Fetal heart ultrasound image-oriented adaptive classification deep model based on differentiable architecture search

Xianhua Zeng (zengxh@cqupt.edu.cn)
Chongqing University of Posts and Telecommunications
Yunjiu Zhang
Chongqing University of Posts and Telecommunications
Wei Huang
Chongqing University of Posts and Telecommunications

Research Article

Keywords: fetal, classification, ultrasound, model, heart, task

Posted Date: August 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-847933/v1

Fetal heart ultrasound image-oriented adaptive classification deep model based on differentiable architecture search

Xianhua Zeng[1,2,*], Yunjiu Zhang[1,2],Wei Huang[1,2]

[1]School of Computer Science and Technology,Chongqing University of Posts and Telecommunications, Chongqing, China.

[2]Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing, China.

[*]zengxh@cqupt.edu.cn

ABSTRACT

Prenatal ultrasound examination is used for screening congenital heart defects and fetal genetic diseases. Unfavorable factors such as low signal-to-noise ratio, artifact and poor fetal posture in ultrasound images make it a very complicated task to identify and interpret the standard scan plane of the fetal heart in prenatal ultrasound examinations. Deep learning related methods are widely used to process and analyze medical images. However, designing an effective network structure for a specific task is a time-consuming and relies on expert knowledge. In order to obtain an effective fetal ultrasound image classification model in a short time, this paper collects and organizes the Fetal Heart Standard Plane(FHSP) level III screening dataset, and we use the Differentiable Architecture Search(DARTS) method for FHSP classification task to automatically obtain an efficient adaptive classification deep model called Ultrasound Image Adaptive Classification model(UIAC) for assisting the diagnosis of fetal congenital heart disease. This new model is a deep neural network consisting of two automatically searched optimal blocks. Our UIAC model has fewer parameters than the mainstream manned classification networks. Moreover, it has achieved the best recognition results on the FHSP classification task: top1-accuracy 89.84%, macro-f1 89.72%, kappa score 88.82%.

Introduction

Congenital Heart Disease (CHD) is one of the most common congenital malformations in newborns**Error! Reference source not found.**. There are many types of CHD, which can easily lead

to miscarriage, stillbirth, and neonatal death. It is one of the important causes of infant deaths, which seriously affects the quality of newborn population[2]. Ultrasound is a routine screening tool offered to all pregnant women because of its safety, relatively low cost and real-time manner[3]. Prenatal ultrasound examination is used for screening congenital heart defects and fetal genetic diseases. It can not only reduce the birth rate of defective newborns caused by congenital factors to a certain extent, but also has a guiding significance for the postpartum treatment of defective fetuses, which is conducive to improving the quality of the population**Error! Reference source not found.**.

Prenatal ultrasound examination generally includes image scanning, standard plane search, structure observation, parameter measurement and diagnosis. Determining of the standard plane is the prerequisite for structural observation, parameter measurement and final diagnosis, and it is an important part of prenatal diagnosis**Error! Reference source not found.**. Factors such as low signal-to-noise ratio of ultrasound images, image artifacts and poor fetal posture make it a very complicated task to identify and interpret the standard scanning plane of the fetus in prenatal ultrasound examinations**Error! Reference source not found.**. Ultrasound imaging is affected by noise and shadows, which will lead to poor imaging results and affect recognition accuracy. Recognition of related structures in ultrasound images is also difficult for inexperienced doctors and non-professionals. Therefore, constructing a FHSP classification model for auxiliary diagnosis can not only improve the efficiency of prenatal ultrasound examinations, but also reduce the burden on doctors. It is of great significance and a very challenging task[3].

In recent years, deep learning research has made amazing achievements in many fields (such as image classification, image segmentation, image restoration, object detection, natural language processing, etc). Deep learning related methods are also widely used to process and analyze medical images for auxiliary diagnosis and other related tasks. It is well applied in the diagnosis of skin, fundus, lung, breast, thyroid, liver and other organ diseases**Error! Reference source not found.**-**Error! Reference source not found.**. In terms of fetal standard plane recognition, Chen et al[3]. used a transferable deep neural network to identify fetal abdominal standard plane (FASP) in ultrasound video. Zhen et al[13]. proposed an automatic recognition of fetal facial standard plane

(FFSP) through a deep convolution neural network structure. Kong et al**Error! Reference source not found.**. constructed a multi-scale dense networks called MSDNet to automatically detect several standard planes from ultrasound scanned images or videos. Baumgartneret et al**Error! Reference source not found.**. proposed a method based on convolution neural network, which can automatically detect 13 fetal standard planes in hand-drawn 2D ultrasound data, and positioning of fetal structure through bounding box. Cai et al**Error! Reference source not found.**. proposed a new multi-task convolution neural network that learns to use the gaze tracking data of the ultrasound machine on the input ultrasound video frames to generate clinically relevant visual attention maps to assist with standardized abdominal circumference (AC) Plane detection. With the rapid development of deep learning technology, a large number of computer vision problems have been significantly improved in performance. Because deep learning methods can learn more abstract and complex representations directly from the original data. Therefore, deep learning methods have stronger applicability and better performance than traditional machine learning methods in complex image recognition tasks**Error! Reference source not found.**. However, the existing image classification models are all artificially designed network structures. The network architecture hyperparameters are complex, discrete, and disordered, which need to be adjusted from multiple dimensions such as depth, width, and jumper for specific tasks. Furthermore, designing an effective network structure for a specific task requires a lot of experimentation, which is time-consuming and relies on expert knowledge[16]. Neural Network Structure Search (NAS) is a very competitive method to solve the above problems[16,17]. It is an automated method that acquires a neural network model for a specific dataset. The obtained model by this method has achieved good performance in image classification, object detection and video understanding. In order to obtain an effective fetal heart ultrasound image classification model in a short time, this paper collects and organizes the FHSP level III screening dataset, and we automatically obtain a fetal heart ultrasound image-oriented adaptive classification model called Ultrasound Image Adaptive Classification model. This model is used to aid the diagnosis of fetal congenital heart disease.

Our main contributions in this work are as follows: 1)According to the specification of fetal heart ultrasound classification, the FHSP level III screening dataset was constructed, with a total of

eleven standard planes. 2)Aiming at the task of fetal heart ultrasound image classification , we use the DARTS method to automatically obtain fetal heart ultrasound image-oriented two optimal block. 3)We obtain a new classification network model (UIAC) consisting of two automatically searched optimal blocks to realize the standard plane classification of fetal heart ultrasound. 4)Experiments on the FHSP dataset show that the proposed UIAC model has the least amount of parameters and achieves the best recognition results than mainstream manned classification networks.

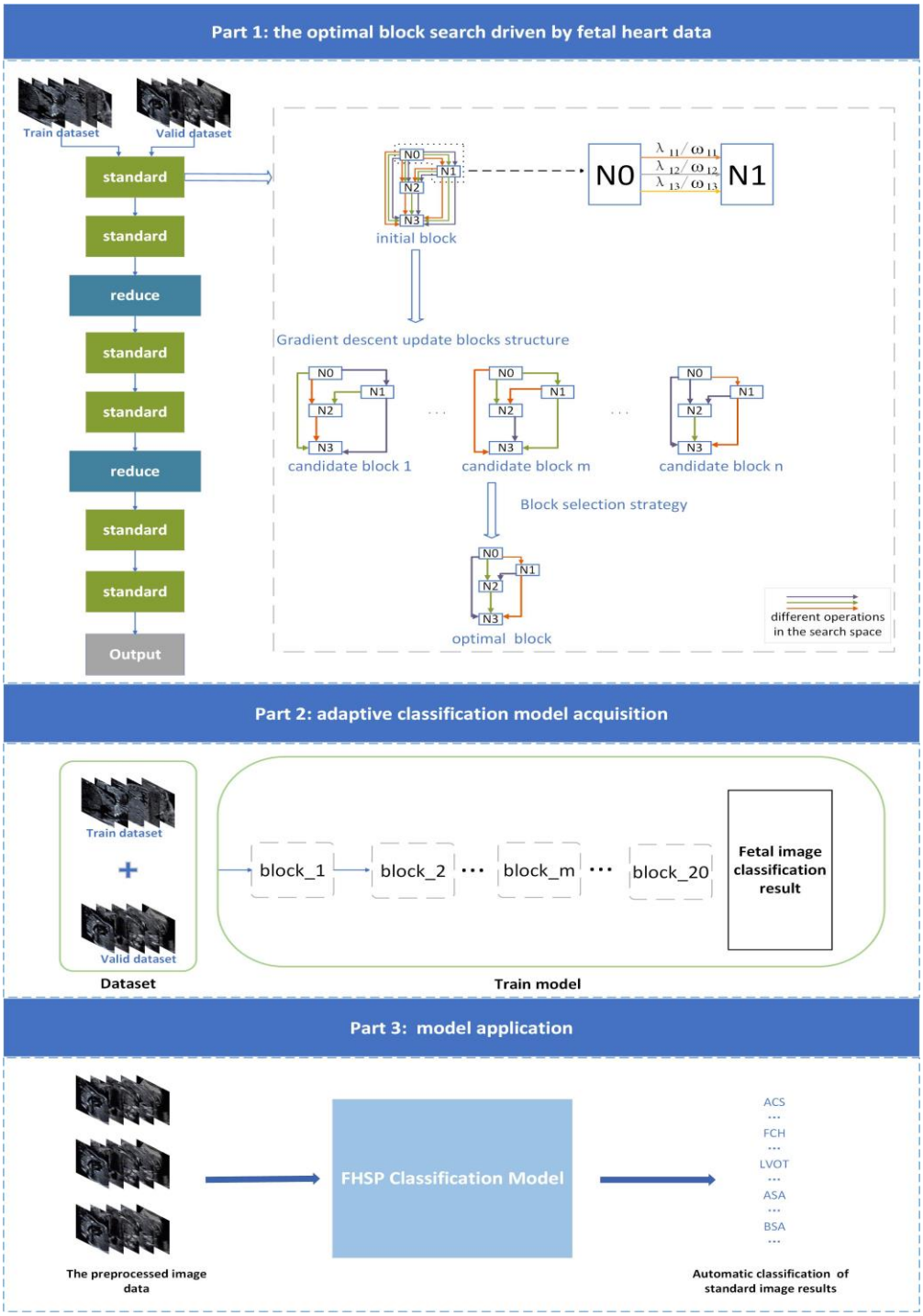


Figure 1. The FHSP adaptive classification model framework.

Method

The FHSP adaptive classification model framework is shown in Figure 1. It mainly includes three steps: the optimal block search driven by fetal heart data, adaptive classification model acquisition and model application. In this section, we introduce in detail the FHSP adaptive

classification model framework.

The optimal block search

First, we use the DARTS method to automatically generate an adaptive FHSP optimal block. It can be seen from the part 1 of Figure 1 that the network is composed of two different blocks. The standard block keeps the size of feature map unchanged. The reduction block halves the size of feature map. There is no difference in the overall structural design of the two optimal blocks. These two blocks are directed acyclic graphs composed of N ordered nodes and edges between nodes. The node represent the feature map T and each node with a lower number is connected to a node with a higher number. For example, $N0, N1, N2$ will be connected to $N3$. The edge E_{ij} between nodes represents the operation of the feature map T_i to T_j. In the part 1 of Figure 1, directional arrows with different colors between N0 and N1 indicate different operations o^m in the search space O. Each operation has two parameters: the network weight parameter ω^m and the network structure parameter λ^m that is the weight of each operation. The conversion of feature maps between nodes can be expressed as a mixed operation:

$$\bar{E}_{ij} = \sum_{o^m \in O} \lambda^m \ o^m\left(T_i\right) \tag{1}$$

where $o^m\left(T_i\right)$ means that the m-th operation in the search space is applied to the feature map of node i. And the feature map of node can be expressed as:

$$\begin{aligned} T_j &= \sum_{i<j} \bar{E}_{ij} T_i \\ &= \sum_{i<j} \sum_{o^m \in O} \lambda_{ij}^m \ o_{ij}^m\left(T_i\right) \end{aligned} \tag{2}$$

where λ_{ij}^m represents the weight of the m-th operation between node i and node j. $o_{ij}^m\left(T_i\right)$ means the m-th operation between node i and node j. In order to make the discrete search space continuous, all operation weights $\lambda^m, \lambda^m \in \lambda$ are relaxed using the softmax function. So that the network structure parameter λ can be optimized. For the FHSP classification task, to obtain the adaptive classification model, we must obtain the optimal blocks corresponding to the optimal network structure parameters λ. Therefore, we need to find the best network structure

parameters λ . It is can be expressed as a matrix:

$$\lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1M} \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ \lambda_{N1} & \mathrm{L} & \lambda_{NM} \end{pmatrix}_{N\times M} \tag{3}$$

where M is the number of operations in the search space, N is the number of edges between nodes. We need to continuously optimize the network structure parameters λ . The data in the training set is divided into training set D_{train} and validation set D_{val} . We define training loss L_{train} and verification loss L_{val} . First, the D_{train} is used to calculate the training loss L_{train} updating the network weights ω^* by gradient descent. Then, the D_{val} is used to calculate the verification loss L_{val} after the network weight is updated, and the network structure parameters λ are optimized by gradient descent. The above two-step optimization can be expressed as:

$$\begin{aligned} &\lambda^* = \arg\min_{\lambda} L_{val}\left(\omega^*(\lambda), \lambda\right) \\ &s.t.\ \omega^*(\lambda) = \arg\min_{\omega} L_{train}(\omega, \lambda) \end{aligned} \tag{4}$$

where $\omega^*(\lambda)$ is the optimal network weight when the structure parameter is λ . If the network weight ω^* is adjusted to the optimal for each network structure parameter λ , it will consume a lot of computing resources and time. Therefore, the network structure parameters λ are optimized by the approximate optimal network weights $\omega^*(\lambda)$ optimized ω one step . The derivation process can be seen in DARTS[18]. After the above two-step optimization, the optimal network architecture parameters λ^* are obtained. Using the maximum weight operation in λ^* instead of the mixing operation to obtain the optimal block B based on FHSP classification. The gray dotted box of the part 1 shows the search process of the block. The top is the initial block with mixed operation that the operations between each two nodes are all operations in the search space. The operation between nodes has two parameters: the network structure weight λ updating with validation set and the network weight ω updating with training set. The middle are candidate blocks that obtained according to the optimized network structure parameters during the search process. In each candidate block, the operation with the maximum network

structure weight is reserved among the nodes. Below is the best optimal block B for the FHSP classification task. Algortithm 1 gives the process of optimal block search by fetal heart data driven.

Algorithm 1 the process of optimal block search by fetal heart data driven.

Input: The FHSP data $D_{train} = \{(X_1, Y_1), ...(X_N, Y_N)\}, D_{val} = \{(X_1^{'}, Y_1^{'}), ...(X_M^{'}, Y_M^{'})\}$, where $N = M$, number of search network blocks $sear\_number$, the learning rate lr_w, lr_λ, the number of iterations $epoch$, the initial block ini_block.

Output: The optimal blocks for FHSP classification tasks B.

Process:

0: Stack $sear\_number$ initial blocks to obtain the block search network.

1: Initialization the network structure parameters λ with $zero$.

2: Use softmax function to make the network structure parameters λ continuous.

3: Random initialize network weights ω

4: for iter $\rightarrow$ epoch do

5: Input D_{train} into the block search network to get the predicted classification label l_p.

6: Calculate the cross-entropy loss L_{train} between the predicted classification label l_p and the real label l_t.

7: $\frac{\partial L_{train}}{\partial \omega} = \nabla_\omega L_{train}(\omega, \lambda)$ calculate training gradient.

8: $\omega \leftarrow \omega - lr_{train} \cdot \frac{\partial L_{train}}{\partial \omega}$ update network weight.

9: Input D_{val} into the block search network to get the predicted classification label $l_p^{'}$.

10: Calculate the cross-entropy loss L_{val} between the predicted classification label $l_p^{'}$ and the real label l_t.

11: $\frac{\partial L_{val}}{\partial \lambda} = \nabla_\lambda L_{val}(\omega^*(\lambda), \lambda)$ calculate verification gradient.

12: $\lambda \leftarrow \lambda - lr_{val} \cdot \frac{\partial L_{val}}{\partial \lambda}$ update network structure parameter.

13: $end\ for$

14: Obtain the optimal block architecture parameters matrix λ^*.

15：Use the maximum weight operation in λ^* instead of the mixing operation to obtain the optimal block B.

16：Output the optimal block B.

The initial block is shown in Figure Figure 2 It includes two input $Node1$ and $Node2$ which are the outputs of the first two blocks, an output $Node7$ which is the concat output of all intermediate nodes, the remaining $Node3, Node4, Node5$ and $Node6$ are intermediate nodes, which are the sum of all pre-order feature maps after operations. The search space O in this paper is consistent with DARTS[18], including zero operation, max pooling, average pooling, 3×3 and 5×5 separable convolution, 3×3 and 5×5 dilated separable convolution, skip connection 8 candidate operations. The conversion of the feature map between the intermediate nodes of the initial block is a mixed operation represented by formula **Error! Reference source not found.**. Among them, the operations in the mixed operation are 8 operations of the search space. The network is composed of a chain structure of 8 blocks.

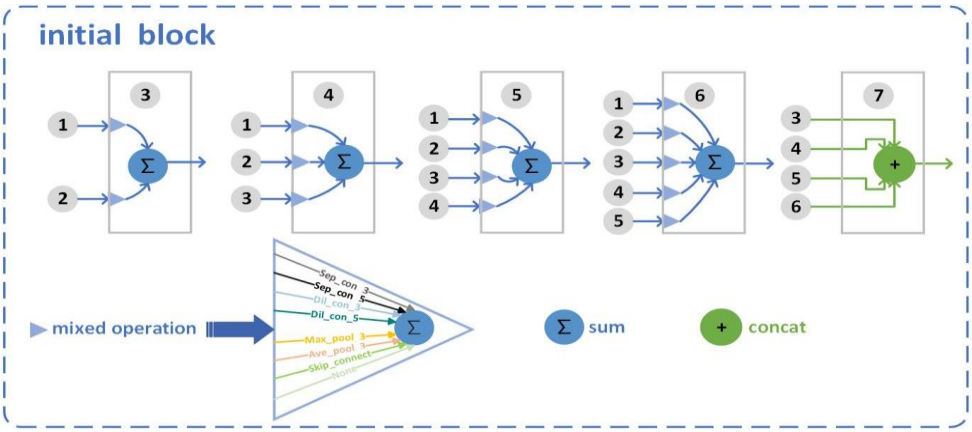


Figure 2. The initial block.

In the optimal block search stage, 8800 FHSP images in the training set are randomly divided into two equal data blocks D_{train}, D_{val}. The network is composed of a chain structure of 8 blocks. The numbers 2,5 of blocks are reduction blocks, and the other blocks are normal blocks. Network weights ω are randomly initialized. D_{train} used to optimize the network weights. The initial learning rate $lr_w = 0.0125$. The network structure parameters λ are initialized with all zeros and then relaxed using the softmax function. D_{val} used to optimize network structure parameters. The initial learning rate $lr_\lambda = 10^{-3}$. Alternating iterative training $50\ epoch$. According to the

accuracy of the validation set during the training process, the two optimal blocks searched based on FHSP data are screened out as shown in Figure 3Figure 4 .

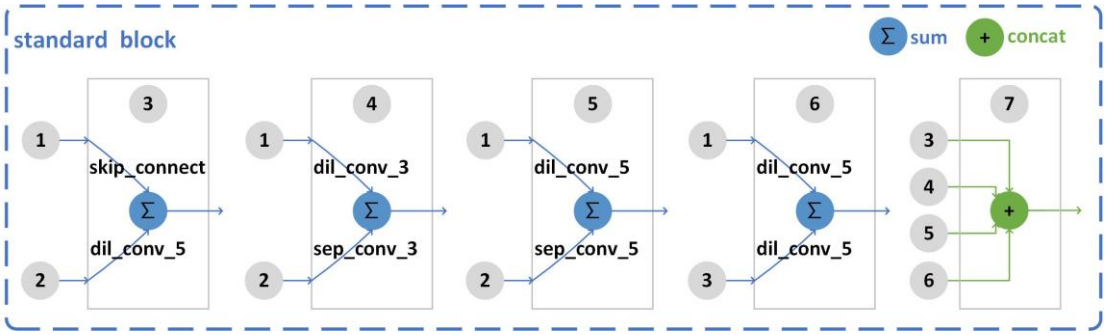


Figure 3. Standard block.

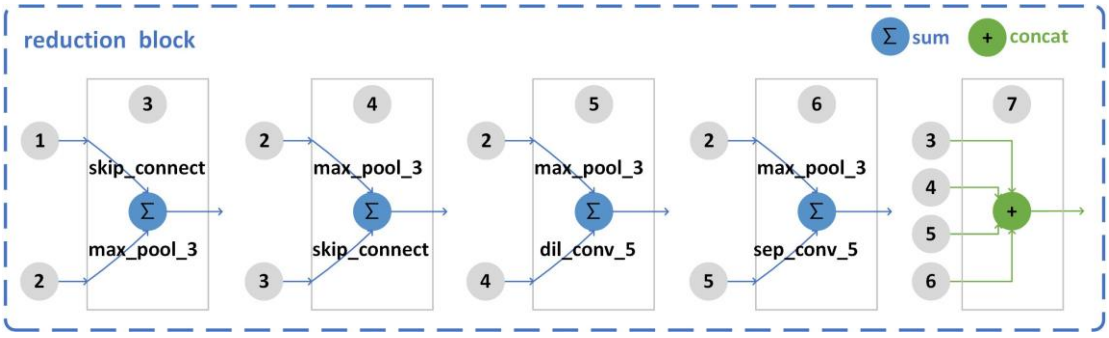


Figure 4. Reduction block.

For each intermediate node, the two operations with the maximum weights in the pre-order mixing operation are used instead of the mixing operation. In Figure 3, the two most weighted operations of the $Node3$ are the skip connection with $Node1$ and the 5×5 dilated separable convolution operation with $Node2$. For $Node4$, the two most weighted operations are the 3×3 dilated separable convolution operation with $Node1$ and the 3×3 separable convolution operation with $Node2$. In $Node5$, the two operations are the 5×5 dilated separable convolution operation with $Node1$ and the 5×5 separable convolution operation with $Node2$. For $Node6$, the two operations are the 5×5 dilated separable convolution operation with $Node1$ and the 5×5 dilated separable convolution operation with $Node3$.Each intermediate node sums the corresponding elements of the calculated feature map. Then the $Node7$ connects the output features of all intermediate nodes as output. The reduction block in Figure 4 is processed in the same way as the standard block in Figure 3, except that the operation between each node is different.

Adaptive classification model acquisition

Next, we obtain an adaptive classification deep model called Ultrasound Image Adaptive Classification model consists of two automatically searched optimal blocks. The entire network is composed of 20 blocks. The numbers 5, 10, 15 of blocks are reduction blocks, and the other blocks are standard blocks. In the training phase, an auxiliary classifier is used to train the network at the number 15 of blocks. Finally, the classification results are output through a global average pooling layer and a fully connected layer. The UIAC model is a deep neural network shown in Figure 5.

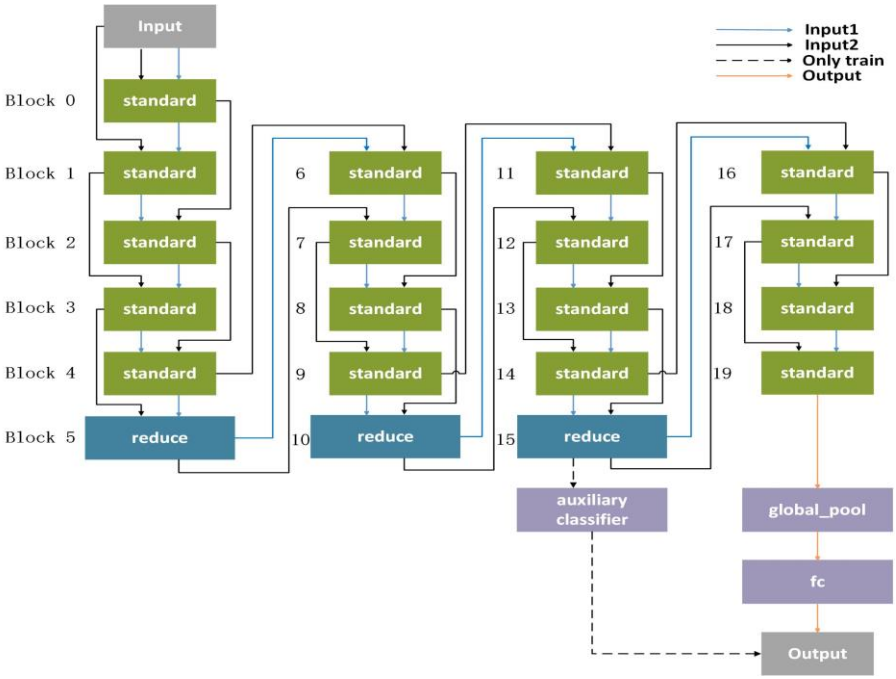


Figure 5. The model of UIAC.

Training the network model from scratch to obtain a new effective classification model for FHSP that can be used to the diagnosis of fetal congenital heart disease. The part 2 of Figure 1 shows Adaptive classification model acquisition. The new UIAC model architecture can be obtained by stacking the two best optimal blocks obtained in the search process. Then the train set of FHSP is used to train the network weights. The early stopping strategy is introduced in the process of training the network. After a certain round of network training, whenever the training loss dose not improve, the network is considered to be converged. Then the network training is stopped.Algorithm 2 gives the training process of UIAC model.

Algorithm2 the process of construction and training fetal heart classification model

Input：The FHSP data $D=\{(X_1,Y_1),.....(X_N,Y_N)\}$, the optimal building block B, number of classification network blocks $train\_number$, the learning rate lr, the early stop starting point $STEP$, number of steps waiting for early stop P.

Output：Trained UIAC model M.

Process：

0: Construct a new classification network model M according to the number of classification network layers c_layers and the optimal building block discrete structure B.

1：Stack $train\_number$ optimal blocks obtained B from Algorithm 1 to obtain an untrained UIAC model.

2：Random initialize network weights ω.

3：$Step\_counter+=1$.

4：$whilr\ True\ do$.

5：$Step\_counter+=1$.

6：Input the ultrasound data D into the classification model to get the predicted classification label l_p.

7：Calculate the cross-entropy loss L between the predicted classification label l_p and the real label l_t.

8：*if* Step_counter>STEP then

9：*if* L dose not improve

10：$P--$.

11：*end* if

12：*if* P==0 then

13：*break*.

14：*end* if

15：*end* if

16: $\frac{\partial L}{\partial \omega}=\nabla_{\omega}L(\omega)$ calculate the gradient.

17：$\omega=\omega-lr\cdot\frac{\partial L}{\partial \omega}$ update network weight.

18：*end while*

19：Trained UIAC model M.

Trained FHSP adaptive classification model can be used to identify the pictures or video files. Due

to the early stopping strategy added in the training model stage, the model of training epochs are not uniform.

Model application

Finally, the part 3 in Figure 1 shows the model application. The trained UIAC model in the second part can be used to identify the pictures or video files generated by the clinical ultrasound workstation. Then input the preprocessed image data into the UIAC model to get the recognition result. It can effectively help inexperienced doctors in clinical diagnosis or help non-professionals to identify FHSP.

Results

FHSP dataset: The original FHSP data was provided by a professional obstetrics and gynecology hospital in Chongqing, which provided the ultrasound images of the fetal heart after desensitization during the prenatal examination of pregnant women in the hospital from 2019 to 2020. For each pregnant woman, we collected 40 common plane image data (including original plane image, color Doppler, etc.). The original dataset obtained from the hospital included 1328 prenatal examination pictures of pregnant women. According to the fetal heart ultrasound grading examination specifications, eleven standard planes abdominal cross section(ACS), four chamber heart(FCH), left ventricular outflow tract(LVOT), right ventricular outflow tract(RVOT), three vessels(TV), three vessel trachea(TVT), aortic arch long axis(AALA), arterial duct arch long axis(ADALA), vena cava long axis(VCLA), aorta short axis(ASA) and biventricular short axis(BSC) were used to construct the fetal heart ultrasound level III screening dataset. Each standard plane selected 1,000 ultrasound images, which contained a total of 11,000 ultrasound images. The fetal heart ultrasound dataset sample of is shown in the Figure 6. During the experiment, we use a random division method, using 80% of the data as training data and 20% of the data as testing data.

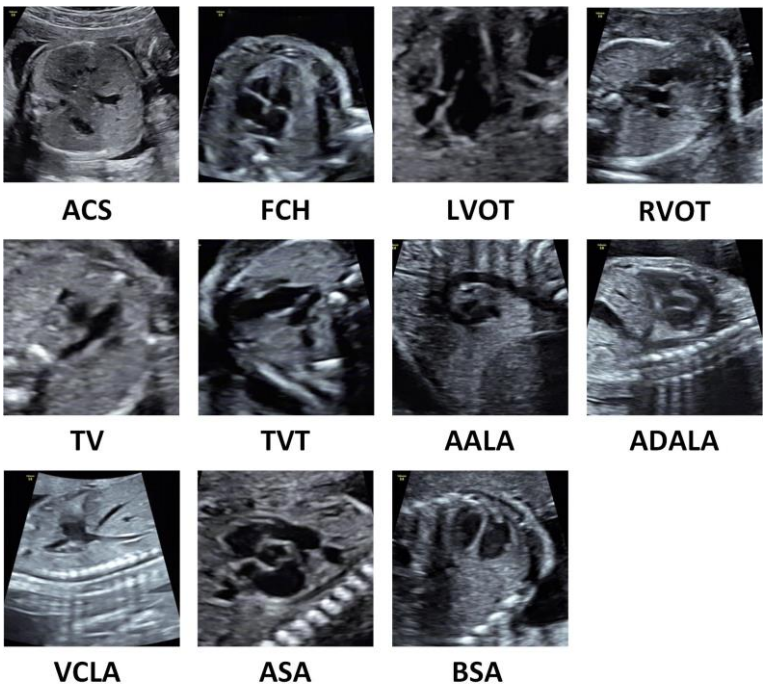


Figure 6. The dataset sample.

We evaluated the UIAC model and the mainstream manually designed classification models on the FHSP dataset. Model evaluation uses common indicators in multiple classifications: top1 accuracy, macro-f1 score, kappa score. In order to evaluate the performance of different classic classification models on FHSP image classification tasks. Training from scratch the Vgg series, Resnet series, Densenet series models, taking the highest top1 accuracy value and the macro-f1 score, the kappa score of the same epoch of the test set for evaluation. All experiments were completed on the Inspur server platform using a Tesla V100-SXM2-32GB graphics card.

In model training phase, the training set contains 8800 FHSP images, and the test set contains 2200 FHSP images. Image size 64 * 64, batch_size = 96. The auxiliary classifier weight is 0.25. The initial learning rate 0.02. As the number of training rounds increases, using the CosineAnnealingLR method to adjust the learning rate. Momentum 0.9. Weight decay 3×10^{-4}. The cutout is used to method to enhance the training set data, the cutout length is 6. The initial number of channels is 32 to ensure our model size (around 6.8M) is comparable with other mainstream manually designed classification models. The training early stop strategy is added in the training phase. After the network is trained for 500 epochs, the network training loss is judged in each subsequent round. If the network training loss does not decrease for 50

consecutive rounds, the model is considered to be converged and training is stopped. Therefore, each model training epochs will be different.

Table 1 shows the comparison of the evaluation indicators of the FHSP image classification results of the Vgg[19] series model. As can be seen from the table, Vgg13-bn gets the highest score on accuracy, macro-f1 and kappa score. For the FHSP classification tasks, Vgg13-bn has achieved best classification results in the Vgg series model.

Model	Accuracy(%)	Macro-f1(%)	Kappa score(%)	Parameter(M)	Training rounds	Model obtain method
Vgg11	86.6±0.13	86.37±0.15	85.27±0.15	128.8	561	manual
Vgg11-bn	87.85±0.08	87.64±0.08	86.63±0.09	128.8	613	manual
Vgg13	87.81±0.16	87.72±0.17	86.59±0.17	128.9	556	manual
Vgg13-bn	**89.02±0.19**	**88.91±0.18**	**87.92±0.21**	129	677	manual
Vgg16	87.05±0.03	86.83±0.04	86.83±0.04	134.3	559	manual
Vgg16-bn	88.44±0.62	88.39±0.6	87.29±0.68	134.3	680	manual
Vgg19	87.07±0.15	86.96±0.17	85.78±0.16	128.9	553	manual
Vgg19-bn	88.75±0.46	88.65±0.43	87.62±0.51	139.6	609	manual

Table 1. Vgg series models evaluation results

Table 2 presents the comparison of the evaluation indicators on the FHSP image classification results of the Resnet[20] series models. Where Resnet18 using the least amount of parameters achieved the highest scores on all three evaluation indicators: 83.59±0.13%, 83.23±0.12%, 81.95±0.15 %.

Model	Accuracy(%)	Macro-f1(%)	Kappa score(%)	Parameter(M)	Training rounds	Model obtain method
Resnet18	**83.59±0.13**	**83.23±0.12**	**81.95±0.15**	11.1	637	manual
Resnet34	83.34±0.22	83.04±0.26	81.67±0.24	21.2	688	manual
Resnet50	81.39±0.21	81.39±0.21	79.53±0.24	23.5	700	manual

Resnet101	81.08±0.3	80.63±0.32	79.53±0.88	42.5	560	manual
Resnet152	80.18±0.39	79.58±0.39	78.2±0.43	58.1	650	manual

Table 2. Resnet series models evaluation results

Results in Table 3. Densenet series models evaluation results show that the recognition effect of Densenet121[21] is better than Densenet161, Densenet169, Densenet201 where accuracy 88.00±0.2%, macro-f1 87.75±0.19%, kappa score 86.8±0.22%. Its parameter amount is 6.9M less than Resnet18 and Vgg13-bn. Its performance is better than Resnet18, but lower than Vgg13-bn.

Model	Accuracy(%)	Macro-f1(%)	Kappa score(%)	Parameter(M)	Training rounds	Model obtain method
Densenet121	**88.00±0.2**	**87.75±0.19**	**86.8±0.22**	6.9	629	manual
Densenet161	87.23±0.15	87.04±0.16	85.95±0.17	26.4	597	manual
Densenet169	86.51±0.23	86.33±0.24	85.17±0.25	12.5	560	manual
Densenet201	85.95±0.18	85.72±0.2	84.55±0.2	18.1	622	manual

Table 3. Densenet series models evaluation results

Compare the best model in Table 1Table 2Table 3 and the Alexnet[23] model with the UIAC model designed by FHSP data-driven automatically. The detailed experimental results are shown in Table 4. We observed that the UIAC model achieves the best results on three evaluation indicators with the least amount of parameters. It is 0.82-10.75% higher than the classic classification network on accuracy, 0.81-10.73% higher than other models on macro-f1, and 0.9-11.82% higher than other methods on the kappa score. The results show that the UIAC model using fewest parameters can learn more effective FHSP features and can distinguish FHSP more accurately.

Model	Accuracy(%)	Macro-f1(%)	Kappa score(%)	Parameter(M)	Training rounds	Model obtain method
Alexnet	79.09±0.09	78.99±0.09	77±0.1	57	561	manual
Vgg13-bn	89.02±0.19	88.91±0.18	87.92±0.21	129	677	manual
Resnet18	83.59±0.13	83.23±0.12	81.95±0.15	11.1	637	manual
Densenet121	88.00±0.2	87.75±0.19	86.8±0.22	6.9	629	manual

UIAC	89.84±0.22	89.72±0.24	88.82±0.24	6.8	574	automatic

Table 4. Each model evaluation results

Discussion

Figure 7. Classification accuracy of each model. shows the classification accuracy of each plane of the FHSP image by each model in Table 4. The UIAC model achieves the highest accuracy in the recognition of six planes of the ACS, LVOT, TV, BSA, ASA,ADALA. It can be seen from the figure that the five models are not effective in classifying four planes of the RVOT, TV, TVT, ASA. However, the UIAC model has better overall recognition performance for these planes. In addition, the confusion matrix of UIAC model on test set is shown in Figure 8. From the confusion matrix, we can observe the misclassification is mainly concentrated in four planes of the RVOT, TV, TVT, ASA that lead to UIAC model poor classification effect. An important reason is that these planes are not very recognizable. The doctor judges that most of these planes are recognized by dynamic video. Another important reason is that some standard planes have a high degree of similarity. For example, the angle of the RVOT and the TVT is only changed a little when the ultrasound probe is scanned. This leads to similarities between different planes.

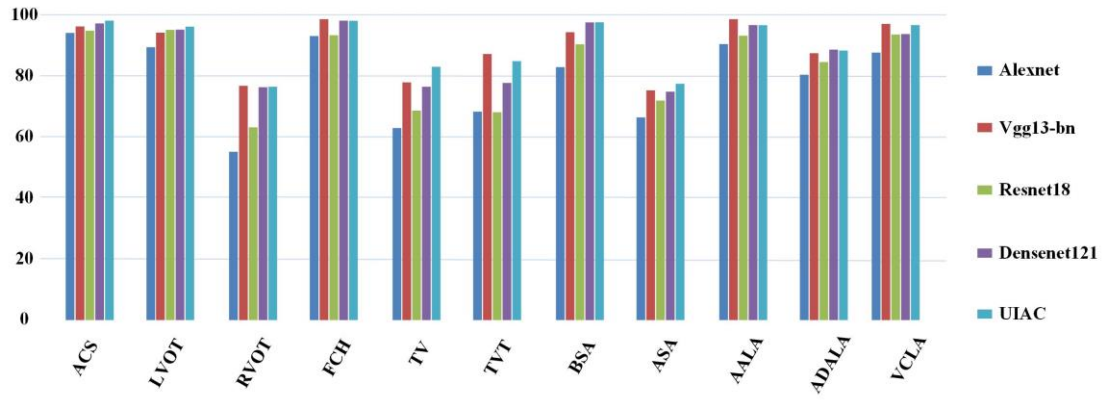


Figure 7. Classification accuracy of each model.

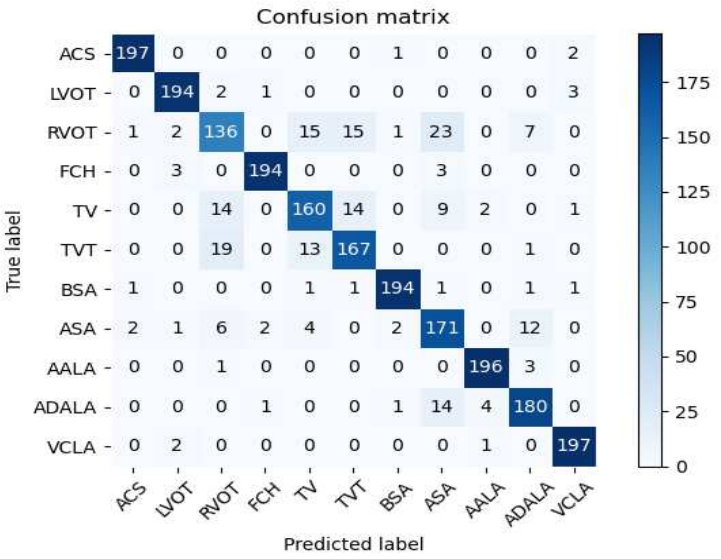


Figure 8. Confusion matrix for UIAC model.

In order to observe the effect of the UIAC model more intuitively, understanding which feature information of the image has made a greater contribution to the final classification result of the model. We use the Class Activation Map (CAM)[23] to visualize the specific feature regions of the image class that the trained model pays attention. The visualization results of 11 standard planes after classification by UIAC are shown in Figure 9. The uncolored picture on the left is the original image, and the colored image on the right is the CAM after the original image input model recognizes. The red is the area that the model pays more attention to when classifying the image. It can be seen from the figure that the color of the ultrasound image is single, and the image category cannot be judged by color information. The red area that the model pays attention to is more concentrated in the structural area of the heart, so the model's focus on the FHSP image is mainly the structure information. It shows that the UIAC model can effectively extract the structural features of the image to classification the image.

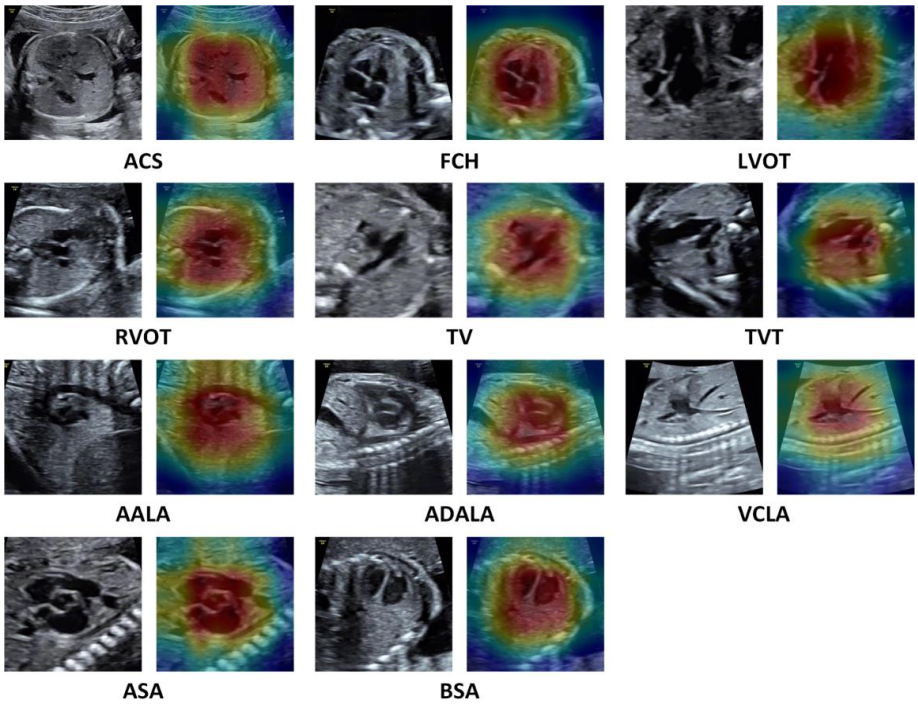
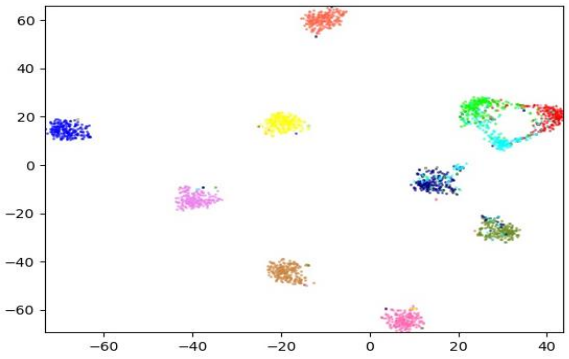
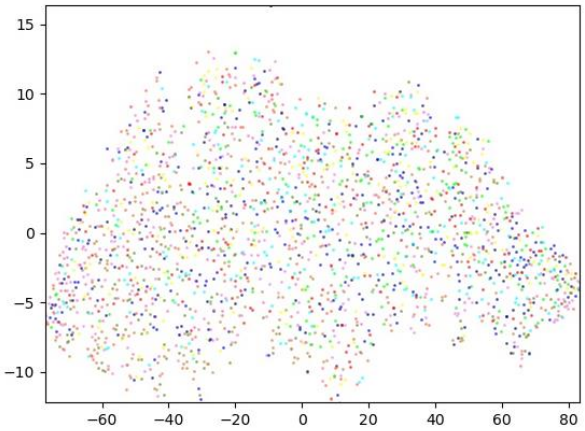


Figure 9. CAM of the UIAC mode.

For deep learning models, feature representation has a great influence on the recognition results**Error! Reference source not found.**. In order to more intuitively demonstrate feature representation of the UIAC model on the FHSP classification task.We use the t-SNE[24] method to visualize the network feature map of the test data extracted from the network. Extract the output feature maps of the last layer of the global average pooling layer of the untrained and trained UIAC models, and then use the t-SNE method to reduce the dimensionality of feature and visualize. The visualization results are shown in Figure 10. The Figure 10(a) is the visualization result of the untrained network extracting the network feature map, and the Figure 10(b) is the visualization result of the network extracting the network feature map after training. The image uses different colors to indicate the data from different labels. From the figure, it can be found that the network feature maps extracted by the untrained network are scattered. However, the network feature maps extracted by the network after training make the samples obviously separable.This shows that the new UIAC model is very effective for the FHSP feature classification.



(a) Untrained model features



(b) Trained model features

Figure 10. the t-SNE visualization results.

Conclusions

Prenatal ultrasound examination is of great significance to the screening of defective fetuses. The determination of the standard plane of fetal ultrasound is an important part of prenatal diagnosis. In this paper, a new classification model UIAC is obtained for the fetal ultrasound standard plane classification using differentiable architecture search method. Experiments show that the new UIAC model can effectively identify each fetal heart ultrasound standard plane with fewest parameters, improving the classification accuracy. The UIAC model can help inexperienced doctors and non-professionals to judge the FHSP. Classifying the FHSP is the first step in the identification of fetal CHD. Due to the lack of abnormal fetal data, we cannot yet identify fetal CHD. Therefore, future work may also be carried out from the following aspects: 1)Collect and organize more abnormal fetal data to identify fetal CHD. 2)Optimize the search phase of the microstructured search method to reduce memory consumption.

References

1. Tegnander, E., Williams, W., Johansen, O. J., Blaas, H. G., & Eik-Nes, S. H. Prenatal detection of heart defects in a non-selected population of 30 149 fetuses—detection rates and outcome.Ultrasound in Obstetrics & Gynecology, **27(3),** 252-265. https://doi.org/10.1002/uog.2710 (2006).

2. Rosamond, W. et al. American Heart Association statistics committee and stroke statistics subcommittee. Heart disease and stroke statistics-2007 update: a report from the American Heart Association statistics committee and stroke statistics subcommittee. Circulation, **115,** e69-e171.https://doi.org/10.1161/CIRCULATIONAHA.106.179918(2007).

3. Chen, H. et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE journal of biomedical and health informatics, **19(5),** 1627-1636.https://doi.org/10.1109/JBHI.2015.2425041(2015).

4. Reddy, P. P., & Mandell, J. Prenatal diagnosis: therapeutic implications. Urologic Clinics of North America, **25(2),** 171-180. https://doi.org/10.1016/S0094-0143(05)70005-7(1998).

5. Kong, P. et al. Automatic and efficient standard plane recognition in fetal ultrasound images via multi-scale dense networks. In Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis (pp. 160-168).https://doi.org/10.1007/978-3-030-00807-9_16(2018).

6. Baumgartner, C. F. et al. SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE transactions on medical imaging, **36(11),** 2204-2215. https://doi.org/10.1109/TMI.2017.2712367(2017).

7. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. nature, **542(7639),** 115-118.https://doi.org/10.1038/nature21056 (2017).

8. Li, L., Xu, M., Wang, X., Jiang, L., & Liu, H. Attention based glaucoma detection: A large-scale database and cnn model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10571-10580). https://doi.org/10.1109/CVPR.2019.01082(2019).

9. Harsono, I. W., Liawatimena, S., & Cenggoro, T. W. Lung nodule detection and classification from thorax ct-scan using retinanet with transfer learning. Journal of King Saud University-Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2020.03.013(2020).

10. Fujioka, T. et al. Classification of breast masses on ultrasound shear wave elastography using convolutional neural networks. Ultrasonic Imaging, **42(4-5),** 213-220. https://doi.org/10.1177/0161734620932609(2020).

11. Song, J. et al. Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules. Medicine, 98(15).https://doi.org/10.1097/MD.0000000000015133(2019).

12. Cao, W. et al. Application of Deep Learning in Quantitative Analysis of 2-Dimensional Ultrasound Imaging of Nonalcoholic Fatty Liver Disease. Journal of Ultrasound in Medicine, **39(1),** 51-59. https://doi.org/10.1002/jum.15070(2020).

13. Yu, Z. et al. A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. IEEE journal of biomedical and health informatics, **22(3),** 874-885.https://doi.org/10.1109/JBHI.2017.2705031(2017).

14. Cai, Y., Sharma, H., Chatelain, P., & Noble, J. A. Multi-task sonoeyenet: detection of fetal standardized planes assisted by generated sonographer attention maps. In International conference on medical image computing and computer-assisted intervention (pp. 871-879).https://doi.org/10.1007/978-3-030-00928-1_98(2018).

15. Litjens, G. et al. A survey on deep learning in medical image analysis. Medical image analysis, **42,** 60-88.https://doi.org/10.1016/j.media.2017.07.005(2017).

16. Zhou, Z. et al. Differentiable architecture search for aeroengine bevel gear fault diagnosis. In 2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD) (pp. 270-274). https://doi.org/10.1109/ICSMD50554.2020.9261641(2020).

17. Peng, C., Li, Y., Jiao, L., & Shang, R. Efficient convolutional neural architecture search for remote sensing image scene classification. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2020.3020424(2020).

18. Liu H, Simonyan K and Yang Y. Darts: Differentiable architecture search. Preprint at http://arxiv.org/abs/1806.09055 (2018).

19. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. Preprint at http://arxiv.org/abs/14091556(2014).

20. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

https://doi.org/10.1109/CVPR.2016.90(2016).

21. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).https://doi.org/10.1109/CVPR.2017.243(2017).

22. Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, **25,** 1097-1105. https://doi.org/10.1145/3065386(2012).

23. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).https://doi.org/10.1109/CVPR.2016.319(2016).

24. Van der Maaten, L., & Hinton, G. Visualizing data using t-SNE. Journal of machine learning research, 9(11). (2008).

Acknowledgements

This work was supported by the Key Project of Chongqing Natural Science Foundation, China (Grant No. cstc2019jcyjzdxmX0011) and the National Natural Science Foundation of China (No. 62076044).

Author contributions statement

X.Z., Y.Z. designed and realized the whole system, made the code, conducted experiments, and wrote this paper. W.H. helped with parts of experiments and gave comments during experiments. All authors reviewed the manuscript.

Additional information

Competing interests

The authors declare no competing interests.