◻   3214

Comparative evaluation of left ventricle segmentation using improved pyramid scene parsing network in echocardiography

Jin Wang[1,2], Sharifah Aliman[1], Shafaf Ibrahim[1]
[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Malaysia
[2]Department of Electrical Engineering, Taiyuan Institute of Technology, Taiyuan, China

Article Info	**ABSTRACT**
	Automatic segmentation of the left ventricle is a challenging task due to the presence of artifacts and speckle noise in echocardiography. This paper studies the ability of a fully supervised network based on pyramid scene parsing network (PSPNet) to implement echocardiographic left ventricular segmentation. First, the lightweight MobileNetv2 was selected to replace ResNet to adjust the coding structure of the neural network, reduce the computational complexity, and integrate the pyramid scene analysis module to construct the PSPNet; secondly, introduce dilated convolution and feature fusion to propose an improved PSPNet model, and study the impact of pre-training and transfer learning on model segmentation performance; finally, the public data set challenge on endocardial three-dimensional ultrasound segmentation (CETUS) was used to train and test different backbone and initialized PSPNet models. The results demonstrate that the improved PSPNet model has strong segmentation advantages in terms of accuracy and running speed. Compared with the two classic algorithms VGG and Unet, the dice similarity coefficient (DSC) index is increased by an average of 7.6%, Hausdorff distance (HD) is reduced by 2.9%, and the mean intersection over union (mIoU) is improved by 8.8%. Additionally, the running time is greatly shortened, indicating good clinical application potential.

Corresponding Author:

Sharifah Aliman
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
Email: sharifahali@uitm.edu.my

1. INTRODUCTION

Due to the portability and low cost of ultrasound images, accurate segmentation of the left ventricle of echocardiography can help doctors quickly and effectively analyze cardiac function, which is of great significance for clinical real-time monitoring and diagnosis [1]. However, echocardiography itself has the characteristics of low image edge contrast and high speckle noise [2]. In addition, individual physiological differences bring about structural differences. Irregular changes make it a well-known challenge to achieve fully automated real-time segmentation of the left ventricle in echocardiography.

With the development of deep learning network models, advanced technologies based on deep learning are good at discovering complex features in an end-to-end manner by learning continuously directly from the data, injecting new vitality into the automatic segmentation of the left ventricle in echocardiography. The deep learning architectures of two dimensional (2D) single-frame echocardiogram segmentation, including full convolutional neural networks (FCN) [3]–[5] and Unet architectures [6]–[8], have received extensive attention and have proven their effectiveness. In order to improve performance, some

studies have enhanced the segmentation capabilities of the FCN network structure by expanding the training dataset [9], refining the loss function [10], [11], and utilizing pre-training to initialize the model [12], achieving good segmentation results. On the other hand, Unet, tuning Unet architecture, wide Unet, and Unet++ show obvious advantages in accuracy and real-time performance for echocardiography segmentation [13], [14]. To further improve the performance, the researchers improved the Unet network by extracting global and local features of the image and proposed a powerful end-to-end solution, such as the the batch normalized Unet (BNU-Net) model, which employs exponential linear unit (ELU) as activation functions at successive layers in the coding path and batch normalization after the convolutional filters [15]. Combining the advantages of ResNet and Unet, the Res_U model expands the original data from the previous layer to each block of the current layer, thereby propagating and enhancing features throughout the model [16]. Residual dilated Unet (ResDNUnet) model that utilizes Unet, cascade dilated convolution, and residual blocks rich in squeeze-and-excitation operations to extract global and multi-scale features [17]. The pyramid network and Unet were combined to construct multi-feature pyramid Unet (MFP-Unet) [18]. An attention mechanism was introduced into the Unet model to avoid extracting many similar features during parameter calculation [19]. The residual residual of residual-Unet (ROR-Unet) was proposed to solve the vanishing gradient problem and improve segmentation performance [20]. Dense-Unet through data augmentation strategy [21]. The results show that these models outperform Unet in data denoising and provide reliable and stable segmentation results.

In addition to FCN and Unet deep learning network structures, some studies combine traditional segmentation techniques with morphological methods [22], snake models [23], active shape models (ASM) [24], [25] and convolutional neural network (CNN) model are combined to improve the performance of left ventricular segmentation in echocardiography. Some studies have significantly contributed to the expansion of deep learning network architectures by mixing model structures. For instance, SegNet is built using 17 stacked convolutional layers [26], cascaded segmentation and regression network (CSRNet) combines segmentation CNN models with quantized regression networks [27], and the VGGNet model, based on transfer learning [28], demonstrates obvious advantages in segmentation efficiency when using a combined network of an encoder and long short-term memory (LSTM) [29], [30]. Based on an improved and variant model of the Unet structure, feature fusion is performed through skip connections to maximize the utilization of deep semantic information and shallow detail information. This approach shows significant advantages in terms of computational accuracy, sensitivity, and efficiency. However, it also faces some limitations. For example, all semantic advantages of features at different scales during segmentation are ignored, which hinders the learning capabilities of deep learning networks. Consequently, these networks rely on large amounts of annotated data and powerful storage and computing units. Additionally, the large number of model parameters results in extended prediction times.

In view of the inability of FCN and Unet to handle targets of different sizes and multi-scale information well, researchers built a pyramid scene parsing network (PSPNet) network to use different areas to aggregate global context information and learn global image-level features and local multi-scale features at the same time. More and more researchers are applying this kind of network to the field of medical image segmentation, showing certain advantages in pixel-level segmentation and achieving excellent segmentation performance on various data sets [31]–[33]. When applying the PSPNet deep learning network to the prediction of tumor markers, a dice similarity coefficient (Dice) index of 91.3% was achieved, along with a faster processing speed [34], Similarly, when applying the PSPNet model to prostate magnetic resonance imaging (MRI) segmentation, a leading segmentation accuracy of 98.65% was achieved [35]. The PSPNet network based on DenseNet was used for breast cancer image segmentation, achieving a segmentation accuracy of 94.68% higher than the existing method [36]. The semantic segmentation of natural images was achieved based on PSPNet, and the segmentation performance was well verified on public data sets [37]. Some studies have used PSPNet for network fusion and achieved good results for medical image segmentation tasks [38]–[40]. However, feasibility studies of left ventricular segmentation in echocardiography are still scarce.

At this stage, deep learning-based methods are faced with the dual challenges of improving processing accuracy and accelerating processing speed when applied to 2D echocardiographic left ventricle segmentation. This study aims to overcome the limitations of current segmentation models such as FCN and Unet in dealing with segmentation tasks. It also seeks to explore an algorithm that balances network learning depth and performance, and it provides a new solution for real-time segmentation of the left ventricle in 2D echocardiography. This model can effectively extract multi-scale global features from images, fully utilize the position and shape priors of the image, fuse global and local information, and improve segmentation accuracy and speed.

This paper mainly focuses on the following three aspects of research: i) to explore the use of PSPNet network to achieve real-time segmentation of the left ventricle, adjust the backbone feature extraction structure to improve MobileNetv2 instead of ResNet for better performance, and comprehensively

evaluate the segmentation accuracy and efficiency; ii) the effects of two different model initialization methods, pre-training and transfer learning, on segmentation performance, such as algorithmic accuracy and learning efficiency, of the network model in the left ventricular segmentation task were studied; and iii) an exhaustive comparison of segmentation performance was conducted between the improved PSPNet and the optimal models of classical segmentation networks, including VGG [41], Unet [42], and Res_U [16]. Experimental results show that the improved PSPNet significantly outperforms the Res_U network (based on Unet) in feature extraction. Specifically, it enhances the Dice index by 2.3%, reduces model parameters, shortens processing time by 33.4%, and thus boosts real-time performance and accuracy.

2. METHOD

Based on the existing research on echocardiography segmentation methods, this paper proposes a PSPNet network model for segmenting the left ventricle in 2D echocardiography. The overall framework of the algorithm is illustrated in Figure 1: i) in processing the public three-dimensional (3D) challenge on endocardial 3D ultrasound segmentation (CETUS) dataset, 2D slices are obtained by sampling along the short axis, and the images undergo preprocessing to enhance details and reduce noise, without altering the shape of the heart [43]. ii) the processed 2D image is passed to the PSPNet segmentation model, which is a pyramid and deep convolutional network, for automatic feature extraction and to predict the left ventricular segmentation results. iii) evaluate the segmentation results.
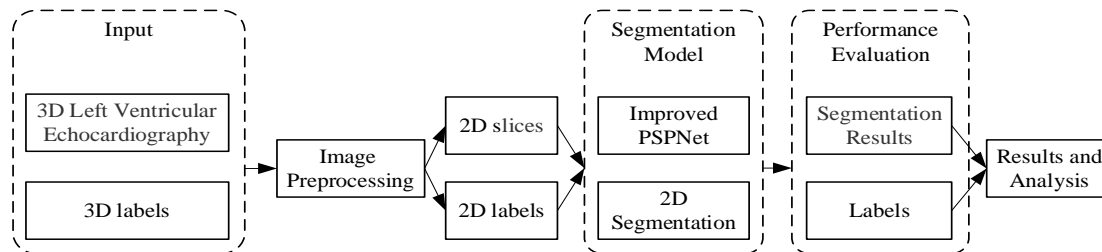

Figure 1. Overall algorithm framework

2.1. Input and image preprocessing

The dataset for this experiment utilizes the public CETUS dataset, which comprises 45 3D echocardiography sequences. These sequences form an echocardiography dataset that is evenly distributed among three different subgroups: healthy subjects, patients with previous muscle damage, and patients with dilated cardiomyopathy. This dataset has been extensively validated in numerous classic and state-of-the-art algorithms [44]. In this experiment, the 45 3D volume data were sliced into 2D images along the short axis. Slices that did not contain any cardiac information were filtered out and deleted. Due to the axial symmetry of the 2D slices, only half of them were selected. Ultimately, 3616 valid 2D slice images were obtained.

2.2. Segmentation models
2.2.1. PSPNet model

The echocardiography input, main structure, and segmentation process of the PSPNet network model are illustrated in Figure 2, which comprises four main parts: feature extraction, pyramid pooling, feature fusion, and deep supervision. First, the feature extraction module of the CNN is used to obtain the feature map of the input image. Then, this feature map is input into the pyramid pooling module (PPM) to obtain a 4-layer pooled feature map. Finally, the pooled feature map is concatenated with the backbone feature map and entered into the FCN module to obtain the predicted segmentation results. The PPM is the core of the PSPNet network. This module aggregates four feature layers of different dimensions, uses 1×1 convolution to reduce the dimensionality, and then performs an upsampling operation to superimpose the restored features with the initial features, thereby forming richer global information and characteristic representations of sub-region information.

In Figure 2, the obtained feature layer is divided into four sub-regions of different dimensions of 1×1, 2×2, 3×3, and 6×6, and then average pooling within the sub-regions is performed. Compared with direct global pooling, which will cause the loss of part of the location information, pyramid pooling is used to take into account the global information and the relationship between each sub-region and realize the aggregation of context information in different regions. The PPM module enables the PSPNet network to fully obtain semantic information at all levels and scales in echocardiography and has strong application potential.
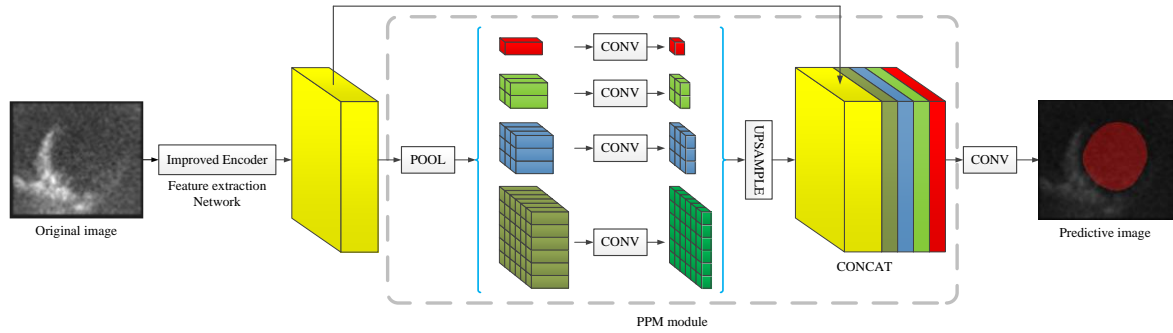
Figure 2. PSPNet segmentation model algorithm structure diagram

2.2.2. Improved PSPNet model

The improved PSPNet network is mainly focused on the feature extraction part, as shown in Figure 2. The lightweight MobileNet is used to replace the ResNet. Atrous convolution and feature fusion are further introduced to achieve the goal of effectively extracting features and shortening the model running time.

2.2.3. Feature extraction network

The traditional PSPNet uses a ResNet-based CNN for backbone feature extraction, which features many layers and a large receptive field, but it has limitations in capturing global information [45]. In contrast, MobileNet employs depthwise separable convolutions, reducing computational complexity and ensuring it is lightweight, swift, and precise [46]. In this paper, ResNet50 and MobileNetv2 are utilized as backbone feature extraction networks for addressing segmentation details and facilitating performance comparisons.

a) ResNet50 as encoder

ResNet, as an encoder, helps improve the accuracy of segmentation networks by retaining spatial information through increased parameters. It features various layers with differing convolutional and batch normalization counts, ResNet34 and ResNet50 being two widely used typical architectures that achieve feature map dimensionality reduction through spatial convolution, as shown in Figure 3 [47]. ResNet50 demonstrates significant advantages in segmentation accuracy and mitigates the issue of vanishing gradients in deep networks. In this paper, ResNet50 is selected as the backbone network for comparative experiments.
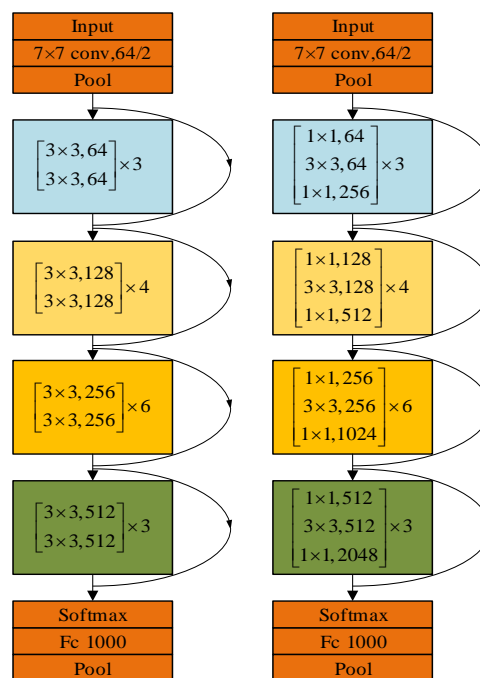

Figure 3. ResNet34 and ResNet50 network structure diagram

b) Improved MobileNetv2 as encoder

The backbone for feature extraction employs the lightweight neural network MobileNet, which minimizes the requirement for network parameters and enhances the network's real-time performance. MobileNet primarily exists in three versions: V1, V2, and V3. To balance algorithm accuracy and speed, this paper adopts MobileNetV2 as the backbone feature extraction network for the encoder. To this end, atrous convolutions are introduced into the Conv6 feature layer, while features from different layers of Conv5 and Conv6 are blended and superimposed. This approach strengthens inter-layer feature fusion without increasing the computational load, expands the receptive field, and reduces the loss of ambiguous information at edges. Figure 4 illustrates the structure of the improved encoder.
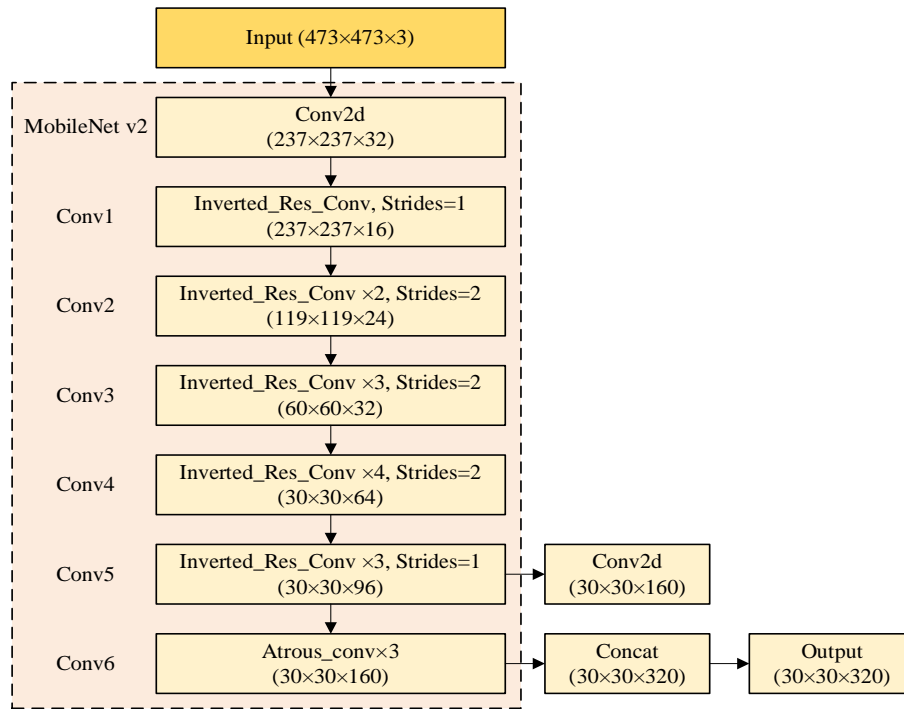

Figure 4. Improved MobileNetv2 encoder structure

To prevent the gradient from vanishing, the MobileNetv2 network is designed with an inverted residual structure. Figure 5 shows the schematic diagram of the inverted residual convolution structure when the stride is 1. The process begins with 1×1 convolution to increase the dimension of the input features, followed by 3×3 spatial convolution to obtain more feature information, and finally concludes with a 1×1 point-wise convolution to reduce the dimension and perform feature compression. This method effectively reduces the number of parameters by first increasing and then decreasing the dimension. In addition, at the end of the inverted residual structure, a linear activation function is used instead of the traditional rectified linear unit 6 (ReLU6) activation layer, which avoids the loss of information caused by the activation function in lower dimensions and improves the performance of the network. Furthermore, the combination of this inverted residual structure with depthwise separable convolutions further enhances computational efficiency and reduces model size, making MobileNetv2 highly suitable for echocardiographic segmentation tasks.

This paper focuses on the short-axis 2D slice of the left ventricle in echocardiograms, which has natural spatial order and rich low-dimensional features. To boost MobileNetv2's feature extraction, atrous convolutions are introduced. These expand the convolution kernels' coverage without increasing complexity, capturing more detailed global information. Considering the size of the input feature map and the design principles of hybrid dilated convolution (HDC), the paper utilizes a series of three atrous convolutions [48]. The structure is shown in Figure 6. The dilation rate (r) represents the spacing between each pixel when the convolution kernel performs convolution calculations, and is set to 2, 3, and 5, respectively. This setup avoids the gridding effect while enhancing information utilization. In addition, combining MobileNetv2 with atrous convolutions facilitates the capture of subtle changes in left ventricular shape and function.
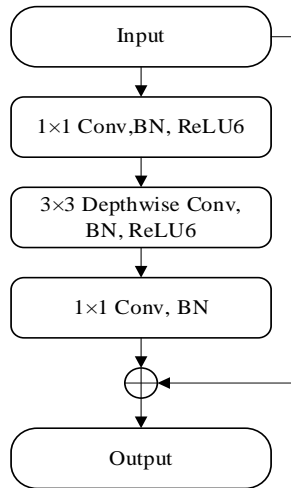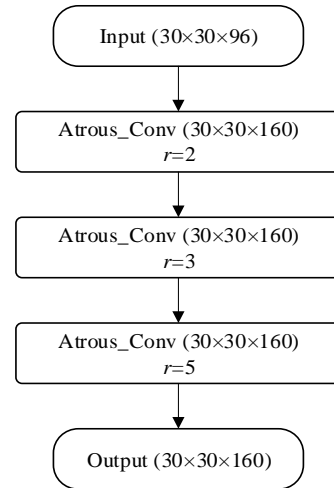
Figure 5. Inverted residual convolution structure

Figure 6. Atrous convolution structure

2.3. Performance evaluation

To measure the accuracy of echocardiographic left ventricular segmentation, we used three different metrics: Dice, Hausdorff distance (HD), and intersection over union (mIoU), as reported in [49]–[51]. These metrics were used to evaluate the segmentation accuracy. Let U={u1,u2,...,um} be the prediction area. Let R={r1,r2,...,rm} be the reference area.

Dice is a measure of the similarity between two sets. It evaluates the similarity between the network prediction structure and the human annotation result. The segmentation task classifies the pixels in the image. Set similarity evaluates the similarity between two contours, generally requires the index to be greater than 0.7, and the segmentation effect is relatively good.

$$\text{Dice} = \frac{2|U \cap R|}{|U| + |R|} \tag{1}$$

HD is the maximum distance from one set to the nearest point in another set. Notably, this distance is directional; specifically, h (U, R) is not equal to h (R, U). H takes the larger of the two distances. For parameters that are sensitive to differences in location information, the smaller the value, the higher the degree of repetition. The calculation formula is as follows: HD=max[h (U, R),h (R, U)], set U={u_1,u_2,..., u_m}, R={r_1,r_2,..., r_m}, where h used to calculate the one-way Hough distance between two surfaces.

$$h(\text{R,U}) = \max_{u \in U} \left\{ \min_{r \in R} \|\text{u-r}\| \right\} \tag{2}$$

$$h(\text{U,R}) = \max_{r \in R} \left\{ \min_{u \in U} \|\text{r-u}\| \right\} \tag{3}$$

mIoU is the average of the intersection and union ratios across two categories: heart area and background area. Intersection over union (IoU) is used to measure the overlapping area of each category, IoU=intersection area of a certain category/union area of a certain category. mIoU is then computed as the sum of the IoUs of all categories divided by the number of categories.

$$\text{mIoU} = \frac{1}{2} \times \left(\frac{n_{\text{ff}}}{t_f + n_{\text{bf}}} + \frac{n_{\text{bb}}}{t_b + n_{\text{fb}}} \right) \tag{4}$$

Among them, n_{ff} represents the number of correctly classified foreground pixels, t_f represents the number of pixels belonging to the foreground, n_{bf} represents the number of incorrectly classified background pixels, n_{bb} represents the number of correctly classified background pixels, t_b represents the number of pixels belonging to the background, and the number of n_{fb} represents the number of misclassified foreground pixels.

3.   RESULTS AND DISCUSSION

This study replaced different backbone feature networks based on the traditional PSPNet algorithm and selected the convolution-based ResNet50 backbone network and the improved lightweight neural network MobileNetv2 to train the segmentation model. Two methods were used to initialize the weights of

the model: left ventricular echocardiography pre-training and natural image transfer. The impact of the two initialization methods of pre-training and transfer learning on model performance indicators was comparatively analyzed through experiments.

The experiment is conducted on Kaggle using the PyTorch deep learning framework. The initial learning rate is 0.01, and the minimum learning rate is 0.0001. The optimizer employs stochastic gradient descent (SGD) with a momentum parameter of 0.9. The batch size is set to 8, and a cosine learning rate strategy is selected. The weight decay is configured as 0.0001. All these parameter settings are chosen based on tracking the model training process to enhance segmentation performance. Model segmentation performance evaluation is done through analysis of the Epoch_loss curve and Epoch_Miou for both the training set and validation set.

During the model training process, the pre-trained model was first utilized to initialize the weights of the backbone feature extraction network of the segmentation model. The images in the data set are stored in VOC format and all images are resized uniformly to 473×473. The data set is divided into a training set and a validation set according to the ratio of 9:1. When ResNet50 is selected as the backbone network, the epoch=30 and epoch=50 model training results (Epoch_loss and Epoch_Miou) are shown in Figures 7(a) and 7(b). When the backbone network uses the improved MobileNetv2, the epoch=30 and epoch=50 model training results (Epoch_loss and Epoch_Miou) are shown in Figures 8(a) and 8(b). The resultant diagram shows that when MobileNetV2, which incorporates dilated convolution and feature fusion, is used as the backbone feature extraction network, the segmentation model achieves convergence within 30 epochs. This is equivalent to the training effect of 50 epochs when ResNet50 is used as the backbone network. Additionally, a higher mean Miou value is achieved within 5 epochs. These results demonstrate significant performance advantages and faster convergence capabilities.
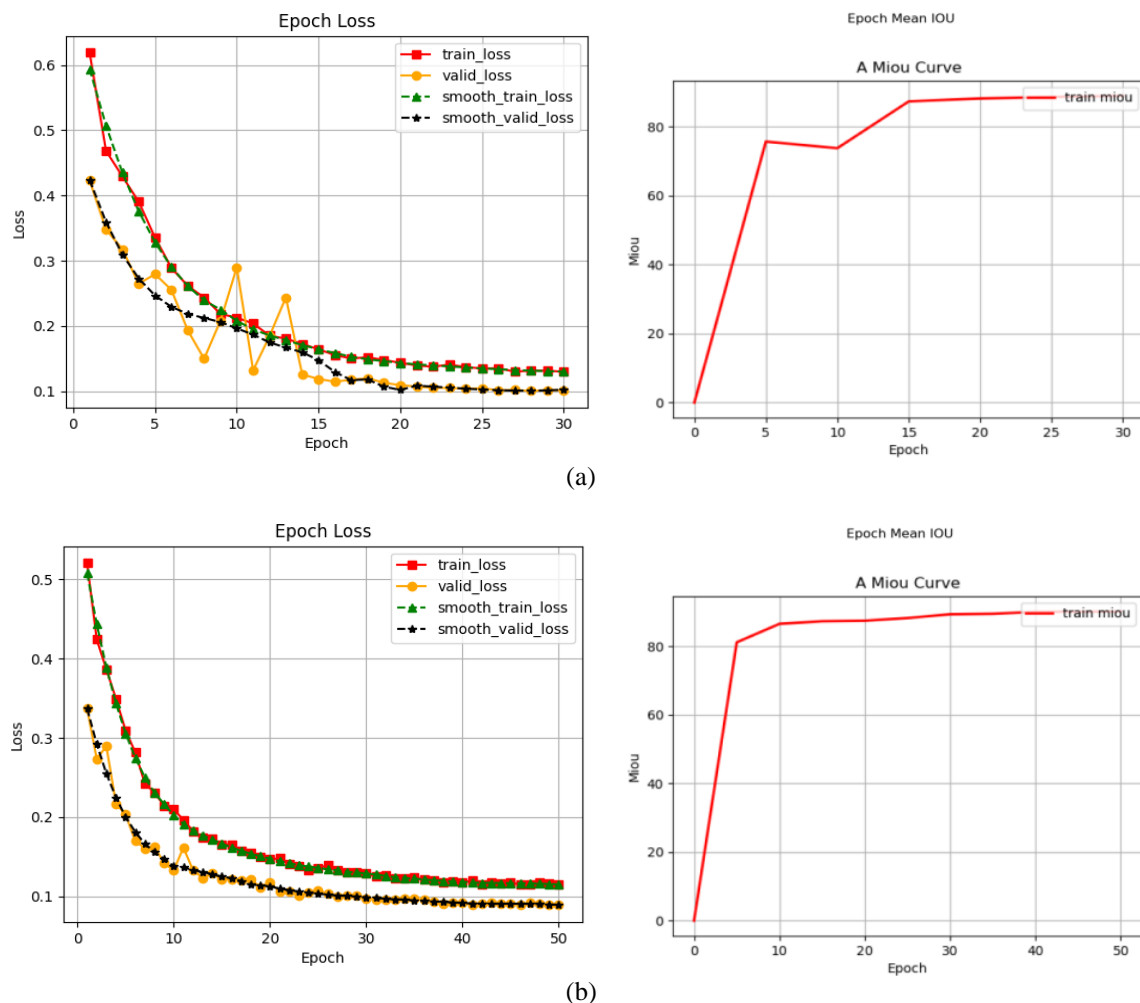

(a)

(b)

Figure 7. ResNet50 backbone network for (a) Epoch=30 and (b) Epoch=50
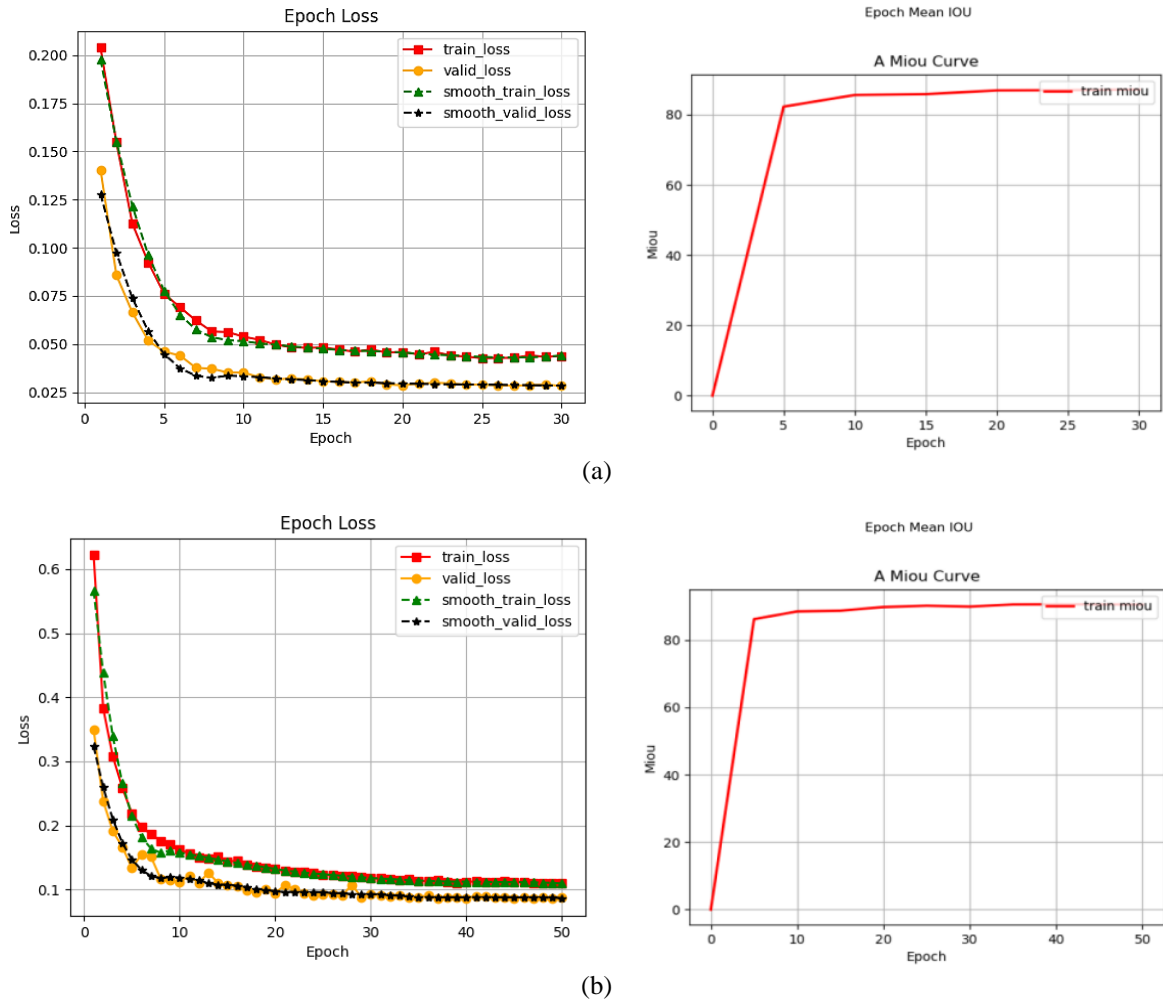
(a)

(b)

Figure 8. MobileNetv2 backbone network for (a) Epoch=30 and (b) Epoch=50

To further explore the specific impact of two different weight initialization schemes on the performance of the segmentation model, the initialization method of transfer learning was then used. Specifically, the model weights obtained by pre-training on a wide range of natural image datasets were used to conduct fine-weight initialization of the backbone feature extraction network in the segmentation model. The above model training experiments were continued to be repeated using both ResNet50 and the improved MobileNetv2 backbone networks, and when the backbone network selects ResNet50 with epochs set to 30, the random initialization and transfer initialization model training results (Epoch_loss and Epoch_Miou) are shown in Figures 9(a) and 9(b). When the backbone network uses the improved MobileNetv2 with epochs set to 30, the random initialization and transfer initialization model training results (Epoch_loss and Epoch_Miou) are shown in Figures 10(a) and 10(b).

It can be observed from the training results of segmentation models initialized with different weights that the PSPNet segmentation model built using ResNet50 as the backbone network has significantly improved model performance, including convergence speed and segmentation accuracy, supported by two initialization schemes: pre-training and transfer learning. However, when the improved MobileNetv2 is used as the backbone network, the initialization method has little effect on the performance and convergence speed of the segmentation model. Even with random initialization, it outperforms the ResNet50-based PSPNet, achieving similar performance with 30 epochs compared to 50 for ResNet50. Due to its lightweight design, the improved MobileNetv2-based PSPNet reduces overall running time by 40% compared to the ResNet50-based model.

To verify the segmentation model's superiority, Table 1 summarizes the performance of VGG [41], Unet [42], Res_U [16], and improved PSPNet in left ventricular echocardiography segmentation, focusing on segmentation accuracy (assessed by Dice, HD, mIoU), processing time (a real-time performance measure), and segmentation effect (displayed visually). Notations 'T' and 'Q' denote pre-training and transfer learning

initializations, respectively. PSPNet_R, PSPNet_M, and PSPNet_MK use ResNet, MobileNetv2, and improved MobileNetv2 as backbones.
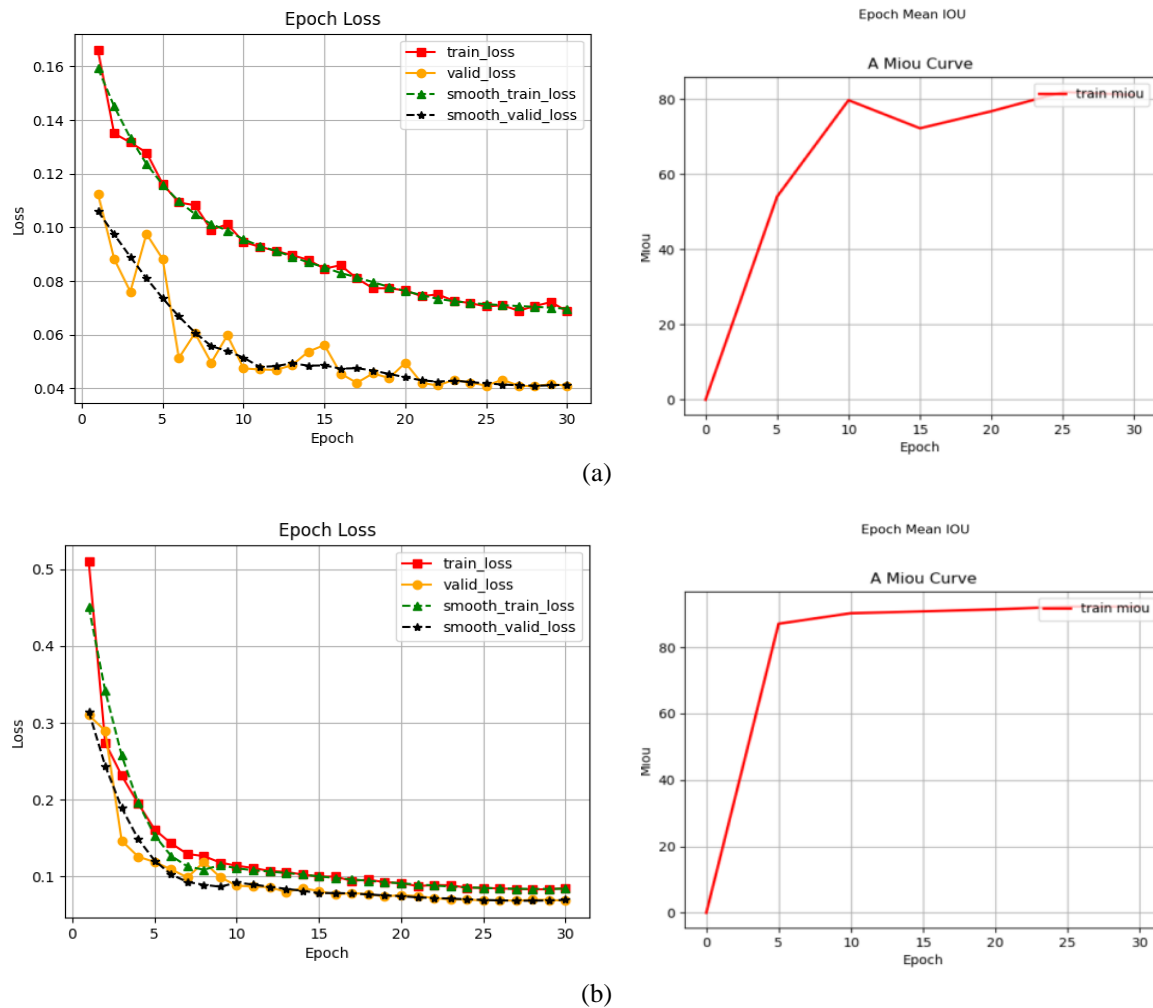

(a)

(b)

Figure 9. ResNet50 backbone network for (a) random initialization and (b) transfer initialization

The Res_U algorithm combines ResNet and Unet to optimize the network structure of the feature extraction part and capture more effective features that are beneficial to segmentation. Compared with the classic algorithms Unet and VGG, the segmentation effect is better, with Dice reaching 83% and mIoU reaching 84%. The paper studies the PSPNet_MK algorithm, which optimizes the encoding part of PSPNet's MobileNetv2 by incorporating a lightweight network model. Additionally, it introduces atrous convolution to integrate contextual information, thereby acquiring richer global information and achieving the highest Dice and mIoU values. Res_U offers a slight accuracy edge but the fastest prediction, balancing segmentation accuracy and efficiency. Notably, Pre-training and transfer learning enhance PSPNet_R's performance, with Dice increasing by 3.7%, HD decreasing by 0.6%, and mIoU increasing by 3.6%. When initialized with the same weights, MobileNetv2 runs significantly faster but achieves lower Dice (-2.5%) and mIoU (-1.2%) compared to ResNet50. By enhancing MobileNetv2 with atrous convolution, feature fusion, and an improved backbone, segmentation performance improved over ResNet with Dice +3.7%, mIoU +3.6%, and HD -0.6%. This achieved optimal segmentation without performance loss and reduced processing time by 33.4%.

Real-time echocardiographic left ventricular segmentation is extremely challenging due to artifacts and speckle noise in images. Figure 11 provides a qualitative visual comparison of different image qualities, their corresponding segmentation masks, and prediction results. The Res_U model and the improved PSPNet model with MobileNetV2 as the backbone network achieve satisfactory segmentation results in the left ventricular region of echocardiography. It closely matches the segmentation mask in boundary accuracy and area overlap, especially outperforming classic VGG and Unet models in images with artifacts, noise, and

blurred anatomical boundaries. Improved PSPNet, studied in this paper, runs fastest, while VGG is slowest. Compared to Res_U, PSPNet shortens calculation time by 33.4% for the same results, balancing segmentation performance and speed, with good clinical potential.
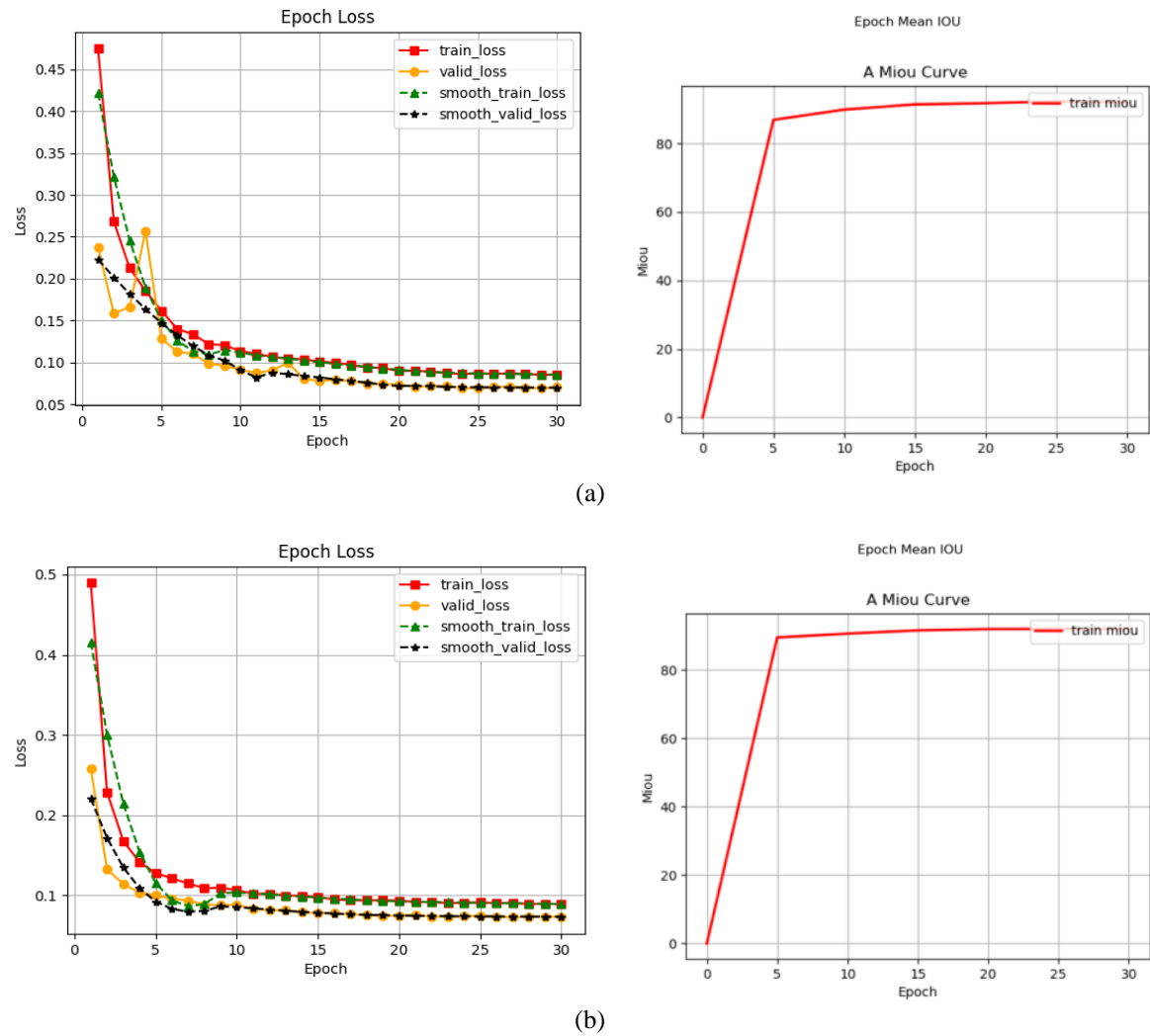

(a)

(b)

Figure 10. MobileNetv2 backbone network for (a) random initialization and (b) transfer initialization

Table 1. Comparison of segmentation results under different configurations of PSPNet segmentation network

| Segmentation architecture		Dice	HD	mIoU	Processing time(s)
VGG	VGG	0.77	4.65	0.76	34.6
	VGG_T	0.79	4.61	0.79	33.2
	VGG_Q	0.79	4.63	0.78	33.7
Unet	Unet	0.78	4.69	0.76	27.4
	Unet_T	0.80	4.65	0.80	26.5
	Unet_Q	0.79	4.67	0.78	26.8
Res_U	Res_U	0.81	4.62	0.82	30.2
	Res_U_T	0.83	4.59	0.84	28.6
	Res_U_Q	0.82	4.60	0.83	29.4
PSPNet	PSPNet_R	0.79	4.61	0.82	27.8
	PSPNet_R_T	0.82	4.59	0.85	28.5
	PSPNet_R_Q	0.81	4.58	0.84	29.1
	PSPNet_M	0.77	4.55	0.81	19.6
	PSPNet_M_T	0.80	4.60	0.82	19.8
	PSPNet_M_Q	0.81	4.59	0.82	21.3
	PSPNet_MK	0.82	4.52	0.85	18.5
	PSPNet_MK_T	0.84	4.51	0.86	19.4
	PSPNet_MK_Q	0.83	4.50	0.85	20.2
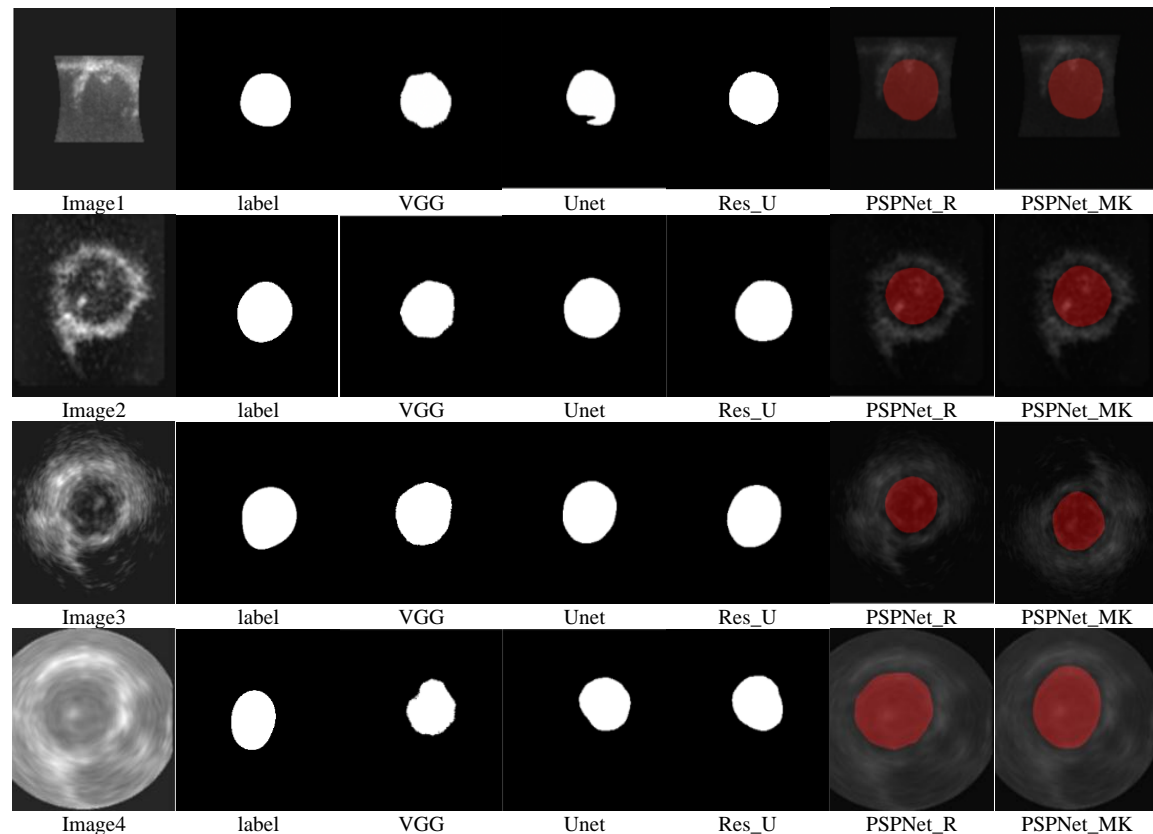
Figure 11. Echocardiographic left ventricular segmentation results

4. CONCLUSION

This study is dedicated to solving the computational efficiency bottleneck problem caused by the pursuit of high accuracy in echocardiographic left ventricular segmentation algorithms. MobileNetv2 and atrous convolution are used to improve the PSPNet model, enhance global and local information acquisition, and optimize network performance and size for left ventricular segmentation in echocardiography. Compared with the classic segmentation algorithms FCN and Unet for echocardiography, the PSPNet_MK segmentation algorithm shows significant improvements in Dice coefficient and mIoU on the test set, is more accurate for speckle noise images, performs well in blurred boundaries and low-contrast scenes, and increases processing speed by 40%. Future research will expand the PSPNet network to the field of 3D echocardiography segmentation and optimize segmentation accuracy and efficiency, At the same time, image correlation is used to implement pre-training and transfer learning, and the impact of initialization methods on model segmentation capabilities is deeply explored. Despite the limitations faced when processing low-quality images, the initialization strategy still needs to be optimized to enhance the robustness and generalization ability of the PSPNet algorithm and adapt to echocardiograms of different qualities and scenarios. Despite its limitations in processing low-quality images, optimization of the initialization strategy is still necessary to improve the robustness and generalization ability of the PSPNet algorithm and ensure that it can adapt to echocardiographic data of different qualities and scenarios.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Jin Wang	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	
Sharifah Aliman	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓
Shafaf Ibrahim	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	

C	:	**C**onceptualization	I	:	**I**nvestigation	Vi	:	**Vi**sualization
M	:	**M**ethodology	R	:	**R**esources	Su	:	**Su**pervision
So	:	**So**ftware	D	:	**D**ata Curation	P	:	**P**roject administration
Va	:	**Va**lidation	O	:	Writing - **O**riginal Draft	Fu	:	**Fu**nding acquisition
Fo	:	**Fo**rmal analysis	E	:	Writing - Review & **E**diting			

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] V. A. Chenarlogh *et al.*, "Fast and accurate U-Net model for fetal ultrasound image segmentation," *Ultrasonic Imaging*, vol. 44, no. 1, pp. 25–38, Jan. 2022, doi: 10.1177/01617346211069882.

[2] J. Zhang *et al.*, "Fully automated echocardiogram interpretation in clinical practice," *Circulation*, vol. 138, no. 16, pp. 1623–1635, 2018, doi: 10.1161/CIRCULATIONAHA.118.034338.

[3] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017, doi: 10.1109/TPAMI.2016.2572683.

[4] N. Savioli, M. S. Vieira, P. Lamata, and G. Montana, "Automated segmentation on the entire cardiac cycle using a deep learning work - flow," in *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, IEEE, 2018, pp. 153–158, doi: 10.1109/SNAMS.2018.8554962.

[5] H. Abdeltawab *et al.*, "A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine MR images," *Computerized Medical Imaging and Graphics*, vol. 81, 2020, doi: 10.1016/j.compmedimag.2020.101717.

[6] P. F. Ferreira *et al.*, "Automating in vivo cardiac diffusion tensor postprocessing with deep learning–based segmentation," *Magnetic Resonance in Medicine*, vol. 84, no. 5, pp. 2801–2814, 2020, doi: 10.1002/mrm.28294.

[7] X. Zou, Q. Wang, and T. Luo, "A novel approach for left ventricle segmentation in tagged MRI," *Computers and Electrical Engineering*, vol. 95, 2021, doi: 10.1016/j.compeleceng.2021.107416.

[8] T. Wech, M. J. Ankenbrand, T. A. Bley, and J. F. Heidenreich, "A data-driven semantic segmentation model for direct cardiac functional analysis based on undersampled radial MR cine series," *Magnetic Resonance in Medicine*, vol. 87, no. 2, pp. 972–983, 2022, doi: 10.1002/mrm.29017.

[9] W. Bai *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, 2018, doi: 10.1186/s12968-018-0471-x.

[10] Z. Ma *et al.*, "An iterative multi-path fully convolutional neural network for automatic cardiac segmentation in cine MR images," *Medical Physics*, vol. 46, no. 12, pp. 5652–5665, 2019, doi: 10.1002/mp.13859.

[11] M. Chen, L. Fang, and H. Liu, "FR-NET: Focal loss constrained deep residual networks for segmentation of cardiac MRI," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 764–767, doi: 10.1109/ISBI.2019.8759556.

[12] H. J. Koo *et al.*, "Automated segmentation of left ventricular myocardium on cardiac computed tomography using deep learning," *Korean Journal of Radiology*, vol. 21, no. 6, pp. 660–669, 2020, doi: 10.3348/kjr.2019.0378.

[13] V. Zyuzin and T. Chumarnaya, "Comparison of Unet architectures for segmentation of the left ventricle endocardial border on two-dimensional ultrasound images," in *2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, IEEE, 2019, pp. 110–113, doi: 10.1109/USBEREIT.2019.8736616.

[14] S. Hossain *et al.*, "Automated breast tumor ultrasound image segmentation with hybrid UNet and classification using fine-tuned CNN model," *Heliyon*, vol. 9, no. 11, p. e21369, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21369.

[15] W. Chu *et al.*, "BNU-Net: a novel deep learning approach for LV MRI analysis in short-axis MRI," in *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, IEEE, 2019, pp. 731–736, doi: 10.1109/BIBE.2019.00137.

[16] Y. Ali, F. Janabi-Sharifi, and S. Beheshti, "Echocardiographic image segmentation using deep Res-U network," *Biomedical Signal Processing and Control*, vol. 64, 2021, doi: 10.1016/j.bspc.2020.102248.

[17] A. Amer, X. Ye, M. Zolgharni, and F. Janan, "ResDUnet: Residual dilated UNet for left ventricle segmentation from echocardiographic images," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020, pp. 2019–2022, doi: 10.1109/EMBC44109.2020.9175436.

[18] S. Moradi *et al.*, "MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography," *Physica Medica*, vol. 67, pp. 58–69, 2019, doi: 10.1016/j.ejmp.2019.10.001.

[19] H. Cui, C. Yuwen, L. Jiang, Y. Xia, and Y. Zhang, "Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images," *Computer Methods and Programs in Biomedicine*, vol. 206, 2021, doi: 10.1016/j.cmpb.2021.106142.

[20] M. Zarvani, S. Saberi, R. Azmi, and S. V. Shojaedini, "Residual learning: A new paradigm to improve deep learning-based segmentation of the left ventricle in magnetic resonance imaging cardiac images," *Journal of Medical Signals and Sensors*, vol. 11, no. 3, pp. 159–168, 2021, doi: 10.4103/jmss.JMSS_38_20.

[21] C. Wang, Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, "Dense U-net based on patch-based learning for retinal vessel segmentation," *Entropy*, vol. 21, no. 2, 2019, doi: 10.3390/e21020168.

[22] X. Li, Y. Wang, W. Yan, R. J. Van der Geest, Z. Li, and Q. Tao, "A multi-scope convolutional neural network for automatic left ventricle segmentation from magnetic resonance images: deep-learning at multiple scopes," in *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, IEEE, 2018, pp. 1–5, doi: 10.1109/CISP-BMEI.2018.8633185.

[23] Y. Lan and R. Jin, "Automatic segmentation of the left ventricle from cardiac MRI using deep learning and double snake model," *IEEE Access*, vol. 7, pp. 128641–128650, 2019, doi: 10.1109/ACCESS.2019.2939542.

[24] W.-Y. Hsu, "Automatic left ventricle recognition, segmentation and tracking in cardiac ultrasound image sequences," *IEEE Access*, vol. 7, pp. 140524–140533, 2019, doi: 10.1109/ACCESS.2019.2920957.

[25] M. Salvi, K. M. Meiburger, and F. Molinari, "Softmax-driven active shape model for segmenting crowded objects in digital pathology images," *IEEE Access*, vol. 12, no. 2, pp. 30824–30838, 2024, doi: 10.1109/ACCESS.2024.3369916.

[26] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.

[27] W. Wang, Y. Wang, Y. Wu, T. Lin, S. Li, and B. Chen, "Quantification of full left ventricular metrics via deep regression learning with contour-guidance," *IEEE Access*, vol. 7, pp. 47918–47928, 2019, doi: 10.1109/ACCESS.2019.2907564.

[28] J. Shedbalkar and K. Prabhushetty, "Deep transfer learning model for brain tumor segmentation and classification using UNet and chopped VGGNet," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 3, pp. 1405–1415, 2024, doi: 10.11591/ijeecs.v33.i3.pp1405-1415.

[29] X. Du, S. Yin, R. Tang, Y. Zhang, and S. Li, "Cardiac-DeepIED: automatic pixel-level deep segmentation for cardiac bi-ventricle using improved end-to-end encoder-decoder network," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019, doi: 10.1109/JTEHM.2019.2900628.

[30] L. Zhang *et al.*, "Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1114–1126, 2020, doi: 10.1109/TMI.2019.2943841.

[31] M. Y. Ansari, Y. Yang, P. K. Meher, and S. P. Dakua, "Dense-PSP-UNet: A neural network for fast inference liver ultrasound segmentation," *Computers in Biology and Medicine*, vol. 153, 2023, doi: 10.1016/j.compbiomed.2022.106478.

[32] P. P. Jena, D. Mishra, K. Das, S. Mishra, and M. P. Behera, "PSPNet EPO-SEB: a novel attention-enhanced hybrid model for accurate histopathological image segmentation," *Connection Science*, vol. 37, no. 1, pp. 268–275, Dec. 2025, doi: 10.1080/09540091.2025.2508357.

[33] N. Sharma, S. Gupta, A. Rajab, M. A. Elmagzoub, K. Rajab, and A. Shaikh, "Semantic segmentation of gastrointestinal tract in MRI scans using PSPNet model with ResNet34 feature encoding network," *IEEE Access*, vol. 11, no. 4, pp. 132532–132543, 2023, doi: 10.1109/ACCESS.2023.3336862.

[34] L. Y. Ye, X. Y. Miao, W. S. Cai, and W. J. Xu, "Medical image diagnosis of prostate tumor based on PSP-Net+VGG16 deep learning network," *Computer Methods and Programs in Biomedicine*, vol. 221, 2022, doi: 10.1016/j.cmpb.2022.106770.

[35] L. Yan *et al.*, "PSP net-based automatic segmentation network model for prostate magnetic resonance imaging," *Computer Methods and Programs in Biomedicine*, vol. 207, 2021, doi: 10.1016/j.cmpb.2021.106211.

[36] S. Samudrala and C. K. Mohan, "Semantic segmentation of breast cancer images using DenseNet with proposed PSPNet," *Multimedia Tools and Applications*, vol. 83, no. 15, pp. 46037–46063, 2023, doi: 10.1007/s11042-023-17411-5.

[37] H. Kim, Y. J. Kim, J. Hong, H. Yang, S. Euna, and K. Lee, "Local context aggregation for semantic segmentation: a novel PSPNet approach," in *2024 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC)*, IEEE, 2024, pp. 1–5, doi: 10.1109/ITC-CSCC62988.2024.10628341.

[38] S. Wang *et al.*, "DPAM-PSPNet: ultrasonic image segmentation of thyroid nodule based on dual-path attention mechanism," *Physics in Medicine and Biology*, vol. 68, no. 16, 2023, doi: 10.1088/1361-6560/ace6f1.

[39] X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, "Coronary angiography image segmentation based on PSPNet," *Computer Methods and Programs in Biomedicine*, vol. 200, Mar. 2021, doi: 10.1016/j.cmpb.2020.105897.

[40] P. Wen *et al.*, "A-PSPNet: A novel segmentation method of renal ultrasound image," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2021, pp. 40–45, doi: 10.1109/SMC52423.2021.9658740.

[41] A. Dozen *et al.*, "Image segmentation of the ventricular septum in fetal cardiac ultrasound videos based on deep learning using time-series information," *Biomolecules*, vol. 10, no. 11, p. 1526, Nov. 2020, doi: 10.3390/biom10111526.

[42] V. Zyuzin *et al.*, "Identification of the left ventricle endocardial border on two-dimensional ultrasound images using the convolutional neural network Unet," in *2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, IEEE, 2018, pp. 76–78, doi: 10.1109/USBEREIT.2018.8384554.

[43] H. H. Xue, B. J. Chen, K. Sun, and J. G. Yu, "Methodological study on filtering and segmentation of three-dimensional echocardiographic images by virtual endoscopy," *Journal of Shanghai Jiaotong University (Medical Science)*, vol. 31, no. 2, pp. 169–172, 2011, doi: 10.3969/jssn.1674-8115.2011.02.011.

[44] S. Leclerc *et al.*, "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography," *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2198–2210, 2019, doi: 10.1109/TMI.2019.2900516.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[46] B. Singh, D. Toshniwal, and S. K. Allur, "Shunt connection: An intelligent skipping of contiguous blocks for optimizing MobileNet-V2," *Neural Networks*, vol. 118, pp. 192–203, 2019, doi: 10.1016/j.neunet.2019.06.006.

[47] Y. Liu, W. Gao, P. He, J. Tang, and L. Hu, "Apple phenological period identification in natural environment based on improved ResNet50 model," *Smart Agriculture*, vol. 5, no. 2, pp. 13–22, 2023, doi: 10.12133/j.smartag.SA202304009.

[48] P. Wang *et al.*, "Understanding convolution for semantic segmentation," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2018, pp. 1451–1460, doi: 10.1109/WACV.2018.00163.

[49] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, p. 210, Dec. 2022, doi: 10.1186/s13104-022-06096-y.

[50] H. Kim *et al.*, "Quantitative evaluation of image segmentation incorporating medical consideration functions," *Medical Physics*, vol. 42, no. 6, pp. 3013–3023, 2015, doi: 10.1118/1.4921067.

[51] M. Wang, H. Jiang, T. Shi, and Y. Yao, "HD-RDS-UNet: Leveraging spatial-temporal correlation between the decoder feature maps for lymphoma segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 1116–1127, 2022, doi: 10.1109/JBHI.2021.3102612.

BIOGRAPHIES OF AUTHORS

Jin Wang received a bachelor's degree in Electronic Information and Science Technology from Shanxi Normal University in 2007 and the master's degree in Communication and Information System from Yanshan University in 2010. She has been working as an Associate Professor in Taiyuan Institute of Technology, China. She is also studying for her Ph.D. degree in College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia since 2021. Her research focuses on medical image segmentation. She can be contacted at email: wangjin@studyedus.cn.

Dr. Sharifah Aliman received a Bachelor of Computer Science, Midwestern State University, Texas, USA. Her Master of Science (Information Technology), Universiti Putra Malaysia and Ph.D. in Information Technology and Quantitative Science from Universiti Teknologi MARA, Shah Alam, Selangor. She has been serving as a computer science lecturer at Universiti Teknologi MARA since 1993. Her research areas of interest include data analytics, web analytics and computer science. She is also currently head of academic management, the office of the deputy vice chancellor (academic and international), UiTM, Shah Alam. Selangor. She can be contacted at email: sharifahali@uitm.edu.my.

Dr. Shafaf Ibrahim received a Bachelor of Computer Science, Master of Science (Computer Science), Ph.D. in Information Technology and Quantitative Science from Universiti Teknologi MARA, Shah Alam, Selangor. She is currently lecturing with Department of Computer Science, Universiti Teknologi MARA, Shah Alam. Her current academic post is the coordinator of postgraduate at College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam. Her research areas of interest in artificial intelligence include evolutionary algorithms, swarm intelligence, and image processing. She can be contacted at email: shafaf2429@uitm.edu.my.