Deep learning Cardiac Cine MRI Segmentation

Zhehao Shen*
Shanghaitech University
2021533110
shenzhh@shanghaitech.edu.cn

Yihui Cao*
Shanghaitech University
2021533029
caoyh3@shanghaitech.edu.cn

Yiqing Zhang*
Shanghaitech University
2022591020
zhangyq22022@shanghaitech.edu.cn

Zhijie Huang*
Shanghaitech University
2020533147
huangzhj1@shanghaitech.edu.cn

ACM Reference Format:
Zhehao Shen*, Yihui Cao*, Yiqing Zhang*, and Zhijie Huang*. 2024. Deep learning Cardiac Cine MRI Segmentation.

1 INTRODUCTION

Cardiac cine MRI is a non-invasive imaging technique that captures dynamic sequences of the heart, providing valuable information on cardiac structure and function. Accurate segmentation of cardiac structures, including the left ventricle (LV), right ventricle (RV), and myocardium (MYO), is critical for diagnosing cardiovascular diseases, assessing treatment efficacy, and planning surgical interventions. Manual segmentation, however, is time-consuming and subject to inter-observer variability, necessitating the development of automated methods to achieve consistent and accurate results.

Deep learning, particularly convolutional neural networks (CNNs) [3], has revolutionized image analysis tasks, including medical image segmentation. Among various architectures, the U-Net has gained prominence due to its effectiveness in biomedical image segmentation. The U-Net architecture features a symmetric encoder-decoder structure with skip connections that allow the network to capture both local and global context, which is essential for precise segmentation.

In this project, we focus on leveraging the U-Net architecture to perform segmentation of cardiac cine MRI images. The primary goals are to design a robust U-Net model, evaluate its segmentation performance using cross-entropy loss and Dice coefficient metrics, and investigate the impact of various modifications and training strategies on the model's performance.

Our specific objectives are as follows:

- Designing the U-Net Architecture: We will develop a customized U-Net model tailored for cardiac cine MRI segmentation. The model's architecture will be defined by key parameters such as the number of channels in the first convolutional layer and the depth of the encoder. These parameters are crucial as they influence the model's capacity to learn intricate features from the input images.
- Baseline Evaluation: The initial U-Net model will be trained using the cross-entropy loss function. We will monitor the training process by plotting the training and validation losses and evaluate the model's performance on a test set using the Dice coefficient for LV, RV, and MYO.
- Impact of Architectural Changes: To understand the role of architectural components, we will modify the U-Net by removing the skip connections and retrain the model. This experiment aims to reveal how skip connections contribute to the segmentation performance.
- Effect of Data Augmentation: Data augmentation techniques are employed to artificially expand the training dataset, thereby improving the model's generalization ability. We will train the U-Net with augmented data and compare its performance to the baseline model to assess the benefits of data augmentation.
- Alternative Loss Functions: The choice of loss function significantly impacts the training dynamics and final performance of the model. We will replace the cross-entropy loss with the soft Dice loss, which directly optimizes the Dice coefficient. This experiment will help us determine whether optimizing for the segmentation metric itself yields better results.
- Exploring Advanced Techniques: Beyond the standard U-Net, we will explore more sophisticated neural network architectures and loss functions from recent literature. This additional investigation aims to push the boundaries of segmentation accuracy and identify potential improvements over the baseline model.

By systematically conducting these experiments, we aim to gain a comprehensive understanding of the factors that influence the performance of deep learning models for cardiac cine MRI segmentation. The insights gained from this project will contribute to the development of more accurate and reliable automated segmentation tools, ultimately aiding in the advancement of cardiac healthcare.

2 METHOD

2.1 Loading the Data

The dataset utilized for this project originates from the publicly available ACDC cardiac dataset. The cine MRI images and segmentation labels were loaded from the provided .npz file. The data consists of multiple slices of MRI images, each labeled with the respective segmentations for LV, RV, and MYO.

2.2 Data Preparation

The dataset utilized for this project originates from the publicly available ACDC cardiac dataset. This dataset provides comprehensive cardiac cine MRI images along with corresponding segmentation labels for three key cardiac structures: the left ventricle (LV),

right ventricle (RV), and myocardium (MYO). However, some samples in this dataset lack complete labels for the left ventricle(LV), right ventricle or(RV) myocardium(MYO). These samples were excluded, and the refined dataset will be used for subsequent training.

2.3 Data Splitting

The dataset was divided into three subsets: training, validation, and testing, with a split ratio of 4/7, 1/7, and 2/7, respectively. This splitting ensures that the model has a robust set of images to learn from, a separate set to validate and tune hyperparameters, and an unseen set to evaluate the final performance.

2.4 Model Architecture

The U-Net architecture was chosen for this project due to its effectiveness in medical image segmentation tasks. The U-Net is composed of an encoder-decoder structure with symmetric skip connections, which help retain spatial information throughout the network. The design specifics of our U-Net model include:

2.4.1 Encoder. The encoder consists of multiple convolutional layers and max-pooling layers. At each downsampling stage, a max-pooling layer is first applied to reduce the spatial dimensions, followed by two convolutional operations. Each convolutional layer is followed by a batch normalization layer and a ReLU activation function. The number of channels in the first convolutional layer is set to 64 and doubles in each subsequent layer, gradually increasing the depth and complexity of the feature maps.

2.4.2 Decoder. The decoder mirrors the encoder but uses upsampling layers instead of max-pooling layers. The upsampling layers, implemented via transposed convolutions, restore the spatial resolution of the feature maps. After upsampling, two convolutional operations are performed to further process the feature maps. In each layer of the decoder, the upsampled feature map is concatenated with the corresponding feature map from the encoder to retain more high-resolution information.

2.4.3 Skip Connetions. Skip connections play a critical role in the decoder. They concatenate the feature maps from the encoder with the upsampled feature maps in the decoder. These connections help preserve high-resolution features during the decoding process, thereby enhancing segmentation accuracy. This design allows the network to leverage both high-level abstract features and low-level detailed features simultaneously.

2.4.4 Output Layer. The final layer uses a 1x1 convolution to map the high-dimensional feature maps to the desired number of output channels (3 channels for LV, RV, and MYO). A softmax activation function is applied to produce probability maps for each class.

2.5 Training Procedure

The U-Net model was trained with the following configurations to optimize its performance

2.5.1 Loss Function. Cross-entropy loss was employed as the primary loss function for training the initial U-Net model. This loss function is suitable for multi-class classification problems and helps in minimizing the difference between predicted and true labels.

2.5.2 Optimizer. The Adam optimizer was chosen due to its adaptive learning rate capabilities, which efficiently handles sparse gradients and requires minimal tuning.

2.5.3 Learning Rate. An initial learning rate of 0.001 was set, with potential adjustments based on the performance observed during training.

2.5.4 Batch Size and Epochs. The model was trained with a batch size of 10, balancing computational efficiency and learning stability. Training was conducted over 50 epochs, allowing sufficient time for the model to converge.

2.5.5 Regularization and Callbacks. Early stopping was implemented to halt training if the validation loss did not improve for a set number of epochs, preventing overfitting.

2.5.6 Data Augmentation. Various data augmentation techniques, such as rotation, scaling, and flipping, were applied to the training set to increase data variability and enhance model generalization.

2.6 Evaluation Metrics

The model's performance was rigorously evaluated using the Dice coefficient, a metric that quantifies the overlap between predicted and ground truth segmentations. The evaluation process included:

2.6.1 Dice Coefficient Calculation. The Dice coefficient was computed for each of the three cardiac structures (LV, RV, and MYO) on the testing set. This metric ranges from 0 to 1, with 1 indicating perfect overlap and 0 indicating no overlap.

2.6.2 Visualization. Segmentation results were visually inspected by overlaying predicted masks on the original MRI images to qualitatively assess the model's performance.

3 EXPERIMENTS

To thoroughly understand the impact of various factors on the model's performance, a series of experiments were conducted.

3.1 Baseline U-Net with Cross-Entropy Loss

The initial experiment involved training the U-Net with the cross-entropy loss function to establish a baseline performance.

The U-Net model was trained over 200 epochs using the cross-entropy loss function and the Adam optimizer. Training and validation losses were recorded for each epoch to monitor the model's convergence.

During training, we recorded the training and validation losses over the epochs and plotted the loss curves as shown in Figure 1 and Figure 2

These figure shows the change in training and validation losses over the 50 epochs and 200 epochs. As training progressed, both training and validation losses decreased and stabilized, indicating that the model was converging.

On the test set, we calculated the Dice coefficients for the left ventricle (LV), right ventricle (RV), and myocardium (MYO). The results, including the mean and standard deviation of the Dice coefficients, are presented below:

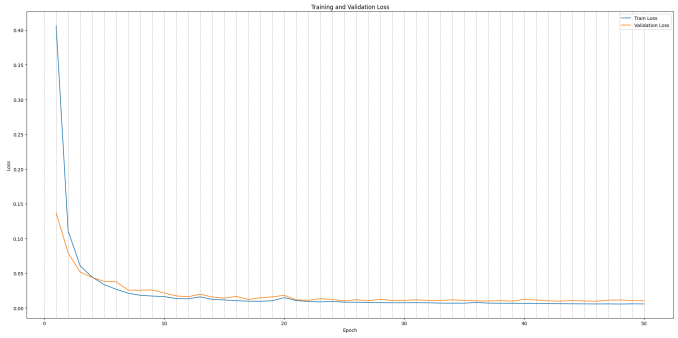


Figure 1: (a)50epoch loss curve

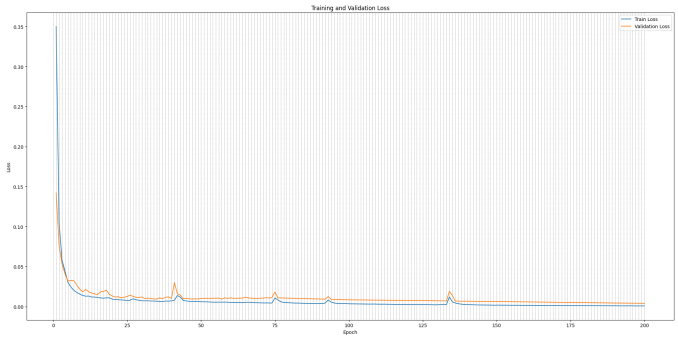


Figure 2: (a)200epoch loss curve

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.8463	0.2217
Right Ventricle (RV)	0.8427	0.1927
Myocardium (MYO)	0.8248	0.1762

Table 1: 50 epoch performance of the baseline U-Net model

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.9096	0.1452
Right Ventricle (RV)	0.8763	0.1655
Myocardium (MYO)	0.8667	0.1418

Table 2: 200 epoch performance of the baseline U-Net model

The results show that the U-Net model performs exceptionally well in segmenting cardiac cine MRI images, with the highest performance observed for the left ventricle, followed by the myocardium and right ventricle. This indicates that the U-Net architecture is highly effective in preserving the details of cardiac structures.

3.2 U-Net without Shortcut Connections

In this experiment, the skip connections in the U-Net were removed to evaluate their importance. The modified model was trained and its performance was compared to the baseline.

During training, the training and validation losses were recorded over the epochs and plotted as shown in Figure 3 and Figure 4

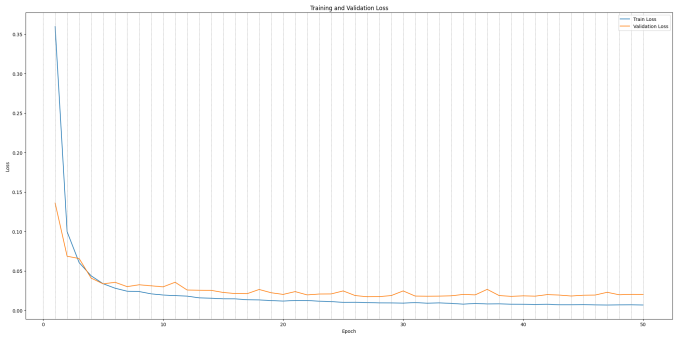


Figure 3: (b)50epoch loss curve

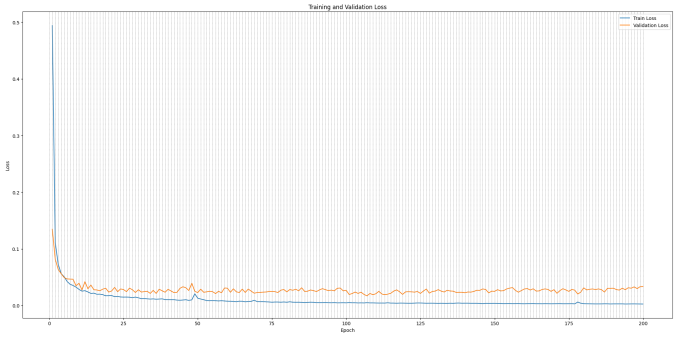


Figure 4: (b)200epoch loss curve

These figure shows the change in training and validation losses over the 50 epochs and 200 epochs. Compared to the baseline model, the loss curves provide insights into how the removal of skip connections affects the training process.

On the test set, the Dice coefficients for the left ventricle (LV), right ventricle (RV), and myocardium (MYO) were calculated. The results, including the mean and standard deviation of the Dice coefficients, are presented below:

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.8050	0.2082
Right Ventricle (RV)	0.7764	0.2255
Myocardium (MYO)	0.6798	0.2341

Table 3: 50 epoch performance of the U-Net model without short connections

The results indicate that the U-Net model without skip connections exhibits a decline in segmentation performance compared to the baseline model. Specifically, the mean Dice coefficients for the left ventricle, right ventricle, and myocardium are lower, and the standard deviations are higher, suggesting that skip connections play a crucial role in retaining spatial information and achieving accurate segmentation.

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.8313	0.1914
Right Ventricle (RV)	0.7991	0.1362
Myocardium (MYO)	0.8533	0.1503

Table 4: 200 epoch performance of the U-Net model without short connections

The decline in performance when removing skip connections can be attributed to several factors:

- Loss of Detailed Information: Skip connections in the U-Net architecture help in preserving high-resolution details by directly connecting corresponding layers in the encoder and decoder. This direct transfer of feature maps helps the decoder reconstruct fine details in the segmentation masks. Without these connections, the model struggles to retain and use detailed spatial information, leading to less accurate segmentations.
- Contextual Information: The skip connections allow the model to access both local and global context simultaneously. This is particularly important in medical image segmentation, where accurate boundary delineation requires understanding both fine-grained details and broader anatomical structures. Without skip connections, the model may miss important contextual cues, reducing segmentation accuracy.

3.3 U-Net with Data Augmentations

Data augmentation techniques were applied to the training data to assess their impact on model generalization. The performance of the augmented model was compared to the baseline.

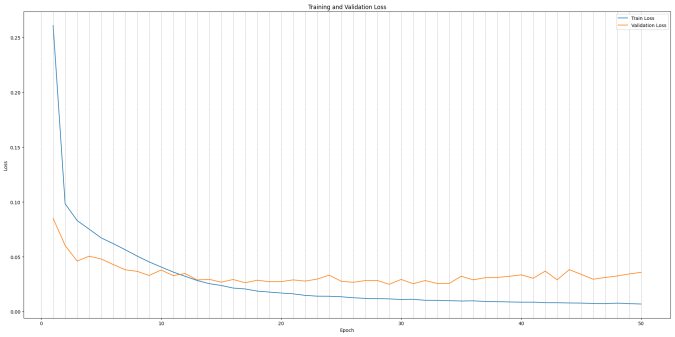


Figure 5: (c)50epoch loss curve with data augmentation

During the training process of the U-Net model, a notable increase in the validation loss is observed starting from the 50th epoch, indicating the onset of overfitting. This phenomenon is evident in the loss curves provided, where the training loss continues to decrease steadily while the validation loss begins to rise. Overfitting occurs when the model starts to learn the noise and intricate details in the training data to an extent that negatively impacts its performance on new, unseen data. This is typically characterized by a divergence between the training and validation loss curves, as depicted in the figures.

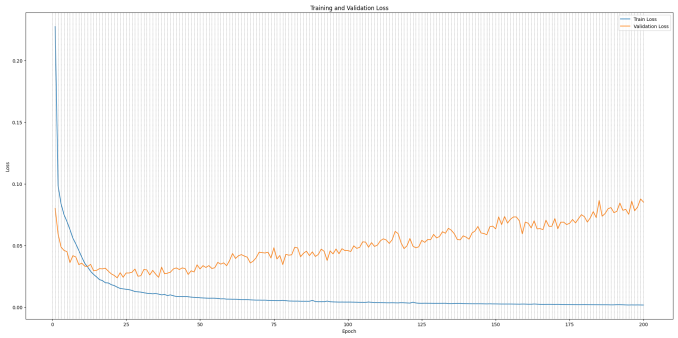


Figure 6: (c)200epoch loss curve with data augmentation

In the initial phase, up to approximately 50 epochs, both training and validation losses decrease, indicating that the model is learning relevant patterns. However, beyond this point, the validation loss starts to increase despite the continuous decline in training loss. This divergence suggests that the model becomes overly specialized to the training data, thereby losing its generalization ability. To address this issue, it is necessary to implement regularization techniques or early stopping to mitigate overfitting and enhance the model's generalization performance.

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.7900	0.2491
Right Ventricle (RV)	0.7505	0.2507
Myocardium (MYO)	0.7080	0.2528

Table 5: 50 epoch performance of the U-Net model with data augmentation

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.8896	0.1433
Right Ventricle (RV)	0.7877	0.2496
Myocardium (MYO)	0.8086	0.1842

Table 6: 200 epoch performance of the U-Net model with data augmentation

Despite the increase in validation loss, the Dice coefficient continues to improve. This suggests that the model is still learning to better segment the relevant structures. Upon closer examination of several samples, it is evident that as the number of training epochs increases, the model indeed starts to learn noise, leading to the appearance of undesired segmented regions. However, it is also apparent that the correctly segmented areas are becoming more accurate. This indicates that the model is enhancing its ability to segment the intended structures, even though it is also capturing noise. Thus, while the overall generalization performance as indicated by the validation loss may decline, the segmentation quality for the correctly identified regions improves, reflecting in the higher Dice coefficient.

3.4 U-Net with Soft Dice Loss

The loss function was changed to the soft Dice loss, which directly optimizes the Dice coefficient.

$$\text{Multiclass Soft Dice Loss} = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{2\sum_{i=1}^{N} p_{i,c}g_{i,c} + \epsilon}{\sum_{i=1}^{N} p_{i,c}^2 + \sum_{i=1}^{N} g_{i,c}^2 + \epsilon}$$

C: The total number of classes in the classification task.
$p_{i,c}$: The prediction of the i-th sample belonging to the c-th class(0 or 1).
$g_{i,c}$: The actual label of the i-th sample for the c-th class(0 or 1).
ϵ: A small smoothing term used to prevent division by zero.In this project it is set to 1.

The model was trained and its performance was compared to the model trained with cross-entropy loss.

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.8056	0.2875
Right Ventricle (RV)	0.8473	0.2087
Myocardium (MYO)	0.8258	0.1830

Table 7: 50 epoch performance of the U-Net model with soft dice loss

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.9199	0.1268
Right Ventricle (RV)	0.8776	0.1378
Myocardium (MYO)	0.8955	0.0610

Table 8: 200 epoch performance of the U-Net model with soft dice loss

Upon training for 200 epochs, overfitting was not observed. Notably, the average Dice coefficient increased, and the standard deviation decreased. However, the loss curve exhibited greater fluctuations. This discrepancy can be attributed to the inherent differences between cross-entropy loss and Dice loss.

Cross-entropy loss treats each pixel as an independent sample for prediction, optimizing the model on a pixel-by-pixel basis. This approach often results in a smoother loss curve but may not adequately capture the spatial relationships and overall structure within the segmentation task. In contrast, Dice loss considers the prediction output in a more holistic manner. By evaluating the overlap between the predicted segmentation and the ground truth, Dice loss takes into account the overall shape and continuity of the structures being segmented.

This holistic view enables Dice loss to better capture the overall structure of the segmentation, leading to improved performance metrics, such as a higher average Dice coefficient and a lower standard deviation, despite the more erratic loss curve. The fluctuations in the loss curve suggest that while Dice loss is effective in improving segmentation accuracy, it may introduce variability during the optimization process. This variability is indicative of the model's adjustments to better align with the true spatial configurations of the segmented structures, ultimately resulting in more accurate and reliable segmentation outcomes.

In summary, the training results indicate that Dice loss, by emphasizing the overall prediction structure, enhances segmentation performance even if it leads to a more irregular loss curve. This highlights the importance of selecting appropriate loss functions tailored to the specific goals of the segmentation task, balancing between pixel-level accuracy and the preservation of structural integrity.

4 ADVANCED TECHNIQUES

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.941	0.096
Right Ventricle (RV)	0.904	0.157
Myocardium (MYO)	0.908	0.123

Table 9: 200 epoch performance of the U-Net model with weighted cross entropy loss and soft dice loss

4.1 Weighted Cross Entropy Loss

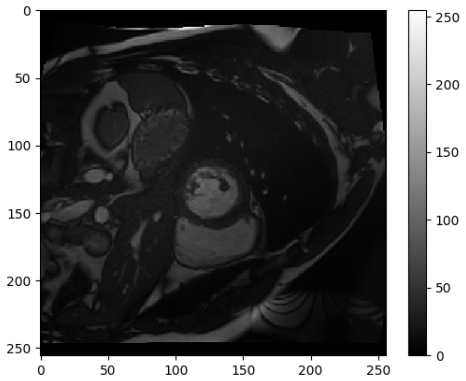
Inspired by [1], we believe that the foreground and background regions should not share the same weights when computing the cross-entropy loss. The foreground, due to its intricate edges and smaller area, deserves a higher weight in classification. Therefore, we used a weight of 0.1 for the background class and 0.3 for the foreground classes (MYO, LV, RV) to achieve better results. The numerical results are shown in Table 9.

4.2 Add SAM Encoder

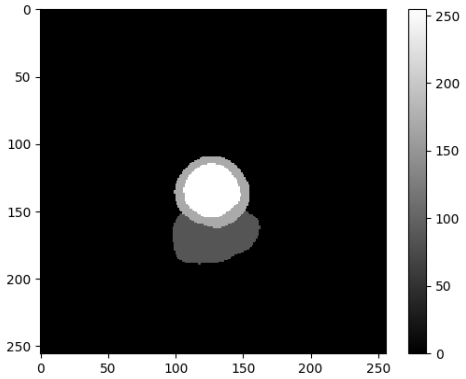
Segment Anything [2] demonstrates transcendent segmentation performance across a variety of extensive tasks. Thus, I aim to leverage such a robust model to enhance our segmentation capabilities. SAM itself evaluates pixel scores through an image encoder and a prompt-based mask encoder. Due to the high memory demand required to finetune the entire SAM model directly, we adopted a cost-effective finetuning approach, which is shown in Fig. 10. We froze the parameters of SAM's image encoder and attached a U-net based CNN decoder, focusing primarily on optimizing the decoder's parameters. The training was conducted on an A6000 GPU, consuming approximately 40GB of memory. Based on the results, the SAM model finetuned through our method did not achieve satisfactory performance, as seen in Fig. **??**. We speculate that this may be due to the absence of heart-related training data in SAM's original dataset, which might hinder its ability to encode the correct features accurately. The numerical results are shown in Table 10.

Structure	Mean Dice Coefficient	Std Deviation
Left Ventricle (LV)	0.216	0.396
Right Ventricle (RV)	0.404	0.257
Myocardium (MYO)	0.715	0.193

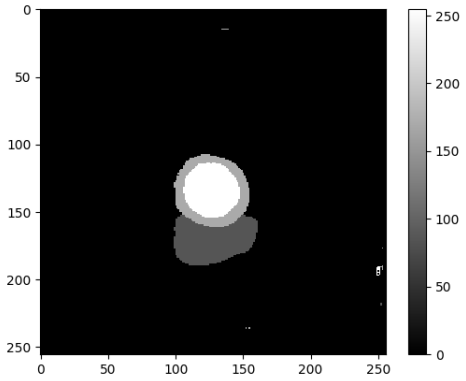
Table 10: 200 epoch performance of the modified SAM model



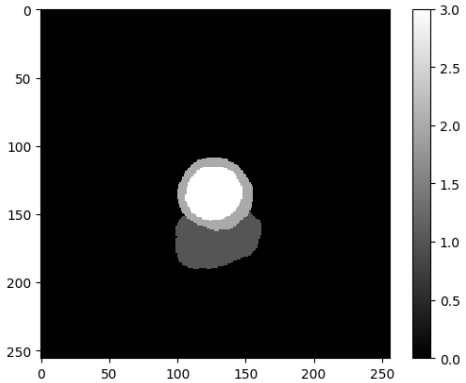
(a) Origin



(b) 50epochs



(c) 200epochs



(d) Label

Figure 7: One sample with data augmentation

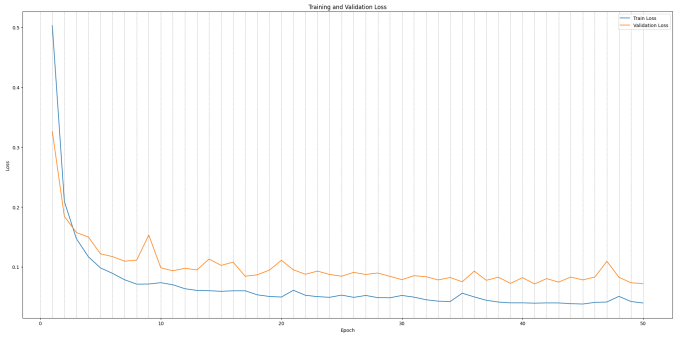


Figure 8: (d)50epoch loss curve with soft dice loss

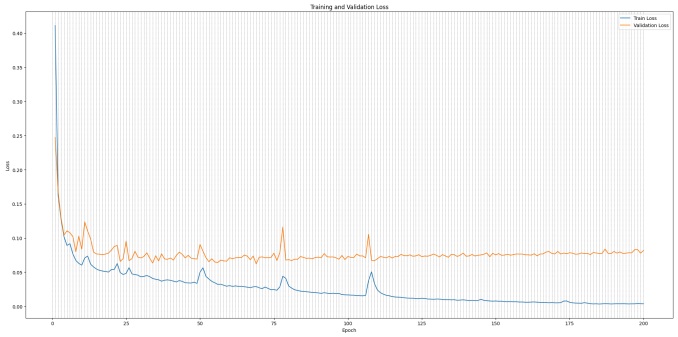


Figure 9: (d)200epoch loss curve with soft dice loss

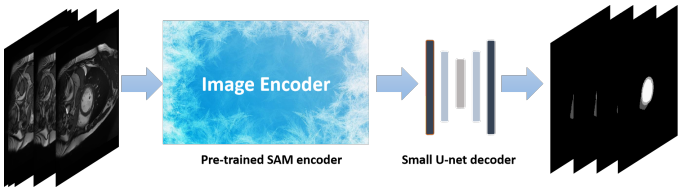


Figure 10: Our modified pipeline

REFERENCES

[1] Christian F Baumgartner, Lisa M Koch, Marc Pollefeys, and Ender Konukoglu. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. *arXiv preprint arXiv:1709.04496*, 2017.

[2] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.