Article:
Guo, L, Lei, B, Chen, W et al. (7 more authors) (2021) Dual attention enhancement feature fusion network for segmentation and quantitative analysis of paediatric echocardiography. Medical Image Analysis, 71. 102042. ISSN 1361-8415

https://doi.org/10.1016/j.media.2021.102042

Dual Attention Enhancement Feature Fusion Network for Segmentation and Quantitative Analysis of Paediatric Echocardiography

Libao Guo[a]†, Baiying Lei[a]†, Weiling Chen[b], Jie Du[a], Alejandro F. Frangi[c,d], Jing Qin[e], Cheng Zhao[a], Pengpeng Shi[b], Bei Xia[b*], Tianfu Wang[a]*

[a] *National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Centre, Shenzhen University, Shenzhen, China Tel: 86-755-86713997 (Correspondence Tianfu Wang:email:tfwang@szu.edu.cn, Bei Xia:xiabeimd@qq.com).*

[b] *Department of Ultrasound Department, Shenzhen Children Hospital, Hospital of Shantou University, 518050*

[c] *Centre for Computational Imaging & Simulation Technologies in Biomedicine (CISTB), School of Computing, University of Leeds, Leeds LS29JT, UK.*

[d] *Leeds Institute of Data Analytics, School of Medicine, University of Leeds, Leeds LS29JT, UK.*

[e] *Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong.*

Abstract Paediatric echocardiography is a standard method for screening Congenital Heart Disease (CHD). The segmentation of paediatric echocardiography is essential for subsequent extraction of clinical parameters and interventional planning. However, it remains a challenging task due to (1) the considerable variation of key anatomic structures, (2) poor lateral resolution affecting accurate boundary definition, (3) the existence of speckle noise and artefacts in echocardiographic images. In this paper, we propose a deep network to address these challenges. We first present a dual-path feature extraction module (DP-FEM) to extract rich features via channel attention mechanism. A high and low-level feature fusion module (HL-FFM) is devised based on spatial attention, which selectively fuses rich semantic information from high-level features with spatial cues from low-level features. In addition, a hybrid loss is designed to deal with pixel-level misalignment and boundary ambiguities. Based on the segmentation results, we derive key clinical parameters for diagnosis and treatment planning. The proposed method is extensively evaluated on 4,485 two-dimensional (2D) paediatric echocardiograms from 127 echocardiographic videos. The proposed method achieves better segmentation performance than other state-of-the-art networks. We demonstrate excellent potential for automatic segmentation and quantitative analysis of paediatric echocardiography. Our code is publicly available at https://github.com/end-of-the-century/Cardiac.

Keywords: Paediatric echocardiography segmentation and quantitative analysis; attention mechanism; feature fusion

1 Introduction

Congenital Heart Disease (CHD) is a disease that exists at birth and affects cardiac structure and function in babies. CHD is one of the most common types of birth defect, which leads to severe physiological consequences and even life-threatening events (Howell *et al.*, 2019; Mendis *et al.*, 2011). The incidence rate of CHD in the world is about 1%, and its incidence rate in China is even higher (Howell *et al.*, 2019; Lang *et al.*, 2015; Zhao *et al.*, 2019). Although Magnetic Resonance Imaging and Computed Tomography have received broad attention from researchers because of the excellent accuracy in CHD diagnosis (Metaxas *et al.*, 2004; Peng *et al.*, 2016; Wang *et al.*, 2019), echocardiography is the most commonly used screening imaging method for heart disease due to its low cost, safety (no radiation), and real-time performance (Lopez *et al.*, 2010). In the diagnosis of CHD, the four-chamber (4CH) view of paediatric echocardiography is the most widely used. This view exhibits not only multiple structures but also various functional parameters are derived from it (Copel *et al.*, 1987). Clinical parameters (e.g., left ventricle long diameter (LVD), the mitral valve inner diameter (MVD), the area of the left ventricle (LVA) and the left atrium (LAA)) are used to evaluate the size and function of the heart (Lang *et al.*, 2015; Schiller *et al.*, 1989). The area-length method is one of the most commonly used methods to measure LVV (Parisi *et al.*, 1979). Routinely, echocardiography is manually performed by an experienced ultrasonographer, which is time-consuming, labour-intensive, repetitive, and highly subjective. Automatic quantification of of paediatric echocardiography is thus highly desirable.

To realise the automatic analysis of paediatric echocardiography, we first address two segmentation tasks. The first task is to segment left ventricle (LV) and left atrium (LA) in 4CH view. The second row of Figure 1 show the LV enclosed in a red line, and the LA defined by a purple line. The second task is to segment apical triangle (APT) illustrated in the third row of Figure 1. The APT is the triangle comprised by the apex of the LV and the two endpoints of the mitral valve. Based on these segmentation results, we can further calculate the cardiac functional parameters such as LVA, LAA, LVD (the line AD in the third row of Figure 1), the MVD (the line BC in the third row of Figure 1) and LVV.

Most of the existing works aim at analysing adult echocardiographic images. Paediatric echocardiography, similarly to adult echocardiography, has inherent difficulties in automatic segmentation of ultrasound images (Gahungu *et al.*, 2020; Leclerc *et al.*, 2019b; Liu *et al.*, 2019). The problems are illustrated in the first row of Figure 1: 1) contrast between the myocardium and the blood pool is low, and the image illumination is different; 2) textures in the trabecula and papillary muscles are similar to

that of the myocardium, and the heart tissue has significant echo variability; 3) cardiac shape, texture, and motion vary from patient to patient; 4) ultrasound images contain speckle noise and artefacts. Compared to adult echocardiography, the segmentation task in paediatric echocardiography presents additional challenges: 1) children's heart size changes changes dramatically with age; 2) children's heart rate is faster, and cardiac boundaries in echocardiography are blurrier than in adults. In the second segmentation task, boundaries are less apparent in paediatric ultrasound making the segmentation even harder. In recent years, Convolutional Neural Network (CNN) have been increasingly used to analyse medical images (Greenspan *et al.*, 2016), and promising results have been achieved in the segmentation (Andreassen *et al.*, 2019; Hu *et al.*, 2020; Leclerc *et al.*, 2019a; Liu *et al.*, 2020; Mishra *et al.*, 2018), classification and localisation of critical structures of ultrasound images (Dong *et al.*, 2019; Lin *et al.*, 2019; Mishra *et al.*, 2018; Pu *et al.*, 2020; Wu *et al.*, 2017). Fully Convolutional Networks (FCN) (Long *et al.*, 2015) are commonly used in encoder-decoded networks, which are widely used network for semantic image segmentation. These networks extract rich features by widening/deepening the network or using feature fusion methods. For example, the U-Net (Ronneberger *et al.*, 2015) network and its 3D extended structure (Çiçek *et al.*, 2016) leverages skip connections to fuse more spatial features, and Bisenet (Yu *et al.*, 2018) explored dual paths to extract features at different scales. The Deeplab series (Chen *et al.*, 2014; Chen *et al.*, 2017) and PSPNet (Zhao *et al.*, 2017) utilised the dilated convolution to increase the receptive field. DANet (Fu *et al.*, 2019) adopted the attention mechanism to extract more informative features. Choosing a proper loss function can make the network performance even better. The Sobel loss function (Cheng *et al.*, 2020) can make the output adhere to the boundary of the object better at the pixel-level, and get more accurate boundary segmentation. Sub-pixel convolution (Shi *et al.*, 2016) can fuse the information of each channel to make the image larger, which can better integrate the spatial information of each channel. U-Net has been widely used in medical image segmentation because of its flexible structure and classic feature fusion method (Andreassen *et al.*, 2019; Leclerc *et al.*, 2019a; Liu *et al.*, 2020). To measure the desired cardiac functional parameters in the echocardiogram, segmenting cardiac structures is essential, which has received numerous attentions (Leclerc *et al.*, 2019a; Smistad and Østvik, 2017). To obtain a more precise segmentations, different feature fusion methods and attention mechanisms are applied to the neural networks. They enable the networks learn more useful features (Hu *et al.*, 2020; Leclerc *et al.*, 2020; Moradi *et al.*, 2019; Xu *et al.*, 2020).

On the other hand, in our task, structure boundary segmentation without cardiac functional parameters measurement still cannot reduce the burden on clinicians for evaluating the cardiac function. Based on the LV boundary segmentation, the researchers have made many explorations on the automatic cardiac functional parameters measurement (Arafati *et al.*, 2020; Hu *et al.*, 2020; Leclerc *et al.*, 2019b; Moradi *et al.*, 2019). However, only relying on segmenting the boundary of the LV cannot accurately measure the LVD and MVD. To measure the LVD and MVD more accurately, we propose to segment the APT and measure LVD and MVD by locating critical points on the boundary of the triangle.

We propose a novel network to comprehensively address the challenges in segmentation and quantitative analysis of paediatric echocardiography, which comprises two key modules: a dual-path feature extraction module (DP-FEM) with channel attention, and a high- and low-level feature fusion module (HL-FFM) with spatial attention. The DP-FEM drives the network to pay more attention to channel information and extract richer low-level features. The HL-FFM aims at extracting more contextual high-level features and embeds these features into the low-level features. To better integrate spatial information for precise segmentation, we harness sub-pixel convolution (Sub-Pixel UP) in the first up-sampling. Finally, high-level and low-level features are seamlessly fused to improve further the segmentation performance of the LV, LA and APT in paediatric echocardiography. When training the network, the Sobel loss function is used to enable the network to tackle boundary ambiguities. We propose a new method to extract key points, by locating the three vertices of APT, find the apex of the heart and the two endpoints of the mitral valve. For poor segmentation results,optimise the positioning method for further improving segmentation and quantitative analysis of paediatric echocardiography. Overall, our contributions are summarised as below:

- We propose a novel network for paediatric echocardiographic segmentation. Specifically, in the encoding part, a dual-path feature extraction module with channel attention is used to strengthen the feature extraction ability. In the decoding part, a high- and low-level feature fusion module (HL-FFM) with spatial attention is adopted to better fuse the high-level and low-level features. Also, the sub-pixel convolution method is explored to enhance spatial feature fusion.
- We devise a new hybrid loss function taking the complementary advantages of cross-entropy loss and Sobel loss to make the network simultaneously tackle pixel-level misalignment and boundary ambiguities.

- We propose a new parameter measurement method for paediatric echocardiographic analysis, which can achieve automatic and robust measurements and reduce the impact of localisation errors.

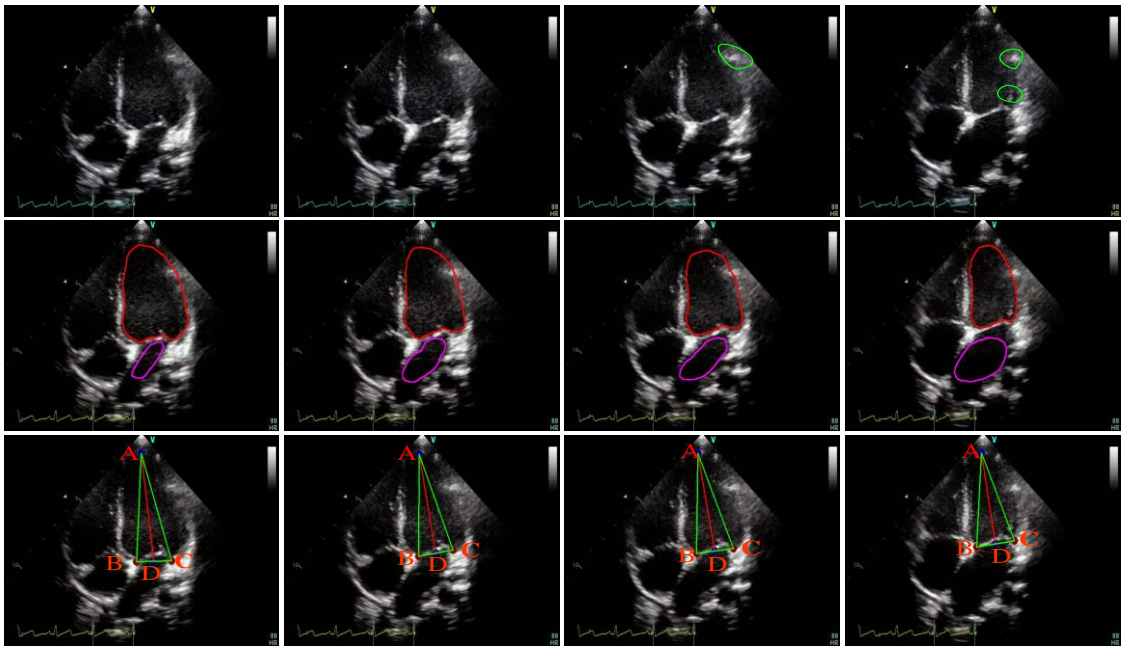


Figure 1: Paediatric echocardiogram in 4CH view (first row) with green regions depicting artefacts. Manual segmentation of LV and LA (second row), LV in red, and LA in purple. Apical triangle (third row), where the AD line is LVD, and the BC line is MVD.

2 Related work

2.1 Feature Fusion in Segmentation Methods

To obtain better segmentation performance, low-level and high-level features need to complement each other (Zhang *et al.*, 2018). Generally, low-level features are with rich spatial information but lack semantic information, while high-level features are with rich semantic information but lack of spatial detail information. The classic FCN segmentation network (Long *et al.*, 2015) consists of a convolutional layer, a pooling layer and an upsampling layer. The convolutional layer learns the features of the input image, the pooling layer optimises the feature information and reduces the calculation, the deconvolutional layer upsamples the learned features. Finally, the softmax function achieves image segmentation through a pixel-level classification. The SegNet (Badrinarayanan *et al.*, 2017) network is symmetrical, where the upsampling is used to make the image larger, and deconvolution is not used in the upsampling part of this network. In this network, the features are rearranged through convolution, which can make the boundary of the segmentation smoother. The U-Net (Ronneberger *et al.*, 2015) network adds the skip connection layers between the original image and the upsampling part so the network can learn more spatial information and improve the network performance. Deeplab v2 (Chen *et al.*, 2014) proposed

Atrous Spatial Pyramid Pooling (ASPP), which used different dilation rates in different branches to obtain features of different scales. This integrates information of these other scales to enable the network to learn more useful features. Deeplab v3 (Chen *et al.*, 2017) improved the spatial pyramid pooling block, added image-level features, and used a multi-grid method in the residual block. This network introduces different dilation ratios and enriches features of different scales. To address the problem of small perception field and loss of spatial information in semantic segmentation, Bisenet (Yu *et al.*, 2018) proposed to use dual paths, one context path, and attention module optimisations to stabilise the maximum receptive field. ExFuse (Zhang *et al.*, 2018) introduces semantic information into the low-level features, embeds spatial information in the high-level features, optimises feature fusion, and improves network performance.

2.2 Attention Mechanisms

Attention mechanisms have become a hot research topic due to the great potential for improving network performance. They are now widely used in image recognition, classification and segmentation (Chaudhari *et al.*, 2019). An attention module is usually an additional neural network that can rigidly select specific parts of the input, or assign different weights according different factors of interest. Attention mechanisms can be widely divided into channel attention, spatial attention, and mixed attention. Channel attention associates a weight to each channel according to their relevance for the task. For example, SENet (Hu *et al.*, 2018) proposes to use the Squeeze-and-Excitation module to obtain the importance of each channel through squeeze and excitation operations. . ECANet (Wang *et al.*, 2020) contains an effective channel attention (ECA) module, which only adds a few parameters with a significant improvement in network performance. The network effectively learns channel attention through non-dimensionality reduction and cross-channel learning. Spatial attention transforms the spatial information of the original image into another space while retaining essential information or properties. Hybrid attention is used to improve the shortcomings of spatial and channel attention. The proposed non-local block in the literature (Wang *et al.*, 2018) calculates the response of a specific position as the weighted sum of the features of all positions and introduces spatial attention. GCNet (Cao *et al.*, 2019) combines the advantages of non-local (Wang *et al.*, 2018) modules and SENet, which can realise context modelling without more parameters. Hu *et al.* (Hu *et al.*, 2020) proposed a dual-path network, which used

dual convolutional block attention module (Woo *et al.*, 2018) to guide the network to learn features. They achieved good results in the segmentation of primary echocardiograph.

2.3 Echocardiography Segmentation and Quantitative Analysis

To accurately evaluate the function of the heart, it is necessary to measure the cardiac functional parameters of the heart, such as LVV, MVD, tricuspid valve distance, and ejection fraction. To measure these parameters, the boundaries of the anatomical structures or locate the key points of structures need segmenting. Current work is mostly focused on automatic segmentation of echocardiography. Some researchers have explored direct and automatic clinical parameter measurements (Andreassen *et al.*, 2019; Du *et al.*, 2018; Sultan *et al.*, 2018). More ordinarily, however, cardiac functional parameters are derived from segmentation. Some papers, however, have researched the automatic quantification of the LV in echocardiography. For example, Leclerc *et al.* (Leclerc *et al.*, 2019b) segmented the LV wall, and LA endocardium from two-chamber (2CH) heart views and corresponding 4CH heart views at end-systole and end-diastole. Ventricular volumes and ejection fraction are then measured based on the segmentation results. Only the automatic measurement of LVV was done, and the result was below the gold standard. Ge *et al.* (Ge *et al.*, 2019) proposed a network automatically and directly estimating multiple LV-related cardiac functional parameters from paired echocardiograms (4CH and 2CH). Moradi *et al.*(Moradi *et al.*, 2019) proposed a novel method for measuring LV volume.

First, the boundary of the LV is segmented, the smallest circumscribed triangle of this boundary is found, the points on the periphery are traversed, and the three points closest to the three fixed points of the triangle are on the perimeter. These three points are the apex of the heart and the two endpoints of the mitral valve. The long diameter is further calculated to get the LV volume. This method uses the circumscribed triangle to find the mitral valve breakpoint, which has apparent errors, resulting in inaccurate ventricular volume measurements. Arafati *et al.*(Arafati *et al.*, 2020) used a generative adversarial network to segment the LV, LA, right ventricle (RV), and right atrium (RA) boundaries and measured the LV volume. The long diameter is calculated to find the leftmost point and the rightmost point on the LV boundary point. By the intersection of the bisector of these two points and the mitral valve, the point farthest from the boundary is the long axis. The first point found by this method is not necessarily the midpoint of the mitral valve, which does not match the standard LVD definition. Ouyang *et al.* (Ouyang *et al.*, 2020) proposed a dual network structure. One network is a segmentation network to segment the LV

region. The other network uses a residual network and a spatial-temporal convolution CNN model to predict ejection fraction and segment it in a 2D echocardiography video. The LV area and prediction of ejection fraction can be calculated.

3 Methodology

3.1 Overview

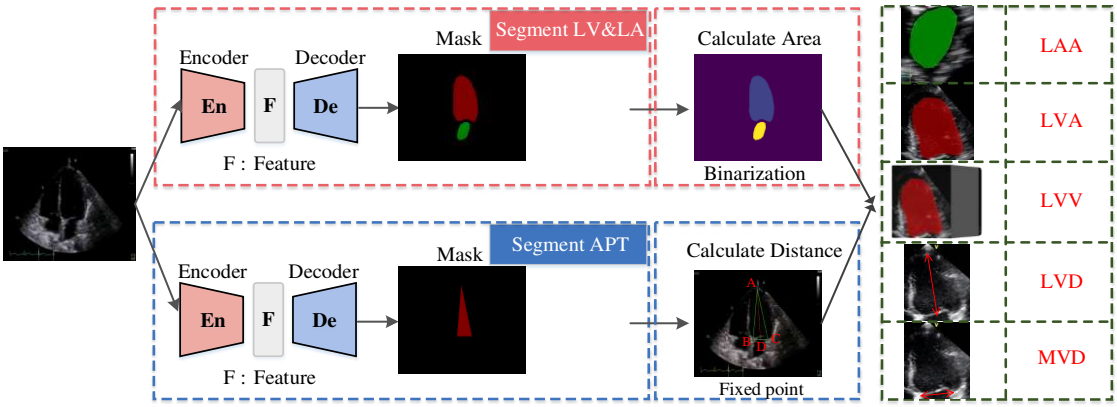


Figure 2: The flowchart of the segmentation and quantitative analysis of paediatric echocardiography. The upper segmentation network separates the LV membrane and the LA endometrium boundary and calculates the area. The lower network segments the apical triangle locates three vertices and calculates the long diameter and the mitral valve distance, calculate the ventricular volume through these two parts.

Figure 2 shows the overall architecture of automatic segmentation and quantitative analysis of paediatric echocardiography. There are two sub-tasks. The first task is to segment the boundary of the LV and the boundary of the LA, which further calculates the LVA and LAA. The second task is to segment the APT, locate the three vertices, calculate MVD and LVD, and calculate the LVV through the ellipsoid formula. The ellipsoid formula is described as:

$$V = \frac{8A^2}{3\pi L}, \tag{1}$$

where V is the LVV, A is the LVA, and L is the LVD. The same segmentation network is used in the two segmentation tasks.

Figure 3 depicts the proposed segmentation network, which is an encoding-decoding network. The lower network layer provides a feature map describing spatial structure information, the higher network contains rich semantic information. We design different modules in the encoding part and the decoding part to improve the segmentation performance of the network. In the encoding part, the DP-FEM consists of two branches extracting complementary features. In the decoding part, HL-FFM combines high-level information with low-level descriptors. The high-level features are sampled to the size of the

upper-level feature map, and then a semantic module is introduced.The Sub-Pixel UP rearranges the pixels of each channel through pixel shuffling, then make the features of each channel merge together. In this way, the network learns more spatial features between channels.

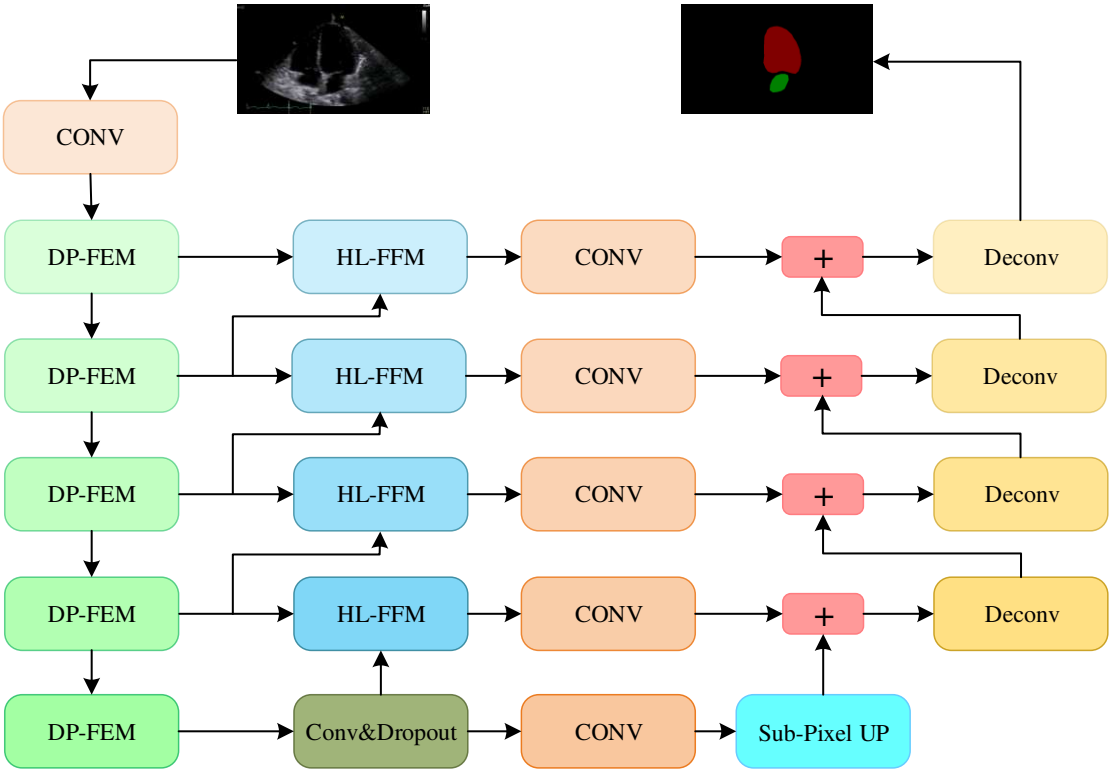


Figure 3: The segmentation network architecture. The DP-FEM has two branches, which can extract complementary features and merge them. HL-FFM can effectively fuse rich semantic information in high-level features, and spatial information in low-level features. Sub-Pixel directly integrates the characteristics of each channel and effectively incorporates spatial features.

3.2 Dual-Path Feature Extraction Module

In the CNN, the convolutional layer extracts features, and the pooling layer removes redundant information and identifies essential features. As shown in Figure 4, the DP-FEM module consists of two branches. In the first branch of this network, the max pooling layer is selected to extract representative features. Since some information is inevitably lost in the pooling process, we design another branch, which uses convolution to extract features and learns the interactive information between channels through an ECA module. It merges the two sets of features through addition as the input of the next convolution layer.

The performance of the CNN is complemented with an attention mechanism widely used in other applications (Chaudhari *et al.*, 2019). Previous work mainly developed complex attention modules, which inevitably increases the calculation (Cao *et al.*, 2019; Hu *et al.*, 2018; Li *et al.*, 2019). The

trade-off between model performance and complexity, Wang *et al.* proposed ECANet (Wang *et al.*, 2020), which adds a few parameters to obtain significant performance gains. The corresponding relationship between the channel and its weight is indirect. ECA is a module for cross-channel information interaction without reducing the channel dimensionality, which can be defined as

$$\omega = \sigma\left(\mathrm{C1}D_k(y)\right), \tag{2}$$

where $\mathrm{C1}D_k$ indicates 1D convolution,σ is a sigmoid function, k is the size of the convolution kernel for 1D convolution, y is the aggregation feature without dimensionality reduction. In the case of a gi-ven channel size C, the size of the convolution kernel k can be adaptively changed by

$$k = \psi(\mathrm{C}) = \left|\frac{log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{odd}, \tag{3}$$

where $|t|_{odd}$ indicates the nearest odd number of t. In this paper, we set γ and b to 2 and 1, respectively, in all experiments. Through the mapping ψ, high-dimensional channels have longer-range interactions, while low-dimensional channels use nonlinear mapping for shorter-range interactions.

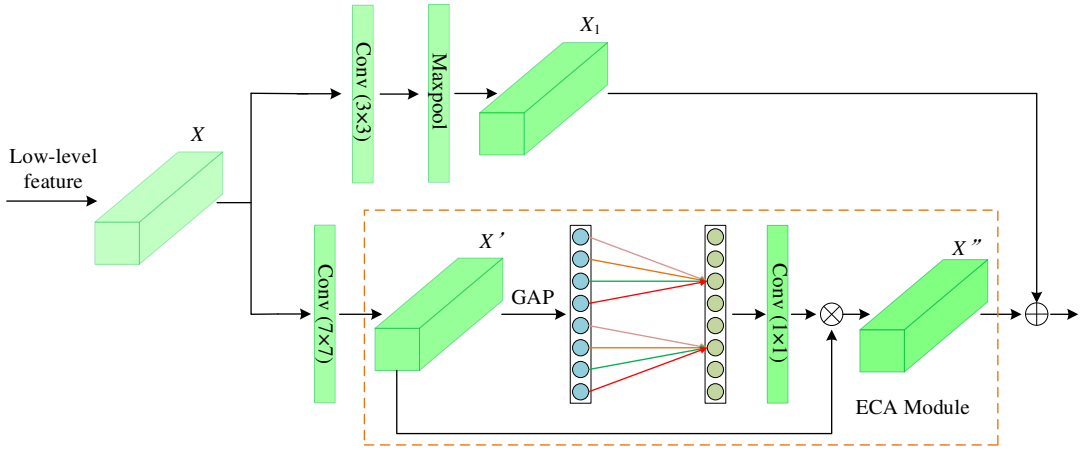


Figure 4: Dual-path feature extraction module, the upper-path feature map extracts salient features through convolution and maximum pooling layers, and the lower path uses a convolution and an ECA module to learn co-channel information, and then the two features are merged by addition. X is the original feature, X_1 is the feature after the first path, X' is the feature after a 7×7 convolution, and X'' is the feature after the ECA module.. Low-level feature is a feature from a network lower than the current module. GAP represents global average pooling.

3.3 High and Low-Level Feature Fusion Module

To embed more semantic features in low-level features, we use a global semantic module for high-level feature fusion in each upsampling part. Figure 5 shows the HL-FFM structure. In this module, the representative and semantic information are extracted from high-level semantic features through a global semantic feature module and add them to low-level features, which can be expressed as

$$z_i = x_{low\_i} + W_{v2}ReLU(LN(W_{v1}\sum_{j=1}^{N_p}\frac{e^{w_k x_j}}{\sum_{m=1}^{N_p} e^{w_k x_m}} x_j)), \tag{4}$$

where x_{low_i} is the i-th element in the low-level feature, $\alpha_j = \sum_{m=1}^{N_p} e^{w_k x_m}$ is the weight for global attention pooling, N_p is the number of positions in the feature map, w_k, W_{v1} and W_{v2} denote linear transformation matrices. LN indicates LayerNorm (Ba *et al.*, 2016), ReLU denotes Rectified Linear Unit (Nair and Hinton, 2010). The bottleneck is transformed to capture channel-wise dependencies. Finally, the semantic features extracted from the high-level and low-level features are added element by element to fuse the features. Sub-pixel convolutions are used to capture more spatial context as the first step in upsampling. Sub-Pixel UP is shown in Figure 6.

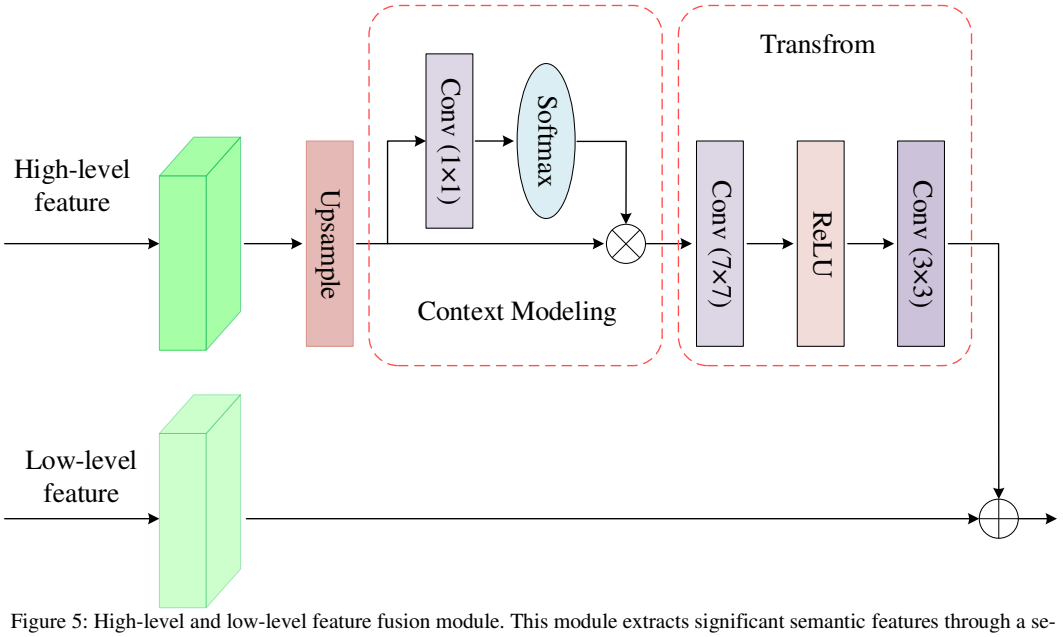


Figure 5: High-level and low-level feature fusion module. This module extracts significant semantic features through a semantic feature extraction module and embeds them into low-level features through addition. High-level features are features from a network deeper than the current module. Transform is to ensure that the two features are of the same type.

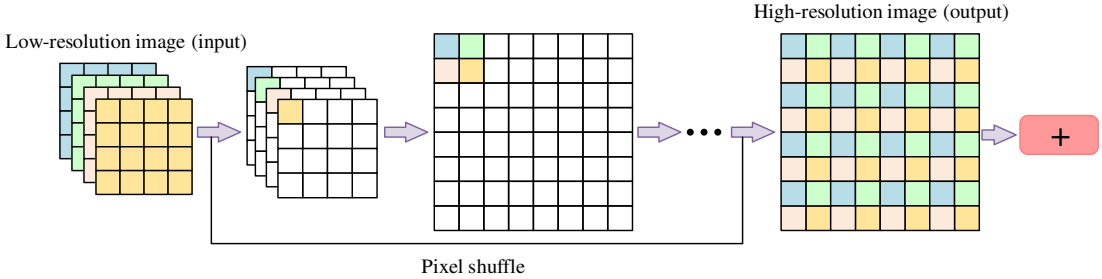


Figure 6: The sub-pixel convolution layer, the feature map can be enlarged only by reshaping the space and channel dimensions so more spatial information can be embedded in the advanced features of the network.

3.4 Loss Function

The standard cross-entropy and gradient loss function is used in our model. Different loss functions have different weights. The l_1 loss on the segmentation gradient amplitude is used to refine segmentation boundaries. The segmentation gradient is estimated by the 3×3 mean filter and the subsequent Sobel

operator (Kanopoulos *et al.*, 1988), and the Sobel loss can make the output better adhere to the boundary of the object at the pixel-level. The gradient loss can be described as

$$L_{Sobel} = \frac{\alpha}{n}\sum_i \left\| \nabla\left(f_m(x_i)\right) - \nabla(f_m(y_i)) \right\|_1, \quad (5)$$

where $f_m(.)$ indicates the 3×3 mean filter, ∇ denotes the gradient operator approximated by the Sobel operator, n represents the total number of pixels, x_i and y_i represent the *i*-th pixel of the label and segmentation result, respectively. Our final loss can be written as

$$L = L_{CE} + L_{Sobel}, \quad (6)$$

where L_{CE} denotes cross-entropy loss, L_{Sobel} denotes Sobel loss. In the experiment, α is set to 0.5.

3.5 Optimising Localisation

LVA and LAA can be calculated by pixel counting. The LVD and MVD are computed by dividing the APT. The apex of the heart is denoted by P_{apex}, and the two endpoints of the mitral valve are represented by P_{mvd1} and P_{mvd2}. The automatic calculation process is: we first segment the boundary of the APT, and use the *findContours* function in OpenCV to get the boundary coordinates. The coordinates on the perimeter of the triangle are (x_{APT_j}, y_{APT_j}), where *j* represents the *j*-th coordinate value clockwise from the apex of the heart. By calculating the maximum y_{APT_j} value (apex of the heart), denoted as P_{ymax}, and the minimum y_{APT_j} value (an end of the mitral valve) of the coordinates, denoted as P_{ymin}, the values of the two vertices of the APT can be determined. P_{ymax} is the apex of the heart (P_{apex}), P_{ymin} is an end of the mitral valve (P_{mvd1}or P_{mvd2}). P_{ymax} and P_{ymin} can determine a vector, then calculate the distance from the point on the APT boundary to this vector. We find the maximum value of the distance to this vector, and the corresponding point, that is another point of the APT. Also, we calculate the midpoint of the P_{mvd1} and P_{mvd2}, and calculate the distance from the midpoint to the apex of the heart to be the LVD. The two end points of the triangle base are the MVD.

Because there is no obvious boundary in APT, the border of the segmentation result is easy to be distorted. According to the observation result, the lowest point of the boundary coordinate value of the segmentation result may not be end points of the mitral valve, which will have larger measurement error. We propose a correction method to make the measurement results more accurate. First find the maximum point P_{ymax} of y_{APT_j} on the APT boundary, and then find the minimum point P_{ymin} of y_{APT_j}. Find the maximum value of the APT boundary coordinate x_{APT_j}, denoted as P_{xmax}, and then the minimum

value of the APT boundary coordinate x_{APT_j}, denoted as P_{xmin}. Determine whether point P_{ymin} and point P_{xmax} are the same point, if not, then determine point P_{xmax} and point P_{ymax} are the same point, if not, point P_{xmax} is used to replace P_{ymin}, if it is, point P_{xmin} replace point P_{ymin}. The optimised pseudo code can be expressed as follows:

Algorithm for locating the vertices of APT

Input: APT segmentation image

Output: APT vertex coordinates

Steps of learning sequence information

1. Return a set of APT boundary coordinates through the cv2.findContours function.
2. Find the maximum value of P_{ymax} in the set (x_{APT_j}, y_{APT_j}). # Look for apex vertices in triangle APT
3. Find the minimum value of P_{ymin} in the set (x_{APT_j}, y_{APT_j}). # Look for one end of the mitral valve in the triangle APT.
4. **if** P_{ymin} != P_{xmax}**:**

   else if P_{xmax} != P_{ymax}:

   P_{ymin} =P_{xmax}

   else:

   P_{ymin} = P_{xmin} # It is found that in some APT segmentation results, the lowest point is not the end point of the mitral valve, and an optimisation method is proposed to make the lowest point close to the end point of the mitral valve.
5. Determine a vector based on the two known points.
6. Calculate the distance from the point to the vector on the APT boundary.
7. The maximum point of the distance is the third point of APT

4 Experiments

4.1 Dataset Description

The data for our training and testing are collected from Department of Ultrasound Department, Shenzhen Children Hospital, Hospital of Shantou University. There are 127 2D paediatric echocardiography 4CH videos. These data are collected from GE Vivid E8 and E9 (GE Healthcare, Horten, Norway) Ultrasound equipment. The videos meet the standards of the American Society of Echocardiography. Specifically, the video collection standard of the dataset is the echocardiogram video of healthy children aged 0-10 years, and each video contains at least 24 frames and a complete cardiac cycle. The dataset comprises two sub-datasets, one is the segmentation of the LV and LA, and the other is the

segmentation of the apical triangle (the triangle formed by the apex and the two endpoints of the mitral valve). We randomly select 100 paediatric echocardiographic videos and extract 3,654 images frame by frame as the training set. The remaining 27 4CH paediatric echocardiograms extract 831 images frame by frame as the testing set. The labelling was done by two sonographers and confirmed by another senior sonographer

4.2 Implementation Protocol and Data Augmentation

Our experiments are conducted on a computer workstation with Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz, 4 GPU NVIDIA Titan Xp, and 64G of RAM. The deep learning framework we use to train the network is PyTorch, and the version is 1.0.1. In the training phase, we set the initial learning rate to 10^{-3}, which is gradually reduced during the training of the network. We use the stochastic gradient descent optimiser to optimise and update the weights. We set the momentum=0.99. We use PyTorch pre-trained VGG (Simonyan and Zisserman, 2014) network as a backbone network to save training time. All other components in the network are randomly initialised using PyTorch default configuration.

When training and testing the network, to save memory, all image sizes are resized to 512×512. To evaluate our approach, we use the Dice index, Jaccard similarity coefficient, Recall (sensitivity), Precision and Accuracy as the evaluation metrics.

Clinically, the parameters needed to evaluate the function of the heart include LVA, LAA, LVV, MVD, LVD (the line between the apex of the heart and the midpoint of the mitral valve). To evaluate the similarity between the predicted values of these cardiac functional parameters and the gold standard, we use Pearson's coefficient (PC) and mean absolute error (MAE) metrics. The PC and MAE are defined as

$$p_i = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}, \tag{7}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i|, \tag{8}$$

where X_i represents the i-th value of the gold standard, $\bar{X}$ represents the average value of the gold standard. Y_i represents the i-th value of the predicted value, and $\bar{Y}$ represents the average value of the predicted value, n indicates the total number of values.

We evaluate the performance of our proposed network through these sets of experiments. 1) Ablation experiments of different modules; 2) Comparison of standard segmentation methods in computer vision;

3) Comparison of predicting clinical evaluation of cardiac function through various segmentation networks.

4.3 Results and Analysis

4.3.1 Ablation Analysis

To test the influence of different modules on network performance, we performed ablation experiments using DP-FEM, HL-FFM and Sub-Pixel UP modules. We also explore whether to use hybrid loss function with Sobel loss. The checkmark indicates that the module is used. In the experiment, the skeleton of the network is the VGG16 network. Our backbone structure is similar to an FCN but with additional jump connections. We verified the influence of each module on the network and report the results inTable 1, Table 2, and Table 3. Table 1 is the segmentation performance evaluation of the LV, and Table 2 is the segmentation performance evaluation of the LA. It can be seen that high-level and low-level features have a significant impact on Precision, and the dual-path feature extraction module can extract rich features, which significantly improves the overall segmentation performance. Sub-pixel convolution and Sobel loss function can better fuse spatial information, and adjustment of edges will further improve segmentation accuracy. Figure 7 is the visualisation result of the feature map. The network gradually deepens from left to right. The feature map of the low-level network contains rich spatial information, which can observe the heart outline. The high-level features of deep networks include rich semantic features to reveal pixel relationship.

Table 3 shows the segmentation results of the APT. Since the APT has no obvious boundary, the segmentation result is slightly worse. Each module also improves the segmentation performance. Figure 8 shows the visualisation results of two segmentation tasks, where the red curve is the label, and the green curve is the prediction result. The network segmentation performance is generally better, but there are subtle differences at the boundary. It can be seen that the prediction curve of our method is closest to the label. In this paper, (A) indicates the addition of the ECA module during feature extraction to form a DP-FEM. (B)means HL-FFM, (C) indicates Sub-Pixel UP and (D) indicates training the network using the Sobel function.

Table 1: Comparison of LV segmentation performance of different modules in 4CH view.

A	B	C	D	Accuracy	Precision	Recall	Jaccard	Dice
				0.9895	0.9592	0.9299	0.8935	0.9431
				0.9910	0.9572	0.9484	0.9092	0.9517

	✓			0.9908	**0.9602**	0.9450	0.9089	0.9514
	✓	✓		0.9912	0.9578	0.9510	0.9123	0.9533
	✓	✓	✓	**0.9915**	**0.9602**	**0.9512**	**0.9149**	**0.9547**

Table 2: Comparison of LA segmentation performance of different modules in 4CH view.

A	B	C	D	Accuracy	Precision	Recall	Jaccard	Dice
				0.9895	0.9275	0.8808	0.8196	0.8977
✓				0.9910	0.9030	**0.9259**	0.8386	0.9094
✓	✓			0.9908	**0.9222**	0.9044	0.8370	0.9082
✓	✓	✓		0.9912	0.9150	0.9173	0.8419	0.9111
✓	✓	✓	✓	**0.9915**	0.9108	0.9245	**0.8451**	**0.9130**

Table 3: Comparison of APT segmentation performance of different modules in 4CH view

i	ii	iii	iv	Accuracy	Precision	Recall	Jaccard	Dice
				0.9938	**0.9458**	0.8640	0.8217	0.9009
✓				0.9945	0.8614	**0.9542**	0.8257	0.9032
✓	✓			0.9946	0.8701	0.9477	0.8282	0.9047
✓	✓	✓		0.9947	0.8843	0.9371	0.8331	0.9077
✓	✓	✓	✓	**0.9947**	0.8964	0.9293	**0.8371**	**0.9100**

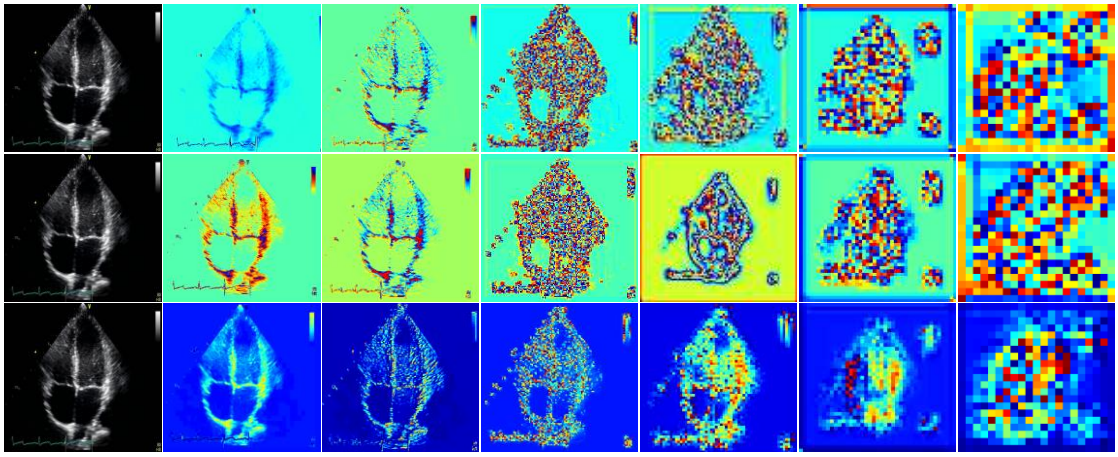
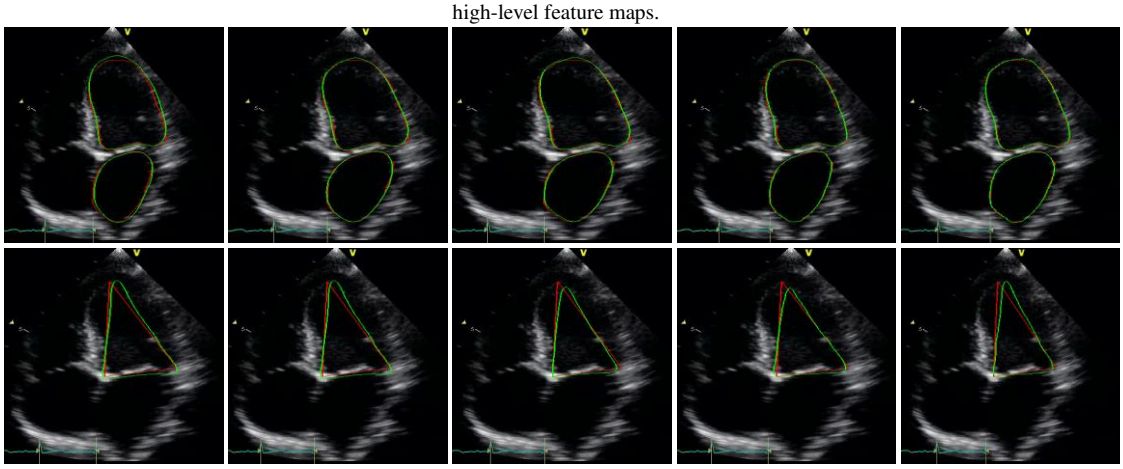


Figure 7: Feature map visualisation, the first column is the original input image, from left to right are low-level feature maps to high-level feature maps.



backbone backbone+A backbone+A+B backbone+A+B+C backbone+A+B+C+D

Figure 8: The visualisation of segmentation results. The red curve is the label, and the green curve is the prediction result. backbone+A adds a dual-path feature extraction module, backbone+A+B adds a high- and low-level feature fusion module, backbone+A+B +C adds a sub-pixel convolution, and backbone+A+B +C +D results with the Sobel loss function training.

4.3.2 Comparison with State-of-the-art Methods

To evaluate the performance of the network, we use the network that has made outstanding achievements in computer vision including FCN (Long *et al.*, 2015), U-Net (Ronneberger *et al.*, 2015), U-Net+ ASPP, Bisenet (Yu *et al.*, 2018), Deeplab_V3 (Chen *et al.*, 2017), PSPNet (Zhao *et al.*, 2017), SegNet (Badrinarayanan *et al.*, 2017), DANet (Fu *et al.*, 2019), AIDAN (Hu *et al.*, 2020). These networks are well-known for natural image segmentation in computer vision. They have also been used with good results in medical image segmentation. Table 4 and Table 5 compare state-of-the-art methods for LA and LV segmentation, respectively. Table 6 compares APT segmentation performance with state-of-the-art methods. Our method achieves the best results according to all evaluation metrics. When segmenting the APT, the overall results are poor, because there is no obvious boundary, and more advanced semantic features are needed to guide the segmentation network. In the first segmentation task, compared with the AIDAN method, considering the memory limitation, we use single-card training and set batch-size to one. To ensure the originality of the image, we did not crop the original image. When we use our training code to train AIDAN, the result is slightly lower than the original author's method. Figure 9 is a visual comparison of the results with state-of-the-art segmentation networks. The red curve is the ground truth, and the green curve is the network prediction result. The image in Figure 9 is the same frame. It can be seen that the prediction result of our proposed network is closer to the ground truth. Especially in the segmentation of boundaries, the better are the results, the smoother the edges. In the second segmentation task, the segmentation results are generally worse than that of the first segmentation task. That is because there are no noticeableedges and corners when the three vertices of the triangle appear.

Table 4: Comparison of LV segmentation performance of state-of-the-art methods in 4CH view.

Network	Accuracy	Precision	Recall	Jaccard	Dice
FCN	0.9896	0.9414	0.9472	0.8922	0.9420
U-Net	0.9880	0.9122	0.9488	0.8685	0.9271
U-Net+ASPP	0.9862	0.8903	0.9472	0.8473	0.9135
Bisenet	0.9851	0.9204	0.9274	0.8549	0.9197
DeepLab	0.9879	0.9247	0.9484	0.8791	0.9345
PSPNet	0.9857	0.9385	0.9290	0.8752	0.9322
Segnet	0.9859	0.9132	0.9287	0.8515	0.9173
DANet	0.9867	0.8866	0.9558	0.8496	0.9152
AIDAN	0.9886	0.9350	0.9447	0.8852	0.9376
OURS	**0.9915**	**0.9602**	**0.9512**	**0.9149**	**0.9547**

Table 5: Comparison of LA segmentation performance of state-of-the-art methods in 4CH view

Network	Accuracy	Precision	Recall	Jaccard	Dice
FCN	0.9896	0.8972	0.9078	0.8167	0.8947
U-Net	0.9880	0.8694	0.9041	0.7915	0.8777
U-Net+ASPP	0.9862	0.8651	0.8827	0.8081	0.8610
Bisenet	0.9851	0.7845	0.8914	0.6999	0.8053
DeepLab	0.9879	0.7933	0.9104	0.7422	0.8382

PSPNet	0.9857	0.8872	0.8478	0.7705	0.8580
Segnet	0.9859	0.8006	0.9145	0.7411	0.8407
DANet	0.9867	0.8174	0.9103	0.7537	0.8471
AIDAN	0.9886	0.8463	0.9174	0.7805	0.8693
OURS	**0.9915**	**0.9108**	**0.9245**	**0.8451**	**0.9130**

Table 6: Comparison of APT segmentation performance of state-of-the-art methods in 4CH view.

Network	**Accuracy**	**Precision**	**Recall**	**Jaccard**	**Dice**
FCN	0.9932	0.8274	0.9497	0.7874	0.8782
U-Net	0.9935	0.8705	0.9162	0.8007	0.8876
U-Net+ASPP	0.9886	0.6470	0.9454	0.6123	0.7209
Bisenet	0.9923	0.8168	0.9254	0.7582	0.8569
DeepLab	0.9933	0.8462	0.9312	0.7911	0.8812
PSPNet	0.9924	**0.8027**	**0.9405**	0.7577	0.8540
Segnet	0.9917	0.7927	0.9286	0.7391	0.8433
DANet	0.9927	0.8194	0.9325	0.7663	0.8636
AIDAN	0.9938	0.8575	0.9342	0.8039	0.8884
OURS	**0.9947**	0.8964	0.9293	**0.8371**	**0.9100**

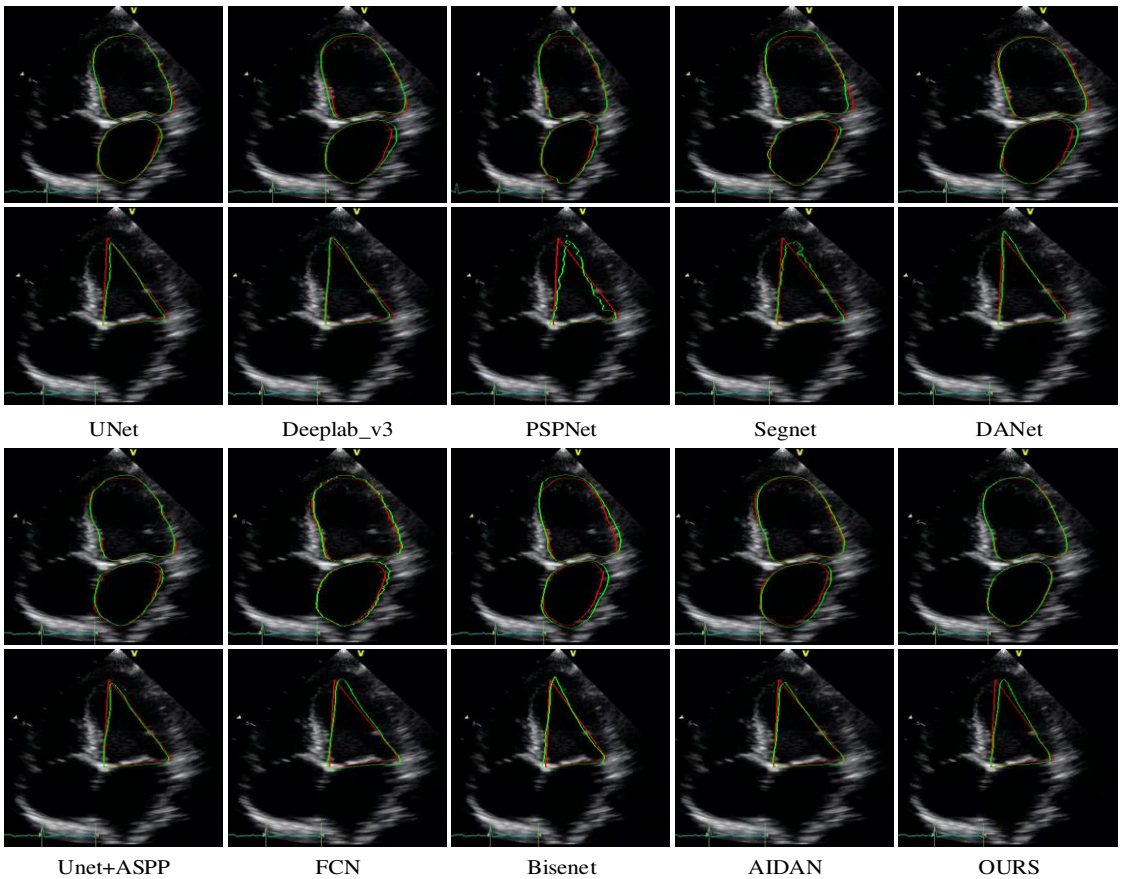


Figure 9: Visualisation of the segmentation results. The red curve is the label, and the green curve is the prediction result.

4.3.3 Estimation of Clinical Parameters

To clinically evaluate the heart function for CHD screening, LVA, LAA, MVD, LVD, LVV are essential in assessing cardiac function. Here we select several representative parameters for automatic measurement. The units of these parameters in this article are: LVA (cm^2) , LAA (cm^2), MVD (cm), LVD(cm), LVV (ml). The LVA and LAA are automatically measured by automatically segmenting the

boundary of the LV and the boundary of the LA, and filling the area of the LV and LA with pixels. The LV is filled with red and LA with green. Then the image performs binarisation and calculates the LVA and LAA by counting the number of pixels. LVD and MVD are calculated by locating key points on APT. Tables 7 and 9 compare PC between the basic network architecture and other state-of-the-art methods with various modules bolt on the basic network. The blacked numbers represent the best results. Table 8 and Table 10 are respectively, the MAE of the draw between adding different network modules and the proposed model and the state-of-the-art method. The blacked numbers represent the best results. The above tables that our method has achieved better performance. To show the similarity between the prediction result and the label more intuitively, we drew correlation and Bland Altman plots. Figure 10 and Figure 11 similarly report on the analysis of cardiac functional parameters. The label and the prediction result are in good agreement overall. Still, there are also points with large deviations, such as the three points in Figure 10 (d) and Figure 11 (d), which are due to the segmentation results not good enough leads to errors in the prediction of cardiac functional parameters.

Table 7: Comparison of different modules (PC).

A	B	C	D	LVA	LAA	LVD	MVD	LVV
				0.9790	0.9623	**0.9655**	0.8262	0.9640
✓				0.9824	0.9639	0.9475	0.8401	**0.9651**
✓	✓			0.9800	0.9582	0.9513	**0.8466**	0.9600
✓	✓	✓		0.9822	0.9621	0.9450	0.8463	0.9600
✓	✓	✓	✓	**0.9828**	**0.9653**	0.9457	0.8350	0.9636

Table 8: Comparison of different modules (MAE).

A	B	C	D	LVA	LAA	LVD	MVD	LVV
				0.7644	0.4976	**0.3018**	**0.1796**	**3.9517**
✓				0.6067	0.4348	0.7983	0.2424	5.7154
✓	✓			0.6162	0.4573	0.7851	0.2335	5.7474
✓	✓	✓		0.5760	0.4207	0.6666	0.2228	4.6976
✓	✓	✓	✓	**0.5426**	**0.4086**	0.5592	0.1971	4.0967

Table 9: Comparison of state-of-the-art methods (PC).

Network	LVA	LAA	LVD	MVD	LVV
FCN	0.9191	0.9420	0.8498	0.6020	0.8795
U-Net	0.9465	0.9248	0.8918	0.7598	0.9214
U-Net+attention	0.9226	0.8958	0.3122	0.4564	0.5286
Bisenet	0.8969	0.7862	0.8780	0.5436	0.8458
DeepLab	0.9668	0.8960	0.8096	0.6968	0.6920
PSPNet	0.9583	0.7373	0.8488	0.5220	0.9071
Segnet	0.9412	0.9077	0.8083	0.6795	0.7123
DANet	0.9220	0.9132	0.8641	0.6731	0.9017
AIDAN	0.9535	0.9331	0.8997	0.7851	0.9359
OURS	**0.9828**	**0.9653**	**0.9457**	**0.8350**	**0.9636**

Table 10: Comparison of state-of-the-art methods (MAE).

Network	**LVA**	**LAA**	**LVD**	**MVD**	**LVV**
FCN	0.8515	0.4362	0.6576	0.3101	4.9225
U-Net	1.2074	0.6586	0.6550	0.2671	4.4444
U-Net+attention	1.2924	0.7598	1.5702	0.5638	19.1984
Bisenet	1.0715	0.8996	0.6334	0.4006	5.6709
DeepLab	0.9191	0.8719	0.9057	0.3197	6.5757
PSPNet	0.6072	0.8125	0.7162	0.3434	4.5860
Segnet	1.1577	0.8332	0.9291	0.3286	7.4552
DANet	1.3820	0.7723	0.6823	0.3501	4.6793
AIDAN	0.8975	0.6641	0.6209	0.2468	4.6072
OURS	**0.5426**	**0.4086**	**0.5592**	**0.1971**	**4.0967**

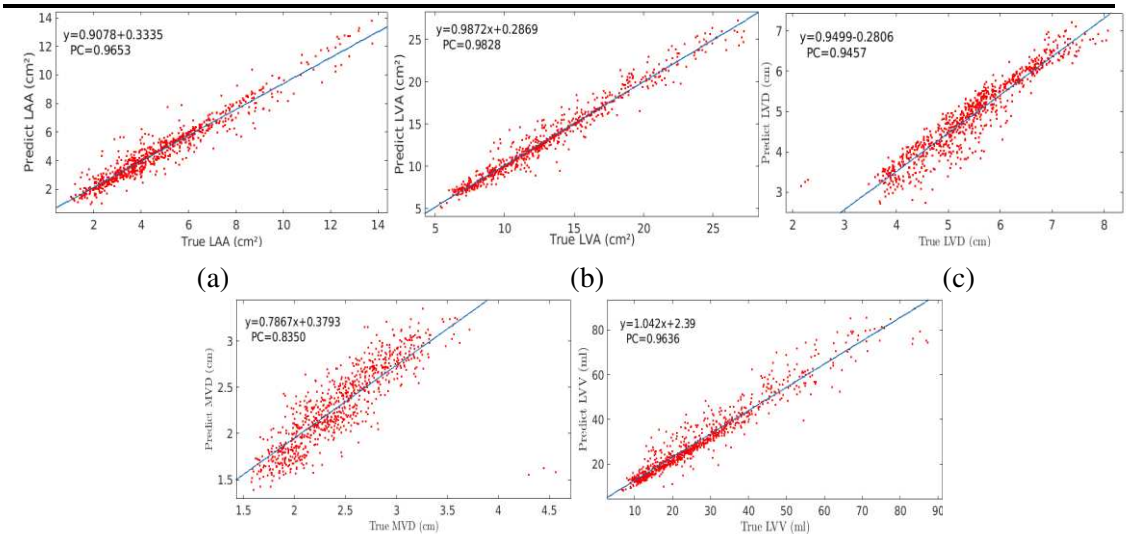


Figure 10: Correlation plots for clinical parameters (a) Correlation between actual and predicted LAA. (b) Correlation between actual and predicted LVA. (c) Correlation between true and predicted LVD. (d) Correlation between True and predicted MVD. (e) Correlation between actual and predicted LVV.

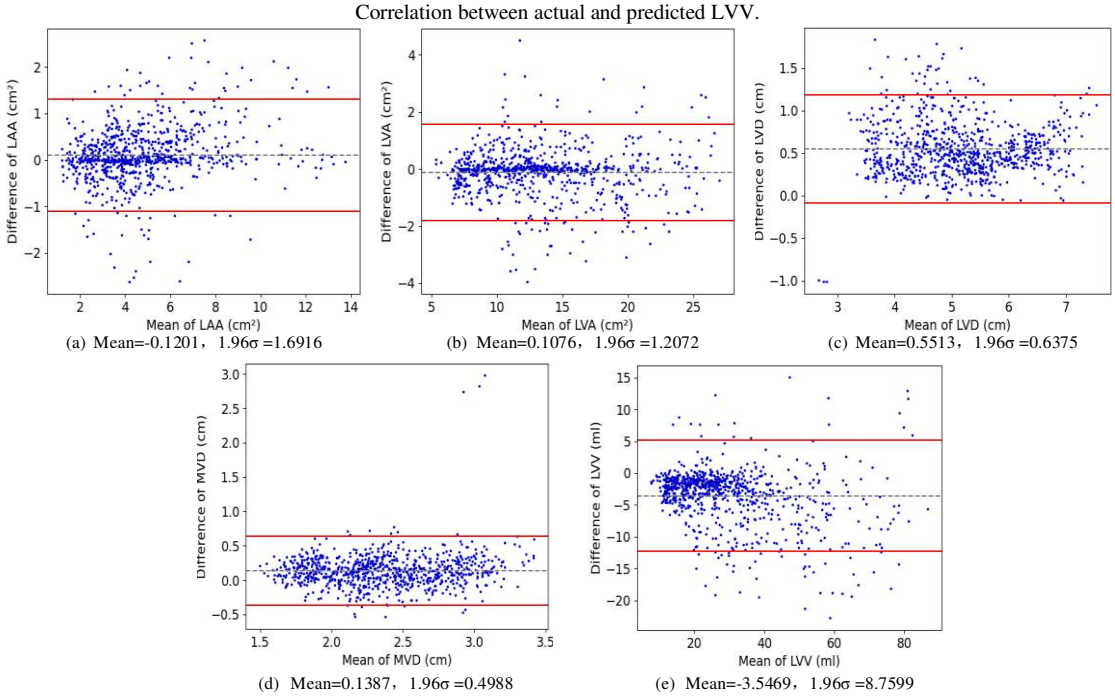


Figure 11: Bland Altman graphs for cardiac functional parameters. (a) Bland Altman plot between actual and predicted LVA. (b) Bland Altman plot between actual and predicted LAA. (c) Bland Altman plot between actual and predicted LVD. (d) Bland Altman plot between actual and predicted MVD. (e) Bland Altman plot between actual and predicted LVV. The x-axis represents

the mean of actual and predicted clinical parameters, and the y-axis represents the difference between actual and predicted clinical parameters. σ: standard deviation of bias between actual and estimated clinical parameters. Some points far away from the red line are the reason for the poor segmentation result, which leads to large errors in the measurement results.

5 Discussion

To accurately segment the 4CH view of paediatric echocardiography and automatically measure clinical parameters, we propose a new FCN based segmentation method to segment the LA, LV and APT. We calculate the cardiac functional parameters according to the segmentation results. These methods have achieved good results in segmenting LV and LA because the area of the ventricle and the atrium is more extensive and the cut has a clear boundary. However, the LA segmentation result is worse than the LV segmentation result because the size changes throughout the cardiac cycle, and there are more noise and artefacts on the border.

Figure 12 is a radar chart of the comparison of network segmentation performance. The first row and column are the segmentation performance comparisons of the LV and LA segmentation tasks using various modules of our method. Our network has achieved the best segmentation performance. The first row and the second column are the segmentation performance comparisons of various network modules in the APT segmentation task. The Dice coefficient is gradually increasing, but the Recall and Precision changes are relatively large. The first column of the second row is the comparison between the proposed network and the classic segmentation network in the LV and LA segmentation task. The second row and column are the comparison of the segmentation performance of the proposed network in the APT segmentation task using the classic network. The proposed network has achieved the best segmentation performance.

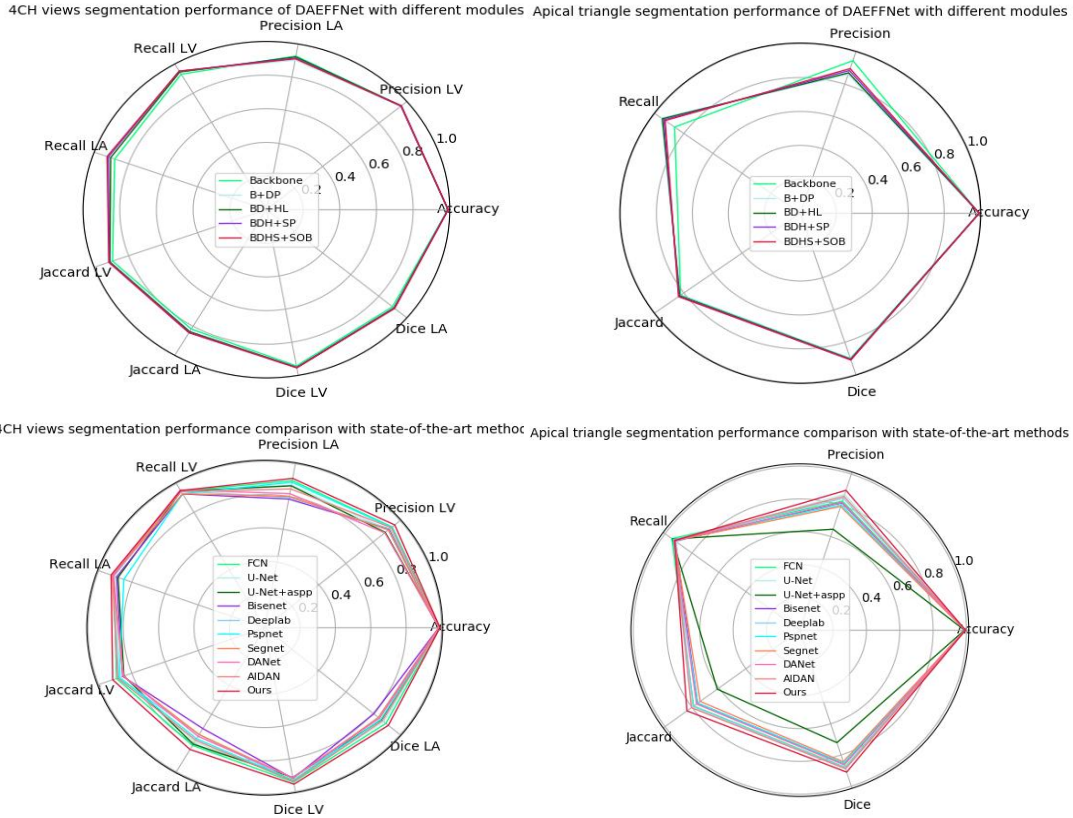


Figure 12: Radar chart results of the segmentation performance comparison using different networks in different segmentation tasks.

Figure 13 is a line chart of clinical parameters, a is LVA, b is LAA, c is LVD, d is MVD, and e is LVV. The red square is the prediction result, and the blue diamond is the labelled data. The test data is a video of a subject with 47 frames. The changes in cardiac functional parameters with the cardiac cycle can be seen in the figure. The prediction result of LVA is better, and LAA is a little bit worse. It can be seen that the size of the LA changes more obviously with the cardiac cycle. The prediction result of LVD is generally lower than the labelling result because the endpoint segmentation is not very good when segmenting APT, which also leads to the change of MVD prediction result not entirely obvious. The Sen coefficient is relatively low. The predicted value of LVV is not significantly different from the labelled data, which indicates that good results have been achieved.

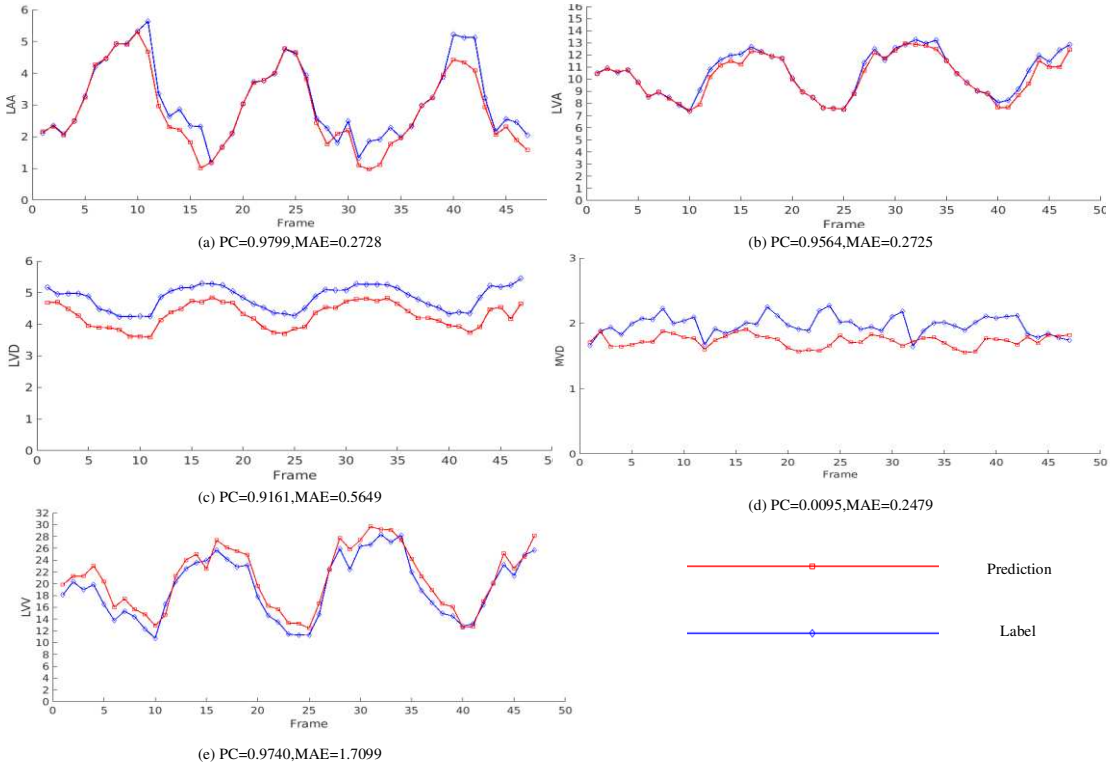


Figure 13: Line chart of clinical parameters. The red square is the prediction result, and the blue diamond is the labelled data.

Figure 14 shows the positioning APT and its optimisation process. The first column is the segmentation result, the second column is the three vertices positioned according to the original positioning method, and the third column is the visualisation result of the three vertices. We found that the lowest point of some segmentation results was not an endpoint of the mitral valve, so the original positioning method was optimised. The fourth column is the three vertices of the APT after optimisation, the fifth column is the visualisation of the three vertices after optimisation, and the last column is the visualisation of the APT based on the label.

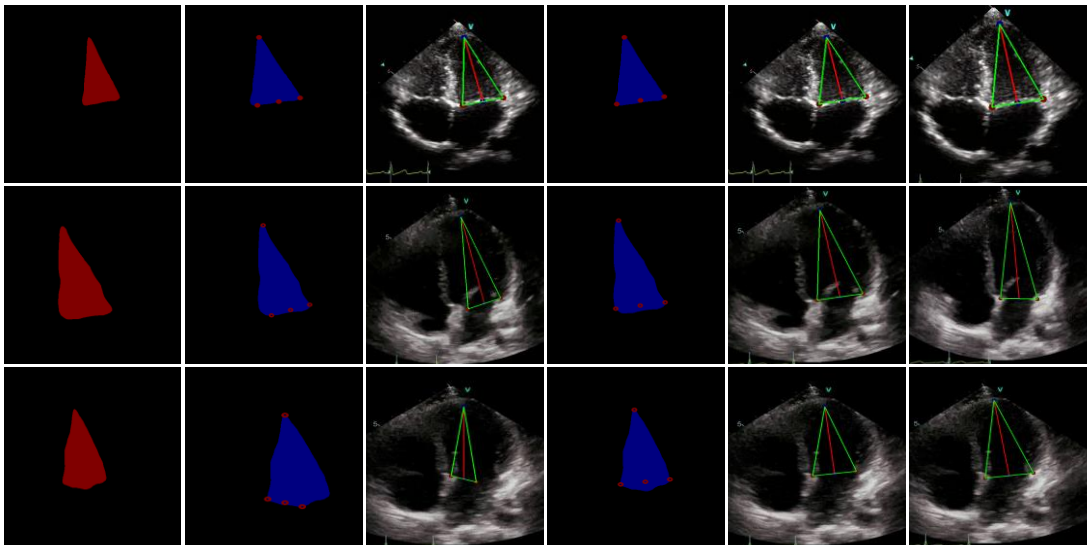


Figure 14: Correction diagram results. The first column is the segmentation result. The second column is the original method locating the three vertices and the base midpoint. The third column is the APT using these three fixed points; the red line is the long

neck; the bottom side is the inner diameter of the mitral valve annulus. The fourth column is the three vertices positioned after correction, and the fifth column is the APT drawn based on the corrected points.

6 Conclusions

In this paper, a network with enhanced feature fusion is proposed for paediatric echocardiographic segmentation. In the encoding part, a DP-FEM is presented, which can learn more channel feature information while widening the network and enhance the feature extraction ability of the network. In the decoding part, we propose the use of an HL-FFM. This module uses a semantic feature extraction module to learn the rich semantic information in the high-level features, and then merge it with the spatial information in the low-level features. At the bottom of the network, the Sub-Pixel convolution is used for upsampling, and spatial information is further merged with semantic information through spatial reshaping. Finally, the Sobel loss function is used to adjust the boundary. In the two segmentation tasks, the proposed method achieves the best Dice coefficient results.

To further realise the intelligent analysis of paediatric echocardiography, we calculate the cardiac functional parameters necessary for measurement based on the segmentation results. We propose a new critical point location method, which can measure LVD and MVD according to the segmentation results, and optimise the process of parameter measurement. Although we achieve good measurement results, these results depend heavily on the quality of the segmentation results of APT. Therefore, improving the accuracy of the segmentation results is still the most critical requirement of automatic measurement.

Acknowledgements

This work was supported partly by National Natural Science Foundation of China (Nos.62071309, 61871274, 61801305 and 81571758), National Natural Science Foundation of Guangdong Province (No.2019A1515111205), Shenzhen Key Basic Research Project (Nos. JCYJ20170818094109846, JCYJ20180507184647636, JCYJ20190808155618806, GJHZ20190822095414576, and JCYJ201908 08145011259). AFF is partially supported by a Royal Academy of Engineering Chair in Emerging Technologies Scheme (CiET1819/19) and a Pengcheng Visiting Scholars Programme from the Shenzhen Government. † indicates joint first author. The asterisk indicates the corresponding author.

References

Andreassen, B.S., Veronesi, F., Gerard, O., Solberg, A.H.S., Samset, E., 2019. Mitral Annulus Segmentation Using Deep Learning in 3-D Transesophageal Echocardiography. IEEE Journal of Biomedical and Health Informatics 24, 994-1003.

Arafati, A., Morisawa, D., Avendi, M.R., Amini, M.R., Assadi, R.A., Jafarkhani, H., Kheradvar, A., 2020. sGeneralisable fully automated multi-label segmentation of four-chamber view echocardiograms based on deep convolutional adversarial networks. Journal of the Royal Society Interface 17, 20200267.

Ba, JL, Kiros, J.R., Hinton, G.E., 2016. Layer normalisation. arXiv preprint arXiv:1607.06450.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39, 2481-2495.

Cao, Y., Xu, J., Lin, S., Wei, F., Hu, H., 2019. Gcnet: Non-local networks meet squeeze-excitation networks and beyond, Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 0-0.

Chaudhari, S., Polatkan, G., Ramanath, R., Mithal, V., 2019. An attentive survey of attention models. arXiv preprint arXiv:1904.02874.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

Cheng, H.K., Chung, J., Tai, Y.-W., Tang, C.-K., 2020. CascadePSP: Toward Class-Agnostic and Very High-Resolution Segmentation via Global and Local Refinement, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8890-8899.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation, International conference on medical image computing and computer-assisted intervention. Springer, pp. 424-432.

Copel, J.A., Pilu, G., Green, J., Hobbins, J.C., Kleinman, C.S., 1987. Fetal echocardiographic screening for congenital heart disease: the importance of the four-chamber view. American journal of obstetrics and gynecology 157, 648-655.

Dong, J., Liu, S., Liao, Y., Wen, H., Lei, B., Li, S., Wang, T., 2019. A generic quality control framework for fetal ultrasound cardiac four-chamber planes. IEEE Journal of Biomedical and Health Informatics 24, 931-942.

Du, X., Tang, R., Yin, S., Zhang, Y., Li, S., 2018. Direct segmentation-based full quantification for left ventricle via deep multi-task regression learning network. IEEE journal of biomedical and health informatics 23, 942-948.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3146-3154.

Gahungu, N., Trueick, R., Bhat, S., Sengupta, P.P., Dwivedi, G., 2020. Current Challenges and Recent Updates in Artificial Intelligence and Echocardiography. Current Cardiovascular Imaging Reports 13, 5.

Ge, R., Yang, G., Chen, Y., Luo, L., Feng, C., Zhang, H., Li, S., 2019. PV-LVNet: Direct left ventricle multitype indices estimation from 2D echocardiograms of paired apical views with deep neural networks. Medical image analysis 58, 101554.

Greenspan, H., Van Ginneken, B., Summers, R.M., 2016. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35, 1153-1159.

Howell, H.B., Zaccario, M., Kazmi, S.H., Desai, P., Sklamberg, F.E., Mally, P., 2019. Neurodevelopmental outcomes of children with congenital heart disease: A review. Current Problems in Pediatric and Adolescent Health Care 49, 100685.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141.

Hu, Y., Xia, B., Mao, M., Jin, Z., Du, J., Guo, L., Frangi, A.F., Lei, B., Wang, T., 2020. AIDAN: An Attention-Guided Dual-Path Network for Pediatric Echocardiography Segmentation. Ieee Access 8, 29176-29187.

Kanopoulos, N., Vasanthavada, N., Baker, R.L., 1988. Design of an image edge detection filter using the Sobel operator. IEEE Journal of solid-state circuits 23, 358-367.

Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., 2015. 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 16, 233-271.

Leclerc, S., Smistad, E., Grenier, T., Lartizien, C., Ostvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.-M., 2019a. RU-Net: A refining segmentation network for 2D echocardiography, 2019 IEEE International Ultrasonics Symposium (IUS). IEEE, pp. 1160-1163.

Leclerc, S., Smistad, E., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Belhamissi, M., Israilov, S., Grenier, T., 2020. LU-Net: a multi-stage attention network to improve the robustness of segmentation of left ventricular structures in 2D echocardiography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.

Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.-M., Grenier, T., 2019b. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE transactions on medical imaging 38, 2198-2210.

Li, X., Wang, W., Hu, X., Yang, J., 2019. Selective kernel networks, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 510-519.

Lin, Z., Li, S., Ni, D., Liao, Y., Wen, H., Du, J., Chen, S., Wang, T., Lei, B., 2019. Multi-task learning for quality assessment of fetal head ultrasound images. Medical image analysis 58, 101548.

Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T., 2019. Deep learning in medical ultrasound analysis: a review. Engineering 5, 261-275.

Liu, T., Tian, Y., Zhao, S., Huang, X., Wang, Q., 2020. Residual Convolutional Neural Network for Cardiac Image Segmentation and Heart Disease Diagnosis. IEEE Access 8, 82153-82161.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440.

Lopez, L., Colan, S.D., Frommelt, P.C., Ensing, G.J., Kendall, K., Younoszai, A.K., Lai, W.W., Geva, T., 2010. Recommendations for quantification methods during the performance of a pediatric echocardiogram: a report from the Pediatric Measurements Writing Group of the American Society of Echocardiography Pediatric and Congenital Heart Disease Council. Journal of the American Society of Echocardiography 23, 465-495.

Mendis, S., Puska, P., Norrving, B., Organization, W.H., 2011. Global atlas on cardiovascular disease prevention and control. World Health Organization.

Metaxas, D., Chen, T., Huang, X., Axel, L., 2004. Cardiac segmentation from MRI-tagged and CT images, 8th WSEAS International Conf. on Computers, special session on Imaging and Image Processing of Dynamic Processes in biology and medicine, p. 1.

Mishra, D., Chaudhury, S., Sarkar, M., Soin, AS, 2018. Ultrasound image segmentation: a deeply supervised network with attention to boundaries. IEEE Transactions on Biomedical Engineering 66, 1637-1648.

Moradi, S., Oghli, M.G., Alizadehasl, A., Shiri, I., Oveisi, N., Oveisi, M., Maleki, M., Dhooge, J., 2019. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. Physica Medica 67, 58-69.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines, ICML.

Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., 2020. Video-based AI for beat-to-beat assessment of cardiac function. Nature 580, 252-256.

Parisi, A., Moynihan, P., Feldman, C., Folland, E., 1979. Approaches to determination of left ventricular volume and ejection fraction by real‐time two‐dimensional echocardiography. Clinical cardiology 2, 257-263.

Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F., 2016. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magnetic Resonance Materials in Physics, Biology and Medicine 29, 155-195.

Pu, B., Zhu, N., Li, K., Li, S., 2020. Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework. Future Generation Computer Systems.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234-241.

Schiller, N.B., Shah, P.M., Crawford, M., DeMaria, A., Devereux, R., Feigenbaum, H., Gutgesell, H., Reichek, N., Sahn, D., Schnittger, I., 1989. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. Journal of the American Society of Echocardiography 2, 358-367.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1874-1883.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Smistad, E., Østvik, A., 2017. 2D left ventricle segmentation using deep learning, 2017 IEEE international ultrasonics symposium (IUS). IEEE, pp. 1-4.

Sultan, M.S., Martins, N., Costa, E., Veiga, D., Ferreira, M.J., Mattos, S., Coimbra, M.T., 2018. Virtual m-mode for echocardiography: A new approach for the segmentation of the anterior mitral leaflet. IEEE Journal of Biomedical and Health Informatics 23, 305-313.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. ECA-net: Efficient channel attention for deep convolutional neural networks, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11534-11542.

Wang, W., Wang, Y., Wu, Y., Lin, T., Li, S., Chen, B., 2019. Quantification of full left ventricular metrics via deep regression learning with contour-guidance. IEEE Access 7, 47918-47928.

Wang, X., Girshick, R., Gupta, A., He, K., 2018. Non-local neural networks, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7794-7803.

Woo, S., Park, J., Lee, J.-Y., So Kweon, I., 2018. Cbam: Convolutional block attention module, Proceedings of the European conference on computer vision (ECCV), pp. 3-19.

Wu, L., Cheng, J.-Z., Li, S., Lei, B., Wang, T., Ni, D., 2017. FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks. IEEE transactions on cybernetics 47, 1336-1349.

Xu, L., Liu, M., Zhang, J., He, Y., 2020. Convolutional-Neural-Network-Based Approach for Segmentation of Apical Four-Chamber View from Fetal Echocardiography. IEEE Access 8, 80437-80446.

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N., 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation, Proceedings of the European conference on computer vision (ECCV), pp. 325-341.

Zhang, Z., Zhang, X., Peng, C., Xue, X., Sun, J., 2018. Exfuse: Enhancing feature fusion for semantic segmentation, Proceedings of the European Conference on Computer Vision (ECCV), pp. 269-284.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881-2890.

Zhao, Q.-M., Liu, F., Wu, L., Ma, X.-J., Niu, C., Huang, G.-Y., 2019. Prevalence of congenital heart disease at live birth in China. The Journal of pediatrics 204, 53-58.