Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Research on safety helmet wearing YOLO-V3 detection technology improvement in mine environment

To cite this article: Fuchuan and Wang RongXin 2019 *J. Phys.: Conf. Ser.* **1345** 042045

Research on safety helmet wearing YOLO-V3 detection technology improvement in mine environment

Fuchuan[1*], WangRongXin[2]

[1]School of Civil Engineering, North China University of Technology, Beijing 100144, China

[2]Hewlett-Packard Enterprise (China) Co., Ltd, Beijing 100102, China

[*]Corresponding author's e-mail:fcjobsuch@sina.com

Abstract. The existing AI object detection technology cannot meet the demand of safety helmet wearing detection accuracy in mine environment. This paper studies YOLO correlation algorithm, establishing an optimal model based on YOLO-V3, combines the deep residual network technology with the multi-scale convolution feature based on the YOLO-V3 detection algorithm, combines the multi-scale detection training and adjusts the loss function in the training process. The experimental results show that with satisfying the detection speed, safety helmet wearing detection accuracy in mine environment is significantly improved.

1. Introduction

Mines always attach great importance to safety production, and safety helmet wearing is one of the important measures to prevent safety risks. At present, video image recognition technology based on shape and colour features is affected by dynamic changes of workers and the light of working scene. The detection speed (second level) and accuracy (less than 50%) are difficult to meet the actual demand of safety helmet wearing detection in mine harsh environment.

In recent years, due to the continuous progress of deep learning research, Girshick [1] and Ren [2] have respectively proposed Fast Regional Convolution Neural Network (Fast R-CNN) and Faster Regional Convolution Neural Network (Faster R-CNN), which have improved the accuracy and real-time detection speed, but there is a certain gap between them. In 2015, Redmon J [3] proposed YOLO object detection algorithm, which balanced the accuracy and detection speed. In 2016, Redmon created YOLO-V2 [4] and YOLO-V3 [5] detection algorithms through improvement. YOLO-V2 focused on small object detection, increased the mean accuracy of mAP (mean average precision) by 2%. The latest YOLO-V3 further strengthened multi-label classification and network architecture, taking into account both accuracy and detection speed, which has good detection effect in construction and other fields. But there are still some deficiencies in detection accuracy in mine environment. In this paper, based on the YOLO-V3 detection algorithm, cluster algorithm is used to predict the target frame of the helmet, and then the accuracy is optimized through the combination of depth residual network and multi-scale detection training in the training process. Finally, the target frame is optimized by adjusting the weight of the loss function. Experiments show that this improvement can significantly improve safety helmet wearing detection accuracy in mine environment.

2. Principle and Improvement of YOLO-V3

YOLO-V3 uses Darknet-53 network structure based on V1 and V2 to extract features through 53 convolution layers and 5 maximum pooling layers, adding batch normalization and dropout removal operations to prevent over-fitting. YOLO-V3 pre-detection system uses classifiers to perform detection tasks many times, and applies the model to multiple positions and proportions of images. Those areas with higher scores are regarded as detection results. Although YOLO-V3 enhances the accuracy of small object detection, the recognition accuracy of safety helmet (mAP-50 between 50 and 60) cannot meet the commercial requirements under complex mine background and various motion postures. Considering that the application scenario mainly recognizes the fixed recognition features of the helmet and human body, it is improved through the following three aspects.

2.1. Predict and obtain a priori box of safety helmet by clustering algorithm, and classify scale features.

In the repeated training process of supervised learning, by increasing the number of iterations, candidate boxes can be adjusted to approximate the real boxes gradually. In this paper, K-means clustering algorithm is used to find the distribution rule by clustering statistical characteristics [7]. By clustering analysis of the size of candidate boxes, the actual candidate boxes are evenly distributed to three different sizes of feature maps, as shown in the figure 1.

Figure 1. Candidate box size to detect

Generally, Euclidean distance is easy to cause large box errors in clustering. In this paper, overlap IOU is used to eliminate the error caused by the size of candidate box [8]. The overlap degree algorithm divides the intersection of candidate box and real box by union. The distance formula is

$$\mathrm{d}(\mathrm{box}, \mathrm{centroid}) = 1 - \mathrm{IOU}(\mathrm{box}, \mathrm{centroid}) \tag{1}$$

where, centroid denotes the center of cluster; box denotes sample; IOU (box, centroid) denotes the intersection and parallelism of the center frame of cluster and cluster frame. The intersection and merge ratio IOU is used to measure the accuracy of detection box.

$$\mathrm{IOU}\left(b_{gt}, b_{dt}\right) = \frac{b_{gt} \cap b_{dt}}{b_{gt} \cup b_{dt}} \tag{2}$$

where b_{gt} represents the real box and b_{dt} represents the prediction box.

2.2. Optimizing accuracy by combining deep residual network with multi-scale detection training

The difficulty of mine safety helmet recognition lies in the distinction between safety helmet and background edge under insufficient light. In this paper, DF-YOLOv3 depth residual network combined with multi-scale detection training is used to optimize the accuracy. DF-YOLOv3 adds several convolution layers according to the size of the residual network. Considering the semantic relevance, DF-YOLOv3 superimposes two feature maps (prediction network and depth residual network). Finally, the maximum method is used to remove the repeated target selection box and calculate the optimal detection results. Considering that multi-scale detection method has better detection effect for high-resolution images, this paper balances detection speed and accuracy, and selects different resolution video pictures for comparison. Figure 2 shows multi-scale detection training.

Journal of Physics: Conference Series **1345** (2019) 042045 doi:10.1088/1742-6596/1345/4/042045

Figure 2. Multi-scale detection training by 13*13, 26*26, 52*52

2.3. Optimize the selection of target box by adjusting the weight of loss function

In the process of screening and predicting targets in the above complex number criteria selection box, the selection of repetitive boundaries is mainly suppressed by non-maxima. In particular scenarios, the original loss function is not very effective. In this paper, we use the mean square and error values of dimension vectors and image truth values to improve the loss function quantization method [9]. Loss function coefficients with and without target border can be set separately, loss coefficients of prediction frame and loss coefficients of judgment category. Through these optimization adjustments, loss coefficients of prediction frame containing detection object can be given higher weight to achieve the purpose of balancing different prediction scenarios. The formula of specific loss function is given

$$loss = \lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}[(x_i-\hat{x}_i)^2+(y_i-y_i)^2]+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\left(\sqrt{w_i}-\sqrt{\widehat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\right]$$

$$+\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left(C_i-\hat{C}_i\right)^2+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{noobj}\left(C_i-\hat{C}_i\right)^2+\sum_{i=0}^{S^2}1_{ij}^{obj}\sum_{c\in classes}\left(p_i(c)-\hat{p}_i(c)\right)^2 \quad (3)$$

The function of parameter λ_{coord} is to strengthen the weight of the prediction box containing the target in the loss function; the function of parameter λ_{noobj} is to weaken the weight of the prediction box without the target in the loss function; C_i is the reliability score of the prediction box; $p_i(c)$ is the class probability of the target condition. 1_{ij}^{obj} denotes that the detected object appears in the No.j box of the No.i grid with a value of 1, otherwise 0. The convolution layer weight is enhanced by controlling the predicted loss value, that is, the tan loss is used to control the gradient change to control the loss weight more effectively. Its derivative form is

$$\text{LossGrad} = \left(sec^2(x)*tan(x)\right)/\left(sec^2(1)*tan(1)\right) \quad (4)$$

3. Experimental results and comparative analysis

3.1. Comparison of experimental results

(1) *Detection at different scales*

Firstly, based on safety helmet wearing detection requirement, five typical working scene images of underground mining process in coalmine and test with different scale network model are selected. The test accuracy and detection results of YOLO-V3-416 are show in Figure 3 and in Figure 4.

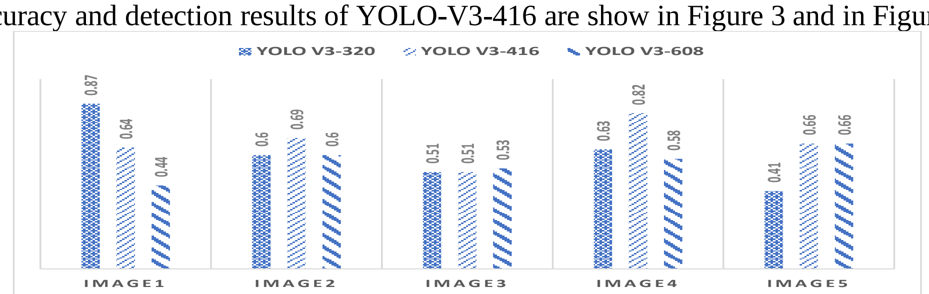

Figure 3. Analysis of the results of different scales

Figure 4. Image Detection results at YOLO-V3-416 scale

From the above results, it can be seen that the recognition accuracy of detection image 1 varies greatly. The input image scale is the highest at 320*320 and the lowest at 608*608. The recognition accuracy of detection image 2 and 4 is the highest at 416*416. The overall recognition accuracy of detection image 3 and 5 does not change much, while that of detection image 5 is lower at 320*320.

(2) Accuracy comparison of algorithms in different scenarios

Based on the above verification analysis of different scales and resolutions, aiming at the detection requirements of different scenarios in mine environment, this paper compares the effect of the improved algorithm for single object detection, multi-object detection and complex scenario detection based on the conclusion that the overall accuracy of YOLOV3-416 is the best.

Figure 5. Accuracy comparison of algorithms in different scenarios

Figure 6-8 show the detection results of the same scene before and after improvement. For single object detection the detection accuracy has been improved from 0.58 to 0.73, which achieves the desired results. For multi- detection the detection accuracy of the three main objects in the image has changed from 0.77, 0.88, 0.87 to 0.92, 0.86 and 0.73, respectively. Although the overall average accuracy is still 0.84, through the improvement, the near recognition rate has been significantly improved, and the accuracy has changed from far and near to rising trend. For object detection in complex scenes, it can be seen that the recognition accuracy increases from 0.62, 0.85 to 0.91 and 0.87 respectively when there is interference from the same or similar colour objects or partial overlap of characters.

3.2. Analysis of results

(1) Analysis of the results of different scales

According to the number of feature layers and the results of detection, when the feature layer is large, it is suitable for detecting larger objects or objects, and when the feature layer is small, it is suitable for detecting smaller objects or objects. The safety helmet of this detection is a medium and long distance panoramic shooting in the surveillance video, belonging to the target of small and medium objects. Therefore, 416*416 has the highest recognition accuracy in the overall detection results of image input scale. Based on the results of this verification, the input scale can be selected according to the distance between the detected object and the camera in the surveillance video.

Figure 6. Single object detection results comparison before and after improvement

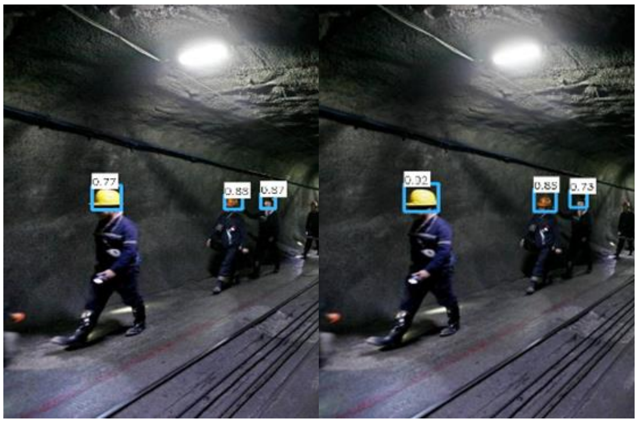

Figure 7. Multi-object detection results comparison before and after improvement

Figure 8. In complex scenes object detection results comparison before and after improvement

(2) Comparing and Analyzing the Accuracy of Algorithms in Different Scenarios

The single object detection results validate the optimization effect of YOLO-V3. The identification accuracy of the safety helmet is improved obviously by means of the method of enhancing and reducing the weight parameters of the loss function with or without targets in the detection frame. The detection accuracy is increased from 0.58 to 0.73 and 15%.

The multi-object detection scenario chooses the channel entrance, and focuses on the distance detection to improve the detection of close points close to the camera. By clustering statistical target distribution and size, overlap IOU is used in clustering analysis to eliminate the error caused by the size of candidate box. In order to ensure that the recognition rate of the close point is obviously improved, but the recognition effect of the long-distance point will be sacrificed, and the recognition ability will be improved when the target of the next frame of long-distance point recognition enters. In this test, the precision of short distance detection is improved from 0.77 to 0.92, which is 19.5%, and the improvement effect is obvious.

In complex scene detection scenarios, it is difficult to recognize objects with the same or similar colours or partially overlapping characters. For one thing, we should strengthen the training of multi-dimensional models such as colour, exposure and rotation, and for another thing, we should enhance the recognition accuracy of small objects or minimal objects on the working face. It can be seen from the test results that although the recognition accuracy of the close point is improved from 0.85 to 0.87, only 2%, the recognition accuracy of the long-distance point is improved from 0.62 to 0.91, and the improvement effect is obvious.

4. Conclusion

This paper improves and validates the real-time detection method of safety helmet wearing in mine environment. Based on YOLO-V3 detection model, the accuracy of safety helmet wearing detection is improved by combining deep residual network technology with multi-scale convolution feature, combining multi-scale detection training in the training process and adjusting the weight of loss function, on the premise of satisfying detection rate, especially in complex underground environment. The recall rate of detection has been significantly improved, and the detection accuracy of single object in mine environment has been improved from 0.58 to 0.73, with an increase of 15%. The recognition accuracy of long-distance points in complex scenes is improved from 0.62 to 0.91, which is 29%. It meets the commercial safety requirement of real-time monitoring of safety helmet wearing in complex working environment of mines.

Acknowledgments

We sincerely appreciate the Fellowships received from the Fundamental Research Business Fee of Beijing Universities (110052971921/064).

References

[1] Girshick, R. (2015) Fast R-CNN[C]//IEEE InternationalConference on Computer Vision. IEEE, 2015:1440-1448.

[2] Ren, S., He, K., Girshick, R., et al. (2015) Faster R-CNN: towardsreal-time object detection with region proposalnetworks[C]//International Conference onNeural Information Processing Systems. MIT Press, 2015: 91-99.

[3] Redmon, J., Divvala, S., Girshick, R., et al. (2015) You OnlyLook Once: Unified, Real-Time Object Detection. J.2015: 779-788.

[4] Redmon, J., Farhadi. A., (2017) YOLO9000: Better, Faster,Stronger[C]//IEEE Conference on Computer Visionand Pattern Recognition. IEEE, 2017:6517-6525.

[5] Redmon, J., Farhadi, A., (2018) YOLOv3: An IncrementalImprovement.J. IEEE Conference on ComputerVision and Pattern Recognition. IEEE, 2018:89-95.

[6] Krizhevsky, A., Sutskever, I., Hinton, G.E., (2012) ImageNetclassification with deep convolutional neural networks[C]//International Conference on Neural InformationProcessing Systems. Curran AssociatesInc. 2012:1097-1105.

[7] Shi, H.,Chen, X.Q.,Yang, Y.,Safety Helmet Wearing Detection Method of Improved YOLO v3. J. Computer Engineering and Applications.2019.3.

[8] Wang, D.W,H, Y.H., Li, D.X., Liu, Y., Xu, Z.J.,Wang, J.. (2018) An Improved infrared Video Image Pedestrian Algorithm. J.Journal of Xian University of Posts and Telecommunications. J.23(4):48-52.

[9] Ma, Q.,Zhu, B,Zhang, H.W.,Zhang, Y., Jiang, Y.C..(2019)Low-altitude UAV Detection and Recognition Method Based on Optimized YOLOv3.J. Laser & Optoelectronics Progress, 56(20).