Face Mask Recognition System with YOLOV5 Based on Image Recognition

Guanhao Yang[1]*, Wei Feng[2]*, Jintao Jin[1], Qujiang Lei[1]†, Xiuhao Li[1], Guangchao Gui[1], Weijun Wang[1]
Intelligent Robot & Equipment Center
[1]Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences
Guangzhou, 511458, China
[2]University of Chinese Academy of Sciences
Beijing, 100039, China
*Co-first authors
†Corresponding author
e-mail: qj.lei@aaia-ai.org

Abstract—**The rapid development of computer vision makes human-computer interaction possible and has a wide application prospect. Since the discovery of the first case of COVID-19, the global fight against the epidemic has begun. In addition to various studies and findings by medical and health care experts, people's daily behaviors have also become key to combating the epidemic. In China, the government has taken active and effective measures of isolation and closure, as well as the active cooperation of the general public, such as it is unnecessary to stay indoors and wear masks. China, as the country with the first outbreak of the epidemic, has now become the benchmark country of epidemic prevention in the world. Of course, it is not enough for people to wear masks consciously. Wearing masks in all kinds of public places still needs supervision. In this process, this paper proposes to replace manual inspection with a deep learning method and use YOLOV5, the most powerful objection detection algorithm at present, to better apply it in the actual environment, especially in the supervision of wearing masks in public places. The experimental results show that the algorithm proposed in this paper can effectively recognize face masks and realize the effective monitoring of personnel.**

Keywords-computer vision; COVID-19; YOLOV5; wear masks; deep learning; in public places

I. INTRODUCTION

Computer vision technology uses a variety of imaging systems instead of visual organs as input means, using computers to replace the brain to complete the processing and interpretation of visual information. With the continuous development of computer vision technology, computers can recognize all kinds of faces and give feedback. At present, face recognition of masks is most widely used in public places.

Since the first case of pneumonia of unknown cause appeared in Wuhan, China, in late 2019, the world has been gripped by a new pandemic. The World Health Organization (WHO) named mycoplasma pneumoniae 19 caused by the novel coronavirus on 12 January 2020, as “COVID-19” and raised the global risk level for COVID-19 to the highest level [1].

The difference between the COVID-19 epidemic and previous infectious diseases is that it can spread through aerosols and infect people within a very short period by contact. According to a report by Beijing SATELLITE TV on June 25, a couple in Beijing's Hai Dian district became infected simply by visiting a public toilet. In addition, COVID-19 lurks in the body of an infected person until symptoms appear 20 days later [2].

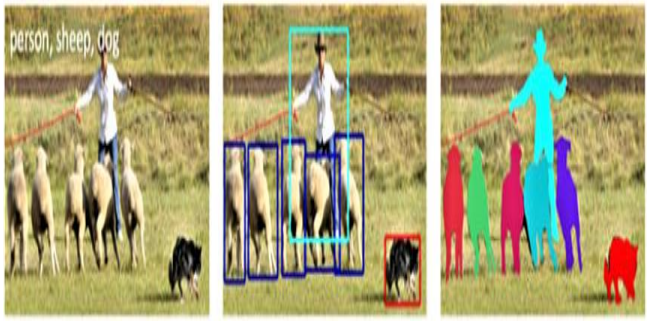
During this time, the patients will not show any symptoms, which means that asymptomatic patients cannot be detected and the government cannot isolate them in time to avoid further disaster. So far, a vaccine and a specific for Covid19 are on the way, but until the virus is eradicated, wearing masks has become the most effective way for people around the world to avoid infection. Around the world, masks were once in short supply, and merchants even raised the price of masks by dozens of times until the market supply could meet consumer demand, so masks have become an essential item for personal travel. As masks have become a necessity, many public places where crowds gather, such as malls and swimming pools, need to be monitored for wearing them, but most Chinese shopping malls today still use manual and restricted access to monitor the flow of people wearing them. Although manual inspection has certain defects, it can be avoided by adding more manpower. Although The epidemic prevention and control in China is excellent, the occurrence of infection cannot be completely avoided in places with a large and scattered population, such as shopping malls, because it will not only waste a lot of resources and manpower to assign staff to supervise at each entrance of shopping malls. In addition, if one of the entrances has a large flow of people, it may cause staff to miss the inspection. In addition, human supervision also wastes time, causing a large number of people to gather at the entrance of shops and supermarkets, and creating the risk of infection. In this paper, of course, there will also be temperature detection and photo-taking in the identification mask. Only when these are met at the same time, the gate will be opened and people can enter the site, otherwise, it will not be opened. This paper mainly addresses the identification of masks.

To solve this problem, this paper proposes a wearable detection algorithm based on deep learning YOLOV5. The algorithm can identify whether faces in public places such as malls and factories are wearing masks.

II. Related Work

A. Object Detection

Object detection tasks can be understood via Fig. 1.



(a)classification (b)detection (c)segmentation

Figure 1. Three steps about object detection

1. Classification, which is to structure the image into a certain type of information, with a predefined category (String) or instance ID to describe the image.

2. Detection, the classification task is concerned with the whole. Compared with classification, detection gives the understanding of picture foreground and background. The output of the detection model is a list, and each item of the list used a data group to give the category and position of the detected target (commonly used coordinate representation of rectangular detection box).

3. Segmentation, which is a pixel-level description of an image, which gives meaning to each pixel category (instance) and is suitable for scenes requiring high understanding.

B. One-Stage and Two-Stage

The one-stage network is represented by the YOLO series network, while the two-stage network is represented by faster-RCNN [3].

C. YOLOV5

Joseph Redmon, the original author of the YOLO algorithm, announced that he would stop all research in the CV field after he became dissatisfied with the military and privacy applications of his open-source algorithm. Then, when the YOLO series went nowhere, Alexey Bochkovskiy published a paper and carried on the development of the YOLO series on April 23, 2020, and subsequently received the official approval of YOLO. In the heat of the YOLOV4 continues, on May 30, issued by YOLOV5 Ultralytics LLC team, although the original author's official website has not issued an acknowledged, and many also will be as YOLO4.5 but do not affect its usefulness, compared with other YOLO series, by darknet into PyTorch YOLOV5 realization way, and more in YOLOV4, YOLOV5 can achieve 140 FPS on Tesla P100 rapid detection, YOLOv4 is only 50 FPS. Meanwhile, the size of YOLOV5 is only 27 MB, while the size of YOLOv4 using Darknet architecture is 244 MB. YOLOV5 also has the same accuracy as YOLOV4.

YOLOV5 has inherited the advantages of YOLOV4, namely adding SPP-NET (as shown in the figure below), modifying the SOTA method, and putting forward new data enhancement methods, such as Mosaic training, self-adversary training (SAT), and multi-channel feature replacing FPN fusion with PANet [4].

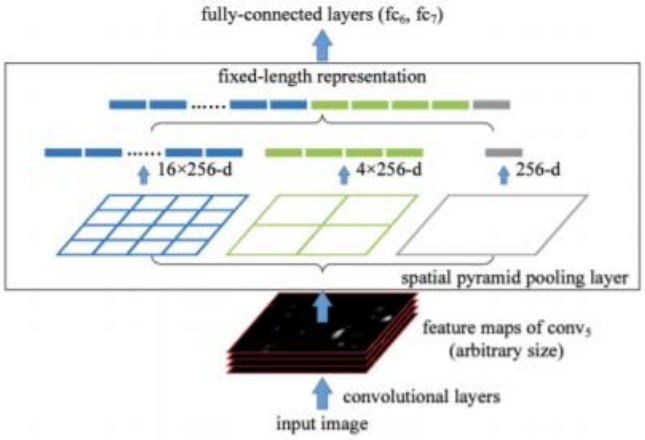


Figure 2. SPP-net

The network structure diagram of YOLOV5 is shown in Fig. 3(mainly the structure diagram of YOLOV5S).

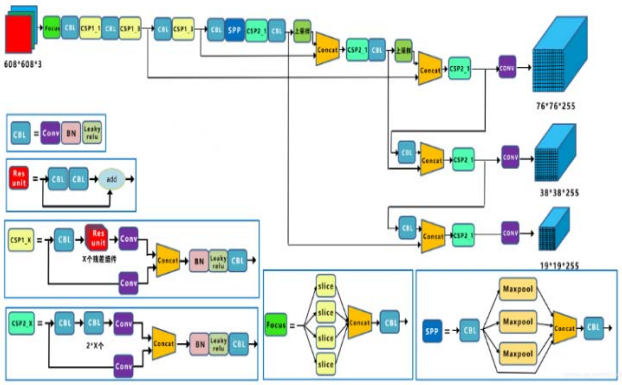


Figure 3. YOLOV5 network

D. Center Loss [5]

In recent years, in addition to the improvement of network structure, there are also a group of people studying the improvement of the loss layer. Wen Yandong introduced the monitoring method of Center Loss [5] in a novel way. It can effectively enhance the recognition ability of deep learning features in the neural network. The formulated in Eq. 1

$$L_c = \frac{1}{2}\sum_{i=1}^{m}\left\|x_i - c_{y_i}\right\|_2^2 \qquad (1)$$

The gradients of LC for x_i and update equation of c_{y_i} are computed as:

$$\frac{\partial L_c}{\partial x_i} = x_i - c_{y_i} \tag{2}$$

$$\Delta_{c_j} = \frac{\sum_{i=1}^{m} \delta(y_i=j)\cdot(c_j - x_i)}{1+\sum_{i=1}^{m} \delta(y_i=j)} \tag{3}$$

where δ(condition) = 1 if the condition is satisfied, and δ(condition) = 0 if not. α is restricted in [0, 1].

After soft-max loss added center loss[5], the inter-class distance increased and the intra-class distance decreased. In his paper, Wen Yandong presented the comparison results of soft-max loss and soft-Max loss + Center loss, the direct impact is shown in Figure 4.

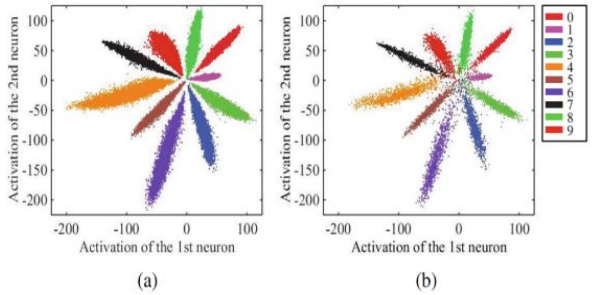


Figure 4. The distribution of deeply learned features in (a) training set (b) testing set, both under the supervision of soft-max loss, where we use 50K/10K train/test splits. The points with different colors denote features from different classes. Best viewed in color. (Color figure online)

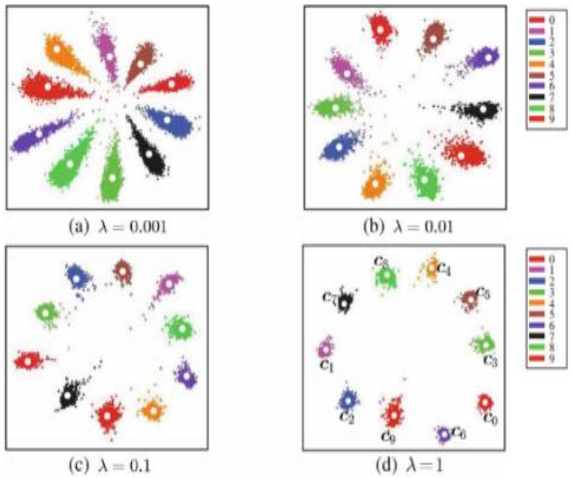


Figure 5. The distribution of deeply learned features under the joint supervision of soft-max loss and center loss [5]. The points with different colors denote features from different classes. Different λ lead to different deep feature distributions ($\alpha = 0.5$). The white dots ($c_0, c_1, \ldots, c_9$) denote 10 class centers of deep features. Best viewed in color. (Color figure online)

Here are some test images from the web that show the results in Fig. 6.

Figure 6. The detection results in markets

III. Proposed System

The system presented in our study is shown in Fig. 7. First of all, people entering the mall will take pictures through the camera, and the images will be sent to the interface for face mask recognition. If the face identified within two seconds is a face with a mask, the mall gate will be opened and displayed to pass, otherwise, it will be returned to face mask recognition until success.

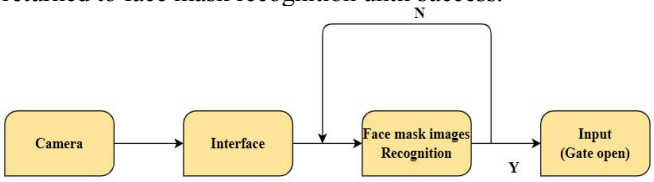


Figure 7. Working system

We divided the whole system into four parts: facial mask image enhancement, facial mask image segmentation, facial mask image recognition, and interface interaction. Facial mask image enhancement is used to improve the resolution of the mask worn for easy detection. Face mask image segmentation is used to extract mask information. The facial mask recognition part classifies the extracted mask information. The final interface output can make the gate open smoothly and help customers to enter. The recognition model is shown in Fig. 8.

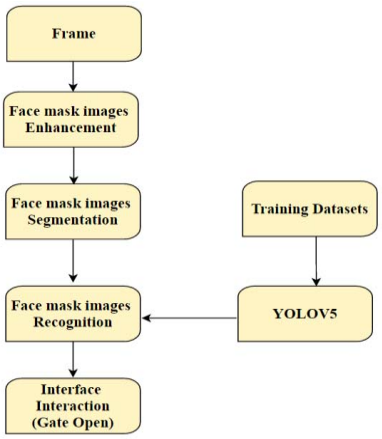


Figure 8. Recognition model

A. Enhancement of Facial Mask Images

In real life, when entering the mall, the image of face masks is often affected by the complex and unfavorable environment such as lights, stains, and colors, which reduces the image quality and weakens the features of masks. Therefore, it is very important to improve the image quality of digital images. This operation requires the enhancement of the part of the face with the mask.

The main purpose of image smoothing is to reduce image noise and extract useful information. A smoothing filter is to enhance the low-frequency component of the image, weaken the high-frequency component of the image, eliminate random noise, and play a smoothing role. Common smoothing filtering methods include mean filtering, median filtering, and Gaussian filtering [6-7].

(1). Gaussian filtering

Gaussian filter is a kind of linear smoothing filter, which is suitable for eliminating Gaussian noise. It is widely used in image processing [8]. Generally speaking, Gaussian filtering is a weighted average process of the whole image. The value of each pixel is obtained by the weighted average of itself and other pixel values in the neighborhood. The specific operation of Gaussian filtering is to scan every pixel in the image with a template (or convolution or mask) and replace the value of the central pixel of the template with the weighted average gray value of the pixel in the neighborhood determined by the template.

In Equation (1),x^2 and y^2 represent the distance between other pixels in the neighborhood and the center pixel in the neighborhood, respectively, and represent the standard deviation.

$$G_\sigma = \frac{1}{2\pi\sigma} e^{-(x^2+y^2)/2\sigma^2} \quad (4)$$

(2) Median filtering [9]

Median filtering method [9] is a nonlinear signal smoothing technology based on the sorting statistics theory that can effectively suppress noise. A two-dimensional sliding template is used to sort the pixels in the board according to the size of pixel values, and a monotonic ascending (or descending) two-dimensional data sequence is generated. The output of the two-dimensional median filter is

$$G(x, y)=\text{med}\{f(x-k, y-l),(k, l\in W) \quad (5)$$

Among them, f (x, y), g (x, y) of the original image and processed image respectively.

(3). Mean filtering [10]

Mean filtering [10] is a traditional image denoising method in the spatial domain and its application in image denoising is mainly to use various image smoothing templates for image convolution processing, to suppress or remove noise.

The basic idea of mean filtering is to replace the gray value of a pixel with the gray value of several pixels. For a pixel point (x, y) in a given image with f (x, y), its neighborhood S contains M pixels, the image after mean filtering and smoothing is g(x, y), g(x, y) is defined by the following formula(3):

$$G(x,y) = \frac{1}{M}\sum_{(i,j\in S)} f(x,y)(x,y) \notin S \quad (6)$$

The results of the comparison are shown in Fig 9.

Figure 9. Comparison of three methods

B. Segmentation of Facial Mask Images

Face segmentation is one of many image segmentation algorithms, and it is the basic premise to realize Robustness and practical face recognition system. The image segmentation algorithm can accurately locate the contour of each organ of the face and extract the target area of interest from the whole image, thus establishing a description of the face image that is easier to analyze and more expressive [11].

At present, the face segmentation algorithm has become mature, and a large number of image segmentation algorithms based on good robustness are proposed.

The Model-driven segmentation algorithm is the most widely used face segmentation algorithm, but in the actual face recognition, the face image is often affected by some interference factors. Therefore, to extract more accurate target segmentation from the background image, model-based segmentation algorithms are used in face images. The active contour model [12], also known as the Snakes model, has proved to be an effective image segmentation framework. The active contour model can be roughly divided into two categories: the active contour model based on edge information [13][14][15] and the active contour model based on area information [16][17]. The C-V model proposed by Chan and Vese [16], the most popular regional active contour model, can naturally handle topological structure changes and global segmentation. However, the C-V model itself also has some shortcomings. The model assumes that the gray values of image pixels in the same region have good continuity, and the gray values of pixels in different regions differ greatly. This assumption is basically not true for many real images and the segmentation effect is bound to be poor. Wang et al. [18] proposed an effective local C-V model, which is insensitive to the selection of initial contour wave position and control parameters and has less time complexity.

Therefore, this paper adopts the C-V model based on the tending to the active contour model to operate.

The energy functional of the C-V model [16] is:

$$E(C, c_1, c_2) = \mu \oint_{\Omega} ds + \lambda_1 \iint_{\Omega_1} (I(x,y) - c_1)^2 dx\, dy + \lambda_2 \iint_{\Omega_2} (I(x,y) - c_2)^2 dx\, dy \quad (7)$$

$$M_1(\phi) = H(\phi) \quad M_2(\phi) = H(\phi) \quad (8)$$

$$c_i = \frac{\iint_{\Omega_i} M_i(\phi)\, I(x,y) dx\, dy}{\iint_{\Omega_i} M_i(\phi)\, dx\, dy} \quad i = 1,2 \quad (9)$$

Among them, I (x, y) is for image segmentation, C is the evolution curve, Ω_i (I = 1, 2) as the curve of the internal and external area, ϕ is the level set function, H (ϕ) is the Heaviside function, $\mu \geq 0$, $\lambda_1, \lambda_2 > 0$ is the weight coefficient. By the representation of the level set and gradient descent flow, the partial differential equation controlling the level set can be derived[19]:

$$\frac{\partial \phi}{\partial t} = \delta(\phi)[\mu div(|\nabla \phi|)] + \delta(\phi)[-\lambda_1 (I - c_1)^2 + \lambda_2 (I - c_2)^2] \quad (10)$$

$\delta(x)$ is the Dirac function.

C. Face Mask Image Recognition

As the latest target Detection method of YOLO (YOU ONLY LOOK ONCE: Unified, real-time Object Detection) series, YOLOV5 combines the functions of YOLOV4, and at the same Time proposes new data enhancement methods, mosaics training, self-antagonistic training (SAT), and PANet[4] instead of FPN. In this paper, we use YOLOV5 to carry out the training data set, which can quickly and accurately identify whether a face mask is worn or not.

D. Interface Interaction

Human-computer Interaction (HCI) refers to the process of information exchange between a person and a Computer to complete certain tasks by using a certain dialogue language and in a certain way. In the current era of artificial intelligence, human-computer interaction has a very important impact and significance [20].

In this paper, we design a man-machine interface for the recognition of face to wear masks, when people enter the store into the door of the mall will be prompted to "face look at the camera," since then, the computer whether to wear masks to customers, which can identify if the customer does not wear masks gate won't open, considering the various complex environment can affect the accuracy of the results, we will identify the time of design for 2 seconds, after the success of the recognition, voice system will be prompted to "identify success", since gate open, customers can enter the market.

Figure 10 shows the process as people enter the mall.

Figure 10. The process of people entering the mall

IV. Experiment

A. Datasets

In this paper, the datasets we used are from the. (https://github.com/AIZOOTech/FaceMaskDetection) AIZOOTe team's FaceMaskDetection. The datasets contain images that can detect faces and determine whether a mask is worn, and open source 7,959 facial mask annotation data. We select 92% in the data set for training and 8% for testing. In the classification, "0" stands for "mask" and "1" stands for "no mask".The experimental data are shown in Figure 11.

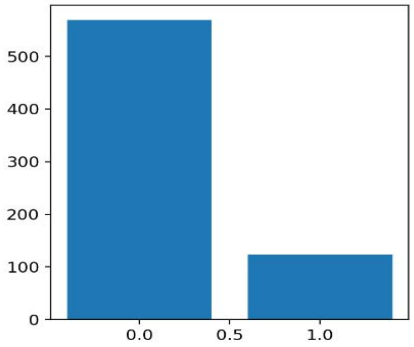


Figure 11. The different classes about “no mask” and “mask”

B. GIOU Loss [21]

The formula for GIOU is as follows

Algorithm 1: Generalized Intersection over Union

Input: Two arbitrary convex shapes: A, B $\subseteq \mathbb{S} \in \mathbb{R}^n$

Output: GIoU

1.For ***A*** and ***B***, find the smallest enclosing convex object ***C***, where C $\subseteq \mathbb{S} \in \mathbb{R}^n$

2.IoU=$\frac{|A \cap B|}{|A \cup B|}$

3.GIoU=IoU-$\frac{|C(A \cup B)|}{|C|}$

Algorithm 2: IoU and GIoU as bounding box losses

Input: Predicted $\boldsymbol{B}^{p}$ and ground truth $\boldsymbol{B}^{g}$ bounding box coordinates:

$\boldsymbol{B}^{p}=(x_1^{p},y_1^{p},x_2^{p},y_2^{p}),\boldsymbol{B}^{g}=(x_1^{g},y_1^{g},x_2^{g},y_2^{g})$

Output: $\mathcal{L}_{IoU},\mathcal{L}_{GIoU}$.

1.For the predicted box $\boldsymbol{B}^{p}$, ensuring $x_2^{p}>x_1^{p}$ and $y_2^{p}>y_1^{p}$:

$$\hat{x}_1^{p}=\min(x_1^{p},x_2^{p}),\hat{x}_2^{p}=\max(x_1^{p},x_2^{p})$$
$$\hat{y}_1^{p}=\min(y_1^{p},y_2^{p}),\hat{y}_2^{p}=\max(y_1^{p},y_2^{p}).$$

2.Calculating area of $\boldsymbol{B}^{g}$: $\boldsymbol{A}^{g}=(x_2^{g}-x_1^{g})\times(y_2^{g}-y_1^{g})$.

3.Calculating area of $\boldsymbol{B}^{p}$: $\boldsymbol{A}^{p}=(\hat{x}_2^{p}-\hat{x}_1^{p})\times(\hat{y}_2^{p}-\hat{y}_1^{p})$.

4.Calculating intersection $\mathcal{L}$ between $\boldsymbol{B}^{p}$ and $\boldsymbol{B}^{g}$:

$$x_1^{\mathcal{L}}=\max(\hat{x}_1^{p},x_1^{g}),x_2^{\mathcal{L}}=\min(\hat{x}_2^{p},x_2^{g}),$$
$$y_1^{\mathcal{L}}=\max(\hat{y}_1^{p},y_1^{g}),y_2^{\mathcal{L}}=\min(\hat{y}_2^{p},y_2^{g}).$$
$$\mathcal{L}=\begin{cases}\left(x_2^{\mathcal{L}}-x_1^{\mathcal{L}}\right)\times\left(y_2^{\mathcal{L}}-y_1^{\mathcal{L}}\right) & \text{if } x_2^{\mathcal{L}}>x_1^{\mathcal{L}},y_2^{\mathcal{L}}>y_1^{\mathcal{L}}\\ 0 & otherwise\end{cases}$$

5.Finding the coordinate og smallest enclosing box:$\boldsymbol{B}^{c}$:

$$x_1^{c}=\min(\hat{x}_1^{p},x_1^{g}),x_2^{c}=\max(\hat{x}_2^{p},x_2^{g}).$$
$$y_1^{c}=\min(\hat{y}_1^{p},y_1^{g}),y_2^{c}=\max(\hat{y}_2^{p},y_2^{g}).$$

6.Calculating area of $\boldsymbol{B}^{c}$:$\boldsymbol{A}^{c}=(x_2^{c}-x_1^{c})\times(y_2^{c}-y_1^{c})$.

7.IoU=$\frac{\mathcal{L}}{\mathcal{U}}$,where$\mathcal{U}=\boldsymbol{A}^{p}-\boldsymbol{A}^{g}-\mathcal{L}$

8.GIoU=IoU-$\frac{A^{c}-\mathcal{U}}{A^{c}}$.

9.$\mathcal{L}_{IoU}$=1-IoU, $\mathcal{L}_{GIoU}$=1-GIoU.

In this paper, we adopted the method of combining GIOU Loss [21] and Center Loss [5] to identify whether a face mask is worn or not. For face-covering recognition, masks can be regarded as covering objects, and Center Loss [5] can recognize faces, so we used this method to carry out experiments. The test results of the model used in the experiment are shown in Fig. 12.

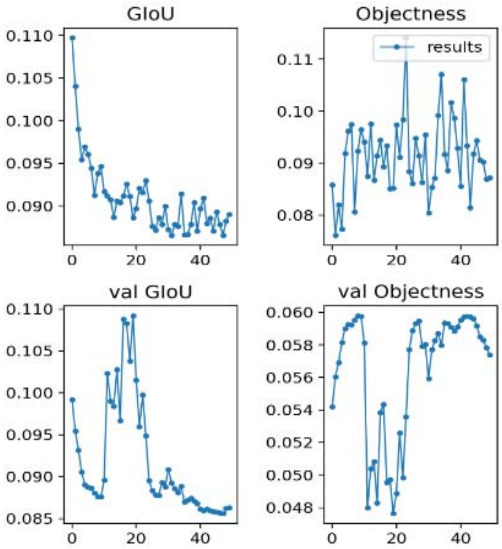


Figure 12. The GIOU Loss curve of training

V. Conclusion

This paper puts forward a new method based on YOLOV5 for application to recognize faces whether wearing masks, people into the store, they just need to stand in front of the camera, people can be identified, if recognition success and interface display can enter the gate open, this method is no longer need to use the human crowd control, greatly saves time and waste. Through testing, our experiment has a success rate of about 97.9%, we select some other classic machine learning models for comparison. The result is shown in Fig.13. Also, there is a picture wearing a mask but not covering the nose, which can also be well recognized. The experimental results are shown in Fig.14.Because of the global impact of COVID-19, we believe that this design can effectively reduce exposure distances and implement effective surveillance.

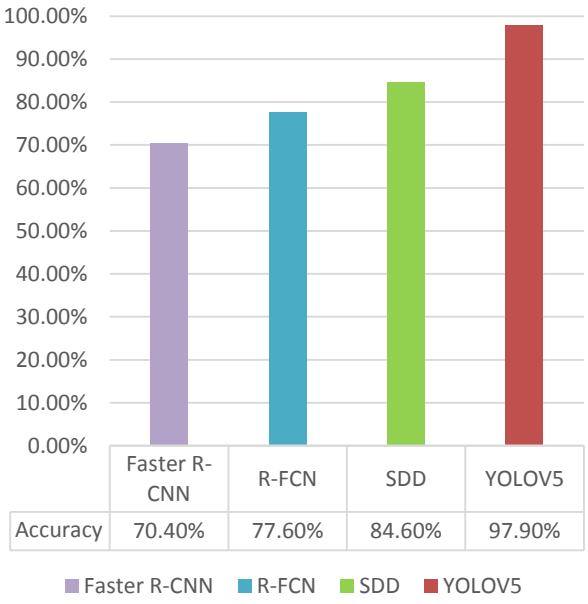


Figure 13. Accuracy comparison

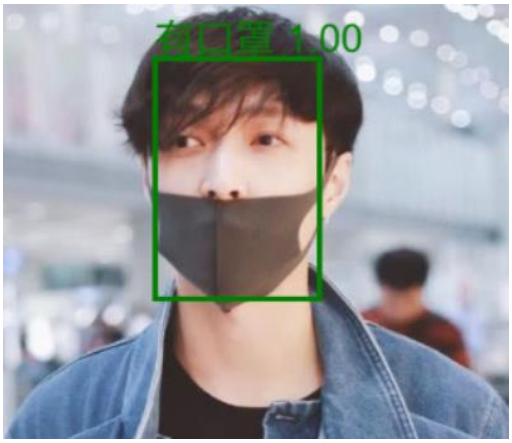


Figure 14. The detection results of wearing the mask but not covering the nose

VI. Future Work

In this paper, YOLOV5 is applied to identify whether a face mask is worn so that the gate at the entrance of the shopping mall can be opened and closed successfully. However, this kind of recognition is only for mask recognition, if in some special circumstances, the customer covers part of the mask with his hand, it will not be recognized successfully. In the future, we will improve the situation where masks covered by hands or other shielding objects cannot be recognized, making it more convenient for people to enter the shopping mall in this special environment, and the identification system will be more intelligent.

Acknowledgement

This research is supported by the Guangdong Province Key Field R&D Program Project (Grant No. 2020B090925002)， the Major Projects of Guangzhou City of China (Grant No. 201907010012).

References

[1] World Health Organization. (2020). Coronavirus disease 2019 (COVID-19): situation report, 72.

[2] Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S. M., Hayashi, K., ... & Linton, N. M. (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). International Journal of infectious diseases, 94, 154.

[3] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[4] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Pathaggregation network for instance segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8759-8768).

[5] Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016, October). A discriminative feature learning approach for deep face recognition. In European conference on computer vision (pp. 499-515). Springer, Cham.

[6] Li Wei. Application and research of curve fitting in steel pipe counting [D]. Southwest Jiaotong University, 2010.

[7] Milan Sonka, AI Haizhou. Image processing, analysis, and machine vision [M]. Tsinghua University Press, 2016.

[8] Wei Ying Gaussian filter OpenMP parallelization [J].Telecom World, 2015(10) 194-194.

[9] Gong Shengrong, Liu Chunping, Wang Qiang, Digital image processing, and analysis [M].Beijing ; Tsinghua University Press, 2006.

[10] YAN Bing, WANG Jin-he， ZHAO Jing. Research of Image De-noising Technology Based on Mean Filtering and Wavelet Transformation[J]. Computer Technology and Development.2011,21(2). (PP51-53,57)

[11] Li Dong mei. Research of Masked Face Recognition based on Image segmentation [D].Wuhan: South-central University For Nationalities, 2015 ; (P5-P8)

[12] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models [J]. International journal of computer vision, 1988, 1(4): 321-331.

[13] Caselles V, Catté F, Coll T, et al. A geometric model for active contours in image processing [J]. Numerische Mathematik, 1993, 66(1): 1-31.

[14] Caselles V, Kimmel R, Sapiro G. Geodesic active contours [J]. International journal of computer vision, 1997, 22(1): 61-79.

[15] Malladi R, Sethian J A, Vemuri B C. Shape modeling with front propagation: A level set approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(2): 158-175.

[16] Chan T F, Vese L A, Active contours without edges [J]. IEEE Transactions on Image Process. 2001,10 (2): 266-277.

[17] Tsai A, Yezzi A, Willsky A S. Curve evolution implementation of the Mumford–Shah functional for image segmentation, denoising, interpolation, and magnification [J]. IEEE Transactions on Image Process. 2001,10 (8) :1169-1186.

[18] Liu C, Wechsler H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition [J]. IEEE Transactions on Image processing, 2002, 11(4): 467-476.

[19] Xu Dong Peng Zhenming. Improved image segmentation method based on fast level set and C-V model [J]. High Power Laser and Particle Beams. 2012,24(12); 2817-2821.

[20] Lin H I, Hsu M H, Chen W K. Human hand gesture recognition using a convolution neural network[C]//2014 IEEE International Conference on Automation Science and Engineering (CASE). IEEE, 2014: 1038-1043.

[21] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, Silvio Savarese. (2019) Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression[D]/ https://arxiv.org/pdf/1902.09630.pdf