A Railway Similarity Multiple Object Tracking Framework Based on Vehicle Front Video

Lirong Lian[1,2,3], Yong Qin[1,3](✉), Zhiwei Cao[1,2,3], Yang Gao[1,2,3], Jie Bai[1,2,3], Xuanyu Ge[1,2,3], and Hang Yu[1,2,3]

[1] State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China
{lianlirong,yqin,zhiwei,gaoyang_,baijie,gexuanyu, yuhang_}@bjtu.edu.cn

[2] School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

[3] Key Laboratory of Railway Industry of Proactive Safety and Risk Control, Beijing 100044, China

Abstract. Real-time awareness of the service status of facilities and equipment in metro tunnels is essential for the safe movement of trains. For low illumination underground tunnel environments where objects have high similarity characteristics and are difficult to distinguish, this paper proposes a framework for tracking high similarity objects along the rail line based on on-board visual perception devices. The framework mainly consists of two stages: object detection and object association. In the first stage, an object detection algorithm is used to detect potentially faulty objects in the video, and this paper takes the continuous occurrence of insulated steel supports on the line as the high similarity research object. In the second stage a matching strategy is used to associate the same object in the video sequence and assign the same ID (identity document). Finally, the proposed framework was tested on metro tunnel video and achieved an accuracy of 95.64 and a detection performance of 25 FPS (frames per second), 82.17 MOTA (multiple object tracking accuracy) and 88.46 MOTP (multiple object tracking precision) tracking performance.

Keywords: On-board vision sensors · Similarity objectives · Object detection · Multiple object tracking

1 Introduction

By the end of 2022, 53 cities in China had opened a total of 290 operational urban rail lines, with an operational mileage of 9,584 km. At present, the monitoring of the safety of the metro operating environment is still mainly based on manual inspection. The long distances, high frequency of inspection requirements and the low illumination level of the tunnel environment lead to frequent cases of missed inspections. Therefore, it is necessary to apply vision sensors to monitor the train operating environment around the clock in order to reduce the pressure on inspectors and improve inspection efficiency.

M. Gong et al. (Eds.): EITRT 2023, LNEE 1138, pp. 73–81, 2024.
https://doi.org/10.1007/978-981-99-9319-2_9

Monitoring the service status of equipment and facilities is a key task in the inspection of metro tunnels. With the large-scale opening of metro lines in cities, the cost of operation and maintenance and the safety of the environment pose higher requirements and greater challenges. The sensors used for inspection tasks need to sense the service status of the equipment and facilities in the tunnel in real time, so intelligent analysis algorithms based on on-board video data are developing rapidly. Cameras can capture large amounts of video stream data, but intelligent analysis algorithms that fail to correlate the same object in time series can lead to repeated alarms for the same faulty object. To solve this problem, we had to introduce object association algorithms and optimize the algorithms based on scene adaptation to achieve fast and accurate multi-object tracking. Figure 1 shows a common railway tunnel scene. As shown in the figure, the tunnel has the scene characteristics of a single background, low illumination, winding lines and a closed field of view, which increases the difficulty of object feature extraction. At the same time the high similarity of objects in the line is difficult to distinguish, which is a key problem for the object tracking algorithm to be solved.

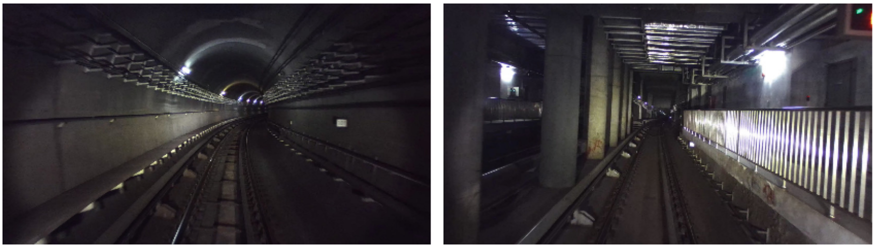

Fig. 1. Underground tunnel scene

This paper proposes a framework for tracking similarity objects along the track based on on-board vision sensors, which can accurately and quickly track facility objects along the track. The framework is shown in Fig. 2. The contributions of this paper are as follows.

A framework for track line similarity object tracking based on on-board vision sensors is proposed. The framework consists of an object detection algorithm and an association algorithm.

A dataset is constructed for the study of similarity object tracking along the line in an underground tunnel scenario.

The algorithmic framework in this paper achieves 95% accuracy and 25FPS detection performance, 97.18MOTA and 94.85MOTP tracking performance when tested on in-vehicle video data captured in an underground tunnel scenario.

2 Related Work

2.1 Intelligent Video Analytics for Rail Transit

Great progress has been made in recent years in object recognition and object tracking based on deep learning. In the field of rail transportation, iterative learning based on monitoring data to construct detection models of research objects is widely used in inspection business scenarios.

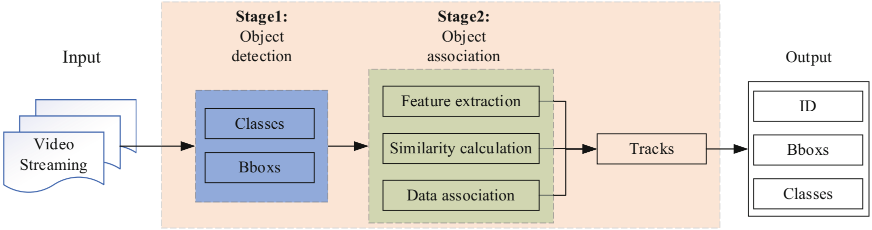

Fig. 2. Subway similarity object tracking framework based on on-board vision sensors

References [1–5] implement foreign object intrusion detection methods in complex railway environments. Reference [6] propose a real-time detection method based on improved YOLOv4, which is able to recognize locomotive signal lights of different colors and railway pedestrians under bad illumination conditions. References [7–9] propose a series of deep learning approaches, which are fast and accurate at the same time, are proposed to inspect the key components of railway track including rail, bolt, track fastener and clip. References [10, 11] research on deep learning method for rail surface defect detection. Reference [12] proposes a dehazing model based on an end-to-end multi-scale residual network, which is effective for dehazing on railway monitoring images.

Intelligent video analysis technology for rail transportation based on deep learning sends timely alert messages in abnormal situations to improve the efficiency of drivers in handling emergencies. The technology effectively overcomes the limitations of manual techniques in terms of scale, accuracy and scene complexity.

2.2 Multi-object Tracking at Railway

Multi-Object Tracking (MOT) uses image processing techniques to steadily track multiple objects of interest in a video sequence. The technique obtains the correct position of the same object in successive frames and achieves its trajectory tracking by giving it a unique identity document (ID). With the improvement of object detection performance in recent years, the detection-based multi-object tracking paradigm has made a major breakthrough and is rapidly becoming the mainstream framework for MOT tasks. In areas such as intelligent surveillance systems, autonomous driving and behavior recognition, tracking objects in surveillance video is beneficial for improving safety in security and transportation.

Reference [13] proposed a railway crossing intrusion detection algorithm, designed three non-parametric background models with different learning rates to detect moving and stationary objects. Which incorporated an object tracking strategy to reduce false alarms in detection. Reference [14] proposed a kernel correlation filtered object tracking dimensionality reduction algorithm combining dark channel prior and scale estimation, which effectively improves the accuracy of object tracking under foggy sky. In the underground tunnel scenario, the object detector loses the object resulting in an ID Switch when the object reappears. The object is extremely similar and indistinguishable, leading to false matches. In summary, how to achieve accurate data association is a key problem to be solved for in-vehicle video-based multi-object tracking.

Smart railway is an important development in the world of railway systems, where active safety technology is the most effective way to address the security challenges of railway systems [15].

3 Methodology

In this section, YOLOv5(You Only Look Once) [16] is used to detect object in vehicle video streams. Then the objects are associated based on the proposed matching strategy to finally achieve real-time and accurate object tracking.

3.1 Object Detection

The network structure of the YOLOv5 Contains three parts: Backbone, Neck and Prediction Head, the network structure is shown in Fig. 3.

The combined Convolution, BN (batch normalization) and SiLU module is used for feature extraction and down sampling of the input image. The C3 module increases the number of network layers through residual units, which facilitates the extraction of high-level semantic information and enables the fusion of multi-level features without changing the size of the feature map. SPPF improves the spatial pyramid pooling structure by serially connecting the three maximum pooling layers to obtain a better performance improvement without compromising the pooling effect. The Neck network layer uses a twin-tower structure combining FPN (feature pyramid networks) and PAN (path aggregation network) for feature fusion; the FPN structure passes semantic information from top to bottom to enhance the features of large objects, and the PAN passes semantic information again from bottom to top to improve the extraction of features of small objects. Finally Head outputs three scales of feature images to predict large, medium and small objects.

In the first stage, the object detector extracts information on the position and dimensions of the rail insulation bracket, which is used as the observation input for the subsequent association algorithm.

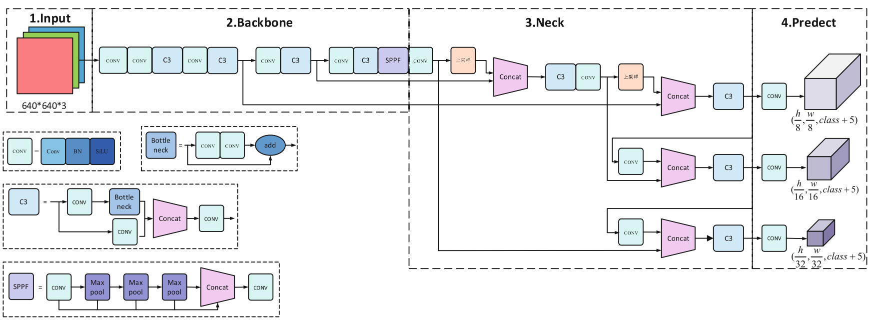

Fig. 3. The network structure of YOLOv5

3.2 Object Association

According to the train running direction, the insulated bracket approaches the bottom of the camera field of view from a distance, which is represented as the object position downward under the image coordinate system, as shown in Fig. 4. We propose a object tracking algorithm based on a priori information of the position of the insulation bracket in the next frame, and the algorithm framework is shown in Fig. 5.

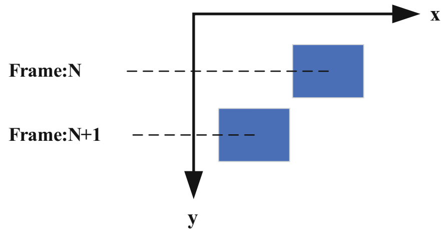

Fig. 4. A priori location information

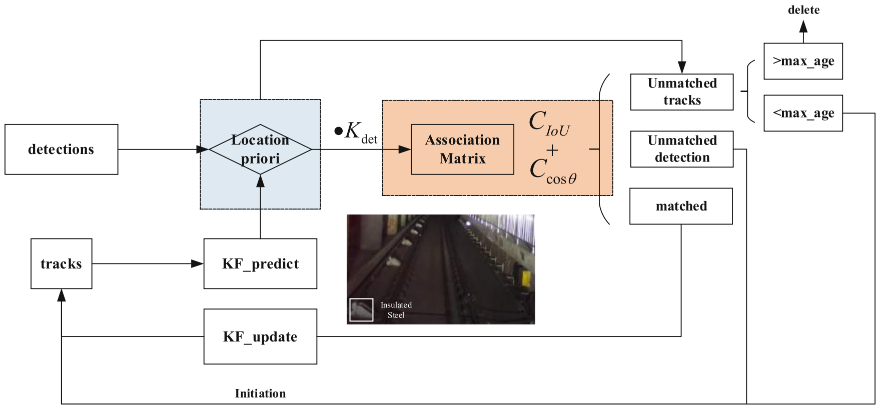

Fig. 5. A priori position-based post-detection tracking framework

In the second stage, given n existing tracks and m detections on the new-coming time step, the association cost matrix is formulated as:

$$C^{m \times n} = K_{\det}(C_{IoU} + C_{\cos\theta})^{m \times n} \quad (1)$$

$$K_{\det} = \begin{cases} k_{i,j} = 1, & y_{\det} \geq y_{track} \\ k_{i,j} = \infty, & y_{\det} < y_{track} \\ i = 1, 2, \ldots, m \; j = 1, 2, \ldots, n \end{cases} \quad (2)$$

where $K_{\det}$ is the position suppression matrix, and existing tracks in non-a priori positions are suppressed by an infinite value of ∞; n existing tracks and m detections compute Intersection over Union and cosine similarity, and the cost matrix is obtained from

$C_{IoU} = 1 - IoU$, $C_{\cos\theta} = 1 - \cos(\theta)$. The matching cost matrix is designed using a priori information from real application scenarios to improve the adaptability and robustness of the tracking algorithm in the underground tunnel scenario.

4 Experiment

4.1 Dataset

The on-board visual perception equipment collects video data from the front view on the metro line species to construct a sample database of metro tunnel scenes. This paper takes the rail insulation bracket as the research object, selects 1981 data images to build the sample database, and uses LabelImg to obtain the label information of the rail insulation bracket, as shown in Fig. 6.

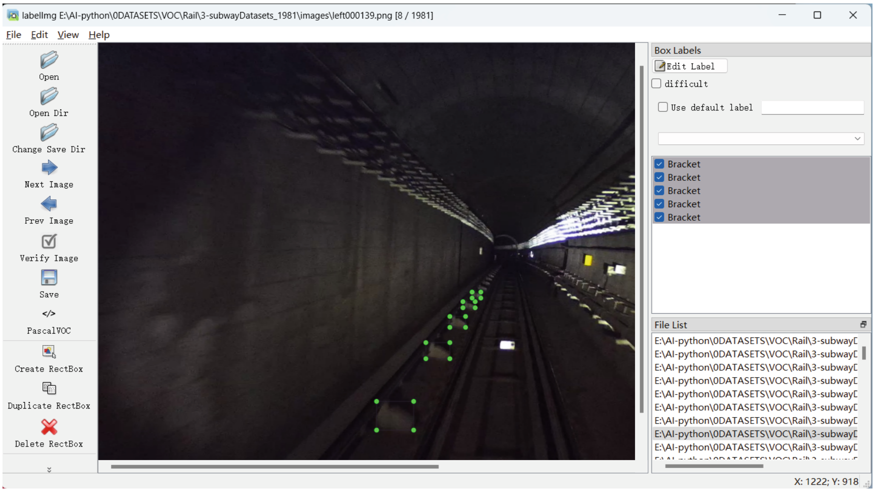

Fig. 6. Diagram of LabelImg labeling window

4.2 Experimental Parameters

YOLOv5 was trained and tested on CPU (Intel(R) Core(TM) i7-8700K) CPU@3.70 GHz × 12) and GPU NVIDIA GeForce RTX 2080. The training set is split 4:1 for training and validation. YOLOv5 is trained with a batch size of 16 for 300 epochs, the learning rate was set to 0.01, with a weight decay of 0.0005, a momentum from 0.937, and a maximum learning rate of 0.01. Unless otherwise specified, all experiments would follow the default training setting.

4.3 Experimental Results

YOLOv5 has a 95.64 accuracy rate and an operating speed of 25FPS, which can fully meet the needs of field applications.

CLEARMOT metrics [17] were used to verify the tracking performance of the proposed algorithm. We present the tracking performance of MOTA (multiple object tracking accuracy), MOTP (multiple object tracking precision) and IDS (identity document switch). Then compare the proposed approach to Deep SORT [18] on test set, as shown in Table 1.

Table 1. Evaluation results on test set.

Method	MOTA↑	MOTP↑	IDS↓
Proposed method	82.17	88.46	16
Deep SORT	72.84	84.17	29

Fig. 7. Tracking results on test set.

As shown in Fig. 7, the first column shows the Deep SORT tracking results and the second column of our method. One detection is matched to two trajectories, trajectories are tracked incorrectly and Significant ID switchover occurs in the processing result of Deep SORT. By contrast our method has more accurate and robust tracking results.

5 Conclusion

To explore trackers suitable for the underground tunnel scene, we propose an efficient but accurate similarity object tracking architecture by the paradigm of post-detection tracking with improved object association algorithms. Evaluation results demonstrate the superiority and competitiveness of our method: compared with Deep SORT, it can significantly improve detection accuracy by 9.33 MOTA and 4.29 MOTP, with down by almost half IDSWR. In future work, we would enhance the robustness and practicality of our algorithm in more data from real-world scene.

Acknowledgments. This work is supported by the National Key R&D Program of China (Contract No. 2022YFB4300601).

References

1. Zheng, J., Chen, Y., Zhang, H., Liu, D.: Foreign object detection algorithm based on multi-scale convolutional network. J. Phys.: Conf. Ser. **1952**(2), 022017 (2021)
2. Chen, W., Meng, S., Jiang, Y.: Foreign object detection in railway images based on an efficient two-stage convolutional neural network. Comput. Intell. Neurosci. (2022)
3. Cao, Z., et al.: An effective railway intrusion detection method using dynamic intrusion region and lightweight neural network. Measurement **191**, 110564 (2022)
4. Meng, C., Wang, Z., Shi, L., Gao, Y., Tao, Y., Wei, L.: SDRC-YOLO: a novel foreign object intrusion detection algorithm in railway scenarios. Electronics **12**(5), 1256 (2023)
5. Li, Y., Qin, Y., Xie, Z., Cao, Z., Jia, L., Yu, Z., Zheng, J., Zhang, E.: Efficient SSD: a real-time intrusion object detection algorithm for railway surveillance. In: International conference on sensing, diagnostics, prognostics, and control, pp. 391–395. IEEE Beijing (2020)
6. Wang, H., Pei, H., Zhang, J.: Detection of locomotive signal lights and pedestrians on railway tracks using improved YOLOv4. IEEE Access **10**, 15495–15505 (2022)
7. Wei, X., Yang, Z., Liu, Y., Wei, D., Jia, L., Li, Y.: Railway track fastener defect detection based on image processing and deep learning techniques: a comparative study. Eng. Appl. Artif. Intell. **80**, 66–81 (2019)
8. Chandran, P., Asber, J., Thiery, F., Odelius, J., Rantatalo, M.: An investigation of railway fastener detection using image processing and augmented deep learning. Sustainability **13**(21), 12051 (2021)
9. Wang, T., Yang, F., Tsui, K.: Real-time detection of railway track component via one-stage deep learning networks. Sensors 20(15), 4325(2021)
10. Feng, J., Yuan, H., Hu, Y., Lin, J., Liu, S., Luo, X.: Research on deep learning method for rail surface defect detection. IET Electr. Syst. Transp. **10**(4), 436–442 (2020)
11. Wang, H., Li, M., Wan, Z.: Rail surface defect detection based on improved Mask R-CNN. Comput. Electr. Eng. **102**, 108269 (2022)
12. Cao, Z., et al.: Haze removal of railway monitoring images using multi-scale residual network. IEEE Trans. Intell. Transp. Syst. **22**(12), 7460–7473 (2020)
13. Cai, N, Chen, H, Li, Y, Peng, Y.: Intrusion detection and tracking at railway crossing. In: Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, pp. 1–6. (2019)
14. Qu, Z., Yi, W., Zhou, R., Wang, H., Chi, R.: Scale self-adaption tracking method of defog-PSA-Kcf defogging and dimensionality reduction of foreign matter intrusion along railway lines. IEEE Access **7**, 126720–126733 (2019)

15. Qin, Y., Cao, Z., Sun, Y., Kou, L., Zhao, X., Wu, Y., Jia, L.: Research on Active Safety Methodologies for Intelligent Railway Systems. Engineering (2022)
16. yolov5, https://github.com/ultralytics/yolov5. Last accessed 17 Nov 2022
17. Bernardin, K., Stiefelhagen, R.: Evaluating multiple object tracking performance: the clear mot metrics. EURASIP J. Image Video Process. **2008**, 1–10 (2008)
18. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. In: International Conference on Image Processing, pp. 3645–3649. IEEE (2017)