Technology and Health Care 32 (2024) S403–S413
DOI 10.3233/THC-248035
IOS Press

Optimizing cardiovascular image segmentation through integrated hierarchical features and attention mechanisms

Shijia Liao[a,1], Bin Wang[b,1] and Shiming Lin[a,c,*]
[a] *School of Informatics, Xiamen University, Xiamen, Fujian, China*
[b] *Emergency Department, Xiamen Cardiovascular Hospital of Xiamen University, Xiamen University, Xiamen, Fujian, China*
[c] *School of Information Engineering, Changji University, Changji, Xinjiang Uygur Autonomous Region, China*

Abstract.
BACKGROUND: Cardiovascular diseases are the top cause of death in China. Manual segmentation of cardiovascular images, prone to errors, demands an automated, rapid, and precise solution for clinical diagnosis.
OBJECTIVE: The paper highlights deep learning in automatic cardiovascular image segmentation, efficiently identifying pixel regions of interest for auxiliary diagnosis and research in cardiovascular diseases.
METHODS: In our study, we introduce innovative Region Weighted Fusion (RWF) and Shape Feature Refinement (SFR) modules, utilizing polarized self-attention for significant performance improvement in multiscale feature integration and shape fine-tuning. The RWF module includes reshaping, weight computation, and feature fusion, enhancing high-resolution attention computation and reducing information loss. Model optimization through loss functions offers a more reliable solution for cardiovascular medical image processing.
RESULTS: Our method excels in segmentation accuracy, emphasizing the vital role of the RWF module. It demonstrates outstanding performance in cardiovascular image segmentation, potentially raising clinical practice standards.
CONCLUSIONS: Our method ensures reliable medical image processing, guiding cardiovascular segmentation for future advancements in practical healthcare and contributing scientifically to enhanced disease diagnosis and treatment.

Keywords: Cardiovascular image segmentation, self-attention mechanism, medical image processing, diagnostic accuracy

1. Introduction

Cardiovascular diseases encompass heart and vascular conditions, including coronary heart disease, cerebrovascular diseases, arterial diseases, and venous thrombosis. The global incidence of cardiovascular diseases has been steadily increasing in recent years, posing a significant public health challenge. In China, with economic development and accelerated aging, the prevalence of cardiovascular diseases is becoming more pronounced. Therefore, it is crucial to enhance prevention and treatment strategies for

[1] Shijia Liao and Bin Wang contribute equally to this paper.
[*] Corresponding author: Shiming Lin, School of Informatics, Xiamen University, Xiamen, Fujian 36100, China; School of Information Engineering, Changji University, Changji, Xinjiang Uygur Autonomous Region 831100, China. E-mail: xmulsm@xmu.edu.cn.

0928-7329 © 2024 – The authors. Published by IOS Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (CC BY-NC 4.0).

cardiovascular diseases. Despite advancements in medical imaging technology, manual segmentation by experts is time-consuming and labor-intensive. The complexity of cardiovascular structures leads to subjective interpretations and significant impacts on optimal treatment strategy decisions. Therefore, there is an urgent need for the development of automatic segmentation techniques [1].

In recent years, deep learning-based medical image analysis has become a research hotspot due to its ability to rapidly and accurately process massive data, aiding physicians in improving diagnostic efficiency. Deep neural networks, especially in medical image segmentation, have shown outstanding performance, influencing modern cardiovascular disease diagnosis and treatment models. However, the heterogeneity, complex structures, and contrast variations present in medical imaging data, especially in cardiovascular contexts [2], require large amounts of high-quality annotations for model training.

To enhance model accuracy, we propose the integration of effective feature extraction methods [3], attention mechanisms [4], and multi-scale fusion strategies [5]. Additionally, we introduce a novel polarized self-attention strategy to further improve spatial and channel features, enhancing the segmentation model's performance. Automatic segmentation of cardiovascular images can significantly reduce the workload of physicians, save diagnostic time, predict cardiovascular diseases in patients in advance, and facilitate timely treatment, effectively reducing the mortality rate and mitigating the threat of cardiovascular diseases to human life and health. This research holds substantial significance for the diagnosis and treatment of cardiovascular diseases.

2. Literature review

Medical image segmentation for cardiovascular disease research has gained attention, historically relying on traditional methods like thresholding [6], edge detection [7], and region-growing algorithms [8]. While artificial intelligence has entered the field, deep learning applications in cardiovascular diagnosis from large-scale datasets are limited. Current deep learning focuses on convolutional neural networks, lacking efficiency exploration. Challenges include significant feature variations across medical image locations. Medical image segmentation can improve efficiency, but the shortage of high-quality annotated data prompts exploration of traditional supervised learning methods, like machine learning and attention mechanisms.

2.1. Evolution of machine learning in segmentation

Compared to traditional methods, deep learning iteratively improves segmentation accuracy through continuous training, especially with uniform characteristic images. Deep learning offers faster computation and adaptability for batch processing. U-Net's [9] paired encoder-decoder structures with skip connections address limitations and enhance pixel-level and semantic information for better segmentation. Wu et al. [10] introduced multiple supervised pathways for richer scale features. Angio-Net [11] enables continuous preprocessing modification, while SD-Net [12] simplifies the network for efficiency. CE-Net [13] captures advanced semantic information with a multiscale context encoding module. Despite successes, challenges persist due to cardiovascular image heterogeneity. Researchers explore multiscale features, attention mechanisms, and novel architectures [14] to address these challenges.

2.2. Attention mechanisms in segmentation

Inspired by human visual attention, attention mechanisms crucially enhance segmentation accuracy. The self-attention mechanism proposed by Vaswani et al. [15] allows focusing on relevant image regions,

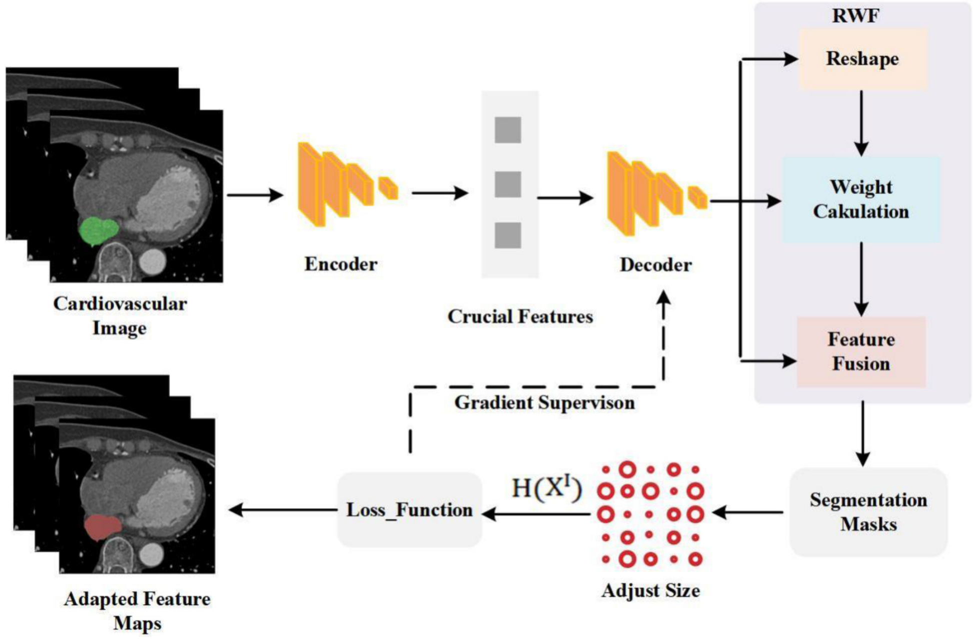

Fig. 1. The flowchart of proposed segmentation model.

while polarized self-attention [16] selectively retains essential features. These innovations aim to provide adaptive and robust segmentation models, vital for clinical practice. The effectiveness of attention mechanisms has garnered attention, with many attention mechanisms proposed and proven effective in improving network performance and reducing computational costs, among which the most widely used is CBAM (Convolutional Block Attention Module) [17] and its variants. Gou et al. [18] introduced SA-UNet based on SD-UNet, using dropblock instead of dropout and adding a spatial attention module. The improved network can better extract small vessels and reduce noise. ESA-UNet [19] introduced an enhanced spatial attention module (ESA), which can better adaptively redistribute features based on spatial context content, enabling the network model to understand more contextual information about the image. Automated segmentation methods contribute to increased efficiency, reducing the time burden on clinical practitioners for timely interventions. Overall, the transition from traditional to advanced machine learning technologies in cardiovascular image segmentation emphasizes addressing image heterogeneity and complex anatomical structures, aiming for better healthcare outcomes.

3. Materials and methods

3.1. Animal experiments to measure ABR

The suggested network architecture, shown in Fig. 1, adopts a novel method to represent the links between inter-class homogeneity to concentrate on cardiovascular image segmentation. Using an encoder to extract important features from cardiovascular pictures, the method generates segmentation masks for each kind of tissue by accurately separating cardiovascular tissues from the backdrop using a decoder. The method ensures adaptability by adjusting the feature map sizes to fit the feature map dimensions of the input cardiovascular image. By improving the precision and effectiveness of cardiovascular picture segmentation, this method seeks to offer a more dependable clinical medical image processing solution.

HRNetV1 is selected as the spine for feature extraction during the encoding step of the cardiovascular image segmentation procedure. There are two main benefits that this backbone has. First off, the network structure as a whole maintains high-resolution characteristics, which helps identify small tissue blocks in the images with accuracy. Second, HRNetV1 ensures efficient use of information contained in each resolution's feature map by combining feature maps of various resolutions, facilitating effective information interchange. Each input image yields four feature maps from the backbone.

During the decoding process, a Region Weighted Fusion (RWF) module is employed for multi-scale feature fusion. Subsequently, the fused feature maps are input into the Shape Feature Refinement (SFR) module to fine-tune learning of shape features, enabling interaction and influence among features. Finally, a classifier with 1 × 1 convolutions is used to adjust the channel numbers of the input image to match the desired classes, and the results are down sampled to the input image's size, yielding the semantic segmentation results for cardiovascular pathological images.

Introducing and integrating a Self-Attention module into the suggested RWF module is a creative departure from traditional feature fusion techniques. This module, which is based on the polarized self-attention mechanism, correlates channel features and learns the significance of spatial information at various scales. It then adaptively and selectively keeps important features while removing unnecessary ones. The model performs much better in cardiovascular image segmentation tasks thanks to its innovative design. For a thorough schematic of the RWF structure, see Fig. 2.

Three phases make up the RWF module in cardiovascular picture segmentation: feature fusion, weight computation, and reshaping. The first stage involves adjusting the feature map sizes ($t \in \{1, 2, 3, 4\}$) to yield (E'_O) by using nearest-neighbor interpolation. Then, a 3 × 3 convolution (U_3) yields enhanced feature information, which becomes $U_3\left(E'_O\right)$. The weights of $U_3\left(E'_O\right)$ are determined by introducing a self-attention mechanism in the second stage. Subsequently, the four weighted feature maps undergo 1 × 1 convolutions to modify their channel numbers, yielding the output weighted feature map At ($t \in \{1, 2, 3, 4\}$). In the third and final stage, the feature map with modified weights is obtained by multiplying $U_3\left(E'_O\right)$ and At pixel-by-pixel. The final feature map is created by concatenating these fused feature maps.

Inspired by the properties of polarizing lenses, a polarized self-attention mechanism is utilized in the second step of the RWF module. This mechanism, allows filtering of light in random directions, orthogonal to the horizontally passing light. In the attention calculation, a polarized filtering mechanism is established, maintaining high-resolution attention computation dimensions and reducing information loss. This innovative design enhances the model's performance in cardiovascular image segmentation tasks.

Equaattion (1) is used to calculate the feature tensor of a sample for cardiovascular image segmentation. This feature tensor is represented as $E_t^O \in R^{h \times w \times c}$, where h, w, and c stand for the feature tensor's height, width, and channel number, respectively. The channel attention weight $A_C\left(X\right)$ plus the spatial attention weight $A_S\left(X\right)$ add up to the attention weight H(·) that the attention mechanism outputs.

$$H\left(\cdot\right) = A_C\left(X\right) + A_S\left(X\right) \tag{1}$$

More specifically, Eq. (2) is used to determine the channel attention weight $A_C\left(X\right)$:

$$A_C\left(X\right) = F_{Si}\left(Y\left(U_3\left(B\left(U_1\left(X\right)\right) \times F_{So}\left(B_2\left(U_2\left(X\right)\right)\right)\right)\right)\right) \tag{2}$$

And the weight of spatial attention Eq. (2). Equation (3) is utilized to calculate $A_S\left(X\right)$:

$$A_S\left(X\right) = F_{Si}\left(B_3\left(F_{So}\left(B_1\left(G\left(U_2\left(X\right)\right)\right)\right) \times B_2\left(U_1\left(X\right)\right)\right)\right) \tag{3}$$

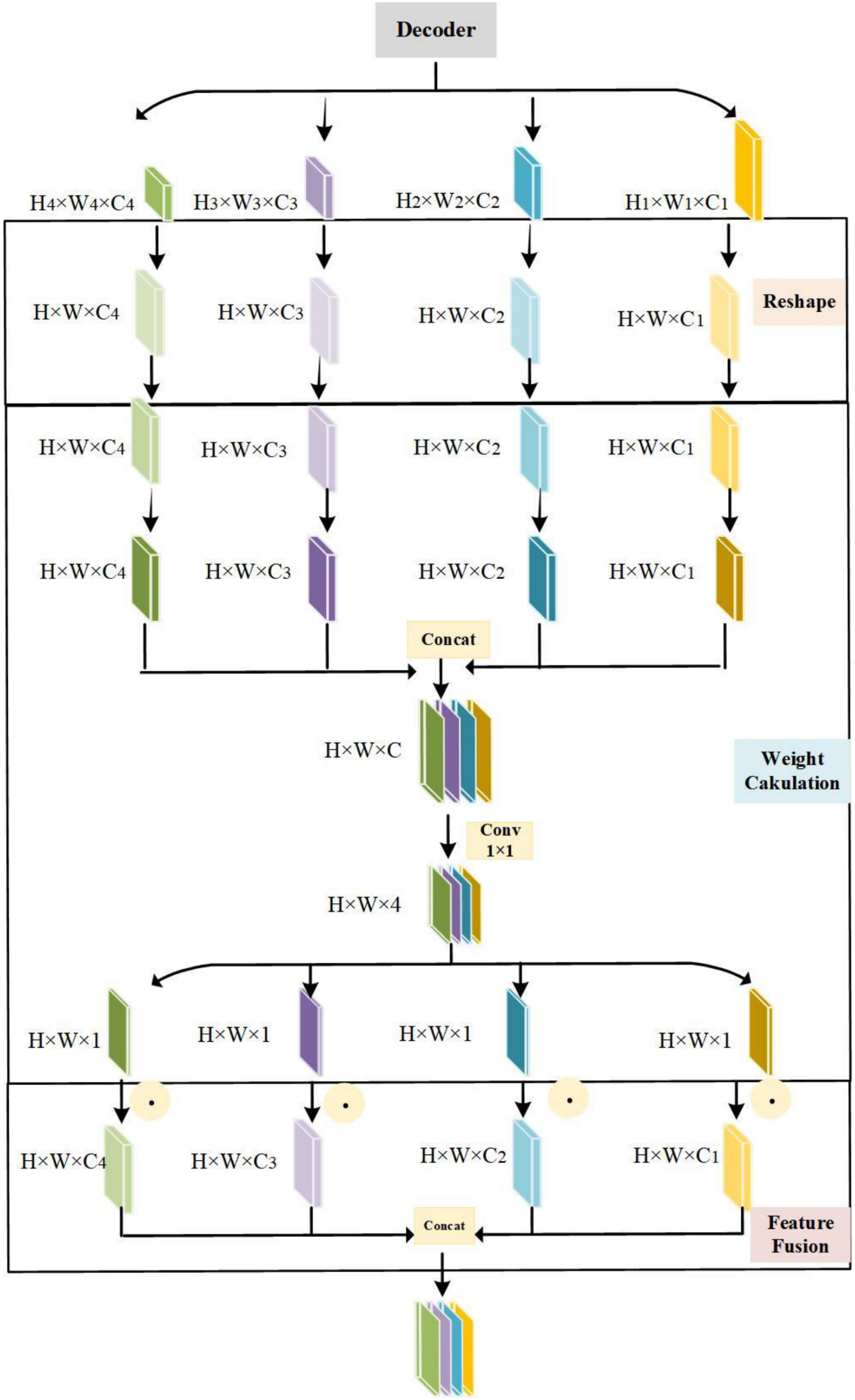

Fig. 2. Schematic diagram of the structure of the RWF module. It mainly includes three steps of remodeling, weight calculation and feature fusion.

The softmax operation, which increases the dynamic range of attention by normalizing, is represented by F_{so} in the functions above. The sigmoid operation is represented by F_{si}, and the probability distribution function that makes use of the high-resolution data kept in the attention branch is the softmax-sigmoid combination. G stands for global average pooling, X is the cross product, B_i is tensor reshaping, U_i is the $i \times i$ convolution operation, and Y is layer normalization.

The weights from each feature map are concatenated and fused in the third stage, known as "Feature Fusion," in order to preserve the weights of particular features within a single feature map as well as the significance of weights from separate feature maps. By removing unnecessary characteristics, this procedure helps to retain important feature information for tissue segmentation in an adaptable and selective manner.

3.2. Loss function

For comprehensive training, an effective loss function is crucial in image segmentation tasks. The Dice Loss is commonly used, defined as the sum of predicted and ground truth binary values divided by the total number of pixels. For multi-class segmentation, adding Cross-Entropy Loss is beneficial. To prevent overfitting, an L2 norm term for the network's parameters can be introduced. The final weighted loss function combines these components, ensuring efficient training and generalization of the cardiovascular image segmentation model. Adjusting weights allows fine-tuning based on the significance of each loss component for specific segmentation tasks.

$$\mathit{DiceLoss} = 1 - \frac{2 \times \sum_i^n (p_i \times g_i)}{\sum_i^n p_i^2 + \sum_i^n g_i^2} \tag{4}$$

$$\mathit{Cross} - \mathit{EntropyLoss} = -\sum\nolimits_i^n (g_i \times \log(p_i)) \tag{5}$$

$$\mathit{L2Regularization} = \lambda \times \sum\nolimits_i^N ||W_i||^2 \tag{6}$$

$$\mathit{TotalLoss} = w_{\mathit{Dice}} \times \mathit{DiceLoss} + w_{\mathit{CE}} \times \mathit{Cross} - \mathit{EntropyLoss} + w_{L2} \times \mathit{L2Regularization} \tag{7}$$

Where p_i and g_i are the predicted and ground truth binary values for each pixel, respectively, and n is the total number of pixels. W_i denotes the network's parameters, N is the total number of parameters, and λ is the regularization strength. The weights w_{Dice}, w_{CE}, and w_{L2} are modifiable to equalize the contributions of each loss component.

4. Experiments

4.1. Datasets

Our dataset for coronary computed tomography (CCTA) includes the proximal descending aorta (DA), inferior vena cava (IVC), coronary sinus (CS), right ventricular wall (RVW) and left atrial wall (LAW). These annotations were created and validated by board-certified cardiologists and were used as "ground truth" for the deep learning model.

We used standard segmentation evaluation metrics, including Hausdorff Distance (HD), Intersection over Union (IoU), and Dice Coefficient (Dice), to quantify the performance of our method and baseline

Table 1
Comparative performance of segmentation methods

Methods	Dice	IoU	HD
Proposed method	0.85	0.78	12.3
U-Net	0.78	0.69	15.2
DeepLab	0.79	0.72	14.5
Attention U-Net	0.81	0.75	13.8

Table 2
Deconstructing the impact of RWF module components

Model	Dice coefficient	IoU
Baseline Model (BM)	0.85	0.78
No Self-Attention (NSA)	0.82	0.75
No Polarized Self-Attention (NPSA)	0.83	0.76
No Feature Fusion (NFF)	0.81	0.74
No Reshaping (NR)	0.80	0.73

approaches. These metrics measure boundary dissimilarity, segmentation accuracy, and spatial overlap, respectively.

$$Dice = \frac{2\left|A \cap B\right|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \tag{8}$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \tag{9}$$

$$HD\left(A, B\right) = \max\left(supa \in A \text{ inf } b \in B\ d\left(a, b\right), \text{sup } b \in B \text{ inf } a \in A\ d\left(a, b\right)\right) \tag{10}$$

The distance between midpoints $(d(a, b))$ measures the dissimilarity between sets A and B, representing two segmentation masks. *TP, TN, FP*, and *FN* denote true positives, true negatives, false positives, and false negatives, respectively.

4.2. Results and comparative analysis

We evaluated our suggested approach's performance by contrasting it with a number of cutting-edge techniques for segmenting cardiovascular images, such as Attention U-Net [21], DeepLab [14], and U-Net [9]. These techniques serve as the foundation for our comparison study and are the industry standards as of right now. The suggested approach was trained using the same training parameters and in the same experimental setup as the comparison model to guarantee the experiment's fairness.

The experimental findings, which are shown in Table 1, demonstrate how successful our suggested strategy is in comparison to baseline techniques. In every evaluation indicator, the suggested approach surpassed the current ones with consistent results. The suggested approach outperformed U-Net by 0.08 and 0.09, respectively, and showed a notable increase in segmentation accuracy, obtaining a Dice of 0.85 and an IoU of 0.78. When compared to baseline approaches, this shows improved boundary alignment and spatial overlap. Furthermore, the HD is 2.9% less than U-Net's. The reduced HD provides more evidence of our method's effectiveness in precisely capturing anatomical details.

4.3. Qualitative analysis

A visual comparison of the segmentation findings for CS, DA, IVC, LAW, and RVW is presented in Fig. 3. Red indicates the segmentation outcomes of various techniques, while green indicates the

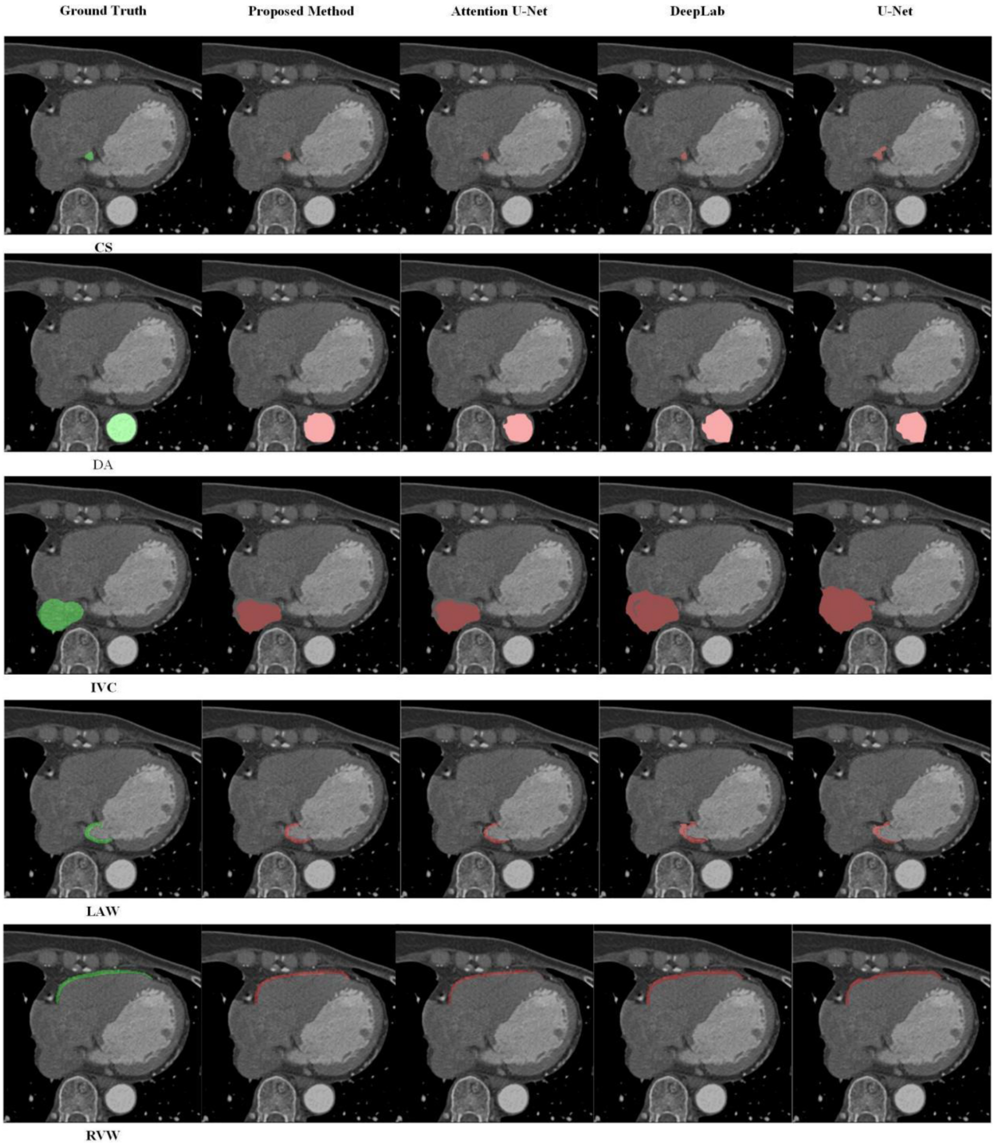

Fig. 3. Qualitative segmentation results.

ground truth in the segmentation findings. The suggested technique shows that it can correctly identify cardiac components even under difficult conditions with fluctuating contrast and intricate anatomical features. The unique incorporation of the polarized self-attention mechanism, which selectively preserves relevant aspects while suppressing irrelevant information, is credited with our suggested method's higher performance. The robustness of this approach is further demonstrated by its capacity to adapt to various imaging situations.

In these sampled regions, all models used for comparison returned some errors and missed detection results. For the CS structure, both Attention U-Net and DeepLab failed to completely identify the CS structure, and U-Net struggled to differentiate CS structure from other relevant matrix parts, resulting in the segmentation of other relevant matrix parts as CS. In the DA structure, except for our proposed model, other comparative models failed to correctly segment the boundaries of the DA. For the IVC structure, DeepLab and U-Net enlarged the segmented part of the IVC structure, significantly impacting the segmentation results. In the LAW structure, although parts of the LAW structure are relatively small, noticeable defects and distorted segmentation shapes are evident in the results of Attention U-Net, DeepLab, and U-Net. Finally, in the RVW structure, our proposed method, along with Attention U-Net and DeepLab, achieved segmentation results closely resembling the Ground Truth, while U-Net led to incomplete segmentation structures.

The results show, in summary, that our suggested method outperforms existing methods in cardiovascular image segmentation. Experimental results demonstrate that the proposed improvement method effectively enhances the segmentation outcomes of cardiovascular images. Comparative analysis with other state-of-the-art segmentation methods indicates the superiority of the proposed segmentation approach, showcasing excellent segmentation results.

4.4. Deconstructing the impact of RWF module components

To validate the effectiveness of our method, ablation experiments deconstructed the effect of RWF module components on cardiovascular image segmentation. The experimental configuration includes baseline model (BM), no self-retention (NSA), no polarization self-retention (NPSA), no feature fusion (NFF), and no remodeling (NR) modules. Evaluation metrics such as Dice coefficient and IoU on the validation set show that the results of the ablation experiments indicate that each element of the RWF module contributes significantly to the segmentation performance of the model. The effective capture of spatial features depends heavily on the self-attention mechanism, especially its polarized version. The feature fusion stage also illustrates the importance of retaining important data for organizing segmentation. Although the contribution of the remodeling stage was not significant, good results were achieved.

In conclusion, the accuracy and efficiency of the cardiovascular image segmentation model was improved by the new components of the proposed RWF module. The thorough analysis makes it easier to understand the different functions of each module and also provides suggestions for future iterations of the model for greater improvement and refinement.

5. Conclusions and future work

5.1. Conclusions and future directions

Cardiovascular disease, a significant health threat, necessitates advanced segmentation methods. Our study introduces an innovative cardiovascular image segmentation approach, optimizing segmentation through hierarchical features and attention mechanisms. Utilizing a polarized self-attention strategy, the method achieves refined segmentation, excelling in distinguishing cardiac structures. Key contributions and conclusions include:

Improved Segmentation Accuracy: Experimental results show enhanced precision in metrics like Intersection over Union (IoU) and Dice similarity coefficient.

Adaptability to Varied Imaging Conditions: The method exhibits high segmentation capabilities under contrast, resolution, and anatomical structure variations.

Enhanced Understanding of Image Features: Integration of polarized attention improves the model's recognition and utilization of cardiovascular image features.

5.2. Future work

Future research will explore:

Vascular Segmentation: Addressing cardiovascular vessel segmentation and extending the method to other vascular segmentation tasks.

Data Challenges: Tackling segmentation challenges in cardiovascular image data with small samples or limited annotations.

Enhanced Accuracy: Improving segmentation accuracy for tasks like organ size assessment through more powerful feature extraction networks.

Clinical Integration: Exploring how segmentation outcomes can specifically contribute to clinical-assisted diagnostic processes.

Integrated Approaches: Implementing an integrated approach involving multiple segmentation models to establish a comprehensive diagnostic platform for cardiovascular diseases.

Acknowledgments

This research was supported by the Natural Science Foundation of Xinjiang Uyghur Autonomous Region (2023D01C55) and Scientific Research Program of the Higher Education Institution of XinJiang (XJEDU2023P127).

Conflict of interest

The authors have no conflicts of interest to declare.

References

[1] Bai W, Sinclair M, Tarroni G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of Cardiovascular Magnetic Resonance. 2018; 20(1): 1-12.

[2] Burton RAB, Plank G, Schneider JE, et al. Three-dimensional models of individual cardiac histoanatomy: tools and challenges. Annals of the New York Academy of Sciences. 2006; 1080(1): 301-319.

[3] Boulares M, Alotaibi R, AlMansour A, et al. Cardiovascular disease recognition based on heartbeat segmentation and selection process. International Journal of Environmental Research and Public Health. 2021; 18(20): 10952.

[4] Zhuang X, Shen J. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Medical image analysis. 2016; 31: 77-87.

[5] Wang W, Ye C, Zhang S, et al. Improving whole-heart CT image segmentation by attention mechanism. IEEE Access. 2019; 8: 14579-14587.

[6] Lee HY, Codella NCF, Cham MD, et al. Automatic left ventricle segmentation using iterative thresholding and an active contour model with adaptation on short-axis cardiac MRI. IEEE Transactions on Biomedical Engineering. 2009; 57(4): 905-913.

[7] Singleton HR, Pohost GM. Automatic cardiac MR image segmentation using edge detection by tissue classification in pixel neighborhoods. Magnetic Resonance in Medicine. 1997; 37(3): 418-424.

[8] Yi J, Ra JB. A locally adaptive region growing algorithm for vascular segmentation. International Journal of Imaging Systems and Technology. 2003; 13(4): 208-214.

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

[10] Wu Y, Song Y, et al. Vessel-Net: Retinal vessel segmentation under multi-path supervision[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. Springer International Publishing, 2019: 264-272.

[11] Iyer K, Najarian CP, Fattah AA, et al. Angionet: a convolutional neural network for vessel segmentation in X-ray angiography. Scientific Reports. 2021; 11(1): 18066.

[12] Guo C, Szemenyei M, Pei Y, et al. SD-UNet: A structured dropout U-Net for retinal vessel segmentation[C]//2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE. 2019: 439-444.

[13] Gu Z, Cheng J, Fu H, et al. Ce-net: Context encoder network for 2d medical image segmentation. IEEE Transactions on Medical Imaging. 2019; 38(10): 2281-2292.

[14] Chen LC, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017; 40(4): 834-848.

[15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in neural information processing systems, 2017; 30.

[16] Liu H, Liu F, Fan X, et al. Polarized self-attention: Towards high-quality pixel-wise regression. ar**v preprint ar**v2107.00782, 2021.

[17] Jobson DJ, Rahman Z, Woodell GA. Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing. 1997; 6(3): 451-462.

[18] Jobson DJ, Rahman Z, Woodell GA. Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing. 1997; 6(3): 451-462.

[19] Jobson DJ, Rahman Z, Woodell GA. Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing. 1997; 6(3): 451-462.

[20] Baskaran L, Al'Aref SJ, Maliakal G, Lee BC, Xu Z, Choi JW, et al. Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning. PLoS ONE. 2020; 15(5): e0232573.

[21] Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: Learning where to look for the pancreas. ar**v preprint ar**v1804.03999, 2018.