



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A depthwise separable dense convolutional network with convolution block attention module for COVID-19 diagnosis on CT scans

Qian Li^{a,b,*}, Jiangbo Ning^a, Jianping Yuan^a, Ling Xiao^a

^a Research Center for Ultrasonics and Technologies, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China

^b University of Chinese Academy of Sciences, Beijing, 100049, China

ARTICLE INFO

Keywords:

COVID-19 diagnosis
CT scan
Dataset light-weighted
DenseNet
Convolution block attention module
Depthwise separable convolution

ABSTRACT

Coronavirus disease 2019 (COVID-19) has caused more than 3 million deaths and infected more than 170 million individuals all over the world. Rapid identification of patients with COVID-19 is the key to control transmission and prevent depletion of hospitals. Several networks have been proposed to assist radiologists in diagnosing COVID-19 based on CT scans. However, CTs used in these studies are unavailable for other researchers to do deeper extensions due to privacy concerns. Furthermore, these networks are too heavy-weighted to satisfy the general trend applying on a computationally limited platform. In this paper, we aim to solve these two problems. Firstly, we establish an available dataset COVID-CTx, which contains 828 CT scans positive for COVID-19 across 324 patient cases from three open access data repositories. To our knowledge, it has the largest number of publicly available COVID-19 positive cases compared to other public datasets. Secondly, we propose a light-weighted hybrid neural network: Depthwise Separable Dense Convolutional Network with Convolution Block Attention Module (AM-SdenseNet). AM-SdenseNet synergistically integrates Convolutional Block Attention Module with depthwise separable convolutions to learn powerful feature representations while reducing the parameters to overcome the overfitting problem. Through experiments, we demonstrate the superior performance of our proposed AM-SdenseNet compared with several state-of-the-art baselines. The excellent performance of AM-SdenseNet can improve the speed and accuracy of COVID-19 diagnosis, which is extremely useful to control the spreading of infection.

1. Introduction

Coronavirus disease 2019 (COVID-19), named SARS-CoV-2 by the International Committee on Taxonomy of Viruses (ICTV), is a highly infectious respiratory disease. More than 170 million confirmed COVID-19 cases and 3 million deaths have been reported in roughly 200 different countries and territories, as of June 7 in 2021. Rapid identification of patients with COVID-19 has been recommended by World Health Organization (WHO) to control transmission and prevent depletion of hospitals. Reverse transcription-polymerase chain reaction (RT-PCR) serves as the gold standard for COVID-19 diagnosis. However, the total positive rate of RT-PCR for throat swab samples was reported to be approximately 30%–60% at the initial presentation [1]. Additionally, it takes 4–6 h to provide results, which is much slower than the speeding of the COVID-19. As a result, some infected patients cannot be identified early and continue to infect others unintentionally.

To mitigate the inefficiency of RT-PCR, chest computed tomography

(CT) has been used to supplement RT-PCR testing of patients with suspected COVID-19 [2]. Several studies [3,4] have shown that CT scan manifests clear radiological findings of COVID-19 cases such as ground-glass opacities and consolidations. And CT scans are promised as a more efficient testing tool in serving for COVID-19 diagnosis with higher sensitivity [5]. However, chest CT contains hundreds of slices, which takes a long time for specialists to diagnose. Time pressure, heavy workload, and lack of experienced radiologists result in challenges in the imaging-based analysis of COVID-19 [6]. Fortunately, Artificial Intelligence (AI) has been widely used in the field of medical images, such as lung nodule [7], tuberculosis [8], breast cancer [9], tumor [10], etc. Rapidly developed AI-based automated CT image analysis tools can achieve high accuracy in the detection of Coronavirus positive patients. To alleviate the burden of medical professionals, there have been increasing efforts on developing computer-aided detection (CAD) systems to assist radiologists in diagnosing COVID-19 based on CT scans [11–20].

* Corresponding author. Research Center for Ultrasonics and Technologies, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China.

E-mail addresses: liqian@mail.ioa.ac.cn (Q. Li), ningjiangbo@mail.ioa.ac.cn (J. Ning), yuanjianping@mail.ioa.ac.cn (J. Yuan), xiaoling@mail.ioa.ac.cn (L. Xiao).

The CAD systems for the detection of COVID-19 are mainly divided into two modules: segmentation and classification. Segmentation aims to segment lung regions where lesions are located, then the segmented regions could be further fed to a classification module for COVID-19 detection. Segmentation is an essential step in image processing and analysis for the assessment and quantification of COVID-19. It delineates the regions of interest (ROIs), e.g., lung and infected regions, in the CT images. To segment ROIs in CT, U-Net [21], U-Net++ [22], and V-Net [23] are widely used. In the classification module, there are several studies aiming to separate COVID-19 patients from Normal subjects. ResNet [24] is the most widely used network for diagnosing COVID-19 from CT scans [14–18]. Besides being based on a ResNet, some networks further combine with the attention mechanism and feature pyramid network (FPN) [25] to focus on important features [14,16]. Given the problem of limited datasets, studies adopt strategies such as weak label [11], transfer learning [19], human-in-the-loop [13], and data augmentation [20] to improve the evaluation indicators of the CAD systems. Although these studies have achieved good results, there are still two major hurdles: (1) CT scans used in these studies are unavailable for public access and use due to privacy concerns. Consequently, these works are difficult to copy and adapt, which greatly hinders the research and development of deep learning methods; (2) These networks are too heavy-weighted to satisfy the general trend applying on a computationally limited platform.

To address the first problem, we establish a publicly available dataset COVID-CTx. It is composed and modified by three open access data repositories: COVID-SIRM [26], COVID-Seg [27], and COVID19-CT [19]. To the best of our knowledge, COVID-CTx has the largest number of publicly available COVID-19 positive cases compared to other public datasets. All CT scans on these three datasets are confirmed by senior radiologists. For the second, we propose a light-weighted hybrid neural network: Depthwise Separable Dense Convolutional Network with Convolutional Block Attention Module (AM-SdenseNet). AM-SdenseNet uses a Dense Convolutional Network (DenseNet) [28] as the basic network. Compared with ResNet, DenseNet can better satisfy the general trend of application on a limited platform. At the same time, AM-SdenseNet synergistically applies Convolutional Block Attention Module (CBAM) [29] on DenseNet in each block with residual learning. This operation can greatly focus on important features and suppress unnecessary ones. In addition, we replace some regular convolutions in DenseNet with depthwise separable convolutions [30] to reduce training parameters while keeping network performance.

The paper is organized as follows. Section 2 introduces related works regarding existing accessible COVID-19 datasets, COVID-19 diagnosis systems, depthwise separable convolution, and attention module. Section 3 presents the COVID-CTx dataset, data preprocessing, and AM-SdenseNet architecture. In Section 4, we compare the performance of AM-SdenseNet with six state-of-the-art networks: MobileNet [31], InceptionV3 [32], ResNet50, VGG16 [33], DenseNet169, and DenseNet121. Finally, Section 5 concludes the paper and discusses future directions.

2. Related works

2.1. Existing accessible COVID-19 datasets

Due to privacy issues, few data sets with massive CT scans on COVID-19 are available for public access and use until now. Existing larger data sets on COVID-19 are COVID-SIRM, COVID-Seg, and COVID19-CT. The Italian Society of Medical Radiology (SIRM) has publicly provided 100 CT scans across 68 patient cases [26]. Each case in the COVID-SIRM provides detailed information of the lesions in the CT images. In addition, some cases are followed up and treated at intervals of about 4 days. COVID-Seg [27] contains 40 labeled COVID-19 CT scans. Left lung, right lung, and infections are labeled by two radiologists and verified by an experienced radiologist. The University of California constructs a

publicly available COVID19-CT data set [19], which contains 349 CT scans that are positive for COVID-19 from 216 COVID-19 cases. All CTs were collected from public websites such as medRxiv, bioRxiv, journals, or papers related to COVID-19. Other existing data sets on COVID-19 are mainly X-ray images. For example, COVID-19 Radiography Database, generated by Chowdhury et al. [20], is comprised of 1200 COVID-19 positive images and 1341 normal images. It is worth noting that the current public data sets still have a very limited number of images for the training and testing of models, and the quality of data sets is insufficient.

2.2. COVID-19 diagnosis systems

During the outbreak of COVID-19, there have been increasing efforts on COVID diagnosis systems to perform screening COVID-19 based on CT scans. Segmentation is an essential step in COVID-19 diagnosis systems to assess and quantify COVID-19. The popular segmentation networks for CT scans in COVID-19 applications include U-Net, U-Net++, and V-Net. The U-Net, a U-shape network, is a popular technique for segmenting lung regions in medical images. For example, Zheng et al. [11] obtain all CT lung masks through a pre-trained U-Net in COVID-19 applications. Various U-Net, meanwhile, has been developed, reaching better segmentation results in COVID-19 image segmentation. Jin et al. [18] propose a two-stage pipeline for diagnosing COVID-19 in CT images, in which the whole lung is first detected by an efficient network based on U-Net++. The U-Net++ can greatly improve the performance of segmentation, as the network inserts a nested convolutional structure between the encoding and decoding path based on U-Net. In addition, CT provides high-quality 3D images for detecting COVID-19. Significantly, V-Net is a 3D image segmentation approach, where volumetric convolutions were applied instead of processing the input volumes slice-wise. To fully utilize the spatial information of CT, Shan et al. [13] propose a 3D segmentation system VB-Net, which combines V-Net with the bottle-neck structure [24]. Obviously, all of these networks can achieve better performance of segmentation. However, their training is difficult without adequate labeled data. In COVID-19 CT segmentation, sufficient labeled data for segmentation tasks is often unavailable since manual delineation for lung regions is labor-intensive and time-consuming.

Segmentation could be used to preprocess the CTs, and classification takes advantage of those segmentation results in the diagnosis. There are many studies aiming to separate COVID-19 patients from non-COVID-19 subjects. For example, Jin et al. [15] establish a deep learning system for COVID-19 detection, which outperforms five radiologists in more challenging tasks at a speed of two orders of magnitude above them. It fully proves that CAD systems can assist radiologists to accelerate the speed of diagnosing COVID-19. ResNet is the most widely used network in CAD systems. For example, Xu et al. [14] propose a diagnostic model, based on a ResNet18, to classify normal, viral, and COVID-19 from CT scans. Furthermore, they add a location-attention mechanism to improve the overall accuracy, which reaches 86.7%. Li et al. [34] preprocess the 2D slices to extract lung regions using U-Net, and a ResNet50 model combined with max-pooling for diagnosis. The model achieves results with a specificity of 96%, a sensitivity of 90%, and an AUC of 0.96 in identifying COVID-19. Similarly, Jin et al. [18] propose a U-Net++ based segmentation model for locating lesions and a ResNet50 based classification model for diagnosis. The specificity and sensitivity using the proposed U-Net++ and ResNet50 combined model are 92.2% and 97.4%.

Given the problem of limited datasets, He et al. [19] propose a self-supervision transfer learning method (Self-Trans). The network integrates contrastive self-supervised learning with transfer learning to learn powerful and unbiased feature representations for reducing over-fitting. The testing dataset shows an AUC of 0.94, even though the number of training CT scans is just a few hundred. Zheng et al. [11] propose a weakly-supervised network: 3D deep CNN (DeCoVNet), which can accurately predict the COVID-19 infectious probability in chest CT

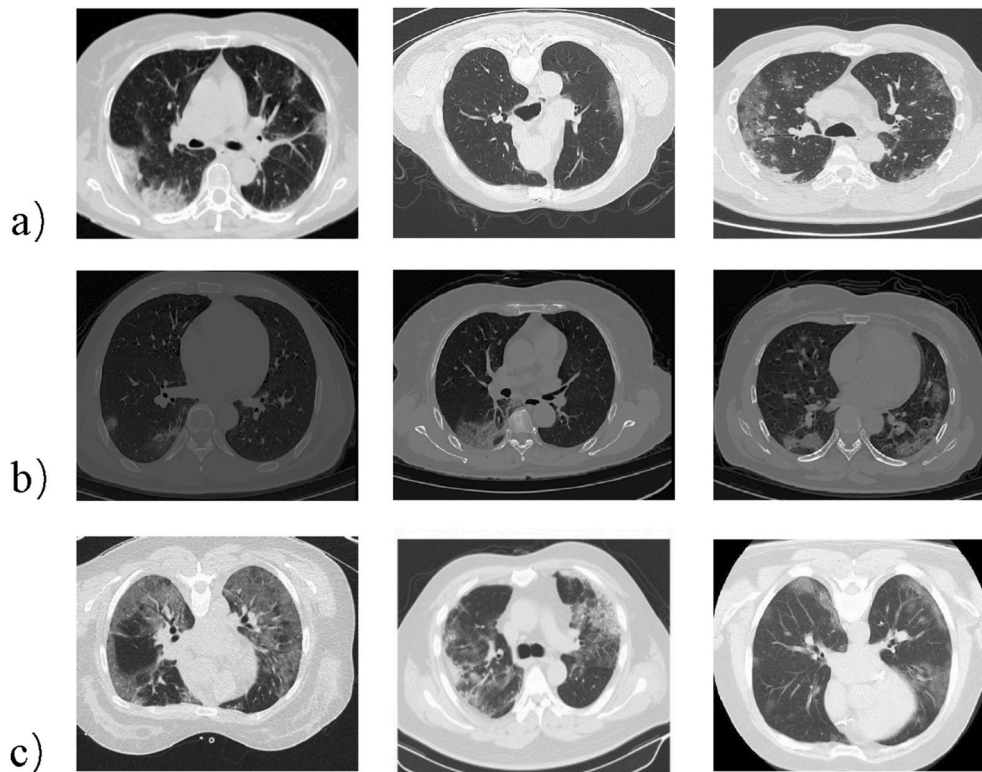


Fig. 1. Example CT images from the COVID-CTx dataset, which comprises 828 images across 324 patient cases from three open access data repositories: a). COVID-SIRM, b). COVID-Seg, c). COVID19-CT.

volumes without labeled lesions for training. DeepPneumonia, proposed by Song et al. [16], uses a ResNet50 backbone with the FPN and attention module as the feature-extracting part. The network can distinguish the COVID-19 patients from others with an excellent AUC of 99% and sensitivity of 93%. Additionally, those researches [11,19,20] all apply data augmentation to overcome the overfitting problem.

In summary, the above approaches have been proposed for CT-based COVID-19 diagnosis with generally promising results. But most of the mentioned COVID-19 classification techniques were training on large datasets. CT scans used in these studies are unavailable for public access and use, which greatly hinders the research and development of deep learning methods. In addition, these networks are too heavy-weighted to satisfy the general trend applying on a computationally limited platform.

2.3. Depthwise separable convolution

To achieve higher accuracy, the network has become much deeper and more complicated [33]. However, the general trend has been to achieve the recognition tasks in a timely fashion on a computationally limited platform [31]. Depthwise separable convolutions, firstly proposed by Sifre et al. [30], aim to reduce training parameters. The operation is to split the corresponding area and channel of the processed image. Later, Inception V1 and Inception V2 use a depthwise separable convolution as the first layer to lighten the network [35,36]. Howard et al. [31] introduce MobileNets based on depthwise separable convolutions. This network is mostly used in mobile and embedded vision applications, largely due to its lightness. Jin et al. [37] and Wang et al. [38] also do related work aiming at reducing the size and computational cost of convolutional neural networks. Chollet et al. [39] present a novel architecture based on depthwise separable convolutions, named Xception. It shows large gains on the JFT dataset [40].

2.4. Attention module

It's well known that attention plays a crucial role in human perception [41]. Presently, there have been several studies about attention mechanisms to improve the performance of CNNs. Wang et al. [40] propose an encoder-decoder style attention module: Residual Attention Network. The network not only performs well but is also robust to noisy inputs. But directly computing the 3d attention, this network has more computational and parameter overhead. Hu et al. [42] propose a Squeeze-and-Excitation module to exploit the inter-channel relationship. Although using a global average-pooled to compute channel-wise attention, they miss spatial attention. Woo et al. [43] introduce CBAM, which learns channel attention and spatial attention separately. For the qualitative analysis, the authors prove that CBAM outperforms other attention modules, comparing the accuracy improvement from others. Notably, the attention mechanism is reported as an efficient localization method in screening, which can be adopted in COVID-19 applications [14,44].

3. Material and methods

Dataset plays a critical and essential role in deep learning models. However, few datasets with massive CT scans for COVID-19 are available for public access and use until now. To address this problem, we establish a publicly available dataset COVID-CTx. Though the largest of its kind, COVID-CTx still can't meet the training of large networks. For this problem, we propose a light-weighted hybrid neural network AM-SdenseNet for combating overfitting. In the next subsections, we will introduce the COVID-CTx, data preprocessing, and AM-SdenseNet architecture in turn.

3.1. COVID-CTx dataset

COVID-CTx contains 828 CT scans that are positive for COVID-19

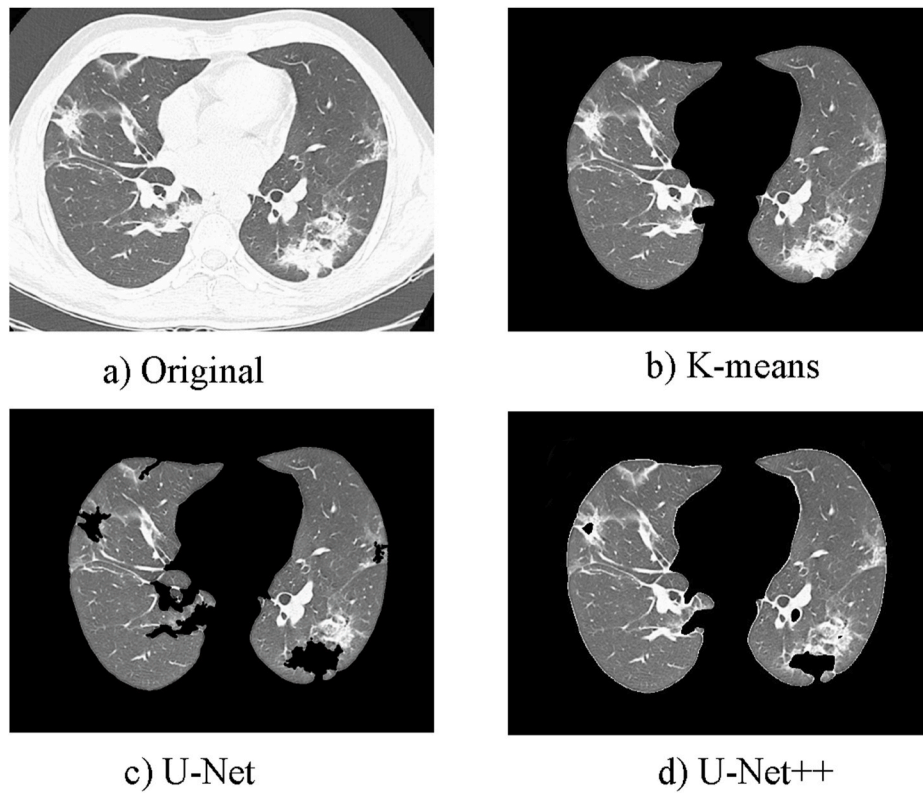


Fig. 2. Results achieved by three lung segmentation methods. a). Original CT; b). Lung achieved by K-means; c). Lung achieved by U-Net; d). Lung achieved by U-Net++.

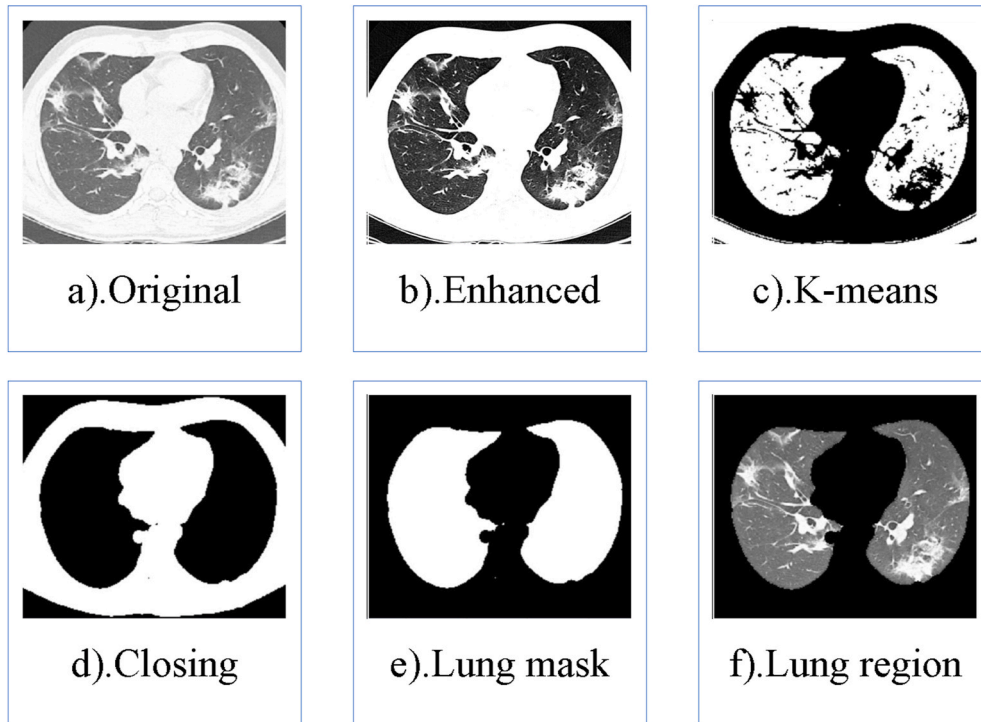


Fig. 3. The process of lung segmentation. a). Original CT; b). Contrast-enhanced CT; c). CT after k-means segmentation; d). CT after closing operation; e). Lung mask; f). Lung region.

across 324 patient cases. The CT scans from the same patient are visually similar. Hence, our dataset has the largest number of publicly available COVID-19 positive cases and richer features about COVID-19 compared

to other public datasets. To build the COVID-CTx, we collected positive CTs for COVID-19 from three open access data repositories: COVID-SIRM, COVID-Seg, and COVID19-CT. Our dataset contains 828 CT

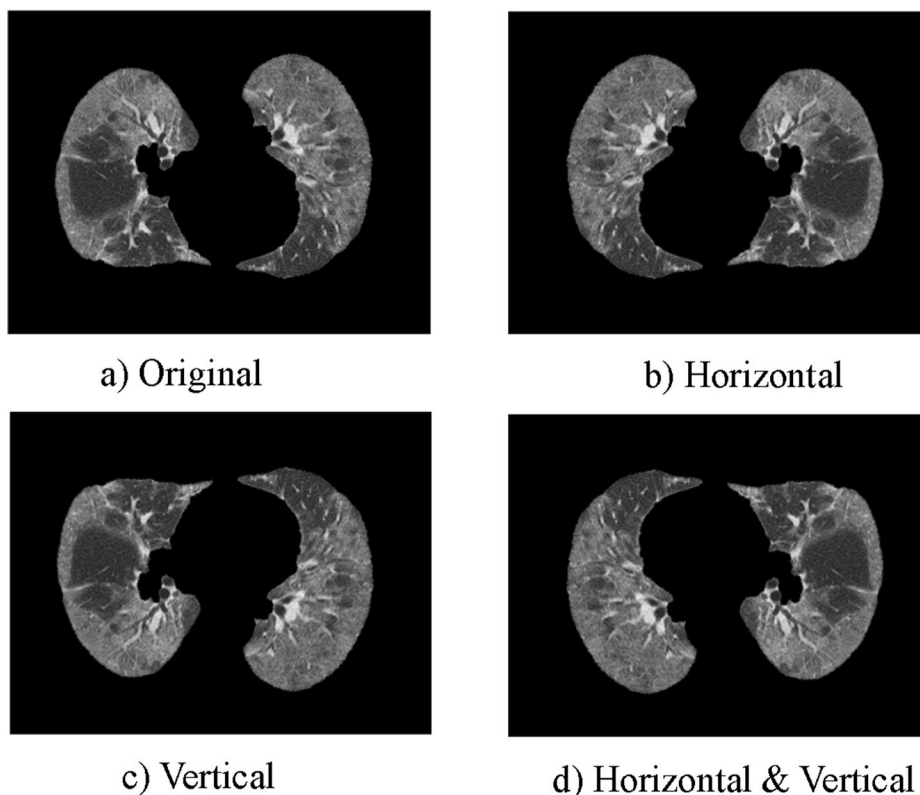


Fig. 4. The methods of augmentation. a). Original CT; b). CT after horizontal mirror; c). CT after vertical mirror; d). CT after horizontal and vertical mirrors.

scans positive about COVID-19, of which 100 CTs are from COVID-SIRM, 379 CTs are from COVID-Seg, and 349 CTs are from COVID19-CT. In addition, COVID-CTx also contains 1000 negative CT scans for COVID-19. Those CT scans, obtained from LUNA16 [45], are normal or containing lung nodules. Fig. 1 shows example CT images from COVID-CTx, of which a) is from COVID-SIRM, b) is from COVID-Seg, and c) is from COVID19-CT.

3.2. Data preprocessing

3.2.1. Lung segmentation

Lung segmentation is an essential step in COVID-19 diagnosis systems to assess and quantify COVID-19. But the training of segmentation networks for CT scans in COVID-19 applications is difficult. Sufficient labeled data for segmentation tasks is often unavailable since manual delineation for lung regions is labor-intensive and time-consuming. Here, U-Net and U-Net++ are first trained from scratch on the LUNA16 and fine-tuned on the COVID-CTx, of which the lung masks are produced by us. In addition, we chose K-means clustering [46] to obtain lung regions due to its simple principle and easy implementation [47, 48]. Fig. 2 shows results achieved by the above three lung segmentation methods. Obviously, K-means reaches a better segmentation result on COVID-19 image segmentation. This demonstrates the effectiveness of K-means, which is more suitable for the segmentation task with insufficient labeled data. Consequently, we segment lung regions by K-means clustering in this study.

The process of lung segmentation is shown in Fig. 3, and the specific

method is as follows.

- (1) The original CT scan is preprocessed to enhance the contrast.
- (2) We separate the foreground (opaque tissue) and background (transparent tissue, the lungs) in the image based on the K-means.
- (3) We use the morphological closing operation to eliminate the residual trachea.
- (4) The hole filling operation algorithm is used to fill the maximum connectivity area in the reverse closing image. Subsequently, the filling image subtracts the reverse closing image to get the lung mask.
- (5) The lung region is extracted by its multiplication with the original CT scan.

3.2.2. Data augmentation

Though the largest of its kind, COVID-CTx may still have the overfitting problem for data-hungry deep learning models. Therefore, two different image augmentation techniques (horizontal mirror and vertical mirror) are utilized to generate COVID-19 training images, as shown in Fig. 4. The

horizontal and vertical mirrors used for image augmentation are done by mirroring the training images in the horizontal and vertical directions, respectively. Table 1 summarizes the number of images per class used for training, validation, and testing for each fold. The number of training samples increases from 1098 to 4392.

Table 1

Number of CTs per class before and after data augmentation.

Dataset	Total number	Training without augmentation			Training with augmentation		
		Train	Validation	Test	Train	Validation	Test
COVID-CTx	1828	1098	365	365	4392	365	365

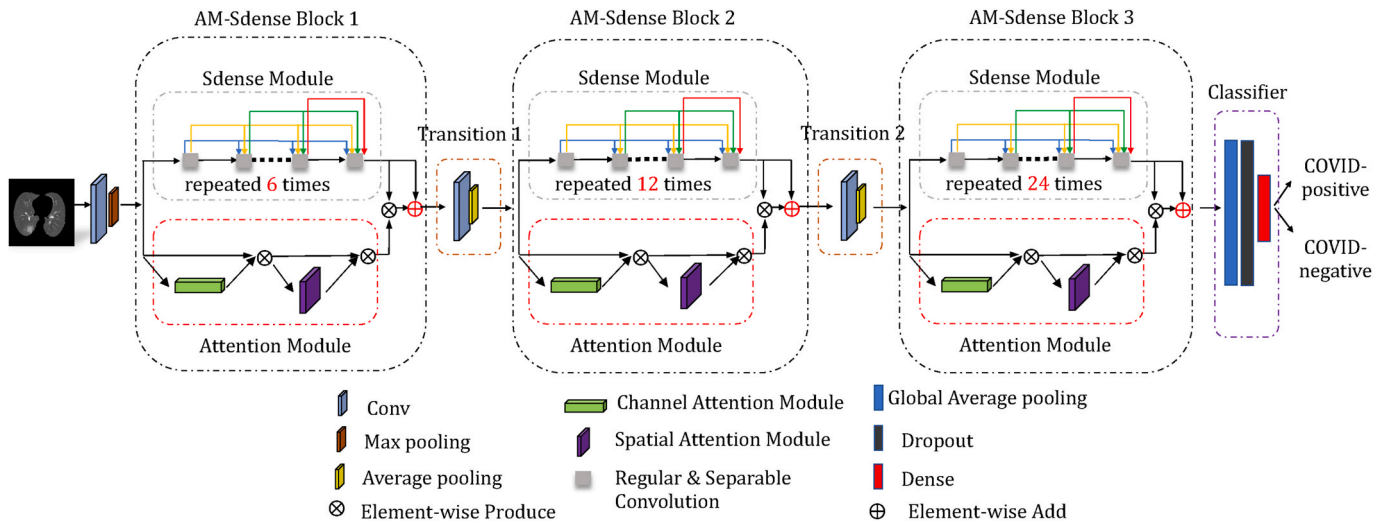


Fig. 5. AM-SdenseNet architecture.

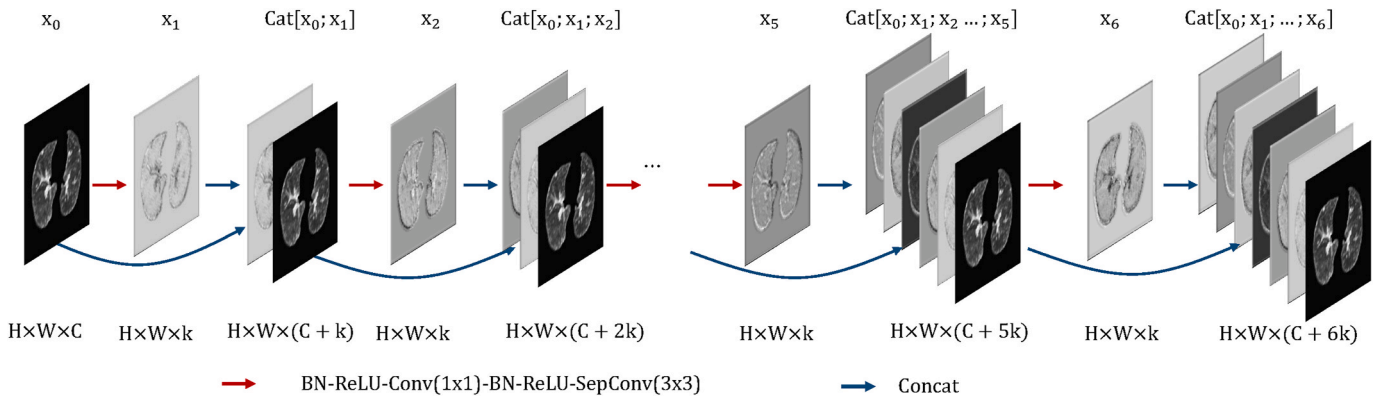


Fig. 6. A 6-layer dense block with a growth rate of k .

3.3. AM-SdenseNet architecture

As shown in Fig. 5, AM-SdenseNet is mainly composed of three AM-Sdense blocks, two transition layers, and one classifier. The AM-Sdense blocks use a DenseNet as the basic network structure to extract image features. Simultaneously, we replace some regular convolutions in the DenseNet with depthwise separable convolutions to reduce training parameters while keeping network performance. In addition, we also apply CBAM on the DenseNet in each AM-Sdense block with residual learning. This operation aims to improve the representation of objective

lesion features and suppress the less relevant ones. The layers between two contiguous dense blocks are transition layers, which do convolution and pooling. It consists of batch normalization (BN) [36] layer and 1×1 convolution layer followed by 2×2 average pooling layer. The output of the last block is fed to the classifier for the final prediction of lung segmentation. The classifier contains a global average pooling layer, a dropout layer with probability $p = 0.5$, and a dense layer with the sigmoid activation function. We now detail the key components of AM-Sdense blocks.

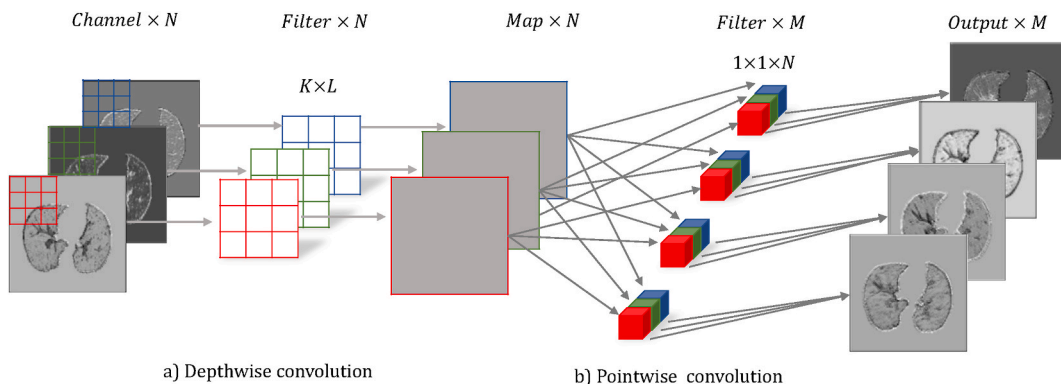


Fig. 7. Depthwise separable convolution.

Table 2
Parameter count comparison across convolution types.

Convolution type	Parameters
Regular	$M \times K \times L$
Depthwise Separable	$M + K \times L$

3.3.1. a.m.-sdense block

In our network, the number of layers in the three AM-Sdense blocks are 6, 12, 24, respectively. Besides, the growth rate of our network is $k = 24$ and the compression of our network is $\theta = 1$. A 6-layer dense block with a growth rate of $k = 24$ is shown in Fig. 6. Our proposed network takes all preceding feature-maps with direct connections as input. Consequently, the channel count used for the input of the following layer increases with the growth rate k . The X_i^{th} layer can be received in Equation (1).

$$X_i = H_i(\text{Cat}[X_0; X_1; \dots; X_{i-1}]) \quad (1)$$

Where $(X_0, X_1, \dots, X_{i-1})$ are the feature-maps produced in layers $(0, 1, \dots, i-1)$. $\text{Cat}(\cdot)$ means to concatenate all input in the channel axis. $H_i(\cdot)$ means four consecutive operations: BN, rectified linear unit (ReLU) [49], convolution (Conv), and depthwise separable convolution (SepConv). The version of H_i is BN-ReLU-Conv(1×1)-BN-ReLU-SepConv(3×3).

3.3.2. Depthwise separable convolution

Many networks show large gains by using depthwise separable convolutions [35,36,39]. Therefore, we replace some regular convolutions in DenseNet with depthwise separable convolutions to reduce the complexity of the model as much as possible. A depthwise separable convolution consists of two parts: a depthwise convolution and a pointwise convolution. The depth-wise convolution is a spatial convolution performed independently over every channel of input. And the pointwise convolution is a regular convolution with 1×1 windows. It projects the channels computed by the depthwise convolution into a new channel space. Fig. 7 shows the schematic diagram.

The specific calculation formulas of the two convolutions are as follows.

$$\text{Conv}(W, y)_{(i,j)} = \sum_{k,l,m}^{K,L,M} W(k, l, m) \cdot y(i+k, j+l, m) \quad (2)$$

$$\begin{aligned} \text{SepConv}(W_p, W_d, y)_{(i,j)} &= \text{PConv}(i, j)_{(i,j)}(W_p, \text{DConv}(i, j)(W_d, y)) \\ &= \sum_m^M W_p \cdot \left(\sum_{k,l}^{K,L} W_d(k, l) \cdot y(i+k, j+l) \right) \end{aligned} \quad (3)$$

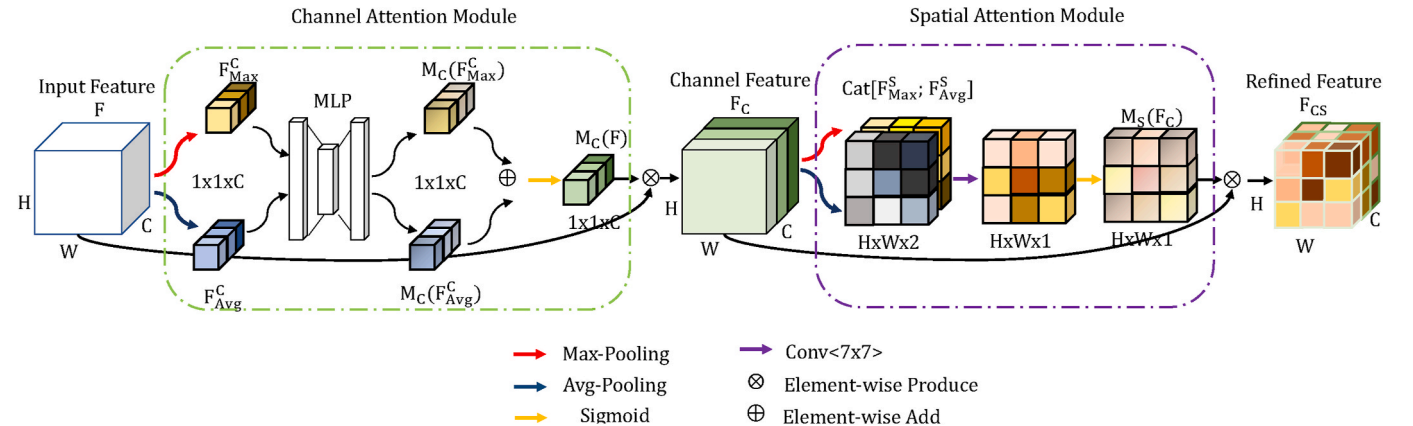


Fig. 8. The overview of CBAM.

Where y is an image, and (i, j) represents pixels. The $W(k, l, m)$ is the kernel with size (K, L, M) . W_p and W_d are the parameters of each separable convolution.

The parameters of a regular convolution and a depthwise separable convolution are shown in Table 2. As can be seen above, the numbers of parameters for a regular convolution and a depthwise separable convolution are $M \times K \times L$ and $M + K \times L$, respectively. When M is much bigger than 1 (as is usually the case), the parameters of separable convolution are much smaller than those of a regular convolution. In this paper, we replace some regular convolutions in DenseNet with depthwise separable convolutions. This operation reduces network parameters by 17.9%.

3.3.3. CBAM

Considering the limited data and computationally limited platform, we apply the CBAM on DenseNet to focus on important features and suppress the less relevant ones. The overview of CBAM is shown in Fig. 8. The channel attention module exploits the inter-channel relationship of features. And the spatial attention module subsequently generates a spatial attention map by utilizing the inter-spatial relationship of features. Given the input feature map $F \in R^{H \times W \times C}$, the channel attention module uses a global average pooling and a global max pooling to extract the average-pooled features F_{Avg}^C and the max-pooled features F_{Max}^C respectively. Both features are passed through a multi-layer perceptron (MLP), and then summed element-by-element to get a channel attention map $M_C(F) \in 1 \times 1 \times C$. The channel attention map $M_C(F)$ and input feature map F are multiplied element-wise to exploit the inter-channel relationship of the features map F . So that the channel feature map $F_C \in R^{H \times W \times C}$ is being generated as Equation (4).

$$\begin{aligned} F_C &= M_C(F) \otimes F \\ &= \sigma(M(F_{Max}^C) + M(F_{Avg}^C)) \otimes F \end{aligned} \quad (4)$$

Where $\sigma(\cdot)$ denotes a sigmoid function and \otimes denotes element-wise multiplication.

In the spatial attention module, the average pooling and max pooling along the channel axis are used to extract two feature maps F_{Avg}^S and F_{Max}^S . Subsequently, two feature maps concatenated in the channel axis are forward to a convolution layer to produce a spatial attention map $M_S(F) \in H \times W \times 1$. The channel attention map $M_S(F)$ and input feature map F_C are multiplied element-wise to exploit the inter-spatial relationship of the features map F_C . Finally, the refined feature map $F_{CS} \in H \times W \times C$ is computed as Equation (5).

$$\begin{aligned} F_{CS} &= M_S(F_C) \otimes F_C \\ &= \sigma(\text{Conv}(\text{Cat}[F_{Max}^S; F_{Avg}^S])) \otimes F_C \end{aligned} \quad (5)$$

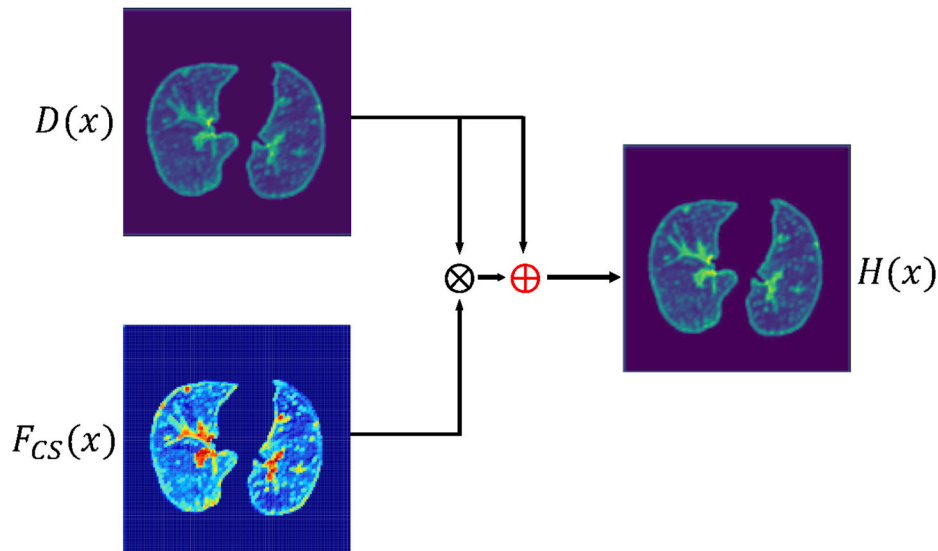


Fig. 9. The diagram of residual learning method.

Table 3
The parameters of AM-SdenseNet architecture.

layers	Output size	AM-SdenseNet
Convolution	256×256	7×7 conv, $s = 2$
Pooling	128×128	3×3 maxpool, $s = 2$
AM-Sdense Block 1	128×128	$\begin{bmatrix} 1 \times 1\text{-conv} \\ 3 \times 3\text{-sconv} \end{bmatrix} \times 6$
Transition 1	128×128	1×1 conv
	64×64	2×2 avgpool, $s = 2$
AM-Sdense Block 2	64×64	$\begin{bmatrix} 1 \times 1\text{-conv} \\ 3 \times 3\text{-sconv} \end{bmatrix} \times 12$
Transition 2	64×64	1×1 conv
	32×32	2×2 avgpool, $s = 2$
AM-Sdense Block 3	32×32	$\begin{bmatrix} 1 \times 1\text{-conv} \\ 3 \times 3\text{-sconv} \end{bmatrix} \times 24$
Classification	1×1	global average pool dropout, $p = 0.5$ Dense, sigmoid

Where $Conv(\cdot)$ represents a convolution operation with a filter size of 7×7 .

In our research, we apply CBAM on DenseNet with residual learning, as shown in Equation (6).

$$H(x) = (1 + F_{CS}(x)) \otimes D(x) \quad (6)$$

Where $H(x)$ is the output of the AM-Sdense block, x is the input of CBAM and dense block. $F_{CS}(x)$ is the output of CBAM. And $D(x)$ is the output of the dense block.

It is worth noting that residual learning can greatly avoid the problem of interference by the attention mechanism. Even when the CBAM output $F_{CS}(x)$ is 0, $H(x)$ still retains main features in $D(x)$. As clearly seen from Fig. 9, CBAM can greatly focus on lesion features in $F_{CS}(x)$.

Table 4
Evaluation metrics used in COVID-19 detection.

Metrics	Equations	Notes
Accuracy	$ACC = (TP + TN)/(TP + TN + FP + FN)$	
Precision	$PPV = TP/(TP + FP)$	Positive Predicted Value (PPV)
Recall	$SE = TP/(TP + FN)$	True Positive Rate (TPR) or Sensitivity
F1-score	$F1 = 2TP/(2TP + FP + FN)$	Harmonic mean of Precision and Recall

Subsequently, lesions in $H(x)$ perform more prominent than those in $D(x)$.

3.3.4. The details of AM-SdenseNet architecture

When dataset is limited, conventional shallow CNN models produce better results as compared to deeper models [50]. Therefore, AM-SdenseNet only has three dense blocks. The number of layers in the three AM-Sdense blocks are 6, 12, 24, respectively. In addition, the growth rate for our networks is $k = 24$ and the compression for our network is $\theta = 1$. AM-SdenseNet takes the lung regions as the input images with size $512 \times 512 \times 3$. Before the data enters the first AM-Sdense block, a convolution layer and a max pooling layer with 48 output channels are performed on the input images. In addition, we use 1×1 convolution followed by 2×2 average pooling between two contiguous dense blocks to change feature-map sizes. At the end of the thirdly dense block, a global average pooling is performed and then a dropout layer with probability $p = 0.5$ is attached. Finally, a dense layer with the sigmoid activation function directly outputs the probabilities of being COVID-positive and COVID-negative. Table 3 shows the details of AM-SdenseNet architecture, where each ‘‘conv’’ layer corresponds to the sequence BN-ReLU-Conv and ‘‘sconv’’ corresponds to the sequence BN-ReLU-SepConv.

4. Experiments and discussions

We empirically demonstrate the effectiveness of AM-SdenseNet on the COVID-19 detection and compare it with state-of-the-art architectures, such as DenseNet and its variants. In the following subsections, we will introduce experimental settings, the influence of CBAM and depthwise separable convolution on the AM-SdenseNet, and comparisons between our approach with several state-of-the-art baselines on the COVID-19 classification task.

4.1. Experimental settings

Our COVID-CT_x is used in all approaches. Besides COVID-CT_x, we also evaluate seven networks trained on COVID-19 Radiology Database. All images are resized to $512 \times 512 \times 3$. The ratio of training set, validation set, and test set is 0.6 : 0.2 : 0.2. It is worth noting that each patient belongs to a single set. BN is used through all models, and binary cross-entropy serves as the loss function. The optimizer is Adam [51] with an initial learning rate of $5e - 5$ and a weight decay of $1e - 7$. The models are implemented in Keras and trained with Tesla V100. All

Table 5
Metrics for three networks with and without augmentation.

Schemes	Models	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Parameters(M)
Without augmentation	denseNet	96.71	97.40	95.89	96.52	3.44
	AM-denseNet	97.53	97.73	97.15	97.42	4.12
	AM-SdenseNet	98.36	98.19	98.40	98.29	3.38
With augmentation	denseNet	97.21	97.60	97.24	97.41	3.44
	AM-denseNet	97.55	97.29	97.60	97.44	4.12
	AM-SdenseNet	99.18	99.32	98.97	99.14	3.38

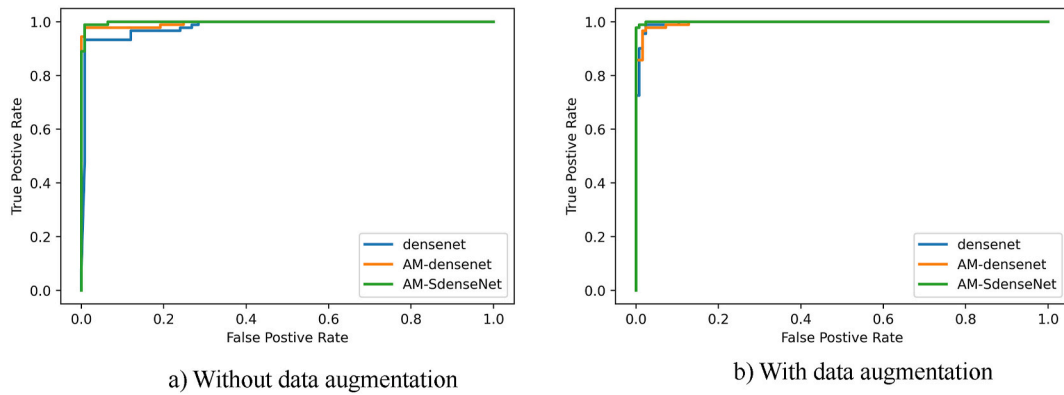


Fig. 10. Comparison of the ROC curve for COVID-19 classification using CNN based models without a) and with b) image augmentation.

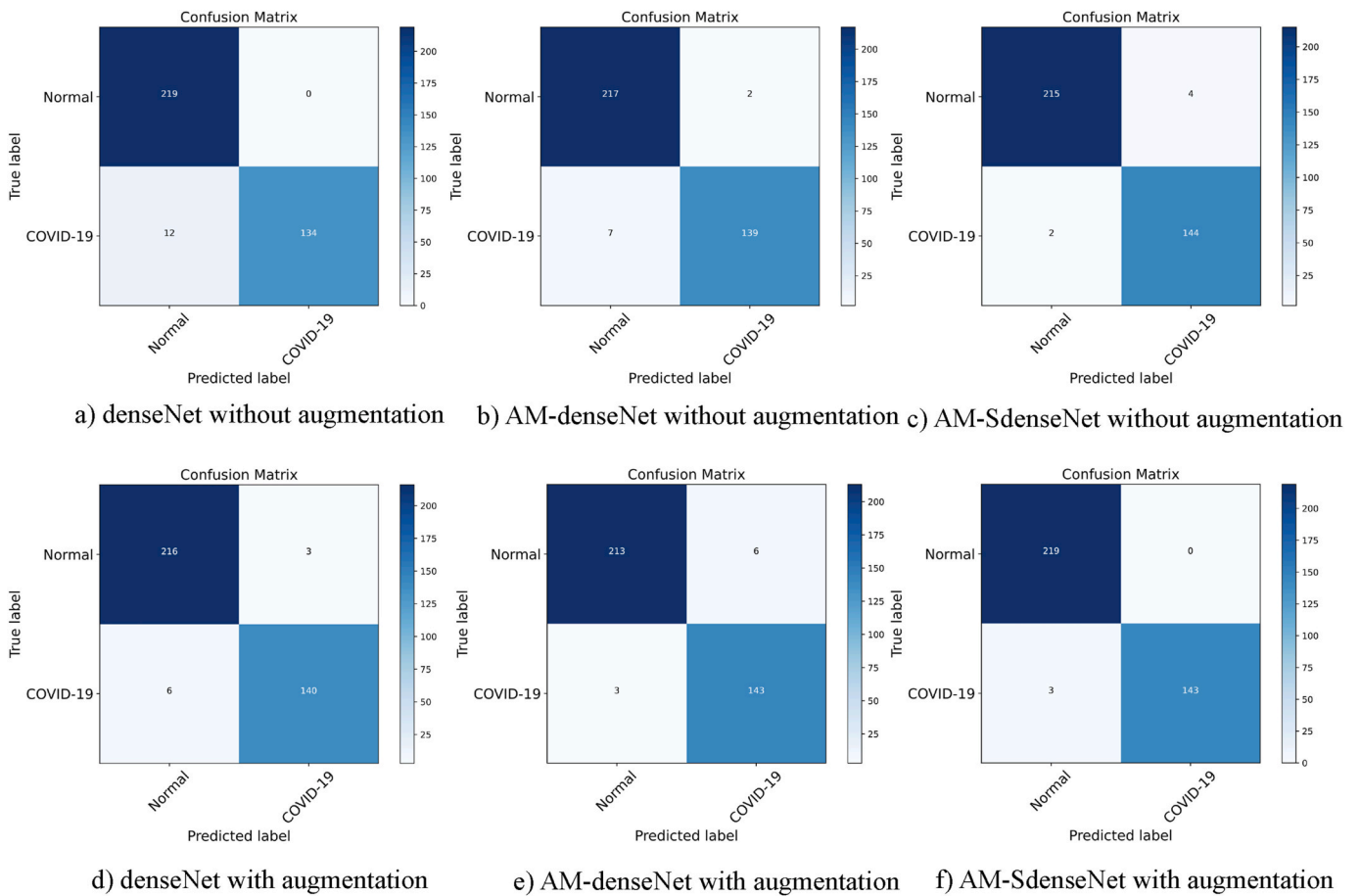


Fig. 11. Confusion matrices for three models with and without data augmentation.

Table 6
Metrics for seven networks trained on COVID-CTx and COVID-19 Radiography Database.

Data sets	Models	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Parameters(M)
COVID-CTx	MobileNet	92.05	93.19	90.53	91.49	3.23
	InceptionV3	94.79	95.80	93.61	94.46	21.80
	ResNet50	95.07	95.19	94.52	94.83	24.97
	VGG16	97.26	97.66	96.69	97.12	14.72
	DenseNet169	97.53	97.38	97.49	97.43	12.64
	DenseNet121	98.08	97.85	98.17	98.00	7.04
	AM-SdenseNet	99.18	99.32	98.97	99.14	3.38
COVID-19 Radiography Database	MobileNet	94.09	93.35	92.26	92.86	3.23
	InceptionV3	97.24	95.25	96.86	96.23	21.80
	ResNet50	96.44	95.06	94.85	95.25	24.97
	VGG16	97.05	97.35	96.35	96.92	14.72
	DenseNet169	98.03	98.45	97.59	98.03	12.64
	DenseNet121	97.62	97.32	96.01	96.84	7.04
	AM-SdenseNet	98.62	98.66	97.56	98.33	3.38

models are trained with 40 epochs and a mini-batch size of 16. We evaluate different networks using four metrics: accuracy, precision, recall, and F1-score, which are shown in Table 4. Where TP and TN mean true positive and negative parameters, respectively. FP and FN are false positive and false negative values respectively. 3-Folds cross-validation is applied by all approaches.

4.2. Evaluation of two modules

To demonstrate the efficacy of AM-SdenseNet and investigate the effects of CBAM and depthwise separable convolution, we first experiment on the denseNet (initial network), AM-denseNet (combined with CBAM), and AM-SdenseNet. The comparative performance for three CNNs for COVID-19 classification problem with and without augmentation is shown in Table 5 and comparative AUC curves are shown in Fig. 10. Additionally, we also built the confusion matrices for three models, as shown in Fig. 11.

As shown in Table 5, three models with augmentation all show some increase in four performance metrics. It proves that the benefits of augmentation are highly significant when the data set is limited. Comparing denseNet with AM-denseNet, we can see that the network combined with CBAM yields higher classification performance. Comparing AM-denseNet with AM-SdenseNet, depthwise separable convolution reduces network parameters by 17.9% and improves precision by 2.03% at the same time. We empirically demonstrate depthwise separable convolution can reduce training parameters while keeping network performance. In summary, the performance of AM-SdenseNet benefits a lot from CBAM and depthwise separable convolution. Trained with augmented images, AM-SdenseNet produces the highest accuracy of 99.18%, precision of 99.32%, recall of 98.97%, and F1 score of 99.14% with the least parameters 3.38 M. Fig. 10 shows AM-SdenseNet achieves the best curve through data augmentation.

According to the confusion matrices presented in Fig. 11, AM-SdenseNet can classify the Normal class with 0 misclassifications and classify the COVID-19 class with 3 misclassifications.

4.3. Comparison of seven models

Few data sets with massive CT scans on COVID-19 are available for public access and use until now due to privacy concerns. Hence, to illustrate the effectiveness of AM-SdenseNet in COVID-19 diagnosis, we separately investigate the performance of networks trained on COVID-CTx and COVID-19 Radiography Database with different backbones, including MobileNet, InceptionV3, ResNet50, VGG16, DenseNet169, DenseNet121, and AM-SdenseNet.

Table 6 shows the metrics for seven networks trained on COVID-CTx and COVID-19 Radiography Database. Trained on COVID-CTx, DenseNet169 and VGG16 present a better performance than InceptionV3 and ResNet50 with fewer parameters. What's more, DenseNet121 outperforms DenseNet169 on the COVID-19 classification task, achieving higher accuracy and precision. This significantly proves that conventional shallow CNN models produce better results as compared to deeper models when the data set is limited. But the network is too simple to extract features well. For instance, MobileNet lacks stronger feature representation learning capabilities, so as to produce low classification performance. Compared with MobileNet, DenseNet can strengthen feature propagation and substantially reduce the number of parameters. The performance benefits a lot from dense connections. For example, DenseNet121 and DenseNet169 achieve good performance on the COVID-19 classification. AM-SdenseNet, however, has a much smaller number of parameters but the performance is better than deeper networks such as DenseNet121 and DenseNet169. The reason is that large-sized networks are more prone to over-fitting, especially considering that our data set is fairly small. Under such circumstances, CBAM and

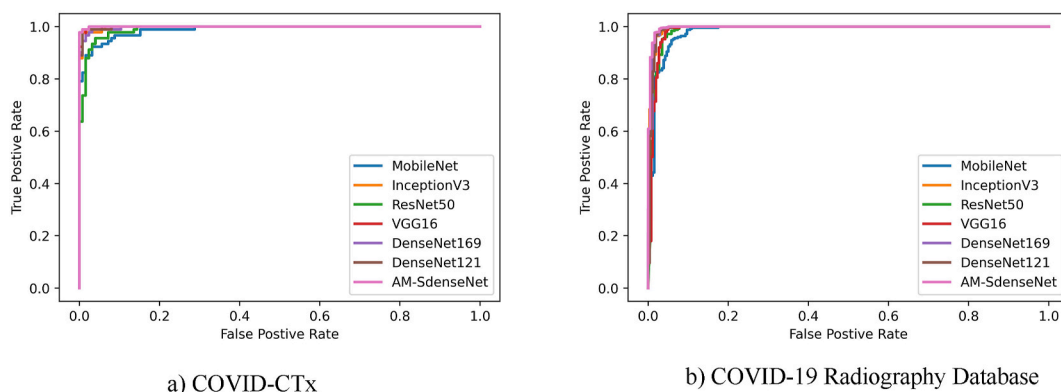


Fig. 12. Comparisons of the ROC curve for COVID-19 classification on seven different CNNs.

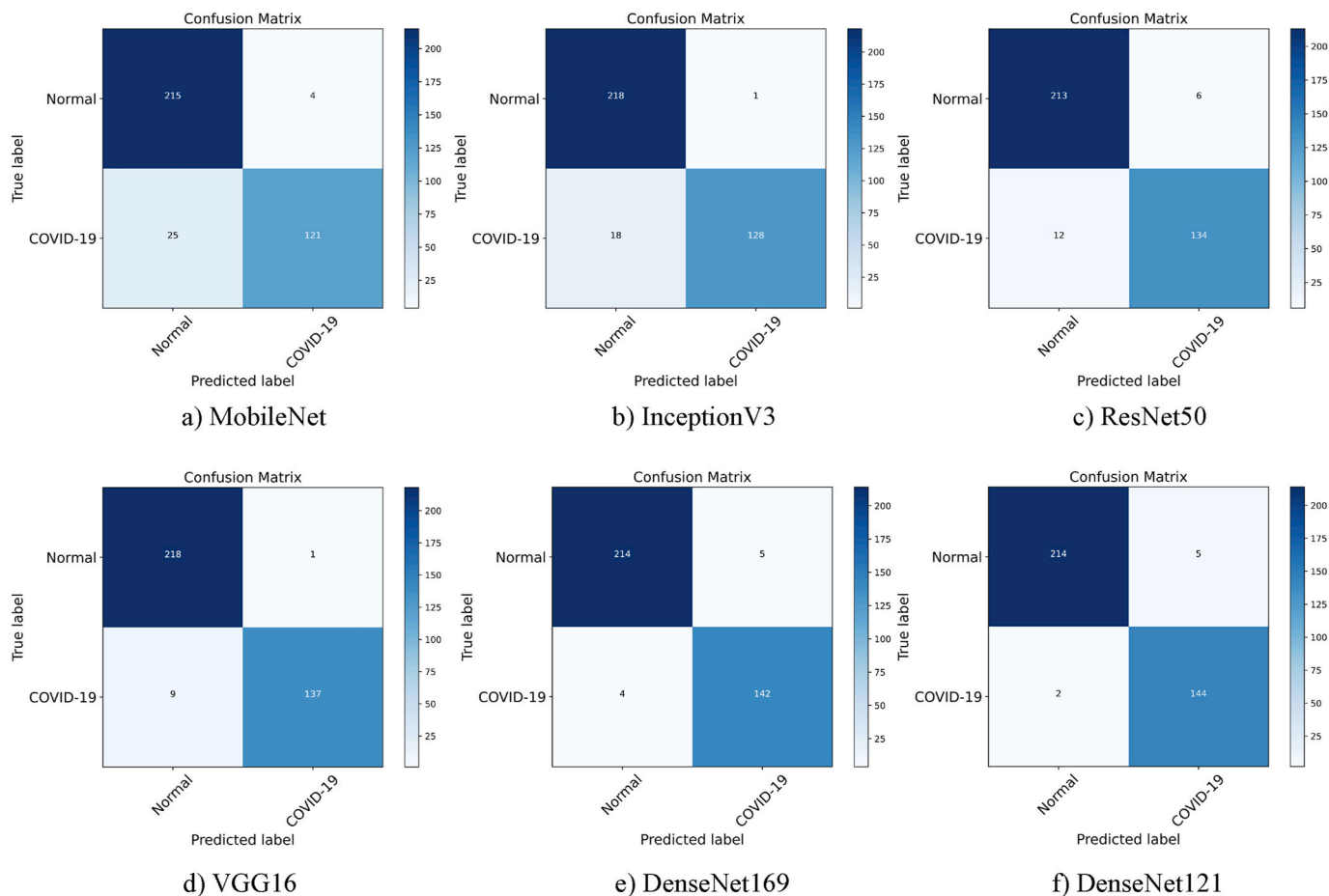


Fig. 13. Confusion matrices for six models.

Table 7

Related studies with medical images for COVID-19 diagnosis.

Model	Modality	COVID-19	Open-source	Parameters (M)	Accuracy (%)
DeCoVNet [11]	CT	313 cases	No	–	90.01
DSAE [52]	CT	317 cases	No	–	97.14
AM-SdenseNet	CT	324 cases	Yes	3.38	99.18
MobileNetV2 + InceptionV3 [50]	X-ray	1000 images	Yes	28.50	98.45
Covid-ResNet [53]	X-ray	1200 images	Yes	25.60	96.23
AM-SdenseNet	X-ray	1200 images	Yes	3.38	98.62

depthwise separable convolution have a better chance to play their value. According to the confusion matrices presented in Figs. 11 and 13, AM-SdenseNet presents the highest ratio of classifying the COVID-19 class and the Normal class. Other models have achieved good performance in the classification of the Normal class, but their performance in the classification of COVID-19 is not good enough.

Compared with CT, X-ray is more easily accessible around the world. However, due to the ribs projected onto soft tissues in 2D and thus confounding image contrast, the classification of X-ray images is even more challenging. Trained on the COVID-19 Radiography Database, AM-SdenseNet reaches the best results with an accuracy of 98.62%. It is confirmed that AM-SdenseNet is universal in COVID-19 diagnosis. In

Fig. 12, AM-SdenseNet also achieves the best curve on the COVID-19 classification task. Through the above experiments, AM-SdenseNet achieves the best performance on the COVID-19 classification task compared to other models. Such experimental results not only illustrate the effectiveness of COVID-CTx, but also provide concrete evidence that AM-SdenseNet has stronger capabilities to improve the speed and accuracy of COVID-19 diagnosis.

4.4. Comparison with the latest methods

The comparison of AM-SdenseNet with the latest approaches for COVID-19 classification is presented in this section. Table 7 shows that only limited COVID-19 images are used in most of these approaches. Given the problem of insufficient samples, studies adopt models such as DenseNet, ResNet, and ensemble networks to combat over-fitting. While considering the performance metrics in Table 7, our approach outperforms the considered state-of-the-art approaches, achieving the best classification performance from CT scan and chest X-ray images, respectively. It is further confirmed that AM-SdenseNet is superior to other algorithms in COVID-19 diagnosis and useful to control the spreading of infection.

5. Conclusions and future works

In our research, we aim to develop a sample and efficient CAD system to diagnose COVID-19 from CT scans. To accelerate the open study in this area, we establish a publicly available data set COVID-CTx. To our knowledge, it has the richest features about COVID-19 compared to other public data sets to date. Although the largest informative, it still has a risk of overfitting for data-hungry deep learning models. For this

problem, we propose a light-weighted hybrid neural network: AM-SdenseNet. The network synergistically applies CBAM on DenseNet in each block with residual learning, which can greatly improve the representation of objective lesion features and suppress the less relevant ones. In addition, we replace some regular convolutions in DenseNet with depthwise separable convolutions to reduce training parameters while keeping network performance. Through experiments, it has been proved that AM-SdenseNet can greatly improve the speed and accuracy of COVID-19 diagnosis, which is extremely useful to control the spreading of infection.

Though the largest of its kind, COVID-CTx still can't meet the training of large networks, which need a larger dataset. However, deep learning with small samples is still an important research direction in the future. Unsupervised learning methods such as stack auto-encoder [54] and restricted boltzmann machine [55] can be used for reference. Considering the actual clinical need, the CAD system can be combined with the hospital's imaging system and electronic medical records to achieve the follow-up treatment of patients in the future.

Funding

No funds, grants, or other support was received.

CRedit authorship contribution statement

Qian Li: Conceptualization of this study, methodology, dataset, experiments, and writing paper. **Jiangbo Ning:** Investigation, validation, writing, review and editing. **Jianping Yuan:** Data processing. **Ling Xiao:** Validation and formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in China: a report of 1014 cases, *Radiology* 296 (2020) E32–E40.
- [2] J.P. Kanne, B.P. Little, J.H. Chung, B.M. Elicker, L.H. Ketaj, Essentials for Radiologists on Covid-19: an Update—Radiology Scientific Expert Panel, 2020.
- [3] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z.A. Fayad, et al., Ct imaging features of 2019 novel coronavirus (2019-ncov), *Radiology* 295 (2020) 202–207.
- [4] P. Lomoro, F. Verde, F. Zerboni, I. Simonetti, C. Borghi, C. Fachineetti, A. Natalizi, A. Martegani, Covid-19 pneumonia manifestations at the admission on chest ultrasound, radiographs, and ct: single-center study and comprehensive radiologic literature review, *European journal of radiology open* 7 (2020) 100231.
- [5] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest ct for covid-19: comparison to rt-pcr, *Radiology* 296 (2020) E115–E117.
- [6] M. Wang, C. Xia, L. Huang, S. Xu, C. Qin, J. Liu, Y. Cao, P. Yu, T. Zhu, H. Zhu, et al., Deep learning-based triage and analysis of lesion burden for covid-19: a retrospective study with external validation, *The Lancet Digital Health* 2 (2020) e506–e515.
- [7] K.-L. Hua, C.-H. Hsu, S.C. Hidayati, W.-H. Cheng, Y.-J. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, *OncoTargets Ther.* 8 (2015).
- [8] P. Lakhani, B. Sundaram, Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks, *Radiology* 284 (2017) 574–582.
- [9] B.E. Bejnordi, M. Veta, P.J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J.A. Van Der Laak, M. Hermesen, Q.F. Manson, M. Balkenhol, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, *Jama* 318 (2017) 2199–2210.
- [10] H. Mohsen, E.-S.A. El-Dahshan, E.-S.M. El-Horbaty, A.-B.M. Salem, Classification using deep learning neural networks for brain tumors, *Future Computing and Informatics Journal* 3 (2018) 68–71.
- [11] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep Learning-Based Detection for Covid-19 from Chest Ct Using Weak Label, *MedRxiv*, 2020.
- [12] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography, *Sci. Rep.* 10 (2020) 1–11.
- [13] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, Y. Shi, Lung Infection Quantification of Covid-19 in Ct Images with Deep Learning, *arXiv Preprint arXiv:2003.2020*, 04655.
- [14] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, et al., A deep learning system to screen novel coronavirus disease 2019 pneumonia, *Engineering* 6 (2020) 1122–1129.
- [15] C. Jin, W. Chen, Y. Cao, Z. Xu, Z. Tan, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, et al., Development and evaluation of an artificial intelligence system for covid-19 diagnosis, *Nat. Commun.* 11 (2020) 1–14.
- [16] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Zha, et al., Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images, *IEEE ACM Trans. Comput. Biol. Bioinf* (2021).
- [17] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Using artificial intelligence to detect covid-19 and community-acquired pneumonia based on pulmonary ct: evaluation of the diagnostic accuracy, *Radiology* 296 (2020) E65–E71.
- [18] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, et al., Ai-assisted ct imaging analysis for covid-19 screening: building and deploying a medical ai system in four weeks, *MedRxiv* (2020).
- [19] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, P. Xie, Sample-efficient Deep Learning for Covid-19 Diagnosis Based on Ct Scans, *MedRxiv*, 2020.
- [20] M.E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al Emadi, et al., Can ai help in screening viral and covid-19 pneumonia? *IEEE Access* 8 (2020) 132665–132676.
- [21] O. Ronneberger, P. Fischer, T. Brox, U-net, Convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.
- [22] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: a nested u-net architecture for medical image segmentation, in: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 3–11.
- [23] F. Milletari, N. Navab, S.-A. Ahmadi, V-net, Fully convolutional neural networks for volumetric medical image segmentation, in: *2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, 2016, pp. 565–571.
- [24] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.
- [26] The Italian Society of Medical Radiology, Covid-19 cases. <https://www.eurorad.org/advanced-search?search=COVID/gt.html>, 2021. (Accessed 17 June 2021) accessed.
- [27] M. Jun, C. Ge, W. Yixin, A. Xingle, G. Jiantao, Z. Minqing, L. Xin, D. Xueyuan, He Jian, Covid-19 ct lung and infection segmentation dataset. <https://zenodo.org/record/3757476/#.YMX8zhP7TDI/gt.html>, 2021. (Accessed 17 June 2021) accessed.
- [28] 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.
- [29] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: convolutional block attention module, in: *Proceedings of the European Conference on Computer Vision, ECCV*, 2018, pp. 3–19.
- [30] L. Sifre, S. Mallat, Rigid-motion Scattering for Texture Classification, 2014 *arXiv preprint arXiv:1403.1687*.
- [31] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets, Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017, 04861 *arXiv preprint arXiv:1704*.
- [32] 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.
- [33] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014, p. 1556, *arXiv preprint arXiv:1409*.
- [34] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Artificial Intelligence Distinguishes Covid-19 from Community Acquired Pneumonia on Chest Ct, *Radiology*, 2020.
- [35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [36] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: *International Conference on Machine Learning, PMLR*, 2015, pp. 448–456.
- [37] J. Jin, A. Dundar, E. Culurciello, Flattened Convolutional Neural Networks for Feedforward Acceleration, 2014, p. 5474, *arXiv preprint arXiv:1412*.
- [38] M. Wang, B. Liu, H. Foroosh, Factorized convolutional neural networks, in: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 545–553.
- [39] 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.
- [40] G. Hinton, O. Vinyals, J. Dean, Distilling the Knowledge in a Neural Network, 2015 *arXiv preprint arXiv:1503.02531*.
- [41] M. Corbetta, G.L. Shulman, Control of goal-directed and stimulus-driven attention in the brain, *Nat. Rev. Neurosci.* 3 (2002) 201–215.

- [42] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.
- [43] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
- [44] G. Gaál, B. Maga, A. Lukács, Attention U-Net Based Adversarial Architectures for Chest X-Ray Lung Segmentation, 2020, p. 10304, arXiv preprint arXiv:2003.
- [45] S.G. Armato III, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A. P. Reeves, B. Zhao, D.R. Aberle, C.I. Henschke, E.A. Hoffman, et al., The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans, *Med. Phys.* 38 (2011) 915–931.
- [46] S.Z. Selim, M.A. Ismail, K-means-type algorithms: a generalized convergence theorem and characterization of local optimality, *IEEE Trans. Pattern Anal. Mach. Intell.* (1984) 81–87.
- [47] S. Zahra, M.A. Ghazanfar, A. Khalid, M.A. Azam, U. Naeem, A. Prugel-Bennett, Novel centroid selection approaches for kmeans-clustering based recommender systems, *Inf. Sci.* 320 (2015) 156–189.
- [48] K. Peng, V.C. Leung, Q. Huang, Clustering approach based on mini batch kmeans for intrusion detection system over big data, *IEEE Access* 6 (2018) 11897–11906.
- [49] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.
- [50] F. Ahmad, M.U.G. Khan, K. Javed, Deep learning model for distinguishing novel coronavirus from other chest related infections in x-ray images, *Comput. Biol. Med.* 134 (2021) 104401.
- [51] D.P. Kingma, J. Ba, Adam, A Method for Stochastic Optimization, 2014 arXiv preprint arXiv:1412.6980.
- [52] J. Cheng, W. Zhao, J. Liu, X. Xie, S. Wu, L. Liu, H. Yue, J. Li, J. Wang, J. Liu, Automated diagnosis of covid-19 using deep supervised autoencoder with multi-view features from ct images, *IEEE ACM Trans. Comput. Biol. Bioinf* (2021).
- [53] M. Farooq, A. Hafeez, Covid-resnet: A Deep Learning Framework for Screening of Covid19 from Radiographs, 2020, p. 14395, arXiv preprint arXiv:2003.
- [54] X. Yuan, S. Qi, Y. Wang, Stacked enhanced auto-encoder for data-driven soft sensing of quality variable, *IEEE Transactions on Instrumentation and Measurement* 69 (2020) 7953–7961.
- [55] Y. Bengio, Deep learning of representations for unsupervised and transfer learning, in: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, JMLR Workshop and Conference Proceedings, 2012, pp. 17–36.