2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-7281-8470-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMLA51294.2020.00040

RAM-Net: A Residual Attention MobileNet to Detect COVID-19 Cases from Chest X-Ray Images

Md Aminur Rab Ratul, Maryam Tavakol Elahi, Kun Yuan, WonSook Lee
School of Electrical Engineering and Computer Science (SITE)
University of Ottawa, Ottawa, Canada
{mratu076, mtava020, kyuan033, wslee}@uottawa.ca

***Abstract*—In the last century, we have passed two severe pandemics; the 1957 influenza (Asian flu) pandemic and the 1918 influenza (Spanish flu) pandemic with a high fatality rate. In the last few months, we have been again facing a new epidemic (COVID-19), which is a frighteningly high-risk disease and is globally threatening human lives. Among all attempts and presented solutions to tackle the COVID-19, a publicly available dataset of radiological imaging using chest radiography, also called chest X-ray (CXR) images, could efficiently accelerate the detection process of patients infected with COVID-19 through presented abnormalities in their chest radiography images. In this study, we have proposed a deep neural network (DNN), namely RAM-Net, a new combination of MobileNet with Dilated Depthwise Separable Convolution (DDSC), Residual blocks, and Attention augmented convolution. The network has been learned and validated using the COVIDx dataset, one of the most popular public datasets comprising the chest X-ray (CXR) images. Using this model, we could accurately identify the positive cases of COVID-19 viral infection while a new suspicious chest X-ray image is shown to the network. Our network's overall accuracy on the COVIDx test dataset was 95.33%, with a sensitivity and precision of 92% and 99% for COVID-19 cases, respectively, which are the highest results on the COVIDx dataset to date, to the best of our knowledge. Finally, we performed an audit on RAM-Net based on the Grad-CAM's interpretation to demonstrate that our proposed architecture detects SARS-CoV-2 (COVID-19) viral infection by focusing on vital factors rather than relying on irrelevant information.**

***Index Terms*—Dilated Depthwise Separable Convolution, Residual Blocks, Attention Augmented Convolution, MobileNet, COVID-19, Medical Image Analysis, CXR, Grad-CAM**

I. Introduction

The rapid spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also known as 2019-nCov or COVID-19) has caused a global alarm since December 2019. There are some credible published papers [6], [7] that have provided beneficial information about clinical features of infected patients with this viral pneumonia, their epidemiological and radiological characteristics besides presenting several treatment outcomes. Moreover, in [8], the authors investigated the diagnostic value and consistency of chest CT compared with RT-PCR examination. They concluded that chest CT has a high sensitivity for the diagnosis of COVID-19, and might be considered as a primary tool for COVID-19 screening in epidemic areas. Fang et al. [9] have also compared the sensitivity of chest CT and viral nucleic acid assay at the initial patient presentation. They have reported that chest CT exhibits higher sensitivity than that of RT-PCR (98% vs 71%, respectively, p<.001), that makes radiography images an acceptable complement or even better alternative than RT-PCR examination.

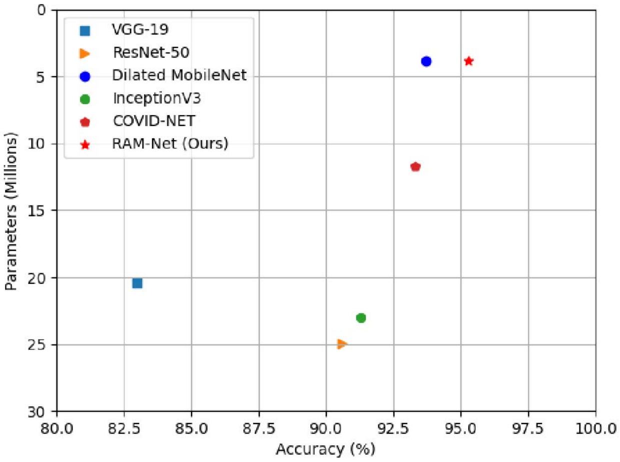


Fig. 1. Our proposed Residual Attention MobileNet (RAM-Net) presents superior accuracy with the fewest number of parameters than ResNet-50 [1], VGG-19 [2], InceptionV3 [3], Dilated MobileNet [4] and COVID-Net [5].

Therefore, radiological examination such as chest X-ray (CXR) imaging could be considered an alternative to detect viral infection of COVID-19. According to some early studies [6], [7], [10], chest radiography images can exhibit the abnormalities of COVID-19, and it could be employed as one of the foremost tools for the virus screening. Influenced by a dire need of automated solutions to fight against COVID-19, and public availability of chest radiography images, we propose a deep neural network, namely RAM-Net (Residual Attention MobileNet), to detect viral infection of COVID-19 from chest X-ray images. To train the neural network, we have used 13675 chest X-ray (CXR) images of the three classes (COVID-19, Normal (no Pneumonia), Other Pneumonia) from the COVIDx dataset [5]. Furthermore, in order to test the Residual Attention MobileNet, 300 X-ray images of the COVIDx test dataset have been utilized. The results are then compared with the state of the art method COVID-Net [5]. COVID-Net is a deep convolutional neural network (DCNN) designed to detect COVID-19 cases from chest X-ray images. The CXR dataset

DOI 10.1109/ICMLA51294.2020.00040

(COVIDx) comprising 13,975 chest radiography images across 13,870 patient cases has been used to train their model.

The **contributions** of this work are three folds. (1) A novel RAM friendly model (mobile vision application) is proposed to efficiently and effectively detect the infected COVID-19 patients. It greatly releases the computational burden and makes the model applicable to the real-world. (2) An attention-based residual module is constructed to force the network focus on the lesion within the X-ray, which offers the network more interpretability to rely on. Besides, the dilated convolution is applied to enlarge the field of view and save the computational cost simultaneously. Together dilated convolution and attention augmented convolution assist the network in achieving the global information from images. (3) Extensive quantitative and qualitative experiments have demonstrated that the proposed RAM-Net can more accurately recognize the positive sample and outperforming state-of-the-art methods.

The remaining sections of the paper are arranged as follows: in section 2, we present our proposed deep neural network (DNN) based architecture that is designed to learn the chest X-ray images. We present and discuss our experimental results that are achieved after employing the model to the unseen (new) chest X-ray images, in Section 3. Finally, we conclude the paper and recommend the potential future directions in Section 4.

II. Methodology and Dataset

For embedded mobile vision applications, MobileNet has been built with a small number of parameters, less computational complexities, and very low latency [11]. It holds three different convolution operations, including pointwise convolution, depthwise convolution, and standard convolution. In this work, we proposed a new version of MobileNet called RAM-Net (in Fig. 2) which based on attention augmented convolution [12], residual block [1], and dilated convolution [13]. The network pre-trained on ImageNet dataset [14] to bolster the accuracy of this classification task.

A. Dilated Depthwise Separable Convolution (DDSC)

The proposed dilated depthwise separable convolution (DDSC) is different from the standard convolution operation. One significant weakness of the traditional convolutional layer is that it works with a small receptive field. Thus, this characteristic prevents it from achieving global information during classification [12]. We attempt to tackle this problem by employing the dilated convolution and attention augmented convolution in our architecture. Firstly, to incorporate dilated depthwise separable convolution, we select five depthwise layers of this network where all of these depthwise convolutional layers have a $stride = 2$. The output of depthwise convolution with one filter, $stride = 1$, and padding can be reckoned as:

$$\hat{Z}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m} \tag{1}$$

where, Z= output feature map, F= input feature map, and K= convolutional kernel. Besides, F is a discrete function, and $\hat{K} : \omega_r \to R$ is a discrete depthwise convolution kernel of size $(2r+1)^2$ where, $\omega_r = [-r, r]^2 \cap Z^2$ and n^{th} filter in $\hat{K}$ is registered to the n^{th} channel of feature map in F to provide the n^{th} channel of the filtered output feature map $\hat{z}$. Next, if h is an element, q is a dilation factor, then $*q$ will be:

$$F *_q \hat{K}(h) = \sum_{s+qt=p} F(s)\hat{K}(t) \tag{2}$$

In equation (2), $*q$ is a dilated convolution where * is referred to as 1-dilated convolution. In dilated depthwise separable convolution (DDSC), we combine this depthwise dilated convolution with pointwise convolution (1×1 convolution). There are five DDSC layers in our proposed RAM-Net, and every layer has a $stride = 2$. In our network, the first two depthwise layers have a dilation rate of 1; however, for the third and fourth depthwise convolution layers, we placed a dilation rate of 2. Furthermore, we create three depthwise convolutions parallelly with a dilation rate of 4, 8, and 16, respectively. Finally, we concatenated these three depthwise layers together to produce the fifth dilated depthwise separable convolution layer, which can be formulated as:

$$DC_5 = CONCAT[(F_4\hat{K})(h), (F_8\hat{K})(h), (F_{16}\hat{K})(h)] \tag{3}$$

Here, DC_5 is the 5th depthwise convolution layer, which obtains the output from a concatenation operation. Additionally, F_4, F_8, and F_{16} are feature maps from three different DDSC layer where 4, 8, and 16 are different dilation rate.

B. Residual Blocks

The vanishing/exploding gradients problem is a colossal issue in the neural networks that happened because of the combination of an extensive amount of layers [1]. It occurs during the backpropagation and hinders convergence from very early. So, We apply identity residual blocks and convolutional residual blocks in our residual part. Firstly, the identity blocks have the same input activation dimensions ($a^{[l]}$) and the output activation dimension ($a^{[l+2]}$). In the residual blocks, “skip connection” has been utilized to backpropagate the gradient to the earlier layers so that the network can comfortably learn the identity mapping [1]. We implement three layers in each identity block. The first and last layer consists of (1×1) kernel size with $ReLU$ activation and BatchNormalization steps to speed up the training. The second layer has a (3×3) convolution operation with $ReLU$ activation and BatchNormalization. We employ a $stride$ rate of 1 throughout the identity block.

The convolutional residual blocks have dissimilar input and output of residual mapping tensor shape. The main difference with the identity block is that there is one convolutional layer in the shortcut path. This convolutional layer is used to reshape the input tensor to a different dimension. We can denote this layer as a shortcut convolutional layer. Therefore, the first layer and the shortcut convolutional layer has (1×1) convolutional operation with the $stride = 2$. However, the other layers have kernel size (3×3) with a $stride = 1$. All the layers have

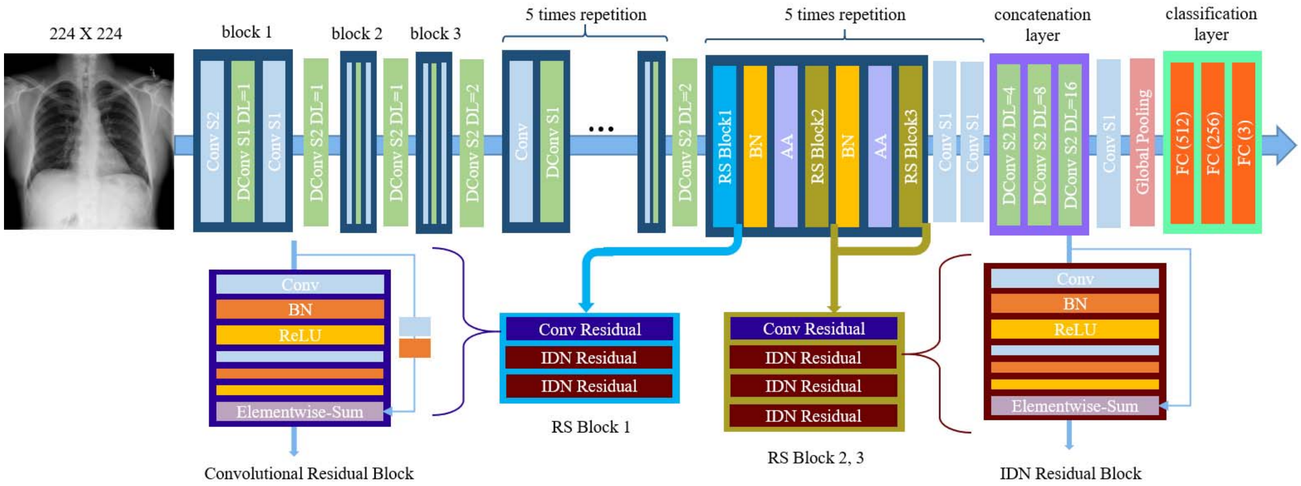


Fig. 2. Block 1, Block 2, and Block3 in Residual Attention MobileNet (RAM-Net) have the same structures. DConv, S, and DL mean Depthwise Convolution, Stride, and Dilation Rate, respectively. Furthermore, RS denotes the residual block, which holds Convolutional (Conv) Residual operation and Identity (IDN) Residual operation. Then, BatchNormalization (BN) and Attention Augmented Convolution (AA) followed by the first two Residual (RS) Blocks. Finally, the classification part has three fully connected (FC) layers.

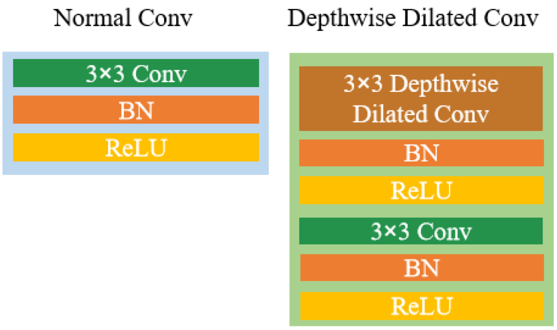


Fig. 3. Left: Normal Conv have standard convolutional layer with (3x3) kernel, BatchNormalization (BN), and Rectified Linear Unit (ReLU). Right: Depthwise Dilated Conv contains (3x3) dilated depthwise separable convolutional (DDSC) operation, (3x3) standard convolutional operation, BatchNormalization (BN), and Rectified Linear Unit (ReLU).

the $ReLU$ activation except the shortcut convolutional layer, though all the layers have BatchNormalization steps.

We utilized three different sets of residual blocks: The first set contains two identity residual blocks, followed by one convolutional residual block, and The second and third set includes three identity blocks, followed by one convolutional residual block. The identity operation can be defined as:

$$y = ReLU(f(x,(\omega_i)) + x) \tag{4}$$

Here, y is the output tensor, x is the input tensor, and function $f(x,(\omega_i))$ can be represented as the learned residual mapping. For the identity residual block, the dimensions of x and $f(x,(\omega_i))$ must be equal. Next, the convolutional residual block can be formulated as:

$$y = ReLU(f(x,(\omega_i)) + \omega_s x) \tag{5}$$

Where to match the output tensor shape of shortcut connection, we have to execute projection ω_s. In both types of residual blocks, the $ReLU$ activation function has been applied after the element-wise addition on two feature maps.

C. Attention Augmented Convolution

Though the convolutional operation is a popular classification method, it has notable frailty because it only concentrates on the local neighborhood via limited receptive field [12]. Nevertheless, this fondness of convolutional kernel on the local environment prevents it from attaining global understanding during the image classification. Previously, we proposed dilated depthwise separable convolution to avoid this problem, and now we apply attention augmented convolutional with residual blocks. Besides, attention augmented convolution increases the classification accuracy with the same architectural complexities as it does not enhance the parameters of the model [12]. Overall we have two attention augmented convolutional layers. Each of these is followed by the first and second sets of the residual block and the BatchNormalization layer. Like Bello et al. [12], we implement attention augmented convolution after the BatchNormalization layer. After getting our output tensor y from a residual block, we pass this tensor to the multi-head self-attention attention augmented layer. If the single head can be assumed by h, then the multi-head self-attention for the input matrix y can be computed as:

$$MA(y) = CONCAT[SA_1, \ldots, SA_{Nh}]W^0 \tag{6}$$

Where SA_h is for single head attention, and W^0 is a learned linear transformation. During the training phase, multi-head self-attention feature maps produce high accuracy when it is concatenated with the convolutional feature maps. So the final output from this layer can be defined as:

$$AAC(y) = CONCAT[MA(y), CONV(y)] \tag{7}$$

D. RAM-Net Architecture

We replaced all the present depthwise convolutional layers in traditional MobileNet with dilated depthwise separable convolution (DDSC). In RAM-Net, we employ the three residual blocks, two attention augmented convolutional layers, and two standard convolutional layers with $stride = 1$ in between the fourth and the fifth DDSC layer. In the classifier part, we apply global-average max pooling to convert the tensor shape from $h{\times}w{\times}d$ to $1{\times}1{\times}d$, where $h{\times}w$ = spatial dimensions and d referred to as the number of feature maps. There are three fully connected (FC) layers with filters number 512, 256, and 3 (for three classes), respectively. We apply a dropout rate of 0.50 for the first two FC layers that substantially debilitates the overfitting. Moreover, The first two layers have the $ReLU$ activation, and the last one has the $SOFTMAX$ activation.

E. Dataset

Lately, there have been noteworthy endeavors to detect COVID-19 as fast and automatic as possible [15]–[17]. In this regard, Wang et al. [5] generate the COVIDx dataset after integrate five open-access datasets: 1) COVID-19 image data collection [15], 2) Actualmed COVID-19 chest X-ray data initiative [16], 3) Figure-1 COVID-19 chest X-ray data initiative [18], 4) COVID-19 radiography database [19], 5) RSNA Pneumonia Detection Challenge dataset [17]. The dataset is divided into the training and test parts. The dataset contains 13675 chest X-ray images in the training part, where we have 7966 Normal, 5451 Non-COVID Pneumonia, and 258 COVID-19 X-ray images. The COVIDx test dataset holds 100 X-ray images from each class (300 images in total). COVIDx includes three different categories of X-ray images:

- Normal: X-ray images do not have Pneumonia infection
- Non-COVID Pneumonia: different kinds of pneumonia infection other than COVID-19
- COVID-19: comprise COVID-19 positive cases

F. Data Augmentation

All the images resized into $(224{\times}224)$, and to achieve faster convergence, we apply the min-max normalization technique to re-scale each pixel between 0 and 1.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}, i = 1, 2, ..., n \tag{8}$$

Here z_i is the output for the i^{th} data point, and x_i is the input for the same data point. Several data augmented approaches have been followed to produce transformed images. At first, we have applied a horizontal and a vertical flip with a probability of 0.40. Then, the zoom range ±0.20 is employed to randomly zooming inside each image. Height and width shifting with range 0.20 have been utilized to manage off-centered objects. Finally, vertical shear mapping range $(x, y{+}0.20)$ and horizontal shear mapping range $(x{+}0.20, y)$ has been used to supplant the image.

G. Implementation Details

The proposed RAM-Net is pre-trained on the ImageNet [14] and trained on the COVIDx dataset for the 22 epochs with Adam optimizer. Moreover, for the training part, the learning rate was $1e - 4$, utilized cross-entropy loss, and the mini-batch size 32. We have used a callback function to lessen the learning rate factor by $\sqrt{0.1}$ if learning stagnates for five epochs (patience). The lower bound for the learning rate was $0.5e - 6$. Hence, if x= current learning rate, then the new learning rate x_{new} will be:

$$x_{new} = x * \sqrt{0.1} \tag{9}$$

III. Experiment Results

Numerous combinations of RAM-Net are investigated to discover the proposed one. We perform various experimental analyses to evaluate the efficacy of our method. Through several evaluation criteria, we attempt to understand the decision-making and classification performance of RAM-Net.

A. Ablation Study

We have inspected several different setups with new dilated depthwise separable convolution (DDSC) in MobileNet to pick the appropriate one. That means in all our combinations, we include the DDSC part but attach or detach the residual blocks and attention augmented convolution. In Table I, we have displayed the outcome of several combinations for different evaluation criteria such as test accuracy, and average sensitivity and precision of all classes to choose the best combinations. The best result has been found by selecting the dilated depthwise separable convolution (DDSC) in MobileNet with three residual blocks and two attention augmented convolutions.

B. Test Accuracy and Architectural Complexities

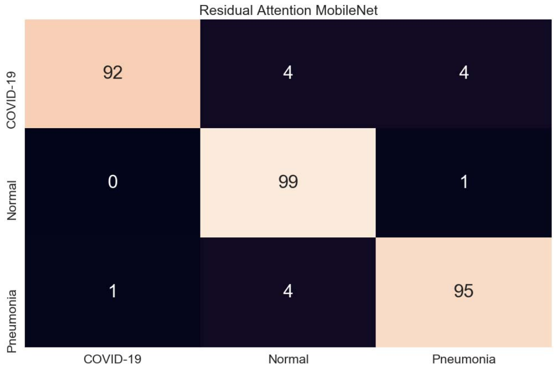
The proposed architecture provides superior accuracy with minimal parameters than many popular models such as ResNet-50 [1], VGG-19 [2], and InceptionV3 [3] on the COVIDx test dataset. Furthermore, we compare our experimental results with Dilated MobileNet [4] and COVID-Net [5], which is a state-of-the-art model for the COVIDx dataset. Our proposed network has only 3.88 million parameters, similar to Dilated MobileNet but exhibits less architectural complexities with higher accuracy than COVID-Net (11.75 million). According to Fig. 1, RAM-Net achieves 95.33% test accuracy. To the best of our knowledge, this efficacy is higher than any state-of-the-art methods tested on the COVIDx test set, including COVID-Net (93.3%).

C. Recall, Precision, and Confusion Matrix

We inspect our proposed network critically by studying recall (sensitivity), precision (positive predictive value), and confusion matrix for separate infections. In Tables II and III, we compare the recall and precision, respectively, for different infections with some existing methods. At first, we can notice that RAM-Net obtains a decent recall rate for COVID-19 instances (92%). A decent recall rate is necessary

TABLE I
THE PERFORMANCE OF SEVERAL COMBINATIONS ON THE COVIDX TEST DATASET AND THE BEST OUTCOMES (PROPOSED METHOD) MARKED IN **BOLD**

Setups	Test Accuracy (%)	Average Recall (%)	Average Precision (%)
DDSC in MobileNet + 2 residual blocks	94.3	94.3	94.3
DDSC in MobileNet + 3 residual blocks	93.7	93.7	94.0
DDSC in MobileNet + 2 residual blocks + 2 attention augmented convolution	94.7	94.7	94.7
DDSC in MobileNet + 3 residual blocks + 2 attention augmented convolution (RAM-Net)	**95.3**	**95.7**	**95.3**
DDSC in MobileNet + 3 residual blocks + 3 attention augmented convolution	94.0	94.00	94.0



	COVID-19	Normal	Pneumonia
COVID-19	92	4	4
Normal	0	99	1
Pneumonia	1	4	95

Fig. 4. Confusion Matrix of RAM-Net for the COVIDx test dataset.

TABLE II
RECALL (SENSITIVITY) FOR EACH VIRAL INFECTION. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**.

Model	Normal (%)	Non-COVID Pneumonia (%)	COVID-19 (%)
VGG-19 [2]	98.0	90.0	58.7
ResNet-50 [1]	97.0	92.0	83.0
Dilated MobileNet [4]	98.0	89.0	91.0
InceptionV3 [3]	93.0	93.0	88.0
COVID-Net [5]	95.0	94.0	91.0
RAM-Net (Ours)	**99.0**	**95.0**	**92.0**

because we do not want to miss a considerable amount of COVID-19 cases during the recognition process. Secondly, we examine that our network attains a high 99% precision rate for COVID-19 infection. This highly accurate positive predictive value demonstrates the low amount of false-positive COVID-19 recognition. From the confusion matrix (Fig. 4), we can see only one patient of COVID-19 viral infections was misclassified. In order to minimize PCR testing, a low false-positive rate is necessary for any healthcare system [5]. Thus, based on the experimental result, we can state that our proposed RAM-Net operates well to detect COVID-19 viral infections from chest X-ray (CXR) images. The proposed method generalizes and performs better than all baseline models for the COVIDx test dataset to the best of our knowledge.

D. Explainability of RAM-Net:

This study's primary goal is to provide a clinical application in response to a critical problem like COVID-19 viral infection. The performance of this application will directly affect the health of infected patients. Thus, we design RAM-Net with sheer accountability and transparency. Here, we displayed that RAM-Net determines the outcome based on apposite knowledge rather than unacceptable facts such as imaging artifacts, incorrect visual indicators outside of the patient body, embedded markup signs, etc. Here, we employ the Gradient-weighted Class Activation Mapping (Grad-CAM) [20] technique, which provides a visual explanation for the outcome of the Deep Neural Network to make it more transparent. Grad-CAM utilizes the gradients of any target concept, flowing into the last convolutional layer to provide a coarse localization map indicating the crucial regions in the image for forecasting the concept. We apply the Grad-CAM method on our proposed RAM-Net to understand how this network inspects the CXR images and makes decisions. Moreover, we validate the result produced from Grad-CAM to examine whether RAM-Net makes a decision based on relevant facts or takes some erroneous bias decisions based on unrelated visual indicators. Based on the interpretation of Grad-CAM in Fig. 5, we can observe that RAM-Net focuses on some particular areas in the lungs on detecting COVID-19 viral infections from chest X-Ray (CXR) images of the patients. In Fig. 5, the red mark can recognize these critical factors, which validates that our proposed RAM-Net focuses on essential aspects rather than depending on irrelevant details.

TABLE III
PRECISION (PPV) FOR EACH VIRAL INFECTION. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**.

Model	Normal (%)	Non-COVID Pneumonia(%)	COVID-19 (%)
VGG-19 [2]	83.1	75.0	98.4
ResNet-50 [1]	88.2	86.8	98.8
Dilated MobileNet [4]	90.0	**95.0**	95.0
InceptionV3 [3]	86.0	**95.0**	94.0
COVID-Net [5]	90.5	91.3	98.9
RAM-Net (Ours)	**93.0**	**95.0**	**99.0**

IV. CONCLUSION

In this paper, we have proposed a deep convolutional neural architecture to tackle the adverse upsurge of COVID-19 cases. Our proposed RAM-Net is learned by employing the chest X-ray (CXR) examples of publicly available COVIDx dataset. Inside our system, we apply the new Dilated Depthwise Separable Convolution (DDSC) layer instead of the basic depthwise layer of MobileNet. Additionally, we integrate three sets of

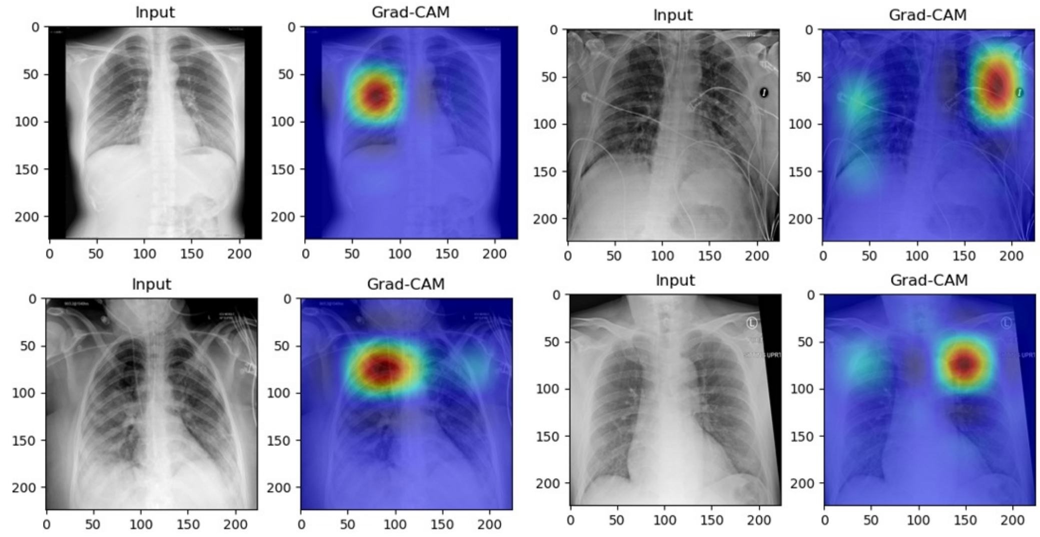


Fig. 5. Chest X-ray (CXR) images of SARS-CoV-2 (COVID-19) cases from four different patients and their associate infected area marked in red, which is identified by Grad-CAM.

residual blocks and two sets of the Attention Augmented Convolution layer. We evaluate our approach on the COVIDx test dataset and achieve superior outcomes for each evaluation metric, which is higher than baseline methods to the best of our knowledge. This study's primary motivation is to assist radiologists in the diagnosis step and accelerate the treatment process of infected patients who urgently require it. However, the future directions of this work will include learning our model with a higher number of CXR images of COVID-19 cases from all over the world. The principal purpose behind the future data collection process is to raise the sensitivity, the precision of COVID-19 infections, and avoid biases. We also hope to increase the explainability of our network in the future. By increasing the network's interpretability, we would be able to indicate the critical factors that have a significant effect on making the right detection decision based on CXR.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[4] M. A. R. Ratul, M. H. Mozaffari, W. Lee, and E. Parimbelli, "Skin lesions classification using deep learning based on dilated convolution," *bioRxiv*, p. 860700, 2020.

[5] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *arXiv preprint arXiv:2003.09871*, 2020.

[6] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," *The lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[7] M.-Y. Ng, E. Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M. M.-s. Lui, C. S.-Y. Lo, B. Leung, P.-L. Khong *et al.*, "Imaging profile of the covid-19 infection: radiologic findings and literature review," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2020.

[8] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, p. 200642, 2020.

[9] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," *Radiology*, p. 200432, 2020.

[10] W.-j. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. Hui *et al.*, "Clinical characteristics of coronavirus disease 2019 in china," *New England journal of medicine*, vol. 382, no. 18, pp. 1708–1720, 2020.

[11] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[12] I. Bello, B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le, "Attention augmented convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3286–3295.

[13] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[15] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, 2020.

[16] A. Chung, "Actualmed covid-19 chest x-ray data initiative," https://github.com/agchung/Actualmed-COVID-chestxray-dataset.

[17] "Rsna pneumonia detection challenge," https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data.

[18] L. Wang, A. Wong, Z. Q. Lin, J. Lee, P. McInnis, A. Chung, M. Ross, B. van Berlo, and A. Ebadi, "Figure 1 covid-19 chest x-ray dataset initiative."

[19] 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?" *arXiv preprint arXiv:2003.13145*, 2020.

[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.