Khulna University Studies Special Issue ICSTEM4IR: 531-538, 2022

ORIGINAL ARTICLE SCIENCE & ENGINEERING

IDENTIFICATION OF FACE MASK USING CONVOLUTIONAL NEURAL NETWORK-BASED EFFICIENTNET MODEL

Bhugol Bijoy Chakma*, M. A. Masud, Tafsir Ahamed, Mahadi Hasan Tusher

Faculty of Computer Science and Engineering, Patuakhali Science and Technology University, Patuakhali-8602, Bangladesh

KUS: ICSTEM4IR-22/0096

Manuscript submitted: June 27, 2022 Accepted: September 26, 2022

Abstract

The highly contagious coronavirus wreaked havoc around the globe. There has been a rapid spread of the virus throughout the world. According to the World Health Organization (WHO), wearing a facemask will help keep the virus from infecting others. Consequently, many governments have adopted the solution of wearing facemasks. In this paper, we use a convolutional neural network (CNN), and a scaling method, namely, EfficientNet with Adam optimizer, for detecting face masks in real-time. A dataset including 10,000 colored images was collected from a public data platform, Kaggle, for training, and testing of the model. Image augmentation is also investigated on the dataset to improve the training, and testing accuracy. Then the binary classifier model is used to detect masks after detecting faces using single shot detection (SSD). From the experimental results, the EfficientNet model outperforms the existing CNN-based methods in terms of accuracy, efficiency, and the validation accuracy of EfficientNet models is above 99 percent. This efficient and highly accurate model can be used to detect facemask anywhere in a real time video surveillance system.

Keywords: Convolutional Neural Network, Coronavirus, EfficientNet, Face Mask Identification, Image Augmentation, Single Shot Detection.

Introduction

The coronavirus epidemic, also known as COVID-19, began in Wuhan, China, and has since spread to numerous nations. Countries with a huge population are having difficulty keeping the coronavirus from spreading. COVID-19 may be regulated in two unique ways in society. The first way is to create social distance, which is a simple and cost-free strategy. The second, and most effective method is to use a facemask to cover the nose, and mouth (Kwon). Sanitizers are also a good technique to cut down on transmission. These have shown promising results in terms of reducing disease transmission. The major portion of the time, it is transferred via surfaces in an indirect manner. Droplets from a person's lungs can spread the virus from person to person, and the incubation period can be quite long, extending up to 14 days in severe instances (Coronavirus disease). To prevent the transmission of

*Corresponding author: <bhugol14@cse.pstu.ac.bd>
DOI: https://doi.org/10.53808/KUS.2022.ICSTEM4IR.0096-se

illness, a variety of measures were used, including social isolation, forced indoor mask wear, quarantine, restricting people's movement inside and outside the state boundaries, self-isolation, and the restriction and suspension of big social meetings and conferences.

COVID-19 is becoming a major public health and economic concern due to the virus's negative impacts on people's quality of life, leading to acute respiratory illnesses, death, and financial crises throughout the world. COVID-19 infected above 485 million people, resulting in six million lives, according to the (Worldometer). COVID-19 spreads quickly in crowded places and through close contact. Despite the fact that vaccines have been developed, poor countries are still having difficulty obtaining adequate vaccines for their citizens. Most of the time, people in many countries have to wear face masks when they go out in public. This means that masked face identification is still important for applications like object recognition. Governments will need help and surveillance of people in public places to fight and win the COVID-19 pandemic, particularly those that are populated, to ensure that laws prohibiting the use of face masks are followed. This may be done by merging surveillance technology with AI models.

In this study, we present EfficientNet, a face mask detection approach based on deep transfer learning and classical machine learning binary classifiers that detects face masks in real time utilizing single shot detection (SSD)[14]. The suggested model may be used in conjunction with security cameras to prevent the spread of COVID-19 by detecting people who are not wearing face masks. In order to assess the accuracy of each design suited for detecting the usage of a mask, we examined numerous pre-trained convolutional neural network architectures. We introduced the most appropriate method, which obtained the maximum accuracy while consuming the least amount of time throughout the training and detection processes.

The following is the structure of the next section of this paper: The associated works are briefly described in Related Work Section. The working technique is described in Face Mask Identification Using Efficientnet with Adam Optimizer Section along with a full description of model implementation. Results and discussion Section contains the results and an explanation of the working model. The conclusion and future work are drawn in the final portion of the study.

Related work

In the recent past, numerous researchers and analysts largely concentrated on object identification from an image owing to its being widely applied in many fields. One of them is a real-time face mask detection system. Much study has previously been undertaken concerning the issue. As a consequence, the accuracy and precision of the model are increasing day by day. Researchers have demonstrated the importance of wearing masks to decrease the spread of coronavirus. Our purpose in this study is to identify persons who don't use masks in crowded places.

Vivekanadam Balasubramaniam has just published a paper on face mask detection using the CNN-based EfficientNet architecture (Vivekanadam, 2021). In this research, accuracy, precision, recall, and f1 score are studied among the EfficientNet, VGG19, Inceptionv3, and ResNet50 classifiers. Among them, the EfficientNet obtained an accuracy of 97.12 percent. But there were no optimizers applied. This accuracy can be enhanced with optimizers. Besides, just the trained model is presented, but no mechanism involving face identification is mentioned in the study.

The paper (Sikakulya, 2021) provided by the authors, investigated how frequently individuals of various ages, genders, educational levels, and jobs wear face masks. In a face-to-face interview, they gathered comments about COVID-19 related topics. But it didn't have any implementation of the mask detecting mechanism. It just examined the data obtained from the participants.

Khulna University Studies Special Issue (ICSTEM4IR): 531-538:2022
DOI: https://doi.org/10.53808/KUS.2022.ICSTEM4IR.0096-se

In (Singh, 2021), the authors used a dual shot face detector (dsfd) for detecting faces and a MobileNetv2 based binary classifier for training the model, which performs a lot slower than a single shot face detector. Over the course of 78 frames, the average prediction time was 7.12 seconds, or 10.96 frames per second. There is no graphics processing unit (GPU) in a surveillance camera, and the processing power is significantly lower than that of a typical personal computer. However, it is better to use a single-shot detector for this purpose.

The authors of (Goyal, 2022) suggested a convolutional neural network model for the identification of face masks in real time. Additionally, it outperforms DenseNet-121, VGG-19, MobileNet-V2, and Inception-V3. After utilizing 4000 image dataset labeled 'with_mask' and 'without_mask' they achieved a performance accuracy rate of 98 percent during training.

Face Mask Identification Using EfficientNet with Adam Optimizer

The proposed work focuses on identifying face masks using convolutional neural network based architecture EfficientNet with adam optimizer (Kingma, 2014). The suggested work is split into two phases: the first consists of the building of the EfficientNet model, and the second stage consists of the identification of the facemask. The work is separated into two sections because of the nature of the proposed work.

Dataset description

The dataset for this article was downloaded from kaggle, a public data platform (Jangra). The dataset included over 10,000 RGB images with and without masks. There were 5000 images used in total for the mask and another 5000 total images used without the mask. Eighty percent of these were used to train the model, while the remaining twenty percent were used to test and validate the EfficientNet model.

Pre-processing

Real data always contains some noise. So they are essential to preprocess (García, 2015) to make them suitable for the training. So some of the photographs had to be deleted manually. Then the photos were downsized to 224x224 and necessary to preprocess using imagenet utils to adapt the image to the format the model demands. Finally the data is supplemented with such things as rotation, zooming, shearing, flipping etc. It helps to lower the chance of overfitting.

Data augmentation

Data augmentation (Van, 2001) is a technique for generating fresh training data from an existing dataset. The method was used to reduce overfitting. The enrichments included a 20-degree rotation range, 20% zoom range, 20% width/height shift, 20% shearing, and horizontal flip. These processes are applied randomly. The procedures have been demonstrated to be quick, consistent, and trustworthy.

Face mask classification

EfficientNet is a CNN-based architecture that was designed with a limited resource budget in mind. Then, it is scaled up for greater accuracy in accordance with the available resources. Eight EfficientNet models are available, ranging from EfficientNetB0 to EfficientNetB7. We used EfficientNetB0 for building the model. In Figure 1, we can see the flow diagram of the categorization model. After preprocessing the photos, they are split into training and testing components. The dataset is used for training the model 80 percent of the time, with the remaining 20 percent being used for testing and validation of the model. The initial learning rate utilized in training the model is 0.0001, batch size of 32, and EPOCHS 20. The learning rate of a model determines how fast it adapts to a

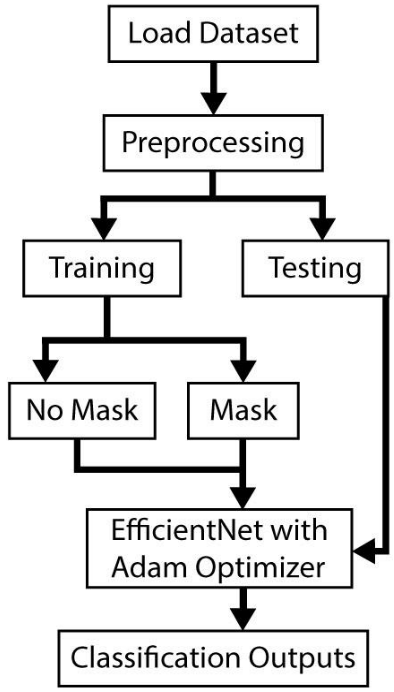

Figure 1. Flow Diagram of Face Mask Classifier.

particular environment. Before the model is updated, the batch size is the number of samples handled. The number of epochs is the total number of times the training dataset has been traversed. During training, the Adam optimizer is used to boost the efficiency and accuracy of the model. The Adam optimizer creates adaptive learning rates for each parameter. It utilizes an adaptive gradient approach that never underperforms in unusual scenarios (Choi, 2019).

Face mask detection

In this section, there are two parts. Detecting faces in a frame is the first step in the process. The model then determines if a mask is present or not. For face detection, an OpenCV DNN was built utilizing the ResNet-10 architecture and the Single Shot Multi-Box Detector. In terms of speed and accuracy, the face detector is highly accurate and efficient (Modern-Day F ace Detection with the OpenCV Library). The single-shot method, on the other hand, is thought to be the best compromise between speed and accuracy (Soviany, 2018). An image of just the face is created after the face has been detected. 224x224 resizing and post-processing are applied to the newly created image. Afterwards, the model determines whether or not the face has a mask and displays the results. In figure 2, the flow diagram of the face mask identification model is presented.

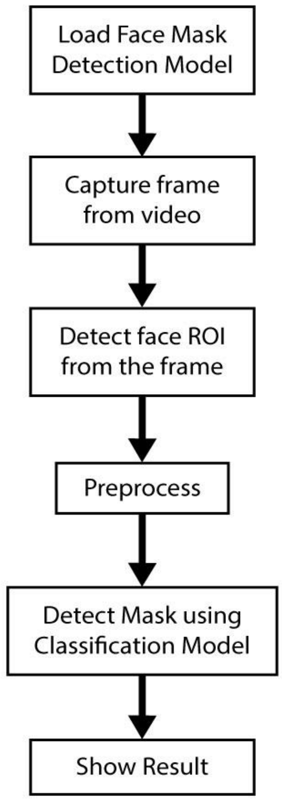

Figure 2. Face Mask Detection Model.

Result and discussion

The following Equations (1),(2).(3), and (4) are used to calculate the accuracy, recall, precision, and f1 score respectively. True positives, false positives, true negatives and false negatives are all abbreviated as TP, FP, TN, and FN respectively. The number of times each class appears in y_true is the support.

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

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+Fn} \quad (3)$$

$$\text{F1 score} = 2 \times \frac{Precision \times Recall}{Pression + Recall} \quad (4)$$

Table 1. Face mask detection model comparison

Models	Accuracy	Precision	Recall	F1 Score
MobilenetV2	93.2	94.6, 93.80	95.7, 94.1	95.1, 93.9
EfficientNet	97.12	99.10	94.42	96.70
EfficientNet with Adam Optimizer	99.46	99.73	99.19	99.46

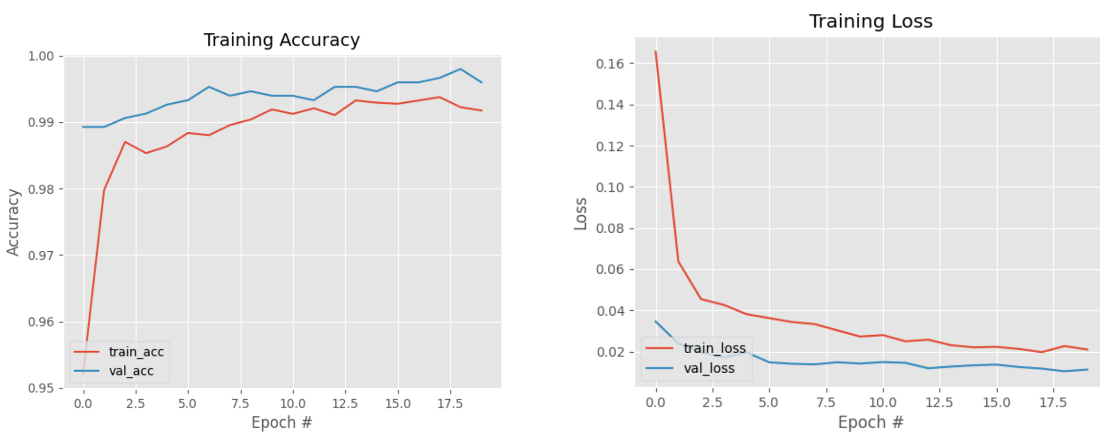

Figure 3. Learning Curves for Training Accuracy and Training Loss.

Table 1 compares the EfficientNet and MobilenetV2 model to our proposed model in terms of accuracy, precision, recall, and f1 score. Here we can see that our proposed model with Adam optimizer achieves an accuracy of 99.46, precision 99.73, recall 99.19 and f1 score of 99.46. Which are better than EfficientNet and MobilenetV2. In Figure 3, the training loss and training accuracy graphs are given. In this case, the model is generalized fine because the validation accuracy is higher than the training accuracy.

Table 2 summarizes the findings of the investigation. After 20 epochs of training with this model, it was possible for us to acquire a training accuracy rate of 99.46 percent. After training the classification model it could detect masks from the video stream of the webcam. The system is depicted in Figure 4.

Figure 4. Mask and No mask output.

Table 2. Classification report

	Precision	Recall	F1-score	Support
Mask	1.00	0.98	0.99	1000
No_mask	0.98	1.00	0.99	1000
Accuracy			0.99	2000
Macro avg	0.99	0.99	0.99	2000
Weighted avg	0.99	0.99	0.99	2000

Conclusion

The EfficientNet architecture with the Adam optimizer is used in this paper to demonstrate a real-time face mask identification system based on CNN for establishing surveillance at crowded places. The dataset of 5000 images with masks and the 5000 images without masks used in this work are taken from kaggle. Those images are used to train the EfficientNet model. Then the model is used to detect masks in video streams. The system also outperforms in terms of accuracy, recall, precision, and f1 score. It offers a 99.46% accuracy to the problem of identifying people in public places who don't wear masks.

References

Coronavirus disease (COVID-19). (n.d.). WHO | World Health Organization. Retrieved from https://www.who.int/health-topics/coronavirus#tab=tab_2

COVID Live - Coronavirus Statistics. (n.d.). Worldometer. Retrieved from https://www.worldometers.info/coronavirus/

Balasubramaniam, V. (2021). Facemask Detection Algorithm on COVID Community Spread Control using EfficientNet Algorithm. *Journal of Soft Computing Paradigm*, *3*(2), 110-122.

Jangra, A. (n.d.). *Face Mask Detection ~12K Images Dataset*. Kaggle. Retrieved from https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset

Kwon, S., Joshi, A. D., Lo, C. H., Drew, D. A., Nguyen, L. H., Guo, C. G., & Chan, A. T. (2021). Association of social distancing and face mask use with risk of COVID-19. *Nature Communications*, 12(1), 1-10.

Sikakulya, F. K., Ssebuufu, R., Mambo, S. B., Pius, T., Kabanyoro, A., Kamahoro, E., & Kyamanywa, P. (2021). Use of face masks to limit the spread of the COVID-19 among western Ugandans: Knowledge, attitude and practices. *Plos one*, 16(3), e0248706.

Singh, R. R., Nayak, N., Srinivasan, S., & Biradar, R. (2021). COVID-19 Monitoring System using Social Distancing and Face Mask Detection on Surveillance video datasets. arXiv e-prints, arXiv-2110.

Goyal, H., Sidana, K., Singh, C., Jain, A., & Jindal, S. (2022). A real time face mask detection system using convolutional neural networks. Multimedia Tools and Applications, 1-17.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In the International conference on machine learning (pp. 6105-6114). PMLR.

Choi, D., Shallue, C. J., Nado, Z., Lee, J., Maddison, C. J., & Dahl, G. E. (2019). On empirical comparisons of optimizers for deep learning. arXiv preprint arXiv:1910.05446.

P. Soviany and R. T. Ionescu, Optimizing the Trade-Off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction. *20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 2018, pp. 209-214, doi: 10.1109/ SYNASC.2018.00041.

Modern-Day Face Detection with the OpenCV Library. (n.d.). INSCALE. Retrieved from https://www.inscale.net/news/modern-face-detection-opencv/

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In European conference on computer vision, 21-37. Springer, Cham.

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*, 72, 59-139). Cham, Springer International Publishing.

Van Dyk, D. A., & Meng, X. L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1), 1-50.