Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Safety Helmet Detection Based on SSD

Kebin Cui [a] , Zhiping Liu [b, *]

Department of Computer Science, North China Electric Power University, China

[a]ncepuckb@163.com, [b] zhiping_95@163.com

Abstract: Aiming at the backward security monitoring measures in some high-risk work scenes, a helmet wearing detection method based on SSD algorithm is proposed. The algorithm integrates low-level features and high-level semantic features to have better detection effect on small targets, and uses multi-scale default box to make the prediction box closer to the real box and better prediction effect. The experimental results show that it can achieve real-time detection even in the case of large number of small targets and mutual occlusion. It can meet the detection of workers' wearing of safety helmets in construction sites, mines and power plants, and realize safe and efficient production.
Keywords: Helmet wearing test, real time, SSD, convolutional neural network.

---

1. INTRODUCTION

With the progress of society, enterprises pay more and more attention to production safety. In production environments such as construction sites, mining areas and power plants, there are safety risks such as falling objects, head electric shock and falling from high altitude. Safety helmets are very important protective measures. Therefore, supervising staff to wear safety helmets in these work scenes is an important part of safety production. However, compared with the modern high production level, the safety protection supervision mode of enterprises is still in a relatively backward stage. Many enterprises still use the original manual supervision mode, which has great potential safety hazards.

At present, there are some intelligent security systems, but most security systems are not intelligent enough to meet the security needs of dangerous work areas. For example, there are few applications of helmet wearing detection in intelligent security systems. At present, the traditional detection methods are mainly used to detect the helmet, such as using the color feature to detect the helmet, or first detect the face, and then classify the helmet feature through the classifier. However, these methods require subjective design features, large amount of calculation and high time complexity. At the same time, in the actual production environment, there are many colors of safety helmets, large background color interference and changeable personnel posture, which makes the detection accuracy not high.

In this paper, the SSD algorithm is used to train the helmet detection model under the tensorflow framework. The algorithm integrates low-level features and high-level semantic features to have

better detection effect on small targets, and uses multi-scale default box to make the prediction box closer to the real box and better prediction effect. Finally, the model is used to detect the content of input pictures, videos and cameras. Whether it is input pictures, videos or real-time detection, it has good detection results.

2. SSD INTRODUCTION

SSD algorithm is a one-step detection algorithm. It introduces the concept of a priori box, and then combines the regression idea with anchors mechanism, so as to simplify the computational complexity and directly predict the target category and bounding box coordinates. When detecting targets with different sizes and shapes, SSD algorithm does not need to convert the images into different sizes and process them separately, as the traditional target detection algorithm does, but integrates the feature maps of different convolution layers to obtain the same results.

The main design concept of the algorithm mainly includes the following three points:

Multi-scale feature map is used for detection, so that targets of different sizes have corresponding feature maps to detect them, so as to improve the detection accuracy. In the characteristic graphs of different feature layers, each cell will produce different a priori boxes, so different boundary boxes will be obtained, and then these boundary boxes will be combined, and then the final required boundary box will be obtained through non maximum suppression processing. Figure 1 shows the characteristic diagrams of different scales.

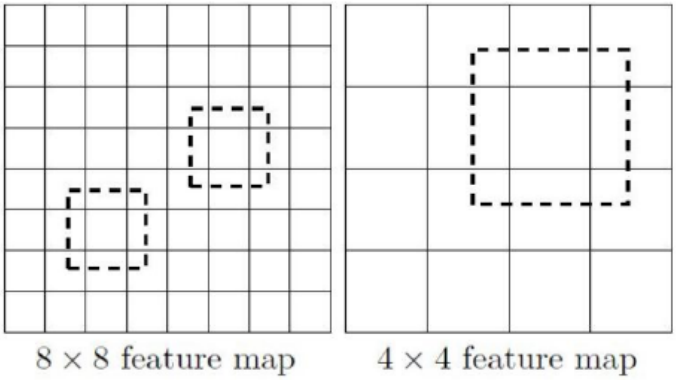

Fig. 1 Feature maps of different scales

(2) Convolution is used for detection, that is, SSD algorithm uses convolution to directly extract detection results from different feature images.

(3) Set a priori box, similar to anchor mechanism in fast r-cnn. In general, a plurality of prior frames with different scales or aspect ratios are preset in each cell, and then their boundary frames are predicted according to these prior frames. As shown in Figure 2, four different a priori frames are set for each unit, and then the appropriate a priori frame is selected for training according to the shape and size of the detection target in the figure. In training, the matching between a priori frame and real target is mainly divided into two parts: the first part is to find the a priori frame with the largest intersection and union ratio for each real target; The second part is for the unmatched a priori frame. When the intersection union ratio with the real target is greater than a certain threshold, the threshold is usually 0.5, and the a priori frame matches it.

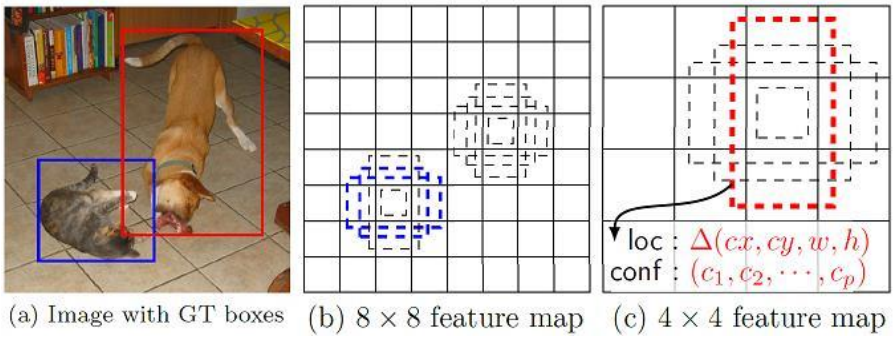

Fig. 2 Anchors

For the feature map with the size of M * n, k a priori boxes are set for each unit, and each a priori box predicts the scores of C categories and the values of 4 offsets. Therefore, a total of (c + 4) * k prediction values are required, and the feature map will produce (C + 4) * n * m * k prediction values. As shown in Figure 3, the SSD uses vgg16 as the basic feature extraction layer, uses its first five layers, and then converts the full connection layers FC6 and fc7 into two convolution layers, adds three convolution layers and a mean pooling layer, and the size decreases step by step to extract features. The offset of the bounding box and the score of each category are predicted according to the characteristic diagrams of different levels, and the final detection results are obtained by non maximum suppression. This network structure balances the detection speed and accuracy.

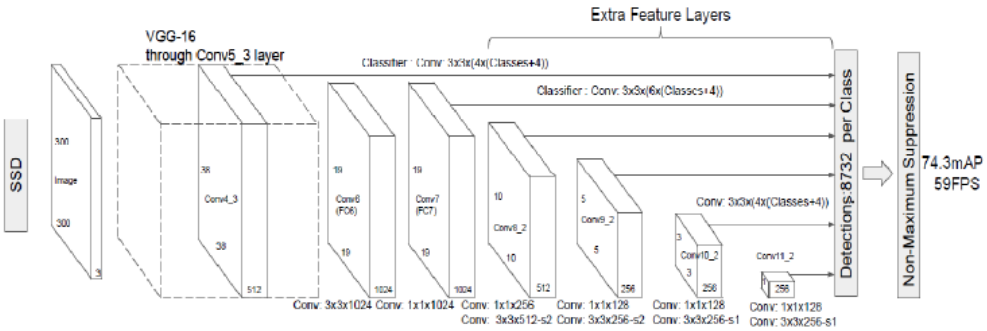

Fig. 3 SSD network structure

The generation rules of a priori box are as follows:

Set M feature maps of different sizes for prediction, which is the ratio between the size of a priori box and the input picture, and the ratio between the top layer and the bottom layer and the picture, usually 0.95 and 0.2, then the other layers are:

$$S_k = S_{min} + \frac{S_{max} - S_{min}}{m-1}(k-1), k\epsilon[1, m] \tag{1}$$

For different aspect ratios α， The actual width W and height h of the a priori box are calculated by the following formula:

$$\omega_k^{\alpha} = S_k\sqrt{\alpha_r} \tag{2}$$

$$h_k^{\alpha} = S_k/\sqrt{\alpha_r} \tag{3}$$

The coordinates of the center point of the a priori frame are: $(\frac{i+0.5}{|f_k|}, \frac{j+0.5}{|f_k|})$, where $|f_k|$ is the size of the characteristic graph, the value range of i and j is$[0,|f_k|]$, that is the center point of the a priori frame of each unit is in the center of the unit. Thus, all a priori boxes can be obtained.

3. EXPERIMENTAL RESULTS

3.1 Data sets

Using the open source SafetyHelmetWearing-Dataset (SHWD), there are 7581 images in total, including 9044 bounding boxes (positive class) with safety helmets and 111514 bounding boxes (negative class) without safety helmets. All images are marked with labelimg to indicate the target area and category. The label of each bounding box: "hat" means wearing a helmet, and "person" means not wearing a helmet. Annotations are in XML form. Most of the data on the person tag in this dataset comes from the scut-head dataset, which is used to judge the person who is not wearing a helmet.

3.2 Model training

The experimental platform of this project is pycham, using tensorflow 2.3 framework, through the helmet detection of the contents in the input pictures, videos and cameras,

Input the picture into SSD network, extract the picture features using multilayer convolution neural network, and take the features of the third convolution of conv4, the features of fc7, the second convolution of conv6, the second convolution of conv7, the second convolution of conv8 and the second convolution of conv9 as effective feature layers to obtain the prediction results. Then, for each effective feature layer obtained, we perform num once respectively_ The convolution of priors * 4 is used to predict the change of each a priori box on each grid point on the feature layer, num_ priors * num_ The convolution of classes is used to predict the type corresponding to each prediction frame at each grid point on the feature layer (num_priors is the number of a priori frames and num_classes is the number of prediction types), and calculate the a priori frame corresponding to each effective feature layer. The prior frame corresponding to each effective feature layer corresponds to a plurality of preset frames on each grid point on the feature layer.

After adjustment, the SSD decoding process will be carried out to convert the prediction result into the position of the prediction frame, and the center point of each grid plus its corresponding X_ Offset and Y_ Offset, the result after adding is the center of the prediction frame, and then the length and width of the prediction frame are calculated by combining a priori frame with H and W. So you can get the position of the whole prediction box. Finally, the selected prediction frame is sorted by score and non maximum suppression to screen the final prediction frame. Through the above steps, the position of the prediction frame on the original graph has been obtained, and these prediction frames have been filtered. These filtered boxes can be drawn directly on the picture. The prediction box and its category (person, hat) can be drawn on the picture by using the imagefont and ImageDraw methods in the PIL library for result visualization.Figure 4 is the result show.

Fig. 4 result show

4. CONCLUSION

In this paper, SSD algorithm is used to realize the real-time detection of helmet wearing, which can achieve good detection effect even in the case of multi-target and occlusion, and provide a new method for security monitoring in some working scenes where helmet must be worn. Of course, there is still room to improve the current accuracy. Sometimes there will be some false positives or missing positives. The accuracy of detection can be further improved by improving the algorithm of the detection module.

REFERENCES

[1]Cuiwen Zhang, Changlun Zhang, Qiang He, Hengyou Wang. Research on Frame Regression Loss Function in Target Detection[J]. Computer Engineering and Application,,:1-8.

[2]Yanqin Zhao, Donglai He, Zihan Zhu. Helmet Wearing Detection Algorithm Based on Integrated Learning in Coal Mine[J]. Journal of Heilongjiang University of Science and Technology,2021,31(02):240-246.

[3]Shuai Jin. Construction and Application of Intelligent Helmet System Based on Safety Risk Management and Control[J]. Hydropower and New Energy,2021,35(03):74-78.

[4]Qinyuan Li, Zhaohong Deng, Xiaoqing Luo, Xin Gu, Shitong Wang. SSD Target Detection Algorithm Based on Attention and Cross Scale Fusion[J]. Computer Science and Exploration,,:1-14.

[5]Libo Sun, Ming Zhang. Intelligent Recognition of Helmet Wearing Based on Target Detection Algorithm[J]. Cement Technology,2021,(02):85-88.

[6]Xiaona Li. Detection Method of Safety Dress Based on Mot Model[J]. Scientific and Technological Innovation and Application,2021,(11):130-132+135.

[7]Shiqin Yue, Qian Zhang, Dingqin Shao, Yu Fan, Jinhua Bai. Research on Helmet Wearing State Detection Based on Resnet50 SSD[J]. Yangtze River Information and Communication,2021,34(03):86-89.

[8]Zhenyu Yin, Chao Fan, Zhihao Zhao, Zhe Huang, Feiqing Zhang. Target Detection Algorithm Based on Multi-Scale Feature Map Classification and RE Extraction[J]. Minicomputer System, 2021, 42(03):536-541.