Human Face Mask Detection Based on Deep Learning Using YOLOv7+CBAM

Xinyi Gao, Minh Nguyen and Wei Qi Yan
Auckland University of Technology, 1010 New Zealand

ABSTRACT

COVID-19 and its variants have affected millions of people around the world. Wearing a mask is an effective way to reduce the spread of the epidemic. While wearing masks is a proven strategy to mitigate the spread, monitoring compliance remains a challenge. In this paper, we propose a mask detection method based on deep learning and Convolutional Block Attention Module (CBAM). In this paper, we extract representative features from input images through supervised learning. In order to improve the recognition accuracy under limited computing resources. We choose YOLOv7 network model and incorporate CBAM into its network structure. Compared with the original version of YOLOv7, our proposed network model improves the mean Average Precision (mAP) up to 0.3% in face mask detection process. Meanwhile, our method improves the detection speed of each frame 73ms. These advancements have significant implications for real-time, large-scale monitoring systems, thereby contributing to public health and safety.

Keywords: Face mask detection, Convolutional block attention module, Deep learning, YOLOv7

INTRODUCTION

Currently, there are variants of COVID-19 virus (Ciotti et al., 2020) in the throes of the pandemic. In 2022, Monkeypox virus (Rizk et al., 2022) began to appear globally. The current known routes of transmission for both viruses are dropped alot. Wearing a mask is an effective way to stop the spread of the viruses. Wearing masks is required in closed and public spaces. Manually testing whether a mask is worn requires a lot of costs and increases the risk of testing personnel being infected. The mask monitoring through digital cameras can reduce the chance of inspectors being infected (Yan, 2019). In recent years, researchers have proposed several effective mask classification and monitoring algorithms that can detect faces with and without masks (Balaji et al., 2021). In this paper, we group mask detection into three cases: Wearing a mask correctly, wearing a mask incorrectly, and not wearing a mask.

Visual object detection (Zou et al., 2023) has always been an enduring research direction in the field of computer vision. The conventional object detection algorithms consist of three stages. Object proposals are firstly generated in the input image. The features in each proposal box are then extracted. Finally, different visual features were extracted by designing a classifier. However, the algorithms were not ideal in terms of accuracy and speed in visual object recognition. In recent years, visual object detection algorithms based on deep learning (Shen et al., 2018) have performed well in terms of accuracy and speed. Among them, You Only Look Once (YOLO) series of algorithms are Superior to others in visual object detection (Redmon et al., 2016).

In this paper, we propose a deep learning-based face mask object detection algorithm CBAM-YOLOv7. This model is based on the existing YOLOv7 model (Wang et al., 2023) with the addition of CBAM (Woo et al., 2018). YOLO is a one-stage detector model based on convolutional neural networks. It applies a neural network to the entire image. The network model firstly segments an image into regions and then predicts the bounding box of each region. CBAM consists of two modules: The channel attention module

(CAM) (Huang et al., 2020) and the spatial attention module (SAM) (Wang et al., 2019). CAM can make the network pay more attention to meaningful ground truth regions. On the other hand, SAM allows the network to focus on context-rich locations throughout the image (Yin et al., 2023). Through this supervised learning method, the accuracy of mask recognition is effectively improved.

The rest of the paper is structured as follows. The second part of this paper introduces the related work on face mask detection. In the third part, we introduce the datasets, methods and models used. In Section 4, we compare and analyse our experimental results. Finally, we summarize our work in Section 5.

LITERATURE REVIEW

Deep learning (Yan, 2021) has now been widely used in the field of computer vision. Especially in the recognition of various images, deep learning (Lu et al., 2021) (Liang et al., 2022) is playing an increasingly important role. For mask recognition, a consortium of deep learning (Wang & Yan, 2022) (Lu et al., 2018) models are widely used a number of representative methods are Faster R-CNN (Lin et al., 2020), InceptionV3 (Jignesh Chowdary et al., 2020), MobileNet (Venkateswarlu et al., 2020), YOLO, etc. Among them, the YOLO series are taken account for a large proportion which is the current mainstream.

In recent years, affected by the epidemic, almost all countries require people to wear masks frequently during travel to prevent the spread of the virus. To detect those who are not wearing masks, Samuel Ady Sanjaya et al. proposed a mask recognition model based on MobileNetV2. MobileNetV2 is a convolutional neural network (CNN) based method. The model firstly detects the video frame by frame. When there is a face in the detection process, the trained MobileNetV2-based model (Sanjaya & Adi Rakhmawan, 2020) is employed for recognition. Determine whether people wear masks by identifying face image frames. The final experimental results show that the model has a detection accuracy of 96.85% in distinguishing between people who wear masks and people who do not wear masks. And help the government to count these data more conveniently.

To identify people who are not wearing masks in public places, a transfer learning-based InceptionV3 recognition model is proposed. In this method, the number of samples that can be utilized is limited. In order to obtain better experimental results, the author proposed to solve the problem of limited data availability through image enhancement technology. Pattern classification is then performed by using the modified InceptionV3 model. The improved model removes the last layer of original InceptionV3 model and adds 5 average pooling layers with a pool size of 5×5 to the network. The model improvement is very effective, and the proposed transfer learning model attains up to 100% accuracy in testing (Jignesh Chowdary et al., 2020).

Since YOLO was proposed, it has been rapidly developed and applied to the field of visual object detection and recognition. YOLO is also suitable for mask recognition. Loy et al. proposed a model based on YOLOv2 and ResNet-50 for detecting medical masks. The model is composed of a feature extraction network and a detection network. The feature extraction network consists of ResNet-50 and the detection network consists of YOLOv2 (Loey et al., 2021). In addition, Intersection over Union (IoU) was harnessed to estimate anchor boxes for testing so as to increase the diversity of the dataset through data augmentation. The final experimental results show that the average recognition accuracy using the improved model is as high as 81%.

Real-time face mask detection is equally important. To this end, Jiang et al. proposed a YOLOv3-based squeeze and excitation YOLOv3 (SE-YOLOv3) object detection model. In this model, YOLOv3 was upgraded by adding an attention mechanism. The specific improvement is the addition of Squeeze and Excitation blocks between the convolutional layers of Darknet53. Generalized Intersection over Union (GIoU) loss was taken into account. Additionally, a larger dataset of masked faces was created. The

experimental results prove that the proposed method can not only locate the face in real time, but also evaluate whether the mask is worn correctly. In terms of accuracy, the new model improved mAP by 6.7% than the base YOLOv3 (Jiang et al., 2021).

The same was employed to use YOLOv3-based model for mask recognition. A different approach was proposed. A novel mask recognition framework, namely, FMD-Yolo (Wu et al., 2022) was proposed. In this framework, Im-Res2Net-101 with deep residual network was propounded as the main feature extractor. The features were fully fused by using enhanced path aggregation network. The IoU loss and IoU-aware were also introduced based on the YOLOv3 loss function to determine the performance of the model. The final experimental results showed that FMD-Yolo achieves the best accuracies of 92.0% and 88.4% based on the two datasets. They outperform the final results of other advanced detectors.

In order to compare the difference in recognition accuracy of different models. Singh et al. took use of YOLOv3 and Faster R-CNN for face mask recognition respectively. They created one dataset to train both models and compared the results of the two models. In the experimental results, both models performed well. But the accuracy of Faster R-CNN model is slightly better. In addition, a new method was proposed to generate the bounding boxes of different colours around a face based on whether a mask is worn or not, recording the proportion of people wearing masks every day (Singh et al., 2021).

In order to solve the influence of complex environment on the accuracy of mask recognition. An improved model based on YOLOv4 is proposed. The model can not only identify masks, but also detect whether the wearing of masks is standard. Firstly, CSPDarkNet53 is added to the backbone feature extraction network of YOLOv4 (Yu & Zhang, 2021). Then, an adaptive image scaling algorithm was considered at the algorithm level. The addition of PANet makes the semantic information of the feature layer more. These improvements effectively reduced the computational cost and amount of computation of the network, and improve the learning ability of the model. In order to verify the effect of the model, a lot of comparisons were conducted with other models by using the improved model. The results show that the improved mask recognition mAP reaches up to 98.3%, which is more accurate than the existing algorithms.

Yang et al. proposed a mask recognition method with an interactive interface based on YOLOv5 (Yang et al., 2020). In this method, they divided the entire recognition system into four parts. Three filters are employed to increase the resolution of the input image. A model was designed to extract the corresponding information. The information was classified by using the YOLOv5 model. After the recognition result was generated, the corresponding information was input into the interactive interface. The final accuracy of the experiments is 97.9%. This model was also compared with other classic models, such as SSD, and the results are all due to the classic model.

Novelty of This Work: Unlike the existing models, our research work incorporates the Convolutional Block Attention Module (CBAM) into the YOLOv7 framework. This innovative approach not only improves the mAP by 0.3% but also enhances the detection speed, makes it highly applicable for real-time, large-scale monitoring systems.

METHODOLOGY

In this paper, we make modifications on the network model of YOLOv7 and add CBAM to ELAN. We train the model through three loss functions: Coordinate loss, object confidence loss, and classification loss.

Dataset

In this paper, the dataset we take into account is the Face Mask Detection Dataset. The dataset contains a total of 853 images. We split it into training set and test set with the ratio of 80% and 20%. Since the model is the YOLO series, the dataset is re-labelled especially for YOLO models. We still set the label to three

classes, namely wearing a mask, wearing a mask incorrectly, and not wearing a mask. There are a total of 4,072 labels in the dataset, including 3,232 labels with masks, 123 labels with incorrect masks, and 717 labels without masks.

The mask images in this dataset all have complex backgrounds, which makes this dataset very suitable for this project. The addition of CBAM can effectively reduce the influence of background on the recognition process. We put 683 images into the training set and 170 images into the test set respectively. The exemplar images of the dataset are shown in Fig. 1.

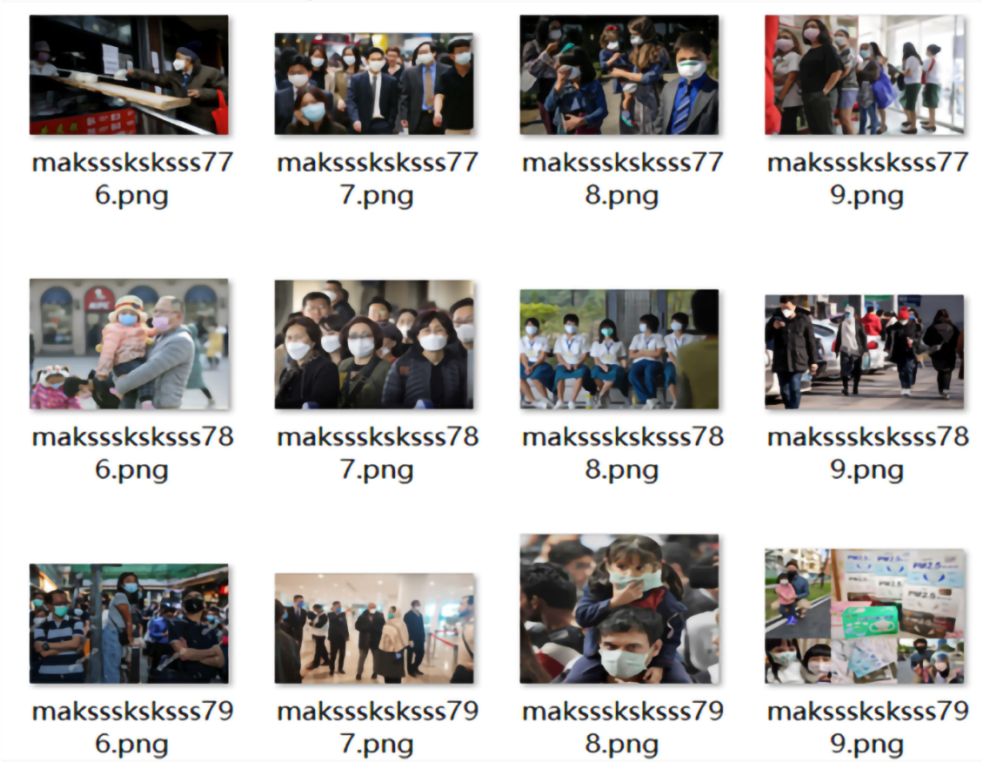

Figure 1. An example of Face Mask Detection Dataset.

Network structure

The main body of our network structure is similar to YOLOv7. We modify the ELAN in YOLOv7 to the CBAM-ELAN module. The proposed YOLOv7+CBAM model is anchored on the YOLOv7 architecture, which is an extension of the YOLO family known for real-time object detection. The YOLOv7 architecture is characterized by its convolutional layers, pooling layers, and skip connections that form the backbone of the network. The Convolutional Block Attention Module (CBAM) is integrated into our YOLOv7 framework at specific stages. CBAM employs both channel and spatial attention to refine the feature maps, thus enhancing the model's ability to focus on salient features.

The YOLOv7+CBAM architecture is formulated as follows:

Initial Layers: The model commences with a series of convolutional layers configured with various kernel sizes and strides, aiming to extract rudimentary features from the input images.

Intermediate Layers and Attention Mechanism: After the initial convolutional blocks, CBAM modules are inserted to refine the generated feature maps. Specifically, these modules are placed after convolutional blocks with 1×1, 3×3 and 5×5 kernels to maximize their impact.

Advanced Blocks: The model incorporates SPPCSPC blocks in the detection head, which combine spatial pyramid pooling and concatenated skip pathways to augment the receptive field and facilitate feature fusion.

Concatenation and Upsampling: Feature maps are upsampled and concatenated at various layers, enhancing the spatial resolution and richness of the feature representation.

Output Layer: The final layer employs a softmax function for class probabilities and a sigmoid activation function for bounding box coordinates. The complete structure is shown in Figure 2.

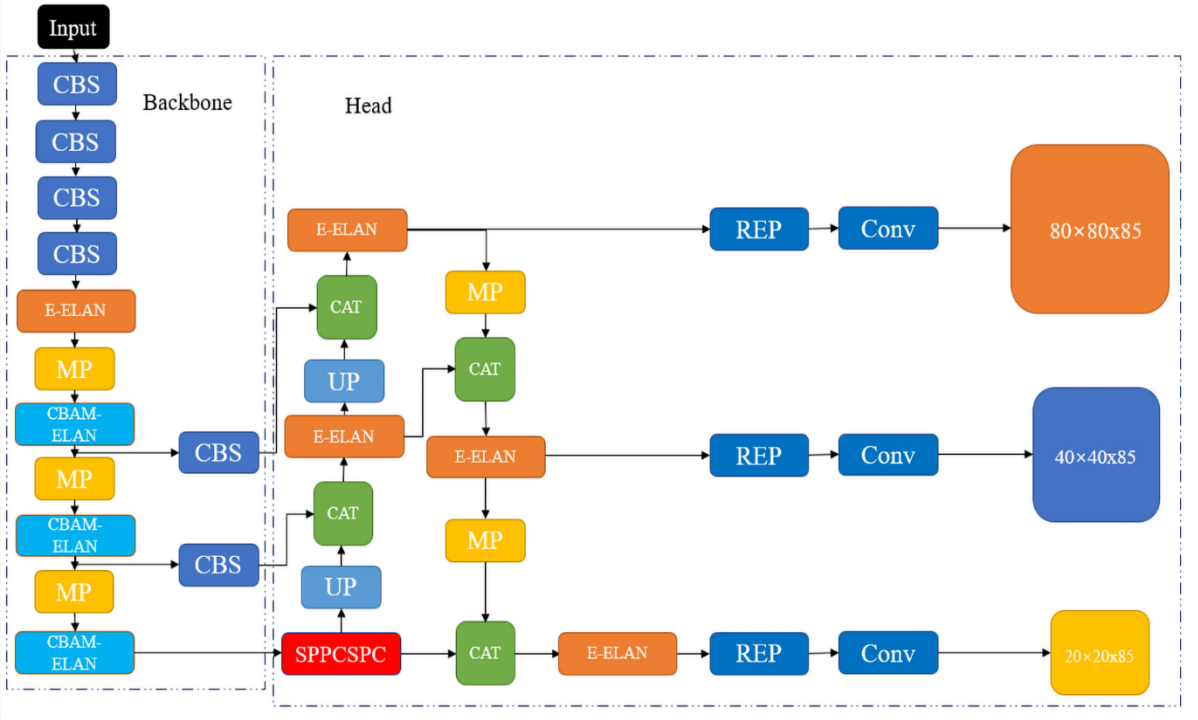

Figure 2. YOLOv7+CBAM model structure.

The ELAN module is an efficient network structure, which enables the network to learn more features and has stronger robustness by controlling the shortest and longest gradient paths. ELAN has two branches. The first branch is to change the number of channels through a 1×1 convolution. The second branch is more complicated. It firstly passes through a 1x1 convolution module to change the number of channels, then goes through four 3×3 convolution modules for feature extraction. Finally, the four features are superimposed to obtain the final feature extraction result.

The overall process of CBAM is divided into two parts. Firstly, the global max pooling and mean pooling are performed on the input. The pooled two one-dimensional vectors are sent to the fully connected layer for operations and added to obtain a one-dimensional channel attention. The next step is to multiply the channel attention with the input elements to obtain the channel attention adjusted features. The features obtained in the first part are globally max-pooled and mean-pooled according to the space. The 2D vectors produced by pooling are concatenated, then rolled and manipulated. The resulting 2D spatial attention is then multiplied by the previously obtained features to complete the process. The addition of CBAM+ELAN to the model can effectively improve the accuracy of target detection.

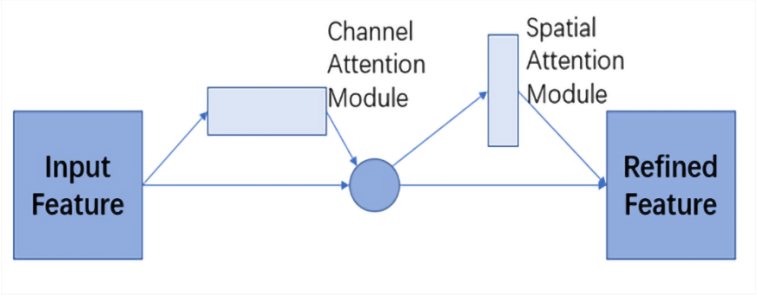

Figure 3. The CBAM module.

The model in this paper is the CBS, CBAM-ELAN and Maxpooling (MP) modules (Christlein et al., 2019). Among them, the CBS module is widely used in YOLOv5 and YOLOv7. The structure of CBS is Convolution + Batch Normalization + SiLU. Our model also uses the CBS module, so the activation function in this paper is SiLU.

$$SiLU(x) = x \cdot sigmoid(x) \tag{1}$$

Our proposed CBAM-ELAN module is a CBAM-based CNN (Liu et al., 2018) architecture. CBAM is an attention module for convolutional neural networks. Since CBAM is a lightweight general-purpose module, it can be seamlessly integrated into convolutional neural networks. The CBAM-ELAN module consists of 6 convolutional neural networks and one convolutional neural network with CBAM.

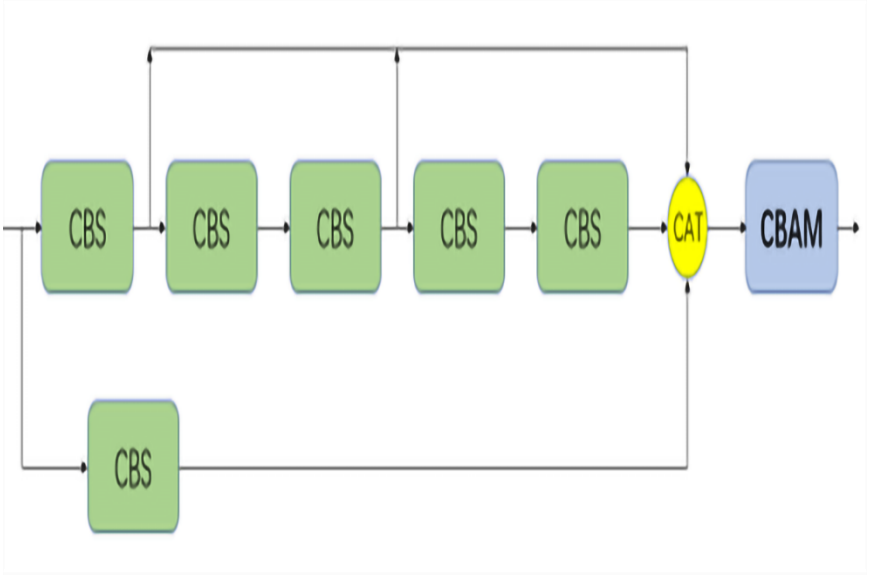

Figure 4. The CBAM-ELAN structure.

Method

In this paper, our method obtains 9 anchor boxes arranged from small to large based on the ground truth (GT) boxes in the training set through *k*-means clustering algorithm. Then, we match each GT box with 9 anchor boxes. According to the centre position of the GT box, the two nearest neighbour grids were also employed as the prediction network. We get the ten largest IOU results with the current GT box and add these ten results. We calculate each GT and candidate anchor loss according to the loss function and keep the smallest *k* loss functions. Finally, the same anchor box is assigned to multiple GTs. Through this method, we obtain better results of human face mask recognition.

RESULT ANALYSIS

We are use of the Mask Dataset, which includes 853 images. The mask images in this dataset all have complex backgrounds, which makes this dataset very suitable for this project. The addition of CBAM can effectively reduce the influence of background on the recognition process. We put 683 images into the training set and 170 images into the test set respectively.

Table 1. Training environment parameters

GPU Name	Processor	CUDA Version
NVIDIA GeForce RTX 3060	Intel Core i7-10700F	CUDA 11.2

We firstly adjust the pixel size of the image to 640×640. In the model training, we are use of three loss functions: Coordinate loss, object confidence loss (GT is the ordinary IoU in the training phase), and classification loss function. The trained model can distinguish whether the mask is correctly worn or not. The experimental results are shown in Figure 5.

In order to more intuitively reflect the performance of object detection, we are use of the performance indicators: Precision (P), Intersection over Union (IoU) (Zhou et al., 2019), F1 score, average precision (AP), and mean average precision (mAP) to test the results. IoU is a measure of the accuracy of detecting corresponding objects in a specific dataset. A standard IoU is a simple measure, as long as the task of getting a bounding box in the output can be measured by IoU. Precision represents the ratio of the number of correct predictions to the number of all predictions. The area enclosed by the PR curve is AP. The larger the area, the better the object detection effect.

$$Precision = \frac{Ture\ Positive}{Ture\ Positives + False\ Positives} \tag{2}$$

$$Recall = \frac{Ture\ Positives}{Ture\ Positives + False\ Negatives} \tag{3}$$

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{4}$$

The IoU obtained by our method is slightly higher than that of YOLOv7. The average IoU of the mask-wearing increases by 0.03. The average IoU without a mask improves up to 0.02. The average IoU for not wearing a mask correctly boosts up to 0.02.

Table 2. Comparison of IoU and F1 score

Method	***IoU***	**F1 score**
CBAM-YOLOv7	0.89	0.92
YOLOv7	0.87	0.92

F1 score refers to the weighted average of precision and recall. F1 score values range from 0 to 1.0, with 1.0 being the highest precision. The F1 score of our method is as same as that of YOLOv7.

$$F1_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

AP is the average of the Precision values on the PR curve. For the PR curve, we are use of the integral for the calculation. We judge the size of AP by comparing the size of the area enclosed by the PR curve. Generally, the larger the area, the larger the AP value, and the better the object detection effect. By superimposing the two images, we found that the PR curves obtained by our method occupy a larger area.

To further determine the effectiveness of the CBAM-YOLOv7 algorithm in the task of face mask detection, we compare the algorithm with YOLOv7, YOLOv5, and YOLOX. All tests were based on the Mask Dataset. As shown in the table, CBAM-YOLOv7 outperforms several other algorithms in mAP. There is also a significant improvement in the recognition speed.

Table 3. Performance comparison of CBAM-YOLOv7, YOLOv7, YOLOv5 and YOLOX on the Mask Dataset.

Algorithm	***mAP@.5***	***Detection time spent per frame(ms)***
CBAM-YOLOv7	0.954	515ms
YOLOv7	0.951	588ms
YOLOv5	0.824	798ms
YOLOX	0.861	810ms

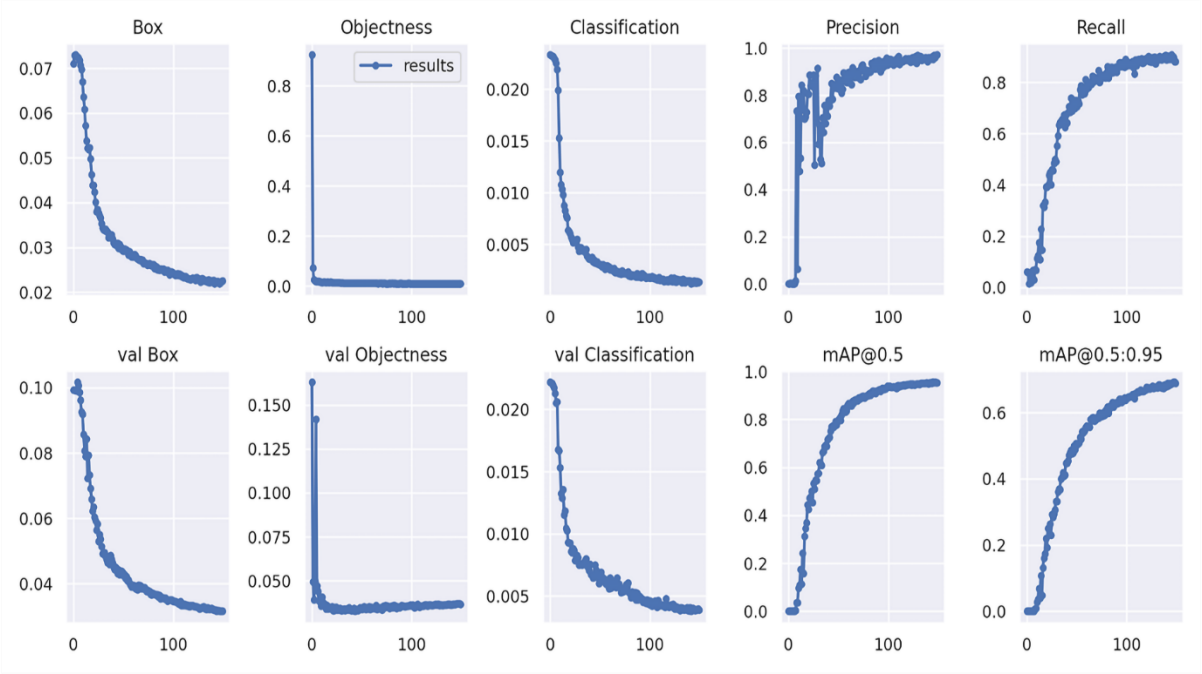

Figure 5. Results of our model.

CONCLUSION

Due to the global outbreak of COVID-19 and Monkeypox, wearing a mask has become a method of epidemic prevention. Monitoring whether people are wearing masks correctly through deep learning methods is a possible solution. We develop the CBAM-YOLOv7 mask detection algorithm in this paper. The algorithm can detect three classes: Wearing a mask, not wearing a mask, and not wearing a mask correctly. The mAP of the proposed CBAM-YOLOv7 algorithm is 0.3% higher than that of the YOLOv7 version, its mAP value is 15.7% and 10.8% higher than that of YOLOv5 and YOLOX, respectively. The proposed algorithm not only improved the accuracy, but also significantly uplifted the recognition speed. Compared with the YOLOv7 version, our algorithm improved the recognition speed per frame up to 73 ms. The current work is of great importance during the global pandemic, and the development of these detectors to monitor whether people are wearing masks is also necessary. In addition, extending CBAM to the adversarial networks is also a viable approach.

REFERENCES

Balaji, S., Balamurugan, B., Kumar, T. A., Rajmohan, R., Kumar, P. P. (2021, March 29). A brief survey on AI based Face Mask Detection System for public places. SSRN. https://ssrn.com/abstract=3814341

Christlein, V., Spranger, L., Seuret, M., Nicolaou, A., Kral, P., Maier, A. (2019). Deep generalized Max pooling. International Conference on Document Analysis and Recognition (ICDAR). https://doi.org/10.1109/icdar.2019.00177

Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W.-C., Wang, C.-B., Bernardini, S. (2020). The COVID-19 pandemic. Critical Reviews in Clinical Laboratory Sciences, 57(6), 365–388. https://doi.org/10.1080/10408363.2020.1783198

Cui, W., Yan, W. (2016) A scheme for face recognition in complex environments. International Journal of Digital Crime and Forensics (IJDCF) 8 (1), 26-36.

Cui, W. (2015) A Scheme of Human Face Recognition in Complex Environments. Master's Thesis, Auckland University of Technology.

Gao, X., Nguyen, M., Yan, W. (2021) Face image inpainting based on generative adversarial network. International Conference on Image and Vision Computing New Zealand.

Gao, X., Nguyen, M., Yan, W. (2022) A face image inpainting method based on autoencoder and adversarial generative networks. Pacific-Rim Symposium on Image and Video Technology.

Gao, X. (2022) A Method for Face Image Inpainting Based on Generative Adversarial Networks. Masters Thesis, Auckland University of Technology, New Zealand.

Gowdra, N., Sinha, R., MacDonell, S., Yan, W. (2021) Maximum Categorical Cross Entropy (MCCE): A noise-robust alternative loss function to mitigate racial bias in Convolutional Neural Networks (CNNs) by reducing overfitting. Pattern Recognition.

Gowdra, N. (2021) Entropy-Based Optimization Strategies for Convolutional Neural Networks. PhD Thesis, Auckland University of Technology, New Zealand.

Huang, G., Zhu, J., Li, J., Wang, Z., Cheng, L., Liu, L., Li, H., Zhou, J. (2020). Channel-attention U-Net: Channel attention mechanism for semantic segmentation of esophagus and esophageal cancer. IEEE Access, 8, 122798–122810. https://doi.org/10.1109/access.2020.3007719

Jiang, X., Gao, T., Zhu, Z., Zhao, Y. (2021). Real-time face mask detection method based on yolov3. Electronics, 10(7), 837. https://doi.org/10.3390/electronics10070837

Jiao, Y., Weir, J., Yan, W. (2011) Flame detection in surveillance. Journal of Multimedia 6 (1).

Jignesh Chowdary, G., Punn, N. S., Sonbhadra, S. K., Agarwal, S. (2020). Face mask detection using transfer learning of inceptionv3. Big Data Analytics, 81–90. https://doi.org/10.1007/978-3-030-66665-1_6

Liang, C., Lu, J., Yan, W. Q. (2022). Human action recognition from digital videos based on Deep Learning. The International Conference on Control and Computer Vision. https://doi.org/10.1145/3561613.3561637

Lin, K., Zhao, H., Lv, J., Li, C., Liu, X., Chen, R., Zhao, R. (2020). Face detection and segmentation based on improved mask R-CNN. Discrete Dynamics in Nature and Society, 2020, 1–11. https://doi.org/10.1155/2020/9242917

Liu, M., Yan, W. (2022) Masked face recognition in real-time using MobileNetV2. ACM ICCCV.

Liu, Z., Yan, W. Q., Yang, M. L. (2018). Image denoising based on a CNN model. International Conference on Control, Automation and Robotics (ICCAR). https://doi.org/10.1109/iccar.2018.8384706

Loey, M., Manogaran, G., Taha, M. H., Khalifa, N. E. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-V2 with resnet-50 for medical face mask detection. Sustainable Cities and Society, 65, 102600. https://doi.org/10.1016/j.scs.2020.102600

Lu, J., Nguyen, M., Yan, W. Q. (2021). Sign language recognition from digital videos using deep learning methods. Communications in Computer and Information Science, 108–118. https://doi.org/10.1007/978-3-030-72073-5_9

Lu, J., Yan, W. Q., Nguyen, M. (2018). Human behaviour recognition using deep learning. IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). https://doi.org/10.1109/avss.2018.8639413

Pan, C., Yan, W. (2018) A learning-based positive feedback in salient object detection. International Conference on Image and Vision Computing New Zealand.

Pan, C., Yan, W. (2020) Object detection based on saturation of visual perception. Multimedia Tools and Applications, 79 (27-28), 19925-19944.

Pan, C., Liu, J., Yan, W., Zhou, Y. (2021) Salient object detection based on visual perceptual saturation and two-stream hybrid networks. IEEE Transactions on Image Processing.

Radhakrishna, A., Yan, W., Kankanhalli, M. (2006) Modelling intent for home video repurposing. IEEE MultiMedia 13 (1), 46-55.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You Only Look Once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. https://doi.org/10.1109/cvpr.2016.91

Rizk, J. G., Lippi, G., Henry, B. M., Forthal, D. N., Rizk, Y. (2022). Prevention and treatment of Monkeypox. Drugs, 82(9), 957–963. https://doi.org/10.1007/s40265-022-01742-y

Sanjaya, S. A., Adi Rakhmawan, S. (2020). Face mask detection using MobileNetv2 in the era of COVID-19 pandemic. International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI). https://doi.org/10.1109/icdabi51230.2020.9325631

Shen, H., Kankanhalli, M., Srinivasan, S., Yan, W. (2004) Mosaic-based view enlargement for moving objects in motion pictures. IEEE ICME'04.

Shen, J., Yan, W., Miller, P., Zhou, H. (2010) Human localization in a cluttered space using multiple cameras. IEEE International Conference on Advanced Video and Signal Based Surveillance.

Shen, D., Chen, X., Nguyen, M., Yan, W. Q. (2018). Flame detection using deep learning. International Conference on Control, Automation and Robotics (ICCAR). https://doi.org/10.1109/iccar.2018.8384711

Singh, S., Ahuja, U., Kumar, M., Kumar, K., Sachdeva, M. (2021). Face mask detection using yolov3 and faster R-CNN models: COVID-19 environment. Multimedia Tools and Applications, 80(13), 19753–19768. https://doi.org/10.1007/s11042-021-10711-8

Song, C., He, L., Yan, W., Nand, P. (2019) An improved selective facial extraction model for age estimation. International Conference on Image and Vision Computing New Zealand.

Venkateswarlu, I. B., Kakarla, J., Prakash, S. (2020). Face mask detection using MobileNet and global pooling block. 2020 IEEE 4th Conference on Information Communication Technology (CICT). https://doi.org/10.1109/cict51604.2020.9312083

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y. M. (2023). YOLOv7: Trainable Bag-of-Freebies sets new state-of-the-art for real-time object detectors. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7464–7475. https://doi.org/10.48550/arXiv.2207.02696

Wang, H., Yan, W. Q. (2022). Face detection and recognition from distance based on deep learning. Advances in Digital Crime, Forensics, and Cyber Terrorism, 144–160. https://doi.org/10.4018/978-1-6684-4558-7.ch006

Wang, J., Yan, W., Kankanhalli, M., Jain, R., Reinders, M. (2003) Adaptive monitoring for video surveillance. International Conference on Information, Communications and Signal Processing.

Wang, J., Kankanhalli, M., Yan, W., Jain, R. (2003) Experiential sampling for video surveillance. *ACM SIGMM International Workshop on Video surveillance* (pp.77-86).

Wang, X., Hu, H.-M., Zhang, Y. (2019). Pedestrian detection based on spatial attention module for outdoor video surveillance. IEEE International Conference on Multimedia Big Data (BigMM). https://doi.org/10.1109/bigmm.2019.00-17

Woo, S., Park, J., Lee, J.-Y., Kweon, I. S. (2018). CBAM: Convolutional block attention module. ECCV, pp.3–19. https://doi.org/10.1007/978-3-030-01234-2_1

Wu, P., Li, H., Zeng, N., Li, F. (2022). FMD-YOLO: An efficient face mask detection method for COVID-19 prevention and control in public. Image and Vision Computing, 117, 104341. https://doi.org/10.1016/j.imavis.2021.104341

Yan, W., Kankanhalli, M., Wang, J., Reinders, M. (2003) Experiential sampling for monitoring. ACM SIGMM Workshop on Experiential Telepresence, 70-72.

Yan, W., Kankanhalli, M. (2015) Face search in encrypted domain. Pacific-Rim Symposium on Image and Video Technology, 775-790.

Yan, W. Q. (2019). Introduction to Intelligent Surveillance: Surveillance Data Capture, Transmission, and Analytics. Springer.

Yan, W. Q. (2021). Computational Methods for Deep Learning: Theoretic, Practice and Applications. Springer Nature.

Yang, G., Feng, W., Jin, J., Lei, Q., Li, X., Gui, G., Wang, W. (2020). Face mask recognition system with YOLOv5 based on image recognition. IEEE International Conference on Computer and Communications (ICCC). https://doi.org/10.1109/iccc51575.2020.9345042

Yin, M., Chen, Z., & Zhang, C. (2023). A CNN-transformer network combining CBAM for change detection in high-resolution remote sensing images. Remote Sensing, 15(9), 2406. https://doi.org/10.3390/rs15092406

Yu, J., Zhang, W. (2021). Face mask wearing detection algorithm based on improved YOLO-V4. Sensors, 21(9), 3263. https://doi.org/10.3390/s21093263

Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., Yang, R. (2019). IOU loss for 2D/3D object detection. International Conference on 3D Vision (3DV). https://doi.org/10.1109/3dv.2019.00019

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE, 111(3), 257–276. https://doi.org/10.1109/jproc.2023.3238524.