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, and the performance difference between Cross-Entropy loss and Soft Dice loss.

1 Introduction

This project aims to:

1. Implement a 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 Cross-Entropy loss with Soft Dice loss for training the segmentation model.

2 Baseline U-Net

2.1 Structure

The U-Net architecture 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.

- **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 by the solver. Example segmentation results were saved.

2.2 Experiments and Results

The training loss and validation loss of the baseline U-Net are shown in figure 1

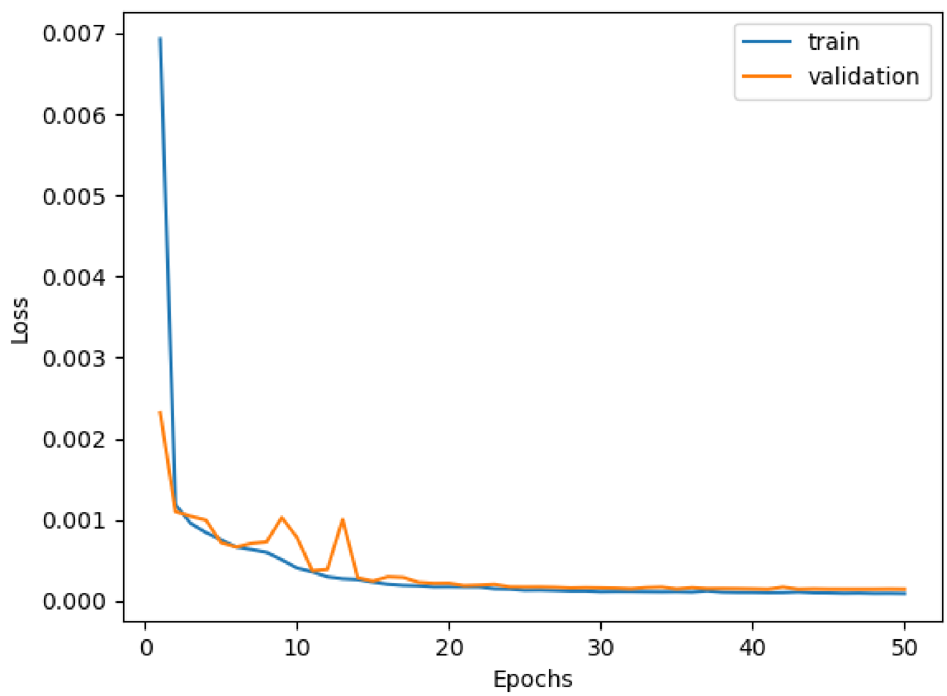

Figure 1: Training and Validation Loss Curves for the Baseline UNet

And here are example segmentation results in figure 2, 3, 4

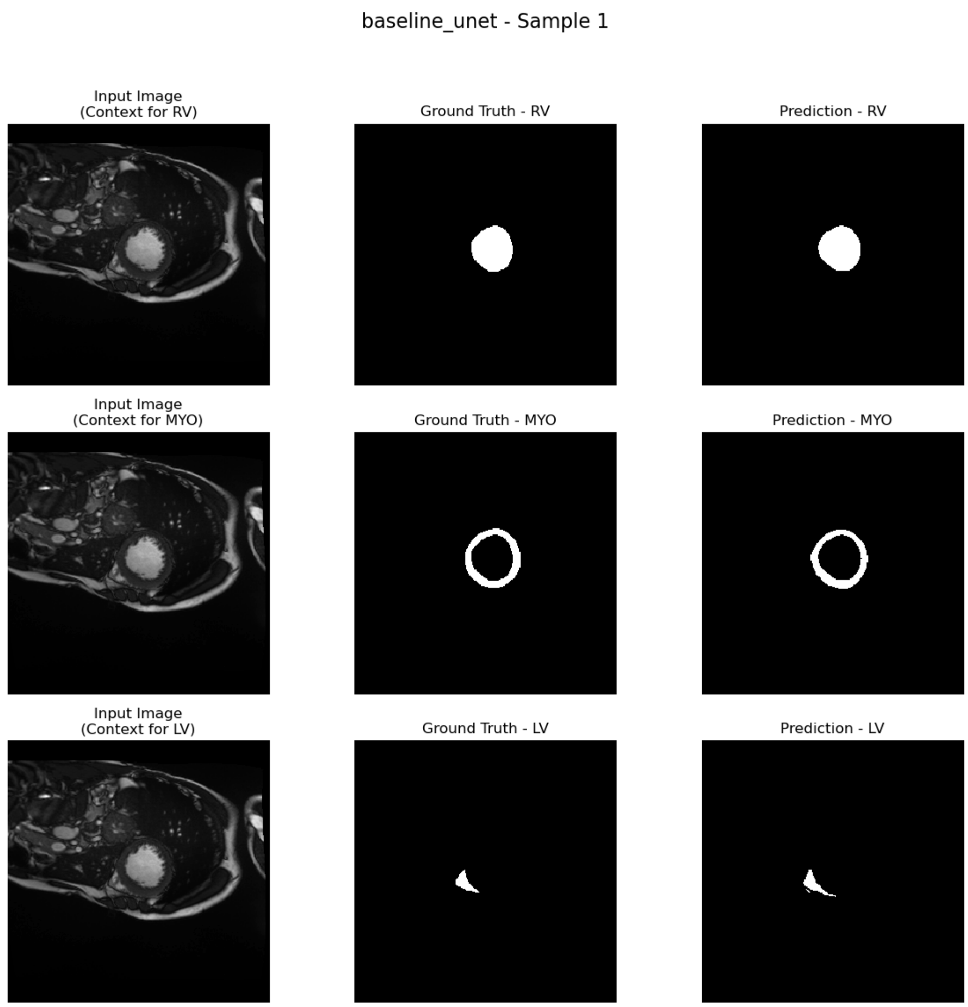

Figure 2: An Example Segmentation for the Baseline UNet (1/3)

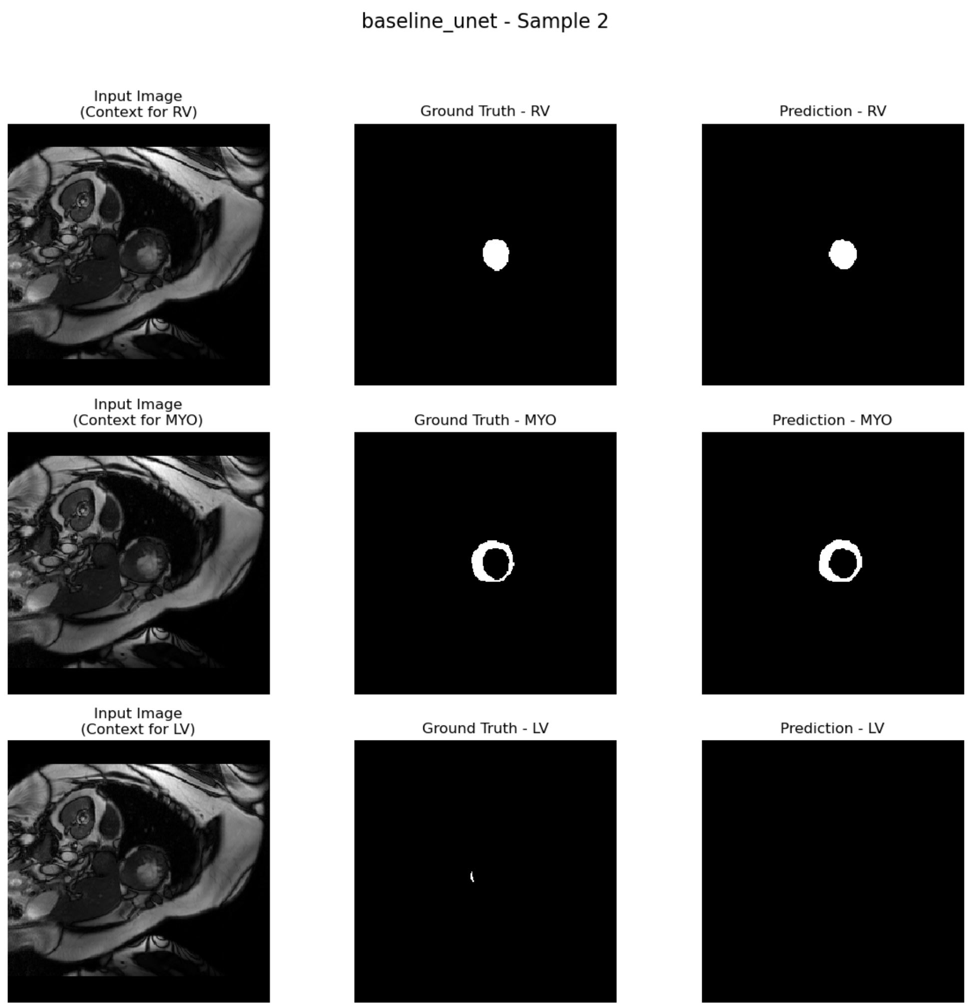

Figure 3: An Example Segmentation for the Baseline UNet (2/3)

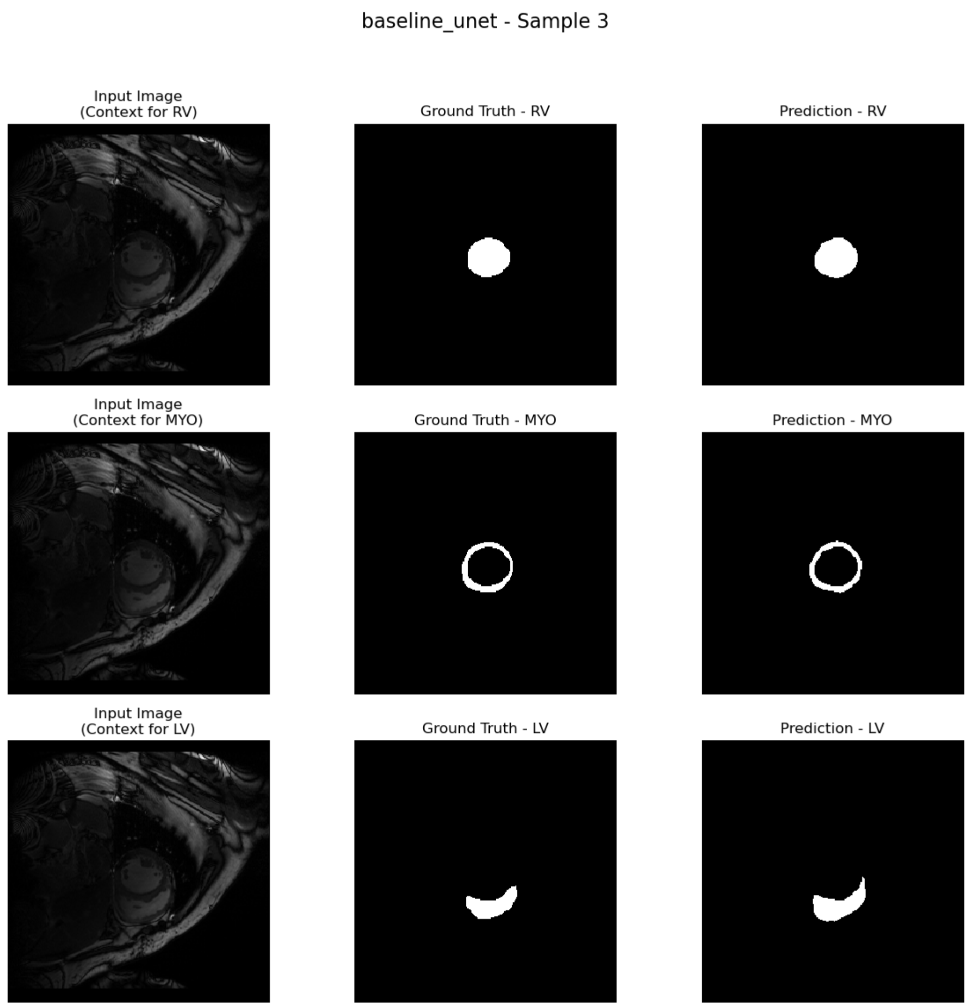

Figure 4: An Example Segmentation for the Baseline UNet (3/3)

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

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

Structure	Mean	Standard Deviation
RV	0.9498	0.0089
MYO	0.8755	0.0120
LV	0.8960	0.0361

2.3 Discussion

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

- **RV Segmentation**: Achieved the highest mean Dice score (0.9498). 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 0.8960. The LV cavity is usually clearly visible.

- **MYO Segmentation**: Had the lowest mean Dice score (0.8755). 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, including the same dataset split, number of epochs, optimizer, learning rate, and `MyBinaryCrossEntropy` loss.

3.2 Performance

The U-Net variant without skip connections, trained under identical conditions, yielded:

Table 2: Dice Coefficients for U-Net without Skip Connections (Task b)

Structure	Mean Dice	Standard Deviation
RV	0.9100	0.0171
MYO	0.8044	0.0185
LV	0.8499	0.0306

3.3 Discussion

Comparing Table 2 with Table 1, removing skip connections from the U-Net architecture resulted in a significant degradation of performance across all three structures:

- RV Dice: $0.9498 \rightarrow 0.9100$
- MYO Dice: $0.8755 \rightarrow 0.8044$
- LV Dice: $0.8960 \rightarrow 0.8499$

This substantial drop highlights the critical role of skip connections. Skip connections allow the decoder to reuse high-resolution feature maps from the encoder, which contain fine-grained spatial information lost during downsampling. This helps in precise localization and reconstruction of segmentation boundaries. Additionally, they facilitate better gradient flow during backpropagation, mitigating vanishing gradient problems in deeper networks and aiding convergence. The MYO, being the most intricate structure, suffered the largest relative drop in performance.

4 U-Net with Data Augmentation

4.1 Structure

- **Data Augmentation Techniques**: Augmentations were applied only to the training dataset using `torchvision.transforms`. 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 baseline U-Net architecture (with skip connections) was retrained using the augmented training data. The validation set remained un-augmented. The loss function was `MyBinaryCrossEntropy`, and the learning rate was 0.01.

4.2 Performance

The baseline U-Net architecture trained with data augmentation on the training set resulted in:

Table 3: Performance of U-Net with Data Augmentation (Task c)

Structure	Accuracy		Dice Coefficient	
	Mean	SD	Mean	SD
RV	0.9988	0.0002	0.9350	0.0109
MYO	0.9971	0.0003	0.8512	0.0160
LV	0.9978	0.0005	0.8673	0.0353

4.3 Discussion

Comparing the Dice coefficients from Table 3 with the baseline U-Net in Table 1:

- RV Dice: $0.9498 \rightarrow 0.9350$
- MYO Dice: $0.8755 \rightarrow 0.8512$
- LV Dice: $0.8960 \rightarrow 0.8673$

We can find out that the 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. While data augmentation is generally expected to improve generalization and robustness, several factors could contribute to this outcome:

-

Despite the drop in Dice scores, the pixel-wise accuracy for the augmented model (Table 3) is very high (e.g., RV accuracy 0.9988). However, accuracy can be misleading in segmentation tasks, especially with class imbalance (background pixels dominating), whereas Dice is more sensitive to boundary overlap.

5 U-Net with Soft Dice Loss

5.1 Structure

- **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 (same architecture as baseline) was trained using the augmented training dataset (as in Task c) but with the `SoftDiceLoss`. The learning rate for the Adam optimizer was set to 0.001 for this task.

5.2 Performance

The U-Net trained with data augmentation and Soft Dice Loss achieved the following pixel-wise accuracies:

Table 4: Accuracy for U-Net with Soft Dice Loss (Task d)

Structure	Mean Accuracy	Standard Deviation
RV	0.9992	0.0001
MYO	0.9979	0.0002
LV	0.9980	0.0007

5.3 Impact of Soft Dice Loss

The U-Net trained with Soft Dice Loss (and data augmentation) was primarily evaluated using pixel-wise accuracy (Table 4). Comparing these accuracies to those from the U-Net trained with BCE loss and data augmentation (Table 3):

- RV Accuracy: $0.9988 \rightarrow 0.9992$ (Soft Dice)
- MYO Accuracy: $0.9971 \rightarrow 0.9979$ (Soft Dice)
- LV Accuracy: $0.9978 \rightarrow 0.9980$ (Soft Dice)

The Soft Dice Loss resulted in slightly higher pixel-wise accuracies for all structures. This is a positive indication. The Soft Dice Loss directly optimizes a surrogate of the Dice coefficient, which is often beneficial for segmentation tasks, particularly when dealing with class imbalance, as it focuses on the overlap between prediction and ground truth rather than just classifying individual pixels.

6 Improvement

7 Conclusion

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

1. The baseline U-Net provides strong segmentation performance for RV, LV, and MYO, with RV 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 resulted in a slight decrease in Dice coefficients compared to the baseline. This suggests that augmentation parameters need careful tuning and their effectiveness can be dataset-dependent.
4. Training with Soft Dice Loss yielded slightly higher pixel-wise accuracies compared to Binary Cross-Entropy loss when both were used with data augmentation. This indicates its potential for optimizing segmentation metrics directly.