Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090408
ISSN 2371-8838 Vol. 9 Num. 4

A U-Net Baseline for Left Atrial Tumor Segmentation: Performance Analysis and Limitations

Yuhong Li

Shenzhen Wisdom Nebula AI Technology Co., Ltd., Xili Street, Nanshan District, Shenzhen, China
odina2006@126.com

Keywords: Left Atrial Tumor, Segmentation, U-Net, Cardiac MRI, Baseline Model

Abstract: Accurate and robust automated segmentation of Left Atrial (LA) tumors is essential for clinical diagnosis and treatment planning. Due to the inherent challenges in cardiac imaging, such as low tumor-to-background contrast and subtle boundaries, high-precision segmentation remains difficult. This study proposes and evaluates a standard 2D U-Net architecture for effective LA tumor segmentation. We address class imbalance using the Dice Loss function and enhance generalization through critical data augmentation, including elastic deformation. Evaluated on an independent cardiac MRI dataset, the U-Net model achieves a Dice Similarity Coefficient (DSC) of 0.8145, demonstrating its strong capability as a reliable baseline for this challenging task.

1. Introduction

Left Atrial (LA) tumors are a critical type of cardiac mass requiring precise delineation for accurate volume quantification and surgical planning. Automated segmentation is necessary to reduce the subjectivity and time associated with manual delineation. While the U-Net architecture[1] has become the standard in medical image analysis, segmenting LA tumors presents unique challenges due to their small size, low contrast, and high susceptibility to artifacts. The objective of this study is to propose an optimized 2D U-Net model and comprehensively evaluate its segmentation performance against these specific clinical challenges.

Our main contributions include: (1) Proposing and implementing a 2D U-Net model optimized for LA tumor segmentation, utilizing Dice Loss for class imbalance. (2) Employing essential data augmentation, such as elastic deformation, to enhance model robustness against non-rigid shape variations. (3) Providing a thorough quantitative and qualitative analysis of the model's performance on a challenging independent test set.

2. Related Work

The segmentation of cardiac structures, particularly Left Atrial (LA) tumors, sits at the intersection of deep learning and medical image analysis, leveraging advancements across several domains.

2.1. Deep Learning for Medical Image Segmentation

The development of Deep Convolutional Neural Networks (DCNNs) marked a paradigm shift in medical image analysis. The U-Net architecture, introduced by Ronneberger et al., remains the cornerstone of many segmentation tasks due to its effective encoding-decoding structure coupled with skip connections. This architecture allows for the fusion of high-resolution feature maps (from the contracting path) with semantic information (from the expanding path), which is crucial for achieving high boundary accuracy. Subsequent variants, such as the V-Net[2] and Attention U-Net [3], have been introduced to address 3D volumetric data and enhance feature selection, respectively. For instance, the V-Net extended the U-Net concept to directly process 3D volumes, demonstrating superior consistency in the slice-to-slice prediction dimension.

2.2. Cardiovascular Image Segmentation

In cardiovascular imaging, DCNNs have been widely adopted for segmenting critical structures like the ventricles, atria, and major vessels. Segmentation of the Left Atrium (LA) itself is a well-studied area, often focusing on identifying the LA wall for atrial fibrillation (AF) analysis[4]. Recent work has explored deep supervision and multi-scale feature aggregation to cope with the complex, thin boundaries of the LA structure[5]. However, the segmentation of pathologies such as LA tumors presents an additional challenge: the tumors are small, irregularly shaped objects, often obscured by weak contrast against the blood pool or adjacent myocardium. This difficulty requires models to achieve high precision (low False Positives) while maintaining high recall (low False Negatives).

2.3. Addressing Challenges in Segmentation

A key challenge in segmenting small lesions like LA tumors is the severe class imbalance between the large background area and the small target region. This has necessitated the adoption of specialized loss functions. The Dice Similarity Coefficient (DSC) has become the de facto metric and, consequently, the primary loss function (Dice Loss)[6] in medical segmentation to directly optimize the model for overlap accuracy, offering robustness against class imbalance. Furthermore, techniques such as aggressive data augmentation, including geometric transformations and non-linear transformations like elastic deformation, are essential for improving the model's generalization ability and robustness against imaging artifacts and morphological variations inherent in clinical data.

3. Methodology

The segmentation of LA tumors is implemented using a 2D U-Net architecture [1], chosen for its proven efficacy in extracting both local and global features necessary for accurate delineation in medical images. The model processes each MRI slice independently, treating the task as a pixel-wise binary classification problem.

3.1. U-Net Architecture

The U-Net structure comprises a symmetric Contracting Path (Encoder) and an Expanding Path (Decoder), connected by Skip Connections.

The Contracting Path (Encoder) is responsible for capturing contextual information. It follows a typical convolutional network structure, consisting of five resolution levels. At each level, the

feature maps undergo two sequential 3×3 convolutional layers, each followed by a Rectified Linear Unit (ReLU) activation function. This is succeeded by a 2×2 max-pooling operation with stride 2 for downsampling, which effectively halves the spatial dimensions and doubles the number of feature channels. This process enables the network to extract increasingly abstract and high-level features of the tumor and surrounding tissue.

The Expanding Path (Decoder) is tasked with precise localization and upsampling the compressed feature maps back to the original input size. It also consists of five corresponding levels. At each stage, the feature maps are first up-sampled using a 2×2 transpose convolution (or up-convolution), which doubles the spatial dimensions and halves the number of feature channels. The key architectural feature of the U-Net is the Skip Connection, where the corresponding high-resolution feature maps extracted from the contracting path are concatenated with the upsampled features. This concatenation is vital, as it allows the decoder to efficiently integrate fine-grained spatial information (lost during pooling) with the abstract semantic information, which is critical for accurately defining the subtle and often ambiguous boundaries of the LA tumor. Following concatenation, two 3×3 convolutional layers and ReLU activations are used to refine the reconstructed features.

3.2. Output Generation

The final layer of the expanding path utilizes a single 1×1 convolution to map the final feature set to the desired number of output channels, which is one (representing the probability of belonging to the tumor class). A Sigmoid activation function is applied to the output to produce a continuous probability map, $P(x) \in [0,1]$, over the entire image. The final binary segmentation mask is then obtained by applying a threshold of 0.5 to the probability map $P(x)$.

4. Experiments

4.1. Dataset Description

The study utilized a Left Atrium (LA) Segmentation Dataset[7]. The dataset contains 30 volumetric MRI scans, partitioned into 20 cases for training and 10 independent cases for testing. All volumes feature a 320×320 lateral resolution. However, the axial depth (slice count) varies across cases,with counts typically ranging from approximately 90 to 220 slices. For instance, the volume la_018.nii.gz has 220 slices.

Each 3D MRI volume was processed slice by slice. Z-score Standardization and Min-Max Normalization were applied to ensure consistent pixel intensity distribution and scale the data to [0, 1]:

$$\hat{I}_{std} = \frac{I-\mu}{\sigma} \text{ and } \hat{I}_{norm} = \frac{\hat{I}_{std}-I_{min}}{I_{max}-I_{min}}.$$

Where I is the original pixel intensity μ and σ are the mean and standard deviation of the pixels, respectively; and I_{max} and I_{min} are the minimum and maximum pixel values after standardization.

4.2. Data Augmentation and Implementation

To enhance generalization, extensive data augmentation was used, including affine transformations, random rotation, scaling, and critically, elastic deformation. Elastic deformation is key to simulating non-rigid tissue changes and provides an effective regularization against overfitting.

The primary loss function was the Dice Similarity Coefficient Loss ($1-\mathrm{DSC}$) to mitigate severe

class imbalance. The model was trained using the Adam optimizer for 200 epochs with a batch size of 12.

To maintain stable training and prevent premature convergence, a dynamic learning rate scheduler (ReduceLROnPlateau) was employed to maintain stable convergence. Specifically, the learning rate was reduced when the training loss showed no significant improvement over a set number of epochs, effectively preventing the model from oscillating and maintaining convergence towards the minimum.

The TensorBoard visualization tool was used throughout the process to monitor and track the evolution of the training and testing losses in real-time.

5. Results and Discussion

5.1. Quantitative Evaluation

The proposed U-Net model was evaluated on the 10 unseen MRI volumes of the test set. The primary metric used was the Dice Similarity Coefficient (DSC), Jaccard Index (IoU), and metrics derived from the confusion matrix (Sensitivity, Specificity, and Accuracy).

The performance metrics averaged over the entire test set are summarized in Table 1:

Table 1: Summary of quantitative segmentation results on the left atrium test set.

Metrics	Mean score	Standard deviation
Dice similarity soefficient (DSC)	0.8145	0.3288
Jaccard index (IoU)	0.7773	0.3297
Sensitivity	0.9299	0.2064
Specificity	0.9994	0.0008
Accuracy	0.9993	0.0008

The achieved DSC of 0.8145 demonstrates a strong capability of the 2D U-Net architecture to accurately identify and segment the Left Atrium cavity in diverse MRI scans. This high overlap score is complemented by a robust Jaccard Index (IoU) of 0.7773, further validating the segmentation quality. The high standard deviation in DSC ($\pm$ 0.3288) indicates significant performance variance on a few challenging cases.

Furthermore, the training dynamics are illustrated in Figure 1. The stable convergence, despite initial fluctuation in the test loss curve, confirms the effectiveness of the dynamic learning rate scheduler in preventing overfitting.

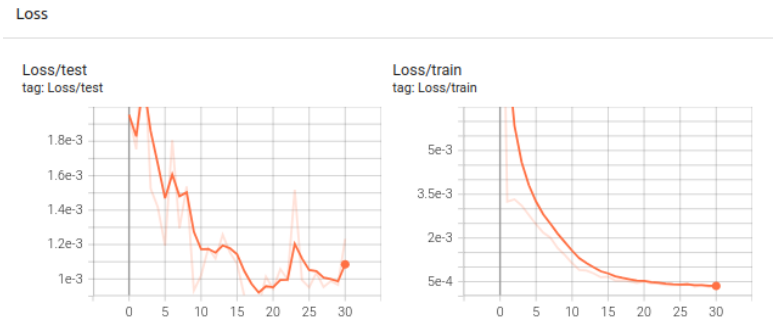

Figure 1: Convergence plot of training and testing losses over 30 epochs.

This figure illustrates the training dynamics over the first 30 epochs, showcasing the evolution of the Dice Loss on both the training set (right panel) and the independent test set (left panel).

Training Loss (Right): The training loss exhibits a rapid, smooth decline, indicating that the U-

Net model is effectively learning to segment the Left Atrium during the early phases of training.

Test Loss (Left): The test loss demonstrates a more volatile but overall decreasing trend, reaching its minimum value around epoch 20–25. This loss is the primary criterion for model selection, and its fluctuation justifies the use of the dynamic learning rate scheduler to prevent overfitting and ensure stability.

This visualization confirms the model's convergence and the effectiveness of the training protocol.

5.2. Qualitative Analysis (Visual Results)

Visual inspection of the predicted masks provides crucial insight into the model's performance, particularly concerning boundary precision. Figure 2 shows visual segmentation results from the test set.

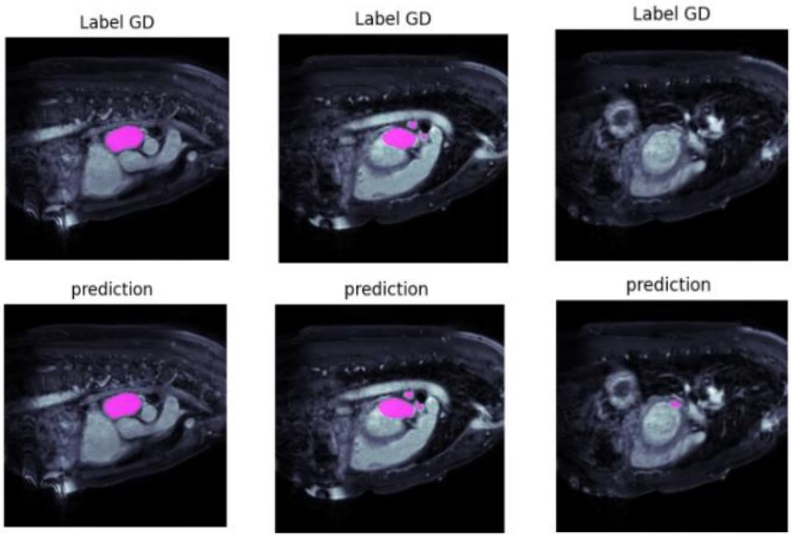

Figure 2: Qualitative segmentation results on representative test cases: a comparison of Ground Truth (top row) and predicted masks (bottom row), both overlaid on the original MRI slices, across three scenarios.

As shown in Figure 2, the qualitative segmentation results across three representative cases provide crucial insights into the model's performance range and failure modes. The first column illustrates the Best Case Scenario, where the predicted mask achieves Near-perfect overlap and highly precise boundary delineation against the Ground Truth (GT), typically observed in slices with high signal quality and distinct contrast. The second column demonstrates the volume effects; despite minor visible discrepancies, the model successfully maintains the overall morphology and localization of the target tumor. Conversely, the third column highlights an Error Case, revealing the primary limitation of the current system, often manifested as significant segmentation errors (e.g., False Positives or False Negatives) in slices corrupted by severe noise or highly ambiguous tumor-tissue boundaries. These visual observations complement the quantitative metrics in Table 1, confirming the model’s strong performance under clear conditions while indicating areas for future focus, particularly in noise robust boundary extraction.

The effectiveness of the extensive data augmentation strategy[8], which includes affine transformations, rotation, scaling, and the crucial Elastic Deformation, is visually confirmed in Figure 3.

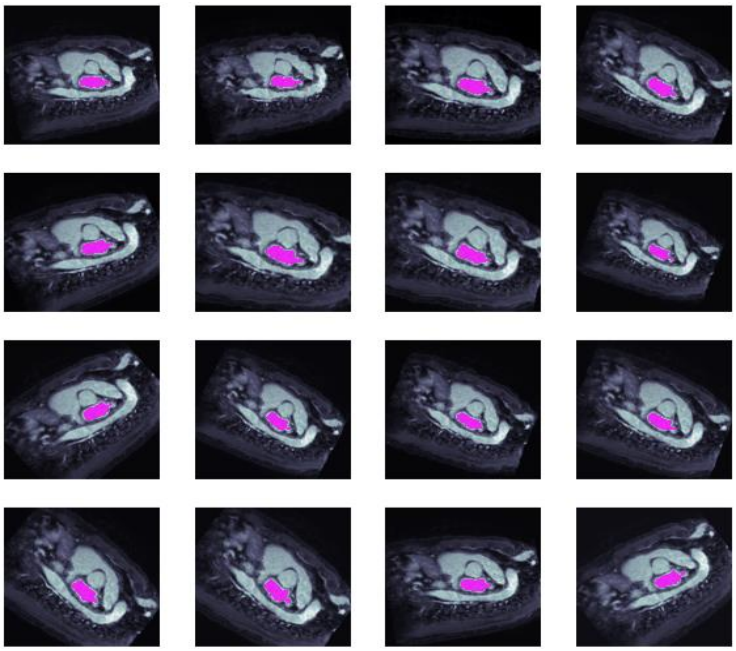

Figure 3: Illustration of on-the-fly data augmentation effects.

This figure visualizes the extensive geometric and non-linear transformations applied to a single, specific 2D magnetic resonance image (MRI) slice during the training phase. The original slice is repeatedly augmented 16 times (a 4×4 grid) using combinations of affine transformations (rotation, scaling) and elastic deformation. The visualization method involves overlaying the segmentation mask (pink/magenta) onto the MRI image, where pixels with a mask value of zero (background) are masked out to ensure transparency and clear display of the segmentation target (Left Atrium). This augmentation strategy is critical for enhancing the model's robustness against varying anatomical orientations and non-rigid shape changes, thereby improving generalization on the limited training set.

5.3. Discussion and Future Work

The competitive performance shown in Table 1 is primarily due to the effective Dice Loss function and the boundary-preserving skip connections of the U-Net architecture. Crucially, the implemented elastic deformation was vital for ensuring model robustness against non-rigid shape changes.

The primary limitation of this study is the use of 2D slice processing, which inherently ignores spatial coherence across adjacent slices. This may lead to inconsistencies in the predicted mask along the Z-axis. Future work will address this by exploring 3D U-Net (or V-Net) architectures to leverage the full 3D spatial relationship of the heart structure. We will also investigate advanced loss functions that incorporate boundary-aware terms to further refine segmentation borders.

6. Conclusions

This paper presented and evaluated a 2D U-Net model for Left Atrial tumor segmentation, achieving a DSC of 0.8145 on an independent test set. By optimizing the model with Dice Loss and elastic deformation, the model demonstrated good robustness and accuracy. The comprehensive quantitative and qualitative analysis establishes a solid baseline for U-Net application in this challenging cardiac tumor segmentation task.

Acknowledgements

The author thanks Mr. Enpei Zhao for providing the experimental dataset and a portion of the implementation code used in this study.

References

[1] Ronneberger, O., Fischer, P., and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer international publishing, 234-241.

[2] Milletari F., Navab N., and Ahmadi S.A. (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. The Fourth International Conference on 3D Vision (3DV). IEEE, 565-571.

[3] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Jammerla, N.Y., Kainz, B., Glocker, B., and Rueckert, D. (2018) Attention u-net: Learning Where to Look for The Pancreas. arXiv preprint arXiv, 1804.03999.

[4] Jin, C., Feng, J., Wang, L., Yu, H., Liu, J., Lu, J., and Zhou, J. (2018) Left Atrial Appendage Segmentation and Quantitative Assisted Diagnosis of Atrial Fibrillation Based on Fusion of Temporal-Spatial Information. Computers in biology and medicine, Vol. 96, 52-68.

[5] Li, J., Yu, Z. L., Gu, Z., Liu, H., and Li, Y. (2019) Dilated-Inception Net: Multi-Scale Feature Aggregation for Cardiac Right Ventricle Segmentation. IEEE Transactions on Biomedical Engineering, 66(12), 3499-3508.

[6] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M. (2017) Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. International Workshop on Deep Learning in Medical Image Analysis, 240-248.

[7] Li, He. (2025) Left Atrium 2018, IEEE Dataport, doi:10.21227/5y8c-wf96.

[8] Jung, A. (2020) imgaug: Library for Image Augmentation in Machine Learning. Version 0.4.0. Available at: https://github.com/aleju/imgaug.