



Article

Algorithm Improvement for Mobile Event Detection with Intelligent Tunnel Robots

Li Wan ¹, Zhenjiang Li ¹, Changan Zhang ¹, Guangyong Chen ¹, Panming Zhao ^{2,*} and Kewei Wu ²

¹ Shandong Provincial Communications Planning and Design Institute Group Co., Ltd., Jinan 250101, China; 13583121175@126.com (L.W.); jnlzj@163.com (Z.L.); sdjxzca@163.com (C.Z.); 51338031@sohu.com (G.C.)

² Beijing Zhuoshi Zhitong Technology Co., Ltd., Beijing 100080, China; 13810497654@139.com

* Correspondence: zhaopanming_2023@163.com; Tel.: +86-150-5189-2101

Abstract: Mobile inspections conducted by intelligent tunnel robots are instrumental in broadening the inspection reach, economizing on inspection expenditures, and augmenting the operational efficiency of inspections. Despite differences from fixed surveillance, mobile-captured traffic videos have complex backgrounds and device conditions that interfere with accurate traffic event identification, warranting more research. This paper proposes an improved algorithm based on YOLOv9 and DeepSORT for intelligent event detection in an edge computing mobile device using an intelligent tunnel robot. The enhancements comprise the integration of the Temporal Shift Module to boost temporal feature recognition and the establishment of logical rules for identifying diverse traffic incidents in mobile video imagery. Experimental results show that our fused algorithm achieves a 93.25% accuracy rate, an improvement of 1.75% over the baseline. The algorithm is also applicable to inspection vehicles, drones, and autonomous vehicles, effectively enhancing the detection of traffic events and improving traffic safety.

Keywords: YOLOv9+DeepSORT; edge computing; mobile event detection; smart tunnels



Citation: Wan, L.; Li, Z.; Zhang, C.; Chen, G.; Zhao, P.; Wu, K. Algorithm Improvement for Mobile Event Detection with Intelligent Tunnel Robots. *Big Data Cogn. Comput.* **2024**, *8*, 147. <https://doi.org/10.3390/bdcc8110147>

Academic Editor: Moulay A. Akhloufi

Received: 30 July 2024

Revised: 9 September 2024

Accepted: 24 September 2024

Published: 28 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

China leads globally in the number and scale of highway tunnels, with 21,316 in total, featuring 1394 extra-long tunnels over 10 km and 5541 long tunnels from 3 to 10 km. Tunnels have higher accident rates due to their enclosed, narrow design and lack of emergency lanes. Specifically, the enclosed structure of tunnels results in significant internal humidity and poor ventilation for exhaust fumes from vehicles. The leakage of coolants and lubricating oils from vehicles, which mix and accumulate on the road surface, further contribute to a reduction in the road's coefficient of friction within the tunnel. This, combined with the suboptimal braking performance of freight vehicles, makes rear-end collisions and other traffic accidents more likely to occur inside tunnels. Moreover, the only points connecting the tunnel interior to the outside are the entrance and exit at each end. If a traffic accident occurs, it can impact the entire stretch of the road. In the event of an accident, if the detection of incident information within the tunnel is not prompt, the risk of secondary accidents increases, which can lead to severe consequences.

Current manual inspection and reactive management lead to issues like poor safety awareness, delayed repairs, inefficient rescues, and lack of real-time traffic monitoring [1–3]. Traditional tunnel monitoring methods primarily consist of three approaches: inspection vehicles, fixed detectors, and manual inspections. Inspection vehicles are associated with high costs, fixed detectors suffer from incomplete coverage, and manual inspections are characterized by low efficiency and a propensity for oversight. This struggles to keep up with increasing operational demands. Thus, integrating AI into mobile inspection robots is crucial to smarten up tunnel inspections effectively.

Additionally, common deep learning techniques like CNNs, R-CNN, Fast R-CNN, Faster R-CNN, and YOLO effectively extract image features [4–7]. In [8], the vehicle detection and tracking model employs the YOLOv5 object detector in conjunction with the DeepSORT tracker to identify and monitor vehicle movements, assigning a unique identification number to each vehicle for tracking purposes. In [9], YOLOv8 has been enhanced for highway anomaly detection using the CBAM attention mechanism, MobileNetV3 architecture, and Focal-EIoU Loss.

Target recognition and tracking algorithms for tunnel robots, essential for traffic safety and incident inspection, have been researched and implemented globally in recent years [10–12]. Among them, the method based on deep learning is more advantageous [13,14]; Papageorgiou C. P. et al. proposed a general trainable framework suitable for complex scenarios, in which the system performs feature learning from examples, and does not rely on prior (artificial labeling) models [15]. Yang B et al. proposed an improved convolutional neural network based on Faster RCNN [16], which enhances the detection performance of small vehicles by fine-tuning the model using vehicle samples specifically for vehicle detection. Lee W. J. et al. developed a selective multi-stage feature detection framework based on a convolutional neural network, which fully extracts the feature information of vehicle images and exhibits strong robustness to noise [17]. Li L. et al. considered the underbody shadow as the candidate area and designed edge enhancement and adaptive feature segmentation algorithms for vehicle detection, which effectively distinguish between underbody shadow and non-underbody shadow interferences, thereby improving the accuracy and reliability of vehicle detection [18,19].

In summary, the current application of tunnel robots primarily focuses on firefighting, rescue operations, and crack detection [20–22]. However, there are limitations in event detection for tunnel robots [23], as well as insufficient consideration of event characteristics from the perspective of mobile devices. The video background and target pixel position change during robot movement, increasing the computational complexity of the detection process, and if the fixed detection equipment algorithms are still employed [24,25], it would be challenging to ensure the accuracy and real-time nature of event analysis [26–28]. This research aims to enable smart event detection in mobile devices by integrating target tracking and temporal feature extraction. The YOLOv9+DeepSORT algorithm is used for this purpose, along with revised abnormal event detection rules tailored for mobile scenarios. The aim is to improve real-time detection and response to events like tunnel fires, accidents, and pedestrian incursions, thus boosting the intelligence of smart tunnel systems.

2. Materials and Methods

The system's workflow includes the following steps for mobile sensing: high-definition cameras, infrared cameras, and gas sensors mounted on tunnel inspection robots transmit data to the robots' edge computing devices. Algorithms deployed on these devices use AI to identify traffic events and send the event information to an AI-based event warning platform. For fixed surveillance equipment, fixed edge computing devices recognize traffic events in the video feeds and relay the output to the AI event warning platform. The AI event warning platform then assesses the events and responds by either using the tunnel robot to issue on-site announcements or dispatching rescue personnel for on-site handling. As shown in Figure 1.

2.1. Intelligent Tunnel Robot

The intelligent tunnel robot has two main parts: the robot itself and the carrying system. The carrying system includes cameras, sensors, and an AI edge computing device. The robot has a 5G antenna, wheels, radar, and other features. Its main jobs are smart monitoring, inspection, control, and emergency communication, as shown in Figure 2. Edge Computing specifications: the processor used is an ARM Cortex-A53 8-core, with an AI computing power of 17.6 TOPS INT8. It is equipped with 12G DDR4X memory and has a power consumption within 20 W.

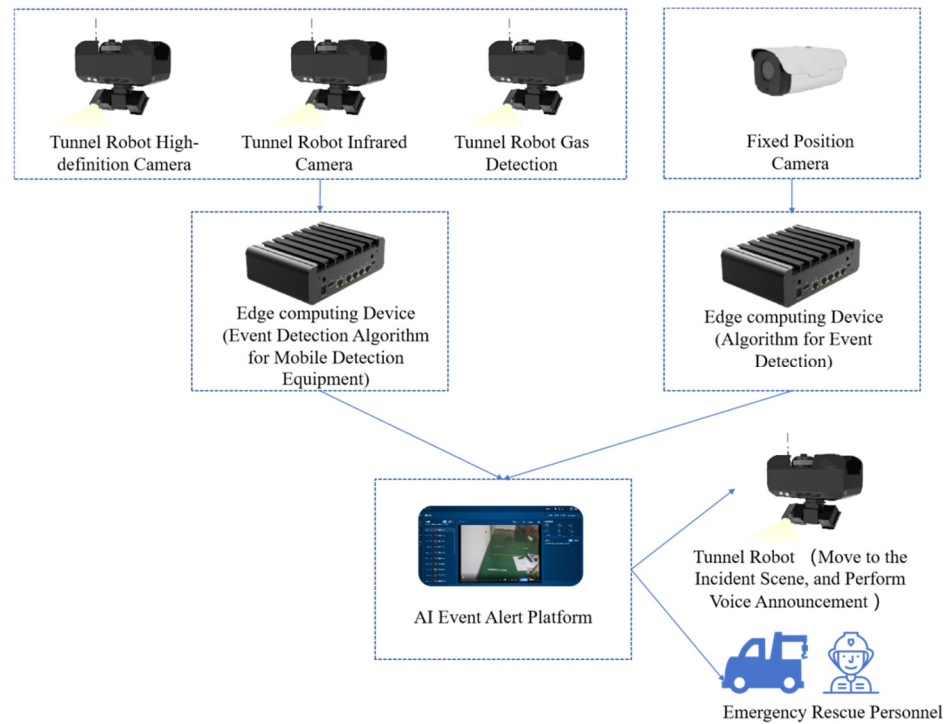


Figure 1. Methodological workflow of the proposed system.

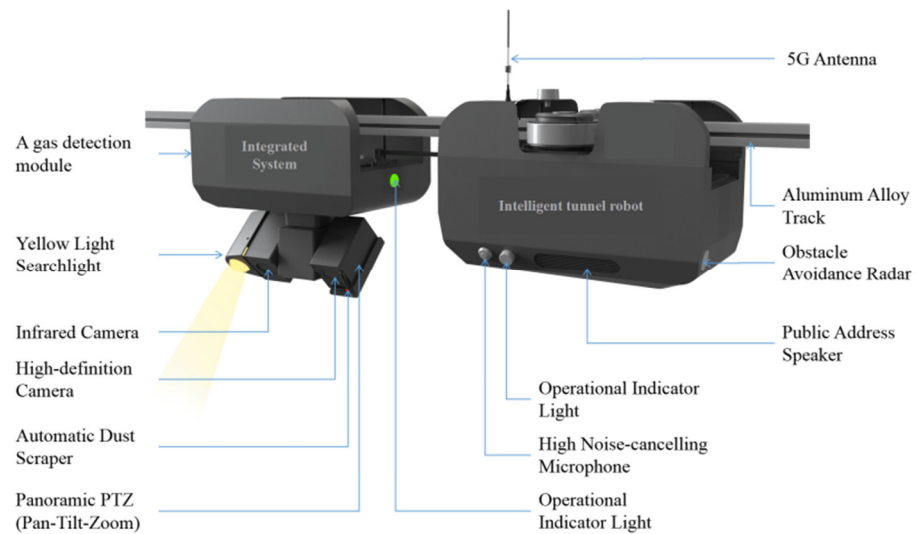


Figure 2. Intelligent tunnel robot module composition.

Intelligent monitoring uses AI for real-time detection and surveillance of issues like abnormal parking, pedestrians, and flames. It gathers data on brightness and environmental conditions, including temperature, humidity, gases, and smoke levels. The system provides 360-degree coverage to eliminate blind spots and improve awareness of the tunnel’s status.

Intelligent inspection offers daily and emergency modes. In daily mode, the robot autonomously patrols at 1–1.5 m/s, identifying issues like abnormal vehicles and unauthorized entries. In emergency mode, it can quickly reach incidents by speeding up to 7 m/s. It uses a light projector to identify and alert fire and deploys a sound and light warning system for effective communication, which aids in preventing additional accidents.

2.2. Designing Algorithms

To achieve high accuracy and efficiency in traffic event detection for mobile device states, considering the limited computing power of edge devices, a lightweight model is required. This paper employs YOLOv9 for object detection in extracted video frames. Detected objects undergo convolution, self-attention mechanisms, and temporal sequence movement to enhance the extraction of channel, spatial, and temporal features, resulting in feature maps. DeepSORT is then used for tracking targets in the feature maps, employing a Kalman filter prediction model and association metrics that use Mahalanobis distance and a Deep Appearance Descriptor with Hungarian Assignment to track trajectories of different objects. Logical rules are applied to determine the occurrence of traffic events, including the identification of four types of traffic events: abnormal parking, pedestrians, wrong-way driving, and flames. The event detection results are then outputted. Comprehensive algorithm framework diagram is shown in Figure 3.

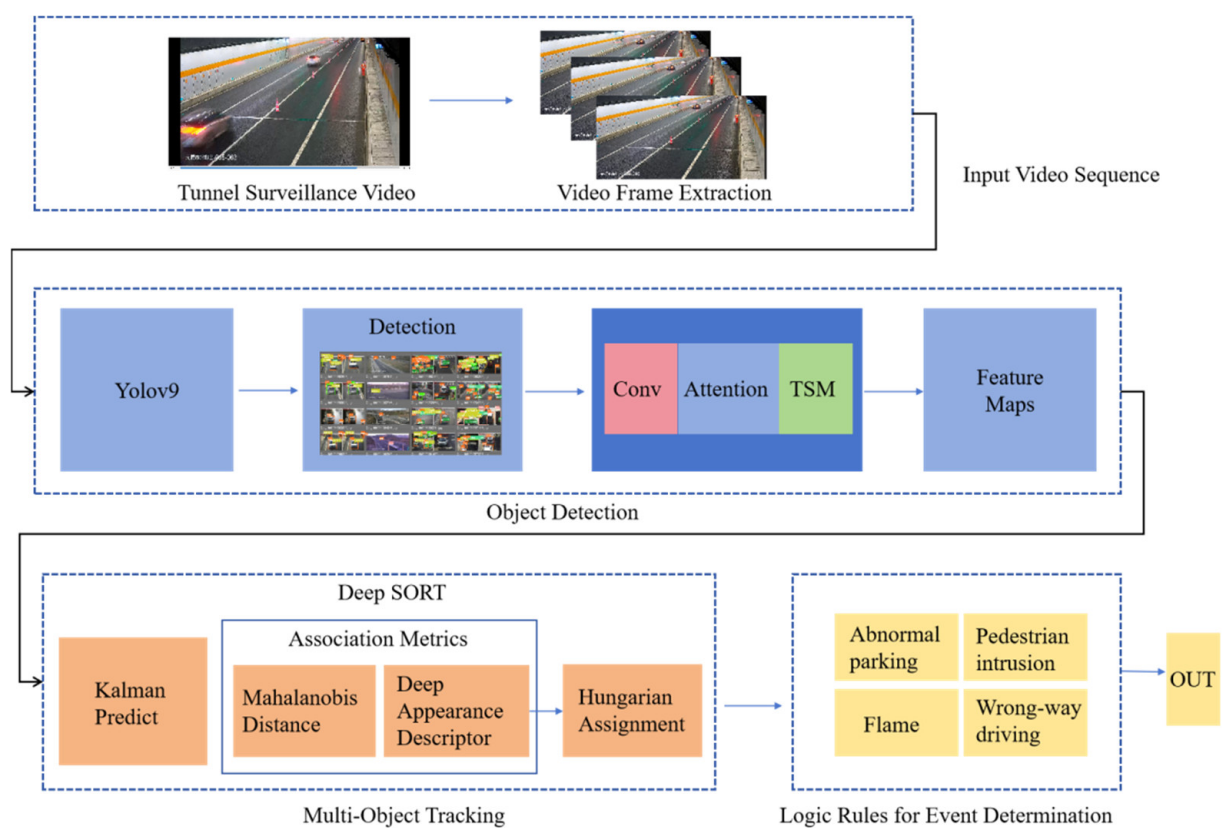


Figure 3. Comprehensive algorithm framework diagram.

2.2.1. Object Detection Algorithms

Traditional deep learning methods often suffer from significant information loss during the feature extraction and spatial transformation processes as the input data propagates through the layers of deep networks. To address the issue of information loss during data transmission, YOLOv9 proposes two solutions: (1) It introduces the concept of Programmable Gradient Information (PGI) to cope with the various transformations required for a deep network to achieve multiple objectives. (2) It designs a novel lightweight network structure based on gradient path planning—Generalized Efficient Layer Aggregation Network (GELAN). Validation on the MS COCO dataset for object detection demonstrates that GELAN can achieve better parameter utilization using only conventional convolutional operators, and PGI can be applied to models of various sizes, from lightweight to large-scale. Moreover, it can be used to obtain comprehensive information, enabling models

trained from scratch to achieve better results than state-of-the-art models pre-trained on large datasets [29].

2.2.2. Temporal Shift Module

This paper further enhances the algorithm’s sensitivity to timing features in traffic videos by introducing weighted time-shifted operations on video frame features. By partially shifting the feature map along the time dimension, information exchange and fusion between adjacent video frames are facilitated. This improves the expression capacity of features in capturing temporal relationships within the time series. As shown in Figure 4, the cells of different colors represent the features of different video frames, and the cells of the same color represent the features of the channels of a certain video frame. The features of the video frame are divided into 4 parts (c_1, c_2, c_3, c_4) in the channel dimension. The features of each channel c_1 are then shifted one time block backward along the time dimension, while keeping the characteristics of the c_2, c_3 channels unchanged. Additionally, the features of the c_4 channel are shifted one time block forward along the time dimension, and the weights (w_1, w_2, w_3, w_4) are assigned to the movement of each channel. The relationship between the time-shifted feature $c'_i (i = 1, 2, 3, 4)$ and the initial feature $c_i (i = 1, 2, 3, 4)$ is expressed as shown in the formula. Among them, T^{-1}, T^0, T^{+1} are the previous moment, the current moment, and the next moment, respectively. The formula is as follows:

$$\begin{aligned} c'_1 &= w_1 c_1^{T^{-1}}, c'_2 = w_2 c_2^{T^0}, \\ c'_3 &= w_3 c_3^{T^0}, c'_4 = w_4 c_4^{T^{+1}} \end{aligned} \tag{1}$$

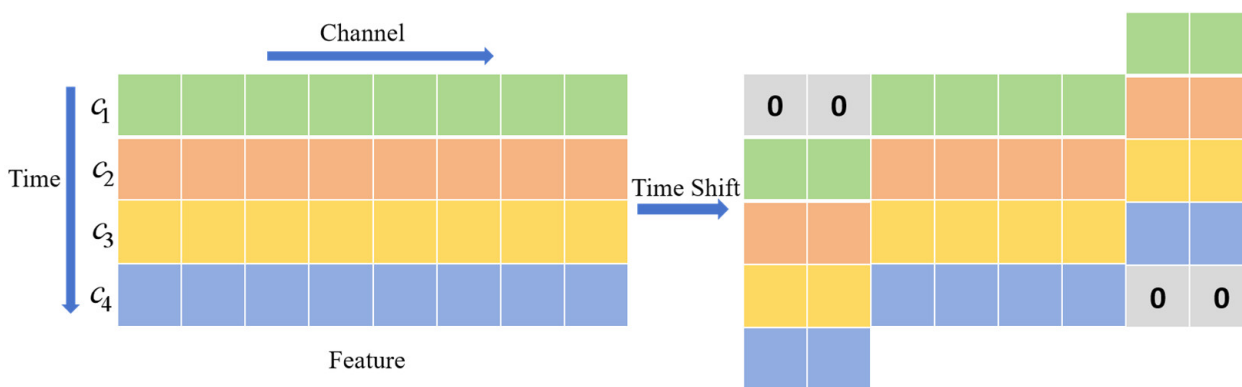


Figure 4. Temporal shift operation of features between video frames.

2.2.3. DeepSORT Multi-Object Tracking Algorithm

The DeepSORT algorithm builds upon the SORT algorithm by introducing a Matching Cascade and confirmation of new trajectories. It integrates the Kalman filter to estimate the mean and covariance of the track. By utilizing a gate matrix, it constrains the cost matrix to limit the distance between the state distribution of the Kalman filter and the measured value, thereby reducing matching errors. Additionally, the algorithm incorporates the image depth feature within the detection frame range of the YOLOv9 target detection algorithm. To extract features, it leverages the ReID domain model, reducing the occurrence of tracking target ID switches and enhancing target matching accuracy.

It is evident that the overall process unfolds as follows:

- (1) Use the Kalman filter to predict where objects will move;
- (2) Apply the Hungarian algorithm to match these predictions with actual objects seen in the current frame, leverage Mahalanobis distance for measuring the similarity between detected and tracked objects;
- (3) Utilize the Deep Appearance Descriptor for the extraction of target appearance features, re-identification of targets, and association of targets, thereby optimizing the matching performance.

2.3. Logic Rules for Event Determination in Mobile Device

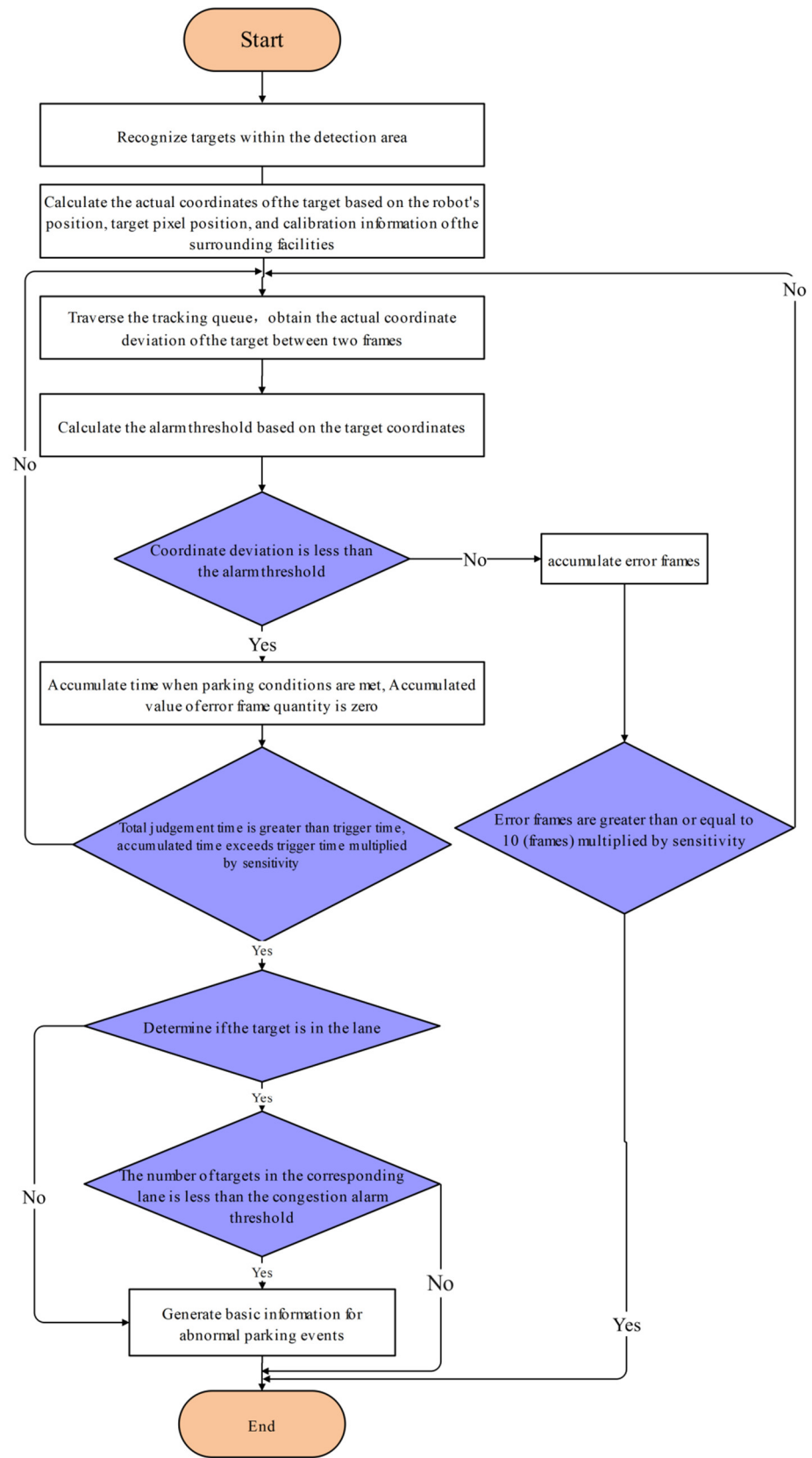
This paper focuses on the development of traffic event detection logic for four specific scenarios within tunnel environments, utilizing the unique characteristics of mobile device surveillance footage. The scenarios addressed are Abnormal parking, Pedestrian intrusion, Wrong-way driving, and Flame detection.

The detection process is primarily object recognition-based, aimed at mitigating the impact of the dynamic and complex environment encountered during the movement of the monitoring devices. Additionally, the paper discusses the conversion of pixel coordinates within the video to real-world coordinates, while also accounting for the movement of the device itself. This is crucial to avoid misclassifications, such as incorrectly identifying stationary vehicles as wrong-way drivers due to relative motion between the vehicle and the moving device.

The standard duration for event detection is set at 15 s. For flame detection, to enhance the reliability of the recognition, a threshold for identifying high-temperature areas is proposed, with the stipulation that the area must be at least 0.4 m by 0.4 m.

Here are the definitions for the listed traffic events: (1) Abnormal parking: the event where a vehicle comes to a stop within a tunnel and remains stationary for more than 15 s is classified as abnormal parking in the tunnel. (2) Pedestrian intrusion: the detection of a pedestrian within a tunnel, with the presence of the pedestrian being continuously detected for over 15 s, is classified as pedestrian intrusion into the tunnel. (3) Wrong-way driving: the event where a vehicle's trajectory within a tunnel is in the opposite direction of the normal driving direction, and this condition persists for 15 s, is classified as wrong-way driving in the tunnel. (4) Flame detection: within a tunnel, the detection of an object's surface temperature exceeding a threshold for a continuous period of 15 s is classified as a fire in the tunnel (this event is identified by the infrared thermal imaging camera of the tunnel robot).

The detailed process for event determination is outlined below, using abnormal parking as an illustrative example: the position coordinates of events in different video frame states are deduced from the real-time position of camera motion and pixel coordinates in camera lenses. Furthermore, whether a vehicle is moving or stationary within a certain period of time is determined by calculating the distance between positions in consecutive N frames (representing a specific time interval), as shown in Figure 5a. As for flame detection, considering the high accuracy of infrared cameras in detecting object temperatures, it integrates the results of infrared cameras and AI algorithms on the basis of visible light cameras. By fusing detection data in two ways and calibrating each of them, the overall accuracy of fire detection is improved, as shown in Figure 5b.



(a)

Figure 5. Cont.

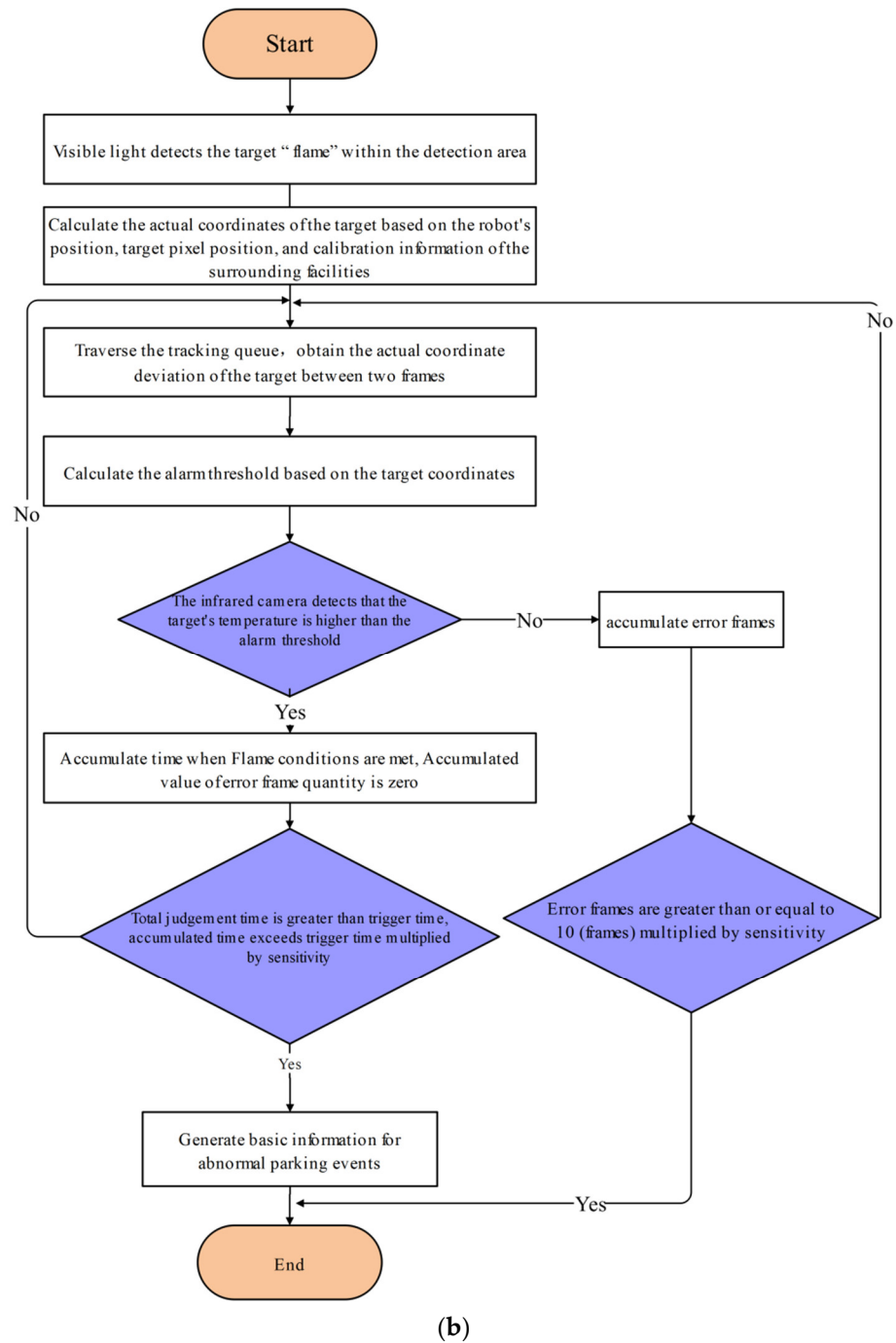


Figure 5. (a) Flowchart of rules for abnormal parking events with mobile cameras. (b) Flowchart of rules for flame events with mobile cameras.

3. Results

3.1. Datasets

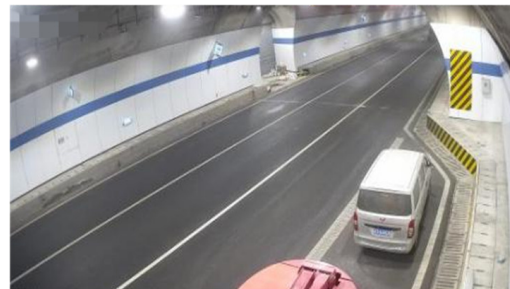
This algorithm uses a proprietary training set consisting of 210,000 images, which includes: 55,000 images for abnormal parking detection, 50,000 images for wrong-way driving detection, 50,000 images for pedestrian detection, and 55,000 images for flame detection.

To validate the experiment’s reliability, the intelligent tunnel robot’s mobile state algorithm underwent testing in two environments: simulated tunnels and mobile video capture. With a priority on replicating real tunnel conditions, the majority of tests were conducted in the simulated tunnel setup, with each event class-tested 80 times. In contrast, mobile video testing was performed 20 times per event class. The test event pictures

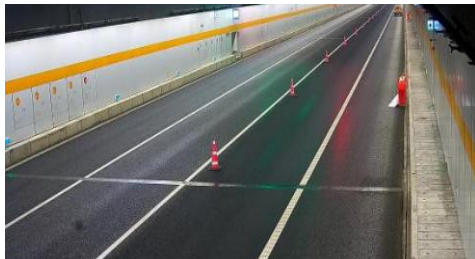
are shown in Figure 6a; the simulated environment and real environment are shown in Figure 6b.



Wrong-way driving



Abnormal parking



Pedestrians



Flames

(a)



Simulated environment (indoor simulated tunnel environment, inspection robot track installation height: 5 m)



Real tunnel environment (newly constructed tunnel environment, inspection robot track installation height: 6 m)

(b)

Figure 6. (a) Different event recognition image effects. (b) Simulated environment and real tunnel environment.

In the paper, the simulation of the tunnel scenario is primarily achieved by deploying a tunnel robot to patrol along a track indoors at the same height and angle, simulating the moving video angle of the robot in the tunnel. At the same time, by simulating the occurrence of events such as reverse wrong-way driving, abnormal parking, flames, and pedestrians, the real traffic events in the tunnel are simulated.

3.2. Ablation Study

This paper conducts ablation studies on YOLOv9, YOLOv9+SORT, YOLOv9+DeepSORT, and YOLOv9+DeepSORT+TSM. The ablation study comprises the following key points:

A preliminary evaluation of the YOLO algorithm suite was performed, encompassing YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, YOLOv8, and YOLOv9, prior to the commencement of ablation studies on the integrated algorithmic framework.

In the course of the ablation study, a module incorporating logical rules for event judgment was added to the YOLOv9+DeepSORT+TSM configuration.

The experimental design involved segmenting the testing area based on distance, with multiple independent sampling tests conducted within each segment to ensure the reliability of the findings; as depicted in Figure 7, the test area is divided into three effective regions: the proximal region (T1), the middle region (T2–T4), and the distal region (T5).

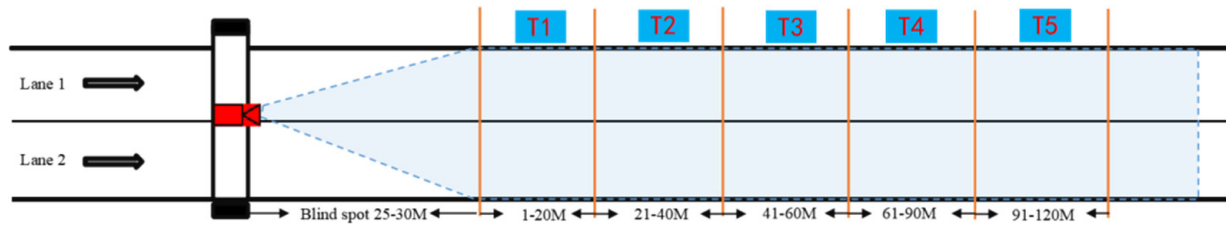


Figure 7. Test viewshed division reference diagram.

The evaluation criteria for the experimental outcomes are delineated by the following metrics: *Accuracy*, *Precision*, *Recall*, *MAP* (Mean Average Precision), and *FPS* (Frames Per Second). As shown in Equations (2)–(4) and (6).

For a given sample, there are four possible outcomes of classification predictions: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). By aggregating these outcomes across all test samples, one can derive evaluation metrics such as Accuracy, Precision, and Recall.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{AP} = \sum_{n=1}^N P_n \cdot \omega_n \quad (5)$$

In the formula, P_n is the precision at the n -th recall level, ω_n is the weight corresponding to the recall level (usually the recall rate itself), and N is the number of recall levels. In the formula, N represents the total number of classes, and AP_i denotes the Average Precision for the i -th class.

$$\text{MAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (6)$$

3.3. Results Analysis

This article compares the performance of YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, YOLOv8, and YOLOv9. The outcomes of this comparison are detailed in Table 1 below. Based on Table 1, the YOLOv5s, YOLOv5l, and YOLOv5m algorithms exhibit higher computational efficiency compared to YOLOv8 and YOLOv9. In terms of Recall, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv8 slightly outperform YOLOv9. However, YOLOv9 surpasses other YOLO algorithms in both MAP and Precision. Considering the high complexity of the tunnel inspection scenario and the limited computing power of edge devices, this study selects YOLOv9 for object detection.

Table 1. Comparison table of YOLO algorithm.

Model Name	Epoch/Iteration	Batch Size	MAP	Processing Frame Rate (fps)	Precision	Recall
YOLOv5s	300 epoch	512	57.1	142	84.77%	76.44%
YOLOv5m	300 epoch	512	58.7	105	85.71%	75.98%
YOLOv5l	300 epoch	512	61.6	68	85.93%	77.29%
YOLOv5x	300 epoch	512	61.7	55	89.17%	77.54%
YOLOv8	300 epoch	512	61.9	45	89.23%	77.32%
YOLOv9	300 epoch	512	62.1	60	90.08%	76.59%

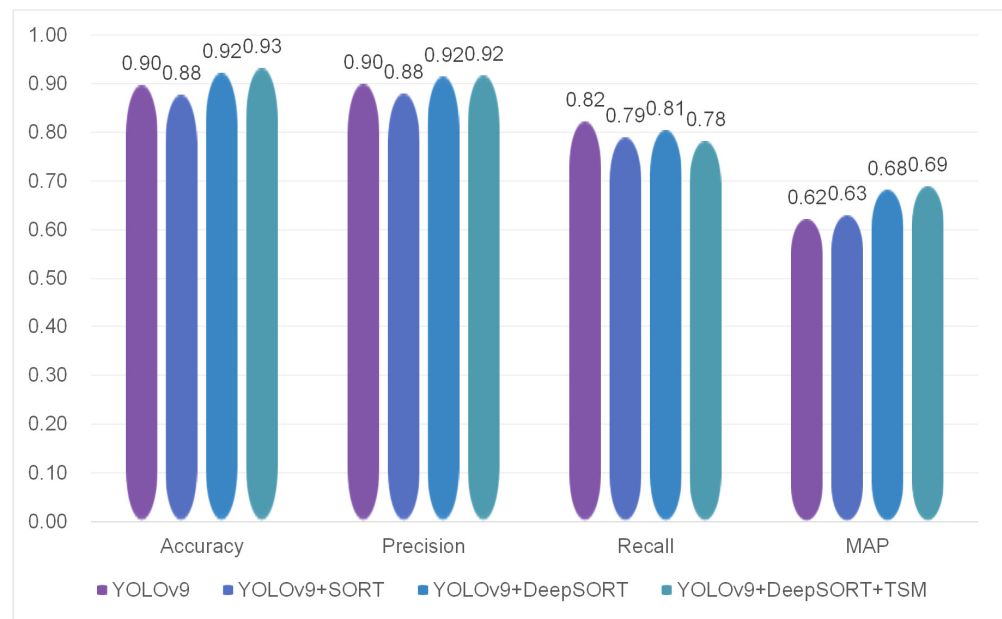
Based on the results of the ablation study in Table 2, the YOLOv9+DeepSORT+TSM algorithm proposed in this paper achieves higher Accuracy, Precision, and MAP, outperforming YOLOv9, YOLOv9+SORT, and YOLOv9+DeepSORT. Specifically, the Accuracy is 3.5% higher than the baseline algorithm YOLOv9, the Precision is 1.75% higher, and the MAP is 6.75% higher. However, in terms of Recall and FPS, YOLOv9+DeepSORT+TSM is slightly inferior to the baseline algorithm, with a 4% lower Recall and 4.75 frames per second (FPS) less, which is not significantly noticeable in practical applications.

Table 2. The ablation study results.

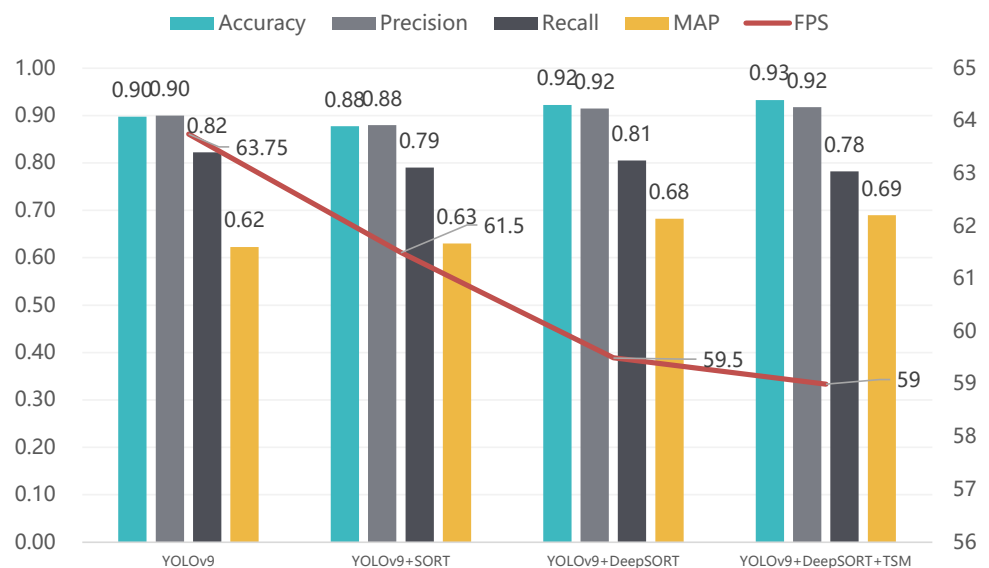
Algorithms	Traffic Events	Accuracy	Precision	Recall	MAP	FPS
YOLOv9	Abnormal parking	0.92	0.90	0.82	0.66	66
	Pedestrians	0.92	0.88	0.80	0.65	64
	Wrong-way driving	0.85	0.92	0.84	0.58	65
	Flames	0.90	0.90	0.83	0.60	60
	Average	0.8975	0.9000	0.8225	0.6225	64
YOLOv9+SORT	Abnormal parking	0.89	0.92	0.86	0.68	66
	Pedestrians	0.90	0.87	0.74	0.60	64
	Wrong-way driving	0.84	0.85	0.84	0.64	58
	Flames	0.88	0.88	0.72	0.60	58
	Average	0.8775	0.8800	0.7900	0.6300	62
YOLOv9+DeepSORT	Abnormal parking	0.92	0.90	0.80	0.70	65
	Pedestrians	0.93	0.94	0.78	0.66	60
	Wrong-way driving	0.94	0.92	0.82	0.68	55
	Flames	0.90	0.90	0.82	0.69	58
	Average	0.9225	0.9150	0.8050	0.6825	60
YOLOv9+DeepSORT+TSM	Abnormal parking	0.94	0.92	0.85	0.72	60
	Pedestrians	0.95	0.94	0.76	0.70	58
	Wrong-way driving	0.92	0.90	0.78	0.66	60
	Flames	0.92	0.91	0.74	0.68	58
	Average	0.9325	0.9175	0.7825	0.6900	59

The experimental results are analyzed and displayed through bar charts and line charts. As shown in Figure 8a, it can be observed that the recognition effect achieved by integrating object recognition algorithms with multiple object tracking algorithms

is superior to that obtained by using only object recognition algorithms. Furthermore, the addition of a Time Shift Module (TSM) to the fusion algorithm further enhances the recognition effect compared to the simple integration of object recognition and multiple object tracking algorithms. In Figure 8b, it is evident that all four algorithm structures based on YOLOv9 achieved an Accuracy of 90% or higher, indicating that the YOLOv9 algorithm has demonstrated applicable performance in the experiments. Additionally, with the addition of fusion algorithm modules, there is a slight decrease in frames per second (FPS). However, the YOLOv9+DeepSORT+TSM shows a negligible decrease in processing speed compared to YOLOv9+DeepSORT, suggesting that the incorporation of the TSM module has a minimal impact on computational power.



(a)



(b)

Figure 8. (a) Comparison of the same evaluation metrics for different fusion algorithms. (b) Performance evaluation of different fusion algorithms.

Figure 9 demonstrates the event recognition effects of the fusion algorithm YOLO9+DeepSORT+TSM. Specifically, Figure 9a presents a captured image of a vehicle driving in the opposite direction in a tunnel, Figure 9b presents an event recognition image of a vehicle abnormally parked in the tunnel, Figure 9c shows a captured image of a pedestrian intrusion event in the tunnel, and Figure 9d displays the recognition effect image of a simulated fire event in the tunnel (using a brazier inside a vehicle to mimic the fire incident).

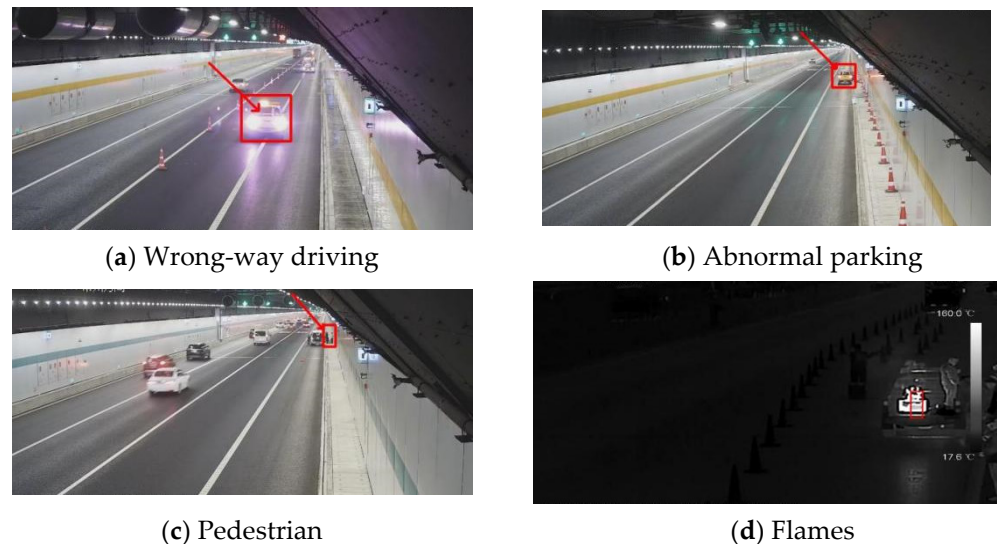


Figure 9. Traffic event detection effectiveness.

4. Conclusions

This research paper aims to meet the demands of daily inspection, emergency handling, and intelligent operation by focusing on event detection in the mobile state. Specifically, it focuses on event detection based on intelligent tunnel robots in smart tunnels operating in a mobile state. The core contributions and conclusions of this study are as follows:

Design an event detection algorithm in a moving state. To effectively detect events in the dynamic state of an intelligent tunnel robot, this paper utilizes YOLOv9 for object detection and identification within tunnels. Through comprehensive improvements in detection speed and recognition accuracy, real-time and accurate detection of tunnel traffic events is achieved. The DeepSORT algorithm is employed to detect image depth features within a specific frame range, while leveraging the ReID domain model for feature extraction. This helps reduce target ID jumps and enhance target matching accuracy. Time Shift Module (TSM) is utilized to model temporal information from video frame features, capturing time series characteristics and enhancing the algorithm's ability to detect time-related features.

Redesign of event analysis rules, such as abnormal parking in mobile device states. By employing the real-time position of camera motion and pixel coordinates, the position coordinates of events in different video frame states are calibrated and derived. Based on the distance in the position of continuous N-frame pictures over a period of time, calculations determine whether the vehicle is moving or stationary. If the vehicle's coordinate position does not change within the detection frame and time frame, it is considered abnormal parking.

The training and testing of algorithms for mobile inspection equipment provide a significant reference value for practical applications. The research constructs tracks for tunnel robots in both indoor simulated and real tunnel environments, and the appropriate AI edge computing equipment for the tunnel robot was selected, thereby enabling the inspection robots to detect four types of events—wrong-way driving, abnormal parking, pedestrians, and flames, while the inspection robot is in motion. The research holds significant practical implications for mobile inspection within actual tunnel environments.

Furthermore, the algorithm could be considered for application in intelligent driving vehicles, thereby enhancing the perception of traffic events in tunnels and along the entire road section, effectively improving the level of road safety.

Although the proposed improved method for event detection algorithms in the mobile detection state based on intelligent tunnel robots demonstrates superior performance in recognizing mobile device states, there remains scope for further enhancement. Future studies can explore the implementation of transformer algorithms in the moving state, optimize data acquisition and algorithm training based on actual tunnel scene events, and further improve the real-time accuracy of event detection. These advancements will contribute to enhanced intelligent operation and traffic safety management of key road sections such as tunnels, thereby elevating the construction and operational standards of smart tunnels.

Author Contributions: Conceptualization, L.W. and K.W.; Methodology, L.W. and Z.L.; Software, K.W.; Validation, C.Z., G.C. and P.Z.; Formal Analysis, C.Z., G.C. and P.Z.; Investigation, C.Z., G.C. and P.Z.; Resources, L.W.; Data Curation, K.W. and P.Z.; Writing—Original Draft Preparation, L.W., C.Z. and G.C.; Writing—Review & Editing, L.W. and K.W.; Visualization, K.W. and P.Z.; Supervision, K.W.; Project Administration, L.W.; Funding Acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Shandong Provincial Department of Transportation (a government agency) in China, Project Name: Research on Intelligent Management and Control Technology of Tunnels Based on Improving Traffic Efficiency and Camp Safety (June 2019 to June 2022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. And researchers have entered into a contract with a third party, which includes confidentiality clauses requiring that the relevant data shall not be disclosed within a specified period. Requests to access the datasets should be directed to authors.

Acknowledgments: This work was supported by the project “Research on Intelligent Management and Control Technology of Tunnels Based on Improving Traffic Efficiency and Camp Safety”. The authors would like to thank our colleagues at Shandong Provincial Communications Planning and Design Institute Group Co., Ltd. and Beijing Zhuoshi Zhitong Technology Co., Ltd. for their excellent technical support and valuable suggestions.

Conflicts of Interest: Authors Li Wan, Zhenjiang Li, Changan Zhang, and Guangyong Chen were employed by the company Shandong Provincial Communications Planning and Design Institute Group Co., Ltd. Authors Panming Zhao and Kewei Wu were employed by the company Beijing Zhuoshi Zhitong Technology Co., Ltd. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Li, W.F.; Yuan, C.; Li, K.; Ding, H.; Liu, Q.Z.; Jiang, X.H. Application of Inspection Robot on Operation of Highway Tunnel. *Technol. Highw. Transp.* **2021**, *37*, 35–40.
2. Wu, W.B. Automatic Tunnel Maintenance Detection System based on Intelligent Inspection Robot. *Autom. Inf. Eng.* **2021**, *42*, 46–48.
3. Wang, Z.J. Application of robotics in highway operation. *Henan Sci. Technol.* **2020**, *19*, 98–100.
4. Lee, C.; Kim, H.; Oh, S.; Doo, I. A Study on Building a “Real-Time Vehicle Accident and Road Obstacle Notification Model”. Using AI CCTV. *Appl. Sci.* **2021**, *11*, 8210. [[CrossRef](#)]
5. Nancy, P.; Dilli Rao, D.; Babuaravind, G.; Bhanushree, S. Highway Accident Detection and Notification Using machine Learning. *Int. J. Comput. Sci. Mob. Comput.* **2020**, *9*, 168–176.
6. Ghosh, S.; Sunny, S.J.; Roney, R. Accident Detection Using Convolutional Neural Networks. In Proceedings of the 2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India, 1–2 March 2019. [[CrossRef](#)]
7. Zhang, X. Research on Traffic Event Recognition Method Based on Video Classification and Video Description. Master’s Dissertation, ShanDong University, Jinan, China, 2022.

8. Basheer Ahmed, M.I.; Zaghdoud, R.; Ahmed, M.S.; Sendi, R.; Alsharif, S.; Alabdulkarim, J.; Albin Saad, B.A.; Alsabt, R.; Rahman, A.; Krishnasamy, G. A Real-Time Computer Vision Based Approach to Detection and Classification of Traffic Incidents. *Big Data Cogn. Comput.* **2023**, *7*, 22. [[CrossRef](#)]
9. Ren, A.H.; Li, Y.F.; Chen, Y. Improved Detection of Unusual Highway Traffic Events for YOLOv8. *Laser J.* 2024. Available online: <https://link.cnki.net/urlid/50.1085.TN.20240628.0926.002> (accessed on 28 June 2024).
10. Li, Y. Research on Autonomous Mobile Platform of Intelligent Tunnel Inspection Robot Based on Laser SLAM. Master's Dissertation, Chang'an University, Xi'an, China, 2021.
11. Manana, M.; Tu, C.; Owolawi, P.A. A survey on vehicle detection based on convolution neural networks. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; IEEE: Chengdu, China, 2017; pp. 1751–1755.
12. Rosso, M.M.; Giulia, M.; Salvatore, A.; Aloisio, A.; Chiaia, B.; Marano, G.C. Convolutional networks and transformers for intelligent road tunnel investigations. *Comput. Struct.* **2023**, *275*, 106918. [[CrossRef](#)]
13. Liu, J.; Zhao, Z.Y.; Lv, C.S.; Ding, Y.; Chang, H.; Xie, Q. An image enhancement algorithm to improve road tunnel crack transfer detection. *Constr. Build. Mater.* **2022**, *348*, 128583. [[CrossRef](#)]
14. Zhang, G.; Yin, J.; Deng, P.; Sun, Y.; Zhou, L.; Zhang, K. Achieving Adaptive Visual Multi-Object Tracking with Unscented Kalman Filter. *Sensors* **2022**, *22*, 9106. [[CrossRef](#)] [[PubMed](#)]
15. Papageorgiou, C.P.; Oren, M.; Poggio, T. A general framework for object detection. In Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), Bombay, India, 7 January 1998; IEEE: Bombay, India, 1998; pp. 555–562.
16. Yang, B.; Zhang, Y.; Cao, J.; Zou, L. On road vehicle detection using an improved faster RCNN framework with small-size region up-scaling strategy. In Proceedings of the Image and Video Technology: PSIVT 2017 International Workshops, Wuhan, China, 20–24 November 2017; Revised Selected Papers 8. Springer International Publishing: Wuhan, China, 2018; pp. 241–253.
17. Lee, W.J.; Pae, D.S.; Kim, D.W.; Lim, M.T. A vehicle detection using selective multi-stage features in convolutional neural networks. In Proceedings of the 17th International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea, 18–21 October 2017; IEEE: Jeju, Republic of Korea, 2017; pp. 1–3.
18. Li, L.H.; Lun, Z.M.; Lian, J. Road Vehicle Detection Method based on Convolutional Neural Network. *J. Jilin Univ. (Eng. Technol. Ed.)* **2017**, *47*, 384–391.
19. Fu, Z.Q. Research on Key Technology for Tunnel Monitoring of Inspection Robot Based on Image Mosaic. Master's Dissertation, Ningbo University, Ningbo, China, 2020.
20. Ma, H.J. Design and research of highway tunnel inspection robot system. *West. China Commun. Sci. Technol.* **2022**, *10*, 122–124.
21. Tan, S. Research on Object Detection Method Based on Mobile Robot. Master's Dissertation, Beijing University of Civil Engineering and Architecture, Beijing, China, 2020.
22. Li, J.Y. Research on the application of automatic early warning system for safety monitoring of cut-and-cover tunnel foundation pit based on measurement robot technology. *J. Highw. Transp. Res. Dev. Appl. Technol. Ed.* **2020**, *6*, 267–269.
23. Tian, F.; Meng, C.L.; Liu, Y.C. Research on the construction scheme of intelligent tunnel in Jiangsu Taihu Lake for generalized vehicle-road coordination. *Highw. Transp. Res. Dev. Appl. Technol. Ed.* **2020**, *16*, 268–272.
24. Jin, Y.; Xu, Y.; Han, F.Y.; He, S.Y.; Wang, J.B. Pixel-level Recognition Method of Bridge Disease Image Based on Deep Learning Semantic Segmentation. *Highw. Transp. Res. Dev. Appl. Technol. Ed.* **2020**, *16*, 183–188.
25. Song, Z.L. Intelligent Analysis of Traffic Events Based on Software-defined Cameras. Master's Dissertation, Southwest Jiaotong University, Chengdu, China, 2021.
26. Hu, M. Design and Implementation of Intelligent Video Analysis System Based on Deep Learning. Master's Dissertation, Huazhong University of Science & Technology, Wuhan, China, 2021.
27. Zhang, C.S.; Zhou, D. Application of event intelligent detection system based on radar and thermal imaging technology in extra-long tunnels. *J. Highw. Transp. Res. Dev. Appl. Technol. Ed.* **2020**, *16*, 313–316.
28. Dong, M.L. Research on Highway Traffic Incident Detection Based on Deep Learning. Doctoral Dissertation, Xi'an Technological University, Xi'an, China, 2022.
29. Wang, C.Y.; Yeh, I.H.; Liao, H.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.