



ORIGINAL RESEARCH

4D foetal cardiac ultrasound image detection based on deep learning with weakly supervised localisation for rapid diagnosis of evolving hypoplastic left heart syndrome

Gang Wang^{1,2,3}  | Weisheng Li^{1,4} | Mingliang Zhou⁵  | Haobo Zhu⁶ |
Guang Yang^{3,7,8} | Choon Hwai Yap³

¹Chongqing Key Laboratory of Image Recognition, Chongqing University of Posts and Telecommunications, Chongqing, China

²School of Computing and Data Engineering, NingboTech University, Ningbo, China

³Department of Bioengineering, Imperial College London, London, UK

⁴Key Laboratory of Cyberspace Big Data Intelligent Security (Chongqing University of Posts and Telecommunications), Ministry of Education, Chongqing, China

⁵School of Computer Science, Chongqing University, Chongqing, China

⁶University of Oxford, Oxford, UK

⁷Cardiovascular Research Centre, Royal Brompton Hospital, London, UK

⁸National Heart and Lung Institute, Imperial College London, London, UK

Correspondence

Weisheng Li and Mingliang Zhou.

Email: liws@cqupt.edu.cn and mingliangzhou@cqu.edu.cn

Funding information

Natural Science Foundation of Chongqing Municipality, Grant/Award Numbers: 2023NSCQ-LZX0045, CSTB2022NSCQ-MSX0436 and cstc202; China Postdoctoral Science Foundation, Grant/Award Number: 2023M740741; National Natural Science Foundation of China, Grant/Award Numbers: 62331008, 62176027, 62027827, 62221005 and 622760; Natural Science Foundation of Ningbo Municipality, Grant/Award Number: 2023J280; Zhejiang Province Postdoctoral Research Funding Project, Grant/Award Number: ZJ2023008; Ningbo Key R&D Program, Grant/Award Number: 2023Z231

Abstract

Hypoplastic left heart syndrome (HLHS) is a rare, complex, and incredibly foetal congenital heart disease. To decrease neonatal mortality, evolving HLHS (eHLHS) in pregnant women should be critically diagnosed as soon as possible. However, diagnosis is currently heavily dependent on skilled medical professionals using foetal cardiac ultrasound images, making it difficult to rapidly and easily examine for this disease. Herein, the authors propose a cost-effective deep learning framework for rapid diagnosis of eHLHS (RDeH), which we have named RDeH-Net. Briefly, the framework implements a coarse-to-fine two-stage detection approach, with a structure classification network for 4D human foetal cardiac ultrasound images from various spatial and temporal domains, and a fine detection module with weakly-supervised localisation for high-precision nidus localisation and physician assistance. The experiments extensively compare the authors' network with other state-of-the-art methods on a 4D human foetal cardiac ultrasound image dataset and show two main benefits: (1) it achieved superior average accuracy of 99.37% on three categories of foetal ultrasound images from different cases; (2) it demonstrates visually fine detection performance with weakly supervised localisation. This framework could be used to accelerate the diagnosis of eHLHS, and hence significantly lessen reliance on experienced medical physicians.

KEYWORDS

deep learning, medical image processing

Guang Yang and Choon Hwai Yap are Co last authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *CAAI Transactions on Intelligence Technology* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology and Chongqing University of Technology.

1 | INTRODUCTION

Congenital heart disease (CHD) involves a congenital malformation caused by abnormal development of the heart and large blood vessels during the prenatal stage. Infant mortality is most commonly caused by CHD, which occurs in about 1% of live births and is the leading cause of birth defect-related deaths [1, 2]. Currently, CHD detection is performed via mass screening with 2D ultrasound, with potential cases then being referral to specialist cardiologists. However, a recent international survey reported that such screening can only achieve prenatal detection of critical CHD in about 50% of cases [3]. Prenatal detection is important because if detection occurs only after birth, there is less time to plan for and execute treatment and surgery, leading to reduced survival rates and adverse long-term neurological outcomes. CHD disrupts normal blood flow and can cause abnormal brain development, and delayed post-natal treatment contributes to poor neurological outcomes. Prenatal diagnosis is also important because it prompts a search for genetic abnormalities and extra-cardiac abnormalities, which can lead to discovery of the need for other procedures. There is also a need to accurately predict disease progression and birth outcomes, so as to accurately advise and counsel parents.

Foetal aortic stenosis involves an abnormality in the foetal heart. When this condition is coupled with specific abnormalities like monophasic mitral inflow and antegrade aortic arch flow, there is a significant risk of the foetus developing hypoplastic left heart syndrome (HLHS) malformation [4]. Such diseased hearts are thus termed as ‘evolving HLHS’ (eHLHS). These hearts have a more globular left ventricular (LV) shape and reduced LV motion, and are used in the current study as a sample of disease conditions for deep learning detection.

Compared with other state-of-the-art medical imaging technologies (e.g., MRI and CT) [5, 6], ultrasound imaging is widely used in prenatal examinations owing to its relative safety, low cost, operator comfort, and operator experience [7, 8]. Thus, cardiac ultrasound imaging is the main examination for CHD during pregnancy [9, 10]. Currently, ultrasound imaging includes 2D, 3D, and advanced 4D methods. Compared with 2D and 3D ultrasound scans, 4D scans can provide a wider range of spatial-temporal foetal heart information for diagnosis.

In previous decades, evaluation of foetal hearts was performed primarily through echocardiography. However, such evaluation is challenging because of the small size of foetal hearts, limited resolution on ultrasound scanners, and limited understanding of the implications of some measured parameters. With the rapid development of artificial intelligence [11], deep learning presents enormous promise for medical image analysis tasks, such as brain segmentation from MRI, lesion detection on ultrasound, and liver segmentation on CT. Deep learning can learn representations of images and automatically discover similar representations for detection or classification [12, 13].

Deep learning computational approaches can alleviate some of aforementioned difficulties by providing higher accuracy than manual measurements and avoiding human errors, enabling automatic interpretations that hasten traditionally slow manual processes, and providing additional predictive and interpolative capabilities that augment user expertise. However, compared with adult ultrasound images, the foetal ultrasound images studied here pose additional difficulties, including the small size of the heart, obstructions from foetal bones, and a greater number of signals from tissue outside the epicardium that can be difficult to distinguish from the myocardium.

Here, we develop a deep learning framework for rapid detection of eHLHS (named herein as RDeH-Net) that can act as a starting framework for future expansion to the detection of a wider range of diseases and abnormalities. The main contributions of our work are as follows.

- We develop a cost-effective labelling scheme for rapid eHLHS diagnosis, which uses bounding box labels to cover regions of interest in the foetal heart without providing explicit directions. Our approach simplifies the labelling task while avoiding the use of overly complex annotations, which distinguishes it from previous methods.
- We propose a two-stage deep network for accurate foetal cardiac nidus region detection, which incorporates a fine-detection module with weakly supervised localisation. This enhances localisation precision, which aids segmentation and provides valuable assistance to physicians.
- We establish a unique 4D foetal cardiac ultrasound dataset with samples from early eHLHS cases and normal foetuses, each sample with specific slice and timing information. This dataset fills a crucial gap in CHD research because no public dataset of 4D foetal cardiac ultrasound images currently exists.

The remainder of this paper is organised as follows. Related work is discussed in Section 2. In Section 3, the framework of our proposed approach is outlined. Subsequently, the proposed coarse and fine detection methods are elaborated on in detail in Section 3. The experimental results and analysis are described in Section 4. Finally, in Section 5, we summarise the content of this study.

2 | RELATED WORKS

2.1 | Traditional methods for analysing foetal cardiac ultrasound images

To analyse and interpret foetal cardiac ultrasound images, traditional automatic analysis approaches primarily use classic image processing algorithms and models, with traditional automatic analysis techniques relying primarily on thresholding and manually determined characteristics [14, 15].

Nageswari et al. [14] proposed the transverse dyadic wavelet transform algorithm to retain the border and

curvature of four chambers on foetal cardiac ultrasound images. Then, Nageswari et al. [16] proposed a spatially constrained-distance regularised level set evolution segmentation algorithm for foetal cardiac ultrasound images. To detect the borders of the foetal heart, Song et al. [17] first extracted ROI regions using a Gaussian filter and texture analysis, and then applied an energy-based active contour model. Li et al. [18] proposed a Hough-forest-based detector for detection and localisation in cardiac ultrasound diagnosis and therapy recommendations. Sushma et al. [19] proposed a method for automated recognition of CHD from ultrasonic 2D imaging, which made use of statistical features and pattern classifiers such as the support vector machine. Bridge et al. [20] proposed an automated description framework for tracking key variables that described the content of each frame of freehand 2D ultrasound scanning videos of healthy foetal heart. Vargas-Quintero et al. [21] designed a novel multi-texture active appearance model based on the Hermite transform to perform segmentation of the LV. Mahmood et al. [22] used the fundamental principle that the area closest to the apex in a four-chamber view is the LV region to provide a new way of robustly finding the LV. However, it remains challenging for traditional approaches to attain higher accuracy and better performance because of the complex properties of the embryonic heart.

3 | DEEP LEARNING METHODS FOR ANALYSING FOETAL CARDIAC ULTRASOUND IMAGES

With the rapid development of deep learning, a large number of excellent deep learning networks have been proposed [23–28], including AlexNet [29], ResNet [30], Yolo [31–34], and SSD [35]. In medical image analysis, numerous deep learning networks have recently been applied to diagnostic support [36–39], with these including U-Net [40], nnU-Net [41], and Swin UNETR [42].

For human cardiac ultrasound image diagnosis, automated analysis methods can be coarsely classified into adult cardiac ultrasound image analysis methods and foetal cardiac ultrasound image analysis methods. Many studies have been performed on adult cardiac ultrasound image analysis methods. Ghorbani et al. [43] applied deep learning to echocardiography to identify local cardiac structures. By using a deep learning model, their approach was able to estimate cardiac function and predict systemic phenotypes. Bello et al. [44] proposed a deep-learning cardiac motion analysis method for human survival prediction, which involved training a fully convolutional network on anatomical shape priors to predict cardiac motion. Dezaki et al. [45] proposed a deep residual recurrent neural network to automatically recognise cardiac cycle phases.

Foetal cardiac ultrasound image analysis methods combine the advantages of adult heart analysis methods while taking into account foetal heart disease characteristics and

foetal posture when designing classification, detection, and segmentation algorithms. Arnaout et al. [46] proposed neural network-based prenatal detection of complicated CHD. In this detection, a view classifier categorised the ultrasound view, and then three CHD diseases could be detected independently by successive detection classifiers. Komatsu et al. [2] proposed a convolutional neural network to detect cardiac substructures and structural abnormalities on foetal ultrasound videos. This detection method only adopted a supervised learning scheme from a dataset of normal cases. Pu et al. [47] proposed an AI-based method for segmenting anatomical structures in the foetal apical four-chamber view. They divided a MobileNet backbone network into four stages and designed an FPN network to enhance multi-scale semantic information.

4 | PROPOSED APPROACH

Our proposed RDeH-Net approach can be visualised in Figure 1. In our approach, we devised a two-stage deep learning framework. To begin with, we designed a binary deep learning neural network to achieve coarse structure classification on human foetal ultrasound images. Then, we designed a fine-detection module with weakly-supervised localisation. This module includes a fine detection network and weakly-supervised localisation.

4.1 | Human foetal cardiac substructure classification

According to the theory of human cardiac anatomy, the human cardiac substructure is made up primarily of aorta (AO), pulmonary artery, left atrium, right atrium, LV, and right ventricle. A normal human foetal heart is shown in Figure 2. However, an abnormal human foetal heart with HLHS has an LV pattern variation differing from the substructure of a normal human foetal heart. Compared with a normal human foetal heart, the LV of a human foetal heart with eHLHS appears larger and brighter. Clinicians can use this knowledge to distinguish an eHLHS foetal heart.

Using the knowledge of these substructural differences between normal and eHLHS human foetal heart, we can find the features on their ultrasound images, as shown in Figure 2. From 4D ultrasound images of the human foetal heart, we can further detect two different types of image substructures of the eHLHS human foetal heart. One is a classic salient four-chamber human cardiac substructure that is referred to as HLHS-1 in this paper, as shown in Figure 2b. The substructure of this HLHS-1 human heart is similar to that of a normal human heart with four-chambers, as shown in Figure 2a. For clinicians with little training in clinical diagnosis, this presents a significant barrier. The other type is a non-salient four-chamber human cardiac substructure, called HLHS-2 in this paper, which is shown in Figure 2c.

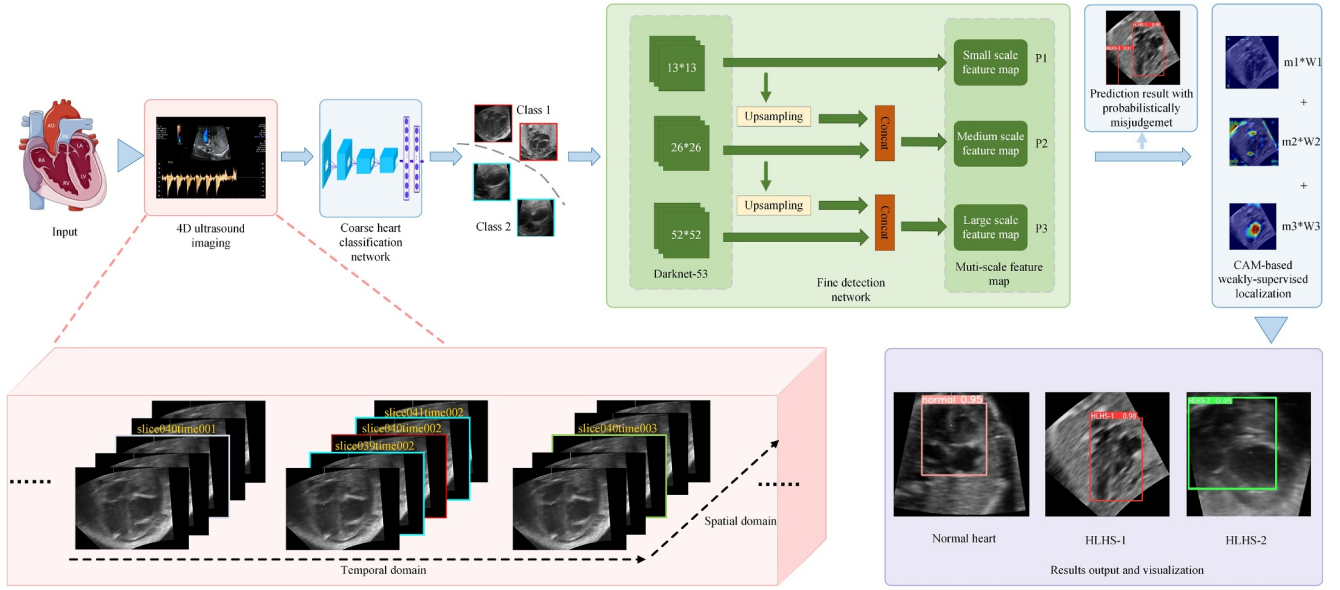


FIGURE 1 Framework of the proposed approach.

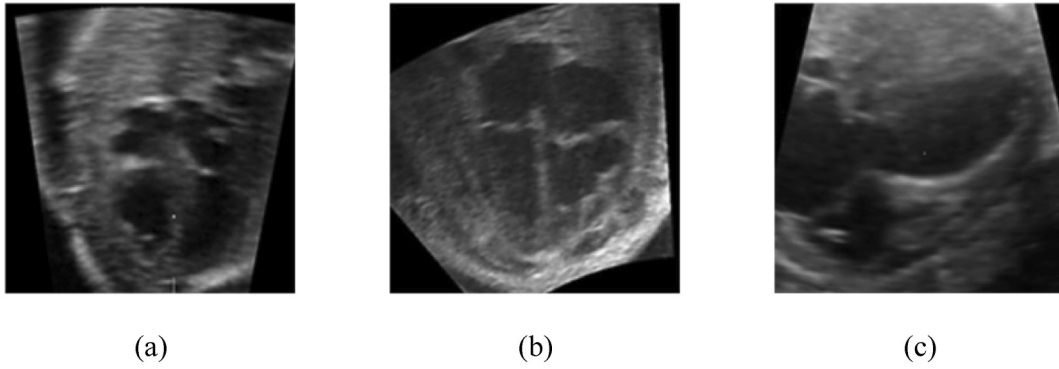


FIGURE 2 Structural representation characteristics of three foetal cardiac ultrasound image cases: (a) normal foetal heart, (b) HLHS-1, and (c) HLHS-2.

Applying this substructural knowledge of the eHLHS human foetal heart, we designed a binary deep learning neural network to achieve coarse structure classification on human foetal ultrasound images. The binary network can classify four-chamber human cardiac substructure images into two groups: normal human cardiac images and HLHS-1 human cardiac images forming one group, and non-salient four-chamber human cardiac substructure images from the HLHS-2 type forming the other group.

To design the binary network, we first defined 4D ultrasound images as follows:

$$V = \{v_{t,s}\}_{t=1,s=1}^{T,S} \in \mathfrak{R}^{T \times H \times W \times S} \quad (1)$$

where $v_{t,s}$ represents the s -th $H \times W$ ultrasound slice image at the t -th time, H represents the height of the ultrasound image, and W represents the width of the ultrasound image.

Then, we defined our binary network as follows:

$$Z_{v_{t,s}} = \Phi_{Binary}(v_{t,s}, t, s) = \begin{cases} C_1 \\ C_2 \end{cases} \quad (2)$$

where Φ_{Binary} is the binary network classification function, C_1 represents a category of normal human foetal cardiac and HLHS-1 human foetal cardiac ultrasound images, and C_2 represents a category of HLHS-2 human foetal cardiac ultrasound images.

We employed ResNet-50 as the fundamental binary network in our binary approach. Because of our incorporation of structural prior knowledge for the two-group US images, even conventional simple networks can demonstrate outstanding classification performance in this task.

4.2 | Fine detection with weakly-supervised localisation

After finishing the first stage of coarse classification based on the cardiac substructure saliency in 4D ultrasound images, we

designed a fine-detection module with weakly-supervised localisation. As seen in Figure 1, this module includes a fine detection network and weakly-supervised localisation.

The fine detection network uses YOLOv5 as the backbone network, and can achieve output of detection results with prediction labels and bounding boxes. Then, we employ a basic focal loss function as our loss function. To obtain the high-precision localisation of cardiac nidus regions, our approach introduces CAM-based weakly-supervised localisation. By using the weakly-supervised localisation, we can optimise prediction bounding boxes.

The fine detection network can output three scale feature maps, which are defined as follows:

$$P_{v_{t,s}}^k = \Gamma(v_{t,s}, Z_{v_{t,s}}) \quad (3)$$

where $P_{v_{t,s}}^k$ represents the k -th feature map of $v_{t,s}$ under the classification of $Z_{v_{t,s}}$.

Each feature map generates prediction bounding boxes according to a non-maximum suppression (NMS) threshold, and the optimal prediction bounding boxes of the network can be given by the following:

$$B_{pre} = \Theta_{k \in N}(P_{v_{t,s}}^k, B^k) \quad (4)$$

where $P_{v_{t,s}}^k$ represents the k -th feature map of $v_{t,s}$, N is 3, and B^k represents the prediction bounding boxes on the k -th feature map.

As NMS is a simple bounding box selection strategy using a threshold, it can result in misjudged bounding boxes. Thus, we use a weakly-supervised localisation algorithm to enhance the selection strategy. Our approach can calculate the importance weight on three feature maps with different scales. According to the importance weight of each feature map, we obtain the final output prediction bounding box.

To calculate the importance weight, we first give the gradient of the class score

$$Y = \sum_k \delta_k \lambda \sum_i^H \sum_j^W P_{v_{t,s}}^k(i, j) \quad (5)$$

where δ_k is the feature weight in the k -th feature map, and λ is a proportionality constant that gets normalised out during visualisation.

The importance weight on the k -th feature map can be given by the following:

$$W_k = \lambda \sum_W \sum_H \frac{\partial Y}{\partial P_{v_{t,s}}^k} \quad (6)$$

The total importance weight of a foetal cardiac ultrasound image can be given by

$$W_{total} = \sum_k m_k W_k \text{ s.t. } \sum_k m_k = 1 \quad (7)$$

where m_k represents the weight of the importance weight on the k -th feature map. In our approach, the small-scale feature map and middle-scale feature map cannot efficiently obtain the important weight values, as can be seen in Figure 1. However, the large-scale feature map can obtain a better important value, so we set $m_3 = 1$, $m_1 = 0$, $m_2 = 0$.

Finally, we can obtain the ultimate prediction bounding box as follows:

$$B_{out} = \begin{cases} W_{total} \cap B_{pre} & q \geq 2 \\ B_{pre} & \text{others} \end{cases} \quad (8)$$

where q represents the number of output prediction bounding boxes from the fine detection network according to the NMS threshold.

Our proposed approach combines the detection network with weakly-supervised localisation to achieve high-precision localisation of foetal cardiac nidus regions. In the first stage, the binary deep learning neural network is able to achieve coarse structure classification of human foetal ultrasound images. In the second stage, the fine-detection module can obtain the ultimate prediction bounding box by computing the importance weight from the feature map of the deep learning network. A detailed description of our proposed approach is shown in Algorithm 1.

Algorithm 1 Coarse-to-fine detection with weakly-supervised localisation for 4D foetal cardiac ultrasound images

Input: The s -th ultrasound slice image at t -th time $V = \{v_{t,s}\}_{t=1, s=1}^{T, S} \in \mathfrak{R}^{T \times H \times W \times S}$

Output: The prediction box B_{out}

1: Initialize two group classifications $c \in \{C1, C2\}$, feature map $N=3$, the number of the output prediction bounding boxes $q=0$;

2: **for** ($i=1$ to s) **do**

3: $Z_{v_{t,s}} \leftarrow \Phi_{Binary}(v_{t,s}, t, s)$;

4: $P_{v_{t,s}}^k \leftarrow \Gamma(v_{t,s}, Z_{v_{t,s}})$;

5: **for** ($k=1$ to N) **do**

6: $B_{pre} \leftarrow \Theta_{k \in N}(P_{v_{t,s}}^k, B^k)$;

7: $Y \leftarrow \sum_k \delta_k \lambda \sum_i^H \sum_j^W P_{v_{t,s}}^k(i, j)$;

8: $W_k \leftarrow \lambda \sum_W \sum_H \frac{\partial Y}{\partial P_{v_{t,s}}^k}$;

9: **end for**

10: $W_{total} \leftarrow \sum_k m_k W_k$ and *s.t.* $\sum_k m_k = 1$;

11: $q \leftarrow$ number of the output prediction bounding boxes;

12: **if** $q \geq 2$ **then**

13: $B_{out} = W_{total} \cap B_{pre}$;

14: **else**

15: $B_{out} = B_{pre}$;

16: **end if**

17: **end for**

5 | EXPERIMENTS

5.1 | Datasets and training set

Establishing datasets was essential for the training and verification of our proposed RDeH-Net approach. Since no publicly available 4D human foetal cardiac datasets for eHLHS can be found, we used our previously collected 4D human cardiac ultrasound images (with diagnostic results for other tasks) as our original source images [15, 48, 49]. All images were acquired at the National University Hospital in Singapore using procedure number 2014/00056, which was authorised by the Domain Specific Review Board and patients. A Voluson 730 ultrasound system with an RAB 4–8L transducer (GE Healthcare, Chicago, IL, USA) was used to acquire raw 4D ultrasound images in the Spatio-Temporal Image Correlation (STIC) mode. With the four-chamber view displayed, STIC imaging was captured. The STIC sweep took place across 10–15 s with an image capture rate of 70–90 frames per second, giving 29–37 volumes for one cardiac cycle. The volumetric images were exported from each time point as a stack of 30–40 slice images parallel to the four-chamber view, and separated by 0.5 mm. All foetal cardiac images were acquired at more than 21 weeks gestation.

To establish our training dataset and train our detection network model, we labelled 4D human foetal cardiac ultrasound images. We started by manually labelling 2676 original source images, including normal and patient cases. Then, we used a data augmentation tool to augment our data, a process that can improve the robustness of a deep learning network model. This augmentation included rotation, scaling, flipping, shifting, and adding noise. After finishing data augmentation, our 4D human foetal cardiac ultrasound image dataset contained 16056 images. In the training stage, our proposed approach was to train using our dataset. We randomly partitioned the dataset into a training set, validation set, and test set in the ratio of 8:1:1.

5.2 | Experimental setting

We adopted common indicators to evaluate the proposed RDeH-Net approach. Then, we set an experimental training and test environment.

To evaluate the performance of the method on all classes, we adopted the mAP indicator as the measurement with a threshold score = 0.5 and score = 0.5:0.95.

The recall indicator was selected to evaluate the miss rate:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

where TP represents the number of sample images truly belonging to a class assigned to that same class by the model, and FN represents the number of sample images of a

particular class not appearing in that class in the output of the model.

All experiments were conducted on the same computer with a single NVIDIA GeForce GTX 2080TI GPU. Benefiting from the GPU, it took around 40 h to train our deep learning network with 150 epochs.

5.3 | Performance evaluation

In this part, we compare the performance of our RDeH-Net method on our validated dataset with other state-of-the-art methods in terms of mAP@0.5, mAP@0.5:0.95, precision, and recall. Moreover, we thoroughly compare the detection results of all methods applied to various scenarios by visualising them.

We first evaluated our RDeH-Net approach and the other state-of-the-art methods on our established datasets. Table 1 lists the results of mAP@0.5, mAP@0.5:0.95, and precision on the dataset. As shown in Table 1, our RDeH-Net approach achieved mAP@0.5 of 99.37%, and mAP@0.5:0.95 of 74.57% on the dataset, values that were significantly better than the other state-of-the-art methods. In particular, while the mAP@0.5 value of our approach was 99.37%, the mAP@0.5 value of the second-best method, YOLOv3, was 98.78%. The better performance with our RDeH-Net was due to the fact that it adopts a coarse-to-fine detection approach, whereas the other state-of-the-art methods directly use an end-to-end approach. Additionally, we made full use of prior knowledge to guide the learning of the network. Figure 3b shows intuitive performance comparisons for mAP@0.5 and mAP@0.5:0.95, and another performance indicator comparison for different methods. From Figure 3a,b, we can see that the results of our proposed method are superior to those of the other methods. Additionally, our method achieved high precision of 94.99%.

TABLE 1 Test results of our method and other state-of-the-art methods.

Methods	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Precision (%)
YOLOv5l [50]	94.60	74.16	92.14
YOLOv5m [50]	97.38	78.00	96.08
YOLOv5s [50]	97.88	77.54	91.66
YOLOv5x [50]	94.42	73.82	89.60
YOLOv3 [34]	98.78	74.42	96.53
YOLOv3-tiny [51]	95.26	64.54	93.93
SSD [35]	97.47	73.36	94.21
FLDS [52]	64.81	-	-
YOLOv8 [53]	98.34	75.86	95.55
RDeH-net (ours)	99.37	74.57	94.99

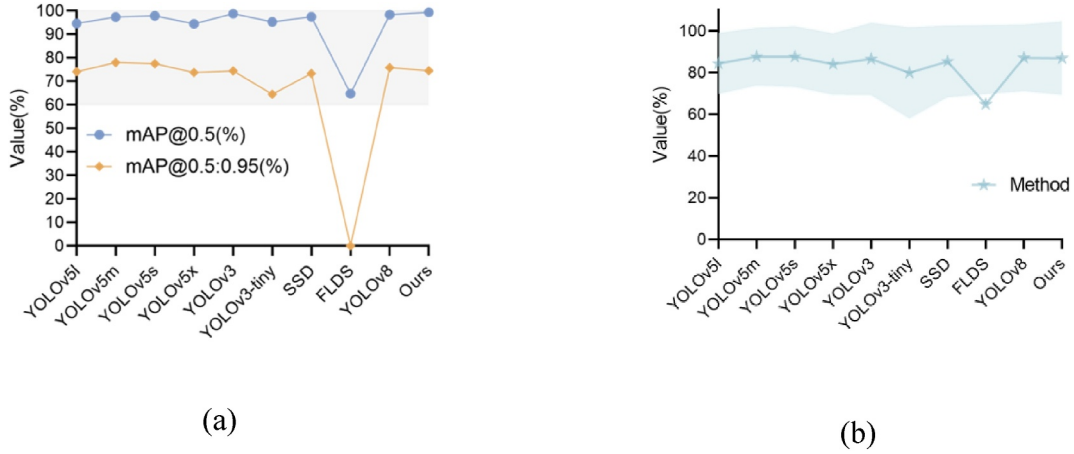


FIGURE 3 Performance comparisons according to the mAP indicator: (a) mAP@0.5 and (b) mAP@0.5:0.95.

For medical images, a low miss rate is a crucial target, meaning the probability of escaping diagnosis will be small. To evaluate the miss rate, we adopted the recall indicator. To obtain recall for each class, we initially randomly chose 269 samples with detection results from the test set. Then, we counted the TP and FN values of the sample images. Table 2 shows the TP and FN values of each class, comparing the TP and FN of nine methods on each class. From Figure 4, we can see that our RDeH-Net approach achieved better TP and FN performance than the other methods.

After completing the TP and FN statistics, we used Eq. (9) to calculate the recall indicator. Table 3 lists the results of the recall indicator on each class in the test samples. Compared with the other methods, our RDeH-Net approach maintained a higher recall rate. Notably, our RDeH-Net approach obtained the highest $mAP@0.5$ under this higher recall rate. An intuitive comparison of the recall indicators of the nine methods is shown in Figure 5. Figure 6 shows our model's performance in the prediction of three foetal cardiac ultrasound image classes in horizontal and vertical directions. Figure 7 shows a comparison between the prediction locations of our RDeH-Net approach and the ground truth locations.

We also tested the detection inference time cost of our RDeH-Net approach. We implement the deployment on a laptop with an NVIDIA GeForce RTX3060 GPU. Our RDeH-Net approach took an average of 39 ms for each ultrasound image. This implies that our RDeH-Net approach for daily diagnosis is convenient.

5.4 | Visual comparisons

Figure 8 shows visual comparisons for the normal case detection performance of our proposed RDeH-Net approach, YOLOv5l, YOLOv5m, YOLOv5s, YOLOv5x, YOLOv3, YOLOv3-tiny, SSD, FLDS, and YOLOv8. As can be seen in Figure 8, both our RDeH-Net approach and YOLOv3

TABLE 2 True positives and false negatives of the nine methods on each class.

Methods	Classes	TP	FN
YOLOv5l [50]	Normal	30	0
	HLHS-1	200	0
	HLHS-2	35	4
YOLOv5m [50]	Normal	30	0
	HLHS-1	200	0
	HLHS-2	29	10
YOLOv5s [50]	Normal	30	0
	HLHS-1	200	0
	HLHS-2	29	10
YOLOv5x [50]	Normal	30	0
	HLHS-1	200	0
	HLHS-2	33	6
YOLOv3 [34]	Normal	30	0
	HLHS-1	196	4
	HLHS-2	38	1
YOLOv3-tiny [51]	Normal	30	0
	HLHS-1	200	0
	HLHS-2	21	18
SSD [35]	Normal	30	0
	HLHS-1	200	0
	HLHS-2	39	0
FLDS [52]	Normal	23	7
	HLHS-1	175	15
	HLHS-2	27	2
YOLOv8 [53]	Normal	30	0
	HLHS-1	178	0
	HLHS-2	38	0
RDeH-net (ours)	Normal	30	0
	HLHS-1	200	0
	HLHS-2	39	0

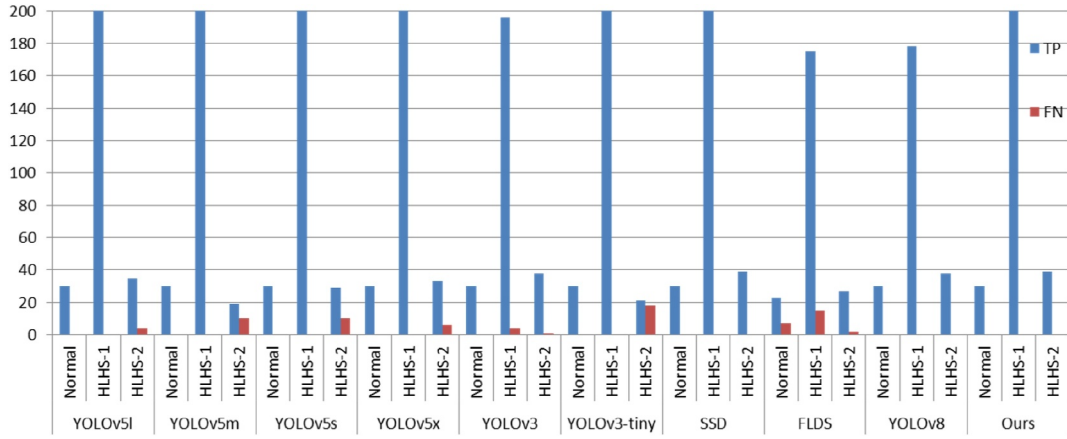


FIGURE 4 Comparison of true positives and false negatives of nine methods on each class.

TABLE 3 Comprehensive analysis of the nine methods in terms of the recall rate.

Methods	Classes	TP	TP + FN	Recall (%)
YOLOv5l [50]	Normal	30	30	100.00
	HLHS-1	200	200	100.00
	HLHS-2	35	39	89.74
YOLOv5m [50]	Normal	30	30	100.00
	HLHS-1	200	200	100.00
	HLHS-2	19	29	65.52
YOLOv5s [50]	Normal	30	30	100.00
	HLHS-1	200	200	100.00
	HLHS-2	29	39	74.36
YOLOv5x [50]	Normal	30	30	100.00
	HLHS-1	200	200	100.00
	HLHS-2	33	39	84.62
YOLOv3 [34]	Normal	30	30	100.00
	HLHS-1	196	200	98.00
	HLHS-2	38	39	97.44
YOLOv3-tiny [51]	Normal	30	30	100.00
	HLHS-1	200	200	100.00
	HLHS-2	21	39	53.85
SSD [35]	Normal	30	30	100.00
	HLHS-1	200	200	100.00
	HLHS-2	39	39	100.00
FLDS [52]	Normal	23	30	76.67
	HLHS-1	175	190	92.11
	HLHS-2	27	29	93.10
YOLOv8 [53]	Normal	30	30	100.00
	HLHS-1	178	178	100.00
	HLHS-2	38	38	100.00
RDeH-net (ours)	Normal	30	30	100.00
	HLHS-1	200	200	100.00
	HLHS-2	39	39	100.00

achieved the highest performance, with accurate localisation and high scores.

For eHLHS foetal heart ultrasound images, Figures 9 and 10 show visual comparisons of detection performance between HLHS-1 and HLHS-2 cases for our proposed RDeH-Net approach, YOLOv5l, YOLOv5m, YOLOv5s, YOLOv5x, YOLOv3, YOLOv3-tiny, SSD, FLDS, and YOLOv8. As shown in Figures 9 and 10, our approach achieved obviously good detection performance in complex foetal heart images. However, some of the state-of-the-art methods showed rather low scores and localisation accuracy for the nidus heart. Moreover, YOLOv3 frequently generated the error concerning prediction box. This result is because our approach includes a weakly-supervised localisation module for our detection network to adaptively select the optimal prediction bounding box, while the other methods only depend on the constant NMS threshold. Additionally, our weakly-supervised localisation module would be useful for clinicians, aiding them to reduce the percentage of missed diagnoses. As shown in Figure 10j, YOLOv3-tiny missed the detection of the nidus region. Therefore, our weakly-supervised localisation can achieve visualisation-aided localisation, enabling less experienced or weary doctors to comfortably and rapidly visually capture the nidus region.

5.5 | Generalisation performance evaluation on the CAMUS dataset

We also validated the performance of our proposed method on the CAMUS dataset [54] on both end diastole (ED) and end systole (ES) frames (Table 4). Because of the lack of public foetal heart echo image datasets, we performed testing on the CAMUS adult heart ultrasound dataset for cross-validation.

The performance of our proposed model on the CAMUS dataset was comparable to its performance on the foetal dataset, as shown in Table 5 and Figure 11, which suggests that it is also capable of detection on adult ES and ED ultrasound images.

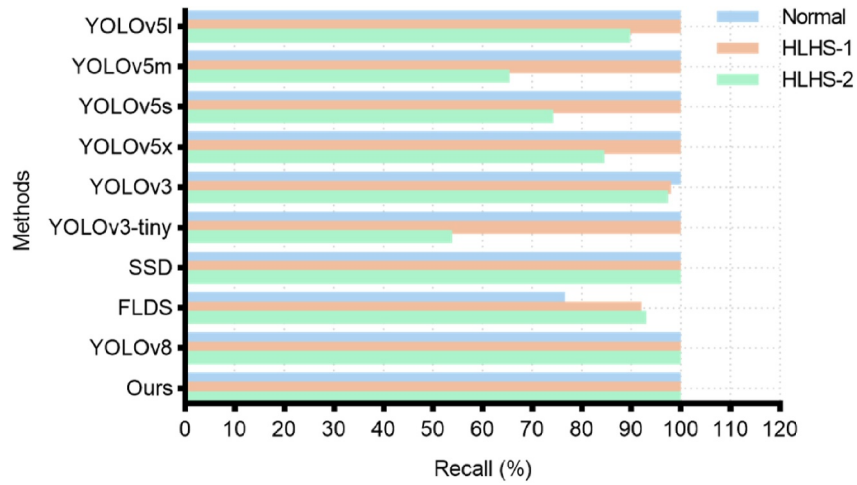


FIGURE 5 Comparison of different methods in terms of the recall rate on each class.

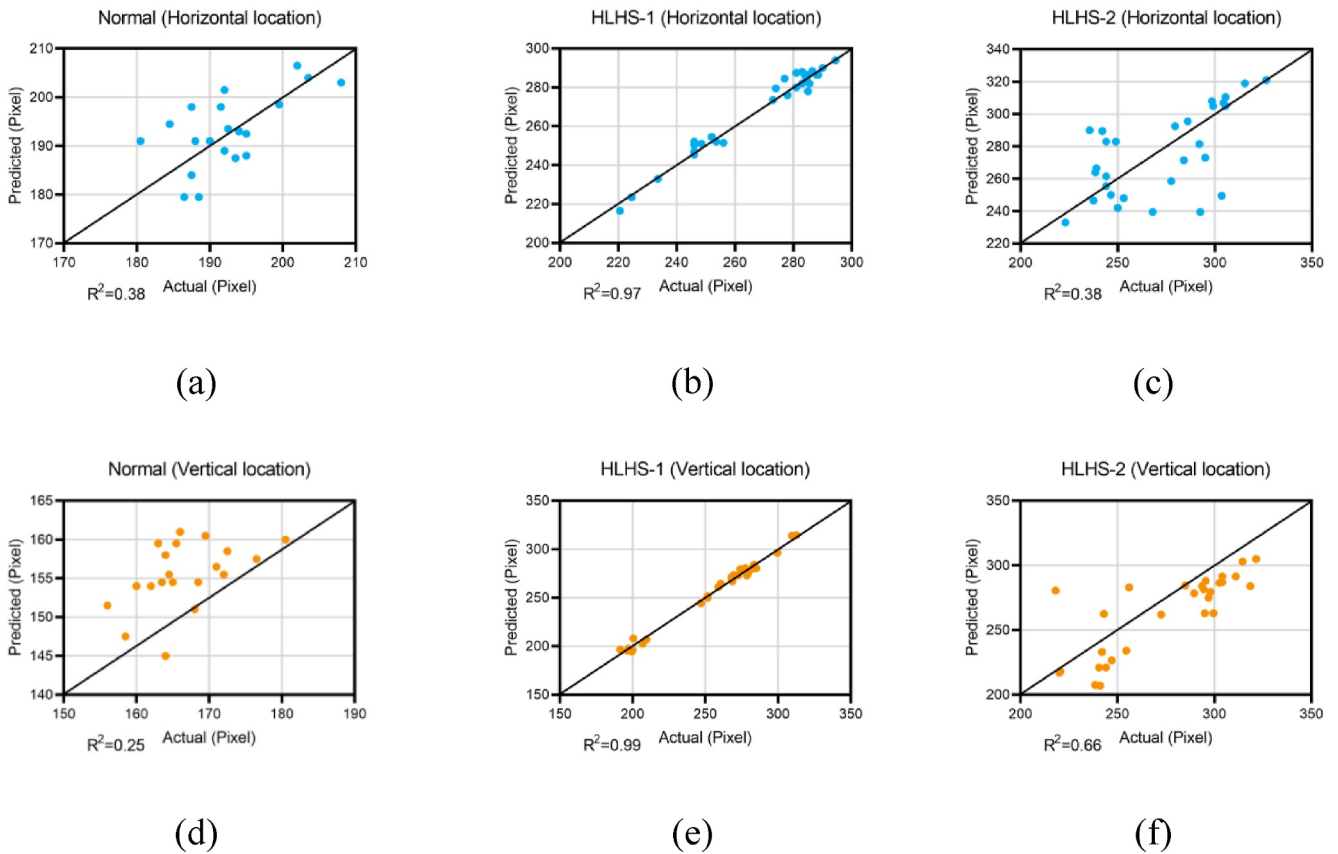


FIGURE 6 Our performance for prediction of three foetal cardiac ultrasound image classes in horizontal and vertical directions: (a) horizontal location of normal cases, (b) horizontal location of HLHS-1 cases, (c) horizontal location of HLHS-2 cases, (d) vertical location of normal cases, (e) vertical location of HLHS-1 cases, and (f) vertical location of HLHS-2 cases.

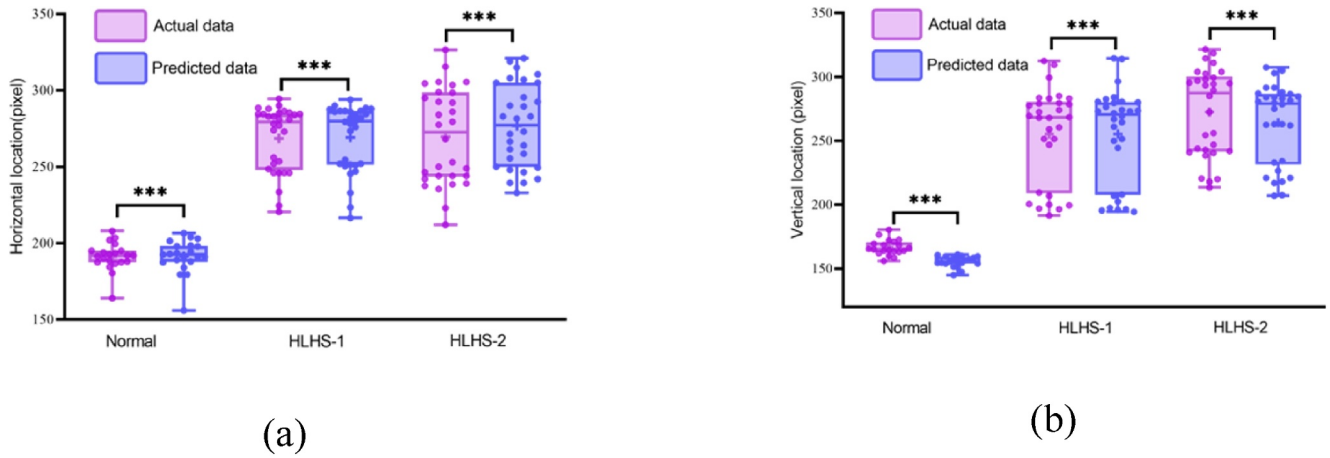


FIGURE 7 Comparisons between the prediction locations of our RDeH-Net approach and the ground truth locations: (a) horizontal direction, and (b) vertical direction.

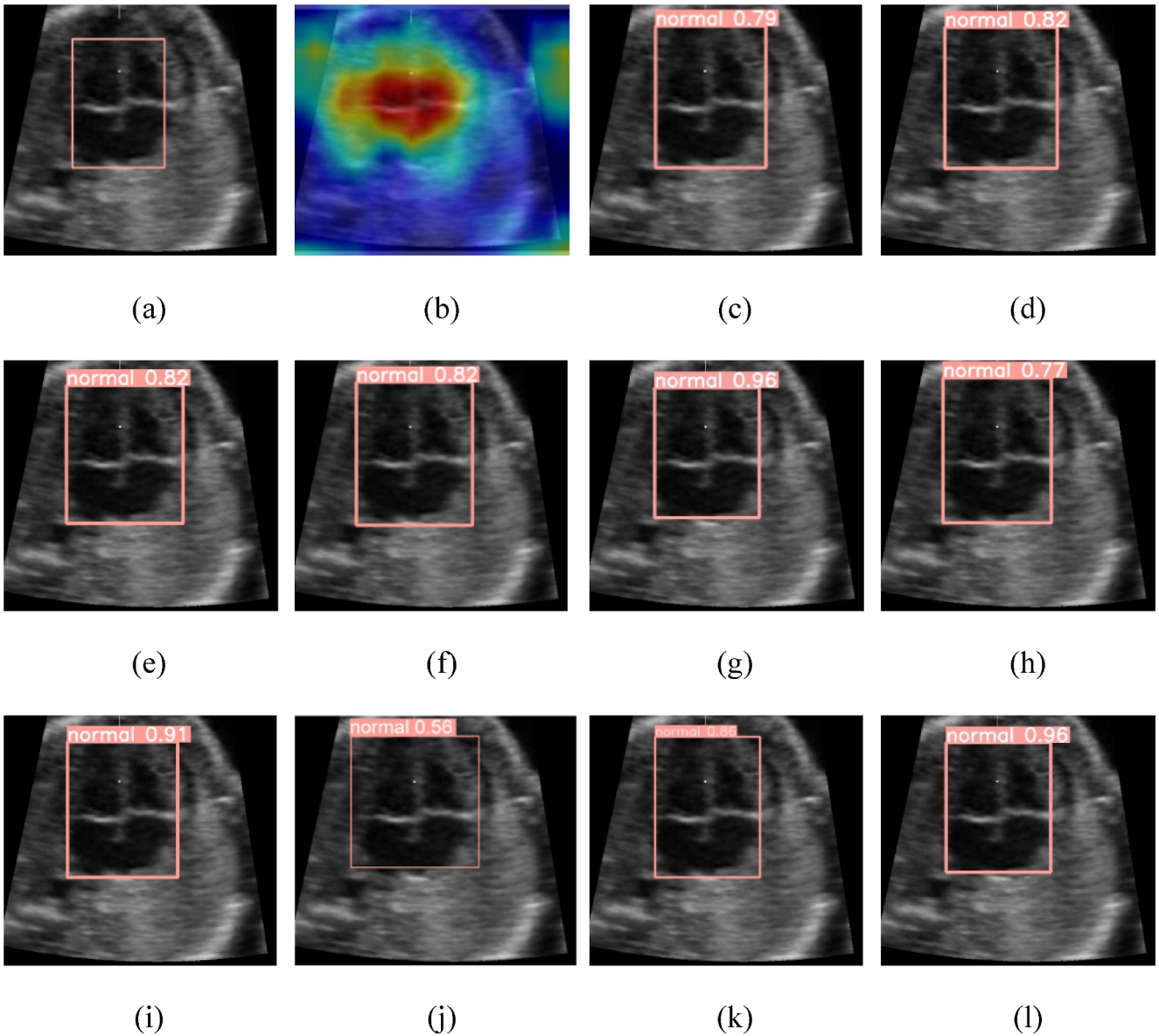


FIGURE 8 Visual comparisons of the detection performance of our proposed approach (RDeH-Net), and YOLOv5l, YOLOv5m, YOLOv5s, YOLOv5x, YOLOv3, YOLOv3-tiny, SSD, FLDS, and YOLOv8 for a normal case. Visual performance includes (a) ground truth, (b) weakly-supervised localisation, (c) YOLOv5l, (d) YOLOv5m, (e) YOLOv5s, (f) YOLOv5x, (g) YOLOv3, (h) YOLOv3-tiny, (i) SSD, (j) FLDS, (k) YOLOv8, and (l) RDeH-Net (ours).

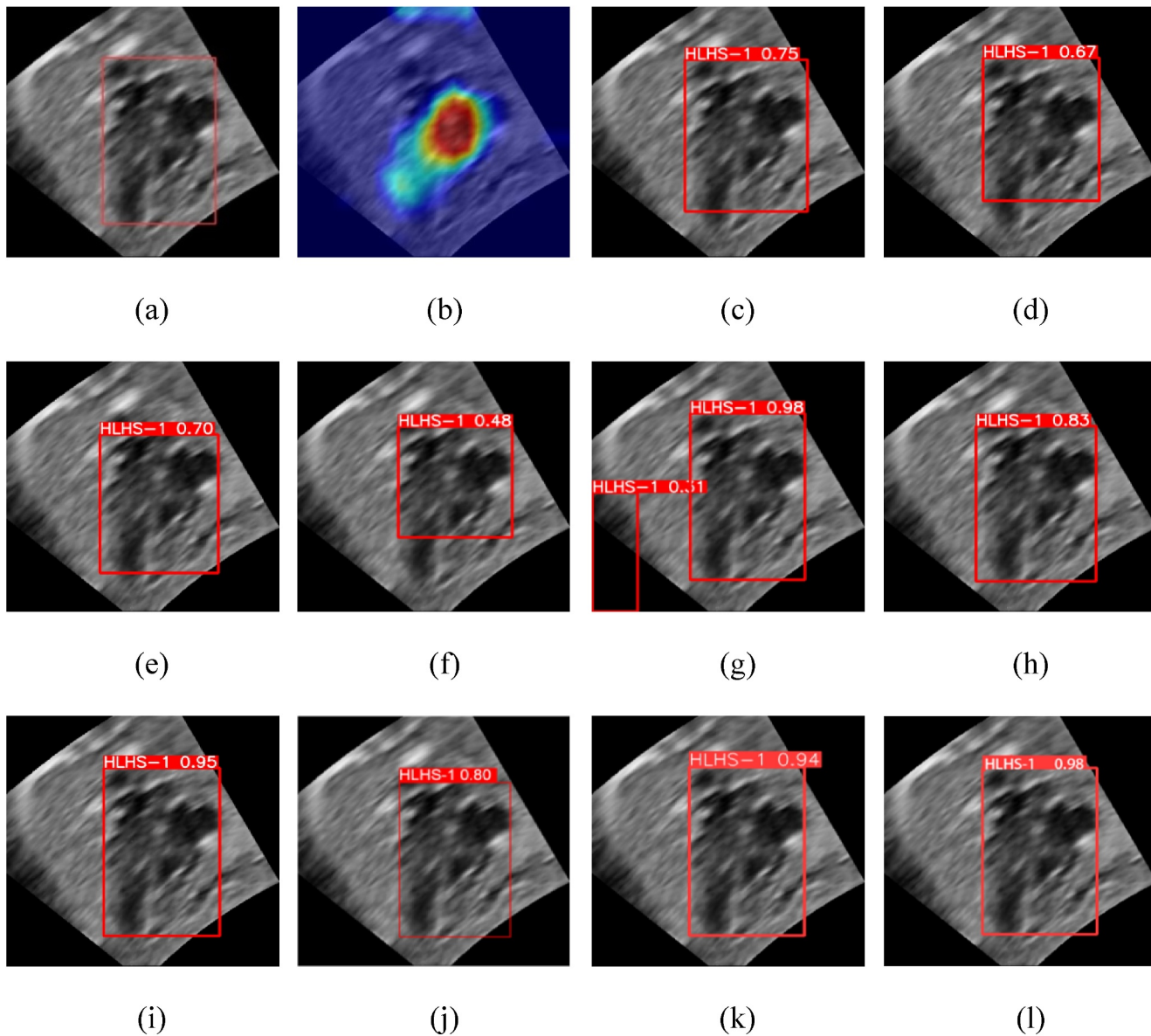


FIGURE 9 Visual comparisons of the detection performance of our proposed approach (RDeH-Net) and YOLOv5l, YOLOv5m, YOLOv5s, YOLOv5x, YOLOv3, YOLOv3-tiny, SSD, FLDS, and YOLOv8 for an HLHS-1 case. Visual performance includes (a) ground truth, (b) weakly-supervised localisation, (c) YOLOv5l, (d) YOLOv5m, (e) YOLOv5s, (f) YOLOv5x, (g) YOLOv3, (h) YOLOv3-tiny, (i) SSD, (j) FLDS, (k) YOLOv8, and (l) RDeH-Net (ours).

6 | CONCLUSION

Our proposed RDeH-Net detection framework can classify a large number of 4D ultrasound images and detect foetal cardiac nidus regions and their localisation, thereby facilitating rapid detection of eHLHS. In our proposed detection approach, a cost-effective labelling scheme was designed to support general labelling tasks without labour expenditure. Additionally, the detection framework that we developed introduces prior knowledge and weak-localisation to increase

its accuracy. Furthermore, we constructed a dataset of 4D foetal cardiac ultrasound images for deep learning tasks, which should be extremely useful for CHD studies. Extensive experimental results demonstrate that our RDeH-Net approach achieved superior average accuracy on three categories of foetal ultrasound images compared with eight other state-of-the-art methods. For future work, we plan to detect the full range of the abnormalities afflicting foetal hearts by collecting more ultrasound image cases from all over the world.

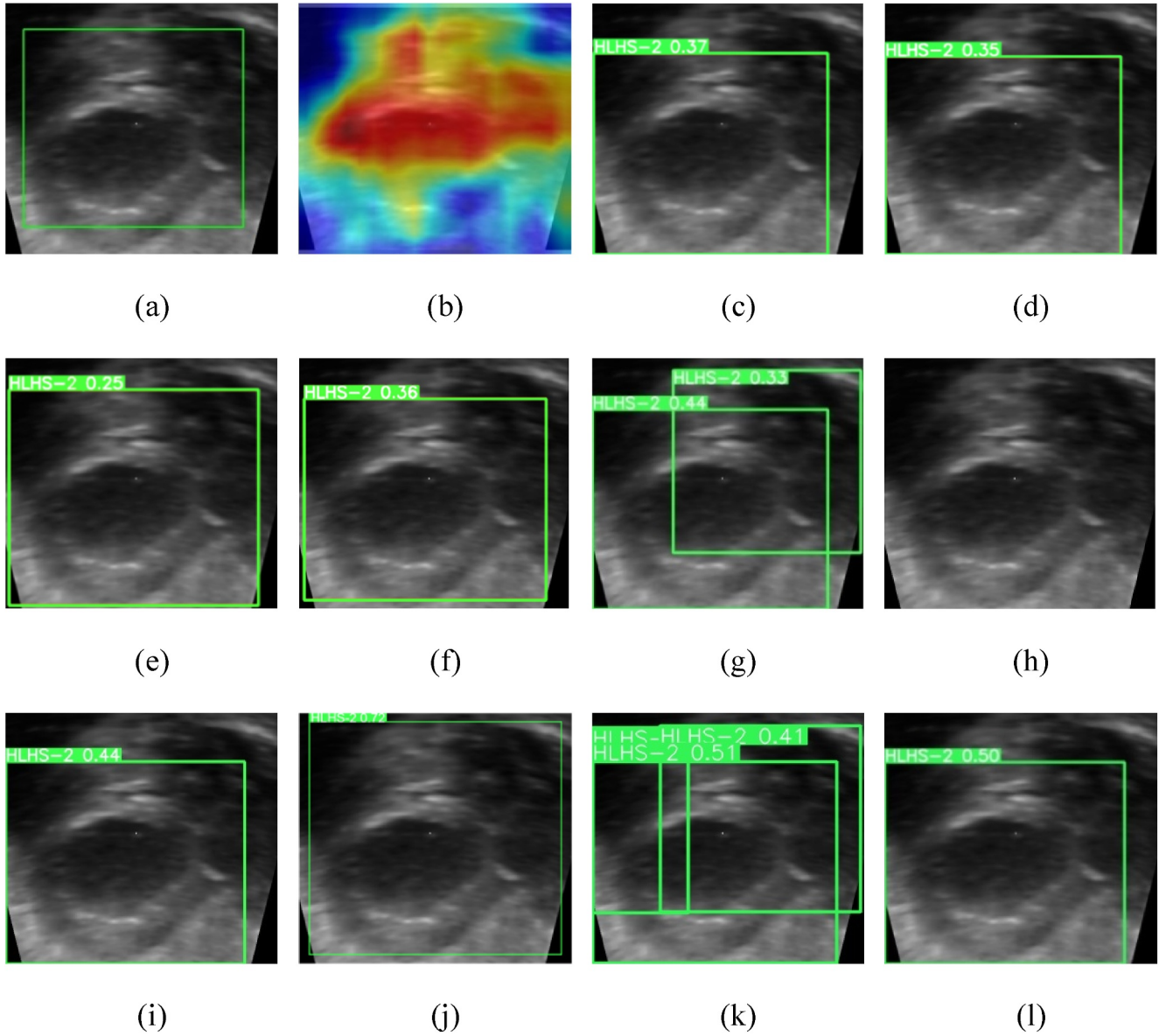


FIGURE 10 Visual comparisons of the detection performance of our proposed approach (RDeH-Net) and YOLOv5l, YOLOv5m, YOLOv5s, YOLOv5x, YOLOv3, YOLOv3-tiny, SSD, FLDS, and YOLOv8 for an HLHS-2 case. Visual performance includes (a) ground truth, (b) weakly-supervised localization, (c) YOLOv5l, (d) YOLOv5m, (e) YOLOv5s, (f) YOLOv5x, (g) YOLOv3, (h) YOLOv3-tiny, (i) SSD, (j) FLDS, (k) YOLOv8, and (l) RDeH-Net (ours).

TABLE 4 Description of the CAMUS dataset.

Classes	Cases
ED	500
ES	500

TABLE 5 Performance of the proposed method on the CAMUS dataset.

Method	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Precision (%)
Ours	96.14	51.45	90.57

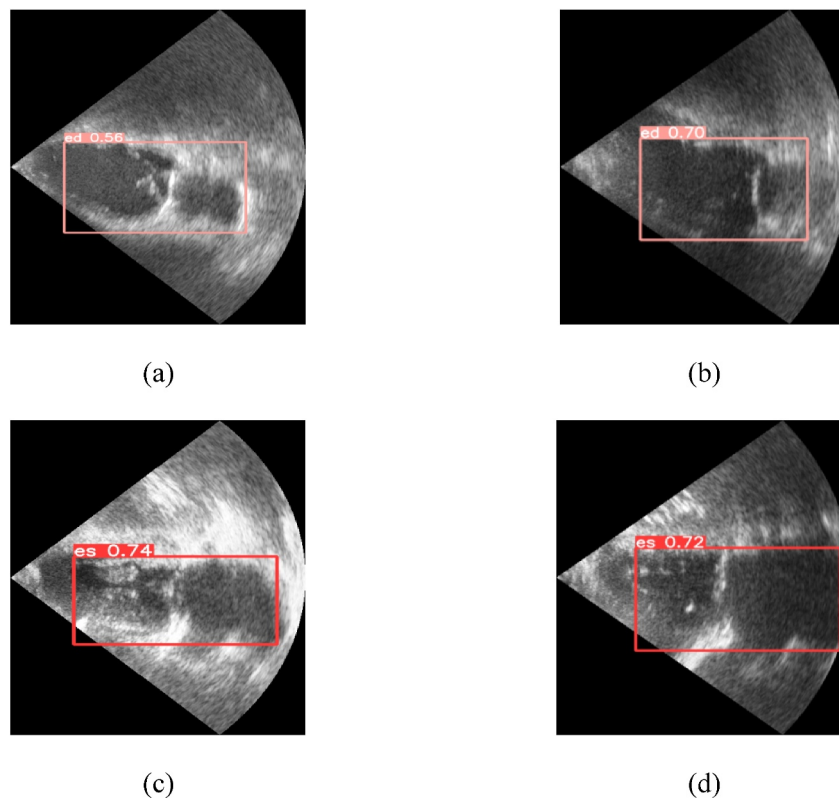


FIGURE 11 Detection performance on ED and ES images from the CAMUS dataset: (a) the ED image of Patient 31, (b) the ED image of Patient 45, (c) the ES image of Patient 04, and (d) the ES image of Patient 82. ED, end diastole; ES, end systole.

ACKNOWLEDGEMENTS

National Natural Science Foundation of China [Nos. 62331008, 62176027, 62027827, 62221005 and 62276040], Natural Science Foundation of Chongqing (Nos. 2023NSCQ-LZX0045, CSTB2022NSCQ-MSX0436 and cstc2020jcyjmsxmX0790), Ningbo Natural Science Foundation under Grant 2023J280, Ningbo Key R&D Program (No.2023Z231), Zhejiang Province Postdoctoral Research Funding Project (No. ZJ2023008), China Postdoctoral Science Foundation (No.2023M740741), Human Resources and Social Security Bureau Project of Chongqing (No. cx2020073); Guangdong Oppo Mobile Telecommunications Corporation Ltd (No. H20221694); UK Research and Innovation Future Leaders Fellowship (No. MR/V023799/1).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All images were taken at the National University Hospital in Singapore using procedure number 2014/00056, which was authorised by the Domain Specific Review Board.

ORCID

Gang Wang  <https://orcid.org/0000-0001-5283-7126>

Mingliang Zhou  <https://orcid.org/0000-0001-8302-4435>

REFERENCES

- Gong, Y., et al.: Fetal congenital heart disease echocardiogram screening based on dgacnn: adversarial one-class classification combined with video transfer learning. *IEEE Trans. Med. Imag.* 39(4), 1206–1222 (2020). <https://doi.org/10.1109/tmi.2019.2946059>
- Komatsu, M., et al.: Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. *Appl. Sci.* 11(1), 371 (2021). <https://doi.org/10.3390/app11010371>
- Xu, X., et al.: Uncompensated mitochondrial oxidative stress underlies heart failure in an ipsc-derived model of congenital heart disease. *Cell Stem Cell* 29(5), 840–855.e7 (2022). <https://doi.org/10.1016/j.stem.2022.03.003>
- Eghtesady, P.: Hypoplastic left heart syndrome: rheumatic heart disease of the fetus? *Med. Hypotheses* 66(3), 554–565 (2006). <https://doi.org/10.1016/j.mehy.2005.09.013>
- Rehman, M.U., et al.: An Efficient Deep Learning Model for Brain Tumour Detection with Privacy Preservation. *CAAI Trans. Intell. Technol.*, 1–16, (2023). <https://doi.org/10.1049/cit2.12254>
- Jia, Y., et al.: An attention-based cascade R-CNN model for sternum fracture detection in X-ray images. *CAAI Transactions on Intelligence Technology* 7(4), 658–670 (2022). <https://doi.org/10.1049/cit2.12072>
- Liu, S., et al.: Deep learning in medical ultrasound analysis: a review. *Engineering* 5(2), 261–275 (2019). <https://doi.org/10.1016/j.eng.2018.11.020>
- Stortz, G., et al.: Quantification of wave reflection in the human umbilical artery from asynchronous Doppler ultrasound measurements. *IEEE Trans. Med. Imag.* 39, 3749–3757 (2020). <https://doi.org/10.1109/tmi.2020.3004511>
- Chaoui, R., et al.: Recent development in three and four dimension fetal echocardiography. *Fetal Diagn. Ther.* 47(5), 345–353 (2019). <https://doi.org/10.1159/000500454>

10. Liu, R., et al.: Nhbs-net: a feature fusion attention network for ultrasound neonatal hip bone segmentation. *IEEE Trans. Med. Imag.* 40, 3446–3458 (2021). <https://doi.org/10.1109/tmi.2021.3087857>
11. Ren, Z., Wang, S., Zhang, Y.: Weakly supervised machine learning. *CAAI Transactions on Intelligence Technology* 8(3), 549–580 (2023). <https://doi.org/10.1049/cit2.12216>
12. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* 521(7553), 436–444 (2015). <https://doi.org/10.1038/nature14539>
13. Wu, X., et al.: Classification of Alzheimer's disease based on weakly supervised learning and attention mechanism. *Brain Sci.* 12, 1601 (2022). <https://doi.org/10.3390/brainsci12121601>
14. Nageswari, C.S., Prabha, K.H.: Preserving the border and curvature of fetal heart chambers through tdywt perspective geometry wrap segmentation. *Multimed. Tool. Appl.* 77(8), 10235–10250 (2017). <https://doi.org/10.1007/s11042-017-5428-9>
15. Liu, B., et al.: A denoising and enhancing method framework for 4d ultrasound images of human fetal heart. *Quant. Imag. Med. Surg.* 11(4), 1567–1585 (2021). <https://doi.org/10.21037/qims-20-818>
16. Nageswari, C.S., HelenPrabha, K.: Spatially constrained distance regularized level set evolution method for segmentation of hydrops fetalis from ultrasound fetal heart images. *Des. Autom. Embed. Syst.* 22(1–2), 45–64 (2018). <https://doi.org/10.1007/s10617-017-9199-3>
17. Song, C., et al.: Fully automatic ultrasound fetal heart image detection and segmentation based on texture analysis. *Invest. Clin.* 61, 600–608 (2020). <https://researchonline.ljmu.ac.uk/id/eprint/12813>
18. Li, Z., et al.: Automatic image analysis and recognition for ultrasound diagnosis and treatment in cardiac, obstetrics and radiology. *Medical and Biological Image Analysis* (2018). <https://doi.org/10.5772/intechopen.76284>
19. Sushma, T.V., et al.: Classification of fetal heart ultrasound images for the detection of chd. *Innovative Data Communication Technologies and Application*, 489–505 (2021). https://doi.org/10.1007/978-981-15-9651-3_41
20. Bridge, C.P., Ioannou, C., Noble, J.A.: Automated annotation and quantitative description of ultrasound videos of the fetal heart. *Med. Image Anal.* 36, 147–161 (2017). <https://doi.org/10.1016/j.media.2016.11.006>
21. Vargas-Quintero, L., et al.: Left ventricle segmentation in fetal echocardiography using a multi-texture active appearance model based on the steered hermite transform. *Comput. Methods Progr. Biomed.* 137, 231–245 (2016). <https://doi.org/10.1016/j.cmpb.2016.09.021>
22. Mahmood, R., Syeda-Mahmood, T.: Automatic detection of dilated cardiomyopathy in cardiac ultrasound videos. *Annual Symposium proceedings. AMIA Symposium* 2014, 865–871 (2014)
23. Kosuge, A., et al.: mmwave-yolo: a mmwave imaging radar-based real-time multiclass object recognition system for adas applications. *IEEE Trans. Instrum. Meas.* 71, 1–10 (2022). <https://doi.org/10.1109/tim.2022.3176014>
24. Peng, J., et al.: Autonomous recognition of multiple surgical instruments tips based on arrow obb-yolo network. *IEEE Trans. Instrum. Meas.* 71, 1–13 (2022). <https://ieeexplore.ieee.org/abstract/document/9743445>
25. Feng, M., et al.: Relation graph network for 3d object detection in point clouds. *IEEE Trans. Image Process.* 30, 92–107 (2021). <https://doi.org/10.1109/tip.2020.3031371>
26. Guan, L., et al.: A lightweight framework for obstacle detection in the railway image based on fast region proposal and improved yolo-tiny network. *IEEE Trans. Instrum. Meas.* 71, 1–16 (2022). <https://doi.org/10.1109/tim.2022.3150584>
27. Wang, W., et al.: Robust object detection via adversarial novel style exploration. *IEEE Trans. Image Process.* 31, 1949–1962 (2022). <https://doi.org/10.1109/tip.2022.3146017>
28. 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). <https://doi.org/10.1016/j.media.2018.10.003>
29. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60(6), 84–90 (2012). <https://doi.org/10.1145/3065386>
30. He, K., et al.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
31. Redmon, J., et al.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). <https://arxiv.org/pdf/1506.02640.pdf>
32. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
33. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: Yolov4: optimal speed and accuracy of object detection. <https://arxiv.org/abs/2004.10934> (2020)
34. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. [arXiv.org https://arxiv.org/abs/1804.02767](https://arxiv.org/abs/1804.02767) (2018)
35. Liu, W., et al.: Ssd: single shot multibox detector. *Computer Vision – ECCV*, 21–37 (2016). <https://arxiv.org/abs/1512.02325>
36. Chen, C., et al.: Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine* 7 (2020). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7066212/>
37. Shen, D., Wu, G., Suk, H.-I.: Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19(1), 221–248 (2017). <https://doi.org/10.1146/annurev-bioeng-071516-044442>
38. Fiorentino, M.C., et al.: A review on deep-learning algorithms for fetal ultrasound-image analysis. *Med. Image Anal.* 83, 102629 (2023). <https://doi.org/10.1016/j.media.2022.102629>
39. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88 (2017). <http://www.frontendvision.net/2018/wp-content/uploads/2018/11/Litjens2017-SurveyDeepLearning.pdf>
40. Oksuz, I., et al.: Deep learning-based detection and correction of cardiac mr motion artefacts during reconstruction for high-quality segmentation. *IEEE Trans. Med. Imag.* 39(12), 4001–4010 (2020). <https://doi.org/10.1109/tmi.2020.3008930>
41. Isensee, F., et al.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18(2), 203–211 (2020). <https://doi.org/10.1038/s41592-020-01008-z>
42. Hatamizadeh, A., et al.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 272–284 (2022). https://doi.org/10.1007/978-3-031-08999-2_22
43. Ghorbani, A., et al.: Deep learning interpretation of echocardiograms. *npj Digital Medicine* 3(1), 10 (2020). <https://doi.org/10.1038/s41746-019-0216-8>
44. Bello, G.A., et al.: Deep-learning cardiac motion analysis for human survival prediction. *Nat. Mach. Intell.* 1(2), 95–104 (2019). <https://core.ac.uk/download/pdf/196351764.pdf>
45. Dezaqi, F.T., et al.: Deep residual recurrent neural networks for characterisation of cardiac cycle phase from echocardiograms. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 100–108 (2017). https://doi.org/10.1007/978-3-319-67558-9_12
46. Arnaout, R., et al.: An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nat. Med.* 27(5), 882–891 (2021). <https://doi.org/10.1038/s41591-021-01342-5>
47. Pu, B., et al.: Mobileunet-fpn: a semantic segmentation model for fetal ultrasound four-chamber segmentation in edge computing environments. *IEEE Journal of Biomedical and Health Informatics* 26(11), 5540–5550 (2022). <https://doi.org/10.1109/jbhi.2022.3182722>
48. Chan, W.X., et al.: Full cardiac cycle asynchronous temporal compounding of 3d echocardiography images. *Med. Image Anal.* 74, 102229 (2021). <https://doi.org/10.1016/j.media.2021.102229>
49. Github.io, (2020). <https://yaplab.github.io/Research%20FetalHeart%20ImageProc.htm>
50. Jocher, G.: ultralytics/yolov5. GitHub (2020). <https://github.com/ultralytics/yolov5>
51. Adarsh, P., Rathi, P., Kumar, M.: Yolo v3-tiny: object detection and recognition using one stage improved model. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (2020)

52. Qiao, S., et al.: Flds: an intel-ligent feature learning detection system for visualizing medical images supporting fetal four-chamber views. *IEEE Journal of Biomedical and Health Informatics* 26, 4814–4825 (2022). <https://doi.org/10.1109/jbhi.2021.3091579>
53. Jocher, G., Chaurasia, A., Qiu, J.: *Yolo by Ultralytics* (2023). <https://github.com/ultralytics>
54. Leclerc, S., et al.: Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans. Med. Imag.* 38(9), 2198–2210 (2019). <https://doi.org/10.1109/tmi.2019.2900516>

How to cite this article: Wang, G., et al.: 4D foetal cardiac ultrasound image detection based on deep learning with weakly supervised localisation for rapid diagnosis of evolving hypoplastic left heart syndrome. *CAAI Trans. Intell. Technol.* 1–15 (2024). <https://doi.org/10.1049/cit2.12354>