Sensors and Materials, Vol. 34, No. 8 (2022) 3271–3284
MYU Tokyo

S & M 3040

Design of Intelligent Detection System for Smoking Based on Improved YOLOv4

Dong Wang,[1,2] Jian Yang,[1] and Feng-Hsiung Hou[3*]

[1]Foshan University, 33 Guang-yun-lu, Shishan, Nanhai, Foshan, Guangdong, P.R. China
[2]Foshan University, Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology, 33 Guang-yun-lu, Shishan, Nanhai, Foshan, Guangdong, P.R. China
[3]Guangzhou College of Technology and Business, 23, San-hua lu, Leping, Sanshui, Foshan, Guangdong, P.R. China

(Received February 25, 2022; accepted July 27, 2022)

Keywords: YOLOv4, smoking detection, k-means++, small-target detection, attention mechanism

In an elevator car, subway, bus, high-speed railway, and other public places, smoking harms the smokers themselves as well as others near them, and seriously affects the public order, and may even cause fire. Thus, smoking detection is important in some public places. The traditional target detection algorithm based on neural networks is used for detection, but it suffers the problems of low speed, high false detection rate, easy loss of the characteristics of small targets, poor generalization, and low robustness. To solve these problems, the improved You Only Look Once version 4 (YOLOv4)-based real-time smoking detection model is designed. For the common target detection algorithm for a small target, the detection effect is not ideal. Therefore, by k-means++ clustering, the a priori box is recalculated and designed to enhance the adaptability of scale and obtain a better anchor. An attention mechanism is also added to improve the accuracy of small-target (cigarette) detection. The detection performance of the model is improved by eliminating certain erroneous detection by using the intersection over union (IoU) of the cigarette and a pedestrian. Because of the lack of international public datasets, we prepare a self-made dataset about smoking by trawling the Internet using crawler technology and combining other self-made datasets. For the dataset used in this study, our improved algorithm model based on YOLOv4 performs better than other one-stage algorithms.

1. Introduction

According to the World Health Organization, there are 1.1 billion smokers globally and more than 360 million in China, the highest number in the world. Tobacco contains more than 70 carcinogens; smoking or passive smoking will cause great harm to people's health. Smoking is not only a health hazard but also a fire hazard. About 20% of fires are caused by smoking each year.

Forced exposure to secondhand smoke in poorly ventilated public places such as elevator cars and buses is severe. In addition, the rapid deceleration of high-speed trains due to smoking causes operational delays. Accidents such as fires caused by smoking are common. No smoking signs have been posted in most public places. Regardless, there are still many smokers in China who

*Corresponding author: e-mail: 1409709996@qq.com
https://doi.org/10.18494/SAM3878

ISSN 0914-4935

smoke in public places. If a cigarette in hand can be identified before the start of smoking, warnings can be issued to halt smoking. Recording smoking behavior can significantly reduce smoking behavior in public places.

1.1 Definition of research problem

At present, the standard traditional smoking detection system uses smoke sensor equipment to detect smoke and judge whether someone is smoking. The detection of smoke triggers sound and light alarms, but in open and well-ventilated public places, the stability of smoke detection cannot be guaranteed. The accuracy and reaction speed of this method are very low, so there are specific requirements for the concentration of smoke. Because smoking produces little smoke, it is generally difficult to detect it. Hence, this is not a reliable method of determining whether anyone is smoking. Another possibility is to use a wearable detection device that can identify human limb movements and speed. Moreover, correlating smoking actions to classifier classifications enables us to judge whether the person is smoking. However, because the cost is high and the device is difficult to carry, the accuracy is unreliable. Therefore, this method for YOLOv4 is viable.

With the continuing development of artificial intelligence, machine vision, and deep learning and the improvement of computing power, deep learning has been fully applied to computer vision. Face recognition, license plate recognition, and, recently, the recognition of mask-wearing faces because of the Covid-19 epidemic have become increasingly reliable. In this study, target detection is applied to detect smoking. Cigarette recognition is tested to determine whether there is smoking behavior. There are two standard convolutional neural network target detection algorithms: the one-stage target detection algorithm and the two-stage target detection algorithm. In two-stage detection, the target candidate box must first be extracted. Then, the detection model, such as FAST Regions with CNN features (R-CNN) or Faster R-CNN,[1,2] must be trained. In one-stage target detection, the idea of regression has been adopted. There is no target candidate box extraction, and the category and position of the target can be detected directly. Compared with the two-stage target detection, the speed is high, but the accuracy is slightly low. Standard single-stage target detection algorithms include You Only Look Once (YOLO),[3,4] SSD,[5] and Retina-net.[6] Since smoke is detected to identify smoking, the method is mainly based on the color of smoke. Compared with the traditional smoke sensor, the detection effect is improved. When the smoke concentration is low, identification has low accuracy and the recognition rate is low.[7,8]

Smoking is judged by targeting cigarettes. In the same way as described above, cigarette color features are segmented and detected to identify smoking. In a single environment, the detection effect is considerable, but the effect becomes poor when faced with more complex scenes.[9] Firstly, deep learning, which breaks through traditional target detection defects and has a particular generalization ability to complex environmental changes, is applied to smoking detection. However, the training dataset is insufficient; the accuracy in general and the detection efficiency are low.[10] Some scholars proposed a fast smoking detection algorithm based on Faster R-CNN.[11] Combined with the face detection algorithm, HSV image segmentation is used to make a preliminary judgment of whether the target is smokeless or not. Then, according to the preliminary results of the next step of detection, the false detection rate is relatively low. However, when the cigarette is not close to the person's face, the detection is not fast enough to recognize the cigarette.

1.2 Purpose of research

Because of the shape and size of cigarettes, the purpose of this research was to improve the network structure based on the YOLOv4 algorithm. We introduced some innovative functions for tuning, added an attention mechanism, and finally improved the accuracy of small-target detection and the model prediction speed. Firstly, k-means clustering was reperformed for nine different priority boxes on the three feature layers, and some large-scale branches in the original network were cut off appropriately in accordance with the clustering situation. Then, the remaining small- and medium-scale branches were added. After subsampling, each branch was fused with high-level semantic information to construct a new scale detection layer. Because various types of noise easily interfere with cigarette detection based on computer vision, false detection often occurs. Since cigarette detection is small-target detection, in which it is difficult to identify the target, there are few cigarette detection methods based on target detection at present, and relevant work and theories are few. In this study, on the basis of the basic idea of target detection theory, cigarettes are taken as the detection and recognition targets. The algorithm uses the self-made data set based on the improved YOLOv4 algorithm for training. Compared with standard target detection algorithms such as YOLOv3, SSD, and barrister RCNN, this algorithm has high accuracy and detection speed.

2. Methodology

We used YOLOv4 in our study. In this section, we discuss the YOLOv4 target detection framework and explain the replicates and statistical methods used in the study.

2.1 Algorithm model structure configuration

YOLOv4 is the latest real-time one-stage target detection algorithm, which is improved from the classical YOLOv3 target detection in many aspects, such as data prepossessing, trunk feature network, feature fusion, activation function, and loss function. There are also many additional tricks compared with the original YOLOv3.

Using the original V3 Darknet, the backbone network integrates the CSPn*et al*gorithm to form CSPDarknet, which includes five CSP modules and 3 × 3 convolution, reducing the computing time of the network and enhancing the learning ability of the network.[12] The neck changes from the characteristic pyramid FPN to the spatial pyramid pooling (SPP) +PAN, and the receptive field increases. The structure adds bottom-up path enhancement to reduce the loss of shallow features in feature transmission to some extent, and the information features obtained after the fusion of multichannel features are abundant.[13] In contrast, the head continues to be the YOLO-head in YOLOv3. Its network structure is shown in Fig. 1.

An essential feature of PANet is repeated feature extraction. The structure of PANet is shown in Fig. 2. Figure 2(a) shows the traditional feature pyramid FPN. The feature pyramid comprises feature extraction from bottom to top, feature fusion, and feature extraction from top to bottom shown in Fig. 2(b). After the PANet structure, different feature layers are fully integrated to effectively improve the target's feature extraction ability.

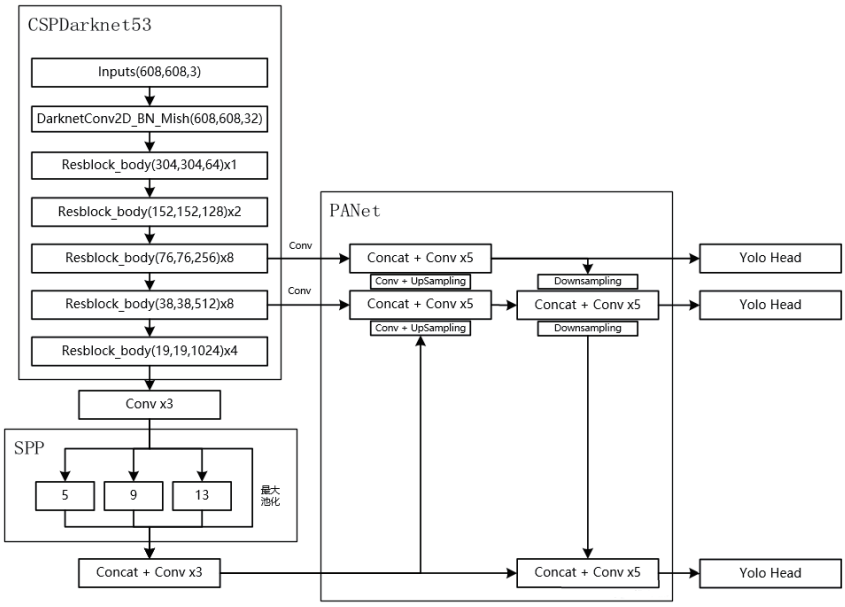

Fig. 1. YOLOv4 network structure.

YOLOv4 has many tricks One example is the use of a new mosaic data enhancement method, which is based on cut-mix data enhancement, to join two pictures together followed by joining four pictures together to dramatically enrich the detection background of the object, strengthen the diversity of input images to a certain extent, and make the image information features richer. The trained detection model can obtain higher robustness. Another example is the use of label smoothing to give a minor penalty to the classification accuracy to prevent the classification from being very precise, as well as overfitting. A third example is that CIOU considers the distance, overlap, scale, and penalty factors between anchors referring to the classic IOU. It makes the target box regression more stable.[14]

2.2 Algorithm detection process

As shown in Fig. 1, YOLOv4 sets the image size as 608 × 608 and inputs it into the network for training. The CSP module is formed by the combination of convolution and residual modules; this can prevent gradient disappearance and explosion to a certain extent. In the backbone network, the convolution layer with a step size of 2 is used to reduce dimensions and carry out subsampling. In the neck module, upper sampling is carried out, and the information fusion of shallow and deep features is completed using PANet and SPP models. Lower sampling is dimensionality reduction, shrinking the image. In this way, the shallow information features can be better utilized, and the problem of feature information loss of small targets can be effectively prevented. The head module of YOLOv4 uses three feature layers processed using PANet in the prediction of smoking. In accordance with the idea of regression classification, target detection is carried out using the three feature layers extracted: the middle layer, the middle and lower layers, and the bottom layer. On the basis of these feature layers, the images are divided into three grid graphs of 76 × 76, 38 × 38,

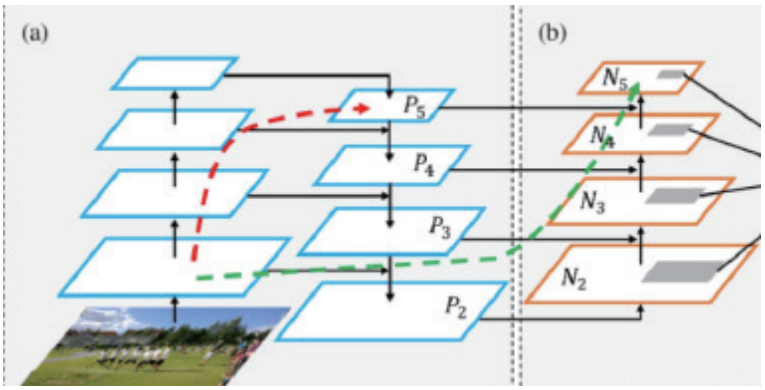

Fig. 2. (Color online) Structure of PANet.

and 19 × 19, changing the shape of the three effective feature layers into (52, 52, 256), (26, 26, 512), and (13, 13, 1024), respectively.

3. Experiment Design and Process

The main purpose of the experiment design is to improve the YOLOv4 algorithm. In this section, we discuss the improved YOLOv4 algorithm and explain the replicates and statistical methods used in the improved YOLOv4 algorithm in this study.

3.1 K-means clustering algorithm

The YOLOv4 algorithm divides detection into three feature layers. Each feature layer contains three prior boxes of different sizes and are used for detecting heads of 13 × 13, 26 × 26, and 52 × 52 to obtain bounding boxes.[15] The last box is detected by the labeled data clustering class of the k-means clustering algorithm. The priority box provided by the original YOLOv4 algorithm was obtained using the COCO data clustering class. The dataset contains up to 80 types of objects, and the original size of an anchor can be seen to be applicable to most scenes.

In this study, the k-means++ clustering algorithm was used to conduct cluster analysis on smoking datasets. As the improved version of k-means, k-means++ has an improved selection of initial points. The core idea of the algorithm improvement is that the distance between the initial clustering centers should be as large as possible. The k-means++ algorithm can reduce the error of classification results to a certain extent, obtain a better initial point, and obtain an anchor that is more suitable for small-target detection.

This algorithm first randomly selects a sample target box as the first clustering center point and then calculates the minimum distance $D(X)$ between each sample and the current clustering center area. The greater the value, the greater the probability of being selected as the clustering center. Then, the roulette wheel method is used to select the second clustering center. Repeat the above steps until all nine clustering centers are selected. The following steps are similar to those of the classical k-means algorithm. However, since using the classical Euclidean distance to measure

labels will lead to more significant errors in large-scale labels and affect the final clustering results, the intersection and association between the clustering center and the label box are used as measurement parameters rather than IOU. The annotation boxes are divided in accordance with the size of the IOU for cluster analysis. After the partition is completed, the above process is repeated until the clustering center no longer changes, and the final nine anchor boxes are obtained.

3.2 Improvement of loss function

The quality of the loss function plays a significant role in improving the accuracy of the model. A good loss function can make the predicted value more accurate. The loss function is specially used to measure the difference between the predicted and actual values, and is an essential model optimization technology. When the neural network propagates forward, the forward propagation function of each network layer will be called, and then the output of each layer will be obtained. The last layer will be compared with the target function, and the error value will be updated through the loss function. Then, the neural network propagates backward to the initial layer and each weight will cause the update.

The error value will be updated through the loss function. Then, it will be propagated back to the initial layer by layer. The loss function of target detection is generally divided into two parts: classification loss function and border regression loss function. For the classification loss function, Lin proposed focal loss,[16] and there are two significant categories of target detection algorithms: one-stage and two-stage algorithms. The accuracy of the one-stage algorithm is always inferior to that of the two-stage algorithm. The motivation behind Lin's proposal of focal loss was to make the one-stage target detection algorithm have two-stage accuracy at the original speed. He believed that the one-stage accuracy problem was caused by the highly unbalanced proportion of positive and negative samples. This function (1) is used to reduce the weight of easily classified samples. It makes the model more focused on the samples that are difficult to classify. The loss function is

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t). \tag{1}$$

For the border regression loss function, we retain the improved CIOU, DIoU,[17] that is, the border regression loss function of the original network of YOLOv4. The CIOU loss function is defined as

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v. \tag{2}$$

3.3 Attention mechanism

The mechanism of attention is the mechanism by which resources are allocated. The function of attention is to focus on the critical features and suppress the unnecessary features. Previous

studies have revealed that attention mechanisms can improve the model's ability to detect small objects such as cigarettes, making it easier for the model to catch cigarettes and improving the model's robustness.[18,19] The essence of this mechanism is to give more weight to what is essential and to divide less for less critical items. Generally, the attention mechanism can be divided into intricate attention and soft attention, and the model structure of the soft attention mechanism can be divided into three types: spatial domain, channel domain, and mixed domain. The spatial domain applies the spatial transformation to the spatial information of the image and then extracts primary information. The channel domain gives a certain weight to different channels, indicating the relevance of the channel information. The newest Image Net image classification model is SENE.[20] In the representation of the channel domain, the model learns the correlation between channels and assigns attention to channels, which signifies the successful application of the attention mechanism in CV. However, the spatial domain treats all channel images the same and ignores the channel domain information, so its interpretation of the channel image in other layers of the neural network is weak. The channel domain's attention directly pools all the information within a channel, ignoring the local information of each channel. By combining the two schemes, we propose the attention mechanism of the mixed domain. This attention mechanism combines both channel attention and spatial attention, as shown in the attention model CBAM.[21] in Figs. 3–5.

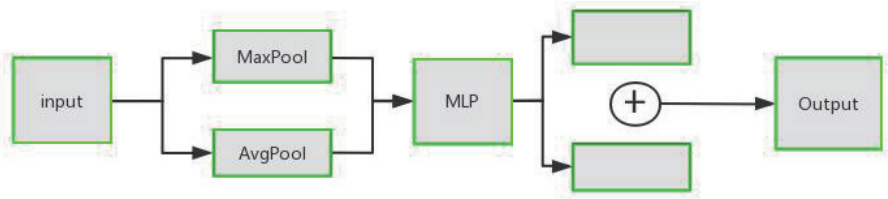

Fig. 3. (Color online) Channel attention feature.

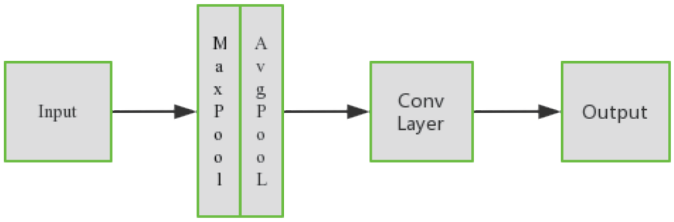

Fig. 4. (Color online) Spatial attention feature.

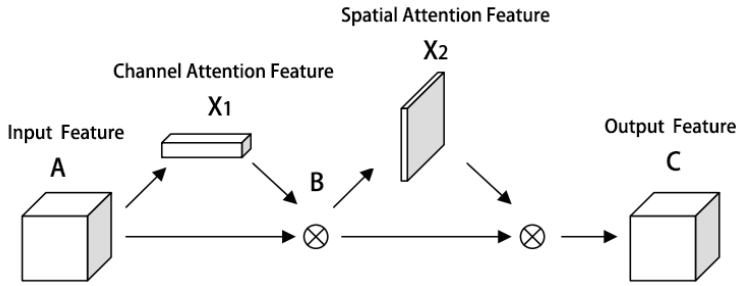

Fig. 5. CBAM.

When A is first input into the channel attention module, it is processed simultaneously by MAXPOOL and AVGPOOL, and the two outputs are combined to obtain the channel feature weight. Finally, the output vector sum is applied to the original feature graph by convolution and multiplication. Then, the output B is input to the spatial attention module, AVGPOOL and MAXPOOL are performed, and the pooled results are fused. B is the feature graph output after channel attention enhancement. It is also the input of the spatial attention module in the next stage. Finally, the result is combined with the original feature image and multiplied to obtain the output feature image C, thus enhancing the spatial attention. The formulae used are

$$B = X_1(A) \otimes A, \tag{3}$$

$$C = X_2(B) \otimes B. \tag{4}$$

A represents the input feature graph; X1 represents channel attention operation; B represents the output after the channel attention mechanism and is also the input of the spatial attention module; X2 stands for spatial attention operation; and C denotes the spatial attention module output.

In this study, we chose the CBAM attention module. Introducing the attention module into the feature extraction network enables the network to capture and extract features more selectively. Thus, the feature expression ability of the strong feature extraction residual network for cigarette targets can be added, making it easier to identify cigarette targets. The improvement diagram of the residual structure is shown in Fig. 6.

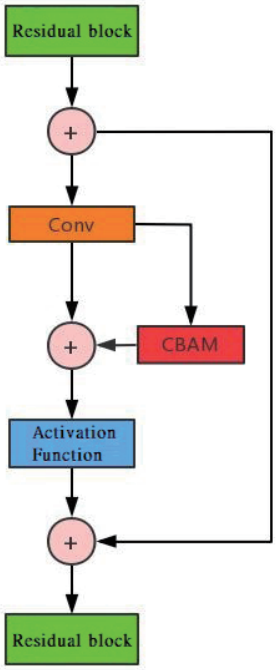

Fig. 6. (Color online) Residual structure diagram.

4. Experiment and Analysis

4.1 Data acquisition

Because of the lack of internationally open smoking datasets, the dataset used in the training and testing of the smoking detection model in this study was self-made. Most of them were obtained from tens of thousands of smoking-related data on the Internet using crawler technology. After manual screening, more than 6000 smoking-related datasets were obtained. Others were obtained as images taken using mobile phones in public places for a total of 7000 pictures. Mosaic data enhancement was adopted in network training, and four pictures were randomly selected from the dataset, enlarged, and spliced randomly. To increase the diversity of the sample, four images are read each time; this method enriches the background of the detection target, making the model generalization ability stronger. In this study, there are two categories to be detected, namely, pedestrian and cigarette.

Owing to the limitation of the number of homemade datasets collected, the number of existing image backgrounds is unbalanced. For example, there are 10 pictures of smoking in an elevator and thousands of pictures of smoking with an outdoor background. In most cases, the model trained in this way will have an excellent detection effect within an outdoor background, while for other scenes, it may be mediocre or even weak (see Figs. 7 and 8).

4.2 Data annotation

Labeling is used to label the pictures before training the model. The software will store the image label information into an XML file and then convert the XML file format into a TXT file through a script. The file consists of six items (path, xmin, ymin, xmax, ymax, and class), which are the picture path, the coordinate information of the real marked box, and the category, as shown in Figs. 9 and 10.

Fig. 7. (Color online) Original image (left).

Fig. 8. (Color online) After improvement and enhancement (right).

Fig. 9. (Color online) Data annotation.

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
<annotation>
        <folder>VOC2007</folder>
        <filename>1.jpg</filename>
        <path>D:\yolov4-pytorch-smopeople\VOCdevkit\VOC2007\1.jpg</path>
        <source>
                <database>Unknown</database>
        </source>
        <size>
                <width>1107</width>
                <height>974</height>
                <depth>3</depth>
        </size>
        <segmented>0</segmented>
        <object>
                <name>person</name>
                <pose>Unspecified</pose>
                <truncated>0</truncated>
                <difficult>0</difficult>
                <bndbox>
                        <xmin>336</xmin>
                        <ymin>117</ymin>
                        <xmax>636</xmax>
                        <ymax>879</ymax>
                </bndbox>
        </object>
        <object>
                <name>smoke</name>
                <pose>Unspecified</pose>
                <truncated>0</truncated>
                <difficult>0</difficult>
                <bndbox>
                        <xmin>526</xmin>
                        <ymin>218</ymin>
                        <xmax>560</xmax>
                        <ymax>232</ymax>
                </bndbox>
        </object>
</annotation>
```

Fig. 10. XML tag file.

4.3 Evaluation indicators

In target detection, it is usually necessary to calculate the precision P and recall R of the detected target to calculate its average precision AP. A single AP value is used to evaluate the advantages and disadvantages of the trained model. The average value of AP for detecting multiple objects and targets is MAP, also known as the mean precision, which is used for the comprehensive evaluation of the model. The larger the MAP value, the higher the detection accuracy of the model. The formulae used are

$$P = \frac{TP}{TP + FP}, \tag{5}$$

$$R = \frac{TP}{TP + FN}, \tag{6}$$

$$MAP = \frac{\sum AP}{class}. \tag{7}$$

TP represents the number of correctly detected targets; FP represents the number of misdirected targets; and FN represents the number of missed targets. The APR curve is a curve with P and R as the horizontal and vertical coordinates, respectively. The essence of AP is the area under the PR curve and is the combination of the precision and recall rates.

4.4 Cigarette detection process

In this study, the real-time detection of smoking was carried out. Video frames were extracted continuously from a video shot by a camera at a certain number of intervals. First, the obtained video frame information is prepossessed. Then, it is input into the smoking detection model. After the detection by the model, pedestrians and cigarettes can be detected with a high level of confidence. Then, the pedestrian frame to cigarette frame intersection ratio is calculated, the threshold value X is set, and the smoking video frame image is saved to the local memory as evidence of smoking. When $IOU < X$, it is judged that there is no smoking and no alarm is given. The detailed flow chart is shown in Fig. 11.

4.5 Experimental setup

The hardware configuration used in this research model training was Core i7-10875H, 16G GeForce RTX 2070 GPU, 8 Gb video memory, and CUDA10.0, CUDNN7.4.1.5, and PyTorch 1.2.0. The learning rate was set to 0.001, the learning rate to 10% of the original after every 50 epochs, the momentum to 0.9, and the attenuation coefficient to 0.0005.

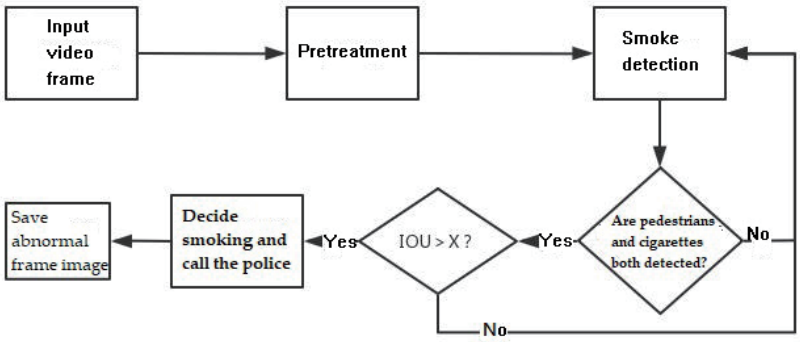

Fig. 11. Flow chart of smoking detection.

Table 1
Test results in this study.

Algorithm	*MAP* (%)
YOLOv4	80.1
YOLOv4 + k-means++	81.5
YOLOv4 + Improved loss function	80.6
YOLOv4 + Attention mechanism	81.9
YOLOv4 + k-means++ + Improved loss function + Attention mechanism	83.5

Table 2
Training test results for smoking dataset.

	MAP (%)	*FPS*
SSD	75.3	22
YOLOv3	77.8	29
YOLOv4	80.1	33
Ours	83.5	33

4.6 Experimental analysis and results

To test the effects of the three improved parts of YOLOv4 on the model's accuracy, in this study, datasets on each module were used in ablation experiments. Specific test results are shown in Table 1. As shown in Table 1, the improved k-means++ clustering algorithm led to an improvement of 1.4% compared with the original algorithm MAP. Compared with the original algorithm, the MAP value obtained with the improved loss function was increased by 0.5%. The MAP value of the attention mechanism was 1.8% higher than that of the original algorithm. Therefore, the three improvements proposed in this study are effective for smoking detection using YOLOv4.

The training configuration remained unchanged after training the present study model using the smoking dataset. SSD, YOLOv3, and YOLOv4 algorithms were used to train detection models using this dataset. The results of experiments were analyzed and compared. MAP and frames per second (FPS) values of different models were compared, as shown in Table 2. After training, the model was used for accurate scene detection, as shown in Figs. 12 and 13. It can be seen from Table 2 that the MAP and FPS values of the improved target detection algorithm

Fig. 12. (Color online) Rendering Fig. 1.

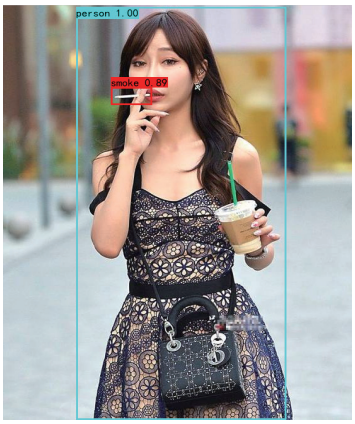

Fig. 13. (Color online) Rendering Fig. 2.

based on YOLOv4 proposed in this study are better than those of other one-stage detection algorithms. This proved that the algorithm proposed in this study is superior for cigarette target detection.

5. Conclusions

This study was motivated by the demand for smoking cessation in public places. An algorithm for detecting smoking behavior was proposed. This study was based on the improved YOLOv4 target detection model. k-means++ clustering was used to optimize the size of the anchor box for the target sample. TheYolov4 target detection model eliminates the unbalanced ratio of positive and negative samples, making the model more focused on the samples that are difficult to classify. The CPAM attention mechanism was added to optimize the algorithm, making the target characteristics of cigarettes easier to be learned, thus making cigarette targets easier to detect. Additionally, invalid cigarettes were excluded in accordance with the intersection ratio of pedestrian and cigarette detection target boxes, which significantly reduced the frequency of misjudgment.

Finally, using the self-made dataset in this study, we proved that the improved algorithm based on YOLOv4 proposed in this study has a favorable effect in terms of accuracy and speed. However, because of the limitations of the self-made dataset in this study, the effect of applying our method in actual public smoking prohibited areas may not be perfect, and there are still some cases of missed detection and misselection. Because of the lack of a dataset on cigarette smoke, the combination of cigarette smoke testing misselection has not been considered for the determination of smoking. In the future, we will continue to expand the dataset and further train and improve the model's detection accuracy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61871129) and the Research Fund of Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology (No. 2020B1212030010).

References

1 J. Moran, L. Haibo, and W. Zhongbo: Acta Optica Sinica 39 (2019) 0715004.
2 Z. Tian, C. Shen, H. Chen, and T. He: Proc. IEEE/CVF Int. Conf. Computer Vision (IEEE 2019) 9627–9636.
3 J. Redmon and A. Farhadi : arXiv preprint arXiv (2018) 1804.02767.
4 A. Bochkovskiy, C. Wang, and H. Liao: arXiv preprint arXiv (2020) 10934.
5 W. Liu, D. Anguelov, and D. Erhan: European Conf. Computer Vision (ECCV 2016) 21–37.
6 T. Y. Lin, P. Goyal, and R. Girshick: Proc. IEEE Int. Conf. Computer Vision (Berlia, Germany) (2017) 2999–3007.
7 Q. Yuan, C. X. Fan, H. F. Qiao, and Z. H. Wang: Computer Engineering and Design (2015) 5.
8 Z. Y. Wang, H. M. Liao, R. D. Zhang, P. Y. Liu, and K. B. Ja: 12th Int. Conf. Signal and Intelligent Information Processing and Applications (NCSII 2018) 35.
9 W. C. Wu and C. Y. Chen: Int. Conf. Genetic and Evolutionary Computing (ICGEC 2018) 127.
10 G. Poonam, B. N. Shashank, and G. R. Athri: Indonesian J. Eng. Comput. Sci. 13 (2019) 113.
11 K. J. Han and Q. Li: J. Xi 'an University of Posts and Telecommun. 25 (2020) 85.
12 C. Y. Wang, H. Y. Mark Liao, and Y. H. Wu: Proc. IEEE/CVF Int. Conf. Computer Vision and Pattern Recognition Workshops (IEEE 2020) 390–391.
13 X. Zhang and S. Ren: IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 1904.
14 A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao: YOLOv4: Optimal Speed and Accuracy of Object Detection (2020) 25.
15 R. Q. Jiang, Y. P. Peng, W. X. Xie, and G. R. Xie: J. Graphics 42 (2021) 1.
16 T. Y. Lin: IEEE Trans. Pattern Anal. Mach. Intell. 42 (2020) 318.
17 Z. H. Zheng, P. Wang, and W. Liu: 34th AAAI Conf. Artifificial Intelligence (AAAI 2020) 12993–13000.
18 J. Hu, L. Shen, and S. Albanie: IEEE Trans. Pattern Anal. Mach. Intell. 42 (2020) 2011.
19 S. Woo, J. C. Park, and J. Y. Lee: 15th European Conf. Computer Vision (ECCV 2018) 3–19.
20 E. Lee and D. Kim: Image Vision Comput. 87 (2019) 24.
21 J. Hu, L. Shen, and S. Gang: Proc. IEEE/CVF Int. Conf. Computer Vision and Pattern Recognition Workshops (IEEE 2018) 7132–7141.

About the Authors

Dong Wang received his B.S. degree from Northeast Petroleum University, China, in 1994 and his M.S. and Ph.D. degrees from the Central South University, China, in 2001 and 2008, respectively. Since 2019, he has been an associate professor at Foshan University and a vice dean of Guangdong Taiwan Institute of Artificial Intelligence. His research interests are in intelligent calculation and optimization calculation.

Jian Yang received his B.S. degree from Foshan University, China, in 2018. His research interests are in intelligent calculation and optimization calculation.

Feng-Hsiung Hou received his Ph.D. degree from the University of the Incarnate Word, USA, in 2002. Since 2022, he has been an associate professor at Guangzhou College of Technology and Business and a vice dean of GCTB-NSU Joint Institute of Technology. His research interests are in artificial intelligence and big data application.