Human intrusion detection for high-speed railway perimeter under all-weather condition

Pengyue Guo, Tianyun Shi and Zhen Ma
Institute of Electronic Computing Technology, China Academy of Railway Sciences Corporation Limited, Beijing, China, and
Jing Wang
School of Information and Electronics, Beijing Institute of Technology, Beijing, China

Received 12 November 2023
Revised 19 December 2023
Accepted 19 December 2023

Abstract

Purpose – The paper aims to solve the problem of personnel intrusion identification within the limits of high-speed railways. It adopts the fusion method of millimeter wave radar and camera to improve the accuracy of object recognition in dark and harsh weather conditions.

Design/methodology/approach – This paper adopts the fusion strategy of radar and camera linkage to achieve focus amplification of long-distance targets and solves the problem of low illumination by laser light filling of the focus point. In order to improve the recognition effect, this paper adopts the YOLOv8 algorithm for multi-scale target recognition. In addition, for the image distortion caused by bad weather, this paper proposes a linkage and tracking fusion strategy to output the correct alarm results.

Findings – Simulated intrusion tests show that the proposed method can effectively detect human intrusion within 0–200 m during the day and night in sunny weather and can achieve more than 80% recognition accuracy for extreme severe weather conditions.

Originality/value – (1) The authors propose a personnel intrusion monitoring scheme based on the fusion of millimeter wave radar and camera, achieving all-weather intrusion monitoring; (2) The authors propose a new multi-level fusion algorithm based on linkage and tracking to achieve intrusion target monitoring under adverse weather conditions; (3) The authors have conducted a large number of innovative simulation experiments to verify the effectiveness of the method proposed in this article.

Keywords High-speed rail perimeter, Personnel invasion, Object detection, All-weather, Radar-camera fusion

Paper type Research paper

1. Introduction

The 20th National Congress of the Communist Party of China explicitly stated the need to improve the capacity of disaster prevention, mitigation and relief, and the handling of emergencies, dangers and public emergencies regarding the aspect of social stability. In the field of railway transportation, China has successfully established the most advanced high-speed rail network globally. By the end of 2022, the operational mileage of high-speed railways has reached 42,000 kilometers, with passenger transportation exceeding 25.9bn individuals over the past decade. Consequently, the safety of high-speed rail operations constitutes a pivotal component in safeguarding national security and social stability.

"The Regulations on Security and Protection Management of High-Speed Railways" (Measures for the Administration of High-speed Railway Safety Protection, 2020) came into

This research was supported by the National Natural Science Foundation of China [U2268217].

Railway Sciences
Vol. 3 No. 1, 2024
pp. 97-110
Emerald Publishing Limited
e-ISSN: 2755-0915
p-ISSN: 2755-0907
DOI 10.1108/RS-11-2023-0043

effect on July 1, 2020, requiring railway transport enterprises to install perimeter intrusion alarm systems in key sections along high-speed railway lines, such as bridgeheads, tunnel entrances, and track bed areas, where access is relatively easy. Due to the diverse and changeable climate types, intricate terrain along railway lines, and the subjective malintent of individuals, uncontrollable risk points exist along the railway lines, resulting in occasional perimeter intrusions Despite employing multiple technological measures for high-speed railway perimeter protection, single technologies still suffer from issues such as high rate of false positives and false negatives, and high comprehensive costs, leading to limitations in their application. Therefore, it is imperative to conduct research on high-speed railway perimeter intrusion alarm front-end monitoring technology that integrates multiple sensor technologies, specifically tailored to typical intrusion scenarios in high-speed railways.

In recent years, railway management authorities at home and abroad have been arranging train protection and alarm systems. The annual report published by the European Union Railway Agency in 2022 (European Union Agency for Railways, 2022) revealed that in railway systems across Germany, Spain, Italy, Luxembourg, the Netherlands, and Romania, more than 90% of the tracks are equipped with the highest-level train protection systems, including warning, automatic stopping, and train speed supervision. In Japan, the Shinkansen (bullet train) system has developed a computer-assisted traffic control system (Hachiga, 1993) to ensure the safety of train operations, which includes monitoring and alerting of foreign objects on the tracks, and stopping. China has also placed a high priority on railway operational safety and has made initial progress in establishing disaster monitoring systems and perimeter intrusion alarm systems (China National Railway Group Co., LTD, 2015). However, the effectiveness of perimeter intrusion monitoring sensor solutions still faces challenges under adverse weather conditions.

High-speed railway operation security is related to billions of people's safety of their lives and properties. Currently, most of the intrusion detection methods of railway lines rely on vision. However, due to the complexity of railway environment, the vision-based approach faces the following problems: firstly, the classes of targets are diverse, including not only people and animals, but also large and small obstacles. Secondly, the operation environment is complex, wind, rain, snow, fog and other bad weather conditions occur from time to time, resulting in the poor performance of the existing vision-based detection approach. Thirdly, vision-based detection approach is usually unable to accurately identify targets at low illumination and far distance. Therefore, we adopt the radar and vision fusion to improve the accuracy of target detection. In order to deal with the detection of small targets, we adopt the fusion strategy of radar and camera linkage to focus and amplify long-distance targets, and solve the problem of low illumination by filling laser light on the focus point. In order to improve the recognition effect, we use the YOLOv8 algorithm for multi-scale target recognition. In addition, for the image distortion caused by bad weather, we propose a linkage and tracking fusion strategy to output the correct alarm results. Simulated intrusion tests show that the proposed method can effectively detect human intrusion within 0–200 m during the day and night in sunny weather, and can achieve more than 80% recognition accuracy under extreme severe weather conditions.

The main contributions of this paper can be summarized in four aspects:

(1) Introduce a personnel intrusion monitoring method based on the fusion of millimeter-wave radar and cameras, enabling round-the-clock intrusion detection.

(2) Propose a multi-level fusion algorithm based on a linkage and tracking strategy to achieve intrusion target monitoring under adverse weather conditions.

(3) Conduct extensive on-site simulation experiments to validate the effectiveness of the methods proposed in this paper.

The remainder of this paper is organized as follows: In Section 2, we provide an introduction of the relevant background, including the conventional millimeter-wave radar processing procedures and an overview of target detection. In Section 3, we present the algorithms proposed in this paper. In Section 4, we describe the radar-camera linkage experiments and the training of the model. Section 5 provides a detailed description of the testing methods and includes a wide range of experiments. In Section 6, we conclude this paper and discuss future prospects.

2. Related work

Both domestic and international research institutions have conducted studies on various sensor fusion technologies regarding perimeter intrusion detection methods (Mockel, Scherer, & Schuster, 2003; Wang, Yu, Wu, Li, & Li, 2021; Garcia-Dominguez *et al.*, 2008). However, challenges persist, including low monitoring accuracy under adverse weather conditions and the inability to recognize small distant targets. Currently, for mobile target detection, a popular approach involves combining sensors such as cameras, millimeter-wave radar, and lidar. Cameras can precisely measure edges, colors, and brightness and can classify and locate objects on image plane. Lidar offers high positioning accuracy, while millimeter-wave radar can operate effectively under adverse weather conditions. In addition to the inherent characteristics of individual sensors, the method of fusing multiple sensors also significantly impacts the final detection performance. Previous research has designed various fusion strategies for cameras, millimeter-wave radar, and lidar. One fusion method (Han, Wang, Lu, & Zhao, 2017) combines lidar and cameras by projecting lidar point clouds onto images by way of cross-calibration, generating sparse depth images for classification and recognition. Another approach involves using image detection algorithms to create a series of detection boxes and then projecting a series of 3D pyramids generating by detection boxes into 3D point cloud space to facilitate recognition (Wang & Jia, 2019) Moreover, in scenarios with poor visibility or lighting conditions, Shuai etc. (Shuai *et al.*, 2021) explored fusion methods for millimeter-wave radar and cameras, significantly improving detection capabilities in low-light scenes. Recently, Chen, Zhang, Wang, Wang, and Zhao (2023) investigated a general fusion framework for millimeter-wave radar, lidar, and cameras in the field of advanced driving assistance system and achieved favorable target detection results on publicly available multi-source data fusion datasets. However, camera and radar fusion heavily rely on pre-training data and require a substantial amount of manually labeled intrusion data to train models effectively. Existing training datasets may not fully reflect real-world railway scenarios.

In recent years, there has been extensive research on fusion algorithms of millimeter-wave radar and cameras in the field of advanced driver assistance systems (ADAS). Based on the levels of fusion and substance, these algorithms can be categorized into three main types (Wei *et al.*, 2022): data-level fusion, feature-level fusion, and decision-level fusion. Decision-level fusion involves the independent recognition of targets by two different sensors, acquiring the information of target and subsequently fusing the detection results from different sensors. Feature-level fusion typically projects the radar features onto the corresponding 2D image plane, thereby further enhancing the recognition performance between modes. Data-level fusion initially generates regions of interest (ROIs) based on radar points and subsequently employs feature extractors and classifiers within ROIs to detect targets.

Zhang, Zhou, Qiu, Huang, and Li (2019) proposed an obstacle recognition algorithm that combines millimeter-wave radar and cameras. Initially, the algorithm utilizes millimeter-wave radar for target localization. Then, through the coordinate transformation of the camera, the ROI of the target in the image is obtained, followed by object recognition. Yao, Wang, and Qian (2021) employed a multi-layer perceptron model. The camera and radar

simultaneously detects targets, then the detection results of camera are associated with radar data. Lukas, Philipp, Jason, and Didier (2022) introduced a novel fusion point pruning method, which automatically finds the optimal fusion point for radar and image features within a neural network structure. Liu and Liu (2021) proposed a three-level fusion architecture that includes sensor data selection, sensor data association, and target trajectory prediction.

While the aforementioned algorithms have shown promising experimental results in the field of autonomous driving assistance system, the detection distance of railway scenarios is longer than that of automobiles, and it is necessary to study fusion detection algorithms for longer distances.

The fusion algorithm proposed in this paper combines millimeter-wave radar and cameras to address the problem of long-distance detection by leveraging the millimeter-wave radar to drive the dome camera. Furthermore, it achieves target recognition under low-light conditions by applying supplemental lighting within a small area around the target. Additionally, to tackle distortion of data captured by the camera in adverse weather conditions, this method utilizes a fusion evaluation strategy based on radar-camera linkage and radar tracking. This strategy effectively ensures precise target identification under adverse weather conditions.

3. Method design

In this section, we will primarily introduce the fusion method of millimeter-wave radar and cameras, as well as the overall structure of the alarm system.

3.1 Overall architecture

Railway perimeter intrusion protection requires all-weather and all-time monitoring of the railway perimeter, regardless of weather and adverse environmental factors. The wavelength of millimeter-wave radar reaches the millimeters level, while the length of rain, snow, and dust is shorter than the wavelength of millimeter-wave radar which makes it a perception sensor that remains unaffected by adverse weather conditions. However, millimeter-wave radar data is inherently sparse. On the other hand, cameras can provide the contours and colors of the scene and the target but exhibit low accuracy in adverse weather conditions. Therefore, the fusion of millimeter-wave radar and cameras realizes the complementary advantages of sensor detection. The fusion architecture of millimeter-wave radar and cameras is illustrated in Figure 1.

Radar and camera collect target data. Radar data includes target position, velocity, and spatial coordinates, while video information includes video stream video. Cameras decode the collected video streams. Subsequently, each sensor, relying on its unique sensor technology, generates distinct feature information for the detected targets. This information is then integrated in the same world coordinate system to identify whether the recognized objects by both radar and camera sensors are the same. Following this integration, demo cameras are

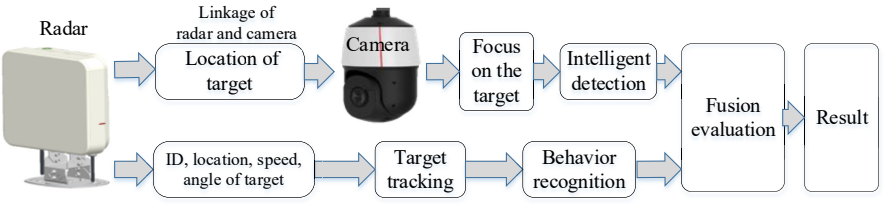


Figure 1.
Fusion architecture of millimeter-wave radar and cameras

Source(s): Authors own work

employed to focus, magnify, and supply illumination for target localization. Subsequently, intelligent recognition algorithms and fusion algorithms are applied to output the alarm results.

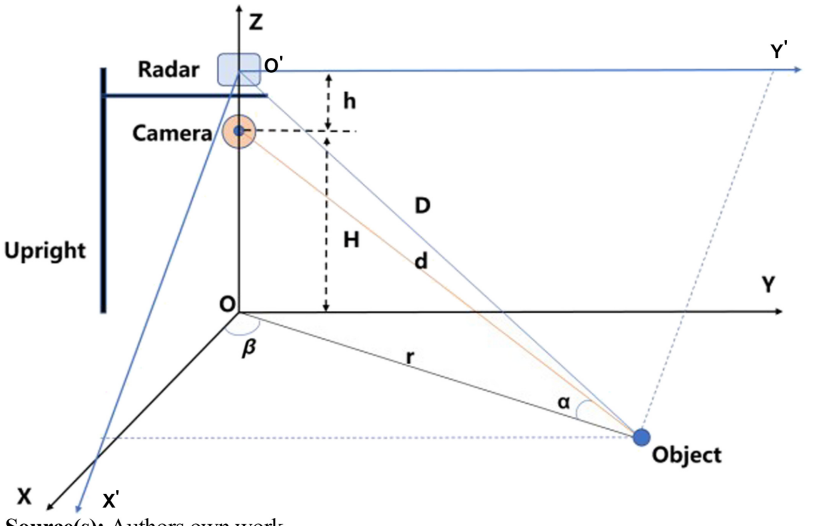
Regarding the fusion of millimeter-wave radar and cameras, firstly, data-level fusion is applied to combine radar data with camera video, enhancing both visual recognition and localization. Secondly, on the basis of improving video target recognition, cameras may still struggle to identify objects at night and under adverse weather conditions. In such scenarios, the paper utilizes target tracking algorithms of millimeter-wave radar to facilitate object recognition.

3.2 The radar-camera linkage method

The linkage detection method involving millimeter-wave radar and spherical cameras operates as follows: millimeter-wave radar obtains data such as distance, angle, and velocity of the target. Subsequently, the spherical camera system swiftly locates the target and the area where the target is located is magnified and focused, realizing the clear display of long-distance small targets.

In this context, the coordinate system of millimeter-wave radar is the polar coordinate system $X'O'Y'$, and the coordinate system of the spherical camera is the world coordinate system $OXYZ$. Initially, based on the target information acquired by the millimeter-wave radar, including the distance D and angle β, the coordinates of the target in the radar's polar coordinate system are transformed into the world coordinate system. This coordinate transformation is illustrated in Figure 2.

Here, the radar is located directly above the camera, and the center of the camera's field of view is at a height H relative to the ground. The distance between the center of the radar transmitting face and the center of the camera's field of view is h. To determine the camera's position, the target's distance d, horizontal angle β, and pitch angle α are required. Among these parameters, H, h, D, and β are known parameters before the camera linkage, while the target distance d and pitch angle α are the parameters to be solved.



Source(s): Authors own work

Figure 2.
The principle of radar and camera linkage

First, the projection r of the target distance in the horizontal plane is calculated by formula (1)

$$r = \sqrt{D^2 - (H + h)^2} \tag{1}$$

The target distance d of the spherical camera can be obtained using the coordinate transformation formula (2).

$$d = \sqrt{r^2 + H^2} \tag{2}$$

The pitch angle α in the spherical camera coordinate system can be calculated by formula (3):

$$\alpha = \arctan(H/r) \tag{3}$$

The premise of the above model assumes an ideal scenario where there is no angular deviation between the radar normal X' axis and the spherical camera's 0-degree X-axis. In practical installations, if there is an angular deviation between the radar normal X' axis and the spherical camera's 0-degree X-axis, then this situation needs to be considered in the coordinate transformation.

Suppose there is an angular difference in the horizontal direction between the radar normal X' axis and the spherical camera's X' axis and an angular difference in the pitch direction. In that case, the final target angle in the spherical camera coordinates would be formula (4) and (5):

$$\beta_c = \beta - \Delta\beta \tag{4}$$

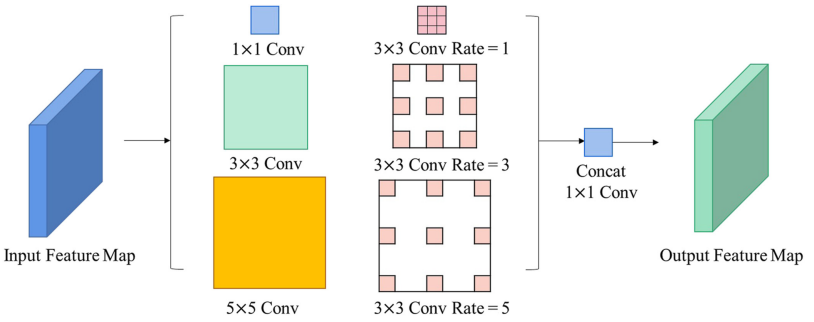
$$\alpha_c = \arctan(H/r) - \Delta\alpha \tag{5}$$

From this, the parameters required for target localization by the camera can be obtained, enabling the functionality of radar-camera linkage positioning.

3.3 Video intelligent recognition algorithm

For the video intelligent recognition method applied to targets after radar-camera linkage, this paper introduces an improved version of the YOLOv8 algorithm (Dillon, Jordan, Jacqueline, & Ahmad, 2023). This algorithm can directly extract features within the network to predict object categories and positions without the need for candidate bounding box extraction. Furthermore, this algorithm enhances recognition performance by adding receptive field block (RFB) modules according to the multi-scale imaging characteristics after linkage. These modules not only reduce information loss but also lower network parameters, thus improving the training and inference speed of the network. The structure diagram of the RFB module is depicted in Figure 3.

The RFB module is a multi-branch convolutional module, where each branch corresponds to a different receptive field. Features from different convolutions and outputs are cascaded in the channel dimension, followed by convolutional operations to integrate and adjust the number of channels, resulting in rich feature information. The smaller convolution kernels have a smaller receptive field, catering to the requirements of small-scale targets, thus enabling the network to detect small-sized objects more effectively, while the larger convolution kernels have a larger receptive field, meeting the receptive field requirements of larger-scale targets. This allows the network to learn features of the large and small targets simultaneously during the learning process. Thus, the feature information in the decoder's output is rich for both small and large targets. The detection branch can then regress the center points of targets of different scales, thereby enhancing the detection rate for targets of different scales.



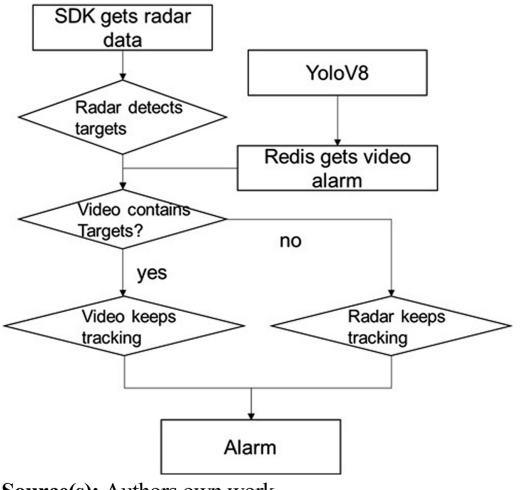
Source(s): Authors own work

Figure 3.
Structure of the RFB module

3.4 The target tracking and fusion algorithm

Based on the acquired radar target information and video data, this paper presents a linkage and tracking fusion decision strategy. This method is capable of making full use of radar information when visual perception is hindered, such as heavy fog. It achieves effective target detection while reducing false positives. The logical diagram of the algorithm is illustrated in Figure 4.

This algorithm begins by acquiring radar sensor data using the SDK. Simultaneously, it conducts video intelligent analysis. When the radar detects a target, the fusion algorithm retrieves video processing data and judges whether the video intelligent analysis has identified the target. Upon detection of the target by the video intelligent analysis, the target information is transmitted to the fusion algorithm through the Redis component. The fusion algorithm, relying on the target information, continuously tracks and judges the target. When the target can be continuously tracked, the algorithm will recognize that there is a target intrusion and give an alarm. Due to the sensitivity of visible light cameras to adverse weather conditions, such as rain and fog, clear video images cannot be obtained in such weather conditions. This limitation results in the inability of video recognition under rainy or foggy



Source(s): Authors own work

Figure 4.
The target tracking and fusion algorithm

weather conditions. Therefore, when the radar can continuously track a target and behavior recognition based on the tracked target indicates human intrusion, the algorithm outputs an alarm result.

In order to solve the problem that millimeter-wave radar is easy to produce false alarms for plant shaking, the feature processing module is added to the fusion strategy to remove false alarms. Firstly, the interference information generated by plant shaking and the information of intruders walking along the fence were collected, and their characteristics were analyzed. Secondly, a new defense zone was established in the vegetation disturbance area, and the standard deviation of the moving distance of the alarm point in the world coordinate system was calculated in this area, and the human invasion and vegetation interference were distinguished according to the different standard deviations.

4. Test and train

In this section, we conducted radar-camera linkage experiments and model training in the actual scenario. The test was conducted at the testing grounds of the East Sub-Institute of the China Academy of Railway Sciences, which has a total length of 212 m. The millimeter-wave radar and video monitoring equipment were installed on a pillar at one end of the testing grounds. The pillar extended upwards with a crossbar, and the equipment was installed as a whole using brackets and clamps on the crossbar, as depicted in Figure 5. Additionally, a rain simulation system and a fog simulation system were installed above to simulate rain and fog conditions in the testing area of the millimeter-wave radar equipment.

The protected area for the video + millimeter-wave radar covers a strip-shaped region with a length of approximately 200 m and a width of about 10 m. It encompasses the double-track railway on both the left and right sides. The near-end boundary starts at the beginning of the railway tracks, with the left boundary along the edge of the track bed and the right boundary marked by a protective fence, as illustrated in Figure 6.

4.1 Radar-camera linkage calibration experiment

Firstly, the camera's installation height H was obtained by the measuring tape as 2.8 m, and the distance h between the millimeter-wave radar and the camera was measured as 0.16 m. Secondly, as shown in Figure 7, as personnel moved within the protected area, the millimeter-wave radar was able to determine the position of a moving target as 74.57 m, with the azimuth angle of the millimeter-wave radar measured as −4.72°. The horizontal and pitch angles of the

Source(s): Authors own work

Figure 5.
Installation diagram for video + millimeter-wave radar

moving ball camera were determined to be −8.74° and −7.67°, respectively. Using the formula from Section 3.2, the deviation angles α_c and β_c were calculated as 3.5° and 2.7°, respectively.

4.2 Video intelligent recognition training

This study employed 4585 real railway scene images as training data, with 3668 images used for training and 917 images for testing. The training experiments were conducted on an RTX3090Ti GPU, utilizing Pytorch as the deep learning framework and the Adam optimizer for training. The batch size was set to 12, the number of training epochs was set to 30, and the initial learning rate was set to 0.0001. The loss function used was the same as the original model. The training results are depicted in Figure 8.

5. Real-world scenario testing

In this section, we conducted numerous experiments to demonstrate the performance and advantages of the proposed algorithm in this paper. The tests covered anti-interference performance to typical interference factors during both daytime and nighttime periods, performance in recognizing typical intrusion behaviors, and intrusion behavior recognition performance under the influence of various interference factors.

The typical interference tests for video + millimeter-wave radar include four categories: rain interference, fog interference, light and shadow interference, and vegetation interference, with a total of 166 tests conducted, as shown in Table 1.

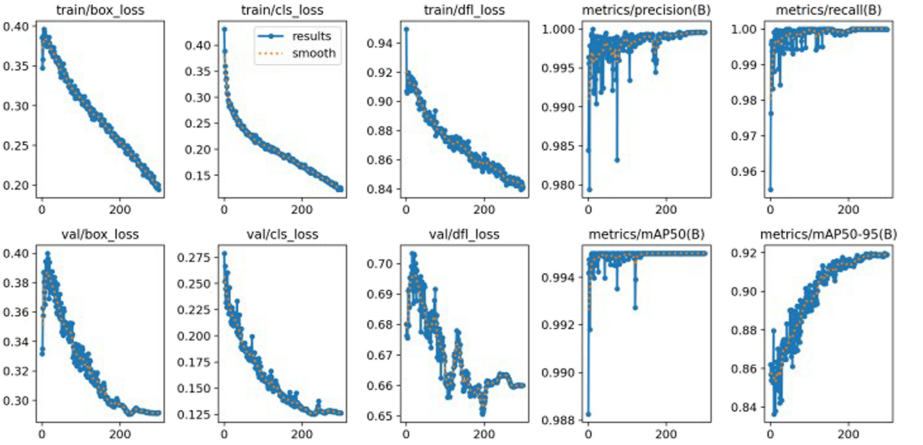
The intrusion tests included two categories: individuals crossing the railway tracks and walking along the inner edge of the fence, with a total of 83 tests conducted. Additionally, intrusion tests under rain, fog, light and shadow interference, and vegetation interference conditions were conducted a total of 392 times. The details of the test are shown in Table 2.

Source(s): Authors own work

Figure 6.
Layout of the protected area

Source(s): Authors own work

Figure 7.
Radar-camera linkage calibration experiment



Source(s): Authors own work

Figure 8.
YOLOv8 training results

Interference factors	Times
Rain	10
Fog	35
Vegetation	17
Light and Shadow	10
Light and Shadow + Vegetation	12
Fog + Vegetation	30
Rain + Vegetation	20
Rain + Light and Shadow	10
Rain + Light and Shadow + Vegetation	12
Fog + Light and Shadow + Vegetation	10
Total	166

Source(s): Authors own work

Table 1.
Test times of video + millimeter-wave radar interference

Through the interference tests, it was observed that under rain, fog, light and shadow, and vegetation interference conditions, a total of 30 false alarms were generated. Among these, vegetation interference resulted in 18 false alarms, primarily due to the presence of abundant vegetation in the test area. Heavy rain or torrential rain led to 6 false alarms, while extremely thick fog also resulted in 6 false alarms. Light and shadow interference did not produce any false alarms.

The intrusion tests show that under sunny weather, except for occasional false negatives caused by the obstruction of tower cranes and small vehicles within the protected area, the proposed method achieved zero false negatives. Under light rain, moderate rain, and heavy rain conditions, there were a few false negatives, and as the intensity of rainfall increased, the number of false negatives gradually rose, with an overall false negative rate of less than 20%. Under fog, heavy fog, and extremely heavy fog conditions, there were even fewer false negatives compared to rainfall, with an overall false negative rate of less than 15%.

The results of false negatives were summarized and it shows that when both interference and intrusion occurred simultaneously, there will be a false positive before the positive report, and this test method categorized such cases as false negative. Another reason for false

Intrusion behavior	Distance	Sunny weather	Rainy weather	Fog	Light and shadow interference	Vegetation interference	Total
Crossing the Track Laterally	50m	10	/	/	/	11	21
Crossing the Track Laterally	100m	11	/	/	13	10	34
Crossing the Track Laterally	150m	10	32	25	10	10	87
Crossing the Track Laterally	200m	10	30	32	10	10	92
Walking Along the Fence	50m	12	/	/		10	22
Walking Along the Fence	100m	10	/	/	11	10	31
Walking Along the Fence	150m	10	31	34	10	10	95
Walking Along the Fence	200m	10	30	31	12	10	93
Total		83	123	122	66	81	475

Source(s): Authors own work

Table 2. Test times for video + millimeter-wave radar intrusion

negatives was the randomness of simulated intruder activities. When intruders had limited movement within the protected area and quickly exited the area, it resulted in false negatives. This issue can be addressed by adjusting tracking threshold based on the specific on-site conditions.

Furthermore, under extreme heavy rainfall conditions, the rain curtain has an attenuating effect on the millimeter-wave radar signal, and the video cannot capture clear images, leading to the inability to recognize distant crossings. Fog and lighting conditions have a relatively small impact on the recognition of video + millimeter-wave radar. The results of the simulated intrusion test are shown in Figure 9.

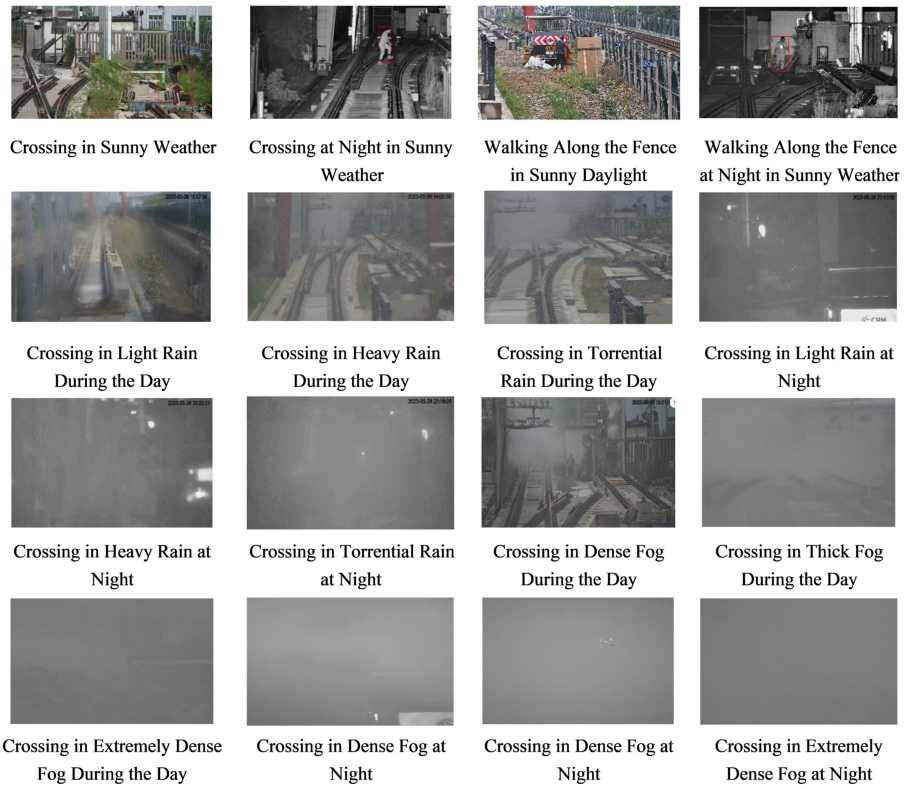
In addition, test was conducted on an actual operational line, and the results were obtained by monitoring the route for one month. A total of 259 positive reports were generated, with zero false negatives and zero false positives. The positive reports were triggered by the presence of patrol personnel and maintenance operations on the route. The alarm images are shown in Figure 10.

6. Conclusion

This paper presents a personnel intrusion detection method for railway perimeter security. The method proposes a combination of video and millimeter-wave radar equipment and designs a linkage and tracking fusion strategy based on this scheme. Deployment and testing experiments were conducted, and the results show:

(1) The video + millimeter-wave radar technology solution can realize personnel intrusion detection in the railway perimeter domain, with an effective monitoring range of up to 200 m.

(2) Tests under extreme adverse weather conditions revealed that the video + millimeter-wave radar equipment and algorithm are sensitive to heavy rainfall but less affected by fog.



Source(s): Authors own work

Figure 9.
Simulated intrusion effect image

Source(s): Authors own work

Figure 10.
Positive line intrusion test image

(3) Tests on the main railway line demonstrated that the proposed algorithm meets the requirements for on-site applications.

Future work will focus on three aspects: the first is to continue to develop feature-level fusion algorithms to improve algorithm performance; the second is to conduct research on interference filtering for adverse environments and vegetation; and the third is to conduct tests for different types of risk points to meet the requirements of personnel intrusion detection in railway scenarios.

Reference

China National Railway Group Co., LTD. (2015). High speed railway perimeter intrusion alarm system general technical scheme (interim) (TJ/QT003-2015).

Chen, X., Zhang, T., Wang, Y., Wang, Y., & Zhao, H. (2023). FUTR3D: A unified sensor fusion framework for 3D detection. In *2023 IEEE/CVf conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 172–181). Vancouver, BC: IEEE.

Dillon, R., Jordan, K., Jacqueline, H., & Ahmad, D. (2023). Real-time flying object detection with YOLOv8. doi: 10.48550/arXiv.2305.09972.

European Union Agency for Railways (2022). Retrieved from Report on railway safety and interoperability in the EU. Available from: https://www.era.europa.eu/library/corporate-publications_en

Garcia-Dominguez, J., Urena-Urena, J., Hernandez-Alonso, A., Mazo-Quintas, M., Vazquez, J., & Jesus Diaz, M. (2008), Multi-sensory system for obstacle detection on railways. In *2008 IEEE instrumentation and measurement technology conference*, (pp. 2091–2096), Victoria, BC: IEEE.

Hachiga, A. (1993). The concepts and technologies of dependable and real-time computer systems for Shinkansen train control. In *Dependable Computing and Fault-Tolerant Systems* (pp. 225–252). Vienna: Springer.

Han, X., Wang, H., Lu, J., & Zhao, C. (2017). Road detection based on the fusion of Lidar and image data. *International Journal of Advanced Robotic Systems*, *14*(6), 172988141773810. doi: 10.1177/1729881417738102.

Liu, Y., & Liu, Y. (2021). A data fusion model for Millimeter-Wave radar and vision sensor in advanced driving assistance system. *International Journal of Automotive Technology*, *22*(6), 1695–1709. doi: 10.1007/s12239-021-0146-8.

Lukas, S., Philipp, H., Jason, R., & Didier, S. (2022). Fusion point pruning for optimized 2D object detection with radar-camera fusion. In *2022 IEEE/CVF winter conference on applications of computer vision (WACV)* (pp. 1275–1282). Waikoloa, HI: IEEE.

Mockel, S., Scherer, F., & Schuster, P. (2003). Multi-sensor obstacle detection on railway tracks. In *IEEE IV2003 intelligent vehicles symposium. Proceedings (Cat. No.03TH8683)* (pp. 42–46). IEEE.

Measures for the Administration of High-speed Railway Safety Protection (2020). Beijing: Ministry of Transport of the People's Republic of China.

Shuai, X., Shen, Y., Tang, Y., Shi, S., Ji, L., & Xing, G. (2021). Millieye: A lightweight mmwave radar and camera fusion system for robust object detection. In *Proceedings of the international conference on internet-of-things design and implementation* (pp. 145–157). Charlottesvle, VA: Association for Computing Machinery.

Wang, Z., & Jia, K. (2019). Frustum ConvNet: Sliding frustums to aggregate local point-wise features for amodal 3D object detection. In *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 1742–1749). Macau.

Wang, Z., Yu, G., Wu, X., Li, H., & Li, D. (2021). A camera and LiDAR data fusion method for railway object detection. *IEEE Sensors Journal*, *21*(12), 13442–13454. doi: 10.1109/jsen.2021.3066714.

Wei, Z., Zhang, F., Chang, S., Liu, Y., Wu, H., & Feng, Z. (2022). MMWave radar and vision fusion for object detection in autonomous driving: A review. *Sensors*, *22*(7), 2542. doi: 10.3390/s22072542.

Yao, T., Wang, C., & Qian, Y. (2021). Camera-radar fusion sensing system based on multi-layer perceptron. *Journal of Shanghai Jiaotong University (Science)*, *26*(5), 561–568. doi: 10.1007/s12204-021-2345-x.

Zhang, X., Zhou, M., Qiu, P., Huang, Y., & Li, J. (2019). Radar and vision fusion for the real-time obstacle detection and identification. *Industrial Robot*, *46*(3), 391–395. doi: 10.1108/ir-06-2018-0113.

Further reading

Holger, C., Varun, B., Alex, H. L., Sourabh, V., Venice Erin, L., Xu, Q., & Beijbom, O. (2020). nuScenes: A multimodal dataset for autonomous driving. In *2020 IEEE/CVF Conference on computer vision and pattern recognition (CVPR)* (pp. 11618–11628). Seattle, WA: IEEE.

Nabati, R., & Qi, H. (2021). CenterFusion: Center-based radar and camera fusion for 3D object detection. In *2021 IEEE Winter conference on applications of computer vision (WACV)* (pp. 1526–1535). Waikoloa, HI: IEEE.

Corresponding author
Pengyue Guo can be contacted at: 878768842@qq.com