2021 International Electrical Engineering Congress (iEECON2021)
March 10-12, 2021, Pattaya, THAILAND

Deep Learning-based Face Mask Detection Using YoloV5

Jirarat Ieamsaard
Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University
Phitsanulok, Thailand
jirarati@nu.ac.th

Surapon Nathanael Charoensook
Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University
Phitsanulok, Thailand
suraponc@nu.ac.th

Suchart Yammen
Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University
Phitsanulok, Thailand
sucharty@nu.ac.th

2021 9th International Electrical Engineering Congress (iEECON) | 978-1-7281-9584-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/iEECON51072.2021.9440346

***Abstract*— Ongoing Corona virus disease 2019 (Covid-19) pandemic, face mask wearing in public could reduce number of COVID-19 infected by minimizing the release of respiratory droplet from infected people. This paper is to study an effective method for face mask detection using a deep learning model created by "Yolov5". Comparative model developed with a different number of epochs: 20, 50, 100, 300 and 500. The experimental results show that the deep learning model created with 300 epochs has the highest performance with an accuracy of 96.5%.**

Keywords—face mask detection, deep learning, Yolov5

I. Introduction

Covid-19 is caused by the 2019 novel coronavirus (2019-nCoV), a new virus belonging to the same family of viruses causing severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). The number of infected and deceased patients are rapidly increasing. According to the World Health Organization (WHO) [1], in December 2020, the number of total cases worldwide was more than 82.7 million people, and the total number of deaths was approximately 1.8 million people. The virus is transmitted via direct contact with respiratory droplets of an infected person. To limit the spreading of COVID-19 pandemic, face mask wearing is highly encouraged, and in some countries, required.

Research studies have shown that face mask wearing can decrease community transmission of COVID-19, peak hospitalization and even deaths. Steffen E. Eikenberry et. al [2] suggested that wearing a mask in public can potentially reduce burden of the pandemic. Cornelia Betscha et. al [3] conducted an experiment to study social and behavioral consequences of mask policies during the COVID-19 pandemic. The study concluded that mask wearing was positively correlated with other protective behaviors, and countries or communities preferred people to wear masks.

During the surging COVID-19 pandemic, governments of many countries require face mask wearing in public area, such as public transport. Computer vision is often utilized in security system to detect body temperature and mask wearing before entering a building. Face mask detection system through artificial intelligence has been shown to be an effective solution; Sammy V. Militante and Nanette V. Dionisio [4,5] proposed a face mask wearing detection and physical distancing measurement system using deep learning, and Sharma Vinay [6] proposed a convolutional neural network (CNN) to detect if an individual is wearing a face mask.

In this paper, the YoloV5 was applied to perform the CNN to recognize whether a person wears a face mask properly, wears a face mask incorrectly, or does not wear a face mask. The experimental results for face mask detection obtained from the deep learning models with different epochs, including 20, 50, 100, 300 and 500, were examined and discussed in Section IV, and concluded in Section V.

II. Face Mask Detection Using CNN

A. Face mask dataset

The public face mask dataset consisted of 853 images with three labels, including "With_Mask", "Without_Mask" and "Incorrect_Mask", as used in the previous studies [7-8]. The 853 images from the face mask dataset were divided into three groups: 682 images for model training, 85 images for result validation, and 86 images for model testing. Each image had various sizes, its label and bounding boxes in the PASCAL VOC format. Examples of images in the face mask dataset are shown in Fig. 1.

Fig. 1. Examples of face mask image dataset

B. Face mask detection framework

In this paper, the framework of the developed YoloV5 model, as presented in Fig. 2, was divided into two parts: the training model and the face mask detection model. In the training model, 682 images from the face mask dataset were used. Original images from the face mask dataset were inputs of the YoloV5 face mask detection model, which processed a prediction score of three classes: "With_Mask", "Without_Mask" and "Incorrect_Mask", and then provided the output images with their predicted classes and detection scores.

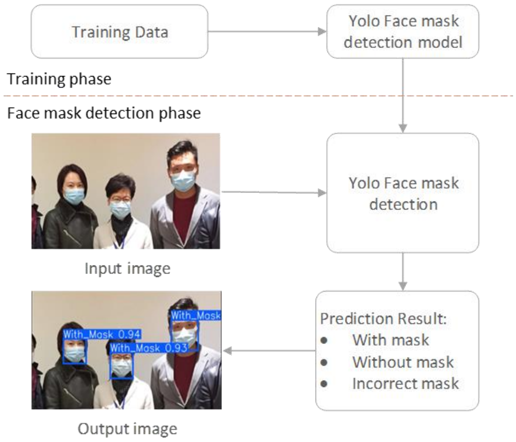


Fig. 2. A diagram of face mask detection framework

C. Face mask deep learning model using YoloV5

YoloV5 is a one-stage detector and a region-based object detection network. The Yolo redefines an object detection as a regression problem leading to a high processing speed. Recently, YoloV5 has been applied in real-time person search [9], and also in vision system for apple-harvesting robot.

YoloV5 has three main components: the backbone, the head and the detection. The backbone is a CNN that collects and shapes image features at different granularities. The YoloV5 implements the center and scale prediction (CSP) Bottleneck to formulate image features. The head is a series of layers to combine image features for throwing them forward to a prediction process. The YoloV5 also implements the PA-NET for feature aggregation. The detection is a process that utilize features from the head and takes box and class prediction steps. A diagram of the YoloV5 architecture is shown in Fig. 3.

To develop the face mask detection model using YoloV5, the training dataset consisting of 685 images with labels in the VOC format were implemented. In the training model process [11], an epoch, a number of passes of the entire training dataset, has been shown to affect the performance of the model. The goal of this paper was to determine the optimal epoch for the developed training model via the YoloV5.

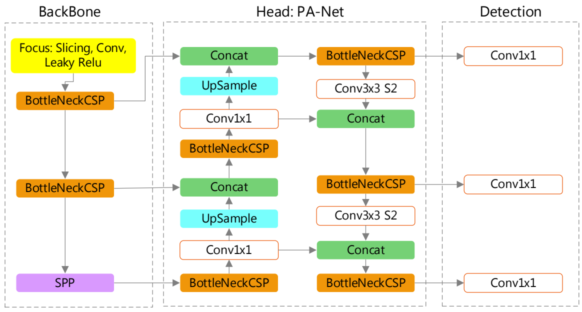
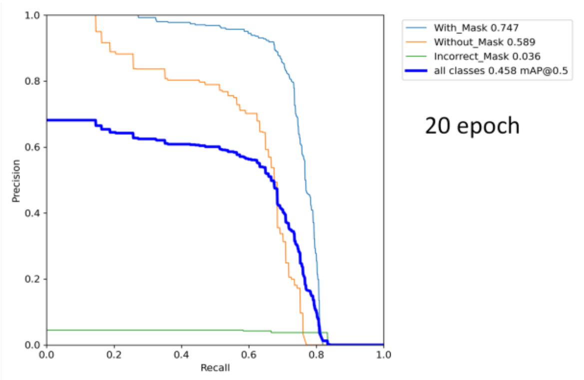


Fig. 3. Overview of YoloV5 architecture

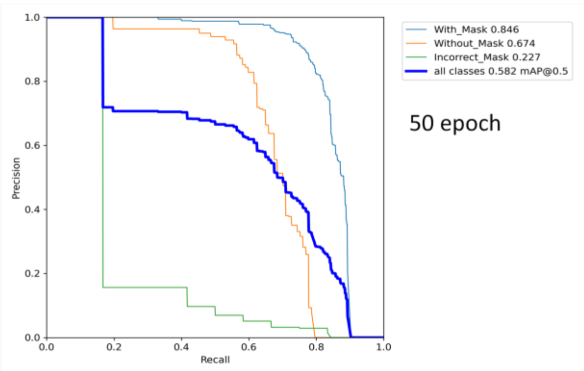
III. EXPERIMENTAL RESULTS

A. Model Training

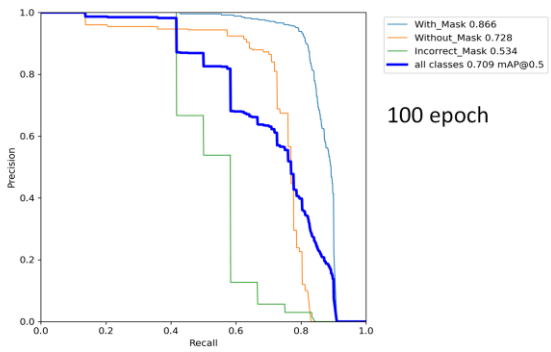
To find the optimal epoch for the developed training model using the YoloV5, the training face mask dataset of 682 images divided into three classes, including "With_Mask", "Without_Mask" and "Incorrect_Mask" are run with five different epochs: 20, 50, 100, 300, and 500. The precision and recall were also computed. Each model was validated with 85 face mask images. The validation process provided the precision and recall of each individual class. The plots of the precision and recall of each model are shown in Fig. 4. Specifically, Fig. 4 (d) indicates that the training model with 300 epoch provided the highest performance among all of the training models.



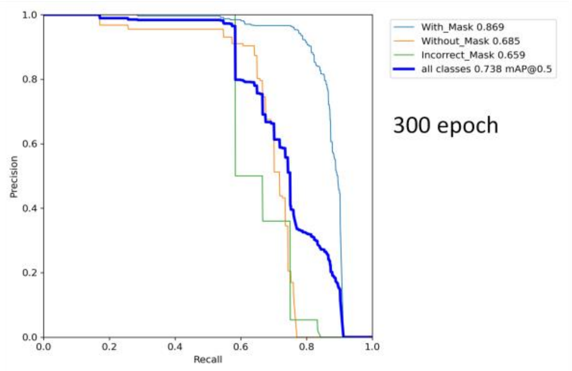
(a)



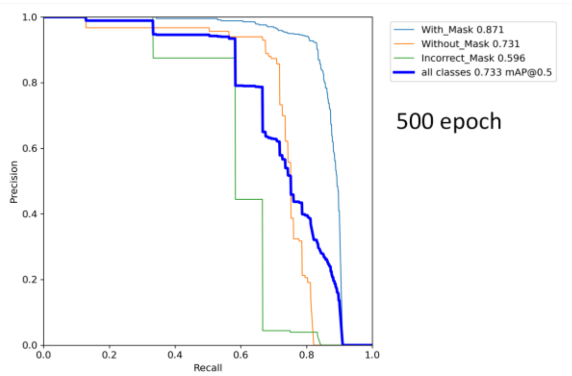
(b)



(c)



(d)



(e)

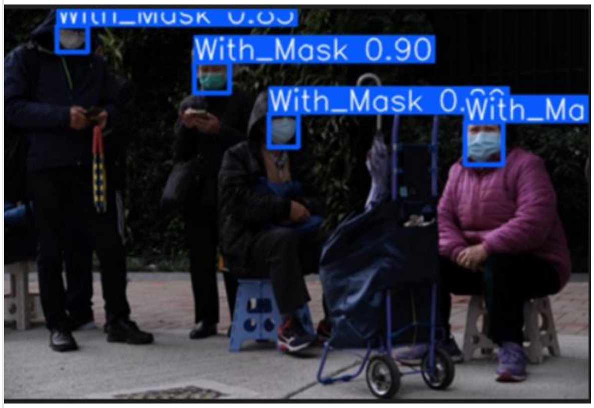
Fig. 4. Precision and recall of training model at different epoch; (a) 20 epoch, (b) 50 epoch, (c) 100 epoch, (d) 300 epoch, and (e) 500 epoch

By experiment represented in Fig. 4, it is seen that the training model for each class at 500 epochs gives more preferable results than that in case of 20 epochs, 50 epochs and 100 epochs. Furthermore, comparing to the training model for each class at five different epochs, the training model with 300 epochs provides the highest mean average precision (mAP) for all class; that is, the training model with 300 epochs is the best performance.

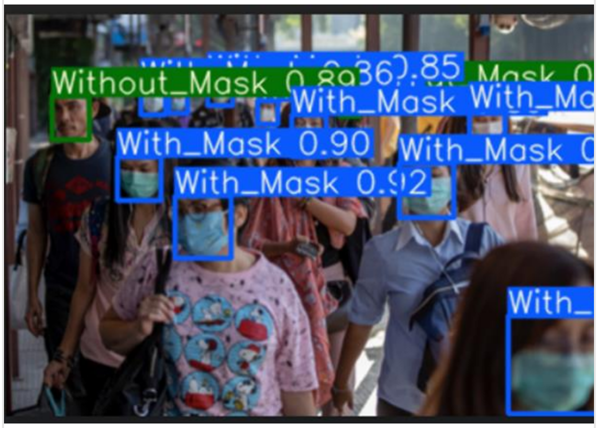
B. Face mask detection results

The developed deep learning models are also tested with the face mask dataset of 86 images. The face mask detection results can be shown in Fig. 5 and given in Table 1.

(a)



(b)



(c)

Fig. 5. Face mask image detection results with three cases: (a) incorrect mask in one-person image, (b) with mask in people image, and (c) with/without mask in crowd image.

TABLE I. ACCURACY OF FACE MASK DETECTION RESULTS

Cases	*Number of*	
	Correct Detections	*Incorrect Detections*
20 epoch	79	7
50 epoch	81	5
100 epoch	81	5
300 epoch	83	3
500 epoch	82	4

To illustrate the face mask detection performance of the developed deep learning model with YoloV5 to with three image cases: 1) incorrect mask in a one-person image, 2) with mask in a people image, and 3) with/without mask in a crowd image as shown in Fig 5, the developed model with the three cases can detect face mask well. Moreover, when testing with eighty-six images, the developed model can provide eighty-three correct detections and only three incorrect detection in case of 300 epochs as shown in Table I; that is, the accuracy of the model is very high up to 96.5% corresponding to the highest precision and recall obtained from the training model with 300 epochs in each class as shown in Fig. 4 (d).

C. Result discussion

Higher number of epochs increases the number of processing steps and therefore, tends to improve the performance in mask detection. However, the experimental results showed that the model trained with 300 epochs yielded the highest performance, even better than the model trained with 500 epochs, with only three incorrect detections as presented in Table 1.

(a)

(b)

(c)

Fig. 6. Incorrect detection resulted from using model trained with 300 epochs (a) "Incorrect_Mask" detected as "Without_Mask", (b) people using white cloth instead of white mask, and (c) "Incorrect_Mask" detected as "Without_Mask" in a partial facial/upper body image

Fig. 6 (a) and (b) demonstrate incorrect detection caused by distorted shape of objects; in Fig 6. (a), one face mask does not look like a face mask, and in Fig 6. (b), white cloth was used and detected as a face mask. Fig 6. (c) on the top right corner shows a woman wearing a face mask incorrectly but detected as "Without_Mask". The false detection was a result of an incomplete facial/upper body image.

IV. CONCLUSION

In this paper, the deep learning model for face mask detection is studies and developed. The model is trained by the YoloV5 at five different number of epochs. When comparing to the 86 tested images, the deep learning model for face mask detection with 300 epochs has the best performance with an accuracy of 96.5% related to the highest precision and recall. In the future work, the incorrect / correct wearing mask could be detected in real time by using a computer vision system.

ACKNOWLEDGMENT

The authors wish to acknowledge the Kaggle and MakeM for providing face mask dataset used in the study experiment.

REFERENCES

[1] WHO, (December 31, 2020). WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/

[2] Eikenberry, Steffen E., Marina Mancuso, Enahoro Iboi, Tin Phan, Keenan Eikenberry, Yang Kuang, Eric Kostelich, and Abba B. Gumel. "To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic." *Infectious Disease Modelling* (2020).

[3] Betsch, Cornelia, Lars Korn, Philipp Sprengholz, Lisa Felgendreff, Sarah Eitze, Philipp Schmid, and Robert Böhm. "Social and behavioral consequences of mask policies during the COVID-19 pandemic." *Proceedings of the National Academy of Sciences* 117, no. 36 (2020): 21851-21853.

[4] Militante, Sammy V., and Nanette V. Dionisio. "Real-Time Facemask Recognition with Alarm System using Deep Learning." In *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, pp. 106-110. IEEE, 2020.

[5] Militante, Sammy V., and Nanette V. Dionisio. "Deep Learning Implementation of Facemask and Physical Distancing Detection with Alarm Systems." In *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, pp. 1-5. IEEE, 2020.

[6] Sharma, Vinay. "Face Mask Detection using YOLOv5 for COVID-19." (2020).

[7] MakeML, 2020. Mask Dataset https://makeml.app/datasets/mask

[8] Kaggle Dataset. 2020. Face Mask Detection. https://www.kaggle.com/andrewmvd/face-mask-detection

[9] Li, Ye, Kangning Yin, Jie Liang, Chunyu Wang, and Guangqiang Yin. "A Multi-task Joint Framework for Real-time Person Search." *arXiv preprint arXiv:2012.06418* (2020).

[10] Kuznetsova, Anna, Tatiana Maleva, and Vladimir Soloviev. "Detecting Apples in Orchards Using YOLOv3 and YOLOv5 in General and Close-Up Images." In *International Symposium on Neural Networks*, pp. 233-243. Springer, Cham, 2020.

[11] Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Prashant Rai. (2020, October 29). ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements (Version v3.1). Zenodo. http://doi.org/10.5281/zenodo.4154370

[12] Kaggle. 2020. YOLOv5 Facemask. https://www.kaggle.com/rkuo2000/yolov5-facemask