ASA-CoroNet: Adaptive Self-attention Network for COVID-19 Automated Diagnosis using Chest X-ray Images

Fujun Wu [1] **Jianjun Yuan** [1] **Yuxi Li** [1] **Jinyi Li** [1] **Ming Ye** [1]

Abstract

Computer-assisted imagery analysis based on chest X-ray images plays a crucial role in the clinical diagnosis and screening of COVID-19. However, the radiographic features of chest X-rays are highly complex and irregular in shape. Moreover, the size and location of the lesion regions vary greatly with infection stages and patients, thus dramatically increasing the difficulty of COVID-19 identification. A lightweight adaptive self-attention network is developed to address this problem, namely ASA-CoroNet. It firstly extracts underlying features using a depthwise separable convolution-based backbone, then further identifies lesion regions through an adaptive self-attentive module, and finally utilizes a homogeneous vector capsule layer to map the obtained features into capsule vectors to instantiate detection objects accurately. Extensive experimental results demonstrate that the proposed model outperforms the state-of-the-art methods and obtains competitive results on limited datasets. More importantly, the trainable params of the proposed model are reduced by 7x compared to the state-of-the-art capsule network. In addition, we also interpret the proposed model using different class activation techniques and confirm the validity of the three components through numerous ablation studies.

1. Introduction

Coronavirus disease (COVID-19) has rapidly spread across the globe since December 2019, and the repeated epidemics still severely influence people's work, study, and life. It is urgent to develop new detection technologies. For a long time, reverse transcription-polymerase chain reaction (RT-PCR) has been considered the gold standard for COVID-19 detection (Huang et al., 2020). However, its detection cycle is long, and its sensitivity is low (Kovács et al., 2021). Therefore, computer-assisted imagery analysis becomes the key to solving this problem, such as computed tomographic (CT) scans and chest X-rays (CXR). Compared with CT imaging, CXR has a shorter diagnostic time and lower cost. Therefore, more and more researchers attempt to introduce deep learning (DL), especially convolutional neural network (CNN), to improve COVID-19 detection efficiency and accuracy on CXR images (Öksüz et al., 2021; Abraham & Nair, 2020; Khan et al., 2020).

1.1. Literature Review

As in other healthcare, to escape the dilemma of data scarcity, some research uses the transfer learning (TL) strategy, that is, fine-tunes models trained on large-scale data using COVID-19 datasets. (Loey et al., 2020) introduced generative adversarial network (GAN) based on AlexNet (Krizhevsky et al., 2012) to solve the problem of insufficient COVID-19 samples, the effect of which is better than single transfer learning methods, such as GoogleNet (Szegedy et al., 2015) and ResNet18 (He et al., 2016). In addition, (Apostolopoulos & Mpesiana, 2020) combined CNN and TL to detect CXR images. (Abbas et al., 2021) 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. Furthermore, (Wang et al., 2021b) combined ResNet with the feature pyramid network to improve ResNet to obtain better performance. Unlike this, (Serte & Demirel, 2021) used multiple image levels to diagnose COVID-19 at the 3D CT volume level. Its detection effect outperformed a single 3D-Resnet. Other transfer methods include InceptionV3 (Szegedy et al., 2016), DenseNet (Huang et al., 2017), etc. Transfer learning has achieved satisfactory results in COVID-19 identification, but with its complex model and high computational overhead. Therefore, DL frameworks specifically for COVID-19 were proposed, such as COVID-net (Wang et al., 2020b), (Ozturk et al., 2020), etc. Nevertheless, the performance of such models on multi-classification needs to be further improved.

[1]College of Artificial Intelligence, Southwest University, Tiansheng Road No.2, 400715, Chongqing, China.. Correspondence to: Jianjun Yuan <jianjuny@sina.com>.

Proceedings of the 1st Workshop on Healthcare AI and COVID-19, ICML 2022.

Additionally, since CNN has some potential defects, especially it cannot capture the relative positional relationship between features. So, (Hinton et al., 2011) exploited a new architecture, referred to as a capsule network, as a powerful alternative to CNN. This structure uses capsules (vectors containing feature information) to effectively avoid the loss of high-level feature information through routing mechanisms, demonstrating certain advantages in medical image processing (Mobiny & Nguyen, 2018; Adu et al., 2019). Inspired by this, (Afshar et al., 2020) proposed a capsule network for detecting COVID-19 using CXR images, named COVID-CAPS, and achieved an accuracy of 95.7%. Similarly, (Toraman et al., 2020) proposed an artificial neural network approach to detect COVID-19 disease. Recently, (Li et al., 2022) designed a new capsule network using multi-head attention routing and obtained the optimal effect on COVID-19 CXR image classification. However, the routing follows vast computation overhead and is limited to a certain extent by the entanglement of capsule dimensions (Byerly et al., 2021). Inspired by this, in the transition from the lower-layer capsule to the upper-layer capsule, we abandon the routing method and only rely on the weights learned between capsule layers in the process of backpropagation. This makes the realizable precision of the model less dependent on the fine-tuned hyperparameters.

1.2. Contributions

Current research has made significant progress in the identification accuracy of COVID-19. However, these methods still face tough challenges in clinical application. 1) Most studies directly extract features from CXR images, only considering image-level features. This easily makes the important information in the infection site neglected and may also make the network learn more redundant information, such as background and noise. 2) It is challenging to learn radiographic features from COVID-19 robustly. Firstly, the CXR has diverse COVID-19 radiographic features, such as lung consolidation, ground glass opacities, lung opacities, and peripheral lung involvement (Jacobi et al., 2020). Secondly, COVID-19 lesion shapes are complex, such as diffuse, reticular nodular (Stogiannos et al., 2020). Moreover, the size and location vary greatly with infection stages and patients. 3) The mainstream methods also have some problems. DL transfer learning models are highly complex. Capsule networks rely on expensive routing calculations and will follow a capsule shedding problem with training samples increases. In view of the above analysis, we propose a novel adaptive self-attention network, namely ASA-CoroNet, to realize the automatic diagnosis of COVID-19. It first constructs a lightweight CNN feature extraction backbone and further adaptively captures the infected regions. Finally, the extracted features are mapped into capsules to instantiate the classification object. The main contributions are summarized as follows:

1) This paper develops a novel adaptive self-attention network. The network subtly incorporates the advantages of CNN and the capsule network. It not only achieves the adaptive learning of COVID-19 complex infection regions and global contextual information interaction of COVID-19 pathological features, but also fully considers the relative position information of COVID-19 radiographic features.

2) The adaptive self-attention module is proposed, which can adaptively adjust the receptive field according to the complex and diverse COVID-19 CXR images while non-locally interacting with the global context information. The aim is to assist the model in perceiving infected regions.

3) Homogeneous vector capsules are designed as the classification layer. It tactfully avoids traditional matrix multiplication between capsule layers or expensive routing computation to deal with the entanglement of capsule dimensions. Compared to fully connected layers, the design can map the extracted features more comprehensively and effectively to improve the discriminative ability of the model.

4) The proposed model obtains excellent performance on a limited training dataset and does not require pre-training. Moreover, experimental results demonstrate that our model outperforms the state-of-the-art transfer learning models and capsule networks. Notably, its trainable params are reduced by 7x compared with the state-of-the-art capsule network. In addition, we perform model interpretation and ablation studies to confirm the feasibility of our method.

The remainder of this paper is as follows. In the section 2, we expound the proposed network architecture in detail. Experimental results and analysis are in section 3. The conclusions are in section 4.

2. Proposed Model

As shown in Figure 1, the proposed model contains three main components. Firstly, the backbone extracts underlying features. On this basis, the adaptive self-attention (ASA) module further captures complex and diverse lesion sites. Finally, the homogeneous vector capsule layer combines valuable features from the ASA module to instantiate normal, pneumonia, and COVID-19 CXR image objects. The framework fully considers the COVID-19 feature distribution, shape, infected regions, and relative location information.

2.1. Feature Extraction Backbone

This paper adopts depthwise separable convolution (Chollet, 2017) (DSC) to design a lightweight feature extraction backbone. It first uses depthwise convolution to extract different channel features of the input chest X-ray image, then uti-

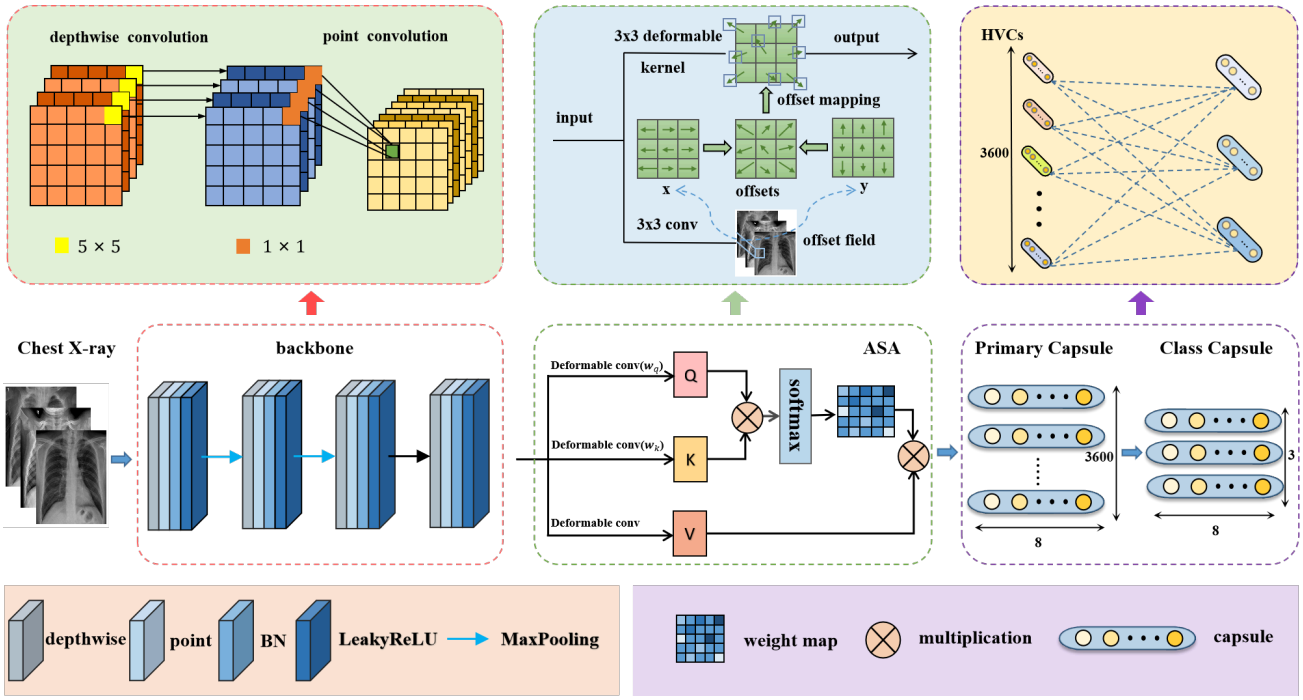

Figure 1. The architecture of the ASA-CoroNet. BN and HVCs denote batch normalization and homogeneous vector capsules, respectively.

lizes point convolution to weight and combine the obtained features. Compared with standard convolution, the DSC can realize the separation of feature map channels and regions, thus accelerating model training. Assuming the width and height of the convolution kernel are K_w and K_h, respectively. Besides, C_{in} and C_{out} represent input channels and output channels, respectively, then for standard convolution, its parameters are:

$$P_{st} = K_w \times K_h \times C_{in} \times C_{out}. \tag{1}$$

For the DSC, parameters can be calculated as:

$$P_{ds} = K_w \times K_h \times C_{in} + 1 \times 1 \times C_{in} \times C_{out}. \tag{2}$$

So,

$$P_{ds}/P_{st} = 1/C_{out} + 1/(K_h \times K_w) < 1. \tag{3}$$

According to (3), the DSC can effectively reduce the parameters of the model compared to standard convolution. To conclude, the lightweight feature extraction framework is beneficial in reducing computational overhead, thus making it easier to deploy on COVID-19 detection equipment. Further, we combine the DSC, batch normalization (BN), and LeakyReLU activation function to design 4 sets of feature extractors. The first and second extractors are followed by the MaxPooling. The pooling unit is 2×2. Convolution kernel sizes are all set to 5×5. In addition, all strides are set to 1. After each extractor operates, the channels of the obtained feature maps are 16, 32, 64, and 128 in turn.

2.2. Adaptive Self-attention Module

To enable the network to further adaptively detect the geometrical changes of COVID-19 abnormal regions and effectively distinguish COVID-19 from other pneumonia symptoms, we exploit an adaptive self-attention module. It first captures the shape changes of the input features using deformable convolutions (Dai et al., 2017), and then realizes non-locally interaction of the global context information of COVID-19 through a self-attention mechanism to further capture infected regions. This operation can further enhance the features of important regions while suppressing irrelevant information such as background and noise, thereby assisting the homogeneous vector capsule layer to instantiate objects. Unlike standard convolutions, which can only sample fixed-position features, deformable convolution can learn the offset of the sampling position from the target task, thus dynamically adjusting the receptive field to control the spatial support region. Specifically, assuming that the position of the input feature $x \in \mathbb{R}^{C \times W \times H}$ is p, the output feature y is:

$$y(p) = \sum_{p_k \in \mathcal{R}} w_k \cdot x\left(p + p_k + \Delta p_k\right), \tag{4}$$

where the sampling grid $\mathcal{R}$ with learnable offsets $\{\Delta p_k\}_{k=1}^{N=|\mathcal{R}|}$, and w_k, and p_k describe the weights of the k_{th} positions and prespecified offsets, respectively. In particular, the learnable spatial offset Δp_k contains the offsets in the horizontal and vertical directions and are learned by

additional convolutional layers, with the corresponding kernel size set to 3. In addition, global contextual information is critical as COVID-19 pathological features tend to be diffuse or localized (Jacobi et al., 2020), and the diffuse trend is strengthened with an increasing degree of infection. We further design the self-attention mechanism. Specifically, x from deformable conv is mapped to spaces $K(x)$, $Q(x)$, $V(x) \in \mathbb{R}^{C\times N}$. The attention map is computed by $K(x)$ and $Q(x)$:

$$\boldsymbol{\beta} = softmax(K(x)^{\top} \times Q(x)). \tag{5}$$

Note that,

$$\sum_{i} \beta_{i,j} = 1. \tag{6}$$

Finally, $V(x)$ is updated using $\boldsymbol{\beta}$ to obtain the output feature map F_{out}, which is expressed as follows.

$$F_{out} = \sum_{i=1}^{N} V(x)\beta_{j,i}. \tag{7}$$

In summary, the ASA module further adaptively samples the lesion region from the feature map, thus effectively capturing its contour information. Further, it enhances the abnormal region features from the global context information while suppressing the background and irrelevant features to capture useful features, aiming to facilitate the discrimination of COVID-19 by the capsule layer.

2.3. Homogeneous Vector Capsules

The classification layer is constructed with homogeneous vector capsules. It first uses each different x and y coordinate of the feature maps to construct capsule vectors, which effectively combines meaningful features to instantiate the capsules. After this operation, the primary capsule layer $P_{n,d}$ is created, where $n = 3600$ and $d = 8$ denote its amount and dimension, respectively. Secondly, we utilize element-wise multiplication to map primary capsules to class capsules, that is:

$$W_i \odot P_i = C_j, \tag{8}$$

where W_i denotes the learnable weight corresponding to the primary capsule P_i, and $i = 0, ..., 3600$. C_j $(j = 1, 2, 3)$ represents the class capsule. The primary capsule and the class capsule layers have the same dimensions through the formula (8), so they are called homogeneous vector capsules, and their visualization form can be referred to as the HVCs in Figure 1. This method has two advantages. The first is in comparison with the routing process. 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 capsule dimension for the dynamic routing method proposed by Sabour et al. (Sabour et al., 2017). Furthermore, this method can model feature vectors flexibly. While the vector dimension must meet the perfect square in the paper (Hinton et al., 2018), which dramatically limits its application in COVID-19 detection. Secondly, compared with the fully connected layer mapping method, it contains richer feature representations, such as a specific entity type and how the entity is instantiated.

After homogeneous operations, the class capsule layer $C_{3,8}$ is obtained with 3 capsules with 8 dimensions. Moreover, class capsules also contain instantiated parameters for normal, pneumonia, and COVID-19 objects. In order to make the length of the activity vector corresponding to each capsule represents the probability that each class exists, the class capsules are activated by a non-linear "squashing" function. It is expressed as follows:

$$squash(\mathbf{C}_n) = (1 - \frac{1}{e^{||\mathbf{C}_n||}})\frac{\mathbf{C}_n}{||\mathbf{C}_n||}. \tag{9}$$

A single capsule is defined as $\mathbf{C}_n \in \mathbb{R}^8$ in the proposed model. After the squash activation function operates, the obtained class capsules have a length "squashed" between zero and one.

2.4. Margin Loss

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

$$\begin{aligned} L_c = &\sum_{k\in CN_{um}} T_k max(0, m^{+} - ||\mathbf{u}_k||)^2 \\ &+ \lambda(1 - T_k)max(0, ||\mathbf{u}_k|| - m^{-})^2. \end{aligned} \tag{10}$$

where T_k is the sample class label. If k class exists, $T_k = 1$. $\lambda = 0.6$ is the balance coefficient, which lowers the weights of the loss for non-existing classes. The two parameters are used to 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 $m^{-} = 0.2$ 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.

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

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

Figure 2. Example images from dataset-2 with three categories: (a) Normal, (b) Pneumonia, and (c) COVID-19.

3. Experiments and Analysis

3.1. Experimental Preparation

Datasets: Table 1 shows two datasets containing normal, pneumonia, and COVID-19 labeled CXR images. Their examples are displayed in Figure 2. COVID-19 images in dataset-1 came from different patients and were created by Dr. Joseph Cohen (Cohen, 2020). In addition, the 350 normal and non-COVID-19 pneumonia CXR images are provided by Kaggle (Mooney, 2020). Dataset-2 contains 1200 COVID-19 positive images, 1341 normal images, and 1345 viral pneumonia images, which come from the COVID-19 radiography database (Chowdhury et al., 2020). These original labeled CXR images have varying lengths and width sizes, so we rescale them to 128×128 pixels. Moreover, we also normalize the pixel values of the images from $[0, 255]$ to $[0, 1]$ with the Min-Max scaling method.

Model Training and Testing: Firstly, experiments are performed in the Google Colab cloud experimental environment with Python 3.7, Keras 2.4.3, and TensorFlow-GPU 2.8.0. Secondly, we train all models using the graphical processing unit (GPU) Tesla T4 with 16 GB. In addition, the optimization uses the Adam with an initial learning rate of 0.001, and it is lowered with a 0.5 decay rate and a 15 decay step. Finally, we set the batch size and epoch to 16 and 100. The training and test sets are divided by 3:1 for a.

Model Evaluation: The accuracy, precision, recall, and F-measure are used as evaluation indicators. They are defined as follows. 1) Accuracy (Acc.) = (TP + TN)/(TP + FP + TN + FN). 2) Precision (Pre.) = TP/(TP + FP). 3) Recall (Rec.) = TP/(TP + FN). 4) F-measure = 2TP/(2TP + FP + FN). where TP, FP, TN, and FN respectively denote true positive, false positive, true negative, and false negative. In addition, we also use AUC, macro average, and weighted average.

3.2. Comparison with the State-of-the-art Models

To verify the performance of the proposed model, we compare it with the current state-of-the-art capsule networks and transfer learning models, both of which have achieved competitive results for COVID-19 detection. Capsule networks include convolutional capsnet (Toraman et al., 2020), COVID-CAPS (Afshar et al., 2020), and MHA-CoroCapsule (Li et al., 2022). Transfer learning methods use VGG16 (Al-Bawi et al., 2020), ResNet50 (Mishra et al., 2021), DenseNet121 (Ezzat et al., 2021), InceptionV3 (Wang et al., 2021a), and MobileNet (Gupta et al., 2021). The results are shown in Table 2. Our model both outperforms other methods on both dataset-1 and dataset-2. For dataset-1, compared with the suboptimal model MHA-CoroCapsule, the acc, pre, rec, F-measure, and AUC of the ASA-CoroNet are improved by 1.2%, 1.19%, 1.14%, 1.14%, and 0.13%, respectively. It is emphasized that the proposed model has the lowest trainable params, reduced by 7x compared to MHA-CoroCapsule. Further analysis demonstrates that the transfer learning methods involved in the experiments are all pre-trained on ImageNet and then fine-tuned on dataset-1 and dataset-2, respectively. Moreover, the source and target tasks are both classification. Theoretically, they have robust performance, but they do not escape from sample dependence with the high complexness of the model. Different from this, the effect of the convolutional capsnet rises a substantial decrease on dataset-2, which could be attributed to the increase in training samples leading to capsule shedding. In contrast, our model also has better generalization even under limited training samples. In addition, we visualize ROC for each class in Figure 3 and the training procedure in Figure 4 of the proposed model. According to Figure 3, the proposed network has an excellent generalization for normal, pneumonia, and COVID-19 objects, among which the AUC of COVID-19 is as high as 0.9955. Simultaneously, Figure 4 indicates that it also has good convergence and stability, and the training can be accomplished in 20 epochs iterations.

There are some similarities between the CXR medical image features of COVID-19 and other pneumonia. Consequently, the discriminative ability for COVID-19 and pneumonia is an essential indicator in assessing the performance of the model. Therefore, we implement comparison experiments. The results are reflected in Figure 5. The proposed model still obtains the best effects, which demonstrates its robust discriminative ability for COVID-19 and other pneumonia.

3.3. Model Interpretation and Robustness Analysis

When samples are small, K-fold cross-validation is a crucial method to evaluate the robustness of the model. The method can effectively reduce the dependence between the accuracy estimates, thereby ensuring the reliability of the results. In the current related COVID-19 identification research, only a few apply this strategy (Aggarwal et al., 2022). The dataset-1 has only 994 CXR cases, so on which we deploy 4-fold cross-validation to test the robustness of the proposed model. As shown in Table 3, our model achieves the best performance on fold-1, and the corresponding means of

Table 2. Performance among our model, capsule networks, and pre-trained models for dataset-1 and dataset-2 (In percentage %).

	Dataset-1					Dataset-2					Trainable params
	Acc.	Pre.	Rec.	F-measure	AUC	Acc.	Pre.	Rec.	F-measure	AUC	
Convolutional Capsnet	91.97	92.27	92.09	92.10	97.27	78.19	83.12	78.44	78.50	93.87	57,002,160
COVID-CAPS	95.18	95.30	95.35	95.29	98.35	95.67	95.81	95.76	95.77	98.79	295,488
MHA-CoroCapsule	96.39	96.61	96.43	96.52	99.52	97.02	97.07	97.08	97.07	99.00	329,232
MobileNet	90.76	90.78	91.03	90.87	98.21	95.37	95.51	95.47	95.47	99.29	3,210,051
ResNet50	90.36	90.77	90.60	90.60	96.65	95.47	95.55	95.58	95.56	99.46	23,540,739
DenseNet121	91.16	91.33	91.41	91.37	97.87	96.91	97.00	96.98	96.98	99.60	6,956,931
VGG16	91.16	91.42	91.64	91.33	98.28	93.00	93.30	93.18	93.11	98.93	50,352,131
InceptionV3	91.96	92.99	91.75	92.13	98.79	95.58	95.57	95.69	95.61	98.30	21,774,499
Ours	**97.59**	**97.80**	**97.57**	**97.66**	**99.65**	**97.53**	**97.54**	**97.56**	**97.55**	**99.49**	**38,024**

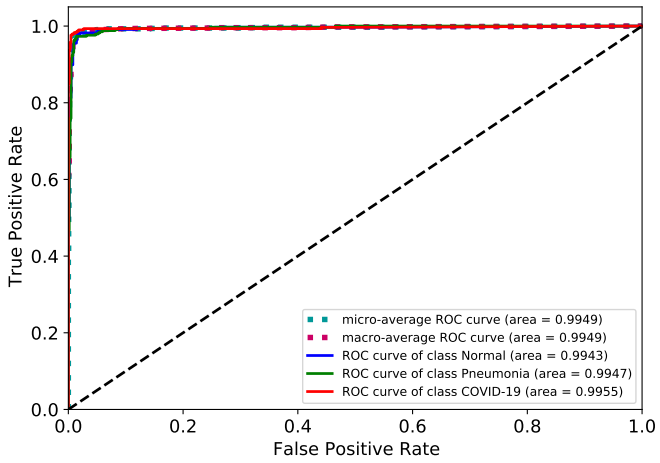

Figure 3. ROC of the proposed model on dataset-2.

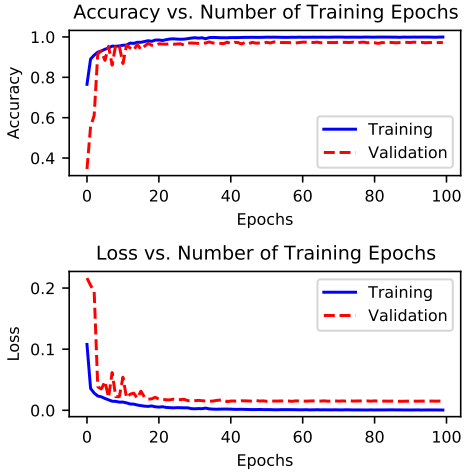

Figure 4. Visualization of the training process for the proposed model on dataset-2.

Table 3. The stability test results using the 4-fold cross-validation method on our model (In percentage %).

Folds	Acc.	Pre.	Rec.	F-measure	AUC
Fold-1	97.59	97.75	97.56	97.65	99.60
Fold-2	97.19	97.28	97.20	97.23	99.20
Fold-3	96.37	96.46	96.42	98.04	99.38
Fold-4	96.77	96.87	96.95	96.90	99.48
Mean	96.98	97.09	97.03	97.05	99.48
Std	0.46	0.48	0.42	0.45	0.20

acc, pre, rec, F-measure, and AUC on all folds are 96.98%, 97.09%, 97.03%, 97.05%, and 99.48%, respectively. More importantly, the std of each indicator is less than 0.5, further illustrating the strong performance of the proposed model on limited samples, which is particularly important in the data-poor medical field. In addition, we also show the confusion matrix on each fold in Figure 6. To further prove this conclusion, this paper implements a transfer application study by training all networks on dataset-1 and then deploying them directly on dataset-2 for classification. According to Table 4, our method still obtains the optimal results, and each evaluation index exceeds 95.19%. This result is uplifting. Firstly, the model is learned the feature space from limited samples and then transfers it to larger samples. Secondly, the datasets are from different patients and databases, and there is wide heterogeneity in image quality. Both of these aspects greatly increase the difficulty of the model prediction.

Model transparency is essential when DL models are used for life-threatening COVID-19 disease detection. This paper adopts three class activation techniques to achieve the interpretation and behavioral understanding of the ASA-CoroNet, including GradCAM++ (Chattopadhay et al., 2018), LayerCAM (Jiang et al., 2021), and ScoreCAM (Wang et al., 2020a). According to Figure 7, even though the

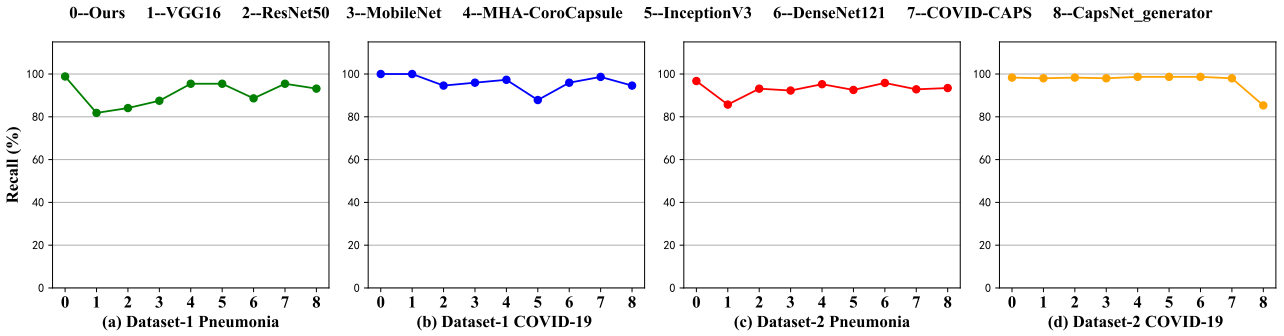

Figure 5. The recall for COVID-19 and pneumonia CXR images on dataset-1 and dataset-2.

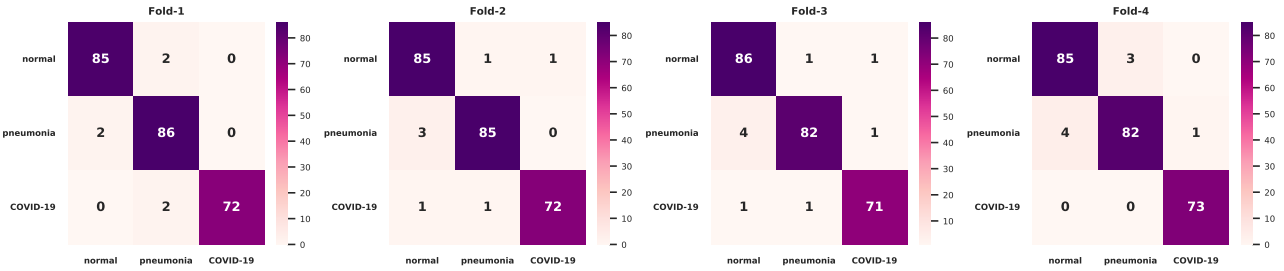

Figure 6. 4-fold confusion matrices of our model on dataset-1. From left to right are fold-1, fold-2, fold-3, and fold-4.

Table 4. Generalization performance comparison between the proposed network and pre-trained models (In percentage %).

Methods	Acc.	Pre.	Rec.	F-measure	AUC
Convolutional Capsnet	90.81	91.07	91.00	91.00	96.66
COVID-CAPS	93.62	93.75	93.71	93.71	97.94
MHA-CoroCapsule	94.67	95.00	94.79	94.77	98.96
MobileNet	90.12	90.22	90.31	90.23	97.51
ResNet50	90.02	90.43	90.03	90.13	97.55
DenseNet121	91.92	92.04	92.02	92.02	98.25
VGG16	90.07	91.03	90.37	90.14	98.29
InceptionV3	92.25	92.72	92.16	92.34	98.67
Ours	**95.19**	**95.23**	**95.26**	**95.23**	**99.14**

proposed model only obtains image-level labels, it can detect COVID-19 lesions even their shape, which can greatly assist doctors in fast screening and diagnosing COVID-19. In conclusion, the COVID-19 solution based on the ASA-CoroNet is acceptable and effective.

Additionally, we further explore the decision-making effect of the proposed model through misclassification images and their prediction scores, as displayed in Figure 8. The fully connected layer maps the extracted features to neuron scalars and then obtains the class probabilities by softmax operations. Conversely, the ASA-CoroNet constructs capsule vectors to instantiate class objects and then predicts the class scores through the length of class capsules. According to the blue bars, for COVID-19 misclassified images, the ASA-CoroNet can still learn the most instantiated features even though their symptoms are vague.

3.4. Ablation Study

In this section, we perform ablation studies on the three key components of the proposed model, aiming to analyze the role of each component in COVID-19 recognition. Results are recorded in Table 5. Moreover, the experiments utilize the backbone constructed by standard convolutional and the fully connected layer as a baseline. Its performance is shown in No.1 of Table 5, and the relevant settings of the backbone except the convolution method remain the same as the proposed model.

Effects of the depthwise separable convolution: To investigate the positive impact of depthwise separable convolution on COVID-19 identification, we replace the standard convolution of No.1 with depthwise separable convolution. According to No.2, the performance of the transformed baseline is improved on both dataset-1 and dataset-2. For example, acc, pre, rec, F-measure, and AUC on dataset-1 are improved by 2.41%, 2.61%, 2.47%, 2.46%, and 1.56%, respectively. Moreover, trainable params are reduced by nearly 12x. Results indicate that DSC plays a key role in reducing model complexity while extracting useful features.

Table 5. Ablation studies for the three key modules involved in the proposed model (In percentage %). SC and FC denote standard convolution and fully connected layer, respectively. No. 1 is used as the baseline.

No.	SC	DSC	ASA	FC	HVCs	Dataset-1					Dataset-2					Trainable params
						Acc.	Pre.	Rec.	F-measure	AUC	Acc.	Pre.	Rec.	F-measure	AUC	
1	√			√		89.96	90.36	90.07	90.19	97.18	93.21	93.29	93.33	93.30	98.73	271,107
2		√		√		92.37	92.97	92.54	92.65	98.74	94.03	94.10	94.16	94.11	99.04	20,785
3	√		√	√		95.18	95.25	95.37	95.28	99.47	95.68	95.76	95.79	95.77	99.38	291,861
4	√				√	96.79	96.83	96.81	96.82	99.08	96.50	96.64	96.59	96.58	99.37	270,720
5		√	√		√	**97.59**	**97.80**	**97.57**	**97.66**	**99.65**	**97.53**	**97.54**	**97.56**	**97.55**	**99.49**	**38,024**

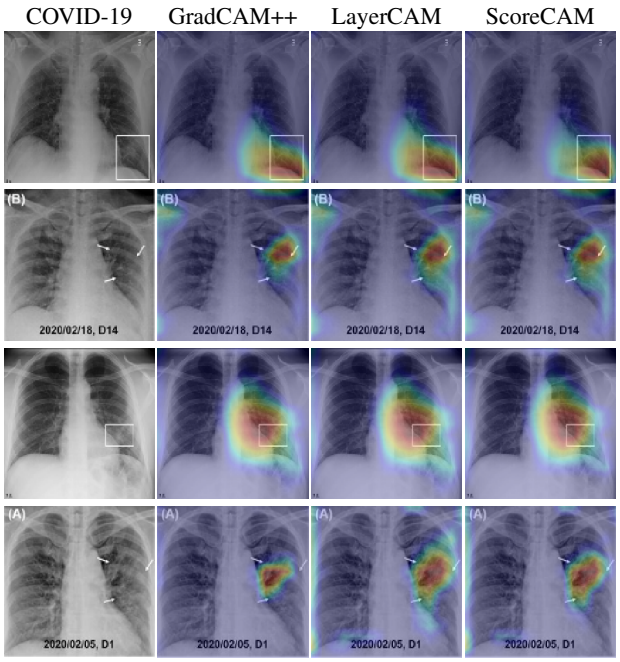

Figure 7. Interpretation of the proposed model for four COVID-19 cases using different class activation maps. Pneumonia sites have been identified in the arrow and box sections.

Effects of the ASA module: Furthermore, we introduce the ASA module into the baseline to further explore its effectiveness. It can be seen from No. 3 that the performance is greatly improved on both datasets compared to the baseline. The reason is that the ASA module can adaptively adjust the receptive field and simultaneously realize the non-local interaction of context information, thus accurately locating the infected region. In this way, the network can better learn COVID-19 radiographic features to facilitate its diagnosis.

Effects of the HVCs: Further, we compare the effect of fully connected and HVCs at the classification layer. Results of No.4 confirm that the classification layer design of HVCs can promote the discrimination of the model compared to the fully connected method. This is because HVCs map the captured features into vectors and instantiate objects

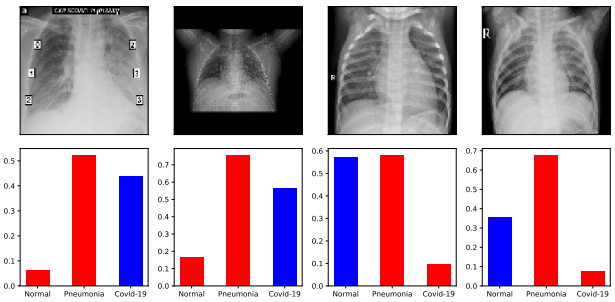

Figure 8. Examples of the ASA-CoroNet misclassified images. Blue bars represent correct labels and their corresponding capsule length.

in the form of capsules, which subtly consider the relative positional relationship and obtain a richer representation of entity features.

4. Conclusions

This paper constructs a lightweight and efficient detection framework, namely ASA-CoroNet, which concludes three components: a depthwise separable convolution-based backbone, an adaptive self-attention module, and a homogeneous vector capsule layer, subtly incorporating the advantage of CNN and the capsule network. Extensive experimental results confirm the superiority and feasibility of the proposed model. In summary, our method has excellent generalizability in COVID-19 scenarios. Firstly, it greatly reduces computational overhead with a lightweight architecture. Secondly, it can adaptively capture lesion areas to better assist doctors in diagnosis and decision-making. Finally, the proposed model can be deployed on limited datasets, thus making it more suitable for epidemic screening needs. Our work is to promote automated diagnosis of COVID-19.

Acknowledgements

This work is supported by Fundamental Research Funds for the Central Universities (No.XDJK2020B033).

References

Abbas, A., Abdelsamea, M. M., and Gaber, M. M. Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. *Applied Intelligence*, 51(2):854–864, 2021.

Abraham, B. and Nair, M. S. Computer-aided detection of covid-19 from x-ray images using multi-cnn and bayesnet classifier. *Biocybernetics and Biomedical Engineering*, 40(4):1436–1445, 2020.

Adu, K., Yu, Y., Cai, J., and Tashi, N. Dilated capsule network for brain tumor type classification via mri segmented tumor region. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 942–947. IEEE, 2019.

Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K. N., and Mohammadi, A. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. *Pattern Recognition Letters*, 138:638–643, 2020.

Aggarwal, P., Mishra, N. K., Fatimah, B., Singh, P., Gupta, A., and Joshi, S. D. Covid-19 image classification using deep learning: Advances, challenges and opportunities. *Computers in Biology and Medicine*, pp. 105350, 2022.

Al-Bawi, A., Al-Kaabi, K., Jeryo, M., and Al-Fatlawi, A. Ccblock: an effective use of deep learning for automatic diagnosis of covid-19 using x-ray images. *Research on Biomedical Engineering*, pp. 1–10, 2020.

Apostolopoulos, I. D. and Mpesiana, T. A. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 43(2):635–640, 2020.

Byerly, A., Kalganova, T., and Dear, I. No routing needed between capsules. *Neurocomputing*, 463:545–553, 2021.

Chattopadhay, A., Sarkar, A., Howlader, P., and Balasubramanian, V. N. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847, 2018. doi: 10.1109/WACV.2018.00097.

Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, 2017.

Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Al Emadi, N., et al. Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8:132665–132676, 2020.

Cohen, J. Covid chest x-ray dataset. *Github https://github.com/ieee8023/covid-chestxray-dataset (accessed on 05 September 2020)*, 2020.

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 764–773, 2017.

Ezzat, D., Hassanien, A. E., and Ella, H. A. An optimized deep learning architecture for the diagnosis of covid-19 disease based on gravitational search optimization. *Applied Soft Computing*, 98:106742, 2021.

Gupta, A., Gupta, S., Katarya, R., et al. Instacovnet-19: A deep learning classification model for the detection of covid-19 patients using chest x-ray. *Applied Soft Computing*, 99:106859, 2021.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Hinton, G. E., Krizhevsky, A., and Wang, S. D. Transforming auto-encoders. In *International Conference on Artificial Neural Networks*, pp. 44–51. Springer, 2011.

Hinton, G. E., Sabour, S., and Frosst, N. Matrix capsules with em routing. In *International Conference on Learning Representations*, 2018.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*, 395(10223):497–506, 2020.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

Jacobi, A., Chung, M., Bernheim, A., and Eber, C. Portable chest x-ray in coronavirus disease-19 (covid-19): A pictorial review. *Clinical Imaging*, 64:35–42, 2020.

Jiang, P.-T., Zhang, C.-B., Hou, Q., Cheng, M.-M., and Wei, Y. Layercam: Exploring hierarchical class activation maps for localization. *IEEE Transactions on Image Processing*, 30:5875–5888, 2021.

Khan, A. I., Shah, J. L., and Bhat, M. M. Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, 196:105581, 2020.

Kovács, A., Palásti, P., Veréb, D., Bozsik, B., Palkó, A., and Kincses, Z. T. The sensitivity and specificity of chest ct in the diagnosis of covid-19. *European Radiology*, 31(5): 2819–2824, 2021.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2012.

Li, F., Lu, X., and Yuan, J. Mha-corocapsule: Multi-head attention routing-based capsule network for covid-19 chest x-ray image classification. *IEEE Transactions on Medical Imaging*, 41(5):1208–1218, 2022. doi: 10.1109/TMI.2021.3134270.

Loey, M., Smarandache, F., and M Khalifa, N. E. Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning. *Symmetry*, 12(4):651, 2020.

Mishra, N. K., Singh, P., and Joshi, S. D. Automated detection of covid-19 from ct scan using convolutional neural network. *Biocybernetics and Biomedical Engineering*, 41 (2):572–588, 2021.

Mobiny, A. and Nguyen, H. V. Fast capsnet for lung cancer screening. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 741–749. Springer, 2018.

Mooney, P. Kaggle chest x-ray images (pneumonia) dataset, 2020.

Öksüz, C., Urhan, O., and Güllü, M. K. Covid-19 detection with severity level analysis using the deep features, and wrapper-based selection of ranked features. *Concurrency and Computation: Practice and Experience*, pp. e6802, 2021.

Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., and Acharya, U. R. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine*, 121:103792, 2020.

Sabour, S., Frosst, N., and Hinton, G. E. Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, 30, 2017.

Serte, S. and Demirel, H. Deep learning for diagnosis of covid-19 using 3d ct scans. *Computers in Biology and Medicine*, 132:104306, 2021.

Stogiannos, N., Fotopoulos, D., Woznitza, N., and Malamateniou, C. Covid-19 in the radiology department: what radiographers need to know. *Radiography*, 26(3):254–263, 2020.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.

Toraman, S., Alakus, T. B., and Turkoglu, I. Convolutional capsnet: A novel artificial neural network approach to detect covid-19 disease from x-ray images using capsule networks. *Chaos, Solitons & Fractals*, 140:110122, 2020.

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 24–25, 2020a.

Wang, L., Lin, Z. Q., and Wong, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1):1–12, 2020b.

Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *European Radiology*, 31(8):6096–6104, 2021a.

Wang, Z., Xiao, Y., Li, Y., Zhang, J., Lu, F., Hou, M., and Liu, X. Automatically discriminating and localizing covid-19 from community-acquired pneumonia on chest x-rays. *Pattern Recognition*, 110:107613, 2021b.