Cardiac MR Scan Segmentation Using U-Net Architectures

Susanna Weber
Biomedical Engineering
Columbia University
NY, USA
smw2251@columbia.edu

***Abstract*—Quantifying the morphology of the heart, such as cardiac chamber volume and myocardial mass, is crucial for diagnosing and monitoring cardiovascular disease. Currently, clinicians rely on manual segmentation of the the heart during analysis, but this is prone to subjective error, variability between clinicians, and is time-consuming. Deep learning has the potential to improve this process through automatic cardiac segmentation. In this report, two U-Net based models are used to segment the right ventricle (RV) of the heart in short-axis view cardiac magnetic resonance (MR) scans, a task which is key to estimating the ejection fraction of the heart. Both models are first trained to segment the entire heart in axial magnitude MR scans of the heart, and these weights are used to initialize the weights for the RV segmentation task.**

I. Introduction

Medical imaging has revolutionized diagnosis and treatment of heart disease, giving clinicians and researchers the ability to both qualitatively and quantitatively evaluate heart health. Understanding the morphology of a patient's heart and its components is crucial for cardiac care- for example, tracking thickening of the walls of the ventricles, or determining the ejection fraction, which is the percentage of blood in the heart pumped with each heartbeat, which can be determined by observing how much the ventricles expand and contract during the cardiac cycle. The reliability of quantitative metrics that determine cardiac function through deformation depend on how accurately the tissues of the heart can be segmented. Manual segmentation is time consuming and requires trained clinicians to spend a significant amount of time going through data. Deep learning segmentation has the potential to provide a fast, automated, and reliable method of segmenting cardiac scans, improving the speed and accuracy of diagnosis for patients.

The focus of this project is magnetic resonance imaging (MRI) cardiac data, since cardiac MR scans are one of the main ways that cardiac disease is diagnosed and monitored. MRI has excellent sensitivity and specificity at detecting ischaemia in the heart, but has lower contrast than computed tomography (CT) scans. Cardiac MR scans are also more time consuming, and require patients to hold their breath for periods of time. In patients who are unable to stay completely still or hold their breath for the required amount of time, images may be blurry or contain artifacts. In these cases, it is especially useful to have a robust segmentation pipeline that is able to identify key features even in lower-quality data.

U-Net based architectures were used to segment first the entire heart in fairly low-resolution MR cardiac magnitude scans. Then, using the weights from the whole-heart model, a network was trained to segment the right ventricle of the heart in a separate set of cardiac cine scans. A vanilla U-Net, as described in [1] for biomedical image segmentation, is used as the baseline for segmentation accuracy. This is then compared to the Attention U-Net first used in [2] to segment the pancreas in abdominal CT scans. This architecture seemed a natural fit for this project because its purpose is to be able to automatically identify relevant parts of an object in different images, even if the object varies in shape

and size, as the heart does. .

II. Background

A. Cardiac MRI

Cardiac MRI is a standard tool for diagnosing cardiac disease, and has excellent sensitivity and specificity in detecting cardiac ischaemia [3]. Unlike CT imaging, MRI is a non-ionizing modality, which is a key advantage in treating younger patients with cardiovascular disease, who will require regular scans over the course of their life to monitor disease progression. The datasets in this report include two different types of cardiac MR scans. The first dataset are thoracic scans acquired with axial or "top-down" slicing. The second dataset consists of short axis cine scans, which means the image slices are perpendicular to the long axis of the heart. This view is commonly used in MR cardiac imaging as it provides a view of both the left and right ventricles.

B. Ventricle Segmentation

Segmentation of the left and right ventricles is particularly important because the volume of the ventricles is used to estimate ejection fraction, the percentage of blood in the heart that is pumped out when the heart relaxes and then contracts during the cardiac cycle. The left ventricle is very identifiable, in 1, it is the circular area surrounded by a dark wall of tissue. The right ventricle is the crescent-shaped area adjacent to it. In general, segmentation of the ventricles can be challenging because of overlap of pixel intensity between the ventricles and the surrounding area, the variation in shape during the cardiac cycle, and the time and expense of acquiring high-resolution cardiac MRI data in general limiting data set size. Segmenting the right ventricle is a more difficult task than segmenting the left ventricle since it is irregularly shaped and has much thinner walls, giving it a less clearly defined border.

III. Approach

My approach to this project has changed several times based on the results from the models. Initially, the project was only going to focus on whole heart segmentation, since I had access to an annotated whole-heart dataset. However, the models I used achieved a 90% accuracy within a very small number of epochs, indicating that applying the existing models to a more difficult task would be a more interesting project than developing more complicated models for whole-heart segmentation. Therefore, the focus of my project changed to be on using the weights from the model trained on the simpler task (whole heart segmentation) to train on ventricle segmentation data. This approach is interesting from a research perspective since many high resolution cardiac datasets are quite small, including the one used for this project. The whole-heart segmentation data is low-resolution and faster to acquire, so this approach is useful in setting where it is not possible to create a large, high-quality dataset, but a model can be pre-trained on data that is easier and less expensive to get. This approach is also interesting in terms of limiting the number of parameters and model complexity when performing a more complicated task.

IV. Architectures

A. U-Net

The U-Net was first introduce in [1] as an architecture for medical image segmentation. It consists of an encoder, a decoder, and several skip connections that concatenate information from the encoder and decoder. The encoding path is also referred to as the contracting path, because its purpose is to capture contextual information about the input while reducing spatial resolution. It consists of a series of convolutions, similar to the feedforward layers in a connvolutional neural network (CNN). The decoder, or expansive path, decodes the lower spatial resolution data and combines it with the information it receives from the skip connections to generate the segmentation map. It upsamples the feature maps while using the skip connections to preserve spatial information that would be lost in the contracting steps otherwise. The U-Net used in this project takes in a 1 channel (gray scale) image, and output a segmentation mask of the same size as the original input. A binary cross entropy loss (BCE) function was used, and the DICE score was used to determine model accuracy.

B. Attention U-Net

The attention U-Net is variation of the U-Net introduced in [2]. It utilizes the concept of network attention that was first described in [4]. One downside of U-Nets is that while the skip-connections give valuable spatial information to the decoder, since the feature representations of the model are very poor in the shallower levels, the skip-connections also send over irrelevant or redundant information. The authors of [2] implement "soft attention" at the skip connections, meaning areas of high relevance in the image are weighted more heavily than areas with low relevance. As the model trains and the weights are refined, the model learns to only focus on the parts of the image that contain relevant information for the segmentation task. This model also used BCE loss and DICE scores for evaluation of performance.

V. Datasets

The whole heart segmentation dataset contains 79 3D axial magnitude scans, each consisting of 36 64x64 slices. Each slice has a manually annotated mask covering the whole heart, creating a dataset of 2844 2D axial slices. The data were separated into training, validation, and test data using a 70/20/10 split. All data were acquired by members of the Juchem lab at Columbia University using a 3T GE MRI scanner.

The dataset of annotated MRI cine short axis slices was sourced from the The Right Ventricle Segmentation Challenge. The dataset consists of 243 physician-segmented scans showing contours for the epicardium (outside wall of the right ventricle) and endocardium (interior wall of the right ventricle). Each scan is 216x256 pixels in size. This dataset was split into 166 training scans, 47 validation scans, and 30 test scans. Data were acquired on 16 patients on a 1.5 T Siemens scanner.

An example of the annotatations and masks can be seen in Figure 1. Before training, data were normalized and randomly augmented by horizontal or vertical flips. Initially, data were also augmented through random cropping of the images, but this consistently produced both worse validation and test DICE scores.

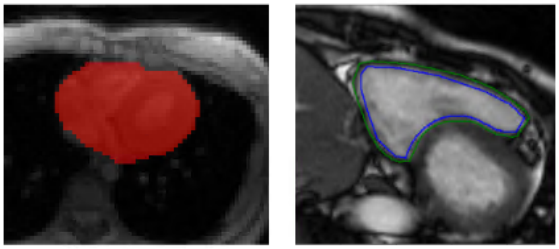

Fig. 1. An axial scan with a whole heart mask (left) and a short axis scan with endocardium and epicardium contours (right).

VI. Results

First a standard 2D U-Net was trained on the whole heart mask dataset. This model was trained for 30 epochs, but reached an validation DICE score 0.90 within 5 epochs. As discussed earlier, this is because U-Nets generally perform well at "blob" segmentation tasks, and applying this network to a more challenging shape is more interesting. The final validation DICE score was 0.93, and the test DICE score was 0.94 and the weights from this model were stored to be used to initialize weights for the right ventricle segmentation task. The U-Net architectures was also modified for 3D inputs and output, to see how well the model could perform with larger inputs. The model performed slightly worse at this task, reaching a final validation accuracy of 0.90. However, training this model was less practical than the 2D variant, because of the high computational load.

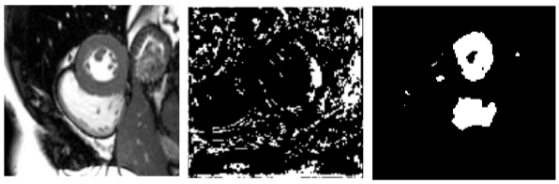

Fig. 2. From left to right are the image, the predicted masks without pre-trained weights, and the predicted mask with pre-trained weights. Note that the original dataset contains only the contours for the right ventricle, but these are transformed into masks for training.

The 2D attention U-Net was also trained on the whole heart segmentation for 30 epochs. This model benefitted significantly from increased data augmentation, as it initially overfit the data, and the validation loss increased over the course of training. The final validation score for this model was 92%, and the final test score was 93%. Overall, the more complicated model only performed on par with the basic U-Net, indicating that for this particular task a basic U-Net is sufficient. Therefore, a 3D version of the attention net was not explored.

Next, the model was applied to the more difficult task of right ventricle segmentation. Since there are two masks in this case (epicardium and endocardium) the model was trained on the two datasets separately, both times using the weights loaded from the whole-heart iteration of the model. The initial predictions of the for the epicardium mask using the whole heart weights from the U-Net and attention U-Net are shown in Figure 2, both without pre-training and with pre-training. While the pre-training mask is of not correct, there are features from the original scan that seem to stand out, while the prediction without pre-training is entirely random.

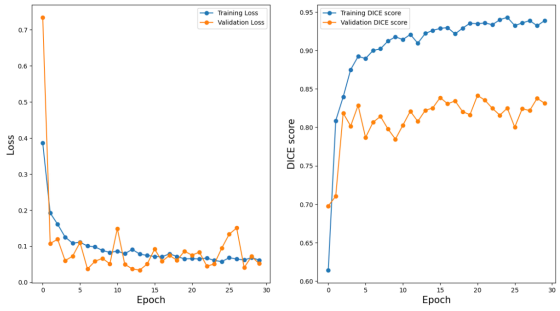

Fig. 3. Binary cross entropy loss and DICE score for training and validation on the epicardium contours.

Here, the model with the pre-trained weights and the model without are compared. Scores are averaged between epicardium and endocardium training. For the not pre-trained U-Net, the validation DICE score is 0.76, and the test DICE score is 0.77. With pre-trained weights, the validation DICE score is 0.78, and the test DICE score is 0.80. While the results look good qualitatively, only 80% of the true mask is being captured. Using the not pre-trained attention U-Net, the validation DICE score is 0.81, and the test DICE score is 0.82. Using the attention model with pre-trained weights, the validation DICE score is 0.82, and the test DICE score is 0.83. There is not a very large improvement from un-trained to pre-trained weight models, and potential reasons for this are discussed in the conclusion. Epicardium mask training curves for the attention U-Net and an example of predicted contours are shown in Figures 3 and 4, respectively. It is worth noting the epicardium and endocardium contours significantly overlap - the model is able to identify the general shape of the right ventricle, but it is not really sophisticated enough to distinguish between the inner and outer part of the wall. There is not a very large improvement from un-trained to pre-trained weight models, and potential reasons for this are discussed in the conclusion.

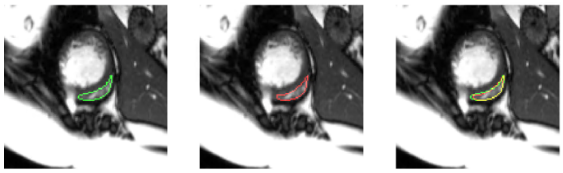

Fig. 4. From left to right: Predicted endocardium contours, epicardium contours, and both contours, overlayed on scan.

VII. Conclusions and Discussion

A U-Net and an attention U-Net were pretrained on a whole heart segmentation task, and then applied to right ventricle segmentation, a more difficult but more clinically relevant task. The attention U-Net model achieved validation and test DICE scores above 0.80 for the latter task. Clearly, there is still room for improvement. One important point is that the use of pre-trained weights provides a small improvement, but for the attention U-Net, the improvement is negligible. This could indicate that for the attention net, the pre-training task is too different from the right ventricle segmentation task to provide an advantage. Other U-Net architectures

that incorporate attention mechanisms, such as dilated U-Nets [5], might perform better. Additionally, pre-training on a more similar but still easier task, such as left ventricle segmentation, could also be beneficial. Experimenting with other loss functions could also improve training. Overall, this approach shows the value of using pre-training on large but easy to acquire datasets in order to improve performance for medical image segmentation.

ACKNOWLEDGMENT

Thank you to Yun Shang for providing the axially slices cardiac MR scans.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, arXiv:1505.04597 [cs] version: 1, May 2015. DOI: 10.48550/arXiv.1505.04597. [Online]. Available: http://arxiv.org/abs/1505.04597 (visited on 04/02/2024).

[2] O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, *Attention U-Net: Learning Where to Look for the Pancreas*, arXiv:1804.03999 [cs], May 2018. DOI: 10.48550/arXiv.1804.03999. [Online]. Available: http://arxiv.org/abs/1804.03999 (visited on 04/13/2024).

[3] M. Dewey, t. l. w. o. i. a. n. t. Link to external site, M. Siebes, *et al.*, "Clinical quantitative cardiac imaging for the assessment of myocardial ischaemia," English, *Nature Reviews. Cardiology*, vol. 17, no. 7, pp. 427–450, Jul. 2020, Num Pages: 427-450 Place: London, United States Publisher: Nature Publishing Group, ISSN: 17595002. DOI: 10.1038/s41569-020-0341-8. [Online]. Available: https://www.proquest.com/docview/2413790383/abstract/7AFA031715904922PQ/1 (visited on 11/17/2023).

[4] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, *Attention Is All You Need*, arXiv:1706.03762 [cs], Aug. 2023. DOI: 10.48550/arXiv.1706.03762. [Online]. Available: http://arxiv.org/abs/1706.03762 (visited on 05/07/2024).

[5] D. Saadati, O. N. Manzari, and S. Mirzakuchaki, *Dilated-UNet: A Fast and Accurate Medical Image Segmentation Approach using a Dilated Transformer and U-Net Architecture*, en, Apr. 2023. [Online]. Available: https://arxiv.org/abs/2304.11450v1 (visited on 05/08/2024).