Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers in Biology and Medicine 149 (2022) 106065

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

COVID-19 diagnosis via chest X-ray image classification based on multiscale class residual attention

Shangwang Liu [a,b,*], Tongbo Cai [a,b], Xiufang Tang [a,b], Yangyang Zhang [a,b], Changgeng Wang [a,b]

[a] College of Computer and Information Engineering, Henan Normal University, Xinxiang, 453007, China
[b] Engineering Lab of Intelligence Business & Internet of Things, Henan Province, China

ARTICLE INFO

Keywords:
COVID-19 diagnosis
Chest X-ray (CXR)
Multiscale class residual attention (MCRA)
Pixel-level mixing
Heat maps

ABSTRACT

Aiming at detecting COVID-19 effectively, a multiscale class residual attention (MCRA) network is proposed via chest X-ray (CXR) image classification. First, to overcome the data shortage and improve the robustness of our network, a pixel-level image mixing of local regions was introduced to achieve data augmentation and reduce noise. Secondly, multi-scale fusion strategy was adopted to extract global contextual information at different scales and enhance semantic representation. Last but not least, class residual attention was employed to generate spatial attention for each class, which can avoid inter-class interference and enhance related features to further improve the COVID-19 detection. Experimental results show that our network achieves superior diagnostic performance on COVIDx dataset, and its accuracy, PPV, sensitivity, specificity and *F*1-score are 97.71%, 96.76%, 96.56%, 98.96% and 96.64%, respectively; moreover, the heat maps can endow our deep model with somewhat interpretability.

1. Introduction

Corona Virus Disease 2019 (COVID-19) is a respiratory pandemic caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [1]. It is reported that COVID-19 patients have a higher mortality rate due to the speed of transmission and high infection rate of the virus [2]. If the immune system is unable to fight the virus, the white blood cells will release cytokines and inflammatory mediators, making more immune cells, causing lung damage and possibly impacting other organs. Symptoms of COVID-19 include nasal congestion, sore throat, diarrhea, and drowsiness, in addition to fever and cough. Critically ill patients may present with respiratory distress, acute respiratory distress syndrome (ARDS), organ failure, etc [3]. From the available epidemiological characteristics, researchers have explored that the main routes of COVID-19 transmission are droplet and close contact transmission with an incubation period of 1–14 days [4]. Patients usually have no obvious adverse reactions or even symptoms during the incubation period, which makes disease detection challenging.

Nowadays, the main practice for diagnosing COVID-19 is Reverse Transcription-Polymerase Chain Reaction (RT-PCR) [5,6], which combines Ribonucleic Acid (RNA) reverse transcription and polymerase chain reaction (PCR) techniques to detect viral RNA fragments, and only a positive nucleic acid test can confirm the diagnosis. However, RT-PCR depends upon expensive equipments and takes at least 24 h to generate findings [7]. Moreover, it is not reliable enough and cannot satisfiedly rule out the chance that the patient is infected with 2019-nCoV. How to improve the efficiency of COVID-19 diagnosis and lower its cost become more emergent. To save limited medical resources, early diagnosis of COVID-19 can be detected by the cheap radiographic images instead of expensive RT-PCR indeed. Chest X-ray (CXR) and computed tomography (CT) are two typical kinds of radiographic images of the lungs. These two types of images are yielded because different organs have different ability to absorb X-rays, and then the abnormalities can be detected based on the contrast in these images. Comparing with CXR, CT has multiple levels of grayscale; but CXR is a more accessible, affordable, and popular method for diagnosing lung infections [8]. Furthermore, artificial visual interpretation of these images is time-consuming and relies heavily on the subjective judgment of the physician. For example, medical images are first annotated by a physician to generate a radiological findings report, and subsequently, the imaging findings are analyzed in conjunction with clinical experience to obtain a diagnosis. With an increase in the number of patients, they cannot be ensured to receive timely and effective treatment. Therefore, an effective and efficient CXR image understanding method of COVID-19 diagnosis becomes

* Corresponding author. College of Computer and Information Engineering, Henan Normal University, Xinxiang, 453007, China.
E-mail address: shwl2012@hotmail.com (S. Liu).

https://doi.org/10.1016/j.compbiomed.2022.106065
Received 15 May 2022; Received in revised form 7 August 2022; Accepted 27 August 2022
Available online 1 September 2022

urgent to prevent the virus continues to spread.

With the emerge of deep neural networks (DNNs) [9–14], especially convolutional neural networks (CNNs), they leverage multi-level layer neural networks for representational learning and are widely used for image classification [15,16], object detection [17,18] and semantic segmentation [19]. Naturally, DNNs are very good at detecting COVID-19 [20–25]. However, mainly due to lack of interpretability and practical skills, applying DNNs into CXR images for COVID-19 clinical diagnosis has run into obstacles [26]. Some literatures have reported that DNNs have achieved competitive performance in lung cancer [27,28], corneal endothelial cells [29], pancreas [30,31], polyp segmentation [32,33], etc. But, they are always prone to overfitting [34,35], poor robustness [36] and lack of generalization [37–39]. Furthermore, the lesion area of COVID-19 is usually small and blurred in CXR images, which needs us to pay more attention to these critical regions. More importantly, the "black box" [40–42] prevents DNNs from being plausible. For effective detection of COVID-19, Huang et al. [43] proposed a lightweight network LightEfficientNetV2, by using fewer parameters to overcome data shortages and obtain higher performance. Kumar et al. [44] combined graph convolutional network and convolutional neural network for determining the presence of COVID-19 infection in CXR images. Tang et al. [45] introduced the additional momentum method in the traditional BP neural network and enhanced the feature representation by adaptive histogram equalization, morphological processing, etc., which is beneficial to reduce the noise but fails to deal with the large scale focal lesion interference effectively. In this study, we propose the CNN-based multi-stage framework with multiscale class residual attention (MCRA) composed of feature representation enhancement and class spatial attention, which is an auxiliary inspection means for high-precision and automatic detection of COVID-19. Our major contributions are as follows:

1. Our MCRA network is with pixel-level mixing of local region. By mixing samples and focusing on critical regions, it can improve the feature area localization and the robustness of our model.
2. MCRA owes to multi-scale feature fusion and class residual attention. Multi-scale feature fusion facilitates the network to capture global contextual information at different scales and enhance feature representation, and class residual attention focuses more on category space region assignment to make more effective prediction.
3. The gradient-weighted class activation map is introduced, which takes into account of patch-wise disease probability to generate global heat map and endows our deep model with somewhat interpretability.

This paper is structured as follows: section 2 shows the works related to detecting COVID-19. Section 3 describes the proposed network. Section 4 provides the experiments, which contains the setup, experimental results and comparisons. Section 5 summarizes the full text.

2. Related work

This section presents the works related to our detecting COVID-19 method, including X-ray of COVID-19, pixel-level mixing of local region, attention mechanism and multi-scale feature fusion.

2.1. COVID-19 X-ray diagnosis

CXR is a common imaging modality that can provide an effective medical diagnosis. To achieve effective screening of patients with COVID-19, many experiments using deep learning to detect COVID-19 have been conducted [46–52]. However, the shortage of labeled data may affect the detection performance of COVID-19. Oh et al. [53] achieved volume expansion by using data augmentation, classifies the lung region after segmentation, and diagnoses COVID-19 with fewer training parameters to overcome the shortage of labeled X-ray images. Meanwhile, Wang et al. [54] proposed a novel CNN network, called "COVID-Net", and create a dataset of 13,975 X-ray medical images from 13,870 patients for the classification of CXR images in three categories: normal, pneumonia and COVID-19, reaching COVID-19 detection accuracy of 93.3%. Lu [55] applied the neural network to recognize the endoscopic image of upper digestive tract, by regarding 1335 cases of digestive tract endoscopic images as the dataset, and achieve an accuracy rate of 94.20%. Su et al. [56] put forward a multi-level thresholding image segmentation method, which introduces horizontal and vertical search mechanisms into Multi-Verse Optimizer, and achieves the improvement of global search and the ability to jump out of local optimum, but it is more time-consuming and ignores classification and prediction of lesions. Therefore, Ieracitano et al. [57] integrated CXR images with fuzzy features to overcome the uncertainty of CXR edge images, achieving COVID-19 classification accuracy rate of 81%. He et al. [58] employed artificial neural network (ANN) to construct a lung cancer recognition model and applied it into lung cancer lesion areas segmentation, assisting in the diagnosis of lung cancer.

2.2. Local region pixel-level mixing

Using cut regions for training images and mixing pixel-level information from other sample enables the network to identify targets from local views, which can improve localization and generalization capabilities of the model [59–63]. Classical representative local region pixel-level mixing methods include Mixup [64], Cutout [65] and CutMix [66]. Specifically, CutMix crops a part of the region and randomly fills the region pixels of other value in the training sample, which can obtain a more accurate mixed sample than traditional Mixup and Cutout, and the classification results are proportionally distributed.

2.3. Attention for DNNs

Attention mechanisms are extensively adopted in computer vision community, such as classification [67–72], detection [73–76] and segmentation [28,77–79]. On the basis of deep learning, CXR image classification can discriminates different pathologies by feature learning. Li et al. [80] integrated DenseNet with Graph Attention Network to reduce the amount of parameters in the network, achieving image classification accuracy rate of 94.8%. Feng et al. [81] presented condense attention (CDSE) and multi-convolution spatial attention (MCSA) to increase the redundancy of feature maps, and take full use of the relationship of the feature maps. Lin et al. [82] proposed an adaptive attention network (AANet), which pays attention to the context information via nonlocal interactions modeling, and needs less amount of parameters. Hence, attention mechanism could assist the deep network in quickly extracting key region information and improving visual representation.

2.4. Multi-scale feature fusion

Because CNN has a pyramidal multi-scale feature structure, it can build high-level semantic feature maps and improve the semantic representation of visual information with feature fusion. Doubtlessly, semantic information plays an important part in image classification [83], and performance of the image classification can be enhanced by associating semantic information between locations and attributes [84]. FPN (Feature Pyramid Networks) [85] is a top-down network, which can enhance semantic information by upsample and integrate the bottom-up ResNet feature layers with the same spatial size by laterally connecting them, and finally the contextual information can effectively improve the image classification. In addition, U-Net [86] performs well in channel dimension by concatenating the feature skip connection. Wang et al. [87] improved the performance of FCN in GI Tract lesion segmentation by fusing global contextual and local spatial information of images, merely using an average fusion strategy on multi-scale features instead of considering the differences between features at different scales. Bai

et al. [88] proposed a multi-feature dictionary representation and ensemble learning method based on symbolic aggregate approximation, which is different from extracting diverse-shapelet for early classification of time series [89]. Muralidharan et al. [90] adopted multiscale deep CNN with different combinations of modes, which is combined and merged by the fully connected layer for all features. He et al. [91] put forward an integrated framework COVIDNet, regarding ResNet [92] as the backbone and the spatial pyramid pooling(SSP) to enhance the middle-level features extraction; and the NetVLAD was employed to aggregate features from the low-level and context gating for learning. In this paper, we only leverage class residual attention integrated with multi-scale feature fusion to detect Covid-19 in CXR, and no any other level of feature fusion module is required.

3. Method

In this section, we proposed the framework of class residual attention (CRA) and multiscale feature fusion. The overview of our network structure is described in section 3.1. Class residual attention and multi-scale feature fusion are introduced in sections 3.2 and 3.3, respectively, and the class activation interpretable method is put forward in section 3.4.

3.1. Overview

The overall pipeline of our network is shown in Fig. 1.

As shown in Fig. 1, the proposed neural network algorithm is based on CRA mechanism. First, the input images are pre-processed, include data normalization and augmentation. Next, the model pre-trained on the ImageNet dataset are fine-tuned by transfer learning for feature extraction. Specifically, we propose multi-scale feature fusion to capture contextual information in CNN. Then, the feature maps obtained are feed into class residual attention module to obtain the spatial region assignment. Subsequently, the reassigned feature maps are input to the classifier to create three types of CXR image classification: COVID-19, Pneumonia and Normal. Finally, we conduct an interpretability study via the LayerCAM [93] visualization method.

3.2. Class residual attention

In COVIDx [54] dataset, each CXR image may contain one or more semantic feature lesion regions. While the location of the target lesion region usually cannot be detected directly or easily, we have to focus more on related regions for feature classification. With values ranging from 0 to 1, the attention module assigns an attention score to each category for allocating weights to the distinct feature spaces of each category. The higher the score value is, the higher the weight of the related position in the feature map is, and the feature representation at this position is thus enhanced, and vice versa. In short, the attention score can highlight relevant features while suppressing irrelevant ones.

We predict the relevant region for each class by using residual attention, which is consist of 1×1 convolutional layers (Conv2d) and a non-linear normalization layer (SoftMax), and then output a class channel attention score and a spatial attention to generate a certain class feature. We predict the attention scores of each class and use them for feature weight assignment directly via feature maps F. Our class residual attention structure diagram is illustrated in Fig. 2.

From Fig. 2, we can see the working procedure of CRA implementation. Specifically, we resize input image to 224×224; after convolution and feature extraction, the volume of feature map X is $256 \times 56 \times 56$ ($d \times h \times w$). Then, after a 1×1 convolution and flattening in the spatial dimension, the features are represented as $x_1, x_2, x_3, \ldots, x_{3136}$ ($x \in R^{256}$). Subsequently, a spatial pooling layer is employed to obtain the score attention, then the features are weighted to extract an

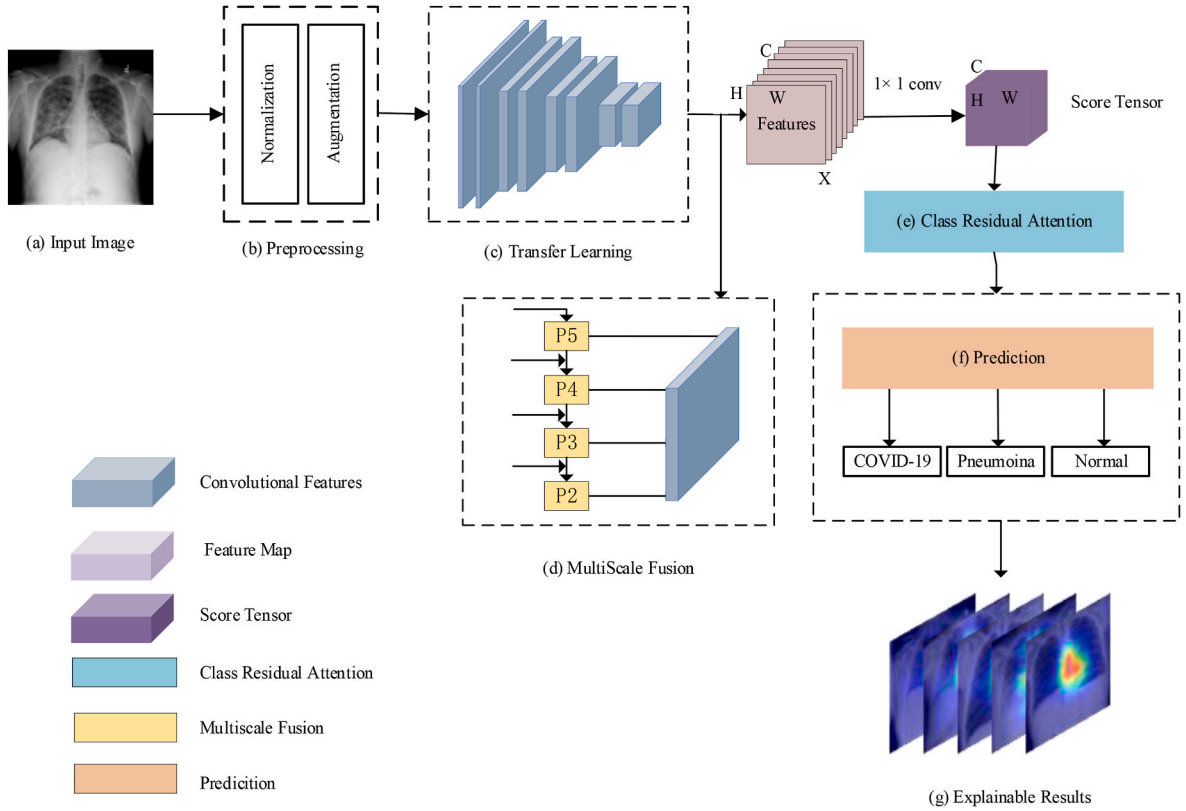


Fig. 1. Network framework. (a): Input image. (b): Preprocessing, which augments data of input image. (c): Transfer Learning, which is CNN (ResNet) or Transformer (ViT, Swin) model. (d): Multiscale Fusion, generating multi-scale features by CNNs. (e) Class Residual Attention, which generates spatial attention for each class. (f) Prediction, inferring image category probabilities. (g) Explainable Results, visualizing key features using class activation maps.

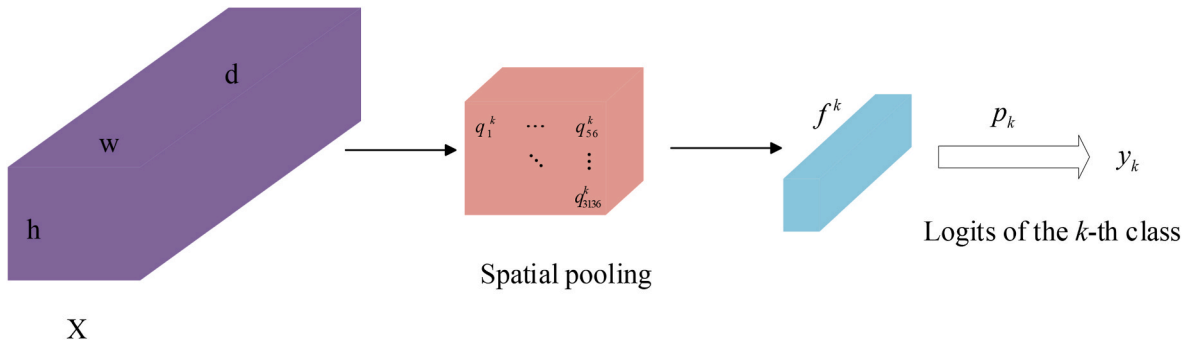


Fig. 2. The proposed class residual attention (CRA) is used to obtain the feature map of the input and predict the *k*-th classification result.

attention weight feature; p_k is the classifier for the *k*-th class, and finally, the feature is regarded as the input to their classifier to yield the final logits. For the *k*-th class and *j*-th position, the class attention score values can be computed from equation (1).

$$q_j^k = \frac{\exp(x_j p_k)}{\sum_{i=1}^{3136} \exp(x_i p_k)} \tag{1}$$

where q_j^k is the probability for class *k* at position *j*, and $\sum_{j=1}^{3136} q_j^k = 1$.

After obtaining the attention score values of the *k*-th category in each spatial location, its category features can be calculated from equation (2).

$$f^k = \sum_{i=1}^{3136} q_i^k x_i \tag{2}$$

where f^k is the feature of each category with size d*1.

Thus, the *k*-category features f^k acquired by weighting the attention score are utilized in the output of logits classification, as described in equations (3) and (4).

$$y^k = p_k f^k \tag{3}$$

$$\hat{y} \triangleq (y^1, y^2, y^3, \ldots, y^n) = (p_1 f^1, p_2 f^2, \ldots, p_n f^n) \tag{4}$$

where *n* means the number of classes.

3.3. Multi-scale feature fusion

For the input image *I*, the previous network structure usually does not take into account the spatial location relationship among local features, and ignores the semantic feature at the lower level. In our viewpoint, for the small size of COVID-19 pathology, it is the smaller target region information that is often missing after multiple convolutions that results in the decrease of accuracy.

Therefore, we add a multi-scale convolutional network to the convolutional layer of ResNet, connecting laterally in the {C1, C2, C3, C4} convolutional layers with top-down feature pyramid layers {P2, P3, P4, P5}. For ResNet-50, the feature maps with 256 * 56 * 56, 512 * 28 * 28,

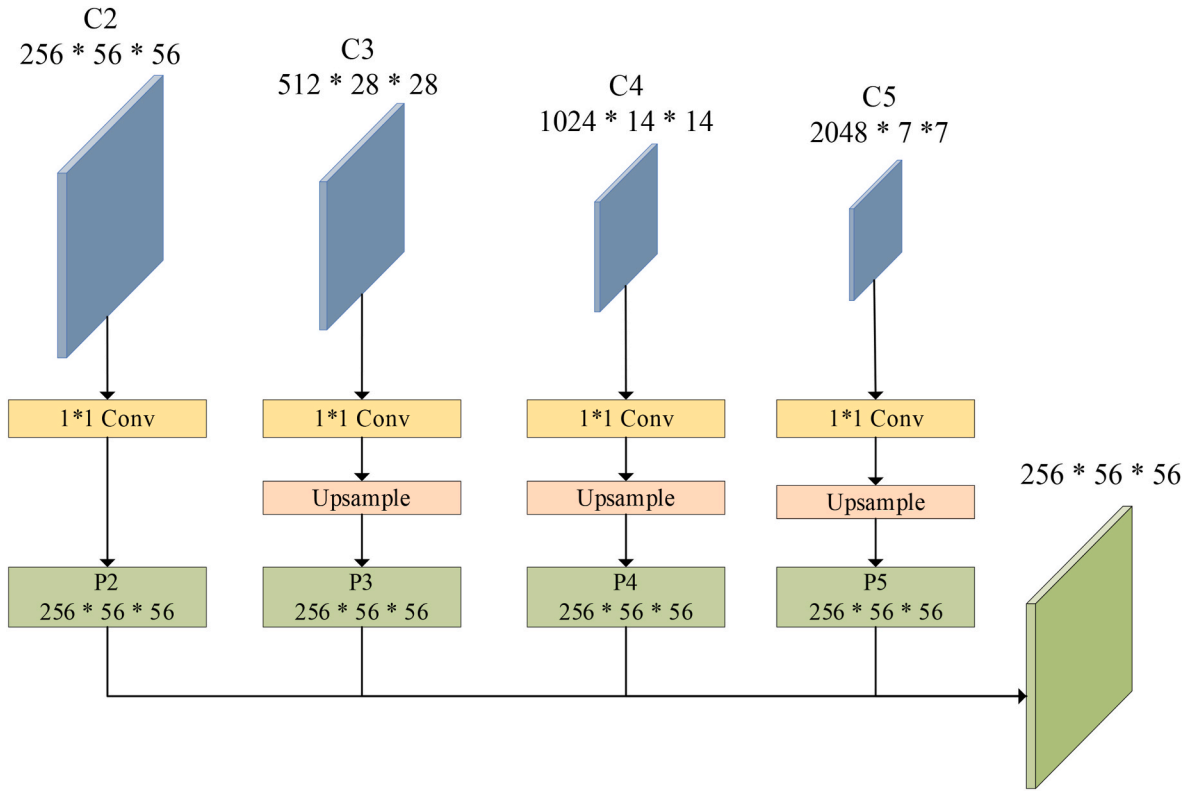


Fig. 3. Structure of feature fusion.

1024 * 14 * 14 and 2048 * 7 *7 resolutions are utilized for fusion; the structure of feature fusion is demonstrated in Fig. 3.

As shown in Fig. 3, C2, C3, C4 and C5 are different scale feature maps. The feature map is indicated by $F \in R^{c\times h\times w}$, where c, h and w denote the number of the channel, height and width, respectively. F can be calculated from equation (5).

$$F = \varphi(I;\theta) \tag{5}$$

where θ is the parameter in the convolutional neural network.

Attention map is generated by operating 1*1 convolutional layer on each feature map, then upsample it to the resolution of 256 * 56 * 56. The final multi-scale feature maps are obtained by P2, P3, P4 and P5 via residuals connection with resolution of 256 * 56 * 56. Especially, we take ResNet-50 as the backbone, the resolution of input image is 224 × 224, and the resolution of the output feature vector is 256 × 56 × 56. Thus, $F \in R^{c\times h\times w}$ is input into classifier, and class residual attention learns spatial weight allocation for information enhancement of related features.

3.4. Visualizing heatmaps of class activation

Deep model cannot be applied into clinical diagnosis plausibly for mainly lack of interpretability, so we utilize LayerCAM, an attribution-based (back propagation, gradient) visualization method, to make our deep model plausible. Unlike existing class activation maps GradCAM [94] and GradCAM++ [95], they can only generate class activation maps from deeper layers of the network, LayerCAM can obtain class activations for all layers of convolutional network, and endow our model with interpretability via hierarchical class activation maps. Therefore, the last convolutional layer of backbone network is replaced by LayerCAM's target-layer to create the final gradient-weighted class activation mapping in our model.

Since positive gradient is the predication of the class, it can be regarded as weights for each location in a feature map; if locations with negative gradients is set to be 0, then the weight of a coordinate (m, n) in the d-th feature map can be calculated as follows:

$$w_{mn}^{dc} = ReLu\left(g_{mn}^{dc}\right) \tag{6}$$

where c means the category of targets, g_{mn}^{dc} denotes the variance of the gradients.

The weight w^d is multiplied by the activation value of each position M^d, and the class activation map for each layer can be thus generated (see equation (7)).

$$T_{mn}^{d} = w_{mn}^{dc} \cdot M_{mn}^{d} \tag{7}$$

where T^d stands for the activation value with weight.

Finally, the activation weight values of all layers are accumulated to yield class activation map, which can be descripted by equation (8).

$$L^c = ReLu\left(\sum_k T^d\right) \tag{8}$$

where L^c refers to the class activation map by fusing channel dimension.

Hence, a variety of channels and spatial locations are obtained, and the weight can represent the relevance of distinct locations on a multi-class feature map.

4. Experiments

In this section, amounts of experiments are carried out to evaluate the performance of the proposed MCRA method. The dataset and the data augmentation strategy are introduced in section 4.1. The evaluation metrics are described in section 4.2. In section 4.3, the setup of the experiments is provided. The evaluation results of the MCRA method are presented in section 4.4. In section 4.5, a comparison with other state-of-the-art (SOTA) methods is discussed. In section 4.6, the ablation experiment is conducted to verify the performance of MCRA. The interpretability of our model is further analyzed by the visualizing heat map of the classification features in section 4.7.

4.1. Dataset and preprocess

We use the COVIDx dataset, which contains COVID-19, Pneumonia and Normal lung images from five different sources. In addition, we divided the dataset into training, validation and test sets in the ratio of 6:2:2 and then conducted several experiments to evaluate our model. Some example images of the COVIDx dataset are demonstrated in Fig. 4, and the number of images in each category is listed in Table 1.

As shown in Table 1, the total number of images in the COVIDx dataset is 30530. For the training dataset, the total number of images is 18318, with 4911 in Normal, 3393 in Pneumonia and 10014 in COVID-19. In either validation set or test set, the numbers of Normal, Pneumonia and COVID-19 images are 1637, 1131 and 3338, respectively.

The image names, paths and labels of the dataset are stored in arrays, and the labels are arranged for each image. It should be noted that the labels of Normal, Pneumonia and COVID-19 are represented as '0', '1' and '2', respectively, and the images and labels are then converted into arrays.

More the number of training images is, less the overfitting occurs during the training period. With respect to vertical flip, horizontal flip, random affine, color of brightness and contrast variation of the image, let the probability p = 0.5. Data augmentation strategy involves TrivialAugment [96] and CutMix [66]. TrivialAugment doesn't require any retrieval, just a simple augmentation strategy chosen at random; while CutMix can crop a part of one image and overlay it on another image to achieve the aim of the image enhancement. The final mixed results of Normal, Pneumonia and COVID-19 is resized to 224 × 224 and regarded as train samples, and some representative mixed results are shown in Fig. 5.

In Fig. 5, there are some example images' preprocessing, which consists of geometric transformation, cropping and mixing of images. While the image is fused with other parts, its soft label can be generated because the labels are statistically distributed probabilistically according to the image resolution. Please note that soft labels are one-hot vectors that consist of a series of successive floating-point values, rather than integer values such as 0, 1, 2, and so on.

4.2. Evaluation metric

Accuracy, *Positive Predictive Value* (*PPV*), *Sensitivity*, *Specificity* and *F*1-Score are adopted as evaluation metrics to verify COVID-19 diagnosis of the proposed method.

Accuracy is employed in medical analysis to measure the model's recognition effect, and it is defined by equation (9).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

where *TP*, *TN*, *FP* and *FN* are the numbers of true positives, true negatives, false positives and false negatives, respectively.

The true positive result vs. the whole positive result is called the *PPV* (see Equation (10)). If the *PPV* value is smaller, it is impossible to confirm the diagnosis of COVID-19, and a more accurate test is needed to make a definite diagnosis.

$$PPV = \frac{TP}{TP + FP} \tag{10}$$

Sensitivity of a disease is the percentage of people who successfully identify a disease, as described in equation (11). If the *Sensitivity* value is too low to screen COVID-19 cases effectively, patients will miss the early

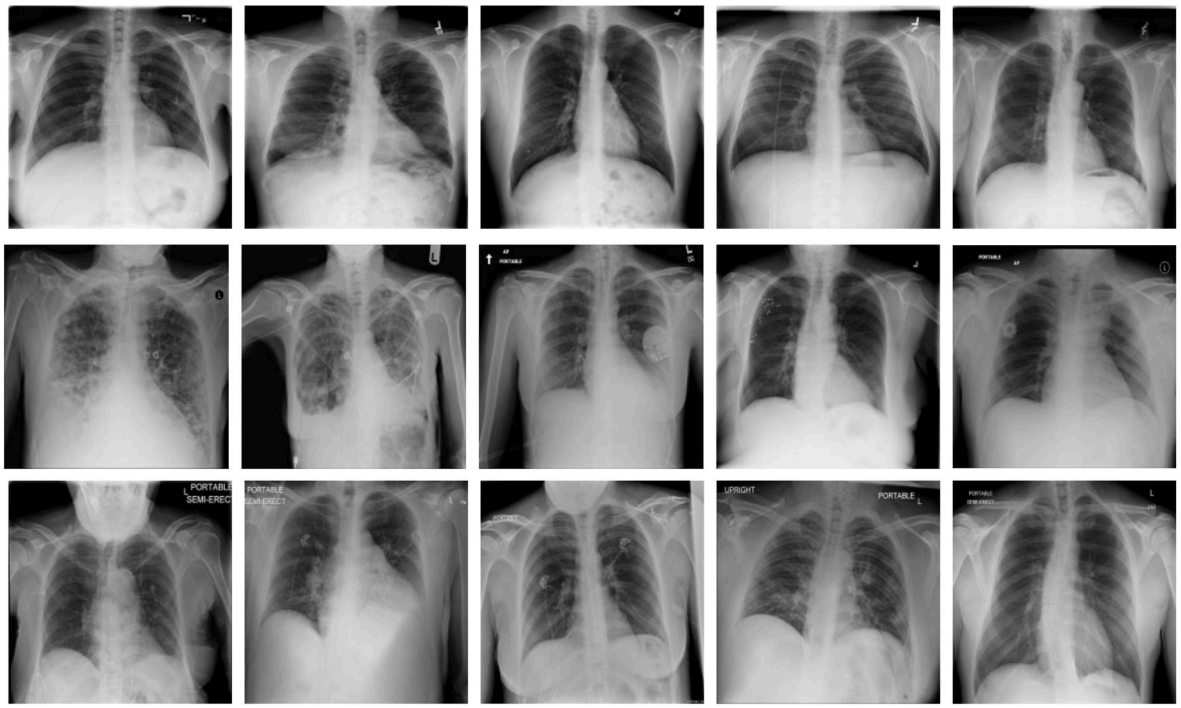


Fig. 4. Data samples: normal (top row), Pneumonia (middle row), and COVID-19 (bottom row).

Table 1
The COVIDx dataset.

Type	Normal	Pneumonia	COVID-19	Total
Train	4911	3393	10014	18318
Valid	1637	1131	3338	6106
Test	1637	1131	3338	6106

proper treatment.

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

Specificity denotes the proportion of people who have the correct diagnosis among those ones who do not have the disease, as described in equation (12).

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

*F*1-Score measures the accuracy of the classification model by calculating the summed average of PPV and Sensitivity, which can be calculated from equation (13).

$$F1 - score = \frac{2 \times PPV \times Sensitivity}{PPV + Sensitivity} \tag{13}$$

4.3. Experimental setup

Software environment: operating system, Ubuntu 22.04; programming language, Python 3.9; deep learning framework, PyTorch 1.11.0. Hardware environment: CPU, Intel i7-4790; GPU, NVIDIA RTX 3090 Ti (video memory, 24 G); RAM, 32 G.

The optimizer is Ranger, which is a co-optimizer combining RAdam (Rectified Adam) [97], LookAhead [98] and Gradient Centralization (GC) [99]; Loss is CutMixCrossEntropyLoss; learning rate (lr) is 1e-4 and weight decay is 1e-3. Furthermore, we resize the input image to 224 × 224, with a batch size of 64, for 50 epochs.

4.4. Experimental results

We regard ResNet-50, ViT-T/16 and Swin-T as backbone and load pre-trained models for training, respectively, while incorporating the proposed CRA into their own classification networks; specifically, we introduce a multi-scale feature fusion module in the ResNet. Moreover, we plot the loss and accuracy curves to evaluate the model performance. Loss is the value calculated by the loss function, and accuracy is the evaluation result of the model on the dataset according to the labels. The accuracy and loss curves of training and validation are shown in Fig. 6, where the horizontal coordinates stand for the training epoch and the vertical coordinates refer to the loss or accuracy values.

From Fig. 6, we can see that the learning curve of the model has a high training loss and a low accuracy at the beginning. As the training epoch increases, the loss value decreases, suggesting that the model is converging; and the accuracy value increases, showing that the accuracy of the model is improving. However, with the training loss and validation loss approaching each other, the training and validation losses will flatten out after a certain time, and the validation loss is slightly higher than the training loss. Comparing to Transformer (ViT, Swin) training process with oscillations for loss and accuracy, CNN is more robust, indicating that CNN model has better adaptability to the training set. After training completed, we select the best models from ResNet + CRA, ViT + CRA, Swin + CRA, and ResNet + MCRA for subsequent evaluation.

To evaluate the generalization ability of related models, we further evaluate the trained model on the test set, computing the evaluation metrics for COVID-19, Normal and Pneumonia images in terms of *accuracy*, *PPV*, *sensitivity*, *specificity*, and *F*1-Score. So, these numerical indices are listed in Table 2.

As shown in Table 2, CRA mechanism along with CNN (ResNet) improves the detection accuracy of COVID-19 from 96.87 to 97.25%, obtaining a gain of 0.38%; along with Transformer (ViT and Swin), the improvements of the accuracy are 0.36% and 0.26%, reaching 97.18% and 97.20%, respectively. But most importantly, the improvement of the accuracy reaches up to 0.84% in ResNet if we further introduce the multiscale feature fusion module: MCRA. It is worth noting here that

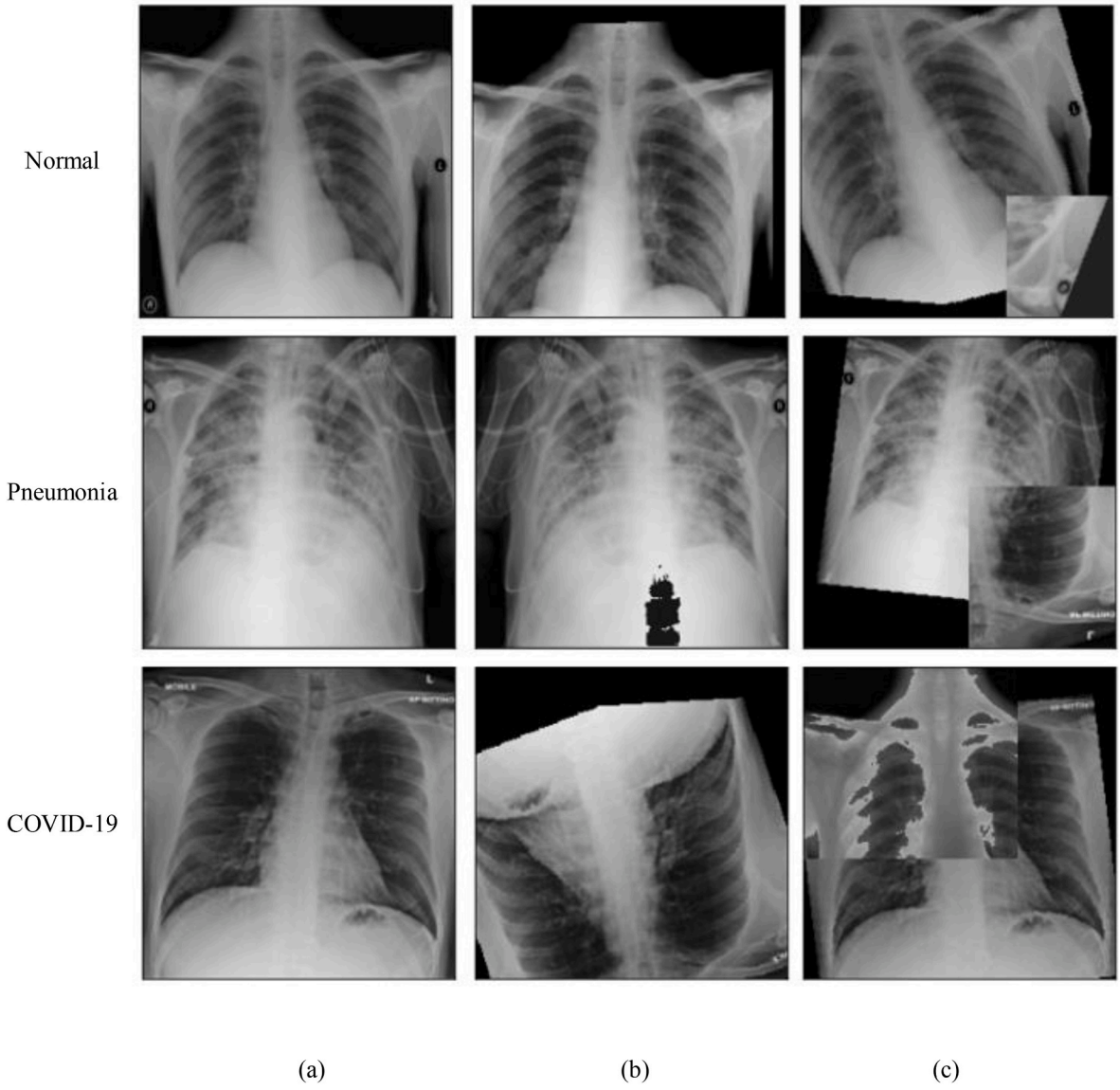


Fig. 5. (a) Original images. (b) Transformed images. (c) Mixed images.

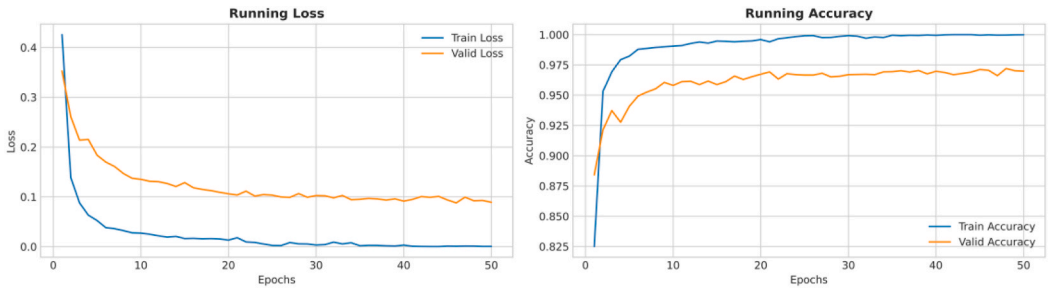
MCRA cannot be added to transformer methods (ViT and Swin) because of their patches structure. Why does CNN (ResNet) achieve superior performance than Transformer (ViT and Swin)? This is because that CNN extracts semantic features from local to global information by using stacking convolutional layers, and the stacked convolutional layers can keep expanding the field of perception until the whole image is covered. In contrast, Transformer extracts from global information, which is more difficult; furthermore, it also requires a larger amount of labeled data and a stronger data augmentation strategy to achieve better training results. For the three types of recognition, COVID-19 can achieve 99.85%, 99.52%, 99.81% and 99.68% for PPV, sensitivity, Specificity and F1, respectively. Comparing with Pneumonia and Normal, COVID-19 has the highest recognition accuracy and the best detection effect.

The ROC (Receiver Operating Characteristic) curve is a visual representation to evaluate the classification ability of a model. In our study, the ROC curves are for three categories. On the basis of four feature extraction networks (ResNet, ResNet-FPN, ViT, and Swin), the ROC curves of the proposed CRA for image classification are shown in Fig. 7.

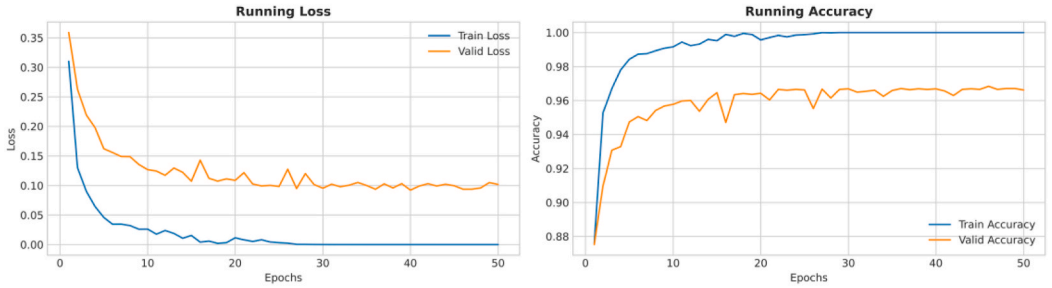
As shown in Fig. 7, the ROC results are almost consistent, all close to the upper left, with higher sensitivity. A larger ROC means better classification of the model and more accurate detection. Hence, all of the four models perform well on the COVIDx dataset, with AUC (Area Under Curve) of at least 0.99. Especially for COVID-19 diagnosis, it has the maximum AUC of 1.00, proving that the models are able to recognize COVID-19 correctly from other pneumonia and uninfected individuals.

As far as image classification task is concerned, confusion matrix is the most popular way to analyze misclassification, where true positives, false negatives, true negatives and false positives are adopted together to verify the relationship between actual and predicted values. For the multi-classification, it is a 3×3 table to record the number of samples from different categories that are correctly and incorrectly classified. The confusion matrices for each of the proposed ResNet + CRA, ResNet + MCRA, ViT + CRA and Swin + CRA models, to evaluate their image classification performance on the test set, are illustrated in Fig. 8.

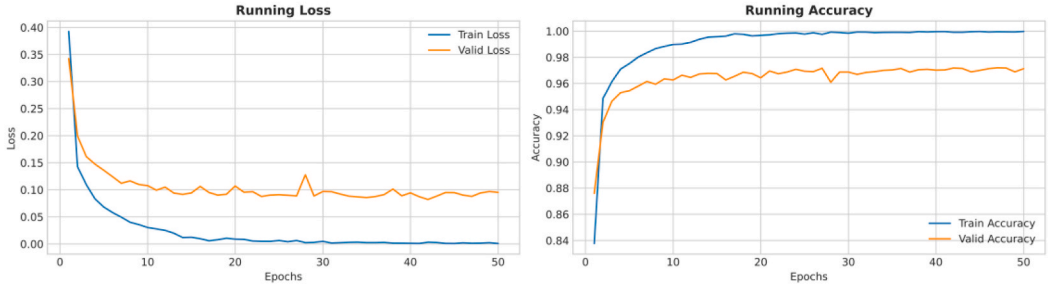
As shown in Fig. 8, the main diagonal element indicates the number of samples of different categories that are correctly classified. Comparing with ResNet + CRA, the ResNet + MCRA correctly predicted 3322 COVID-19 patients, which is higher than ResNet (3315 patients); ViT + CRA and Swin + CRA correctly predicted 3315 and 3313 patients, respectively, suggesting that MCRA has a better representation in a sample of 3338 COVID-19. Furthermore, ResNet + CRA and ResNet + MCRA also have lower errors than ViT + CRA and Swin + CRA for the



(a)



(b)



(c)

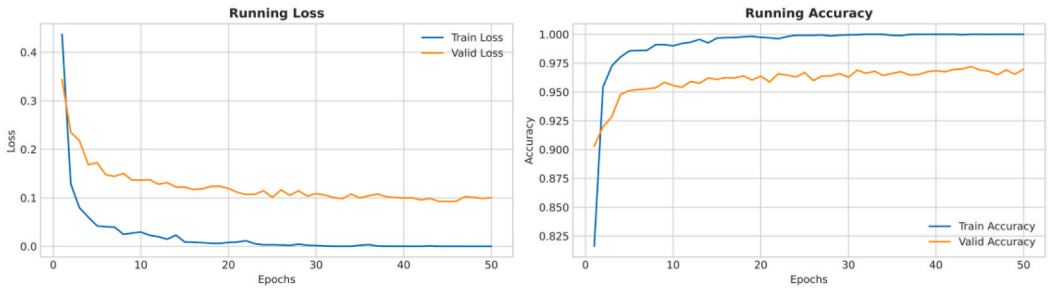


Fig. 6. Loss and Accuracy. (a) ResNet + CRA network with epochs. (b) ViT + CRA network with epochs. (c) Swin + CRA network with epochs. (d) ResNet + MCRA network with epochs.

detection of pneumonia. The reason that Pneumonia is often misclassified as Pneumonia is because their CXR images are extreamly similar, and so is COVID-19. Moreover, especially for normal, comparing with others, ResNet + MCRA has a maximum of 1596 correct classifications, which means that the misdiagnosis rate would obtain a notable reduction. As a result, ResNet + MCRA achieves the best performance, demonstrating the multi-scale feature fusion module constructs adequate feature representations.

The values of *Accuracy*, *PPV*, *Sensitivity*, *Specificity*, and *F*1-Score of seven models mentioned above, four of which were derived from CRA, are listed in Table 3.

From Table 3, we can see that our MCRA method achieves the highest accuracy in the ResNet, and its *accuracy*, *PPV*, *sensitivity*, *specificity* and *F1-score* reach up to 97.71%, 96.76%, 96.56%, 98.96% and

Table 2
Accuracy, *PPV*, *Sensitivity*, *Specificity*, and *F*1-Score of seven related models on each classes (i.e., COVID-19, Pneumonia and Normal).

Method	*Acc* (%)	COVID-19				Pneumonia				Normal			
		PPV	*Sen*	*Spe*	*F*1	*PPV*	*Sen*	*Spe*	*F*1	*PPV*	*Sen*	*Spe*	F1
ResNet-50 [92]	96.87	99.55	98.98	99.43	99.26	93.78	90.63	98.63	92.18	93.62	96.88	97.57	95.23
ViT-S [100]	96.82	99.58	99.31	99.46	99.45	92.01	91.69	98.19	91.85	94.55	95.30	97.97	94.92
Swin-T [101]	96.94	99.67	99.25	99.58	99.46	93.61	90.72	98.59	92.14	93.71	96.52	97.62	95.09
ResNet-50+CRA(Ours)	97.25	99.82	99.31	99.77	99.56	93.57	92.66	98.55	93.11	94.59	96.21	97.98	95.40
ViT-S + CRA(Ours)	97.18	99.73	99.31	99.66	99.52	94.48	90.80	98.79	92.61	93.92	97.25	97.68	95.56
Swin-T + CRA(Ours)	97.20	99.79	99.25	99.73	99.52	93.96	92.13	98.65	93.04	94.22	96.52	97.82	95.35
ResNet-50+MCRA(Ours)	**97.71**	**99.85**	**99.52**	**99.81**	**99.68**	**95.53**	**92.66**	**99.01**	**94.08**	**94.89**	**97.50**	**98.07**	**96.17**

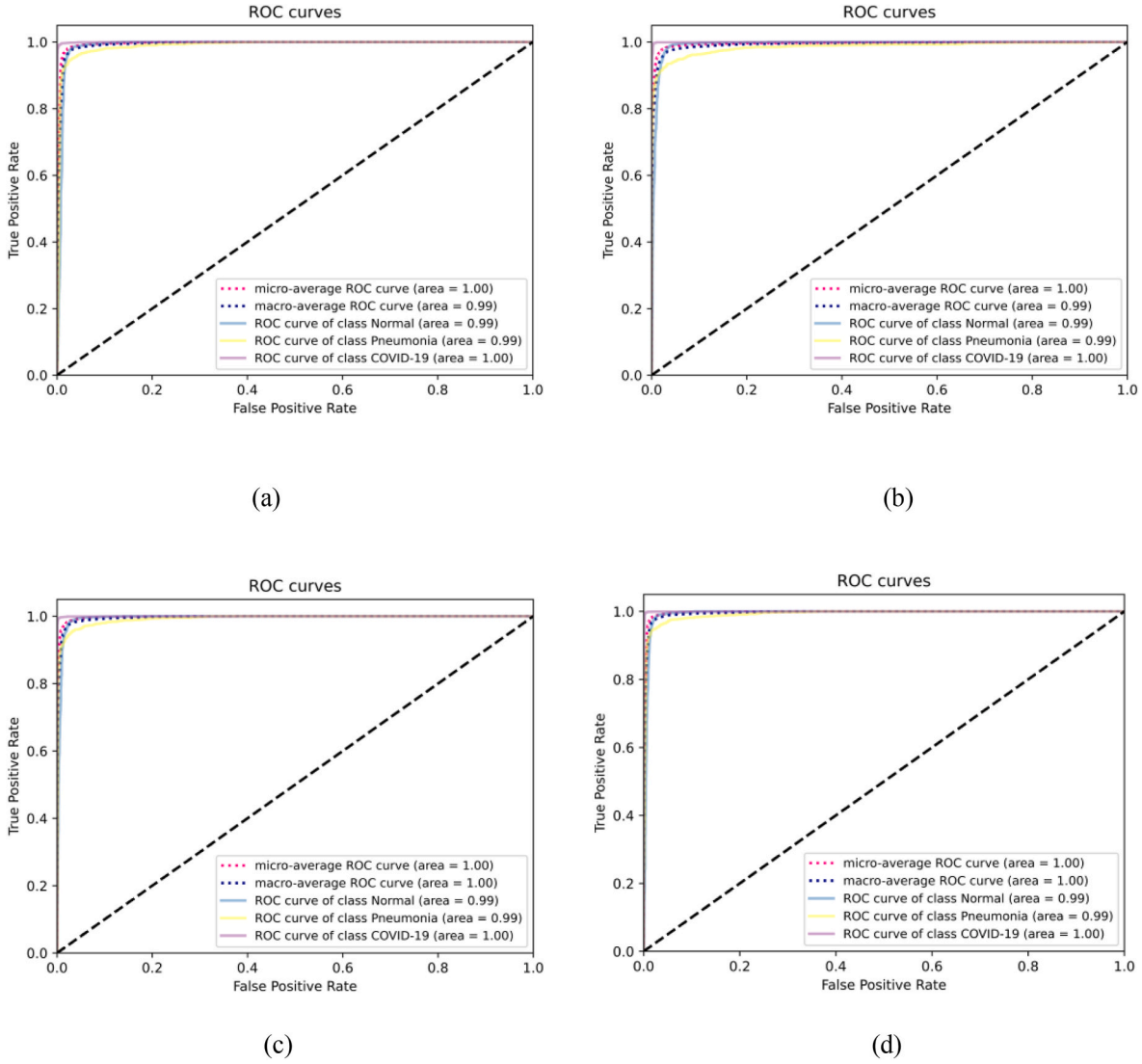


Fig. 7. ROC curves: (a) ResNet + CRA. (b)ViT + CRA. (c) Swin + CRA. (d) ResNet + MCRA.

96.64%, respectively. Without CRA, ResNet has the accuracy of 96.87%; Swin has 96.94%; and Vit has 96.82%. With CRA, ResNet has the accuracy of 97.25%, which is improved the most; ViT and Swin have 97.18% and 97.20%, respectively. We can also see that the CRA unit and multi-scale feature fusion module are effective in image classification and can improve the accuracy while lowering misclassification rates. It is worth noting again that MCRA cannot be integrated with transformer methods (ViT and Swin) because of their patches structure.

4.5. Comparing with SOTA methods

To compare the performance of ResNet-50-MCRA with previous state-of-the-art methods, we have established a comprehensive benchmark. The benchmark contains 17 state-of-the-art classification methods, including COVID-Net [54], VGG [102], DenseNet [103], Xception [104], Inception [105], ResNet [92], RegNet [106], EfficientNet [107], ViT [100], TNT [108], DeiT [109], Swin [101], VOLO [110], ConvNeXt [111], ConvMixer [112], CSWin [113] and PoolFormer [114]. The quantitative results are presented in Table 4.

As shown in Table 4, We calculated and compared the parameters

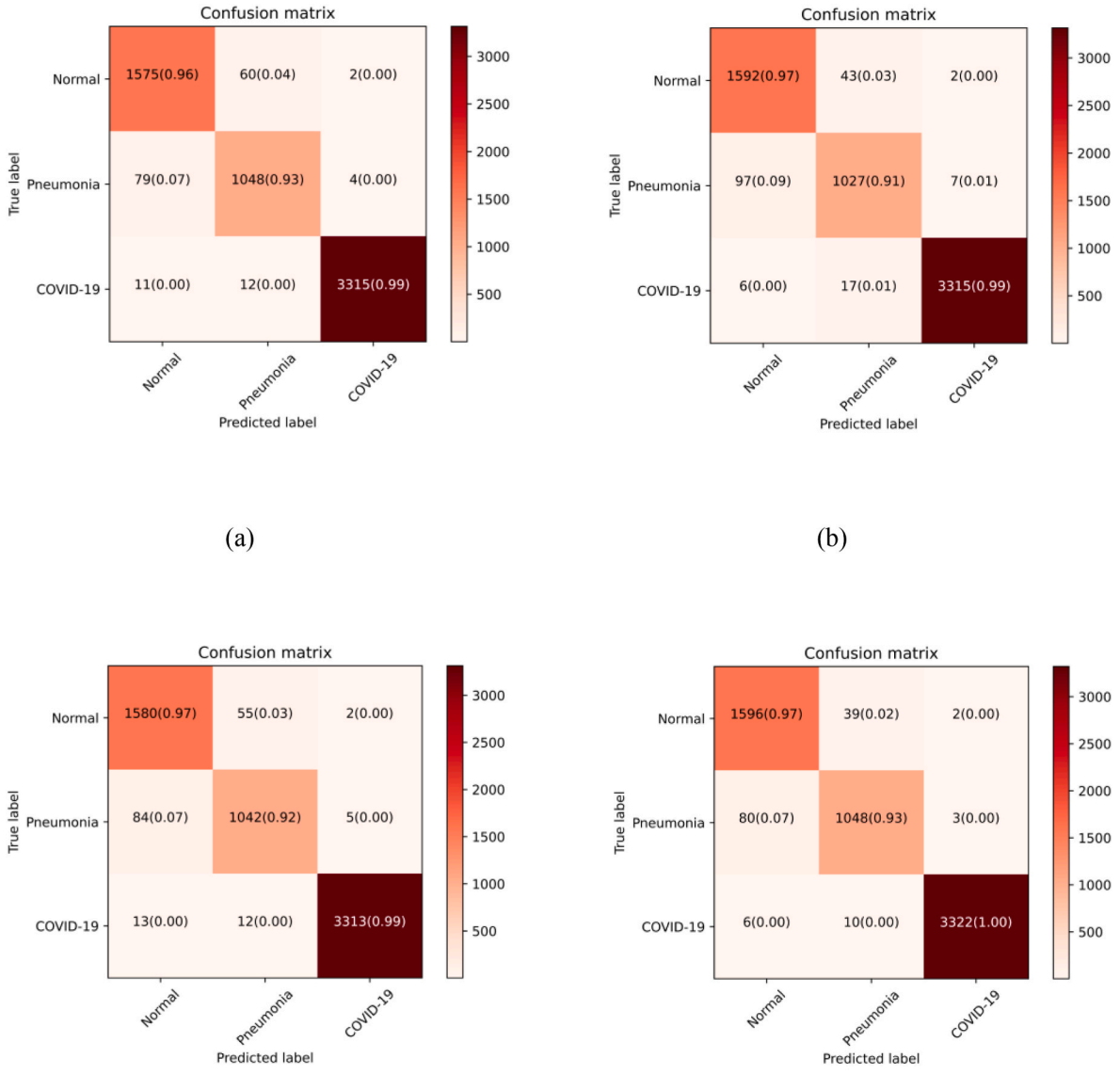


Fig. 8. Confusion matrix: (a) ResNet + CRA. (b) ViT + CRA. (c) Swin + CRA. (d) ResNet + MCRA.

Table 3
Accuracy, PPV, Sensitivity, Specificity, and F1-Score of seven models.

Method	Params(M)	FLOPs(G)	Accuracy (%)	PPV (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
ResNet-50 [92]	23.52	4.11	96.87	95.65	95.50	98.54	95.56
ViT-S [100]	21.67	4.61	96.82	95.38	95.43	98.54	95.41
Swin-T [101]	27.52	4.51	96.94	95.67	95.50	98.60	95.57
ResNet-50+CRA (Ours)	23.52	4.11	97.25	96.00	96.06	98.77	96.03
ViT-S + CRA (Ours)	**21.67**	**4.61**	97.18	96.04	95.79	98.71	95.89
Swin-T + CRA (Ours)	27.52	4.51	97.20	95.99	95.97	98.73	95.97
ResNet-50+MCRA (Ours)	26.85	6.96	**97.71**	**96.76**	**96.56**	**98.96**	**96.64**

and FLOPs of ResNet-50+MCRA with other methods, FLOPs means the efficiency of floating point operations, and networks with similar FLOPs do not necessarily perform at the same speed, but provide some reference since model complexity can be measured. For CNNs, the accuracy of VGG-19 is 96.54%, with the highest parameters and FLOPs of 139.58 M and 19.63G, owing to its 3 fully connected layers. As DenseNet uses concatenate instead of ResNet's addition operation for skip Layer, its number of params is minimum with merely 6.96 M while its accuracy reaches up to 96.56%. The smallest FLOPs model is COVID-Net, for 0.42G, with an accuracy rate of 94.76%. The parameters and FLOPs of ResNet-50+MCRA with 97.71% accuracy are 26.85 M and 6.96G, respectively, similar to VOLO with 97.49% accuracy. Comparing to

Table 4
Comparison of quantitative results of seventeen state-of-the-art classification methods on COVIDx.

Method	Params(M)	FLOPs(G)	Accuracy (%)	PPV (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
EfficientNet-B5 [107]	28.35	2.40	94.15	91.86	92.10	97.22	91.96
CSwin-T [113]	21.81	4.34	94.51	92.65	92.35	97.29	92.44
COVID-Net [54]	12.07	**0.42**	94.76	92.93	92.28	97.42	92.55
VGG-19 [102]	139.58	19.63	96.54	95.27	95.00	98.41	95.08
DenseNet-121 [103]	**6.96**	2.86	96.56	95.08	95.16	98.44	95.10
Xception [104]	20.81	4.57	96.66	95.23	95.43	98.46	95.33
Inception v3 [105]	21.79	2.85	96.74	95.66	95.12	98.49	95.31
ViT-S [100]	21.67	4.61	96.82	95.38	95.43	98.54	95.41
ResNet-50 [92]	23.51	4.11	96.87	95.65	95.50	98.54	95.56
Swin-T [101]	27.52	4.51	96.94	95.67	95.50	98.60	95.57
DeiT-S [109]	21.67	4.61	96.95	95.53	95.57	98.61	95.55
TNT-S [108]	23.37	5.24	97.02	96.05	95.30	98.58	95.62
ConvNeXt-T [111]	27.82	4.47	97.13	96.02	95.73	98.65	95.86
ConvMixer [112]	20.34	19.55	97.22	96.18	95.83	98.70	95.99
RegNet-8GF [106]	37.17	8.0	97.33	96.31	95.94	98.77	96.10
PoolFormer -S36 [114]	30.35	5.0	97.36	96.20	96.19	98.82	96.18
VOLO-D1 [110]	25.86	6.87	97.49	96.34	96.22	98.87	96.28
ResNet-50+MCRA(Ours)	26.85	6.96	**97.71**	**96.76**	**96.56**	**98.96**	**96.64**

ResNet-50, in terms of parameters and FLOPS of 23.51 M and 4.11G, ResNet-50+MCRA increases by 3.34 M and 2.85G, gaining the benefit of multi-scale feature fusion. Furthermore, our method is superior to most transformer-based methods, not only in terms of efficiency, but also regarding performance. In general, these methods obtain a larger perceptual field and contextual information by enlarging the depth of the network. However, the COVID-19 lesion regions are small and blurred in CXR images, many useful information would be lost. Nevertheless, our proposed multi-scale feature fusion network can contribute to capture global contextual information at different scales to improve feature representation. Meanwhile, The CRA assigns attention weights to enhance relevant features and suppress irrelevant features for specific categories, making our ResNet-50+MCRA excellent in terms of Accuracy, PPV, Sensitivity, Specificity, F1-Score metrics on the COVIDx. We attribute the performance improvement to our class residual attention and multi-scale feature fusion modules, which provide robust feature representation. Hence, our method achieves superior performance than previous methods.

4.6. Ablation experiments

To analyze the impact of each module within our method, we conducted ablation experiments by taking ResNet-50 of CNN network as backbone, which includes using CutMix, CRA and multi-scale fusion module FPN. The ablation experimental results are listed in Table 5.

In Table 5, First, to utilize all training data, we use CutMix to execute data augmentation, which includes converting the dataset from hard labels to soft labels and adding other objects to the cut region so that the model can recognize objects from the local view to enhance the model's localization capability, achieving an increase in accuracy rate from 96.50% to 96.87%. Secondly, the introduction of CRA leads to the accuracy rate achieves 97.25%, gaining from the focus on category space region assignment, which resulting in more accurate predictions. Thirdly, the multi-scale fusion module FPN fuses low-level details with high-level semantic information to capture global contextual information at different scales to improve the performance of classification, reaching up to 97.71% accuracy rate in detecting the COVID-19. Last but least, the *PPV*, *sensitivity*, *Specificity* and *F*1-Score have all improved after each module.

4.7. Interpretability of deep model

CRA is good at multiple object image classification, which can be also verified by heat map, a kind of attentional visualization. We leveraged class activation maps to illustrate the heat map. The heat map can find several object areas of an image easily by using different colors, indicating the area is infected with COVID-19 or not; obviously, heat map is of somewhat interpretability for a deep model. The heat maps of ResNet, ResNet + CRA and ResNet + MCRA are shown in Fig. 9.

In Fig. 9, on the basis of the attention score of CRA, we can visualize CXR image with different colors. For the deep neural model, the key infected area usually tends to be redder around with different colors, and often has dark red centers. The higher the attention score is, the darker the red center is and the greater the risk of the patient's lungs infecting in that area is. On the fringes of infected area, the blue-green colors still distinguish the relevant area more carefully till the purple indicates the body organ or part has less influence on the model prediction. From Fig. 9 (c) & (d), we can see that our CRA unit and multi-scale feature fusion module can more accurately locate the feature regions which is very important to the diagnosis of COVID-19.

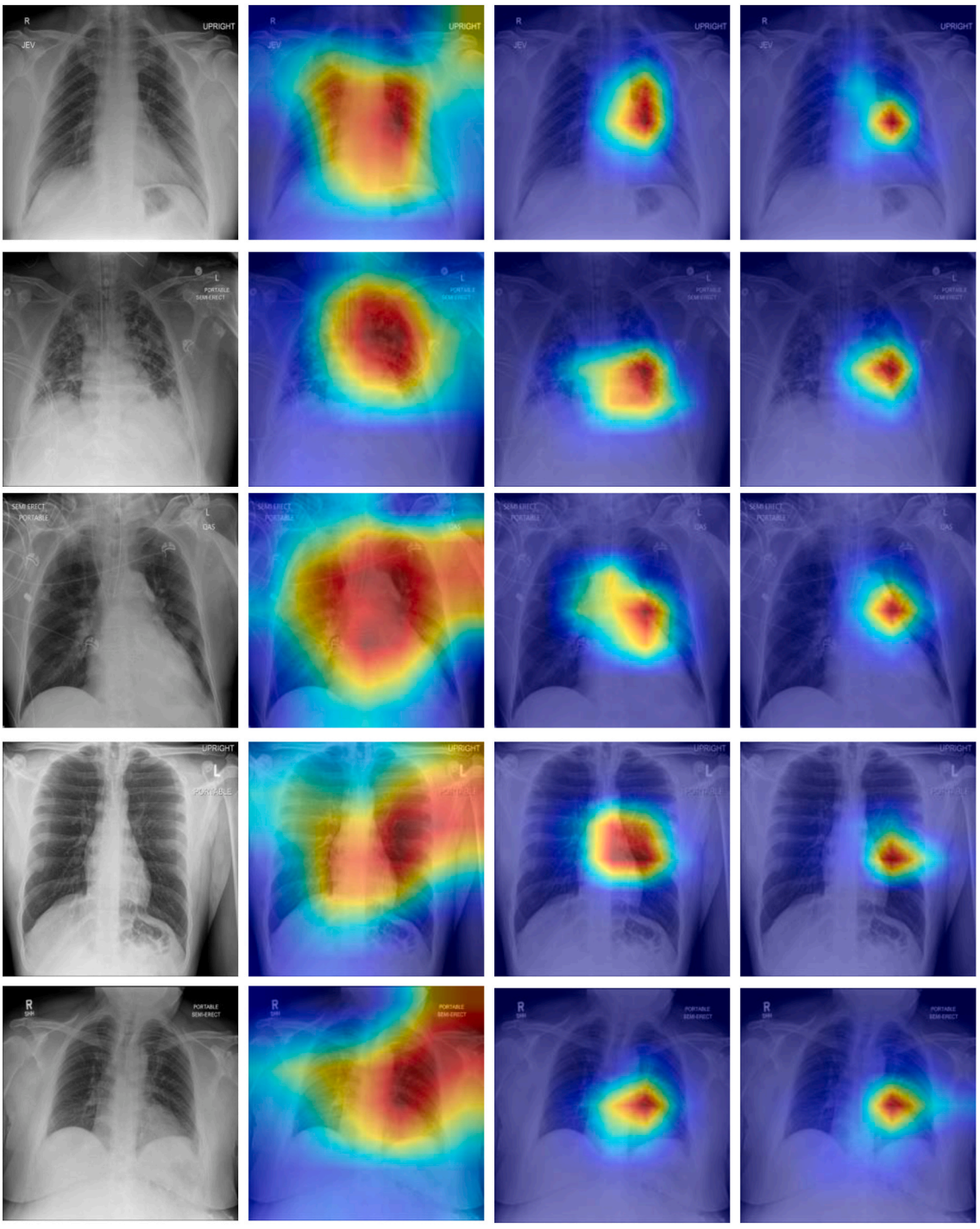
That is to say, these heat maps can make our deep model plausible to some extent, because it is visual, accurate, generalized and traceable for its diagnosis processes or results.

5. Conclusion

In this study, we propose a class attention called CRA for the detection and classification of COVID-19 patients from X-ray medical images. Unlike previous work, our method utilizes attention for classification to minimize the disturbance of irrelevant categories, and focus more on the target category to improve multi-classification task processing performance. The attention is applied to both CNN (ResNet) and Transformer (ViT and Swin) models. We also introduce strategies such as CutMix and multiscale feature fusion FPN (only for CNN) to solve the

Table 5
Impact of CutMix, CRA and FPN on COVIDx image classification for ResNet-50+MCRA(Ours).

Method	CutMix	CRA	FPN	Accuracy (%)	PPV (%)	Sensitivity (%)	Specificity (%)	F1 (%)
ResNet-50				96.50	94.84	95.38	98.42	95.10
	✓			96.87	95.65	95.50	98.54	95.56
	✓	✓		97.25	96.00	96.06	98.77	96.03
	✓	✓	✓	**97.71**	**96.76**	**96.56**	**98.96**	**96.64**



(a) (b) (c) (d)

Fig. 9. Examples of heat maps for COVID-19 diagnosis. (a) Original images. (b) Heat map of ResNet. (c) Heat map of ResNet + CRA. (d) Heat map of ResNet + MCRA.

problem of low detection or classification accuracy of COVID-19 caused by small data samples. Numerous experimental results show that the image classification of CXR by using CRA and FPN is effective. Our method reaches the highest accuracy of COVID-19 CXR image classification, whose *accuracy*, *PPV*, *sensitivity*, *specificity* and *F*1-Score are 97.71%, 96.76%, 96.56%, 98.96% and 96.64%, respectively. More importantly, our deep model is of a little interpretability because its visible and feasible working procedure or results. However, only three categories of chest images are recognized: COVID-19, Pneumonia, and Normal. As far as COVID-19 in concerned, once its clinical shape and characteristics have changed, more training data is required. Therefore, we are going to explore more varieties of COVID-19 diagnosis in future.

Credit authorship contribution statement

Shangwang Liu: Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Software, Funding acquisition. Tongbo Cai: Writing – original draft, Writing – review & editing,

Conceptualization, Methodology, Software. Xiufang Tang: Supervision, Methodology, Funding acquisition, Writing – original draft, Writing – review & editing. Yangyang Zhang: Supervision, Methodology, Writing – original draft, Writing – review & editing. Changgeng Wang: Methodology, Software, Conceptualization, Writing – original & draft, Writing – review & editing.

Funding

This research was supported by the Key Scientific Research Project of Higher School of Henan Province, grant number 21A520022.

Declaration of competing interest

None Declared.

Acknowledgement

Not applicable.

References

[1] P. Aggarwal, N.K. Mishra, B. Fatimah, P. Singh, A. Gupta, S.D. Joshi, COVID-19 image classification using deep learning: advances, challenges and opportunities, Comput. Biol. Med. 144 (2022), 105350, https://doi.org/10.1016/j.compbiomed.2022.105350.

[2] M. Cascella, M. Rajnik, A. Aleem, S.C. Dulebohn, R. Di Napoli, Features, Evaluation, and Treatment of Coronavirus (COVID-19), 2022. Statpearls [Internet].

[3] B. Yang, W. Bao, J. Wang, Active disease-related compound identification based on capsule network, Briefings Bioinf. 23 (2022) bbab462.

[4] B. Hu, H. Guo, P. Zhou, Z.-L. Shi, Characteristics of SARS-CoV-2 and COVID-19, Nat. Rev. Microbiol. 19 (2021) 141–154.

[5] World Health Organization, Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: interim guidance, in: Clinical Management of Severe Acute Respiratory Infection when Novel Coronavirus (2019-NCoV) Infection Is Suspected, Interim Guidance, 2020, p. 21, 21.

[6] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, R.-G. Zhang, M.-M. Cheng, JCS: an explainable COVID-19 diagnosis system by joint classification and segmentation, IEEE Trans. Image Process. 30 (2021) 3113–3126, https://doi.org/10.1109/TIP.2021.3058783.

[7] J. Gu, L. Yang, T. Li, Y. Liu, J. Zhang, K. Ning, D. Su, Temporal relationship between serial RT-PCR results and serial chest CT imaging, and serial CT changes in coronavirus 2019 (COVID-19) pneumonia: a descriptive study of 155 cases in China, Eur. Radiol. 31 (2021) 1175–1184, https://doi.org/10.1007/s00330-020-07268-9.

[8] S. Dong, Q. Yang, Y. Fu, M. Tian, C. Zhuo, RCoNet: deformable mutual information maximization and high-order uncertainty-aware learning for robust COVID-19 detection, IEEE Transact. Neural Networks Learn. Syst. 32 (2021) 3401–3411, https://doi.org/10.1109/TNNLS.2021.3086570.

[9] M. Chen, X. Shi, Y. Zhang, D. Wu, M. Guizani, Deep feature learning for medical image analysis with convolutional autoencoder neural network, IEEE Trans. Big Data 7 (2021) 750–758, https://doi.org/10.1109/TBDATA.2017.2717439.

[10] N. Feng, X. Geng, B. Sun, Study on neural network integration method based on morphological associative memory framework, Neural Process. Lett. 53 (2021) 1–31, https://doi.org/10.1007/s11063-021-10569-9.

[11] D. Xintao, D. Guo, C. Qin, Image information hiding method based on image compression and deep neural network, Comput. Model. Eng. Sci. 124 (2020) 1–17, https://doi.org/10.32604/cmes.2020.09463.

[12] J. Zhou, Real-time task scheduling and network device security for complex embedded systems based on deep learning networks, Microprocess. Microsyst. 79 (2020), 103282, https://doi.org/10.1016/j.micpro.2020.103282.

[13] L. Chen, J. Li, M. Chang, Cancer diagnosis and disease gene identification via statistical machine learning, Curr. Bioinf. 15 (2020) 956–962.

[14] X.-F. Wang, P. Gao, Y.-F. Liu, H.-F. Li, F. Lu, Predicting thermophilic proteins by machine learning, Curr. Bioinf. 15 (2020), https://doi.org/10.2174/1574893615666200207094357.

[15] J. Chen, K. Liao, Y. Wan, D.Z. Chen, J. Wu, Danets: deep abstract networks for tabular data classification and regression, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 3930–3938.

[16] R. Riad, O. Teboul, D. Grangier, N. Zeghidour, Learning strides in convolutional neural networks, in: International Conference on Learning Representations, 2022. https://openreview.net/forum?id=M752z9FKJP.

[17] S. Zhang, Z. Yu, L. Liu, X. Wang, A. Zhou, K. Chen, Group R-CNN for weakly semi-supervised object detection with points, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9417–9426.

[18] P. Zhang, Z. Kang, T. Yang, X. Zhang, N. Zheng, J. Sun, LGD: label-guided self-distillation for object detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 3309–3317.

[19] W. Liao, Progressive minimal path method with embedded CNN, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 4514–4522.

[20] S. Kundu, H. Elhalawani, J.W. Gichoya, C.E. Kahn Jr., How might AI and chest imaging help unravel COVID-19's mysteries? Radiology: Artif. Intell. 2 (2020), e200053.

[21] N.A. Baghdadi, A. Malki, S.F. Abdelaliem, H. Magdy Balaha, M. Badawy, M. Elhosseini, An automated diagnosis and classification of COVID-19 from chest CT images using a transfer learning-based convolutional neural network, Comput. Biol. Med. 144 (2022), 105383, https://doi.org/10.1016/j.compbiomed.2022.105383.

[22] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, C. Zhang, Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning, IEEE Trans. Med. Imag. 39 (2020) 2606–2614.

[23] X. Ouyang, J. Huo, L. Xia, F. Shan, J. Liu, Z. Mo, F. Yan, Z. Ding, Q. Yang, B. Song, Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia, IEEE Trans. Med. Imag. 39 (2020) 2595–2605.

[24] A. Abbas, M.M. Abdelsamea, M.M. Gaber, 4S-DT: self-supervised super sample decomposition for transfer learning with application to COVID-19 detection, IEEE Transact. Neural Networks Learn. Syst. 32 (2021) 2798–2808.

[25] E. Jangam, A.A.D. Barreto, C.S.R. Annavarapu, Automatic detection of COVID-19 from chest CT scan and chest X-Rays images using deep learning, transfer learning and stacking, Appl. Intell. 52 (2022) 2243–2259, https://doi.org/10.1007/s10489-021-02393-4.

[26] I. De Falco, G. De Pietro, G. Sannino, Classification of Covid-19 chest X-ray images by means of an interpretable evolutionary rule-based approach, Neural Comput. Appl. (2022) 1–11.

[27] A. Konwer, X. Xu, J. Bae, C. Chen, P. Prasanna, Temporal context matters: enhancing single image prediction with disease progression representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 18824–18835.

[28] G. Wang, S. Zhai, G. Lasio, B. Zhang, B. Yi, S. Chen, T.J. Macvittie, D. Metaxas, J. Zhou, S. Zhang, Semi-Supervised segmentation of radiation-induced pulmonary fibrosis from lung CT scans with multi-scale guided dense attention, IEEE Trans. Med. Imag. 41 (2022) 531–542, https://doi.org/10.1109/TMI.2021.3117564.

[29] Y. Zhang, R. Higashita, H. Fu, Y. Xu, Y. Zhang, H. Liu, J. Zhang, J. Liu, A multi-branch hybrid transformer network for corneal endothelial cell segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 99–108.

[30] Y. Tang, R. Gao, H. Lee, Q. Yang, X. Yu, Y. Zhou, S. Bao, Y. Huo, J. Spraggins, J. Virostko, Pancreas CT segmentation by predictive phenotyping, Others, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 25–35.

[31] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, Attention U-Net: Learning where to Look for the Pancreas, Others, 2018. ArXiv Preprint ArXiv:1804.03999.

[32] N.K. Tomar, D. Jha, U. Bagci, S. Ali, TGANet: Text-Guided Attention for Improved Polyp Segmentation, 2022. ArXiv Preprint ArXiv:2205.04280.

[33] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Pranet: parallel reverse attention network for polyp segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 263–273.

[34] A. Elazab, M.A. Elfattah, Y. Zhang, Novel multi-site graph convolutional network with supervision mechanism for COVID-19 diagnosis from X-ray radiographs, Appl. Soft Comput. 114 (2022), 108041, https://doi.org/10.1016/j.asoc.2021.108041.

[35] A.M. Tahir, M.E.H. Chowdhury, A. Khandakar, T. Rahman, Y. Qiblawey, U. Khurshid, S. Kiranyaz, N. Ibtehaz, M.S. Rahman, S. Al-Maadeed, S. Mahmud, M. Ezeddin, K. Hameed, T. Hamid, COVID-19 infection localization and severity grading from chest X-ray images, Comput. Biol. Med. 139 (2021), 105002, https://doi.org/10.1016/j.compbiomed.2021.105002.

[36] Y. Tang, Y. Tang, Y. Zhu, J. Xiao, R.M. Summers, A disentangled generative model for disease decomposition in chest X-rays via normal image synthesis, Med. Image Anal. 67 (2021), 101839, https://doi.org/10.1016/j.media.2020.101839.

[37] J. Liu, H. Shao, Y. Jiang, X. Deng, CNN-based hidden-layer topological structure design and optimization methods for image classification, Neural Process. Lett. (2022), https://doi.org/10.1007/s11063-022-10742-8.

[38] Y. Xu, H.-K. Lam, G. Jia, MANet: a two-stage deep learning method for classification of COVID-19 from Chest X-ray images, Neurocomputing 443 (2021) 96–105.

[39] N.S. Shaik, T.K. Cherukuri, Transfer learning based novel ensemble classifier for COVID-19 detection from chest CT-scans, Comput. Biol. Med. 141 (2022), 105127, https://doi.org/10.1016/j.compbiomed.2021.105127.

[40] C.M. Dasari, R. Bhukya, Explainable deep neural networks for novel viral genome prediction, Appl. Intell. 52 (2022) 3002–3017, https://doi.org/10.1007/s10489-021-02572-3.

[41] H. Wang, Y. Li, N. He, K. Ma, D. Meng, Y. Zheng, DICDNet: deep interpretable convolutional dictionary network for metal artifact reduction in CT images, IEEE Trans. Med. Imag. 41 (2022) 869–880, https://doi.org/10.1109/TMI.2021.3127074.

[42] I. De Falco, G. De Pietro, G. Sannino, Classification of Covid-19 Chest X-Ray Images by Means of an Interpretable Evolutionary Rule-Based Approach, Neural

Computing and Applications, 2022, https://doi.org/10.1007/s00521-021-06806-w.
[43] M.-L. Huang, Y.-C. Liao, A lightweight CNN-based network on COVID-19 detection using X-ray and CT images, Comput. Biol. Med. 146 (2022), 105604, https://doi.org/10.1016/j.compbiomed.2022.105604.
[44] A. Kumar, A.R. Tripathi, S.C. Satapathy, Y.-D. Zhang, SARS-Net: COVID-19 detection from chest x-rays by combining graph convolutional network and convolutional neural network, Pattern Recogn. 122 (2022), 108255.
[45] S. Tang, F. Yu, Construction and verification of retinal vessel segmentation algorithm for color fundus image under BP neural network model, J. Supercomput. 77 (2021) 3870–3884.
[46] M.F. Aslan, K. Sabanci, A. Durdu, M.F. Unlersen, COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization, Comput. Biol. Med. 142 (2022), 105244, https://doi.org/10.1016/j.compbiomed.2022.105244.
[47] S. Tang, C. Wang, J. Nie, N. Kumar, Y. Zhang, Z. Xiong, A. Barnawi, EDL-COVID: ensemble deep learning for COVID-19 case detection from chest x-ray images, IEEE Trans. Ind. Inf. 17 (2021) 6539–6549.
[48] P. Bhowal, S. Sen, J.H. Yoon, Z.W. Geem, R. Sarkar, Choquet integral and coalition game-based ensemble of deep learning models for COVID-19 screening from chest X-ray images, IEEE J. Biomed. Health Inf. 25 (2021) 4328–4339.
[49] O. Attallah, ECG-BiCoNet: An ECG-based pipeline for COVID-19 diagnosis using Bi-Layers of deep features integration, Comput. Biol. Med. 142 (2022), 105210, https://doi.org/10.1016/j.compbiomed.2022.105210.
[50] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning, Med. Image Anal. 65 (2020), 101794.
[51] H. Chen, Y. Jiang, M. Loew, H. Ko, Unsupervised domain adaptation based COVID-19 CT infection segmentation network, Appl. Intell. 52 (2022) 6340–6353, https://doi.org/10.1007/s10489-021-02691-x.
[52] N. Sobahi, A. Sengur, R.-S. Tan, U.R. Acharya, Attention-based 3D CNN with residual connections for efficient ECG-based COVID-19 detection, Comput. Biol. Med. 143 (2022), 105335, https://doi.org/10.1016/j.compbiomed.2022.105335.
[53] Y. Oh, S. Park, J.C. Ye, Deep learning COVID-19 features on CXR using limited training data sets, IEEE Trans. Med. Imag. 39 (2020) 2688–2700.
[54] L. Wang, Z.Q. Lin, A. Wong, Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, Sci. Rep. 10 (2020) 1–12.
[55] B. Lu, Algorithm improvement of neural network in endoscopic image recognition of upper digestive tract system, Expet Syst. (2021), e12912.
[56] H. Su, D. Zhao, H. Elmannai, A.A. Heidari, S. Bourouis, Z. Wu, Z. Cai, W. Gui, M. Chen, Multilevel threshold image segmentation for COVID-19 chest radiography: a framework using horizontal and vertical multiverse optimization, Comput. Biol. Med. (2022), 105618.
[57] C. Ieracitano, N. Mammone, M. Versaci, G. Varone, A.-R. Ali, A. Armentano, G. Calabrese, A. Ferrarelli, L. Turano, C. Tebala, A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images, Neurocomputing (2022).
[58] B. He, W. Hu, K. Zhang, S. Yuan, X. Han, C. Su, J. Zhao, G. Wang, G. Wang, L. Zhang, Image segmentation algorithm of lung cancer based on neural network model, Expet Syst. 39 (2022), e12822.
[59] R. Singh, V. Bharti, V. Purohit, A. Kumar, A.K. Singh, S.K. Singh, MetaMed: few-shot medical image classification using gradient-based meta-learning, Pattern Recogn. 120 (2021), 108111.
[60] Z. Huang, J. Zhang, Y. Zhang, H. Shan, DU-GAN: Generative adversarial networks with dual-domain U-Net-Based discriminators for low-dose CT denoising, IEEE Trans. Instrum. Meas. 71 (2022) 1–12, https://doi.org/10.1109/TIM.2021.3128703.
[61] Y. Shen, S. Zhu, T. Yang, C. Chen, D. Pan, J. Chen, L. Xiao, Q. Du, BDANet: multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images, IEEE Trans. Geosci. Rem. Sens. 60 (2022) 1–14, https://doi.org/10.1109/TGRS.2021.3080580.
[62] S.K. Yadav, S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H.M. Pandey, M. Mathur, CSITime: privacy-preserving human activity recognition using WiFi channel state information, Neural Network. 146 (2022) 11–21.
[63] X. Xia, X. Chai, N. Zhang, T. Sun, Visual classification of apple bud-types via attention-guided data enrichment network, Comput. Electron. Agric. 191 (2021), 106504.
[64] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: beyond Empirical Risk Minimization, 2017. ArXiv Preprint ArXiv:1710.09412.
[65] T. DeVries, G.W. Taylor, Improved Regularization of Convolutional Neural Networks with Cutout, 2017. ArXiv Preprint ArXiv:1708.04552.
[66] S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: Regularization Strategy to Train Strong Classifiers with Localizable Features, 2019, pp. 6023–6032.
[67] J. Li, D. Wang, X. Liu, Z. Shi, M. Wang, Two-Branch attention network via efficient semantic coupling for one-shot learning, IEEE Trans. Image Process. 31 (2022) 341–351, https://doi.org/10.1109/TIP.2021.3124668.
[68] S.-B. Chen, Q.-S. Wei, W.-Z. Wang, J. Tang, B. Luo, Z.-Y. Wang, Remote sensing scene classification via multi-branch local attention network, IEEE Trans. Image Process. 31 (2022) 99–109, https://doi.org/10.1109/TIP.2021.3127851.
[69] S. Zhu, B. Du, L. Zhang, X. Li, Attention-based multiscale residual adaptation network for cross-scene classification, IEEE Trans. Geosci. Rem. Sens. 60 (2022) 1–15, https://doi.org/10.1109/TGRS.2021.3056624.
[70] F. Yang, H. Zhang, S. Tao, Semi-supervised classification via full-graph attention neural networks, Neurocomputing 476 (2022) 63–74, https://doi.org/10.1016/j.neucom.2021.12.077.
[71] Y. Dong, Q. Liu, B. Du, L. Zhang, Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification, IEEE Trans. Image Process. 31 (2022) 1559–1572, https://doi.org/10.1109/TIP.2022.3144017.
[72] Y. Liu, J. Zhou, L. Liu, Z. Zhan, Y. Hu, Y.Q. Fu, H. Duan, FCP-net: a feature-compression-pyramid network guided by game-theoretic interactions for medical image segmentation, IEEE Trans. Med. Imag. (2022) 1, https://doi.org/10.1109/TMI.2021.3140120, 1.
[73] S. Wang, Y. Zhu, S. Lee, D.C. Elton, T.C. Shen, Y. Tang, Y. Peng, Z. Lu, R. M. Summers, Global-Local attention network with multi-task uncertainty loss for abnormal lymph node detection in MR images, Med. Image Anal. 77 (2022), 102345, https://doi.org/10.1016/j.media.2021.102345.
[74] X. Wang, L. Zhu, S. Tang, H. Fu, P. Li, F. Wu, Y. Yang, Y. Zhuang, Boosting RGB-D saliency detection by leveraging unlabeled RGB images, IEEE Trans. Image Process. 31 (2022) 1107–1119, https://doi.org/10.1109/TIP.2021.3139232.
[75] W. Ji, G. Yan, J. Li, Y. Piao, S. Yao, M. Zhang, L. Cheng, H. Lu, DMRA: depth-induced multi-scale recurrent attention network for RGB-D saliency detection, IEEE Trans. Image Process. 31 (2022) 2321–2336, https://doi.org/10.1109/TIP.2022.3154931.
[76] C. Shang, H. Li, F. Meng, H. Qiu, Q. Wu, L. Xu, K.N. Ngan, Instance-level context attention network for instance segmentation, Neurocomputing 472 (2022) 124–137, https://doi.org/10.1016/j.neucom.2021.11.104.
[77] G. Tao, H. Li, J. Huang, C. Han, J. Chen, G. Ruan, W. Huang, Y. Hu, T. Dan, B. Zhang, S. He, L. Liu, H. Cai, SeqSeg: a sequential method to achieve nasopharyngeal carcinoma segmentation free from background dominance, Med. Image Anal. 78 (2022), 102381, https://doi.org/10.1016/j.media.2022.102381.
[78] X. Lu, W. Wang, J. Shen, D. Crandall, J. Luo, Zero-Shot video object segmentation with Co-attention siamese networks, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2022) 2228–2242, https://doi.org/10.1109/TPAMI.2020.3040258.
[79] Y. Liu, Y. Chen, P. Lasang, Q. Sun, Covariance attention for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2022) 1805–1818, https://doi.org/10.1109/TPAMI.2020.3026069.
[80] J. Li, D. Zhang, Q. Liu, R. Bu, Q. Wei, COVID-GATNet: a Deep Learning Framework for Screening of COVID-19 from Chest X-Ray Images, IEEE, 2020, pp. 1897–1902.
[81] Y. Feng, X. Yang, Q. Dawei, H. Zhang, D. Wei, L. Jing, PCXRNet: pneumonia diagnosis from Chest X-Ray Images using Condense attention block and Multiconvolution attention block, IEEE J. Biomed. Health Inf. (2022).
[82] Z. Lin, Z. He, S. Xie, X. Wang, J. Tan, J. Lu, B. Tan, AANet: adaptive attention network for COVID-19 detection from chest X-ray images, IEEE Transact. Neural Networks Learn. Syst. 32 (2021) 4781–4792.
[83] Z. Wu, H. Zhu, G. Li, Z. Cui, H. Huang, J. Li, E. Chen, G. Xu, An efficient Wikipedia semantic matching approach to text document classification, Inf. Sci. 393 (2017) 15–28.
[84] Z. Wu, G. Li, S. Shen, X. Lian, E. Chen, G. Xu, Constructing dummy query sequences to protect location privacy and query privacy in location-based services, World Wide Web 24 (2021) 25–49.
[85] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature Pyramid Networks for Object Detection, 2017, pp. 2117–2125.
[86] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional Networks for Biomedical Image Segmentation, Springer, 2015, pp. 234–241.
[87] S. Wang, Y. Cong, H. Zhu, X. Chen, L. Qu, H. Fan, Q. Zhang, M. Liu, Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract, IEEE J. Biomed. Health Inf. 25 (2020) 514–525.
[88] B. Bai, G. Li, S. Wang, Z. Wu, W. Yan, Time series classification based on multi-feature dictionary representation and ensemble learning, Expert Syst. Appl. 169 (2021), 114162.
[89] W. Yan, G. Li, Z. Wu, S. Wang, P.S. Yu, Extracting diverse-shapelets for early classification on time series, World Wide Web 23 (2020) 3055–3081.
[90] N. Muralidharan, S. Gupta, M.R. Prusty, R.K. Tripathy, Detection of COVID19 from X-ray images using multiscale deep convolutional neural network, Appl. Soft Comput. 119 (2022), 108610.
[91] L. He, P. Tiwari, R. Su, X. Shi, P. Marttinen, N. Kumar, COVIDNet: an Automatic Architecture for COVID-19 Detection with Deep Learning from Chest X-Ray Images, IEEE Internet of Things Journal, 2021.
[92] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[93] P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, Y. Wei, Layercam: exploring hierarchical class activation maps for localization, IEEE Trans. Image Process. 30 (2021) 5875–5888.
[94] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual Explanations from Deep Networks via Gradient-Based Localization, 2017, pp. 618–626.
[95] A. Chattopadhyay, A. Sarkar, P. Howlader, V. Balasubramanian, Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks, arXiv, 2017. ArXiv Preprint ArXiv:1710.11063. (n.d.).
[96] S.G. Müller, F. Hutter, Trivialaugment: Tuning-free yet State-Of-The-Art Data Augmentation, 2021, pp. 774–782.
[97] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, J. Han, On the Variance of the Adaptive Learning Rate and beyond, 2019. ArXiv Preprint ArXiv:1908.03265.
[98] M. Zhang, J. Lucas, J. Ba, G.E. Hinton, Lookahead optimizer: k steps forward, 1 step back, Adv. Neural Inf. Process. Syst. 32 (2019).
[99] H. Yong, J. Huang, X. Hua, L. Zhang, Gradient Centralization: A New Optimization Technique for Deep Neural Networks, Springer, 2020, pp. 635–652.

[100] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2020. ArXiv Preprint ArXiv: 2010.11929.
[101] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows, 2021, pp. 10012–10022.
[102] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014. ArXiv Preprint ArXiv:1409.1556.
[103] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[104] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
[105] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[106] I. Radosavovic, R.P. Kosaraju, R. Girshick, K. He, P. Dollár, Designing network design spaces, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10428–10436.
[107] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
[108] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, Y. Wang, Transformer in transformer, Adv. Neural Inf. Process. Syst. 34 (2021) 15908–15919.
[109] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, Training data-efficient image transformers & distillation through attention, in: International Conference on Machine Learning, PMLR, 2021, pp. 10347–10357.
[110] L. Yuan, Q. Hou, Z. Jiang, J. Feng, S. Yan, Volo: Vision Outlooker for Visual Recognition, 2021. ArXiv Preprint ArXiv:2106.13112.
[111] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.
[112] D. Ng, Y. Chen, B. Tian, Q. Fu, E.S. Chng, ConvMixer: feature interactive convolution with curriculum learning for small footprint and noisy far-field keyword spotting, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 3603–3607.
[113] X. Dong, J. Bao, D. Chen, W. Zhang, N. Yu, L. Yuan, D. Chen, B. Guo, Cswin transformer: a general vision transformer backbone with cross-shaped windows, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12124–12134.
[114] W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, S. Yan, Metaformer is actually what you need for vision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10819–10829.