A Cascade-SEME network for COVID-19 detection in chest x-ray images

Dailin Lv
The Key Lab of RF Circuits and Systems of Ministry of Education, Microelectronics CAD Center, Hangzhou Dianzi University, Hangzhou 310018, China

Yaqi Wang[a]
College of Media Engineering, Communication University of Zhejiang, Hangzhou 310018, China

Shuai Wang
Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, National Institutes of Health Clinical Center, Bethesda, MD 20892, USA

Qianni Zhang
School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK

Wuteng Qi, Yunxiang Li and Lingling Sun[a]
The Key Lab of RF Circuits and Systems of Ministry of Education, Microelectronics CAD Center, Hangzhou Dianzi University, Hangzhou 310018, China

Purpose: The worldwide spread of the SARS-CoV-2 virus poses unprecedented challenges to medical resources and infection prevention and control measures around the world. In this case, a rapid and effective detection method for COVID-19 can not only relieve the pressure of the medical system but find and isolate patients in time, to a certain extent, slow down the development of the epidemic. In this paper, we propose a method that can quickly and accurately diagnose whether pneumonia is viral pneumonia, and classify viral pneumonia in a fine-grained way to diagnose COVID-19.
Methods: We proposed a Cascade Squeeze-Excitation and Moment Exchange (Cascade-SEME) framework that can effectively detect COVID-19 cases by evaluating the chest x-ray images, where SE is the structure we designed in the network which has attention mechanism, and ME is a method for image enhancement from feature dimension. The framework integrates a model for a coarse level detection of virus cases among other forms of lung infection, and a model for fine-grained categorisation of pneumonia types identifying COVID-19 cases. In addition, a Regional Learning approach is proposed to mitigate the impact of non-lesion features on network training. The network output is also visualised, highlighting the likely areas of lesion, to assist experts' assessment and diagnosis of COVID-19.
Results: Three datasets were used: a set of Chest x-ray Images for Classification with bacterial pneumonia, viral pneumonia and normal chest x-rays, a COVID chest x-ray dataset with COVID-19, and a Lung Segmentation dataset containing 1000 chest x-rays with masks in the lung region. We evaluated all the models on the test set. The results shows the proposed SEME structure significantly improves the performance of the models: in the task of pneumonia infection type diagnosis, the sensitivity, specificity, accuracy and F1 score of ResNet50 with SEME structure are significantly improved in each category, and the accuracy and AUC of the whole test set are also enhanced; in the detection task of COVID-19, the evaluation results shows that when SEME structure was added to the task, the sensitivities, specificities, accuracy and F1 scores of ResNet50 and DenseNet169 are improved. Although the sensitivities and specificities are not significantly promoted, SEME well balanced these two significant indicators. Regional learning also plays an important role. Experiments show that Regional Learning can effectively correct the impact of non-lesion features on the network, which can be seen in the Grad-CAM method.
Conclusions: Experiments show that after the application of SEME structure in the network, the performance of SEME-ResNet50 and SEME-DenseNet169 in both two datasets show a clear enhancement. And the proposed regional learning method effectively directs the network's attention to focus on relevant pathological regions in the lung radiograph, ensuring the performance of the proposed framework even when a small training set is used. The visual interpretation step using Grad-CAM finds that the region of attention on radiographs of different types of pneumonia are located in different regions of the lungs. © *2021 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14711]

Key words: Cascade-SEMEnet, chest x-rays, COVID-19, grad-CAM, regional learning

1. INTRODUCTION

Recently, a novel virus SARS-CoV-2, also referred to as COVID-19, swiftly spread over the world, leading to a global pandemic, posing a deadly threat to people. As indicated by latest research,[1] SARS-CoV-2 is a more infectious virus than SARS and MERS. As of May 15, 2020, more than 4.52 million people have been diagnosed with pneumonia in the world, and more than 303,000 people died of it. The resulting death rate is as high as 6.7%. According to recent studies,[2] SARS-CoV-2 has not only a devastating impact on human lung tissue, it also attacks other vital organs such as the heart, blood vessels, kidneys, gastrointestinal tract, eyes, and brain with significant consequences. The surviving patients with severe COVID-19 may also be at risk of disability.[3]

To address the urgent demand for automated solutions to detect COVID-19 cases in x-ray images, in this paper a dedicated network structure called Squeeze-Excitation and Moment Exchange (SEME) is designed. Following this structure, a Cascade-SEMEnet is developed for COVID-19 diagnosis using chest radiographs. The patients with suspicious symptoms of pneumonia will receive an x-ray scan. Among these patients, some are fortunately normal, while some will present pneumonia caused by either bacteria or virus. COVID-19 is a type of viral pneumonia and the pathological features are similar to those of other viral pneumonia. Following these observations, Cascade-SEMEnet takes a two-level paradigm, by first conducting a coarse classification of viral pneumonia images from bacterial pneumonia and normal images, and then performing a fine-grained classification of COVID-19 images against other types of viral pneumonia images. Besides the two level structure, we propose a Regional Learning approach to mitigate the impact of non-lesion features on network training.

To enable the medical experts interpret the prediction results of the Cascade-SEMEnet and therefore assist their visual assessment, the Grad-CAM method is employed to calculate the weighted sum of the feature map in each convolution layer of the input image and obtain a thermal map to visualise the detection results with a confidence level. The main contribution of this paper lies in three aspects:

1. A novel SEME structure is designed dedicated to the task of COVID-19 detection in x-ray images, based on which the Cascade-SEMEnet is developed implementing a two-level detection process: a SEME-ResNet50 model for viral pneumonia detection and a SEME-DenseNet169 model for COVID-19 classification.
2. A regional learning method is proposed, which can effectively avoid learning unnecessary misleading information during neural network training and enhance the classification accuracy of network.
3. To facilitate the experts' interpretation of the output from Cascade-SEMEnet, the Grad-CAM method[4] is employed to visualise a thermal map of COVID-19 detection confidence on top of the original image.

2. RELATED WORK

In the field of chest x-ray analysis, machine learning has been playing an important role. The work in Chandra et al[5] presents a method to detect and locate disease in lung x-ray images based on pre-clustered small image patches using Fuzzy C-Means and K-Means. Deep neural networks are also employed in related tasks such as detecting areas with abnormal density in chest radiographs using a fusion of multiple CNNs.[6] These studies show that computer vision and machine learning methodologies can achieve promising results when they are appropriately applied in these tasks. Sharma et al[7] propose a deep convolutional neural network (CNN) architecture, which extracts features from chest radiographs and classifies the images of pneumonia cases. According to previous studies, deep learning performs remarkably in tasks related to chest imaging, such as detecting pneumothorax.[8]

With the recent outbreak of COVID-19, a large number of chest radiographs of suspect cases are being generated on a daily basis and they all require urgent analysis by medical experts.[9–12] This process costs a massive amount of resources and can be error-prone. The lungs of most COVID-19 patients show consolidation and Ground-glass opacification. Different from many other types of viral pneumonia, the chest CT of COVID-19 patients will show multiple tiny pulmonary nodules.[10] The lab led by Zhong Nanshan at the Chinese Academy of Engineering investigate the data of 1099 lab-confirmed COVID-19 patients from 552 hospitals in mainland China as of January 29, 2020, and find that approximately 86% of these patients present relevant abnormal features in their chest imaging.[13] These studies demonstrate the unique morphological characteristics of COVID-19 in chest radiographs and lay the foundation for the employment of deep learning techniques to automate the examination of chest radiographs and detect COVID-19.

Extensive machine learning based research has emerged in the last a few months since the virus outbreak to help coping with the global challenge in different ways. For instance, the work in Bontempi et al[14] proposes a weakly-supervised framework to detect COVID-19 in chest CT while locating possible lesion areas, the methods in S. Wang et al[15] and J. Wang et al[16] are similar.

However, the chest CT data acquisition cost in these methods is high, the number of images is enormous, and the analysis of them is more complicated compared with the chest radiograph. Cohen et al[17] and Zhu et al[18] used chest radiographs of patients with COVID-19 to assess the severity of infection in patients with COVID-19, which was useful and effectively reduced the burden of doctors but did not play a role in the diagnosis of COVID-19. In the works of Minaee et al[19] and Azemin et al,[20] COVID-19 patients' chest radiographs were diagnosed using deep learning networks, and good test results had been obtained. These works demonstrated the potential of such methods for improved model design and the resulting performance. Das et al[21] improved the structure of the model and improved the test results

greatly, but in their experiments, the network was only used to distinguish COVID-19 from healthy chest radiographs, and some patients with non-COVID-19 pneumonia were likely to be misdiagnosed. Yoo et al[22] proposed three binary decision trees to more accurately identify the type of pneumonia. However, the construction and evaluation details of these models have not been enumerated in detail, and the decision-making basis for the classification of COVID-19 chest radiographs has not been examined. Thus in this research, we focus on addressing the challenge of COVID-19 detection and localisation in radiograph images using deep learning and computer vision techniques.

3. MATERIALS

3.A. Dataset

All data in this article are made publicly available online. The data consists of three parts:

1. **Dataset 1** Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification[23] contains 5859 chest radiographs taken from patients with normal lungs, bacterial pneumonia, and viral pneumonia, accounting for 27%, 47% and 26% respectively. It is initially divided into a training set of (1341, 2538, 1345) and a test set of (250, 242, 148). In the process of training for Dataset1, we sub-divided the initial test set into a valuation subset and a test subset according to 1:1 ratio [(125, 121, 74): (125, 121, 74)].
2. **Dataset 2** COVID-19 image data collection[24] collects and arranges the chest radiographs of patients with pneumonia, such as MERS, SARS, COVID-19, etc. from various publications. The train set and test set are separated by nearly 1: 1 ratio, while each class retain the same proportion in each subset.

3. **Unet Training Data** The Dataset of Lung Segmention[25] contained 1000 chest x-rays with masks in the lung region. In the experiment, we use U-Net to segment the lung area in the chest radiographs, and the training data for U-Net segmentation are the masks of the lung area and the chest radiographs.

Figure 1 shows some examples from these datasets, and Table I gives some statistics and basic facts about the dataset.

3.B. Contrast limited adaptive histogram equalization

Due to the technical limitations, almost all chest radiographs contain noise and unclear contour areas. Contrast-limited histogram equalization (CLAHE)[26] can effectively mitigate the effects of these noises on neural networks. The use of this technique in chest radiographs is very effective in enhancing the chest radiographs. It solves the problems of normal histogram equalization including leading to loss of target details, excessive background enhancement, and noise amplification. After using CLAHE, the boundaries between bone and bone or bone and organ tissues become clearer, and some detailed textures can be clearly captured.

4. METHODS

We design a two-level cascade framework, which first coarsely classifies chest radiographs into three classes: normal, bacteria pneumonia and viral pneumonia, and then performs a fine-grained screening to identify COVID-19 cases out of all viral pneumonia ones. An illustration of the designed CascadeSEMEnet framework architecture is shown in Fig. 2.

As depicted in Fig. 2, each SEME network includes three components: a Moment Exchange (MoEx)[27] component [(a) in Fig. 2], which is an algorithm that normalizes and
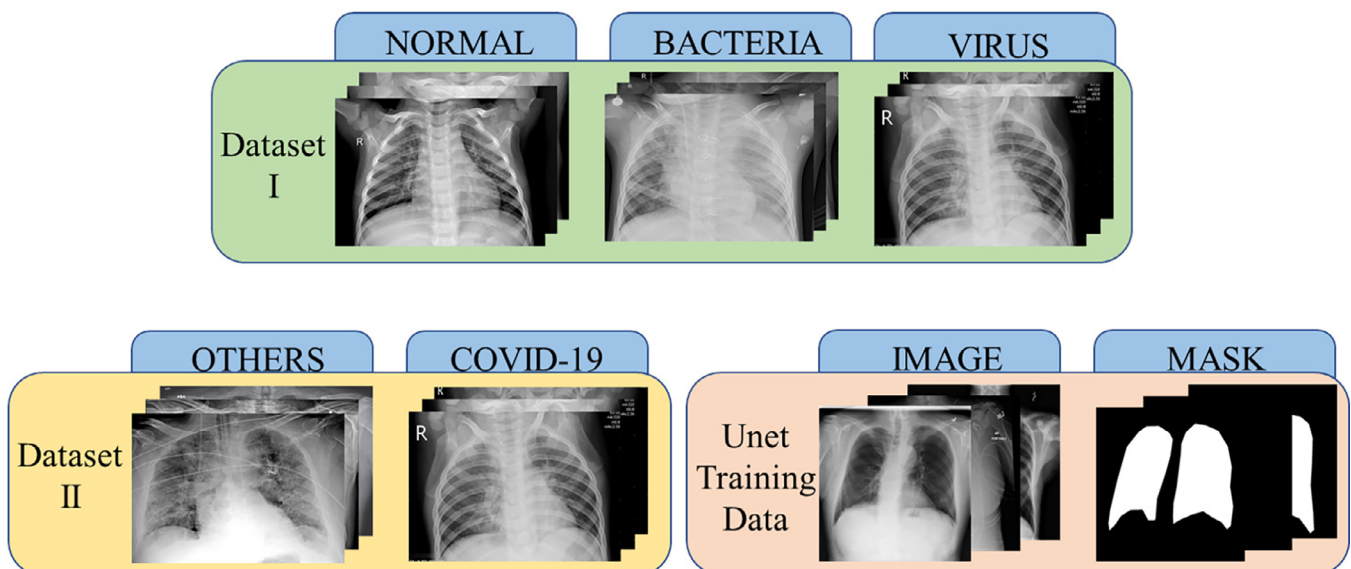

FIG. 1. Display of dataset1 and dataset2. [Color figure can be viewed at wileyonlinelibrary.com]

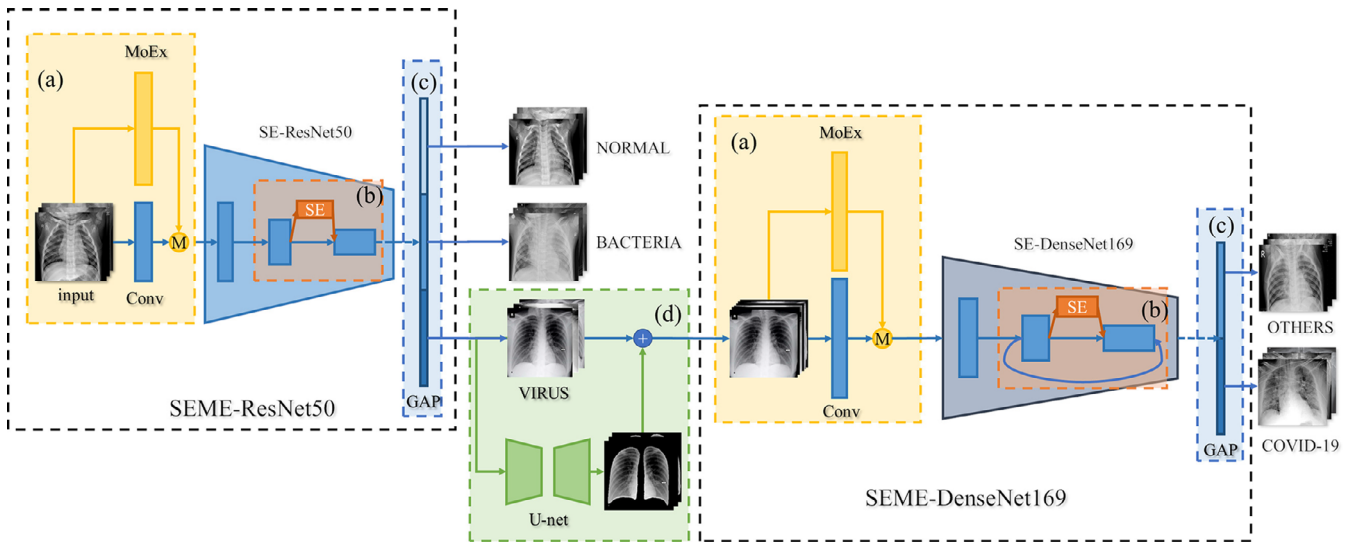|	Dataset I			Dataset II		U-Net Training Data	
	Normal	Bacteria	Virus	COVID-19	Others virus	Image	Mask
Training	1341	2529	1345	105	165	900	900
Validation	125	121	74	10	76	50	50
Test	125	121	74	321	530	50	50
Overall	1591(27%)	2771(47%)	1493(26%)	436(37%)	755(63%)	1000	1000

FIG. 2. A new cascade network structure with the most potential. (a) is the MoEx structures, (b) is the SE structures, (c) is the Global Average Pooling (GAP) layers, and (d) is the process of Regional Learning. [Color figure can be viewed at wileyonlinelibrary.com]

enhances image features in the process of network reasoning by increasing the generalization ability of data distribution and model; a Squeeze-Excitation (SE) component (b), which exists between the fusion layers of the model and is used to assign values to each characteristic channel during network training, directing the attention of the network to the channel; and a Global Average Pooling (GAP) component (c), which exists in the last layer of the model convolution layer. The GAP structure guarantees that the size of the network receptive field changes with the input size of the network, and the ability of the network to utilize picture information also increases. At the same time, the addition of GAP can effectively reduce the number of full connection layer parameters and improve the performance of the network.

In the cascade process, we average all the output results of the first network according to the category to determine a threshold for judging whether the classification is accurate. Set the n-th picture as the input to the first layer network, and $[Bac_n, Nor_n, Vir_n]$ represents the probability of the bacterial, normal, and viral classes of its output, respectively. The threshold of the final viral classes can be expressed as:

$$Th_{Vir} = average\left(\sum Vir_n\right) \quad (1)$$

Next, before each picture is entered into the subordinate network, its Vir_n and Th_{Vir} are compared, and only images

associated to values greater than this threshold are sent to the second network. In the next subsection, I'll go into details about network implementation.

4.A. Feature normalization and data enhancement

Data promotion is a good way to improve network performance and increase the generalization ability. However, at present, most of the mainstream data enhancement approach is only carried out at the original image level, which limits the feature extraction ability of the network to some extent. In order to break this limitation, we design a Moment Exchange (MoEx) structure in the network, which is an algorithm to enhance image features in the network reasoning process. Unlike traditional data promotion methods, MoEx mixes two different parts: the standardized features of one instance and the feature matrix of the other. In the feature space, this asymmetric combination enables the network to capture and smooth the different directions of the decision boundary, which is not covered by the previous data promotion methods. Suppose the network inputs two different kinds of samples x_A and x_B and the Normalization function can be denoted as F, the MoEx formula can be expressed as:

$$h_A^{(B)} = F^{-1}\left(h_A, \mu_B, \sigma_B\right) \quad (2)$$

where $\widehat{h}_A$ represents the normalized features of x_A, μ_B and σ_B are the mean and standard of x_B features, respectively, and $h_A^{(B)}$ contains the moments of image B hidden inside the features of image A. In order to encourage the neural network to pay attention to the injected features of B, MoEx uses a modified loss function to predict the class label y_A and y_B with a mixing constant $\lambda \in [0,1]$. The loss becomes a straight-forward combination:

$$\lambda * l(h_A^{(B)}, y_A) + (1 - \lambda) * l(h_A^{(B)}, y_B) \qquad (3)$$

Figure 3(a) illustrates the process of MoEx based feature representation. Figure 3(b) shows the role of MoEx in the network. It is verified by experiments that this network structure can significantly enhance the generalization performance of the model.

Considering that VGG is a kind of network that directly tries to learn the mapping between input and output by using the parameter layer, it has no merge layer. If SE is added to such a structure, the structure of the whole network model would be changed, and network may lose its stability. Therefore, we did not add the MoEx to VGG, including the SE structure following.

4.B. The Squeeze-Excitation structure

To examine a chest radiography, even the most experienced doctors need to analyze the image in a highly detailed level in order to reach a relatively accurate outcome. This means that the key features of pneumonia including COVID-19 are present in fine details in radiography. Thus, the neural networks need to exploit such fine details so as to achieve accurate results. Squeeze-Excitation (SE) is a structure that can change the network's attention,[28] as shown in Fig. 4(a). It consists of the two operations: Squeeze and Excitation, and can be applied to any mapping.

First, Squeeze encodes the entire channel using global average pooling, then adds two fully connected (FC) layers, activates the second FC layer by ReLU, and finally, Excitation activates the non-linear relationship between channels through Sigmoid.

Let Z be the feature vector after the GAP of the feature map, C be the number of channels of the vector, and F_{ex} the Excitation operation, which is performed with Sigmoid Gating mechanism g, leading the network to learn the nonlinear relationship between the channels. This process is defined as:

$$s = F_{ex}(\mathbf{Z}, \mathbf{W}) = \sigma(g(\mathbf{Z}, \mathbf{W})) = \sigma(\mathbf{W}_2 ReLU(\mathbf{W}_1 \mathbf{Z})),$$
$$\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}, \mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}} \qquad (4)$$

Here, $\mathbf{W}_1$ and $\mathbf{W}_2$ are the weights of the two fully connected layers of Excitation operation, and σ is the Sigmoid operation.

Finally, this will be multiplied by the activation value (sigmoid activation, value 0-1) of each channel, learned by the original features on U to change the network's attention to each channel. As presented in Figs. 4(b) and 4(c), the SE structure is employed in each fusion layer of SE-ResNet50 and SE-DenseNet169, to enhance the attention to pathological features and increase the credibility of detection.

4.C. Input size and receptive field

In previous studies, large images need to be down-scaled to 224×224 pixels to be applicable to network structures such as ResNet50 and DenseNet169. However, such resizing operations will modify some pathological features in the radiography images, and result in a significant loss of details, such as the example shown in Fig. 5. These changes will

(a) Feature extraction process   **(b)** Role in network
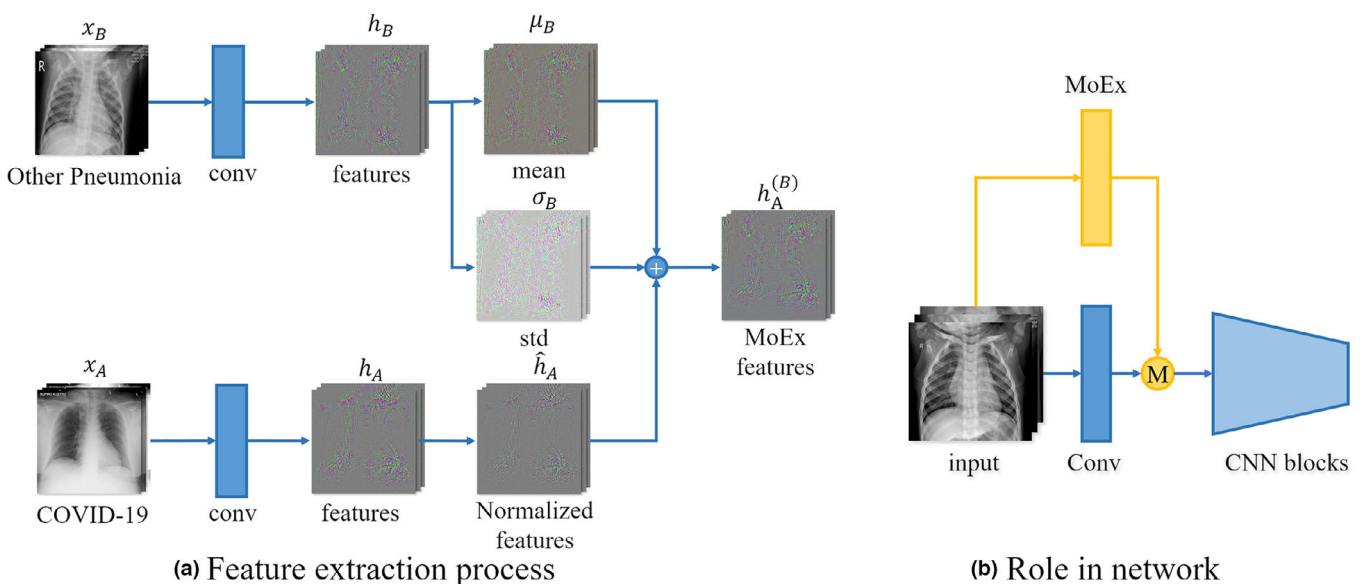
FIG. 3. An illustration of the MoEx component, its feature extraction process (a) and its role in the network (b). [Color figure can be viewed at wileyonlinelibrary.com]
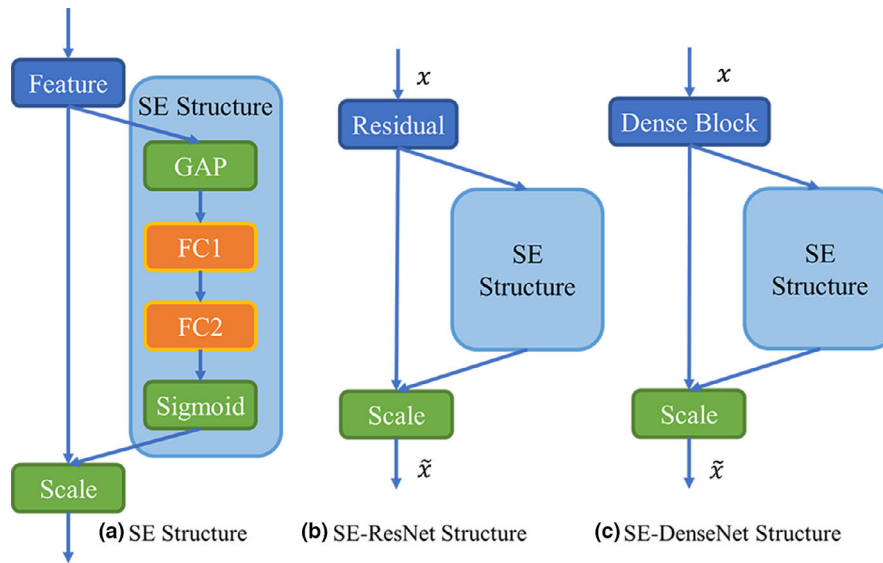
FIG. 4. Illustrations of (a) the Squeeze-and-Exception (SE) structure, (b) the SE-ResNet structure and (c) the SE-DenseNet structure. [Color figure can be viewed at wileyonlinelibrary.com]
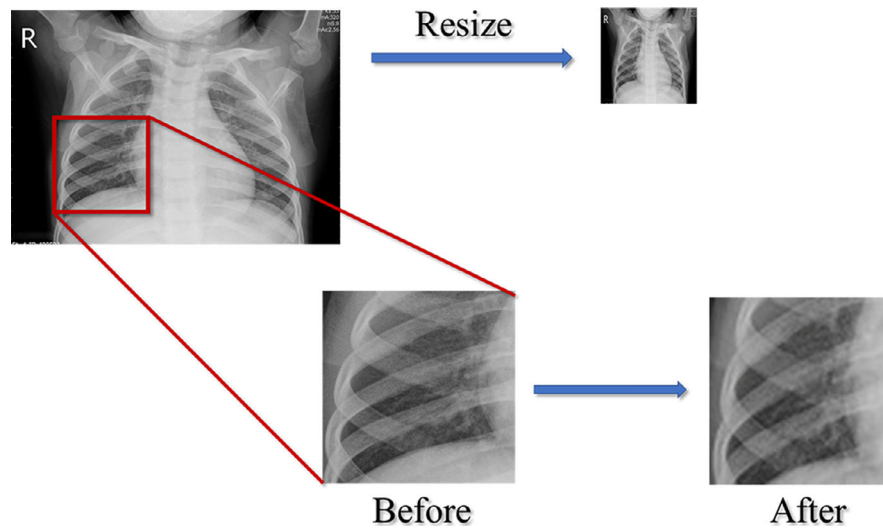

FIG. 5. Resize operation leads to loss of lesion details in the image. [Color figure can be viewed at wileyonlinelibrary.com]

adversely affect the learning and prediction abilities of the neural network.

To solve this problem, the Global Average Pooling (GAP) component is designed in the last layer of convolution. If f' is set to be the result of the feature in a channel f passing through GPA, i and j are the coordinates of feature pixels in channel k, the GAP formula is then expressed as:

$$f' = average(f_{i,j}) \qquad (5)$$

According to (5), after the features pass through this structure, the size of each channel will become 1×1, so that the input size of the classifier level will be greatly reduced and the number of parameters will remain stable. This process is illustrated in Fig. 6(a). After adding the GAP, the size of the newly formed receptive field will also increase with the size of the input image. The enlarging process of the receptive field is shown in Figs. 6(b) and 6(c). Therefore, after the convolutional layer, this adaptive method can meet the high-efficiency requirements for the neural network model.[29]

4.D. Regional learning

Regional Learning is a method suitable for small sample data. This method uses a pre-segmented image, referred to as a Target Region (TR) containing only the lesion area and the original image superimposed in a batch to guide the network to learn the correct information, as shown in Fig. 7. Denote a TR as X_T and the original figure as X_O. After the two images are superimposed in the network, the network output is represented as $(O_T, O_O) = F(X_T, (X_O))$. Denote the Cross Entropy as CE, then the overall loss L can be calculated by the combination of the losses of the TR and the original image:
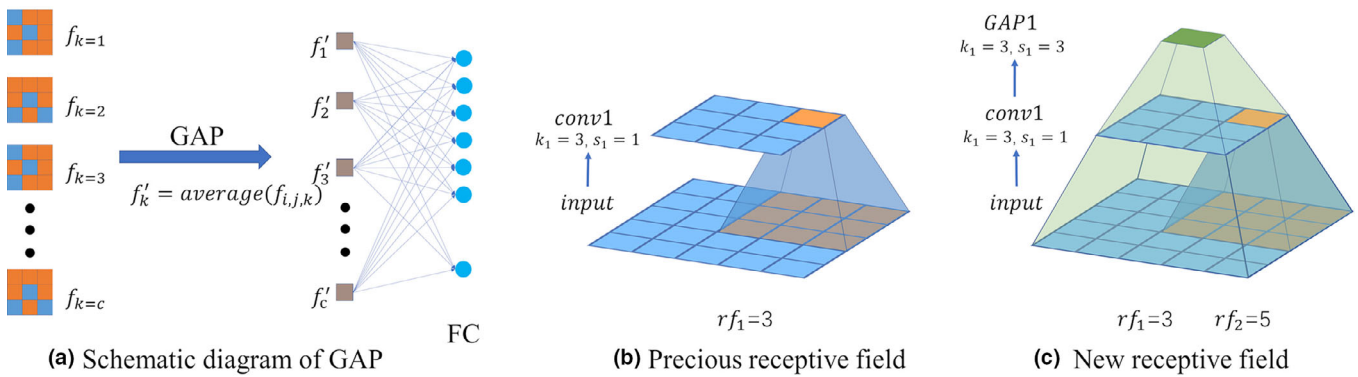
FIG. 6. (a) A schematic diagram of GAP, (b) the receptive field area before the GAP structure is increased, (c) the receptive field area after the GAP structure is increased. [Color figure can be viewed at wileyonlinelibrary.com]
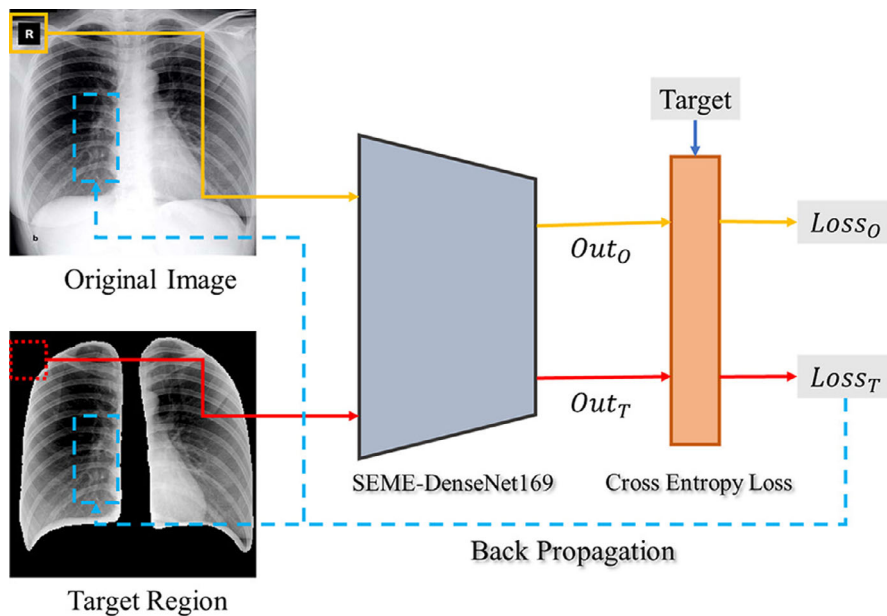

FIG. 7. When the network identifies a non-target area as the basis for decision, Regional Learning uses the corresponding Target Region to guide the network to focus on a true target area. [Color figure can be viewed at wileyonlinelibrary.com]

$$L = L_T + L_O = CE(O_T, Label) + CE(O_O, Label). \quad (6)$$

If the network notices some features in X_T during the training process, that is, the characteristics of the lesion area, then the output O_T and O_O will have little difference because the *Label* is the same. The inverse gradient of the parameters will update normally. If the decision in the network training process is based on features other than X_T, and the classification is wrong, then the loss will also be updated normally under the influence of L_T and L_O. If the decision is based on features other than X_T and the classification of the original image is correct, the network must not find the corresponding judgment basis in TR. The difference between O_T, O_O will increase significantly, which will result in L_T being significantly larger than L_O. At this time, when updating the parameters of the network in the reverse gradient, more attention will be paid to extracting the features of the common areas of the image.

In this paper, the lung region is segmented as TR by using a trained U-Net, which is superimposed on top of the original image for training. If the network uses non-lesion features other than the lung region during the training process, only retaining the image of the lung region will inevitably lead to a wrong decision and increase the loss of the network, thus guiding the network's attention to the lesion feature area of the lung. As demonstrated by the experiment results, this method works effectively in enhancing the overall detection performance.

5. EXPERIMENT

5.A. Implementation details

The building and training of Cascade-SEMEnet are divided into two parts. In the first part, we evaluate the baseline performance of VGG19,[30] ResNet50,[31] and

DenseNet169,[32] which have achieved promising results in several tasks during recent years, and compare the impact of structures: GAP, SE, MoEx on network performance. All the Experiment use SGD as an optimizer, which learning rate has been set to 1e-4 and the momentum to 0.9.

In the second part, due to the small amount of data, we perform histogram equalization on dataset2 and use MoEx to enhance image features, randomly rotate the images and limiting the rotation amplitude to $\pm 30°$ during the training. This operation increases the diversity of images and reduce the occurrence of overfitting. We train four different models: MoEx-ResNet50, MoEx-DenseNet, SEME-ResNet50, SEME-DenseNet169 in this part with Regional Learning, and compare the performance of these models. Adam optimizer and cosine annealing are used in the training process of U-Net. Set η_t as the learning rate in the current epoch, η_{min} and η_{max} as the minimum and maximum of the learning rate, cosine annealing formula can be expressed as:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min}^i)(1 + cos(\frac{T}{T_t}))$$

where T is the period of cosine annealing, and T_t is the time of current epoch. In our experiment, η_{max}^i is set to 0.1 and η_{min}^i is set to $1e-8$. Since chest radiographs are largely different from natural images as those in ImageNet, we did not use the pre-trained weights from ImageNet for initializing the network. Instead, the He-Uniform[33] is adopted to initialize the convolution layers of the network:

$$\mathbf{W} \sim \text{Uniform}(-\sqrt{\frac{6}{n_{in}}}, \sqrt{\frac{6}{n_{in}}}),$$

where $\{\mathbf{W}\}$ is the weight of every layer, and the n_{in} is the number of input units in the weight tensor.

5.B. Evaluation metrics

In the evaluation phase, we evaluated different models in detail. The evaluation indexes include sensitivity, specificity, precision, F1 score, accuracy and AUC. Usually, these indicators are used to measure the performance on both positive and negative categories. However, it should be pointed out that dataset1 is a three-class classification dataset. Therefore, for each category, we have calculated the indicators once, regarding this category as a positive sample and other categories as negative samples, and finally obtain three indicators. Then we average these indicators to get an overall evaluation of the network. For ROC curve, we encoding the target with the one-hot method, and every element in the processed targets can be seen into a two classification problem. In this way, the ROC curves of the three categories can be drawn by the method of a two classification problem.

To evaluate the response level of the model to the addition of improved structures, we used F-test to test the statistical significance of the model's output. Set F-test be calculated as $F = Ftest(N_1, N_2)$, where F is the test statistic, N_1 and N_2 are the two distribution to be evaluated.

For Dataset1, we collect the output distributions of all 10 models according to normal, bacteria and virus respectively: $(N_{1N}, ..., N_{10N}), (N_{1B}, ..., N_{10B}), (N_{1V}, ..., N_{10V})$, and for each category, we perform a significance test with F-test to assess the extent of the effect on the improved model by every two models' output response. That can be described as: $F_{i,j} = Ftest(N_i, N_j)$ where i, j is the indexes of these models and $F_{i,j}$ is the test statistic of N_i and N_j. For Dataset2, this method is also applicable.

We also use the Grad-CAM[4] to visualize the classification basis of the network. Grad-CAM is an improved version of CAM, by calculating the weights of feature graphs based on the gradient information flowing into the last convolutional layer of CNN. Thus, Grad-CAM is applicable to any convolutional neural network. In the diagnosis of chest radiography, Grad-CAM is used to visualise the classification basis of the network, interpreting the system decision to doctors for their better understanding.

6. RESULT

6.A. Classification of pneumonia infection type

We added improved structure in our experiment to different basic models and enhanced the data to determine the type of pneumonia infection. After training, we used the test set to evaluate these models. The specific training and evaluation processes are described in details in the following:

6.A.1. Baseline

In the first stage, chest radiographs from the three classes, *bacterial pneumonia*, *viral pneumonia* and *normal*, are used to train the neural network model. We first use un-enhanced data to train three classic network structures (VGG19, ResNet50 and DenseNet169) to determine their baseline performance. Figure 8 presents the curve comparisons of all training processes in this paper, and (a) is the result of this experiment.

Overall, the losses of the three kinds of curves converge well, showing that the networks can capture the distinctive characteristics from the chest radiograph to decide on the types of pneumonia infection. Comparing the training curves of the three networks, it can be found that the overfitting degree of VGG is higher than that of the other two networks.

6.A.2. Adding GAP and SE structure to the network

We add GAP to the last convolution layer of VGG19, ResNet50 and DenseNet169, adjust the input size of the networks to 512 * 512, and retrain them. The training curve is shown in Figs. 8(b)–8(d). With GAP added, the input images contain abundant details, the networks' sense fields are enlarged, and the convergence speed of ResNet50 and VGG19 are significantly improved. However, the convergence rate of DenseNet169 does not change significantly. In the evaluation of the test set, after the adding of GAP, some
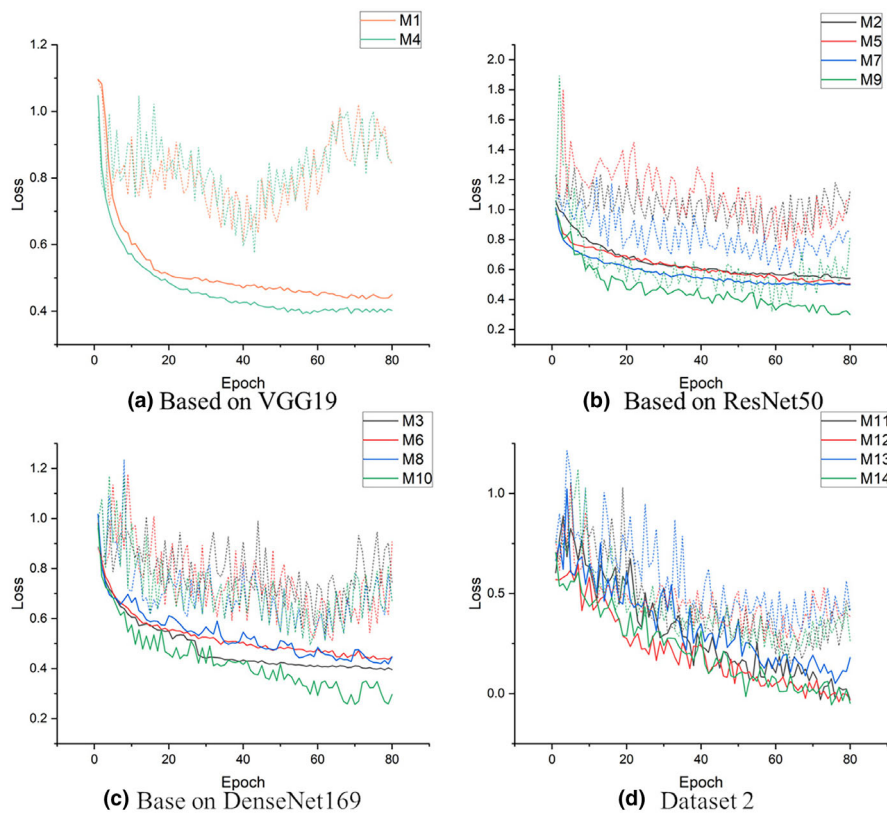
FIG. 8. (a), (b), (c) training curves on Dataset1. (d) training curves on Dataset2. M1: Base-VGG19, M2: Base-ResNet50, M3: Base-DenseNet169, M4: VGG19, M5: ResNet50, M6: DenseNet169, M7: SE-ResNet50, M8: SE-DenseNet169, M9: SEME-ResNet50, M10: SEME-DenseNet169, M11: MoEx-ResNet50, M12: MoEx-DenseNet169, M13: SEME-ResNet50, M14: SEME-DenseNet169. **line: Training curves, dots: Validation curves**. [Color figure can be viewed at wile yonlinelibrary.com]

statistical indicators in the single category evaluation will decrease as shown in Table II, for example, the sensitivity of ResNet50 in the normal category is reduced from 0.61 to 0.50, but the overall evaluation indicators are improved.

Then, the SE structure is applied in ResNet50 and Dense-Net169 networks, and the networks are trained with the same data. For ResNet50 with SE structure, the speed of convergence is significantly improved, as it can be seen in Fig. 8(c), but there is still no significant change in DenseNet169. Test set evaluation shows that after adding SE module to the model, the sensitivities of some single categories are decreased, as shown in Table II, while the specificity is significantly improved. For example, the sensitivity of resnet50 in viral pneumonia decreases from 0.97 to 0.93, but the specificity increases from 0.76 to 0.86. The overall evaluation indicators also improve significantly. Figure 9(a) shows the ROC curves of the VGG19 network before and after applying GAP and Fig. 10 M1, M2 present its confusion matrices. It can be seen that after adding GAP to the VGG19 network, the Area under the Curve (AuC) of ROC rises from 0.776 to 0.824, the confusion matrix shows a clearer concentration in the diagonal region, and the the average F1-score rises from 0.69 to 0.77 (Table II). Figures 9(b) and 9(c) and Fig. 10 M3-M10 are the ROC curves and confusion matrices of the basic ResNet50 and DenseNet169, with GAP, SE-Structure added, and with the data enhancement algorithm (MoEx) applied. It

can be seen that after adding GAP and the SE-Structure to the two networks, the AuC of ResNet50 and DenseNet169 increase from 0.781 to 0.850 and from 0.785 to 0.842, respectively.

6.A.3. Data Augmentation and enhancement

On the basis of previous experiments, CLAHE is randomly performed on 40% of the training set, at the same time, we add the MoEx structure to SE-ResNet50 and SE-DenseNet169 networks and use the algorithm of feature normalization to fuse the features of different kinds of lesions.

After the MoEx structure is introduced, the convergence speed of the two networks is further improved. As shown in Fig. 8(d). It can be seen from the evaluation of the test set in Table II, that the sensitivity and specificity of SEME-ResNet50 and SEME-DenseNet169 for the normal lung class are greatly improved. For bacterial pneumonia, the sensitivity is greatly improved, the specificity is reduced, while the F1 score still demonstrates improvement. For viral pneumonia, the sensitivity is slightly decreased, but the other indicators of SEME-ResNet50 show clear advantage over other networks. The AuC of SEME-ResNet50 and SEME-Dense-Net169 increase to 0.926 and 0.897, respectively [Figs. 9(b) and 9(c)], and the confusion matrices further converge towards the diagonal regions (Fig. 10).

TABLE II. The test indicators of the networks on Dataset 1.

Class	Model	Recall	Specificity	Precision	F1-score	accuracy	AUC
Overall	M1	0.72	0.86	0.74	0.69	0.6969	0.776
	M2	0.73	0.87	0.74	0.72	0.7281	0.781
	M3	0.80	0.90	0.80	0.77	0.7750	0.785
	M4	0.79	0.90	0.78	0.77	0.7781	0.824
	M5	0.77	0.88	0.78	0.73	0.7406	0.796
	M6	0.82	0.91	0.82	0.80	0.8094	0.851
	M7	0.93	0.91	0.81	0.81	0.8156	0.850
	M8	0.83	0.92	0.83	0.81	0.8187	0.842
	M9	**0.89**	**0.95**	**0.89**	**0.88**	**0.8906**	**0.921**
	M10	0.82	0.92	0.83	0.81	0.8281	0.890
Normal	M1	0.4876	0.9698	0.9077	0.6344		
	M2	0.6116	0.9447	0.8706	0.7184		
	M3	0.6364	0.9749	0.9390	0.7586		
	M4	0.6281	0.9548	0.8941	0.7379		
	M5	0.4959	0.9799	0.9375	0.6486		
	M6	0.6694	0.9749	0.9419	0.7826		
	M7	0.7521	0.9246	0.8585	0.8018		
	M8	0.6694	0.9849	0.9643	0.7902		
	M9	**0.8182**	0.9749	0.9519	**0.8800**		
	M10	0.7025	**0.9950**	**0.9884**	0.8213		
Bacteria	M1	0.8080	0.9026	0.8417	0.8245		
	M2	0.8080	0.8769	0.8080	0.8080		
	M3	0.8080	0.9487	0.9099	0.8559		
	M4	0.8720	0.9179	0.8720	0.8720		
	M5	0.8400	0.9026	0.8468	0.8434		
	M6	0.8720	0.9282	0.8862	0.8790		
	M7	0.8080	**0.9538**	**0.9182**	0.8596		
	M8	0.9040	0.9385	0.9040	0.9040		
	M9	**0.9600**	0.9231	0.8889	**0.9231**		
	M10	**0.9600**	0.9026	0.8633	0.9091		
Virus	M1	0.8514	0.7073	0.4667	0.6029		
	M2	0.7838	0.7886	0.5273	0.6304		
	M3	0.9459	0.7683	0.5512	0.6965		
	M4	0.8649	0.8130	0.5818	0.6957		
	M5	**0.9730**	0.7561	0.5455	0.6990		
	M6	0.9324	0.8293	0.6216	0.7459		
	M7	0.9324	0.8577	0.6635	0.7753		
	M8	0.9189	0.8252	0.6126	0.7351		
	M9	0.8919	**0.9390**	**0.8148**	**0.8516**		
	M10	0.8108	0.8577	0.6316	0.7101		

The bold numbers are used to represent the best values in each evaluation index.

*M1: Base-VGG19, M2: Base-ResNet50, M3: Base-DenseNet169, M4: VGG19, M5: ResNet50, M6: Dense-Net169, M7: SE-ResNet50, M8: SE-DenseNet169, M9: SEME-ResNet50, M10: SEME-DenseNet169.

6.A.4. Subjective model validation and diagnostic basis of pneumonia infection types

To further verify the effectiveness of the network, we use the Grad-CAM method to visualize the prediction results of the SEME-ResNet50 network, as shown in Fig. 11, (a) shows a few original images and (b) presents the Grad-CAM thermal diagram of the SEME-ResNet50 output. From the visualization results, we can see that the focus areas picked up by the networks are basically concentrated in the position of the lung in the chest radiographs. In the lung region, the basis of chest radiographs of bacterial pneumonia mainly locate in the middle and lower region. This might be because the network picks up key features of pleural effusion[34] and mediastinal lymph node enlargement[35] as the judgment basis. In the chest radiograph of viral pneumonia, the network uses the central region of the entire lung as a basis for judgment. The significance test of every categories between networks is shown in
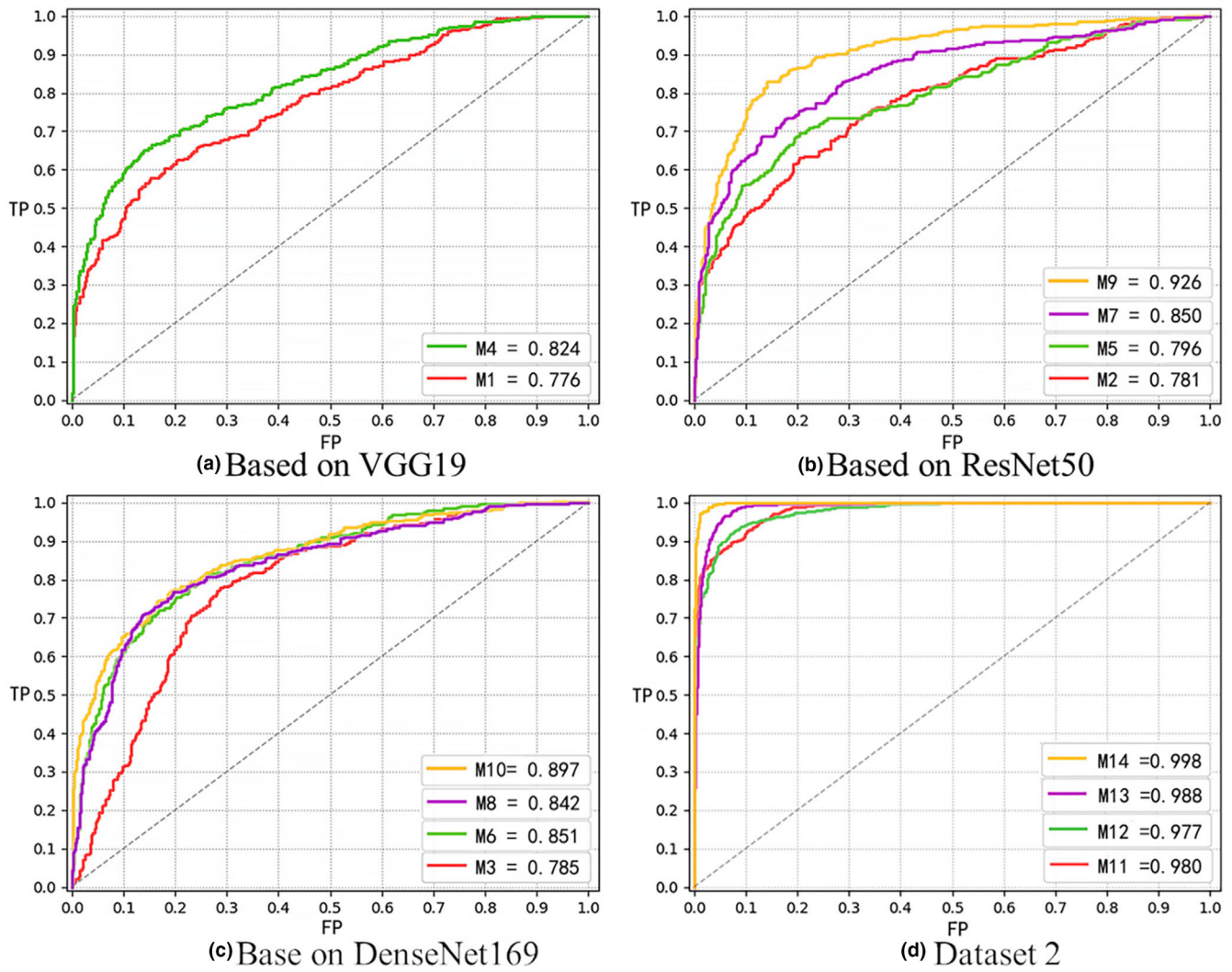
FIG. 9.   ROC curves of different models and different Datasets. (a) (b) (c) ROC curves of Dataset1, (d) ROC curves of Dataset2. M1: Base-VGG19, M2: Base-ResNet50, M3: Base-DenseNet169, M4: VGG19, M5: ResNet50, M6: DenseNet169, M7: SE-ResNet50, M8: SE-DenseNet169, M9: SEME-ResNet50, M10: SEME-DenseNet169, M11: MoEx-ResNet50, M12: MoEx-DenseNet169, M13: SEME-ResNet50, M14: SEME-DenseNet169. [Color figure can be viewed at wileyonlinelibrary.com]
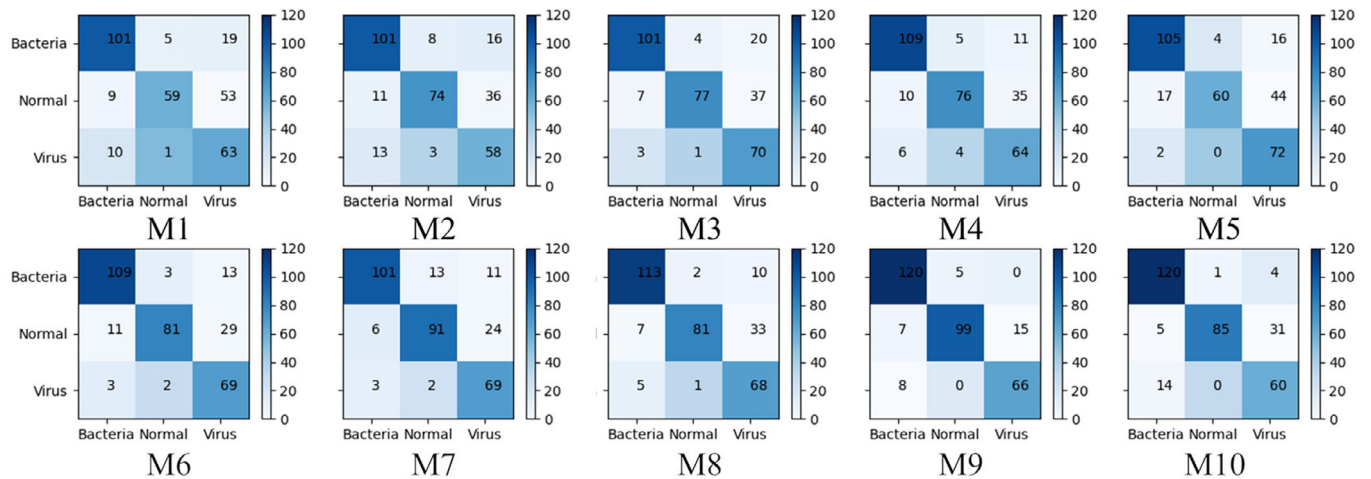

FIG. 10.   The confusion matrices of Models in Dataset1. M1: Base-VGG19, M2: Base-ResNet50, M3: Base-DenseNet169, M4: VGG19, M5: ResNet50, M6: DenseNet169, M7: SE-ResNet50, M8: SE-DenseNet169, M9: SEME-ResNet50, M10: SEME-DenseNet169. [Color figure can be viewed at wileyonlinelibrary.com]
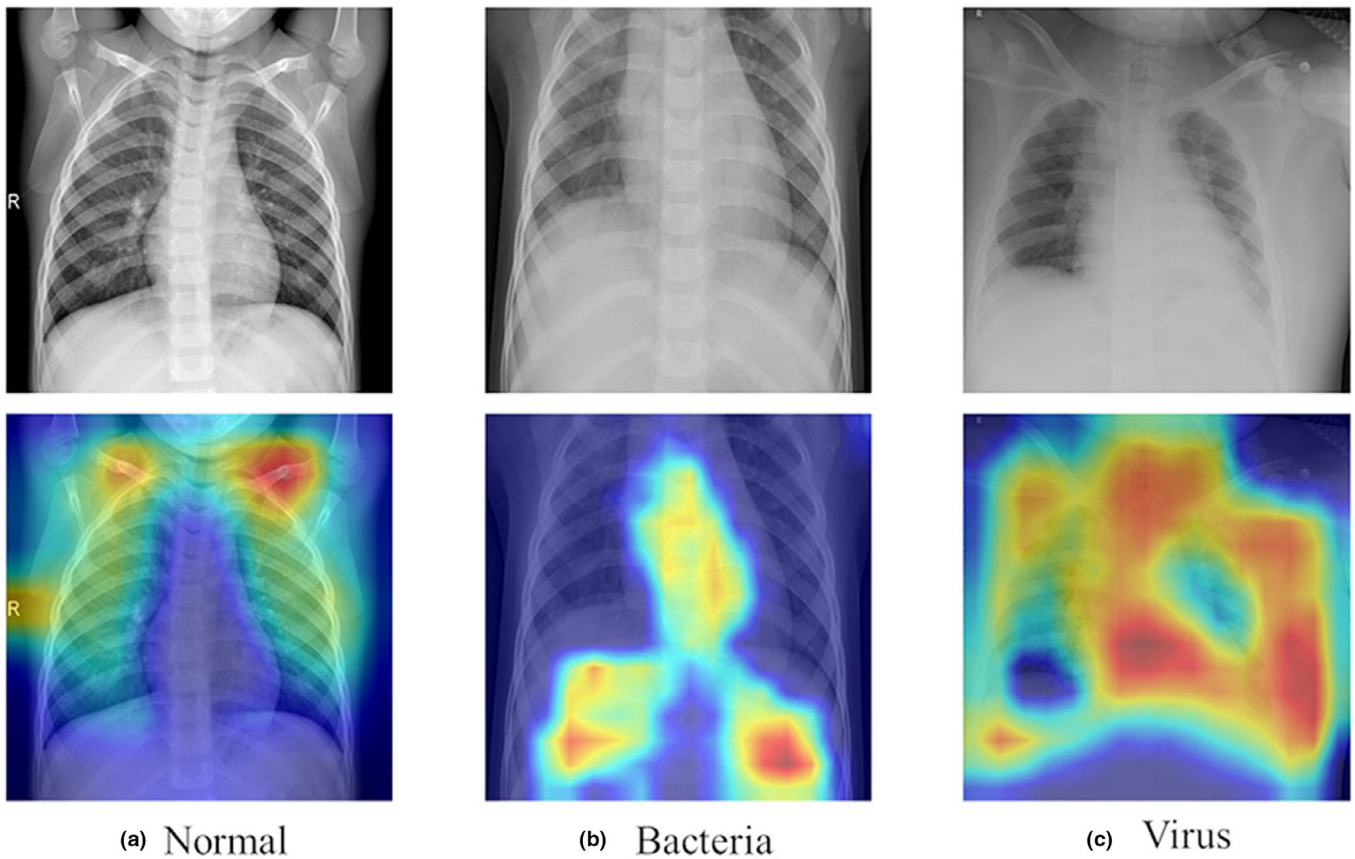
FIG. 11. The result of data visualization in Dataset1 by Grad-CAM. [Color figure can be viewed at wileyonlinelibrary.com]
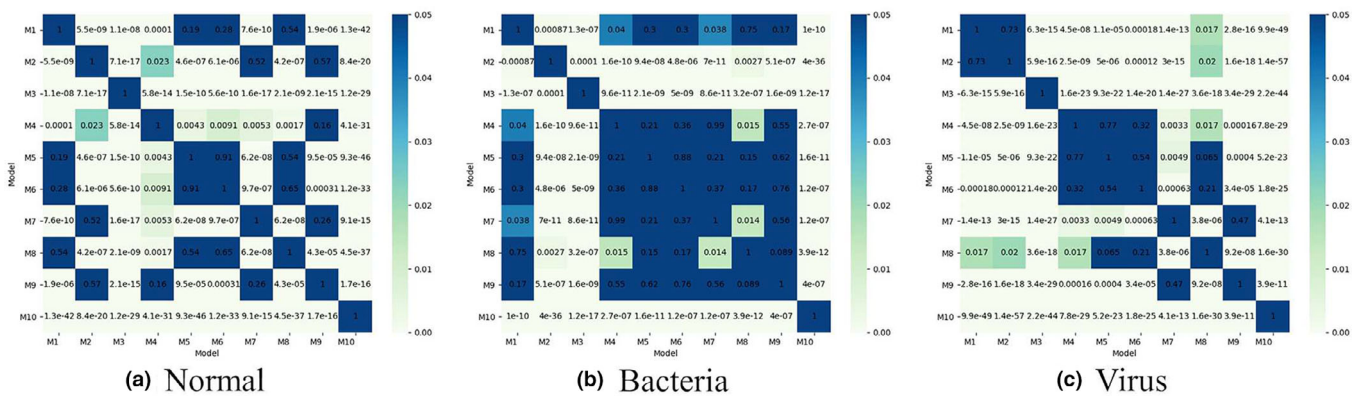

FIG. 12. The significance test of models in Dataset1. M1: Base-VGG19, M2: Base-ResNet50, M3: Base-DenseNet169, M4: VGG19, M5: ResNet50, M6: Dense-Net169, M7: SE-ResNet50, M8: SE-DenseNet169, M9: SEME-ResNet50, M10: SEME-DenseNet169. [Color figure can be viewed at wileyonlinelibrary.com]

Fig. 12, expect bacterial pneumonia, the addition of different improved structures lead to significant changes to the output of the network.

6.B. Fine-grained classification of COVID-19

The next step of Cascade-SEMEnet in identifying viral pneumonia is to perform fine-grained classification of viral pneumonias to diagnose COVID-19. We use Regional Learning in this process to mitigate the impact of non-lesion areas on the network.

6.B.1. Baseline

First, the VGG19, ResNet50, DenseNet169 networks are trained on the original data. The VGG19 results are shown in Fig. 13(b) as an example. The curves show a high fitting speed, the accuracy of the verification set reaches 97% and the AUC of the test set reaches 1.0. To explain this abnormal result, the data is visualised and it is noticed that the network converges to the label area in the chest radiographs, as shown in the example in Fig. 13(a). We speculate that this situation is due to the small number of chest radiographs in dataset2,
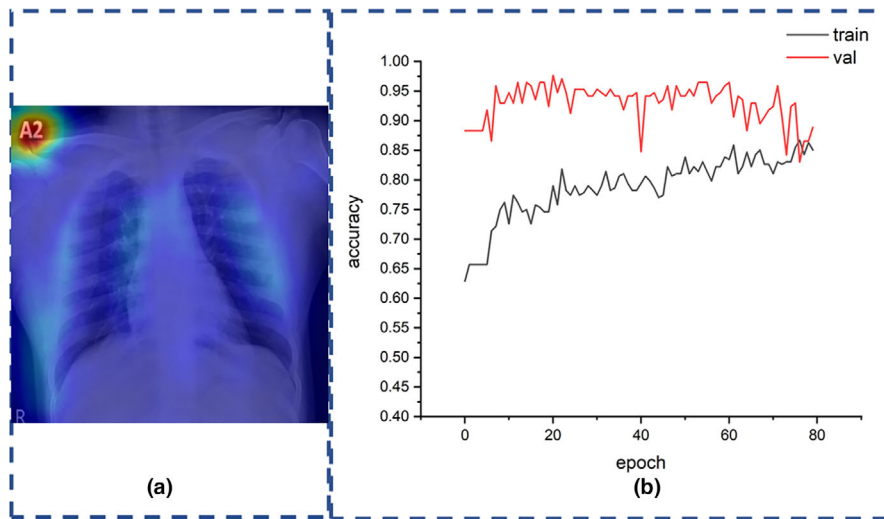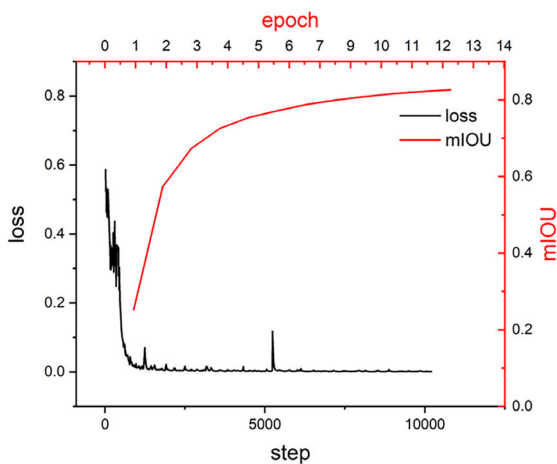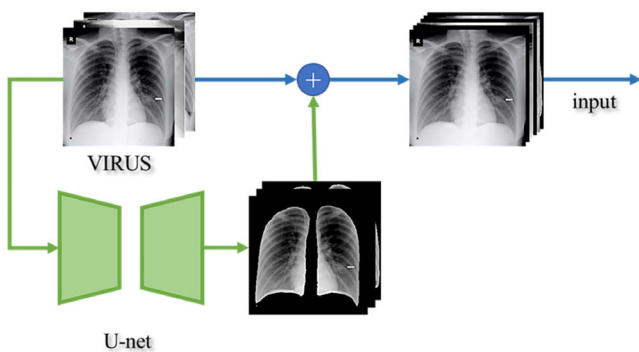
FIG. 13. (a) Grad-CAM based visualization result of the VGG19 network trained on dataset2 original data, (b) the corresponding training and validation curves. [Color figure can be viewed at wileyonlinelibrary.com]

(a) Train curve

(b) Flow chart of Regional learning

FIG. 14. (a) is the loss and mean IOU in the verification set during the U-Net training, (b) flow chart of Regional Learning. [Color figure can be viewed at wileyonlinelibrary.com]

which results in the interference of embedded text and numbers in the chest radiograph, and mislead the network to learn irrelevant non-lesion features.

TABLE III. The test indicators of the networks on Dataset 2.

Model	Recall	Specificity	Precision	F1-score	Accuracy	AUC
M11	**0.9938**	0.8547	0.8051	0.8895	0.9071	0.980
M12	0.8188	**0.9849**	0.9704	0.8881	0.9223	0.977
M13	0.9313	0.9717	0.9521	0.9415	0.9565	0.988
M14	0.9688	0.9736	**0.9568**	**0.9627**	**0.9718**	**0.998**

The bold numbers are used to represent the best values in each evaluation index.
*M11: MoEx-ResNet50, M12: MoEx-DenseNet169, M13: SEME-ResNet50, M14: SEME-DenseNet169.

6.B.2. U-Net segmentation and Regional Learning

To exclude influences of irrelevant features and direct the network to focus on the key characteristics of lung lesions, a U-Net model is trained and used to segment the lung area of the chest radiographs in Dataset2. The training loss and the mean-IoU of the verification set are shown in Fig. 4(a). Figure 14(b) presents a flow chart of Regional Learning in the network.

From the training curves of the networks [Fig. 8(d)], it can be seen that the convergence speed of SEME-ResNet50 and SEME-DenseNet169 are significantly faster than that of MoEx-ResNet50 and MoEx-DenseNet169. The evaluation results of the test set (Table III) show that before SE structure is added, the sensitivity and specificity of ResNet50 and DenseNet169 can not reach a good balance, limiting the F1-score to only 0.89. However, after adding the structure, the sensitivity and specificity of the model are both increased to above 0.9, which give the F1-score a boost. The confusion matrix is shown in Fig. 15, and the ROC curve is shown in Fig. 9(d), which is also increased from 0.98, 0.977 to 0.988, 0.998, respectively.

6.B.3. Subjective assessment of the model output and the diagnostic basis of COVID-19

The performance of the network is subjectively interpreted using Grad-CAM, as shown in Fig. 16. It can be observed that
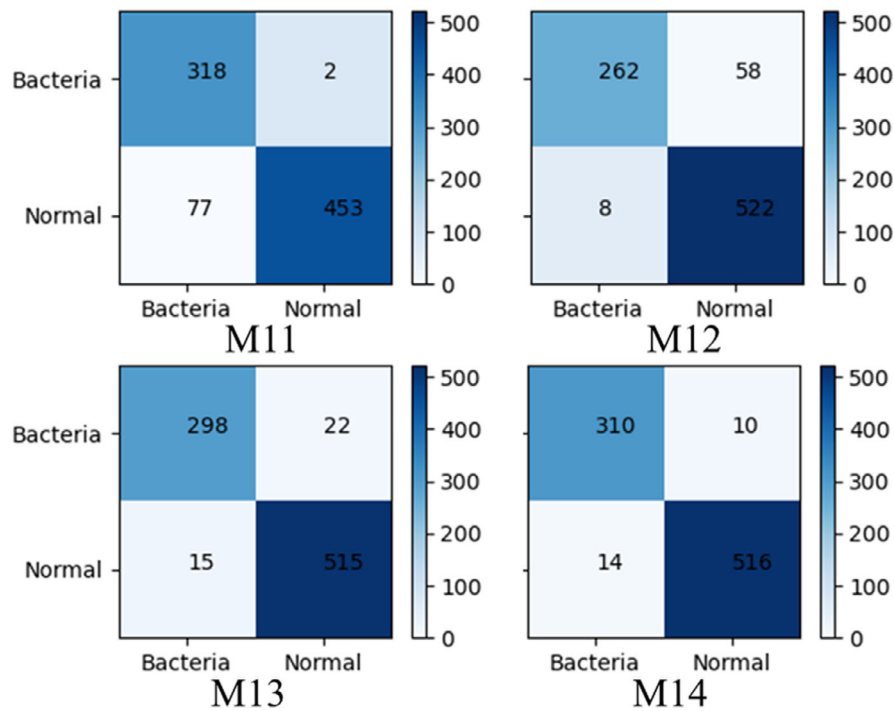
F_IG. 15. The confusion matrices of Models in Dataset2. M11: MoEx-ResNet50, M12: MoEx-DenseNet169, M13: SEME-ResNet50, M14: SEME-DenseNet169. [Color figure can be viewed at wileyonlinelibrary.com]
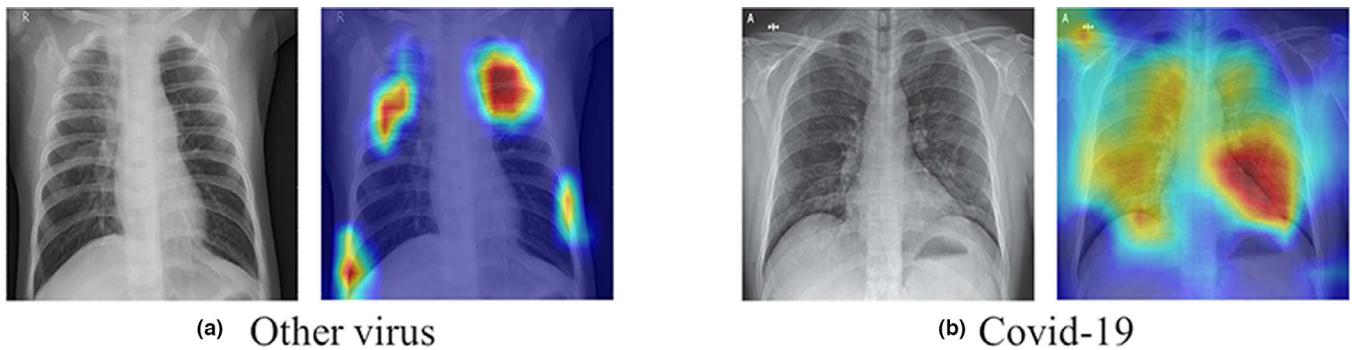

F_IG. 16. The result of data visualization in Dataset2 by Grad-CAM. [Color figure can be viewed at wileyonlinelibrary.com]

when the network categorises COVID-19 radiographs, the discriminant criterion occupies almost the entire lung. The heatmaps of COVID-19 radiograph in Fig. 16(d) shows that the network focuses on the lower right lung region in the process of deciding the type of chest radiographs. Recent research finds that the most common features present in radiographs of COVID-19 patients is airspace opacities,[9] often described as consolidation or frosted glass shadow. These features are also noticed by the networks in training. It can be seen from Fig. 16(c) that a similar description of the lesion appears in the lower right area, corresponding to the highlighted region in the heatmaps (d). These results show that the proposed framework can quickly detect COVID-19 cases and locate suspected lesion areas in the radiograph. Such methods provides reliable and objective evidences to help the physicians better analyse the chest radiographs of patients with suspicious COVID-19 symptoms. At this stage,

the significance test of every categories between networks is shown in Fig. 17, where it can be clearly seen that the improved structures have significantly changed the networks' output.

7. DISCUSSION

In this work, we focus on revealing the influence of optimization methods on network performance and the role they play in the task of chest radiograph assessment. Besides, different enhancement methods are considered to enhance the data and captured key features. After adding a global average pooling (GAP) layer, the network can effectively use more information contained in a larger sized image for learning, reducing the probability of false-negative and false positive. Adding an attention structure in the network can direct the network's attention to the specific feature channels in the task
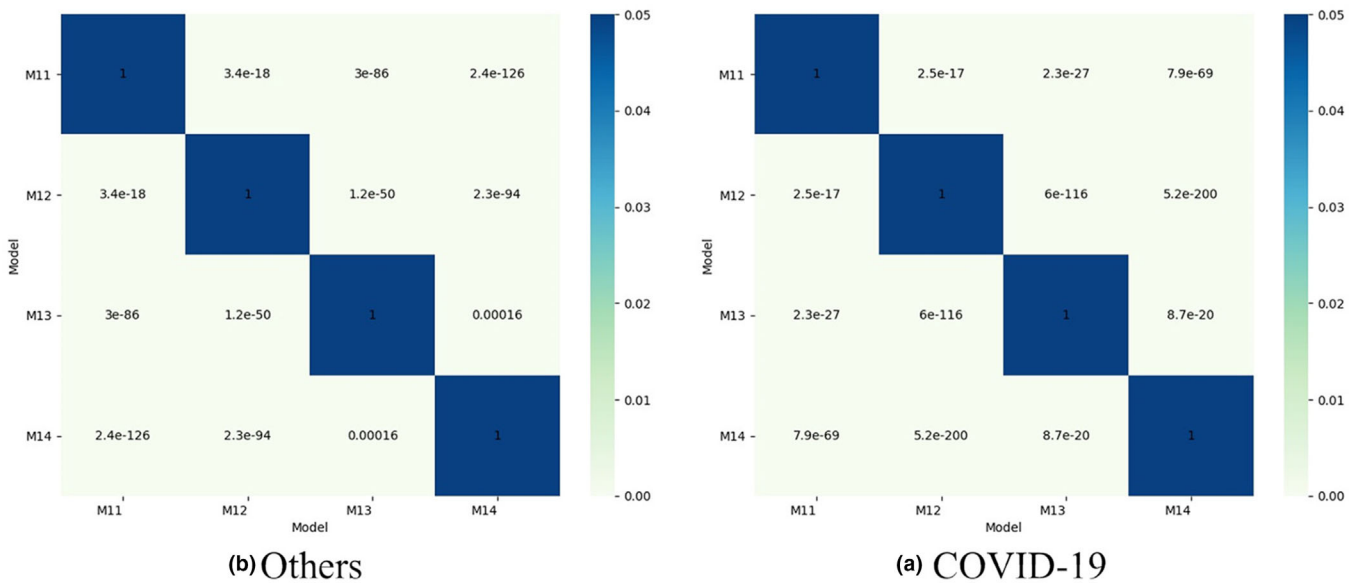
FIG. 17. The significance test of models in Dataset2. [Color figure can be viewed at wileyonlinelibrary.com]

of chest radiograph assessment, and improve the network's classification accuracy. In the experiment, we perform a significance test and several comparative analyses on the output of the network, and finally determine that ResNet and DenseNet are the most effective as the main structure of the proposed pneumonia detection network. It is worth noting that the convergence rate of DenseNet169 is not significantly improved when this structure is applied. We speculate that the dense merge layer of DenseNet itself may dilute the attention of it to details and affect the convergence speed of the network. In the evaluation phase, there are three categories in Dataset1. In order to get an overall indicator of the networks, we calculate the statistical indicators of each category respectively, and then average these indicators to get an overall evaluation. In addition, the results of the significance test among the outputs of each network show our improvement in the network can be reflected in the output of the results, further indicating that the upgrading of the model structure is beneficial to the final decision-making process of the network.

We notice that in dealing with small sized datasets, some non-pathological areas in the image becomes the focus region during network learning. Even with traditional image enhancement such as CLAHE, the network will still learn some wrong information from these regions, This information may lead the network to over-judge. An extreme example is that the accuracy of the verification set in VI.B.1 is higher than that of the training set. Regional Learning solves this problem well. Taking the diagnosis of COVID-19 as an example, when the lung area segmentation mask is superimposed on the original image retaining only the lung area, non-pathological features will be filtered out when they are extracted by networks as decision criteria. Such a regional learning approach will inevitably increase the loss, guiding the network to pay attention to relevant pathological features existing in the lung.

We reorganize Dataset 1 and evaluate our networks following the same evaluation methods used by Kermany et al[36] and Hu et al.[37] A comparative study is presented in Table IV. It can be seen that the accuracy of our proposed network structure works better in classifying pneumonia and normal cases by 0.2% compared to Hu et al.[37] Although the sensitivity is not the highest, the specificity is superior. In the task of classifying viral pneumonia and bacterial pneumonia cases, the proposed network outperforms Hu et al[37] by 2.9% in accuracy, and the sensitivity and specificity are both significantly higher. In addition, unlike the works by Minaee et al[19] and Das et al,[21] our model can diagnose other types of pneumonia besides COVID-19, avoiding mixing other types of pneumonia with COVID-19 cases. Compared with Minaee et al,[19] our proposed model improves the structure and achieves higher sensitivity and specificity (0.960/0.946 vs 0.968/0.974). Compared with Yoo et al,[22] our model has a higher accuracy (95% vs 97%) and theinfluence of different optimized structures on the network are also evaluated in detail.

TABLE IV. Comparative results of our method against methods proposed by Kermany et al.[36] and Hu et al.[37]

Class	Method	Accuracy (%)	Sensitivity (%)	Sepcificity (%)
Pneumonia vs Normal	Kermany et al[36]	92.8	93.2	90.1
	Hu et al[37]	93.4	**97.4**	86.8
	Ours	**93.6**	92.2	**94.9**
Bacterial vs Viral	Kermany et al[36]	90.7	88.6	90.9
	Hu et al[37]	91	88.5	92.6
	Ours	**93.9**	**94.1**	**96.2**

The bold numbers are used to represent the best values in each evaluation index.
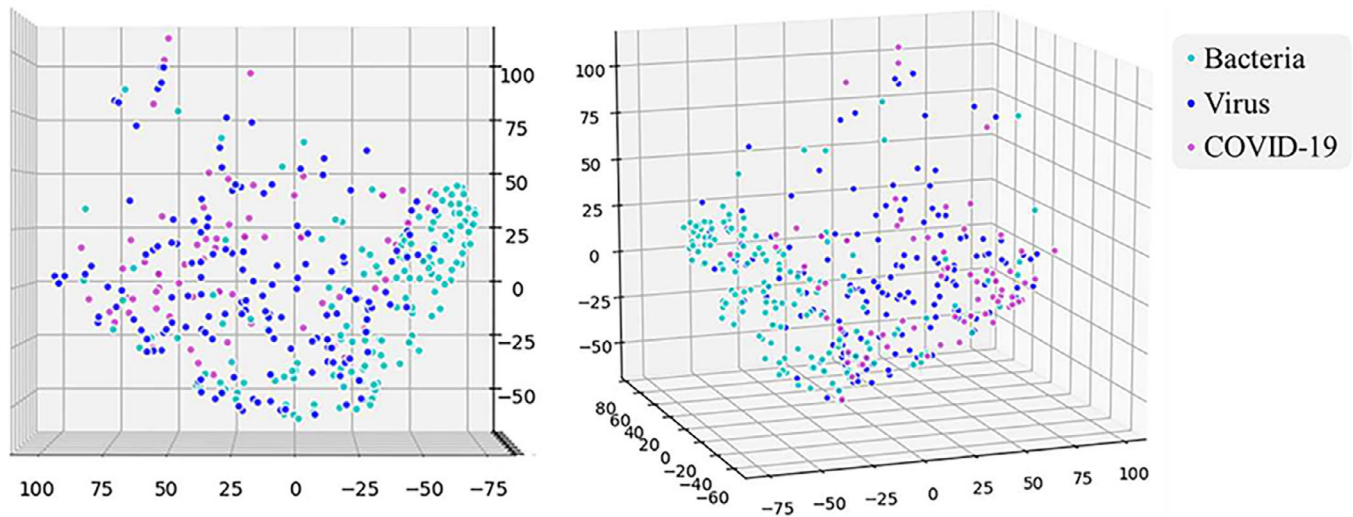
FIG. 18. Characteristic distribution in two different angles of chest radiograph focus areas in bacterial pneumonia, COVID-19 and other viral pneumonia. [Color figure can be viewed at wileyonlinelibrary.com]

We are in long-term collaboration with the Department of Radiology of the Second Affiliated Hospital of Zhejiang University and have produced important join research as in.[4] They conducted comprehensive subjective assessment on the output of Grad-CAM and conclude that the heatmaps of Grad-CAM demonstrate high agreement with expert radiologists in judging the area of pulmonary infection. We use Grad-CAM to segment the focus areas of bacterial pneumonia, COVID-19 and other viral pneumonia, apply hash-trac to perform feature dimensionality reduction on these focus areas, and then visualize the resulting characteristics in three-dimensional space as presented in Fig. 18. The results show that there are obvious boundaries between the three-dimensional feature map of bacterial pneumonia focus area (Turquoise dots) and viral pneumonia focus area (including COVID-19), while the feature map of COVID-19 (violet dots) and other viral pneumonia focus area (blue dots) overlap, with some minor differences in distribution. This is compliant with the fact that bacterial pneumonia radiographs show clear visual distinctions to that of viral pneumonia, and COVID-19 radiographs present a high visual similarity to those of other viral pneumonia cases. The underlying distinctive features of COVID-19 radiographs can be picked up by the proposed Cascade-SEME network, as evidenced by the evaluation results in Section 4.

8. CONCLUSION

Experiments show that the optimized structure proposed in this paper can effectively improve the performance of the network. After adding GAP to the network, the convergence speed of VGG19, ResNet50, and Densenet169 is improved, and the overfitting problem is alleviated to a certain extent. After joining the MoEx and SE structures, the indicators of ResNet50 and DenseNet169 on the test set s are also significantly improved. The proposed regional learning method effectively directs the network's attention to focus on relevant

pathological regions in the lung radiograph, ensuring the performance of the proposed framework even when a small training set is used. The visual interpretation step using Grad-CAM presents additional interesting findings. The region of attention on radiographs of normal, bacterial pneumonia and viral pneumonia are located in different regions of the lungs.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

a)Authors to whom correspondence should be addressed. Electronic mails: wangyaqi@cuz.edu.cn; sunll@hdu.edu.cn

REFERENCES

1. T. N. C. P. E. R. E. Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chinese J Epidemiol*. 2020;145–151.
2. Wadman M, Couzin-Frankel J, Jocelyn Kaiser CM. How does coronavirus kill? Clinicians trace a ferocious rampage through the body from brain to toes. *Science*. 2020;38:1885–1898.
3. Servick K. For survivors of severe COVID-19, beating the virus is just the beginning. *Science*. 2020;368(6489):359.
4. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2020;128:336–359. http://dx.doi.org/10.1007/s11263-019-01228-7.
5. Chandra TB, Verma K, Jain D, Netam SS. Localization of the Suspected Abnormal Region in Chest Radiograph Images. IEEE; 2020:204–209.

6. Kieu PN, Tran HS, Le TH, Le T, Nguyen TT. Applying Multi-CNNs model for detecting abnormal problem on chest x-ray images. *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, Ho Chi Minh City, Vietnam, 2018;300–305.

7. Sharma H, Jain JS, Bansal P, Gupta S. Feature Extraction and Classification of Chest X-Ray Images Using CNN to Detect Pneumonia. *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2020, 227-231. https://doi.org/10.1109/Confluence47617.2020.9057809.

8. Yaqi Wang QJ, Sun L. Enhanced Diagnosis of Pneumothorax with an Improved Real-time Augmentation for Imbalanced Chest X-rays Data Based on DCNN. *IEEE/ACM Trans Comput Biol Bioinf*. 2019;1–1.

9. Haouimi A, Bell DJ. COVID-19, radiopaedia; 2020.

10. Assistant R. COVID-19 Imaging findings. 2020.

11. Healthcare-In-Europe. Imaging the coronavirus disease COVID-19, 2020.

12. Kampalath R, Anju Goel M. Chest X-ray and CT Scan for COVID-19 When is medical imaging appropriate? Verywellhealth. 2020.

13. Jie W, Guan Z, Ni Y. et al. Clinical Characteristics of Corona virus Disease 2019 in China: 2020.

14. Bontempi D, Benini S, Signoroni A, Svanera M, Muckli L. CEREBRUM: a fast and fully-volumetric Convolutional Encoder-decodeR for weakly-supervised sEgmentation of BRain strUctures from out-of-the-scanner MRI. *Med Image Anal*. 2020;62:101688.

15. Wang S, Kang B, Ma J, et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19); 2020.

16. Wang J, Bao Y, Wen Y, et al. Prior attention residual learning for more discriminative COVID-19 screening in CT images. *IEEE Trans Med Imaging*. 2020;39:2572–2583.

17. Cohen JP, Dao L, Roth K, et al. Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning, 2020.

18. Zhu J, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, Duong TQ. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs; 2020.

19. Minaee S, Kafieh R, Sonka M, Yazdani S, Soufi GJ. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Med Image Anal*. 2020;65:101794.

20. Azemin MZC, Hassan R, Tamrin MIM, Ali MAM. COVID-19 deep learning prediction model using publicly available radiologist-adjudicated chest x-ray images as training data: preliminary findings. *Int J Biomed Imaging*. 2020;2020:1–7.

21. Das D, Santosh KC, Pal U. Truncated inception net: COVID-19 outbreak screening using chest X-rays, 2020.

22. Yoo SH, Geng H, Chiu TL, et al. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest x-ray imaging. *Front Med*. 2020;7:427.

23. Kermany D, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*. 2018;172:1122-1131.

24. Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv 2003.11597; 2020.

25. Linming W. The Dataset of Lung Segmention. 2018.

26. Zuiderveld K. Contrast Limited Adaptive Histogram Equalization. 1994.

27. Li B, Wu F, Lim S-N, Belongie S, Weinberger KQ. On Feature Normalization and Data Augmentation, arXiv:2002.11102 [cs.LG]; 2020.

28. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks, arXiv:1709.01507 [cs.CV]; 2019.

29. Cao X. A practical theory for designing very deep convolutional neural networks. 2015.

30. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV]; 2014.

31. He K, Zhang X, Ren S, Sun J. *Deep Residual Learning for Image Recognition*, arXiv:1512.03385 [cs.CV]; 2015.

32. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks, arXiv:1608.06993 [cs.CV]; 2016.

33. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers. Surpassing Human Level Performance on ImageNet Classification, arXiv:1608.06993 [cs.CV]; 2015.

34. Bell DJ, Jones J, Hacking C. *Pleural Effusion*. Radiopaedia; 2020.

35. Muzio BD, Weerakkody Y, Seikh Y. *Mediastinal Lymph Node Enlargement*. Radiopaedia; 2021.

36. Kermany DS, Goldbaum M, Cai W. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172:1122–31.

37. Hu M, Lin H, Fan Z, et al. Learning to recognize chest-xray images faster and more efficiently based on multi-kernel depthwise convolution. *IEEE Access*. 2020;8:37265–37274.