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, Right Ventricle, and Myocardium. 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 and Hybrid Loss.

1 Introduction

The primary goal of this project is to segment key cardiac structures, namely the Left Ventricle, Myocardium, and Right Ventricle, from cine MRI scans. A significant challenge lies in the accurate and robust delineation of these structures, which can exhibit considerable variability in shape and appearance across different patients and image acquisitions.

To address this, we employ a U-Net based deep learning framework. Our approach involves several stages:

- Implementing a baseline U-Net model.
- Investigating the impact of removing U-Net skip connections.
- Evaluating the effect of data augmentation techniques.
- Comparing the performance of Binary Cross-Entropy (BCE) loss versus Soft Dice Loss.
- Exploring improvements to the U-Net architecture by incorporating attention mechanisms (Attention U-Net) and a Hybrid Loss function.

The performance of all models is primarily evaluated using the Dice Similarity Coefficient (DSC).

2 Data Preprocessing

Prior to training, all input images underwent **MinMax** normalization to ensure consistent pixel intensity ranges across the dataset. This preprocessing step is crucial for deep learning models as it:

- Accelerates convergence during training by maintaining input values within a consistent range
- Prevents gradient vanishing/exploding problems that can occur with unnormalized intensity values
- Ensures stability across different MRI acquisition parameters and patient variations

3 Baseline U-Net

3.1 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 RV, LV, and MYO were calculated on the test set. Training and validation loss curves were plotted. Example segmentation results were saved.

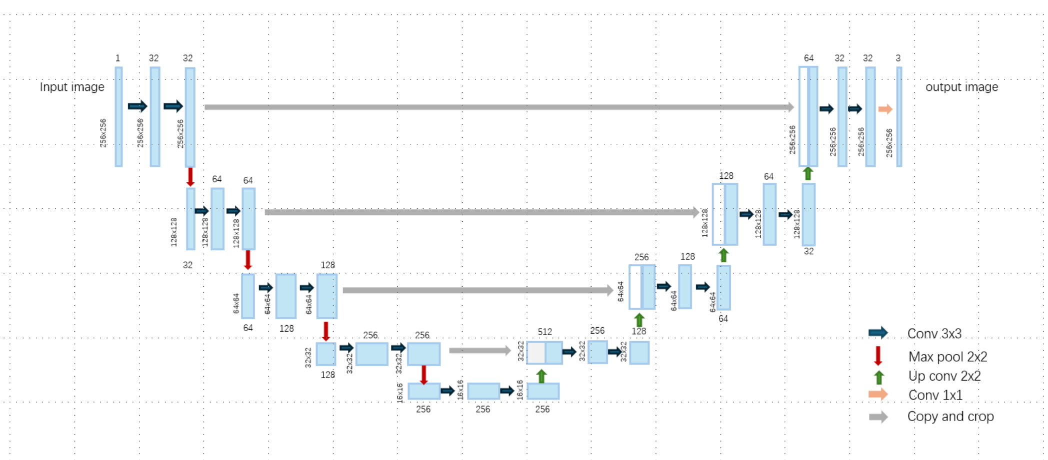

Figure 1: U-Net Network Architecture.

3.2 Experiments and Results

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

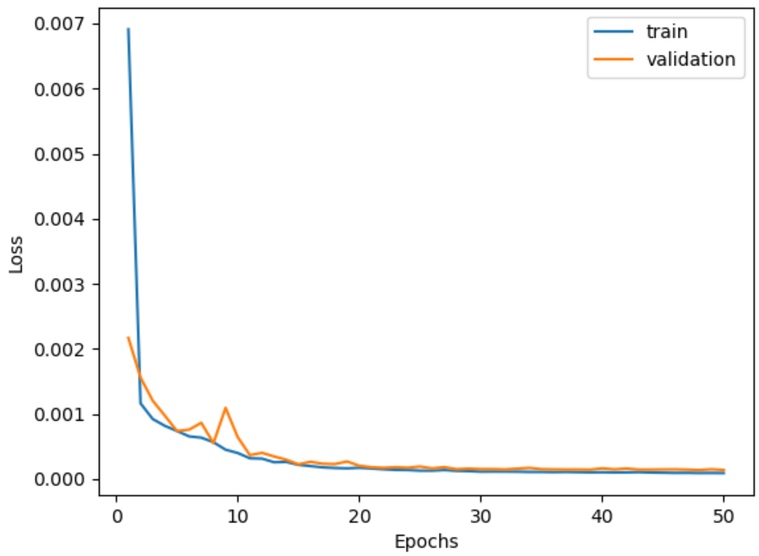

Figure 2: Training and Validation Loss Curves for the Baseline U-Net

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

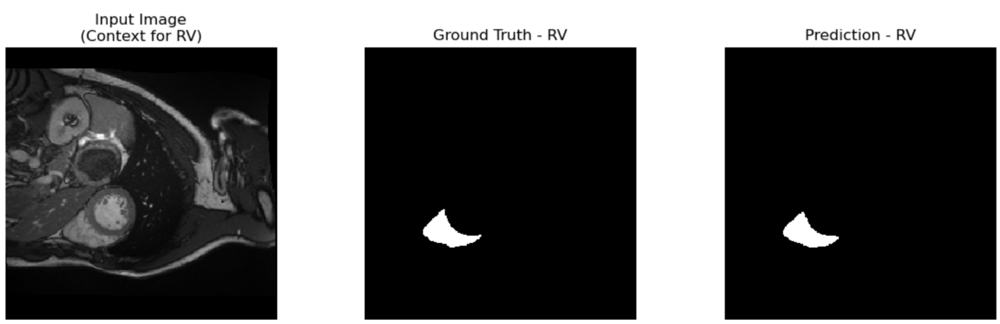

Figure 3: Baseline U-Net Segmentation Example: RV

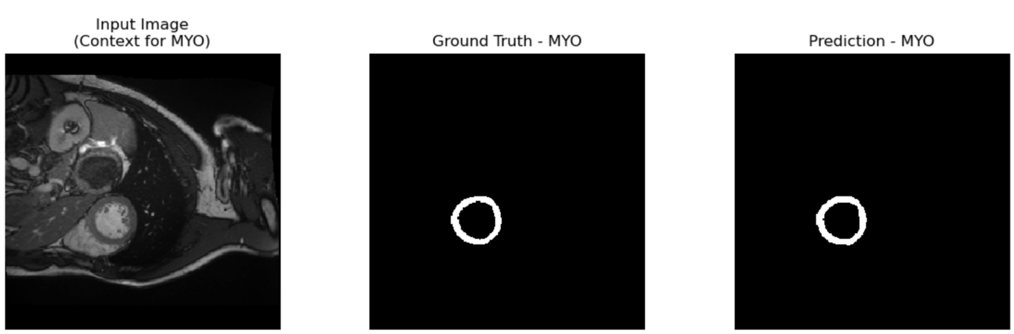

Figure 4: Baseline U-Net Segmentation Example: MYO

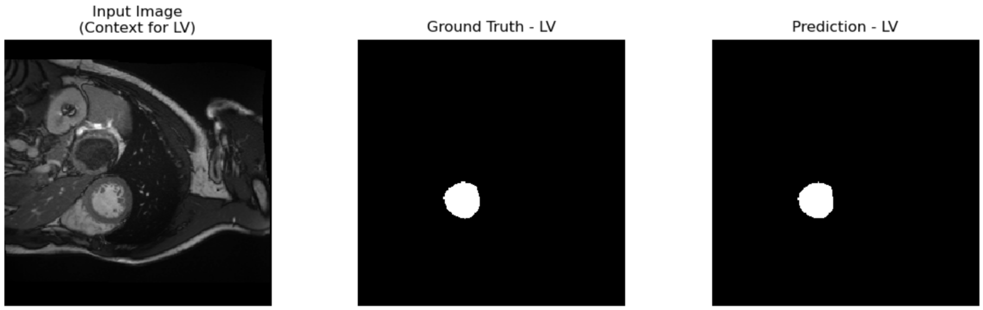

Figure 5: Baseline U-Net Segmentation Example: LV

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
LV	0.9519	0.0086
MYO	0.8734	0.0161
RV	0.8920	0.0310

3.3 Discussion

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

- **LV Segmentation**: Achieved the highest mean Dice score (0.9519). This is often expected as the LV is typically a large, relatively well-defined structure with good contrast against surrounding tissues in many MRI sequences.
- **RV Segmentation**: Also showed good performance with a mean Dice of 0.8920. The RV 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 RV, and its boundaries, especially with the RV 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.

4 U-Net without Skip Connections

4.1 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.
- **Purpose**: Evaluate the importance of skip connections.

4.2 Performance

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

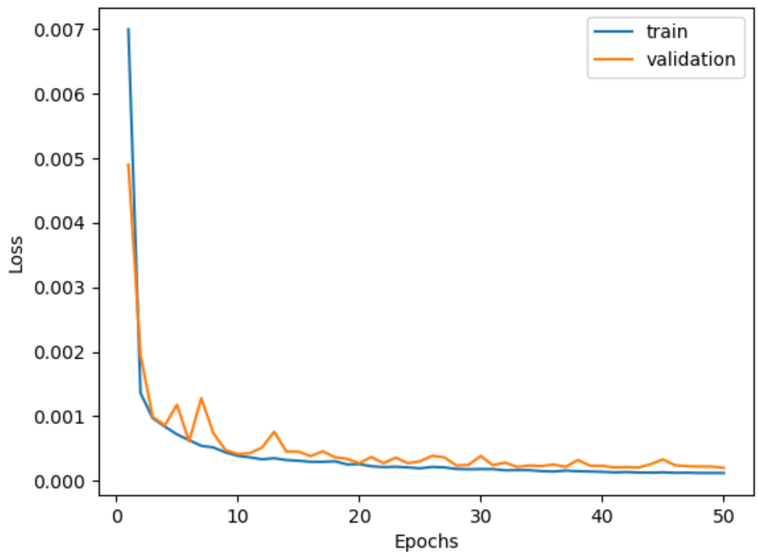

Figure 6: Training and Validation Loss for U-Net without Skip Connections.

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.
LV	0.9519	0.0086	0.9260	0.0111
MYO	0.8734	0.0161	0.8223	0.0168
RV	0.8920	0.0310	0.8588	0.0296

4.3 Discussion

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

- **Significant Drop in Performance**: All structures showed a noticeable decrease in 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.

5 U-Net with Data Augmentation

5.1 Structure

- **Network**: The baseline U-Net architecture (with skip connections) was used.
- **Data Augmentation Techniques (Training Set Only)**: 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, using BCE Loss and a learning rate of 0.01. The validation set remained un-augmented.

5.2 Performance

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

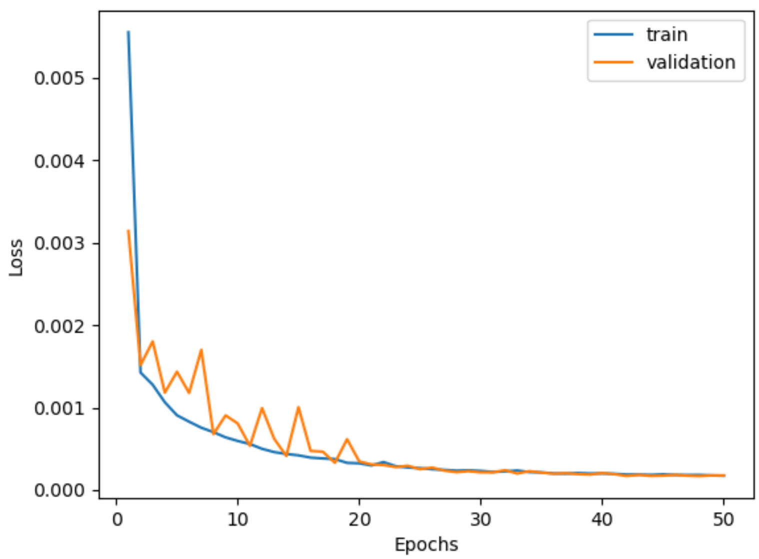

Figure 7: Training and Validation Loss for U-Net with Data Augmentation.

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.
LV	0.9519	0.0086	0.9276	0.0107
MYO	0.8734	0.0161	0.8469	0.0149
RV	0.8920	0.0310	0.8635	0.0384

5.3 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.

- **DSC Decrease**: The specific augmentation strategy led to slightly lower Dice scores.
- **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. Maybe the relative location of structures was altered too much.
- **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. The relative location of structures is crucial for segmentation tasks, and the specific augmentations used may not have been beneficial for this dataset. More careful selection, parameter tuning, or domain-specific augmentations would be necessary to potentially see benefits.

6 U-Net with Soft Dice Loss

6.1 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 $1 - \text{mean(Dice\_per\_class)}$. 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.

6.2 Performance

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

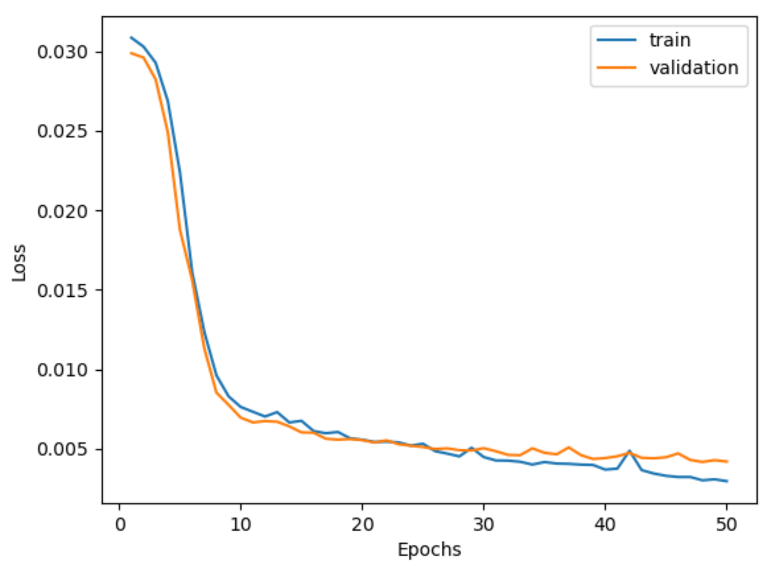

Figure 8: Training and Validation Loss for U-Net with Soft Dice Loss.

The U-Net trained with Soft Dice Loss on the non-augmented dataset achieved the Dice coefficients shown in Table 4 and pixel-wise accuracies in Table 5.

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.
LV	0.9519	0.0086	**0.9566**	0.0100
MYO	0.8734	0.0161	**0.8962**	0.0100
RV	0.8920	0.0310	**0.8998**	0.0371

Table 5: Pixel-wise Accuracy: Baseline U-Net (BCE Loss) vs. U-Net (Soft Dice Loss)

Structure	Baseline U-Net (BCE Loss)		U-Net (Soft Dice Loss)	
	Mean Accuracy	Std. Dev.	Mean Accuracy	Std. Dev.
LV	0.9991	0.0002	**0.9992**	0.0002
MYO	0.9977	0.0003	**0.9980**	0.0002
RV	0.9983	0.0005	0.9983	0.0006

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

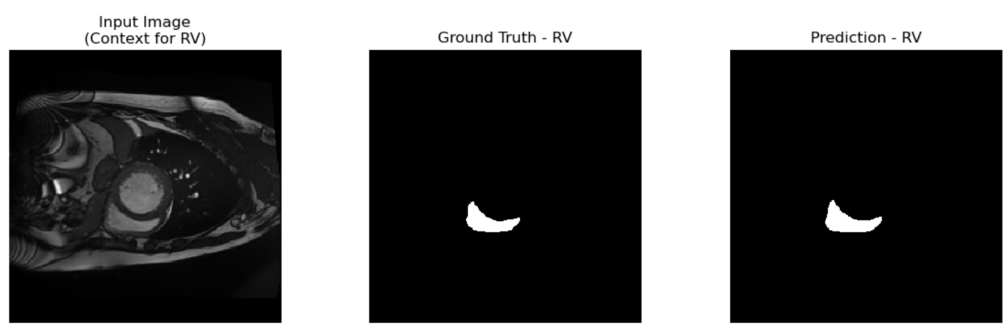

Figure 9: U-Net with Soft Dice Loss Segmentation Example: RV

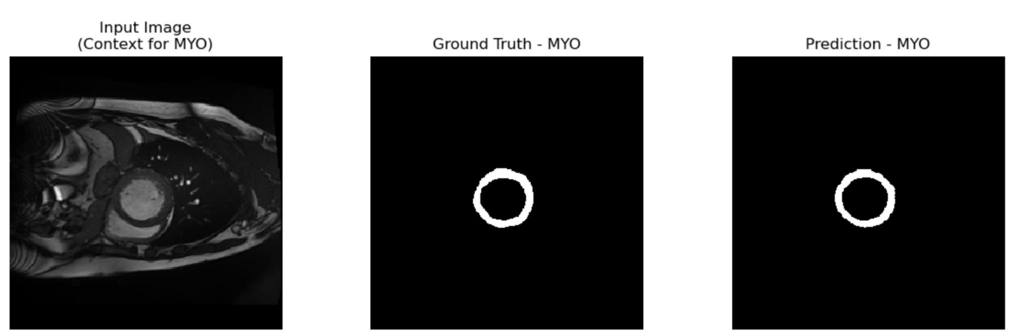

Figure 10: U-Net with Soft Dice Loss Segmentation Example: MYO

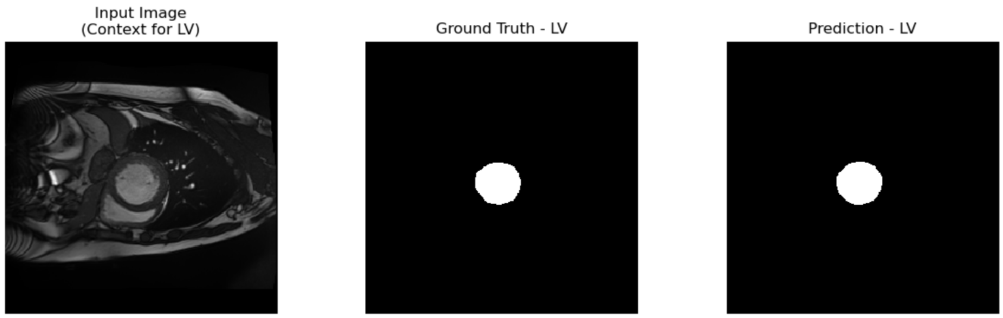

Figure 11: U-Net with Soft Dice Loss Segmentation Example: LV

6.3 Discussion

As shown in Table 4 and Table 5:

- **Segmentation Accuracy (Dice)**: Using Soft Dice Loss resulted in noticeably better Dice coefficients for all cardiac structures compared to BCE Loss when trained on the same non-augmented data. The improvement for MYO was particularly significant (0.8734 to 0.8962).

- **Segmentation Accuracy (Pixel-wise)**: Pixel-wise accuracy also showed slight improvements or remained comparable at very high levels.

- **Conclusion for Task (d)**: Changing the training loss from cross-entropy (BCE) to Soft Dice Loss improved overall segmentation accuracy, especially when evaluated by the Dice coefficient, which is more sensitive to segmentation overlap and robust to class imbalance.

7 Improvements

To further enhance segmentation performance, an Attention U-Net architecture and a Hybrid Loss function were implemented and evaluated.

7.1 Attention U-Net

7.1.1 Structure

- **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. The architecture is depicted in Figure 12.

- **Loss Function:** `BCE` 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.

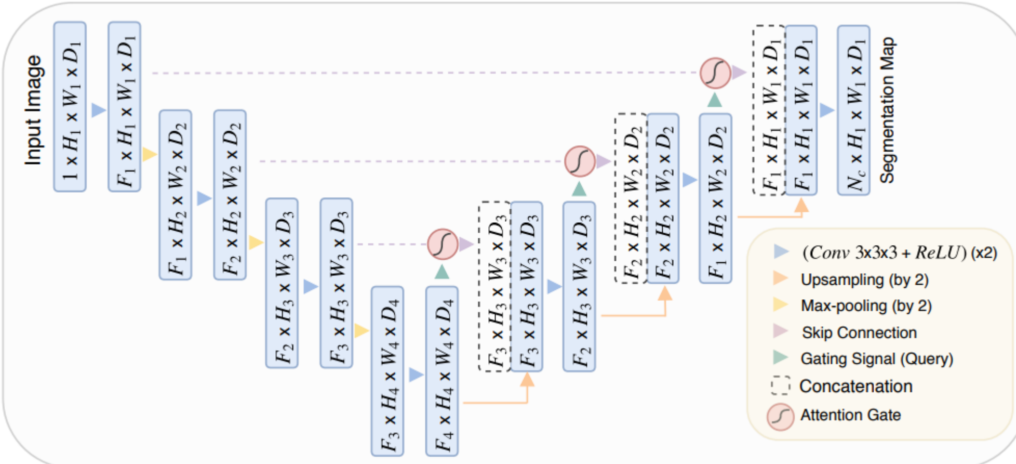

Figure 12: Attention U-Net Architecture.

7.1.2 Performance

The Attention U-Net trained with BCE achieved the Dice coefficients shown in Table 6.

Table 6: Dice Coefficients: Baseline (BCE), Baseline (Soft Dice Loss), and Attention U-Net (BCE) (Task e.1)

Structure	Baseline U-Net (BCE)		U-Net (Soft Dice)		Attention U-Net	
	Mean Dice	Std. Dev.	Mean Dice	Std. Dev.	Mean Dice	Std. Dev.
LV	0.9519	0.0086	0.9566	0.0100	**0.9568**	0.0095
MYO	0.8734	0.0161	0.8962	0.0100	**0.8963**	0.0120
RV	0.8920	0.0310	0.8998	0.0371	**0.9029**	0.0370

Example segmentation results for the Attention U-Net are shown in Figures 13, 14, and 15.

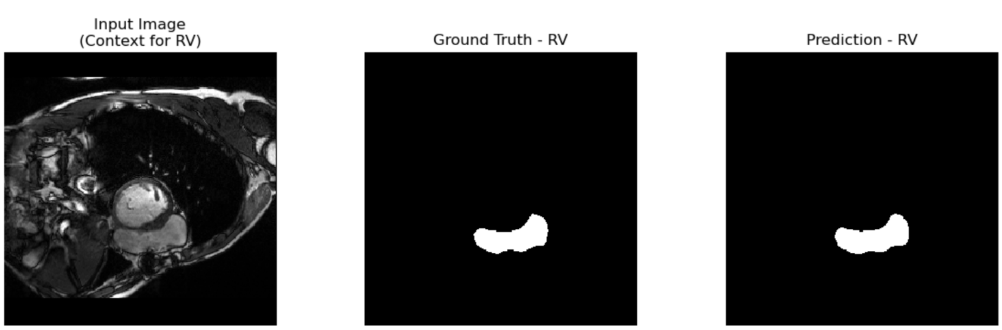

Figure 13: Attention U-Net Segmentation Example: RV

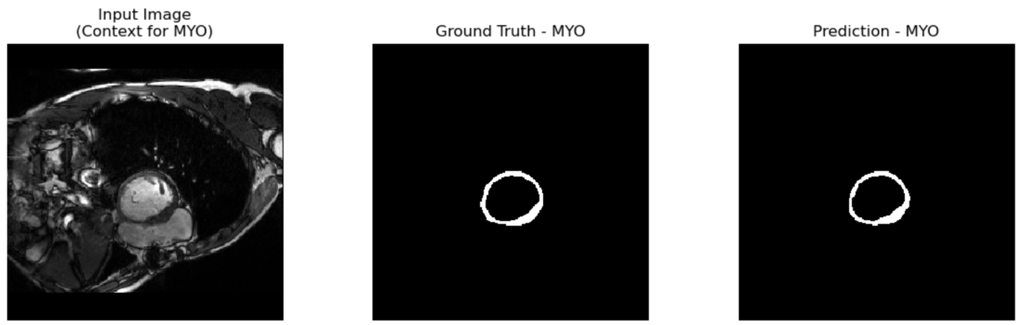

Figure 14: Attention U-Net Segmentation Example: MYO

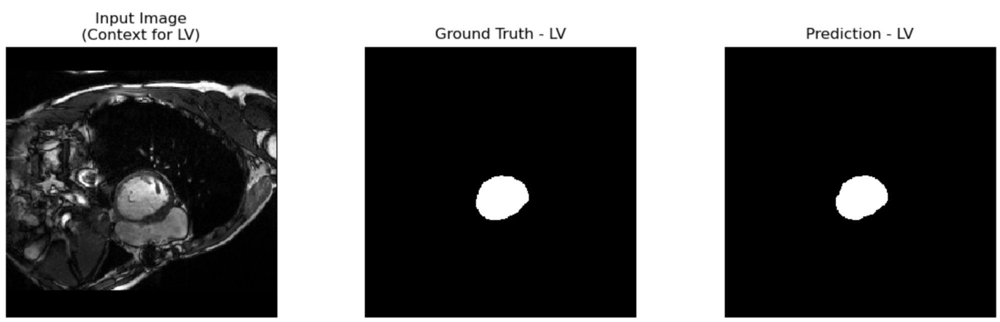

Figure 15: Attention U-Net Segmentation Example: LV

7.1.3 Discussion

- The Attention U-Net showed improved Dice scores compared to the baseline U-Net with BCE loss and the U-Net with Soft Dice Loss.
- This suggests that the attention mechanism effectively helps the model to focus on more complex structures or finer details, leading to better boundary delineation for LV and MYO.
- Accuracy scores (not shown in table, but generally high) are very high across all structures, which is common in segmentation tasks with large background areas. The Dice coefficient remains a more informative metric for evaluating overlap.

7.2 Hybrid Loss

7.2.1 Motivation

To further improve segmentation, especially at boundaries and for complex structures, by combining multiple complementary loss objectives. This aims to leverage the strengths of different loss types for a more holistic optimization.

7.2.2 HybridLoss Structure

The HybridLoss adaptively weights four distinct loss components:

1. **Dice Loss** (Overlap)
2. **Binary Cross-Entropy (BCE) Loss** (Pixel-wise accuracy)
3. **Boundary Loss** (Edge definition)
4. **Hausdorff Distance Loss (Approximation)** (Shape similarity)

The implementation features adaptive weighting of these components using learnable uncertainty parameters. Both the standard U-Net and Attention U-Net were trained with this HybridLoss on non-augmented data.

The U-Net + HybridLoss learned Loss parameters:

1. **Dice Loss**: $\log(\sigma) = -0.7169$
2. **BCE Loss**: $\log(\sigma) = -0.7040$
3. **Boundary Loss**: $\log(\sigma) = -0.6851$
4. **Hausdorff Distance Loss**: $\log(\sigma) = -0.0468$

The Attention U-Net + HybridLoss learned Loss parameters:

1. **Dice Loss**: $\log(\sigma) = -0.7316$
2. **BCE Loss**: $\log(\sigma) = -0.6962$
3. **Boundary Loss**: $\log(\sigma) = -0.6799$
4. **Hausdorff Distance Loss**: $\log(\sigma) = -0.1023$

7.2.3 Performance with HybridLoss

The mean Dice scores for models trained with HybridLoss are presented in Table 7.

Table 7: Mean Dice Scores with HybridLoss (Task e.2, e.3)

Model Configuration	LV Dice (SD)	MYO Dice (SD)	RV Dice (SD)
U-Net + HybridLoss	0.9504 (0.0276)	0.8839 (0.0275)	**0.9061** (0.0573)
Baseline U-Net (BCE)	*0.9519 (0.0086)*	*0.8734 (0.0161)*	*0.8920 (0.0310)*
Attention U-Net + HybridLoss	0.9507 (0.0235)	0.8875 (0.0247)	0.9033 (0.0703)
Attention U-Net	*0.9568 (0.0095)*	*0.8963 (0.0120)*	*0.9029 (0.0370)*

8 Overall Performance Summary

Table 8 provides a consolidated view of the mean Dice Similarity Coefficients (DSC) for all evaluated models across the three cardiac structures.

Table 8: Overall Performance Summary: Mean Dice Coefficients + Std

Model Configuration	LV DSC (M ± SD)	MYO DSC (M ± SD)	RV DSC (M ± SD)
(a) Baseline U-Net	0.9519 ± 0.0086	0.8734 ± 0.0161	0.8920 ± 0.0310
(b) U-Net without Skip Connections	0.9260 ± 0.0111	0.8223 ± 0.0168	0.8588 ± 0.0296
(c) U-Net + Data Augmentation	0.9276 ± 0.0107	0.8469 ± 0.0149	0.8635 ± 0.0384
(d) U-Net (Soft Dice Loss)	0.9566 ± 0.0100	0.8962 ± 0.0100	0.8998 ± 0.0371
(e.1) Attention U-Net	**0.9568 ± 0.0095**	**0.8963 ± 0.0120**	0.9029 ± 0.0370
(e.2) U-Net + HybridLoss	0.9504 ± 0.0276	0.8839 ± 0.0275	**0.9061 ± 0.0573**
(e.3) Attention U-Net + HybridLoss	0.9507 ± 0.0235	0.8875 ± 0.0247	0.9033 ± 0.0703

9 Overall Discussion

The results from Table 8 highlight several key observations:

- The Attention U-Net (e.1) achieved the best performance for LV and MYO segmentation.
- The standard U-Net combined with HybridLoss (e.2) yielded the highest Dice score for RV segmentation.
- No Universal Superiority of HybridLoss: Despite its sophisticated multi-component design with adaptive weighting, HybridLoss did not prove to be a universally superior loss function across all structures and base architectures in these experiments. For LV and MYO, the simpler Soft Dice Loss with Attention U-Net performed better.

- RV Segmentation Strength with HybridLoss: A consistent observation is the relative strength of HybridLoss (or its components) in improving or maintaining high performance for RV segmentation, even when LV/MYO performance might not be optimal compared to other configurations. This suggests that the combined objectives in HybridLoss might be particularly beneficial for the characteristics of the RV.

10 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 LV, RV, and MYO.
2. Skip connections are essential for U-Net's performance. Their removal led to a significant decline in Dice scores.
3. The specific data augmentation strategy employed resulted in a decrease in Dice coefficients, highlighting that augmentation strategies must be carefully chosen and tuned.
4. Training with Soft Dice Loss significantly improved performance over Binary Cross-Entropy loss, particularly for the more challenging MYO structure.
5. The Attention U-Net yielded the best segmentation performance for LV and MYO
6. The U-Net with HybridLoss achieved the best result for RV segmentation (0.9061 DSC).
7. The choice of loss function and architectural enhancements like attention mechanisms can lead to notable improvements, but their effectiveness can be structure-dependent. A simpler model (U-Net) with a well-chosen, targeted loss function (Soft Dice Loss) can still be highly effective.
8. The performance of HybridLoss models suggests that further optimization (e.g., training duration, hyperparameter tuning of the loss components or solver) could potentially lead to even better results.

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