

Article

Not peer-reviewed version

A Novel Machine Vision-Based Collision Risk Warning Method for Unsignalized Intersections on Arterial Roads

Zhongbin Luo , [Yanqiu Bi](#) ^{*} , [Qing Ye](#) , [Yong Li](#) , [Shaofei Wang](#) ^{*}

Posted Date: 14 February 2025

doi: 10.20944/preprints202502.1123.v1

Keywords: YOLOv8; Deep SORT; machine vision; collision risk; unsignalized intersection



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

A Novel Machine Vision-Based Collision Risk Warning Method for Unsignalized Intersections on Arterial Roads

Zhongbin Luo ^{1,2}, Yanqiu Bi ^{3,4,*}, Qing Ye ², Yong Li ¹, Shaofei Wang ²

¹ College of Computer Science, Chongqing University, Chongqing 400044, China

² China Merchants Chongqing Communications Research and Design Institute Co, Ltd., Chongqing 400067, China

³ National & Local Joint Engineering Research Center of Transportation Civil Engineering Materials, Chongqing Jiaotong University, Chongqing 400074, China

⁴ School of Civil Engineering, Chongqing Jiaotong University, Chongqing 400074

* Correspondence: biyanqiu@cqjtu.edu.cn

Abstract: Addressing the critical need for collision risk warning at unsignalized intersections, this study proposes an advanced predictive system combining YOLOv8 for object detection, Deep SORT for tracking, and Bi-LSTM networks for trajectory prediction. To adapt YOLOv8 for complex intersection scenarios, several architectural enhancements were incorporated. The RepLayer module replaced the original C2f module in the backbone, integrating large-kernel depthwise separable convolution to better capture contextual information in cluttered environments. The GIoU loss function was introduced to improve bounding box regression accuracy, mitigating issues related to missed or incorrect detections due to occlusion and overlapping objects. Furthermore, a Global Attention Mechanism (GAM) was implemented in the neck network to better learn both location and semantic information, while the ReContext gradient composition feature pyramid replaced the traditional FPN, enabling more effective multi-scale object detection. Additionally, the CSPNet structure in the neck was substituted with Res-CSP, enhancing feature fusion flexibility and improving detection performance in complex traffic conditions. For tracking, the Deep SORT algorithm was optimized with enhanced appearance feature extraction, reducing identity switches caused by occlusions and ensuring stable tracking of vehicles, pedestrians, and non-motorized vehicles. The Bi-LSTM model was employed for trajectory prediction, capturing long-range dependencies to provide accurate forecasting of future positions. Collision risk was quantified using the Predictive Collision Risk Area (PCRA) method, categorizing risks into three levels (danger, warning, and caution) based on predicted overlaps in trajectories. In the experimental setup, the dataset used for training the model consisted of 30,000 images, annotated with bounding boxes around vehicles, pedestrians, and non-motorized vehicles. Data augmentation techniques such as Mosaic, Random_perspective, Mixup, HSV adjustments, Flipud, and Fliplr were applied to enrich the dataset and improve model robustness. In real-world testing, the system was deployed as part of the G310 highway safety project, where it achieved a mean Average Precision (mAP) of over 90% for object detection. Over a one-month period, 120 warning events involving vehicles, pedestrians, and non-motorized vehicles were recorded. Manual verification of the warnings indicated a prediction accuracy of 97%, demonstrating the system's reliability in identifying potential collisions and issuing timely warnings. This approach represents a significant advancement for enhancing safety at unsignalized intersections in urban traffic environments.

Keywords: YOLOv8; Deep SORT; machine vision; collision risk; unsignalized intersection

1. Introduction

1.1 Background and Motivation

Pedestrians and non-motorized vehicles are essential components of urban transportation systems and constitute a vulnerable group among road users, as shown in Figure 1. This type of scenario is common in arterial roads and rural roads, where the main road typically consists of two-way, two-lane traffic, while the secondary road is usually a single-lane roadway. The main road primarily serves motor vehicles, while the secondary road is mainly used by pedestrians, non-motorized vehicles, and agricultural vehicles. Traffic volume is generally low, but the lack of signal control results in significant conflicts over right-of-way between pedestrians and motor vehicles. This lack of regulation exacerbates safety risks at intersections, as pedestrians and non-motorized vehicles often struggle to navigate the roads alongside faster-moving motor vehicles. Globally, one-third of all traffic accidents involve pedestrians [1]. According to the World Health Organization's (WHO) Global Status Report on Road Safety 2018, approximately 1.35 million people die annually in road traffic accidents worldwide, with pedestrians and cyclists accounting for 26% of these fatalities [2]. In China, 273,098 road traffic accidents were reported in 2021 [3]. Statistics indicate that pedestrians accounted for 25.15% of fatalities and 17.21% of injuries in these accidents [4]. Ensuring the safety of vulnerable road users (VRUs), particularly pedestrians and non-motorized vehicle users, has become a key research focus in both academia and industry.

Unsignalized intersections present a significant challenge for traffic safety due to the absence of regulatory control devices such as traffic signals and stop signs. This lack of control requires road users to rely heavily on their judgment to navigate these intersections safely. Consequently, these intersections have a higher risk of accidents compared to signalized intersections [5]. According to the Federal Highway Administration (FHWA), intersection-related crashes account for approximately 50% of all traffic accidents, with unsignalized intersections being a critical contributor [6]. These accidents often result in severe injuries and fatalities, emphasizing the need for improved safety measures.

Human factors are a major cause of accidents at unsignalized intersections. Driver behavior, including failure to yield, distracted driving, and speeding, significantly increases the risk of accidents [7]. Similarly, pedestrians and cyclists may also exhibit unpredictable behavior, further complicating the safety dynamics of these intersections [8]. Studies have shown that driver misjudgment of gaps and speeds is a common cause of accidents at unsignalized intersections [9].

Environmental and infrastructural factors also play a crucial role in the safety of these intersections. Poor visibility due to inadequate lighting or obstructed sightlines, adverse weather conditions, and roadway design flaws contribute to the risk of accidents [10]. For instance, intersections located on curves or with multiple entry points pose additional challenges for road users [11].

Traditional safety measures at unsignalized intersections include road signs, pavement markings, and public awareness campaigns. However, these measures often fall short in preventing accidents due to their static nature and the reliance on road users' compliance and attentiveness [12]. The advent of intelligent transportation systems (ITS), machine learning, and computer vision technologies offers new opportunities for enhancing traffic safety at these intersections [13]. Systems that can monitor traffic in real-time, detect potential collision risks, and provide timely warnings to road users are emerging as promising solutions. Trajectory prediction models can provide real-time assessment of potential traffic conflicts, issuing early warnings and enabling timely interventions to prevent accidents. In high-risk areas like unsignalized intersections, predicting conflicts between vehicles and pedestrians not only helps prevent accidents but also provides technical support for future intelligent transportation systems and autonomous driving technologies. Therefore, there is an urgent need to develop conflict trajectory prediction models that adapt to the characteristics of unsignalized intersections in order to address the complex and dynamic traffic environments at these locations. Although trajectory prediction has made significant progress in fields like traffic flow and

autonomous driving, there is still a notable gap in research on vehicle-pedestrian conflict trajectory prediction at unsignalized intersections, which are complex traffic environments. At unsignalized intersections, due to the lack of signal control, road users (such as pedestrians, non-motorized vehicles, and motor vehicles) must rely heavily on their own judgment, leading to poorer traffic flow and safety. Currently, research on trajectory prediction for conflicts at unsignalized intersections is limited, especially concerning the interaction and prediction of conflicts between vehicles and pedestrians. Existing trajectory prediction models typically focus on predicting the trajectory of individual road users (e.g., vehicles or pedestrians alone), but there is insufficient research on the complex interactions and potential conflicts between vehicles and pedestrians or vehicles and non-motorized vehicles[14].

Furthermore, existing studies do not fully consider the impact of dynamic traffic environments on prediction accuracy. Factors such as traffic participants' decision-making behaviors (e.g., whether a pedestrian decides to cross the road, whether a vehicle slows down or changes lanes) significantly influence collision risk. In addition, most current models rely on historical trajectory data without sufficiently addressing the effect of environmental factors (such as lighting conditions, weather, and road obstructions) on behavior prediction.

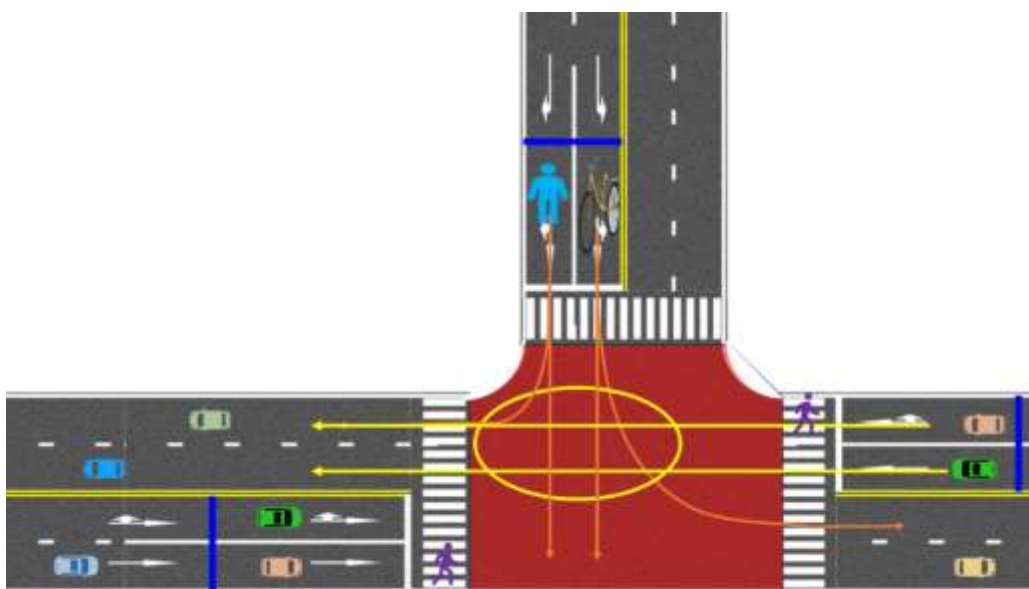


Figure 1. Unsignalized Intersection Scenario.

1.2 Problem Statement

Despite advancements in traffic management technologies, the inherent risk at unsignalized intersections persists due to the absence of control mechanisms. Traditional methods, such as static signs and road markings, are often inadequate as they depend heavily on the compliance and alertness of road users [15]. Existing traffic management systems primarily focus on signalized intersections, where traffic signals help regulate the flow and reduce the likelihood of collisions. However, these systems are not directly applicable to unsignalized intersections, where the absence of signals requires a different approach [16].

The dynamic nature of traffic flow at unsignalized intersections requires real-time monitoring and proactive intervention to effectively mitigate collision risks. This involves accurately detecting and tracking the movements of various road users and predicting their trajectories to identify potential conflicts [17]. Advanced object detection and tracking technologies, such as YOLOv8 [18] and Deep SORT [19], provide the necessary tools to monitor traffic in real-time. These technologies can identify and track multiple road users simultaneously, offering a comprehensive view of the traffic dynamics at unsignalized intersections.

Deep learning-based trajectory prediction models can analyze historical movement patterns of road users to forecast their future positions. By predicting potential collision points, these models enable the system to issue timely warnings and prevent accidents [20]. The integration of these technologies into a cohesive system designed for unsignalized intersections represents a significant advancement in traffic safety.

- The primary objectives of this research are:
- To develop a real-time collision warning system for unsignalized intersections using YOLOv8 and Deep SORT for object detection and tracking.
- To implement deep learning models for accurate trajectory prediction of road users.
- To evaluate the effectiveness of the proposed system in various traffic scenarios and assess its potential for reducing accident risks.

1.3 Organization of the Paper

This research aims to develop a predictive collision warning system for unsignalized intersections that integrates advanced object detection, tracking, and trajectory prediction technologies. Specifically, the system will utilize YOLOv8 for high-precision object detection and Deep SORT for robust tracking of Pedestrians and non-motorized vehicles, and then predicts their trajectories by using deep Bi-LSTM networks [21]. With use of predicted trajectories, this system infers collision risk areas statistically, further, severity of levels is categorized as danger, warning, and relative safe.

The remainder of this paper consists of three chapters as follows:

1. Materials and methods: Descriptions of a predictive collision risk area estimation system based on Bi-LSTM networks and statistical inference method.
2. Experiments and results: Validating feasibility and applicability of the proposed system, and discussion of results and limitations.
3. Conclusions: Summary of our study and future research directions.

2 Literature Review

2.1 Traffic Accident Prevention Systems

Recent advancements in intelligent transportation systems (ITS) have introduced dynamic and proactive measures to enhance traffic safety. ITS applications, such as adaptive traffic control systems, vehicle-to-infrastructure (V2I) communication, and automated enforcement technologies, have shown promise in reducing accidents and improving traffic flow [22]. Adaptive signal control technologies, for example, can significantly reduce delays and improve safety at intersections by adjusting signal timings based on real-time traffic conditions [23].

Collision avoidance systems, utilizing sensors and communication technologies, are increasingly integrated into modern vehicles. Systems such as forward collision warning (FCW) and automatic emergency braking (AEB) have been shown to reduce rear-end collisions and other types of accidents [24,25]. Integrating these systems into broader traffic management infrastructures remains a vital area of ongoing research and development [26].

2.2 Object Detection and Tracking Technologies

Object detection and tracking are crucial for modern traffic safety systems. The evolution of computer vision technologies has enabled accurate and efficient detection and tracking of objects [27]. Fusion techniques for different scenes and images have significantly enhanced the accuracy and efficiency of visual object detection technology. By combining multiple sources of information or integrating various imaging techniques, these methods enable the model to better understand complex environments, leading to more precise detections under diverse conditions. The fusion of data from different perspectives, such as from multiple cameras or sensor modalities, helps address challenges like occlusion, varying lighting, and different object scales, ultimately improving the

robustness and performance of object detection systems [28]. YOLO (You Only Look Once) and its variants, like YOLOv8, are noted for their high-speed and high-accuracy object detection capabilities [29–31].

YOLOv8 uses deep convolutional neural networks to detect objects in real-time, making it particularly suitable for traffic monitoring applications where real-time performance is crucial [32]. Deep SORT combines deep learning for appearance feature extraction with a traditional Kalman filter and Hungarian algorithm for data association, providing robust tracking of detected objects [33].

The combination of YOLOv8 and Deep SORT has been explored in various studies, demonstrating their effectiveness in tracking multiple objects simultaneously in dynamic environments [34,35]. This combination is particularly useful in traffic scenarios where accurate detection and continuous tracking of vehicles and pedestrians are essential for collision prevention. Many studies have proposed signal denoising and fault detection methods based on wavelet packets and machine learning algorithms to improve data quality and decision support accuracy in intelligent transportation systems and industrial processes [36]. In this paper, the tracking algorithm employs the Deep SORT approach, utilizing the linear Kalman filter for state estimation and prediction in dynamic environments. The advantage of this method lies in its ability to effectively handle uncertainty and noise in both object motion and sensor measurements, allowing for reliable and accurate object tracking even in challenging conditions such as occlusions and fast-moving targets. The Kalman filter's prediction and update steps enable robust tracking by continuously refining state estimates, making it highly efficient for real-time applications.

2.3 Trajectory Prediction Methods

Trajectory prediction is essential for proactive traffic safety systems, allowing for the forecasting of future positions of road users based on historical movement patterns. Existing trajectory prediction methods, including physics-based models, machine learning techniques (e.g., regression, decision trees, neural networks), and deep learning approaches (e.g., RNNs, LSTMs, GANs), each have their advantages and limitations. Physics-based models are simple and computationally efficient but lack the ability to model complex interactions in dynamic environments. Data-driven methods, particularly machine learning and deep learning models, excel at capturing complex patterns but require large datasets and significant computational resources, and may struggle with generalizing to unseen environments [37,38]. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are effective at modeling temporal dependencies but are sensitive to noise and require substantial training data.

LSTM-based models, for instance, outperform traditional methods in predicting pedestrian trajectories by capturing the temporal dependencies in their movements more effectively [39,40]. Similarly, Transformers, with their attention mechanisms, demonstrate superior performance in capturing long-range dependencies and complex motion patterns [41,42]. Applying these deep learning models to traffic safety systems is a rapidly growing area of research, highlighting their potential to enhance collision prediction and prevention [43–45].

Generative models like GANs can handle uncertainty by predicting multiple trajectories, but their computational cost and complexity are high [46]. Reinforcement learning (RL) offers potential for dynamic environments but faces challenges in stability and real-time application [47]. Hybrid models, combining the strengths of different approaches, offer promising solutions but are complex to implement [48]. For unsignalized intersections, where traffic behavior is highly dynamic and unpredictable, a hybrid approach that integrates data-driven techniques with physics-based models or reinforcement learning could provide more accurate and efficient predictions, balancing accuracy, computational efficiency, and real-time applicability.

3. Methodology

The proposed predictive collision warning system integrates cutting-edge object detection, tracking, and trajectory prediction technologies, specifically designed for deployment at unsignalized intersections. The system architecture is designed to operate in real-time, processing data from multiple sensors to monitor and predict potential collision risks among various road users, as shown in Figure 2.

3.1 System Architecture

The system comprises several key components:

1. Data sources: High-definition cameras and edge computing unit installed at strategic locations around the intersection.
2. Preprocessing: (1) Object Detection: Utilizing YOLOv8 for detecting vehicles, pedestrians, and cyclists. (2) Object Tracking: Implementing Deep SORT to track detected objects over time. (3) Trajectory Prediction: Using a Bi-LSTM model for predicting future positions of tracked objects.
3. Risk estimation: Evaluating the likelihood of collisions based on predicted trajectories.
4. Dangers warning: Through variable message signs (VMS) and audio-visual warning devices installed on the roadside, warnings are issued to vehicles on the main road and to pedestrians and non-motorized vehicles on the side roads.

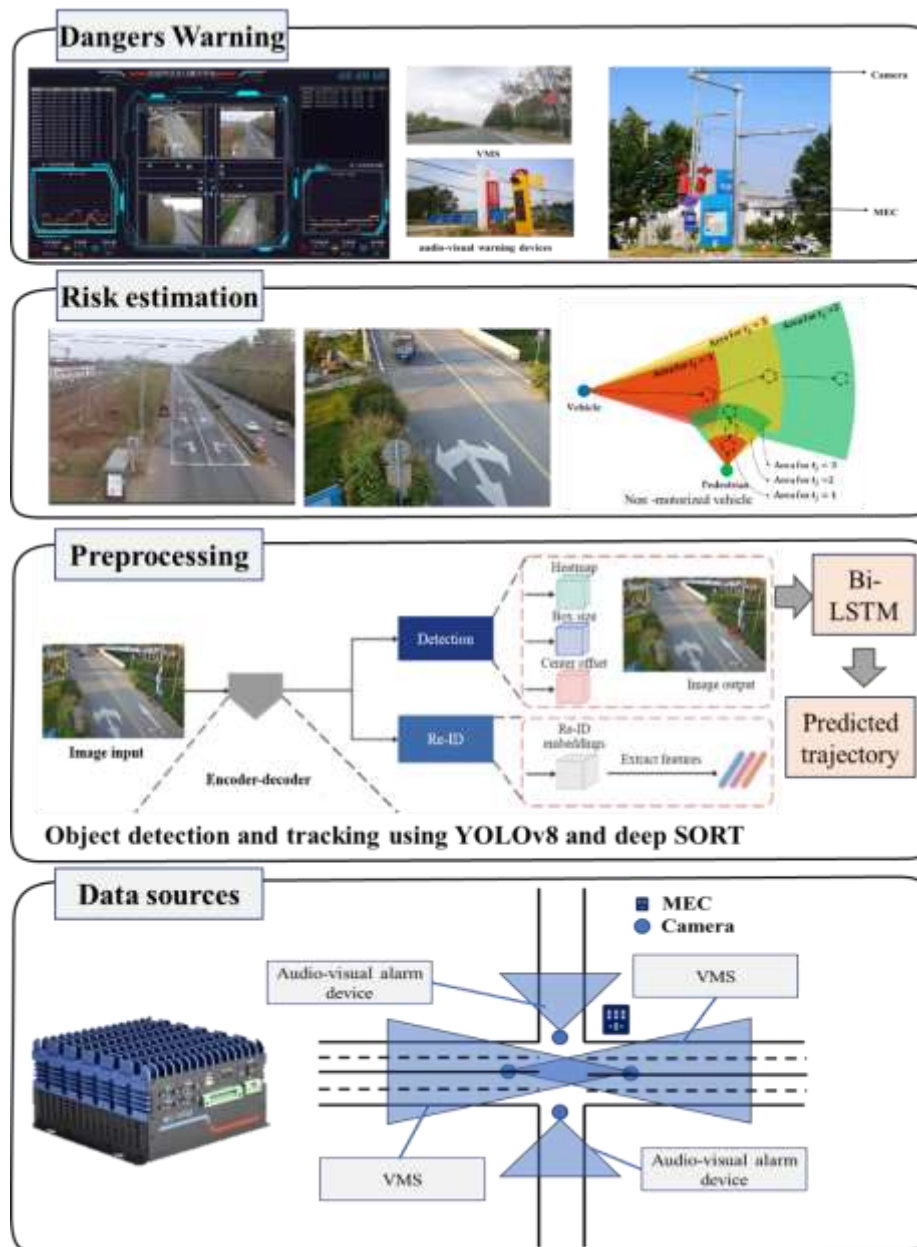


Figure 2. Overall structure of the proposed system.

3.2 Object Detection

YOLOv8 is employed for real-time object detection due to its high accuracy and speed. YOLOv8 is particularly suited for real-time applications, such as traffic monitoring at unsignalized intersections.

This paper proposes a small-target vehicle detection model for road traffic (RGGE-YOLOv8) based on YOLOv8, incorporating large kernels and multi-scale gradient composition. The network structure is shown in Figure 3.

In the Backbone module, the RepLayer module replaces the original C2f module, introducing large-kernel depthwise separable convolution to capture contextual information more effectively.

To address issues such as missed or incorrect detections caused by overlapping and occlusions during vehicle operations, the GIoU loss function is introduced. By integrating the minimum enclosing box information into the loss calculation, the model can accurately evaluate and correct the deviation between the predicted bounding box and the ground truth, thereby improving detection performance.

Additionally, the GAM (Global Attention Mechanism) is added to the neck network to enhance learning of both location and semantic information for key features.

Finally, the ReContext gradient composition feature pyramid is introduced, replacing the original FPN with CSPNet. Furthermore, the traditional 3x3 convolutional feature fusion in CSPNet is replaced by Res-CSP, which not only retains the advantages of CSPNet's partial connection structure but also enables more efficient and flexible feature fusion, enhancing the model's ability to detect multi-scale objects in complex scenarios.

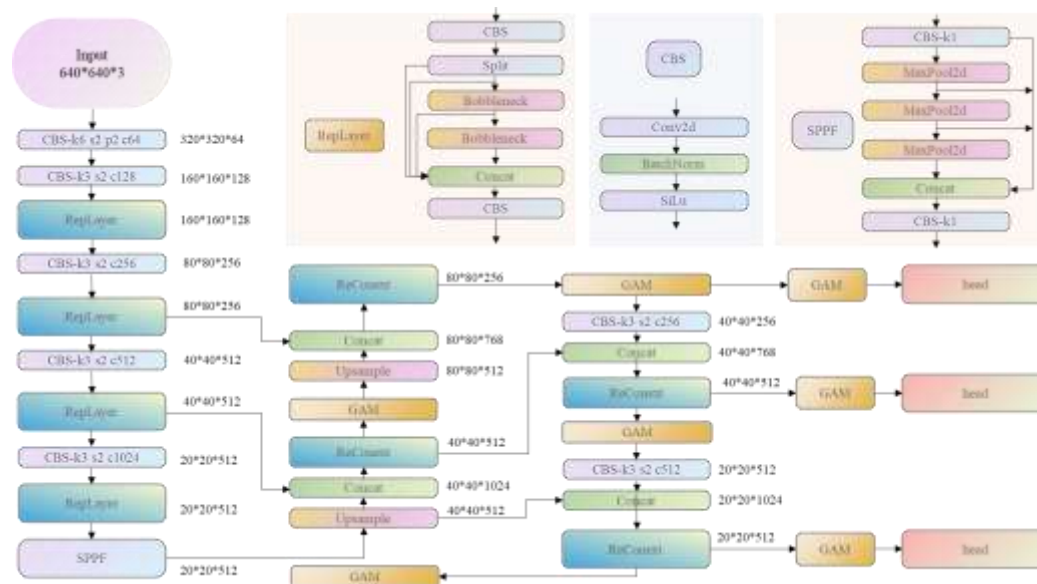


Figure 3. Overall framework diagram of RGGE-YOLOv8.

3.3 Object Tracking

Deep SORT is used for tracking detected objects across frames. Deep SORT enhances the original SORT algorithm by incorporating appearance information through a deep learning-based feature extractor.

Multi-Target Tracking enables the continuous and dynamic tracking of the movement of vehicles, non-motorized vehicles, and pedestrians within a scene, allowing for more accurate capture and analysis of their trajectory behavior. The Deep SORT algorithm is widely used for multi-target tracking, and its workflow is shown in Figure 4 and Figure 5.

The process begins with YOLOv8 detecting target locations. A Kalman filter then predicts the object's position in the next frame. The predicted position is compared with the actual detected position, and their similarity is calculated based on Intersection over Union (IoU). Finally, the Hungarian matching algorithm assigns corresponding IDs between consecutive frames, enabling consistent tracking of target types and behaviors.

As Deep SORT compares the features of objects across frames and incorporates cascade matching and new trajectory confirmation, it enhances tracking accuracy. As illustrated in Figure 5, after the detection network extracts each object's unique features, the algorithm strengthens these features through processes like prediction, observation, and update. This ensures that the object's ID remains consistent throughout tracking, allowing for more precise detection and analysis of its movement behavior.

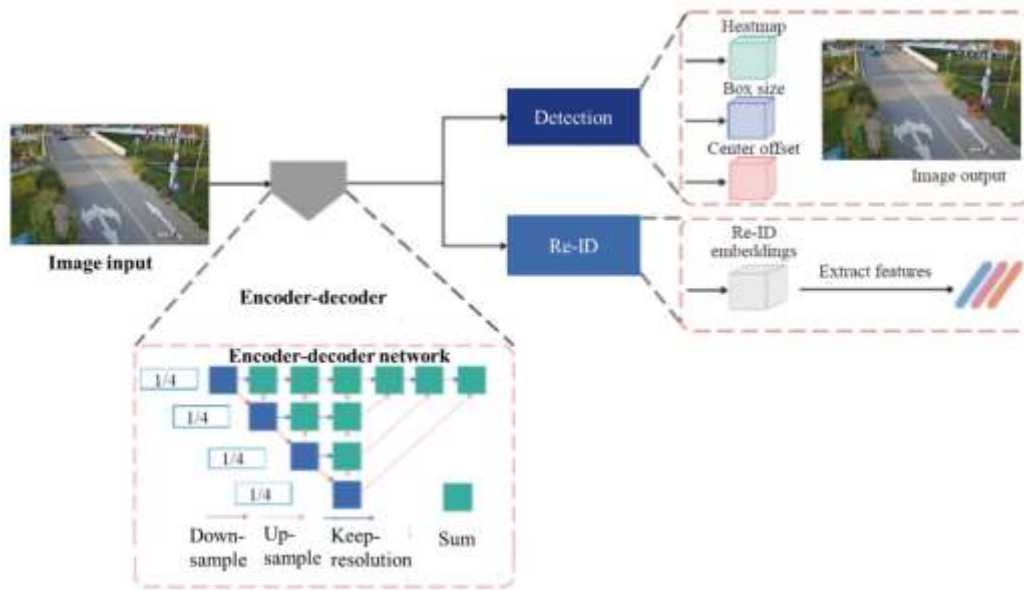


Figure 4. Overall framework diagram of RGGE-YOLOv8 structure of Deep SORT.

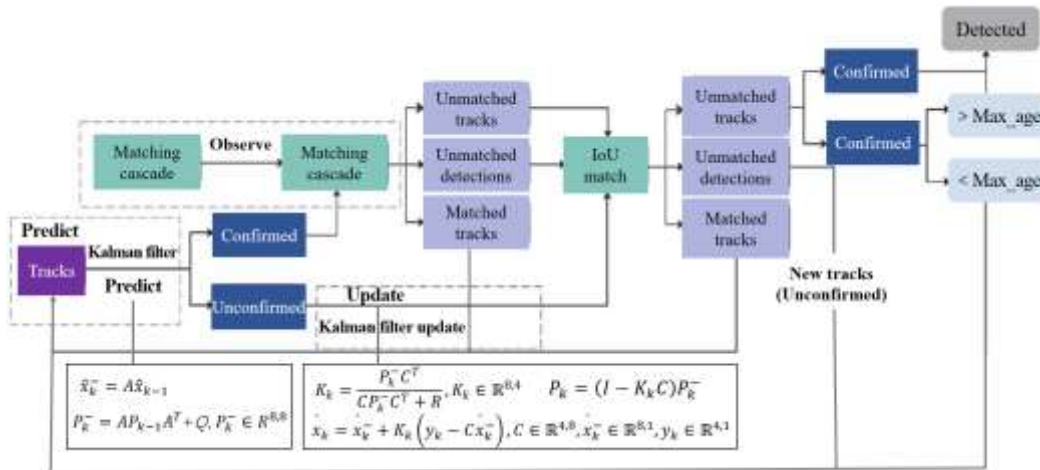


Figure 5. Workflow diagram of Deep SORT algorithm.

In the Deep SORT (Simple Online and Realtime Tracking) algorithm, the Kalman filter is a crucial tool for estimating the state of each tracked object. It predicts the object's future position and updates this prediction using new measurements. The Linear Kalman Filter used in Deep SORT is ideal for dynamic tracking applications, where object motion can be effectively modeled with linear equations. The performance of the Kalman filter depends on the proper tuning of two key parameters: the process noise covariance Q and the measurement noise covariance R . In this implementation, Q and R are dynamically adjusted based on certain weight coefficients reflecting the uncertainties in position and velocity, which are applied during the prediction and update steps. This dynamic adjustment allows the filter to adapt to varying conditions and ensure more accurate tracking results in real-time applications.

Prediction Phase:

In the prediction phase, the Kalman filter predicts the future state based on the previous state estimate and the motion model. The predicted state \hat{x}_k is calculated as:

$$\hat{x}_k = A\hat{x}_{k-1}, \quad (1)$$

where \hat{x}_k is the predicted state at time k , \hat{x}_{k-1} is the state estimate at time $k-1$, and A is the state transition matrix. This matrix A defines how the previous state evolves into the predicted state based on the motion model.

The predicted error covariance P_k^- is updated as:

$$P_k^- = AP_{k-1}A^T + Q, \quad (2)$$

where P_{k-1} is the error covariance from the previous step, and Q is the process noise covariance matrix. In this implementation, Q is dynamically adjusted based on the uncertainties in the predicted position and velocity. The adjustment is based on the standard deviation weights for position and velocity, which are calculated using the target's size (bounding box height). The weight coefficients such as std-weight-position and std-weight-velocity control how much uncertainty is assigned to the predicted position and velocity.

For instance, when the target's size or motion becomes unpredictable (such as fast movement or significant changes in direction), the system increases the value of Q to account for higher uncertainty in the motion model. Conversely, if the target is moving smoothly or predictably, Q can be reduced.

Update Phase:

In the update phase, the Kalman filter corrects the predicted state using new measurements. The Kalman gain K_k is computed to determine how much to adjust the predictions based on the new measurement:

$$K_k = \frac{P_k^- C^T}{C P_k^- C^T + R}, \quad (3)$$

where C is the observation matrix, and R is the measurement noise covariance matrix. In this implementation, R is also dynamically adjusted based on the uncertainties in the measurements. Specifically, the uncertainty in the position measurements (which depends on the target's height) influences R .

When measurements are noisy or unreliable, R increases, causing the Kalman filter to rely more on the model's predictions. Conversely, if the measurements are of high quality, R is reduced, allowing the filter to place more trust in the measurements.

The updated state estimate is calculated as:

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-), \quad (4)$$

where y_k is the new measurement at time k , and $C\hat{x}_k^-$ is the predicted measurement based on the predicted state. This step adjusts the predicted state by incorporating the new measurement, and the Kalman gain K_k determines how much weight to give to the prediction versus the measurement.

Finally, the error covariance P_k is updated as:

$$P_k = (I - K_k C)P_k^-, \quad (5)$$

where I is the identity matrix. This update reduces the uncertainty in the state estimate by incorporating the new measurement and the Kalman gain.

3.4. Vehicle Speed Measurement

3.4.1. Model Assumptions

To establish a reliable vehicle speed detection model, the study proposes the following foundational assumptions based on highway monitoring scenario characteristics [49]: (1) Road Plane Mapping Property: Through camera calibration technology, all target positions in monitoring images can be mapped to the $Z_w = 0$ plane of the world coordinate system (as shown in Figure 6). This assumption relies on the high flatness and absence of significant undulations characteristic of highway road surfaces. (2) Vehicle Motion Continuity: Within the time interval between consecutive video frames, vehicle trajectories can be approximated as linear changes. This assumption not only applies to straight-moving vehicles but is also compatible with curved driving scenarios through piecewise linearization processing. (3) Temporal Consistency Constraint: The time interval between

adjacent frames in the video stream remains constant, ensuring unified time reference for speed calculations.

3.4.2 Model Design and Implementation

Based on the aforementioned assumptions, the vehicle speed detection model is constructed through the following workflow: First, the YOLO object detection algorithm is employed to extract vehicle bounding boxes. Geometric parameters of the detection boxes are calculated to obtain the bottom-edge center coordinates as the vehicle position reference point (providing closer proximity to the actual ground contact point compared to traditional vertex coordinates). For each target vehicle in every video frame, a displacement vector relationship between adjacent frames is established (as shown in Equation 6).

$$d_i = u_i(t) - u_i(t - \Delta t), \quad (6)$$

Here, $u_i(t)$ represents the bottom-edge center coordinates of the vehicle detection box in the current frame; $u_i(t - \Delta t)$ denotes the corresponding coordinates in the previous frame; Δt is the fixed frame interval; i indicates the trajectory tracking point sequence. To convert pixel displacement d_i into actual speed values, a coordinate mapping model is required to transform the image pixel coordinate system into the world coordinate system. Traditional methods rely on camera parameter calibration (e.g., focal length, installation height, intrinsic matrix), but such calibration processes are operationally complex and environmentally sensitive. This study reduces calibration error impacts on speed calculation by optimizing the position reference of the bottom-edge center point of detection boxes, while simplifying the coordinate mapping process based on highway scenario characteristics. The model ultimately outputs instantaneous speed sequences between consecutive frames, providing data support for subsequent traffic state analysis.

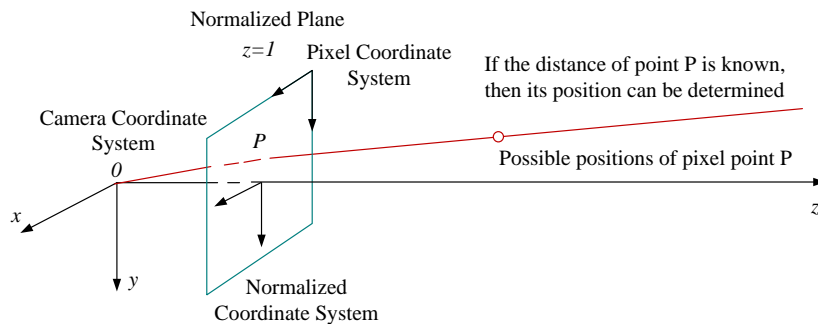


Figure 6. Pixel coordinate conversion diagram.

3.4.3. Vehicle Speed Measurement

Through coordinate transformation, the pixel coordinates $u_i(t)$ and $u_i(t - \Delta t)$ are projected onto the world coordinate system to derive the physical displacement vector S_i , as formalized in Equation (7). The Euclidean norm $\|S_i\|$, measured in meters, quantifies the actual distance traveled by the target vehicle between consecutive frames. Vehicle speed v_i is then computed using the relationship:

$$S_i = \varphi(a, b, c) \cdot u_i(t) - \varphi(a, b, c) \cdot u_i(t - \Delta t), \quad (7)$$

$$v_i = \frac{\|S_i\|}{\Delta t} = \frac{\|\varphi(a, b, c) \cdot u_i(t) - \varphi(a, b, c) \cdot u_i(t - \Delta t)\|}{\Delta t}, \quad (8)$$

where Δt denotes the fixed time interval between frames (in seconds), determined as the reciprocal of the video frame rate. For instance, in highway surveillance systems operating at 25 frames per second (fps), Δt equals 1/25 seconds. This framework leverages prior camera calibration results to bypass complex real-time geometric computations, ensuring efficient and robust speed estimation while maintaining compatibility with the assumptions outlined in Section 3.4.1.

For a vehicle trajectory spanning m consecutive frames, instantaneous speeds v_1, v_2, \dots, v_{m-1} between adjacent frames are derived from sequential displacement vectors. As formalized in Equations (9)-(11).

$$v_1 = \frac{\|S_1\|}{\Delta t} = \frac{\|\varphi(a, b, c) \cdot u_2(t) - \varphi(a, b, c) \cdot u_1(t)\|}{\Delta t}, \quad (9)$$

$$v_2 = \frac{\|S_2\|}{\Delta t} = \frac{\|\varphi(a, b, c) \cdot u_3(t) - \varphi(a, b, c) \cdot u_2(t)\|}{\Delta t}, \quad (10)$$

$$v_{m-1} = \frac{\|S_{m-1}\|}{\Delta t} = \frac{\|\varphi(a, b, c) \cdot u_m(t) - \varphi(a, b, c) \cdot u_{m-1}(t)\|}{\Delta t}, \quad (11)$$

Therefore, the target vehicle's average speed \bar{v} across the m -frame sequence is then computed through temporal aggregation:

$$\bar{v} = \frac{\sum_{i=1}^{m-1} v_i}{m-1}, \quad (12)$$

This multi-frame averaging strategy mitigates instantaneous measurement noise and transient tracking errors, enhancing speed detection robustness. The framework aligns with the motion continuity assumption by treating piecewise linear displacements as approximations of continuous trajectories.

3.5 Trajectory Prediction

For trajectory prediction, we employ a Bi-LSTM model, which is highly effective in capturing long-range dependencies and complex movement patterns of road users, as shown in Figure 7. LSTM, known for their attention mechanisms, provide a robust solution for modeling the temporal and spatial dynamics of object trajectories.

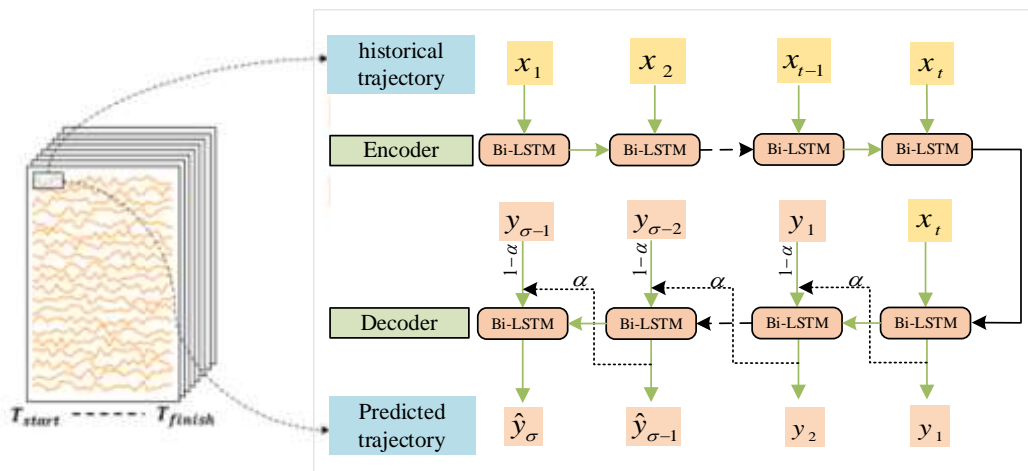


Figure 7. Bi-LSTM network architecture.

Currently, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) modules are commonly used to capture the memory retention and partial forgetting characteristics of vehicle position information during operation [50]. These models are designed to predict trajectories by unfolding along the time dimension. Through the internal state memory of the network, short-term prediction results from previous time steps are stored and fed into the current prediction module. This ensures that each output (i.e., risk level) in the network is a function of the previous state, improving both the continuity and efficiency of predictions.

LSTM employs a gated output mechanism with three gates (input gate, forget gate, and output gate) and two states (Cell State and Hidden State). The core of the LSTM is the Cell State, which functions as the network's memory, propagating information through the sequence. This is

represented by the direct path from (C_{t-1}) to (C_t) , enabling the model to carry essential information over long time steps, as shown in Figure 8.

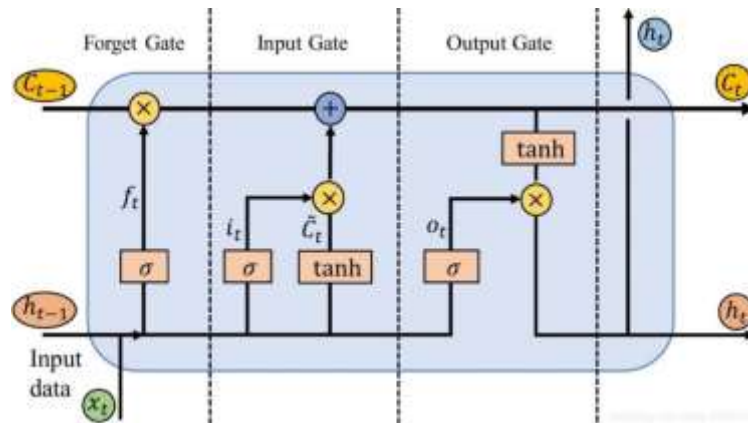


Figure 8. LSTM modules.

The Memory Cell (as shown in Figure 9) receives two inputs: the previous output value h_{t-1} and the current input value x_t . These two parameters first pass through the forget gate, where the information to be discarded, f_t (i.e., low-weight information), is determined. Next, the inputs are processed through the input gate, which determines the information to be updated, i_t (i.e., high-weight information compared to the previous cell). At this stage, the current Cell State \tilde{C}_t (a candidate vector, serving as an intermediate variable to store the information of the current Cell State) is also generated. The outputs of the forget gate and the input gate (i.e., f_t , i_t , \tilde{C}_t) are then combined. Specifically, the previous Cell State C_{t-1} is multiplied by the activation value f_t (the information to be forgotten), while the current Cell State \tilde{C}_t is multiplied by the activation value i_t (the information to be remembered). These two components are added together to produce the long-term state C_t and short-term state h_t , which are stored and serve as inputs for the next neuron. The detailed calculation process is as follows:

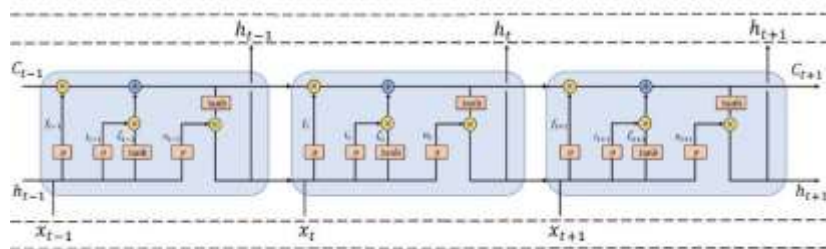


Figure 9. LSTM memory cell processing workflow.

(1) Forget Gate

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f), \quad (13)$$

(2) Input Gate

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i), \quad (14)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c), \quad (15)$$

The equation for the Cell State at time t (long-term state) is:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \quad (16)$$

(3) Output Gate

$$\sigma_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) , \quad (17)$$

$$h_t = \sigma_t \cdot \tanh(C_t), \quad (18)$$

f : Activation function of the gate; g : Activation function of the Cell input; h : Activation function of the Cell output.

In a unidirectional LSTM, the predicted output at the next time step is influenced only by inputs from previous time steps. However, during feature extraction, valuable information may be lost. In many cases, predictions are affected not only by past inputs but also by future ones. Incorporating both directions can lead to more accurate predictions. This study utilizes BI-LSTM (Bidirectional LSTM) to capture data features from both directions. Since future data also plays a role in prediction, a backward recurrent neural network (RNN) is introduced to complement the traditional forward RNN, which only processes historical data. By combining the forward and backward networks, both past and future data can be leveraged. The forward network computes the hidden vector \vec{h} in the forward direction, while the backward network computes the hidden vector \vec{h} in the reverse direction [19]. These two hidden vectors are then combined as shown in Equation (19). Figure 3 illustrates an unfolded Bidirectional LSTM, and Figure 10 shows the BI-LSTM model structure.

- In the forward network layer, computations are performed sequentially from the beginning to the end, yielding the forward hidden state outputs for each time step.
- In the backward network layer, the sequence is processed in reverse, from the end to the beginning, producing the backward hidden state outputs at each time step.
- Finally, the outputs from both the forward and backward layers are combined to provide the comprehensive output for each time step. This dual-direction approach ensures that the model captures richer temporal dependencies, resulting in more accurate predictions.

$$y_t = \vec{h} + \vec{h}, \quad (19)$$

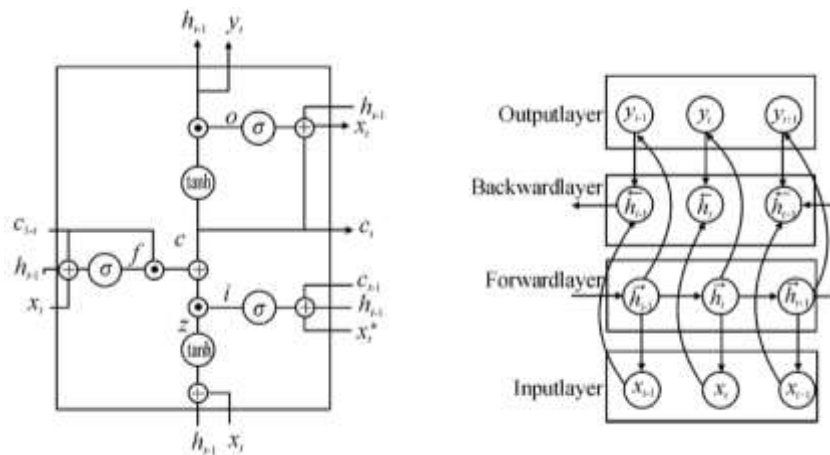


Figure 10. Bi-LSTM model schematic diagram.

3.6 Collision Risk Estimation

Collision risk estimation involves evaluating the likelihood of a collision based on predicted trajectories. This is done by computing the intersection points of predicted paths and determining the proximity and timing of these intersections. Based on the predicted trajectories, the Predictive Collision Risk Area (PCRA) [51] is estimated using confidence intervals for speed and angle. By combining these confidence intervals, the PCRA is represented as an area reflecting the predicted collision risk. In this experiment, the trajectories of objects were predicted for approximately 1, 2, and 3 seconds into the future. Therefore, multiple PCRA are generated for 1, 2, and 3 seconds, as illustrated in Figure 11.

The PCRA levels are then defined to measure the severity of risk based on the degree of overlap between the PCRA of vehicles and pedestrians, as follows:

1. Danger: When the PCRA of vehicles and pedestrians overlap 1 second into the future.
2. Warning: When the PCRA overlap 2 seconds into the future.
3. Caution: When the PCRA overlap 3 seconds into the future.
4. Relatively Safe: When there is no overlap between the PCRA.

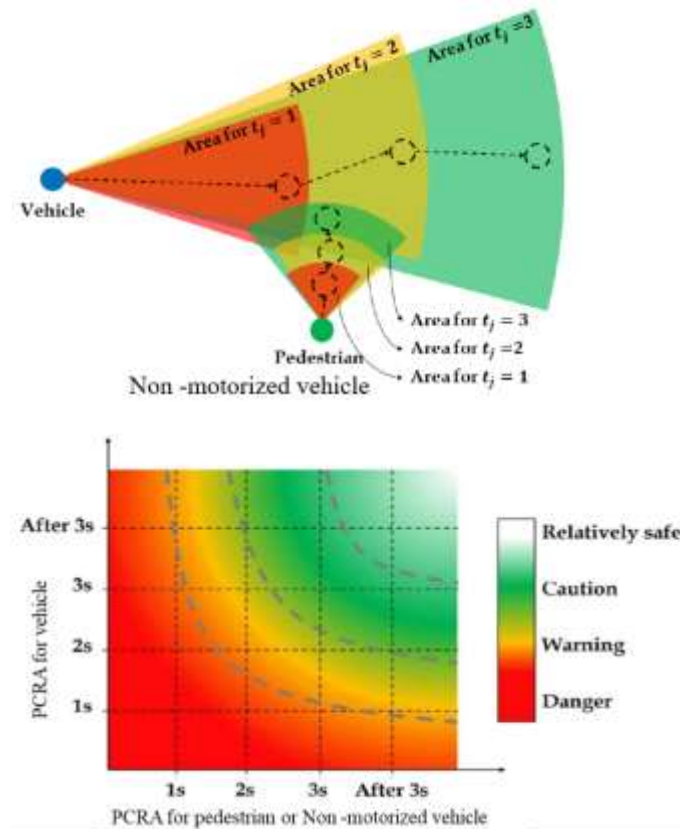


Figure 11. PCRA levels.

4. Experimental Analysis

4.1. Experimental Setup and Model Training

The experiments were conducted on a Windows 10 64-bit system equipped with an NVIDIA GeForce RTX3080ti GPU (24GB VRAM) and 64GB RAM. The software stack included CUDA V11.2 and OpenCV 4.5.3 to support deep learning frameworks and image processing tasks.

4.1.1. Dataset Annotation

A custom dataset comprising 30,000 annotated images was developed, covering diverse traffic scenarios with variations in lighting, viewing angles, and occlusions. Annotation guidelines included:

(1) Bounding Box Criteria: Tightly fitted rectangular boxes around vehicle contours, excluding instances with >50% occlusion, unidentifiable vehicle types, or objects smaller than 10×10 pixels.

(2) Augmentation Strategy: Synthetic data enrichment via geometric transformations (scaling, cropping, rotation) and photometric adjustments (exposure, saturation) to enhance model robustness against environmental variability..

4.1.2 Dataset Training

During model training, the dataset was divided into an 80:20 ratio for training and testing. Data augmentation techniques were applied to enhance the model's generalization ability, including Mosaic augmentation, random scaling, cropping, and arrangement. Additional augmentations included random rotation, random exposure, and saturation adjustment. The initial learning rate was set to 0.001 and adjusted to 0.0005 after 40,000 iterations to optimize model convergence. Training was conducted for a total of 50,000 iterations, with input image dimensions of 416×416 and a batch size of 8. Furthermore, to better adapt to traffic scene characteristics, additional augmentation methods such as Random_perspective, Mixup, HSV adjustments, Flipud, and Fliplr were employed to improve model accuracy and generalization performance.

4.2 Selection of Evaluation Metrics

Model performance was quantified using standard object detection metrics:

(1) Intersection over Union (*IoU*): Measures localization accuracy between predicted and ground-truth bounding boxes. The calculation formula is shown as Equation (23).

$$IoU = \frac{Detection\ Result \cap Ground\ Truth}{Detection\ Result \cup Ground\ Truth}, \quad (20)$$

(2) Precision-Recall Analysis (PR): In object detection, Precision and Recall are key evaluation metrics used to assess model performance. Precision measures the proportion of detected objects that are actual targets, while Recall represents the proportion of real targets successfully identified by the model. These metrics are calculated using the following formulas:

$$Precision = \frac{TP}{TP + FP}, \quad (21)$$

$$Recall = \frac{TP}{TP + FN}, \quad (22)$$

where TP (True Positive) represents correctly detected targets, FP (False Positive) denotes incorrectly identified objects, and FN (False Negative) refers to missed targets.

(3) Average Precision (*AP*): To evaluate a model's detection performance, a Precision-Recall (PR) curve is plotted based on computed Precision and Recall values. Average Precision (*AP*), a widely used performance metric, is derived by averaging Precision values across the PR curve. To enhance accuracy, the PR curve is smoothed, and the area under the curve (AUC) is calculated using integral methods to obtain the final *AP* score. The calculation formula is shown as Equation (23).

$$AP = \int_0^1 P_{smooth}(r)dr, \quad (23)$$

(4) mean Average Precision (*mAP*): Another critical performance metric in object detection is *mAP* (mean Average Precision). It is determined by calculating the *AP* (Average Precision) for each category and then computing the mean across all categories. This metric provides a comprehensive evaluation of the model's detection accuracy across different object classes.. The calculation formula is shown in Equation (24).

$$mAP = \frac{\sum AP_i}{n}, \quad i = 1, 2, \dots, n, \quad (24)$$

5 Results and Discussion

5.1 Detection and Tracking Performance

The algorithm proposed in this paper consists of several improved modules. To better assess the impact of each module on the model, ablation experiments were conducted on the same dataset. The results are shown in Table 1.

Table A1. Ablation results.

Type	Fps(frame/s)	P-value(%)	R(%)	mAp50 (%)	mAp50-95(%)
------	--------------	------------	------	-----------	-------------

YOLOv8n	256	80.2	84.6	92.5	68.3
+ RepLayer	243	82.5	85.3	93.3	72.6
+ GIoU	226	84.6	86.3	94.6	75.8
+ GAM	182	90.6	89.3	95.6	79.8
+ ReContext	165	92.3	90.7	96.7	84.5

As shown in Table 1, although the processing efficiency has decreased due to model improvements, the P-value, R, mAP50, and mAP50-95 have increased by 12.1%, 6.1%, 4.2%, and 16.2%, respectively.

5.2 Trajectory Prediction Accuracy

To further validate the accuracy of the trajectory prediction algorithm proposed in this paper, the target trajectory detected by the radar-vision fusion device is used as ground truth. A specific target's trajectory was selected for precision comparison and analysis. Additionally, the prediction performance for the trajectories of vehicles, pedestrians, and non-motorized vehicles was compared with the CNN-LSTM and LSTM models. The comparative results for the X and Y axis trajectories are shown in Figure 12.

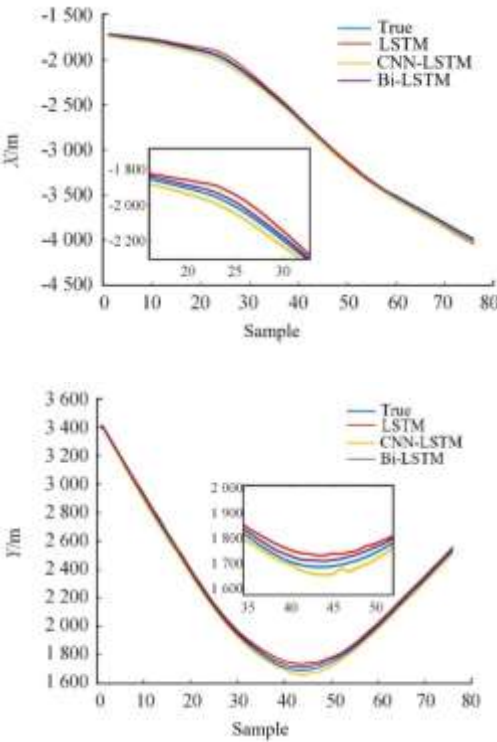


Figure 12. Trajectory tracking performance of different models.

From Figure 12, it can be observed that the predicted trajectories of the target by the three models maintain the same trend as the actual trajectories. However, the model constructed in this paper achieves higher accuracy in single-step predictions of the target trajectory and demonstrates excellent performance as the prediction horizon increases. As shown in the subplots of Figure 14, when the target performs maneuvers (with significant variations in X and Y), the model maintains a smaller prediction error and outperforms both the LSTM and CNN-LSTM models. This highlights the ability of the bidirectional self-attention mechanism to extract features at different time steps, thereby improving prediction performance.

Further analysis is conducted on trajectory prediction for different time horizons. Figure 13 illustrates that the longer the prediction horizon, the lower the prediction accuracy. This is because

driving behavior is inherently unpredictable, and as the prediction time increases, the error also grows. Moreover, in sequence prediction tasks, neural networks tend to accumulate errors over time, leading to larger deviations between the predicted trajectory points further in the sequence and the vehicle's actual trajectory.

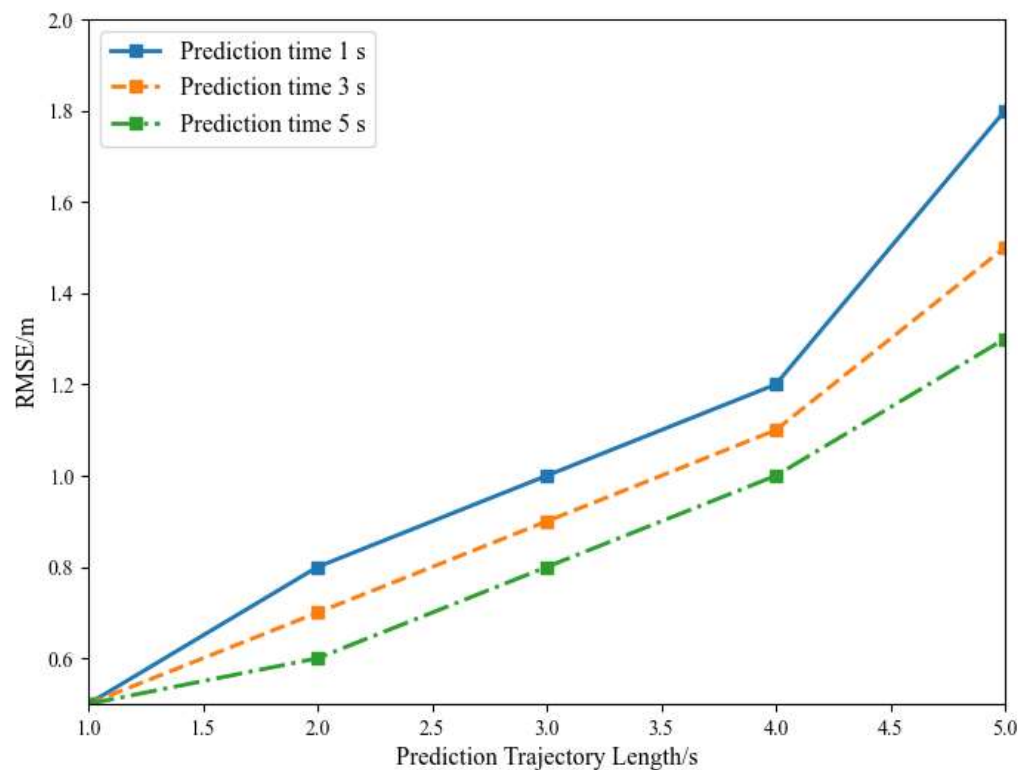


Figure 13. The impact of different prediction trajectory lengths on prediction accuracy.

5.3 System Performance

To further validate the effectiveness of the early warning system, a proactive safety warning system for unsignalized intersections was developed as part of the Science and Technology Safety Demonstration Project on the G310 highway in the Donghai section of Lianyungang as shown in Figure 14. This system integrates video surveillance, real-time data analysis, and digital twin technology to automatically detect incidents such as traffic accidents, vehicle anomalies, and congestion. It immediately triggers alerts on the interface and uploads the results to the management platform in real time. The front-end warning system predicts risks based on the movement of vehicles, pedestrians, and non-motorized vehicles on both main and side roads. It issues variable message board alerts to drivers on the main road and provides localized rapid warnings through audible and visual alarms for pedestrians and non-motorized vehicles on side roads. To verify the effectiveness of trajectory prediction and the warning system, warning logs and video records from September 1 to September 30, 2024, were analyzed. A total of 120 alerts involving vehicles, pedestrians, and non-motorized vehicles were recorded. Through manual comparison and analysis, the warning system achieved an accuracy rate of 97%.



Figure 14. Proactive safety warning system for unsignalized intersections.

6 Conclusions

The background and problem statement sections underscore the critical safety challenges at unsignalized intersections and the limitations of existing solutions. By leveraging state-of-the-art object detection, tracking, and trajectory prediction technologies, this research aims to develop a proactive collision warning system that enhances the safety of unsignalized intersections. This article proposes an improved YOLOv8+DeepSORT-based collision risk warning system for unsignalized intersections, integrating vehicle and pedestrian detection, tracking, and trajectory prediction to enhance traffic safety. The main conclusions are as follows:

(1) YOLOv8 is employed as the primary detector, optimized with the RepLayer module, GIoU loss, and Global Attention Mechanism (GAM) to improve object detection accuracy, especially in complex traffic environments. The model demonstrated high detection precision with mAP50 exceeding 96.7% and significant improvements in detecting multi-scale objects due to the ReContext gradient composition feature pyramid.

(2) Deep SORT was modified to enhance tracking robustness, reducing identity switches caused by occlusions and ensuring continuous monitoring of vehicles, pedestrians, and non-motorized users. This allows for reliable multi-target tracking, which is essential for accurate collision prediction at intersections.

(3) A Bi-LSTM network was implemented for trajectory prediction, effectively capturing long-range dependencies and predicting future movements with high precision. The Predictive Collision Risk Area (PCRA) approach was used to assess collision risk levels (danger, warning, and caution), based on the spatial overlap of predicted paths.

(4) In the experimental setup, the dataset used for training the model consisted of 30,000 images, annotated with bounding boxes around vehicles, pedestrians, and non-motorized vehicles. Data augmentation techniques such as Mosaic, Random_perspective, Mixup, HSV adjustments, Flipud, and Fliplr were applied to enrich the dataset and improve model robustness.

(5) In field tests on the G310 highway project, the system achieved 97% accuracy in collision risk prediction across 120 recorded events involving vehicles and pedestrians. The system's reliable performance in real-time collision warnings demonstrates its applicability and effectiveness in reducing accident risks at unsignalized intersections.

This system offers a scalable and effective solution for traffic safety, with significant potential for further enhancement through multi-sensor fusion technology. By integrating radar and video data, the system can complement each other to improve accuracy under diverse environmental conditions. In the future, this technology holds great promise for enhancing safety at unsignalized intersections, a critical challenge in urban traffic management. The integration of multiple sensors will enable more precise detection of potential collisions and enable timely warnings, thereby contributing to a substantial improvement in traffic safety and reducing the risk of accidents at high-risk, uncontrolled intersections.

Author Contributions: methodology, Y.B.; software, Q.Y and S.W.; validation, Y.L.; writing—original draft preparation, Z.L.; funding acquisition, Y.B and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National R&D Program of China (Grant No. 2023YFB2504704); The National Natural Science Foundation of China (52208424); The Natural Science Foundation of Chongqing (2022NSCQ-MSX1939); Chongqing Municipal Education Commission Foundation (KJQN202300728); Chongqing Talent Innovation Leading Talent Project (CQYC20210301505); The Key Research and Development Program of Guangxi, China, (Grant No. AB21196034).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We sincerely appreciate the invaluable support of our colleagues throughout the data acquisition process, as well as their assistance in proofreading and reviewing this work. Their contributions have significantly enhanced the quality and accuracy of this study. Additionally, we extend our heartfelt gratitude to the editor and reviewers for their insightful suggestions and constructive feedback, which have greatly contributed to improving this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fernando, D.M.; Tennakoon, S.U.; Samaranayake, A.N.; Wickramasinghe, M. Characteristics of Road Traffic Accident Casualties Admitted to a Tertiary Care Hospital in Sri Lanka. *Forensic Sci Med Pathol* **2017**, *13*, 44–51, doi:10.1007/s12024-016-9828-3.
2. Peličić, D.; Ristić, B.; Radević, S. EPIDEMIOLOGY OF TRAFFIC TRAUMATISM. *Sanamed* **2024**, *19*, doi:10.5937/sanamed0-51914.
3. Yang, P.; Yang, R.; Luo, Y.; Zhang, Y.; Hu, M. Hospitalization Costs of Road Traffic Injuries in Hunan, China: A Quantile Regression Analysis. *Accident Analysis & Prevention* **2024**, *194*, 107368, doi:10.1016/j.aap.2023.107368.
4. Song, L.; (David) Fan, W. Exploring Truck Driver-Injury Severity at Intersections Considering Heterogeneity in Latent Classes: A Case Study of North Carolina. *International Journal of Transportation Science and Technology* **2021**, *10*, 110–120, doi:10.1016/j.ijtst.2020.12.006.
5. Shao, Y.; Luo, Z.; Wu, H.; Han, X.; Pan, B.; Liu, S.; Claudel, C.G. Evaluation of Two Improved Schemes at Non-Aligned Intersections Affected by a Work Zone with an Entropy Method. *Sustainability* **2020**, *12*, 5494, doi:10.3390/su12145494.
6. Sharafeldin, M.; Farid, A.; Ksaibati, K. Examining the Risk Factors of Rear-End Crashes at Signalized Intersections. *Journal of Transportation Technologies* **2022**, *12*, 635, doi:10.4236/jtts.2022.124037.
7. Singh, H.; Kathuria, A. Analyzing Driver Behavior under Naturalistic Driving Conditions: A Review. *Accident Analysis & Prevention* **2021**, *150*, 105908, doi:10.1016/j.aap.2020.105908.

8. Schepers, P.; Hagenzieker, M.; Methorst, R.; van Wee, B.; Wegman, F. A Conceptual Framework for Road Safety and Mobility Applied to Cycling Safety. *Accident Analysis & Prevention* **2014**, *62*, 331–340, doi:10.1016/j.aap.2013.03.032.
9. Pawar, D.S.; Patil, G.R. Response of Major Road Drivers to Aggressive Maneuvering of the Minor Road Drivers at Unsignalized Intersections: A Driving Simulator Study. *Transportation Research Part F: Traffic Psychology and Behaviour* **2018**, *52*, 164–175, doi:10.1016/j.trf.2017.11.016.
10. Luo, Z.; Shi, H.; Liu, W.; Jin, Y. HMM-Based Traffic Situation Assessment and Prediction Method | Proceedings | Vol , No. **2020**.
11. Cheng, H.T.; Shan, H.; Zhuang, W. Infotainment and Road Safety Service Support in Vehicular Networking: From a Communication Perspective. *Mechanical Systems and Signal Processing* **2011**, *25*, 2020–2038, doi:10.1016/j.ymssp.2010.11.009.
12. Advancing Traffic Safety through the Safe System Approach: A Systematic Review. *Accident Analysis & Prevention* **2024**, *199*, 107518, doi:10.1016/j.aap.2024.107518.
13. Yuan, T.; Da Rocha Neto, W.; Rothenberg, C.E.; Obraczka, K.; Barakat, C.; Turletti, T. Machine Learning for Next-Generation Intelligent Transportation Systems: A Survey. *Transactions on Emerging Telecommunications Technologies* **2022**, *33*, e4427, doi:10.1002/ett.4427.
14. Geng, M.; Cai, Z.; Zhu, Y.; Chen, X.; Lee, D.-H. Multimodal Vehicular Trajectory Prediction With Inverse Reinforcement Learning and Risk Aversion at Urban Unsignalized Intersections. *IEEE Transactions on Intelligent Transportation Systems* **2023**, *24*, 12227–12240, doi:10.1109/TITS.2023.3285891.
15. Mihalj, T.; Li, H.; Babić, D.; Lex, C.; Jeudy, M.; Zovak, G.; Babić, D.; Eichberger, A. Road Infrastructure Challenges Faced by Automated Driving: A Review. *Applied Sciences* **2022**, *12*, 3477, doi:10.3390/app12073477.
16. A Review of Different Components of the Intelligent Traffic Management System (ITMS) Available online: <https://www.mdpi.com/2073-8994/15/3/583> (accessed on 31 October 2024).
17. Luo, Z.; Bi, Y.; Lei, Q.; Li, Y.; Song, L. Method for Identifying and Alerting to Operational Risks of En-Route Vehicles on Arterial Road. In Proceedings of the Ninth International Conference on Electromechanical Control Technology and Transportation (ICECTT 2024).
18. YOLOv8: Advancements and Innovations in Object Detection | SpringerLink Available online: https://link.springer.com/chapter/10.1007/978-981-97-1323-3_1 (accessed on 31 October 2024).
19. Wojke, N.; Bewley, A.; Paulus, D. Simple Online and Realtime Tracking with a Deep Association Metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP); September 2017; pp. 3645–3649.
20. DDSS: Driver Decision Support System Based on the Driver Behaviour Prediction to Avoid Accidents in Intelligent Transport System. *International Journal of Cognitive Computing in Engineering* **2024**, *5*, 1–13, doi:10.1016/j.ijcce.2023.12.001.
21. Predictions for COVID-19 with Deep Learning Models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals* **2020**, *140*, 110212, doi:10.1016/j.chaos.2020.110212.
22. Understanding the Potential of Emerging Digital Technologies for Improving Road Safety. *Accident Analysis & Prevention* **2022**, *166*, 106543, doi:10.1016/j.aap.2021.106543.
23. Wang, Y.; Yang, X.; Liang, H.; Liu, Y. A Review of the Self-Adaptive Traffic Signal Control System Based on Future Traffic Environment. *Journal of Advanced Transportation* **2018**, *2018*, 1096123, doi:10.1155/2018/1096123.
24. Cicchino, J.B. Effects of Forward Collision Warning and Automatic Emergency Braking on Rear-End Crashes Involving Pickup Trucks. *Taylor & Francis* **2023**.
25. Effectiveness of Forward Collision Warning and Autonomous Emergency Braking Systems in Reducing Front-to-Rear Crash Rates. *Accident Analysis & Prevention* **2017**, *99*, 142–152, doi:10.1016/j.aap.2016.11.009.
26. Smart Cities: Advances in Research—An Information Systems Perspective. *International Journal of Information Management* **2019**, *47*, 88–100, doi:10.1016/j.ijinfomgt.2019.01.004.
27. Zhu, Z.; He, X.; Qi, G.; Li, Y.; Cong, B.; Liu, Y. Brain Tumor Segmentation Based on the Fusion of Deep Semantics and Edge Information in Multimodal MRI. *Information Fusion* **2023**, *91*, 376–387, doi:10.1016/j.inffus.2022.10.022.

28. Liu, Y.; Qi, Z.; Cheng, J.; Chen, X. Rethinking the Effectiveness of Objective Evaluation Metrics in Multi-Focus Image Fusion: A Statistic-Based Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2024**, *46*, 5806–5819, doi:10.1109/TPAMI.2024.3367905.
29. Hussain, M. YOLOv1 to v8: Unveiling Each Variant—A Comprehensive Review of YOLO. *IEEE Access* **2024**, *12*, 42816–42833, doi:10.1109/ACCESS.2024.3378568.
30. Enhancing Vehicle Detection in Intelligent Transportation Systems via Autonomous UAV Platform and YOLOv8 Integration. *Applied Soft Computing* **2024**, *164*, 112015, doi:10.1016/j.asoc.2024.112015.
31. Wang, H.; Liu, C.; Cai, Y.; Chen, L.; Li, Y. YOLOv8-QSD: An Improved Small Object Detection Algorithm for Autonomous Vehicles Based on YOLOv8. *IEEE Transactions on Instrumentation and Measurement* **2024**, *73*, 1–16, doi:10.1109/TIM.2024.3379090.
32. Utilizing YOLOv8 for Enhanced Traffic Monitoring in Intelligent Transportation Systems (ITS) Applications. *Digital Signal Processing* **2024**, *152*, 104594, doi:10.1016/j.dsp.2024.104594.
33. Adžemović, M.; Tadić, P.; Petrović, A.; Nikolić, M. Beyond Kalman Filters: Deep Learning-Based Filters for Improved Object Tracking 2024.
34. Hu, D.; Al Shafian, S. Segmentation and Tracking of Moving Objects on Dynamic Construction Sites | Proceedings | Vol, No. **2021**.
35. Zhao, J.; Chen, J. YOLOv8 Detection and Improved BOT-SORT Tracking Algorithm for Iron Ladles. In Proceedings of the Proceedings of the 2024 7th International Conference on Image and Graphics Processing; Association for Computing Machinery: New York, NY, USA, May 3 2024; pp. 409–415.
36. Mercorelli, P. Denoising and Harmonic Detection Using Nonorthogonal Wavelet Packets in Industrial Applications. *Jrl Syst Sci & Complex* **2007**, *20*, 325–343, doi:10.1007/s11424-007-9028-z.
37. He, Y.; Huang, P.; Hong, W.; Luo, Q.; Li, L.; Tsui, K.-L. In-Depth Insights into the Application of Recurrent Neural Networks (RNNs) in Traffic Prediction: A Comprehensive Review. *Algorithms* **2024**, *17*, 398, doi:10.3390/a17090398.
38. Machine Learning for Autonomous Vehicle's Trajectory Prediction: A Comprehensive Survey, Challenges, and Future Research Directions. *Vehicular Communications* **2024**, *46*, 100733, doi:10.1016/j.vehcom.2024.100733.
39. Korbmacher, R.; Tordeux, A. Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-Based Approaches. *IEEE Transactions on Intelligent Transportation Systems* **2022**, *23*, 24126–24144, doi:10.1109/TITS.2022.3205676.
40. Zhang, C.; Ni, Z.; Berger, C. Spatial-Temporal-Spectral LSTM: A Transferable Model for Pedestrian Trajectory Prediction. *IEEE Transactions on Intelligent Vehicles* **2024**, *9*, 2836–2849, doi:10.1109/TIV.2023.3285804.
41. Pereira, G.A.; Hussain, M. A Review of Transformer-Based Models for Computer Vision Tasks: Capturing Global Context and Spatial Relationships 2024.
42. Hussain, A.; Hussain, T.; Ullah, W.; Baik, S.W. Vision Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance Videos. *Computational Intelligence and Neuroscience* **2022**, *2022*, 3454167, doi:10.1155/2022/3454167.
43. Berhanu, Y.; Alemayehu, E.; Schröder, D. Examining Car Accident Prediction Techniques and Road Traffic Congestion: A Comparative Analysis of Road Safety and Prevention of World Challenges in Low-Income and High-Income Countries. *Journal of Advanced Transportation* **2023**, *2023*, 6643412, doi:10.1155/2023/6643412.
44. Predicting Real-Time Traffic Conflicts Using Deep Learning. *Accident Analysis & Prevention* **2020**, *136*, 105429, doi:10.1016/j.aap.2019.105429.
45. Enhancing Road Safety in Internet of Vehicles Using Deep Learning Approach for Real-Time Accident Prediction and Prevention. *International Journal of Intelligent Networks* **2024**, *5*, 212–223, doi:10.1016/j.ijin.2024.05.002.
46. Hegde, C.; Dash, S.; Agarwal, P. Vehicle Trajectory Prediction Using GAN. In Proceedings of the 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC); October 2020; pp. 502–507.

47. Pang, Y.; Kashiya, T.; Yabe, T.; Tsubouchi, K.; Sekimoto, Y. Development of People Mass Movement Simulation Framework Based on Reinforcement Learning. *Transportation Research Part C: Emerging Technologies* **2020**, *117*, 102706, doi:10.1016/j.trc.2020.102706.
48. Li, C.; Liu, Z.; Lin, S.; Wang, Y.; Zhao, X. Intention-Convolution and Hybrid-Attention Network for Vehicle Trajectory Prediction. *Expert Systems with Applications* **2024**, *236*, 121412, doi:10.1016/j.eswa.2023.121412.
49. Luo, Z.; Bi, Y.; Yang, X.; Li, Y.; Yu, S.; Wu, M.; Ye, Q. Enhanced YOLOv5s + DeepSORT Method for Highway Vehicle Speed Detection and Multi-Sensor Verification. *Front. Phys.* **2024**, *12*, doi:10.3389/fphy.2024.1371320.
50. Xing, L.; Liu, W. A Data Fusion Powered Bi-Directional Long Short Term Memory Model for Predicting Multi-Lane Short Term Traffic Flow. *IEEE Transactions on Intelligent Transportation Systems* **2022**, *23*, 16810–16819, doi:10.1109/TITS.2021.3095095.
51. Analyzing Vehicle–Pedestrian Interactions: Combining Data Cube Structure and Predictive Collision Risk Estimation Model. *Accident Analysis & Prevention* **2022**, *165*, 106539, doi:10.1016/j.aap.2021.106539.

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.