IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 26, 2019, accepted July 7, 2019, date of publication July 15, 2019, date of current version September 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928603

A Target Detection Model Based on Improved Tiny-Yolov3 Under the Environment of Mining Truck

DONG XIAO[1,4], **FENG SHAN**[1], **ZE LI**[1], **BA TUAN LE**[2], **XIWEN LIU**[1], **AND XUERAO LI**[3]

[1]School of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[2]Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
[3]School of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China
[4]Liaoning Key Laboratory of Intelligent Diagnosis and Safety for Metallurgical Industry, Northeastern University, Shenyang 110819, China

Corresponding author: Dong Xiao (xiaodong@ise.neu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773105, Grant 61374147, Grant 41371437, Grant 61203214, Grant 61473072, and Grant 61733003, in part by the Fundamental Research Funds for the Central Universities, China, under Grant N182608003, Grant N182008004, Grant N180404012, and Grant N160404008, in part by the Northeastern University Undergraduate Education and Teaching Reform Project under Grant 02100021301026, in part by the Liaoning Natural Science Foundation, in part by the Fundamental Research Funds for the Central Universities under Grant N180704013, Grant N150402001, and Grant N172410005-7, in part by the Fundamental Research Funds for Liaoning Natural Science Foundation under Grant 2019-MS-120, in part by the National Key Research and Development Plan under Grant 2016YFC0801602, and in part by the National Training Program of Innovation and Entrepreneurship for Undergraduates under Grant 201810145161.

ABSTRACT In the open pit mine production systems, a certain number of trucks transport mine and rock between the power shovel and the unloading point. Due to the mining truck has characteristics of high height, long width and big size, it has a large blind zone and a long braking distance. Therefore, the probability of accidents in mining trucks is high, which results in huge loss of manpower, material resources and financial resources. In this paper, tiny-yolov3 is used to detect obstacles in the mine, its real-time performance is high enough, but the detection accuracy is not ideal. Therefore, this paper proposes an improved target detection model based on tiny-yolov3. The residual network structure based on convolutional neural network is added to the tiny-yolov3 structure, and the accuracy of obstacle detection is improved under the condition of real-time detection. The experimental results show that compared with tiny-yolov3, the detect precision of tiny-yolov3 with residual structure is improved, and the detection speed is reduced slightly, there is no particular impact on the real-time nature of the entire algorithm.

INDEX TERMS Convolutional neural network, real-time, residual network, target detection, tiny-yolov3.

I. INTRODUCTION

In most open pit mines, due to labor shortages and high labor costs, on the one hand, large number of mining trucks are sought to reduce the demand of personnel, and on the other hand, better solutions of this problem are sought. With the development of technology, unmanned mining trucks [1] came into being. Unmanned mining trucks increase safety and productivity while saving labor costs and become an integral part of mine digitization. In unmanned mining trucks technology, active anti-collision technology [2] is a very important part. Due to the high, wide and large characteristics of the mining truck, the blind zone is large and its braking distance is long [3]. The truck driver will inevitably have traffic accidents due to fatigue or negligence, so the implement of truck active anti-collision technology is also necessary in the case of manned driving. At present, measures to reduce the "blind zone" of mining trucks include equipping high flagpoles on assisting vehicles, using convex lenses, using camera probes and reversing radars [4], using anti-collision warning electronic devices and so on. Anti-collision warning electronic devices use GPS to locate the position of targets and remind the driver to avoid danger [5]. However, using convex lens can enlarge the field of view, but there is still a blind spot; using radar to detect target only applicable to the flat, it is not suitable for strip-shaped objects, and it can only measure objects on the straight line. What's more, large mining trucks need to install multiple radars or which increased the cost. Nowadays early warning system which using ultrasonic and image are used in mining trucks either, although it can measure the distance of the target, it cannot classify the class of the target. Due to the deep pit of the open

The associate editor coordinating the review of this article and approving it for publication was Kemal Polat.

pit mine, the use of GPS positioning may result in positioning failure or false alarm due to the presence of signal dead zones. Therefore, this paper uses computer vision to achieve the detection and location of obstacles in the mine to achieve active anti-collision technology.

Target detection [6] is a premise for advanced visual tasks such as scene content understanding, and has been applied to tasks such as intelligent video surveillance, content-based image retrieval, robot navigation, and enhanced implementation. In recent years, deep convolutional networks have made breakthroughs in the field of computer vision. The deep convolutional network mainly deepens the network level through the weight sharing strategy, which makes the network have stronger resolution capabilities. VGG network, GoogleNet [7] and residual network [8] push the convolution network to a deeper level, which greatly improve the performance of the network, and increase the accuracy of large-scale image classification to a very high level. The Region Based Convolutional Neural Network (RCNN) raises the accuracy of target detection to a new level. Since the RCNN is divided into three separate processes, the detection efficiency is very low. YOLO (You Only Look One) [10] and SSD (Single Shot Multibox Detector) [11] are proposed to improve the detection efficiency of target detection, and try to achieve the degree of real-time detection of target detection. Compared with SSD, the newly proposed tiny-yolov3 [12] model greatly improves the detection accuracy under the condition of ensuring real-time performance. Although the detection accuracy of yolov3 and SSD is very high, their real-time performance on low performance devices or PCs is not ideal. Tiny-yolov3 is a simplified model of YOLOV3, which reduces the depth of the convolutional layer. Although the detection accuracy has decreased, the running speed has been greatly improved. Therefore, this paper proposes an improved tiny-yolov3 model to improve the accuracy of model detection when the real-time performance is slightly reduced, but still meets the requirements.

II. DATASET OBTAINING AND DATA ENHANCEMENT WITH SPECIFIC CLASS

The dataset used in this paper consists of VOC2007 [13], VOC2012 [14], the image of the target collected from a bird's eye view camera and the mining truck. The annotation file of the image containing the target captured from a bird's eye view camera and the mining truck need to be made for model training. Due to the limited number of images of mining trucks collected at the mine, data enhancement is also required for the class of mining truck.

A. USING LABELIMG TO MAKE XML FILES

The tiny-yolov3 model used in this paper obtains the storage location of the training image and the marked pixel location of the target in the image and the class of the target by reading the txt text. This txt file is generated by reading a dataset which using the VOC format. Therefore, the pictures collected in the scene are made into a dataset in the form of VOC

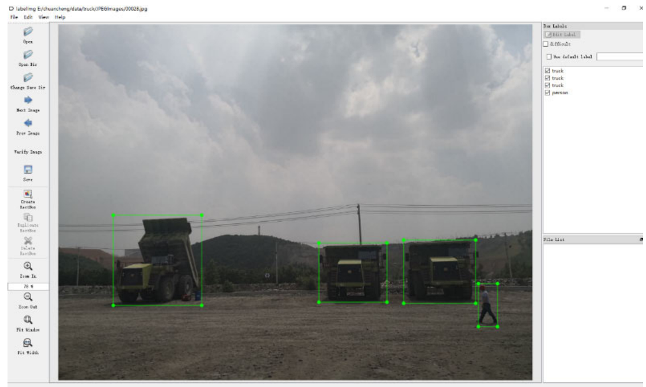


FIGURE 1. Using LabelImg to label the target and then a xml file will be saved which contains position information of the target.

FIGURE 2. Image data enhancement.

by using a software called LabelImg. LabelImg is an image annotation tool, which saves the object class information and position information marked in the image as a file in xml format for training. The figure 1 shows the process of labeling the image. After using the rectangle to mark the target and select the target class, we need click Save and the software will generate the xml format label text with the same name as the image.

B. DATA ENHANCEMENT

Due to the limited number of pictures of the mining trucks obtained at the mine, the number of samples in this class is very different from other targets. When the number of one sample is too small during training, the accuracy of detection is very random, so data enhancement is only used in the mining truck class to extend the dataset. Common image data enhancement methods include flipping, rotating, panning, zooming, random contrast, etc. [15]. In this paper, the image data is enhanced by 90°, 180°, 270° rotation and horizontal flipping of the original picture and the corresponding rectangular frame of the annotation file, and the number of the data is expanded to 5 times. The data enhancement effect diagram is shown in Figure 2.

C. IMPROVED TINY-YOLOV3 MODEL

This paper proposes a tiny-yolov3 model based on deep convolutional neural network with residual network structure. Based on the tiny-yolov3 model, the model adds residual network structures between the original convolutional layers, which helps the network to further extract features from the target and reduce information loss when information is transmitted between deep convolutional layers. In this way, the accuracy of the model detection will increase.

III. IMPROVED TINY-YOLOV3 MODEL

A. K-MEANS ALGORITHM FOR CALCULATING YOLO ANCHORS

The K-means algorithm [16] is an unsupervised clustering algorithm, which is easy and has a good clustering effect, so it is used widely. It uses distance as the criterion for the measure of similarity between data objects. The smaller the distance between data objects, the higher their similarity, the more likely they are in the same cluster. The steps of the K-means algorithm are:

(1) Selecting an initial cluster center for each cluster;

(2) Assign the sample set to the nearest neighbor cluster according to the principle of minimum distance;

(3) Updating the cluster center using the mean of each cluster;

(4) Repeat steps (2) and (3) until the cluster center no longer changes;

(5) Output the final cluster center and k cluster divisions;

The K-means algorithm usually uses the Euclidean distance to calculate the distance between data objects. In this paper, if the Euclidean distance is calculated by anchors, calculating a large anchor box will produce more errors than the small prediction box. So there are other ways to calculate the distance between data objects. Intersection over Union (IOU) is a standard for measuring the accuracy of detecting corresponding objects in a specific data set. Its calculation formula is shown in equation (1):

$$IOU = \frac{TP}{FP + TP + FN} \tag{1}$$

where TP, FP and FN represent true positive, false positive and false negative counts, respectively. In this paper, the probability value is used to approximate the IOU, and the calculation formula is shown in equation (2):

$$IOU = \frac{I(X)}{U(X)} \tag{2}$$

where I(X) and U(X) are calculated as shown in equations (3) and (4):

$$I(X) = \sum_{\mathrm{v} \in V} X_{\mathrm{v}} * Y_{\mathrm{v}} \tag{3}$$

$$U(X) = \sum_{\mathrm{v} \in V} (X_{\mathrm{v}} + Y_{\mathrm{v}} - X_{\mathrm{v}} * Y_{\mathrm{v}}) \tag{4}$$

where V $= \{1, 2, \ldots, N\}$ is the set of all pixels of all images in the training set, X is the output of the network of the probability of pixels on the set V, and $Y \in \{0, 1\}^V$ is the ground truth assignment of the set V, 0 represents the background pixels, 1 represents the object pixel.

The purpose of clustering is that the anchor boxes and the adjacent ground truth have larger IOU values, so the method of calculating the distance in this paper is shown in equation (5):

$$d(box, centriod) = 1 - IOU(box, centriod) \tag{5}$$

In addition, tiny-yolov3 is different from yolov3 and requires 6 anchors, so the k value in the K-means algorithm is chosen to be 6 instead of 9.

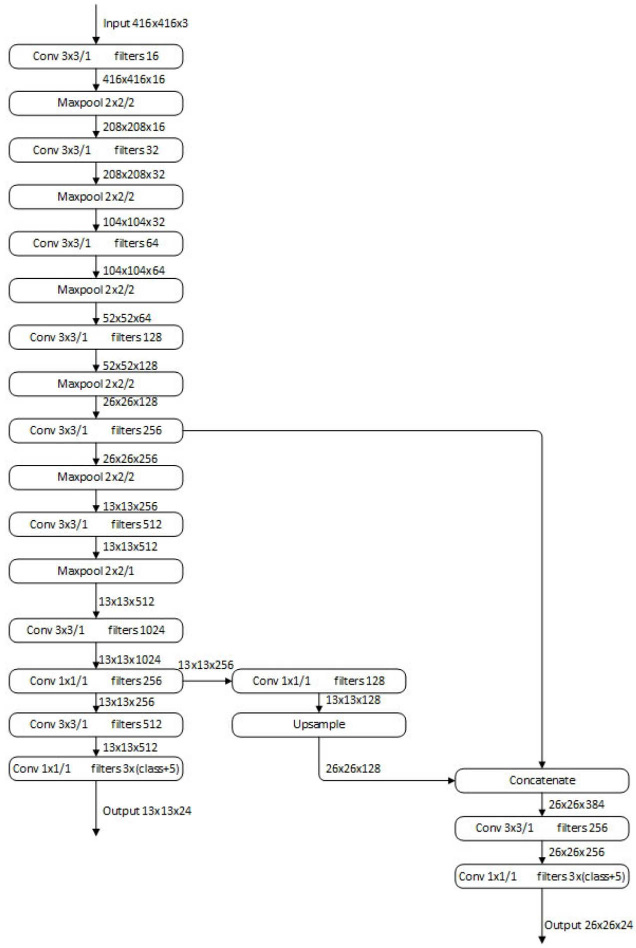


FIGURE 3. The structure of tiny-yolov3.

B. ORIGINAL TINY-YOLOV3 MODEL

The tiny-yolov3 model is a simplified version of the YOLOV3 model. YOLOV3 uses the architecture of darknet53, and then uses many 1x1 and 3x3 convolution kernels to extract features. Tiny-yolov3 reduces the number of convolutional layers, its basic structure has only 7 convolutional layers, and then features are extracted by using a small number of 1x1 and 3x3 convolutional layers. Tiny-yolov3 uses the pooling layer instead of YOLOV3's convolutional layer with a step size of 2 to achieve dimensionality reduction. However, its convolutional layer structure still uses the same structure of Convolution2D+BatchNormalization+LeakyRelu as YOLOV3. The structure of tiny-yolov3 is shown in Figure 3.

When the model is trained, the loss function used by tiny-yolov3 is the same as that of YOLOV3, which mainly composed by the position of the prediction frame (x, y), the prediction frame size (w, h), the prediction class (class), and the prediction confidence (confidence). The expression of tiny-yolov3 is shown in equation (6):

$$loss = \frac{1}{n}\sum_{i=1}^{n} loss_{xy} + \frac{1}{n}\sum_{i=1}^{n} loss_{wh} + \frac{1}{n}\sum_{i=1}^{n} loss_{class} + \frac{1}{n}\sum_{i=1}^{n} loss_{confidence} \quad (6)$$

where n is the total number of targets trained, and the loss function for each factor in equation (6) is as follows:

$$loss_{xy} = objectmask * (2 - w * h) * binarycrossentropy(truexy, predxy) \quad (7)$$

$$loss_{wh} = objectmask * (2 - w * h) * 0.5 * square(truewh, predwh) \quad (8)$$

$$loss_{class} = objectmask * binarycrossentropy(trueclass, predclass) \quad (9)$$

$$loss_{confidence} = objectmask * binarycrossentropy(objectmask, predmask) + (1 - objectmask) * binarycrossentropy(objectmask, predmask) * ignoremask \quad (10)$$

where objectmask refers to the point of the object, (w, h) is the width and height of the prediction box, bianrycrossentropy is a binary cross entropy function, and square is a function of variance. Truexy is the actual target position, predxy is the predicted position; truewh is the actual groundtruth box size, predwh is the prediction frame size; trueclass is the actual target class, predclass is the prediction class; predmask is the predicted object point, ignoremask is related to IOU, if IOU is less than the threshold, ignoremask is 0.

C. THE TINY-YOLOV3 MODEL AFTER ADDING THE RESIDUAL NETWORK STRUCTURE

Although tiny-yolov3 has a much higher speed than YOLOV3, its detection accuracy is greatly reduced. Therefore, this paper proposes an improved tiny-yolov3 model, which improves the accuracy of detection when the detection speed is slightly reduced. Although the frame rate of the program is slightly reduced, it still satisfies the real-time performance of the system. The most common way to improve the detection accuracy of the deep convolutional network is to increase the depth of the network, that is, increase the number of the convolutional layers. The deeper the network, the less likely it is to converge, and the features of small objects extracted by shallow networks are diluted as the network deepens. If the network is too deep, it will also cause feature information losing when they were passed between layers. Therefore, this paper adds residual network structures between the 4th to 7th layers of the original network. The residual network uses 1x1 convolutional layer and 3x3 convolutional layer to extract features. The feature map before inputting the structure is added to the feature map generated after the residual structure, and the shallow information and the deep information are simultaneously transmitted to the next convolution layer to extract feature. In this way, such loss of feature information when passing between layers will reduce, and the accuracy of network detection will improve. The structure of the improved tiny-yolov3 model is shown in Figure 4:

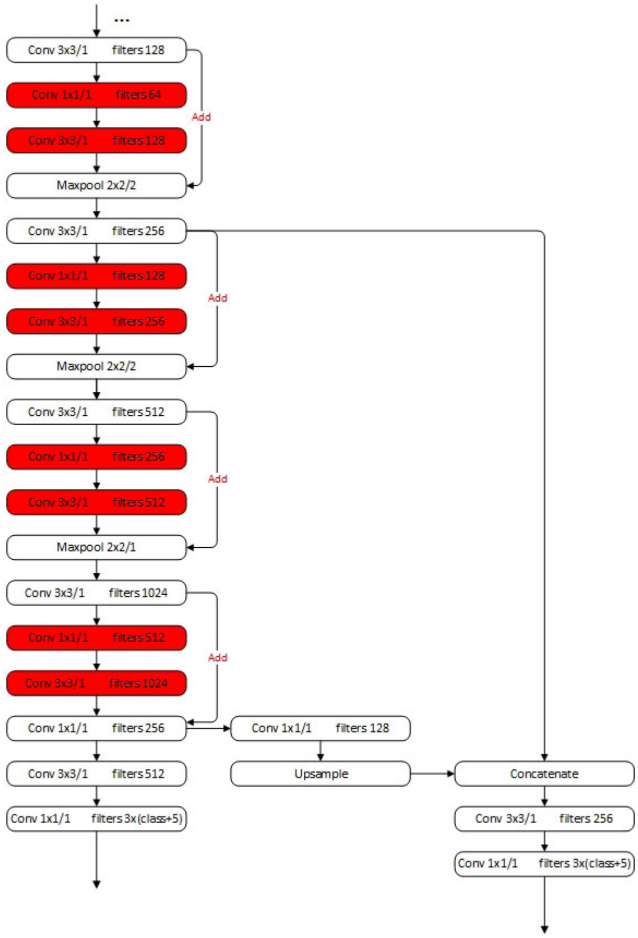


FIGURE 4. Improved tiny-yolov3 structure.

IV. EXPERIMENT

A. DATASET, TRAINING METHODS AND ENVIRONMENT

The dataset used in this paper is composed of: extracting people and cars from VOC2007 and VOC2012, and randomly selecting 50% of them as part of the dataset; pictures of people and cars taken by the camera in a bird's eye view; mining trucks, people and cars pictures at the mine. Among them, it is necessary to make data enhancements to the pictures containing the mining trucks to expand the number of training targets.

The hardware used in this experiment is a PC with Intel Core i5-7500 3.40GHz processor, NVIDIA GTX 1050 Ti graphics card, 16 GB RAM, 500GB Western Digital mechanical hard disk, Windows 10 64-bit system.

Programming language is Python, GPU accelerated library is CUDA9.0 and CUDNN7.0.

The training method used in this paper is: randomly take 90% of the pictures in the above data set as the training set, and the remaining 10% as the verification set. At first, the tiny-yolov3 preloaded model should be loaded for training which downloaded from the yolo official website. The training process is divided into two parts. The first part trains the training set for 50 epochs. In this part, the batchsize is 32. The training optimizer uses the adam optimization algorithm. The learning rate is 0.001, beta_1=0.9, beta_2=0.999; In the second part, the training set is retrained for 50 epochs. In this part, the batchsize is 16. The training optimizer uses the adam optimization algorithm. The training learning rate is 0.0001, the parameter beta_1=0.9, beta_2=0.999. After each training period of the training set, AP, mAP and loss of the verification set are calculated, the weights models are saved every five epochs, the model with the highest mAP and the lowest loss are also saved.

B. EVALUATION CRITERIA: MAP AND FPS

The mAP (Mean Average Precision) is an average AP value for multiple verification sets, and it is used as an indicator for measuring detection accuracy in target detection. AP (Average Precision) is the average accuracy. The value is the area enclosed by the Precision-recall curve and the coordinate axis. The Precision-recall curve is a two-dimensional curve with precision and recall as the vertical and horizontal axis coordinates. It is drawn by selecting the corresponding precision and recall rate for different thresholds. The calculation of precision is shown in equation (11):

$$precision = \frac{TP}{TP + FP} \tag{11}$$

where TP is the number that is correctly divided into positive examples, and FP is the number that is erroneously divided into positive examples.

The recall is calculated as shown in equation (12):

$$recall = \frac{TP}{TP + FN} = \frac{TP}{precision} \tag{12}$$

where FN is the number that is erroneously divided into negative examples. Through the above formula, the P-R curve can be drawn to calculate the AP value of a single class, and the average value of the AP values of all categories can calculated to obtain the mAP value of the entire model.

In addition to detection accuracy, another important performance indicator of the target detection algorithm is speed. Only fast speed can achieve real-time detection, which is extremely important for some application scenarios. A common indicator for evaluating speed is Frame Per Second (FPS), which is the number of pictures that can be processed per second.

C. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, the position where the residual network structure is added in the model and whether the add function is used in the residual structure are variables. First, the residual network structure is added after the third to sixth convolutional layers of the tiny-yolov3 model. Compared with the original model, the mAP and FPS of the model are shown in Table 1:

TABLE 1. The test results with whether add resnet in tiny-yolov3.

Network structure	FPS	mAP(%)
tiny-yolov3	25	62.0
tiny-yolov3(conv3-6+resnet)	24	65.4

It can be known from Table 1 that the residual network structure is added after the third layer convolution layer to the sixth layer convolution layer, the detection accuracy is improved from 62.0 to 65.4. Since the most important convolutional layer of tint-yolov3 is the first seven layers, and as the number of layer increases, the number of convolution kernels in the convolutional layer also increases. Since the number of convolution kernels is increased, the amount of calculation increases during detection, and the FPS is reduced from 25 frames to 24 frames. When using the same size of convolution kernels the more convolution kernels, the richer the extraction features. Therefore, this paper also adds the residual network structure from the fourth convolutional layer to the seventh convolutional layer. The mAP and FPS of the detection model are shown in Table 2:

TABLE 2. The test results with different layers add resnet.

Network structure	FPS	mAP(%)
tiny-yolov3	25	62.0
tiny-yolov3(conv3-6+resnet)	24	65.4
tiny-yolov3(conv4-7+resnet)	23	66.5

It can be known from Table 2 that the addition residual structure in the fourth to seventh convolution layers extracts more features than that in the third convolutional layer to the sixth convolutional layer. The accuracy is further improved from 65.4 to 66.5. Although the FPS is reduced by 2 frames compared with the original model, it still meets the requirements of real-time performance.

Further, based on the improved model, this paper also try to delete add function in residual structure, it means only 1x1 and 3x3 convolution kernels are added to further extract features. The experimental results are shown in Table 3:

TABLE 3. The test results with whether add resnet in tiny-yolov3.

Network structure	FPS	mAP(%)
tiny-yolov3(conv4-7+resnet)	23	66.5
tiny-yolov3 (conv4-7+conv)	23	65.1

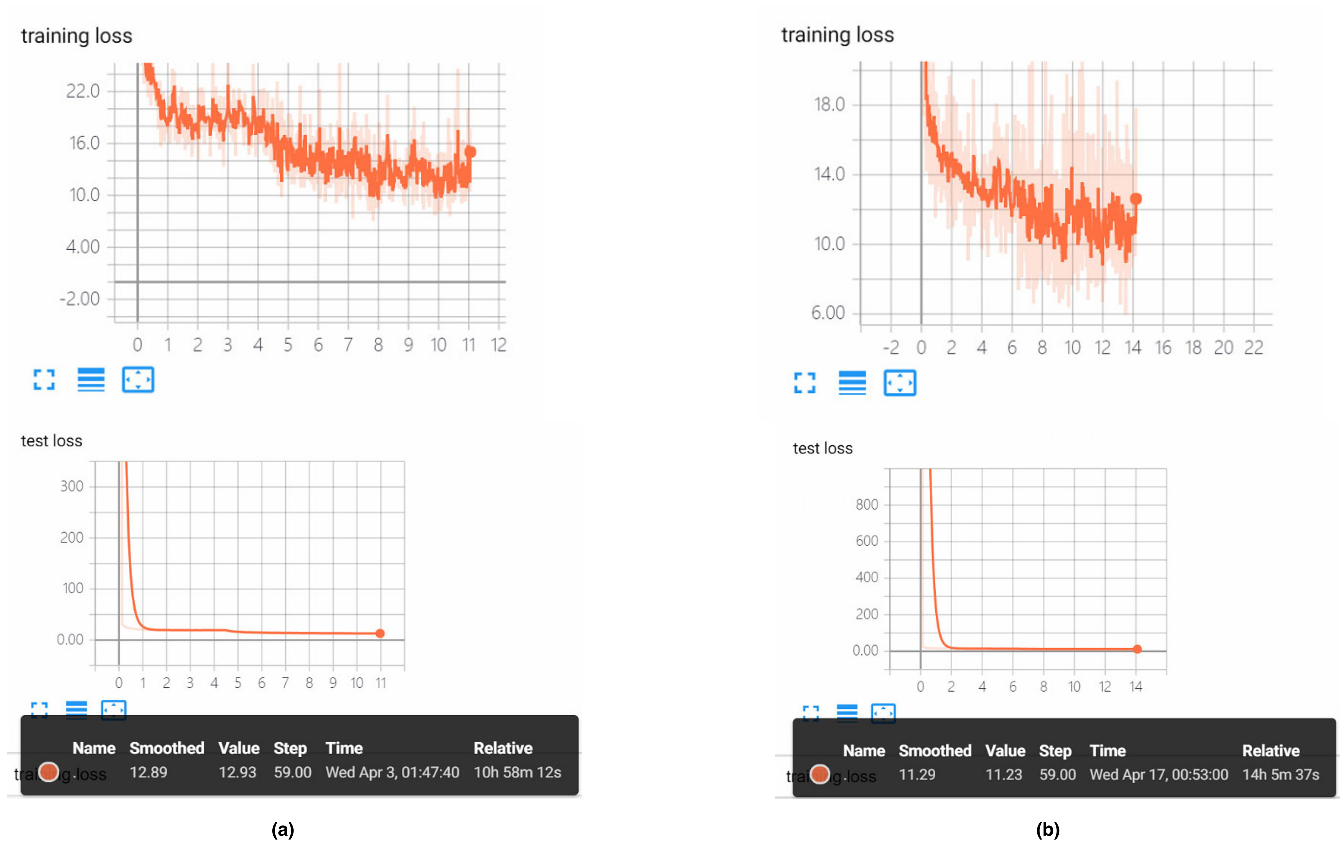


FIGURE 5. Training loss curves and testing loss curves. (a) Tiny-yolov3's training loss curve and testing loss curves. (b) Improved tiny-yolov3's training loss curve and testing loss curve.

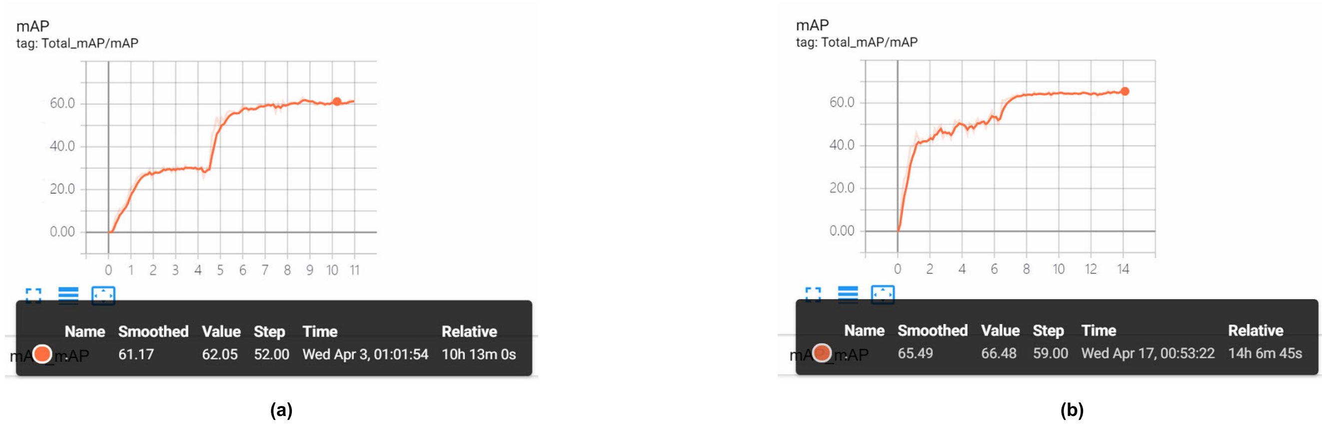
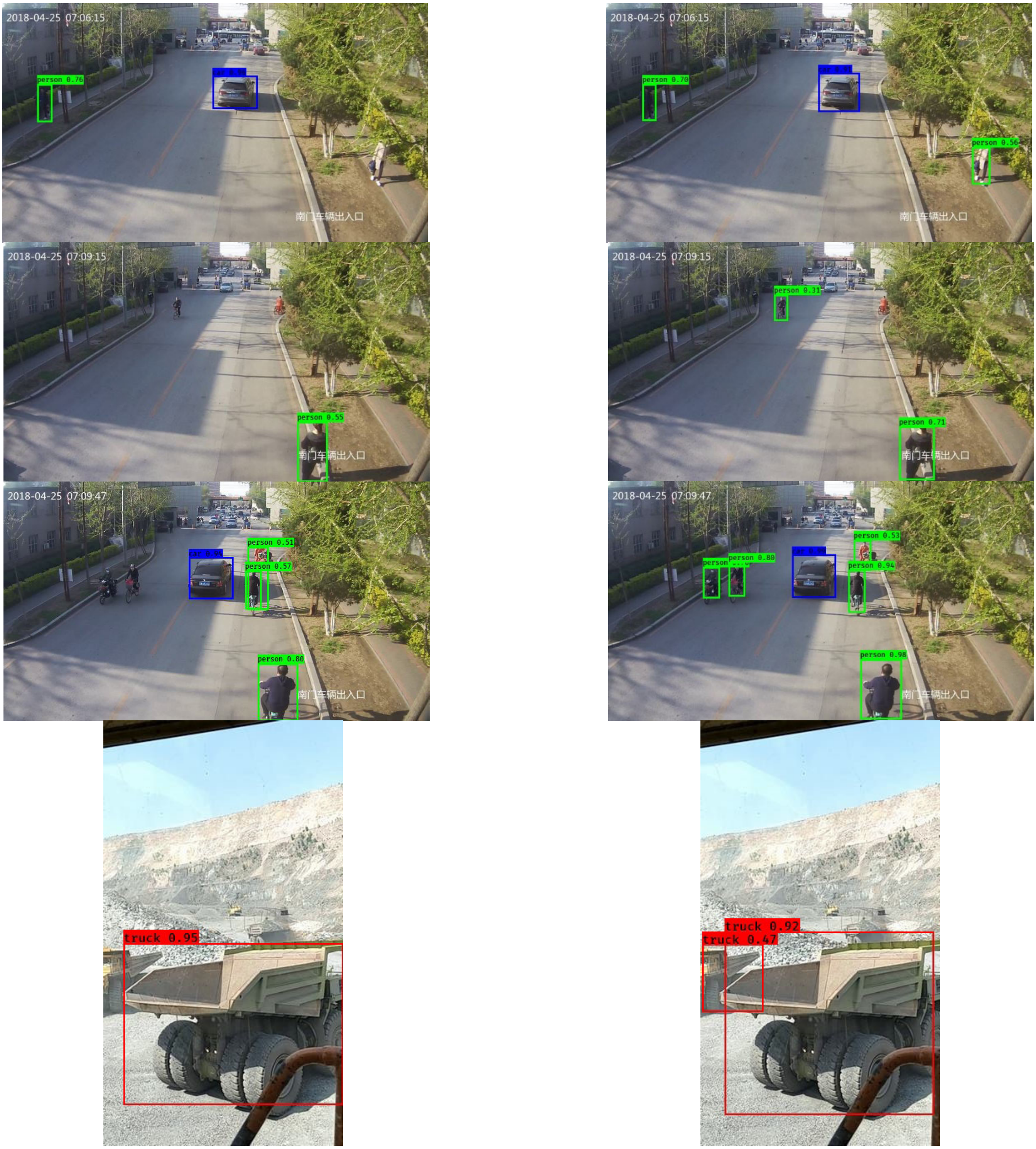


FIGURE 6. The mAP curves. (a)Tiny-yolov3's mAP curve. (b) Improved tiny-yolov3'smAP curve.

It can be known from Table 3 that only adding convolution kernel compared with adding residual structure has a greatly reduce in detection accuracy. The mAP reduces from 66.5 to 65.1. It also shows that the add function transmit this layer's feature information and information extracted by the two layers of the convolution layer to the next layer, which can reduce the information loss that occurs when the feature information is transmitted. Then the detection accuracy is improved and the FPS remain same because the amount of calculation during detection has hardly increased. Therefore, this paper adds residual structure from the fourth to seventh convolution layer to the tiny-yolov3 model as the improved model. The training loss curves and the testing loss curves are shown in Figure 5. The horizontal axis is the training time and the vertical axis is the loss value. Figure 5 shows that in the end, improved tiny-yolov3 model's loss value is lower than that of tiny-yolov3. The mAP curve of the model is shown in Figure 6. The horizontal axis is the training time and the vertical axis is the mAP value. Figure 6 shows that the highest mAP value of improved tiny-yolov3 model is 66.5, while the



(a)

(b)

FIGURE 7. Tiny-yolov3 test results and improved tiny-yolov3 test results. (a) Using tiny-yolov3 to detect object in the image; (b) Using improved tiny-yolov3 to detect object in the image.

highest mAP value of tiny-yolov3 model is 62.0. It means that improved tiny-yolov3 detect better than tiny-yolov3. The improvement effect of the model and the effect of original model is shown in Figure 7. Figure 7 shows that more targets can be detected by using improved tiny-yolov3 model.

V. CONCLUSION

This paper proposes a target detection model based on the improved tiny-yolov3 under mining truck environment. This model uses data enhancement on mining trucks and uses K-means clustering algorithm to calculate the most suitable anchors for the dataset used in this paper. It also proposes an improved model based on the tiny-yolov3 model which adds residual network structure to the original model to improve the accuracy of the detection. The experimental results show that the new model proposed in this paper is superior to the tiny-yolov3 model in terms of target detection, and the detection speed is not much decreased. However, the detection accuracy of this model is much different from that of the large model such as YOLOV3, and it is not as good as YOLOV3 for the detection of particularly small targets. The next step will be focus on the accuracy of the detection and the detection of small targets.

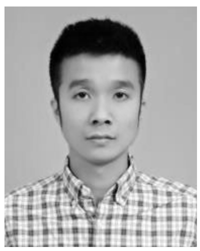
REFERENCES

[1] Y. Yuan, "Unmanned technology for mining trucks," *Mining Equip.*, vol. 10, pp. 72–74, Oct. 2013.
[2] H. Y. Su, "Research on security assistance system of mining heavy truck," *Opencast Mining Technol.*, vol. 31, no. 12, pp. 72–76, 2016.
[3] Q. Hu, Y. Zhang, and D. Y. Shen, "Emergency alarming equipment used in truck transportation in open-pit coal mine," *Opencast Mining Technol.*, vol. 2, no. 1, p. 34, 2002.
[4] H. L. Wei, A. B. Guo, W. B. Dong, X. J. Wang, and G. K. Huang, "Research and application of millimeter wave radar collision avoidance system for mining truck," *Saf. Coal Mines*, vol. 46, no. S1, pp. 43–47, 2015.
[5] H. P. Shi, "Application of truck collision warning system based on GPS in open-pit mine," *Opencast Mining Technol.*, vol. 31, no. 7, pp. 75–78, 2016.
[6] H. P. Yin, B. Chen, Y. Chai, and Z.-D. Liu, "Vision-based object detection and tracking: A review," *ACTA Automatica Sinica*, vol. 42, no. 10, pp. 1466–1489, 2016.
[7] Y. Yu and Y. Mo, "Deep learning principles and tensorflow practice," Publishing House Electron. Ind., Beijing, China, 2017, pp. 88–98.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.
[11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.
[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: https://arxiv.org/abs/1804.02767
[13] M. Everingham and J. Winn, "The PASCAL visual object classes challenge 2007 (VOC2007) development kit," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, 2006.
[14] S. Shetty, "Application of convolutional neural network for image classification on Pascal VOC challenge 2012 dataset," 2016, *arXiv:1607.03785*. [Online]. Available: https://arxiv.org/abs/1607.03785
[15] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*. [Online]. Available: https://arxiv.org/abs/1712.04621
[16] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *J. Roy. Stat. Soc. C, Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.

DONG XIAO received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2009, where he has been a Professor with the College of Information Science and Engineering, since 2006. His research interests include neural networks, extreme learning machine algorithms, partial least-squares algorithms, and modified median minimum distance algorithms.

FENG SHAN is currently pursuing the degree with the School of Information Science and Engineering, Northeastern University, Shenyang, China. His research interests include computer vision, neural networks, and pattern recognition.

ZE LI is currently pursuing the bachelor's degree in automation with Northeastern University, Shenyang, China. His research interests include deep learning and pattern recognition.



BA TUAN LE was born in Viet Tri, Vietnam. His research interests include intelligence information process, artificial neural networks, extreme learning machine, remote sensing exploration, and spectral analysis techniques.

XIWEN LIU is currently pursuing the degree with the College of Information Science and Engineering, Northeastern University, Shenyang, China. Her research interests include machine learning, deep learning, and image processing

XUERAO LI is currently pursuing the degree with the College of Resources and Civil Engineering, Northeastern University, Shenyang, China. Her research interests include deep learning and image processing.