A Lightweight Bilateral Left Ventricle Segmentation Method for Echocardiography

Zi Ye[1], Dan Wang[2], Lijun Zhang[1]*

[1] Institute of Intelligent Software Guangzhou, China
[2] Hangzhou Institute for Advanced Study of UCAS, China
yezi1022@gmail.com, wangdan233@mails.ucas.ac.cn, zhanglj@ios.ac.cn

Abstract

Echocardiographic images, often vulnerable to speckle and noise distortion, have a fuzzy property, presenting particular difficulties in manual segmentation. In this study, we introduced an improved bilateral segmentation network named Echo-BiSeNet, which offered a high-performance and low-computation solution. We implemented four types of convolutional neural network architectures, i.e., Echo-BiSeNet with backbones Xception and ResNet18, U-Net, BiSeNetV1 (Res18), and DeepLabV3 (Res18), that were trained, validated, and tested using 19,882 images. The experimental results indicated that Echo-BiSeNet (Res18) showed the highest DSC value and the least skewness for the left ventricle segmentation in the four-chamber view (0.9257±0.0384). This model also outperformed other conventional models regarding IoU, Recall, and Accuracy (0.8639, 0.9280, and 0.9868, respectively). In addition, the proposed Echo-BiSeNet (Res18) only has half of the number of parameters of the U-Net model, which results in a good balance between accuracy and efficiency.

Keywords: Ultrasound image segmentation, Left ventricle, Bilateral lightweight model, Cardiac

1 Introduction

Ultrasound (US) imaging is a prevalent modality for diagnosing cardiovascular diseases due to its non-invasive, radiation-free, cost-effective, and user-friendly nature compared to other imaging techniques [1]. The crucial step in evaluating the cardiac health status is segmenting the left ventricular (LV) endocardial border, quantifying the ejection fraction, and assessing regional wall motion. Usually, trained cardiac experts perform manual or semi-automated LV segmentation from the US images with the utilization of appropriate image analysis algorithms and techniques [2].

However, manual or semi-automatic segmentation of LV can be laborious and time-consuming, requiring the expertise of specialized clinical physicians. Moreover, the reproducibility of segmentation results is poor, necessitating repeated segmentation for different cardiac patients [3]. To mitigate these issues, an automated left ventricle segmentation system utilizing ultrasound (US) can offer a degree of solution.

Therefore, it is imperative to investigate rapid and accurate automatic segmentation of left ventricles from US images to facilitate research and medical applications. This paper proposes a novel method for segmenting the left ventricle in echocardiography with irregular and varied morphology based on an improved BiSeNet model, termed Echo-BiSeNet. Contributions made by the present work to the LV segmentation area are shown below.

(1) We achieved precise LV segmentation in four-chamber view images with a bilateral-structured network, where both high- and low-level features were taken into account. The two network branches are designed explicitly for separately capturing spatial and semantic information, contributing to the high precision of the segmentation results.

(2) In this study, a multi-scale feature pyramid module was incorporated into the early stages of the semantic path network, thereby enabling the network to extract and integrate semantic information from objects of varying scales.

(3) We validated our model using the EchoNet-Dynamic dataset [4], which consists of a large sample of 10,036 patients with varying image quality and pathology. Our results demonstrate that the proposed model outperforms traditional BiSeNet and other classic segmentation models.

2 Related Work

So far, deep learning (DL) development has promoted automatic medical image segmentation, and several well-known deep learning frameworks have provided good ideas for echocardiography segmentation with outstanding performance [6]. Table 1 summarizes recent research on AI's application based on mainstream DL frameworks for LV segmentation.

Amer et al. [7] proposed ResDUnet, an innovative U-Net-based approach for deep learning segmentation, where feature extraction is incorporated at varying scales by integrating cascaded dilated convolution. In addition, YOLOv3 is adopted by Zhuang et al. [8] to determine the constraints and positioning of the LV summit and bottom, as well as the LV region. Furthermore, encouraged by the success of transformer architecture, Deng et al. [11] have proposed

*Corresponding Author: Lijun Zhang; E-mail: zhanglj@ios.ac.cn
DOI: https://doi.org/10.70003/160792642024072504003

TransBrdige. With this efficient and lightweight model, an encoder-decoder framework for convolutional neural network (CNN) is integrated with transformer architecture for fulfilling LV segmentation tasks.

Most recently, Ye et al. [10] presented Bi-DCNet, a DL methodology that benefited from a Bilateral Segmentation Network (BiSeNet), to achieve automatic 4-chamber view segmentation of LV. By introducing dilated convolutions, this model enlarges its receptive field by incorporating feature extraction, thereby enabling multi-scale data acquisition. Mask R-CNN [12] is a renowned dual-stage segmentation algorithm with immense success in image segmentation. Lei et al. [9] put forward an anchorless Mask CNN-based technique, termed Cardiac-SegNet, to outline LV rapidly and precisely. It consists of 3 subnetworks: a backbone, a mask head, as well as a fully convolutional one-stage object detector (FCOS) head.

In most studies concerning the deep learning-based segmentation of ultrasound LV, only the end-systole (ES) and end-diastole (ED) images were trained and assessed for accuracy, while the remaining cardiac cycle was unexplored. Smistad et al. [5] investigated using neural networks with convolutional long short-term memory (ConvLSTM) layers for real-time temporal coherent LV segmentation. The results indicate that increasing the number of frames annotated from the cardiac cycle helps.

Table 1. Studies of AI's application based on mainstream DL frameworks for LV segmentation

Authors	Year	Base framework	Model name	Dataset	Results
Smistad, E. et al. [5]	2021	LSTM	ConvLSTM NN	HUNT; CAMUS	LV Segmentation (Dice 0.93 ± 0.03)
Amer, A. et al. [7]	2021	U-Net	ResDUnet	CAMUS	LV Segmentation (Dice 0.95 ± 0.03)
Zhuang, Z. et al. [8]	2021	YOLOv3	Proposed YOLOv3	Private Dataset	LV Segmentation (Dice 0.94 ± 0.20)
Lei, Y. et al. [9]	2021	Mask R-CNN	Cardiac-SegNet	CAMUS	LV endocardium Segmentation (mDice 0.92) LV epicardium Segmentation (mDice 0.95)
Ye, Z. et al. [10]	2023	BiSeNet	Bi-DCNet	EchoNet-Dynamic	LV Segmentation (Dice 0.92 ± 0.04)
Deng, K. et al. [11]	2021	Transformer	TransBridge	EchoNet-Dynamic	LV Segmentation (Avg.Dice 0.92)

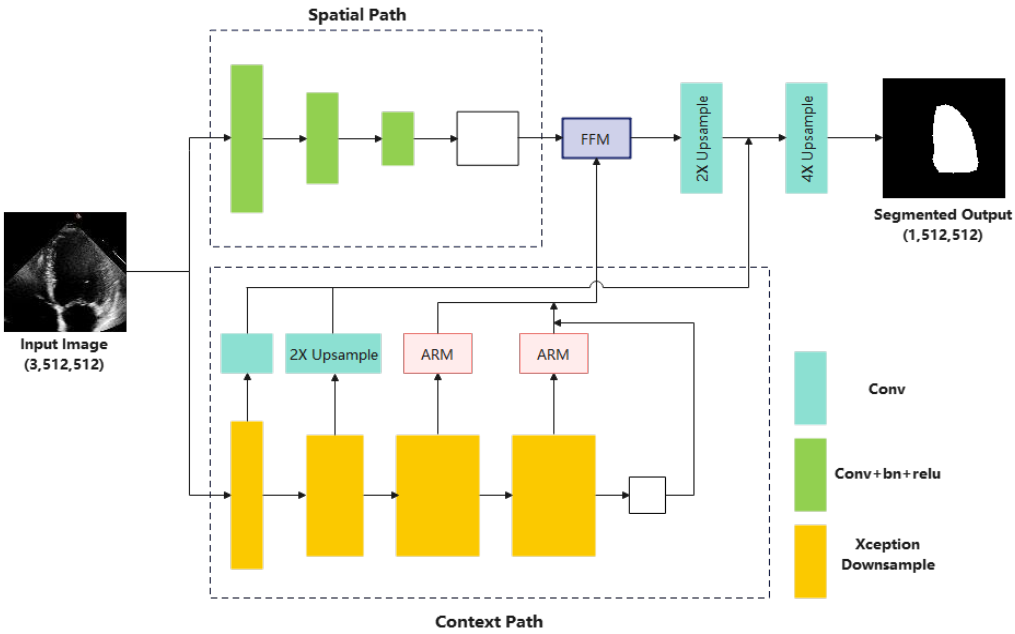

Figure 1. An overview of the proposed Echo-BiSeNet

3 Methodology

The Echo-BiSeNet model we propose inherits the foundational principles of the classical BiSeNet model, which encompasses two paths, one spatial and the other contextual. The spatial path is responsible for creating high-resolution feature maps by employing convolutions with small strides, thereby ensuring the preservation of spatial information. Conversely, the context pathway uses a rapid downsampling strategy to attain a broader receptive field. As a result, the model extracts high-level contextual and low-level spatial details by capitalizing on these dual branch networks. Furthermore, the network's performance is enhanced through feature fusion and optimization using auxiliary loss functions, ultimately yielding accurate segmentation results.

The general structure of the proposed Echo-BiSeNet is shown in Figure 1. The details are discussed in the following section.

3.1 Spatial Path

The Spatial Path (SP) module aims to acquire high-resolution feature maps for more precise spatial data. Its structure consists of three convolutional layers, each comprising a convolutional layer whose stride is 2, followed by a batch normalization layer and a ReLU activation layer [13]. As a result, the output image is diminished by this pathway to 1/8 of the original one. The details of the structure are presented in Figure 2.

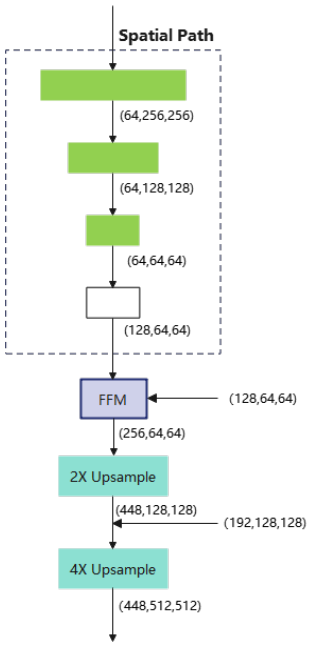

Figure 2. The detailed architecture of the Spatial Path of Echo-BiSeNet

3.2 Context Path

The context path is used in this paper to get a sizeable receptive field with efficient computation simultaneously. The context path uses the lightweight model and global average pooling to downsample the feature map fast to obtain a large receptive field and encode high-level semantic context information [14]. The detailed architecture of the Context Path is shown in Figure 3.

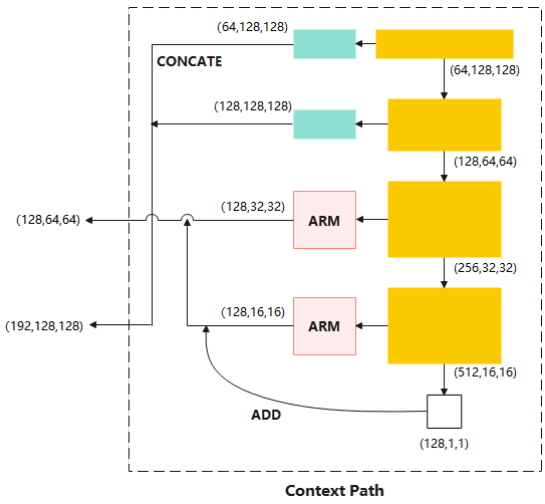

Figure 3. The detailed architecture of the Context Path of Echo-BiSeNet

Furthermore, this study integrates a Multi-scale Feature Pyramid module into the initial embedding of its semantic path network, facilitating the acquisition of semantic information pertaining to objects of diverse scales. The Feature Pyramid module undergoes convolutional processing on the feature maps that have been downsampled by a factor of 4 and 8 within the original BiSeNet semantic path, yielding two low-level feature maps. This augmentation significantly enhances the model's precision in segmenting small objects. As formulated in Equation (1) and (2), a 3D tensor is adopted as output.

$$\mathrm{MFP}\left(C^{H\times W\times C}\right)\rightarrow C^{H\times W\times C}. \tag{1}$$

$$\mathrm{MFP}\left(C^{H\times W\times C}\right)\rightarrow C^{H'\times W'\times C}. \tag{2}$$

The same attention refinement module (ARM) from BiSeNet is added to refine the features of each downsample stage. ARM is used to optimize the feature maps in the down-sampling stage of the Context Path. The final layer of the Context Path employs global average pooling (GAP) [15] to fuse the output features from each stage that have passed through ARM, thereby concatenating low-level and deeper-level semantic information. ARM belongs to the channel attention mechanism, which first utilizes GAP to obtain global semantic information. After normalization, a sigmoid activation layer calculates an attention vector to re-weight the importance of the features. This module optimizes the output features of each stage in the down-sampling process, and it can integrate global feature information without the need for upsampling, making the computational cost negligible.

3.3 Feature Fusion Model

Similarly, the integration of SP and CP features is

accomplished via the classic FFM technique. First, an FFM connects the components generated from both SP and CP. The computational process of low- and high-level information syndication is formulated by Equation (3).

$$F_c = F^L \oplus F^H. \tag{3}$$

where F^L and F^H stand respectively for low- and high-level features; $\oplus$ symbolizes element-level addition.

Subsequently, the fused features are subjected to scale normalization through batch normalization. The pooling of connected features proceeded into a single feature vector, followed by the computation of a weight vector. Finally, feature re-weighting is accomplished with the utilization of the derived weight vector, enabling the integration of SP and CP feature outputs.

4 Dataset and Evaluation

4.1 Dataset

EchoNet-Dynamic provides echocardiograms in apical four-chamber (A4C) view obtained from Stanford University Hospital, which constitutes the most significant publicly available dataset of this class. This dataset comprises 10,030 echocardiogram clips sourced from 10,030 patients selected randomly, who underwent echocardiography between 2006 and 2018 as part of their clinical care at the hospital [16]. This dataset was gathered from patients with an average age of 57, of whom 55% were female. Additionally, it includes cases of hypertension (39%), heart failure (29%), and coronary artery disease (23%). Each video clip represents a distinct individual within the dataset. To identify the specific video clip showcasing the apical four-chamber view, Digital Imaging and Communications in Medicine (DICOM) files associated with ventricular volume measurements used for ejection fraction calculations were extracted from the apical 4-chamber view.

After the data cleansing process, a total of 9,989 video samples out of the initial 10,030 videos were included in the analysis. Each video was examined solely for the end-diastolic and end-systolic frames. Expert cardiologists and sonographers annotated the left ventricular (LV) region during the routine clinical workflow. Since the corresponding labels of the left ventricle endocardium were given by X coordinates of the left-most point and line segment and Y coordinates of the right-most point of line segment, a filled polygon was drawn based on lists of point coordinates X Y to obain the ground truth for each sample image.

Ninety-six images were excluded from the experiment due to the poor quality of ground truth generated. Consequently, 14,846 out of 19,882 images were selected for the training set, while the remaining images were divided for separate use in the validation (2,563 images) and testing (2,473 images) sets. Additionally, the end-diastolic and end-systolic frames from the same individuals were grouped together.

All images underwent center cropping and resizing to a dimension of 512×512 [17]. Data augmentation techniques were applied during the training phase, including random flipping with a 0.5 probability and random rotation within a range of ±25 degrees.

4.2 Network Training

The complete experimental setup employed a computer system consisting of an Intel(R) Xeon(R) CPU E5-2678 v3 @2.5GHz, four NVIDIA RTX 2080TI GPUs, along with 32GB RAM. Throughout the experimentation process, open-source toolboxes PyTorch (version 1.9.1) and MMSegmentation were utilized.

During the training process, network optimization was accomplished using stochastic gradient descent (SGD) by setting the momentum to 0.9 and the weight decay to 0.0005. Noticeably, a "poly" learning rate scheme was also applied in which $\eta = \eta_0 \left(1 - \frac{n}{N}\right)^{\beta}$, with η denoting learning rate; η_0 = 0.0001 standing for initial learning rate; n and N severally denoting the present epoch and overall epochs; and β having a value of 0.9.

For the remaining components of EchoNet-Dynamic, random initialization was applied using the default PyTorch settings. Each model was subjected to 160,000 training iterations with a batch size of 96. Subsequently, the model configuration that yielded the best validation results was selected for evaluating the test dataset to assess model performance.

4.3 Evaluation Metrics

For the performance assessment of our approach, intensive experiments are conducted based on the EchoNet-Dynamic dataset. Binary masks were generated by trained models for test images through segmentation. Comparison of these masks against the corresponding ground truth binary masks proceeded. Subsequently, the performance of our model was assessed and contrasted against other models of semantic segmentation. This evaluation was conducted using well-established evaluation metrics, including specificity, accuracy, recall, precision, dice similarity coefficient (DSC), and intersection over union (IoU) [18]. Computational formulas for every evaluation metric are provided below:

$$Accuracy = (TP + TN) / (TP + FN + TN + FP). \tag{4}$$

$$Recall = TP / (TP + FN). \tag{5}$$

$$Specificity = TN / (TN + FP). \tag{6}$$

$$Precision = TP / (TP + FP). \tag{7}$$

$$DSC = 2TP / (2TP + FN + FP). \tag{8}$$

$$IoU = DSC / (2 - DSC). \tag{9}$$

where TP and TN refer severally to the quantities of correctly

identified LV and background pixels; *FP* and *FN* stand severally for the background and LV pixels mistakenly assigned as the LV and background pixels.

5 Performance Evaluation

5.1 Evaluation Results

In this section, we compared the performance of Echo-BiSeNet with state-of-the-art approaches. In particular, a comparison was undertaken against the basic BiSeNetV1 model. Moreover, a comparison is made with the DeepLabV3 [19]. U-Net, a well-known convolutional framework for segmenting medical images quickly and accurately, was also applied to the EchoNet-Dynamic dataset. Finally, Xception and ResNet-18 were studied to investigate the effect of different backbones on our proposed structures.

The model that exhibited the highest validation mDice during the training phase was chosen to assess the final test outcomes. Detailed information on the peak validation mDice may be observed in Table 2. The validation mDice of the Echo-BiSeNet (Res18) model peaked at iteration 120,000, with a value of 0.9239. In comparison, the base BiSeNet model attained a 0.9122 maximum validation at iteration 140,000, while U-Net and DeepLabV3 (Res18) achieved a validation mDice of 0.9104 and 0.9116, respectively, at iteration 160,000.

Table 2. Performance of different models on peak validation mDice

Models	Peak validation mDice	Iteration
Echo-BiSeNet (Xception)	0.9226	100,000
Echo-BiSeNet (Res18)	0.9239	120,000
U-Net	0.9104	160,000
BiSeNetV1 (Res18)	0.9122	140,000
DeepLabV3 (Res18)	0.9116	160,000

The experimental index of various approaches on the EchoNet-Dynamic dataset is presented in Table 3. The table illustrates the superior performance of our Echo-BiSeNet architecture in accurately segmenting the left ventricle in a four-chamber image, particularly when utilizing the ResNet-18 backbone. The DSC, IoU, Recall, and Accuracy metrics are all at the forefront. It is worth mentioning that it even performs better than the EchoNet-Dynamic baseline, with a Dice coefficient of 0.9211 [20].

From a macroscopic perspective, it has been determined that the bilateral structure yields advantageous outcomes, and there is observable evidence of progress. The bilateral structural models, i.e., Echo-BiSeNet (Xception), Echo-BiSeNet (Res18), and BiSeNetV1 (Res18), revealed improved IoU and DSC values over DeepLabV3 (Res18) and U-Net. Besides, the BiSeNetV1 (Res18) model demonstrated exceptional performance in terms of Precision (0.9482) and Specificity (0.9951).

Finally, DeepLabV3 performed better than U-Net in the segmentation task. The Dice coefficient associated with DeepLabV3 is 0.9084, far surpassing the Dice coefficient associated with U-Net, which stands at a mere 0.8612. In conclusion, the U-Net model exhibits inferior performance compared to the other approaches employed in this study.

5.2 Discussion

In the present work, various segmentation networks are subjected to performance contrast. It proposes an innovative architecture named Echo-BiSeNet, which outperformed U-Net, BiSeNetV1 (Res18), and DeepLabV3 (Res18) regarding the trade-off between segmentation efficiency and parameter quantity.

Figure 4 provides a quantitative comparative analysis of the DSC across various segmentation models alongside the corresponding parameter count associated with each model. The findings suggest that Echo-BiSeNet (Res18) achieves the greatest DSC of 0.9257, marginally superior to the second-ranked Echo-BiSeNet (Xception). This observation shows that the Echo-BiSeNet structure we have developed ensures a high level of segmentation performance. In the meanwhile, it should be noticed that Echo-BiSeNet (Res18) requires only half the amount of parameters of U-Net, and the network can ensure better accuracy when encountered with fewer parameters. In brief, Echo-BiSeNet (Res18) effectively mitigated the challenges associated with training so that it can run smoothly on low-cost hardware and is friendly for clinical diagnosis due to the minimal parameter.

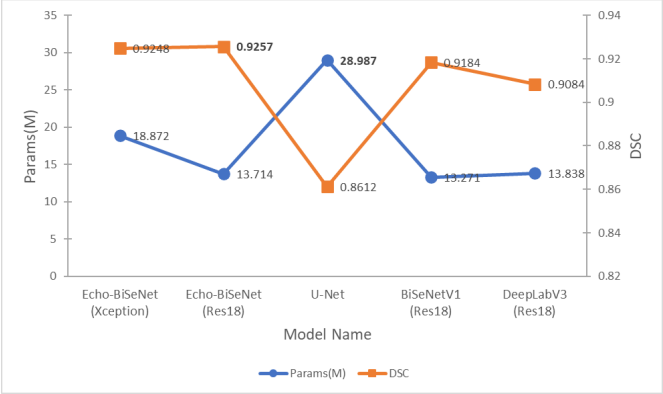

Figure 4. Comparison of DSC and parameter scale of different models

To further confirm the performance of the proposed model, Figure 5 depicts the ultrasound image of the patient, the manually annotated contours provided by the physician as the ground truth, and the visual segmentation outcomes generated by several methodologies.

Based on the analysis of the four chosen four-chamber view images, each depicting distinct variations in size and form of the left ventricle, it is obvious that our model exhibits excellent segmentation efficacy, while other models might display inadequate segmentation.

In the first condition, denoted as “a,” the left ventricle was successfully segmented using bilateral structured networks in specific ultrasound images of subpar quality and incompleteness. However, alternative methods were unsuccessful in achieving the same outcome.

Tile (c) displays the predicted outcomes of the modified BiSeNet and the original BiSeNet models for tiny targets in the test data. The segmentation results of the original BiSeNet offer insufficient edge features of the end-systolic left ventricle, which shrunk to minimum size, causing the omission of specific areas within the left ventricle. The reduced segmentation accuracy of small objects in the original BiSeNet semantic path is driven by its reliance solely on downsampled feature maps at 16 times, 32 times, and global pooling, neglecting the inclusion of local information from low-level features.

Visualization analysis demonstrated the effectiveness of Echo-BiSeNet concerning semantic segmentation of the left ventricle with high variation in shape and size.

Moreover, the statistical analysis of the evaluation indicators is also provided. Figure 6 illustrates the boxplots employed to analyze and contrast the distribution of evaluation metric values across a total of 2,473 test images. The boxplots once again confirmed the superior performance of the suggested model in comparison to other segmentation approaches. The model highlights enhanced precision in detecting small target objects by including the feature pyramid structure in the original BiSeNet.

Additionally, the U-Net model had the least satisfactory performance overall, as evidenced by its lowest scores across all metrics and the highest degree of skewness found in the boxplots. Also, it can be shown that DeepLabV3 (Res18) exhibits a higher frequency of outliers than the alternative model, suggesting that its performance lacks consistency across the entirety of the test dataset.

Table 3. Mean and standard deviation values of evaluation metrics

Evaluation metrics	IoU	DSC	Precise	Recall	Accuracy	Specificity
Echo-BiSeNet (Xception)						
Mean	0.8624	0.9248	0.9303	0.9259	0.9866	0.9933
Standard deviation	0.0640	0.0397	0.0678	0.0607	0.0077	0.0071
Echo-BiSeNet (Res18)						
Mean	**0.8639**	**0.925**7	0.9295	**0.9280**	**0.9868**	0.9930
Standard deviation	0.0622	0.0384	0.0674	0.0561	0.0076	0.0076
U-Net						
Mean	0.7668	0.8612	0.9230	0.8194	0.9766	0.9939
Standard deviation	0.1250	0.0987	0.0869	0.1283	0.0173	0.0062
BiSeNetV1 (Res18)						
Mean	0.8515	0.9184	**0.9482**	0.8967	0.9858	**0.9951**
Standard deviation	0.0639	0.0394	0.0579	0.0694	0.0080	0.0063
DeepLabV3(Res18)						
Mean	0.8373	0.9084	0.9264	0.8991	0.9835	0.9922
Standard deviation	0.0879	0.0648	0.0875	0.0782	0.0135	0.0124

6 Conclusions

For rapid diagnosis of abnormalities and to study LV function, the segmentation of LV is a critical task. A lightweight CNN named Echo-BiSeNet is developed herein, demonstrated in the EchoNet-Dynamic dataset for LV segmentation in US echocardiograms. The suggested model is derived from the BiseNet segmentation network and integrates the feature pyramid structure into its bilateral network path. Additionally, it introduces an enhanced version of BiSiNet for left ventricle segmentation, which aims to increase the model's capacity to recognize irregular shapes and small targets. Our model attains higher accuracy with the most minor parameters than other advanced techniques. Prospective studies are needed to investigate how the potential benefits of AI translate into clinical practice.

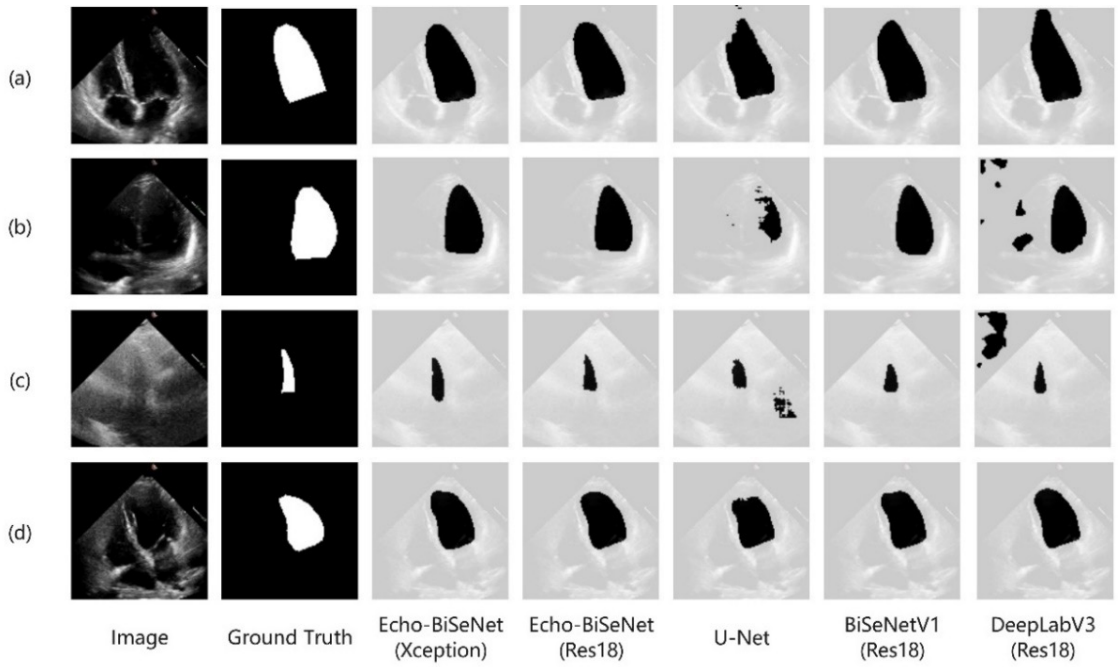

Figure 5. Visual results of the proposed Echo-BiSeNet method and other SOTA algorithms
(Selected frames from different videos with many motions of the left ventricle are presented from top to bottom.**)**

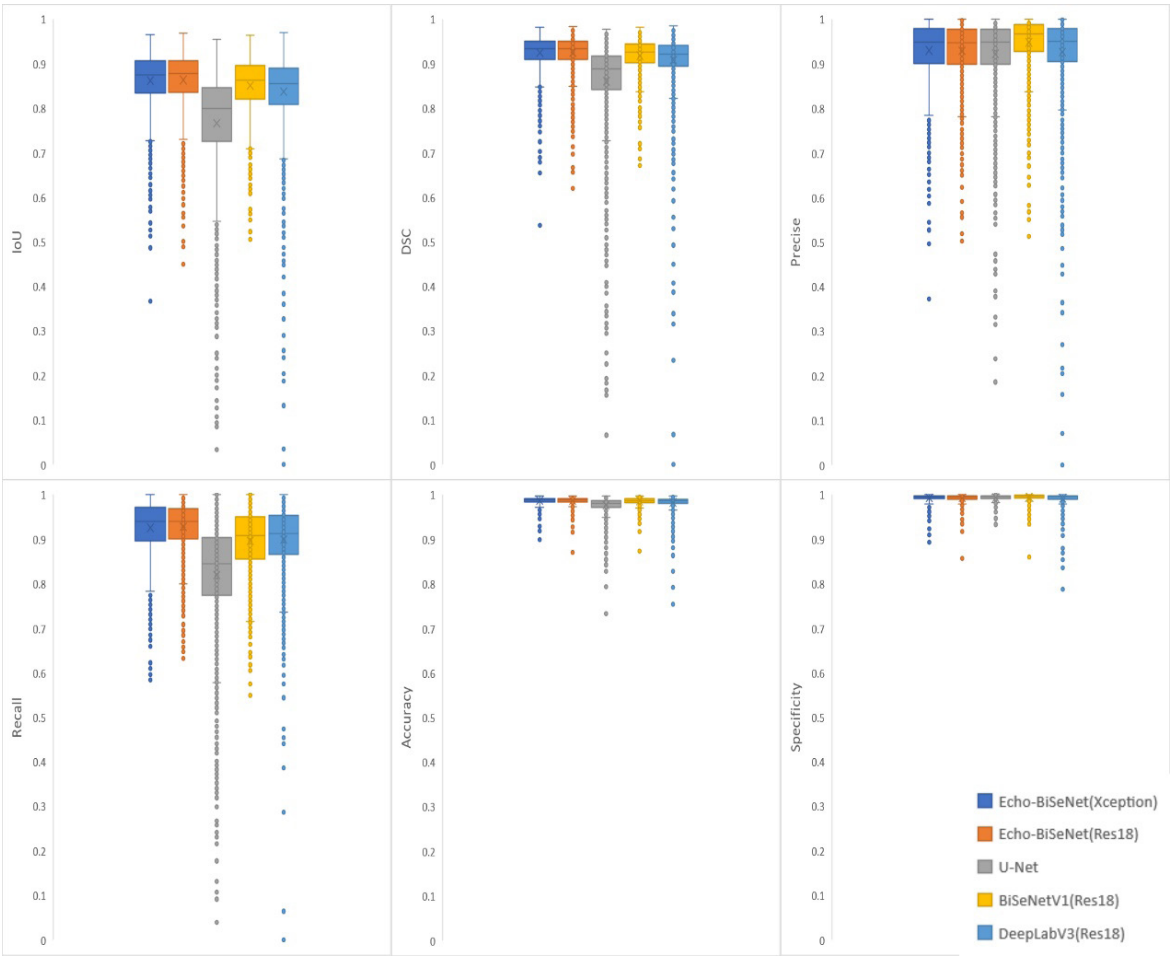

Figure 6. The boxplots of IoU, DSE, Precise, Recall, Accuracy, and Specificity of each model

References

[1] A. Tarighatnia, M. R. Fouladi, N. D. Nader, A. Aghanejad, H. Ghadiri, Recent trends of contrast agents in ultrasound imaging: a review of the classifications and applications, *Materials advances*, Vol. 3, No. 9, pp. 3726–3741, March, 2022.

[2] J. Tromp, P. J. Seekings, C.-L. Hung, M. B. Iversen, M. J. Frost, W. Ouwerkerk, Z. Jiang, F. Eisenhaber, R. S. M. Goh, H. Zhao, W. Huang, L.-H. Ling, D. Sim, P. Cozzone, A. M. Richards, H. K. Lee, S. D. Solomon, C. S. P. Lam, J. A. Ezekowitz, Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study, *The Lancet Digital Health*, Vol. 4, No. 1, pp. e46–e54, December, 2022.

[3] Y. Zeng, P.-H. Tsui, K. Pang, G. Bin, J. Li, K. Lv, X. Wu, S. Wu, Z. Zhou, MAEF-Net: Multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography, *Ultrasonics*, Vol. 127, Article No. 106855 , January, 2023.

[4] D. Ouyang, B. He, A. Ghorbani, M. P. Lungren, E. A. Ashley, D. Liang, J. Zou, EchoNet-Dynamic: a Large New Cardiac Motion Video Data Resource for Medical Machine Learning, *NeurIPS ML4H Workshop*, Vancouver, BC, Canada, 2019, pp. 24–26.

[5] E. Smistad, I. M. Salte, H. Dalen, L. Lovstakken, Real-time temporal coherent left ventricle segmentation using convolutional LSTMs, *2021 IEEE International Ultrasonics Symposium (IUS)*, Xi'an, China, 2021, pp. 11–16.

[6] J. Wang, C. Jin, Q. Tang, N. N. Xiong, G. Srivastava, Intelligent Ubiquitous Network Accessibility for Wireless-Powered MEC in UAV-Assisted B5G, *IEEE Transactions on Network Science and Engineering*, Vol. 8, No. 4, pp. 2801–2813, October-December, 2021.

[7] A. Amer, X. Ye, F. Janan, ResDUnet: a deep learning-based left ventricle segmentation method for echocardiography, *IEEE Access*, Vol. 9, pp. 159755–159763, October, 2021.

[8] Z. Zhuang, P. Jin, A. N. J. Raj, Y. Yuan, S. Zhuang, Automatic segmentation of left ventricle in echocardiography based on YOLOv3 model to achieve constraint and positioning, *Computational and Mathematical Methods in Medicine*, Vol. 2021, pp. 1–11, May, 2021.

[9] Y. Lei, Y. Fu, J. Roper, K. Higgins, J. D. Bradley, W. J. Curran, T. Liu, X. Yang, Echocardiographic image multi-structure segmentation using Cardiac-SegNet, *Medical Physics*, Vol. 48, No. 5, pp. 2426–2437, March, 2021.

[10] Z. Ye, Y. J. Kumar, F. Song, G. Li, S. Zhang, Bi-DCNet: Bilateral Network with Dilated Convolutions for Left Ventricle Segmentation, *Life*, Vol. 13, No. 4, Article No. 1040, April, 2023.

[11] K. Deng, Y. Meng, D. Gao, J. Bridge, Y. Shen, G. Lip, Y. Zhao, Y. Zheng, Transbridge: A lightweight transformer for left ventricle segmentation in echocardiography, *Simplifying Medical Ultrasound: Second International Workshop*, Strasbourg, France, 2021, pp. 63–72.

[12] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, *Proceedings of the IEEE international conference on computer vision*, Venice, Italy, 2017, pp. 22–29.

[13] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, Z. Liu, Dynamic ReLU, *Proceedings of the European conference on computer vision (ECCV)*, Glasgow, UK, 2020, pp. 23–28.

[14] Y. Liu, F. Yi, Y. Ma, Y. Wang, ASA-BiSeNet: improved real-time approach for road lane semantic segmentation of low-light autonomous driving road scenes, *Applied Optics*, Vol. 62, No. 19, pp. 5224–5235, July, 2023.

[15] R. L. Kumar, J. Kakarla, B. V. Isunuri, M. Singh, Multi-class brain tumor classification using residual network and global average pooling, *Multimedia Tools and Applications*, Vol. 80, No. 9, pp. 13429–13438, April, 2021.

[16] Z. Liu, Y. Huang, H. Li, W. Li, F. Zhang, W. Ouyang, S. Wang, Z. Luo, Y. Chen, R. Xia, Y. Li, X. Pan, A generalized deep learning model for heart failure diagnosis using dynamic and static ultrasound, *Journal of Translational Internal Medicine*, Vol. 11, No. 2, pp. 138–144, April-June, 2023.

[17] J. Wang, Y. Liu, S. Rao, R. S. Sherratt, J. Hu, Enhancing Security by Using GIFT and ECC Encryption Method in Multi-Tenant Datacenters, Computers, *Materials & Continua*, Vol. 75, No. 2, pp. 3849–3865, March, 2023.

[18] J. Wang, Y. Zou, P. Lei, R. S. Sherratt, L. Wang, Research on Recurrent Neural Network Based Crack Opening Prediction of Concrete Dam, *Journal of Internet Technology*, Vol. 21, No. 4, pp. 1161–1170, July, 2020

[19] Z. Hu, J. Zhao, Y. Luo, J. Ou, Semantic SLAM Based on Improved DeepLabv3^{+} in Dynamic Scenarios, *IEEE Access*, Vol. 10, pp. 21160–21168, February, 2022.

[20] M. Saeed, R. Muhtaseb, M. Yaqub, Contrastive pretraining for echocardiography segmentation with limited data, *Annual Conference on Medical Image Understanding and Analysis*, Cambridge, UK, 2022, pp. 680–691.

Biographies

Zi Ye received Master's degree in Applied Statistics from the University of Oxford, UK, in 2010 and PhD at Universiti Teknikal Malaysia Melaka in 2022. She is now a postal doctor at the Institute of Intelligent Software, Guangzhou, China. Her research interests involve Artificial Intelligence & Machine Learning.

Dan Wang received B.S. degree in Software Engineering from Nanchang University, China, in 2023 and is now a current graduate student in Artificial Intelligence at University of Chinese Academy of Sciences. Her research interests involve medical artificial intelligence and machine learning.

Lijun Zhang is the director of the Sino-European Joint Laboratory for Reliable and Intelligent Software at the Guangzhou Intelligent Software Industry Research Institute. His research area focuses on model checking for probabilistic concurrent systems based on Markov models and their extensions.