Usformer: A Light Neural Network for Left Atrium Segmentation of 3D LGE MRI

Hui Lin[a*], Santiago Lopez Tapia[a], Florian Schiffers[a], Yunan Wu[a], Huili Yang[a], Nikolay Iakovlev[a]
Bradley D. Allen[a], Ryan Avery[a], Daniel C. Lee[a], Daniel Kim[a], Aggelos K. Katsaggelos[a*]
[a] Northwestern University
* huilin2023@u.northwestern.edu, a-katsaggelos@northwestern.edu

***Abstract*—Left atrial fibrosis is an important mediator of atrial fibrillation and atrial myopathy. Late gadolinium-enhancement (LGE) MRI is a proven non-invasive test for the evaluation of left atrial (LA) fibrosis. However, manual segmentation is labor-intensive. Automatic segmentation is challenging due to varying intensities of data acquired by different vendors, low contrast between the LA and surrounding tissues, and complex LA shapes. Current approaches based on 3D networks are computationally expensive and time-consuming due to the large size of 3D LGE MRIs and networks. To address this, most approaches use two-stage methods to first locate the LA center using a down-scaled version of the MRIs and then crop the full-resolution MRIs around the LA center for final segmentation. We propose a light transformer-based model to accurately segment LA volume in one stage, avoiding errors introduced by sub-optimal two-stage training. Transposed attention in transformer blocks can capture long-range dependencies among pixels in large 3D volumes without significant computation requirements. Our proposed model achieved a promising dice similarity coefficient of 92.6% in the 2018 Atrial Segmentation Challenge, with only 611k parameters, which is about 1% of the method ranked 3rd in the challenge but with similar performance.**

***Index Terms*—Left atrium segmentation, Late gadolinium enhanced magnetic resonance image, Transformer**

I. Introduction

The quantification of left atrial (LA) fibrosis from late gadolinium enhanced (LGE) MRI currently requires labor-intensive manual segmentation, which can result in significant variance. Therefore, automatic LA segmentation with high accuracy and robustness is of high interest. However, this task is challenging due to complex LA shapes, varying shapes and sizes among patients, low contrast, and background noise [1].

Convolutional neural networks (CNNs) have presented a promising performance in multiple applications [2], [3]. Notably, 15 CNN-based methods outperformed the other two traditional methods by about 7% in terms of dice score during the 2018 Atrial Segmentation Challenge [4]. Among these, U-Net [5] variants demonstrated the best performance due to the skip connections in the U-Net architecture. These connections not only recover spatial information for fine-grained segmentation but also alleviate the potential vanishing gradient problem during training.

CNN-based methods for LA segmentation can be divided into 2D or 3D approaches. 2D approaches independently segment each slice of a 3D scan along the Z-axis and stack them for the final 3D prediction [6]–[8]. For example, Wong et al. [6] developed a 2D U-Net variant called GCW-UNet, which achieved a dice score of 93.57% on the testing set of the 2018 Atrial Segmentation Challenge. For each individual slice's segmentation, the input is its three Gaussian blurred images with different blur degrees. GCW-UNet employs a channel weight module and Gaussian blurring to capture both details and overall outlines of the LA. Bian et al. [7] combined the PSPNet [9] and the ResNet [10] with dilated convolution. The spatial pyramid pooling combines multi-scale features for clear boundary outlining.

In contrast, 3D methods directly segment the entire 3D LGE MRI. However, current 3D methods are inefficient in terms of time and memory due to the large size of 3D scans, even though they preserve the correlation among the surrounding slices. Vesal et al. [11] proposed a single 3D CNN using the U-Net architecture, which was ranked 4th in the 2018 Atrial Segmentation Challenge. They used dilated convolution to enlarge the receptive field and residual connections to incorporate the features extracted at different layers. However, the model is the largest in the challenge with $104M$ parameters, which is 50 times larger than the smallest one.

Many researchers have focused on two-stage approaches to reduce the burden on memory and computation [12]–[14]. In the first stage, the LA center is located from a down-sampled version of the LGE MRIs, and a fixed area around the detected center is cropped as the region of interest (ROI). The second stage is to segment the LA from the ROI. Xia et al. [13] trained two V-Net-based networks with the same architecture separately for coarse and fine LA segmentation. The coarse LA segmentation determines the coordinates of the LA center. In contrast, Jamart et al. [12] applied a 2D version of V-net [15] for the regression of LA center coordinates in the first stage. However, two-stage methods have an issue: it is difficult to train two networks jointly, and the errors from the first network might badly affect the accuracy of the second one.

In this work, we propose a novel and light 3D transformer model to accurately segment LA volume. Our approach is inspired by the U-Net architecture, where 3D convolutions are utilized to capture the inter-slice correlations at higher scales. At lower scales, we employ transposed attention to extract long-range interactions in 3D volumes while significantly reducing the number of computations required for normal attention. Additionally, carefully selected data augmentation techniques further improve the segmentation performance. The

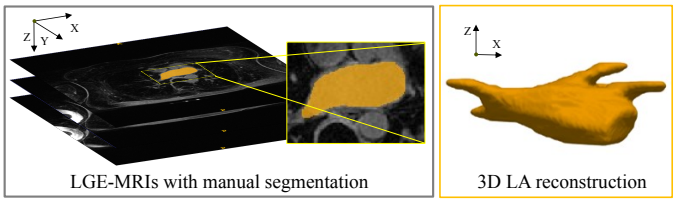

Fig. 1. A patient's 3D LGE MRI with manual segmentation (denoted in orange) and the manual segmentation's 3D reconstruction. An LGE MRI was segmented manually slice-by-slice from the axial view (XY-plane), and all segmentation results were stacked in the Z direction for the 3D LA geometry.

remainder of the paper is organized as follows: The details of the proposed network are given in Section II, including information on input, architecture, and loss function. Experimental results and analysis are shown in Section III. Section IV presents the conclusions.

II. Methods

A. The input and challenges

Fig. 1 presents an example of a short-axis LGE scan and its corresponding manual LA segmentation. When analyzing the scan, automatic LA segmentation of LGE scans faces the following challenges: (1) Class imbalance, as the LA is a small part of the whole volume. (2) Blurry boundaries, which make it difficult to distinguish the LA from surrounding tissues. (3) Background noise, which can impact image quality and make it harder to identify the LA. According to the 2018 Atrial Segmentation Challenge [4], the Signal-to-Noise Ratio (SNR) of all data was evaluated for image quality. Results showed that less than 15% of MRI data were of high quality. (4) Complex anatomy, which includes thin and long parts such as the left atrial appendage (LAA), mitral valve (MV), and pulmonary vein (PV). These structures are often sources of segmentation errors. (5) Varied shapes and sizes among patients, making it challenging to develop a general model for LA segmentation.

B. The network architecture

To overcome the challenges mentioned above, a transformer-based segmentation model with a U-Net architecture named Usformer is proposed. As depicted in Fig. 2, Usformer comprises encoder and decoder networks on the left and right sides, respectively. The encoder network extracts high-level features from the input volume, while the decoder network reconstructs these features to produce the segmentation maps at the original size. To improve spatial accuracy, skip connections are used to connect high-level and low-level features. However, the U-Net architecture has a limited receptive field and fails to capture contextual information in areas with background noise. To overcome this limitation, transformer blocks are introduced to capture global context through their self-attention mechanism. Thus, Usformer consists of the first two convolutional stages and the last three transformer stages. Each transformer stage consists of one transformer block followed by a max pooling or upsampling layer.

The segmentation output is a probability map representing the probability of each pixel belonging to LA. Pixels with a probability higher than a certain threshold are classified as LA. Our experiments demonstrate that the proposed method is robust, as we observe only a 0.01% difference in the 3D dice score when varying the threshold from 0.1 to 0.9.

C. Transposed attention module

The transposed attention module [16] in the transformer block is shown in the yellow box of Fig. 2. From a layer-normalized input of size $\hat{H} \times \hat{W} \times \hat{Z} \times \hat{C}$, Query ($Q$), Key ($K$), and Value ($V$) are generated through bias-free convolutional layers. $\hat{H}, \hat{W}, \hat{Z}$ represent the size in the X, Y, and Z directions, respectively, and n is the number of the input's voxels, which is equal to $\hat{H} \times \hat{W} \times \hat{Z}$. Then the matrix K is transposed to ensure the size of the attention map generated by K and Q is $\hat{C} \times \hat{C}$ rather than $n \times n$. Therefore, the transposed attention is calculated as:

$$A(Q, K, V) = V\text{softmax}(KQ) \tag{1}$$

where $Q, V \in \mathbb{R}^{n \times \hat{C}}$ and $K \in \mathbb{R}^{\hat{C} \times n}$.

The computation complexity of KQ is $O(\hat{C}^2 n)$. In a conventional self-attention module [17], $Q, K, V \in \mathbb{R}^{n \times \hat{C}}$ and attention is computed using $A(Q, K, V) = \text{softmax}(QK^T)V$. The computation complexity of QK^T is $O(n^2\hat{C})$. Since $\hat{C} \ll n$ and $O(\hat{C}^2 n) \ll O(n^2\hat{C})$, the computation complexity of transposed attention is considerably lower than the conventional one.

D. Loss function

The total segmentation loss $\mathcal{L}_{seg}$ is calculated as the weighted sum of binary cross entropy loss (BCE) and dice loss, using Equation 2. The BCE loss $\mathcal{L}_{seg}^{BCE}$ treats all pixels' loss equally, but the LA pixels' contributions to the training process are badly harmed due to the significant class imbalance between the LA and the background. On the other hand, the dice loss $\mathcal{L}_{seg}^{dice}$ is area-based and remains constant regardless of the background's size, addressing the class imbalance issue in the LA segmentation dataset. However, if only the dice loss is used, the training process becomes unstable when the foreground is small since minor changes can cause significant changes in the dice loss. Therefore, $\mathcal{L}_{seg}$ combines both losses to achieve a stable and effective training process. $\mathcal{L}_{seg}$ is given by:

$$\begin{aligned}\mathcal{L}_{seg} &= 0.5\mathcal{L}_{seg}^{BCE} + \mathcal{L}_{seg}^{dice} \\ &= -0.5(Y\log\hat{Y} + (1-Y)\log(1-\hat{Y})) + (1 - \frac{2Y \cap \hat{Y}}{Y \cup \hat{Y}})\end{aligned} \tag{2}$$

where Y is the manual segmentation, $Y \in \{0, 1\}$ and $\hat{Y}$ is the model output, $\hat{Y} \in [0, 1]$, where 0 represents the background, and 1 represents the LA.

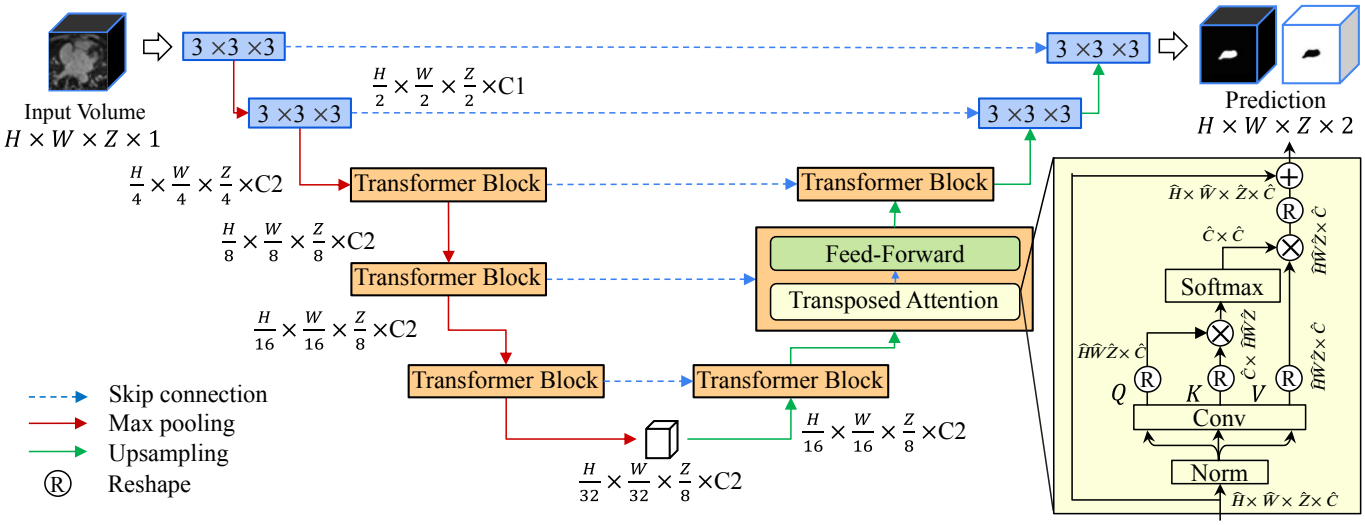

Fig. 2. The proposed Usformer for LA segmentation. Usformer incorporates a U-Net architecture with efficient transposed transformer blocks that significantly decrease computation complexity.

III. Experiments and Results

A. Dataset and Evaluation Metrics

The dataset used in this paper is from the 2018 Atrial Segmentation Challenge [4], which is commonly used by researchers studying LA segmentation. The original size of each slice in this dataset is 576×576 or 640×640 pixels, and each patient has 88 slices. Each slice is from the axial view. The spatial resolution of each scan is 0.625×0.625×0.625 mm^3. The LA cavity was manually segmented in agreement by three trained observers, which included all pixels contained within the LA endocardial surface, left atrial appendage (LAA), mitral valve (MV), and pulmonary vein (PV). The original training set in the challenge is randomly split into 80 3D LGE MRIs for training and 20 for validation. The testing set is the same used by all researchers participating in the challenge. Since the LA only occupies a small part of the whole slice, as shown in Fig. 1, each original LGE MRI scan was cropped in the volume center to the same size of $288 \times 288 \times 88$ pixels to reduce the computational complexity. The largest dimensions of the LA in the x, y, and z directions in the original training set are 209, 128, and 73, respectively, so $288 \times 288 \times 88$ is sufficient to cover the whole LA. Data augmentation (DA) was employed to prevent overfitting and improve generalizability. Three transformation methods-scaling, rotation, and translation- were applied to each XY-plane with a probability of 50% for data augmentation. The scaling factor, rotation angle, and translation were randomly selected within $(0.5, 1.5)$, $(-25°, 25°)$, and $(-10, 10)$ pixels for the X and Y axes, respectively. Our experiments presented that data augmentation increased the 3D dice score by 2.1%.

The dice score is the most commonly used metric for LA segmentation [18]. The dice score for each individual LGE MRI segmentation in 3D is calculated using Equation 3. The average 3D dice score across all cases in the testing set is then used to compare the performance of different models.

$$dice = \frac{2TP}{2TP + FN + FP} \tag{3}$$

where TP, FN, TN, and FP represent respectively the number of true positives, false negatives, true negatives, and false positives over each patient's entire volume.

B. Performance Evaluation

(1) Quantitative results

State-of-art methods for LA segmentation in the 2018 challenge are not publicly accessible, making it difficult to replicate their results. Their publicly disclosed results are compared in Table I. We implemented a 3D two-stage method using the nnU-Net framework [19], which achieved a 93.1% 3D dice score, similar to the best result in the challenge. In addition to nnU-Net, we also implemented popular architectures such as U-Net [5] and UNeXt [20] for comparison purposes.

Our proposed Usformer outperforms nnU-Net in the number of parameters with a competitive 3D dice score. The total number of parameters of Usformer is 611K, much smaller than nnU-Net's 16.2M. Moreover, Usformer also outperforms U-Net and UNeXt, achieving about 8% higher and 2% higher 3D dice scores, respectively, in LA segmentation.

The transposed attention module in Usformer enhances efficiency by reducing computational complexity and enabling the learning of global information. Table I demonstrates that our model is significantly lighter than other models without compromising the 3D dice score. Additionally, it has the lowest standard deviation, making it more robust.

(2) 3D visualization

Four cases are selected from the testing set with the best and worst two performances in terms of the 3D dice score by the proposed method for 3D and 2D visualization, as shown in Figs. 3 and 5. Fig. 3 presents that our proposed method has promising performance in LA segmentation even though a large variation exists in LA shapes among patients. The

TABLE I
EVALUATION OF THE PROPOSED USFORMER AND STATE-OF-ART METHODS ON THE TESTING SET IN THE 2018 ATRIAL SEGMENTATION CHALLENGE [4]. THE FIRST FOUR ROWS ARE OUR IMPLEMENTATIONS, AND THE OTHERS ARE THE AUTHORS' DISCLOSED RESULTS.

Method	dice (%)	Number of parameters (M)
Usformer	92.6 ± 2.0	**0.611**
nnUNet [19]	93.1 ± 2.2	16.2
UNeXt [20]	90.5 ± 3.4	26.5
U-Net [5]	84.1 ± 3.9	1.9
Xia et.al [13]	$\mathbf{93.2 \pm 2.2}$	21
Huang et.al [4]	93.1 ± 2.2	2
Bian et.al [7]	92.6 ± 2.2	45
Vesal et.al [11]	92.5 ± 2.7	3
Yang et.al [21]	92.5 ± 2.3	104
Li et.al [22]	91.9 ± 2.3	5.14

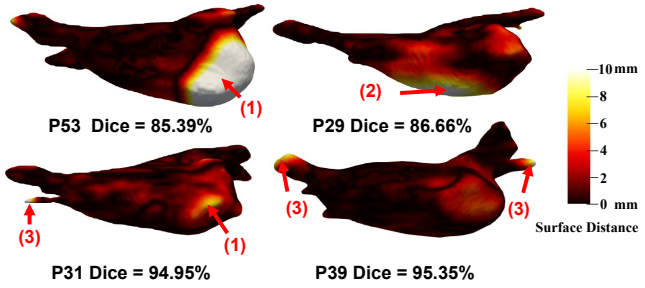

Fig. 3. 3D visualization of the two worst and best LA segmentation results achieved by the proposed method in terms of 3D dice score. The color on the surface indicates the distance from the prediction to the manual segmentation. The surface distances are scaled between 0 and 10 mm for better visualization. Arrows (1-3) highlight the errors in MV, the regions between LA and RA, and PV, respectively.

predictions are generally smooth and accurate. The main errors are on the MV (highlighted by arrow(1)), the regions between LA and the right atrium (RA) (highlighted by arrow(2)), and the PV (highlighted by arrow(3)), which are also shown in 2D visualization. The errors in the MV and the boundary regions can be attributed to the unclear boundary between LA and LV and the flat shape labeled by observers. The errors in the PV are mainly due to its long, thin, and varied shapes.

(3) 2D visualization

Fig. 4 displays three examples of the LA segmentation results obtained by our proposed method, Usformer, and two other methods, nnUNet and UNeXt. Our results demonstrate high accuracy in outlining LA segments. Furthermore, our method outperforms nnUNet and UNeXt with higher 2D dice scores, and the segmentation results are closer to the manual segmentation. These findings highlight the potential of our proposed Usformer for accurate LA segmentation.

Fig. 5 shows the 2D visualization of our LA segmentation results in the axial view. Each row displays three slices of a patient, with the 3D dice score denoted on the left and the slice index and 2D dice score at the top of each slice's visualization.

The proposed method achieves promising results in capturing complex segments and detecting slices without LAs. As shown in the last two rows of Fig. 5, our proposed method

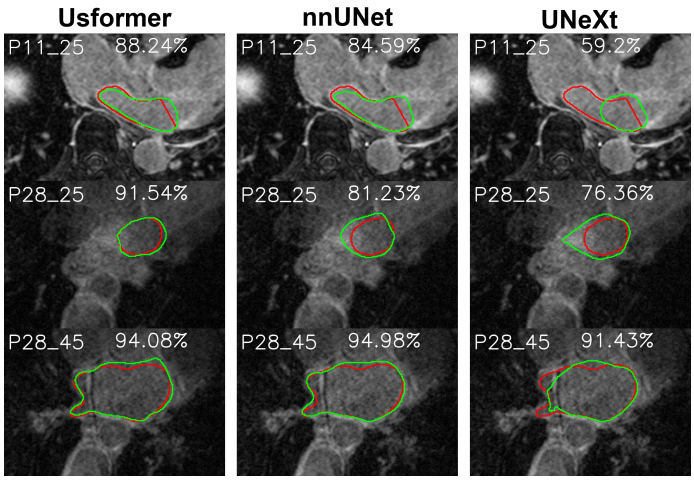

Fig. 4. Axial view of LA segmentation results by our proposed method Usformer, nnUNet [19], and UNeXt [20]. Contours of manual and predicted segmentation are denoted in red and green.

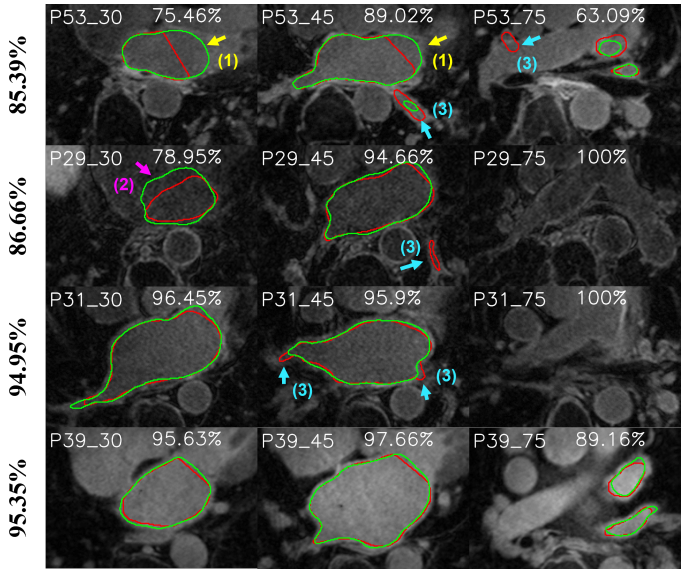

Fig. 5. Axial view of the two worst and best LA segmentation results in terms of 3D dice score by the proposed method. Contours of manual and predicted segmentation are denoted in red and green. The arrows (1-3) indicate the errors in MV, regions between LA and RA, and PV, respectively.

successfully outlines the LA segments, despite the complex shapes and low contrast between the LA and its surroundings. Additionally, the 75^{th} slices of patients $P29$ and $P31$ are successfully detected without LAs.

Our method has difficulty in predicting the MV and the boundary between LA and RA, as denoted by arrows (1) and (2). In the first case $P53$, the MV labeled by observers is a flat plane, but the proposed method predicted it as a circle, causing many false positives. The contrast around the error area is poor, and observers might segment the region with large variability, which confuses the network. Arrow (3) points out errors in PV. It is difficult to segment the thin and long PVs with varied shapes and lengths among patients.

IV. Conclusions

This study proposes Usformer, a light, fast, and accurate network for LA segmentation. Usformer adopts a UNet-like architecture that integrates abstract features with precise spatial control. Transformer blocks using transposed attention learn the long-range dependencies between pixels in large 3D volumes without incurring a high computational cost. Despite the poor image quality of LGE MRI, our method shows promising results in the 2018 Atrial Segmentation Challenge testing set with an average 3D dice of 92.6% and only 611K parameters in the network.

The main limitation is that the generalizability of the proposed method is not validated outside of the 2018 challenge dataset. The challenge dataset is limited in numbers, patients, vendors, etc. More experiments could be done to explore the model's generalization ability. Another potential area for future work is to combine different imaging modalities, such as other MRI sequences and computed tomography (CT), to provide a more comprehensive view of the LA anatomy.

Overall, the presented light Usformer accurately outlines the left atrium from LGE MRI with low computation memory, offering a flexible and viable alternative to costly manual segmentation. It is envisioned that the proposed network will be successful in other segmentation tasks.

Acknowledgment

This work was partially supported by funding from the National Institutes of Health (R01HL116895, R01HL151079, R21EB030806, R01HL167148) and American Heart Association (949899).

References

[1] K. Jamart, Z. Xiong, G. D. Maso Talou, M. K. Stiles, and J. Zhao, "Mini review: Deep learning for atrial segmentation from late gadolinium-enhanced mris," *Frontiers in Cardiovascular Medicine*, vol. 7, p. 86, 2020.

[2] H. Lin, B. Li, X. Wang, Y. Shu, and S. Niu, "Automated defect inspection of led chip using deep convolutional neural network," *Journal of Intelligent Manufacturing*, vol. 30, pp. 2525–2534, 2019.

[3] Y. Mao, H. Lin, C. X. Yu, R. Frye, D. Beckett, K. Anderson, L. Jacquemetton, F. Carter, Z. Gao, W.-k. Liao *et al.*, "A deep learning framework for layer-wise porosity prediction in metal powder bed fusion using thermal signatures," *Journal of Intelligent Manufacturing*, vol. 34, no. 1, pp. 315–329, 2023.

[4] Z. Xiong, Q. Xia, Z. Hu, N. Huang, C. Bian, Y. Zheng, S. Vesal, N. Ravikumar, A. Maier, X. Yang, P.-A. Heng, D. Ni, C. Li, Q. Tong, W. Si, E. Puybareau, Y. Khoudli, T. Géraud, C. Chen, W. Bai, D. Rueckert, L. Xu, X. Zhuang, X. Luo, S. Jia, M. Sermesant, Y. Liu, K. Wang, D. Borra, A. Masci, C. Corsi, C. de Vente, M. Veta, R. Karim, C. J. Preetha, S. Engelhardt, M. Qiao, Y. Wang, Q. Tao, M. Nuñez-Garcia, O. Camara, N. Savioli, P. Lamata, and J. Zhao, "A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging," *Medical Image Analysis*, vol. 67, p. 101832, 2021.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[6] K. K. Wong, A. Zhang, K. Yang, S. Wu, and D. N. Ghista, "Gcw-unet segmentation of cardiac magnetic resonance images for evaluation of left atrial enlargement," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106915, 2022.

[7] C. Bian, X. Yang, J. Ma, S. Zheng, Y.-A. Liu, R. Nezafat, P.-A. Heng, and Y. Zheng, "Pyramid network with online hard example mining for accurate left atrium segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 237–245.

[8] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 515–524, 2018.

[9] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2881–2890.

[10] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[11] S. Vesal, N. Ravikumar, and A. Maier, "Dilated convolutions in neural networks for left atrial segmentation in 3d gadolinium enhanced-mri," in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*. Springer, 2019, pp. 319–328.

[12] K. Jamart, Z. Xiong, G. M. Talou, M. K. Stiles, and J. Zhao, "Two-stage 2d cnn for automatic atrial segmentation from lge-mris," in *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers 10*. Springer, 2020, pp. 81–89.

[13] Q. Xia, Y. Yao, Z. Hu, and A. Hao, "Automatic 3d atrial segmentation from ge-mris using volumetric fully convolutional networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2018, pp. 211–220.

[14] X. Yang, N. Wang, Y. Wang, X. Wang, R. Nezafat, D. Ni, and P.-A. Heng, "Combating uncertainty with novel losses for automatic left atrium segmentation," in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*. Springer, 2019, pp. 246–254.

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[16] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5728–5739.

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[18] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, pp. 1–28, 2015.

[19] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[20] J. M. J. Valanarasu and V. M. Patel, "Unext: Mlp-based rapid medical image segmentation network," *arXiv preprint arXiv:2203.04967*, 2022.

[21] X. Yang, N. Wang, Y. Wang, X. Wang, R. Nezafat, D. Ni, and P.-A. Heng, "Combating uncertainty with novel losses for automatic left atrium segmentation," in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*. Springer, 2019, pp. 246–254.

[22] F. Li, W. Li, X. Gao, R. Liu, and B. Xiao, "Comprehensive information integration network for left atrium segmentation on lge cmr images," *Biomedical Signal Processing and Control*, vol. 81, p. 104537, 2023.