

Received 19 July 2022, accepted 8 August 2022, date of publication 16 August 2022, date of current version 19 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3198958

RESEARCH ARTICLE

Thorax Disease Classification Based on Pyramidal Convolution Shuffle Attention Neural Network

KAI CHEN^{ID}, XUQI WANG, AND SHANWEN ZHANG

College of Electronic Information, Xijing University, Xian, Shanxi 710123, China

Corresponding author: Kai Chen (chen_kai@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62172338 and Grant 62072378, in part by the Xijing University High-Level Talent Special Fund Project under XJ21B14, and in part by the Henan Science and Technology Research Project under Grant 212102210005.

ABSTRACT Chest X-ray is one of the most common radiological examinations for screening thoracic diseases. Despite the existing methods based on convolution neural network that have achieved remarkable progress in thoracic disease classification from chest X-ray images, the scale variation of the pathological abnormalities in different thoracic diseases is still challenging in chest X-ray image classification. Based on the above problems, this paper proposes a residual network model based on a pyramidal convolution module and shuffle attention module (PCSANet). Specifically, the pyramid convolution is used to extract more discriminative features of pathological abnormality compared with the standard 3×3 convolution; the shuffle attention enables the PCSANet model to focus on more pathological abnormality features. The extensive experiment on the ChestX-ray14 and COVIDx datasets demonstrate that the PCSANet model achieves superior performance compared with the other state-of-the-art methods. The ablation study further proves that pyramidal convolution and shuffle attention can effectively improve thoracic disease classification performance. The code is published in <https://github.com/Warrior996/PCSANet>.

INDEX TERMS Chest X-ray, pyramidal convolution, shuffle attention, thoracic disease classification.

I. INTRODUCTION

Chest X-ray (CXR) is an economical and affordable medical imaging technology, which is mostly used for early screening of thorax, lung tissue, heart, and other diseases, such as pneumonia, heart failure, lung cancer, and so on. More than 1 million adults are hospitalized each year for pneumonia, and approximately 50,000 patients died each year from this disease only in the United States [1]. Typically, most chest X-ray images primarily rely on manual observation by a professional radiologist. Unfortunately, a radiologist with professional clinical training can make mistakes due to the complex pathological abnormalities and subtle structural changes of various thoracic diseases. Approximately 20-50% of lung nodules are reported to be missed or misdiagnosed [2]. Even excellent radiologists will have 3-6% of serious misdiagnosis [3]. Therefore, the precise classification and localization

of chest X-ray images have great significance to assist the clinical diagnosis of thoracic diseases.

The previous works made significant progress on the chest X-ray images classification. For example, wang *et al.* [4] evaluated the classification effects of four classical CNN models, AlexNet [5], GoogleNet [6], VGGNet-16 [7], and ResNet-50 [8] on the chest X-ray images while publishing the ChestX-ray14 dataset. The CheXNet model proposed by Rajpurkar *et al.* [9] achieved the accurate classification of thoracic diseases by modifying the fully connected layer of DenseNet-121 [10] and loss function, especially the accuracy of the diagnosis of pneumonia reaches 88.87%, which exceeds the current human diagnosis level. However, the chest X-ray images classification is still a challenge due to the complexity of disease pathological abnormalities. On the one hand, there is an obvious scale variation of the pathological abnormalities in various thoracic diseases. The example is shown in Fig.1, “pneumonia” and “cardiomegaly” cover almost the half of left lung and the whole heart, but “nodule”

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo^{ID}.

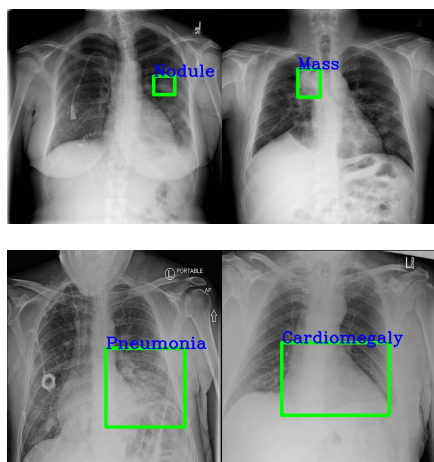


FIGURE 1. Four chest radiographs with different diseases from the ChestX-ray14 dataset. Pathological findings and pathological abnormalities are highlighted by blue text and green bounding boxes.

and “mass” are much smaller than it. Consequently, the standard 3×3 convolution may not adapt to the scale variation of discriminative features when facing pathological abnormalities of various thoracic diseases. On the other hand, it is difficult for convolution neural network to learn discriminative representations of the complicated pathology abnormalities in thoracic diseases since there are similar organs and tissues in each chest X-ray image. Thus, chest X-ray image classification is a challenging task due to the significant scale variance of the pathological abnormalities in various thoracic diseases and the similarities of discriminative features in each chest X-ray image.

In recent years, deep learning have made significant advances in many areas of medical image analysis, such as lesion region segmentation or detection [11], [12], [13], [14], [15], disease classification [4], [16], [17], [18], [19], [20], image alignment [21], [22], [23]. The previous works were used for pathological abnormality detection and disease classification by high-resolution feature maps and standard convolution neural network. For example, Huang *et al.* [24] proposed a high-resolution network (HRNet) to extract abnormal abnormality features from four high-resolution feature maps. Guan and Huang [25] proposed a category-wise residual attention learning (CRAL) framework, in which the feature embedding module uses convolutional neural network (CNN) to learn high-level features. In this paper, we propose a residual network model based on pyramidal convolution and shuffle attention for a multi-label chest X-ray image classification. The example of the PCSANet model for the diagnosis of thoracic disease is illustrated in Fig.2. However, the high-resolution feature maps and standard convolution neural network only learn single-scale features of pathological abnormality. Therefore, the PCSANet model does not follow the standard 3×3 convolution, but employs pyramid convolution to extract discriminative features of pathological abnormality from chest X-ray images. In addition, the latest shuffle attention mechanism is also introduced to focus

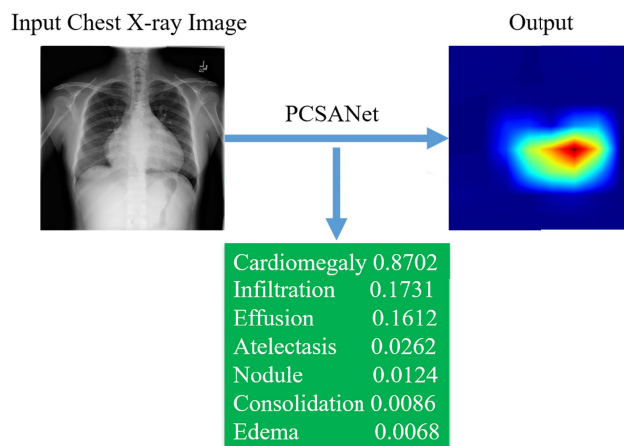


FIGURE 2. Example of the PCSANet model for thoracic disease diagnosis. The input is a chest X-ray image and the output is prediction scores and a heatmap of the pathological abnormality.

on more pathological abnormality features. The numerous experiments demonstrate that the model effectively improves the classification performance of thoracic diseases from chest X-ray images, with the AUC can reach 82.5% on the Chest X-ray14 dataset. The main contributions of this paper are as follows:

(1) This paper proposes a residual network model based on pyramidal convolution for disease feature extraction, in which pyramidal convolution can extract multi-scale discriminative features of pathological abnormality compared with the standard 3×3 convolution.

(2) The shuffle attention is introduced into the PCSANet model to focus on more pathological abnormality features.

The structure of this paper is as follows: section II mainly introduces the related work of chest X-ray image classification and attention mechanisms in the field of medical imaging. Section III describes the proposed PCSANet model. Section IV shows the dataset and experiments. Section V summarizes the whole work.

II. RELATED WORK

A. THORACIC DISEASE CLASSIFICATION ON THE CHEST X-Ray IMAGE

In recent years, the research on the computer-aided diagnosis of thoracic diseases has attracted much attention. With the release of the ChestX-ray14 dataset, the method based on deep learning for thoracic diseases classification has become a research hotspot, and achieved excellent classification performance. Kumar *et al.* [26] proposed an enhanced cascade network for thoracic disease classification by studying the most suitable loss function. Guan *et al.* [16] proposed a two-branch architecture ConsultNet to learn discriminative features by extracting critical disease-specific features and enhancing the potential semantic dependencies in the feature space. Chen *et al.* [27] introduced a graph convolution network (GCN) into the classification of thoracic diseases to explore the relevant information of pathology.

Rehman *et al.* [28] proposed a self-activated CNN approach for the detection of multiple thoracic diseases, including COVID-19. Due to the expensive manual annotation of chest X-ray images, Li *et al.* [29] proposed a unified approach for thoracic disease identification and localization using class information and limited positional annotation. Rozenberg *et al.* [30] inspired by [29], proposed a method for thoracic disease localization with limited annotation based on a well-posed loss function and an architecture for patch non-independence and shift invariance. It is worth noting that [30] uses much cheaper image annotations instead of very few bounding box annotations. Compared with previous works, the PCSANet model incorporates a pyramidal convolution module and a shuffle attention module to achieve superior performance.

B. ATTENTION MECHANISM IN CHEST X-ray IMAGE ANALYSIS

Attention mechanisms have been successfully explored in the field of natural language processing and computer vision. The multi-label chest X-ray image classification task requires learning more discriminative features of pathological abnormality to distinguish various thoracic diseases. Ma *et al.* [31] proposed a multi-attention convolution neural network using a multi-attention mechanism and fusion of global and local information for thoracic disease classification and localization. Yan *et al.* [32] proposed a weakly supervised deep learning framework equipped with the SE-blocks, multi-map transfer, and max-min pooling for classifying thoracic diseases as well as localizing suspicious lesion regions. Wang *et al.* [17] proposed a triple attention learning model for thoracic disease classification, which integrates three attention modules of channel level, element level, and scale level into a unified framework. Ouyang *et al.* [33] proposed an attention-driven weakly supervised algorithm for the problem of abnormality localization, which incorporates a hierarchical attention mining framework that unifies activation- and gradient-based visual attention in a holistic manner. Different from the previous works, this paper introduces a shuffle attention mechanism that focuses on more discriminative features of pathological abnormality to improve the classification performance of thoracic diseases from chest X-ray images.

III. PYRAMIDAL CONVOLUTIONAL SHUFFLE ATTENTION RESIDUAL NETWORK MODEL

The overview of the proposed PCSANet model is illustrated in Fig.3, which consists of a backbone network for feature extraction and the shuffle attention module. The backbone network learns the multi-scale discriminative features of pathological abnormality through a residual network model with pyramidal convolution. Then, the discriminative features are entered into the shuffle attention module to increase representation power: focusing on more discriminative features of pathological abnormality. The partial layers of the proposed PCSANet model are illustrated in Table.1, where output

TABLE 1. Partial layers of the proposed PCSANet model.

Layer	Output shape	Kernel size	Stride	Padding
PyConvBlock	(256, 56, 56)	/	/	/
Conv2d	(64, 112, 112)	(1, 1)	(1, 1)	/
BatchNorm2d	(64, 112, 112)	/	/	/
ReLU	(64, 112, 112)	/	/	/
PyConv4	(64, 56, 56)	/	/	/
Conv2d	(16, 56, 56)	(3, 3)	(2, 2)	(1, 1)
Conv2d	(16, 56, 56)	(5, 5)	(2, 2)	(2, 2)
Conv2d	(16, 56, 56)	(7, 7)	(2, 2)	(3, 3)
Conv2d	(16, 56, 56)	(9, 9)	(2, 2)	(4, 4)
BatchNorm2d	(64, 56, 56)	/	/	/
Conv2d	(256, 56, 56)	(1, 1)	(1, 1)	/
BatchNorm2d	(256, 56, 56)	/	/	/
SALayer	(256, 56, 56)	/	/	/
AdaptiveAvgPool2d	(2, 1, 1)	/	/	/
Sigmoid	(2, 1, 1)	/	/	/
GroupNorm	(2, 56, 56)	/	/	/
Sequential	(256, 56, 56)	/	/	/
MaxPool2d	(64, 56, 56)	(3, 3)	(2, 2)	(1, 1)
Conv2d	(256, 56, 56)	(1, 1)	(1, 1)	/
BatchNorm2d	(256, 56, 56)	/	/	/

shape, kernel size, stride, and padding show how the features information are collected and passed to the next block.

A. BACKBONE NETWORK

The PCSANet model adopts the ResNet50 as the backbone network to extract the discriminative feature mapping $F \in \mathbb{R}^{C \times W \times H}$ of the input image and outputs corresponding category information of the image. ResNet consists of five residual blocks, an average pooling layer, and a fully connected layer. The output vector of the fully connected layer passes through a nonlinear activation layer—the sigmoid layer:

$$\tilde{p}(c | I) = 1 / (1 + \exp(-p_g(c | I))) \quad (1)$$

where I denote the input image. $\tilde{p}(c | I)$ indicates the probability score of I belongs to the category $c \in \{1, 2, \dots, C\}$. The global branching parameters W are optimized by minimizing the binary cross entropy loss function(BCE):

$$L(W) = -\frac{1}{C} \sum_{c=1}^C l_c \log(\tilde{p}(c | I)) + (1 - l_c) \log(1 - \tilde{p}(c | I)) \quad (2)$$

where l_c denotes the true label of category c and C denotes the total number of disease categories.

B. PYRAMIDAL CONVOLUTION MODULE

The bounding box of lesion areas in the ChestX-ray14 dataset provided by the National Institutes of Health Clinical Center (NIHCC) shows that the lesion areas of thoracic diseases such as Atelectasis, Infiltration, and Pneumonia are larger, while the lesion areas of thoracic diseases such as Mass and Nodule are smaller. The pyramidal convolution can extract the multi-scale discriminative features to improve classification performance. The structure of pyramidal convolution is shown in Fig.4.

The pyramidal convolution contains n convolution kernels of different scales, which have different spatial resolutions

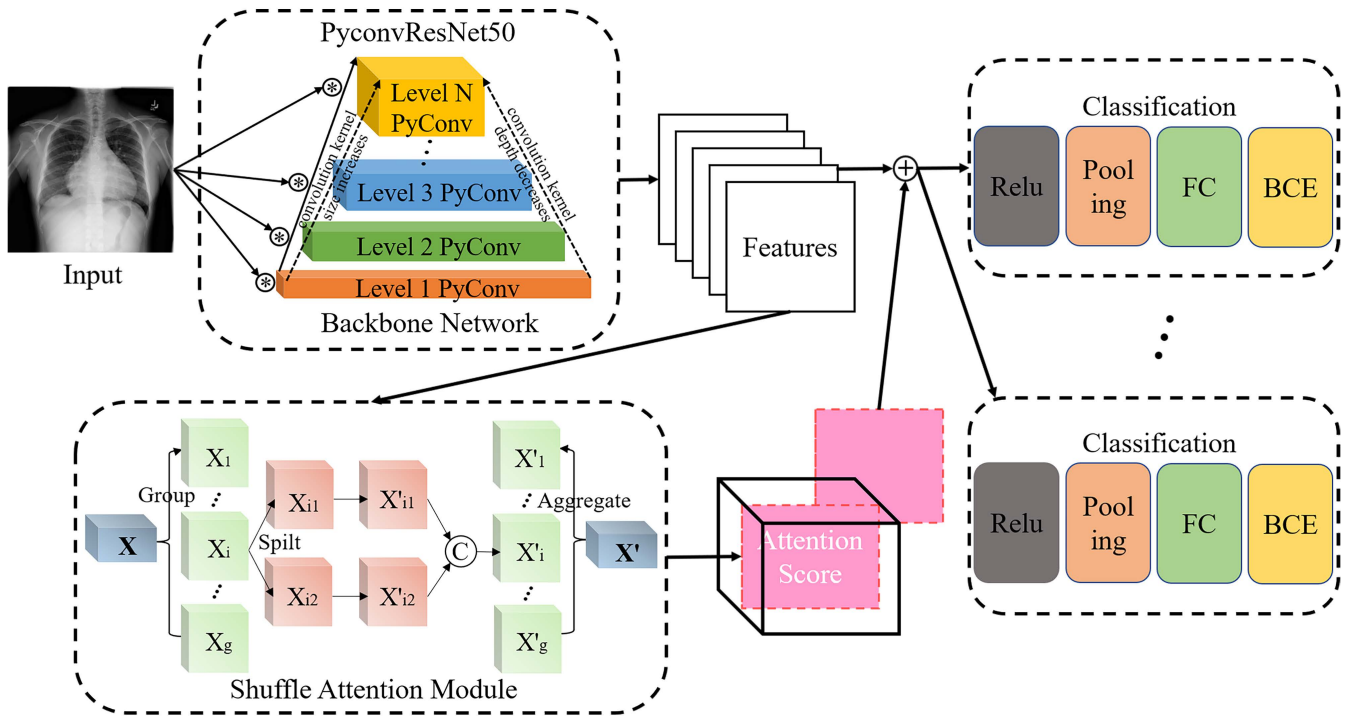


FIGURE 3. Overview of the PCSANet model. The PCSANet model consists of a Backbone Network and a Shuffle Attention module. First, we feed a input image into the Backbone Network and obtain the multi-scale discriminative features. Then we introduce shuffle attention to focus on more pathological abnormality features. At last, both of them are combined to classify the input image. Note that the “Relu”, “Pooling”, “FC”, and “BCE” represent the relu layer, average pooling layer, fully connected layer, and binary cross entropy loss function respectively.

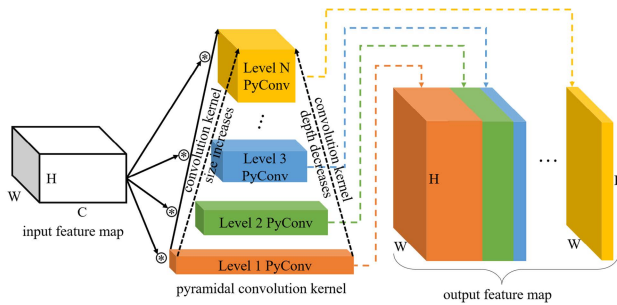


FIGURE 4. Structure of the pyramidal convolution module. Pyramidal convolution kernels increase in size and decrease in depth from bottom to top.

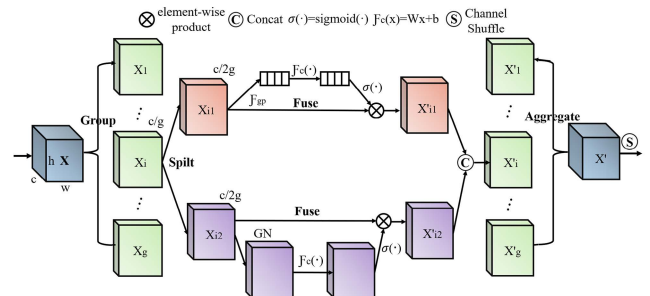


FIGURE 5. Structure of the shuffle attention module.

and depths to capture more detailed feature information without increasing the computational cost. The convolution kernels of different scales lead to different receptive fields in each layer, which can extract multi-scale features at different levels, both 1×1 and 3×3 sized convolution kernels can learn detailed information, both 5×5 and 7×7 sized convolution kernels with larger receptive fields can learn semantic information. In addition, the pyramidal convolution can perceive the spatial feature relationships at different levels, which also makes the convolution layer have the ability to retain detailed features. Meanwhile, it also effectively alleviates the problem of local information loss due to down-sampling. The feature mapping is grouped and convolved by different convolution branches. At last, the features of different branches are cascaded and stitched to form a fused feature. In this paper, the

residual bottleneck block based on the pyramidal convolution module is used to replace the residual block of ResNet, and the 3×3 convolution kernel in the standard residual block is replaced with the pyramidal convolution kernel of different kernel levels (9×9 , 7×7 , 5×5 , 3×3), where the depth of the convolution kernels varies at each level.

C. SHUFFLE ATTENTION MODULE

There are two broad types of attention mechanisms: spatial attention and channel attention, which are used to capture the pixel-level pairwise relationships and channel dependence. Different from previous attention mechanisms, shuffle attention employs shuffle units to efficiently combine spatial attention and channel attention to obtain better performance without increasing the computational cost.

The structure of the shuffle attention module is shown in Fig.5, the entire attention mechanism consists of four parts:

(1) Feature grouping: given an input feature mapping $X \in R^{C \times W \times H}$, where C , W , and H represent the channel, width, and height of the feature map, respectively. First, the feature map X is split into G groups $X = [X_1, \dots, X_G]$, $X_i \in R^{(C/G) \times W \times H}$. Then, each group is divided into two branches along the direction of the channel $X_{i1}, X_{i2} \in R^{((C/G)/2) \times W \times H}$. One branch utilizes interrelationships between channels to generate a channel attention map, while the other branch utilizes the spatial relationship between a feature to generate a spatial attention map.

(2) Channel attention: First, the channel statistics is obtained according to the avgpool function. Then, the feature representation is enhanced by using the fuse linear function. At last, the global information is embedded by multiplying with the original feature values after being activated by the sigmoid function to obtain the category representation containing the channel attention weights X_{i1}' to enhance the features. The specific calculation formula is as follows:

$$X_{i1}' = \sigma(f_c(s)) \cdot X_{i1} = \sigma(W_1 s_1 + b_1) \cdot X_{i1} \quad (3)$$

where f_c denotes the linear function, σ denotes the sigmoid activation function, s_1 denotes the average pooled feature, and the two parameters $W_1, b_1 \in R^{((C/G)/2) \times 1 \times 1}$ can be obtained by network training.

(3) Spatial attention: Spatial attention can be regarded as a supplement to channel attention. First, spatial statistics are obtained according to the groupnorm normalization function. Then, the feature representation is enhanced by using the fuse linear function. At last, the global information is embedded by multiplying with the original feature value after being activated by the sigmoid function to obtain the one containing spatial attention weights X_{i2}' to strengthen the importance of features for a specific region. The specific formula is as follows:

$$X_{i2}' = \sigma(W_2 \cdot GN(X_{i2}) + b_2) \cdot X_{i2} = \sigma(W_2 s_2 + b_2) \cdot X_{i2} \quad (4)$$

where GN denotes the groupnorm normalization function, s_2 denotes the normalized feature, and the two parameters $W_2, b_2 \in R^{((C/G)/2) \times 1 \times 1}$ can be obtained through network training.

(4) Aggregation: Combining features $X_{i1}' X_{i2}'$ weighted by channel attention and spatial attention, back to the grouped dimension $CG^{-1} \times H \times W$, where C is the channel dimension, H and W are the length and width of the feature map. Merge the grouped blocks again and return to the original dimension $C \times H \times W$. After completing both attention learning and recalibrating the features, the two branches are spliced and aggregated $X_{i2}' = [X_{i1}', X_{i2}'] \in R^{(C/G) \times W \times H}$. Then all sub-features are aggregated. In the end, the channel grouping operation is performed.

IV. EXPERIMENT

A. DATASET AND EVALUATION METRICS

ChestX-ray14 is one of the early large-scale chest X-ray datasets published by the National Institutes of Health. The data are collected from the National Institutes of Health Clinical Center's Picture Archiving and Communication System. Wang et al. [4] designed an automatic annotation algorithm based on natural language processing techniques to mine keywords from medical examination reports and automatically annotate a sample of them. The dataset includes 112,120 X-ray frontal view images of 30,805 patients. The example images and the corresponding labels are shown in Fig.6. If none of the 14 types of diseases are found on an image, it is labeled as "No Findings". However, a "No Findings" sample simply means that it does not contain any of the 14 types of diseases mentioned above, it is not possible to determine whether it contains other diseases. In addition, the ChestX-ray14 dataset is divided into a training validation set (80%) and a test set (20%) at the patient level, in which bounding boxes of lesion areas in 984 images for 8 types of diseases are labeled for performance evaluation of the lesion area localization task.

COVIDx Dataset¹ [34] is a new dataset for the detection of COVID-19 with over 30,000 CXR images from a multinational cohort of over 16,400 patients, which contains 15884 confirmed positive COVID-19 cases, 13992 normal cases. The COVIDx dataset is collected from eight open source chest radiography datasets: COVID-19 Image Data Collection² [35], Figure1-COVID-chestxray-dataset,³ Actualmed-COVID-chestxray-dataset,⁴ COVID-19 radiography database⁵ [36], [37], rsna-pneumonia-detection-challenge dataset,⁶ RICORD COVID-19 dataset⁷⁸ and bimcv-covid19 dataset⁹ [38]. In order to avoid overfitting, we divide the training set into a training set(80%) and a validation set(20%), and the test set is still the COVIDx Dataset test set, consisting of only 400 images. Furthermore, the COVIDx dataset is still being updated.

1) MULTI-LABEL SETTINGS

Using a C -dimensional vector $\mathbf{L} = [l_1, l_2, \dots, l_c]$, where $l_c \in \{0, 1\}$. l_c denotes the presence or absence of the c disease, 1 indicates presence, and 0 indicates absence. C is the number of all diseases in the dataset. If \mathbf{L} is a zero vector, this means

¹https://www.kaggle.com/datasets/andyczhao/covidx-cxr?select=competition_test

²<https://github.com/ieee8023/covid-chestxray-dataset>

³<https://github.com/agchung/Figure1-COVID-chestxray-dataset>

⁴<https://github.com/agchung/Actualmed-COVID-chestxray-dataset>

⁵<https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>

⁶<https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>

⁷<https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70230281>

⁸<https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=89096912>

⁹<https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/>

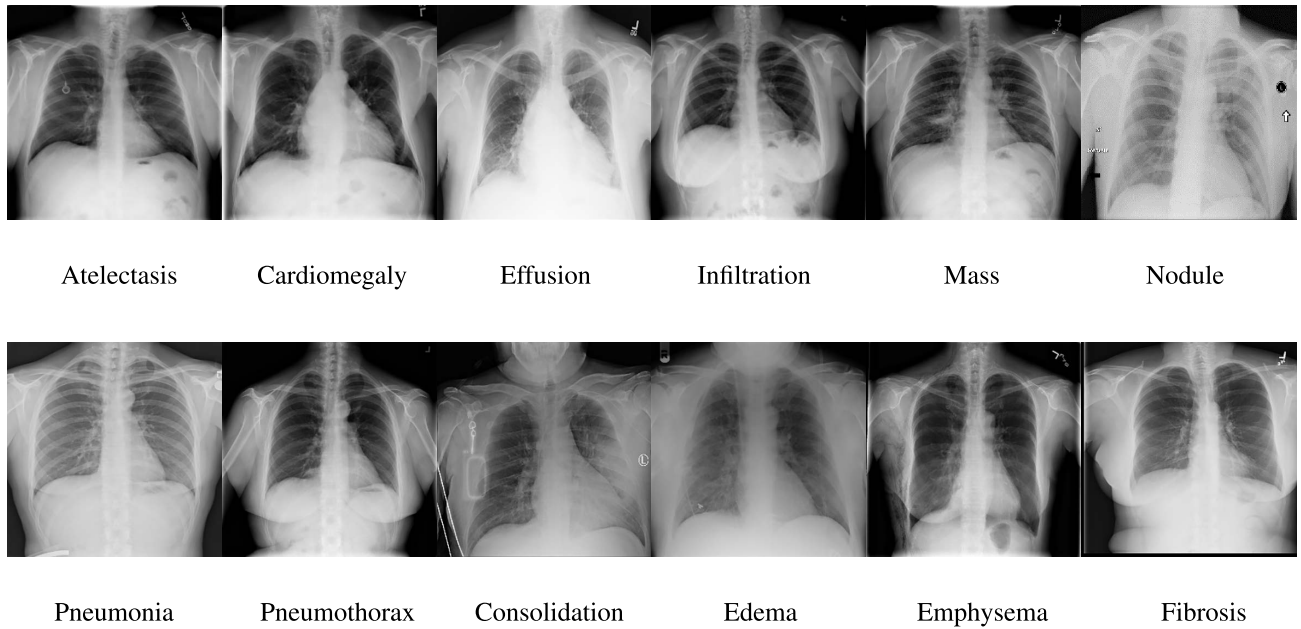


FIGURE 6. Example images and the corresponding labels in the Chest X-ray14 dataset.

that none of the above 14 types of diseases are exists in the image.

2) EVALUATION METRICS

To evaluate the PCSANet model on the ChestX-ray14 dataset, we have used Area Under Curve scores (AUC) and Receiver Operating Characteristic curves (ROC). The horizontal coordinate of the ROC curve is the False Positive Rate (FPR) and the vertical coordinate is the True Positive Rate (TPR). According to the multi-label setting, the FPR is the proportion of samples with the true label “0” and predicted “1”, and the TPR is the proportion of samples with the true label “1” and predicted “1”. The formula is as follows:

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (5)$$

We have used confusion matrix, accuracy, sensitivity, precision, specificity and F1-score to evaluate the PCSANet model on the Covidx dataset. The formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

$$F1 - score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (10)$$

where TP and TN indicate the number of rightly predicted positive and negative samples, FP and FN represent to the number of mistakenly predicted positive and negative samples.

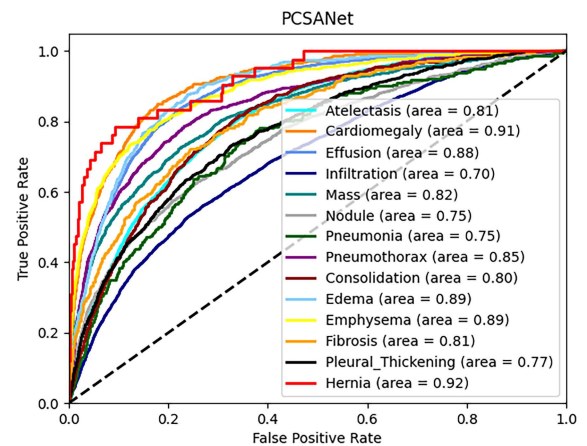


FIGURE 7. ROC curves of the PCSANet model over the 14 types of pathologies.

B. EXPERIMENTAL SETUP

The PCSANet model was implemented on the PyTorch framework and trained on a computer with NVIDIA TITAN GPU for 120 epochs from scratch with a batch size of 64. During training, we perform image preprocessing by resizing the input images to 256×256 , randomly rotating and horizontal flipping, randomly resized center cropping to 224×224 , and normalizing by the mean and standard deviation of the ImageNet dataset [42]. We use ResNet-50 as the backbone network and shuffle data only in the training stage. We optimize the network using Adam [43] with a weight decay of $1e-5$, a betas of (0.9, 0.999), a eps of $1e-8$, and the learning rate starts from 0.001 and is divided by 10 after 30 epochs. During validation and testing, we also resize the original image to 256×256 , resize center cropping to 224×224 , and normalize with the same as the training stage.

TABLE 2. Comparison results of previous works on the ChestX-ray14 dataset. We compute the AUC score of each class and the average AUC scores across the 14 types of diseases.

Method	Wang [4]	Li [29]	Baltruschat [39]	Yao [40]	Wang [41]	Huang [24]	PCSANet
Atel	0.700	0.727	0.755	0.733	0.751	0.794	0.807
Card	0.810	0.836	0.877	0.856	0.871	0.902	0.910
Effu	0.759	0.789	0.818	0.806	0.818	0.839	0.879
Infi	0.661	0.672	0.694	0.673	0.681	0.714	0.698
Mass	0.693	0.776	0.810	0.777	0.800	0.827	0.824
Nodu	0.669	0.696	0.736	0.724	0.715	0.727	0.750
Pnel	0.658	0.649	0.703	0.684	0.694	0.703	0.750
Pne2	0.799	0.808	0.819	0.805	0.825	0.848	0.850
Cons	0.703	0.720	0.742	0.711	0.742	0.773	0.802
Edem	0.805	0.806	0.842	0.806	0.835	0.834	0.888
Emph	0.833	0.888	0.875	0.842	0.843	0.911	0.890
Fibr	0.786	0.771	0.800	0.743	0.804	0.824	0.812
PT	0.684	0.737	0.742	0.724	0.746	0.752	0.768
Hern	0.872	0.693	0.916	0.775	0.902	0.916	0.915
Mean	0.745	0.755	0.795	0.761	0.788	0.812	0.825

The 14 types of pathologies are Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening, and Hernia

C. RESULTS ANALYSIS

To prove the effectiveness of the PCSANet model on the chest X-ray image, the PCSANet model is evaluated on both ChestX-ray14 and COVIDx datasets.

1) RESULTS ON CHESTX-ray14 DATASET

We report the performance of the PCSANet model on the ChestX-ray14 dataset in the section. The AUC scores and ROC curves of each pathology are shown in Table.2 and Fig.7. Notably, the PCSANet model achieve a high AUC score of 0.825. The results of comparing with previous works are summarized in Table.2. Wang *et al.* [4] evaluated the classification of thoracic diseases in CNN classic models AlexNet [5], GoogleNet [6], VGGNet-16 [7], and ResNet-50 [8], respectively. Baltruschat *et al.* [39] fully considered the influence of non-image features on thoracic disease classification and add angle, gender and other features to the model. Yao *et al.* [40] achieved excellent classification results by learning multi-scale features while generating higher resolution salient maps. Li *et al.* [29] proposed a method for thoracic disease identification and localization using class information and limited positional annotation. Wang *et al.* [41] proposed a two-branch CNN architecture: the global branch and local branch learning features from global images and the local regions, which is guided by heatmaps produced by class activation mapping (CAM). Huang *et al.* [24] proposed a fused high-resolution network (FHRNet) concatenating global and local feature extractors with a global average pooling layer for diagnosing chest X-ray images. Lietal The PCSANet model has made great progress compared with previous works. The overall performance of the PCSANET model is improved by 3.6% and 7.8% compared to [39] and [40]. Compared to the second-best model (Huang *et al.* [24]), the PCSANet model performs better in 9 types of disease classifications: “Atelectasis”, “Cardiomegaly”, “Effusion”, “Nodule”, “Pneumonia”, “Pneumothorax”, “Consolidation”, “Edema” and “Pleural Thickening”, and it has increased by 6.3% and 6.1% in “Pneumonia” and “Edema”. For some challenging

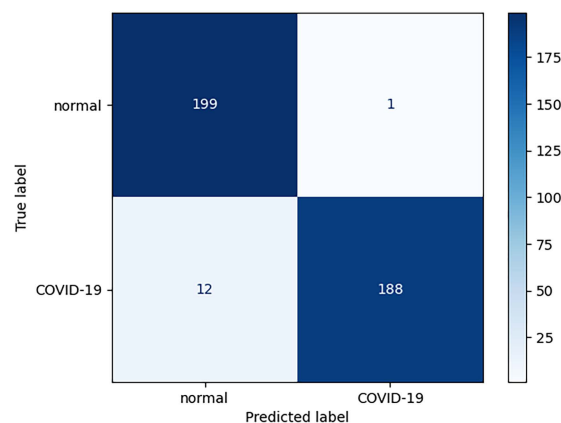


FIGURE 8. Confusion matrix of the PCSANet model on the COVIDx dataset.

diseases, such as “Effusion” and “Edema”, the accuracy is dramatically improved by a margin of at least 6.0% contrast to the second highest AUC in Table.2. The overall classification performance is improved by learning more multi-scale discriminative features of pathological abnormality. In addition, the data clearly show that the AUC score of infiltration is relatively low because the diagnosis of infiltration mainly depends on the changes in lung regional texture, which is difficult to recognize.

As can be seen from Table.2, the horizontal comparison shows that the existing methods and our model have different classification effects. Among them, 9 types of thoracic diseases have the best AUC scores with the PCSANet model. The vertical comparison shows that the existing methods and the PCSANet model have different classification effects in the 14 types of thoracic diseases. The most accurately thoracic disease is “Cardiomegaly”, with an AUC score of 0.910, and the least accurately identified disease is “Infiltration”, with an AUC score of 0.698.

As shown in Fig.7, we also map the ROC curves of the PCSANET model for each of the 14 types of thoracic diseases, and it is worth noting that the ROC curve

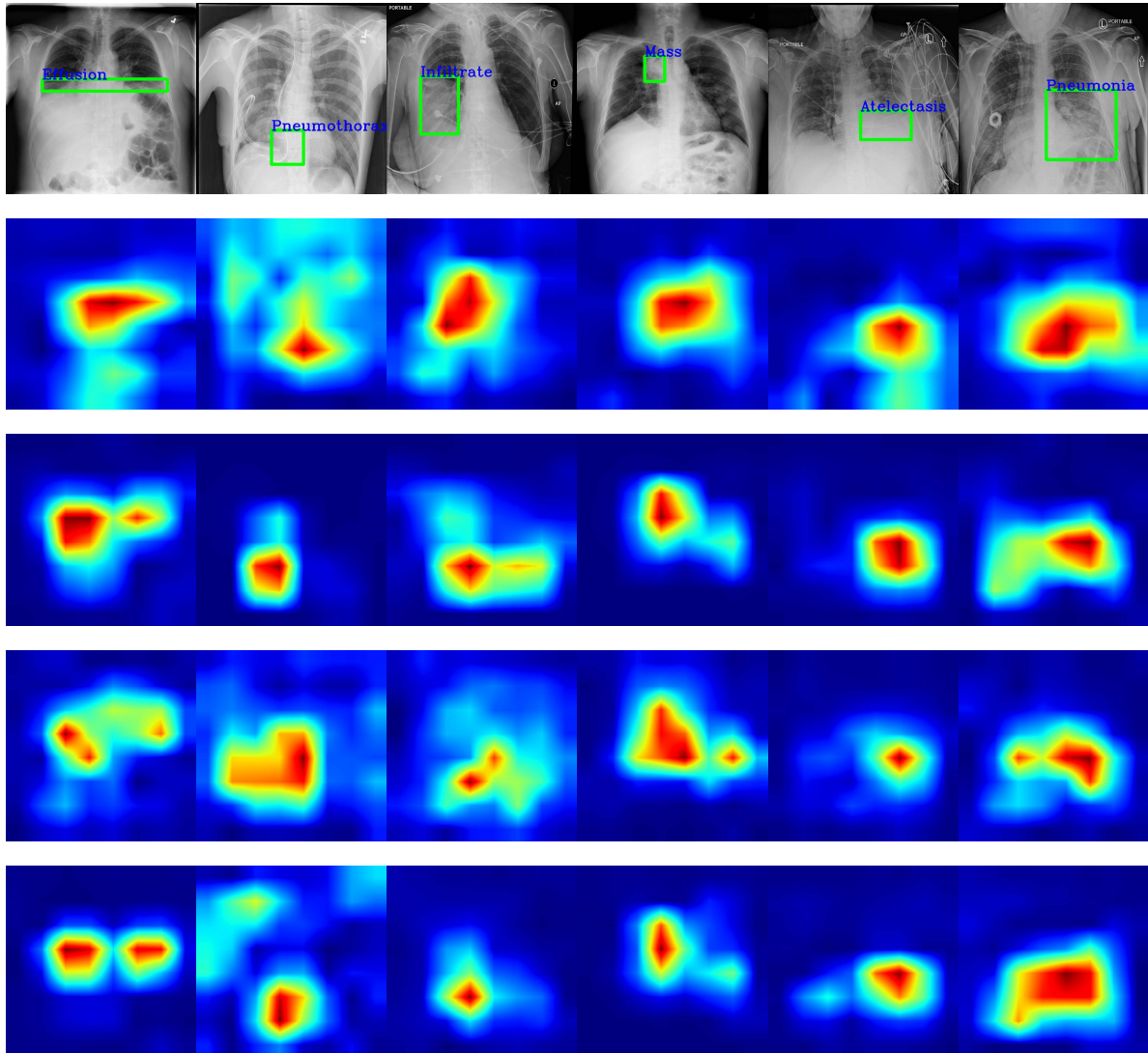


FIGURE 9. Example of a heat map generated by Grad-cam. The first row is a bounding box provided by the National Institutes of Health Clinical Center (NIHCC), the second-row models resnet-50, the third row is the addition of the pyramidal convolution module, and the fourth and fifth rows are the PCSANet model with the addition of the shuffle attention module and the pyramidal convolution module, respectively.

TABLE 3. Comparison of different methods on the COVIDx dataset(%). For each column, the best results are highlighted in bold.

Method	Accuracy	Sensitivity	Precision	Specificity	F1-score
Deep-COVID [44]	81.33	66.00	90.41	96.50	76.30
nCOVnet [45]	87.33	82.00	91.11	96.00	86.32
DarkCovidNet [46]	88.67	89.00	94.68	97.50	91.75
COVID-Net [34]	93.33	91.00	98.91	99.47	94.79
PCSANet	96.75	94.00	99.47	99.50	96.66

of “Infiltration” is flatter than that of “Cardiomegaly”, which indicates that “Infiltration” is less effective than “Cardiomegaly” in classification performance.

2) RESULTS ON COVIDx DATASET






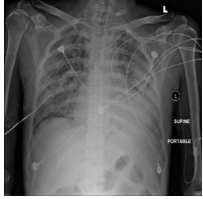




We compare the performance of the PCSANet model on the COVIDx dataset with several state-of-the-art COVID-19 detection methods [34], [44], [45], [46]. As shown in Table.3, it is obvious that the PCSANet model outperforms the state-of-the-art methods. The PCSANet model has significant

TABLE 4. Comparison of ablation study with the same experimental settings. The average AUC scores are reported. For each column, the best results are highlighted in bold.

Method	PCSANet	w/o attention	w/o pyconv	backbone
Atel	0.807	0.803	0.783	0.782
Card	0.910	0.911	0.893	0.891
Effu	0.879	0.880	0.870	0.869
Infi	0.698	0.698	0.687	0.686
Mass	0.824	0.830	0.795	0.796
Nodu	0.750	0.752	0.721	0.716
Pnel	0.750	0.750	0.724	0.735
Pne2	0.850	0.853	0.826	0.831
Cons	0.802	0.802	0.794	0.793
Edem	0.888	0.883	0.880	0.878
Emph	0.890	0.893	0.839	0.847
Fibr	0.812	0.805	0.779	0.774
PT	0.768	0.763	0.747	0.744
Hern	0.915	0.895	0.844	0.837
Mean	0.825	0.823	0.799	0.798

performance improvements on all metrics, with the accuracy increased by 3.53%, sensitivity increased by 3.19%, precision

TABLE 5. Classification examples of the PCSANet model. We present the top-8 predicted categories and the corresponding probability scores. The ground truth labels are highlighted in red or blue.

Images					
	Atelectasis 0.8886 Infiltration 0.3991 Effusion 0.2667 Consolidation 0.1135 Pneumonia 0.0339 Nodule 0.0240 Pneumothorax 0.0097 Edema 0.0092	Cardiomegaly 0.9711 Effusion 0.4259 Infiltration 0.2713 Atelectasis 0.0419 Edema 0.0287 Nodule 0.0208 PT 0.0202 Consolidation 0.0151	Effusion 0.8966 Infiltration 0.1517 Atelectasis 0.0987 Edema 0.0797 Nodule 0.0383 Consolidation 0.0333 Cardiomegaly 0.0324 Mass 0.0314	Infiltration 0.6114 Nodule 0.4292 Edema 0.1522 Consolidation 0.0929 Pneumonia 0.0854 Mass 0.0343 Effusion 0.0269 Fibrosis 0.0103	Emphysema 0.8918 Pneumothorax 0.8154 Atelectasis 0.1361 Effusion 0.0824 Infiltration 0.0557 PT 0.0422 Mass 0.0190 Nodule 0.0133
Images					
	Infiltration 0.6114 Nodule 0.4294 Edema 0.1522 Consolidation 0.0929 Pneumonia 0.0854 Mass 0.0343 Effusion 0.0260 PT 0.0121	Mass 0.9079 Nodule 0.8595 Infiltration 0.0293 Effusion 0.0150 Atelectasis 0.0124 Consolidation 0.0115 PT 0.0067 Fibrosis 0.0052	Nodule 0.8640 Infiltration 0.1925 Mass 0.1579 Effusion 0.0586 Fibrosis 0.0357 PT 0.0254 Consolidation 0.0254 Pneumonia 0.0224	Effusion 0.8915 Atelectasis 0.2436 Infiltration 0.0993 Consolidation 0.0727 Mass 0.0383 Nodule 0.0292 Pneumothorax 0.0288 PT 0.0208	Effusion 0.4511 Atelectasis 0.3270 Infiltration 0.1765 Cardiomegaly 0.0570 Consolidation 0.0507 Edema 0.0264 Pneumonia 0.0178 Nodule 0.0129

increased by 0.56%, specificity increased by 0.03%, and FI-score increased by 1.93% over the second-best model (COVID-Net [34]).

Moreover, the confusion matrix is used to demonstrate the performance of the PCSANet model in the COVIDx dataset, as shown in Fig.8. The PCSANet model only misclassified 12 COVID-19 cases, and it can correctly predict 188 COVID-19 cases. At the same time, only 1 normal case was misclassified as COVID-19 case. The experimental results on the COVIDx dataset confirm that the PCSANet model can robustly handle various complex thoracic diseases on the chest X-ray image.

D. ABLATION STUDY

To evaluate the effectiveness of the pyramidal convolution module and shuffle attention module, ablation studies are conducted on the ChestX-ray14 dataset. The PCSANet model uses ResNet-50 as the backbone network and combines the pyramidal convolution module and shuffle attention module. We remove the pyramidal convolution module and the shuffle attention module from the PCSANet model in turn. The “backbone” indicates the backbone network ResNet-50, and the PCSANet model indicates the “complete” model, “w/o pyconv” means that only the shuffle attention module is added to the backbone network, and “w/o attention” means that only the pyramidal convolution module is added in the backbone network. The average AUC scores are shown in Table.4. The corresponding

feature heat map is shown in Fig.9. It is obvious that the AUC scores of the PCSANet model and “w/o attention” are very similar, and “w/o attention” has a slight advantage over the PCSANet model in 6 types of thoracic diseases: “Cardiomegaly”, “Effusion”, “Mass”, “Nodule”, “Pneumothorax”, “Emphysema”. We think that the shuffle attention not only focuses on important discriminative features of pathological abnormality but also focuses on some useless features.

E. VISUALIZATION

The section demonstrates the feature heat map and the classification results are shown in Fig.9 and Table.5. The feature heat map is generated in two steps: first, the absolute value of the feature value at each location is obtained from a specific layer (the conv_5 layer of ResNet-50), and then the maximum value along the feature channel is calculated. It can be observed from Fig.9 that local regions of the image are activated. This indicates that the PCSANet model can learn to focus on the pathological abnormality and thus be able to identify the disease accurately. There are the first 8 probability scores for each sample in Table.5. The real labels are highlighted in red or blue. It can be seen that there is a big gap between the scores of real thoracic disease and other thoracic diseases. For example, the predicted score of “Atelectasis” (row 1, column 1) is 0.8886, about 97 times that of “Cardiomegaly” (0.0092). Only in some special

cases (highlighted in blue) is the PCSANet model unable to accurately identify the thoracic disease.

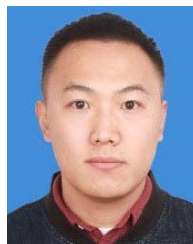
V. CONCLUSION

In this paper, we propose the PCSANet model for multi-label chest X-ray image classification. Numerous experiments have shown that the PCSANet model can effectively classify chest X-ray images with the AUC score that can reach 82.5% on the Chest X-ray14 dataset. The experiment results also demonstrate that our PCSANet model is capable of extracting multi-scale discriminative features of various pathological abnormalities using pyramidal convolution, focusing on more pathological abnormality features using shuffle attention. In the future, the research will be continued in two directions. First, weakly supervised precision localization methods will be investigated due to the expensive bounding box. Second, image segmentation methods are used to enable a better combination of global and local features to achieve more accurate localization and identification of pathological abnormality.

REFERENCES

- [1] (Oct. 2021). *Pneumonia Can be Prevented-Vaccines Can Help*. National Center for Immunization and Respiratory Diseases. [Online]. Available: <https://www.cdc.gov/pneumonia/prevention.html>
- [2] H. Chen, C. Shen, J. Qin, D. Ni, L. Shi, J. C. Cheng, and P.-A. Heng, "Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Nov. 2015, pp. 515–522.
- [3] M. A. Bruno, E. A. Walker, and H. H. AbuJudeh, "Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction," *Radiographics*, vol. 35, no. 6, pp. 1668–1676, 2015.
- [4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 2097–2106.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.
- [6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2016, pp. 770–778.
- [9] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-Rays with deep learning," 2017, *arXiv:1711.05225*.
- [10] G. Huang, Z. Liu, and L. van der Maaten, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [11] H. Wang, D. Zhang, S. Ding, Z. Gao, J. Feng, and S. Wan, "Rib segmentation algorithm for X-ray image based on unpaired sample augmentation and multi-scale network," *Neural Comput. Appl.*, pp. 1–15, Sep. 2021.
- [12] M. F. Rahman, Y. Zhuang, T.-L. Tseng, M. Pokojovy, P. McCaffrey, E. Walsler, S. Moen, and A. Vo, "Improving lung region segmentation accuracy in chest X-ray images using a two-model deep learning ensemble approach," *J. Vis. Commun. Image Represent.*, vol. 85, May 2022, Art. no. 103521.
- [13] K. Kondo, M. Ishii, S. Fujimoto, M. Tanaka, M. Kiyono, H. Itoh, and H. Kimura, "Chest X-ray anomaly detection based on changes in anatomical structures due to disease," *Med. Imag. Technol.*, vol. 39, no. 5, pp. 229–242, Nov. 2021.
- [14] C. Xue, L. Zhu, H. Fu, X. Hu, X. Li, H. Zhang, and P.-A. Heng, "Global guidance network for breast lesion segmentation in ultrasound images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101989.
- [15] M. A. Khan, K. Muhammad, M. Sharif, T. Akram, and V. H. C. D. Albuquerque, "Multi-class skin lesion detection and classification via teledermatology," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 12, pp. 4267–4275, Dec. 2021.
- [16] Q. Guan, Y. Huang, Y. Luo, P. Liu, M. Xu, and Y. Yang, "Discriminative feature learning for thorax disease classification in chest X-ray images," *IEEE Trans. Image Process.*, vol. 30, pp. 2476–2487, 2021.
- [17] H. Wang, S. Wang, Z. Qin, Y. Zhang, R. Li, and Y. Xia, "Triple attention learning for classification of 14 thoracic diseases using chest radiography," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101846.
- [18] S. Z. Y. Zaidi, M. U. Akram, A. Jameel, and N. S. Alghamdi, "Lung segmentation-based pulmonary disease classification using deep neural networks," *IEEE Access*, vol. 9, pp. 125202–125214, 2021.
- [19] B. Chen, J. Li, G. Lu, and D. Zhang, "Lesion location attention guided network for multi-label thoracic disease classification in chest X-rays," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 2016–2027, Jul. 2020.
- [20] Y. Mao, Y. He, L. Liu, and X. Chen, "Disease classification based on synthesis of multiple long short-term memory classifiers corresponding to eye movement features," *IEEE Access*, vol. 8, pp. 151624–151633, 2020.
- [21] A. Bessadok, M. A. Mahjoub, and I. Rekek, "Brain graph synthesis by dual adversarial domain alignment and target graph prediction from a source graph," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101902.
- [22] M. Blendowski, L. Hansen, and M. P. Heinrich, "Weakly-supervised learning of multi-modal features for regularised iterative descent in 3D image registration," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101822.
- [23] F. Zhu, X. Zhu, Z. Huang, M. Ding, Q. Li, and X. Zhang, "Deep learning based data-adaptive descriptor for non-rigid multi-modal medical image registration," *Signal Process.*, vol. 183, Jun. 2021, Art. no. 108023.
- [24] Z. Huang, J. Lin, L. Xu, H. Wang, T. Bai, Y. Pang, and T.-H. Meen, "Fusion high-resolution network for diagnosing ChestX-ray images," *Electronics*, vol. 9, no. 1, p. 190, Jan. 2020.
- [25] Q. Guan and Y. Huang, "Multi-label chest X-ray image classification via category-wise residual attention learning," *Pattern Recognit. Lett.*, vol. 130, pp. 259–266, Feb. 2020.
- [26] P. Kumar, M. Grewal, and M. M. Srivastava, "Boosted cascaded Convnets for multilabel classification of thoracic diseases in chest radiographs," in *Proc. Int. Conf. Image Anal. Recognit. (ICIAR)*. Cham, Switzerland: Springer, Jun. 2018, pp. 546–552.
- [27] B. Chen, J. Li, G. Lu, H. Yu, and D. Zhang, "Label co-occurrence learning with graph convolutional networks for multi-label chest X-ray image classification," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 8, pp. 2292–2302, Aug. 2020.
- [28] N.-U. Rehman, M. S. Zia, T. Meraj, H. T. Rauf, R. Damaševičius, A. M. El-Sherbeeney, and M. A. El-Meligy, "A self-activated CNN approach for multi-class chest-related COVID-19 detection," *Appl. Sci.*, vol. 11, no. 19, p. 9023, Sep. 2021.
- [29] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei, "Thoracic disease identification and localization with limited supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8290–8299.
- [30] E. Rozenberg, D. Freedman, and A. A. Bronstein, "Learning to localize objects using limited annotation, with applications to thoracic diseases," *IEEE Access*, vol. 9, pp. 67620–67633, 2021.
- [31] Y. Ma, Q. Zhou, X. Chen, H. Lu, and Y. Zhao, "Multi-attention network for thoracic disease classification and localization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1378–1382.
- [32] C. Yan, J. Yao, R. Li, Z. Xu, and J. Huang, "Weakly supervised deep learning for thoracic disease classification and localization on chest X-rays," in *Proc. ACM Int. Conf. Bioinf., Comput. Biol., Health Informat.*, Aug. 2018, pp. 103–110.
- [33] X. Ouyang, S. Karanam, Z. Wu, T. Chen, J. Huo, X. S. Zhou, Q. Wang, and J.-Z. Cheng, "Learning hierarchical attention for weakly-supervised chest X-ray abnormality localization and diagnosis," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2698–2710, Oct. 2021.

- [34] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020.
- [35] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*.
- [36] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132665–132676, 2020.
- [37] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. A. Maadeed, S. M. Zughaier, M. S. Khan, and M. E. H. Chowdhury, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, May 2021, Art. no. 104319.
- [38] M. de la Iglesia Vayá, J. M. Saborit, J. A. Montell, A. Pertusa, A. Bustos, M. Cazorla, J. Galant, X. Barber, D. Orozco-Beltrán, F. García-García, M. Caparrós, G. González, and J. M. Salinas, "BIMCV COVID-19+: A large annotated dataset of RX and CT images from COVID-19 patients," 2020, *arXiv:2006.01174*.
- [39] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest X-ray classification," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Dec. 2019.
- [40] L. Yao, J. Prosky, E. Poblenz, B. Covington, and K. Lyman, "Weakly supervised medical diagnosis and localization from multiple resolutions," 2018, *arXiv:1803.07703*.
- [41] H. Wang, H. Jia, L. Lu, and Y. Xia, "Thorax-Net: An attention regularized deep neural network for classification of thoracic diseases on chest radiography," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 475–485, Feb. 2020.
- [42] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.
- [43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [44] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101794.
- [45] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, and V. Singh, "Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVNet," *Chaos, Solitons Fractals*, vol. 138, Sep. 2020, Art. no. 109944.
- [46] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103792.



KAI CHEN was born in Shandong, China. He received the degree from the Weifang Vocational College of Engineering, Weifang, China, in 2016. He is currently pursuing the M.S. degree in electronic information with Xijing University, Xi'an, China.

His research interests include pattern recognition and medical image processing.



XUQI WANG received the B.S. degree in computer science and technology from Xidian University, Shaanxi, China, in 1999, the M.S. degree in software engineering from Xi'an Jiaotong University, Shaanxi, in 2005, and the Ph.D. degree in communication and information system from the School of Mechanical Electronic and Information Engineering, China University of Mining and Technology (Beijing), Beijing, China, in 2020.

He is currently an Associate Professor with Xijing University. His research interests include wireless sensor networks and artificial intelligence.



SHANWEN ZHANG was born in Shanxi, China. He received the B.S. degree in mathematics from Northwest University, China, in 1988, the M.S. degree in applied mathematics from Northwest Polytechnic University, China, in 1995, and the Ph.D. degree in electromagnetic field and microwave from Air Force Engineering University, China, in 2001.

He is currently a Professor with Xijing University. He is also a Visiting Scholar with the Department of Computer Science, Virginia Tech. His research interests include machine learning and its application in data mining, including machine learning, leaf image processing, data reduction, data mining, feature selection, wavelet transforms, and their application in the sonar image recognition.

• • •