Chapter 5

Cascaded Model (Conventional+Deep Learning) for Weakly Supervised Segmentation of Left Ventricle in Cardiac Magnetic Resonance Images

5.1 Introduction

According to the World Health Organization statistics, cardiovascular diseases (CVDs) are one of the leading causes of death worldwide, taking approx. 18 million lives in 2016. Among these, four out of five deaths were due to CVD due to heart attacks or strokes. In the last two decades, the number of deaths due to CVDs has increased considerably [102]. Timely recognition of CVDs is the first and most important step in the treatment and management of patients. Cardiac function analysis provides important information related to an abnormality in the heart such as but not limited to valve abnormality and cardiac tumor as it provides structural information of the heart. To extract this information accurately, precise segmentation of different parts of the heart is required [102, 103].

In the last decade, several methods have been developed to segment cardiac MR images. The conventional methods of MR image segmentation include but are not limited to threshold-based methods, pixel classification methods, deformable models, atlas-based methods, and graph-based methods. More recently, with the development of advanced deep learning neural network-based algorithms, many more automatic segmentation methods have been developed. U-Net and its modification have been widely used to develop and improve segmentation techniques for Cardiac MR images.

The main drawback of the traditional methods is they are easily affected by surrounding organs, noise, weak edges, and inherent contrast inhomogeneity. Some of them work on the detection of similarity in structures or depend on seed point selection. These factors make these methods vulnerable. Other drawbacks include low accuracy and robustness and require user interaction

which is not desirable. On the other hand, deep learning-based methods give more accurate and robust results and are less affected by noise or weak edges. The drawback of deep learning-based methods is the requirement of ground truth.

The application of deep learning-based methods for the segmentation of cardiac MR images was very much limited till 2015 when Kaggle organized a challenge where several deep learning-based methods were presented to the community. Since then, many algorithms have been developed to segment the Cardiac MR images. Several variants of U-Net, Dense Net, and others were applied to solve the problem of Cardiac MR image segmentation. In MICCAI 2017 Automatic Cardiac Diagnosis Challenge, 9 out of 10 methods are deep learning based methods and out of them, eight were variants of U-Net.Avendi et al.[104] proposed a cascade of deformable models and a deep learning method for the segmentation of CMRI. Level set combined with deep learning method is proposed by Luo et al. [105]for segmentation of the Left ventricle in CMRI. Isensee et al.[102] proposed an ensemble of modified 2D and 3DU-Net and the results obtained were highly accurate. Tripathi et al. and Bhattacharjee et al.[106, 107] proposed a neural network architecture based on U-Net to segment the noise-free as well as noisy Cardiac MR images and for 2D/3D abdominal image registration, respectively. In [106], the author reported that the proposed modification not only overcomes the noise-induced in the images but also increases the segmentation accuracy. Yang et al. [108] reviewed GAN-based fast MRI segmentation methods on various anatomical datasets to show the generalizability of the methods. Further, a modification to U-Net is also proposed to segment the images and conclude a strong correlation between segmentation obtained using the proposed method and ground truth obtained by experienced personnel.

Also, Zhou et al.[109] developed a retrospective study to segment Ventricle images in elderly patients and showed to produce excellent results. Chen et al[110] proposed inter-cascade GAN for segmentation task of unbalanced atrial targets from late gadolinium-enhanced cardiac magnetic resonance images. It first investigates an adaptive attention cascade to automatically correlate segmentation of unbalanced tasks and then applies adversarial regularization to segment the scars [109, 111] .Multiview recurrent aggregation network (MV-RAN) proposed by M. Li et al. utilized a unified framework for segmentation and classification of echocardiographic sequences segmentation [112, 113] .A modified two-stage U-net proposed by M. Pop et al. simultaneously learned to detect ROI in the full volume and classified voxels without compromising the original resolution. This approach achieved better segmentation performance as compared to state-of-theart methods [36, 114] . A. F. Frangi et al. proposed a probabilistic deep voxel-wise dilated residual network for segmentation of whole heart in 3D MR images [97, 115]. An iterative algorithm for segmentation of

left ventricle (LV) was proposed by Yuanhan Mo. Fusion of deep learning network with the novel dynamic-based labeling scheme was utilized in this approach [116, 117] . G. Yang, J. Chen, Z Gao et al. proposed a joint segmentation approach which utilizes multiview two-task (MVTT) recursive attention model. This method is used to segment the left atrium (LA) in 3D CMR images and in addition it also delineates the scar on the same dataset [118]. For segmentation of the peripheral zone (PZ) and transition zone (TZ) of the prostate with uncertainty estimation, Liu et al. proposed spatial attentive Bayesian deep learning network. The overall uncertainty was found to be the highest with the actual model performance[119] . A fully CNN with a novel feature pyramid attention mechanism was developed by Y. Liu et al. for automatic segmentation of prostate zones. This automatic segmentation method achieved highest accuracy in segmentation of PZ and TZ [120].

The availability of a massive biomedical dataset nowadays is relatively easy as several datasets were available due to different challenges carried out in recent years. The problem in the way of segmentation is to annotate the images correctly and accurately which still requires medical practitioners and biomedical experts. The situation reached the bottleneck in the segmentation of biomedical images as supervised learning methods require manually segmented ground truth of the image portion required to be segmented. Semi-supervised deep learning methods hold the promise of segmenting large unlabeled images with improved segmentation accuracy.

Recently,many researchers focused on developing strategies to segment the biomedical images in a minimal or unsupervised way. Lin Gu et al. [120] focused on semi supervised learning so that large unlabeled data can be used for training. The proposed use of forest-oriented super pixels to augment the standard random forest to cope with low confidence homogeneous connected areas in biomedical images. Karvadec et al. [111, 121] investigated curriculum-style strategy for semi-supervised segmentation. The proposed method achieved competitive results leveraging unlabeled data more effectively. Yi Zhou et al. [112, 122] used collaborative learning and attention model for Semi-supervised segmentation of biomedical images. The model was tested on diabetic retinopathy images and validated effectiveness in improvements over state-of-the- art methods. LuyiHan et al. [91, 122] introduced BUSGAN consisting of a segmentation base network – BUS-S and an evaluation base network–BUS-E. Through adversarial training, the BUS-GAN model achieved higher accuracy in the segmentation of images. Li et al. [97, 123] train a GAN that captures the joint image-label distribution and is trained using a large set of unlabeled data. The proposed model was built on top of Style GAN2 and demonstrated strong in-domain performance compared to baselines.

Wu et al.[116, 124] reviewed the conventional and current state-of-the-art methods to segment cardiac fibrosis and scar accurately utilizing different modalities. Yung et al. [33] in their paper described recent advances in deep learning and how it is improving cardiac imaging. Huang et al [125] proposed a deep multi-tasking learning framework and integrated multi-depth fusion module in the framework to accurately segment brain tumors. The group used distance transform decoder-based V-Net using contour generated by the mask.

In this paper, we propose to use a cascade of conventional segmentation techniques and more recent deep neural network-based architecture for weakly supervised segmentation. Conventional methods (seed region growing, random walker, and K-means clustering) works on pixel classification based on the intensity level. These methods give good segmentation results provided no sudden change in pixel intensity is there or edges are sharp. But in biomedical images, due to contrast inhomogeneity, instrument handling, and operational variations, low edges or sudden pixel intensity change occurs even for the same class of pixels. This makes the conventional methods vulnerable. To overcome the issue, the binary masks obtained from the conventional methods are used to train a deep neural network to increase the segmentation accuracy and reduce the vulnerability of the algorithm while giving it more generalization and robustness.

5.2 Method

Accurate segmentation of Cardiac MR images is required to extract useful information about the structural details of the heart. Currently, the majority of algorithms available are supervised learning-based which means they require manual segmentation or ground truth. In this manuscript, an autoencoder with a U-Net style skip connection is used to perform the task of segmentation. The model architecture of the autoencoder is the same as U-Net with a depth of four layers and skips connections from the encoder part to the decoder part to retain the tempo-spatial feature maps. The basic difference between the two neural networks is that in U-Net, upsampling is used in the decoder part to match the feature map size whereas, for autoencoder, pooling is used. The network consists of a convolution layer, ReLU activation, and Batch Normalization [126, 127] .The contracting path consists of two 3×3 convolutions with ReLU activation and a 2×2 max-pooling layer. The max-pooling operation reduces the factor by n (n = stride) and the number of channels is increased by a factor of n. At every step of the expansive path, upsampling of the feature map is carried followed by a 2×2 convolution to half the number of feature channels and concatenated with the corresponding cropped feature map of the contraction path. The concatenation is followed

by two 3×3 convolutional layers, each with ReLU activation. It is required to crop because the border pixels were lost at every convolution. At the final output layer, a 1×1 convolution is used to map each 64- component feature map to the desired class. A total of 23 convolutional layers are there in the network. The use of dense layers is avoided so that it can accept images of any size.

Autoencoder provides accurate segmentation results and requires the involvement of a medical practitioner or biomedical expert for manual annotation of datasets. Also, the time required for annotating the images makes it more tedious. To overcome this, here we propose weakly-supervised or minimally supervised techniques to segment the cardiac MR images.

5.2.1 Seed Region Growing Based Segmentation

Seed region growing is one of the popular segmentation approaches which is used in the proposed methodology for extracting the region of interest (ROI) from the cardiac MR images. This method works on the seed indication, or in other words, to choose a point or pixel within the particular region that is going to be extracted. Primarily the idea is to constitute a region by accumulating similar pixels together [109]. Hence the seed point selection should be done very carefully as the segmentation results are directly influenced by this. In this procedure, first, the seed point or the starting point of growth in the image is precisely chosen. Researchers have reported a lot of approaches regarding the initiation of the seed value [113]. A manual seed initiation process has been executed in the proposed methodology based on the required region of interest which in this case is the left ventricle in cardiac cine MR images. Once the first seed point has been selected, the neighbor pixels are automatically compared by the algorithm in search of similar characteristics while satisfying the predefined growth criteria (threshold value). Once the criteria are matched, the seed will grow by adopting the neighbor pixel into the region squad and the grouped pixels become the new seed as a whole. The process is repeated till all the pixels of the same threshold level are grouped and no more pixel is left to add to the region. It is very important to choose the threshold minutely as it is the only growing criterion for the region. In general cases, the value of the threshold is calculated based on the gray-level information of the images. Improper threshold selection leads to inappropriate merging of the pixels, while some pixels are not added to the region, resulting in holes in the constructed area of interest.

Suppose if pixel similarity is defined by F, then at a certain level, the function $F(R)$ will be defining the regional characteristic for the region $R\cdot(x, y)$ as the next pixel with eigenvector $F(p)$ on which the threshold criteria is going to be tested. Now, if $|F(R)F(p)| < T$, where T specifies the

predefined threshold level, then the pixel is merged, otherwise, the search is carried on for the next neighboring pixel. The process continues till all the similar pixels are grouped and then generates a binary image of the region of interest for further analysis.

5.2.2 Clustering Based Segmentation

Clustering is an unsupervised method to split up a set of data into a specific number of groups. One of the popular clustering method is the K-means technique, which is used in this proposed work to generate a primary segmentation of the region of interest from the cardiac MR images. A collection of data points is partitioned into k number of groups in this technique; the idea is simple and the method has a relatively medium computational complexity (based on the number of elements, clusters, and iterations). Since biomedical image segmentation using unsupervised techniques are pretty challenging due to the presence of organs and tissues of similar intensities, minimizing the number of data points will be a smart choice. Hence a manually cropped portion of the images inclusive of the region of interest is provided for clustering which produces the initial (rough) segmentation.

The algorithm is executed in several phases. First Random cluster centers are initialized for the dataset; based on that, the dataset is divided into K clusters randomly, while assigning each cluster almost the same number of nearest data points. Once the first stage of grouping is completed, the new centroids for each cluster are recalculated. Then, Euclidean distance is computed from each of the data points to each cluster centroid, and based on that, the data point is assigned to the nearest cluster with the minimum Euclidean distance to its centroid [115].

The process continues until the shifting of the centroids becomes less than a predefined threshold or the iteration limit has been reached. The Euclidean distance is defined by $d = ||p(x,y)c_k||$, where c_k is the centroid of the k^{th} cluster and $p(x,y)$ is any random pixel of the image.The new centroid position is recalculated using: $c_k = \frac{1}{k}\sum_{y \in c_k} \sum_{y \in c_k} p(x,y)$.

So, in this unsupervised iterative clustering algorithm, the sum of the distances from each element to its cluster center is minimized over all the clusters. Although random, the initial cluster centroids should be chosen carefully, as the final clustering quality depends on it.

5.2.3 Random Walker Based Segmentation

The Random Walker is a well-known algorithm for performing multi-label interactive segmentation. With a small number of predefined pixel labels, the random walking probability to reach a pre-labeled pixel starting from the unlabeled ones can be quickly calculated. Hence the segmentation can be performed by assigning each pixel with the greatest probability. Basically, in this algorithm, the image is represented by an undirected weighted graph, where the pixels of the image constitutes the vertices and the intensities are mapped to the edge weights. As the algorithm works based on pre-defined labels or seed points, a manual seed point selection work (a few pixels for both foreground and background) have been performed [117] .

So at first, a small number of pixels are interactively pointed as "object" seed points (inclusive of the region of interest) and a few "background" seed points are selected from the other portion of the image. Now the unlabeled pixels should release the random walker. The random walker leaving any particular unlabeled pixel will be arriving at a seed point. The probability of the seed point being an "object" rather than a "background" is computed . By using the graph Laplacian matrix, a system of linear equations can be solved which may analytically determine the value of these probabilities. Once all the probabilities for each of the pixels are calculated, they are assigned to the label (object/background) corresponding to the most probable random walker generation (i.e. "1" for the object and "0" for the background) and finally, it produces an image with the primary segmentation of the region of interest.

5.3 Experiments

In this study, the Automatic Cardiac Diagnosis Challenge dataset (ACDC) and Sunnybrook Cardiac dataset (SCD) are used. The images in the ACDC dataset are first preprocessed to crop the region of interest and then normalized between 0 and 1 and saved as a NumPy array. The dataset is then randomly split into training and testing datasets in an 80:20 ratio which is then fed to the segmentation pipeline. The segmentation pipeline first includes the segmentation of pre-processed unlabeled images using conventional methods and then applying a the morphological operation to obtain a rough segmentation the mask which is then used as ground truth in deep neural network training. *Cross entropy* and *dice loss* combination used as loss function and *Dice score, Jaccard Index*, and *Hausdorff* distance is used to monitor the performance of the algorithm. Experiments- (i)

comparison of all segmentation methods and (ii) model training with fewer images, were performed and the results are further explained in detail in the next sections.

5.3.1 Dataset

5.3.1.1 Automatic Cardiac Diagnosis Challenge [Dataset 1]

Automatic Cardiac Diagnosis Challenge (ACDC) dataset is magnetic resonance image data in a short-axis view (SAX) of 100 patients with five different pathophysiological conditions namely Dilated cardiomyopathy (DCM), Hyper Cardiomyopathy (HCM), Myocardial infarction (MINF), Abnormal right ventricle (RV), and Normal (NOR). Manual segmentation of the left ventricle (LV), the myocardium (MYO), and the right ventricle (RV) for the The end-Diastolic phase and End-Systolic phase are provided. Images were acquired over 6 years using two different MRI scanners in breadth-hold. Slices of image cover from the base of LV to the apex with slice thickness varying from 5 to 8mm with an interslice gap of 5mm and spatial resolution varying from 1.37 to $1.68 mm^2/pixel$ with 28 to 40 images per patient.

5.3.1.2 Sunnybrook Cardiac Data [Dataset 2]

The Sunnybrook Cardiac MR Left Ventricle Segmentation Challenge – MICCAI 2009 dataset consists of 45 CMR images from patients having four pathological conditions heart failure with ischemia, heart failure without ischemia, hypertrophic cardiomyopathy, and normal subjects. Manually drawn contours for the endocardium and epicardium are provided in text format which consists of contour points. CMR images were acquired with 1.5 T GE Signa MRI with a thickness of 8 mm, an interslice gap of 8 mm, a FOV of 320 × 320 mm, and a size of 256 × 256. Six to 12 SAX images were obtained from the atrioventricular ring to the apex.

ACDC dataset (Dataset 1) is used for training and testing of the model while the SCD dataset (Dataset 2) is used only for testing in every experiment.

5.3.2 Pre-processing

The ACDC dataset is first cropped to 128×128 pixel size using ROI center and radius. This size reduction not only reduced the computation load on the system but also eliminated training

issues that come due to the large background, pixel intensity interference of other organs, and class imbalance. As the images are taken from different scanners at different sites, there is a wide range of intensity variation. To overcome the effect of intensity variation during training, images were then normalized in the range of 0–1. SCD data is also processed following a similar protocol used for the ACDC dataset. SCD dataset is only used for testing the proposed method.

5.3.3 Training Protocol

The segmentation network is implemented in Python using the TensorFlow framework and training on the GPU provided by Google Collaboratory. Pre-processed Images from dataset 1 were first segmented with unsupervised learning methods like Seed Region Growing and then the result is used as the target to train the Autoencoder network. For supervised training, the original mask provided with the dataset was used to train the network. The ADAM optimizer is used with the initial learning rate of 0.0001 and is trained for 50 epochs with a batch size of 16. For testing on the SCD dataset, saved model weights were loaded and predicted on images without processing on the dataset images.

5.3.4 Loss function and evaluation metrics

For a set of events, a measure of the difference between the probability distribution is known as cross-entropy. Using cross-entropy as a loss function instead of other loss functions like the sum of squares leads to faster training and better generalization of the model. The loss calculated by cross-entropy is the average per-pixel loss and the per-pixel loss is calculated discreetly. As a result, cross-entropy considers loss in the micro sense. Cross entropy is defined by Equation.5.1:

$$Loss_{CE} = -\frac{1}{m}(y^i \log(p^i) + (1 - y^i) \log(1 - p^i)) \tag{5.1}$$

where m is the number of samples, y^i is the label of the sample, and p^i is the predicted probability value, $p^i \in (0, 1)$. If there is a class imbalance in the training set, training will be difficult. To overcome this, the loss function is conjugated with dice loss which is a statistic developed in the 1940s to gauge the similarity between two samples. It was brought to the computer vision community by Milletari et al. [117] in 2016 for 3D medical image segmentation. Dice loss is defined by Equation.5.2:

$$DiceLoss = 1 - \frac{2 \times \sum_{pixels} \times y_{true} \times y_{pred}}{\sum_{pixels} y_{true}^2 + \sum_{pixels} y_{pred}^2} \tag{5.2}$$

The segmented results were evaluated using well-known metrics namely Jaccard Index, Dice score, and Directed Hausdorff Distance.

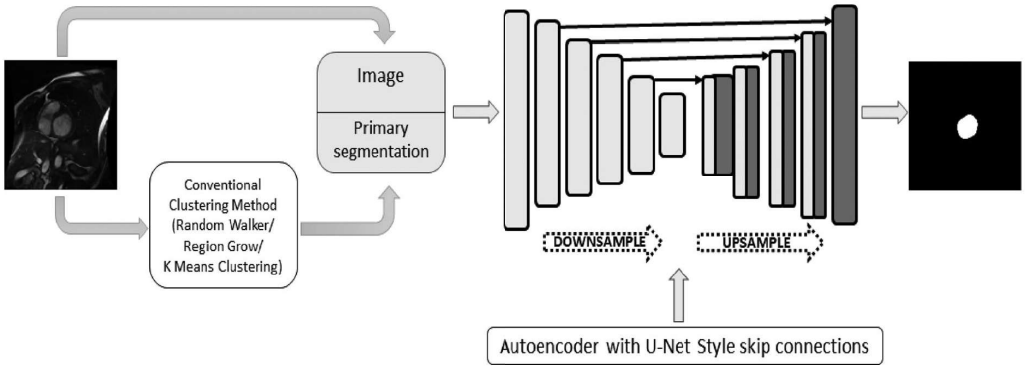

FIGURE 5.1: Graphical Representation of method

5.4 Results

5.4.1 Comparison of all segmentation methods

This section describes the experimentation of comparing the weakly supervised segmentation methods and the supervised one using the autoencoder network, tested on the Cardiac MR images dataset (for both Dataset 1 and 2, shown in Figure.5.1). The ground truth images are used as labels for the supervised segmentation process, whereas, in the case of weakly supervised, primary segmentations are derived from the training images with Seed Region Growing, Random Walker, and K-means clustering methods and those are utilized as initial labels for the autoencoder. The CNN model is trained with a total of 1419 images (from Dataset 1). It is to be noted that only Dataset 1 images are taken for the training, while 355 images from Dataset 1 and 666 images of Dataset 2 are kept as testing images for evaluating the overall performance. The performance metrics for all the methods in terms of Dice score, Jaccard index, and Hausdorff distance is compared with the supervised techniques and are shown in Table.5.1 and Table.5.2 for Dataset 1 and 2, respectively.

	Weakly Supervised			Supervised
All Images	**SRGA**	**RWA**	**KMCA**	**Autoencoder**
Dice	0.87794 (0.142408)	0.80313 (0.183653)	0.85215 (0.138007)	0.92106 (0.157554)
Jaccard	0.8016 (0.169149)	0.70296 (0.212712)	0.76457 (0.171016)	0.87801 (0.169526)
Hausdorff	2.30347 (0.529502)	2.46827 (0.657614)	2.3555 (0.585234)	1.84001 (0.434773)

TABLE 5.1: Segmentation metric for different methods in dataset 1; $mean(\pm Std)$

	Weakly Supervised			Supervised
All Images	**SRGA**	**RWA**	**KMCA**	**Autoencoder**
Dice	0.8208 (0.16878)	0.70304 (0.21727)	0.76529 (0.19855)	0.833136 (0.161659)
Jaccard	0.72477 (0.2029)	0.58133 (0.23743)	0.65571 (0.22611)	0.741082 (0.197559)
Hausdorff	2.31379 (0.74281)	2.37598 (0.56868)	2.24461 (0.57776)	2.164891 (0.69708)

TABLE 5.2: Segmentation metric for different methods in dataset 2; $mean(\pm Std)$

5.4.2 Model training with fewer images

Although the proposed methodology does not require the original ground truth images for the training, manual intervention to have the primary segmentation of more than 1400 images is undoubtedly a tiresome job for obvious reasons. Hence, this experiment is performed to train the autoencoder with a smaller number of images with respect to Experiment 1 and observe the outcomes precisely (shown in Figure.5.2-5.4) for each of the initial segmentation techniques. For a particular segmentation method, three levels of training are done; Degree I, Degree II, andDegree III. The basis of image selection for each level is based on the Dice score criteria of the previous experiment. For example, if the selection is going on for the Seed Region growing technique, all the images will undergo a score check for the model developed in the previous experiment (using Seed Region growing segmentation), and the images with a score of more than 0.7 will be selected for Degree I, more than 0.8 for Degree II and more than 0.9 for Degree III. A similar approach has been followed for the remaining initial segmentation methods (Figure.5.5).

Now, a weakly supervised segmentation, trained with a set of images can claim to achieve good performance when the test results are competitive enough with the supervised one. In this case, at each degree of weakly supervised training, using the reduced number of training images, the same images are utilized to train a supervised network using the original ground truth. As the same images (with and without ground truth) are utilized to execute the supervised and weakly supervised training, hence the performance metrics can be compared precisely now for each degree. These comparisons have been done independently for the different initial segmentation approaches. Just like the Experiment 1, all the training at every step is done with the images from dataset 1,

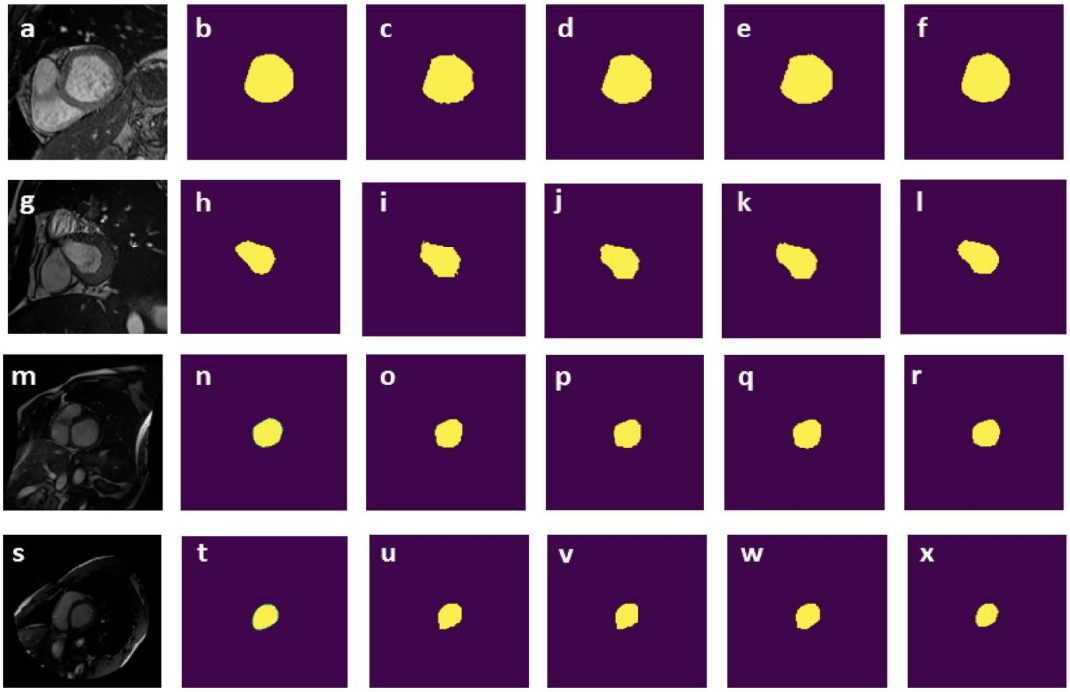

FIGURE 5.2: Segmentation result for unsupervised methods compared to supervised training: (a) original image, (b) mask, (c) seed region growing segmentation, (d) random walker segmentation, (e) K-means clustering segmentation, (f) autoencoder supervised learning

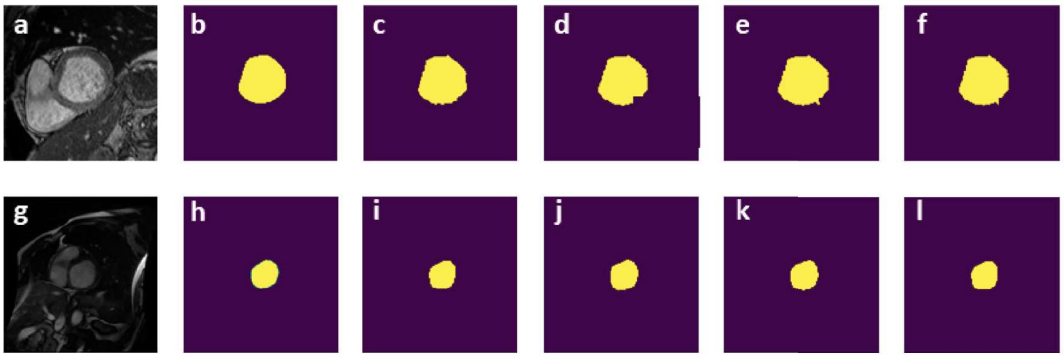

FIGURE 5.3: Segmentation result for seed region growing method: (a,g) image, (b,h) target, (c, i) segmentation result for all images, (d,j) segmentation result for Degree I, (e,k) segmentation result for Degree II (f,l) segmentation result for Degree III

while images from both datasets 1 and 2 are utilized for the performance testing (shown in Table.5.3-5.8). Also, the decreasing/increasing differences between the supervised and weakly supervised techniques are reflected in Figure.5.6 and Figure.5.7(For each approach, for each degree).

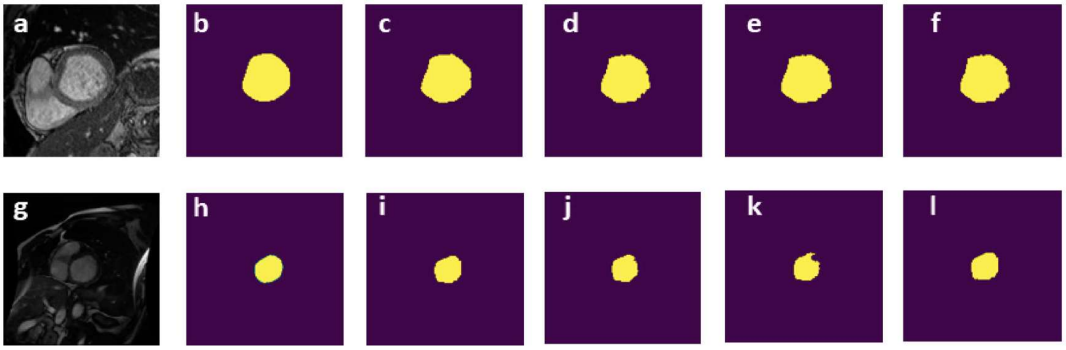

FIGURE 5.4: Segmentation result for random walker method: (a,g) image; (b,h) target; (c, i) segmentation result for all images, (d,j) segmentation result for Degree I (e,k) segmentation result for Degree II, (f,l) segmentation result for Degree III

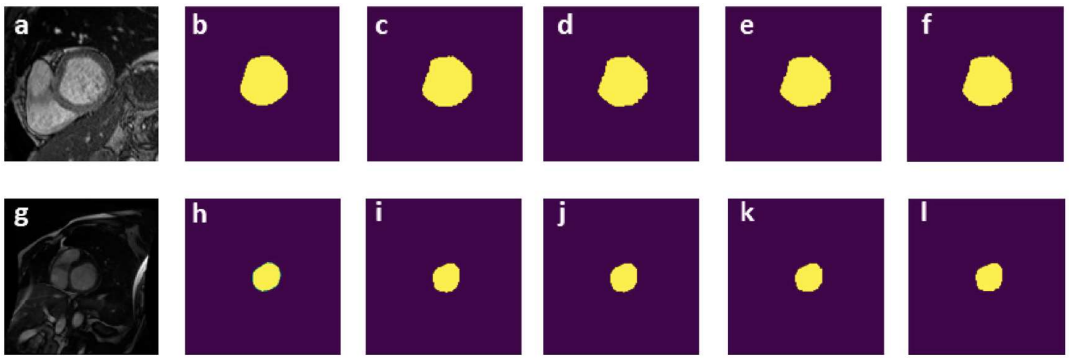

FIGURE 5.5: Segmentation result for K-means clustering method: (a,g) image, (b,h) target, (c, i) segmentation result for all images; (d,j) segmentation result for Degree I, (e,k) segmentation result for Degree II, (f,l) segmentation result for Degree III

5.5 Discussion

In the first experiment of comparing all the segmentation methods, it is evident from Table.5.2 that the use of the unsupervised Seed Region growing procedure to generate initial segmentation of the region of interest is competitive enough against the supervised segmentation with original ground truth images. The output using the the k-means clustering algorithm is also comparable to the supervised one, which is followed by the Random Walker method. A similar trend has been observed in Table.5.2 also, comprising the testing results of dataset 2. It is reflected in Figure.5.2 that with proper seed selection, Seed Region growing technique is almost perfectly able to capture the ROI from the cardiac MR images. The K-means clustering has also produced decent enough results when the images are manually cropped for the ROI including a limited neighborhood. Though the segmentation of these medical MR images is a pretty hard task, manual point selection for Random Walker has precisely given the capacity to segregate the ROI with satisfactory results.

Finally, it can be inferred from the listed figures and quantitative values, that the performances of Seed Region growing (with manual seed selection) and k-means clustering (with manual cropping of ROI) methods are close enough to the results of the supervised segmentation, while the performance of random walker is a bit lesser than the former ones.

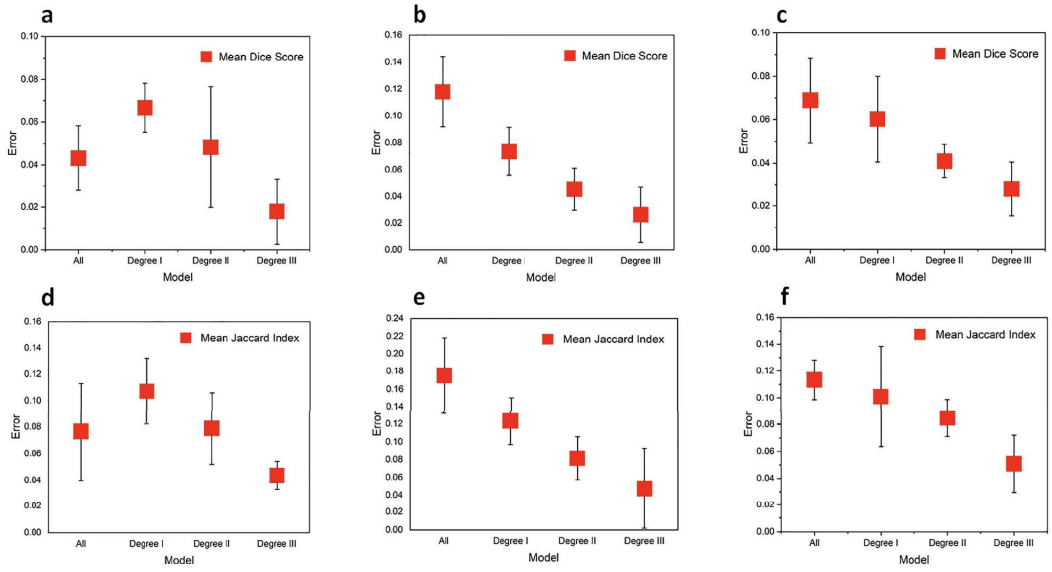

FIGURE 5.6: Error plot for Dataset 1 in (a) dice score of seed region growing, (b) dice score of random walker, (c) dice score of K-means clustering, (d) Jaccard Index of seed region growing, (e) Jaccard Index of a random walker, (f) Jaccard Index of K-means clustering

Seed region growing method on dataset 1				
	All (1342 images used)	**Degree I (1192 images used)**	**Degree II (1007 images used)**	**Degree III (653 images used)**
Dice	0.87794 (0.142408)	0.88105 (0.05634)	0.90537 (0.060176)	0.94182 (0.069287)
Jaccard	0.8016 0.169149)	0.80007 (0.089448)	0.83598 (0.093021)	0.89136 (0.086401)
Hausdorff	2.30347 (0.529502)	2.48109 (0.50791)	2.35445 (0.618121)	2.1906 (0.558301)
Dice_sup	0.92106 (0.157554)	0.94779 (0.067826)	0.95364 (0.031752)	0.95977 (0.084652)
Jaccard_sup	0.87801 (0.169526)	0.90732 (0.090945)	0.91493 (0.055768)	0.935 (0.09692)
Hausdorff_sup	1.84001 (0.434773)	1.89475 (0.51256)	1.93368 (0.505686)	1.9412 (0.601946)

TABLE 5.3: Segmentation metric for Seed Region growing method for all images and lesser images in dataset 1; $mean(\pm Std)$

In the second experiment, for each unsupervised segmentation technique, the network has been trained in three levels with a reduced number of images. Correspondingly, supervised training is

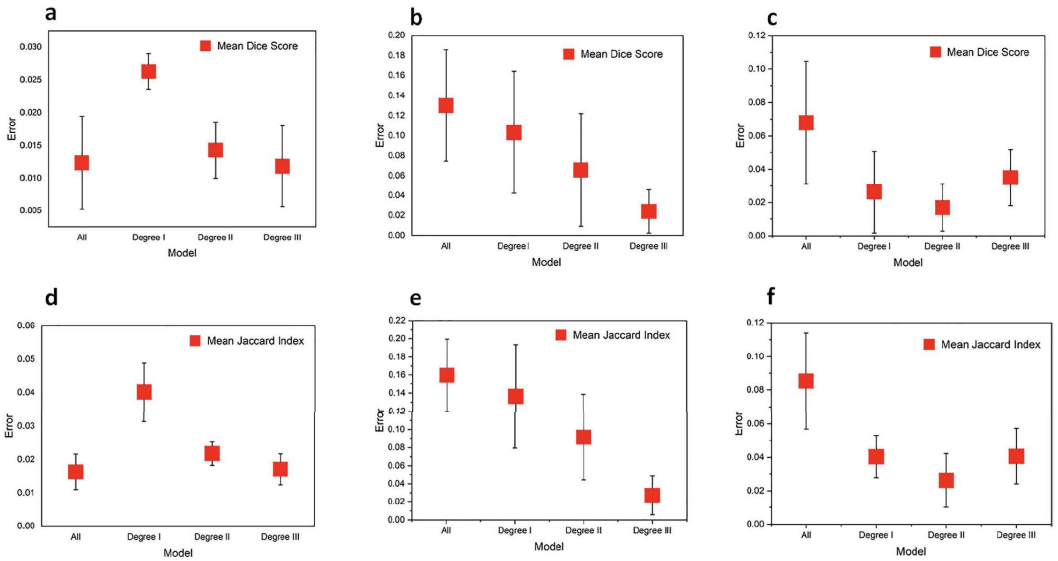

FIGURE 5.7: Error plot for Dataset 2 in (a) dice score of seed region growing, (b) dice score of random walker, (c) dice score of K-means clustering, (d) Jaccard Index of seed region growing, (e) Jaccard Index of the random walker, (f) Jaccard Index of K-means clustering

Random walker method on dataset 1				
	All (1332 images used)	**Degree I (958 images used)**	**Degree II (783 images used)**	**Degree III (474 images used)**
Dice	0.80313 (0.183653)	0.89076 (0.033615)	0.914 (0.030489)	0.94105 (0.023235)
Jaccard	0.70296 (0.212712)	0.80906 (0.056177)	0.84474 (0.051808)	0.88979 (0.040394)
Hausdorff	2.46827 (0.657614)	2.46158 (0.692343)	2.36876 (0.592624)	2.2497 (0.504001)
Dice_sup	0.92106 (0.157554)	0.96443 (0.015761)	0.95943 (0.014628)	0.96718 (0.019159)
Jaccard_sup	0.87801 (0.169526)	0.93248 (0.029752)	0.92636 (0.02772)	0.93701 (0.035841)
Hausdorff_sup	1.84001 (0.434773)	1.8853 (0.453864)	1.90723 (0.464985)	1.92368 (0.518198)

TABLE 5.4: Segmentation metric for Random walker method for all images and lesser images in dataset 1; $mean(\pm Std)$

also performed with the same images (and their ground truths) and the difference of metric values is plotted in Figure.5.5 and Figure.5.6.

The three different error plots are for the different segmentation techniques and each of the plots is defined with four data points (defined by the number of images), starting from "all images" and then reduced to a degree I, II, and III. As the training of the network is done with better images in

k-means clustering method on dataset 1				
	All (1323 images used)	**Degree I (1188 images used)**	**Degree II (1017 images used)**	**Degree III (648 images used)**
Dice	0.85215 (0.138007)	0.89372 (0.06975)	0.90922 (0.047531)	0.93952 (0.058939)
Jaccard	0.76457 (0.171016)	0.81644 (0.107593)	0.83995 (0.077653)	0.88677 (0.070125)
Hausdorff	2.3555 (0.585234)	2.35099 (0.579833)	2.32081 (0.54198)	2.23858 (0.487581)
Dice_sup	0.92106 (0.157554)	0.95398 (0.049977)	0.9501 (0.048318)	0.96751 (0.046429)
Jaccard_sup	0.87801 (0.169526)	0.91746 (0.070281)	0.92483 (0.063879)	0.93759 (0.065004)
Hausdorff_sup	1.84001 (0.434773)	1.87104 (0.618468)	1.87258 (0.470071)	1.8746 (0.645104)

TABLE 5.5: Segmentation metric for k.means clustering method for all images and lesser images in dataset 1; $mean(\pm Std)$

Seed Region growing method on dataset 2				
	All (666 images used)	**Degree I (554 images used)**	**Degree II (476 images used)**	**Degree III (295 images used)**
Dice	0.820801 (0.168783)	0.855234 (0.111655)	0.879061 (0.092663)	0.85239 (0.189638)
Jaccard	0.724776 (0.202897)	0.761431 (0.148392)	0.794727 (0.128115)	0.719063 (0.18085)
Hausdorff	2.313792 (0.742815)	2.347937 (0.714361)	2.269776 (0.691738)	2.530202 (0.842135)
Dice_sup	0.833136 (0.161659)	0.881531 (0.108924)	0.892356 (0.096964)	0.865206 (0.127721)
Jaccard_sup	0.741082 (0.197559)	0.801535 (0.139633)	0.816491 (0.12564)	0.761706 (0.164097)
Hausdorff_sup	2.164891 (0.69708)	2.057529 (0.584714)	2.077978 (0.631542)	2.500205 (0.869477)

TABLE 5.6: Segmentation metric for Seed Region growing method for all images and lesser images in dataset 2; $mean(\pm Std)$

the corresponding levels, the difference in metric values between supervised and weakly supervised should decrease and the ideal trend for all the plots should have been downward-moving with a slightly reduced slope. The error plots are shown for both datasets.

The error plot of Seed Region growing segmentation takes a little bounce in the Degree I data point (shown in Figure.5.5(a)–(d) and Figure.5.6(a)–(d)), but then it gets corrected in the next points and keeps moving downwards. For the case of K-Means clustering also, the downward moving curve is clearly visible, with a slight bulge present in the Degree III point of dataset 2. Apparently, the most relatable plot to the ideal one can be seen in the case of random walker segmentation

Random Walker method on dataset 2				
	All (638 images used)	**Degree I (384 images used)**	**Degree II (280 images used)**	**Degree III (116 images used)**
Dice	0.70304 (0.21727)	0.777012 (0.181319)	0.83116 (0.137664)	0.88527 (0.144408)
Jaccard	0.58133 (0.23743)	0.664851 (0.202106)	0.729563 (0.159263)	0.815554 (0.167058)
Hausdorff	2.37598 (0.56868)	2.349262 (0.513969)	2.365991 (0.534772)	2.22494 (0.557234)
Dice_sup	0.833136 (0.161659)	0.880141 (0.120482)	0.896708 (0.081243)	0.909335 (0.088586)
Jaccard_sup	0.741082 (0.197559)	0.801294 (0.145119)	0.820911 (0.112084)	0.842738 (0.113205)
Hausdorff_sup	2.164891 (0.69708)	1.882101 (0.467262)	2.012453 (0.499304)	2.051793 (0.53865)

TABLE 5.7: Segmentation metric for Random walker method for all images and lesser images in dataset 2; $mean(\pm Std)$

k-means clustering method on dataset 2				
	All (680 images used)	**Degree I (500 images used)**	**Degree II (402 images used)**	**Degree III (194 images used)**
Dice	0.76529 (0.19855)	0.861775 (0.095297)	0.892748 (0.062242)	0.842898 (0.183002)
Jaccard	0.65571 (0.22611)	0.767731 (0.128593)	0.811345 (0.091154)	0.761205 (0.209525)
Hausdorff	2.24461 (0.57776)	2.109581 (0.500889)	2.061978 (0.488003)	2.301189 (0.56674)
Dice_sup	0.833136 (0.161659)	0.888072 (0.092839)	0.909624 (0.048132)	0.877916 (0.136096)
Jaccard_sup	0.741082 (0.197559)	0.808147 (0.115943)	0.837522 (0.075144)	0.801888 (0.162908)
Hausdorff_sup	2.164891 (0.69708)	1.862171 (0.396635)	1.808729 (0.327025)	2.03432 (0.532902)

TABLE 5.8: Segmentation metric for k.means clustering method for all images and lesser images in dataset 2; $mean(\pm Std)$

which depicts a crisp downward slope for the corresponding data points. Hence, it can be clearly inferred that the use of Seed Region growing and k-means clustering for initial segmentation gives the exact and precise output, while the random walker is a steady and consistent approach for the primary segmentation of cardiac MR images.

5.6 Conclusion

In this study, we present a novel approach for weakly supervised segmentation by combining established conventional segmentation techniques with contemporary deep neural network architectures. Our methodology entails a cascade of traditional methods including seed region growing, random walker, and K-means clustering, which operate on pixel classification based on intensity levels. These traditional methods are employed to extract initial patch-based segmentation, yielding binary masks that are subsequently input into a newly proposed deep neural network. This network integrates an autoencoder with a U-Net style skip connection to accomplish the segmentation task. The autoencoder architecture aligns with U-Net's structure, comprising four layers and encoder-to-decoder skip connections to retain spatial-temporal feature maps. Distinctively, U-Net employs upsampling in the decoder to match feature map sizes, whereas the autoencoder employs pooling. Thus, our approach leverages an adapted U-Net architecture, customarily referred to as a variant of U-Net, for our segmentation task. It's noteworthy to mention that this adaptation while utilizing the Unet framework, does not encompass pre-existing weights or parameters from previous tasks. Consequently, this process diverges from transfer learning, which specifically encompasses the utilization of prior knowledge to enhance performance on new tasks. Accurate segmentation of biomedical images is required to precisely extract the features related to pathophysiological conditions. Manual segmentation is a tedious and time-consuming process that requires an expert. To eliminate the existing problems, a weakly supervised segmentation algorithm is proposed where a cascade of conventional (seed region growing/random walker/K-means clustering) and deep learning method is utilized. The original images were first segmented using a conventional method to obtain the primary segmentation of ROI and then the obtained segmentations are used to train the deep learning network. The training split of dataset 1 is used to train the proposed model while the results are predicted on both dataset 1 and dataset 2 separately. For the experiments presented in the manuscript, dataset 2 is always used for cross-validation only. It can be clearly inferred from the obtained results that the proposed algorithm of weakly supervised segmentation is comparable enough to the state-of-the-art supervised method, used earlier in the literature.