Mask Detection and Epidemic Prevention System

Can Wang*1,a, LePing Bu2,b, HuanYu Xiao3,c, Kui Qin4,d
1Naval University of Engineering Wuhan, China
2Naval University of Engineering Wuhan, China
3Naval University of Engineering Wuhan, China
4Naval University of Engineering Wuhan, China
a* Corresponding author: 442776630@qq.com
bblpzxs_wuhan@126.com
c2398167715@qq.com
d100425497@qq.com

Abstract—The rapid spread of the coronavirus (COVID-19) has led the world into a huge pandemic crisis. As of November, more than 50 million people have been infected. In response to the transmission characteristics of COVID-19, the World Health Organization (WHO) has identified wearing masks as the most effective anti-epidemic measure in public places and crowded places. In order to ensure that people entering these areas wear masks, this paper designs a complete epidemic prevention system. Utilizes deep learning, embedded technology and LoRa wireless communication. It is characterized by low cost and flexible communication methods, which can well meet actual epidemic prevention needs.

1. Introduction
In less than 11 months, COVID-19 has infected more than 50 million patients worldwide and caused more than 1 million deaths, making it the most serious public health emergency in history. Most of the positive cases occurred in crowded and overcrowded areas, such as train stations, subway stations, and schools. Because COVID-19 is mainly spread through respiratory droplets and aerosols, the World Health Organization recommends that masks should be worn in densely populated places, and it has been recognized as the most effective means of epidemic prevention in public places[1]. However, people often do not wear masks when entering public places, and it is very difficult to monitor this behavior artificially.

In order to ensure that people entering public places wear masks, a special video surveillance system needs to be designed for this situation. Deep learning is the hottest research field in the 21st century, and its core research direction is computer vision. The current target detection technology has been widely used in life, such as face recognition, security monitoring, and drone scene analysis. The object detection method can effectively detect whether the outlet cover is worn.

The main work of this paper is to develop a mask detection and anti-epidemic system, which monitors the mask wearing status of people entering public places in real time through the camera and the EfficienDet-D0[2] object detection algorithm. If you are not wearing a mask, an alarm number will be sent to the lower computer through the LoRa wireless communication technology. At this time, the
access control will be closed and the buzzer will send a warning signal to remind the entering personnel to wear the mask. It has been verified that the system can effectively meet the actual prevention and control needs and has strong practical application significance.

2. Data and detection algorithm analysis

2.1. Data set analysis

In this article, the data sets used to train the deep learning model are Simulated Masked Face Dataset (SMFD) [3] and Wider Face [4]. The SMFD data set consists of 1570 images, with half of the face images wearing masks and those without masks. Wider Face is currently the world's largest face detection data set. We selected nearly 6000 photos from them and corrected their mislabeling. In the end, we randomly select 60% of the 7570 images as the training set, 10% as the validation set, and 30% as the test set. The annotated images in the data set are shown in Figure 1 and Figure 2.

![Figure 1](image1.png) An example of labeling in SFMD dataset

![Figure 2](image2.png) An example of annotation in the Wider Face dataset

2.2. Algorithm selection

Training a complex deep neural network is very expensive because it requires high computing power and memory, and it is very time-consuming. In order to save training costs, this article chooses to fine-tune the trained model. Even when training on a small data set, this method can improve the performance of the new model to a certain extent and reduce the training time.

Considering the performance indicators of Table 1 and the performance comparison of Table 2, we choose to fine-tune the pre-trained EfficienDet-D0 [5]. The model is pre-trained on the COCO data set. You only need to change the final output fully connected layer to two nodes, and keep the previous EfficiencyNet-B0 [6] weights unchanged during training, and only perform backpropagation to update the parameters of the classification layer.
EfficienDet uses EfficienNet as the backbone network, Bi-directional Feature Pyramid Network (BiFPN) as the feature network, and introduces a joint scaling method. BiFPN can quickly perform multi-scale feature fusion, in which weighting is applied. Joint scale scaling can uniformly scale the depth, width and resolution of the model to achieve the best results. The EfficiencyDet model parameters combined with these methods are reduced by 4 times compared with the most advanced model of the year, and the detection speed is increased by 3 times. Great savings in training costs.

Table 1 Main evaluation indicators of COCO data set

| Performance | meaning |
|-------------|---------|
| AP          | The value of AP when IOU=0.5 : 0.05 : 0.95 |
| AP50        | AP value when IOU=0.5 |
| AP75        | AP value when IOU=0.75 |
| APs         | AP value of small objects |
| APm         | AP value of medium object |
| APL         | AP value of large objects |

Table 2 Performance comparison of some target detection algorithms on the COCO data set

| model         | Backbone network | speed (fps^-1) | AP | AP50 | AP75 | ApS | APm | APL |
|---------------|------------------|----------------|----|------|------|-----|-----|-----|
| Fast R-CNN    | VGG-16           | 3.0            | 19.7| 35.9 | -    | -   | -   | -   |
| Faster R-CNN  | VGG-16           | 6.0            | 36.2| 59.1 | 39.0 | 18.2| 39.0| 48.2|
| R-FCN         | ResNet           | 9.0            | 29.9| 51.9 | -    | 10.8| 32.8| 45.0|
| Mask R-CNN    | ResNet-101       | 11.0           | 39.8| 62.3 | 43.4 | 22.1| 43.2| 51.2|
| SNIP          | DPN-98           | 2.5            | 45.7| 67.3 | 51.1 | 29.3| 48.8| 57.1|
| TridentNet    | ResNet-101       | 0.7            | 48.4| 69.7 | 35.5 | 31.8| 51.3| 60.3|
| D2Det         | ResNet-101       | -              | 50.1| 69.4 | 54.9 | 32.7| 52.7| 62.1|
| YOLO-v1       | DarkNet-53       | 40.0           | 33.0| 57.9 | 34.4 | 18.3| 35.4| 41.9|
| YOLO-v3       | DarkNet-19       | 20.6           | 21.6| 44.0 | 19.2 | 5.0 | 22.4| 35.5|
| YOLO-v4       | CSPDarkNet-53    | 31             | 43.0| 64.9 | 46.5 | 24.3| 46.1| 55.2|
| SSD512        | VGG-16           | 22.0           | 28.8| 48.5 | 30.3 | 10.9| 31.8| 43.5|
| DSSD321       | ResNet-101       | 9.5            | 28.0| 46.1 | 29.2 | 7.4 | 28.1| 47.6|
| FSSD512       | VGG-16           | -              | 31.8| 52.8 | 33.5 | 14.2| 35.1| 45.0|
| RefineDet320  | VGG-16           | 38.7           | 39.4| 49.2 | 31.1 | 10.0| 32.0| 44.4|
| M2Det512      | ResNet-101       | 15.8           | 38.8| 59.4 | 41.7 | 20.4| 43.9| 53.4|
| RetinaNet500  | ResNet-101       | 5.4            | 34.4| 53.1 | 36.8 | 14.7| 38.5| 49.1|
| CornerNet512  | Hourglass        | 4.1            | 40.5| 57.8 | 45.3 | 20.8| 44.8| 56.7|
| CenterNet-HG  | Hourglass-104    | 7.8            | 42.1| 61.1 | 45.9 | 24.1| 45.5| 52.8|
| EfficienDet-D0| EfficienNet-B0   | -              | 33.8| 52.2 | 34.4 | -   | -   | -   |
| CentripetalNet| Hourglass-104    | 48.0           | 65.1| 51.8 | 29.0 | 50.4| 59.9|

Figure 3 EfficienDet structure diagram
3. Principle analysis and hardware circuit diagram

3.1. Monitoring module
The monitoring function is mainly to shoot the target area in real time through the image acquisition module (camera). And pass the image to the analysis and judgment module (central control computer) through the transmission line, and then pass the trained target detection model, and finally return the result of whether the entry person is wearing a mask.

3.2. Wireless communication module
The biggest feature of LoRa wireless communication is that it has a longer propagation distance than other wireless methods under the same circumstances, and realizes the agreement of low power and long distance. Under the same power consumption, it is 3-5 times longer than the traditional radio frequency communication distance.[7] According to actual needs, choose ATK-LoRa-01-V3.0. The module design uses the ISM band radio frequency SX1278 expansion chip, the module working frequency is 410Mhz-441Mhz, and the communication distance is about 2000 meters. The connection between the module and MCU/ARM equipment is shown in Figure 4. This module mainly plays the role of receiving signals for wireless transmission.

3.3. Access control module
After the image enters the analysis and judgment module, the final result is transmitted to the receiving end through the transmitting end of the LoRa wireless transmission module. The access control motor and buzzer complete corresponding actions according to their results. If the judgment result is that someone is not wearing a mask, the access control is closed and the buzzer emits a warning sound. Otherwise, the door is open, allowing people to enter. The overall process is shown in Figure 5.
3.4. Overall circuit and physical diagram

The signal sending end is mainly composed of STM32F103 chip, LoRa wireless communication module and USB to serial port. The USB to serial port module is responsible for transmitting the data processed by the computer to the F103 chip, and the F103 controls the LoRa wireless communication module to send the data to the signal receiving end through the written program. The specific circuit diagram is shown in Figure 6.

The signal receiving end is mainly composed of STM32F103 chip, LoRa wireless communication module, motor module and buzzer module. LoRa wireless communication module The Lora module is responsible for receiving the signal sent from the signal transmitter. When the signal indicates that the person wearing a mask, the motor will start to rotate, and the buzzer will not sound; when the signal indicates that the person is not wearing a mask, the motor will not Turn, the buzzer will sound an early warning; when the signal indicates that no one is visiting, the buzzer will not sound and the motor will not rotate. The specific circuit diagram is shown in Figure 7 below.

According to the specific circuit diagram, we drew the PCB board ourselves and soldered the corresponding components on the PCB board. The specific physical diagram is as follows.

![Figure 6 Physical map of the wireless transmitter](image1)

![Figure 7 Circuit diagram of wireless transmitter](image2)
4. System testing and error analysis

In order to evaluate the performance of the model, several performance indicators need to be used, namely accuracy and recall, as shown in the following formula:

$$\text{Precision} = \frac{TP}{TP + FP}$$  \hspace{1cm} (1)
The performance metrics are formulated in terms of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

\[
Recall = \frac{TP}{TP + FN}
\]  

Figure 11 Confusion Matrix.

Because of the equipment performance, the network model did not choose the EffectiveDet-D7, but chose the EffectiveDet-D0. Compared with the 52M parameters of the D7 type, the D0 type has only 3.9M parameters. Although the network performance is much lower than that of the D7, it can already meet most of the performance needs of the system, and it can be very multi-person face mask recognition.

5. Conclusion

In this article, we systematically summarize the performance comparison of target detection algorithms, and select the best algorithm for the research situation in this article. Using deep learning target detection algorithms, we have proposed a high-performance face mask recognition and epidemic prevention system, which can ensure that people who enter public places and those who are already in public places wear masks. The model is trained on the Wider Face data set after SMFD and artificial correction, and the recognition rate can reach 95%. This system can well meet the actual epidemic prevention needs at a low cost, and has a wide range of application value and strong practical significance.

References

[1] Liu X, Zhang S. COVID-19: Face masks and human-to-human transmission[J]. Influenza and Other Respiratory Viruses, 2020.
[2] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
[3] Wang Z, Wang G, Huang B, et al. Masked face recognition dataset and application[J]. arXiv preprint arXiv:2003.09093, 2020.
[4] Yang S, Luo P, Loy C C, et al. Wider face: A face detection benchmark[C]//Proceedings of the
IEEE conference on computer vision and pattern recognition. 2016: 5525-5533.

[5] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of
the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10781-10790.

[6] Tan M, Le Q V. Efficientnet: Rethinking model scaling for convolutional neural networks[J].
arXiv preprint arXiv:1905.11946, 2019.

[7] Cattani M, Boano C A, Römer K. An experimental evaluation of the reliability of lora long-range
low-power wireless communication[J]. Journal of Sensor and Actuator Networks, 2017, 6(2): 7.