Journal of Multimedia Information System VOL. 8, NO. 3, September 2021 (pp. 159-166): ISSN 2383-7632 (Online)
http://doi.org/10.33851/JMIS.2021.8.3.159

A Mask Wearing Detection System Based on Deep Learning

Shilong Yang[1], Huanhuan Xu[1], Zi-Yuan Yang[2*], Changkun Wang[3]

Abstract

COVID-19 has dramatically changed people's daily life. Wearing masks is considered as a simple but effective way to defend the spread of the epidemic. Hence, a real-time and accurate mask wearing detection system is important. In this paper, a deep learning-based mask wearing detection system is developed to help people defend against the terrible epidemic. The system consists of three important functions, which are image detection, video detection and real-time detection. To keep a high detection rate, a deep learning-based method is adopted to detect masks. Unfortunately, according to the suddenness of the epidemic, the mask wearing dataset is scarce, so a mask wearing dataset is collected in this paper. Besides, to reduce the computational cost and runtime, a simple online and real-time tracking method is adopted to achieve video detection and monitoring. Furthermore, a function is implemented to call the camera to real-time achieve mask wearing detection. The sufficient results have shown that the developed system can perform well in the mask wearing detection task. The precision, recall, mAP and F1 can achieve 86.6%, 96.7%, 96.2% and 91.4%, respectively.

Key Words: Mask wearing detection system, Deep learning, Image detection, Video detection, Real-time detection.

I. INTRODUCTION

People's daily lives have been dramatically changed for the COVID-19 outbreak. The epidemic remains severe, as of April 20, 2021, the number of new confirmed cases of COVID-19 has increased for eight consecutive weeks. More than 5.2 million new confirmed cases over the world in this week, and the death rate of this week also increases compared to the previous weeks 0. The virus-containing respiratory fluid from an infected COVID-19 person can remain airborne for several hours, which is also the main transmission way of the epidemic 0. Masks are effective at defending the virus particles, so wearing masks scientifically is a simple and effective way to prevent the spread of COVID-19.

In the context of the epidemic, how to supervise wearing masks in the public areas is a key problem that needs to be urgently solved. Unfortunately, manual testing methods are still used in many places, which means inspectors are in infection danger and the detection rates cannot be guaranteed. Compared with other technologies, vision-based computer technologies have many advantages, such as low cost and high detection rate 0-0 . However, traditional method-based detection systems cannot overcome the complex problems in real environments, such as tiny targets, various targets, and complex backgrounds. Compared with them, deep learning-based methods can keep high accuracy even in complex backgrounds 0-0.

To alleviate these difficult problems mentioned above, a deep learning-based mask wearing system is developed in this paper, sufficient experiments have shown that the proposed system can achieve promising results in the real environments. However, deep learning-based methods always need many data to train, but because of the suddenness of the epidemic, the datasets are scarce for mask detection. Hence, a mask wearing dataset is collected in this paper.

The main contributions can be summarized as follows:

1. A powerful deep learning-based mask wearing detection system is developed to reduce the burden and the infection risks of the inspectors. The system can satisfy all the three requirements: image detection, video detection and real-time detection.
2. A deep learning-based multi-object detection is adopted, and this method has achieved promising results. In particular, some other detection methods can

Manuscript received August 26, 2021; Revised September 15, 2021; Accepted September 17, 2021. (ID No. JMIS-21M-08-027)

Corresponding Author (*): Zi-Yuan Yang, Sichuan University, Section 4, Southern Frist Ring Road, Chengdu, 610065, P. R. China, cziyuanyang@gmail.com.

[1]School of Software, Nanchang Hangkong University, No. 696, Fenghenan Road, Nanchang, 330063, P. R. China

[2]School of Computer Science, Sichuan University, No. 24, Section 4, Southern Frist Ring Road, Chengdu, 610065, P. R. China

[3]School of Information Engineering, Nanchang Hangkong University, No. 696, Fenghenan Road, Nanchang, 330063, P. R. China

also be used in our system, here YOLO v3 is used as an instance.

3. A mask wearing dataset is collected to alleviate the scarcity of datasets.

The rest of this paper is organized as follows: Related works are introduced in Section II. Section III introduces the developed system in detail. Experiments are shown in Section IV to confirm the effectiveness of the proposed system. Finally, conclusions and future works are discussed in Section V.

II. RELATED WORKS

In recent years, deep learning has achieved remarkable performances in different important computer vision tasks, such as classification, detection, tracking, denoising and segmentation 0-0. Besides, deep learning is also widely adopted in face-related tasks and has achieved remarkable performances 0-0.

The target detection methods can be briefly categorized into two classes, which are regression-based one-stage and multi-class classifier-based two-stage methods. Two-stage methods are typically designed for some multi-class detection tasks. Redmon et al. 0 proposed You Only Look Once (YOLO) and Liu et al. 0 proposed Single Shot Detector (SSD). Inspired by SSD, Yang et al. 0 proposed a cascaded tiny face detection method, which can achieve a satisfactory performance. The classical two-stage methods are the family of Fast R-CNN 0. In this paper, a powerful detector YOLOv3 0 is adopted to detect masks. Compared with two-stage methods, the detection speeds of one-stage methods are commonly faster and YOLOv3 can also achieve a significant performance with a satisfactory detection speed.

In the real environments, the input of the monitoring system is a video rather than a simple image. Compared with images, videos are more difficult to recognize, because there are many fuzzy relationships between frames and the data size of the video is greater. Hence, a real-time tracking method is necessary for reducing the computation costs. Yuan et al. 0 proposed a scale-adaptive object-tracking with occlusion detection. Park and Kim 0 proposed an interactive system based on efficient eye detection and pupil tracking method. Bewley et al. 0 proposed a simple online and real-time tracking (SORT) method, which established the relationship between different frames, and reduced the computational time.

There are also deep learning-based mask detection methods. 0 used transfer learning with three popular baseline models: ResNet50, AlexNet and MobileNet, in order to conduct mask detection. In 0, the mask detection algorithm can be used in real-time application.

III. SYSTEM DEVELOPMENT

The proposed deep learning-based mask wearing detection system is developed based on Web. At first, the system calls the mask wearing detection model to detect and mark the masks on the uploaded images and videos. After that, the labeled images and videos are shown in the Web interface. The proposed system can be divided into three parts by functions, which are image detection, video detection and real-time detection. The whole framework of the proposed system is shown in Fig. 1.

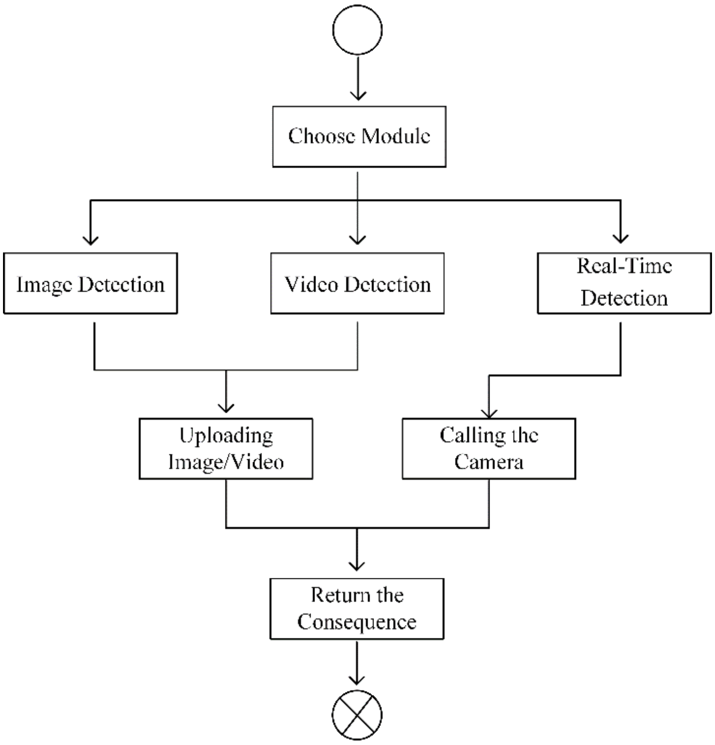
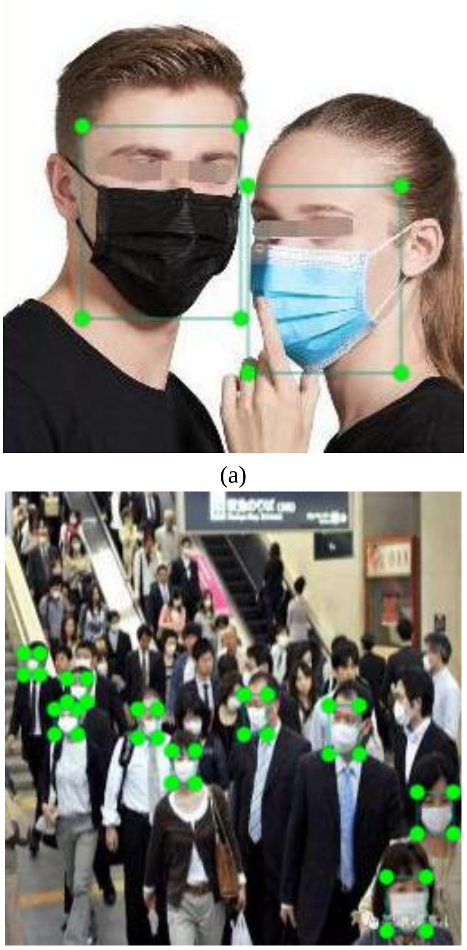


Fig. 1. The framework of the proposed mask wearing system.

2.1. Dataset Collection

In order to alleviate the scarcity of wearing mask dataset, a dataset is collected in this paper. This dataset contains total 2370 samples, 1182 wearing mask samples and 1188 without mask samples, these images are collected. There are total 2 classes, which are “mask” and “unmask”, respectively. LabelImg is used to label these images. In order to conform the real environments, the collected images contain a variety of different scenarios. The collected and labelled images are shown in Fig. 2.



(a)

(b)

Fig. 2. The collected samples. (a)-(b) represent the images in simple and complex environments, respectively.

2.2. Image Detection Module

The image detection module is implemented based on a deep learning-based target detection algorithm. As introduced above, detection methods can be divided into two classes, which are one-stage-based methods and two-stage-based methods. Two-stage-based methods are suitable for a large amount of category detection, one-stage-based methods are suitable for a small amount of category detection. There are only two categories of targets in our system, face and mask. Meanwhile, one-stage-based methods are always faster, which can ensure the proposed system can achieve real-time performance. Hence, a one-stage-based method, YOLOv3 is adopted in this system to detect mask wearing, the backbone of YOLOv3 is ResNet53.

In this method, the image is divided into $N \times N$ grid cells, 3 scale boxes are predicted in each cell. The upper-left and low-right coordinates of anchors and the class are predicted. The detection task in this paper is a 2-class task, so the tensor is $N \times N \times [3 \times (4 + 1 + 2)]$.

The activation function used is Leaky ReLU, the function is shown as follows:

$$f(x) = \max(\alpha x, x) \qquad (1)$$

where x is the input of this function, α is a predefined parameter in the range of $(0,1)$.

2.3. Video Detection Module

Video is composed of many frames, so the video detection problem can be treated as the image target detection. However, it is unacceptable that each frame is detected based on the image detection method, which definitely causes numerous computational costs, and the real-time performance cannot be guaranteed. It is important to extract the implicit relationships between different frames to reduce the computational cost. In this paper, the simple online and real-time tracking (SORT) is adopted to achieve multi object tracking. In SORT, the inter-frame displacements of each target with a linear constant velocity model are approximated. The state of each target is modelled as:

$$\mathbf{x} = [u, v, s, r, u', v', s']^T, \qquad (2)$$

where u and v represent the horizontal and vertical locations of the center of the target, respectively. s and r present the scale and the aspect ratio of the bounding box of the target, respectively. u', v' and s' are the predicted state values by Kalman filter framework 0.

The video detection module is composed of detector and tracker, which can achieve high accuracy and satisfactory real-time performance. The framework of video detection is shown in Fig. 3, the input video is divided into video frames, and then these images are detected by the detector, and then the detected targets are tracked by SROT, and finally the labeled targets are passed to a Web interface as the output.

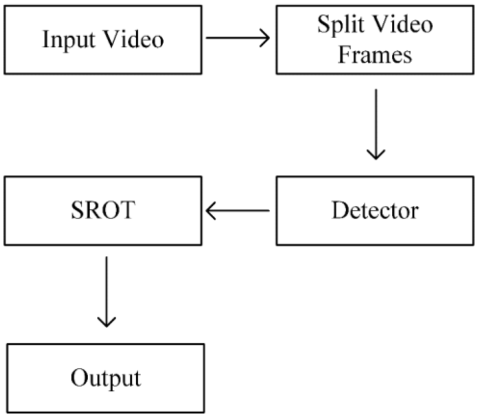


Fig. 3. The framework of video detection module. The video is passed to split the frames, then the split frames are passed to the detector and the targets are tracked by SROT, the results are output at last.

In order to achieve real-time monitoring function, an OpenCV based module is implemented. If the user chooses to real-time monitor, the module calls the camera to real-time detect mask wearing.

IV. EXPERIMENTS

In this section, experiments and implementation environments are introduced in detail. At the beginning of this section, we would introduce the experiment environments, including the versions of the equipment and the versions of the packages. The experiments are tested with Google Colab, the related environments are shown in Table 1.

Table 1. Experiment environments.

Environment	Version
GPU	Tesla P100 (16G)
CUDA Version	10.1
Operation System	Ubuntu 16.04

As introduced above, deep learning needs a large amount of data, but mask wearing dataset is scarcity. To alleviate this problem, a dataset is collected in this paper. This dataset contains total 2370 samples, 1182 wearing mask samples and 1188 without mask samples, these images are collected from Internet, 80% data is random selected as the training data, and the rest is selected as the testing data.

The proposed system is coded by PyTorch 0 and OpenCV, the main related versions of the used packages are shown in Table 2.

Table 2. Relate versions of the used packages.

Package	Version
PyTorch	≥1.5.1
numpy	1.17
Python	3.6.0

At first, we tested the image detection module of the proposed system. As the results shown in Table 3, the image detection module can achieve satisfactory performance in image detection. Meanwhile, the training process is shown in Fig. 4, the proposed system can converge quickly and the detect performance is promising.

Table 3. Testing results of different classes in different indicators.

Class	Precision	Recall	mAP@0.5	F1
all	0.866	0.967	0.962	0.914
mask	0.846	0.971	0.952	0.906
unmask	0.884	0.963	0.972	0.922

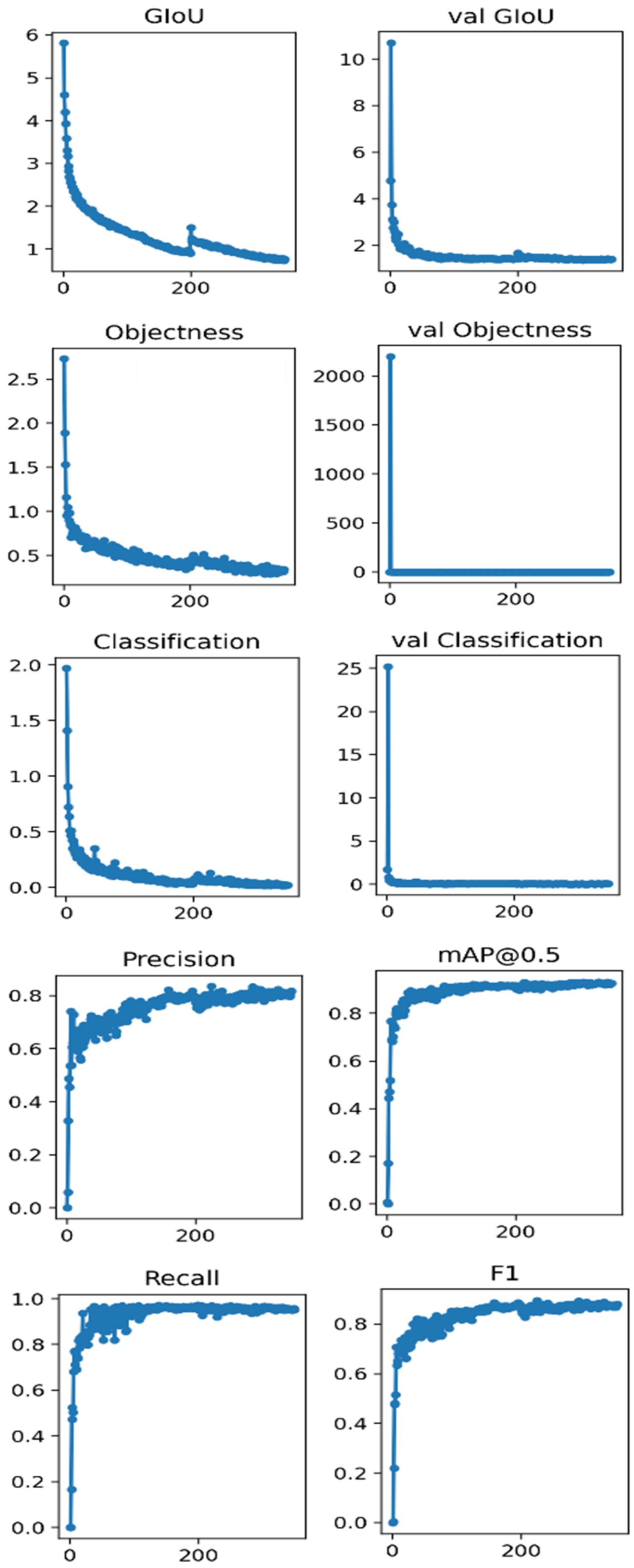


Fig. 4. The training results under different indicators, such as GIOU, classification loss, detection loss, precision, recall, mAP and F1 score.

The detected samples are shown in Fig. 5. It can be seen that the proposed system can detect targets accurately in multi-objectiveness and complex environments, and the detection speed is 12.7 ms per image, about 80.65 frame per second (FPS), which can achieve the real-time requirement.

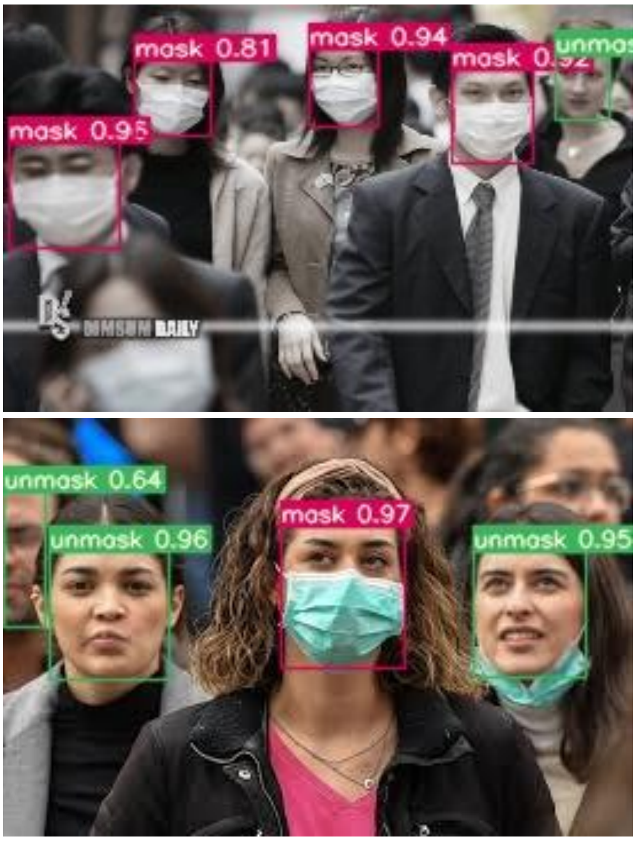


Fig. 5. The detected samples in different complex environments.

Meanwhile, we test the real-time monitoring and video detection modules, the results are shown in Fig. 6. The proposed system can achieve real-time detection and monitoring. Furthermore, the outpour of the system can be displayed in the Web interface accurately.

V. CONCLUSION

In this paper, a deep learning-based mask wearing system is implemented, which can work on personal computer without any high equipment requirement. Meanwhile, a mask wearing dataset, is collected to alleviate the scarcity of the similar datasets. The proposed system can achieve remarkable mask detection performance, and mAP is 96.2%, which is promising. Besides, the proposed system can achieve real-time detection based on low equipment. In the future works, we will increase the expansibility of the system, such as adding crowd-counting module and dangerous action recognition module. In addition, we will try to modify the model to improve the performance.

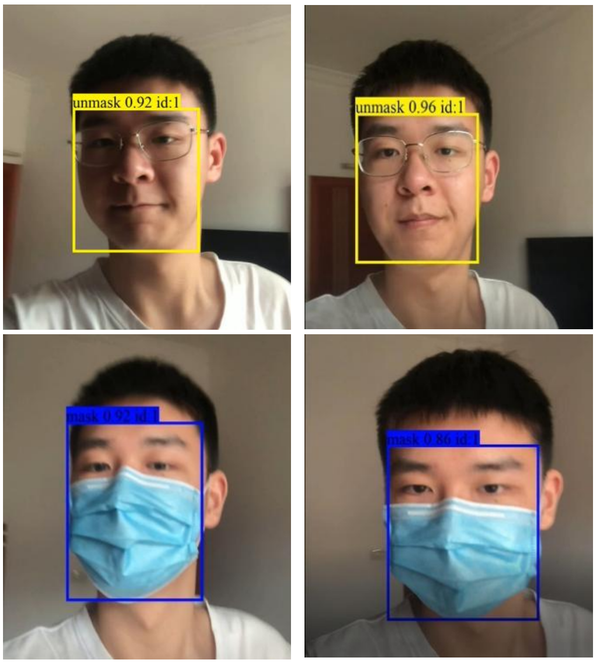


Fig. 6. Real-time detection and monitoring samples.

Acknowledgement

This research was funded by the National Natural Science Foundation of China (61866028), Key Program Project of Research and Development (Jiangxi Provincial Department of Science and Technology) (20192BBE50073).

References

[1] Weekly epidemiological update on COVID-19 - 20 April 2021, https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---20-april-2021, 2021.

[2] B. Asadi, N. Bouvier, A. S. Wexler, et al. "The coronavirus pandemic and ae-rosols: Does COVID-19 transmit via expiratory particles?" *The Lancet Respiratory Medicine*, vol. 8, no. 5, pp. 434-436, 2020.

[3] L. Leng, J. Zhang, M. K. Khan, et al. "Dynamic weighted discrimination power analysis: a novel approach for face and palmprint recognition in DCT domain," *International Journal of the Physical Sciences*, vol. 5, no. 17, pp. 2543-2554, 2010.

[4] Z. Yang, L. Leng and W. Min "Extreme Downsampling and Joint Feature for Coding-Based Palmprint Recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 1-12, 2021.

[5] L. Leng, S. Zhang, X. Bi, et al. "Two-dimensional cancelable biometric scheme, " in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition*, Xi'an, July 2012.

[6] L. Leng, J. Zhang, G. Chen, et al. "Two-directional two-

dimensional random projection and its variations for face and palmprint recognition," in *Proceedings of the International Conference on Computational Science and its Applications*, Berlin, pp.458-470, June 2011.

[7] Y. Zhang, J. Chu, L. Leng, et al. "Mask-refined R-CNN: A network for refining object details in instance segmentation, " *Sensors*, vol. 20, no. 4, pp. 1010.

[8] Z. Yang, L. Leng and B. G. Kim. "StoolNet for color classification of stool medical images," *Electronics*, vol. 8, no. 12, pp. 1464, 2019.

[9] Y. J. Heo, B. G. Kim and P. P. Roy. "Frontal face generation algorithm from multi-view images based on generative adversarial network," *Journal of Multimedia Information System*, vol. 8, no. 2, pp. 85-92, 2021.

[10] J. Chu, Z. Guo and L. Leng. "Object detection based on multi-layer convolution feature fusion and online hard example mining," *IEEE Acces*s, vol. 6, pp. 19959-19967, 2018.

[11] H. J. Kwon, G. P. Lee, Y. J. Kim, et al. "Comparison of pre-processed brain tumor MR images using deep learning detection algorithms," *Journal of Multimedia Information System*, vol. 8, no. 2, pp. 79-84.

[12] L. Leng, Z. Yang, C. Kim, et al. "A light-weight practical framework for feces detection and trait recognition," *Sensors*, vol. 20, no. 9, pp. 2644, 2020.

[13] J. H. Kim, B. G. Kim, P. P. Roy, et al., “Efficient facial expression recognition algorithm based on hierarchical deep neural network structure,” *IEEE Access*, vol.7, pp. 41273-41285, 2019.

[14] Y. J. Heo, B. G. Kim, P. P. Roy, “Frontal Face Generation Algorithm from Multi-view Images Based on Generative Adversarial Network,” *Journal of Multimedia Information System*, vol. 8, no. 2, pp. 85-92, 2019.

[15] A. Bhattacharyya, R. Saini, P. P. Roy, et al., “Recognizing gender from human facial regions using genetic algorithm,” *Soft Computing*, vol. 23, no. 17, pp. 8085-8100, 2019.

[16] J. H. Kim, G. S. Hong, B. G. Kim, et al., “deepGesture: Deep learning-based gesture recognition scheme using motion sensors,” *Displays*, vol. 55, pp. 34-45, 2018.

[17] J. Redmon, S. Divvala, R. Girshick, et al. "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 779-788, July 2016.

[18] W. Liu, D. Anguelov, D. Erhan, et al. "SSD: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, Amsterdam, pp. 21-37, Oct. 2016.

[19] Z. Yang, J. Li, W. Min, et al. "Real-time pre-identification and cascaded detection for tiny faces," *Applied Sciences*, vol. 9, no. 20, pp. 4344, 2019.

[20] R. Girshick. "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, pp. 1440-1448, Dec. 2015.

[21] J. Redmon and A. Farhadi. "Yolov3: An incremental improvement," arXiv preprint, arXiv:1804.02767, 2018.

[22] Y. Yuan, J. Chu, L. Leng, et al. "A scale-adaptive object-tracking algorithm with occlusion detection," *EURASIP Journal on Image and Video Processing*, vol. 1, pp. 1-15, 2020.

[23] S. J. Park and B. G. Kim. "Development of low-cost vision-based eye tracking algorithm for information augmented interactive system," *Journal of Multimedia Information System*, vol. 7, no. 1, pp. 11-16, 2020.

[24] A. Bewley, Z. Ge, L. Ott, et al. "Simple online and real-time tracking," in Proceedings of the *IEEE International Conference on Image Processing*, Phoenix, pp. 3464-3468, Sep. 2016.

[25] S. Sethi, M. Kathuria and T. Kaushik. “Face mask detection using deep learning: An approac h to reduce risk of Coronavirus spread,” *Journal of Bio medical Informatics*, vol. 120, pp. 103848, 2021.

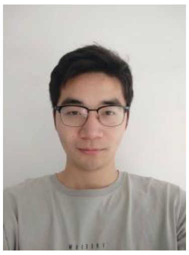
[26] S. Susanto, F. A. Putra, R. Analia, and I. K. L. N. Suciningtyas, “The face mask detection of preventing the spread of COVID-19 at politeknik negeri batam,” in *Proceeding of the 3-rd International Conference on Applied Engineering (I CAE)*, pp. 1-5, 2020.

[27] R. Kalman. "A new approach to linear filtering and prediction problems, " *Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[28] A. Paszke, S. Gross, F. Massa, et al. "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026-8037, 2019.

Authors

Shilong Yang is pursuing his B.E. degree in School of Software from Nanchang Hangkong University, P. R. China. His research interests include object detection and deep learning.



Huanhuan Xu obtained his B. E. degree from Nanchang Hangkong University, P. R. China. He is pursuing his M. E. degree with School of Software, Nanchang Hangkong University. His research interests include palmprint recognition, transfer learning and deep learning.

Zi-Yuan Yang obtained his B.E. degree and M. E. degree from Nanchang Hangkong University and Nanchang University, respectively. He is pursuing his Ph.D. degree with School of Computer Science, Sichuan University, P. R. China. He is a reviewer of several international journals. His research interests include pattern recognition, medical imaging, deep learning and security analysis.

Changkun Wang obtained his M.E. degree from Harbin Institute of Technology, P. R. China. He is currently an associate professor of School of Information Engineering, Nanchang Hangkong University.

He has presided over 30 engineering subjects and published over 30 papers. His research interests include control theory, control engineering and automatics.