Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2021.015807

ARTICLE

ANC: Attention Network for COVID-19 Explainable Diagnosis Based on Convolutional Block Attention Module

Yudong Zhang[1,3,*], Xin Zhang[2,*] and Weiguo Zhu[1]

[1]Jiangsu Key Laboratory of Advanced Manufacturing Technology, Huaiyin Institute of Technology, Huai'an, 223003, China

[2]Department of Medical Imaging, The Fourth People's Hospital of Huai'an, Huai'an, 223002, China

[3]School of Informatics, University of Leicester, Leicester, LE1 7RH, UK

[*]Corresponding Authors: Yudong Zhang. Email: yudongzhang@ieee.org; Xin Zhang. Email: hasyzx@njmu.edu.cn

Received: 15 January 2021 Accepted: 24 February 2021

ABSTRACT

Aim: To diagnose COVID-19 more efficiently and more correctly, this study proposed a novel attention network for COVID-19 (ANC). **Methods:** Two datasets were used in this study. An 18-way data augmentation was proposed to avoid overfitting. Then, convolutional block attention module (CBAM) was integrated to our model, the structure of which is fine-tuned. Finally, Grad-CAM was used to provide an explainable diagnosis. **Results:** The accuracy of our ANC methods on two datasets are 96.32% ± 1.06%, and 96.00% ± 1.03%, respectively. **Conclusions:** This proposed ANC method is superior to 9 state-of-the-art approaches.

KEYWORDS

Deep learning; convolutional block attention module; attention mechanism; COVID-19; explainable diagnosis

1 Introduction

COVID-19 (also known as coronavirus) pandemic is an ongoing infectious disease [1] caused by severe acute respiratory syndrome (SARS) coronavirus 2 [2]. As of 7/Feb/2021, there are over 106.22 m confirmed cases and over 2.31 m deaths attributed to COVID-19 (See Fig. 1). The main symptoms of COVID-19 are a low fever, a new and ongoing cough, a loss or change to taste and smell [3].

In UK, the vaccines approved were developed by Pfizer/BioNTech, Oxford/AstraZeneca, and Moderna [4]. The joint committee on vaccination and immunization (JCVI) [5] determines the order in which people will be offered the vaccine. Currently, people aged over 80, people living or working in the care homes, and health care providers are being offered.

Two COVID-19 diagnosis methods are available. The first method is viral testing to test the existance of viral RNA fragments [6]. The shortcomings are two folds: (i) the swab may be contaminated and (ii) it needs to wait from several hours to several days to get the results. The other method is chest imaging. The chest computed tomography (CCT) [7] is one of the best chest imaging techniques. CCT is operator independent. Besides, it provides the highest sensitivity compared to ultrasound and X-ray [8]. It provides real 3D volumetric image of the chest region [9].

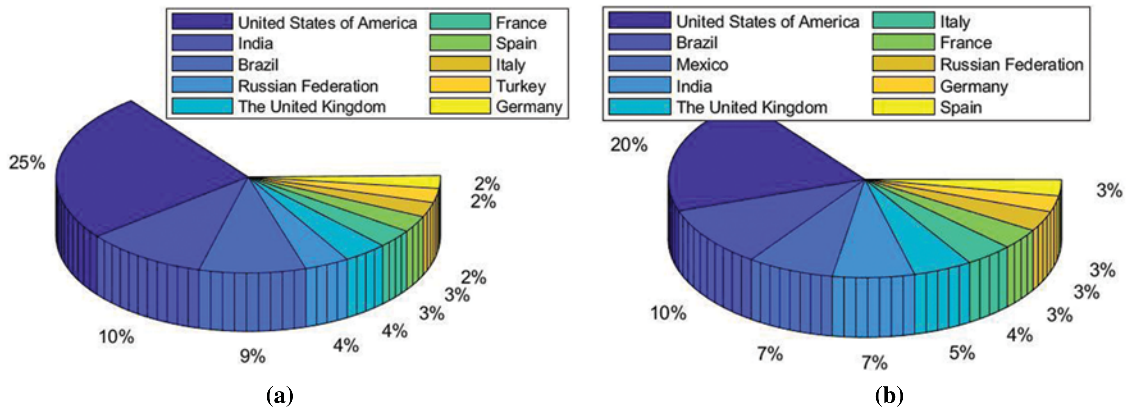


Figure 1: The cumulative data till 7/Feb/2021. (a) Cumulative cases, (b) cumulative deaths

Nevertheless, the manual labelling by human experts are time-consuming, tedious, labor-intensive, and easily influenced by human factors (fatiguing or lethargic emotions). In contrast to manual labelling, computer vision techniques [10] are now gaining promising results on automatic labelling of COVID-19 and other medical images with the help of artificial intelligence (AI) [11].

For non-COVID-19 images, Lu [12] proposed a bat algorithm-based extreme learning machine (BA-ELM) approach, and applied it to pathological brain detection. Lu [13] presented a radial basis function (RBFNN) to recognize pathological brain types. Fulton et al. [14] employed ResNet-50 for identifying Alzheimer's disease. Guo et al. [15] utilized ResNet-18 to identify Thyroid ultrasound plane images. Although those above methods were not directly developed for COVID-19 diagnosis, they are chosen and used as comparison methods in this study.

For COVID-19 images, Yu [16] proposed GoogleNet-COD model. The authors first replaced the last two layers with four new layers, which included the dropout layer, two fully connected layers and the output layer. Satapathy [17] proposed a five-layer deep convolutional neural network with stochastic pooling (DCNN-SP). Yao [18] combined wavelet entropy (WE) and biogeography-based optimization (BBO) techniques. Wu [19] used wavelet Renyi entropy (WRE) to extract features from chest CT images. Li et al. [20] presented a COVID-19 detection neural network (COVNet). Akram et al. [21] proposed a four-step procedure to handle COVID-19: (i) data collection & normalization; (ii) feature extraction; (iii) feature selection;, and (iv) feature classification. Khan et al. [22] used one class kernel extreme learning machine to predict COVID-19 pneumonia. Khan et al. [23] used DenseNet and firefly algorithm for classification of positive COVID-19 CT scans.

To further improve the COVID-19 diagnosis performance, this paper proved a novel AI model, attention network for COVID (ANC), for providing an explainable diagnosis. Here "attention" means it can tell neural network which region should focus. Compared to ordinary neural networks, the advantages of ANC are four folds:

(i) A novel 18-way data augmentation was proposed to avoid overfitting;
(ii) Convolutional block attention module was integrated so our model can infer attention maps;

(iii) Our model was fine-tuned, and its performances were better than 9 state-of-the-art approaches;
(iv) Grad-CAM was used to provide an explainable heatmap so the users can understand our model.

2 Dataset

Two datasets are used in this study. The first dataset contains 148 COVID-19 images and 148 healthy control (HC) images [19]. The second dataset contains 320 COVID-19 images and 320 HC images [24]. Tab. 1 provides the descriptions of these two COVID-19 CCT datasets, where $a+b$ stands for a COVID-19 subject/images and b HC subjects/images.

Table 1: Descriptions of two COVID-19 CCT datasets

	No. of subjects	No. of images
Dataset-1 [19]	$66+66$	$148+148$
Dataset-2 [24]	$142+142$	$320+320$

Preprocessing were applied to all the images. Let R_0 and R stand for the raw dataset (Dataset-1 or Dataset-2) and the final preprocessed set, and R_1, R_2, R_3 denote three temporary sets. The flowchart of our pre-processing is displayed in Fig. 2.

Th raw dataset set $R_0=\{r_0(k),\ k=1,2,\ldots,|R|\}$, where $|R|$ means the number of images. The size of each image is $size[r_0(k)]=1024\times1024\times3$. Although the raw images appear grayscale, but they are stored in RGB format at hospitals' servers.

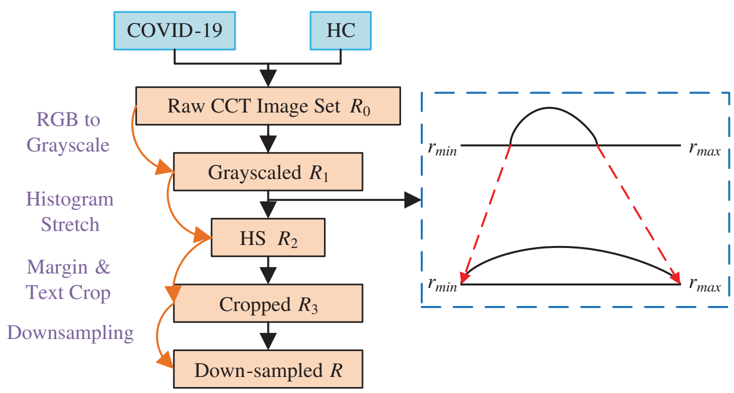


Figure 2: Flowchart of preprocessing

First, we need to grayscale all those raw images. The grayscale transformation is described as

$$gray=0.2989*R+0.5870*G+0.1140*B, \tag{1}$$

which utilized the standard RGB to grayscale transformation [25].

Second, the histogram stretching (HS) was used to improve the contrast of all grayscaled images $R_1 = \{r_1(k)\}$. For the k-th image $r_1(k)$, suppose its upper bound (UB) and lower bound (LB) grayscale values of $r_1(k)$ are $r_1^U(k)$ and $r_1^L(k)$. They can be obtained as

$$\begin{cases} r_1^U(k) = \max\limits_{x=1}^{W1} \max\limits_{y=1}^{H1} r_1(x, y \mid k) \\ r_1^L(k) = \min\limits_{x=1}^{W1} \min\limits_{y=1}^{H1} r_1(x, y \mid k) \end{cases}, \tag{2}$$

where (x, y) are the indexes of width and height dimension, respectively. $(W1, H1)$ stand for the width and height of image r_1, respectively.

The new HS-enhanced image $r_2(k)$ is calculated as

$$r_2(k) = \frac{r_1(k) - r_1^L(k)}{r_1^r(k)}, \tag{3}$$

where $r_1^r(k)$ is the grayscale range of the image $r_1(k)$. It is defined as

$$r_1^r(k) = r_1^U(k) - r_1^L(k), \tag{4}$$

The HS-enhanced image will occupy the full grayscale range as $[r_{min}, r_{max}]$, where r_{min} and r_{max} stand for the minimum and maximum gray values, respectively.

Third, we cropped the texts at the right region and the check-up bed at the bottom region. The crop values (e_1, e_2, e_3, e_4) are pixels to be cropped from four directions: top, left, bottom, and right, respectively. The cropped image $r_3(k)$ can be written as

$$r_3(k) = r_2(x', y'), \tag{5}$$

where

$$\begin{cases} x' = e_1 : H3 - e_3 \\ y' = e_2 : W3 - e_4 \end{cases}, \tag{6}$$

where $(W3, H3)$ stand for the weight and height of any image r_3, respectively. The range $a{:}b$ stands for the range from the integer a to the integer b.

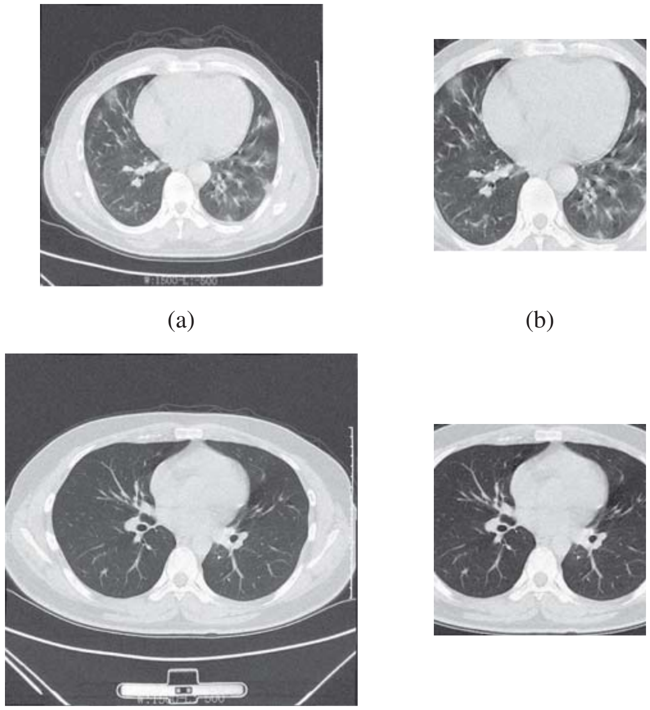
Finally, downsampling was carried out to further reduce the image size and remove redundant information. Suppose the final size is (W, H), the final image,

$$r(k) = f_{ds}[r_3(k), (W, H)], \tag{7}$$

where f_{ds} is the downsampling function defined as

$$b = f_{ds}[a, (W_b, H_b)], \quad size(a) = (W_a, H_a), \quad size(b) = (W_b, H_b), \quad W_b < W_a, \quad H_b < H_a. \tag{8}$$

Figs. 3a and 3b show the raw and pre-processed images of a COVID-19 case, while Figs. 3c and 3d illustrates the raw and pre-processed images of a HC case. As can be observed from Fig. 3, the pre-processed images have better contrast, remove the irrelevant information, down-sample to a smaller size, and take less storage than the raw images. Tab. 2 itemizes the abbreviation list, which will help readers to understand the following sections.



(a) (b)

(c) (d)

Figure 3: Raw images and pre-processed images selected from our datasets. (a) Raw COVID-19 image, (b) pre-processed image of (a), (c) raw HC image, (d) pre-processed image of (c)

3 Methodology

3.1 18-Way Data Augmentation

Data augmentation (DA) is an important tool over the training set to avoid overfitting of classifiers, and to overcome the small-size dataset problem. Recently, Wang [26] proposed a novel 14-way data augmentation (DA), which used seven different DA techniques to the preprocessed image $r(x, y \mid k),\ x = 1, \ldots, W, y = 1, \ldots, H, k = 1, \ldots, |R|$ and its horizontal mirrored image $r^M(x, y \mid k)$, respectively.

This study enhances the 14-way DA method [26] to 18-way DA, by adding two new DA methods: salt-and-pepper noise (SAPN) and speckle noise (SN) on both $r(x, y \mid k)$ and $r^M(x, y \mid k)$. In our future studies, we will research on including more types of DAs. Fig. 4 displays the illustration example of using two new DAs.

For a given image $r(x, y)$, the SAPN altered image [27] is defined as $r^{SAPN}(x, y)$ with its values are set as

$$\begin{cases} \mathrm{P}\left(r^{SAPN} = r\right) = 1 - \beta_d^{sa}, \\ \mathrm{P}\left(r^{SAPN} = r_{\min}\right) = \dfrac{\beta_d^{sa}}{2}, \\ \mathrm{P}\left(r^{SAPN} = r_{\max}\right) = \dfrac{\beta_d^{sa}}{2}, \end{cases} \tag{9}$$

where β_d^{sa} stands for noise density, and P the probability function. $r_{\min}$ and $r_{\max}$ correspond to black and white colors, respectively. The definition of r_{min} and r_{max} can be seen at Section 2.

Table 2: Abbreviation list

Meanings	Abbreviations
AI	Artificial intelligence
AM	Activation map
AP	Average pooling
AUC	Area under curve
BN	Batch normalization
CAM	Channel attention module
CB	Convolutional block
CBAM	Convolutional block attention module
CNN	Convolutional neural network
DA	Data augmentation
DL	Deep learning
FCL	Fully-connected layer
FMI	Fowlkes–Mallows index
HC	Healthy control
JCVI	Joint committee on vaccination and immunization
MCC	Matthews correlation coefficient
MLP	Multi-layer perceptron
MP	Max pooling
MSD	Mean and standard deviation
NA	No attention
ReLU	Rectified linear unit
ROC	Receiver operating characteristic
SAM	Spatial attention module
SAPN	Salt-and-pepper noise
SARS	Severe acute respiratory syndrome
SN	Speckle noise

(a) (b) (c)

Figure 4: Illustration of two new DAs. (a) Raw image, (b) SAPN altered, (c) SN altered

On the other side, the SN altered image [28] is defined as

$$r^{SN}(x, y) = r(x, y) + \mathbb{N} \times r(x, y), \tag{10}$$

where $\mathbb{N}$ is uniformly distributed random noise, of which the mean and variance are symbolized as β_m^s and β_v^s, respectively.

Let N_a stands for the number of DA techniques to the preprocessed image $r(x, y \mid k)$, and N_b stands for the number of new generated images for each DA. Thus, our $(2 \times N_a)$-way DA algorithm is a four-step algorithm depicted below:

First, N_a geometric/photometric/noise-injection DA transforms are utilized on preprocessed image $r(x, y)$, as shown in Fig. 5. We use $h_{(m)}^{DA}$, $m = 1, \ldots, N_a$ to denote each DA operation. Note each DA operations $h_{(m)}^{DA}$ will yield N_b new images. So, for a given image $r(x, y)$, we will yield N_a different data set $h_{(m)}^{DA}[r(x, y)]$, $m = 1, \ldots, N_a$, and each dataset has N_b new images.

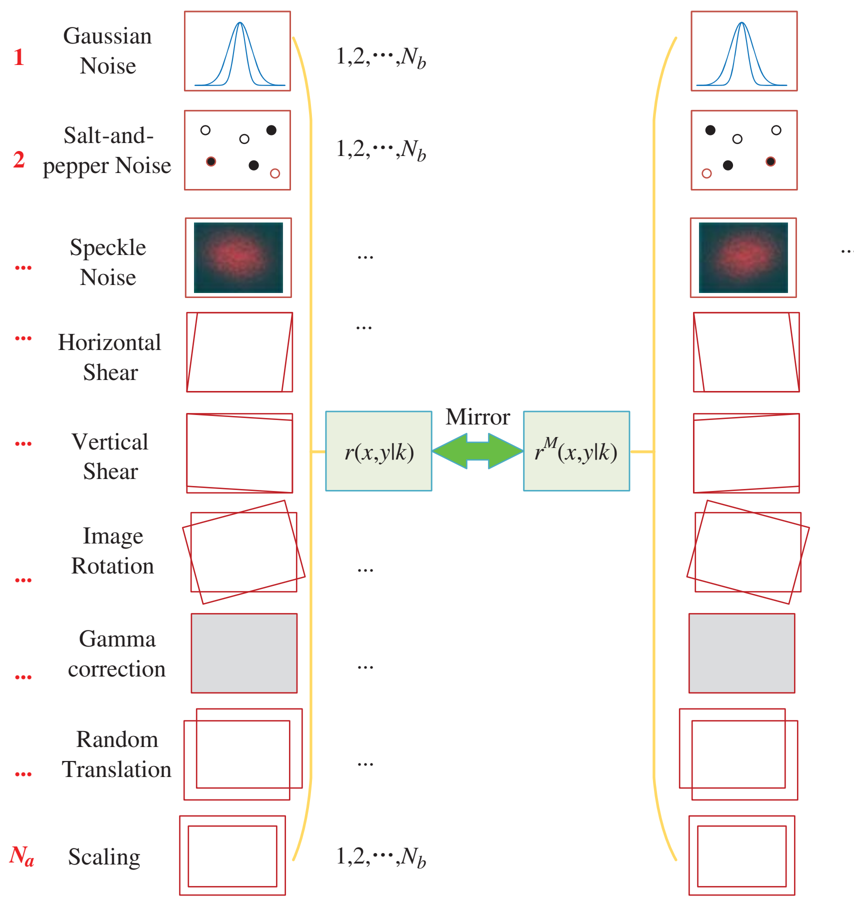


Figure 5: Schematic of proposed 18-way DA

Second, horizontal mirror image is generated as

$$r^M(x, y) = h_M[q(x, y)], \tag{11}$$

where h_M means horizontal mirror function.

Third, all the N_a different DA methods are carried out on the mirror image $q^M(x, y)$, and generate N_a different dataset $h_{(m)}^{DA}\left[r^M(x, y)\right]$, $m = 1, \ldots, N_a$.

Four, the raw image $r(x, y)$, the mirrored image $r^M(x, y)$, all the above N_a-way results of preprocessed image $h_{(m)}^{DA}[r(x, y)]$, $m = 1, \ldots, N_a$, and N_a-way DA results of horizontal mirrored image $h_{(m)}^{DA}\left[r^M(x, y)\right]$, $m = 1, \ldots, N_a$, are fused together using concatenation function h_{CON}.

The final combined dataset is defined as Γ

$$r(x, y) \mapsto \Gamma = h_{con}\begin{Bmatrix} r(x, y) & r^M(x, y) \\ \underbrace{h_{(1)}^{DA}[r(x, y)]}_{N_b} & \underbrace{h_{(1)}^{DA}\left[r^M(x, y)\right]}_{N_b} \\ \ldots & \ldots \\ \underbrace{h_{(N_a)}^{DA}[r(x, y)]}_{N_b} & \underbrace{h_{(N_a)}^{DA}\left[r^M(x, y)\right]}_{N_b} \end{Bmatrix} \tag{12}$$

Therefore, one image $r(x, y)$ will generate

$$|\Gamma| = 2 \times N_a \times N_b + 2 \tag{13}$$

images (including the original image $r(x, y)$). Algorithm 1 shows the pseudocode of this proposed 18-way data augmentation on one image $r(x, y)$.

Algorithm 1: Pseudocode of proposed 18-way DA on one training image $r(x, y)$

Inputs Import raw training image $r(x, y)$.
Step 1 N_a geometric or photometric or noise-injection DA transforms are utilized on raw image $r(x, y)$,
We obtain $h_{(m)}^{DA}[r(x, y)]$, $m = 1, \ldots, N_a$.
Step 2 Horizontal mirror image is generated as $r^M(x, y) = h_M[r(x, y)]$. See Eq. (11).
Step 3 Other N_a-way data augmentation methods are carried out on $r^M(x, y)$,
we obtain $h_{(m)}^{DA}\left[r^M(x, y)\right]$, $m = 1, \ldots, N_a$.
Step 4 The raw image, the mirrored image, all the above N_a-way DA results of raw image, and N_a-way DA results of horizontal mirrored image are combined together. See Eq. (12).
Output A new dataset Γ is generated with $|\Gamma| = 2 \times N_a \times N_b + 2$. See Eq. (13).

3.2 Convolutional Block Attention Module

Deep learning (DL) has gained successes in prediction/classification tasks. There are many DL structures, such as deep neural network [29], deep belief network, convolutional neural network (CNN) [30], recurrent neural network [31], graph neural network, etc. Among all those DL structures, CNN is particularly suitable for analyzing visual images.

To further improve the performance of CNN, a lot of researches are done with respects to either depth, or width, or cardinality of CNN. In recent times, Woo et al. [32] proposed a novel convolutional block attention module (CBAM), which improves the traditional convolutional block (CB) by integrating attention mechanism. There are many successful applications of CBAM.

For example, Mo et al. [33] combined self-attention and CBAM. They proposed a light-weight dual-path attention block. Chen et al. [34] proposed a 3D spatiotemporal convolutional neural network with CBAM for micro-expression recognition.

Fig. 6a displays the structure of a traditional CB. The output of previous block was sent to n-repetitions of convolution layer, batch normalization (BN), rectified linear unit (ReLU) layer, and followed by a pooling layer. The output is called activation map (AM), symbolized as $\mathbf{D} \in \mathbb{R}^{C \times H \times W}$, where (C, H, W) stands for the sizes of channel, height, and width, respectively.

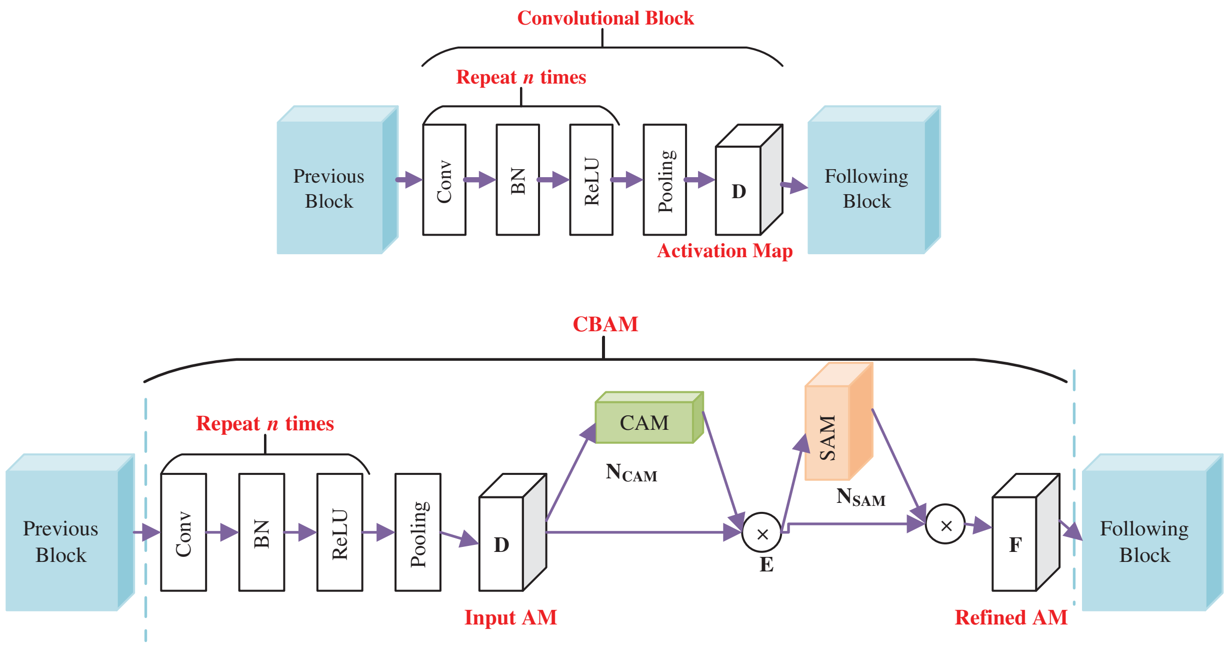


Figure 6: Structural comparison with CB against CBAM. (a) Traditional CB, (b) our CBAM

In contrast to Figs. 6a and 6b displays the structure of CBAM, in which two modules: channel attention module (CAM) and spatial attention module (SAM) are added to refine the activation map $\mathbf{D}$. Suppose the CBAM applies a 1D CAM $\mathbf{N_{CAM}} \in \mathbb{R}^{C \times 1 \times 1}$ and a 2D SAM $\mathbf{N_{SAM}} \in \mathbb{R}^{1 \times H \times W}$ in sequence to the input $\mathbf{D}$. Thus, we have the channel-refined activation map

$$\mathbf{E} = \mathbf{N_{CAM}}(\mathbf{D}) \otimes \mathbf{D} \tag{14}$$

And the final refined AM

$$\mathbf{F} = \mathbf{N_{SAM}}(\mathbf{E}) \otimes \mathbf{E} \tag{15}$$

where $\otimes$ stands for the element-wise multiplication. This refined AM $\mathbf{F}$ will be sent to the next block.

Note if the two operands are not with the same dimension, then the values are broadcasted (copied) in such ways the spatial attentional values are broadcasted along the channel dimension, and the channel attention values are broadcasted along the spatial dimension.

3.3 Channel Attention Module

We define the CAM here. Both max pooling (MP) [35] z_{mp} and average pooling (AP) [36] z_{mp} are utilized, generating two features $\mathbf{G_{ap}}$ and $\mathbf{G_{mp}}$ as shown in Fig. 7.

$$\begin{cases} \mathbf{G_{ap}} = z_{ap}\left(\mathbf{D}\right) \\ \mathbf{G_{mp}} = z_{mp}\left(\mathbf{D}\right) \end{cases} \tag{16}$$

Both $\mathbf{G_{ap}}$ and $\mathbf{G_{mp}}$ are then forwarded to a shared multi-layer perceptron (MLP) to generate the output features, which are then merged using element-wise summation $\oplus$. The merged sum is finally sent to the sigmoid function σ. Mathematically

$$\mathbf{N_{CAM}}\left(\mathbf{D}\right) = \sigma\left\{MLP\left[\mathbf{G_{ap}}\right] \oplus MLP\left[\mathbf{G_{mp}}\right]\right\} \tag{17}$$

To reduce the parameter resources, the hidden size of MLP is set to $\mathbb{R}^{C/r\times1\times1}$, where r is defined as the reduction ratio. Suppose $\mathbf{W_0} \in \mathbb{R}^{C/r\times C}$ and $\mathbf{W_1} \in \mathbb{R}^{C\times C/r}$ stand for the MLP weights, respectively, we can rephrase Eq. (17) as

$$\mathbf{N_{CAM}}\left(\mathbf{D}\right) = \sigma\left\{\mathbf{W_1}\left[\mathbf{W_0}\left(\mathbf{G_{ap}}\right)\right] \oplus \mathbf{W_1}\left[\mathbf{W_0}\left(\mathbf{G_{mp}}\right)\right]\right\} \tag{18}$$

Note $\mathbf{W_0}$ and $\mathbf{W_1}$ are shared by both $\mathbf{G_{ap}}$ and $\mathbf{G_{mp}}$. Fig. 7 shows the flowchart of CAM. Note that squeeze-and-excitation (SE) [37] method is similar to CAM.

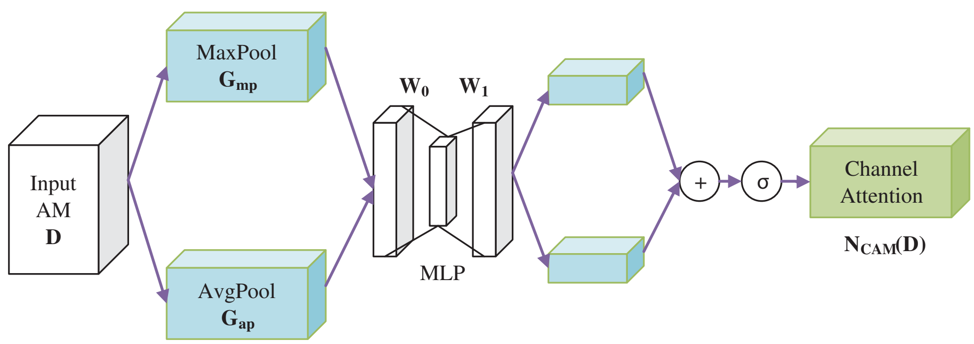


Figure 7: Flowchart of CAM

3.4 Spatial Attention Module

Now we define the spatial attention module (SAM) as shown in Fig. 8. The spatial attention module $\mathbf{N_{SAM}}$ is a complementary step to the previous channel attention module $\mathbf{N_{CAM}}$. The average pooling z_{ap} and max pooling z_{mp} are applied again to the channel-refined activation map $\mathbf{E}$, and we get

$$\begin{cases} \mathbf{H_{ap}} = z_{ap}\left(\mathbf{E}\right) \\ \mathbf{H_{mp}} = z_{mp}\left(\mathbf{E}\right) \end{cases} \tag{19}$$

Both $\mathbf{H_{ap}}$ and $\mathbf{H_{mp}}$ are two dimensional AMs: $\mathbf{H_{ap}} \in \mathbb{R}^{1\times H\times W} \wedge \mathbf{H_{mp}} \in \mathbb{R}^{1\times H\times W}$. They are concatenated together along the channel dimension as

$$\mathbf{H} = \text{concat}\left(\mathbf{H_{ap}}, \mathbf{H_{mp}}\right) \tag{20}$$

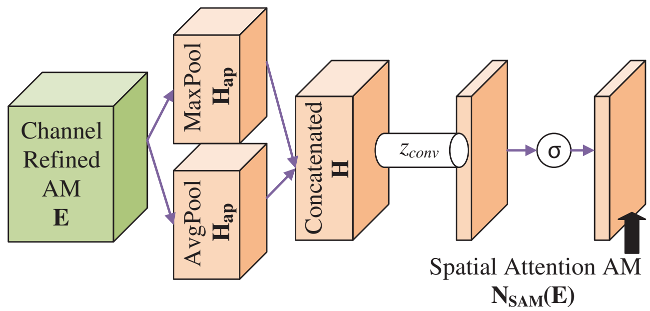


Figure 8: Flowchart of SAM

The concatenated activation map is then passed into a standard 7×7 convolution z_{conv} and followed by a sigmoid function σ. In all, we have

$$\mathbf{N_{SAM}}(\mathbf{E}) = \sigma\left\{z_{conv}[\mathbf{H}]\right\} \tag{21}$$

The $\mathbf{N_{SAM}}(\mathbf{E})$ is then element-wisely multiplied by $\mathbf{E}$ to get the final refined AE $\mathbf{F}$. See Eq. (15). The flowchart of SAM is drawn in Fig. 8.

3.5 Proposed Attention Network for COVID-19

In this study, we proposed a novel Attention Network for COVID-19 (ANC) based on CBAM (See Fig. 6b). The structure of this proposed ANC is determined by trial-and-error method. The variable n in each block varies, and we found the best values are chosen in the range from [1, 3]. We tested values larger than 3, which increase the computation burden, but the performances do not increase.

The structure of proposed shown below in Tab. 3, which is a 15-layer deep CNN. The Input I is a $256 \times 256 \times 1$-size preprocessed image. The first block "CBAM-1" contains 3 repetitions of 32 filters, each filter is with size of 3×3. The AM after CBAM-1 is symbolized as $T1$ with size of $128 \times 128 \times 32$. We can observe the size of AMs after CBAM-2, CBAM-3, CBAM-4, and CBAM-5 are $size(T2) = 64 \times 64 \times 32$, $size(T3) = 32 \times 32 \times 64$, $size(T4) = 16 \times 16 \times 64$, and $size(T5) = 8 \times 8 \times 128$, respectively. The output of CBAM-5 $T5$ is flattened to a vector symbolized as FL with size of $1 \times 1 \times 8192$.

Table 3: Structure of proposed ANC

Block index	Name	Kernel parameter	Variable of AM	Size of AM
1	Input		I	$256 \times 256 \times 1$
2	CBAM-1	$[3 \times 3, 32] \times 3$	$T1$	$128 \times 128 \times 32$
3	CBAM-2	$[3 \times 3, 32] \times 3$	$T2$	$64 \times 64 \times 32$
4	CBAM-3	$[3 \times 3, 64] \times 2$	$T3$	$32 \times 32 \times 64$
5	CBAM-4	$[3 \times 3, 64] \times 2$	$T4$	$16 \times 16 \times 64$
6	CBAM-5	$[3 \times 3, 128] \times 2$	$T5$	$8 \times 8 \times 128$
7	Flatten		FL	$1 \times 1 \times 8192$
8	FCL-1	$200 \times 8192,\ 200 \times 1$	$F1$	$1 \times 1 \times 200$
9	FCL-2	$50 \times 200,\ 50 \times 1$	$F2$	$1 \times 1 \times 50$
10	FCL-3	$2 \times 50,\ 2 \times 1$	$F3$	$1 \times 1 \times 2$
11	Softmax			$1 \times 1 \times 2$

Here $T5$ is used to generate heatmap by the gradient-weighted class activation mapping (Grad-CAM) [38] method, which shows explanations (i.e., heatmaps) which region the ANC model pays more attention to and how our ANC model makes prediction.

The variable FL is then sent to three consecutive fully-connected layers (FCLs). The neurons of those three FCLs are set to 200, 50, and 2, respectively. The output of those three FCLs are symbolized as $F1$, $F2$, and $F3$, respectively, as shown in Fig. 9. Finally, $F3$ is sent through a softmax layer to output the probabilities of all classes.

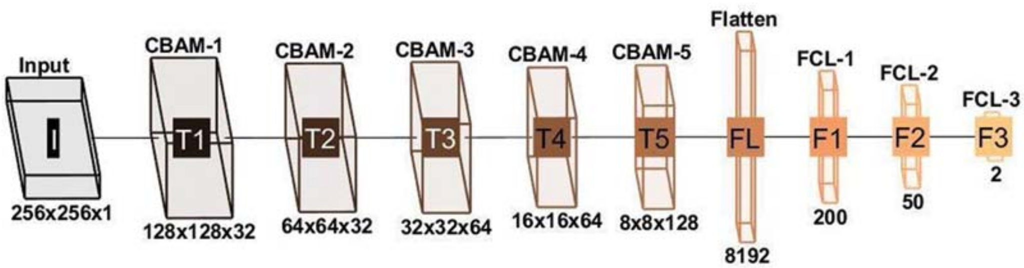


Figure 9: Variables and sizes of AMs of proposed ANC

3.6 Implementation and Evaluation

F-fold cross validation is employed on both datasets. Suppose confusion matrix $\mathbf{L}$ over r-th run and f-th fold is defined as

$$\mathbf{L}(r,f) = \begin{bmatrix} l_{11}(r,f) & l_{12}(r,f) \\ l_{21}(r,f) & l_{22}(r,f) \end{bmatrix} \tag{22}$$

where (l_{11}, l_{12}, l_{21}, l_{22}) stand for TP, FN, FP, and TN, respectively. f represents the index of trial, and r stands for the index of run ($1 \le r \le R$). At f-th trial, the f-th fold is used as test, and all the left folds are used as training, $1 \le f \le F$.

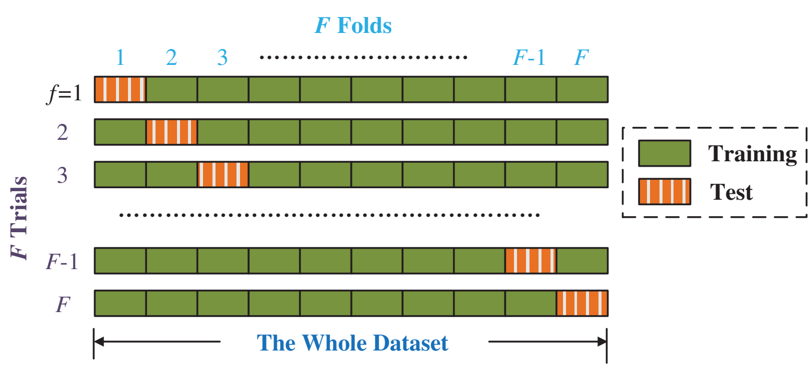


Figure 10: Diagram of one run of F-fold cross validation

Note that $\mathbf{L}(r,f)$ is calculated based on each test fold, and are then summarized across all F trials, as shown in Fig. 10. Afterwards, we get the confusion matrix at r-th run $\mathbf{L}(r)$

$$\mathbf{L}(r)=\sum_{f=1}^{F}\mathbf{L}(r,f) \tag{23}$$

Seven indicators $\vec{\tau}(r)$ could be computed based on the confusion matrix over r-th run $\mathbf{L}(r)$

$$\mathbf{L}(r)\mapsto\vec{\tau}(r)=[\tau_1(r),\tau_2(r),\ldots,\tau_7(r)], \tag{24}$$

where first four indicators are: τ_1 sensitivity, τ_2 specificity, τ_3 precision, and τ_4 accuracy. We have

$$\begin{cases}\tau_1(r)=\dfrac{l_{11}(r)}{l_{11}(r)+l_{12}(r)}\\[2ex]\tau_2(r)=\dfrac{l_{22}(r)}{l_{22}(r)+l_{21}(r)}\\[2ex]\tau_3(r)=\dfrac{l_{11}(r)}{l_{11}(r)+l_{21}(r)}\\[2ex]\tau_4(r)=\dfrac{l_{11}(r)+l_{22}(r)}{l_{11}(r)+l_{12}(r)+l_{21}(r)+l_{22}(r)}\end{cases} \tag{25}$$

τ_5 is F1 score.

$$\tau_5(r)=2\times\frac{\tau_3(r)\times\tau_1(r)}{\tau_3(r)+\tau_1(r)}=\frac{2\times l_{11}(r)}{2\times l_{11}(r)+l_{12}(r)+l_{21}(r)} \tag{26}$$

τ_6 is Matthews correlation coefficient (MCC)

$$\tau_6(r)=\frac{l_{11}(r)\times l_{22}(r)-l_{21}(r)\times l_{12}(r)}{\sqrt{[l_{11}(r)+l_{21}(r)]\times[l_{11}(r)+l_{12}(r)]\times[l_{22}(r)+l_{21}(r)]\times[l_{22}(r)+l_{12}(r)]}} \tag{27}$$

and τ_7 is the Fowlkes–Mallows index (FMI).

$$\tau_7(r)=\sqrt{\frac{l_{11}(r)}{l_{11}(r)+l_{21}(r)}\times\frac{l_{11}(r)}{l_{11}(r)+l_{21}(r)}} \tag{28}$$

There are two indicators τ_4 and τ_6 utilizing all the four basic measures $(l_{11},l_{12},l_{21},l_{22})$. Considering the range of τ_4 is $0\le\tau_4\le1$, and the range of τ_6 is $-1\le\tau_6\le+1$, we finally choose τ_6 as the most important indicator.

Above procedure is one run of F-fold cross validation. In the experiment, we run the F-fold cross validation R runs. The mean and standard deviation (MSD) of all seven indicators τ_m ($m = 1, \ldots, 7$) are calculated over all R runs.

$$\begin{cases} v_1\left(\tau_m\right) = \dfrac{1}{R} \times \displaystyle\sum_{r=1}^{R} \tau_m(r) \\ v_2\left(\tau_m\right) = \sqrt{\dfrac{1}{R-1} \times \displaystyle\sum_{r=1}^{R} \left|\tau_m\left(r\right) - v_1\left(\tau_m\right)\right|^2} \end{cases} \tag{29}$$

where v_1 stands for the mean value, and v_2 stands for the standard deviation. The MSDs are reported in the format of $v_1 \pm v_2$. The pseudocode of our evaluation is displayed in Algorithm 2. In the experiment, the evaluations of two datasets will be reported.

Algorithm 2: Pseudocode of evaluation

Input: prediction labels and actual labels.
for $r = 1: R$
  for $f = 1: F$
    Calculate the confusion matrix $\mathbf{L}(r, f)$. See Eq. (22).
  end
  Summarize the r-th run confusion $\mathbf{L}(r) = \sum_{f=1}^{F} \mathbf{L}(r, f)$. See Eq. (23).
  Evaluate seven indicators over r-th run and obtain $[\tau_1(r), \tau_2(r), \ldots, \tau_7(r)]$. See Eq. (24).
end
Output: MSD $v_1 \pm v_2$. See Eq. (29).

4 Experiments, Results, and Discussions

4.1 Parameter Setting

Tab. 4 shows the parameter setting. Here the crop values of top, bottom, left, and right are all set to 200 pixels. The size of preprocessed image is set to (256, 256). The minimum and maximum gray values are set to 0 and 255, respectively. The noise density of SAPN is set to $\beta_d^{sa} = 0.05$. The mean and variance of uniformly distributed random noise in SN are set to 0 and 0.05, respectively. Both the number of folds and the number of runs are set to 10.

Table 4: Parameter setting

Parameter	Value
(e_1, e_2, e_3, e_4)	(200, 200, 200, 200)
(W, H)	(256, 256)
(r_{min}, r_{max})	(0, 255)
β_d^{sa}	0.05
β_m^s	0
β_v^s	0.05
F	10
R	10

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)

Figure 11: N_a DA results on original preprocessed image. (a) Gaussian noise, (b) SAPN, (c) SN, (d) horizontal shear, (e) vertical shear, (f) rotation, (g) gamma correction, (h) random translation, (i) scaling

4.2 Results of Proposed 18-Way DA

Take Fig. 3b as an example, Fig. 11 displays the N_a results of this original image. Due to the page limit, the horizontal mirrored image and its corresponding N_a results are not displayed in Fig. 11. As it is calculated, one image will generate $2 \times N_a \times N_b + 2 = 542$ images.

4.3 Statistics of Proposed ANC

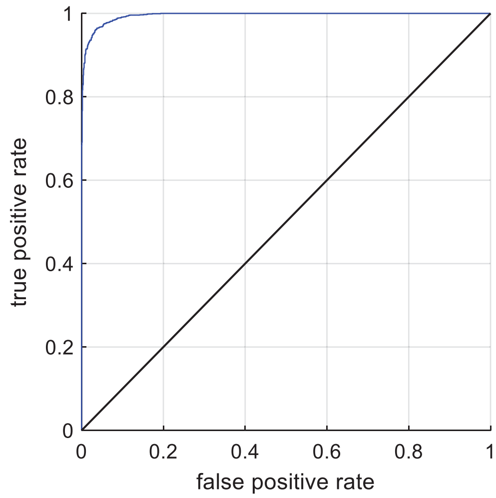
First, the statistical results on the first dataset and the second dataset are shown in Tabs. 5 and 6 respectively. We can observe that for Dataset-1, the sensitivity is 96.49 ± 1.18, the specificity is 96.15 ± 1.56, the precision is 96.18 ± 1.50, and accuracy is 96.32 ± 1.06, the F1 score is 96.33 ± 1.05, MCC is 92.65 ± 2.11, and FMI is 96.33 ± 1.05. For dataset-2, the performances are slightly lower than those of dataset-1. The sensitivity is 95.88 ± 1.28, the specificity is 96.13 ± 1.16, the precision is 96.12 ± 1.15 , the accuracy is 96.00 ± 1.03, F1 score is 95.99 ± 1.04, MCC is 92.01 ± 2.07, and FMI is 96.00 ± 1.04. Fig. 12 display the receiver operating characteristic (ROC) curves on datasets D1 and D2. The area under curve (AUC) of proposed ANC model on two datasets are 0.9951 and 0.9911, respectively.

Table 5: Statistical results of proposed ANC model on D1

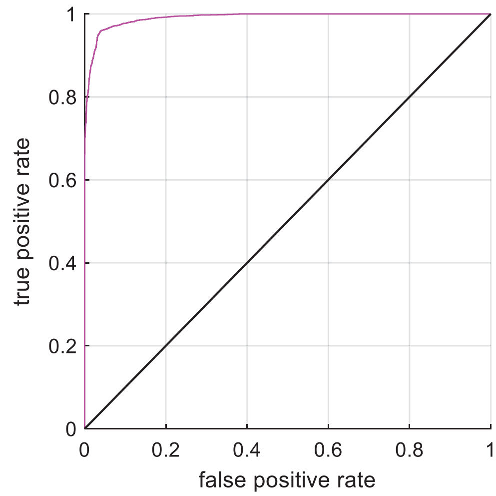
Run	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
1	95.95	96.62	96.60	96.28	96.27	92.57	96.27
2	96.62	97.30	97.28	96.96	96.95	93.92	96.95
3	97.30	94.59	94.74	95.95	96.00	91.93	96.01
4	96.62	93.92	94.08	95.27	95.33	90.57	95.34
5	93.92	96.62	96.53	95.27	95.21	90.57	95.21
6	95.95	95.27	95.30	95.61	95.62	91.22	95.62
7	96.62	95.95	95.97	96.28	96.30	92.57	96.30
8	96.62	97.97	97.95	97.30	97.28	94.60	97.28
9	96.62	94.59	94.70	95.61	95.65	91.23	95.66
10	98.65	98.65	98.65	98.65	98.65	97.30	98.65
MSD	96.49 ± 1.18	96.15 ± 1.56	96.18 ± 1.50	96.32 ± 1.06	96.33 ± 1.05	92.65 ± 2.11	96.33 ± 1.05

Table 6: Statistical results of proposed ANC model on D2

Run	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
1	97.19	98.13	98.11	97.66	97.65	95.32	97.65
2	93.44	95.94	95.83	94.69	94.62	89.40	94.63
3	96.88	95.00	95.09	95.94	95.98	91.89	95.98
4	95.00	95.31	95.30	95.16	95.15	90.31	95.15
5	94.06	94.69	94.65	94.38	94.36	88.75	94.36
6	95.94	97.81	97.77	96.88	96.85	93.77	96.85
7	96.88	96.88	96.88	96.88	96.88	93.75	96.88
8	96.56	95.94	95.96	96.25	96.26	92.50	96.26
9	96.56	96.25	96.26	96.41	96.41	92.81	96.41
10	96.25	95.31	95.36	95.78	95.80	91.57	95.80
MSD	95.88 ± 1.28	96.13 ± 1.16	96.12 ± 1.15	96.00 ± 1.03	95.99 ± 1.04	92.01 ± 2.07	96.00 ± 1.04



(a)



(b)

Figure 12: ROC curves of proposed ANC model (10 runs). (a) Dataset D1 (AUC = 0.9951), (b) dataset D2 (AUC = 0.9911)

4.4 Effect of Attention Mechanism

This ablation study removed attention blocks (CAM and SAM) from our ANC model. Here NA in Tab. 7 means no attention. Tab. 7 shows that removing attention blocks will reduce the performance significantly. For example, the MCCs, i.e., τ_6 values, of ANC-NA and ANC on the first dataset are 86.37% ± 1.92% and 92.65% ± 2.11%, respectively. For the second dataset, the MCC values of ANC-NA and ANC are 84.54% ± 1.93%, and 92.01% ± 2.07%, respectively.

Fig. 13 shows the error bars of different models, where D1 and D2 stand for the 1st and 2nd dataset, respectively, as introduced in Section 2. It clearly indicates that using attention mechanism can help improve the classification performances on both datasets.

Table 7: Comparison of models with attention against without attention

Data	Method	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
D1	ANC-NA	93.38 ± 1.23	92.97 ± 1.66	93.03 ± 1.52	93.18 ± 0.97	93.19 ± 0.95	86.37 ± 1.92	93.20 ± 0.95
	ANC	96.49 ± 1.18	96.15 ± 1.56	96.18 ± 1.50	96.32 ± 1.06	96.33 ± 1.05	92.65 ± 2.11	96.33 ± 1.05
D2	ANC-NA	92.06 ± 0.95	92.47 ± 1.20	92.44 ± 1.16	92.27 ± 0.97	92.25 ± 0.96	84.54 ± 1.93	92.25 ± 0.96
	ANC	95.88 ± 1.28	96.13 ± 1.16	96.12 ± 1.15	96.00 ± 1.03	95.99 ± 1.04	92.01 ± 2.07	96.00 ± 1.04

Note: (NA = no attention).

4.5 Comparison to State-of-the-Art Approaches

We compare our method "ANC" with 9 state-of-the-art methods: ELM-BA [12], RBFNN [13], ResNet-50 [14], ResNet-18 [15], GoogleNet-COD [16], DCNN-SP [17], WE-BBO [18], WRE [19], and COVNet [20]. All the methods were evaluated via 10 runs of 10-fold cross validation. The MSD results $\vec{\tau}$ on ten runs of 10-fold cross validation are displayed in Tab. 8, where the definitions abbreviations can be found in Section 1. This proposed ANC outperforms all the other 9 comparison baseline methods in terms of all indicators.

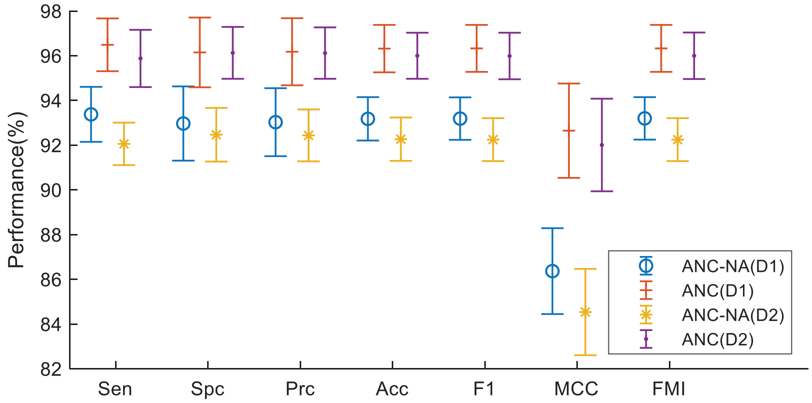
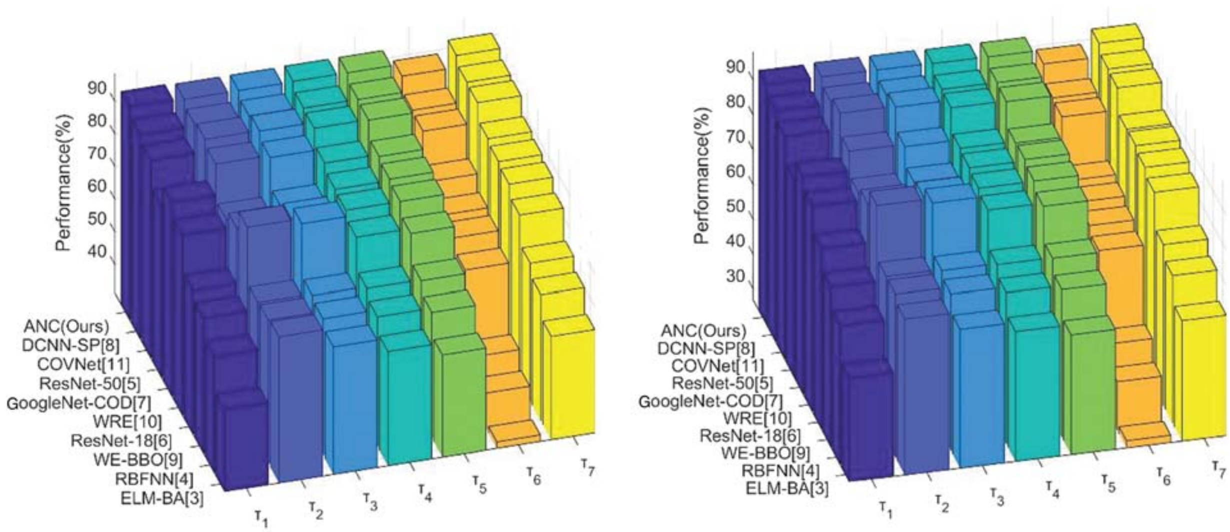


Figure 13: Effect of attention mechanism on two datasets. (D1: the first dataset; D2: the second dataset)

Table 8: Comparison with SOTA approaches in two datasets (unit: %)

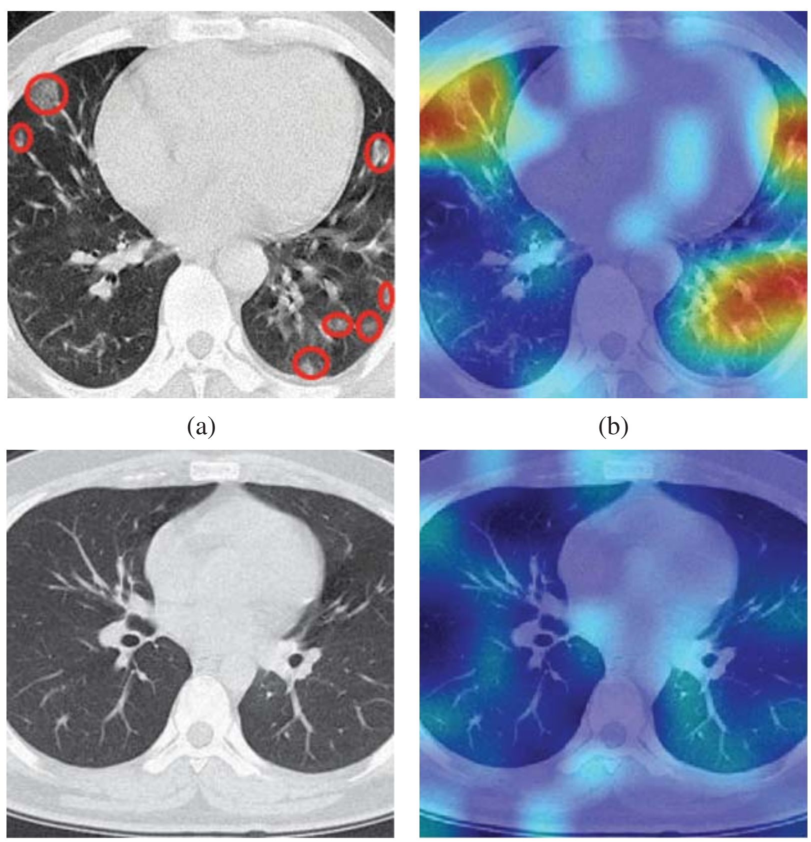
Dataset	Approach	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
D1	ELM-BA [12]	56.55 ± 2.99	76.22 ± 2.56	70.45 ± 1.61	66.39 ± 1.05	62.69 ± 1.76	33.46 ± 2.07	63.09 ± 1.60
	RBFNN [13]	66.89 ± 2.43	75.47 ± 2.53	73.23 ± 1.48	71.18 ± 0.80	69.88 ± 1.08	42.56 ± 1.61	69.97 ± 1.04
	ResNet-50 [14]	83.85 ± 2.31	93.85 ± 1.25	93.19 ± 1.21	88.85 ± 1.15	88.25 ± 1.34	78.12 ± 2.16	88.39 ± 1.28
	ResNet-18 [15]	75.34 ± 2.51	92.03 ± 1.71	90.48 ± 1.71	83.68 ± 1.01	82.18 ± 1.34	68.37 ± 1.88	82.54 ± 1.23
	GoogleNet-COD [16]	90.54 ± 2.16	82.77 ± 2.65	84.07 ± 1.93	86.66 ± 1.14	87.15 ± 1.06	73.59 ± 2.25	87.23 ± 1.07
	DCNN-SP [17]	93.78 ± 1.98	94.26 ± 1.47	94.25 ± 1.36	94.02 ± 1.10	94.00 ± 1.14	88.07 ± 2.18	94.01 ± 1.13
	WE-BBO [18]	72.84 ± 3.00	75.00 ± 1.99	74.47 ± 1.20	73.92 ± 1.18	73.61 ± 1.57	47.89 ± 2.34	73.63 ± 1.56
	WRE [19]	86.40 ± 3.00	85.81 ± 3.14	86.14 ± 3.03	86.12 ± 2.75	86.16 ± 2.77	72.42 ± 5.55	86.15 ± 2.76
	COVNet [20]	91.76 ± 2.74	95.61 ± 0.91	95.45 ± 0.82	93.68 ± 1.15	93.54 ± 1.28	87.47 ± 2.18	93.57 ± 1.25
	ANC (Ours)	96.49 ± 1.18	96.15 ± 1.56	96.18 ± 1.50	96.32 ± 1.06	96.33 ± 1.05	92.65 ± 2.11	96.33 ± 1.05
D2	ELM-BA [12]	56.91 ± 1.21	71.94 ± 2.17	67.01 ± 1.52	64.42 ± 0.88	61.53 ± 0.77	29.19 ± 1.88	61.74 ± 0.77
	RBFNN [13]	66.97 ± 2.42	74.00 ± 2.14	72.07 ± 1.25	70.48 ± 0.81	69.39 ± 1.18	41.10 ± 1.60	69.46 ± 1.12
	ResNet-50 [14]	83.66 ± 2.43	89.84 ± 0.82	89.18 ± 0.73	86.75 ± 1.13	86.31 ± 1.34	73.67 ± 2.17	86.36 ± 1.30
	ResNet-18 [15]	78.88 ± 2.57	89.28 ± 0.90	88.05 ± 0.79	84.08 ± 1.14	83.19 ± 1.44	68.55 ± 2.10	83.32 ± 1.36
	GoogleNet-COD [16]	89.44 ± 1.59	82.91 ± 1.64	83.98 ± 1.16	86.17 ± 0.67	86.61 ± 0.68	72.53 ± 1.32	86.66 ± 0.68
	DCNN-SP [17]	93.28 ± 1.50	94.00 ± 1.56	93.96 ± 1.54	93.64 ± 1.42	93.62 ± 1.42	87.29 ± 2.83	93.63 ± 1.42
	WE-BBO [18]	72.94 ± 0.96	73.97 ± 1.02	73.70 ± 0.79	73.45 ± 0.69	73.31 ± 0.71	46.91 ± 1.38	73.32 ± 0.71
	WRE [19]	85.94 ± 1.68	84.75 ± 2.42	84.96 ± 2.16	85.34 ± 1.81	85.44 ± 1.74	70.71 ± 3.61	85.44 ± 1.73
	COVNet [20]	91.00 ± 1.89	95.72 ± 0.93	95.52 ± 0.91	93.36 ± 0.91	93.19 ± 0.98	86.84 ± 1.76	93.23 ± 0.96
	ANC (Ours)	95.88 ± 1.28	96.13 ± 1.16	96.12 ± 1.15	96.00 ± 1.03	95.99 ± 1.04	92.01 ± 2.07	96.00 ± 1.04

Fig. 14 shows the 3D bar plot of our ANC method and 9 state-of-the-art algorithms. All the algorithms are sorted in terms of τ_6, i.e., MCC indicator. As is shown here, the best is our ANC. The second best and the third best are DCNN-SP [17] and COVNet [20], respectively. The reason why our ANC model gives the best results are three folds: (i) The proposed 18-way DA help resist overfitting; (ii) The CBAM help build our ANC model integrate attention mechanism; and (iii) The structure of ANC is fine-tuned (See Tab. 3).



(a) (b)

Figure 14: Comparison to state-of-the-art approaches over two datasets (D1 and D2). (a) D1, (b) D2



(a) (b)

(c) (d)

Figure 15: Manual delineation and heatmap results of COVID-19 sample image. (a) Manual delineation of COVID-19, (b) heatmap of (a), (c) HC image, (d) heatmap of (c)

4.6 Explainable Heatmap

Fig. 15 shows the manual delineation and heatmap results of Fig. 3. Figs. 15a–15c show the delineation, where the healthy control image has no lesion (Fig. 15c), so the radiologist does not delineate the lesion regions in Fig. 15c. Figs. 15b–15d show the corresponding heatmaps, where we can observe that this proposed ANC model can find the COVID-19 related regions correctly.

5 Conclusion

This paper proposed a novel attention network for COVID-19 (ANC) model. The classification results of ANC model are better than 9 state-of-the-art approaches in terms of seven indicators: sensitivity, specificity, precision, accuracy, F1 score, MCC, and FMI.

There are two disadvantages of our proposed method: (i) The dataset is relatively small. We expect to expand the dataset to have more than 1k images. (ii) Graph neural network technologies will be attempted to integrate to our model.

Funding Statement: This paper is partially supported by Open Fund for Jiangsu Key Laboratory of Advanced Manufacturing Technology (HGAMTL-1703); Guangxi Key Laboratory of Trusted Software (kx201901); Fundamental Research Funds for the Central Universities (CDLS-2020-03); Key Laboratory of Child Development and Learning Science (Southeast University), Ministry of Education; Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Medical Research Council Confidence in Concept Award, UK (MC_PC_17171); Hope Foundation for Cancer Research, UK (RM60G0680); British Heart Foundation Accelerator Award, UK.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Rauf, H. T., Lali, M. I. U., Khan, M. A., Kadry, S., Alolaiyan, H. et al. (2021). Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks. *Personal and Ubiquitous Computing, 9(3),* 674. DOI 10.1007/s00779-020-01494-0.
2. Marquez, R., Tolosa, L., Celis, M. T. (2021). Understanding COVID-19 effect on the U.S. supply chain of strategic products: Important factors, current situation, and future perspective. *Revista Ciencia e Ingeniería, 42(1),* 53–62. http://erevistas.saber.ula.ve/index.php/cienciaeingenieria/article/viewFile/16645/21921927789.
3. Avci, H., Karabulut, B. (2021). The relation between otolaryngology-specific symptoms and computed tomography findings in ambulatory care COVID-19 patients. *Ear Nose Throat Journal, 100(2),* 79–85. DOI 10.1177/0145561320975508.
4. Mahase, E. (2021). Covid-19: UK approves moderna vaccine to be given as two doses 28 days apart. *British Medical Journal, 372,* 1. DOI 10.1136/bmj.n74.
5. Hall, A. J. (2010). The United Kingdom joint committee on vaccination and immunisation. *Vaccine, 28,* 54–57. DOI 10.1016/j.vaccine.2010.02.034.
6. Goncalves, J., Koritnik, T., Mioc, V., Trkov, M., Boljesic, M. et al. (2021). Detection of SARS-CoV-2 RNA in hospital wastewater from a low COVID-19 disease prevalence area. *Science of the Total Environment, 755,* 7. DOI 10.1016/j.scitotenv.2020.143226.
7. Teichgraber, U., Malouhi, A., Ingwersen, M., Neumann, R., Reljic, M. et al. (2021). Ruling out COVID-19 by chest CT at emergency admission when prevalence is low: The prospective, observational SCOUT study. *Respiratory Research, 22(1),* 11. DOI 10.1186/s12931-020-01611-w.
8. Nayak, S. R., Nayak, D. R., Sinha, U., Arora, V., Pachori, R. B. (2021). Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. *Biomedical Signal Processing and Control, 64,* 12. DOI 10.1016/j.bspc.2020.102365.

9. Leonard-Lorant, I., Severac, F., Bilbault, P., Muller, J., Leyendecker, P. et al. (2021). Normal chest CT in 1091 symptomatic patients with confirmed COVID-19: Frequency, characteristics and outcome. *European Radiology, 6,* pp. 1–6. DOI 10.1007/s00330–020-07593-z.
10. Khemasuwan, D., Sorensen, J. S., Colt, H. G. (2020). Artificial intelligence in pulmonary medicine: Computer vision, predictive model and COVID-19. *European Respiratory Review, 29(157),* 16. DOI 10.1183/16000617.0181-2020.
11. Kabra, R., Singh, S. (2021). Evolutionary artificial intelligence based peptide discoveries for effective COVID-19 therapeutics. *Biochimica et Biophysica Acta (BBA)—Molecular Basis of Disease, 1867(1),* 13. DOI 10.1016/j.bbadis.2020.165978.
12. Lu, S. (2017). A pathological brain detection system based on extreme learning machine optimized by bat algorithm. *CNS & Neurological Disorders-Drug Targets, 16(1),* 23–29. DOI 10.2174/1871527315666161019153259.
13. Lu, Z. (2016). A pathological brain detection system based on radial basis function neural network. *Journal of Medical Imaging and Health Informatics, 6(5),* 1218–1222. DOI 10.1166/jmihi.2016.1901.
14. Fulton, L. V., Dolezel, D., Harrop, J., Yan, Y., Fulton, C. P. (2019). Classification of Alzheimer's disease with and without imagery using gradient boosted machines and ResNet-50. *Brain Sciences, 9(9),* Article ID: 212. DOI 10.3390/brainsci9090212.
15. Guo, M. H., Du, Y. Z. (2019). Classification of thyroid ultrasound standard plane images using ResNet-18 networks. *IEEE 13th International Conference on Anti-Counterfeiting, Security, and Identification, Proceedings of the International Conference on Anti-counterfeiting Security and Identification*, pp. 324–328. Xiamen, China.
16. Yu, X. (2020). Detection of COVID-19 by GoogLeNet-COD. *Lecture Notes in Computer Science, 12463,* 499–509. DOI 10.1007/978-3-030-60799-9_43.
17. Satapathy, S. C. (2021). A five-layer deep convolutional neural network with stochastic pooling for chest CT-based COVID-19 diagnosis. *Machine Vision and Applications, 32*.
18. Yao, X. (2020). COVID-19 Detection via wavelet entropy and biogeography-based optimization. In: Santosh, K. C., Joshi, A. (Eds.), *COVID-19: Prediction, decision-making, and its impacts*, pp. 69–76. Berlin/Heidelberg, Germany: Springer.
19. Wu, X. (2020). Diagnosis of COVID-19 by wavelet Renyi entropy and three-segment biogeography-based optimization. *International Journal of Computational Intelligence Systems, 13(1),* 1332–1344. DOI 10.2991/ijcis.d.200828.001.
20. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X. et al. (2020). Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy. *Radiology, 296(2),* E65–E71. DOI 10.1148/radiol.2020200905.
21. Akram, T., Attique, M., Gul, S., Shahzad, A., Altaf, M. et al. (2021). A novel framework for rapid diagnosis of COVID-19 on computed tomography scans. *Pattern Analysis and Applications, 382(2),* 727. DOI 10.1007/s10044-020-00950-0.
22. Khan, M. A., Kadry, S., Zhang, Y. D., Akram, T., Sharif, M. et al. (2021). Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine. *Computers & Electrical Engineering, 90,* 106960. DOI 10.1016/j.compeleceng.2020.106960.
23. Khan, M. A., Hussain, N., Majid, A., Alhaisoni, M., Bukhari, S. A. C. et al. (2021). Classification of positive COVID-19 CT scans using deep learning. *Computers, Materials & Continua, 66(3),* 2923–2938. DOI 10.32604/cmc.2021.013191.
24. Zhang, Y. D. (2020). A seven-layer convolutional neural network for chest CT based COVID-19 diagnosis using stochastic pooling. *IEEE Sensors Journal,* DOI 10.1109/JSEN.2020.3025855.
25. Sowmya, V., Govind, D., Soman, K. P. (2017). Significance of contrast and structure features for an improved color image classification system. *2017 IEEE International Conference on Signal and Image Processing Applications*. pp. 210–215. Kuching, Malaysia.
26. Wang, S. H. (2021). Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Information Fusion, 67,* 208–229. DOI 10.1016/j.inffus.2020.10.004.

27. Sagar, P., Upadhyaya, A., Mishra, S. K., Pandey, R. N., Sahu, S. S. et al. (2020). A circular adaptive median filter for salt and pepper noise suppression from MRI images. *Journal of Scientific and Industrial Research, 79(10),* 941–944. http://nopr.niscair.res.in/handle/123456789/55470.
28. Karimi, N., Taban, M. R. (2021). A convex variational method for super resolution of SAR image with speckle noise. *Signal Processing-Image Communication, 90,* 11. DOI 10.1016/j.image.2020.116061.
29. Jeong, Y., Lee, B., Han, J. H., Oh, J. (2021). Ocular axial length prediction based on visual interpretation of retinal fundus images via deep neural network. *IEEE Journal of Selected Topics in Quantum Electronics, 27(4),* 7. DOI 10.1109/jstqe.2020.3038845.
30. Lee, Y. Y., Halim, Z. A. (2020). Stochastic computing in convolutional neural network implementation: A review. *PeerJ Computer Science, 35,* Article ID: e309. DOI 10.7717/peerj–cs.309.
31. Hassan, A. A., El-Habrouk, M., Deghedie, S. (2020). Inverse kinematics of redundant manipulators formulated as quadratic programming optimization problem solved using recurrent neural networks: A review. *Robotica, 38(8),* 1495–1512. DOI 10.1017/s0263574719001590.
32. Woo, S., Park, J., Lee, J. Y., Kweon, I. S. (2018). CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision*. pp. 3–19. Munich, Germany: Springer.
33. Mo, J. Y., Xu, L. (2020). Weighted cluster-range loss and criticality-enhancement loss for speaker recognition. *Applied Sciences—Basel, 10(24),* 20. DOI 10.3390/app10249004.
34. Chen, B. Y., Zhang, Z. H., Liu, N., Tan, Y., Liu, X. et al. (2020). Spatiotemporal convolutional neural network with convolutional block attention module for micro-expression recognition. *Information-an International Interdisciplinary Journal, 11(8),* 14. DOI 10.3390/info11080380.
35. Akbas, G. E., Kozubek, M. (2020). Condensed U-Net (CU-Net): An improved u-net architecture for cell segmentation powered by 4 × 4 max-pooling layers. *17th International Symposium on Biomedical Imaging. IEEE International Symposium on Biomedical Imaging*. pp. 446–450. Iowa, IA, USA: IEEE.
36. Kumar, R. L., Kakarla, J., Isunuri, B. V., Singh, M. (2021). Multi-class brain tumor classification using residual network and global average pooling. *Multimedia Tools and Applications, 10,* DOI 10.1007/s11042-020-10335-4.
37. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E. H. (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(8),* 2011–2023. DOI 10.1109/TPAMI.2019.2913372.
38. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. et al. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision, 128(2),* 336–359. DOI 10.1007/s11263-019-01228-7.