ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

FULLY SUPERVISED AND SEMI-SUPERVISED SEMANTIC SEGMENTATION OF CARDIAC MR USING DEEP LEARNING

M.Sc. THESIS

MAHYAR BOLHASSANI

Department of Electronics and Communication Engineering

Biomedical Engineering Programme

SEPTEMBER 2021

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

FULLY SUPERVISED AND SEMI-SUPERVISED SEMANTIC SEGMENTATION OF CARDIAC MR USING DEEP LEARNING

M.Sc. THESIS

MAHYAR BOLHASSANI
(504191413)

Department of Electronics and Communication Engineering

Biomedical Engineering Programme

Thesis Advisor: Prof. Dr. ILKAY OKSUZ

SEPTEMBER 2021

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

TAM DENETIMLI VE YARI DENETIMLI SEMANTIK SEGMENTASYON DERIN ÖĞRENMEYI KULLANAN KARDIYAK MR'NIN

YüKSEK LİSANS TEZİ

MAHYAR BOLHASSANI
(504191413)

Elektronik ve Haberleşme Mühendisliği

Biomedikal Mühendisliği Programı

Tez Danışmanı: Prof. Dr. ILKAY OKSUZ

TEZİN SAVUNULDUĞU EYLÜL 2021

MAHYAR BOLHASSANI, a M.Sc. student of ITU Graduate School student ID 504191413, successfully defended the thesis entitled "FULLY SUPERVISED AND SEMI-SUPERVISED SEMANTIC SEGMENTATION OF CARDIAC MR USING DEEP LEARNING", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. ILKAY OKSUZ**
Istanbul Technical University

Jury Members : **Prof. Dr. MUSTAFA KAMASAK**
Istanbul Technical University

Prof. Dr. SEZER ULUKAYA
Trakya University

Date of Submission : 01 August 2021
Date of Defense : 02 September 2021

To my family,

FOREWORD

I would like to show my appreciation and gratitude to my supervisor, Dr. Ilkay Oksuz. I am grateful for his detailed guidance, patience, and accessibility. Also, Thanks to be a member of PIMI lab, I could finish this thesis research using the online infrastructure of the lab.

SEPTEMBER 2021

MAHYAR BOLHASSANI
(Master's Degree student)

TABLE OF CONTENTS

Page

ABBREVIATIONS

Tr	: Training set
Val	: Validation set
Tst	: Test set
CNN	: Convolutional Neural Network
LV	: Left Ventricle
RV	: Right Ventricle
Myo	: Myocardium
EF	: Ejection Fraction
HFpEF	: Preserved Ejection Fraction
HFrEF	: Reduced Ejection Fraction
DCS	: Dice Coefficient
CE	: Cross Entropy Loss Function
MMS	: M&Ms 2020 challenge dataset
STACOM	: STACOM LVQuan 18 challenge dataset
aug	: Data Augmentation
ES	: End Systole
ED	: End Diastole
SS	: Semi Supervised
FS	: Fully Supervised
HM	: Histogram Matching
FSH	: Fully Supervised with Histogram Matching
CMR	: Cardiac MRI
MRI	: Magnetic Resonance Imaging
CT	: Computed Tomography
US	: Ultrasound
CAD	: Computer Aided Detection
ANNs	: Artificial Neural Networks
DNNs	: Deep Neural Networks
CNNs	: Convolutional Neural Networks
SD	: Standard Deviation

LIST OF TABLES

Page

LIST OF FIGURES

Page

FULLY SUPERVISED AND SEMI-SUPERVISED SEMANTIC SEGMENTATION OF CARDIAC MR USING DEEP LEARNING

SUMMARY

Heart diseases are one of the primary causes of death worldwide. A key factor to accurately and effectively treat heart diseases is to have quantified measures like high-quality images of the organ. When we provided physicians with medical scans, they can pinpoint the kind of abnormality in the heart. Cardiac Ultrasound, CT, and MRI scans are some of the modalities that we can leverage, while each modality has both advantages and disadvantages. Depending on the situation and the patients' condition, we can choose a preferred modality. We concentrate on cardiac MRI, which is a non-ionizing modality that constructs high-quality images. Segmentation of different heart areas in CMR scans such as myocardium mass, wall thickness, left ventricle (LV), right ventricle volume, and ejection fraction (EF) is a quantitative measure that assists cardiologists in diagnosing cardiac failures. Thanks to computer-aided detection (CAD) advancements, the automatic segmentation of the heart cavity for diagnosis purposes alleviates the burden of quantitative interpretation of large numbers of cardiac scans for cardiologists. The ultimate goal of training an automatic model is to predict correctly on unseen data. Therefore, we need a large number of labeled data which is a tedious and expensive task. However, the variation of CMR data acquisition from different centers or vendors demands us to have training data from almost all centers and vendors for a robust model, which is almost impossible.

To address this issue, this thesis proposes a semi-supervised segmentation setup to leverage unlabeled data to segment left ventricle, right ventricle, and myocardium regions. We utilize an enhanced version of residual U-Net architecture on a large-scale cardiac MRI dataset. Handling the class imbalanced data issue using dice loss, and the improved supervised model can achieve better dice scores than a vanilla U-Net model. We applied standard augmentation techniques as well as histogram matching techniques to increase the performance of our model in the multi-domain setup. Also, we introduce a simple but efficient semi-supervised segmentation method to improve segmentation results without the need for extensive labeled data. Finally, we applied our approach on two benchmark datasets, STACOM LVQuan 18 and M&Ms 2020 challenges, to show the potency of the proposed model. The quantitative results demonstrate the effectiveness of our proposed model. The model achieves average dice scores of 0.926, 0.933, and 0.892 for left ventricle, right ventricle, and myocardium respectively.

TAM DENETIMLI VE YARI DENETIMLI SEMANTIK SEGMENTASYON DERIN ÖĞRENMEYI KULLANAN KARDIYAK MR'NIN

ÖZET

Dünya Sağlık Örgütü'ne (WHO) göre kalp hastalıkları dünya çapında başlıca ölüm nedenlerinden biridir. Kalp hastalıklarını doğru ve etkili bir şekilde tedavi etmenin kilit bir faktörü, organın yüksek kaliteli görüntüleri gibi nicel ölçümlere sahip olmaktır. Doktorlara tıbbi taramalar sağladığımızda, kalpteki anormalliğin türünü saptayabilirler. Kardiyak Ultrason, BT ve MRI taramaları, yararlanabileceğimiz yöntemlerden bazılarıdır, ancak her bir yöntemin hem avantajları hem de dezavantajları vardır. Duruma ve hastanın durumuna göre tercih edilen bir yöntem seçebiliriz. Yüksek kaliteli görüntüler oluşturan iyonlaştırıcı olmayan bir modalite olan kardiyak MRG'ye odaklanıyoruz. Segmentasyon, bilgisayarlı görü alanındaki en zorlu süreçlerden biridir. Aynı grupta, görüntü piksellerini doku, mesafe, renk ve yoğunluk gibi benzer faktörlere dayalı olarak kategorize etmeliyiz. Bu durumda her gruba bir segment diyebiliriz.

Miyokard kütlesi, duvar kalınlığı, sol ventrikül (LV), sağ ventrikül hacmi ve ejeksiyon fraksiyonu (EF) gibi CMR taramalarında farklı kalp alanlarının segmentasyonu, kardiyologlara kalp yetmezliği teşhisinde yardımcı olan nicel bir ölçüdür. Bilgisayar destekli algılama (CAD) gelişmeleri sayesinde, teşhis için kalp boşluğunun otomatik olarak segmentasyonu, kardiyologlar için çok sayıda kardiyak taramanın kantitatif yorumlanmasının yükünü hafifletir. Derin öğrenme yaklaşımlarının ortaya çıkmasından önce yaygın olarak kullanılan makine öğrenimi tabanlı yöntemler, nicelleştirilmiş ilgi alanını başarılı bir şekilde tanımlayabiliyordu. Bununla birlikte, bu yöntemler, veriler ve ağır özellik mühendisliği prosedürleri hakkında önceden bilgi sahibi olmayı gerektiriyordu. Öte yandan, derin öğrenme yöntemleri, doğrudan verilerden basit ve karmaşık özellikleri otomatik olarak öğrenir. Ve GPU kapasitesindeki artış ve veri kullanılabilirliğinde bir artış olmazsa, derin öğrenme algoritmalarının yaygın kullanımı elde edilemezdi.

Derin öğrenme algoritmaları kullanarak otomatik bir model yetiştirmenin nihai amacı, görünmeyen verileri doğru bir şekilde tahmin etmektir. Bu nedenle, özellikle tıbbi veriler söz konusu olduğunda, açıklama için tıbbi uzmanlara ihtiyaç duyulduğundan, sıkıcı ve pahalı bir iş olan birçok etiketli veriye ihtiyacımız var. Ancak, farklı merkezlerden veya satıcılardan alınan CMR verilerinin varyasyonu, neredeyse tüm merkezlerden ve satıcılardan sağlam bir model için eğitim verileri almamızı gerektiriyor ki bu neredeyse imkansız.

Sınırlı sayıda etiketlenmiş veriyle derin bir modelin eğitilmesi, fazla uydurmaya neden olabilir. Bu durumda model, test ve görünmeyen verilerde başarısız olurken

eğitim verilerini doğru bir şekilde tahmin edebilir. Bu araştırmada ilk olarak etiketli konserve yöntemiyle bazı standart büyütme tekniklerini uyguluyoruz. Veri büyütme, bir dereceye kadar aşırı uyumun üstesinden gelebilir.

Bu konuyu ele almak için bu tez, sol ventrikül, sağ ventrikül ve miyokard bölgelerini segmentlere ayırmak için etiketlenmemiş verilerden yararlanmak için yarı denetimli bir segmentasyon kurulumu önermektedir. Bildiğimiz gibi, etiketlenmemiş tıbbi veriler bol ve ucuzdur. Sınırlı etiketli verilere sahip olduğumuzda, bu verileri kullanarak daha iyi tahminler elde etmek için modelimizi geliştirebiliriz. Bu nedenle, farklı merkezlerden etiketlenmemiş verileri ve farklı zaman dilimlerinden etiketlenmemiş görüntü taramalarını içeren bir veri seti kullanıyoruz. Ek açıklamaların çoğu sistol sonu ve diyastol sonu zaman çerçevelerini içerirken diğer zaman çerçeveleri için açıklamalardan yoksundur. Sonuç olarak, önerilen modelimizin sağlamlığını güçlendirebiliriz.

Büyük ölçekli bir kardiyak MRI veri setinde artık U-Net mimarisinin geliştirilmiş bir sürümünü kullanıyoruz. Bahsedilen mimariyi seçmek için, her birinin performansını değerlendirmek için iki farklı derin mimariyi karşılaştırıyoruz. Ayrıca dikkat modülünün takılması modele resimlerde nereye bakacağı konusunda yardımcı olacaktır. Bu dikkat modülü, kardiyak MR taramalarında ilgi alanının arka plana kıyasla küçük olması bakımından etkilidir. Bu dengesizlik sorunu da ele alınmalıdır.

Zar kaybı kullanarak sınıf dengesiz veri sorununu ele alan geliştirilmiş denetimli model, çapraz entropi kaybı işlevi göz önüne alındığında vanilya U-Net modelinden daha iyi zar puanları elde edebilir.

Döndürme, çevirme ve elastik dönüşüm gibi standart veri büyütme tekniklerini uyguladık. Bu büyütme teknikleri, etiketli koruma yaklaşımlarına dayanmaktadır, böylece eğitim verilerinin sayısını artırmak için etiketli verileri yapay olarak artırabiliriz. Ek olarak, verilerimiz farklı merkezlerden veya satıcılardan geldiğinde veri değişkenliğini ele almak için etki alanı uyarlama teknikleri gereklidir. Histogram eşleştirme teknikleri, çok alanlı kurulumda modelimizin performansını artırmak için alan uyarlama tekniklerinden biridir.

Ayrıca, kapsamlı etiketlenmiş verilere ihtiyaç duymadan segmentasyon sonuçlarını iyileştirmek için basit ama verimli bir yarı denetimli segmentasyon yöntemi sunuyoruz. Veri büyütme ile zar kaybını göz önünde bulundurarak Artık U-Net mimarisini uygulayan son denetimli modelimiz ile başlıyoruz. İlk olarak, eğitilmiş, denetlenen modeli kullanarak ES ve ED zaman çerçevelerindeki etiketlenmemiş veriler için segmentasyon maskesini tahmin ediyoruz. Düşük kaliteli ek açıklamaları kullanmaktan kaçınmak için, bu tahminlerin gürültülü olup olmadığını kontrol etmek için tahmini bir döngü boyunca değerlendirir ve eğitilmiş, denetlenen model tarafından bunları göz ardı ederiz. Başka bir deyişle, tahminlerin kalitesini yinelemeli bir şekilde inceliyoruz ve düşük kaliteli ek açıklamaları kaldırıyoruz. Ardından, iki veri kümesi, manuel olarak etiketlenen veriler ve sözde etiketli veriler birleştirilir. Bu yöntemin kullanılması, modelin farklı satıcılardan veya merkezlerden gelen görüntüler üzerinde eğitim almasına olanak tanır.

Ayrıca modelimiz için hiper parametre ayarlama kurulumunu da açıklıyoruz. Derin bir modeli eğitmek için anahtar parametrelerden biri Öğrenme oranıdır. Kayıp gradyanı ile ilgili olarak eğitim modelimizin ağırlık ayarlama miktarını kontrol eder. En iyi sonuçları elde etmek için uyarlanabilir bir öğrenme oranına ihtiyacımız var, böylece

RMSprop ve Momentum ile Stokastik Gradient Descent'in bir kombinasyonu olan Adam optimizer'ı kullandık. Bu optimize edici, farklı parametreler için uyarlanabilir öğrenme oranı yöntemi olarak bilinir.

bireysel öğrenme oranlarını hesaplar. Bu nedenle, uygulamamızda Adam optimizer kullanıyoruz. Ayrıca, hem test hem de eğitim tahminlerinde en iyi doğruluğu elde etmek için erken durdurma yöntemini kullandık. Son olarak, önerilen modelin gücünü göstermek için yaklaşımımızı iki kıyaslama veri kümesine, STACOM LVQuan 18 ve M&Ms 2020 zorluklarına uyguladık. Önceki veri kümesi, A, B ve C olmak üzere üç farklı satıcıdan gelen verileri içerir. Satıcı A ve B etiketlenirken satıcı C etiketlenmez. Satıcı C'den alınan verileri yarı denetimli yöntem için etiketlenmemiş veriler olarak değerlendirdik. Nicel sonuçlar, önerilen modelimizin etkinliğini göstermektedir. Model, sol ventrikül, sağ ventrikül ve miyokard için sırasıyla 0,926, 0,933 ve 0,892 olan ortalama zar puanlarına ulaşır.

1. INTRODUCTION

In the computer vision domain, image segmentation is one of the most challenging and prominent concepts. For segmentation, image pixels are categorized based on similar factors such as texture, distance, color, and intensity in the same group. Each group is called a segment. Image segmentation has been utilized in different tasks in computer vision, from boundary detection [5] to medical image segmentation. Traditional machine learning-based medical image segmentation techniques like the atlas-based method and model-based methods (such as graph-cut [6, 7], clustering [8], active boundary [9]) have been widely used before the emergence of deep learning methods. Despite achieving good segmentation results using the traditional segmentation techniques, having prior knowledge and heavy feature engineering were the essential ingredients of successful outcomes. On the contrary, the deep learning-based methods automatically learn simple to complicated features directly from the data. Two major changes encouraged researchers worldwide to pay more attention to applying deep learning methods to medical image segmentation, increased GPU capacity within the past few years and a spike in data availability [6]. Therefore, the DL-based methods, eventually, outperformed the previous segmentation techniques [10]. Thus, in this work, we focus on applying deep learning methods to medical image segmentation, particularly cardiac MRI image segmentation.

According to World Health Organization (WHO), cardiovascular diseases (CVD) are a significant cause of death worldwide [1]. WHO anticipates that almost 23.6 million people will die from heart diseases by 2030 [11]. Hence, it is quite essential to diagnose and cure CVD accurately. Therefore, medical imaging that is considered a branch of medicine [12] assist experts in an accurate disease diagnosis and treatment in a non-invasive way. There are various medical imaging modalities based on the technologies that they are using. Magnetic resonance imaging (MRI), computed tomography (CT), and Ultrasound (US) are some of the examples of medical imaging techniques that provide non-invasive assessments of the heart. Figure 1.1 illustrates

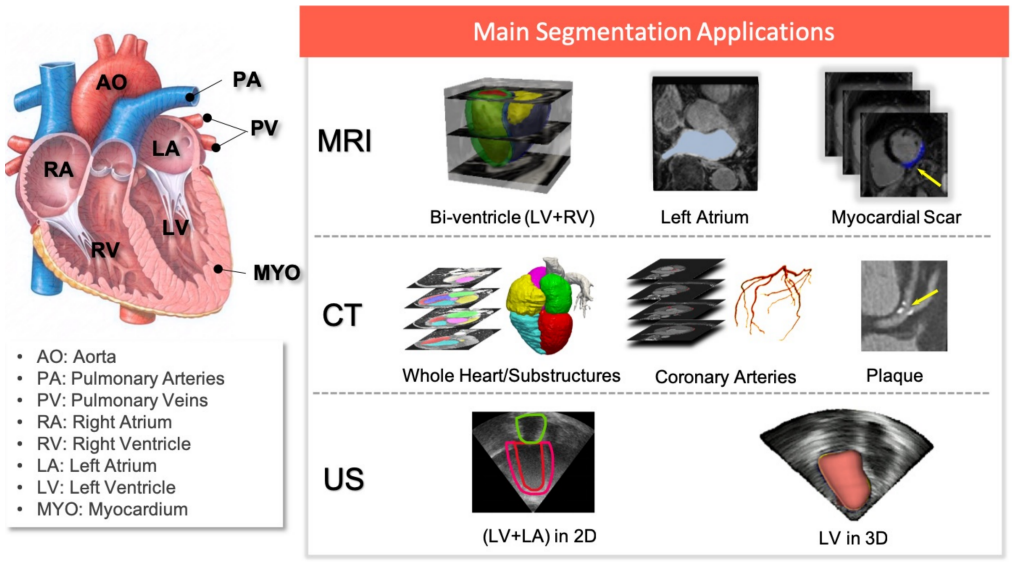

Figure 1.1 : Overview of different imaging modality and cardiac segmentation task. Image on the left shows anatomical regions of the heart [1].

different cardiac imaging modalities and the cardiac image segmentation task in detail. Also, on the left side of Figure 1.1, the anatomical regions of the heart has been shown.

1.1 Imaging Systems

1.1.1 Computed tomography(CT)

A computed tomography scan (CT scan) consists of a series of X-Ray scans from different angles by combining them using computer software. CT scans show us more detailed information about patients than X-ray scans, though the radiation dose may be harmful to patients [13].

1.1.2 Ultrasound

Cardiac Ultrasound or echocardiogram uses high-frequency sound waves to get moving pictures of the heart, such as images that show how well the blood flow works in the heart [14]. This imaging modality has been considered the general prenatal checkups due to being non-invasive, safe, painless, and cheap. Even though Ultrasound imaging is harmless for patients since there is no radiation during scanning, some drawbacks lead us to utilize other imaging modalities in some applications. The low signal-to-noise ratio and low image contrast are considered as problems that limit the achieved accuracy. In addition, it needs a skilled operator to get high-quality images.

Also, the limitation of the field of view and the trouble of distinguishing between different structures in the body -like between gas and soft tissue- are other significant problems we face in the ultrasound imaging system [15].

1.1.3 Cardiac MRI

CMR (Cardiovascular Magnetic Resonance) is complementary to other mentioned imaging techniques. CMR technique uses the same physics and fundamentals as MRI modality to acquire and reconstruct heart images. MRI imaging modality is based upon nuclear magnetic resonance. To elaborate on this fundamental, hydrogen protons, abundant in the human body, have a magnetic axis that randomly rotates when there is no external magnetic field. When we apply a strong magnetic field, all protons aligned in the direction of the external domain rotate around this new axis. This rotation is called precession. The precession frequency (Larmor frequency) is proportional to the applied magnetic field. At equilibrium state, the net longitudinal magnetization is zero. Thus, to change it, an RF wave with precession frequency is transmitted to the tissue, which disturbs the equilibrium of protons' magnetization. The amplitude and duration of the RF (Radio Frequency) wave determine the degree of excitation of protons [16]. The next state is relaxation; when the RF wave is removed, the excited protons prone to go back to the equilibrium state. This process has two parameters T1-relaxation and T2-relaxation. The former refers to when longitudinal magnetization recovers back to equilibrium. While the latter indicates the time that transverse magnetization decays. The key element of MRI images is that T1 and T2 relaxation times vary by the type of tissue, for example, T1 and T2 of lipid differs from water [16].

Based on the motion regarding the heart and respiration system, the general MRI technique cannot create high-quality images. CMR technique overcomes these problems using ECG gating and high-resolution protocols. The first solution is to ask patients to hold their breath; this eliminates the artifact referring to breathing movements. Physiologists can ask patients to hold breath, breathe again, and normal breaths. During these intervals, the heartbeat rate is low, which means the resting time in the cardiac cycle (between T-wave and next P-wave) is longer, we can start the image acquisition process using an MRI technique. Therefore, We aim to synchronize data acquisition with respect to the cardiac cycle by applying ECG gating, which is different

from ECG triggering (initialized the acquisition process, pulse sequence, when it detects R-wave). Therefore, we can acquire about thirty images considering current technologies during one cardiac cycle, which means we have a temporal resolution of 30 ms [6]. Consequently, unlike the Cardiac CT imaging system, CMR is a non-ionization technique that is considered a safe modality for patients. Also, CMR constructs high-quality images compared to Ultrasound protocol.

1.2 Description of Short-axis CMR Images

The standard imaging plane for typical MR examination is perpendicular to the heart's long axis or apex-base axis. This axis is called the short-axis plane (or two-chamber view), that the Figure 1.2 shows the details. The red rectangle indicates the region of interest (ROI) that contains two chambers of the heart, LV, RV, and the myocardium.

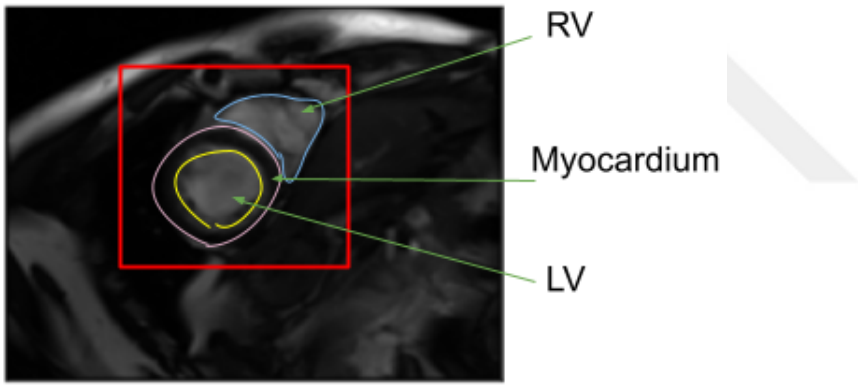

Figure 1.2 : A short-axis cardiac MR image. The red rectangle illustrates the region of interest. RV, Myocardium, and LV regions are shown in blue, purple, and yellow scribbles, respectively.

According to Figure 1.2, myocardium or heart muscle with normal values of thickness range from 6 to 16 mm surrounds the left ventricle or LV cavity region, which has an ellipsoid shape [6]. The function of LV is to pump oxygenated blood into the aorta. In contrast, the right ventricle (RV), with a convoluted crescent shape, ejects deoxygenated blood to the lungs. Physicians are interested in calculating RV and LV mass and volume at two specific moments at ES and ED. End systole (ES)is the moment or the time that the two chambers of the heart have the most contraction, while the end of diastole (ED) is the time that the two chambers have the maximum

filling [10]. Another measurement that significantly affects the diagnosis of heart failure is ejection fraction (EF). Ejection fraction defines as a percentage of the pumped blood volume out of LV with each contraction. The normal range for ejection fraction is between 52%-72% [6].Although EF is an essential gauge for heart disease diagnosis purposes, some patients with heart failure show normal EF percentages. Doctors call this condition preserved ejected fraction, referring to either preserved ejection fraction (HFpEF) or Reduced ejection fraction (HFrEF) [17]. The former refers to diastolic heart failure when the heart muscles contract during ejection but fail to relax while filling the ventricle. The latter, however, indicates systolic heart failure when the heart muscles do not effectively contract so that the heart provides less oxygenated blood for the body.

1.3 Overview of Segmentation Methods

One of the challenges of using cardiac MR scans is the number of images that radiologists need to analyze. For a cine cardiac MR scan, we need to interpret almost 260 2D images of the heart if the scan consists of 20 time-frame and 13 apex-bottom 2D images. Thus, computer-aided detection (CAD) systems can play a crucial role in automatizing the burdensome radiologist task.

1.3.1 Automatic segmentation

Before the rise of DL-based methods, researchers utilized machine learning-based approaches to automatize the annotation and interpretation of the medical scans. The emergence of Deep Neural Networks (DNNs) resulting from computational capacity advancements encouraged researchers to adapt DNNs to medical image segmentation. Using DNNs freed us from prior solid knowledge about image features, unlike machine learning-based methods. DNNs are, in fact, the stack of several Artificial Neural Networks (ANN) layers which we consider the first layer as the input layer and the last layer as the output layer, while the rest of the layers are called hidden layers. Figure 1.3 shows a DNN network in details.

The structure of DNNs allows the model to train complicated features by the combination of the simple features. If we consider the Figure 1.3, the first hidden layer may learn horizontal lines while the last hidden layer may learn the curve-like

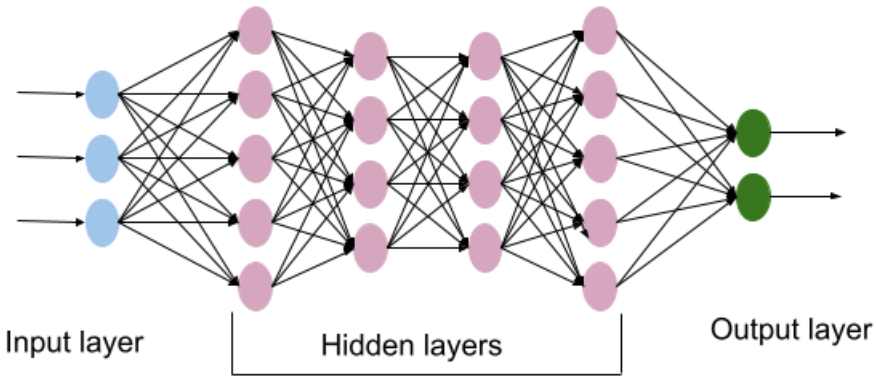

Figure 1.3 : Schematic of a Deep Neural Network (DNN).

structures as a complex feature. Then, the weights of each feature will optimize through back-propagation using optimization approaches such as gradient descent, stochastic gradient descent. The loss function evaluates the difference between the prediction (DNN output) and the labeled class by applying the training samples in each step [18].

In the computer vision domain, it is vital to capture spatial and temporal dependencies of features of the image. However, applying a DNN algorithm needs to flatten the image (for instance, with a size of 3×3) to a 9×1 array. Therefore, not only does the model lose the local information, but also the number of learnable parameters is overwhelming in medical image segmentation. Convolutional neural networks, introducing to solve the mentioned issue, are deep learning algorithms that consist of convolution layers. These convolution layers are filter functions that combine the information of neighbor information of each pixel using learnable weights – we also called them kernel or receptive field. Figure 1.4 illustrates the schematic of a CNN. It consists of convolution, pooling, and fully connected layers. Pooling layers reduce the feature map to capture a larger field of view so that it remains robust to the shape and position of the detected features [18]. We review more advanced DL-based algorithms in ***Chapter 2***.

Thus, we can train a DL-based model by feeding a large number of input training data (including the original images and their manually segmented masks) to a CNN. Based on the training samples, we need to choose an appropriate loss function (such as Cross entropy loss function) to quantify the model predictions in each iteration. However,

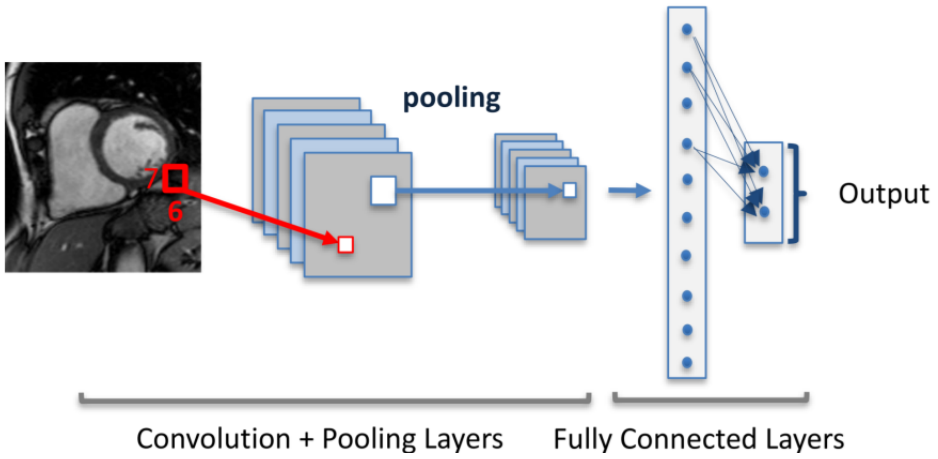

Figure 1.4 : Schematic of a Convolutional Neural Network (CNN). A CNN architecture takes a patch of a short-axis CMR image as an input. Then CNN learns features using the convolution and pooling operations. Afterward, fully connected layers flatten and reduce these spatial feature maps into a vector. Thus, the prediction can be made based on these feature maps.

to achieve high accuracy results we need considerably large numbers of annotated training data.

1.3.2 Semi-automatic segmentation

Deciphering and interpreting the medical images required medicine practitioners or expert clinicians. Although manually extracting vital information like segmentation mask images from medical images helps clinicians ensure a disease diagnosis, this process is tedious and time-consuming. In addition, it presents uncertainties of expert variabilities (inter- and intra-observer variability). Therefore, an automatic approach can eliminate those limitations. Unlike natural images, we face the lack of extensive annotated medical data; in the natural domain, ImageNet and many others are available. Also, the specialized image preprocessing for medical images based on the modality can be considered another challenge in using deep learning techniques in the medical domain. The distribution shift - common in cardiac MR scan images- when they come from different centers or acquired using various technologies (like Philips or Siemens) or various acquisition parameters exacerbates the annotation task. Consequently, distribution shifts prevent us from achieving a robust model.

Hence, semi-supervised methods are developed to leverage the unlabeled data. Adding the unlabeled data to the labeled training samples will assist the DL model in getting

familiar with features from different distributions. Figure 1.5 illustrates the schematic of a semi-supervised method.

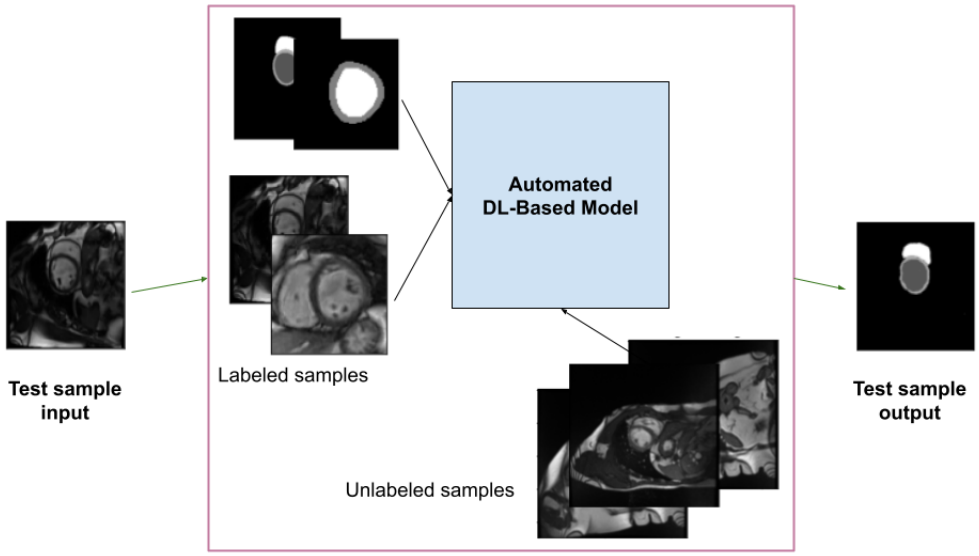

Figure 1.5 : Schematic of a semi-supervised method. Utilizing unlabeled datasets which are cheap and abundant will increase the generalizability of fully supervised methods.

1.4 Purpose of Thesis

In this work, we are concentrating on Cardiac MRI scans. Cardiac MRI segmentation divides the image into anatomically meaningful areas like Myocardium, LV volume, RV volume, etc. Fully supervised automatic segmentation methods are proposed to address the challenge of cardiac MR segmentation [1, 6, 10], though the lack of large numbers of data prone these deep models to overfitting. This limitation will result in low accurate segmented masks on unseen datasets. Furthermore, the variance of the data acquired from different centers or various technologies (vendors) imposes an extreme challenge to the deep model in terms of generalization [19]. Therefore, when we have the unseen data from a new center or vendor, the trained deep model may fail.

To achieve a robust approach for automatic segmentation of cardiac MRI scans, we assess supervised and semi-supervised deep learning methods. To address the lack of large labeled data, we first introduce both standard data augmentation techniques (rotation, translation) and domain adaptation data augmentation techniques to overcome this limitation. Despite the effectiveness of the augmentation methods, the model still may fail on unseen cardiac MRI data. Unlabeled medical data is abundant,

easy to access, and cheap. Hence, semi-supervised deep learning methods that leverage unlabeled medical data can alleviate the problem mentioned above and improve the robustness of the deep model on unseen datasets. Another solution for the domain adaptation issue that we apply in this research is histogram matching. This approach allows the deep model to mimic the intensity distribution of unlabeled data. As a result, the model can achieve better predictions. Finally, we aim to evaluate and assess both supervised and semi-supervised methods. We are trying to enhance the methods to gain accurate and robust segmentation masks for unseen medical images.

We aim to find answers to these research questions at the end of this report:

- ***Question 1*** How different deep learning architectures perform in a fully supervised manner?
- ***Question 2*** How much a deep model is sensitive to choosing the right loss function for the segmentation task?
- ***Question 3*** What is the effect of the attention module in the cardiac MR segmentation task?
- ***Question 4*** How can we leverage unlabeled images to train a deep model to have a robust model?
- ***Question 5*** How to reduce the sensitivity of our proposed semi-supervised method on the supervised model?

  We will answer the mentioned research questions in **Chapter 5**.

1.5 Overview

We organize the thesis as follow:

- **Chapter 2** reviews related work. We provide background information about DL-based algorithms in cardiac MRI segmentation.
- **Chapter 3** explains supervised medical image segmentation. Specifically we implement supervised segmentation methods on CMR datasets.

- **Chapter 4** demonstrate semi-supervised medical image segmentation to overcome the limitations of fully supervised segmentation methods. We implement a semi-supervised segmentation algorithm on CMR datasets to enhance the generalizability of the proposed model.

- **Chapter 5** summarize the contributions of the thesis and discuss future research. We answer the raised questions regarding this research in the contribution section.

2. LITERATURE REVIEW

This chapter reviews background information about medical image segmentation techniques that researchers explore after emerging deep learning approaches. We survey the literature on cardiac MR segmentation algorithms since our research focuses on CMR image segmentation methods. First, we go over the supervised methods and the way they evolve. Then, explain the approaches and methodologies that have been introduced on data augmentation. Finally, we list a number of significant yet related researches on semi-supervised medical image segmentation methods.

2.1 Supervised Methods

Supervised learning is a machine learning approach that we utilize to teach models to give us the desired output using a training set. In deep supervised learning, we use neural network algorithms to train a model. The training set is composed of input images and manually segmented mask images. Many researchers have utilized deep supervised learning methods for medical image segmentation. Authors in [20] present a fully automatized segmentation method to segment the left ventricle in short axis cine cardiac MR images. To achieve better results, they incorporate image processing techniques such as thresholding and edge detection. [21] employs a fully supervised convolutional neural network to segment the left ventricle of the heart in the CMR short axis view and combines it with a deformable model to enhance the quality of the segmentation task and increase the robustness of the model. They also validated the proposed approach on the CMR dataset provided by MICCAI 2009 LV segmentation challenge.

2.1.1 Fully convolutional networks (FCN)

Long et al. introduced Fully Convolutional Networks (FCN) in 2015 for natural image segmentation [22] . Since the fully connected layers produce large numbers of learning parameters and lose the local connectivity between image pixels, FCNs proposed as a type of CNNs that do not have any fully connected layers. Figure 2.1 illustrates the

schematic of an FCN. This model consists of an encoder-decoder structure that we can feed images with arbitrary sizes and reconstruct the output images with the same size. Using convolutional and pooling layers in the encoder path, we create the input image into a high-level latent space. Then, we can reconstruct the output image in each upsampling operation by utilizing the transposed convolution layers in the decoder path [1].

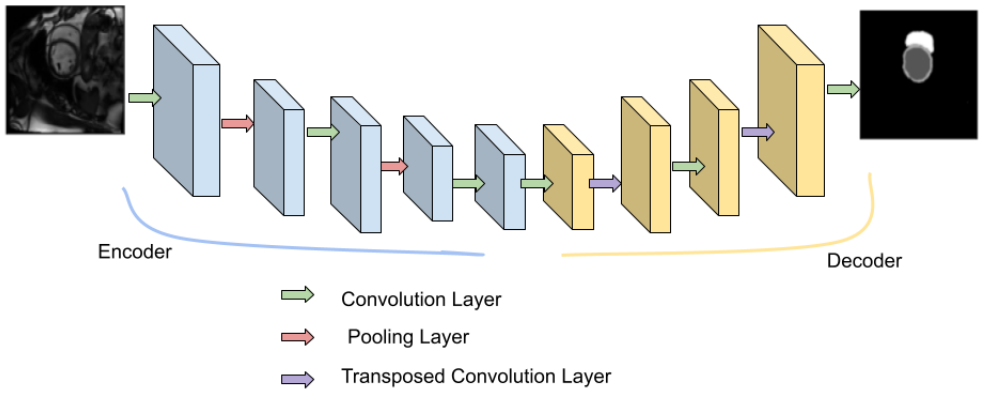

Figure 2.1 : Schematic of a Fully Convolutional Network (FCN). FCN uses an encoder-decoder structure for image segmentation. It takes the whole image as input, then through the learning process, it learns the most salient features in the encoder path. The spatial dimension is then recovered through the decoder path.

[23] utilizes a recurrent fully convolutional network to learn the image representations from 2D slices of short-axis cardiac MR scans. This approach allows the model to leverage the interslice spatial dependency information to predict better segmentation on unseen data. Their method decreases the computational time by combining both anatomical detection modules and segmentation into one architecture. [24] leverage a 3D fully convolutional network to segment CT images. They illustrate the power that a trained model using a multi-class 3D FCN deep architecture on labeled CT scans of various anatomical structures can obtain high accuracy segmentation results.

2.1.2 U-Net

The main contribution of a CNN architecture is to extract salient information of the image through the feature map. Therefore, in a CNN, the dimension of the image decreases, and it converts to a vector that contains the most distinguished information drawn from the image. This vector, then, is used for a classification task. However, this

architecture is not appropriate for the segmentation task since we need to reconstruct the image. Reconstructing an image from a vector is a super complicated task.

Thus, U-Net [25] is introduced to solve the mentioned problem. In U-Net, the expansion path reconstructs the image from the feature map vector. Also, the skip connections from the contraction path pass the local information of the image to the expansion path. Figure 2.2 illustrates a U-Net architecture.

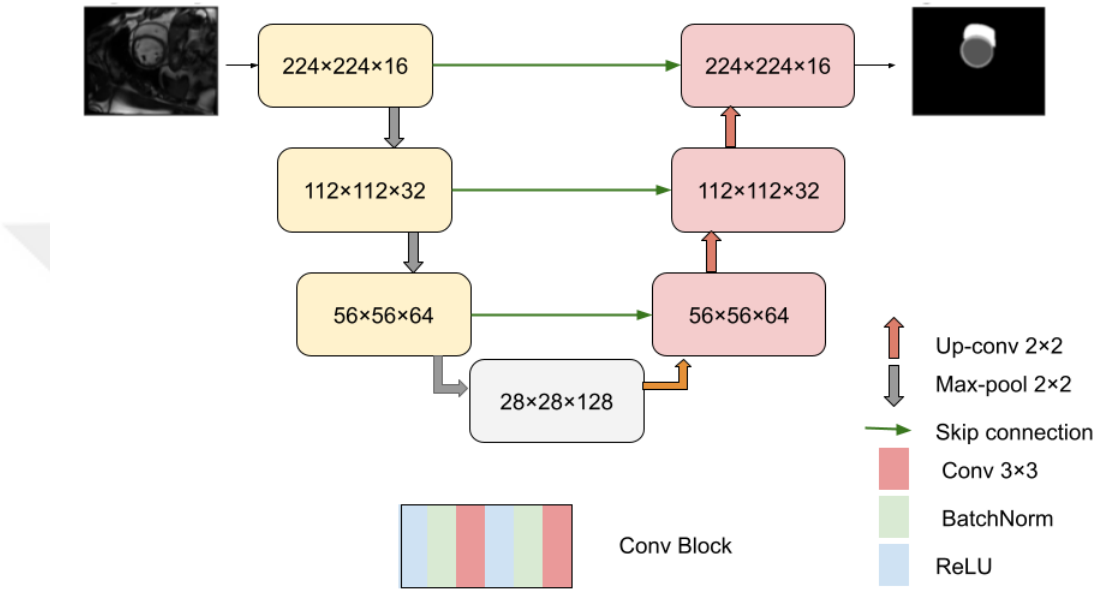

Figure 2.2 : U-Net Architecture is based on FCN, but it utilizes skip-connections (the green arrows) that convey the spatial information of each block in the contraction path to the expansion path.

The name is this architecture comes from its shape, which is like "U." It consists of three major parts, contraction (downsampling) path, bottleneck, and expansion (upsampling) path. The contraction path is made of several downsampling blocks. Each block in the downsampling path is composed of two 3x3 convolution layers followed by a 2x2 max-pooling layer. For the increase in learning complex structures of the image, the number of feature maps doubles after each contraction block. The same structure holds for the upsampling path, except using 2x2 up convolution instead of 2x2 max pooling. The vital part of this architecture that distinguished it compares to other architectures is the skip connections between upsampling and downsampling paths. These skip connections ensure that the feature maps learned through the downsampling path will be utilized by upsampling and reconstructing the image. Therefore, in each block in the upsampling path, the feature map of the

same level will append to the input from the previous block. [26] developed a fully automatic segmentation on CMR scan images using U-Net architecture. They trained the deep model on an extensive multivendor, multicenter cohort dataset. Therefore, the proposed model could achieve segmentation results on unseen data. [27] proposes an approach to segment the 3D atrial region of the heart from Gadolinium-enhanced CMR scans using U-Net architecture. The proposed network consists of two networks. The first one tries to locate the atrial center by the latent space information. While the second network segments the atrial region from a cropped sub-region achieved from the first network.

2.1.3 Residual network

In 2009, [28] theoretically shows the deeper network, the more accurate results, though the practical results didn't support that theory. When we have a very deep network, given more than 25 layers, the accuracy is likely to drop. [2] demonstrated the problem mentioned above, which wasn't overfitting, and proposed a solution to reduce vanishing gradient when the network is very deep. The solution was a residual connection or unity matrix, which connects the previous layer's output to the output of the new layer. Therefore, the back-propagation path is through the residual link, and the gradient will be multiplied by one. This process helps preserve information and avoid vanishing the gradient. More details of the structure of a residual block are shown in Figure 2.3 that how the identity matrix adds to the output layer before feeding into the ReLU layer.

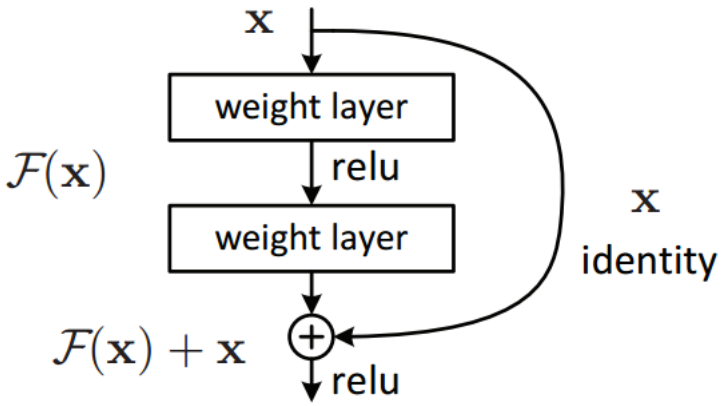

Figure 2.3 : Residual block [2]. Skip-connections or residual-connections help learn an identity function that is a solution for vanishing gradient in deep networks.

2.1.4 Residual U-Net

Residual U-Net is inspired by the combination of residual blocks and U-Net architecture. In the field of remote sensing image analysis, [3] first introduced Residual U-Net for the road extraction from high-resolution images. The effectiveness of their proposed architecture, then utilized in other domains like medical imaging. Thus, in medical image segmentation, we can utilize residual blocks in U-Net architecture. This modification allows the model to take advantage of deeper networks to learn complicated structures in the image without worrying about the vanishing gradient effect. Figure 2.4 shows the difference between plain neural block and residual block in a U-Net architecture.

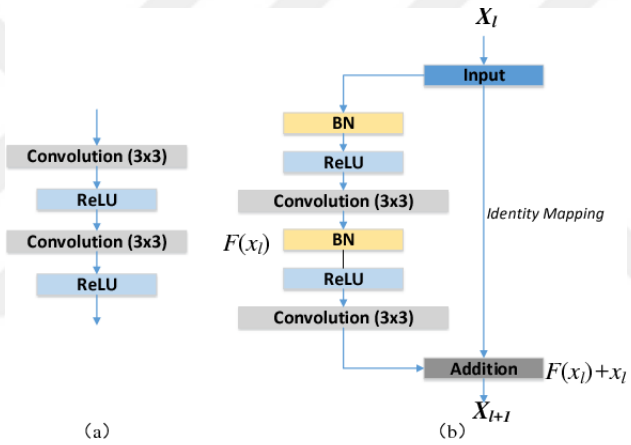

Figure 2.4 : (a) Plain neural unit block in U-Net architecture, (b) Residual unit in the proposed ResU-Net [3].

This architecture reduces the number of parameters like the vanilla U-Net architecture and allows a deeper neural layer in each block in both contraction and expansion paths. Figure 2.5 illustrates the residual U-Net architecture in detail.

2.1.5 Attention module

Attention is one of the most effective mechanisms in deep neural learning. This module, first, developed in the context of Neural Machine Translation. This approach was highly successful in deciphering long texts. This module works well in the context of deep learning because it tells the network where to look in the input data. Thus, the effectiveness of this method encouraged researchers to apply it to other domains like image recognition and medical imaging.

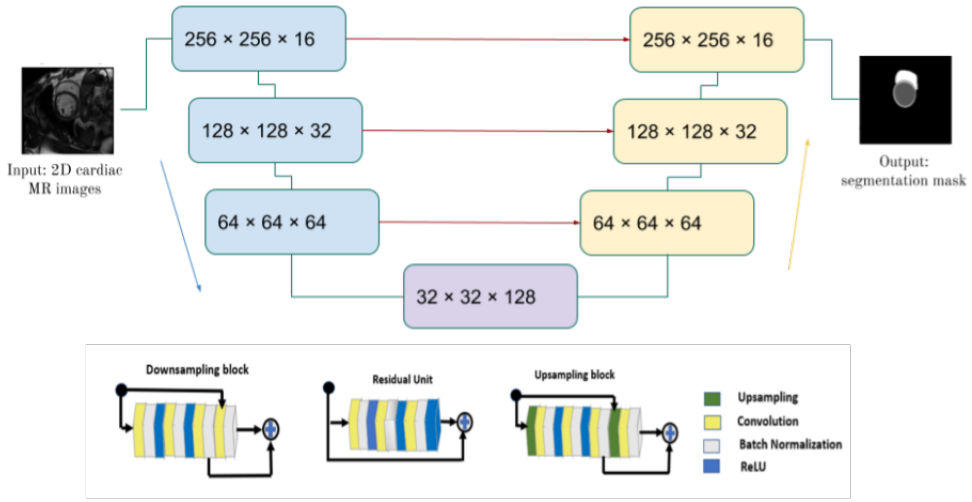

Figure 2.5 : Residual U-Net architecture. The architecture is based on U-Net while incorporated the residual-connections in each block to avoid vanishing gradient in deep networks.

In medical image segmentation, U-Net or Residual U-Net architectures perform reliable results. However, the repetitive use of low-level features extracted by those models burdens different computation processes on the resources. [4] introduced an attention gate in a standard U-Net architecture that allows the model to be specific in local regions. In this work, we implemented this module in Residual U-Net architecture. According to Figure 2.6(a) the attention gate adds just before the concatenation. Accordingly, the salient features from skip connection pass through to construct the image in the upsampling path.

Figure 2.6(b) illustrates the schematic of the attention gate. As shown in the figure, for having the same number of channels, both g and x inputs are fed into a 1x1 convolution layer. Then, They are summed and passed into a ReLU layer. The output of the Sigmoid function converts the output into 0-1 scale. Each tells us how much importance to give that part of the image.

One of the most famous extensions of fully connected networks (FCNs) is U-Net presented in [25]. The U-Net is an auto-encoder that takes advantage of skip connections between the encoder and decoder module. This modification allows the model to retrieve spatial context in the contraction path, it the expansion path. Taking advantage of this spatial context improves the accuracy of the segmentation. Many state-of-the-art methods employ U-Net as their architecture or an enhanced

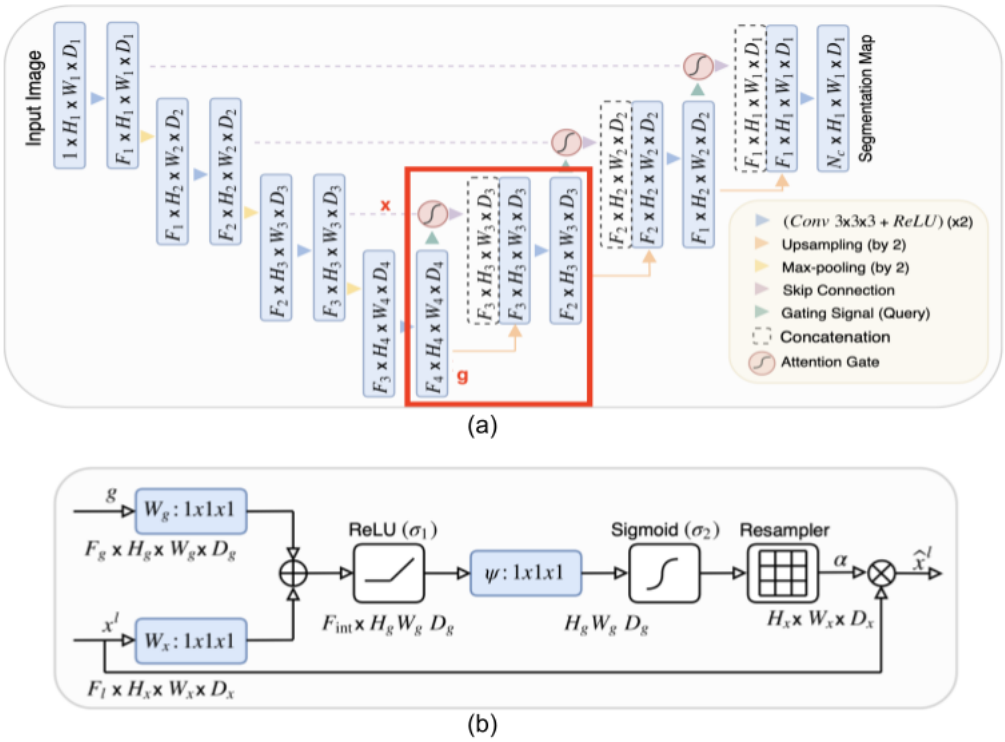

Figure 2.6 : (a) A block diagram of the attention module, Attention gates in this Figure filter the features transmitted through the skip connections. , (b) the schematic of the attention gate [4].

version of it. Despite the promising results from supervised methods, the lack of large datasets still hinders them from achieving reasonable accuracy and robustness. Therefore, several data augmentation approaches have been introduced. In order to better exploitation of the spatial information from the contraction path in a U-Net [29] introduces residual blocks instead of simple convolutional layers. Residual blocks facilitate information propagation and ease the training process. Also, [4] proposes a novel attention gate that assists the deep model where to look for the most relevant information. This attention gate can be easily integrated into U-Net architecture.

2.2 Augmentation Methods

Providing large labeled data is a challenging task. It is time-consuming and expensive, especially when it comes to medical image annotation. Insufficient training data cause overfitting. This phenomenon will deteriorate the predictions on test data. One of the most influential and straightforward approaches to cope with this issue is to produce artificial data using label-preserving transformations. [30] utilizes two forms of transformation that produce transformed data from original

data with the least computation. The first form translates the original image and the labeled image, while the second form tries to change the color intensity of the data. In [31], authors introduce an intelligent augmentation technique that learns how to generate augmented data during training to reduce network loss. The main goal of the network that they have introduced is to increase the model's regularization to handle overfitting. Finally, [32] proposes a novel augmentation technique named Fast Auto Augment that searches for effective augmentation policies using a density matching-based approach. For policy exploration, they apply the Bayesian Optimization technique, which significantly decreases the computation time. To tackle the labeled medical data scarcity, [33] applies a so-called mixup algorithm that uses linear combinations of the training images and the labels as an augmentation technique. However, the typical medical imaging augmentations consist of applying small transformations to the training images. Thus, they succeeded in enhancing the segmentation performance of the medical images using the proposed augmentation method [33]. Also, [34] introduces an automatic search-based algorithm that seeks optimal augmentation strategies. They present a formulation that allows them to apply alternative gradient-based methods on medical data.

2.3 Semi-Supervised Methods

Fully automatic segmentation of medical images requires a large number of annotated data. Though acquiring large labeled data is a significant challenge in medical image processing. First, the annotation of images is a tedious task. Also, for medical image annotation, we need medical experts, which makes the process highly expensive. Finally, labeled images from one imaging modality or protocol may not be compatible with another, so that they can't be transferable to another study. Many researchers aim to solve the problem by introducing semi-supervised methods. The goal of these methods is to fully utilize the unlabeled data to enhance the accuracy of the existing methods. [35] proposes a technique to combine a large number of weakly labeled data (bounding box) and few strongly labeled data. They develop an Expectation-Maximization approach for semantic image segmentation. [36] introduces a semi-supervised learning network for cardiac MR segmentation. In this approach, both labeled and unlabeled data are alternatively updated. However, one drawback

of their method is that the errors during initial segmentation will accumulate during training and will be learned. To address the shortage of annotated cardiac MR images and the variance of data belonged to different centers, [19] designs a method to take advantage of unlabeled cardiac MR sequence time-frame. They develop an algorithm based upon label propagation when several time frames are segmented while remaining unlabeled. This approach which applies an image registration technique successfully increases the number of labeled data. As explained earlier, domain shift in medical data is a common challenge. Therefore many researchers have recently concentrated on the domain adaptation issue. [37] presents a domain generalization model to segment multi-domain source data. They regularize the semantic structure of the latent space using a model-agnostic leaning by introducing two losses. To help their model accurately predict unseen domain data, they propose an optimization objective that leads the model to learn semantically consistent features across training domains [37]. [38] seeks another approach to tackle this domain shift challenge by presenting the Histogram Matching technique. Histogram matching as a data augmentation technique eliminates the domain gap between training data from different domains. They develop histogram matching to give the model to learn the intensity distribution of the test data from a different domain. They also showcase promising results that indicate the effectiveness of their method. And finally, [39] introduces a domain-adversarial learning approach to train a multi-domain U-Net architecture based on both labeled and unlabeled data. They show that an adversarial domain network improves the generalizability of a deep model.

2.3.1 Contributions

According to the literature reviewed above, the authors have proposed different deep architectures in the fully supervised section. Also, many researchers introduced various data augmentation techniques to prevent overfitting during training for lack of enough annotated data. Although those methods improved the accuracy of the segmentation predictions, the results still need further improvements. Especially in cardiac MR scans, because of variance in the intensity distribution of scans from different vendors and centers, we need paradigms to address those conditions. Therefore, fully supervised segmentation methods may fail on unseen datasets. As a result, augmentation techniques proposed alleviating the lack of annotated medical

data by introducing image transformations to both original images and correspondence label masks. In addition, semi-supervised segmentation methods emerged to increase medical segmentation accuracy by modeling more variations in data. From our point of view, medical image segmentation accuracy can still be improved.

In this thesis, first, we implement some of the mentioned methods introduced in this chapter to compare the effectiveness of each on cardiac MR scans. Also, we consider scenarios that combine them to investigate the impact of their combination on the CMR images. For example, we add the attention module to the U-Net and Residual U-Net architecture to enhance the segmentation accuracy. In addition, our proposed semi-supervised method combined with histogram matching (augmentation technique) increases the segmentation accuracy and model generalizability.

3. SUPERVISED MEDICAL IMAGE SEGMENTATION

After the introduction of Alex-Net by [30] for image classification and the advancements in the capacity of GPUs, deep learning methods for image recognition and classification have been dominated. Also, many researchers have been applying those methods to medical images [1, 21, 22, 33].

In this chapter, we describe the application of deep learning in cardiac MR image segmentation. We utilize two architectures that we explained in **Chapter 2** and compare the performance of each to check which one is suitable for CMR segmentation. In addition, we introduce the loss functions and the evaluation metrics that we have been utilized in this research. The detailed implementation will be discussing step by step. Finally, we expand the results and analysis them in detail.

3.1 Deep Learning Architectures for CMR Scans

This section discusses two state-of-the-art deep learning architectures that are dominant in medical image segmentation. We start with U-net architecture. Next, we meticulously explain the Residual U-Net.

3.1.1 U-Net

In this section, we explain the U-Net architecture that we use for CMR segmentation. We considered five blocks in both contraction and expansion paths. Each block in the downsampling path is composed of a 3x3 convolution layer followed by a 2x2 max-pooling layer. To overcome overfitting, we applied a Dropout layer in the last two blocks. The numbers of filters start with 16 and doubles after each contraction block for the increase in learning complex structures of the image. The same structure holds for the upsampling path, except using 2x2 up convolution instead of 2x2 max pooling and no dropout layer. We implemented the U-Net architecture both in Keras and PyTorch environments.

3.1.2 Residual U-Net

Based on the explanation of Residual U-Net in ***Section 2.1.4***, we implemented this architecture like the U-Net architecture except that we added the skip connection to each block in both contraction and expansion paths. [40] applied Residual U-Net for each slice in the short-axis view segmentation.

3.2 Loss Function and Evaluation Metrics

This section explains the loss functions and evaluation metrics that have been used in this research. To train a neural network for segmentation, we must identify a dataset consisting of paired images and annotation masks for both training and testing. In addition, it is crucial to separate training and testing sets before the beginning of the preprocessing to avoid picking phenomena. Another required key point to define for training a neural network is an optimizer such as Adam, Stochastic gradient descent, etc. Finally, a loss function is a must to update the model parameters. The loss function compares the prediction results with the ground truth masks and calculates the error in each iteration. This error with the optimizer updates the learnable weights through backpropagation. Thus, we aim to minimize the loss function by finding proper values for the learnable weights [10].

3.2.1 Loss function

In deep learning methods, loss function plays an essential role. In this research, we have compared two loss functions, cross-entropy loss and dice loss.

3.2.1.1 Categorical cross entropy loss function

The most commonly used loss function in the deep learning approaches is Cross-Entropy loss. This loss function is based on pixel-wise examination. It compares the probability of the Sigmoid function with the one-hot-encoded label samples in order to achieve segmentation results. Based on the definition, it predicts each pixel

and then averages over all pixels. The Cross-Entropy loss is defined as equation 3.1 [40]:

$$CE = L(\hat{y}, y) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}(y_i^c log(\hat{y}_i^c)) \quad \textbf{(3.1)}$$

Where N is the number of samples, $\hat{y}$ is the predicted value of the segmentation task, and y is the ground truth value. The segmentation task we aimed was multi-class segmentation, so that we used categorical cross-entropy loss instead of the binary one. Although applying cross-entropy loss function for segmentation task on balanced labeled distributions results in reasonable segmentation masks, it performs considerably poorly when dealing with unbalanced labeled distribution.

3.2.1.2 Dice loss

Dice loss has developed based upon the dice coefficient that measures the similarities between two samples. This objective function was first introduced by [41] to segment 3D medical images. Dice loss can alleviate the strong imbalance between foreground and background pixels in medical images, especially in cardiac MRI images that we are concentrating on them. Equation 3.2 illustrates detail implementation of this loss function.

$$DiceLoss = 1 - \frac{2\sum_{i=1}^{N}\sum_{c=1}^{C}(y_i^c \hat{y}_i^c)}{\sum_{i=1}^{N}\sum_{c=1}^{C}(y_i^c + \hat{y}_i^c)} \quad \textbf{(3.2)}$$

Where $\hat{y}$ is the predicted value of the segmentation task, and y is the ground truth value for each class C.

3.2.2 Segmentation evaluation metrics

After a deep neural network trains a model and has a prediction, we need an evaluation metric to check how well the trained model works. This research utilizes the dice coefficient metric or dice score as our evaluation metric for the segmentation results. According to Figure 3.1, dice score measures the overlap between the predicted mask and the ground truth mask. Therefore, in a good segmentation, we expect to have maximum overlap, which means a dice score of one is the most desired. Also, equation 3.3 shows more detail about dice score implementation.

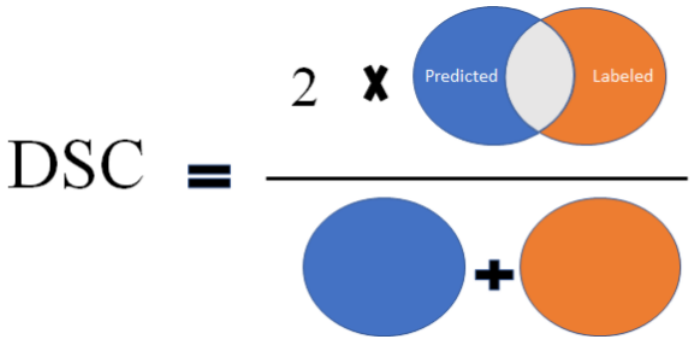

Figure 3.1 : Dice Score is an evaluation metric suitable for image segmentation. The blue region is the predicted mask, while the orange area regards the ground truth mask. DSC defines as the intersection of the two regions multiply by two over the union of them.

$$DC = \frac{2|A \bigcap B|}{|A| + |B|} \tag{3.3}$$

3.3 Experimental Setup

We trained our networks using the Keras library on NVIDIA Quadro GV100 GPU. The training took 3 hours and 48 minutes (28 seconds for each epoch) for 300 epochs for the Residual U-Net architecture. However, the U-Net architecture was significantly slower, taking 5 hours (33 seconds for each epoch.)

3.3.1 Parameter setups

One of the key parameters for training a deep model is the learning rate. It controls the amount of weight adjustment of our training model regarding the loss gradient. If we choose the learning rate too slow, it leads to slow convergence of our model. Especially in a plateau region, we might get stuck because of choosing a too low learning rate. Typically, researchers configure the learning rate based on past experiences. The Figure 3.2 demonstrates the impact of different selected values on the learning rate [42].

A deep learning algorithm may stick in the local minimum during the optimization instead of converging to the global minimum. One way to cope with the mentioned issue is to use the adaptive learning rate. We can start with a relatively high learning rate and then gradually decrease it. At the same time, another approach could be using Adam optimizer [43], which is a combination of RMSprop and Stochastic Gradient

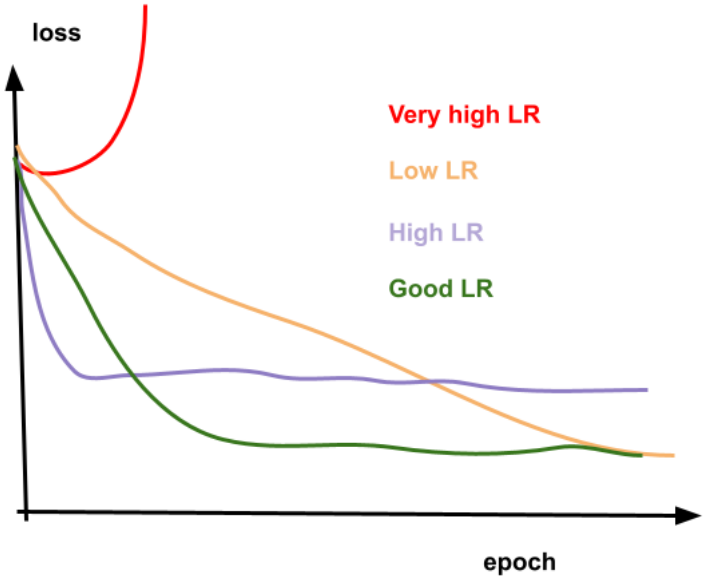

Figure 3.2 : Different scenarios regarding the selection of learning rate values in convergence of a deep model.

Descent with Momentum. This optimizer is known as an adaptive learning rate method since, for different parameters, it calculates individual learning rates. Therefore, we use Adam optimizer in our implementation.

Furthermore, another critical parameter is the number of epochs. If we try to train a neural network with few epochs, we face underfitting phenomena, which means the model cannot learn enough. Therefore, the accuracies regarding both training and testing set are low. On the other hand, training a neural network for large epochs causes overfitting, which means the model memorizes the training samples' features. Thus, the model achieves a high training accuracy, while the model performs poorly on the test set. As a result, we need to determine the epoch numbers to overcome over or under-fitting. Applying an early stopping approach [44], we can find a balance between the achieved accuracy between training and test sets. Figure 3.3 shows the importance of early stopping. When the accuracy of the training set increases in contrast to the decrease of the test set accuracy, we need a trade-off to balance the achieved accuracy.

3.3.2 Datasets

We used two datasets for our analysis, the STACOM LVQuan 18 challenge dataset, which for simplicity, we will call STACOM, and M&Ms 2020 challenge dataset. We will mention it as MMS from now on. The former dataset is released by [45] as a

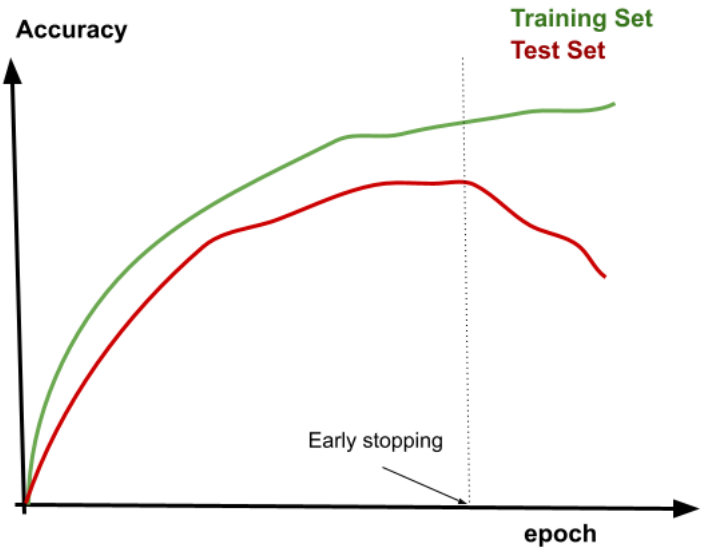

Figure 3.3 : The relation between test and train accuracies according to the number of epochs. It demonstrates the importance of early stopping.

challenge dataset that focuses on a full quantification of the left ventricle of the heart (LV). This dataset consists of cardiac MRI sequences of 145 patients. The annotated masks delineate the left ventricle and myocardium regions. Therefore, we need a three-class model. On the other hand, the organizers of M&Ms 2020 challenge [46] released our second dataset, which contains CMR sequences of 150 patients from two different vendors (75 samples from each). The scans have annotations of the left ventricle, myocardium, and the right ventricle regions of the heart for end-diastole (ED) and end-systole (ES) sequences. The aim of this challenge is to leverage the unlabeled data from different vendors to introduce generalizability to the models. In addition, they provided unlabeled CMR sequences of 25 patients from a third vendor. This dataset is more challenging and demanding a robust deep model to handle multi-vendor and multi-center samples while procuring unlabeled data as extra data to improve the performance of our model. Figure 3.4(a) illustrates a 2D sample of the STACOM dataset and the ground truth image, and Figure 3.4(b) shows a sample image and its ground truth mask of the MMS dataset.

3.3.3 Detail implementation

We implemented our model in a 2D fashion so that we needed to load the input data slice by slice. We only have annotated images of the MMS dataset of end-systole

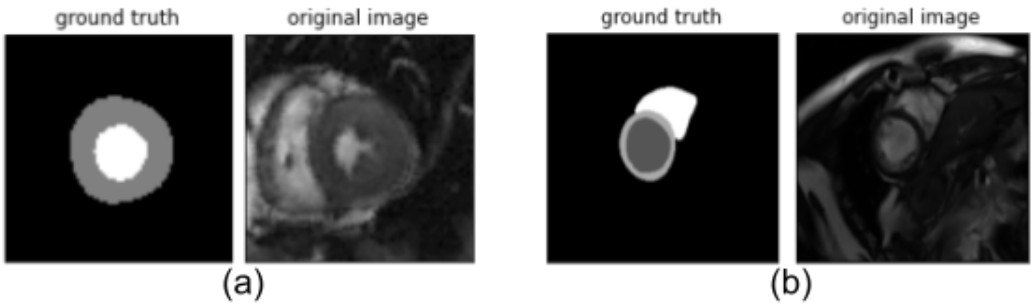

Figure 3.4 : (a) a sample 2D image of STACOM dataset and the annotated mask, (b) sample 2D image of MMS dataset of vendor A and the annotated mask.

(ES) and end-diastole (ED), so that we only use those slices as labeled samples. In the STACOM dataset, we divided 2620 2D labeled images into 2340 training and 280 testing images while considering 20% of the training set for validation. This dataset is clean, and all original images and the labeled images are the same sizes so that we have not done any preprocessing process on this dataset. We split the total 2439 2D annotated samples from two different vendors into 1711 training, 428 validation (20% of the training data), and 300 test set for the MMS dataset. To avoid having the same sequence that belongs to a specific patient in both training and testing sets during splitting data, we consider the numbers of 2D images of each patient (z-axis and time-axis slides are different) to assure correct splitting. This splitting process is essential to avoid over-fitting during training. In addition, we used 478 2D images of the unlabeled data from vendor C of the MMS dataset. Because of inconsistency during data acquisition, we need to apply preprocessing approach to ensure the model receives uniform input data. We cropped the images with 224×224 pixels using the center point of each slice. Also, we normalized the images by subtracting the minimum pixel intensity value from each pixel in the image and then dividing by the maximum pixel intensity value of the image. We have done this normalization process in both the training and test sets separately to avoid the picking phenomenon. Finally, We choose an ADAM optimizer with an initialization of the learning rate equal to 0.001.

3.3.4 Augmentation

A large number of training samples is crucial to enhance the performance of a deep model. Using data augmentation can alleviate the lack of large number of labeled data by bringing variations to the training dataset, which can significantly improve the model's performance on unseen samples. We have applied several

augmentation techniques in this work, including rotation of -45 to 45-degree, -90 to 90-degree rotation, horizontally flipping with the rate 50% of the images, and elastic transformations. Figure 3.5 illustrate two examples of the results of augmentation techniques that we have done for both datasets.

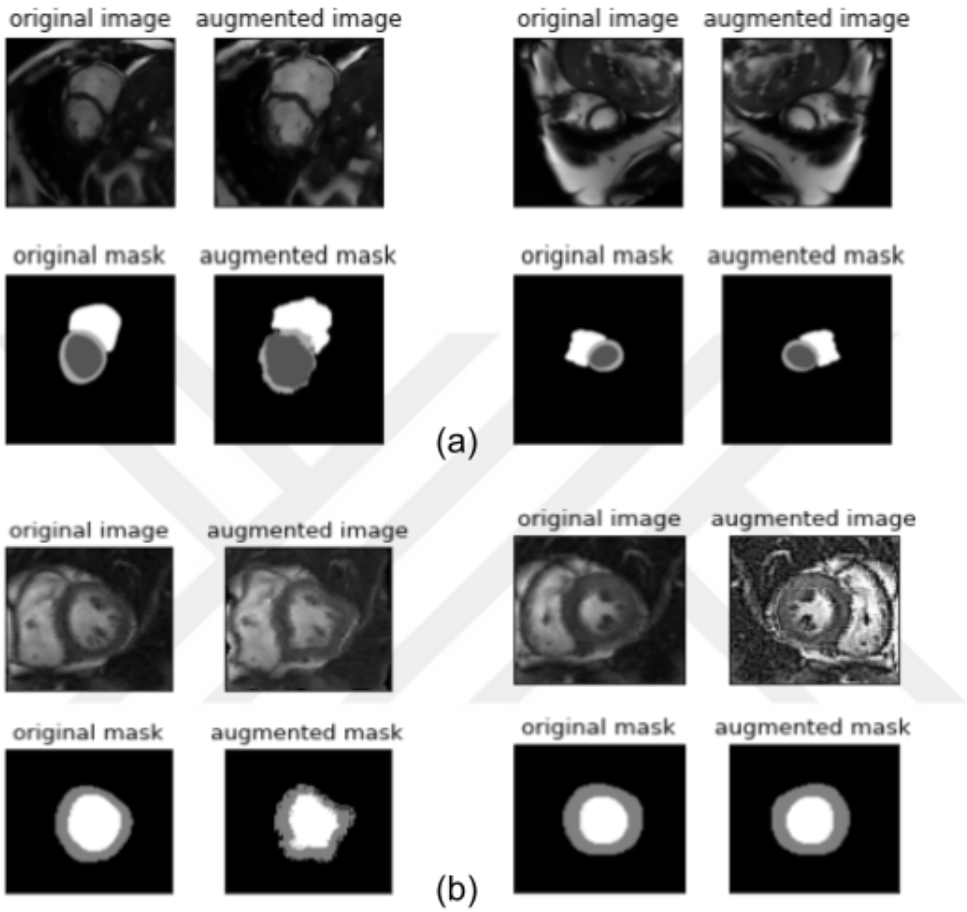

Figure 3.5 : (a) two sample images from MMS dataset and the augmented images , (b) two sample images from STACOM dataset and the augmented images.

3.4 Results and Analysis for Supervised Method

We train different fully supervised models to compare them and achieve the best hyperparameters. We utilized both categorical cross-entropy loss function and dice loss. Also, we train our model with and without augmentation techniques. This approach allows us to evaluate the effectiveness of each of the parameters.

3.4.1 Architecture comparison

We report the average and standard deviation dice score (DSC $\pm SD$) of all regions of interest, LV, RV, and Myocardium of the heart in the STACOM dataset in Table 3.1 considering categorical cross-entropy loss function. We calculated the standard deviation by applying 5-fold cross validation process. Also, we report the dice scores of all regions of interest, LV, RV, and Myocardium of the heart, according to the MMS dataset in Table 3.2.

Table 3.1 : The average DSC($\pm SD$) results of fully supervised method using CE loss on STACOM dataset. The performance comparison between U-Net and Residual U-Net while considering learning process with and without data augmentation. The results contain the DCS on train-set, validation-set, and test-set.

DSC	U-Net		Res-U-Net	
	no-aug	with-aug	no-aug	with-aug
Tr-LV	0.784±.026	0.778±.031	0.985 ±.021	0.983±.03
Val-LV	0.768±.041	0.772±.043	0.983±.032	0.985±.039
Tst-LV	0.748±.044	0.761±.052	0.941±.053	**0.957**±.041
Tr-Myo	0.842±.012	0.868±.019	0.992±.022	0.994±.023
Val-Myo	0.802±.035	0.832±.029	0.962±.038	0.962±.027
Tst-Myo	0.803±.067	0.815±.056	0.923±.043	**0.942**±.052

Table 3.2 : The average DSC($\pm SD$) results of fully supervised method using CE loss on MMS dataset. The performance comparison between U-Net and Residual U-Net while considering learning process with and without data augmentation.The results contain the DCS on train-set, validation-set, and test-set.

DSC	U-Net		Res-U-Net	
	no-aug	with-aug	no-aug	with-aug
Tr-LV	0.781±.037	0.769±.031	0.921±.029	0.914±.025
Val-LV	0.725±.048	0.751±.053	0.82±.032	0.878±.039
Tst-LV	0.702±.064	0.711±.072	0.795±.053	**0.801**±.061
Tr-RV	0.74±.043	0.742±.057	0.866±.04	0.872±.048
Val-RV	0.656±.062	0.674±.059	0.705±.074	0.791±.069
Tst-RV	0.623±.083	0.661±.081	0.683±.064	**0.776**±.078
Tr-Myo	0.787±.044	0.766±.027	0.747±.038	0.761±.036
Val-Myo	0.697±.067	0.721±.081	0.703±.071	0.73±.063
Tst-Myo	0.614±.077	0.622±.082	0.638±.087	**0.648**±.076

Considering Table 3.1 and Table 3.2, we can compare the performance of U-Net and Residual U-Net. In addition, we compare the effectiveness of the augmentation

techniques that we have chosen for each architecture. Based on the results, data augmentation improves the outcomes in all regions of interest in both architectures. Therefore, we will utilize the data augmentation techniques for the rest of this thesis. Also, the results indicate the superiority of Residual U-Net architecture over vanilla U-Net. The Residual U-Net with data augmentation techniques outperforms other scenarios. The bold numbers in the tables emphasize the improvements of the dice score in the test samples compare with other scenarios.

3.4.2 Loss function comparison

Considering **Section 3.2.1**, loss function is a significant factor in the learning process. Therefore, we compare the performance of our model when we use Dice loss and CE loss. Table 3.3 concentrates on the comparison between the performance of Dice loss and CE loss on Residual U-Net architecture considering data augmentation. In Table 3.3, we report the average and standard deviation of dice scores (DSC$\pm$*SD*) based on cross entropy loss and Dice loss functions, respectively. Also, it is important to mention that the evaluation metric for the loss function comparison is dice coefficient. Therefore, the values in the Table 3.3 are dice score while we either using CE loss function or Dice loss. We do not provide the values of loss function here.

Table 3.3 : Comparison of CE loss and Dice loss on Residual U-Net architecture while using data augmentation techniques.

DSC(SD)	CE Loss	Dice Loss
Tr-LV	0.914$\pm$.025	**0.938**$\pm$.018
Val-LV	0.801$\pm$.061	**0.825**$\pm$.033
Tr-RV	0.872$\pm$.048	**0.93**$\pm$.021
Val-RV	0.792$\pm$.069	**0.810**$\pm$.035
Tr-Myo	0.761$\pm$.036	**0.893**$\pm$.02
Val-Myo	0.648$\pm$.063	**0.706**$\pm$.041

According to the results in Table 3.3, we can conclude that the model using dice loss outperforms the network utilizing categorical cross-entropy loss function. Therefore, we will use a supervised model with Residual U-Net architecture employing dice loss function as the objective function.

To showcase the results in a fully supervised manner, Figure 3.6 illustrates the best and worst predictions with our final scenario. Figure 3.6 (a) shows two best-predicted samples of the STACOM dataset, while Figure 3.6(b) illustrates the two worst

predicted samples of the STACOM dataset. Figure 3.6 (c) and (d) indicate the best and worst predictions, respectively.

Looking at the images in Figure 3.6, we can observe that segmentation accuracy for the MMS dataset, which is multi-vendor, multi-center, and need to segment RV region too compare to STACOM dataset, still requires enhancements.

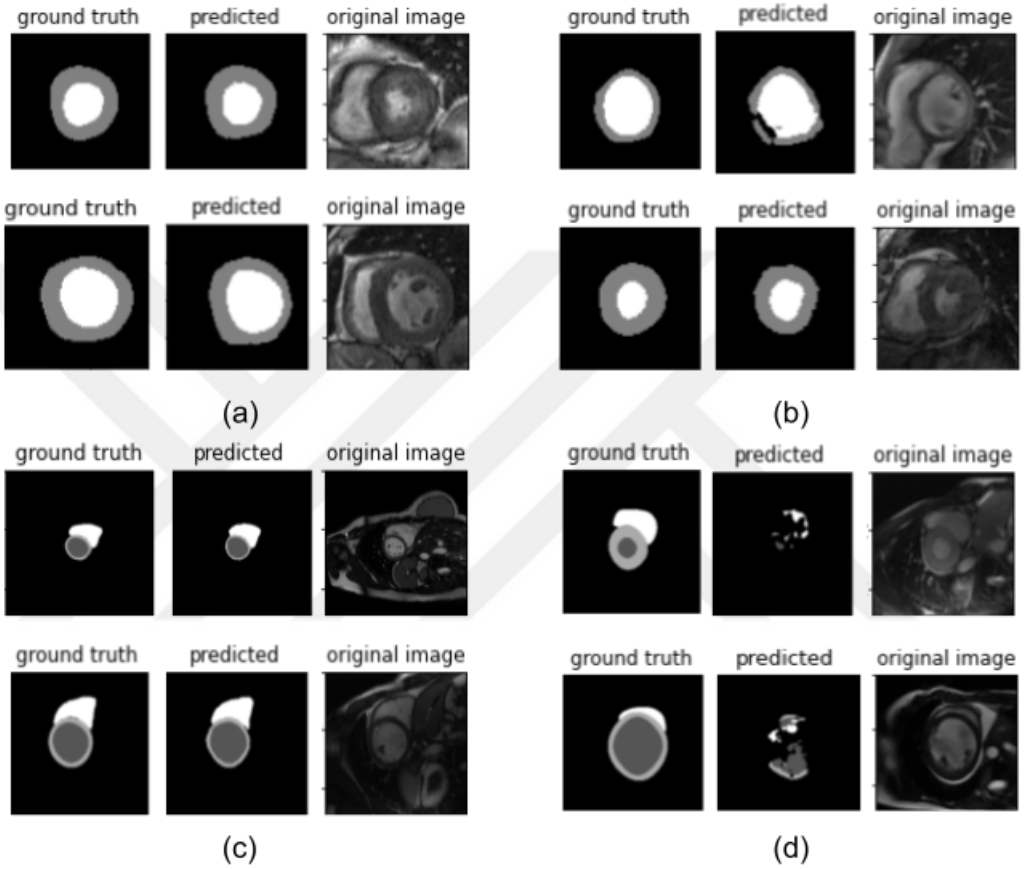

Figure 3.6 : Fully supervised learning using Residual U-Net considering data augmentation. (a) Best results for STACOM dataset , (b) Worst results for STACOM dataset, (c) Best results for MMS dataset, (d) worst results for MMS dataset.

4. SEMI-SUPERVISED MEDICAL IMAGE SEGMENTATION

In order to train a deep network using a fully supervised approach, we need a large annotated dataset. In addition, manual annotation of medical images is a tedious task and requires professional clinicians for the labeling task, which makes the task expensive. Although the data augmentation technique that we applied in **Section 3.3.3** slightly alleviates the problem, we still need to address distribution shifts in medical images. We know that distribution shifts are standard in medical images when they come from different centers or acquire different equipment (like Siemens, GE, or Philips) or take images with various acquisition parameters. The standard data augmentation that we have done in **Chapter 3** cannot resolve the domain shift in various datasets. Therefore, a supervised model may fail on unseen datasets. On the other hand, there are huge unlabeled data from different domains. A semi-supervised method tries to leverage the unlabeled data to enhance the performance of a deep model on unseen data. We proposed a simple yet effective semi-supervised method to improve the robustness of the model and utilize the unlabeled data.

4.1 Proposed Semi-Supervised Method

In the MMS dataset, we have access to 25 unlabeled sequence samples. In this chapter, we will work on the MMS dataset because of the available unlabeled data. As described earlier in **Chapter 3**, the MMS dataset contains data from 3 vendors. Vendors A and B are annotated by expert clinicians, 75 samples each, and the third vendor C is unlabeled and includes 25 patients. We aim to leverage the unlabeled data from vendor C to enhance the generalizability of our model.

We start with our final supervised model, applying Residual U-Net architecture considering dice loss with data augmentation. First, using the trained, supervised model, we predict the segmentation mask for the unlabeled data on ES and ED time frames. To avoid using low-quality annotations, we evaluate the prediction throughout a loop to check if those predictions are noisy and disregard them by the trained,

supervised model. In other words, we study the quality of the predictions in an iterative manner and remove the low-quality annotations. Then, two datasets, the manually labeled data and the pseudo labeled data, are merged. The detailed implementation of the method is shown in Figure 4.1. Using this method allows the model to train on images from different vendors or centers.

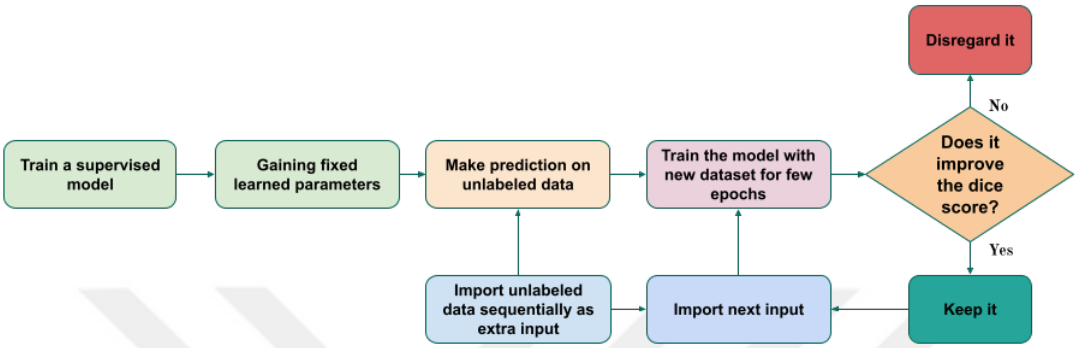

Figure 4.1 : Proposed Semi-Supervised Diagram.

4.2 Proposed SS Method Considering Histogram Matching

Even though the proposed semi-supervised method in the previous section improves the robustness of our model, the noisy predictions can deteriorate the performance, and the pseudo labels have heavily relied on the fully supervised model. To tackle this issue, we introduce the histogram matching method as an augmentation method. This approach considerably improves the performance of the model on unseen data.

4.2.1 Histogram matching

One of the most common approaches for gaining a processed image using the histogram of the images is known as histogram matching [38]. This method perturbs the intensity distribution of an image considering a reference image. The matched image is like the first image with the intensity distribution of the second image. To explain the method in more detail, we need two images, a source image, and a reference image. In our case, we can choose the source image from either vendor A or B. The reason for this selection is that we don't want to change the image in such a way as to distort the segmentation mask. This means we want intact annotation masks. On the other hand, we can choose the reference image from vendor C. In this way, the labeled images from vendors A and B have an opportunity to mimic the intensity distribution

of vendor C, the unlabeled images. Therefore, we will have a training set with different variations in the intensity distributions using this augmentation technique. As a result, the trained model on this enhanced dataset will achieve robust predictions. Figure 4.2 illustrates a matched image using histogram matching. In this Figure, the source image is a 2D sample image from vendor B, and the reference image is a 2D sample image from vendor C. We implement histogram matching using Scikit-image, a library in Python.

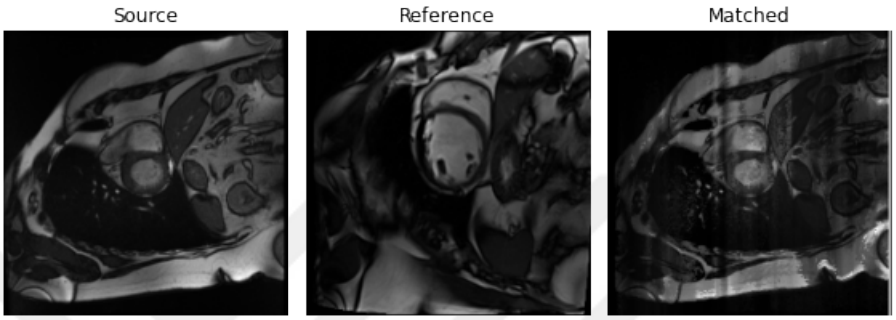

Figure 4.2 : Histogram matching result. From the left to right respectively: the source image, reference image, the matched image.

Now that we know about histogram matching and how it affects the images. We apply the histogram matching technique on both datasets from vendor A and B to mimic the intensity distribution of the dataset from vendor C. Thus, we have a more diverse training dataset to train a deep model. This time, the pseudo labels are satisfactory. As a result, we could overcome the low-quality predictions on unlabeled datasets based on the fully supervised model using histogram matching. Hence, the proposed semi-supervised model predictions significantly improve according to the dice scores. The rest of the method is the same as we described earlier in this section, according to Figure 4.1.

4.3 Results and Analysis for Semi-Supervised Method

Table 4.1 shows the comparison between the performance of the fully supervised model and the initial semi-supervised model. In Table 4.1, we report the average and standard deviation of dice scores (DSC$\pm SD$) based on fully supervised and semi-supervised methods, respectively.

Table 4.1 : Comparison of fully supervised and semi-supervised learning methods based on MMS dataset. FS indicates of fully supervised considering Res-U-Net with data augmentation. While, SS defines the proposed semi-supervised method. The (DSC$\pm$*SD*) results are provided.

DSC	FS	SS
Tr-LV	0.938$\pm$.018	**0.939**$\pm$.046
Val-LV	0.825$\pm$.033	**0.897**$\pm$.052
Tr-RV	0.93$\pm$.021	**0.941**$\pm$.038
Val-RV	0.810$\pm$.035	**0.813**$\pm$.053
Tr-Myo	0.893$\pm$.02	**0.898**$\pm$.029
Val-Myo	**0.706**$\pm$.041	0.675$\pm$.048

As stated by Table 4.1, the proposed semi-supervised method improves the dice scores in all regions in both validation and training sets, except on validation of the Myocardium region. This result shows that the model needs modification to be robust on unseen data. Figure 4.3 shows the differences between the predictions on image samples when we apply histogram matching (FSH) and when we use the fully supervised (FS) method.

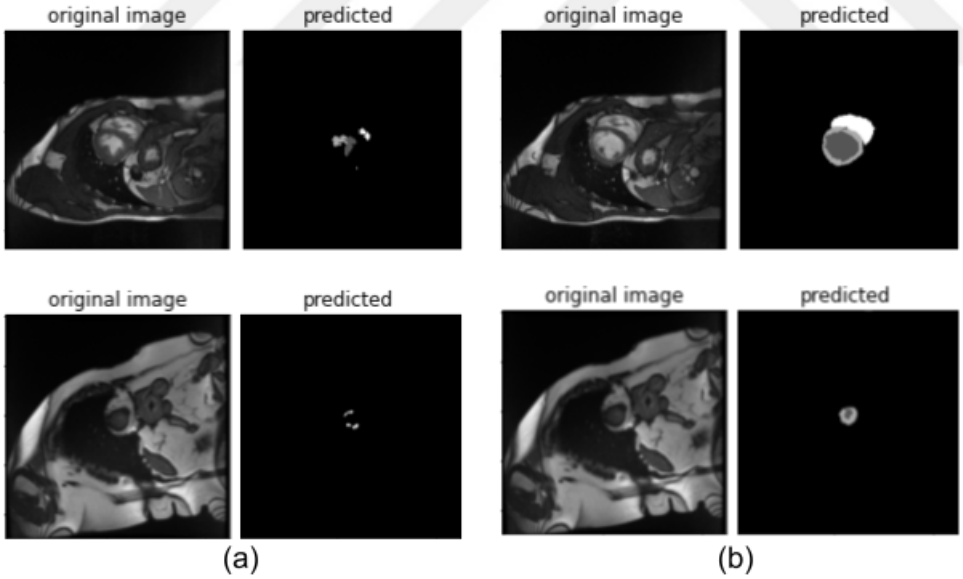

Figure 4.3 : predicted masks for an image in vendor C (a) using FS model, (b) using FSH model. The comparison between images (a) and (b) indicates of a remarkable improvement in the generalizability of our model.

According to the Figure 4.3, histogram matching remarkably improves the predictions on unseen data. In Figure 4.3(a) shows the prediction of the supervised method (FS) on the unlabeled unseen dataset (vendor C). The sample results are not satisfactory. While

Figure 4.3(b) illustrate the prediction of the supervised model after using histogram matching technique (FSH). The sample predicted masks indicate the improvements on the performance of our model on the vendor C dataset.

Table 4.2 and Table 4.3 explain the prediction dice scores on training and test sets, respectively. In this table, we consider eight scenarios. Four of them are related to the fully supervised methods, while the other four scenarios concern the semi-supervised techniques. The first row of the table shows the results of the fully supervised model, but the second row illustrates what happens if we only use 50% of the labeled data. The third and fourth rows explain the effects of applying histogram matching while using either 100% or 50% of the labeled data. On the other hand, the four last rows show the results of the semi-supervised method either using 100% or 50% labeled data. Also, the previous two rows illustrate how the histogram matching affects the results.

Table 4.2 : The training (DSC$\pm$*SD*) results of semi-supervised method for labeled and unlabeled MMS data. Scenarios: fully supervised (FS) considering 100% of labeled data, FS using 50% of labeled data, FS with histogram matching (FSH) using 100% of labeled data, and FSH50 using 50% of labeled data. The last four rows are the same but for semi-supervised mehtod.

DSC	Tr-LV	Tr-RV	Tr-Myo
FS	0.938$\pm$.018	0.93$\pm$.021	0.893$\pm$.02
FS50	0.913$\pm$.023	0.901$\pm$.017	0.879$\pm$.026
FSH	0.963$\pm$.022	0.951$\pm$.019	0.904$\pm$.021
FSH50	0.941$\pm$.018	0.936$\pm$.026	0.911$\pm$.026
SS	0.966$\pm$.046	0.96$\pm$.038	0.899$\pm$.029
SS50	0.950$\pm$.054	0.941$\pm$.058	0.912$\pm$.049
SSH	**0.971**$\pm$.02	**0.977**$\pm$.031	**0.946**$\pm$.027
SSH50	0.951$\pm$.044	0.968$\pm$.051	0.944$\pm$.038

According to the dice scores from Table 4.2 and 4.3, the semi-supervised scenarios outperform the fully supervised approaches. Also, we can observe that the semi-supervised method with histogram matching gives the best results compare to other scenarios.

Figure 4.4 explains one of the best and worst results from the fully supervised method and semi-supervised method, respectively.

Table 4.3 : The test (DSC±*SD*) results of semi-supervised method for labeled and unlabeled MMS data. Scenarios: fully supervised (FS) considering 100% of labeled data, FS using 50% of labeled data, FS with histogram matching (FSH) using 100% of labeled data, and FSH50 using 50% of labeled data. The last four rows are the same but for semi-supervised mehtod.

DSC	Val-LV	Val-RV	Val-Myo
FS	0.779±.033	0.810±.035	0.706±.041
FS50	0.77±.051	0.787±.039	0.690±.067
FSH	0.782±.034	0.793±.03	0.714±.054
FSH50	0.773±.043	0.733±.032	0.714±.061
SS	0.873±.052	0.892±.053	0.778±.048
SS50	0.801±.058	0.821±.056	0.765±.063
SSH	**0.921**±.32	**0.926**±.039	**0.891**±.034
SSH50	0.851±.04	0.858±.043	0.839±.037

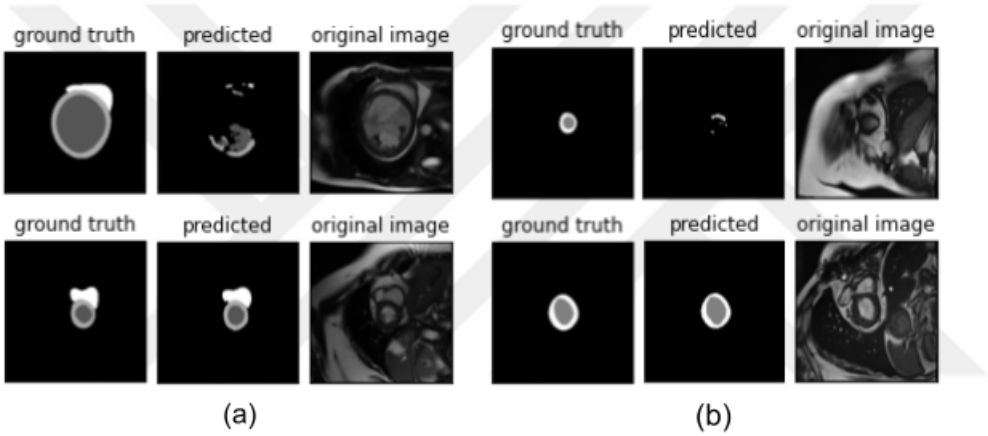

Figure 4.4 : Semi-supervised method results considering labeled and unlabeled MMS data. first row: worst results, and second row: best results. (a) FS results, (b) SS results.

4.3.1 Results for the model with attention module

Finally, we append the attention module that we explain in chapter 3 to our central architecture. The results compare to our best results from the semi-supervised method are shown in Table 4.4. In the Table, the Baseline method means our enhanced proposed semi-supervised method. At the same time, the attention module indicates in appending the attention module to the Residual U-Net architecture of the baseline model.

Figure 4.5 illustrates a sample example of the results of each scenario. Figure 4.5(a) shows a sample result with adding the attention module, Figure 4.5(b) indicates a sample result without attention module.

Table 4.4 : Comparison of adding attention module to the Baseline model. We consider the semi-supervised method with histogram matching as the baseline.

DSC	Baseline	attention module
Tr-LV	0.971±.02	**0.973**±.011
Val-LV	0.921±.032	**0.926**±.027
Tr-RV	0.977±.031	**0.983**±.028
Val-RV	0.926 ±.039	**0.933**±.034
Tr-Myo	0.946±.027	**0.948**±.021
Val-Myo	0.891±.034	0.892±.033

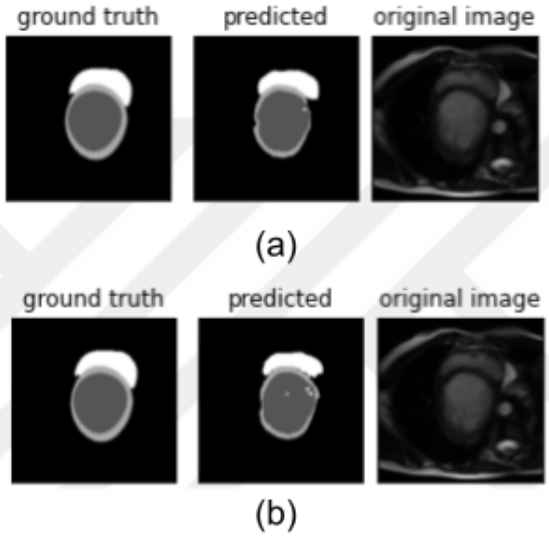

Figure 4.5 : The results (a) with attention module, (b) without attention module.

According to Table 4.3 and Figure 4.5, we can conclude that appending the attention module can slightly improve the performance of the model. Especially it enhances the dice score related to the RV regions.

In order to make a comparison between our proposed method with the state-of-the-art method, we compare our results with two baseline methods which were accepted in M&Ms challenge 2020, [19], and [38] respectively. Even though, parameters and hyper-parameter tuning, data splitting, and other factors may not be compatible. The comparison results are shown in the Table 4.5.

Based on the results shown in Table 4.5, our proposed method achieves better dice score compare to other methods.

Table 4.5 : Comparison of our proposed method with two accepted methods on M&Ms challenge 2020 on validation data

DSC	Our method	[19]	[38]
LV	**0.926**±.027	0.917±.07	0.9166 ±.019
RV	**0.933**±.034	0.893±.062	0.893 ±.024
Myo	**0.892**±.033	0.865 ±.042	0.866 ±.017

5. CONCLUSION

In this chapter, we first overview the main questions regarding our research and explain our approach to reach the answers. Then, we summarize the contributions of this thesis in detail. Finally, we discuss the practical future work related to the current research.

5.1 Answers

We faced several questions when we started the research to try to find answers for questions step by step. The answers of the raised questions in the **Chapter 1** are listed below:

- **Question 1** How different deep learning architectures perform in a fully supervised manner?

  Our answer to this question is described in **Section 3.1**, where we compare two architectures on two different datasets. We have chosen U-Net and Residual U-Net as our selected architectures. According to the results, Residual U-Net outperformed the standard U-Net.

- **Question 2** How much a deep model is sensitive to choosing the right loss function for the segmentation task?

  We compare the two most common loss functions in medical image segmentation in **Section 3.2**. Categorical cross-entropy loss function and dice loss are the two loss functions that we decided to utilize for our research. Dice loss showed that it is suitable for imbalanced CMR data.

- **Question 3** What is the effect of the attention module in the cardiac MR segmentation task?

  we added an attention module to the supervised baseline model to get the answer in **Section 3.3**. The attention module guides the deep model where to look during training so it has a positive effect on the output results.

- **Question 4** How can we leverage unlabeled images to train a deep model to have a robust model?

  we introduced our proposed semi-supervised method in **Chapter 4**. We considered an approach to predict pseudo masks for the unlabeled data samples from the vendor C using an iterative manner. This method significantly enhanced the generalizability of our proposed model by providing our model a larger training dataset.

- **Question 5** How to reduce the sensitivity of our proposed semi-supervised method on the supervised model?

  We explained the histogram matching in **Section 4.2.1** as a solution to improve the generalizability of our model and to reduce the dependency of our model to the trained, supervised model. The reason for the effectiveness of this approach is that the deep model can mimic the intensity distribution of the unlabeled datasets.

5.2 Summary and Contributions

Medical image segmentation using deep learning has been recently dominant in the related research areas. However, for having a robust deep model to achieve accurate segmentation on unseen images, we need many labeled data that is both tedious and expensive. To overcome this issue mentioned above, we proposed a semi-supervised method to leverage the unlabeled medical data, abundant and accessible. First, we introduced fully supervised methods in **Chapter 3**. We reviewed U-Net and Residual U-Net architectures in detail and compared each of the two benchmark datasets presented in **Section 3.3**. According to the results provided in **Section 3.4.1**, Residual U-Net showed better dice scores compare to the U-Net architecture.

Also, a comparison between categorical cross-entropy loss function and dice loss in **Section 3.4.2** showed us that dice loss could cope with the imbalance in the medical images. In contrast, categorical cross-entropy loss function, which is based on pixel-wised examinations, is not a proper loss function for cardiac MR images with a high imbalance between background and foreground regions.

Furthermore, we append the attention module that [4] have applied on Pancreas segmentation to our fully supervised method. Because the region of interest in a CMR

image is relatively small, an attention module can guide the deep model where to look. According to the results in **Section 4.3.1**, the attention module slightly improves the dice score.

In order to increase the number of labeled data, we used standard data augmentation techniques such as translation and rotation, which we described in **Section 3.3.4**. Applying the data augmentation increases the numbers and variance of the training data, which helps avoid overfitting. Results that we provided led us to the best-supervised model, using Residual U-Net architecture with dice loss considering data augmentation techniques.

Next, we proposed a semi-supervised method to utilized the unlabeled data. We use the trained fully supervised method to generate pseudo labels for the unlabeled dataset from vendor C. We articulately explained this method in **Section 4.1**. We merged both manually labeled data, and the pseudo labeled data to retrain the supervised model. Although this approach improves the performance of the model on unseen images, it still is substantially sensitive to the supervised baseline model. Histogram matching was the solution to the problem. We explained histogram matching as a data augmentation technique in **Section 4.2**. By applying histogram matching, the performance of the semi-supervised model was significantly boosted. The main reason for this performance enhancement is that the model has the chance to mimic the intensity distribution of the unlabeled dataset. Therefore, we can achieve plausibly accurate pseudo labels. The results in **Section 4.3** provided detailed information about the effectiveness of the proposed model.

5.3 Future Work

Semi-supervised segmentation is an open research area that many researchers are concentrating. Our results show performance improvements of the deep model to have robust predictions on unseen data, though still the accuracy of the model has the capacity of more improvements. First, we aim to validate our technique in a larger cohort with more variations (unlabeled data from a different vendor). We aim to train our framework in a continuous learning scheme in datasets, where the manual annotations are limited.

Also, we plan to apply the label propagation approach. Since the CMR data are four-dimensional, containing a sequence of several 2D scans in the vertical direction (z-axis). Usually, the manual annotations have been done on ES and ED time frames. Thus, other time frames remain unlabeled. In order to leverage them in a semi-supervised manner, we aim to use label propagation to increase the variation of the training dataset in terms of time frames.

REFERENCES

[1] **Chen, C.**, **Qin, C.**, **Qiu, H.**, **Tarroni, G.**, **Duan, J.**, **Bai, W. and Rueckert, D.** (2020). Deep learning for cardiac image segmentation: A review, *Frontiers in Cardiovascular Medicine*, *7*, 25.

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

[3] **Zhang, Z.**, **Liu, Q. and Wang, Y.** (2018). Road extraction by deep residual u-net, *IEEE Geoscience and Remote Sensing Letters*, *15*(5), 749–753.

[4] **Oktay, O.**, **Schlemper, J.**, **Folgoc, L.L.**, **Lee, M.**, **Heinrich, M.**, **Misawa, K.**, **Mori, K.**, **McDonagh, S.**, **Hammerla, N.Y.**, **Kainz, B.** *et al.* (2018). Attention u-net: Learning where to look for the pancreas, *arXiv preprint arXiv:1804.03999*.

[5] **Yuheng, S. and Hao, Y.** (2017). Image segmentation algorithms overview, *arXiv preprint arXiv:1707.02051*.

[6] **Petitjean, C. and Dacher, J.N.** (2011). A review of segmentation methods in short axis cardiac MR images, *Medical image analysis*, *15*(2), 169–184.

[7] **Lin, X.**, **Cowan, B. and Young, A.** (2006). Model-based graph cut method for segmentation of the left ventricle, *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, pp.3059–3062.

[8] **Cocosco, C.A.**, **Netsch, T.**, **Sénégas, J.**, **Bystrov, D.**, **Niessen, W.J. and Viergever, M.A.** (2004). Automatic cardiac region-of-interest computation in cine 3D structural MRI, *International Congress Series*, volume1268, Elsevier, pp.1126–1131.

[9] **Stegmann, M.B.**, **Fisker, R. and Ersboll, B.** (2001). Extending and applying active appearance models for automated, high precision segmentation in different image modalities, *Proceedings of the Scandinavian Conference on Image Analysis*, pp.90–97.

[10] **Peng, P.**, **Lekadir, K.**, **Gooya, A.**, **Shao, L.**, **Petersen, S.E. and Frangi, A.F.** (2016). A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging, *Magnetic Resonance Materials in Physics, Biology and Medicine*, *29*(2), 155–195.

[11] **WHO**, Cardiovascular disease, `https://www.who.int/cardiovascular_diseases/about_cvd/en/`.

[12] **Chow, L.S. and Paramesran, R.** (2016). Review of medical image quality assessment, *Biomedical signal processing and control*, *27*, 145–154.

[13] **Singh, G.**, **Al'Aref, S.J.**, **Van Assen, M.**, **Kim, T.S.**, **van Rosendael, A.**, **Kolli, K.K.**, **Dwivedi, A.**, **Maliakal, G.**, **Pandey, M.**, **Wang, J.** *et al.* (2018). Machine learning in cardiac CT: basic concepts and contemporary data, *Journal of cardiovascular computed tomography*, *12*(3), 192–201.

[14] **Van Sloun, R.J.**, **Cohen, R. and Eldar, Y.C.** (2019). Deep learning in ultrasound imaging, *Proceedings of the IEEE*, *108*(1), 11–29.

[15] **Wang, Z.** (2020). Deep learning in medical ultrasound image segmentation: A review, *arXiv preprint arXiv:2002.07703*.

[16] **Biederman, R.W.**, **Biederman, R.W.**, **Doyle, M. and Yamrozik, J.** (2008). *The cardiovascular MRI tutorial: lectures and learning*, Lippincott Williams & Wilkins.

[17] **Lorell, B.H. and Carabello, B.A.** (2000). Left ventricular hypertrophy: pathogenesis, detection, and prognosis, *Circulation*, *102*(4), 470–479.

[18] **Lee, J.G.**, **Jun, S.**, **Cho, Y.W.**, **Lee, H.**, **Kim, G.B.**, **Seo, J.B. and Kim, N.** (2017). Deep learning in medical imaging: general overview, *Korean journal of radiology*, *18*(4), 570–584.

[19] **Zhang, Y.**, **Yang, J.**, **Hou, F.**, **Liu, Y.**, **Wang, Y.**, **Tian, J.**, **Zhong, C.**, **Zhang, Y. and He, Z.** (2020). Semi-supervised Cardiac Image Segmentation via Label Propagation and Style Transfer, *arXiv preprint arXiv:2012.14785*.

[20] **Huang, S.**, **Liu, J.**, **Lee, L.**, **Venkatesh, S.**, **Teo, L.**, **Au, C. and Nowinski, W.** (2011). An Image-Based Comprehensive Approach for Automatic Segmentation of Left Ventricle from Cardiac Short Axis Cine MR Images, *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, *24*, 598–608.

[21] **Avendi, M.**, **Kheradvar, A. and Jafarkhani, H.** (2016). A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI, *Medical image analysis*, *30*, 108–119.

[22] **Long, J.**, **Shelhamer, E. and Darrell, T.** (2015). Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.3431–3440.

[23] **Poudel, R.P.**, **Lamata, P. and Montana, G.**, (2016). Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation, Reconstruction, segmentation, and analysis of medical images, Springer, pp.83–94.

[24] **Roth, H.R.**, **Oda, H.**, **Zhou, X.**, **Shimizu, N.**, **Yang, Y.**, **Hayashi, Y.**, **Oda, M.**, **Fujiwara, M.**, **Misawa, K. and Mori, K.** (2018). An application of cascaded 3D fully convolutional networks for medical

image segmentation, *Computerized Medical Imaging and Graphics*, *66*, 90–99.

[25] **Ronneberger, O.**, **Fischer, P. and Brox, T.** (2015). U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp.234–241.

[26] **Tao, Q.**, **Yan, W.**, **Wang, Y.**, **Paiman, E.H.**, **Shamonin, D.P.**, **Garg, P.**, **Plein, S.**, **Huang, L.**, **Xia, L.**, **Sramko, M.** *et al.* (2019). Deep learning–based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study, *Radiology*, *290*(1), 81–88.

[27] **Xia, Q.**, **Yao, Y.**, **Hu, Z. and Hao, A.** (2018). Automatic 3D atrial segmentation from GE-MRIs using volumetric fully convolutional networks, *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, pp.211–220.

[28] **Bengio, Y.** (2009). *Learning deep architectures for AI*, Now Publishers Inc.

[29] **Zhang, Z.**, **Liu, Q. and Wang, Y.** (2018). Road Extraction by Deep Residual U-Net, *IEEE Geoscience and Remote Sensing Letters*, *15*(5), 749–753.

[30] **Krizhevsky, A.**, **Sutskever, I. and Hinton, G.E.** (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, *25*, 1097–1105.

[31] **Lemley, J.**, **Bazrafkan, S. and Corcoran, P.** (2017). Smart augmentation learning an optimal data augmentation strategy, *Ieee Access*, *5*, 5858–5869.

[32] **Lim, S.**, **Kim, I.**, **Kim, T.**, **Kim, C. and Kim, S.** (2019). Fast autoaugment, *arXiv preprint arXiv:1905.00397*.

[33] **Eaton-Rosen, Z.**, **Bragman, F.**, **Ourselin, S. and Cardoso, M.J.** (2018). Improving data augmentation for medical image segmentation.

[34] **Xu, J.**, **Li, M. and Zhu, Z.** (2020). Automatic data augmentation for 3D medical image segmentation, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp.378–387.

[35] **Papandreou, G.**, **Chen, L.C.**, **Murphy, K.P. and Yuille, A.L.** (2015). Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation, *Proceedings of the IEEE international conference on computer vision*, pp.1742–1750.

[36] **Bai, W.**, **Oktay, O.**, **Sinclair, M.**, **Suzuki, H.**, **Rajchl, M.**, **Tarroni, G.**, **Glocker, B.**, **King, A.**, **Matthews, P.M. and Rueckert, D.** (2017). Semi-supervised learning for network-based cardiac MR image segmentation, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp.253–260.

[37] **Dou, Q.**, **Castro, D.C.**, **Kamnitsas, K. and Glocker, B.** (2019). Domain generalization via model-agnostic learning of semantic features, *arXiv preprint arXiv:1910.13580*.

[38] **Ma, J.** (2020). Histogram Matching Augmentation for Domain Adaptation with Application to Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Image Segmentation, *arXiv preprint arXiv:2012.13871*.

[39] **Scannell, C.M.**, **Chiribiri, A. and Veta, M.** (2020). Domain-Adversarial Learning for Multi-Centre, Multi-Vendor, and Multi-Disease Cardiac MR Image Segmentation, *arXiv preprint arXiv:2008.11776*.

[40] **Elif, A. and Ilkay, O.** (2020). Accurate Myocardial Pathology Segmentation with Residual U-Net, ***X. Zhuang and L. Li****, editors, Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images*, Springer International Publishing, Cham, pp.128–137.

[41] **Milletari, F.**, **Navab, N. and Ahmadi, S.A.** (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation, *2016 fourth international conference on 3D vision (3DV)*, IEEE, pp.565–571.

[42] **Zeiler, M.D.** (2012). Adadelta: an adaptive learning rate method, *arXiv preprint arXiv:1212.5701*.

[43] **Kingma, D.P. and Ba, J.** (2014). Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.

[44] **Caruana, R.**, **Lawrence, S. and Giles, L.** (2001). Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping, *Advances in neural information processing systems*, 402–408.

[45] **STACOM2018**, Left Ventricle Full Quantification Challenge MICCAI 2018, `https://lvquan18.github.io/`.

[46] **M&MsChallenge**, Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Image Segmentation Challenge, `https://www.ub.edu/mnms/`.

CURRICULUM VITAE

Name Surname : Mahyar Bolhassani

- **M.Sc. (Currently)** : 2021, Istanbul Technical University, Faculty of Electronics and Communication Engineering, Department of Biomedical Engineering
- **M.Sc.** : 2012, Science and Research branch of Azad University, Faculty of Electrical Engineering, Department of Power Engineering
- **B.Sc.** : 2005, Amirkabir University of Technology, Faculty of Electrical and Electronics, Department of Electrical Engineering

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **M. Bolhassani, Ilkay Oksuz** 2021. Semi-Supervised Segmentation of Multi-vendor and Multi-center Cardiac MRI *SIU 2021- 29th IEEE CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATIONS APPLICATIONS*, June, 2021, Turkey.

M. Bolhassani	FULLY SUPERVISED AND SEMI-SUPERVISED SEMANTIC SEGMENTATION OF CARDIAC MR USING DEEP LEARNING	2021