



OPEN

# Graph matching and deep neural networks based whole heart and great vessel segmentation in congenital heart disease

Zeyang Yao<sup>1,2</sup>, Wen Xie<sup>1,2</sup>, Jiawei Zhang<sup>3</sup>, Haiyun Yuan<sup>2</sup>, Meiping Huang<sup>2</sup>, Yiyu Shi<sup>4</sup>, Xiaowei Xu<sup>2</sup>✉ & Jian Zhuang<sup>1,2</sup>✉

Congenital heart disease (CHD) is one of the leading causes of mortality among birth defects, and due to significant variations in the whole heart and great vessel, automatic CHD segmentation using CT images has been always under-researched. Even though some segmentation algorithms have been developed in the literature, none perform very well under the complex structure of CHD. To deal with the challenges, we take advantage of deep learning in processing regular structures and graph algorithms in dealing with large variations and propose a framework combining both the whole heart and great vessel segmentation in complex CHD. We benefit from deep learning in segmenting the four chambers and myocardium based on the blood pool, and then we extract the connection information and apply graph matching to determine the categories of all the vessels. Experimental results on 68 3D CT images covering 14 types of CHD illustrate our framework can increase the Dice score by 12% on average compared with the state-of-the-art whole heart and great vessel segmentation method in normal anatomy. We further introduce two cardiovascular imaging specialists to evaluate our results in the standard of the Van Praagh classification system, and achieves well performance in clinical evaluation. All these results may pave the way for the clinical use of our method in the incoming future.

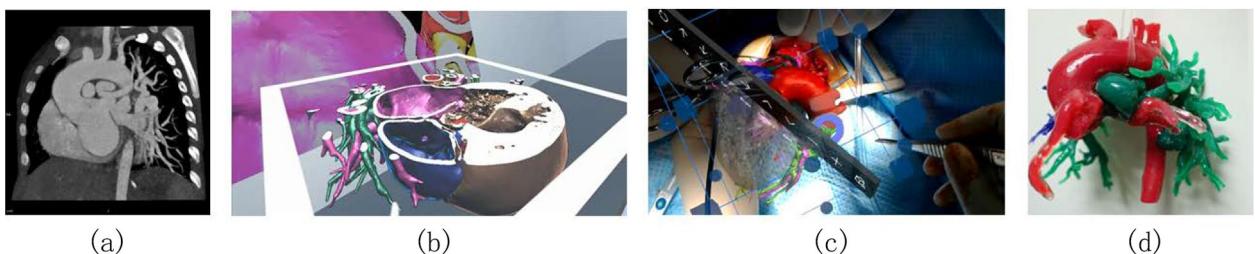
Congenital heart diseases (CHD) are defects in the structure of the heart or great vessels that present at birth<sup>1</sup>. The structure defects can involve the walls of the heart, the valves of the heart, and the arteries and veins near the heart. Each of these significant structure variations can disrupt the normal flow of blood through the heart, which cause serious health problems or death.

Clinical decision-making and surgery planning are the main challenges in the treatment of CHD. Currently, 2D (e.g., 2D echocardiography) and 3D (e.g., 3D echocardiography, magnetic resonance imaging (MRI), and computed tomography (CT)) imaging techniques are widely used to tackle these challenges. Furthermore, if surgery is required, CT and MRI are usually mandatory to help surgeons to have an informative knowledge of the heart structure. However, such a process is time-consuming and non-intuitive which introduces many difficulties to especially less experienced surgeons.

In recent years, 3D virtualization techniques (e.g., virtual reality (VR) and augmented reality (AR)) and 3D printing have been widely adopted for clinical decision-making and surgery planning of CHD<sup>2–5</sup>, greatly making up for the limitations of the traditional examination methods. As shown in Fig. 1, 3D medical images (usually CT or MRI) of hearts are usually used to produce a digital model of the heart which is then used for VR, AR, and 3D printing. Such virtualization of the heart can help surgeons have an intuitive sense of the structure and connections of the heart. To obtain the digital models, whole heart and great vessel segmentation are usually performed manually which is a time-consuming and costly process.

Nowadays, the development of deep learning has shown its advantages in the performance of segmentation<sup>6,7</sup>. There are some works in this area trying to solve the above problems<sup>8–11</sup>, and most of these works are based on

<sup>1</sup>School of Medicine, South China University of Technology, Guangzhou 510006, China. <sup>2</sup>Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou 510080, China. <sup>3</sup>School of Computer Science, Fudan University, Shanghai 200433, China. <sup>4</sup>Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA. ✉email: xiao.wei.xu@foxmail.com; zhuangjian5413@163.com



**Figure 1.** Illustration of 3D virtualization and 3D printing in CHD surgery: (a) is a typical coronal CT scanning of CHD; (b) is a horizontal section of heart segmentation results using virtualization reality (VR) for CHD surgery planning; (c) is a view in augmented reality (AR) for guiding real-time surgery; (d) is a printed model using 3D printing.

MRI data and focus on chambers and the whole heart only<sup>12–15</sup>. For congenital heart disease, Mukhopadhyay et al.<sup>16</sup> used total variation random forest to perform fully automatic congenital heart disease in magnetic resonance images (MRI) sequences. Li et al.<sup>11</sup> proposed a deep learning method for automatic whole-heart segmentation in cardiac magnetic resonance (CMR) images with CHD. The state-of-the-art segmentation performance in computed tomography (CT) images is obtained by Payer et al.<sup>8</sup>, they combine 3D U-Net for segmentation and a simple convolutional neural network for label position prediction. Recently there are also some works about blood pool segmentation of CHD images<sup>17,18</sup>, which only handles the blood pool and myocardium.

Moreover, Pace et al.<sup>19</sup> adopted a semi-automated approach for left ventricle (LV) and aorta (Ao) segmentation in CHD, they use a recurrent neural network recursively evolves segmentation in several steps, which requires user interaction to locate an initial seed for segmentation. Yoshida et al.<sup>20</sup> using U-net for investigating the usefulness of deep learning methods for segmenting the whole heart region and the cardiac cavity region in pediatric cardiac CT images. However their results are achieved on a small dataset and only the Dice similarity coefficient (DSC) is used for evaluating the segmentation accuracy, which is not suitable for CHD diagnosis to some extent. Pace et al.<sup>21</sup> proposed an iterative segmentation model and show that it can be accurately learned from a small MRI dataset, the model they proposed evolves a segmentation over multiple steps and achieved great performance on CHD segmentation. However, the characteristics of MRI images they used are completely different from those of CTA. In developing countries, the accessibility of CTA is far greater than that of MRI, and the study of the application of CTA in congenital heart disease is equally significant. All in all, fully automated segmentation of the whole heart and great vessel segmentation of CHD using CTA images is still an under-researched piece in the literature.

In this paper, we propose an algorithm that combines deep neural networks and graph matching for the whole heart and great vessel segmentation in CHD based on our previous work<sup>22</sup>. Note that graph matching has been applied in a variety of applications<sup>23–25</sup>, but not yet to congenital heart disease segmentation.

Particularly, we first use deep neural networks to segment four chambers and cardiac muscle, where variations are usually small. Then We extract the connection information and apply graph matching to determine the category of all the vessels. For making up for the shortcomings of Dice scores in evaluation segmentation results, two cardiovascular imaging specialists engaged to evaluate our results. Experiment results show that our method performs better than the state-of-the-art method for multi-modality whole heart and great vessel segmentation on both segmentation criterion and clinical evaluation. And as far as we know, this is the first work to perform the whole heart and great vessel segmentation and reconstruction in CHD. We hope all these contributions may pave the way to the clinical use of our method in the incoming future.

## Materials and methods

**Patient demographics.** The ages of the associated patients range from 1 month to 21 years, and the average age at repair is  $79.75 \pm 59.07$  weeks, with the majority between 1 month and 2 years. There are 40 females in the patients, approximately 60% of the total. Based on the degree of morphological variability of CHD<sup>26</sup>, we have divided our dataset into two categories: simple and complex. And there are 14 types of CHD, six simple types and eight complex ones, which all have been discussed in Table 1. Specifically the eight complex ones including Tetralogy of Fallot (TOF), transposition of great arteries (TGA), pulmonary artery sling (PAS), Anomalous Pulmonary Venous Connection (APVC), common arterial trunk (CAT), aortic arch anomalies (AAH), single ventricle (SV), pulmonary atresia ( $PA_{atresia}$ )). The number of images associated with each is summarized in Table 2.

**Image and acquisition parameters.** Our dataset consists of 68 3D CTA images, and all images were obtained by a SOMATOM Definition Flash Dual-source CT scanner, using the following protocol: collimation,  $(96\text{--}128) \times 0.625$  mm; rotation time, 270 ms, which corresponds to a 135 ms standard temporal resolution; slice thickness, 0.9 mm; reconstruction interval, 0.45 mm. Adaptive axial z-collimation was used to optimize the cranio-caudal length. Data were obtained at 40–50% of the RR interval, utilizing a 5% phase tolerance around the 45% phase. The dosage is 100 kVp/651 mAs (routine-dose). The size of the images is  $512 \times 512 \times (130\text{--}340)$ , and the typical voxel size is  $0.25 \times 0.25 \times 0.5$  mm<sup>3</sup>.

**Ground-truth labels generation.** All labeling was performed by experienced radiologists, and the time for labeling each image was 1–1.5 h. Within the range from the superior plane of the clavicle to the plane of the

| Types     | Description   |  |  |  |  |  |  |  |  |  |  |  |  |  |
|-----------|---|--|--|--|--|--|--|--|--|--|--|--|--|--|
| ASD       | A hole in the wall (septum) that divides the upper chambers (atria) of the heart  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| VSD       | A hole in the wall (septum) that separates the two lower chambers (ventricles) of the heart   |  |  |  |  |  |  |  |  |  |  |  |  |  |
| AVSD      | A large hole in center of the heart affecting all four chambers   |  |  |  |  |  |  |  |  |  |  |  |  |  |
| PDA       | The ductus arteriosus fails to close after birth  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CoA       | A part of the aorta, the tube that carries oxygen-rich blood to the body, is narrower than usual  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| PS        | A stenosis of main pulmonary artery and/or its branches   |  |  |  |  |  |  |  |  |  |  |  |  |  |
| TOF       | A heart birth defect that includes 4 defects: ventricular septal defect, right ventricle outflow tract stenosis, aorta overriding and secondary right ventricular hypertrophy |  |  |  |  |  |  |  |  |  |  |  |  |  |
| TGA       | The main pulmonary artery and the aorta are switched in position, or “transposed”   |  |  |  |  |  |  |  |  |  |  |  |  |  |
| PAS       | Left pulmonary artery originates from the right pulmonary artery, encircles the right main-stem bronchus, and distal trachea before entering the hilum of the left lung       |  |  |  |  |  |  |  |  |  |  |  |  |  |
| APVC      | The lungs pulmonary veins don't connect completely or partly to the left atrium like usual  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CAT       | A single common blood vessel comes out of the heart, instead of the main pulmonary artery and aorta   |  |  |  |  |  |  |  |  |  |  |  |  |  |
| IAA       | Aorta is not completely developed, part of the aorta is missing, leaving a gap  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| SV        | One lower chamber (ventricle) does not develop, and the heart has only one pumping chamber  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| PAatresia | The valve that controls blood flow from the heart to the lungs doesn't form at all  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Table 1.** Types of CHD and their descriptions.

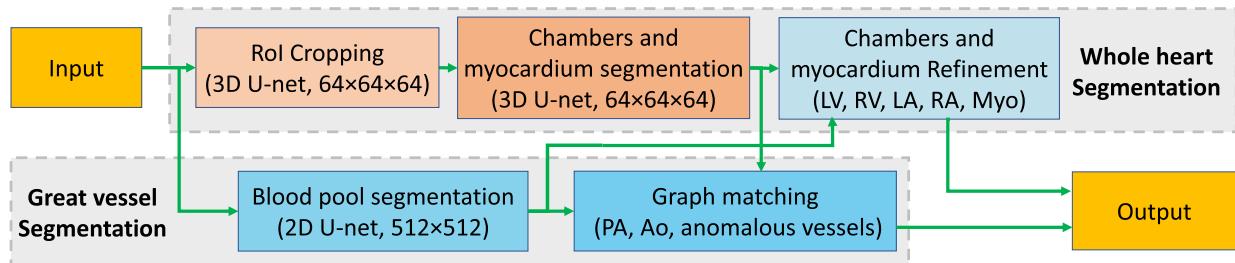
| Simple CHD 37 |      |     |     |     |    |     | Complex CHD 31 |     |      |     |     |    |                       | Normal |
|---------------|------|-----|-----|-----|----|-----|----------------|-----|------|-----|-----|----|-----------------------|--------|
| ASD           | AVSD | VSD | PDA | CoA | PS | TOF | TGA            | PAS | APVC | CAT | AAH | SV | PA <sub>atresia</sub> |        |
| 17            | 4    | 26  | 7   | 4   | 4  | 7   | 4              | 3   | 20   | 4   | 8   | 2  | 7                     | 2      |

**Table 2.** The types of CHD in our dataset and the case number associated with them. Note that some images may correspond to more than one type of CHD.

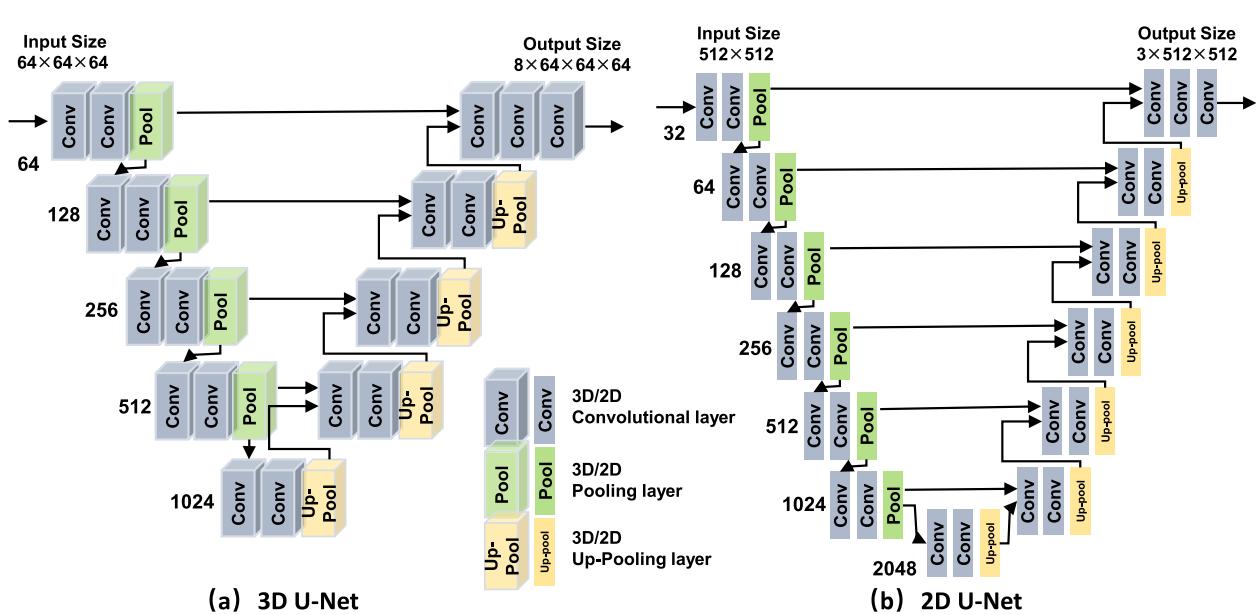
aortic hiatus of the diaphragm, we identified seven target cardiovascular structures: the left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), myocardium (Myo), aorta (Ao), and pulmonary artery (PA). During the procedure, the CTA images were imported into the Mimics software and labeled using thresholding. We assumed that the target and background of the image occupy different ranges of gray levels, with small differences in the gray values between adjacent pixels within the target and background. However, the pixels on both sides of the target-background interface have a large difference in gray value. Based on the prior knowledge of the anatomy recognized by the radiologist, we selected seven different gray threshold values to mark the anatomical structures. Then, we compared the gray values of each pixel in the image with these threshold values, and according to the comparison results, the pixels were classified into different categories. This process produced a continuous gray threshold image, and the target was extracted from the background. Next, we used the differences in characteristics such as gray level, color, and texture between the target and background to achieve cardiovascular edge detection. We first performed threshold segmentation and then mask editing. After these preprocessing steps, we used a three-dimensional reconstruction algorithm combining shear deformation to obtain the three-dimensional reconstruction of the cardiovascular model at different levels. Due to the large variability of the cardiovascular structures in congenital heart disease, vena cavae (VC) were also labeled as part of RA, and pulmonary veins (PV) were labeled as part of LA for ease of processing, as they are connected, and their boundaries are relatively hard to define. Anomalous vessels such as the delayed joining of the subclavian vein to the superior vena cava and abnormally enlarged pulmonary veins were also identified and classified as one of the seven aforementioned cardiovascular substructures based on their connections.

**Framework overview.** The whole heart and great vessel segmentation method have two sub-tasks: heart segmentation and great vessel segmentation. The overall framework is shown in Fig. 2.

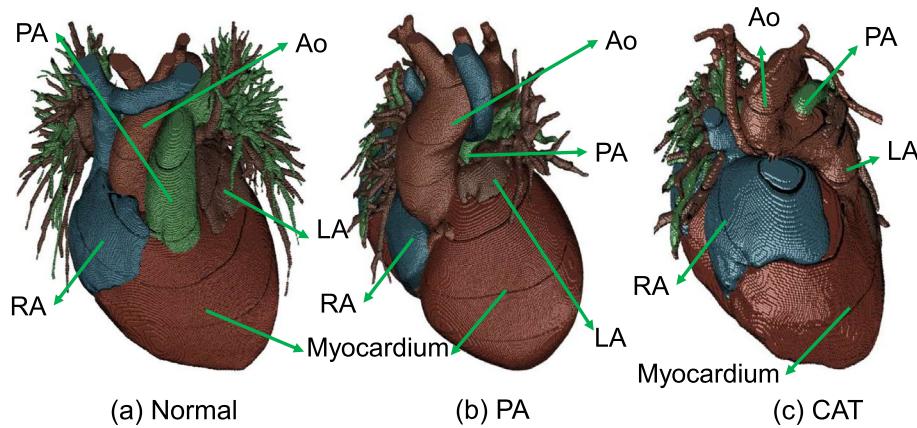
**Whole heart segmentation.** Segmentation was performed with multiple U-Nets<sup>27</sup>, and the network structure of the adopted 3D U-net and 2D U-net is shown in Fig. 3. There were several steps in segmentation: *Region of interest (RoI) cropping* extracts the area that includes the heart and its surrounding vessels. We resized the input image to a low resolution of  $64 \times 64 \times 64$ , and then adopt the same segmentation-based extraction as<sup>8</sup> to get the RoI. *Chambers and myocardium segmentation* resized the extracted RoI to  $64 \times 64 \times 64$  which had fed to a 3D U-net for segmentation. *Chambers and myocardium refinement* refined the boundaries of chambers and myocardium based on the outputs of chambers and myocardium segmentation and blood pool segmentation. Due to the limited GPU memory<sup>8</sup>, the input of 3D U-net was usually limited to low resolution or small size, and accordingly, the chambers and myocardium segmentation results may lose boundary information. This was critical for CHD where significant variations exist. To address this issue, we refined the boundary of chambers



**Figure 2.** Overview of the proposed framework combining deep learning and graph matching for whole heart and great vessel segmentation in CHD.

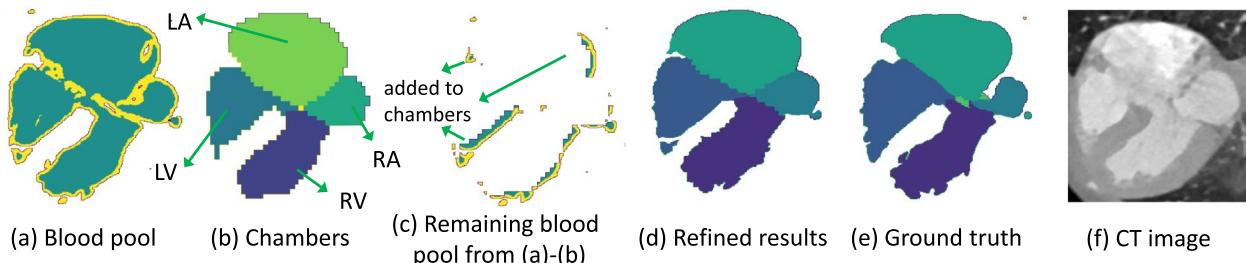


**Figure 3.** Network structures of the adopted 3D U-Net and 2D U-Net.

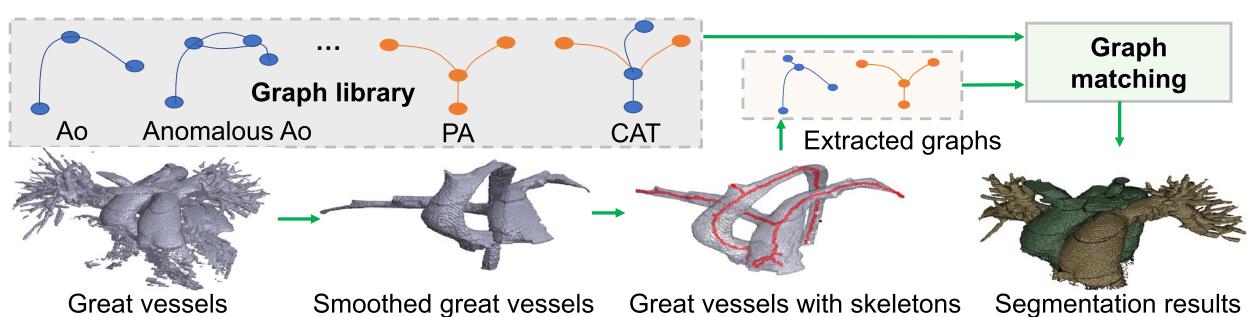


**Figure 4.** Pulmonary atresia ( $PA_{atresia}$ ) and common arterial trunk examples (CAT) in our dataset, with large variations from normal heart anatomy.

and myocardium by reusing the blood pool segmentation results, which were in high resolution. Specifically, we removed the portion of the blood pool that corresponded to the chambers from the results of blood pool segmentation, and the remaining blood pool was added to its surrounding chambers to refine the boundaries. With the refined boundary of chambers, the boundary of the myocardium was also refined as the chambers and the myocardium share a large portion of boundaries as shown in Fig. 4. An illustration of the refinement process



**Figure 5.** Illustration of chambers and myocardium refinement. (a) Is obtained from blood pool segmentation (high resolution). (b) Is from chambers and myocardium segmentation (low resolution). (c) Is the remaining blood pool by subtracting chambers (b) from blood pool (a). It is added to the surrounding chambers to refine the boundaries (d). (e,f) Are the ground truth and CT image, respectively.



**Figure 6.** Illustration of great vessel segmentation with graph matching. With smoothing, the skeleton of great vessels can be easily extracted, and then its corresponding graph is obtained for graph-matching-based classification of Ao, PA, and anomalous vessels.

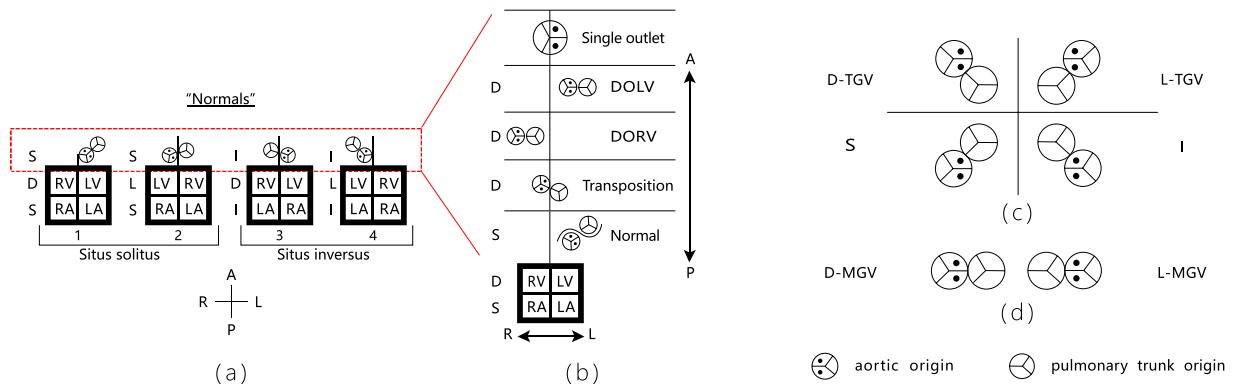
is shown in Fig. 5. Comparing (b) with (e), we can notice that part of the boundary information is lost, and the boundary is indeed refined after the process as shown in (d).

**Great vessel segmentation.** First, *blood pool segmentation* was conducted on each 2D slice of the input using a 2D U-net<sup>27</sup> with an input size of  $512 \times 512$ . Note that to detect the blood pool boundary for easy graph extraction in graph matching later, we added another class blood pool boundary in the segmentation. Second, *graph matching* identified Ao, PA, and anomalous vessels using the outputs of blood pool segmentation and chambers and myocardium segmentation. Great vessels could be obtained by removing the chamber areas from the blood pool, which need to be segmented to identify Ao, PA as well as anomalous vessels. To address significant variations that occurred in CHD, we adopt a surface thinning algorithm<sup>28</sup> to obtain skeletons of blood vessels for graph matching, and the workflow is shown in Fig. 6. A graph library is built based on medical knowledge to represent all the possible connections between great vessels and anomalous vessels. First, we illustrate all the possible combinations of PA and Ao, and the corresponding key points and their categories (Ao and PA) are labeled. We then extracted the graphs corresponding to Ao, PA, and anomalous vessels or their mixtures.

However, due to inaccurate blood pool segmentation or small anomalous connections, sometimes only one large graph corresponds to Ao and PA, which makes the matching difficult. Thus, these extracted graphs (corresponding to Ao and PA) should be disconnected from each other, and Ao and PA have their own corresponding graphs which can be matched with the ones in the library. To tackle this issue, we applied multiple smoothing in various scales (e.g., using a  $3 \times 3 \times 3$  kernels (all 1) 1 to 7 times iteratively for smoothing) to extract several candidate graphs. Then we matched these graphs with the ones in the library to identify the most similar pairs.

Particularly, we model the graphs as distributions, and such graph similarity is calculated using the earth mover's distance (EMD) which is a widely used similarity metric for distributions<sup>29</sup>. Two factors need to be modeled: the *weight* of each bin in the distribution, and the *distance* between bins. We model each sampled point in the sampled skeleton as a bin, the Euclidean distance between the points as the distance between bins, and the volume of blood pool around the sampled point as the weight of its corresponding bin. Particularly, the weight is defined as  $r^3$  where  $r$  is the radius of the inscribed sphere in the blood pool centered at the sampled point. With graph matching, the categories of the extracted graphs as well as the categories of the corresponding vessels in these graphs could be determined (based on the labeled graphs in the library). The vessels that left out in the smoothing process were finally classified by a simple region growing technique<sup>30</sup>.

**Van Praagh classification system.** A brief illustration of the Van Praagh classification system is shown in Fig. 7. In the Van Praagh system, a three-part notation consisting of letters separated by commas and encompassed by a set of braces is used to succinctly describe the visceroatrial situs, the orientation of the ventricular loop, and the position and relation of the great vessels. Van Praagh's symbolic representation may be combined



**Figure 7.** Illustration of clinical evaluation in the standard of Van Praagh classification system. (a) Is a model of four so-called normal hearts—that is, hearts with atrioventricular and ventriculoarterial concordant connections. The vertical line above the box represents the position of the ventricular septum. Note that in situs inversus, the aortic origin lies to the left of pulmonary trunk origin. The red dotted line represents the location of ventriculoarterial connections. (b) Is a model of varieties of ventriculoarterial connection. Aortic origin in the transposition of the great arteries and the double outlet ventricles (DORV, DOLV) is indicated by D when it lies to the right of the pulmonary artery origin and L when it lies to the left. (c) Schematically show the great vessel configurations on axial cross-sectional images at the level of valves when in a normal position (left column) or inversion (right column) and in normal relation (bottom row) or transposition (top row). (d) Schematically show d-malposition (top) and l-malposition (bottom) of the great vessels.

with those of atrioventricular and ventricular arterial connections as shown in Fig. 7a–d. In Van Praagh's convention, the first letter (S or I) refers to the atrial position (solitus or inversus), the second letter (D or L) to the ventricular loop, and the third letter to the position of the origin of the aorta (recognized by its two coronary ostia) in relation to the origin of the pulmonary trunk. The arrangement of boxes and abbreviations is identical in all similar models presented in Fig. 7.

**Experiments.** All the experiments were run on an Nvidia GTX 1080Ti GPU with 11 GB memory. We implemented our 3D U-net using Pytorch based on<sup>8</sup>. For 2D U-net, most configurations remain the same as those of the 3D U-net except that 2D U-net adopts 5 levels and the number of filters in the initial level was 16. Both Dice loss and cross-entropy loss were used, and the loss function is as follows:

$$L = \left( 1 - \frac{1}{C} \sum_{j=1}^C \frac{2 \sum_{i=1}^N p_i^j g_i^j}{\sum_{i=1}^N (p_i^j + g_i^j)} \right) + \frac{1}{NC} \sum_{j=1}^C \sum_{i=1}^N -g_i^j \log(p_i^j), \quad (1)$$

where  $p_i^j$  and  $g_i^j$  are the predicted and ground truth label maps at pixel  $i$  for the  $j$ th class. and the training epochs were 6 and 480 for 2D U-net and 3D U-net, respectively. Data augmentation was also adopted with the same configuration as in Ref.<sup>8</sup> for 3D U-net. Data normalization was the same as<sup>8</sup>. The learning rate was 0.0002 for the first 50% epochs, and then 0.00002 afterward.

Due to the scarcity of related CT CHD segmentation algorithms, our work is one of the pioneers in exploring and addressing this challenging task, we adopted Seg-CNN<sup>8</sup> that achieves the state-of-the-art performance in whole heart and great vessel segmentation of hearts in normal anatomy for comparison. The configuration was the same as that in Ref.<sup>8</sup>.

For both methods, fourfold cross-validation was performed (17 images for testing and 51 images for training). The split of our dataset considers the structures of CHD so that any structure in the testing dataset also had a similar presence in the training dataset, though they may be not of the same type of CHD.

The Dice score was used for segmentation evaluation. For evaluation, we adopt both general segmentation criteria and clinical evaluation for the assessment of our method. The Dice score was used as a general segmentation criterion for fourfold cross-validation. For clinical evaluation, two cardiovascular imaging specialists were introduced to evaluate our segmentation results in the standard of Van Praagh classification system<sup>31–33</sup>.

**Ethics declarations.** All data analyzed were collected as part of routine examination and diagnosis, and this data collection involves no procedures for which written consent is normally required outside the research context. All analysis procedures performed in studies involving human participants' data were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. And the informed consent is waived under the approval of the Research Ethics Committee of Guangdong General Hospital, Guangdong Academy of Medical Science under Protocol No. 20140316.

| Method     | LV                                | RV                                | LA                               | RA                               | Myo                              | Ao                               | PA                                | Overall                           |
|------------|-----------------------------------|-----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------------|-----------------------------------|
| Seg-CNN    | $67.3 \pm 13.9$                   | $65.0 \pm 12.0$                   | $70.2 \pm 7.8$                   | $76.0 \pm 7.5$                   | $71.5 \pm 8.3$                   | $63.0 \pm 13.3$                  | $52.3 \pm 12.3$                   | $66.5 \pm 10.7$                   |
| Our method | <b><math>82.4 \pm 10.5</math></b> | <b><math>77.6 \pm 14.3</math></b> | <b><math>78.6 \pm 7.4</math></b> | <b><math>82.7 \pm 7.5</math></b> | <b><math>77.3 \pm 8.3</math></b> | <b><math>82.2 \pm 8.1</math></b> | <b><math>67.1 \pm 19.8</math></b> | <b><math>78.3 \pm 10.8</math></b> |
| t-value    | 7.15                              | 5.57                              | 6.44                             | 5.21                             | 4.07                             | 10.17                            | 5.24                              | 6.40                              |
| P          | < 0.05                            | < 0.05                            | < 0.05                           | < 0.05                           | < 0.05                           | < 0.05                           | < 0.05                            | < 0.05                            |

**Table 3.** Mean and standard deviation of Dice score, and paired t-test of between the state-of-the-art method Seg-CNN<sup>8</sup> and our method (in %) for seven substructures of the whole heart and great vessel segmentation. Significant values are in bold.

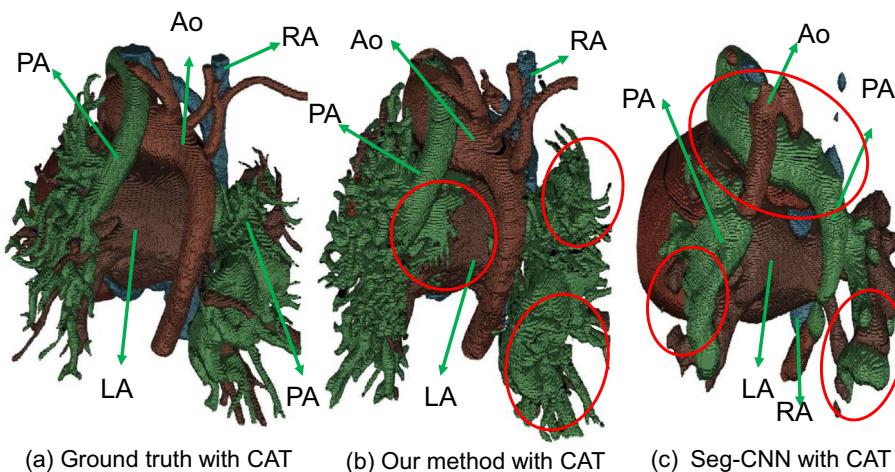
## Discussion

**General segmentation results.** The comparison with Seg-CNN<sup>8</sup> is shown in Table 3. Our method achieves a 5.8–19.2% higher mean Dice score across all the seven substructures (12% higher on average). The highest improvement is attained in Ao, which is due to its simple graph connection with successful graph matching. The least improvement is obtained in the myocardium, which is because the myocardium is not well considered in the high-resolution blood pool segmentation. Visualization of CAT segmentation using our method and Seg-CNN is shown in Fig. 8. Our method is capable of accurately segmenting the Ao and PA, with only minor instances of incorrect segmentation between the PA and LA. However, Seg-CNN wrongly segments the main part of the Ao as PA due to the limitations of pixel-level segmentation, in which the U-net-based framework only takes into account the surrounding pixels and fails to adequately incorporate connection information.

The segmentation performances of Seg-CNN<sup>8</sup> and our method in different scenarios (simple and complex CHD)<sup>26</sup> are shown in Table 4. Both methods achieve higher mean Dice with lower standard deviations in simple CHD than in complex CHD, as complex CHD has more complicated structure variations. Compared with Seg-CNN, our method achieves about 12% higher average mean Dice score on both simple and complex CHD. Our method also achieves a 1.9% reduction on the standard deviation of Dice score in simple CHD compared with Seg-CNN<sup>8</sup>.

Finally, visualizations of segmentation results from our method with the best, median, and worst Dice scores among all the test images are shown in Fig. 9. The segmentation results in Fig. 9a achieve the best accuracy, and most of the structures are segmented correctly, with some errors in the connections between RA and Ao, and the wrong connections are indicated by the red circle. The segmentation results in Fig. 9b,c have some severe wrong segmentations: the one in Fig. 9b has an anomalous vein from RA which is segmented as part of Ao due to the boundary extraction error in blood pool segmentation, and the one in Fig. 9c suffers from the boundary extraction error between LA and PA. This type of error also results in the worst Dice score as shown in Fig. 9d, which corresponds to ground truth provided in Fig. 9d. In the ground truth, a thick anomalous vein from RA crosses Ao, and PA has no trunk vessels and is of a very small volume. Compared with the ground truth, the thick anomalous vein from RA is misclassified as PA, and the majority of PA is misclassified as LA.

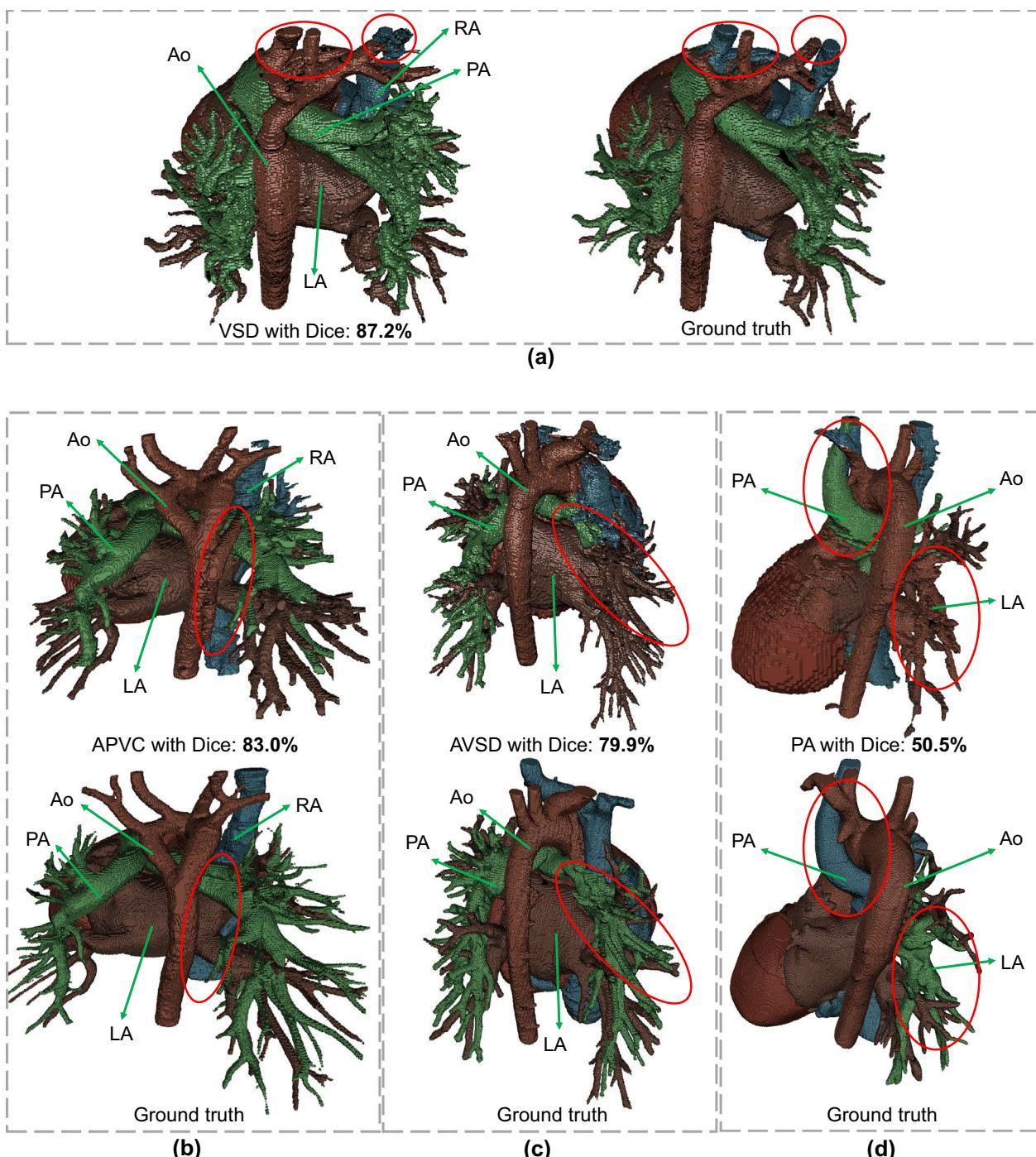
**Clinical evaluation results.** The detail of clinical evaluation results are shown in Table 5. All results were assessed independently by two cardiovascular imaging specialists, and discordant reads were reviewed and discussed until a consensus was reached. Our method had not made mistakes in some simple CHDs like ASD, AVSD, VSD, PS, and CoA. Specifically, we found the position of the atrium of a PDA case changed from situs sol-



**Figure 8.** Visualized comparison between the state-of-the-art method Seg-CNN<sup>8</sup> and our method. The differences from the ground truth are highlighted by the red circles.

| Types       | Seg-CNN         | Our method                        | t-value | P      |
|-------------|-----------------|-----------------------------------|---------|--------|
| Simple CHD  | $70.3 \pm 8.3$  | <b><math>82.6 \pm 6.2</math></b>  | 6.2     | < 0.05 |
| Complex CHD | $62.7 \pm 14.4$ | <b><math>74.1 \pm 14.5</math></b> | 14.5    | < 0.05 |

**Table 4.** Mean and standard deviation of Dice score, and paired t-test of between the state-of-the-art method Seg-CNN<sup>8</sup> and our method (in %) in simple and complex CHDs. Significant values are in bold.



**Figure 9.** Visualization of our segmentation results with (a) best, (b,c) median, and (d) worst Dice score among all the test images. Their corresponding ground truth is also shown, and red circles indicate the segmentation error.

| Types             | Total cases | Wrong cases | Description                     |
|-------------------|-------------|-------------|---------------------------------|
| ASD               | 17          | 0           |                                 |
| AVSD              | 4           | 0           |                                 |
| VSD               | 26          | 0           |                                 |
| PDA               | 7           | 1           | $\{S \rightarrow I, D, S\}$     |
| CoA               | 4           | 0           |                                 |
| PS                | 4           | 0           |                                 |
| TOF               | 7           | 1           | $\{S, D \rightarrow L, S\}$     |
| TGA               | 4           | 1           | $\{S, D, S \rightarrow L-TGV\}$ |
| PAS               | 3           | 0           |                                 |
| APVC              | 20          | 0           |                                 |
| CAT               | 4           | 0           |                                 |
| AAH               | 8           | 0           |                                 |
| SV                | 2           | 1           | $\{S, X \rightarrow D, S\}$     |
| <i>PA</i> atresia | 7           | 0           |                                 |
| Normal            | 2           | 0           |                                 |

**Table 5.** Results of clinical evaluation. *Total cases* gives all case number of each type CHD, *Wrong cases* shows wrong segmentation results number, and *Description* presents the specific type of the wrong segmentation, with bold and black italics indicating specific wrong parts. Divided by types according to Table 2.

itus to situs inversus. More common wrong segmentations happened in complex CHD. In a TOF case, we found the ventricular loop changed from curved rightward to curved leftward. Interestingly we found the position of the aortic origin changed with pulmonary trunk origin, the aorta was anterior to and leftward of the pulmonary trunk origin, the anomaly is described as levotransposition (l-transposition) or congenitally corrected transposition, which is shown as L-TGV in Fig. 5. In a single ventricle case, we found the ventricle was segmented into the left and right two parts. The overall percentage of correct connections in atrioventricular, ventriculoarterial, and the relationship of the great vessels is 0.986 (70/71), 0.97 (69/71), and 0.986 (70/71).

Our method achieves much better performance than the state-of-the-art segmentation method. However, in the 3D segmentation process, although each type of CHD shown in Fig. 9 achieves good accuracy under the general segmentation criterion, there are still some tiny wrong-connected segmentations. Thus far, we have considered connection features in the blood pool and considered the shapes of the vessels to improve our segmentation results. However, the presence of certain defects may be attributed to the failure of boundary extraction in blood pool segmentation. From a data perspective, the initial threshold-based construction of ground-truth labels also limits the differentiation of boundary and boundary extraction, which is also a limitation of our study and may contribute to the occurrence of these errors in segmentation. In order to enhance performance, additional structural features should be taken into account, such as local tissue changes. This requires innovative approaches from the deep learning community and deeper collaboration between computer scientists and radiologists.

Unless we can achieve excellent performance in every segmentation result, using the Dice score to evaluate our method is not a satisfactory choice. Due to the uncertainty and incomprehensibility of the U-Nets network, pixel-level segmentation criterion may result in some segmentation parts being consistent with the ground truth, while some critical parts are not. It is easy to understand from Fig. 5. In chambers and myocardium refinement, either boundaries from high-resolution segmentation results or low-resolution segment wrong will cause refined boundaries do not consistent with the ground truth, and introduce connection mutations to segmentation results. When it comes to 3D printing and virtual surgery planning, the impact of those connection mutations is serious. The existence of some basic anatomical errors is not acceptable clinically, it is necessary to introduce another criterion clinically to guarantee there is not an atrium defect from segmentation compared with the ground truth.

However, there are still no recognized clinical criteria to instruct the training process of the segmentation network. In the standard of Van Praagh classification system, our two cardiovascular imaging specialists introduced a new clinical evaluation to assess our method. Because the Van Praagh notation imposes on its users a systematic approach to anatomic description, it is a helpful device for structuring the interpretation of segmentation results as well as the reporting of evaluation. By manual evaluation of the segmentation results of each case, we roughly proved the clinical availability of our method. From detailed check results from Table 5, we can notice that most types of CHD achieve correct segmentation, which gives us the confidence to further complete our method. In view of wrong segmentation still existing, it is not realistic to let clinicians evaluate every automatic segmentation result. Even though specialists find some wrong connections in PDA, will only affect imaging analysis to some degree. However, wrong connections happened in complex CHD like TOF, TGA, and SV, are very different from the ground truths. As our goal is to achieve fully automatic segmentation and reconstruction of the CHD model that can guide surgery, the lack of clinical-level criterion will seriously affect the future development of our method. Although our clinical evaluation used in this paper is not excellent, we will try to solve those problems in future work.

## Conclusion

In this paper, we proposed a whole heart and great vessel segmentation framework for CT images in CHD. We first used deep learning to segment the four chambers and myocardium followed by the blood pool, where variations were usually small. We then extracted the connection information and applied graph matching to determine the categories of all the vessels. We collected a CHD dataset in CT with 68 3D images, and the ground truth has seven categories: LV, RV, LA, RA, myocardium, Ao, and PA. 14 types of CHD are included in this dataset which is made publicly available. Compared with the state-of-the-art method for whole heart and great vessel segmentation in normal anatomy, our method has achieved 12% improvement in Dice score on average, and well performance when evaluated by clinical evaluation. We hope these results may pave the way for the clinical use of our method in the incoming future. Due to the limited GPU memory, currently, we used a 3D U-net to process a down-sampled CT image, which cannot segment the related structures precisely. In the future, with more advanced GPUs, we may try to process CT images with much larger dimensions. A fusion (or an ensemble) network should be further adapted to take the outputs of the 2D U-net and the 3D U-net and output the final segmentation. In addition, a two-stage training strategy should be applied in which the 2D U-net and the 3D U-net are firstly trained, and then all three networks are fine-tuned together. In this way, the segmentation accuracy can be further improved potentially.

## Data availability

The datasets generated during and/or analyzed during the current study are not publicly available due to patient privacy protection but are available from the corresponding author upon reasonable request.

## Code availability

Correspondence and requests for materials and code could be addressed to Dr. Xiaowei Xu upon reasonable request.

Received: 17 May 2022; Accepted: 22 April 2023

Published online: 09 May 2023

## References

- Hoffman, J. I. & Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39**, 1890–1900 (2002).
- Yoo, S.-J. *et al.* 3d printing in medicine of congenital heart diseases. *3D Print. Med.* **2**, 3 (2016).
- Sun, Z., Lau, I., Wong, Y. H. & Yeong, C. H. Personalized three-dimensional printed models in congenital heart disease. *J. Clin. Med.* **8**, 522 (2019).
- Bramlet, M., Olivieri, L., Farooqi, K., Ripley, B. & Coakley, M. Impact of three-dimensional printing on the study and treatment of congenital heart disease. *Circ. Res.* **120**, 904–907 (2017).
- Hussein, N. *et al.* Development and validation of a procedure-specific assessment tool for hands-on surgical training in congenital heart surgery. *J. Thorac. Cardiovasc. Surg.* **160**, 229–240 (2020).
- Xu, X. *et al.* Edge segmentation: Empowering mobile telemedicine with compressed cellular neural networks. In *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* 880–887 (IEEE, 2017).
- Qi, C. R., Su, H., Mo, K. & Guibas, L. J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 652–660 (2017).
- Payer, C., Štern, D., Bischof, H. & Urschler, M. Multi-label whole heart segmentation using cnns and anatomical label configurations. In *International Workshop on Statistical Atlases and Computational Models of the Heart* 190–198 (Springer, 2017).
- Wang, C., MacGillivray, T., Macnaught, G., Yang, G. & Newby, D. A two-stage 3d unet framework for multi-class segmentation on full resolution image. Preprint at <http://arxiv.org/abs/1804.04341> (2018).
- Xu, Z., Wu, Z. & Feng, J. Cfnn: Combining faster r-cnn and u-net network for efficient whole heart segmentation. Preprint at <http://arxiv.org/abs/1812.04914> (2018).
- Li, J., Zhang, R., Shi, L. & Wang, D. Automatic whole-heart segmentation in congenital heart disease using deeply-supervised 3d fcn. In *Reconstruction, Segmentation, and Analysis of Medical Images* (eds Poudel, R. P. *et al.*) 111–118 (Springer, 2016).
- Zhuang, X. & Shen, J. Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. *Med. Image Anal.* **31**, 77–87 (2016).
- Torrents-Barrena, J. *et al.* Segmentation and classification in mri and us fetal imaging: Recent trends and future prospects. *Med. Image Anal.* **51**, 61–88 (2019).
- Jang, Y., Hong, Y., Ha, S., Kim, S. & Chang, H.-J. Automatic segmentation of lv and rv in cardiac mri. In *International Workshop on Statistical Atlases and Computational Models of the Heart* 161–169 (Springer, 2017).
- Heinrich, M. P. & Oster, J. Mri whole heart segmentation using discrete nonlinear registration and fast non-local fusion. In *International Workshop on Statistical Atlases and Computational Models of the Heart* 233–241 (Springer, 2017).
- Mukhopadhyay, A. Total variation random forest: Fully automatic mri segmentation in congenital heart diseases. In *Reconstruction, Segmentation, and Analysis of Medical Images* (eds Poudel, R. P. *et al.*) 165–171 (Springer, 2016).
- Wolterink, J. M., Leiner, T., Viergever, M. A. & Isgum, I. Dilated convolutional neural networks for cardiovascular mr segmentation in congenital heart disease. In *Reconstruction, Segmentation, and Analysis of Medical Images* (eds Poudel, R. P. *et al.*) 95–102 (Springer, 2016).
- Yu, L., Yang, X., Qin, J. & Heng, P.-A. 3d fractalnet: Dense volumetric segmentation for cardiovascular mri volumes. In *Reconstruction, Segmentation, and Analysis of Medical Images* (eds Poudel, R. P. *et al.*) 103–110 (Springer, 2016).
- Pace, D. F. *et al.* Iterative segmentation from limited training data: Applications to congenital heart disease. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (eds Stoyanov, D. P. *et al.*) 334–342 (Springer, 2018).
- Yoshida, A. *et al.* Automated heart segmentation using u-net in pediatric cardiac ct. *Meas. Sens.* **18**, 100127 (2021).
- Pace, D. F. *et al.* Learned iterative segmentation of highly variable anatomy from limited data: Applications to whole heart segmentation for congenital heart disease. *Med. Image Anal.* **80**, 102469 (2022).
- Xu, X. *et al.* Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 477–485 (Springer, 2019).
- Lajevardi, S. M., Arakala, A., Davis, S. A. & Horadam, K. J. Retina verification system based on biometric graph matching. *IEEE Trans. Image Process.* **22**, 3625–3635 (2013).
- Das, D. & Lee, C. G. Graph matching and pseudo-label guided deep unsupervised domain adaptation. In *International Conference on Artificial Neural Networks* 342–352 (Springer, 2018).

25. Emmert-Streib, F., Dehmer, M. & Shi, Y. Fifty years of graph matching, network alignment and network comparison. *Inf. Sci.* **346**, 180–197 (2016).
26. Powell, A. J. & Geva, T. Simple and complex congenital heart disease: Infants and children. In *Cardiovascular Magnetic Resonance* (eds Kwong, R. Y. et al.) 469–485 (Elsevier, 2019).
27. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241 (Springer, 2015).
28. Lee, T.-C., Kashyap, R. L. & Chu, C.-N. Building skeleton models via 3-d medial surface axis thinning algorithms. *Graph. Models Image Process.* **56**, 462–478 (1994).
29. Rubner, Y., Tomasi, C. & Guibas, L. J. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* **40**, 99–121 (2000).
30. Adams, R. & Bischof, L. Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 641–647 (1994).
31. Van Praagh, R., Van Praagh, S., Vlad, P. & Keith, J. D. Anatomic types of congenital dextrocardia: Diagnostic and embryologic implications. *Am. J. Cardiol.* **13**, 510–531 (1964).
32. Van Praagh, R. The importance of segmental situs in the diagnosis of congenital heart disease. *Semin. Roentgenol.* **20**, 254–271 (1985).
33. Van Praagh, R. Terminology of congenital heart disease. Glossary and commentary. *Circulation* **56**, 139–143 (1977).

## Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2018YFC1002600), the Science and Technology Planning Project of Guangdong Province, China (No. 2017B090904034, No. 2017B030314109, No. 2018B090944002, No. 2019B020230003), Guangdong Peak Project (No. DFJH201802), the National Natural Science Foundation of China (No. 62006050).

## Author contributions

X.X. and J.Z. designed the study. Z.Y., W.X. and J.Z. conducted the whole experiment(s). H.Y., M.H. and Y.S. analyzed the results and prepared all tables. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to X.X. or J.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023