Journal of Intelligent & Robotic Systems (2022) 106: 9
https://doi.org/10.1007/s10846-022-01652-x

REGULAR PAPER

Application and Evaluation of Image-based Information Acquisition in Railway Transportation

Haifeng Song[1] · Xiying Song[2] · Hairong Dong[3]

Received: 10 October 2021 / Accepted: 22 April 2022 / Published online: 22 August 2022

Abstract
In railway transportation, the information is provided in literature and graphic styles. Generally, quite a lot information can not be obtained directly from the images. As a result, an artificial intelligence system, which can obtain information and perceive the environment, has to be established. In the driving equipment monitoring system, there is a lack of comprehensive analysis and utilization of the multiple monitoring data. This paper briefly introduces the research ideas and optimization directions of image-based data acquiring, such as template matching, support vector machine (SVM), and convolutional neural network (CNN) from the perspective of image detection. Then the characteristics, application scenarios, and possible future research directions of these three types of algorithms are compared and analyzed.

Keywords Computer vision · Railway transportation · Intelligent perceive

1 Introduction

Train modeling, control and scheduling lack consideration of line status and train status in abnormal scenarios, and it is difficult to achieve timely control and scheduling in equipment failure scenarios. Based on this, we started with two aspects of high-speed train line data perception and train internal state perception, and carried out related research, aiming to form a "vehicle-line" multi-domain high-speed railway operation information perception architecture.

Computer vision aims to automatically recognize and understand the images or videos and imitate biological vision to achieve effective and accurate environmental perception. Computer vision emerged in the 1950s and flourished in the 1980s, which has gradually matured with the development of artificial intelligence algorithms after the 21st century.

In recent years, computer vision has gradually penetrated security, transportation, medical and other fields under its real-time, fast, and accuracy, helping to reduce the inefficiency and unreliability caused by manual operations. The security field mainly includes three directions: face recognition [1, 2], pedestrian detection [3], and anomaly detection [4, 5]. License plate recognition [6, 7], automatic driving [8], and traffic sign recognition [9, 10] have been effectively applied in the transportation field. Besides, various auxiliary medical treatments including medical image recognition, cancer detection have also become a research focus [11, 12].

The structure of the computer vision system largely depends on the application direction. However, regardless of whether the system works independently in detection or measurementn or as a component of a large-scale complicated system, it should have some essential functions: image acquisition, pretreatment, feature extraction, and advanced task processing such as target segmentation, target detection, target recognition, target tracking.

Based on this structure, algorithms in the field of computer vision have undergone three stages of evolution.

✉ Hairong Dong
hrdong@bjtu.edu.cn

Haifeng Song
songhaifeng2011@gmail.com

Xiying Song
19120244@bjtu.edu.cn

[1] School of Electronic and Information Engineering, Beihang University, Beijing, 100191, China

[2] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, 100044, China

[3] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China

The first stage is based on image processing algorithms, such as geometric transformation, image rotation, to achieve segmentation and feature matching. However, such algorithms require massive prior knowledge, and in the case of the complex background, the detection accuracy is greatly reduced [13, 14]. Therefore, to improve detection efficiency and accuracy, the second phase is expanded based on traditional machine learning algorithms. The basic research idea is to extract features through HOG, SIFT, or other algorithms, and then train the classifier to achieve target detection or recognition. What is worthy of confirmation is that compared with the simple image processing in the first stage, the detection accuracy is improved due to the use of artificial design features or statistical features. However, different features need to be designed for different goals, a large amount of expert knowledge is required, and the robustness of the algorithm is low [15, 16]. So it enters the third stage, deep learning model based on convolution neural network has appeared. Through the model's self-learning ability, it can obtain richer and abstract general features, enhance the representation of image semantic information, and realize the leap-forward development of computer vision [17].

This paper takes image detection technology as the core, integrates the latest research results of computer vision, and selects typical examples from the above-mentioned three-stage algorithm development for comparison and explanation. the template matching and support vector machine (SVM) combining with deep learning are described in Sections 2 and 3 respectively. Section 4 describes the detection algorithm of the CNN. The comparative analysis is presented in Section 5. Finally, Section 6 concludes this study.

2 Based on Template Matching Algorithm

Template matching is one of the most typical methods in image recognition. It extracts several eigenvectors from the images to be recognized. This method compares images with the eigenvectors corresponding to the template and calculating the relative distance. Then by employing the minimum distance method, the image category could be determined. Therefore, template matching is usually used for detection with many consistent internal characteristics and is not affected by the background, such as pedestrian detection. Because early template matching algorithms are sensitive to rotation or scaling, matching based on feature points gets more attention. Among the feature-based matching algorithms, SIFT has high uniqueness and can maintain matching robustness under different viewing angles, affine transformations, and lighting conditions.

2.1 Principle of SIFT

In the traditional SIFT algorithm, to achieve the scale invariance use different Gaussian kernel functions to generate the corresponding scale-space. Then get the difference pyramid and extreme points of the discrete space. Since the discrete space is obtained by sampling, the extreme points sought maybe not true features. So low-contrast and unstable edge response points need to be delete. To achieve rotation invariance, the histogram is used to count the gradient direction and amplitude of all points in the neighborhood of the feature point. Then the gradient direction corresponding to the peak value is set as the main direction of the feature. Next, the eigenvector (x, y, σ, θ) can be established for each feature point with position, scale, and direction as dimensions. Finally, a descriptor is established to describe features, which will be used in the subsequent feature matching. However, the traditional SIFT has high computational complexity and time-consuming, cannot be applied in real-time engineering. Therefore, how to improve the algorithm to reduce the complexity and calculation time has become the main research direction.

2.2 Optimization of SIFT

Although SIFT is a powerful algorithm, it still needs further expansion and enhancement in the face of various application scenarios. The following four aspects of optimization are the main research directions.

- High computational complexity and time-consuming. In the traditional SIFT algorithm, each feature point is described by a 128-dimensional eigenvector, which cannot be applied in real-time engineering. To solve the problem, scholars have proposed methods to reduce the dimension of descriptors. PCA-SIFT (principle component analysis sift) [18], SURF (speed up robust feature) [19], and SSIFT (simplified SIFT) [20] have been successively proposed.
- Insufficient extraction of image features. Although the overall performance of SIFT is relatively high, it only uses the grayscale information of the image and does not make full use of the color features. Therefore, many scholars believe that the stability of SIFT can be further optimized to improve the matching effect. To this end, some scholars have proposed methods such as CSIFT (colored SIFT) [21], GLOH (gradient location and orientation histogram) [22], 3DSIFT (3-dimension SIFT) [23], and HOG (histogram oriented gradients) [24] from the perspectives of color feature extraction, adding dimensions, and expanding histogram statistics.

- The processing effect is unsatisfactory under similar background environments. The SIFT is based on the gradient statistics of local information. If there are a large number of similar local areas, the eigenvectors of feature points at different locations will be largely similar, thereby increasing the mismatch rate. Mortensen added global context information to SIFT, proposed SIFT+GC (SIFT + global context), and established a global context vector for each feature point in addition to a descriptor representing local characteristics [25].

In addition to these classic studies, a method for optimizing descriptors based on region division has also been proposed to reduce the computational complexity in recent years [26, 27]. And for different design requirements in actual engineering, there are also some corresponding improvement methods [28, 29]. So far, these algorithms' advantages are only reflected in some aspects, and specific problems still need related optimization.

3 Based on SVM

SVM is a generalized linear classifier based on the statistical VC dimension theory and the minimum structural risk principle, which classifies data in a supervised learning method. SVM can be divided into support vector regression (SVR) and support vector classification (SVC) according to purpose. The core idea is to find the optimal hyperplane in the feature space that meets the classification or regression requirements. For a nonlinear problem, its input can be mapped to a high-dimensional Hilbert space, so that the problem can be transformed into a linearly separable one. And the kernel function is introduced to solve the problem that the inner product of the mapping function is difficult to calculate.

3.1 Optimization of Kernel Function Parameters

SVM uses the kernel function to directly transform the inner product of two eigenvectors, which is equivalent to mapping the vector first and then doing the inner product, which simplifies the calculation. Common kernel functions include polynomial kernels, Gaussian kernels, and so on. Although the SVM theory is complete, there still exists that the kernel function parameters are difficult to choose in actuality. And different parameters have a significant impact on accuracy. Therefore, the optimization of kernel function parameters has always been one of the research hotspots of SVM. There are experimental and gradient descent methods in the early stage for kernel function parameters optimization. To solve the time-consuming and large errors of traditional algorithms, intelligent algorithms have been widely used, including but not limited to particle swarm optimization (PSO), fruit fly optimization, genetic algorithm [30–32]. Through the parameter optimization algorithm, the accuracy of SVM is significantly improved.

3.2 Multi-class SVM

SVM mainly aims at two classification problems in the early phase, but it is generally a multi-classification problem in practical applications. Therefore, how to use SVM in multi-classification problems has become an improvement direction of SVM. The multi-class SVM is proposed as a result, and there are two main research ideas:

- Optimize the objective function of the classic SVM and construct a multi-classification model to realize the multi-classification problem. This method is a one-time solution. Because of its high computational complexity and low efficiency in practical applications, it is not commonly used.
- The multi-classification problem is reduced to multiple two-classification problems so that a complex problem is transformed into many simple problems. This type of algorithm has been greatly developed. Commonly used are one-to-many [33], one-to-one [34], guided acyclic graph [35], and binary tree [36]. The four algorithms are analyzed and verified theoretically from training complexity, test complexity, and classification accuracy [37]. It can be found that the classification performance of the guided acyclic graph is the best, and the one-to-many is the worst. Besides, in the training and testing phase, the binary tree takes the shortest time, and the one-to-one and one-to-many time-consuming is relatively long.

3.3 Fuzzy SVM

Fuzzy support vector machine (FSVM) is proposed to mainly solve the influence of data noise on the prediction model [38]. The core idea is to combine fuzzy mathematics with SVM to give the support vector a higher subjection degree and a lower subjection degree to the noise, thus realize the separation of noise. FSVM can improve the classification accuracy and solve the over-learning problem caused by abnormal data. However, in practical applications, FSVM also has some shortcomings. If there are too many abnormal data, or they all obey a certain distribution, using FSVM to eliminate data will cause information loss, which makes the training model's generalization ability insufficient. Besides, FSVM also requires significant computation for kernel function, storage resources, and long training time. So the optimization of

FSVM has become the direction of some researchers [39, 40].

3.4 SVM Combines Deep Learning

As a machine learning model, SVM has unsatisfactory data representation capabilities. But deep learning can extract richer and more abstract features. Therefore, the deep model integrating deep learning and SVM has become a research focus in recent years. At present, there are mainly deep belief networks (DBN) and convolutional neural networks (CNN) that have been introduced as the underlying models. DBN is suitable for processing single-dimensional time-series data, while CNN is more suitable for processing image and multi-dimensional time-series forecasting problems [41, 42]. Usually, the output layer of DBN and CNN is the SoftMax classifier. Using SVM to replace SoftMax can effectively improve the accuracy of the model. However, the current research is not good enough to solve the parameter optimization of SVM.

4 Based on Convolutional Neural Network

With the emergence of backpropagation algorithms and high-performance computing systems, neural networks have gradually gained more attention. And achieve significant progress in the design of network structure and training strategy. Neural networks of different structures have emerged, such as AlexNet, VGG, GoogleNet, and ResNet.

The CNN is the most representative model of deep learning. Compared with traditional algorithms, CNN can learn multi-level representations from pixels to high-level semantic features through a multi-layer hierarchical structure, thereby obtaining richer hidden information and increasing expression capabilities. Secondly, the CNN architecture makes multi-task learning possible. Thus, deep convolutional neural network (DCNN) has gradually become a new direction for target detection. The mainstream deep learning image detection are mainly divided into two-stage and one-stage detection, and they are mainly different in research ideas and optimization targets.

4.1 Two-stage Algorithm

The two-stage algorithm mainly selects candidate regions that may contain the target through selective search or Edge Boxes, then classifies and regresses the candidate regions to obtain the detection results. Typical algorithms are R-CNN, R-FCN, Mask R-CNN.

1) **R-CNN:** Girshck proposes R-CNN that combines region proposal and CNN for achieving target location, which created a precedent for the neural network to achieve target detection [43]. R-CNN has an mAP of 58.5% on the VOC2007 data set. Compared with the traditional algorithm, R-CNN has been improved, but there are also significant problems. Including long training time and difficulty in optimization, large space consumption, and slow detection speed. In response to these problems, based on R-CNN, SPP-Net (Spatial Pyramid Pooling Network) is proposed to greatly increase the training speed [44]. Then Fast R-CNN, Faster R-CNN improve the training speed and accuracy [45, 46].
2) **Mask R-CNN:** In 2017, He improved again based on Faster R-CNN and proposed the Mask R-CNN [47]. By adding the Mask branch, target detection and semantic segmentation can be achieved simultaneously. The advantage of the Mask R-CNN is that it adds the target mask as the output based on the Faster R-CNN, which makes the spatial layout extraction finer. The detection accuracy on the COCO dataset has increased from 19.7% of Fast R-CNN to 39.8%. The disadvantage is that the segmentation branch increases calculation, resulting in Mask R-CNN has a slower detection speed than Faster R-CNN.

4.2 One-stage Algorithm

The one-stage algorithm uses regression analysis, omits the step of selecting candidate regions, and directly obtains the target position. Take the YOLO and SSD series as typical representatives.

1. **YOLO:** Because of the inefficiency of the two-stage algorithm, Joseph Redmon proposed YOLO (You Only Look Once) in 2016 [48]. In real-time conditions, the detection speed of YOLO is 45 FPS, and the average detection accuracy mAP is 63.4%. But YOLO's detection effect on small-scale targets is not good, and it is easy to miss the detection in the environment of overlapping targets. Corresponding improvements of YOLO have been successively proposed for the problems of inaccurate positioning, poor detection of small targets, and low detection accuracy [49–52].
2. **SDD:** Combining Faster RCNN and YOLO, Liu proposed the SSD (single shot multibox detector) algorithm to balance detection accuracy and speed [53]. Adopt the target prediction mechanism, which reduces calculation and can effectively detect groups of small target. The running speed of SSD on Nvida Titan X

is increased to 59 FPS, which is significantly better than YOLO, and the mAP on the VOC2007 data set reaches 79.8%. Because the feature maps of different scales are independent of each other, the SSD has a poor classification effect on small targets. To further improve the model, DSSD (deconvolutional single shot detector), F-SSD (feature fusion SSD), DSOD (deeply learning supervised object detectors) have been proposed successively [54–56].

5 Image Pretreatment for Multi-level Railway Operation Information Perception

5.1 Literature Recognition

Generally, quite a lot information is provided in literature styles. Hence, the image-based sequences recognition technologies are indispensable. Traditional Optical Character Recognition (OCR) can be utilized to extract those image-based sequence and it does performs well in recognizing texts that are clear and simple. However, the literature information recognition performance seems getting worse when the noise increases. As a result, some other methods using CNN are proposed. Some CNN methods carry out the recognition by segmenting the text into individual parts first, and then recognizing each part. This method may perform well, but it takes more time to prepare for the data acquired. Some other methods utilize the combination of convolutional layer and the transcription layer like connectionist temporal classification to implement an end-to-end recognition. This way can achieve the recognition of imaged-based text without wasting too much time on data preparation. Besides, the identification result is not acceptable.

5.2 DMI Based Train Speed and Location Data Acquisition

In railway on-board equipment, as the bridge between the train driver and train control system, the driver machine interface (DMI) has both text and various icons, and there are many types. However, when extracting the essential operation information, it is only necessary to selectively extract valuable one for the control and scheduling analysis. By analyzing the actual requirements of vehicle equipment testing, it is determined that there are a total of four different parts, which are speed, mile mark, gradient, and train ID, as shown in Fig. 1.

Since the original image contains both the DMI and the background, it is necessary to extract the entire DMI interface first, and then segment each interface element to be identified from the DMI. Commonly used image segmentation methods are threshold, region-growing and edge detection based ones. Because both threshold-based and region-growing-based image segmentation methods require pixels in the foreground image to have similar grayscale features, the DMI has a large area of black background in addition to displaying some surrounding images. Hence, a method based on the combination of edge detection and Hough transform is used to locate the DMI. The tilt correction and recognition of DMI is shown in Fig. 2.

If the colors of the target and the background are the same. If only the threshold segmentation method is used, the digital part of the image will be lost, so special processing is required. The processing steps are as follows:

1) Use the threshold segmentation method to obtain the binary image without the digital part;
2) Invert the color of the binary image, the digital part becomes the foreground, but the original background

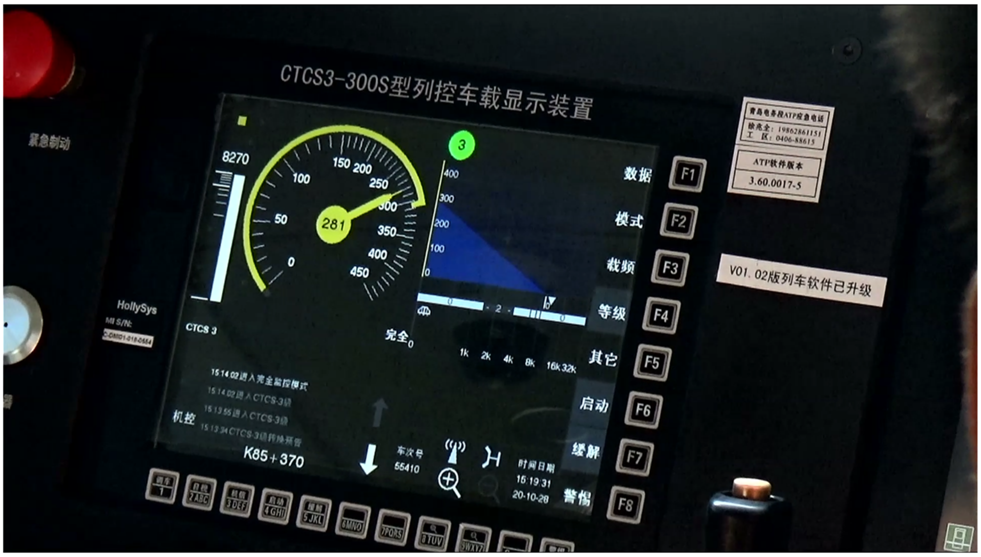


Fig. 1 Actual on-board DMI equipment

part also becomes the foreground, and the digital part cannot be segmented;

3) Using the regional growth method, take the four corners of the image as seed growth points, find the background part and set its pixel value to zero.

5.3 Foreign Body Intrusion

As the speed of high-speed railway trains increases, the kinetic energy of trains increases. If there is abnormal intrusion, it is difficult for the train to brake and stop in a short time. Therefore, it is necessary to set up an intrusion monitoring device to detect the intrusion in time and take countermeasures in advance.

Foreign body intrusion on railway tracks is the largest type of railway accidents, accounting for about 20% of the total number of railway accidents. Among them, the intrusion of pedestrians and motor vehicles is the most serious, which is a relatively large accident recognized by the railway department. The research difficulty lies in the sparsely sampled video data and the moving vehicle-mounted camera, which leads to a dynamically changing intrusion scene within the field of view, which increases the difficulty of identifying the intrusion limit of foreign objects. It can be solved by designing a feature extraction, feature cropping and feature compression modules, as shown in Fig. 3. The backbone consists of different neural network, and this part is utilized to extract feature. Then, spatial feature fusion (SFF) and deep feature flow (DFF) were used to process the feature extracted from backbone. After that, Fuzzy C-means (FCM) was used to process features from SFF and DFF and obtained the cropped feature. Through the region proposal network (RPN) and region of interest (ROI) pooling, picture with anchor boxes were obtained. Finally, after the intersection of two binary plot, the intrusion detection can be obtained. The performance indicators of the proposed algorithm surpass the existing algorithms in segmentation and detection tasks.

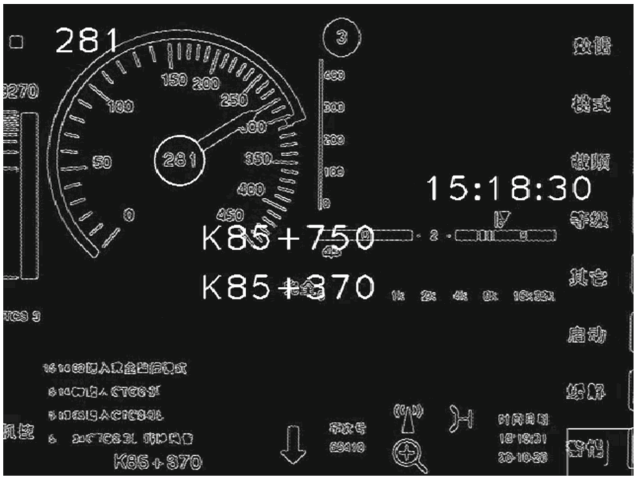


Fig. 2 DMI image with tilt correction and recognition

6 Comparison and Analysis

SIFT-based image matching is currently widely used in many scenarios, such as mobile robot positioning recognition and map generation, panoramic puzzles, iris recognition, license plate recognition, and 3D reconstruction. Although the SIFT has been successful in many fields, it still has some difficulty solving problems, such as low real-time performance and the inability to extract features for smooth edges. The main development direction of SIFT in the future may focus on two aspects:

- Combine with deep learning algorithms to obtain better application effects.
- Research on faster image retrieval and matching algorithms is of great significance in occasions with high real-time requirements.

SVM is a machine learning method based on the statistical theoretical framework and has shown many superior performances in practical applications. Under the idea of minimizing structural risks, SVM has excellent generalization ability and can better solve problems such as nonlinear data, small samples, and dimensionality disasters. Because of these strengths, SVM has been developed in pattern recognition, regression analysis, time-series prediction, and is widely used in traffic (passenger flow prediction, incident detection), speech and face recognition, gene classification. But SVM also has its limitations. For example, it does not work well for large-scale data training; the kernel function parameters are difficult to determine; theoretically, SVM can only provide sub-optimal solutions. So SVM can continue to explore the following aspects:

- Optimize the kernel function and its parameters to improve the training efficiency and generalization ability of SVM.
- Integrate SVM with other disciplines. New algorithms such as FSVM and CNN-SVM are proposed by combining SVM with fuzzy mathematics and CNN in recent years. Although there are improvements in training time and generalization performance, they also have their shortcomings. How to improve existing theories, put forward more reasonable models, and integrate new disciplines to achieve further optimization of SVM is worth considering.

At present, the target detection algorithm based on deep learning has received extensive attention and research. It is mainly divided into two categories: two-stage based on

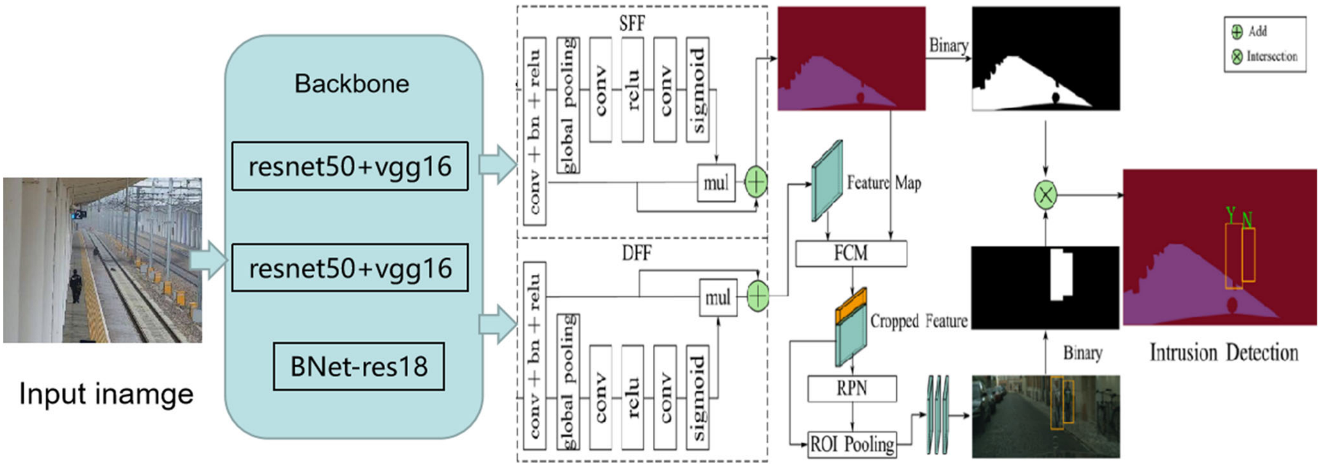
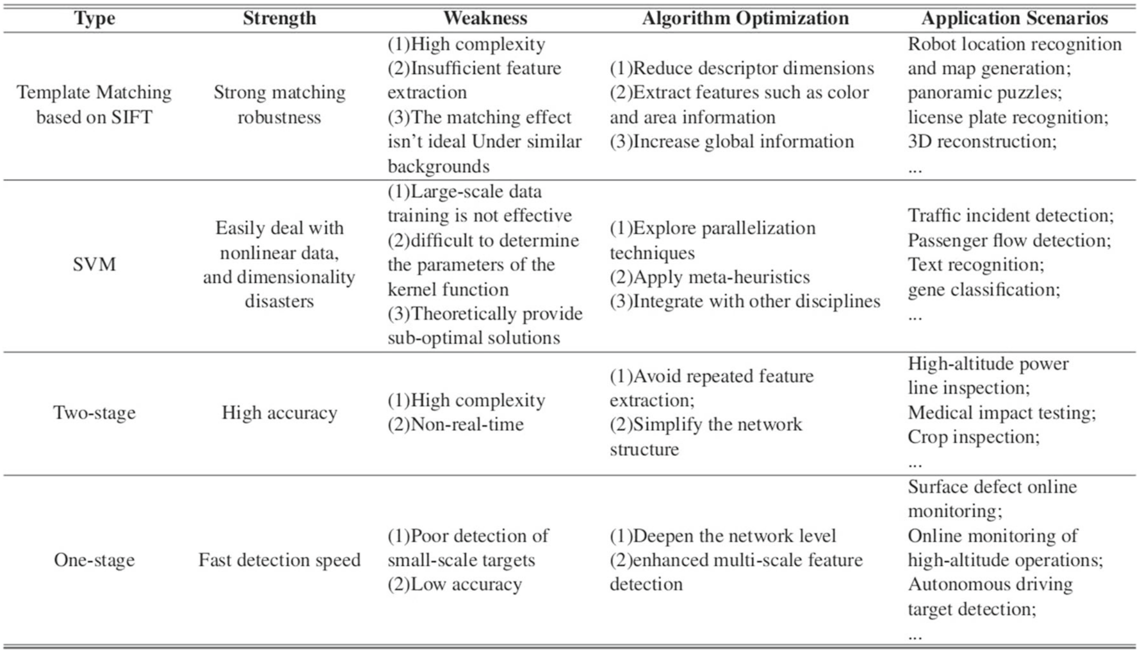


Fig. 3 Model for pedestrian intrusion detection in designated areas

candidate regions and one-stage based on regression. Two-stage aims at higher detection accuracy and sacrifice speed. It is more suitable for applications in scenes with low real-time detection. One-stage takes the detection speed as the optimal index to improve the accuracy as much as possible and is used in high real-time scenarios, such as automatic driving target detection and surface defect online detection. Both types of algorithms have derived different neural network mechanisms and characteristics, and have made corresponding achievements. But there still face some problems waiting to be solved.

- Small target detection. Currently, UAV aerial photography, satellite telemetry, and other fields propose higher requirements for small target detection capabilities. But the existing algorithms are still unable to meet.
- Weakly supervised target detection method. At present, the accuracy of most algorithms relies on a large number of labeled complete data sets. However, large-scale data annotation requires plenty of time. The implementation cost is too high. Therefore, it is crucial to study how to use a small amount of labeled data to



Type	Strength	Weakness	Algorithm Optimization	Application Scenarios

Template Matching based on SIFT	Strong matching robustness	(1)High complexity
		(2)Insufficient feature extraction
		(3)The matching effect isn't ideal Under similar backgrounds
			(1)Reduce descriptor dimensions
			(2)Extract features such as color and area information
			(3)Increase global information
				Robot location recognition and map generation;
				panoramic puzzles;
				license plate recognition;
				3D reconstruction;
				...

SVM	Easily deal with nonlinear data, and dimensionality disasters	(1)Large-scale data training is not effective
		(2)difficult to determine the parameters of the kernel function
		(3)Theoretically provide sub-optimal solutions
			(1)Explore parallelization techniques
			(2)Apply meta-heuristics
			(3)Integrate with other disciplines
				Traffic incident detection;
				Passenger flow detection;
				Text recognition;
				gene classification;
				...

Two-stage	High accuracy	(1)High complexity
		(2)Non-real-time
			(1)Avoid repeated feature extraction;
			(2)Simplify the network structure
				High-altitude power line inspection;
				Medical impact testing;
				Crop inspection;
				...

One-stage	Fast detection speed	(1)Poor detection of small-scale targets (2)Low accuracy	(1)Deepen the network level (2)enhanced multi-scale feature detection	Surface defect online monitoring; Online monitoring of high-altitude operations; Autonomous driving target detection; ...

Fig. 4 Comparison and Summary of Different Algorithms

automatically detect unlabeled data and achieve model training.

- Multi-domain target detection. At present, most algorithms design specific models for specific scenarios and targets. It can achieve better detection performance on the specified data set but cannot be used in multiple fields. Therefore, exploring multi-domain detectors has great high significance.

Figure 4 collectively shows the comparison between the described algorithms.

7 Conclusion

This paper briefly introduces three kinds of classical target detection algorithms from the research ideas and related optimization. Then the characteristics, application scenarios, and possible future research directions of these algorithm are compared and analyzed. Although these algorithms have achieved excellent results in specific fields respectively, no general detectors suitable for various image detection fields have been found. How to combine deep learning to achieve cross-domain image detection is a challenging subject and remains to be further research.

Author Contributions Haifeng Song contributes to make the algorithm for target detection and integrate the algorithm to our railway transportation. Xiying Song and Hairong Dong contributes to review the article and supervise our research.

Funding This work is supported jointly by the National Natural Science Foundation of China under Grant 61925302, 61903021, and Beijing Natural Science Foundation L211021.

Declarations

Consent for Publication The authors declare that they consent to publish this article.

Conflict of Interests The authors declare that they have no competing interest.

References

1. Ranjan, R., Patel, V.M., Chellappa, R.: Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition[J]. IEEE Trans Pattern Anal Mach Intell **41**(1), 121–135 (2017)
2. He, R., Wu, X., Sun, Z., et al: Wasserstein CNN: Learning invariant features for nirvis face recognition[J]. IEEE Trans Pattern Anal Mach Intell **41**(7), 1761–1773 (2018)
3. Braun, M., Krebs, S., Flohr, F., et al: Eurocity persons: A novel benchmark for person detection in traffic scenes[J]. IEEE Trans Pattern Anal Mach Intell **41**(8), 1844–1861 (2019)
4. Barz, B., Rodner, E., Garcia, Y.G., et al: Detecting regions of maximal divergence for spatio-temporal anomaly detection[J]. IEEE Trans Pattern Anal Mach Intell **41**(5), 1088–1101 (2018)
5. Sabokrou, M., Fathy, M., Hoseini, M.: Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder[J]. Electron. Lett. **52**(13), 1122–1124 (2016)
6. Li H, Wang, P, Shen, C.: Toward end-to-end car license plate detection and recognition with deep neural networks[J]. IEEE Trans Intell Transp Syst **20**(3), 1126–1136 (2018)
7. Shivakumara, P., Tang, D., Asadzadehkaljahi, M., et al: CNN-RNN based method for license plate recognition[J]. CAAI Trans Intell. Technol. **3**(3), 169–175 (2018)
8. Lu, W., Zhou, Y., Wan, G., et al: L3-net: Towards learning based lidar localization for autonomous driving[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6389–6398 (2019)
9. Arcos-García, Á., Alvarez-Garcia, J.A., Soria-Morillo, L.M.: Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods[J]. Neural Netw. **99**, 158–165 (2018)
10. Zhou, S., Liang, W., Li, J., et al: Improved VGG model for road traffic sign recognition[J]. Computers, Materials and Continua **57**(1), 11–24 (2018)
11. Li, L., Xu, M., Wang, X., et al: Attention based glaucoma detection: A large-scale database and cnn model[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10571–10580 (2019)
12. Liu Q, Fang, L., Yu, G., et al: Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data[J]. Nature Commun **10**(1), 1–11 (2019)
13. Li, Q., Ren, S.: A real-time visual inspection system for discrete surface defects of rail heads[J]. IEEE Trans Ins Meas **61**(8), 2189–2199 (2012)
14. Han, Y., Liu, Z., Han, Z., et al: Research on detection of ear piece fracture of catenary support device of high-speed railway based on SIFT feature matching[J]. J China Railw Soc **36**(2), 31–36 (2014)
15. Liu, L., Zhou, F., He, Y.: Automated status inspection of fastening bolts on freight trains using a machine vision approach[J]. In: Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 230, pp. 1629–1641 (2016)
16. Liu, L., Zhou, F., He, Y.: Automated visual inspection system for bogie block key under complex freight train environment[J]. IEEE Trans Instrum Meas **65**(1), 2–14 (2015)
17. Sun, J., Xiao, Z., Xie, Y.: Automatic multi-fault recognition in TFDS based on convolutional neural network[J]. Neurocomputing **222**, 127–136 (2017)
18. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors[C]. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. II–II (2004). CVPR 2004. IEEE, 2004
19. Bay H, Tuytelaars T, Van Gool L: Surf: Speeded up robust features[C] European Conference on Computer Vision, pp. 404–417. Springer, Heidelberg (2006)
20. Liu, L., Peng, F., Zhao, K., et al: Simplified SIFT algorithm for fast image matching[J]. Infrared Laser Eng **37**(1), 181–184 (2008)
21. Abdel-Hakim, A.E., Farag, A.A.: CSIFT: A SIFT descriptor with color invariant characteristics[C]. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Ieee, vol. 2, pp. 1978–1983 (2006)
22. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors[J]. IEEE Trans Pattern Anal Mach Intell **27**(10), 1615–1630 (2005)
23. Scovanner, P., Ali, S., Shah, M.: A 3-dimensional sift descriptor and its application to action recognition[C]. In: Proceedings of the 15th ACM International Conference on Multimedia, pp. 357–360 (2007)

24. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection[C]. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, vol. 1, pp. 886–893 (2005)
25. Mortensen, E.N., Deng, H., Shapiro, L.: A SIFT descriptor with global Context[C]. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, vol. 1, pp. 184–190 (2005)
26. Feng, W., Liu, B.: Research on improved SIFT algorithm for image matching[J]. Comput Eng Appl **54**(03), 200–205+232 (2018)
27. Zhou, X., Wang, K., Fu, J.: A method of SIFT simplifying and matching algorithm improvement[C]. In: 2016 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration. IEEE, pp. 73–77 (2016)
28. Zhou, D., Wu, Y., Yao, Y.u.: Medical image retrieval based on feature fusion of global feature and scale-invariant feature conversion[J]. J Comput Appl **35**(04), 1097-1100+1105 (2015)
29. Geng, Q., Zhao, H., Wang, Y., Zhao, H.: Licence plate recognition based on improved SIFT feature extraction[J]. Opt Precis Eng **26**(05), 1267–1274 (2018)
30. Lu, W.X., Li, C.: Forecasting of short-time tourist flow based on improved PSO algorithm optimized LSSVM model[J]. Comput Eng Appl **55**(18), 247–255 (2019)
31. Cong, Y.L., Wang, J.W., Li, X.L.: Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm[J]. Procedia Engineering **137**(1), 59–68 (2016)
32. Fang, Z., Yu, B., Xiao, W., et al: Identifying travel mode with GPS data using support vector machines and genetic algorithm[J]. Information **6**(2), 212–227 (2015)
33. Polat, K., Güneş S: A novel hybrid intelligent method based on C4. 5 decision tree classifier and one-against-all approach for multi-class classification problems[J]. Expert Syst Appl **36**(2), 1587–1592 (2009)
34. Chaudhuri, A., De, K., Chatterjee, D.: A comparative study of kernels for the multi-class support vector machine[C]. In: 2008 Fourth International Conference on Natural Computation. IEEE, vol. 2, pp. 3–7 (2008)
35. Manikandan, J., Venkataramani, B.: Study and evaluation of a multi-class SVM classifier using diminishing learning technique[J]. Neurocomputing **73**(10-12), 1676–1685 (2010)
36. Wu, D.: Research on intelligent aided quality diagnosis based on multi-class support vector machines[J]. J Syst Simul **21**(6), 1689–1693 (2009)
37. Xue, N.: Comparative research on multi-class support vector machine classifier[J]. Compu Eng Design **32**(5), 1792–1795 (2011)
38. Lin, C.F., Wang, S.D.: Fuzzy support vector machines[J]. IEEE Trans Neural Netw **13**(2), 464-471 (2002)
39. Liu, Y., Huang, H.: Fuzzy support vector machines for pattern recognition and data mining[J]. Int J Fuzzy Syst **4**(3), 826–835 (2002)
40. Samma, H., Lim, C.P., Saleh, J.M., et al: A memetic-based fuzzy support vector machine model and its application to license plate recognition[J]. Memetic Computing **8**(3), 235–251 (2016)
41. Niu, X.X., Suen, C.Y.: A novel hybrid CNN–SVM classifier for recognizing handwritten digits[J]. Pattern Recognit **45**(4), 1318–1325 (2012)
42. Zhu, L., Chen, L., Zhao, D., et al: Emotion recognition from Chinese speech for smart affective services using a combination of SVM and DBN[j]. Sensors **17**(7), 1694 (2017)
43. Girshick, R., Donahue, J., Darrell, T., et al: Rich feature hierarchies for accurate object detection and semantic segmentation[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587 (2014)
44. He, K., Zhang, X., Ren, S., et al: Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Trans Pattern Anal Mach Intell **37**(9), 1904–1916 (2015)
45. Girshick, R.: Fast r-cnn[C]. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
46. Ren S, He, K., Girshick, R., et al: Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell **39**(6), 1137–1149 (2016)
47. He, K., Gkioxari, G., Dollár, P., et al: Mask R-CNN[C]. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969 (2017)
48. Redmon, J., Divvala, S., Girshick, R., et al: You only look once: Unified, real-time object detection[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
49. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)
50. Redmon, J., Farhadi, A.: Yolov3:, An incremental improvement[J]. arXiv:1804.02767 (2018)
51. Bochkovskiy, A., Wang, C.Y., Liao, H.YM.: Yolov4:, Optimal speed and accuracy of object detection[J]. arXiv:2004.10934 (2020)
52. Wang, X., Liu, M., Raychaudhuri, D.S., et al: Learning person re-identification models from videos with weak supervision[J]. IEEE Trans. Image Process. **30**, 3017–3028 (2021)
53. Liu, W., Anguelov, D., Erhan, D., et al: Ssd: Single shot multibox detector[C] European Conference on Computer Vision, pp. 21–37. Springer, Cham (2016)
54. Fu, C.Y., Liu, W., Ranga, A., et al: Dssd:, Deconvolutional single shot detector[J]. arXiv:1701.06659 (2017)
55. Li, Z., Zhou, F.: FSSD:, feature fusion single shot multibox detector[J]. arXiv:1712.00960 (2017)
56. Shen Z, Liu Z, Li J, et al: Dsod: Learning deeply supervised object detectors from scratch[C]. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1919–1927 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Haifeng Song received his B.Sc. degree in 2011 from Beijing Jiaotong University (China), the Master degree in Traffic Information Engineering and Control from the same university in 2014. He received his Dr.-Ing degree from Technische Universität Braunschweig, Germany in 2018. He has joined the Institute for Traffic Safety and Automation Engineering, Technische Universität Braunschweig (Germany) as a visiting scientist since 2014. He is an Associate Professor with the School of Electronic and Information Engineering, Beihang University, China. He specializes in safety and security of transportation systems, his current research interests include railway control system, formal method, intelligent control, and transportation modeling. He has been involved in several national and international research projects dealing with system safety and system evaluation. He serves as an associate editor for *IEEE Transactions on Intelligent Vehicles*. He is a member of IEEE Intelligent Transportation Systems Society, Chinese Association of Automation, China Institute of Communications and a reviewer for international journals.

Xiying Song received the B.S. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China, in 2019. She is currently working toward the Ph.D. degree in traffic information engineering and control with the State Key Laboratory of Rail Traffic Control and Safety. Her research interests include computer vision and automomous train technologies.

Hairong Dong earned her Ph.D. degree from Peking University in 2002. She is the deputy director of the National Engineering Research Center for Rail Transportation Operation Control Systems and a professor in the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China. Her research interests include intelligent transportation systems, automatic train operation, intelligent dispatching, and complex network applications. She serves as an associate editor for *IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Intelligent Vehicles, IEEE Intelligent Transportation Systems Magazine, and Journal of Intelligent and Robotic Systems*. She is a fellow of the Chinese Automation Congress, a Senior Member of IEEE, and co-chair of the IEEE Intelligent Transportation Systems Society Technical Committee on Railroad Systems and Applications.