Safety Helmet Wearing Detection Based on Image Processing and Deep Learning

Wei Zhang
Harbin Institute of Technology
Harbin, China
18S103172@stu.hit.edu.cn

Chi-fu Yang
Harbin Institute of Technology
Harbin, China
cfyang@hit.edu.cn

Feng Jiang
Harbin Institute of Technology
Harbin, China
fjiang@hit.edu.cn

Xian-zhong Gao
National University of Defense Technology
Changsha, China
gaoxianzhong@nudt.edu.cn

Xiao Zhang
Chengde Chenggang Engineering Technology Co., Ltd.
Chengde, China
hit_cvzhangwei@163.com

***Abstract*—The environment of the steel factory workshop is complex, and there may be a variety of unexpected potential dangers, so wearing a helmet to enter the workshop is a prerequisite for the factory. In order to supervise this situation, it is necessary for employees to wear helmets for testing, which is a key part of the overall intelligent monitoring system for steel plant personnel. In this paper, through the crawler to collect high-definition employees wearing helmets and no helmet pictures, using manual labeling, proposed a helmet detection framework based on computer vision deep learning detection framework Faster-RCNN. The actual testing results produce convincing experimental results, which proves the effectiveness and practicability of the proposed framework.**

Keywords-helmet detection; image processing; Faster-RCNN; deep learning

I. INTRODUCTION

As we all know, the monitoring system is very important to the safety of the steel workshop. In the past few decades, some artificial intelligence technologies, such as computer vision and machine learning, have been rapidly applied to factory intelligent monitoring [1]. It can not only avoid time-consuming labor-intensive work, but also point out equipment failures and workers' illegal and timely operation to accurately prevent accidents. In addition to the safety of the equipment, the intelligent monitoring system can also monitor whether workers wear helmets in compliance with safety regulations. As the most common safety operation in factories, real-time helmet wear detection for employees is an important task related to workers' safety. Therefore, it is of great significance to develop factory employees who can automatically detect whether they wear helmets or not. Unfortunately, there is little work, and most of it is testing for motorcyclists wearing helmets. For example, Waran [1], etc., use moving object extraction and K-nearest neighbor (KNN) classifier to develop a system that can automatically classify motorcyclists and determine whether they are wearing helmets or not. In [2], Silva et al. The Hough transform and directional gradient histogram descriptor are applied to extract features, and a multi-layer perceptron classifier is used to identify motorcyclists without helmets. In [3], Kalman filter and cam shift algorithm are used to track pedestrians and determine moving objects. At the same time, the color information of the helmet is used to detect the wearing of the helmet.

The main purpose of this paper is to develop a safety helmet wearing detection system for walkers in factories. Because this work is of great significance to all factories, to avoid personal injuries caused by careless rubbing, employees can better protect themselves. After comparing the traditional target detection methods and deep learning target detection methods, the Faster-RCNN with the best performance is selected as the basis of the network framework. In view of the lack of helmet detection data sets, through the design of crawler framework, high-definition data sets of employees wearing helmets and employees without helmets are collected from Google, and then manually tagged, trained and tested later.

II. RESEARCH ON TARGET DETECTION ALGORITHM FOR DEEP LEARNING

A. A Survey of the Development of Deep Learning Target Detection algorithm

At present, the mainstream target detection algorithms are mainly based on deep learning model, which can be divided into two categories: (1) two-stage detection algorithm, which divides the detection problem into two stages, first generates candidate regions, and then classifies candidate regions. The typical algorithms of this kind of algorithms represent R-CNN algorithms, such as Rmur CNN Fast Rmur CNNQuest Faster R-CNN and so on. (2) one-stage detection algorithm, which

978-1-7281-9761-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CISCE50729.2020.00076

does not need region proposal stage, directly produces the class probability and position coordinate values of objects, which are more typical algorithms such as YOLO and SSD[4-7]. The main performance indicators of the target detection model are detection accuracy and speed. In general, the two-stage algorithm has an advantage in accuracy, while the one-stage algorithm has an advantage in speed. However, with the development of research, the two kinds of algorithms are improved in two aspects[8] .

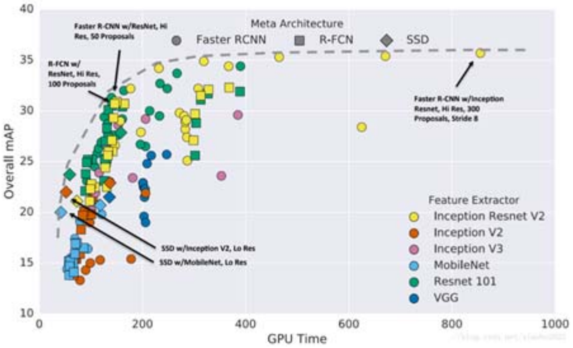


Figure 1. comparison of the performance of Faster RmurCNN dint and SSD algorithms on MS COCO datasets

B. datasets and performance metrics

Data sets commonly used in target detection include data sets such as PASCAL VOC, ImageNet, MS COCO, which are used by researchers to test the performance of algorithms or for competitions. The performance index of target detection should take into account the location of the detected object and the accuracy of the prediction category[9] . We will talk about some commonly used performance evaluation indicators below.

PASCAL VOC (The PASCAL Visual Object Classification) is a well-known data set in the fields of target detection, classification, segmentation and so on. From 2005 to 2012, eight different challenges were held. PASCAL VOC contains about 10000 pictures with bounding boxes for training and verification. However, the, PASCAL VOC dataset contains only 20 categories, so it is regarded as a benchmark dataset for target detection problems.

ImageNet released a target detection dataset with bounding boxes in 2013. The training data set contains 500000 pictures and belongs to 200 categories of objects. Because the data set is too large, the amount of calculation required for training is very large, so it is rarely used. At the same time, due to the large number of categories, target detection is also quite difficult. The comparison between the 2014 ImageNet dataset and the 2012 PASCAL VOC dataset is here.

Another well-known dataset is Microsoft the established MS COCO dataset. This data set is used in a variety of competitions: image title generation, target detection, key point detection and object segmentation. For target detection tasks, COCO contains a total of 80 categories. The training and verification data set of the annual competition contains more than 120000 images and more than 40000 test images. Test sets have recently been divided into two categories, test-dev datasets for researchers and test-challenge datasets for competitors. The label data of the test set is not disclosed to avoid over-fitting on the test set[10-12]. In COCO 2017 Detection Challenge, the scientific team won the championship with the proposed Light-Head R-CNN model (AP is 0.526). It seems that the two-stage algorithm is more accurate.

III. FASTER-RCNN TARGET DETECTION ALGORITHM

Faster-R-CNN algorithm consists of two modules: PRN candidate box extraction module and Fast R-CNN detection module. The RPN (Region Proposal Network) area suggests that the network is used to extract the detection area, which can share the convolution features of the whole graph with the whole detection network, so that the area recommendation takes almost no time. The core idea of RPN is that the method used to generate Region Proposal, directly using CNN convolution neural network is essentially a sliding window (only need to slide once on the last convolution layer), because the anchor mechanism and frame regression can obtain multi-scale Region Proposal with multi-aspect ratio. RPN network is also a full convolution network (FCN, fully-convolutional network), can train end-to-end for the task of generating detection suggestion box, and can predict the boundary and score of object at the same time. Only two additional convolution layers (full convolution layer cls and reg) have been added to CNN.

The function for an image in Faster-R-CNN is defined as:

$$L\quad p_i\ ,\ u_i\quad = \frac{1}{N_{cls}}\sum_i L_{cls}\quad p_i, p_i^*\quad + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}\quad t_i, t_i^* \qquad (1)$$

In the above formula, i denotes anchors index, p_i means that foreground softmax probability, p_i^* represents the corresponding GT predict probability (that is, when the IoU between the i anchor and GT is IoU > 0.70, it is considered that the anchor is foreground, otherwise, the anchor is considered to be background,; as for those anchor with 0.3 < IoU < 0.7, they do not participate in training); t represents predict bounding box . t^* represents the GT box corresponding to the positive anchor. As you can see, the whole Loss is divided into two parts:

Cls loss, that is, the softmax loss, calculated by the rpn_cls_ loss layer, is used to classify anchors as positive and negative network training.

Reg loss, the soomth L1 loss, calculated by the rpn_loss_ Bbox layer, is used for bounding box regression network training.

Because in the actual process, the gap between N_{cls} and N_{reg} is too large, the parameter λ is used to balance them (for example, $N_{cls} = 256$, $N_{reg} = 2400$, set $\lambda = \frac{N_{reg}}{N_{cls}} \approx 10$), so that two kinds of Loss can be evenly considered in the calculation of total network Loss. The important thing here is that the soomth L1 loss, calculation formula used by L_{reg} is as follows:

$$L_{reg}\ \ t_i,t_i^* \ = \sum_{i\in\ x,y,w,h} smooth_{L1}\ \ t_i - t_i^* \tag{2}$$

$$smooth_{L1}\ \ x\ = \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \tag{3}$$

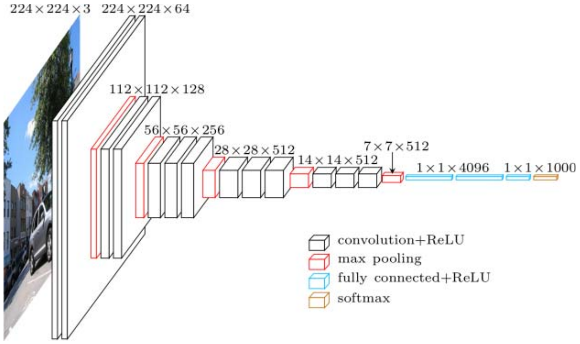
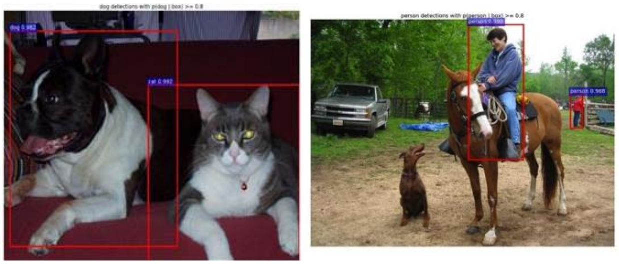


Figure 2. Schematic diagram of Faster-RCNN structure

IV. Helmet Detection based on Deep Learning

A. VOC 2007 dataset testing

Before training and testing with experimental data sets, training and testing using existing official VOC2007 data sets have achieved good results. Using two 1080XP graphics cards for training and verification, the following figure 5.1 as an example, the main targets in the picture are cats and dogs, from the detection results, we can see that the detection frame almost completely frames the whole target object, and the detection rate of puppies is 0.982, and the detection rate of cats is 0.992. It can be seen that the detection rate of the whole neural network is very high. Figure 5.2 shows the final test results. It includes the detection rate of all kinds of target objects in VOC2007 data sets. It can be seen that the car with the best average detection effect has an average detection accuracy of 0.80, while the pots, bottles and boats have a relatively poor detection effect, and their average detection rates are only 0.41, 0.51 and 0.55, respectively.



dog detections with p(dog | box) >= 0.8
dog 0.982
cat 0.992
person detections with p(person | box) >= 0.8
person 0.998
person 0.988

Figure 3. VOC2007 detection image

As can be seen from figure 4, the average detection accuracy mAP (mean Average Precision) of all kinds of targets is 0.676.

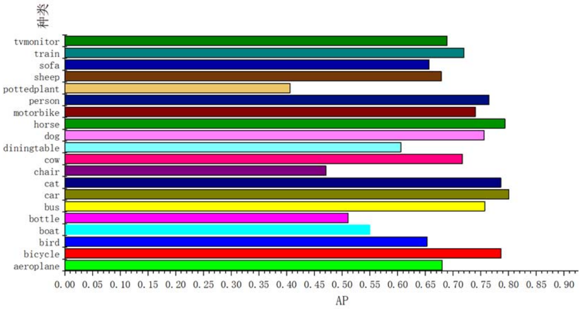


Figure 4. Statistical chart of detection accuracy of VOC 2007

B. Build a helmet data set

In order to collect data, it is processed by a web crawler to collect appropriate pictures on Google. Collect 4500 pictures, enter the LabelImg tag, and generate the PASCAL dataset.

1. Select the suitable image for the data set (including the image of the helmet) from the image obtained by the crawler;

2. Rename the selected image to 6 digits in a format of 00*;

3. Open the selected images in LabelImge and mark them one by one;

4. Using the compiled program, the XML generated by the tag is grouped and divided into three groups: training, testing and verification;

5. Put the image, the XML file containing the tag result, and the txt file with the image name in the same folder.

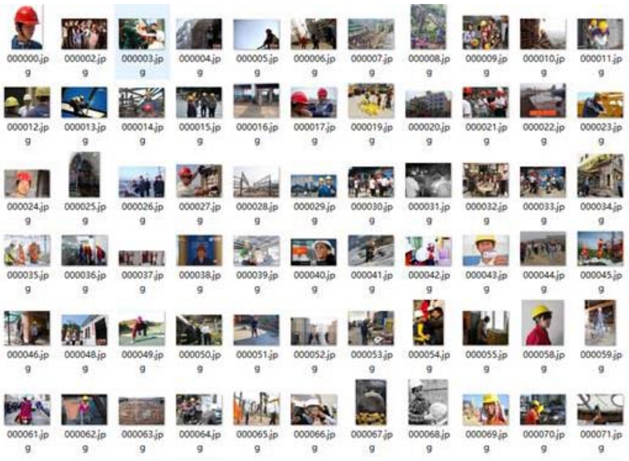


Figure 5. Image data set obtained from Google crawler

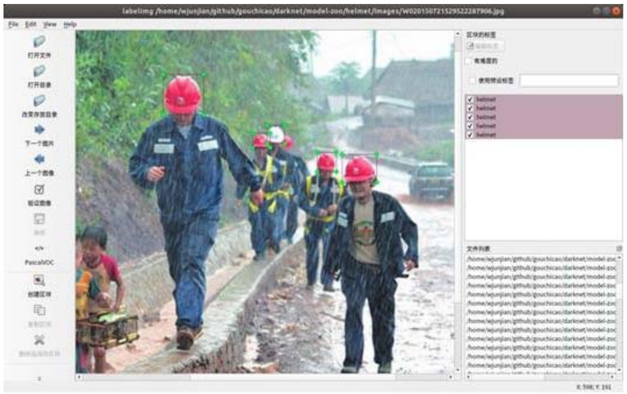

labelImg /home/wjunjian/github/gouchicao/darknet/model-zoo/helmet/images/W020150721529522287906.jpg
File Edit View Help
PascalVOC
helmet
helmet
helmet
helmet
helmet

Figure 6. Using LabelImge for Image Annotation

The data set is inputted into the convolution neural network, and the model with the highest accuracy is obtained after training and screening, and the average detection accuracy is obtained by testing the test set. The initial feature extraction network is initialized using the classified samples of ImageNet, and the rest of the new layers are initialized randomly. Each mini-batch contains 256 anchor, foreground background samples extracted from an image. The first 60K iteration, the learning rate is 0.001, and the last 10K iteration, the learning rate is 0.0001. One training model is saved for every 5000 times of training.

According to the loss curve and operation ability, the training times was determined to be 70 000 times. The map, statistics of each target are tested with 50% of the test data set, as shown in the following table. It can be seen that the map, with map higher than the official VOC data set proves that the training effect is better. Add tensorboard to the training and test, record the changes of various parameters, and count the rpn_loss,box_loss and total loss curves as shown in the following figure. Finally, the training and test flow chart is generated.

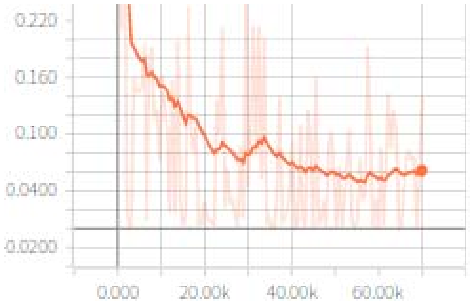


Figure 7. Rpn_loss_box curve of training data set

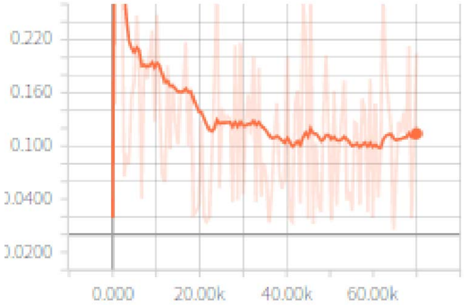


Figure 8. Loss_box curve of training data set

The data set is divided according to the proportion of training, verification and testing at 6:2:2. After the final test, the AP value of employees without helmets is 0.8087, the AP value of employees with helmets is 0.6155, and the final map value is 0.7121, which is higher than the detection effect on PASCAL data sets. The following picture shows the test effect of the image captured by the camera on the spot, which shows that the detection effect of the helmet is excellent.

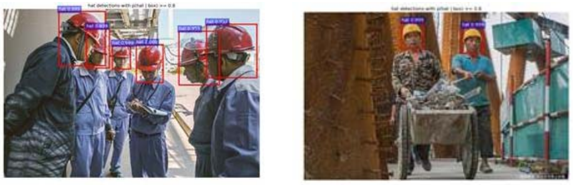


Figure 9. Faster-RCNN data test effect diagram

V. CONCLUSION

Based on the method of deep learning, a practical safety helmet wearing detection system is developed in this paper, which can be used to judge whether the operators of steel mills wear safety helmets or not. Good results have been obtained in the actual test. A large number of experimental results show that the safety helmet wearing detection system is effective and efficient. Future work will focus on helmet detection in scenes such as weak light and complex backgrounds.

ACKNOWLEDGMENT

This work is partly funded by National Key Research and Development Program of China via grant 2018YFC0806802 and 2018YFC0832105.

REFERENCES

[1] R. Waranusast, N. Bundon, V. Timtong, C. Tangnoi and P. Pattanathaburt, "Machine vision techniques for motorcycle safety helmet detection," 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), Wellington, 2013, pp. 35-40, doi: 10.1109/IVCNZ.2013.6726989.

[2] R. R. V. e. Silva, K. R. T. Aires and R. d. M. S. Veras, "Helmet Detection on Motorcyclists Using Image Descriptors and Classifiers," 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, Rio de Janeiro, 2014, pp. 141-148, doi: 10.1109/SIBGRAPI.2014.28.

[3] S. Q. Huang, Research and application of intelligent video analysis algorithm in substation, 2012.

[4] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[5] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[7] S. Zhang, Y. Wu, C. Men and X. Li, "Tiny YOLO Optimization Oriented Bus Passenger Object Detection," in Chinese Journal of Electronics, vol. 29, no. 1, pp. 132-138, 1 2020, doi: 10.1049/cje.2019.11.002.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[9] D. Forsyth, "Object Detection with Discriminatively Trained Part-Based Models," in Computer, vol. 47, no. 2, pp. 6-7, Feb. 2014, doi: 10.1109/MC.2014.42.

[10] C. P. Papageorgiou, M. Oren and T. Poggio, "A general framework for object detection," Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 1998, pp. 555-562, doi: 10.1109/ICCV.1998.710772.

[11] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu and Qi Tian, "Statistical modeling of complex backgrounds for foreground object detection," in IEEE Transactions on Image Processing, vol. 13, no. 11, pp. 1459-1472, Nov. 2004, doi: 10.1109/TIP.2004.836169.

[12] O. Sharma, "Deep Challenges Associated with Deep Learning," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 72-75, doi: 10.1109/COMITCon.2019.8862453.