Mathematical Biosciences and Engineering

MBE, 20(2): 2081–2093.
DOI: 10.3934/mbe.2023096
Received: 01 September 2022
Revised: 18 October 2022
Accepted: 27 October 2022
Published: 14 November 2022

http://www.aimspress.com/journal/MBE

Research article

Semantic segmentation method for myocardial contrast echocardiogram based on DeepLabV3+ deep learning architecture

Huan Cheng[1], Jucheng Zhang[2], Yinglan Gong[3], Zhaoxia Pu[4], Jun Jiang[4], Yonghua Chu[2] and Ling Xia[1,*]

[1] Key Laboratory for Biomedical Engineering of Ministry of Education, Institute of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China
[2] Department of Clinical Engineering, School of Medicine, The Second Affiliated Hospital, Zhejiang University, Hangzhou 310009, China
[3] Institute of Wenzhou, Zhejiang University, Wenzhou 325036, China
[4] Department of Cardiology, School of Medicine, The Second Affiliated Hospital, Zhejiang University, Hangzhou 310009, China

* **Correspondence:** Email: xialing@zju.edu.cn.

Abstract: Myocardial contrast echocardiography (MCE) has been proposed as a method to assess myocardial perfusion for the detection of coronary artery diseases in a non-invasive way. As a critical step of automatic MCE perfusion quantification, myocardium segmentation from the MCE frames faces many challenges due to the low image quality and complex myocardial structure. In this paper, a deep learning semantic segmentation method is proposed based on a modified DeepLabV3+ structure with an atrous convolution and atrous spatial pyramid pooling module. The model was trained separately on three chamber views (apical two-chamber view, apical three-chamber view, and apical four-chamber view) on 100 patients' MCE sequences, divided by a proportion of 7:3 into training and testing datasets. The results evaluated by using the dice coefficient (0.84, 0.84, and 0.86 for three chamber views respectively) and Intersection over Union(0.74, 0.72 and 0.75 for three chamber views respectively) demonstrated the better performance of the proposed method compared to other state-of-the-art methods, including the original DeepLabV3+, PSPnet, and U-net. In addition, we conducted a trade-off comparison between model performance and complexity in different depths of the backbone convolution network, which illustrated model application feasibility.

Keywords: myocardial contrast echocardiogram; semantic segmentation; DeepLabV3+; deep learning

1. Introduction

Coronary artery disease is the leading cause of cardiovascular mortality. Cardiac imaging has a pivotal role in preventing, diagnosing and treating ischemic heart disease. In recent years, non-invasive clinical cardiac imaging techniques have been rapidly developed; they are commonly used to assess myocardial ischemia and quantitative perfusion parameters, including single-photon emission computed tomography (SPECT), magnetic resonance imaging, computed tomography, myocardial contrast echocardiography (MCE), etc. Compared with the other cardiac imaging techniques, MCE has the advantages of being radiation-free, convenient and inexpensive (about 3–4 times less expensive than SPECT) [1]. MCE has been validated as an effective myocardial perfusion imaging method- to evaluate myocardial perfusion and infarction size [2,3], the microvascular changes after coronary revascularization [4] and the outcome in those undergoing a heart transplant [5].

In the clinical application of MCE, the ultrasound contrast agents (UEAs) containing the gas cores and lipid shells are injected intravenously into the myocardium. Contrast imaging can be generated from the signals produced by the resonance of microbubbles; blood containing microbubbles appear to be a bright white region [6]. Shown in Figure 1, after the distribution of the UEA reaches a stable state, high mechanical index impulses will be applied to clear all of the microbubbles, and then several end-systolic frames during the destruction-replenishment of UEA are selected to fit the time-intensity curve (shown in Figure 2), from which myocardial perfusion parameters such as blood volume, and flux rate can be obtained.

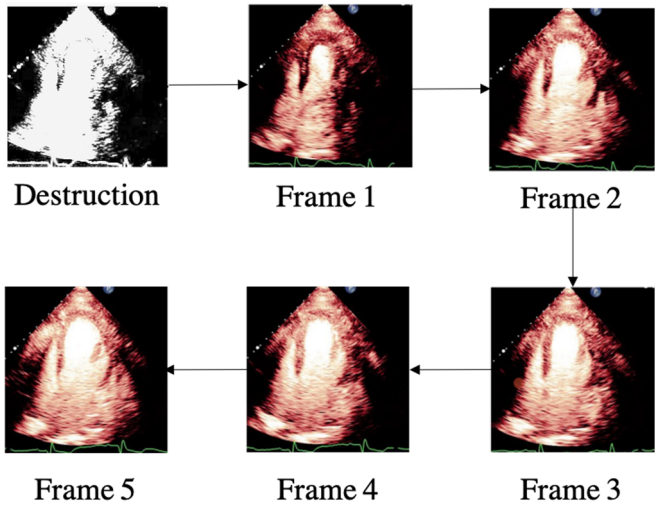

Figure 1. MCE destruction and replenishment imaging.

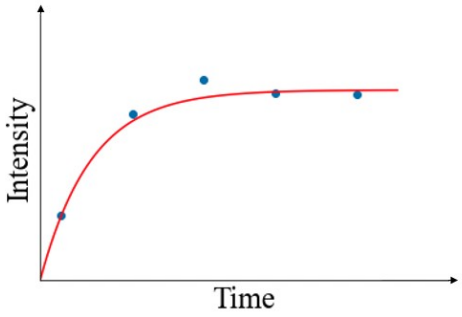

Figure 2. Time-intensity curve of MCE destruction and replenishment.

However, the analysis of MCE is very time-consuming, as it includes two steps that need to be done manually by experienced echocardiographers. Several end-systolic frames need to be extracted, and then the myocardium, i.e., the region of interest, needs to be segmented in the cardiac frames too. Therefore, automatic myocardial segmentation methods are desired for efficiency and operator-independence of the MCE perfusion analysis. Nevertheless, myocardial segmentation faces the following challenges. First, the concentration of UEA changes over time during destruction and replenishment, causing great intensity variations in images [7]. Second, the variations in shape and position of the myocardium according to different chambers, heart motions, patient individual differences, etc. Moreover, unclear myocardial borders and misleading structures such as papillary muscle have similar appearances to the myocardium.

Existing myocardial segmentation methods could be broadly classified into traditional image segmentation algorithms and machine learning algorithms. Traditional image segmentation algorithms define the segmentation task as a contour finding problem by using optimization methods based on image information, such as active contour [8] and an active shape model [9]. Malpica et al. [10] proposed a coupled active contour model guided by optical flow estimates to track the myocardium in MCE. Pickard et al. [11] applied principal component analysis with an active shape model algorithm to model the shape variability; they proposed a specialized gradient vector flow field to guide the contours to the myocardial borders, Guo et al. [12] proposed an automatic myocardial segmentation method based on an active contours model and neutrosophic similarity score; they applied a clustering algorithm to detect the initial ventricle region to speed up the evolution procedure and increase accuracy. However, due to the low complexity of the traditional image segmentation algorithm, it does not perform well on the MCE myocardial segmentation task with a large intensity variation [13] and it still needs manual tracing of myocardial contours; in addition, the optimization algorithm would be easily stuck in the local optimal solution without a good manual initial contour. Machine learning algorithms for myocardial segmentation tasks are often defined as pixel-level classification tasks, known as semantic segmentation. Li et al. [13] combined a random forest with a shape model, achieving notable improvement in segmentation accuracy compared with the classic random forest and active shape model. In recent years, deep learning has shown superior performance and great potential in medical image analysis; the majority of these deep learning approaches in cardiac ultrasound focus on left ventricle segmentation. Azarmehr et al. [14] experimented with three deep learning left ventricle (LV) segmentation models (U-Net, SegNet and fully connected DenseNets) on 992 echocardiograms, the U-net model outperformed the other models and achieved an average dice coefficient of 0.93. Veni et al. [15] proposed a U-net combined with a shape-driven deformable model in

the form of a level set. The U-net model is used to produce the segmentation of LV, which is considered as a prior shape; then, the prior shape drives the level set to converge the final shape; the model produced a 0.86 dice coefficient on a private 2D echocardiographic dataset. Hu et al. [16] proposed a segmentation model based on a bilateral segmentation network (BiSeNet); it consists of two paths, a spatial path for capturing low-level spatial features and a context path for exploiting high-level context semantic features; they also used a fusion module to fuse the features of those two paths, achieving a dice coefficient of 0.932 and 0.908 in the left ventricle and left atrium, respectively. To the best of our knowledge, Li et al. [17] was the first to apply deep learning methods to MCE segmentation; they proposed an encoder-decoder architecture based on a U-net, introduced a bi-directional training schema incorporating temporal information in MCE sequences and achieved the highest segmentation precision compared to the traditional U-net model.

However, we believe that the MCE segmentation accuracy still has great improvement space due to the rapid development of new deep learning algorithms. Among all of the deep learning algorithms, DeepLabv3+ [18] has become an excellent algorithm in the field of medical segmentation by virtue of its ability to extract multi-scale information and its encoder-decoder structure. Thus, in this paper, we propose a semantic segmentation method based on DeepLabV3+ to solve the segmentation problem in the MCE automatic perfusion quantification.

2. Materials and methods

2.1. Dataset description

Li et al. [17] have made the MCE dataset publicly available; it consists of MCE data from 100 patients from Guangdong Provincial People's Hospital. Apical two-chamber view (A2C), apical three-chamber view (A3C) and apical four-chamber view (A4C) MCE data were collected from each patient. Every MCE sequence has 30 end-systolic frames. In summary, there are 100 (patients) × 3 (chamber views) × 30 (end-systolic frames) = 9000 frames. The manual annotations of the myocardium were performed by an experienced echocardiographer. We split the dataset into the training dataset and test dataset at a proportion of 7:3. The segmentation models were trained for each chamber view separately; the detailed data information is illustrated in Table 1.

Table 1. Detailed dataset information. Each patient has three chamber views (apical two-chamber view, apical three-chamber view, apical four-chamber view), and each chamber view has 30 frames.

	Patient number	MCE sequence number	Frame number
training data	70	210	6300
testing data	30	90	2700

2.2. Segmentation model

As shown in Figure 3, the segmentation model was modified based on Deeplabv3+; it consists of a dilated ResNet backbone to extract feature maps, an atrous spatial pyramid pooling (ASPP) module to convert feature maps into multi-scale information and a decode module to generate the final predictions.

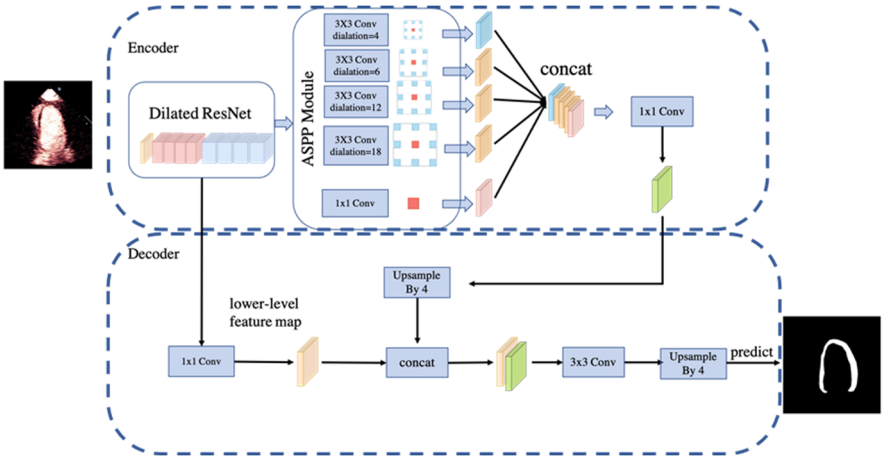

Figure 3. Architecture of modified segmentation model; the figure was adapted from DeepLabV3+.

2.2.1. Backbone

The backbone network was based on a modified 101 depth ResNet [19]. First, we replaced the 7 × 7 convolution in the input stem with a 3 × 3 convolution to improve the performance and accelerate the training process [20]. the standard ResNet uses downsampling operations such as a convolutional layer with a stride greater than 1 to increase feature maps. However, it would cause receptive field reduction; thus, DeepLabV3+ utilizes dialated convolutions [21] to alleviate spatial information losses from downsampling operations, also known as atrous convolution. The implementation of atrous convolution involves adding zeros between weights in the convolutional kernel with a stride of 1, as shown in Figure 4. In this way, features can be extracted across pixels, increasing the receptive field without introducing redundant parameters that need to be learned. Figure 5 shows a comparison of the original ResNet and dilated ResNet in the final two groups of the ResNet.

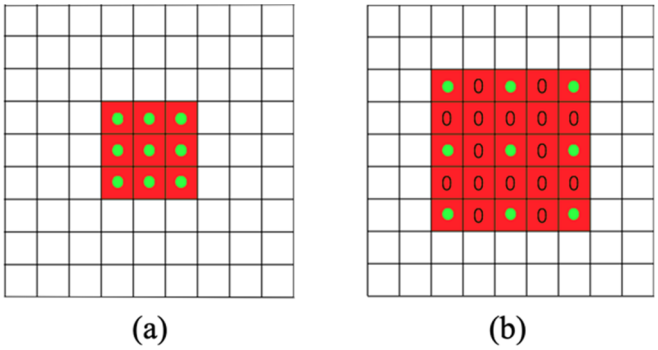

Figure 4. (a) Standard convolutional kernel and (b) convolutional kernel with dilation factor = 2; the figure was adapted from DeepLabV3+.

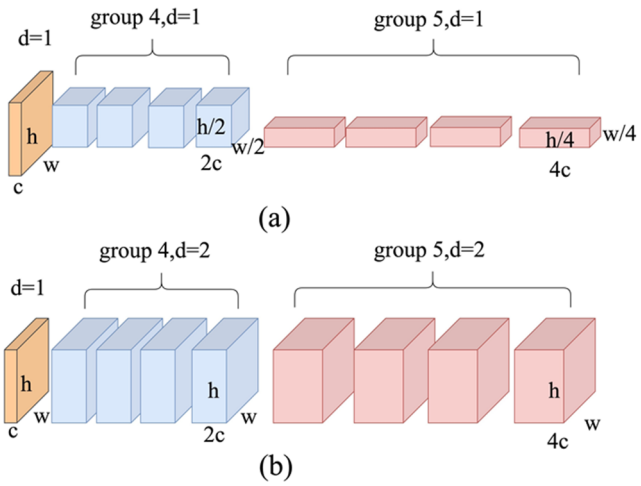

Figure 5. Final two groups (G4 and G5) of original ResNet (a) and dilated ResNet (b), where c, h and w respectively stand for the channel, height and weight of the feature map; d is the dilation factor, the dilation factor of standard convolution equals to 1; the figure was adapted from DeepLabV3+.

2.2.2. Atrous spatial pyramid pooling module

The ASPP module applies atrous convolution to extract multi-scale information by using atrous convolution with different dilation factors. The ASPP module consists of one 1 × 1 standard convolution and 3 × 3 atrous convolutions in parallel. The original DeepLabV3+ model proposed dilation factors of 6, 12 and 18 for atrous convolutions; however, the original structure may not be suitable because the myocardial border is not very clear due to the huge intensity variation in MCE; so, we added convolution with dilation factor 4 to the ASPP module to obtain more detailed spatial information. In conclusion, the ASPP module gets five feature maps from five parallel atrous convolutions and concatenates them together; it then sends them to the decoder. The detailed architecture of the ASPP module is shown in Figure 3.

2.2.3. Decoder

The decoder decodes features aggregated by the encoder at multiple levels and generates a semantic segmentation mask from high dimensional feature vectors. The decoder is simple but effective, and it is the most most significant improvement of the DeepLabV3+ compared to the predecessor DeepLabV3 [22]; in this way, the detailed boundaries of the myocardium can be recovered faster and stronger [18]. In the decoder module, the feature map from the encoder is first upsampled by 4 bilinearly and then concatenated with the lower-level feature map from the backbone after channel reduction from 1 × 1 convolution; since the lower-level feature map contains more spatial information, the fusion of the lower-level feature map and high-level feature map improves the segmentation accuracy. After a 3 × 3 convolution, the segmentation prediction is obtained by upsampling by 4 bilinearly.

2.3. Experiment setup

The model was implemented by using a Pytorch 1.11.0 deep learning framework and trained using NVIDIA RTX 2060 SUPER with 8 GB of memory. All images were center cropped to 256 × 256 and RGB pixels were normalized using the following: mean = [0.485, 0.456, 0.406], standard deviation = [0.229, 0.224, 0.225].

For data augmentation, during the training phase, all images were randomly scaled by [0.8, 1.2], rotated by [-5◦, 5◦] and randomly flipped by a probability of 0.5. During the testing phase, we did not apply any augmentations.

Every model was trained for 80 epochs that contained 84000 iterations. Stochastic gradient descent was used as the optimizer, where the momentum was set to 0.9 and the weight decay was set to 3e-5. The initial learning rate was set to 0.01, and the minimum learning rate was set to 0.001 and followed the polynomial decay policy, which is defined as

$$lr = initial\_lr \times \left(1 - \frac{iteration}{num\_iteration}\right)^{power} \tag{1}$$

where *iteration* represents the current iteration, *num_iteration* represents the total iteration, *initial_lr* = 0.01, *power* = 0.9.

Moreover, we used dice loss [23] as our loss function due to the imbalance problem, because the myocardium is small compared to the large heart chamber; it is defined as

$$loss_{dice}(P,T) = 1 - \frac{2|P \cap T|}{|P|+|T|} \tag{2}$$

where P is the predicted myocardium area and T is the ground truth of the myocardium area. In addition, we added an auxiliary loss [24] only in the training phase to help optimize the learning process.

3. Results and discussion

3.1. Evaluation criterion

The Dice coefficient and intersection over union (IoU) were used as evaluation criteria to evaluate the performance of the model; they are defined as

$$dice(P,T) = \frac{2|P \cap T|}{|P|+|T|} \tag{3}$$

$$IOU(P,T) = \frac{|P \cap T|}{P \cup T} \tag{4}$$

where P is the predicted myocardium area and T is the ground truth.

3.2. Results and comparison

The visualization of the segmentation results are illustrated in Figure 6; six apical four-chamber view MCE images were randomly selected from a subject in the test dataset and input into our trained model. It can be seen from the figure that the proposed model gets the correct prediction of

myocardium in the presence of the misleading structure, papillary muscle. The boundaries of the predicted segmentation area have a great match with the ground truth.

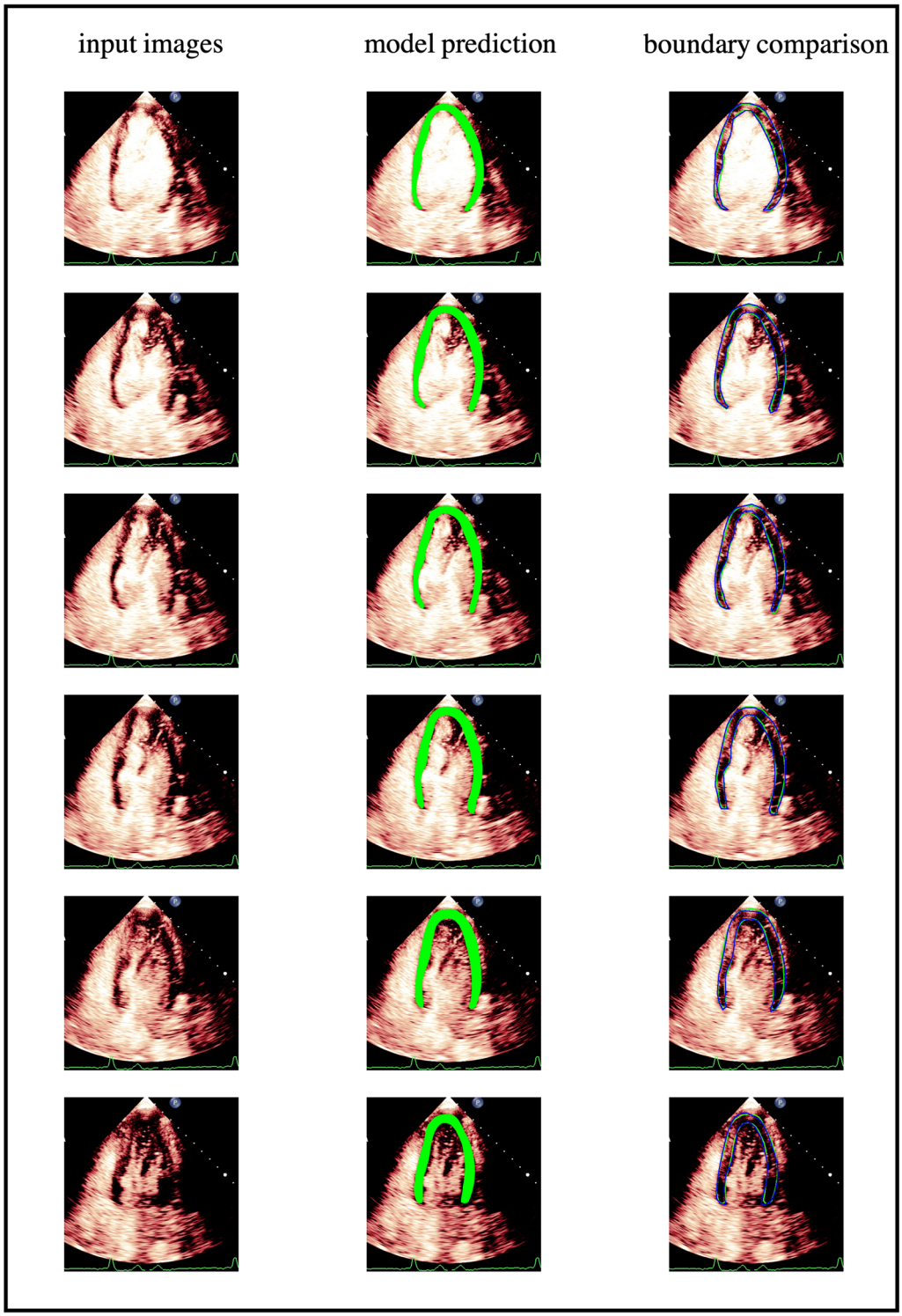

Figure 6. Randomly selected input images from test dataset, model prediction, and boundary comparison. (The blue and green outlines refer to the ground truth and prediction respectively.)

Moreover, we compared the modified DeepLabV3+ to the original DeepLabV3+, the results of Li et al. [17] and other state-of-the-art models, e.g., a U-net [25] with a Deeplabv3 backbone and PSPnet [24] with a ResNet-101 backbone; the results are shown in Table 2.

Table 2. Dice, IoU of modified, original DeepLabV3+, Li's model, PSPnet and U-net,where A2C represents apical two-chamber view, A3C represents apical three-chamber view, A4C represents apical four-chamber view.

	Modifed DeepLabV3+	Original DeepLabV3+	Li's model	PSPnet	U-net
Dice					
A2C	**0.84**	**0.84**	0.81	0.83	0.82
A3C	**0.84**	0.83	0.81	**0.84**	0.82
A4C	**0.86**	0.84	0.82	0.83	0.80
IoU					
A2C	**0.74**	0.72	0.69	0.73	0.69
A3C	**0.72**	0.71	0.65	**0.72**	0.70
A4C	**0.75**	**0.75**	0.71	0.72	0.72

The modified DeepLabV3+ improved the segmentation results for the dice coefficient in A3C by 0.01 and in A4C by 0.02, and for the IoU in A2C by 0.02 and A3C by 0.01. Moreover, the modified DeepLabV3+ also outperformed other state-of-the-art models in both metrics.

To see the trade-off between model performance and complexity, we also tested the model with different depths of ResNet, i.e., 18 and 50, and compared it with the PSPnet and U-net. The number of parameters and GFlops (giga floating-point operations per second) were selected to indicate the model complexity, and the average IoU of all chamber views was selected to evaluate the model performance. In addition, the MCE frame per second processed (FPS) was evaluated on a personal computer with the RTX 2060 super and an Intel® Core™ i5-9600 processor; the results are shown in Table 3 and the comparisons of the number of parameters and GFlops to the average IoU are illustrated in Figure 7 and Figure 8, respectively.

Table 3. Number of parameters, GFlops, Average IoU and FPS of different depths of the modified DeepLabV3+, PSPnet and U-net.

	No. of parameters	GFlops	Average IoU (%)	FPS
modified DeepLabV3 + (ResNet18)	12.47	54.21	70.94	39.6
modified DeepLabV3 + (ResNet50)	43.58	176.25	72.81	21.2
modified DeepLabV3 + (ResNet101)	62.68	255.14	74.23	15.2
PSPnet (ResNet101)	68.07	256.44	72.68	15.7
U-net	29.06	203.43	71.02	20.5

Although the parameter count of PSPnet (ResNet101) increased 36.87% and GFlops

increased 31.27% relative to the modified DeepLabV3+ (ResNet50), DeepLabV3+ (ResNet50) still outperformed PSPnet, which proves the efficiency of the proposed model. Comparing different ResNet backbone depths of the modified DeepLabV3+ and 18 depth only had 28.61% and 19.89% of the parameter count of the 50 depth and 101 depth, respectively, and 30.76% and 21.25% of GFlops of the 50 depth and 101 depth, respectively; it still had a 97.43% IoU for the 50 depth and 95.56% IoU for the 101 depth.

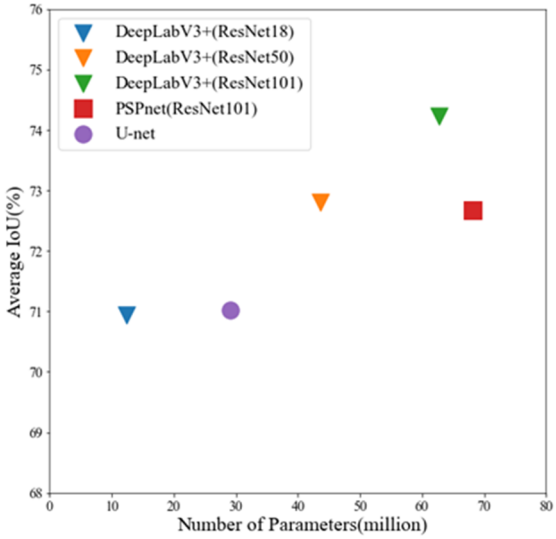

Figure 7. Number of parameters versus the average IoU of different models.

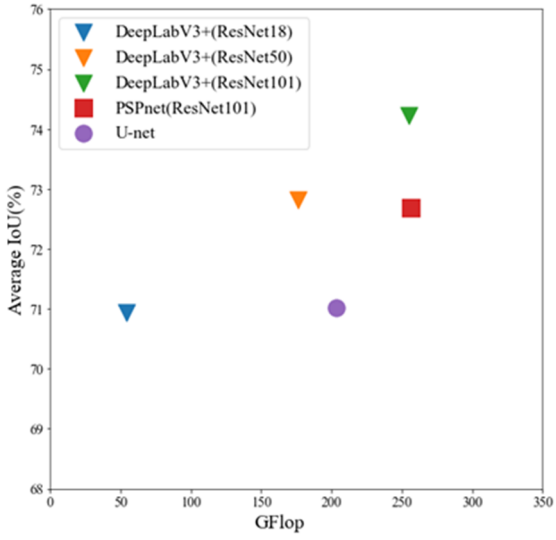

Figure 8. GFlop versus the average IoU of different models.

The feasibility of model application takes both performance and computational complexity into consideration; we believe balance can be made based on the depth of the backbone ResNet.

4. Conclusions

This paper proposed a modified architecture DeepLabV3+ model for MCE segmentation. The model consists of three main modules: the backbone, ASPP module and decoder. The backbone utilizes ResNet with atrous convolution, which allows the algorithm to find the best balance between the receptive field from a large field of view and the resolution of the feature map from a small field of view. The modified ASPP module also applies atrous convolution with different dilation factors to resample multi-scale patterns from the feature maps extracted from the backbone. The decoder combines the lower-level feature map from the encoder and high-level feature map from the decoder to generate the final prediction. A comparison between the proposed model and other state-of-the-art models, i.e., the PSPnet and U-net was conducted; the proposed model has achieved the best scores for both the dice and IoU. Moreover, we also did a performance and complexity analysis for all of the models, including the proposed model with different backbone depths, PSPnet and U-net; the results show the efficiency of the proposed architecture, and a comparison of the different depths of ResNet backbone illustrated the application feasibility of the proposed model. In the future study, we will focus on the balance of performance and complexity of the model to seek opportunities for the application in clinical analysis. Moreover, the Li et al. data only provides MCE frames rendered by a coloring mapping procedure; however, the rendering schema depends on settings from different companies and radiologists, so it will reduce the objectives and make the segmentation lose detail, which might bring variation to the algorithm's performance. Thus, we hope to experiment with our model on original MCE frames to improve the robustness and generalization ability in the future.

Acknowledgments

This work was supported by the Natural Science Foundation of China (NSFC) under grant number 62171408, and the Key Research and Development Program of Zhejiang Province (2020C03060, 2020C03016, 2022C03111).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. M. Dewey, M. Siebes, M. Kachelriess, K. Kofoed, P. Maurovich-Horvat, K. Nikolaou, et al., Clinical quantitative cardiac imaging for the assessment of myocardial ischaemia, *Nat. Rev. Cardiol.*, **17** (2020), 427–450. http://dx.doi.org/10.1038/s41569-020-0341-8
2. G. Q. Du, J. Y. Xue, Y. Guo, S. Chen, P. Du, Y. Wu, et al., Measurement of myocardial perfusion and infarction size using computer-aided diagnosis system for myocardial contrast echocardiography, *Ultrasound Med. Biol.*, **41** (2015), 2466–2477. https://doi.org/10.1016/j.ultrasmedbio.2015.04.012
3. Y. Guo, G. Q. Du, W. Q. Shen, C. Du, P. He, S. Siuly, Automatic myocardial infarction detection in contrast echocardiography based on polar residual network, *Comput. Methods Programs Biomed.*, **198** (2021), 105791. https://doi.org/10.1016/j.cmpb.2020.105791

4. J. Cho, S. Her, H. Youn, C. Kim, M. Park, G. Kim, et al., Usefulness of the parameters of quantitative myocardial perfusion contrast echocardiography in patients with chronic total occlusion and collateral flow, *Echocardiography*, **32** (2015), 475–482. https://doi.org/10.1111/echo.12663
5. T. Rutz, S. F. de Marchi, P. Roelli, S. Gloekler, T. Traupe, H. Steck, et al., Quantitative myocardial contrast echocardiography: a new method for the non-invasive detection of chronic heart transplant rejection, *Eur. Heart J. Cardiovasc. Imaging*, **14** (2014), 1187–1194. https://doi.org/10.1093/ehjci/jet066
6. A. L. Klibanov, P. T Rasche, M. S. Hughes, J. K. Wojdyla, K. P. Galen, J. H.Wible, et al., Detection of individual microbubbles of ultrasound contrast agents: imaging of free-floating and targeted bubbles, *Invest. Radiol.*, **39** (2004), 187–195. https://doi.org/10.1097/01.rli.0000115926.96796.75
7. M. X. Tang, H. Mulvana, T. Gauthier, A. Lim, D. Cosgrove, R. Eckersley, et al., Quantitative contrast-enhanced ultrasound imaging: a review of sources of variability, *Interface Focus*, **1** (2011), 520–539. http://doi.org/10.1098/rsfs.2011.0026
8. M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, *Int. J. Comput. Vision*, **1** (1988), 321–331. https://doi.org/10.1007/BF00133570
9. T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, Active shape models-their training and application, *Comput. Vision Image Understanding*, **61** (1995), 38–59. https://doi.org/10.1006/cviu.1995.1004
10. N. Malpica, M. Ledesma-Carbayo, A. Santos, E. Perez-David, M. Garcia-Fernandez, M. Desco, A coupled active contour model for myocardial tracking in contrast echocardiography, *Image Understanding and Anal.*, **2004** (2004).
11. J. E. Pickard, J. A. Hossack, S. T. Acton, Shape model segmentation of long-axis contrast enhanced echocardiography, in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro*, (2006), 1112–1115. https://doi.org/10.1109/ISBI.2006.1625117
12. Y. Guo, G. Du, J. Xue, R. Xia, Y. Wang, A novel myocardium segmentation approach based on neutrosophic active contour model, *Comput. Methods Programs Biomed.*, **142** (2017), 109–116. https://doi.org/10.1016/j.cmpb.2017.02.020
13. Y. Li, C. Ho, M. Toulemonde, N. Chahal, R. Senior, M. Tang, Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model, *IEEE Trans. Med. Imaging*, **37** (2017), 1081–1091. https://doi.org/10.1109/tmi.2017.2747081
14. N. Azarmehr, X. Ye, S. Sacchi, J. Howard, D. Francis, M. Zolgharni, Segmentation of left ventricle in 2D echocardiography using deep learning, in *Annual Conference on Medical Image Understanding and Analysis*, (2019), 497–504. https://doi.org/10.1007/978-3-030-39343-4_43
15. G. Veni, M. Moradi, H. Bulu, G. Narayan, T. Syeda-Mahmood, Echocardiography segmentation based on a shape-guided deformable model driven by a fully convolutional network prior, in *2018 IEEE 15th International Symposium on Biomedical Imaging*, (2018), 898–902. https://doi.org/10.1109/ISBI.2018.8363716
16. Y. Hu, L. Guo, B. Lei, M. Mao, Z. Jin, A. Elazab, et al., Fully automatic pediatric echocardiography segmentation using deep convolutional networks based on BiSeNet, in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2019), 6561–6564. https://doi.org/10.1109/EMBC.2019.8856457

17. M. Li, D. Zeng, Q. Xie, R. Xu, Y. Wang, D. Ma, et al., A deep learning approach with temporal consistency for automatic myocardial segmentation of quantitative myocardial contrast echocardiography, *Int. J. Cardiovasc. Imaging*, **37** (2021), 1967–1978. https://doi.org/10.1007/s10554-021-02181-8
18. L. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentationm, preprint, arXiv:1802.02611.
19. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, preprint, arXiv:1512.03385.
20. T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, M. Li, Bag of tricks for image classification with convolutional neural networks, preprint, arXiv:1812.01187.
21. F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, preprint, arXiv:1511.07122.
22. L. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, preprint, arXiv:1706.05587.
23. F. Milletari, N. Navab, S. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *2016 fourth international conference on 3D vision (3DV)*, (2016), 565–571. https://doi.org/10.1109/3DV.2016.79
24. H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, preprint, arXiv:1612.01105.
25. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)