PAM-DenseNet: A Deep Convolutional Neural Network for Computer-Aided COVID-19 Diagnosis

Bin Xiao, Zeyu Yang, Xiaoming Qiu, Jingjing Xiao, Guoyin Wang, *Senior Member, IEEE*, Wenbing Zeng, Weisheng Li, Yongjian Nian, and Wei Chen

***Abstract*—Currently, several convolutional neural network (CNN)-based methods have been proposed for computer-aided COVID-19 diagnosis based on lung computed tomography (CT) scans. However, the lesions of pneumonia in CT scans have wide variations in appearances, sizes, and locations in the lung regions, and the manifestations of COVID-19 in CT scans are also similar to other types of viral pneumonia, which hinders the further improvement of CNN-based methods. Delineating infection regions manually is a solution to this issue, while excessive workload of physicians during the epidemic makes it difficult for manual delineation. In this article, we propose a CNN called dense connectivity network with parallel attention module (PAM-DenseNet), which can perform well on coarse labels without manually delineated infection regions. The parallel attention module automatically learns to strengthen informative features from both channelwise and spatialwise simultaneously, which can make the network pay more attention to the infection regions without any manual delineation. The dense connectivity structure performs feature maps reuse by introducing direct connections from previous layers to all subsequent layers, which can extract representative features from fewer CT slices. The proposed network is first trained on 3530 lung CT slices selected from 382 COVID-19 lung CT scans, 372 lung CT scans infected by other pneumonia, and 200 normal lung CT scans to obtain a pretrained model for slicewise prediction. We then apply this pretrained model to a CT scans dataset containing 94 COVID-19 CT scans, 93 other pneumonia CT scans, and 93 normal lung scans, and achieve patientwise prediction through a voting mechanism. The experimental results show that the proposed network achieves promising results with an accuracy of 94.29%, a precision of 93.75%, a sensitivity of 95.74%, and a specificity of 96.77%, which is comparable to the methods that are based on manually delineated infection regions.**

***Index Terms*—Computer-aided diagnosis, convolutional neural network (CNN), coronavirus disease 2019 (COVID-19), lung computed tomography (CT) scans.**

Manuscript received 18 August 2020; revised 14 November 2020; accepted 2 December 2020. Date of publication 24 August 2021; date of current version 17 October 2022. This work was supported in part by the National Key Research and Development Project under Grant 2016YFC1000307-3 and Grant 2019YFE0110800; in part by the National Natural Science Foundation of China under Grant 61976031; in part by the Chongqing Research Program of Application Foundation Advanced Technology under Grant cstc2018jcyjAX0117; and in part by the Scientific Technological Key Research Program of Chongqing Municipal Education Commission under Grant KJZD-K201800601. *(Bin Xiao and Zeyu Yang are co-first authors.)* This article was recommended by Associate Editor L. Jiao. *(Corresponding authors: Bin Xiao; Yongjian Nian; Wei Chen.)*

Bin Xiao, Zeyu Yang, Guoyin Wang, and Weisheng Li are with the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: xiaobin@cqupt.edu.cn; yangzy0215@163.com; wanggy@cqupt.edu.cn; liws@cqupt.edu.cn).

Xiaoming Qiu is with the Department of Radiology, Huangshi Central Hospital (Affiliated Hospital of Hubei Polytechnic University) Edong Healthcare Group, Huangshi 435002, China (e-mail: qxm2020cov@163.com).

Jingjing Xiao is with the Xinqiao Hospital, Army Medical University, Chongqing 400038, China (e-mail: shine636363@sina.com).

Wenbing Zeng is with the Department of Radiology, Chongqing Three Gorges Center Hospital, Chongqing 404000, China (e-mail: zengwenbing123@163.com).

Yongjian Nian is with the School of Biomedical Engineering and Imaging Medicine, Army Medical University, Chongqing 400038, China (e-mail: yjnian@tmmu.edu.cn).

Wei Chen is with the Department of Radiology, Southwest Hospital, Army Medical University, Chongqing 400038, China (e-mail: landcw@hotmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2020.3042837.

Digital Object Identifier 10.1109/TCYB.2020.3042837

I. Introduction

CORONAVIRUS disease 2019 (COVID-19) is a novel coronavirus disease, which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1], and has become a global pandemic that broke out in early 2020. Patients usually present respiratory infection symptoms in the early stage, some rapidly developing acute respiratory distress syndrome, acute respiratory failure, and other serious complications [2]. It has infected about 18 902 735 people and caused about 709 511 deaths in more than 200 countries up to August 7, 2020 [3]. Presently, the diagnosis methods of COVID-19 mainly include medical imaging methods, such as computed tomography (CT) [4], [5] and chest X-ray [6], [7], and an etiological method, such as reverse-transcriptase polymerase chain reaction (RT-PCR) [8], [9]. Currently, the real-time RT-PCR is regarded as the gold standard for confirmation of the SARS-CoV-2 infection [10]. However, some studies have shown that the diagnostic sensitivity of RT-PCR is lower than lung CT especially in the initial presentation of COVID-19 [10], [11]. It is essential to consider CT as a primary tool for the current COVID-19 detection [12]. But observation of the lung CT scans by clinicians is time consuming and laborious, especially with the current surge in the number of patients, which greatly increases the workload of radiologists and reduces the diagnostic efficiency. Therefore, for fast and accurate detection, we aim to develop a computer-aided COVID-19 diagnosis method based on lung CT scans.

In recent years, machine learning methods have been used in the computer-aided diagnosis of pneumonia, such as the support vector machine (SVM) [13], [14]; the k-nearest neighbors algorithm (KNN) [15], [16]; the Bayesian method [17], etc. However, these traditional machine learning methods usually

require a complex procedure of handcrafted feature extraction and dimension reduction. By the deep learning methods, the convolutional neural network (CNN) can automatically learn features from high-dimensional medical images in an end-to-end multilevel manner, which does not require handcrafted feature extraction [18]. The CNN extracts high-level features progressively via several layers of feature representations [19], and has achieved outstanding performance in many image processing tasks [20]–[22] and medical image analysis [23]–[26], as well as in the computer-aided COVID-19 diagnosis [27]–[32].

Currently, most existing CNN-based studies on computer-aided COVID-19 diagnosis have made some achievements, but the lesions of pneumonia in CT scans have wide variations in appearances, sizes, and locations in the lung regions, and the manifestations of COVID-19 in CT scans are also similar to other types of viral pneumonia, which hinders the further improvement of CNN-based methods. To enable the network to focus more on the infection regions and improve the classification accuracy, some researchers first delineate infection regions manually, then they extract patches from lung CT scans according to the delineation and input these patches into the network for training to improve recognition performance. However, manual delineation is time consuming and laborious, which is made more difficult by the excessive workload of physicians during the epidemic. Therefore, the researchers usually obtain lung CT scans with coarse labels that lack manual delineation.

In this article, we propose a CNN-based method called dense connectivity network with parallel attention module (PAM-DenseNet) that uses a dense connectivity structure and attention module to distinguish COVID-19 from other types of pneumonia and normal lung cases, which can perform well on coarse labels without manually delineated infection regions. The attention module is a mechanism that enables the network to enhance the extraction of informative features through self-learning [23], [33]. We propose an attention module called the parallel attention module (PAM). It contains channel attention and spatial attention that can learn to strengthen or suppress features from both channelwise and spatialwise simultaneously. The PAM makes the network pay more attention to the infection regions in lung CT scans, which can improve the performance without any manual delineation. The dense connectivity structure is introduced in densely connected convolutional networks (DenseNets) [34], which has been experimentally demonstrated to be particularly well suited for small-scale datasets [35]. It can perform effective feature map reuse by introducing direct connections from previous layers to all subsequent layers, which not only benefits gradient backpropagation in the deep network but also extracts representative features from fewer CT samples.

To achieve patientwise prediction, we first train the PAM-DenseNet on selected lung CT slices to obtain a slicewise prediction model. Then, we use this pretrained model to predict each slice in a CT scan of a patient, and use a voting mechanism to synthesize the slicewise prediction results to output the final patientwise prediction results. The lung CT slices dataset contains 3530 slices selected from 382 COVID-19 lung CT scans, 372 lung CT scans infected by influenza-B virus pneumonia or bacterial pneumonia, and 200 normal lung CT scans. The CT scans dataset includes 94 COVID-19 CT scans, 93 other pneumonia CT scans containing influenza-B virus pneumonia and bacterial pneumonia cases, and 93 normal lung scans. The experimental results show that the proposed network achieves promising results with an accuracy of 94.29%, a precision of 93.75%, a sensitivity of 95.74%, and a specificity of 96.77%, which is comparable to the methods that are based on manually delineated infection regions.

The remainder of this article is organized as follows. We briefly introduce the most relevant studies in Section II. In Section III, we present the proposed PAM-DenseNet in detail. The dataset, experimental settings, comparative methods, experimental results, and analysis are provided in Section IV. Finally, we conclude this article in Section V.

II. Related Work

Recently, many CNN models have been proposed in computer-aided COVID-19 diagnosis based on lung CT scans. Shi *et al.* [36] reviewed and concluded several recent studies in his paper. Song *et al.* [28] constructed a CNN model based on a pretrained ResNet [37] with a feature pyramid network [38] to recognize COVID-19 on 777 COVID-19 CT scans, 505 bacterial pneumonia CT scans, and 708 normal CT scans, and they achieved an accuracy of 86%. They also test VGG [39] on their dataset which achieved an accuracy of 84%. Xu *et al.* [29] first used a CNN-based segmentation model pretrained from a pulmonary CT image dataset to extract candidate abnormal patches, then they used a ResNet to perform classification on 2301 COVID-19 patches, 2244 influenza-A viral pneumonia patches, and 5616 no infection patches with an overall accuracy of 86.7%. Chen *et al.* [30] first manually delineated infection regions on CT scans by researcher, then they used the U-Net++ [40] to segment COVID-19- related lesions and detect COVID-19 on 4382 COVID-19 CT slices and 9369 non-COVID-19 CT slices. The evaluation results of COVID-19 classification using the proposed model are 95.2% (accuracy), 100% (sensitivity), and 93.6% (specificity). Zheng *et al.* [41] proposed a U-Net for lung infection regions segmentation and used a 3-D CNN model for predicting the probability of COVID-19, and the proposed model achieves a sensitivity of 90.7% and a specificity of 91.1% on 313 COVID-19 subjects and 229 other subjects. Wang *et al.* [31] sketched the region of interest on CT slices manually and used a CNN model with inception structure [42] to diagnose COVID-19 on 195 COVID-19 CT scans and 258 non-COVID-19 CT scans, and showed a total accuracy of 73.1%, along with a specificity of 67.0% and a sensitivity of 74.0%. Shan *et al.* [27] introduced a CNN-based segmentation model to detect COVID-19 infection regions with manual delineation of physicians and obtained 91.6% Dice similarity coefficients between automatic and manual segmentations. Jin *et al.* [43] used a 2-D CNN model to segment the lung image and then identified slices of COVID-19 cases, and the sensitivity of 94.1% and the specificity of 95.5% are achieved on 496 COVID-19 subjects and

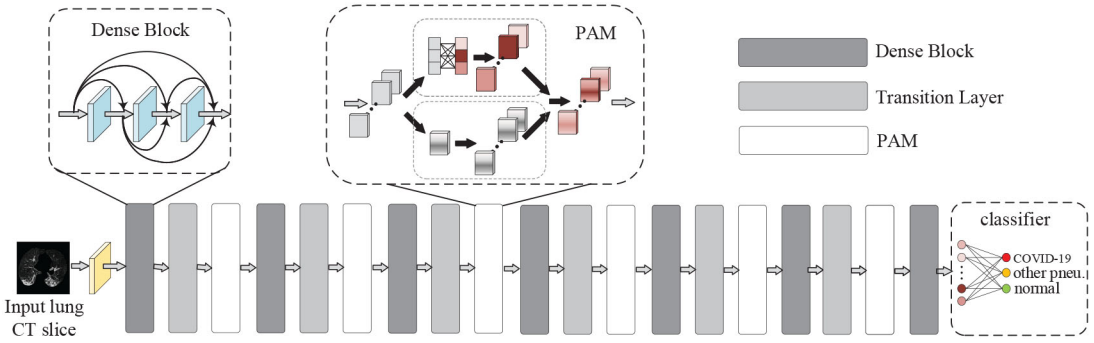


Fig. 1. Framework of the PAM-DenseNet. The PAM-DenseNet is a cascaded structure of dense block, transition layer, and PAM. The dense block extracts features from CT slices through a dense connectivity structure, and the transition layer performs downsampling, then the PAM self-recalibrates features from channelwise and spatialwise. A fully connected layer with softmax operation outputs the prediction as a classifier. The "other pneu." stands for other types of pneumonia.

1385 other subjects. Jin *et al.* [44] used a segmentation model based on U-Net++ to extract lesion regions. This is followed by a ResNet-50 classification model on 723 COVID-19 subjects and 413 other subjects, and the final sensitivity is 97.4% and specificity is 92.2%, respectively. Li *et al.* [45] used a ResNet-50 model to discriminate COVID-19 from community-acquired pneumonia and nonpneumonia on 468 COVID-19 subjects, 1551 other subjects, and 1445 nonpneumonia subjects, and the results show a sensitivity of 90% and a specificity of 96%, etc. Many studies used manually delineated infection regions to improve the recognition accuracy due to the variability of infection regions that hampers network performance, but it also increases the workload of radiologists, especially in the current period of severe epidemic. Therefore, our goal is to develop a well-performed CNN model for computer-aided COVID-19 diagnosis based on lung CT slices without any manual delineation of lung infection regions, and to utilize this model to achieve patientwise prediction through a voting mechanism.

III. Proposed PAM-DenseNet

In this article, we propose a CNN called the dense connectivity network with PAM (PAM-DenseNet) for the slicewise classification of COVID-19, other pneumonia, and normal cases. As illustrated in Fig. 1, the PAM-DenseNet mainly employs PAM and the dense connectivity structure. First, an initial 7×7 convolution is performed on the input CT slice, and the feature maps are input into a dense block to extract features by the dense connectivity structure. The dense connectivity structure iteratively concatenates the input features with the output features and enables each convolution layer to receive raw information from all previous layers, which can realize the reuse of feature maps so that more features can be extracted from fewer CT slices. After that, a transition layer is introduced to perform downsampling. The transition layer consists of a batch normalization (BN), a rectified linear unit (ReLU), 1×1 convolution, and an average pooling layer. Then, a PAM is followed to implement channel attention and spatial attention on the feature maps to make the network pay more attention to the lung infection regions from "what" and "where", respectively. At the end of the last dense block, the feature maps are sent to a fully connected layer after applying the global average pooling operation, followed by a softmax classifier that outputs the prediction. The input CT slice and feature maps in all convolution layers are zero padded to keep the size fixed. The whole PAM-DenseNet employs seven dense blocks and six PAMs. The dense block and PAM will be introduced in the following sections.

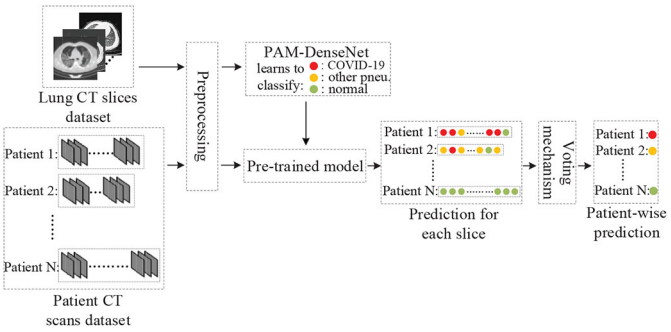


Fig. 2. Process of predicting patientwise labels. We first train the PAM-DenseNet model on the lung CT slices dataset, then apply this pretrained model to the patient CT scans dataset, and obtain the patientwise prediction through the voting mechanism. The "other pneu." stands for other types of pneumonia.

The process of achieving patientwise prediction is shown in Fig. 2. First, we construct a lung CT slices dataset where slices are selected by physicians from COVID-19 CT scans, other types of pneumonia CT scans, including influenza-B virus pneumonia and bacterial pneumonia, and CT scans of normal lungs. Then, the lung CT slices dataset is preprocessed, which mainly includes using a pretrained U-Net model to segment lung tissue and normalization. We use this lung CT slices dataset to train our PAM-DenseNet and obtain a pretrained model to distinguish COVID-19 CT slices, other types of pneumonia CT slices, and normal lung CT slices. We then use the pretrained model to predict the CT slices of each patient CT scan. The patient CT scans dataset contains COVID-19 cases, other types of pneumonia cases, and normal cases. We perform the same preprocessing operation on the patient CT scans dataset. After we obtain the CT slices prediction labels for each patient, we use a voting mechanism to synthesize these slicewise labels to obtain a patientwise

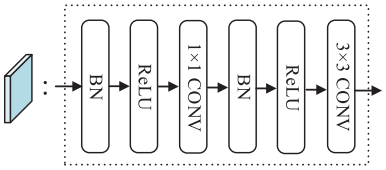


Fig. 3. Structure of the dense layer.

prediction label. The voting mechanism we use is that when all CT slices of a patient are predicted to be normal, the label of the patient is predicted to be normal. When there are slices predicted to be COVID-19 or other pneumonia, the category with a larger number of slices is taken as the final label of the patient. When the number of slices predicted to be COVID-19 is the same as that of other pneumonia, the prediction probabilities of these two types of slices are summed, respectively, and the larger one is the final prediction label of the patient. The above datasets and preprocessing methods will be described in Section IV.

A. Dense Block

The dense block provides the function of extracting features, and each dense block is composed of three layers that are connected densely. The dense layer shown in Fig. 3 is designed to act as a composite of six operations [34]: 1) BN; 2) a ReLU; 3) a 1×1 convolution (CONV); 4) BN; 5) a ReLU; and 6) 3×3 convolution. For reducing the number of input feature maps, 1×1 convolution is introduced as a bottleneck layer before each 3×3 convolution, which can improve the computational efficiency. The number of feature map channels output by each dense layer is called the growth rate. The 1×1 bottleneck convolution first compresses the number of input feature map channels to equal the growth rate, and then the 3×3 convolution further extracts features from the feature maps and keeps the number of channels unchanged.

The dense block in the PAM-DenseNet is illustrated in Fig. 4. Three dense layers are used in each dense block. Let $F_\ell(\cdot)$ denote a dense layer transformation, where ℓ indexes the dense layer, and denote the output of the ℓth dense layer as X_ℓ. Within each dense block, for improving the information flow between dense layers, the dense block utilizes direct connections from any dense layer to all subsequent dense layers. That is, the ℓth dense layer receives the feature maps of all preceding dense layers, $X_0, \ldots, X_{\ell-1}$

$$X_\ell = F_\ell\left(\left[X_0, X_1, \ldots, X_{\ell-1}\right]\right) \tag{1}$$

where $[X_0, X_1, \ldots, X_{\ell-1}]$ denotes the concatenation of feature maps produced in the dense layers $0, \ldots, \ell - 1$. This kind of dense connectivity structure achieves the reuse of feature maps, which can mine more information from the limited CT slices to improve classification accuracy.

B. Parallel Attention Module

To improve the performance of classification without manually delineated infection regions of lungs, a natural idea is that if the network can autonomously pay more attention to the infection regions, the network will improve its ability to distinguish the interclass variations between COVID-19 and other pneumonia, followed by improving the classification accuracy. The attention module in CNNs mainly includes channel attention and spatial attention. It is a mechanism that can make the network strengthen the critical information adaptively according to different tasks. Recently, many attention modules have been proposed, such as the squeeze-and-excitation networks (SE block) [33], the bottleneck attention module (BAM) [46], the convolutional block attention module (CBAM) [47], SCA-CNN [48], etc. The idea of these attention modules is to aggregate and self-recalibrate the feature maps via channelwise and spatialwise. For distinguishing the variations of infection regions in the lung CT slices more effectively, we propose a PAM in our network, which employs a channel attention module and a spatial attention module (SAM) in a parallel manner.

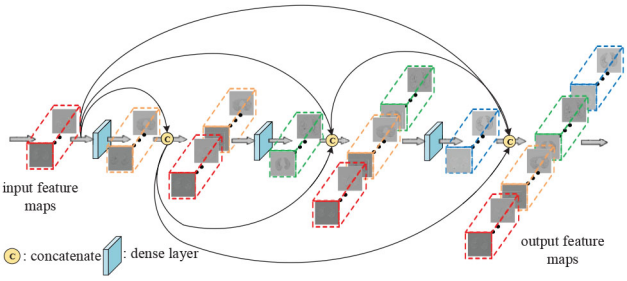


Fig. 4. Structure of the dense block. The output feature maps of each dense layer are densely concatenated to preform feature reuse.

The structure of PAM is illustrated in Fig. 5. For a given intermediate feature map $M \in \mathbb{R}^{C\times H\times W}$, PAM simultaneously infers a 1-D channel attention map $A_c \in \mathbb{R}^{C\times 1\times 1}$ and a 2-D spatial attention map $A_s \in \mathbb{R}^{1\times H\times W}$. However, to obtain the weights that indicate the importance of the feature, the attention map usually needs to be processed by sigmoid function to compress the weights between zero and one. Multiplying the feature map with the two attention maps will cause a significant decrease in value, which will lead to performance drop. Therefore, we first do an elementwise addition operation on A_c and A_s before sigmoid, and then use the sigmoid function to obtain a weight map indicating the information that should be enhanced. Considering that the parameters of the PAM in the early stage of network training have not been trained well, the weight map generated at this time is also inaccurate, and direct multiplication to the original feature map will cause problems, such as the decline of network performance and difficulty in training. To ease the above problems, we design identical mapping to enable the network to bypass the effect of the weight map on the original feature map. The weight map is elementwise multiplied with the original feature map M and the results then elementwise add with the feature map M to obtain the final channel and spatial refined feature map $M' \in \mathbb{R}^{C\times H\times W}$. The overall process of PAM can be summarized as

$$M' = M \oplus M \otimes [\sigma(A_c \oplus A_s)] \tag{2}$$

where $\otimes$ denotes the elementwise multiplication, σ denotes a sigmoid function, and $\oplus$ denotes the elementwise addition.

Since the two attention maps A_c and A_s have different shapes, we expand the attention maps to $\mathbb{R}^{C\times H\times W}$ before adding them.

We analyze the PAM from both forward and backward processes. Given a dense block output $D(x,\theta)$ with input x, θ is the trainable parameter in the dense block. Then, $D(x,\theta)$ is fed into the channel attention module and the SAM to calculate the weight map $A(D(x,\theta),\phi)$, where ϕ is the trainable parameter in the PAM. The final output y of PAM is computed as

$$y = A(D(x,\theta),\phi)\cdot D(x,\theta) + D(x,\theta). \tag{3}$$

During the forward inference, the weight map $A(D(x,\theta),\phi)$ serves as a soft feature selector to indicate "what" and "where" to attend in the channelwise and spatialwise, respectively, which can help the network learn the feature that needs to be emphasized or suppressed. During the back propagation, the PAM can also serve as a gradient update filter. The gradient of the dense block is

$$\begin{aligned}\frac{\partial y}{\partial \theta} = &\frac{\partial A(D(x,\theta),\phi)}{\partial D(x,\theta)}\cdot\frac{\partial D(x,\theta)}{\partial \theta}\cdot D(x,\theta)\\ &+A(D(x,\theta),\phi)\cdot\frac{\partial D(x,\theta)}{\partial \theta}+\frac{\partial D(x,\theta)}{\partial \theta}.\end{aligned} \tag{4}$$

We can find that the weight map $A(D(x,\theta),\phi)$ also participates in the gradient propagation process of the previous dense block, and to some extent has played a role in selectively strengthening or suppressing the gradient.

1) Channel Attention Module: The structure of the channel attention module is illustrated in Fig. 6. To obtain channelwise attention information, we adopt the SE block structure and make slight adjustments to adapt to our network structure. The SE block is an effective channel attention mechanism proposed in squeeze-and-excitation networks [33], which can effectively recalibrate channelwise feature responses by explicitly modeling interdependencies between channels. For a given feature map $M \in \mathbb{R}^{C\times H\times W}$, we first use average pooling to aggregate spatial information for each channel, generating a spatial context descriptor $M^c_{\text{avg}} \in \mathbb{R}^{C\times 1\times 1}$, which denotes the average-pooled features. Then, the descriptor is forwarded to a multilayer perceptron (MLP) with one hidden layer to produce the channel attention map $A_c \in \mathbb{R}^{C\times 1\times 1}$. To lower the number of parameters in the attention module, the size of the hidden layer in MLP is set to $\mathbb{R}^{C/r\times 1\times 1}$, where r is the reduction ratio. The channel attention module is computed as

$$\begin{aligned}A_c &= \text{MLP}(\text{Avgpool}(M))\\ &= W_1\left(W_0\left(M^c_{\text{avg}}\right)\right)\end{aligned} \tag{5}$$

where MLP$(\cdot)$ denotes the operation of the MLP and AvgPool$(\cdot)$ denotes the average pooling. W_0 and W_1 denote the MLP weights $W_0 \in \mathbb{R}^{C/r\times C}$ and $W_1 \in \mathbb{R}^{C\times C/r}$, respectively. $W_0(\cdot)$ and $W_1(\cdot)$ denote the fully connected operation. Compared with the original SE block, our modification is to cancel the sigmoid in the SE block and make it after the addition of the channel attention map and the spatial attention map, so as to prevent the value of the feature map from dropping significantly.

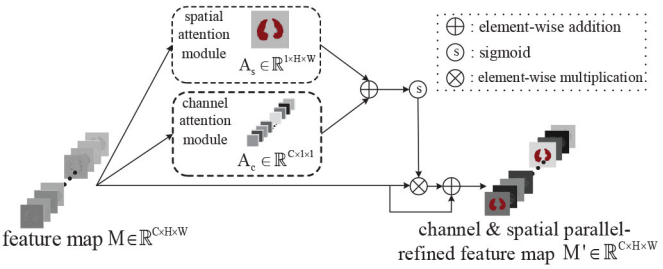


Fig. 5. Structure of the PAM. The channel attention module and SAM are placed in parallel to infer attention maps, and the attention maps are fused and applied to the original feature map by elementwise multiplication.

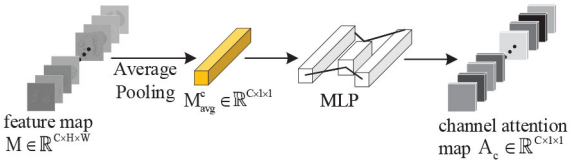


Fig. 6. Structure of the slightly modified SE block.

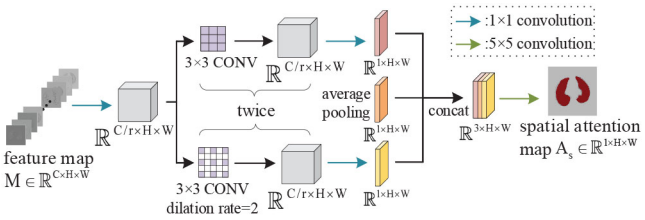


Fig. 7. Structure of the SAM.

2) Spatial Attention Module: The structure of the SAM is illustrated in Fig. 7. Given an intermediate feature map $M \in \mathbb{R}^{C\times H\times W}$ as input, the SAM produces a spatial attention map $A_s \in \mathbb{R}^{1\times H\times W}$ to strengthen or suppress features in different spatial locations, which is complementary to the channel attention. For saving both the number of parameters and computational overhead, we first apply a 1×1 convolution on the input feature map to compress channels, which reduce the dimension of feature map M to $\mathbb{R}^{C/r\times H\times W}$. We use the same reduction ratio r with the channel attention module for simplicity. Utilizing contextual information adequately is crucial to identify which areas of lung CT slices should be focused on. Therefore, we design two different convolutional branches to encode spatial information: one branch adopts two consecutive 3×3 standard convolution operations and the other branch uses two consecutive 3×3 dilated convolution [49] with a dilation rate of 2. These two branches have different priorities: the dilated convolution can enlarge the receptive fields by a sparse convolution kernel, which can leverage contextual information more effectively, while the standard convolution can extract more information about details with a smaller receptive field by a compact convolutional kernel. The dimension of the feature maps in these two branches is maintained at $\mathbb{R}^{C/r\times H\times W}$ and then we integrate and compress the feature maps to $\mathbb{R}^{1\times H\times W}$ by 1×1 convolution, respectively. To aggregate the original spatial information of the input feature map M from spatialwise, we apply the average-pooling operation

along the channel axis to obtain a one-channel feature map M^s_{avg} and concatenate it with the output of the previous two branches to obtain a spatial feature descriptor in dimension $\mathbb{R}^{3\times H\times W}$. Finally, we apply a 5×5 convolution layer on the spatial feature descriptor to generate a spatial attention map $A_s \in \mathbb{R}^{1\times H\times W}$ encoding where to emphasize or suppress. In short, the spatial attention is computed as

$$A_s = f^{5\times 5}\Big[M^s_{\text{avg}}; f_2^{1\times 1}\Big(f_1^{3\times 3}\Big(f_0^{3\times 3}\Big(f_0^{1\times 1}(M)\Big)\Big)\Big); f_3^{1\times 1}\Big(d_1^{3\times 3}\Big(d_0^{3\times 3}\Big(f_1^{1\times 1}(M)\Big)\Big)\Big)\Big] \tag{6}$$

where $f^{k\times k}$ represents a convolution operation with a filter size of $k\times k$, and $d^{3\times 3}$ denotes a dilated convolution operation with a filter size of 3×3. [;] denotes the concatenate operation. Subscripts indicate different convolution operations.

IV. Experiments and Discussion

A. Dataset and Preprocess

This study is based on the reliable lung CT scans that are collected by the cooperative hospitals in Chongqing and Hubei, China. There are a total of 476 COVID-19 lung CT scans and 465 other types of pneumonia lung CT scans including 268 bacterial pneumonia CT scans and 197 influenza-B virus pneumonia CT scans, and 293 normal lung CT scans. The slice thickness of the COVID-19 CT scans contains 1, 2, and 5 mm. The slice thickness of the other types of pneumonia is 2 mm. The slice thickness of the normal lung CT scans contains 1.25, 2, and 5 mm. Based on these CT scans we construct two datasets: 1) a lung CT slices dataset and 2) a patient CT scans dataset. The patient CT scans dataset randomly selects 94 COVID-19 CT scans, 93 other types of pneumonia lung CT scans, including 41 bacterial pneumonia CT scans and 52 influenza-B virus pneumonia CT scans, and 93 normal lung CT scans. Excluding the CT scans of the patient CT scans dataset, we select slices to construct the lung CT slices dataset from the remaining CT scans. The lung CT slices dataset contains 3530 slices, of which 1016 slices are selected from 382 COVID-19 CT scans, 1012 slices are selected from 372 other types of pneumonia CT scans, including 227 bacterial pneumonia CT scans and 145 influenza-B virus pneumonia CT scans, and 1502 slices are selected from 200 normal lung CT scans. In the selection of slices, we pay attention to the slices with typical lesions and take into account the diversity of lung morphology on different slices. These slices only consist of the transverse plane, and slice size is 512×512. The original CT scans contain a lot of content unrelated to lung tissue, such as clothes, bones, etc., which will cause unnecessary interference in the network. Therefore, we first employ a pretrained U-Net model to segment the lung tissue. The pretraining segmentation model we used is proposed by Hofmanniger *et al.* [50], which is trained on a large and diverse lung CT dataset that covers a wide range of visual variability, and our experiments have shown that this model can accurately segment lung tissue in the case of severe lung infection. Several segmentation results are shown in Fig. 8. After segmenting the lung tissue, we normalize these datasets between zero and one.

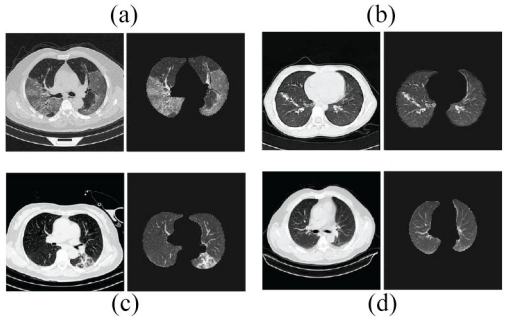


Fig. 8. Several segmentation results of lung CT slices. (a) COVID-19. (b) Influenza-B virus pneumonia. (c) Bacterial pneumonia. (d) Normal lung slice.

B. Evaluation Metric

For evaluating the performance of the proposed PAM-DenseNet, five evaluation metrics are introduced for evaluation: 1) accuracy (Acc); 2) sensitivity (Sen); 3) specificity (Spec); 4) precision (Prec); and 5) F1 score (F1), which are given as follows:

$$\text{Acc} = \frac{\text{Num}_{\text{correct}}}{\text{Num}_{\text{total}}} \tag{7}$$

$$\text{Sen} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{8}$$

$$\text{Spec} = \frac{\text{TN}}{\text{FP}+\text{TN}} \tag{9}$$

$$\text{Prec} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP}+\text{FN}+\text{FP}} \tag{11}$$

where TP denotes the true positive, FP is the false positive, TN is the true negative, and FN stands for false negative. $\text{Num}_{\text{correct}}$ represents the number of samples that are correctly predicted, and $\text{Num}_{\text{total}}$ represents the total number of test samples. The COVID-19, other types of pneumonia, and normal case are regarded as positive class, respectively, in the calculation of Sen, Spec, Prec, and F1 score.

C. Implementation Details

In our PAM-DenseNet, the growth rate of the dense block is 16 and the kernel size of convolution in dense block is 3×3 with one padding. The kernel size of initial convolution is 7×7 with three padding. The reduction ratio of PAM is 8. The implementation details of the PAM-DenseNet are shown in Table I. The PAM-DenseNet is trained from scratch on a Tesla V100 GPU. Adam is used as the optimizer with the learning rate of 3e-4. The cross-entropy function is used as the loss function, and the proposed network is trained for 400 epochs with a batch size of 8. The cross-entropy loss function L_{CE} can be formulated as follows:

$$L_{CE} = -\sum_{i=1}^{3} y_i \log(p_i) \tag{12}$$

where i ranges from 1 to 3 that denotes three categories, y_i represents the ground truth, and p_i denotes the predicted probability.

TABLE I
IMPLEMENTATION DETAILS OF OUR PROPOSED PAM-DENSENET

Layers	Output size
Initial convolution	512×512×32
Dense block	512×512×80
Transition layer, PAM	256×256×80
Dense block	256×256×128
Transition layer, PAM	128×128×128
Dense block	128×128×176
Transition layer, PAM	64×64×176
Dense block	64×64×224
Transition layer, PAM	32×32×224
Dense block	32×32×272
Transition layer, PAM	16×16×272
Dense block	16×16×320
Transition layer, PAM	8×8×320
Dense block	8×8×368
Global average pooling	1×368
Fully connected layer, softmax	3

TABLE II
EVALUATION RESULTS FOR DIFFERENT TYPES OF ATTENTION MODULES

Architecture	Acc	Prec	Sen	Spec	F1
DenseNet	0.9735	0.9680	0.9705	0.9871	0.9693
DenseNet+SE	0.9750	0.9691	0.9755	0.9875	0.9722
DenseNet+SAM	0.9811	**0.9850**	0.9706	**0.9941**	0.9786
DenseNet+PAM	**0.9844**	0.9755	**0.9851**	0.9901	**0.9803**

D. Experiments and Analysis

1) Effectiveness of Parallel Attention Module: To compare the performance of PAM with the single channel attention module and single SAM, we conduct experiments to evaluate PAM, SE block, and the SAM of PAM. We adopt DenseNet as our baseline, and combine the above attention modules to evaluate through five-fold cross-validation on the lung CT slices dataset. The DenseNet utilizes seven dense blocks and six transition layers, three dense layers in each dense block, and the attention modules are placed behind the transition layers. The reduction ratio of the SE block is set to 8. Experimental settings, such as epoch, batch size, etc., are the same as those in Section IV-C above. The test results are shown in Table II, and the receiver operating characteristic curves (ROC) and area under the curve (AUC) are shown in Fig. 9. COVID-19 and other types of pneumonia are regarded as the positive classes, respectively, in the calculation of these evaluation metrics. We can find that compared with the "DenseNet" without attention module, the channel attention module and SAM can effectively improve network performance. The PAM outperforms that of using only the channel attention module or SAM and has a better diagnostic performance for pneumonia, which demonstrates that utilizing both attention modules simultaneously produce a more significant promotion for improving the classification accuracy in computer-aided COVID-19 diagnosis.

To demonstrate the effectiveness of PAM from the perspective of visualization, we apply the gradient-weighted class activation mapping (Grad-CAM) [51] method on the "DenseNet," "DenseNet+SE," "DenseNet+SAM," and "DenseNet+PAM," respectively. The Grad-CAM utilizes the gradient information

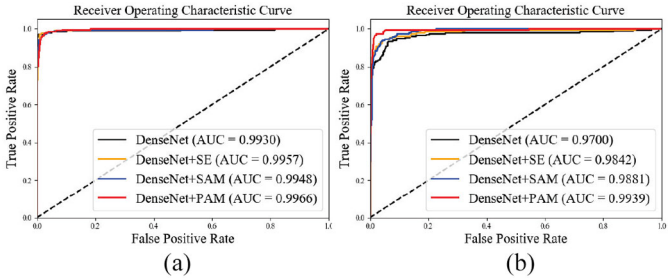


Fig. 9. ROC and AUC of different attention modules. (a) The ROC and AUC of the COVID-19 category. (b) The ROC and AUC of the other types of pneumonia.

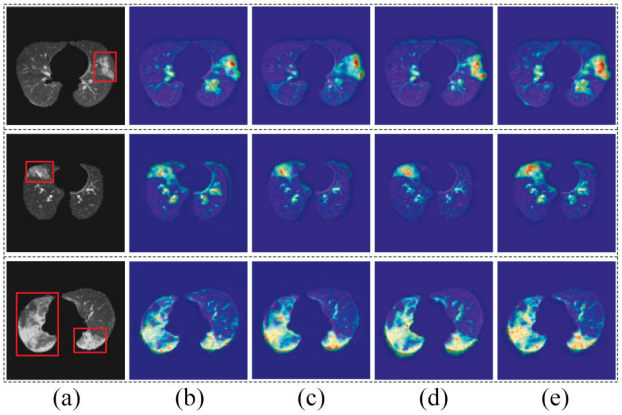


Fig. 10. Class activation maps of the "DenseNet," "DenseNet+SE," "DenseNet+SAM," and "DenseNet+PAM." The warmer color represents a higher value of weights, which denotes that the network is more focused on the regions. (a) input lung CT slices. (b) DenseNet. (c) DenseNet + SE. (d) DenseNet + SAM. (e) DenseNet + PAM.

that has flowed into a convolution layer of the CNNs to understand the importance of each neuron for a decision of interest; thus, the regions focused by the network can be visualized. We randomly select three COVID-19/other types of pneumonia CT slices as input. The visual results are shown in Fig. 10. The three slices are displayed in column Fig. 10(a) and the regions marked by the red rectangle are the infection regions, such as the peripheral ground-glass opacities, lung consolidation, etc. The column in Fig. 10(b)–(e) shows the visualization of the "DenseNet," "DenseNet+SE," "DenseNet+SAM," and "DenseNet+PAM," respectively. We can see that the networks with PAM have a higher value of weights in the relevant infection regions of lungs than the networks with other attention modules. This indicates that the PAM enables the network to pay more attention to the infection regions that are conducive to the pneumonia recognition tasks, thereby improving performance without manual delineation.

To show the effect of PAM more intuitively, we display the intermediate feature maps before and after PAM in Fig. 11. We randomly select three COVID-19/other types of pneumonia CT slices as input, which are shown in the first column. To facilitate visual observation, we chose the first four PAMs with high resolution. We show the change of the feature maps after passing through the PAM, and we can see that the feature response value of the lesion area is significantly enhanced.

2) Experiment of Slice-Wise Classification and Visual Explanation of PAM-DenseNet: We train and compare the

TABLE III
PERFORMANCE COMPARISON AMONG DIFFERENT NETWORKS ON THE LUNG CT SLICES DATASET

Networks	Params. (Million)	Acc	COV vs. (pneu and Nor)				pneu vs. (COV and Nor)				Nor vs. (COV and pneu)			
			Prec	Sen	Spec	F1	Prec	Sen	Spec	F1	Prec	Sen	Spec	F1
VGG16	14.72	0.9448	0.9360	0.9179	0.9739	0.9268	0.8919	0.9340	0.9514	0.9124	0.9929	0.9721	**0.9952**	0.9824
SE-VGG16	14.88	0.9476	0.9140	0.9758	0.9619	0.9439	0.9492	0.8821	0.9798	0.9144	0.9722	0.9756	0.9809	0.9739
PAM-VGG16	15.30	0.9603	0.9356	0.9643	0.9745	0.9497	0.9388	0.9246	0.9763	0.9316	0.9903	0.9807	0.9924	0.9855
VGG19	20.04	0.9462	0.9248	0.9676	0.9653	0.9457	0.9497	0.8673	0.9824	0.9067	0.9601	0.9830	0.9709	0.9714
SE-VGG19	20.19	0.9582	0.9641	0.9586	0.9849	0.9612	0.9422	0.9063	0.9784	0.9239	0.9647	0.9918	0.9727	0.9781
PAM-VGG19	20.61	0.9618	0.9487	0.9439	0.9804	0.9463	0.9347	0.9347	0.9744	0.9347	0.9872	0.9904	0.9899	0.9888
ResNet18	11.17	0.9533	0.9674	0.9630	0.9857	0.9652	0.9508	0.8878	0.9824	0.9182	0.9448	0.9898	0.9587	0.9668
SE-ResNet18	11.35	0.9575	**0.9902**	0.9398	**0.9959**	0.9644	0.9239	0.9286	0.9706	0.9262	0.9572	0.9898	0.9684	0.9732
PAM-ResNet18	11.50	0.9688	0.9810	0.9583	0.9918	0.9696	0.9442	0.9490	0.9784	0.9466	0.9765	0.9898	0.9830	0.9831
ResNet34	21.28	0.9589	0.9733	0.9286	0.9902	0.9504	0.9303	0.9397	0.9724	0.9350	0.9686	0.9904	0.9747	0.9793
SE-ResNet34	21.60	0.9618	0.9841	0.9490	0.9941	0.9662	0.9436	0.9246	0.9783	0.9340	0.9596	0.9936	0.9671	0.9763
PAM-ResNet34	21.61	0.9717	0.9692	0.9643	0.9882	0.9668	0.9497	0.9497	0.9803	0.9497	0.9872	0.9904	0.9899	0.9888
DenseNet	0.41	0.9735	0.9680	0.9705	0.9871	0.9693	0.9686	0.9368	0.9882	0.9562	0.9826	0.9949	0.9868	0.9851
SE-DenseNet	0.48	0.9750	0.9691	0.9755	0.9875	0.9722	0.9628	0.9502	0.9855	0.9564	0.9868	0.9910	0.9901	0.9889
PAM-DenseNet	0.68	**0.9844**	0.9755	**0.9851**	0.9901	**0.9803**	**0.9803**	**0.9660**	**0.9920**	**0.9731**	**0.9933**	**0.9966**	0.9951	**0.9950**

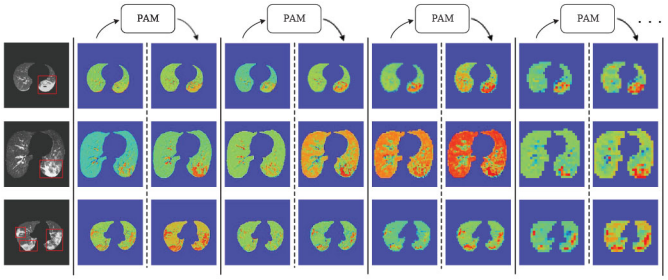


Fig. 11. Visual comparisons of the intermediate feature maps produced by PAM in the "DenseNet+PAM". The infection regions are marked by the red rectangle and warmer color represents a higher response value.

PAM-DenseNet with several typical CNN architectures on the lung CT slices dataset, including VGG16 [39], VGG19 [39], ResNet18 [37], and ResNet34 [37]. For the comparison of the attention module, we combine the PAM or SE block into these networks. We evaluate these networks with five-fold cross-validation on our lung CT slices dataset. The reduction ratio of the SE block is set to 8. Experimental settings, such as epoch, batch size, etc., are the same as those in Section IV-C above. The number of parameters and the classification results are shown in Table III. COVID-19 (COV), other types of pneumonia (pneu), and normal lung (Nor) are regarded as the positive class, respectively, in the calculation of these evaluation metrics. We can find that the PAM-DenseNet combined with the dense connectivity structure and PAM achieves the best performance among all those networks. The PAM performs better than SE block because it has spatial and channel attention modules but the SE block only has channel attention while missing the importance of spatial attention. The PAM arranges the channel attention module and SAM in parallel and simultaneously calculates channel and spatial attention maps to fuse, and achieves better classification performance. For the feature extraction structure, the networks using dense connectivity structure perform significantly better than those with residual connection and plain structure. It is worth noting that DenseNet without any attention module performs better than the PAM-ResNet and PAM-VGG, both of which are with PAM. This also proves the effectiveness of the dense connectivity structure. Moreover, we can find that the networks with dense connection structure have a significantly lower number of parameters. The attention modules can steadily improve the classification performance of networks with a small number of parameters required.

We also conduct experiments to explain the PAM-DenseNet from a visual perspective through Grad-CAM. It is crucial to track the attention of the proposed network and provide valuable information about which lung regions are particularly related to the diagnosis of COVID-19 or other types of pneumonia. To visually demonstrate the effectiveness of the proposed PAM-DenseNet in COVID-19/other pneumonia recognition, and to show the lung regions that the proposed network focuses on, we use Grad-CAM on a trained PAM-DenseNet model to generate the class activation map. We randomly select six samples of COVID-19 and other types of pneumonia from the lung CT slices dataset and feed them into the Grad-CAM to visualize the regions that the networks focused on. As illustrated in Fig. 12, the PAM-DenseNet accurately focuses on the peripheral ground-glass opacities, lung consolidation, etc., which are in accord with clinical studies of COVID-19 [4]. Therefore, the proposed PAM-DenseNet can focus on these infection regions accurately to make predictions, which demonstrates the network's effectiveness, and meanwhile confirms that these regions are associated with COVID-19.

3) Experiment of Patient-Wise Prediction: To achieve patientwise prediction, we first obtain slicewise pretrained models trained from the lung CT slices dataset, and then apply these models to the patient CT scans dataset. The pretrained models predict each slice in a CT scan of a patient, and we use a voting mechanism to synthesize the slicewise prediction results to output the final patientwise prediction. The experimental results are reported in Table IV. We can find that our PAM-DenseNet achieves the best accuracy in the

TABLE IV
PERFORMANCE COMPARISON AMONG DIFFERENT NETWORKS ON THE PATIENT CT SCANS DATASET

Networks	Acc	COV vs.(pneu and Nor)				pneu vs.(COV and Nor)				Nor vs.(COV and pneu)			
		Prec	Sen	Spec	F1	Prec	Sen	Spec	F1	Prec	Sen	Spec	F1
VGG16	0.8393	0.8889	0.9362	0.9409	0.9119	0.7227	0.9247	0.8235	0.8113	0.9839	0.6559	0.9947	0.7871
SE-VGG16	0.8536	0.8936	0.8936	0.9462	0.8936	0.7368	0.9032	0.8396	0.8116	0.9861	0.7634	0.9947	0.8606
PAM-VGG16	0.8839	0.8714	0.9788	0.9221	0.9197	0.8366	0.8979	0.9064	0.8620	0.9855	0.7312	0.9947	0.8395
VGG19	0.8571	0.7603	0.9787	0.8441	0.8558	0.8864	0.8387	0.9465	0.8619	0.9859	0.7527	0.9947	0.8537
SE-VGG19	0.8786	0.9263	0.9362	0.9624	0.9312	0.7699	0.9355	0.8610	0.8447	0.9861	0.7634	0.9947	0.8606
PAM-VGG19	0.9036	0.8762	0.9787	0.9301	0.9246	0.8750	0.9032	0.9358	0.8889	0.9747	0.8280	0.9893	0.8953
ResNet18	0.8500	0.8302	0.9362	0.9032	0.8800	0.7788	0.8710	0.8770	0.8223	0.9857	0.7419	0.9947	0.8466
SE-ResNet18	0.8643	0.7876	0.9468	0.871	0.8599	0.8471	0.7742	0.9305	0.8090	0.9878	0.8710	0.9947	0.9257
PAM-ResNet18	0.8964	0.9263	0.9362	0.9624	0.9312	0.8073	0.9462	0.8877	0.8713	0.9868	0.8065	0.9947	0.9070
ResNet34	0.8643	0.8224	0.9362	0.8978	0.8756	0.8163	0.8602	0.9037	0.8377	0.9867	0.7957	0.9947	0.8876
SE-ResNet34	0.8982	0.9175	0.9149	0.9570	0.9158	0.8187	**0.9463**	0.8957	0.8779	0.9880	0.8334	0.9947	0.9034
PAM-ResNet34	0.9197	0.9312	0.9575	0.9624	0.9429	0.8648	0.9033	0.9278	0.8815	0.9766	0.8979	0.9893	0.9354
DenseNet	0.8929	0.8476	0.9468	0.9140	0.8945	0.8696	0.8602	0.9358	0.8649	0.9759	0.8710	0.9893	0.9205
SE-DenseNet	0.9286	**0.9375**	0.9574	**0.9677**	0.9474	0.8788	0.9355	0.9358	0.9062	0.9765	0.8925	0.9893	0.9326
PAM-DenseNet	**0.9429**	0.9118	**0.9894**	0.9516	**0.9490**	**0.9341**	0.9140	**0.9679**	**0.9239**	**0.9885**	**0.9247**	**0.9947**	**0.9556**

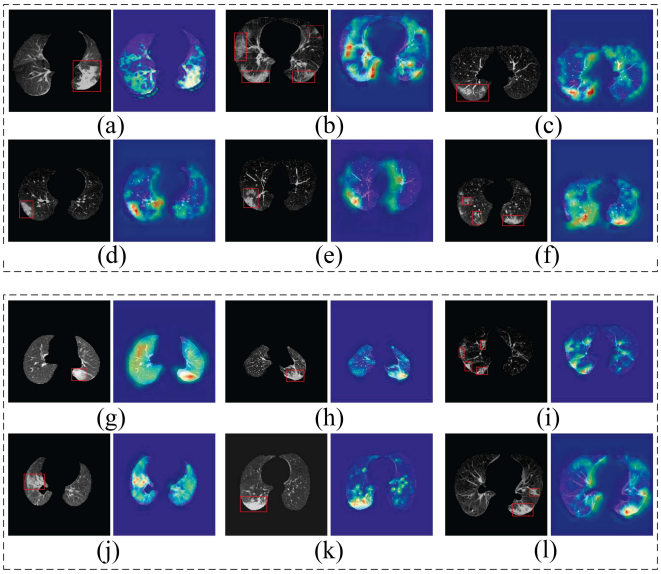


Fig. 12. Visualization of focused lung regions by the proposed PAM-DenseNet. For each visualization result, the one with the black background is an input CT image, followed by a class activation map shown as a heat map. The warmer color represents a higher value of weights, which denotes that the network is more focused on the regions. The regions marked by the red rectangle are the infection regions, such as the peripheral ground-glass opacities, lung consolidation, etc. (a - f): Class activation maps of six COVID-19 samples. (g - l): Class activation maps of six other pneumonia samples.

TABLE V
COMPARISON OF EXISTING COMPUTER-AIDED COVID-19 DIAGNOSIS METHODS BASED ON CNN

Literature	Subjects	Method	Result
Chen *et al.* [30]	51 COVID-19 55 Other pneumonia	U-Net++ Manually delineating	95.2% (Acc) 100% (Sen) 93.6% (Spec)
Zheng *et al.* [41]	313 COVID-19 229 Other pneumonia	U-Net CNN	90.7% (Sen) 91.1% (Spec)
Jin *et al.* [43]	496 COVID-19 1385 Other pneumonia	CNN	94.1% (Sen) 95.5% (Spec)
Jin *et al.* [44]	723 COVID-19 413 Other pneumonia	U-Net++ ResNet-50	97.4% (Sen) 92.2% (Spec)
Wang *et al.* [31]	44 COVID-19 55 Other pneumonia	CNN	82.9% (Acc)
Song *et al.* [28]	88 COVID-19 100 Other pneumonia 86 Healthy	ResNet-50	86.0% (Acc)
Xu *et al.* [29]	219 COVID-19 224 Other pneumonia 175 Healthy	CNN	86.7% (Acc)
Li *et al.* [45]	468 COVID-19 1551 Other pneumonia 1445 Non-pneumonia	ResNet-50	90.0% (Sen) 96.0% (Spec)
PAM-DenseNet	476 COVID-19 465 Other pneu. 293 Normal	Dense connectivity Attention module	94.29%(Acc) 91.18%(Prec) 98.94%(Sen) 95.16%(Spec)

three categories. In the calculation of evaluation metrics for each category, most of the indicators, especially the integrated index F1 score, also reach the highest value, which demonstrates the promising performance of our proposed network in patientwise prediction. In addition, we can see that in the patientwise prediction, the dense connection structure is superior to the residual connection structure and the plain structure. Moreover, the attention module has a stable improvement effect on performance. The PAM with spatial attention and channel attention is superior to SE block with only channel attention.

4) Comparison With Existing Methods: There are some existing methods in computer-aided COVID-19 diagnosis: the method of Chen *et al.* [30], which obtained 95.2% (accuracy), 100% (sensitivity), and 93.6% (specificity) on 51 subjects with COVID-19 and 55 subjects with other, the method of Zheng *et al.* [41], which achieved the sensitivity of 90.7% and specificity of 91.1% on 313 subjects with COVID-19 and 229 subjects with others, the method of Jin *et al.* [43], which achieved the sensitivity of 94.1% and specificity of 95.5% on 496 subjects with COVID-19 and 1385 subjects with others, the method of Jin *et al.* [44], which obtained the

sensitivity is 97.4% and specificity is 92.2% on 723 subjects with COVID-19 and 413 subjects with others, the method of Wang *et al.* [31], which achieved a total accuracy of 73.1%, along with a specificity of 67.0% and a sensitivity of 74.0% on 44 subjects with COVID-19 and 55 subjects with others, the method of Song *et al.* [28], which achieved an accuracy of 86.0% for pneumonia classification (COVID-19 or bacterial pneumonia), and an accuracy of 94.0% for pneumonia diagnosis (COVID-19 or healthy) on 88 subjects with COVID-19, 100 subjects with other pneumonia, and 86 healthy subjects, the method of Xu *et al.* [29], which achieved an overall accuracy of 86.7% on 219 subjects with COVID-19, 224 subjects with other pneumonia, and 175 healthy subjects, and the method of Li *et al.* [45], which obtained the sensitivity of 90% and a specificity of 96% on 468 subjects with COVID-19, 1551 subjects with other pneumonia, and 175 healthy subjects. The classification results of PAM-DenseNet and some state-of-the-art methods are listed in Table V. We can see that compared with existing methods, our method achieves promising results without manually delineated infection regions under a small number of parameters, which shows the application value in computer-aided COVID-19 diagnosis.

V. Conclusion

In this article, we proposed a CNN architecture PAM-DenseNet for computer-aided COVID-19 diagnosis on lung CT scans. The proposed PAM-DenseNet employs the dense connectivity structure and PAM attention module to perform feature maps reuse and self-recalibrating features, which can improve performance significantly on limited lung CT scans without manually delineated infection regions. Our dataset consists of three categories: 1) COVID-19; 2) other types of pneumonia; and 3) normal case. We train our PAM-DenseNet on selected lung CT slices to obtain a slicewise prediction model, then we use the pretrained model to predict each slice in patient CT scans dataset, and use a voting mechanism to calculate the final patientwise prediction results. The experimental results on 280 CT scans demonstrate that our proposed network achieves promising patient-wise results with an accuracy of 94.29%, a precision of 91.18%, a sensitivity of 98.94%, and a specificity of 95.16%, which is superior to most existing computer-aid COVID-19 diagnosis methods and is also comparable to the methods that are based on manually delineated infection regions.

References

[1] W.-J. Guan *et al.*, "Clinical characteristics of coronavirus disease 2019 in China," *New England J. Med.*, vol. 382, pp. 1708–1720, Apr. 2020.

[2] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, 2020.

[3] *Coronavirus Disease 2019 (COVID-19): Situation Report, 200*, World Health Org., Geneva, Switzerland, 2020.

[4] A. Bernheim *et al.*, "Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection," *Radiology*, vol. 295, no. 3, 2020, Art. no. 200463.

[5] H. Shi *et al.*, "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: A descriptive study," *Lancet Infect. Dis.*, vol. 20, no. 4, pp. 425–434, 2020.

[6] H. Y. F. Wong *et al.*, "Frequency and distribution of chest radiographic findings in patients positive for COVID-19," *Radiology*, vol. 296, Mar. 2020, Art. no. 201160.

[7] Z. Xu *et al.*, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respiratory Med.*, vol. 8, no. 4, pp. 420–422, 2020.

[8] J. Lim *et al.*, "Case of the index patient who caused tertiary transmission of COVID-19 infection in Korea: The application of lopinavir/ritonavir for the treatment of COVID-19 infected pneumonia monitored by quantitative RT-PCR," *J. Korean Med. Sci.*, vol. 35, no. 6, p. e79, 2020.

[9] Z. Y. Zu *et al.*, "Coronavirus disease 2019 (COVID-19): A perspective from China," *Radiology*, vol. 296, no. 2, 2020, Art. no. 200490.

[10] T. Ai *et al.*, "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, vol. 296, Aug. 2020, Art. no. 200642.

[11] Y. Fang *et al.*, "Sensitivity of chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, vol. 296, Aug. 2020, Art. no. 200432.

[12] J. P. Kanne, B. P. Little, J. H. Chung, B. M. Elicker, and L. H. Ketai, "Essentials for radiologists on COVID-19: An update—Radiology scientific expert panel," *Radiology*, vol. 296, no. 2, 2020, Art. no. 200527.

[13] Y.-H. Liao, Z.-C. Wang, F.-G. Zhang, M. F. Abbod, C.-H. Shih, and J.-S. Shieh, "Machine learning methods applied to predict ventilator-associated pneumonia with pseudomonas aeruginosa infection via sensor array of electronic nose in intensive care unit," *Sensors*, vol. 19, no. 8, p. 1866, 2019.

[14] N. Singh, R. Sharma, and A. Kukker, "Wavelet transform based pneumonia classification of chest X- ray images," in *Proc. Int. Conf. Comput. Power Commun. Technol. (GUCON)*, New Delhi, India, 2019, pp. 540–545.

[15] K. M. Kuo, P. C. Talley, C. H. Huang, and L. C. Cheng, "Predicting hospital-acquired pneumonia among schizophrenic patients: A machine learning approach," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, p. 42, 2019.

[16] R. T. Sousa *et al.*, "Evaluation of classifiers to a childhood pneumonia computer-aided diagnosis system," in *Proc. IEEE 27th Int. Symp. Comput.-Based Med. Syst.*, New York, NY, USA, 2014, pp. 477–478.

[17] W. W. Chapman, M. Fizman, B. E. Chapman, and P. J. Haug, "A comparison of classification algorithms to automatically identify chest X-ray reports that support pneumonia," *J. Biomed. Informat.*, vol. 34, no. 1, pp. 4–14, 2001.

[18] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights Imag.*, vol. 9, no. 4, pp. 611–629, 2018.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[20] Y. Ding *et al.*, "DeepEDN: A deep learning-based image encryption and decryption network for Internet of medical things," *IEEE Internet Things J.*, early access, Jul. 28, 2020, doi: 10.1109/JIOT.2020.3012452.

[21] B. Fan, H. Liu, H. Zeng, J. Zhang, X. Liu, and J. Han, "Deep unsupervised binary descriptor learning through locality consistency and self distinctiveness," *IEEE Trans. Multimedia*, early access, Aug. 17, 2020, doi: 10.1109/TMM.2020.3016122.

[22] W. Li *et al.*, "Progressive multistage learning for discriminative tracking," *IEEE Trans. Cybern.*, early access, Aug. 5, 2020, doi: 10.1109/TCYB.2020.2985398.

[23] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: A review," *J. Med. Syst.*, vol. 42, no. 11, p. 226, 2018.

[24] J.-G. Lee *et al.*, "Deep learning in medical imaging: General overview," *Korean J. Radiol.*, vol. 18, no. 4, pp. 570–584, 2017.

[25] D. Nie, L. Wang, E. Adeli, C. Lao, W. Lin, and D. Shen, "3-D fully convolutional networks for multimodal isointense infant brain image segmentation," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1123–1136, Mar. 2019.

[26] X. Wang *et al.*, "Weakly supervised deep learning for whole slide lung cancer image analysis," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3950–3962, Sep. 2020.

[27] F. Shan *et al.*, "Lung infection quantification of COVID-19 in CT images with deep learning," 2020. [Online]. Available: arXiv:2003.04655.

[28] Y. Song *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *medRxiv*, 2020, doi: 10.1101/2020.02.23.20026930.

[29] X. Xu *et al.*, "Deep learning system to screen coronavirus disease 2019 pneumonia," 2020. [Online]. Available: arXiv:2002.09334.

[30] J. Chen *et al.*, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study," *medRxiv*, 2020, doi: 10.1101/2020.02.25.20021568.

[31] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," *medRxiv*, 2020, doi: 10.1101/2020.02.14.20023028.
[32] N. Zheng *et al.*, "Predicting COVID-19 in China using hybrid AI model," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 2891–2904, Jul. 2020.
[33] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 7132–7141.
[34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 4700–4708.
[35] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.
[36] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," 2020. [Online]. Available: arXiv:2004.02731.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 770–778.
[38] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 2117–2125.
[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: arXiv:1409.1556.
[40] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.
[41] C. Zheng *et al.*, "Deep learning-based detection for COVID-19 from chest CT using weak label," *medRxiv*, 2020, doi: 10.1101/2020.03.12.20027185.
[42] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1–9.
[43] C. Jin *et al.*, "Development and evaluation of an AI system for COVID-19 diagnosis," *medRxiv*, 2020, doi: 10.1101/2020.03.20.20039834.
[44] S. Jin *et al.*, "AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks," *medRxiv*, 2020, doi: 10.1101/2020.03.19.20039354.
[45] L. Li *et al.*, "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy," *Radiology*, vol. 296, Mar. 2020, Art. no. 200905.
[46] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," 2018. [Online]. Available: arXiv:1807.06514.
[47] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[48] L. Chen *et al.*, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 5659–5667.
[49] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015. [Online]. Available: arXiv:1511.07122.
[50] J. Hofmanninger, F. Prayer, J. Pan, S. Rohrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," 2020. [Online]. Available: arXiv:2001.11767.
[51] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, 2017, pp. 618–626.

Bin Xiao received the B.S. and M.S. degrees in electrical engineering from Shaanxi Normal University, Xi'an, China, in 2004 and 2007, and the Ph.D. degree in computer science from Xidian University, Xi'an, in 2012.

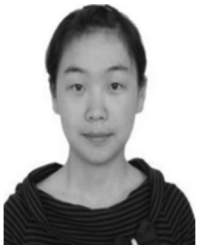
He is currently a Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include image processing and pattern recognition.

Zeyu Yang was born in 1994. He is currently pursuing the postgraduate degree with the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China.

His research interest is deep learning in medical image analysis.

Xiaoming Qiu received the B.S. and M.S. degrees in clinical medicine from Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, in 1998 and 2006, respectively.

He is currently a Chief Radiologist with the Department of Radiology, Huangshi Central Hospital (Affiliated Hospital of Hubei Polytechnic University) Edong Healthcare Group, Huangshi, China. His research interests include thoracic imaging diagnosis and abdominal imaging diagnosis.



Jingjing Xiao received the bachelor's and master's degrees from the College of Mechatronics Engineering and Automation, National University of Defence Technology, Changsha, China, in 2010 and 2012, respectively, and the Ph.D. degree from the University of Birmingham, Birmingham, U.K., in 2016.

She is currently an Honorary Research Fellow with the University of Birmingham, and a Full-Time Researcher with the Xinqiao Hospital, Chongqing, China. Her current research interest includes object tracking with computer vision and medical image processing.

Guoyin Wang (Senior Member, IEEE) received the bachelor's degree in computer software, the master's degree in computer software, and the Ph.D. degree in computer organization and architecture from Xi'an Jiaotong University, Xi'an, China, in 1992, 1994, and 1996, respectively.

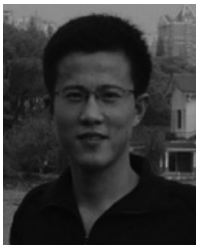
He has authored over 200 reviewed research publications. His books and papers have been cited over 6000 times. His research interests include data mining, machine learning, and pattern recognition.

Wenbing Zeng received the bachelor's and master's degrees from the North Sichuan Medical College of Sichuan, Chongqing University of Medical Sciences, Chongqing, China, in 1994 and 2007, respectively.

He is currently the Chief Radiologist with the Chongqing University affiliated Three Gorge Hospital, Chongqing. His current research interest is neuroimaging.

Weisheng Li received the B.S. degree from the School of Electronics and Mechanical Engineering, Xidian University, Xian, China, in 1997, and the M.S. and Ph.D. degrees from the School of Electronics and Mechanical Engineering and School of Computer Science and Technology, Xidian University in 2000 and 2004, respectively.

He is currently a Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research focuses on intelligent information processing and pattern recognition.



Yongjian Nian received the Ph.D. degree in information and communication engineering from the National University of Defense and Technology, Changsha, China, in 2010.

He is an Associate Professor with the School of Biomedical Engineering and Imaging Medicine, Army Medical University, Chongqing, China. His research fields include hyperspectral images processing, pattern recognition, and machine learning.

Wei Chen received the Ph.D. degree from Third Military Medical University, Chongqing, China, in 2001.

He is currently a Professor with Southwest Hospital, Army Medical University, Chongqing. His research interests includes medical image processing and radiomics.