Journal of Digital Imaging (2023) 36:988–1000
https://doi.org/10.1007/s10278-023-00791-3

DPDH-CapNet: A Novel Lightweight Capsule Network with Non-routing for COVID-19 Diagnosis Using X-ray Images

Jianjun Yuan[1] · Fujun Wu[1] · Yuxi Li[1] · Jinyi Li[1] · Guojun Huang[1] · Quanyong Huang[2]

Received: 31 August 2022 / Revised: 26 January 2023 / Accepted: 29 January 2023 / Published online: 22 February 2023

Abstract

COVID-19 has claimed millions of lives since its outbreak in December 2019, and the damage continues, so it is urgent to develop new technologies to aid its diagnosis. However, the state-of-the-art deep learning methods often rely on large-scale labeled data, limiting their clinical application in COVID-19 identification. Recently, capsule networks have achieved highly competitive performance for COVID-19 detection, but they require expensive routing computation or traditional matrix multiplication to deal with the capsule dimensional entanglement. A more lightweight capsule network is developed to effectively address these problems, namely DPDH-CapNet, which aims to enhance the technology of automated diagnosis for COVID-19 chest X-ray images. It adopts depthwise convolution (D), point convolution (P), and dilated convolution (D) to construct a new feature extractor, thus successfully capturing the local and global dependencies of COVID-19 pathological features. Simultaneously, it constructs the classification layer by homogeneous (H) vector capsules with an adaptive, non-iterative, and non-routing mechanism. We conduct experiments on two publicly available combined datasets, including normal, pneumonia, and COVID-19 images. With a limited number of samples, the parameters of the proposed model are reduced by 9x compared to the state-of-the-art capsule network. Moreover, our model has faster convergence speed and better generalization, and its accuracy, precision, recall, and F-measure are improved to 97.99%, 98.05%, 98.02%, and 98.03%, respectively. In addition, experimental results demonstrate that, contrary to the transfer learning method, the proposed model does not require pre-training and a large number of training samples.

Keywords COVID-19 · Capsule networks · Chest X-ray images · Homogeneous vector capsules

Introduction

Coronavirus disease (COVID-19) has rapidly spread across the globe since December 2019, claiming tens of thousands of lives. Three years have passed, the repeated epidemics still severely influence people's work, study, and life. It is urgent to develop new detection technologies. COVID-19 has similarities with other pneumonia diseases, such as severe acute respiratory syndrome (SARS) or viral pneumonia (VP), which requires a large number of professional radiologists, thus significantly increasing the pressure on hospital emergency departments and emergency centers. Compared with CT imaging, chest X-ray (CXR) has a shorter diagnostic time and lower cost. Thus more and more researchers attempt to introduce deep learning (DL), especially convolutional neural network (CNN), to improve COVID-19 detection efficiency and accuracy on chest X-ray images [1–4].

The current state-of-the-art CNN models are highly complex in structure, which determines their "data-starved" property. This heavily limits their application in COVID-19 detection. This problem is effectively addressed by using transfer learning, that is, fine-tunes models trained on large-scale data using COVID-19 datasets. To solve the problem of insufficient COVID-19 samples, Loey et al. [5] introduced Generative Adversarial Network (GAN) based on the classical DL framework AlexNet, the effect of which is better than other transfer learning methods, such as GoogleNet and

✉ Jianjun Yuan
jianjuny@sina.com

✉ Quanyong Huang
huangguojun@swu.edu.cn

[1] College of Artificial Intelligence, Southwest University, Chongqing 40075, China

[2] College of Machinery and Automation, Wuhan University of Science and Technology, Heping Avenue No. 947, Wuhan, Hubei Province 430091, China

ResNet18. Abbas et al. [6] adopted VGGNet to design a decomposition, transfer, and synthesis method for classifying CXR images into three categories: normal, COVID-19, and SARS. Its performance outperformed the traditional VGG19 pre-trained model. In addition, many research attempts to improve ResNet to obtain better classification performance, such as the paper [7] combined ResNet with the feature pyramid network. Unlike this, [8] used multiple image levels to diagnose COVID-19 at the 3D CT volume level. Its detection performance is superior to that of a single 3D-Resnet. Other transfer methods include Inception [9], DenseNet [10]. Transfer learning has achieved satisfactory results in COVID-19 identification, but it cannot be the preferred method for clinical diagnosis of COVID-19 due to its complex model and high computational overhead. Therefore, some researchers proposed DL frameworks specifically for COVID-19. The COVID-net proposed by Wang et al. [11] achieved 83.5% accuracy in the classification of COVID-19, normal, pneumonia-bacterial, and pneumonia-viral. Moreover, Ozturk et al. [12] used 17 convolutional layers to design a model based on DarkNet, with an accuracy of 98.08% for binary classification and 87.02% for multi-class classification. But the performance of this type of model on multi-classification needs to be further improved. Additionally, since CNN has some potential defects, especially it cannot capture the relative positional relationship between features, Hinton et al. [13] exploited a new architecture, referred to as Capsule Network, as a powerful alternative to CNN. This structure uses capsules (that is, vectors containing feature information) to build capsule layers, and effectively avoids the loss of high-level feature information through routing mechanisms, demonstrating certain advantages in medical image processing [14, 15]. Inspired by this, Afshar et al. [16] proposed a capsule network for identifying COVID-19 using X-ray images, named COVID-CAPS, and achieved an accuracy of 95.7% and an area under the curve (AUC) of 0.97. Similarly, Toraman et al. [17] proposed an artificial neural network approach to detect COVID-19 disease. On their basis, Fudong Li et al. [2] recently designed a new capsule network using multi-head attention routing and obtained the optimal effect on COVID-19 chest X-ray image classification. In response to the current problems in COVID-19 detection:

1. DL transfer learning models are highly complex;
2. CNNs specially designed for COVID-19 underperform in multi-classification;
3. Capsule networks rely on expensive routing calculations. A novel capsule network called DPDH-CapNet is exploited to promote COVID-19 automatical diagnosis. It utilizes depthwise convolution, point convolution, and dilated convolution to design a new feature extraction backbone while abandoning the traditional routing design between adjacent capsule layers, aiming to build a more lightweight and efficient architecture for COVID-19 diagnosis. The main contributions are as follows:

1. This paper exploits a novel capsule network. Capsule layers are constructed by homogeneous vector capsules, which tactfully avoids traditional matrix multiplication between capsule layers or expensive routing computation to deal with the entanglement of capsule dimensions. This operation makes the realizable precision of the model less dependent on the fine-tuned hyperparameters with the non-adaptive optimizer. At the same time, it can effectively capture the relationship between the lower-level capsule and the higher-level capsule to promote COVID-19 detection.
2. We adopt depthwise convolution, point convolution, and dilated convolution to design a new feature extractor. It can validly capture the local and global dependencies of feature maps with fewer parameters and lower computational overhead, further extract more abundant representation features from X-ray images, and thus improve the pathology discrimination of the model.
3. The proposed model can be trained end-to-end on a limited training dataset. Moreover, it does not require external datasets for pre-training and transfer learning. More importantly, its parameters are 29,750, reduced by 9x compared with the state-of-the-art capsule network.
4. Experimental results on the CXR dataset demonstrate that the proposed model outperforms the state-of-the-art capsule networks (shown in Fig. 1) and transfer learning methods. In addition, it also has faster convergence and better generalization.

The remainder of this paper is as follows. "Related Work'' discusses the work related to our model. "Proposed Model'' describes the proposed network architecture in detail. Experimental results and analysis are in "Experimental Preparation''. The conclusions and future work are in "Conclusion''.

Related Work

The capsule network was first proposed by Hinton et al. [13], aiming to overcome the defect of CNN ignoring relative position information. General neural networks are composed of neurons, but capsule networks are composed of capsules, which are a set of neurons [18] that can be represented by feature vectors. The capsule not only represents a specific entity type, but also describes how the entity is instantiated, such as pose, texture, deformation, and the existence of these features themselves [19].

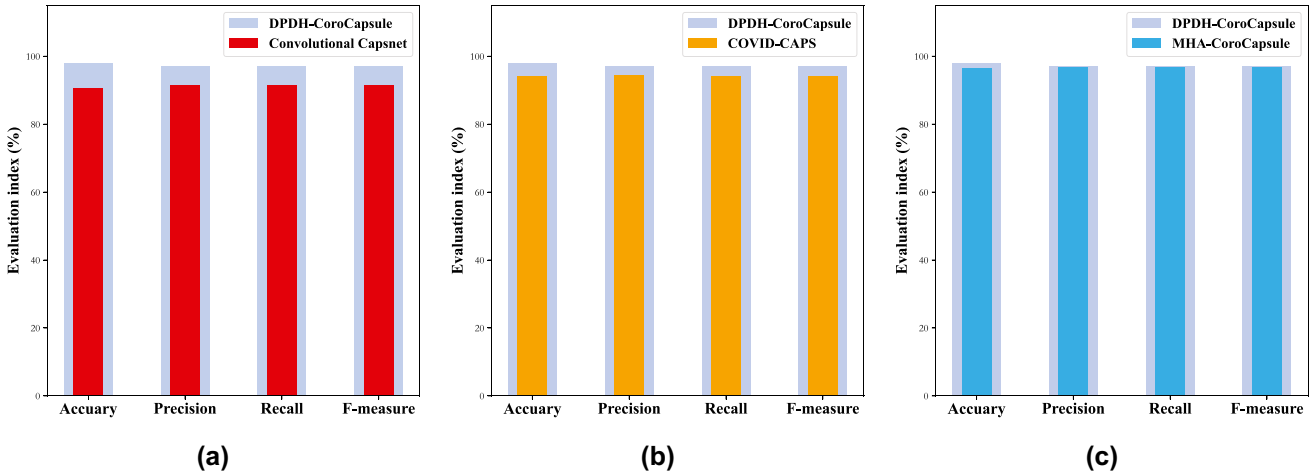


Fig. 1 Comparison of the proposed model (DPDH-CapNet) and the state-of-the-art capsule networks on different evaluation metrics. **a** Convolutional capsnet. **b** COVID-CAPS. **c** MHA-CoroCapsule

Sabour et al. [18] first adopted a dynamic routing mechanism to train the weights between different capsule layers, thus allowing the output of the current sub-capsule to be mapped to the appropriate parent capsule. It achieved 99.75% accuracy on MNIST classification. After that, Hinton et al. [20] proposed a new routing iterative mechanism (EM algorithm) by changing the sub-capsule activation method while incorporating a gaussian mixture model. Similarly, most research on capsule networks mainly focused on improving dynamic routing mechanisms. Rajasegaran et al. [21] proposed a novel dynamic routing (DeepCaps) based on 3D convolution to reduce the complexity. In addition, F. Ribeiro et al. [22] designed a routing by variational bayesian (VB) and combined it with a gaussian mixture model of the fitted transform. This method outperformed EM in terms of convergence speed, stability, and final test error. Venkataraman et al. [23] introduced another degree-centrality-based equal-variable routing. These operations further improved the performance of the capsule network. Unlike the above approaches, Choi et al. [19] (AR CapsNet) and Tsai et al. [24] used an attention mechanism to design routing to capture the relationship of adjacent capsule layers. Mazzia et al. [25] extended the research by adopting non-iterative, highly parallel self-attentive instead of dynamic routing, which conclusively reduced the parameters. Recently, Fudong Li et al. [2] confirmed the effectiveness of the multi-head attention mechanism on routing.

The routing mechanism is often iterative, parameterized, and limited to a certain extent by the entanglement of capsule dimensions, so the computational overhead is vast. We abandon the routing process and only rely on the weights learned between capsule layers in the process of backpropagation, which aims to implement the connection of the capsule layer using an adaptive optimization strategy.

Proposed Model

As shown in Fig. 2, the proposed model consists of convolutional layers and capsule layers. The convolutional layers are exploited using a new feature extractor to obtain richer feature representation. The capsule layers consist of primary and class capsule layers. They are connected with an element-wise multiplication method instead of matrix multiplication conversion or routing algorithm, thus reducing computational complexity. The advantage of this framework is that the convolutional layers extract the local feature and global feature of COVID-19 from the input chest X-ray image. The extracted features are fed into the capsule layers to instantiate different objects, thereby improving the model's discriminative ability.

Convolutional Layer

Adam Jacobi et al. [26] pointed the lung consolidation and ground glass opacities features of COVID-19 tend to be diffuse (e.g., bilateral lower lung) or localized (e.g., left upper lung) on chest X-ray images, and the diffuse trend is strengthened with increasing degree of infection. Therefore, it is crucial to capture the local and global dependencies of its pathological features. So this paper explores a new feature extraction architecture to implement this function. CNNs [27, 28] and Vision Transformers (ViTs) [29, 30] have become two mainstream frameworks in computer vision. Recently, some studies have combined CNN and ViTs to design new feature extraction backbones, aiming to subtly incorporate the advantages of both [31–34] to reduce image local redundancy (CNNs) while capturing long-range correlations (ViTs). However, such models have not yet fully escaped from the "data starvation" paradigm, especially in

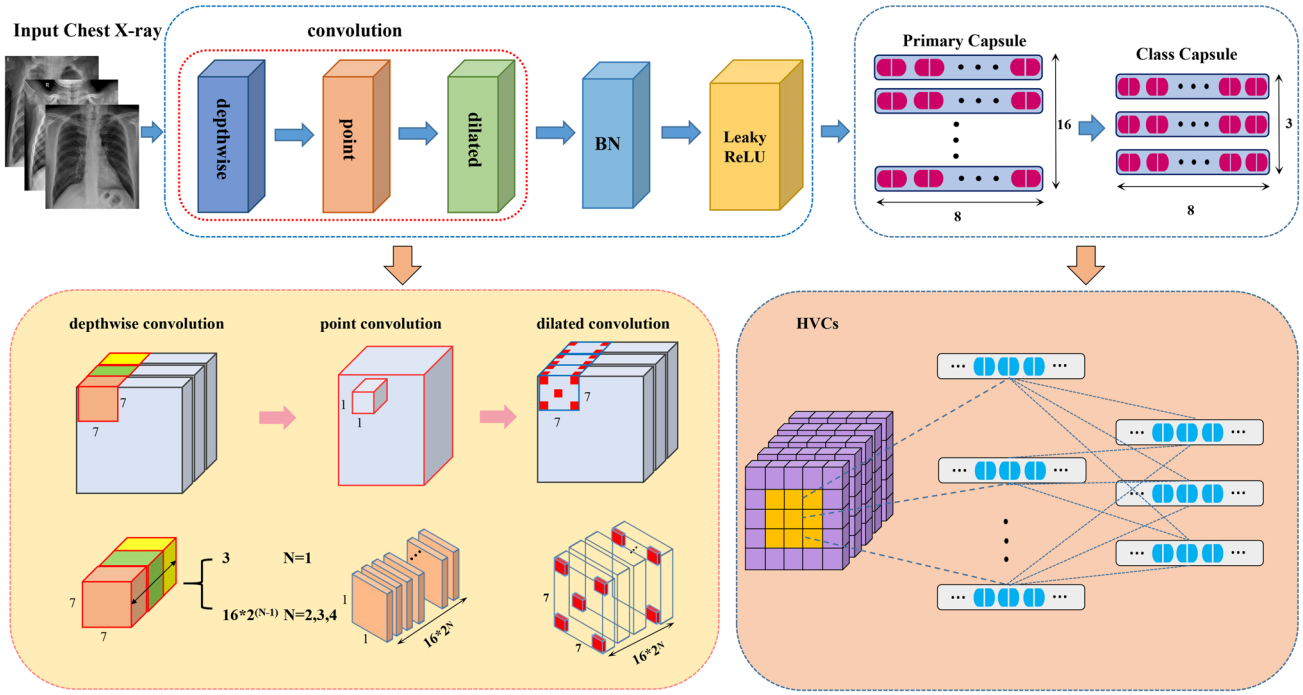


Fig. 2 Schematic representation of the overall architecture of DPDH-CapNet. *N* (equal to 1, 2, 3, 4) represents the number of convolutional layers. BN is batch normalization. HVCs denotes homogeneous vector capsules

the medical field. By contrast, the proposed feature extractor can effectively break through this limitation and combine the advantages of CNN+ViTs. It consists of depthwise convolution, point convolution, and dilated convolution.

In depthwise convolution, one filter corresponds to one channel, which can be used to extract the channel features of COVID-19 chest X-ray images. The size of the point convolution kernel is $1 \times 1 \times M$, and M represents channels corresponding to the previous layer. It weights and combines the features obtained by depthwise convolution. Assuming that W and H represent the width and height of the convolution kernel, respectively. C_{in} and C_{out} are the number of input channels and output channels, respectively, then the parameters of standard convolution can be expressed as:

$$P_{st} = W \times H \times C_{in} \times C_{out} \tag{1}$$

For combined depth wise convolution and point convolution, its parameters are:

$$P_{dp} = W \times H \times C_{in} + 1 \times 1 \times C_{in} \times C_{out} \tag{2}$$

So,

$$P_{dp}/P_{st} = 1/C_{out} + 1/(H \times W) < 1 \tag{3}$$

It can be seen that compared with standard convolution, depthwise convolution and point convolution can significantly reduce parameters. In addition, we introduce injection holes based on the depthwise convolution to construct a dilated convolution, aiming to capture global information of feature maps. Assuming its kernel size is k and the number of holes is d, and $d = 3$ in this paper, then its equivalent standard convolution kernel size k' is:

$$k' = k + (k-1) \times (d-1) > k \tag{4}$$

Moreover, let RF_{i+1}, RF_i respectively be the receptive field of the current layer and the previous layer, then:

$$RF_{i+1} = RF_i + \left(k' - 1\right) \times S_i \tag{5}$$

where S_i represents the product of the stride of all previous layers (excluding this layer), it is calculated as follows:

$$S_i = \prod_{i=1}^{i} \text{Stride}_i \tag{6}$$

Equations (4) and (5) indicate that, compared to standard convolution, dilated convolution can acquire a larger receptive field with the same kernel size, thus effectively capturing contextual information. To conclude, our feature extractor has a more robust performance than standard convolution in capturing the local and long-range dependencies of COVID-19 pathological features. Further, we design 4 sets of such feature extractors, each group sets the size of all convolution kernels to 7×7. In addition, all strides of convolution operations are all set to 1. The batch normalization (BN) process and LeakyReLU activation function follow

each group of feature extractors. Finally, after each group of feature extractors, the number of channels of the obtained feature maps is 16, 32, 64, and 128 in turn.

Homogeneous Vector Capsules

The primary capsule layer is created by constructing capsule vectors using each different x and y coordinate of the feature maps, which fully considers meaningful feature combinations. The aim is to obtain different feature vector formations to instantiate the capsules. After this operation, the obtained primary capsule layer can be described as $P_{16,8}$, where $n = 16$ and $d = 8$ represent the number of primary capsules and their respective dimensions, respectively.

To avoid the problem of the overdetermined system caused by traditional matrix multiplication and expensive routing calculation overhead. A new approach is applied to map primary capsules to class capsules, namely element-wise multiplication. This operation can be described as the Eq. (7):

$$W_i \odot P_i = C_j \tag{7}$$

where $i = 0, ..., 16$. W_i represents the learnable weight corresponding to the primary capsule P_i, both have equal dimensionality. C_j (j=1, 2, 3) represents the class capsule. The primary capsule layer and the class capsule layer have the same dimensions, so they are called homogeneous vector capsules, and their visualization form can be referred to as the HVCs in Fig. 2. This method has the following advantages. On the one hand, the training weight parameters are few. The training weight parameters of each capsule are equal to the dimensionality of the capsule. However, they are the square of the dimension of the capsule for the dynamic routing method proposed by Sabour et al. [18]. On the other hand, the primary and class capsule layers have the same dimensions, which makes it flexible to model feature vectors. While the dimension of the vector must meet the perfect square in paper [20], which dramatically limits its application in COVID-19 detection.

After homogeneous operations, the class capsule layer $C_{3,8}$ has 3 capsules with 8 dimensions, where the length of the activity vector of each capsule represents the probability that each class exists, which is used to calculate the classification loss. Moreover, they also contain instantiated parameters for normal, pneumonia, and COVID-19 chest X-ray images.

Margin Loss

We adopt the margin loss L_c in capsnet as the loss function. It is calculated as follows:

$$\begin{aligned} L_c = & \sum_{k \in CN_{um}} T_k max(0, m^+ - ||\mathbf{u}_k||)^2 \\ & + \lambda(1 - T_k) max(0, ||\mathbf{u}_k|| - m^-)^2 \end{aligned} \tag{8}$$

Table 1 Chest X-ray image's distribution for normal, pneumonia, and COVID-19

Dataset	Normal	Pneumonia	COVID-19	Total
dataset-1	350	350	294	994
dataset-2	1341	1345	1200	3886

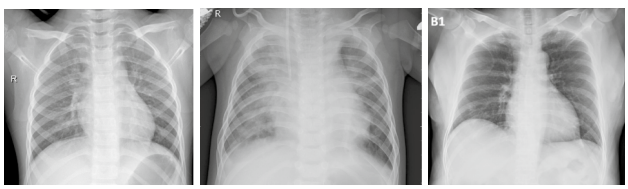
where T_k is the sample class label. If k class exists, $T_k = 1$. $\lambda = 0.5$ is the balance coefficient, which lowers the weights of the loss for non-existing classes. These two parameters can prevent the initial learning from shrinking the length of the activity vector of all class capsules. CN_{um} and k represent the number of classes in the dataset and the sequence number of classes, respectively. $m^+ = 0.9$ and $m^- = 0.4$ are class prediction thresholds used to control the class response value of the actual computed output. In particular, $L_c = 0$ when the prediction vector u_k of the class capsule layer is consistent with T_k.

Experiments and Analysis

Experimental Preparation

Datasets

The datasets as shown in Table 1. Each dataset contains three types of CXR images with normal, pneumonia, and COVID-19 labeled, its example images are shown in Fig. 3. COVID-19 images in dataset-1 come from the database created by Cohen [35], for a total of 294 positive chest X-ray images, each from a different patient. In addition, we also obtain 350 normal and non-COVID-19 pneumonia chest X-ray images from Mooney [36]. Dataset-2 is from the COVID-19 radiography database [37], mainly consisting of 1200 COVID-19 positive images, 1341 normal images, and 1345 viral pneumonia images. The original labeled normal, pneumonia, and COVID-19 X-ray images have varying length and width sizes. Moreover, they are all high-resolution. In such a case, all images in dataset-1 and dataset-2 are rescaled to 128 × 128 pixels. Furthermore, we employ a normalization strategy to scale the pixel values of the images from [0, 255] to [0, 1].



(a) Normal **(b)** Pneumonia **(c)** COVID-19

Fig. 3 Example images from datasets with three categories: **a** Normal, **b** Pneumonia, and **c** COVID-19

Table 2 Performance for different folds on the proposed capsnet. Dataset-1 is partitioned into 5 different 4-fold data

Folds	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)	AUC(%)
Fold-1	97.59 ± 0.56	97.67 ± 0.58	97.59 ± 0.55	97.61 ± 0.55	99.41 ± 0.07
Fold-2	97.03 ± 1.82	97.10 ± 1.80	97.11 ± 1.76	97.07 ± 1.83	99.43 ± 0.06
Fold-3	95.89 ± 0.36	95.97 ± 0.36	95.99 ± 0.44	95.95 ± 0.38	98.94 ± 0.03
Fold-4	95.65 ± 0.60	95.82 ± 0.56	95.73 ± 0.60	95.74 ± 0.57	98.95 ± 0.15

Model Training and Testing

Firstly, with the Google Colab cloud experimental environment, experiments are executed using Python 3.7, Keras 2.4.3, and TensorFlow-GPU 2.8.0. Second, the proposed model and all the compared models are implemented using the graphical processing unit (GPU) Tesla-P100-PCIE with 16 GB. In addition, the training set and test set are divided according to the ratio of 3:1. Model optimization uses the Adam with an initial learning rate of 0.001. Besides, we also apply an exponential decay function with the decay rate of 0.5 and the decay step of 15 to lower the learning rate. The batch size and epoch are set to 16 and 100, respectively.

Model Evaluation

Since some of the comparison experiments are carried out on small samples, we use the 4-fold cross-validation method on dataset-1, thus guaranteeing the reliability of the experimental results. Furthermore, we employ accuracy, precision, recall, and F-measure as evaluation metrics for all models. The calculation formulas are as follows:

1. Accuracy = (TP + TN)/(TP + FP + TN + FN)
2. Precision = TP/(TP + FP)
3. Recall = TP/(TP + FN)
4. F-measure = 2TP/(2TP + FP + FN)

where TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively. In addition, AUC, macro avg, and weighted avg are adopted as another three evaluation metrics in our experiment. AUC represents the area under the ROC. Macro avg and weighted avg are calculated as the average and weighted average of all categories corresponding to precision, recall, F-measure, and AUC, respectively.

Model Performance Test

We first use the 4-fold cross-validation to test the performance of the proposed model on dataset-1. The experiment takes COVID-19 images as positive samples. To make the experimental results more accurate, we divide dataset 1 into 5 different 4-fold data, and show the average results and std of different indicators in Table 2. In all folds, the average value of all evaluation indexes is not less than 95.65%, and the fluctuation is slight according to std. The best detection performance is obtained in fold-1, with accuracy, precision, recall, F-measure, and AUC of 97.59%, 97.67%, 97.59%, 97.61%, and 99.41%, respectively. Based on the above results, it is persuasive that the overall performance of the proposed model is excellent. More importantly, it has a powerful detection ability for COVID-19 based on the evaluation indexes precision, recall, and F-measure. In addition, Table 3 shows the identification effects for each class in fold-1. The precision of the COVID-19 class is 100.00%, which illustrates the proposed network does not misclassify normal and pneumonia samples as COVID-19 in the dataset-1. This is crucial for epidemic prevention and control of the COVID-19, because it can identify asymptomatic infected persons to the greatest extent possible. The visualizations of the confusion matrix of our model in each fold are presented in Fig. 4.

Figure 5 shows the ROC of the proposed capsule network in fold-1. The AUC of COVID-19 classification is as high as 0.9992, significantly higher than that of normal, pneumonia, which indicates that our model is highly competitive for COVID-19 classification. Moreover, we further explore its misclassification images and their prediction scores, as displayed in Fig. 6. In fact, DPDH-CapNet utilizes the length of the class capsule layer output vector (i.e., the probability of the existence of each class.) to predict the class scores. Furthermore, the length also represents the probability of the presence of each class. For the misclassified images in fold-1, we can see from the blue bars that even though the COVID-19 symptoms are not obvious, the proposed model can still learn most of the instantiated features of COVID-19.

Routing mechanisms play an essential role in connecting low-level capsules and high-level capsules. It is necessary to conduct a series of exploratory experiments to analyze

Table 3 Performance obtained from the proposed model in fold-1

	Precision(%)	Recall(%)	F-measure(%)	AUC(%)
Normal	96.67	100.00	98.31	99.65
Pneumonia	98.86	98.86	98.86	99.65
COVID-19	100.00	95.95	97.93	99.92
Macro avg	98.51	98.27	98.37	99.74
Weighted avg	98.43	98.39	98.39	99.72

Table 4 Performance comparison for different routing algorithms. The evaluation metrics are obtained from the average values on the 4-fold cross-validation. The best results are bolded

Routing Algorithm	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)	AUC(%)	Parameters
Dynamic routing (r=3)	96.88 ± 0.86	96.94 ± 0.88	96.96 ± 0.86	96.93 ± 0.87	99.55 ± 0.14	304,528
Dynamic routing (r=5)	96.98 ± 0.93	97.06 ± 0.93	97.05 ± 0.92	97.04 ± 0.93	**99.58 ± 0.10**	304,528
Dynamic routing (r=7)	96.88 ± 0.88	96.94 ± 0.88	96.96 ± 0.86	96.93 ± 0.87	99.55 ± 0.14	304,528
MHA routing (h=2)	96.38 ± 0.50	96.46 ± 0.48	96.48 ± 0.43	96.45 ± 0.45	99.41 ± 0.13	329,520
MHA routing (h=4)	95.98 ± 1.02	96.17 ± 0.99	96.01 ± 1.10	96.05 ± 1.05	99.40 ± 0.14	329,712
MHA routing (h=8)	96.58 ± 0.35	96.63 ± 0.39	96.64 ± 0.33	96.61 ± 0.36	99.42 ± 0.08	330,096
MHA routing (h=16)	96.38 ± 0.76	96.43 ± 0.78	96.48 ± 0.70	96.43 ± 0.74	99.14 ± 0.11	330,864
Attention Routing	96.98 ± 0.45	97.09 ± 0.48	97.04 ± 0.40	97.04 ± 0.43	99.52 ± 0.20	320,896
Ours	**97.08 ± 0.78**	**97.15 ± 0.83**	**97.16 ± 0.72**	**97.14 ± 0.77**	99.27 ± 0.17	**29,750**

the performance of different routings and our method in COVID-19 detection. The experiment selects the current typical routing designs, including the dynamic routing, attention routing, and multi-head attention routing. In the experiment, the iterations of dynamic routing are set to 3, 5, and 7. The headers of multi-head attention routing take 2, 4, 8, and 16, respectively. The performance comparisons are in Table 4. Our method has great performance improvement over dynamic routing, attention routing, and multi-head attention routing. Moreover, the parameters of our routing are nearly 9x lower than other routings, which further demonstrates that our routing design can better quantify the correlation between capsules. Finally, we can conclude that our routing method can effectively instantiate COVID-19 features.

Model transparency is essential when DL models are used for life-threatening COVID-19 disease detection. To confirm that the proposed DPDH-CapNet can provide the contributing regions in X-ray images, we adopt three class activation techniques to achieve the interpretation and behavioral understanding of the DPDH-CapNet, including GradCAM++ [38], LayerCAM [39], and ScoreCAM [40]. According to Fig. 7, even though the proposed model only obtains image-level labels, it can detect COVID-19 lesions, which can greatly assist doctors in fast screening and diagnosing COVID-19.

Comparison with the State-of-the-Art Networks

We compare performance of the proposed model and the state-of-the-art capsule networks and transfer learning methods. They are all designed for detecting COVID-19. Evaluation indicators are all taken from the average of 4-fold cross-validation on dataset-1. They are all designed for detecting COVID-19. It needs to note that experimental settings are kept the same, i.e., initial learning rate = 0.001, decay rate = 0.5, decay step = 15, batch size = 16, and epochs = 100. We first discuss the comparison results between the proposed model and the capsule networks (COVID-CAPS, Convolutional capsnet, and MHA-CoroCapsule). According to Table 5, it is obvious that our network has the lowest parameters, only 11.46% of COVID-CAPS (second-lowest parameters). Simultaneously, it performs best on the accuracy, precision, recall, and F-measure evaluation metrics. Further analysis reveals that convolutional capsnet uses additional reconstruction losses to encourage the capsule

Table 5 Performance comparison between our model and other state-of-the-art networks used to identify COVID-19 from X-ray images. The results come from datasets-1

Method	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)	AUC(%)	Parameters
Convolutional capsnet	90.54 ± 1.98	91.51 ± 1.74	91.40 ± 1.98	91.37 ± 1.87	96.45 ± 1.12	57,002,640
COVID-CAPS	94.06 ± 1.01	94.36 ± 1.15	94.18 ± 0.92	94.17 ± 1.00	97.87 ± 0.30	295,616
MHA-CoroCapsule	96.58 ± 0.35	96.63 ± 0.39	96.64 ± 0.33	96.61 ± 0.36	**99.42 ± 0.08**	330,096
Vgg16	92.86 ± 0.99	93.17 ± 1.06	93.13 ± 0.87	92.99 ± 0.97	98.53 ± 0.22	65,066,819
ResNet50	89.33 ± 0.94	89.91 ± 0.78	89.59 ± 0.92	89.61 ± 0.87	96.98 ± 0.22	23,593,859
InceptionV3	95.07 ± 1.48	95.29 ± 1.44	95.10 ± 1.47	95.16 ± 1.45	99.09 ± 0.27	21,808,931
DenseNet121	90.04 ± 1.33	90.60 ± 1.40	90.34 ± 1.24	90.24 ± 1.41	97.99 ± 0.37	7,040,579
MobileNet	92.66 ± 1.14	92.81 ± 1.20	92.83 ± 1.07	92.77 ± 1.13	98.35 ± 0.15	3,231,939
Ours	**97.08 ± 0.78**	**97.15 ± 0.83**	**97.16 ± 0.72**	**97.14 ± 0.77**	99.27 ± 0.17	**29,750**

Table 6 Performance comparison between the proposed DPDH-CapNet and MHA-CoroCapsule on dataset-2

Method	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)	AUC(%)
MHA-CoroCapsule	97.02	96.43	96.72	96.57	**99.69**
DPDH-CapNet	**97.43**	**97.01**	**97.01**	**97.01**	99.41

classes to encode the instantiated parameters of chest X-ray images, which dramatically increases the complexity of the model. This makes the performance of the model rely on the amount of data to some extent. COVID-CAPS and MHA-CoroCapsule respectively adopt an agreement process and a multi-head attention design routing. These two methods can efficiently capture the relationship between the bottom capsule and the top capsule. Nevertheless, our approach is still more efficient and conducive to detecting COVID-19. This is because our network can adaptively capture the local and global dependencies of COVID-19 pathology features, which makes it more sensitive to discriminate normal, pneumonia, or COVID-19 chest X-ray images, even for asymptomatic infected persons. We also display the training process of MHA-Corocapsule and DHDP-CapNet to analyze the convergence. According to Fig. 8, it is evident that the proposed model has faster convergence and better stability. Moreover, when the epochs are over 20, the loss and accuracy on the training and test sets are closer than other models, which fully proves that our model has better generalization.

According to Table 5, the proposed DPDH-CapNet and MHA-CoroCapsule models are superior to other models in various indexes on small sample dataset-1. The better capsule structure design of these two models makes it better at capturing fine-grained information and feature spatial relationships of COVID-19, so as to detect COVID-19 more effectively. However, the MHA-CoroCapsule detection performance is inferior than DPDH-CapNet. The proposed DPDH-CapNet, possibly because it has more parameters

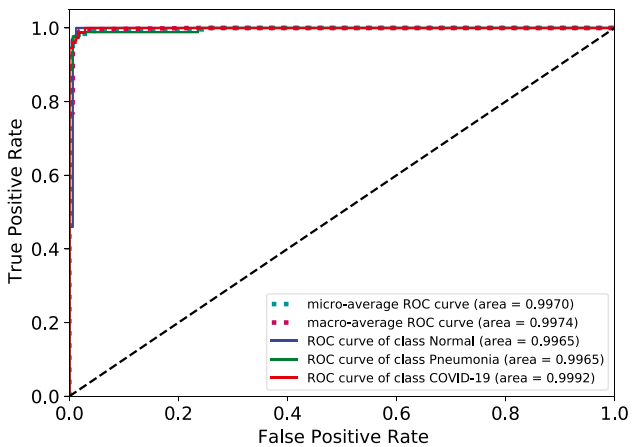


Fig. 4 ROC of the proposed capsnet for 3-class classification in fold-1

that make its performance dependent on more training dataset. So we carry out comparative experiments on dataset-1 and record the results in Table 6. Obviously, the proposed network still achieves the best effects. This fully demonstrates that global information interaction design and non-routing architecture of our model are effective in COVID-19 detection.

Additionally, due to the extremely high complexity of current state-of-the-art deep learning models, large-scale datasets are required to train the models for accurate feature extraction. However, the existing COVID-19 database is small, which seriously limits their applications in COVID-19 recognition. To address this problem, many researchers have used transfer learning strategies, i.e., pre-training on ImageNet or other large-scale datasets first and then fine-tuning the weights of the model on the training sets. Typical methods include VGG [41], ResNet [42], DenseNet [43], Inception [44], and MobileNet [45]. We compare these models with the proposed model. The aim is to further highlight the advantages of our model in identifying COVID-19. The source and target tasks of the transfer learning models involved in the experiment are similar: image classification. Theoretically, the deep learning frameworks used for comparison have a better capability of feature transfer representation after being fine-tuned by weights. Nevertheless, according to experimental results of Table 5, the proposed model still obtains the best performance.

Analysis for Model Generalization

To further validate model generalization of different models in migrating applications, we pre-train them on dataset-1 and then migrate them directly to dataset-2 for COVID-19 prediction. Experimental results are shown in Table 7. The proposed model also obtains the best effect, which strongly confirms excellent generalization of our model in COVID-19 detection. In addition, we also display the ROC of suboptimal network MHA-Corocapsule and the proposed model for 3-class classification in Fig. 9, from which we can see that our model is the most competitive for normal, pneumonia, and COVID-19 classification. Based on the above comparison results, we can conclude that the design of the proposed model is more efficient in identifying COVID-19 compared with the capsule networks designed by routing and transfer learning models. It utilizes depthwise convolution, point convolution, and dilated convolution to design feature

Table 7 Generalization comparison between the proposed network and capsule networks, pre-trained CNN models. All models are previously trained on dataset-1, then directly migrated to dataset-2 for COVID-19 classification

Method	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)	AUC(%)
Convolutional capsnet	93.81	93.81	94.67	94.23	99.27
COVID-CAPS	93.90	95.86	96.58	96.22	98.85
MHA-CoroCapsule	94.39	**97.43**	94.75	96.07	99.59
Vgg16	90.09	94.48	**98.42**	96.41	**99.68**
ResNet50	89.94	96.17	90.08	93.02	99.10
InceptionV3	93.46	96.88	95.67	96.27	99.46
DenseNet121	92.07	94.75	94.75	94.75	99.33
MobileNet	90.07	93.46	95.33	94.39	99.20
Ours	**94.42**	96.46	97.75	**97.10**	**99.68**

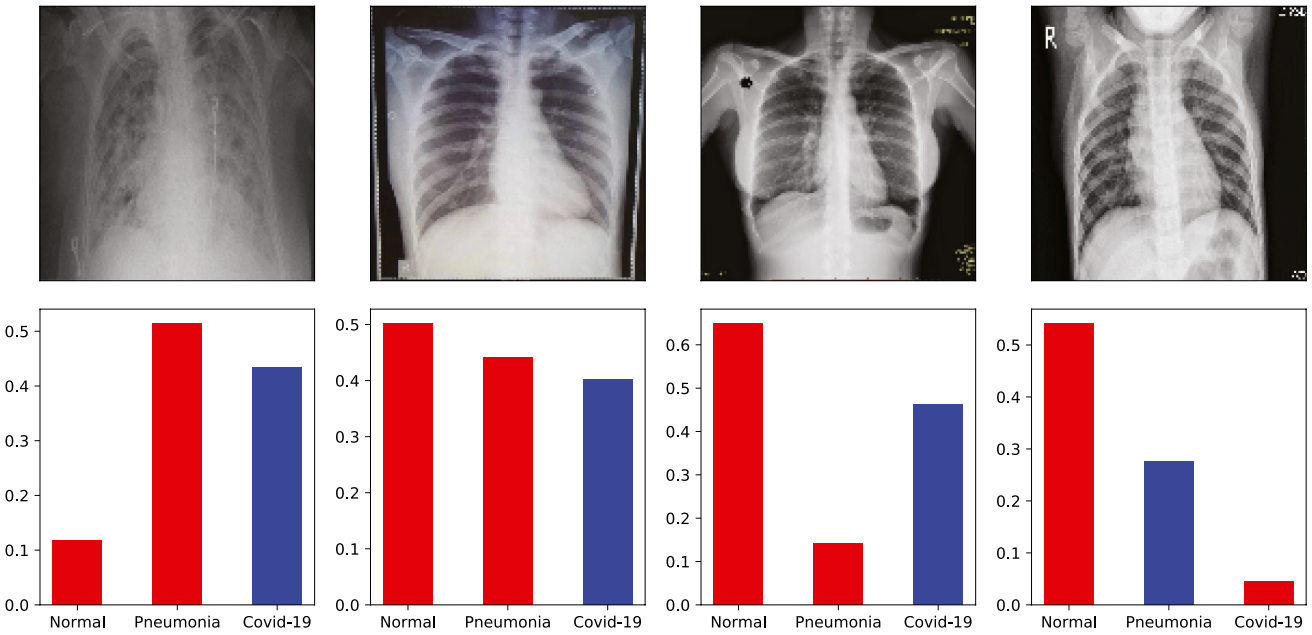


Fig. 5 Examples of the DPDH-CapNet misclassified images in fold-1. Blue bars represent correct labels and their corresponding capsule length

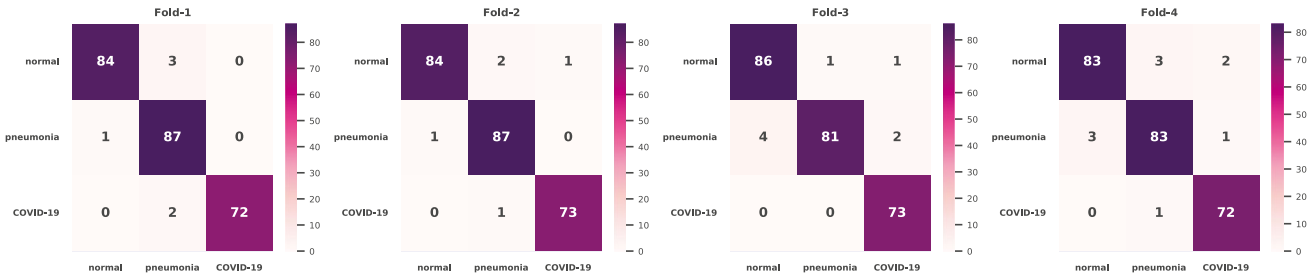


Fig. 6 4-fold confusion matrices of the multi-class classification task. From left to right are fold-1, fold-2, fold-3, and fold-4

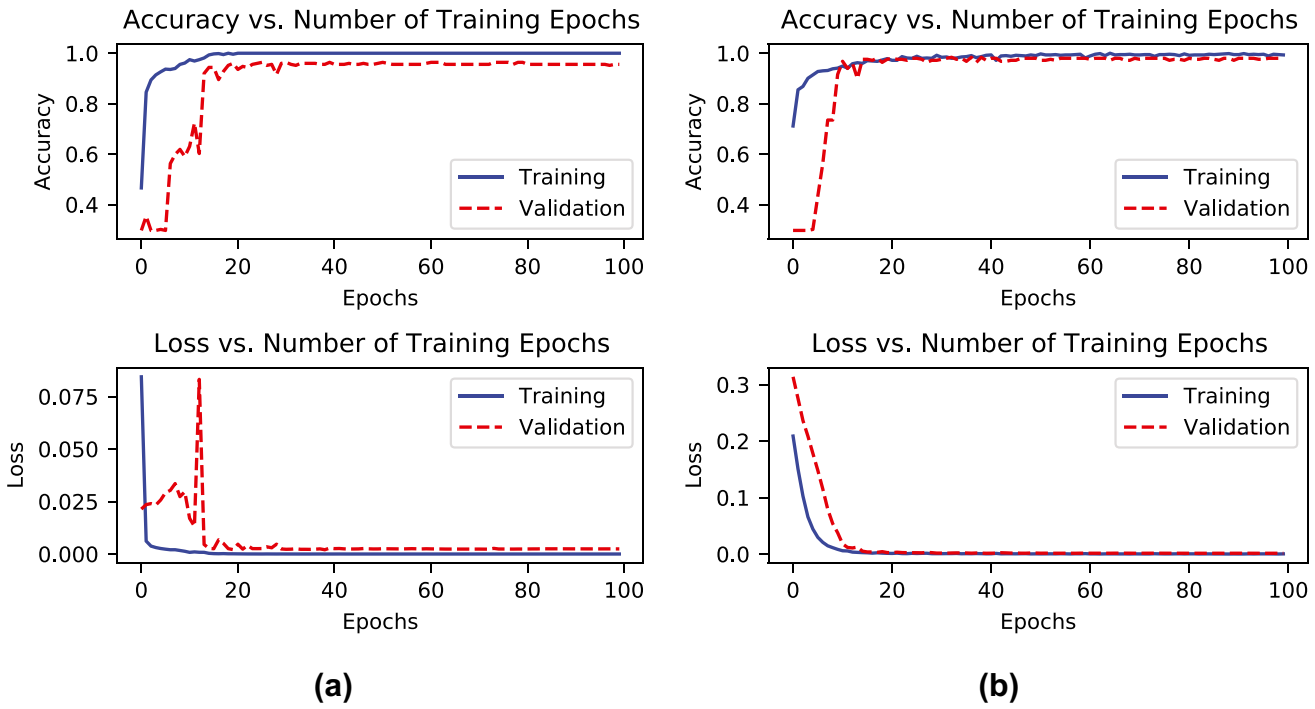


Fig. 7 Visualization of the training process on different capsnets in fold-1. **a** MHA-CoroCapsule. **b** Proposed model

extractors, which can effectively capture the local dependency and global dependency of COVID-19 features. We also construct homogeneous vector capsules to build the classification layer. Such a design significantly reduces the complexity of the model and achieves the best performance simultaneously. To conclude, our network is simple and efficient and plays a crucial role in preventing and controlling COVID-19.

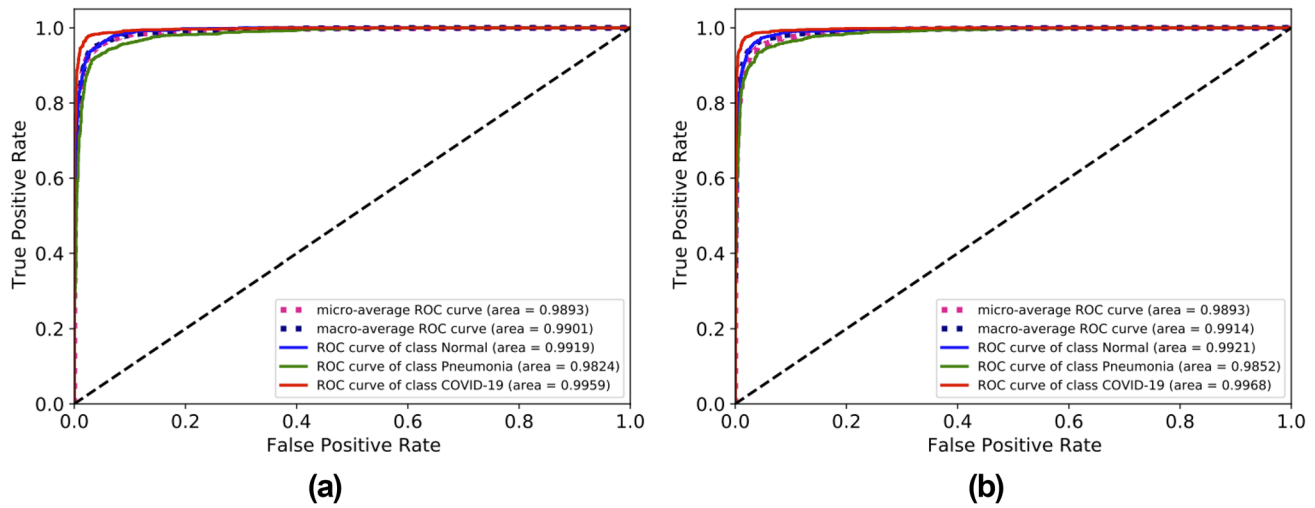


Fig. 8 ROC for 3-class classification on dataset-2 using trained capsule networks on dataset-1. **a** MHA-CoroCapsule. **b** Proposed model.

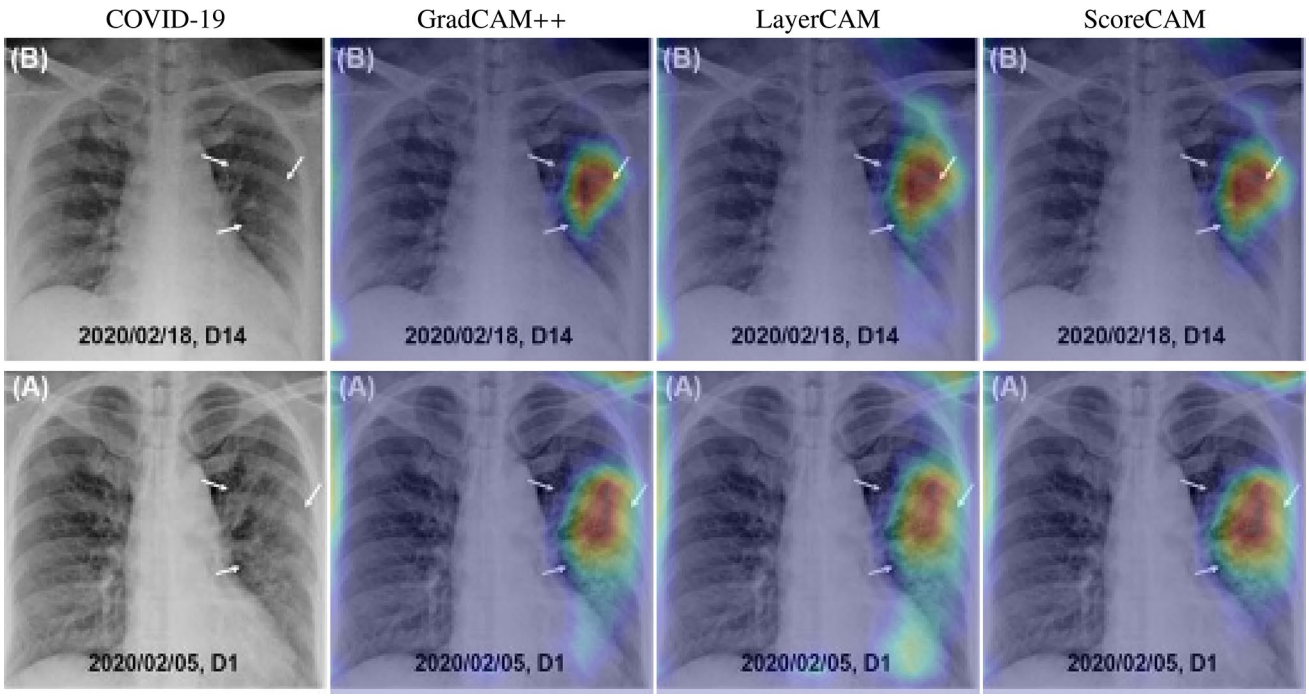


Fig. 9 Interpretation of the proposed model for four COVID-19 cases using different class activation maps. Pneumonia sites have been identified in the arrow and box sections.

Conclusions and Future Work

Mainstream DL frameworks mainly use deep feature extraction approaches or transfer learning to detect COVID-19. However, their performance often depends on massive labeled samples. This paper proposes a more lightweight capsule network, called DPDH-CapNet, which mainly consists of convolutional layers and capsule layers. The convolutional layer uses depthwise convolution, point convolution, and dilated convolution as a set of feature extractors, which can effectively capture local and long-range dependencies for the pathological features of COVID-19. The capsule layer is constructed with homogeneous vector capsules. It can effectively avoid the traditional matrix multiplication and expensive computational routing mechanisms dealing with the capsule dimensional entanglement between capsule layers. At the same time, it obtains competitive results in comparison with different routing mechanisms. In comparison with the state-of-the-art capsule networks, the parameters of our model are reduced by 9x, and it achieves the best performance. Moreover, it has faster convergence and better generalization. In addition, the proposed model also shows great advantages for COVID-19 recognition compared to current state-of-the-art transfer learning methods, and it does not require any pre-training. Extensive experiments also indicate that our model can achieve an accurate diagnosis for COVID-19 even under limited samples, with a lower computational overhead. Despite the encouraging results, the proposed model still requires clinical research and testing. Due to its higher accuracy and sensitivity for COVID-19 cases, the DPDH-CapNet contributes to a deeper understanding of critical aspects of COVID-19 cases for radiologists and health professionals.

Acknowledgements This work is supported by Fundamental Research Funds for the Central Universities (No. XDJK2020B033), National Key Research and Development Program (No. 2021YFB3101504), and Innovation and Entrepreneurship Training Program Project for College Students (No. S202210635186).

Author Contributions All authors contribute equally. All authors read and approved the manuscript.

Declarations

Ethics Approval The manuscript did not require ethics approval.

Consent to Participate There are not involving human subjects in this research. So, informed consent to participate in the study did not require.

Consent for Publication An open dataset was used to research, so, consent for publication did not require.

Conflict of Interest The authors declare no competing interests.

References

1. A. I. Khan, J. L. Shah, M. M. Bhat, Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images. Computer Methods and Programs in Biomedicine 196 (2020) 105581.
2. F. Li, X. Lu, J. Yuan, Mha-corocapsule: Multi-head attention routing-based capsule network for covid-19 chest x-ray image classification, IEEE Transactions on Medical Imaging (2021). doi:10.1109/TMI.2021.3134270.
3. B. Abraham, M. S. Nair, Computer-aided detection of covid-19 from x-ray images using multi-cnn and bayesnet classifier, Biocybernetics and Biomedical Engineering 40 (2020) 1436–1445.
4. A. I. Khan, J. L. Shah, M. M. Bhat, Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images, Computer Methods and Programs in Biomedicine 196 (2020) 105581.
5. M. Loey, F. Smarandache, N. E. M Khalifa, Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning, Symmetry 12 (2020) 651.
6. A. Abbas, M. M. Abdelsamea, M. M. Gaber, Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network, Applied Intelligence 51 (2021) 854–864.
7. Z. Wang, Y. Xiao, Y. Li, J. Zhang, F. Lu, M. Hou, X. Liu, Automatically discriminating and localizing covid-19 from community-acquired pneumonia on chest x-rays, Pattern Recognition 110 (2021) 107613.
8. S. Serte, H. Demirel, Deep learning for diagnosis of covid-19 using 3d ct scans, Computers in Biology and Medicine 132 (2021) 104306.
9. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al., A deep learning algorithm using ct images to screen for corona virus disease (covid-19), European Radiology 31 (2021) 6096–6104.
10. A. J. DeGrave, J. D. Janizek, S.-I. Lee, Ai for radiographic covid-19 detection selects shortcuts over signal, Nature Machine Intelligence 3 (2021) 610–619.
11. L. Wang, Z. Q. Lin, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, Scientific Reports 10 (2020) 1–12.
12. T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, U. R. Acharya, Automated detection of covid-19 cases using deep neural networks with x-ray images, Computers in Biology and Medicine 121 (2020) 103792.
13. G. E. Hinton, A. Krizhevsky, S. D.Wang, Transforming auto-encoders, in: International Conference on Artificial Neural Networks, Springer, 2011, pp. 44–51.
14. K. Adu, Y. Yu, J. Cai, N. Tashi, Dilated capsule network for brain tumor type classification via mri segmented tumor region, in: 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, 2019, pp. 942–947.
15. A. Mobiny, H. V. Nguyen, Fast capsnet for lung cancer screening, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 741–749.
16. P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, A. Mohammadi, Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images, Pattern Recognition Letters 138 (2020) 638–643.
17. S. Toraman, T. B. Alakus, I. Turkoglu, Convolutional capsnet: A novel artificial neural network approach to detect covid-19 disease from x-ray images using capsule networks, Chaos, Solitons & Fractals 140 (2020) 110122.
18. S. Sabour, N. Frosst, G. E. Hinton, Dynamic routing between capsules, Advances in Neural Information Processing Systems 30 (2017).
19. J. Choi, H. Seo, S. Im, M. Kang, Attention routing between capsules, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, pp. 0–0.
20. G. E. Hinton, S. Sabour, N. Frosst, Matrix capsules with em routing, in: International Conference on Learning Representations, 2018.
21. J. Rajasegaran, V. Jayasundara, S. Jayasekara, H. Jayasekara, S. Seneviratne, R. Rodrigo, Deepcaps: Going deeper with capsule networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10725–10733.
22. F. D. S. Ribeiro, G. Leontidis, S. Kollias, Capsule routing via variational bayes, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 3749–3756.
23. S. Venkatraman, S. Balasubramanian, R. R. Sarma, Building deep, equivariant capsule networks, arXiv preprint http://arxiv.org/abs/1908.01300 (2019).
24. Y.-H. H. Tsai, N. Srivastava, H. Goh, R. Salakhutdinov, Capsules with inverted dot-product attention routing, arXiv preprint https://doi.org/10.48550/arXiv.2002.04764 (2020).
25. V. Mazzia, F. Salvetti, M. Chiaberge, Efficient-capsnet: Capsule network with self attention routing, Scientific Reports 11 (2021) 1–13.
26. A. Jacobi, M. Chung, A. Bernheim, C. Eber, Portable chest x-ray in coronavirus disease-19 (covid-19): A pictorial review, Clinical Imaging 64 (2020) 35–42.
27. 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.
28. 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, arXiv preprint http://arxiv.org/abs/1704.04861 (2017).
29. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint. http://arxiv.org/abs/2010.11929 (2020).
30. H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, H. J´egou, Going deeper with image transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 32–42.
31. D. Li, J. Hu, C. Wang, X. Li, Q. She, L. Zhu, T. Zhang, Q. Chen, Involution: Inverting the inherence of convolution for visual recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12321–12330.
32. K. Li, Y. Wang, J. Zhang, P. Gao, G. Song, Y. Liu, H. Li, Y. Qiao, Uniformer: Unifying convolution and self-attention for visual recognition,arXiv preprint. http://arxiv.org/abs/2201.09450 (2022).
33. A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, A. Vaswani, Bottleneck transformers for visual recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16519–16529.
34. Z. Peng, W. Huang, S. Gu, L. Xie, Y. Wang, J. Jiao, Q. Ye, Conformer: Local features coupling global representations for visual recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 367–376.
35. J. Cohen, Covid chest x-ray dataset, Github https://github.com/ieee8023/covid-chestxray-dataset (Accessed on 05 September 2020) (2020).
36. P. Mooney, Kaggle chest x-ray images (pneumonia) dataset, 2020.
37. 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.
38. A. Chattopadhay, A. Sarkar, P. Howlader, V. N. Balasubramanian, Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 839–847. https://doi.org/10.1109/WACV.2018.00097.
39. P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, Y. Wei, Layercam: Exploring hierarchical class activation maps for localization, IEEE Transactions on Image Processing 30 (2021) 5875–5888.
40. H.Wang, Z.Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, X. Hu, Score-cam: Score-weighted visual explanations for convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 24–25.
41. I. D. Apostolopoulos, T. A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Physical and Engineering Sciences in Medicine 43 (2020) 635–640.
42. G. Jain, D. Mittal, D. Thakur, M. K. Mittal, A deep learning approach to detect covid-19 coronavirus with x-ray images, Biocybernetics and Biomedical Engineering 40 (2020) 1391–1405.
43. Y. Oh, S. Park, J. C. Ye, Deep learning covid-19 features on cxr using limited training data sets, IEEE Transactions on Medical Imaging 39 (2020) 2688–2700.
44. K. Hammoudi, H. Benhabiles, M. Melkemi, F. Dornaika, I. Arganda-Carreras, D. Collard, A. Scherpereel, Deep learning on chest x-ray images to detect and evaluate pneumonia cases at the era of covid-19, Journal of Medical Systems 45 (2021) 1–10.
45. M. Toğaçar, B. Ergen, Z. Cömert , Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches, Computers in Biology and Medicine 121 (2020) 103805.