Diagnosis of Fetal Congenital Heart Disease Based on Deep Learning

Fang Li[1], Rui Yang[1], Xueqin Ji[3], Wei Tang[2, 4], Xinrong Chen[2, 4]*

[1] School of Computer Science and Technology, Shanghai University of Electric Power, China
[2] Academy for Engineering and Technology, Fudan University, China
[3] Peking University First Hospital Ningxia Women and Children's Hospital
(Ningxia Hui Autonomous Region Maternal and Child Health Hospital), China
[4] Shanghai Key Laboratory of Medical Image Computing and Computer Assisted Intervention, China
fang_li@shiep.edu.cn, suepyangrui@163.com, wtang22@m.fudan.edu.cn, jinian54@163.com, chenxinrong@fudan.edu.cn

Abstract

Fetal congenital heart disease is the most common birth defect. In early diagnosis, the use of echocardiography is an important means of diagnosis, but due to the unique structure of the fetal heart, there are still many challenges in the early screening process. Hence, this study proposes a diagnosis model called ConvNeXt based on Attention Mechanism and Transfer Learning (ConvNeXt-AMTL) for congenital heart disease, which utilizes a large-kernel convolutional neural network to extract features from fetal echocardiography, and using attention mechanisms to focus and optimize key features. At the same time, in order to alleviate the problem that image data samples are too few to train the model well, this study proposes to use transfer learning to train the model. Numerous experiments have shown that the proposed model can efficiently diagnose fetal congenital heart disease, achieving an accuracy of 98.8% on the test set, effectively promoting prenatal screening of fetal congenital heart disease.

Keywords: Fetal congenital heart disease, Deep learning, Transfer learning, Convolutional neural networks, Attention mechanism

1 Introduction

Fetal congenital heart disease (CHD) is the most common type of congenital malformation, and in recent years, the mortality rate of fetal congenital heart disease ranks first among all congenital disabilities. In addition, fetal CHD is still the main cause of infant death, which has brought more and more burdens to many families and society. Early detection and diagnosis of diseases can not only help doctors to diagnose specific conditions in time, but also allow patients to obtain cure of diseases through interventional treatment of CHD in the early stages of the disease. However, in areas where medical resources are relatively scarce, especially in remote and backward areas such as high-altitude areas, due to the limitation of medical level, it is difficult for patients with CHD to obtain effective early detection and diagnosis.

In recent decades, the diagnosis of CHD by echocardiography has been one of the main means of early screening. Echocardiography is a non-invasive inspection technique that uses the special echo physical characteristics of ultrasound to receive and process echo signals, and then obtain the anatomical structure of the heart's internal cardiovascular system. Because of its relatively safe and non-invasive characteristics to the fetus, it is widely used in prenatal diagnosis and disease screening. And echocardiography can effectively evaluate the structure and function of the fetal heart.

Although rapid advances have been made in fetal ultrasound imaging, the prenatal detection rate of fetal CHD remains low based on clinical population studies, mainly due to the following challenges:

First of all, fetal ultrasound views usually have low resolution, more spots, and more artifacts, which bring great obstacles to the diagnosis of cardiologists. In addition, the experience of the physician probing the sonogram and the different positions of the fetus in the uterus may lead to inconsistency and non-reproducibility in obtaining echocardiograms, which poses a huge problem for cardiologists in diagnosing fetal CHD. Finally, cardiologists must be familiar with fetal cardiac anatomy when analyzing fetal echocardiograms to diagnose fetal CHD. However, due to the complex structure of the fetal heart, accurate identification of fetal CHD is a demanding task, which leads to a very long learning curve for doctors in this process. In this case, it is very expensive to train an excellent congenital heart disease diagnostic expert, both in terms of time cost and resource cost. Therefore, using deep learning to assist cardiologists can largely help doctors analyze and diagnose more effectively.

In the current field of computer vision, deep learning is arguably a state-of-the-art technique for image analysis. Therefore, deep learning is also widely used in the research of medical imaging, such as medical image segmentation [1-5], image classification and prediction. For example, in terms of medical image classification and prediction, work in [6-7] used deep neural networks for Alzheimer's disease classification tasks, and the experimental results showed good performance. In [8-13], the task of using deep learning for pneumonia has also made good progress.

In the study of adult cardiology diagnosis using echocardiography, Arnaout [14] proposed that deep neural networks can be used to better interpret echocardiography,

*Corresponding Author: Xinrong Chen; Email: chenxinrong@fudan.edu.cn
DOI: https://doi.org/10.70003/160792642025092605004

thus assisting physicians in diagnosis. Madani et al. [15] proposed the use of convolutional neural networks for multi-view classification of echocardiography, which had a high accuracy in the experimental results, with an overall prediction accuracy of 97.8%, promoting the diagnosis of artificial intelligence-assisted echocardiography. Ouyang et al. [16] proposed a beat-to-beat deep learning method, using convolutional neural networks to label the left ventricle and extract features on echocardiographic video data, so as to achieve the evaluation of cardiac function.

Deep learning is also involved in the diagnosis of fetal congenital heart disease using echocardiography. Tan et al. [17] proposed a screening system for the diagnosis of hypoplastic left heart syndrome (HLHS), which includes the extraction of different standard planes and the identification of whether the subject has HLHS. Gong et al. [18] proposed a one-class classification network to classify patients with fetal CHD and healthy subjects, utilizing cycle adversarial learning and transfer learning to further improve the recognition accuracy and robustness of the model. Experiments have shown that the model achieves a recognition rate of 84% in identifying fetal CHD. Sundaresan et al. [19] proposed to use fully convolutional neural network to segment the fetal heart plane in the ultrasound video frame, and then further realize the heart detection and the classification of the heart plane. Komatsu et al. [20] used convolutional neural networks to detect cardiac structural abnormalities in fetal ultrasound videos. Its model can achieve an AUC of 0.787 in diagnosing cardiac abnormalities. S. Nurmaini et al. [21] proposed a computer-assisted echocardiographic examination of fetal CHD, using the deep learning network Mask-RCNN to segment and detect defects in standard heart views. The results show that the average accuracy of this method in classification tasks is 98.30%. S. Qiao et al. [22] proposed a simple yet effective residual learning diagnosis system and the experimental results show that the accuracy of this method on the test set is 93%. S. Nurmaini et al. [23] proposed a novel multi classification method for fetal CHD based on DenseNet201 and the results show that the sensitivity, the specificity and the accuracy are 100% for the intra-patient scenario. R. Arnaout et al. [24] proposed the use of ensemble deep neural network to classify five types of heart views, and used this to classify normal hearts and complex fetal CHD. This model achieved the area under curve (AUC) of 0.99, the sensitivity of 95%, the specificity of 96% and 100% negative predictive value in distinguishing normal from fetal CHD.

Although some previous studies have achieved good results in the diagnosis of fetal CHD, there are still some shortcomings: 1) It is easy to ignore the key information of small lesions. 2) The ability to extract feature association information between different regions is insufficient.

In order to solve the above problems, this paper proposes a more efficient deep learning-based image classification model for fetal CHD, which is improved on ConvNeXt V1 [25] and combined with attention mechanism to better diagnose fetal CHD. Meanwhile, due to the small number of samples of fetal CHD, this study uses the method of transfer learning to pre-train the model and fine-tune it on a small sample of fetal CHD before predicting and diagnosing it. The contributions of this paper are summarized as follows:

1) This paper proposes an effective deep learning model to diagnose and predict fetal congenital heart disease, using ConvNeXt V1 and attention mechanism to process feature information in more detail, and adopting transfer learning to solve the small samples of fetal CHD. Compared with previous models, this model has better performance.
2) This paper provides visual explanations for the model to increase its interpretability. The confusion matrix provides a global explanation on diagnosing fetal CHD and the feature map visualization explains the local process of the model learning features.

2 Related Work

2.1 Convolutional Neural Network Model (CNN)

With the continuous deepening and optimization of deep learning algorithm models, the computer's ability to process data and large amounts of data determine the development of deep learning. Driven by the high-speed parallel computing technology of large-scale graphics processing units (GPU), in the field of deep learning, AlexNet [26] stood out in the competition of the ImageNet [27] data set, and AlexNet pushed the depth of the world for the world. The door to learning, followed by VGGNet [28], GoogleNet [29], ResNet [30], DenseNet [31], ShuffleNet [32] and other representative deep learning models have also been used by many studies to this day.

However, since most models expand the receptive field by stacking smaller convolutions, each output contains a smaller area of information. In the process of continuous hardware development, convolutional neural networks using large convolution kernels have also been proposed in the field of deep learning to improve model efficiency, such as ReplkNet [33] and ConvNeXt. Compared with a small convolution kernel network, the use of a large convolution kernel can improve the effective receptive field more efficiently, which is conducive to the extraction of contextual information. Compared with small-kernel CNNs, large-kernel CNNs have higher-level shape biases rather than texture biases. Therefore, this paper proposes to use a large-kernel convolution method to extract detailed features from congenital heart disease images, so as to learn context information more efficiently.

2.2 Attention Mechanism

The attention mechanism was first used in the field of Natural Language Processing (NLP), and was later widely used in the field of Computer Vision (CV). Generally speaking, the purpose of the attention mechanism is to enable the model to pay more attention to the key information in the image, thereby improving the ability of feature representation. For example, CBAM [34] is a simple and effective attention module, which not only considers the channel dimension, but also the spatial dimension. The CBAM module can be seamlessly incorporated into any CNN model and can improve the performance of the backbone network.

In the medical field, experts can clearly understand the correlation of lesion characteristics based on years of clinical experience. Therefore, when analyzing images, they can independently ignore information that is not important for diagnosis, and focus on those key information features, and the attention mechanism coincides with this idea, so it is feasible to use the attention mechanism for diagnosis.

In recent years, attention modules have also been widely used in medical imaging. In [35], CBAM was used to extract the most informative features from pneumonia images. In order to improve the model's ability to extract feature information, Zhan et al. [36] proposed a new CNN structure based on the combination of DenseNet and attention for arrhythmia diagnosis.

In the study of CHD, because the fetal heart is small, it is not conducive to the key information extraction of the image, so this paper proposes an improved CBAM module to analyze the features of the image more efficiently.

2.3 Transfer Learning

With the continuous development of deep learning, transfer learning, as a main learning method of deep learning, has become an indispensable part of many application fields, especially in medical imaging research. There are fewer datasets, which may lead to overfitting of the deep learning model during training. Transfer learning can reduce the occurrence of this situation. Therefore, the learning method using transfer learning is widely used in the field of medical imaging. For example, Gajendran et al. [37] utilized transfer learning to transfer features learned from general natural image classification to ECG classification. Saito et al. [38] used a large number of heart disease images to pre-train the model in a simple CNN model. Y. Gao et al. [39] demonstrated that a CNN initialized with a large-scale pre-trained network outperforms a CNN trained directly with small-scale ultrasound data. Swati et al. [40] used Vgg19 for transfer learning, which can have a higher classification effect in MRI brain tumors. They also proposed in their research that because medical images are different from natural images, if only the last few layers of the model are fine-tuned, it will be difficult for the model to learn medical image features, and the performance will be improved by deep fine-tuning. Therefore, this paper also proposes to use deep fine-tuning to train the model.

3 Models

In this section, the basic network architecture of the proposed model ConvNeXt-AMTL is firstly described in detail, and then the Large-kernel Convolution Module (LM) and the Attention Module (AM) of the network is introduced, finally the transfer learning and the deep fine-tuning method of ConvNeXt-AMTL is discussed.

3.1 Overall Structure

The architecture of ConvNeXt-AMTL is shown in Figure 1. The framework of ConvNeXt-AMTL is improvement on the ConvNeXt network, and which combines the attention mechanism to extract the key information of the lesion area. ConvNeXt-AMTL takes the fetal echocardiography that need to be classified into categories as the input of the network. First, the input image goes through a convolution layer with a convolution kernel size of 4×4, and the Layer Normalization (LN) is down sampled by 4 times; then it is passed into the network composed of a large-kernel convolution module (LM) and an attention module (AM) to further process the feature information; in the procedure of processing the feature information, three down-sampling operations of 2 times will be performed, its purpose is to let ConvNeXt-AMTL learn feature information of different scales; finally, the feature information is feed to the global average pooling (GAP) layer, fully connected layer, and softmax classifier to complete the final classification operation.

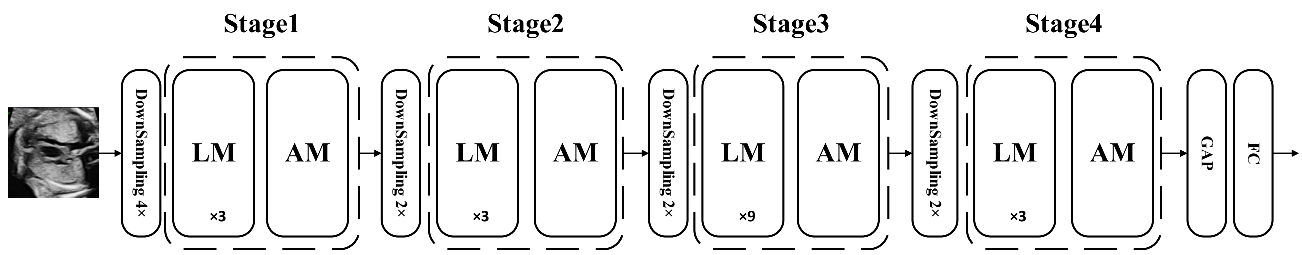

Figure 1. The architecture of ConvNeXt-AMTL

3.2 Large-kernel Convolution Module (LM)

In 2022, ConvNeXt V1 is proposed to use large convolution kernels to enhance the ability of convolutional neural networks to extract features. Therefore, this study also introduces the large-kernel convolutional network of ConvNeXt V1 into this module, but because the effective receptive field of the large convolution kernel is single and huge, this study adds two connection layers to the model to generate a combined receptive field. At the same time, more small local features are retained, which promotes the effective transmission and utilization of feature information, thereby avoiding the phenomenon of smooth transition. The Large-kernel Convolution Module (LM) is shown in Figure 2. This module is mainly composed of a 7×7 depth-separable convolutional layer, two connection layers, and two fully-connected layers.

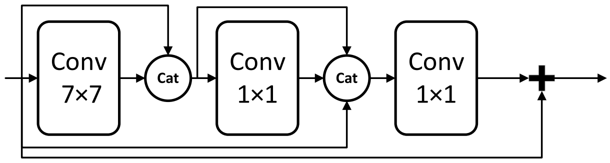

Figure 2. Large-kernel convolution module (LM)

3.3 Attention Module (AM)

Spatial attention can make the neural network pay more attention to the key information of the image and reduce the attention to the non-key information. Channel attention is an attention mechanism that considers the relationship between feature map channels. CBAM combines both attentions to achieve a more comprehensive attention structure. Hence, CBAM has been introduced and improved in this module.

In this module, the Channel Shuffle proposed by ShuffleNet is adopted to ensure that different semantic information between different channels can be exchanged before being input into CBAM, and the residual connections are used to ensure long-distance dependencies. The Attention Module (AM) is shown in Figure 3. This module is divided into two stages: the first is the channel attention stage, the second is the spatial attention stage. The detailed introduction to the two stages is as follows:

In the first stage, channel shuffle is conducted on the input feature map F to obtain the feature map F'. Then, the feature map F' is proceeded by average-pooling and max-pooling in spatial dimensions to obtain two spatial context descriptors. The two spatial context descriptors are feed to MLP, and then the two outputs are combined and activated through the sigmoid activation function to obtain the weight coefficient M_c of the channel. This channel weight coefficient is multiplied with the input feature, and the resulting map is merged with the original input feature map by residual connection to obtain a new feature map $F_c^{'}$. The process of the first stage is explained in equation (1).

$$\begin{cases} M_c = \sigma\left(MLP\left(AvgPool\left(F'\right) + MaxPool\left(F'\right)\right)\right) \\ F_c' = M_c \otimes F' + F \end{cases} \tag{1}$$

F is input feature map, F' is feature map processed by channel shuffle, σ is sigmoid activation function, M_c is weight coefficient of the channel, F_c' is final feature of channel attention stage, $\otimes$ is weighted multiplication of feature map.

In the second stage, the feature map F_c' is proceeded by average-pooling and max-pooling in channel dimensions to obtain two feature maps. The two feature maps are processed by concatenating in channel dimension and then the convolution operation is performed. The feature map obtained from convolution operation is activated through the sigmoid activation function to obtain the spatial weight coefficient M_s. Finally, this weight coefficient is multiplied with the input feature, and the resulting map is merged with the original feature map by residual connection to obtain the final feature map F_S' as output. The process of the second stage is explained in equation (2).

$$\begin{cases} M_S = \sigma\left(f^{3\times3}\left(AvgPool\left(F_c'\right) + MaxPool\left(F_c'\right)\right)\right) \\ F_S' = M_S \otimes F_c' + F \end{cases} \tag{2}$$

F is input feature map, F_c' is final feature of channel attention stage, σ is sigmoid activation function, M_s is weight coefficient of the spatial, F_S' is final feature of spatial attention stage, $\otimes$ is weighted multiplication of feature map.

3.4 Transfer Learning and Deep Fine-tuning of ConvNeXt-AMTL

First, use the Mini-ImageNet [41] dataset to pre-train ConvNeXt-AMTL, and then fine-tune this model in depth. The fetal CHD dataset is not similar to natural images. If this study only fine-tunes the last few layers of fully connected layers, the model will be very difficult to learn the image features of fetal CHD, therefore this study does not freeze the weights of the shallow layer, and uses the pre-trained ConvNeXt-AMTL to re-learn the fetal echocardiography, so as to achieve the purpose of deep fine-tuning.

4 Experimental Analysis

4.1 Dataset Source and Preprocessing

The fetal echocardiography dataset for this study was obtained from a hospital. The experimental data set randomly selected in this paper includes five views of the fetal heart from different angles, including three-vessel tracheal view (3VT), three-vessel view (3VV), apical four-chamber view (A4C), right ventricular outflow tract view (RVOT), and left ventricular outflow tract view (LOVT). Figure 4 and Figure 5 show some of the experimental images used. The experimental dataset includes 185 echocardiograms of healthy fetuses and 123 echocardiograms of fetuses with congenital heart disease. These echocardiograms have varying degrees of artifact noise, which is conducive to verifying the effectiveness and reliability of the model in handling the above diagnostic tasks.

In addition, this experiment split the entire data set and used 226 image data for model training, including 125 echocardiograms of healthy fetuses and 101 fetal echocardiograms with CHD. The test set and the training set are separate and independent. In the test set, 82 image data are used for test experiments, including 60 echocardiograms of healthy fetuses and 22 echocardiograms of fetuses with CHD.

In the process of image preprocessing, in order to keep the size of the image consistent when inputting the model, this experiment adjusts the size of the image to uniformly specify the size of the image as 224 pixels × 224 pixels. In addition, this experiment enhances the fetal echocardiography dataset through random cropping and random horizontal flipping operations, the purpose of which is to improve the generalization ability of the model to a certain extent.

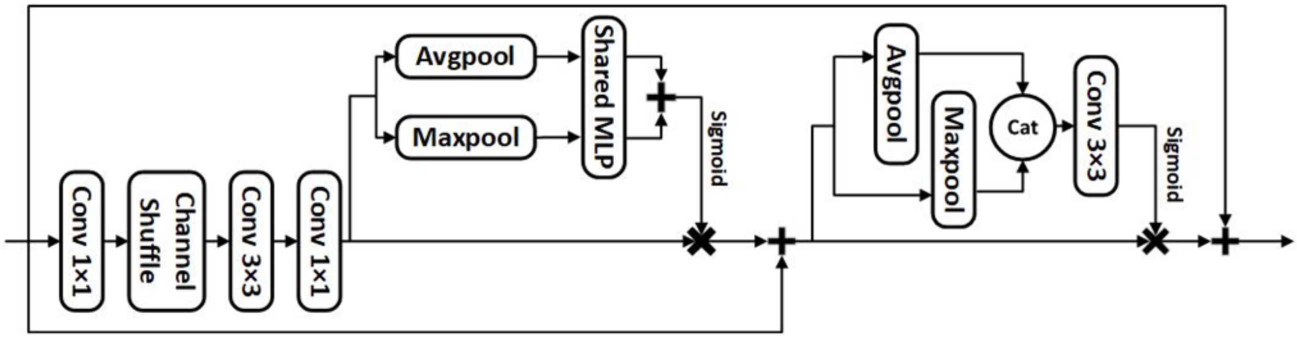

Figure 3. Attention Module (AM)

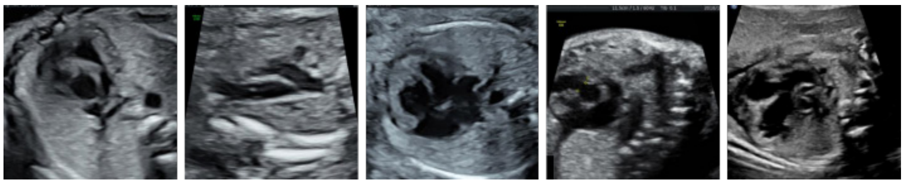

Figure 4. Samples of fetal CHD images

Figure 5. Samples of fetal healthy heart images

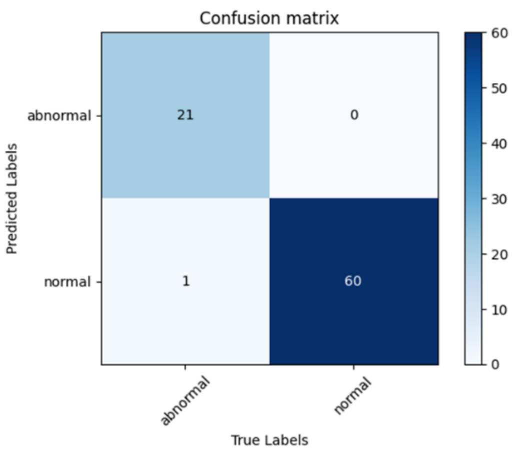

Figure 6. Confusion matrix of ConvNeXt-AMTL

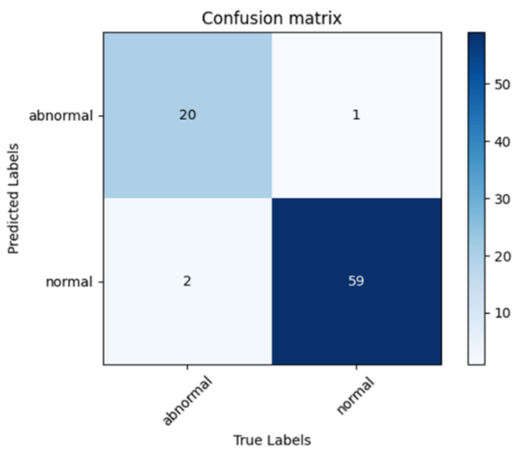

Figure 7. Confusion matrix of ConvNeXt-AMTL without TL

4.2 Evaluation Indicators

This article uses a confusion matrix to visualize the performance of the model. The performance of the model proposed in this paper is evaluated by obtaining four main indicators in the confusion matrix, including accuracy, precision, recall and F_1 score:

$$Accuracy = \frac{N_{Tp} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \tag{3}$$

$$Precision = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{4}$$

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{5}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

F_1 score is a combination of precision and recall, which is a comprehensive evaluation of the model in this paper. Therefore, the greater the precision of the model, the better the diagnostic effect on fetal congenital heart disease; the higher the recall value, the more sensitive the model is to fetal congenital heart disease. Range of F_1 score from 0 to 1, where 1 means the model has the best performance and 0 means the worst performance.

In addition, this paper also uses the ROC curve and AUC to measure the diagnostic performance of our model at different thresholds. ROC is a probability curve, and AUC means that it is a quantitative processing of the ROC curve, and they can well evaluate the ability of the model to distinguish between normal samples and abnormal samples.

4.3 Experimental Discussion

4.3.1 Visualization of Model Classification Results

This experiment uses a test set of 82 sample images (22 abnormal samples, 60 normal samples) to verify the effect of the model. The confusion matrix predicted by ConvNeXt-AMTL in this study is shown in Figure 6. It can be seen that a total of 21 fetal echocardiograms labeled as abnormal were diagnosed as abnormal, while only one of the samples marked as abnormal was predicted as normal. Among the sample images labeled as normal, they are all accurately predicted as normal. From the above experimental results, it can be seen that the use of this model and transfer learning is very efficient in the diagnosis of fetal congenital heart disease.

At the same time, this experiment also evaluated the performance of ConvNeXt-AMTL without transfer learning (TL). The confusion matrix predicted by ConvNeXt-AMTL without transfer learning (TL) is shown in Figure 7. It can be seen that there are 20 fetal echocardiograms with abnormal labels was diagnosed as abnormal, and only 2 of the samples marked as abnormal were predicted to be normal. Fifty-nine fetal echocardiograms labeled normal were diagnosed as normal, and only 1 sample was predicted to be abnormal. It can be seen that the model ConvNeXt-AMTL without transfer learning (TL) also can maintain a high accuracy without transfer learning.

From the above experimental results, it can be seen that the research model still has high sensitivity to whether the fetus has CHD, which is conducive to the timely detection of patients in the diagnosis process, and can ensure that over-screening will not be performed. Therefore, in summary, it can be seen that ConvNeXt-AMTL has good performance in the diagnosis of fetal CHD.

4.3.2 Model Feature Map Visualization

In the previous section, the performance of ConvNeXt-AMTL in diagnosing fetal CHD was illustrated through the confusion matrix. In this section, the learning ability of the response model will be more intuitive in the form of feature map visualization. In this experiment, several channels were randomly selected from the feature maps of the Large-kernel Convolution Module (LM) and Attention Module (AM) for visualization. As shown in Figure 8 and Figure 9, the Large-kernel Convolution Module (LM) can capture low-level feature, including texture and shape, and can clearly outline feature information; while the Attention Module (AM) can further focus on key feature information in the stage1.

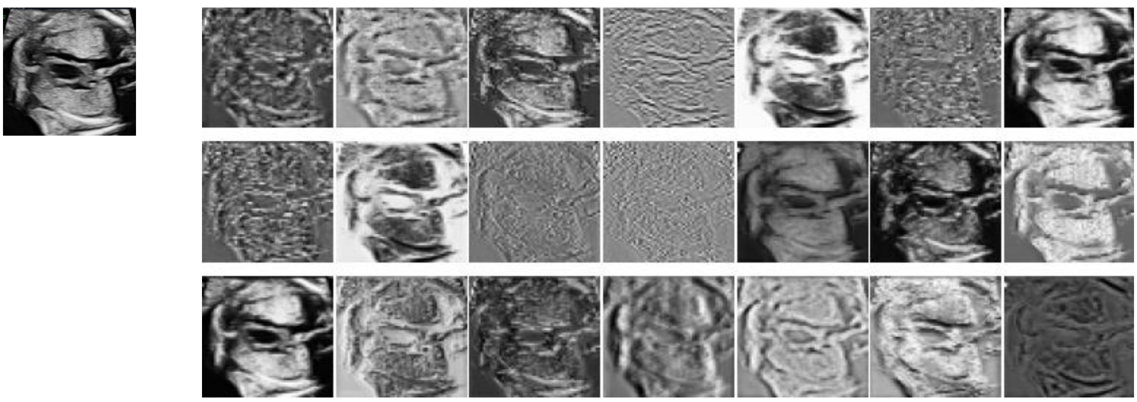

Figure 8. Visualization of feature maps in LM

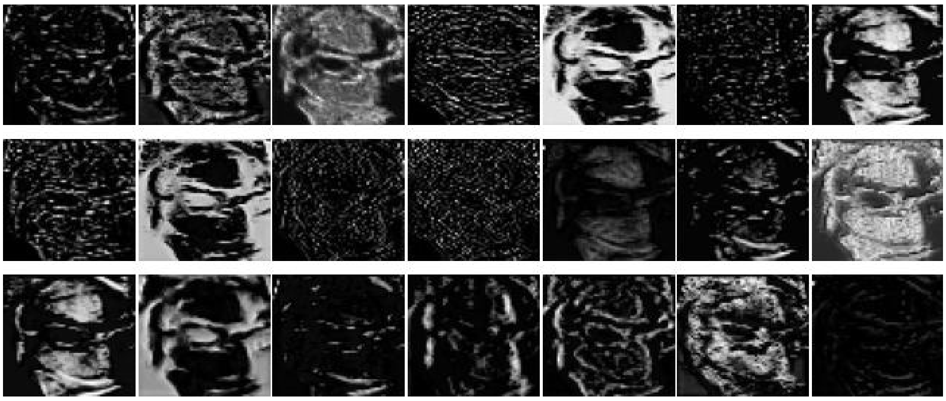

Figure 9. Visualization of feature maps in AM

4.3.3 Model Comparison

This experiment compared the performance indicators between different models. As shown in Table 1, ConvNeXt-AMTL performed the best in diagnosing fetal congenital heart disease, achieving the precision of 99.2%, the recall of 97.8%, F1 score of 98.5% and an accuracy rate of 98.8%.

Table 1. Comparison of performance indicators of different network models

	Precision	Recall	F_1 score	Accuracy
ResNet34				
Abnormal	0.95	0.864	0.905	
Normal	0.952	0.983	0.967	
Average	0.951	0.924	0.937	0.951
ShuffleNetV2				
Abnormal	0.938	0.682	0.790	
Normal	0.894	0.983	0.936	
Average	0.916	0.833	0.872	0.902
ConvNeXtV1				
Abnormal	0.516	0.727	0.604	
Normal	0.882	0.750	0.811	
Average	0.699	0.739	0.718	0.744
DenseNet121				
Abnormal	1.0	0.773	0.872	
Normal	0.923	1.0	0.960	
Average	0.962	0.887	0.922	0.939
ConvNeXt-AMTL without TL				
Abnormal	**0.952**	**0.909**	**0.930**	
Normal	**0.967**	**0.983**	**0.975**	
Average	**0.960**	**0.946**	**0.953**	**0.963**
ConvNeXt-AMTL				
Abnormal	**1.0**	**0.955**	**0.977**	
Normal	**0.984**	**1.0**	**0.992**	
Average	**0.992**	**0.978**	**0.985**	**0.988**

In addition, this experiment also compared the ROC curves between different models. As shown in Figure 10, ConvNeXt-AMTL reached the best AUC of 0.999. Finally, this experiment also analyzed the loss curves and accuracy curves between different models. As shown in Figure 11 and Figure 12, it can be seen from the learning curve that ConvNeXt-AMTL has the best convergence. Therefore, ConvNeXt-AMTL proposed in this paper is more effective in diagnosing fetal CHD.

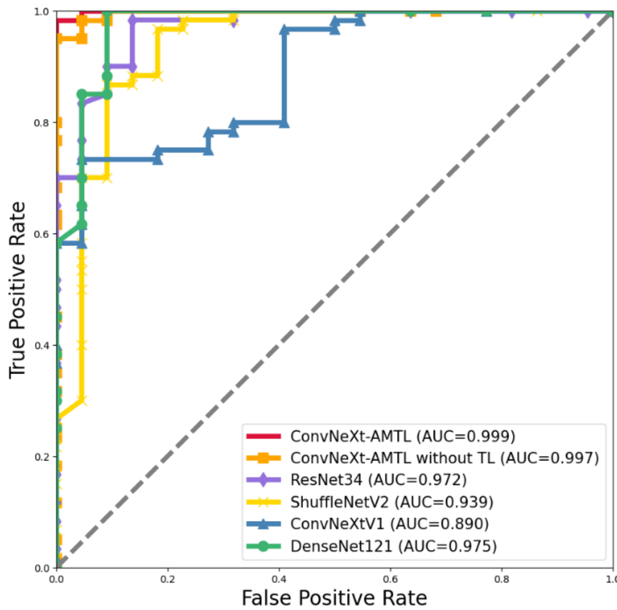

Figure 10. ROC curves of different models

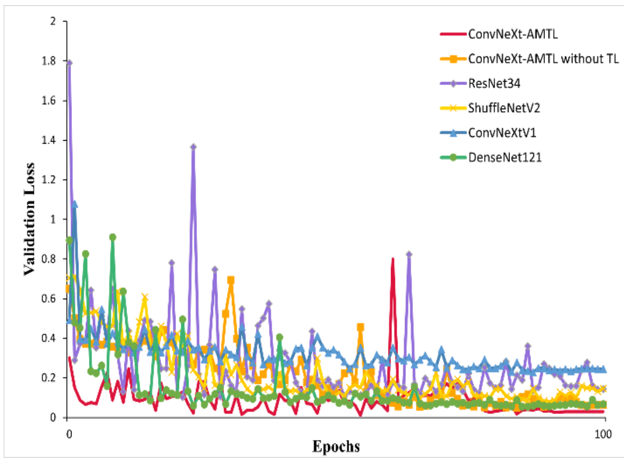

Figure 11. Loss curves of different models

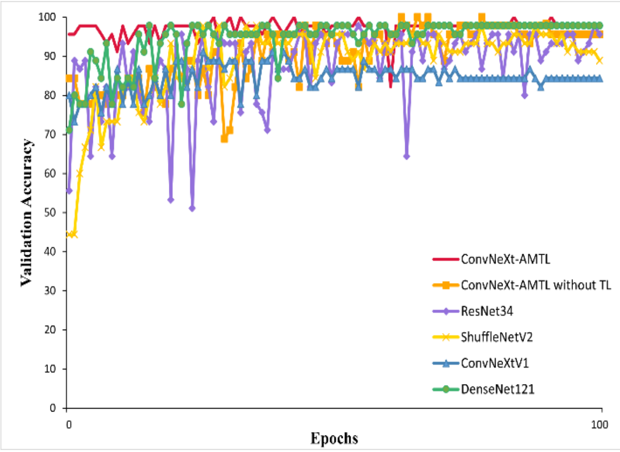

Figure 12. Accuracy curves of different models

5 Conclusion

In this paper, a deep learning model ConvNeXt-AMTL is proposed to diagnose fetal congenital heart disease. In this model a module using a large-kernel convolution is proposed to learn the correlation between different heart regions. In addition, in order to focus on key information and capture detailed features, this paper also proposes a module using attention mechanism. Furthermore, the pre-trained model is deeply fine-tuned through transfer learning on the CHD dataset. Experiments show that the accuracy rate of ConvNeXt-AMTL in the diagnosis of fetal CHD reached 98.8%, and the recall rate reached 97.8%, which significantly improved the accuracy of diagnosis of fetal CHD.

However, there are still some limitations in this study, because there are few abnormal samples of fetal heart, and the potential of the model cannot be more powerfully explained due to the limitation of image scale, and it is a great challenge to obtain enough images in a short time. Not only clinical acquisition is required, but also rigorous labeling of images is required. Therefore, more fetal CHD images will continue to be added in the future to improve the potential of the model.

Acknowledgements

Fang Li, Xueqin Ji contributed equally.

References

[1] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A nested U-Net architecture for medical image segmentation, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Granada, Spain, 2018, pp. 3-11.

[2] M. Y. Durrani, S. Yasmin, S. Rho, An Internet of Medical Things Based Liver Tumor Detection System using Semantic Segmentation, *Journal of Internet Technology*, Vol. 23, No. 2, pp. 363-375, March, 2022.

[3] F. Xie, P. Zhang, T. Jiang, J. She, X. Shen, P. Xu, W. Zhao, G. Gao, Z. Guan, Lesion Segmentation Framework Based on Convolutional Neural Networks with Dual Attention Mechanism, *Electronics*, Vol. 10, No. 24, Article No. 3103, December, 2021.

[4] H. C. Park, R. Ghimire, S. Poudel, S. W. Lee, Deep Learning for Joint Classification and Segmentation of Histopathology Image, *Journal of Internet Technology*, Vol. 23, No. 4, pp. 903-910, July, 2022.

[5] N. Yamanakkanavar, B. Lee, A novel M-SegNet with global attention CNN architecture for automatic segmentation of brain MRI, *Computers in Biology and Medicine*, Vol. 136, Article No. 104761, September, 2021.

[6] A. Farooq, S. M. Anwar, M. Awais, S. Rehman, A deep CNN based multi-class classification of Alzheimer's disease using MRI, *2017 IEEE International Conference on Imaging Systems and Techniques*, Beijing, China, 2017, pp. 111-116.

[7] A. Farooq, S. M. Anwar, M. Awais, M. Alnowami, Artificial intelligence based smart diagnosis of alzheimer's disease and mild cognitive impairment, *2017 International Smart Cities Conference*, Wuxi, Jiangsu, China, 2017, pp. 1-4.

[8] S. H. Kassania, P. H. Kassanib, M. J. Wesolowskic, K. A. Schneidera, R. Detersa, Automatic Detection of Coronavirus Disease(COVID-19) in X-ray and CT Images: A Machine Learning Based Approach, *Biocybernetics and Biomedical Engineering*, Vol. 41, No. 3, pp. 867-879, July-September, 2021.

[9] K. Gao, J. Su, Z. Jiang, L. Zeng, Z. Feng, H. Shen, P. Rong, X. Xu, J. Qin, Y. Yang, W. Wang, D. Hu, Dual-branch combination network (DCN): Towards accurate diagnosis and lesion segmentation of COVID-19 using CT images, *Medical Image Analysis*, Vol. 67, Article No. 101836, January, 2021.

[10] M. Shorfuzzaman, M. S. Hossain, MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients, *Pattern Recognition*, Vol. 113, Article No. 107700, May, 2021.

[11] S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora, R. B. Pachori, Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study, *Biomedical Signal Processing and Control*, Vol. 64, Article No. 102365, February, 2021.

[12] K. E. Asnaoui, Y. Chawki, Using X-ray images and deep learning for automated detection of coronavirus disease, *Journal of biomolecular structure & dynamics*, Vol. 39, No. 10, pp. 3615-3626, 2021.

[13] A. I. Khan, J. L. Shah, M. M. Bhat, CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images, *Computer Methods and Programs in Biomedicine*, Vol. 196, Article No. 105581, November, 2020.

[14] R. Arnaout, Toward a clearer picture of health, *Nature Medicine*, Vol. 25, No. 1, Article No. 12, January, 2019.

[15] A. Madani, R. Arnaout, M. Mofrad, R. Arnaout, Fast and accurate view classification of echocardiograms using deep learning, *NPJ digital medicine*, Vol. 1, No. 1, Article No. 6, March, 2018.

[16] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C. P. Langlotz, P. A. Heidenreich, R. A. Harrington, D. H. Liang, E. A. Ashley, J. Y. Zou, Video-based AI for beat-to-beat assessment of cardiac function, *Nature*, Vol. 580, No. 7802, pp. 252-256, April, 2020.

[17] J. Tan, A. Au, Q. Meng, S. FinesilverSmith, J. Simpson, D. Rueckert, R. Razavi, T. Day, D. Lloyd, B. Kainz, Automated Detection of Congenital Heart Disease in Fetal Ultrasound Screening, in: Y. Hu, R. Licandro, J. A. Noble, J. Hutter, S. Aylward, A. Melbourne, E. A. Turk, J. T. Barrena (Eds.), *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis*, Vol. 12437, Springer, Cham, 2020, pp. 243-252.

[18] Y. Gong, Y. Zhang, H. Zhu, J. lv, Q. Cheng, H. Zhang, Y. He, S. Wang, Fetal congenital heart disease echocardiogram screening based on DGACNN: adversarial one-class classification combined with video transfer learning, *IEEE transactions on medical imaging*, Vol. 39, No. 4, pp. 1206-1222, April, 2020.

[19] V. Sundaresan, C. P. Bridge, C. Ioannou, J. A. Noble, Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks, *2017 IEEE 14th International Symposium on Biomedical Imaging*, Melbourne, VIC, Australia, 2017, pp. 671-674.

[20] M. Komatsu, A. Sakai, R. Komatsu, R. Matsuoka, S. Yasutomi, K. Shozu, A. Dozen, H. Machino, H. Hidaka, T. Arakaki, K. Asada, S. Kaneko, A. Sekizawa, R. Hamamoto, Detection of Cardiac Structural Abnormalities in Fetal Ultrasound Videos Using Deep Learning, *Applied Sciences*, Vol. 11, No. 1, Article No. 371, January, 2021.

[21] S. Nurmaini, M. N. Rachmatullah, A. I. Sapitri, A. Darmawahyuni, B. Tutuko, F. Firdaus, R. U. Partan, N. Bernolian, Deep Learning-Based Computer-Aided Fetal Echocardiography: Application to Heart Standard View Segmentation for Congenital Heart Defects Detection, *Sensors*, Vol. 21, No. 23, Article No. 8007, December, 2021.

[22] S. Qiao, S. Pang, G. Luo, S. Pan, Z. Yu, T. Chen, Z. Lv, RLDS: An explainable residual learning diagnosis system for fetal congenital heart disease, *Future Generation Computer Systems*, Vol. 128, pp. 205-218, March, 2022.

[23] S. Nurmaini, R. U. Partan, N. Bernolian, A. I. Sapitri, B. Tutuko, M. N. Rachmatullah, A. Darmawahyuni, F. Firdaus, J. C. Mose, Deep Learning for Improving the Effectiveness of Routine Prenatal Screening for Major Congenital Heart Diseases, *Journal of clinical medicine*, Vol. 11, No. 21, Article No. 6454, November, 2022.

[24] R. Arnaout, L. Curran, Y. Zhao, J. C. Levine, E. Chinn, A. J. Moon-Grady, An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease, *Nature Medicine*, Vol. 27, No. 5, pp. 882-891, May, 2021.

[25] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrel, S. Xie, A ConvNet for the 2020s, *2022 IEEE/CVF Conference on computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 11966-11976.

[26] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Communications of the ACM*, Vol. 60, No. 6, pp. 84-90, June, 2017.

[27] J. Deng, W. Dong, R. Socher, L. Li, K. Li, F. Li, ImageNet: A Large-Scale Hierarchical Image Database, *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248-255.

[28] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *The 3rd International Conference on Learning Represen-tations*, San Diego, CA, USA, 2015, pp. 1-14.

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 1-9.

[30] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 770-778.

[31] G. Huang, Z. Liu, L. V. D. Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 2261-2269.

[32] N. Ma, X. Zhang, H. Zheng, J. Sun, ShuffleNet V2: Practical guidelines for efficient CNN architecture design, *The 15th European Conference on Computer Vision*, Munich, Bavaria, Germany, 2018, pp. 122-138.

[33] X. Ding, X. Zhang, J. Han, G. Ding, Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs, *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 11963-11975.

[34] S. Woo, J. Park, J. Y. Kee, I. S. Kweon, CBAM: Convolutional Block Attention Module, *The 15th European Conference on Computer Vision*, Munich, Bavaria, Germany, 2018, pp. 3-19.

[35] M. Canayaz, C+EffxNet: A novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet, *Chaos Solitons & Fractals*, Vol. 151, Article No. 111310, October, 2021.

[36] Q. Zhan, P. Li, Y. Wu, J. Huang, X. Dong, A novel CNN model with dense connectivity and attention mechanism for arrhythmia classification, *2022 IEEE 35th International Symposium on Computer-Based Medical Systems*, Shenzhen, Guangdong, China, 2022, pp. 50-55.

[37] M. K. Gajendran, M. Z. Khan, M. A. K. Khattak, ECG Classification using Deep Transfer Learning, *2021 4th International Conference on Information and Computer Technologies*, Kahului, HI, USA, 2021, pp. 1-5.

[38] K. Saito, Y. Zhao, J. Zhong, Heart Diseases Image Classification Based on Convolutional Neural Network, *2019 International Conference on Computational Science and Computational Intelligence*, Las Vegas, NV, USA, 2019, pp. 930-935.

[39] Y. Gao, M. A. Maraci, J. A. Noble, Describing ultrasound video content using deep convolutional neural networks, *2016 IEEE 13th International Symposium on Biomedical Imaging*, Prague, Czech Republic, 2016, pp. 787-790.

[40] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu, Brain tumor classification for MR images using transfer learning and fine-tuning, *Computerized Medical Imaging and Graphics*, Vol. 75, pp. 34-46, July, 2019.

[41] S. Ravi, H. Larochelle, Optimization as a Model for Few-Shot Learning, *The 5th International Conference on Learning Representations*, ICLR 2017, Toulon, Var, France, 2017, pp. 1-11.

Biographies

Fang Li Assistant professor in Shanghai University of Electric Power. Her current research focuses on image processing, pattern recognition and machine learning.

Rui Yang Student of Shanghai University of Electric Power. His research interests include computer vision and deep learning.

Xueqin Ji Director of Ultrasound Department, Ningxia Women and Children's Hospital, Peking University First Hospital. Her current research interests include prenatal ultrasound diagnosis and prognosis evaluation of fetal malformations, maternal fetal medical research, etc.

Wei Tang received a bachelor's degree in engineering and graduated from Chongqing University in 2020 with a bachelor's degree in measurement, control technology and instrumentation. Currently studying for a master's degree in electronic information from Fudan University. Main research interests include computer vision and machine learning.

Xinrong Chen received the B.S. degree in electronic and information engineering from the Nanjing University of Science and Technology, in 2004, and the M.S. degree in signal and information processing from Southeast University, Nanjing, China, in 2007, and the Ph.D. degree in biomedical engineering from Fudan University, China, in 2014. He is currently a Professor with the Academy for Engineering and Technology, Fudan University, China. His current research interests include medical image analysis, computer vision, and image-guided surgery.