

Research Article

The Application of Knowledge Distillation toward Fine-Grained Segmentation for Three-Vessel View of Fetal Heart Ultrasound Images

Qiwen Cai ¹, Ran Chen,² Lu Li,³ Chao Huang,⁴ Haisu Pang,⁵ Yuanshi Tian,² Min Di,² Mingxuan Zhang,² Mingming Ma,² Dexing Kong,¹ and Bowen Zhao ²

¹School of Mathematical Sciences, Zhejiang University, Hangzhou, China

²Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

³College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China

⁴Ningbo First Hospital, Ningbo, China

⁵The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Correspondence should be addressed to Bowen Zhao; zbwjcp@zju.edu.cn

Received 29 March 2022; Revised 29 May 2022; Accepted 30 June 2022; Published 14 July 2022

Academic Editor: Guangming Zhang

Copyright © 2022 Qiwen Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objectives. Measuring anatomical parameters in fetal heart ultrasound images is crucial for the diagnosis of congenital heart disease (CHD), which is highly dependent on the clinical experience of the sonographer. To address this challenge, we propose an automated segmentation method using the channel-wise knowledge distillation technique. **Methods.** We design a teacher-student architecture to conduct channel-wise knowledge distillation. ROI-based cropped images and full-size images are used for the teacher and student models, respectively. It allows the student model to have both the fine-grained segmentation capability inherited from the teacher model and the ability to handle full-size test images. A total of 1,300 fetal heart ultrasound images of three-vessel view were collected and annotated by experienced doctors for training, validation, and testing. **Results.** We use three evaluation protocols to quantitatively evaluate the segmentation accuracy: Intersection over Union (IoU), Pixel Accuracy (PA), and Dice coefficient (Dice). We achieved better results than related methods on all evaluation metrics. In comparison with DeepLabv3+, the proposed method gets more accurate segmentation boundaries and has performance gains of 1.8% on mean IoU (66.8% to 68.6%), 2.2% on mean PA (79.2% to 81.4%), and 1.2% on mean Dice (80.1% to 81.3%). **Conclusions.** Our segmentation method could identify the anatomical structure in three-vessel view of fetal heart ultrasound images. Both quantitative and visual analyses show that the proposed method significantly outperforms the related methods in terms of segmentation results.

1. Introduction

It is reported that the annual birth rate of congenital heart disease (CHD) in children is 0.8–1% [1, 2], and the incidence rate has ranked first among all birth defects. At present, fetal echocardiography is an important medical imaging technology and most used for prenatal detection and diagnosis of CHD. Well-trained and experienced doctors can make a reliable diagnosis of 80–90% CHD by fetal echocardiography, while the diagnostic accuracy of doctors who lack experience in fetal

echocardiography operation and diagnosis is significantly decreased [3, 4].

The three-vessel and trachea (3VT) view is one of the fetal echocardiography views, which was introduced by Yagel et al. [5] as a complementary cardiac view to easily assess the aortic arch anomalies. Gireadă et al. [6] performed a retrospective study on 1,596 unselected pregnant patients presenting at 11–37 weeks of gestation for a routine anomaly scan and analyzed the performance of the four-chamber (4C) view and 3VT view in detecting CHD. The results demonstrated 4C view detected 47.8% of all CHD, going up

to 71.7% by adding grayscale 3VT view. The 3VT view is deemed desirable if technically feasible by both International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) and American Institute of Ultrasound in Medicine (AIUM) screening guidelines [7, 8], especially due to its utility in detecting outflow tract anomalies [9–11].

Automatic image segmentation can guide operators to standardize and serialize the display and analysis of fetal echocardiography data. Semantic segmentation has been extensively studied for natural images and is expected to be applied to medical images to assist in diagnosis. However, unlike natural images, medical images are often more difficult to acquire and annotate, which results in a small amount of annotated data. In addition, the special imaging modality (e.g., ultrasound) of medical images produces images of low quality with noises and blurred boundaries. These factors often lead to unsatisfactory segmentation of images. For the segmentation of three-vessel view of fetal heart ultrasound images, we collected a total of 1,300 annotated images, which is completely incomparable to large natural image segmentation datasets.

Recently, machine learning methods have widely used in fetal image processing or ultrasound image processing, such as quality assessment [12–14], detection, and segmentation [15–17]. In the field of vessel image processing, physical models or properties are sometimes introduced to simulate realistic environment to obtain higher evaluation accuracy [18–20].

In this paper, we propose to utilize the channel-wise knowledge distillation [21, 22] technique toward fine-grained segmentation of three vessels in the 3VT view of fetal heart ultrasound images. Particularly, we first train a teacher model whose training data is precisely cropped to a region of three vessels. Such a teacher model enables fine-grained segmentation because it focuses only on the target region. However, such a model cannot be directly applied to the test data because we are not able to crop the test image to the target region before prediction. To this end, we train the student model by distilling the knowledge of the pretrained teacher model using full-size training data. This allows the student model to have both the fine-grained segmentation capability of the teacher model and the ability to handle full-size test images. Experiments show that the proposed method outperforms the most widely used existing methods by a significant margin.

2. Methods

2.1. Channel-Wise Knowledge Distillation. Knowledge distillation was originally designed for model compression: a compact student model is trained to perform better under the supervision of a large teacher model [23–25]. A subsequent study [26] has also shown that a student model with the same configuration as the teacher model can even exceed the performance of the teacher model. In this work, we employ the channel-wise knowledge distillation method, as it was shown to be more effective than the pixel-wise knowledge distillation that simply aligns point-wise classification scores per pixel [27, 28].

Channel-wise knowledge distillation was first introduced by Zhou et al. [21]. Its basic idea is to convert the feature map on each channel to a probability map and then align the channel-wise probabilities of the teacher model and the student model. The corresponding channel-wise distillation loss $\mathcal{L}_{CD}(\cdot, \cdot)$ is defined in the form of a KL divergence:

$$\mathcal{L}_{KD}(y^T, y^S) = \frac{\tau^2}{C} \sum_{c=1}^C \sum_{i=1}^{W \cdot H} \phi(y_{c,i}^T) \cdot \log \left[\frac{\phi(y_{c,i}^T)}{\phi(y_{c,i}^S)} \right], \quad (1)$$

where T and S denote the teacher and student models, $y^{(\cdot)}$ denotes the predicted logits of size $H \times W \times C$; i indexes the spatial location of pixels, where W and H are the width and height of the predicted logits; $c = 1, 2, \dots, C$ indexes the channel, where C is the number of classes including the background; τ is a hyperparameter called distillation temperature; $\phi(\cdot)$ is a channel-wise softmax function which converts the logits on each channel into a soft probability distribution, defined as

$$\phi(y_{c,i}) = \frac{\exp(y_{c,i}/\tau)}{\sum_{i=1}^{W \cdot H} \exp(y_{c,i}/\tau)}. \quad (2)$$

2.2. Our Method. In this paper, we use a model of the DeepLabv3+ [29] architecture with channel-wise knowledge distillation to segment three vessels in 2D fetal heart ultrasound images.

DeepLabv3+ is an encoder-decoder network designed for segmentation tasks. A modified aligned Xception model is used as the encoder, while the Atrous Spatial Pyramid Pooling (ASPP) layer is used as the decoder as the default setting of DeepLabv3+.

As shown in Figure 1, the teacher and student networks use the same architecture. According to the groundtruth label masks, we crop the region of interest (ROI, i.e., the region of three vessels) from full-size images to train the teacher model. Meanwhile, we record the coordinates of cropped regions to restore them to the corresponding spatial location in the full-size image afterwards. After training teacher network with cropped inputs, its logits output is used to distill knowledge on channels for student model training. This allows the student model to have the same ability to segment the ROI as the teacher model. For the training of the student model, we use full-size images as input and align the logits of the teacher model to the full-size ones as additional knowledge. The loss of our network is made up of three components:

$$L = \alpha \cdot \mathcal{L}_{CE}(y_{gt}, y_{pred}) + \beta \cdot \mathcal{L}_{Dice}(y_{gt}, y_{pred}) + \gamma \cdot \mathcal{L}_{KD}(y^T, y^S). \quad (3)$$

Here, $\mathcal{L}_{CE}(\cdot, \cdot)$ and $\mathcal{L}_{Dice}(\cdot, \cdot)$ denote cross-entropy loss and dice loss, respectively, y_{gt} is the groundtruth label mask, and y_{pred} is the predicted probability map showing the probability that each pixel being categorized in each class. $\mathcal{L}_{KD}(\cdot, \cdot)$ is defined in Section 2.1. Note that for the teacher network, γ is set to 0, since no knowledge distillation is used.

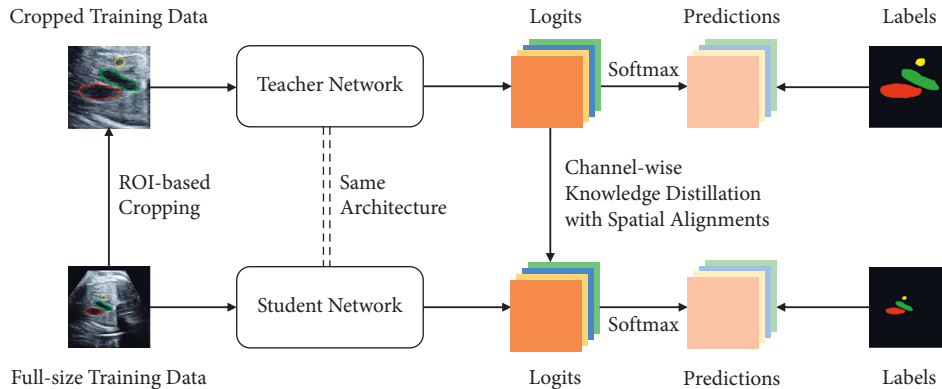


FIGURE 1: Overview of the proposed channel-wise knowledge distillation method.

2.3. Materials. 1300 pregnant women who underwent fetal echocardiography examination in Sir Run Run Shaw Hospital from 2016 to 2021 were randomly recruited. The gestational age was 20 to 40 weeks. The inclusion criteria were as follows: (1) normal fetus without heart or noncardiac malformations, especially there has no abnormality in the nine standard fetal echocardiography views posted by ISUOG; (2) low risks of chromosome abnormalities were confirmed by early NT and maternal serological examination, or amniocentesis and umbilical cord blood puncture; (3) gestational weeks estimated by ultrasound were consistent with those calculated by menopause history (difference < 2 weeks); and (4) the mothers who had no diabetes, hypertension, or pregnancy complications. All pregnant women were informed the purpose of this study and agreed that their fetal heart ultrasound related data should be used for scientific research, and consents should be signed.

Philips IE33 (Philips Medical Systems, Bothell, WA, USA) color ultrasound diagnostic instrument was used, with S5-1 2D imaging probe and a frequency of 1~6 MHz. Firstly, the fetal echocardiography mode was used to comprehensively evaluate the structure and function of the fetal heart, clearly displaying the standard three-vessel view images and storing the original data. Pregnant women are requested to hold their breath or reduce the range of breathing as much as possible during the whole collection process.

Seven experienced doctors annotated and reviewed the collected fetal heart ultrasound images. They annotated the pulmonary artery (PA), aorta (AO), and superior vena cava (SVC) regions on each image as the groundtruth. The resolution of the original image is 1024×768 . All images are simply center-cropped and resized to 512×512 as full-size training data in the student network. The ROI-based cropped images described in Section 2.2 are generated by padding the annotated three-vessel region with 50 pixels per edge and resizing them to a square region of size 512×512 . We selected 1040, 130, and 130 images in all experiments for training, validation, and testing, respectively.

Figure 2 shows two examples of the fetal heart ultrasound images, as well as their corresponding full-size labeled images and ROI-based cropped images.

2.4. Training. The proposed method is implemented on an NVIDIA RTX 3090 GPU using the Keras and TensorFlow frameworks and trained using the Adam optimizer. We train the network through two stages:

- (1) During the first stage, the initial learning rate is 0.0005 and the parameters are initialized by pre-training on the PASCAL VOC 2012 [30] dataset. All parameters of the Xception backbone are frozen. We set $\alpha = 1$, $\beta = 1$, and $\gamma = 0$. Since the main purpose of this stage is to speed up the training, we do not use the channel-wise knowledge distillation here. The batch size is set to 8.
- (2) During the second stage, the initial learning rate is 0.00005 and all layers are trainable. We set $\alpha = 1$, $\beta = 1$ for both teacher and student models, while the hyperparameters of the channel-wise knowledge distillation are $\gamma = 3$ and $\tau = 4$ and are used only in the student model. In equation (1), $C = 4$ as we have 4 classes including the background, PA, AO, and SVC. The batch size is set to 2, since backpropagation to the backbone layers requires a larger amount of GPU memory.

In both stages, we set the decayed learning rate as $lr = lr_{\text{init}} \cdot 0.92^{\text{iterations}-1}$ and train for 50 epochs with an early stopping setting when the validation loss stops decreasing.

The following random data augmentation methods are used in the training: scale transformation, displacement, flip, rotation, and color jittering.

3. Results

To evaluate the performance of the proposed method, we compare with the most used existing methods, including U-Net [31] and the original DeepLabv3+. U-Net is a classical network and is widely used in medical image segmentation. The original DeepLabv3+ is also adopted as our teacher network; we train it on full-size images for performance comparison. For ablation study, we train the models for comparison by using a combination of cross-entropy loss and dice loss without the knowledge distillation.

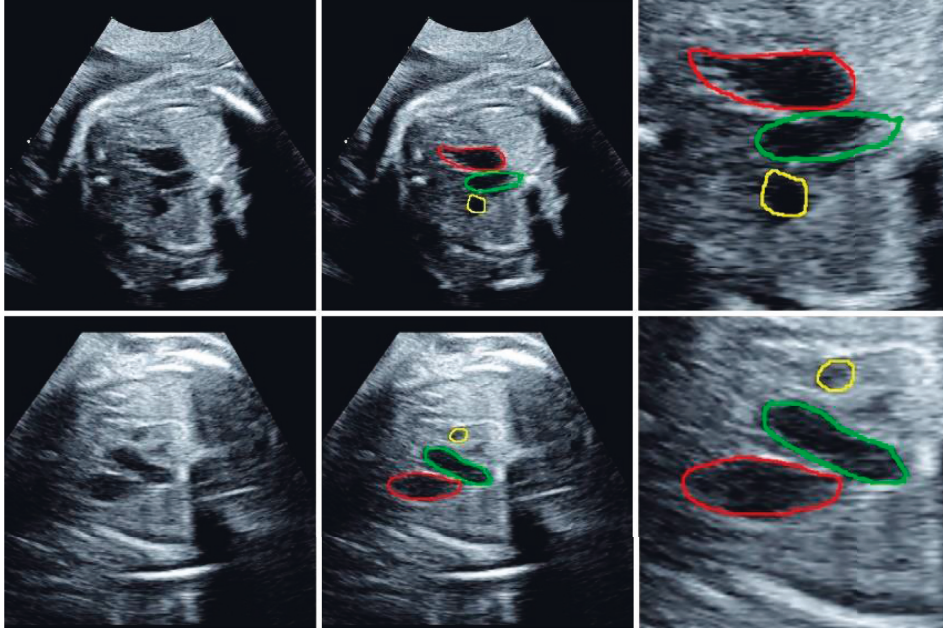


FIGURE 2: The 1st column shows original 3VT view of fetal heart ultrasound images. The 2nd column shows full-size labelled images, the red region is pulmonary artery (PA), the green region is aorta (AO), and the yellow region is superior vena cava (SVC). The 3rd column shows ROI-based cropped images, which are cropped from full-size images.

To quantitatively evaluate the segmentation accuracy, we use three evaluation protocols: Intersection over Union (IoU), Pixel Accuracy (PA), and Dice coefficient (Dice).

IoU:

$$\text{IoU} = \frac{|S_a \cap S_b|}{|S_a \cup S_b|} \times 100\%. \quad (4)$$

PA:

$$\text{PA} = \frac{|S_a \cap S_b|}{|S_a|} \times 100\%. \quad (5)$$

Dice:

$$\text{Dice} = \frac{2 \cdot |S_a \cap S_b|}{|S_a| + |S_b|} \times 100\%. \quad (6)$$

Here, S_a is the predicted segmentation result of the network and S_b is the groundtruth.

Table 1 shows the performance of our method. We present the IoU, PA, and Dice for each class evaluated on the test data. The mean value over three vessels' segmentation accuracies is shown in the bottom row. The mean IoU, PA, and Dice of the proposed method are 68.6%, 81.4%, and 81.3%, respectively.

Table 2 shows the comparison of three methods. The displayed measurements are the average of the three vessels' segmentation accuracies from all test data. It is evident that the proposed method performs better than the other methods.

To intuitively demonstrate the effectiveness of our results, we plot segmentation contours of the listed methods in Figure 3.

TABLE 1: Performance of the proposed method.

	IoU (%)	PA (%)	Dice (%)
Pulmonary artery (PA)	71.2	83.5	83.2
Aorta (AO)	69.7	82.8	82.1
Superior vena cava (SVC)	64.9	77.8	78.7
Mean	68.6	81.4	81.3

The segmentation accuracies of three vessels and their mean values are shown.

TABLE 2: Performance comparison to existing methods.

	IoU (%)	PA (%)	Dice (%)
U-Net	62.4	77.5	76.9
DeepLabv3+	66.8	79.2	80.1
Ours	68.6	81.4	81.3

The mean segmentation accuracy of three vessels over all test data is shown.

Specifically, we observed that when using existing segmentation methods, the SVC vessel is sometimes undetected or incorrectly segmented due to its small size. In addition, since the three vessels occupy only a small portion of the image, the existing methods only produce rough and incorrectly segmented boundaries.

4. Discussion

The incidence of birth defects in China is about 5.6% and the number of new birth defects is about 900,000 every year [1]. Amongst them, there are about 250,000 cases of birth defects that can be observed clinically at birth. Birth defects are the main causes of early abortion, stillbirth, perinatal death, infant death, and congenital disability, which not

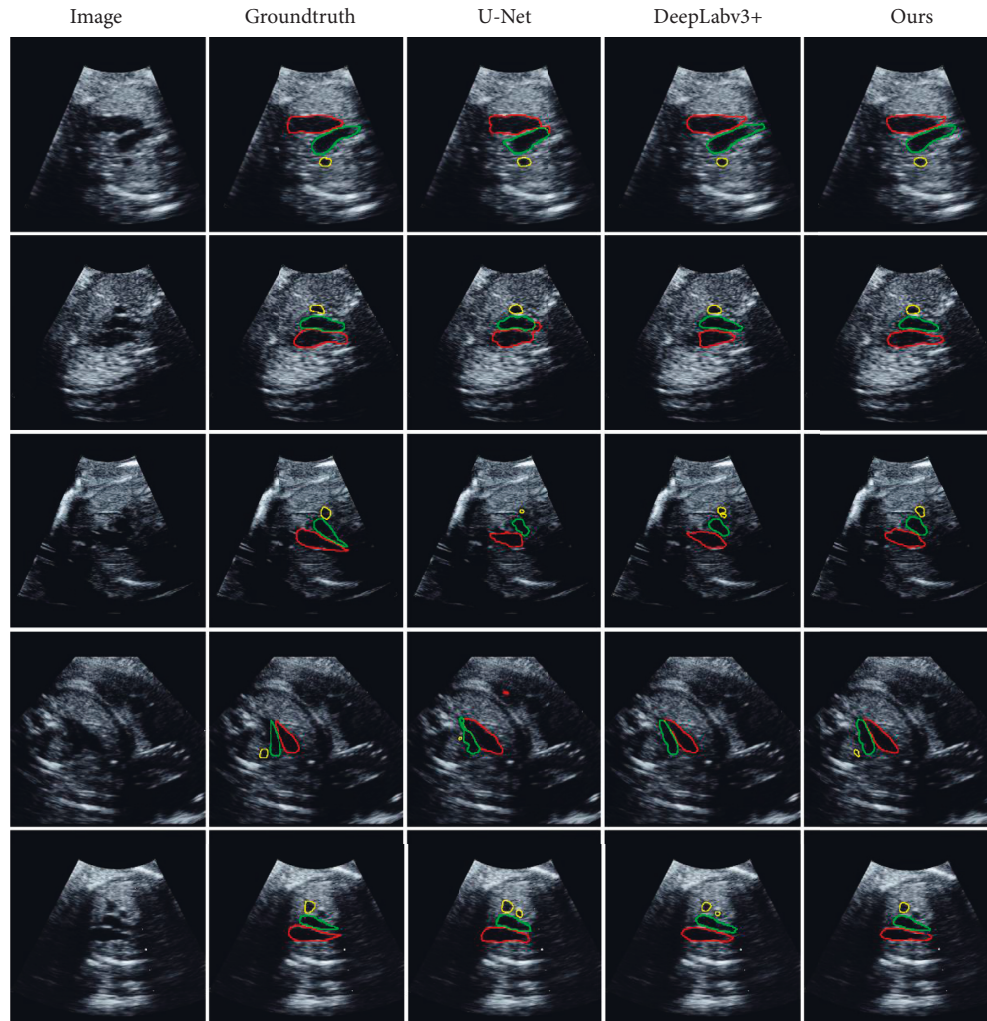


FIGURE 3: Segmentation results of different methods. The original 3VT view of fetal heart ultrasound images are shown in the first column and the groundtruth are shown in the second column. The segmentation results of U-Net, DeepLabv3+ and our proposed method are shown in the last three columns, respectively.

only seriously harm the survival and quality of life of children but also affect the happiness and harmony of families. It will also lead to potential life loss and increasing social and economic burden. Accurate prenatal diagnosis can significantly improve the perioperative treatment effect of CHD and the success rate of operation and reduce neonatal mortality.

Many studies have shown that improving the ability of routine obstetricians to recognize CHD is the most important issue to improve the successful rate of CHD prenatal diagnosis. However, doctors who complete obstetrical ultrasound diagnoses in their daily medical work are always lacking the diagnostic basis and evaluation experience of complex CHD, and it is difficult to obtain all basic views needed for CHD diagnosis [32–35], including four-chamber cardiac view, left and right ventricular outflow tract view, and three-vessel view, so it is difficult to make reliable display and diagnosis of CHD. As a result, CHD has become the most easily missed structural abnormality in prenatal routine ultrasound examination.

As we know, the 4C view is the most commonly used and easily obtained basic view in fetal heart examination. The acquisition rate at 16–40 weeks is around 95–99.5%, but four-chamber abnormalities only account for 48–63% of congenital cardiovascular malformations. A variety of fetal congenital cardiovascular malformations do not show abnormal shape of the four-chamber heart, including tetralogy of fallot, persistent truncus arteriosus, aortic valve stenosis, pulmonary valve stenosis, transposition of great arteries, and double outlet of ventricle. The extended basic views include three-vessel view, three-vessel trachea view, aortic arch view, pulmonary artery-ductus arteriosus arch view, and vein-atrium connection view. The detection rate of congenital cardiovascular malformation can be increased from 48–63% to 83–86%. Amongst which, the number, internal diameter, course, vascular arrangement, and abnormal blood flow direction of the major vessels can be clearly observed in the three-vessel and three-vessel trachea view, which plays an important role in the screening of fetal cardiac macrovascular malformations. Therefore, it is clinically important

to collect and identify the three vessels, and the automatic segmentation algorithm could help doctors to easily operate and accurately diagnose CHD.

Automated segmentation of fetal heart ultrasound images is a challenging task due to low signal-to-noise ratio, low contrast, and blurred boundaries. Collecting and accurately annotating the data is also a difficult task. We believe that the performance of the proposed model can be further improved by using a larger and more standard dataset. Nonetheless, we achieved better results than DeepLabv3+ on all evaluation metrics, with performance gains of 1.8% on mean IoU (66.8% to 68.6%), 2.2% on mean PA (79.2% to 81.4%), and 1.2% on mean Dice (80.1% to 81.3%).

5. Conclusion

In summary, we propose a fully automated segmentation method for fine-grained segmentation of three vessels in the 3VT view of fetal heart ultrasound images using a channel-wise knowledge distillation technique. We design a teacher-student architecture to distill channel-wise knowledge from ROI-based cropped images to full-size images. The logits output of the teacher model empowers the student model with fine-grained segmentation capability. In this way, we obtain more accurate segmentation boundaries. Both quantitative and visual analyses show that the proposed method significantly outperforms other methods in terms of segmentation results of the three vessels in the 3VT view of fetal heart ultrasound images.

Data Availability

The image data used to support the findings of this study have not been made available due to information that could compromise the privacy of research participants.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by Major Program of the National Natural Science Foundation of China (Grant no. 12090025) and the Natural Science Foundation of Zhejiang Province, China (Grant no. LGF22H180008).

References

- [1] C. Li, "Report on the prevention and treatment of birth defects in China," *Ministry of Health PRC*, vol. 3, no. 37, 2012.
- [2] S. C. Mitchell, S. B. Korones, and H. W. Berendes, "Congenital heart disease in 56,109 births incidence and natural history," *Circulation*, vol. 43, no. 3, pp. 323–332, 1971.
- [3] U. Gembruch and A. Geipel, "Indications for fetal echocardiography: screening in low-and high-risk populations," *Fetal cardiology*, vol. 96, 2003.
- [4] D. Paladini, "Expert consensus of fetal heart ultrasonography standardization," *Chinese Journal of Ultrasound Imaging*, vol. 20, no. 10, pp. 904–909, 2011.
- [5] S. Yagel, R. Arbel, E. Y. Anteby, D. Raveh, and R. Achiron, "The three vessels and trachea view (3VT) in fetal cardiac scanning," *Ultrasound in Obstetrics and Gynecology*, vol. 20, no. 4, pp. 340–345, 2002.
- [6] R. Gireadă, D. Socolov, E. Mihălceanu et al., "The additional role of the 3-vessels and trachea view in screening for congenital heart disease," *Medicina*, vol. 58, no. 2, p. 262, 2022 Feb 10.
- [7] J. Carvalho, L. D. Allan, R. Chaoui et al., "ISUOG Practice Guidelines (updated): sonographic screening examination of the fetal heart," *Ultrasound in Obstetrics and Gynecology*, vol. 41, no. 3, pp. 348–359, 2013.
- [8] J. Pellerito, B. Bromley, S. Allison et al., "AIUM-ACR-ACOG-SMFM-SRU practice parameter for the performance of standard diagnostic obstetric ultrasound examinations," *Journal of Ultrasound in Medicine*, vol. 37, no. 11, pp. E13–E24, 2018.
- [9] M. Pasternok, A. Nocun, A. Knafel et al., "'Y sign' at the level of the 3-vessel and trachea view: an effective fetal marker of aortic dextroposition anomalies in the first trimester," *Journal of Ultrasound in Medicine*, vol. 37, no. 8, pp. 1869–1880, 2018.
- [10] A. Palatnik, W. A. Grobman, L. S. Cohen, J. S. Dungan, and N. L. Gotteiner, "Role of the 3-vessel and trachea view in antenatal detection of Tetralogy of Fallot," *Journal of Ultrasound in Medicine*, vol. 35, no. 8, pp. 1799–1809, 2016.
- [11] H. Edwards and R. Hamilton, "Single centre audit of early impact of inclusion of the three vessel and trachea view in obstetric screening," *Ultrasound*, vol. 26, no. 2, pp. 93–100, 2018.
- [12] C. P. Bridge, C. Ioannou, and J. A. Noble, "Automated annotation and quantitative description of ultrasound videos of the fetal heart," *Medical Image Analysis*, vol. 36, pp. 147–161, 2017.
- [13] J. Dong, S. Liu, Y. Liao et al., "A generic quality control framework for fetal ultrasound cardiac four-chamber planes," *IEEE journal of biomedical and health informatics*, vol. 24, no. 4, pp. 931–942, 2020.
- [14] Z. Lin, S. Li, D. Ni et al., "Multi-task learning for quality assessment of fetal head ultrasound images," *Medical Image Analysis*, vol. 58, Article ID 101548, 2019.
- [15] X. Li, Y. Hong, D. Kong, and X. Zhang, "Automatic segmentation of levator hiatus from ultrasound images using U-net with dense connections," *Physics in Medicine and Biology*, vol. 64, no. 7, Article ID 075015, 2019.
- [16] S. Dong, G. Luo, C. Tam et al., "Deep atlas network for efficient 3D left ventricle segmentation on echocardiography," *Medical Image Analysis*, vol. 61, Article ID 101638, 2020.
- [17] L. Wang, L. Zhang, M. Zhu, X. Qi, and Z. Yi, "Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks," *Medical Image Analysis*, vol. 61, Article ID 101665, 2020.
- [18] Z. Gao, J. Chung, M. Abdelrazek et al., "Privileged modality distillation for vessel border detection in intracoronary imaging," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1524–1534, 2020.
- [19] Z. Gao, X. Wang, S. Sun et al., "Learning physical properties in complex visual scenes: an intelligent machine for perceiving blood flow dynamics from static CT angiography imaging," *Neural Networks*, vol. 123, pp. 82–93, 2020.
- [20] Z. Gao, H. Xiong, X. Liu et al., "Robust estimation of carotid artery wall motion using the elasticity-based state-space approach," *Medical Image Analysis*, vol. 37, pp. 1–21, 2017.

- [21] Z. Zhou, C. Zhuge, X. Guan, and W. Liu, "Channel Distillation: Channel-Wise Attention for Knowledge Distillation," 2020, <https://arxiv.org/abs/2006.01683>.
- [22] C. Shu, Y. Liu, J. Gao, Z. Yan, and C. Shen, "channel-wise knowledge distillation for dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5311–5320, Seoul Korea, June 2021.
- [23] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," 2015, <https://arxiv.org/abs/1503.02531>.
- [24] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer," 2016, <https://arxiv.org/abs/1612.03928>.
- [25] F. Tung and G. Mori, "Similarity-preserving knowledge distillation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1365–1374, Las Vegas, June 2019.
- [26] T. Furlanello, Z. Lipton, M. Tschannen, L. Itti, and A. Anandkumar, "Born again neural networks," *International Conference on Machine Learning*, PMLR, vol. 3, , pp. 1607–1616, 2018.
- [27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Las Vegas, June 2018.
- [28] Y. Wang, W. Huang, F. Sun, T. Xu, Y. Rong, and J. Huang, "Deep multimodal fusion by channel exchanging," *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [29] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, Seoul Korea, June 2018.
- [30] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge 2012 results," *Pattern Anal. Stat. Model. Comput. Learn*, vol. 5, 2012.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Cham, 2015 October 5.
- [32] J. Rychik, N. Ayres, B. Cuneo et al., "American Society of Echocardiography guidelines and standards for performance of the fetal echocardiogram," *Journal of the American Society of Echocardiography*, vol. 17, no. 7, pp. 803–810, 2004.
- [33] "Cardiac screening examination of the fetus: guidelines for performing the "basic" and "extended basic" cardiac scan," *Ultrasound in obstetrics & gynecology the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, vol. 27, no. 1, pp. 107–113, 2006.
- [34] W. Lee, L. Allan, J. S. Carvalho et al., "ISUOG consensus statement: what constitutes a fetal echocardiogram?" *Ultrasound in Obstetrics and Gynecology*, vol. 32, no. 2, pp. 239–242, 2008.
- [35] F. E. Force, "American college of obstetricians and gynecologists, society for maternal-fetal medicine. AIUM practice guideline for the performance of fetal echocardiography," *Journal of Ultrasound in Medicine: Official Journal of the American Institute of Ultrasound in Medicine*, vol. 30, no. 1, pp. 127–136, 2011.