Optimizacija algoritama dubokog učenja za obradu slika kardiovaskularnog sustava korištenjem rezidualnih jedinica

Habijan, Marija

MARIJA HABIJAN

OPTIMIZATION OF DEEP LEARNING ALGORITHMS FOR THE CARDIOVASCULAR SYSTEM IMAGE PROCESSING USING RESIDUAL UNITS

Optimization of Deep Learning Algorithms for the Cardiovascular System Image Processing Using Residual Units

Optimizacija algoritama dubokog učenja za obradu slika kardiovaskularnog sustava korištenjem rezidualnih jedinica

Optimalisatie van deep learning-algoritmen voor de verwerking van cardiovasculaire beelden gebruik makend van residu-eenheden

GHENT UNIVERSITY

Department of Telecommunications and Information Processing
Faculty of Engineering and Architecture
Ghent University

Department of Software Engineering
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Josip Juraj Strossmayer University of Osijek

Marija Habijan
Optimization of deep learning algorithms for the cardiovascular system image processing using residual units, Joint Doctoral Thesis, © January 2022

DEGREES:
Doctor of Computer Science Engineering

SUPERVISOR [UNIOS]:
Prof. Irena Galić, PhD

SUPERVISOR [UGENT]:
Prof. Aleksandra Pižurica, PhD
Dr. Danilo Babin, PhD

MEMBERS OF THE JURY:
prof. dr. sc. Krešimir Nenadić (University of Osijek, chairman)
prof. dr. sc. Ronny Verhoeven (Ghent University, co-chairman)
prof dr. ir. Jan Aelterman (Ghent University)
prof. dr. ir. Charlotte Debbaut (Ghent University)
prof. dr. sc. Robert Cupec (University of Osijek)
prof. dr. sc. Sonja Grgić (University of Zagreb)
prof. dr. md. Lazar Velicki (University of Novi Sad)

AFFILIATIONS:
Department of Software Engineering (ZPI)
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, J. J. Strossmayer University of Osijek
Department of Telecommunications and Information Processing (TELIN), Faculty of Engineering and Architecture, Ghent University

"*It was a large, loose, pluralistic affair without any clear unifying principle. It encompassed superhuman beings and forces, witches and wise men and a mass of low-grade magical and superstitious practices. The whole was less than the sum of its parts - for it was not a cosmos to be contemplated or worshipped but a treasury of separate and specific resources to be used or applied in concrete situations.*"

James Obelkevich

- This puts it extremely well.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Irena Galić, for allowing me to conduct this research in the first place and for numerous constructive discussions about my work. I would especially like to thank Prof. Galić for her patience, constant support and encouragement. I feel very fortunate to have learned challenging but great lessons from her.
I would also like to express my gratitude to Prof. Aleksandra Pižurica and dr. ir. Danilo Babin who gave me the unique opportunity to pursue research at TELIN and who guided me with their excellent scientific knowledge. I would also like to thank them for numerous constructive discussions about my research topic and for allowing me great freedom to follow my interests.
I would like to thank the president of my Ph.D. Examination Board and all its members for the effort they have put into my Ph.D. examination. Additionally, I would like to thank the Croatian Science Foundation for financially supporting this research through the project ImagineHeart.
I would also like to thank all my FERIT colleagues for creating a nice atmosphere and for making the work environment very pleasant. Great appreciation to my lab mate Marin for all the encouragement and wise words.
Finally, I would like to thank my family and relatives for their love and support. I am grateful to my parents - mother Ivanka and dad Vladimir that left us too early. Special thanks to my sister Jelena for all the laughs and fun when I needed to relax from work.
Last but not least, I would like to thank my friends Ana and Boris for their support and everlasting patience.

Marija Habijan
January, 2022

Abstract

The term cardiovascular disease (CVD) refers to numerous dysfunctions of the heart and circulatory system. Cardiovascular disease accounts for nearly one-third (33%) of all deaths in the modern world, which is the highest proportion of all diseases. Early diagnosis and appropriate treatment can significantly reduce mortality and improve quality of life. The diagnosis of heart disease is based on the complete cardiovascular picture, including anatomy and physiology. The diagnostic process usually consists of two main parts. The first part refers to obtaining images of the heart using imaging devices. Numerous invasive and noninvasive imaging techniques have been developed to characterize the anatomy and functionality of the heart. The second part of the diagnostic process is the quantification and interpretation of the images using advanced image processing methods. Developing efficient medical image processing and analysis methods is a complex task, mainly because it involves processing large amounts of high-dimensional data. Advances in the development of image processing, computer vision, and artificial intelligence, as well as the widespread availability of powerful graphical processing units (GPUs), have made this challenging task manageable.

Medical image segmentation plays an important role in the assessment, diagnosis, and prognosis of various cardiovascular diseases. Extensive research and clinical applications have shown that computed tomography (CT) and magnetic resonance imaging (MRI) play an important role in the noninvasive assessment of cardiovascular disease. They help quantify disease, measure the volume of structures, and analyze organ morphology. Therefore, segmentation of whole heart is an important step for a variety of clinical applications. For example, it is used for modeling and analyzing the anatomy and function of the heart and for localizing pathologies. The creation of a patient-specific 3D heart model holds excellent potential for improving surgical planning for patients with congenital heart defects. It requires delineation of all cardiac structures, including heart chambers, epicardial surface, entire blood pool, and great vessels. Segmentation of the left and right ventricles plays a critical role in quantitative analysis of global and regional information, i.e., indicators of cardiac function, such as end-diastolic volume (EDV), end-systolic volume (ESV), ejection fraction (EF), wall thickness, and mass. For example, ventricular hypertrophy

is caused by abnormal enlargement of the myocardium surrounding the left or right ventricle. Therefore, segmentation of the whole heart and heart chambers from volumetric medical images plays an essential role in cardiac assessment. In addition, radiologists often need to delineate the aorta to obtain its morphology, which is essential for the detection and diagnosis of aortic aneurysms. Manual segmentation of cardiac structures is a time-consuming process that depends on observer variability. Therefore, the development of accurate and robust automatic segmentation algorithms is critical for clinical practice.

Deep learning has emerged as a state-of-the-art method for various image processing tasks such as recognition, segmentation, and classification. Deep learning methods are based on deep artificial neural networks. The most common type of deep neural network is convolutional neural networks (CNNs). Fully convolutional neural networks (FCNs) are a special type of CNNs that do not have a fully connected layer and are trained and applied to the entire image so that no patch selection is required. Several variants of FCNs have been proposed to transfer features from the encoder to the decoder to increase segmentation accuracy. The most widely used FCNs for biomedical image segmentation are the U-net architecture and its corresponding three-dimensional counterpart, the 3D U-net architecture. The ability of U-Net architecture to capture low-level features makes them very useful in scenarios with a small amount of training data. Although it has strong representational power, long-range relationships are weak due to the inherent localization of convolutional operations, so more advanced mechanisms and building blocks are required. Techniques and building blocks such as residual connections and deep supervision enable the construction of deeper architectures that provide more abstract learning results and higher accuracy for medical segmentation tasks. The increment in the number of layers provides larger parameter space enabling learning of more abstract features. Therefore, deeper architectures could provide more abstract learning that results in better performance and higher accuracy in medical segmentation tasks. Nevertheless, when the depth of CNN increases, information about the gradient passes through many layers, and it can vanish or accumulate large errors by the time it reaches the end of the network. This leads to common obstacles of training deep neural network architectures such as appearance of vanishing gradients, accuracy degradation, and extensive parameter growth, which results in computationally intensive models.

In this Thesis, we propose a set of deep learning methods for automatic heart and heart chambers segmentation. We focus on improving deep learning segmentation methods for the whole heart, both ventricles, myocardium, and abdominal aortic aneurysm. Several unique challenges and issues arise in developing deep learning methods for medical image segmentation and analysis. For example, the high image dimensionality leads to trained models with a high number of

parameters, and the lack of expert annotation makes the models more susceptible to overfitting. Therefore, we aim to alleviate these challenges by proposing new and robust CNNs that reduce the number of parameters so that they can be trained with smaller training sets and are less prone to overfitting.

One of the most important scientific contributions of this work is the novel connectivity structure of residual units, which we call the feature merge residual unit (FM-Pre-ResNet). The FM-Pre-ResNet unit attaches two convolution layers at the top and at the bottom of the pre-activation residual block. The top layer balances the parameters of the two branches, while the bottom layer reduces the channel dimension. The proposed connectivity allows the construction of notably deeper models while maintaining the same or smaller number of parameters than the pre-activation residual units.

Following that, the second scientific contribution is a novel three-dimensional (3D) encoder-decoder architecture that successfully integrates FM-Pre-ResNet units and is additionally guided with variational autoencoders (VAE) for the task of whole heart segmentation from CT and MRI images. The architecture includes three stages. First, in an encoding stage, FM-Pre-ResNet units learn a low-dimensional representation of the input. Second, in the VAE stage, an input image is reduced to a low-dimensional latent space and reconstructs itself to provide a strong regularization of all model weights. This ensures that all model weights are strongly regularized while avoiding overfitting the training data. Third, the decoding stage creates the final whole heart segmentation. We evaluate our method on the 40 test subjects of the MICCAI Multi-Modality Whole Heart Segmentation (MM-WHS) Challenge. Our method achieves an average Dice score (DSC), Jaccard index (JI), surface distance (SD), and Hausdorff distance (HD) for WHS of 90.39%, 82.24%, 1.1093, and 15.3621 on CT images and 89.50%, 80.44%, 1.8599, 25.6558 on MRI images, respectively. The proposed approach obtains highly comparable DSC to the state-of-the-art for whole heart segmentation tasks on CT images while outperforming the current state-of-the-art on the MRI images.

The third scientific contribution is a new automatic method for left ventricle (LV), right ventricle (RV), and myocardium (Myo) segmentation and quantification from cine-MRI images. We introduce a new architecture that incorporates SERes blocks into 3D U-net architecture (3D SERes-U-Net). The SERes blocks incorporate squeeze-and-excitation operations into residual learning. The adaptive feature recalibration ability of squeeze-and-excitation operations boosts the network's representational power while feature reuse utilizes effective feature learning, which improves segmentation performance. We evaluate the proposed method on the MICCAI Automated Cardiac Diagnosis Challenge (ACDC) testing dataset. Our method obtains an average DSC for LV, RV, and Myo at end-diastole of 95%, 90%, 83%, respectively. Similarly, we obtain an average DSC for LV, RV, and

Myo at end-systole of 86%, 83%, 85%, respectively. Additionally, we calculate significant clinical metrics, i.e., indicators of hearts' function, including volume of the left ventricle at end-diastole (LVEDV), the volume of the left ventricle at end-systole (LVESV), left ventricles' ejection fraction (LVEF), the volume of the right ventricle at end-diastole (RVEDV), volume of the right ventricle at end-systole (RVESV), right ventricles' ejection fraction (RVEF), myocardium volume at end-systole (MyoLVES), and myocardium mass at end-diastole (MyoMED). The Bland-Altman analysis shows a high correlation coefficient of R=0.99 for LVEDV and LVESV, while R=0.95 for LVEF. Correlations of RVEDV, EVESV and RVEF are R=0.97, R=0.93, R=0.69, respectively. Finally, R=0.96 for MyoLVES and R=0.95 for MyoMED further show our proposed methods' strength of accuracy and precision.

Finally, the fourth scientific contribution includes a new automatic approach for robust and reproducible abdominal aortic aneurysm (AAA) segmentation. The 3D U-Net network is adapted by introducing residual units in the contracting pathway and a deep supervision mechanism in the expanding pathway. We conduct an ablation study to demonstrate the effect of the addition of residual units and deep supervision for this particular clinical application. To increase the robustness of the results, networks are trained, validated, and evaluated on 19 pre-operative CTA volumes from different patients using a 4-fold cross-validation approach. Our pipeline achieves a Dice score of 91.03% for AAA segmentation.

The work conducted during this Thesis resulted in 5 journal publications (of which 3 as the first author), 10 papers are published at international conferences (of which 5 as the first author), and 1 publication in book chapters (as co-author).

Samenvatting

Hart en vaatziekten (HVZ) verwijst naar talrijke functionele afwijkingen van het hart en de bloedsomloop. Hart en vaatziekten zijn verantwoordelijk voor bijna een derde (33%) van alle sterfgevallen in de moderne wereld, het hoogste percentage van alle ziekten. Vroege diagnose en passende behandeling kunnen de mortaliteit aanzienlijk verminderen en de kwaliteit van het leven verbeteren. De diagnose van hartziekten is gebaseerd op het volledige cardiovasculaire beeld, inclusief anatomie en fysiologie. Het diagnostisch proces bestaat meestal uit twee hoofdonderdelen. Het eerste deel verwijst naar het verkrijgen van beelden van de hartstructuur met behulp van beeldvormingsapparatuur. Er zijn talloze invasieve en niet-invasieve beeldvormingstechnieken ontwikkeld om de anatomie en functionaliteit van het hart te karakteriseren. Het tweede deel van het diagnostisch proces is het kwantificeren en interpreteren van de beelden met behulp van geavanceerde beeldverwerkingsmethoden. Het ontwikkelen van efficiënte medische beeldverwerkings- en analysemethoden is een complexe taak, vooral omdat het gaat om het verwerken van grote hoeveelheden hoogdimensionale gegevens. Vooruitgang in de ontwikkeling van beeldverwerking, computervisie en kunstmatige intelligentie, evenals de wijdverbreide beschikbaarheid van krachtige grafische verwerkingseenheden (GPU's), hebben deze uitdagende taak haalbaar gemaakt.

Medische beeldsegmentatie speelt een belangrijke rol bij de beoordeling, diagnose en prognose van verschillende hart- en vaatziekten. Uitgebreid onderzoek en klinische toepassingen hebben aangetoond dat computertomografie (CT) en magnetische resonantiebeeldvorming (MRI) een belangrijke rol spelen bij de niet-invasieve beoordeling van hart- en vaatziekten. Ze helpen bij het kwantificeren van ziekten, het meten van het volume van structuren en het analyseren van de morfologie van organen. Segmentatie van afbeeldingen van het volledige hart is dus een belangrijke stap voor een breed scala aan klinische toepassingen. Het wordt bijvoorbeeld gebruikt voor het modelleren en analyseren van hartanatomie en functie- en pathologielokalisatie. De creatie van een patiëntspecifiek 3D-hartmodel heeft uitstekende mogelijkheden voor het verbeteren van de chirurgische planning voor patiënten met aangeboren hartafwijkingen. Het vereist afbakening van alle hartstructuren, inclusief hartkamers, het epicardiaal oppervlak, de volledige bloedplas en de grote bloedvaten. Segmentatie van de linker en rechterventrikels

speelt een cruciale rol bij de kwantitatieve analyse van de globale en regionale informatie, d.w.z. indicatoren van de hartfunctie, zoals einddiastolisch volume (EDV), eindsystolisch volume (ESV), ejectiefractie (EF), wanddikte en massa. Ventriculaire hypertrofie wordt bijvoorbeeld veroorzaakt door een abnormale vergroting van de hartspier rond de linker- of rechterventrikel. Segmentatie van het hele hart en de hartkamers van volumetrische medische beelden speelt dus een essentiële rol bij cardiale beoordeling. Bovendien moeten radiologen de aorta afbakenen om zijn morfologie te verkrijgen, wat essentieel is voor het detecteren en diagnosticeren van aorta-aneurysma's. Handmatige segmentatie van hartstructuren is een tijdrovend proces, vatbaar voor variabiliteit van waarnemers. De ontwikkeling van nauwkeurige en robuuste automatische segmentatie-algoritmen is daarom cruciaal voor de klinische praktijk.

Deep learning is naar voren gekomen als een state-of-the-art methode voor verschillende beeldverwerkingstaken zoals herkenning, segmentatie en classificatie. Deep learning-modellen zijn gebaseerd op diepe kunstmatige neurale netwerken. Het meest voorkomende type van diepe neurale netwerken zijn de convolutionele neurale netwerken (CNN's). Volledig convolutionele neurale netwerken (FCN's) zijn een speciaal type CNN's die geen volledig verbonden laag hebben en worden getraind en toegepast op het hele beeld, zodat er geen patch-selectie vereist is. Er zijn verschillende varianten van FCN's voorgesteld om functies van de encoder naar de decoder over te dragen om de nauwkeurigheid van de segmentatie te vergroten. De meest gebruikte FCN's voor biomedische beeldsegmentatie zijn het U-net en de bijbehorende driedimensionale tegenhanger, het 3D U-net. Het vermogen van U-net om functies op laag niveau vast te leggen, maakt ze erg handig in scenario's met een kleine hoeveelheid trainingsgegevens. Hoewel het een sterke representatiekracht heeft, zijn langetermijnrelaties zwak vanwege de inherente lokalisatie van convolutionele operaties, dus zijn meer geavanceerde mechanismen en bouwstenen vereist. Technieken en bouwstenen zoals restverbindingen en deep supervision maken de constructie van diepere architecturen mogelijk die meer abstracte leerresultaten en een hogere nauwkeurigheid voor medische segmentatietaken opleveren. De toename in het aantal lagen zorgt voor een grotere parameterruimte waardoor het mogelijk wordt om meer abstracte functies te leren. Daarom kunnen diepere netwerk architecturen abstracter leren, wat resulteert in betere prestaties en hogere nauwkeurigheid bij medische segmentatietaken. Niettemin, wanneer de diepte van CNN toeneemt, gaat informatie over de gradiënt door vele lagen, en het kan verdwijnen of grote fouten ophopen tegen dat het het einde van het netwerk bereikt. Dit leunt op veelvoorkomende obstakels bij het trainen van diepere neurale netwerkarchitecturen, zoals het verschijnen van vanishing gradients, verslechtering van de nauwkeurigheid en extensieve groei van parameters, wat leidt tot rekenintensieve modellen.

In dit proefschrift stellen we een reeks diepgaande leermethoden

voor voor automatische hart- en hartkamerssegmentatie. We richten ons op het verbeteren van deep learning-segmentatiemethoden voor segmentatie van het volledige hart, bi-ventrikels en de myocardiumsegmentatie en kwantificering, evenals segmentatie van aneurysma's van de abdominale aorta. Bij het ontwerpen van diepgaande leermethoden voor medische beeldanalyse doen zich enkele unieke uitdagingen en problemen voor. Een hoge beelddimensionaliteit resulteert bijvoorbeeld in getrainde modellen met een groot aantal parameters, en het gebrek aan deskundige annotaties maakt modellen vatbaarder voor overfitting. Daarom willen we deze uitdagingen verlichten door nieuwe en robuuste netwerken voor te stellen die het aantal parameters verminderen, waardoor deze getraind kunnen worden met kleinere trainingssets en deze minder vatbaar zijn voor overfitting.

Een van de belangrijkste wetenschappelijke bijdragen van dit werk is de nieuwe connectiviteitsstructuur van residuele eenheden, waarnaar we verwijzen als een feature merge residual unit (FM-Pre-ResNet). De FM-Pre-ResNet-eenheid bevestigt twee convolutielagen aan de boven- en onderkant van het pre-activatie residueel blok . De bovenste laag balanceert de parameters van de twee takken, terwijl de onderste laag de dimensies van het kanaal verkleint. De voorgestelde connectiviteit maakt de constructie van met name diepe modellen mogelijk met behoud van hetzelfde of een kleiner aantal parameters als bij de pre-activatie residuele eenheden.

Daarna tweede wetenschappelijke bijdrage is een nieuwe driedimensionale (3D) encoder-decoder-architectuur die met succes FM-Pre-ResNet-eenheden integreert en die bovendien wordt begeleid met variabele autoencoders (VAE) voor de taak van segmentatie van het volledige hart van CT en MRI afbeeldingen. De architectuur omvat drie fasen. Ten eerste leren FM-Pre-ResNet-eenheden in een coderingsfase een laagdimensionale weergave van de invoer. Ten tweede wordt in de VAE-fase een invoerbeeld gereduceerd tot een laagdimensionale latente ruimte en reconstrueert het zichzelf wat leidt tot een sterke regularisatie van de gewichten van het model. Dit zorgt ervoor dat alle modelgewichten sterk worden geregulariseerd, terwijl overfitting van de trainingsgegevens wordt vermeden. Ten derde creëert de decoderingsfase de uiteindelijke segmentatie van het volledige hart. We evalueren onze methode op de 40 proefpersonen van de MICCAI Multi-Modality Whole Heart Segmentation (MM-WHS) Challenge. Onze methode behaalt een gemiddelde Dice-score (DSC), Jaccard-index (JI), oppervlakteafstand (SD) en Hausdorff-afstand (HD) voor WHS van respectievelijk 90,39%, 82,24%, 1,1093 en 15,3621 op CT-beelden en 89,50%, 80,44%, 1,8599, 25,6558 MRI-beelden. De resulterende netwerkarchitectuur bereikte state-of-the-art resultaten met hoge nauwkeurigheid zonder te vertrouwen op trial-and-error architectuurontwerpmethodologieën of nauwgezette monitoring van hyperparameterveranderingen. De voorgestelde aanpak verkrijgt zeer vergelijkbare Dice-scores met de state-of-the-art voor segmentatietaken van het volledige hart op

CT-beelden, terwijl het beter presteert dan de huidige state-of-the-art op de MRI-beelden.

De derde wetenschappelijke bijdrage is een nieuwe automatische methode voor segmentatie van de linkerventrikel (LV), de rechterventrikel (RV) en het myocard (Myo) van MRI-beelden. We introduceren een nieuwe architectuur die SERes-blokken opneemt in de 3D U-net-architectuur (3D SERes-UNet). De SERes-blokken nemen squeeze-and-excitation operaties op in residueel leren. Het adaptieve herkalibreringsvermogen van squeeze-and-excitation bewerkingen verhoogt de representatiekracht van het netwerk, terwijl het hergebruik van functies gebruikmaakt van effectief leren van de functies, wat de segmentatieprestaties verbetert. We evalueren de voorgestelde methode op de MICCAI Automated Cardiac Diagnosis Challenge (ACDC) testdataset. Onze pijplijn behaalt een gemiddelde DSC voor LV, RV en Myo bij einddiastole van respectievelijk 95%, 90%, 83%. Evenzo verkrijgen we een gemiddelde DSC voor LV, RV en Myo bij eindsystole van respectievelijk 86%, 83%, 85%. Daarnaast berekenen we significante klinische meetwaarden, dat wil zeggen indicatoren van de hartfunctie, inclusief het volume van de linker hartkamer bij de einddiastole (LVEDV), het volume van de linker hartkamer bij de eindsystole (LVESV), de ejectiefractie van de linker hartkamer (LVEF), het volume van de rechter ventrikel aan de eind-diastole (RVEDV), het volume van de rechter ventrikel aan de eind-systole (RVESV), de ejectiefractie van de rechter ventrikel (RVEF), het myocardvolume aan de eind-systole (MyoLVES), en myocardmassa bij einddiastole (MyoMED). De Bland-Altman en analyse tonen een hoge correlatiecoëfficiënt van R=0,99 voor LVEDV en LVESV, met R=0,95 voor LVEF. Correlaties van RVEDV, EVESV en RVEF zijn respectievelijk R=0,97, R=0,93, R=0,69. Ten slotte tonen R = 0,96 voor MyoLVES en R = 0,95 voor MyoMED verder de nauwkeurigheid en precisie van onze voorgestelde pijplijn.

Ten slotte omvat de vierde wetenschappelijke bijdrage een nieuwe automatische benadering voor robuuste en reproduceerbare segmentatie van abdominaal aorta-aneurysma (AAA). Het 3D U-Net segmentatienetwerk is aangepast door het introduceren van resteenheden in het contracterende pad en een diepgaand supervisiemechanisme in het uitbreidende pad. We voeren een ablatie-onderzoek uit om het effect van de toevoeging van resteenheden en diepgaande supervisie voor deze specifieke klinische toepassing aan te tonen. Om de robuustheid van de resultaten te vergroten, worden netwerken getraind, gevalideerd en geëvalueerd op 19 pre-operatieve CTA-volumes van verschillende patiënten met behulp van een 4-voudige cross-validation benadering. Onze pijplijn behaalt een Dice-score van 91,03% voor preoperatieve aneurysmasegmentatie.

Het werk tijdens dit proefschrift resulteerde in 5 journaalpublicaties (waarvan 3 als eerste auteur), 10 papers gepubliceerd op internationale conferenties (waarvan 5 als eerste auteur), en 1 publicatie in boekhoofdstukken (als co-auteur).

Sažetak

Izraz kardiovaskularne bolesti (KVB) odnosi se na brojne funkcionalne abnormalnosti srca i krvožilnog sustava. KVB uzrokuju gotovo jednu trećinu (33%) smrtnosti u suvremenom svijetu, što predstavlja najveći udio u odnosu na sve druge bolesti. Rana dijagnoza i odgovarajuće liječenje kardiovaskularnih bolesti mogu značajno smanjiti smrtnost i poboljšati kvalitetu pacijentova života. Postavljanje dijagnoze temelji se na cjelokupnoj slici kardiovaskularnog sustava, uključujući anatomiju i fiziologiju srca. Dijagnostički proces obično se sastoji od dva glavna dijela. Prvi dio odnosi se na prikuplanje slika srca pomoću medicinskih uređaja. Razvijene su brojne invazivne i neinvazivne tehnike medicinskog snimanja koje omogućuju uvid u anatomiju i funkcionalnost srca. Drugi dio dijagnostičkog procesa je kvantifikacija i interpretacija prethodno dobivenih slika pomoću naprednih metoda obrade slike. Razvoj učinkovitih metoda za obradu medicinskih slika je složen zadatak, s obzirom da podrazumijeva obradu ogromne količine visokodimenzionalnih podataka. Napredak u razvoju algoritama obrade slike, računalnog vida i umjetne inteligencije, kao i dostupnost grafičkih procesorskih jedinica (GPU-a), značajno su olakšale i ubrzale razvoj takvih metoda.

Segmentacija medicinskih slika ima važnu ulogu u procjeni, dijagnozi te postavljanju prognoze različitih kardiovaskularnih bolesti. Opsežna istraživanja i kliničke primjene pokazale su da računalna tomografija (CT) i magnetska rezonanca (MRI), kao osnovne tehnike prikupljanja medicinskih slika, imaju izrazito važnu ulogu u procjeni kardiovaskularnih bolesti. Njima je omogućeno kvantificiranje bolesti, mjerenje volumena kao i analizira morfologije različitih organa. Prema tome, segmentaciju srca i srčanih struktura predstavlja osnovu za širok spektar kliničkih primjena. Primjerice, često se koristi se za modeliranje i analizu anatomije i funkcionalnosti kao i za lokalizaciju različitih patologija. Izrada trodimenzionalnog (3D) modela srca specifičnog za pojedinog pacijenta predstavlja izrazit potencijal za poboljšanje kirurškog planiranja za pacijente s urođenom srčanom manom. Kako bi se takvi 3D modeli mogli izraditi, potrebno je imati segmentirane različite srčane strukture, uključujući pojedine srčane komore, epikardijalnu površinu, aortu kao i pojedine žile kardiovaskularnog sustava. Segmentacija lijeve i desne klijetke ima izrazito važnu ulogu u kvantitativnoj analizi globalnih i regionalnih informacija, odnosno pokazatelja rada srca, poput

volumena na kraju dijastole (VKD), volumena na kraju sistole (VKS), frakcije izbacivanja (FI), debljine stijenke ili mase. Primjerice, ventrikularna hipertrofija uzrokovana je abnormalnim povećanjem srčanog mišića koji okružuje lijevu ili desnu klijetku. Prema tome, segmentacija cijelog srca i srčanih komora iz volumetrijskih medicinskih slika igraju bitnu ulogu u procjeni cjelokupnog kardiovaskularnog zdravlja. Nadalje, radiolozi često trebaju ocrtati aortu kako bi dobili njezinu morfologiju, što je bitno za otkrivanje i dijagnosticiranje aneurizme aorte. Ručna segmentacija srca i srčanih struktura je vremenski veoma zahtijevan posao, podložan subjektivnosti. Prema tome, razvoj točnih i robusnih automatskih algoritama za segmentaciju je neophodan za primjenu u kliničkoj praksi.

Duboko učenje predstavlja najsuvremeniju metodu za različite zadatke obrade slike poput raspoznavanja, segmentacije i klasifikacije. Metode dubokog učenja temelje se na umjetnim neuronskim mrežama. Najčešće upotrebljena vrsta neuronske mreže su konvolucijske neuronske mreže (CNN). FCNs predstavljaju specifičnu vrstu CNN-a bez potpuno povezanog sloja, kojima se obrađuje cijela slika te nije potrebno korištenje patcheva. Razvijene su različite varijante FCN-a, od kojih su najznačajnije varijante koje koriste koder-dekoder arhitekture. U biomedicinskoj obradi slika, za segmentaciju, najčešće se koristi U-Net arhitektura neruonske mreže kao i njezina odgovarajuća 3D verzija. U-Net arhitektura ima snažnu reprezentativnu snagu te je u mogućnosti zabilježiti značajke niskih razina što je izrazito važno prilikom treniranja mreže sa malom količinom podataka. Iako U-Net ima snažnu reprezentativnu snagu, dugoročni odnosi između značajki su slabi zbog upotrebe konvolucijskih operacija. Prema tome, potrebno je razvijati naprednije mehanizme kao i dodatne blokove koji će biti u mogućnosti ispraviti nedostatke U-Net arhitekture. Tehnike i blokovi poput veza za preskakivanje ili dubokog nadzora, omogućuju izgradnju dubljih arhitektura neuronskih mreža koje pružaju apstraktnije rezultate učenja te postižu veću točnost prilikom segmentacije medicinskih slika. S obzirom da povećanje broja slojeva osigurava veći prostor parametara koji omogućuje učenje apstraktnijih značajki, dublje arhitekture neuronskih mreža pružaju apstraktnije učenje koje rezultira boljim performanse i većom točnost u zadacima medicinske segmentacije. Unatoč tome, kako se dubina mreže povećava, informacije o gradijentu prolaze kroz mnogo slojeva te mogu nestati ili nakupiti velike pogreške do trenutka kada gradijet dosegne kraj mreže. To dovodi do uobičajenih prepreka treninga dubokih arhitektura neuronskih mreža kao što su problem nestajajućih gradijenta, ekstenzivnog rasta parametara, kao i smanjenja točnosti, što dovodi do računalno zahtjevnih modela.

U ovoj doktorskoj disertaciji, predložen je niz metoda dubokog učenja za automatsku segmentaciju srca i srčanih komora. Fokus disertacije je na poboljšanju metoda dubokog učenja za segmentaciju cijeloga srca, lijeve i desne klijetke i miokarda kao i aneurizme abdominalne aorte. S

obzirom na karakteristične probleme koji se javljanju prilikom dizajniranja metoda dubokog učenja za segmentaciju medicinskih slika, poput problema visoke dimenzionalnosti slika koje rezultiraju treniranim modelima s velikim brojem parametara kao i nedostatkom anotiranih podataka za treniranje, cilj ove disertacije je ublažiti navedene izazove predlaganjem novih i robusnih arhitektura neuronskih mreža koje smanjuju broj korištenih parametara, ali zadržavaju izrazito visoku točnost krajnjih rezultata segmentacije.

Prvi i najvažniji znanstveni doprinos predstavlja nova struktura povezivanja rezidualnih jedinica, koju nazivamo rezidualna jedinica za spajanje značajki (FM-Pre-ResNet). FM-Pre-ResNet struktura povezivanja rezidualnih jedinica dodaje konvolucijski sloj na vrh i na dno već postojećih prethodno aktivirajućih rezidualnih jedinica. Pri tome, gornji sloj uravnotežuje parametre dviju grana rezidualne jedinice, dok donji sloj smanjuje dimenzije kanala. Na ovaj način predložena struktura povezivanja rezidualnih jedinica omogućuje kreiranje značajno dubljih modela uz održavanje iste ili čak manje količine parametara u odnosu na originale rezidualne jedinice.

Nakon toga, u drugom znanstvenom doprinosu, predložena je nova 3D arhitektura neuronske mreže bazirana na koder-dekoder arhitekturi koja uspješno integrira FM-Pre-ResNet jedinice s varijacijskim autokoderima (VAE) za segmentaciju srca i srčanih komora iz CT i MRI slika. Metoda se sastoji od tri osnovna dijela. U prvom dijelu, prethodno predložene FM-Pre-ResNet jedinice koriste se za učenje nisko-dimenzionalnog prikaza ulaza u fazi kodiranja. U drugom dijelu, VAE rekonstruira ulaznu sliku iz nisko-dimenzionalnog latentnog prostora, osiguravajući da su sve težine modela snažno regulirane, kako bi se izbjegnula neželjena pojava pretreniranja. VAE dio koristi se samo tijekom treniranja mreže. Konačno, u trećoj fazi dekodiranja ponovno su integrirane FM-Pre-ResNet jedinice pomoću kojih se stvaraju konačne segmentacije srca. Predložena nova arhitektura evaluirana je na testnom skupu podataka koji se sastoji od 40 različitih pacijenata dostupnih kroz MICCAI Multi-Modality Whole Segmentation Challenge (MM-WHS) izazov. Naša metoda ostvarila je prosječni DSC, JI, SD i HD za cijelo srce od 90,39%, 82,24%, 1.1093 i 15,3621 na CT snimkama, odnosno 89,50%, 80,44%, 1,8599, 25,6558 na MRI snimkama. Predloženi pristup ostvario je približno slične rezultate kao i najsuvremenije metode za segmentaciju cijelog srca na CT slikama dok su rezultati na MRI slikama bolji od rezultata prethodno objavljenih najsuvremenijih metoda.

Treći znanstveni doprinos, predstavlja novu automatsku metodu za segmentaciju miokarda (MiO), lijeve (LK) i desne klijetke (DK) iz cineMRI slika. Predstavljena je nova arhitekturu koja integrira SERes blokove u 3D U-net arhitekturu (3D SERes-U-Net). SERes blokovi upotrebljavaju operacije stiskanja i uzbude u rezidualne jedinice. Sposobnost ponovne kalibracije značajki operacija stiskanja i uzbude povećava reprezentativnu snagu mreže, dok ponovna upotreba značajki

koristi učinkovito učenje o značajkama, što poboljšava performanse segmentacije. Predloženu metodu evaluirali smo na testnom skupu podataka MICCAI Automated Cardiac Diagnosis Challenge (ACDC). Naša predložena metoda za segmentaciju pomoću 3D SERes-U-Net ostvarila je prosječni DSC za LK, DK i MiO na kraju dijastole od 95%, 90%, 83%. Slično, prosječni DSC za LK, DK i MiO na kraju sistole je 86%, 83%, 85%. Dodatno, izračunati su volumeni LK, DK i MiO na temelju kojih su dalje računate značajne kliničke metrike te su uspoređeni rezultati s referentnim rezultatima. Navedeno uključuje kliničke metrike, odnosno pokazatelje funkcionalnosti srca, uključujući volumen lijeve klijetke na kraju dijastole (VLKKD), volumen lijeve klijetke na kraju sistole (VLKKS), frakciju izbacivanja lijeve klijetke (FILK), volumen desne klijetke na kraju dijastole (VDKKD), volumen desne klijetke na krajnjoj sistoli (VDKKS), frakciju izbacivanja desne klijetke (FIDK), volumen miokarda na krajnjoj sistoli (VMiOKS) kao i masu miokarda na kraju dijastole (MiOKD). Bland-Altman analiza pokazuje visoki koeficijent korelacije od R = 0,99 za VLKKD i VLKKD, dok je R = 0,95 za FILK. Korelacije VDKKD, VDKKS i FIDK su R = 0,97, R = 0,93, R = 0,69. Konačno, R = 0,96 za VMiOKS i R = 0,95 za MiOKD dodatno pokazuju snagu točnosti i preciznosti naše predložene metode.

Konačno, četvrti znanstveni doprinos predstavlja novi automatski pristup za segmentaciju aneurizme abdominalne aorte (AAA). 3D U-Net arhitektura modificirana je uvođenjem rezidualnih jedinica u koder dijelu kao i mehanizmom dubokog nadzora u dekoder dijelu. Kako bi se povećala točnost rezultata, mreža je trenirana i validirana na 19 preoperativnih AAA CTA volumena različitih pacijenata primjenom 4-ostrukog pristupa unakrsne provjere valjanosti. Naša metoda postiže DSC rezultat od 91,03% za segmentaciju aneurizme abdominalne aorte.

Tijekom rada na ovoj doktorskoj disertaciji, objavljeno je 5 radova u časopisima (od čega 3 kao prvi autor), 10 radova objavljeno je na međunarodnim konferencijama (od čega 5 kao prvi autor) te 1 rad kao dio knjige (ko-autor).

Contents

List of Figures

List of Tables

List of Abbreviations

AAA	Abdominal Aortic Aneurysm
APV	Anomalous Pulmonary Venous
AAH	Aortic Arch Hypoplasia
Ao	Aortic Valve
AS	Aortic Stenosis
APW	Aortopulmonary Window
ARVCD	Arrhythmogenic Right Ventricular Dysplasia
ASD	Atrial Septal Defect
AVSD	Atrioventricular Junction
ACDC	Automated Cardiac Diagnosis Challenge
CO	Cardiac Output
CVDs	Cardiovascular Diseases
CoA	Coarctation of the Aorta
CAT	Common Arterial Trunk
CT	Computed Tomography
CHD	Congenital Heart Disease
CAEs	Contractive Autoencoders
CNNs	Convolutional Neural Networks
CTS	Cor Triatriatum Sinister
CTA	Coronary CT Angiography
DBN	Deep Belief Network
DVT	Deep Vein Thrombosis
DSC	Dice Similarity Coefficient
DCM	Dilated Cardiomyopathy
DRN	Dilated Residual Network
DORV	Double Outlet Right Ventricle
DSVC	Double Superior Vena Cava
EVAR	Endovascular Aneurysm Repair
ELBO	Evidence Lower Bound
FM-Pre-ResNets	Feature Merge Pre-activation Residual Units
FPN	Feature Pyramid Network
FFN	Feed Forward Network
FCN	Fully Convolutional Neural Network
GANs	Generative Adversarial Networks
GPUs	Graphics Processing Units
TGA	Great Artery Transposition
HD	Hausdorff Distance

HR	Heart Rate
HCM	Hypertrophic Cardiomyopathy
IDC	Idiopathic Dilated Cardiomyopathy
IHD	Ischemic Heart Disease
JI	Jaccard Index
KL	Kullback-Leibler
KL	Late Gadolinium Enhancement
LA	Left Atrium
LV	Left Ventricle
LVEF	Left Ventricle ejection fraction
LVEDV	Left Ventricle volume at end-diastole
LVESV	Left Ventricle volume at end-systole
LVH	Left Ventricular hypertrophy
LVM	Left Ventricular mass
LVWT	Left Ventricular wall thickness
LAX	Long Axis
MRI	Magnetic Resonance
MLE	Maximum Likelihood Estimate
MSE	Mean Squared Error
MC-dropout	Monte Carlo dropout
mDSC	Multi-Class Dice Similarity Coefficient
MLP	Multi-Layer Perceptron
MMWHS	Multi-Modality Whole Heart Segmentation
MO-MP-CNN	Multi-Object Multi-Planar CNN
MVL	Multi-View Learning
Myo	Myocardium
MyoMED	Myocardium Mass at End-Diastole
MyoVES	Myocardium Volume at End-Systole
PDA	Patent Ductus Arteriosus
PET	Positron Emission Tomography
PCA	Principal Component Analysis
PAS	Pulmonary Artery Sling
PV	Pulmonary Valve
PVS	Pulmonary Valve Stenosis
ReLU	Rectified Linear Unit
ROI	Region of Interest
RPN	Region Proposal Network
ResNets	Residual Networks
RCM	Restrictive Cardiomyopathy
RA	Right Atrium
RV	Right Ventricle
RVEF	Right Ventricle ejection fraction
RVEDV	Right Ventricle volume at end-diastole
RVESV	Right Ventricle volume at end-systole
RVCO	Right Ventricular cardiac output
RVH	Right Ventricular hypertrophy
RVOT	Right Ventricular outflow tract

SAX	Short Axis
SPECT	Single Photon Emission Computed Tomography
SA node	Sinoatrial Node
SD	Surface Distance
TOF	Tetralogy of Fallot
3D	Three Dimensional
US	Ultrasound
VAE	Variational Autoencoder
VSD	Ventricular Septal Defect
WRN	Weighted Residual Networks

CHAPTER 1

Introduction

Cardiovascular diseases (CVDs) cause significant health complications. They are responsible for more than 17.9 million deaths per year, making them the leading cause of death worldwide [32]. Automatic diagnosis and treatment of cardiovascular diseases have improved thanks to advances in cardiovascular imaging technologies, mathematical algorithms for medical image processing and widely available graphics processing units (GPUs). Traditional semi-automatic segmentation algorithms are limited in their ability to capture image information, especially for low-quality images. With the rise of artificial intelligence, deep learning in image processing has attracted tremendous attention. Its high efficiency in automatic feature extraction and learning ability makes it very accurate for image segmentation. However, using deep learning for medical image analysis has its challenges and limitations. First, the high dimensionality of medical images requires a vast number of parameters in convolutional neural networks (CNNs), which leads to a substantial computational resource requirement that is not affordable with current hardware technology. The second challenge is to develop an algorithm to find the optimal CNN architecture. Third, optimizers need to be studied as the engine of deep learning methods and optimized for medical image segmentation. Therefore, the high complexity of deep learning models requires continuous improvements, such as reducing the number of network parameters and the training time of the model.

This chapter provides a general introduction to the Thesis. The outline of the chapter is as follows. Section 1.1 provides the motivation for this research. Section 1.2 summarizes the main objectives to be achieved by this Thesis. Section 1.3 summarizes the major contributions of the Thesis. Section 1.4 lists the publications produced during the work on this Thesis. Finally, section 1.5 gives an overview of the structure of the Thesis.

1.1 Motivation

High-resolution three-dimensional (3D) images achieved by recent advances in medical imaging technology have enabled the creation of accurate patient-specific models of individual heart structures and the calculation of related quantitative measurements. This has led to the development of a wide range of new applications for computer-aided diagnosis, intervention and follow-up of cardiovascular disease.

Image-based analysis of heart and heart structures provides valuable information for planning and navigation during surgical procedures. Segmentation is a crucial pre-processing step in medical image analysis, as it enables the acquisition of relevant quantitative data for medical interpretation. In this processing stage, the pixels of an image are divided into groups corresponding to the objects on the image. It has enabled the non-invasive acquisition of important information about the anatomy of the target structures. For example, segmentation of the whole heart is critical for pathology localization, anatomy and functional analysis [83]. The construction of patient-specific 3D heart models and surgical implants is of great benefit for preoperative planning of patients with atherosclerosis, congenital heart disease (CHD), cardiomyopathy, or even for the study of various cardiac infections during postoperative treatment [77, 167].

More specifically, by fusing the 3D surface of the heart created from anatomical and real-time images, geometric information about the whole heart can be used to guide interventional procedures. In addition, clear delineations of the myocardium (Myo), left ventricle (LV) and right ventricle (RV) are required for quantitative assessment and calculation of clinical indicators such as volume measurements at end-systole and end-diastole, ejection fraction, thickening measurements and chamber mass. Most of the markers described above are required for determination of cardiac contractile function [80, 52]. Extraction of these features is required for detection and prevention of myocardial infarction and for diagnosis of ischemic heart disease (IHD) [137] and hypertrophic cardiomyopathy (HCM) [31, 138]. For instance, a decrease in the ejection fraction impairs the LVs' ability to pump. Patients with dilated cardiomyopathy have an ejection fraction less than 40%, a left ventricular volume greater than 100 $mL/m2$ and a diastolic wall thickness less than 12mm. Certain patients with dilated LV also have a dilated RV or a large LV mass due to dilated LV. Ventricular hypertrophy may develop in chronic lung disease, congenital heart defect with a left-to-right shunt (patent ductus arteriosus or ventricular septal defect), hypoxia at high altitude, or idiopathic pulmonary hypertension.

In recent years, a plethora of methods and procedures dealing with cardiovascular segmentation and analysis have emerged and are becoming increasingly complex [54, 25, 102]. To enable faster research in this area, it is necessary to identify the components or building

blocks of the various methods, their influence, their relationship and the impact and sensitivity of their parameters to develop robust and efficient methods suitable for use in a clinical setting.

1.2 Objectives

The objective of this thesis is related to the development of new, robust and accurate methods for cardiac image segmentation and analysis. This thesis focuses on improving deep learning-based methods for whole heart segmentation, bi-ventricle, myocardium segmentation and quantification as well as abdominal aortic aneurysm segmentation. The main objectives can be summarized as follows:

- Develop a novel and optimized connectivity structure of residual units that will significantly reduce number of network parameters.
- Develop a novel method for the automated segmentation of the whole heart and heart chambers from 3D CT and 3D MRI images.
- Develop a novel method for the automated segmentation and quantification of the LV, RV and Myo from cine-MRI images (2D + time).
- Develop a novel method for the automated segmentation of abdominal aortic aneurysms in 3D CTA images.
- Contribute to the field of cardiovascular image segmentation and analysis by providing insight into existing methods through critical review.
- Contribute to the field of cardiovascular image segmentation by providing new, robust and highly accurate deep-learning based methods.

1.3 Contributions

In this thesis, we introduce one theoretical improvement of deep learning mechanisms by introducing a novel connectivity structure. We introduce three novel deep learning-based methods for cardiovascular image segmentation that increase training performance, efficiency and final segmentation accuracy. The proposed methods include:

- A novel connectivity structure of residual units named feature merge pre-activation residual units (FM-Pre-ResNets) that allow the creation of distinctly deeper models without an increase in the number of network parameters compared to the pre-activation residual units. FM-Pre-ResNets adds the two additional convolutional layers at the top and the bottom of the pre-activation

residual block. The top convolution layer balances the parameters of the two branches, while the bottom layer reduces the channel dimension. In this way, it is possible to construct a deeper model with similar or fewer parameters than the original pre-activation residual unit.

- A 3D encoder-decoder architecture based on FM-Pre-ResNets and variational autoencoder (VAE) is proposed for the task of the whole heart segmentation from CT and MR images. FM-Pre-ResNet units are used to learn a low-dimensional representation of the input during the encoding stage. Following that, the variational autoencoder (VAE) reconstructs the input image from the low-dimensional latent space, ensuring that the model weights are strongly regularized while avoiding over-fitting the training data. The decoding stage generates the final segmentation of the whole heart and heart chambers.

- Modified 3D U-Net architecture that incorporates SERes blocks into 3D U-Net architecture (3D SERes-U-Net) for the task of LV, RV and Myo segmentation. The SERes blocks incorporate channel-wise squeeze and excitation operations into residual learning. An adaptive feature re-calibration ability of squeeze and excitation operations boosts the network's representational power, while feature reuse utilizes effective learning of the features, which improves segmentation performance. Additionally, based on obtained segmentations, LV, RV and Myo volumes are calculated. Significant indicators of hearts' function, including volume of the left ventricle at end-diastole (LVEDV), the volume of the left ventricle at end-systole (LVESV), left ventricles' ejection fraction (LVEF), the volume of the right ventricle at end-diastole (RVEDV), volume of the right ventricle at end-systole (RVESV), right ventricles' ejection fraction (RVEF), myocardium volume at end-systole (MyoVES) and myocardium mass at end-diastole (MyoMED) are calculated and compared to referent values.

- Modified 3D U-Net architecture with the addition of residual units in the contracting pathway and deep supervision in expanding pathway for the task of an abdominal aortic aneurysm segmentation (AAA). The addition of residual units in the contracting pathway preserves information. It significantly increases network performance, while the addition of deep supervision in expanding pathway injects gradient signals deep into the network.

In this thesis, cardiac images have been chosen as a target organ for analysis; however, the proposed methods can be applied to any other organs and image modalities.

1.4 Publications

The research work reported in this Thesis (as a first author) appears in two journals in the Science Citation Index Expanded (SCIE), one journal in the Emerging Sources Citation Index (ESCI) and five proceedings of international conferences:

- Marija Habijan, Irena Galić, Hrvoje Leventić, and Krešimir Romić. Whole Heart Segmentation Using 3D FM-Pre-ResNet Encoder–Decoder Based Architecture with Variational Autoencoder Regularization. *Applied Sciences*, 11(9), 3912, 2021

- Marija Habijan, Danilo Babin, Irena Galić, Hrvoje Leventić, Krešimir Romić, Lazar Velicki, and Aleksandra Pižurica. Overview of the whole heart and heart chamber segmentation methods. *Cardiovascular Engineering and Technology*, pages 1–23, 2020.

- Marija Habijan, Hrvoje Leventić, Irena Galić, and Danilo Babin. Neural network based whole heart segmentation from 3D CT images. *International journal of electrical and computer engineering systems*, 11(1), 25–31, 2020.

- Marija Habijan, Irena Galić, Hrvoje Leventić, Krešimir Romić, Danilo Babin. Segmentation and Quantification of Bi-Ventricles and Myocardium Using 3D SERes-U-Net. *2021 International Conference on Systems, Signals and Image Processing (IWSSIP)* Bratislava, Springer, 2021.

- Marija Habijan, Irena Galić, Hrvoje Leventić, Krešimir Romić, and Danilo Babin. Abdominal aortic aneurysm segmentation from CT images using modified 3D U-Net with deep supervision. *2020 International Symposium ELMAR*, pages 123–128, IEEE, 2020.

- Marija Habijan, Danilo Babin, Irena Galić, Hrvoje Leventić, Lazar Velicki, and Milenko Cankovic. Centerline tracking of the single coronary artery from x-ray angiograms. *2020 International Symposium ELMAR*, pages 117–121, IEEE, 2020.

- Marija Habijan, Hrvoje Leventić, Irena Galić, Danilo Babin. Whole heart segmentation from CT images using 3D U-Net architecture. *Proceedings of 2019 international conference on systems, signals and image processing*, pages 121-126, IEEE, 2019.

- Marija Habijan, Hrvoje Leventić, Irena Galić, Danilo Babin. Estimation of the left ventricle volume using semantic segmentation. *2019 61st International Symposium ELMAR*, pages 39-44. IEEE, 2019.

Additionally, the research work during this Thesis that contributions to other peoples' work (as a co-author) resulted in the two journals in the Science Citation Index Expanded (SCIE) and five proceedings of international conferences and is listed below:

- Marin Benčević, Irena Galić, Marija Habijan, and Danilo Babin. Training on Polar Image Transformations Improves Biomedical Image Segmentation. *IEEE Access* vol. 9, pages 133365-133375, 2021.

- Krešimir Romić, Irena Galić, Hrvoje Leventić, Marija Habijan. Pedestrian Crosswalk Detection Using a Column and Row Structure Analysis in Assistance Systems for the Visually Impaired. *Acta Polytechnica Hungarica*, pages 25-45, vol 18(7), 2021.

- Marin Benčević, Marija Habijan, Irena Galić. Epicardial Adipose Tissue Segmentation from CT Images with A Semi-3D Neural Network. Epicardial Adipose Tissue Segmentation from CT Images with A Semi-3D Neural Network. *2021 International Symposium ELMAR*, pages 87-90, IEEE, 2021.

- Danilo Babin, Daniel Devos, Ljiljana Platiša, Ljubomir Jovanov, Marija Habijan, MHrvoje Leventić, Wilfried Philips. Segmentation of Phase-Contrast MR Images for Aortic Pulse Wave Velocity Measurements. *International Conference on Advanced Concepts for Intelligent Vision Systems* New Zealand, pages 77-86, Springer International Publishing, 2020.

- Marin Benčević, Hrvoje Leventić, Danilo Babin, Marija Habijan, Irena Galić. A survey of Left Atrial Appendage Segmentation and Analysis in 3D and 4D medical images. *2021 International Conference on Systems, Signals and Image Processing (IWSSIP)* Bratislava: Springer, 2021.

- Krešimir Vdovjak, Hrvoje Leventić, Marija Habijan, Irena Galić. Adaptive Thresholding for Single Click Left Atrial Appendage Segmentation. *2019 International Symposium ELMAR*, Zadar, pages 35-38, 2019.

- Kresimir Romic, Irena Galic, Hrvoje Leventic, Marija Habijan. SVM based column-level approach for crosswalk detection in low-resolution images. *2020 International Symposium ELMAR*, Zadar, pages 133-137, IEEE, 2020.

To summarize, the work conducted during this Thesis resulted in 5 journal publications (of which 3 as the first author), 10 papers are published at international conferences (of which 5 as the first author) and 1 publication in book chapters (as co-author).

1.5 Organization of the Thesis

This Thesis introduces one theoretical improvement of deep learning mechanisms by introducing a novel connectivity structure of residual units. Further, we introduce a series of deep-learning methods for heart and heart chambers segmentation. We focus on improving deep learning segmentation methods for whole heart segmentation, bi-ventricle and myocardium segmentation, as well as abdominal aortic aneurysm segmentation. This section presents an overview of the content of the chapters of the thesis.

CHAPTER 2: MEDICAL BACKGROUND.

In this chapter, we introduce the medical background concerning the cardiovascular system and heart anatomy. We give an overview of the cardiovascular anatomy with a focus on the LV, RV and Myo anatomy and function. We describe aorta anatomy and four types of aortic aneurysms. Several medical imaging modalities are routinely used for specific cardiovascular structure assessment - computer tomography (CT) and magnetic resonance (MRI). We explain their advantages and disadvantages in the context of specific cardiovascular structure assessments.

CHAPTER 3: RELATED RESEARCH

This chapter briefly reviews the most relevant deep learning mechanisms and prior research in cardiovascular image segmentation. In the first part of the chapter, we give an overview of the deep learning CNN network building blocks and commonly used architecture - 3D U-Net architecture. We further overview significant residual network variants and autoencoders relevant to our research to further highlight the main focus of the thesis. In the second part of the chapter, we give an overview of the state-of-the-art deep-learning-based approaches for cardiovascular segmentation. First, we present an overview of the whole heart segmentation methods from CT and MRI images. Second, we provide an overview of the LV, RV and Myo segmentation and quantification methods from Cine-MRI images. Third, we give an overview of segmentation methods for abdominal aortic aneurysm segmentation from CT images. The approaches for the segmentation and analysis of the whole heart, bi-ventricles and aorta have been intensively researched both with traditional segmentation methods and artificial intelligence. Nevertheless, the need for increasing accuracy, robustness and optimality in performance remains a challenge that needs to be addressed, motivating the development of the methods proposed in this thesis.

CHAPTER 4: WHOLE HEART AND HEART CHAMBERS SEGMENTATION

This chapter describes our new method for segmenting the whole heart and its chambers. The suggested approach for segmenting the whole heart employs a novel three-dimensional (3D) encoder-decoder architecture that successfully includes a novel connectivity structure for

a residual unit - FM-Pre-ResNet - and variational autoencoder (VAE). FM-Pre-ResNet enables the construction of robust models without increasing the number of parameters in comparison to pre-activation residual units. By incorporating two convolutional layers at the top and bottom of the pre-activation residual block, the parameters of the two branches are balanced. In comparison, the bottom layer reduces the dimension of the channel. This allows for the construction of a more detailed models with a similar or smaller number of parameters to the initial pre-activation residual unit. As indicated previously, the general architecture is encoder-decoder based. The method consist of three main steps. In first step, FM-Pre-ResNet units are used to learn a low-dimensional representation of the input during the encoding stage. Following that, the variational autoencoder (VAE) reconstructs the input image from the low-dimensional latent space, ensuring that the model weights are strongly regularized while also avoiding over-fitting on the training data. The decoding stage generates the final segmentation of the whole heart. The validation of the extraction method is performed by measuring the Dice score (DSC), Jaccard index (JI), surface distance (SD) and Hausdorff distance (HD) between our segmentation results and ground truth segmentations by radiologists.

chapter 5: Bi-Ventricles and Myocardium Segmentation and Quantification

This chapter presents our method for the LV, RV and Myo segmentation and quantification. The proposed method relies on modifying 3D U-Net architecture and incorporates SERes blocks into 3D U-net architecture (3D SERes-U-Net). The SERes blocks incorporate channel-wise squeeze and excitation operations into residual learning. An adaptive feature re-calibration ability of squeeze and excitation operations boosts the network's representational power, while feature reuse utilizes effective learning of the features, which improves segmentation performance. The validation of the segmentation method is performed by measuring DSC and HD between our segmentation results and ground truth segmentations by radiologists.

chapter 6: Abdominal aortic aneurysm segmentation

In this chapter, we introduce our method for the task of abdominal aortic aneurysm segmentation. The proposed method is based on 3D U-Net architecture and incorporates residual learning and deep supervision in encoder and decoder paths, respectively. The addition of residual units in the contracting pathway preserves information. It significantly increases network performance, while the addition of deep supervision in expanding pathway inject gradient signals deep into the network. The validation of is performed by measuring DSC, JI, SD and HD distance between our segmentation results and ground truth segmentations by radiologists.

chapter 7: conclusions

The final chapter states the global conclusions of the thesis and points to some directions for further research continuing on this work.

CHAPTER 2

Medical Background

This chapter introduces the medical background of the cardiovascular system, the heart, and heart structures. We give a brief overview of their anatomy and functional indices. The description of LV and RV functional indices includes their definitions, calculation methods, corresponding normal ranges and exemplary applications for the diagnosis of CVDs. CVDs affect the structures of the cardiovascular system and significantly disrupt its proper function. We focus primarily on an overview of CVDs relevant to our research: cardiomyopathy, congenital heart disease, ventricular hypertrophy and aortic aneurysm. Medical imaging techniques can produce detailed images that depict human anatomy in vivo. The images produced reveal structural and functional information about organs and tissues. Therefore, cardiac imaging plays an essential role in informing the nature of pathological conditions. Invasive and noninvasive imaging modalities such as echocardiography, computed tomography (CT), magnetic resonance imaging (MRI) and single photon emission computed tomography (SPECT) have been developed to assess the anatomy and functionality of the cardiovascular system.

The outline of the chapter is structured in the following manner. Section 2.1 introduces the medical background of the cardiovascular system, heart and heart structures. In Section 2.2 we briefly describe CVDs relevant to our research. Section 2.3 gives a brief overview and characteristics of CT, MRI and Cine-MRI imaging modalities. The representation of different heart structures on each imaging modality is discussed as well.

2.1 Cardiovascular System

The cardiovascular system is a collection of organs that regulate blood circulation throughout the body. The heart, vessels, arteries, capillaries, veins and blood comprise the cardiovascular system. The main function of the circulatory system is to provide a continuous supply of oxygen and nutrients such as amino acids and glucose to every cell in the human body. The vascular system makes this circulation possible [4]. The circulatory system consists of two closed circuits: the pulmonary circuit and the systemic circuit. The pulmonary circulation carries deoxygenated blood to the lungs, where it picks up oxygen, and then returns to the heart via the pulmonary veins [121]. The systemic circulation transports oxygenated blood from the heart to the tissues and cells and back again. All cells of the body are supplied with blood and oxygen in this manner. In addition to distributing oxygen, the cardiovascular system also collects and distributes carbon dioxide and other waste products of metabolism to the lungs, liver and kidneys, where they are removed from the body. An illustration of this process is shown in Figure 2.1.

2.1.1 Heart Anatomy and Physiology

The human heart is a muscular organ with four chambers. The two upper chambers (atria) are separated by an atrial septum (septum interatriale), which resembles a wall. A similar structure, called the interventricular septum, separates the two lower chambers (ventricles). Valves connect the atria and ventricles, allowing blood to flow in one direction and preventing backflow. The usual procedure of blood flow is as follows [130]. Through two major veins, the superior vena cava and inferior vena cava, deoxygenated blood from all of the body's tissues reaches a relaxed right atrium. The right atrium contracts and blood flows into the relaxed right ventricle via the tricuspid valve. The right ventricle then contracts and blood is forced into the pulmonary artery, which delivers it to the lungs for oxygenation via the pulmonary valve. The left atrium receives blood that has been oxygenated by the lungs and is returning to the heart. The four pulmonary veins supply this blood to the relaxed left atrium. The left atrium contracts and blood flows into the relaxed left ventricle via the mitral valve. When the left ventricle contracts, blood is forced into the aorta, the body's biggest artery, via the aortic valve. The aorta is responsible for transporting blood to all regions of the body. An illustration of blood flow through the heart is shown in Figure 2.2.

The heart is a muscular pump made up of cardiac muscle fibers. The heart is placed in the mediastinum in the chest cavity's center; however, it is not perfectly centered; more of the heart is on the left side than on the right. The heart's interior is divided into four hollow chambers – two atria and two ventricles – that are frequently referred

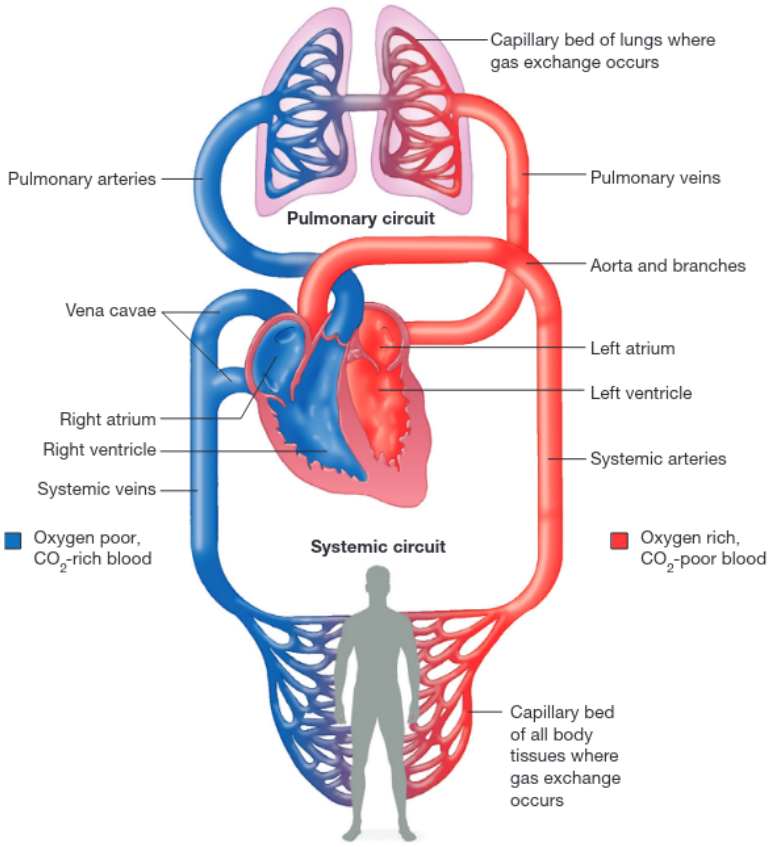


Figure 2.1: An illustration of the circulatory system. The pulmonary circulation picks up oxygen from the lungs and the systemic circulation delivers oxygen to the body. Image source: Quizlet Plus [132]

to as the left and right hearts. The left atrium (LA) and left ventricle (LV) are the chambers of the left heart, whereas the right atrium (RA) and right ventricle (RV) are the chambers of the right heart [4]. An illustration of the heart anatomy is shown in Figure 2.3(a).

The pericardium surrounds the entire heart, a thin sac that protects it and prevents it from becoming overly enlarged. The pericardial space or pericardial cavity is the space between the heart and the pericardium. It contains the interstitial fluid that serves as a lubricant. When the cardiac muscle contracts, blood is expelled from the heart and pumped through the arteries into the body. The heart muscle, also called the myocardium, is the force generator that causes the heart to contract. The myocardium is located within the walls of the heart chambers. Two layers surround the myocardium: the endocardium on the inside of the chambers and the epicardium on the outside. Figure 2.3(b) illustrates the composition of the heart wall. Strands and clusters of specialized cardiac muscle are found throughout the heart. They comprise only a few myofibrils. These areas initiate and distribute impulses throughout the heart. They form the cardiac conduction

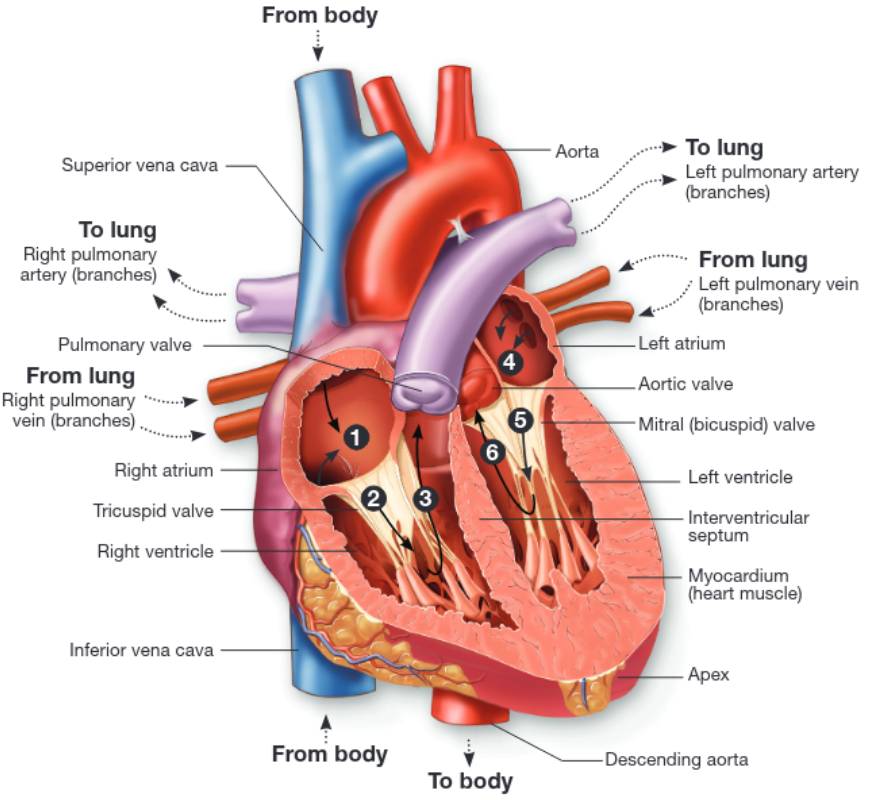
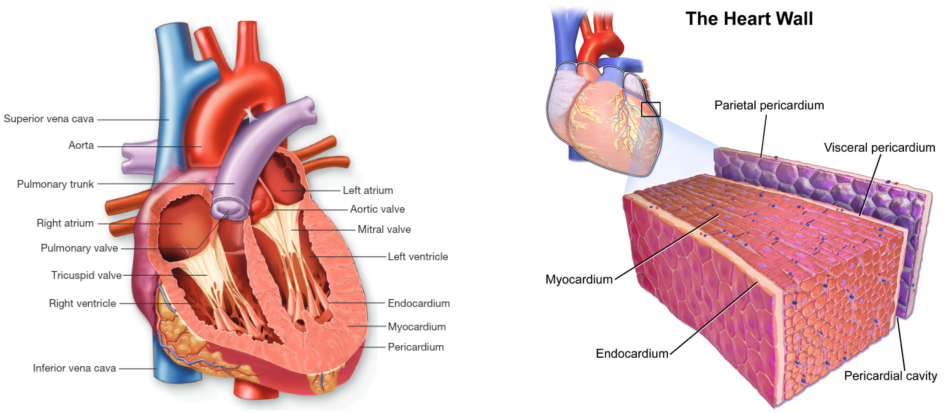


Figure 2.2: The path of blood flow through the chambers of the left and right side of the heart. Image source: Lumen Learning [95]

system, which is responsible for coordinating the cardiac cycle.



(a) (b)

Figure 2.3: Heart anatomy. (a) Diagram of the human heart. Image source: Wikimedia [131]. (b) Illustration of the heart wall. Image source: Medical gallery of Blausen Medical[16]

Cardiac Cycle

The sinoatrial node (SA node) is a small collection of specialized tissue located under the right atrium. It is located near the entrance of the superior vena cava and has fibers connected to those of the atrial syncytium. The cells of the SA node can reach the threshold of stimulation independently, triggering impulses through the heart and driving contraction of the cardiac muscle fibers. A cardiac impulse travels from the SA node to the atrial syncytium, where it causes the atria to contract virtually simultaneously. The impulse travels via the junctional fibers of the excitation conduction system to a collection of specialized tissue called the atrioventricular node (AV node). Av node is located under the endocardium in the inferior atrial septum (septum interatriale). The AV node is the only natural conduit between the atrial and ventricular synapses. Due to the small diameter of the junctional fibers, impulses are slightly delayed. This gives the atria more time to contract and release blood into the ventricles before the ventricles contract. For an illustration of the excitation conduction system of the heart, see Figure 2.4. The cardiac cycle is defined as the alternate contraction and relaxation of the atria and ventricles to pump blood throughout the body. Each cardiac cycle consists of a diastolic phase (also called diastole) and a systolic phase (also called systole) [126, 63].

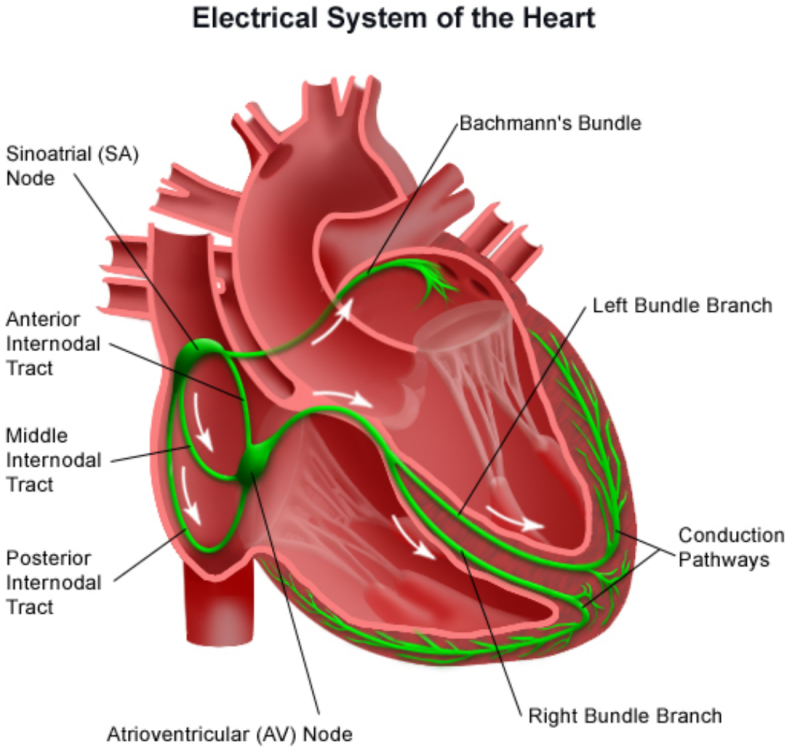


Figure 2.4: Diagram of the heart conduction system. Image source: Wikimedia [131]

During the diastolic phase, the heart chambers relax and are filled with blood from the veins or atria. In systole, the ventricles contract and blood is pumped to the periphery through the arteries. Both the atria and the ventricles alternate between the phases of systole and diastole. In other words, the ventricles are in systole when the atria are in diastole and vice versa.

Atrial diastole is the first event of the cardiac cycle. It occurs a few milliseconds before the electrical signal from the SA node reaches the atria. The atria act as conduits and primers for blood flow to the ipsilateral ventricle and for pumping residual blood to the ventricles. During atrial diastole, blood enters the right atrium via the superior and inferior vena cava, coronary sinus and the left atrium via the pulmonary veins. The atrioventricular valves are closed in the early stages of this phase and blood pools in the atria. When the pressure in the atrium is greater than the pressure in the ventricle on the same side, the pressure difference causes the atrioventricular valves to open, allowing blood to flow into the ventricle [126, 63].

The SA node triggers an action potential that propagates throughout the atrial myocardium during atrial systole. The electrical depolarization causes the atria to contract simultaneously, pushing the remaining blood from the upper chambers into the lower chambers of the heart. Contraction of the atria results in an additional increase in pressure in the atria. Both the atrioventricular and semilunar valves are closed during the early stages of ventricular diastole. The blood volume in the ventricle remains constant during this phase, whereas intraventricular pressure drops precipitously. This phenomenon is called isovolumetric relaxation. The ventricular pressure eventually falls below the atrial pressure, whereupon the atrioventricular valves open. This leads to rapid filling of the ventricles with blood, often referred to as rapid ventricular filling. It is responsible for most of the blood in the ventricle before contraction. A small amount of blood flows directly into the ventricles from the venae cavae. The remaining blood in the atria is forced into the ventricle towards the end of the ventricular diastole. End-diastolic volume or preload refers to the total volume of blood that is in the ventricle at the end of diastole.

Ventricular systole refers to the period during which the ventricles contract. The AV node receives the electrical impulse shortly after the atria depolarize. At the AV node, there is a short delay that allows the atria to fully contract before the ventricles depolarize. The action potential is conducted through the AV node and then through the left and right bundle branches. The electrical impulses are transmitted from these fibers through the respective ventricles and cause the ventricles to contract. When the ventricle begins to contract, the pressure in the ventricle exceeds the pressure in the corresponding atrium, causing the atrioventricular valves to close. At the same time, the pressure is insufficient to open the semilunar valves. As a result, the ventricles contract isovolumetrically - the end-diastolic volume does not change.

When the pressure inside the ventricle exceeds the pressure outside the ventricle, the semilunar valves open and blood can leave the ventricle. This is the ejection phase of the cardiac cycle. End-systolic volume refers to the amount of blood remaining in the ventricle at the end of systole. The ventricles re-enter isovolumetrically relaxed while the atria continue to fill. The cycle begins again and continues indefinitely as long as the individual is alive. For an illustration of the whole cardiac cycle process, see Figure 2.5.

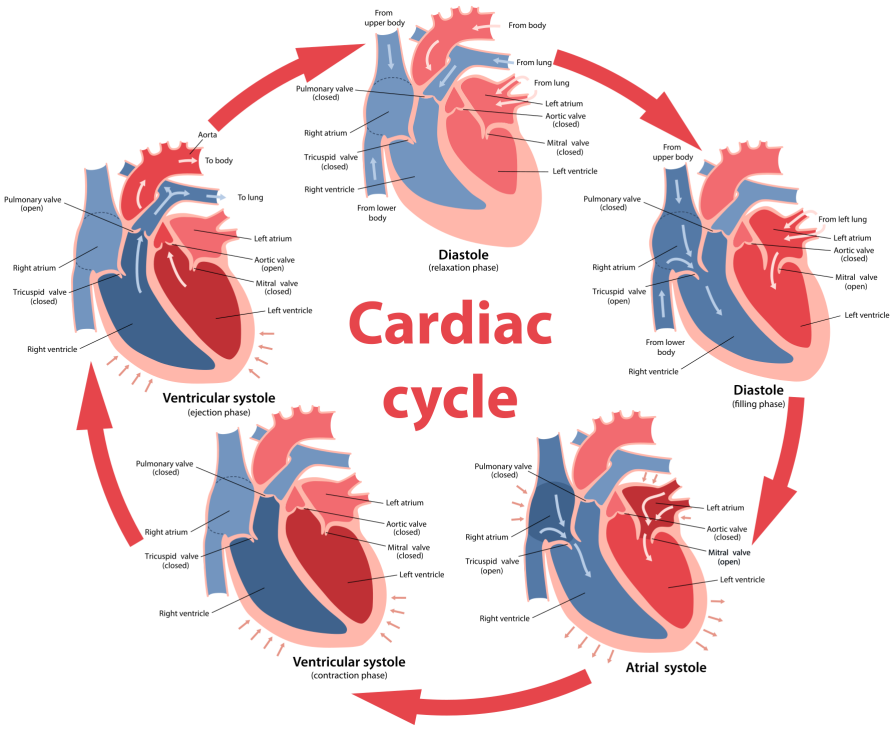


Figure 2.5: Cardiac cycle. Image source: Wikimedia [131]

2.1.2 Anatomy and Physiology of Ventricles

The left ventricle is critical for maintaining pulsatile blood flow in the despite of relatively high pressures in the systemic circulation. It is a ventricle that is muscular and receives oxygenated blood from the left atrium. The walls of the left ventricle are three times as thick as those of the right ventricle. The base of the left ventricle originates at the left atrioventricular valve and extends toward the apex of the heart. Subsequently, the ventricular canal curves towards the aortic valve, where blood is expelled into the aorta and the systemic circulation [128]. The inner surface of the left ventricle is often inconspicuous. In the right ventricle, the lumen is oval and trabeculated toward the apex. On the left side of the heart, the valves are more tightly interconnected than on the right side. The mitral, aortic and tricuspid valves are connected by the fibroelastic cardiac skeleton (formed by the left and

right fibrous trigones). The mitral valve is separated from the aortic valve by a fibroelastic subaortic curtain that descends from the left and right posterior arches of the aortic valve. In addition, the left ventricle has anterior and posterior papillary muscles associated with the chordae tendineae. The papillary muscles of the left ventricle are significantly larger than those of the right ventricle. The free edge of each mitral leaflet receives multiple chordae tendineae from both papillary muscles. This is most likely because these papillary muscles must withstand increased pressure in order to keep the mitral valve closed during ventricular systole. The anatomy of the left ventricle is depicted in Figure 2.6. The interventricular septum is a crucial structure that divides the two ventricles. It is separated anatomically into two sections: a dense, muscular section and a relatively thin, membranous section—the interventricular septum's muscular portion of the majority of the left and right ventricles.

The membranous part of the interventricular septum, located posteriorly and superiorly in the left ventricle, separates the right ventricles from the subaortic region. The right ventricle is the smallest of the two lower chambers of the heart, measuring less than one-third the thickness of its counterpart. Despite its smaller size, the right ventricle pumps the same amount of blood as the left ventricle [166]. However, it has to exert less effort because the pulmonary circulation has much less resistance than the systemic circulation. The right ventricle is located in front of the left atrium and in front of the right atrium. It begins at the opening of the tricuspid valve orifice (right atrioventricular valve) and continues inferolaterally to the apex of the heart's apex [36, 9]. The chamber's natural contour of the chamber then curves upward into the conus arteriosus (also called the infundibulum) and ends at the pulmonary valve orifice (right semilunar valve). It is somewhat difficult to determine the geometric shape of the right ventricle.

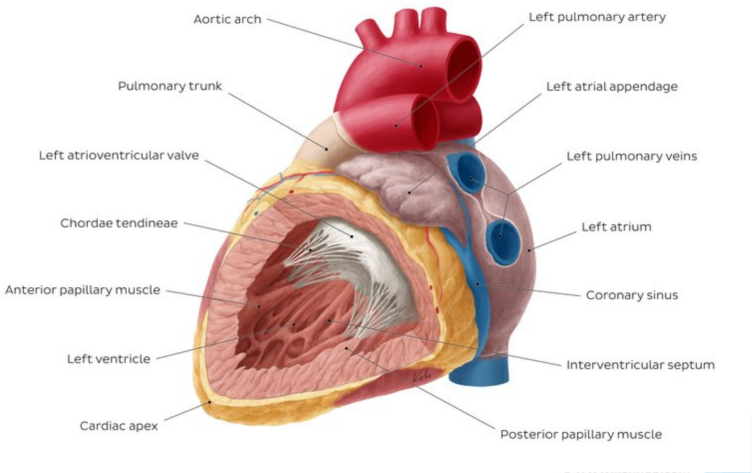


Figure 2.6: Left ventricle anatomy. Image source: KenHub [85]

When viewed from the side, the structure appears triangular. However, when viewed transversely, it appears crescent-shaped because the free wall (the part that is not connected to the apex or septum) curves inward. The atrioventricular septa separate the atria from the ventricles. This is a fibrous elastic structure that prevents inadvertent blood flow from the atrium to the ventricles and inadvertent electrical conduction from the atrial myocardium to the ventricles. Without this arrangement, blood and electrical activity would flow backward through the myocardium. For an illustration of the anatomy of the right atrium and ventricles, see Figure 2.7.

Clinical Indices

The analysis of cardiac function begins with the computation of a set of indices for different heart structures. Due to the fact that the LV and RV have significantly different volumes during distinct phases of the cardiac cycle, structural and functional indices are calculated at both phases of the cardiac cycle: at the end of diastole and at the end of systole [159]. The volumetric variations between the left and right ventricles during relaxation and contraction are shown in Figure 2.8. Apart from volume calculations, other frequently used approaches include the single area-length method, the bi-plane area-length method, Simpson's method and direct measurement. Nonetheless, we limited ourselves to volumetric calculations in this Thesis. Indices of ventricular morphology and function fall into two categories: global and regional. Global indices include ventricular volume, stroke volume, ejection fraction (EF), cardiac output (CO) and myocardial mass. Regional or local indices include myocardial wall thickness (WT) and wall thickening (WTK). Strain analysis can be either global or local [82].

Left ventricular end-diastolic and end-systolic volumes (LVEDV and LVESV, respectively) are measures of the amount of blood in the chamber enclosed by myocardial tissue when the myocardium is relaxed (LVEDV) or contracted (LVESV). The amount of blood ejected from the heart after each contraction is called the left ventricular stroke volume (LVSV), which is the difference between LVEDV and LVESV. Left ventricular ejection fraction refers to the amount of blood ejected from the heart (LVEF). It subtracts the LVSV from the LVEDV. The change in myocardial wall thickness during systole is called left ventricular wall thickening (LVWT)[82, 171].

Left ventricular strain (LVS) is a measure of the degree of ventricular deformation and the rate of deformation is the left ventricular strain rate (LVSR). The amount of systemic blood flowing through the heart each minute is called cardiac output (CO). It is calculated by multiplying the LVSV by the heart rate (HR), which reflects the frequency of heartbeats (beats per minute). Left ventricular mass (LVM) is a measure of myocardial tissue. The volume of the myocardium is obtained by subtracting the endocardial volume from the volume within the epicardial border. Then calculate the mass as the product of the

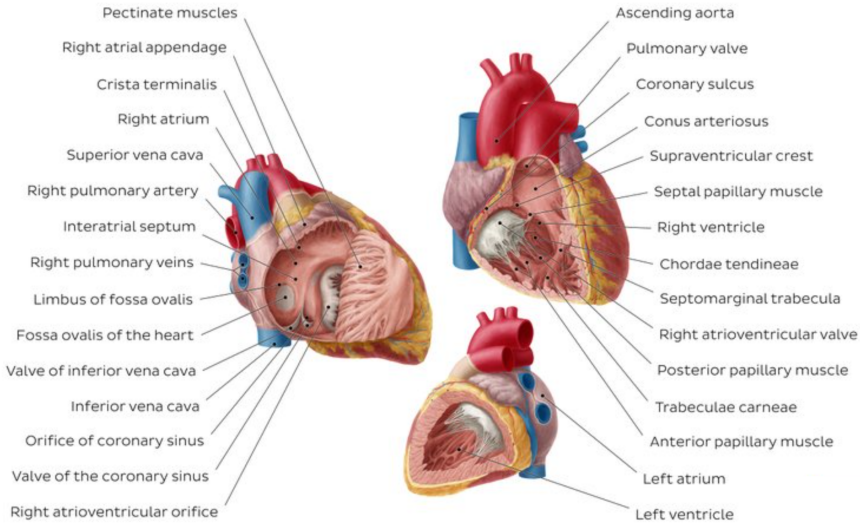


Figure 2.7: Right atrium and right ventricle anatomy. Image source: KenHub [85]

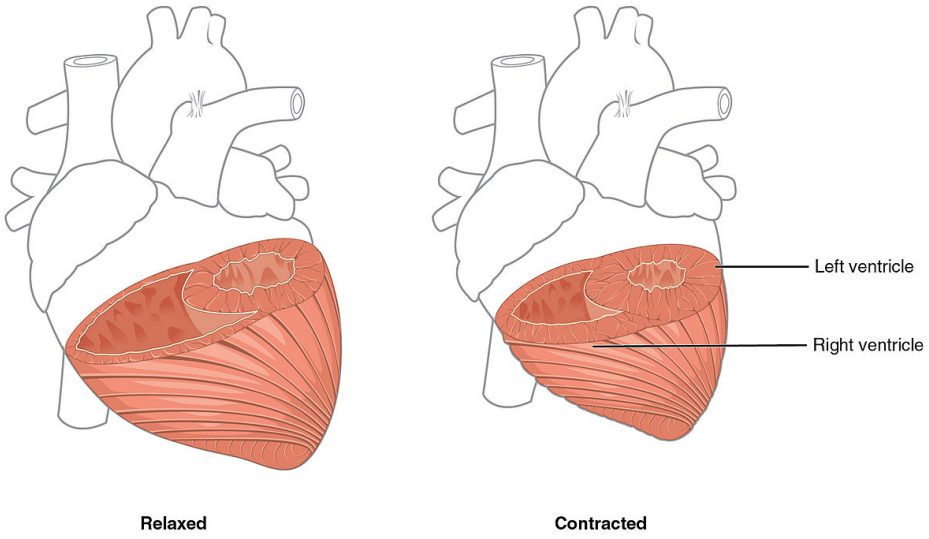


Figure 2.8: An illustration of ventricles at contraction and relaxation. Image source: Wikimedia [28]

cardiac volume and the muscle density. The mean distance between the endocardial and epicardial contours can be used to determine left ventricular wall thickness (LVWT) or myocardial thickness. Most right ventricular (RV) indices, such as right ventricular end-diastolic and end-systolic volumes (RVEDV and RVESV), right ventricular stroke volume (RVSV), right ventricular ejection fraction (RVEF) and right ventricular cardiac output (RVCO), are defined similarly to the corresponding values for the left ventricle (LV) [82]. For a summary of the LV and RV function indices, their calculation methods, normal ranges and common clinical applications, see Table 2.1.

Table 2.1: Summary of LV and RV functional indices. Table shows calculation methods, normal ranges and common clinical diagnostic applications for each functional index. Values for normal ranges taken from Kang et al.[82].

Indices	Calculation methods	Normal range	Applications
LVEDV	Single area-length method Bi-plane area-length method Simpson's method Direct measurement	$F : 128 \pm 21mL$ $M : 156 \pm 21mL$	Dilated cardiomyopathy
LVESV	Single area-length method Bi-plane area-length method Simpson's method Direct measurement	$F : 42 \pm 9.5mL$ $M : 53 \pm 11mL$	Dilated cardiomyopathy
LVM	$(LVV_{epi} - LVV_{endo}) \times 1.05$	$F : 108 \pm 18g$ $M : 146 \pm 20g$	Hypertension Hypertrophic cardiomyopathy
LVSV	LVEDV - LVESV	$F : 86 \pm 14mL$ $M : 104 \pm 14mL$ $M : 104 \pm 14mL$	Aortic stenosis Aortic insufficiency
EF	$\frac{LVEDV-LVESV}{LVEDV} \times 100\%$	$F : 67 \pm 4.6\%$ $M : 67 \pm 4.5\%$	Heart failure Hypertrophic cardiomyopathy
LVCO	$LVSV \times HR$	$4\breve{}8L/min^a$	Hypertension Congestive heart failure
RVEDV	Simpson's method	$F : 148 \pm 35mL$ $M : 190 \pm 33mL$	Arrhythmogenic right ventricular cardiomyopathy Congenital heart disease
RVESV	Simpson's method	$F : 56 \pm 18mL$ $M : 78 \pm 20mL$	Arrhythmogenic right ventricular cardiomyopathy
RVSV	RVEDV - RVESV	$F : 90 \pm 19mL$ $M : 113 \pm 19mL$	Pulmonary arterial hypertension
RVEF	$\frac{RVSV}{RVEDV} \times 100\%$	$F : 63 \pm 5.0\%$ $M : 59 \pm 6.0\%$	Pulmonary arterial hypertension Congenital heart disease
RVCO	$RVSV \pm HR$	$5.25L/min^b$	Ventricle failure with cardiomyopathy Pulmonary hypertension
WT	Radial method Centreline method	$F : 6.4 \pm 0.9mm$ $M : 7.8 \pm 1.1mm$	Myocardial infarction
WTK	$\frac{(WT_{ed}-WT_{es})}{WT_{ed}} \times 100\%$		Average ES WT Average ED WT

2.1.3 Aorta Anatomy

The aorta is the largest artery in the body, responsible for supplying nutrient-rich blood to the systemic circulation. It is classified according to its course and location in relation to the other organs and the body. The thoracic aorta begins at the heart, at the level of the aortic valves. It becomes the abdominal aorta at the diaphragm, just proximal to the celiac artery origin.

The thoracic aorta is further divided into the ascending aorta, the aortic arch and the descending aorta, as shown in Figure 2.9. The ascending aorta arches posteriorly and to the left and passes through the left pulmonary root to form the aortic arch. The aortic arch continues to curve backward and downward to form the descending aorta. The descending aorta is further divided into the thoracic and abdominal aortas. The thoracic aorta is a segment of the descending aorta located in the posterior mediastinal cavity. Numerous branches arise from the thoracic aorta [33]. The pericardial, bronchial, esophageal and mediastinal branches are all visceral branches. Coronary arteries are tiny vessels that travel to the posterior surface of the pericardium.

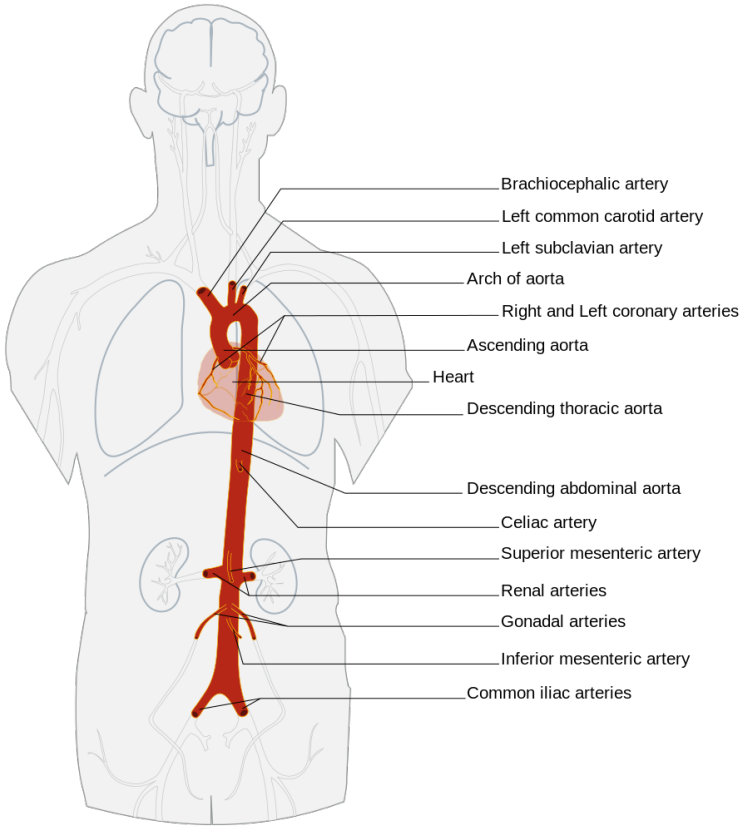


Figure 2.9: Segments of the aorta, including: thoracic aortaa ascending aorta, aortic arch, descending aorta, abdominal aorta (suprarenal abdominal aorta, infrarenal abdominal aorta). Image source: Wikimedia [40]

Moreover, the left bronchial arteries, of which there are usually two, arise from the thoracic aorta [92]. These arteries supply blood to the bronchial airways, the pulmonary area and the esophagus. The esophageal arteries arise in the anterior aorta and run downward to the esophagus, where they anastomose with a number of other arteries. The mediastinal branches of the thoracic aorta continue to supply the lymph glands and areolar tissue in the posterior mediastinum. The intercostal, subcostal and superior phrenic branches are all parietal branches. There are nine pairs of intercostal arteries arising from the posterior segment of the aorta. These arteries divide further to form the intercostal artery, the lateral cutaneal artery, the mammary artery and the spinal artery, to name a few. The superior phrenic artery also arises from the thoracic aorta and eventually forms an anastomosis with the pericardiophrenic and musculophrenic arteries. The lowest branching arteries of the thoracic aorta are the subcostal arteries, which eventually branch into a posterior branch. Five primary branches arise from the abdominal aorta: the celiac artery, the superior mesenteric artery, the left and right renal arteries and the inferior mesenteric artery. The celiac trunk supplies primarily the organs of the foregut, while the superior mesenteric artery and inferior mesenteric artery supply the organs of the midgut and hindgut, respectively [153]. The abdominal aorta terminates at its branch into the common iliac arteries, which then supply the pelvis and lower limbs with arterial blood.

2.2 Cardiovascular Diseases

CVDs claim more lives each year than any other cause: with an estimated 17.9 million deaths in 2019, CVDs account for 31% of all deaths worldwide [32]. The term CVDs refers to a group of diseases of the heart and blood vessels. These include coronary artery disease, rheumatic heart disease, congenital heart disease, stroke, aortic aneurysm and dissection, peripheral arterial disease, deep vein thrombosis (DVT) and pulmonary embolism. We describe the following CVDs and cardiovascular disorders relevant to our research:

- Cardiomyopathy - a heart muscle disease that hardens blood pumping to the rest of the body.
- Congenital heart disease - malformations of heart structures existing at birth may be caused by genetic factors or by adverse exposures during gestation.
- Aortic aneurysm and dissection - dilatation and rupture of the aorta.
- Ventricular hypertrophy - a condition defined by an abnormal enlargement of the cardiac muscle surrounding the left or right ventricle.

2.2.1 Cardiomyopathy

Cardiomyopathy is a heart muscle disease in which the heart muscle is physically or functionally damaged, making it difficult for the heart to pump blood throughout the body. Cardiomyopathy is divided into the following subtypes: dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy (RCM), arrhythmogenic right ventricular dysplasia (ARVCD) and unclassified cardiomyopathy [17, 109, 157].

DCM is the most common form of myocardial disease and accounts for approximately 60% of all cardiomyopathies [17, 72]. In DCM, the left ventricle becomes dilated (enlarged). In addition, decreased LVEF leads to impaired pump function - the LV is unable to properly move blood out of the heart, as shown in Figure 2.10. Patients with dilated cardiomyopathy have an ejection fraction less than 40%, a left ventricular volume greater than 100 $mL/m2$ and a diastolic wall thickness less than 12 mm [141, 107]. Certain patients with dilated LV also have a dilated RV or a large LV mass as a result of dilated LV [157]. This condition may be seen in the presence of other known CVDs. However, to be classified as DCM, the extent of myocardial dysfunction cannot be explained solely by abnormal loading conditions (hypertension or valvular disease) or ischemic heart disease [146, 119]. Numerous cardiac and systemic conditions can lead to systolic dysfunction and LV dilatation, but the etiology is often unknown [112]. This has led to the term idiopathic dilated cardiomyopathy (IDC) [58]. Similar to hyperdynamic LV systolic function as evidenced by a high LVEF in HCM, hyperdynamic LV systolic function is abnormal myocardial thickening that makes the heart work harder, as shown in Figure 2.10.

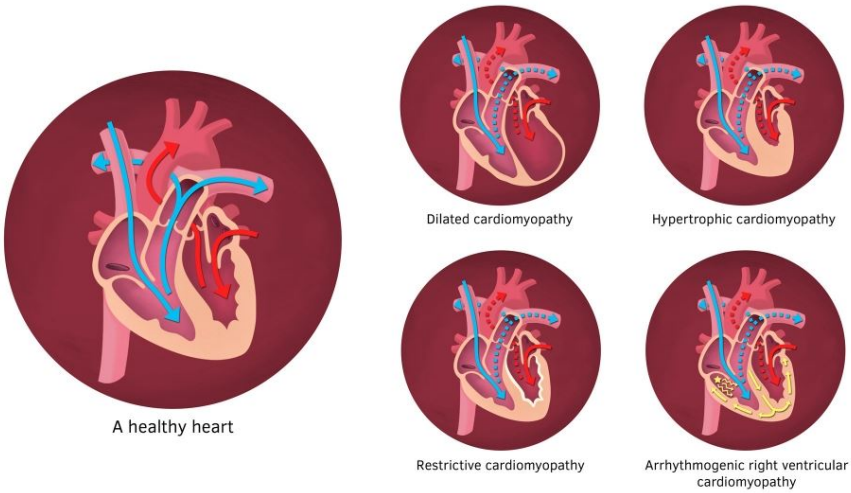


Figure 2.10: An illustration of a healthy heart and heart affected by different cardiomyopathy types. Image source: Healthand [61]

In contrast, patients with HCM are defined as having normal cardiac function (ejection fraction greater than 55%) but diastolic myocardial segments thicker than 15 mm. According to Girolami et al. [50] patients in this category may have an abnormal cardiac mass index greater than 110 $g/m2$. HCM is thus defined by asymmetric or symmetric hypertrophy of LV associated with an increase in LV mass. Asymmetric hypertrophy is demonstrated by comparing the thickness of the septum with the thickness of the LV free wall and by the presence of a septal-to-free wall thickness ratio greater than 1.3. The most common form of HCM is asymmetric hypertrophy of the interventricular septum [6].

RCM is defined by decreased diastolic volume in one or both ventricles and impaired ventricular filling, but with normal or near-normal systolic function [120, 7]. RCM causes the heart muscle to stiffen and become less flexible, preventing it from expanding and filling with blood between heartbeats (Figure 2.10). Because of the increased diastolic pressure, restrictive filling occurs, resulting in passive venous congestion. Cardiac output can be increased by increasing heart rate, but this becomes ineffective because filling time is shortened. RCM may occur for no apparent reason (idiopathic) or as a result of another disease affecting the heart, such as amyloidosis [169].

ARVCD is a relatively rare form of cardiomyopathy in which the muscle in the lower right chamber of the heart (RV) is replaced by scar tissue, resulting in arrhythmias [157]. It is often caused by genetic mutations and is characterized by fibrosis and fatty infiltration of the RV myocardium and ventricular tachycardia/fibrillation [157].

2.2.2 Congenital Heart Disease

Congenital heart disease (CHD) is a condition in which the structure and function of the heart are abnormal due to abnormal development of the heart before birth [114]. According to Khoshnood [87], it is a leading cause of birth defects and the second leading cause of infant mortality. Pathophysiologically, it is defined by the presence of a shunt between arterial and venous blood, cyanosis and postnatal circulatory changes. A shunt is a connection between two heart chambers or vessels through which blood can flow from one to the other. There is a left-to-right shunt, a right-to-left shunt, or a bidirectional shunt. The direction is entirely determined by the pressure gradient across the shunt as it affects the state of pulmonary blood flow, which can be normal, increased or decreased. Increased blood flow to the lungs due to left-to-right shunts leads to LV volume overload and the possibility of dilatation followed by heart failure. If left untreated, this eventually leads to pressure overload in the pulmonary artery. This pressure irreversibly alters the arterial wall and increases pulmonary vascular resistance (PVR). When the PVR is greater than the systemic vascular resistance, the shunt becomes bidirectional or predominantly right-to-left. In a right-to-left shunt, venous blood with low oxygen saturation

mixes with arterial blood with high oxygen saturation, resulting in cyanosis. CHDs are often classified as follows: (1) CHDs with a shunt between the systemic and pulmonary circulation, (2) left heart CHDs, (3) right heart CHDs, (4) CHDs with an abnormal origin of the great arteries and (5) miscellanea [43, 139].

An atrial septal defect (ASD) is a common defect involving a shunt between the systemic and pulmonary circulation. It is characterized by an unexpected connection between the atrial chambers. Another type is the ventricular septal defect (VSD), which is defined by an abnormal connection between the two chambers of the heart. Atrioventricular septal defect refers to a group of malformations characterized by abnormal development of the atrioventricular junction (AVSD) [43]. The patent ductus arteriosus (PDA) is a vascular structure located near the origin of the left pulmonary artery that connects the descending aorta to the roof of the pulmonary arterial trunk [113]. Finally, the aortopulmonary window (APW) is a major defect between the ascending aorta and the main pulmonary artery [53].

Aortic stenosis (AS) accounts for 3-6% of all patients with left heart CHD. The stenosis may be valvular (70%), subvalvular (23%), or supravalvular (7%). The bicuspid aortic valve with a fused commissure and an eccentric orifice results in a valvular AS. Subvalvular stenosis may result from a discrete membranous diaphragm most commonly associated with other CHDs such as VSD, PDA, or coarctation. An hourglass-shaped aorta characterizes supravalvular AS [43]. In addition, coarctation of the aorta (CoA) occurs when the aorta narrows circumferentially, whereas aortic arch hypoplasia (AAH) occurs when the aorta is blocked at a specific location. An interrupted aortic arch (IAA) has a distinct shape and its anatomical spectrum ranges from a severe form of CoA to the absence of an arch segment. Congenital mitral valve stenosis, in which one or more components of the valve apparatus are defective and cor triatriatum sinister (CTS), in which the left atrium is divided into two chambers by a fibrous membrane, are extremely rare CHD conditions [11].

The most common type of right heart CHD is pulmonary valve stenosis (PVS), classified into two distinct subtypes. The first type of tricuspid valve is characterized by thin leaflets, cusp fusion and underdeveloped or absent commissures, resulting in a dome-shaped valve with a narrow orifice. In the absence of cusp fusion, the second type is defined by thickened and irregular leaflets with a variable hypoplastic annulus. Another common CHD is the tetralogy of Fallot (TOF), which is characterized by four anatomic abnormalities of the heart: a large malaligned ventricular septum, an anterior shift of the aorta over the ventricular septum, obstruction of the right ventricular outflow tract (RVOT) and right ventricular hypertrophy. Additionally, CHD with an abnormal origin of the great arteries includes great artery transposition (TGA). An atypical ventriculoatrial connection defines TGA. The aorta originates from the RV and is anterior to

and to the right of the PA, whereas the PA originates from the LV. Additionally, the group of CHDs known as double outlet right ventricle must be mentioned (DORV). DORV outflow tracts originate entirely or primarily from the RV and may physiologically behave similarly to a VSD, TGA, TOF, or single ventricle [43]. Figure 2.11 shows an illustration of different CHD types described above.

Apart from the CHDs mentioned previously, the miscellanea category encompasses a variety of conditions. For instance, the common arterial trunk (CAT) is defined by a single great artery that originates at the base of the heart and supplies systemic, coronary and pulmonary blood flow, as well as by VSD. Pulmonary atresia (PuA) is a congenital heart defect in which the valve regulating blood flow from the heart to the lungs does not form properly [133]. Anomalous pulmonary venous (APV) is a condition in which one or more pulmonary veins return to the right atrium rather than the left. A persistent left superior vena cava causes the double superior vena cava (DSVC) as a result of the anterior cardinal vein not regressing embryologically. The left pulmonary artery originates from the right pulmonary artery and encircles the right mainstem bronchus and distal trachea before entering the left lung in a pulmonary artery sling (PAS) [43].

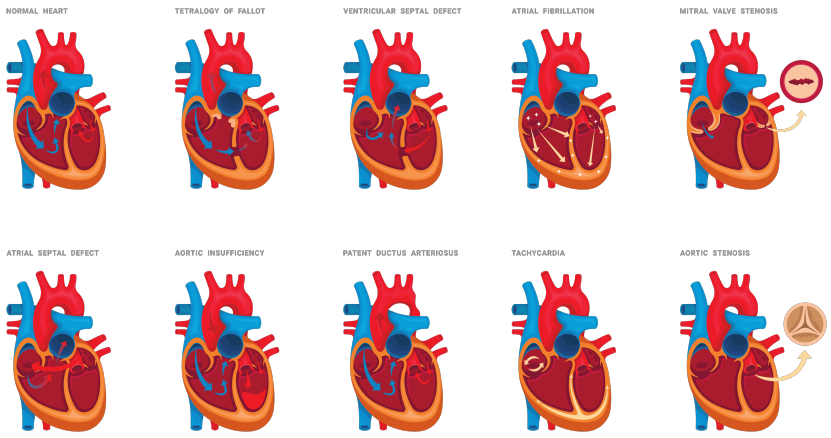


Figure 2.11: Different types of congenital heart disease. Image purchased on Canva Pro platform.

2.2.3 Ventricular Hypertrophy

Ventricular hypertrophy is a serious condition characterized by abnormal enlargement of the heart muscle surrounding the left or right ventricle. Ventricular hypertrophy is divided into two types: left ventricular hypertrophy (LVH) and right ventricular hypertrophy (RVH). LVH is a condition in which the mass of the left ventricle increases,

either due to increased wall thickness or increased cavity size, or both [13]. The wall of the left ventricle thickens in response to pressure overload, while the chamber dilates in response to volume overload [174]. RVH is defined as either concentric or eccentric enlargement of the right ventricle as a result of increased workload and subsequent hypertrophy of the cardiac muscle cells that make up the right ventricle [15]. If left untreated, the disease progresses from an adaptive state (compensatory hypertrophy) to a maladaptive state (progressive loss of contractility), eventually leading to right heart failure. RVH is caused by a chronic overload of the right ventricle, which is most commonly caused by the pulmonary hypertension.

2.2.4 Aortic Aneurysm

An aneurysm is defined as a focal and persistent dilatation of an artery more than 50% larger than its expected normal diameter [30]. The exact definitions of aneurysms vary depending on their anatomic location, with pathology varied and unique to that location, including the limb, splanchnic and cerebrocervical arteries. Extracranial arterial aneurysms can occur at any site in the aorta, but are most commonly found in the infrarenal segment and account for approximately 30% of aortic aneurysms[155].

The aortic wall consists of three layers: the intima, the media and the adventitia. The intima is composed of endothelial cells. The medial wall of the artery consists of smooth muscle cells surrounded by elastin, collagen and proteoglycans. It is responsible for the structural and elastic properties of the artery. The adventitia is composed mainly of collagen. Aneurisms result from degradation of the major structural proteins of the aorta (elastin and collagen), usually as a result of degeneration of the medial layer, leading to dilatation of the vessel lumen and loss of structural integrity. Although an imbalance between collagen formation and degradation is thought to be responsible for rupture of the aortic wall, many of the underlying physiopathological mechanisms underlying aneurysm formation and rupture remain unknown. If left untreated, the aortic wall continues to deteriorate and eventually becomes unable to withstand the forces of luminal blood pressure, leading to progressive dilatation and rupture [90, 184]. The risk of aortic rupture increases with increasing aortic diameter and this catastrophic event is associated with a 50-80% mortality rate [35, 155]. Aortic aneurysms can develop in both the thoracic and abdominal aorta, as shown in Figure 2.12. AAAs are further subdivided into suprarenal or paravisceral aneurysms when they involve the visceral arteries [81], pararenal aneurysms when they involve the origins of the renal arteries and infrarenal aneurysms when they begin lower than the renal arteries, as shown in Figure 2.13.

AAAs are typically asymptomatic and aneurysms are frequently discovered during another examination. Due to the clarity of the

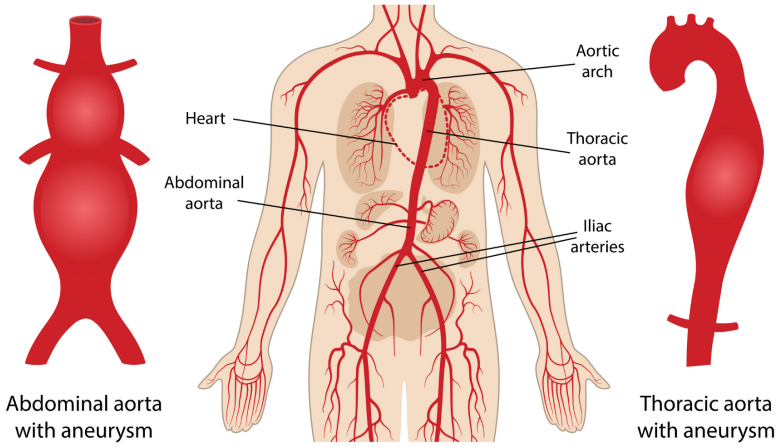


Figure 2.12: An illustration of normal aorta, thoracic aortic aneurysm and abdominal aortic aneurysm. Image purchased on Canva Pro platform.

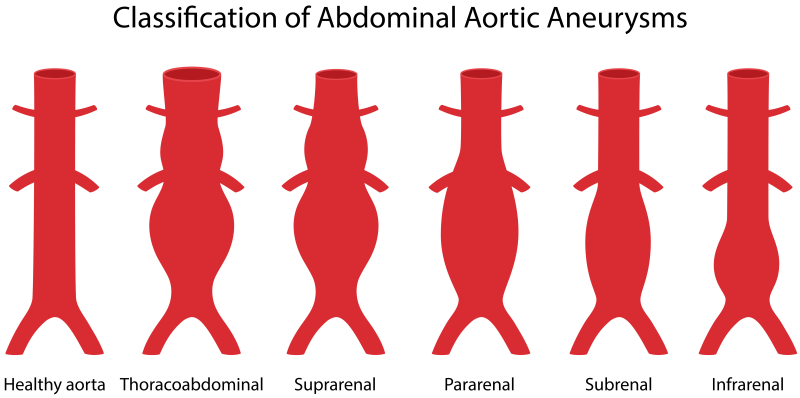


Figure 2.13: Different types of an abdominal aortic aneurysms. Image purchased on Canva Pro platform.

images of the aorta and the ability to detect the size and shape of an aneurysm, the CTA scan is the gold standard imaging modality used in AAA management. Once detected, the subsequent management of an abdominal aortic aneurysm is determined by the aneurysm's size or diameter and the risk of aneurysm rupture versus the risk of operative mortality. There are widespread agreement that the risk of rupture is negligible for very small (3-3.9 cm) aneurysms. As a result, these aneurysms do not require surgical intervention and are monitored [110]. In general, the size threshold for intervention is 5.5 cm for men and 5 cm for women, or a 12-month growth rate of at least 10 mm in both sexes [155]. Patients who are symptomatic or have experienced a rupture are seen immediately. There are two possible types of interventions: open repair and endovascular aneurysm repair (EVAR) [8].

AAAs Treatment

Traditionally, repair of an aortic aneurysm is performed through open surgery. This involves cutting out the AAA and implanting a prosthesis or stent-graft. The stent-graft is a metal tube made of synthetic material that replaces the AAA and restores continuity to the aorta. To repair AAA, an abdominal incision is made and then the artery's blood flow is stopped with clamps above and below the aneurysm [29]. The aneurysm is then opened and the stent graft is sewn in. The prosthesis is covered by the same wall of the aneurysm sac. Although the open repair requires a longer recovery period, the patient is unlikely to need further surgery, no follow-up is required and the patient's long-term life expectancy is high. It is also a technique that can be adapted to complicated anatomies. However, it is a major procedure that often results in significant perioperative morbidity and mortality due to hemodynamic instability, patient comorbidities, surgical exposure and aortic clamping with associated lower body ischemic injury. A minimally invasive procedure called endovascular aneurysm repair is an alternative to open surgery (EVAR) [140]. In the EVAR procedure, an endograft is inserted and fixed in place using a catheter inserted through the femoral arteries. An endograft is a self-expanding type of metallic vascular stent that is encased in tissue to form a closed tube. The proximal and distal ends of the endografts are equipped with hooks for attachment to the inner wall of the blood vessel. The EVAR procedure isolates the damaged aneurysm wall from the blood circulation, which flows through the body of the endograft and the isolated walls, forming an intraluminal thrombus that shrinks if the procedure is successful. Accurate aortic and aneurysm sizing and preoperative planning are critical to successful initial and long-term outcomes after EVAR, for which CTA is the imaging modality of choice.

EVAR has proven advantages over open surgery in terms of reduced preoperative morbidity and mortality and shorter procedure and recovery time. In addition, it allows intervention of high-risk patients who cannot be treated by open surgery. However, the postoperative 2-year survival rates are almost identical for both procedures [105]. While open repair removes the aneurysm wall and replaces it with a stent graft, EVAR excludes the aneurysm from the circulation but does not remove it. In the long term, this can lead to EVAR-specific complications called endoleaks. Endoleaks occur when blood continues to flow within the excluded aneurysm sac or thrombus after endovascular aneurysm repair. If left untreated, the aneurysm may re-expand due to the pressure, increasing the risk of rupture and requiring reintervention in some cases. Thus, some endoleaks may occur as a result of poor EVAR planning or endograft defects. These problems could be solved by a better understanding of the different endograft models and their modes of operation, as well as by more accurate preoperative sizing of the endografts.

2.3 Cardiac Imaging Modalities

Cardiovascular imaging is a critical component of the diagnosis and prognosis of CVDs. The most commonly used devices for performing imaging are ultrasound (US), magnetic resonance (MR), computed tomography (CT), positron emission tomography (PET), single-photon emission computed tomography (SPECT). They all have a great capacity to capture detailed information about the anatomy and soft tissues of the body. This significantly improves our understanding of the healthy and pathological anatomy of various organs. In this Thesis, we focus on images examined by CT and MRI. Although both: CT and MRI contain detailed information about the anatomy and soft tissues of the body, the details vary according to the acquisition techniques used.

2.3.1 Computed Tomography

CT is a type of imaging that uses X-rays to collect structural and functional data about the human body. Reconstruction of the CT image is based on an X-ray absorption profile. X-rays are electromagnetic waves used in diagnosis due to the property that all substances and tissues absorb X-rays differently. CT scans use tiny X-rays aimed at a patient and rapidly rotated around the body to produce a digital signal. A computer connected to the X-ray machine processes this signal to produce a sequence of cross-sectional images of the patient's body [150].

As shown in Figure 2.14, a patient lies in a tunnel equipped with a scanner. The scanner or rotating gantry (ring) consists of an X-ray emitter at 180 degrees across the receiver. The patients' bed slowly moves through the tunnel and stops, after which the scanner circles the patient and X-rays are beamed and received at many points along the tunnel. Each time the bed moves, the scanner circles again. In this way, while the patient remains in one position, large amounts of data can be acquired quickly and correctly.

The image data are provided as grey levels depending on the amount of ionizing radiation absorbed or attenuated. The linear attenuation coefficient expresses the amount of radiation lost through an absorbing material of a given thickness [57, 149]. This value grows in proportion to the atomic number and density of the material. This difference in linear attenuation coefficients between tissues produces contrast in X-ray images. Tissues with low attenuation (e.g., air) appear dark because they absorb very little radiation, so most of the radiation is transmitted to the detector. Tissues with a high attenuation coefficient (e.g., bone) appear light on the image because most of the radiation is absorbed and only a small portion is transmitted. When interpreted, the left side of the image corresponds to the patient's right anatomy and vice versa. CT images can be viewed individually or in a volume

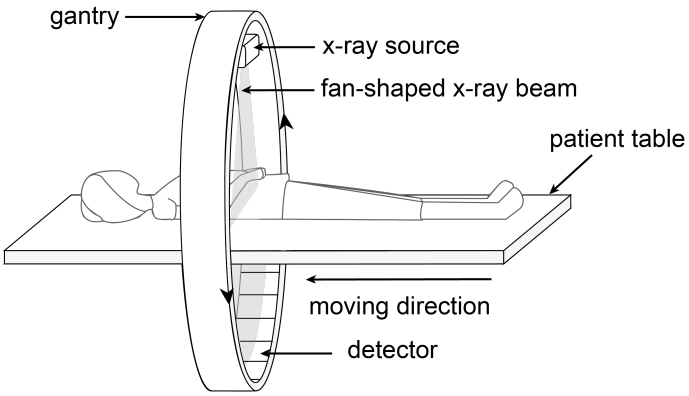


Figure 2.14: A diagrammatic representation of computed tomography (CT). The absorption of numerous x-ray projections from various angles is used to rebuild a CT image slice. The spinning gantry and patient table both move in unison to acquire CT slices. By repeating the image acquisition method, a sequence of CT images is obtained. Image source: Clara Tam [164]

or three-dimensional (3D) format. When viewing a CT volume, the individual volume elements are referred to as voxels rather than pixels (image elements) in a two-dimensional image slice [164].

CT can provide detailed anatomical information about ventricles, vessels, coronary arteries and coronary calcium score thanks to recent rapid developments in CT technology. Cardiovascular CT imaging consists of two procedures: (1) coronary calcium scoring using noncontrast CT and (2) noninvasive coronary artery imaging using contrast-enhanced CT. Non-contrast CT imaging generally uses the density inherent in the tissue. This makes it easy to separate different densities with different attenuation values, such as air, calcium, fat and soft tissue. Noncontrast CT imaging is a low-radiation approach to identify the presence of coronary artery calcification within a single breath [38]. In contrast, contrast-enhanced CT imaging of the coronary arteries, valves, effusions, pericardium or pacemaker leads is performed using contrast media, such as a bolus or continuous infusion of a high concentration of iodine-containing contrast agent [148, 116]. Coronary CT angiography (CTA) can image both the arterial lumen and wall, allowing noninvasive assessment of the presence and size of noncalcified coronary plaques [39]. AAAs are often asymptomatic and aneurysms are frequently detected by another examination. Because the CTA scan provides detailed images of the aorta and is able to detect the size and shape of an aneurysm, it is the gold standard in AAA care. It is used as a screening test when ultrasound images are inadequate, as a diagnostic test when rupture is suspected and as part of the preoperative workup for AAA repair [64, 162]. Although CT offers a wealth of benefits, it is still quite expensive. Nonetheless, CT is the

gold standard for diagnosing a variety of diseases.

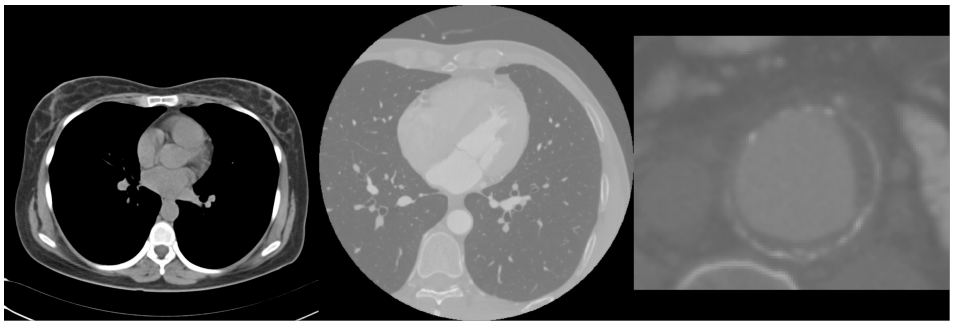


Figure 2.15: Example of non-constrast cardiac CT image, contrast CT cardiac image and CT image with cropped AAA.

2.3.2 Magnetic Resonance Imaging

Cardiac MRI is a noninvasive imaging technique that, unlike CT, does not require ionizing radiation. MRI images are formed when radiofrequency energy (RF) is exchanged between the patient's body and the imaging device. This is possible due to the inherent magnetic properties of the human body, more specifically hydrogen atoms (protons), which are particularly abundant in tissues. A proton has an inherent spin angular momentum that rotates about its axis at a constant speed. When a strong external static magnetic field is applied, the rotational axes of the protons align with the applied field. Since protons have angular momentum, they will rotate perpendicular to the applied field. The precession rate of the proton, shown in Figure 2.16, is directly proportional to the intensity of the magnetic field, as shown by the Larmor frequency equation (resonance frequency):

$$\omega_o = \frac{\gamma \cdot B_o}{2 \cdot \pi} \tag{2.1}$$

where ω_o is the Larmor frequency in megahertz, B_o is the magnetic field in tesla and γ is the nuclear gyromagnetic ratio in radians per second per tesla [164].

The technician administers an RF pulse during the imaging procedure to perturb the protons and bring them into 90 or 180 degree alignment with the static magnetic field. When the precession rate or frequency of the protons matches the frequency of the applied RF pulse, the protons begin to resonate and absorb some of the energy of the RF pulse, placing them in an excited state. At the same time, the protons are in the excited state and their electromagnetic energy increases. When the radiofrequency pulse is interrupted or turned off, the protons realign with the magnetic field, releasing additional electromagnetic energy. This is called relaxation. The RF radiation is then picked up by a receiver in the MRI scanner to create the magnetic resonance images. Because different tissues in the body contain different amounts

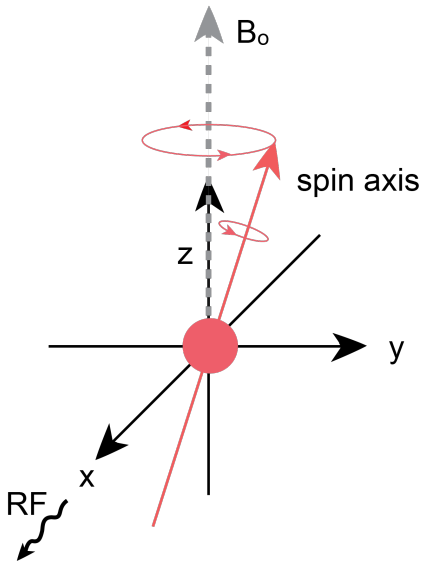


Figure 2.16: Precession of protons in a static magnetic field. Image source: Clara Tam [164]

of water, MRI uses the electromagnetic fields from the nuclei (protons) to determine the density and shape of the tissues in the body. One can automatically distinguish between different tissues in the body based on how quickly the protons release their extra energy after the applied RF pulse is turned off. The relaxation rate (or time) is the most important factor contributing to the contrast between different tissue types in an image. In addition, the strength of the RF signal is critical to the quality of the image.

In MRI, relaxation times are divided into T1 (longitudinal relaxation time) and T2 (transverse relaxation time). T1 is the time constant used to calculate the time it takes for the spinning protons to realign with the external magnetic field. That is, the time required for the longitudinal signal to recover 63% of its magnetization value. T2 is the time constant at which excited protons lose phase coherence with each other [14, 93]. This is the time required for the transverse magnetization to drop to 37% of its original value [24]. Image sequences are obtained by changing the RF pulses delivered to the same image slice during T1 and T2 relaxation pulse sequences. The time to echo (TE) is the time interval between the emission of an RF pulse and the reception of the signals emitted by the patient's body, called echoes. The repetition time (TR) indicates the interval between successive pulse trains applied to the same image section. Time to echo (TE) is the time interval between the administration of an RF pulse and the reception of echoes emitted by the patient's body. T1-weighted and T2-weighted images are generated using MRI sequences that exploit intrinsic T1 and T2 relaxation features. T1-weighted sequences have short TR and TE times, whereas T2-weighted sequences have longer TR and TE times. T1 images show proton density in the fatty tissues of the body, such

as the bone marrow of the vertebral bodies. Because cerebrospinal fluid does not contain fat, it appears black on T1-weighted images. T2 images, on the other hand, show the proton density of tissues containing fat and water. Therefore, cerebrospinal fluid appears bright on T2-weighted scans.

Cardiovascular MRI allows precise assessment of cardiac structure and function (e.g., cine imaging) and diseased tissue such as scarring. Because images can be acquired in any orientation, they can be acquired in specific anatomical planes. Because of these features, experts have developed a variety of techniques that provide varying amounts of information. Cine-CMR, flow-CMR, tagged-CMR, late gadolinium enhancement (LGE) and perfusion-CMR are the most commonly used [129]. Cine-CMR aims to achieve excellent spatial and temporal resolution while maintaining high contrast between tissues. Typically, a single sample has 20-30 consecutive images, each corresponding to 20-30 time periods during the cardiac cycle. Each image contains between ten and fifteen sections from base to apex. Images are usually acquired along two axes: the long axis and the short axis, as shown in Figure2.17. The long axis (LAX) runs throughout LV from base to apex. The cuts of the short axis (SAX) are perpendicular to the cuts of the long axis (LAX). Cine-CMR is usually used to calculate global functional indices such as stroke volume and ejection fraction because the image sequence loop accurately captures the dynamic process of an entire cardiac cycle during a respiratory pause. On the other hand, cine MRI has the following disadvantages: it is not real-time, it is expensive and it has lower resolution than CT.

2.4 Conclusion

In this chapter, we have given an overview of the cardiovascular system, the heart and the cardiac structures. We gave a brief overview of their anatomy and functional indices. The description of the LV and RV functional indices included their definitions, calculation methods, corresponding normal ranges and exemplary applications for the diagnosis of CVD. We briefly described CVDs relevant to our research and provided a brief overview of the characteristics of CT, MRI, and Cine-MRI imaging modalities. In particular, we have attempted to introduce the medical concepts necessary to understand the need for the segmentation methods presented in this work. The main challenge in cardiac imaging is the constant motion of the heart due to its contraction and respiration. Therefore, images must be acquired quickly and synchronously with the heart rhythm to avoid motion blur. Although both CT and MRI provide valuable information about the anatomy and soft tissues of the body, the details differ depending on the imaging technique used.

A variety of imaging techniques combined with some form of visualization provide valuable information about the anatomy and function of

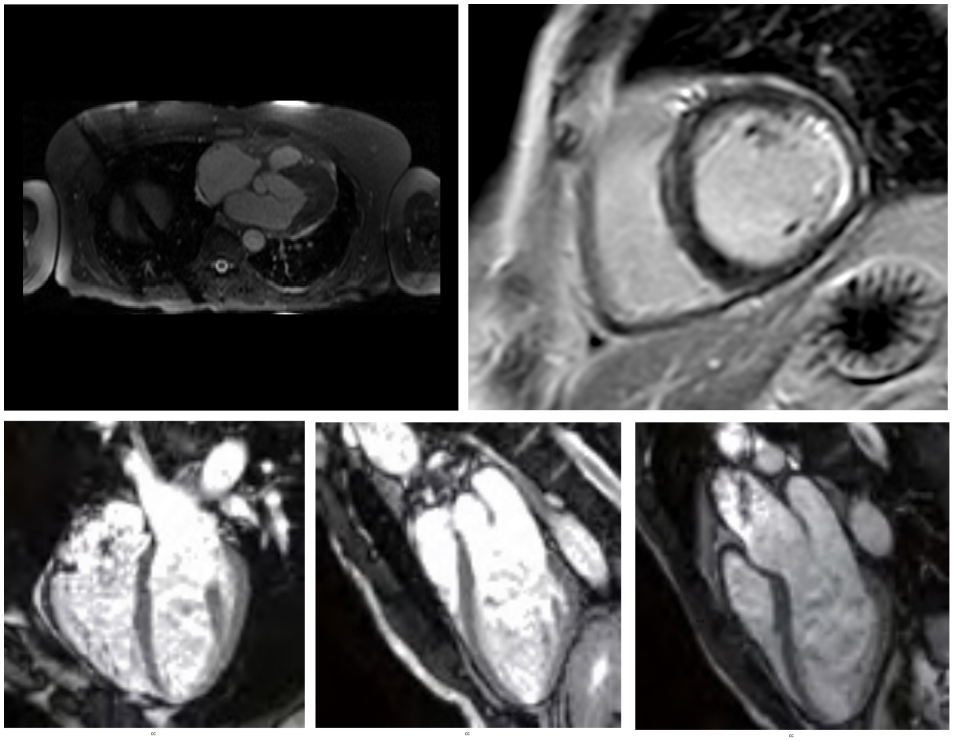


Figure 2.17: Top row left: Steady-state free precession MRI (SSFP). Top row right: Late gadolinium enhancement magnetic resonance imaging (LGE-MRI). Bottom row, from left to the right: cine MRI 4-chamber view, 2-chamber view, 3-chamber view.

the heart. Nevertheless, this process is often manual or semi-automated and requires a lot of physicians' time. In order to provide a complete assessment of the heart, more sophisticated algorithms are required that allow automatic extraction and analysis of cardiac structures. This is made possible by image segmentation algorithms. The development of image processing algorithms is challenging because images acquired with different imaging devices differ significantly. Thus, it is important to summarize and highlight some challenges in cardiac structure segmentation that arise from the differences between imaging devices.

Segmentation of the whole heart and ventricles in CT and MRI images is challenging because:

- The cardiac organ with its numerous chambers and large vessels is geometrically complex.
- The geometry of the heart varies considerably between people and between different heart states within the same subject. Segmentation using a previous model created from a training set of data from healthy subjects worked well for healthy subjects. However, when separating complex data, the same segmentation algorithm gave significantly worse results.

- Due to the intensity distributions of certain anatomical substructures, such as texture patterns, some boundaries between anatomical substructures are visually ambiguous.
- Due to the complicated motion and blood flow within the heart, imaging data often contains significant motion artifacts, intensity inhomogeneity and noise. This contributes to the boundary delineation not being smooth and unsatisfactory.

Furthermore, segmentation of the LV, MR and Myo in cine MRI images is challenging due:

- Insufficient contrast between the myocardium and surrounding tissues (high contrast between blood and myocardium).
- Due to blood flow, there are brightness differences in the cavities of the left and right ventricles.
- Due to limited CMR resolution along the longitudinal axis, there are inhomogeneous partial volume effects.
- There is inherent noise due to motion artifacts and cardiac dynamics.
- Variability in shape and intensity of cardiac structures in different patients and diseases.

Segmentation of AAA in CTA images is challenging due:

- Similarity between the intensity values of the aneurysm thrombus and those of some adjacent tissues, leading to segmentation leaks due to the blurred boundaries of the thrombus.
- In some cases, the thrombotic surface is locally obscured due to its noncontrast nature.
- The geometric structure of the aneurysm is non-uniform, making it impossible to approximate the thrombus with a simple geometric model.

Knowledge of the above characteristics provides guidance in the development of image processing methods for segmenting the heart from different imaging modalities.

CHAPTER 3

Related Research

Advances in information technology are having a significant impact on healthcare, particularly through the incorporation of artificial intelligence. Machine learning is a subset of artificial intelligence that consists of algorithms and statistical models that are used to perform a task without explicit instructions. They create mathematical models from a set of data that includes inputs and provides the desired output. Machine learning is divided into two main categories: unsupervised learning and supervised learning. Unsupervised machine learning algorithms learn the desired output from the data without using predefined labels. They learn the inherent structure of the data by discovering certain patterns through repeated experiences, such as grouping data points or clustering. Unsupervised learning uses cluster analysis techniques to form these groups with common attributes. Autoencoders [10], deep belief networks [68], k-means and generative adversarial networks [51] are examples of unsupervised learning algorithms. Supervised machine learning methods transform the input data of an algorithm into corresponding outputs by identifying the correlation between input and output based on statistical or data-driven rules learned from the data. In unsupervised learning, the goal is to derive relationships and patterns without prior knowledge, i.e., the dataset does not contain information about the output labels. In contrast, supervised learning methods learn the underlying patterns and representations from the data to identify their valuable properties and latent spaces. In other words, supervised learning is based on the presence of previously labelled ground truths that are used for model training.

The outline of the chapter is structured in the following manner. Section 3.1 gives an introduction to deep learning mechanisms and networks. It presents an overview of the most significant deep learning CNN network architectures and mechanisms: U-Net architecture,

ResNets variants, feature reuse mechanisms and autoencoders to highlight novelties and the main focus of this Thesis. Section 3.2 presents an overview of related researches. It is directed to prior methods that deal with the whole heart segmentation from CT and MRI images, segmentation and quantification methods for LV, RV and Myo as well as AAA segmentation. The most commonly used evaluation metrics for medical image evaluation are also described. Section 3.3 recaps known challenges and limitations of the previous researches. Finally Section 3.4 gives concluding remarks.

3.1 Deep Learning Mechanisms and Networks

Deep learning algorithms are a subset of machine learning that uses multiple layers of neural networks while processing data. Their ability to map multiple levels of semantic information and their scalability to large amounts of data quickly established them as the leading technology for medical image analysis [54, 25]. Unlike other machine learning methods, deep learning algorithms will continue to improve when given more data. When the dataset is larger, the algorithm is more sensitive to slight variations. Deep learning algorithms either learn from present outcomes based on reference datasets or directly from the data. Until the advent of deep learning algorithms, feature extraction was performed manually. The challenge with manually managed features is choosing which are most appropriate for specific applications. Testing the many types of feature descriptors and their numerous combinations would require a large amount of time and effort. A typical machine learning algorithm is trained to process and extract manually specified image attributes (e.g., texture, color, shape, edge patterns, pixel spatial relationships, pixel intensities). On the other hand, deep learning automates the feature extraction process, obviating the requirement for pre-training manually constructed feature extractors. Another distinctive property of deep learning is its hierarchical architecture for feature learning. A deep learning algorithm emulates this behaviour by taking a complicated and abstract task, such as differentiating a triangle and breaking it down into numerous levels of simpler tasks, much like the human brain does when it learns to correlate distinct qualities to identify a specific item.

Deep learning algorithms are statistical models composed of deep artificial neural networks. In the following few subsections, we detail the components of a neural network and review important deep learning techniques relevant to our work.

3.1.1 Feed Forward Neural Network

The concept of a neural network was inspired by neuroscience, where a single node (artificial neuron) in a neural network is representative and mimics aspects of a biological neuron [170]. A neural network consists of a collection of nodes interconnected in various architectures to enable communication between them. It can be viewed as a mathematical function that attempts to approximate the function represented by our data. It computes the error between the predicted and expected outputs and minimizes this error during the training process.

Let $y = f(x)$ be the underlying function that outlines the relation between variable y and explanatory variable x, the neural network is a non-linear mapping $f(x;\theta)$ that approximates this value of y by learning parameters θ. When a neuron receives an input satisfying the specific threshold value, it gets activated and forwards the input to the adjacent neurons. If the sum of the inputs that is collected from multiple adjacent neurons exceeds their threshold, they are activated too. Information flows through the network following such a schema. Another way of describing a neuron is the perceptron. In statistical terms, the perceptron is a linear and binary classifier. Figure 3.1 shows a perceptron with inputs coming from the left-hand side.

Now, lets consider an input vector x of size d. The linear function of the percepton processes the sum of the input and weights w with a bias added to it. The intercept of the linear function is now determined. A combination of many perceptrons produces a network that resembles a neural network. The resulting model is usually either called a feed forward network (FFN) or Multi-Layer Perceptron (MLP). Figure 3.2 shows a simple neural network of multiple layers with multiple perceptrons.

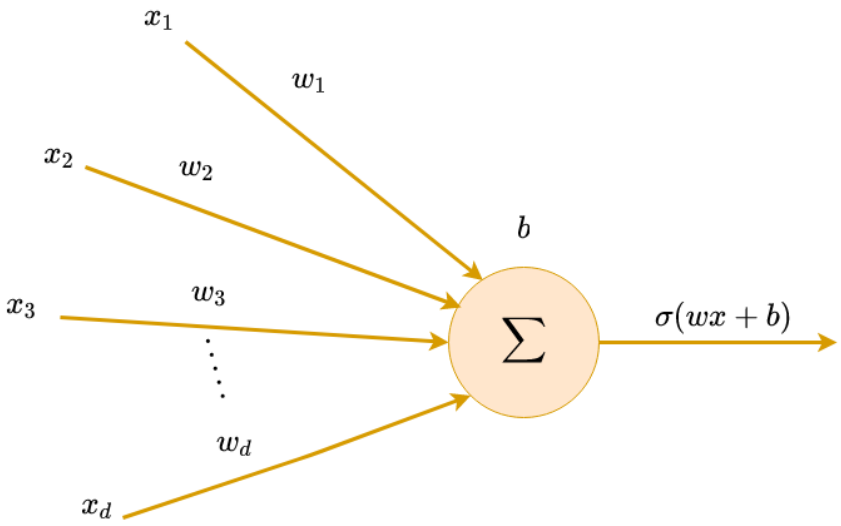


Figure 3.1: An illustration of perceptron.

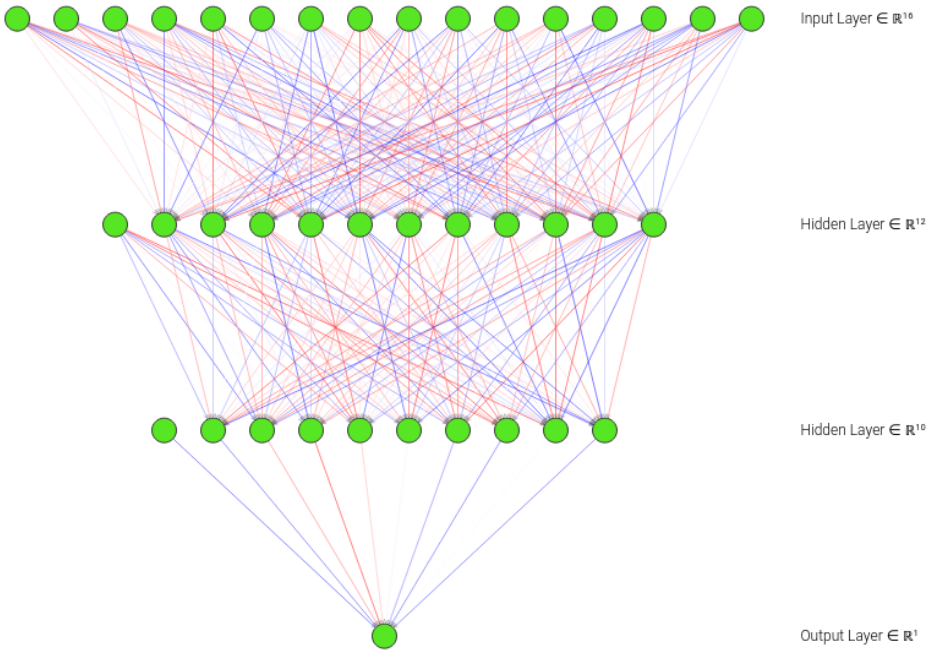


Figure 3.2: An illustration of Multi-Layer perceptron.

The first layer of the neural network is called the input layer, the last layer that contains the output is called the output layer and the layers in between are called hidden layers. The computation is vectorised, therefore the variable x is of the actual dimensions $n \times d$, with n denoting the number of observations and d the dimensionality of the input. The forward pass can be mathematically written as:

$$h(x) = \sigma(x^T \cdot w_1 + b) \tag{3.1}$$

and

$$y(h) = \sigma(h^T \cdot w_2 + b) \tag{3.2}$$

where w_1 and w_2 denote the weights and b_1 and b_2 denote the bias in the respective layers. The *Sigmoid* activation function denoted by σ is a mapping of the output between $(0, 1)$ in the following equation:

$$\sigma(x) = \frac{1}{1 + exp(-x)} = \frac{exp(x)}{exp(x) + 1} \tag{3.3}$$

In the subsequent step, we define a loss function and apply backpropagation to update the weights. The error in the output of the neural network is defined by the loss function. The loss function used depends on the application, for instance, a classification, regression or as in our case, segmentation problem. An assumption can be made that the error is normally distributed about the prediction y. Hence, The Maximum Likelihood Estimate (MLE) for the normal distribution can be optimised by maximising the Mean Squared Error (MSE), where y

denotes the true value and $\hat{y}$ denotes the estimated value.

$$\epsilon(\hat{y}) = \frac{1}{n}\sum i = 1^n(\hat{y} - y)^2 \tag{3.4}$$

The gradient can now be computed with respect to the corresponding weights and updated. The learning rate α decides the magnitude of change in the weights after each iteration or epoch, i.e, one pass of the neural network training over the training dataset. For the purpose of application more complex optimisers along with momentum are used for the training to converge faster to a solution. One gradient update can be shown in the gradient of the error function with respect to its weights, with mathematical definition as follows:

$$\nabla\epsilon = (\frac{\delta\epsilon}{\delta w_n}, \frac{\delta\epsilon}{\delta w_{n-1}}, ..., \frac{\delta\epsilon}{\delta w_1}) \tag{3.5}$$

while the chain rule of partial derivatives can be used to compute the gradient. For the w_2 , gradient is

$$\frac{\delta\epsilon}{\delta w_2} = \frac{\delta\epsilon}{\delta\hat{y}}\frac{\delta\hat{y}}{\delta h}\frac{\delta h}{\delta w_2} \tag{3.6}$$

Hence, the weight can be updated according to the following expression:

$$w_2' = w_2 - \alpha \odot \frac{\delta\epsilon}{\delta w_2} \tag{3.7}$$

where $\odot$ defines the element-wise product and α is the learning rate which determines the magnitude of change of the weight update. This training procedure is shown in Figure 3.3.

Sometimes, the updates to the structure of the neural network may be required which changes the analytical solution for gradient updates. The application of numerical optimisation techniques can allow for such a change without requiring the derivation of the analytical solution for such a gradient update. Deep neural networks are susceptible to the vanishing gradient/exploding gradient problem [59], which is the reason for the application of more complex types of neural networks.

3.1.2 Convolutional Neural Network

A CNN is a special type of feed forward neural network that uses grid-like data with a spatial correlation between neighborhood data points, like 2-D images, 3-D volumes, or time-series data. Instead of using a weight for each pixel of the image, a CNN uses a filter over the inputs, i.e, a sliding window with a certain stride over the pixel intensities to create an intermediary output that is a function of the input and the filter, this process is known as convolution. The intensity of each pixel of the image (voxel in 3D case), is generally rescaled and mapped to a $(0, 1)$ range. It is noted that, zero-padding may be performed on the input to allow for convolution involving all the input

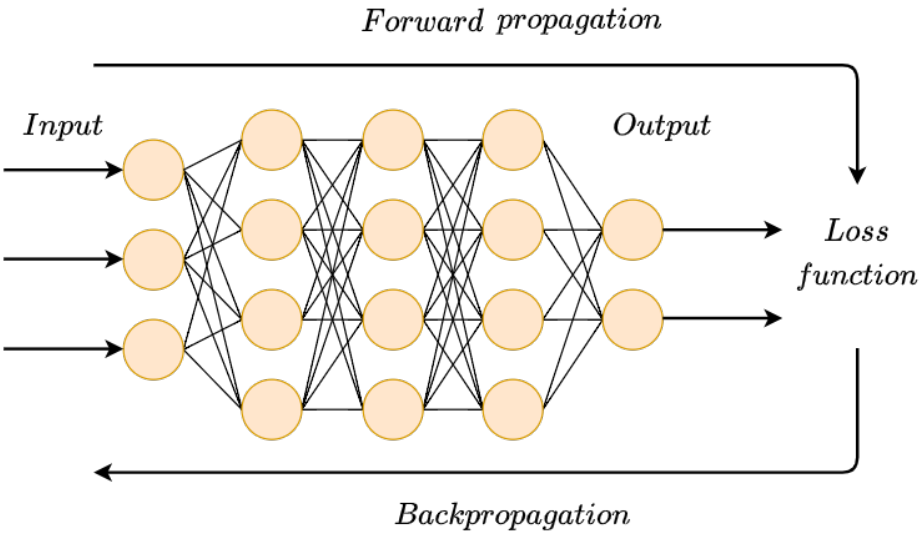


Figure 3.3: A neural network's training procedure. In an iterative process, neural networks learn by propagating information forward and backward from a loss function. By updating the weights during backpropagation, the network aims at minimizing the loss function. Forward propagation returns the information from the output back to the loss function.

pixel or voxel intensities, i.e., 0 intensity pixels are added around all the edges and corners. Other forms of padding are also adopted depending on the application. During each convolution step as described above, a product of the input and filter is computed. Figure 3.4 shows an example of 2D convolution (can be seen as a one image pixel). Here, input shape is 1×4 and 4×1 which results in an output of shape 1×1. The stride of 1, moves the sliding window of 2×2 to the next position with the output being computed on the right hand-side. Increasing the strides will result in a smaller output size. Also, it can be noted that the output will always be smaller than the input size.

Furthermore, the convolution operation in the 2D space can be mathematically expressed as:

$$y_{ij} = \sum_{a=0}^{[m/s]} \sum_{b=0}^{[m/s]} x_{1+as,j+bs} F_{a,b} \tag{3.8}$$

where x denotes the input at indices i and j, F denotes filter, m denotes size of the filter and s is the stride. Due to convolutional operation's weight sharing capability, distinct sets of features inside an image can be retrieved by sliding a kernel with the same set of weights on the image.

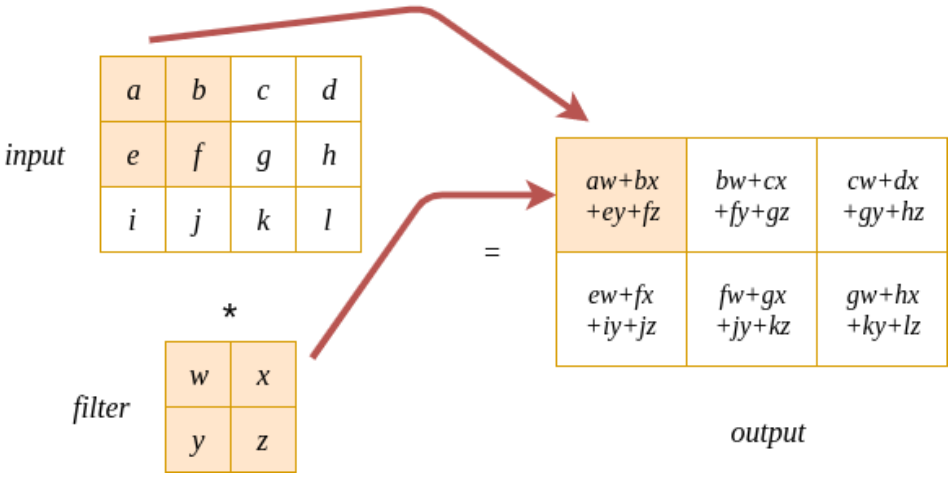


Figure 3.4: Illustrative 2D convolution. The dot product of input and filter results in the output.

Pooling Layer

For the purpose of extracting features from images, edges are of interest. Therefore, a difference between adjacent pixels/voxels can determine the edge, as in the case of foreground vs background. Size reduction of the input is also a concern for faster image processing, can be done by a method called pooling. Pooling or down-sampling is a local operation that use addition to acquire similar information in the neighborhood of the receptive field and outputs the dominant response within this local region. Mathematically, this can be expressed with:

$$P_{ij} = p_f(y_{ij}) \tag{3.9}$$

where P_{ij} is pooled feature-map of $i-th$ layer for $j-th$ input feature map y_{ij} and p_f is pooling operation. An illustrative example of max pooling is shown in Figure 3.5.

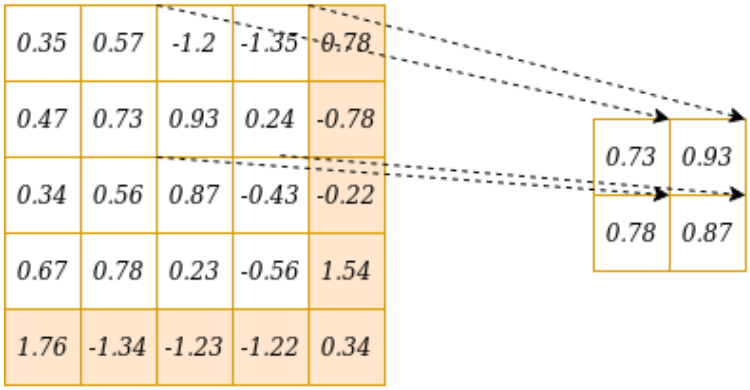


Figure 3.5: Max pooling is being performed on the input by a window of size 2×2. The cases with edges are either padded with zeros or just ignored.

The concept of pooling can also be described as a sliding window of mathematical operation such as *max* applied in each case. Generally, no overlap is considered between the region selected. This can result in the edge cases being overlooked, but padding may also be used. Settings related to pooling and filter size are both considered arbitrarily decided and hence, considered to be a hyperparameter of the neural network. The general structure of CNN architecture is shown in Figure 3.5.

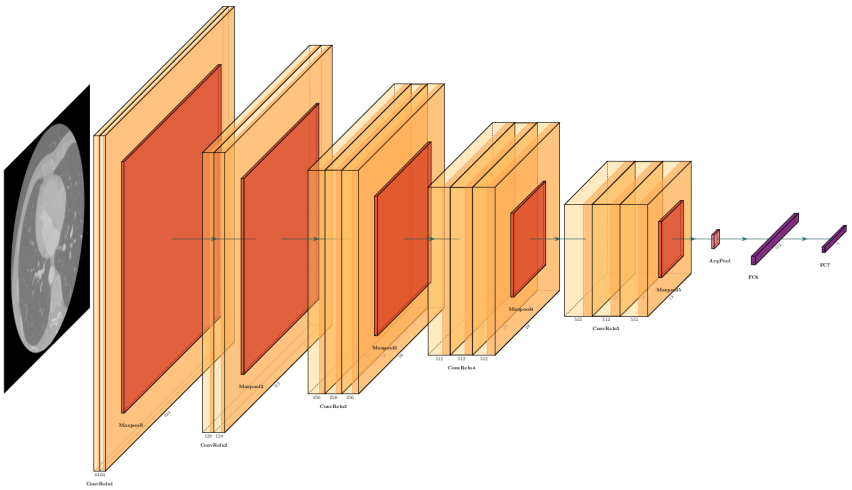


Figure 3.6: An example of general structure of CNN architecture.

Upsampling and Transponse Convolution

For a neural network to generate images or image maps such as in the case of semantic segmentation, it generally involves the application of upsampling from a low resolution to a higher resolution. There are many different methods to perform an upsampling operation such as nearest neighbor interpolation, linear interpolation, bilinear interpolation, trilinear interpolation, bicubic interpolation and various others found in literature [49].

Lets observe what happens if we associate value 1 to 9 other values in a matrix from Figure 3.5. It will be termed a one-to-many relationship. This is similar to going backwards in a convolution operation as shown in Figure 3.7.

Further, a convolution operation can be represented as a kernel matrix that is arranged in the form of matrix multiplication to perform convolution operations, as illustrated in Figure 3.8. This kernel can be arranged as shown in Figure 3.9. Here, each row is defined by one convolution operation and represents a rearranged kernel matrix with zero padding at different places. The matrix multiplication from Figure 3.9 and the flattened column vector of the input leads to the vector of output (4×1) which can be reshaped so that it is (2×2). To increase the size from (2×2) to (4×1), a (16×4) matrix is used. However, the 1 to 9 relationship has to be maintained. The transpose of the

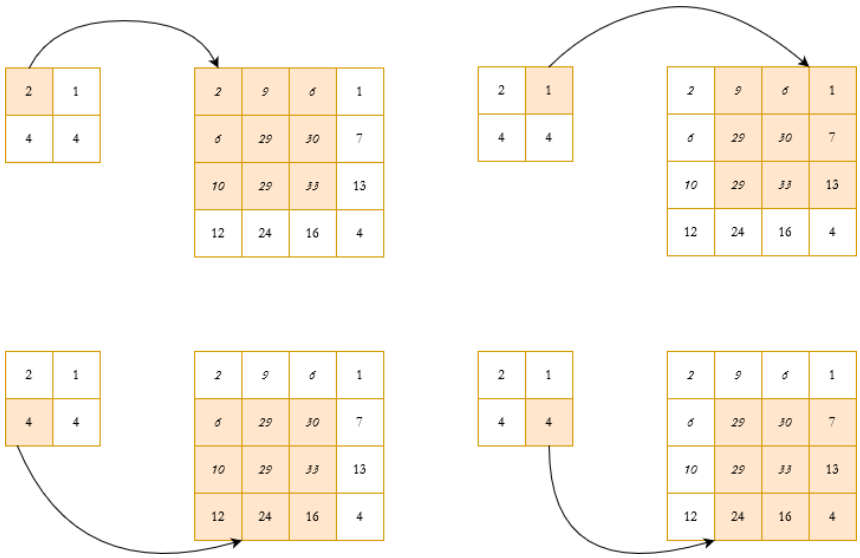


Figure 3.7: Illustration of upsampling.

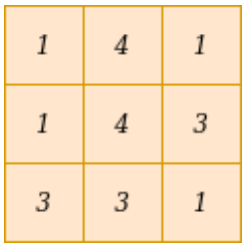


Figure 3.8: Illustration of convolution kernel.

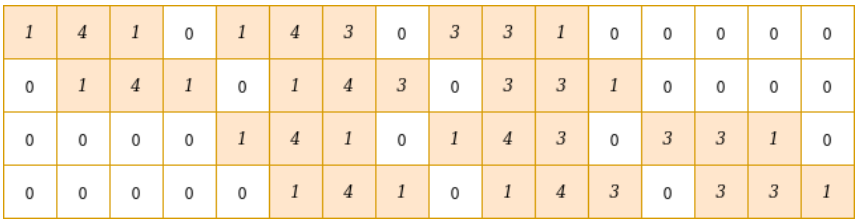


Figure 3.9: Illustration of convolution matrix (4,16).

convolution matrix in Figure 3.9 from (4×1) to (16×4), this can be matrix multiplied with the column vector of the output to generate the output matrix as seen in the Figure 3.7.

Activation Function

Layers of nodes are present within neural networks that learn the mapping of input instances to outputs. In any given node, the sum of products of the inputs and weights is found, are the summed activation of the node. Transformation of this value using an activation function defines the specific output of the node, also known as the activation of the node. The selection of an appropriate activation function can significantly accelerate the learning process. A relatively simple activation function is a linear activation function that involves no transformation.

A neural network consisting of only linear activation functions is simple to train but cannot learn more complicated mapping functions which are required for solving real life problems. The output layer of the network uses linear activation for prediction, such as in the case of regression problems.

The activation function for a convolved feature-map can be mathematically expressed with:

$$A_{ij} = a_f(y_{ij}) \tag{3.10}$$

where y_{ij} is an output of a convolution, to which is assigned activation function $a_f(\cdot)$ which adds non-linearity and returns a transformed output A_{ij} for $i-th$ layer. A non-linear activation function can learn complex features within the data. Some examples of non-linear activation functions are: sigmoid, ReLU, PReLU and hyperbolic tangent [163].

For effective application of backpropagation of the errors along with stochastic gradient descent in order to train neural networks, the activation function must behave similarly to a linear activation function. However, only non-linear activation functions allow learning of complex structures and relationships within the data. Sensitivity to the activation function's summed input is important to avoid running out of values to output. A rectified linear activation (ReL)can be used to solve this problem. A unit or a node that implements ReL activation function is called ReLU or Rectified Linear Unit. Neural networks that use rectifier functions are often called rectified networks. The adoption of ReLU has allowed researchers to implement deep neural networks efficiently.

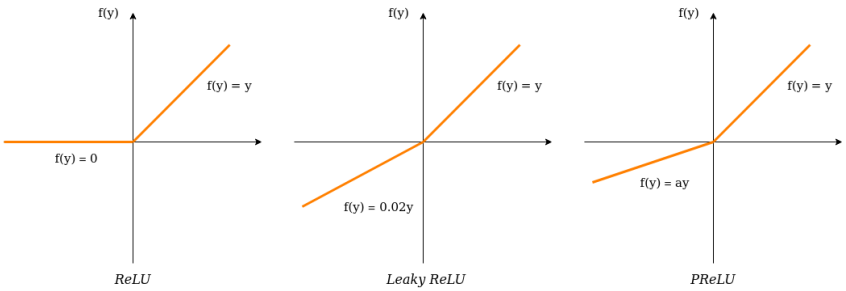


Figure 3.10: An example of general structure of CNN architecture.

ReLU has its own limitations, such as the problem with large weight updates where the summation of the input to the activation function remains negative. A node with such a problem will have an output value of 0 and this is also known as a dying ReLU. Small negative values can be allowed into the function, which effectively reduces the non-linearity of ReLU, such as Leaky ReLU (LReLU), which is a modified ReLU function where the input is < 0. The Parametric ReLU (PReLU) can train to learn parameters that control how leaky the activation function will be and is mathematically defined as follows:

$$loss = \begin{cases} y_i & y_i > 0 \\ a_i y_i & y_i \leqslant 0 \end{cases}$$

where y_i represents any input on the $i - th$ channel and a is the negative slope which is a learnable parametar. If $a = 0$, then f becomes leaky ReLU, if a_i is a learnable parameter then f becomes PReLU.

A CNN will be of limited use if it could only learn with one filter, i.e, it will extract only one type of feature from the input. For instance, an edge detection filter works where the intensity changes drastically (for example: from black to white). Hence, multiple independently operating filters are used for training. Therefore, the CNN consists of an input layer, an output layer and multiply functional layers that transform an input into output in a specific form. These functional layers often contain alternating convolutional layers, pooling layers and/or fully connected layers, as shown in Figure 3.10. Nevertheless, for efficient, end-to-end pixel-wise segmentation, a variant of CNNs called fully convolutional neural network (FCN) is more commonly used, which are discussed in the Subsection 3.1.3.

Loss Functions and Optimization Algorithms

The loss function determines the ability to approximate the ground truth labels for all training inputs. It takes as inputs samples from the training set, weights and biases. A network's goal is to minimize the loss function to as close to zero as possible during training. If the loss value is large, the loss function penalizes the network by frequently changing the weights. Contrary, if the loss value is low, the weights will only slightly change since the network is performing well. Different loss functions are frequently used for specific tasks. For example, in numerical/regression tasks, the mean squared error would be used as the loss function to calculate the differences between continuous variables. On the other hand, categorical tasks would use the cross-entropy loss function to compute the differences between probability distributions. Different tasks have different outputs and are thus modelled by different loss functions.

Using an optimization function, such as stochastic gradient descent (SGD), is the most efficient way to determine the weights and biases. SGD is the oldest and most basic optimization method. SGD algorithm is a stochastic optimization algorithm that employs and generates random variables. It is an iterative method for optimizing a neural network model's objective or loss function. The algorithm iteratively traverses the training set until it converges. The training set is randomly shuffled after each pass. Other optimization algorithms (for example, Adam [88] and Adadelta [180]) use more advanced techniques such as momentum and adaptive learning rates. These techniques allow faster convergence and make hyperparameter tuning algorithms like

grid search or random search easier to implement. They do, however, require more processing time and memory consumption.

Regularization Approaches

Deep neural network's accuracy can continuously converge to perfection during training but degrade during validation. This is because the network has memorized the data, including the noise, too closely. This is referred to as overfitting. Regularization techniques are used to prevent overfitting and improve generalizability in neural network models. Regularization methods that are most commonly used are L1 and L2. Both methods include a regularization term in the loss function of the model. L1 penalizes the weights' absolute value, whereas L2, also known as weight decay, forces the weights to decay towards zero. Data augmentation, dropout, early stopping and batch normalization are some regularization techniques. Data augmentation is a technique for increasing the size of training samples so that the model does not memorize every variation of the data. As augmented samples, invariant properties or expected distortions of the data, such as flips, rotations, scaling, or intensity changes, can be introduced. Dropout works by randomly dropping a certain percentage of neurons in each layer during training and setting the activation of the dropped neurons to zero. All neurons will be active during testing or validation. Different variations of the model can be obtained by dropping some neurons during the training phase. Dropout aids in reducing interdependent learning between nodes. The training of a network can be halted before it begins to overfit. In practice, early stopping is commonly used and is implemented by measuring the accuracy or loss on an isolated test set. When the test performance no longer improves, the training is terminated. Batch normalization is a technique for enhancing the performance, speed and stability of trained neural network models [73]. To achieve batch normalization, a normalization step is performed to fix each layer's input, means and variances for each mini-batch during training. Batch normalization, like dropout, forces each layer of a network to be resistant to variations in its input. At each training step, each hidden unit is multiplied by the standard deviation and subtracted from the mean of the mini-batch. Because the random samples chosen in each mini-batch are different at each step, the standard deviation and mean fluctuate randomly.

3.1.3 Fully Convolutional Neural Network

Fully convolutional neural network (FCN) is introduced by work of Long et al. [104] and represents the most successful and advanced deep learning technique for semantic segmentation. FCNs are a subset of CNNs that do not have any fully connected layers. They are constructed with an encoder-decoder structure that allows them to accept inputs of any size and outputs of the same size.

Let us observe what happens to an input image when it is brought to network input. The encoder converts the input image to a high-level feature representation, while the decoder reads the feature mappings and restores spatial details to the image space for pixel-by-pixel prediction via a sequence of upsampling and convolution operations. Upsampling is accomplished in this case by applying transposed convolutions to the up-scaled feature maps. Additionally, these transposed convolutions can be substituted with unpooling and upsampling layers. FCN with the simple encoder-decoder structure is shown in Figure 3.11.

In comparison to a patch-based CNN for segmentation, FCN is trained on the complete image and applied to it, eliminating the necessity for patch selection. FCN, on the other hand, may be limited in its ability to capture extensive context information in an image for exact segmentation, as some features may be deleted by the encoder's pooling layers. Numerous variants of FCNs have been proposed for feature propagation from the encoder to the decoder in order to improve segmentation accuracy. The U-Net is the most well-known and widely used variant of FCNs for medical image segmentation. The U-Net architecture [145], which is based on the FCN, uses skip connections between the encoder and decoder to recover spatial context loss throughout the downsampling process, resulting in more accurate segmentation.

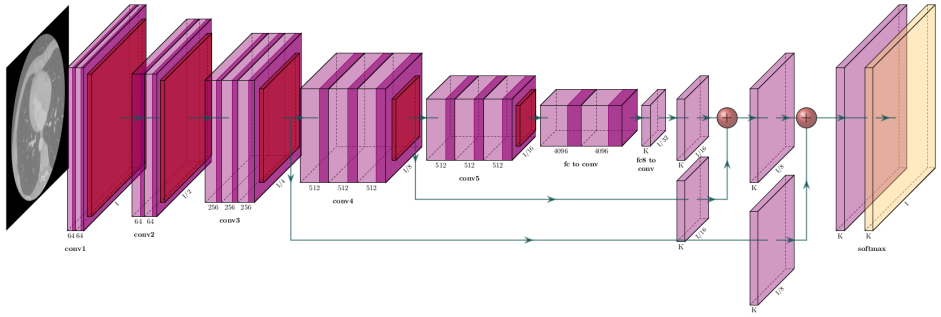


Figure 3.11: An example of general structure of FCN architecture.

U-Net Architecture

Similar to FCN [104], U-Net architecture [145] performs semantic segmentation. The network architecture is symmetrical, with an encoder extracting spatial information from the image and a decoder creating the segmentation map using the encoded features. The encoder is constructed in the conventional manner of a convolutional network. It begins with two 3×3 convolution operations, followed by a max-pooling operation with a pooling size of 2×2 and a stride of 2. This process is repeated four times, with the number of filters in the convolutional layers being doubled after each downsampling. Finally, the encoder and decoder are connected by a series of two 3×3 convolution operations. The decoder first up-samples the feature

map using a 2×2 transposed convolution operation, thus halving the number of feature channels. Then, a sequence of two 3×3 convolution operations is conducted once again. As with the encoder, this sequence of upsampling and two convolution operations is repeated four times, thereby halving the number of filters at each stage. Finally, the final segmentation map is generated using a 1×1 convolution operation. Except for the final convolutional layer, this architecture employs the ReLU activation function. The final convolutional layer employs the Sigmoid activation function to produce the final result. The introduction of skip connections is perhaps the most innovative component of the U-Net architecture. At each of the four levels, the output of the convolutional layer is transferred to the decoder prior to the encoder's pooling operation. These feature maps are then concatenated with the output of the upsampling procedure and the resulting feature map is propagated to subsequent layers. These skip connections enable the network to recover spatial information that was lost during pooling operations. The network architecture is illustrated in Figure 3.12.

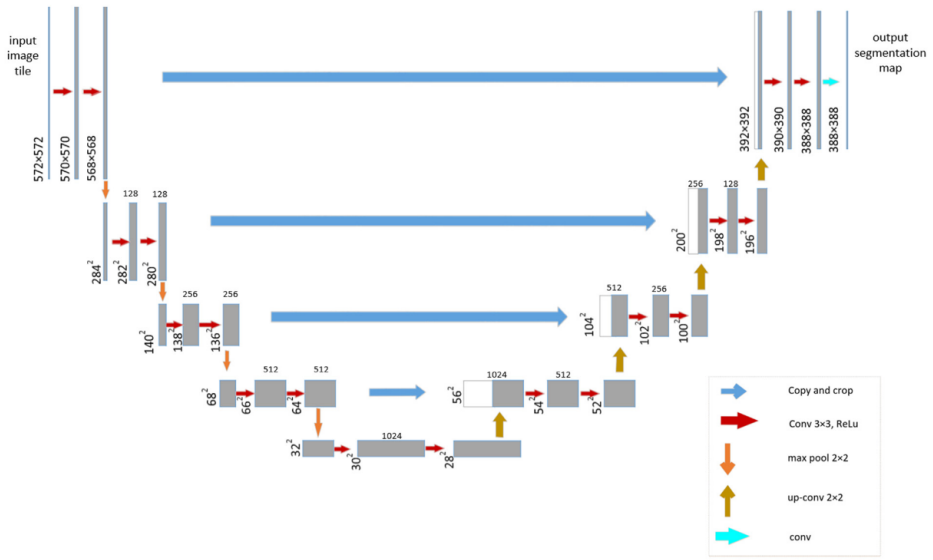


Figure 3.12: The structure of U-Net architecture. Image source: Ronneberger et al. [145]

3D U-Net Architecture

Cicek et al. [186] introduced 3D counterpart of the U-Net architecture. The network structure is similar to U-Net, with one encoding path and one decoding path. Each path has four resolution levels. Each layer in the encoding path contains two $3 \times 3 \times 3$ convolutions and is followed by a ReLU. It uses a maximum pooling layer to reduce dimensionality. In the decoding path, each layer contains a $2 \times 2 \times 2$ deconvolution layer with a stride of 2, followed by two $3 \times 3 \times 3$ convolution layers. Through a skip connection, the layer with same resolution in encoding path is

passed to the decoding path, providing it with original high-resolution features. This network can not only train on a sparsely labeled data set and predict other unlabeled places on this data set, but also train on multiple sparsely labeled data set and then predict new data.

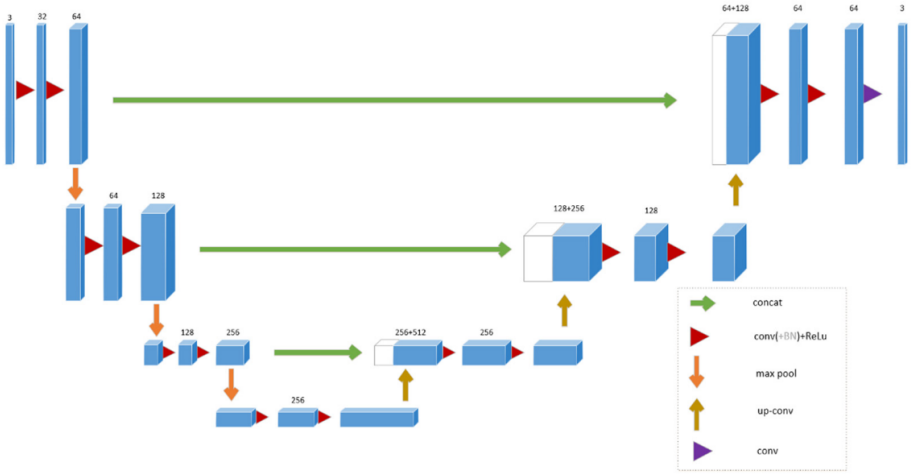


Figure 3.13: The structure of 3D U-Net architecture. Image source: Cicek et al. [186]

3.1.4 Residual Learning

Deep convolutional neural networks (DCNNs) have significantly increased accuracy for various segmentation and classification tasks. However, a common obstacle in training DCNNs is the appearance of vanishing or exploding gradients. As the depth of CNN increases, information about the gradient passes through many layers and it can vanish or accumulate large errors by the time it reaches the end of the network. This problem has been largely addressed using activation functions with a small derivate such as rectified linear unit (ReLU), implementation of gradient clipping, intermediate normalization layers, or careful weight initialization. Nevertheless, with the increasing network depth, accuracy gets saturated and then degrades rapidly. This problem was addressed with the introduction of shortcut connections in residual networks (ResNets) [59].

The ResNet contains multiple stacked residual units. Generally, each residual unit can be expressed with the following two formulations:

$$y_l = H(x_l) + F(x_l, W_l), \tag{3.11}$$

and

$$x_{l+1} = f(y_l), \tag{3.12}$$

where F is residual function, x_l and x_{l+1} denote the input and output of the $l-th$ residual unit in the network, while the output of the $l-th$

residual unit is denoted with y_l. The parameters of the $l-th$ residual unit are denoted as W_l, while the function f refers to the ReLU.

The identity mapping, by which ResNets learn residual function F in regard to $H(x_l)$, can be written as:

$$H(x_l) = x_l \tag{3.13}$$

Therefore, the identity mapping of original residual block attaches an identity skip connection allowing information flow within a residual unit. Numerous ResNet variants have been produced dependent on the amount of layers (starting with 34 layers and going up to 1202 layers). The most widely used form is ResNet50, which consists of 49 convolutional layers and one FC layer.

Nevertheless, when the depth of the network goes very deep, ResNets become challenging to converge. These difficulties were addressed in Pre-ResNets [60] by introducing forward and backward signals that directly propagate from one block to any other using identity mappings after-addition activation and as the skip-connections. This ultimately constructs a new residual block with the BN-ReLU-Conv order. Zagoruyko [181] introduces level-wise shortcut connections to alleviate the learning capability and significantly boost network performance. Moreover, the deep network initialization problem and incompatibility between ReLU and element-wise summation were addressed in weighted residual networks (WNR) [151]. Although deeper residual networks showed performance improvement, diminishing feature reuse slows down network training. This was addressed by increasing and decreasing the width and depth, respectively, in improved WNRs [182].

Furthermore, another efficient way to alleviate network performance is by reusing features. DenseNet [70] introduces connections between all successive layers in a feed-forward manner where features from each preceding layer are used as inputs to every other layer. This means that each layer is receiving cumulative knowledge from all prior layers, i.e., it reuses features. A variety of compelling benefits are obtained with the introduction of direct connections between layers. First, it allows more depth of the network while simultaneously alleviating the vanishing and exploding gradient problems. Second, the use of features from all layers leads to improvements in the performance. Finally, it efficiently utilizes parameters. This allows for less propensity to overfitting and leads to a reduction of computational costs. CondenseNet [69] combines dense connectivity with a group convolution to further facilitate feature reuse through the network. Here, the group convolutions aim at removing direct connections between layers allowing distinctly smooth feature reuse.

3.1.5 Autoencoders

Autoencoders (AEs) are a group of unsupervised learning algorithms that produce a reconstruction of their input. Autoencoders are made

up of two symmetric neural networks: an encoder and a decoder.

Let us assume that encoder is function f which maps input data $x \in \mathbb{R}^{d_x}$ to a latent representation $z \in \mathbb{R}^{d_z}$. It then can be mathematically expressed as:

$$z = f(x) = s_f(W \cdot x + b_z) \tag{3.14}$$

where s_f denotes an activation function, $W \in \mathbb{R}^{d_z \times d_x}$ is a weight matrix and $b_z \in \mathbb{R}^{d_z}$ is a bias vector.

The decoder is function g that reconstructs latent representations back to input space. It can be mathematically expressed as follows:

$$\hat{x} = g(z) = s_g(W'z + b_x) \tag{3.15}$$

where s_g is an activation function of decoder, $W' \in \mathbb{R}^{d_z \times d_x}$ is a weight matrix and $b_x \in \mathbb{R}^{d_z}$ is a bias vector.

The AE training procedure consist of finding the set of parameters $\Theta = (W, b_z, b_x)$ which are minimizing a loss function given with:

$$L(x, g(f(x))) \tag{3.16}$$

The goal of the reconstruction step is to minimize the difference between x and $\hat{x}$. This is also known as reconstruction loss. In order to obtain this, the latent space must learn the most important feature variations of the original data so that the reconstruction is sufficiently similar to the original data. This is achieved by simultaneously training both; the encoder and the decoder. An example of an AE network is illustrated in Figure 3.14.

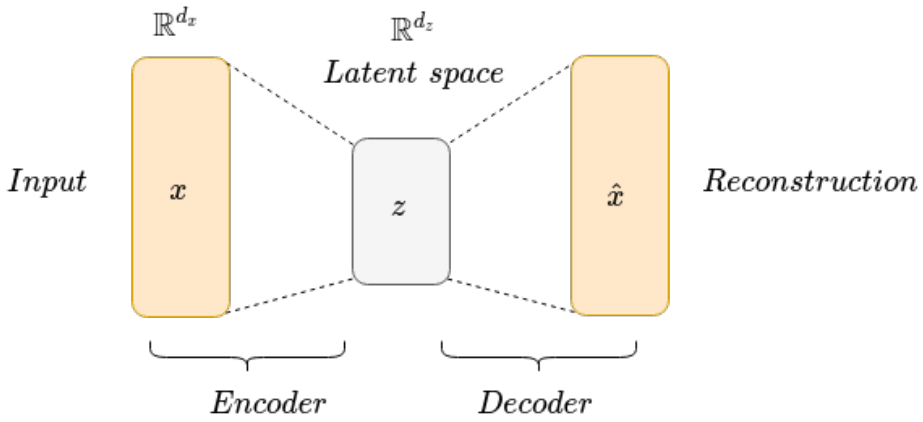


Figure 3.14: AE network illustration.

There are four types of autoencoders. Denoising autoencoders [168] recover a clean image from a partially corrupted input. Here, besides the identity function, the hidden layer learns robust features to ensure the result is not another corrupted image. In sparse autoencoders [111] hidden layers have greater dimensions than the input. They overcome

the challenge of avoiding the network for learning the identity function by permitting only a small number of hidden neurons to be active concurrently. In contractive autoencoders (CAEs) [143], there is an addition of a new term of their loss function. This addition ensures the robustness of a model to slight variations of the input values. Finally, variational autoencoders (VAEs) [89] are distinctive by the use of a variational approach for latent space learning while having the same architecture as autoencoders.

3.1.6 Variational Autoencoders

VAEs are a specific type of autoencoders that have probabilistic and generative nature. This means that the input and the latent space are supposed to be random variables with probability distributions. The problem formulation can be seen from a graphical model perspective, using graph theory to show the dependency between random variables.

Lets assume there is a dataset $X = x_i^N$ with a random variable x and N samples. The relationship of random variable z and dataset X can be mathematically expressed using the probabilistic graph model:

$$p_\theta(x, z) = p_\theta(x|z)p(z) \tag{3.17}$$

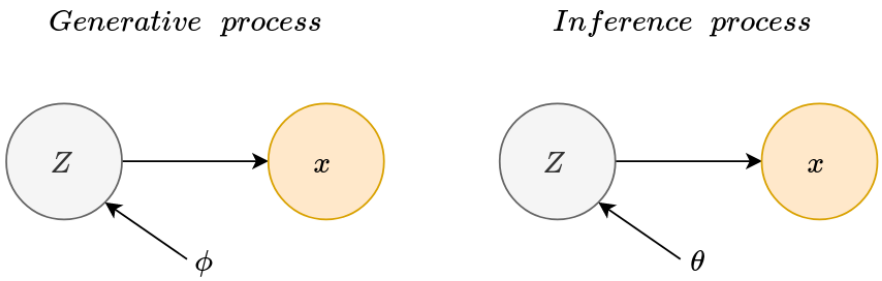


Figure 3.15: Generative and inference process of VAE expressed through a graphical model.

Using a generative probabilistic model, sampling a random variable z_i from a prior distribution $p(z)$ results in generation of the latent variables. Here, the data points x_i are obtained from conditional distribution over z, $p(x|z$. The prior and the likelihood are Gaussian distributions, that can be written as:

$$p(z) = N(z|0, I) \tag{3.18}$$

and

$$p_\theta(z|x) = N(x|f(z, \theta)), \sigma^2 I) \tag{3.19}$$

where $f(x, \theta)$ is a neural network and θ are the generative model parameters of the network.

Since the objective is to achieve a correct latent space z in regard to the observed data, we need to calculate the posterior probability $p(z|x)$ which according to Bayes can be expressed as:

$$p(z|x) = \frac{p(x,z)p(z)}{p(x)} \tag{3.20}$$

where marginal likelihood of $p_\theta(x)$ is expressed with:

$$p(x) = \int p_\theta(x,z) = \int p_\theta(x|z)p(z)dz \tag{3.21}$$

The Equation 3.21 requires all possible values of z, consequently requiring exponential time to compute. This is solved by approximating it with simpler distributions such as a Gaussian distribution, expressed with:

$$q_\theta(z|x) = N(z|\mu(x,\phi), \sigma^2(x,\phi)I) \tag{3.22}$$

Nevertheless, the total loss increases in this case and there is an addition of latent loss into the reconstruction loss term. That latent loss is calculated using the Kullback-Leibler divergence, which is further explained.

Kullback-Leibler Divergence

Kullback-Leibler (KL) divergence is a measure of difference or similarity between two probability functions. KL divergence can be mathematically described with:

$$KL(p(x)||q(x)) = -\sum q(x) log \frac{q(x)}{p(x)} \tag{3.23}$$

There are two main properties of KL divergence. The first property postulates when $p = q$ then

$$KL(p||q) \geqq 0 \tag{3.24}$$

The second property is that KL divergence is asymmetric, therefore:

$$KL(p||q) \neq KL(q||p) \tag{3.25}$$

In the case of second property, i.e., when KL divergence is asymmetric it is calculated using the following expression:

$$KL(q_\theta(z|x)||p(q_\theta(z|x))) = E_{q\phi(z|x)}[logq_\phi(z|x)] - E_{\phi(z|x)}[logp(z|x)] + logp(x) \tag{3.26}$$

The goal then is to minimize the KL divergence by finding the optimal variational parameters. However, it can be noted that the unknown $p(x)$ appears in that divergence. In order to solve this

problem, the posterior inference can be approximated by combining this KL divergence with the Evidence Lower Bound (ELBO).

Evidence Lower Bound

If we factorize the marginal likelihood as in:

$$logp(X) = log \prod_{i=1}^{N} p(x_i) = \sum_{i=1}^{N} log\, p(x_i) \tag{3.27}$$

then for each of the data points this likelihood can be mathematically written as:

$$\begin{aligned} log\, p_\theta(x_i) &= log \int p_\theta(x_i, z)dz \\ &= log \int \frac{q_\theta(z|x_i)p_\theta(x_i, z)}{q_\theta(z|x_i)}dz \\ &= log\, E_{q\psi(z|x_i)}\left[\frac{p_\theta(x_i), z}{q_\phi(z|x_i)}\right] \end{aligned} \tag{3.28}$$

Further, if we apply Jensen's inequity, defined with:

$$\psi(E[x]) \geqq E[\psi(x)] \tag{3.29}$$

then lower bound can be written as:

$$logE_{q\psi(z|x_i)}\left[\frac{p_\theta(x_i), z}{q_\phi(z|x_i)}\right] \geqq E_{q\psi(z|x_i)}\left[log\frac{p_\theta(x_i), z}{q_\phi(z|x_i)}\right] \tag{3.30}$$

Given that, ELBO can be written as:

$$ELBO(x_i) = E_{q\psi(z|x_i)}[log\, p_\theta(x_i|z) + log\, p(z) - log\, q_\theta(z|x_i)] \tag{3.31}$$

If we assimilate terms from previous Equation 4.7 with Equation 3.23, then for each data point ELBO can be written as:

$$ELBO(x_i) = E_{q\psi(z|x_i)}[log\, p_\theta(x_i|z) - KL(q_\theta(z|x_i)||p(z))] \tag{3.32}$$

Finally, the ELBO for the whole dataset can be written as:

$$ELBO(X) = \sum_{i=1}^{N} E_{q\psi(z|x_i)}[log\, p_\theta(x_i|z) - KL(q_\theta(z|x_i)||p(z))] \tag{3.33}$$

Therefore, the model's objective is to maximize the objective function through stochastic gradient descent optimization. However, derivatives of a distribution with respect to its parameters are not possible. To

tackle with this problem, a reparametrization trick is introduced as described in the following subsection.

Reparametrization Trick

Until now, we have achieved the samples z from distribution $q_\theta(z, x)$. To extract the derivates of a function of z with respect to ϕ, reparametrization of z is needed. In this way, the stochasticity is independent of the parameters of the distribution. This is done by using an auxiliary noise variable $\epsilon - N(0, 1)$, and we can write:

$$z = \mu(x, \phi) + \sigma(x, \phi)\epsilon \tag{3.34}$$

By taking Monte Carlo estimates, the expectation would be:

$$E\hat{LBO} = \sum_{i=1}^{N} \left[\frac{1}{L} \sum_{l=1}^{L} [log\, p_\theta(x_i|z_{i,l})] - KL(q_\phi(z|x_i)||p(z)) \right] \tag{3.35}$$

where $z_{i,l} = \mu(x_i, \phi) + \sigma(x_i, \phi)\epsilon_{i,l}$

As the two distributions in the KL-divergence term are Gaussian distributions, it can be calculated as:

$$KL(q_\phi(z, x_i)||p(z)) = -\frac{1}{2} \sum_{k=1}^{K} (1 + log\, (\sigma_k^2(x_i, \phi) - \mu_k^2(x_i, \phi) - \sigma_k^2(x_i, \phi)) \tag{3.36}$$

The Gaussian likelihood reconstruction term is:

$$log\, p_\theta(x_i|z_{i,l}) = -\frac{1}{2\sigma^2}(x_i - f(z_{i,l}, \theta))^2 + const \tag{3.37}$$

Finally, the estimation of the ELBO from a random data batch of size B would be:

$$ELBO(X) = E\hat{LBO}(X^B) = \frac{N}{B} \sum_{i=1}^{B} E\hat{LBO}(x_i) \tag{3.38}$$

3.2 Deep Learning for Medical Image Segmentation

The field of medical image processing is characterized by the use of mathematical algorithms that process and analyze multidimensional (2D, 3D, 4D) images to achieve recognition, segmentation, extraction, 3D reconstruction, or 3D visualization of various human organs [45, 12]. As the first step of medical image analysis, semantic segmentation plays a key role by extracting regions of interest in medical images and providing more intuitive medical information than raw images.

Semantic segmentation divides the image into several regions based on the similarity or differences between regions [104]. Generally, it can be described by a set theory model in the following manner. Given a medical image I and a set of similarity constraints $C_i(i = 1, 2, ...)$, the segmentation of I is to obtain a division of it, namely:

$$\bigcup_{x=1}^{N} R_x = I, R_x \bigcap R_y = \varnothing, \forall x \neq y, x, y \in [1, N] \tag{3.39}$$

where R_x satisfies both sets of all pixels in communication similarity constraint $C_i(i = 1, 2, ...)$, i.e., the image areas. The same is true for $Ry.x, y$ which are used to distinguish the different regions, N is a positive integer not less than 2, indicating the number of regions after division.

Deep learning algorithms based on convolutional neural networks (CNN) have shown excellent feature extraction capabilities when performing image segmentation operations. They have been extensively used in medical image segmentation [25]. The process of medical image segmentation using CNNs can be divided into the following stages:

- Obtain medical imaging dataset, generally including training set, validation set and testing set. The dataset is often divided into three parts. Among them, the training set is used to train the network model, the validation set is used to adjust the model's hyperparameters and the testing set is used to verify the final efficiency of the model.

- Preprocess and expand the image, generally including standardization of input image, perform appropriate data augmentation techniques on the input image to increase the size of the data set.

- Use the appropriate medical image segmentation method to segment the medical image and output the segmented images.

- Performance evaluation. In order to verify the effectiveness of medical image segmentation, effective performance indicators need to be set to be verified.

Before the advent of deep learning, traditional image processing techniques such as model-based methods (e.g., active shape and appearance models) or atlas-based methods had performed well in segmenting cardiovascular and medical images [82]. However, they frequently require significant feature engineering to achieve acceptable accuracy. In contrast, algorithms based on deep learning are able to automatically discover complex data features for object detection and segmentation. These features are extracted directly from the data using a general learning procedure in an end-to-end manner. This makes it easy to apply deep learning-based algorithms to other image analysis applications. Benefiting from advances in computer hardware, GPUs and tensor processing units (TPUs) and the increasing availability of training data, deep learning-based segmentation algorithms have gradually surpassed the state-of-the-art of conventional methods and are gaining popularity in research. Comprehensive review papers of cardiac segmentation methods can be found in [25, 54]. However, since the methods in this Thesis focus on deep learning, in the following subsections we provide an overview of the main deep learning algorithms used for segmentation of whole heart, left and right ventricles and abdominal aortic aneurysms.

3.2.1 Whole Heart Segmentation Methods

Deep learning-based whole heart segmentation methods can be divided into three groups: (1) two-stage segmentation methods consisting of localization and segmentation networks, (2) FCNs with deep supervision, (3) multi-view CNNs and (4) residual network variants.

Two-stage Segmentation

Two-stage segmentation methods consist of localization and segmentation networks. These methods characterize the extraction of a region of interest (ROI), which is then fed into a CNN for subsequent classification. In this two-step procedure, two CNNs are used. The first, localization CNN, approximates the center of all cardiac structures' bounding boxes. The predicted center is then cropped to create a fixed-size ROI that includes all relevant cardiac substructures. The second CNN segmentation algorithm predicts the label for each pixel. The may advantage of such approaches is enabling the segmentation network to concentrate only on anatomically significant regions and has been demonstrated to be successful for whole heart segmentation.

The work of Payer et al. [18] introduces a framework consisting of two separate CNN networks as shown in Figure 3.21. The first CNN network localizes the approximate center of the heart using landmark localization [19, 122]. This network is trained to regress the bounding box center around all heart structures using a U-Net-based network with heatmap regression. Input images are downsampled to a lower resolution due to memory restrictions and after cropping a

fixed predicted bounding box, voxels are resampled to a higher resolution. After that, three-stage segmentation CNN network inspired by SpatialConfiguration-Net is employed [19]. Here, the intermediate label predictions are generated in the first stage using a U-Net-like architecture. A sigmoid activation function is utilized to constrain the values between 0 and 1 for each output voxel, resulting in a voxel-wise probability prediction of all labels. The network then converts these probabilities to the placements of additional labels in the second stage, allowing the network to learn possible anatomical label configurations by suppressing infeasible intermediate predictions. In the third stage, the combined label predictions are obtained by multiplying the intermediate predictions from the U-Net with the modified predictions. The combination of localization and segmentation layers in the CNNs minimizes memory and computation needs.

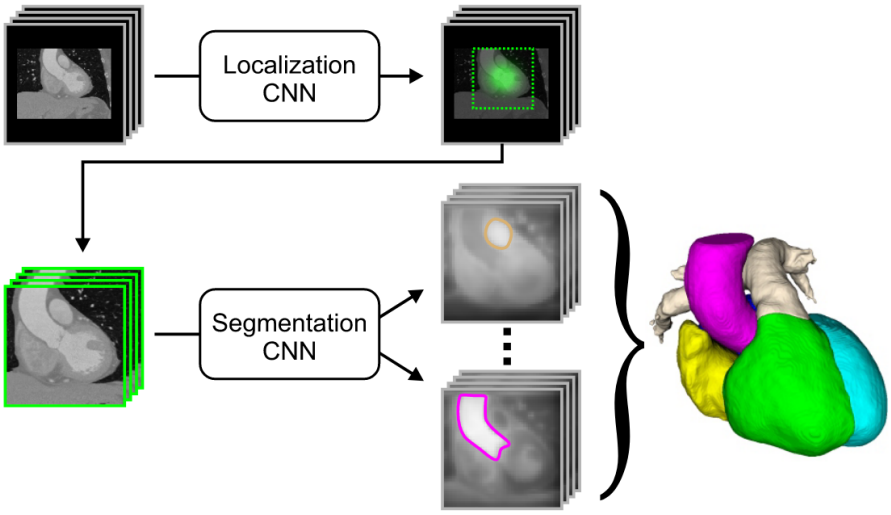


Figure 3.16: An illustration of an automatic multi-label segmentation framework composed of two CNN networks. The first CNN finds the center of the bounding box around all heart substructures. The second CNN crops the area surrounding this center and performs multi-label segmentation. Image source: Payer et al. [18]

In another two-stage framework, named CFUN, Xu et al. [179] utilized a modified 3D Faster R-CNN [142] network for localization named Region Proposal Network (RPN) and modified 3D U-Net for segmentation. First, modified 3D Faster R-CNN allows detection of one bounding box around heart structures. Originally proposed ResNet structure in Faster R-CNN is replaced with P3D ResNet structure [136] to better handle different information of images. After P3D ResNet, feature pyramid network (FPN) [99] combines feature maps in different resolutions. Second, 3D U-Net architecture with deep supervision in the decoder path generated final segmentation predictions. Further, the CFUN framework adopts a new loss function based on edge information named 3D edge-loss that significantly accelerates the convergence of

training and improves the segmentation results, which can be considered the biggest strength of this method.

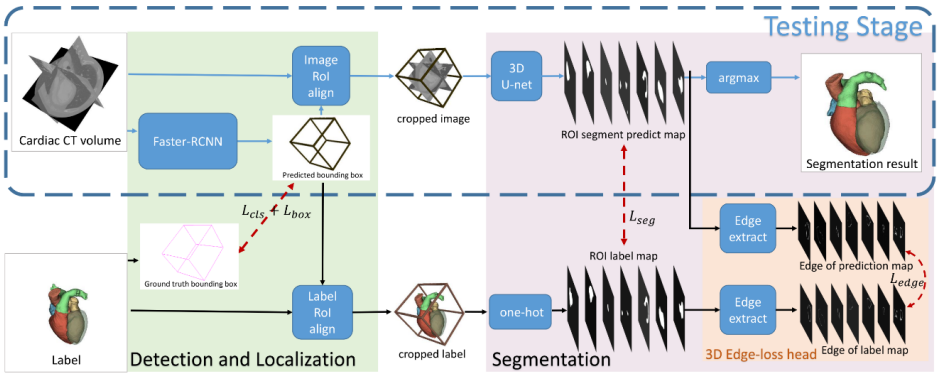


Figure 3.17: An illustration of CFUN framework. The localization 3D Faster R-CNN network outputs ROI containing the whole heart and the following modified 3D U-Net architecture provides fine segments of all heart structures. Image source: Xu et al. [179]

Similarly, the framework proposed by Tong et al. [135] consists of localization and segmentation networks. In the localization stage, they use U-Net to detect whole heart ROI coarsely. Following that, they augment the training set by extracting additional regions of interest and fusing CT and MRI images to use full multimodality information. In the segmentation stage, they employ 3D U-Net with deep supervision for obtaining final heart segmentation predictions. This method demonstrates how integrating a 3D U-Net with ROI detection mitigates the effect of neighbouring tissues and simplifies the computational process. Nevertheless, this method yields poor segmentation performance, especially for MRI images. Liu et al. [101] propose the two-stage U-Net framework that consists of an ROI detection and fine whole heart segmentation. The adaptive threshold window method is utilized to minimize noise and the weight map is enhanced to compel the network to learn the heart structure's boundary sections. The second network is fed with the equivalent ROI from the original data in the first stage.

FCN with Deep Supervision

Deep supervision [84] is the design where multiple segmentation maps are generated at different resolutions levels. The feature maps from each network level are transposed by $1 \times 1 \times 1$ convolutions to create secondary segmentation maps. It has been established that a deep supervision mechanism can effectively increase the convergence speed of training by driving the early hidden layers to prioritize discriminative characteristics for explicit predictions aggressively. Simultaneously, it has been proved that a deep supervision mechanism substantially

favors discriminative characteristics for explicit predictions in the early hidden layers.

One example of using deep supervision is in the work of Yang et al. [175]. They construct a framework based on a 3D FCN. The framework is strengthened in the following aspects. First, the network is initialized by inheriting the knowledge from 3D CNNs trained on the large-scale Sports-1M video dataset. The gradient flow is then applied by condensing the back-propagation and utilizing numerous auxiliary loss functions on the network's shallow levels. Direct training of the deep 3D FCN overcomes the issue of over-fitting and low efficiency in this manner. Taking into account the clear volume imbalance between different classes, they employ a multi-class form of the dice similarity coefficient-based loss function (mDSC) to balance the training for all classes. The method's primary strength is its use of a pre-trained network, which ensures proper initialization and minimizes overfitting. Utilizing auxiliary loss functions facilitates gradient flow and simplifies the training procedure. The primary disadvantage of this method is that hyperparameters are determined empirically and segmentation performance on MRI pictures is low.

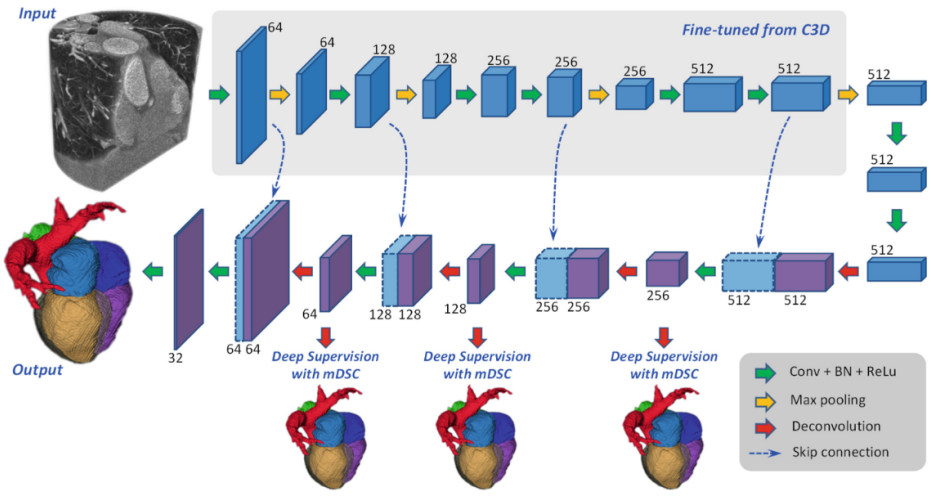


Figure 3.18: An illustration of segmentation framework with deep supervision mechanism. Image source: Yang et al. [179]

Ye et al. [21] base their work on 3D deeply-supervised U-Net. They extend the multi-branch residual network with muti-depth fusion. Multi-depth fusion allows better feature aggregation and extraction of the context information. After that, they apply focal loss that helps capture more advanced features to provide more precise boundary identification. In this way, the network application is extended for multi-category segmentation. Similarly, Yu et al. [91] introduce DenseVoxNet, a novel densely connected volumetric CNN that utilizes 3D FCN for accurate volume-to-volume prediction. The densely-connected

mechanism maintains the maximum amount of information flow between layers, which simplifies the training process. Additionally, it obviates the need to learn redundant feature maps that boost performance. The superiority of their method is presented in using focal loss, which successfully addresses the class imbalance issue.

Dou et al. [134] present another 3D FCN approach based on a 3D deep supervision mechanism (3D DSN). This network provides volume-to-volume learning, which eliminates redundant computations and mitigates the effects of overfitting on sparse training data. Additionally, it turns out that the 3D deep supervision mechanism efficiently alleviates the usual optimization problem of vanishing gradients that occurs during the training of 3D deep models. This enhances discrimination capabilities and speeds up convergence. Additionally, the fully connected conditional random field model is used to refine the segmentation findings as a post-processing step. Their strategy considerably enhances discrimination capabilities while also speeding up convergence.

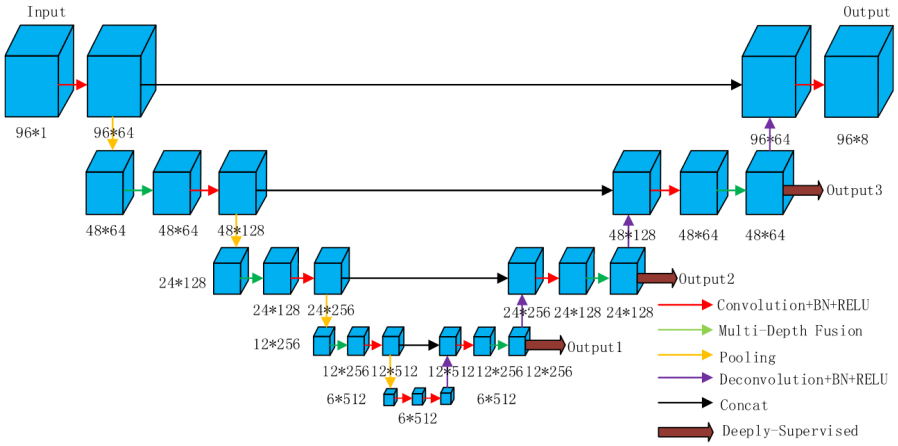


Figure 3.19: An illustration of segmentation framework with deep supervision mechanism. Image source: Ye et al. [179]

Similarly, Li et al. [76] propose training with 3D FCN and the addition of dilated convolution layers (3D- HOL layers) to increase the receptive field and make better use of spatial information. The introduction of deeply-supervised paths allows the use of multi-scale information at multiple levels, which accelerates the training process. The importance of the method stems from its capacity to finely segment healthy cardiac structures and severely defective heart structures such as those found in CHD. This is one of the few techniques that can successfully segment diseased hearts.

Multi-view CNNs

Generally, multi-view learning (MVL) aims to learn the common feature spaces or shared patterns by combining multiple distinct features

or data sources. Multi-view CNNs are frameworks that combine information from different views into fully connected layers to classify the voxel where different planes intersect.

Wang and Smedby [20] introduce a framework consisting of two concatenated U-Net networks. The framework includes three stages: scout segmentation with orthogonal 2D U-Nets, shape context estimation and final segmentation with U-Net and shape context. In the first stage, they adapt three independently trained U-Net networks for segmenting heart structures in different orthogonal projections. The final segmentation map is obtained by averaging the outputs of these three U-Nets. In the second stage, the statistical shape model [97] is created by taking the mean of the signed distance functions of each segmented region and extracting prominent variations using Principal Component Analysis (PCA). After estimating the shape models that fit the probability map of the scout segmentation, in the third stage, the distance maps of the heart structures are fed into another three U-Net networks, similar to those used in the first stage. The final segmentation map is obtained by averaging the outputs of these three U-Nets. Combining shape context information with orthogonal U-Nets obtains more consistent segmentation, which yields higher segmentation accuracy. Nevertheless, the biggest downside of this method is the use of weighting factors of the shape context generation, which are determined empirically.

Mortazi et al. [3] present another solution based on an encoder-decoder CNN architecture: a multi-object multi-planar CNN (MO-MP-CNN) method. First, multiple 2D CNNs are trained from three different views, i.e., axial, sagittal and coronal views. Second, an adaptive fusion method is employed to refine the delineation by combining different results. Finally, connected component analysis (CCA) is utilized to determine which regions are reliable and which are not. The distinctions between these zones are utilized to determine the segmentation process's reliability (a higher difference corresponds to a more reliable segmentation). The primary advantage of this approach is that it requires less processing power, as numerous 2-D CNNs require less memory than a 3-D CNN. However, the softmax function in the network's last layer may result in data loss due to class normalization.

Residual Networks Variants

Shi et al. [178] proposed a probabilistic deep voxelwise dilated residual network named Bayesian VoxDRN that can predict voxelwise class labels with a measure of model uncertainty. By utilizing the dropout process, the model is able to learn weight distributions with a higher degree of data explanation. This considerably reduces the likelihood of over-fitting. Another enhancement is optimizing a binary segmentation using an iterative switching training technique. This method is significant because it enables the optimization of binary segmentation

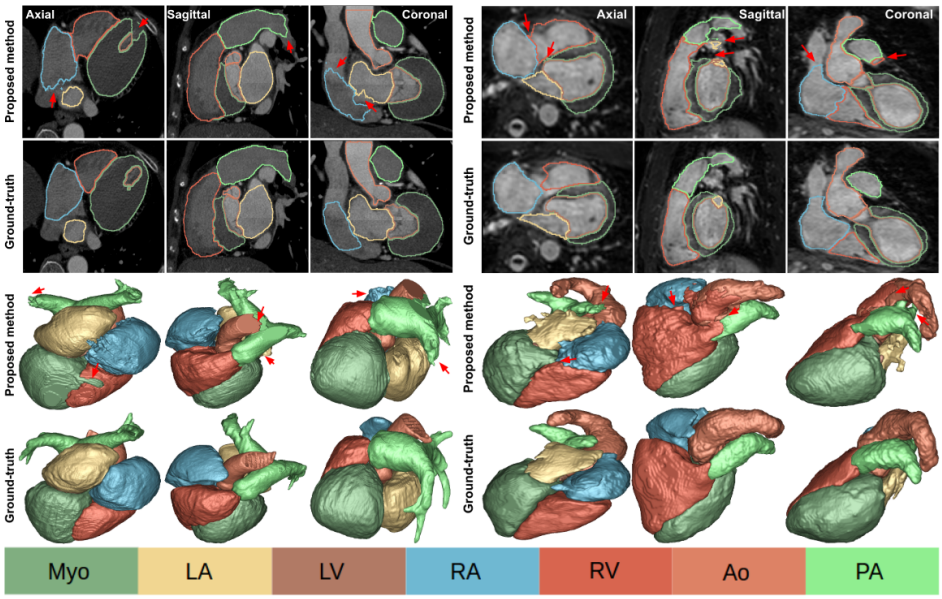


Figure 3.20: The first two rows show axial, sagittal and coronal planes of the CT (first three columns) and MR images (last three columns), annotated cardiac structures and their corresponding surface renditions (last two rows). Red arrows indicate unsuccessful segmentations. Image source: Mortazi et al. [3]

using an iterative switching training strategy. The clinical utility of the uncertainty measurements, on the other hand, is unknown.

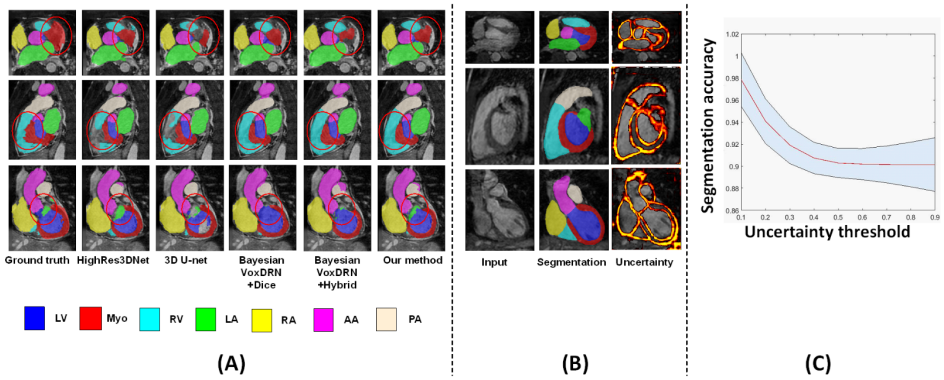


Figure 3.21: An example of obtained results. (a) Red circles highlight the major differences among various methods. (b) Visualization of uncertainty achieved with a dropout-based Monte Carlo sampling, the brighter the color, the higher the uncertainty. (c) The relationship between the segmentation accuracy and the uncertainty threshold where the shaded area shows standard errorsc. Image source: Shi et al. [178]

3.2.2 Bi-ventricles and myocardium segmentation methods

Deep learning-based methods for the bi-ventricles and myocardium segmentation task can be divided into four groups: (1) U-Net and its variants, (2) U-Net with deep supervision, (3) U-Net with residual connections and (4) U-Net with transformers.

U-Net Architecture

As discussed in subsection 3.1.3, U-Net and 3D U-Net have symmetrical architecture, with an encoder extracting spatial information from the image and a decoder creating the segmentation map using the encoded features.

Few works use U-Net architecture to provide experimental analysis observing the influence of different parameters for final segmentation results. For example, Baumgartner et al. [12] investigate 2D U-Net and 3D U-Net with various hyperparameters. They compare the performance of 2D and 3D convolutional layers and training with Dice loss versus training with cross-entropy loss. Their optimal architecture was determined to be a U-Net with two-dimensional convolutional layers trained with cross-entropy loss. Patravali et al. [127] evaluated a 2D and 3D U-Net trained with varying Dice and cross-entropy losses. According to their experiments, the optimal architecture was a two-dimensional U-Net with a Dice loss. Jang et al. [78] implemented an M-Net architecture [115] in which the decoding layers' feature maps are concatenated with those of the previous layer. A weighted cross-entropy loss was used to train the matching network.

Luo et al. [108] propose a method based on U-Net and combined with image sequence information. They introduce two modules: the contextual extraction module and the segmentation module. The context extraction module can fully extract the context features of the image to be segmented and effectively combines the sequence features. The segmentation module is an encoder-decoder module and input image can directly predict a segmented image. The module effectively learns the characteristics of the original image and avoids feature loss and gradient dispersion by the design of the skip connection.

The work of Galea et al. [47] introduces a method that consists of three parts: data preprocessing, ROI localization and segmentation using U-Net and DeepLab architectures [27]. In the preprocessing stage, the data is normalized using each particular slice's mean value and standard deviation. After that, preprocessed input images are fed to the networks. Further, they explore if models would benefit by focusing only on ROIs. They compare segmentation with and without using ROI and as expected, they obtain higher accuracies when using ROI. An illustration of their method is shown in Figure 3.22.

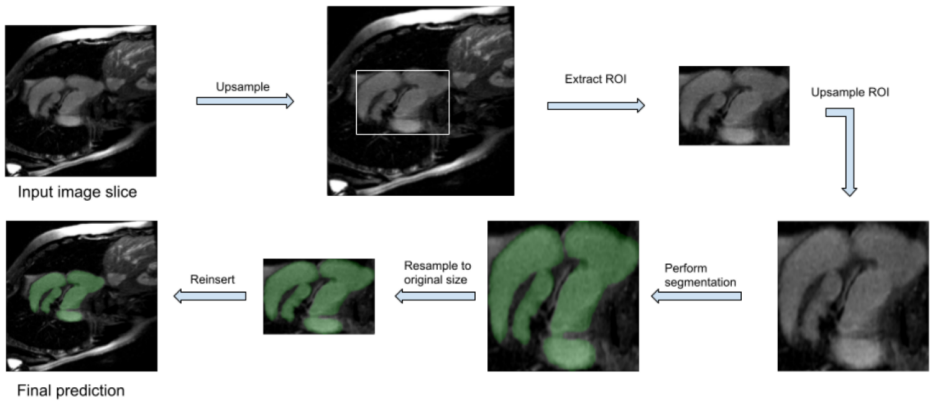


Figure 3.22: An illustration of three-step method proposed by Galea et al. Image source: Galea et al. [47]

U-Net with Deep Supervision

Khened et al. [86] implemented a dense U-Net. Their method starts by finding the region of interest with a Fourier transform followed by a Canny edge detector on the first harmonic image and compute an approximate radius and center of the LV with a circular Hough transform on the edge map previously generated. After that, they use a U-Net with dense blocks instead of a basic convolution block to make the system lighter. The first layer of this network also corresponds to an inception layer.

Snauw et al. [158] combine DenseNet and U-Net in order to solve the segmentation and classification tasks. They include three branches into the network: main branch, segmentation branch and diagnosis branch. The composite function for every operation in the model consists of BN-ReLU operation. The network is trained using a loss function that is a convex combination of segmentation and classification losses.

U-Net with Residual Connections

Isensee et al. [74] implemented an ensemble of 2D and 3D U-Net architectures (with residual connections along with the upsampling layers). Concerning the 3D network, due to the large interslice gap on the input images, pooling and upscaling operations are carried out only in the short-axis plane. Moreover, due to memory requirements, the 3D network involves a smaller number of feature maps. Both networks were trained with a Dice loss. Yang et al. [176] implemented a 3D U-Net with residual connections instead of a commonly used concatenation operator. They also used pre-trained weights for the downsampling path using the C3D network known to work well on video classification tasks [165]. Their network was trained with a multi-class Dice loss.

The work of Sander et al. [147] combines automatic segmentation and assessment of segmentation uncertainty. They train three networks:

dilated CNN, dilated residual network (DRN) and U-Net for automatic segmentation of LV, RV and Myo. Spatial uncertainty maps of the obtained segmentations are generated to detect failures in segmentation masks (to investigate uncertainty). They use two measures of predictive uncertainty: entropy and a measure derived by Monte Carlo dropout (MC-dropout)[46]. With simulated and manual correction of detected segmentation failures, this combined approach increases performance compared to an approach with only a segmentation. An illustration of their network architecture is shown in Figure 3.23.

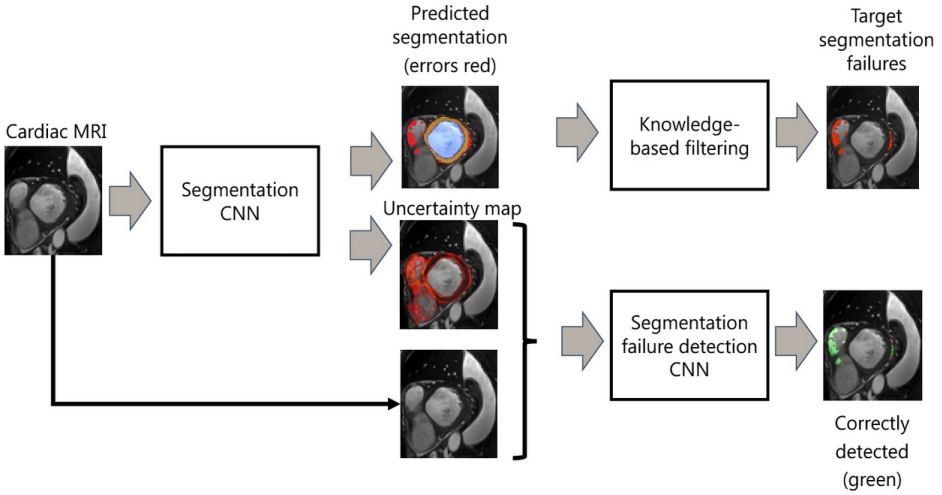


Figure 3.23: An illustration of the proposed two-step method. In the first step, MR images are automatically segmented, while the second step distinguishes acceptable mistakes from segmentation failures using distance transform maps. Image source: Sander et al. [147]

U-Net with Transformers

Chen et al. [26] proposed architecture, TransUNet, that incorporates Transformers [103], with inherent global self-attention mechanisms into U-Net. As the input sequence for extracting global contexts, the Transformer encodes tokenized image patches from a CNN feature map. On the other hand, the decoder upsamples the encoded features before combining them with the high-resolution CNN feature maps to enable exact localization. Transformers overcome U-Net limited localization ability due to insufficient low-level details. An illustration of TransUnet is shown in Figure 3.24.

Cao et al. [22] propose U-Net based architecture named Swin-Unet. They use hierarchical Swin Transformer [103] with shifted windows as the encoder to extract context features and a symmetric Swin Transformer-based decoder with patch expanding layer designed to perform the up-sampling operation to re-store the spatial resolution of the feature maps. Their architecture outperforms those methods with full-convolution or the combination of transformer and convolution.

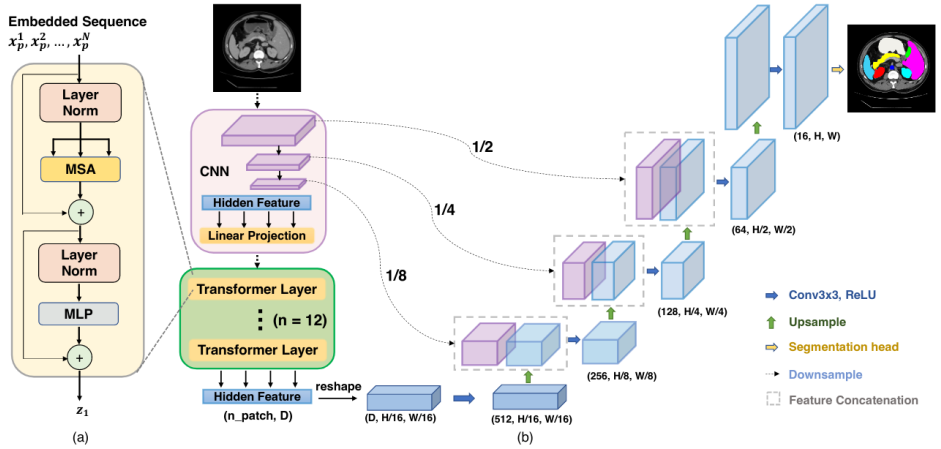


Figure 3.24: An illustration of the framework. (a) Transformer layer, (b)architecture of the proposed TransUNet. Image source: Chen et al. [26]

Moreover, this is the first U-Net based architecture that leverage the power of pure Transformer for medical image segmentation. An illustration of Swin-Unet is shown in Figure 3.25.

3.2.3 Abdominal Aortic Aneurysm Segmentation Methods

Traditionally, AAA segmentation has been addressed with intensity-based semi-automatic methods (level-sets, active shape models, graph cuts) combined with shape priors [44, 96, 5, 41, 156], deformable models [37, 123]. Even if some of these algorithms provide reasonably good results, they require the optimization of many parameters and are dataset-dependent, which reduces the robustness and the reproducibility required in a real clinical setting.

Given that the annotated datasets of AAA have only recently emerged, the development of specific methods based on deep learning is gaining momentum. Deep learning-based methods for the AAA segmentation task can be divided into two groups:(1) FCNs and (2) various CNN variants.

FCNs

The work of Lopez-Linares et al. [100] introduces a framework that includes adapted Detect Net for AAA region of interest (ROI) detection. They introduce a segmentation network based on FCN and holistically-nested edge detection for thrombus segmentation. They use contract enhancement, resizing and ROI extraction to alleviate image quality in a pre-processing stage. Such images are then fed as input to the network. In the post-processing stage, they binarize obtained predicted segmentation to remove an unnecessary small objects and

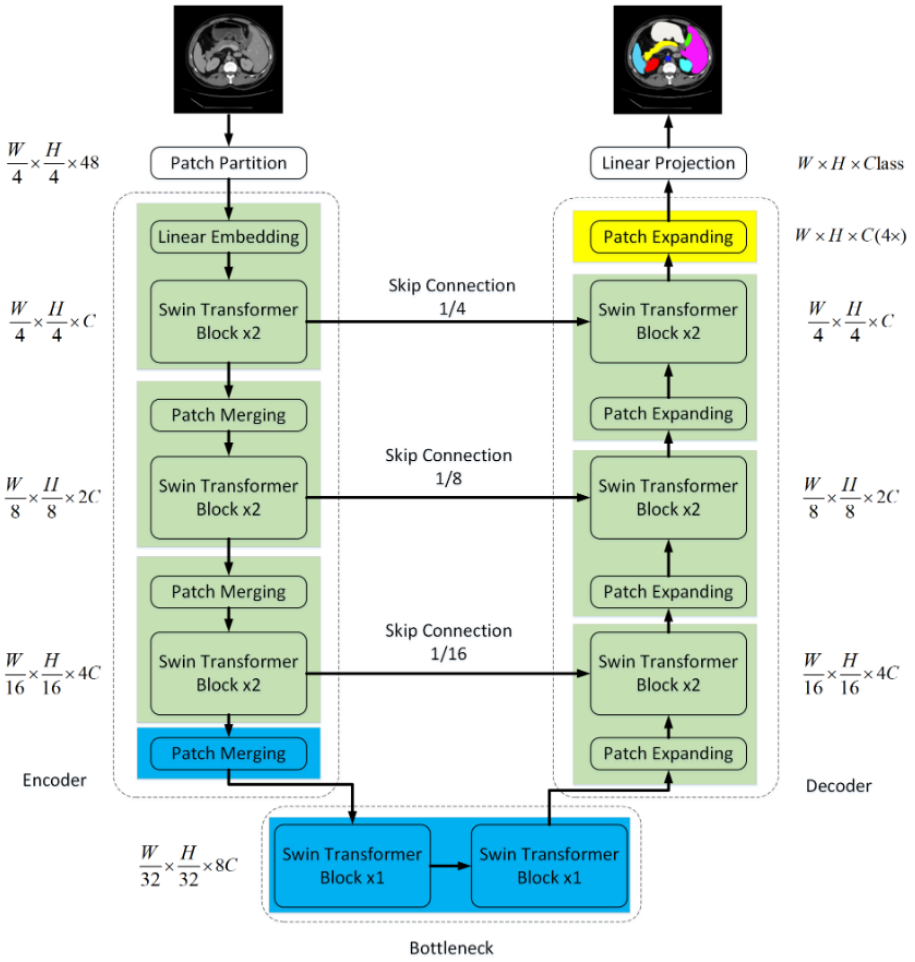


Figure 3.25: The architecture of Swin-Unet, which is composed of encoder, bottleneck, decoder and skip connections. Encoder, bottleneck and decoder are all constructed based on swin transformer block. Image source: Cao et al. [22]

obtain high segmentation accuracy. Zheng et al. [183] investigate the impact of using an extremely small number of datasets for AAA segmentation. They trained the U-Net network on just two CT datasets while maintaining high accuracy using strong data augmentation. They also use a linear mapping transformation to eliminate the inter-subject variation of image contrast. An illustration of the framework is shown in Figure 3.26.

Lu et al. [106] propose AAA segmentation with DeepAAA. Their framework consists of two steps: aorta segmentation and aorta contour fitting. In the first step, they develop a variant of a 3D U-Net that accepts data with varying numbers of images. The second step uses elliptical fitting to determine the greatest aortic diameter using segmented aortic outlines. Because they developed a general AAA detector that worked with both contrast and non-contrast CT scans, they trained the model on both types of CT images. Thus, they achieve

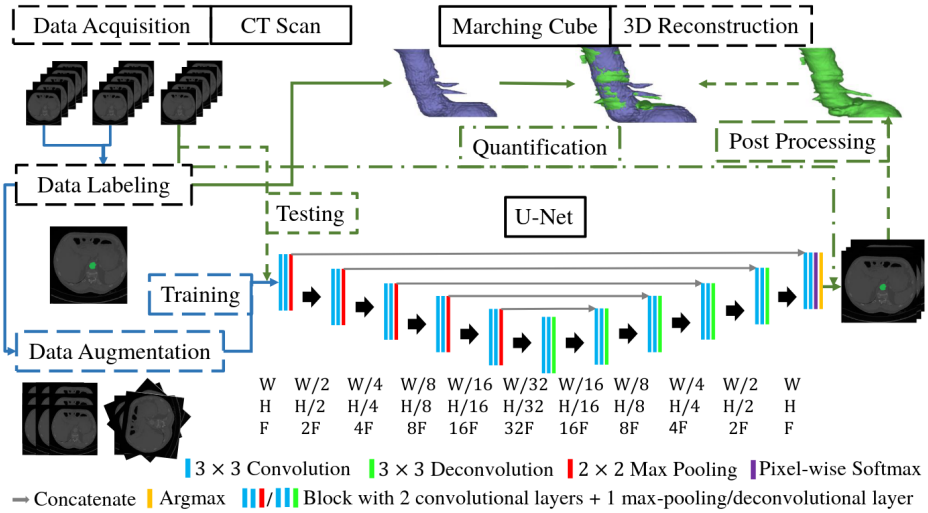


Figure 3.26: An illustration of a framework of 3D AAA reconstruction with modified U-Net and strong data augmentation. Image source: Zheng et al. [183]

a high detection rate for contrast and non-contrast CT scans on images with different resolutions and slice thicknesses. Their approach shows high generalization ability and performance highly similar to literature-reported values for radiologist sensitivity.

Wang et al. [20] propose a U-Net-like architecture that fuses the high-level part of the MR and CT images, allowing successful multi-modal segmentation. The main benefits of fusing higher-level feature layers are as follows. The validation accuracy of the fusion model increases faster than that of separate models during training, while maintaining the same number of model parameters. While shared layers can learn higher feature representations from both visual modalities, individual models can only learn from one. The fusion models allow a shared representation to be learned for all image modalities. That means that the representations of infused layers are similar to CT and MR images showing similar parts of the aorta. Such a network can be trained end-to-end with non-registered CT and MR images using a shorter training time. An illustration of framework is shown in Figure 3.27.

CNN variants

Hong et al. [65] use a deep belief network (DBN) for AAA ROI detection, segmentation, classification and measurement of AAA. ROI detection step uses two DBNs. The first DBN detects large aneurysm patches, while the second detects small ones, bones, organs and air. Another separate DBN is trained with patches containing an aneurysm for segmentation purposes. The use of DBN has shown efficient in solving this task as DBN allows a significant reduction in training complexity while leveraging high segmentation accuracy.

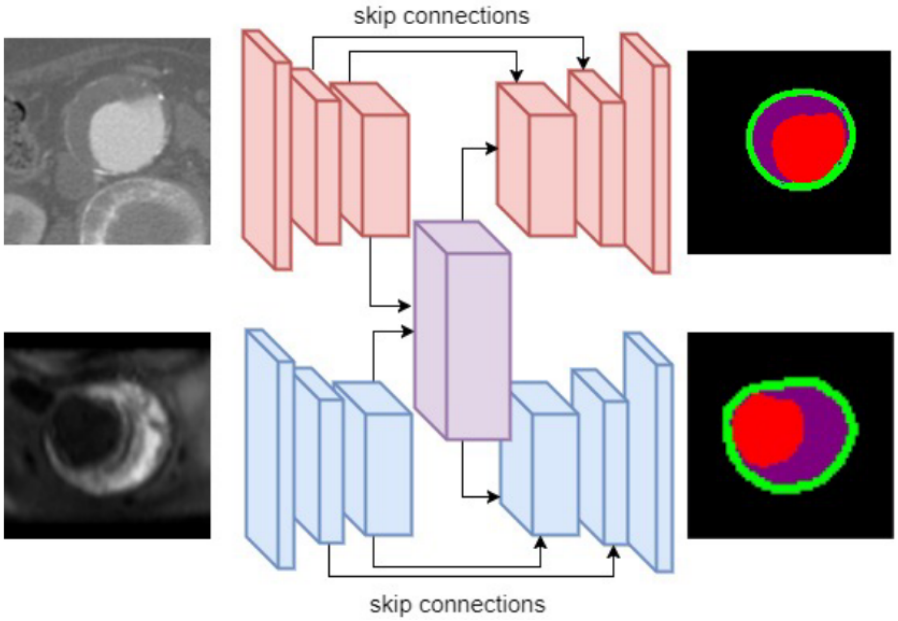


Figure 3.27: An illustration of fusion model for CT and MR image modalities. The top layers in an encoder and decoder are fused from two separate streams into one stream. Image source: Wang et al. [20]

Fantazzini et al. [42] propose framework that consist of three stages. First, a 2D U-Net is trained on down-sampled axial slices to localize and extract an initial aortic mask from CTA data. The extraction of an ROI serves to reduce the amount of required memory. After that, 2D U-Nets trained on different planes are used to process the identified ROI. These planes are acquired by extracting 2D slices along the axial, sagittal and coronal axes of the CTA scan at a higher resolution. The three-plane U-Net predictions are then concatenated to produce a spatially coherent final segmentation, overcoming the constraints of single-plane CNNs.

Jiang et al. [79] propose method to cope with the limited medical follow-up dataset of AAAs. Their method uses a vascular growth and remodeling computational model. It is able to capture the variations of actual patient's AAA geometries and is used to generate a limited in silico dataset. After that, the probabilistic collocation method reproduces a large in silico dataset by approximating simulation outputs. A DBN is then trained to provide fast predictions of patient-specific AAA expansion, using both in silico data and patients' follow-up data.

Caradu [23] presents open-sourced software PRAEVAorta with dedicated deep learning-based algorithms for fully automatic segmentation of AAAs. This software automatically detects AAAs morphology (including an inner lumen and the thrombus, which might easily be put into clinical practice. They facilitate image segmentation and analysis by enabling investigators to more quickly discover aneurysms, define their anatomic properties (including the presence of intraluminal thrombus) and calculate the aneurysm's diameters, lengths and

volumes automatically. It presents the advantage of reducing segmentation time and user interaction. Nevertheless, their analysis was limited to strict infra-renal AAAs since the precise characterization of patients with complex aneurysms, including the para-renal and visceral segments, still requires future validation.

3.2.4 Common Evaluation Metrics

Comparing images to assess the accuracy of segmentation is critical for evaluating progress in this field of research. The following four metrics are most commonly used to evaluate segmentation performance: the Dice similarity coefficient (DSC), Jaccard Index (JI), surface distance (SD) and Hausdorff distance (HD). DSC and JI measure the level of overlap between the ground truth and predicted segmentations, while SD and HD examine boundary distances. The DSC metric measures the degree of overlap between the ground truth and predicted segmentation. It is a commonly used metric for evaluating segmentation quality and can be written as:

$$DSC(G,P) = \frac{2|G \cap P|}{|G| + |P|} \tag{3.40}$$

where G is the ground truth and P is the predicted mask.

Similarly, the Jaccard Index (JI) emphasizes the size of the intersection divided by the size of the union of the sample sets. The mathematical representation of the JI can be written as:

$$JI(G,P) = \frac{|G \cap P|}{|G \cup P|} \tag{3.41}$$

where G is the ground truth and P is the predicted mask.

SD measures an average of the minimum voxel-wise distance between the ground truth and predicted object boundaries and can be written as:

$$SD(G,P) = \frac{1}{n_G + n_P} \left\{ \sum_{x_P \in P} \bar{d}(x_P, G) + \bar{d}(x_G, P) \right\} \tag{3.42}$$

where n_G and n_P denote the number of voxels on the object boundaries in the ground truth and predicted segmentations, respectively.

Furthermore, HD represents the maximum of the minimum voxel-wise distances between the ground truth and predicted object boundaries and can be written as:

$$HD(G,P) = \max_{g \in G} \left\{ \min_{p \in P} \left\{ \sqrt{g^2 - p^2} \right\} \right\} \tag{3.43}$$

where g is the ground truth and p is the predicted mask.

3.3 Challenges and Limitations

As reviewed in previous sections, various approaches and methods are introduced for the task of cardiovascular structures segmentation. Here we highlight common challenges and limitations of previously proposed methods.

While huge collections of general-purpose images are easily available and accessible to researchers, collecting and utilizing medical images is a considerable hurdle to developing new deep learning-based systems. Although every hospital has image archiving and transmission systems that regularly store millions of images, medical image databases are typically small and confidential for research purposes. There are two reasons why this massive amount of stored data cannot be directly used for medical image analysis:

- Ethical, privacy, security and legal issues. The transmission, storage and use of medical data are subject to specific regulations. Informed consent of the patient is usually required for the use of an image in a study, as well as data anonymization methods to protect the patient's privacy.

- Insufficient expert annotations for the images. Training an algorithm to segment medical images usually requires labelling each pixel in the image according to its class, i.e., object or background, which is often time consuming and subject to observer error. As a result, the number of publicly available annotated datasets is limited.

Efficient learning from limited annotated data is an important area of research. In developing deep learning-based segmentation approaches, the following strategies are usually used to increase the size of the dataset:

- Data augmentation is the technique of extending a dataset by generating new images, either through simple operations such as translations and rotations, or with advanced techniques such as principal component analysis [144], histogram matching, or elastic deformations [62, 118].

- Data augmentation using synthetic image creation methods to augment the database, for example, alleviate existing dataset using GANs [152, 154].

- To train the network with more data, many applications convert 3D medical volumes into stacks of independent 2D images instead of using the full-size images [34, 100]. However, the obvious drawback is that the anatomical context in orthogonal directions to the slice plane is completely ignored. To increase the amount of data, partial volumes or image patches are often extracted from the images.

3.4 Conclusion

Segmentation of medical images is a crucial step in extracting meaningful information from body structures. It allows complex analysis of segmented regions and extraction of quantitative information that could be useful in the development of computer-aided diagnosis systems. Deep learning techniques have increased the accuracy of complex segmentation tasks that could not be solved using conventional image processing algorithms. However, in order for deep learning systems to be used in clinical practice, several challenges inherent to the field of medical imaging must be overcomed. These mainly relate to the generalization of segmentation approaches, which require large amounts of annotated medical image data obtained with different protocols in different environments, covering most anatomical and pathological variations in patients. As a result, researchers use intelligent data augmentation techniques and specially designed loss functions to compensate for the lack of data and annotations. According to literature, the most commonly used supervised deep learning segmentation networks in medical imaging are variants of encoder-decoder architectures. Developing new building blocks to improve the efficiency and accuracy of these networks is a current research topic.

CHAPTER 4

Whole Heart and Heart Chambers Segmentation

This chapter presents a new, fully automatic approach for robust and accurate whole heart and heart chambers segmentation. We present a novel connectivity structure of residual unit, which we refer to as a feature merge residual unit (FM-Pre-ResNet). The proposed connectivity allows the creation of distinctly deep models without an increase in the number of parameters compared to the pre-activation residual units. Following that, we present a novel 3D encoder-decoder-based architecture that successfully integrates FM-Pre-ResNet units with VAEs. FM-Pre-ResNet units are used to learn a low-dimensional representation of the input during the encoding stage. The VAE reconstructs the input image from the low-dimensional latent space, ensuring that all model weights are strongly regularized while avoiding overfitting on the training data. Finally, during the decoding stage, the final segmentations are created. The proposed method is evaluated on the 40 test subjects of the MICCAI Multi-Modality Whole Heart Segmentation (MM-WHS) Challenge. Our method achieves an average DSC, JI, SD and HD for WHS of 90.39%,82.24%, 1.1093 and 15.3621 on CT images and 89.50%, 80.44%, 1.8599, 25.6558 on MRI images, respectively.

The outline of the chapter is structured as follows. Section 4.1 gives the main objectives of conducted research. Section 4.2 gives a theoretical background of used methods and describes our proposed method for whole heart and heart chambers segmentation. Section 4.3 describes the experimental setup, gives network training details and presents obtained results. Finally, discussion and concluding remarks are provided in Section 4.5.

4.1 Objectives

This research aims to develop an efficient method for fully automatic segmentation of heart structures in CT and MRI images. Deep neural network architectures provide more abstract learning, resulting in better performance and higher accuracy in cardiovascular segmentation tasks. For example, in 3D U-Net architecture features from contracting and expanding pathways are concatenated with skip connections to retrieve lost image information that occurs during the down-sampling process. Intuitively, this indicates that part of the information is lost during the encoding process and can not be recovered when decoding. Variational autoencoders enable regularization during the training to ensure that the latent space, i.e., encoded space, keeps the maximum of information when encoding, which results in the minimum reconstruction error during the decoding. Furthermore, since the number of features in the contracting pathway is significantly lower than the number in the expanding pathway, direct concatenation of these features may not produce the most optimal results. The increment in the number of layers provides larger parameter space enabling learning of more abstract features. Nevertheless, with the increasing depth, information about the gradient passes through many layers and it can vanish or accumulate large errors by the time it reaches the end of the network resulting in saturated accuracy that degrades rapidly. Since some features are best constructed in shallow networks and others require more depth, the introduction of skip connections allows residual learning and increases the network's capability, flexibility and performance. Nevertheless, even residual learning networks with extremely large depth are challenging to converge.

Therefore, the objectives of this research can be summarized as below:

1. To develop a novel residual unit connectivity structure (FM-Pre-ResNet) that enables the construction of a deeper models with a less or equal number of parameters to the original pre-activation residual unit.

2. To propose a novel 3D encoder-decoder based architecture that efficiently incorporates FM-Pre-ResNet units and is additionally guided with variational autoencoders (VAE) for the task of whole heart segmentation.

3. To compare the performance and result obtained from the proposed method with existing methods.

Hereby, we present a novel 3D encoder-decoder-based architecture with variational autoencoder regularization. Our intention is to achieve high optimization in training performance, efficiency and final segmentation result accuracy for the whole heart segmentation task.

4.2 Methodology

This section presents the proposed segmentation method for whole heart and heart chambers segmentation from CT and MRI images. We present a theoretical background of the proposed FM-PreResNet units and VAEs. We give an overall design of the proposed encoder-decoder-based architecture with VAE and introduce its main building blocks and purpose. We give dataset description, implementation details, present conducted experiments and obtained results. Finally, we provide some concluding remarks.

4.2.1 Feature Merge Residual Units

As discussed in Section 3.1.4, residual learning reformulates the layers as learning residual functions with reference to the layer inputs instead of learning unreferenced functions. Generally, each residual unit can be expressed with the following two expressions:

$$\boldsymbol{y_l} = H(\boldsymbol{x_l}) + F(\boldsymbol{x_l}, W_l), \tag{4.1}$$

and

$$\boldsymbol{x_{l+1}} = f(\boldsymbol{y_l}), \tag{4.2}$$

where F is residual function, $\boldsymbol{x_l}$ and $\boldsymbol{x_{l+1}}$ are vectors that denote the input and output of the $l-th$ residual unit in the network, while the output of the $l-th$ residual unit is denoted with vector $\boldsymbol{y_l}$. The parameters of the $l-th$ residual unit are denoted as W_l, while the function f refers to the rectified linear unit (ReLU).

Multiple stacked residual units form ResNets. The identity mapping, by which ResNets learn residual function F in regard to $H(\boldsymbol{x_l})$, can be written as:

$$H(\boldsymbol{x_l}) = \boldsymbol{x_l} \tag{4.3}$$

Therefore, the identity mapping of the original residual unit attaches an identity skip connection allowing information flow within a residual unit as shown in Figure 4.1a. As introduced in Pre-ResNets (Figure 4.1b), if $H(\boldsymbol{x_l})$ and f are both an identity mapping, the direct propagation of information through the entire network in forward and backward fashion can be written as:

$$\boldsymbol{x_{l+1}} = H(\boldsymbol{x_l} + F(\boldsymbol{x_l}, W_l)) \tag{4.4}$$

To obtain more features from an original image and to improve the networks' capacity, we try to merge the feature from a previous layer in the residual signal branch and add the residual blocks of earlier groups based on the original Pre-ResNet. As shown in (Figure 4.1b), the original Pre-ResNet consist of two parts: the identity mapping and the residual signal branch. We separately add a weight layer

(convolution) at the top and bottom of the original residual signal. The output of the top weight layer is concatenated with the output of the residual signal. In this way, it merges the feature from the previous layer into a subsequent one. After that, the concatenated result passes through another weight (convolution) layer to reduce the feature map dimension as illustrated in (Figure 4.1c). Therefore, following the concept described in Equation (4.4), we propose a new residual structure named feature merge residual unit that can be written as follows:

$$Z(\boldsymbol{x_l}) = F(z(\boldsymbol{x_l}), W_l) \circ z(\boldsymbol{x_l})) \tag{4.5}$$

and

$$\boldsymbol{x_{l+1}} = H(\boldsymbol{x_l} + g'(Z(\boldsymbol{x_l}), W_l') \tag{4.6}$$

where $\circ$ presents the concatenation operation, $Z(\boldsymbol{x_l})$ denotes the concatenated result, while the functions z and g' denote the convolution layers, added at the top and at the bottom of the residual unit, respectively. In this manner, the top convolution layers' output is concatenated with the residual signals' output, which allows the merge of features from preceding layers. After that, the concatenated result is passed through a bottom convolution layer to reduce channel dimension, as shown in Figure 4.1c.

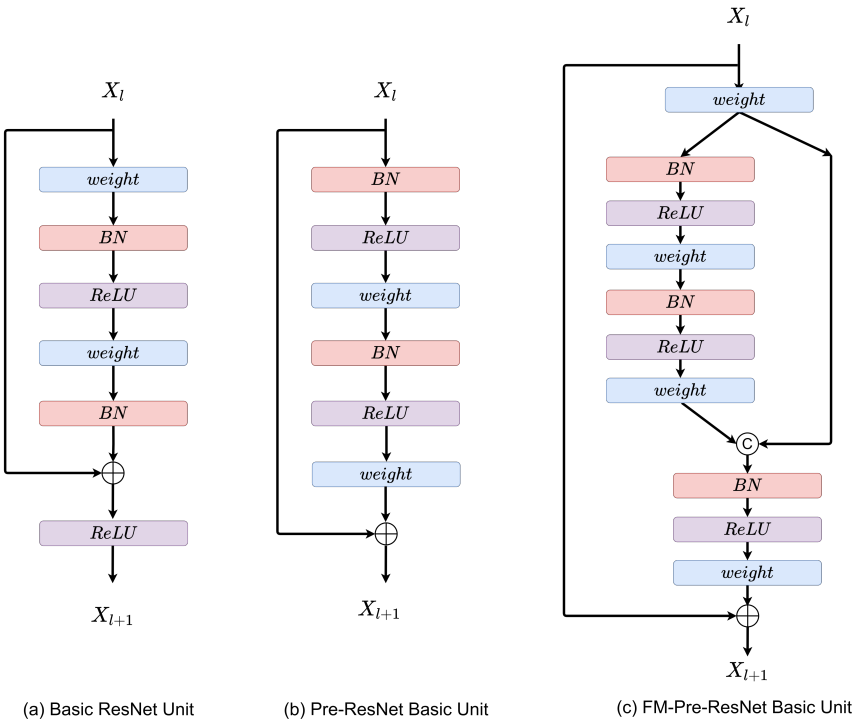


Figure 4.1: An illustration of different connectivity types of residual units. (a) Original residual unit. (b) Pre-ResNet unit. (c) Proposed FM-Pre-ResNet unit.

4.2.2 Variational Autoencoder

As discussed in Section 3.1.6, VAEs are able to capture latent representations, which makes them ideal for use in generative settings [89, 124]. Since VAEs optimization objective ELBO can be written as:

$$\mathcal{L}_{VAE}(\boldsymbol{i}, \hat{\boldsymbol{i}}) = \mathcal{L}_{REC}(\boldsymbol{i}, \hat{\boldsymbol{i}}) + \mathcal{K}, \mathcal{L}[q_\Lambda(\boldsymbol{r}|\boldsymbol{i})||p(\boldsymbol{r})] \tag{4.7}$$

where the term $\mathcal{L}_{REC}(\boldsymbol{i}, \hat{\boldsymbol{i}})$ denotes reconstruction loss and can be further written as:

$$\mathcal{L}_{REC}(\boldsymbol{i}, \hat{\boldsymbol{i}}) = -\mathbb{E}_{q_r|i}[log(p_\Delta(\boldsymbol{i}|\boldsymbol{r}))] \tag{4.8}$$

where $\hat{\boldsymbol{i}}$ denotes the reconstructed input.

The term $\mathcal{K}, \mathcal{L}[q_\Lambda(\boldsymbol{r}|\boldsymbol{i})||p(\boldsymbol{r})]$ from Equation (4.7) defines the KL divergence of the approximating variational density, which can be expressed as:

$$q_\Lambda(\boldsymbol{r}|\boldsymbol{i}) = \mathcal{N}(\boldsymbol{r}; \mu_\Lambda, \sigma^2_\Lambda) \tag{4.9}$$

The standard prior on the latent variable can be written as:

$$p(\boldsymbol{r}) = \mathcal{N}(\boldsymbol{r}; 0.1) \tag{4.10}$$

where the aligned Gaussian $(\mu_\Lambda, \sigma^2_\Lambda)$ is expressed by the encoder network $V_\Lambda(\cdot)$.

Following this, the low dimensional representations of the input data $\boldsymbol{i}$ can be obtained by introducing the latent random variable $\boldsymbol{r}$. The input images are mapped to a low dimensional space using VAE encoder $V_\Lambda(\cdot)$. After that, the output of the encoder of the segmentation network, $E_\Omega(\cdot)$, takes samples from the latent space as shown in Figure 4.2. In this manner, segmentation encoder and VAE jointly share the decoder $D_\Delta(\cdot)$, which can be also written as:

$$\mathcal{L}(\boldsymbol{o}, \hat{\boldsymbol{o}}) = \mathcal{L}_{REC}(\boldsymbol{o}, \hat{\boldsymbol{s}}) + \mathcal{K}\mathcal{L}[q_\Lambda(r|i)||p(r)] \tag{4.11}$$

The final segmentation, $\hat{o}$, is obtained from the decoder using following expression:

$$\hat{\boldsymbol{o}} = D_\Delta[E_\Omega(i) \circ V_\Lambda(i)] \tag{4.12}$$

which can be written as:

$$\hat{\boldsymbol{o}} = D_\Delta[\boldsymbol{h} \circ \boldsymbol{r}] \tag{4.13}$$

where $\circ$ denotes concatenation, $H = E_\Omega(i)$ is the output of the segmentation encoder and $r \sim q_\Lambda(r|i)$ is a sample from the latent space that is learnt by VAE.

Loss Function

The loss function plays an important role in improving the models' performance. In this work, we employ total loss function that is the addition of soft dice loss, $L2$ loss and standard VAE penalty term [125] and can be written as:

$$L = L_{dice} + 0.1 \cdot L_{L2} + 0.1 \cdot L_{KL} \tag{4.14}$$

The term that represents the soft dice loss, L_{dice} [117], can be written as:

$$L_{dice} = \frac{2 \cdot \sum P_{gt} \cdot P_{pred}}{\sum p_{gt}^2 + \sum p_{pred}^2 + \epsilon} \tag{4.15}$$

where ϵ denotes a small constant used for computational stability, i.e., to avoid
zero division.

The loss $L2$ represents loss on the VAE encoder output and can be written as:

$$L_{L2} = ||I_{input} - I_{pred}||_2^2 \tag{4.16}$$

The standard VAE penalty term, L_{KL}, represents KL divergence between a prior distribution $N(0, 1)$ and the estimated normal distribution $N(\mu, \sigma^2)$, which can be written as:

$$L_{KL} = \frac{1}{N} \sum \mu^2 + \sigma^2 - log\, \sigma^2 - 1 \tag{4.17}$$

where N represents the entire set of image voxels. Finally, the hyper-parameter weight of 0.1 is empirically found to provide a good balance between VAE loss term and soft dice loss in Equation (4.14).

4.2.3 Architecture Overview

An image segmentation task can be written as mapping:

$$g(\cdot) : I \rightarrow O \tag{4.18}$$

where $i \in I$ denote input images, while $o \in O$ denote their corresponding segmentations. For an encoder-decoder based architecture, the same mapping function can be written as:

$$g(\cdot) = E_\Omega(D_\Delta(\cdot)) \tag{4.19}$$

where E_Ω, D_Δ are an encoder and the decoder networks parametrized by Ω and Δ, respectively. Introduction of shared VAE, expressed with $V_\Lambda(\cdot)$, at the encoders' endpoint allows mapping of input images to a lower-dimensional latent, i.e., encoded, space. The output of an encoder E_Ω contains the samples from the latent space.

Therefore, our proposed architecture consists of three main stages: (1) encoding stage, (2) reconstruction of the input with variational autoencoder and (3) decoding stage. An encoding stage incorporates feature merge residual units by which the network learns a low-dimensional representation of the input. The variational autoencoder reconstructs the input image from low-dimensional latent space to regularize all model weights and adds additional guidance to the encoding stage. Finally, in the decoding stage, the network learns high-level features and creates the final segmentations. The decoder consists of FM-Pre-ResNet units. Every FM-Pre-ResNet unit in decoder doubles the spatial dimension while reduces the feature numbers by a factor of 2. Each decoder level is concatenated with the corresponding encoder output. The final layer of the decoder provides whole heart segmentation and has the same number of features and spatial size as the original input image. An illustration of the proposed architecture is shown in Figure 4.2.

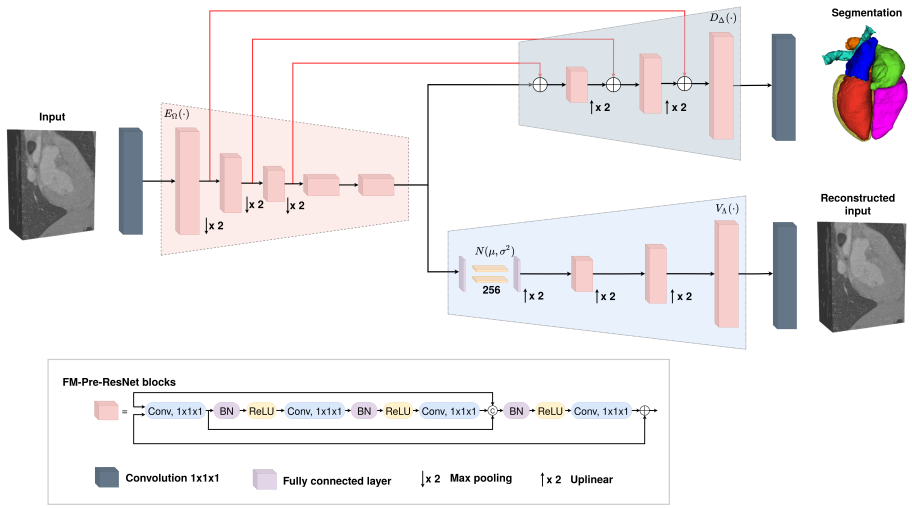


Figure 4.2: An illustration of proposed network architecture for the 3D whole heart segmentation. Input is a volumetric CT or MRI image. Each red block is the FM-Pre-ResNet block. The VAE branch is added at encoders' output and is used only during training. The decoder stage creates the final whole heart segmentation. Image source: Habijan et al. [56]

4.3 Implementation Details

In this section, we give a dataset description on which we conducted our experiments. After that, we give details about network training and implementation. We train four different networks to provide a successful ablation study: Pre-ResNet, Pre-ResNet with VAE, FM-Pre-ResNet and FM-Pre-ResNet with VAE. We evaluate the proposed

method using Multi-Modality Whole Heart Segmentation Challenge (MM-WHS) dataset and presently conducted experiments and results. Finally, we compare our results to the state-of-the-art research and provide concluding remarks.

4.3.1 Dataset Description

In this work, we use 3D CT and 3D MRI datasets provided by Multi-Modality Whole Heart Segmentation (MMWHS) Challenge organized by Zhuang, Yang and Li in conjunction with STACOM and MICCAI 2017 [160]. The CT data were acquired from 64-slice CT scanners using CT angiography at two different states. The in-plane resolution of the axial slices is (0.78×0.78) mm and the average slice thickness is 1.60 mm. The MRI data is acquired using a 1.5T Philips scanner and Siemens Magnetom Avanto 1.5T scanner. Whole heart imaging is done using 3D balanced steady-state free precession (b-SSFP) sequence and realized free-breathing scans by enabling a navigator beam before data acquisition for each cardiac phase. All the data were collected from clinical environments, so the image quality was variable. This enables an assessment of the validation and robustness of the developed algorithms with representative clinical data rather than selected best quality images.

The dataset includes 60 CT and 60 MRI whole heart images, where 20 volumes of each modality have corresponding delineations of seven heart structures performed by clinicians or by students majoring in biomedical engineering or medical physicists using the semi-automatic ITK-SNAP software [177]. Each manual segmentation result was examined by senior researchers specialized in cardiac imaging with experience of more than five years and modifications have been taken if a revision was necessary. These seven structures include the following: LV, RV, LA, RA, the myocardium of the LV (Myo), the trunk from the aortic valve to the superior level of the atria (Ao) and the trunk from the pulmonary valve to the bifurcation point (PA). The remaining 40 volumetric images are used for testing purposes. Ground truths of the testing dataset are provided in encrypted form and can be decoded to evaluate algorithms using the procedure described in [160]. An example of input images in different views with corresponding manual segmentation is shown in Figure 4.3.

4.3.2 Data Preprocessing and Augmentation

To alleviate the irregularities of variable contrast in some MRI images, we normalize all input images (both CT and MRI) to have zero mean and unit std. The volumes were cropped and zero-padded to a fixed size of $176 \times 224 \times 144$ to provide a fine ROI for the network input while making sure all heart structures are inside the selected ROI. We apply three different data augmentation methods to increase the sample

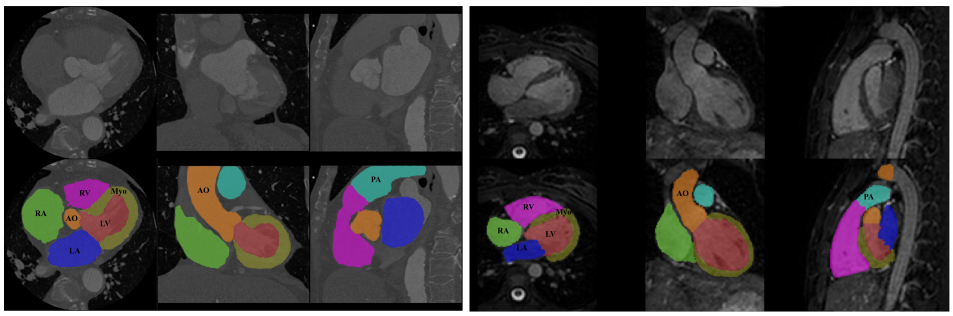


Figure 4.3: An example of one slice with corresponding ground truth from 3D volume across axial, coronal and sagittal planes. The ground truths include seven heart structures: LV (red), RV (magenta), LA (blue), RA (green), Myo (yellow), Ao (orange) and PA (cyan). Image source: Habijan et al. [56]

size of training data and enhance the robustness and generalization ability, namely random axis mirror flip, random scaling and intensity shift. Random axis mirror flip creates a mirror reflection of an original image along one (or more) selected axis and is commonly flipped at a rate of 50%. Random scaling operation S scales input image and performs independently in different directions. Intensity shift performs an element-wise addition of a scalar to the image and affects the brightness of the original image. Details about parameters of used data augmentation methods are presented in Table 4.1, while examples of input images after applying different data augmentation methods are shown in Figure 4.4. Moreover, we empirically found that advanced augmentation techniques, such as random histogram matching, or random image filtering, do not show any additional improvements to the final segmentation result.

Table 4.1: Data augmentation parameters.

Method	Parameters
Random flip along all axis	with probability 0.5
Random scale	$S \in [0.9, 1.1]$
Intensity shift	between $[-0.1, 0.1]$

4.3.3 Network Implementation and Training

In our experiments, we train four encoder-decoder based architectures: (1) 3D Pre-ResNet without VAE regularization, (2) 3D Pre-ResNet with VAE regularization, (3) FM-Pre-ResNet without VAE regularization and (4) FM-Pre-ResNet with VAE regularization. All four networks are trained from scratch and separately for CT and MRI images. The whole experimental procedure is implemented in Pytorch and trained on two NVIDIA Titan V100 GPUs, simultaneously. Our training and validation dataset consists of 20 CT volumes and 20 MRI volumes,

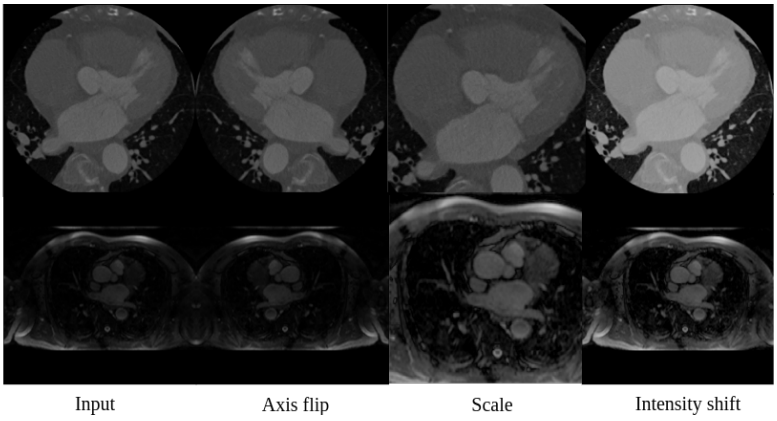


Figure 4.4: An example of different augmentation methods. Top row, from left to right on: input CT image, image after axis flip, image after scale, image after intensity shift. Bottom row, from left to right on: input MRI image, image after axis flip, image after scale, image after intensity shift.

with $80\% - 20\%$ training and validation split, respectively. We use adaptive learning rate optimizer Adam with an initial learning rate of $\alpha_0 = 10^{-4}$ and gradually decrease it according to following expression:

$$\alpha = \alpha_0 * \left(1 - \frac{c}{T_c}\right)^{0.9} \tag{4.20}$$

where T_c is a total number of epochs (200 in our case) and c is an epoch counter.

Furthermore, to ensure our models generalizes well on unseen data, i.e. to reduce the effect of overfitting or underfitting, we employ $L2$ norm regularization with a weight of 10^{-5} and the spatial dropout with a rate of 0.2 after the initial encoder convolution. Since early stopping aims at finding the network parameters at the point of the lowest validation loss we implement early stopping with patience set to 50. All trained networks are evaluated using a testing dataset that includes 40 subjects for both CT and MRI images [161].

Moreover, the network is trained for 200 epochs since further training appears not to decrease validation loss as shown in Figure 4.5 and Figure 4.6. Moreover, Figure 4.5 and Figure 4.6 indicate decrease in loss value when number of epochs increases. This is a clear indication that the network is successfully learning from the input data. We can also see significant improvement regarding training and validation accuracies of networks with VAEs, which confirms our initial assumption that VAE successfully provides regulation during training.

Furthermore, Pre-ResNet has demonstrated that increasing the depth of the network improves model performance significantly. The addition of two convolutional layers at the top and bottom of the pre-activation residual block introduced in our FM-Pre-ResNet unit allows for the feature fusion block to reach the same depth with fewer

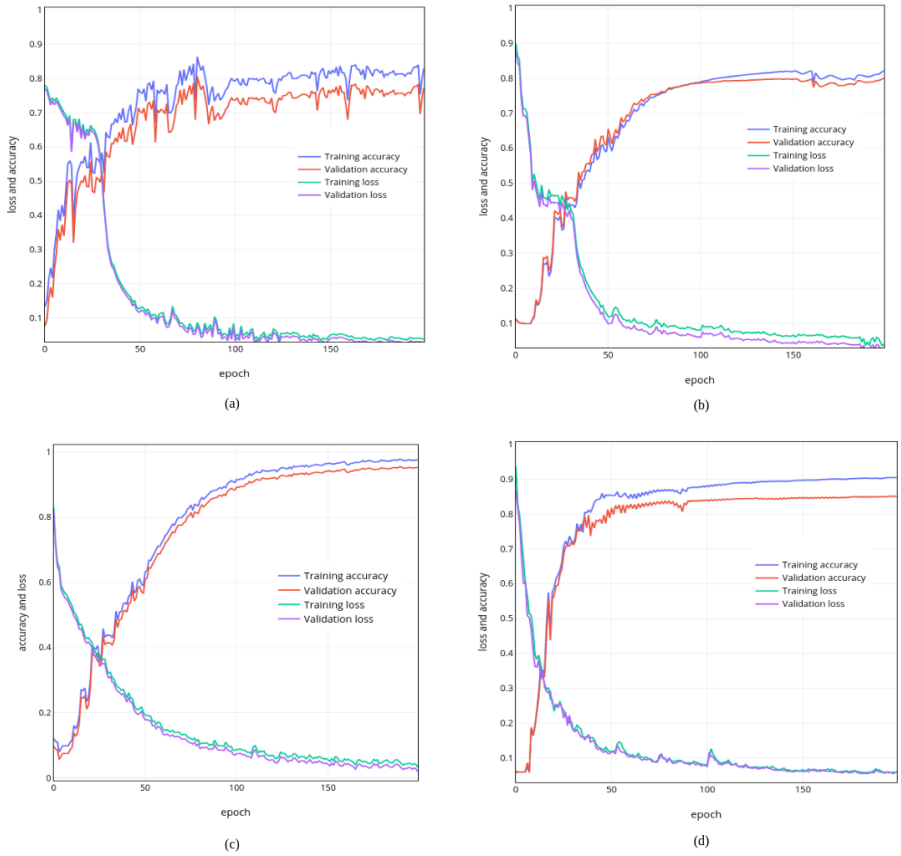


Figure 4.5: Training and validation accuracies on CT dataset. (a) 3D Pre-ResNet network architecture, (b) 3D FM-Pre-ResNet network architecture, (c) 3D Pre-ResNet + VAE network architecture, (d) 3D FM-Pre-ResNet + VAE network architecture.

parameters which benefits model performance. Therefore, the proposed type of connectivity of the FM-Pre-ResNet unit in terms of depth and number of parameters regarding Pre-ResNet implies no increase in the number of parameters compared to the Pre-ResNet. Time-wise, each training epoch (200 cases) and prediction times on two GPU-s (NVIDIA Titan V) are significantly reduced with architectures with VAE. This shows the computational efficiency of our choice for VAE introduction. Comparison of depth, number of parameters, training times per epoch and prediction time of one volume for different architectures is shown in Table 4.2.

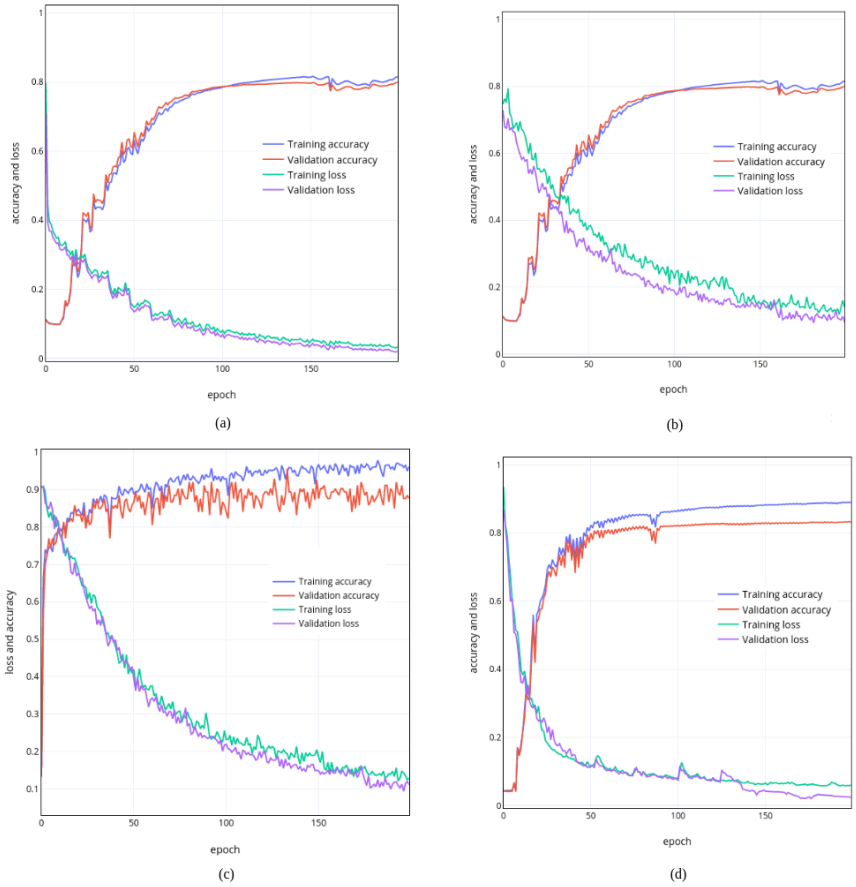


Figure 4.6: Training and validation accuracies and losses on MRI dataset. (a) 3D Pre-ResNet network architecture, (b) 3D FM-Pre-ResNet network architecture, (c) 3D Pre-ResNet + VAE network architecture, (d) 3D FM-Pre-ResNet + VAE network architecture.

Table 4.2: Comparison of depth, number of parameters ($\times 10^6$), training times per epoch (min) and prediction time (sec) for one volume for different architectures: Pre-ResNet, 3D Pre-ResNet + VAE, FM-Pre-ResNet and FM-Pre-ResNet + VAE.

Architecture	Depth	Number of Parameters	Training Time	Prediction Time
3D Pre-ResNet	110	23.48	10	0.7
3D Pre-ResNet + VAE	110	26.18	8	0.6
3D FM-Pre-ResNet	218	22.54	9	0.5
3D FM-Pre-ResNet + VAE	218	25.14	7	0.4

4.4 Experiments and Results

To demonstrate the effectiveness of the proposed approach and our design choice for the new FM-ResNet unit, we train encoder-decoder-based architecture using 3D Pre-ResNet without and with VAE regularization and the proposed method FM-Pre-ResNet without and with VAE regularization. To evaluate the proposed methodology performance, we compare ground truth masks with obtained segmentations for each CT and MRI volume. We used commonly used four metrics to evaluate segmentation accuracy, namely, DSC, JI, SD and HD, which are discussed in subsection 3.2.4. For ease of understanding, it is helpful to mention that DSC and JI should be high as possible, while SD and HD should be as low as possible. Table 4.3 summarizes an average whole heart segmentation results on CT and MRI images.

On CT images, the 3D Pre-ResNet network achieves average WHS segmentation results for DSC, JI, SD and HD of 87.11%, 80.16%, 1.71 mm and 24.44 mm, respectively. The addition of VAE at Pre-ResNet segmentation encoders' endpoint improve DSC, JI, SD and HD values for 2.12%, 1.0%, 0.2039 mm and 2.704 mm, respectively. The 3D FM-Pre-ResNet network achieves DSC, JI, SD and HD values of 90.03%, 82.14%, 1.43 mm and 18.82 mm, respectively. Compared to 3D Pre-ResNet, it achieves improvement in DSC, JI, SD and HD values of 2.92%, 1.98%, 0.2789 mm and $5,6307$ mm, which means that the proposed FM-PreResNet unit significantly improves segmentation accuracy. Moreover, the highest DSC, JI, SD and HD are achieved using 3D FM-Pre-ResNet + VAE network and report values of 90.39%, 82.24%, 1.1093 mm and 15.3621 mm, respectively. Similarly, on MRI images, the 3D Pre-ResNet network achieves average WHS segmentation results for DSC, JI, SD and HD of 83.06%, 75.54%, 5.9201 mm and 42.5578 mm, respectively. The addition of VAE at Pre-ResNet segmentation encoders' endpoint improve DSC, JI,SD and HD values for 2.28%, 0.09%, 2.15 mm 3.6766 mm. The 3D FM-Pre-ResNet network achieves average DSC, JI, SD and HD values of 88.40%, 78.55%, 2.4558 mm and 32.0451 mm, respectively. Compared to 3D Pre-ResNet, it achieves improvement in DSC, JI, SD and HD values of 5.34%, 3.01%, 3.4643 mm and 10.5127 mm, which means that the proposed FM-PreResNet unit significantly improves segmentation accuracy.

Moreover, the highest DSC, JI, SD and HD are achieved using 3D FM-Pre-ResNet + VAE network and report values of 89.50%, 80.44%, 1.8599 mm and 25.6558 mm, respectively. These results highlight the improvement in segmentation accuracy afforded by the introduction of FM-Pre-ResNet units and VAE. Boxplots showing the distribution of the DSC for WH, LV, Myo, LA, RA, RV, AO and PA using different segmentation networks on MMWHS CT and MRI testing datasets are presented in Fig 4.7 and Fig 4.8, respectively. Boxplots illustrate interquartile range (bounds of box), mean (X inside a box), median

Table 4.3: Comparison of an average WHS results in terms of DSC, JI, SD and HD on different architectures for CT and MRI testing dataset

	CT				MRI			
Architecture	DSC	JI	SD	HD	DSC	JI	SD	HD
3D Pre-ResNet	0.8711 ±0.0721	0.8016 ±0.0609	1.7110 ±0.4991	24.4421 ±17.8355	0.8306 ±0.9254	0.7554 ±0.0581	5.9201 ±0.4421	42.5578 ±21.6645
3D Pre-ResNet + VAE	0.8923 ±0.0209	0.8116 ±0.0358	1.5071 ±1.407	21.7381 ±16.8850	0.8534 ±0.0441	0.7545 ±0.0583	3.7701 ±0.9100	38.8812 ±23.5812
3D FM-Pre-ResNet	0.9003 ±0.0148	0.8214 ±0.0271	1.4321 ±0.0518	18.8114 ±12.4032	0.8840 ±0.0701	0.7855 ±0.0455	2.4558 ±0.7956	32.0451 ±17.5508
3D FM-Pre-ResNet + VAE	**0.9039 ±0.0517**	**0.8224 ±0.0571**	**1.1093 ±0.0215**	**1.1093 ±12.3737**	**0.8950 ±0.0215**	**0.8044 ±0.0757**	**1.8599 ±0.6740**	**25.6558 ±16.4001**

(centerline), maximum and minimum values (whiskers) and outliers (circles outside whiskers).

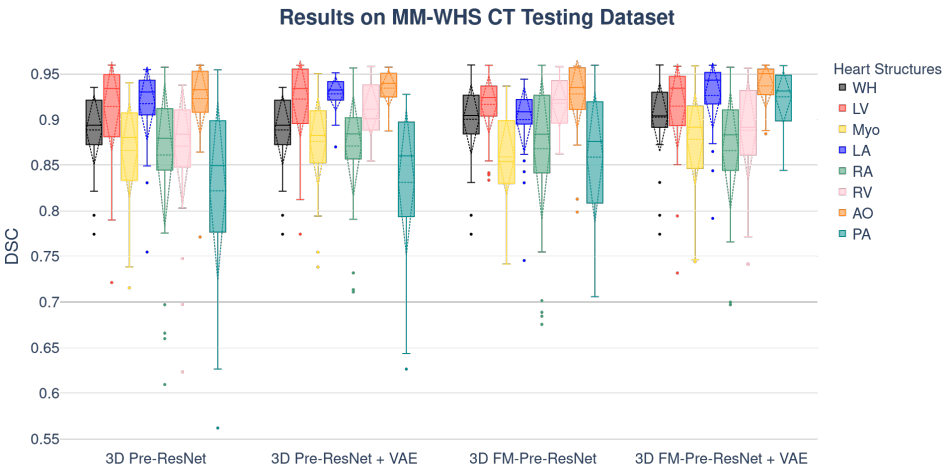


Figure 4.7: Boxplots showing the DSC dispersion for WH, LV, Myo, LA, RA, RV, AO and PA using different segmentation networks on the MMWHS CT testing dataset.

The p-values have been calculated using a Wilcoxon rank-sum test to show the significant difference of used architectures. Bonferroni correction was used for controlling the family-wise error rate. Figures 4.9 and 4.10 show the comparisons of p-values for CT and MRI testing datasets, respectively. The visual inspection of the obtained segmentations using each network investigated in this work are presented in Figure 4.11 for the CT dataset and Figure 4.12 for the MRI dataset.

For example, Figure 4.12(d) shows clear improvements regarding LV segmentation that is obtained using FM-Pre-ResNet compared to missed segmentation of LV parts while using Pre-ResNet without proposed feature merge residual unit as shown in Figure 4.12(b). Moreover, Figure 4.12(f) shows a significant reduction in segmentation error than all other presented networks. These further highlights the benefits of the proposed FM-Pre-ResNet + VAE approach. Nonetheless, in both modalities, PA and Myo's segmentation results are significantly lower

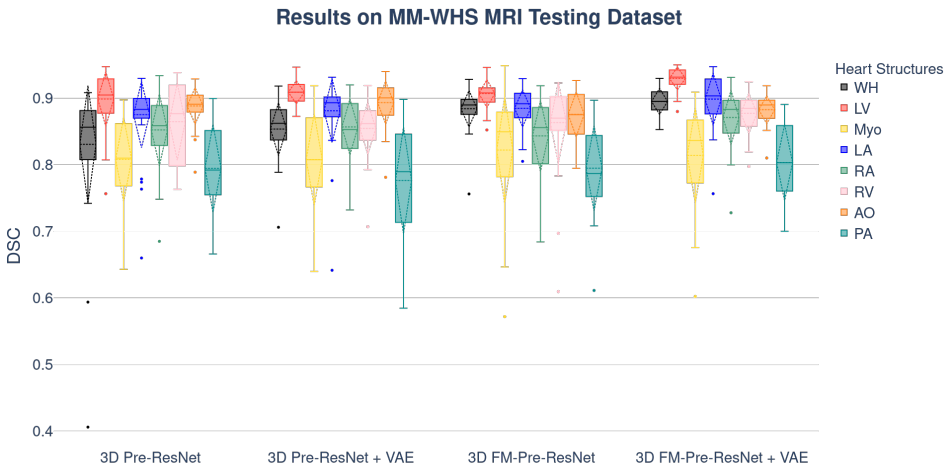


Figure 4.8: Boxplots showing the DSC dispersion for WH, LV, Myo, LA, RA, RV, AO and PA using different segmentation networks on the MMWHS MRI testing dataset.

than other substructures due to high shape variations and heterogeneous intensity of blood fluctuations. Figure 6.7 shows 3D visualization of the best and the worse segmentation cases on the CT and MRI test dataset obtained using the proposed FM-Pre-ResNet + VAE approach. Furthermore, Pre-ResNet has demonstrated that increasing the depth of the network improves model performance significantly. The addition of two convolutional layers at the top and bottom of the pre-activation residual block introduced in our FM-Pre-ResNet unit allows for the feature fusion block to reach the same depth with fewer parameters which benefits model performance.

Additional, structure-wise segmentation accuracies for the LV, RV, LA, RA, Myo, Ao and PA, for both CT and MRI images, are summarized in Table 4.4 and 4.5.

Comparison with Other Methods

The proposed approach was compared with other similar deep learning approaches in terms of image segmentation accuracy as shown in Table 4.6 and Table 4.7. An approach that combines atlas registration with CNNs' [48] provides an incremental segmentation that allows user interaction, which can be beneficial in a clinical setting. Nevertheless, the challenges of accurate atlas registration resulted in low accuracy on MRI images. Deep supervision mechanism [135] and use of transfer learning [98] result in an increase of trainable parameters and overall network complexity. In contrast, we aim to introduce a lightweight network that results in a significantly deep network without increasing parameter number. Moreover, in [98] report an average WHS DSC of 0.914 ± 0.075 on CT images and $.890 \pm 0.054$ on MRI images using a hold-out set of 10% of training data and evaluate their method with 10-fold cross-validation. Our results report 0.9039 ± 0.0517 on CT

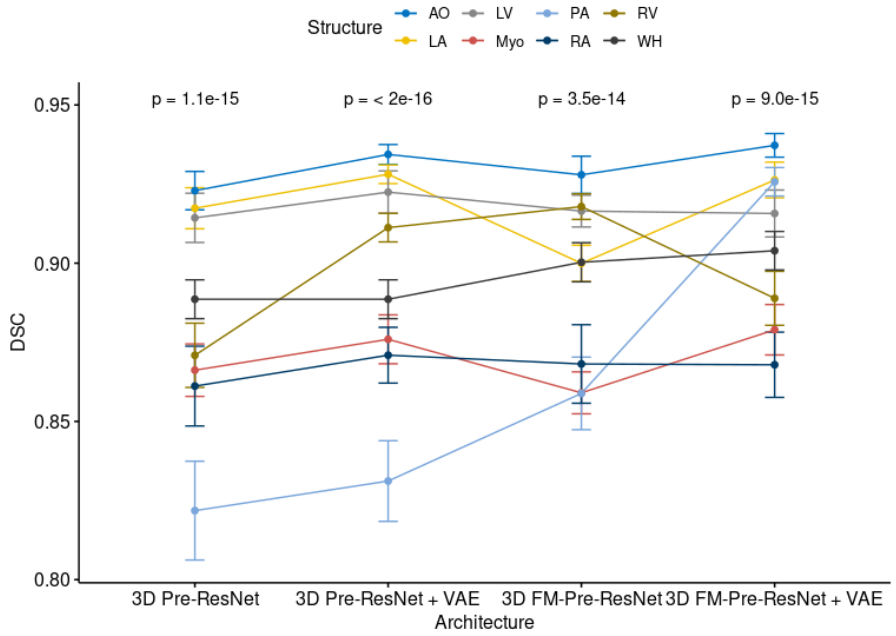


Figure 4.9: Comparison of Wilcoxon rank sum test of each heart structure for different architectures on the MM-WHS CT testing dataset.

images and 0.8950 ± 0.0215 on MRI images and are evaluated on all unseen 40 subjects, which shows that the VAE stage's introduction significantly helps in overcoming overfitting problems. Therefore, these results highlight the advantages of our proposed method.

4.5 Conclusion

Accurate heart and its substructures segmentation enable faster visualization of target structures and data navigation, which benefits clinical practice by reducing diagnosis and prognosis times. This chapter introduced an encoder–decoder-based architecture for whole heart segmentation on CT and MRI images. Our proposed method introduces a novel connectivity structure of residual unit that we refer to as feature merge residual unit (FM-Pre-ResNet). The proposed connectivity allows the creation of distinctly deep models without an increase in the number of parameters compared to the Pre-ResNet units. FM-Pre-ResNet enables the construction of profound models without increasing the number of parameters in comparison to pre-activation residual units. By incorporating two convolutional layers at the top and bottom of the pre-activation residual block, the parameters of the two branches are balanced. In comparison, the bottom layer reduces the dimension of the channel. This allows for the construction of a more detailed model with a similar number of parameters to the initial pre-activation residual unit.

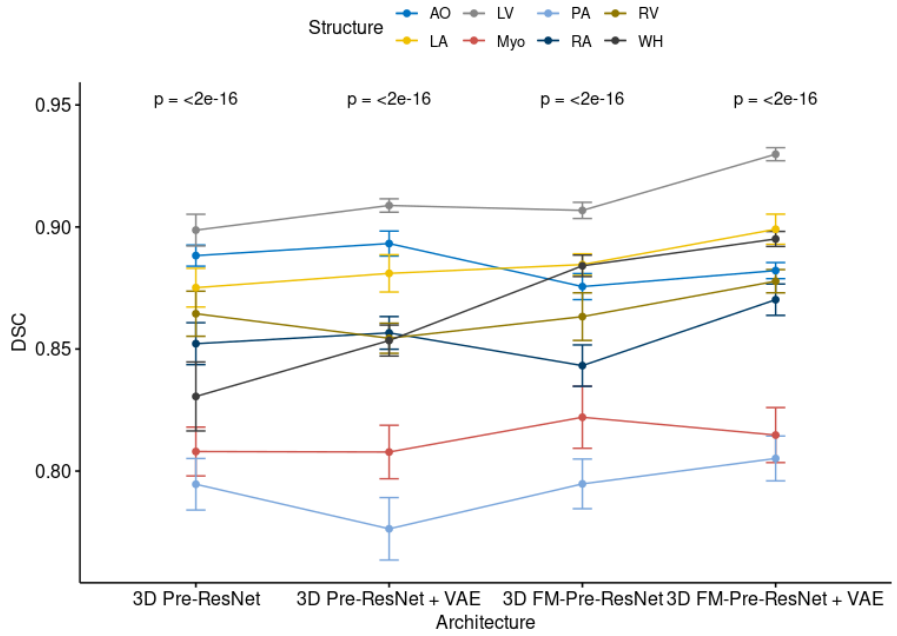


Figure 4.10: Comparison of Wilcoxon rank sum test of each heart structure for different architectures on the MMWHS MRI testing dataset.

Furthermore, we construct an encoder–decoder-based architecture that incorporates the VAE encoder at the segmentation encoder output to have a regularizing effect on the encoder layers. FM-Pre-ResNet units are used to learn a low-dimensional representation of the input during the encoding stage. Following that, VAE reconstructs the input image from the low-dimensional latent space, ensuring that the model weights are strongly regularized while also avoiding over-fitting on the training data. The VAE acts as a regulator of model weights, adds additional guidance and exploits the encoder endpoint features. In the end, the segmentation decoder learns high-level features and creates the final segmentations. We evaluated the proposed approach on MMWHS CT and MRI testing datasets and obtain average WHS DSC, JI, SD and HD values of 90.39%, 82.24%, 1.1093, 15.3621 for CT images and 89.50%, 80.44%, 1.8599, 25.6558 for MRI images, respectively. Results for CT datasets are highly comparable to the state-of-the-art.

Table 4.4: Structure-wise DSC evaluation of proposed architecture and other 3D based architectures in terms of DSC, JI, HD, SD on CT testing dataset for LV, RV, LA, RA, Myo, Ao and PA

		Heart Structure						
Metrics	Architecture	LV	Myo	RV	LA	RA	AO	PA
DSC	3D Pre-ResNet	0.9165 ±0.0512	0.8662 ±0.0524	0.8709 ±0.0642	0.9181 ±0.0417	0.8609 ±0.08	0.9251 ±0.4404	0.8093 ± 0.1331
	3D Pre-ResNet +VAE	0.9245 ±0.0176	0.8762 ±0.0212	0.9124 ±0.009	0.9281 ±0.0392	0.8709 ± 0.0532	0.935 ±0.0149	0.8311 ± 0.0199
	3D FM-Pre-ResNet	0.9165 ±0.0125	0.851 ±0.015	0.9179 ±0.0063	0.899 ±0.0277	0.8683 ± 0.0376	0.9326 ±0.0105	0.9272 ± 0.0141
	3D FM-Pre-ResNet + VAE	**0.9177 ±0.049**	**0.8791 ±0.0504**	**0.8882 ±0.0546**	**0.931 1 ±0.0396**	**0.8617 ±0.0802**	**0.9449 ±0.0404**	**0.8271 ± 0.1331**
JI	3D Pre-ResNet	0.8501 ±0.0814	0.7675 ±0.0786	0.7764 ±0.0914	0.8511 ±0.0668	0.7635 ±0.1121	0.863 ±0.0666	0.6973 ± 0.163
	3D Pre-ResNet +VAE	0.8601 ±0.0762	0.7775 ±0.0329	0.7863 ±0.0155	0.8611 ±0.068	0.7734 ± 0.089	0.873 ±0.0266	0.7073 ± 0.0338
	3D FM-Pre-ResNet	0.8699 ±0.0398	0.7873 ±0.0488	0.7961 ±0.0338	0.8709 ±0.0323	0.7832 ± 0.0592	0.8828 ±0.0601	0.7171 ± 0.0756
	3D FM-Pre-ResNet + VAE	**0.8709 ±0.0573**	**0.7883 ±0.0725**	**0.7971 ±0.0834**	**0.8719 ±0.0736**	**0.7842 ±0.1131**	**0.8838 ±0.0568**	**0.7181 ± 0.1449**
SD	3D Pre-ResNet	0.1078 ±0.5188	1.3061 ±0.6522	1.4767 ±0.764	1.2568 ±0.7873	1.7143 ±0.8301	0.8131 ±0.4853	1.8828 ± 2.5626
	3D Pre-ResNet +VAE	1.0778 ±0.4210	1.2544 ±0.6003	1.3574 ±0.5321	1.2047 ±0.5504	1.6980 ± 0.4321	0.6251 ±0.7001	1.6320 ± 1.0848
	3D FM-Pre-ResNet	0.9321 ±0.7701	1.1178 ±0.5987	1.2047 ±0.4895	1.0157 ±0.7754	1.5534 ± 0.3305	0.5220 ±0.0653	1.5884 ± 1.0012
	3D FM-Pre-ResNet + VAE	**0.7455 ±0.8905**	**1.0057 ±0.3210**	**0.9907 ±0.2078**	**1.1775 ±0.6055**	**1.3544 ±0.5587**	**0.4444 ±0.3217**	**1.735 ± 1.0997**
HD	3D Pre-ResNet	9.5403 ±4.8047	13.573 ±4.5287	14.3229 ±13.1375	10.3919 ±6.7654	13.0453 ±6.9765	8.0746 ±4.2339	10.3851 ± 13.1497
	3D Pre-ResNet +VAE	7.5402 ±4.0019	12.4457 ±3.9210	13.5571 ±11.2474	9.0781 ±5.4880	14.210 ±5.7871	9.7758 ±5.4421	12.8835 ±15.5432
	3D FM-Pre-ResNet	7.0037 ±3.5707	10.7785 ±3.7500	10.0787 ±9.457	9.4743 ±4.7171	11.0375 ±3.8810	8.1170 ±3.5778	10.5532 ±3.4210
	3D FM-Pre-ResNet + VAE	**5.5011 ±2.3088**	**8.3257 ±2.9901**	**7.3854 ±7.7809**	**8.7555 ±3.2089**	**9.5777 ±3.5432**	**6.5781 ±6.5001**	**9.5587 ±8.5578**

Table 4.5: Structure-wise DSC evaluation of proposed architecture and other 3D based architectures in terms of DSC, JI, HD, SD on MRI testing dataset for LV, RV, LA, RA, Myo, Ao and PA

		Heart Structure						
Metrics	Architecture	LV	Myo	RV	LA	RA	AO	PA
DSC	3D Pre-ResNet	0.9014 ±0.0342	0.8088 ±0.0178	0.8644 ±0.0457	0.8751 ±0.0111	0.8521 ±0.1089	0.8891 ±0.2701	0.7945 ± 0.3002
	3D Pre-ResNet +VAE	0.9121 ±0.0458	0.8077 ±0.0388	0.8544 ±0.1882	0.8810 ±0.0157	0.8566 ± 0.0501	0.8932 ±0.0327	0.7763 ± 0.0497
	3D FM-Pre-ResNet	0.9080 ±0.0102	0.8220 ±0.0245	0.8632 ±0.0233	0.8846 ±0.0589	0.8432 ± 0.0799	0.8755 ±0.0301	0.7947 ± 0.0243
	3D FM-Pre-ResNet + VAE	**0.9313 ±0.0885**	**0.8147 ±0.0119**	**0.8777 ±0.0154**	**0.9017 ±0.0867**	**0.8702 ±0.0146**	**0.8821 ±0.0137**	**0.8020 ± 0.1102**
JI	3D Pre-ResNet	0.8005 ±0.1155	0.6222 ±0.1235	0.7129 ±0.1494	0.7419 ±0.1084	0.7051 ±0.1453	0.7208 ±0.1395	0.6076 ± 0.1286
	3D Pre-ResNet +VAE	0.8344 ±0.0587	0.7178 ±0.0441	0.7123 ±0.0328	0.7942 ±0.0758	0.7108 ± 0.0107	0.8155 ±0.0789	0.6328 ± 0.0977
	3D FM-Pre-ResNet	0.8001 ±0.06732	0.7244 ±0.0483	0.7732 ±0.1652	0.8155 ±0.0559	0.7201 ± 0.1551	0.8053 ±0.1344	0.6855 ± 0.0266
	3D FM-Pre-ResNet + VAE	**0.8159 ±0.0.0891**	**0.7388 ±0.0552**	**0.7244 ±0.0341**	**0.8053 ±0.0322**	**0.7221 ±0.2175**	**0.8147 ±0.0285**	**0.7095 ± 0.1532**
SD	3D Pre-ResNet	3.1154 ±4.2951	4.1305 ±4.4141	3.8078 ±5.6198	1.9685 ±1.8108	3.1319 ±3.0756	1.7262 ±1.8632	1.9394 ± 0.8231
	3D Pre-ResNet +VAE	2.0102 ±3.0051	3.7214 ±3.2708	2.5699 ±4.3201	1.5421 ±1.3037	2.6542 ±2.7822	1.2201 ±1.2447	1.5572 ±0.6241
	3D FM-Pre-ResNet	1.4425 ±0.6055	2.1778 ±4.2871	2.8321 ±3.5542	1.7728 ±1.4002	2.4880 ±2.3551	1.0027 ±1.1998	2.3571 ±0.7581
	3D FM-Pre-ResNet + VAE	**0.9789 ±1.7757**	**1.7562 ±1.3321**	**1.2552 ±1.9947**	**1.8853 ±1.5570**	**1.99722 ±1.8771**	**0.6799 ±0.7844**	**2.0774 ± 0.8231**
HD	3D Pre-ResNet	33.6531 ±23.5248	38.8297 ±29.8463	31.2102 ±27.1629	17.6381 ±15.0182	31.2076 ±27.6534	9.5942 ±7.5978	10.3042 ±4.1532
	3D Pre-ResNet +VAE	31.5542 ±18.2863	35.5541 ±25.8371	28.8105 ±21.4779	17.5428 ±11.3571	24.7579 ±23.8901	8.7709 ±6.3481	8.5721 ±2.7799
	3D FM-Pre-ResNet	29.8821 ±14.5887	36.4528 ±27.3378	25.7773 ±19.8421	18.5789 ±9.2297	26.8832 ±25.7892	7.2027 ±3.5599	11.2577 ±6.7987
	3D FM-Pre-ResNet + VAE	**26.5428 ±11.4450**	**34.1750 ±18.2889**	**23.5771 ±14.543**	**19.7750 ±9.4798**	**16.7750 ±6.9543**	**5.5897 ±3.4201**	**9.4477 ±3.5947**

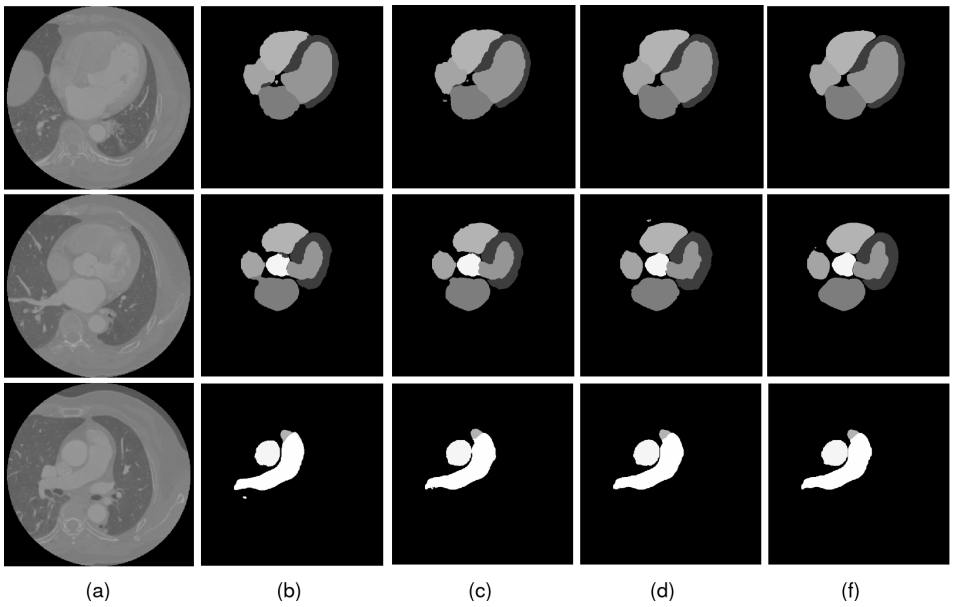


Figure 4.11: Comparison of the results of four different network architectures. (a) The input original CT image. (b) Segmentation result of Pre-ResNet without VAE. (c) Segmentation result of Pre-ResNet with VAE. (d) Segmentation result of FM-Pre-ResNet without VAE. (f) Segmentation result of proposed FM-Pre-ResNet with VAE obtains the most accurate results on the testing dataset. Image source: Habijan et al. [56]

Table 4.6: Comparison of an average DSC, JI, SD and HD of the state-of-the-art whole heart segmentation methods on CT images.

Authors	Method	DSC	JI	SD(mm)	HD(mm)
Galisot et al. [48]	Multi Atlas + CNN	0.838 ± 0.152	0.742 ± 0.161	4.812 ± 13.604	34.634 ± 12.351
Payer et al. [18]	Localization + segmentaiton CNN	0.908 ± 0.086	0.832 ± 0.037	1.117 ± 0.250	25.242 ± 10.813
Mortazi et al. [3]	multi planar CNN	0.879 ± 0.079	0.792 ± 0.106	1.538 ± 1.006	28.481 ± 11.434
Wang et al. [20]	Statistical shape priors + CNN	0.894 ± 0.030	0.810 ± 0.048	1.387 ± 0.516	31.146 ± 13.203
Tong et al. [135]	Deeply supervised 3D U-Net	0.849 ± 0.061	0.742 ± 0.086	1.925 ± 0.924	44.880 ± 16.084
Liao et al. [98]	multi planar 2D CNN	0.914 ± 0.075	0.840 ± 0.075	1.42 ± 0.46	28.042 ± 12.142
Ours	FM-Pre-ResNet + VAE	0.9039 ± 0.0517	0.8224 ± 0.0571	1.1093 ± 0.0215	15.362 ± 12.3737

Table 4.7: Comparison of an average DSC, JI, SD and HD of the state-of-the-art whole heart segmentation methods on MRI images.

Authors	Method	DSC	JI	SD(mm)	HD(mm)
Galisot et al. [48]	Multi Atlas + CNN	0.817 ± 0.059	0.695 ± 0.081	2.420 ± 0.925	30.938 ± 12.190
Payer et al. [18]	Localization + segmentaiton CNN	0.863 ± 0.043	0.762 ± 0.064	1.890 ± 0.781	30.227 ± 14.046
Mortazi et al. [3]	multi planar CNN	0.818 ± 0.096	0.701 ± 0.118	3.040 ± 3.097	40.092 ± 21.119
Wang et al. [20]	Statistical shape priors + CNN	0.855 ± 0.069	0.753 ± 0.094	1.963 ± 1.012	30.201 ± 13.2216
Tong et al. [135]	Deeply supervised 3D U-Net	0.674 ± 0.182	0.532 ± 0.178	9.776 ± 0.924	44.880 ± 16.084
Liao et al. [98]	multi planar 2D CNN	0.914 ± 0.075	0.840 ± 0.075	1.42 ± 0.46	28.042 ± 12.142
Ours	FM-Pre-ResNet + VAE	0.8950 ± 0.0215	0.8044 ± 0.0757	1.8599 ± 0.6740	25.6558 ± 16.4001

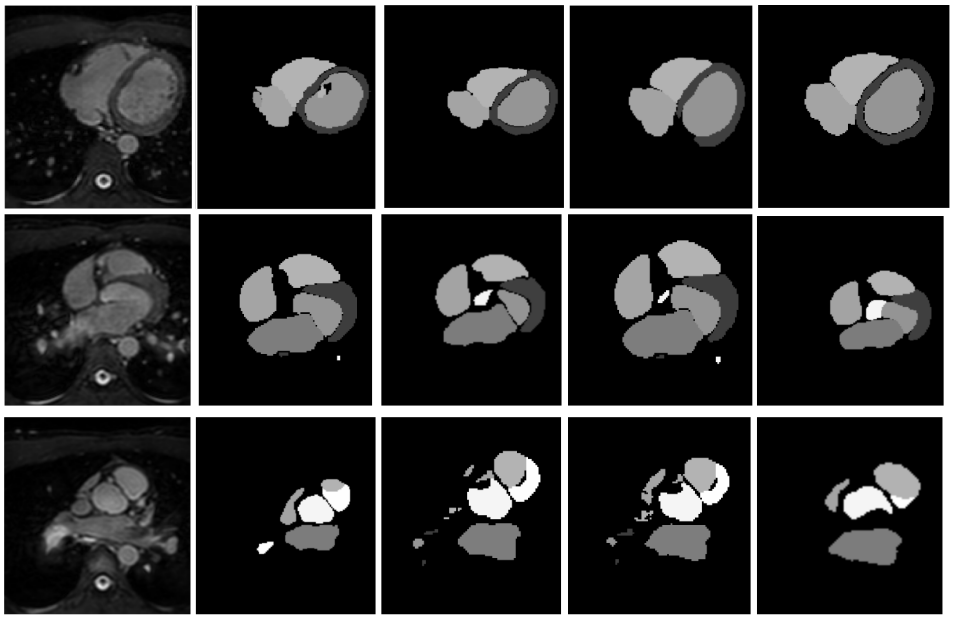


Figure 4.12: Comparison of the results for four different network architectures. (a) The input original MRI images. (b) Segmentation result of Pre-ResNet without VAE. (c) Segmentation result of Pre-ResNet with VAE. (d) Segmentation result of FM-Pre-ResNet without VAE. (f) Segmentation result of proposed FM-Pre-ResNet with VAE obtains the most accurate results on the testing dataset. Image source: Habijan et al. [56]

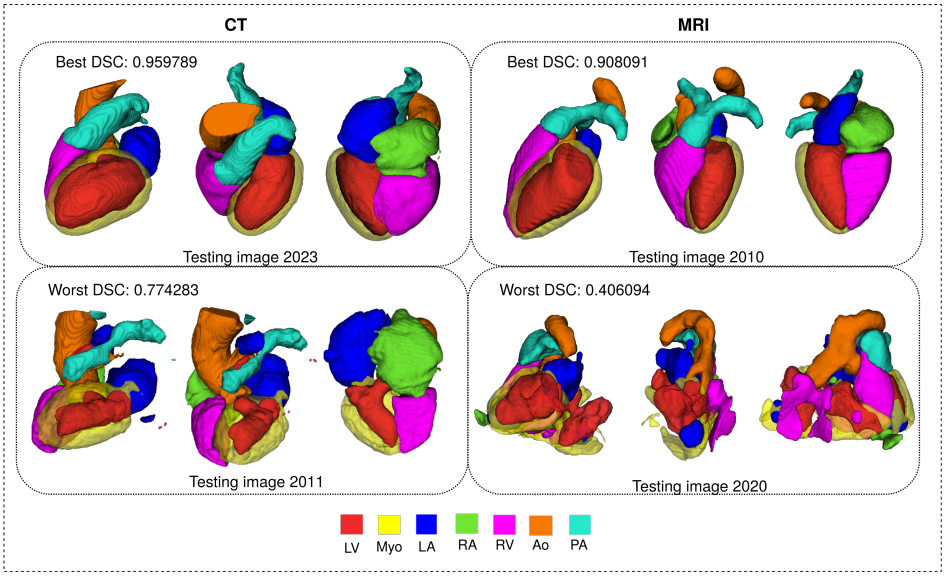


Figure 4.13: 3D visualization of the best and worse cases of WHS results in the CT and MRI test dataset. Image source: Habijan et al. [56]

CHAPTER 5

Bi-Ventricles and Myocardium Segmentation

This chapter presents a new automatic method for LV, RV and Myo segmentation from Cine MRI images. We introduce a new architecture that incorporates SERes blocks into 3D U-Net architecture (3D SERes-U-Net). The SERes blocks incorporate squeeze-and-excitation operations into residual learning. The adaptive feature recalibration ability of squeeze-and-excitation operations boosts the network's representational power while feature reuse utilizes effective learning of the features, which improves segmentation performance. We evaluate the proposed method on the MICCAI Automated Cardiac Diagnosis Challenge (ACDC) testing dataset. Our pipeline obtains an average DSC for LV, RV and Myo at end-diastole of 95%, 90%, 83%, respectively. Similarly, we obtain an average DSC for LV, RV and Myo at end-systole of 86%, 83%, 85%, respectively. We calculate significant clinical metrics, i.e., indicators of hearts function, including volume of the left ventricle at end-diastole (LVEDV), LVESV, LVEF, RVEDV, RVESV, RVEF, MyoLVES and MyoMED. The Bland-Altman and analysis show a high correlation coefficient of R=0.99 for LVEDV and LVESV, while R=0.95 for LVEF. Correlations of RVEDV, EVESV and RVEF are R=0.97, R=0.93, R=0.69, respectively. Finally, R=0.96 for MyoLVES and R=0.95 for MyoMED further show the strength of accuracy and precision of our proposed method.

The outline of the chapter is structured in the following manner. Section 5.1 gives the main objectives of conducted research. Section 5.2 gives a theoretical background of used methods and describes our proposed method for LV, RV and Myo segmentation. Section 5.4 describes the experimental setup, gives network training details, and presents obtained results. Finally, concluding remarks are provided in Section 5.5.

5.1 Objectives

This research aims to develop an efficient method for fully automatic segmentation of LV, RV and Myo from Cine MRI images. As discussed in Section 3.1, the basic building block for most CNN architectures is the convolution layer. The convolutional layer learns by capturing local spatial patterns along all the input channels and generates feature maps jointly encoding the spatial and channel information. While much effort is put into improving this encoding of spatial and channel information, encoding of the spatial and channel-wise patterns independently is less explored. Recent work attempted to address this issue by explicitly modeling the interdependencies between the channels of feature maps. An architectural component called squeeze and excitation (SE) block [67] was introduced, which can be seamlessly integrated within any CNN model. The SE block factors out the spatial dependency by global average pooling to learn a channel-specific descriptor, which is used to recalibrate the feature map to emphasize useful channels. Further, we incorporate SE block into residual learning, obtaining a new structure that we call SERes block. Specifically, 3D SERes-U-Net is a U-Net-like architecture including an encoder and a decoder with four skip connection paths. The encoder and decoder of 3D SERes-U-Net contain SERes blocks for learning high-level semantic features and model the long-range dependencies among different channels of the learned feature maps. The encoder and decoder are connected by the skip connections for feature concatenation.

Therefore, the objectives of this research can be summarized as below:

1. To introduce a SERes building block that uses adaptive feature recalibration ability of squeeze-and-excitation operations boosts the network's representational power while feature reuse utilizes effective learning of the features, which improves segmentation performance.

2. To propose a new 3D encoder-decoder based architecture that efficiently incorporates SERes blocks into 3D U-Net-like architecture named 3D SERes-U-Net.

3. To compare the performance and results obtained from the proposed method with existing methods.

Hereby, we present a new 3D encoder-decoder-based architecture that efficiently incorporates SERes blocks into 3D U-Net-like architecture and we name it 3D SERes-U-Net. We intend to optimize training performance, efficiency and final segmentation result accuracy for the LV, RV, and Myo segmentation tasks.

5.2 Methodology

This section presents the proposed method for LV, RV and Myo heart segmentation from Cine MRI images. We present a theoretical background of the proposed SERes units and give an overall design of the proposed 3D SERes-U-Net architecture. We give dataset description, implementation and training details, present conducted experiments, and obtained results. Finally, we provide some concluding remarks.

5.2.1 Sqeeze and Excitation

The SERes block takes the advantages of the squeeze and excitation operations [66] for adaptive feature recalibration and residual learning for feature reuse [59]. The 3D SERes block can be expressed with the following expression:

$$\mathbf{X}^{res} = F_{res}(\mathbf{X}) \tag{5.1}$$

where $\mathbf{X}$ refers to the input feature, $\mathbf{X}^{res}$ is the residual feature, and $F_{res}(\mathbf{X})$ is residual mapping that needs to be learned.

$$p_n = F_{sq}(\mathbf{x}_n^{res}) = \frac{1}{L \times H \times W} \sum_{i=1}^{L} \sum_{j=1}^{H} \sum_{k=1}^{W} x_n^{res}(i,j,k) \tag{5.2}$$

where $\mathbf{p} = [p_1, p_2, ..., p_n]$ and p_n is the $n-th$ element of $\mathbf{p} \in R^n$, F_{sq} refers to the squeeze function which groups global spatial information into channel-wise statistics using global average pooling, $L \times H \times W$ is the spatial dimension of $\mathbf{F}^{res}$, $x_n^{res} \in R^{L \times H \times W}$ represents the feature map of the $n-th$ channel from the feature $\mathbf{X}^{res}$ and N is the number of channels of the residual mapping. Scale values for the residual feature channels $\mathbf{s} \in R^N$ can be expressed with:

$$\mathbf{s} = F_{ex}(\mathbf{p},\mathbf{W}) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{p})) \tag{5.3}$$

where F_{ex} is the excitation function which generates them. It is parameterized by two fully connected layers with parameters $\mathbf{W}_1 \in R^{\frac{N}{r} \times N}$ and $\mathbf{W}_2 \in R^{N \times \frac{N}{r}}$, the $ReLU$ function δ and the sigmoid function σ, and reduction ration r. The channel-wise multiplication between feature map and learned scale value s_n can be expressed with:

$$\widetilde{\mathbf{X}}_n^{res} = F_{scale}(\mathbf{X}_n^{res}, s_n) = s_n \cdot \mathbf{X}_n^{res}, \in R^{H \times W \times L} \tag{5.4}$$

Finally, applying the squeeze and excitation operations obtains the calibrated residual feature, which can be expressed with:

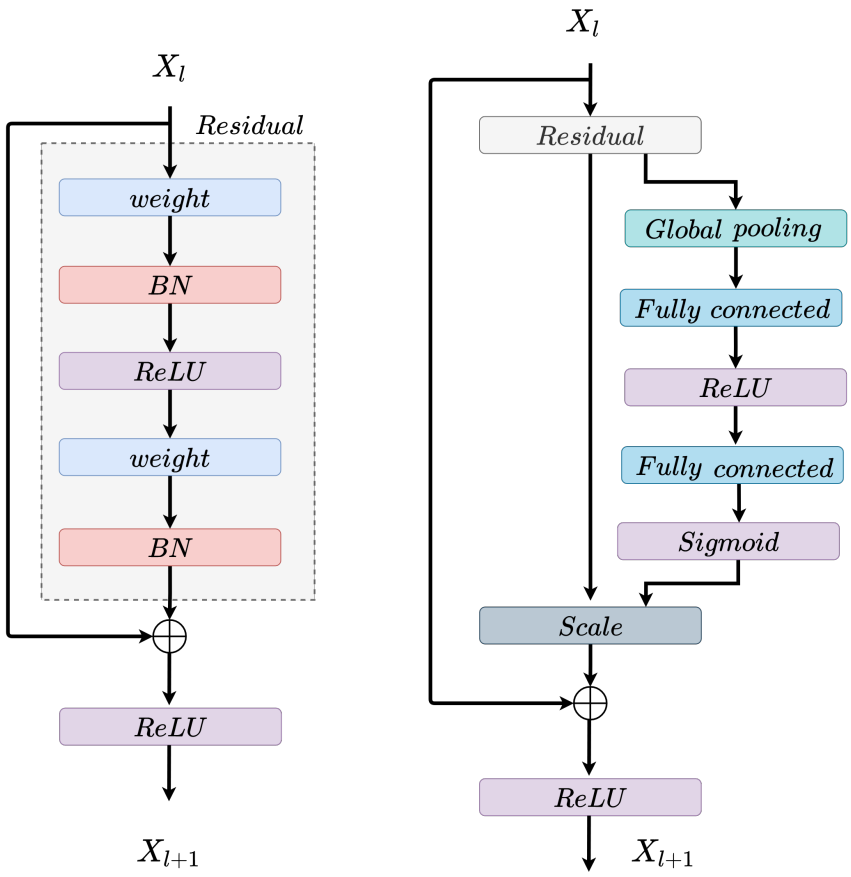
$$\widetilde{\mathbf{X}}^{res} = [\widetilde{\mathbf{X}}_1^{res}, \widetilde{\mathbf{X}}_2^{res}, ..., \widetilde{\mathbf{X}}_n^{res}] \tag{5.5}$$

The output feature $\mathbf{Y}$ after the $ReLU$ function δ is obtained as:

$$\mathbf{Y} = \delta(\widetilde{\mathbf{X}}^{res} + \mathbf{X}) \tag{5.6}$$

where $(\widetilde{\mathbf{X}}^{res} + \mathbf{X})$ refers to element-wise addition and a shortcut connection.

An illustration of the 3D ResNet block and 3D SERes block is shown in Figure 5.1.



(a) Basic residual unit (b) SERes unit

Figure 5.1: An illustration used residual blocks. (a) The original 3D ResNet block and (b) structure of the 3D SERes block.

5.2.2 Architecture Overview

Our proposed network architecture is based on the standard 3D U-Net [186] which follows encoder-decoder architecture. The encoder or contracting pathway encodes the input image and learns low-level features, while the decoder or expanding pathway learns high-level features and gradually recovers original image resolution. Like 3D U-Net, our contracting pathway consist of three downsampling layers. We replace initially used pooling layers in the original 3D U-Net with convolutional layers with stride equal to 2. Instead of plain units, we adopt SERes blocks consisting of squeeze and excitation operations followed by a residual block to accelerate convergence and training. Each residual unit inside the SERes block consists of two convolutional layers followed by batch normalization and ReLU activation. Similarly,

three SERes blocks are used in the expanding path. This pathway has three up-sampling layers, each of which doubles the size of the feature maps. Moreover, a $2 \times 2 \times 2$ convolutional layer is adopted after each upsampling layer. The network can acquire the importance degree of each residual feature channel through the feature recalibration strategy. This enhances useful channel features according to the importance degree and suppresses less useful ones. Therefore, by modeling the interdependencies between channels, the 3D SERes block performs dynamic recalibration of residual feature responses in a channel-wise manner. In this way, the network can capture every residual feature channel's importance degree, which improves its representational power. An overall network SERes-U-Net architecture is presented in Figure 5.2.

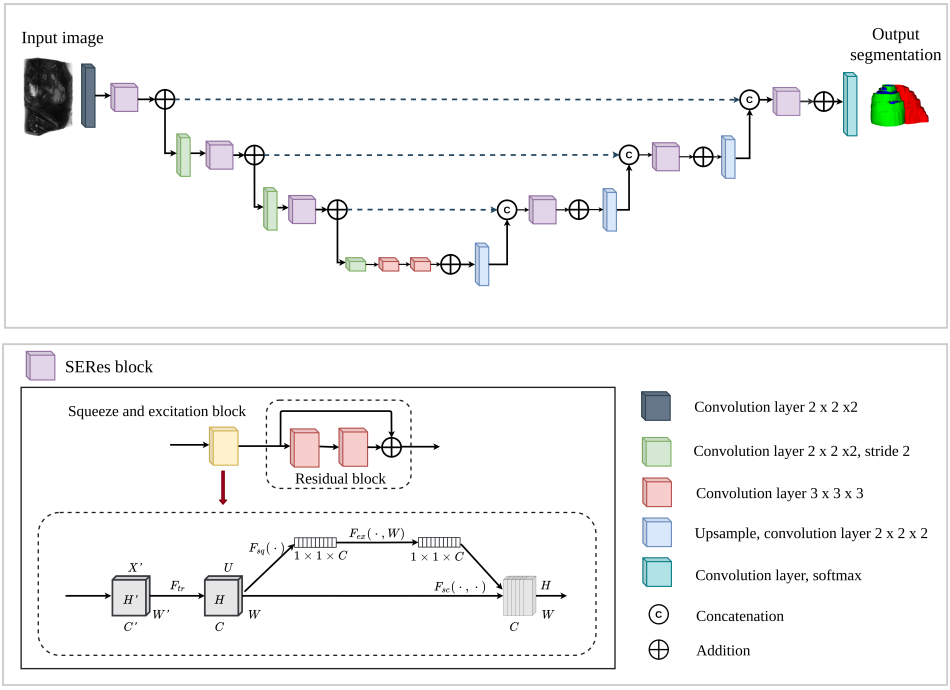


Figure 5.2: Illustration of SERes-U-Net architecture for LV, RV, Myo segmentation. Image source: Habijan et al. [55].

5.3 Implementation Details

In this section, we give a dataset description on which we conducted our experiments. After that, we give details about network training and implementation. We train two different networks to provide a successful ablation study: 3D U-Net and proposed 3D SERes-U-Net. We evaluate the proposed method using Automated Cardiac Diagnosis Challenge (ACDC) dataset [1] and present conducted experiments and results. Finally, we compare our results to the state-of-the-art research and provide concluding remarks.

5.3.1 Dataset Description

The Automated Cardiac Diagnosis Challenge (ACDC) dataset [1] consists of real-life clinical cases obtained from an everyday clinical setting at the University Hospital of Dijon (France). The dataset includes cine-MRI images (2D + time) of patients suffering from different pathologies, including myocardial infarction, hypertrophic cardiomyopathy, dilated cardiomyopathy, abnormal right ventricle, and normal cardiac anatomy. Dataset has been evenly divided based on the pathological condition and includes 100 cases with corresponding ground truth for training and 50 cases for testing through an online evaluation platform. Clinical experts manually annotated LV, RV and Myo at systolic and diastolic phases, for which the weight and height information was provided as well. Images are acquired as a series of short-axis slices covering the LV from the base to the apex. The spatial resolution goes from 1.37 to 1.68 $mm^2/pixel$, slice thickness is between 5-8 mm, while an inter-slice gap is 5 or 10 mm. An example of input images in different views with corresponding manual segmentation is shown in Figure 5.3.

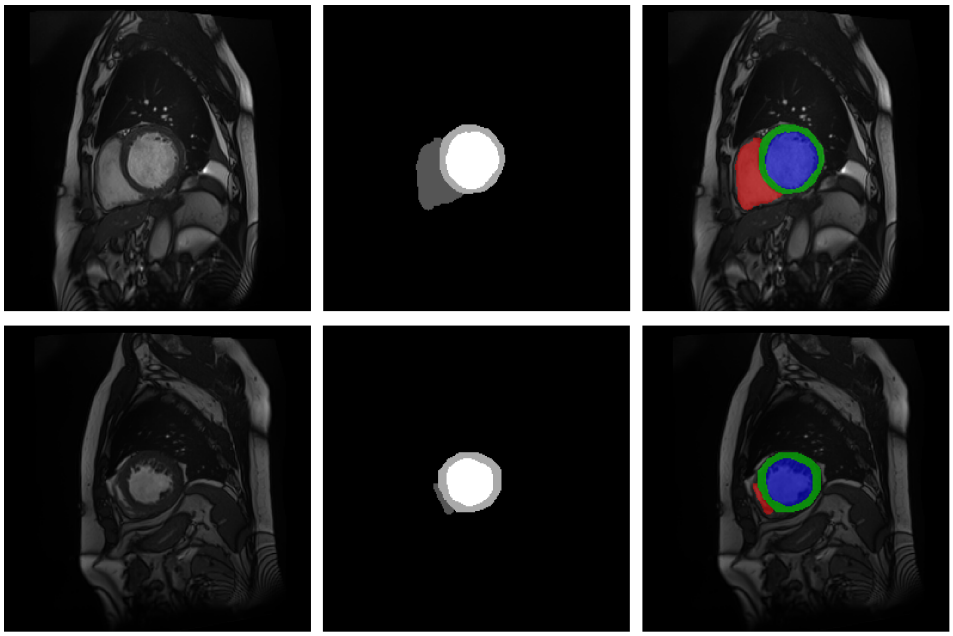


Figure 5.3: An example of the ACDC dataset. Top row (from left to right): original input image at ED, corresponding GT and input image with GT overlay. Bottom row (from left to right): original input image at ES, corresponding GT and input image with GT overlay. RV is represented in red color, Myo in green color, and LV in blue color.

5.3.2 Data Preprocessing and Augmentation

To overcome high-intensity irregularities of MRI images, we normalize each volume based on the standard and mean deviation of their

intensity values. The volumes were center-cropped to a fixed size and zero-padded to provide fine ROI for the network input. For data augmentation, we apply a random axis mirror flip with a probability of 0.5, random scale and intensity shift on input images.

5.3.3 Network Implementation and Training

We use $L2$ norm regularization with a weight of 10^{-5} and employ the spatial dropout with a rate of 0.2 after the initial encoder convolution. We use Adam optimizer with initial learning rate of $\alpha_0 = 10^{-4}$ and gradually decrease it according to following expression:

$$\alpha = \alpha_0 * \left(1 - \frac{e}{T_e}\right)^{0.9} \tag{5.7}$$

where T_e is a total number of epochs and e is an epoch counter. We employ a smoothed negative Dice score [106] loss function, defined with:

$$D_{loss} = -\frac{2\sum_{i=1}^{N} p_i g_i + 1}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + 1} \tag{5.8}$$

where p_i is probability of predicted regions, g_i is the ground truth classification for every i voxel.

We train two networks to provide a successful ablation study: 3D Res-U-Net and proposed 3D SERes-U-Net. The networks are trained separately for each of cardiac phase. For both network architectures, we use 80%-20% training and validation split, respectively, i.e., we use 80 patient images for training and 20 for validation. Final segmentation accuracy testing was done through an online ACDC Challenge submission page on 50 patient subjects [2]. The total training time took approximately 34 hours for 200 epochs since further training appears not to decrease validation loss. We used two NVIDIA Titan V100 GPU simultaneously. Moreover, Figure 5.4 and Figure 5.5 indicate decrease in loss value when number of epochs increases. This is a clear indication that the network is successfully learning from the input data. We can also see significant improvement regarding training and validation accuracies and faster and smoother convergence of the 3D SERes-U-Net network architecture. On MRI images of the ED cardiac phase, the 3D Res-U-Net model has an average efficiency of 96.18% of trained accuracy while validation error was on average 94.24%. On the other hand, the 3D SERes-U-Net obtains an average training accuracy of 99.35%, while validation accuracy is 89.71%. A clear improvement in training is obtained with the addition of the SE block, i.e., using the proposed 3D SERes-U-Net architecture. Comprehensive training on MRI images at the ES phase has lower accuracy in comparison to the ED phase. For example, 3D Res-U-Net yields an average training accuracy of 88.07% and validation accuracy of 85.13%. The inclusion

of SE blocks in the proposed 3D SERes-U-Net architecture improves training and validation accuracies for 4.69% and 3.57%, respectively.

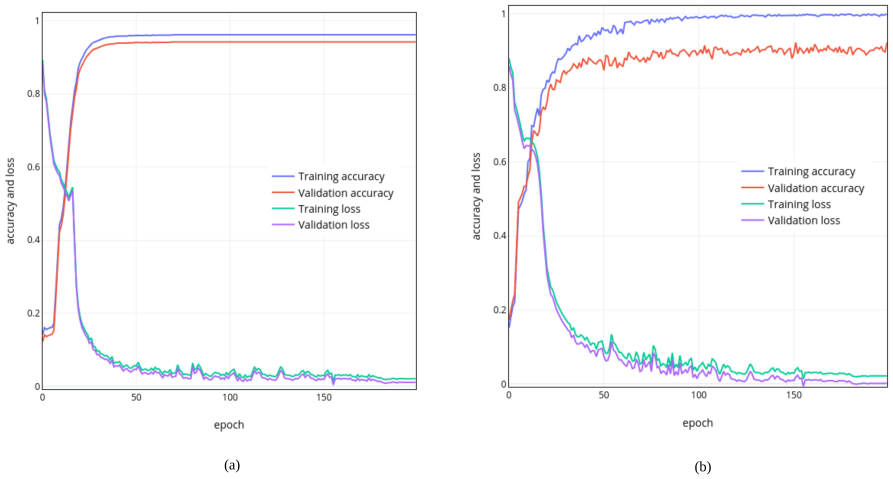


Figure 5.4: Training and validation accuracies on Cine MRI dataset at ED cardiac phase. (a) 3D Res-U-Net network architecture and (b) 3D SERes-U-Net network architecture.

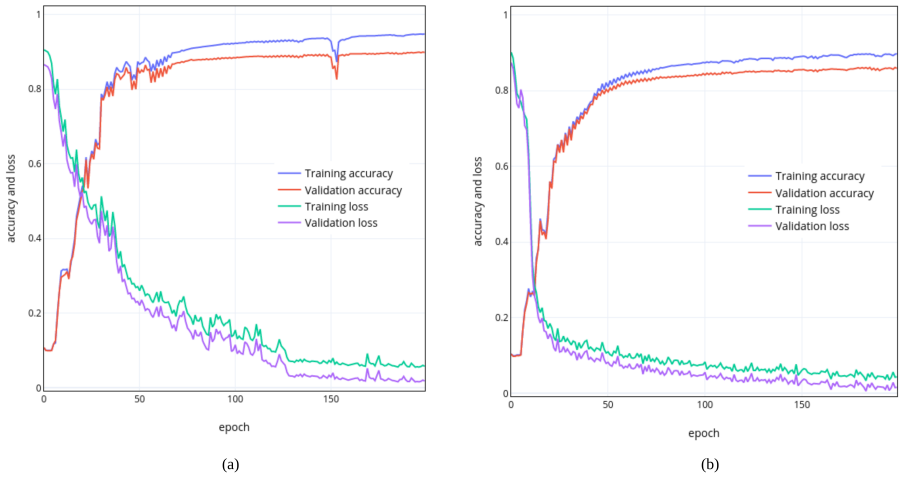


Figure 5.5: Training and validation accuracies on Cine MRI dataset at ES cardiac phase. (a) 3D Res-U-Net network architecture and (b) 3D SERes-U-Net network architecture.

5.4 Experiments and Results

To evaluate the segmentation performance of the proposed method, we observe distance and clinical indices metrics. Distance measures include calculation of DSC and HD, which provides information on similarity between obtained segmentations for LV, RV and Myo with their reference ground truth. Based on these results, it is shown

that the inclusion of SE blocks into the proposed 3D SERes-U-Net architecture yields slightly better results than plain 3D Res-U-Net architecture. Detailed qualitative segmentation results are presented in Table 5.1 and Table 5.2.

Table 5.1: The segmentation accuracy results for LV, RV and Myo expressed in Dice score (DSC) and Hausdorff distance (HD) for the proposed method at ED for 3D Res-U-Net and proposed 3D SERes-U-Net.

Methods	ED					
	LV		RV		Myo	
	D_{sc}	H_d	D_{sc}	H_d	D_{sc}	H_d
3D Res-U-Net	0.93 (0.063)	38.2 (4.872)	0.86 (0.091)	52.9 (12.441)	0.8 (0.063)	32.95 (5.600)
3D SERes-U-Net	0.95 (0.007)	11.53 (0.410)	0.9 (0.021)	23.41 (12.357)	0.83 (0.007)	13.77 (1.987)

Table 5.2: The segmentation accuracy results for LV, RV and Myo expressed in Dice score (DSC) and Hausdorff distance (HD) for the proposed method at ES for 3D Res-U-Net and proposed 3D SERes-U-Net.

Methods	ES					
	LV		RV		Myo	
	D_{sc}	H_d	D_{sc}	H_d	D_{sc}	H_d
3D Res-U-Net	0.86 (0.028)	29.77 (1.774)	0.77 (0.042)	36.99 (5.395)	0.81 (0.028)	30.29 (1.103)
3D SERes-U-Net	0.86 (0.127)	11.94 (8.499)	0.83 (0.028)	21.49 (5.755)	0.85 (0.007)	15.00 (1.979)

Our proposed 3D SERes-U-Net obtains an average DSC for LV, RV and Myo at end-diastole of 95%, 90%, 83%, respectively. Similarly, we obtain an average DSC for LV, RV and Myo at end-systole of 86%, 83%, 85%, respectively. The 3D Res-U-Net network achieves an average DSC for LV, RV and Myo at end-diastole of 93%, 86%, 80%, respectively. The addition of squeeze and excitation operations, i.e., use of proposed SERes blocks, improves DSC and HD for 2%, 4% and 3%, respectively. Similarly, the 3D Res-U-Net network achieves an average DSC for LV, RV and Myo at end-systole of 86%, 77, 81, respectively. The addition of squeeze and excitation operations, i.e.,

use of proposed SERes blocks, improves DSC for 0.2%, 6% and 4%, respectively. Therefore, obtained results using the proposed 3D SERes-U-Net shows significant improvements in DSC in comparison to the network without squeeze and excitation operations (3D Res-U-Net). Boxplots showing the distribution of the DSC for LV, RV and Myo in ES and ED cardiac cycle phases are presented in Fig 5.6, while Figure 5.7 and Figure 5.8 shows visual examples of obtained segmentation predictions.

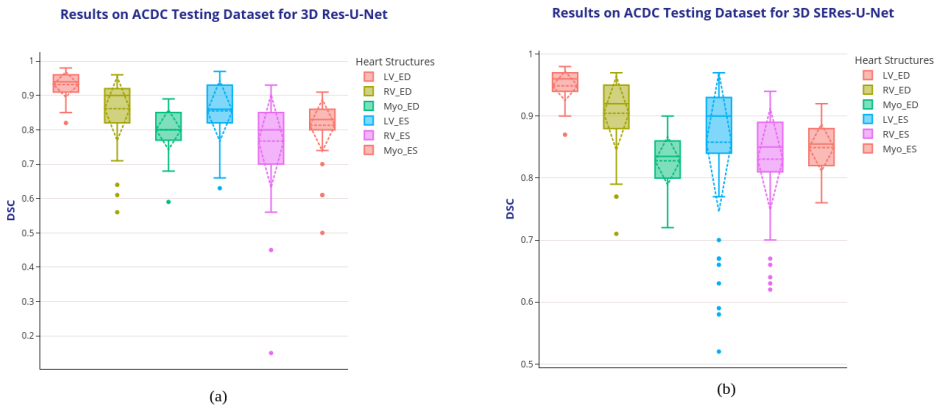


Figure 5.6: Boxplots showing the DSC dispersion for LV, RV and Myo using (a) 3D Res-U-Net segmentation network and proposed (b) 3D SERes-U-Net on the ACDC testing dataset. Boxplot illustrates interquartile range (bounds of box), mean (X inside a box), median (centerline), maximum and minimum values (whiskers) and outliers (circles outside whiskers). Image source: Habijan et al. [55].

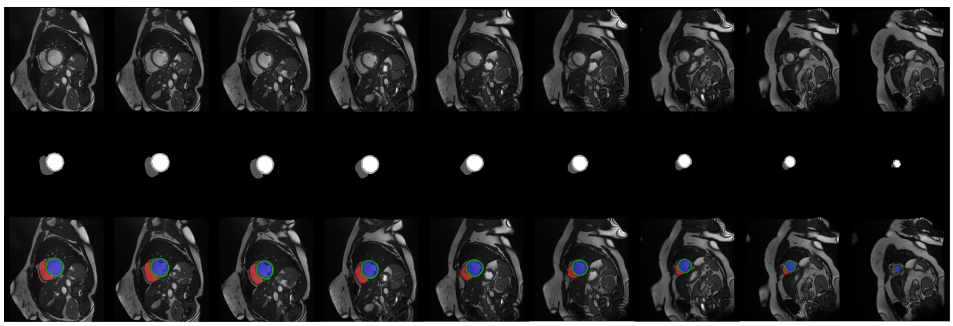


Figure 5.7: An example of obtained results. Top row: an original MRI image at the end-diastolic phase of the cardiac cycle. Middle row: Obtained segmentation. Bottom row: an overlay of the original image and obtained segmentation prediction. Image source: Habijan et al. [55].

Next, we observe and discuss the best and worst segmentation results. Figure 5.9 shows an example of the most successful segmentation that yields DSC of 98%, 93%, 87% for LV, RV and Myo at ED phase, respectively. The results of the same patient image in the ES phase

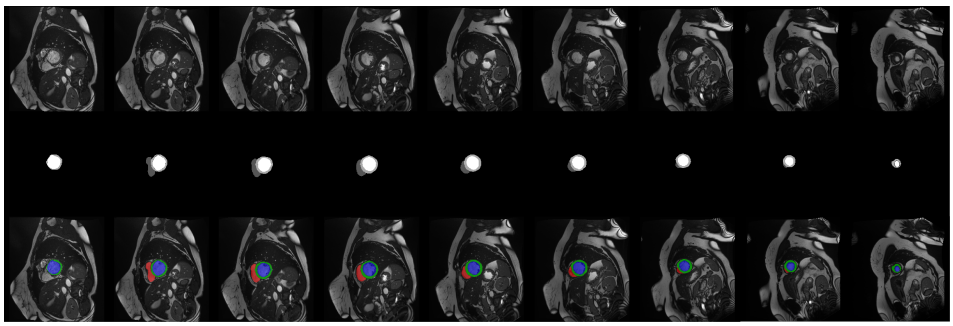


Figure 5.8: An example of obtained results. Top row: original MRI image at the end-systolic phase of the cardiac cycle. Middle row: Obtained segmentation. Bottom row: an overlay of the original image and obtained segmentation prediction. Image source: Habijan et al. [55].

are somewhat lower; DSC is 97%, 91%, and 86% for LV, RV and Myo, respectively. As DSC, segmentation of the LV in the ED phase commonly has high accuracy for different patient images compared to the ES phase. This is due to the high contrast in the ED images, where structures can be more easily distinguished even with the naked eye.

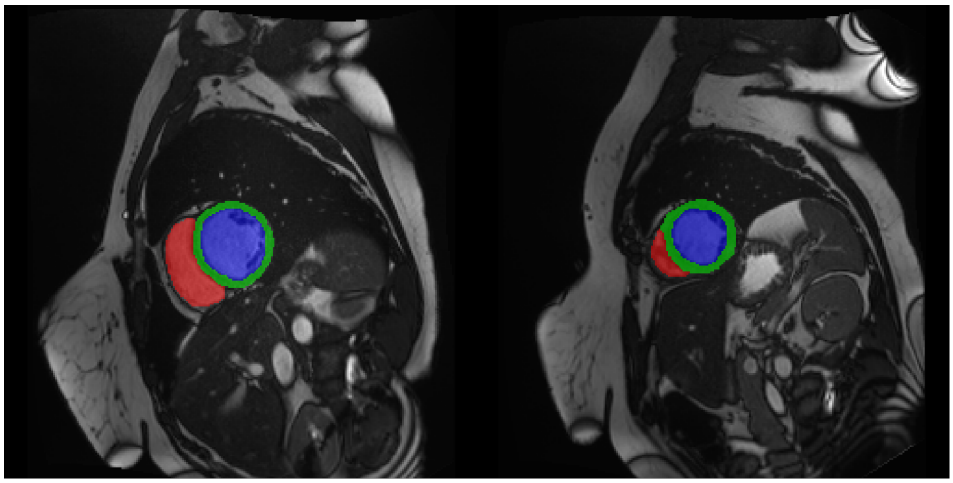


Figure 5.9: An example of most successful segmentation in ED (left) and ES (right) phases.

Another common problem was segmentation failure, mostly for Myo and RV in both phases of the cardiac cycle. For example, we noticed that our model has difficulties in correctly segmenting RV and Myo as shown in Figure 5.10. This may be partially explained by the fact that an accurate myocardium segmentation requires the precise delineation of two walls instead of one, which was the case for the LV and RV.

The overfitting issue is successfully overcomed in the most cases. Still, we observe an overfitting in basal part of RV where model hardly distinguishes between RV and RA. This failure is characteristic only

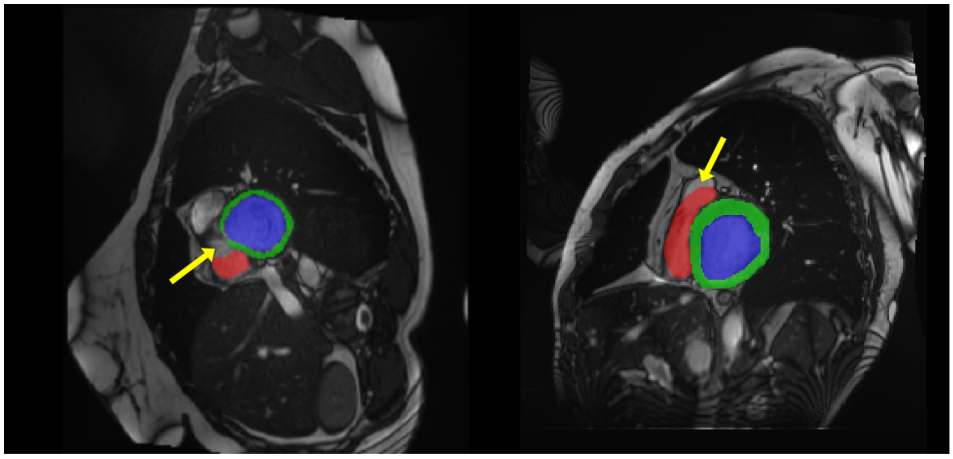


Figure 5.10: An example of RV and Myo segmentation failure in ED and ES phases.

in ES phase, which resulted in significantly lower DSC in comparison to ED phase. An example of such failure is shown in Figure 5.11.

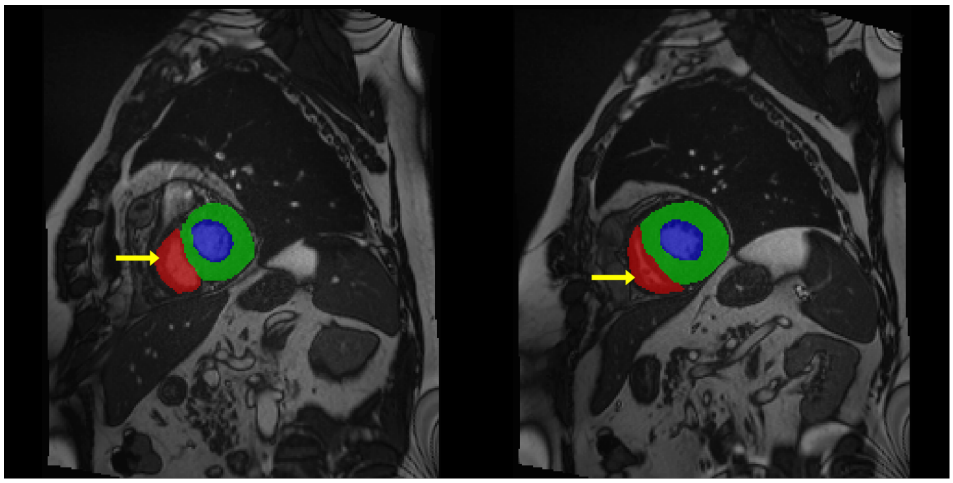


Figure 5.11: Comparison of the automatically obtained segmentations and the reference volumes of the MRI scans. The image shows correlation and Bland-Altman plots for the LV volumes at and diastole and at the end-systole as well as ejection fraction.

Figure 5.12 shows 3D visualization of the best and the worse segmentation cases on the MRI test dataset obtained using the proposed 3D SERes-U-Net architecture. From visual inspection, we may see that failures in worst segmentation cases are primarily due to over-segmented Myo and missing parts of the RV.

After we successfully obtained segmentations, we calculated significant clinical metrics. These metrics are significant indicators of hearts' function and include the volume of the left ventricle at end-diastole (LVEDV), the volume of the left ventricle at end-systole (LVESV), left ventricles' ejection fraction (LVEF), the volume of the right ventricle at end-diastole (RVEDV), volume of the right ventricle at end-systole

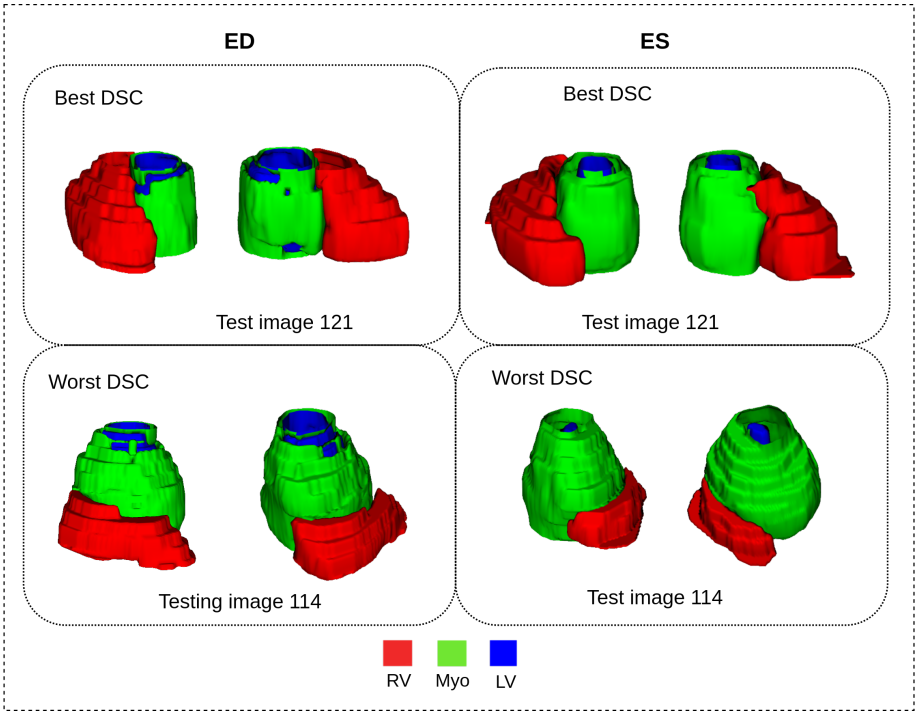


Figure 5.12: 3D visualization of the best and worse cases for LV, RV and Myo at ED and ES in different rotation views.

(RVESV), right ventricles' ejection fraction (RVEF), myocardium volume at end-systole (MyoLVES), and myocardium mass at end-diastole (MyoMED). The Bland-Altman and analysis show a high correlation coefficient of R=0.99 for LVEDV and LVESV, while R=0.95 for LVEF. Correlations of RVEDV, EVESV and RVEF are R=0.97, R=0.93, R=0.69, respectively. Finally, R=0.96 for MyoLVES and R=0.95 for MyoMED further show the strength of accuracy and precision of our proposed pipeline. Table 5.3 gives detailed representation of obtained results. For more convenient representation, in Figure 5.13 we show correlation and Bland-Altman plots for the LV volumes at ED and at ES cardiac phases. Figure 5.14 shows correlation and Bland-Altman plots for the RV volumes at ED and at ES cardiac phase as well as EF. Figure 5.15 shows correlation and Bland-Altman plots for the Myo volumes at ED and at ES cardiac phase as well as EF.

5.4.1 Comparison with Other Methods

The proposed approach was compared with other similar deep learning approaches in terms of image segmentation accuracy, as shown in Table 5.4 and Table 5.5.

Most of the previous work includes modifications and experiments using 2D or 3D U-Net architecture. For example, Isensee et al. [74] implemented an ensemble of 2D and 3D U-Net architectures. Baumgartner et al. [12] tested influence of using different hyperparameters on the U-Net and the FCN for this particular application. They also

Table 5.3: Calculated clinical indexes. R is correlation coefficient, while *mae* is mean absolute error.

Network	3D Res-U-Net			3D SERes-U-Net		
Clinical index	R	bias + σ	mae	R	bias + σ	mae
LVEDV	0.985	-19.66 + 28.70	4.52	0.990	10.44 + 17.04	1.25
LVESV	0.980	-30.92 + 30.98	0.03	0.989	-18.20 + 20.70	-4.91
LVEF	0.938	-13.51 + 16.69	1.59	0.949	-28.01 + 18.19	3.30
RVEDV	0.819	-60.88 + 92.32	15.72	0.966	-16.12 + 28.50	-6.0
RVESV	0.829	54.07 + 59.19	2.56	0.925	-40.55 + 28.41	-13.16
RVEF	0.627	-22.10 + 30.20	4.05	0.721	52.53 + 26.21	6.19
MyoMED	0.945	-17.54 + 60.14	21.30	0.956	-20.84 + 46.62	12.89
MyoLVES	0.909	-29.05 + 65.49	18.22	0.963	-26.96 + 34.16	3.60

explore the impact of using 2D and 3D convolution layers and a training Dice loss versus a cross-entropy loss. Their best architecture ended up being a U-Net with 2D convolution layers trained with a cross-entropy loss. Jang et al. [78] use M-Net, which is a modification of U-Net, which has the feature maps of the decoding layers which are concatenated with those of the previous layer. Khened et al. [86] use a dense U-Net. Their method begins by locating the region of interest on the first harmonic image using a Fourier transform followed by a Canny edge detector. They next compute the approximate radius and center of the LV using a circular Hough transform on the previously obtained edge map. They then replace the convolutional blocks in a U-Net with dense blocks to make the system lighter. This network's initial layer also corresponds to an inception layer. The network was trained using a weighted average of dice and cross-entropy losses. Zotti et al. [185] modified U-Net by using convolutional layers along the skip connections to create Grid Net. Additionally, the architecture records a shape prior to completing the final decision, which is employed as an additional features map. Wolterink et al. [173] instead, the encoder-decoder architecture uses a sequence of convolutional layers with increasing levels of kernel dilation. This ensures that sufficient image context was used for each pixel's label prediction. This CNN was fed simultaneously with spatially corresponding ED and ES 2D slices while the output of the network was split in two, one softmax for ED and one for ES. The only exception to using a U-Net-like network is Tziritas and Grinias [71], which implemented a Chan-Vese level-set method followed by an MRF graph cut segmentation method and spline fitting to smooth out the resulting boundaries. Experimental results show that the proposed 3D SERes-U-Net has better performance for RV, LV-Myo and LV segmentation than other current state-of-the-art models.

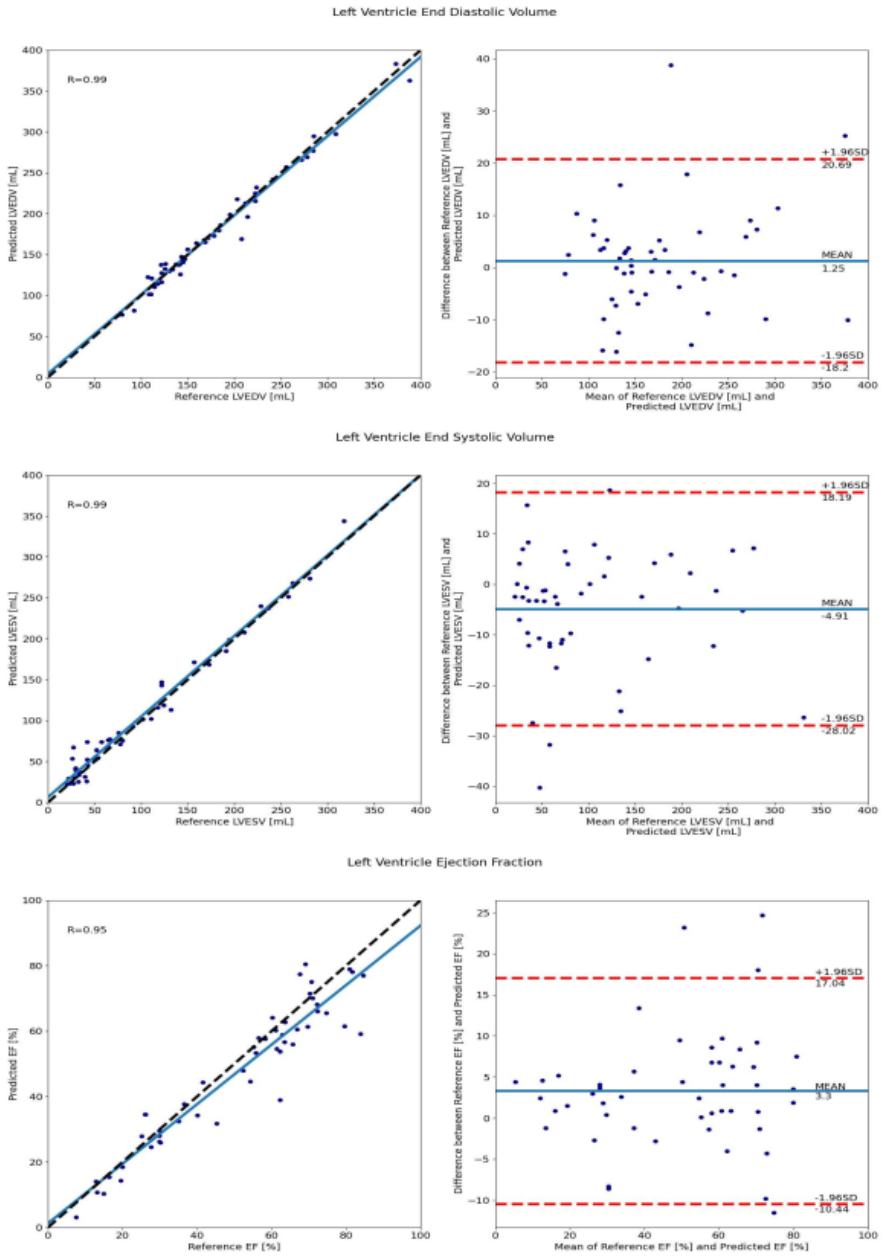


Figure 5.13: Comparison of the automatically obtained segmentations and the reference volumes of the MRI scans. The image shows correlation and Bland-Altman plots for the LV volumes at end-diastole and at the end-systole as well as ejection fraction. Image source: Habijan et al. [55].

5.5 Conclusion

In this chapter, we presented a new automatic method for LV, RV and Myo segmentation from Cine MRI images. For this purpose, we used the ACDC challenge dataset. As a result of the different breath-holds used during the acquisition of the dataset, it contains a large image shift and slice thickness. Therefore, we assumed that considering 3D

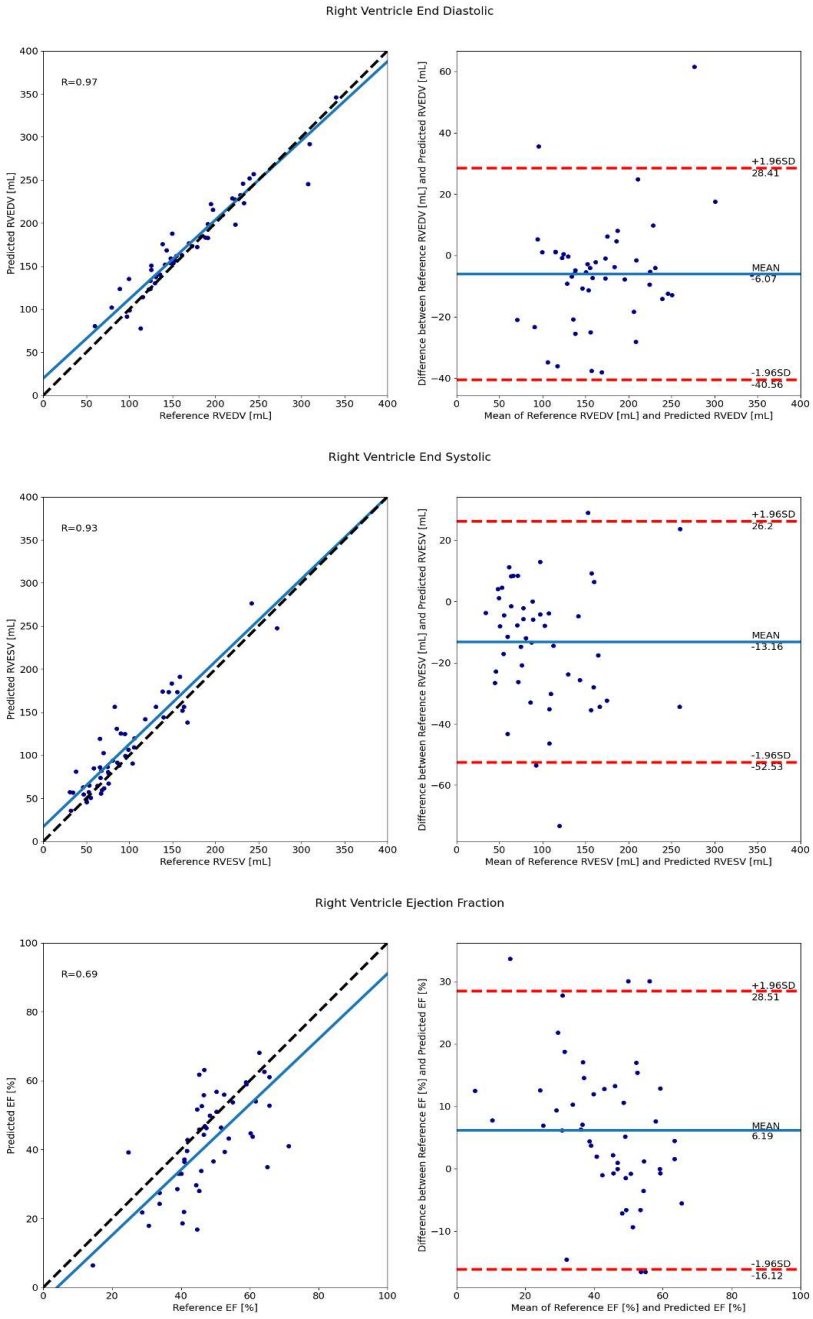


Figure 5.14: Comparison of the automatically obtained segmentations and the reference volumes of the MRI scans. The image shows correlation and Bland-Altman plots for the RV volumes at end-diastole and at the end-systole as well as ejection fraction. Image source: Habijan et al. [55].

information can impair model generalization. This was the main reason for choosing to develop the 3D method over the 2D method.

In this research, we introduced a new deep neural network architecture named 3D SERes-U-Net for automatic segmentation of LV, RV and Myo from Cine MRI images. The 3D SERes-U-Net incorporates SERes blocks into 3D U-net architecture. The SERes blocks use

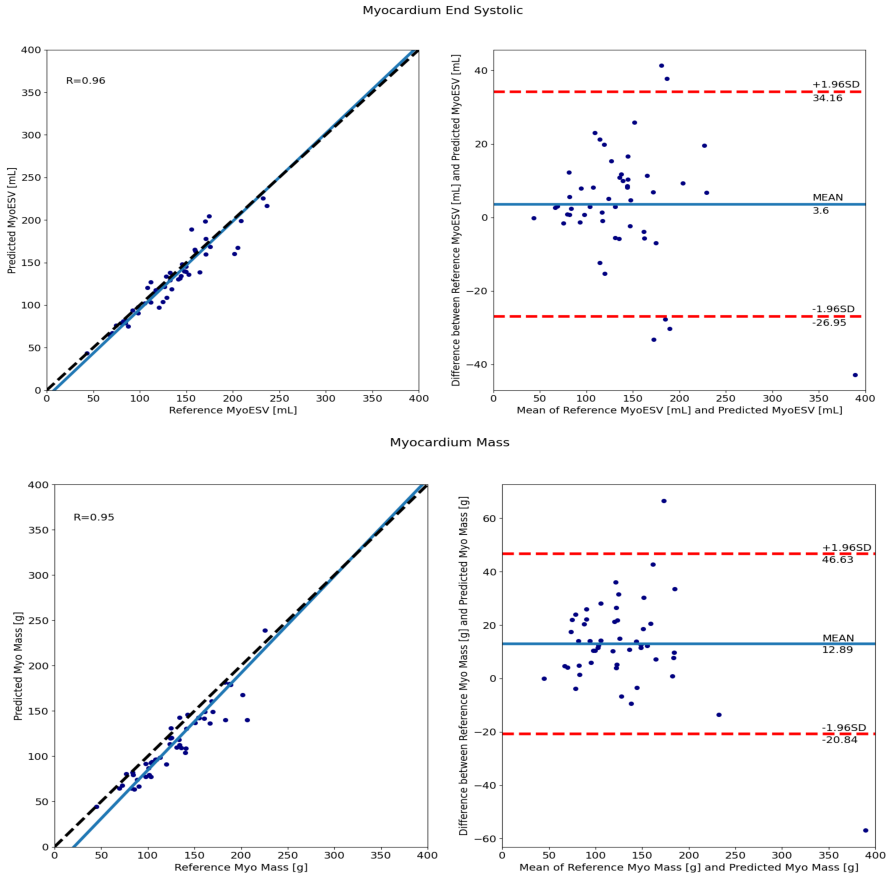


Figure 5.15: Comparison of the automatically obtained segmentations and the reference volume of the myocardium end systolic volume and myocardium mass. The image shows correlation and Bland-Altman plots to compare automatically obtained segmentation and the reference values. Image source: Habijan et al. [55].

Table 5.4: Comparison of the segmentation accuracy of the proposed method and the state-of-the-art methods at ED cardiac phase. LV: Endocardial contour of the left ventricle; RV: Endocardial contour of the right ventricle;Myo: Epicardial contour of the left ventricle (myocardium); DSC: Dice Index; HD: Hausdorff distance.

	LV		RV		Myo	
Authors	DSC	HD	DSC	HD	DSC	HD
Isense et al. [74]	0.968	7.4	0.946	10.1	0.902	8.7
Baumgartner et al. [12]	0.963	6.5	0.932	12.7	0.892	8.7
Jang et al. [78]	0.959	7.7	0.929	12.9	0.875	9.9
Zotti et al. [185]	0.957	6.6	0.941	10.3	0.884	8.7
Khened et al. [86]	0.964	8.1	0.935	14.0	0.889	9.8
Wolternik et al. [173]	0.961	7.5	0.928	11.9	0.875	11.1
Tziritas-Grinias et al. [71]	0.948	8.9	0.863	21.0	0.794	12.6
Proposed	0.95	11.53	0.90	23.41	0.83	13.77

Table 5.5: Comparison of the segmentation accuracy of the proposed method and the state-of-the-art methods at ES cardiac phase.

	LV		RV		Myo	
Authors	DSC	HD	DSC	HD	DSC	HD
Isense et al. [74]	0.931	6.9	0.899	12.2	0.919	8.7
Baumgartner et al. [12]	0.911	9.2	0.883	14.7	0.901	10.6
Jang et al. [78]	0.921	7.1	0.885	11.8	0.895	8.9
Zotti et al. [185]	0.905	8.7	0.882	14.1	0.896	9.3
Khened et al. [86]	0.917	9.0	0.879	13.9	0.898	12.6
Wolternik et al. [173]	0.918	9.6	0.872	13.4	0.894	10.7
Tziritas-Grinias et al. [71]	0.865	11.6	0.743	25.7	0.801	14.8
Proposed	0.86	11.94	0.83	21.49	0.85	15.00

squeeze-and-excitation operations together with residual learning. The adaptive feature recalibration ability of squeeze-and-excitation operations boosts the network's representational power while feature reuse utilizes effective learning of the features, which improves segmentation performance. We evaluate the proposed method on the Automated Cardiac Diagnosis Challenge (ACDC) testing dataset. Our pipeline obtains an average DSC for LV, RV and Myo at end-diastole of 95%, 90%, 83%, respectively. Similarly, we obtain an average DSC for LV, RV, and Myo at end-systole of 86%, 83%, 85%, respectively. We calculate significant clinical metrics, i.e., indicators of hearts function, including LVEDV, LVESV, LVEF, RVEDV, RVESV, RVEF, MyoLVES, and MyoMED. The Bland-Altman and analysis show a high correlation coefficient of R=0.99 for LVEDV and LVESV, while R=0.95 for LVEF. Correlations of RVEDV, EVESV and RVEF are R=0.97, R=0.93, R=0.69, respectively. Finally, R=0.96 for MyoLVES and R=0.95 for MyoMED further show the strength of accuracy and precision of our proposed method. Our proposed 3D SERes-U-Net obtains competitive results for all three structures. The results on the LV segmentation are highly comparative. Nevertheless, it appears that the same level of accuracy is still challenging to obtain for the RV and the MYO. The RV often has the highest Hausdorff distances, the lowest Dice scores, the lowest correlation values and the largest biases. Furthermore, we have seen that most segmentation failures and errors appeared in RV and Myo due to overfitting problems. We have noticed that Myo segmentations, particularly at ES, vary the most. This may partly be explained because correct myocardium segmentation necessitates the precise delineation of two walls rather than just one for the LV and RV.

CHAPTER 6

Abdominal Aortic Aneurysms Segmentation

This chapter presents a new automatic approach for robust and reproducible abdominal aortic aneurysm (AAA) segmentation. The 3D U-Net segmentation network is adapted by introducing residual units in an encoder part and a deep supervision mechanism in the decoder part. We train four different architectures from stretch: (1) 3D U-Net network architecture, (2) 3D U-Net with residual blocks in encoder pathway (3D U-Net RE), (3) 3D U-Net with deep supervision (3D U-Net DS) and (4) 3D U-Net with residual blocks in an encoder pathway and deep supervision in decoder pathway (3D U-Net RE + DS). In this way, we demonstrate the effect of residual units and deep supervision for this particular clinical application. Networks are trained, validated, and evaluated on 19 pre-operative CTA volumes from different patients using a 4-fold cross-validation approach to increase the results' robustness. Our pipeline achieves a Dice score of 91.03% for AAA segmentation.

The outline of the chapter is structured in the following manner. Section 6.1 gives the main objectives of conducted research. Section 6.2 gives a theoretical background of the used methods and describes our proposed method for AAA segmentation. Section 6.4 describes the experimental setup, gives network training details and presents obtained results. Finally, concluding remarks are provided in Sections 6.5.

6.1 Objectives

This research aims to develop an efficient method for the fully automatic segmentation of the AAA regions in CTA images. Generally, deep learning approaches require a large amount of data for a good generalization. Nevertheless, a limited annotated dataset complicates training and testing. Thus, we take advantage of the feature reuse mechanism to alleviate these obstacles. The proposed network relies on a modified 3D U-Net architecture. We introduce residual connections in the encoder pathway and deep supervision in the decoder pathway. The proposed method reduces the time of the segmentation process compared to the conventional manual method. It is comparable or equivalent in terms of performance and accuracy.

Therefore, the main objectives can be summarized and listed as below:

1. To develop an automatic method for detecting and identifying AAA regions from CTA images using deep learning methodology.

2. To alleviate common obstacles of constructing very deep neural network architectures using feature reuse mechanism and by combining multiple segmentation maps created at different scales.

3. To compare the performance and result obtained from the proposed method with existing methods.

Hereby, we present a new 3D U-Net with residual blocks in an encoder pathway and deep supervision in decoder pathway and we name it 3D U-Net (RE + DS). We intend to optimize training performance, efficiency and final segmentation result accuracy for the task of AAA.

6.2 Architecture Overview

Motivated by the high success of 3D U-Net, we propose a modification to 3D U-Net architecture by adding residual units in the contracting pathway and deep supervision in expanding the pathway for the task of an abdominal aortic aneurysm segmentation. The addition of residual units in the contracting pathway preserves information. It significantly increases network performance, while the addition of deep supervision in expanding pathway injects gradient signals deep into the network.

The proposed architecture has encoding and decoding pathways. In the encoding pathway, the input representations are increasingly encoded as they advance deeper within the network. In contrast, the decoding pathway reincorporates obtained representations with superficial features resulting in precise localization of the interest structures. For simplicity purposes, all processing blocks in the encoding pathways are referred to as the encoding module. Likewise, all processing blocks in the decoding pathways are referred to as the decoding module.

Each encoding module represents a residual block consisting of two $3x3x3$ convolution layers with a dropout layer in between. They are reciprocally connected with stride 2, 3×3×3 convolutions. Similarly, the decoding pathway consists of up-scaling with stride 2, after which comes a $3x3x3$ convolution. The up-sampling module halves the number of feature maps that are further concatenated with the encoding pathway and subsequently passed to the decoding module. A decoding module consists of a 3×3×3 convolution and a 1×1×1 convolution, which again halves the number of feature maps. Additionally, we apply deep supervision in the decoding pathway by integrating segmentation layers at different network levels. These segmentation layers are further combined with element-wise summation and form the final network output, i.e., AAA segmentations. Auxiliary supervision paths were added so that intermediate feature maps could get supervision to restore details further and improve segmentation accuracy. An illustration of the network is shown in Figure 6.1.

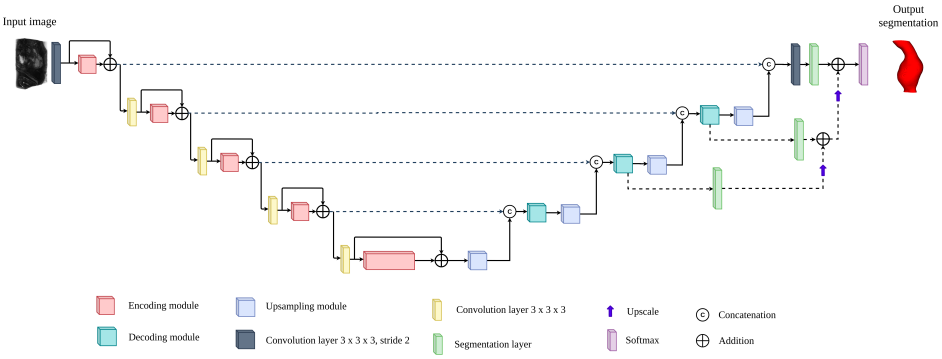


Figure 6.1: Illustration of 3D U-Net (RE + DS) architecture for AAA segmentation.

The deep supervision approach forces the output from the decoder units to yield meaningful segmentation maps. This technique is mainly introduced for obtaining transparency and robustness of the features extracted in the middle of the network and helps to address the vanishing gradient problem. It allows gradient information to flow back directly from the loss to every decoder block. The feature maps from each network level are transposed by $1 \times 1 \times 1$ convolutions to create secondary segmentation maps. These are then combined in the following way. First, the segmentation map with the lowest resolution is upsampled with bilinear interpolation to have the same size as the second-lowest resolution segmentation map. The element-wise sum of the two maps is then upsampled and added to the third-lowes segmentation map and so on until we reach the highest resolution level. These additional segmentation maps do not primarily serve for any further refinement of the final segmentation map created at the last layer of the model because the context information is already provided by long skip connections. The secondary segmentation maps help in the speed of convergence by encouraging earlier layers of the network

to produce better segmentation results. A similar principle has been used by Kayalibay et al. [84] and Isensee et al. [75].

6.3 Implementation Details

In this section, we give a dataset description on which we conducted our experiments. After that, we give details about network training and implementation. We train four different networks to provide a successful ablation study: (1) 3D U-Net network architecture, (2) 3D U-Net with residual blocks in encoder pathway (3D U-Net RE), (3) 3D U-Net with deep supervision (3D U-Net DS) and (4) 3D U-Net with residual blocks in an encoder pathway and deep supervision in decoder pathway (3D U-Net RE + DS).

6.3.1 Dataset Description

In our experiments, we use 3D CT images of unruptured AAAs acquired as a part of standard medical procedures and treatments from patients at the University Hospitals Leuven, Belgium. The dataset is publicly available for research purposes [172]. It consists of 19 volumetric CT images in the nearly raw raster data with corresponding AAA ground truth (GT) segmentations. GTs were segmented using BioPARR software and 3D Slicer. The remaining unwanted segmentation artifacts were manually corrected from the resulting label maps. Images have different dimensions, with a spacing of 0.625 mm in each direction. Additionally, the dataset also includes geometries of the ITL internal surface, AAA internal and external surface and finite element meshes of the AAA and ITL in the stereolithography (STL) format and characterization of the boundary conditions external load for finite element model and material properties. An example of one image slice along with three different views and corresponding ground truth images are presented in Figure 6.2.

6.3.2 Preprocessing and Data Augmentation

With CT intensity values being non-standardized, normalization is critical to allow for data from different institutes, scanners and acquired with varying protocols to be processed by one single algorithm. We normalize each input CT image by subtracting the mean and dividing by the standard deviation of the AAA region. Since we use a very limited training dataset, we try to alleviate the training dataset size using extensive data augmentation techniques. The following augmentation techniques were applied on the fly during training: random rotations, random scaling, random elastic deformations, gamma correction augmentation and mirroring. Nevertheless, the low resolution of input images distortion and random rotations significantly worsen final

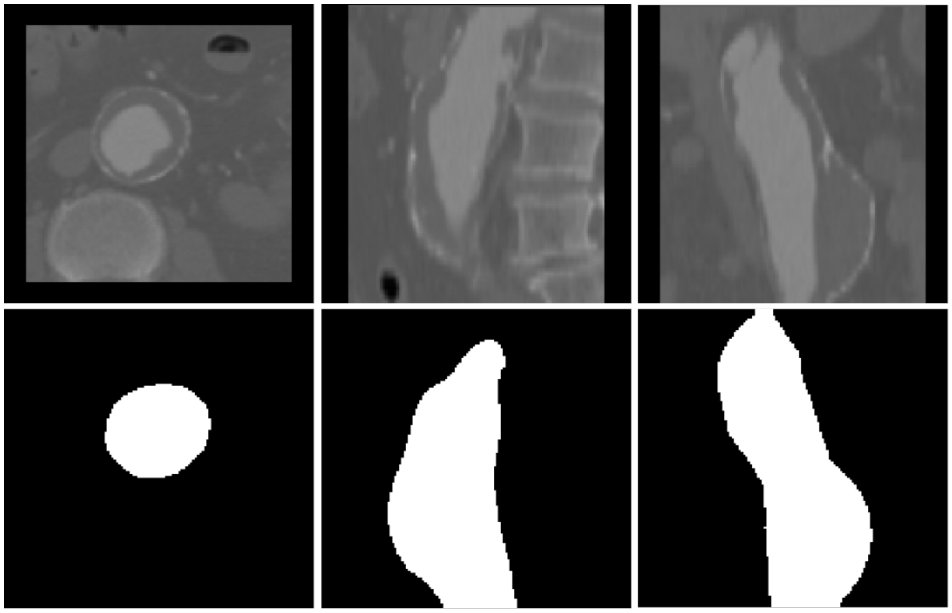


Figure 6.2: Example images from used AAA dataset. Up row, from left to right: cropped axial, coronal and sagittal image slices within the AAA ROI. Bottom row, from left to right: corresponding ground truth masks for axial, coronal and sagittal image slices.

segmentation results. Therefore, we decided to use just permutations as a data augmentation technique.

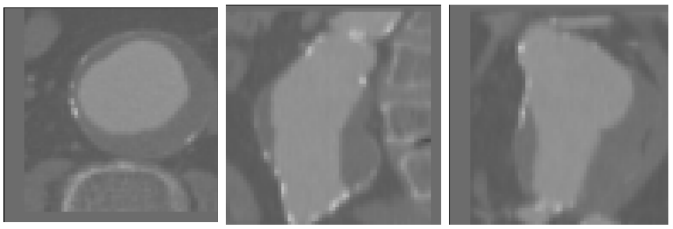


Figure 6.3: Example of input images after normalization.

6.3.3 Network Implementation and Training

We trained and evaluated our network on the training dataset via four-fold cross-validation. The network architecture is trained with sampled patches of size 64×64×64, the leaky ReLu, with batch size of 6, validation batch size of 12, padding=same, with instance normalization instead of commonly used batch normalization and without validation patch overlap. Training is done using the Adam optimizer with an initial learning rate $lr_{init} = 5 * 10^4$, and the learning rate schedule: $l2$ weight decay of 10^5 and $lr_{init} = 0.985$ epoch.

We use a dice loss function to cope with class imbalances. For the loss function, we employ a smoothed negative dice score [106], defined with:

$$D_{loss} = -\frac{2\sum_{i=1}^{N} p_i g_i + 1}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + 1} \tag{6.1}$$

where p_i represents predicted AAA probability, while g_i represents the ground truth classification for every i voxel. As seen in Equation 6.1, the summation is done over all N voxels in the CT image. Division with zero is avoided with additional ones in the denominator and numerator. All the experiments were implemented using the Keras and Tensorflow deep learning libraries. We trained our network on NVidia Geforce Titan V GPU, and training took approximately 1 hour.

The network is trained for 200 epochs since further training appears not to decrease validation loss. Moreover, Figure 6.4 indicates a decrease in loss value when the number of epochs increases. This is a clear indication that the network is successfully learning from the input data. We can also see significant improvement regarding training and validation accuracy and faster and smoother convergence of the 3D U-Net with residual blocks in an encoder pathway and deep supervision in the decoder pathway.

The 3D U-Net model has an average efficiency of 91.58% of trained accuracy while validation error was on average 85.43%. On the other hand, the addition of residual blocks in an encoder pathway obtains an average training accuracy of 93.18%, while validation accuracy is 83.89%. Similarly, the addition of only deep supervision in the decoder pathway yields training accuracy of 96.74% and validation error an average of 85.67%. A clear improvement in training is obtained with the proposed 3D U-Net with residual blocks in an encoder pathway and deep supervision in the decoder pathway. The inclusion of residual connections in an encoder pathway and deep supervision in the decoder pathway yields an average training accuracy of 96.07% and a validation accuracy of 95.03%.

6.4 Experiments and Results

In our experiments, we train four different architectures: (1) 3D U-Net network architecture, (2) 3D U-Net with residual blocks in encoder pathway (3D U-Net RE), (3) 3D U-Net with deep supervision (3D U-Net DS) and (4) 3D U-Net with residual blocks in an encoder pathway and deep supervision in decoder pathway (3D U-Net RE+DS). In this way, we demonstrate the effect of residual units and deep supervision for this particular clinical application. To evaluate the segmentation performance of the proposed method, we calculate the DSC of predicted AAA segmentations. Based on these results, it is shown that the inclusion of residual blocks into U-Net's encoder pathway and deep supervision in the decoder pathway yields significantly better results than plain 3D U-Net architecture. Detailed qualitative segmentation results is presented in Table 6.1.

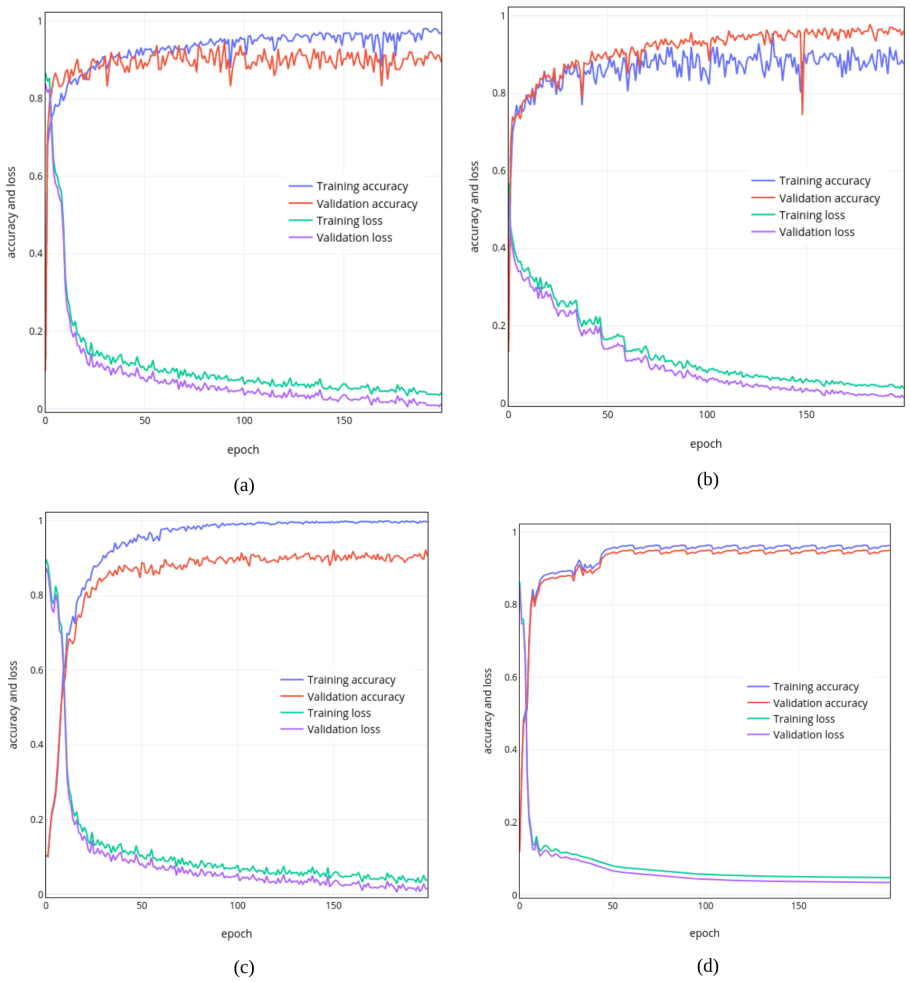


Figure 6.4: Training and validation accuracies for different networks. (a) 3D U-Net network architecture, (b) 3D U-Net with residual blocks in encoder pathway, (c) 3D U-Net with deep supervision and (d) 3D U-Net with residual blocks in an encoder pathway and deep supervision in decoder pathway.

For plain U-Net, we obtain an average DSC of 74.53%. An addition of residual connections in the encoding pathway yields improvement of 2.36%, while the addition of deep supervision obtains an improvement of 12.7% in comparison to plain 3D U-Net. The most significant improvements in DSC overlap are noticed when both residual connections and deep supervision are applied. We obtained a DSC of 91.03% using the proposed modified 3D U-Net with deep supervision. The reason behind this improvement is in addition of auxiliary supervision branches in the decoding pathway of the network. Boxplots showing the distribution of the DSC for AAA for different networks are shown in Figure 6.5.

Figure 6.6 show visual examples of obtained segmentation predictions. Visual comparisons between the ground truth and obtained segmentations are shown in Figure 6.8. Some observed difficulties in

Table 6.1: Obtained results for AAA segmentation.

Network	Average DSC
3D U-Net	0.7453 ± 0.112
3D U-Net (RE)	0.7689 ± 0.208
3D U-Net (DS)	0.8723 ± 0.124
3D U-Net RE+DS	0.9103 ± 0.156

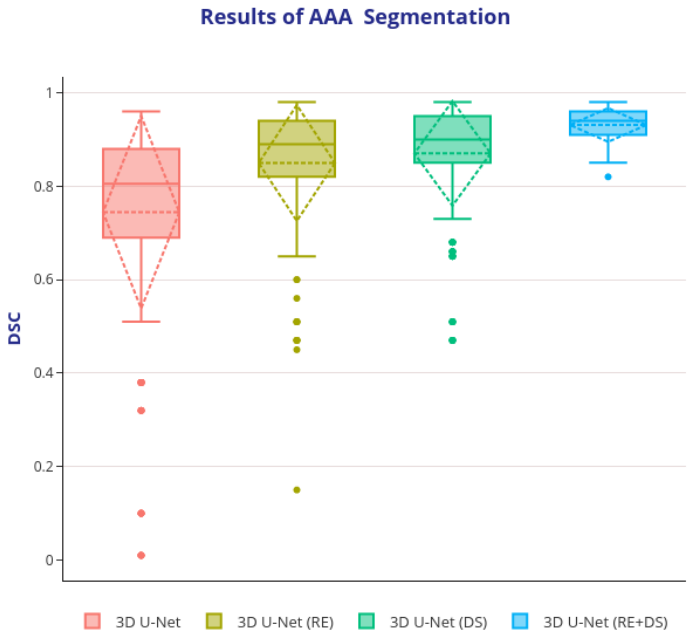


Figure 6.5: Boxplots showing the DSC dispersion for AAA using different segmentation networks. Boxplot illustrates interquartile range (bounds of box), mean (X inside a box), median (centerline), maximum and minimum values (whiskers) and outliers (circles outside whiskers).

segmentation are caused by the low image quality, contrast differences, and the highly anatomical complexity of the structures. Here, we may also observe the influence of the overfitting issue on the final segmentation result in the coronal view. This failure is due to the model hardly distinguishing between background and AAA structure due to low contrast.

Figure 6.7 shows 3D visualization of the best and the worse AAA segmentation cases obtained using our proposed 3D U-Net architecture with residual blocks in the encoding pathway and deep supervision in the decoding pathway.

6.4.1 Comparison with Other Methods

In Table 6.2, we compare the average obtained DSC of the proposed method to the current state-of-the-art that deals with AAA segmentation. Our method produces comparative results and shows higher

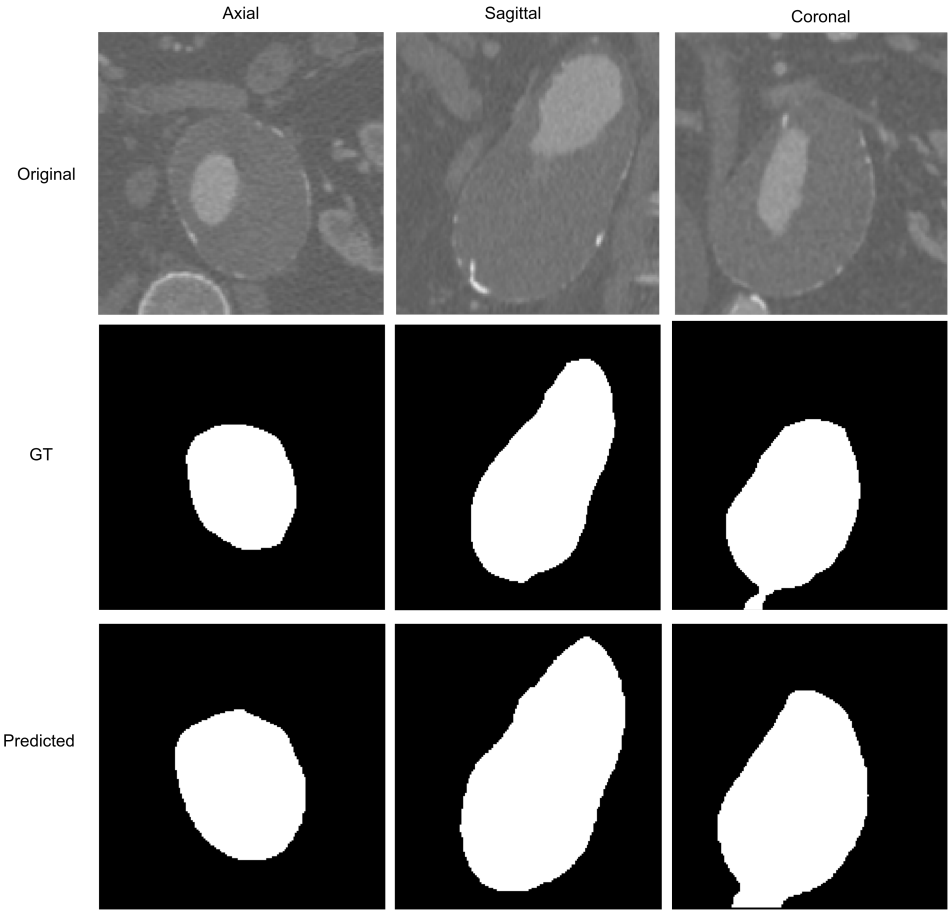


Figure 6.6: An example of obtained AAA segmentations. Top row: an original image. Middle row: ground truth. Bottom row: obtained segmentation predictions.

DSC overlap for AAA segmentation than the current state-of-the-art. Nevertheless, a few things need to be addressed. The main downside of our approach is that we, in lack of different, publically available datasets, train on already cropped images within the AAA ROI, while the state-of-the-art methods use full cardiac CTA images. Comparison to the state-of-the-art may not be fully adequate as the rest of the methods, besides segmentation, deal with AAA ROI detection. Therefore, the training on our dataset was much easier.

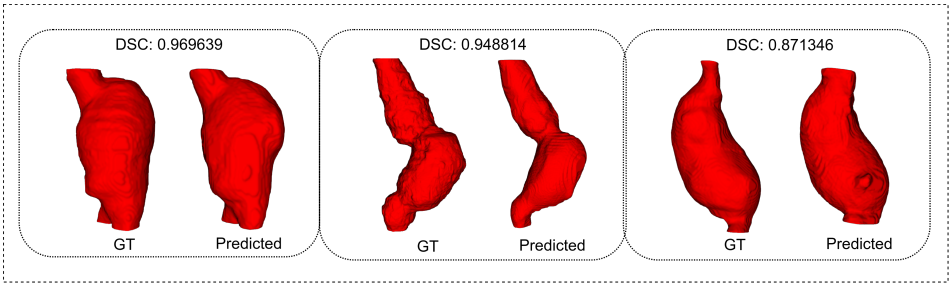


Figure 6.7: 3D visualization of the AAA results of the CT test datasets.

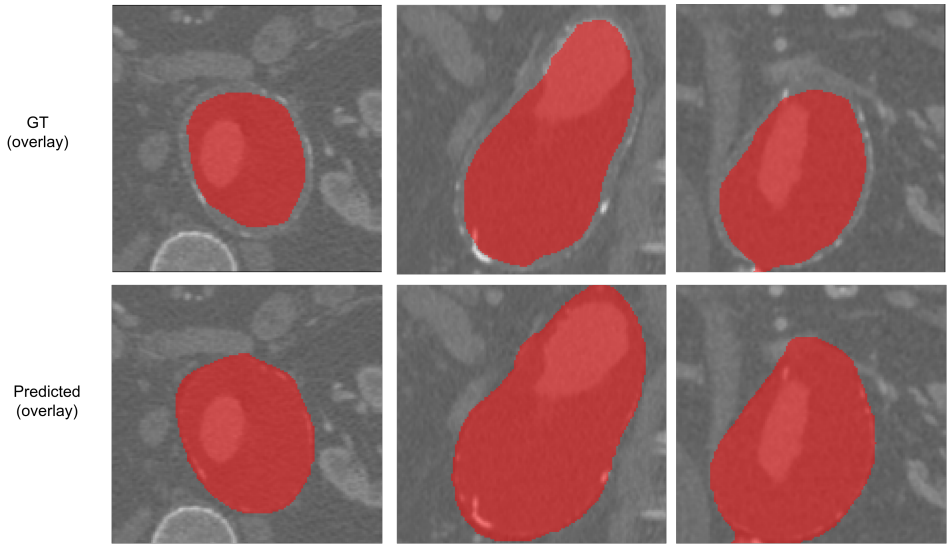


Figure 6.8: Comparison between ground truth and obtained AAA segmentations. Top row: an original AAA images with GT overlay. Bottom row: AAA images with obtained segmentation predictions.

Table 6.2: Comparison of proposed method with the state-of-the-art

Method	**Authors**	**Average DSC**
Active contours + thresholding	Lareyre et. al [94]	0.88 ± 0.12
Deep AAA	Lu et. al [106]	0.873 ± 0.129
Adapted DetectNet + FCN + Holistically-Nested Edge Detection Network	Linares et. al [100]	0.82 ± 0.07
U-Net with small training dataset	Zheng et al. [183]	0.824 ± 0.131
3D U-Net with deep supervision	Proposed method	0.9103 ± 0.156

6.5 Conclusion

The risk of aneurysm rupture is a critical unmet requirement in the assessment of AAA disease. The first step in providing such an assessment is in obtaining precise aneurysm segmentations. This segmentation method enables a more detailed examination of the AAA, which may aid in more precisely estimating the rupture risk.

In this chapter, we presented our proposed method for AAA segmentation. The proposed network relies on a modified 3D U-Net architecture. We introduce residual connections in the encoder pathway and deep supervision in the decoder pathway. Here, encoding units capture context information while decoding units restore details and spatial dimensions to enable pixel-wise classification. We train four different architectures: (1) 3D U-Net network architecture, (2) 3D U-Net with residual blocks in encoder pathway (3D U-Net RE), (3) 3D U-Net with deep supervision (3D U-Net DS), and (4) 3D U-Net with residual blocks in an encoder pathway and deep supervision in decoder pathway (3D U-Net RE+DS). All four networks are trained from scratch. In this way, we demonstrate the effect of residual units and deep supervision for this particular clinical application. To increase the robustness of the results, all four networks are trained, validated, and evaluated on 19 pre-operative CTA volumes from different patients using a 4-fold cross-validation approach. In the first experiment, we train the original 3D U-Net, which we use as a baseline. In the second experiment, we add residual connections into the 3D U-Net encoding pathway to explore the benefits of shortcut connections. This modification improved results by 2.36%. In the third experiment, we add only deep supervision to our baseline. This modification showed a vast improvement of 12.7% DSC in comparison to the original 3D U-Net. In the final experiment, our proposed solution for AAA segmentation - 3D U-Net with residual connections in the encoder pathway and deep supervision in the decoding pathway obtains a DSC of 91.03%. This shows an improvement of 16.5% in comparison to the baseline 3D U-Net architecture. Finally, we compared obtained results to state-of-the-art methods. We obtained a DSC of 91.03% using a modified 3D U-Net with deep supervision. Overall, results obtained with modified 3D U-Net with deep supervision show much higher DSC than those obtained using the original 3D U-Net.

CHAPTER 7

Conclusion

7.1 Conclusion

Recent advances in medical imaging have been facilitated by the widespread application of deep learning techniques. Supervised machine learning with CNNs has been a major driver of this change. Typically, CNNs work with images and produce a single prediction per image pattern, such as an image class label or disease burden quantification. FCNs, similar networks to CNNs, are also commonly used. They predict values for each pixel or voxel rather than a single value for the entire image. Precise segmentation models can provide fast and reliable quantification of tissue volume, eliminating the need for tedious manual labeling. These networks have a large number of parameters that can be optimized or trained by frequently providing training patterns and changing the network parameters to reduce the discrepancy between the expected and desired output values.

This thesis introduced one theoretical improvement of deep learning mechanisms by introducing a novel connectivity structure of residual units. Further, we introduced a series of deep-learning methods for heart and heart chambers segmentation. We focus on improving deep learning segmentation methods for whole heart segmentation, bi-ventricle and myocardium segmentation, and abdominal aortic aneurysm segmentation. Different cardiovascular structures are chosen to show the applicability of the deep learning methods to various segments of the cardiovascular system. Although cardiac images have been chosen as a target organ for analysis, the proposed methods can be applied to any other organs and image modalities. In this chapter, we first review our main contributions and then outline a few directions for future research.

7.1.1 Review of our Contributions

This Thesis developed new, robust, and accurate methods for cardiac image segmentation and analysis. This Thesis focuses on improving deep learning-based cardiovascular segmentation methods for whole heart segmentation, bi-ventricle, myocardium segmentation and quantification, and abdominal aortic aneurysm segmentation. During the literature review, we have determined that there is still a need to improve the performance of deep neural networks while maintaining high accuracy in medical segmentation tasks. Since common obstacles in training deeper neural network architectures are the appearance of vanishing gradients, accuracy degradations and extensive parameter growth that lead to computationally expensive models, we mainly focus on developing methods that would help alleviate previously mentioned problems.

In Chapter 4, we introduced a novel connectivity structure of residual units named feature merge pre-activation residual units (FM-Pre-ResNets) that allow the creation of distinctly deeper models without an increase in the number of network parameters compared to the pre-activation residual units. FM-Pre-ResNets adds the two additional convolutional layers at the top and the bottom of the pre-activation residual block. The top convolution layer balances the parameters of the two branches while the bottom layer reduces the channel dimension. In this way, it is possible to construct a deeper model with similar or fewer parameters than the original pre-activation residual unit. After that, we incorporate new FM-Pre-ResNets into a new 3D encoder-decoder architecture based on FM-Pre-ResNets and variational autoencoder (VAE) is proposed for the task of segmenting the entire heart from CT and MR images. Here, FM-Pre-ResNet units are used to learn a low-dimensional representation of the input during the encoding stage. Following that, the variational autoencoder (VAE) reconstructs the input picture from the low-dimensional latent space, ensuring that the model weights are strongly regularized while also avoiding overfitting on the training data. The decoding stage generates the final segmentation of the complete heart.

In Chapter 5 we present modified 3D U-Net architecture that incorporates SERes blocks into 3D U-Net architecture (3D SERes-U-Net) for the task of LV, RV, and Myo segmentation and quantification. The SERes blocks incorporate channel-wise squeeze and excitation operations into residual learning. An adaptive feature re-calibration ability of squeeze and excitation operations boosts the network's representational power, while feature reuse utilizes effective learning of the features, which improves segmentation performance.

In Chapter 6, we present a modified 3D U-Net architecture with the addition of residual units in the contracting pathway and deep supervision in expanding pathways for the task of an abdominal aortic aneurysm segmentation. The addition of residual units in the contracting pathway preserves information. It significantly increases network

performance, while the addition of deep supervision in expanding pathway inject gradient signals deep into the network.

In this Thesis, cardiac images have been chosen as a target organ for analysis; however, the proposed methods can be applied to any other organs and image modalities.

In terms of publications, so far this work resulted in two journals in the Science Citation Index Expanded (SCIE), one journal in the Emerging Sources Citation Index (ESCI) and five proceedings of international conferences (as a first author). Additionally, the research work during this thesis that contributions to other people's work (as a co-author) resulted in the two journals in the Science Citation Index Expanded (SCIE), and five proceedings of international conferences. To summarize, the work conducted during this Thesis resulted in 5 journal publications (of which 3 as the first author), 10 papers are published at international conferences (of which 5 as the first author), and 1 publication in book chapters (as co-author).

7.1.2 Future Research

The research presented in this thesis opens up several directions for future work. First, medical images, such as those from CT or MRI, are often very large. Spatial 3D data naturally requires 3D models, which consume a large amount of GPU resources. Additional research and development to create lightweight models for both training and inference are needed. Second, deep learning models have high capacity and a high degree of complexity, resulting in low generalization capacity for outlier samples. In addition, abnormal tissues can differ significantly in size and shape, leading to significant differences in test images. Third, there is an increasing need to alleviate and solve key challenges for the potential use and transition of deep learning methods into real clinical practice.

Deep learning models usually require a large amount of annotated samples to train neural networks. This is one of the main challenges in applying deep learning-based methods for medical image segmentation and analysis. Medical image datasets are usually minimal due to patient privacy and lack of annotation by trained radiologists. Moreover, it is challenging to acquire training samples in many real-world scenarios, especially for cardiovascular data. To address this challenge, the focus should be on self-supervised methods, the use of deep generative models and the development of advanced data augmentation strategies to enlarge the number of training samples. It is also necessary to collaborate with different hospitals and physicians to produce as much annotated data as possible from different clinical settings. The potential increase in data diversity (data obtained from different imaging devices, different ethnological groups, and geographical areas) would enable the development of more accurate methods, accelerating the transition of deep learning into actual clinical practice. Other constraints for

incorporating deep learning methods into real clinical practice are those inherent in the deep learning field, such as logistical issues in deployment, consideration of adoption barriers and necessary route modifications. While various clinical evaluations are conducted as part of controlled trials and are often seen as the gold standard for evidence production, they are not always appropriate or practical. Additional performance indicators should be set to convey real-world clinical relevance and be easily understood by physicians. It is critical to conduct robust clinical evaluations that utilize criteria that are obvious to physicians and go beyond only technical accuracy, i.e., to include measures of quality of care. Moreover, legal regulation that strikes a balance between the speed of innovation and the potential for harm is necessary to guarantee that patients are not exposed to harmful interventions or denied access to helpful advanced procedures.

Finally, although numerous methods have been developed in the literature for segmenting the whole heart and heart chambers, they are ineffective for images with severe CHD, which have significant heterogeneity in heart shape and great vessel connections. In future research, we will aim to combine the capabilities of deep learning for processing regular structures with those of graph algorithms for dealing with large deviations and provide a framework for segmenting the entire heart and large vessels in congenital heart disease, as graph matching has already shown success in a number of applications with large deviations.

Acknowledgement

This work has been supported in part by the Croatian Science Foundation under the project UIP-2017-05-4968

Bibliography

[1] Automated Cardiac Diagnosis Challenge (ACDC) MICCAI challenge 2017. *Post-2017-MICCAI-challenge testing phase*. URL: https://acdc.creatis.insa-lyon.fr/challenges. (Accessed: 12.01.2021).

[2] Automated Cardiac Diagnosis Challenge (ACDC) MICCAI challenge 2017. *Post-2017-MICCAI-challenge testing phase*. URL: https://acdc.creatis.insa-lyon.fr/challenges. (Accessed: 12.01.2021).

[3] Mortazi A., Burt J., and Bagci U. "Multi-Planar Deep Segmentation Networks for Cardiac Substructures from MRI and CT". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*. Ed. by Descoteaux M. et al. Cham: Springer International Publishing, 2017, pp. 287–295.

[4] Philip I. Aaronson, Jeremy P.T. Ward, and Charles M. Wiener. "The Cardiovascular System at a Glance". In: 1999.

[5] Anthony Adam Duquette et al. "3D segmentation of abdominal aorta from CT-scan and MR images". In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 36 (Jan. 2012), pp. 294–303. DOI: 10.1016/j.compmedimag.2011.12.001.

[6] P. Alvarez and W. Tang. "Recent Advances in Understanding and Managing Cardiomyopathy". In: *F1000Research* 6 (2017).

[7] Gregory Artz and Joshua Wynne. "Restrictive cardiomyopathy". In: *Current Treatment Options in Cardiovascular Medicine* 2 (2000), pp. 431–438.

[8] Takeshi Baba et al. "Clinical Outcomes of Total Endovascular Aneurysm Repair for Aortic Aneurysms Involving the Proximal Anastomotic Aneurysm following Initial Open Repair for Infrarenal Abdominal Aortic Aneurysm." In: *Annals of vascular surgery* 49 (2018), pp. 123–133.

[9] Ana Maria Balahura, Daniela Bartoș, and Elisabeta Bădilă. "Right Ventricular Normal Function". In: 2018.

[10] Pierre Baldi. "Autoencoders, Unsupervised Learning, and Deep Architectures". In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. Ed. by Isabelle Guyon et al. Vol. 27. Proceedings of Machine Learning Research. Bellevue, Washington, USA: PMLR, 2012, pp. 37–49.

[11] Ishita Banerji. "Cor Triatriatum Sinister". In: *Journal of The Indian Academy of Echocardiography & Cardiovascular Imaging* 4 (2020), pp. 45 –48.

[12] Christian F. Baumgartner et al. "An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Mihaela Pop et al. Cham: Springer International Publishing, 2018, pp. 111–119.

[13] Daniel Bell. "Left ventricular hypertrophy". In: *Radiopaedia.org* (2019).

[14] Abi Berger. "Magnetic resonance imaging". In: *BMJ* 324.7328 (2002), p. 35. DOI: 10.1136/bmj.324.7328.35.

[15] Priyanka T Bhattacharya and Sandeep Sharma. "Right Ventricular Hypertrophy". In: 2019.

[16] Bruce Blausen. *Heart Wall by Blausen - Own Work, CC BY-SA 3.0*. 2013. URL: https://commons.wikimedia.org/wiki/File:Blausen_0470_HeartWall.png.

[17] Eugene Braunwald. "Cardiomyopathies". In: *Circulation Research* 121.7 (2017), pp. 711–721. DOI: 10.1161/CIRCRESAHA.117.311812.

[18] Payer C. et al. "Multi-label Whole Heart Segmentation Using CNNs and Anatomical Label Configurations". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Pop M. et al. Cham: Springer International Publishing, 2018, pp. 190–198.

[19] Payer C. et al. "Regressing Heatmaps for Multiple Landmark Localization Using CNNs". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*. Ed. by Ourselin S. et al. Cham: Springer International Publishing, 2016, pp. 230–238.

[20] Wang C. and Smedby O. "Automatic Whole Heart Segmentation Using Deep Learning and Shape Context". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Pop M. et al. Cham: Springer International Publishing, 2018, pp. 242–249.

[21] Ye C. et al. "Multi-Depth Fusion Network for Whole-Heart CT Image Segmentation". In: *IEEE Access* 7 (2019), pp. 23421–23429.

[22] Hu Cao et al. *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*. 2021. arXiv: 2105.05537 [eess.IV].

[23] Caroline CARADU et al. "Fully automatic volume segmentation of infra-renal abdominal aortic aneurysm CT images with deep learning approaches versus physician controlled manual segmentation". In: *Journal of Vascular Surgery* (2020). ISSN: 0741-5214. DOI: https://doi.org/10.1016/j.jvs.2020.11.036. URL: https://www.sciencedirect.com/science/article/pii/S0741521420325106.

[24] Govind B. Chavhan et al. "Principles, techniques, and applications of T2*-based MR imaging and its special applications." In: *Radiographics : a review publication of the Radiological Society of North America, Inc* 29 5 (2009), pp. 1433–49.

[25] Chen Chen et al. "Deep Learning for Cardiac Image Segmentation: A Review". In: *Frontiers in Cardiovascular Medicine* 7 (2020), p. 25. ISSN: 2297-055X. DOI: 10.3389/fcvm.2020.00025. URL: https://www.frontiersin.org/article/10.3389/fcvm.2020.00025.

[26] Jieneng Chen et al. *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. 2021. arXiv: 2102.04306 [cs.CV].

[27] Liang-Chieh Chen et al. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (2018), pp. 834–848.

[28] OpenStax College. *Illustration from Anatomy and Physiology*. 2013. URL: https://en.wikipedia.org/wiki/File:2007_Ventricular_Muscle_Thickness.jpg.

[29] Jesse A. Columbo et al. "Design of The PReferences for Open Versus Endovascular Repair of Abdominal Aortic Aneurysm (PROVE-AAA) Trial." In: *Annals of vascular surgery* (2019).

[30] Johnston KW Cronenwett JL. *Rutherford's Vascular Surgery*. 2010.

[31] Kristopher S. Cunningham, Danna A. Spears, and Melanie Care. "Evaluation of cardiac hypertrophy in the setting of sudden cardiac death". In: *Forensic Sci. Res.* 4 (2019), pp. 223–240.

[32] World Health Organization Cardiovascular diseases (CVDs). *Fact Sheet*. 2018.

[33] François Dagenais. "Anatomy of the thoracic aorta and of its branches." In: *Thoracic surgery clinics* 21 2 (2011), pp. 219–27, viii.

[34] Cirean Dan et al. "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images". In: *Proceedings of Neural Information Processing Systems* 25 (Jan. 2012).

[35] Frank M Davis, D. Rateri, and A. Daugherty. "Mechanisms of aortic aneurysm formation: translating preclinical studies into clinical therapies". In: *Heart* 100 (2014), pp. 1498 –1505.

[36] Louis J. Dell'Italia. "Anatomy and physiology of the right ventricle." In: *Cardiology clinics* 30 2 (2012), pp. 167–87.

[37] Stefanie Demirci, Guy Lejeune, and Nassir Navab. "Hybrid deformable model for aneurysm segmentation". In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2009, pp. 33–36. DOI: 10.1109/ISBI.2009.5192976.

[38] D. Dey et al. "Comprehensive Non-contrast CT Imaging of the Vulnerable Patient". In: 2011.

[39] Damini Dey et al. "Automated Three-dimensional Quantification of Noncalcified Coronary Plaque from Coronary CT Angiography: Comparison with Intravascular US". In: *Radiology* 257 (Nov. 2010), pp. 516–22. DOI: 10.1148/radiol.10100681.

[40] Edoarado. *Aorta artery and it's branches in anterior views*. 2012. URL: https://commons.wikimedia.org/wiki/File:Aorta_scheme_en.svg.

[41] Jan Egger et al. *Aorta Segmentation for Stent Simulation*. 2011. arXiv: 1103.1773 [cs.CV].

[42] Alice Fantazzini et al. "3D Automatic Segmentation of Aortic Computed Tomography Angiography Combining Multi-View 2D Convolutional Neural Networks". In: *Cardiovascular Engineering and Technology* 11 (Aug. 2020), pp. 1–11. DOI: 10.1007/s13239-020-00481-z.

[43] Francesca Flocco et al. *Congenital Heart Diseases*. Jan. 2019, p. 303. ISBN: 978-3-319-78421-2. DOI: 10.1007/978-3-319-78423-6.

[44] Moti Freiman et al. "AN iterative model-constrained graph-cut algorithm for Abdominal Aortic Aneurysm thrombus segmentation". In: *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2010, pp. 672–675. DOI: 10.1109/ISBI.2010.5490085.

[45] Litjens G. et al. "State-of-the-Art Deep Learning in Cardiovascular Image Analysis". In: *JACC: Cardiovascular Imaging* 12 (2019), pp. 1549–1565.

[46] Yarin Gal and Zoubin Ghahramani. *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*. 2016. arXiv: 1506.02142 [stat.ML].

[47] Raul-Ronald Galea et al. "Region-of-Interest-Based Cardiac Image Segmentation with Deep Learning". In: *Applied Sciences* 11 (Feb. 2021), p. 1965. DOI: 10.3390/app11041965.

[48] Gaetan Galisot, Thierry Brouard, and Jean-Yves Ramel. "Local Probabilistic Atlases and a Posteriori Correction for the Segmentation of Heart Images". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Cham: Springer International Publishing, 2018, pp. 207–214.

[49] Pascal Getreuer. "Linear Methods for Image Interpolation". In: *Image Processing On Line* 1 (Sept. 2011). DOI: 10.5201/ipol.2011.g_lmii.

[50] Francesca Girolami et al. "Clinical Features and Outcome of Hypertrophic Cardiomyopathy Associated With Triple Sarcomere Protein Gene Mutations". In: *Journal of the American College of Cardiology* 55 (Apr. 2010), pp. 1444–53. DOI: 10.1016/j.jacc.2009.11.062.

[51] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].

[52] Dorn G.W. and Molkentin J.D. "Manipulating cardiac contractility in heart failure". In: *Circulation* 109 (2004), pp. 150–158.

[53] Chouaib H, Fellat N, and Hatem S. "Aortopulmonary Window Associated with a Patent Ductus Arteriosus in an Adult". In: 2020.

[54] Marija Habijan et al. "Overview of the Whole Heart and Heart Chamber Segmentation Methods". In: *Cardiovascular Engineering and Technology* (Nov. 2020). DOI: 10.1007/s13239-020-00494-8.

[55] Marija Habijan et al. "Segmentation and Quantification of Bi-Ventricles and Myocardium Using 3D SERes-U-Net". In: 2021.

[56] Marija Habijan et al. "Whole Heart Segmentation Using 3D FM-Pre-ResNet Encoder–Decoder Based Architecture with Variational Autoencoder Regularization". In: *Applied Sciences* 11.9 (2021). ISSN: 2076-3417. DOI: 10.3390/app11093912. URL: https://www.mdpi.com/2076-3417/11/9/3912.

[57] John T. Hathcock and Russ L. Stickle. "Principles and Concepts of Computed Tomography". In: *Veterinary Clinics of North America: Small Animal Practice* 23.2 (1993), pp. 399–415. ISSN: 0195-5616. DOI: https://doi.org/10.1016/S0195-5616(93)50034-7. URL: https://www.sciencedirect.com/science/article/pii/S0195561693500347.

[58] Mark Hazebroek, Robert Dennert, and Stephane Heymans. "Idiopathic dilated cardiomyopathy: possible triggers and treatment strategies". In: *Netherlands heart journal : monthly journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation* 20 (May 2012), pp. 332–5. DOI: 10.1007/s12471-012-0285-7.

[59] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.

[60] Kaiming He et al. "Identity Mappings in Deep Residual Networks". In: *Computer Vision ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 630–645.

[61] Healthand. *Cardiomyopathy*. 2021. URL: https://healthand.com/hr/topic/general-report/cardiomyopathy.

[62] Larissa Heinrich et al. "Synaptic Cleft Segmentation in Non-Isotropic Volume Electron Microscopy of the Complete Drosophila Brain". In: *MICCAI*. 2018.

[63] F. Hodges. "Anatomy of the ventricles and subarachnoid spaces". In: *Seminars in Roentgenology* 5 (1970), pp. 101–121.

[64] H. Hong et al. "Imaging of Abdominal Aortic Aneurysm: the present and the future." In: *Current vascular pharmacology* 8 6 (2010), pp. 808–819.

[65] Ho Aik Hong and U. U. Sheikh. "Automatic detection, segmentation and classification of abdominal aortic aneurysm using deep learning". In: *2016 IEEE 12th International Colloquium on Signal Processing Its Applications (CSPA)*. 2016, pp. 242–246. DOI: 10.1109/CSPA.2016.7515839.

[66] J. Hu, L. Shen, and G. Sun. "Squeeze-and-Excitation Networks". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7132–7141. DOI: 10.1109/CVPR.2018.00745.

[67] Jie Hu et al. "Squeeze-and-Excitation Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP (Sept. 2017). DOI: 10.1109/TPAMI.2019.2913372.

[68] Yuming Hua, Junhai Guo, and Hua Zhao. "Deep Belief Networks and deep learning". In: *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things*. 2015, pp. 1–4. DOI: 10.1109/ICAIOT.2015.7111524.

[69] G. Huang et al. "CondenseNet: An Efficient DenseNet Using Learned Group Convolutions". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[70] Gao Huang et al. "Densely Connected Convolutional Networks". In: July 2017. DOI: 10.1109/CVPR.2017.243.

[71] Grinias Ilias and Georgios Tziritas. "Fast Fully-Automatic Cardiac Segmentation in MRI Using MRF Model Optimization, Substructures Tracking and B-Spline Smoothing". In: Jan. 2018, pp. 91–100. ISBN: 978-3-319-75540-3. DOI: 10.1007/978-3-319-75541-0_10.

[72] Kyoko Imanaka-Yoshida. "Inflammation in myocardial disease: From myocarditis to dilated cardiomyopathy". In: *Pathology International* 70 (2019).

[73] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, 2015, pp. 448–456. URL: https://proceedings.mlr.press/v37/ioffe15.html.

[74] Fabian Isensee et al. "Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features". In: *ArXiv* abs/1707.00587 (2017).

[75] Fabian Isensee et al. "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge". In: *ArXiv* abs/1802.10508 (2017).

[76] Li J. et al. "Automatic Whole-Heart Segmentation in Congenital Heart Disease Using Deeply-Supervised 3D FCN". In: *Reconstruction, Segmentation, and Analysis of Medical Images*. Ed. by Zuluaga M. A. et al. Cham: Springer International Publishing, 2017, pp. 111–118.

[77] Stephan Jacobs et al. "3D-Imaging of cardiac structures using 3D heart models for planning in heart surgery: a preliminary study". In: *Interactive CardioVascular and Thoracic Surgery* 7.1 (Feb. 2008), pp. 6–9. ISSN: 1569-9293. DOI: 10.1510/icvts.2007.156588.

[78] Yeonggul Jang et al. "Automatic Segmentation of LV and RV in Cardiac MRI". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Mihaela Pop et al. Cham: Springer International Publishing, 2018, pp. 161–169.

[79] Zhenxiang Jiang et al. "A Deep Learning Approach to Predict Abdominal Aortic Aneurysm Expansion Using Longitudinal Data". In: *Frontiers in Physics* 7 (2020), p. 235. ISSN: 2296-424X. DOI: 10.3389/fphy.2019.00235. URL: https://www.frontiersin.org/article/10.3389/fphy.2019.00235.

[80] MacLeod K. "Recent advances in understanding cardiac contractility in health and disease". In: *F1000research* 5 (2016), p. 1770.

[81] Loay S. Kabbani et al. "Survival after repair of pararenal and paravisceral abdominal aortic aneurysms." In: *Journal of vascular surgery* 59 6 (2014), pp. 1488–94.

[82] Dongwoo Kang et al. “Heart chambers and whole heart segmentation techniques: review”. In: *J. Electronic Imaging* 21 (2012), p. 010901.

[83] Theodoros D. Karamitsos et al. “The Role of Cardiovascular Magnetic Resonance Imaging in Heart Failure”. In: *Journal of the American College of Cardiology* 54.15 (2009), pp. 1407–1424. DOI: 10.1016/j.jacc.2009.04.094.

[84] Baris Kayalibay, Grady Jensen, and Patrick van der Smagt. “CNN-based Segmentation of Medical Imaging Data”. In: *ArXiv* abs/1701.03056 (2017).

[85] Yousun Koh KenHub. *I Atria of the heart*. 2021. URL: https://www.kenhub.com/en/library/anatomy/the-atria-of-the-heart.

[86] Mahendra Khened, Alex Varghese, and Ganapathy Krishnamurthi. “Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest”. In: *STACOM@MICCAI*. 2017.

[87] Babak Khoshnood et al. “Prevalence, timing of diagnosis and mortality of newborns with congenital heart defects: A population-based study”. In: *Heart (British Cardiac Society)* 98 (Aug. 2012). DOI: 10.1136/heartjnl-2012-302543.

[88] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *CoRR* abs/1412.6980 (2015).

[89] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2014. arXiv: 1312.6114 [stat.ML].

[90] Helena Kuivaniemi et al. “Understanding the pathogenesis of abdominal aortic aneurysms”. In: *Expert Review of Cardiovascular Therapy* 13 (2015), pp. 975 –987.

[91] Yu L. et al. “Automatic 3D Cardiovascular MR Segmentation with Densely-Connected Volumetric ConvNets”. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*. Ed. by Descoteaux M. et al. Cham: Springer International Publishing, 2017, pp. 287–295.

[92] Alicia Cerezo Lajas, María del Carmen Rodríguez Guzmán, and Javier de Miguel Díez. “Left Bronchial Artery Aneurysm”. In: *Archivos de Bronconeumología (English Edition)* (2019).

[93] Luigi Landini, Vincenzo Positano, and Maria Filomena Santarelli. “Advanced Image Processing in Magnetic Resonance Imaging”. In: 2005.

[94] Fabien Lareyre et al. “A fully automated pipeline for mining abdominal aortic aneurysm using image segmentation”. In: *Scientific Reports* 9 (Sept. 2019), pp. 1–14. DOI: 10.1038/s41598-019-50251-8.

[95] Lumen Learning. *Anatomy and Physiology*. URL: https://courses.lumenlearning.com/nemcc-ap/chapter/heart-anatomy/. (Accessed: 12.4.2019).

[96] Kyungmoo Lee et al. "Three-dimensional thrombus segmentation in abdominal aortic aneurysms using graph search based on a triangular mesh". In: *Computers in biology and medicine* 40 3 (2010), pp. 271–278.

[97] M. Leventon, W. Grimson, and O. Faugeras. "Statistical shape influence in geodesic active contours". In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)* 1 (2000), 316–323 vol.1.

[98] Xiangyun Liao et al. "MMTLNet: Multi-Modality Transfer Learning Network with adversarial training for 3D whole heart segmentation". In: *Computerized Medical Imaging and Graphics* 85 (2020), pp. 101–785. ISSN: 0895-6111.

[99] Tsung-Yi Lin et al. *Feature Pyramid Networks for Object Detection*. 2017. arXiv: 1612.03144 [cs.CV].

[100] Karen Linares-Lopez et al. "Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using Deep Convolutional Neural Networks". In: *Medical Image Analysis* 46 (Mar. 2018). DOI: 10.1016/j.media.2018.03.010.

[101] Tao Liu et al. "Automatic Whole Heart Segmentation Using a Two-Stage U-Net Framework and an Adaptive Threshold Window". In: *IEEE Access* 7 (2019), pp. 83628–83636. DOI: 10.1109/ACCESS.2019.2923318.

[102] Xiangbin Liu et al. "A Review of Deep-Learning-Based Medical Image Segmentation Methods". In: *Sustainability* 13.3 (2021). ISSN: 2071-1050. DOI: 10.3390/su13031224. URL: https://www.mdpi.com/2071-1050/13/3/1224.

[103] Ze Liu et al. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. 2021. arXiv: 2103.14030 [cs.CV].

[104] Jonathan Long, Evan Shelhamer, and Trevor Darrell. *Fully Convolutional Networks for Semantic Segmentation*. 2015. arXiv: 1411.4038 [cs.CV].

[105] Karen López-Linares et al. "Image-Based 3D Characterization of Abdominal Aortic Aneurysm Deformation After Endovascular Aneurysm Repair". In: *Frontiers in Bioengineering and Biotechnology* 7 (2019).

[106] Jen-Tang Lu et al. "DeepAAA: Clinically Applicable and Generalizable Detection of Abdominal Aortic Aneurysm Using Deep Learning". In: Oct. 2019, pp. 723–731. ISBN: 978-3-030-32244-1. DOI: 10.1007/978-3-030-32245-8_80.

[107] A Luk et al. "Dilated cardiomyopathy: a review". In: *Journal of Clinical Pathology* 62.3 (2009), pp. 219–225. ISSN: 0021-9746. DOI: 10.1136/jcp.2008.060731. eprint: https://jcp.bmj.com/content/62/3/219.full.pdf. URL: https://jcp.bmj.com/content/62/3/219.

[108] Chao Luo et al. "Cardiac MR segmentation based on sequence propagation by deep learning". In: *PLoS ONE* 15 (2020).

[109] Thomas F. Luscher. "Cardiomyopathies: definition, diagnosis, causes, and genetics". In: *European Heart Journal* 37.23 (June 2016), pp. 1779–1782. ISSN: 0195-668X. DOI: 10.1093/eurheartj/ehw254. eprint: https://academic.oup.com/eurheartj/article-pdf/37/23/1779/6734344/ehw254.pdf. URL: https://doi.org/10.1093/eurheartj/ehw254.

[110] I. Macía et al. "Chapter 15 - Preoperative Planning of Endovascular Procedures in Aortic Aneurysms". In: *Computing and Visualization for Intravascular Imaging and Computer-Assisted Stenting*. Ed. by Simone Balocco et al. The Elsevier and MICCAI Society Book Series. Academic Press, 2017, pp. 413–444. ISBN: 978-0-12-811018-8. DOI: https://doi.org/10.1016/B978-0-12-811018-8.00015-1. URL: https://www.sciencedirect.com/science/article/pii/B9780128110188000151.

[111] Alireza Makhzani and Brendan J. Frey. "k-Sparse Autoencoders". In: *CoRR* abs/1312.5663 (2014).

[112] Peter Marstrand et al. "Hypertrophic Cardiomyopathy With Left Ventricular Systolic Dysfunction". In: *Circulation* 141 (2020), pp. 1371 –1383.

[113] Juan Antonio Martínez-Mera et al. "A hybrid method based on level set and 3D region growing for segmentation of the thoracic aorta". In: *Computer Aided Surgery* 18 (2013), pp. 109 –117.

[114] In Maternal. "[Book] Ultrasound Of Congenital Fetal Anomalies Differential Diagnosis And Prognostic Indicators Series". In: 2021.

[115] Raghav Mehta and Jayanthi Sivaswamy. "M-net: A Convolutional Neural Network for deep brain structure segmentation". In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. 2017, pp. 437–440. DOI: 10.1109/ISBI.2017.7950555.

[116] W.B. Meijboom et al. "Diagnostic Accuracy of 64-Slice Computed Tomography Coronary Angiography. A Prospective, Multicenter, Multivendor Study". In: *Journal of the American College of Cardiology* 52 (Jan. 2009), pp. 2135–44. DOI: 10.1016/j.jacc.2008.08.058.

[117] F. Milletari, N. Navab, and S. Ahmadi. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: *2016 Fourth International Conference on 3D Vision (3DV)*. 2016, pp. 565–571. DOI: 10.1109/3DV.2016.79.

[118] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. 2016. arXiv: 1606.04797 [cs.CV].

[119] Anthony P Morise. "Exercise testing in nonatherosclerotic heart disease: hypertrophic cardiomyopathy, valvular heart disease, and arrhythmias." In: *Circulation* 123 2 (2011), pp. 216–225.

[120] Eli Muchtar, Lori A. Blauwet, and Morie A. Gertz. "Restrictive Cardiomyopathy: Genetics, Pathogenesis, Clinical Manifestations, Diagnosis, and Therapy." In: *Circulation research* 121 7 (2017), pp. 819–837.

[121] Alan B. Noble et al. "The Cardiovascular System, 2nd Edition Systems of the Body Series". In: 2010.

[122] Ronneberger O., Fischer P., and Brox T. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *LNCS* 9351 (2015), pp. 234–241.

[123] Silvia Olabarriaga et al. "Segmentation of Thrombus in Abdominal Aortic Aneurysms from CTA with Non-Parametric Statistical Grey Level Appearance Modelling". In: *IEEE transactions on medical imaging* 24 (May 2005), pp. 477–85. DOI: 10.1109/TMI.2004.843260.

[124] Daniel O'Malley, John K. Golden, and Velimir V. Vesselinov. *Learning to regularize with a variational autoencoder for hydrologic inverse analysis*. 2019. eprint: 1906.02401.

[125] Daniel O'Malley, John K. Golden, and Velimir V. Vesselinov. "Learning to regularize with a variational autoencoder for hydrologic inverse analysis". In: *ArXiv* abs/1906.02401 (2019).

[126] Santosh K Padala, José Angel Cabrera, and Kenneth A. Ellenbogen. "Anatomy of the cardiac conduction system". In: *Pacing and Clinical Electrophysiology* 44 (2020), pp. 15 –25.

[127] Jay Patravali, Shubham Jain, and Sasank Chilamkurthy. "2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation". In: *ArXiv* abs/1707.09813 (2017).

[128] Daniel J. Penny and Andrew N. Redington. "Function of the Left and Right Ventricles and the Interactions Between Them". In: *Pediatric Critical Care Medicine* 17 (2016), S112–S118.

[129] C. Petitjean and J. Dacher. "A review of segmentation methods in short axis cardiac MR images". In: *Medical image analysis* 15 2 (2011), pp. 169–84.

[130] Thuy Pham et al. "Quantification and comparison of the mechanical properties of four human cardiac valves." In: *Acta biomaterialia* 54 (2017), pp. 345–355.

[131] Eric Pierce. *Diagram of the Human Heart by By Wapcaplet - Own Work, CC BY-SA 3.0*. 2006. URL: https://commons.wikimedia.org/w/index.php?curid=830253.

[132] Quizlet Plus. *Pulmonary and systemic circuit*. URL: https://quizlet.com/268072929/pulmonary-and-systemic-circuit-diagram/. (Accessed: 12.4.2019).

[133] Laura B Presnell et al. "An Overview of Pulmonary Atresia and Major Aortopulmonary Collateral Arteries". In: *World Journal for Pediatric and Congenital Heart Surgery* 6 (2015), pp. 630–639.

[134] Dou Q. et al. "3D Deeply Supervised Network for Automated Segmentation of Volumetric Medical Images". In: *Med Image Anal* 41 (2017), pp. 40–54.

[135] Tong Q. et al. "3D Deeply-Supervised U-Net Based Whole Heart Segmentation". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Pop M. et al. Cham: Springer International Publishing, 2018, pp. 224–232.

[136] Zhaofan Qiu, Ting Yao, and Tao Mei. *Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks*. 2017. arXiv: 1711.10305 [cs.CV].

[137] Hamburger R. "Left ventricular dysfunction in ischemic heart disease". In: *Cardiovascular Innovations and Applications* 3 (2019).

[138] Sultana R. et al. "Cardiac arrhythmias and left ventricular hypertrophy in systemic hypertension". In: *J Ayub Med Coll Abbottabad* 22 (2010), pp. 155–158.

[139] Ramiah Rajeshkannan, Vimal Raj, and Sanjaya Viswamitra. *CT and MRI in Congenital Heart Diseases*. Jan. 2021. ISBN: 978-981-15-6754-4. DOI: 10.1007/978-981-15-6755-1.

[140] Paola de Rango. "Is open surgery for AAA repair a reason for concern in the EVAR era?" In: *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery* 42 2 (2011), pp. 185–6.

[141] D. Reichart et al. "Dilated cardiomyopathy: from epidemiologic to genetic phenotypes". In: *Journal of Internal Medicine* 286.4 (2019), pp. 362–372. DOI: https://doi.org/10.1111/joim.12944.

[142] Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2015).

[143] Salah Rifai et al. "Contractive Auto-Encoders: Explicit Invariance During Feature Extraction". In: *ICML*. 2011.

[144] Karen Lopez-Linares Roman et al. "3D Pulmonary Artery Segmentation from CTA Scans Using Deep Learning with Realistic Data Augmentation". In: *Image analysis for moving organ, breast, and thoracic images : third International Workshop, RAMBO 2018, fourth International Workshop, BIA 2018, and first International Workshop, TIA 2018, held in conjunction with MICCAI 2018, Granada.* 11040 (2018), pp. 225–237.

[145] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation.* 2015. arXiv: 1505.04597 [cs.CV].

[146] Wayne Rosamond et al. "Heart Disease and Stroke Statistics—2008 Update A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee". In: *Circulation* 117 (Feb. 2008), e25–146. DOI: 10.1161/CIRCULATIONAHA.107.187998.

[147] Jog Sander, B. D. Vos, and I. Isgum. "Automatic segmentation with detection of local segmentation failures in cardiac MRI". In: *Scientific Reports* 10 (2020).

[148] U. Schoepf. "CT of the Heart: Principles and Applications". In: 2005.

[149] Beth Schueler. "The AAPM/RSNA Physics Tutorial for Residents: Clinical Applications of Basic X-ray Physics Principles". In: *Radiographics : a review publication of the Radiological Society of North America, Inc* 18 (May 1998), 731–44; quiz 729. DOI: 10.1148/radiographics.18.3.9599394.

[150] Neeraj Sharma and Lalit Mohan Aggarwal. "Automated medical image segmentation techniques". In: *Journal of Medical Physics / Association of Medical Physicists of India* 35 (2010), pp. 3 –14.

[151] Falong Shen and Gang Zeng. "Weighted Residuals for Very Deep Networks". In: *CoRR* abs/1605.08831 (2016). arXiv: 1605.08831. URL: http://arxiv.org/abs/1605.08831.

[152] Hoo-Chang Shin et al. "Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks". In: *SASHIMI@MICCAI*. 2018.

[153] C H Shivarama et al. "MULTIPLE VARIATIONS OF BRANCHES OF ABDOMINAL AORTA : A CASE STUDY". In: 2012.

[154] Ashish Shrivastava et al. "Learning from Simulated and Unsupervised Images through Adversarial Training". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 2242–2251.

[155] Perler BA Sidawy AN. *Rutherford's vascular surgery and endovascular therapy*. 2018.

[156] T. Siriapisith, Worapan Kusakunniran, and P. Haddawy. "Outer Wall Segmentation of Abdominal Aortic Aneurysm by Variable Neighborhood Search Through Intensity and Gradient Spaces". In: *Journal of Digital Imaging* 31 (2018), pp. 490–504.

[157] Hamayak S. Sisakian. "Cardiomyopathies: Evolution of pathogenesis concepts and potential for new therapies." In: *World journal of cardiology* 6 6 (2014), pp. 478–94.

[158] Gerard Snaauw et al. *End-to-End Diagnosis and Segmentation Learning from Cardiac Magnetic Resonance Imaging*. 2018. arXiv: 1810.10117 [cs.CV].

[159] Scott D. Solomon et al. "Influence of Ejection Fraction on Cardiovascular Outcomes in a Broad Spectrum of Heart Failure Patients". In: *Circulation* 112 (2005), pp. 3738–3744.

[160] STACOM and MICCAI 2017. *MM-WHS: Multi-Modality Whole Heart Segmentation*. 2017. eprint: http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/ ((accessed: 23.6.2018)).

[161] STACOM and MICCAI 2017. *MM-WHS: Multi-Modality Whole Heart Segmentation*. 2017. eprint: http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/ ((accessed: 23.6.2018)).

[162] Zhonghua Sun. "Abdominal aortic aneurysm: Treatment options, image visualizations and follow-up procedures". In: *Journal of geriatric cardiology : JGC* 9 (Mar. 2012), pp. 49–60. DOI: 10.3724/SP.J.1263.2012.00049.

[163] Tomasz Szandala. "Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks". In: Jan. 2021, pp. 203–224. ISBN: 978-981-15-5494-0. DOI: 10.1007/978-981-15-5495-7_11.

[164] Clara Tam. "Machine Learning towards General Medical Image Segmentation". In: *Electronic Thesis and Dissertation Repository* (2020), pp. 1–100.

[165] Du Tran et al. *Learning Spatiotemporal Features with 3D Convolutional Networks*. 2015. arXiv: 1412.0767 [cs.CV].

[166] Pia Trip, Nico Westerhof, and Anton Vonk Noordegraaf. "Function of the Right Ventricle". In: 2014.

[167] Israel Valverde et al. "Three-dimensional patient-specific cardiac model for surgical planning in Nikaidoh procedure". In: *Cardiology in the Young* 25.4 (2015), 698–704. DOI: 10.1017/S1047951114000742.

[168] Pascal Vincent et al. "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion". In: *J. Mach. Learn. Res.* 11 (2010), pp. 3371–3408.

[169] Riccardo Vio et al. "Hypertrophic Cardiomyopathy and Primary Restrictive Cardiomyopathy: Similarities, Differences and Phenocopies". In: *Journal of Clinical Medicine* 10 (2021).

[170] Steven Walczak and Narciso Cerpa. "Artificial Neural Networks". In: *Encyclopedia of Physical Science and Technology (Third Edition)*. Ed. by Robert A. Meyers. Third Edition. New York: Academic Press, 2003, pp. 631–645. ISBN: 978-0-12-227410-7. DOI: https://doi.org/10.1016/B0-12-227410-5/00837-1. URL: https://www.sciencedirect.com/science/article/pii/B0122274105008371.

[171] Harvey D. White et al. "Left ventricular end-systolic volume as the major determinant of survival after recovery from myocardial infarction." In: *Circulation* 76 1 (1987), pp. 44–51.

[172] Adam Wittek et al. "Image, Geometry and Finite Element Mesh Datasets for Analysis of Relationship Between Abdominal Aortic Aneurysm Symptoms and Stress in Walls of Abdominal Aortic Aneurysm". In: *Data in Brief* 30 (Mar. 2020), p. 105451. DOI: 10.1016/j.dib.2020.105451.

[173] Jelmer Wolterink et al. "Automatic Segmentation and Disease Classification Using Cardiac Cine MR Images". In: (Aug. 2017).

[174] Justina Wu et al. "Cardiovascular manifestations of Fabry disease: relationships between left ventricular hypertrophy, disease severity, and alpha-galactosidase A activity." In: *European heart journal* 31 9 (2010), pp. 1088–97.

[175] Yang X. et al. "3D Convolutional Networks for Fully Automatic Fine-Grained Whole Heart Partition". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Pop M. et al. Cham: Springer International Publishing, 2018, pp. 181–189.

[176] Xulei Yang, Zeng Zeng, and Yi Su. "Deep convolutional neural networks for automatic segmentation of left ventricle cavity from cardiac magnetic resonance images". In: *IET Computer Vision* 11 (June 2017). DOI: 10.1049/iet-cvi.2016.0482.

[177] P. Yushkevich et al. "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability". In: *NeuroImage* 31 (2006), pp. 1116–1128.

[178] Shi Z. et al. "Bayesian VoxDRN: A Probabilistic Deep Voxelwise Dilated Residual Network for Whole Heart Segmentation from 3D MR Images". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2018*. Ed. by Frangi A.F., Schnabel J.A., and Davatzikos C. Cham: Springer International Publishing, 2018, pp. 569–577.

[179] Xu Z., Wu Z., and Feng J. "CFUN: Combining Faster R-CNN and U-net Network for Efficient Whole Heart Segmentation". In: *CoRR* abs/1812.04914 (2018).

[180] Matthew D. Zeiler. "ADADELTA: An Adaptive Learning Rate Method". In: *ArXiv* abs/1212.5701 (2012).

[181] Ke Zhang et al. "Residual Networks of Residual Networks: Multilevel Residual Networks". In: *CoRR* abs/1608.02908 (2016). URL: http://arxiv.org/abs/1608.02908.

[182] Ke Zhang et al. "Residual Networks of Residual Networks: Multilevel Residual Networks". In: *IEEE Transactions on Circuits and Systems for Video Technology* 28.6 (2018), 1303–1314. ISSN: 1558-2205. DOI: 10.1109/tcsvt.2017.2654543. URL: http://dx.doi.org/10.1109/TCSVT.2017.2654543.

[183] Jian-Qing Zheng et al. "Abdominal Aortic Aneurysm Segmentation with a Small Number of Training Subjects". In: *ArXiv* abs/1804.02943 (2018).

[184] Bulat A Ziganshin and John A. Elefteriades. "Triggers of Aortic Dissection". In: *Surgical Management of Aortic Pathology* (2019).

[185] Clément Zotti et al. "GridNet with Automatic Shape Prior Registration for Automatic MRI Cardiac Segmentation". In: *STACOM@MICCAI*. 2017.

[186] Özgün Çiçek et al. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*. 2016. arXiv: 1606.06650 [cs.CV].