Research on Smoking Detection Based on Deep Learning

Zhen Zhang¹, Huiying Chen²*, Ruijie Xiao², Qingjie Li²

¹ School of Electrical Engineering, Zhengzhou University, Zhengzhou, Henan Province, 450001, China
² School of Information Engineering, Zhengzhou University, Zhengzhou, Henan Province, 450001, China
*Corresponding author’s e-mail: ksh1985357098@gs.zzu.edu.cn

Abstract. Smoking detection in public places is an important means to protect people's health and safety of life and property, but small target detection will have problems such as low detection accuracy and missed detection. In response to this problem, from the perspective of single-stage detection, a smoking detection model suitable for real-time monitoring is proposed. Based on the custom attention mechanism module and the improved residual network, design a backbone network that reuses underlying features while fusing features at different stages to enhance the ability to extract small target features and improve the accuracy of target detection; the feature fusion part uses FPN The structure and PAN structure are integrated, and a lightweight Neck layer network structure is designed to retain the semantic information and location information of the target feature; the DIOU_nms algorithm is selected to improve the missed detection problem. The self-made smoking data set was detected, the average accuracy rate (mAP) reached 86.32%, and the detection speed reached 55f/s. The detection model in this paper improves the accuracy of smoking detection, improves the supervision effect of smoke-free areas to a certain extent, and provides help for eliminating fire hazards.

1. Introduction
Smoking in public places not only causes harm to the health of oneself and others, but also has a great safety risk. Many fire incidents are caused by smoking in sensitive areas. Therefore, more and more public places begin to detect and control smoking behavior. Airports, high-speed trains, gas stations, flammable and explosive warehouses and other smoke-free areas need to be equipped with equipment that can accurately and efficiently monitor smoking behavior to ensure that firefighters and site management personnel can detect fire hazards in a timely manner. With the development of science and technology, the detection of smoking has been improved. Traditional detection methods mostly use various smoke detectors [1], but in open areas such as airports, gas stations and shopping malls, the smoke concentration will decrease rapidly, and the smoke sensing equipment cannot be triggered, so it is difficult to achieve the effect of monitoring and warning. Some researchers have also designed wearable detection devices, but they need to be worn by everyone. The production cost is high and the service life of the devices is short. In addition to physical detection equipment, image processing technology is also gradually playing an important role in smoking detection. Smoking detection based on image recognition is roughly divided into detection for smoke and detection for cigarette. Compared with smoke sensor, smoke detection method can realize smoking detection in a large range and a long distance, and the detection effect is better. However, when the background light is weak and smoke is thin, the detection accuracy is low. The cigarette detection method solves the influence
of smoke concentration on the detection accuracy, but the detection accuracy is still not ideal due to the small cigarette target in the image captured by the surveillance camera and the overlap of occlusion.

Therefore, in order to realize the rapid and accurate detection of smoking behavior in public places, this paper designs a single-stage detection model by drawing on the idea of YOLOV5 algorithm. In this model, the semantic information and location information of small targets were extracted by the backbone network based on the custom attention mechanism module and the improved residual network module. The FPN structure and PAN structure of Yolov5 were used for the fusion of features at different scales to improve the detection accuracy of small targets, and the DIOU_NMS algorithm was used to improve the missed detection of targets.

2. Model building

2.1. Custom attention mechanism module

This paper uses the design idea of CBAM module for reference, and defines a lightweight attention mechanism module named CBAM-Tiny. This method can not only save parameters and calculation, but also ensure the accuracy of small target detection.

Given that the feature of the original image of $H \times W \times G$ is $F$, the global maximum pooling and global average pooling of the space are carried out first, and the eigenvalues of the input images are compacted to obtain two $1 \times 1 \times G$ channel features $F_1$ and $F_2$, and the two features are added. Then, the $1 \times 1$ convolution is used to reduce the network dimension of the new feature, and the ReLU activation function is used to learn the nonlinear relationship between the features. Finally, the $1 \times 1$ convolution is used to reconstruct the dimensions before computational compression, and the channel feature $F_3$ is obtained.

In the right half, the original image of $H \times W \times G$ is firstly pooled through channel average pooling and channel maximum pooling to obtain the location information of the features in the original image and obtain two channel features of $H \times W \times 1$. The two features are splice together, and then through a convolution layer with the size of the convolution kernel of $7 \times 7$, a spatial feature $F_4$ with channel 1 is obtained.

Finally, the channel feature $F_3$ in the left part and the space feature $F_4$ in the right part are spliced. The Sigmoid activation function is used to obtain the weight coefficient $M$. Multiply the original image feature $F$ with the weight coefficient $M$ to get the final feature $F_5$. The module structure of the custom attention mechanism is shown in Figure 1. This paper integrates the custom attention mechanism module into the backbone network.

![CBAM-tiny structure diagram](image)

**Figure 1. CBAM-tiny structure diagram.**

2.2. Residual network improvement

The more layers the neural network has, the richer the features can be extracted and the stronger the representation ability is. However, with the increase of network depth, there will be problems of gradient explosion and dissipation. Another significant problem is model degradation, and the performance of network will get worse and worse. To solve this problem, YOLOV5 network uses the
design idea of CSPNET for reference. The CSP1 structure of the backbone network adopts ResUnit residual network structure to eliminate repeated gradient information, and the gradient transformation process is integrated in the feature map[2].

In 2017, Huang G et al. proposed a new network structure, DENSENET, which refers to the ideas of ResNet and Inception network, realizing the direct pair connection between all layers[3]. In a traditional convolutional neural network, if there are k layers, there are k connections. In Densenet, there will be K (K +1) /2 connections to ensure that the input of each layer in the network comes from the output of all previous layers[4].

In this paper, the improvement of residual component of CSP1_x structure in backbone network is based on the idea of DenseNet network. Input of each layer of custom DenseBlock module is derived from output of all previous layers. Such a tight connection structure makes each layer directly connected to input and loss. The network structure of the custom DenseBlock module is shown in Figure 2.

In the improved residual network, the ResUnit residual network structure of CSP1 structure is replaced with a custom DenseBlock module. The improved residual module is named CSP_Dense, and its structure is shown in Figure 3.

2.3. Selection of non-maximum suppression algorithms

Non-maximum Suppression (NMS) is a method to ensure that the algorithm only gets one detection for each object, namely "cleaning detection"[5]. On the basis of Diou_loss, YOLOV4 algorithm adopts the Diou_NMS method to carry out prediction box screening [6], which has a good improvement effect on the missed detection of occluded targets. Therefore, this paper draws on the Diou_loss idea in YOLOV4 and uses the Diou_NMS algorithm to screen the target box of the output images.

DIOU-NMS takes DIOU as the criterion of NMS, because in the suppression criterion, not only the overlapping region but also the center distance between two boxes should be considered, while DIOU considers both the overlapping region and the center distance between two boxes. For the prediction Box M with the highest score, the SI update formula of DIOU-NMS can be formally defined as:

\[
S = \begin{cases} 
    s_i, & IOU - R_{DIOU}(M, B_i) < \epsilon \\
    0, & IOU - R_{DIOU}(M, B_i) \geq \epsilon 
\end{cases}
\]

Box Bi is deleted by considering the distance between IOU and the center points of the two boxes simultaneously, \(s_i\) is the classification score, and \(\epsilon\) is the NMS threshold. Diou-NMS believes that two boxes with far center points may be located on different objects and should not be deleted, which is the biggest difference between Diou-NMS and NMS.
2.4. Model structure design
This section describes in detail the overall structure of the smoking detection model proposed in this paper. First of all, the backbone network responsible for feature extraction slits the original image using the Focus structure. The original 608×608×3 image is input into the Focus structure and slits into a 304×304×12 feature image first. After a convolution operation of 32 convolution cores, it finally becomes a 304×304×32 feature image. Next, the backbone network is composed of convolution layer, custom attention mechanism module, residual network and SPP module.

Neck network structure adopts FPN structure and PAN structure. The FPN layer conveys strong semantic features from top to bottom, and transfers the strong semantic features from high level to lower level. The semantic information of the whole pyramid is enhanced, but the location information is not transmitted. The PAN structure conveys the strong positioning feature information from bottom to top, and the parameters of different detection layers are aggregated from different trunk layers. Finally, the results of the three scales are output, and the non-maximum suppression operation is carried out on the output results through the DIOU_NMS algorithm to eliminate the interference target box and solve the problem of missing detection of small targets. Finally, the position information of the target to be measured is obtained. Figure 4 shows the overall network architecture of the smoking detection model.

![Figure 4. The overall architecture of the model.](image)

3. Experiment and Analysis

3.1. Experimental platform and network training
The experimental platform in this paper uses i5 processor, Nvidia GeForce 2080TI graphics card, Ubuntu 18.04 operating system, CUDA9.0 configuration, and PyTorch deep learning framework. A total of 100 epochs were trained, in which the initial learning rate in the training process was 0.016. Equal interval was adopted to adjust the learning rate Steplr [7], the interval (Step_size) was adjusted to 25, and the momentum parameter (Momentum) was 0.847. Adam gradient optimizer is used, and the regularization method adopts Dropblock[8]. Figure 5 shows the change of the loss value when the model is trained on the self-made data set. The detection effect of the trained model in the test set is shown in Figure 6.
3.2. Test model performance analysis

In order to improve the smoking detection performance of the detection model in this paper, a comparative test was conducted. Faster RCNN, SSD, YOLOV5 and the present model were trained on the labeled data sets respectively, and the experimental results after training 100 EPOCH were compared and analyzed. The algorithm performance evaluation indexes are detection accuracy mAP and algorithm deduction speed FPS, and the input image pixels are 608*608. The performance index results of the four algorithms are shown in Table 1, and the detection effect is shown in Figure 7.

| The algorithm name | Backbone network     | mAP%  | FPS/(f/s) |
|--------------------|----------------------|-------|-----------|
| Faster RCNN        | ResNet-10            | 84.75 | 6         |
| SSD                | VGG16                | 76.28 | 78        |
| YOLOv5             | CSP                  | 81.62 | 53        |
| Custom model       | CSP-Dense+CBAM-tiny  | 86.32 | 55        |

![The loss curves](image1)

Figure 5. Loss loss graph during training.

![Figure 6. Test effect diagram.](image2)

Figure 6. Test effect diagram.

As can be seen from the above experimental data, the detection accuracy and detection speed of the proposed model in the smoking detection data set have been improved. Compared with Faster RCNN,
SSD and YOLOV5, the detection accuracy MAP value has been increased by 1.57%, 10.04% and 4.7%, respectively. Compared with the YOLOV5 algorithm, the deduced speed of FPS improves by 2F/s. In order to meet the actual demand that the frame rate of real-time detection of surveillance video is greater than 25 frames, the detection model proposed in this paper improves the detection accuracy and speed, and improves the effect of smoking prevention and control to a certain extent. As can be seen from the detection effect diagram, when the cigarette target in the image is small, the detection model in this paper can identify more targets, and SSD and YOLOV5 algorithms will miss the detection problem for small targets, and SSD algorithm has the worst detection effect.

4. Conclusion
According to the actual application scenarios of smoking detection in public places, this paper proposes a lightweight detection model, which has a good detection effect for cigarette detection. First of all, in order to improve the problem of low detection accuracy of small targets, draw lessons from the dense connection idea of the DenseNet network, improve the residual network, and design a feature extraction backbone network based on a custom attention mechanism module. Secondly, in order to improve the detection speed, drawing on the FPN structure and PAN structure in YOLOv5, a lightweight Neck layer network is designed to reduce the amount of model calculation. Finally, the prediction layer selects the DIOU_nms algorithm as the non-maximum value suppression operation to screen the prediction frame, and improves the robustness of the model through the Mosaic data enhancement method. The model in this paper shows good detection performance on the self-made data set, the training accuracy reaches 86.32%, and the detection speed reaches 55f/s. To a certain extent, it solves the problem of low detection accuracy and missed detection caused by small cigarettes in smoking detection. Such problems can help managers accurately and timely monitor smoking problems in public places and reduce the occurrence of safety hazards such as fires.

References
[1] Jang Hyo Yeon,Hwang Cheol Hong. Methodology for DB construction of input parameters in FDS-based prediction models of smoke detector[J]. Journal of Mechanical Science and Technology,2021,34(12).
[2] Shu Lang,Zhang Zhijie,Lei Bo.A Dense-Yolov5 algorithm for infrared target detection[J].Optics and Optoelectronic Tech-nology,2021,19(01):69-75.
[3] Montalbo Francis Jesmar P. Diagnosing Covid-19 Chest X-Rays with a Lightweight Truncated DenseNet with Partial Layer Freezing and Feature Fusion[J]. Biomedical Signal Processing and Control,2021(prepublish).
[4] Yang Jihao. Research on the Detection Method of Dim and Small Targets in Infrared Images[D]. University of Electronic Science and Technology of China, 2020.
[5] Qiu Shaohua. Research on high-resolution remote sensing image occlusion target detection method based on partial con-figuration model[D]. National University of Defense Tech-
ology, 2018.
[6] Cai Shuping,Sun Zhongming,Liu Hui,Wu Hongxuan,Zhuang Zhenzhen.A real-time detection method for obstacles in or-chard based on improved YOLOv4[J].Transactions of the Chinese Society of Agricultural Engineer-ing,2021,37(02):36-43.
[7] Geng Lei,Wang Xuebin,Xiao Zhitao,Zhang Fang,Wu Jun,Li Yueling,Su Jingjing.Fast compression tracking algorithm combining feature selection and secondary positioning[J].Acta Automatica Sinica,2016,42(09):1421-1431.
[8] Ding Yifeng, Ma Zhanyu,Wen Shaoguo,Xie Jiyang,Chang Dongliang,Si Zhongwei,Wu Ming,Ling Haibin. AP-CNN: Weakly Supervised Attention Pyramid Convolutional Neural Network for Fine-Grained Visual Classification.[J]. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society,2021,30.