Echocardiographic Left Ventricular Segmentation Using Double-layer Constraints on Spatial Prior Information

Original Scientific Paper

Jin Wang
[1]College of Computing,
Informatics and Mathematics, Shah Alam, Malaysia
[2]Department of Electrical Engineering,
Taiyuan Institute of Technology, Taiyuan, China
wangjin@studyedus.cn

Sharifah Aliman
Universiti Teknologi MARA,
College of Computing, Informatics and Mathematics
Shah Alam, Malaysia
sharifahali@uitm.edu.my

*Corresponding author

Shafaf Ibrahim*
Universiti Teknologi MARA,
College of Computing, Informatics and Mathematics
Shah Alam, Malaysia
email: shafaf2429@uitm.edu.my

Yanli Tan
Universiti Teknologi MARA,
College of Computing, Informatics and Mathematics
Shah Alam, Malaysia
email: tanyanli@studyedus.cn

Abstract *– Real-time segmentation of echocardiograms is of great practical significance for doctors' clinical diagnosis. This paper addresses the existing echocardiogram segmentation models' pursuit of high segmentation accuracy in insufficient training data, which leads to high model complexity and low learning efficiency. This paper fully exploits the spatial prior characteristics of the image itself. It proposes an echocardiographic left ventricular segmentation algorithm that utilizes double-layer constraints of prior information on spatial anatomical structures. The algorithm is based on the following two principles. Firstly, the segmentation model is initialized using a self-supervised sorting model based on the spatial anatomy to fully learn the orderly image features of the left ventricular spatial anatomy and achieve same-domain transfer of images, allowing the segmentation network to learn segmentation information more effectively; Secondly, the segmentation network is subjected to mask shape constraints, and the output space is limited by imposing anatomical shape priors to expand the global training goals of the CNN model. Finally, the algorithm proposed in this paper was verified using three classic segmentation models. The experimental results showed that on the public echocardiography dataset CETUS (Challenge on Endocardial Three-dimensional Ultrasound Segmentation), compared with the classic Resnet, Unet, and VGG segmentation models, the double-layer constrained segmentation model that introduces prior features has increased the segmentation accuracy (Dice index) by 5.6%, 4.9%, and 4.8%, respectively. The MIOU (Mean Intersection over Union) index increased by 7%, 5.5%, and 6.8%, respectively, demonstrating robustness to slice misalignment.*

Keywords: *echocardiographic segmentation, deep learning, spatial prior, CETUS*

Received: January 20, 2025; Received in revised form: May 1, 2025; Accepted: May 2, 2025

1. INTRODUCTION

Due to the portability, cost-effectiveness, non-radiation, and real-time nature of echocardiography, accurate segmentation of the left ventricle from ultrasound images can help doctors with less clinical experience to analyze cardiac images conveniently and accurately to serve actual clinical diagnosis [1]. However, due to the ultrasonic imaging mechanism, echocardiography has characteristics such as considerable dynamic noise, low image contrast, and loss of edges [2]. This makes achieving fully automated real-time segmentation of the left ventricle in echocardiography a well-known challenge. In recent years, the most advanced deep learning technology has been used for cardiac image segmentation to automatically measure size and functional assessment of the left ventricle, effectively improving the diagnostic efficiency of echocardiography

[3, 4]. However, it also faces some limitations. For example, deep learning networks depend on the learning capabilities and results of a large amount of annotated data and powerful storage computing units; there are currently very few publicly available datasets, and the scale is difficult to meet research needs.

To solve this problem, some researchers have proposed deep network fusion algorithms to improve segmentation accuracy and convergence speed, especially when training datasets are limited. Literature [5-8] combines deep learning networks with deformable models, and features extracted by trained deep neural networks are used instead of handcrafted features to improve accuracy and robustness. Literature [9] proposed a method combining convolutional neural networks and ASM (Active Shape Model) to achieve automatic segmentation of the left ventricle of echocardiograms. It uses the Nakagami distribution to integrate the shape prior of the image to provide preprocessing classification. The results show that the segmentation accuracy and convergence are improved at the same time. Literature [10-11] uses generative adversarial networks to make segmentation masks, and image frame structures correspond one-to-one, increasing the number of training samples and improving segmentation accuracy. Literature [12] fuses two convolutional neural networks, YOLOv7 and U-Net, to automatically segment echocardiographic images. Some researchers have effectively utilized unlabeled data and proposed semi-supervised and unsupervised deep learning methods to improve the segmentation performance of the model by combining multiple strategies [13-16].

At this stage, deep network fusion algorithms have performed well in left ventricular segmentation tasks on ultrasound cardiac images. Algorithms that introduce prior information about intensity, shape, time, topology, and atlas show obvious advantages in improving the accuracy and efficiency of segmentation. However, most deep learning networks are based on feature classification of pixel sets, ignore the structural characteristics and related prior knowledge of echocardiograms, and lack the learning of global features related to segmentation target structures, resulting in limited feature learning capabilities of the model. Some researchers realize the importance of prior knowledge, such as image anatomy and imaging information, and try to utilize prior features better to optimize deep learning models. Literature [17] incorporates the perceptual similarity information between the generated and original frames into the segmentation model as prior knowledge. It uses unlabeled data for semi-supervised learning to improve segmentation performance. Literature [18] introduces a prior information encoding module, and the results show that the accuracy of this method is close to the segmentation result of the current optimal nnU-Net, with the convergence speed increased by 145%. Literature [19] proposes a Unet network model (MCCT-Unet) based on a multi-channel cross-fusion transformer. By effectively combining deep information with shallow information in the encoding stage, the segmentation performance of the network is improved. Literature [20] constructs a multi-fusion residual attention U-Net (MURAU-Net) automatic segmentation model by strengthening the connection of spatial features. Literature [21] introduced spatial and temporal prior features and achieved excellent segmentation results through deep network fusion. These research results demonstrate the effectiveness of introducing prior left ventricular anatomical structure features in improving the deep network fusion algorithm's segmentation accuracy and convergence speed.

This paper proposes a segmentation algorithm for the left ventricle in echocardiography using double-layer constraints on the prior information of spatial anatomy. The algorithm uses the orderliness of the anatomical position of the left ventricle to construct a self-supervised sorting model to initialize the segmentation network. The shape prior is incorporated into the mask part of the segmentation network to constrain the output space, reduce the extraction depth of the feature layer of the segmentation network, and improve the segmentation performance. Experimental results show that with relatively limited training data, the model has achieved significant improvements in both the Dice and MIOU indicators of segmentation accuracy, fully verifying its excellent performance and practicality. Specifically, the model brings the following benefits: (1) By utilizing the strong correlation between different positions of the image simultaneously, efficient model pre-training is achieved based on same-domain transfer, effectively solving the problem of insufficient generalization ability in different domain transfer learning. (2) By analyzing the imaging characteristics of echocardiography and the anatomical structure of the left ventricle and taking advantage of the natural order of the short-axis section of the left ventricle in spatial position, a self-supervised sorting model is constructed, aiming to explore a reasonable model initialization method to improve the performance and efficiency of the model. (3) The short-axis section of the left ventricle presents a fixed position relationship from top to bottom at any time in the cardiac cycle. This spatial anatomical prior knowledge is not only applicable to echocardiography. Still, it can also be extended to image segmentation of other modalities, providing a new way to solve the training requirements of medical image segmentation problems.

2. METHOD

The overall framework of the echocardiographic left ventricular segmentation algorithm using double-layer constraints of prior information on spatial anatomy is shown in Fig. 1. It mainly consists of two parts: a self-supervised ranking model and a shape-constrained image segmentation model. The self-supervised rank-

ing model aims to learn the anatomical prior features of the image. It uses pre-training of the self-supervised ranking model to initialize the segmentation network. It encourages the model to learn more about the segmentation task by obtaining anatomical position features when training the segmentation task. This information helps improve the prediction accuracy and convergence speed of segmentation models. The shape-constrained image segmentation model incorporates the shape prior of the label structure into the deep learning network. In this way, by constraining the training process of the neural network, the network is guided to make more anatomically meaningful predictions. This study uses the prior spatial structure features and integrates high-dimensional and low-dimensional features into the deep learning segmentation network simultaneously. It is expected to achieve better segmentation results, reduce the segmentation network learning model's complexity, and reduce the scale of the training dataset.

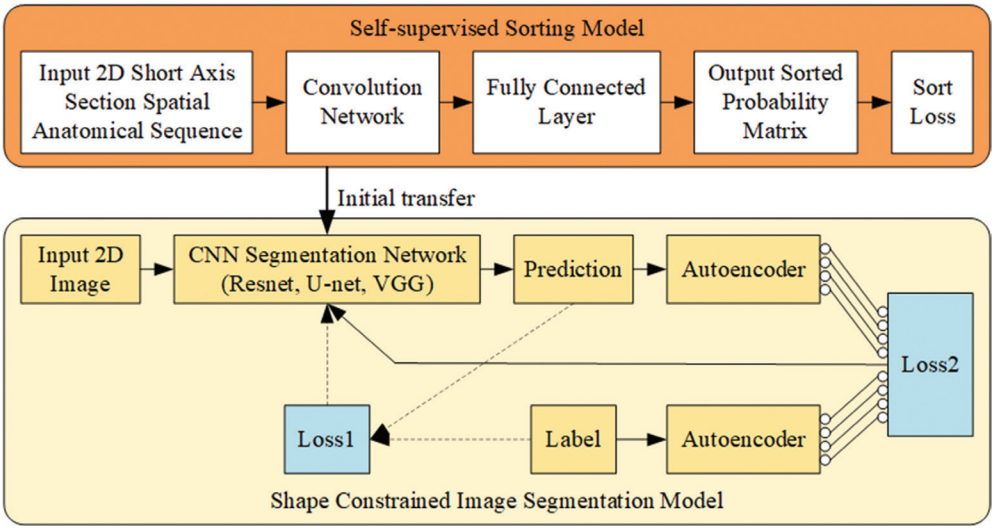


Fig. 1. Overall algorithm framework

2.1. IMAGE DESCRIPTION

The dataset for the pre-training model and the shape-constrained segmentation model are both from the public CETUS dataset. CETUS comprises 45 3D echocardiography sequences, which are evenly distributed in three different subgroups: healthy subjects, patients with muscle damage, and patients with dilated cardiomyopathy, and has been widely verified for its superior algorithm [22]. First, the 3D volume data is sliced along the short axis to obtain 2D slices. Second, 2D slices of the left ventricle from the apex to the base were obtained manually, and other parts were removed.

Finally, the acquired left ventricular short-axis slice sequence is normalized to ensure that all data have the exact spatial resolution along the short axis, and the left ventricular short-axis slice data are sampled according to the sorting scale, 10,658 2D short-axis slice images were obtained, of which 8,276 were used for training and 2,382 for testing.

The self-supervised sorting pre-training model uses slice sequences as input data. The input sequence is randomly shuffled to prevent the problem of the absolute position of the left ventricular 2D slice feature map. The shape constraint segmentation model uses a single 2D slice as input for subsequent segmentation tasks.

2.2. SELF-SUPERVISED RANKING MODEL BASED ON SPATIAL PRIOR

The puzzle problem trains a deep learning network to identify the components of the target [23]. This paper analyzes echocardiography's imaging characteristics and the left ventricle's short-axis structure based on this concept. Its spatial anatomy resembles a cone, with these slices appearing in a fixed order from top to bottom in spatial position. This paper takes advantage of the natural spatial ordering of short-axis slices of the left ventricle to build a self-supervised sorting pre-training model, initializes the parameters of the segmentation network, and effectively integrates spatial anatomy prior knowledge into the Segmentation network for deep learning. As shown in Fig. 2, the self-supervised sorting pre-training model includes four parts: input, backbone feature extraction, output, and loss module. The input is an echocardiographic left ventricular short-axis slice aerial anatomical structure sequence, a set of short-axis slices from the apex to the base. The number of slices N defines the sorting scale ($20>N>2$), based on the selected N Slices, with different sorting for different inputs. The backbone feature extraction module is a general convolutional network structure. In this paper, three classical structures, VGG [24], Unet [25], and Resnet [26], are chosen for the performance comparison of self-supervised sorting models. The output is an $N*N$-dimensional probability matrix, and the loss module is mainly used to evaluate the accuracy of N-slice sorting.

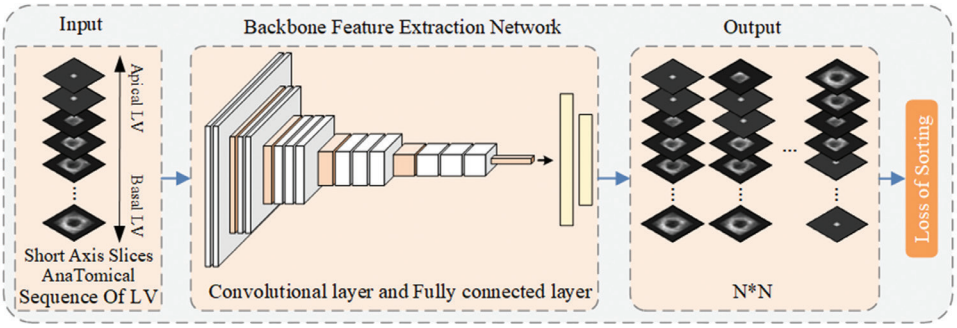


Fig. 2. Self-supervised ranking pre-training model based on spatial prior

The loss module is mainly used to evaluate the sorting accuracy of echocardiographic left ventricular short-axis section image sequences. The optimization goal of the spatial anatomical structure self-supervised sorting task is the multi-level Softmax loss function. The specific loss function is as follows:

$$\text{Loss} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{N} y_{ji} p_{ji} \tag{1}$$

Among them, the formula of p_{ji} is as follows:

$$p_{ji} = \frac{e^{Z_{ji}}}{\sum_{c}^{N} Z_{jc}} \tag{2}$$

N represents the number of categories, y_{ji} p_{ji} indicating that the j-th image belongs to the i-th category, y_{ji}=1 indicating that the j-th image belongs to the i-th category, y_{ji}=0indicating that the j-th image does not belong to the i-th category, p_{ji} indicating that the input j-th image belongs to the i-th category Probability, Z represents the input of the softmax activation function.

2.3. SHAPE-CONSTRAINED IMAGE SEGMENTATION MODEL

The shape-constrained segmentation model implements the shape constraint function by adding convolutional autoencoding, applying anatomical shape priors to the predicted images of the segmentation model to constrain the output space, expanding the global training goals of the CNN segmentation model, and using two loss functions to adjust the feedback of the segmentation network. This approach improves sub-pixel segmentation accuracy by training an upsampling layer with high-resolution ground truth maps.

Fig. 3 shows the structure of the shape-constrained segmentation model. The convolutional autoencoding constraint model AE (autoencoder) [27] is integrated into the basic segmentation network based on deep learning and predicts image class label shape constraints. It fully uses the anatomical low-dimensional features of 2D echocardiographic left ventricular images to improve model segmentation accuracy.

The basic segmentation network uses a cross-entropy loss function to run predictions at the single-pixel level, since the backpropagation gradient is only parameterized by the individual probability divergence term at the pixel level, it provides little global context. It cannot guarantee the consistency of the overall anatomical shape. Class label prediction obtains the parameters and underlying structure of the lower-dimensional segmentation by performing AE-based nonlinear low-dimensional projections of the predicted image and the true label [28]. This paper builds a segmentation network with a double-constraint loss function to obtain more global information and local features, thereby improving the performance of the segmentation model.

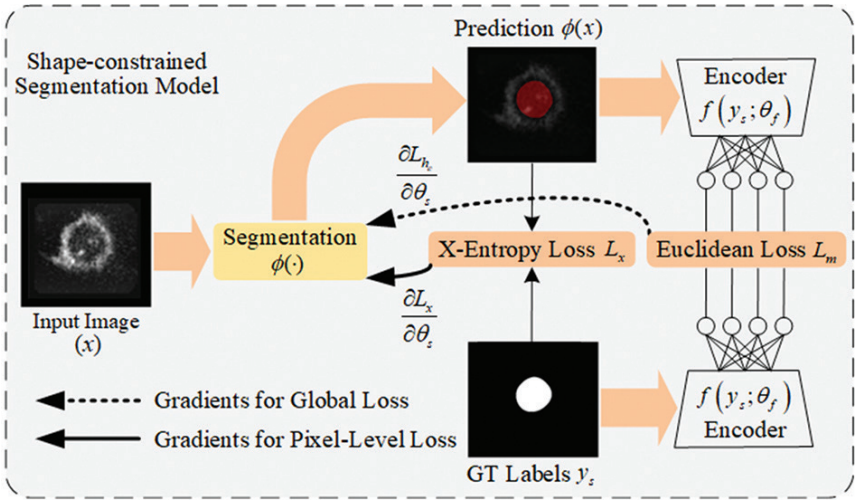


Fig. 3. Shape-constrained segmentation network model

The shape Constrained Segmentation Network via Cross-Entropy Loss of Basic CNN Segmentation Network L_x (ϕ(x; θ_s), y_s)and the linear combination of the shape loss L_m from AE to train the objective function, as shown in Equation 3. ω is the weight of the convolution filter of the segmentation network. The third term corresponds to weight decay, which limits the number of free parameters in the model to avoid over-fitting, weight decay to restrict the number of free parameters in the model to avoid overfitting. θ_s represents all trainable parameters of the segmentation model, θ_f represents all trainable parameters of the AE model, which are updated during training. The coupling parameters λ_1 and λ_2 determine the weight of the shape loss and the weight decay terms used in the training. In this equation, the second term L_m ensures that the generated segmentations are in a similar low-dimensional space as the ground-truth labels.

$$L_m = \left\| f\left(\phi(x);\theta_f\right) - f\left(y;\theta_f\right)\right\|_2^2$$
$$\min_{\theta_s}\left(L_x\left(\phi\left(x;\theta_s\right),y_s\right) + \lambda_1 \cdot L_{h_e} + \frac{\lambda_2}{2}\|\omega\|_2^2 \right) \quad (3)$$

2.4. PERFORMANCE EVALUATION

To measure the accuracy of echocardiographic left ventricular segmentation, this paper used three different metrics, namely Dice, two-dimensional HD (Hausdorff Distance), and MIOU to evaluate the segmentation accuracy [29][30][31]. Let U={u_1, u_2,…u_m} be the prediction area, and R={r_1, r_2,…r_m } 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 and generally requires the index to be greater than 0.7 for the segmentation effect to be considered relatively good.

$$\text{Dice}=\frac{2|U\cap R|}{|U|+|R|} \quad (4)$$

HD is the maximum distance from one set to the nearest point in another set. This distance is directional, meaning that $h_{(U, R)}$ is not equal to $h(_{R, U})$. H takes the larger of the two distances. A smaller value indicates a higher degree of similarity for parameters sensitive to differences in location information. The calculation formula is as follows:

$$h(\text{R,U}) = \max_{u\in U}\left\{\min_{r\in R}\|\text{u-r}\|\right\} \quad (5)$$

$$h(\text{U,R}) = \max_{r\in R}\left\{\min_{u\in U}\|\text{r-u}\|\right\} \quad (6)$$

MIOU is the average intersection and union ratio, including the heart and background areas. IOU is used to test the overlapping area of each category, calculated as the intersection area of a specific category divided by the union area of a particular category. The MIOU is calculated as the sum of the Io of all categories divided by the total number of categories.

$$\text{MIOU}=\frac{1}{2}\times\left(\frac{n_{ff}}{t_f+n_{bf}}+\frac{n_{bb}}{t_b+n_{fb}}\right) \quad (7)$$

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

3. RESULTS AND DISCUSSION

During model training, the classic CNN network structures, including VGG [24], Unet [25], and Resnet [26], were selected for medical image segmentation tasks. The echocardiographic left ventricle segmentation algorithm using double-layer constraints of spatial prior information was verified to be effective. The model was implemented using the PyTorch deep learning framework, with Kaggle selected as the running platform.

3.1. RESULTS AND ANALYSIS OF SELF-SUPERVISED SORTING PRE-TRAINING MODEL

First, the feasibility of the self-supervised ranking model based on spatial priors for different deep-learning networks was verified under various input image sequence sizes. During the self-supervised model training process, the hyperparameters Epoch=20, batch size=16, and learning rate=1e-5 were set, with all parameter size limits chosen based on tracking and error to provide higher accuracy. VGG [24], Unet [25], and Resnet [26] were used as the basic network structures for deep learning, and the performance of the self-supervised model was evaluated through average ranking accuracy. Table 1 shows the ranking accuracy of the self-supervised ranking model on the test set using different basic network structures and input image sequence sizes. Experimental results indicate that for sorting tasks with a sorting vector of less than 10, the sorting model can achieve an accuracy higher than 50%, which validates the results to a certain extent. This paper illustrates the rationale behind constructing a feasible sorting task for a self-supervised sorting model based on spatial anatomical priors. It demonstrates that developing a self-supervised sorting model has significant potential for application in pre-training left ventricular segmentation tasks.

Table 1. Accuracy of different deep learning network structures

Method	Accuracy at different sorting scales								
	2	3	4	5	6	7	8	9	10
Unet	0.80	0.77	0.72	0.71	0.70	0.68	0.62	0.55	0.53
VGG	0.85	0.80	0.76	0.74	0.73	0.64	0.62	0.61	0.58
ResNet	0.88	0.85	0.83	0.80	0.73	0.70	0.64	0.63	0.61

Next, the effect of self-supervised ranking based on spatial anatomy priors on the pre-trained segmentation model is verified. The Models used for comparison and verification include: Resnet, Resnet_S; Unet, Unet_S; VGG, VGG_S, where S represents self-supervised sorting based on spatial anatomy prior, which is used for the pre-training of segmentation models.

Given the performance analysis results of the pre-training model, the input scale of the ranking model in the experiment utilized 8 2D image slices, and the ranking output was an 8×8 probability matrix. After the training of the ranking task is completed, the model that performed best on the test set is selected, and the segmentation network is initialized. The ranking model and segmentation network use the same base network to facilitate simple and effective model parameter migration. The initial learning rate (Lr) is set to 0.01, the minimum learning rate is 1e-5, the optimizer used is AE, the batch size is 8, and the limits of all parameter sizes are selected based on tracking and error to improve accuracy. The model is evaluated through the Loss_epoch curve of the training set, and the experimental results are shown in Fig. 4.

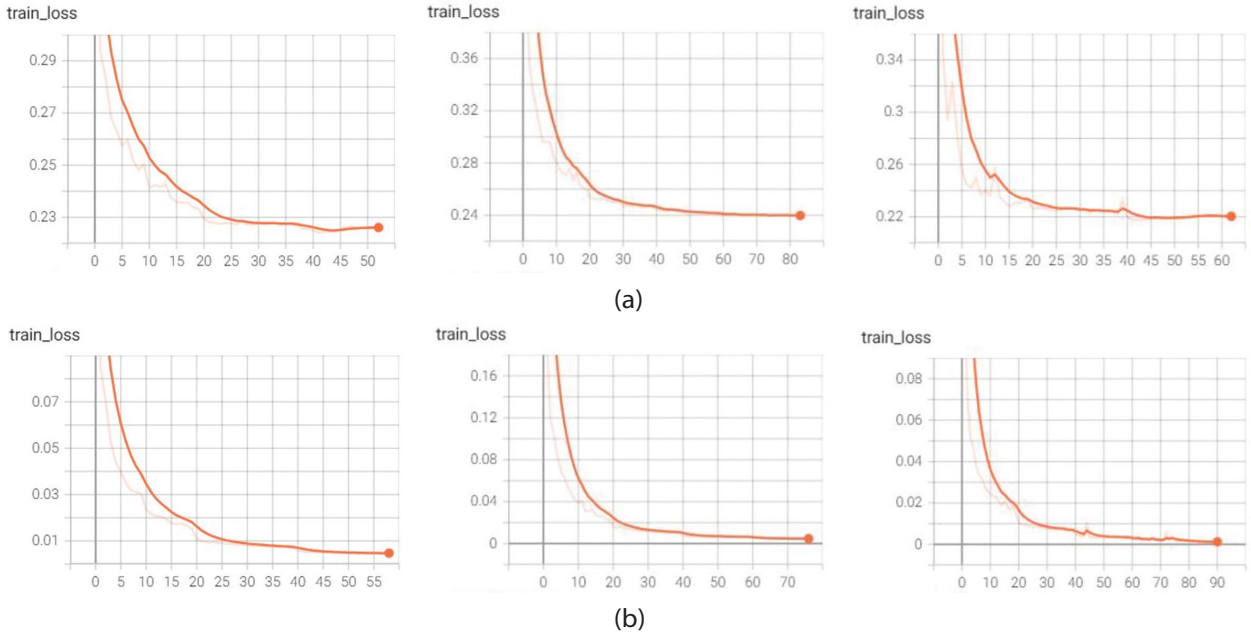


Fig. 4. Loss_epoch curves at different model training stages. (**a**) Resnet(loss-epoch), Unet(loss-epoch) and VGG(loss-epoch), (**b**) Resnet_S(loss-epoch), Unet_S(loss-epoch) and VGG_S(loss-epoch)

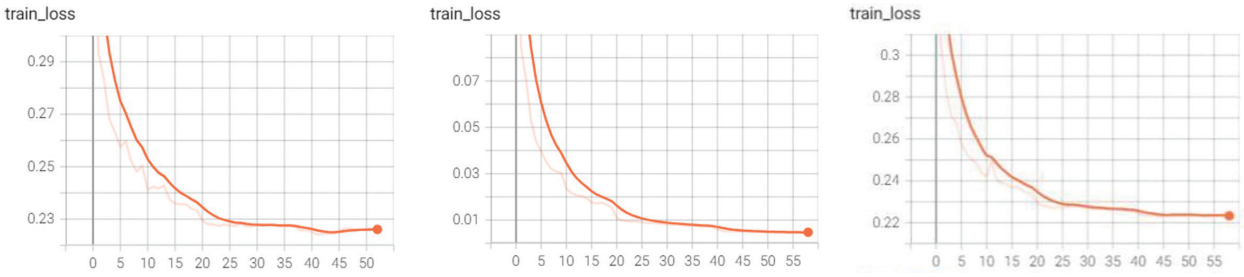
Fig. 4 shows the Loss-Epoch curves of different base network model training stages. By comparison, it is found that the segmentation models Resnet_S, Unet_S, and VGG_S, which are based on spatial anatomy prior self-supervised sorting pre-training, exhibit faster convergence capabilities. The reason is that during pre-training, the model learns effective prior information naturally ordered in the spatial dimensions of left ventricular ultrasound images. When the segmentation task training is completed, it will learn more task-related information and perform segmentation based on the learned prior features, which helps improve segmentation accuracy and speeds up training convergence.

3.2. RESULTS AND ANALYSIS OF THE DOUBLE-LAYER CONSTRAINT SEGMENTATION NETWORK MODEL

First, the impact of adding anatomical constraints on the convergence performance of the segmentation network was verified during the model training phase. Resnet was selected as the representative backbone network for evaluation in the experiment, and the performance of the segmentation network before and after the introduction of anatomical constraints was comparatively analyzed. Here, S represents the use of pre-training, and L represents the introduction of anatomical constraints. The Model parameter settings included a batch size of 10 and a learning rate of 2e-4, which was reduced to 1e-5 in the later stage of training. The loss function used was the cross-entropy loss function. Fig. 5 shows the Loss_epoch curve of the model training stage represented by the Resnet base network. It was found that pre-training effectively improves the convergence speed of the model. After the anatomical constraints are introduced, since two losses limit the model, it increases the learning difficulty of the model, making the convergence speed slower and consistent with the convergence of the model without pre-training.

Next, a comprehensive performance evaluation of the double-constraint left ventricular segmentation model based on self-supervised sorting pre-training proposed in this paper was conducted from two aspects: segmentation accuracy and model segmentation effect. The model's accuracy is evaluated through three indicators: Dice, HD, and MIOU. The experiment

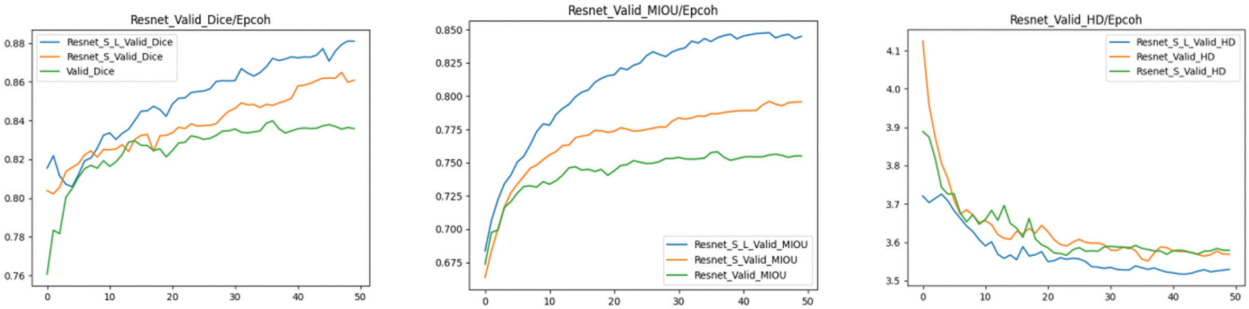
was implemented using the PyTorch deep learning framework, and the running platform used was Kaggle. The test set comprises 2382 2D slices from 36-45 patients in the CETUS dataset. 9 segmentation models, including Resnet, Resnet_S, Resnet_S_L, Unet, Unet_S, Unet_S_L, VGG, VGG_S, and VGG_S_L were tested. The Segmentation effects were intuitively assessed through qualitative visual comparisons of different image qualities, segmentation masks produced by different models, and corresponding ground truth values.



(a) (b) (c)

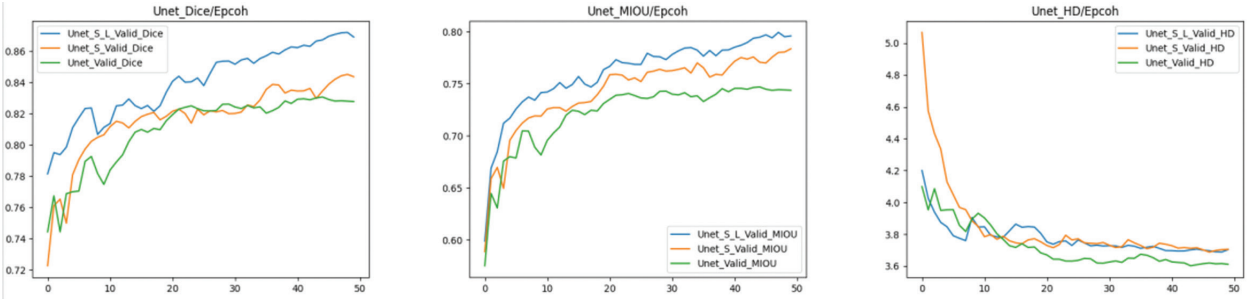
Fig. 5. Loss-epoch curves of Resnet, Resnet_S, and Resnet_S_L models. (**a**) Resnet(Loss-epoch), (**b**) Resnet_S(Loss-epoch), (**c**) Resnet_S_L(Loss-epoch)

Three pre-training models, Resnet_S, Unet_S, and VGG_S, refer to the performance analysis results of the pre-training models, using eight image slices as input, with a batch size of 16, a learning rate of 1e-5, and an epoch count of 10. The loss function used is multi-level Softmax. After completing the self-supervised sorting task, the model with the best performance on the test set is selected as the pre-training model for the segmentation task. The parameters for the three basic network models of Resnet, Unet, and VGG are set with a batch size of 10 and a learning rate of 2e-4, which is reduced to 1e-5 in the later stage of training, with an epoch count of 50. The loss function used is the cross-entropy loss function. Based on anatomical prior pre-training, the shape-constrained segmentation models Resnet_S_L, Unet_S_L, and VGG_S_L proposed in this paper incorporate a linear combination of shape-constrained Loss1 and Loss2. The objective function is trained using a linear combination of the cross-entropy loss of the basic CNN segmentation network and the AE shape loss. The experimental results are shown in Fig. 6, Fig. 7, Fig. 8, and Table 2.



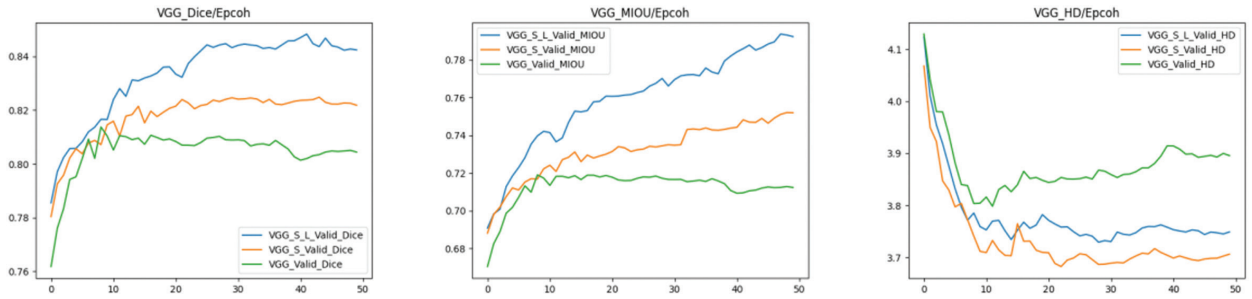
(a) (b) (c)

Fig. 6. The performance of Resnet, Resnet_S, and Resnet_S_L models using (**a**) Dice, (**b**) MIOU, and (**c**) HD methods



(a) (b) (c)

Fig. 7. The performance of Unet, Unet_S, and Unet_S_L models using (**a**) Dice, (**b**) MIOU, and (**c**) HD methods



(a) (b) (c)

Fig. 8. The performance of VGG, VGG_S, and VGG_S_L models using (**a**) Dice, (**b**) MIOU, and (**c**) HD methods

Fig. 6, Fig. 7, and Fig. 8 visually show the changes in test indicators of the model. By comparison, it is found that the Resnet_S, Unet_S, and VGG_S models based on self-supervised pre-training have significantly higher accuracy than the randomly initialized Resnet, Unet, and VGG basic segmentation models. In the sorting task, the model learns many basic features, such as the spatial anatomical structure of the image. This effective prior information is suitable for migrating the weights of the segmentation model, allowing it to learn based on the acquired prior features once the segmentation model training is completed. More information related to segmentation tasks can enhance model accuracy and speed up training convergence. Resnet_S_L, Unet_S_L, and VGG_S_L, which incorporate shape constraints, have further improved segmentation accuracy compared to Resnet_S, Unet_S, and VGG_S. The segmentation network learns low-dimensional position and shape information using sticky note shape constraints, significantly improving segmentation accuracy.

Table 2. Accuracy of different segmentation models

Accuracy	Resnet	Resnet_S	Resnet_S_L	Unet	Unet_S	Unet_S_L	VGG	VGG_S	VGG_S_L
Dice	0.827	0.843	0.874	0.813	0.827	0.853	0.805	0.821	0.844
MIOU	0.746	0.773	0.816	0.725	0.747	0.765	0.714	0.734	0.772
HD	3.617	3.605	3.565	3.796	3.774	3.759	3.871	3.761	3.727

Table 2 shows the segmentation results of the model on the test set. The Dice index for Resnet_S is improved by 1.8% compared to Resnet, Unet_S is improved by 1.7% compared to Unet, and VGG_S is improved by 1.9% compared with VGG; the MIOU index for Resnet_S is improved by 3.6% compared with Resnet, and Unet_S is 3% higher than Unet, and VGG_S is 2.8% higher than VGG; The HD parameters have not changed significantly. These results fully demonstrate that the self-supervised sorting tasks can be used to pre-train deep learning-based segmentation tasks. A double-layer constrained segmentation model using spatial prior information, the Dice index for Resnet_S_L further improved by 3.7% compared to Resnet_S, Unet_S_L further improved by 3.9% based on Unet_S, and VGG_S_L improved by 2.8% based on VGG_S; the MIOU index for Resnet_S_L further increased by 5.6% compared to Resnet_S, Unet_S_L further increased by 2.4% based on Unet_S, and VGG_S_L further increased by 5.2% based on VGG_S. The model constructed in this paper shows excellent segmentation performance, with Dice and MIOU indicators as high as 0.874 and 0.816, respectively, but it still has significant potential for performance improvement. Future research will focus on optimizing the segmentation network, incorporating spatial prior information, introducing cutting-edge attention mechanisms, integrating multi-scale feature fusion technology, and deep mining contextual information. These advancements are expected to improve the Dice and MIOU indicators, ensuring the model provides more accurate and reliable segmentation results across various image segmentation application scenarios.

Fig.9 shows the segmentation experimental results of each model under different image qualities. Comparing the segmentation masks produced by various models against the corresponding ground truth values allows for an intuitive evaluation of each model's segmentation performance. The results show that the Resnet_S, VGG_S, and Unet_S models based on self-supervised sorting pre-training perform better than the randomly initialized Resnet, VGG, and Unet models in the echocardiography left ventricle segmentation task. The pre-trained models showed satisfactory segmentation results in the face of segmentation challenges such as artifacts, speckle noise, and blurred anatomical boundaries. When evaluating the two key indicators of boundary accuracy and region overlap, the segmentation results of these models matched the ground truth values very well, significantly outperforming the basic segmentation network models. However, the Resnet_S_L, VGG_S_L, and Unet_S_L models that further incorporated the double-layer prior information constraints of the shape mask did not show significant performance improvements.

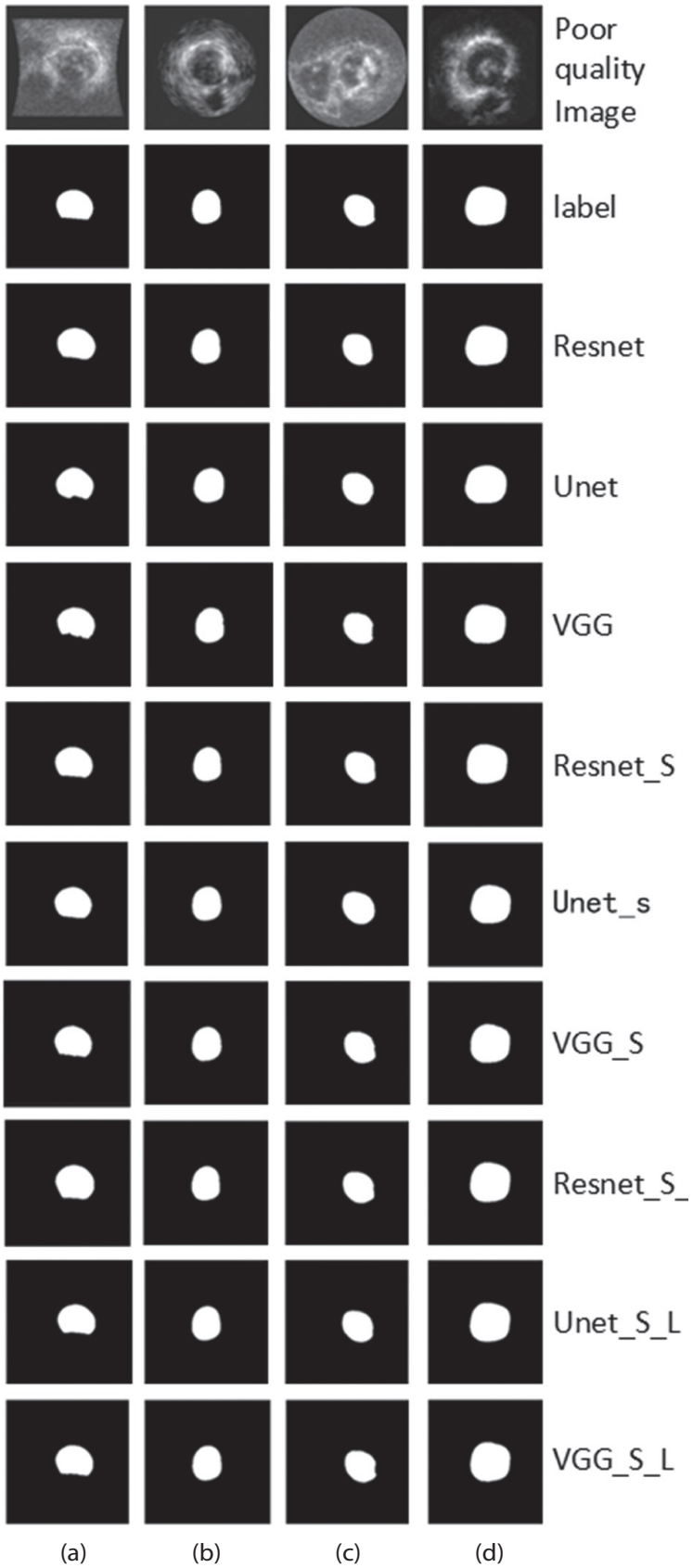


Fig. 9. Echocardiographic left ventricular segmentation renderings in different scenarios

The segmentation model based on self-supervised sorting pre-training uses a sequence of adjacent slices as input. During the pre-training process of the model, the basic structural features of the left ventricular image are learned. This prior information has a high reuse potential in the segmentation task, which helps to improve the segmentation accuracy and accelerate the convergence of the model. It is worth noting that fur-

ther integrating shape-constrained AE on top of these pre-trained models did not show significant performance improvements. The reason is that, on the one hand, the spatial anatomical prior of the image in the pre-training stage has already implied some shape information, thus weakening the gain effect brought by the additional shape constraint.

On the other hand, the AE model is mainly trained based on the left ventricle segmentation mask to capture the anatomical variation of the left ventricle accurately. Considering that the shape variation of the heart's left ventricle is relatively limited in the public dataset used in this study, the shape constraint of the second layer improves the performance. This improvement is more reflected in the subtle optimization of the existing performance. It fails to achieve the significant improvement brought by pre-training.

4. CONCLUSION

To solve the problem that the fully supervised segmentation algorithm for the left ventricle in echocardiography needs to deepen the network learning depth to improve segmentation accuracy due to insufficient training data, this paper constructs a segmentation model that integrates image prior information to achieve an effective combination of low-dimensional and high-dimensional features, as well as global and local features. On the one hand, through self-supervised sorting pre-training of the left ventricle from apex to base based on the spatial anatomical structure, the weights of the segmentation model are initialized, so that the model can fully obtain more local information related to the segmentation when performing the segmentation task, thereby improving the segmentation accuracy and accelerating the convergence speed. On the other hand, the segmentation network model is implemented to predict the shape constraints of image class labels, and the anatomical low-dimensional features of the left ventricle image of the two-dimensional echocardiogram are used to capture more global information and further improve the segmentation accuracy of the model. The model can extract basic features from related images, which are common to similar image analysis tasks, and can improve the performance of subsequent tasks.

Future research will continue to explore the sorting relationships implied by other spatial anatomical prior knowledge and try to model these relationships into a self-supervised sorting framework, study the specific impact of different sorting modes on the performance of the segmentation model, and find better sorting strategies. In addition, by optimizing the data input method, the sorting model can learn richer knowledge, significantly shorten the model training time, and improve the sorting model's learning effect and generalization ability, providing more solid technical support for applications in medical image processing.

5. ACKNOWLEDGMENT

The study was supported by the Ministry of Higher Education Malaysia (MoHE) and Universiti Teknologi MARA through the Fundamental Research Grant Scheme (FRGS/1/2022/ICT02/UITM/02/2).

6. REFERENCES

[1] R. M. Lang et al. "Recommendations for Cardiac Chamber Quantification by Echocardiography in Adults: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging", European Heart Journal - Cardiovascular Imaging, Vol. 16, No. 3, 2015, pp. 233-271.

[2] J. Zhang et al. "Fully Automated Echocardiogram Interpretation in Clinical Practice", Circulation, Vol. 138, No. 16, 2018, pp. 1623-1635.

[3] M. Balasubramani, C.-W. Sung, M.-Y. Hsieh, E. P.-C. Huang, J.-S. Shieh, M. F. Abbod, "Automated Left Ventricle Segmentation in Echocardiography Using YOLO: A Deep Learning Approach for Enhanced Cardiac Function Assessment", Electronics, Vol. 13, No. 13, 2024, p. 2587.

[4] S. Ferraz, M. Coimbra, J. Pedrosa, "Deep Learning for Segmentation of the Left Ventricle in Echocardiography", Proceedings of the IEEE 7th Portuguese Meeting on Bioengineering, Porto, Portugal, 22-23 June 2023, pp. 159-162.

[5] 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", Proceedings of the IEEE 15th International Symposium on Biomedical Imaging, Washington, DC, USA, 4-7 April 2018, pp. 898-902.

[6] G. Carneiro, J. C. Nascimento, "Combining Multiple Dynamic Models and Deep Learning Architectures for Tracking the Left Ventricle Endocardium in Ultrasound Data", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 11, 2013, pp. 2592-2607.

[7] S. Dong, G. Luo, G. Sun, K. Wang, H. Zhang, "A left ventricular segmentation method on 3D echocardiography using deep learning and snake", Proceedings of the Computing in Cardiology Con-

ference, Vancouver, BC, Canada, 11-14 September 2016, pp. 473-476.

[8] S. Dong, G. Luo, K. Wang, S. Cao, Q. Li, H. Zhang, "A Combined Fully Convolutional Networks and Deformable Model for Automatic Left Ventricle Segmentation Based on 3D Echocardiography", BioMed Research International, Vol. 2018, No. 1, 2018, p. 5682365.

[9] Y. Ali, S. Beheshti, F. Janabi-Sharifi, "Echocardiogram segmentation using active shape model and mean squared eigenvalue error", Biomedical Signal Processing and Control, Vol. 69, 2021, p. 102807.

[10] V. Zyuzin, J. Komleva, S. Porshnev, "Generation of echocardiographic 2D images of the heart using cGAN", Journal of Physics: Conference Series, Vol. 1727, No. 1, 2021, p. 012013.

[11] A. Gilbert, M. Marciniak, C. Rodero, P. Lamata, E. Samset, K. Mcleod, "Generating Synthetic Labeled Data From Existing Anatomical Models: An Example With Echocardiography Segmentation", IEEE Transactions on Medical Imaging, Vol. 40, No. 10, 2021, pp. 2783-2794.

[12] M. J. Mortada, S. Tomassini, H. Anbar, M. Morettini, L. Burattini, A. Sbrollini, "Segmentation of Anatomical Structures of the Left Heart from Echocardiographic Images Using Deep Learning", Diagnostics, Vol. 13, No. 10, 2023.

[13] J. Liang et al. "Echocardiographic segmentation based on semi-supervised deep learning with attention mechanism", Multimedia Tools and Applications, Vol. 83, No. 12, 2024, pp. 36953-3697.

[14] S. Zhuang, H. Zhang, W. Ding, Z. Zhuang, J. Zhang, Z. Gao, "Semi-supervised domain adaptation incorporating three-way decision for multi-view echocardiographic sequence segmentation", Applied Soft Computing, Vol. 155, 2024, p. 11144.

[15] Y. Wan et al. "A Semi-supervised Four-Chamber Echocardiographic Video Segmentation Algorithm Based on Multilevel Edge Perception and Calibration Fusion", Ultrasound in Medicine & Biology, Vol. 50, No. 9, 2024, pp. 1308-1317.

[16] G. F. Cacao, D. Du, N. Nair, "Unsupervised Image Segmentation on 2D Echocardiogram", Algorithms, Vol. 17, No. 11, 2024, p. 515.

[17] M. H. Jafari et al. "Semi-Supervised Learning For Cardiac Left Ventricle Segmentation Using Conditional Deep Generative Models as Prior", Proceedings of the IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8-11 April 2019, pp. 649-652.

[18] D. Cao, J. Dang, Y. Zhong, "Real-Time Segmentation of Echocardiograms with Geometric Information Assistance", Journal of Computer-Aided Design & Computer Graphics, Vol. 34, No. 8, 2022, pp. 1252-1259.

[19] C. Liu, S. Dong, F. Xiong, L. Wang, B. Li, H. Wang, "Echocardiographic mitral valve segmentation model", Journal of King Saud University - Computer and Information Sciences, Vol. 36, No. 9, 2024, p. 10221.

[20] K. Wang, H. Hachiya, H. Wu, "A Multi-Fusion Residual Attention U-Net Using Temporal Information for Segmentation of Left Ventricular Structures in 2D Echocardiographic Videos", International Journal of Imaging Systems and Technology, Vol. 34, No. 4, 2024, p. e23141.

[21] Z. Feng, J. A. Sivak, A. K. Krishnamurthy, "Two-Stream Attention Spatio-Temporal Network For Classification Of Echocardiography Videos", Proceedings of the IEEE 18th International Symposium on Biomedical Imaging, Nice, France, 13-16 April 2021, pp. 1461-1465.

[22] 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, 2019, pp. 2198-2210.

[23] M. Noroozi, P. Favaro, "Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles", Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 11-14 October 2016, pp. 69-84.

[24] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556, 2015.

[25] V. Zyuzin et al. "Identification of the left ventricle endocardial border on two-dimensional ultrasound images using the convolutional neural network Unet", Proceedings of the Ural Symposium on Biomedical Engineering, Radioelectronics and

Information Technology, Yekaterinburg, Russia, 7-8 May 2018, pp. 76-78.

[26] A. Amer, X. Ye, M. Zolgharni, F. Janan, "ResDUnet: Residual Dilated UNet for Left Ventricle Segmentation from Echocardiographic Images", Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, Montreal, QC, Canada, 20-24 July 2020, pp. 2019-2022.

[27] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion", The Journal of Machine Learning Research, Vol. 11, 2010, pp. 3371-3408.

[28] A. Sharma, O. Grau, M. Fritz, "VConv-DAE: Deep Volumetric Shape Learning Without Object Labels", Proceedings of Computer Vision - ECCV 2016 Workshops, Amsterdam, The Netherlands, 8-10 October 2016, pp. 236-250.

[29] Y. Yu, C. Wang, Q. Fu, R. Kou, W. Wu, T. Liu, "Survey of Evaluation Metrics and Methods for Semantic Segmentation", Computer Engineering and Applications, Vol. 59, No. 6, 2023, p. 57.

[30] C. Wang, Z. Zhao, Q. Ren, Y. Xu, Y. Yu, "Dense U-net Based on Patch-Based Learning for Retinal Vessel Segmentation", Entropy, Vol. 21, No. 2, 2019, p. 168.

[31] X. Li, Y. Wang, W. Yan, R. J. Van der Geest, Z. Li, Q. Tao, "A Multi-Scope Convolutional Neural Network for Automatic Left Ventricle Segmentation from Magnetic Resonance Images: Deep-Learning at Multiple Scopes", Proceedings of the 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, Beijing, China, 13-15 October 2018, pp. 1-5.