FACE MASK AND SOCIAL DISTANCE DETECTION SYSTEM

MUHAMMAD ALIFF DANIAL BIN MOHD ZAHIRUDDIN

UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

FACE MASK AND SOCIAL DISTANCE DETECTION SYSTEM

MUHAMMAD ALIFF DANIAL BIN MOHD ZAHIRUDDIN

FACE MASK AND SOCIAL DISTANCE DETECTION SYSTEM SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Muhammad Aliff Danial Bin Mohd Zahiruddin

I.C/Passport No: 980607-10-5703

Matric No: 17121897/2

Name of Degree: Computer Science and Information Technology

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Face Mask and Social Distance Detection System

Field of Study: Computer Vision, Image Processing

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature　　　　　　　　Date: 6 January 2021

Subscribed and solemnly declared before,

Witness's Signature　　　　　　　　Date:

Name:

Designation:

FACE MASK AND SOCIAL DISTANCE DETECTION SYSTEM

ABSTRACT

The coronavirus (COVID-19) is causing the world to face a global health crisis. This is due to the rapid transmission of COVID-19. Some guidelines were issued by the World Health Organization (WHO) to avoid coronavirus from spreading. According to WHO, wearing a face mask and distance to each other by 1 meter is very effective to prevent COVID-19 from spreading. In this work, a transfer learning model for face mask detection and an algorithm to calculate the distance between two persons is presented. The proposed model is built by fine-tuning the pre-trained state-of-the-art deep learning model, MobileNetV2. The proposed model is trained and tested on the dataset that is collected by https://github.com/chandrikadeb7. Image augmentation technique is adopted to address the limited availability of data for better training and testing of the model. The model outperformed the other recently proposed approaches by achieving an accuracy of 93%.

Keywords: COVID-19, Face Mask, Social Distance

ACKNOWLEDGEMENTS

With the name of ALLAH (most gracious and merciful), I am grateful to Him in blessing me with the knowledge, giving me the courage to tackle the problems and always help me in each step of my life.

In successful completion of this work, I owe an enormous debt of gratefulness to my lecturer, Dr Woo Chaw Seng. I cannot find the right words to express the admiration and sincere gratitude towards my supervisors who gave me the opportunity to be a part of this work. They helped, encouraged, and motivated me in every step of research, during the time I have spent with them. Their suggestions were always valuable and their technical comments lead to the completion of this research project.

Lastly, I would also like to thank all my family members for their encouragement and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

COVID19	:	Coronavirus
SDLC	:	Software Development Live Cycle

CHAPTER 1: INTRODUCTION

The outbreak of COVID-19 has forced many countries to initiate new rules for face mask-wearing. In Malaysia, Governments have started working on new strategies to manage spaces, social distancing, and supplies for medical staff and normal citizens. Government is forcing everyone to always keep a 1-meter distance and wearing a face mask in a public area. A fine up to RM 1000 will be imposed on the people that failed to follow the rules. The goal of wearing face masks is to reduce the transmission and spreading rate. The World Health Organization (WHO) has recommended the usage of personal protective equipment (PPE) among people and in medical care. However, the ability of most of the countries to expand the production of PPE is very limited.

Today, COVID-19 is a significant public health and economy issue due to the harmful effects of the virus on people's quality of life, contributing to acute respiratory infections, mortality, and financial crises worldwide. According to WHO, more than six million cases were infected by COVID-19 in more than 180 countries with a death rate of 3%. The COVID-19 spreads easily in crowded environments and close contact. Governments are facing extraordinary challenges and risks to protect people from coronavirus in many countries. As people are forced by laws to wear face masks in public in many countries, masked face detection is a key to face applications, such as object detection. To fight and win in the battle against COVID-19 pandemic, Governments need guidance and surveillance on people in public areas, especially the crowded to ensure that wearing face masks laws are applied. This could be applied through the integration between surveillance systems and Artificial Intelligence models.

In this paper, we introduce a mask face detection model that is based on transfer learning and an algorithm to detect the distance between faces. The proposed model can be integrated with surveillance cameras to impede the COVID-19 transmission by allowing the detection of people who are not wearing face masks. The model use MobileNetV2 as the model is a lightweight model so that it can be integrated with any surveillance camera and even Raspberry Pi. We have used transfer learning for feature extractions and classification.

The main objective of this research is to detect and locate a medical face mask in an image and classify whether the people applying social distance or not as illustrated in Fig. 1. In this paper, the face mask and social distance detection are the main focus of research to reduce the spreading and transmission of Coronavirus specially COVID-19. Given an image, a region of the face mask and social distance indicator on the input image based on MobileNetV2 will be illustrated in the output image. The other objective is to measure the accuracy of the proposed model.

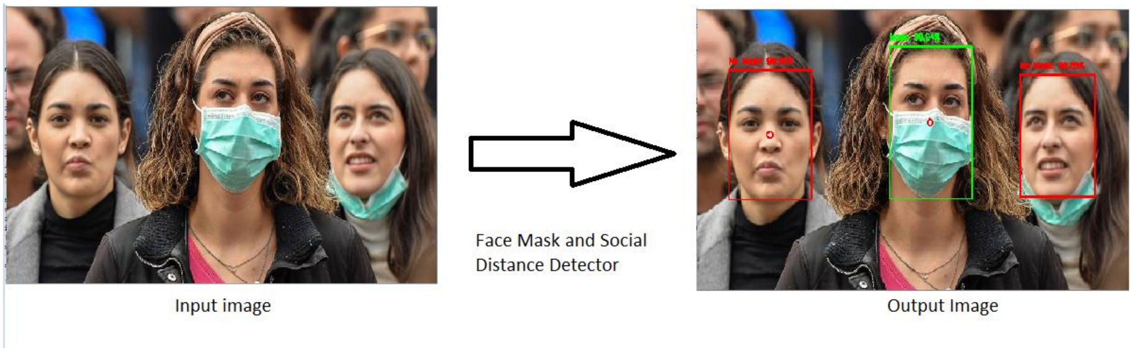

Figure 1. The outcome of the proposed model.

This paper conducted its experiments based on a dataset that has been collected by https://github.com/chandrikadeb7. She collects the data by using a web scraper. She uses Bing API image search to find all the images with a face mask. The dataset consists of 3868 pictures of faces. From the total, 1938 is the picture with a face mask and the rest are pictures without a face mask.

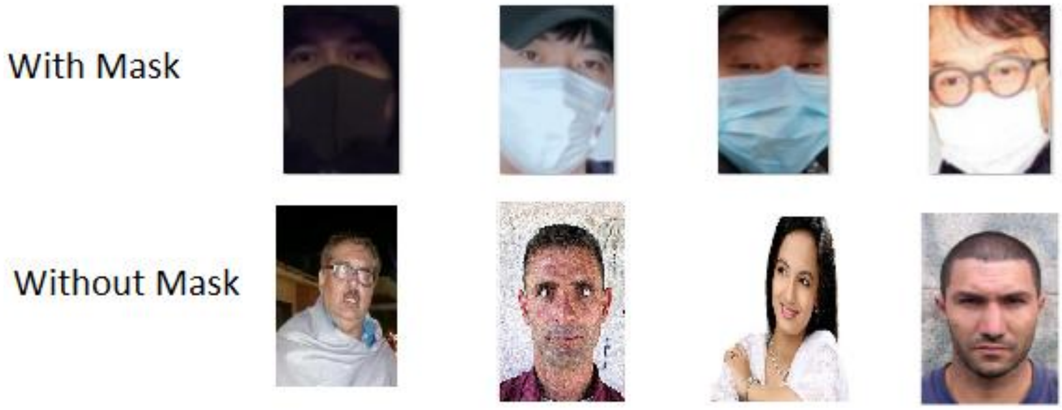

Figure 2. Dataset image samples.

CHAPTER 2: METHODOLOGY

2.1 Overview

This chapter describes the method performed for Face Mask and Social Distance Detection System. In completing this chapter, consideration for the SDLC model, analysis and design, and the technical implementation is proposed.

2.2 SDLC Model

To achieve my objectives, I am loosely following the Waterfall Model methodology for this work. The waterfall model is a classical model used in system development life cycle to create a system with a linear and sequential approach. It is termed as a waterfall because the model develops systematically from one phase to another in a downward fashion. I spent earlier of my time doing the requirement analysis. After finalizing the requirements, I planned for the system design to decide what model needs to be used that can be catered in every device. Lastly, I build the system and test the system.

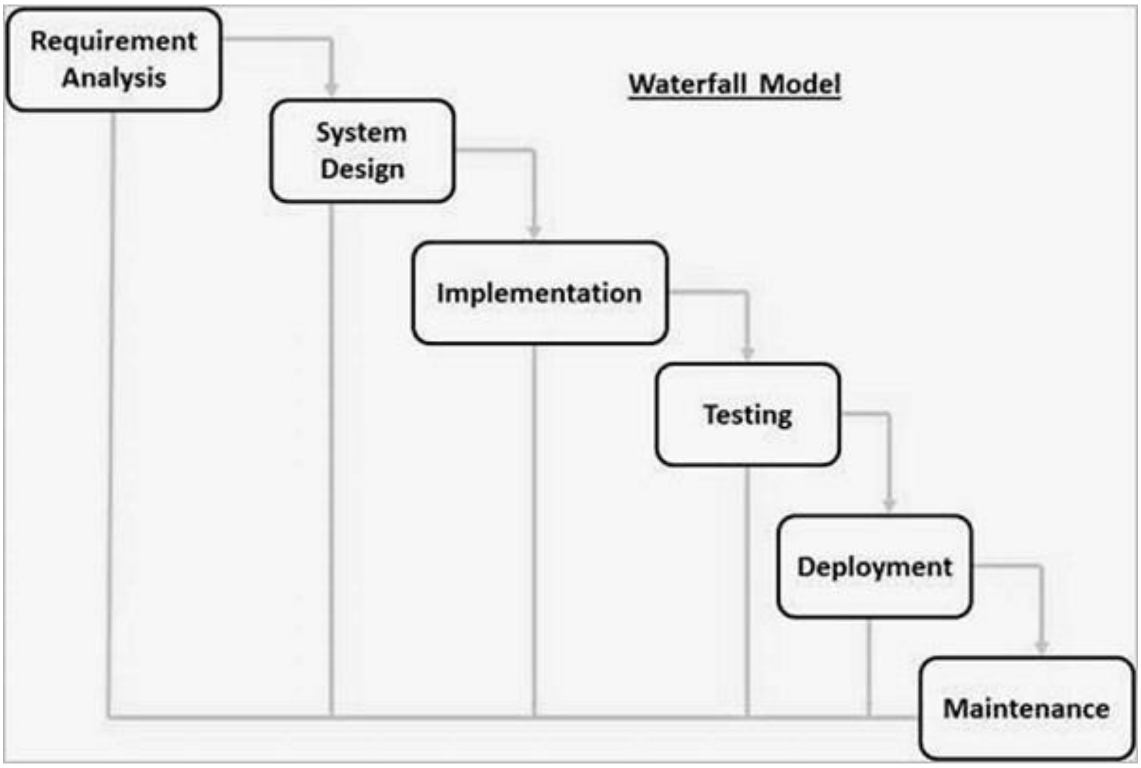

Figure 3. Waterfall Model.

2.3 Analysis and Design

Figure 4. presents the architecture diagram of the proposed model. The introduced model includes three main components: the first components is the image preprocessing, the second components are the data augmentation, the final main component is the detector. Figure 4 illustrates the proposed model. Mainly, the detector used MobileNetV2 for feature extraction and detection in training, validation, and testing phase.

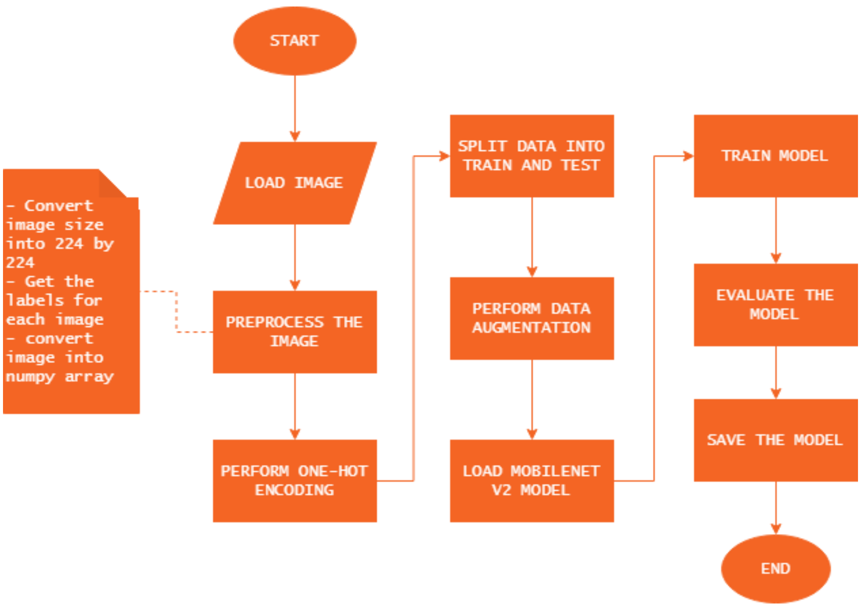

Figure 4. The proposed detector model.

2.3.1 Image Preprocessing

Preprocess the image is an important stage to train a model. The image needs to be in the same size before being fed into the training phase. After that, the labels for each image need to be extracted. Once the labels extracted, then the image and the labels will be converted into NumPy array.

2.3.2 Data Augmentation

Data augmentation is a method that can be used to artificially increase the diversity of datasets for training detectors. By transforming the original masked face images during training. Data augmentation improves the performance of the detector in training. Data were flipped horizontally to increase the unmasked face dataset as shown in Figure 5.

Figure 5. Sample of data augmentation.

2.3.3 MobileNetV2 Detector

The proposed model is MobileNetV2. MobileNetV2 object detection deep network is composed of feature extraction and detection network as shown in Figure 7. In the proposed model, MobileNetV2 used as a deep transfer model for feature extraction and classification. There are mainly two types of blocks one is bottleneck block with stride 1 and other with stride 2. There are 3 layers for each of the blocks as mentioned above. If stride=2 is used for depthwise convolution, the bottleneck block will not have a residual connection.

Input	Operator	Output
$h \times w \times k$	1x1 conv2d , ReLU6	$h \times w \times (tk)$
$h \times w \times tk$	3x3 dwise s=s, ReLU6	$\frac{h}{s} \times \frac{w}{s} \times (tk)$
$\frac{h}{s} \times \frac{w}{s} \times tk$	linear 1x1 conv2d	$\frac{h}{s} \times \frac{w}{s} \times k'$

Table 1: Bottleneck residual block transforming from k to k 0 channels, with stride s, and expansion factor t.

In most layers, the Expansion factor used is 6. If the input has 64 channels output has 64×6=384 channels.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Table 2: The expansion factor t is always applied to the input size as described in Table 1.

2.3.4 Social Distance Detector

The proposed model is using an algorithm from Scipy to calculate the Euclidean distance. Firstly, after detecting the face with or with a face mask, a bounding box is drawn. From the bounding box, a midpoint is calculated by using the formula shown in Eq. (1). Once the midpoint is detected, the Euclidean distance is calculated from each midpoint. Then, a threshold needs to be set to classify whether the distance is safe or not.

$$midpoint = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right) \quad (1)$$

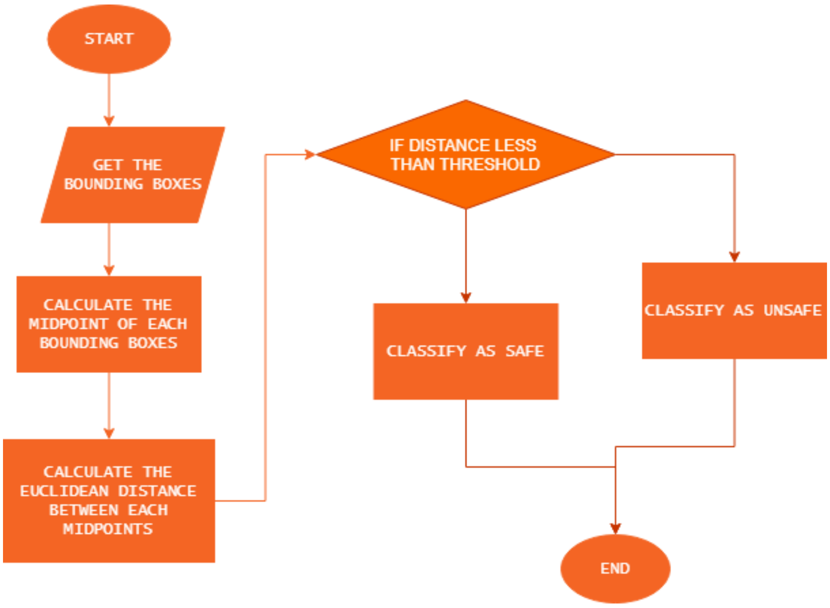

Figure 6. The proposed model for social distance detector.

CHAPTER 3: EXPERIMENTAL RESULTS

To evaluate the MobileNetV2 performance to detect the medical masked face, some performance metrics have been conducted throughout this research. The proposed model was implemented on the system having the following specifications: The GPU used NVIDIA RTX 2060 with the CUDA with Tensorflow and Deep Neural Network library (CuDNN) for GPU learning. The experiment configuration is presented in Table 3.

Model	Batch Size	Epoch	Learning Rate	Optimizer
Detector	32	20	0.0001	Adam

Table 3. Configuration of the proposed model.

Dataset split up to 75% training images and 25% testing images. The configuration of YOLO v2 with ResNet-50 with initial learning rate, 0.0001 and the number of epochs equal to 20 as illustrated in Table 3. The mini-batch size of the detector is set to 32. In terms of optimizer technique, Adam is chosen to be our optimizer technique to improve detector performance.

Table 4 illustrate the training process using Adam. To finish up to 20 epochs, the training duration is up to 1 hour and 40 minutes. From the table, the convergence happens around epoch 10.

Epoch	Time Elapsed	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
5	0:25:13	0.1687	0.9340	0.3074	0.8737
10	0:50:25	0.1273	0.9516	0.2681	0.8919
15	1:15:15	0.0972	0.9618	0.3278	0.8867
20	1:40:25	0.0920	0.9683	0.3923	0.8815

Table 4. The training and validation process based on Adam.

3.1 Validation, testing accuracy, and performance metrics for the proposed model.

To evaluate the performance of the proposed model, performance metrics need to be investigated through this work. The most common performance measures to be calculated are Accuracy, Precision, Recall, and F1 score. These metrics are represented from Eqs (2), (3), (4), (5).

$$Accuracy = \frac{TP + TN}{(TP + FP) + (TN + FN)} \tag{2}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{3}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{4}$$

$$F1score = 2 * \frac{Precision * Recall}{(Precision + Recall)} \tag{5}$$

TP is the count of True Positive samples, TN is the count of True Negative samples, FP is the count of False Positive samples, and FN is the count of False Negative from a confusion matrix.

For the dataset, the proposed model achieved a validation accuracy with performance metrics up to 88%. The precision for detecting image with face mask is 99%, while for the image without a face mask is 81%. Table 5 illustrates the overall performance metrics.

	Accuracy	Precision	Recall	F1 – Score
With mask	88%	99%	78%	87%
Without mask	88%	81%	99%	89%

Table 5. Overall Performance Metrics.

Based on Figure 7, we can see the True Positive and True Negative is on a higher percentage compared to False Positive and False Negative. This shows that the system has high accuracy in detecting the face mask.

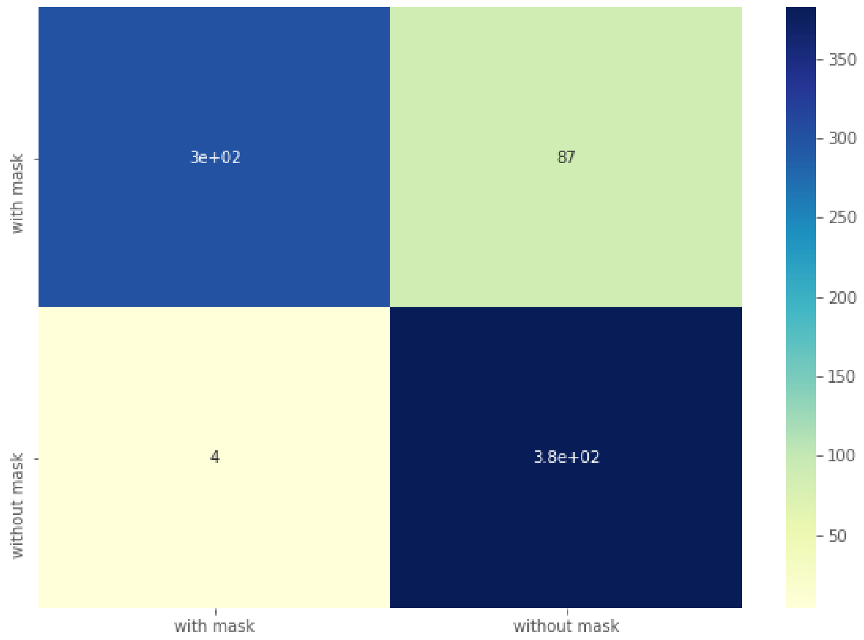

Figure 7. Confusion Matrix

3.2 Sample Input and Output

Here are some examples of input and output image for this work.

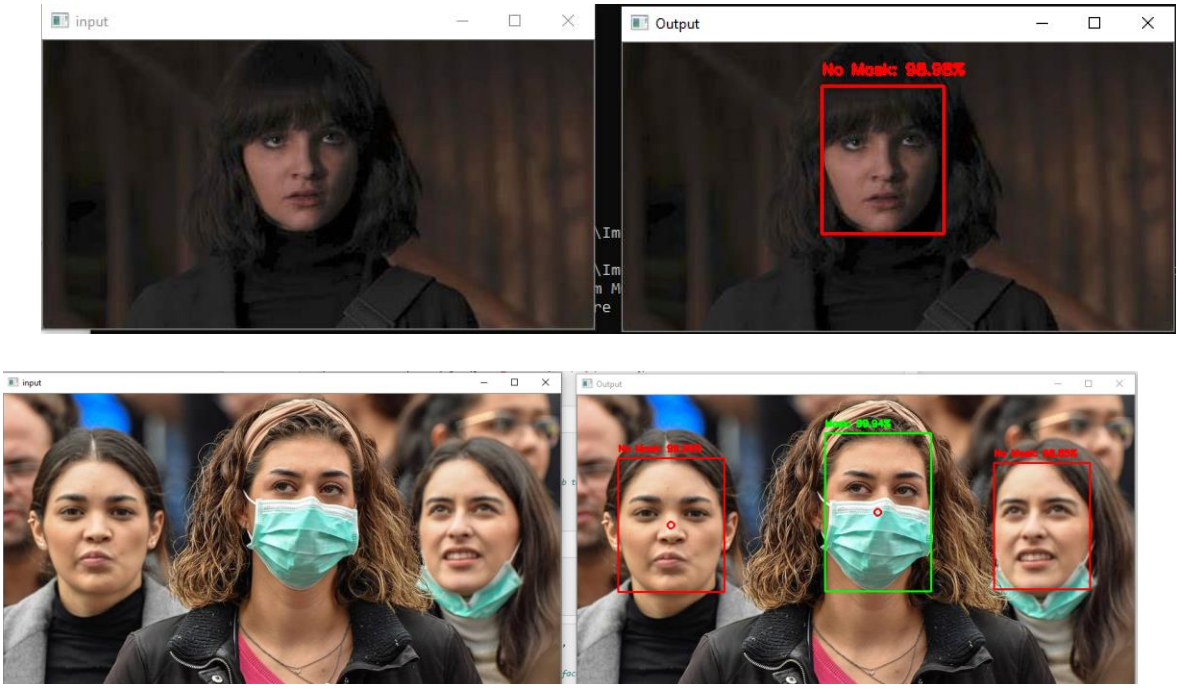

Figure 8. Samples of input and output image.

In the second example, we can see a red dot in the middle of the bounding box. This shows that the person in the middle is not in a safe distance with the person on the left.

CHAPTER 4: DISCUSSION

The strength of my work is I can detect the face mask with an accuracy of 88%. Other than that, I can detect the face mask with a different colour. Figure 9 illustrates the detection of a black face mask. My model can detect more than one faces in a single frame. Lastly, I can classify whether the distance is considered a safe social distance or not.

The limitation of my work, to detect the social distance, I need to set the threshold manually. So, in different camera angle or camera distance, the threshold to consider it as not safe distance, I must set it manually.

In the future, I am planning to make a model to learn to choose the appropriate threshold for social distance given any camera distance or angle. I want to make the model more robust.

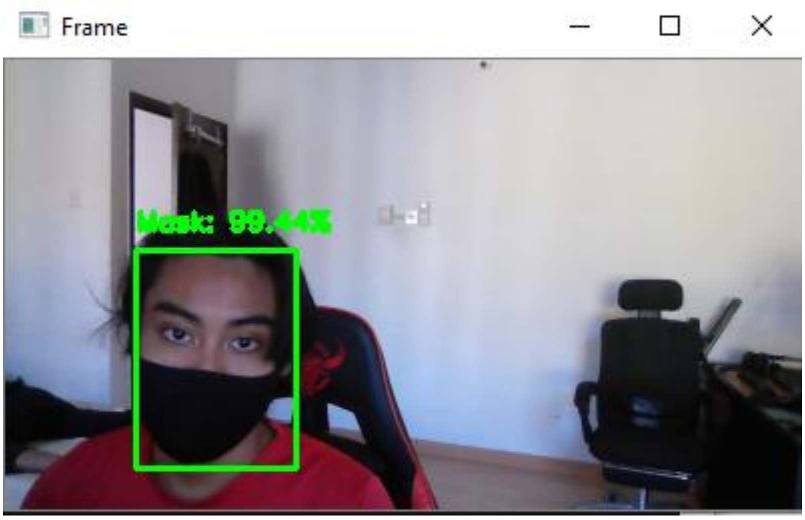

Figure 9. Detection of a black face mask.

CHAPTER 5: CONCLUSION

In this work, I have applied some of the image processing methods that I have learned during the class. One of them is Image Transformation. This process is used in the data augmentation. To increase the number the data, the image is being flipped horizontally and turned. The transformation used is geometrical transformation. Figure 10 illustrates the image has been going through image transformation.

Figure 10. Example of image transformation.

Other than that, image blurring is also being used in data augmentation. The image is being blurred to make the model more general and able to detect the image in any situation. A Gaussian kernel is used in this process.

Lastly, Neural Network. In this work, the neural network is used to train the model to detect the face mask based on the datasets. The type of learning used here is transfer learning and supervised learning. It is supervised learning because it learns based on a labelled dataset. It is a transfer learning because it learns on top of MobileNet V2 Model. The learning process is by updating the weights repeatedly until the stopping condition is met.

REFERENCES

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). 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.

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.

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society, 102600.

Pai, A. (2020, May 18). Build your Social Distancing Detection Tool using Deep Learning. Retrieved January 11, 2021, from https://www.analyticsvidhya.com/blog/2020/05/social-distancing-detection-tool-deep-learning/

Deb, C. (n.d.). Chandrikadeb7/Face-Mask-Detection. Retrieved January 11, 2021, from https://github.com/chandrikadeb7/Face-Mask-Detection

APPENDIX

- https://github.com/alfdnl/face-mask-and-social-distance-detector