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.
DDCNNC: Dilated and depthwise separable convolutional neural Network for diagnosis COVID-19 via chest X-ray images

Xiang Li\(^a\), Mengyao Zhai\(^b\), Junding Sun\(^a,\,*\)

\(^a\) College of Computer Science and Technology, Henan Polytechnic University China
\(^b\) College of Humanities and Education, Hebi Polytechnic China

**A R T I C L E   I N F O**

**Keywords:**
Convolutional neural networks
COVID-19
Depthwise separable convolution
Deep learning
Dilated convolution

**A B S T R A C T**

**Purpose:** As of December 21, 2020, a total of 77,670,400 cases of coronavirus disease 2019 (COVID-19) have been confirmed worldwide, 53,825,243 cases have been cured and 1,693,253 cases have died. Among the diagnostic methods of COVID-19, chest X-ray images have the advantages of fast imaging, low cost and high accuracy of single plane lesions recognition. The current COVID-19 detection models have shortcomings such as weak robustness, unreliable generalization ability, and long training time.

**Methods:** To solve the above problems, our team proposed two novel frameworks and five methods to diagnose COVID-19 based on chest X-ray images. (i) A novel framework – depthwise separable convolutional neural network (DCNN), and we tested Three methods, viz., using LeNet-5, VGG-16, and ResNet-18 as backbones. (ii) A novel framework – dilated and depthwise separable convolutional neural network (DDCNN), and we tested Two methods, viz., using VGG-16 and ResNet-18 as backbones.

**Results:** Experiment results show that our models not only improve the detection accuracy, but also reduce the training time.

**Conclusions:** Our methods are superior to state-of-the-art methods in both above aspects.

**Introduction**

The pathogen of COVID-19 is different from that of pneumonia. COVID-19 is caused by coronavirus (Toapanta et al., 2021) and is highly contagious (Satapathy, 2021). Pneumonia is mainly caused by bacterial infection and is usually not infectious. The symptoms of COVID-19 and pneumonia are also different. After infected COVID-19, symptoms such as fever, fatigue, and dry cough will appear (Marquez et al., 2021). Critically ill patients will have dyspnea or even respiratory failure within 7 days, leading to death. After infected pneumonia, symptoms such as fever and cough will appear Satapathy (2021). The progress of pneumonia is slow and symptoms can be quickly relieved after treatment.

The main diagnostic methods for COVID-19 are: (i) nucleic acid detection. This method not only has high requirements for test equipment, but also high requirements for laboratory cleanliness and operators (Lerner et al., 2020). The quality of sample collection could directly affect the test results. And it takes a longer time from sampling to issuance of the inspection report than other methods. (ii) serum antibody detection. The accuracy of this method is lower than nucleic acid detection (Kulkarni et al., 2021). For the patient who is negative after cured, the serum antibody detection result may still be positive (Mendoza et al., 2021). Therefore, it cannot be used as the only reference standard for diagnosis. (iii) chest computed tomography (CCT) Wang (2021). The cost of this method is higher than others. For lesions with a density similar to normal tissues, the plain scan is likely to be missed, and enhanced scanning is required (Wang, 2021, Fernandes, 2021). (iv) chest X-ray images. The cost of this method is lower than CCT (Pham, 2021), and it has high quality contrast and sharpness of the images (Ismael & Sengur, 2020). Besides, the chest X-ray images can be stored for a long time. It is convenient for doctors to compare and consult the images when patients come back for consultation. Therefore, our team chose chest X-ray images as the experimental dataset in this paper.

With the rapid development of deep learning, more and more scholars have begun to apply deep learning technology, especially convolutional neural network (CNN) to the artificial intelligence aided diagnosis system. Shanthi and Sabeenian (Shanthi & Sabeenian, 2019) improved the AlexNet and applied it to the classification of diabetic retinopathy. The classification accuracy of this method can reach 96.6%. Li, Shen (Li et al., 2019) proposed a multi-resolution network based on VGG to detect benign and malignant pulmonary nodules based on chest X-ray images. The experimental results show that the classification accuracy of this method is 99%. Although the above models can achieve high classification accuracy in the binary classification problems, they have unreliable generalization ability when dealing with the multi-classification problems.
problems. Zhang (2020) proposed a seven-layer customized CNN to classify COVID-19 images.

Lin et al. (2017) proposed RetinaNet model to detect dense objects. And an improved RetinaNet model is also proposed to detect breast cancer via X-ray images. Based on the RetinaNet model, Mercan et al. (2019) proposed a detection method for epithelial cells in breast cancer tissues. Harsono et al. (2020) proposed a method for detection pulmonary nodules based on chest CT images. Zlocha et al. (2019) proposed a lesion detection method via CT images based on RetinaNet model. The above models have shortcomings such as: long training time, easy overfitting, and weak robustness.

To solve above problems, our team proposed two novel frameworks and five methods to diagnose COVID-19 based on chest X-ray images. The experimental results show that our models have higher detection accuracy and less running time than the state-of-the-art methods, and the models have strong robustness and reliable generalization. In this paper, we proposed five novel contributions as follows:

(i) A novel framework – Depthwise separable Convolutional Neural Network (DCNN) – is proposed, and we tested three configurations, viz., using LeNet-5, VGG-16, and ResNet-18 as backbones.

(ii) A novel framework – Dilated and Depthwise separable Convolutional Neural Network (DDCNN) – is proposed, and we tested two configurations, viz., using VGG-16 and ResNet-18 as backbones.

(iii) Three novel methods of DCNN (DCNNC-I, DCNNC-II, and DCNNC-III) are proposed for COVID-19, which are based on DCNN for COVID-19 dataset.

(iv) Two novel methods of DDCNN (DDCNNC-I, DDCNNC-II) are proposed for COVID-19, which are based on DDCNN for COVID-19 dataset.

(v) We find DDCNNC-I has the highest accuracy and DCNNC-I costs the least training time.

**Dataset**

To make the experimental results more contrast and reliability, the dataset in this paper is a public dataset from the Kaggle website (COVID19 with Pneumonia & Normal Chest Xray(PA) Dataset, 2020). There are 6939 images in the dataset, which are divided into three categories: COVID-19, Normal, and Pneumonia. Each category has 2313 images. We randomly select 20% of each category of images as the test set and the remaining 80% as the training set. Therefore, there are 1389 images in the test set, 463 images of each category. The equipment used in the experiment is NVIDIA QUADRO RTX 8000. Total board power is 295W. Total graphics power is 260W. GPU memory is 48GB GDDR6 with ECC. NVIDIA tensor cores are 576. NVIDIA RT cores are 72. Figure 1 shows samples of chest X-ray images from our dataset. Here, the COVID-19 chest X-ray images show ground glass shadow, patch shadow, invagination shadow, or lung consolidation. And the pneumonia chest X-ray images show that the lobes of the lungs showed large blured shadows or small blured patches, which looked like ground glass.

**Methodology**

To ease the understanding of this paper, Table 13 shows all variables used in our study. Table 14 gives the abbreviation and their full names. Table 13 and Table 14 are in the appendix at the end of the paper.

Tables 1, 4, 5

**Depthwise separable convolution**

The calculation methods of standard convolution output size and the number of parameters in the convolution operation (Sahani & Dash, 2021) are as follows.

\[
O_C = \left[ \frac{I_C - F_C + 2P_C}{S_C} + 1 \right]
\]

\[
A_C = N_C \times F_C
\]

Here, \(O_C\) represents the output size of the convolution operation. \(I_C\) represents the input image size. \(F_C\) represents the size of convolution kernel. \(P_C\) represents the size of padding. \(S_C\) represents the stride size of the convolution kernel. \(A_C\) represents the number of parameters of convolution operation. And \(N_C\) represents the number of convolution kernels. When a standard convolution performs feature extraction to an input image (Alencastre-Miranda et al., 2021), all channels of the image target areas are calculated at the same time, to obtain the extracted feature (Garnett et al., 2021). It can be seen that whenever the number of extracted features increases, the corresponding number of convolution kernels also increases (Li, 2018, Pan, 2018).

The feature extraction method of depthwise separable convolution (DSC) (Vorogunti et al., 2020) is different from standard convolution, which process is divided into two steps (Tsai and Lee, 2020). Step 1, performs convolution operation on each channel in the input image target areas to obtain the corresponding results. Step 2, use a \(1 \times 1\) convolution kernel to perform the standard convolution operation on the result obtained in step 1 to change the number of channels. This step is also an important manifestation of DSC reducing the number of parameters when it performs the convolution operation. Therefore, when DSC performing feature extraction, it needs to perform two convolution operations.

The parameter calculation methods of standard convolution and DSC performing convolution operation are as follows.

\[
A_{SC} = I_C \times I_C \times L_I \times F_{SC} \times F_{SC} \times L_I \times N_C
\]

\[
A_{DC} = I_C \times I_C \times L_I \times F_{DC} \times F_{DC} \times L_I \times N_C
\]

Here, \(A_{SC}\) represents the number of parameters when standard convolution performs convolution operation. \(A_{DC}\) represents the number of parameters when DSC performs convolution operation. \(L_I\) represents the number of channels of input image. \(F_{SC}\) represents the size of standard convolution kernel. \(F_{DC}\) represents the convolution kernel size in step 1 of the DSC performs feature extraction.

According to (3) and (4), when performing convolution operation, DSC needs to calculate more parameters than standard convolution when the number of convolution kernels is 1. But when the number of convolution kernels is greater than 1, standard convolution needs to calculate more parameters than DSC (Lee et al., 2021). If a model needs to obtain more image features through convolution operations, the number of convolution kernels will increase accordingly. Therefore, using DSC instead of standard convolution can significantly both reduce the number of parameters and the training time of model.

In Section 3.2, we elaborated on a novel framework – DCNN and three methods (DCNNC-I, DCNNC-II, and DCNNC-III) to diagnose COVID-19 based on chest X-ray images.

**Improvement I: A Novel Framework – DCNN**

In Section 3.1, we have explained the feature extraction methods of DSC and standard convolution. Based on DSC can effectively reduce the number of parameters when performing convolution operation, our team proposed a novel framework – DCNN – Depthwise separated Convolutional Neural Network. And we applied this framework to three
CNN models (LeNet-5, VGG-16, and ResNet-18) and proposed three methods (DCNNC-I, DCNNC-II and DCNNC-III) to diagnose COVID-19 based on chest X-ray images.

In our methods, DCNNC-I represents using LeNet-5 model with DSC to diagnose COVID-19. DCNNC-II represents using VGG-16 model (Sitaula & Hossain, 2021) with DSC to diagnose COVID-19. DCNNC-III represents using ResNet-18 model (Ayyachamy et al., 2019) with DSC to diagnose COVID-19. Fig. 2, Fig. 3 and Fig. 4 show the structures of DCNNC-I, DCNNC-II, and DCNNC-III. Here, Conv represents the depthwise separable convolutional layer. Pooling represents the pooling layer. BN represents batch normalization. ReLu represents the rectified linear unit. And FL represents the fully connected layer.

Here, the size of input images is 224×224. The output of feature extraction is 32×10×10. Because BBO has the characteristics of fast convergence and good at solving high-dimensional multi-objective optimization. In this model, we used biogeography-based optimization (BBO) to optimize the value of hyperparameters (convolution kernel size, and convolution kernel stride size) in two convolutional layers. We use the default value of LeNet-5 for the remaining hyperparameters.

In the BBO optimization process, the hyperparameters combination to be optimized is regarded as each island, and the hyperparameters to be optimized are regarded as the organisms on the island, and the optimal solution is selected by judging the diversity of species. In each iteration of optimization process, we use elitism to retain the optimal solution. Until the iteration termination condition is met, BBO outputs the optimal hyperparameters value combination.

Here, Table 2 shows the value of hyperparameters of DCNNC-II. The size of input images is 1024×1024. The output size of feature extraction is 512×8×8. In this model, we used BBO to optimize the hyperparameter (convolution kernel size, and convolution kernel stride size) of the first two convolutional layers.

In Fig. 4, GAP represents the global average pooling. Table 3 shows the value of hyperparameters of DCNNC-III. In this model, the size of input images is 224×224. We used BBO to optimize the hyperparameters (convolution kernel size, convolution kernel stride, pooling kernel size and pooling kernel stride) in the first convolutional layer and the first pooling layer. In the structure of DCNNC-III, each block has two convolutional layers, and the convolution kernel size is 3×3.

*Dilated convolution*

Compared with standard convolution, dilated convolution (Kim et al., 2020) can expand the receptive field of convolution kernel by changing the dilated rate. Therefore, dilated convolution can capture multi-scale information by expanding the receptive field of convolution kernel. When the dilation rate is 1, the receptive field...
size of dilated convolution is the same as that of standard convolution. When the dilation rate is greater than 1, dilated convolution can obtain larger receptive field size and capture richer image information than standard convolution (Sooksatra et al., 2020). Calculate the receptive field size of a dilated convolution follows the following formula.

\[ U = F_C + (F_C - 1)(R - 1) \]  

Here, \( U \) represents the receptive filed size of dilated convolution. \( R \) represents the size of dilation rate. When multiple dilated convolutional layers are used in the model, under the premise of setting the dilation rate reasonably, the receptive field size of convolution kernel should be calculated layer by layer according to the structure of the model.

In Section 3.4, we discussed the other novel framework – DDCNN and two methods (DDCNNC-I, DDCNNC-II) to diagnose COVID-19 based on chest X-ray images.

**Improvement II: a novel framework – DDCNN**

In this section, our team proposed a novel framework DDCNN – Dilated and Depthwise separable Convolutional Neural Network. And we applied this framework to two CNN models (VGG-16, ResNet-18) and
proposed two methods (DDCNC-I, DDCNC-II) to diagnose COVID-19 based on chest X-ray images. Compared with DCNN framework, DDCNN framework added the dilated convolution. Which has a large receptive field than DSC, and can catch more information features.

In our methods, DDCNC-I represents using VGG-16 model with dilated convolution and DSC to diagnose COVID-19. DDCNC-II represents using ResNet-18 model with dilated convolution and DSC to diagnose COVID-19. Fig. 5 and Fig. 6 show the structures of DDCNC-I and DDCNC-II.

The structure of DDCNC-I as shown in Fig. 5. We set the first two convolutional layers in DDCNC-I to dilated convolutional layers, and set all remaining convolutional layers to depthwise separable convolutional layers. We expand the receptive field by continuously using the dilated convolutional layer. In this model, we used BBO to optimize the hyperparameter (convolution kernel size, and convolution kernel stride size) of the first two convolutional layers. Here, through trial and error, we determined the dilated rate of Conv1 layer is 2, and the dilated rate of Conv2 layer is 4.

The structure of DDCNC-II as shown in Fig. 6. We used BBO to optimize the hyperparameters (convolution kernel size, convolution kernel stride, pooling kernel size and pooling kernel stride) in the first convolutional layer and the first pooling layer. Here, GAP represents the global average pooling. We set the first convolutional layer in DDCNC-II to dilated convolutional layer to obtain more information, and set all remaining convolutional layers to depthwise separable convolutional layers to reduce the number of parameters. Here, \(\frac{3}{3}\) represents the two sizes of convolution kernel in a block.

**Measure**

To improve the effectiveness and reliability of the experimental results, 10-fold cross validation Chavez & Kiefer (2021) is introduced. The specific steps are as follows: the experimental dataset is divided into 10 subsets, numbered 1–10 Wu (2020). Firstly, select the subset numbered 1 and numbered 2 as the test set, and the remaining 8 subsets as the training set. Combined them into a new experimental dataset numbered I. Secondly, select the subset numbered 2 and numbered 3 as the test set, and the remaining 8 subsets as the training set. Combined them into a new experimental dataset numbered II. Perform the same operations in sequence until the subsets numbered 10 and numbered 1 are

### Table 3

The value of hyperparameters of DDCNC-III.

|            | In Channels | Out Channels | Kernel Size | Stride | Padding |
|------------|-------------|--------------|-------------|--------|---------|
| Conv 1     | 3           | 64           | 4           | 6      | 1       |
| Pooling1   | 64          | 64           | 4           | 3      | 1       |
| Block1     | 64          | 64           | 3           |        |         |
| Block2     | 64          | 64           | 3           |        |         |
| Block3     | 64          | 128          | 3           |        |         |
| Block4     | 128         | 128          | 3           |        |         |
| Block5     | 128         | 256          | 3           |        |         |
| Block6     | 256         | 256          | 3           |        |         |
| Block7     | 256         | 512          | 1           |        |         |
| Block8     | 512         | 512          | 1           |        |         |
| Average Pooling | 512     | 512          | 12          |        |         |
| FL         | 73728       | 3            |             |        |         |
selected as the test set, and the remaining 8 subsets are selected as the training. Combined them into a new experimental dataset numbered X. From this, we get 10 datasets, numbered 1 to X. Finally, experiments are carried out on these datasets (Huang, 2018, Zhao, 2018).

To improve the intuitiveness and contrast of experimental results, the confusion matrix is introduced, as shown in Table 6. Here, TP represents both the predicted class and the actual class are positive. FN represents that the predicted class is negative, but the actual class is positive. FP represents that the predicted class is positive, but the actual class is negative. And TN represents both the predicted class and the actual class are negative.

Besides, we defined five metrics: Accuracy, Precision, Sensitivity, Specificity and F1 score.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (6)
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (7)
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (8)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (9)
\]

\[
F1 = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (10)
\]

According to the properties of 10-fold cross validation and three categories of chest X-ray images, the number of each category in (11) is 463, and that in (12) is 4630. Therefore, an ideal confusion matrix is as follows.

\[
\alpha(T = 1, W = 10) = \begin{bmatrix}
463 & 0 & 0 \\
0 & 463 & 0 \\
0 & 0 & 463
\end{bmatrix}
\quad (11)
\]

Here, \( T \) represents the number of repetitions (\( t \) loops from 1 to \( T \)). \( W \) represents the number of folds (\( w \) loops from 1 to \( W \)). And an ideal

| Table 4 | The value of hyperparameters of DCCNC-II. |
|---------|------------------------------------------|
|         | In Channels | Out Channels | Kernel Size | Stride | Padding | Dilation |
| Conv1   | 3           | 64           | 3           | 2      | 1        | 2         |
| Conv2   | 64          | 128          | 1           | 2      | 1        | 4         |
| Pooling1| 128         | 128          | 2           | 2      | 0        |           |
| Conv3   | 128         | 256          | 3           | 1      | 1        | 1         |
| Conv4   | 256         | 512          | 3           | 1      | 1        | 1         |
| Pooling2| 512         | 512          | 2           | 1      | 1        |           |
| Conv5   | 512         | 512          | 3           | 1      | 1        | 1         |
| Conv6   | 512         | 512          | 3           | 1      | 1        | 1         |
| Conv7   | 512         | 512          | 3           | 1      | 1        | 1         |
| Pooling3| 512         | 512          | 2           | 2      | 0        |           |
| Conv8   | 512         | 512          | 3           | 1      | 1        | 1         |
| Conv9   | 512         | 512          | 3           | 1      | 1        | 1         |
| Conv10  | 512         | 512          | 3           | 1      | 1        | 1         |
| Pooling4| 512         | 512          | 2           | 2      | 0        |           |
| Conv11  | 512         | 512          | 3           | 1      | 1        | 1         |
| Conv12  | 512         | 512          | 3           | 1      | 1        | 1         |
| Conv13  | 512         | 512          | 3           | 1      | 1        | 1         |
| Pooling5| 512         | 512          | 2           | 2      | 0        |           |
| FL1     | 32768       | 4096         | 4096        |       |           |           |
| FL2     | 4096        | 4096         | 4096        |       |           |           |
| FL3     | 4096        | 3            | 3           |       |           |           |
Fig. 6. The structure of DDCNNC-II.

(a) A sample of the original chest X-ray images
(b) The Grad-CAM result of (a)

Fig. 7. A sample of Grad-CAM of a chest X-ray images.

10 runs 10-fold cross validation is shown as follows.

$$\alpha(T = 10, W = 10) = \begin{bmatrix} 4630 & 0 & 0 \\ 0 & 4630 & 0 \\ 0 & 0 & 4630 \end{bmatrix}$$

$$w = 1, 2, 3, \ldots, W$$

(12)

To improve the visibility of model performance, Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2020) is introduced. By generating Grad-CAM (Panwar et al., 2020) results for different models, we can analyze the experimental results more intuitively. Figure 7 shows the Grad-CAM result of the original chest X-ray image (a). Here, the high brightness areas in the Grad-CAM result is the main reference for the model output. The higher the color brightness, the greater the reference weight.

| Table 5 | The value of hyperparameters of DCCNC-III. |
|---------|------------------------------------------|
|         | In Channels | Out Channels | Kernel Size | Stride | Padding | Dilation |
| Conv 1  | 3           | 64           | 4           | 5      | 3       | 2        |
| Pooling1| 64          | 64           | 5           | 5      | 1       | 1        |
| Block1  | 64          | 64           | $\frac{3}{3}$ | 1      | 1       | 1        |
| Block2  | 64          | 64           | $\frac{3}{3}$ | 1      | 1       | 1        |
| Block3  | 64          | 128          | $\frac{3}{3}$ | 1      | 1       | 1        |
| Block4  | 128         | 128          | $\frac{3}{3}$ | 1      | 1       | 1        |
| Block5  | 128         | 256          | $\frac{3}{3}$ | 1      | 1       | 1        |
| Block6  | 256         | 256          | $\frac{3}{3}$ | 1      | 1       | 1        |
| Block7  | 256         | 512          | $\frac{3}{3}$ | 1      | 1       | 1        |
| Block 8 | 512         | 512          | $\frac{3}{3}$ | 1      | 1       | 1        |
| Average Pooling | 512 | 512 | 12 |
| FL      | 73728       | 3            |             |        |         |          |
Table 2
The value of hyperparameters of DCNNC-II.

| In Channels | Out Channels | Kernel Size | Stride | Padding |
|-------------|--------------|-------------|--------|---------|
| Conv1       | 3            | 64          | 3      | 2       | 1       |
| Conv2       | 64           | 128         | 1      | 2       | 1       |
| Pooling1    | 128          | 128         | 2      | 2       | 0       |
| Conv3       | 128          | 256         | 3      | 1       | 1       |
| Conv4       | 256          | 512         | 3      | 1       | 1       |
| Pooling2    | 512          | 512         | 2      | 2       | 0       |
| Conv5       | 512          | 512         | 3      | 1       | 1       |
| Conv6       | 512          | 512         | 3      | 1       | 1       |
| Conv7       | 512          | 512         | 3      | 1       | 1       |
| Pooling3    | 512          | 512         | 2      | 2       | 0       |
| Conv8       | 512          | 512         | 3      | 1       | 1       |
| Conv9       | 512          | 512         | 3      | 1       | 1       |
| Conv10      | 512          | 512         | 3      | 1       | 1       |
| Pooling4    | 512          | 512         | 2      | 2       | 0       |
| Conv11      | 512          | 512         | 3      | 1       | 1       |
| Conv12      | 512          | 512         | 3      | 1       | 1       |
| Conv13      | 512          | 512         | 3      | 1       | 1       |
| Pooling5    | 512          | 512         | 2      | 2       | 0       |
| FL1         | 32768        | 4096        |        |         |
| FL2         | 4096         | 4096        |        |         |
| FL3         | 4096         | 3           |        |         |

Table 6
Confusion Matrix.

| Actual Class | Positive | Negative |
|--------------|----------|----------|
| Confusion Matrix | Predicted Class | Positive | Negative |
| TP           | FN       | FP       | TN       |

Experiment results and discussions

In this part, we compared, analyzed and discussed the experimental results. And we also compared our methods with two state-of-the-art methods.

Confusion matrix of our methods

Confusion matrix of DCNNC-I, DCNNC-II and DCNNC-III

Table 7 shows the four confusion matrix metrics of three methods. Here, COV represents the chest X-ray images of COVID-19. Nor represents the chest X-ray images of Normal. And Pne represents the test X-ray image of Pneumonia.

As can be seen from Table 7, the highest accuracy is DCNNC-II, followed by DCNNC-III, the lowest is DCNNC-I. Therefore, the complex model structure is better at chest X-ray image detection than simple model structure. DCNNC-I and DCNNC-II have the highest precision for COV and DCNNC-III has the highest precision for Pne. All three methods have the highest sensitivity to COV. DCNNC-I and DCNNC-II have the highest specificity for COV, DCNNC-III has the highest specificity for Pne. Here, the DCNNC-III structure is mainly composed of residual blocks which does not show obvious advantages than the structure without residual blocks.

Confusion matrix of DDCNNC-I and DDCNNC-II

Table 8 shows the four confusion matrix metrics of two methods. Here, COV represents the chest X-ray images of COVID-19. Nor represents the chest X-ray images of Normal. And Pne represents the test X-ray image of Pneumonia.

As can be seen from Table 8, the highest accuracy is DDCNNC-I, followed by DDCNNC-II. DDCNNC-I has the highest precision for COV and DDCNNC-II has the highest precision for Nor. Both of them have the highest sensitivity to COV. And DDCNNC-I has the highest specificity for COV. DDCNNC-II has the highest specificity for Nor. Here, the detection effort of DDCNNC-I is better than DDCNNC-II. We believe the reason is DDCNNC-I used dilated convolutional layers continuously. DDCNNC-II only has one dilated convolutional layer, which cannot fully demonstrate the advantages of dilated convolution.

Statistical results

This section listed the overall accuracy (OA) and running time of 10 runs 10-fold cross validation. Detailed data are shown in Table 9, Table 10 and Table 11.

It can be seen from Table 10 that both DCNNC-II and DDCNNC-I are using VGG-16 as the backbone. The OA of DCNNC-II is 0.37± 0.12% lower than that of DDCNN-I, the running time of the DCNNC-II is 7.1±3.95s more than that of DDCNN-I. Our team believes that the reason for the OA of DDCNN-I higher than that of DCNNC-II is DDCNN-I used dilated convolution and DSC.

It can be seen from Table 11 that both DCNNC-III and DDCNNC-II are using ResNet-18 as the backbone. The OA of DCNNC-III is 0.3± 0.24% higher than that of DDCNNC-II. The running time of DCNNC-III is 3.5±0.92s lower than that of DDCNNC-II. Our team believes that the reason for the OA of DDCNN-III higher than that of DDCNNC-II is DDCNN-II using dilated convolution and DSC. However, due to DDCNNC-II using only one dilated convolutional layer, the advantages of dilated convolution cannot be fully utilized.

Grad-CAM

In this section, we analyzed and discussed the Grad-CAM results of five methods. which helps us understand the performance of models more intuitively. Fig. 8 shows the Grad-CAM result of DCNNC-I. The high brightness areas in Grad-CAM visualization image of DCCN-I are basically concentrated on the chest area, some of the high brightness areas are distributed outside the human body.

Fig. 9 shows the Grad-CAM results of DCNNC-II and DDCNNC-I. Since the OA of DDCNNC-I is higher than that of DDCNNC-II, the high brightness color areas in the Grad-CAM result of DDCNNC-II not only concentrated in the chest, but also on the shoulders on both sides. However, the high brightness color areas in the Grad-CAM result of DDCNNC-I is basically the same as that in DCNNC-II. But DDCNNC-I has more bright areas and higher brightness. Therefore, those bright areas of DDCNN-I have higher weight than that of DDCNN-II. Fig. 9 shows Grad-CAM results of the other four methods. Since DCNNC-II and DDCNNC-I used the same structure with different convolution kernel, the original image of them is the same. And the same reason is why DCNNC-III and DDCNNC-II used the same original image. Since the OA of DCNNC-III is higher than that of DDCNNC-II, the Grad-CAM result of DCNNC-III has dense high brightness colors in the chest area. By analyzing the Grad-CAM results of the above two methods, our team believes that the main

DDCNNC-I used dilated convolutional layers continuously. DDCNNC-II only has one dilated convolutional layer, which cannot fully demonstrate the advantages of dilated convolution.

Conclusion

In this paper, we proposed three different architectures of dilated convolutional neural networks. We have evaluated the performance of our proposed methods on chest X-ray images with Pneumonia, COVID-19 and Normal. The experimental results show that our proposed methods perform well in the classification of chest X-ray images.

Future work

In future work, we plan to further improve our proposed methods by incorporating more advanced techniques such as attention mechanisms and generative adversarial networks.
Table 7
Five metrics of three methods based on the DCNN framework (unit: %).

|        | Accuracy | Precision | Sensitivity | Specificity | F1       |
|--------|----------|-----------|-------------|-------------|----------|
|        |          |           | COV Nor Pne | COV Nor Pne | COV Nor Pne |
| I      | 96.03    | 92.81     | 93.46       | 94.54       | 94.69    |
| II     | 97.01    | 94.49     | 94.77       | 94.93       | 94.93    |
| III    | 95.41    | 94.24     | 96.20       | 90.18       | 90.18    |

Table 8
Five metrics of DDCNNC-I and DDCNNC-II (unit: %).

|        | Accuracy | Precision | Sensitivity | Specificity | F1       |
|--------|----------|-----------|-------------|-------------|----------|
|        |          |           | COV Nor Pne | COV Nor Pne | COV Nor Pne |
| I      | 97.09    | 94.82     | 95.72       | 95.34       | 95.21    |
| II     | 95.15    | 93.62     | 95.95       | 90.70       | 95.21    |

Fig. 9. The Grad-CAM results of the other four methods.

(c) The Grad-CAM result of (a)
(d) The Grad-CAM result of (a)
(e) The Grad-CAM result of (b)
(f) The Grad-CAM result of (b)

Table 9
Statistical analysis on the overall accuracy and run time of DCNNC-I.

|        | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Run 6 | Run 7 | Run 8 | Run 9 | Run 10 | Average |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|---------|
| DCNNC-I| 91.23 | 91.92 | 89.87 | 91.25 | 90.06 | 91.15 | 91.46 | 90.57 | 90.76 | 92.46  | 91.07±0.79 |
| Time   | 1335  | 1326  | 1325  | 1324  | 1334  | 1338  | 1331  | 1327  | 1329  | 1331   | 1330±4.62 |

Comparing to State-of-the-art Approaches

In this section, we compared our methods with state-of-the-art methods. We selected VGG-16 (Shah et al., 2020), ResNet-18 (Rahman et al., 2021), and DDCNNC-II is DCNNC-III and DDCNNC-II used DSC in the ResNet-18.
Table 10
Statistical analysis on the overall accuracy and run time of two methods (i).

| Run | DCNNC-II OA | DCNNC-I OA | DCNNC-II Time | DCNNC-I Time |
|-----|-------------|------------|---------------|--------------|
| 1   | 91.11       | 92.60      | 1376          | 1357         |
| 2   | 93.15       | 93.80      | 1358          | 1358         |
| 3   | 93.11       | 92.17      | 1364          | 1358         |
| 4   | 93.16       | 93.94      | 1362          | 1358         |
| 5   | 93.03       | 93.93      | 1369          | 1360         |
| 6   | 94.07       | 94.16      | 1363          | 1358         |
| 7   | 93.02       | 93.99      | 1364          | 1357         |
| 8   | 94.06       | 93.05      | 1367          | 1359         |
| 9   | 94.05       | 94.55      | 1368          | 1359         |
| 10  | 93.85       | 94.14      | 1363          | 1359         |
| Average | 93.26± 0.88 | 93.63± 0.76 | 1365.4± 4.90 | 1358.3± 0.95 |

Table 11
Statistical analysis on the overall accuracies and run time of two methods (ii).

| Run | DCNNC-III OA | DCNNC-II OA | DCNNC-III Time | DCNNC-II Time |
|-----|--------------|-------------|----------------|---------------|
| 1   | 93.17        | 93.01       | 1507           | 1510          |
| 2   | 93.05        | 92.98       | 1506           | 1508          |
| 3   | 93.55        | 93.66       | 1507           | 1509          |
| 4   | 91.72        | 92.17       | 1507           | 1511          |
| 5   | 92.92        | 91.08       | 1507           | 1510          |
| 6   | 92.76        | 92.25       | 1505           | 1513          |
| 7   | 92.66        | 92.16       | 1510           | 1511          |
| 8   | 92.36        | 92.03       | 1508           | 1510          |
| 9   | 93.15        | 93.12       | 1508           | 1511          |
| 10  | 93.12        | 93.08       | 1510           | 1517          |
| Average | 92.85± 0.51 | 92.55± 0.75 | 1507.5± 1.58  | 1511± 2.50    |

Table 12
Comparison to state-of-the-art methods.

| Approach   | OA       | Time      |
|------------|----------|-----------|
| VGG-16     | 88.06± 1.69 | 1520.5± 11.56 |
| ResNet-18  | 90.81± 0.72 | 1582.6± 10.17 |
| DCNNC-I (Ours) | 91.07± 0.79 | 1330± 4.62 |
| DCNNC-II (Ours) | 93.26± 0.88 | 1365.4± 4.90 |
| DDCNNC-I (Ours) | 93.63± 0.76 | 1358.3± 0.95 |
| DDCNNC-II (Ours) | 92.85± 0.51 | 1507.5± 1.58 |
| DDCNNC-II (Ours) | 92.55± 0.75 | 1511± 2.50 |

In this paper, our team has proposed two novel frameworks – DCNN, DDCNN. Three methods based on DCNN framework – DCNNC-I, DCNNC-II, and DCNNC-III to diagnose COVID-19 based on chest X-ray images. Two methods based on DDCNN framework – DDCNNC-I and DDCNNC-II to diagnose COVID-19 based on chest X-ray images. Experiment results showed that the methods proposed by our team not only improve the OA, but also reduce the running time. In all methods, DDCNN-I has the highest OA. And DCCNC-III has the shortest running time.

In future research, we will continue to work on artificial intelligence aided diagnosis system. And we will try to use more deep learning techniques to propose more efficient neural network models. Our research directions are as follows: (i) we will try to use more dilated convolutional layers in ResNet-18. (ii) we will try to build a convolutional neural network that has better diagnosis performance to detect multiclassification problems.

Declaration of Competing Interest

None.

Funding

Research on Optimization Technology of Pathological Image Detection Based on Deep Learning, Key Projects of Science and Technology Department of Henan Province, CN (212102310084).

Ethical approval

"All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Appendix

Table 13 and Table 14
Table 13
Variable definition table.

| Variable Name | Variable Meaning |
|---------------|------------------|
| $A_C$         | Number of parameters in convolution operation |
| $A_{SC}$      | Number of parameters when standard convolution performs convolution operation. |
| $A_{DSC}$     | Number of parameters when DSC performs convolution operation. |
| $F_C$         | Size of convolution kernel. |
| $F_{SC}$      | Size of standard convolution kernel. |
| $F_{DSC}$     | Size of depthwise separable convolution kernel. |
| $I_C$         | Input image size. |
| $I_{SC}$      | Number of channels of input images. |
| $N_C$         | Number of convolution kernel. |
| $O_C$         | Output of convolution operation. |
| $P_C$         | Size of padding. |
| $R$           | Size of dilation rate. |
| $N_T$         | Stride size of convolution kernel |
| $T$           | Total number of runs (each run carries out a $W$-fold cross validation). |
| $i$           | Run index (each run carries out a $W$-fold cross validation). |
| $U$           | Size of receptive field of dilated convolution. |
| $W$           | Number of folds for cross validation. |
| $w$           | Index of fold used as test set. |
| $I$           | Dataset of 10-fold cross validation with number 1. |
| $II$          | Dataset of 10-fold cross validation with number 2. |
| $X$           | Dataset of 10-fold cross validation with number 10. |
| $s$           | Confusion matrix. |

Table 14
Abbreviation table.

| Abbreviation | Full Definition |
|--------------|-----------------|
| BBO          | Biogeography-based optimization. |
| CNN          | Convolutional Neural Network. |
| CCT          | Chest Computed Tomography. |
| DCNN         | Depthwise separable convolutional neural network. |
| DCNCC        | Depthwise separable convolutional neural network for COVID-19. |
| DDCNNC       | Dilated and Depthwise separable Convolutional Neural Network. |
| DSC          | Depthwise Separable Convolution. |
| FL           | Fully Connected Layer. |
| FN           | False Negative. |
| FP           | False Positive. |
| Grad-CAM     | Gradient-weighted Class Activation Mapping. |
| Nor          | The chest X-ray images of Normal. |
| OA           | The overall accuracy. |
| Pne          | The chest X-ray images of Pneumonia. |
| ReLU         | Rectified Linear Unit. |
| TN           | True Negative. |
| TP           | True Positive. |
| WHO          | World Health Organization. |

References

Toapanta, A., et al. (2021). A review of autonomous vehicle technology and its use for the COVID-19 contingency. *Ciencia E Ingenieria*, 42(1), 43–51.
Satapathy, S. C. (2021). Covid-19 diagnosis via DenseNet and optimization of transfer learning setting. *Cognitive Computation* DOI: 10.1007/s12559-020-09776-8.
Marquez, R., Tolosa, L., & Cells, M. T. (2021). Understanding COVID-19 effect on the US supply chain of strategic products: important factors, current situation, and future perspective. *Ciencia E Ingenieria*, 42(1), 53–62.
Satapathy, S. C. (2021). A five-layer deep convolutional neural network with stochastic pooling for chest CT-related COVID-19 diagnosis. *Machine Vision and Applications*, 32, Lerner, D., et al. (2020). Study summary - randomized control trial of omega-3 fatty acid supplementation for the treatment of COVID-19 related olfactory dysfunction. *Trials*, 21(1), 2.
Kulkarni, R., et al. (2021). Anti-SARS-CoV-2 IgG antibody response among Indian COVID-19 patients using beta-propiolactone-inactivated, whole virus-based indirect ELISA. *Journal of Virological Methods*, 287, 6.
Mendoza, P., Lorenzi, J. C. C., & Gaebler, C. (2021). COVID-19 antibody development fueled by HIV-1 broadly neutralizing antibody research. *Current Opinion in HIV and AIDS*, 16(1), 25–35.
Wang, S.-H. (2021). COVID-19 classification by FGNet with deep feature fusion from graph convolutional network and convolutional neural network. *Information Fusion*, 67, 208–229.
Wang, S.-H. (2021). COVID-19 classification by CSGNet with deep feature fusion using transfer learning and discriminant correlation analysis. *Information Fusion*, 68, 131–148.
Fernandes, S. AVNC (2021). Attention-based VGG-style network for COVID-19 diagnosis by CRAM. *IEEE Sensors Journal*. 10.1109/JSEN.2021.3062442.
Pham, T. D. (2021). Classification of COVID-19 chest X-rays with deep learning: new models or fine tuning? *Health Information Science and Systems*, 9(1), 11.
Ismael, A. M., & Sengar, A. (2020). The investigation of mutiresolution approaches for chest X-ray image based COVID-19 detection. *Health Information Science and Systems*, 8(1).
Shanthi, T., & Subennian, R. S. (2019). Modified AlexNet architecture for classification of diabetic retinopathy images. *Computers & Electrical Engineering*, 76, 56–64.
Li, X., et al. (2019). Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection. *Artificial Intelligence in Medicine*, 103, Article 101744.
Zhang, Y. D. (2020). A seven-layer convolutional neural network for chest CT based COVID-19 diagnosis using stochastic pooling. *IEEE Sensors Journal*, 1–1 DOI: 10.1109/JSEN.2020.3025855.
Lin, T.Y., et al., Focal Loss for Dense Object Detection. 2017. *PP*: p. 2999-3007.
Mercan, C., et al. From Point Annotations to Epithelial Cell Detection in Breast Cancer Histopathology using RetinaNet. 2019.
Harsono, I.W., et al., Lung Nodule Detection and Classification from Thorax CT-Scan Using RetinaNet with Transfer Learning. 2020.
Zlocha, M., Dou, Q., & Glocker, B. (2019). Improving RetinaNet for CT lesion detection with dense masks from Weak RECIST Labels.. Medical image computing and computer assisted intervention – MICCAI 2019. Cham: Springer International Publishing.
COVID19 with Pneumonia and Normal Chest Xray(PA) Dataset. 2020 Available from: https://www.kaggle.com/amansulluhasrat/covid19-pneumonia-normal-chest-xray-pa-dataset.
Sahani, M., & Darh, P. K. (2019). FPGA-based deep convolutional neural network of process adaptive VMD data with online sequential BVFLN for power quality events recognition. *IEEE Transactions on Power Electronics*, 36(4), 4006–4015.
Alencastre-Miranda, M., Johnson, R. M., & Krebs, H. I. (2021). Convolutional neural networks and transfer learning for quality inspection of different sugarcane varieties. IEEE Transactions on Industrial Informatics, 17(2), 787–794.

Garnett, R. L., et al. (2021). Event selection and background rejection in time projection chambers using convolutional neural networks and a specific application to the ADEPT gamma-ray polarimeter mission. Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment, 987, 9.

Lv, Y.-D. (2018). Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. Journal of Medical Systems, 42(1).

Pan, C. (2018). Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. Journal of Computational Science, 28, 1–10.

Vorugunti, C. S., et al. (2020). DeepFuseOSV: online signature verification using hybrid feature fusion and depthwise separable convolution neural network architecture. IET Biometrics, 9(6), 259–268.

Tsai, T. H., & Lee, Y. C. (2020). A light-weight neural network for wafer map classification based on data augmentation. IEEE Transactions on Semiconductor Manufacturing, 33(4), 663–672.

Lee, S. K., et al. (2021). Prediction of tire pattern noise in early design stage based on convolutional neural network. Applied Acoustics, 172, 18.

Sitaula, C., & Hossain, M. B. (2021). Attention-based VGG-16 model for COVID-19 chest X-ray image classification. Applied Intelligence, 14.

Ayyachamy, S., et al. (2019). Medical image retrieval using ResNet-18. In P. H. Chen, & P. R. Bak (Eds.), Medical imaging 2019: Imaging informatics for healthcare, research, and applications. Bellingham: Spie-Int Soc Optical Engineering Editors.

Kim, S., et al. (2020). Motion retargeting based on dilated convolutions and skeleton-specific loss functions. Computer Graphics Forum, 39(2), 497–507.

Sookatra, S., et al. (2020). Redesigned skip-network for crowd counting with dilated convolution and backward connection. Journal of Imaging, 6(5), 16.

Chavez, J. M., & Kiefer, B. (2021). MATCOR, a program for the cross-validation of material properties between databases. Computational Materials Science, 187.

Wu, X. (2020). Diagnosis of COVID-19 by Wavelet Renyi entropy and three-segment bio-geography-based optimization. International Journal of Computational Intelligence Systems, 13(1), 1332–1344.

Huang, C. (2018). Multiple Sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. Frontiers in Neuroscience, 12.

Zhao, G. (2018). Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. Journal of Real-Time Image Processing, 15(3), 631–642.

Selvaraju, R. R., et al. (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision, 128(2), 336–359.

Panwar, H., et al. (2020). A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos Solitons & Fractals, 140, 12.

Shah, U., et al., An Efficient Method to Predict Pneumonia from Chest X-Rays Using Deep Learning Approach. 2020.

Rahman, T., et al. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Applied Sciences, 10, 3233.