Medical Image Analysis 58 (2019) 101548

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Multi-task learning for quality assessment of fetal head ultrasound images

Zehui Lin[a,1], Shengli Li[b,1], Dong Ni[a], Yimei Liao[b], Huaxuan Wen[b], Jie Du[a], Siping Chen[a], Tianfu Wang[a,*], Baiying Lei[a,*]

[a] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, 518060, China

[b] Department of Ultrasound, Affiliated Shenzhen Maternal and Child Healthcare Hospital of Nanfang Medical University, 3012 Fuqiang Rd, Shenzhen, 518060, China

ARTICLE INFO

Article history:
Received 21 February 2019
Revised 15 July 2019
Accepted 23 August 2019
Available online 6 September 2019

Keywords:
Fetal head standard plane
Ultrasound image quality assessment
Anatomical structure detection
Multi-task learning

ABSTRACT

It is essential to measure anatomical parameters in prenatal ultrasound images for the growth and development of the fetus, which is highly relied on obtaining a standard plane. However, the acquisition of a standard plane is, in turn, highly subjective and depends on the clinical experience of sonographers. In order to deal with this challenge, we propose a new multi-task learning framework using a faster regional convolutional neural network (MF R-CNN) architecture for standard plane detection and quality assessment. MF R-CNN can identify the critical anatomical structure of the fetal head and analyze whether the magnification of the ultrasound image is appropriate, and then performs quality assessment of ultrasound images based on clinical protocols. Specifically, the first five convolution blocks of the MF R-CNN learn the features shared within the input data, which can be associated with the detection and classification tasks, and then extend to the task-specific output streams. In training, in order to speed up the different convergence of different tasks, we devise a section train method based on transfer learning. In addition, our proposed method also uses prior clinical and statistical knowledge to reduce the false detection rate. By identifying the key anatomical structure and magnification of the ultrasound image, we score the ultrasonic plane of fetal head to judge whether it is a standard image or not. Experimental results on our own-collected dataset show that our method can accurately make a quality assessment of an ultrasound plane within half a second. Our method achieves promising performance compared with state-of-the-art methods, which can improve the examination effectiveness and alleviate the measurement error caused by improper ultrasound scanning.

© 2019 Published by Elsevier B.V.

1. Introduction

Ultrasound (US) has been widely used in prenatal diagnosis because of its advantages, including real-time acquisition, low costs, no radiation, and noninvasive. Prenatal ultrasound is the most commonly used method for comprehensive observation of intrauterine growth and development. It is also an important tool for preventing birth defects and assessing the healthiness of the fetuses (Bucher and Schmidt, 1993; Dudley and Chapman, 2002). However, the precondition is to obtain standard planes for prenatal ultrasound diagnosis (Benacerraf, 2008; Gao et al., 2016). By measuring the physiological parameters of the fetus in the standard plane (e.g., standard transthalamic plane), clinicians can accurately assess the growth and development of the fetus. However, it requires extensive clinical experience and comprehensive knowledge of fetal anatomy for the sonographers to obtain standard planes. This is a challenging task, especially for novice sonographers. In clinical practice, the quality of ultrasound plane scanned by the novice sonographers is further evaluated by experienced sonographers, which tends to be subjective and time-consuming (Carneiro et al., 2008). How to improve the quality of ultrasonic plane of novice sonographers has become the practical demand of clinical examination (Huang et al., 2018; Ni et al., 2014; Yang et al., 2019). There are many automatic methods for detecting and recognizing standard planes (Huang et al., 2018; Li et al., 2017; Ma et al., 2017; Wu et al., 2017). For example, automatic measurements of biparietal diameter (BD) and head circumference (HC) (Salomon et al., 2005; Sinclair et al., 2018) are used to estimate fetal weight. In such a process, obtaining the fetal head standard plane (FHSP) is a prerequisite for physicians to perform parameter measurements

---

[*] Corresponding authors.
E-mail addresses: tfwang@szu.edu.cn (T. Wang), leiby@szu.edu.cn (B. Lei).

[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.media.2019.101548
1361-8415/© 2019 Published by Elsevier B.V.

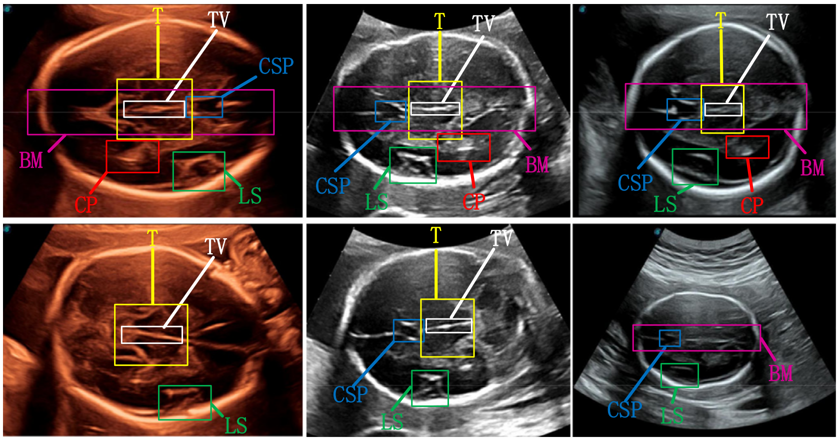


Fig. 1. Illustration of FHSP image in different instruments, top and bottom represent the standard and non-standard planes, respectively. (Lateral sulcus (LS), thalamus (T), choroid plexus (CP), cavum septi pellucidi (CSP), third ventricle (TV), brain midline (BM)).

and anatomical observations. However, to the best of our knowledge, there are few studies on automatic acquisition of FHSP and its subsequent automatic quality assessment.

To obtain FHSP and assess its quality (International Society of Ultrasound in Obstetrics and Gynecology Education Committee, 2010; Pilu et al., 2006; Zhang et al., 2017), it is important to detect the presence of six key anatomical structures (i.e., lateral sulcus (LS), thalamus (T), choroid plexus (CP), cavum septi pellucidi (CSP), third ventricle (TV), brain midline (BM)). It is also vital to determine if the skull is in the middle of the ultrasound plane and occupies larger than 2/3 of overall fan-shape area (FS) (Paladini et al., 2007). Fig. 1 illustrates the FHSP image in different instruments. For FHSP assessment, the clinical protocol needs to specify according to the clinical standards (Salomon et al., 2005). Quality assessment aims to boost the sonographer's judgment whether the obtained ultrasound image is a standard plane or not, because it can output the ultrasound images with the detection of the key anatomical structures (e.g., CSP, LS, CP, etc.), as shown in the following Fig. 1. The protocol refers to the quality control criteria for manually assessing the fetal transthalamic plane. The protocol is reviewed and approved unanimously by the internal research committee at Shenzhen Maternal and Child Health Hospital, Shenzhen, China. The committee is led by a senior radiologist with more than 20 years of experience in the fetal US examination.

The scoring protocol is shown in Table 1, where US images can be scored from 1 to 3 according to the specified protocol. To detect important anatomical structures, CSP and BM are of the utmost importance. If these two anatomical structures cannot be detected, then the ultrasound image is regarded as the non-standard ultrasound image. According to the clinical protocols, ultrasound images of fetal head with a score of 9 and above are considered as a standard one.

Table 1
Quality assessment protocol for fetal head US images. A score is given based on the protocol of stand plane evaluation.

Item	Protocol	Score
LS	Lateral sulcus must be clearly visible	1
CSP	Camera septi pellucidi must be clearly visible	3
CP	The lateral ventricle only shows the choroid plexus, and the posterior horn cannot be clearly and completely displayed.	1
BM	Brain midline shall appear full and clearly visible	3
TV	Third ventricle between the two thalamus	1
T	Symmetrical thalamus on both sides of the brain is clearly visible	1
FS	The skull is in the middle of the ultrasound plane and larger than 2/3 of overall fan-shape area	1
	Scores <9: ×, Scores ≥9: ✔	

There are many challenges involved in quality assessment of the ultrasound images, which are illustrated in Fig. 2. The challenges can be mainly divided into three groups: (1) The quality of ultrasound images is often affected by noise and shadowing effect. (Fig. 2(a)); (2) The scanning angle and the fetal location are unstable due to the rotation of the anatomical structure (Fig. 2(b)); (3) There is a lot of interference with similar tissues or anatomical structures (Fig. 2(c)). It is noting that, the challenge in Fig. 2(a) is caused by the limitation of imaging principle, which is not the primary focus of this work. Hence, in order to tackle challenges in Fig. 2(b) and (c), we propose a multi-task faster regional convolutional neural network (MF R-CNN) (Girshick, 2015; Ren et al., 2015; Xue et al., 2018) architecture for quality assessment of the fetal head US plane. Specifically, we break down the complex quality assessment tasks into a few simple ones, which can be easily handled with CNNs (i.e., six key anatomical structure localization tasks and one FS classification task) (He et al., 2016; Hu et al., 2018; Krizhevsky et al., 2012; Simonyan and Zisserman, 2014). Our proposed method can effectively learn and extract discriminative features from the training images, and is able to perform joint classification and detection tasks simultaneously. Meanwhile, its detection speed is fast enough to fully meet the clinical needs. Inspired by the recent studies on transfer learning (Donahue et al., 2014; Sinno Jialin et al., 2011; Yosinski et al., 2014), we propose a section training method via transfer learning technique, which can improve the speed and accuracy of training (Tajbakhsh et al., 2017). In addition, to further improve the detection results of the network, we also add the prior clinical knowledge module in the network detection process. To sum up, this paper has the following contributions:

- A MF R-CNN architecture is proposed for joint six key anatomical structures detection and FS classification. Our model leverages domain-specific knowledge to train these closely related tasks by sharing low-level features in the early layers, before branching them into independent output streams for each task. The task-specific predictions are then combined to assess the quality of US images.
- Both prior clinical and statistical knowledge are combined to improve the accuracy of test results. Since the location of each anatomical structure is fixed, we compute the total coverage between the anatomical structures and eliminate those with insufficient coverage in the predicted results, which can avoid interference of other structures and tissues in the FHSP.

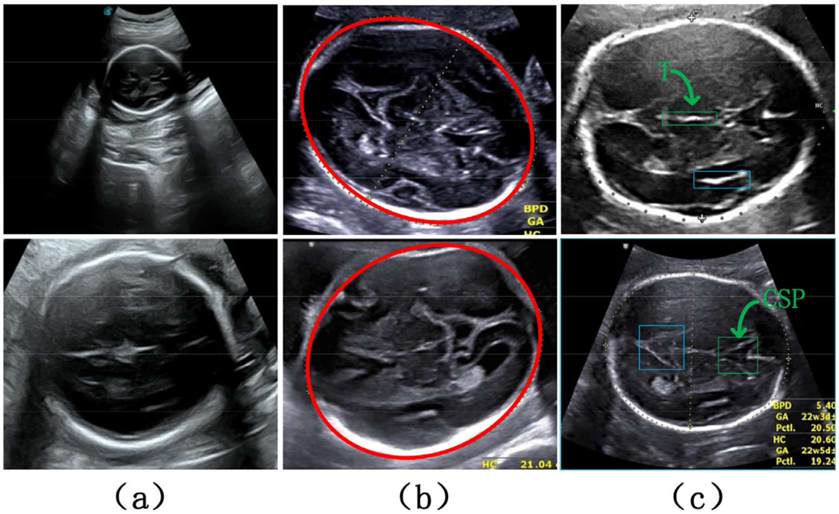


Fig. 2. Illustration of different types of challenges. (a) US images with heavy noise and shadowing effect. (b) Cases of fetal head regions (marked with red circles) with significant variations in size and appearance. (c) US images with two fetal structures which are in similar shape but are actually two different structures in relatively similar positions (green is the target anatomy; blue is the disturbing tissue or anatomy). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- We use section training based on transfer learning to train our network in such way that the classification module with fast convergence rate will not interfere with the detection module with slow convergence rate.

Our proposed method has been evaluated on our self-collected dataset with promising performance. In addition, our model can be extended to the quality control of other planes (e.g., four-chamber-view plane and abdominal plane). It is applicable to a number of ultrasonic instruments with high robustness. The detection speed of the network is fast for clinical auxiliary standard plane acquisition.

2. Related work

In the recent years, with the rapid development of image processing technology, "intelligent ultrasound" (Huang et al., 2018; Li et al., 2017; Ma et al., 2017; Meng et al., 2018; Shin et al., 2018) has become a hot research topic. Powered by the machine learning and deep learning techniques (Shi et al., 2018, 2016), many intelligent auxiliary diagnostic systems have been proposed for fetal US images (Gao et al., 2016; Litjens et al., 2017; Noble, 2016; Yaqub et al., 2016). For instance, Namburete et al. (2018) proposed a multi-task fully convolutional neural network (FCN) architecture to address the problem of 3D fetal brain localization, structural segmentation, and alignment to a referential coordinate system. These methods make simple, computationally inexpensive predictions from 2D slices and are capable of incorporating this information to estimate 3D brain orientation. For example, Xu et al. (2018) proposed a multi-task learning framework for automatic view classification and landmark detection of the organs in the fetal abdominal ultrasound image. The view classification result shows that their proposed method outperforms the human expert. For landmark detection, each landmark-based measurement error is reduced as well. Li et al. (2018) proposed a new iterative transformation network (ITN) for the automatic detection of standard planes in 3D volumes. Under a multi-task learning framework, they introduced additional classification probability outputs to the network to act as confidence measures for the regressed transformation parameters. Li et al. (2017) combined random forests and the prior knowledge to detect the region of interest (RoI) of the fetal head circumference. A non-iterative ellipse fitting method based on geometric distance was utilized for head circumference fitting. The detection speed of this method is quite fast, and the result is similar to the result of manual measurement by the sonographer. Sundaresan et al. (2017) presented a framework for tracking the key variables that describe the content of each frame of freehand 2D ultrasound scanning videos of a healthy fetal heart. They trained classification and regression forests to predict the visibility location and orientation of the fetal heart in the image, and the viewing plane label from each frame. Baumgartner et al. (2017) proposed a SonoNet, which can automatically detect 13 fetal standard views in freehand 2-D ultrasound data as well as provide a localization of the fetal structures via a bounding box. In addition, the network learns to localize the target anatomy using weak supervision based on image-level labels only. Chen et al. (2015, 2017) proposed an automatic framework based on deep learning to detect standard planes. The automatic framework achieved competitive performance and showed potential and feasibility of deep learning for region localization in ultrasound images. The most related work was proposed by Wu et al. (2017), which could automatically detect the existence of two key anatomical structures (stomach bubble and umbilical vein) in the US image of the fetal abdominal through a 4-class classification network architecture. Their method achieved very good classification results. However, it is only suitable for identifying US plane with fewer key anatomical structures, which is unsuitable for six anatomical structure detection of FHSP. Moreover, there are no detailed clinical quality control protocols for FHSP quality assessment of fetal transthalamic plane in the existing methods. Thus, it is desirable to provide clinical assessment of US images with six anatomical structures and one classification task.

3. Methodology

The framework of the proposed method is illustrated in Fig. 3. With the heterogeneous input sources of original US data, MF R-CNN can automatically detect key anatomical structure and classify FS of image simultaneously. Then, the detection results of the anatomical structure are used as input for the prior knowledge module to further improve the detection accuracy. Finally, we can automatically score the US images of the fetal head based on the number of detected anatomical structures and FS classification results to determine whether it is a standard plane.

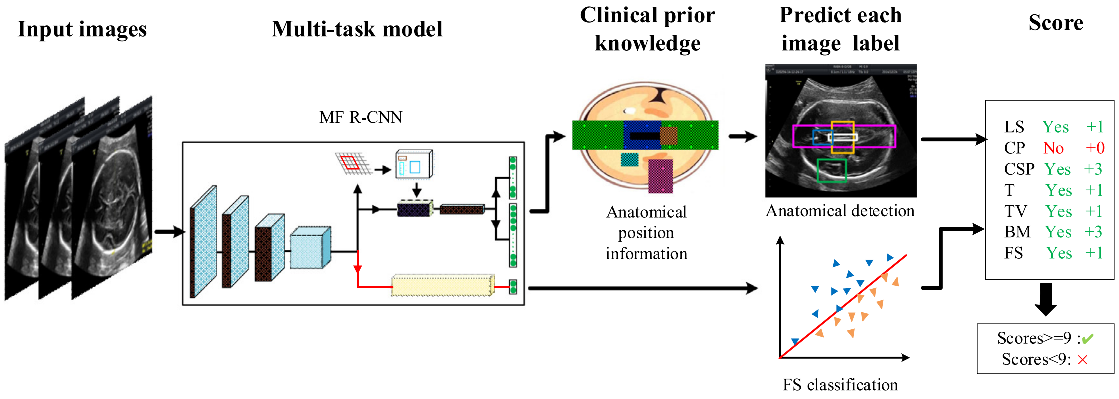


Fig. 3. Framework of the proposed method. The transcerebellar plane to be detected is input into the MF R-CNN to automatically detect key anatomical structures and classify FS. In the detection task, the clinical prior knowledge module is added to further improve the detection accuracy. Finally, the network will score the ultrasound image based on the results of classification and detection, and judge whether it is a standard image based on the score.

3.1. *The multi-task model*

To assess the quality of the fetal head plane, we need to break down complex problems into a few simple ones that can be easily solved with CNNs. We train a multi-task network to simultaneously achieve six localization tasks (whether key anatomical structures are existing) and one classification task (whether FS region is sufficient). Then, using the predicted results of network to decide if a fetal head plane is a standard one or not. The detection of key anatomical structures is in accordance with the doctor's clinical guidelines, which can also be used to eliminate the interference from other ultrasound planes (e.g. transventricular plane, transcerebellar plane, etc.). The classification task enables us to determine whether the FS of US plane is sufficient to meet clinical standards, and it can further determine whether the magnification of the transthalamic plane collected by the sonographer is appropriate or not.

3.1.1. *Convolutional neural network design*

The CNNs are hierarchical models, which are mainly composed of non-linear convolutional layers and pooling layers. By varying the depth and width of the networks, the model power can be controlled, effectively. In a CNN model, the convolutional layers locally calculate pixel dependencies while pooling layers are used to cut down the computational burden by reducing the image resolution, and simultaneously increase the receptive field and invariance. Low-level features obtained from previous layers are more generalized and more task-specific features are found in the high-level layers.

Inspired by the Faster R-CNN, our proposed multi-task model is subdivided into three parts as shown in Fig. 4. Faster R-CNN can learn from the training data to extract useful features, and through the use of joint training and alternative optimization. Next, the Regional Proposal Networks (RPN) module and Fast R-CNN module in Faster R-CNN share the convolution layer features and build a complete end-to-end CNN object detection model to detect object. In our multi-task model, the first part contains low-level features shared by all tasks, which combines these hierarchical features and retains a very distinct and valid deep representation. The second and third parts are the detection and classification module, respectively, which further learn each task-specific feature. This CNN architecture can avoid individual training of CNNs for each task, take advantage of the correlation among different tasks and provide more oversight of learning sharing features. Therefore, in the training process, the joint optimization and alternative training are utilized, which not only ensures the invariance of the low-level features, but also ensures the distinguishability of the seven tasks for the task-specific feature operations.

In our proposed architecture, the input image size is 960 × 720 or 1027 × 813. Inspired by Faster R-CNN (Ren et al., 2015), we rescale the image so that their shorter sides are 600 pixels. All the convolutional layers use fixed kernel size with sliding step size $\delta = 2$. The batch normalization (BN) layers are used after each convolutional layer in our network to resolve the convergence problem and accelerate the training process. We employ rectified linear unit (ReLU) activations after every BN layer. Eventually, our network branches out from the convolutional layer to produce outputs for each specific task.

In the detection module, in order to get accurate location of the anatomical structure, we use the region proposal network (RPN) (He et al., 2015) from faster R-CNN (Girshick, 2015) to generate more accurate region proposals. Here, RPN uses a 3×3 sliding window on the feature map of the shared low-level convolution layer to generate a full connection feature with a length of 512 dimensions. Then, a series of rectangular region proposals are generated. According to the region proposal, RoI pooling layer extracts fixed length feature vector (7×7) from the shared low-level feature map, and then adds two parallel fully connected layers as the output layers. One outputs the bounding-boxes of the position of the anatomy while the other outputs the categories and scores of the predicted anatomy.

In the classification module, because the input image size of the network is relatively large (800 × 600 × 3 or 713 × 600 × 3), as compared to the traditional image size of 224 × 224 × 3 or 227 ×227 × 3, the dimension of the lower layer features for the classification module is still large after the shared low-level layer. To fully utilize the large dimension of features, a cls_block is added to the classification module to further learn the effective features. As shown in Fig. 4(B), cls_block is composed of a different number (0, 1, 2 or 3) of block a or block b in the classification module. Here, block a is inspired from ResNet 50 (He et al., 2016), which is stacked with three residual layers, while block b is inspired by VGG16 (Krizhevsky et al., 2012), which is composed of three convolutional layers stacked together. The most appropriate cls_block and the optimal number of blocks are selected empirically through comparative experiments.

3.1.2. *Training via transfer learning*

Both detection and classification modules are trained independently to modify their convolution layers in different ways. Therefore, we develop a joint learning technique that allows for sharing convolution layers between two modules, rather than learning two modules separately. We design a single network that includes both

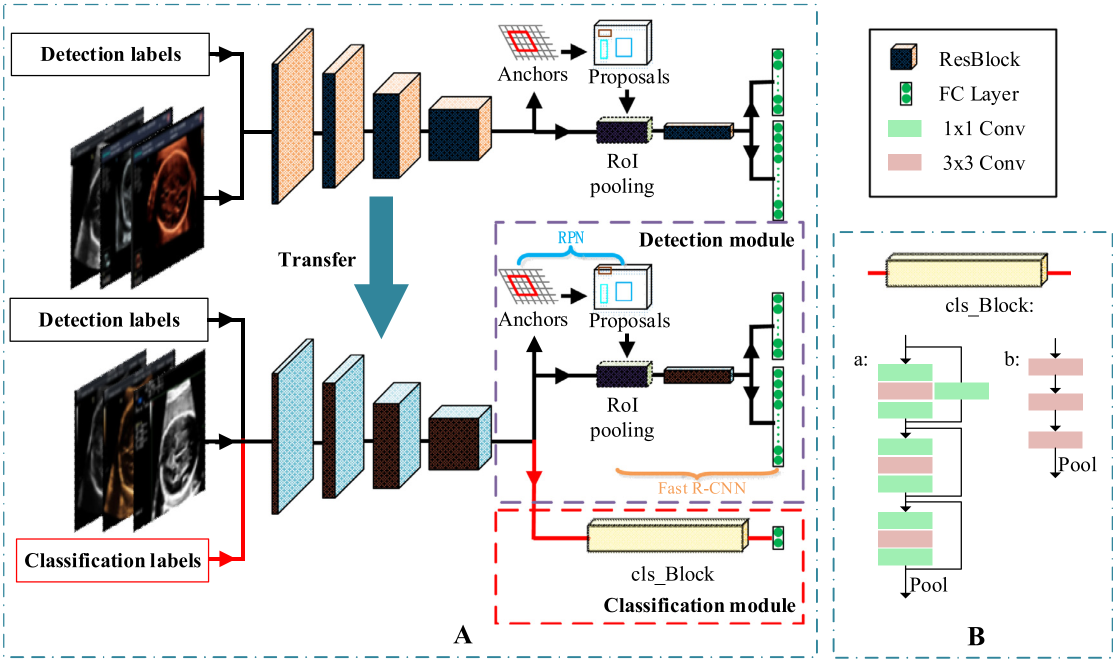


Fig. 4. Proposed architecture of MF R-CNN (A). We only train the detection module first then transfer the trained and shared low-level features from the detection module to MF R-CNN. Afterwards, we jointly train the classification and detection module. MF R-CNN contains three parts, namely, the shared low-level features, the detection, and classification modules. The cyan block is the residual convolution block, and the Fully Connected (FC) layer is represented as the green sphere block. (B) is a zoomed-in schematic of two different blocks for cls_Block in the classification module. Block B(a) is the residual block composed of 1 × 1 Conv. and 3 × 3 Conv. while Block B(b) consists of three 3 × 3 Conv. The purpose of cls_block is to further learn effective features and reduce feature dimensions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

detection and classification module, and then optimize it jointly with back propagation. Inspired by Faster R-CNN (Ren et al., 2015) and transfer learning, we learn shared features via section training and alternative optimization.

Since the loss of the classification module drops faster than that of the detection module, we adopt the section training based on transfer learning. We first train the detection module separately, and then jointly train the detection and classification module later. We develop a practical section training algorithm in the following to train the two modules through alternative optimization.

Step 1: We initialize the shared convolution layer with the pre-trained model via ImageNet, and separately train the network of detection module in the first 10 epochs.

Step 2: In the next step, the parameters of the initially trained low-level feature in the detection module are transferred to MF R-CNN. We then jointly train the detection and classification module using 10 epochs. Before the joint training, the features learned by the shared low-level layer are all from the detection module.

Step 3: To train the classification module, we initialize the shared convolution layer with the training results in Step 2, and only fine-tune the layers that are unique from the classification module so that the two modules share features of the convolution layers.

Step 4: After retraining the detection module and keeping the shared convolution layers fixed, we fine-tune the unique feature from the detection module.

As a result, both classification and detection modules share the same convolution layers and form a unified network via the alternative and iterative training.

3.1.3. Multi-task loss

Our multi-task model is divided into two modules: classification and detection modules. We define loss of the classification module and detection module as L_c and L_d, respectively. The classification module outputs a discrete probability distribution (per US plane), $p = (p^0, \ldots, p^K)$, over $K+1$ categories. In this paper, FS has only two categories, $K = 1$. Here, the classification loss $L_c(p, u)$ is defined as:

$$L_c(p, u) = -\log p_u, \tag{1}$$

where u is the ground-truth label of FS.

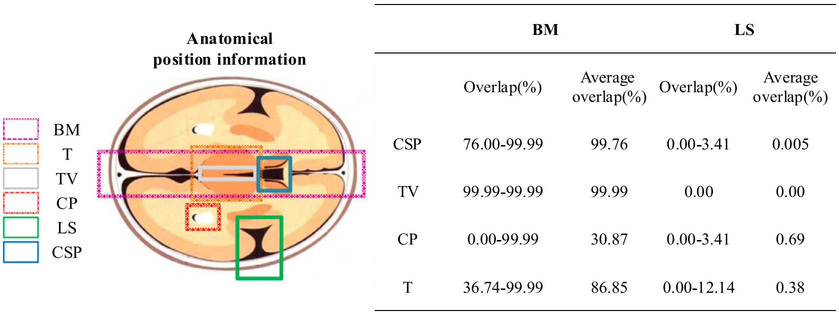
The detection module is composed of RPN and Fast R-CNN, therefore, the loss of the detection module is composed of the losses of these two parts L_r and L_f, respectively. There are two sibling outputs in the RPN: the first one is a discrete probability distribution (per anchor) y_i, i is the index of anchor. The predicted label of anchor is positive when y_i is 1, and is negative when y_i is -1. The second is a vector representing the 4 parameterized coordinates v_i of the predicted bounding box. Like RPN, Fast R-CNN has two sibling outputs. But the first output is the predicted probability of each class, $q = (q^0, \ldots, q^{\mathcal{L}})$, over $\mathcal{L}+1$ classes. In this paper, $\mathcal{L} = 6$ because there are six detection classes. The second sibling output is bounding-box regression offsets t for each $\mathcal{L}$ object class, $t^{\mathcal{L}} = (t_x^{\mathcal{L}}, t_y^{\mathcal{L}}, t_w^{\mathcal{L}}, t_h^{\mathcal{L}})$. The detailed calculation for L_r and L_f are given in Fast R-CNN (Girshick, 2015; Ren et al., 2015; Xue et al., 2018) and Faster R-CNN (Ren et al., 2015). Therefore, the loss function of detection module $\mathrm{L_d(y_i, v_i, q, t)}$ is given by:

$$L_d(y_i, v_i, q, t) = L_r(y_i, v_i) + L_f(q, t). \tag{2}$$

Finally, we design an objective function using the multi-task loss L developed in Fast R-CNN and Faster R-CNN, which is denoted as:

$$L = \alpha_1 L_c(p, u) + \alpha_2 L_d(y_i, t_i, q, v), \tag{3}$$

where α_1 and α_2 refer to the weights for two modules. For the first 10 epochs in the training, only the detection module is trained, so $\alpha_1 = 0, \alpha_2 = 1$; For the last 10 epochs, two modules are trained together, so $\alpha_1 = \alpha_2 = 1$. It is worth noting that in our network, the loss of the classification module converges faster than that of the positioning module. Hence, we develop a section



	BM		LS	
	Overlap(%)	Average overlap(%)	Overlap(%)	Average overlap(%)
CSP	76.00-99.99	99.76	0.00-3.41	0.005
TV	99.99-99.99	99.99	0.00	0.00
CP	0.00-99.99	30.87	0.00-3.41	0.69
T	36.74-99.99	86.85	0.00-12.14	0.38

Fig. 5. The relative position of the clinical anatomical structure (a). The statistical information of the position coverage of two anatomical structures, and the distribution of overlap between BM, LS and CSP, TV, CP, and T (b).

training technique. We do not use joint training by changing the weights α_1 and α_2, because even if α_1 is low, the interference of the classification module in the training process is still very huge.

3.2. Clinical prior knowledge

Due to the similar anatomical structure and tissue effects in the ultrasound image (as shown in Fig. 2(c)), the detection methods are vulnerable to misdetection of individual anatomical structures especially TV. To reduce the false detection of anatomical structure, we add clinical prior knowledge in our algorithm to improve the accuracy of anatomical structure.

Most of the object detection methods based on bounding boxes use non-maximum suppression (NMS) in both training and test stage. However, during the test stage, the NMS role is to sort the bounding boxes by score, then retain the box with the highest score, and delete other boxes with overlapping areas less than the specific intersection-over-union (IoU) value. The IoU is defined between the ground truth (area^{GT}) and predicted image (area^{P}):

$$IoU = \frac{\left|area^{GT} \cap area^{P}\right|}{\left|area^{GT} \cup area^{P}\right|} \tag{6}$$

The IoU value specified in NMS has great influence on the detection result. If the IoU value is too high, it can easily lead to object omission. On contrary, if the IoU value is too low, object misdetection is likely to increase. Although there are many studies (Bodla et al., 2017; Hosang et al., 2017; Jiang et al., 2018) using different state-of-the-art approaches such as Soft NMS (Bodla et al., 2017), and the NMS algorithms in the IoU-net (Jiang et al., 2018) improves detection results, the improvement for our method is still limited. We further improve the detection accuracy from the prior clinical knowledge rather than the specific IoU value, which make the detection results of the network unaffected by the IoU value of the NMS.

Specifically, we identify and analyze FHSP using its informative clinical knowledge as below: (1) in each US plane, there is only one detection result for each anatomical category; (2) the relative position between the anatomical structures is fixed. As shown in Fig. 5(a), majority of the location areas of the BM and CSP, T and TV are overlapped, while that of LS and CP, CSP, T, and TV do not have significant overlap. Hence, we use this clinical knowledge to further improve our detection results. We first use an *overlap* formula to represent the overlapping relationship of location areas:

$$overlap = \frac{|area\ A \cap\ area\ B|}{area\ B}, \tag{7}$$

where *area A* indicates the anatomical area with clear outline and high detection accuracy, and *area B* indicates the anatomical structure area, which is easily interfered by other anatomical structures and tissues. From the preliminary test results of the network, it can be seen that the detection results of LS and BM are the best, reaching 98.5% and 98.4%, respectively. Therefore, we can use the position information of these two anatomical structures to supervise the position information of other anatomical structures, and eliminate the anatomical structure interference and tissue interference results. Hence, the accuracy of detection anatomical structures can be boosted.

As shown in Fig. 5(b), *overlap* is used to calculate the statistical information of the position coverage between the two anatomical structures, which calculates the overlap between BM, LS and CSP, TV, CP, and T, respectively. The coverage range of CP and BM is relatively wide. The initial detection accuracy is high due to its obvious shape. Therefore, we do not use the clinical anatomical position information to improve it. During the detection process, we add the prior knowledge to the initial display result of the network (e.g., a filter to remove the wrong result), and the specific algorithm is implemented in Algorithm 1.

Algorithm 1
An iterative algorithm used to reject high-score interference anatomical structures and tissue detection results.

```
Input:   I: the number of all detection results for a single image.
         '$\mathcal{L}$: category of predicted anatomy. $\mathcal{L} \in$ (BM, LS, CSP, TV, CP, T)
         '$t^{\mathcal{L}}$:the bounding box of each predicted anatomy.$t^{\mathcal{L}} = (t_x^{\mathcal{L}}, t_y^{\mathcal{L}}, t_w^{\mathcal{L}}, t_h^{\mathcal{L}})$
         1: Find the $t^{BM}$ and $t^{LS}$ corresponding to BM, LS in the detection result. (If BM or LS has multiple prediction results, take the result with
         the highest prediction score; if BM or LS does not exist, there is no operation.);
         2: Set the number of iterations $i \in$ I,initialize $i = 0$;
         3: For the detected $t^{\mathcal{L}} \in$ ( $t^{CSP}, t^{TV}, t^{T},$), the following operations are performed :
             1) Calculate overlap(overlap$_{\mathcal{L}_{BM}}$, overlap$_{\mathcal{L}_{LS}}$) between $t^{\mathcal{L}}$ and $t^{BM}$, $t^{LS}$, respectively.
             2) Determine whether overlap$_{\mathcal{L}_{BM}}$, overlap$_{\mathcal{L}_{LS}}$is in the range of Fig. 5(b).
             3) If overlap$_{\mathcal{L}-BM}$ and overlap$_{\mathcal{L}_{LS}}$ are not in the range, we will delete the prediction result $t^{\mathcal{L}}$.
         4:$i = i + 1$
         5:end until stop sign ($i = I$)
         6: Each type of detected result is re-examined. If there are multiple test results in the same class, only the test result with the highest
         predicted score is retained.
Output:  Select bounding boxes results
```

3.3. Implementation details

All experiments are performed on a computer with CPU Inter Xeon E5-2680 @ 2.70 GHz, GPU NVIDIA Quadro K4000, and 128G of RAM. The model is implemented in the Python programming language using the Keras interface to Tensorflow. In our implementation, we initialize the learning rate to $\alpha \mathrm{lr}_1 = 10^{-5}$ for the detection module and $\alpha \mathrm{lr}_2 = 10^{-6}$ for classification module, and the learning momentum is set to $\rho \mathrm{a} = 0.9$. We first separately train 10 epochs in detection module, and then jointly train 10 epochs in detection and classification module. The training takes approximately 2 h/epoch.

4. Experiments and results

4.1. Experimental setup

Our dataset is collected from Shenzhen Maternal and Child Health Hospital, and age of the fetus ranges from 14 to 28 weeks. These images are collected from different types of ultrasonic devices such as Siemens, Mindray, and Philips. Apart from conventional ultrasonic images, there are also pseudo-color ultrasonic images, which increase the diversity of ultrasonic images. After normalizing the ultrasound images, the selected images are first manually labeled by sonographers with 8 years of fetal ultrasound experience. We use about 80% of the images (1451 images) as the training set, and the remaining 20% (320 images) as the test set. Actually, our training data set and test data set are pre-specified. In details, we collect 1–3 images from each patient, and ensure that the data of the subject of the test set and the training set are collected separately. The data distribution of the training and test set for detection and classification task is shown in Tables 2 and 3, respectively. Data augmentation is also implemented (mirror, rotation) in order to expand the dataset.

For the classification module, feature dimension remains large after 4 convolutional blocks. We add blocks after sorting branches to reduce the feature dimension. As shown in Fig. 4(b), we use two different blocks (block a and block b) as cls_block, and the number of cls_blocks added to the classification module is also uncertain. Seven different architectures (O, A1, A2, A3, B1, B2, B3) are tested, where O means no addition of cls_block in the classification module, A1, A2, and A3 represent the addition of 1, 2, and 3 block a in the classification module, respectively. Similarly, B1, B2, and B3 also represent the addition of block b in the classification module. By changing the depth of the classification network, we explore its impact on overall network performance, and aim to find models that balance capabilities and data availability.

Table 2
The number of anatomical structures contained in the training set and the test set.

	Train	Test
LS	1363	317
BM	1383	317
CSP	985	273
CP	1382	314
TV	1416	315
T	1415	315

Table 3
The number of positive and negative samples of FS in the training set and test set.

	Train	Test
Positive	646	180
Negative	805	140

4.2. Evaluation metrics

For detection module, weighted foreground IoU is calculated between the prediction and manual locations. For accurate detection, only IoU value greater than 0.5 is considered as the correct detections. We also use average precision (AP) to evaluate the detection results of each anatomical structure and mean AP (mAP) to evaluate the overall performance of the detection network. For detection models, we qualitatively assess the location results by visualizing the detections of all anatomical structures in an image.

For classification module, popular evaluation metrics such as overall accuracy (ACC), precision (Pre), sensitivity (Sen), specificity (Spec), F1-score (F1), and area of receiver of operation curve (AUC) are computed to quantitatively evaluate all classification models. To demonstrate the efficacy of our proposed method, the intermediate and final FHSP scoring results are compared with the corresponding manual annotations.

4.3. Classification results

In the classification modules, we continue to add the convolution block to learn effective features at the expense of increasing the depth of the network. In addition, the changes in network structure in classification module affect the shared feature and detection results. We set up seven comparative experiments to show effects of architectures on classification and detection result.

From Table 4, it is observed that the specificity results of our proposed method are superior to the existing methods (AlexNet, VGG16, ResNet50). For example, B1 and B2 achieve the best result with a classification accuracy of 96.25%, a precision of 97.76%, an AUC of 98.89% and a F1-score of 95.68%, which shows that adding 1 or 2 block b in the classification module can lead to better classification results. Fig. 8 shows ROC results of different methods, which is consistent with the results in Table 4.

Data visualization can directly characterize the performance of the classifier. We use t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008) technique to visualize our datasets and deep representations extracted from different networks, and deep representational features of FC layers of different classification module extracted from the test datasets. T-SNE is an effective method for data dimension reduction and visualization. T-SNE is a nonlinear dimensionality reduction algorithm, which can effectively map high-dimensional data to low-dimensional space and maintain the local structure of data as in high-dimensional space. Fig. 7 shows the t-SNE visualization for samples of two classes. Red and cyan represent non-FS and FS, respectively. The mixed distribution of the raw test datasets shows that FS and non-

Table 4
Classification result comparisons of our method with different senarios, our method vs. other methods.

Method	Prec	Sen	ACC	F1	Spec	AUC
AlexNet	92.19	92.85	93.43	92.52	93.89	98.66
VGG16	91.94	**97.85**	95.31	92.56	93.33	98.11
VGG19	94.73	90	93.43	92.30	96.11	98.46
ResNet50	94.85	92.14	94.37	93.47	96.11	96.83
O	89.40	96.43	93.43	92.78	91.11	97.81
A1	97.76	93.57	96.25	95.62	98.33	98.29
A2	96.35	94.28	95.93	95.30	97.22	98.17
A3	83.95	97.14	90.62	90.06	85.55	97.86
B1	**97.76**	93.57	96.25	95.62	**98.33**	98.84
B2	96.37	95	**96.25**	**95.68**	97.22	**98.89**
B3	93.57	93.57	94.37	93.57	95	98.31

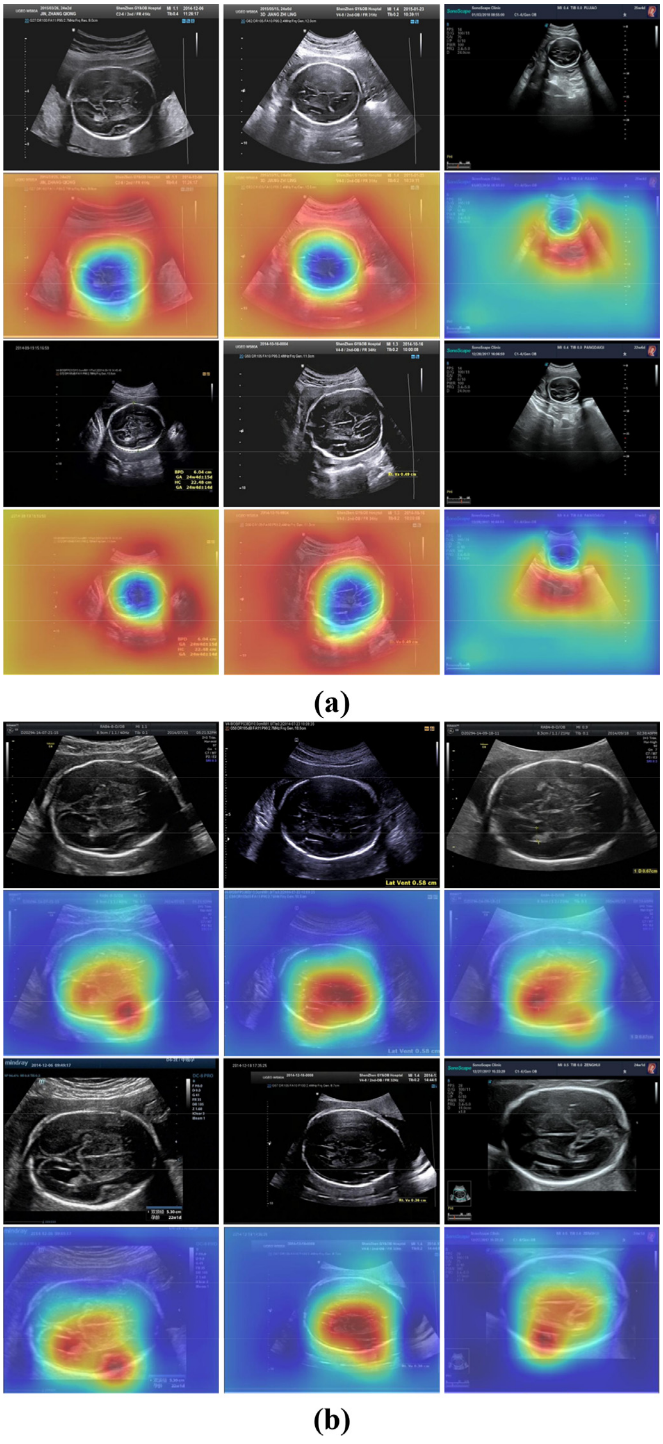


Fig. 6. CAMs for unique layer of classification modules (B1). (a) FS classification result when network is 0 (non-FS), (b) FS classification result when network is 1 (FS).

FS data have high intra-class and low inter-class variations. In contrast, the output features of the FC layer of different classification modules show that the FS and non-FS are in a separated state, which demonstrates the effectiveness of the classification and discriminative ability of the cls_block introduced in this paper.

Fig. 6 shows the class activation maps (CAM) (Chattopadhay et al., 2018; Selvaraju et al., 2017; Zhou et al., 2016) for the unique layer of classification modules (B1). CAM is a tool that helps visualize CNN. With CAM, we can clearly observe which area of the picture the network focuses on. Here, (a) and (b) represent the predicted results of the classification module as 0 and 1 (non-FS and FS), respectively. It can be clearly seen that when the classification result is 0, the network emphasizes the distance from the middle skull area to the scanning fan area. When the classification result is 1, the network emphasizes the area of the brain and the distance from the brain to the bottom of the scanned fan area.

4.4. Detection results

The detection result of the anatomical structure is considered accurate when the detection result of a key anatomical structure and the ground truth's IOU value is higher than 0.7. According to this definition, the detection results of BM, T, CP, CSP, and LS using the test dataset (320 ultrasound images) are quite good. However, the detection result of TV is only 80%, which is lower than the other anatomical structures. The reason is that, the relative smaller and flat anatomical structure of TV makes it difficult to detect.

As shown in Table 5, we observe that our method has the highest accuracy in terms of mAP as compared to other methods (SSD (Liu et al., 2016), YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017), Faster R-CNN (Ren et al., 2015)). Our method only requires approximately 0.6 s to detect an ultrasound plane, which is fast enough to meet clinical needs. Since the changes of cls_block of the classification module affect the detection results, we show the corresponding detection results of 7 architectures. As shown in Table 8, the test results of MF R-CNN using B1 architecture achieve the best performance while the sub-optimal is using B2 architecture. The observation indicates that adding shallow convolutional layers in cls_block could have a better impact on the detection result. In addition, as can be seen from visualization results in Fig. 9, our method has almost no missed detection and false detection rates as compared to other methods. The green, red, yellow, blue, green, and purple bounding boxes indicate the LS, CP, T, CSP, TV, and BM, respectively.

In Table 6, our proposed method compares with a single network in terms of network parameters and average detection speeds of multiple US planes. The parameters of our multi-task network are close to those of two separate networks (Faster R-CNN + ResNet50 or Faster R-CNN +VGG16). When detecting multiple images, our method has the similar speed with the two separate networks. Since our multi-task network is an end-to-end network framework, we only need to load the weights once and can output the results of classification and detection at the same time when only one ultrasonic plane is detected. However, if an ultrasonic image is detected by a classification and location network, the weight needs to be loaded twice to obtain the results of classification and detection, which increases the detection time. Therefore, our method possesses a relatively higher speed in detecting a single image.

4.5. Clinical prior knowledge and transfer learning

As shown in Table 7, we change the IoU in NMS during the test of MF R-CNN (B1). It can be seen that the detection result is greatly influenced by IoU, and the test result improves as IoU decreases. It is worth noting that “ours (B1)+C” represents the network output with the addition of the clinical prior knowledge module. We can see that the detection result of “ours (B1)+C” is the best, which is unaffected by IoU. Moreover, with the prior knowledge, the detection accuracy of each anatomical structure is significantly improved, especially for CSP and TV, which are applicable to other anatomical structures as well.

We summarize the results of detection and classification of different cls_blocks in MF R-CNN in Table 8. We also demonstrate the

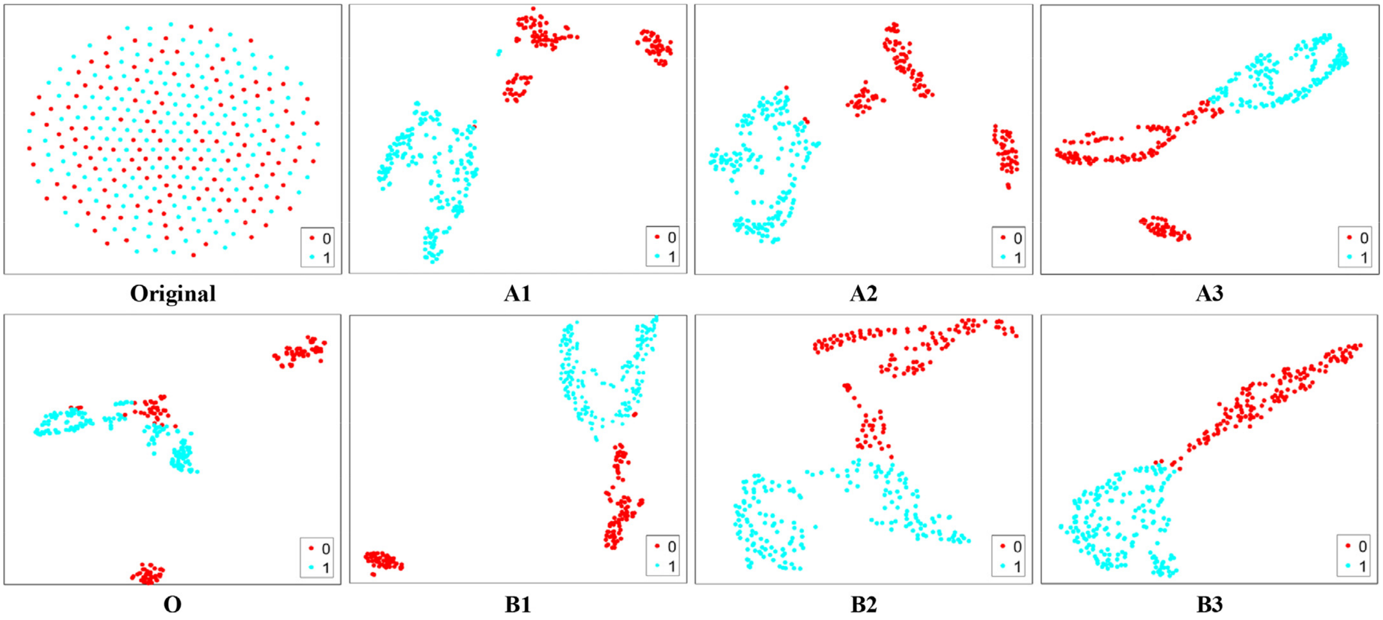


Fig. 7. Visualization with the t-SNE technique for samples of two classes. 0 and 1 represent non-FS and SF, respectively. Original: Raw test data; O, A1-A3 and B1-B3: Extracting the features of the FC layer of different classification modules (O, A1-A3, and B1-B3) from the test dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

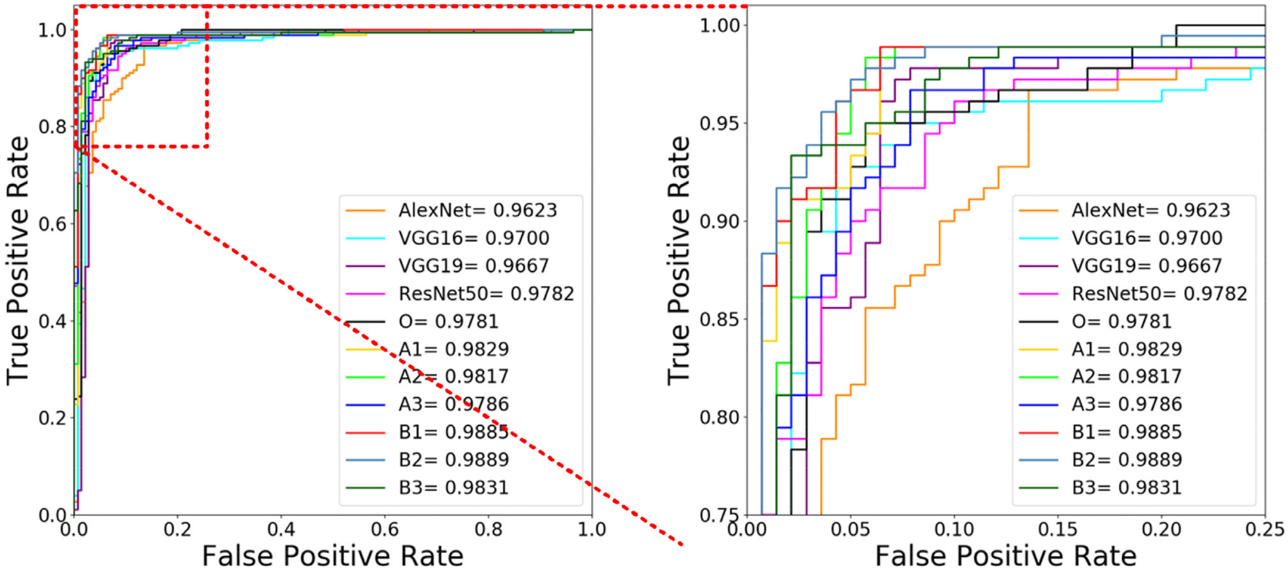


Fig. 8. ROC curves of classification results of our method with other methods.

Table 5
Comparisons about detection results between our method and other methods.

Method	TV	BM	T	CP	CSP	LS	mAP
SSD	32.37	88.47	86.70	61.99	50.50	77.56	66.27
YOLOv2	23.46	82.90	46.02	78.26	73.81	95.37	66.63
Faster R-CNN (VGG16)	78.94	99.20	94.92	87.66	81.26	98.35	90.06
Faster R-CNN (Resnet50)	75.57	98.48	94.81	92.70	85.14	98.53	90.87
O	81.59	98.65	92.63	94.28	89.50	97.96	92.44
A1	80.08	96.85	94.34	94.02	89.44	98.22	92.16
A2	77.35	98.11	92.82	**95.86**	84.86	97.29	91.06
A3	78.68	98.90	94.58	92.42	85.73	**98.73**	91.51
B1	**82.50**	**98.95**	93.89	95.82	89.92	98.46	**93.26**
B2	78.46	98.74	94.94	94.78	**91.47**	98.07	92.74
B3	76.11	98.56	**95.48**	91.97	87.88	98.01	91.34

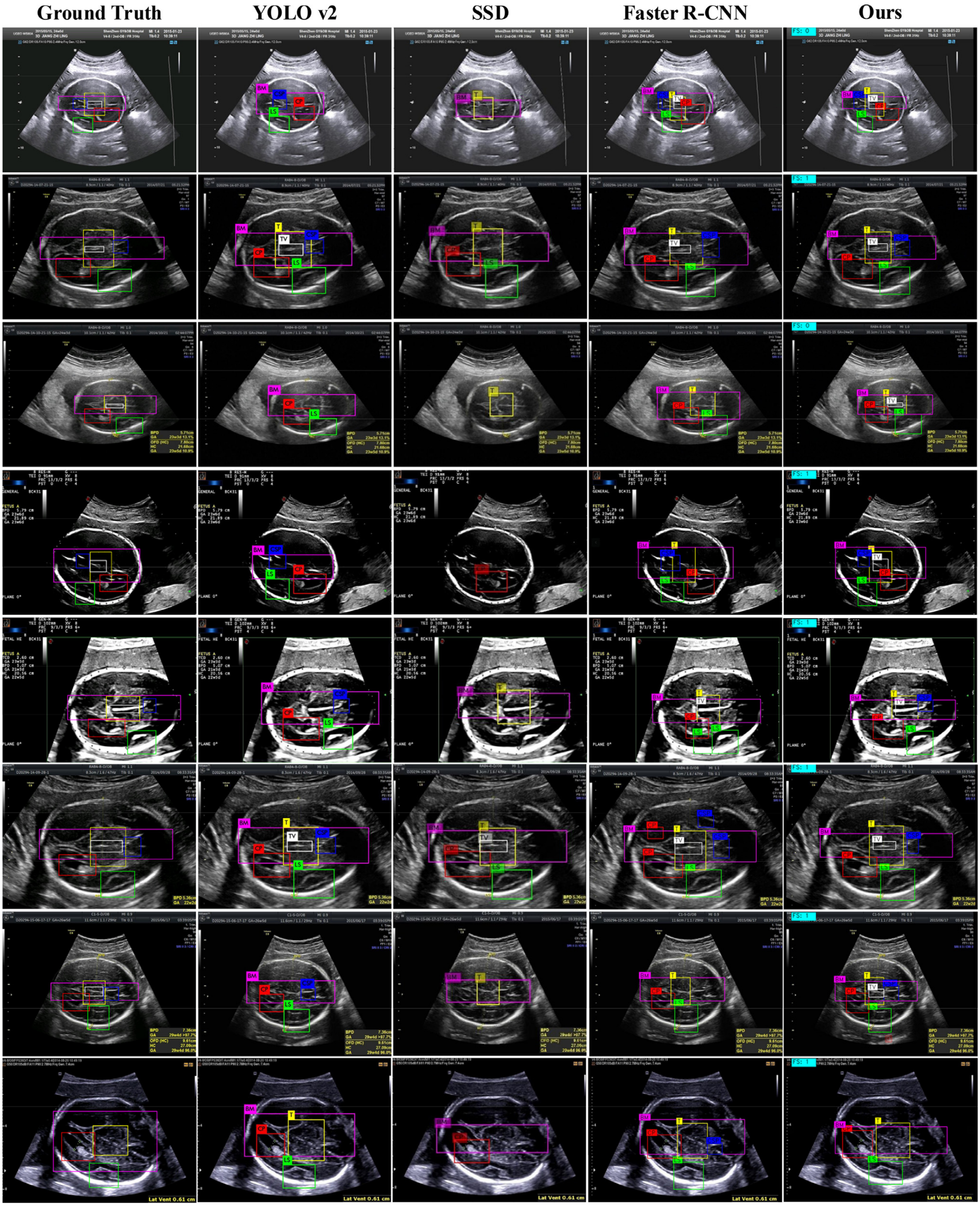


Fig. 9. Visualization of detection results. The left-to-right detection results are: Ground-truth, YOLOv2, SSD, Faster R-CNN, and ours (B1+A). The green, red, yellow, blue, green, and purple bounding boxes indicate the LS, CP, T, CSP, TV, and BM, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Single-task and multi-task methods with different architectures, speed and parameters.

	Method	Speed (s)	Parameters (M)
Single-task	AlexNet	0.015	26.01
	VGG16	0.020	67.04
	VGG19	0.020	72.10
	ResNet50	0.053	22.50
	Faster R-CNN(VGG16)	0.534	130.45
	Faster R-CNN(ResNet50)	0.586	27.07
Multi-task	O	0.601	27.08
	A1	0.658	41.30
	A2	0.655	58.20
	A3	0.653	75.00
	B1	0.604	36.09
	B2	0.632	42.84
	B3	0.647	54.09

Table 7
Results of prior knowledge modules using different IoU.

IOU	TV	BM	T	CP	CSP	LS	mAP
0.9	70.00	86.14	79.95	77.51	77.94	90.07	80.27
0.7	80.04	96.99	91.51	93.45	85.40	97.79	90.86
0.5	82.50	98.95	93.89	95.82	89.92	98.46	93.26
0.3	84.21	98.96	94.08	96.10	90.72	98.48	93.76
0.1	84.29	98.96	94.08	96.11	90.72	98.48	93.78
Our(B1+C)	84.58	98.96	94.34	96.51	91.07	98.75	94.04

effect of different training methods and clinical prior knowledge added to MF R-CNN on the accuracy of network results, where I (initialization) represents MF R-CNN without section training and clinical prior knowledge module. C (clinical prior knowledge) represents MF R-CNN without section training, but with clinical prior knowledge module. S (section training) represents MF R-CNN using section training, but without clinical prior knowledge module. A (all) represents MF R-CNN, which uses both section training and clinical prior knowledge. In terms of mAP, except for A2 and A3, the detection results of A are better than those of C and S, which indicates that adding the clinical prior knowledge module into the network or adopting section training improves accuracy of the classification and detection results. In both A2 and A3, the detection result of C is slightly better than that of A. The classification results are unaffected by clinical prior knowledge module. Furthermore, we observe that the classification results are significantly affected by section training.

Fig. 10 shows the scoring evaluation result of our method. In the US plane, our method simultaneously displays the detection and classification results to facilitate the sonographer's observation. In addition, for comparison with the sonographer's score, the results of our proposed method are shown in the lower left, while scores by sonographers are displayed on the lower right in Fig. 10. It can be seen that our method can effectively evaluate the quality of each ultrasound image by combining clinical criteria.

5. Discussions

In this work, we elaborate the FHSP quality assessment method. Extensive experiments are carried out to demonstrate the effectiveness of the proposed MF R-CNN network. The proposed method is an end-to-end multi-task learning model, which achieves good detection and classification results, simultaneously. In our proposed method, the convergence speed of classification module is faster than that of detection module. We devise the section training method via transfer learning to train the detection module first, and then combine the training of both classification and detection module. Accordingly, the features in the classification mod-

Table 8
Detection and classification results of different cls_blocks are added into the network. I: initialization without section training and prior knowledge; C: clinical prior knowledge without section training; S: section training without prior knowledge; A: all with both section training and clinical prior knowledge.

Method		Detection							Classification					
		TV	BM	T	CP	CSP	LS	mAP	Pre	Sen	ACC	F1_score	Spec	AUC
O	I	72..77	97.08	93.74	85.94	83.88	97.50	88.49	93.57	93.57	94.37	93.57	95.00	98.18
	C	77.10	97.15	94.23	92.02	87.90	98.85	91.21						
	S	81.59	98.65	92.63	94.28	89.50	97.96	92.44	89.40	96.43	93.43	92.78	91.11	97.81
	A	81.81	98.69	92.75	95.04	90.64	98.19	92.85						
A1	I	74.97	98.42	93.90	90.84	88.70	98.31	90.81	87.26	**97.85**	92.81	92.25	88.89	98.28
	C	78.06	98.63	94.17	93.32	92.04	**99.03**	92.54						
	S	80.08	96.85	94.34	94.02	89.44	98.22	92.16	**97.76**	93.57	**96.25**	**95.62**	**98.33**	98.29
	A	80.29	96.91	94.42	94.28	91.02	98.80	92.62						
A2	I	73.83	97.43	93.71	90.67	86.83	96.41	89.81	93.70	95.71	95.31	94.69	95.00	98.13
	C	76.22	97.56	94.35	93.73	91.22	97.63	91.78						
	S	77.35	98.11	92.82	95.86	84.86	97.29	91.06	96.35	94.28	95.93	95.30	97.22	98.17
	A	78.37	98.11	93.01	96.31	86.71	97.59	91.68						
A3	I	76.48	97.68	96.75	93.06	79.66	98.30	90.30	87.50	95.00	91.87	91.05	89.44	97.94
	C	80.00	97.83	**97.30**	95.83	85.49	98.98	92.57						
	S	78.68	98.90	94.58	92.42	85.73	98.73	91.51	83.95	97.14	90.62	90.06	85.55	97.86
	A	78.99	98.97	94.69	92.97	86.99	98.92	91.92						
B1	I	81.63	97.24	94.97	89.48	85.55	97.64	91.09	91.83	96.42	94.68	94.07	93.33	98.57
	C	82.52	97.27	95.10	92.11	87.74	98.38	92.19						
	S	82.50	98.74	93.89	95.82	89.92	98.46	93.26	**97.76**	93.57	**96.25**	**95.62**	**98.33**	98.84
	A	**84.58**	**98.96**	94.34	**96.51**	91.07	98.75	**94.04**						
B2	I	73.92	98.21	93.41	91.55	85.30	97.24	89.94	92.19	92.85	93.43	92.52	93.89	98.31
	C	78.11	98.31	93.81	93.71	89.60	97.83	91.89						
	S	78.46	98.74	94.94	91.97	87.88	98.01	91.34	96.37	95	**96.25**	95.68	97.22	**98.89**
	A	80.01	98.75	95.41	94.96	**92.58**	98.20	93.32						
B3	I	74.04	98.19	93.07	90.51	84.33	96.59	89.46	91.60	93.57	93.43	92.57	93.33	97.76
	C	77.04	98.36	93.69	89.01	89.01	97.95	91.44						
	S	76.11	98.56	95.48	91.97	87.88	98.01	91.34	93.57	93.57	94.37	93.57	95	98.31
	A	76.97	98.56	95.51	92.48	88.70	98.23	91.74						

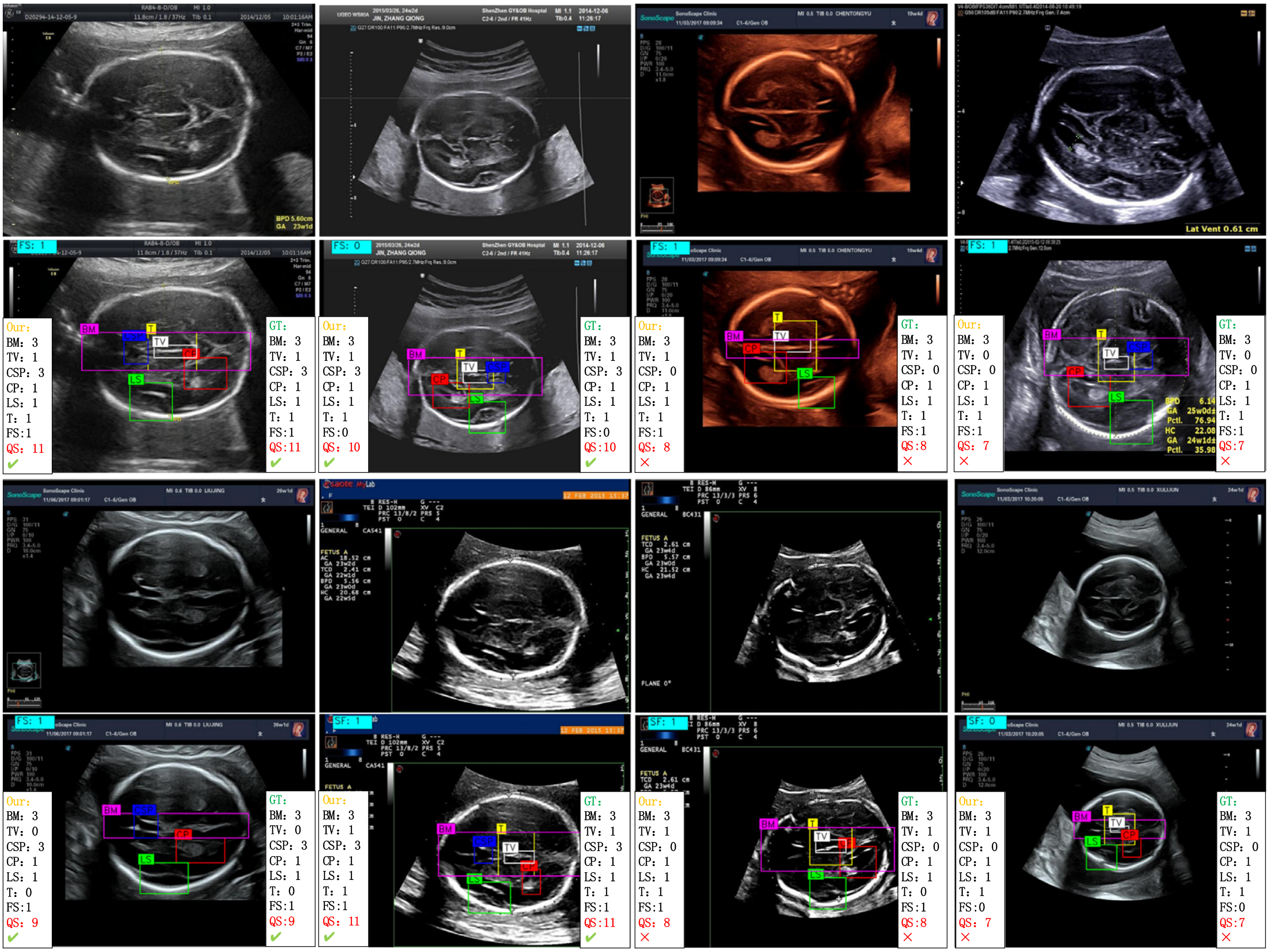


Fig. 10. Scoring process of our proposed method. Our method means that we use both section training and clinical prior knowledge. The FS classification results are shown on the upper left of the US plane, and the anatomical detection results are also shown on the US plane. GT is the result of the sonographer. QS is the total score. ✔: Standard; ×: Non-standard.

ule can avoid interfering with the features in the shared low-level layer and the detection module. In existing methods (e.g., Faster R-CNN, YOLO, SSD, etc.), the detection results are greatly affected by the IoU value between the same class. Our study exploits the clinical prior knowledge since the relative position of the detected anatomical structure is fixed in only one detection task. As a result, the detection results are unaffected by the IoU value, and the detection accuracy is further improved.

Although our method achieves quite impressive results, there are some limitations. First, the current research work focuses on normal clinical US planes from healthy babies and mothers. In future studies, we will validate the framework using more pathological cases. We will collect larger and more representative datasets. Second, there remain some detection and classification errors of FHSP. Fig. 11 shows examples of false FHSP quality assessment by our method. Most of these errors are due to misdetection of individual anatomical structures. For better FHSP quality assessment, the anatomical structures are clear and the test results are similar to the sonographer's results. In addition, if the misdetection is from the BM and CSP regions with high scores, it can affect the overall FHSP assessment quality. This is still a challenging task. To further improve the robustness and accuracy of detection and classification, adding some attention module to the network might boost the performance by detecting key anatomical structures in an attentive way.

The proposed work further boosts the automated analysis of quality assessment in ultrasound images. The method is implemented on 2D ultrasound data due to its ease availability. However, 2D ultrasound is still very challenging for automated prenatal diagnosis due to a number of factors such as noise effects, fetal gestational age differences, fetal position differences etc. To overcome these challenges and to build an applicable and feasible diagnostic system, we will collect more 2D ultrasonic planes and ultrasonic videos to build a larger ultrasound database. To further improve the robustness and accuracy of detection and classification, adding some attention module to the network might boost the performance by detecting key anatomical structures in an attentive way. Finally, there are differences in ultrasonic data collected by different types of ultrasonic instruments, which may affect the accuracy of the results. For example, there are few pseudo-color ultrasonic data, and the test results of pseudo-color ultrasonic data are relatively poor. We will consider combining GAN to transform the data

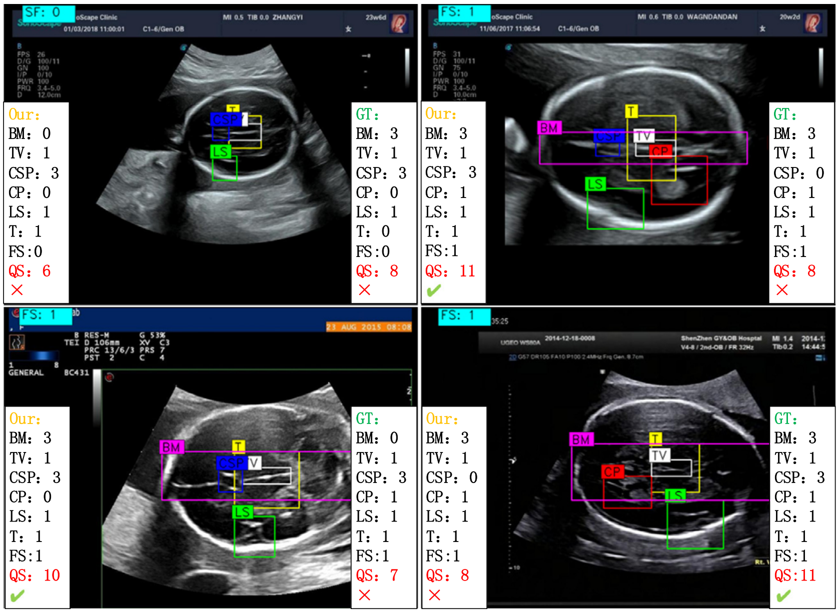


Fig. 11. Examples of false quality assessment of FHSP by our method.

of different instruments to alleviate the deviation caused by different instrument data.

6. Conclusions

In this paper, we propose an automatic technique for quality assessment of fetal head in ultrasound images. Our multi-task model MF R-CNN is based on deep learning for automatic assessment image quality. Our proposed method can automatically locate the six anatomical structures and classify whether the SF of the US plane is sufficient. We also add clinical prior knowledge in the detection process to improve the detection accuracy of the anatomical structure. The clinical knowledge enables our system to score the anatomical structures similar to the clinical protocols. Thus, the results of FHSP detection and classification are closer to manual scoring by sonographers. The experimental results show that our method can achieve quite remarkable performance and outperform related algorithms, which is sufficient to meet clinical needs. Finally, we believe that our method can be extended to other ultrasound image tasks which is yet to be explored.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported partly by National Natural Science Foundation of China (nos. 61871274, 61801305 and 81571758), National Natural Science Foundation of Guangdong Province (no. 2017A030313377), Guangdong Pearl River Talents Plan (no. 2016ZT06S220), Medical Scientific Research Foundation of Guangdong Province, China (no. B2018031), Shenzhen Peacock Plan (nos. KQTD2016053112051497 and KQTD2015033016104926), and Shenzhen Key Basic Research Project (nos. JCYJ20170413152804728, JCYJ20180507184647636, JCYJ20170818142347251 and JCYJ20170818094109846).

References

Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Kainz, B., Rueckert, D., 2017. Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE Trans. Med. Imaging 36, 2204–2215.

Benacerraf, B.R., 2008. Ultrasound of Fetal Syndromes. Elsevier Health Sciences.

Bodla, N., Singh, B., Chellappa, R., Davis, L.S., 2017. Soft-NMS—improving object detection with one line of code. In: IEEE International Conference on Computer Vision, pp. 5562–5570.

Bucher, H.C., Schmidt, J.G., 1993. Does routine ultrasound scanning improve outcome in pregnancy? Meta-analysis of various outcome measures. Br. Med. J. 307, 13–17.

Carneiro, G., Georgescu, B., Good, S., Comaniciu, D., 2008. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Trans. Med. Imaging 27, 1342–1355.

Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision, pp. 839–847.

Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., Heng, P.A., 2015. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J. Biomed. Health Inform. 19, 1627–1636.

Chen, H., Wu, L., Dou, Q., Qin, J., Li, S., Cheng, J.-Z., Ni, D., Heng, P.-A., 2017. Ultrasound standard plane detection using a composite neural network framework. IEEE Trans. Cybern. 47, 1576–1586.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2014. Decaf: a deep convolutional activation feature for generic visual recognition. In: International Conference on Machine Learning, pp. 647–655.

Dudley, M.N.J., Chapman, E., 2002. The importance of quality management in fetal measurement. Ultrasound Obstet. Gynecol. 19, 190–196.

Gao, Y., Maraci, M.A., Noble, J.A., 2016. Describing ultrasound video content using deep convolutional neural networks. In: IEEE International Symposium on Biomedical Imaging, pp. 787–790.

Girshick, R., 2015. Fast R-CNN. In: IEEE International Conference on Computer Vision, pp. 1440–1448.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904.

He, K., Zhang, X., Ren, S., Sun, J. , 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hosang, J., Benenson, R., Schiele, B., 2017. Learning non-maximum suppression. IEEE Conference on Computer Vision and Pattern Recognition.

Hu, X., Xu, X., Xiao, Y., Chen, H., He, S., Qin, J., Heng, P.-A., 2018. SINet: a scale-insensitive convolutional neural network for fast vehicle detection. IEEE Trans. Intell. Transp. Syst. 20, 1010–1019.

Huang, R., Xie, W., Noble, J.A., 2018. VP-Nets: efficient automatic localization of key brain structures in 3D fetal neurosonography. Med. Image Anal. 47, 127–139.

International Society of Ultrasound in Obstetrics and Gynecology Education Committee, 2010. Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the 'fetal neurosonogram'. Ultrasound Obstet. Gynecol. 29, 109–116.

Jiang, B., Luo, R., Mao, J., Xiao, T., Jiang, Y., 2018. Acquisition of localization confidence for accurate object detection. In: Proceedings of the European Conference on Computer Vision, pp. 784–799.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, pp. 1097–1105.

Li, J., Wang, Y., Lei, B., Cheng, J.Z., Qin, J., Wang, T., Li, S., Ni, D., 2017. Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting. IEEE J. Biomed. Health Inform. 22, 215–223.

Li, Y., Khanal, B., Hou, B., Alansary, A., Cerrolaza, J.J., Sinclair, M., Matthew, J., Gupta, C., Knight, C., Kainz, B., 2018. Standard plane detection in 3d fetal ultrasound using an iterative transformation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 392–400.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. SSD: single shot multibox detector. In: European Conference on Computer Vision, pp. 21–37.

Ma, J., Wu, F., Jiang, T.a., Zhao, Q., Kong, D., 2017. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. Int. J. Comput. Assist. Radiol. Surg. 11, 1–16.

Maaten, L.v.d., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Meng, Q., Sinclair, M., Zimmer, V., Hou, B., Rajchl, M., Toussaint, N., Gomez, A., Housden, J., Matthew, J., Rueckert, D., (2018). Weakly supervised estimation of shadow confidence maps in ultrasound imaging. arXiv:1811.08164.

Namburete, A., Xie, W., Yaqub, M., Zisserman, A., Noble, J.A., 2018. Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning. Med. Image Anal. 46, 1–14.

Ni, D., Yang, X., Chen, X., Chin, C.-T., Chen, S., Heng, P.A., Li, S., Qin, J., Wang, T., 2014. Standard plane localization in ultrasound by radial component model and selective search. Ultrasound Med. Biol. 40, 2728–2742.

Noble, J.A., 2016. Reflections on ultrasound image analysis. Med. Image Anal. 33, 33–37.

Paladini, D., Malinger, G., Monteagudo, A., Pilu, G., Timor-Tritsch, I., Toi, A., 2007. Sonographic examination of the fetal central nervous system: guidelines for performing the'basic examination'and the'fetal neurosonogram'. Ultrasound Obstet. Gynecol. 29, 109–116.

Pilu, G., Segata, M., Ghi, T., Carletti, A., Perolo, A., Santini, D., Bonasoni, P., Tani, G., Rizzo, N., 2006. Diagnosis of midline anomalies of the fetal brain with the three-dimensional median view. Ultrasound Obstet. Gynecol. 27, 522–529.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788.

Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger. IEEE Conference on Computer Vision and Pattern Recognition.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks. In: International Conference on Neural Information Processing Systems, pp. 91–99.

Salomon, L.J., Bernard, J.P., Duyme, M., Doris, B., Mas, N., Ville, Y., 2005. Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester. Ultrasound Obstet. Gynecol. 27, 34–40.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad–CAM: visual explanations from deep networks via gradient-based localization. In: IEEE International Conference on Computer Vision, pp. 618–626.

Shi, J., Xue, Z., Dai, Y., Peng, B., Dong, Y., Zhang, Q., Zhang, Y., 2018. Cascaded multi-column RVFL+ classifier for single-modal neuroimaging-based diagnosis of Parkinson's disease. IEEE Trans. Biomed. Eng. 66, 2362–2371.

Shi, J., Zhou, S., Liu, X., Zhang, Q., Lu, M., Wang, T., 2016. Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. Neurocomputing 194, 87–94.

Shin, S.Y., Lee, S., Yun, I.D., Kim, S.M., Lee, K.M., 2018. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. IEEE Trans. Med. Imaging 38, 762–774.

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

Sinclair, M., Baumgartner, C.F., Matthew, J., Bai, W., Martinez, J.C., Li, Y., Smith, S., Knight, C.L., Kainz, B., Hajnal, J., (2018). Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. arXiv:1804.09102.

Sinno Jialin, P., Tsang, I.W., Kwok, J.T., Qiang, Y., 2011. Domain adaptation via transfer component analysis. IEEE Trans. Neural Netw. 22, 199–210.

Sundaresan, V., Bridge, C.P., Ioannou, C., Noble, J.A., 2017. Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks. In: IEEE International Symposium on Biomedical Imaging, pp. 671–674.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2017. Convolutional neural networks for medical image analysis: full training or fine tuning. IEEE Trans. Med. Imaging 35, 1299–1312.

Wu, L., Cheng, J.Z., Li, S., Lei, B., Wang, T., Ni, D., 2017. FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. IEEE Trans. Cybern. 47, 1336–1349.

Xu, Z., Huo, Y., Park, J., Landman, B., Milkowski, A., Grbic, S., Zhou, S., 2018. Less is more: simultaneous view classification and landmark detection for abdominal ultrasound images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 711–719.

Xue, W., Brahm, G., Pandey, S., Leung, S., Li, S., 2018. Full left ventricle quantification via deep multitask relationships learning. Med. Image Anal. 43, 54–65.

Yang, X., Yu, L., Li, S., Wen, H., Luo, D., Bian, C., Qin, J., Ni, D., Heng, P.-A., 2019. Towards automated semantic segmentation in prenatal volumetric ultrasound. IEEE Trans. Med. Imaging 38, 180–193.

Yaqub, M., Rueda, S., Kopuri, A., Melo, P., Papageorghiou, A., Sullivan, P.B., McCormick, K., Noble, J.A., 2016. Plane localization in 3-D fetal neurosonography for longitudinal analysis of the developing brain. IEEE J. Biomed. Health Inform. 20, 1120–1128.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, pp. 3320–3328.

Zhang, L., Dudley, N.J., Lambrou, T., Allinson, N., Ye, X., 2017. Automatic image quality assessment and measurement of fetal head in two-dimensional ultrasound image. J. Med. Imaging 4, 024001.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929.