Five Deep Learning Models for Classification of Chest X-Ray Images of COVID-19

Abdesselam FERDI
Electronics Department
Frères Mentouri Constantine 1 University
Constantine, Algeria
abdesselam.ferdi@gmail.com

Said BENIERBAH
Electronics Department
Frères Mentouri Constantine 1 University
Constantine, Algeria
said_benierbah@umc.edu.dz

Youcef FERDI
National Higher School of Biotechnology
Constantine, Algeria
youcef.ferdi@gmail.com

***Abstract*—In recent years, deep learning algorithms have acquired a great popularity in many fields including medical imaging to achieve different tasks such as image segmentation, classification, detection, and retrieval tasks. Although reverse transcription polymerase chain reaction test (RT-PCR) is considered to be the gold standard for COVID-19 screening, deep learning algorithms may help radiologists to automatically diagnose this disease by processing chest X-ray (CXR) and chest CT images. This paper proposes a comparative study of five deep learning models for classification of CXR COVID-19 images. The five selected models are pre-trained networks, namely, ResNet-50, InceptionResNet-v2, Inception-v3, MobileNet-v2, and GoogleNet. The models were evaluated using CXR COVID-19 and normal images from publicly available datasets. The obtained classification results show that the GoogleNet model achieves the highest classification performance with a validation accuracy of 95.69%.**

Keywords—COVID-19 · Deep Learning · Convolutional Network · Chest X-ray image · Medical Image Processing

I. INTRODUCTION

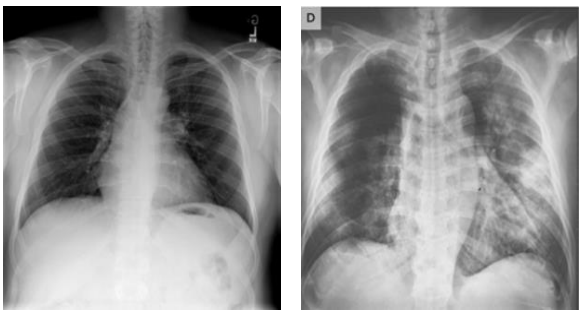
The COVID-19 is an infectious disease caused by a virus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The outbreak has been first identified in Wuhan, China, in December 2019. Since then, the number of infected people is increasing rapidly and, unfortunately, some patients develop difficulty in breathing and require hospitalization. To help combatting the spread of this worldwide pandemic, researchers are working on developing accurate detection and early treatment of infected patients. The reverse transcription polymerase chain reaction test (RT-PCR) is the gold standard for diagnosis of COVID-19. However, this method is not always available because of the possible short supply of test kits and the strict environment requirements for accomplishing screening [1]. Some studies have also shown false positives in PCR testing.

Another fast and effective method for COVID-19 screening is based on chest X-ray (CXR) and chest computed tomography (CT) images analysis. COVID-19 diagnosis systems based on radiological imaging techniques have been considered as a good complement to RT-PCR tests because of the clinical signs of most COVID-19 patients, which suffer from lung infection. In addition, these methods have several advantages such as speed, availability, and accessibility. Medical images are acquired and analyzed by radiologists who look for visual findings associated with lung infections caused by the virus. Visual medical images analysis is time-consuming, and the findings may be observed in other lung infections, such as Influenza (H1N1) and SARS [2]. Figure 1 shows an image example of a normal case and a COVID-19 case.

Deep learning is widely used in many disciplines such as computer vision, medical imaging, and bioinformatics. In medical imaging, deep learning techniques have been used to achieve image segmentation, classification, detection, and retrieval tasks [3-9]. In the context of COVID-19, deep learning is used to help radiologists to interpret radiographic images faster and more accurately to detect COVID-19 cases, in order to fight against its spread. Recently, many deep learning models has been proposed to detect patients infected with COVID-19 using both CXR and chest CT images [3-9]. Deep learning based COVID-19 diagnosis systems are developed either from existing pre-trained models with transfer learning or as customized networks. Most of deep learning models used for COVID-19 detection are based on convolutional neural networks (CNNs) architecture. These include GoogleNet, Inception or Xception, VGGNet, ResNet, AlexNet, MobileNet, and ShuffleNet. Recently, some papers dealing with reviewed studies of deep learning based COVID-19 diagnosis systems have been published [3-9]. Their results for COVID-19 detection from various datasets of CXR and CT images mostly collected from internet sources have been reported. Metrics commonly used to assess the performance of deep learning models are accuracy, sensitivity, specificity, precision, F1-score, and area under the curve (AUC). According to these studies, promising performance has been achieved by the developed systems, but improvements still need to be proposed regarding the datasets of medical images and the design of optimized deep learning algorithms to reduce computational cost and overcome the problem of small amount of data.

In this paper, five deep learning models are fine-tuned by replacing the last layers with new layers adapted to the datasets. Although previous works have dealt with this topic, the aim of the proposed work is to compare these models on new datasets with a larger number of COVID-19 images.

The rest of the paper is organized as follows. Section II provides a literature survey on the most recent works dedicated to the application of deep learning technology to COVID-19 diagnosis. Transfer learning, dataset, image preprocessing, classification models, parameters settings, and metrics are described in Section III. The obtained results are discussed in Section IV. The paper ends with a conclusion and opens up prospects for new research.



(a) (b)

Fig. 1. Examples of a) normal and b) COVID-19 CXR images [2]

II. LITTERATURE SURVEY

Very recently, several works on the application of deep learning models to COVID-19 classification in CXR and chest CT images have been proposed. In this section, we list some of these works. H. X. Bai et al. [3], compared the performance of radiologists in detection of COVID-19 from other pneumonia with and without AI assistance. The model was trained using 1186 chest CT images. On independent testing, this model achieved an accuracy of 87%, a sensitivity of 89%, and a specificity of 86%. K. El. Asnaoui et al. [4], presented a comparative study of seven different models for classification of COVID-19 pneumonia in CXR and CT images. These models were trained using 6087 images. The Inception-ResNet-v2 provided the highest accuracy of 92.18%. U. Ozkaya et al. [5], compared the proposed method as fusing and ranking deep features with three different pre-trained models. 150 CT images were used to train these models to classify COVID-19 cases. The proposed method shows best results with 98.27% accuracy. I. D. Apostolopoulos and T. A. Mpesiana [6], evaluated the classification performance of five different CNNs using transfer learning. To train the models to distinguish COVID-19 from common pneumonia, a collection of COVID-19, normal, and common bacterial and viral pneumonia images was selected. The MobileNet-v2 is the better model with an accuracy and a specificity of 96.78% and 96.46% respectively. A. Narin et al. [7], proposed five pre-trained CNNs for the classification of COVID-19 pneumonia. To train the models, 341 COVID-19, 2800 normal, 1493 viral pneumonia, and 2772 bacterial pneumonia X-ray scans are used. The performance results shows that the ResNet-50 provides the highest classification performance with 99.7% accuracy. E. E.-D. Hemdan, et al. [8], evaluated seven different pre-trained CNNs to classify COVID-19 in X-ray images. The models are trained and tested on 50 chest X-ray images. The VGG19 and DenseNet models showed better and similar classification performance with f1-scores of 0.91 and 0.89 for COVID-19 and normal, respectively. M. Chieregato et al. [9], developed a hybrid machine learning/deep learning model to classify patients in two classes (non-ICU and ICU: Intensive Care Unit admission) using 558 CT images. The proposed method reached a probabilistic AUC of 0.949.

III. METHODS

In this Section, transfer learning, dataset, image preprocessing, classification models, parameters settings, and metrics are described.

A. Transfer Learning

Transfer learning is the enhancement of the learning of a new task by the transfer of knowledge from another similar task that has already been learned. In neural networks, this is done by using parts of the previously trained network in the new one. This operation is, particularly, important when the new task is using a smaller number of training images [10]. In pre-trained CNNs, the convolutional layers extract image features that the last learnable layer and the final classification layer use to classify the input image. To fine-tune the network to classify CXR images into two classes (COVID-19 and normal) using transfer learning, we replace the last layers with three new layers adapted to the new dataset.

B. Dataset

In this study, we used *COVID-19 Radiography Database* [2]. This is a dataset of 13808 CXR images of normal and COVID-19. It contains 3616 COVID-19 positive cases along with 10.192 Normal, 6012 Lung Opacity, and 1345 Viral Pneumonia images. All the images are in PNG format of size 299×299 pixels. Only the COVID-19 and Normal images are selected.

C. Images preprocessing

The selected images has been resized into 224×224 pixels and 299×299 pixels in Jpeg format. The dataset has been randomly divided into two independent sets with 80% for training and 20% and validation. The k-fold cross-validation method is used, where the value of k is 5.

D. Classification models

Five models are used to classify CXR images into two classes: COVID-19 and normal. These models are pre-trained networks, namely, ResNet-50 [11], Inception-ResNet-v2 [12], Inception-v3 [13], MobileNet-v2 [14], and GoogleNet [15]. These models are trained on over millions of images in the dataset ImageNet, and can classify images into 1000 classes. The main characteristics of these networks are given in Table I. These networks are fine-tuned with transfer learning by replacing the last learnable layer with three new layers of GlobalAveragePooling2D and two fully connected layers with the number of outputs equal to 1024 and 2 respectively.

TABLE I. MAIN CHARACTERISTICS OF THE MODELS.

Network	Depth	Parameters (Millions)	Image Input Size
ResNet-50	50	25.6	224×224
Inception-ResNet-v2	164	55.9	299×299
Inception-v3	48	23.9	299×299
MobileNet-v2	53	3.5	224×224
GoogleNet	22	7.0	224×224

E. Performance evaluation

To evaluate the performance of the different algorithms used to classify CXR images into two classes, three criteria are calculated, namely, Accuracy, Sensitivity, and Specificity. In terms of positives and negatives, they are defined as:

$$\text{Accuracy (ACC)} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Sensitivity (SN)} = TP/(TP+FN) \quad (2)$$

$$\text{Specificity (SP)} = TN/(TN+FP) \quad (3)$$

where TP is the true positive in case of COVID-19 case and TN is the true negative in case of normal case, while FP and FN are the incorrect false positive and false negative predictions for COVID-19 and normal cases.

F. Parameters Settings

The work has been carried out in MATLAB R2021a on a computer equipped with an Intel® Core ™ i7-7800X CPU @ 3.50 GHz processor with 64 GB of RAM and RTX 2080 TI GPU. For models training, we used the Adam optimizer as the optimizer with an initial learning rate of 0.0005, 50 epochs, and mini-batch size of 32.

IV. RESULTS COMPARISON AND DISCUSSION

The obtained results of our implementation of the five models are presented here in the form of:

- Training progress curves for fold-5. These are shown in figures 2 to 6.
- Table of the computed performance metrics. These are presented in Table Ⅱ.
- Confusion matrices. These are shown in figure 7.

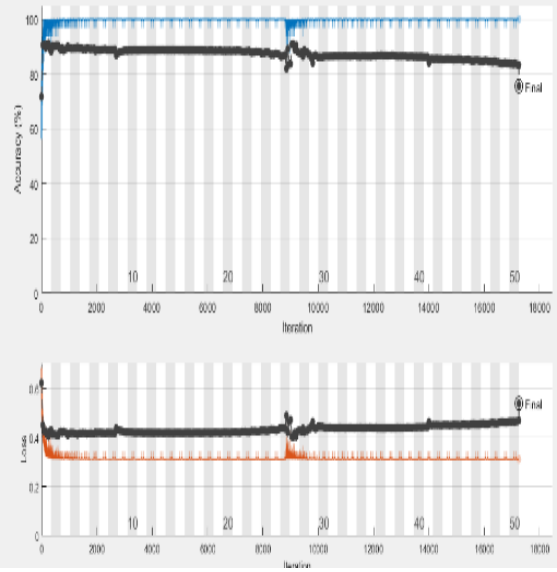


Fig. 2. Training progress curve of the ResNet-50 model.

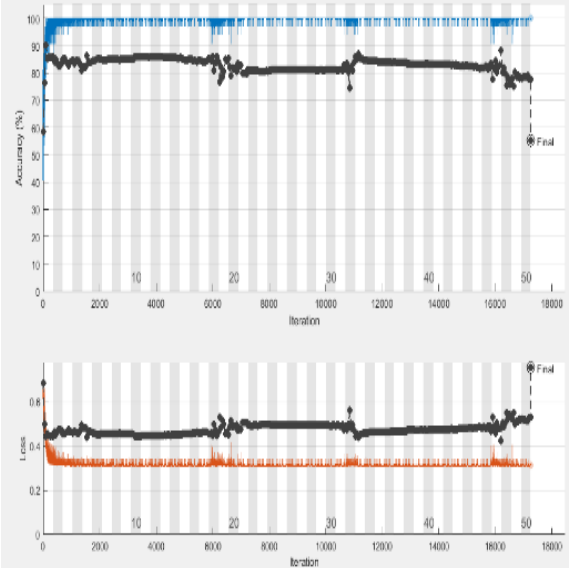


Fig. 3. Training progress curve of the InceptionResNet-v2 model.

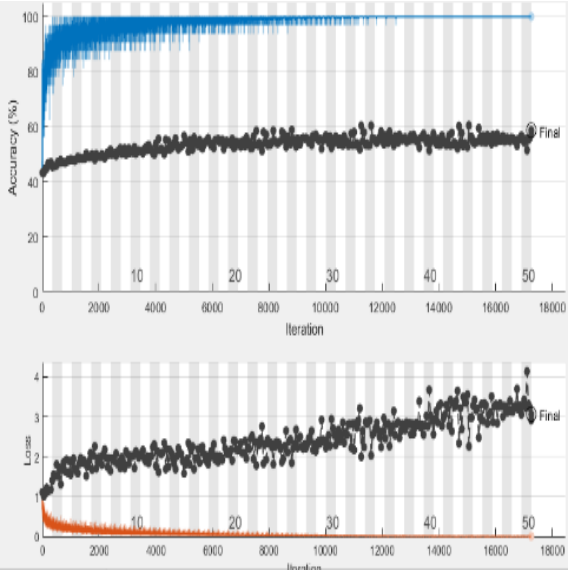


Fig. 4. Training progress curve of the Inception-v3 model.

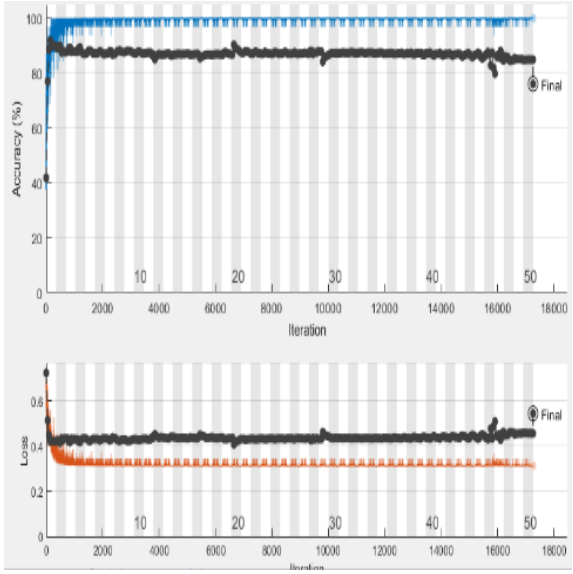


Fig. 5. Training progress curve of the MobileNet-v2 model.

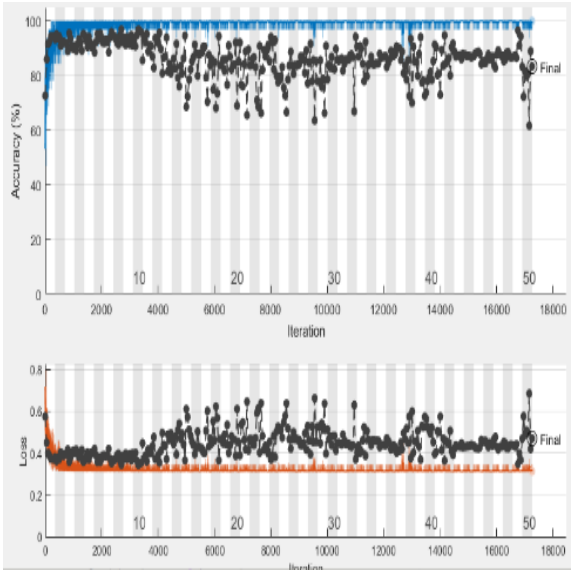


Fig. 6. Training progress curve of the GoogleNet model.

During the training of the models, MATLAB R2021a software can display a progression figure of four curves representing, in the upper part, the training and validation accuracy in blue and black respectively, and in the lower part the loss curves in training and validation in orange and black respectively , depending on the number of iterations (epochs). Figures 2 to 6 show the training progress curves for fold-5 of ResNet-50, Inception-ResNet-v2, Inception-v3, MobileNet-v2, and GoogleNet respectively. We can see from these figures that the highest training accuracy is obtained with the GoogleNet model. The second is ResNet-50 model which is more efficient than MobileNet-v2, Inception-v3, and Inception-ResNet-v2 models. We can also see that the validation loss decreases in five models during the training stage. In figure 4, we can observe that the validation loss values of the Inception-v3 model increase and approaches 3.

Table II shows the values of the three performance criteria given by the implementation of the five models. As can be observed from Table II, The results obtained using the GoogleNet model outperform those obtained from ResNet-50, MobileNet-v2, Inception-v3, and Inception-ResNet-v2 models in accuracy and sensitivity with 95.69% of validation accuracy.

The confusion matrices of COVID-19 and normal validation results of the models for fold-5 are shown in figure 7. The confusion matrix measures the quality of a classification model, with rows representing output classes, and columns representing target classes. The GoogleNet model classified 690 of COVID-19 as true positive and classified 33 images as false positive. In normal case, 1612 images are classified as true negative and 426 images as false negative. The ResNet-50 model classified 704 of COVID-19 as true positive and classified 19 images as false positive. In normal case, 1383 images are classified as true negative and 655 images as false negative.

TABLE II CLASSIFICATION PERFORMANCE RESULTS OBTAINED FROM THE FIVE MODELS.

Models/Folds		Acc	SN	SP
ResNet-50	Fold-1	0.995	0.994	0.995
	Fold-2	0.995	0.992	0.996
	Fold-3	0.992	0.9861	0.9941
	Fold-4	0.9964	0.9917	0.998
	Fold-5	0.7559	0.518	0.9864
	Mean	**0.9469**	**0.8964**	**0.9939**
InceptionRes-Net-v2	Fold-1	0.9848	0.9736	0.9887
	Fold-2	0.9967	0.9917	0.9985
	Fold-3	0.9804	0.9912	0.9769
	Fold-4	0.9942	0.9836	0.998
	Fold-5	0.5516	0.3665	0.9808
	Mean	**0.9015**	**0.8613**	**0.9886**
Inception-v3	Fold-1	0.993	0.994	0.993
	Fold-2	0.99	0.975	0.996
	Fold-3	0.9942	0.9944	0.9941
	Fold-4	0.9971	0.9917	0.999
	Fold-5	0.5867	0.3797	0.978
	Mean	**0.9122**	**0.8669**	**0.992**
MobileNet-v2	Fold-1	0.985	0.988	0.984
	Fold-2	0.99	0.978	0.994
	Fold-3	0.992	0.9861	0.9941
	Fold-4	0.9884	0.9765	0.9926
	Fold-5	0.7606	0.523	0.9866
	Mean	**0.9432**	**0.89**	**0.99**
GoogleNet	Fold-1	0.983	0.983	0.983
	Fold-2	0.99	0.985	0.993
	Fold-3	0.9837	0.9788	0.9854
	Fold-4	0.9898	0.9833	0.9922
	Fold-5	0.8338	0.6183	0.9799
	Mean	**0.9569**	**0.9097**	**0.9867**

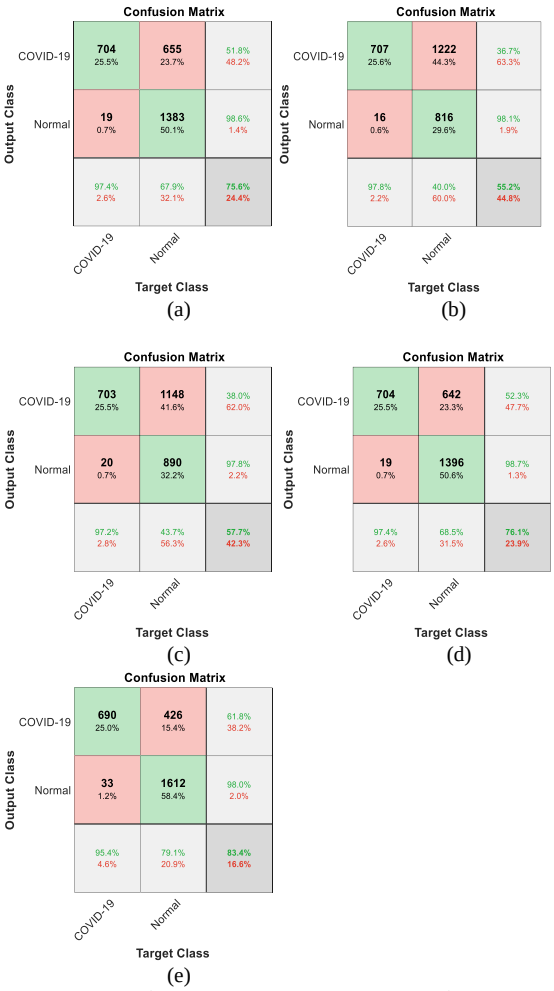


Fig. 7. The confusion matrices obtained using the five models for fold-5 results: a) ResNet-50, b) Inception-ResNet-v2, c) Inception-v3, d) MobileNet-v2, e) GoogleNet.

V. CONCLUSION AND FUTURE WORK

This paper described a comparative study of five deep learning models for CXR COVID-19 images from publically available dataset. These models were used to automatically classify the COVID-19 CXR images into two classes. Among the five models tested, GoogleNet have yielded the highest validation accuracy of 95.69%.

In our future work, we will tackle the problem of generalizing the proposed model to wider range of practical scenarios in order to facilitate diagnosis of more types of diseases from CXR and CT images.

REFERENCES

[1] C. Chen, G. Gao, Y. Xu, L. Pu, Q. Wang, L. Wang, W. Wang, Y. Song, M. Chen, L. Wang, F. Yu, S. Yang, Y. Tang, L. Zhao, H. Wang, Y. Wang, H. Zeng, and F. Zhang, "Sars-cov-2-positive sputum and feces after conversion of pharyngeal samples in patients with COVID-19," Annals of Internal Medicine, 2020.

[2] COVID-19 RADIOGRAPHY DATABASE, https://www.kaggle .com/tawsifurrahman/covid19-radiography-database. Accessed June 15, 2021.

[3] H. X. Bai, "Ai augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest ct," Radiology, 2020.

[4] K. E. Asnaoui and Y. Chawki, "Using x-ray images and deep learning for automated detection of coronavirus disease," J. Biomol. Struct. Dyn., 2020.

[5] U. Ozkaya, S. Ozturk, and M. Barstugan, "Coronavirus (COVID-19) classification using deep features fusion and ranking technique," Springer, 2020.

[6] I. D. Apostolopoulos and T. A. Mpesiana, "COVID-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," Phys. and Eng. Sci. in Med., Jul. 2020.

[7] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks," Pattern Analysis and Applications, 2021.

[8] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in x-ray images," arXiv, 2020, [accessed 15-June-2021]. [Online]. Available: https://arxiv.org/abs/2003.11055.

[9] M. Chieregato, F. Frangiamore, M. Morassi, C. Baresi, S. Nici, C. Bassetti, C. Bnà, and M. Galelli, "A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data," arXiv, 2021. [accessed 20-June-2021]. [Online]. Available: https://arxiv.org/pdf/2105.06141.

[10] S. J. Pan and Y. Qiang, "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, 2009.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Thirty-first AAAI conference on artificial intelligence, 2017.

[13] 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.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.