Development of Face Mask Detection using SSDLite MobilenetV3 Small on Raspberry Pi 4

Nenny Anggraini, Syarif Hilmi Ramadhani, Luh Kesuma Wardhani, Nashrul Hakiem, Imam Marzuki Shofi, M. Tabah Rosyadi
Department of Informatics
Faculty of Science and Technology
UIN Syarif Hidayatullah
Jakarta, Indonesia
{nenny.anggraini, luhkesuma, hakiem, tabah.rosyadi}@uinjkt.ac.id, {syarif.ramadhani, imam}@mhs.uinjkt.ac.id

2022 5th International Conference of Computer and Informatics Engineering (IC2IE) | 978-1-6654-5338-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/IC2IE56416.2022.9970078

***Abstract*— This study aimed to develop a mask detection tool with SSDLite MobilenetV3 Small based on Raspberry Pi 4. SSDLite MobilenetV3 Small is a single-stage object detection. The single-stage object detection method is faster than the two-stage detection method. However, it has the disadvantage as the level of accuracy is not as good as the two-stage detection method. In the experiments, we used some methods to compare with SSDLite MobilenetV3, such as: SSDLite MobilenetV3 Large, SSDLite MobilenetV2, SSD MobilenetV2, SSDLite Mobileedets, and SSDMNV2 models. The result is that SSDLite MobilenetV3 is more powerful than other systems for detecting face masks. While the model with the best detection is the SSDLite MobilenetV2 model, the system with the SSDLite MobilenetV3 Small model still detects the use of masks, with a score of 70% accuracy from model accuracy testing in deployment. The limitation is the system with SSDLite MobilenetV3 Small can't detect incorrect masks.**

Keywords—COVID-19, face mask, SSDMNV2, SSD, SSDlite, Mobilenetv2, Mobilenetv3, Mobiledets, object detection

I. INTRODUCTION

Since December 2019, a new type of coronavirus has been found in humans in Wuhan, China. This new type of coronavirus was later named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV2) and caused Coronavirus Disease-2019 (COVID-19). This disease can spread through tiny droplets from the nose or mouth when coughing, sneezing, or talking. The use of masks is mandatory for everyone, especially for people with respiratory infection symptoms (coughing or sneezing), suspected COVID-19 infection with mild symptoms, and treating people with symptoms such as fever and cough. Healthy people are advised to use cloth masks.

Meanwhile, health workers use surgical masks [1]. A previous study made a mask detection system using a camera to help implement the mandatory use of masks. This system will classify the use of masks with object classification and object detection methods.

There were similar studies about object classification and detection with various objects. Some techniques such as Xception [2], MobileNetV2 [3] were used for classification. In other researches, YOLOV3-Tiny [4], YOLOV2[5], VGG-16 [6] were used for object detection. In [7] the researchers used MobileNetV2 for classifying the use of mask in the human face. Similar research [8], used SSD MobileNetV2 for classifying the use of a mask. The result of [7] and [8] can detect people when using face mask and not using face mask, but the number of FPS that can get in implementation with Raspberry Pi is not provided. Research [3], [7], and [6] were also implemented using Raspberry Pi for the purpose of making face mask detection with low computing unit device. The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse [9].

In the SSDMNV2 algorithm, there are two stages to detect the use of masks. According to [10], object detection methods are divided into 2 types: two stages and one stage. Two-stage object detection has a high level of accuracy, but the detection speed is slower than single-stage object detection. In contrast, single-stage object detection has a lower accuracy rate but has a faster detection speed compared to the two-stage object detection method. Based on the literature study results, there is a single-stage SSDLite MobilenetV3 Small detection model. According to [11], MobilenetV3 Small has a higher accuracy rate than MobilenetV2, with a latency of 20 – 40 ms when tested on Pixel 1. And in research [12], SSDLite was introduced, which is a higher version efficiency than SSD; SSDLite replaced all regular convolutions with separable convolutions. From the previous studies, we believed there is a chance to increase the detection speed of the mask detection using SSDLite MobilenetV3 Small on the Raspberry Pi.

In this study, SSDLite MobilenetV3 Small was compared with other models in its ability to detect the use of face mask based on Raspberry Pi. The contributions of our research are:

1. To develop a face mask detection device using SSDLite MobilenetV3 Small based on Raspberry Pi 4. It has capabilities of detecting the use of a cloth mask, the use of incorrect cloth mask, the use of a medical mask, the use of incorrect medical mask. It also can detect when someone is not using masks or covering his face with hands or a newspaper.
2. To show that SSDLite MobilenetV3 Small, a single-stage object detection, is more powerful than other methods in a face mask detection device based on Raspberry Pi.

II. METHOD

In this study, we used the prototyping method by Pressman and Maxim [13] in creating and developing the system. Using this method, we can find out what method is best used in building a face mask detection system. The following are the steps in the prototype method that have been carried out.

A. Communication

This stage was conducted to gain information about previous research and any other aspects to expand.

B. Quick Planning

At this stage we made a working system plan using several analyzes as below:

1. Analysis of the Classification of the Use of Masks: The model will be able to classify people who are using masks, not using masks, using masks in the wrong way.
2. Dataset Analysis: In the SSDMNV2 algorithm, the model for face detection was already available from OpenCV. We trained the model to classify the use of masks using the MobilenetV2 algorithm. A pre-training model was used from Keras to train the dataset. The pre-training model's purpose is to make the training process faster. The training for the classification model for the use of masks required a dataset in the form of images labelled with folders. Each folder has the label of the classification.

   We also used the SSD MobilenetV2, SSDLite MobilenetV2, SSDLite MobilenetV3 Large, SSDLite MobilenetV3 Small, and SSDLite Mobiledets algorithms. The training models for these algorithms were different from the previous one. It was used for the object detection task. Labels were not created with folders but were necessary for bounding box notation for each target object in the sample image. This notation will be labelled according to its classification.

   Therefore, the dataset used in MobilenetV2 training with a one-stage object detection algorithm differs. For one-stage detection object training, it requires more process than object classification training in preparing the dataset. We used a dataset [14] divided into training and test datasets. The training dataset has 7385 images consisting of 6702 annotations using masks (with mask), 9680 annotations without masks (without_mask) and 320 annotations using incorrect masks (incorrect_mask). Meanwhile, the test dataset has 1820 images consisting of 993 annotations using masks (with mask), 791 annotations without masks (without_mask) and 46 annotations using incorrect masks (incorrect_mask).

   In addition to the dataset, we also use dataset [8] to train the mask classification model with MobileNetV2, consisting of 5521 images using masks with the label with_mask and 5521 images without masks with the label without_mask. Furthermore, we added 5521 images using the incorrect mask from the dataset obtained from [15]. This dataset is named with MaskedFace-Net. There are 70,000 images in the dataset. In this dataset, from photos of human faces, masks are given automatically through the program. To adjust to the number of datasets used by [8]. We took 1841 images with the label Mask_Mouth_Chin where the mask only covers the mouth and chin. 1840 image labelled Mask_ Nose_Mouth where the mask covers only the nose and mouth. And 1840 images labelled Mask_Chin where the mask covers only the chin.
3. System Functional Requirements Analysis: the system requirement is to be able to classify the use of masks and provide output in the bounding box (Fig. 1)

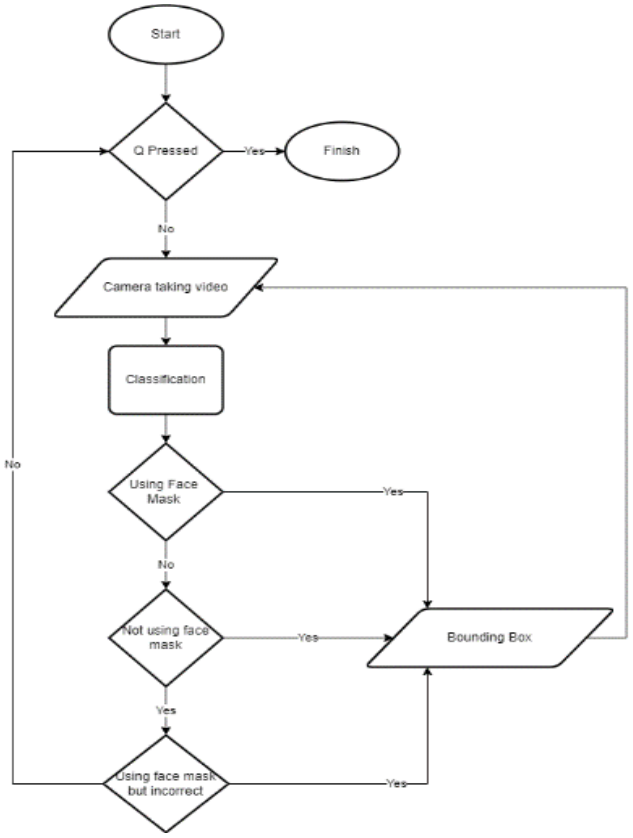

Fig. 1. Program flowchart

4. Hardware Requirements Analysis: The selection of hardware is based on the functional requirements of the system and the literature study that has been carried out. This system required: a Raspberry Pi 4 Model B 4Gb, a Raspberry Pi 4 Model B Cam V.1, a monitor, a push button non-momentary switch, a fan, a diode 1N4001, 3 resistor 470 Ohm, and a transistor 2n2222. (Fig. 2).
5. Quick Modelling: In this rapid modelling stage, we created a block diagram of the mask classification system

   In one package, there is a camera, Raspberry Pi 4 Model B, led, power button, program button, fan, and supporting devices. This packaging is the core of the system. While in a separate section, speakers and a monitor are output devices, and the power source is taken from AC electricity using an adapter. Both the adapter for the monitor and the adapter for the Raspberry Pi 4 Model B.

C. Construction

1. The next stage is construction; at this stage, we divided it into 2, namely making a detection model where the system will use this model to detect the use of masks. Next is the hardware assembly of the system.
   a. Detection Model: In making the detection model, this research is divided into 2, namely training for the one-stage detection model and training for the two-stage detection model. Fig. 3 shows the process of making a one-stage detection model.

      After the training is complete, the model that has been trained will be converted into a frozen graph. After becoming a frozen graph, the conversion

process is repeated to .tflite where this .tflite file will later be used on the Raspberry Pi.

Fig. 4 describes the stages of making a two-stage detection model. In making a two-stage detection model, this study only made a section to classify the use of masks. So in the two-stage detection model, the first stage is recognizing human faces. After the human face is identified, the area of the face will be taken, which will then be classified in that area. This two-stage process occurs on the Raspberry Pi. In this study, we use SSD as a face detection model that already has a model.

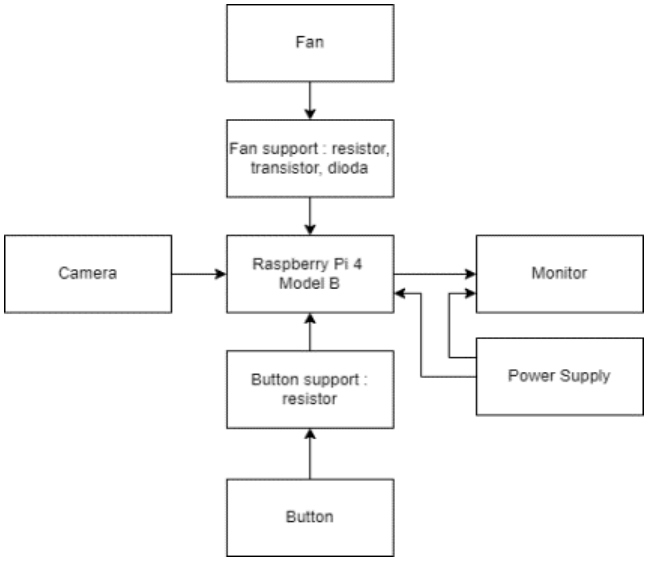

Fig. 2. Block diagram

b. Hardware: The hardware was arranged based on the block diagram in the modelling. At the top, there is a fan mounted using a bolt. The fan was useful for cooling the Raspberry Pi when the temperature gets hot. On the front, a camera is mounted using a bolt; the camera is facing forward to capture objects well. At the bottom, there are no components; only 4 screws are seen that are used to lock the Raspberry Pi to the case. On the back, there is the power button, power port, micro-HDMI port, and audio jack port. On the right is a USB port used as a connector to the keyboard and mouse and an Ethernet port so that you can connect to the network via a LAN cable. On the left, there are no components.

D. Deployment Delivery and Feedback

At this stage, we did several evaluations, and the results are shown in Table I. These evaluations were conducted until the desired output was obtained. After several attempts, the classification still doesn't work, so the results are gained. There are 3 evaluations: detection evaluations; comprise input using a medical mask, incorrect use of medical mask, not wearing mask, using a cloth mask, incorrect use of cloth mask, cover by hand or a newspaper. Using a medical mask and wearing a cloth mask should detect as correct, which means wearing a mask. Improper use of a medical mask or cloth mask was detected as incorrect, which means wearing the mask incorrectly. Not using mask, covering by hand or newspapers were detected as not using mask. The second is FPS evaluations, and the third is power consumption evaluations.

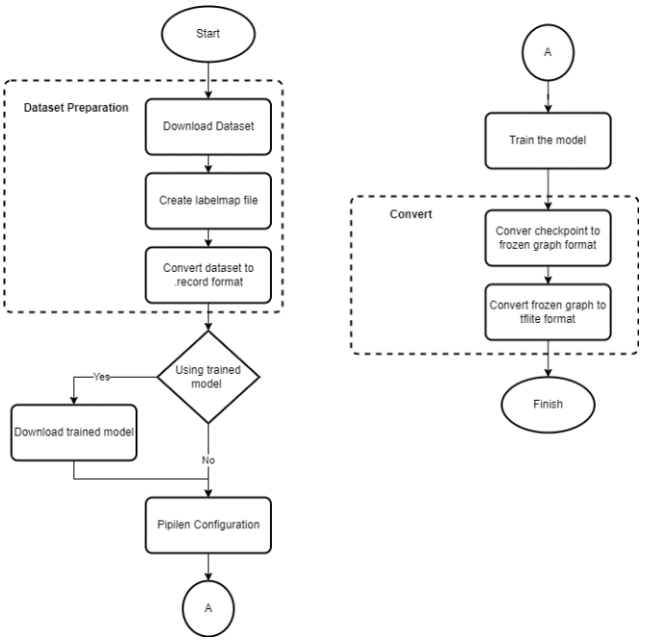

Fig. 3. Onestage model train phase

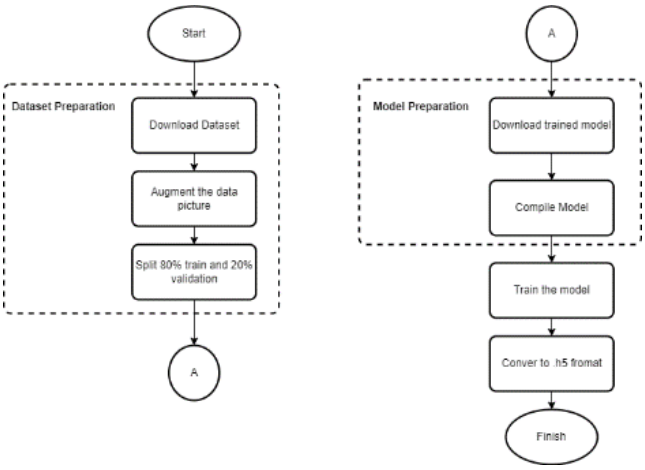

Fig. 4. Classification model train phase

E. Deployment Delivery and Feedback

At this stage, we did several evaluations, and the results are shown in Table I. These evaluations were conducted until the desired output was obtained. After several attempts, the classification still doesn't work, so the results are gained. There are 3 evaluations: detection evaluations; comprise input using a medical mask, incorrect use of medical mask, not wearing mask, using a cloth mask, incorrect use of cloth mask, cover by hand or a newspaper. Using a medical mask and wearing a cloth mask should detect as correct, which means wearing a mask. Improper use of a medical mask or cloth mask was detected as incorrect, which means wearing the mask incorrectly. Not using mask, covering by hand or newspapers were detected as not using mask. The second is FPS evaluations, and the third is power consumption evaluations.

TABLE I. EVALUATION RESULT

	Fine tuning	Max FPS	Min FPS	Power consumption	Result
SSDLite MobilenetV3 Small	Y	9.79	8.67	7.4 – 8.0 Watt	Model has incorrect detection improper use of medical mask and cloth mask. Model can detect other detection evaluation input correctly.
SSDLite MobilenetV3 Small	N	10	8.83	7.4 – 8.1 Watt	Have wrong detection in incorrect use of medical mask and cloth mask and also have struggle to detect cover by hand and cover by a newspaper correctly. But can detect other detection evaluation input correctly.
SSDLite MobilenetV3 Large	Y	4.26	3.81	7.3 – 8.0 Watt	Have wrong detection in incorrect use of medical mask and cloth mask and have struggle to detect cover by a newspaper correctly. But can detect other evaluation correctly.
SSDLite MobilenetV3 Large	N	4.21	4.01	7.3 – 8.0 Watt	Have wrong detection in incorrect use of medical mask. But can detect other evaluation correctly.
SSDLite MobilenetV2	Y	3.57	3.43	7.2 – 7.9 Watt	Can detect all input correctly.
SSDLite MobilenetV2	N	3.49	3.33	7.3 – 8.0 Watt	Have wrong detection in incorrect use of medical mask and sometimes in incorrect use of cloth mask. But can detect other evaluation correctly.
SSD Mobilenet V2	Y	3.42	3.21	7.3 – 8.0 Watt	Have wrong detection in incorrect use of cloth mask. When have input incorrect use of medical mask, sometimes it's detect right and sometimes it's can't detect. But can detect other evaluation correctly.
SSD Mobilenet V2	N	3.42	3.27	7.7 – 8.1 Watt	Have wrong detection in incorrect use of medical mask. But can detect other evaluation correctly.
SSDLite Mobiledets	Y	3.81	3.59	7.3 – 8.1 Watt	Have wrong detection in incorrect use of medical mask and cover by a newspaper. But can detect other evaluation correctly.
SSDLite Mobiledets	N	3.83	3.40	7.3 – 8.0 Watt	Have wrong detection in incorrect use of medical mask and cloth mask. When have input cover by hand and cover by a newspaper the system can't detect. But can detect other evaluation correctly.
SSDMNV2	-	1.42	0.78	8.1 – 9.2	When have input incorrect use of medical mask have to bring face closer to the camera to get correct detection, have struggle when input using incorrect use of cloth mask, and have wrong detection when detect cover by newspaper. But can detect other detection evaluation input correctly.

III. RESULTS AND DISCUSSIONS

In this section, the results of research and evaluations that have been carried out previously will be presented which are divided into several parts.

A. Evaluation Results of Face Mask Detection

The detection evaluation shows that the SSDLite MobilenetV2 model with training using fine-tuning has a better detection ability than the other models tested in this study. However, in this study, this model has a better detection ability than other models because it can detect all inputs easily. Meanwhile, using the same model but trained without fine-tuning, the result was not good, the errors were detected when tested misusing masks. Errors when detecting masks with incorrect use are also encountered in other models besides the SSDLite MobilenetV2 model with fine-tuning. However, for the SSDLite Mobiledets model using fine-tuning, it was easy to detect the improper mask for cloth mask use. For SSDLite MobilenetV3 Large without fine-tuning also easily detected the use of the improper mask when using a medical mask, and the model with SSDMNV2 could detect the use of the incorrect mask, but the result is not too good. There is an inability on the SSDLite Mobiledets model without fine-tuning and the SSD MobilenetV2 model with fine-tuning to detect testing input when the face with a mask covered with a hand or a newspaper.

The model used as the research objective in this research is SSDLite MobilenetV3 Small. This study has two research configurations: training with fine tuning and training without fine-tuning. Both models were equally unsuccessful in detecting the use of the incorrect mask. We found that the model with training without fine-tuning had difficulty in detecting the mask while it was covered with a hand or a newspaper. Therefore, we will use SSDLite MobilenetV3 Small for advanced testing with fine-tuning.

Please note that each training with a different configuration can produce a different final model result. The results of this detection evaluation cannot make the benchmark model A better in detection ability than model B because each training with specific configurations can produce different final results.

B. FPS Evaluation

Based on the FPS evaluation results, SSDLite MobilenetV3 Small has the highest FPS compared to other models, and a two-stage mask detection system has the lowest FPS compared to the others. These results also show that a single-stage detection system is faster than a two-stage detection system.

C. Power Consumption Evaluation Results

The power required is relatively the same. The rise and fall of power were also influenced by the fan on the system, where the fan will automatically turn on when the CPU temperature reaches 60 degrees Celsius and will stop if it is below that.

D. Model Accuracy Testing

The detection model chosen is SSDLite MobilenetV3 Small Fine Tuning. In this accuracy test, we conducted 10 times experiments for each type of input. The types of information include using a mask, using a medical mask, using an incorrect medical mask, using a cloth mask, using an incorrect cloth mask, covering the face with hands, and covering face with a newspaper. Table II shows the results

of prediction for each right usage, wrong usage, and without mask.

TABLE II. RESULTS

	Right Prediction	Wrong Prediction	Without Mask Prediction
Actual Right	19	24	0
Actual Wrong	0	0	0
Actual Without Mask	0	6	30

To calculate the accuracy of the above test, we use the (1).

$$Accuracy = \frac{\sum TP + TN \times 100\ \%}{\sum TP + FP + FN + TN} \quad (1)$$

Based on Table II, the overall accuracy can be calculated, the overall accuracy is obtained by the accuracy equation and can be used as (2).

$$Accuracy = \frac{49\ \times 100\ \%}{70} = 70\ \% \quad (2)$$

Actual predictions with imprecise detection results were found while testing for false predictions. No false prediction result has an actual false either. Meanwhile, predictions without masks have accurate detection results for all tests. And a correct prediction has one actual result that isn't wrong. It happened when it tried to use the input with the face down to the camera and produced the actual result without the mask. Based on these results, the significant role is that this model test's accuracy value is 70% due to the model's inability to detect the use of the incorrect mask.

E. Object Distance Testing

In this test, we tried to find a point where the tool can still detect. After the point is obtained, the we marked it and measured the distance. In this test, we get the results for the detection of using a mask and without a mask, a maximum distance was 169 cm. For detection without a mask by covering the face with hands, the maximum distance is 146 cm. For detection without a mask by covering the face with a newspaper, the maximum distance is 133 cm. For the detection of masks with the incorrect use, as in the previous accuracy test, this was not obtained.

F. Number of Detection Testing

In the experiment, we tried to evaluate how many detections can be conducted in one frame. This experiment was conducted by directing the tool to a monitor screen. On the monitor screen, there was already a photo collage of a person consisting of 8 photos. In this experiment, the device success in detecting 7 photos simultaneously.

G. Lighting Testing

In this test, we evaluated the tool's ability to detect object with inadequate lighting conditions. The light conditions we used dim, dark, and dark but with additional lighting. The results obtained by the tool have difficulty detecting when lighting conditions are inadequate. When the light is dim, the device can only detect medical masks, which tend to be bright and conspicuous, without masks, and cloth masks with a dark blue color are not detected. Furthermore, the device cannot detect anything when it is dark without lighting. When it is dark but has additional lighting, the tool detected as wearing no mask. While we using a cloth mask with a dark blue color, it was detected as wearing no mask. From this test, the device requires adequate lighting to get maximum detection results.

IV. CONCLUSION

This study aimed to develop a mask detection tool with SSDLite MobilenetV3 Small based on Raspberry Pi 4. This research shows that SSDLite MobilenetV3 Small has a higher detection speed than other models tested in this study. The SSDLite MobilenetV3 Small model can classify covering the face with a newspaper without a mask, which in this study [10] was classified as using a mask. However, its detection capability is still inferior to the SSDLite MobilenetV2 model when detecting the use of the incorrect mask. The accuracy value obtained is 70%. Also, because SSDLite MobilenetV3 Small includes a single-stage detection model, the resulting power consumption is slightly less than programs with a two-stage detection model. However, the slightly lower power consumption can also occur because the programs run the single-stage and two-stage detection models are different. The tool has limitations in terms of detection distance, detection capability when light conditions are inadequate, and the number of detections that can be detected is a maximum of 7 objects.

For further research, it will be interesting to develop a face mask detection system which can be expanded in some aspects, namely, using different datasets, using other training model configurations, and using other type of cameras. The capabilities of the Raspberry Pi 4 model B can be further increased using an overclock. We also can use a mini-PC similar to Raspberry Pi 4 model B, namely the NVIDIA Jetson NANO. We can also use other obstacles in object detections, such as people wearing headscarves.

REFERENCES

[1] Ministry of Health, "QnA: Questions and Answers Regarding COVID-19," 2020. https://infeksiemerging.kemkes.go.id/uncategorized/qna-pertanyaan-dan-jawaban-terkait-covid-19 (accessed Feb. 23, 2021).

[2] Darmasita, "Detection of Mask Usage Using Xception Transfer Learning," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 5, pp. 279–288, 2020, doi: https://doi.org/10.24252/instek.v5i2.20132.

[3] M. M. Lambacing and F. Ferdiansyah, "Design of New Normal Covid-19 Detector Mask With Telegram Notification Based on Internet of Things," *Dinamik*, vol. 25, no. 2, pp. 77–84, 2020, doi: 10.35315/dinamik.v25i2.8070.

[4] D. Giancini, E. Y. Puspaningrum, and Y. V. Via, "Identification of Mask Usage Using the CNN YOLOv3-Tiny Algorithm," *Semin. Nas. Inform. Bela Negara*, vol. 1, pp. 153–159, 2020.

[5] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustain. Cities Soc.*, vol. 65, no. June 2020, p. 102600, 2021, doi: 10.1016/j.scs.2020.102600.

[6] S. V. Militante and N. V. Dionisio, "Real-Time Facemask Recognition with Alarm System using Deep Learning," *2020 11th IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2020 - Proc.*, no. August, pp. 106–110, 2020, doi: 10.1109/ICSGRC49013.2020.9232610.

[7] M. Abdul, R. Irham, and D. A. Prasetya, "Mask Detection Prototype in Mask Required Rooms for Automatic Door Control Based on Deep Learning as the Prevention of COVID-19 Transmission," *Simp. Nas. RAPI XIX Tahun 2020 FT UMS*, pp. 47–55, 2020, [Online]. Available: http://hdl.handle.net/11617/12377

[8] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J.

Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," *Sustain. Cities Soc.*, vol. 66, no. December 2020, p. 102692, 2021, doi: 10.1016/j.scs.2020.102692.

[9] Raspberry Pi, "What is a Raspberry Pi?," 2022. https://www.raspberrypi.org/help/what- is-a-raspberry-pi/ (accessed Sep. 11, 2022).

[10] M. Elgendy, *Deep Learning for Vision Systems*. Manning Publications Co, 2020.

[11] A. Howard *et al.*, "Searching for mobileNetV3," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 1314–1324, 2019, doi: 10.1109/ICCV.2019.00140.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNet v2," *Comput. Vis. Pattern Recognit.*, vol. 4, 2019, doi: 10.1007/978-1-4842-6168-2_9.

[13] R. S. Pressman and B. R. Maxim, *Software Engineering : A Practitioner's Approach, Eighth Edition*. 2015. doi: 10.1145/1226816.1226822.

[14] X. Jiang, T. Gao, Z. Zhu, and Y. Zhao, "Real-time face mask detection method based on yolov3," *Electron.*, vol. 10, no. 7, pp. 1–17, 2021, doi: 10.3390/electronics10070837.

[15] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, "MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19," *Smart Heal.*, vol. 19, no. November 2020, p. 100144, 2021, doi: 10.1016/j.smhl.2020.100144.