Medical Image Segmentation for Cardiac MRI Using U-Net

1. Introduction

The importance of MRI image segmentation is rapidly growing in medical imaging. It facilitates tasks like identifying anatomic structures of organs, as well as detecting tumors and illnesses in early stages, facilitating the analysis of anatomy and pathology.

For a long time, clinicians have relied strongly on manual segmentation. Though important, manual segmentation of cardiac MRI images is generally very time-consuming, labor-intensive, and prone to inter-observer variability; thus, this creates a strong need for automated solutions.

In this project, we tried to work on and improve an MRI image segmentation network by using the infamous ACDC dataset, which has become a standard for benchmarking automated cardiac segmentation methods. It provides multi-class ground truth annotations of the **left ventricle, right ventricle, and myocardium**, along with diverse MRI images that capture variations in both anatomy and pathology.

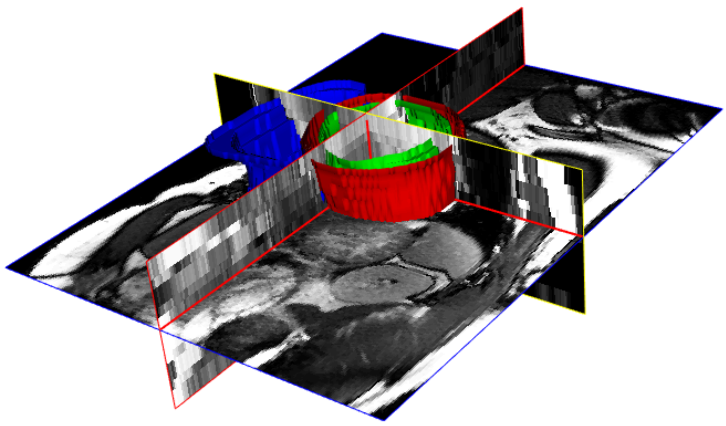

Figure 1: ACDC Dataset
source: https://www.creatis.insa-lyon.fr/Challenge/acdc/

However, automated segmentation in this domain faces several challenges, including:

- **Class imbalance:** Background pixels dominate the masks, while critical regions like the myocardium occupy significantly fewer pixels.
- **Variability in the quality of images**: the difference in MRI acquisition protocols and anatomy introduces variability both in intensity and spatial features.

- **Small region segmentation:** Accurate segmentation of small structures like the myocardium remains challenging due to its size and proximity to other regions.

These challenges have been addressed by deep learning-based approaches, especially CNNs. This work focuses on the application of the U-Net architecture-a well-established model for biomedical image segmentation-to the ACDC dataset. The encoder-decoder structure of **U-Net**, combined with skip connections, enables the extraction of hierarchical features and preserves spatial details, making it well-suited for segmenting complex structures like cardiac anatomy.

In this study, we explore the performance of a U-Net model trained on the ACDC dataset, incorporating different techniques to address the above mentioned challenges:

- **Data augmentation:** Spatial transformations are used, such as rotations, flips, and cropping of images, along with some intensity augmentations, like changes in brightness and contrast.

- **Weighted loss function:** To mitigate the impact of class imbalance, we use a weighted CrossEntropy loss, assigning lower weights to background pixels and higher weights to underrepresented classes like the myocardium.

- **Evaluation metrics:** Pixel wise accuracy is employed as the primary metric to evaluate segmentation quality for both background and foreground pixels.

This report describes the methodology, results, and insights gained from applying the U-Net architecture to the ACDC dataset. Our findings demonstrate the effectiveness of the proposed approach and provide a foundation for further improvements in automated cardiac segmentation.

2. Related Work:

Recent advances in medical image segmentation have been driven by deep learning techniques, especially Convolutional Neural Networks. Among them, U-Net, proposed by Ronneberger et al. (2015), is one of the most successful architectures for pixel-wise segmentation, especially in medical imaging tasks such as organ segmentation in MRI scans. U-Net has seen wide adoption in cardiac MRI segmentation. For example, Zhu et al. (2018) applied U-Net to segment myocardial infarction and obtained large gains in performance. Successful uses of U-Net were also obtained in the challenge ACDC, with very good accuracy in ventricular segmentation, Dice scores being greater than 0.85 or higher.

- **General code work:**

Papers with code: https://paperswithcode.com/dataset/acdc

Kaggle: https://www.kaggle.com/datasets/anhoangvo/acdc-dataset

https://github.com/baumgach/acdc_segmenter?tab=readme-ov-file

https://github.com/creatis-myriad/ASCENT/blob/main/notebooks/acdc.ipynb

3. Methodology

3.1 Neural Network Architecture

The architecture selected for this review is the U-Net, a CNN that is widely used in biomedical image segmentation tasks. Some important features of the architecture are:

- **Encoder**: A set of convolutional blocks with max-pooling operations to extract hierarchical features at different resolutions.
- **Decoder**: Upsampling operations combined with skip connections to recover spatial details lost during encoding.
- **Skip Connections**: Allow spatial information to be transferred directly from encoder layers to decoder layers for better segmentation quality.
- **Output Layer**: A convolutional layer that maps the features to the number of segmentation classes, producing a probabilistic mask for each class.

The U-Net implementation has the following details:

- **Input Channels**: Single-channel grayscale MRI images.
- **Output Channels**: Four classes representing the background, left ventricle, right ventricle, and myocardium.
- **Layer Details**: Convolutional blocks with ReLU activations and Batch Normalization.

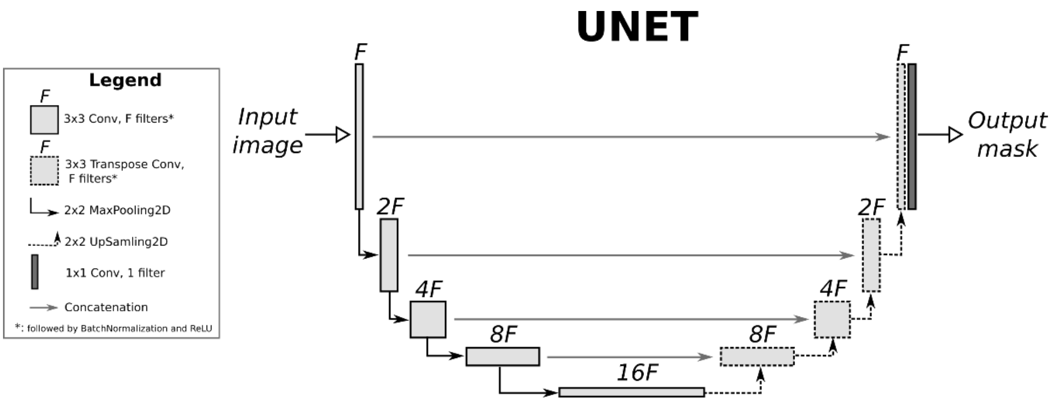

Figure 2: U-Net Model Architecture,
source: https://nchlis.github.io/2019_10_30/page.html

The baseline model was adapted from the following github repository: **cnn_architectures_image_segmentation**, where the owner used it for 1-class image segmentation.

3.2 Data Preprocessing

Preprocessing was done to both the MRI images and their corresponding masks. Preprocessing steps include:

- **Resizing**: The shape of all images and masks is unified to **256×232×256** (height × width × depth) for the unification of input dimensions to be fed into the neural network

- **Data Augmentation**: To help generalization and avoid overfitting, augmentations are used in training:
    - Spatial augmentations: Random cropping, horizontal flipping, and rotation.
    - Intensity augmentations: Random brightness and contrast adjustments.

- **No Normalization**: After research, no normalization is performed in order to maintain the real intensity of mask pixels especially.

3.3 Training Strategy

- **Loss Function**:
  - The main loss function is **CrossEntropyLoss**, which calculates the pixel-wise classification loss over all segmentation classes.
  - First, **class weights** based on pixel distribution in the training set and the Cross Entropy Loss function parameter are derived. Background pixels have lower weights due to domination over other pixels in masks, while weight will increase for underrepresented classes.
- **Optimization**:
  - **Optimizer**: Adam optimizer is used with default parameters
  - **Learning Rate**: The initial learning rate was set to 0.01 then later reduced to 0.001
- **Hyperparameters**:
  - Batch Size: 4 (limited by GPU memory constraints).
  - Epochs: The model is trained for 20 epochs with early stopping to prevent overfitting.
  - Train/Validation Split: 80% of the data is used for training, and 20% is reserved for validation.
- **Class Imbalance Handling**:
  - Weighted loss functions are employed, and accuracy metrics exclude the background class to ensure that performance on key anatomical structures is not overshadowed by dominant background pixels.

3.4 Validation and Testing

Validation is performed at the end of each epoch using the following metrics:

- **Pixel-Wise Accuracy**: Evaluates the proportion of correctly classified pixels.
- **Per-Class Dice score**: computed for each class to identify class-specific accuracy.

For testing, the model is evaluated on a separate test set, with additional qualitative assessments conducted by visualizing segmentation outputs compared to ground truth masks.

3.5 Experimentation:

In this work, the first approach was to train a "vanilla model" with no data augmentations or class weighting, using a standard U-Net architecture and unweighted cross-entropy loss. This was used as a baseline to understand the raw performance of the model on the dataset.

Subsequently, enhancements were made by including weighted cross-entropy loss. Weights were determined as the class distributions in the training dataset. This correction was to account for class imbalance due to the preponderance of background pixels over smaller structures of interest. Next, random rotations, flips, and brightness adjustments were applied so that the model could be made to generalize better.

These enhancements were evaluated quantitatively (Dice scores for each class) and qualitatively through visual inspections of segmentation predictions, demonstrating their impact on performance.

4. Evaluation and Results:

We assume that:

- **Model 1:** un-weighted cross entropy loss, no data transformation is applied on the images and their masks.
- **Model 2:** weighted cross entropy loss, data transformation is applied on the images and their masks.

Evaluation of the segmentation models was performed both quantitatively and qualitatively. All models were tested on a held-out test set of the ACDC MRI dataset to measure their performance in segmenting the cardiac structures. The key metrics that will be used for evaluation include pixel-wise accuracy and Dice scores for each class. Besides, visual comparisons of predicted masks with the ground truth were employed to analyze strengths and weaknesses of the models.

Quantitative Results:

The evaluation highlighted key differences between the two models:

- Model 1 served as a stable baseline with reasonable performance but struggled with class imbalance and small structures.
- The improved model, with weighted loss and augmentations, addressed these limitations by achieving better Dice scores for smaller classes and capturing finer details. However, it introduced variability and occasional over-segmentation.

Summary:

Metric	Model 1	Model 2
Test Accuracy	98%	98%
Dice Score Class 0	99%	99%
Dice Score Class 1	56%	57%
Dice Score Class 2	68%	68%
Dice Score Class 3	83%	84%

Qualitative Results:

Despite showing similar quantitative results, qualitative results make model 2 stand out in terms of mask predictions when compared to the actual masks:

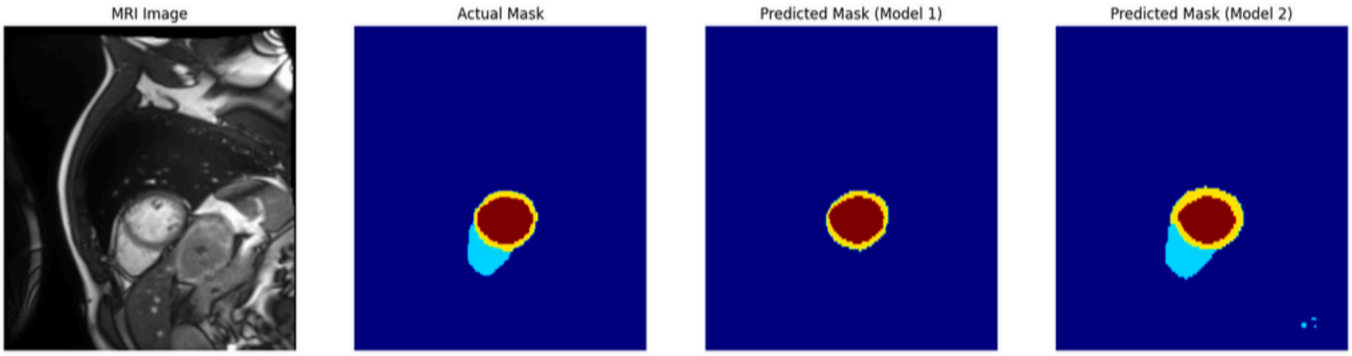

Figure 3: Models Performance

5. Conclusion

In this study, we explored the task of cardiac MRI segmentation using a U-Net architecture on the ACDC dataset. Starting with a baseline "vanilla" model, which used standard cross-entropy loss and no data augmentation, we systematically introduced enhancements such as weighted cross-entropy loss to address class imbalances and data augmentations to improve model generalization. These refinements aimed to tackle the inherent challenges in medical image segmentation, including handling small structures of interest (e.g., myocardium and ventricles) and avoiding over-segmentation or false positives.

Key Findings

1. **Baseline Model Performance:** Baseline performance was pretty decent with overall high accuracy due to the dominance of background pixels, while Dice scores showed that the smaller structures such as myocardium and right ventricle are undersegmented, thus showing the struggles of the model in class imbalance and variability of data.
2. **Improvement with Weighted Loss:** Adding class weights to the cross-entropy loss function did allow for improved capturing of the smaller classes, as reflected in the marginally higher Dice scores observed for these regions. This also resulted in more over-segmentation, in the form of a greater number of false positives, particularly in background regions.
3. **Impact of Data Augmentation**: Data augmentation like random rotation, flipping, and brightness adjustment increased the sensitivity of the model to changes in the input data, hence giving better coverage of the target regions. However, this sometimes introduced noise in the model, resulting in inconsistent performance on edge cases.

Strengths and Limitations

- The improved model demonstrated better sensitivity and generalization than the baseline model, making it more robust for clinical use. However, the improvements in Dice scores were modest, and the model remained prone to over-segmentation in certain cases.
- Qualitative visualizations showed that while the models captured the general shape of cardiac structures, they often struggled with fine boundary details, suggesting room for further refinement in the architecture or training process.

Future Directions

- **Advanced Data Augmentation**: Incorporating more sophisticated augmentations, such as elastic deformations or domain-specific transformations, could provide additional benefits in handling data variability.
- **Loss Function Refinements**: Exploring alternative loss functions, such as a combined Dice and cross-entropy loss, could help balance overlap accuracy with boundary precision.

7. References

- ChatGPT. (2024). Used for code generation, comment generation, text generation, and text correction throughout this project. OpenAI. Available at: https://openai.com/chatgpt.
- An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation
- Semi-Supervised Learning With Fact-Forcing for Medical Image Segmentation
- A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine MR images
- Transformer and group parallel axial attention co-encoder for medical image segmentation
- GPA-TUNet: Transformer and GPA Attention Co-Encoder for Medical Image Segmentation