IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 July 2022, accepted 10 August 2022, date of publication 17 August 2022, date of current version 24 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3199419

RESEARCH ARTICLE

Convolution Neural Network With Coordinate Attention for the Automatic Detection of Pulmonary Tuberculosis Images on Chest X-Rays

TIANHAO XU AND ZHENMING YUAN
School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China

Corresponding author: Zhenming Yuan (zmyuan@hznu.edu.cn)

This work was supported in part by the Science and Technology Research Project of the Corps under Grant 2021AB034-2, and in part by the Natural Science Foundation of Zhejiang Province under Grant LGF20F020009.

ABSTRACT Tuberculosis is a chronic respiratory infectious disease that seriously endangers human health. Diagnosis of pulmonary tuberculosis usually depend on the analysis of chest X-rays by radiologists. However, there is a certain misdiagnosis rate with time consuming. Therefore, the purpose of this study is to propose a low-cost and automatic detection method of pulmonary tuberculosis images on chest X-rays to help primary radiologists. A pulmonary tuberculosis classification algorithm based on convolution neural network is proposed, which uses deep learning to classify chest X-ray images. Our method introduces coordinate attention mechanism into convolutional neural network (VGG16), so that the algorithm can capture not only cross-channel information, but also direction sensing and position sensing information, in order to better identify and classify pulmonary tuberculosis images. During the training process, we use the method of transfer learning and freeze network to make the model fit faster. The performance of our method is evaluated on the public dataset of tuberculosis classification of Shenzhen Third Hospital, China. We take the average data through 5-fold cross validation: accuracy= 92.73%, AUC= 97.71%, precision= 92.73%, recall= 92.83%, F1 score = 92.82%. Compared with the existing end-to-end method based on convolutional neural network (CNN), our method is superior to ConvNet, FPN + Faster RCNN and other methods. The comparison results with other methods show that our method has better accuracy, which can help radiologists make auxiliary diagnosis.

INDEX TERMS Classification, convolutional neural network, CoordAttention, deep learning, tuberculosis.

I. INTRODUCTION

Tuberculosis (TB) is one of the ten leading causes of death worldwide. In China, the incidence and death of pulmonary tuberculosis ranked second. If the early detection is not timely, tuberculosis will spread widely in the body and cause lung tissue necrosis to form cavities, which is prone to massive hemoptysis. Therefore, rapid detection and diagnosis of tuberculosis has become extremely important. Diagnosis of tuberculosis depends on the analysis of chest X-rays [1], [2] by radiologists. The chest X-ray examination can not only detect tuberculosis early, but also smaller or hidden lesions can be found. However, there is a certain misdiagnosis rate and time consuming by manual, and it also requires the radiologists to have rich medical imaging knowledge.

In recent years, deep learning has performed well in the field of image object detection and classification, and also has many good applications in medical imaging. One of the most popular models is the Convolutional Neural Network (CNN) model, which uses feature eaxtraction to identify images. However, the detection and classification of pulmonary tuberculosis is facing challenges. The pulmonary tuberculosis images often have noise and the interference of other tissue, as well as there also exists a problem with a small number of training images, which leads to the simple CNN model can not well identify the features. In medical imaging, some researchers proposed to solve similar problems by adding attention mechanism. Inspired by this, this paper introduces

The associate editor coordinating the review of this manuscript and approving it for publication was Sabu M. Thampi.

the coordinate attention [3] based on the traditional convolutional neural network VGG16 [4]. Although the attention mechanism was first used in natural language processing (NLP), it also has a very ideal effect in image processing. The attention mechanism allows the convolutional neural network to select the focus position when extracting features, so as to produce more discriminative feature representation and attention-aware features, and the features of different modules will change adaptively with the deepening of the network [5]. As a plug-and-play module, the attention mechanism is also very suitable for modification on traditional convolutional neural networks. Compared with Channel Attention (ECANET) [6] and Spatial Attention (CBAM) [7], Coordinate Attention not only captures cross-channel information, but also direction-aware and position-sensitive information, which allows the model to locate and identify the target region more accurately. It has a better help for the detection and classification of pulmonary tuberculosis chest X-ray. At the same time, we train the model by adding transfer learning to freeze the network to speed up the training speed and improve the classification accuracy. Finally, we pass five cross-validations to verify the robustness of the method.

In this paper, an algorithm based on convolution neural network with coordinate attention for the automatic detection of pulmonary tuberculosis images on chest X-rays was proposed. We use different pretrained network models to detect the classification results of pulmonary tuberculosis in Chest X-Ray (CXR) images, compare with the classification results of the transfer learning freeze network to evaluate the effectiveness of the training scheme. Compared with the existing end-to-end schemes, our method achieves better results without artificial feature extraction and lung segmentation.

The rest of the paper is organized as follows: Section 2: related work. Section 3: dataset. Section 4: introduces the proposed model of VGG16-CoordAttention, the training mode, and five cross-validation. Section 5: experimental result. Section 6: conclusion and discussion.

II. RELATED WORK

In this section, we will explore some machine learning or deep learning-based TB diagnostic schemes. The diagnosis of tuberculosis is mainly based on images obtained by examining chest x-rays. However, the diagnostic identification of medical images requires experienced and specialized radiologists, which is a major obstacle to the effective diagnosis of TB. Compared to human diagnosis, computer algorithms can provide more important results with less diagnostic errors and time consuming, and use less resources to achieve efficient image diagnosis of large-scale tuberculosis.

In previous studies, deep learning, such as convolutional neural networks have achieved relatively good results in the classification and recognition of tuberculosis medical images. Eman Showkatian *et al.* [8] proposed a CNN architecture (ConvNet), which achieved 88.0% accuracy, 87.0% sensitivity, 87.0% F1 score, 87.0% Precision and 87.0% AUC on the tuberculosis classification data sets in Shenzhen and Montgomery, China. Qingchen, Zhang *et al.* [9] achieved an accuracy of 87.71% on the Montgomery tuberculosis classification dataset by changing the global average pooling of the network model to an adaptive dropout layer based on the residual network ResNet50. Xie *et al.* [10] proposed a multiclass TB lesion detection method. The algorithm introduces a learning scalable pyramid structure in Fast Region Convolutional Neural Network (Faster RCNN) and achieves good performance on two public datasets, Montgomery dataset: AUC: 97.7.% and accuracy: 92.6%; Shenzhen dataset: AUC: 94.1% and accuracy: 90.2%. Abideen, Zain *et al.* [11] proposed a solution for tuberculosis identification via Bayesian Convolutional Neural Networks (B-CNN). Compared with state-of-the-art machine learning and CNN methods, B-CNN achieves 96.42% (Montgomery dataset) and 86.46% (Shenzhen dataset) accuracy on the two datasets, respectively. In addition, there are also research reports on the preprocessing of medical images by means of data enhancement. Munadi *et al.* [12] achieved 89.92% classification accuracy and 94.8% AUC on ResNet and EfficientNet models through image enhancement methods such as Unsharp Masking (UM), High-Frequency Emphasis Filtering (HEF) and Contrast Limited Adaptive Histogram Equalization (CLAHE). All the results were obtained using the public Shenzhen dataset. Among the traditional methods, there are studies through methods such as machine learning. Jaeger *et al.* [13] calculated texture and shape features in the lung area cut by the graph, and obtained the data set collected by the tuberculosis control project of the local county health department in the United States and the Shenzhen data set through the method of machine learning linear logistic regression (LLR). The area under the ROC curve (AUC) is 87% (78.3% accuracy), and the Shenzhen dataset AUC is 90% (84% accuracy).

In the existing studies, we found that the current end-to-end methods still have limitations in TB image detection. Due to the interference of many other tissues in TB images, a single simple convolutional neural network and data enhancement method cannot capture the key information in medical images, making the detection of TB images still challenging. In order to solve the problem, this paper will incorporate an attention mechanism to better extract image features and so as to improve the classification accuracy of pulmonary tuberculosis images.

III. DATASET

All experiments carried out in this paper used Shenzhen tuberculosis CXR [14]. The Shenzhen dataset was collected in collaboration with Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China. The chest X-rays are from outpatient clinics and were captured as part of the daily hospital routine within a 1-month period, mostly in September 2012, using a Philips DR Digital Diagnost system. The dataset has no exclusion criteria, such as gender, age, or race. The set contains 662 frontal chest X-rays, of which 326 are normal cases and 336 are cases with manifestations of TB, including pediatric X-rays (AP). The

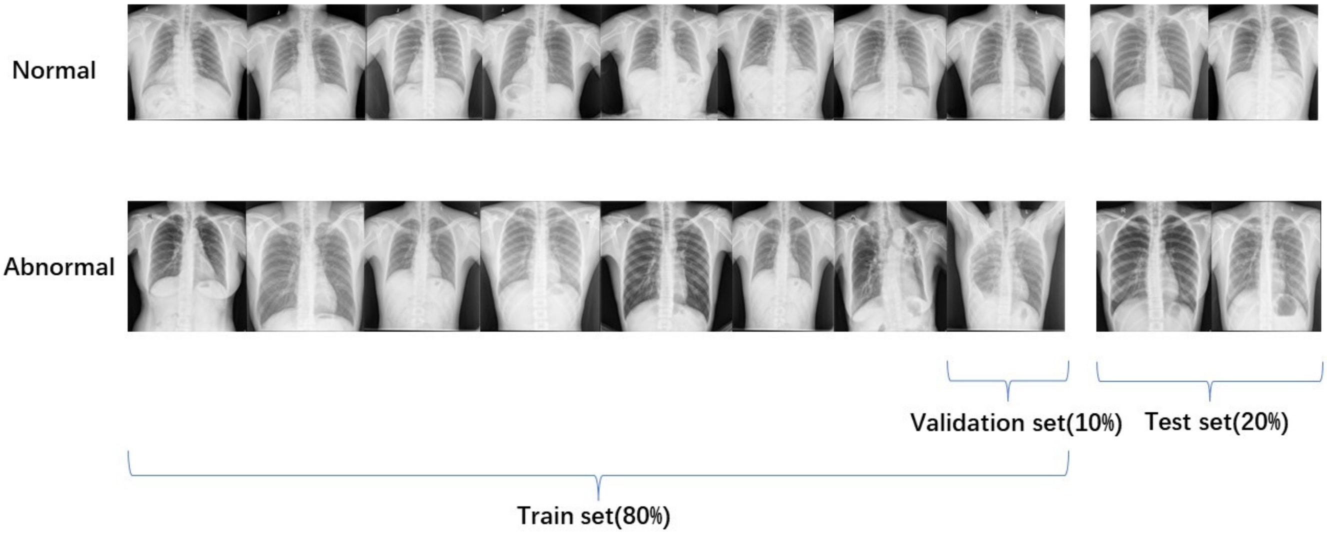

FIGURE 1. Shenzhen pulmonary tuberculosis dataset.

X-rays are provided in PNG format. Their size can vary but is approximately 3K × 3K pixels.

Each network model has a specific input image size. In order to meet the input requirements of this model, we standardize the images of the data set. We uniformly convert the dataset images into 224 $*$ 224 resolution to send them to the network for training. This resolution meets the requirements of the model, can satisfactorily retain the structural details of the image, and significantly reduce the calculation cost of the model. In terms of data set division, we randomly select 20% of the sample size from normal and tuberculosis images as the test set to ensure the balance of positive and negative samples in the dataset. Another 10% is randomly selected from the training set as the validation set for the model adjustment. At the same time, considering the small number of data sets, we enhance the random data of the training set, flip, scale, twist the length and width of the image and transform the color gamut, so as to avoid over fitting the model and enhance the generalization ability of the model. The division of data set is shown in Fig.1.

IV. METHOD

Our method consists of a series of steps, including transfer learning, freezing network, feature extraction and classification tasks using supervised learning. We conducted three experiments to verify the feasibility of our method. In the first study, combining the coordinate attention mechanism with VGG16 network, we propose a VGG16-CoordAttention network model, which is trained by freezing the network to evaluate its classification effect on the Shenzhen dataset. In the second study, we use four representative models and our model as the backbone network to evaluate the effect of frozen network training. In the third study, we conduct five cross validation of the proposed method to verify the robustness of the model.

A. VGG16-COORDATTENTION MODEL

VGG16 network is a traditional deep convolution neural network. The network uses smaller convolution blocks. By increasing the network depth, the classification performance can be effectively improved. In order to improve the classification accuracy of tuberculosis images as much as possible, we combine VGG16 model with CoordAttention to establish a new tuberculosis deep learning network model Vgg16-CoordAttention. The network structure is shown in Fig.2.

VGG16 network is mainly composed of convolution layer, pooling layer and full connection layer. The network requires the input image to be a 3-channel 224 $*$ 224 size image. The network consists of five convolution blocks (Conv). In the conv1, performs 3 $*$ 3 convolution and ReLu activation function processing on the input image twice, the output feature layer is 64, and then through a 2 $*$ 2 maxpooling layer, the maxpool layer compresses the height and width of the image without changing the number of channels to obtain the image of (112,112,64). The Conv2 is the same as Conv1, and the output network is (56, 56,128). The Conv3, Conv4 and Conv5 all carry out three times of 3 $*$ 3 convolution and ReLu activation function processing on the input image, and then carry out global feature extraction through a 2 $*$ 2 maxpool layer. The final network output result is (7, 7,512). In order to achieve the final classification goal, the results are flatten and the classification results we want are realized through the full connection layer.

At the end of the VGG16 backbone network, in front of the maxpool layer of the Conv5, we add the coordinate attention mechanism and the ReLu activation function, which are hereinafter referred to as CA. CA is a simple and efficient attention mechanism. Its network structure is shown in Fig.3. By embedding the location information into the channel attention, the network can obtain the information of a larger

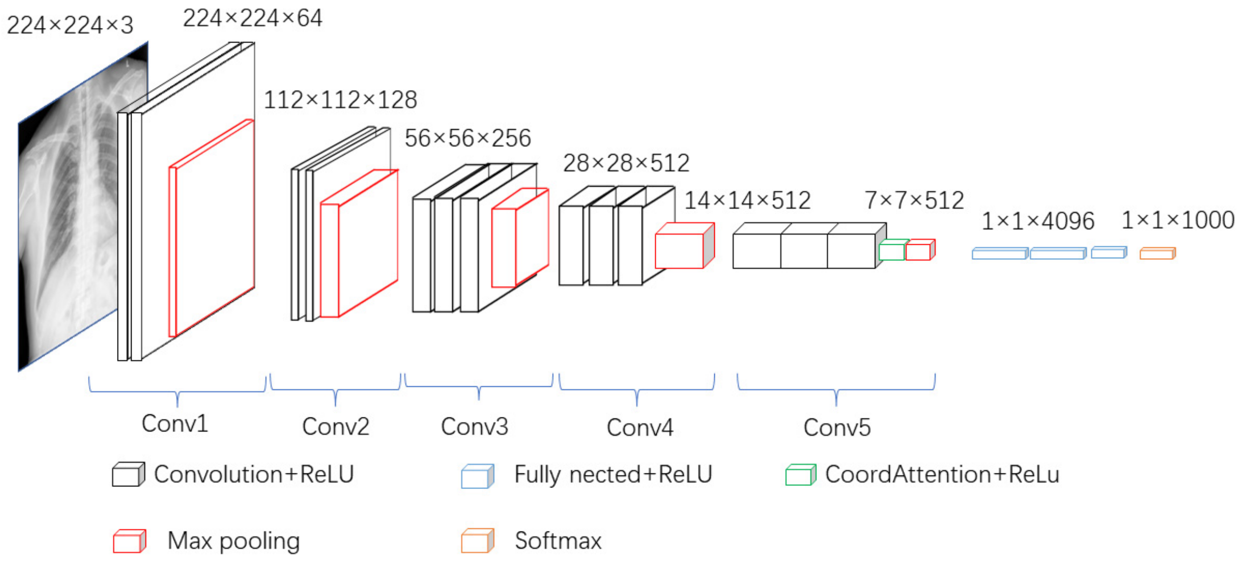

FIGURE 2. VGG16-CoordAttention network structure.

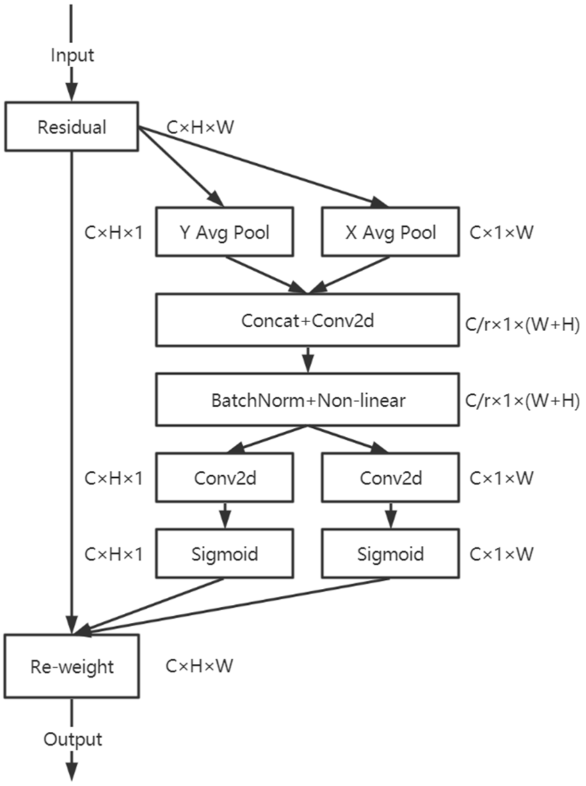

FIGURE 3. CoordAttention structure.

area and avoid excessive overhead. Through the average pooling operation of the vertical and horizontal coordinate positions, it decomposes the channel into two parallel feature codes, so that it can not only capture the information across the channel, but also capture the information of direction perception and position perception, which can help the model locate and identify the interested targets more accurately. At the same time, each attention map captures the long-range dependence of the input feature map along a spatial direction. Finally, the attention maps in two spatial directions are attached to the input feature map through convolution operation to restore the number of input channels, so as to enhance the ability of feature extraction. In addition, CA is flexible and lightweight, which can play the effect of plug and play. It only needs to determine the number of input and output channels. It can be well combined with traditional convolutional neural network and lightweight network, so as to enhance the feature by strengthening the information representation. Compared with the previous ECANET channel attention module and CBAM spatial attention module, ECANET only considers the internal channel information and ignores the importance of location information, lacking cross-channel information interaction; Although CBAM attention module attempts to introduce location information by global average pooling on the channel, this method can only capture local information, but can not obtain long-range dependent information. Due to the interference of other tissues in pulmonary tuberculosis images, CA can extract the features of the images from two directions to better determine the location of lesions, so as to improve the accuracy of classification.

B. TRAINING METHOD BASED ON TRANSFER LEARNING

In this section, we hope to evaluate the effectiveness of the transfer learning freezing network method through experiments. The method of freezing the network belongs to a kind of transfer learning. When we give the pretraining weight to the model, we freeze the backbone network model at the initial stage of training, so that the model does not change the backbone network weight at the initial stage of training and focuses on training the classifier, so as to put more resources

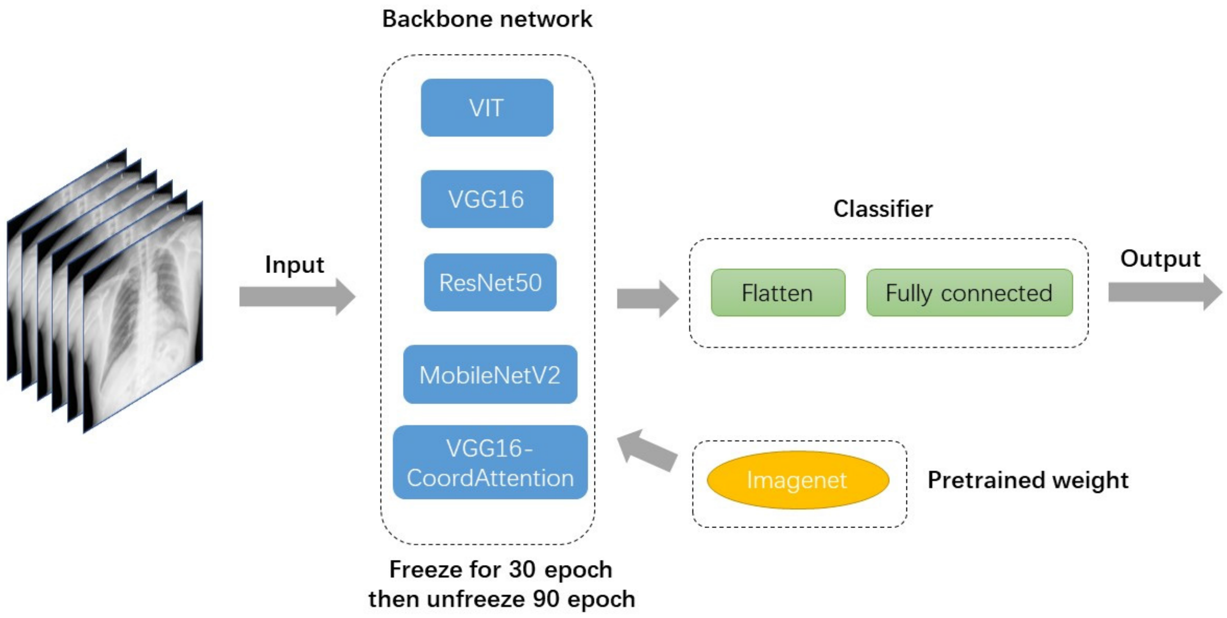

FIGURE 4. Freeze network training process.

on the network parameters in the later part of training, which can greatly improve the utilization of time and resources. At the same time, this can prevent the weight from being destroyed in the early stage of training, so that the model can converge quickly. Our experiments show that on our models, the training method of frozen network can also improve the classification accuracy.

In this study, we selected four representative network models and our model to evaluate the effect of freezing network training methods, Version-Transformer (VIT) [15] represents the network based on attention mechanism, VGG16 represents the traditional convolutional neural network, ResNet50 [16] represents the residual network, and MobileNetV2 [17] represents the lightweight network. The training process is shown in the Fig.4. The above five networks use Imagenet pretrained weight to initialize the model, and use freeze network and non-freeze network to train respectively. Five evaluation indexes, Top1-accuracy, area under ROC curve (AUC), recall(sensitivity), precision and F1 score, are used to evaluate the performance of different backbone networks. Where accuracy, recall(sensitivity), precision and F1 score are calculated as shown in formula (1) (2) (3) (4):

$$ACC = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}} \tag{1}$$

$$Recall(Sensitivity) = \frac{\mathbf{TP}}{\mathbf{TP + FN}} \tag{2}$$

$$Precision = \frac{\mathbf{TP}}{\mathbf{TP + FP}} \tag{3}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

where TP (True Positive) means predicting positive classes as positive numbers, TN (True Negative) means predicting negative classes as negative numbers, FP (False Positive) means predicting negative classes as positive numbers, and FN (False Negative) means predicting positive classes as negative numbers.

In the comparative experiments of the two groups, we run epoch for a total of 120 times for comparison. We freeze the backbone network parameters 30 times and unfreeze the network training 90 times, so as to evaluate the effect of this training method. The learning rate is set to 1e-3, the batch-size is 8, the adam optimizer and the gradient descent algorithm are used, and the overfitting is prevented by weight decay.

C. 5-FOLD CROSS VALIDATION

In order to verify the robustness of the model, we evaluate the model by 5-fold cross validation. We divide the dataset into five parts on average, taking four parts each time as the training set and one part as the verification set. Finally, we average the results of five times to verify the generalization ability of the model.

V. EXPERIMENTAL RESULS

A. RESULT OF TRANSFER LEARNING

Under five different backbone networks, we use the cross entropy loss function and the training method of freeze network to output the loss function curves of training set and validation set, as shown in Fig.5. We can find that no matter which model, in the early stage of thawing training after 30 epochs, the loss rate of training and verification set will increase significantly. This is because unfreeze makes the weight of the model trunk updated and iterated, and the fluctuations that occur in a short period of time occur,

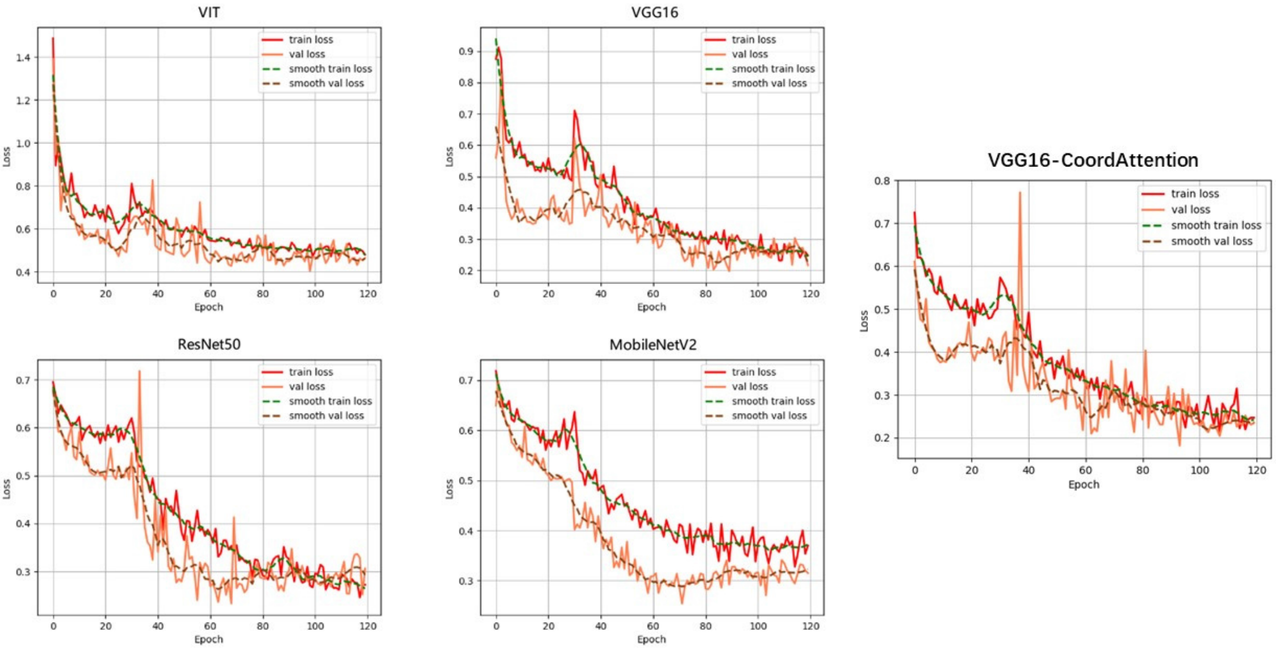

FIGURE 5. The loss curves of the training set and the validation set of the five backbone networks are frozen on the Shenzhen dataset(Red represents train loss, yellow represents val loss, green represents smooth train loss, and brown represents smooth val loss).

TABLE 1. Performance of five backbone networks under non freeze training(The bold ones represent our method).

	Top1-ACC(%)	AUC(%)	Recall(%)	Precision(%)	F1(%)
VIT	78.79	84.32	78.87	79.20	77.78
VGG16	81.82	88.77	82.02	84.29	79.31
ResNet50	89.39	95.66	89.35	89.56	89.86
MobileNetV2	88.64	96.49	88.76	89.66	87.80
Vgg16-CoordAttention	**85.61**	**93.16**	**85.61**	**85.61**	**85.71**

TABLE 2. Performance of five backbone networks under freeze training(The bold ones represent our method).

	Top1-ACC(%)	AUC(%)	Recall(%)	Precision(%)	F1(%)
VIT	80.30	90.56	80.46	81.68	78.33
VGG16	90.91	96.33	90.91	90.91	91.04
ResNet50	90.15	95.98	90.91	90.91	91.04
MobileNetV2	87.12	95.11	87.20	87.50	86.61
Vgg16-CoordAttention	**93.18**	**97.98**	**93.19**	**93.18**	**93.23**

after which the model can quickly converge and gradually stabilize.

B. SELECTION OF BACKBONE NETWORK

We conduct a comparative experiment by whether to use the freeze network to evaluate the effect of this training method. The results are shown in Table 1, 2. From the two tables, we can find that when using VIT, VGG16, ResNet50 and VGG16-CoordAttention as the backbone network, the training method of freezing network can improve the evaluation indexes of predicting classification problems to a certain extent, among which the improvement of traditional convolutional neural network VGG16 is the most obvious. Under this training mode, the classification effect of MobileNetV2 network decreases, which may be due to the fact that the lightweight network has few network parameters and the freeze network makes the model backbone unable to extract key image features.

TABLE 3. The results of 5-fold cross validation on Shenzhen dataset(The bold ones represent average result).

	Top1-ACC(%)	AUC(%)	Recall(%)	Precision(%)	F1(%)
1	93.18	97.98	93.19	93.18	93.23
2	91.67	97.08	91.63	91.77	91.97
3	92.42	97.59	92.47	92.55	92.31
4	94.70	97.96	94.71	95.45	94.74
5	91.67	97.93	91.65	91.18	91.85
AVG	**92.73**	**97.71**	**92.73**	**92.83**	**92.82**

TABLE 4. Compared with VGG16-CoordAttention and the existing methods on the Shenzhen dataset (The bold ones represent the best results).

	ACC(%)	AUC(%)
ConvNet[8]	87	87%
ResNet50+ AdaptiveDropout[9]	81.80	-
Fpn+Faster RCNN[10]	90.20	94.10
B-CNN[11]	86.46	-
EfficientNet+HEF[12]	89.92	94.80
LLR[13]	84.00	90
Our method	**92.73**	**97.71**

As can be seen from Table 2, in Shenzhen dataset, our method can achieve the highest accuracy by freezing network training. Although the accuracy of residual network ResNet50 is also high, from the loss function curve in Fig.5, the effect of the model on the dataset is not stable and fluctuates greatly. In the last few iterations, the loss rate of the verification set began to rise gradually, and there is the phenomenon of over fitting. So we finally choose VGG16 model as our backbone network.

C. ABLATION EXPERIMENT

On the Shenzhen dataset, ablation experiments were done to verify the effectiveness of our method. From Table 1, it can be found that without freezing the network, VGG16-CoordAttention compared with VGG16, due to the addition of coordinate attention, the model can better acquire the direction and position information of the image, which makes the classification effect of the algorithm improved. From Table 2, it can be found that by freezing the training method of the network, the backbone network retains the high-level semantic information of the pre-trained weights at the early stage of training, and devotes more resources to training the classifier, which makes the classification performance of VGG16-CoordAttention further improved. We used vgg16-CoordAttention model with freeze network to get the best results at present, with Top1-ACC: 93.18%, AUC: 97.98%, recall: 93.19%, precision: 93.18% and F1 score: 93.23%. The index of AUC is close to 98%. Fig.6 shows ROC curve on test set. The ROC curve shows the trade-off between recall and specificity [18]. AUC is considered to be an effective method to show the accuracy of ROC generated by each model, which proves that our model has excellent classification ability in tuberculosis task.

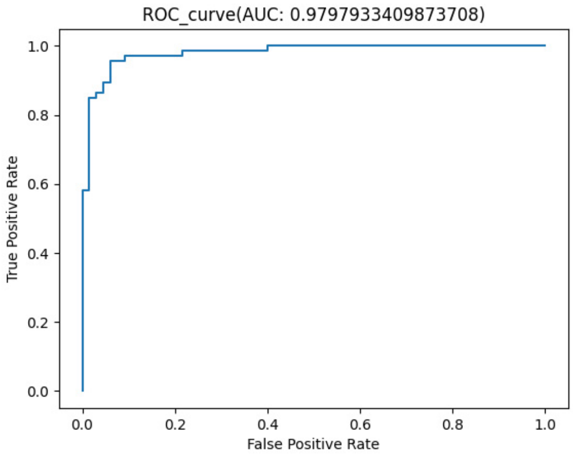

FIGURE 6. ROC curve on test set.

D. COMPARISON EXPERIMENTS

Our method has achieved excellent results in the classification of pulmonary tuberculosis in Shenzhen dataset. In order to verify the robustness and generalization ability of our model, we evaluated the model through 5-fold cross validation methods. The results of 5-fold cross validation are shown in Table 3. Our evaluation indexes are above 91%, which proves that our model has certain generalization ability.

We compare the average results of 5-fold cross validation with other existing end-to-end methods, and evaluate them with the accuracy and AUC indicators that are representative of classification problems. The comparison performance is shown in Table 4. The results show that our method has better results than other end-to-end methods on Shenzhen dataset.

VI. CONCLUSION AND DISCUSSION

In our research, we propose a model of adding coordinate attention mechanism module to convolutional neural network VGG16. Compared with the existing mainstream models, the addition of attention mechanism enables our method to better focus on the location and direction information in the tuberculosis image, so as to obtain better classification accuracy. At the same time, we use the method of freezing the network to speed up the model training process and further improve the performance of the network. Compared with the existing end-to-end methods, our method has better effect. Our method does not need to use ensemble learning for multi model fusion, nor does it need to consume huge computing resources. Compared with the traditional methods that use a large number of data preprocessing methods to improve performance, our method also avoids the preprocessing methods or data expansion of specific tasks. The results of five cross validation also show that our method has certain generalization ability. Besides, the classification accuracy, AUC, precision, recall and F1 score of our model in pulmonary tuberculosis detection are more than 91%. The results obtained by our method have better performance, which can help radiologists diagnose pulmonary tuberculosis images. In the follow-up research, we hope to achieve the same performance through the use of lightweight networks such as mobilenetv2, so that computing can also be carried out on mobile devices, which is further convenient for radiologists to assist in diagnosis.

REFERENCES

[1] K. R. Steingart, M. Henry, V. Ng, P. C. Hopewell, A. Ramsay, J. Cunningham, R. Urbanczik, M. Perkins, M. A. Aziz, and M. Pai, "Fluorescence versus conventional sputum smear microscopy for tuberculosis: A systematic review," *Lancet Infectious Diseases*, vol. 6, no. 9, pp. 570–581, Sep. 2006.

[2] *Chest Radiography in Tuberculosis Detection: Summary of Current WHO Recommendations and Guidance on Programmatic Approaches*, World Health Org., Geneva, Switzerland, 2016.

[3] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13713–13722.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[5] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156–3164.

[6] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1–12.

[7] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[8] E. Showkatian, M. Salehi, H. Ghaffari, R. Reiazi, and N. Sadighi, "Deep learning-based automatic detection of tuberculosis disease in chest X-ray images," *Polish J. Radiol.*, vol. 87, no. 1, pp. 118–124, Feb. 2022.

[9] Q. Zhang, C. Bai, Z. Liu, L. T. Yang, H. Yu, J. Zhao, and H. Yuan, "A GPU-based residual network for medical image classification in smart medicine," *Inf. Sci.*, vol. 536, pp. 91–100, Oct. 2020.

[10] Y. Xie, Z. Wu, X. Han, H. Wang, Y. Wu, L. Cui, J. Feng, Z. Zhu, and Z. Chen, "Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs," *J. Healthcare Eng.*, vol. 2020, pp. 1–12, Aug. 2020.

[11] Z. Ul Abideen, M. Ghafoor, K. Munir, M. Saqib, A. Ullah, T. Zia, S. A. Tariq, G. Ahmed, and A. Zahra, "Uncertainty assisted robust tuberculosis identification with Bayesian convolutional neural networks," *IEEE Access*, vol. 8, pp. 22812–22825, 2020.

[12] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image enhancement for tuberculosis detection using deep learning," *IEEE Access*, vol. 8, pp. 217897–217907, 2020.

[13] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y.-X. Wang, P.-X. Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 233–245, Feb. 2013.

[14] S. Jaeger, S. Candemir, S. Antani, Y. X. Wang, P. X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imag. Med. Surg.*, vol. 4, p. 475, Dec. 2014.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[18] D. L. Streiner and J. Cairney, "What's under the ROC? An introduction to receiver operating characteristics curves," *Can. J. Psychiatry*, vol. 52, no. 2, pp. 121–128, Feb. 2007.

TIANHAO XU was born in Hangzhou, Zhejiang, China, in 1999. He received the B.S. degree in computer science and technology from the Wenzhou University of Technology, China, in 2021. He is currently pursuing the M.S. degree in electronic information with Hangzhou Normal University, Hangzhou. His research interests include medical image processing and deep learning.

ZHENMING YUAN received the Ph.D. degree from the College of Computer Science and Technology, Zhejiang University. He is currently a Professor with the College of Information Science and Engineering, Hangzhou Normal University, and the Vice Dean of the Engineering Research Center of Intelligent Healthcare, Ministry of Education in China. His research interests include artificial intelligent in medical, machine learning, and big data mining.