Automation and Machine Learning (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/autml.2022.030302
ISSN 2516-5003 Vol. 3 Num. 3

# *Application of Deep Learning Face-mask Detection Based on YOLOv5*

**Zixuan Xia[1,*], Ming Zhu[2]**

[1]*School of Software Engineering, Xi'an Jiao Tong University, Xi'an, 710049, China*
[2]*Department of Information Technology, Hunan Police Academy, Changsha, 410138, China*
[*]*Corresponding author: xxiazixuan824@outlook.com.*

***Keywords:*** Computer Vision, COVID-19, Deep Learning, Object Detection, YOLOv5

***Abstract:*** Since the outbreak of coronavirus in 2019, people around the world have started to wear masks to avoid the further spread of the virus. In order to better control the epidemic, it is very necessary to supervise the wearing of masks. In this paper, a deep learning model based on yolov5 is established for mask recognition and detection, and the model is trained and tested through datasets. Finally, mask detection is carried out for people in images and videos using local computer equipment, and the best weight of training and the training accuracy of nearly 95% is obtained. This paper combines the algorithms of binary classification, convolutional neural network, and deep learning object detection to effectively and accurately train and test the model, which has a certain reference value.

## 1. Introduction

With the novel coronavirus (COVID-19) outbreak in 2019, people from all over the world have been suggested by the World Health Organization (WHO) to wear their masks in public places. However, many people have not been used to this behavior due to various reasons [1] (E.g.: cultural, conscious, political). WHO [2] shows that over 80 million people from 72 countries have been diagnosed with this virus and this figure is still rising day by day. Scientists who have done some systematic reviews claimed that face mask use was effective in preventing COVID-19 [3] and governments in various countries are taking positive measures to require strict compliance with mask-wearing regulations in crowded areas.

Therefore, in order to be effective, algorithms based on deep learning and computer vision have been developed to monitor whether people are wearing masks more quickly and accurately. Since the significant boom of Convolutional Neural Network (CNN) brought by Alex Net in 2012, different versions of object detection algorithms have come out, leading a new wave of research in this field. There has been some recent research addressing mask detection. This paper proposed a new object detection method based on YOLO, which can be well qualified for this testing work after you only looked once. It was first introduced by Joseph Redmon and Ali Farhadi in 2015 and versions of this algorithm have been constantly updated and optimized in recent years.

A mask detection method based on YOLOv5 was proposed in this research. Compared to the previous Squeeze and Excitation YOLOv3 [4] (SE-YOLOv3), v5 has a smaller and faster network so its effect on online production is objective, which makes embedded equipment easier to use.

## 2. Research Methodology

### 2.1 Data Collection

As a common content in the field of computer vision and object detection, there are various open-source datasets that have been widely used in previous studies [5-6], such as MAFA, AIFOO, and so on. In order to validate the robustness and universality of this model, making the dataset of this study was considered a good approach. Examples of images in the face mask dataset are illustrated in the following Figure 1.

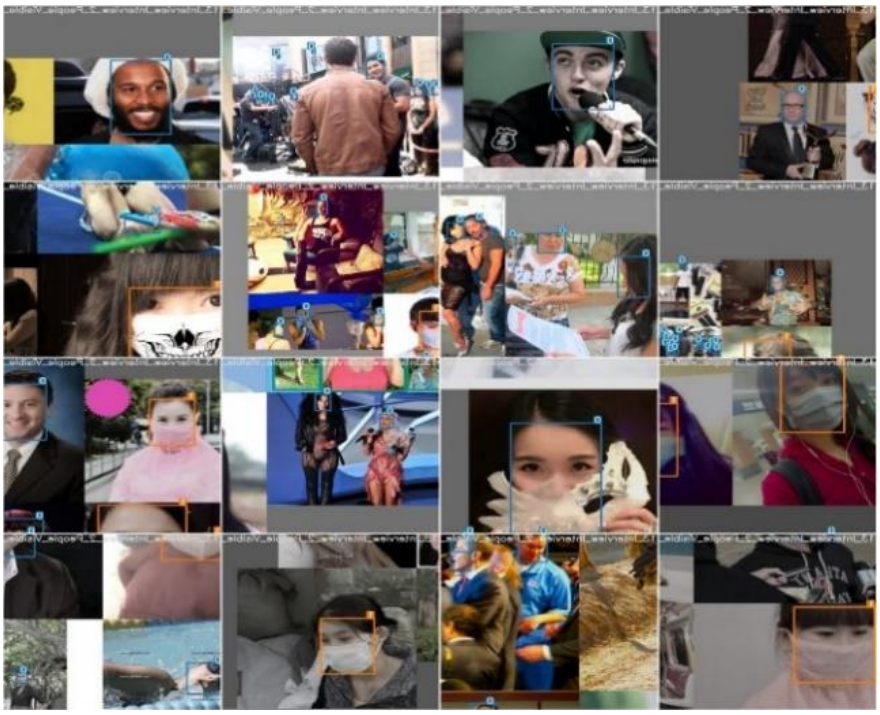


Figure 1: Dataset Examples

For mask detection, the research randomly collected 7958 images of people who were either wearing masks or not, or some of them had masks and others were not from Google. These images from the dataset were divided into two groups: 6307 images for model training, and 1651 images for result validation. For testing, this paper chose to use the computer to detect whether people (objects) were wearing masks in front of the camera in real time.

### 2.2 Preprocessing and Augmentation of Data

Before training, a label-making software named ‘labelimg’ was used to convert the picture into the XML form for which the Yolo tag applies. Since the images in the dataset are not all the same size, preprocessing is needed to resize all of the images to 256 ×256 pixels by Keras' Image Data Generator. After normalizing all images after converting them to 256 ×256, the next step is to label the dataset which processed a result of two classes: “Mask” and “No_mask”. Finally, the output images with their predicted classes and scores were provided to show whether the people in the image were wearing the mask or not. The whole procedure of the research is illustrated in Figure 2.

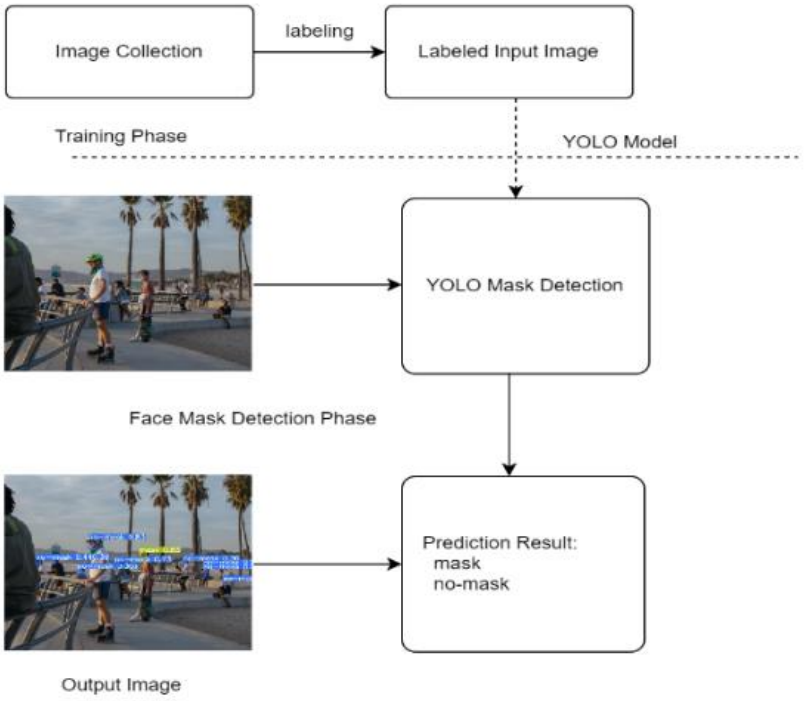



Figure 2: Face Detection Framework

## 2.3 Introduction and Architecture of the YoloV5 Model

As one of the top algorithms for current object detection, the Yolo series of algorithms is composed by R. Joseph et al. proposed in 2015. It was the first single-stage detector in the era of deep learning. YoloV5 adds some new improvements to the previous algorithm, making it a great performance improvement in speed and accuracy which recently has been applied in real-time person search [7], vision systems, and some other fields.

YoloV5 is made up of the following four main components: Input, Backbone, Neck, and Prediction. The Backbone of YoloV5 has its unique "Focus" structure which the previous versions do not contain. The core of this section is the slicing operation, which collects and reshapes the image features at different granularities by Convolutional Neural Network (CNN). In this research, input images were generally resized to 256 × 256 pixels by methods of Open CV. The full architecture of the Yolo V5 [8] is shown in Figure 3.

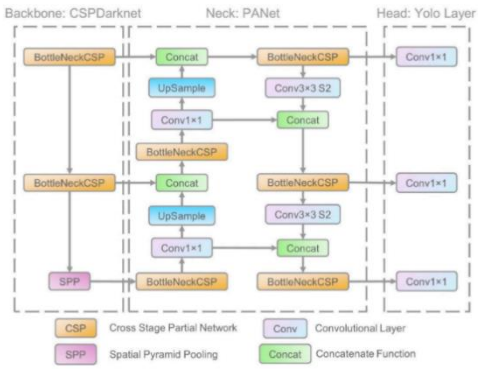



Figure 3: Architecture of YoloV5

## 3. Experimental Results

### 3.1 Model Training

Given the same model and parameters, different data sets, convergence rates, and metric values vary greatly, this study determined to set up an endless epoch value and run down the verification from time to time until finally get a satisfactory indicator. In this study, 7958 images which were divided into two major classes were trained this model with 8 NVidia RTX A40 48G boosted by NV-link.

The experiment illustrated that the training model with 300 epochs provides the highest mean average precision (mAP) for all classes, which can reach a value of 0.913 mAP (@0.5). The result is illustrated in the following Figure 4.

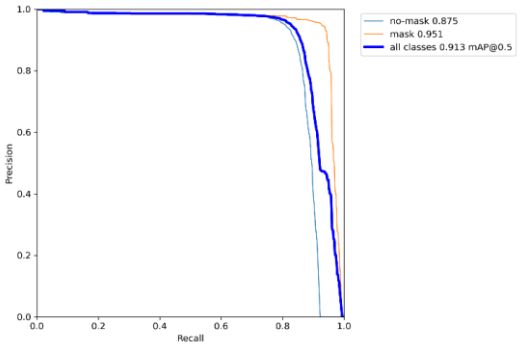



Figure 4: Precision and recall of training model at 300 epochs

### 3.2 Result Presentation

The presentation of the training result can be analyzed from several dimensions. Firstly, the performance of supervised learning algorithms can be visually evaluated by the confusion matrix [9]-[10] of this model. Then, given that the training is a two-classification issue, the F1 Score matrix can better balance this model accuracy with the recall rate. Finally, train batches and test batches can better show the training results.

The confusion matrix in Figure 5 evaluates the correctness of the classification result. The three dimmer squares in the matrix represent the computed number of correctly classified samples, the rest of which are either recognized in the wrong category or failed to be recognized as class examples. It is seen that the true positive of masked samples is 0.93 while the true negative one is 0.85.

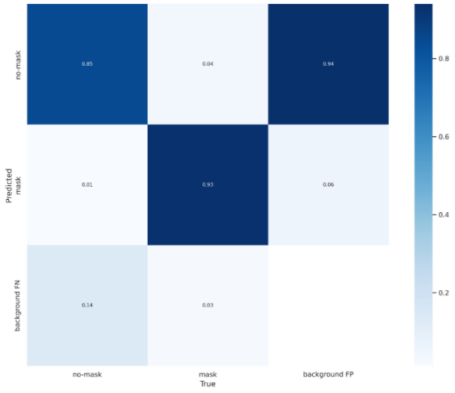
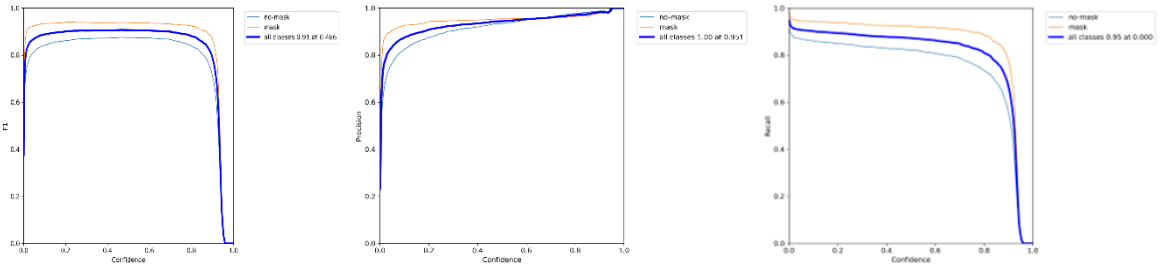



Figure 5: Confusion Matrix of the Model

The represented curve in Figure 6 demonstrates the relationship between the F1-score, the

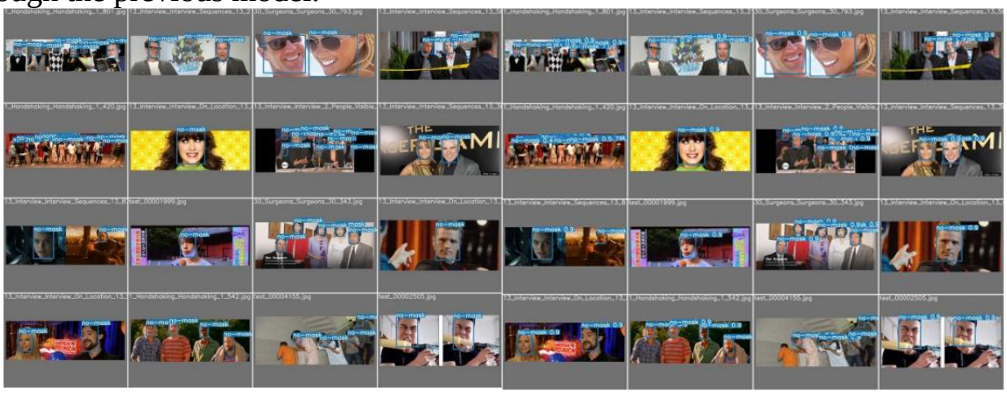
harmonic mean of precision and recall, and confidence level. It is seen that the overall F1 score of the model squeaks through 0.8, which is a comparatively preferable result, for all the no-mask class has a lower F1 score.



(a) F1 Curve of the Model (b) P-Curve of the Model (c) Q-curve of the Model

Figure 6: Relationships between F1, P and Q Curves

Finally, some of the test data and its training results will be displayed, each set of test data contains several pictures in the original data set, Figure 7 mainly shows the process of labeling the test samples, and then corresponding predictions were given about whether each person in the picture was wearing a mask through the previous model.



(a) Test Label of Sample (b) Prediction of the Sample Test

Figure 7: Comparison of Sample Test and Prediction

## 3.3 Correctness Validation

In order to better reflect the robustness and universality of this model, this research used two ways for validation.

The first way is to quickly verify that the person in the picture is wearing a mask by entering an arbitrary picture. The experiments demonstrated that in addition to real people, the model was also suitable for cartoons and comic characters, which was tested by using a comic that appears in the New York Times for training, the results are shown in Figure 8.


Figure 8: Training Result of the Model

Another way of testing is to use the camera of a local computer to find out whether the people in front of the screen are wearing masks or not. This kind of validation has a rather good practicability because, in the work following up, this model can be utilized in various areas, such as full-stack deep learning, deep learning system development, and so on. What's more, the model can not only recognize people but also has a certain degree of identification of the object of masks, which means when wearing some other things to cover your face, the model will still consider the situation as not wearing a mask.

## 4. Conclusion

This paper studied and developed the YoloV5 model to identify and test whether people are wearing masks or not. The YoloV5-based model has the best performance at 300 epochs, with an accuracy of 93.67%. However, there is still room for improvement. For instance, in some previous studies, there were some approaches to detect if the people in the image were wearing their masks properly, which would make the epidemic prevention and control management benefit more. Besides, compared with other systems, this model is not so convenient for ordinary people to use. The code is available on https://github.com/ZixuanSummerXia609/Research_SU22. Hence, in the next stage of the research, this algorithm is going to be combined with software development or web development so that the model can be widely used in different fields which can contribute to the early victory over the Covid -19.

## References


*[1] Lehmann, E.Y., Lehmann, L.S. Responding to Patients Who Refuse to Wear Masks during the Covid-19 Pandemic. J Gen Intern Med 36, 2814–2815 (2021).*

*[2] Who coronavirus disease (covid-19) dashboard, https://covid19.who.int/, online accessed Aug 25, 2022*

*[3] Jingjing Nie, Linna Kang, Yaya Pian & Jihong Hu. The need for more robust research on the effectiveness of masks in preventing COVID-19 transmission. Published Online:19 Apr 2022*

*[4] Jiang, X.; Gao, T.; Zhu, Z.; Zhao, Y. Real-Time Face Mask Detection Method Based on YOLOv3. Electronics 2021, 10, 837.*

*[5] J. Ieamsaard, S. N. Charoensook and S. Yammen, "Deep Learning-based Face Mask Detection Using YoloV5," 2021 9th International Electrical Engineering Congress (iEECON), 2021, pp. 428-431,*

*[6] F. M. J. Mehedi Shamrat, S. Chakraborty, M. M. Billah, M. A. Jubair, M. S. Islam and R. Ranjan, "Face Mask Detection using Convolutional Neural Network (CNN) to reduce the spread of Covid-19," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1231-1237*

*[7] 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)*

*[8] Xu Renjie & Lin Haifeng & Lu Kangjie & Cao Lin & Liu Yunfei. (2021). A Forest Fire Detection System Based on Ensemble Learning. Forests. 12. 217. 10.3390/f12020217.*

*[9] Haghighi S, Jasemi M, Hessabi S, et al. PyCM: Multiclass confusion matrix library in Python [J]. Journal of Open Source Software, 2018, 3(25): 729.*

*[10] M. Heydarian, T. E. Doyle and R. Samavi, "MLCM: Multi-Label Confusion Matrix," in IEEE Access, vol. 10, pp. 19083-19095, 2022*