Research on the Algorithm of Cross-region Target Tracking in the Perimeter of High Speed Railway Based on Deep Learning

Ting Guo[1], Limin Jia[2,3], Zhengyu Xie[1,2,3(✉)], Yong Qin[2,3], and Feng Lin[4]

[1] School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China
20120808@bjtu.edu.cn
[2] State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing, China
[3] Key Lab of Proactive Safety and Risk Control in Railway Operation, Beijing, China
[4] National Railway Test Center, China Academy of Railway Sciences Co., Ltd, Beijing, China

Abstract. With the continuous development of science and technology, more and more fields are applying emerging video surveillance technology to ensure public safety. Compared with the way of manually processing video, the video monitoring technology based on deep learning also has the advantages of higher efficiency, high accuracy and intelligence when processing more and more massive video image data. In this paper, aiming at the detection requirements of personnel intrusion in the perimeter of high-speed railways, it realizes cross-region tracking of suspicious persons. It studies the application of deep learning cross-region tracking algorithm based on ResNet-50 in high-speed railway perimeter intrusion detection, and analyzes its reliability and optimizes accuracy. The degree reached 88%. Compared with the results of other cross-region tracking algorithms, the detection accuracy is still high. Therefore, it can be seen that the high-speed rail perimeter cross-region target tracking based on deep learning is feasible and efficient.

Keywords: High speed railway · Perimeter intrusion · Target tracking

1 Research Background and Significance

With the increase in the speed of trains and the expansion of the scale of the railway network, the importance of driving safety has further increased. If there is an intrusion of people on the perimeter of the high-speed rail, it may seriously endanger driving safety and affect the operating efficiency of the entire road network. The "Opinions on Strengthening the Governance of Safety and Environment Along Railways" issued in 2021 proposes to make full use of video monitoring, artificial intelligence and other means to improve the ability to investigate and manage potential safety and environmental problems along railways. The cross-region target tracking function meets the anti-terrorism needs of the high-speed rail perimeter, and alarms and tracks suspicious persons who have repeatedly entered the perimeter.

At present, manual investigation is mainly used to determine whether there are suspicious persons. This method is time-consuming and labor-intensive, and cannot effectively track suspicious persons, identify terrorists, and solve hidden dangers. When traditional algorithms deal with perimeter intrusions based on video images, they are easily affected by factors such as light, and cannot achieve cross-region tracking of suspicious persons. Therefore, this paper proposes a study on the perimeter cross-region target tracking of high-speed rail based on deep learning algorithm.

2   Research Status at Home and Abroad

The related issues of cross-regional target tracking were first raised by Cai Q et al. at the ICPR meeting in 1996 [1]. In 2006, N. Gheissari et al. first proposed the concept of pedestrian re-identification on CVPR [2]. Since then, related research on pedestrian re-identification has continued to develop.

The traditional manual feature cross-region target tracking has the following classic representatives: the author D. Gray and others combined color and texture features to extract pedestrian image features [3]. Author Farenzena et al. extracted features of three complementary aspects of human appearance [4]. S. Liao and others of the Institute of Automation of the Chinese Academy of Sciences proposed a feature representation method of Local Maximal Occurrence, referred to as LOMO, which is used for pedestrian re-identification [5], which reduces the interference of background noise.

In complex application scenarios, cross-region tracking algorithms based on manual features can no longer meet the needs. The emergence of algorithms based on deep learning is a leap forward in the performance of video image processing technology.

In 2014, the author W.Li et al. [6] proposed a new filter paired neural network (FPNN). In the same year, Simonyan et al. used a 3×3 size convolution filter instead of a 5×5 size filter, and increased the model's fitting ability by deepening the number of network layers [7]. Geng et al. used CNN to extract global features and merged classification loss and verification loss to solve the problem of data sparsity [8]. Zhang et al. only retain global features to calculate the similarity between images [9].

The cross-region tracking model based on deep learning has been well developed in theory, but it will encounter the problem of model generalization performance in the application process. Therefore, this paper studies the application of the ResNet-50 algorithm on the basis of constructing the data set of the perimeter of the high-speed rail, and examines the application effect of the model.

3   Cross-region Target Tracking Algorithm Based on ResNet-50

In order to realize the security monitoring of the perimeter along the high-speed rail, the need to identify, detect and track intruders is proposed. If criminals enter the perimeter of the high-speed rail, they will often obscure their facial information. They may also appear in multiple key areas on the perimeter of the high-speed rail, threatening the safety of high-speed rail traffic. Therefore, this question proposes a cross-region target tracking algorithm based on ResNet-50 to realize the identification and tracking of intruders.
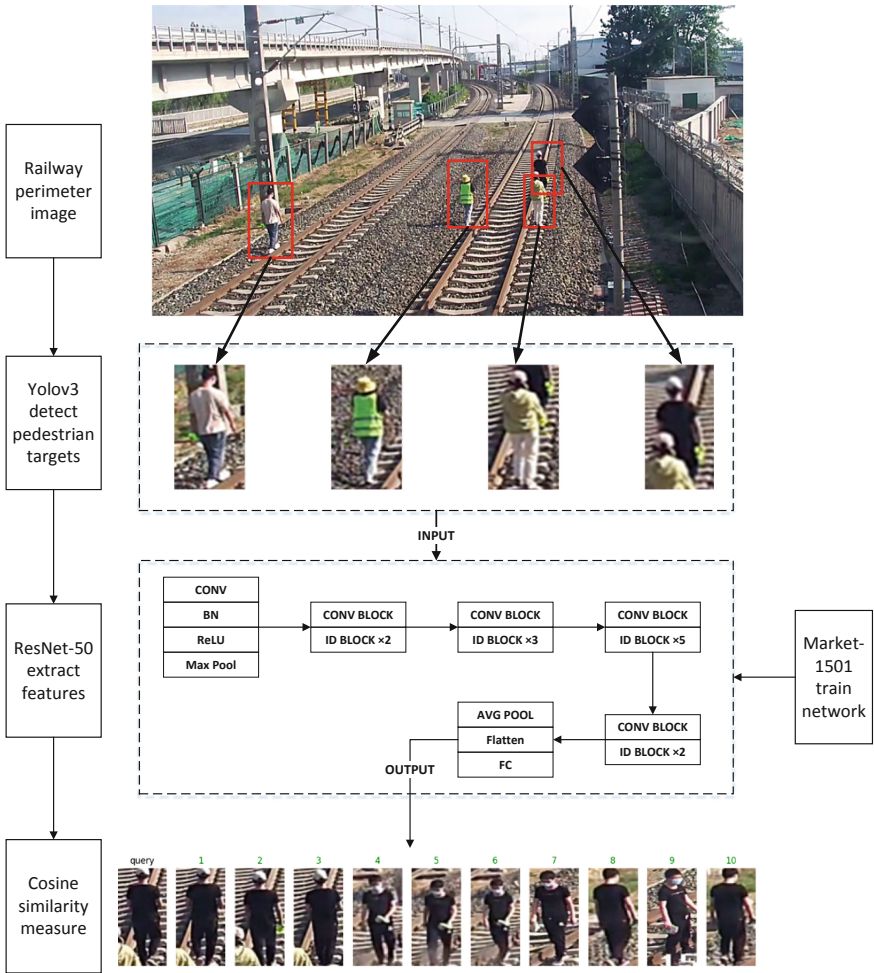
Fig. 1. Flow chart of cross-region target tracking algorithm based on DCGAN and ResNet-50

The research in this paper is mainly divided into three experimental stages, namely the detection of pedestrian targets, the feature extraction of target images, and the similarity matching of image features. Because yolov3 is a target detection network with relatively balanced speed and accuracy, it solves the problem of small object detection and can identify targets at 100m from high-speed rail monitoring. The background false detection rate is low, and it is easy to transplant to other platforms. Therefore, the yolov3 algorithm is selected as the method of pedestrian target detection.

In the second stage, ResNet50 is the basic feature extraction algorithm. The feature extraction model performs semi-supervised learning training to learn the characteristics of different pedestrians in the image. After the training, the target is tested and the features of the image are extracted. Finally, the cosine similarity is used to sort the distances

between the feature vectors from small to large, and the retrieval results are obtained. Compare and analyze the test results of different hyperparameter models (Fig. 1).

4 Experiment and Result Analysis

4.1 Generate Perimeter Data Set

The first stage is to use the yolov3 algorithm to detect the target pedestrian, and generate pedestrian images based on this. This experiment uses video data collected on an experimental railway line. A total of 1911 pedestrian images were intercepted, including 960 images in the query set and 951 images in the candidate set.

The captured image result is shown in the figure below (Fig. 2):

Fig. 2. Part of the perimeter dataset images

And annotated the image results of pedestrian target detection, and each image contains pedestrian id information, camera information, and so on.

4.2 Perimeter Image Feature Extraction and Matching

After the data set is generated, it is used as the input of the ResNet-50 feature extraction model, and the output is the high-level semantic feature value of the target image.

First, we must train the ResNet-50 feature extraction model. The training set uses the Market-1501 data set, a public data set for pedestrian re-identification. The Market-1501 data set is collected on the campus of Tsinghua University. The training set and test set can be set and used in single-target tracking or multi-target tracking tests.

The trained ResNet-50 feature extraction model extracts the image features in the high-speed rail perimeter data set, outputs the following matrix results, and calculates the cosine similarity of the feature values to obtain the similarity ranking between the target images (Fig. 3).

Query represents the target pedestrian image. The top 9 similar images retrieved from the candidate set are listed on the right. The green number indicates that the image retrieval result is correct, and the red number indicates that the image retrieval result is wrong.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	-0.0707	-0.0615	0.0142	-0.0073	0.0287	-0.0122	-0.0185	-0.0465	-0.0253	-0.0709	7.2851...	-0.0187	0.0364	0.0364	0.0371	0.0822
2	0.0020	-0.0116	-0.0427	-0.0058	0.0588	0.0164	-0.0059	-0.0255	0.0405	-0.0370	0.0191	-0.0610	0.0456	0.0324	0.0194	0.0119
3	-0.0270	-0.0482	-0.0297	-0.0364	0.0362	0.0155	-0.0429	-0.0421	0.0202	-0.0244	-0.0218	-0.0529	0.0028	0.0615	0.0245	0.0906
4	-0.0225	0.0144	0.0070	-0.0499	0.0808	0.0178	0.0021	-0.0761	-2.609...	0.0317	-0.0112	0.0182	0.0221	0.0230	0.0290	0.0990
5	-0.0790	-0.0371	-0.0375	-0.0172	0.0951	0.0285	0.0163	-0.0224	-0.0376	-0.0999	-0.0198	-0.0443	0.0100	0.0477	0.0178	0.0863
6	-0.0272	-0.0028	-0.0645	-0.0648	0.0716	0.0398	-0.0268	-0.0510	-0.0089	0.0381	-0.0079	-0.0235	-9.537...	0.0536	-0.0271	0.0848
7	-0.0590	0.0258	0.0495	-5.558...	0.0083	-0.0480	-0.0277	-0.0570	-0.0342	0.0140	0.0614	-0.0192	0.0187	0.0290	0.0656	0.0107
8	-0.0387	0.0042	0.0705	-0.0157	0.0190	-0.0511	-0.0048	-0.0531	-0.0249	0.0124	0.0276	0.0019	0.0294	0.0621	0.0186	0.0204
9	-0.0440	0.0070	0.0060	0.0060	0.0067	-0.0404	-0.0351	-0.0632	-0.0342	-0.0117	0.0709	-0.0187	0.0060	0.0266	-0.0036	0.0090
10	-0.0906	-0.0165	0.1139	0.0145	0.0154	-0.0064	0.0373	-0.0689	-0.0569	-0.0098	0.0445	0.0663	0.0257	0.0243	0.0562	0.0232
11	-0.0020	0.0165	0.0567	-0.0391	0.0148	-0.0381	-0.0164	-0.0298	0.0226	-0.0122	0.0388	-0.0447	0.0062	0.0647	0.0399	0.0070
12	0.0209	-0.0421	0.0984	-0.0019	0.0548	1.9596...	-0.0345	0.0318	0.0200	0.0072	0.0188	-0.0080	0.0337	-0.0119	-0.0364	-0.0172
13	-0.0353	-0.0757	0.0536	-0.0076	-0.0055	-0.0914	-0.0313	0.1019	-0.0255	2.0346...	-0.0245	-0.0192	0.0104	-0.0285	-0.0290	0.0122
14	0.0086	-0.0456	0.0717	-0.0182	0.0467	0.0394	-0.0355	0.0682	0.0109	-0.0274	0.0130	-0.0104	0.0484	-0.0940	0.0140	-0.0035
15	0.0390	-0.0365	0.0897	-0.0311	0.0238	0.0563	-0.0183	0.0537	-0.0289	0.0050	-0.0031	-0.0338	0.0535	-0.0930	0.0270	-0.0038
16	-0.0210	-0.0318	0.1004	0.0199	0.0404	-0.0046	-0.0277	0.1063	0.0140	-0.0450	-0.0279	0.0097	0.0141	-0.0640	-0.0434	-0.0134
17	-0.0332	0.0532	0.0247	-0.0299	0.0134	-0.0108	0.0362	0.0241	-0.0497	-2.627...	0.0051	-0.0092	4.5490...	0.0270	-0.0786	-0.0590
18	-0.0272	0.0118	-0.0659	-0.0713	0.0428	0.0573	0.0441	0.0660	0.0034	0.0303	0.0292	-0.0021	-0.0643	-0.0425	-0.0740	-0.0402
19	-0.0283	0.0979	0.0225	-0.0439	0.0347	0.0041	0.0395	-0.0191	-5.767...	-0.0097	-0.0125	0.0325	-0.0253	-0.0141	-0.0733	-0.0028
20	-0.0690	0.0291	0.0422	-0.0058	-0.0023	-0.0171	0.0491	0.0174	-0.0384	-0.0079	-1.542...	-0.0275	0.0281	0.0382	-0.0735	-8.679...
21	-0.0278	0.0462	-0.0333	-0.0544	0.0298	0.0189	0.0245	0.0597	-0.0455	-0.0288	0.0606	-0.0315	-0.0345	-0.0165	-0.0819	0.0016

Fig. 3. Image feature extraction result

The following pictures show part of the experimental results (Fig. 4):

The Rank-1 value of the ResNet-50 model on the high-speed rail perimeter data set reached 97.90%, and the Rank-10 reached 99.78%, with an average accuracy rate of 67.23%, and achieved good application results (Table 1).
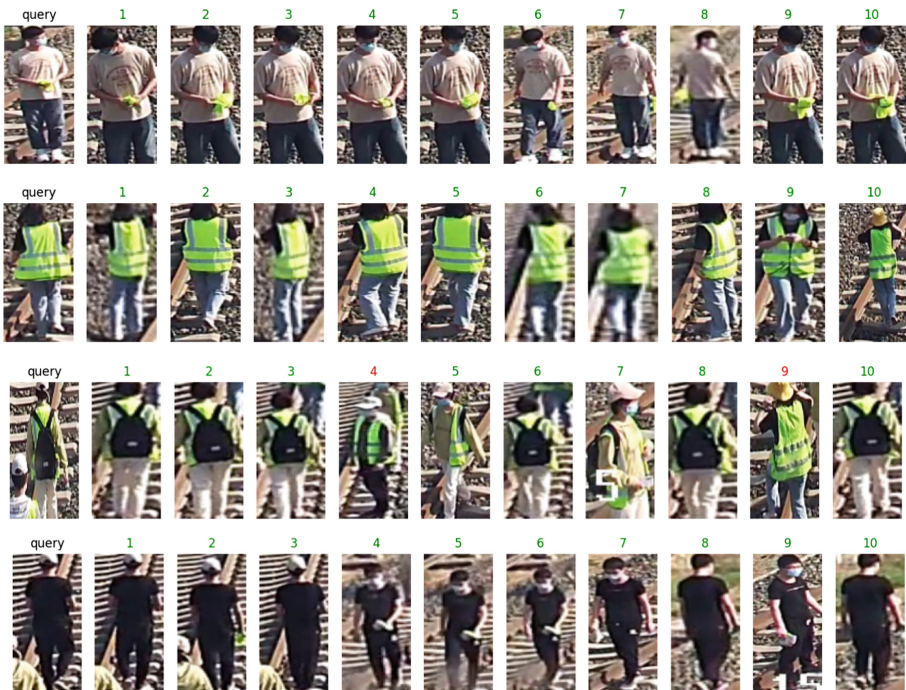

Fig. 4. Visualization results of cross-regional target tracking experiments

Table 1. Cross-region target tracking experimental training results

Evaluation metrics	Market-1501	Baseline
Rank-1	87.50%	97.90%
Rank-5	95.27%	99.37%
Rank-10	96.97%	99.78%
mAP	69.65%	67.23%

4.3   Results and Comparative Analysis

In order to further test the detection accuracy of cross-region tracking of intrusive targets in the high-speed rail perimeter and analyze the factors that affect the detection accuracy of the cross-region tracking algorithm, this paper optimizes the generated high-speed rail perimeter data set, and more strictly distinguishes the images of the two regions. The data is used as a query set and a gallery set, and some interference images in the data set are deleted. For example, different pedestrians wearing yellow vests are easily identified as the same target, which affects the accuracy of the cross-region tracking algorithm.

After optimizing the high-speed rail boundary data set, the ResNet-50 feature extraction model extracts feature values from the No-vest data set, and performs a cosine similarity measurement to evaluate the matching results. The adjusted Rank value is slightly lower than that of the original data set, but it is also above 95%. The result of Rank-10 also reaches 99%, which still has a high matching efficiency. The mAP value of the No-vest data set has increased significantly, reaching 88.84%, which is +21.41% better than the mAP result of the original data set. Therefore, it can be analyzed that the optimized data set has a positive effect on the experimental results (Table 2).

Table 2. Optimized data set ResNet-50 model comparison evaluation results

Evaluation metrics	Baseline	No-vest
Rank-1	97.90%	96.74%
Rank-5	99.37%	99.13%
Rank-10	99.78%	99.13%
mAP	67.23%	88.64%

The following table is the test results of this algorithm and the classic pedestrian re-identification method on the Market-1501 data set (Table 3).

Table 3. Different algorithm results

Method	mAP	Rank-1
TriNet [10]	69.14%	84.92%
DF [11]	63.40%	81.00%
PDC [12]	63.41%	84.14%
APR [13]	66.89%	87.04%

Compared with these algorithms, the experimental results of the algorithm in this article on the public data set show that the values of mAP and Rank-1 can be ranked high among them, and it has already had a good training effect. The application of the ResNet-50 feature extraction model on the high-speed rail perimeter data set, the experimental results are not weaker than these methods, and the optimized data set application is far superior to the results of the algorithms listed above.

5   Conclusion

In this paper, a deep learning-based cross-region target tracking algorithm for high-speed rail perimeters is carried out. The cross-region tracking of suspicious persons invaded by the perimeter of high-speed rail is realized. Good detection results are obtained on the high-speed rail perimeter data set, which can achieve maintenance driving purpose of security.

In future research, first of all, the high-speed rail perimeter data set needs to continue to be expanded, and the optimized network is continuously trained to improve detection accuracy. Secondly, the algorithm calculation process should be further optimized to reduce the calculation amount and complexity of the detection process. In addition, continue to study advanced technologies such as unsupervised learning and generative confrontation networks, and integrate the advantages of various algorithms to make the high-speed rail perimeter monitoring system more efficient and accurate.

References

1. Cai, Q., Aggarwal, J.K.: Tracking human motion using multiple cameras. IEEE (1996)
2. Gheissari, N., Sebastian, T.B., Hartley, R.: Person Reidentification Using Spatiotemporal Appearance: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06).2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (2006)
3. Gray, D., Tao, H.: Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features, pp. 262–275. Springer, Heidelberg (2008)

4. Farenzena, M., Bazzani, L., Perina, A., et al.: Person re-identification by symmetry-driven accumulation of local features. IEEE (2010)
5. Liao, S., Hu, Y., Zhu, X., et al.: Person re-identification by Local Maximal Occurrence representation and metric learning: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
6. Li, W., Zhao, R., Xiao, T., et al.: DeepReID: Deep Filter Pairing Neural Network for Person Re-identification: 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014 IEEE Conference on Computer Vision and Pattern Recognition (2014)
7. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition (2014)
8. Geng, M., Wang, Y., Xiang, T., et al.: Deep Transfer Learning for Person Re-identification (2016)
9. Zhang, X., Luo, H., Fan, X., et al.: AlignedReID: Surpassing Human-Level Performance in Person Re-Identification (2017)
10. Hermans, A., Beyer, L., Leibe, B.: In Defense of the Triplet Loss for Person Re-Identification (2017)
11. Zhao, L., Li, X., Zhuang, Y., et al.: Deeply-Learned Part-Aligned Representations for Person Re-identification. IEEE (2017)
12. Su, C., Li, J., Zhang, S., et al.: Pose-Driven Deep Convolutional Model for Person Re-identification. IEEE (2017)
13. Lin, Y., Zheng, L., Zheng, Z., et al.: Improving person re-identification by attribute and identity learning. Pattern Recogn. **95**, 151–161 (2019)