Al-Rafidain Engineering Journal (AREJ) Vol.31, No.2, June 2026, pp. 38-45

Segmentation Improvement of Cardiac Regions in MRI using Hybrid Early-Late Fusion U-Net (HELFU-Net)

Ula T. Salim ula.tariq@uomosul.edu.iq
Shefa A. Dawwd shefa.dawwd@uomosul.edu.iq
Fakhrulddin H. Ali fhazaa@uomosul.edu.iq

Computer Engineering Department, College of Engineering, University of Mosul, Mosul, Iraq

Received: July 22th, 2025 Received in revised form: October 26th, 2025 Accepted: December 14th, 2025

ABSTRACT

Image fusion and the U-Net architecture have been successfully applied to many real applications, in particular, cardiac segmentation. This paper suggests a new version named Hybrid Early-Late Fusion U-Net(HELFU-Net) to segment the cardiac structure into regions. The design has been built by extending the U-Net with five encoder branches and one decoder branch. The encoder branches take advantage of the adjacency property within the cardiac slice-images stack to boost the accuracy of the target image. The first branch of the encoder in the U-Net is to merge adjacent images using the concept of early-stage fusion. The next three branches apply late-stage fusion to the features of adjacent slices that are processed separately. The last branch is for the target slice of the image, while the decoder branch retrieves the data. This design boosts spatial information collection using only 2D image slices. HELFU-Net is evaluated using a public dataset of the ACDC challenge. The experimental results gave mean dice coefficients of 0.942, 0.856, and 0.893 for left ventricular cavity, right ventricular cavity, and left ventricular myocardium, respectively, on the test dataset. Additionally, the suggested HELFU-Net gives 94.9% comparable predicted accuracy on the test dataset over a test time of 1.065 sec.

Keywords:

Hybrid image fusion; MRI; Segmentation; U-Net.

https://rengj.uomosul.edu.iq/
Email: alrafidain_engjournal3@uomosul.edu.iq

==========

1. INTRODUCTION

Cardiac segmentation refers to the strategy used to transform the medical cardiac image into different meaningful regions or sub-images. It classifies each pixel in the medical cardiac image. Segmentation is the first procedure in understanding and interpreting the common cardiovascular disease types.

Magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound are useful medical imaging techniques for producing high-quality cardiac images to support segmentation [1]. Conventionally, cardiac regions are delineated visually by qualified doctors using medical images.

Consistent and accurate segmentation of target cardiac regions from surrounding overlapping tissues utiusing images is critical for the medical assessment of cardiac disease progression. Since manual segmentation is tedious, not scalable for large image datasets, and time-consuming, several automated solutions have been considered and presented so far. Deep learning models have as much imitation in their data processing as the structure of the human brain does in its processing. Cost design is one of the issues affecting the performance of deep learning models, such as the U-Net [2]. Therefore, it takes advantage of advanced computers, such as graphics processing units (GPUs). The U-Net is one of the most commonly used networks for image segmentation, achieving strong predictive performance [3]. Baumgartner et al. [4] and Patravali et al.[5] and Abdelmaguid et al.[6] used the 2D U-Net and explored several U-Net configurations. Yang et al.[7] used a three-dimensional U-Net and Isensee et al.[8] ensembled both two and three-dimensional U-Net. Nevertheless, the U-Net design was developed and extended into many improved variants with distinct ideas [9-15]. Another work uses a multi-atlas mechanism as a basis, as in the model SVF-Net developed by Roh'e et al. [16]. Other works focus on performing as 3D as proposed by Vesal et

al.[17], where the model is built as three three-dimensional DR-U-Net, which worked in an end-to-end and multi-stage manner, which raises the receptive field, besides gathering local information as well as global. Wang et al.[18] suggested a new ICA-UNet. Sun et al.[19] suggested Stack Attention U-Net (SAUN). Other research embeds transformers to make efficient global learning. For example, Cao et al.[20] suggested Swin-Unet, Xu, Guoping, et al.[21] suggested LeViT-Unet and Zhou et al.[22] presented nnFormer which preserves the same shape of the original U-Net. F. Guo et al. embedded Monte Carlo dropout into a 2D U-Net network to segment unseen samples without requiring manual labels [23].

The image fusion concept played an essential role in achieving effective improvement and alleviating the accuracy limitations of processing a single image; it has the advantage of abstracting important details from several images. This leads to increased acquisition of target information and to the removal of redundant pixels from the images. The manner of Image fusion can be done early before image analysis, which contributes to reducing the complexities of the model, but it needs a good classifier since it cannot describe some features along the levels of variant scaling. Therefore, the late fusion of independent features extracted at the last encoder level might be a better solution. It helps improve accuracy and preserves certain features[24-28].

Due to the patient's movement and image quality, the search for new and more efficient methods continues. In this paper, the new and improved contributions are summarised in a hybrid Early-Late Fusion U-Net (HELFU-Net) design. The HELFU-Net is extended to the LVRV-Net [24] and IMFCN [25] methods. Also, it contains a multipath encoder and a single-path decoder. The basic idea is to apply each early-stage and late-stage fusion on the adjacent image slices at the network encoder, while the decoder restores the information with the help of the nearest neighbor method and without employing the deep supervision (DS) at the decoder levels. This method will enrich and maintain the unique features from the high to low levels while minimizing the model cost.

The full paper is organized according to the following structure: Section 1 provides an introduction to the application and the most models that are used. Section 2 describes how the researchers approached it. Section 3 demonstrates the details of the suggested network construction. Section 4 describes the implementation details. Section 5 presents and discusses the most typical results obtained. Finally, Section 6 concludes with the paper's main points and offers suggestions for improving the use of the proposed method.

2. LITERATURE REVIEW

This section focuses on leveraging spatial details overlapping across slices and on the role of fusion ideas in boosting the performance of cardiac segmentation to address memory limitations.

Zheng et al.[24] segmented MRI cardiac slices using iterative 3D consistent networks (LV-net or LVRV-net). Their network encodes an early fusion of the contiguous top-slice and their manual label, then processes them in a single pass across multiple scales. The features of the last encoder level are fused with the corresponding features of the target slice path using late-stage fusion. Then, the data is recovered using the decoder based on deep supervision.

Ma et al.[25] developed an iterative multipath fully convolutional network (IMFCN). It consists of a late-stage fusion of the four separated paths for the encoder and decoder under deep supervision. The experimental results demonstrate comparable dice scores of (0.935)LV, (0.905)MYO, and (0.920)RV.

Luo et al.[26] developed a network inspired by the structure of the U-Net. It consists of context feature extraction and segmentation modules built with different kernel sizes. The Context features extraction module applies a late-fusion approach across three independent pathways dedicated to neighboring slices and one to the target slice. The proposed model produces a mean dice value of 85.02 ± 0.15.

Al Khalil et al. Proposed using a multiple encoder with a U-Net architecture.[27]. The model handles six transformed forms of the input image and passes them through separate paths. At the last level, the final encoding outputs are combined using the late fusion idea. The suggested approach achieves dice values of 0.962(LV), 0.891(MYO), and 0.934(RV) when evaluated employing ACDC data.

Huang et al. presented the MOSformer model, which handles the issue of 2.5D models. The model includes two encoders, the first one is called moment and used for the neighbor image slice, while the second is dedicated to the target image slice. The features of the double encoder at each level are fused independently by utilizing an inter-slice transformer(IF). The reported results show the capabilities of the presented model through realizing a dice score of 92.19%[28]

3. METHODOLOGIES

UNet becomes an accurate and fast architecture for boosting the success of new segmentation models in medical applications. This

section describes the main components of the suggested HELFU-Net architecture. HELFU-Net was used to segment the heart into four regions: background (BG), left ventricular cavity (LVC), left ventricular myocardium (LVM), and right ventricular cavity (RVC). It considers both small and large target MRIs using a pixel-to-pixel and end-to-end learning approach. The suggested HELFU-Net, which fuses the key properties of the UNet basis and the spatial image fusion concept.

Figure 1 describes the steps applied in the cardiac segmentation algorithm:

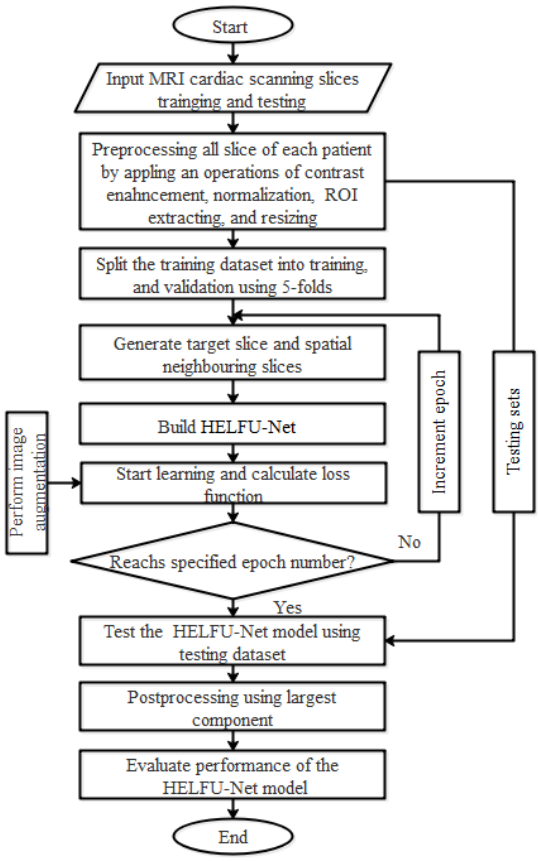


Fig. 1 Cardiac segmentation algorithm

3.1. U-Net

U-Net is a U-shaped version of the convolutional network, which was constructed to segment medical images. Its structure was divided into symmetrical encoder and decoder blocks, combining multilevel local and global data via short- and long-range links. The encoding part includes a set of blocks of two sequence 3x3 unpadded convolutions, rectified linear unit (ReLU), and max pooling. Therefore, the spatial data is reduced by a factor of 2, and the feature count is multiplied by 2. On the contrary, the spatial data is zoomed by a factor of 2, and the feature number is divided by 2 across the decoding levels. This is because the incoming encoding data is passed through a 2x2 up-convolution and concatenated with the shrunk features from the same encoder level, then the resulting features are passed through two symmetric 3x3 convolutions and ReLU. Finally, the data vector is assigned to the specified classes employing a 1x1 convolution with a sigmoid activation function.

3.2. Spatial Image Fusion Concept

The spatial image fusion concept involves gathering features along the channel axis. In the SAX view, a patient's heart is represented as a stack of multiple slices S[i], where each slice corresponds to a different physical ED or ES frame.

The spatial unit exploits the information from the explicit regions of the neighborhood slicers of the target slice S[i]. It consists of two paths, one for the above slice S[i-1] and its mask M[i-1], while the other path is for the below slice S[i+1] as illustrated in Fig. 2. When the first slice of the slice volume is S[i], then both of S[i-1] and M[i-1] will be missing and will be mapped as a null image. In another case, when S[i] is the last slice in that stack, then S[i+1] will not exist and will be taken as a null image. All slices are repeatedly fused depending on the spatial state from base to apical slice according to index (i), which is determined by the level and method of fusion.

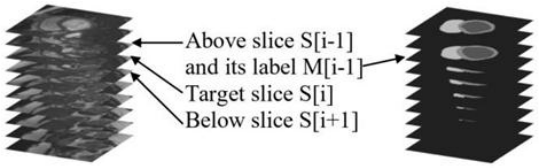


Fig. 2 Spatial approach based on image fusion

3.3. HELFU-Net

HELFU-Net comprises the spatial multipath encoder and decoder modules, as shown in Fig. 3. The spatial multipath encoder module is a new variant built on a hybrid early-late fusion approach. It provides more information by fusing the neighboring slices of the objective segment slice S[i] using four paths. The spatial neighboring paths are concatenated at the encoder input according to the early fusion strategy using the concatenation operator. Also, it avoids the information overlapped issue of the early fusion by processing the information of the slices S[i−1], M[i−1], and S[i+1] using the separated paths. Then, all the extracted features of S[i−1], M[i−1], and S[i+1], the early fusion outcome, and the target slice are fused at the last encoder level by applying a concatenation operator followed by sequence operations of Conv of type 1×1 with stride=1, BN, and ReLU function. The resulting image is restored using the decoder units.

This hybrid strategy boosts the collection of details over several disparate scale levels. However, the architecture of all multipath-encoder and decoder levels is a customized design of the original U-Net architecture[29], but with adding

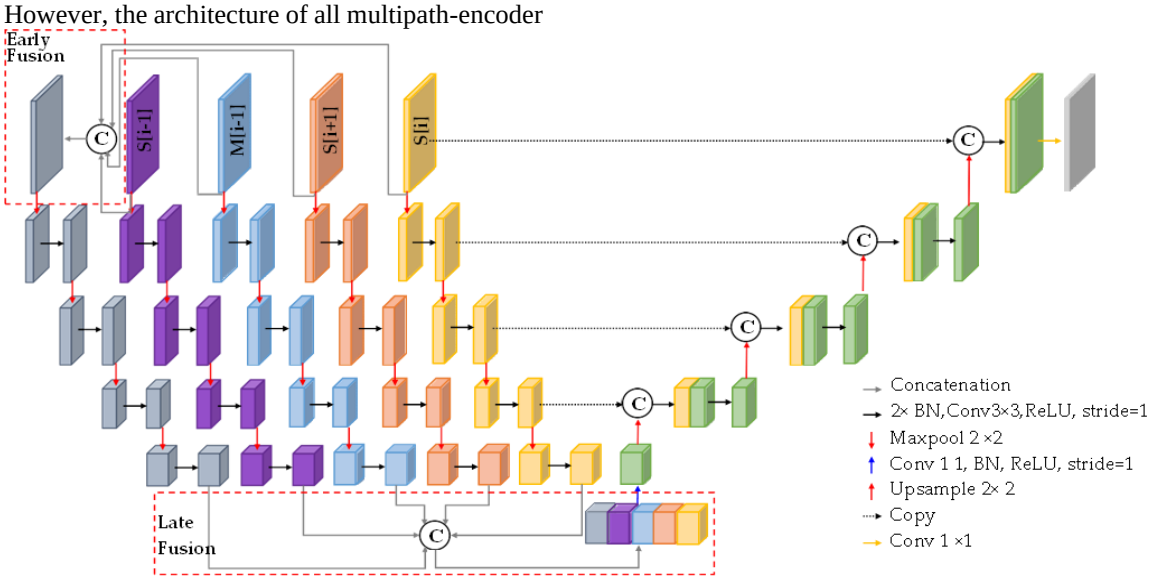


Fig. 3 HELFU-Net architecture represents the output image size at each level. represent the output image size at each level.

Batch normalization (BN) before each convolution operation, as well as using an up-sampling based nearest neighbor algorithm instead of using transpose convolution.

The last layer applies a SoftMax activation function to assign each element of the feature vector to the BG, LVC, LVM, and RVC classes.

4. Experiment Implementation

4.1. Database

In this research, the training and testing data are the popular MRI datasets for cardiac segmentation provided by the Automated Cardiac Diagnosis Challenge (ACDC) [2]. The database contains 100 training patients and 50 testing patients across five classes (normal subjects, abnormal right ventricle, myocardial infarction, dilated cardiomyopathy, and hypertrophic cardiomyopathy). The dataset is a 4D infrastructure, and at each time point, one image frame, either end-diastole (ED) or end-systole (ES), is acquired as a 3D volume with slice thicknesses ranging from 5mm to 10mm and in-plane resolutions of 1.37mm to 1.68mm. The total number of images covering part or all of the cardiac cycle varies from one patient to another, with value ranging from 28 to 40. The reference images and their annotations of LVC, LVM, and RVC at the ES and ED phases are available for the 150 patients within the training database.

4.2. Experimental setup

Python 3.8 and the related packages, including TensorFlow 2.2.0 and Keras 2.2.0, are used to develop the software experiments. On hardware, the used machine is an Intel(R) i7 CPU with 8 cores and 16 GB of RAM, running at 2.9 GHz. The machine is equipped with a GHz RTX 2060 Super, which has 2176 cores and 8 GB of memory. The operating system is Windows-based 64-bit.

The experiments are performed on the ACDC dataset, which is divided into 10% for testing, and the remaining 90% is resampled using a 5-fold cross-validation, with 10% for validation and 80% for training. The data set is processed and artificially augmented in the same way as in reference [24]. To reduce memory usage, a trained ROI network was applied to identify the smallest square overlying the cardiac regions, and the dataset was then rescaled to 128 × 128. The suggested network starts by extracting features with 32 filters. Then, it is trained using the dice loss (DL) [24]. The ADAM optimizer has been adopted for step-wise training with a mini-batch size of 16 and an initial learning rate of 0.0001. The training process ended at the 1000th epoch, when the network's values converged. The testing phase was conducted in an environment similar to the training phase, but with a batch size of 1. Later, the estimated results are sent to a similar post-processing as in [24].

5. Results and Discussion

The success of any algorithm depends on the results obtained. This section discusses the

visual and qualitative results used to evaluate the performance of the suggested network for fusing neighboring image slices to improve segmentation of the target image slice.

5.1. Visual results

To obtain a clear assessment of performance, several experiments were performed on the MRI image. After implementing the prediction and post-processing phases of the suggested network, figures (4) and (5) illustrate the visual outputs for a single patient sample at different slices. In addition, the error (E) in misclassified pixels compared to the reference segmented slice image is presented.

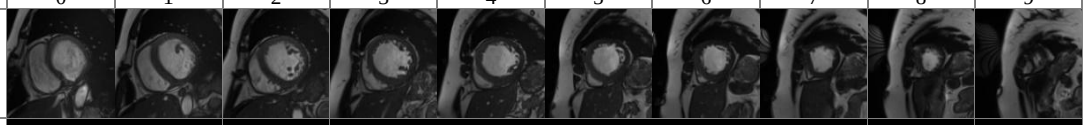
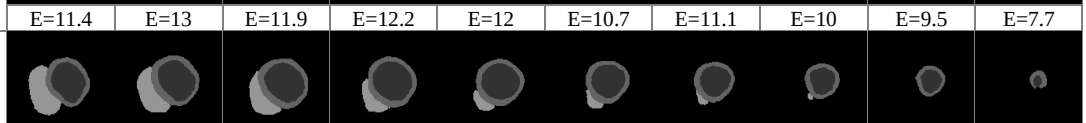
Slice No.	0	1	2	3	4	5	6	7	8	9
Target slice										
Expert ground-truth										
Prediction based on HELFU-Net										
	E=11.4	E=13	E=11.9	E=12.2	E=12	E=10.7	E=11.1	E=10	E=9.5	E=7.7
Prediction based on LVRV-Net[24]										
	E=11.9	E=13.4	E=12.3	E=12.5	E=13	E=11.6	E=8.6	E=9.3	E=7.9	E=8

Fig. 4 Analytical comparisons of cardiac MRI visual results obtained by the proposed networks. All the slices are for one patient of the reserved test set at the ED phase.

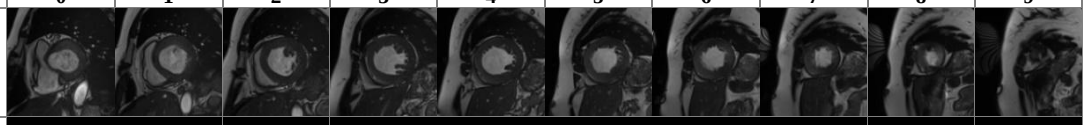
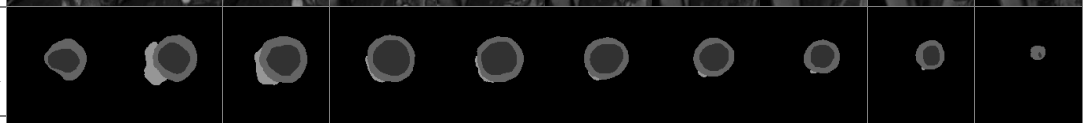
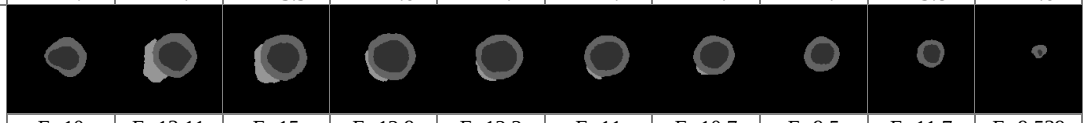
Slice No.	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Target slice										
Expert ground-truth										
Prediction based on HELFU-Net										
	E=11.2	E=12.7	E=13.9	E=12.6	E=11.4	E=12.4	E=12.7	E=11.4	E=9.8	E=7.6
Prediction based on LVRV-Net[24]										
	E=10	E=13.11	E=15	E=12.9	E=12.2	E=11	E=10.7	E=9.5	E=11.7	E=9.539

Fig. 5 Analytical comparisons of cardiac MRI visual results obtained by the proposed networks. All the slices are for one patient of the reserved test set at the ES phase.

In general, across both figures, the suggested network provides a better match with manual masking for LVC than for RVC and MYO in the ED and ES phases. In addition, the predicted visual results showed that HELFU achieved remarkable segmentation accuracy, albeit with some errors, and that the ED phase was more accurate than the ES phase. This is due to overlap between the cardiac infrastructure and surrounding tissues, an imbalance in class pixels, and patient motion. Also, it depends on the patient's case and the location of each cardiac structure, in addition to the type of the device and image slice quality, such as contrast, are affected.

Another notable aspect is that the figures demonstrate that HELFU outperforms LVRV-Net[24] in particular as cardiac regions become smaller. This finding is due to the hybrid early-late fusion at the encoder unit.

5.2. Qualitative metrics

The suggested networks have been trained end-to-end using 90% of the training MRI images, then tested on the remaining 10% of the test images and 50 patients. The suggested network took about 52.4sec and 1.065sec for training and testing, respectively, which is approaching the prediction time of 0.965sec for LVRV-Net[24].

5.2.1. Training metric

The DL has been used to determine appropriate hyperparameters; Fig. 6 demonstrates the DL for the cardiac regions LVC, MYO, and RVC. The results show that HELFU-Net yields good learning performance. It describes the efficiency of the spatial fusion approach for LVC and MYO segmentation, alleviating information loss across level-wise features. This matches the quality noted in the visual predicted results.

5.2.2. Testing metrics

The performance is tested based on a mean successful predicted accuracy of 92.14 on 50 patients, which is considered acceptable compared to 92.414 for LVRV-Net [24]. This small difference is due to the fact that the HELFU-Net has a number of parameters of 25686200 while the LVRV-Net[24] has an estimated number of parameters of 59790988. In addition, the Dice score coefficient (DSC) and the Hasusedroff (HD) metrics are calculated. Compared with the results in the reference IMFCN[25], Table 1 presents comparisons of the LVC, MYO, and RVC structures for the suggested network at both the ED and ES phases. The proposed multipath network can compete with related work and achieve comparable scores at low cost, especially when segmenting the LV region during the ED phase.

On the other hand, in the ES phase, the performance of the proposed network has been affected by reducing the cost, but with a small difference that can be neglected for the segmentation of LVC and MYO structures. This indicates the importance of using the image fusion approach. In summary, the quantitative results match the quality noted in the visually predicted results.

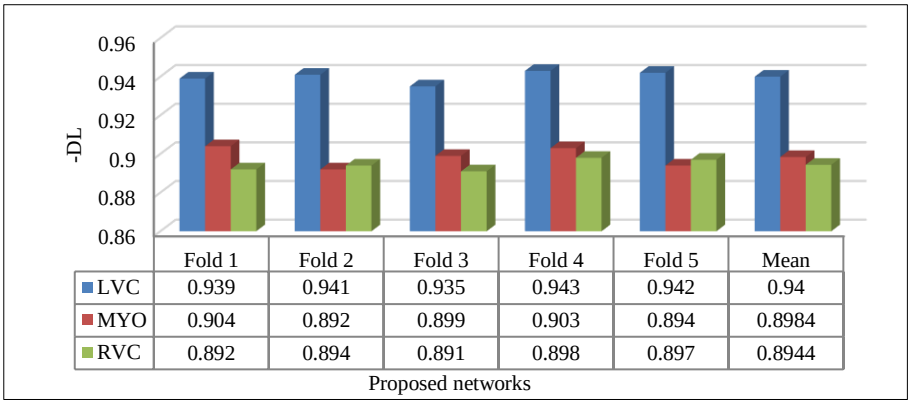


Fig. 6 Dice loss for cardiac regions at five-fold cross-validation and mean score

Table 1: Qualitative comparison of the proposed networks with the previous works in terms of the DSC and HD. The bold values are the best results.

Phase	Method	LV				RV				MYO			
		Dice		HD(mm)		Dice		HD(mm)		Dice		HD(mm)	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
ED	LVRV-Net[24]	0.961	0.025	5.89	3.99	0.930	0.037	13.81	5.44	0.898	0.028	6.87	4.49
	IMFCN[25]	0.963	0.026	5.58	**3.75**	**0.949**	**0.022**	**12.10**	**4.92**	**0.902**	**0.027**	**6.42**	**4.70**
	Our	**0.968**	**0.011**	**5.48**	5.51	0.945	0.023	12.53	5.68	0.886	0.043	7.94	5.39
ES	LVRV-Net[24]	0.924	0.085	7.86	5.68	0.805	0.118	16.62	8.83	0.908	0.035	8.37	4.05
	IMFCN[25]	**0.932**	0.075	**6.92**	4.69	**0.891**	**0.070**	**14.18**	**6.66**	**0.916**	**0.033**	**7.42**	4.06
	Our	0.928	**0.033**	7.13	**2.57**	0.821	0.13	21.47	17.95	0.891	0.073	7.73	**3.42**

6. CONCLUSIONS

In this paper, the HELFU-Net architecture is proposed for computer-aided segmentation of cardiac MRI regions. The suggested architecture provides additional spatial data by applying the fusion propagation flow to 2D images. The HELFU-Net compensates for the rate of accuracy by combining the early fusion and the

late fusion at the U-Net encoder part and then recovering the data at the decoding part. Unlike previous work, it does not apply deep supervision at the decoder levels. The experimentally obtained results demonstrate that most metric values of the suggested architecture outperform those obtained with early-step fusion. In addition, it consumes suitable time during the training and testing phases. Also, the suggested model has fewer parameters.

The main benefit of the current suggested approach is its use of spatial fusion to differentiate regions, alleviating the memory issues that may arise when dealing with 4D data directly. In general, it achieves a suitable cost-to-time ratio and remarkable accuracy in the LVC region, which is considered the most important part in deciding the patient's case. However, the accuracy metrics for both MYO and RV regions need to be enhanced.

In future work, instead of using a hybrid early-late fusion only at the encoder of a U-Net architecture, this method will be implemented by applying one of the early or late fusions at the encoder and the other at the decoder. Also, a multi-scale loss function can capture more details. Furthermore, the design cost will be quantized and used to test other medical applications, such as skin segmentation.

7. REFERENCES

[1] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep Learning for Cardiac Image Segmentation: A Review," *Front. Cardiovasc. Med.*, vol. 7, no. March, 2020, doi: 10.3389/fcvm.2020.00025.

[2] O. Bernard, A. Lalande, C. Zotti, et al, "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?," *IEEE Trans. Med. Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018, doi: 10.1109/TMI.2018.2837502.

[3] U. T. Salim, F. H. Ali, and S. A. Dawwd, "U-Net Convolutional Networks Performance Based on Software-Hardware Cooperation Parameters: A Review," *Int. J. Comput. Digit. Syst.*, vol. 11, no. 1, pp. 977–990, 2022, doi: 10.12785/ijcds/110180.

[4] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation," in *Proc. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, Sep 2017, Québec, Canada. *vol. 10663*, pp. 111–119. https://doi.org/10.1007/978-3-319-75541-0_12.

[5] J. Patravali, S. Jain, and S. Chilamkurthy, "2D-3D Fully Convolutional Neural Networks for Cardiac *MR Segmentation," in Proc. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Sep 2017, Québec, Canada. *vol. 10663*, pp. 130–139. https://doi.org/10.1007/978-3-319-75541-0_14.

[6] E. Abdelmaguid, J. Huang, S. Kenchareddy, D. Singla, L. Wilke, M. H. Nguyen, and I. Altintas, "Left Ventricle Segmentation and Volume Estimation on Cardiac MRI using Deep Learning," Sep. 2018, Accessed: Feb. 28, 2022. [Online]. Available: https://arxiv.org/abs/1809.06247.

[7] X. Yang, C. Bian, L. Yu, D. Ni, and P.-A. Heng, "Class-Balanced Deep Neural Network for Automatic Ventricular Structure Segmentation," *in Proc. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Sep 2017, Québec, Canada. *vol. 10663*, pp. 152–160. https://doi.org/10.1007/978-3-319-75541-0_16.

[8] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features," *in Proc. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Sep 2017, Québec, Canada, *vol. 10663*, pp. 120–129. https://doi.org/10.1007/978-3-319-75541-0_13.

[9] Y. Jang, S. Ha, S. Kim,Y. Hong, and H. J. Chang, "Automatic Segmentation of LV and RV in Cardiac MRI," *in Proc. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Sep 2017, Québec, Canada, *vol. 10663*, pp. 161–169. https://doi.org/10.1007/978-3-319-75541-0_17.

[10] R. Mehta and J. Sivaswamy, "M-net: A Convolutional Neural Network for deep brain structure segmentation," in *Proc. - Int. Symp. Biomed. Imaging*, pp. 437–440, 2017, doi: 10.1109/ISBI.2017.7950555.

[11] M. Khened, V. Alex, and G. Krishnamurthi, "Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest," *in Proc. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Sep 2017, Québec, Canada, *vol. 10663*, pp. 140–151. https://doi.org/10.1007/978-3-319-75541-0_15.

[12] C. Zotti, Z. Luo, O. Humbert, A. Lalande, and P.-M. Jodoin, "GridNet with Automatic Shape Prior Registration for Automatic MRI Cardiac Segmentation," *in Proc. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Sep 2017, Québec, Canada, *vol. 10663*, pp. 73–81. https://doi.org/10.1007/978-3-319-75541-0_8.

[13] S. M. K. Hasan and C. A. Linte, "CondenseUNet: A Memory-Efficient Condensely-Connected Architecture For Bi-ventricular Blood Pool and Myocardium Segmentation," in *Proc.SPIE*, Mar. 2020, p. 113151J. doi: 10.1117/12.2550640.

[14] X. Chen, B. M. Williams, S. R. Vallabhaneni,

G. Czanner, R. Williams, and Y. Zheng, "Learning Active Contour Models for Medical Image Segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11624–11632. doi: 10.1109/CVPR.2019.01190.

[15] G. Snaauw, D. Gong, G. Maicas, A. v. d. Hengel, W. J. Niessen, J. Verjans, and G. Carneiro, "End-To-End Diagnosis And Segmentation Learning From Cardiac Magnetic Resonance Imaging," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 802–805. doi: 10.1109/ISBI.2019.8759276.

[16] M. M. Rohé, M. Sermesant, and X. Pennec, "Automatic Multi-Atlas Segmentation of Myocardium with SVF-Net," in *Proc. STACOM*: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges, Sep 2017, Québec, Canada. *Vol. 10663*, pp.170-177, 10.1007/978-3-319-75541- 0_18.

[17] S. Vesal, A. Maier, and N. Ravikumar, "Fully Automated 3D Cardiac MRI Localisation and Segmentation Using Deep Neural Networks," *J. Imaging*, vol. 6, no. 7, p. 65, 2020, doi: 10.3390/jimaging6070065.

[18] T. Wang, X. Xu, J. Xiong, Q. Jia, H. Yuan, M. Huang, J. Zhuang, and Y. Shi, "ICA-UNET: ICA Inspired Statistical Unet for Real-Time 3D Cardiac Cine MRI Segmentation," in *Proc. MICCAI: Medical Image Computing and Computer Assisted Intervention, Oct 2020, Lima, Peru*, vol 12266. pp. 447–457, https://doi.org/10.1007/978-3-030-59725-2_43

[19] X. Sun, P. Garg, S. Plein, and R. J. van der Geest, "SAUN: Stack attention U-Net for left ventricle segmentation from cardiac cine magnetic resonance imaging," *Med. Phys.*, vol. 48, no. 4, pp. 1750–1763, 2021, doi: 10.1002/mp.14752.

[20] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation," 2021, [Online]. Available: http://arxiv.org/abs/2105.05537

[21] G. Xu, X. Wu, X. Zhang, and X. He, "LeViT-UNet: Make Faster Encoders with Transformer for Medical Image Segmentation," *in Proc. PRCV: Pattern Recognition and Computer Vision, Oct 2023*, Xiamen, China, vol. 14432, pp. 42–53, https://doi.org/10.1007/978-981-99-8543-2_4.

[22] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu, "nnFormer: Interleaved Transformer for Volumetric Segmentation," 2021, [Online]. Available: http://arxiv.org/abs/2109.03201.

[23] F. Guo, M. Ng, I. Roifman, and G. Wright, "Cardiac Magnetic Resonance Left Ventricle Segmentation and Function Evaluation Using a Trained Deep-Learning Model," *Appl. Sci.*, vol. 12, no. 5, 2022, doi: 10.3390/app12052627.

[24] Q. Zheng, H. Delingette, N. Duchateau, and N. Ayache, "3-D Consistent and Robust Segmentation of Cardiac Images by Deep Learning With Spatial Propagation," *IEEE Trans. Med. Imaging*, vol. 37, no. 9, pp. 2137–2148, 2018, doi: 10.1109/TMI.2018.2820742.

[25] Z. Ma, X. Wu, X. Wang, Q. Song, Y. Yin, K. Cao, et al., "An iterative multipath fully convolutional neural network for automatic cardiac segmentation in cine MR images," *Med. Phys.*, vol. 46, no. 12, pp. 5652–5665, 2019, doi: 10.1002/mp.13859.

[26] C. Luo, C. Shi, X. Li, and D. G. Id, "Cardiac MR segmentation based on sequence propagation by deep learning," PLoS One, vol. 15, no. 4, p. e0230415, 2020, [Online]. Available: https://doi.org/10.1371/journal.pone.0230415

[27] Y. Al Khalil, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "Reducing segmentation failures in cardiac MRI via late feature fusion and GAN-based augmentation," *Comput. Biol. Med.*, vol. 161, no. December 2022, p. 106973, 2023, doi: 10.1016/j.compbiomed.2023.106973.

[28] D.-X. Huang, X.-H. Zhou, X.-L. Xie, S.-Q. Liu, Z.-Q. Feng, M.-J. Gui, et al., "MOSformer: Momentum Encoder-based Inter-slice Fusion Transformer for Medical Image Segmentation," 2024, [Online]. Available: http://arxiv.org/abs/2401.11856.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Proc. *MICCAI: Medical Image Computing and Computer-Assisted*, Oct. 2015, Munich, Germany, pp. 234–241. Vol. 9351. https://doi.org/10.1007/978-3-319-24574-4_28.