2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Achieving High Efficiency Heart Image Segmentation in U-Net by Means of Early Fusion and Contextual Information Reconstruction

Huijuan Hao[1,2*] and Wenpeng Wang[1,2]

1 Key Laboratory of Computing Power Network and Information Security, Ministry of Education,Shandong Computer Science Center(National Supercomputer Center in Jinan),Qilu University of Technology (Shandong Academy of Sciences), Jinan, China.
2 Shandong Provincial Key Laboratory of Computer Networks,Shandong Fundamental Research Center for Computer Science, Jinan, China.
haohj@sdas.org

Abstract. Cardiovascular diseases pose a significant threat to global health, making accu- rate cardiac MRI segmentation crucial. However, this task is hindered by complex anatomies, multi - scale integration issues, poor feature handling, long - range dependency problems, and limitations of existing methods. To address these challenges, this study introduces the EFCIR - U architecture based on the clas- sic U - shaped encoder - decoder framework. Confronted with complex cardiac anatomies, it uses Recurrent RSU modules instead of traditional convolutional layers for efficient multi - scale feature extraction and fine - grained detail cap- ture. To solve multi - scale integration issues, EFCIR - U adopts early fusion strategies: in the encoder, Recurrent RSU - generated feature maps are fused with incoming data for the CIR module to integrate multi - scale information, and in the decoder, up - scaled feature maps are fused with relevant ones in the CIR module during up - sampling for better cardiac region reconstruction. To handle poor feature extraction and perception, the CIR module in the U - shaped model integrates context through separation and reconfiguration to optimize these aspects. To tackle long - range dependency problems, the Mamba block in the decoder captures long - range dependencies and fuses multi - scale features. These components together enhance computational efficiency, feature representation, segmentation accuracy, and generalization, enabling EFCIR - U to overcome existing method limitations and provide a more efficient and accurate solution for high - quality cardiac image segmentation, thus contributing to better patient care considering the global threat of cardiovascular diseases.

Keywords: Cardiac MRI segmentation EFCIR - U architecture Multi - scale features

1 Introduction

Cardiovascular diseases pose a significant threat to global health, taking a heavy toll on human well - being. In medical diagnosis and treatment, precisely segmenting high - quality cardiac magnetic resonance imaging (MRI) is of great significance as it offers crucial information about cardiac function, morphology, and potential patholo- gies, enabling medical professionals to intervene in a timely and accurate manner for patients [1, 2]. However, this task is fraught with difficulties. The complexity of car- diac anatomical structures, coupled with low image contrast, partial volume effects, and significant inter - patient anatomical variations, makes it extremely challenging to achieve high - fidelity segmentation. For instance, low contrast blurs the boundaries between different cardiac tissues, making it hard to accurately demarcate the left and right ventricles, and partial volume effects can distort tissue volume representation, leading to inaccuracies in quantitative analysis.

Although the emergence of deep learning, especially convolutional neural networks (CNNs), has revolutionized cardiac MRI segmentation, with architectures like U - Net and its derivatives being popular for their ability to capture local and global con- text [6], contemporary deep - learning - based models still have limitations. Complex anatomical structures in high - quality cardiac images make it difficult to capture long- range dependencies and global context, as seen in models like SwinUNet and Tran- sUNet which struggle in regions with complex anatomical features [4]. Deep network archi- tectures also come with high computational costs, large memory requirements, long training times, and issues like gradient vanishing or explosion that can hinder model convergence. Additionally, the loss of crucial details during network process- ing due to decreasing feature map resolution is a major setback for high - precision medical segmentation, especially for cardiac images where fine - grained details are essential for accurate diagnosis.

To address these challenges, this study presents the EFCIR - U architecture based on the classic U - shaped encoder - decoder framework. EFCIR - U incorporates multiple innovative techniques. It uses Recurrent RSU modules instead of traditional convolutional layers, which are effective in capturing multi - scale features and extract- ing rich contextual information at different stages. By adjusting the sampling rate and fusing features through up - sampling, feature cascading, and convolution, the RSU minimizes detail loss, enhancing feature extraction and enabling precise capture of fine - grained details for accurate identification of cardiac structural details. The architecture also adopts early fusion strategies. In the encoder, the feature maps gen- erated by the Recurrent RSU modules are fused with the incoming data at an early encoding stage, providing the CIR module with more comprehensive multi - scale information. This early integration allows the model to better understand the cardiac image from the start. In the decoder, during the up - sampling process, the up - scaled feature maps are fused with relevant ones in the CIR module, refining the feature rep- resentation for better cardiac region reconstruction. This early fusion in the decoder improves segmentation accuracy and model generalization. The CIR module, embed- ded in the U - shaped model, optimizes feature extraction through separation and recon- figuration operations, enhancing the model's generalization ability and reducing noise.

Moreover, EFCIR - U strategically integrates the Mamba block in the decoder pathway, which has advantages such as capturing long - range dependencies and multi- scale feature fusion. Overall, this paper aims to introduce a novel network architec- ture that over- comes the limitations of existing methods in high - quality cardiac image segmentation, providing a more efficient and accurate solution for medical diagnosis and treatment and contributing to better patient outcomes.

2 METHOD

2.1 Overall Frameworks

This paper explores advanced reconstruction methods for high quality cardiac image segmentation in deep U - shaped networks and introduces the EFCIR - U architecture to enhance accuracy and efficiency with low model complexity (Fig. 1). The EFCIR - U starts with a high quality cardiac image input.

In the encoder part, after the down - sampling stage using Recurrent RSU mod- ules for sequential down sampling and efficient feature extraction (while maintaining resid- ual connections for smooth information flow), the resulting feature maps are fused with the image in the normal process that is meant to enter the CIR module. This early fusion in the encoder helps in better integrating the multi - scale information at an earlier stage, enhancing the model's ability to capture complex features.

Wavelet fusion is then used to integrate pre - sampling and in - process information into the CIR module to address challenges related to high - frequency detail preserva- tion and multi - scale information acquisition. The encoded feature maps then enter the decoder pathway with a Mamba block, which offers advantages such as capturing long - range dependencies, high - efficiency feature processing, excellent multi - scale fea- ture fusion, and strong adaptability to diverse datasets, thus improving segmentation accuracy and generalization.

In the decoder stage, during the up - sampling process with bilinear interpolation, the up - scaled feature maps are again fused with the feature maps in the normal process within the CIR module. This early fusion in the decoder further refines the feature representation, enabling more accurate reconstruction of the cardiac region. Finally, feature fusion by concatenation generates a precise cardiac region mask.

2.2 Residual Structure Unit (RSU)

The RSU plays a vital role in improving cardiac image processing quality. It adeptly captures multi scale features and extracts abundant contextual information at differ- ent stages. When more encoders are incorporated, the RSU structure becomes deeper,and additional pooling operations expand its receptive field, empowering it to capture both local and global features. The RSU first gradually decreases the sampling rate and then fuses features into high resolution maps through up sampling, feature cas- cading, and convolution, thus minimizing the loss of details. In contrast to traditional convolutional

methods, it remarkably improves feature extraction and propagation. Its weight transformation and local feature fusion mechanisms enhance the feature representational ability, enabling the precise capture of fine grained details. In cardiac imaging, a model equipped with RSU can more effectively extract the intricate edge information and overall characteristics of cardiac structures, ensuring the accurate identification of structural details. The EFCIR-U that integrates RSU blocks can efficiently process multi scale features, enhancing the precision of identifying and segmenting various cardiac regions, ranging from minute vascular structures to large anatomical areas.

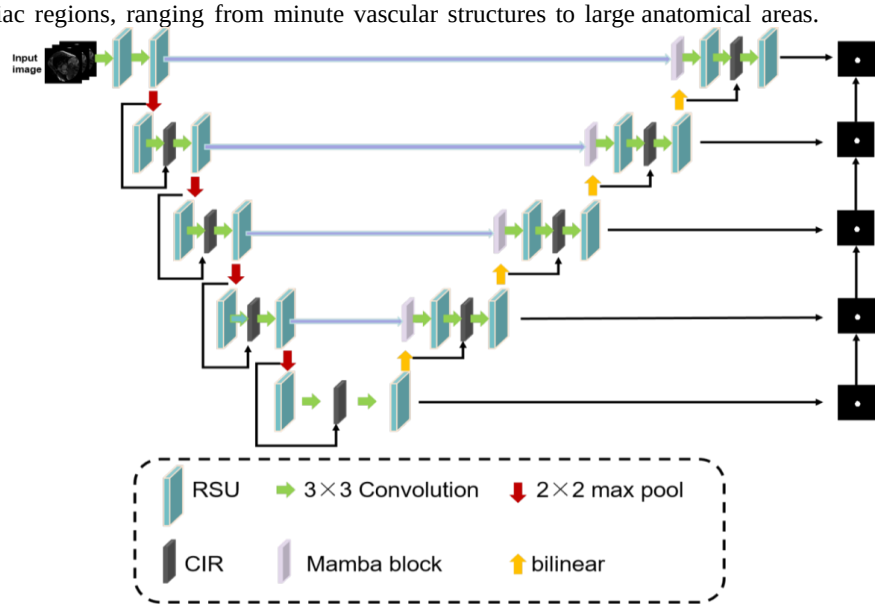

Fig. 1. The architecture of EFCIR-U

2.3 Contextual Information Reconstruction (CIR)

In this section, we will offer a thorough and in-depth introduction to a brand-new CIR module. This module is embedded at certain designated positions within the U-shaped model, aiming to integrate contextual information for the purpose of reconstruction tasks. This particular approach demonstrates remarkable superiority in handling complex and dynamic high-quality cardiac medical images, enabling the extraction of more comprehensive and complete image features. An overall view of the framework is depicted in Figure 2.

The CIR method cleverly combines separation and reconfiguration operations. By integrating these operations, it effectively extracts the vital contextual information from the feature maps. First and foremost, an advanced feature fusion operation is carried out on the feature maps that are fed into the CIR. After that, the fused features are input into the CIR for the separation process. Before this separation takes place, the input feature maps, represented as X, go through processing via Group

Normalization (GN) [10]. This step effectively differentiates between the feature regions that are rich in information and those that are sparse in information, thus

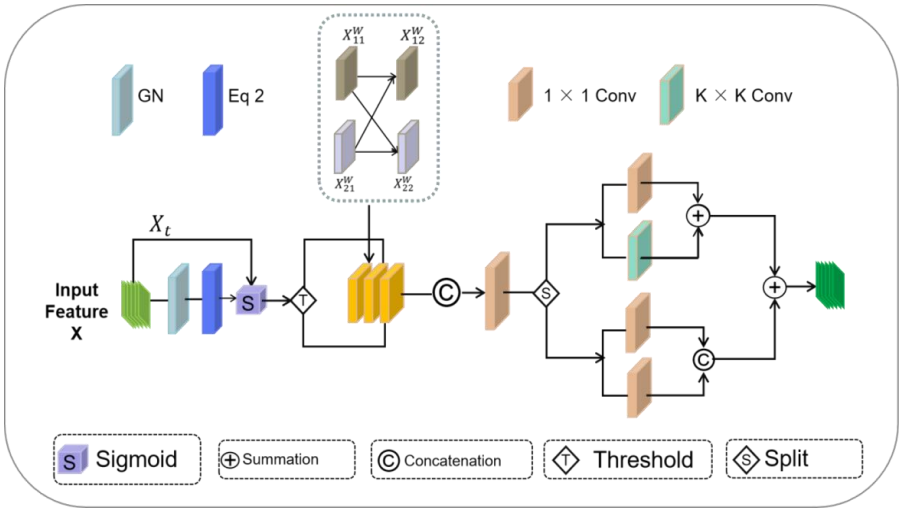

Fig. 2. The architecture of Contextual Information Reconstruction (CIR).

increasing the sensitivity and discriminative ability of the feature representation. The specific formula is as follows:

$$X_i = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta, \tag{1}$$

where, μ and σ represent the mean value and the standard deviation of X respectively. Meanwhile, βrepresents the learnable affine transformation parameter in the GN layer, and ε is a small constant added to ensure numerical stability. Subsequently, the channel weights W_γ, which are calculated by using the normalization scaling factor γ, are used to measure the importance of different feature maps. The specific formula is as follows:

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\Sigma_{j=1}^{C} \gamma_j} X_i, i, j = 1, 2, ..., C \tag{2}$$

Subsequently, for the feature maps fused based on contextual information, a Sigmoid function along with a Threshold mechanism is applied to map the weights to binary values within the range of (0, 1). Specifically, information weights W_1 above the threshold are set to 1, while those below the threshold W_2 are set to 0, as shown in Equation 3.

Next, for the feature maps that are fused based on contextual information, a combination of a Sigmoid function and a Threshold mechanism is applied. This is done to map the weights to binary values within the range of (0, 1). Specifically, the information weights W_1 that are above the threshold are set to 1, while those weights W_2 that are below the threshold are set to 0, as shown in Equation 3.

$$W = \text{Gate}(\text{Sigmoid}(W_\gamma X_t)) \tag{3}$$

After that, these binary weights are utilized to partition the input feature map X into two components: the information-rich component X_1^W and the information-poor component X_2^W. The calculation formulas for these two components are as follows:

$$X_1^W = W_1 \circledR X,$$
$$X_2^W = W_2 \circledR X, \tag{4}$$

where, ® represents the element-wise multiplication. Subsequently, the two feature maps are combined into a refined feature map through a cross-reconstruction oper- ation. The resulting refined feature map then undergoes a simple convolutional operation. This operation helps in efficient feature extraction, optimization of the computational process, enhancement of the model's generalization ability, and aug- mentation of features as well as reduction of noise during context reconstruction. Right after that, the obtained feature map is divided into two parts, denoted as and, according to a preset ratio α. The formula is shown as follows:

$$X_1, X_2 = \text{Split}(X^w, \alpha) \tag{5}$$

The spatially refined feature map is initially split into X_1 and X_2, which are further divided for different operations: the two parts from X_1 undergo 1×1 and $k \times k$ convolutions respectively and the results are summed, while the two segments of X_2 each have 1×1 convolutions and the outputs are concatenated. These operations efficiently extract and fuse multi scale features in cardiac images, capturing detailed and global information, enhancing the model's robustness and expressiveness, especially suitable for high quality medical cardiac image processing in deep neural networks. The CIR method identifies important contextual features and dynamically adjusts feature distribution during reconfiguration based on weights, emphasizing key information, greatly improving the model's ability to perceive complex spatial features and balance between fine grained details and global semantics, thus providing more comprehensive task support.

3 Experiments

3.1 Dataset

This study utilizes two publicly available cardiac imaging datasets—Sunnybrook[11], and RVSC[19] datasets—to evaluate the performance of cardiac MRI image segmentation algorithms. The Sunnybrook dataset, derived from the 2009 Left Ventricle Segmentation Challenge, contains 805 MRI images from 45 patients, covering four pathological states: healthy, left ventricular hypertrophy, heart failure with infarction, and heart failure without infarction. The dataset is randomly sampled into training, validation, and test sets to comprehensively assess algorithm performance. The RVSC dataset provides DICOM-standard cardiac MRI images from 48 patients, covering the entire cardiac cycle, with high-precision segmentation of the endocardium and

epicardium. It is divided into training, validation, and test sets, further supporting algorithm training and evaluation. These datasets provide a rich experimental foundation for the development and performance validation of cardiac image segmentation models.

3.2 Evaluation Metrics

Our evaluation framework incorporates the metrics specified by relevant challenges, including a range of specific segmentation indicators. The details are as follows:

- Dice (Dice Coefficient) = 2 * TP / (FP + FN + 2 * TP),
- IoU (Intersection Over Union) = TP / (FP + FN),
- Acc (Accuracy) = (TN + TP) / (TN + TP + FN + FP),
- HD95(Hausdorff Distance 95%): It is a measure of segmentation boundary quality, evaluating accuracy by calculating the maximum distance between the predicted boundary and the corresponding ground truth within the top 95.

TP, FP, TN and FN were true positive, false positive, true negative and false negative respectively. For the RVSC dataset experiments, we use Accuracy (Acc), Intersection over Union (IoU) and Dice as evaluation metrics. For the Sunnybrook Dataset, we apply Dice, IoU, at the 95th percentile (HD95), aligning with common practices in the field.

3.3 Implementation

All experiments were implemented with Python 3.9.5 and PyTorch 1.9.1, and model training was executed on the Quadro RTX 6000. In the EFCIR-U model, distinct data augmentation techniques and loss functions were applied to different datasets, such as the Sunnybrook dataset and the RVSC dataset. The hyperparameter configurations are as follows:

- For the Sunnybrook dataset, the input size was set to 256 × 256, the batch size was 2, the learning rate was 1e - 5, the optimizer was Adam with default β_1 = 0.9, β_2 = 0.999, ϵ = 1e - 8, and the loss function was Dice Coefficient Loss. The model was trained for 100 epochs.
- Regarding the RVSC dataset, the input size was 128 × 128, the batch size was 64, the learning rate was 4e - 5, the optimizer was Adam with β_1 = 0.9, β_2 = 0.999, ϵ = 1e-8, and the loss function was a combination of Dice Loss and Cross Entropy Loss. The training was carried out for 150 epochs.

3.4 Comparison with state-of-the-art approaches

The EFCIR-U model proposed in this paper has achieved remarkable progress in ventricular segmentation. It reached optimal Dice Similarity Coefficient (Dice) values of 87.59% and 95.62% on the important public datasets RVSC and SunnyBrook respectively, representing a significant performance improvement over existing methods. The

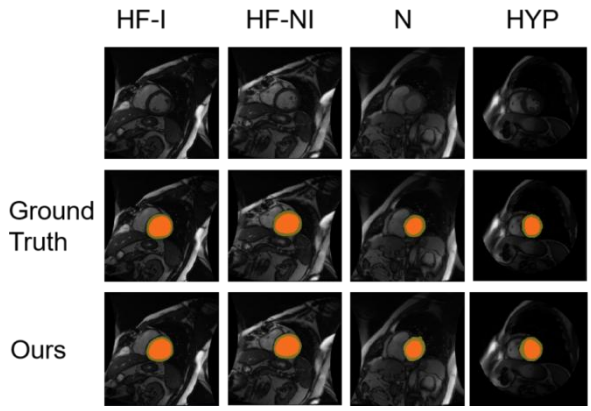

Fig. 3. Visualization of segmentation results for EFCIR-U on the Sunnybrook Dataset. In the images,red represents the endocardial boundary, while yellow indicates the epicardial boundary.

model has been optimized to have only 17.42 million parameters, reducing computational costs and significantly decreasing the risk of overfitting. Despite having 24.89 billion FLOPS, slightly lower than SwinUNet [5], EFCIR-U's testing performance on these two datasets is significantly better, demonstrating its excellent balance between efficiency and performance.

Table 1. In the comparative experiment on the SunnyBrook dataset, we used Dice, IoU, and HD95 metrics to evaluate the performance of the proposed architectures. The best results are indicated in bold.

Methods	Params(M)	Dice	IoU	HD95
UNet[3]	31.04	90.00	86.00	2.33
UNet++[7]	36.17	93.86	89.77	2.29
U^2Net[13]	44.00	92.70	82.38	3.84
TransUNet[6]	66.80	92.66	89.34	2.38
Swin-UNet[5]	27.14	85.11	81.22	5.81
MedFormer[12]	28.07	93.92	90.26	2.26
nnU-Net[18]	30.8	93.52	91.46	2.31
TransCeption[15]	22.25	94.19	89.84	2.23
LeViT-UNet[14]	35.09	92.76	87.29	4.16
PVT_CASCADE[16]	35.27	93.11	88.86	2.28
MERIT[17]	82.50	94.43	89.83	2.37
Ours	**17.42**	**95.62**	**90.33**	**2.21**

Experimental results on the Sunnybrook Dataset

We assessed EFCIR-U using the Sunnybrook MRI dataset, with the results detailed in Table 1. The quantitative results presented in Table 1 illustrate that the proposed

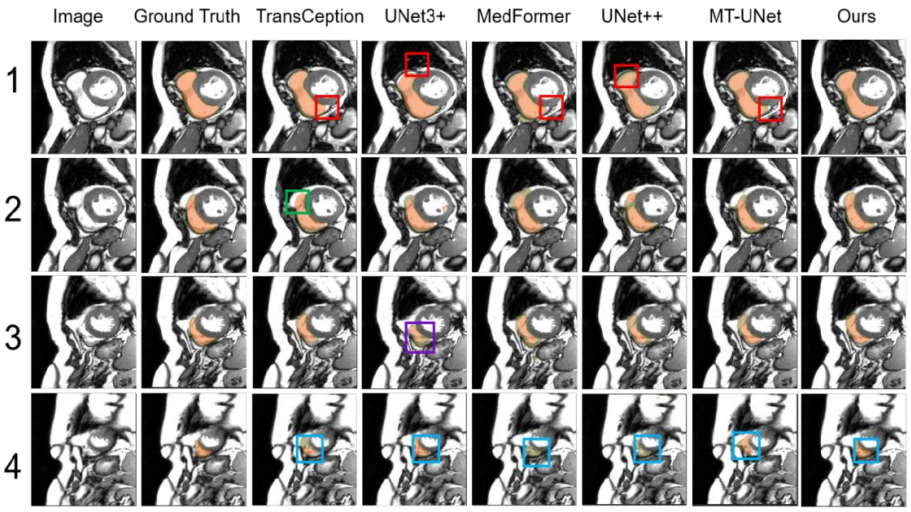

Fig. 4. From top to bottom, the visualization results of the right ventricle on the RVSC dataset for the same patient during the same cardiac cycle are shown for different segmentation methods. Yellow indicates the endocardial boundary, while green represents the epicardial boundary.

EFCIR-U outperforms in three evaluation metrics (Dice, IoU, HD95) for left ventricular tissue on the test set. This finding suggests that EFCIR-U has the potential to effectively manage pathological changes. Specifically, with a model size of 17.42 M, EFCIR-U achieved an average Dice score of 95.62%. EFCIR-U attains the highest aver- age values and the lowest standard deviations across the three segmentation metrics,reflecting its robustness in managing pathological changes in the left ventricle that could complicate morphological feature extraction by neural networks. Additionally, we visualized segmentation results for four distinct cardiac pathological conditions: N, HYP, HF-I, and HF-NI. As illustrated in Figure 3, various diseases can impact patients' hearts, resulting in abnormal morphological changes. Our Global context analysis module, EFCIR-U, has been demonstrated to be especially advantageous for cardiac segmentation.

Experimental results on the RVSC Dataset.

In this study, we divided the RVSC dataset into two independent test sets to evaluate the generalization capability of EFCIR-U. Table 2 summarizes the average experi- mental results of EFCIR-U on these two test sets. EFCIR-U performs exceptionally well on both RVSC test sets, achieving an average Dice coefficient of 87.59%, an IoU of 76.99%, and an accuracy of 93.30%, making it one of the best-performing models. This is primarily due to the model's design, which emphasizes enhancing global fea- ture analysis, reducing the number of parameters, and improving multi-scale feature extraction and fusion capabilities.

Figure 3presents a visual analysis of the RVSC dataset, focusing on MRI images of the same patient across a single cardiac cycle. This analysis contrasts the performance of different segmentation models and provides a clear description of the differences in the

delineation of the right ventricle (RV) boundary. The data in Table 2 and the visual results in Figure 3 together highlight the superior performance of EFCIR-U in terms of segmentation accuracy and boundary precision, significantly outperforming other competing methods.

By leveraging the power of convolution operations for local high-level feature extraction and employing effective strategies to expand the receptive field, we significantly reduce the false positive rate. Figure 5 demonstrates the superiority of our method: in the first row, the misidentification of the RV boundary (red box), in the second row, the error in epicardial identification by TransCeption [15] (green box), and in the third row, the erroneous identification of the RV endocardium by the UNet3+ [8] model (purple box) are all clearly visible. In the fourth row (blue box), our method significantly improves the segmentation and boundary accuracy of small structures during the RV systolic phase. The comparative analysis confirms that EASNet significantly enhances segmentation performance and boundary delineation. Unlike other models that struggle with the complex morphology of the RV, our design optimizes global boundary integration and reduces edge effects.

By utilizing rich feature hierarchies and contextual information, our model excels in accurately locating the RV boundary and performing overall morphological segmentation.

Table 2. In the comparative experiments conducted on the RVSC dataset, we evaluated two separate test sets and recorded their average values. To assess the performance of the proposed architecture, we employed metrics including Dice, IoU, and Accuracy. The best results are indicated in bold.

Methods	Params(M)	Dice	IoU	Accuracy
UNet[3]	31.04	81.72	72.95	92.14
UNet++[7]	36.17	82.65	75.25	93.68
UNet3+[8]	26.97	75.89	67.60	91.65
U^2Net[13]	44.00	80.63	72.69	92.21
TransUNet[6]	66.80	72.82	63.12	85.41
MT-UNet[9]	78.87	82.82	74.66	92.14
MedFormer[12]	88.93	84.68	72.26	88.24
TransCeption[15]	22.25	64.92	60.01	83.92
MERIT[17]	82.5	78.06	60.22	71.94
PVT_CASCADE[16]	34.13	79.76	70.08	90.12
Ours	**17.42**	**87.59**	**76.99**	**93.30**

3.5 Ablation Studies

To comprehensively evaluate the necessity and effectiveness of the enhancements in EFCIR-U, we performed an extensive ablation study with the ACDC dataset. Using the Dice score as the primary metric, we quantified the impact of each proposed modification on segmentation accuracy. This rigorous analysis allowed us to systematically determine the contribution of each improvement, thereby validating their role in enhancing the model's performance and reliability.

Table 3. In the ablation experiments for the ACDC dataset, we used the Dice coefficient as the metric to evaluate the performance of the proposed modules, CIR and Mamba.

Architecture	Module		Params(M)	Flops(G)	Dice %
	CIR	Mamba			
Model			44.00	28.83	87.04
Model_2		√	**14.58**	**24.44**	87.23
Model_3	√		44.37	29.08	92.32
Ours	√	√	17.42	24.89	**94.25**

Effect of CIR and Mamba

In the ablation experiments on the ACDC dataset, as presented in Table 3, the effects of CIR and Mamba are clearly visible. For CIR, when comparing Model 1 (lacking CIR, Dice coefficient 87.04%) with Model 2 (equipped with CIR, Dice coefficient 87.23%), a small performance boost is noted. In Table 5's erosion visualization analysis, the model with CIR shows a 5.28% increase in segmentation accuracy with a marginal parameter rise. CIR optimizes feature extraction and spatial perception by integrating context through separation and reconfiguration. It also contributes to model lightweighting, as seen in the reduction of parameters from 44.00M to 14.58M and FLOPs from 28.83G to 24.44G when moving from Model 1 to Model 2. Regarding Mamba, comparing Model 1 with Model 3 (featuring Mamba, Dice coefficient 92.32%) reveals a significant accuracy jump. Mamba captures long - range dependencies and fuses multi - scale features, crucial for handling cardiac MRI complexity. It only slightly increases parameters (from 44.00M to 44.37M) and FLOPs (from 28.83G to 29.08G). When CIR and Mamba are combined in our proposed method (Ours, Dice coefficient 94.25%), the synergy of CIR's context optimization and Mamba's long - range dependency handling further enhances performance. Our method also maintains a relatively small parameter count (17.42M) and FLOPs (24.89G) compared to Model 1 and Model 3.

Table 4. Effect of EFCIR-U with/without RSU block instead of traditional U-block. The best scores are highlighted.

Methods	Params(M)	Flops(G)	Dice %
with RSU	17.42	**24.89**	**94.25**
without RSU	**16.16**	30.37	88.12

Effect of RSU

The RSU module in EFCIR - U significantly impacts performance. It innovatively integrates receptive fields for efficient multi - scale feature extraction, combining global and local information. Its internal residual connections ensure seamless information flow, alleviating the gradient vanishing problem and enabling learning of intricate feature representations. The computationally - efficient design allows for increased network depth without a large computational burden. Table 4shows that without the RSU module, there's a 6.13% drop in segmentation accuracy and a 5.48% increase in FLOPs, while the change in parameters is negligible. This highlights the RSU module's crucial role in high - quality segmentation, enhancing performance through its architecture and functionality rather than increased complexity, and making a substantial contribution to segmentation tasks in deep - learning models, especially in medical image analysis like cardiac MRI segmentation.

Reasons for Achieving SOTA Performance while Maintaining a Lightweight Structure

The EFCIR - U architecture achieves SOTA performance while remaining lightweight for several reasons. Firstly, it uses Recurrent RSU modules for efficient multi - scale feature extraction and early fusion strategies to integrate multi - scale information. The CIR module further optimizes feature extraction and spatial perception, handling cardiac anatomical complexity well. Secondly, the Mamba block in the decoder captures long - range dependencies and fuses multi - scale features, which is vital for addressing challenges like low image contrast and partial volume effects in cardiac MRI. Finally, the architecture optimally utilizes parameters. Instead of relying on a large number of parameters, it focuses on enhancing feature extraction efficiency, context integration, and long - range dependency capture. This enables the model to achieve high performance with a relatively small parameter count, resulting in a lightweight structure that still attains SOTA performance.

4 CONCLUSION

In this study, we proposed EFCIR-U, a novel deep-learning architecture specifically designed for high-quality cardiac MRI image segmentation. EFCIR-U integrates theRSU and the CIR module. The RSU module enables efficient multi-scale feature extraction, enhancing the network's ability to capture fine details and complex spatial structures. The CIR module plays a crucial role in combining local and global infor-

mation, reducing computational complexity, and improving segmentation accuracy. Ablation studies conducted on the ACDC dataset further validated the contribution of the CIR module. This module significantly enhanced segmentation performance while reducing model complexity.In conclusion, EFCIR-U provides an efficient and effective solution for cardiac MRI image segmentation, addressing issues such as detail recovery, noise sensitivity, and class imbalance. Future research will focus on applying EFCIR-U to other medical imaging modalities and improving its generalization ability for more complex datasets.

ACKNOWLEDGMENT. This study was funded by the following projects: The Shandong Provincial Major Scientific and Technological Innovation Project (2024CXGC010905), the Shandong Provincial Natural Science Foundation (ZR2022MF279), the Shandong Provincial Technol-ogy Innovation Guidance Program (YDZX2024121 and YDZX2023050), the Major Innova-tion Project of Science-Education-Industry Integration Pilot Program at Qilu University of Technology (Shandong Academy of Sciences) (2024ZDZX08-05), the Shandong Provincial Key R&D Program (Innovation Capacity Enhancement Project for Science and Technology SMEs) (2024TSGC0603)，the Shandong Provincial Innovation Capacity Enhancement Project for Science and Technology SMEs (2023TSGC0201 and 2023TSGC0587).

References

1. Nick Townsend, Denis Kazakiewicz, F Lucy Wright, Adam Timmis, Radu Huculeci, Aleksandra Torbica, Chris P Gale, Stephan Achenbach, Franz Weidinger, and Panos Vardas. Epidemiology of cardiovascular disease in europe. Nature Reviews Cardiology, 19(2):133–143, 2022.
2. Adam Timmis, Panos Vardas, Nick Townsend, Aleksandra Torbica, Hugo Katus, Delphine De Smedt, Chris P Gale, Aldo P Maggioni, Steffen E Petersen, Radu Huculeci, et al. European society of cardiology: cardiovascular disease statistics 2021. European heart journal, 43(8):716–799, 2022.
3. Olaf Ronneberger, Philipp Fischer, and Thomas 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, pages 234–241. Springer, 2015.
4. Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In International conference on medical image computing and computer-assisted intervention, pages 263–273. Springer, 2020.
5. Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In European conference on computer vision, pages 205–218. Springer, 2022.
6. Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306, 2021.

7. Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging, 39(6):1856–1867, 2019.
8. Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP), pages 1055–1059.IEEE, 2020.
9. Hongyi Wang, Shiao Xie, Lanfen Lin, Yutaro Iwamoto, Xian-Hua Han, Yen-Wei Chen, and Ruofeng Tong. Mixed transformer u-net for medical image segmentation. In ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP), pages 2390–2394. IEEE, 2022.
10. Yuxin Wu and Kaiming He. Group normalization. In Proceedings of the European conference on computer vision (ECCV), pages 3–19, 2018.
11. Perry Radau, Yingli Lu, Kim Connelly, Gideon Paul, Alexander J Dick, and Graham A Wright. Evaluation framework for algorithms segmenting short axis cardiac mri. The MIDAS Journal, 2009.
12. Yunhe Gao, Mu Zhou, Di Liu, Zhennan Yan, Shaoting Zhang, and Dimitris N Metaxas. A data-scalable transformer for medical image segmentation: architecture, model efficiency, and benchmark. arXiv preprint arXiv:2203.00131, 2022.
13. Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. Pattern recognition, 106:107404, 2020.
14. Guoping Xu, Xuan Zhang, Xinwei He, and Xinglong Wu. Levit-unet: Make faster encoders with transformer for medical image segmentation. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV), pages 42–53. Springer, 2023.
15. Reza Azad, Yiwei Jia, Ehsan Khodapanah Aghdam, Julien CohenAdad, and Dorit Merhof. Enhancing medical image segmentation with transception: A multi-scale feature fusion approach. arXiv preprint arXiv:2301.10847, 2023.
16. Rahman M M, Marculescu R. G-CASCADE: Efficient cascaded graph convolutional decoding for 2D medical image segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 7728-7737.
17. Rahman M M, Marculescu R. Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation[C]//Medical Imaging with Deep Learning. PMLR, 2024: 1526-1544.
18. Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203-211.
19. C. Petitjean, M. A. Zuluaga, W. Bai, J.-N. Dacher, D. Grosgeorge, J. Caudron, S. Ruan, I. B. Ayed, M. J. Cardoso, H.-C. Chen, D. JimenezCarretero, M. J. Ledesma-Carbayo, C. Davatzikos, J. Doshi, G. Erus, O. M. Maier, C. M. Nambakhsh, Y. Ou, S. Ourselin, C.-W. Peng, N. S. Peters, T. M. Peters, M. Rajchl, D. Rueckert, A. Santos, W. Shi, C.-W. Wang, H. Wang, and J. Yuan. Right ven- tricle segmentation from cardiacmri: A collation study. Medical Image Analysis, 19(1):187 – 202, 2015.