
BME1312 Project2 Report: Deep Learning Cardiac Cine MRI Segmentation

Wenye Xiong
2023533141

xiongwy2023@shanghaitech.edu.cn

Renyi Yang
2023533030

yangry2023@shanghaitech.edu.cn

Jiaxing Wu
2023533160

wujx2023@shanghaitech.edu.cn

Boyang Xia
2023533073

xiaby2023@shanghaitech.edu.cn

Fengmin Yang
2023533183

yangfm2023@shanghaitech.edu.cn

Abstract

This report details the implementation and evaluation of U-Net based deep learning models for cardiac segmentation in cine MRI scans. The project focuses on segmenting three key structures: the Left Ventricle (LV), Right Ventricle (RV), and Myocardium (MYO). We explore the standard U-Net architecture, the impact of removing skip connections, the effect of data augmentation, the performance difference between Binary Cross-Entropy loss and Soft Dice loss, and potential improvements using an Attention U-Net.

1 Introduction

This project aims to:

1. Implement a baseline U-Net model for segmenting LV, RV, and MYO from cardiac cine MRI images.
2. Investigate the role of skip connections in the U-Net architecture.
3. Evaluate the impact of data augmentation on segmentation performance.
4. Compare Binary Cross-Entropy loss with Soft Dice loss for training the segmentation model.
5. Explore improvements to the U-Net architecture, such as incorporating attention mechanisms.

2 Baseline U-Net

2.1 Structure

The U-Net architecture, as depicted in Figure 1, consists of:

- **DoubleConv**: A block of two sequential (Conv2D 3x3, BatchNorm2D, ReLU) operations.
- **Down**: Max pooling (2x2) followed by a DoubleConv block for downsampling in the encoder.

- **Up:** Upsampling (bilinear or transpose convolution) followed by concatenation with features from the corresponding encoder layer (skip connection) and a DoubleConv block. Padding is used to handle potential size mismatches during concatenation.

Our baseline is as below:

- **Network:** The standard UNet class as described above, with C_base=32, 1 input channel, and 3 output classes. Bilinear upsampling was used.
- **Loss Function:** A custom MyBinaryCrossEntropy loss was used. This involves applying a Sigmoid function to the model's output logits to get probabilities, then computing nn.BCELoss against the 3-channel binary ground truth masks. The learning rate was 0.01, and training was run for 50 epochs.
- **Evaluation:** Mean and standard deviation of the Dice Similarity Coefficient (DSC) for LV, RV, and MYO were calculated on the test set. Training and validation loss curves were plotted. Example segmentation results were saved.

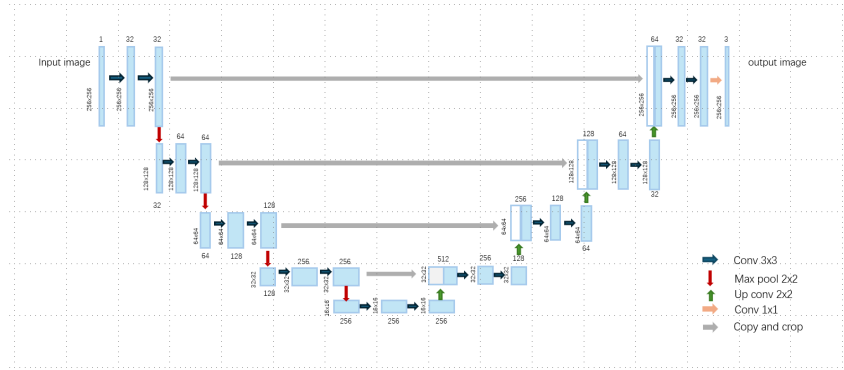


Figure 1: U-Net Network Architecture.

2.2 Experiments and Results

The training loss and validation loss of the baseline U-Net are shown in Figure 2.

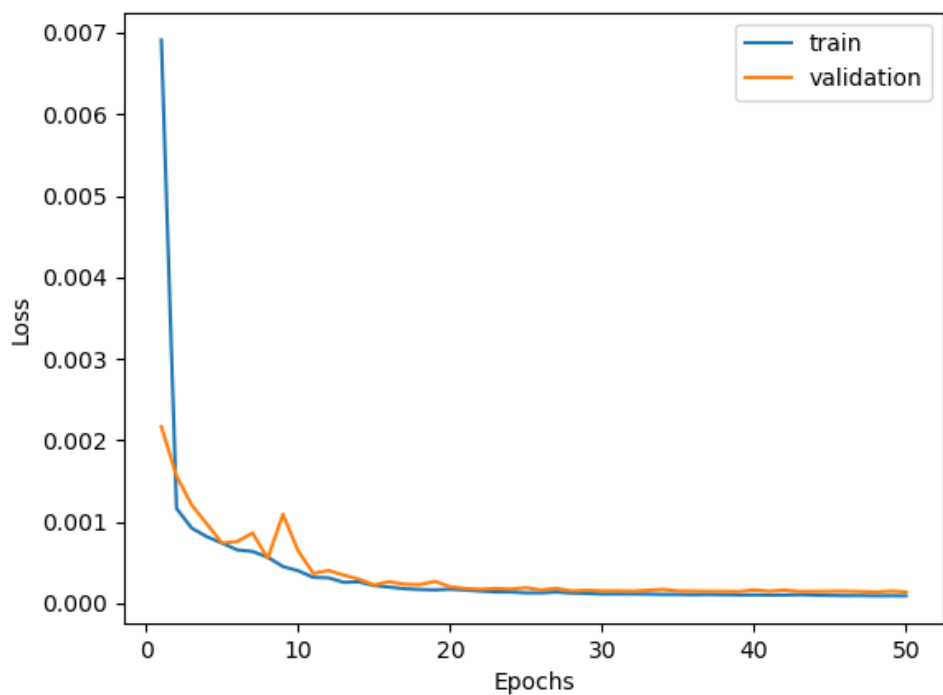


Figure 2: Training and Validation Loss Curves for the Baseline U-Net

And here are example segmentation results in Figure 3, 4, and 5.

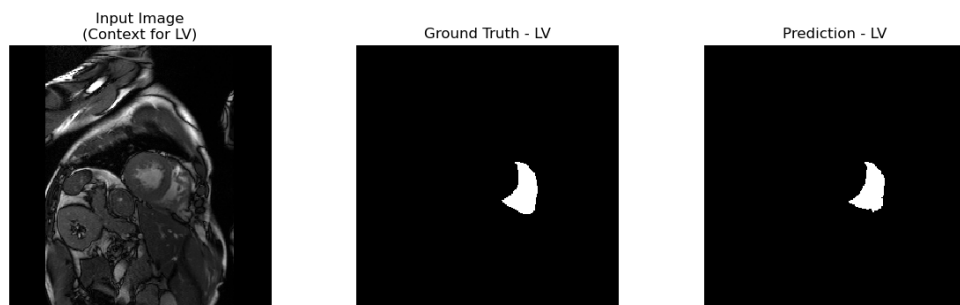


Figure 3: Baseline U-Net Segmentation Example: LV

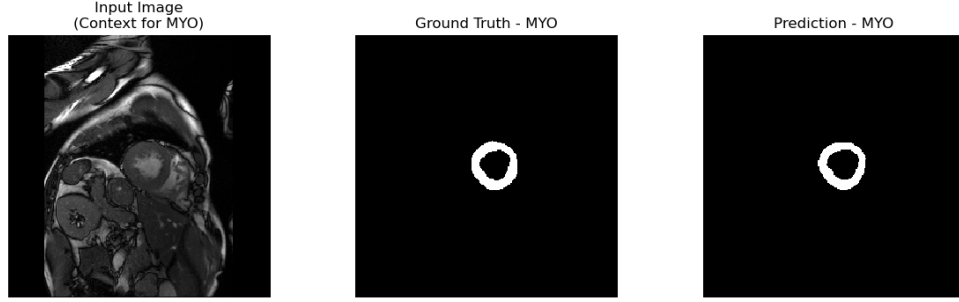


Figure 4: Baseline U-Net Segmentation Example: MYO

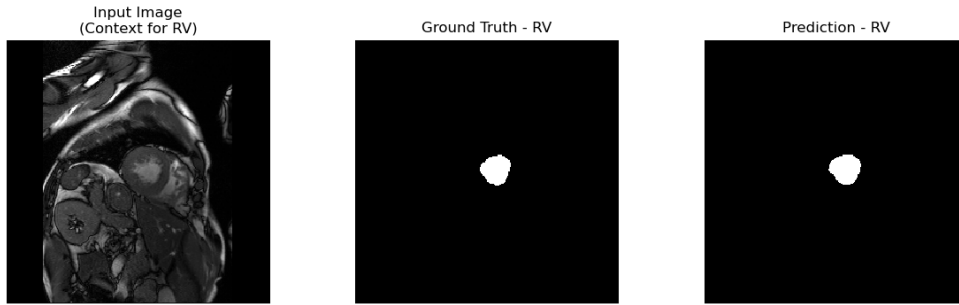


Figure 5: Baseline U-Net Segmentation Example: RV

The baseline U-Net trained with Cross-Entropy loss achieved the following Dice scores (Table 1):

Table 1: Dice Coefficients for Baseline U-Net (Task a)

Structure	Mean Dice	Standard Deviation
RV	0.9519	0.0086
MYO	0.8734	0.0161
LV	0.8920	0.0310

2.3 Discussion

The baseline U-Net (Table 1) demonstrated strong segmentation performance.

- **RV Segmentation:** Achieved the highest mean Dice score (0.9519). This is often expected as the RV is typically a large, relatively well-defined structure with good contrast against surrounding tissues in many MRI sequences.
- **LV Segmentation:** Also showed good performance with a mean Dice of 0.8920. The LV cavity is usually clearly visible.
- **MYO Segmentation:** Had the lowest mean Dice score (0.8734). The myocardium is a thinner, more complex structure surrounding the LV, and its boundaries, especially with the LV cavity (endocardium) and epicardium, can be more challenging to delineate accurately, potentially leading to lower overlap scores.
- The standard deviations are relatively small, indicating consistent performance across the test slices. Overall, the baseline U-Net provides a solid foundation for cardiac segmentation.

3 U-Net without Skip Connections

3.1 Structure

- **Network Modification:** A UNet_NoShortcut model was implemented. This involved creating an Up_NoShortcut module that performs upsampling and convolution but does not concatenate features from the encoder path. The forward method of Up_NoShortcut only processes the feature map from the previous decoder layer.
- **Retraining:** This modified U-Net was trained following the same procedure as the baseline U-Net: 50 epochs, Adam optimizer, learning rate of 0.01, and MyBinaryCrossEntropy loss.

3.2 Performance

The training and validation loss curves for the U-Net without skip connections are shown in Figure 6.

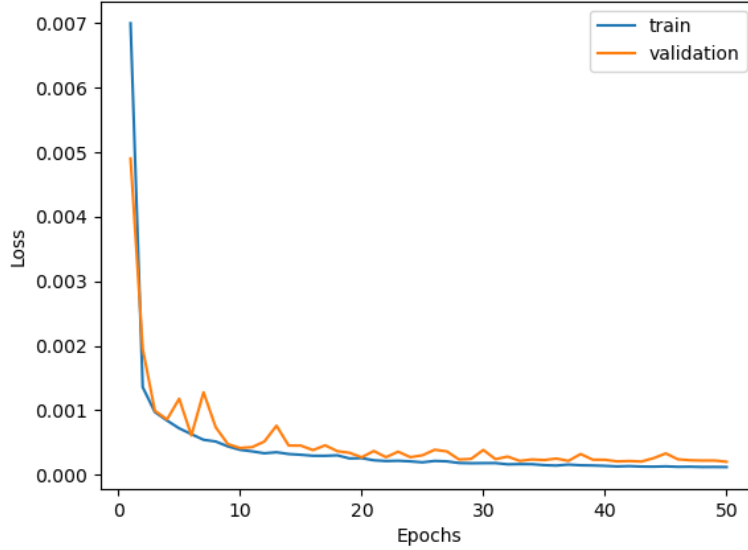


Figure 6: Training and Validation Loss for U-Net without Skip Connections.

The U-Net variant without skip connections, trained under identical conditions, yielded the Dice coefficients shown in Table 2.

Table 2: Dice Coefficients: Baseline U-Net vs. U-Net without Skip Connections (Task b)

Structure	Baseline U-Net (DSC)		U-Net No Shortcut (DSC)	
	Mean Dice	Std. Dev.	Mean Dice	Std. Dev.
RV	0.9519	0.0086	0.9260	0.0111
MYO	0.8734	0.0161	0.8223	0.0168
LV	0.8920	0.0310	0.8588	0.0296

3.3 Discussion

Comparing the performance in Table 2, removing skip connections resulted in a significant drop in performance across all structures.

- All structures showed a noticeable decrease in Dice Similarity Coefficient (DSC).

- **Reason:** Skip connections provide high-resolution spatial information from the encoder to the decoder, which is crucial for accurate boundary localization. They also aid gradient flow during backpropagation, mitigating vanishing gradient problems and improving convergence.
- **Conclusion:** Skip connections are vital for U-Net’s segmentation accuracy in this cardiac MRI segmentation task. The MYO, being the most intricate structure, suffered a substantial relative drop in performance.

4 U-Net with Data Augmentation

4.1 Structure

- **Network:** The baseline U-Net architecture (with skip connections) was used.
- **Data Augmentation Techniques:** Augmentations were applied only to the training dataset using `torchvision.transforms`. The specific augmentations included:
 - `RandomHorizontalFlip`
 - `RandomRotation(15°)`
 - `RandomAffine(degrees=50, translate=(0.1,0.1), scale=(0.9,1.1), shear=5)`

A custom `SegmentationDataset` class was used. In its `__getitem__` method, the image and its corresponding label (both single-channel tensors) were stacked along a new dimension before applying the transformations. This ensures that geometric augmentations are applied identically to both the image and its mask.

- **Retraining:** The U-Net was retrained using the augmented training data for 50 epochs. The validation set remained un-augmented. The loss function was `MyBinaryCrossEntropy`, and the learning rate was 0.01.

4.2 Performance

The training and validation loss curves for the U-Net trained with data augmentation are shown in Figure 7.

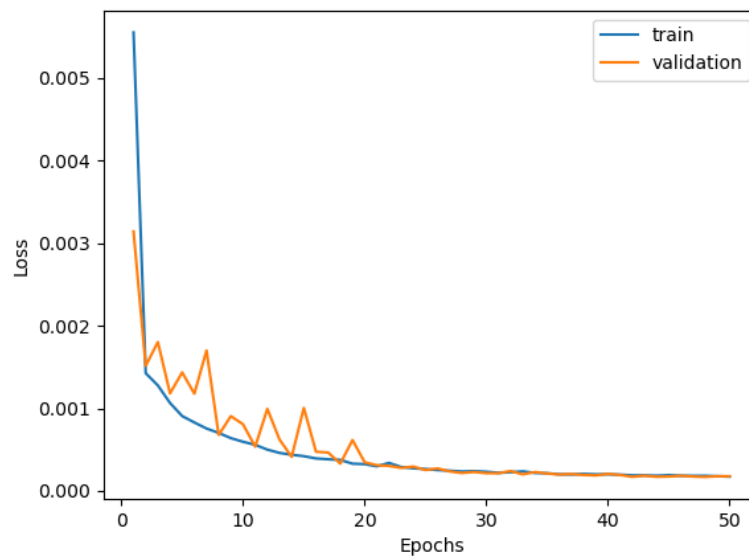


Figure 7: Training and Validation Loss for U-Net with Data Augmentation.

The baseline U-Net architecture trained with data augmentation on the training set resulted in the Dice coefficients shown in Table 3.

Table 3: Dice Coefficients: Baseline U-Net vs. U-Net with Data Augmentation (Task c)

Structure	Baseline U-Net (DSC)		U-Net + Data Aug. (DSC)	
	Mean Dice	Std. Dev.	Mean Dice	Std. Dev.
RV	0.9519	0.0086	0.9276	0.0107
MYO	0.8734	0.0161	0.8469	0.0149
LV	0.8920	0.0310	0.8635	0.0384

4.3 Discussion

As shown in Table 3, the specific data augmentation strategy employed led to a slight decrease in Dice coefficients for all structures compared to the baseline model trained on un-augmented data.

- **Possible Reasons:**

- Some aggressive augmentations (e.g., large rotations or shears in `RandomAffine`) could have distorted anatomical structures or altered their relative locations significantly, making it harder for the model to learn precise boundaries.
- The augmentations might not be perfectly suited for the specific characteristics of cardiac MRI data, where subtle features and consistent anatomy are important.

- **Conclusion:** While data augmentation is often beneficial, its effectiveness is highly dependent on the chosen transformations and the dataset. For this project, the selected augmentations did not improve performance and may have even been detrimental. More careful selection, parameter tuning, or domain-specific augmentations would be necessary to potentially see benefits.

5 U-Net with Soft Dice Loss

5.1 Structure

- **Network:** The baseline U-Net architecture (with skip connections) was used.
- **Training Data:** The original non-augmented training dataset was used for this experiment.
- **Loss Function Modification:** A `SoftDiceLoss` class was implemented. This loss calculates the Dice coefficient directly on the sigmoid probabilities of the model’s output and the multi-channel binary ground truth. The loss is defined as $1 - \text{mean}(\text{Dice_per_class})$. A smoothing factor of 1.0 was added to the numerator and denominator to maintain stability.
- **Retraining:** The U-Net model was trained for 50 epochs using the `SoftDiceLoss`. The Adam optimizer was used with a learning rate of 0.001 and an `ExponentialLR` scheduler.

5.2 Performance

The training and validation loss curves for the U-Net trained with Soft Dice Loss are shown in Figure 8.

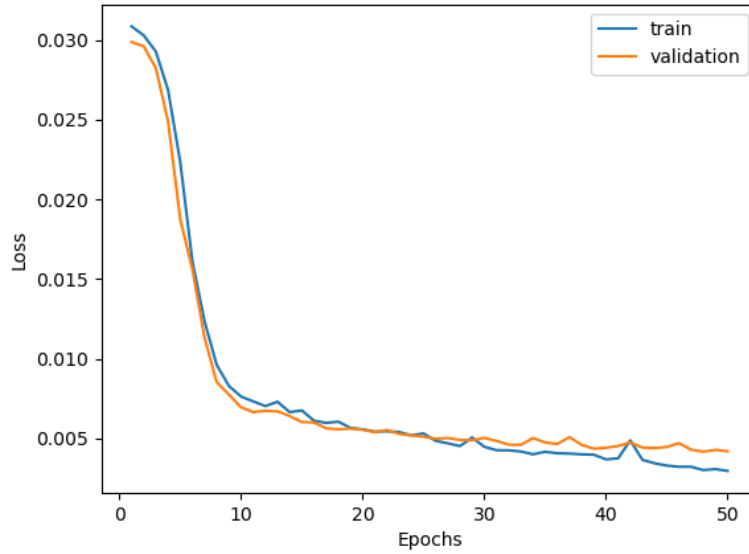


Figure 8: Training and Validation Loss for U-Net with Soft Dice Loss.

The U-Net trained with Soft Dice Loss on the non-augmented dataset achieved the Dice coefficients shown in Table 4.

Table 4: Dice Coefficients: Baseline U-Net (BCE Loss) vs. U-Net (Soft Dice Loss) (Task d)

Structure	Baseline U-Net (BCE Loss)		U-Net (Soft Dice Loss)	
	Mean Dice	Std. Dev.	Mean Dice	Std. Dev.
RV	0.9519	0.0086	0.9566	0.0100
MYO	0.8734	0.0161	0.8962	0.0100
LV	0.8920	0.0310	0.8998	0.0371

Example segmentation results for the U-Net trained with Soft Dice Loss are shown in Figures 9, 10, and 11.

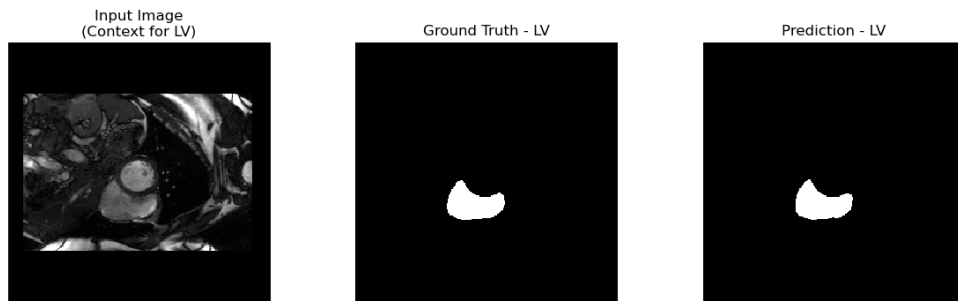


Figure 9: U-Net with Soft Dice Loss Segmentation Example: LV

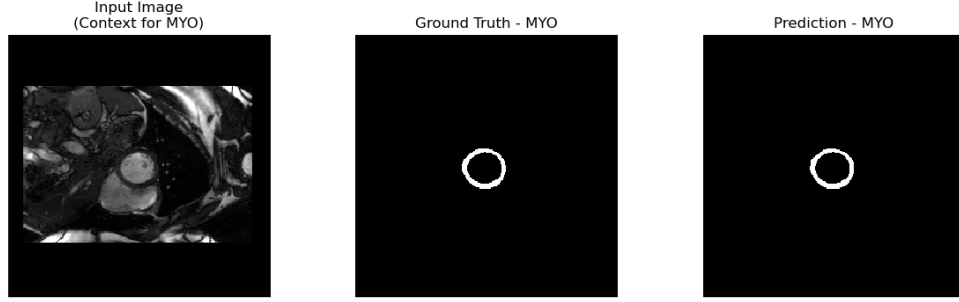


Figure 10: U-Net with Soft Dice Loss Segmentation Example: MYO

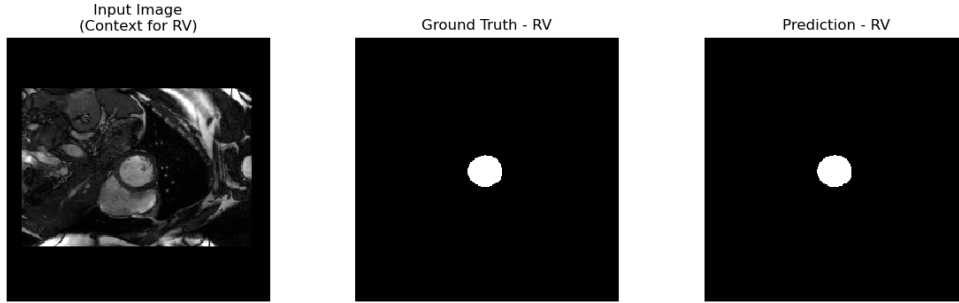


Figure 11: U-Net with Soft Dice Loss Segmentation Example: RV

5.3 Discussion

As shown in Table 4, when trained on the same non-augmented data, the U-Net with Soft Dice Loss significantly outperformed the baseline U-Net with Binary Cross-Entropy (BCE) Loss in terms of Dice Coefficient for all structures.

- The improvement is most notable for MYO segmentation (0.8734 to 0.8962), which is often the most challenging structure.
- LV and RV segmentation also saw improvements.
- This suggests that directly optimizing a Dice-based metric is beneficial for this segmentation task, likely due to its robustness to class imbalance and focus on overlap rather than pixel-wise classification accuracy alone.

6 Improvements: Attention U-Net

6.1 Structure

To further enhance segmentation performance, an Attention U-Net architecture was implemented.

- **Advanced U-Net (Attention U-Net):**
 - **Architecture:** An `AttentionBlock` was introduced in the decoder's Up module. The `AttentionBlock` computes attention coefficients by combining features from the decoder (gating signal) and the corresponding skip connection from the encoder. These coefficients are then applied to the encoder features, effectively allowing the model to focus on more relevant spatial regions during the upsampling and feature fusion process. This helps in refining segmentation boundaries, especially for complex structures.
- **Loss Function:** `SoftDiceLoss` was used, given its superior performance in the previous experiment.

- **Optimizer:** Adam optimizer with a learning rate of 0.001 and an ExponentialLR scheduler.
- **Training:** The model was trained for 50 epochs on the non-augmented training dataset.

6.2 Performance

The Attention U-Net trained with Soft Dice Loss achieved the Dice coefficients shown in Table 5.

Table 5: Dice Coefficients: Baseline (BCE), Baseline (Soft Dice Loss), and Attention U-Net (Soft Dice Loss) (Task e)

Structure	Baseline U-Net (BCE)		U-Net (Soft Dice)		Attention U-Net (Soft Dice)	
	Mean Dice	Std. Dev.	Mean Dice	Std. Dev.	Mean Dice	Std. Dev.
RV	0.9519	0.0086	0.9566	0.0100	0.9588	0.0086
MYO	0.8734	0.0161	0.8962	0.0100	0.8967	0.0109
LV	0.8920	0.0310	0.8998	0.0371	0.9072	0.0292

Example segmentation results for the Attention U-Net are shown in Figures 12, 13, and 14.

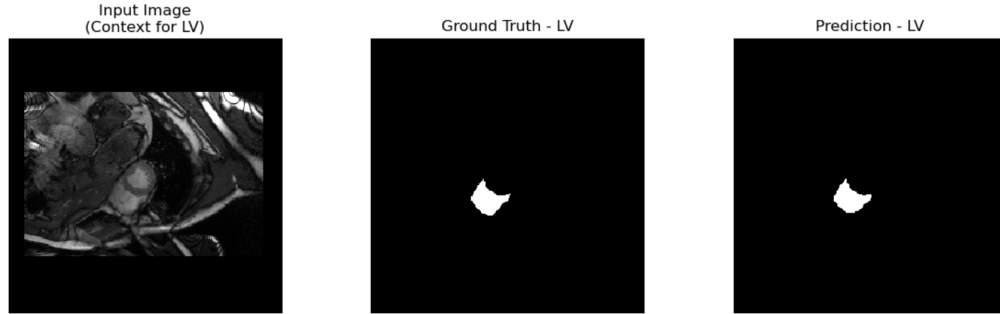


Figure 12: Attention U-Net Segmentation Example: LV

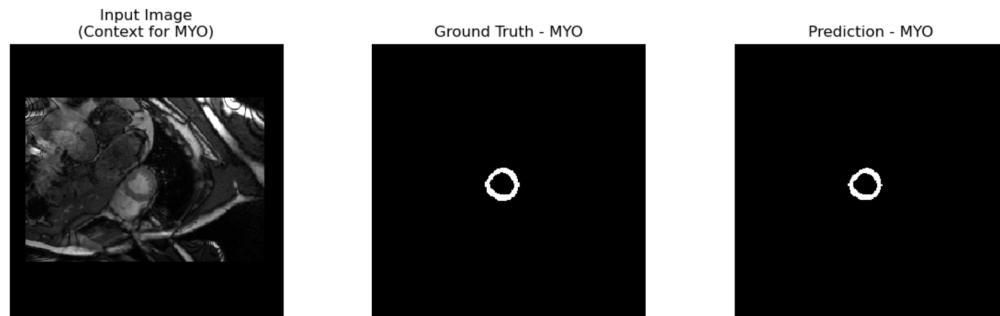


Figure 13: Attention U-Net Segmentation Example: MYO

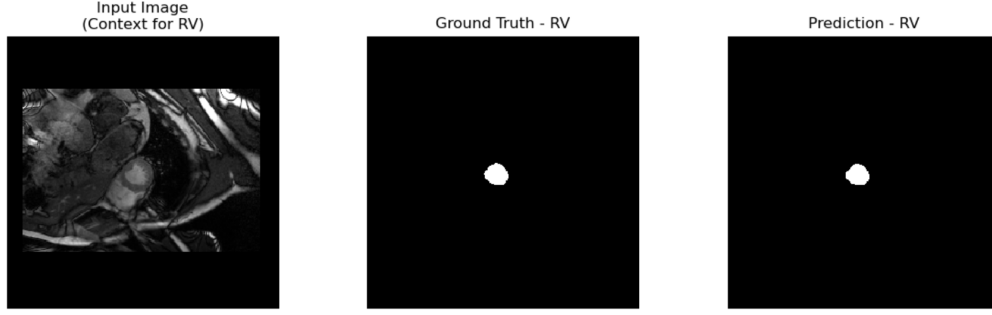


Figure 14: Attention U-Net Segmentation Example: RV

6.3 Discussion

The Attention U-Net, when combined with Soft Dice Loss and trained on non-augmented data, demonstrated further improvements in Dice scores compared to both the baseline U-Net with BCE loss and the U-Net with Soft Dice Loss (Table 5).

- The most significant improvement was observed for LV segmentation (0.8998 to 0.9072). RV and MYO also showed slight gains.
- This suggests that the attention mechanism effectively helps the model to focus on more salient features and complex structural details, leading to better boundary delineation and overall segmentation accuracy.
- The standard deviations for the Attention U-Net are comparable or slightly lower for some structures, indicating robust performance.

7 Overall Performance Summary

Table 6 provides a consolidated view of the mean Dice Similarity Coefficients (DSC) for all evaluated models across the three cardiac structures.

Table 6: Overall Performance Summary: Mean Dice Coefficients

Model Configuration	RV Mean DSC	MYO Mean DSC	LV Mean DSC
(a) Baseline U-Net (BCE Loss)	0.9519	0.8734	0.8920
(b) U-Net without Skip Connections (BCE Loss)	0.9260	0.8223	0.8588
(c) U-Net + Data Augmentation (BCE Loss)	0.9276	0.8469	0.8635
(d) U-Net (Soft Dice Loss, No Augmentation)	0.9566	0.8962	0.8998
(e) Attention U-Net (Soft Dice Loss, No Aug.)	0.9588	0.8967	0.9072

The results clearly indicate that the Attention U-Net trained with Soft Dice Loss on the original non-augmented dataset achieved the highest performance for all three cardiac structures.

8 Conclusion

This project successfully implemented and evaluated several U-Net based models for cardiac cine MRI segmentation. Key findings include:

1. The baseline U-Net provides a strong starting point for segmenting RV, LV, and MYO, with RV generally being the best-segmented structure.
2. Skip connections are essential for U-Net’s performance in this task. Their removal led to a significant decline in Dice scores, confirming their importance for feature reuse and precise localization.

3. The specific data augmentation strategy employed in this study (random flips, rotations, and affine transformations) resulted in a decrease in Dice coefficients compared to training on non-augmented data. This highlights that augmentation strategies must be carefully chosen and tuned for the specific dataset and task.
4. Training with Soft Dice Loss on non-augmented data significantly improved performance over Binary Cross-Entropy loss, particularly for the more challenging MYO structure. This indicates the benefit of directly optimizing a segmentation-specific metric.
5. The Attention U-Net, when combined with Soft Dice Loss and trained on non-augmented data, yielded the best overall segmentation performance across all three structures (RV: 0.9588, MYO: 0.8967, LV: 0.9072). This demonstrates the effectiveness of attention mechanisms in helping the model focus on relevant features for improved boundary delineation.

Future work could explore more sophisticated data augmentation techniques, investigate other loss functions, or experiment with different attention mechanisms or network backbones.