Computer Modeling in Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2022.022322

ARTICLE

A New Childhood Pneumonia Diagnosis Method Based on Fine-Grained Convolutional Neural Network

Yang Zhang[1], Liru Qiu[2], Yongkai Zhu[1], Long Wen[1,*] and Xiaoping Luo[2,*]

[1]School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan, 430074, China

[2]Department of Pediatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430074, China

*Corresponding Authors: Long Wen. Email: wenlong@cug.edu.cn; Xiaoping Luo. Email: xpluo@tjh.tjmu.edu.cn

Received: 04 March 2022 Accepted: 12 May 2022

ABSTRACT

Pneumonia is part of the main diseases causing the death of children. It is generally diagnosed through chest X-ray images. With the development of Deep Learning (DL), the diagnosis of pneumonia based on DL has received extensive attention. However, due to the small difference between pneumonia and normal images, the performance of DL methods could be improved. This research proposes a new fine-grained Convolutional Neural Network (CNN) for children's pneumonia diagnosis (FG-CPD). Firstly, the fine-grained CNN classification which can handle the slight difference in images is investigated. To obtain the raw images from the real-world chest X-ray data, the YOLOv4 algorithm is trained to detect and position the chest part in the raw images. Secondly, a novel attention network is proposed, named SGNet, which integrates the spatial information and channel information of the images to locate the discriminative parts in the chest image for expanding the difference between pneumonia and normal images. Thirdly, the automatic data augmentation method is adopted to increase the diversity of the images and avoid the overfitting of FG-CPD. The FG-CPD has been tested on the public Chest X-ray 2017 dataset, and the results show that it has achieved great effect. Then, the FG-CPD is tested on the real chest X-ray images from children aged 3–12 years ago from Tongji Hospital. The results show that FG-CPD has achieved up to 96.91% accuracy, which can validate the potential of the FG-CPD.

KEYWORDS

Childhood pneumonia diagnosis; fine-grained classification; YOLOv4; attention network; Convolutional Neural Network (CNN)

1 Introduction

Pneumonia is a common lung disease caused by a variety of infectious sources [1], including viruses, bacteria, and fungi. It is one of the leading infectious diseases that cause the death of children all over the world [2]. In 2017, an estimated 808,694 children died from pneumonia, accounting for 15 percent of all deaths in children under five years old [3]. Pneumonia poses a significant threat to children's life. Chest X-ray examination is an important method to diagnose pneumonia. Usually, the chest X-ray image is analyzed by the radiologists. However, it has been found that the imaging changes

of recent viral pneumonia are rapid, and the manifestations of pneumonia are varied [4]. Because of the low Kv used in children to reduce radiation injury, the image quality of children's chest X-rays is less than adults. The interstitial part of the lungs in children is prosperous, the volume of the chest cavity in children is small, and the chest X-ray shows many cardiovascular vessels, which can easily cover up the lesions. Therefore, it is necessary to develop a more robust diagnostic method.

With the rapid development of Deep Learning (DL), the DL-assisted pneumonia diagnosis has been widely researched [5]. As one of the most powerful kinds of DL, Convolutional Neural Network (CNN) can handle the raw images directly and extract the feature of images for further diagnosis [6,7], and it has been widely used in the field of medical image [8]. As shown in Fig. 1, there is very little difference between abnormal sample 1 and normal sample 1, and the difference between normal sample 1 and normal sample 2 is very obvious, which is consistent with the fine-grained characteristics with small differences between classes and large intra-class differences. Therefore, it can be determined that chest X-ray images of children with pneumonia have typically fine-grained characteristics. There is little difference in chest X-ray images between normal people and patients with pneumonia, and the chest X-ray images of patients with pneumonia are diverse [4], which leads to inefficient and unreliable diagnosis methods based on CNN. Therefore, it is of great significance to develop more powerful diagnostic methods to improve the performance of CNNs on children's chest X-ray images.

Figure 1: Children chest X-ray images, from left to right, are abnormal sample 1, normal sample 1, and normal sample 2

Fine-grained (FG) classification is a new paradigm in this field. The spirit of the fine-grained classification is to find the local information in images to expand the small differences between classes. Therefore, fine-grained classification can learn both the local and global features of the images, which are suitable for pneumonia diagnosis using chest X-ray images. As shown in Fig. 2, the local information of the chest X-ray image is located by using an attention mechanism. The attention mechanism finds the discriminative parts denoted as the attention maps and then fuses the attention maps and the raw images to locate the local parts denoted as the attention thermal map. As the attention mechanism can judge the significant parts in the images, the FG method can pay attention to these discriminative parts to train the CNN classifier while ignoring other irrelevant information and promote the classification accuracy of CNN models [9].

In this research, a novel childhood Pneumonia Diagnosis based on a Fine-Grained Convolutional Neural Network (FG-CPD) is proposed. Firstly, since the original chest X-ray images of children have many sparse spaces, the YOLOv4 algorithm is adopted to localize and crop the children's chest X-ray images before introducing the CNN model. Secondly, a novel attention network (SGNet) is used to locate and distinguish local information, and the attention-guided data augmentation is used to learn these distinguished parts, and the feature maps are fused with the attention maps based on feature fusion to improve the performance of FG-CPD. Then, it was further conducted on real chest X-ray

images from children aged 3–12 years old between August 2020 and May 2021 from Tongji Hospital in Tongji Medical College of Huazhong University of Technology and Science. The results show that FG-CPD has achieved good results.

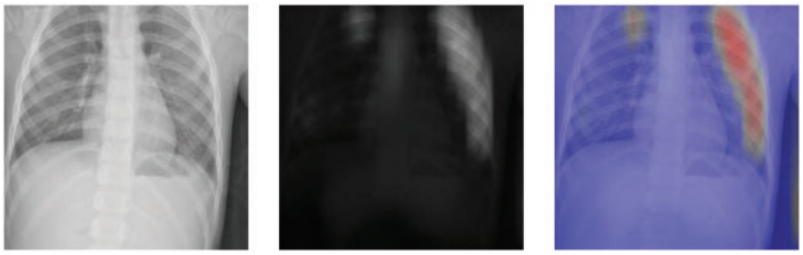

Figure 2: The images from the Chest X-ray 2017 dataset, from left to right, are the original raw image, attention map, and attention thermal map

The rest of this research is as follows. Section 2 presents the related work. Section 3 presents the paradigm of fine-grained CNN. Section 4 explains the methodologies of the FG-CPD. Section 5 presents the case studies and the experimental results. Section 6 concludes and gives future research directions.

2 Related Work

2.1 Convolutional Neural Network Based Pneumonia Diagnosis

CNN has been extensively used in various classifications [10,11]. Most of the existing DL pneumonia diagnostic methods are based on CNN. Sedik et al. [12] proposed a deep learning module based on a Convolutional Neural Network (CNN) and convolutional long-term memory. The experimental result showed that in some cases, the accuracy of pneumonia diagnosis reaches 100%. Yu et al. [13] researched a novel pneumonia diagnosis by the ResGNet framework, and the results showed that the average accuracy of ResGNet-C was up to 96.62%. Ahsan et al. [14] investigated a novel pneumonia diagnosis module using multilayer perceptron and MLP-CNN. The experimental results showed that the overall accuracy of the model reached 94.6%, which was better than existing models. Jaiswal et al. [15] proposed a mask-RCNN model with connection context structure by using chest X-ray images of pneumonia. Rahman et al. [16] used four different pre-trained CNN for migration learning to diagnose bacterial and viral pneumonia and achieved 93.3% accuracy in the diagnosis of mixed pneumonia. Nam et al. [17] studied the diagnosis of malignant pulmonary nodules on chest film on the CNN network. The experimental results showed that the automatic diagnosis algorithm based on deep learning is better than doctors in medical image classification and improves the performance of doctors as second readers. Rajpurkar et al. [18] proposed a CNN model CheXNeXt for detecting a variety of diseases, including pneumonia, the algorithm classifies clinically important chest X-ray abnormalities, and its performance level corresponds to license radiologists. Pham et al. [19] proposed a supervised multi-label classification framework based on CNN, which uses the hierarchical dependence between abnormal labels to improve the diagnosis accuracy of chest X-ray images. Sun et al. [20] proposed a novel framework of BEVGG for diagnosing COVID-19 through chest X-ray images and used a biogeography-based optimization method to optimize hyperparameter values of convolutional neural networks, and the experimental results showed that the framework performs higher than the current methods. Gayathri et al. [21] proposed a computer-assisted diagnosis method using chest X-ray images, sparse autoloader, and feedforward neural network FFNN, which reached high results in diagnosing COVID-19. Showkat et al. [22] evaluated the performance of

the ResNet model to classify pneumonia cases from CXR images, built a custom ResNet model, and evaluated its contribution to performance improvement. Zhang et al. [23] proposed a new pneumonia diagnosis network, which uses the hybrid attention module (CBAM) to obtain more attention information, making the model have better performance in pneumonia diagnosis.

Although several previous studies combined CNN pneumonia diagnosis methods, there are few studies on the diagnosis of pneumonia in children. Kermany et al. [24] collected and labeled a total of 5232 chest X-ray images from children for pneumonia diagnosis. The improved CNN model for pneumonia diagnosis proposed has obtained an accuracy of 92.8%. Then, several researchers follow this work. Yi et al. [25] studied a scalable and interpretable deep convolutional neural network to diagnose pneumonia using chest X-ray images. Chouhan et al. [26] proposed a new CNN model for pneumonia diagnosis based on transfer learning, which improves diagnosis accuracy by integrating multiple pre-training models. Hou et al. [27] proposed a six-layer convolutional neural network combining maximum pooling, batch normalization, and an Adam algorithm that outperforms several state-of-the-art methods with a 10-weight cross-validation method. In this research, the FG-CPD is developed for the children's pneumonia diagnosis. It tests in the Tongji Hospital dataset and the results validate its performance.

2.2 Fine-Grained Convolutional Neural Network Classification

Fine-grained classification has been investigated in image classification. Fu et al. [28] proposed a fine-grained classification method based on a cyclic neural network. The experiments are shown that it is more efficient than a single classification network. Ju et al. [29] proposed a multi-task learning framework named GAPNet, which has improved the accuracy of fine-grained classification by integrating multi-scale and multi-site features. Wang et al. [30] proposed a new weakly supervised fine-grained classification method, which used a gaussian mixture model to locate key regions and finally improved the accuracy of fine-grained classification. The attention mechanism in FG is used to find discriminative parts for the FG classification. Xiao et al. [31] proposed a two-level attention algorithm. It was the first to use an attention mechanism to achieve fine-grained classification. Zhang et al. [32] proposed a weakly supervised fine-grained classification method based on component image representation, which can reach high accuracy just by using image category labels. Zhao et al. [33] proposed a diversified visual attention model to improve classification accuracy by increasing the variety of visual attention. Sun et al. [34] proposed a multi-attention and multi-category constraint model to improve the computational efficiency of the model through the correlation of the discriminant inner region. Zheng et al. [35] proposed a Trilinear Attention Sampling Network (TASN) to achieve fine-grained classification. Ding et al. [36] proposed a fine-grained classification model based on the attention pyramid CNN, and model performance was optimized by enhancing the discriminating regions and eliminating background noise.

In this research, a new model FG-CPD is proposed, which uses the SGNet to expand the search for the discriminative part. In FG-CPD, the raw chest X-ray is used directly, so the YOLOv4 is used to detect and trim the position of the chest on the raw chest X-ray before feeding into the FG classification.

3 The Fine-Grained Convolutional Neural Network

This section presents the Fine-Grained Convolutional Neural Network (FG-CNN) and the detailed descriptions of its three main components.

3.1 The Architecture of FG-CNN

FG-CNN architecture is shown in Fig. 3, and it mainly includes three parts. 1) Feature extraction part. The feature extraction part usually uses a CNN network to extract the features of the images and obtain the image feature map and attention map. 2) Feature fusion part. In this part, Bilinear Attention Pooling (BAP) is used to fuse the attention maps and feature maps, and then the fused results would feed to the fully connected (FC) layer for the FG classification. 3) Attention-guided data augmentation part. Obtained attention maps can reflect the significant parts in the images, so they can be used to guide the data augment to improve the performance of FG-CNN.

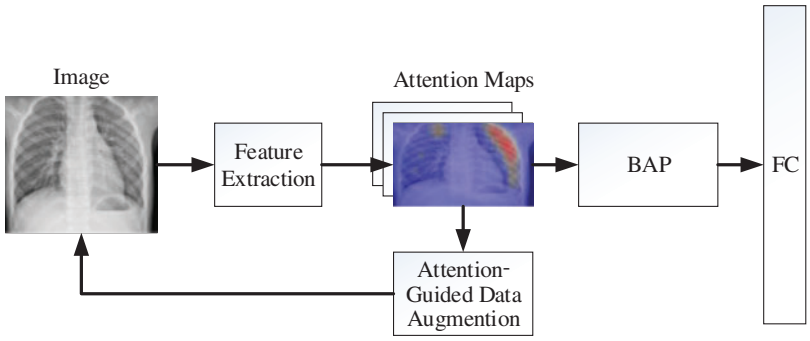

Figure 3: The overall architecture of the FG-CNN

3.2 Feature Extraction Network

In this research, the Inception V3 network is used as a feature extraction network. The Inception V3 is firstly pre-trained on the ImageNet before handling the raw chest X-ray images. Suppose that the feature map obtained by Inception V3 is denoted as $F \in R^{C\times H\times W}$, where C, H, and W represent the number of channels, width, and height of the features respectively.

The attention maps are based on the feature maps. It passes the data features using a convolution operation with a convolution kernel size of 1. Suppose that $A \in R^{M\times H\times W}$ denotes an attention map where M is the number of the attention map. $f(\cdot)$ is the convolution function. The attention map can be expressed as Eq. (1).

$$f(F) = A \tag{1}$$

3.3 BAP Fusion

BAP fusion is to combine the attention maps and feature maps. As attention maps are to find the discriminative parts in the images, so fusing the attention maps and feature maps can strengthen the feature. BAP fusion is the element-wise production of the attention map and feature map which can be denoted in Eq. (2). F is the feature map and A_M is the attention map.

$$P = \begin{Bmatrix} g(F \cdot A_1) \\ g(F \cdot A_2) \\ \cdots \\ g(F \cdot A_M) \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \\ \cdots \\ f_M \end{Bmatrix} \tag{2}$$

In the BAP fusion, the features are multiplied with each layer of attention map element-wise, and then the Global Maximum Pool (GMP) is used for the dimension reduction. Then the fusion feature $P \in R^{M\times N}$ can be obtained for the classification.

3.4 Attention-Guided Data Augmentation

The attention-guide data augmentation has two sub-parts, the augmentation cropping, and the augmentation dropping. Both data augmentation methods are based on attention maps. As shown in Eq. (3), one layer $k \in [1, M]$ in the attention maps is randomly selected and normalized. Then, the augmentation cropping and the augmentation dropping are conducted.

$$A_k^* = \frac{A_k - \min(A_k)}{\max(A_k) - \min(A_k)} \tag{3}$$

1) Augmentation Cropping: The crop mask C_k is generated with the threshold θ_c. When the element in A_k^* is greater than θ_c, then the element in crop mask is set as 1, otherwise to be 0.

$$C_k(i,j) = \begin{cases} 1, & if\ A_k^*(i,j) > \theta_c \\ 0, & otherwise \end{cases} \tag{4}$$

By using the crop mask C_k, the discriminative parts in the images are cropped from the original images, and then those discriminative parts are scaled up to the size of the raw images. Then, these cropped images, as the data augmentation, would be used to train FG-CNN.

2) Augmentation Dropping: The augmentation dropping is the opposite of the augmentation cropping. In augmentation dropping, the drop mask D_k sets the elements to be 0 when the element in A_k^* is greater than θ_d, otherwise to be 1. This will encourage the network to extract other identifiable parts, and the robustness of classification and the accuracy of positioning will be improved.

$$D_k(i,j) = \begin{cases} 0, & if\ A_k^*(i,j) > \theta_d \\ 1, & otherwise \end{cases} \tag{5}$$

Attention-guided data augmentation is to input the original images into the network to obtain the corresponding feature maps and attention maps, combine the two parts of the images, adopt attention cropping and attention dropping, and return the images to the network for training to achieve the effect of data augmentation. The process can be divided into two parts, the first part uses the attention-guided data augmentation to enhance the images, and the second part trains the network using a combination of raw and augmented images.

4 The Fine-Grained Children Pneumonia Diagnosis

This section presents the proposed Fine-Grained Children Pneumonia Diagnosis (FG-CPD), and the mythologies of FG-CPD are shown.

4.1 The Structure of the Proposed FG-CPD

The structure of FG-CPD is shown in Fig. 4. Compared with FG-CNN, the FG-CPD model is improved in two aspects.

Firstly, the data preprocessing part is added which aims to handle the raw chest X-ray directly. In this process, the raw chest X-ray is converted from the DICOM medical data format to JPG format. Then, the YOLOv4 algorithm is trained to detect the position of the chest X-ray in the converted images. Finally, automatic data augmentation is used on the chest X-ray before feeding the FG-CNN network.

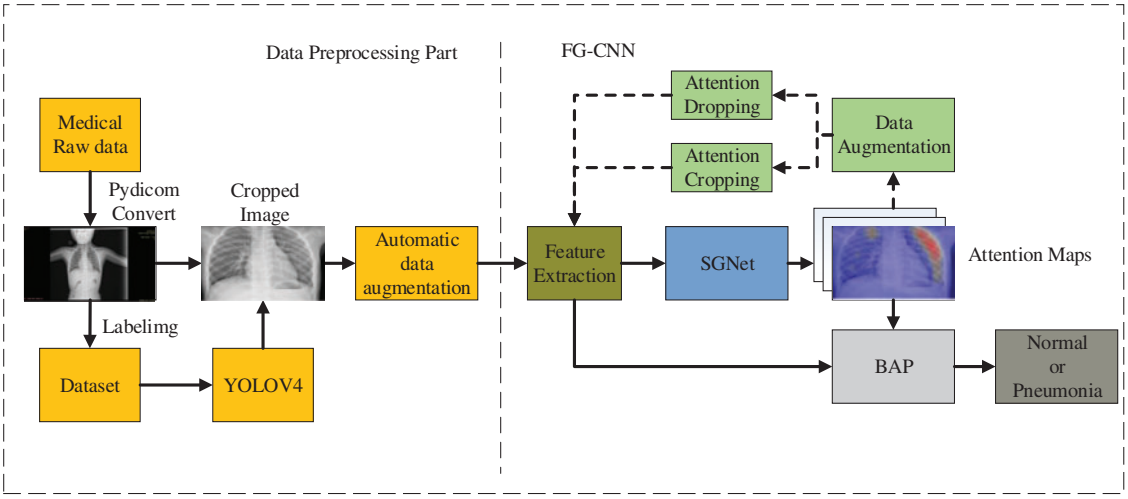

Figure 4: The overall structure of the proposed FG-CPD

Secondly, a new attention network, SGNet, is proposed to improve the representation ability of attention maps. SGNet is an improved attention network that can perform $1 * 1$ convolution operation based on channel attention maps, fuse information between different channels, and generate hybrid attention maps to improve its characterization ability and model performance.

4.2 Data Preprocessing Part

4.2.1 Data Conversion

As the raw chest X-ray data is stored in the database using DICOM format, so they are converted into the JPG format by using the Pydicom library. Fig. 5 shows some converted JPG images of the chest X-ray images.

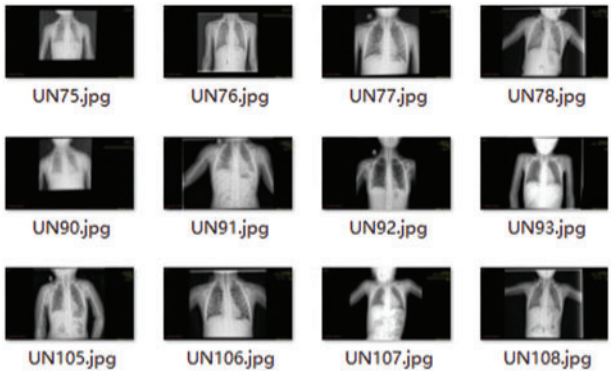

Figure 5: The converted chest X-ray images

4.2.2 The Chest Detection and Positioning using YOLOv4

As the raw images of the chest X-ray contain a lot of spare space, it can interfere with the classification of FG-CPD. Therefore, the YOLOv4 algorithm was used to detect and locate the chest region before using FG-CPD. YOLO series is the representative object detection algorithm, which has

attracted remarkable attention in both academia and industry. YOLOv4 is the fourth version of the YOLO series with the advantages of fast detection speed and high detection accuracy, so the YOLOv4 is used to detect and position the chest in the raw images.

As shown in Fig. 6, the structure of YOLOv4 consists of three parts: feature extraction network, SPP module and PANet feature pyramid network, and YOLO Head prediction network. It should be pointed out that this feature extraction network in YOVOv4 has nothing to do with the feature extraction network in FG-CPD. Firstly, the raw chest X-ray images were fed into a feature extraction network to extract the features of the images. Then, the features are handled by the SPP module and PANet for feature fusion. Finally, the fused features are sent into the YOLO Head network to detect the position of the chest in the raw chest X-ray images.

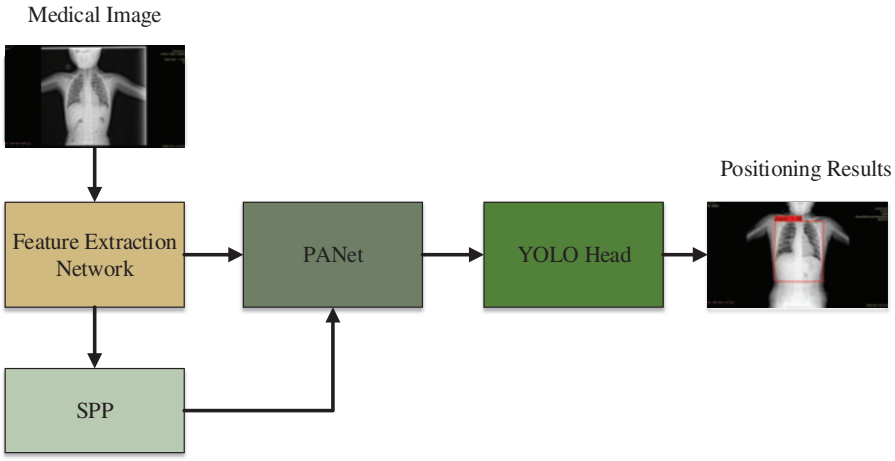

Figure 6: The structure of YOLOv4

During the training process of YOLOv4, 200 images were first annotated with labeling annotation software, as shown in Fig. 7. These 200 images are used as a training dataset to train YOLOv4. The training *mAP* for the chest X-ray of the YOLOv4 has reached 100% accuracy. Then, the chest region detection by YOLOv4 would be located and cropped for further FG-CPD. Several cropped images using YOLOv4 are shown in Fig. 8.

4.2.3 Automatic Data Augmentation

Data augmentation can increase the sample of the training set, effectively alleviate the model overfitting situation and bring stronger generalization ability to the model. After cropping the chest X-ray images, the images were augmented using automated data augmentation. In this research, the augment policy [37], which is trained on the ImageNet, is used for automatic data augmentation. By increasing the data, the diversity of samples, and the performance of the model for the diagnosis of children with pneumonia. Several samples after the automatic data augmentation are shown in Fig. 9.

Figure 7: Labeling annotated for the raw chest X-ray images

Figure 8: The cropped images using YOLOv4 after detection

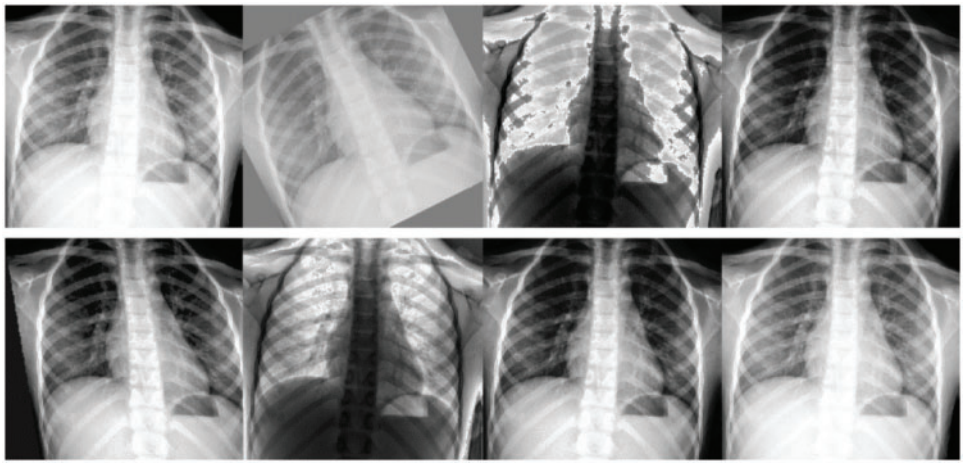

Figure 9: Several samples with automatic data augmentation

4.3 Attention Network

In FG-CPD, the SGNet is used to replace the attention network in FG-CNN. The structure of the proposed SGNet attention network is shown in Fig. 10. Firstly, the image feature maps F are sent to the channel attention branch to generate a channel attention map. The channel attention map contains the Global Average Pooling (GAP) and two fully-connected (FC) layers with different activation functions. Then, the channel attention branch is merged with the feature maps F using element-wise multiplication to generate channel attention maps $A_s \in R^{C\times H\times W}$.

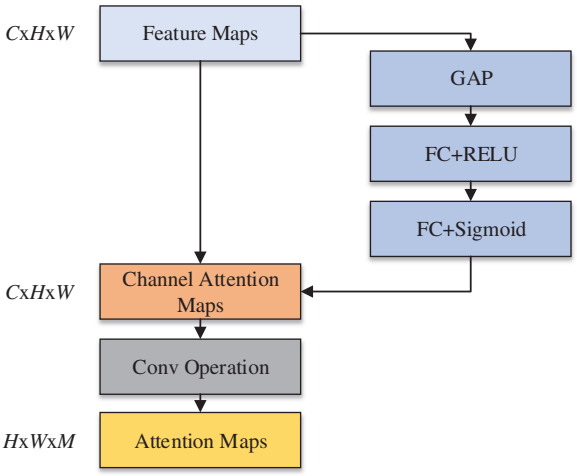

Figure 10: The overall architecture of the SGNet

The dimension of the channel attention maps is the same as the images. Then, the $1 * 1$ convolution operation is followed by the channel attention maps. Therefore, the final attention maps $A \in R^{M\times H\times W}$ of the mixed channel domain can be formulated as Eq. (6).

$$f(A_s) = A \tag{6}$$

4.4 Attention Regularization

In the training of FG-CPD, the loss used is the cross-entropy loss function. To make the attention maps A more centralized, attention regularization is added to the loss of FG-CPD.

The attention regularization is inspired by center loss. It forces attention maps to converse with virtual centers, which can make them more compact. The k-th global feature center is denoted as C_k and the loss of the attention regularization is shown in Eq. (7). The C_k is initialized from 0 and updated by moving average, as shown in Eq. (8), where β is the update factor and f_k is the feature matrix.

$$L_A = \sum_{k}^{M} ||f_k - C_k||_2^2 \tag{7}$$

$$C_k + \beta(f_k - C_k) \rightarrow C_k \tag{8}$$

$$L = L_A + 1/3(L_{\text{cross\_raw}} + L_{\text{cross\_crop}} + L_{\text{cross\_drop}}) \tag{9}$$

Specifically, as shown in Eq. (9), L is the total loss function of the network, Lcross_raw is the cross-entropy loss function of the original image, Lcross_crop is the cross-entropy loss function with augmentation cropping in attention-guided data augmentation, and Lcross_drop is the cross-entropy loss function with augmentation dropping in attention-guided data augmentation.

5 Case Studies and Experimental Results

In this section, the proposed FG-CPD is tested on the public Chest X-ray 2017 dataset and real-world chest X-ray images. The Chest X-ray 2017 dataset is a well-known benchmark dataset and the real-world chest X-ray images are from Tongji Hospital.

In the following experiments, the Inception V3 network which has been pre-trained on ImageNet is selected as feature extraction network, random gradient descent (SGD) as optimizer, cross-entropy loss function as finally loss function, learning rate 0.001, momentum value 0.9, training epoch 100 times, weight attenuation impairment value 1e-5 and batch size 4, using a CPU of Intel Core i9-9900X 3.5 GHz $*$ 20, the GPU is RTX2080Ti, and the Python version is 3.6.4.

To test the performance of the FG-CPD network, four indicators including accuracy, precision, recall rate, and F1 are used to verify the potential of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{13}$$

where TP is the number of correctly classified pneumonia samples, TN is the number of correctly classified normal samples, FP is the number of incorrectly classified pneumonia samples, and FN is the number of incorrectly classified normal samples. This research applied the 8-fold cross-validation to validate the performance.

5.1 Case Study 1: Chest X-Ray 2017

5.1.1 Dataset Description

The Chest X-ray 2017 dataset was provided by Guangzhou Women and Children's Medical Center and has been widely used in pneumonia diagnosis. In this dataset, a total of 5826 chest X-ray images were collected and labeled. The training set consisted of 5232 images (1346 images belong to the normal and 3883 images belong to pneumonia). The remaining 624 images were used to test the FG-CPD performance. Several raw images of the dataset are shown in Fig. 11. The dataset is available at the following: https://data.mendeley.com/datasets/rscbjbr9sj/3.

Figure 11: Raw images of the Chest X-ray 2017 dataset, from left to right, are the bacterial image, viral image, and normal image

5.1.2 Experimental Results and Comparison

To compare the performance of FG-CPD, different kinds of backbones are researched as feature extraction networks. Experiments on the Chest X-ray 2017 dataset were performed 8 times. Except for Inception v3, Resnet34, Resnet50, Resnet101, and VGG19 are used as backbones. The denotation FG-CPD-Inception v3 means that FG-CPD uses Inception v3 as its backbone. The FG-CPD is also compared with the CNN models without the FG part, and the selected CNN models include Resnet34, Resnet50, and Resnet101. The above networks are pre-trained on the ImageNet dataset. The comparison results are presented in Table 1.

Table 1: Comparison results of FG-CPD with different backbones and CNN models on case 1

Methods	Accuracy (%)
FG-CPD-Inception v3	**100**
FG-CPD-Resnet34	98.30
FG-CPD-Resnet50	98.09
FG-CPD-Resnet101	98.76
FG-CPD-VGG19	98.24
Resnet34	96.95
Resnet50	96.63
Resnet101	96.75
Inception v3	98.00

From Table 1, it can be seen that the accuracy of FG-CPD-Inception v3 has achieved remarkable performance and it is the accuracy is 100%. FG-CPD methods using other CNN as backbones have achieved 98.30%, 98.09%, 98.76%, and 98.24%, showing that FG-CPD-Inception v3 is superior to other FG-CPD variants.

Resnet34, Resnet50, and Resnet101 are different variants in Resnet, and they are widely used in image classification. In this research, they are used as the baseline models. From the results, it can be seen that these three models have achieved 96.95%, 96.63%, and 96.75%, respectively, which means that FG-CPD-Inception v3 is superior outperform to these classic CNN models.

Through the above experiments, it is not difficult to see that compared with other CNNs, the accuracy rate of Inception v3 alone is up to 98%, which proves that the feature extraction ability is the strongest in the Chest X-ray 2017 dataset. Therefore, in FG-CPD, Inception v3 is used as a feature extraction module to obtain feature information of chest X-ray images, which is conducive to the accurate diagnosis of pneumonia.

5.1.3 Analysis for FG-CPD

In this subsection, several key components in FG-CPD are analyzed, including the number of the attention map M, the attention-guided data augmentation, and the automatic data augmentation. As the images have been cropped well, the YOLOv4-based chest detection and position have been ignored in this case.

Fig. 12a presents the effect of the value of M on the performance of FG-CPD. It can be seen that M can influence the final accuracy of FG-CPD. When the M value is too low, such as set to be 16, the performance of FG-CPD degrades sharply. But when its value is higher than 32, the effect has been reduced. The performance of FG-CPD shows to be stable the value of M at 32, 48, 64, and 80. The best performance of FG-CPD occurs at the $M = 64$.

To show the effect of the attention-guided data augmentation, four configurations, which are denoted as FG-CPD-o, FG-CPD-w-C, FG-CPD-w-D, and FG-CPD-wCD, are investigated. FG-CPD-o is the configuration of FG-CPD without both attention-guided cropping and dropping. FG-CPD-w-C/D means that FG-CPD only with attention-guided cropping or dropping, and FG-CPD-wCD is that FG-CPD with both attention-guided cropping and dropping. The results are shown in Fig. 12b. It can be seen that the attention-guided cropping and dropping can improve the performance of FG-CPD, and both of them can be used together to improve FG-CPD.

On the automatic data augmentation aspect, the FG-CPD with or without the automatic data augmentation is presented. The random data augmentation is used as the baseline to show the effectiveness of the automatic data augmentation. The results are shown in Fig. 12c. Automatic data augmentation can significantly improve the performance of FG-CPD by expanding the dataset samples.

Fig. 13 gives four attention maps of this case study. From the figures, it can be seen that the subtle differences in pneumonia X-ray chest ray can be accurately found with the FG-CPD. This indicates that FG-CPD has an excellent performance in fine-grained image classification and validates the potential of the proposed FG-CPD.

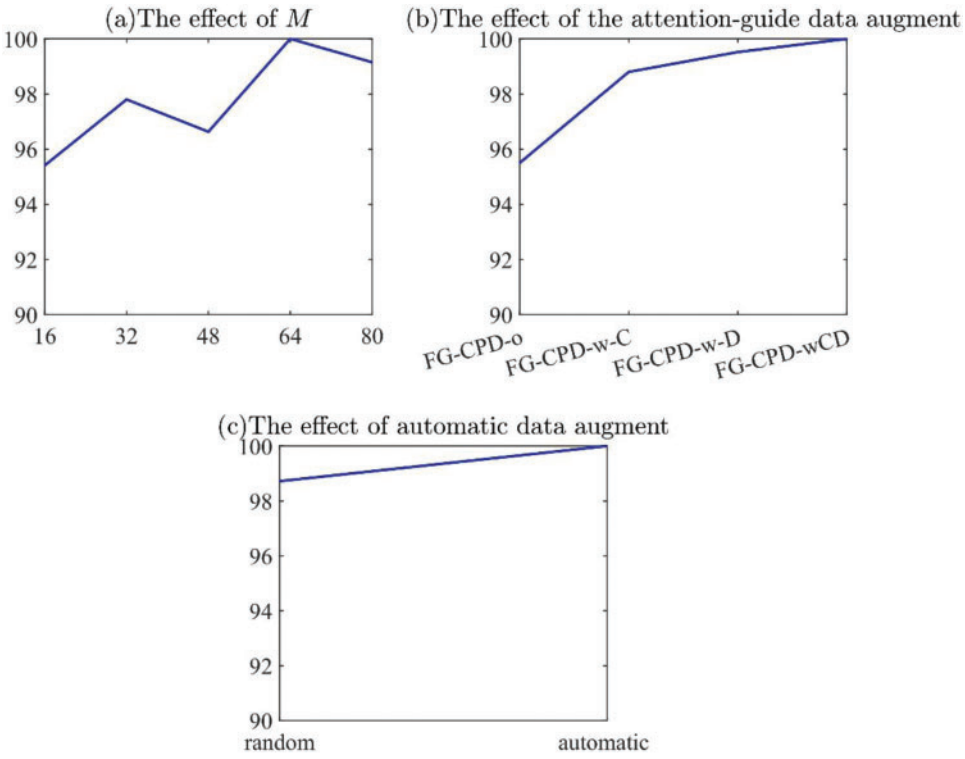

Figure 12: The effect of three key components in FG-CPD in case 1

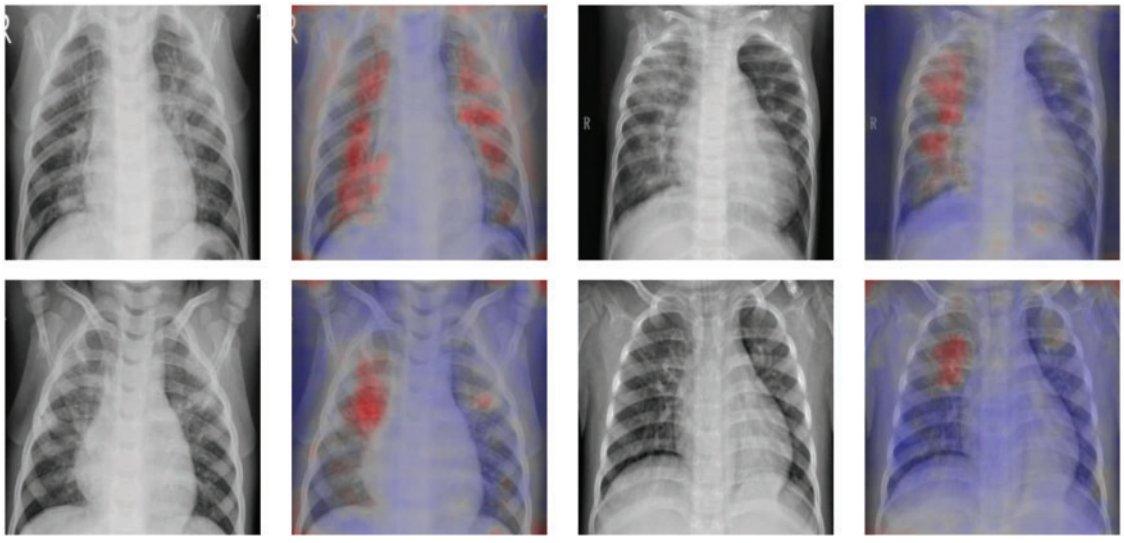

Figure 13: Four attention maps obtained by FG-CPD in case 1

5.1.4 Comparison with Other Published DL Methods

In this subsection, the FG-CPD is compared with other DL methods which are published in recent years. The comparison methods contain image-based deep learning (IMDT) [24], Transfer

learning-based approach for pneumonia detection (TLPD) [26], Deep learning to detect and evaluate pneumonia (DL-DEP) [38], Childhood pneumonia using convolutional neural networks (CP-CNN) [39], Deep convolutional neural network (DCNN) [25] and Adaptive median filter convolutional neural network recognition model based on random forest (ACNN-RF) [40], and the comparison results are shown in Table 2.

Table 2: Comparison with other published DL methods in case 1

Methods	Accuracy (%)
IMDT	92.80
TLPD	96.39
DL-DEP	92.57
CP-CNN	95.30
DCNN	96.09
ACNN-RF	96.90
FG-CPD	**100**

It can be seen that the proposed FG-CPD has achieved the best result among these DL methods. The prediction accuracy of FG-CPD is 100%, while that of other methods such as IMDT, TLPD, DL-DEP, CP-CNN, DCNN, and ACNN-RF is 92.80%, 96.39%, 92.57%, 95.30%, 96.09%, and 96.90%. The results show that compared with other advanced CNN or DCNN methods in fine-grained image classification. FG-CPD, with its excellent feature extraction, and feature fusion module, can identify subtle differences in images and is effective in diagnosing chest X-ray pneumonia.

5.1.5 Comparison with Other Attention Modules

To verify the superiority of the SGNet attention module proposed in this research, the attention module was replaced by Channel Attention Module, Spatial Attention Module, and Hybrid Attention Module for a control experiment, The channel attention module used is Squeeze-and-Excitation Networks (SENet) [41], the spatial attention module used is Spatial Transformer Networks (STN) [42], and the hybrid attention module used is Convolutional Block Attention Module (CBAM) [43]. These three attention networks are used in image recognition and classification widely. Experimental results are shown in Table 3. The accuracy of the Channel Attention Module used was 98.6%, the accuracy of the Spatial Attention Module used was 98.6%, and the accuracy of the Hybrid Attention Module used was 98.804%. SGNet has a superior performance compared to other attention modules.

Table 3: Comparison with another attention module in case 1

Methods	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)
FG-CPD + SENet	98.60	99.18	99.07	99.12
FG-CPD + STN	98.32	98.60	99.29	98.94
FG-CPD + CBAM	98.80	99.65	98.73	99.19
FG-CPD + SGNet	**100**	**100**	**100**	**100**

In addition, three performance indicators (Precision, Recall, F1) of FG-CPD on the Chest X-ray 2017 dataset are 100%, proving that FG-CPD has excellent performance and great potential in fine-grained classification.

5.2 Case Study 2: Chest X-Ray Dataset from Tongji Hospital

5.2.1 Dataset Description

This dataset is a real-world dataset provided by Tongji Hospital in Tongji Medical College of Huazhong University of Technology and Science. It is a chest X-ray dataset for children with or without pneumonia. The chest X-ray dataset includes 7251 images from 3–12 years old children, and it was collected between August 2020 and May 2021. The chest X-ray has been diagnosed by professional doctors, and there are 4726 samples with pneumonia and 2555 samples without pneumonia. During the training process of FG-CPD, the training dataset and testing dataset are randomly divided with a proportion of 8:2. Therefore, there are 5825 images in the training dataset and 1426 images in the testing dataset. The detail of the division of the training and testing dataset can be found in Table 4. Several raw chest images (after being converted from DICOM to JPG format) are shown in Fig. 5.

Table 4: Tongji Hospital chest X-ray dataset for training and testing dataset

Category	Training set (No. of images)	Test set (No. of images)
Normal	2065	490
Pneumonia	3760	966
Total	5825	1426

5.2.2 Experimental Results and Comparison

The comparison of FG-CPD using different backbones and other baseline DL methods is presented in Table 5. Experiments on the Chest X-ray Dataset from Tongji Hospital were performed 8 times. The above networks are pre-trained on the ImageNet dataset.

Table 5: Comparative results of FG-CPD with different backbone and CNN models in case 2

Methods	Accuracy (%)
FG-CPD-Inception v3	**96.91**
FG-CPD-Resnet34	79.30
FG-CPD-Resnet50	80.09
FG-CPD-Resnet101	81.76
FG-CPD-VGG19	80.24
Resnet34	78.32
Resnet50	78.84
Resnet101	79.19
Inception v3	75.00

The results show that the FG-CPD-Inception v3 has achieved the most superior accuracy. Its accuracy is around 96% with a little fluctuation, and the specific accuracy is 96.91%, 97.26%, 96.91%,

97%, 96.6%, 96.6%, 97.26%, and 96.91%, which outperforms FG-CPD-Resnet34, FG-CPD-Resnet50, FG-CPD-Resnet101, and FG-CPD-VGG19 a lot. Table 5 also presents the comparison of FG-CPD with other DL methods which are directly used without the FG parts. It can be seen that FG-CPD is superior to Resnet34, Resnet50, Resnet101, and Inception v3.

Through comparison, it is found that in the Chest X-ray Dataset from Tongji Hospital, Inception v3 still plays a good feature extraction effect, indicating that FG-CPD-Inception v3 has excellent pneumonia diagnosis effect and generalization performance.

5.2.3 Analysis of FG-CPD

In this subsection, four key components in FG-CPD are analyzed. They are the effect of YOLOv4 cropping, the number of attention map M, the attention-guided data augmentation, and the automatic data augmentation. Their effects on FG-CPD have been shown in Fig. 14.

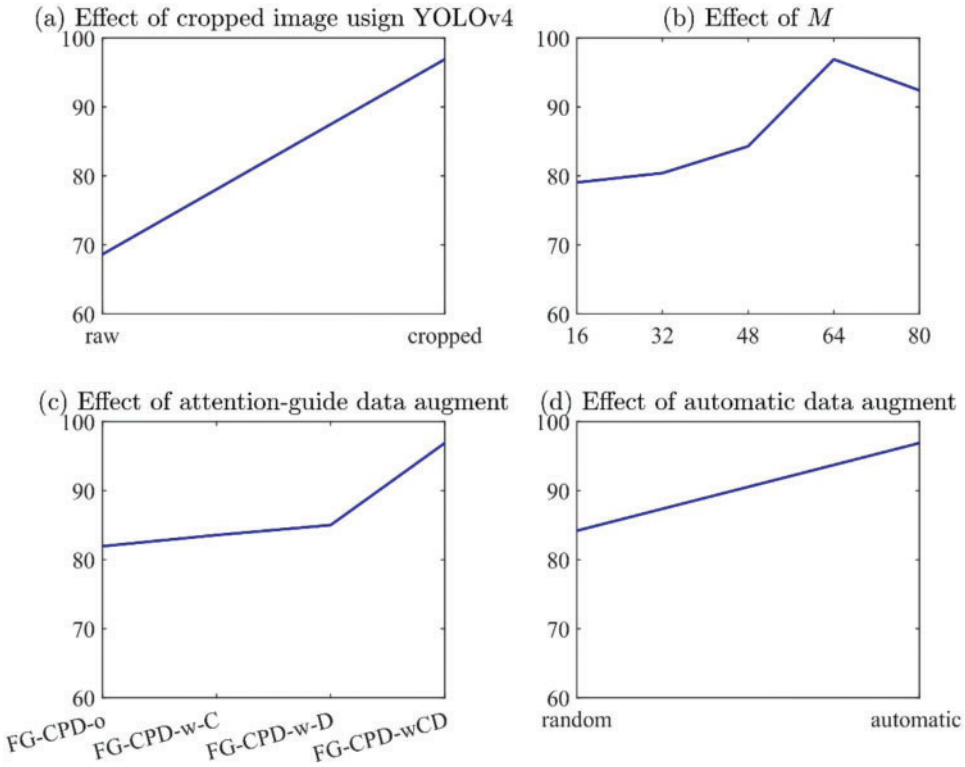

Figure 14: The effect of four key components in FG-CPD in case 2

Fig. 14a shows the effect of YOLOv4 on the FG-CPD. From the result, it can be seen that FG-CPD using raw images have an accuracy of 68.63%, which is obviously inferior to the result of FG-CPD with cropping using YOLOv4. Because the cropping can remove the irrelevant part of the chest images and can ensure the FG-CPD learns more discriminative information from the chest X-ray images.

The number of the attention map has a great influence on the final performance of FG-CPD, as shown in Fig. 14b. It can be seen that the accuracy of FG-CPD is lower than 85% when the value of M is smaller than 48. The best value for M is 64. When M is set to be 80, the accuracy is slightly degraded

to 92.42%. Therefore, the number of attention maps M used in this experiment is 64 to obtain the best effect of FG-CPD in pneumonia diagnosis.

On the effect of attention guided data augmentation aspect, it can be seen that both the attention guide cropping and dropping have positive effects on the performance, as shown in Fig. 14c. FG-CPD with both of them can promote the performance a lot. The accuracy of FG-CPD-o is 81.95%, which is significantly worse than FG-CPD-wCD.

The auto data augmentation also has a positive effect on FG-CPD, as shown in Fig. 14d. It can be seen that FG-CPD with random data augment is 84.21%, while FG-CPD with automatic data augmentation is 96.91%, which means that the automatic data augmentation can improve the performance of FG-CPD a lot.

5.2.4 Comparison with Other Attention Module

To verify the superiority of the SGNet attention module proposed in this research, the SGNet was replaced by Channel Attention Module (SENet), Spatial Attention Module (STN), and Hybrid Attention Module (CBAM) for a control experiment. The remaining parameters were set unchanged and were tested in the Chest X-ray Dataset from Tongji Hospital. Experimental results are shown in Table 6, the accuracy of the Channel Attention Module used was 95.96%, the accuracy of the Spatial Attention Module used was 95.85%, the accuracy of the Hybrid Attention Module used was 96.49%, SGNet has a more superior performance compared to other attention modules.

Table 6: Comparison with another attention module in case 2

Methods	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)
FG-CPD + SENet	96.09	94.29	94.09	94.19
FG-CPD + STN	95.95	94.49	93.54	94.01
FG-CPD + CBAM	96.49	95.51	94.17	94.83
FG-CPD + SGNet	96.91	96.12	94.80	95.44

In addition, three performance indicators (Precision, Recall, and F1) of FG-CPD on the Chest X-ray Dataset from Tongji Hospital all showed high results, proving that FG-CPD has excellent performance and great potential in fine-grained classification.

6 Conclusion and Future Research

In this research, a new fine-grained CNN for childhood pneumonia diagnosis (FG-CPD) is proposed. The main contribution of this research is as follows. Firstly, the fine-grained classification is introduced for the diagnosis of children with pneumonia. To acquire the raw images from the real-world chest X-ray database, the YOLOv4 algorithm is trained to detect the position of the chest and remove the irrelevant place in the raw chest X-ray images. Secondly, a new attention network, SGNet, is adopted in the proposed FG-CPD to enhance the ability to locate the discriminative parts in the raw chest X-ray images. Thirdly, automatic data augmentation and attention guided data augmentation is used to increase the diversity of the training data and avoid the overfitting of the FG-CPD. The proposed FG-CPD has been conducted on the public Chest X-ray 2017 dataset and it has achieved the state-of-the-art performance by comparing with other published methods. Then, FG-CPD was

tested on real-world Chest images from the children from Tongji Hospital, and the results show the potential of FG-CPD.

Although FG-CPD has achieved remarkable results, the limitations of FG-CPD can be summarized as follows. Firstly, the backbone network has a significant effect on the performance of FG-CPD. Therefore, how to select the right backbone network in practice can be investigated. Secondly, several data augment techniques have been applied in this research, and the results show that they can avoid the overfitting of FG-CPD. But these data augmentation techniques are selected by experiments. Based on the limitations, further research can be conducted to promote this work. Firstly, Auto Deep Learning can be introduced to this field, and it can select the right backbone network and the hyper-parameters of FG-CPD automatically. Secondly, as data augmentation is effective in FG-CPD, the investigation of the right data augmentation methods can be done to avoid the overfitting of FG-CPD further.

Acknowledgement: Yang Zhang and Liru Qiu contributed equally. Long Wen and Xiaoping Luo contributed equally, too.

Ethics: All data analysis in this case were approved by the Ethical Committee of Tongji Hospital of Huazhong University of Science & Technology (China) (TJ-IRB20210871) and were conducted in accordance with the Declaration of Helsinki.

Funding Statement: This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant No. 51805192, Major Special Science and Technology Project of Hubei Province under Grant No. 2020AEA009 and sponsored by the State Key Laboratory of Digital Manufacturing Equipment and Technology (DMET) of Huazhong University of Science and Technology (HUST) under Grant No. DMETKF2020029.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Gupta, M., Jain, R., Gupta, A., Jain, K. (2020). Real-time analysis of COVID-19 pandemic on most populated countries worldwide. *Computer Modeling in Engineering & Sciences, 125(3),* 943–965. DOI 10.32604/cmes.2020.012467.
2. Li, W., Deng, X., Shao, H., Wang, X. (2021). Deep learning applications for COVID-19 analysis: A state-of-the-art survey. *Computer Modeling in Engineering & Sciences, 129(1),* 65–98. DOI 10.32604/cmes.2021.016981.
3. Yue, Z., Ma, L., Zhang, R. (2020). Comparison and validation of deep learning models for the diagnosis of pneumonia. *Computational Intelligence and Neuroscience, 2020,* 8. DOI 10.1155/2020/8876798.
4. Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O. et al. (2020). Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: A descriptive study. *The Lancet Infectious Diseases, 20(4),* 425–434. DOI 10.1016/S1473-3099(20)30086-4.
5. Guo, X., Zhang, Y. D., Lu, S., Lu, Z. (2022). A survey on machine learning in COVID-19 diagnosis. *Computer Modeling in Engineering & Sciences, 130(1),* 23–71. DOI 10.32604/cmes.2021.017679.
6. Wen, L., Wang, Y., Li, X. (2022). A new cycle-consistent adversarial networks with attention mechanism for surface defect classification with small samples. *IEEE Transactions on Industrial Informatics.* DOI 10.1109/TII.2022.3168432.

7. He, Z., Shao, H., Zhong, X., Zhao, X. (2020). Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. *Knowledge-Based Systems, 2020(207),* 106396. DOI 10.1016/j.knosys.2020.106396.
8. Deng, X., Shao, H., Shi, L., Wang, X., Xie, T. (2020). A classification–detection approach of COVID-19 based on chest X-ray and CT by using keras pre-trained deep learning models. *Computer Modeling in Engineering & Sciences, 125(2),* 579–596. DOI 10.32604/cmes.2020.011920.
9. Zheng, H., Fu, J., Zha, Z. J., Luo, J., Mei, T. (2019). Learning rich part hierarchies with progressive attention networks for fine-grained image recognition. *IEEE Transactions on Image Processing, 29,* 476–488. DOI 10.1109/TIP.83.
10. Wen, L., Li, X., Gao, L. (2020). A new reinforcement learning based learning rate scheduler for convolutional neural network in fault classification. *IEEE Transactions on Industrial Electronics, 68(12),* 12890–12900. DOI 10.1109/TIE.2020.3044808.
11. Wang, J., Xu, C., Yang, Z., Zhang, J., Li, X. (2020). Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition. *IEEE Transactions on Semiconductor Manufacturing, 33(4),* 587–596. DOI 10.1109/TSM.66.
12. Sedik, A., Hammad, M., El-Samie, A., Fathi, E., Gupta, B. B. et al. (2021). Efficient deep learning approach for augmented detection of coronavirus disease. *Neural Computing and Applications,* 1–18. DOI 10.1007/s00521-020-05410-8.
13. Yu, X., Lu, S., Guo, L., Wang, S. H., Zhang, Y. D. (2021). ResGNet-C: A graph convolutional neural network for detection of COVID-19. *Neurocomputing, 452,* 592–605. DOI 10.1016/j.neucom.2020.07.144.
14. Ahsan, M. M., Alam, T. E., Trafalis, T., Huebner, P. (2020). Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients. *Symmetry, 12(9),* 1526. DOI 10.3390/sym12091526.
15. Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A. et al. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement, 145,* 511–518. DOI 10.1016/j.measurement.2019.05.076.
16. Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F. et al. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences, 10(9),* 3233. DOI 10.3390/app10093233.
17. Nam, J. G., Park, S., Hwang, E. J., Lee, J. H., Jin, K. N. et al. (2019). Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology, 290(1),* 218–228. DOI 10.1148/radiol.2018180237.
18. Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B. et al. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine, 15(11),* e1002686. DOI 10.1371/journal.pmed.1002686.
19. Pham, H. H., Le, T. T., Tran, D. Q., Ngo, D. T., Nguyen, H. Q. (2021). Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing, 437,* 186–194. DOI 10.1016/j.neucom.2020.03.127.
20. Sun, J., Li, X., Tang, C., Chen, S. (2021). BEVGGC: Biogeography-based optimization expert-VGG for diagnosis COVID-19 via chest X-ray images. *Computer Modeling in Engineering & Sciences, 129(2),* 729–753. DOI 10.32604/cmes.2021.016416.
21. Gayathri, J. L., Abraham, B., Sujarani, M. S., Nair, M. S. (2022). A Computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. *Computers in Biology and Medicine, 141,* 105134. DOI 10.1016/j.compbiomed.2021.105134.
22. Showkat, S., Qureshi, S. (2022). Determining the efficacy of transfer learning-based ResNet models in chest X-ray image classification for detecting COVID-19 pneumonia. *Chemometrics and Intelligent Laboratory Systems,* 104534. DOI 10.1016/j.chemolab.2022.104534.

23. Zhang, Y., Zhang, X., Zhu, W. (2021). ANC: Attention network for COVID-19 explainable diagnosis based on convolutional block attention module. *Computer Modeling in Engineering & Sciences, 127(3),* 1037–1058. DOI 10.32604/cmes.2021.015807.
24. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H. et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell, 172(5),* 1122–1131.e1129. DOI 10.1016/j.cell.2018.02.010.
25. Yi, R., Tang, L., Tian, Y., Liu, J., Wu, Z. (2021). Identification and classification of pneumonia disease using a deep learning-based intelligent computational framework. *Neural Computing and Applications,* 1–14. DOI 10.1007/s00521-021-06102-7.
26. Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P. et al. (2020). A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Applied Sciences, 10(2),* 559. DOI 10.3390/app10020559.
27. Hou, S., Han, J. (2022). COVID-19 detection via a 6-layer deep convolutional neural network. *Computer Modeling in Engineering & Sciences, 130(2),* 855–869. DOI 10.32604/cmes.2022.016621.
28. Fu, J., Zheng, H., Mei, T. (2017). Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4438–4446. Honolulu, USA.
29. Ju, M., Ryu, H., Moon, S., Yoo, C. D. (2020). GAPNet: Generic-attribute-pose network for fine-grained visual categorization using multi-attribute attention module. *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 703–707. Abu Dhabi.
30. Wang, Z., Wang, S., Yang, S., Li, H., Li, J. et al. (2020). Weakly supervised fine-grained image classification via Gaussian mixture model oriented discriminative learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9749–9758. Seattle, USA.
31. Xiao, T., Xu, Y., Yang, K., Zhang, J., Peng, Y. et al. (2015). The application of two-level attention models in deep convolutional neural network for fine-grained image classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 842–850. Boston, USA.
32. Zhang, Y., Wei, X. S., Wu, J., Cai, J., Lu, J. et al. (2016). Weakly supervised fine-grained categorization with part-based image representation. *IEEE Transactions on Image Processing, 25(4),* 1713–1725. DOI 10.1109/TIP.2016.2531289.
33. Zhao, B., Wu, X., Feng, J., Peng, Q., Yan, S. (2017). Diversified visual attention networks for fine-grained object classification. *IEEE Transactions on Multimedia, 19(6),* 1245–1256. DOI 10.1109/TMM.2017.2648498.
34. Sun, M., Yuan, Y., Zhou, F., Ding, E. (2018). Multi-attention multi-class constraint for fine-grained image recognition. *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 805–821. Munich, Germany.
35. Zheng, H., Fu, J., Zha, Z. J., Luo, J. (2019). Looking for the devil in the details: Learning trilinear attention sampling network for fine-grained image recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5012–5021. California, USA.
36. Ding, Y., Ma, Z., Wen, S., Xie, J., Chang, D. et al. (2021). AP-CNN: Weakly supervised attention pyramid convolutional neural network for fine-grained visual classification. *IEEE Transactions on Image Processing, 30,* 2826–2836. DOI 10.1109/TIP.83.
37. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., Le, Q. V. (2019). Autoaugment: Learning augmentation strategies from data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 113–123. California, USA.
38. Hammoudi, K., Benhabiles, H., Melkemi, M., Dornaika, F., Arganda-Carreras, I. et al. (2021). Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. *Journal of Medical Systems, 45(7),* 1–10. DOI 10.1007/s10916-021-01745-4.

39. Saraiva, A. A., Ferreira, N. M. F., de Sousa, L. L., Costa, N. J. C., Sousa, J. V. M. et al. (2019). Classification of images of childhood pneumonia using convolutional neural networks. *6th International Conference on Bioimaging*, pp. 112–119. Prague, Czech Republic.
40. Wu, H., Xie, P., Zhang, H., Li, D., Cheng, M. (2020). Predict pneumonia with chest X-ray images based on convolutional deep neural learning networks. *Journal of Intelligent & Fuzzy Systems, 39(3),* 2893–2907. DOI 10.3233/JIFS-191438.
41. Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141. Utah, USA.
42. Jaderberg, M., Simonyan, K., Zisserman, A. (2015). Spatial transformer networks. *Advances in Neural Information Processing Systems, 28,* pp. 2017–2025.
43. Woo, S., Park, J., Lee, J. Y., Kweon, I. S. (2018). Cbam: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19. Munich, Germany.