Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier

Truong Quang Vinh
Ho Chi Minh City University of Technology (HCMUT)
Vietnam National University – Ho Chi Minh (VNU-HCM)
Ho Chi Minh, Vietnam
tqvinh@hcmut.edu.vn

Nguyen Tran Ngoc Anh
Ho Chi Minh City University of Technology (HCMUT)
Vietnam National University – Ho Chi Minh (VNU-HCM)
Ho Chi Minh, Vietnam
anh.nguyenk2017@hcmut.edu.vn

2020 International Conference on Advanced Computing and Applications (ACOMP) | 978-1-7281-8167-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ACOMP50827.2020.00029

***Abstract*—During the pandemic of Covid-19, wearing face mask in some factories, departments, or working offices is required. This paper presents a real-time face mask detector which can alarm when detecting a person without wearing a face mask. Moreover, the system can recognize the person who wears a face mask incorrectly, or wear other things except a face mask. The proposed algorithm for face mask detection in this system utilizes Haar cascade classifier to detect the face and YOLOv3 algorithm to detect the mask. The whole system has been built and demonstrated in a practical application for checking people wearing face mask at the office entrance. The experiment result shows that the accuracy of the system can achieve up to 90.1%.**

Keywords—face mask, YOLOv3, Haar cascade classifier, deep learning, computer vision

I. INTRODUCTION

The use of face masks is very necessary to prevent and limit the spread of certain respiratory viral diseases, including COVID-19 [1]. Face mask can be used to protect either healthy people or prevent infection by infected persons [2]. However, wearing face mask correctly is very important to reduce risks of contamination. The World Health Organization recommends to use face mask at crowded spaces such as station, office, school, etc. In order to check whether a person wears a face mask or not, we need a computer vision system that is able to perform this type of detection.

Face mask detection is a particular problem of general object detection in the research field of computer vision. Applications of object detection can be seen in many areas such as intelligent control system, surveillance, smart home, autonomous driving, etc. Traditional object detectors are based on Haar feature extractors [3], support vector machine (SVM) [4], or Bayesian network [5], etc. Recently, object detection using deep learning are widely used thanks to the excellent performance. YOLOv3 is currently is one of emerging algorithms which provide high performance in detection and classification problems [6-8].

In this paper, we proposed a real-time face mask detection system using YOLOv3 algorithm. In order to increase the processing time and accuracy of the detector, we employ Haar cascade detector to detect the face region in the input images, and then put the region of interest (ROI) into the YOLOv3 to detect the face mask. The deep learning model has been trained with dataset of 7,000 samples. Finally, we build up a whole system to demonstrate an application in which people are checked whether they are wearing face mask or not at the entrance door.

The remain part of this paper is organized as follows. The Section II give the review of the background on YOLO networks. Section III presents the proposed algorithm for face mask detection. Section IV describes experimental results. Finally, Section V gives a conclusion of the paper.

II. BACHGROUND

YOLO is a state-of-art algorithm for real-time object detection by Joseph Redmon et al. The authors released three versions: YOLOv1 in 2016 [6], YOLOv2 in 2017 [7], and YOLOv3 in 2018 [8]. YOLO algorithm has been proved to be effective in many object detection applications such as vehicle detection [9], aerial target detection [10], pedestrian detection [11], etc. The feature extractor of the YOLOv3 contains 53 convolutional layers, and thus it is named Darknet-53. This is a hybrid approach between the previous network v2 and v1. The detail description of Darknet-53 is shown in Table I.

TABLE I. THE DARKNET-53 MODEL

Repeat	Type	Filters	Size	Output
	Convolutional	32	3×3	256×256
	Convolutional	64	3×3/2	128×128
	Convolutional	32	1×1	
×1	Convolutional	64	3×3	
	Residual			128×128
	Convolutional	128	3×3/2	64×64
	Convolutional	64	1×1	
×2	Convolutional	128	3×3	
	Residual			64×64
	Convolutional	256	3×3/2	32×32
	Convolutional	128	1×1	
×8	Convolutional	256	3×3	
	Residual			32×32
	Convolutional	512	3×3/2	16×16
	Convolutional	256	1×1	
×8	Convolutional	512	3×3	
	Residual			16×16
	Convolutional	1024		
	Convolutional	512		
×4	Convolutional	1024		
	Residual			16×16
	Convolutional	1024	3×3/2	8x8
	Avgpool		Global	
	Connected		1000	
	Softmax			
53 convolutional layers				

DOI 10.1109/ACOMP50827.2020.00029

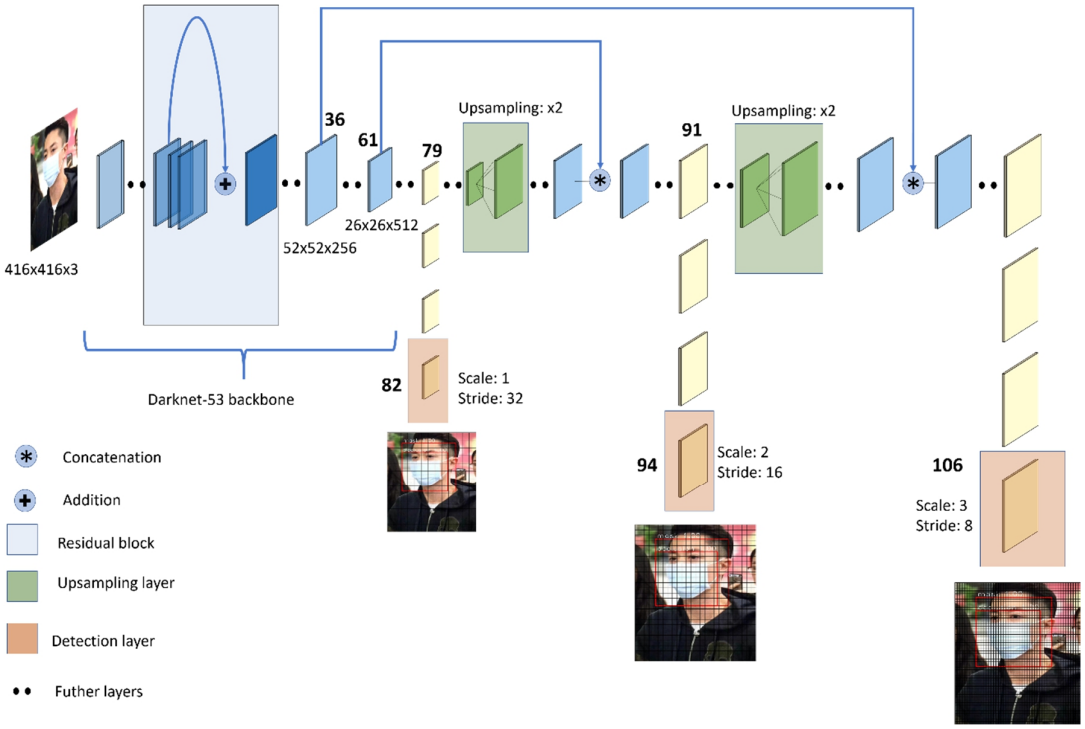

Fig. 1. YOLOv3 network architecture

The YOLO-v3 algorithm uses a Darknet-53 as a backbone for feature detection. The full YOLOv3 consists of 106 layers, including convolutional, residual, and up-sampling layers. The architecture of YOLOv3 network is shown in Fig. 1. The YOLO algorithm uses only one forward propagation pass through to detect objects on a single input image. YOLOv3 is better in detecting small objects thanks to the up-sampled layers which can preserve the fine-grained features of the small objects.

In this paper, we employ YOLOv3 to take its advantages for high accuracy and real-time processing.

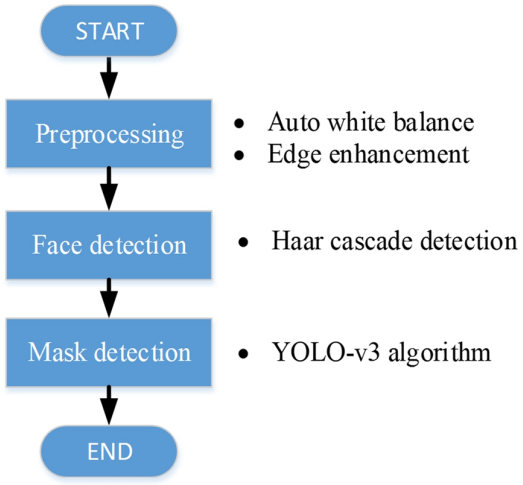

Fig. 2. The flowchart of the proposed algorithm

III. Proposed Algorithm

The proposed algorithm for face mask detection consists of three steps: preprocessing, face detection, mask detection as shown in Fig.2.

The preprocessing step is to enhance the input image quality by using auto white balance, and edge enhancement using unsharp filter. The auto white balance [12] is to ensure the color consistency of the input image frames in a various range of color temperature. The unsharp filter [13] is then applied to enhance the edges in the input images. Some researchers have proved that image enhancement can improve the accuracy of the object detection about 2-5% [14-15].

The face detection step is to detect the face region. We utilize the Haar cascade classifier proposed by Viola-Jones [3] to detect the face region. This classifier performs feature extraction by Haar Wavelet technique with 24x24 window size, uses AdaBoost to remove redundant features, and applies cascade classifiers to detect objects. The detected face regions by Haar cascade classifier is then put into the input data of the YOLOv3 algorithm to detect the regions of face masks.

The last step is to recognize whether the person is wearing the mask or not by using YOLOv3 algorithm. We configure the predicting process at three scales: 13x13, 26x26, and 52x52. The first detection is made by layer 82nd. The output is a feature map which has the size 13x13x27. The second detection is made by layer 94th with the size is 26x26x27. The last detection is made by layer 106th, with the size is 52x52x27. We apply the loss function given by (1).

$$\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2\right]$$
$$+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\left(\sqrt{w}_i-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h}_i-\sqrt{\hat{h}_i}\right)^2\right]$$
$$+\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{noobj}\left(C_i-\hat{C}_i\right)^2 \quad (1)$$
$$+\sum_{i=0}^{S^2}1_{ij}^{obj}\sum_{c\,\epsilon\,classes}(p_i-\hat{p}_i)^2$$

IV. Experimental Results

A. Dataset

For the experiment, we use the dataset MAFA [16], which consists of 35,806 masked faces with a minimum size of 32x32. The face in this dataset have different orientation and occlusion degree. We select 7000 images which contain frontal faces from MAFA. The dataset is divided into 3 parts for training and validation and test set with 5000, 1000, and 1000 images, respectively. Fig. 1 and Fig. 3 show example images in the MAFA dataset.

Fig. 3. Faces with masks in MAFA dataset

Fig. 4. Faces without mask or wearing masks incorrectly.

B. Training

The training process is divided into 2 phases. At the first phase, we use batch size 16, learning rate 0.001, and learning rate schedule factor 0.1. after 3 epochs with improvement (new_learning rate = 0.1*old_learning rate), the batch size is set to 16. This stage helps converge to stable loss field. After 10 epochs, if validation loss cannot be improved, this phase will stop and the second phase will start. At the second phase, batch size is 8, learning rate is 0.0001, and learning rate schedule factor is 0.1. After 10 epochs with no improvements, this phase will stop.

The training process has been done in 75 epochs. The loss value indicates the performance of the model. Fig. 5 shows the training and validation loss. The lower is better result. According to the results, the training loss and validation can achieve down to 5.15% and 5.27, respectively.

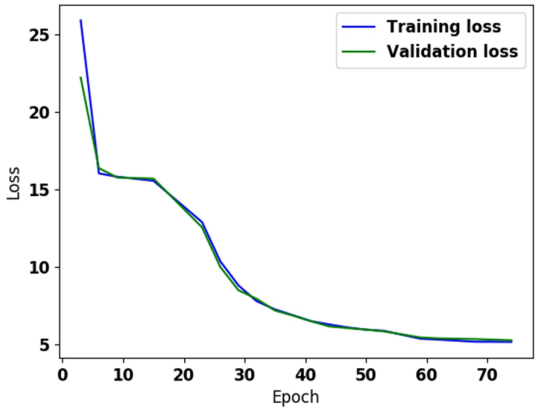

Fig. 5. The training and validation loss.

C. Performance

We evaluate the performance of the proposed algorithm by two metrics: precision and recall, given by (2). Precision metric indicates the accuracy of the algorithm based on classification result. The recall metric presents the ability to find all relevant objects in a dataset.

$$Precision=\frac{True\ Positive}{True\ Positive+False\ Positive} \quad (2)$$
$$Recall=\frac{True\ Positive}{True\ Positive+False\ Negative}$$

Table II shows the performance metrics of the proposed algorithm and another method using locally linear embedding convolutional neural network (LLE-CNN) [16]. The accuracy of the proposed algorithm can achieve up to 90.1%, slightly higher the other method. Moreover, the proposed algorithm utilizes the advantage of YOLOv3 model which can run faster and detect objects at smaller scales.

TABLE II. Comparison between the proposed algorithm and others.

	LLE-CNN [16]	Proposed
Precision	82.8%	83%
Recall	89.0%	90.1%

D. Demonstration

In order to demonstrate the proposed algorithm in a real application, we build a whole system for real-time face mask detection at the office entrance as shown in Fig. 6, is installed at the entrance or gate of offices, buildings, or factories. The system consists of a computer, camera, and a speaker for warning messages. In order to detect person at a far distance, we utilize a 10x optical zoom to magnify the input image frames. Fig. 7 shows the demonstration of the face mask detection system. The video demo has been recorded and uploaded at Youtube https://youtu.be/HdhFdsXkFHQ. The experimental result shows that the system can detect multiple faces and recognize whether people wear face masks or not.

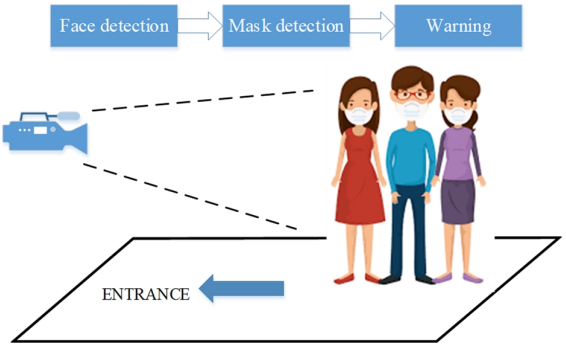

Fig. 6. The system of facemask detection

Fig. 7. Demonstration of facemask detection system.

V. Conclusion

This paper has presented a face mask detection system which uses the YOLOv3 algorithm and Haar cascade classifier. The proposed algorithm employs image enhancement technique to improve the accuracy of the system. Thanks to the advantages of the YOLOv3 network, the system can work in real-time with 30fps. This system can be applied effectively for practical applications to reduce the spread of infectious diseases such as Covid-19.

References

[1] World Health Organization (WHO), “Coronavirus disease 2019 (covid-19): situation report, 96”, 2020

[2] World Health Organization (WHO), "Advice on the use of masks in the context of COVID-19", June 2020.

[3] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001.

[4] Wei Zhu, Qian-Liang Fu, Jun-Qi Bai, “The real-time object detection algorithm based on ORBP and cascade SVM”, 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2016.

[5] Hong Huo, Tao Fang, “Integration of bottom-up and top-down cues in Bayesian network for object detection”, Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC), 2013.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[7] J. Redmon, and Ali Farhadi. "YOLO9000: Better, Faster, Stronger", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[8] J. Redmon, Ali Farhadi, “YOLOv3: An Incremental Improvement”, arXiv preprint arXiv:1804.02767, 2018.

[9] Yunxiang Liu ; Guoqing Zhang ; Yuanyuan Zhang, “Vehicle detection method based on ADE-YOLOV3 algorithm”, International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2019.

[10] Lecheng Ouyang ; Huali Wang, “Aerial Target Detection Based on the Improved YOLOv3 Algorithm”, 2019 6th International Conference on Systems and Informatics (ICSAI), 2019.

[11] Changhao Piao ; Xianhao Wang ; Mingjie Liu, “Pedestrian Detection Using Optimized YOLOv3 in UAV Scenario”, 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS), 2019.

[12] Jintao Jiang, Ming Yang, XinPo Wang, Zhengguan Wu, “Auto White Balance Algorithm Based on Digital Camera”, 2011 International Conference on Internet Technology and Applications, 2011.

[13] R. C. Gonzalez & R. E. Woods, “Digital Image Processing”, AddisonWesley, Publishing Company, Inc., 1992.

[14] Liu, X., Pedersen, M., Charrier, C., Bours, P., “Can image quality enhancement methods improve the performance of biometric systems for degraded face images?”, Colour and Visual Computing Symposium (CVCS), 2018.

[15] M. O. Oloyede, G. P. Hancke and H. C. Myburgh, "Improving face recognition systems using a new image enhancement technique hybrid features and the convolutional neural network", IEEE Access, vol. 6, pp. 75181-75191, 2018.

[16] Shiming Ge, Jia Li, Qiting Ye1, Zhao Luo, “Detecting Masked Faces in the Wild with LLE-CNNs”, IEEE Conference on Computer Vision and Pattern Recognition, 2017.