Al-Nahrain Journal of Science
ANJS, Vol.26 (4), December, 2023, pp. 80-87

Automated Face Mask Detection Using Pretrained CNN

Farah Saad Al-Mukhtar*

Department of Computer Science, College of Science, Al-Nahrain University, Jadriya, Baghdad, Iraq

Article's Information	Abstract
Received: 09.08.2023 Accepted: 07.12.2023 Published: 15.12.2023 **Keywords:** CNN COVID-19 Face Mask Detection (FMD) SARS	In recent times, the use of face masks has emerged as a critical subject. Automated facial mask detection can curb the transmission of the COVID-19 virus and SARS-VIRUS within communal areas by identifying individuals who are not utilizing masks. In this work, a pretrained Convolutional Neural Network (CNN) with model ResNet-50, which is initially trained on the ImageNet competition data, is utilized. This model is augmented with a 300-linear layer network and fine-tuned on a dataset that comprises 1,000 facial images. During the evaluation of the validation dataset consisting of approximately 800 face images, the model achieved an impressive 99% accuracy. Its primary objective is to discover if an individual is wearing a facial mask using a cropped image of their face. By leveraging such advanced technologies, we can contribute significantly to public health and safety measures in the ongoing battle against COVID-19 and SARS-VIRUS.

DOI: 10.22401/ANJS.26.4.12

*corresponding author email: farah.saad@nahrainuniv.edu.iq

1. Introduction

COVID-19 (COronaVIrus Disease of 2019) and SARS (Severe Acute Respiratory Syndrome) are two viral respiratory illnesses caused by coronaviruses, both of which have had important global impacts on public health [1]. These contagious diseases emerged in different time frames but share similarities in their mode of transmission and clinical presentation, prompting the search for effective non-pharmaceutical strategies to mitigate their spread [2]. One essential strategy to reduce the transmission of respiratory droplets, which can carry the viruses, is the widespread adoption of face masks in public spaces. However, monitoring and ensuring compliance with mask-wearing practices in crowded areas can be challenging, particularly in densely populated regions. Automatic face mask detection (FMD) systems powered by advanced technologies offer a potential solution to this challenge [3]. Convolutional Neural Networks (CNNs) and deep learning techniques have demonstrated remarkable success in computer vision tasks, including image classification [4]. In this context, utilizing pretrained CNN models, such as ResNet-50, has shown promising results for identifying faces wearing or not wearing masks. In this work, we explore applying a pretrained ResNet-50 CNN model fine-tuned on a well-balanced dataset of 12,000 cropped photos of faces with and without masks. The goal is to create an accurate and reliable FMD (Face Mask Detection) system that can automatically identify individuals who are not wearing masks in public spaces. Through this work, we aim to contribute to the expanding realm of research on FMD as a crucial tool to prevent the spread of respiratory illnesses. The successful implementation of such technology has the ability to play a significant role in safeguarding public health, improving safety measures, and aiding in the ongoing battle against COVID-19 and SARS in various community settings.

2. Literature Review

In [5], the authors used Deep Learning, Keras, Tensorflow, and OpenCV to create a face cover finder. For the best results, they used the MobileNet V2 classifier with the ADAM analyzer to prepare it to distinguish between people wearing masks and people who are not wearing them. With still images and continuous video transmissions, the model is tested. It focused on each person's face and applied the face veil classifier to it after recognizing the face from the photos or videos. Across two distinct datasets, the method attains accuracies of 95.77% and 94.58%, respectively.

In [6], the authors employed OpenCV, Tensor Flow, Keras, Pytorch, and Deep Learning to ascertain whether individuals were wearing facial masks or not. The models were assessed using both images and live video feeds. The accuracy of the model has been attained, and model optimization is a constant process in which they adjust the hyperparameters to produce an extremely accurate response. This model might be used to show how edge analytics functions. Additionally, the proposed approach produced cutting-edge results using a publicly available face mask dataset.

In [7], the authors proposed a deep learning algorithm to automatically check to see if anyone is wearing a face mask. The transfer learning technique is used to fine-tune the pre-trained Deep learning model Faster R-CNN Inception V2; it was developed and evaluated using the Simulated Masked Face Dataset (SMFD). The TensorFlow environment's trained model was precise enough to find the face mask. As a result, the LabVIEW interface was used to detect face masks and a secure workplace can be maintained by controlling security gaps in public living areas.

In [8], the authors used thermal imaging in their study to examine the potential for locating (detection) masks on faces as well as to see if it was possible to categorize the many types of masks that could be found on faces. The previously proposed thermal imaging collection was expanded and accompanied by a description of the kind of mask and its location on the face. Various deep learning models were modified. The "nano" version of the Yolov5 model, boasting a mAP exceeding 97% and an approximate precision of 95%, emerged as the optimal choice for face mask detection. Notably, it also demonstrated remarkable accuracy in classifying different mask types. The CNN model, which originated from an initial autoencoder trained to address thermal image reconstruction concerns, yielded the most favorable outcomes. A classifier with a 91% accuracy was taught using the pre-trained encoder.

In [9], they employed the YOLOv3 deep learning object detection model in their research. Their objective was to develop an automated system capable of detecting both faces and face masks in crowded scenarios, aiming to streamline the process of identifying individuals wearing masks. The facial region and mask presence detection model was created by concatenating the YOLOv3 object identification algorithm with a variety of backbones, such as ResNet-50 and Darknet-53. The researchers collected datasets from online sources such as Kaggle and GitHub. Then, they curated and properly annotated the images for further use. A total of 4,393 images were utilized for training the models, and their performance was assessed in terms of recall, precision, detection time, and mean average precision. As a result of DarkNet53_YOLOv3's superior performance over ResNet50_YOLOv3 in terms of accuracy (mAP of 95.94%) and speed (detection time of 50 seconds on 776 photos), it was selected as the better model.

In [10], they used three special computer models called VGG16, Resnet50, and MobileNet. First, they prepared the data for the models by setting it up nicely. Then, these models were tested to determine how well they could understand the data. VGG16 and Resnet50 got 96% accuracy, while MobileNet did even better with 98%. These test results showed that MobileNet is the best for recognizing masks. This means that when using MobileNet, they don't need to use as many computer resources as the other models. MobileNet is innovative because it uses something called the Depth-Wise Separable method, which makes it work faster without using too much computer power.

In [11], A deep learning algorithm called MobileNetV2 was used to build the face mask detector model that can tell if someone is wearing a face mask. They also used a special sensor, called MLX90614, and a tiny computer called Arduino to measure body temperature without touching anyone. All the results, like whether someone is wearing a mask and their temperature, are shown on a special screen called a Graphical User Interface (GUI). Additionally, they added a cool feature that's related to the Internet of Things (IoT). This feature can send a message to your phone if someone has a very high temperature. The model is really good, and it's accurate about 98% of the time. But there's a small problem. Sometimes, if people use things to cover their mouth and nose in a funny way, the model might think they're wearing a mask when they're not. However, it's still better than some other systems you can buy because it can even tell if someone is pretending to wear a mask with their hand.

3. Problem Statement

The primary objective of this system is to contribute to preventing and controlling COVID-19 and SARS-VIRUS transmission by providing a reliable and efficient solution for mask detection. By automating this process, the system aims to assist in enforcing mask-wearing guidelines, ensuring compliance, and enhancing public health and safety measures in communal areas. The key challenges to address in developing this system include achieving high accuracy in mask detection, handling variations in lighting conditions, angles, and facial expressions, and ensuring real-time performance for the timely identification of non-compliant individuals. Additionally, the system should be scalable, adaptable, and easily deployable in various settings, such as airports, public transportation, healthcare facilities, and retail establishments.

4. Methodology and Design

Within this section, the approaches employed to address the stated problem are explained, subsequently offering an in-depth account of the accomplishments in this endeavor. We will focus on the architecture we have adopted and the various techniques that are applied to enhance our outcomes.

There are basically four stages in the procedure, which are preparing the dataset for training and testing, building and training the proposed mask detector, and performing mask detection on each Image from the dataset. Finally, we get the result of an image with or without a mask, as shown in Figure 1.

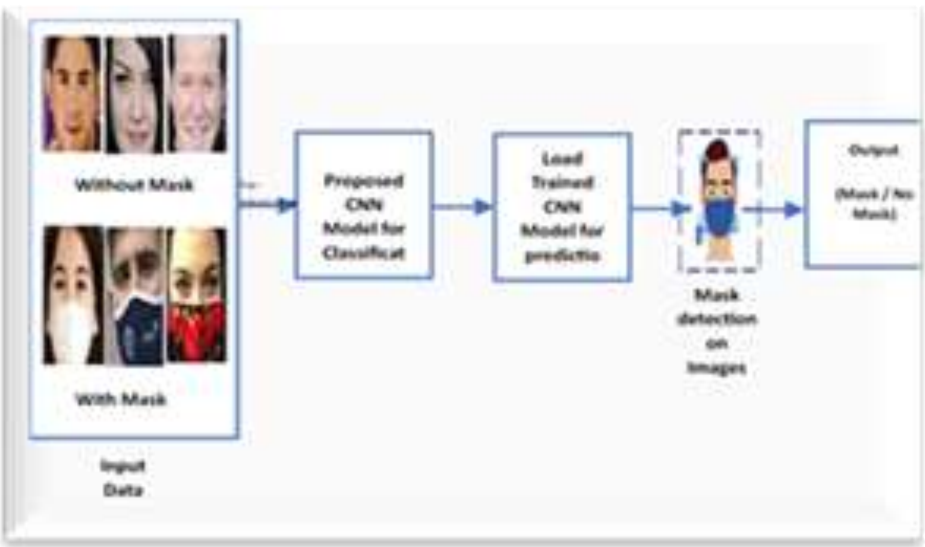


Figure 1. Stages of Image processing.

4.1 Data Exploration

The study utilized a dataset consisting of 12,000 cropped photographs portraying faces both with and without masks.
Visit "https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset/activity" to access the public data set.
The following statistics show how well-balanced this data set is:

- Train: (5000 mask- and non-mask) of 10,000 training photos
- Test: 1000 photos (500 with and 500 without a mask).
- 800 validations were performed (400 with and 400 without a mask).

These folders each have two subfolders:

1. With Mask (Images of people wearing masks)
2. Without Mask (Images of people without masks)

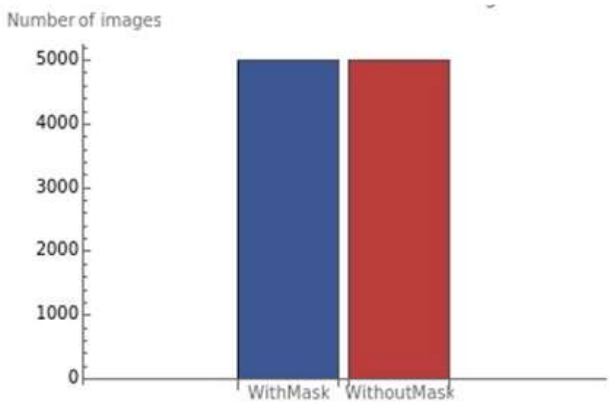
The Images will resemble this:

Figure 2. Pictures in dataset

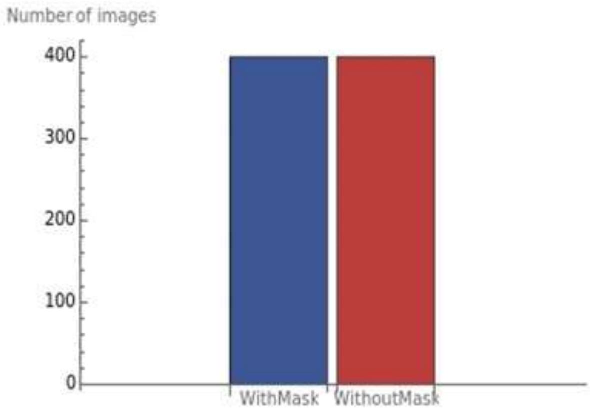
The dataset comprises square-cropped images of faces exhibiting diverse resolutions and facial orientations, as well as a variety of people's ages, genders, and ethnicities (this holds importance during the model training process using individuals' images, as avoiding the introduction of bias is crucial). Additionally, the dataset includes a selection of comic-style images, which can allow the model to classify comics as well.

4.2 Preprocessing

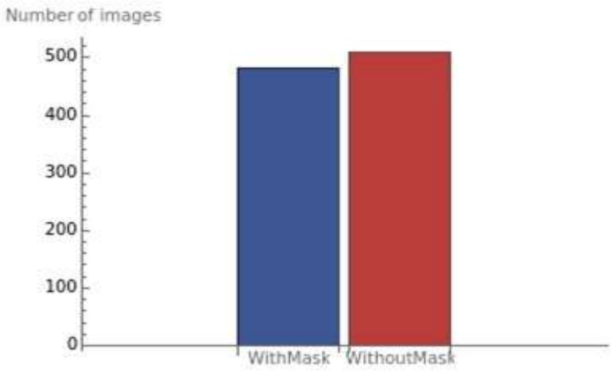
The collected dataset is preprocessed to ensure uniformity and compatibility with the CNN model. This includes resizing the images to a consistent resolution, normalizing pixel values, and converting them to the appropriate format (e.g., RGB).



(a)



(b)



(c)

Figure 3. The distribution of class for (a) train data (b) validation data (c) test data.

4.3 Distribution of the Data

The Data was distributed among train, test, and validation sets in classes (with mask/without mask) to train a custom face mask detector; each dataset's class distribution will be examined to ensure that the data is balanced as anticipated:

4.4 Model Section

The ResNet-50 model, pretrained on the ImageNet competition data, is chosen as the base model for this task. ResNet-50 is known for its excellent performance in image classification tasks.

- **ResNet-50 Architecture:**

ResNet-34 represented a weighted adaptation of the initial ResNet (Residential Energy Services Network) concept. It introduced an innovative approach to bolster the count of convolutional layers within a CNN[12], effectively sidestepping the vanishing gradient issue through the utilization of shortcut connections. These shortcut links circumvent several layers, thus transforming a standard network into a residual network[13]. The VGG neural networks, encompassing VGG-16 and VGG-19, formed the basis of the conventional architecture, featuring 33 filters in each convolutional network. Conversely [14], a ResNet is more streamlined, housing fewer filters than a VGGNet. In contrast to the 19.6 billion FLOPs of a VGG-19 Network, a 34-layer ResNet achieves 3.6 billion FLOPs, with an 18-layer ResNet, even smaller, accomplishing 1.8 billion FLOPs [15].

Two key design ideas govern the ResNet architecture. First, each layer has the same number of filters regardless of the size of the output feature map. Second, it has twice as many filters to maintain the time complexity of each layer even if the size of the feature map is cut in half [16]. The Resnet-50 architecture is a convolutional neural network (CNN) design that includes a new idea known as shortcut connections. These connections imply directing the input from the previous layer directly to the output of that layer. This concept aims to safeguard crucial features from being lost during the convolution process.

- **Pretrained model:**

The pretrained ResNet-50 model, which was trained on the ImageNet dataset http://www.image-net.org. There are almost a million photos and 1000 distinct classifications. The model weights that are generated can be used for additional computer vision tasks, accelerating the learning of the new models. Imported pretrained ResNet-50 model first, and then the final two layers were deleted(To train a new model with a different number of classes in the dataset) [17]. These layers align with the number of classes in the ImageNet dataset. Later, our tailored layers are incorporate to match the specific class count in our dataset, which, in this particular task, amounts to 2 classes.

- **Customized Neural Network**

The final two layers (LinearLayer and SoftmaxLayer) from the pre-trained ResNet-50 model are remFoved, allowing the construction of a customized neural network [18]. This tailored network configuration also involves specifying the necessary number of classes for the intended task. ResNet-50 convolutional pretrained is followed by the addition of Dropout layers and Linear layers. As a result, the model can acquire more sophisticated architecture and learn new features in addition to the ones that were previously transmitted (the pretrained weights). Following are the contributions made by each new layer [19]:

1. LinearLayer: Makes the neural network more sophisticated.
2. DropoutLayer: Stops overfitting by irrationally turning off neurons during training.
3. SoftMaxLayer: The classification task's activation function layer.

4.5 Train Model

The training process will occur in two stages, with a batch size of 16 for each round, using the metrics "Accuracy," "Precision," and "Recall" as metric monitoring [20].

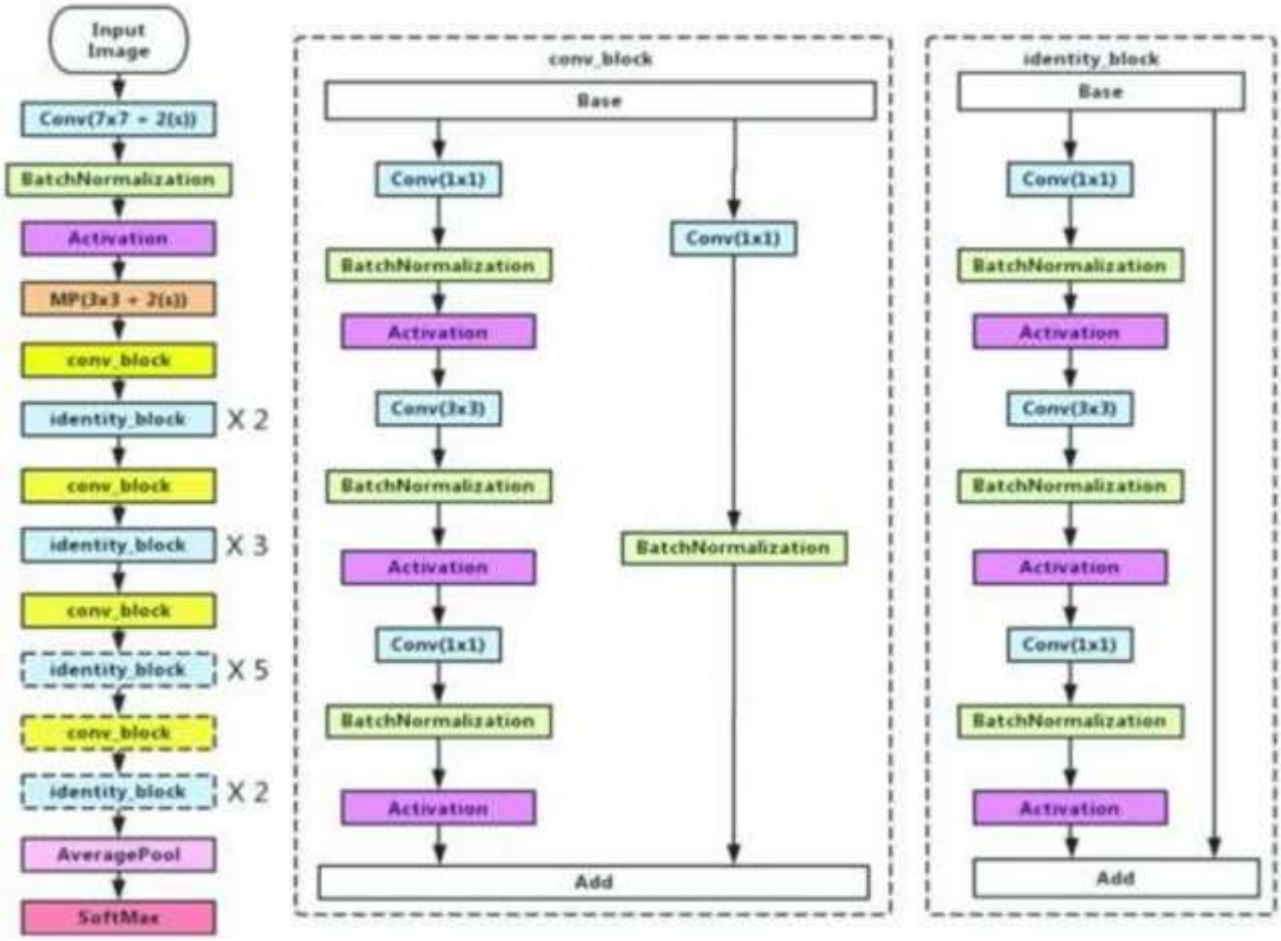


Figure 4. Resnet-50 Architecture

5. Result

The face mask detection system was trained and evaluated on a well-balanced dataset consisting of 12,000 cropped photos of faces, including both masked and non-masked individuals. The training set comprised 5,000 images, while the test set contained 1,000 photos, with 500 of them depicting faces with masks. Additionally, an 800-image validation set was used, evenly distributed between masked and non-masked faces, resulting in an impressive performance on the validation set. The model achieved a precision and recall of 99% over the validation dataset, indicating its ability to accurately classify faces with and without masks. However, it was observed that the model's performance may be affected when the resolution of the image is low. The evaluation metrics on the test set further demonstrated the model's exceptional performance. The accuracy of the model was measured at 98.75%, indicating a high percentage of correct predictions. The F1 score, which considers both precision and recall, was also 98.75% for faces with masks and 98.74% for faces without masks. The precision and recall for both classes were also 98.75%. These outstanding results highlight the model's ability to correctly classify faces with and without masks. The high F1 score indicates a balanced trade-off between precision and recall, ensuring robust performance across both classes, compared with results obtained in researches [5]-[11] that has been discussed in Literature Review. The well-balanced dataset and the model's accuracy showcase great promise for real-world applications in face mask identification, contributing to the efforts to limit the spread of COVID-19 in public places.

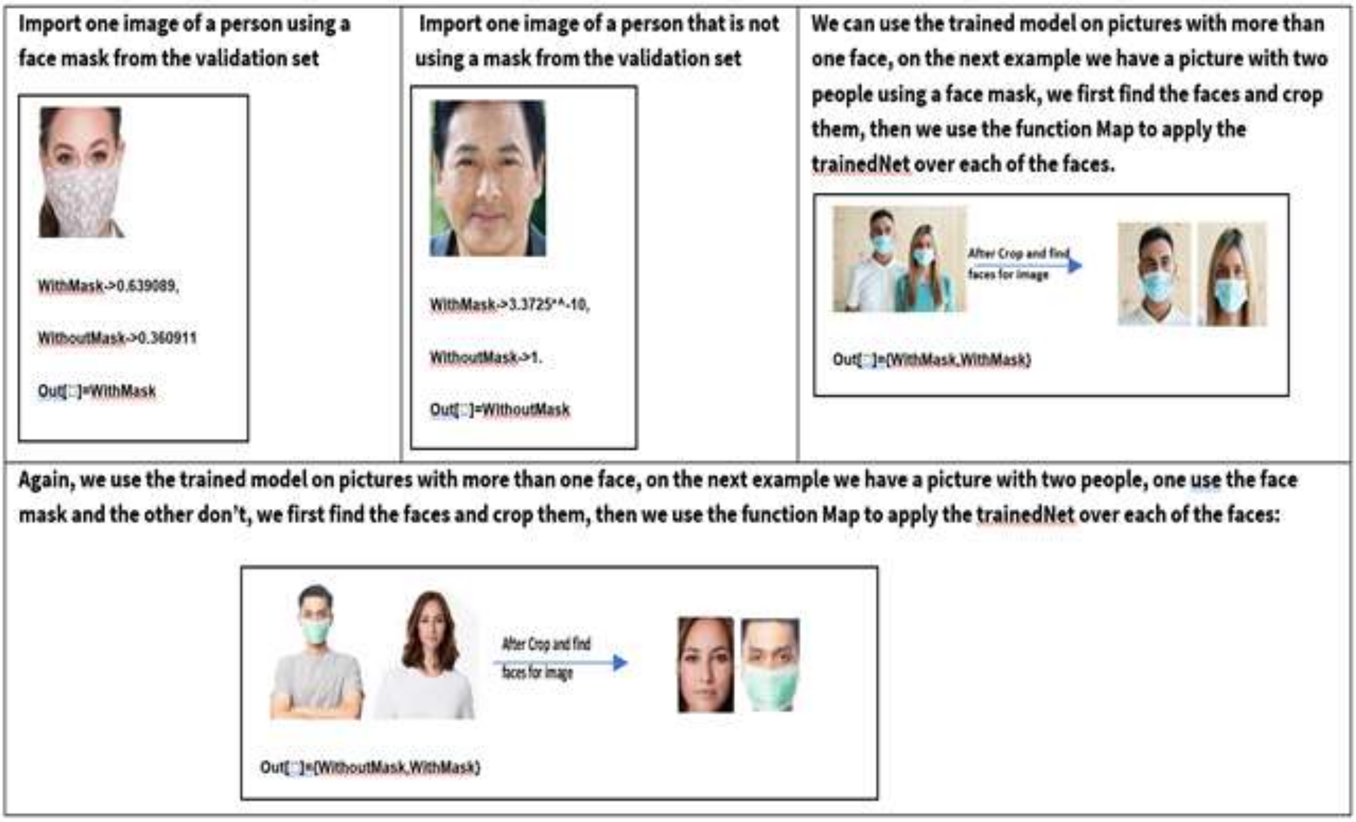


Figure 5. Outputs of proposed mask detector

6. Conclusions

In recent times, the importance of face masks has become a critical subject in combating the spread of contagious respiratory illnesses like COVID-19 and SARS. Automatic face mask detection (FMD) has emerged as a crucial non-pharmaceutical strategy to identify individuals not wearing masks in public spaces. This work utilizes a pretrained Convolutional Neural Network (CNN), specifically ResNet-50, which was fine-tuned on a well-balanced dataset of 12,000 cropped photos of faces with and without masks. The dataset comprises 10,000 training images, 1,000 test images, and 800 validation images.

The model had an outstanding accuracy of 98.75% and an F1 score of 98.75% on the test set, demonstrating its effectiveness in identifying faces with and without masks. The precision and recall for both classes were also consistently high at 98.75%. These results highlight the model's robustness and reliability for real-world face mask detection applications.

By leveraging advanced deep learning technologies like CNNs, face mask detection can play a significant role in safeguarding public health and safety. The successful deployment of such systems can aid in controlling the transmission of respiratory illnesses like COVID-19 and SARS in public spaces, thereby contributing to global efforts in combating infectious diseases.

References

[1] Tandon, R.; Sayed, A.; Hashmi, M.A.; "Face Mask Detection Model Based on Deep CNN Technique Using AWS". Int. J. Eng. Res. Appl., 13(5): 12-19, 2023.

[2] Kumar, P.; Gupta, V.K.; Singh, D.P.; "Face Mask Detection Using Convolution Neural Network". ICICT-2022,IEEE: 1-6, 2022.

[3] He, K.; Zhang, X.; Ren, S.;, Sun, J.; "Deep Residual Learning For Image Recognition".

Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.: 770-778, 2016.
[4] Soni, A.; Rai, A.; "YOLOv4: A Face Mask Detection System". Int. Res. J. Eng. Technol., 9(9), 2022.
[5] Patil, R.; Patil, S.; Kazi, A.A.; "Face Mask Detection Using AI". High Technol. Lett., 28(4): 469-475, 2022.
[6] Shukla, S.K.; Singh, P.; Haswani, D.; Khare, S.; Gupta, A.; "Face Mask Detection Using Machine Learning". Int. J. Mod. Trends Sci. Technol., 8(06): 132-136, 2022.
[7] Öztürk, G.; Eldoğan, O.; Karayel, D.; ATALI, G.; "Face Mask Detection on LabVIEW". AITA, 1(2),: 9-18, 2021.
[8] Kowalczyk, N.; Sobotka, M.; Rumiński, J. J. I. A.; "Mask Detection and Classification in Thermal Face Images". IEEE Access,11: 43349-43359, 2023.
[9] Ge, C.X.; As'ari, M.A.; Sufri, N.A.; "Multiple face mask wearer detection based on YOLOv3 approach". IAES Int. J. Artif. Intell., 12(1): 384-393, 2023.
[10] Handoko; Rozy, F.; Gani, F.E.; Dharma, A.; "Comparative Analysis of Convolutional Neural Network Methods in Detecting Mask Wear". Budapest Int. Res. Critics. Inst. J., 5(2): 16792-16801, 2022.
[11] Zhang, Y.; Al-Ataby, A.; Al-Naima, F.; "A Deep Learning-Based Tool for Face Mask Detection and Body Temperature Measurement". IEEE, 2023, DOI: 10.1109/ICSPIS57063.2022.10002688.
[12] Yapıcı, M.M.; Tekerek, A.; Topaloglu, N.; "Performance comparison ˘ of convolutional neural network models on gpu". AICT: 1–4, 2019.
[13] Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A.; "Vggface2: A Dataset for Recognising Faces Across Pose and Age". IEEE: 67-74, 2018.
[14] Olivas, E.S.; Guerrero, J.D.M.; Sober, M.M.; Benedito, J.R.M.; Lopez, A.J.S.; "Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques, 2 Volumes". Inf. Sci., Imprint of: IGI Publishing, Hershey, PA, 2009.
[15] Kirandeep, K.R.; Dhir, V.; "Image Recognition Using Resnet50". Eur. Chem. Bull., 12(S3): 7533 – 7538, 2023.
[16] Lin, C.; Wu, K.; "Development of Revised ResNet-50 for Diabetic Retinopathy Detection". BMC Bioinform., 24(1): 1-18, 2023.
[17] Celano, G.; "A ResNet-50-based Convolutional Neural Network Model for Language ID Identification from Speech Recordings". NLP: 136-144, 2021.
[18] Suherman, E.; Rahman, B.; Hindarto, D.; Santoso, H.; "Implementation of ResNet-50 on End-to-End Object Detection (DETR) on Objects". Sinkron: J. Penelit. Tek., 8(2): 1085-1096, 2023.
[19] Muhathir, R.M.; Syah, R.; Khairina, N.; Muliono, R.; "Convolutional Neural Network (CNN) of Resnet-50 with Inceptionv3 Architecture in Classification on X-Ray Image". CSOC, pp. 208-221: 2023.
[20] Rachman, S.; Bagaskara, D.; Magdalena, R.; Sa'idah, S.; "Classification of Pneumonia Based on X-Ray Images with ResNet-50 Architecture". ICEBEHI 2022, 5–6 October, pp. 117-130: Springer , 2023.