



Category: STEM (Science, Technology, Engineering and Mathematics)

ORIGINAL

## High response speed and accuracy real time mask detection system for chinese campuses

### Sistema de detección de máscaras en tiempo real de alta velocidad de respuesta y precisión para los campus chinos

Baitong Zhong<sup>1,2</sup> , Johan Bin Mohamad Sharif<sup>2</sup> , Sah Salam<sup>2</sup> , Chengke Ran<sup>1</sup> , Zhuoxi Chen<sup>3</sup>  

<sup>1</sup>College of Information Engineering, Hunan Mechanical Electrical Polytechnic. Changsha, Hunan province, 410000, China.

<sup>2</sup>UTM Johor Bahru, Universiti Teknologi Malaysia. Johor, 81310, Malaysia.

<sup>3</sup>Hunan Vocational College of Electronic and Technology. Changsha, Hunan province, 410000, China.

Cite as: Zhong B, Mohamad Sharif JB, Salam S, Ran C, Chen Z. High response speed and accuracy real-time mask-detection system for chinese campuses. Salud, Ciencia y Tecnología - Serie de Conferencias. 2024; 3:937. <https://doi.org/10.56294/sctconf2024937>

Submitted: 10-02-2024

Revised: 02-05-2024

Accepted: 20-06-2024

Published: 21-06-2024

Editor: Dr. William Castillo-González 

#### ABSTRACT

Due to the increasing number of students studying in universities globally, the need for effective and timely safety measures has become more critical. This study aims to provide a high tech monitoring system that can help universities realize the security they need. The main functions are mask detection. Among them, mask detection is mainly used to determine if students are wearing the right masks. This paper also carried out algorithm provinciation for two kinds of detection. In the mask detection function, YOLOV4-Tiny model is used, and SPP is added and improved on this basis and replace the feature enhancement network with the path aggregation network (PAN). After the experiment, the accuracy was improved, Precision (P) and Recall (R) increase by 1,61 % and 4,14 %.and the response speed of mask detection was improved(The FPS reached 98,67) too. It greatly improves the efficiency of the system and provides security for students.

**Keywords:** Campus Security; Mask Recognition; YOLO; SPP; PAN.

#### RESUMEN

Debido al creciente número de estudiantes que estudian en universidades de todo el mundo, la necesidad de medidas de seguridad efectivas y oportunas se ha vuelto más crítica. Este estudio tiene como objetivo proporcionar un sistema de monitoreo de alta tecnología que pueda ayudar a las universidades a obtener la seguridad que necesitan. Las funciones principales son la detección de máscaras. Entre ellos, la detección de mascarillas se utiliza principalmente para determinar si los estudiantes están usando las mascarillas adecuadas. En este trabajo también se llevó a cabo la provinciación algorítmica para dos tipos de detección. En la función de detección de máscaras, se utiliza el modelo YOLOV4-Tiny y, sobre esta base, se agrega y mejora SPP. Y reemplaza la red de mejora de características por la red de agregación de rutas (PAN). Después del experimento, se mejoró la precisión, la precisión (P) y la recuperación (R) aumentaron en un 1,61 % y un 4,14 % y también se mejoró la velocidad de respuesta de la detección de máscaras (el FPS alcanzó 98,67). Mejora en gran medida la eficiencia del sistema y proporciona seguridad a los estudiantes.

**Palabras clave:** Seguridad del Campus; Reconocimiento de Máscaras; YOLO; SPP; PAN.

## INTRODUCTION

The safety of the people is a vital indicator of a country's political power. A country's development, including its military, scientific, and technological capabilities, should be focused on ensuring the happiness and safety of its citizens. According to a Chinese statesman and modern thinker named Qichao, a strong youth can make a country great.

To ensure the safety of its citizens, the government should prioritize the protection of the youth. One of the most important factors that young people consider when it comes to choosing a university is the availability of high-quality education. In 2021, China's university graduates reached 9 million, which is an increase of about 350 000 from the previous year.

The number of university students worldwide is 125 million. Since there are so many people on campuses and in the country, ensuring their safety has become a major concern. Teachers continue to value issues such as campus fire prevention, thievery prevention, bullying prevention, and psychological safety for students.

Due to the increasing number of students, the safety management procedures have become more complex. Traditional methods of addressing these issues have not been able to fully prevent accidents.

Although the use of manpower can prevent an accident from happening, it is also difficult to solve it in a quick manner.

Society has shifted away from relying on human labor. Instead, it now believes that technology can replace human labor. However, while technology can be useful in solving problems, it can also be very limiting when it comes to handling them. This is why it is important that the combination of technology and human is used to protect students' lives.

The goal of this study is to provide a comprehensive analysis of the various advantages of combining technology and manpower in order to address the issue of campus security.

## Problem Background

In the past few years, various universities and colleges have started using video monitoring equipment. However, this type of equipment is only used for investigations and cannot prevent accidents.

It's a huge manpower investment for schools because they can't have security guards standing in front of every surveillance device 24/7. To ensure a good teaching and research atmosphere in universities, especially to ensure a good public security environment, it is necessary to establish a comprehensive security prevention system on university campuses.<sup>(1)</sup> Various companies and individuals are working to provide solutions to the issues that arise due to the development of AI.

For example, in the case of COVID-19 since 2019, the campus needs to detect key areas such as libraries, cafeterias and other areas for any violation of Covid-19 protocols.<sup>(2)</sup>

An academic in India worked to build a reliable system to detect specific areas on campus, using object detectors such as:<sup>(3)</sup> YOLOv3, SSD<sup>(4)</sup> and the Mobile Net architecture<sup>(5)</sup> to detect violations and provide accurate results.

Through the use of a camera, the system can transmit information about the danger to the security personnel and the general console. However, this method can be very inefficient as many of these incidents happen on an emergency basis, which can prevent the police and security personnel from properly monitoring the situation. It is important that the system's response speed and accuracy are improved to protect the students.

Fast and accurate judgment depends on the accurate identification of risk warning modes for college students' safety prevention, including the construction and integration of functional modules such as data collection systems, information analysis and prediction systems, and decision alarm systems.<sup>(6)</sup> This paper will mainly focus on two kinds of campus safety precautions, mask detection and boundary detection. Among them, mask detection can effectively help students prevent the epidemic, and boundary detection can prevent students from approaching the dangerous area.

Mask detection is developed based on object detection technology. The rapid emergence and evolution of deep learning have changed the way object detection is done.<sup>(7)</sup> The conventional method now involves extracting features. There are two main types of extraction algorithms.

One-stage object detection is a type of algorithm that uses a neural network to identify and classify objects in images. It is very fast and can meet the system's requirements for real-time detection.<sup>(8)</sup> However, its accuracy is still not as good as it should be. Currently, most popular algorithms are the YOLOv3tiny, YOLOv4, YoloV4-Tiny, and SSD.<sup>(9)</sup>

Another type of algorithm is the two-stage object detection. It first makes a set of candidate regions that are likely to contain objects, then it categorizes them, and finally performs a bounding box regression. Some of the popular algorithms include R-CNN<sup>(10)</sup>, Faster R-CNN.<sup>(11)</sup>

Although both of these algorithms can successfully detect the target, they have distinct processes that make them different. One-stage detection is fast, but it can't provide the accuracy that it should. On the other hand, the two-stage method has a high recall rate and accuracy, but its response time is slow, which is unsuitable for real-time monitoring.

The goal of this project is to develop a one-stage algorithm that can accurately detect objects in images. Unfortunately, the general framework for object detection is not suitable for various applications. For instance, its model parameters and computation amount are too large.<sup>(12)</sup>

#### *Accuracy of mask detection*

The first points are the accuracy rate of mask detection is too low, which will lead to false positives or missed positives of the system. If the alarm is not accurately reported, the safety of students will not be guaranteed. So how to maintain a high accuracy is a problem that needs to be studied.

There are two problems in the accuracy of mask detection, the first one is what kind of feature extraction algorithms is used (one stage or two stage). Considering that this project is monitored in real time on campus, one stage is more suitable. However, although one stage feature extraction algorithms has a high response speed, in general, in order to meet the fixed-size input limit of the convolutional neural network classifier, the input image can be cropped, scaled and other operations, but this will often distort the image.

Researchers are working on improving the accuracy of feature extraction methods by ensuring that they can provide the fastest response time. Cabani et al.<sup>(13)</sup> proposed a large-scale mask detection dataset containing 130 000 images, but the image targets and background of people in this dataset are relatively simple, and the coverage of realistic scenes is seriously insufficient. Wang et al.<sup>(14)</sup> proposed a Real-World Masked FaceDataset (RMFD), but the scene coverage of the dataset was still insufficient, and more importantly, the open data annotation was not perfect. It is difficult to apply directly to the testing duties of the mouth mask. By combining transfer learning and Retina-Net network, Deng et al.<sup>(15)</sup> made the AP value of the trained model under the verification set 86,45 %

By using YOLOv3 and YCrCb methods, Xiao et al.<sup>(16)</sup> made the mAP of face mask detection reach 89,04 %. Wang et al.<sup>(17)</sup> proposed a YOLOv4-based lightweight network with better mAP by 3,3 and 4,9 percentage points. Although the above algorithm has achieved lightweight, there is still a relatively large room for improvement in detection accuracy. In the face mask wearing detection in crowded public places such as stations and shopping malls with more complex scenes, difficulties brought by small targets and occluded targets need to be overcome, which requires higher accuracy of the algorithm.

In the research of improved algorithm model, the main focus is on three solutions. The first is to add SPP modules, Cao et al.<sup>(18)</sup> introduces SPP to improve the algorithm. SPP (Spatial Pyramid Pooling)<sup>(19)</sup> can process input images of any size or proportion to generate fixed-scale feature representation, and has good robustness to the size and shape changes of input images. However, the improvement of feature representation and accuracy is not obvious enough.

The second is to add attention mechanisms<sup>(20)</sup> which are derived from the way the human brain processes visual information. When observing a target object, human beings are used to first roughly observe the overall characteristics of the object, and then observe the significant features of the object, find out the symbolic memory points to deepen the impression of the object, and focus on this area to obtain more detailed information. Because of its powerful and effective representation, it has been widely used in deep learning, especially in deep high-performance networks.<sup>(21)</sup>

The last way is improve feature enhancement network<sup>(22)</sup> deep in a neural network's layers, the more abstract elements of the image can be extracted. However, with the deepening of the network layer, after multiple convolution, the feature of the detection target may be weakened and disappeared, and the feature cannot be well characterized. Zhu et al.<sup>(23)</sup> proposed to change the FPN of YOLOv4-tiny into a path convergence network (PAN) as a feature enhancement network. In the feature fusion strategy of PAN, features from different scales are fused with each other and repeatedly enhanced, making full use of the detailed information extracted by the backbone network, which greatly improves the expression ability of the predictive feature layer to the target.

#### *System response speed of mask detection*

The second points are the system response speed. At present, the mainstream practice is to use lightweight human object detection based on yolov3,<sup>(24)</sup> which can more quickly identify whether to wear masks and other issues. But there are still delays. Wang<sup>(25)</sup> introduced a spatial pyramid structure in YOLOv3 that can be used to mask detection tasks. This method led to a slight improvement in FPS and mAP indicators. The algorithm's background is fairly simple, which makes it hard to expand to complex tasks. Also, the real-time response time is not ideal. YOLOv4-tiny, published by Wang et al.<sup>(26)</sup> achieves 42,0 % detection accuracy at 443FPS (on GeForce RTX 2080Ti) real-time detection on the COCO dataset.<sup>(27)</sup> If the accuracy rate can be improved to the standard level on the basis of this algorithm, it will be more suitable for this project. Another reason for limiting reaction time is the transmission speed of the system.<sup>(28)</sup> Even if cloud processing is adopted, transmission speed will affect the upper limit of system response speed.

**Literature review**

*Mask detection*

On March 11, 2020, the WHO announced that COVID-19 has turned into a pandemic. It has affected over 80 000 individuals in 114 different nations.<sup>(29)</sup> To ensure the safety of students on campus, the campus security system should strengthen the monitoring of students wearing masks.

Classification through deep neural networks is regarded as the most accurate method. They can perform better than other techniques in terms of both loss functions and accuracy. Ashwan et al.<sup>(30)</sup> in his paper, he experimented with deep learning neural networks to verify whether to wear a mask. Recorded the precision, the specific data is shown below figure 1.

**Metrics for evaluation the proposed CNN model**

Class	Precision	Recall	F1-measure
Correct	0.99	0.92	0.95
Incorrect	0.96	1.00	0.98
Nomask	0.94	0.97	0.96

**Table 4. The 10-fold cross validation for proposed model**

No of fold	Accuracy of each fold	Model accuracy
1	92.05%	97.59% (+/-2.48%)
2	97.75%	
3	94.85%	
4	99.77%	
5	100.00%	
6	100.00%	
7	99.56%	
8	98.27%	
9	97.69%	
10	95.99%	

Figure 1. Precision

**YOLOv4-tiny**

The YOLO V4-TINY is a lighter variant of the YOLO V4, but it delivers faster response times at the expense of accuracy. The figure shows the network structure figure 2. The feature extraction network is CSPDarknet53-tiny, which consists of two CBLs (Conv2D-BN-Leaky-ReLU) and three CSPblocks. CSPBlock doubles the number of gradient paths by splitting and merging strategies across phases, to reducing the computational cost of the model. First, FPN has two inputs, Feat1 and Feat2. Feat1 is Conv convolved to produce an output P1, and then convolved and upsampled to produce a second output P2. The two outputs are convolved a little to produce two YOLO Head prediction feature layers.

The YOLO v4-tiny FPN module is very simple to use, making it ideal for real-time monitoring. However, due to its oversimplicity, it also has some shortcomings. First, the overall receptive field of the network is not strong, and the feature fusion and extraction are insufficient, which makes it difficult to adapt to the task of detecting small targets and occluded objects in complex scenes.

Based on the above, YOLOv4-tiny, due to its lightweight model, has fast response speed and high relative accuracy. Compared with other algorithms, YOLOV4- TINY is more suitable for small and repetitive scenes such as campus data.

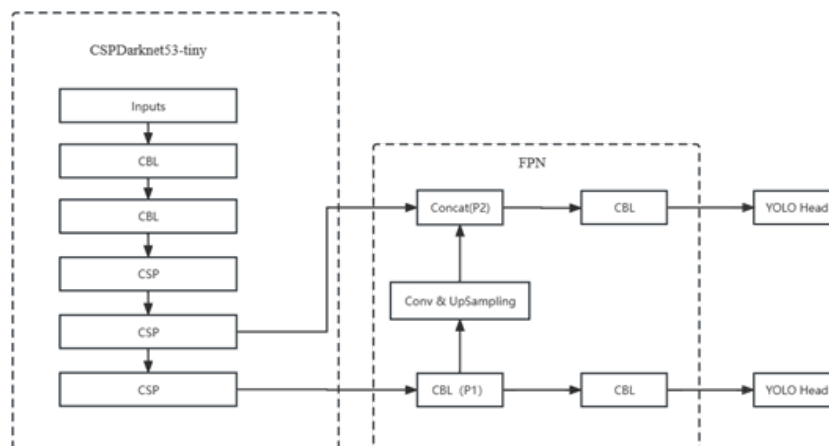


Figure 2. YOLO V4-TINY Network Structure

**Spatial Pyramid Pooling**

Due to the complexity of the classification process, it is usually recommended to use clipping or distortion after the convolution layer has been processed. However, in this way, it is easy for the algorithm to lose important information. Therefore, the Spatial Pyramid Pooling (SPP) module is introduced in this paper, which can pool and fuse the feature layers extracted at multiple scales and obtain the output of fixed size. The specific structure is shown in the figure.

The processing steps are as follows figure 3: the input feature layer is first processed by three convolution processes, and then by three pooling layers for maximum pooling processing. The output of any size can be produced by adjusting the scale, and the use is flexible. The purpose of the SPP module is to increase the network’s receptive field, which enables it to accept and output images of any size. It also has a good degree of robustness to the distortion of objects, extracts the information of the network feature layer more comprehensively and effectively, and learns the features of multi-scale targets.

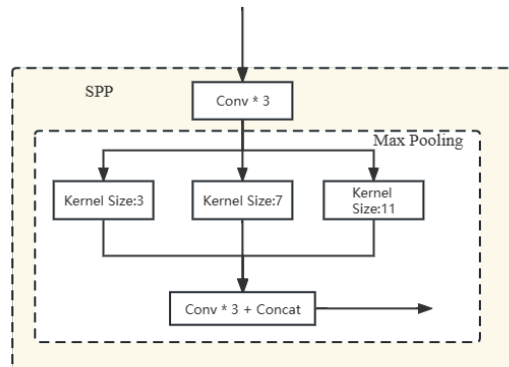


Figure 3. Space pyramid pool

**METHOD**

**Principle of YOLO v4-tiny algorithm**

The YOLO V4-TINY is a lightweight variant of the YOLO v4. It offers a faster response while reducing accuracy. The feature extraction network is CSPDarknet53-tiny, which consists of two CBLs (Conv2D-BN-Leaky-ReLU) and three CSPblocks. CSPBlock doubles the number of gradient paths by splitting and merging strategies across phases, to reducing the computational cost of the model. First, FPN has two inputs, Feat1 and Feat2. Feat1 is Conv convolved to produce an output P1, and then convolved and upsampled to produce a second output P2. The two outputs are convolved a little to produce two YOLO Head prediction feature layers.

**Mask detection -accuracy**

*Improving Space pyramid pool*

Inspired by SPP, this paper optimizes SPP accordingly, and the improved SPP is shown in the figure figure 3. In this case, three pooling layers, each with a step size of 5 × 5, 9 × 9 and 13 × 13, were used to maximize their input features. The three local features of the different scales were also obtained. The network’s detection accuracy is enhanced by stacking the various local features with the global input features. A richer feature table is also created to help improve the network’s performance.

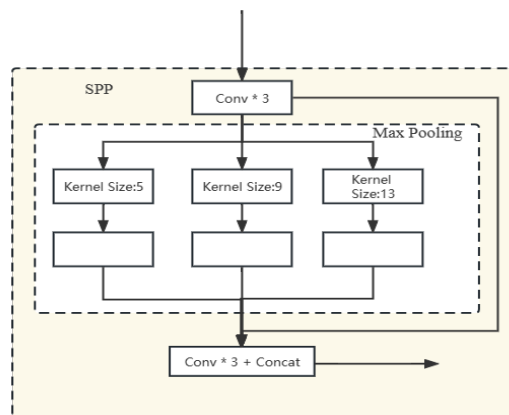


Figure 4. Improved Space pyramid pool

Feature enhancement network PAN

The YOLO V4-tiny’s feature information fusion is carried out by FPN, and this method leads to low receptive field and poor performance when dealing with complex scenes. The paper presents the use of the PAN as the enhancement network. The figure shows the structure of this network.

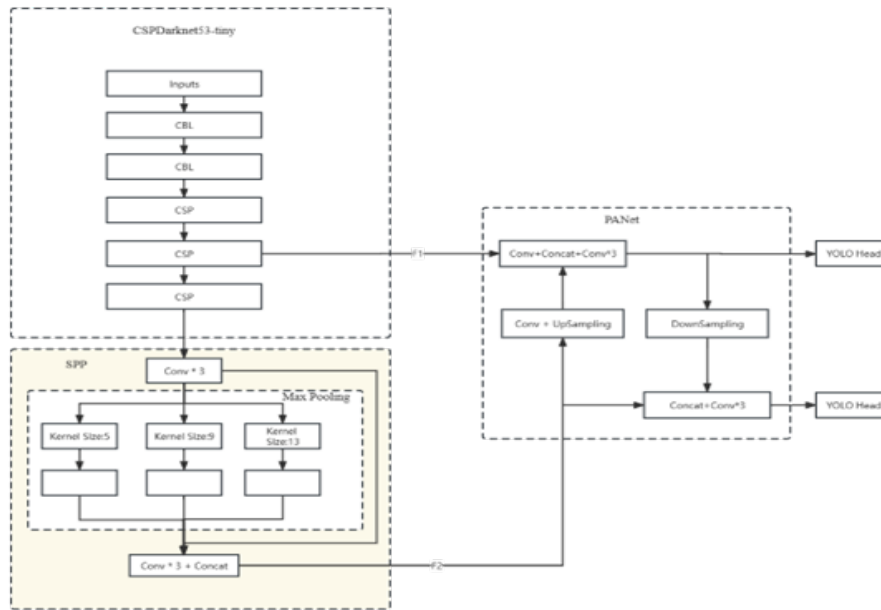


Figure 5. PAN

There are two feature input layers in the figure, F1 is from the backbone network and F2 is from the SPP module. F2, as a bottom-up fusion path input, needs to undergo convolution and Upsampling processing, then is stacked with F1, and is convolved three times again without getting the first output. Similarly, F1, after convolution and Down sampling, is stacked with F2 and convolved three more times to get a second output. In this strategy, because the features from different scales are repeatedly enhanced and fused, the feature information is fully utilized, so that the expression ability of the prediction layer to the target is greatly enhanced.

Evaluation index

The performance evaluation indexes adopted by the training network in this experiment are mainly mAP (Mean Average Precision), Precision (P) and Recall (R), which are calculated by the following formula. In this paper, to detect the target face-mask wearing a mask, TP (True Positive) represents the number of faces wearing a mask correctly identified by the mask wearing detection algorithm as face-mask. The FP stands for False Positive. It shows the number of faces that the algorithm for detecting masks incorrectly identifies as having masks. The value of FN indicates the number of faces whose masks are recognized by the algorithm as the face without masks. The average accuracy of all categories is the average accuracy mAP.

$$Precision = \frac{TP}{TP + FP} \times 100\% \leftarrow$$

$$Recall = \frac{TP}{TP + FN} \leftarrow$$

$$mAP = \frac{\sum_{i=1}^{\epsilon} AP_i}{TP + FP} \times 100\% \leftarrow$$

Figure 6. Algorithm

RESULTS

Mask recognition

A mask recognition system uses a combination of video cameras and cameras to collect images of a person’s face. It then performs various face recognition techniques by automatically detecting and tracking the faces

in the captured images. The process of mask recognition involves defining various key points figure 7 and identifying the mask position in the image or video. It then calculates the position of the point in the matrix. This information can be used to distinguish the mask from the others.



Figure 7. Key Points

After identifying the individual who is not wearing a mask and sending their image to the cloud center, a notification will be sent to the students reminding them of this requirement. This method can minimize the number of attempts to identify people using masks, and it can also help students know if they are not wearing one.

**Experimental Design**

The first experiment is mask recognition, which mainly identifies three conditions. With mask, without mask, and mask wear incorrect. And record the response time and accuracy. A total of ten pictures were used to verify the experiment. Here’s a gallery of images and the results.

After identifying that the mask is not worn or is not correctly worn, the image data of six people need to be sent to the cloud center, and face recognition is carried out, and then SMS reminder, during this period, SMS reminder function is running normally.

*Mask Recognition*

The table compares the FPS and mAP indicators of different target detection algorithms, such as YOLOv4-tiny and the proposed algorithm. For face targets and face targets wearing masks, the R and P values of the proposed algorithm reach 95,32 % and 90,21 %, respectively, which increase by 1,61 % and 4,14 % compared with YOLOv4-tiny, achieving higher detection accuracy. Due to the introduction of SPP enhanced feature extraction network, the network size increases and the speed is slightly lower than YOLOV4-TINY, but it is still faster than other algorithms.



Figure 8. Before recognition

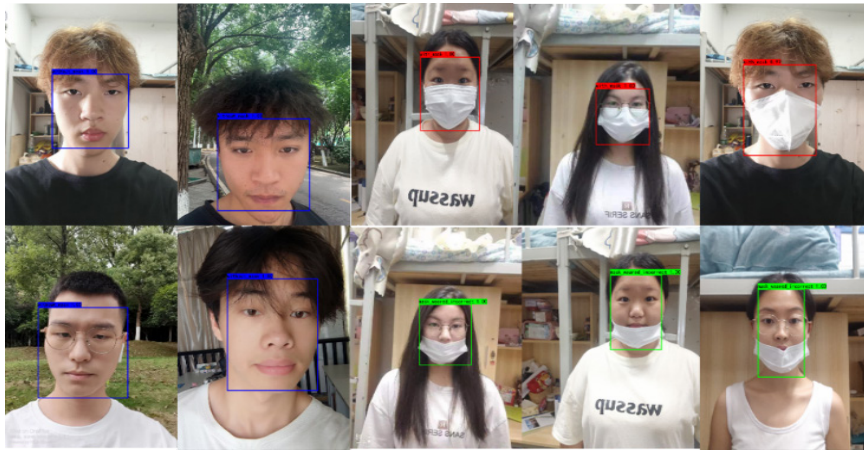


Figure 9. After recognition

**Table 1. Mask recognition Result**

Num	True situation	Precision	Accuracy	Response
1	Without	1,0	Right	Fast
2	Without	1,0	Right	Fast
3	With mask	1,0	Right	Fast
4	With mask	1,0	Right	Fast
5	With mask	0,99	Right	Fast
6	Without	1,0	Right	Fast
7	Without	1,0	Right	Fast
8	Incorrect	1,0	Right	Fast
9	Incorrect	1,0	Right	Fast
10	Incorrect	1,0	Right	Fast

**Table 2. Mask Recognition efficiency**

Algorithm	R (%)	P(%)	mAP(%)	FPS SSD-512
	89,34	85,38	73,19	19,36
YOLOV4	94,25	89,39	95,5	22,5
YOLOV3	92,38	87,02	85,85	17,35
YOLOV4-TINY	93,71	86,07	93,6	112,34
Our Algorithm	95,32	90,21	95,7	98,67

**DISCUSSION**

The experimental results of mask detection show that the precision of the correct wearing of the mask should be improved, but because it does not affect the main function of the system (will not miss the alarm), the impact is not important.

**CONCLUSIONS**

This paper proposes a method that uses a one-stage Deep Learning algorithm to prevent the spread of the disease on campus. By adding and strengthening SPP and PAN modules, the YOLOV4-TINY algorithm is more suitable for this project. Although some speed is sacrificed, the accuracy is improved within the standard range. In the later work, part of the speed will be restored without increasing the hardware requirements while ensuring this accuracy.

**ACKNOWLEDGMENTS**

This work was supported by the Ministry of Higher Education under Fundamental Research Grant Scheme FRGS/1/2018/ICT03/UTM/02/2.



## BIBLIOGRAPHIC REFERENCES

1. Ahmed T. Lung Cancer Detection Using CT Image Based on 3D Convolutional Neural Network. 2020:35-42.
2. Raote N. Campus Safety and Hygiene Detection System using Computer Vision. International Conference on Advances in Computing, Communication, and Control. 2021. doi:10.1109/ICAC353642.2021.9697148.
3. Redmon J. You Only Look Once: Unified Real-Time Object Detection. arXiv:1506.02640 [Cs]. 2016. Available from: <http://arxiv.org/abs/1506.02640>.
4. Liu W. SSD: Single Shot MultiBox Detector. 9905. 2016:21-37.
5. Howard AG. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861 [Cs]. 2017. Available from: <http://arxiv.org/abs/1704.04861>.
6. Rahmatov N. Realtime fire detection using CNN and search space navigation. J Real-Time Image Process. 2021:1331-1340.
7. Zheng YP, Li GY, Li Y. Survey of application of deep learning in image recognition. Comput Eng Appl. 2019;55(12):20-36.
8. Zhou JY, Zhao YM. Application of convolution neural network in image classification and object detection. Comput Eng Appl. 2017;53(13):34-41.
9. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: Single shot multibox detector. In: European conference on computer vision. Springer, Cham; 2016:21-37.
10. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE; 2014:580-587.
11. Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. New York: ACM; 2015:91-99.
12. Ding P, Geng LT, et al. Real-time face mask detection and standard wearing recognition in natural environment. Comput Eng Appl. 2021;57(24):268-275.
13. Cabani A, Hammoudi K, Benhabiles H, Melkemi M. MaskedFace-Net - A dataset of correctly/incorrectly masked face images in the context of COVID-19. Smart Health. 2021;19:100144. doi:10.1016/j.smhl.2020.100144.
14. Wang ZY. Masked face recognition dataset and application. arXiv:2003.09093 [Cs]. 2020.
15. Deng HX. A mask wearing detection method based on migration learning and RetinaNet. Electron Technol Softw Eng. 2020:209-211.
16. Xiao JJ. Face mask detection and wear recognition based on YOLOv3 and YCrCb. Softw Eng. 2020:164-169.
17. Wang B. Mask detection algorithm for improved YOLO lightweight network. Comput Eng Appl. 2020:62-69.
18. Cao XX. Mask wearing real-time detection algorithm based on improved YOLOv4-tiny. J Heilongjiang Univ Technol. 2022.
19. He K, Zhang X, Ren S, Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans Pattern Anal Mach Intell. 2015;37(9):1904-1916. doi:10.1109/TPAMI.2015.2389824.
20. Zhang LW, Liu J, Zhou C. Research on gas station target detection algorithm based on improved Yolov3-Tiny. J Jilin Univ Inf Sci Ed. 2023; doi:10.19292/j.cnki.jdxp.20230707.004.
21. Zhao R, Liu H, Sun Q. Research on Safety Helmet Detection Algorithm Based on Improved YOLOv5s. J Beijing Univ Aeronaut Astronaut. 2021; doi:10.13700/j.bh.1001-5965.2021.0595.

22. Wan Q, Huang Y, Zhang W. Research on depth map restoration algorithm based on layered joint bilateral filtering. *Comput Eng Appl*. 2021.
23. Zhu J, Wei Q, Ma L. Light weight mask detection algorithm based on improved YOLOv4-tiny. *Chin J Liquid Cryst Displays*. 2021.
24. Abed AA, Al-Ali AK, Hameed RA. Real-time multiple face mask and fever detection using YOLOv3 and TensorFlow lite platforms. *Bull Electr Eng Inform*. 2023;12:922-929. doi:10.11591/eei.v12i2.4227.
25. Wang YH. Mask wearing detection algorithm based on improved YOLOv3 in complex scenes. *Comput Eng*. 2020:12-22.
26. Wang CY, Mark Liao HW, Wu YH, Chen PY, Hsieh JW, Yeh IH. Scaled-YOLOv4: scaling cross stage partial network. arXiv:2011.08036 [Cs]. 2020.
27. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: common objects in context. In: 13th European Conference on Computer Vision. Zurich: Springer; 2014:740-755.
28. Ding Y, Yang H, Yao X, Jiang P. JMDC: A joint model and data compression system for deep neural networks collaborative computing in edge-cloud networks. *J Parallel Distrib Comput*. 2022;173:83-93. doi:10.1016/j.jpdc.2022.11.008.
29. World Health Organization. Retrievable at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed May 2, 2020.
30. Ashwan AA, Khan RZ, Uddin M. Deep learning-based masked face recognition in the era of the COVID-19 pandemic. *Int J Electr Comput Eng*. 2023;13(2):1550-1559. doi:10.11591/ijece.v13i2.pp1550-1559.

#### **FINANCING**

“The authors did not receive financing for the development of this research”.

#### **CONFLICT OF INTEREST**

“The authors declare that there is no conflict of interest”.

#### **AUTHOR CONTRIBUTION**

*Conceptualization:* Baitong Zhong.

*Data curation:* Johan Bin Mohamad Sharif.

*Formal analysis:* Sah Salam.

*Research:* Chengke Ran.

*Methodology:* Zhuoxi Chen.

*Drafting - original draft:* Zhuoxi Chen.

*Writing - proofreading and editing:* Zhuoxi Chen.