Detection and Alarm of E-bike Intrusion in Elevator Scene

Hu Huang, Xiaodong Xie, and Luoyu Zhou

Abstract—**Object detection has always been a research topic in computer vision. With the deep learning, conventional object detection can obtain satisfactory results. However, some object detection problems with serious occlusions still need to be improved. This paper proposes a high accuracy detection and alarm method based on an improved YOLOv3 network and multi-frame direction field and solve the serious occlusions problem. Meanwhile, we apply the method into the elevator scene to solve the elevator security. Firstly, an improved YOLOv3 network is proposed for enhancing the feature extraction ability of the network and increasing the recognition accuracy. Then, multi-frame direction field is proposed to decrease false alarm. The experimental results have demonstrated that our proposed method can effectively solve object detection problem with serious occlusions, and has much practicality value in elevator scenes and other scenes.**

Index Terms—**E-bike intrusion; YOLOv3; Multi-frame direction field; Detection and alarm**

I. Introduction

WITH the acceleration of urbanization development, more and more residents choose E-bikes to travel. However, in the process of using E-bikes, it is inevitable to encounter various safety problems, especially the residents living in high-rise buildings when charging accidents of E-bike are frequently reported. Therefore, it is necessary to install intelligent detection and alarm system in some places such as elevators. This can not only prevent the building fire caused by E-bike charging, but also accord with the current intelligent trend. In fact, with the development of computer vision, artificial intelligence, pattern recognition technology, the perimeter intrusion detection and alarming has become a popular topic. However, the detection and alarm of E-bike intrusion in elevator scene face with many challenges, such as variety of E-bikes, structural similarity between E-bike and bicycle and serious occlusion. Therefore, how to quickly and accurately detect and identify the E-bike inside the elevator has become a key research direction in the field of security.

Manuscript received February 01, 2021; revised April 27, 2021.

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61901059 and Hubei Provincial excellent young and middle-aged scientific and technological innovation team project in Colleges and Universities under Grant T2020007.

Hu Huang is a student of Electronic and Information School of Yangtze University, Jingzhou 434023, China, (email 2285279728@qq.com)

Xiaodong Xie is an Engineer of material supply Department of PetroChina Changqing Oilfield Company, Xi'an, China, (email: dong1_cq@petrochina.com.cn)

Luoyu Zhou is an Associate Professor of Electronic and Information School of Yangtze University, Jingzhou 434023, China (Corresponding author, email: luoyuzh@yangtzeu.edu.cn, ORCID: orcid.org/0000-0003-4417-1250)

In the traditional target detection method, its main thought is to extract image texture features through sliding windows. These features include color, shape, statistics etc [1-3]. SIFT (Scale-invariant feature transform) was proposed by David Lowe [4] and applied to target recognition, image matching and other fields [5-7]. This method is to find the key points in different scale space and calculate the direction of the key points, which can easily combine with other feature vectors. However, this method lacks robustness and its real-time performance is not satisfactory because of a vast amount of sub-sampling and interpolation operations in SIFT. In recent years, many researchers have proposed various improved features based on SIFT [8-10], including dense SIFT, PCA-SIFT, SURF etc.

Navneet Dalal proposed a HOG feature in 2005 [11]. It constructs features by calculating the histogram of the gradient direction in the local region of the image. Because of its block and unit processing method, the relationship between local pixel points can be well-characterized. Nevertheless, it will cost too much time to generate the descriptor, which leads to poor real-time performance. Kedong Wang proposed a matching method by fusing HOG features and Zernike moments descriptors and their method was applied to SAR images [12]. Lakshmi D proposed to use modified HOG and LBP features for facial emotion recognition and achieved detection results [13]. In a word, the traditional target detection method is affected by the size and the step of the sliding window. Moreover, the features used in different scenes will be different and it is difficult to obtain the universal features by hand. Therefore, there are many inevitable problems such as window redundancy, poor generalization ability and high time complexity.

With the explosive growth of computational power, deep learning has been applied into many fields, such as image retrieval, image segmentation and image classification and so on [14-16]. Meanwhile, recent advances in target detection based on deep learning have also yielded transformative results.

In 2014, Girshick proposed the R-CNN [17] algorithm based on convolutional neural network for the first time. Thereafter, target detection methods based on deep learning are mainly divided into two categories. The first category algorithm, such as R-CNN [17], SPP-NET [18], FAST R-CNN [19] and Faster R-CNN [20], are all two-stage target detection algorithms based on candidate regions. Firstly, a series of candidate regions are generated by the boundary box search algorithm or selective search algorithm, and then the convolutional neural network is used to extract features from the original image for classification and positioning. The two-stage algorithm needs to detect every possible candidate region, but it has the problem of high time complexity. The second category algorithm, such as YOLO [21], and SSD [22], belongs to the one-stage target detection algorithms based on regression. This kind of algorithm takes target detection as a regression problem and does not need to generate the candidate box in the image. Instead, it directly gets the category probability and position coordinate value of the target through the regression model. Compared with the two-stage algorithm, the one-stage algorithm has the disadvantages in terms of detection accuracy. However, it has the advantages in terms of speed and time efficiency. Because of the superiority of speed and time efficiency, YOLO algorithm was optimized and improved later. YOLOv2 [23] and YOLOv3 [24] were proposed for further improving the performance by anchor frame assisted positioning, multi-scale training, residual structure optimization network and feature pyramid. However, existing object detection algorithms can not obtain the high accurate detection results in several special scenes, such as serious occlusion, too small targets and so on.

Inspired the above discussion, this paper proposed a detection and alarm method for the objects with serious occlusion in the elevator scene. The method is proposed based on an improved YOLOv3 network and multi-frame direction field. The experimental results have demonstrated that our proposed method can achieve satisfactory alarm results. Meanwhile, the method can effectively detect object with serious occlusion and can be extended into other complex scenes.

The structure of the paper is arranged as follows. Section II describes the experimental datasets and explains the whole training-test framework. Section III details the improved YOLOv3 algorithm and the alarm method based on multi-frame direction field. The experimental results are shown in Section IV and the final conclusion is displayed in Section V.

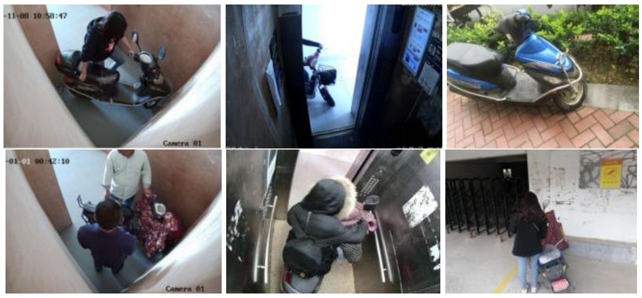
II. Dataset and Training-test Framework

A. Dataset

The E-bikes in the experimental dataset is mainly taken from three scenes, including the simulated elevator scene, the actual elevator scene and the natural scenes. Fig.1 shows some samples of E-bike.

During the testing, it is found that a bicycle is easy to be identified as an E-bike because of the structural similarity between them. On the other hand, in the narrow environment of the elevator, the E-bike may be obscured by pedestrian. These factors will further increase the difficulty of distinguishing between E-bikes and bicycles. Therefore, it is necessary to add bicycles into this dataset. As shown in Fig. 2, this additional bicycles can enrich the dataset and highlight the differences between E-bike and bicycle in subsequent network training.

Through the above acquisition, there are 7000 images in the dataset, including 3500 E-bike images, 3300 bicycle images and 200 mixed images (mixed image: contains both E-bike and bicycle, for example Fig. 2(e)) in the dataset. All images were labeled by LabelImg, which was an annotation tool for deep learning dataset. We used LabelImg to record the category name and location information, and store this information in XML format file. In the following training process, the training set and the testing set were randomly divided according to the ratio of 6:1.



(a) Simulated elevator (b) Actual elevator (c) Natural scene

Fig. 1. Some samples of E-bike in the dataset.

(a) (b) (c)

(d) (e) (f)

Fig. 2. Some samples of E-bike in the dataset.

B. Training-Testing Framework

As shown in Fig. 3, the whole network model framework includes training part and testing part. In the training part, the YOLOv3 network is improved by the modified loss function and activation function, and then trained by using the training set generated in Section II.

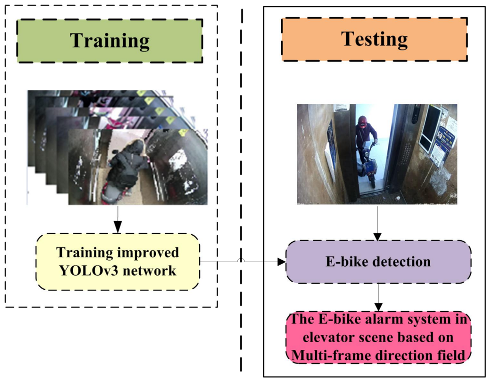


Fig. 3. The training-testing framework based on the improved YOLO v3 and multi-frame direction field.

In the test part, the trained network is firstly used to detect the E-bike in the image frame, and then the alarm method based on multi-frame direction field is used to further improve the accuracy of E-bike intrusion in the elevator scene. In the following section, we will describe the detailed algorithm.

III. The Detailed Algorithm

A. The principle of YOLOv3

YOLOv3 is improved based on YOLOv1 and YOLOv2 and simultaneously integrated with the advantages of other networks. Generally speaking, it is a single-stage detection network with excellent performance in real-time and accuracy. YOLOv3 is a full convolutional network [24], which adopts the first 52 layers of Darknet-53 to reduce the

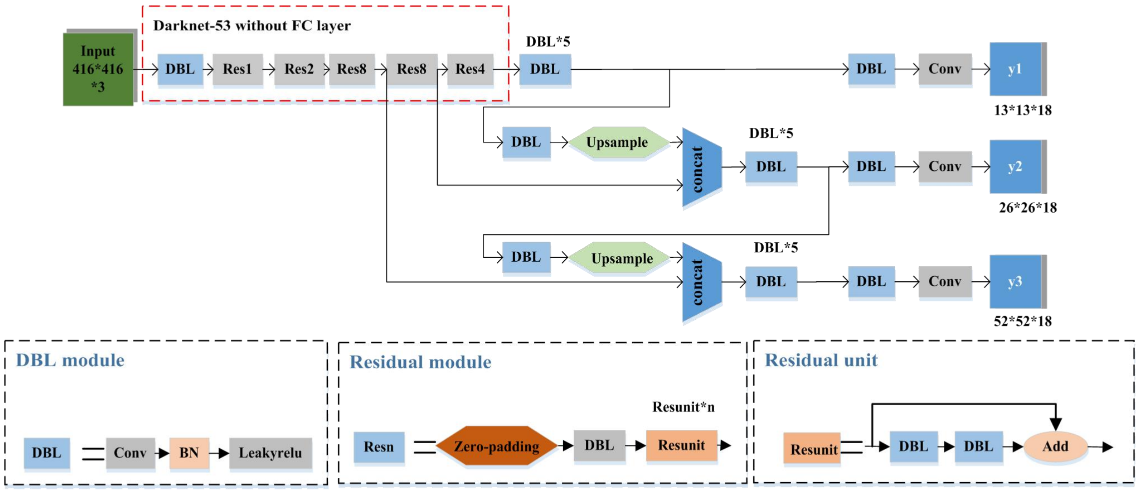


Fig. 4. YOLOv3 network structure in [24].

number of network layers. It greatly improves the computing speed as well as ensures classification accuracy. Moreover, YOLOv3 integrates the output feature map of Darknet-53 by twice up-sampling with shallow output features according to feature pyramid network structure [25]. The whole network structure is shown in Fig. 4. The basic modules of YOLOv3 includes DBL module, residual module, and residual unit.

The DBL module is the basic component of YOLOv3 and includes convolution, batch normalization, and leaky relu. The *n* in the Resn module stands for the number, for example res1, res2 ..., res8, and so on, showing how many res units in this res_block. The residual unit consists of DBL. Because YOLOv3 uses residual structure for reference, it can make the network structure deeper.

As for the evaluation indicators, YOLOv3 uses the commonly used Precision rate and Recall rate to evaluate the performance of network model.

$$P = \frac{T_p}{T_p + F_p} \tag{1}$$

$$R = \frac{T_p}{T_p + F_N} \tag{2}$$

where T and F are true and false respectively, which means that the result is correct or not. P and N are positive and negative respectively, which means that the result is considered as "positive class" or "negative class". T_P is the abbreviation of "True Positive" and denotes the number of instances that actually positive class is categorized into positive class. F_P is the abbreviation of "False Positive" and denotes the number of instances that actually negative class is categorized into positive class. F_N is the abbreviation of "False Negatives" and denotes the number of instances that actually positive class is categorized into negative class. In general, it is difficult to optimize both of them simultaneously, so the two parameters need to be considered comprehensively to obtain the third evaluation value F.

$$F = \frac{2*P*R}{P+R} \tag{3}$$

In addition, Accuracy refers to how closely a measurement or observation comes to " true value ", Therefore, we use Accuracy to visually observe the correct ratio of the proposed network and compare with other approaches. It is defined as follows.

$$Accuracy = \frac{N_{coorect}}{N_{total}} \tag{4}$$

where $N_{correct}$ means the correct results and N_{total} denotes the total samples.

B. E-bike Detection Method Based on Improved YOLO v3

1) Loss function improvement

The purpose of target recognition and positioning is to detect the category and position of the target. In order to effectively measure the two most important factors, The loss function used in YOLOv3, includes the loss function of the width and height of the prediction box, the loss function of the central coordinates of the prediction box, the loss function of confidence and the loss function of category.

$$\begin{aligned} Loss\_yolov3 = {} & wh\_loss + xy\_loss \\ & + confidence\_loss + class\_loss \end{aligned} \tag{5}$$

where wh_loss adopts the total mean square error, and the other three loss functions adopt the binary cross entropy. The mean square error is very sensitive to the scale of the target, especially for the small targets. By taking the square root of width and height, the influence of the scale can be reduced to some extent. However, the influence cannot be completely eliminated. Considering that the offset effects of different target prediction boxes are not consistent, the loss function in this paper draws on the idea of G_IOU loss function and introduces overlapping area. However, this paper does not directly use the overlapping area as the loss function, but uses the normalized overlapping area (actual overlapping area/actual frame area) as the loss function. Therefore, the loss function in this paper is shown as

$$\begin{aligned} Loss\_ebike = {} & wh\_loss + xy\_loss + confidence\_loss \\ & + class\_loss + area\_Loss \end{aligned} \tag{6}$$

By adding this loss function $area\_Loss$, the multi-scale property of the network can be further improved to avoid the missed detection of small targets and improve the prediction accuracy.

2) Activation function improvement

ReLU activation function is one of the most commonly used activation functions at present. As shown in Equation (7), it is not globally differentiable, but we can use sub-gradient. In addition, ReLU also has the following advantages:

- The exploding gradient can be effectively solved;
- The calculation speed is very fast;
- The convergence rate is very fast.

$$ReLU(x) = \begin{cases} x & if\ x > 0 \\ 0 & if\ x \le 0 \end{cases} \tag{7}$$

However, it is easy to stop when ReLU is used for training. For a value less than 0, the gradient of this neuron will always be 0. If the learning rate is very high, many neurons in the network may be useless. Even if the learning rate is small, this situation is also likely to happen. YOLOv3 adopts the Leaky Relu activation function, as shown in the following equation. It assigns a non-zero slope to all negative values, which overcomes the shortcomings of ReLU.

$$LReLU(x) = \begin{cases} x & if\ x_i \ge 0 \\ a_i x & if\ x_i < 0 \end{cases} \tag{8}$$

Nevertheless, the a_i in Leaky ReLU is constant. Then PReLU was proposed for the ImageNet classification [26]. Its slope of the negative part a_i is determined by the data. PReLU only adds a tiny number of parameters, which means that the computational load of the network and the risk of overfitting are only slightly increased. However, it can enhance the feature extraction ability of the network and facilitate the implementation of the sub-pixel convolutional network. Moreover, the introduction of PReLU function can increase the learning ability of the network to some extent, and thus improve the recognition rate.

3) The alarm of E-bike intrusion in elevator scene based on multi-frame direction field

Although the recognition accuracy of E-bike has been improved to a certain extent (Section IV), there are still some image frames that cannot be recognized or even misrecognized due to serious occlusion. As shown in Fig. 5, serious occlusion leads to missed recognition in Fig. 5(a). Moreover, owing to the structural similarity, the E-bike in Fig. 5(b) are recognized as a bicycle. Therefore, if the alarm is determined by the recognition result of a single image frame, a large number of false alarms will inevitably be generated. This paper analyzes numerous original videos in the elevator environment, and finds that the recognition effect is the best when an E-bike enters or leaves the elevator. Therefore, the false alarm of E-bike intrusion can be decreased by the direction field of multi-frame image. The flow chart of the alarm algorithm based on the multi-frame direction field is shown in Fig. 7 and the sets of this algorithm are shown as follows

- The elevator scene is divided into two areas: region A is inside of the elevator and region B is outside of the elevator, as shown in Fig. 6;
- If the E-bike moves from region A to region B, the direction field is outward. In return, the direction field is inward;
- Set a switching factor a to judge the elevator state. If the switching factor a is the closed state (set as 0), the elevator state remains unchanged; If a is the open state (set as 1), the elevator state can be changed.

Fig. 5. Missed recognition and false recognition.

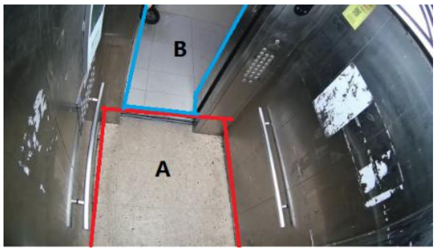


Fig. 6. Region partition

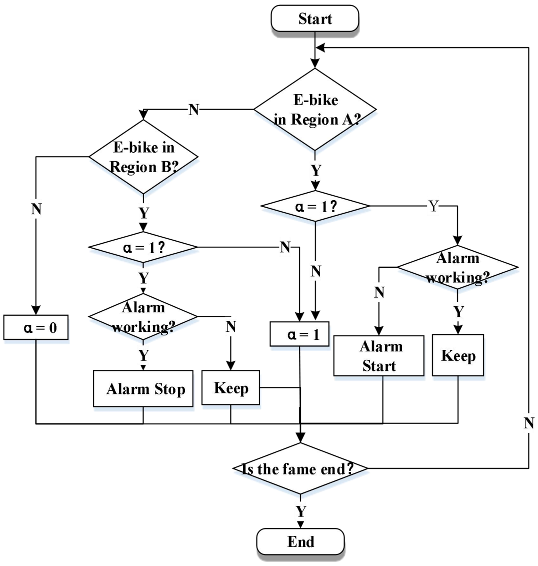


Fig. 7. The flow chart of the alarm algorithm based on the multi-frame direction field.

IV. Experiments and Analysis

The experimental environment of this paper is Ubuntu18.04 version 64-bit operating system, including Intel Core i9 3.6GHz, NVIDIA GTX1080TI, Memory 32GB, CUDA version 10.0 and CUDNN version 7.5. In order to avoid the gradient decreasing too fast and overfitting, the initial learning rate of the network is set as 0.001. Some experimental parameters are shown in Table 1.

A. Data augmentation

In the field of computer vision, to train the model with high precision and strong generalization ability, it is necessary to have a large number of training samples. However, the limited data can be obtained in most cases, so it is very necessary to use a data augmentation method to make up for the lack of data. Data augmentation is based on existing data by applying some small changes to the original sample. The data enhancement methods include zooming, clipping, saturation adjustment and brightness value adjustment, etc.

These data augmentation methods can reduce the sensitivity of network training to spatial position. Different data augmentation method can be superimposed to increase the complexity and diversity of the sample image. Data augmentation methods can be used to prevent over fitting in deep convolution neural network and improve the network performance. Moreover, it can effectively solve the problem of small number of samples and unbalanced samples.

TABLE I
THE EXPERIMENTAL PARAMETERS

Parameter	Value
Learning rate	0.001
Batch size	32
Weight_decay	0.0005
Score_threshold	0.4
Iou_threshold	0.6
Epoch	100

(a) original (b) cropped

(c) mirror (d) distorted

Fig. 8. Data augmentation in pre-processing.

Based on the above analysis, this paper uses data augmentation method to increase the sample size of the dataset and the complexity and diversity of the data. As shown in Fig. 8, it shows the effect of various data augmentation technologies.

B. E-bike Recognition

As described in Section II, we divide the dataset into a training set and a testing set according to the ratio of 6:1. The testing set contains 1000 images (500 E-bike images, 471 bicycle images, and 29 mixed images of both the E-bike and the bicycle). Some samples of test results are shown in Fig. 9.

In order to reflect the superiority of the improved YOLOv3 network, we also used the training set to train original YOLOv3 model. Then the comparison between the test results of original YOLOv3 model and our test results are shown in Table 2. Firstly, the Recall rate of both models is less than the accuracy rate, because the structures of some E-bikes are often very similar to those of bicycles. As a result, these E-bikes are easily recognized as bicycles. Secondly, the Precision rate, Recall rate and F values of the improved YOLOv3 model are about 3 percentage points higher than those of the original YOLOv3 model. The improvement of recognition accuracy will provide a good foundation for the subsequent alarm of E-bike intrusion.

Fig. 9. The test results based on improved network.

TABLE II
UNITS FOR MAGNETIC PROPERTIES

Indicators	Precision rate	Recall rate	F values	Accuracy
Original YOLOv3	0.958	0.946	0.952	0.950
Improved YOLOv3	0.985	0.974	0.979	0.978

C. The alarm of E-bike intrusion in elevator scene

In this section, the improved YOLOv3 model is firstly used to obtain recognition result of each frame of a given video. Then the alarm algorithm described in Part C of Section III is used to achieve alarm results. Fig. 10 shows the detection results of a video.

If the alarm of E-bike intrusion is only determined by the recognition result of a single frame, there will be a lot of incorrect alarm. As shown in Fig.11, the alarm will begin at the 3rd frame, and stop at the 9th frame. Then the alarm will begin at the 11th frame, and stop at the 15th frame again. Obviously, due to the serious occlusion in the 9th and the 10th frames of this video, the method can not detect the E-bike. Undoubtedly, this alarm result contraries to the actual state of the elevator.

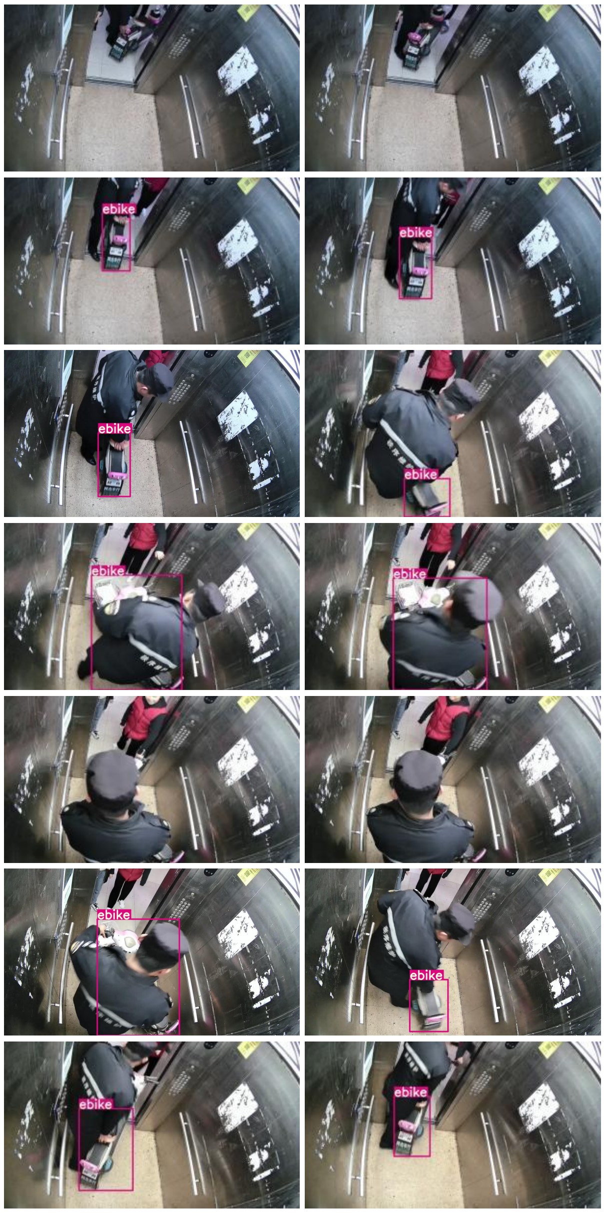


Fig. 10. The detection results of a video.

On the other hand, we use the alarm algorithm based on the multi-frame direction field to achieve the alarm result. There is no detected E-bike in Region B of the 9th frame and the 10th frame. However, there is also no E-bike in region A at present, that is to say, the direction field of the E-bike is not outward. Therefore, the alarm will not be stop at this time. When entering the 15th frame, there is a detected E-bike in region A, that is to say, the direction field is outward at this time, so the alarm will be stop. The alarm result by introducing the direction field is shown in Fig.12, which is consistent with the actual state of the elevator.

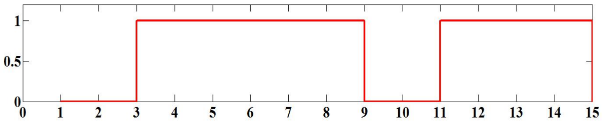


Fig. 11. The alarm result based on each single frame detection result (1: alarm, 0: normal).

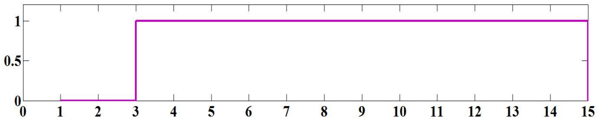


Fig. 12. The alarm result based on the multi-frame direction field (1: alarm, 0: normal).

Similarly, Fig. 13 shows the detection results of another video. The E-bikes in 3rd frame and 7th frame are not detected. If the alarm of E-bike intrusion is only determined by the recognition result of a single frame, there will be also a lot of incorrect alarm, as shown in Fig.14. On the other hand, the alarm result by introducing the direction field is shown in Fig.15, which is consistent with the actual state of the elevator. These instances have demonstrated that our method is effective for detection and alarm of E-bike intrusion in elevator scene.

Fig. 13. The detection results of another video.

V. Conclusion

In this paper, a detection and alarm method is proposed for E-bike intrusion in elevator scene. The obtained satisfactory detection and alarm results benefit from two aspects. On the one hand, the improved YOLOv3 is proposed for the E-bike detection. Then the detection accuracy can increase by three percentage points. On the other hand, in order to achieve the superior alarm results, an alarm algorithm based on the multi-frame direction field is proposed for E-bike intrusion in elevator scene. In a word, the proposed method has much practicality value in elevator security and other scenes.

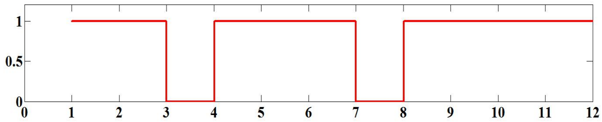


Fig. 14. The alarm result based on each single frame detection result (1: alarm, 0: normal).

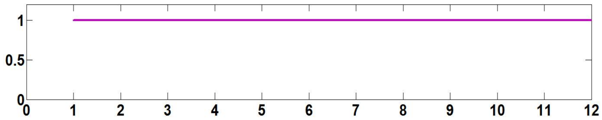


Fig. 15. The alarm result based on the multi-frame direction field (1: alarm, 0: normal).

References

[1] Chapelle O, Haffner P, Vapnik V. “Support Vector Machines for Histogram-Based Image Classification,” *IEEE Transactions on Neural Networks*, 10(5): 1055-1064, 1999.

[2] Maca R, Benes M, Tintera J, “Segmentation of MRI Images by Adaptive Degenerate Diffusion,” *IAENG International Journal of Applied Mathematics*, 45(3): 208-217, 2015.

[3] Salhi K, Jaara E, Alaoui M, et al. “Color-Texture Image Clustering Based on Neuro-morphological Approach,” *IAENG International Journal of Computer Science*, 46(1):134-140, 2019.

[4] Lowe D G. “Object recognition from local scale-invariant features,” *Proceedings of the IEEE International Conference on Computer Vision*, 1150-1157, 1999.

[5] Lenc L, Král P. “Automatic face recognition system based on the SIFT features,” *Computers & Electrical Engineering*, 46:256-272, 2015.

[6] Shi Y, Lv Z, Bi N, et al. “An improved SIFT algorithm for robust emotion recognition under various face poses and illuminations,” *Neural Computing and Applications*, 2020, 32(4):9267-9281.

[7] Li Y, Li Q, Liu Y, et al. “A spatial-spectral SIFT for hyperspectral image matching and classification,” *Pattern Recognition Letters*, 127:18-26, 2018.

[8] Juan L, Gwun O. “A Comparison of SIFT, PCA-SIFT and SURF,” *International Journal of Image Processing*, 3(4):143-152, 2009.

[9] Liu Y, Liu S, Wang Z. “Multi-focus image fusion with dense SIFT,” *Information Fusion*, 23:139-155, 2015.

[10] Holman R A, Brodie K L, Spore N J. “Surf Zone Characterization Using a Small Quadcopter: Technical Issues and Procedures,” *IEEE Transactions on Geoscience and Remote Sensing*, 55(4):2017-2027, 2017.

[11] Dalal N, Triggs B. “Histograms of oriented gradients for human detection,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 886-893, 2005.

[12] Wang K, Wang H, Wang J. “Terrain Matching by Fusing HOG With Zernike Moments,” *IEEE Transactions on Aerospace and Electronic Systems*, 56(2):1290-1300, 2020.

[13] Lakshmi D, Ponnusamy R. “Facial Emotion Recognition using modified HOG and LBP features with Deep Stacked AutoEncoders,” *Microprocessors and Microsystems*, 82, 103834, 2021.

[14] Huang Y, Chen J, Fan S, et al. “Applying Time-Frequency Image of Convolutional Neural Network to Extract Feature on Long-Term EEG Signals to Predict Depth of Anesthesia,” *International Conference of Signal and Image Engineering*, 2019.

[15] Liu G, Wang Y. “Oracle Character Image Retrieval by Combining Deep Neural Networks and Clustering Technology,” *IAENG International Journal of Computer Science*, 47(2):199-206, 2020.

[16] Sun Y, Chen Y, Pan Y, “Android Malware Family Classification Based on Deep Learning of Code Images,” *IAENG International Journal of Computer Science*, 46(4):524-533, 2019.

[17] Girshick R, Donahue J, Darrell T, et al. “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580-587, 2014.

[18] He K, Zhang X, Ren S, et al. "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 37(9): 1904-1916, 2014.

[19] Girshick R. "Fast R-CNN," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1440-1448, 2015.

[20] Ren S, He K, Girshick R, et al. "Faster R-CNN: towards real-time object detection with region proposal networks," *International Conference on Neural Information Processing Systems*, 37(9): 1137-1139, 2017.

[21] Redmon J, Divvala S, Girshick R, et al. "You only look once: unified, real-time object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779-788, 2016.

[22] Liu W, Anguelov D, Erhan D, et al. "SSD: single shot multibox detector," *Proceedings of the European Conference on Computer Vision*, 21-37, 2016.

[23] Redmon J, Farhadi A. "YOLO9000: Better, Faster, Stronger," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recongnition*, 6517-6525, 2017.

[24] Redmon J, Farhadi A. "YOLOv3: an incremental improvement," *arXiv:* 1608.06019, 2018.

[25] Lin T, Dollár P, Girshick R, et al. "Feature Pyramid Networks for Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recongnition*, 936-944, 2017.

[26] Shelhamer E, Long J, Darrell T. "Fully Convolutional Networks for Semantic Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4): 640-651, 2017.

Hu Huang received the B.S. degree in Electronic Information Science and Technology from Wuhan Polytechnic University, Wuhan, China, in 2019. He is currently pursuing the M.E. degree in Yangtze University, Jingzhou, China. His current research interests include image/video processing, high performance computing and parallel computing.

Xiaodong Xie received the B.S. degree in Automation from Yangtze University, Jingzhou, China, and Master degree in oil and gas engineering from Yangtze University, Jingzhou, China, in 2017. He is currently an Engineer with material supply Department of PetroChina Changqing Oilfield Company, Xi'an, China. His current research interests include Internet of things, artificial intelligence

Luoyu Zhou received the B.S. degree in optical information science and technology from University of Science and Technology of China (USTC), Hefei, China, in 2008, and the Ph.D. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, (CIOMP, CAS), Changchun, China, in 2013. He is currently an associate professor with the School of Electronics and Information, Yangtze University, China. His current research interests include image processing, computer vision and artificial intelligence.