FACE MASK-WEARING CLASSIFICATION USING TRANSFER LEARNING TECHNIQUE WITH MOBILENET V2

PHÂN LỚP ẢNH MẶT NGƯỜI ĐEO KHẨU TRANG SỬ DỤNG KỸ THUẬT HỌC CHUYỂN TIẾP TỪ MẠNG MOBILENET V2

Nguyen Minh Nhut Pham, Duc Hien Nguyen, Nga Le Thi Thu
Vietnam - Korea University of Information and Communication Technology, The University of DaNang, DaNang

ABSTRACT: *The COVID-19 pandemic has adversely affected the global economy, politics, society, and other areas. It has dramatically led to a loss of human lives worldwide and is continuing to have a heavy impact on people's health and living conditions. In order to avoid the spread of this pandemic, mask-wearing is required by law or regulation in most places, especially indoors public places. Using Deep Learning techniques to detect and authenticate people wearing masks which can help recognize patterns or behavior of the public, contributing to limiting the rapid spread of COVID-19 pandemic is becoming effective, beneficial, and widespread. In spite of the fact that wearing face masks incorrectly will not protect ourselves, especially children, from virus transmission and will reduce the effectiveness of COVID-19 prevention, a limited study has been done to solve the problem of mask-wearing. We use the technique of transfer learning with MobileNet V2 to train the model from the Face Mask Label Dataset (FMLD) and Flickr Faces HQ (FFHQ) dataset to not only detect if a mask is used or not, but also classify the status of mask wearing over the faces. The results show that the training accuracy of experiment model is 99%.*

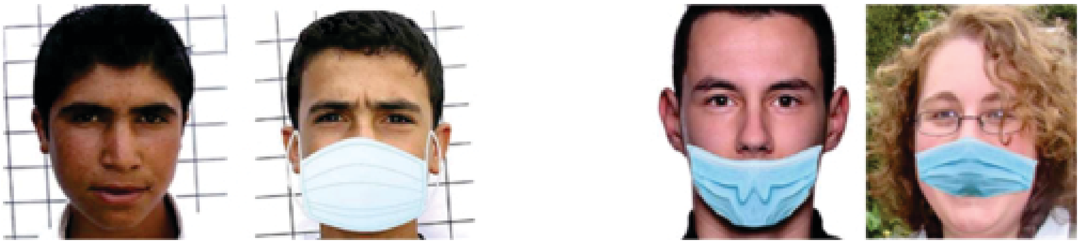
Keywords: *COVID-19, masked face detection, face mask classification, face mask recognition, Deep Learning.*

TÓM TẮT: *Dịch bệnh COVID-19 đã ảnh hưởng nghiêm trọng đến kinh tế, xã hội, chính trị trên toàn cầu. Nhằm hạn chế sự lây lan của dịch bệnh, quy định bắt buộc con người đeo khẩu trang nơi công cộng đang trở nên phổ biến. Sử dụng kỹ thuật Deep Learning để phát hiện khuôn mặt người đeo khẩu trang có thể giúp giám sát hành vi của con người, góp phần hạn chế đại dịch COVID-19. Tuy nhiên, đeo khẩu trang không đúng quy cách sẽ giảm hiệu quả phòng ngừa Dịch COVID-19. Bài báo này, với mục đích sử dụng kỹ thuật học chuyển tiếp đối với mạng Mobilenet V2 trên tập dữ liệu Face-Mask-Label Dataset (FMLD) và tập dữ liệu Flickr-Faces-HQ Dataset (FFHQ) nhằm phân loại các trạng thái đeo khẩu trang trên khuôn mặt người. Kết quả cho thấy, sử dụng kỹ thuật học chuyển tiếp có độ chính xác đạt 99%.*

Từ khóa: *COVID-19, masked face detection, face mask classification, face mask recognition, Deep Learning.*

1. INTRODUCTION

To prevent the rapid spread of the COVID-19 pandemic, the World Health Organization, governments as well as health professionals around the world recommend people to wear face masks in public places. In order to effectively monitor mask-wearing, many studies on face mask-wearing detection have enchanted the consideration of the world's computer vision community. These existing researches not only create machine learning models to detect mask-wearing but also collect data and construct masked face datasets for training machine learning models [1-3], [6].



(a) (b) (c) (d)

Fig.1. The 03 class of Face Mask-Wearing [10][11]
(a) Without mask; (b) Correct mask; (c) and (d) Incorrect mask

For instance, to build face mask detection and classification models, most current studies use Convolutional Neural Network (CNN) network models such as Chen et al. [4] have developed a mobile application using KNN method. With this, three results are obtained. The model they developed achieves 82.87% accuracy. Nagrath et al. [5] used MobileNet V2 architecture as the foundation for model to automatically detect whether a person is wearing a mask or not. The accuracy of their model is 92.64%. Mata et al. [7] offer a CNN model to detect between masked and non-masked person. It is applied a deep learning method that uses image data and video data.

However, most of these studies involve differentiating between the use of a face mask or not. In spite of the fact that wearing face masks incorrectly will not protect ourselves, especially children, from virus transmission and will reduce the effectiveness of COVID-19 prevention, a limited study has been done to solve the problem of face mask-wearing status. Therefore, this paper aims to focus on the following aspects: (1) using transfer learning techniques with MobileNet V2 [18] to create a 3-class classifier including correct mask, incorrect mask, and without mask (Figure 1); (2) use FFHQ dataset [9] and FMLD dataset [10] to train the model.

The rest of the paper is as follows: Section 2 presents related works; Section 3 describes transfer learning method with MobileNet V2; Section 4 exhibits experimental results of model training; Section 5 shows the conclusions along with future work.

2. RELATED WORKS

In the past few years, Artificial Intelligence techniques have proven their significance in monitoring and warning behavior of people when face masks are being worn incorrectly. Existing problems of classifying ways that people wear masks using machine learning techniques are divided into two types: 2-class classification problem solving the detection of a facial mask or not and 3-class classification problem regarding the problem of face mask-wearing status: correct mask, incorrect mask and without mask.

For the former, Chen et al. [4] have developed a mobile application using KNN method that allows defining the life of masks, indicating what stage they are in and how effective it is after a period of use. They use a KNN network, which experimentally obtained with the accuracy of "normal use" and "not recommended" being 92% and 92.59%, corresponding. Nagrath et al. [5] have developed a system that uses the SSMB detector as a face detector and the MobileNet V2 network as the classifier model, the accuracy of this model is 92,64% and F1-score is 93%. Similarly, Mata et al. [7] created a CNN model using the Image Data Generator as a face detector for discrimination with 60% accuracy. Jauhari et al. [10], with the aim of detecting facial patterns and the presence of facial masks in images, they used Cascade Viola Jones face detection system and AdaBoost classification model, the experimental precision of this study is 90.9%. Sen et al. [11], through a series of images and videos, developed a system which based on the MobileNet V2 network, the model accuracy is 79.24%. Kurlekar et al. [12] have developed a deep learning system that can be used in public places.

For the latter, in addition to detecting the use or not of a mask, the need to be able to detect when a person is wearing a mask in a wrong way is also very essential. Following are the most typical works for this type of problem. Rudraraju et al. [14] developed an application which based on two stages. The first stage is to detect the use/non-use of a facial mask using the Haar Cascade Classifier. In second stage, after face detection, it distinguishes between its correct and incorrect use using MobileNet models, the accuracy of this system is about 90%. Wang et al. [15] proposed using a machine learning technique to detect facial masks which using a two-step approach. In the first step, a person wearing a face mask is detected using the FRCNN model. The second step verifies the real face mask using the classifier through the InceptionV2 classifier, the accuracy for simple scenarios and complex scenarios being 97.32% and 91.13%, corresponding. Qin et al. [16] used the Multitask Cascaded CNN for face mask-wearing recognition. To do so, they used a super resolution image processing technology and SRCNet model. Their method is based on the four steps technique: image pre-processing, face image detection and cropping, super resolution image, and masked face image detection. Their model precision using this method is 98.7%.

3. TRANSFER LEARNING WITH MOBILENET V2

3.1. MobileNet V2

MobileNet V2 [18] is the improved version of the MobileNet and MobileNet V1 that achieve higher precision, fewer parameters, and fewer calculations. Therefore, this model can be applied to embedded systems with limited computing resources. These systems have embedded in commercial applications and industrial applications. MobileNet V2 is used a shortcut connection technique and it is called the "bottleneck residual block". The structure of this technique, previous layer blocks are connected directly to the next one. The main function of shortcut connections is to restrain the number of data dimensions (channels) at the input and output of each residual block. The intermediate layers in a residual block is used the nonlinear transformation technique. In between layers in a bottleneck, MobileNet V2 is used depth separated convolution to reduce the number of model parameters. Therefore, the size of the MobileNet V2 model is reduced to a minimum. The architecture of MobileNet V2 is shown in Figure 2.

3.2. Transfer Learning Technique

Transfer learning is a technique that uses a pre-trained model as a way to learn knowledge from data it was not originally trained on. There are two types of transfer learning. One is feature extraction from a pre-trained CNN network. When performing feature extraction, the pre-trained network is treated as a feature extractor which changes the aim of the previous CNN. The other effective method is fine-tuning weights of a CNN network. In this method, we first cut off its final set of fully connected (FC) layers which calling the "head" of the network from a pre-trained CNN. We then replace the head with a new set of FC layers. Therefore, this method can lead to higher accuracy than feature extraction.

In this paper, we use the second method (i.e. the fine-tuning weights of a CNN network) to implement transfer learning with MobileNet V2 model. Therefore, in the following section we will fine-tune the architecture of the MobileNet V2 network. We first load MobileNet V2 model after cutting off

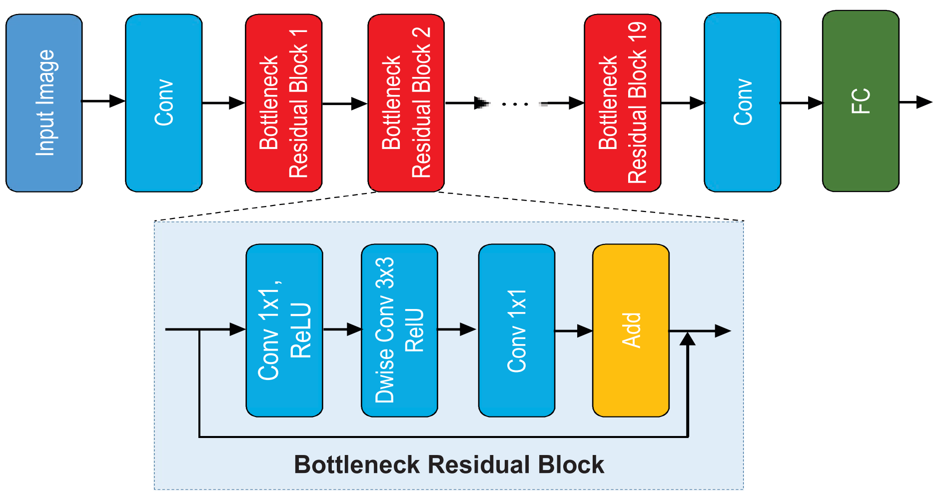


Fig.2. MobileNetV2 model architecture [18]

"head" of network and fine-tune the parameters of model with pre-trained ImageNet weights. Next, we replaced a new FC head in the base classes and freeze the base layers of the MobileNet V2. So, the weights of these base layers will not be updated but the weights of head layer will be tuned. Next, we train the model by a small learning rate so that the new FC layers can learn data from the previously learned convolutional layers in the network. Finally, we unfreeze the rest of the network and continue training until obtaining enough accuracy.

3.3. Data augmentation

Data augmentation is the image processing techniques to generate new training data from the original data so that the class labels remain unchanged. This technique is used when the training data set is not enough. So, applying data augmentation is to increase the generalizability of the model. Currently, the commonly used image augmentation techniques include: image translations, image rotations, changes in scale of image, image shearing, image horizontal or image vertical flips, and noise adding for image. Applying image augmentation allows obtaining substantially higher classification accuracy while helping to dramatically reduces over fitting.

4. EXPERIMENT RESULTS

In order to train the model, we used the FMLD [9] dataset and FFHQ dataset [10]. The dataset of FFHQ publicly made available online by NVIDIA Corporation. This dataset consists of 70,000 high quality images. The dataset of FMLD contains two datasets: the Correctly Masked Face Dataset (CMFD) and the Incorrectly Masked Face Dataset (IMFD). The FMLD datasets contains 137,016 images which generated the FFHQ image datasets. So, dataset for our model training has contains 207.016 images for 3 classes: without mask, correct mask and incorrect mask. We train face mask classifier with the Keras, Tensorflow library, compile the model with the Adam optimizer [18], configure a decreasing learning rate schedule, and use the Categorical Cross-entropy [18] loss function with parameters as follow: start learning rate = 0.0001, Epochs = 10 and batch size = 32.

In this experiment, we will cover the models for image classification that are state of the art and are widely used in the industry as well, consists of the VGG16 [19] is one of the most popular models for image classification. InceptionV3 [20] has a good accuracy. ResNet [21] was increased the network depth so that accuracy increases. DenseNet [22] has the significant improvements and requiring less computation to achieve high performance.

For MobileNet V2 model [18], we have carried out two sets of experiments: (1) Training MobileNet V2 model from the dataset; (2) using the Transfer Learning technique and fine-tuning method (as presented in Section 3.2) to train MobileNet V2 from the dataset generated from data augmentation on the dataset and we have called a MobileNet V2* model. We replaced the head of network with a new set of FC layers that is shown in Table 1, there in:

- Flatten layer: after using the AverageMaxpooling2D layer, we used a flatten layer to flat the whole network. This layer transforms the entire pooled feature map matrix into a single column.
- Dense layer: after the flatten layer, we have used two dense layers. The dense layer is also known as a FC layer. A FC layer provides the correct probabilities for the different classes as it computes the products between the weights and the previous layer.
- Activation function: we have used Softmax function with another dense layer and ReLU activation function which has proved to give best performance with Average Maxpooling2D.
- Dropout layer: dropout technique is a regularized technique which helps prevent over fitting. The input of the dropout layer is the output of the previous layer after the convolution layer.

Table 1. The new set of fully-connected layers

Layer (type)	Output Shape	Parameter
AveragePooling2D	1, 1, 1280	0
Flatten	1280	0
Dense	128	163968
ReLU	128	0
Dropout	128	0
Dense	3	387

The experiment results indicate that MobileNet V2* accuracy achieved with transfer learning technique is 99% and the training time is short (Table 2, Table 3). The training time of MobileNet V2 is shorter than ones of MobileNet V2*. Because the training dataset for MobileNet V2* is larger than ones for MobileNet V2 (using Data augmentation). The time of trainning is calculated by second (s) and is measured by a virtual machine having an Intel(R) Xeon(R) CPU E5-2667 v2 @ 3.30 GHz (2 processors), RAM 96 GB.

5. CONCLUSIONS

In this study, we trained a 3-class model including correct mask, incorrect mask, and without mask. We use the technique of transfer learning with MobileNet V2 to train the model from the FMLD and FFHQ dataset. This results in a more accurate transfer learning model than the mo del training from scratch and achieved 99% accuracy. Using the MobileNet V2 architecture, it is also computationally efficient, and can make it easy to deploy the model to embedded systems with limited computing resources. In future studies, other CNNs should be explored for increasing accuracy for training, and experiments on real datasets are designed to evaluate model performance ensuring real-time requirements

Table 2. The experiment results

Model	Accuracy	Parameters	Epochs	Times(s)
VGG-16	97.50%	14,780,739	10	9504
ResNet50	98.25%	23,850,371	10	4546
InceptionV3	98.70%	22,065,443	10	4105
DenseNet121	98.81%	7,169,091	10	4392
MobileNet V2	97.50%	2,422,339	10	2082
MobileNet V2*	99.00%	2,422,339	10	3054

Table 3. Training evaluation results of MobileNetV2* model for epochs = 30

Labels	precision	recall	f1-score	support
Correct mask	0.99	0.99	0.99	9752
InCorrect mask	0.99	0.99	0.99	9830
Without mask	1.00	1.00	1.00	10611
accuracy			0.99	30193
macro avg	0.99	0.99	0.99	30193
weighted avg	0.99	0.99	0.99	30193

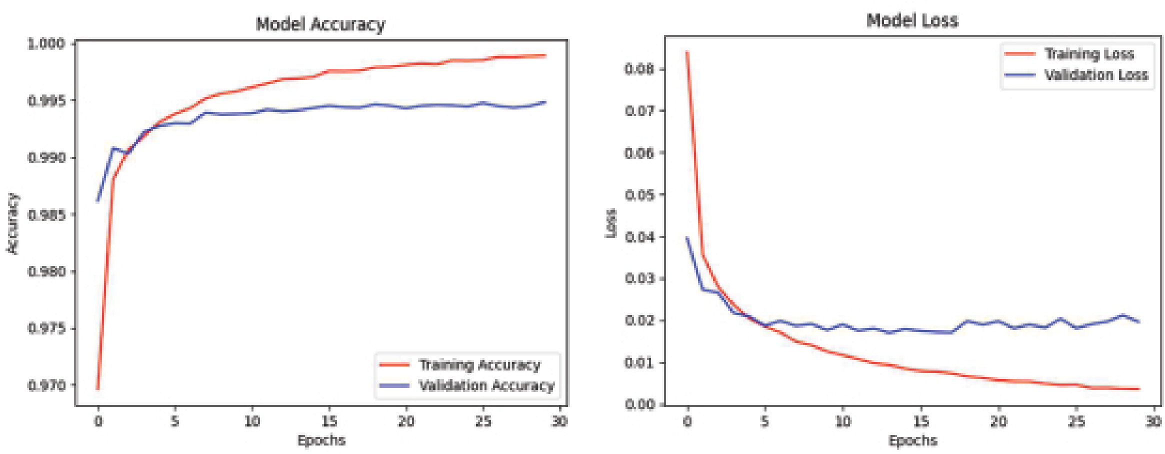


Fig.3. Training plot of MobileNet V2* model for epochs = 30

REFERENCES

[1] Zafeiriou, S., Zhang, C., Zhang, Z. *A (2015), survey on face detection in the wild: past, present and future*. Comput. Vis. Image Underst, Vol. 138, pp. 1-24.

[2] Ge, S., Li, J., Ye, Q., Luo, Z.(2017), *Detecting masked faces in the wild with LLE-CNNs*. In Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition, pp. 2682-2690.

[3] Jiang, M., Fan, X.(2021), *RetinaFaceMask: A Single Stage Face Mask Detector for Assisting Control of the COVID-19 Pandemic*. Computer Vision and Pattern Recognition.

[4] Chen, Y., Hu,M., Hua, C., Zhai, G., Zhang, J., Li, Q., Yang,S.X. (2021), *Face mask assistant: Detection of face mask service stage based on mobile phone*. IEEE Sens. J, Vol. 21, pp. 11084-11093.

[5] Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., Hemanth, J. (2021), *SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNet V2.* Sustain. Cities Soc., Vol. 66.

[6] Batagelj, B., Peer, P., Štruc, V., Dobrišek, S. (2021), *How to Correctly Detect Face-Masks for COVID-19 from Visual Information?*. Appl. Sci., Vol. 11, 2070.

[7] Mata, B.U. (2021), *Face Mask Detection Using Convolutional Neural Network*. Journal of Natural Remedies, Vol. 21, pp. 14-19.

[8] Tero Karras, T., Laine, S., Aila, T. *A (2019), Style-Based Generator Architecture for Generative Adversarial Networks*. Neural and Evolutionary Computing.

[9] Cabani, A., Hammoudi, K., Benhabiles, H., Melkemi, M. (2021), *MaskedFace - Net A dataset of correctly/incorrectly masked face images in the context of COVID-19 Smarth Health*. Vol. 19.

[10] Jauhari,A., Anamisa,D.R., Negara, Y. D. P. (2021), *Detection system of facial patterns with masks in new normal based on the Viola Jones method*. Journal of Physics Conference Series.

[11] Sen, S., Sawant, K.(2021), *Face mask detection for covid_19 pandemic using pytorch in deep learning*. In IOP Conference Series: Materials Science and Engineering, Vol. 1070, pp. 1-6.

[12] Kurlekar, M.S. (2021), *Face Mask Detection System Using Deep Learning*. Turk J. Comput. Math. Educ, Vol.12, pp. 1327-1332.

[13] He, K., Zhang, X., Ren, S., Sun, J. (2016), *Deep Residual Learning for Image Recognition*. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.

[14] Rudraraju,S.R., Suryadevara,N.K., Negi, A.(2020), *Face Mask Detection at the Fog Computing Gateway*. In Proceedings of the 2020 IEEE 15th Conference on Computer Science and Information Systems (FedCSIS), pp. 521-524.

[15] Wang, B., Zhao, Y., Chen, C. P.(2021), *Hybrid Transfer Learning and Broad Learning System for Wearing Mask Detection in the COVID-19 Era*. IEEE Trans. Instrum. Meas., Vol. 70, pp. 1-12.

[16] Qin,B., Li,D. (2020), *Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19*. Sensors 2020, Vol. 20(18).

[17] Diederik, P.K., Jimmy, L.B. (2015), *Adam: A Method for Stochastic Optimization* . In ICLR 2015, pp. 1-15.

[18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L. C. (2018), *MobileNet V2: Inverted Residuals and Linear Bottlenecks*. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4510-4520.

[19] Simonyan, K., Zisserman, A. (2015), *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Computer Vision and Pattern Recognition, pp.1-14, arXiv:1409.1556.

[20] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Zbigniew, W.Z. (2015), *Rethinking the Inception Architecture for Computer Vision*. Computer Vision and Pattern Recognition, arXiv:1512.00567

[21] He, K., Zhang,X.,Ren, S.,Sun, J. (2015), *Deep Residual Learning for Image Recognition*. Computer Vision and Pattern Recognition, pp. 1-12, arXiv:1512.03385.

[22] Huang, G., Liu, Z., Maaten, L. V. D., Kilian Q. Weinberger, Q., K. (2018), *Densely Connected Convolutional Networks*. Computer Vision and Pattern Recognition, arXiv:1608.06.

Liên hệ:

TS. Phạm Nguyễn Minh Nhựt

Trường Đại học CNTT&TT Việt - Hàn, Đại học Đà Nẵng

Địa chỉ: 470 Trần Đại Nghĩa, Quận Ngũ Hoàng Sơn, TP. Đà Nẵng

Email: pnmnhut@vku.udn.vn

Ngày nhận bài: 01/8//2022

Ngày gửi phản biện: 03/8/2022

Ngày duyệt đăng: 11/10/2022