Face Mask Detection Using Deep Learning

Sufia Jasmin Binti Saiful Azian[1], Hu Ng[1(✉)], Timothy Tzen Vun Yap[1], Hau Lee Tong[1], Vik Tor Goh[2], and Dong Theng Cher[3]

[1] Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia
nghu@mmu.edu.my
[2] Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia
[3] SIRIM Berhad, Shah Alam, Malaysia

Abstract. It is vital to remain vigilant during pandemic COVID-19. Wearing a face mask is one of the crucial steps that people must take to ensure that they are a step away from spreading and infecting the virus. However, controlling and monitoring people in a densely crowded place is tough. Hence, a face mask detection system in public area is needed to remotely monitor if one is wearing a face mask or vice versa. In this study, two face masks datasets are downloaded from GitHub with 3834 images and 11800 colour images. Data pre-processing steps are carried out before the classification, which includes image resizing, converting images into array and label encoding. Two deep learning models, MobileNetV2 and VGG19, are developed for detection and evaluation. The experimental results performed by MobileNetV2 outperformed the VGG19 with achieving accuracy of 98.96% and 99.55% on Dataset 1 and Dataset 2 respectively.

Keywords: MobileNetV2 · VGG19 · Accuracy · Face mask detection · Real time camera

1 Introduction

The world is facing an emergency health situation due to the emerging Coronavirus (COVID-19) that started in 2019. COVID-19 is a coronavirus disease produced by a novel strain. The letters 'CO' represent corona, 'VI' for virus, and 'D' for disease. Previously, this illness was known as the '2019-nCoV'. The COVID-19 virus is a novel virus that is related to the severe acute respiratory syndrome (SARS). COVID-19 pandemic became a global health concern because of how quickly it spreads. The virus can spread in the form of small liquid particles when one who is infected coughs, sneezes, speaks or breathes. Infections can cause dyspnoeal, fever, and cough, which can lead to pneumonia, septic shock, organ failure, and death in extreme cases [1]. Wearing a face mask along with other preventive measures such as physical distancing and getting vaccinated can help to reduce the rapid spread of the virus.

When it comes to respiratory diseases, face masks are a type of protective equipment that may be useful in preventing the transmission of respiratory viruses. Surgical masks are composed of numerous layers of non-woven plastic and may successfully filter very small particles, such as SARS-CoV-2 droplets which is the virus that causes COVID-19.

S.-C. Haw and K. Sonai Muthu (Eds.): CITIC 2022, 10, pp. 279–288, 2022.
https://doi.org/10.2991/978-94-6463-094-7_22

Typically, the masks have an external waterproof layer and an inside absorbent layer [2]. According to [3], it is discovered that wearing a face mask can lower the chance of getting infected by more than 80%. Thus, wearing a mask in public areas is recommended by scientists to avoid spreading of the disease.

In 2015, Nieto-Rodriguez [4] developed a technique for surgical face marks. However, this research was focused on the detection of face masks using conventional pattern recognition methods that involved many image processing steps and cannot be implemented in real times. Since then, a considerable amount of research has been carried out by deep learning approaches such as MobileNetV2, VGG16, InceptionV3, ResNet50, and YOLOv3 have been used to perform the detection of face masks in faster speed and even real time.

2 Related Works

Many researchers are working on the problem of detecting face masks. Chowdary et al. [5] proposed an image augmentation technique to improve the testing and training of the model on the dataset Simulated Masked Face Dataset with 1099 images. They applied InceptionV3 that was developed as a Googlenet module to assist in image processing and object detection. An accuracy of 99.92% was reported during training and 100% was achieved during testing.

Nagrath et al. [6] proposed SSDMNV2 approach for the face mask detector, which utilized Single Shot Multibox Detector as the face detector and MobilenetV2 architecture for classifier. The tested dataset was a combination of several datasets which were collected from Kaggle and PyImageSearch (5521 images). The accuracy obtained by the method used in this study is 92.64%.

Vinh et al. [7] applied Haar Cascade classifier to detect faces and YOLOV3 for mask detection on a 5000 images dataset. The model achieves up to 90.1% accuracy and can work in real time camera.

Loey et al. [8] proposed a hybrid deep transfer learning model with ML approaches for face mask detection. The hybrid model is trained and tested on three datasets, Real World Masked Face Dataset (RWMFD), Labeled Faces in the Wild Dataset (LFWD) and Simulated Masked Face Dataset (SMFD). They employed ResNet50 to extract features from images before feeding them to 3 standard ML classifiers which are the decision tree, SVM, and KNN, linear regression, and logistic regression. SVM scored the highest accuracy with of 99.64% in RWMF, 99.49% in SMFD, and 100% in LFWD.

Asif et al. [9] applied OpenCV and ML to recognize and track faces. Then, MobileNetV2 was employed to determine the mask region from the processed face frames. The model was tested on Face Mask ~12K Images Dataset (11,800 images) and obtained 99.8% validation accuracy on 800 images.

Sadeddin et al. [10] utilised a pretrained CNN, ResNet-50 model to connect to a 300 linear layer network for face masks detection system. The model was tested on Face Mask ~12K Images Dataset (11,800 images) and obtained 99% validation accuracy on 800 images.

Table 1. Details of Dataset 1 and Dataset 2.

Dataset	Total number of images	Images with masks	Images without masks
Dataset 1	3834	1915	1919
Dataset 2	11800	5883	5917

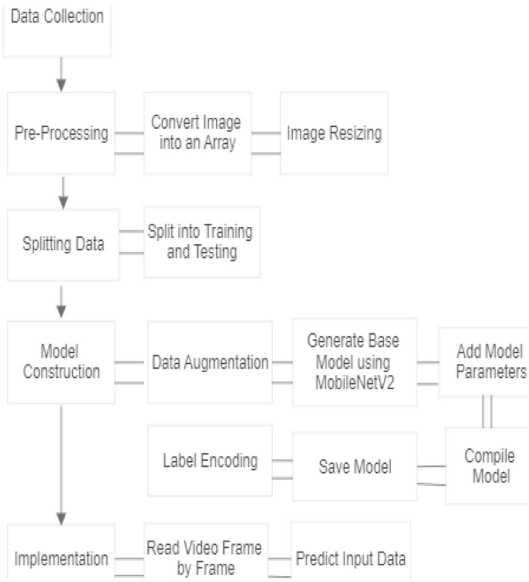

Fig. 1. Process flow

3 Methodology

In this research, six major stages are designed for the proposed research work, which are data acquisition, data pre-processing, splitting data into training and testing, model development, and implementation via a real time camera. Figure 1 depicts the process flow of the proposed method.

3.1 Data Collection

Two well-known face masks datasets have been downloaded, namely Face-Mask-Detection (Dataset 1) [11] and Face Mask ~ 12K Images Dataset (Dataset 2) [12]. All the images in both datasets are in RGB colour mode. The details of datasets are shown in Table 1. As the count of images are considered balance between the images without masks and images with masks. Therefore, data oversampling approach is not required in this work.

Table 2. Number of images before and after image augmentation.

Dataset	Number of images before augmentation	Number of images after augmentation
Dataset 1	3834 images	6230 images
Dataset 2	11800 images	13228 images

3.2 Pre-processing

The pre-processing stage consists of resizing, converting image into an array and label encoding. Resizing every image from the dataset to uniformly distributed with 224 × 224 pixels. Then, the following step is to create an array out of all the images in the dataset. The images are transformed into an array, which is then used by the loop function to call the images.

The final step is to do one-hot encoding on labels. Images are labelled as numerical label, which will allow the algorithm to comprehend and interpret the image in the later stage.

3.3 Splitting Data

The images are divided into two batches, training data and testing data. Each batch contains images of people with masks and without masks. For this paper, the datasets are split into 50–50, 60–40, 70–30 and 80–20 train-test ratios.

3.4 Model Construction

The following phase is model construction. The model is constructed in six steps which are creating the training image generator for augmentation, generating the base model using MobileNetV2, adding model parameters, compiling the model, training the model, and finally, saving the model for future prediction.

ImageDataGenerator is applied to perform data augmentation. It produced 10 different augmented images in each iteration of the training was carried out during model construction. Table 2 shows the number of images before and after augmentation.

3.5 Evaluation Metrics

The evaluation metrics in this work are measured by accuracy, precision, recall, and f1-score. TP, TN, FP, and FN are denoted by the true positive, true negative, false positive, and false negative observations respectively.

$$Accuracy = \frac{Tp + Tn}{(Tp + Fp + Fn + Tn)} \times 100\% \tag{1}$$

$$Precision = \frac{Tp}{(Tp + Fn)} \tag{2}$$

Table 3. Classification result by MobileNetV2 on Dataset 1.

Train-Test ratio	Precision	Recall	F1-Score	Accuracy
50:50	0.9810	0.9805	0.9804	0.9804
60:40	0.9838	0.9835	0.9835	0.9835
70:30	0.9838	0.9835	0.9835	0.9835
80:20	0.9896	0.9896	0.9896	0.9896

$$Recall = \frac{Tp}{(Tp + Fn)} \tag{3}$$

$$f1 - score = 2 * \frac{Recall * Precision}{(Recall + Precision)} \tag{4}$$

3.6 Experimental Setup

The research experiment is carried out using a MacBook Pro 2020 build with an M1 chip CPU and 8 GB RAM. The experiment is implemented in Visual Studio Code development environment. For the purposes of this study, each dataset is separated into several training and testing sets as 50:50, 60:40, 70:30 and 80:20 train-test split ratios.

By conducting a series of experiments with MobileNetV2 and VGG19 as the deep learning model, it is possible to assess the effectiveness of the suggested model for the face mask detection system.

4 Results and findings

4.1 MobileNetV2 and VGG19 Performance Comparison on Colour Images

This section presents the comparison between MobileNetV2 with VGG19. All the original color images from Dataset 1 and Dataset 2 are employed to compare and evaluate the two models. The comparison is evaluated by precision, recall, f1-score, and accuracy without any bias since both models were evaluated on the same dataset. The results are presented in Tables 3, 4, 5, and 6.

From Tables 3, 4, 5, and 6, the 80:20 train-test split ratio produces the highest accuracy score across all other train-test ratios evaluated. This is because the 80% of training size has provide larger sample size for a model to be trained, and a larger dataset results in greater diversity (Dataset 2 over Dataset 1). By training the model on a larger number of samples, it minimises generalisation error. A greater number of training examples results can be a reduced in test-error rate [13].

MobileNetV2 is found outperformed VGG19 by a distance. MobileNetV2 attained a higher accuracy compared to VGG19. This is because MobileNetV2 is low-latency and a low-power model that have been customised to meet the resource restrictions of certain use cases [14]. This also proven on MobileNetV2 performs faster than VGG19.

Table 4. Classification result by VGG19 on Dataset 1.

Train-Test ratio	Precision	Recall	F1-Score	Accuracy
50:50	0.9048	0.9048	0.9048	0.9048
60:40	0.9421	0.9421	0.9421	0.9421
70:30	0.9497	0.9487	0.9487	0.9487
80:20	0.9511	0.9511	0.9511	0.9511

Table 5. Classification result by MobileNetV2 on Dataset 2.

Train-Test ratio	Precision	Recall	F1-Score	Accuracy
50:50	0.9802	0.9802	0.9802	0.9802
60:40	0.9875	0.9875	0.9875	0.9875
70:30	0.9940	0.9940	0.9940	0.9940
80:20	0.9955	0.9955	0.9955	0.9955

Table 6. Classification result by VGG19 on Dataset 2.

Train-Test ratio	Precision	Recall	F1-Score	Accuracy
50:50	0.9696	0.9686	0.9686	0.9686
60:40	0.9762	0.9760	0.9760	0.9760
70:30	0.9772	0.9772	0.9772	0.9772
80:20	0.9774	0.9773	0.9773	0.9773

4.2 MobileNetV2 and VGG19 Performance on Grayscale Images

MobileNetV2 and VGG19 were examined on grayscale images on Datasets 1 and Dataset 2 to determine the classification on grayscale images. The models were evaluated with a train-test ratio of 80:20. Table 7 summarises the outcomes obtained by the models on Dataset 1 and Dataset 2 in terms of accuracy, precision, recall, and F1-score.

As observed from Table 7, The accuracy rates archived by both models on grayscale images are lower about 1–2% as comparing with the original RGB images. Although the computation on grayscale images is less heavy as comparing to colour images, the classification results are poorer. This is because colour images carry a significant quantity of information, which could increase the amount of training data required to attain satisfactory results [15].

Table 7. Result obtained by MobileNetV2 and VGG19 on grayscale images.

Model	Dataset	Precision	Recall	F1-Score	Accuracy
MobileNetV2	Dataset 1	0.9735	0.9735	0.9735	0.9725
	Dataset 2	0.9810	0.9805	0.9804	0.9804
VGG19	Dataset 1	0.9432	0.9426	0.9426	0.9426
	Dataset 2	0.9657	0.9655	0.9655	0.9655

Table 8. Comparison on Dataset 1.

Model	Precision	Recall	F1-Score	Accuracy
Balaji S [11]	0.9200	0.9100	0.9100	0.91
MobileNetV2	0.9896	0.9896	0.9896	0.9896
VGG19	0.9774	0.9773	0.9773	0.9773

Table 9. Comparison on Dataset 2.

Model	Precision	Recall	F1-Score	Accuracy
MobileNetV2 [9]	–	–	–	0.9920
ResNet-50 [10]	–	–	–	0.99
MobileNetV2	0.9955	0.9955	0.9955	0.9955
VGG19	0.9774	0.9773	0.9773	0.9773

4.3 Comparison with Others Research Finding

The comparison of the proposed model MobileNetV2 and VGG19 model with the finding of Balaji [13] on the Dataset 1 is presented in Table 8. The comparison is conducted using a train test ratio of 80:20.

The comparison of the proposed models with the finding of Asif et al. [9] and Sadeddin [10] on Dataset 2 is presented in Table 9. The comparison is conducted using a train test ratio of 80:20.

4.4 Real Time Face Mask Detection System

The MobileNetV2 model is apply in developing the face mask detection system model as it has a higher accuracy score and performs better than VGG19. A bounding box is created to roam around human's face along with the label and the percentage score.

Figure 2 shows the model being implemented in a real time camera under various condition and scenarios. The models can be implemented on real time camera. Green indicates face mask detected. Red indicates no face mask detected. Number indicates confidence level.

Fig. 2. Model being implemented in a real time camera under various conditions and scenarios

5 Conclusion and Future Work

The aim of this paper is to develop a face mask detection system using a deep learning model to classify and detect whether a person is wearing a face mask or vice versa.

Two models, MobileNetV2 and VGG19 are examined and evaluated to finalize the best model in implementing the face mask detection system. The experimental results performed by MobileNetV2 outperformed the VGG19. MobileNetV2 attained an accuracy of 98.96 and 99.55% on Dataset 1 and Dataset 2 respectively. MobileNetV2 is found having smaller parameter size as comparing to VGG19 and it is a powerful feature extractor for detecting objects.

In future, more deep learning methods will be studied in order to enhance the performance of the face mask detection system. Besides that, this research aims to develop Malaysian own face masks dataset for the training and testing on the proposed face mask detection system.

Acknowledgments. The authors would like to acknowledge staff and friends from Multimedia University for providing support in carrying out this research.

Authors' Contributions. Hu Ng, roles: Corresponding author, Conceptualization, Project Administration, Supervision, Validation, Visualization, Writing—Review & Editing.

Sufia Jasmin Binti Saiful Azian, roles: Data Curation, Formal Analysis, Investigation, Resources, Writing—Original Draft Preparation.

Timothy Tzen Vun Yap, roles: Conceptualization, Investigation, Administration, Supervision, Validation, Writing—Review & Editing.

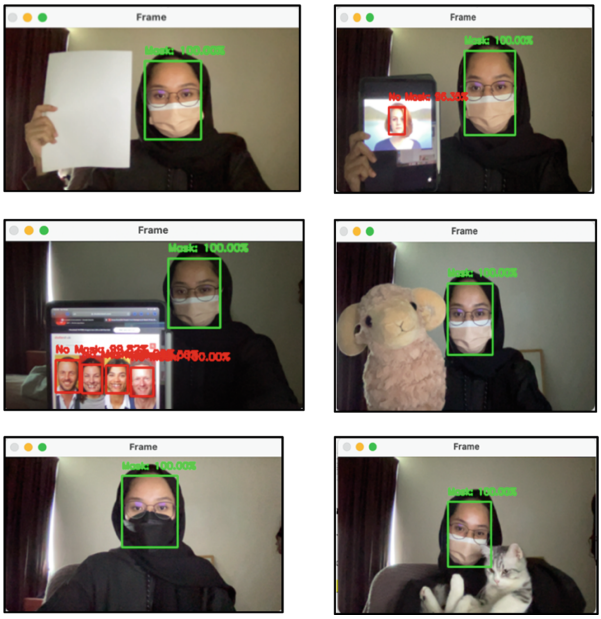

Fig. 2. (*continued*)

Hau Lee Tong, roles: Conceptualization, Validation, Visualization, Writing—Review & Editing.

Vik Tor Goh, roles: Conceptualization, Validation, Visualization, Writing—Review & Editing.

Dong Theng Cher, roles: Conceptualization, Validation, Visualization, Writing—Review & Editing.

References

1. *Coronavirus disease (COVID-19)*. (2021, November 3). World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019?gclid=CjwKCAiAvriMBhAuEiwA8Cs5lb3JozPjrkLba5jsx0cXiz5XrRmoTwOVM59oxoB3q4NTuVMTHoL25hoCVxMQAvD_BwE
2. Desai, A. N., & Mehrotra, P. (2020). Medical masks. *Jama*, *323*(15), 1517–1518.
3. Ueki, H., Furusawa, Y., Iwatsuki-Horimoto, K., Imai, M., Kabata, H., Nishimura, H., & Kawaoka, Y. (2020). Effectiveness of face masks in preventing airborne transmission of SARS-CoV-2. *MSphere*, *5*(5), e00637-20.
4. Nieto-Rodriguez, A., Mucientes, M., & Brea, V. M. (2015, June). System for medical mask detection in the operating room through facial attributes. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 138–145). Springer, Cham.
5. Chowdary, G. J., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020, December). Face mask detection using transfer learning of inceptionv3. In *International Conference on Big Data Analytics* (pp. 81–90). Springer, Cham.

6. 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 MobileNetV2. *Sustainable cities and society*, *66*, 102692.
7. Vinh, T. Q., & Anh, N. T. N. (2020, November). Real-time face mask detector using YOLOv3 algorithm and Haar cascade classifier. In 2020 International Conference on Advanced Computing and Applications (ACOMP) (pp. 146–149). IEEE. doi: https://doi.org/10.1109/ACOMP50827.2020.00029.
8. Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement, 167, 108288.
9. Asif, S., Wenhui, Y., Tao, Y., Jinhai, S., & Amjad, K. (2021, May). Real Time Face Mask Detection System using Transfer Learning with Machine Learning Method in the Era of Covid-19 Pandemic. In *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)* (pp. 70–75). IEEE.
10. Sadeddin, S. (2020). *Face mask detection trained on ResNet - 50*. The Notebook Archive | Wolfram Foundation. https://notebookarchive.org/face-mask-detection-trained-on-resnet-50--2020-12-6z8vw5z/
11. Balaji S. (2020, July 9). balaji srinivasan. Face-Mask-Detection https://github.com/balajisrinivas/Face-Mask-Detection/tree/011f1f7ed02256a43394ef8d518805935c30bc8b/dataset
12. 'Face Mask Detection ~12K Images Dataset' (2020, May 26). Kaggle. https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset
13. Brownlee, J. (2019). Deep learning for computer vision: image classification, object detection, and face recognition in python. Machine Learning Mastery.
14. Hossain, M. U., Rahman, M. A., Islam, M. M., Akhter, A., Uddin, M. A., & Paul, B. K. (2022). Automatic driver distraction detection using deep convolutional neural networks. Intelligent Systems with Applications, 14, 200075.
15. Kanan, C., & Cottrell, G. W. (2012). Color-to-grayscale: does the method matter in image recognition?. PloS one, 7(1), e29740.