images (2D tensors) and 2D convolutional networks. The 3D variant can be inferred by the reader since the extension is the same as going from 1D to 2D and thus will not be explored here.

2.4.3 Architecture and Components

A basic neural network (also known as a perceptron or fully connected neural network) is a network that operates on 1D tensor inputs. It consists of an input layer followed by a few hidden layer and finally an output layer. In Figure 2.7, there are two input nodes in the input layer, two hidden nodes in the one hidden layer and one output node in the output layer. The signal from the any node is multiplied by the respective weights of the outbound edges. All nodes aside from input nodes accept the sum of all the signals incident upon it. There is also a bias term common to all nodes except the input layer nodes, which contributes to the aggregated signal at a non input node. *One can think of an node learning from its previous layers as well as including its own bias or its own knowledge. This is synonymous to the way we as humans act; we are influenced by our predecessors as well as our personal bias in taking an action.* Finally, there is a component of non linearity which is applied to the aggregated signal onto a node which determines if it should fire a signal forward or not and its magnitude. The commonly used activation functions in this dissertation are, sigmoid [97], tanh [78, 50, 85] and Rectified Linear units (Relu) [68].

Convolution Neural Networks The primary difference in the convolutional operations or Convolution Networks (Conv Nets) from basic neural networks involve replacing the input signals with images or 2D tensors and a kernel/operator similar to ones explained in the previous sections are **learned** as weights. In essence the one dimensional weights in a simple fully connected neural network as shown in Figure 2.7 are replaced with two dimensional filters which are learnt during training. The biases are also two dimensional for convolutional networks.

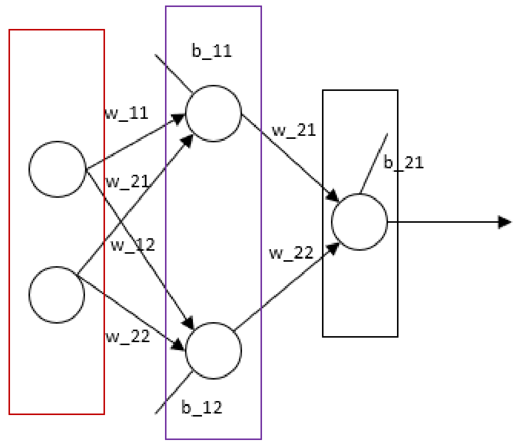


Figure 2.7: A simple neural network with one input layer in red, one hidden layer in purple and one output layer in black. The number of input nodes is two, hidden nodes is two and output node is one. The w_ij represents the weights and the b_ij represents the biases.

2.4.4 Feed Forward and Back Propagation

Feed Forward

After the creation of a neural network architecture, it is initialized to some random weights and biases. The first step involves a dot product of the input with the weights across the edge radiating outward from them. The total amount of incoming signal on any given node in the architecture is the sum of the values coming from the edge's incident upon it. For 2D convolution neural networks this simple ends up being an input image in dot product multiplication with a learned operator/kernel/filter which convolute the input image. This is the feed forward step, where in based on a number of chained multiplications, followed by a non linearity activation function, a final output is produced at the output layer.

2.4.5 Back Propagation

One of the cornerstone of the neural network revolution was the back propagation algorithm developed in [82, 83] by David E. Ruineihart, Geoffrey E. Hinton and Ronald J. Williams. The practical application of this concept for Convolutional Neural Networks was later popularized by Yann LeCun. After the output of a network is obtained, it is measured against the true value, or the expected prediction given an input. The difference between the prediction of a neural network at the current step and the actual output known previously (ground truth) is the error term. Next, the idea of back propagation involves taking the total error at the output node and partially distributing it based on the contribution the incident nodes had on the output. This is a backward step for passing the error gradient starting from the last layer to the first learning/hidden layers. Essentially a set of partial derivatives are computed to calculate each incident node's contribution to the accumulated error of the current node under inspection and then based on learning rules and schemes the weights are updated or adjusted by a percent in the direction opposite to the error gradient. The original error which is now split between the immediate incident nodes is further back-propagated to their own incident layers and so on. The weights are adjusted for all learning layers in the neural network based on the learning rate assigned.

2.4.6 Other Layers and Techniques

Pooling Operator: Pooling layers simple take an input tensor and reduces the dimension by a factor specified by the pooling operator. For example, if an image or 2D tensor of resolution 512×512 pixels is passed through a pooling layer of 4×4 then the new image dimension is 128×128. Similar to Section 2.3.2 , the pooling operator simply moves across the input tensor and keeps the maximum value of the input tensor sub set or average the input tensor subset to generate a singular number. In this dissertation we use maximal pooling operator.

Basic Inverse Convolution Operator: These are replacement for traditional up sampling methods such as bicubic or nearest neighbor interpolation. For 2D neural network this can be a learned operator or a constant operator. The first step involves padding the original input 2D tensor with 0's forming a morphed tensor larger in dimension, such that when a weight or filter/operator/kernel passes over it, it will give the desired final dimension of the original input tensor. Next the filter/operator/kernel passes over the morphed tensor resulting in a final tensor larger in dimension to the original input tensor. The morphed tensor from the input tensor has to take into considerations the size of the filter/operator/kernel and add additional padding layers to ensure the final image generated is of the expected size.

DropOut: [93] introduces a technique where certain nodes in a neural network are selected at random to be turned off and in effect break certain input to output pathways. They do not participate in either the feed forward or the backpropagation steps. This technique helps to reduce overfitting and helps to regularize the network by allowing it to find multiple optimal or quasi optimal pathways in solving the desired problem.

Optimizers: These are the rules which govern how a node should adapt or change its weights with respect to its error gradient obtained via back propagation.

2.5 Neural Network for Image Segmentation

In this section I go over the sequence of landmark neural networks architectures/ideas, building up to the U-Net which is one of the most used architecture for medical image segmentation. Chapter 3 improves it by adopting it to satisfy the complementary constraint.

2.5.1 Auto Encoders

Auto encoders [35], are unsupervised neural network which takes in input and produces a compressed code as shown in Figure 2.8 which is then decoded using a decoder to recon-

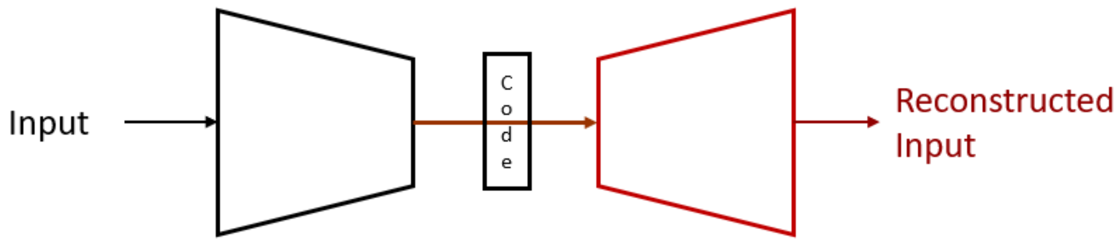


Figure 2.8: Auto encoder structure showing the encoder, generated code and decoder.

struct the input. The base autoencoder has been adopted to serves different applications. Regularizing autoencoders can be used to denoise [29], enforce sparsity [104] to name a few applications. Variational autoencoders [47] on the other hand is useful for image generation.

2.5.2 Fully Convolutional Networks (FCN)

A breakthrough architecture is the Fully Convolutional Network (FCN) [57], which made some changes to an usual auto encoder. Firstly, they introduced the concept of using a convolution layer with kernel/operator size of (1,1) instead of using a fully connected layer at the output node. Secondly, they added up sampling blocks originating at the code producing an output with the resolution of the input image. The authors released multiple versions of this framework with certain frameworks reproducing the input resolution from the code via two and three inverse convolution operations respectively to produce finer details. Another interesting feature of this framework was the skip connection of the corresponding down sample/encoder portion of the framework to the matching resolution of the decoder section. The output of FCN was the segmentation mask of the input. While the auto encoder is unsupervised, FCN is a fully supervised method.

2.5.3 U-Nets

A further refinement of FCN architecture has resulted in the birth of U-Nets [80] which has since then served as the baseline framework for medical image segmentation. Unlike an FCN, the U-Net is an autoencoder whose decoder branch parallels the encoder branch with skip connection between every encoder and decoder sub-blocks at the corresponding levels. This allowed for very fine segmentation masks to be produced. U-Net has found a permanent home in the field of medical image segmentation and serves as a staple benchmark framework. However, the U-Net is sensitive to unseen pathology in the region of interest and may not provide the best segmentation result under such circumstances. **My method in Chapter 3 overcomes these issues by using the idea of complementary learning**.

2.6 Neural Networks for Anomaly/Novelty Segmentation

Anomaly or Novelty segmentation involves first defining a notion of normalcy and then finding deviations from it on new incoming inputs. To appreciate the pioneering work in this area, AnoGAN [87], I first describe the workings of the Generative Adversarial Network or GANs[31] which serves as the core of AnoGAN.

2.6.1 GANs

A GANs framework consists of two neural networks, a generator, and a discriminator. The generator accepts a random input vector and outputs a signal or image. The discriminator is a classifier which is trained on a set of real images or signals and the output of the generator at every stage. The job of the discriminator is to classify the signal produced by the generator as false while the ones coming from the real image set as true.

The training cycle involves first training the discriminator followed by freezing the weights of the discriminator. Next the generator is trained with the aim to produce a signal or image such that when it passes through the discriminator's frozen architecture, produces an output of true value. This causes the generator to learn a mapping between the input vector and the set of real images which were reserved for the discriminator to predict as true.

The termination criterion is said to have reached when the two adversarial networks, generator and discriminator reaches Nash's Equilibrium[69]. However, recent literature [27] suggests that GANs may not satisfy the Nash's equilibrium. The article is yet to be peer reviewed and this discussion is outside the scope of this dissertation. However, the difficulty in reaching the Nash's equilibrium for the generator and discriminator has caused a lot of issues in the utility of GANs and the pioneering method in the field of Anomaly/Novelty segmentation also suffers from the instability issue of GANs training.

2.6.2 AnoGAN and f-AnoGAN

The pioneering work of the anomaly and novelty segmentation can be attributed to [87] and further improved in [86]. The core idea of these frameworks is to learn a manifold of healthy or normal images by first training an GANs generator to only generate normal healthy images as shown in Figure 2.9. The discriminator has a healthy image set to enforce this. Following the training, if an anomalous or unhealthy sample is presented as a query image, AnoGAN does an iterative search to find the nearest healthy sample to the query image and then takes a subtraction or residual between the two to find the anomaly as shown in Figure 2.10. Figure 2.11 shows how the anomaly is highlighted. In f-AnoGAN, the authors connect a second encoder which can map a real image to the input vector of GANs generator in order to bypass the incremental query searching of the original work in AnoGAN. This encoder termed *izi* is shown in Figure 2.12. The primary demerit of this method is the high dependency on the quality of generation via the generator. Due to GANs instability issue this may not be possible even for normal images as shown in Figure 5.5. **My method in Chapter 5 circumvents this issue entirely and provides a novel way to conduct**

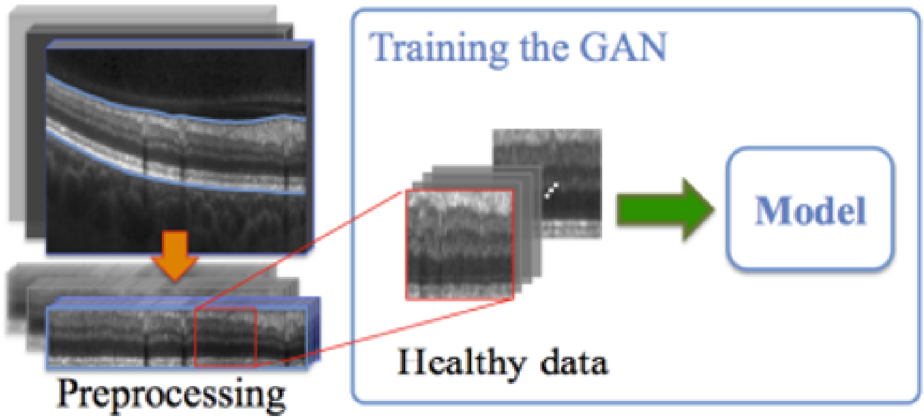


Figure 2.9: Training Scheme on only healthy samples for AnoGAN. Image source: [87]

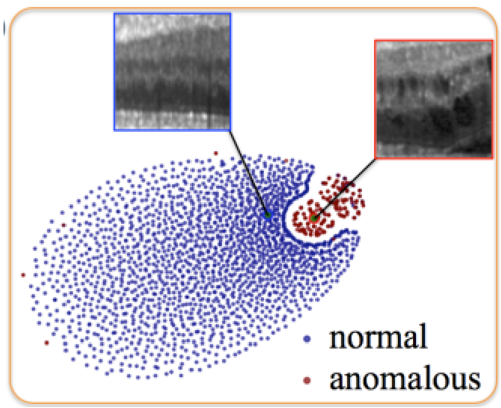


Figure 2.10: Iterative search for finding the closes image in manifold of healthy samples to the query image. Image source: [87]

anomaly/novelty segmentation.

2.7 Motivation for Complementary Learning

In this section, I provide some more insight as to why complementary constraint is instrumental and nonrestrictive unlike current training regimens. If we consider the Figure 2.13 top left,

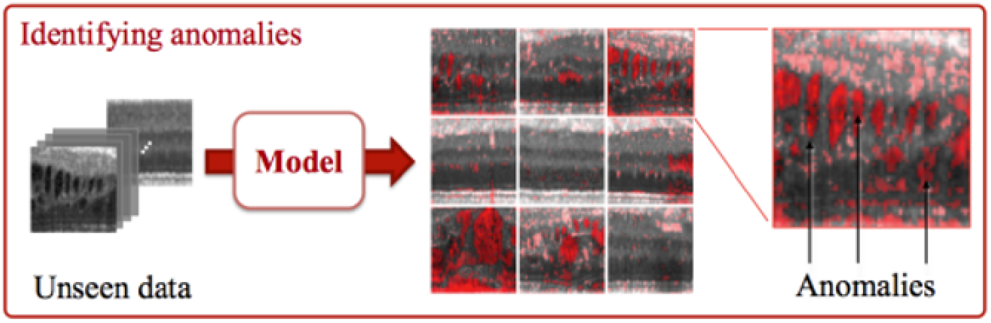


Figure 2.11: Highlighting anomalous regions. Image source [87]

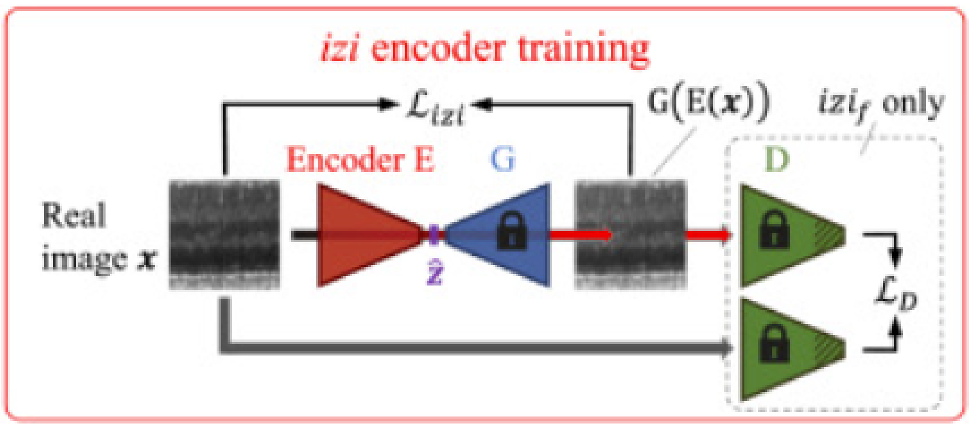


Figure 2.12: The *izi* encoder in red connects an input image to the GANs random vector z. Image source [86]

the task there is to segment out the brain region. A neural network like U-Net trained to segment out the brain region will focus on learning features relevant to characterize just the brain region and in case there is a large perturbation or distortion during testing, such as the image on top right in Figure 2.13, which was not present during training is presented, it would fail to segment the brain region afflicted with the distortion. The first column second third and fourth row in Figure 3.8 demonstrates that. On the contrary the ability to utilize the non brain features such as demonstrated in the second row of Figure 2.13 will be able to handle such distortions during testing. Essentially, the aim of complementary learning is to learn both ROI and non-ROI features in an image space. By knowing the non ROI features we can effectively deduce the ROI regions as well.

The task of defining the non ROI region is non-trivial, and the idea of complementary learning allows us to ascertain that without further annotation or additional work from the annotators. In Section 1.2, I go over the disjoincy loss required for implementing the complementary constraint.

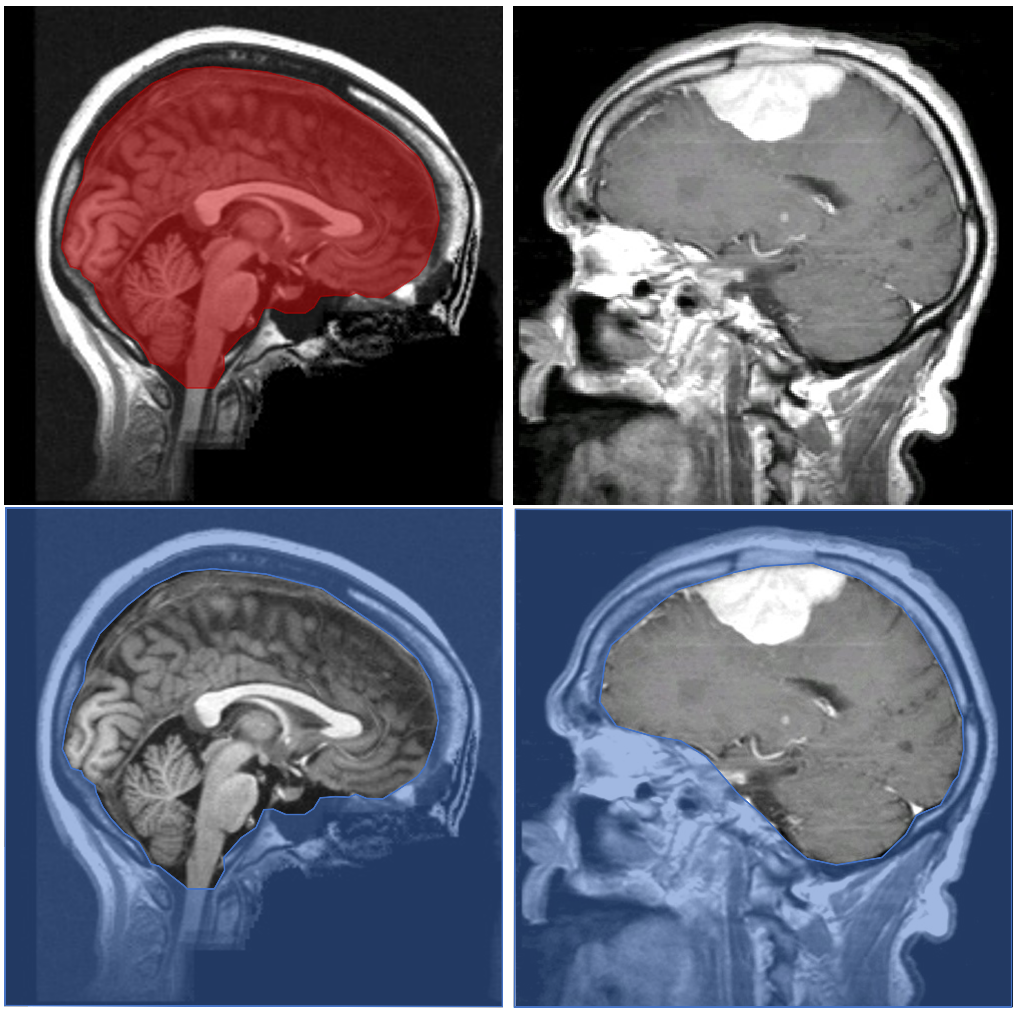


Figure 2.13: Two ways of inferring the same information for a given task. Here the aim is to extract the brain region. It can be done by directly learning the features of the brain region as shown in Row 1. Similarly, the complete knowledge of the non brain regions in Row 2 also indirectly infer the same brain region.

2.8 Medical Imaging

2.8.1 MRI

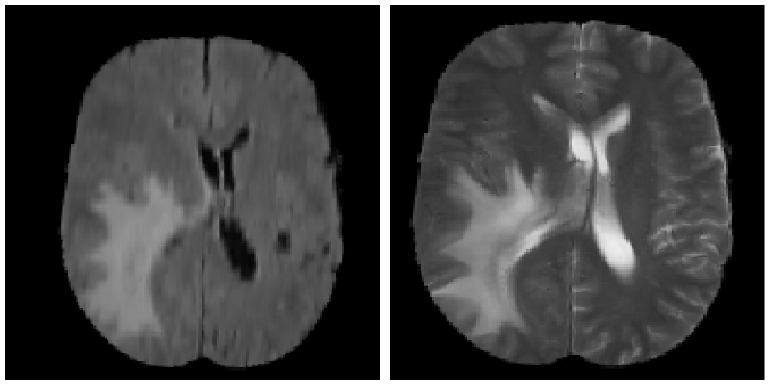


Figure 2.14: T2 Weighted Flair left and T2 right. The CSF is dark in the Flair image vs the T2 image where it is bright.

Magnetic resonance imaging (MRI) is a common test used in neurology and neurosurgery for diagnostic as well as intervention radiology. It is used for detecting strokes, lesions, bleeding, and other ailments that may affect the normal working of the nervous system. A powerful uniform external magnetic field is used to align the randomly oriented water nuclei of the tissue under inspection. In the absence of such a field they are moving in random motion. The magnetic field helps align the water molecule for further analysis. However, even then some water molecules don't move at the same rate as the rest under the influence of the magnetic field and are called low energy water molecules. Next, a strong external radio wave with same frequency as the magnetic field is introduced. This energy is absorbed by the low energy water molecule to move at the same phase as the other water molecules influenced by the magnetic field. After the radio wave source is turned off the low energy water molecules release the additional energy and go back to their original orientation by emitting energy. This energy is then captured by a computer to create the MRI scans. The different slices in an MRI scan are the images of the tissue under inspection at different sections of the magnetic field. The different modalities of MRI used for this dissertation are T1, T2 and T2 weighted Flair (referred here onward as Flair) . They are different measures of the relaxation time of different tissues under inspection. In T1 the CSF is dark and the white matter is light, cortex is gray, fat is bright, and inflammation is dark. In T2 CSF is bright, white matter is dark gray, cortex is light gray, fat is light and inflammation is bright. Flair is similar to T2 however the CSF is dark which sometimes helps in figuring out the boundaries between an inflammation and the CSF region [1]. Figure 2.15 illustrates that point. In the case of t2 the bright CSF could confuse the unsupervised ASC-Net in Chapter 5 to predict it as part of the tumor.

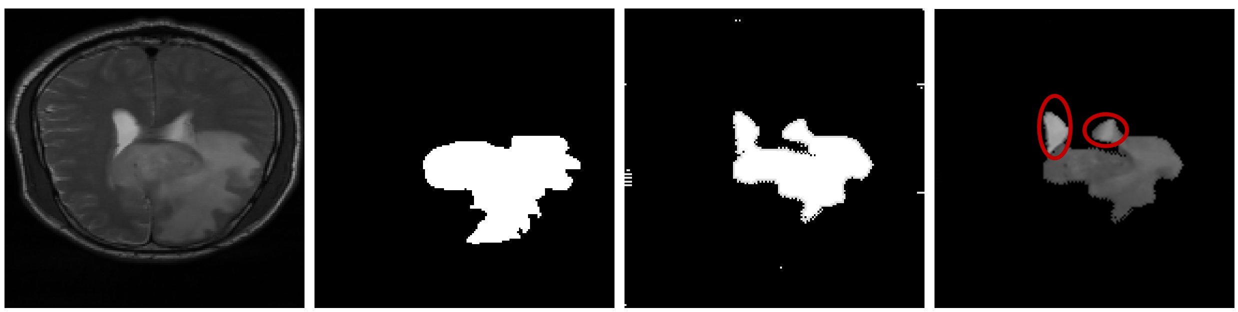


Figure 2.15: A sample from the private dataset experiment with ASC-Net in Chapter 5; Left to right: Input slice, ground truth, the prediction, superimposition of prediction on input image showing the CSF region in red ellipses.

2.8.2 CT

The basic principle of a CT scan involves a revolving X-ray source, which goes through a patient on to a detector which then measures the differences between the x-rays absorbed by the body and the x-rays transmitted through the body. This determines the density of the different tissues. The unit of measurement of the density in a CT scan is defined by Hounsfield Unit where every kind of tissue has its own Hounsfield value ranges.

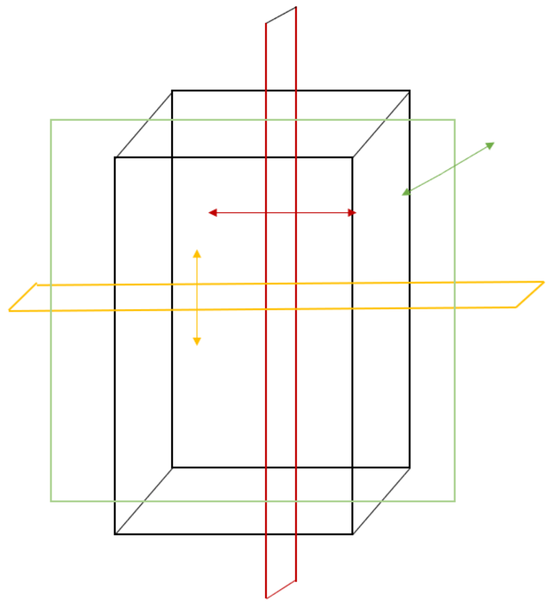


Figure 2.16: A simple representation of the Sagittal plane in red, Coronal in green and Axial in yellow.

2.8.3 Anatomical Planes of Observation

The human body can be analyzed in three anatomical planes which are perpendicular to each other. Figure 2.16 demonstrates the three planes in 3D and Figure 3.1 shows an example of the three planes on a brain MRI scan about a point.

Sagittal Plane

This divides the body on a longitudinal axis into left and right parts. One can imagine this as a plane cutting through the face and out the back and moving left to right or right to left to collect all the 2D slices to form the 3D image.

Coronal Plane

This divides the body into ventral/belly and dorsal/back sections. One can imagine this as a plane cutting through one side and out the other and moving front to back or back to front to collect all the 2D slices to form the 3D image.

Axial Plane

This is the plane that divides the body into superior and inferior part. One can imagine this as a plane cutting through the side as well but perpendicular to the coronal and sagittal plane and going up to down or down to up to get all the 2D slices to construct the 3D image.

Chapter 3

Complementary Segmentation Network for Skull Stripping

3.1 Disclaimer

This chapter borrows from the research conducted by me under the supervision of my doctoral advisor Dr Yi Hong and published under the following titles -

(i) CompNet: Complementary segmentation network for brain MRI extraction [21]

(ii) Diagnostic classification of Lung nodules using 3D neural networks [24]

3.2 Introduction

Skull stripping is the first preprocessing step that is undertaken after a Brain MRI scan is conducted. Figure 3.1 demonstrates the three different planes of an MRI image. The purpose of skull stripping involves the removal of all artifacts present in the MRI except the Cerebrum, Cerebellum, Brain Stem, and the Spinal Cord to conduct diagnostic analysis of the brain, eg, measuring normal brain development and degeneration, uncovering brain disorders such as Alzheimer's disease, or diagnosing brain tumors or lesions.

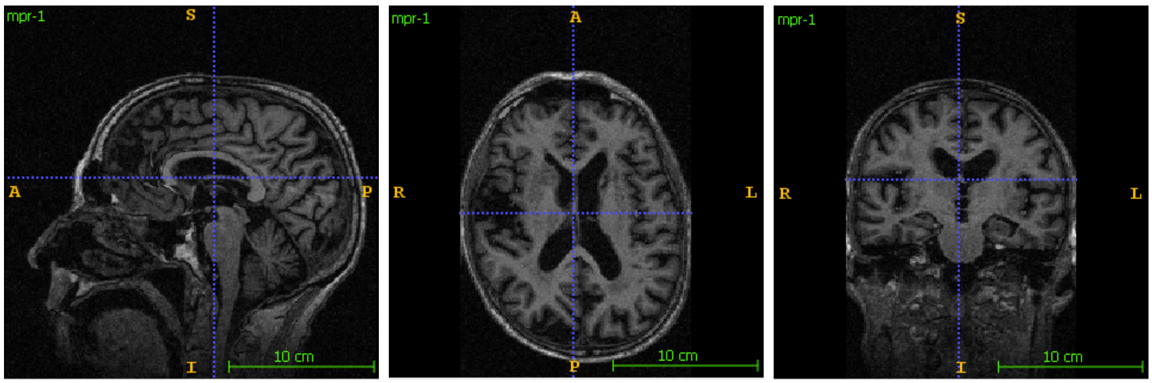


Figure 3.1: Left to Right: Sagittal, axial and coronal planes of a Brain MRI about a point from the Oasis dataset captured using ITK Snap [101]

The current state of data acquisition and labelling for medical imaging data is sparse [51], and lack diversity [98, 75] thus making the generalization of AI based approaches solutions limited. There is a stark disparity in the availability of pathological dataset with non brain artifacts intact. For example public datasets such as BraTS [4, 5, 63] for brain tumor segmentation, MS-Seg2015 [13] for MS lesion segmentation, provides skull stripped imaging studies. In literature, the approaches developed for brain MRI extraction can be divided into two categories: traditional methods (manual, intensity or shape based model , hybrid, and PCA-based methods [34, 92]) and deep learning methods [48, 84]. A majority of the artificial intelligence methods developed for the purpose of skull stripping focuses on learning image features for the brain tissues using a training dataset typically consisting of a collection of normal (or apparently normal) brain MRI. Thus, the model performance is sensitive to unseen pathological tissues.

In the following sections, I introduce a neural network framework based on the complementary constraint in Chapter 1.2 which performs skull stripping from brain MRI, improving the performance of existing methods on brain extraction and more importantly is invariant to brain pathology despite being trained on publicly available regular brain scans. I achieve the state of the art result on two fold cross validation on the publicly available Oasis [60] dataset and also demonstrate the framework's invariance to distortions by introducing synthetic pathologies during testing. Lastly, I conduct an ablation study starting with a base network and add multiple augments such as densely connected blocks [37], complementary constrain inspired augments and another novel multi-output strategy.

3.3 Complementary Constraint via Deep Learning Architecture

For the purpose of skull stripping the current state of the art methods are supervised and the pipeline is composed of an input image which passes through a model and whose output is the binary ground truth mask or label for the brain regions. The value of one in the output image of these models symbolize the presence of a pixel belonging to either cerebrum, cerebellum, brain stem or spinal cord regions in the input image. An encoder-decoder network, like U-Net [80] is often used for the purpose of segmentation. However, a method of this sort tends to primarily focuses on features to identify the ROI and may have difficulty in generalizing to unseen image distortions in the ROI.

Section 1.2 explores the effect of reversing the objective of the dice coefficient metric. It provides two sets/images which do not share any common regions. In the same vein, I propose a novel framework which provides two sub image of an input image that are disjoint. One of the sub image is the Segmentation Output (SO) and is guided by the ground truth similar to other supervised frameworks as discussed above. The complementary constraint now allows us to define a second image, Complementary Output (CO), which is the region complement of the SO. While the SO is bounded by the ground truth loss and operates on a supervised setting which is an essential part to getting a good segmentation mask, the CO is unbounded

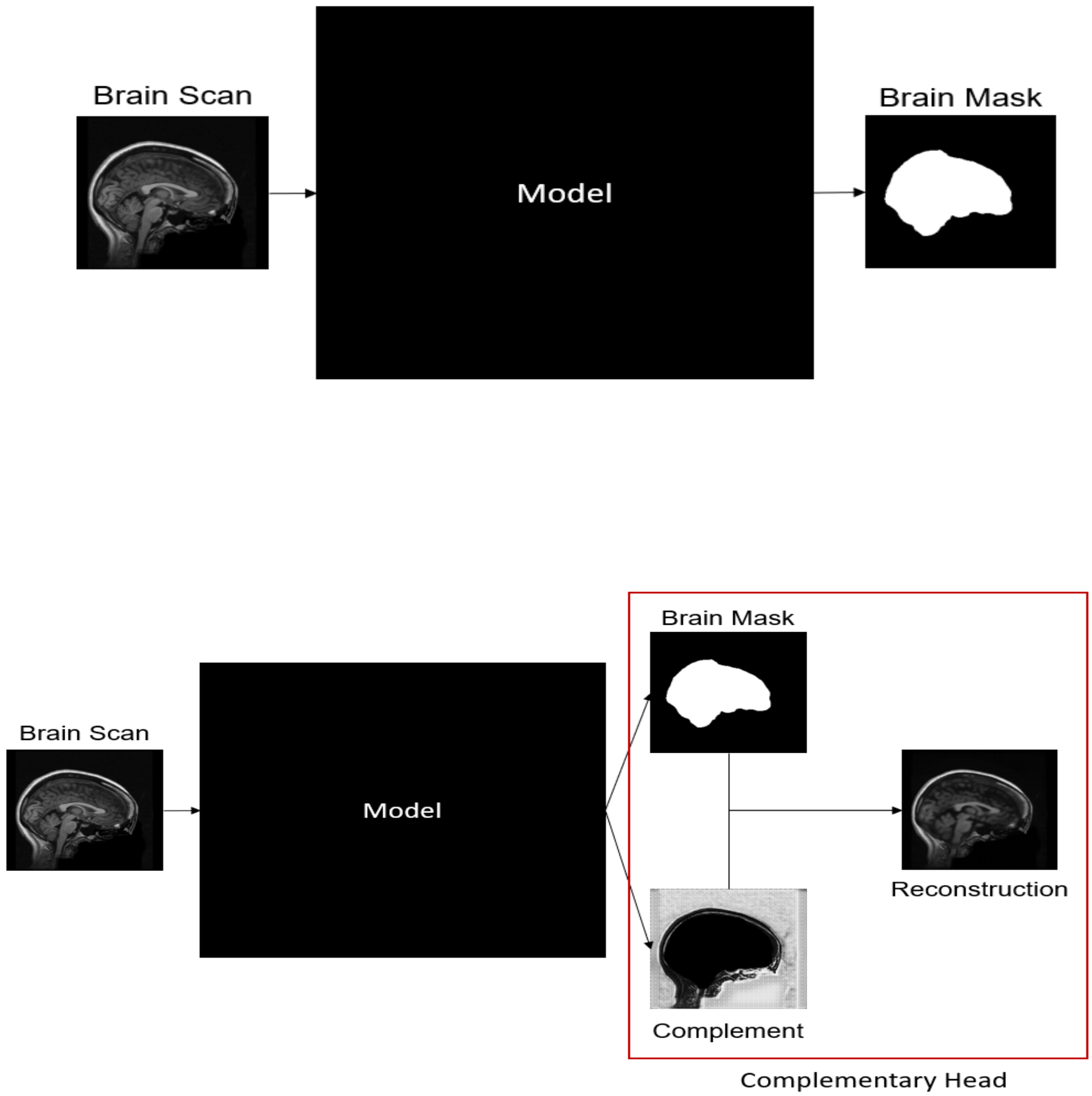


Figure 3.2: Top: Standard pathway for a segmentation problem; Bottom: Pathway including the complementary section for the same segmentation problem. The inclusion of the complementary section to the pipeline can be thought of a Complementary Head, which is composed of a segmentation, a complement of the segmentation and a reconstruction of the input.

by any guidance aside from the complementary constraint. **The trivial solution of the complementary constraint is the null set and since the CO is unbounded it can always produce a null set i.e. it can produce empty images**. Since the trivial solution provides no benefit in solving the task at hand, and effectively reduces the framework to become similar to a standard U-Net type architecture, I introduce a second constraint which requires the two sub images to give us back the original input image, i.e., Reconstruction Output (RO). This requires CO to account for all parts of the input image which do not include the areas covered by the supervised SO. SO, CO, and RO taken together form the Complementary Head and can be augmented to any existing segmentation framework. Figure 3.2 demonstrates how a network augmented with the complementary head differs from the standard segmentation network. The SO and CO satisfy the complementary constraint and the RO further prevents the CO from producing a trivial empty image as output.

Definition 3 (ROI Interactive Pathology)**.** Any distortions in the input image space, absent in the training image space but present in the testing image space and has a negative impact on the quality of segmentation of the ROI.

Central Idea:*A framework designed to satisfy the complementary constraint operates to find feature for both the ROI and its complement, thus being able to segment the ROI even in the presence of ROI interactive pathologies which were absent in the training set.* This further solves an additional problem of having to train a network to segment the ROI in the presence of specific pathologies since that problem is intractable as there may be uncountable pathologies.

3.3.1 Baseline U-Net

The backbone of the Comp-Net is the U-Net. The U-Net base I use is illustrated in Figure 3.3. Each gray block in Figure 3.3 consists of two convolutional blocks with a kernel size

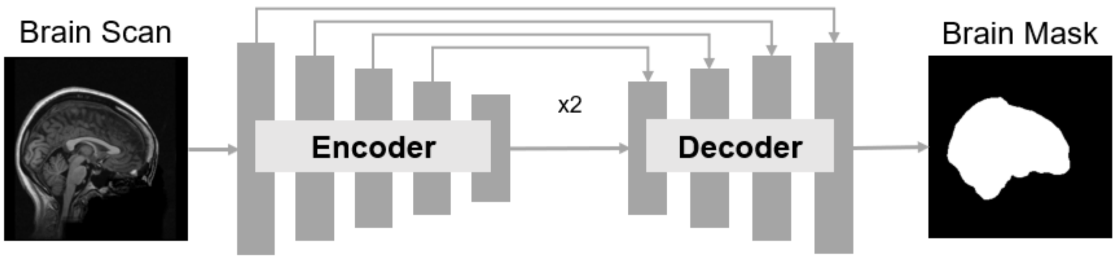


Figure 3.3: The baseline U-Net used in the experiments.

of 3×3. The number of convolutional filters in the encoder starts from 32, followed by 64, 128, 256, and 512, while the number in the decoder starts from 256, followed by 128, 64, and 32. Each convolutional layer is followed by batch normalization [39] and dropout [93]. After each gray bar in the encoder, the feature maps are down sampled by 2 using max pooling, while for the decoder the feature maps are up sampled by 2 using deconvolutional layers. The output of this network is defined by a convolution block with feature map of size one and filter size of 1×1. The activation function of all convolution layers except the last one is relu. The final output layer has an activation of sigmoid. The loss function used is the Dice coefficient $Dice(A, B) = 2|A \cap B|/(|A| + |B|)$ [25] between the output of the U-Net and the ground truth label/mask. For implementation purposes since the neural network is a minimizer, we compute the negative of the Dice loss.

3.3.2 Basic CompNet

The CompNet, depicted in Figure 3.4, builds on top of the U-Net described in Section 3.3.1. To augment the U-Net using the complementary head, a second decoder having a similar architecture is attached in parallel to the existing decoder. The output of the first decoder

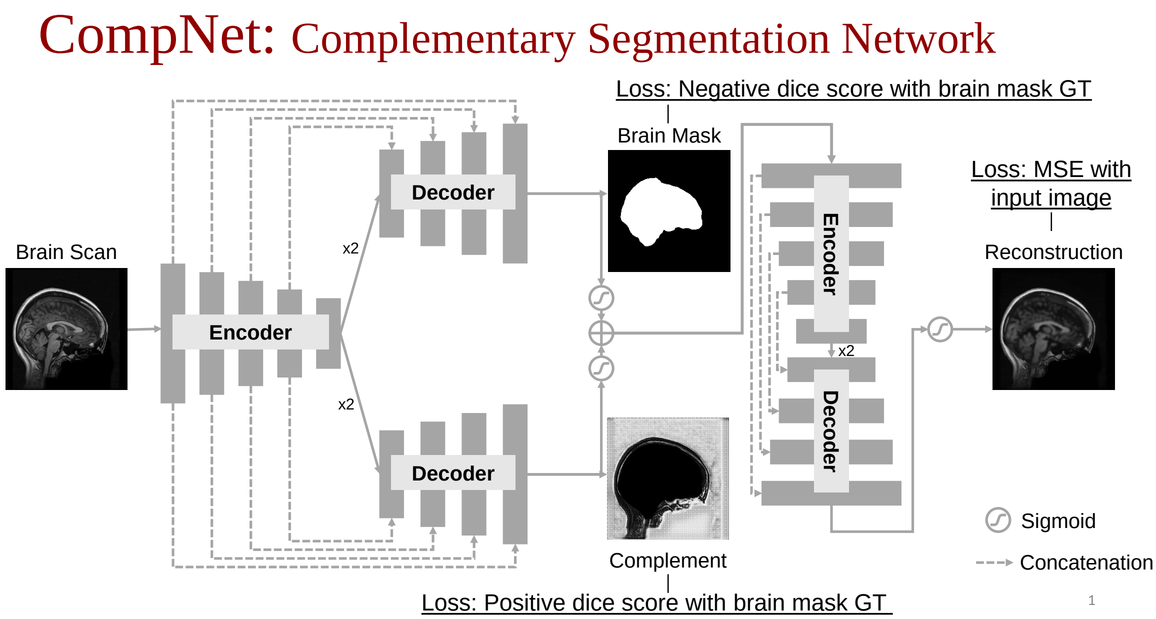


Figure 3.4: The basic CompNet design consisting of two decoder branches in parallel to obtain the segmentation and its complement, followed by a second U-Net to obtain the reconstruction.

branch is the Segmentation Output (SO) and the loss function here is the dice coefficient loss. The output to the second decoder branch is the Complementary Output (CO) and follows the complementary constraint which is implemented via the disjoincy loss i.e., the negative of the dice loss. For a neural network which is a minimizer, this ends up being the positive dice. Thus, the **Disjoincy Loss** defined in Section 1.2 is effectively the positive dice loss of the SO and the CO. The optimal value for this is 0. To prevent the realization of the trivial solution (i.e., null set or empty image), I also connect a second Encoder-Decoder structure whose input is the concatenation of the SO and CO and the output is the reconstruction of input image itself enforced with a MSE loss with the input image. **The overall structure of the Comp-Net is trained end to end and consists of a segmentation branch, a complementary branch and a reconstruction branch**. The input to the network is an Input image, and the outputs are the segmentation mask, the complement, and the reconstruction.

3.4 Augmentations for Improved Gradient Back Propagation

I use two schemes to improve the performance of the framework. The first of the two is the well-studied DenseNet [37]. The second is a novel Multi Output concept. Networks containing both two augmentations are termed Optimal U-Net and Optimal Comp-Net.

3.4.1 Dense Blocks

One of the issues of very deep networks is the vanishing gradient problem and the other one is the increase in the number of parameters to optimize as the network gets deeper. In order to combat the issue, I utilize the idea of DenseNets [37]. Each gray block of the plain U-Net in Section 3.3.1 is replaced with a dense block, illustrated in Figure 3.5. Each dense

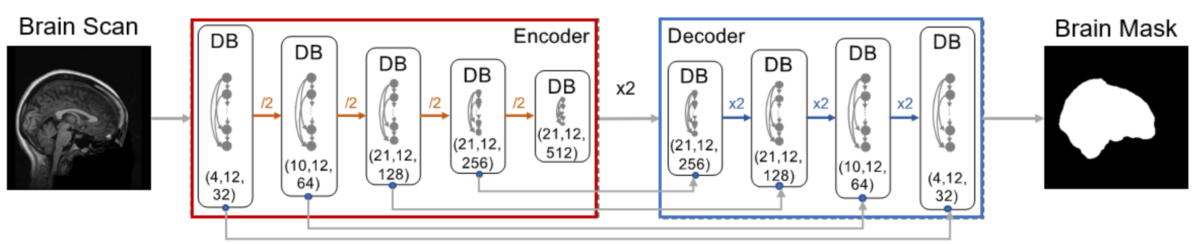


Figure 3.5: Dense U-Net: The baseline U-Net modified with dense blocks.

block has different number of convolutional layers and filters. Specifically, the dense block in each encoder has 4, 10, 21, 21 and 21 convolutional layers, respectively and the ones in each decoder have 21, 21, 10, and 4 layers, respectively. All of them use the 3×3 kernel size.

The convolutional layers have a base feature map size of 12 i.e., each of the non-terminal convolution layers have 12 feature maps. The terminal convolutional layers of the encoder blocks have feature maps of 32, 64, 128, 256, 512 respectively while the terminal decoder layers have feature maps of size 256, 128, 64, 32 respectively. This design aims to increase the amount of information transferred between one dense block and the next. The internal convolutional layers within the block are all connected to each other and thus do not require this increase in the feature maps.

3.4.2 Multi Outputs

Another way to improve the vanishing gradient problem, is to introduce early output blocks. Figure 3.6 demonstrates the intermediate outputs. The early gradient from the intermediate output provides a strong back propagation directly to each block. This can mitigate the vanishing gradient problem by shorting the distance from the input to the output [24]. The

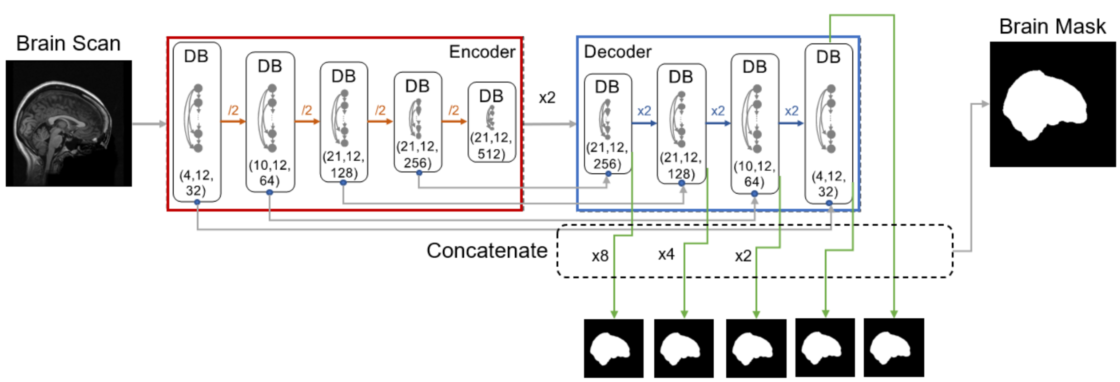


Figure 3.6: Optimal U-Net: Dense U-Nets further modified to include intermediate outputs shown in green.

decoder of the dense U-Net now has intermediate outputs which are generated from the first convolutional layer in each dense block of the decoder. This is important since the first convolution layer is projected from the concatenation of the previous dense block's tensor inverse convoluted, and the tensor coming from the same level of the encoder and thus has a learning component based on both the decoder and encoder input as the same level. The only exception is the final dense block in the decoder which also has an output from the terminal layer of the block since there are no more blocks after it. To obtain the intermediate outputs, I first apply inverse convolution using same number of feature maps as the convolution layer its connected to and a scale factor which is the ratio of the input dimension and the current dimension. This projects their incident input tensor to the dimensions of the input image/output mask, followed by a convolution via moving window size of 1×1 and a single feature map. Finally, the $n + 1$ intermediate output's penultimate layers, where n is the total number of dense blocks in the decoder, are concatenated together to obtain the final output providing an implicit ensemble learning. Since a part of the gradient now originate at every dense block, it provides feedback as to the performance excluding the blocks downstream. This also helps to diagnose the depth of the network since if a shallower dense block attains a lower optimum than the deeper dense block, then the successive dense blocks should be discarded and is detrimental or redundant. The U-Net obtained this way is called the Optimal U-Net.

3.5 Optimal CompNet

The optimal CompNet (Figure 3.7) is obtained via augmenting the CompNet with dense blocks and multi outputs. Similar to the Dense U-Net in Section 3.4.1, each of the blocks in the encoder decoder are converted to dense blocks following the same scheme. Furthermore, each dense block also boasts intermediate outputs. To satisfy the complementary constraints, the pairwise decoder intermediate outputs and the final outputs satisfy pairwise disjoincy loss at their respective levels. Each decoder's penultimate layer of the dense blocks i.e., intermediate SO's penultimate layers, and intermediate CO's penultimate layers are also concatenated

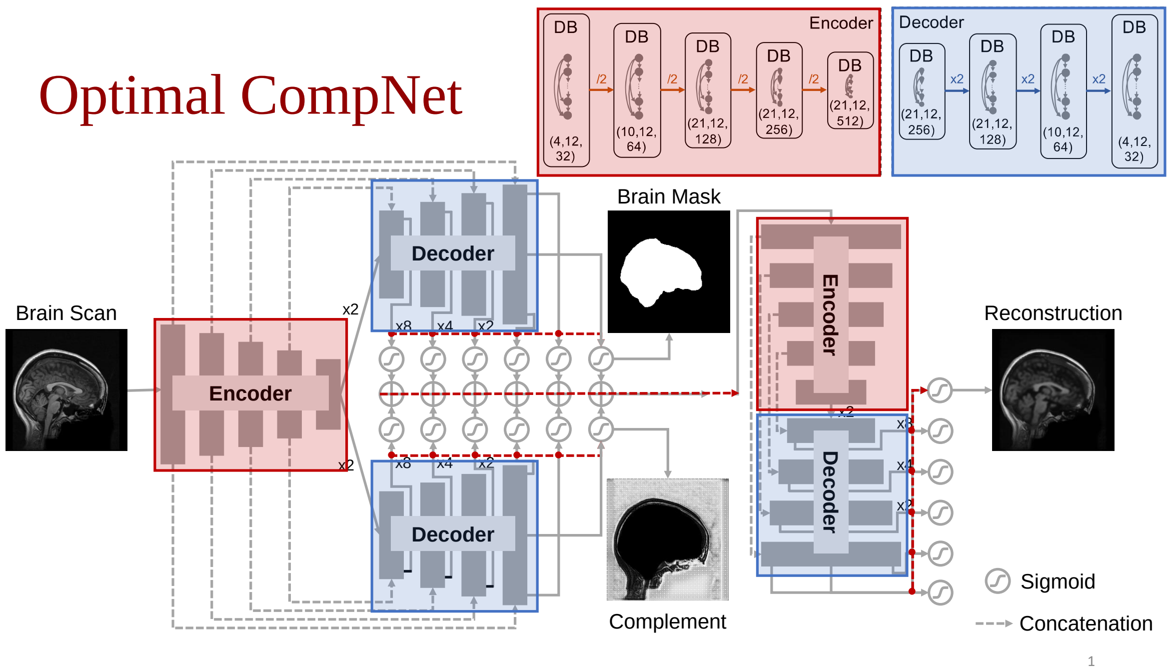


Figure 3.7: Optimal CompNet: The modified CompNet composed of dense blocks as well as intermediate outputs similar to the Optimal U-Net. The Optimal CompNet maintains the basic CompNet's idea of implementing the complementary head composed of a segmentation, complement of the segmentation and the reconstruction of the input.

together to generate the final SO and CO outputs. At the same time a concatenation of the pairwise concatenation of the intermediate and final outputs respectively of the two decoders is fed forward to the reconstruction block. The reconstruction block also follows the same dense multi output architecture of the U-Net in Section 3.4.2.

3.6 Experiments

I conduct an ablation study involving the Baseline U-Net in Section 3.3.1, Basic CompNet in Section 3.3.2, Optimal U-Net in Chapter 3.4 and Optimal Comp-Net in Chapter 3.5.

3.6.1 Datasets

Standard Oasis Dataset

I evaluate CompNets on the OASIS dataset [60], which consists of a collection of T1-weighted brain MRI scans of 416 subjects aged 18 to 96 and 100 of them clinically diagnosed mild to moderate Alzheimer's disease. I use a subset with 406 subjects that have both brain images and masks available, with image dimension of $256 \times 256 \times 256$. These subjects are randomly shuffled and equally divided into two chunks for training and testing with two-fold cross-validation, like [48] for comparison on (apparently) normal brain images.

Synthetic Pathology Induced Oasis Testset

To further evaluate the robustness of the networks, I simulate brain pathologies in one half of the OASIS subset. The simulated pathologies include 3D brain tumors and lesions with different intensity distributions and sizes placed at different locations as well as damaged skull and non-meningeal tissue membranes (Figure 3.8, column 1). The different networks are trained on the other half of unchanged images and tested on the noisy images. The synthetic tumors/distortions created had the following control parameters:

(i) A weighted random bit determining if the distortion should be inside brain region or outside. The odds of distortion inside brain region were set to 90 percent.

(ii) A random number between 20 and 80 which determined the radius of the sphere.

(iii) Three random numbers x, y, z symbolizing the location of the center of the tumor/sphere.

(iv) A random bit determining if the sphere would accentuate or mask the underlying tumor intensity.

(v) A random number between 40 and 100 indicating the value to accentuate/mask the under lying tissue intensity in the brain region.

(vi) A random number between 40 and 200 indicating the value to accentuate/mask the under lying tissue intensity in the non-brain region. A higher maximum value was chosen for the outside since I wanted the skull to vanish in some cases.

3.6.2 Experimental Settings

I use a dropout [93] rate of 0.3, as well as use L_2 regularizer to penalize network weights with large magnitudes and its control hyperparameter λ set to 2e-4. For training, I use the Adam optimizer [44] with a learning rate 1e-3. All networks run up to 10 epochs.

	(Apparently) Normal Images			Pathological Images		
	Dice	Sensitivity	Specificity	Dice	Sensitivity	Specificity
Kleesiek et al.[48]*	95.77±0.01	94.25±0.03	99.36±0.003	–	–	–
Plain U-Net	92.30±6.20	95.60±1.48	96.20±0.09	79.90±8.10	93.80±5.10	95.20±2.15
Optimal U-Net	96.40±4.10	97.50±0.70	96.90±0.01	85.43±5.80	96.13±3.20	97.10±1.27
Plain CompNet	96.70±0.22	97.93±0.62	98.57±0.06	95.21±3.75	96.32±1.03	99.21±0.10
Opti. CompNet	98.27±0.30	98.26±0.58	99.80±0.05	97.62±2.21	97.84±0.80	99.76±0.12

Table 3.1: Quantitative comparison (mean and standard deviation in percentage) among different networks on (apparently) normal and pathological images. *This paper is not directly comparable, because it was evaluated on mixed data samples, including 77 images (57%) from OASIS data set. (Opti.: Optimal)

3.6.3 Results

I compare the CompNets with a 3D deep network proposed in [48], a plain U-Net (the backbone of the plain CompNet), and an optimal U-Net (the backbone of the optimal CompNet). These networks are tested on (apparently) normal images (with two-fold cross-validation) and on pathological images (trained on the other fold with clean images). Given a 3D brain MRI scan of a subject, the networks accept 2D slices and predict brain masks slice by slice, which are stacked back to a 3D mask without any post-processing.

Figure 3.8 demonstrates the qualitative comparison among predicted brain masks. For (apparently) normal brain scans, all networks produce visually acceptable brain masks.

However, the plain and dense U-Nets have difficulties in handling images with pathologies, especially the pathological tissue on or near the boundary of the brain. Part of the pathological tissue in the brain is considered as non-brain tissue. The plain U-Net even over segments part of the skull as the brain when the skull intensity changes, as shown in the case 3. In contrast, the CompNets can correctly recognize the brain, and the optimal CompNet presents the best visual results for all four cases. I then use Dice score, sensitivity, and specificity to quantify the segmentation performance of each network, as reported in Table 3.1. The optimal CompNet consistently performs the best among all networks for either normal (averaged Dice of 98.27%) or pathological (averaged Dice 97.62%) images, although its performance on images with pathologies is slightly downgraded by $< 0.7\%$ on average and $< 2.6\%$ in the worst case.

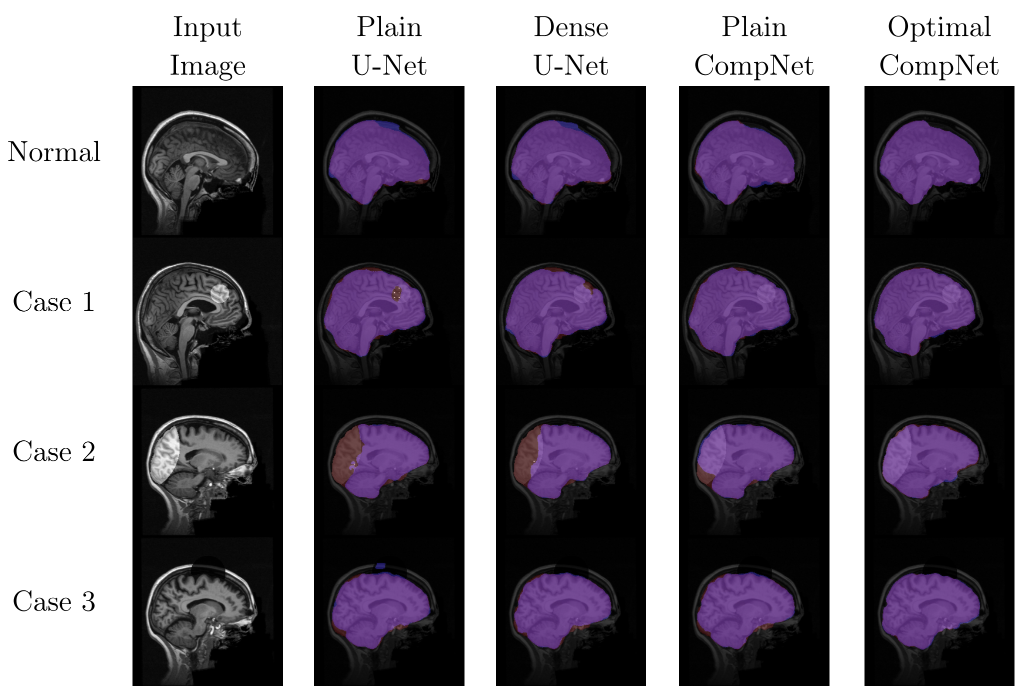


Figure 3.8: Qualitative comparison among all networks, plain and optimal U-Nets, plain, and optimal CompNets, on four image samples: a normal one, one with pathology inside of the brain (case 1), one with pathology on the boundary of the brain (case 2), and one with a damaged skull (case 3). The true (red) and predicted (blue) masks are superimposed over the original images. The purple color indicates a perfect overlap between the ground truth and the prediction.

3.7 Conclusion

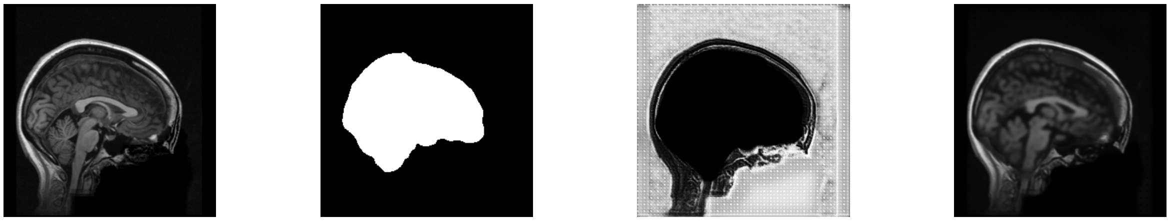
Complementary learning provides robust segmentation.

Table 3.1 demonstrates the advantage of using complementary learning by providing a failsafe in the event the region of interest has features completely different from the training set. The rest of the image can provide cues to aid in segmentation of the region of interest. Essentially a supervised learning method is biased and focuses primarily on features relevant to the region of interest i.e. it is a *"Look Here"* and *"Do This"* kind of framework. Using complementary learning frees the complementary branch to focus on any features required to aid in the segmentation, and at the same time it is tasked with accounting for features which would enable reconstruction of the input image. These two taken together learns features of the entire image space. While the overall pipeline is supervised, the complementary branch has no specific task. It only has constraints which prevents it from predicting items in the area of the segmentation branch. This essentially converts the learning to also say *"Dont Look Here"* and *"Dont Do This"*. The two statements above have a different implication, in the first case the AI must learn to do the task as directed. However, the complementary design allows the AI the freedom to do anything it needs to accomplish the task at hand and the only restriction provided by the complementary constraint and reconstruction branch to prevent the trivial solution. In essence this framework combines the advantages of supervised learning via the segmentation branch which is task focused as well as retains the flexibility of building a general unsupervised code of the input image via the complementary branch.

Complementary learning introduces stability.

The second interesting observation is the standard deviation obtained via adopting a complementary learning strategy. Unlike a standard U-Net, the CompNets prediction are more stable.

Components of the complementary supervised learning framework CompNet.



(a) Input (b) Brain mask (c) Complement (d) Reconstruction

Figure 3.9: Three outputs (b-d) of the optimal CompNet for an input brain scan (a).

In Figure 3.9 we see the three outputs, Segmentation, Complementary and Reconstruction. As previously claimed, this framework can take advantage of the supervised learning structure and obtain a very fine mask of the brain. At the same time the complementary branch is free to produce anything which aids in the task constrained to not "step" inside the segmentation region. The reconstruction constraint further prevents the realization of the trivial "empty image" case of the complementary branch in the event the input image is nonzero.

Chapter 4

Hierarchical application of CompNet for Small Liver Lesion segmentation

4.1 Disclaimer

This chapter borrows from the research conducted by me under the supervision of my doctoral advisor Dr Yi Hong and published under the following titles -

(i) HYBRID CASCADED NEURAL NETWORK FOR LIVER LESION SEGMENTATION [22]

4.2 Introduction

Liver lesion segmentation presents a unique challenge due to the heterogeneous and diffusive appearance of hepatic lesions. Researchers have proposed many segmentation algorithms based on the classical segmentation techniques, e.g., thresholding [65], region growing [66], active contour [61], and ensemble [38]. Recently, deep neural networks have been widely used in the liver lesion segmentation and have shown improved performance in segmenting and detecting

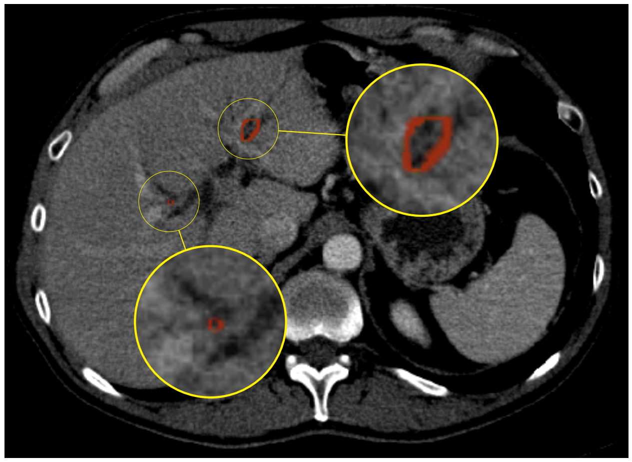


Figure 4.1: False positive examples of small liver lesions on a 2D slice using a 2D network.

liver lesions. These algorithms either use 2D convolutional neural networks [18, 33, 54], 3D networks [40], or a combination of both [19, 56]. One of the issues faced by the existing methods is the segmentation of the small lesions, which are often missed in their segmentation predictions.

The other issue is false positives as shown in Figure 4.1. Since hepatic tissue often time appear like lesions, networks which are sensitive to small tumors may label those tissues as lesions. A 3D network solves the entire problem however it is expensive. I leverage the observation that a tumor location remains relatively constant across slices whereas liver tissues appear to move around as shown in Figure 4.2 to create a hybrid framework, in order to maximize speed and accuracy, which uses 2D Networks for large lesion segmentation and 3D for smaller ones. In the following sections I demonstrate the effectiveness of our complementary based segmentation framework in a hybrid setting, obtaining a state of the art result without post processing on the LiTS [11] Liver Lesion challenge test set.

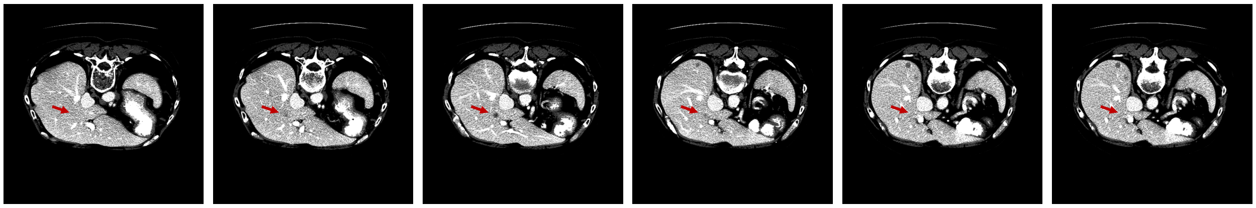


Figure 4.2: The different liver artifact moves around, however the tumor location marked with a red arrow remains constant.

4.3 Hybrid Complementary Solution to the Liver Lesion Segmentation Problem

The overall pipeline as shown in Figure 4.3 can be broken down into three subsections; liver segmentation, large lesion segmentation and small lesion segmentation.

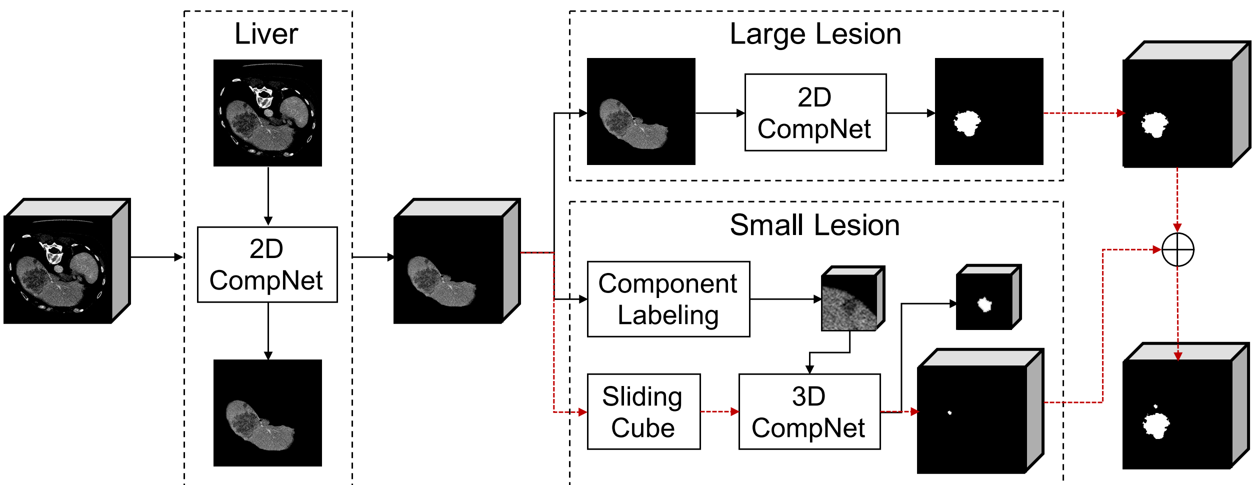


Figure 4.3: Overview of the proposed hybrid cascaded network for the liver lesion segmentation. 2D CompNets is adopted to segment the liver region and large lesions while a 3D CompNet is used to segment the small lesions. The red dashed lines indicate the test phase of the liver lesion segmentation.

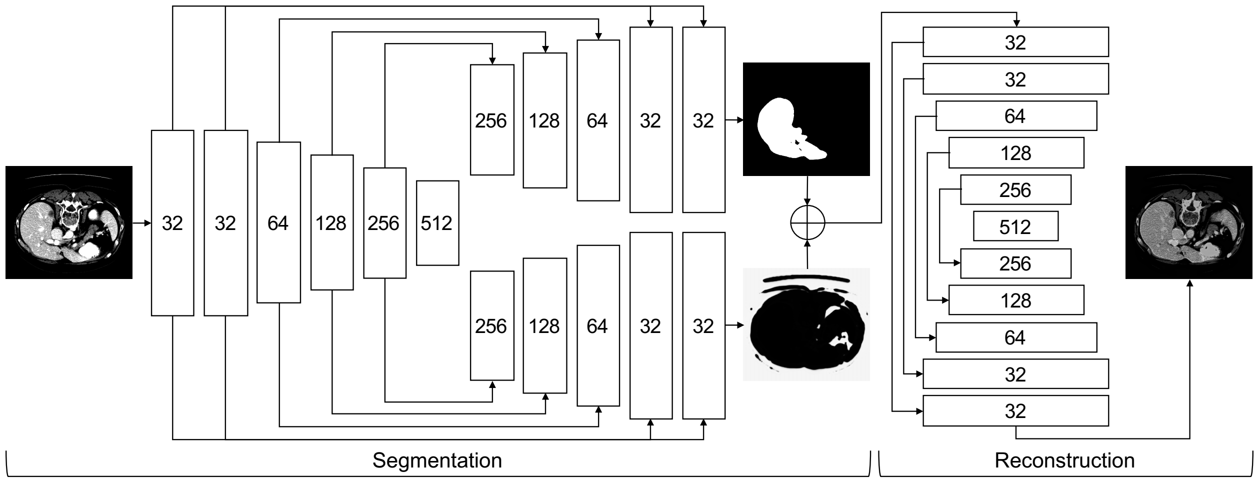


Figure 4.4: Network architecture of the 2D CompNet for the liver/large lesion segmentation. Each block has two convolutional layers with the numbers of feature maps showing in the block box.

4.3.1 Liver Segmentation

I first segment the liver from the original CT scan slice-by-slice using a 2D CompNet. As demonstrated by [18, 33, 55, 100] a 2D framework suffices to faithfully segment out the liver region. The basic CompNet in Section (3.3.2) with the feature maps in the blocks as 32, 32, 64, 128, 256 with a transition block 512 is used. All convolutional layers use a filter size of 3 $\times$ 3 and the following each block there are pooling layer for the encoder sections. The decoder section mirrors the encoder with pooling layers replaced by 2D transposed convolutional layers as shown in Figure 4.4.

4.3.2 Large Lesion Segmentation

Any tumor having a dimension greater than 32 in any of its axis, is considered a large lesion. This threshold point was chosen empirically for the LiTS dataset, to maximize speed and accuracy. Tumors labeled as 'large' in the training set, via component labelling, was processed slice by slice via a 2D CompNet having the same architecture as the 2D CompNet in Section 4.3.1.

4.3.3 Small Lesion Segmentation

To reduce false positives and obtain finer segmentation for small tumors, I used a 3D CompNet. To generate the training samples, I used component labelling to locate the center of a lesion and estimate its dimensions on a 2D slice. If it was less than 32 for both horizontal and vertical axis, I extracted a cube centered on the tumor on the 2D slice of size $32 \times 32 \times 32$. I select 15 slices above and 16 slices below the current slice. While this introduces redundancy, with the same tumor covered multiple times, it allows the framework to benefit from implicit image augmentation via shifts. The 3D CompNet used for this consists of blocks of two 64, 128, 256 feature maps with a transition block of 512. The filter size was set to $3 \times 3 \times 3$. A pooling layer followed each block in the encoder section. The decoder mirrored the encoder with pooling replaced by 3D transposed convolution layers as shown in Figure 4.5. During testing, I slide a cube over the liver volume to predict the small lesions. The stride size for the sliding cube is empirically selected to be 8. Anything higher and parts of smaller tumors may be missed and anything lower doesn't benefit the performance for LiTS dataset and simply increases the prediction time.

During testing, I ignore all predictions of size greater than $32 \times 32 \times 32$ from the large lesion segmentation module and anything greater than $32 \times 32 \times 32$ from the small lesion segmentation module. To fuse the overlapping predictions I take the average of all prediction at a voxel and set any value greater than 0.5 to 1.

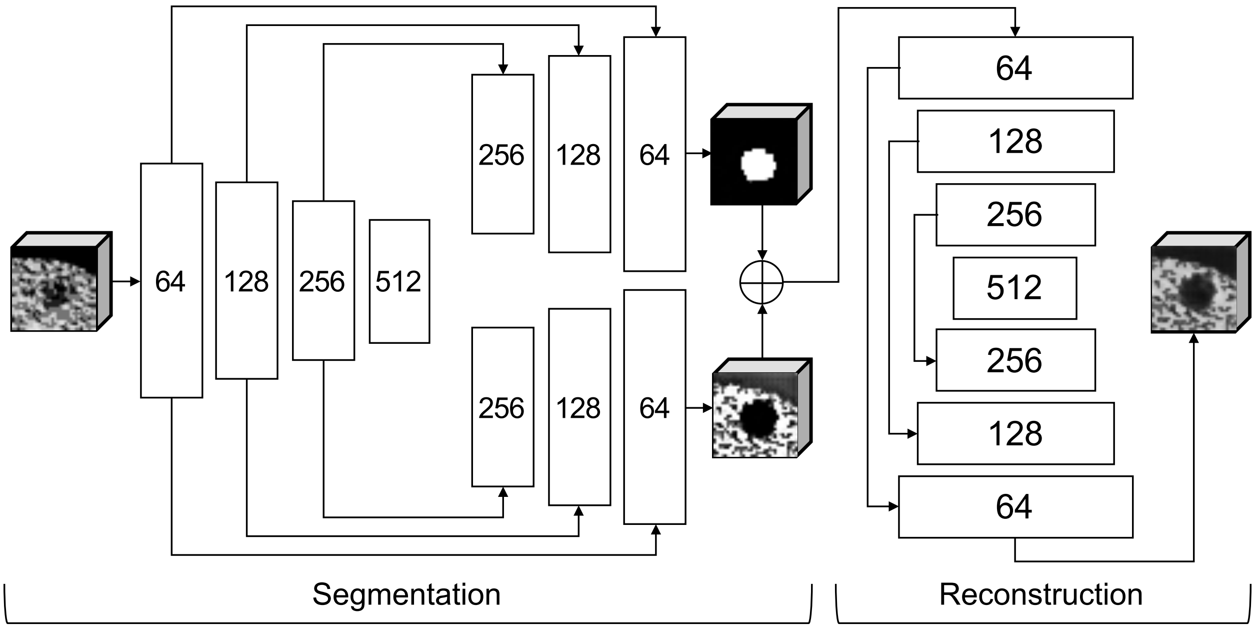


Figure 4.5: Network architecture of the 3D CompNet used for the small lesion segmentation. Each block has two convolutional layers with the numbers of feature maps showing in the block box.

4.4 Experiments

4.5 Dataset and Preprocessing

I use the public LiTS [11] dataset to demonstrate the efficacy of the CompNet in segmenting small tumors. LiTS consists of 130 abdomen CT scans for training and 70 for testing. To train the 2D liver segmentation network, I fully use all training scans with a total of 58,638 2D slices. In the network training of the large liver lesion segmentation, I focus on the slices with the liver present, resulting in 19,163 2D slices in total. The 3D network is trained on 11,503 3D small lesion samples with size of $32 \times 32 \times 32$. I preprocess the liver CT scans using a histogram based thresholding method. I select the rightmost peak of the intensity histogram distribution of a CT scan for normalization and use the histogram equalization algorithm to generate the enhanced images as shown in Figure 4.6.

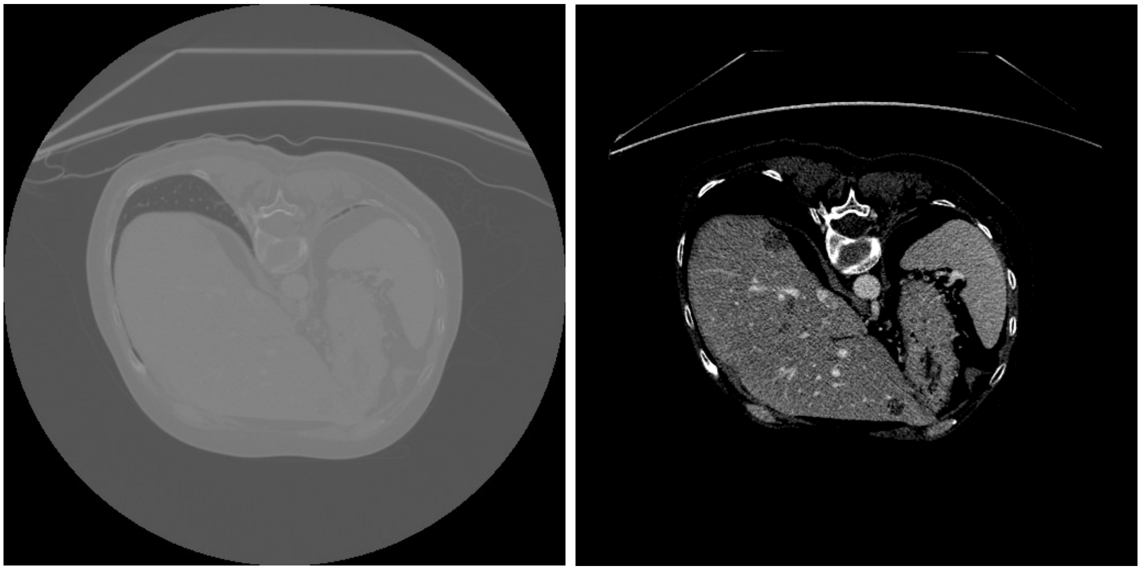


Figure 4.6: An example of a liver CT scan before (left) and after (right) preprocessing.

4.6 Experimental Settings

Keras with the TensorFlow backend is used to implement the proposed network. The 2D CompNet for the liver segmentation is trained for 40 epochs using the Adam optimizer with a learning rate of 5e-5. Next, I train the 2D and 3D CompNets for the large and small lesion segmentation in the same manner, i.e., first train the networks using the Adam optimizer with a learning rate of 5e-5, the same as that used in [18], and having an early stopping scheme with the tolerance being set to 5; then I train the networks with a learning rate of 1e-6 using an early stopping with a tolerance of 10 trials. Both steps have 150 maximum number of epochs for training. In addition, I use an L2 regularization with a parameter of 2e-4 and a dropout with a rate of 0.3 after all pooling and upsampling layers to mitigate overfitting.

Method	Dice Per Case	Dice Overall	No Pre-training	No Post-processing
H-Dense U-Net [55]	0.722	0.824		✓
Multiple U-Nets [18]	0.680	0.796	✓	
2.5 D U-Net [33]	0.670	—	✓	
CDNN [100]	0.657	0.820	✓	✓
FED-Net [17]	0.650	0.766	✓	
AH-Net [56]	0.634	0.834	✓	✓
RA U-Net [41]	0.595	0.795	✓	✓
BS-Unet [54]	0.552	0.729	✓	✓
This method (w/o small lesion)	0.681 (0.610)	0.813 (0.776)	✓	✓

Table 4.1: Comparison among published approaches and the Hybrid Cascaded Comp Net on the LiTS challenge.

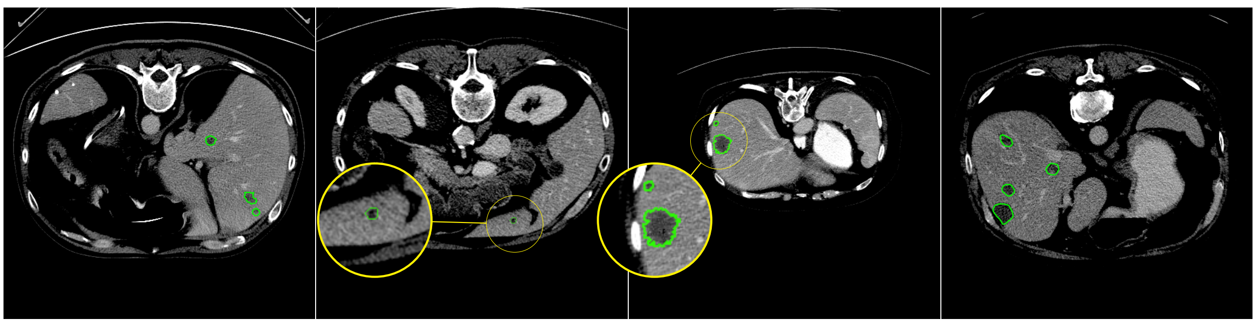


Figure 4.7: Examples of the predictions (indicated by the green lines) on the LiTS test set.

4.6.1 Experimental Results

Since I do not have the ground truth for the LiTS test set, I first perform the two-fold cross-validation on the training set to evaluate the performance of the proposed method quantitatively and qualitatively. I obtain 67.3% Dice per case and some visual samples are presented in Figure 4.8. By evaluating on the test set of the LiTS challenge, I list in Table 4.1 the result with comparison to those of the currently published approaches. According to the Dice score per case, the most important metric for measuring an algorithm's performance on the LiTS challenge, this approach is at the second rank, following the H-Dense U-Net [55], which however needs pre-training. In addition, the small lesion segmentation could be an

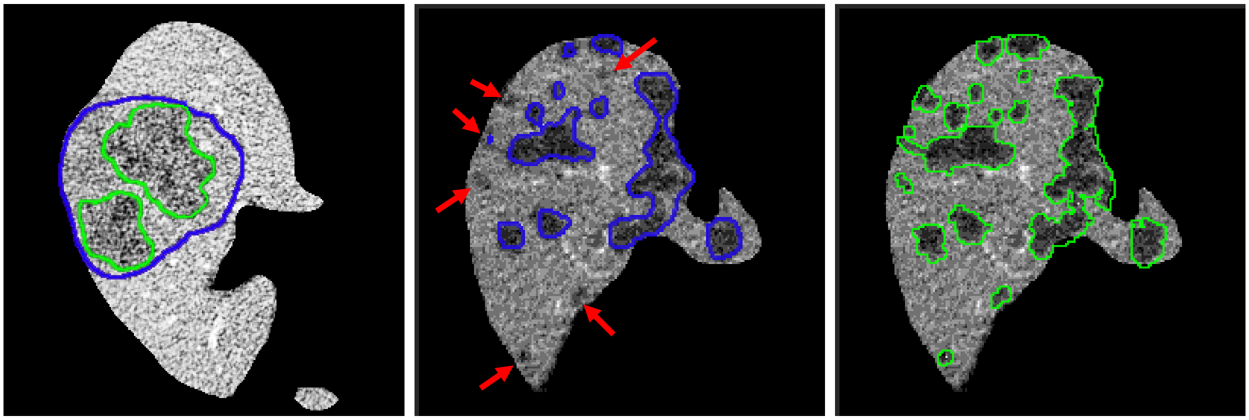


Figure 4.8: Comparison between our predictions (green) and the LiTS annotations (blue) using the two-fold cross-validation on the training set. The red arrows indicate the artifacts which share similar intensity distribution to tumors annotated on the same slice but are missing in the annotations while being predicted by our method.

add-on component to their method for a further improvement in segmenting small tumors as mentioned in their future work. My experiment shows that the dice per case can be improved from 0.61 to 0.681 with the consideration of the small lesion segmentation. Worth to mention that, among the methods without pre-training and post-processing, this method has the best performance in terms of the dice per case score. Figure 4.7 reports some visual results of the predictions on the LiTS test set.

4.7 Conclusion

CompNet can be successfully extended to 3D cases to segment small tumors.

While not using the 3D CompNet on the entire dataset, I have shown an hybrid cascaded approach may be optimal for the case of segmenting a mix of large and small liver lesions. The 3D segmentation network for small liver lesions improves the score by 7% and can be added on to a network that suffers from segmenting small objects.

Observations on LiTS annotation. As shown in Figure 4.8, the LiTS annotations have both over-segmentation and under-segmentation issues. For the large lesions in both cases shown in Figure 4.8, my predictions better fit lesions compared to the ground truth; while for the small lesions, my predictions locate more lesions potentially missing in the ground truth. Similar observations have been reported in [18]. Due to the imperfect ground truth provided by the LiTS challenge, I argue that the metrics computed against the ground truth probably could not be the only way to compare the segmentation results. Visual results could be considered as well, and this indicates the efficiency of our method. One of the reasons the ground truth labels may be over segmented is due to the purpose in the minds of the radiologists. It is always better to operate and cut out some healthy tissues, than risk leaving behind some tumor tissues. However, for the purpose of a machine learning algorithm the performance may drop due to inconsistency in the labelling schemes. Another reason could be the subjectivity of the different radiologists as to what constitutes a tumor. Since the data has been collected from multiple sources there could be additional bias at play here and in the next chapter, I explore potential unsupervised methods to mitigate such issues.

Chapter 5

Constrained Two Cuts via Complementary Learning for Unsupervised Anomaly Segmentation

5.1 Disclaimer

This chapter borrows from the research conducted by me under the supervision of my doctoral advisor Dr Yi Hong and published under the following titles -

(i) ASC-Net: Adversarial-Based Selective Network for Unsupervised Anomaly Segmentation [23]

(ii) ASC-Net: Adversarial-Based Selective Network for Unsupervised Anomaly Segmentation [Submitted to Medical Image Analysis]

5.2 Introduction

This chapter introduces the second application of complementary learning, unsupervised anomaly segmentation or novelty segmentation. In computer vision and medical image analysis, unsupervised image segmentation has been an active research topic for decades [28, 53, 72, 77, 89], due to its potential of applying to many applications without requiring the data to be manually labelled. Recently, advances in GANs [31] have given rise to a class of anomaly detection algorithms, which are inspired by AnoGAN [87] to identify abnormal events, behaviors, or regions in images or videos [20, 26, 88]. The AnoGAN learns a manifold of normal images by mapping from image space to a latent space based on GANs. To detect the anomaly, AnoGAN needs iterative search in the latent space to find the closest corresponding images for a query image. The AnoGAN family, including f-AnoGAN [86] and other works [7, 8, 43, 102, 103], focus on the reconstruction of the corresponding normal images for a query image, but not directly working on the anomaly detection. As a result, their reconstruction quality heavily affects the performance of anomaly detection.

To center the focus on the anomaly without needing faithful reconstruction, I propose an adversarial-based selective cutting neural network (ASC-Net), shown in Figure 5.1. This network aims to decompose an image into two selective cuts based on a reference image distribution. Typically, the reference distribution is defined by a set of images provided by users or experts who have vague knowledge and expectation of normal cases. In this way, one cut will fall into the reference distribution, while other image content outside of the reference image distribution will group into the other cut. These two cuts allow to reconstruct the original input image semantically and perform a simple intensity thresholding to cluster normal and abnormal regions. To consider these two cuts simultaneously, U-Net [80] with two up sampling branches is adopted, as used in CompNet [21], the supervised image segmentation approach. Meanwhile, one branch connects to a GAN's discriminator network,

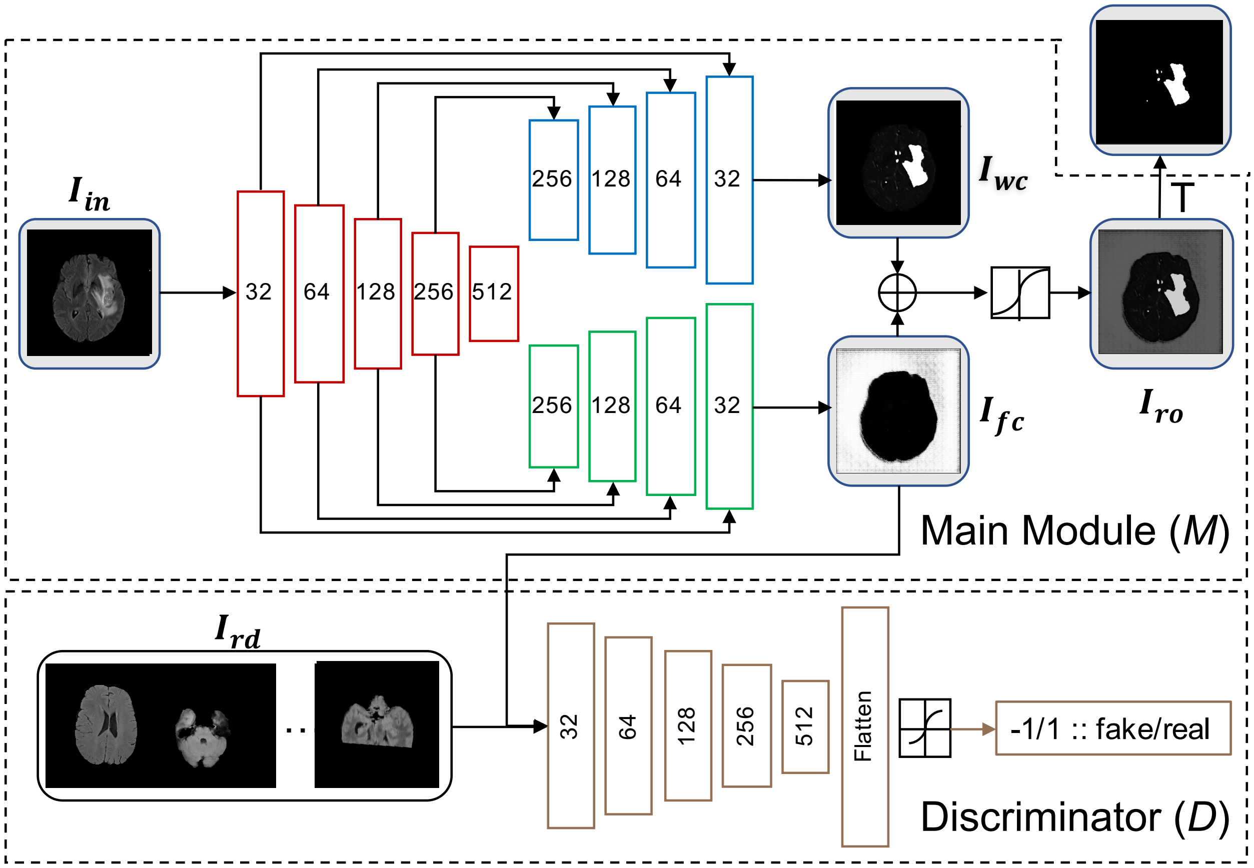


Figure 5.1: Overview of our proposed ASC-Net for unsupervised anomaly segmentation.

which allows introducing the knowledge contained in the reference image distribution. With the discriminator component aiding, the network can separate images into softly disjoint regions; that is, the generation of our selective cuts is under the constraint of the reference image distribution. As a result, a joint estimation of anomaly and the corresponding normal image is obtained, thus bypassing the need for perfect reconstruction. Furthermore, under the constraints of the GAN discriminator and the reconstruction of the original input, the ASC-Net becomes an unsupervised solution for anomaly detection in medical images, since except for a collection of normal images in the reference distribution, there are no other labels at the image level or annotations at the pixel/voxel level for the anomaly. I evaluate the proposed unsupervised anomaly segmentation network on three public datasets, i.e., MS-SEG2015 [13], BraTS-2019 [4, 5, 63], and LiTS [11] datasets and a private brain tumor dataset. For the MS-SEG2015 dataset, an exhaustive study on comparing multiple existing autoencoder-based models, variational-autoencoder-based models, and GAN-based models is performed in [6]. Compared to the best Dice scores reported in [6], significant gains in performance are obtained, with an improvement of 23.24% in mean Dice score without any post-processing and 20.40% with post-processing[1]. For BraTS dataset, the experiments show that f-AnoGAN, has difficulty reconstructing the normal images required for anomaly segmentation. By contrast, under the two-fold cross-validation settings we obtain a mean Dice score of 63.67% for the BraTS brain tumor segmentation before post-processing and 68.01% after post-processing. This result also outperforms another recent study in [70] which uses 50 BraTS 2018 scans for testing and performs comparably to [105] where the authors use an artificial tumor dataset to pretrain their framework to specialize it for tumor segmentation. Similarly, for the LiTS liver lesion segmentation, a mean dice score of 32.24% before post-processing and 50.23% using a simple post-processing scheme of open and closed sets, is obtained. This result outperforms [105] for the same task. I also evaluate our method on a private MRI T2 dataset with skull intact, curated from Zhongnan hospital in Wuhan, obtaining a mean dice score of 85.79%. Since the skull is left intact this dataset provides a more challenging scenario. However, it also allows the access to a healthy image set for reference distribution unlike the public datasets.

Overall, the contributions of our proposed method in this paper are summarized below:

- Proposing an adversarial based framework for unsupervised anomaly segmentation, which bypasses the normal image reconstruction and works on anomaly segmentation directly. This framework presents a general strategy to generate two selective cuts with incorporating human knowledge via a reference image set.

[1]Different from that in [6], I use a simple open-and-closed operation for the post-processing.

- To the best of my knowledge, this work is the first one to apply an unsupervised segmentation algorithm to both BraTS 2019 and LiTS liver lesion public datasets. Besides, the method outperforms the AnoGAN family and other popular methods presented in [6] on the publicly available MS-SEG2015 dataset.

5.3 Related Works

5.3.1 Unsupervised Segmentation Based on Clustering

Before the prevalence of deep neural networks, the primary means of medical image segmentation involve a combination of classical clustering operations followed by edge detection, watershed transformation such as in [32, 71]. W-Net [99] is a recent unsupervised image segmentation approach, which introduces the classical graph clustering method [89] into deep neural networks to find k-clusters in an image. Since W-Net requires the computation of the similarity score matrix for the pixels, it results in high-computational cost for images with high resolutions.

5.3.2 Unsupervised Anomaly Detection and Segmentation

The primary school of thought in the case of unsupervised anomaly detection and segmentation has revolved around the idea of training frameworks to learn the manifold of healthy samples, which helps filter out anomalies in a query image during testing for reconstruction and comparison. GAN-based and AutoEncoder (AE) Based algorithms are the two most represented categories in addressing the unsupervised anomaly detection and segmentation problem.

GAN-Based Methods.

The majority of GANs based application [102, 8, 43, 103] are motivated from the pioneering paper AnoGAN [87] which was later replaced with a faster version f-AnoGANs [86]. The principle idea of all of these frameworks involves training a neural network in an adversarial fashion to learn to generate perfect healthy images. Following which any unhealthy or anomalous image would not be constructive faithfully and thus a residual or subtraction between the regenerated image during testing and its input image would provide the anomaly.

AE-Based Methods. Following the success of AnoGAN idea of anomaly detection and accompanied by advances in Variational Autoencoders (VAE) [46] over the basic VAE [45], Adversarial autoencoders in [59] replace the KL-Divergence loss in VAE with an adversarial network to ensure that generating from any part of prior space results in meaningful samples. Similar combinations have also been proposed in [52] to create a merger between GANs and VAEs, which have been utilized for anomaly segmentation in brain MRI scans [7]. Different variations of VAEs have also been employed in [106, 15, 16] for brain tumor detection.

5.3.3 Other Methods

In [74, 58], the authors train a set of strongly annotated images as well as weakly annotated images in tandem. Recently, [105] introduced a self learning method where they create an artificial dataset to pretrain a network and test on similar datasets.

5.3.4 Differences from Existing Methods

AnoGANs and approaches inspired by AnoGAN [7, 8, 43, 103] including the VAE variants focus on regenerating the corresponding normal images for a query image, but not directly working on the anomaly detection. The performance of these methods on anomaly detection highly depends on the image reconstruction quality which makes the methods susceptible to GANs instability issue. Failing to reconstruct a high quality corresponding normal image will lead to inaccurate prediction results for the anomaly detection and segmentation task. My method does indeed reconstruct a healthy subset of an input image in one of the two cuts; however the corresponding reconstruction quality of the final image is never used as a termination criteria for the algorithm. Instead, a clean histogram separation with 3 or more distinct peaks is the requirement for the termination. This additionally frees the framework from the instability of the adversarial training scheme since termination is not based on reaching any kind of equilibrium between the two adversarial branches.

Unsupervised Segmentation and Anomaly Detection both aim to produce clusters that separate image pixels into disjoint groups. In the anomaly detection algorithms, the AnoGANs family [86, 87] separates a query image into the normal and abnormal groups by generating high-quality normal images based on Generative Adversarial Networks (GANs) [31] and identifying the anomaly as the subtraction between the query image and its generated normal image, while W-Net can perform 2 cut to obtain two clusters one normal and one anomalous. The primary disadvantage of W-Net for the task of anomaly segmentation is the formulation of n-cut which is done explicitly as a soft n-cut loss thus incurring a high computation cost of computing the similarity matrix for the pixels. This also requires additional post processing schemes such as the CRF [49] and Contour Maps[3].

Unlike, [105], this framework is general and doesnt require curation of separate artificial datasets for different tasks. In fact, this framework can segment MS-Lesions as well as Brain Tumors using the non-tumor slices of BraTS 2019 dataset.

5.4 Complementary Constraint for an Unsupervised Framework

One of the primary advantages of unsupervised methods is the lack of bias. In contrast supervised methods provide an accurate solution for a specific problem. CompNet 3.3.2 allows to combine both supervised and unsupervised aspects and provides robust solution to unseen pathologies in a test set while a traditional supervised method such as the U-Net would fail. However, in CompNet the complementary constraint is satisfied between one branch which has guidance from a ground truth, and an unconstrained complementary branch thus making it an supervised approach overall. In contrast the ASC-NET satisfies the complementary constraint between an image which falls in-distribution to the reference distribution and a second image which is unconstrained such that taken together they reconstruct the input image. In the following sections I discuss the architecture and idea in more details.

5.4.1 Network Framework

This network aims to decompose an input image into two discontinuous sub-images, which is achieved by the main module M shown in Figure 4.3. To guide the decomposition, the user knowledge is incorporated by defining a reference distribution of normal images, using the discriminator D. These two modules are the core components of the proposed ASC-Net, which is followed by a simple clustering step T based on thresholding to obtain the segmentation mask of the anomaly.

Main Module M. The main module aims to generate two selective cuts, which separate the normal and abnormal information in an input image. The M follows an encoder-decoder architecture like the U-Net, including one encoder and two decoders. The encoder E extracts features of an input image I_{in}, which could be an image located within or outside of the reference distribution $\{I_{rd}\}$, a collection of normal images.

One decoder in green (the second upsampling branch) is designed to generate a "fence" cut C_f that is constrained by an image fence formed by $\{I_{rd}\}$. The C_f aims to generate an image I_{fc} and tries to fool the discriminator D. The other decoder in blue (the first unsampling branch) is designed to generate another "wild" cut C_w, which captures leftover image content

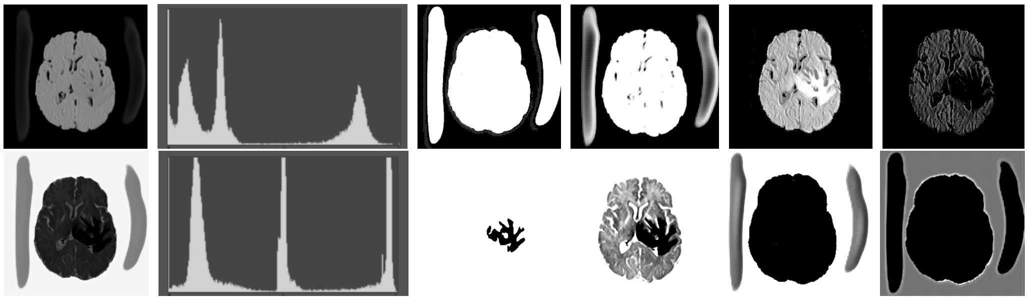


Figure 5.2: Visualization of the "disjoincy" between images I_{fc} (top) and I_{wc} (bottom) generated by two cuts of ASC-Net. From left to right: the generated image, its histogram, and the following four columns representing the histogram equalized images of the thresholded peaks with the first peak being the first image, etc. The first peak of I_{fc} is disjoint with the last peak of I_{wc}, etc.

that is not included in I_{fc}. As a result, the C_w produces another image I_{wc} to complement the fence-cut output I_{fc}. Since the wile-cut output is complementary to the fence-cut output, image information that cannot be covered by the reference distribution would be included in the wild-cut output, like the anomaly. The complementary relation between these two cuts C_f and C_w is enforced by a positive Dice loss, i.e., a "disjoincy" loss. Figure 5.2 demonstrates the "disjoincy" of I_{fc} and I_{wc}, like their complementary histogram distribution and different thresholded images at different peaks.

Aside from a weak guidance of the "disjoincy" loss for the "wild" cut, a reconstruction branch is adopted like the CompNet, to make sure the two selective cuts include enough information to construct a coarse version of the original input. The reconstructor R consists of a 1×1 convolution layer with the Sigmoid as the activation function, which is applied on the concatenation of the two-cut outputs I_{fc} and I_{wc} to regenerate the input image I_{in} back. This reconstructor R ensures that the C_f does not generate an image I_{fc} far from the input image I_{in} and ensures that the C_w does not generate an empty image I_{wc} if the anomaly or novelty exists. Figure 5.3 shows the histogram separation of the reconstructed images, compared to the original input images which present complex histogram peaks and have difficulty in separating the brain tumor from backgorund and other tissues via a simple thresholding. The discontinuous histogram distribution of I_{ro} is inherited from the two generated sub-images I_{fc} and I_{wc} through a simple weighted combination. As a result, the segmentation task becomes relatively easy to be done on the reconstructed image I_{ro}.

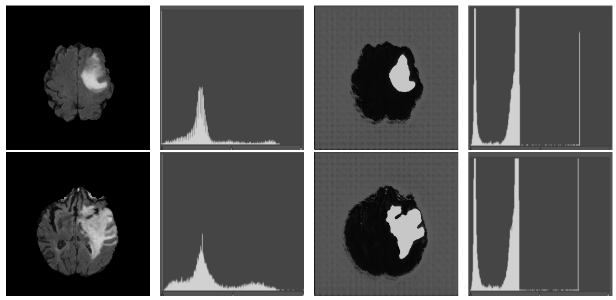


Figure 5.3: Histogram comparison of two sample images. From left to right: the input image, its histogram, its reconstructed image using ASC-Net, and the histogram of the reconstructed image. The histograms of the input images vary greatly, while the ones of their reconstructions show peaks at similar ranges, which enables a thresholding based pixel-level separation.

Discriminator D. The GAN discriminator tries to distinguish the generated image I_{fc}, according to a reference distribution R_d defined by a set of images $\{I_{rd}\}$, which are provided by the user or experts. The R_d typically includes images collected from the same group, for instance, normal brain scans, which share similar structures and lie on a manifold. Introducing D allows to incorporate the vague prior knowledge about a task into a deep neural network. Typically, it is non-trivial to explicitly formulate such prior knowledge; however, it could be implicitly represented by a selected image set. The R_d is an essential component that makes the ASC-Net able to generate selective cuts according to the user's input, without requiring other supervisions or pixel-level annotations.

Thresholding T.

The reconstruction branch consistently provides a reduced version of the original input image, where an unsupervised segmentation is easy to achieve via clustering. To cluster the reconstructed image I_{ro} into two groups at the pixel level, I choose the thresholding approach with the threshold values obtained using the histogram of I_{ro}. It is observed that for an anomaly that is often brighter than the surrounding tissues like the BraTS brain tumor, the intensity value at the rightmost peak of the histogram is a desired threshold; while an opposite case like darker LiTS liver lesions, the value at the leftmost peak would be the threshold. It is also observed that the histograms of the reconstructed images for different inputs reflect the same cut-off point for the left or right peaks, which allows using one threshold for an entire dataset. Thus, in practice on an unknown problem, both left most and right most peaks should be investigated before reaching a conclusion.

Loss Functions. The main module M includes three loss functions: (i) the image generation loss for C_f, $Loss_{C_f}$, (ii) the "disjoincy" loss between C_f and C_w, $Loss_{C_w}$, and (iii) the reconstruction loss, $Loss_R$. In particular, the C_f tries to generate an image I_{fc} that fools the discriminator D by minimizing $Loss_{C_f} = \frac{1}{n}\sum_{i=1}^{n} |D(I_{fc}^{(i)}) - 1|$. Here, n is the number of samples in the training batch. The C_w tries to generate an image I_{wc} that is complement to I_{fc} by minimizing the positive Dice score $Loss_{C_w} = \frac{2|I_{fc} \cap I_{wc}|}{|I_{fc}| + |I_{wc}|}$.

The last reconstruction takes a Mean-Squared-Error (MSE) loss between the input image I_{in} and the reconstructed image I_{ro}: $Loss_R = \frac{1}{n}\sum_{i=1}^{n} \|I_{in}^{(i)} - I_{ro}^{(i)}\|_2^2$. The discriminator D tries to reject the the C_f output I_{fc} but accept the images from the reference distribution R_d, by minimizing the following loss function: $Loss_D = \frac{1}{n+m}\left(\sum_{i=1}^{n} |D(I_{fc}^{(i)}) - (-1)| + \sum_{i=1}^{m} |D(I_{R_d}^{(i)}) - 1|\right)$. Here, m is the number of the images in R_d. Even though D and C_{fc} are tied in an adversarial setup, here I do not use the Earth Mover distance [81] in the loss function, since I would like D to identify both positive samples and negative samples with equal precision. Therefore, Mean Absolute Error (MAE) is used instead.

5.4.2 Architecture Details and Training Scheme

I use the same network architecture for all of the experiments as shown in Figure 4.3. The encoder E consists of four blocks of two convolution layers with a filter size of $(3, 3)$ followed by a max pooling layer with a filter size of $(2, 2)$ and batch normalization after every convolution layer. After every pooling layer I introduce a dropout of 0.3. The number of feature maps in each of the convolution layer of a block are 32, 64, 128, and 256. Following these blocks is a transition layer of two convolution layers with feature maps of size 512 followed by batch normalization layers. The C_{fc} and C_{wc} decoders are connected to the E and mirror the layers with the pooling layers replaced with 2D transposed convolutional layers, which have the same number of feature maps as the blocks mirror those in the encoder. Like a U-Net, I also introduce skip connections across similar levels in the encoder and decoders. The reconstructor R is simply a Sigmoid layer applied to the concatenation of I_{fc} and I_{wc}, resulting in a simplified CompNet [21]. The Discriminator D mimics the architecture of the E, except for the last layer where a dense layer is used for classification. All the intermediate layers have ReLU activation function and the final output layers have the Sigmoid activation. The only exception is the output of the discriminator D, which has a Tanh activation function to separate I_{fc} and images from the R_d to the maximum extent.

Keras with Tensorflow backend and Adam optimizer with a learning rate of 5e-5 is used to implement the architecture. There are two distinct training stages:

- In the *first stage*, I train D and M in cycles. I start training D with $\{R_d\}$ with True labels and $\{I_{fc}\}$ with False labels. These training samples are shuffled randomly. Following D, I train M with $\{I_{in}\}$ as input and the weights of D frozen while preserving the connection between $\{I_{fc}\}$ and D. The objective of the M is to morph the appearance of $\{I_{in}\}$ into $\{I_{fc}\}$ to fool D with the frozen weights. These two steps constitute one cycle, and in each step, there may be more than one epochs of training for M or D.

- In the *second stage*, M and D continue to be trained alternatively; however, the input images to D are changed, since the training purpose at this stage is to focus on the differences between the $\{R_d\}$ and $\{I_{in}\}$, while ignoring the noisy biases created by the M in transforming $\{I_{in}\}$ to $\{I_{fc}\}$. To achieve this, I augment the reference distribution $\{R_d\}$ with its generated images via M, i.e., $\{I_{fc}(R_d)\}$. I treat them as true images, and the union set $\{R_d \cup I_{fc}(R_d)\}$ is used to update D.

5.5 Applications

The model is evaluated on four unsupervised anomaly segmentation tasks: MS lesion segmentation, brain tumor segmentation, liver lesion segmentation and brain tumor segmentation with skull intact. I use the MS-SEG2015 [13] training set, BraTS [4, 5, 63], LiTS [11] and a private curated dataset in these tasks.

5.5.1 Datasets and Experimental Settings

MS-SEG2015. The training set consists of 21 scans from 5 subjects with each scan dimensions of $181 \times 217 \times 181$. The axial slices are resized to 160×160, so that the same network design can be shared between all rest of the experiments.

BraTS 2019. This dataset consists of 335 MRI brain scans collected from 259 subjects with high grade Glioma and 76 subjects with low grade Gliomas in the training set. The 3D dimensions of the images are $240 \times 240 \times 155$.

LiTS. The training set of LiTS consists of 130 abdomen CT scans of patients with liver lesions, collected from multiple institutions. Each scan has a varying number of slices with dimensions of 512×512. I resize these CT slices to 240×240 to share the same network architecture with other tasks.

Private Tumor Dataset. The private curated dataset from **Zhongnan hospital** in Wuhan, consists of non skull stripped brain MRI images of different modalities. To ensure consistency in the modality across all the training and testing phases and to maximize the utility of the images used, I selected the T2 modality which were obtained using Fast Spin Echo (FSE) to reduce imaging time. The reference distribution consisted of 55 subjects without any lesions present in their scans. I then used 26 subjects consisting of both patients with and without tumors and for whom there was no ground truth guidance as the training images. For testing, I used 41 subjects for which I had ground truth guidance to verify the framework's performance. The images were all different resolutions and thus I standardized them to 160×160 resolution. I kept the variable 3rd dimension as this framework operates on 2D. The final 2D slice count after removing near empty images was, 1224 for the reference distribution, 846 for the training images and 1151 for testing images.

For all experiments, the image intensity is normalized to $[0, 1]$ over the 3D volume; however, the 3D segmentation task is performed in the slice-by-slice manner using axial slices. To balance the sample size in I_{in} and R_d, I randomly sample and duplicate the number difference to the respective set.

5.5.2 Experimental Results

MS Lesion Segmentation. In this task, I randomly sample 2870 non-tumor, non-zero, BraTS-2019 training set slices to make the reference distribution R_d like [6], where they use their own privately annotated healthy dataset. Meanwhile, the 2870 non zero 2D slices of the MS-SEG2015 training set are used in the main module M. I train this network using three cycles in the first stage and one cycle in the second stage and take the threshold at 254 intensity, based on the right most peak of the image histogram.

I obtain an average Dice score of 32.94% without any post processing. By using a simple

Model	w/o post	w post
AE(dense)	4.5%	15.0 ±7.5%
Context AE	5.1%	18.8 ±11.6%
VAE	5.1%	20.0 ±12.4%
Context VAE (orig)	6.0%	26.7 ±11.2%
Context VAE (gradient)	5.9%	12.7 ±8.8%
GMVAE(dense)	6.3%	17.4 ±12.1%
GMVAE(dense restoration)	6.3%	22.3 ±12.4%
GMVAE(spatial)	2.8%	6.9 ±7.3%
GMVAE(spatial restoration)	2.7%	11.8 ±11.0%
f-AnoGAN	8.9%	27.8 ±14.0%
Constrained AE	9.7%	20.9 ±10.0%
Constrained AAE	6.8%	19.0 ±17.0%
VAE(restoration)	5.8%	21.1 ±12.2%
AnoVAEGAN	6.6%	20.0 ±13.3%
Ours	32.94±35.98%	48.20 ±47.84%

Table 5.1: Experimental comparison of anomaly segmentation on MS-SEG2015 across different methods, via experiments conducted in [6]

	BraTS 2019	LiTS	Private
Ours w/o post	63.67 ±16.24%	32.24±20.74%	85.79 ±47.63%
Ours w post	68.01±14.63	50.23±32.20%	70.90 ±49.32%

Table 5.2: Experimental comparison of anomaly segmentation on two datasets.

post-processing with erosion and dilation with 5×5 filters, this number improves to 48.20% Dice score. In comparison, a similar study conducted by [6] consisting of a multitude of algorithms including AnoVAEGAN [7] and f-AnoGANS, obtained a best mean score of 27.8% Dice after post processing by f-AnoGANS. Before post processing the best method was Constrained AutoEncoder [14] with a score of 9.7% Dice. An exhaustive list is presented in table 5.1. Sample images of our method is included in Figure 5.4

Brain Tumor Segmentation. In this task, I perform patient-wise two-fold cross-validation on the BraTS-2019 training set. In each training fold, I use a 90/10 split after removing empty slices. The 2D slices from the 90% split without tumors are used to make the reference

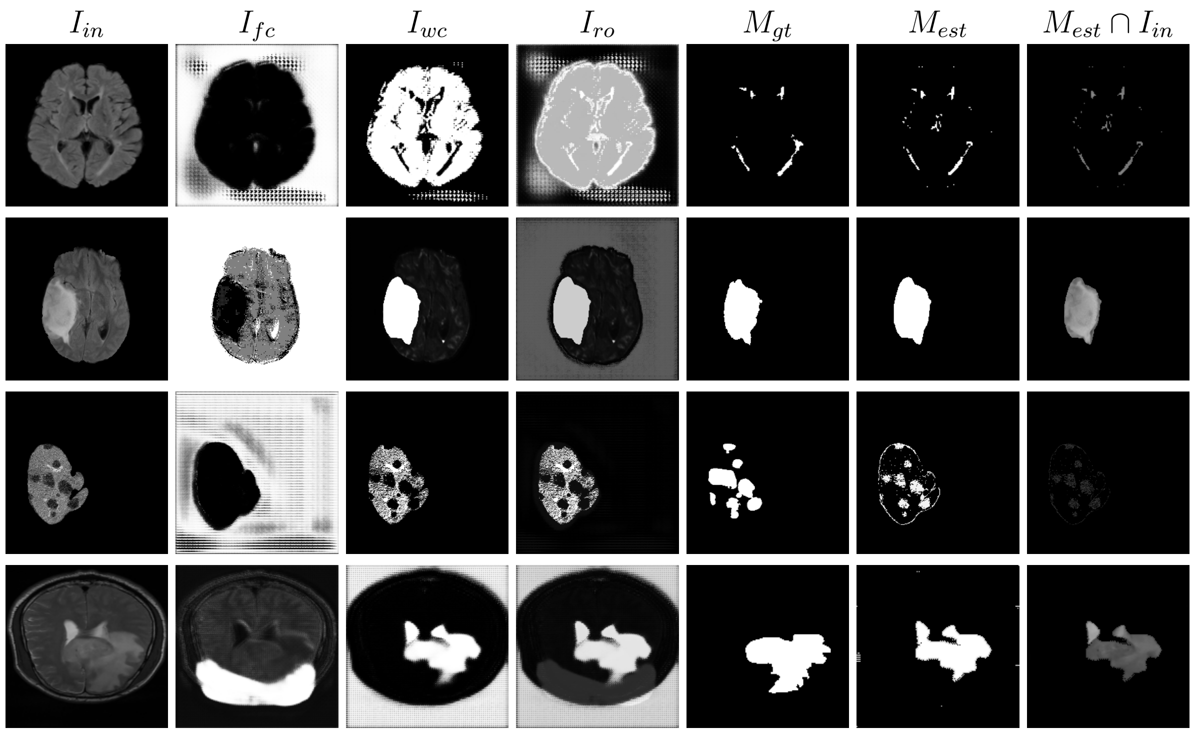


Figure 5.4: Sample results of MS-SEG2015, BraTS-2019, LiTS and Private Dataset (top to bottom) obtained from the various branches of the network. The I_{fc} in the second row is contrast enhanced to present the content contained in the brain region. None of these include any of the post processed images.

distribution R_d; while the 2D slices with tumors from the 90% split and all the slices from the 10% split are used for training the model. As a result, the sample size of R_d for fold one and two amounts to 11,745 and 12,407 respectively, while the size of I_{in} amounts to 11,364 and 10,786, respectively. This network is trained using two cycles in the first stage and one cycle in the second stage.

An mean Dice score of 63.67% for the brain tumor segmentation is obtained. Utilizing a simple post processing scheme of erosion and dilation with 9×9 filter we improve the mean Dice score to 68.01%. Figure 5.4 shows samples generated by the ASC-Net and Table 5.5.1 shows the before and after post processing results. Figure 5.5 shows the attempt to apply f-AnoGANs [86] by following their online instructions. The failure of AnoGANs in the reconstruction brings to light the issue with the regeneration focused methods and the complexity and stability of GAN based image reconstruction. Most recently [70] trained their algorithm using 1,112 healthy scans from the Human Connectome Project Young Adult (HCP) dataset [95] and tested their algorithm on 50 random BraTS 2018 scans obtaining a mean dice score of 67.2% and 15.5 % standard deviation. Following the simple post processing scheme my algorithm performs better on two fold cross validation across 335 scans. Another interesting work in [105] create an artificial dataset to resemble brain tumors which a self supervised networks trains on and test on the BraTS2018 training set obtaining a mean dice score of 71.63% and standard deviation of 0.84%. While their method outperforms my method, it is to be noted that the self supervised method is highly specialized to a particular task, in this case tumor segmentation. It may happen that the object to be segmented is difficult to synthesize artificially thus creating a bottleneck for the pipeline. Furthermore, the assumption of a self supervised learning algorithm is that the object to be learnt is known beforehand. Thus a model trained for brain tumors cannot be applied readily to MS-lesions for example. This method on the other hand suffers from no such limitations and is general. This method is generic and do not need new dataset generation for a new task. Further-more this method performs better than [105] on the liver lesion segmentation task utilizing a simple post processing scheme of erosion/dilation.

Liver Lesion Segmentation. To generate the image data for this task, I remove the non-liver region by using the liver mask generated by CompNet [21] and take all non-zero images. In the end there are 11,926 2D slices without liver lesions used in the reference distribution R_d. The remaining 6,991 images is then used for training the model. I perform slice-by-slice two-fold cross-validation and train the network using two cycles in both first and second stages. To extract the liver lesions, I first mask out the noises in the non-liver region of the reconstructed image I_{ro} and then invert the image to take a threshold value at

242, the rightmost peak of the inverted image.

A mean Dice score of 32.24% is obtained for this liver lesion segmentation, which improves to 50.23% by using a simple post processing scheme of erosion and dilation with 5×5 filter. Sampled results are shown in Figure 5.4. In comparison, a recent study [22] reports a cross-validation result of 67.3% under a supervised setting. It is to be noted that the annotation in the LiTS lesion dataset is imperfect with missing small lesions [18, 22]. Since I use the imperfect annotation to select images for the reference distribution, some slices with small lesions may be included and treated as normal examples. This framework also outperforms [105], which obtains a mean dice score of 40.78% and a standard deviation of 0.43%, following the simple post processing scheme.

Brain Tumor Segmentation from Non Skull Stripped MRI Scans. In this task, the reference distribution is composed of 1224 2D slices obtained from 55 patients and the training input image passed to the main module is composed of 846 2D slices from 26 subjects. The training input images consisted of a mixture of patients with and without tumors. The test set similarly consisted of patients with and without tumors from 41 subjects with 1151 2D slices. The threshold value of 220 was selected based off the rightmost image histogram peak resulting in a mean dice score of 85%. Further post processing resulted in a decline in the overall performance as compared to the other experiments and I address that in the conclusion section. Sample images are presented in Figure 5.4. I trained this framework using 2 stage one and 1 stage two cycles.

5.6 Conclusion

In this chapter I have presented a framework that performs two-cut split in an unsupervised fashion guided by a reference distribution using GANs. Unlike the methods in AnoGAN's family, ASC-Net focuses on the anomaly detection with the normal image reconstruction as a

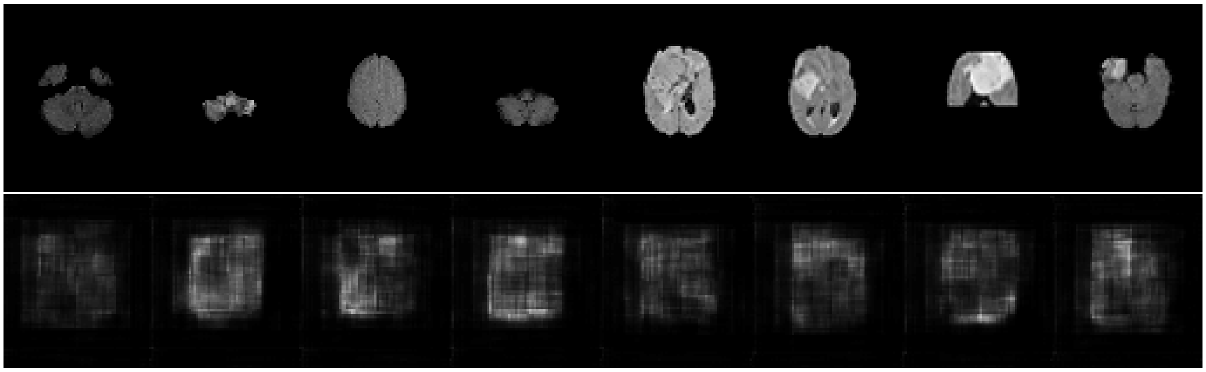


Figure 5.5: Query images (top) and their reconstructions (bottom) using f-AnoGANs [86].

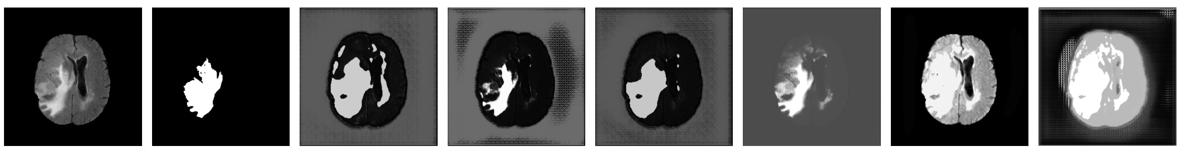


Figure 5.6: **Stability** : The first image is the input image, the second is the ground truth. The rest of images are reconstruction from various re-runs of the framework with variable training cycles and stage. All runs can isolate the anomaly in question.

byproduct, thus still producing competative results where reconstruction dependent methods such as f-AnoGAN fails to work on. The current version of the ASC-Net aims to solve the two-cut problem, which will be tasked to handle more than two selective cuts in the future. Theoretical understanding of the proposed network is also required, which is left as a future work.

Termination and stability. The termination point of this network training is periodic. The general guideline is that the peaks should be well separated, and I terminate the algorithm at three or four peak separations. However, continuing to train further may not always result in the improvement for the purpose of segmentation due to accumulation of holes as shown in Figure 5.7, even though visually the anomaly is captured in more intricate detail. I however encourage training longer as it reduces false positive and provide detailed anomaly reconstruction, though the Dice metric might not account for it. In the experiments,

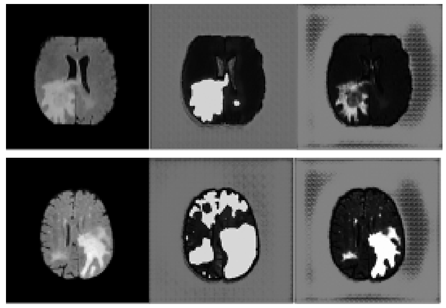


Figure 5.7: Termination of network training affects the reconstruction result. Left to right columns in each image: the input images, the images reconstructed via two cycles in the first stage and one in the second stage, and the images reconstructed via adding one cycle in the second stage.

I specify the number of cycles in each stage. However, due to the random nature of the algorithm and the lack of a particular purpose and guidance, the peak separation may occur much earlier, then training should be stopped accordingly. The reported network in the BraTS-2019 experiments has an mean Dice score of 6% over the network trained longer as shown in Figure 5.7. All the visual results are without post processing. Regarding the stability, Figure 5.6 demonstrates an anomaly estimated by different networks that are trained with different number of training cycles. It is observed that while the appearance of I_{ro} changes, the anomaly is still obtained as a separate cut since the framework works without depending on the quality of reconstruction.

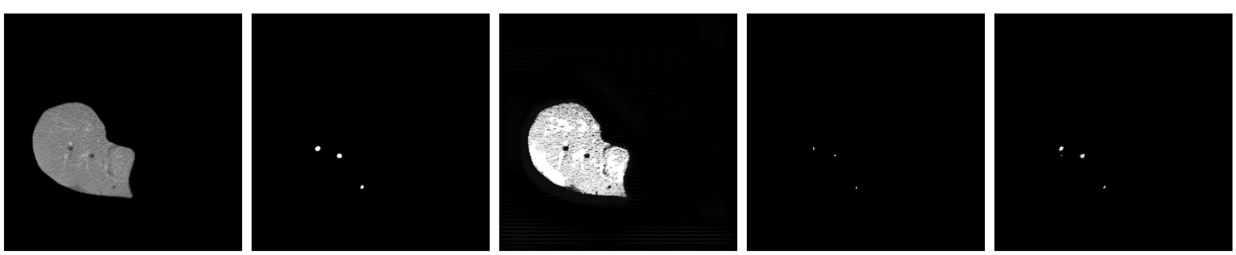


Figure 5.8: The difference in taking the threshold using the overall dataset's extreme peaks vs the individual sample's extreme peak. Left to Right: Input, ground truth, reconstruction, threshold at 242, threshold at 238 on LiTS dataset.

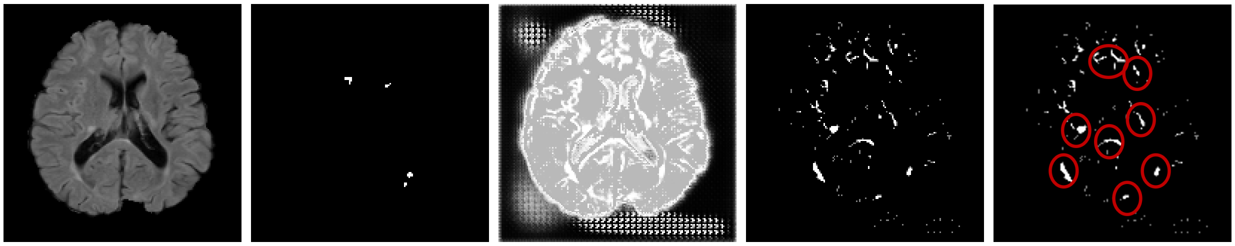


Figure 5.9: Demonstrating potential better interpretation on prediction compared to the ground truth. Left to Right: Input, ground truth, reconstruction, thresholded at rightmost peak, possible missed lesions on the ground truth highlighted in red ellipses on MS-SEG dataset.

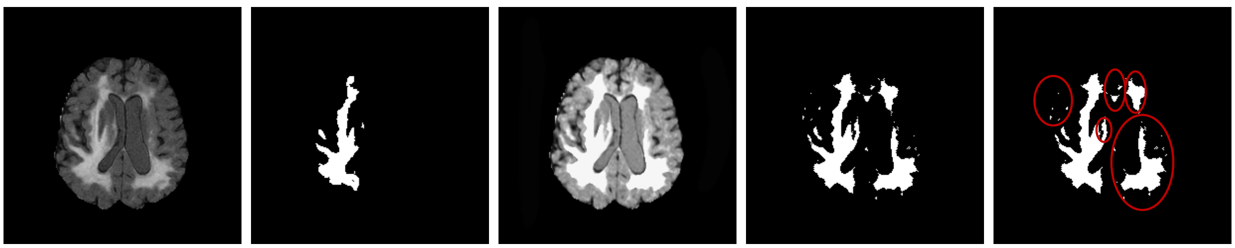


Figure 5.10: Demonstration of additional anomalies obtained in the prediction. Left to Right: Input, ground truth, reconstruction, thresholded at rightmost peak, potential missing labels in ground truth highlighted using red ellipses on BraTS dataset

Limitations and failing cases. The low Dice scores reported for the public datasets could be because I had to select non-tumor as normal slices as the reference distribution, which does not account for other co-morbidities. This affects the performance of the framework as it has no other guidance and would consider co-morbidities as an anomaly as well. This statement is partially supported by the better brain tumor segmentation results on the private dataset, which has better healthy scans for forming the reference distribution. Also, the method provides possibility of bringing other anomalies into the users' attention such as shown in Figure 5.10. Aside from co-morbidities in some samples the network creates a boundary of the organ region which in turn penalizes the framework in every slice e.g. Figure 5.4 third row. The faint liver boundary causes a fixed penalty to be incurred per slice thus reducing the dice score. In practice this kind of noise can be ignored by anyone using this framework for initial observation since it is quite distinct and easy to spot. Another reason could be taking a threshold on the entire test set. For example, in the liver lesion segmentation experiment, if I were to take a threshold at 238 based on the single sample's rightmost peak, I get a better final mask for the sample compared to taking the threshold at 242 based on the entire data set. The small tumors may also disappear under the post processing scheme used. While the result overall improves, due to the removal of the faint liver boundary in samples such as in Figure 5.4 third row for practical purposes non preprocessed version should be analyzed. The post processing simply helps to remove the extra noise which can easily be ignored by a physician or human inspecting the results. This would ensure the small tumors do not disappear even if it costs the performance to go down slightly in terms of dice score for the dataset.

Another point of interest as shown in Figure 5.9 for the MS-SEG challenge, is the presence of more details in prediction in comparison to the corresponding ground truth labels which may miss potential lesions. This however, results in reduced performance since the potential additional details are treated as noise in comparison to the ground truth.

High standard deviation. Unlike [105], where the network is pre-trained on a artificial tumor dataset and hence the pipeline customized for tumor segmentation, this method do not need such information beforehand. It is to be noted that the standard deviation for BraTS 2019 experiment is like [70]. This is because any novelty/anomaly detection algorithm without a pre-defined task would suffer from issues mention in Section 5.6. A second reason for high standard deviation could be because the network considers edge cases of tumors as tumors which may be ignored by radiologists. This is a speculation at this point and should be verified by a medical professional.

Bad post processing results on private dataset. In the case of the private dataset the mask obtained via the threshold is quite solid and contains minimal holes or additional noise after taking the threshold. Following this, the erosion dilation operation may add and subtract additional pixels thus lowering the performance. This is unique to this dataset and could be due to the choice of modality. In the BraTS experiment we use flair while for the private dataset we use T2 modality to maximize the total number of useable slices. It could also be since the healthy samples in the private dataset did not contain any tumors. In the public datasets, I had to compromise on the choice of reference distribution by choosing slices from patients who has tumors or lesions in the overall scans. Often co morbidities accompany brain tumors and the public datasets do not offer any resolution to this issue.

Chapter 6

Conclusion

This chapter summarizes and discusses the contributions of this dissertation and proposes future direction this research may be extended in.

This dissertation explored: 1) A simple way of formulating complementary constraint for two images, 2) Application of complementary constraint under a supervised setting, 3) Application of complementary constraint under an unsupervised setting. It also introduces a novel way to overcome the vanishing gradient problem. In the following sections I restate and discuss of each of the contributions.

(i) *Formulation of an easy to use loss function to enforce complementary constraint*

In Chapter 1, the motivation of complementary learning was explored and a low cost solution of enforcing it for pairs of images was suggested. The simple act of reversing the sign of the dice loss effectively enforces this disjoincy and provides a respite from defining complicated metric to define two clusters where one cluster characterizes a Region of Interest (ROI). The primary challenge in any other method in literature currently lies in not just in the complexity of finding a basis for the cluster but also in the computation of the said clusters where the back propagation steps are heavily bottlenecked via the computation of the similarity matrix such as in [99]. This further limits the overall dimensions of the image that can be supported without requiring an extreme hardware capacity. In contrast the reversed dice loss can be incorporated very easily on any image resolution where a standard dice loss based algorithm such as a U-Net can be used.

The immediate next steps to improve the existing complimentary constraint is to enforce it for multiple images instead of limiting it to pairwise images like proposed in this dissertation. This is not trivial and could be achieved in a few ways. One of which is via chaining pairwise disjoincy across multiple levels in the architecture. What it means for each of the two methods is listed in their separate sections below.

(ii) *Formulation of a neural network model the Comp-Net which enhances the robustness of a supervised segmentation method via utilizing the non ROI features to provide an additional fail safe for ROI Interactive Pathology.*

In Chapter 3 and 4, the complementary constraint is applied via a disjoincy loss using a modified U-Net type architecture where there are two parallel decoders to produce the Segmentation Output (SO), and the Complementary Output (CO), which taken together attempts to reconstruct the original input image (RO). The SO is guided via the ground truth during training and enforces the benefits of a supervised framework, while the CO is its complement and provides the benefit of the lack of bias as seen in unsupervised methods in a local context. The RO prevents the CO from producing an empty image (i.e., the trivial null set) and taken together formed the complementary head. I have shown through the experiments on skull stripping using the modified synthetic subset, that a traditional training scheme utilized for a U-Net like architecture is unable to correctly extract a ROI in the presence of ROI Interactive Pathologies during testing.

The existing framework only supports a binary segmentation and is not directly suitable for a multi class segmentation. One approach would be to break down the initial problem into n segmentation network for n class segmentation with each network dedicated towards segmenting one of the n classes. However, this is inefficient and cannot be trained end to end. An alternative could be to hierarchically chain CompNets in series. Such an implementation could be bottle-necked by the GPU capacity, and it may not be possible to extend it past a few classes. An efficient solution to this could be in figuring out some way to utilize the same base encoder decoder structure and providing an exhaustive pairwise loss for each segmentation level.

Another simple change could be in the removal of the final encoder-decoder section for the RO branch, since the primary purpose of that is to prevent CO from producing an empty image and as such the quality of reconstruction may not be of paramount importance as seen in ASC-NET.

(iii) *A constrained selective two cut framework satisfying disjoincy loss between two images where one image is regulated via a reference distribution specified by the user using a discriminator. Given an input image this framework splits its space into two disjoint entities with one falling in-distribution with the reference distribution and the other being the anomaly and novelty.*

Chapter 5 introduces an extension of the CompNet by replacing the guidance provided by a ground truth mask with a reference distribution enforced via a discriminator. This relaxation of control over the learning objective in the SO branch increases the scope of the framework drastically. Firstly, the discriminator can model the complex manifold of the healthy images which would be quite difficult to characterize otherwise. Secondly, it also removes the biases that may exist in human judgement. I have shown that this framework outperforms all existing anomaly segmentation algorithm for the task of MS-lesion segmentation. For brain tumor segmentation, this method also performs the best amongst all true novelty/anomaly segmentation algorithms. In [105], the self supervised method is tailored for the task of brain tumor segmentation via an artificial dataset and obtain a 3% increase in performance compared to ASC-Net. However, it cannot segment MS-lesions without creating a new synthetic dataset. In contrast my design can segment both MS-lesions and brain tumors using the same set of non tumor slices of the brain tumor dataset. Furthermore, this framework outperforms [105] on the liver lesion segmentation task after applying a simple post processing scheme. The private dataset demonstrates a much better capacity of this work, mainly due to the availability of a good reference distribution obtaining 85.79% mean dice score on a dataset which had the skull intact.

At this point the framework is deemed to have terminated after a clear peak separation in the RO branch is obtained. This involves checking the performance after every cycle to determine a good stopping point. A future work could be in automating this process and stopping the training automatically. However, that may introduce biases as to what constitute a clear peak and may vary from data scientist to data scientist and between application to application, so care must be taken to create a general set of flexible rules governing the termination. An even better solution could be to formulate the peak separation problem as a loss function and incorporating it into the learning objective of the network.

Like the CompNet at this stage all anomalies are classified under a unified label and similar changes as discussed in the previous section can also be incorporated to further differentiate the different classes of anomaly/novelty.

Aside from the similar challenges shared with the CompNet, ASC-Net suffers from its own unique challenges, one of which is to segment an anomaly inside another anomaly. This can be solved in two ways, the first of which is to create a new reference distribution for the secondary task. The second case would be to perform some kind of true unsupervised technique such as wavelet decomposition or principle component analysis. Since the primary region of anomaly/novelty shrinks the original image space these techniques might perform better under such circumstances.

This framework also demonstrates critical insights which may be missing from annotation across datasets curated from different institutions as shown on the public MS-SEG and BraTS datasets in Figures 5.9 and 5.10 respectively. It can also locate very small tumors accurately if a threshold is taken at the level of an individual sample instead of the complete dataset as demonstrated in Figure 5.8. While the performance of the framework may increase with the inclusion of a post processing scheme since it removes background noise from some scans as seen in Figure 5.4, Row 3 Column 6 and 7, important indicators such as the small tumors of Figure 5.8 may get erased at the same time. The purpose of this framework is to assist in the initial discovery of anomalous or novel marker and as such a post processing scheme may be detrimental to the potential interpretation of the results. Following the initial discovery, other segmentation methods or human inspection could provide better interpretability and insights instead.

Each of the above mentioned algorithms could also benefit from utilizing all the different imaging modalities such as T1, T2, T1ce, and Flair simultaneously. Furthermore, utilizing all the different views as mentioned in Section 2.8.3 could also improve the shortcomings of utilizing any one plane. All our experiments have been conducted with only one plane so far and it can be transformed to take an aggregate of all the planes.

Bibliography

[1] Mri basics. https://case.edu/med/neurology/NR/MRI%20Basics.htm.

[2] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[3] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2010.

[4] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117, 2017.

[5] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.

[6] Christoph Baur, Stefan Denner, Benedikt Wiestler, Nassir Navab, and Shadi Albarqouni. Autoencoders for unsupervised anomaly segmentation in brain mr images: A comparative study. *Medical Image Analysis*, page 101952, 2021.

[7] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *International MICCAI Brainlesion Workshop*, pages 161–169. Springer, 2018.

[8] Amanda Berg, Jörgen Ahlberg, and Michael Felsberg. Unsupervised learning of anomaly detection from contaminated image data using simultaneous encoder training. *arXiv preprint arXiv:1905.11034*, 2019.

[9] Serge Beucher. Use of watersheds in contour detection. In *Proceedings of the International Workshop on Image Processing*. CCETT, 1979.

[10] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203, 1984.

[11] Patrick Bilic, Patrick Ferdinand Christ, Eugene Vorontsov, Grzegorz Chlebus, et al. The liver tumor segmentation benchmark (lits). *arXiv:1901.04056*, 2019.

[12] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[13] Aaron Carass, Snehashis Roy, Amod Jog, Jennifer L Cuzzocreo, Elizabeth Magrath, Adrian Gherman, Julia Button, James Nguyen, Ferran Prados, Carole H Sudre, et al. Longitudinal multiple sclerosis lesion segmentation: resource and challenge. *NeuroImage*, 148:77–102, 2017.

[14] Xiaoran Chen and Ender Konukoglu. Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders. *arXiv preprint arXiv:1806.04972*, 2018.

[15] Xiaoran Chen, Nick Pawlowski, Martin Rajchl, Ben Glocker, and Ender Konukoglu. Deep generative models in the real-world: An open challenge from medical imaging. *arXiv preprint arXiv:1806.05452*, 2018.

[16] Xiaoran Chen, Suhang You, Kerem Can Tezcan, and Ender Konukoglu. Unsupervised lesion detection via image restoration with a normative prior. *Medical image analysis*, 64:101713, 2020.

[17] Xueying Chen, Rong Zhang, and Pingkun Yan. Feature fusion encoder decoder network for automatic liver lesion segmentation. *arXiv:1903.11834*, 2019.

[18] Grzegorz Chlebus, Andrea Schenk, Jan Hendrik Moltz, Bram van Ginneken, Horst Karl Hahn, and Hans Meine. Automatic liver tumor segmentation in ct with fully convolutional neural networks and object-based postprocessing. *Scientific reports*, 8(1):1–7, 2018.

[19] Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A Elshaera, et al. Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. *arXiv:1702.05970*, 2017.

[20] Allison Del Giorno, J Andrew Bagnell, and Martial Hebert. A discriminative framework for anomaly detection in large videos. In *European Conference on Computer Vision*, pages 334–349. Springer, 2016.

[21] Raunak Dey and Yi Hong. Compnet: Complementary segmentation network for brain mri extraction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 628–636. Springer, 2018.

[22] Raunak Dey and Yi Hong. Hybrid cascaded neural network for liver lesion segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1173–1177. IEEE, 2020.

[23] Raunak Dey and Yi Hong. Asc-net: Adversarial-based selective network for unsupervised anomaly segmentation. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 236–247, Cham, 2021. Springer International Publishing.

[24] Raunak Dey, Zhongjie Lu, and Yi Hong. Diagnostic classification of lung nodules using 3d neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 774–778. IEEE, 2018.

[25] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[26] Sarah M Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, 58:121–134, 2016.

[27] Farzan Farnia and Asuman Ozdaglar. Gans may have no nash equilibria. *arXiv preprint arXiv:2002.09124*, 2020.

[28] Nathalie Giordana and Wojciech Pieczynski. Estimation of generalized multisensor hidden markov chains and unsupervised image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):465–475, 1997.

[29] Lovedeep Gondara. Medical image denoising using convolutional denoising autoencoders. In *2016 IEEE 16th international conference on data mining workshops (ICDMW)*, pages 241–246. IEEE, 2016.

[30] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[31] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[32] Vicente Grau, AUJ Mewes, M Alcaniz, Ron Kikinis, and Simon K Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE transactions on medical imaging*, 23(4):447–458, 2004.

[33] Xiao Han. Automatic liver lesion segmentation using a deep convolutional neural network method. *arXiv:1704.07239*, 2017.

[34] Xu Han, Roland Kwitt, Stephen Aylward, Spyridon Bakas, Bjoern Menze, Alexander Asturias, Paul Vespa, John Van Horn, and Marc Niethammer. Brain extraction from normal and pathological images: a joint pca/image-reconstruction approach. *NeuroImage*, 176:431–445, 2018.

[35] Geoffrey E Hinton and Richard S Zemel. Autoencoders, minimum description length, and helmholtz free energy. *Advances in neural information processing systems*, 6:3–10, 1994.

[36] Steven L Horowitz. Picture segmentation by a directed split-and-merge procedure. In *IJCPR*, pages 424–433, 1974.

[37] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[38] Weimin Huang, Yongzhong Yang, Zhiping Lin, Guang-Bin Huang, Jiayin Zhou, Yuping Duan, and Wei Xiong. Random feature subspace ensemble based extreme learning machine for liver tumor detection and segmentation. In *EMBC*, pages 4675–4678. IEEE, 2014.

[39] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[40] Huiyan Jiang, Tianyu Shi, Zhiqi Bai, and Liangliang Huang. Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes. *IEEE Access*, 2019.

[41] Qiangguo Jin, Zhaopeng Meng, Changming Sun, Leyi Wei, and Ran Su. Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans. *arXiv:1811.01328*, 2018.

[42] P Kalavathi and VB Surya Prasath. Methods on skull stripping of mri head scan images—a review. *Journal of digital imaging*, 29(3):365–379, 2016.

[43] Masanari Kimura and Takashi Yanagihara. Anomaly detection using gans for visual inspection in noisy training data. In *Asian Conference on Computer Vision*, pages 373–385. Springer, 2018.

[44] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[45] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[46] Diederik P Kingma and Max Welling. Stochastic gradient vb and the variational auto-encoder. In *Second International Conference on Learning Representations, ICLR*, volume 19, page 121, 2014.

[47] Diederik P Kingma and Max Welling. An introduction to variational autoencoders. *arXiv preprint arXiv:1906.02691*, 2019.

[48] Jens Kleesiek, Gregor Urban, Alexander Hubert, Daniel Schwarz, Klaus Maier-Hein, Martin Bendszus, and Armin Biller. Deep mri brain extraction: A 3d convolutional neural network for skull stripping. *NeuroImage*, 129:460–469, 2016.

[49] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011.

[50] Johann Heinrich Lambert. *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, volume 2. Verlag des Buchladens der Realschule, 1770.

[51] Curtis P Langlotz, Bibb Allen, Bradley J Erickson, Jayashree Kalpathy-Cramer, Keith Bigelow, Tessa S Cook, Adam E Flanders, Matthew P Lungren, David S Mendelson, Jeffrey D Rudie, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 nih/rsna/acr/the academy workshop. *Radiology*, 291(3):781–791, 2019.

[52] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1558–1566, New York, New York, USA, 20–22 Jun 2016. PMLR.

[53] Te-Won Lee and Michael S Lewicki. Unsupervised image classification, segmentation, and enhancement using ica mixture models. *IEEE Transactions on Image Processing*, 11(3):270–279, 2002.

[54] Song Li and Geoffrey Kwok Fai Tso. Bottleneck supervised u-net for pixel-wise liver and tumor segmentation. *arXiv:1810.10331*, 2018.

[55] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018.

[56] Siqi Liu, Daguang Xu, S Kevin Zhou, Olivier Pauly, Sasa Grbic, Thomas Mertelmeier, Julia Wicklein, Anna Jerebko, Weidong Cai, and Dorin Comaniciu. 3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes. In *MICCAI*, pages 851–858. Springer, 2018.

[57] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[58] Fei Ma, Fei Gao, Jinping Sun, Huiyu Zhou, and Amir Hussain. Weakly supervised segmentation of sar imagery using superpixel and hierarchically adversarial crf. *Remote Sensing*, 11(5):512, 2019.

[59] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.

[60] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507, 2007.

[61] Laurent Massoptier and Sergio Casciaro. A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from ct scans. *European radiology*, 18(8):1658, 2008.

[62] Francesco Masulli and Andrea Schenone. A fuzzy clustering based segmentation system as support to diagnosis in medical imaging. *Artificial intelligence in medicine*, 16(2):129–147, 1999.

[63] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[64] Fernand Meyer and Serge Beucher. Morphological segmentation. *Journal of visual communication and image representation*, 1(1):21–46, 1990.

[65] Jan Hendrik Moltz, Lars Bornemann, Volker Dicken, and H Peitgen. Segmentation of liver metastases in ct scans by adaptive thresholding and morphological processing. In *MICCAI workshop*, volume 41, page 195, 2008.

[66] Jan Hendrik Moltz, Lars Bornemann, Jan-Martin Kuhnigk, Volker Dicken, et al. Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in ct scans. *IEEE Journal of selected topics in signal processing*, 3(1):122–134, 2009.

[67] J. L. Muerle and D. C. Allen. Experimental evaluation of techniques for automatic segmentation of ob]ecm in a complex scene. pages 3–13. 1968.

[68] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[69] John F Nash. 4. the bargaining problem. In *The Essential John Nash*, pages 37–48. Princeton University Press, 2016.

[70] Sergio Naval Marimont and Giacomo Tarroni. Implicit field learning for unsupervised anomaly detection in medical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 189–198. Springer, 2021.

[71] H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh, and W.L. Nowinski. Medical image segmentation using k-means clustering and improved watershed algorithm. In *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 61–65, 2006.

[72] Robert J O'Callaghan and David R Bull. Combined morphological-spectral unsupervised image segmentation. *IEEE transactions on image processing*, 14(1):49–62, 2004.

[73] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[74] Xi Ouyang, Zhong Xue, Yiqiang Zhan, Xiang Sean Zhou, Qingfeng Wang, Ying Zhou, Qian Wang, and Jie-Zhi Cheng. Weakly supervised segmentation framework with uncertainty: A study on pneumothorax segmentation in chest x-ray. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 613–621. Springer, 2019.

[75] Seong Ho Park and Kyunghwa Han. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*, 286(3):800–809, 2018.

[76] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970.

[77] Jan Puzicha, Thomas Hofmann, and Joachim M Buhmann. Histogram clustering for unsupervised image segmentation. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 2, pages 602–608. IEEE, 1999.

[78] Vincenzo Riccati and Girolamo Saladini. *INSTITUTIONES ANALYTICAE A VINCENTIO RICCATO SOCIETATIS JESU ET HIERONYMO SALADINO MONACHO CAELESTINO COLLECTAE.: TOMUS PRIMUS*, volume 1. ex typographia Sancti Thomae Aquinatis, 1765.

[79] Felix Ritter, Tobias Boskamp, A. Homeyer, Hendrik Laue, Michael Schwier, Florian Link, and H.-O. Peitgen. Medical image analysis. *IEEE Pulse*, 2(6):60–70, 2011.

[80] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[81] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[82] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[83] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[84] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging. *IEEE transactions on medical imaging*, 36(11):2319–2330, 2017.

[85] Jean Sauri. *Cours complet de mathématiques*, volume 1. Ruault, 1774.

[86] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019.

[87] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017.

[88] Philipp Seeböck, Sebastian Waldstein, Sophie Klimscha, Bianca S Gerendas, René Donner, Thomas Schlegl, Ursula Schmidt-Erfurth, and Georg Langs. Identifying and categorizing anomalies in retinal imaging data. *arXiv preprint arXiv:1612.00686*, 2016.

[89] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[90] Irwin Sobel, R Duda, P Hart, and John Wiley. Sobel-feldman operator.

[91] M. Sonka, G.K. Reddy, M.D. Winniford, and S.M. Collins. Adaptive approach to accurate analysis of small-diameter vessels in cineangiograms. *IEEE Transactions on Medical Imaging*, 16(1):87–95, 1997.

[92] Roberto Souza, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170:482–494, 2018.

[93] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[94] G Sudhavani and K Sathyaprasad. Segmentation of lip images by modified fuzzy c-means clustering algorithm. In *International Journal of Computer Science and Network Security IJCSNS*. Citeseer, 2009.

[95] David C Van Essen, Kamil Ugurbil, Edward Auerbach, Deanna Barch, Timothy EJ Behrens, Richard Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[96] Huynh Van Lung and Jong-Myon Kim. A generalized spatial fuzzy c-means algorithm for medical image segmentation. In *2009 IEEE International Conference on Fuzzy Systems*, pages 409–414. IEEE, 2009.

[97] Pierre François Verhulst. *Traité élémentaire des fonctions elliptiques: ouvrage destiné à faire suite aux traités élémentaires de calcul intégral*. Hayez, 1841.

[98] Martin J Willemink, Wojciech A Koszek, Cailin Hardell, Jie Wu, Dominik Fleischmann, Hugh Harvey, Les R Folio, Ronald M Summers, Daniel L Rubin, and Matthew P Lungren. Preparing medical imaging data for machine learning. *Radiology*, 295(1):4–15, 2020.

[99] Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation. *arXiv preprint arXiv:1711.08506*, 2017.

[100] Yading Yuan. Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation. *arXiv:1710.04540*, 2017.

[101] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[102] Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Ramaseshan Chandrasekhar. Efficient gan-based anomaly detection. *arXiv preprint arXiv:1802.06222*, 2018.

[103] Houssam Zenati, Manon Romain, Chuan-Sheng Foo, Bruno Lecouat, and Vijay Chandrasekhar. Adversarially learned anomaly detection. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 727–736. IEEE, 2018.

[104] Nianyin Zeng, Hong Zhang, Baoye Song, Weibo Liu, Yurong Li, and Abdullah M Dobaie. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing*, 273:643–649, 2018.

[105] Xiaoman Zhang, Weidi Xie, Chaoqin Huang, Ya Zhang, and Yanfeng Wang. Self-supervised tumor segmentation through layer decomposition. *arXiv preprint arXiv:2109.03230*, 2021.

[106] David Zimmerer, Simon AA Kohl, Jens Petersen, Fabian Isensee, and Klaus H Maier-Hein. Context-encoding variational autoencoder for unsupervised anomaly detection. *arXiv preprint arXiv:1812.05941*, 2018.