An Integrated Smart Road Stud-Based Vehicle Localization Method With Velocity Estimation

Keyin Wang, *Graduate Student Member, IEEE*, Guoqiang Mao, *Fellow, IEEE*, Xiaojiang Ren, *Member, IEEE*, Haoyuan Du, Baoqi Huang, *Senior Member, IEEE*, Tianxuan Fu, and Zhaozhong Zhang

***Abstract*—The integration of the global navigation satellite system (GNSS), an odometer, and an inertial navigation system (INS) holds significant potential for achieving high-precision vehicle localization. However, the GNSS is vulnerable to obstructions and jamming, and the odometer is unreliable in harsh road conditions. These factors can lead to cumulative positioning errors in GNSS-denied environments. To address these issues, a novel multisource information fusion-based vehicle localization method that integrates an onboard binocular camera, an INS, and smart road studs (SRSs) —Internet of Things (IoT) devices extensively used for road safety and data collection in intelligent transportation systems—is introduced. We construct a position measurement model directly in the camera coordinate system through an enhanced You Only Look Once 8th version (YOLOv8) algorithm for SRS detection, combined with binocular vision measurement and position transformations. Additionally, we propose a method to enhance vehicle localization accuracy by integrating vehicle speed without relying on additional hardware speed sensors. The vehicle's speed is estimated from the image sequences captured by the onboard camera using a deep neural network (DNN), named Speed-Net. The final navigation results are produced by fusing the SRS-aided positioning information, the estimated vehicle speed, and INS data through an error-state extended Kalman filter (ESEKF). Real-world experiments demonstrate the effectiveness of the proposed SRS/velocity/INS integrated vehicle localization method.**

***Index Terms*—Error-state extended Kalman filter (ESEKF), onboard camera, smart road studs (SRSs), vehicle localization, vehicle speed estimation.**

I. Introduction

ACCURATE and robust localization system is essential for connected and autonomous vehicles (CAVs), as it provides crucial information about the vehicle's precise location that is fundamental for safe navigation, collision avoidance, and efficient path planning [1], [2], [3].

Received 17 April 2025; revised 28 July 2025 and 8 January 2026; accepted 9 January 2026. Date of publication 19 January 2026; date of current version 26 March 2026. This work was supported by NSFC under Grant U21A20446. *(Corresponding author: Guoqiang Mao.)*

Keyin Wang, Xiaojiang Ren, Haoyuan Du, and Tianxuan Fu are with the Research Institute of Smart Transportation, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China (e-mail: keyinwang@stu.xidian.edu.cn; xjren@stu.xidian.edu.cn; 23013221187@stu.xidian.edu.cn; futianxuan@stu.xidian.edu.cn).

Guoqiang Mao is with the School of Transportation, Southeast University, Nanjing 210096, China (e-mail: g.mao@ieee.org).

Baoqi Huang is with the Engineering Research Center of Ecological Big Data, Ministry of Education, Inner Mongolia Key Laboratory of Wireless Networking and Mobile Computing, College of Computer Science, Inner Mongolia University, Hohhot 010021, China (e-mail: cshbq@imu.edu.cn).

Zhaozhong Zhang is with the School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China (e-mail: 71033@nchu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2026.3654932

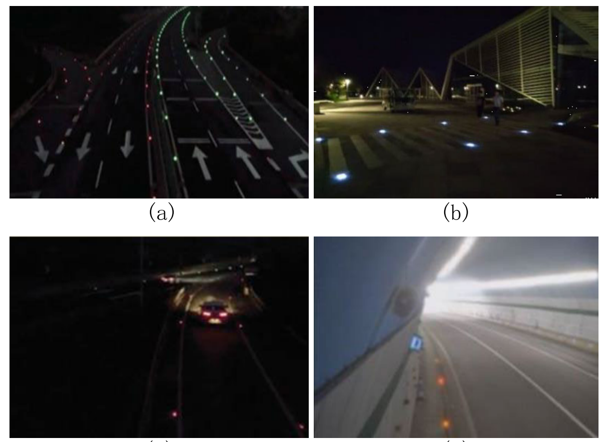


Fig. 1. Examples of SRS applications in different scenarios. (a) Highway, (b) crosswalk, (c) ramp-merge, and (d) tunnel.

In an open-sky environment, the integration of the global navigation satellite system (GNSS) and the inertial navigation system (INS) can provide reliable and accurate positioning for vehicles. Particularly, using real-time kinematic (RTK) technology, it can even achieve centimeter-level positioning accuracy [4], [5], [6], [7]. However, in GNSS-denied environments such as urban canyons, tunnels, and underground parking lots, GNSS signals are easily blocked or affected by multipath effects. Additionally, GNSS signals are susceptible to deliberate spoofing and jamming. These factors make the navigation methods relying on the GNSS unreliable and vulnerable [8].

To improve vehicle positioning accuracy in GNSS-denied environments, an effective approach is to incorporate observation information from other sensors, such as odometers [6], LiDAR [9], radar [10], and cameras [11], and so on. The positioning method incorporates additional observation information to mitigate the divergence of positioning errors, which is often effective during short-term GNSS outages, but cannot avoid the eventual collapse in location estimates when the GNSS is unavailable for a long time period. The positioning solution based on a high-definition (HD) map can reduce reliance on the GNSS to some extent. However, on the one hand, creating a large-scale HD map is costly, and, on the other hand, real-time map updates are challenging [12]. Unlike previous works, this study explores a high-precision vehicle localization method that does not rely on GNSS or HD maps, significantly enhancing the adaptability and robustness of the positioning system in complex environments, thereby supporting the further development of autonomous driving and intelligent transportation. The proposed work relies on the use of a new type of road-embedded Internet of Things (IoT) devices, termed smart road studs (SRSs). As a type of ubiquitously deployed IoT device, SRSs integrating light-emitting diodes (LEDs) and various sensors, such as temperature, humidity, light, vibration, and magnetic sensors, are increasingly being applied in intelligent transportation systems [13], [14]. Fig. 1 shows the actual deployment of SRSs in different scenarios, including highway, pedestrian crosswalk, ramp-merge area, and tunnel. Once SRSs are installed, their positions remain fixed and hence can be readily calibrated, which allows them to be used as a type of excellent auxiliary landmark devices to assist vehicle positioning. The SRSs can actively emit red/white/yellow light with controllable brightness, which can be detected by vehicle onboard cameras. Based on this, we propose a method that integrates SRS-aided positioning information with INS navigation information to achieve vehicle localization.

Accurate and real-time detection of SRSs is a prerequisite for SRS-aided localization. For fast-moving CAVs, SRSs are relatively small, with inconsistent brightness and often blurry backgrounds. These factors make SRS detection a challenging task. Traditional object detection methods based on handcrafted features struggle to overcome these challenges to achieve accurate and robust SRS detection [15], [16]. In recent years, deep neural network (DNN)-based object detection methods have been widely studied [17]. These methods can automatically learn high-level semantic features from complex visual inputs, making them more suitable for detecting small and visually ambiguous targets like SRSs. DNN-based object detectors can be broadly categorized into two groups: two-stage and one-stage detectors. Two-stage detectors, such as region-based convolutional neural network (R-CNN) [18], Fast R-CNN [19], and Faster R-CNN [20], first generate region proposals and then refine them through classification. In contrast, one-stage detectors like the single-shot multibox detector (SSD) [21] and the You Only Look Once (YOLO) series [22] predict object locations and categories in a single step, making them significantly faster. It is noteworthy that in the development of object detection algorithms, YOLO algorithms have gained popularity due to their ability to provide a favorable tradeoff between accuracy and inference speed, which is critical for real-time applications. Although newer versions such as YOLOv9 [23] and YOLO-World [24] have been proposed, their more complex network structures do not lead to significant improvements in the detection accuracy of SRSs. In contrast, YOLO 8th version (YOLOv8) [25] offers a more mature, stable, and deployment-friendly architecture. Therefore, we adopt YOLOv8 for SRS detection. We train the YOLOv8 model on our self-constructed SRS dataset, incorporating the small-object-friendly normalized Wasserstein distance (NWD) loss function [26] to enhance detection performance for SRSs.

Unlike other landmark-aided localization methods, where landmarks usually have unique identifiers, all SRSs employed in the proposed method are identical and do not possess identifiable features. To address this issue, we combine the navigation results from the INS to accurately determine the position of the SRS recognized by the camera. The detailed description of this approach is provided in Section III. In addition, we propose a method to improve vehicle localization accuracy by incorporating vehicle speed without relying on additional hardware-based speed sensors. Specifically, we introduce Speed-Net, a DNN that estimates the ego vehicle's speed directly from image sequences, built upon an enhanced version of DeepVO [27]. The estimated speed is then utilized to refine velocity measurement updates in the proposed filtering algorithm. This approach further improves the accuracy of vehicle localization without the need for additional hardware odometer, which is unreliable under harsh conditions, such as when the vehicle jumps or slips. An ESEKF is further used to fuse the vision-based positioning information from SRSs, the vehicle speed derived from image sequences, and the INS navigation data, achieving high-precision vehicle localization. Specifically, our main contributions are as follows.

1) An integrated camera, INS, and SRSs vehicle localization method based on ESEKF is proposed. In this method, we construct the position measurement directly in the camera coordinate system by combining binocular vision measurement and position transformations to correct the accumulated positioning errors of the INS. Compared to existing methods, this approach eliminates the reliance on GNSS and HD maps, achieving high-precision, continuous, and reliable vehicle localization.
2) A method to improve vehicle localization accuracy by incorporating vehicle speed without requiring additional hardware-based speed measurement devices is proposed. The vehicle speed is estimated from image sequences using Speed-Net. Compared to traditional odometer-based vehicle localization methods, this approach eliminates the need for additional hardware-wheeled odometer and is more robust to varying road conditions.
3) Real-world experiments demonstrate that the proposed SRS/velocity/INS vehicle localization method significantly improves positioning accuracy. The contribution of Speed-Net in improving positioning accuracy is confirmed through ablation experiments.

The rest of this article is organized as follows. Section II introduces related work. Section III gives an overview of the proposed vehicle localization method. Section IV provides the system setting and modeling. The experimental results are given in Section V. Finally, the conclusion and future work are drawn in Section VI.

II. Related Work

Considering the scope of our research, this section primarily focuses on vehicle localization methods utilizing onboard sensor perception, while excluding vehicle localization techniques based on wireless communication.

Simultaneous localization and mapping (SLAM) technology allows robots to construct maps of their surroundings and simultaneously determine their own position by utilizing onboard sensors like LiDAR, cameras, and inertial measurement units (IMUs), without relying on the GNSS. Abdelaziz and El-Rabbany [28] proposed a loosely coupled integration of the INS, LiDAR SLAM, and visual SLAM using an EKF. Experimental results confirmed that the multi-source coupled SLAM system outperforms other single-source positioning schemes in terms of accuracy. Niu et al. [29] proposed a robust INS-centric GNSS–visual–inertial SLAM system, where the IMU, visual, and GNSS measurements are tightly fused within a factor graph optimization framework. Experimental results validate that this method improves the positioning accuracy in GNSS-degraded environments. Wu and Zhao [30] proposed a complete SLAM system integrating the IMU, LiDAR, and GNSS based on the iterated ESEKF and factor graph. This method can effectively handle intermittent GNSS signals to achieve high-precision positioning. An effective method to improve the accuracy of SLAM systems is closed-loop detection. Closed-loop detection is a key part of back-end optimization in SLAM, reducing accumulated errors by recognizing locations previously visited by the robot. However, this is difficult to achieve in the highly dynamic vehicular driving environments, which limits the application of SLAM technology in vehicle localization.

HD map-based vehicle localization commonly utilizes the GNSS to achieve initial alignment to the global coordinate system, after which map-matching techniques combined with vision, LiDAR, or IMU data are adopted for continuous position updating [31]. Alternatively, GNSS measurements with noise can be refined by matching them with a digital map, leading to more accurate vehicle localization. Hansson et al. [32] proposed a hidden Markov model for matching a trajectory of noisy GNSS measurements to the road lanes to achieve lane-level localization. Niu et al. [11] proposed a lane-assisted vehicle positioning method. In this method, the lateral vehicle-to-lane distance is obtained through camera measurement and map-matching and used as a lateral distance measurement correction in GNSS-denied environments. Map-assisted localization methods heavily rely on the accuracy of the map itself. Moreover, the positioning accuracy can decrease in dynamic environments due to a lack of timely updates in HD map [33], [34].

Recently, some researchers have proposed integrated vehicle positioning methods that are aided by vision positioning. Wang et al. [35] proposed a vehicle positioning method that improves the positioning accuracy at traffic intersections by integrating the traffic light position information from an HD map with the navigation data from the INS. In this approach, traffic lights are detected by an onboard camera. Welzel et al. [36] proposed a landmark-based positioning method that uses traffic signs detected by a camera to correct the vehicle's position through a Bayesian filter, improving the positioning accuracy. Hu et al. [37] proposed a method to improve positioning accuracy for vehicles traveling on highways by using kilometer signs to correct errors of the INS. In this approach, the positions of kilometer signs are identified through image recognition, and the visual positioning results are fused with INS data using the EKF. The aforementioned vision positioning-based vehicle localization methods have a significant limitation: they can only correct the vehicle's position in specific scenarios with sufficient landmarks, such as at traffic intersections or near kilometer signs, which means they cannot provide accurate positioning across the entire area. Zhang et al. [38] proposed a GNSS-independent vehicle localization method that utilizes camera-based recognition of road semantics, such as lane lines and pole-like objects, with a lightweight HD map. This method constructs the measurement model of the filtering algorithm in the pixel coordinate system, without taking into account the depth information from visual measurements, and may lead to larger measurement errors.

In GNSS-denied environments, introducing a vehicle speed constraint can effectively suppress the velocity error divergence of the INS, thereby improving positioning accuracy [39]. The most commonly used method for vehicle speed estimation involves measuring the rotational speed of the wheels through wheel speed sensors installed on the wheel hubs. This odometer-based speed estimation method can achieve relatively accurate vehicle speed estimation under normal driving conditions. However, the odometer is susceptible to inaccuracies due to both tire wear and pressure changes, as well as harsh road conditions during vehicle operation [40]. de Araujo et al. [41] mounted the radar at the bottom of the vehicle with an inclined angle toward the ground, enabling it to measure the Doppler speed between the vehicle and the ground. The estimated vehicle speed from the onboard radar is then used to assist vehicle localization. Several studies [42], [43], [44] estimate vehicle speed from onboard IMU data and incorporate it as a constraint to mitigate INS error divergence, thus enhancing positioning accuracy. The main limitation of these methods is that the uncertain IMU measurement noise leads to low accuracy in speed estimation.

Previous studies mitigated error divergence by integrating information from other sensors; however, these approaches ultimately fail in the long-term absence of GNSS or HD maps. To address this gap, we propose an SRS-assisted vehicle localization solution that does not rely on GNSS or HD maps, achieving submeter-level positioning accuracy. To further enhance localization accuracy without the need for additional speed sensors, we estimate vehicle speed from image sequences and utilize it as a velocity constraint to mitigate the velocity error drift of the INS.

III. Overview of the Proposed Methodology

A. Coordinate Frames and Notations

Some coordinate systems are first defined for clearer explanation as shown in Fig. 2. Table I provides the names, notations, and specific roles of the coordinate systems used in this article.

The coordinate systems in the model of binocular vision include the *c*-frame, the *p*-frame, and the *im*-frame. Below, the subscripts l and r will, respectively, represent the left camera and the right camera of the binocular camera. The detailed explanations about these coordinate systems are provided in reference [45].

TABLE I
DESCRIPTION OF COORDINATE SYSTEMS

Coordinate System	Notation	Description
Inertial frame	*i*-frame	Global reference frame for measuring the absolute motion of the vehicle.
Earth frame	*e*-frame	Geocentric coordinate system fixed to the Earth, used for modeling geodetic positions.
Navigation frame	*n*-frame	Local-level frame pointing toward geographic north, east, and down. Used for local navigation.
Body frame	*b*-frame	Vehicle-fixed frame used for describing vehicle motion in its own local space.
Camera frame	*c*-frame	Coordinate system of the camera, used for describing the three-dimensional position of an object relative to the camera.
Pixel frame	*p*-frame	Frame used for describing the pixel coordinates in the image.
Image frame	*im*-frame	Frame used for describing the physical position on the image plane.

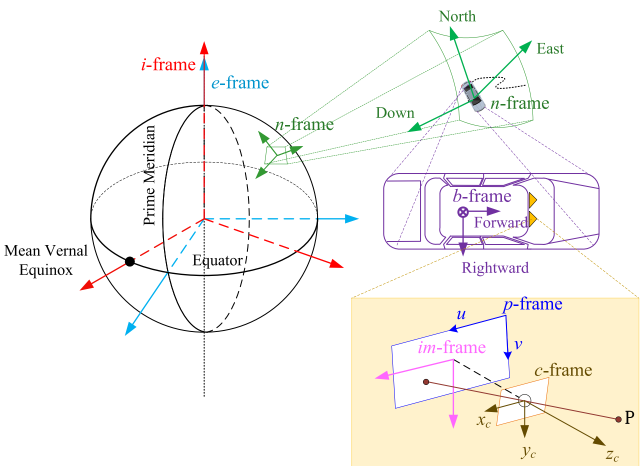


Fig. 2. Coordinate systems involved in the proposed vehicle localization method. The point P in the world frame, observed by the camera in the c-frame, is projected through the optical center onto the im-frame. Subsequently, the im-frame is discretized into the p-frame, enabling practical image-based localization. The IMU measurement data is based on the i-frame, with the transition from the i-frame to the e-frame accounting for Earth' s rotation to establish a geocentric context. The e-frame then transitions to the n-frame, aligning with the vehicle' s local navigation requirements. This n-frame is further refined by the b-frame, which reflects the vehicle' s physical orientation.

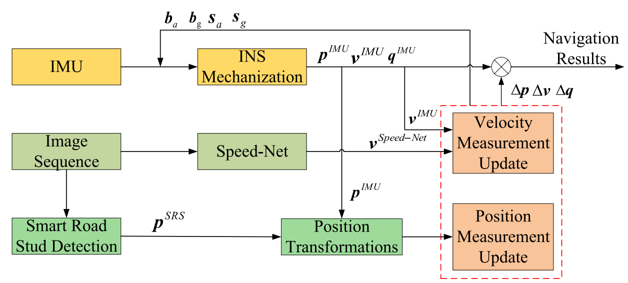


Fig. 3. Framework of the vehicle localization method. $\boldsymbol{p}^{\mathrm{IMU}}$, $\boldsymbol{v}^{\mathrm{IMU}}$, and $\boldsymbol{q}^{\mathrm{IMU}}$ represent the position, velocity, and attitude of the vehicle, respectively, as outputs from the INS mechanization. $\boldsymbol{v}^{\text{Speed-Net}}$ is the vehicle speed output from Speed-Net. $\boldsymbol{p}^{\mathrm{SRS}}$ represents the pixel position of the SRS detected by the object detection model. $\triangle\boldsymbol{p}$, $\triangle\boldsymbol{v}$, $\triangle\boldsymbol{q}$, $\boldsymbol{b}_a$, $\boldsymbol{b}_g$, $\boldsymbol{s}_a$, and $\boldsymbol{s}_g$ represent the errors in position, velocity, attitude, accelerometer bias, gyroscope bias, accelerometer scale factor, and gyroscope scale factor, respectively, as outputs of the filtering algorithm.

B. Framework of the Proposed Vehicle Localization Method

The overall framework of the proposed vehicle localization method is shown in Fig. 3. The ESEKF integrates the INS navigation information, the vehicle speed output from Speed-Net, and the vision-aided positioning information derived from SRSs to obtain the vehicle navigation results. The IMU-based INS mechanization is used for the state update in the proposed filtering algorithm. The measurement update of the filtering algorithm includes both the velocity measurement update and the position measurement update. The vehicle speed is obtained from the image sequences captured by the vehicle's onboard camera using the DNN-based speed estimation model, Speed-Net. The difference between the vehicle velocity estimated from the image sequences and the INS velocity output is taken as the velocity measurement of the ESEKF. The global positions of the SRSs are provided by the RTK device, forming the SRS position database stored in the vehicle. An object detection algorithm based on YOLOv8 is used to detect SRSs from images captured by the onboard camera. The following method is adopted to determine the location of the light-emitting SRS recognized by the binocular camera. First, an estimated position of the SRS can be obtained by combining the vehicle's position calculated by the INS with binocular vision measurement and coordinate transformations. The estimated position of the SRS is compared to the positions of SRSs in the SRS position database, and the closest match in the database to the estimated value identifies the recognized SRS. The accurate position of the SRS is then obtained from the database. If there are multiple SRSs in the image captured by the camera, we select the one closest to the vehicle. Using the SRS position obtained from the SRS position database and the vehicle position from the INS, the position of the SRS relative to the camera can be calculated through coordinate transformations. The difference between the calculated position of the SRS relative to the camera and the position obtained from binocular vision measurement is used as the position measurement of the ESEKF. After the prediction and measurement updates of the ESEKF, the state corrections are fed back to update the vehicle positions and compensate for the IMU errors.

In our proposed method, Speed-Net, an enhanced variant of DeepVO, is tailor-made to meet the specific objectives of this study. Unlike the original DeepVO, which estimates

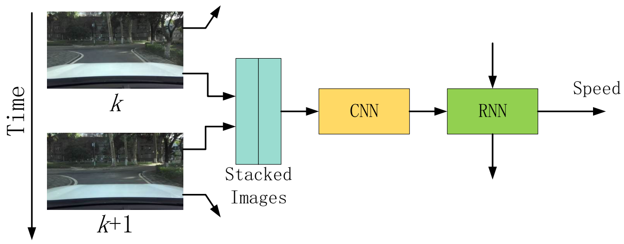


Fig. 4. Overview of the Speed-Net at a given timestep from k to $k+1$.

six-degree-of-freedom (6-DoF) poses, Speed-Net is modified to directly predict the vehicle's forward velocity. To capture the continuous nature of vehicle speed variations, we augment the original mean-squared-error (mse) loss function with a smoothness regularization term, promoting temporal consistency and continuity in the predicted velocity outputs. Speed-Net is trained on collected driving data and subsequently deployed to provide real-time velocity estimates, which serve as velocity constraints within the fusion-based localization framework.

Unlike other velocity acquisition methods used in fusion-based localization, this study employs a DNN to extract vehicle speed from image sequences. This approach not only eliminates the dependency on dedicated speed measurement hardware but also avoids the decline in speed estimation accuracy caused by vehicle jumps and slips under poor road conditions.

IV. SRS/Velocity/INS Integrated Vehicle Localization Method Based on the ESEKF

A. INS Mechanization and the IMU Error Model

The data measured by the IMU includes the specific force of the carrier relative to the i-frame measured by the accelerometers, which allows the estimation of the acceleration, and the rotational angular velocity of the carrier relative to the i-frame measured by the gyroscopes. INS mechanization describes the process of calculating the navigation information of the carrier, such as position, velocity, and attitude, from the raw data obtained from the IMU. The error differential equations of the INS mechanization can be expressed as follows [11]:

$$\dot{\boldsymbol{C}}_b^n = \boldsymbol{C}_b^n\left(\boldsymbol{\omega}_{ib}^b\times\right) - \left(\boldsymbol{\omega}_{\text{in}}^n\times\right)\boldsymbol{C}_b^n \tag{1}$$

$$\dot{\boldsymbol{v}}^n = \boldsymbol{C}_b^n\boldsymbol{f}^b - \left[\left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right)\times\right]\boldsymbol{v}^n + \boldsymbol{g}_p^n \tag{2}$$

$$\begin{cases} \dot{\varphi} = \dfrac{v_N}{R_M + h} \\ \dot{\lambda} = \dfrac{v_E}{(R_N + h)\cos\varphi} \\ \dot{h} = -v_D \end{cases} \tag{3}$$

where $\boldsymbol{C}_b^n$ is the direction cosine matrix from the b-frame to the n-frame, $\boldsymbol{\omega}_{ib}^b$ is the angular rate vector of the b-frame relative to the i-frame in the b-frame, $\boldsymbol{\omega}_{\text{in}}^n$ is the angular rate vector of the n-frame relative to the i-frame in the n-frame, $\boldsymbol{v}^n$ is the velocity in the n-frame, $\boldsymbol{f}^b$ is the specific force vector in the b-frame, $\boldsymbol{\omega}_{ie}^n$ is the angular rate vector of the e-frame relative to the i-frame in the n-frame due to rotation of the Earth, $\boldsymbol{\omega}_{en}^n$ is the angular rate vector of the n-frame relative to the e-frame in the n-frame, $\boldsymbol{g}_p^n$ is the normal gravity in the local position in the n-frame, φ, λ, and h denote the latitude, longitude, and altitude of the vehicle position, respectively, v_N, v_E, and v_D are the projection components of $\boldsymbol{v}^n$ in the North, East, and Down vertical directions of the n-frame, respectively, and R_M and R_N represent the radius of curvature in meridian and prime vertical, respectively. The symbol $\times$ represents the skew-symmetric matrix of a vector.

By performing perturbation analysis on position, velocity, and attitude, and neglecting the second-order error terms, we obtain the IMU error model as follows:

$$\dot{\boldsymbol{\phi}} = -\left(\boldsymbol{\omega}_{\text{in}}^n\times\right)\boldsymbol{\phi} + \delta\boldsymbol{\omega}_{\text{in}}^n - \delta\boldsymbol{\omega}_{ib}^n \tag{4}$$

$$\begin{aligned}\delta\dot{\boldsymbol{v}}^n = {} & \boldsymbol{C}_b^n\delta\boldsymbol{f}^b + \boldsymbol{C}_b^n\left(\boldsymbol{f}^b\times\right)\boldsymbol{\phi} - \left[\left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right)\times\right]\delta\boldsymbol{v}^n \\ & + \left(\boldsymbol{v}^n\times\right)\left(2\delta\boldsymbol{\omega}_{ie}^n + \delta\boldsymbol{\omega}_{en}^n\right) + \delta\boldsymbol{g}_p^n \end{aligned} \tag{5}$$

$$\begin{cases} \delta\dot{\varphi} = -\dfrac{v_N}{(R_M + h)^2}\delta h + \dfrac{1}{R_M + h}\delta v_N \\ \delta\dot{\lambda} = \dfrac{v_E\tan\varphi}{(R_N + h)\cos\varphi}\delta\varphi - \dfrac{v_E}{(R_N + h)^2\cos\varphi}\delta h \\ +\dfrac{1}{(R_N + h)\cos\varphi}\delta v_E \\ \delta\dot{h} = -\delta v_D \end{cases} \tag{6}$$

where $\boldsymbol{\phi}$ is the attitude error vector, $\delta\dot{\boldsymbol{v}}$ represents the velocity error vector, and $\delta\varphi$, $\delta\lambda$, and δh are the errors of φ, λ, and h, respectively.

The bias and the scale factor of the accelerometers and gyroscopes are modeled as first-order Gaussian–Markov processes as follows:

$$\begin{cases} \dot{\boldsymbol{b}}_g(t) = -\dfrac{1}{T_{gb}}\boldsymbol{b}_g(t) + \boldsymbol{\omega}_{gb}(t) \\ \dot{\boldsymbol{b}}_a(t) = -\dfrac{1}{T_{ab}}\boldsymbol{b}_a(t) + \boldsymbol{\omega}_{ab}(t) \\ \dot{\boldsymbol{s}}_g(t) = -\dfrac{1}{T_{gs}}\boldsymbol{s}_g(t) + \boldsymbol{\omega}_{gs}(t) \\ \dot{\boldsymbol{s}}_a(t) = -\dfrac{1}{T_{as}}\boldsymbol{s}_a(t) + \boldsymbol{\omega}_{as}(t) \end{cases} \tag{7}$$

where $\boldsymbol{b}_g(t)$ and $\boldsymbol{b}_a(t)$ are the gyroscopes' bias and accelerometers' bias, respectively, $\boldsymbol{s}_g(t)$ and $\boldsymbol{s}_a(t)$ are the gyroscopes' scale factor and the accelerometers' scale factor, respectively, T_{gb}, T_{ab}, T_{gs}, and T_{as} are correlation times, and $\boldsymbol{\omega}_{gb}$, $\boldsymbol{\omega}_{ab}$, $\boldsymbol{\omega}_{gs}$, and $\boldsymbol{\omega}_{as}$ are driving white noise.

Based on the error model, the integrated navigation system state equation can be expressed as follows:

$$\delta\dot{\boldsymbol{x}} = \boldsymbol{F}\delta\boldsymbol{x} + \boldsymbol{G}\boldsymbol{\omega} \tag{8}$$

where the error state vector $\delta\boldsymbol{x} = \left[\delta\boldsymbol{p}^n\ \delta\boldsymbol{v}^n\ \boldsymbol{\phi}\ \boldsymbol{b}_g\ \boldsymbol{b}_a\ \boldsymbol{s}_g\ \boldsymbol{s}_a\right]^{\mathrm{T}}$, $\delta\boldsymbol{p}^n = \left[\delta\varphi\ \delta\lambda\ \delta h\right]^{\mathrm{T}}$, the driving white noise vector $\boldsymbol{\omega} = \left[\boldsymbol{\omega}_v\ \boldsymbol{\omega}_\phi\ \boldsymbol{\omega}_{gb}\ \boldsymbol{\omega}_{ab}\ \boldsymbol{\omega}_{gs}\ \boldsymbol{\omega}_{as}\right]^{\mathrm{T}}$, $\boldsymbol{\omega}_v$, and $\boldsymbol{\omega}_\phi$ are the measurement white noises of the accelerometers and gyroscopes, respectively, $\boldsymbol{F}$ is the system state matrix, and $\boldsymbol{G}$ is the system noise driving matrix. $\boldsymbol{F}$ and $\boldsymbol{G}$ can be derived from (4) to (7).

Moreover, the measurement updates of the ESEKF come from the velocity measurement based on Speed-Net and the position measurement based on SRSs. In Sections IV-B and IV-C, we will elaborate on these two components in detail.

B. Speed-Net and Velocity Measurement

The vehicle speed estimation model, Speed-Net, proposed in this article, is an improvement upon DeepVO. The original intention of DeepVO is to estimate the 6-DoF poses of a camera from image sequences. We improve upon DeepVO to enable it to accurately and robustly estimate the vehicle speed. The original network output is modified to provide the vehicle's speed, as shown in Fig. 4. The network takes two consecutive frames as input, extracts features from the images using a convolutional neural network (CNN), and then predicts the vehicle's speed through a recurrent neural network (RNN). The CNN consists of nine convolutional layers of different sizes stacked together. Following the CNN, the RNN is composed of two long short-term memory (LSTM) layers, which can find and exploit correlations among images captured over long trajectories. At each time step, the RNN uses visual features extracted by the CNN to estimate the vehicle speed. This process continues as the camera moves and captures new images over time.

During vehicle motion, speed varies continuously. To reduce the output fluctuations of Speed-Net, we design a smoothing regularization term to be added to the mse loss during model training. The specific loss function is defined as follows:

$$L = \frac{1}{N}\sum_{i=1}^{N}(\hat{v}_i - v_i)^2 + \kappa\left[\frac{1}{N}\sum_{i=1}^{N}\left(\hat{v}_{i+1} - \hat{v}_i\right)^2\right] \quad (9)$$

where $1/N\sum_{i=1}^{N}(\hat{v}_i - v_i)^2$ represents the mse loss, with $\hat{v}_i$ as the predicted speed value, v_i as the true speed value, and N as the number of samples. Parameter κ is the regularization coefficient, and $1/N\sum_{i=1}^{N}(\hat{v}_{i+1} - \hat{v}_i)^2$ is the smoothing regularization term. The loss function aims to maintain the accuracy of speed prediction while reducing the fluctuations in the speed sequence, thereby achieving a smooth speed estimation.

Since the ground-truth vehicle speed for training Speed-Net is provided by RTK, the output of Speed-Net represents the speed of the vehicle at the GNSS receiver. Under normal driving conditions, the effects of jump and sideslip can be ignored, so the lateral and vertical speeds of the vehicle are zero. Therefore, the velocity vector in the b-frame can be expressed as follows:

$$\boldsymbol{v}^b = \left[\, v_x^b \; 0 \; 0 \,\right] \quad (10)$$

where v_x^b is the longitudinal speed of the vehicle.

The vehicle velocity vector can also be derived from the INS navigation data. The velocity transformation relationship between the IMU center and the GNSS antenna center can be described as follows:

$$\boldsymbol{v}^b = \boldsymbol{C}_n^b \boldsymbol{v}_I^n + \left(\boldsymbol{\omega}_{ib}^b \times\right) \boldsymbol{l}_{\text{gnss}} \quad (11)$$

where $\boldsymbol{C}_n^b$ is the direction cosine matrix from n-frame to b-frame, $\boldsymbol{v}_I^n$ is the velocity vector of the IMU center, and $\boldsymbol{l}_{\text{gnss}}$ is the lever arm offset vector from the GNSS antenna center to the IMU center in the b-frame.

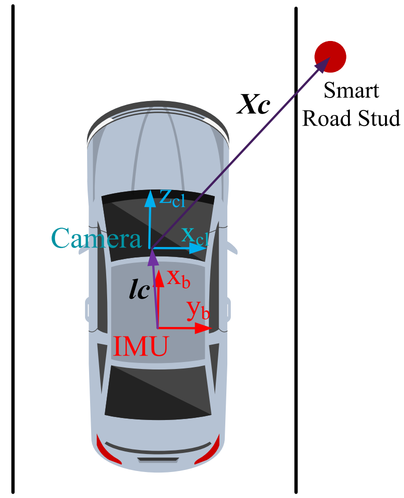


Fig. 5. Illustration of the relationship between the SRS, camera, and IMU. Parameter $\boldsymbol{l}_c$ refers to the lever arm from the left camera of the binocular camera to the IMU, which can be measured. Parameter $\boldsymbol{X}_c$ denotes the 3-D coordinate of the SRS in the c_l-frame, which can be obtained from the binocular visual measurement. The y-coordinate value of $\boldsymbol{X}_c$ can be replaced with the height of the camera from the ground, which can improve the measurement accuracy of the camera in the y-direction.

By using perturbation analysis and simultaneously neglecting the second-order error terms, the estimated vehicle velocity can be obtained as follows:

$$\hat{\boldsymbol{v}}^b = \boldsymbol{v}^b + \boldsymbol{C}_n^b \delta \boldsymbol{v}^n - \boldsymbol{C}_n^b \left(\boldsymbol{v}_I^n \times\right) \boldsymbol{\phi} - \left(\boldsymbol{l}_{\text{gnss}} \times\right) \delta \boldsymbol{\omega}_{ib}^b. \quad (12)$$

The measured velocity considering error can be written as follows:

$$\tilde{\boldsymbol{v}}^b = \boldsymbol{v}^b - \boldsymbol{n}_v \quad (13)$$

where $\boldsymbol{n}_v$ is the velocity measurement error.

The velocity measurement equation can be expressed as follows:

$$\begin{aligned}\delta z_v &= \hat{\boldsymbol{v}}^b - \tilde{\boldsymbol{v}}^b \\ &= \boldsymbol{H}_v \delta \boldsymbol{x} + \boldsymbol{n}_v\end{aligned} \quad (14)$$

where $\boldsymbol{H}_v$ is the velocity measurement matrix.

Considering (12), (13), (14), and $\delta \boldsymbol{\omega}_{ib}^b = \boldsymbol{b}_g + \text{diag}(\boldsymbol{\omega}_{ib}^b)\boldsymbol{s}_g + \boldsymbol{\omega}_\phi$, the velocity measurement matrix can be deduced as follows:

$$\begin{aligned}\boldsymbol{H}_v = \big[\, &\boldsymbol{0}_{3\times 3} \; \boldsymbol{C}_n^b \; -\boldsymbol{C}_n^b \left(\boldsymbol{v}_I^n \times\right) \; -\left(\boldsymbol{l}_{\text{gnss}} \times\right) \\ &\boldsymbol{0}_{3\times 3} \; -\left(\boldsymbol{l}_{\text{gnss}} \times\right) \text{diag}\left(\boldsymbol{\omega}_{ib}^b\right) \; \boldsymbol{0}_{3\times 3} \,\big]\end{aligned} \quad (15)$$

where $\boldsymbol{0}_{3\times 3}$ represents a 3×3 zero matrix, and $\text{diag}(\boldsymbol{\omega}_{ib}^b)$ represents the diagonal matrix generated from the elements of the vector $\boldsymbol{\omega}_{ib}^b$.

C. SRS Detection and Position Measurement

The prerequisite for using SRSs to assist in vehicle positioning is the accurate real-time detection of the SRSs. To enhance detection accuracy and speed, we implement YOLOv8 for detecting SRSs. Given that the SRSs are relatively small, especially for fast-moving CAVs, we replace the original loss function with the NWD loss function [26], a metric specifically designed for small objects. The NWD loss function measures the similarity between the predicted boxes and the ground-truth boxes using the Wasserstein distance, improving the model's robustness in detecting small targets. Finally, we train the YOLOv8 model on our self-built SRS dataset, achieving real-time detection of SRSs and obtaining the pixel coordinates of the SRSs in the images.

We utilize the SRS captured by the onboard camera to perform position measurement updates in the ESEKF. When the camera detects an SRS, the position of the SRS in the camera coordinate system can be calculated based on its known global coordinates, the vehicle' s position predicted by the INS, and the transformation between the vehicle and camera coordinate frames. The difference between the calculated position of the SRS relative to the camera and the position obtained from binocular vision measurement is used as the position measurement input for the ESEKF.

The positional relationship between the SRS, camera, and IMU is shown in Fig. 5. According to the camera model, the 3-D coordinate of the SRS in the c_l-frame can be expressed as follows:

$$\boldsymbol{X}_c = z_c \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \tag{16}$$

where f_x and f_y are the scale factors in the x and y directions, respectively, and (u_0, v_0) is the plane center of the image in the p_l-frame. These are the intrinsic parameters of the binocular camera, which can be obtained through calibration. The parameter (u_l, v_l) is the pixel coordinate of the geometric center of the detected SRS in the p_l-frame, which can be obtained from the output of the object detection model; z_c is the depth value of the observed object, which can be described as follows:

$$z_c = f_x \frac{d}{\Delta p} \tag{17}$$

where d is the distance between the left camera and the right camera of the binocular camera, which can be measured, Δp represents the disparity of the same SRS captured in the left and right images by the binocular camera, which is obtained using the semi-global block matching (SGBM) algorithm [46].

The positional relationship between the SRS, camera, and IMU can be represented as follows:

$$\boldsymbol{p}_I = \boldsymbol{p}_S - \boldsymbol{D}_R^{-1} \boldsymbol{C}_b^n \left(\boldsymbol{C}_c^b \boldsymbol{X}_c + \boldsymbol{l}_c \right) \tag{18}$$

where $\boldsymbol{p}_I$ is the position of the IMU, $\boldsymbol{p}_S$ is the position of the SRS, $\boldsymbol{D}_R^{-1}$ is used to transform the North–East–Down coordinate in the navigation frame into φ, λ, and h, and $\boldsymbol{C}_c^b$ is the direction cosine matrix from the c_l-frame to the b-frame, which can be obtained by prior calibration. $\boldsymbol{D}_R^{-1}$ can be expressed as follows:

$$\boldsymbol{D}_R^{-1} = \operatorname{diag}\left(\left[\frac{1}{R_M + h} \;\; \frac{1}{(R_N + h)\cos\varphi} \;\; -1 \right]^{\mathrm{T}} \right). \tag{19}$$

Fig. 6. Experimental platform.

From (18), the calculated coordinate of the SRS in the c_l-frame can be derived as follows:

$$\hat{\boldsymbol{X}}_c = \boldsymbol{C}_b^c \left[\hat{\boldsymbol{C}}_n^b \boldsymbol{D}_R \left(\boldsymbol{p}_s - \hat{\boldsymbol{p}}_I \right) - \boldsymbol{l}_c \right] \tag{20}$$

where the superscript ˆ denotes the calculated value considering errors, $\boldsymbol{C}_b^c$, $\boldsymbol{C}_n^b$, and $\boldsymbol{D}_R$ are the inverse matrices of $\boldsymbol{C}_c^b$, $\boldsymbol{C}_b^n$, and $\boldsymbol{D}_R^{-1}$, respectively. By using perturbation analysis, $\hat{\boldsymbol{C}}_n^b$ and $\hat{\boldsymbol{p}}_I$ can be expressed as follows:

$$\hat{\boldsymbol{C}}_n^b = \left[\boldsymbol{I} + (\boldsymbol{\phi} \times) \right] \boldsymbol{C}_n^b \tag{21}$$

$$\hat{\boldsymbol{p}}_I = \boldsymbol{p}_I + \boldsymbol{D}_R^{-1} \delta \boldsymbol{p}^n. \tag{22}$$

Ideally, the calculated coordinate of the SRS in the c_l-frame should coincide with the one measured by binocular vision. However, in practice, due to measurement errors in the camera and IMU, there may be deviations between the two. The position measurement vector is defined as the difference between the calculated coordinate of the SRS in the c_l-frame and the measured coordinate of the SRS by the binocular camera in the c_l-frame as follows:

$$\delta \boldsymbol{z}_p = \hat{\boldsymbol{X}}_c - \tilde{\boldsymbol{X}}_c \tag{23}$$

where $\tilde{\boldsymbol{X}}_c$ is the binocular visual measurement considering visual measurement noise, which can be expressed as follows:

$$\tilde{\boldsymbol{X}}_c = \boldsymbol{X}_c - \boldsymbol{n}_p \tag{24}$$

where $\boldsymbol{n}_p$ is the position measurement noise.

Substituting (18), (20), (21), (22), and (24) into (23), and simultaneously neglecting the second-order error terms, the position measurement equation can be written as follows:

$$\delta \boldsymbol{z}_p = -\boldsymbol{C}_b^c \boldsymbol{C}_n^b \delta \boldsymbol{p}^n - \boldsymbol{C}_b^c \left(\boldsymbol{C}_n^b \boldsymbol{D}_R \left(\boldsymbol{p}_s - \boldsymbol{p}_I \right) \times \right) \boldsymbol{\phi} + \boldsymbol{n}_p. \tag{25}$$

The position measurement matrix can be expressed as follows:

$$\boldsymbol{H}_p = \big[-\boldsymbol{C}_b^c \boldsymbol{C}_n^b \;\; \boldsymbol{0}_{3\times3} \;\; -\boldsymbol{C}_b^c \left(\boldsymbol{C}_n^b \boldsymbol{D}_R \left(\boldsymbol{p}_s - \boldsymbol{p}_I \right) \times \right) \;\; \boldsymbol{0}_{3\times3} \;\; \boldsymbol{0}_{3\times3} \;\; \boldsymbol{0}_{3\times3} \;\; \boldsymbol{0}_{3\times3} \big]. \tag{26}$$

V. Performance Evaluation

A. Experimental Setup and the Test Scenario

To validate the proposed fusion positioning scheme, we set up a data acquisition experimental platform as shown in Fig. 6. A binocular camera (Stereolabs, ZED 2i) is mounted at the front of the vehicle, while an integrated navigation system (Xsens, MTi-680G) is installed on an internal platform in the trunk, with a GNSS receiver mounted on the roof of

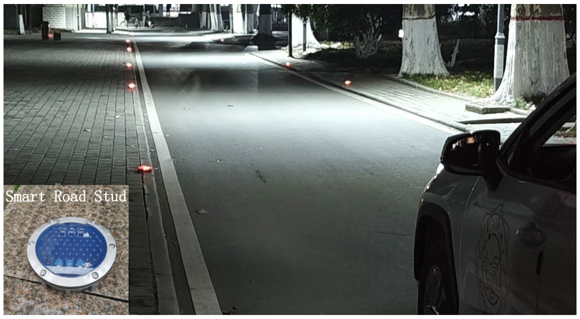


Fig. 7. Experimental scenario.

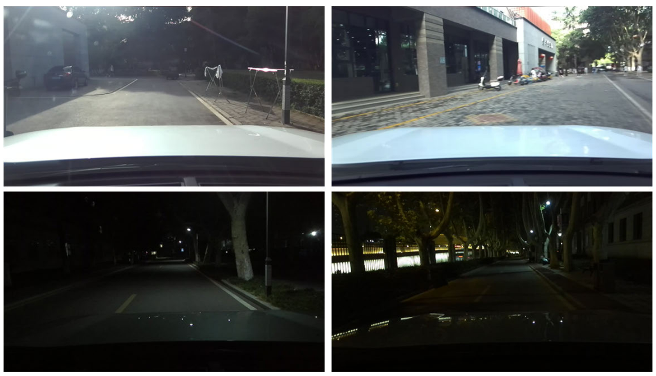


Fig. 8. Data samples used for training Speed-Net.

the vehicle. To collect the global positioning system (GPS) data and LiDAR data needed for comparative experiments, we additionally configure a GPS receiver (Topgnss, GNSS 100G) with a sampling frequency of 1 Hz and a LiDAR (RoboSense Ruby Plus) with a frame rate of 10 Hz. The frame rate of the camera is 30 fps, and it has 1280×720 image resolution with a field of view (FOV) of $90° \times 60°$. The integrated navigation system combines IMU data and RTK data to provide a reference to evaluate the results of the proposed method. The sample rate of the IMU is 100 Hz. The bias stability of the gyroscopes and the accelerometers are 8°/h and 15 μg, respectively, and the random walk of the gyroscopes and the velocity are $0.17°/\sqrt{h}$ and 0.02 m/s/$\sqrt{h}$, respectively.

The experiment is conducted at Xidian University in Xi'an, China. A total of 42 SRSs are deployed along both sides of the experimental route with the distance between adjacent SRSs set to 15 m, as shown in Fig. 7.

B. Training and Testing for Speed-Net

We use the experimental platform to collect data for training and testing Speed-Net. The image sequences are provided by the left camera of the binocular camera, and the corresponding ground truth speeds are provided by the RTK. We collect approximately 180 min of data for training the model. During the data collection, we consider a wider range of driving conditions, including acceleration, deceleration, and turning, as well as various lighting environments. The same collected data samples are shown in Fig. 8. The model is trained by using an NVIDIA GeForce GTX 1660 SUPER GPU. The

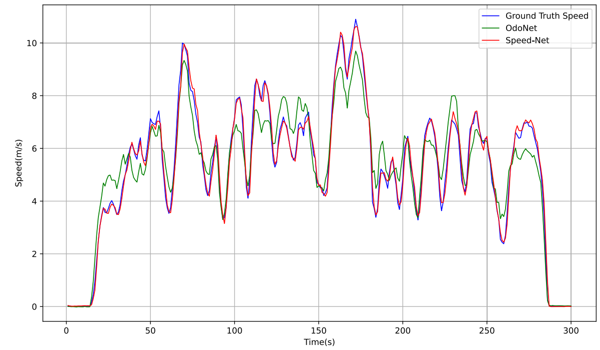


Fig. 9. Speed estimation result of different models.

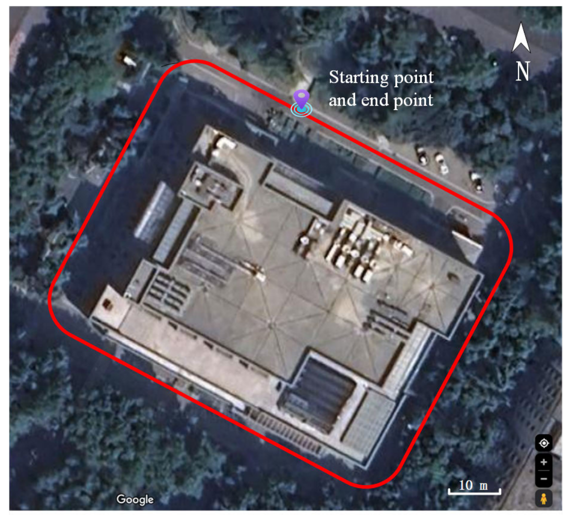


Fig. 10. Vehicle test trajectory.

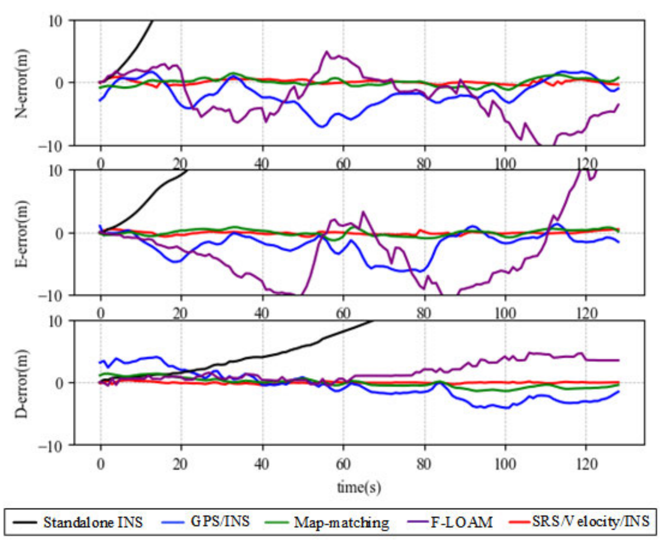


Fig. 11. 3-D position errors for the five algorithms.

Adagrad optimizer is used to train the network for up to 200 epochs with a learning rate of 0.001.

To verify the superiority of the proposed Speed-Net in speed prediction, we compare the prediction results of the

TABLE II
PERFORMANCE OF DIFFERENT LOCALIZATION ALGORITHMS

Methods	Orientation	Mean Abs. Pos. Error (m)	Max. Abs. Pos. Error (m)	RMSE (m)
Standalone INS	N	289.41	880.17	395.80
	E	54.43	269.93	89.97
	D	10.62	33.45	14.56
F-LOAM	N	3.70	11.17	4.58
	E	5.35	15.13	6.47
	D	1.80	4.71	2.24
GPS/INS	N	2.34	7.16	2.85
	E	2.19	6.29	2.74
	D	2.06	4.10	2.35
Map-matching	N	0.50	1.43	0.61
	E	0.44	1.30	0.51
	D	0.65	1.40	0.77
SRS/Velocity/INS	N	0.29	0.88	0.36
	E	0.23	0.72	0.30
	D	0.12	0.44	0.15

IMU-based speed prediction model OdoNet [44] and Speed-Net, as shown in Fig. 9. The mean and standard deviation of the Speed-Net estimated speed error are 0.18 and 0.16 m/s, respectively. For OdoNet, the mean and standard deviation of the estimated speed error are 0.61 and 0.40 m/s. Based on the comparison results, it is evident that Speed-Net can more accurately estimate the vehicle's speed compared to OdoNet.

C. Performance Comparison

To evaluate the performance of different localization algorithms, four methods are tested: the standalone INS algorithm, the SLAM-based localization algorithm (F-LOAM [47]), the GPS/INS integrated localization algorithm (using ESEKF to fuse GPS data and INS), and the map-matching algorithm. We employ an improved particle filter algorithm from the literature [48] to match GPS data with a digital map, achieving map-matching-based localization. The digital map used in this work is constructed by accurately extracting the road boundaries in the experimental area using RTK equipment and the proposed SRS/velocity/INS integration algorithm. Fig. 10 shows the vehicle test trajectory. Fig. 11 shows the positioning errors in the North (N), East (E), and Down (D) directions for the five algorithms. The mean absolute error, the maximum absolute error, and the root-mean-square error (RMSE) of positioning results along different coordinate directions of different algorithms are reported in Table II.

The RMSE is used to measure the difference between positioning results and ground truth, which is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{27}$$

where $\hat{y}_i$ is the output of the localization algorithm, y_i is the ground truth, and n is the total number of data.

As shown in Fig. 11, the positioning errors of the standalone INS in all three directions increase significantly as time progresses. The maximum absolute positioning errors in the N, E, and D directions reach 880.17, 269.93, and 33.45 m, respectively. This demonstrates the inherent error accumulation problem of the INS, which is unacceptable in practical navigation applications. Similarly, due to error accumulation and inaccurate interframe pose estimation, the localization errors of F-LOAM are significant, with the maximum absolute position errors reaching 11.17 and 15.13 m in the N and E directions, respectively, and 4.71 m in the D direction. In contrast, by fusing GNSS data, the positioning errors are significantly reduced and become more stable. The maximum absolute positioning errors in all three directions are within 10 m, and the mean absolute positioning errors in the N, E, and D directions reduce to 2.34, 2.19, and 2.06 m, respectively. But this still falls short of meeting the localization accuracy requirements for CAVs. Map-matching further enhances positioning precision by aligning the vehicle's trajectory with the known road map. According to Table II, compared to the GPS/INS algorithm, map-matching reduces the mean absolute positioning error and RMSE in all directions by over 60%. Finally, the SRS/velocity/INS method achieves the best overall performance by using vision-based positioning information from SRSs to suppress position error divergence of the INS and by using speed information estimated by Speed-Net to control velocity error divergence of the INS. This approach achieves mean absolute positioning errors as low as 0.29 m in the N direction, 0.23 m in the E direction, and 0.12 m in the D direction. The maximum absolute positioning error is limited to 0.88 m across all directions, with RMSE values well below 0.5 m, demonstrating the high accuracy and robustness of this method. It is worth noting that the SRS/velocity/INS vehicle localization algorithm proposed in this article only utilizes GNSS for the initial position. During the subsequent oper-

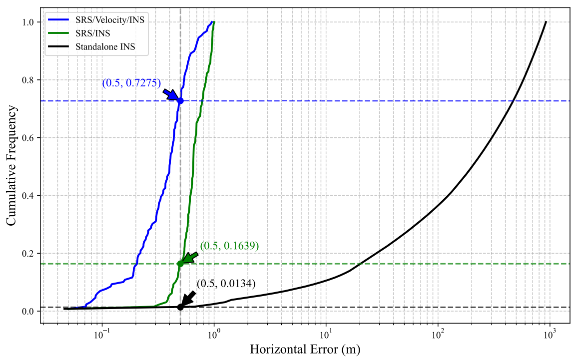


Fig. 12. Cumulative horizontal absolute error comparison of the standalone INS, SRS/INS, and SRS/velocity/INS.

ation, it completely eliminates the need for GNSS, integrates vision-based positioning information assisted by SRSs, and velocity information estimated by Speed-Net from the image sequences, achieving submeter positioning accuracy.

D. Ablation Study

To evaluate the contribution of each module in our proposed positioning framework, we conduct ablation experiments involving the following three configurations: standalone INS, INS aided by SRSs (SRS/INS), and INS aided by both SRSs and vehicle velocity information predicted by Speed-Net (SRS/Velocity/INS). During vehicle movement, we focus primarily on the horizontal positioning accuracy. To intuitively compare the positioning errors of the three methods, we analyze the cumulative absolute error distribution. The horizontal axis is plotted on a base-10 logarithmic scale to better illustrate error distribution across a wide range.

As shown in Fig. 12, the standalone INS suffers from significant drift, with only 1.34% of the horizontal absolute errors falling below 0.5 m, indicating poor positioning reliability. When aided by SRSs, the positioning accuracy improves considerably. Most horizontal errors remain within 1 m, and 16.39% of them are below 0.5 m. The best performance is achieved when both SRSs and predicted velocity are integrated, where the system reaches a cumulative distribution function value of 72.75% at 0.5 m. The statistics indicate that the mean horizontal absolute error for the SRS/velocity/INS method is 0.40 m, with an RMSE of 0.45 m, while the mean horizontal absolute error for the SRS/INS method is 0.67 m, with an RMSE of 0.69 m. These results indicate that the inclusion of velocity information leads to smaller and more concentrated positioning errors. This demonstrates the effectiveness of the proposed Speed-Net velocity estimation method in enhancing vehicle localization accuracy.

VI. Conclusion and Future Work

This article proposed an integrated SRS/velocity/INS vehicle localization method. The method employed an ESEKF to fuse the position information of SRSs—IoT devices increasingly used in intelligent transportation systems—with speed information derived from image sequences by a DNN, Speed-Net, and INS data, achieving submeter-level positioning accuracy. We first trained a YOLOv8-based SRS detection model on a self-built SRS dataset using the NWD loss function. Based on the detection of the SRSs, we directly constructed a position measurement in the camera coordinate system using binocular measurement and position transformations to suppress the divergence of the INS position error. Additionally, we introduced a method to improve vehicle localization accuracy by incorporating vehicle speed without requiring additional hardware-based speed measurement devices. The vehicle speed was estimated using a DNN-based model, Speed-Net, which derives vehicle speed from image sequences acquired by the onboard camera. The estimated speed was subsequently employed to mitigate the divergence of the velocity estimation error in the INS. Experimental results validated the effectiveness of the proposed SRS/velocity/INS localization method, achieving a mean horizontal absolute positioning error of 0.40 m and an RMSE of 0.45 m for horizontal positioning.

However, the method proposed in this article also has its limitations. On the one hand, due to experimental and legal constraints, we only conducted field tests within the campus. In the future, we will validate our method in more complex environments. On the other hand, we will explore other methods to fuse SRS, vehicle speed, and INS information, especially the factor graph-based optimization method.

References

[1] K. Ansari, "Cooperative position prediction: Beyond vehicle-to-vehicle relative positioning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1121–1130, Mar. 2020.

[2] A. Mahmoud, A. Noureldin, and H. S. Hassanein, "Integrated positioning for connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 397–409, Jan. 2020.

[3] X. Tao et al., "A multi-sensor fusion positioning strategy for intelligent vehicles using global pose graph optimization," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2614–2627, Mar. 2022.

[4] Y. Xu, K. Wang, C. Yang, Z. Li, F. Zhou, and D. Liu, "GNSS/INS/OD/NHC adaptive integrated navigation method considering the vehicle motion state," *IEEE Sensors J.*, vol. 23, no. 12, pp. 13511–13523, Jun. 2023.

[5] N. Vagle, A. Broumandan, and G. Lachapelle, "Multiantenna GNSS and inertial sensors/odometer coupling for robust vehicular navigation," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4816–4828, Dec. 2018.

[6] X. Li, Z. Qin, Z. Shen, X. Li, Y. Zhou, and B. Song, "A high-precision vehicle navigation system based on tightly coupled PPP-RTK/INS/odometer integration," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 2, pp. 1855–1866, Feb. 2023.

[7] X. Meng, H. Tan, P. Yan, Q. Zheng, G. Chen, and J. Jiang, "A GNSS/INS integrated navigation compensation method based on CNN–GRU + IRAKF hybrid model during GNSS outages," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–15, 2024.

[8] T. Zhang, L. Zhou, X. Feng, J. Shi, Q. Zhang, and X. Niu, "INS-aided GNSS pseudo-range error prediction using machine learning for urban vehicle navigation," *IEEE Sensors J.*, vol. 24, no. 6, pp. 9135–9147, Mar. 2024.

[9] Q. Xu, X. Yu, X. Li, and X. Liu, "Environmental information-assisted intelligent fusion localization for vehicles in urban area," *Measurement*, vol. 237, Sep. 2024, Art. no. 115144.

[10] B. Yang, L. Xue, H. Fan, and X. Yang, "SINS/odometer/Doppler radar high-precision integrated navigation method for land vehicle," *IEEE Sensors J.*, vol. 21, no. 13, pp. 15090–15100, Jul. 2021.

[11] X. Niu, Y. Peng, Y. Dai, Q. Chen, C. Guo, and Q. Zhang, "Camera-based lane-aided multi-information integration for land vehicle navigation," *IEEE/ASME Trans. Mechatronics*, vol. 28, no. 1, pp. 152–163, Feb. 2023.

[12] Y. Lu, H. Ma, E. Smart, and H. Yu, "Real-time performance-focused localization techniques for autonomous vehicle: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6082–6100, Jul. 2022.

[13] G. Mao, Y. Hui, X. Ren, C. Li, and Y. Shao, "The Internet of Things for smart roads: A road map from present to future road infrastructure," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 6, pp. 66–76, Nov. 2022.

[14] Y. Sun et al., "Smart road stud-empowered vehicle magnetic field distribution and vehicle detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7357–7362, Jul. 2023.

[15] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[16] T. Zhang et al., "HOG-ShipCLSNet: A novel deep learning network with HOG feature fusion for SAR ship classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5210322.

[17] S. Y. Alaba and J. E. Ball, "Deep learning-based image 3-D object detection for autonomous driving: Review," *IEEE Sensors J.*, vol. 23, no. 4, pp. 3378–3394, Feb. 2023.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[19] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[21] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 21–37.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[23] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.

[24] T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, and Y. Shan, "YOLO-world: Real-time open-vocabulary object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 16901–16911.

[25] H. Liu, R. Hu, H. Dong, and Z. Liu, "SFC-YOLOv8: Enhanced strip steel surface defect detection using spatial-frequency domain-optimized YOLOv8," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–11, 2025.

[26] J. Wang, C. Xu, W. Yang, and L. Yu, "A normalized Gaussian Wasserstein distance for tiny object detection," 2021, *arXiv:2110.13389*.

[27] S. Wang, R. Clark, H. Wen, and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 2043–2050.

[28] N. Abdelaziz and A. El-Rabbany, "LiDAR/visual SLAM-aided vehicular inertial navigation system for GNSS-denied environments," in *Proc. 5th Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Dec. 2022, pp. 1–5.

[29] X. Niu, H. Tang, T. Zhang, J. Fan, and J. Liu, "IC-GVINS: A robust, real-time, INS-centric GNSS-visual-inertial navigation system," *IEEE Robot. Autom. Lett.*, vol. 8, no. 1, pp. 216–223, Jan. 2023.

[30] Y. Wu and J. Zhao, "A robust and precise LiDAR-inertial-GPS odometry and mapping method for large-scale environment," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 6, pp. 5027–5036, Dec. 2022.

[31] H. Wang, C. Xue, Y. Zhou, F. Wen, and H. Zhang, "Visual semantic localization based on HD map for autonomous vehicles in urban scenarios," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Xi'an, China, May/Jun. 2021, pp. 11255–11261.

[32] A. Hansson, E. Korsberg, R. Maghsood, and E. Norden, "Lane-level map matching based on HMM," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 3, pp. 430–439, Sep. 2021.

[33] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1697–1716, Aug. 2019.

[34] S. R. Bhavsar, A. Vatavu, T. Rehfeld, and G. Krehl, "Sensor fusion-based online map validation for autonomous driving," in *Proc. Intell. Veh. Symp. Workshops*, virtual, Oct. 2020, pp. 77–82.

[35] C. Wang, H. Huang, Y. Ji, B. Wang, and M. Yang, "Vehicle localization at an intersection using a traffic light map," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1432–1441, Apr. 2019.

[36] A. Welzel, P. Reisdorf, and G. Wanielik, "Improving urban vehicle localization with traffic sign recognition," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 2728–2732.

[37] H. Hu, K. Li, W. Liang, Q. Li, and Z. Xie, "Kilometer sign positioning-aided INS/odometer integration for land vehicle autonomous navigation," *IEEE Sensors J.*, vol. 23, no. 4, pp. 4143–4158, Feb. 2023.

[38] Q. Zhang, Y. Dai, T. Zhang, C. Guo, and X. Niu, "Road semantic-enhanced land vehicle integrated navigation in GNSS denied environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 12, pp. 20889–20899, Dec. 2024.

[39] K.-W. Chiang et al., "Assessment for INS/GNSS/odometer/barometer integration in loosely-coupled and tightly-coupled scheme in a GNSS-degraded environment," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3057–3069, Mar. 2020.

[40] S. Bai, J. Lai, P. Lyu, B. Wang, X. Sun, and W. Yu, "An enhanced adaptable factor graph for simultaneous localization and calibration in GNSS/IMU/odometer integration," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 11346–11357, Sep. 2023.

[41] P. R. M. de Araujo, M. Elhabiby, S. Givigi, and A. Noureldin, "A novel method for land vehicle positioning: Invariant Kalman filters and deep-learning-based radar speed estimation," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 9, pp. 4275–4286, Sep. 2023.

[42] X. Wang et al., "SdoNet: Speed odometry network and noise adapter for vehicle integrated navigation," *IEEE Internet Things J.*, vol. 10, no. 21, pp. 19328–19343, Nov. 2023.

[43] Z. Yu, J. Jiang, P. Yan, Y. Li, J. Wu, and D. Xie, "A lightweight odometry network for GNSS/INS integration during GNSS outages," *Appl. Soft Comput.*, vol. 151, Jan. 2024, Art. no. 111143.

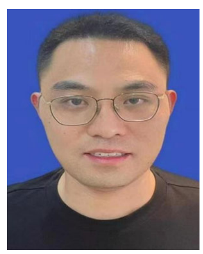
[44] H. Tang, X. Niu, T. Zhang, Y. Li, and J. Liu, "OdoNet: Untethered speed aiding for vehicle navigation without hardware wheeled odometer," *IEEE Sensors J.*, vol. 22, no. 12, pp. 12197–12208, Jun. 2022.

[45] S. Lin, W. Li, C. Wang, and Y. Tang, "Distance measurement of underwater target based on stereo vision," in *Proc. IEEE 7th Annu. Int. Conf. CYBER Technol. Autom., Control, Intell. Syst. (CYBER)*, Jul. 2017, pp. 97–102.

[46] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[47] H. Wang, C. Wang, C.-L. Chen, and L. Xie, "F-LOAM: Fast LiDAR odometry and mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 4390–4396.

[48] W. Li, T. Chen, J. Han, L. Wang, and J. Huang, "A high-precision map matching algorithm based on nonlinear filtering optimization for IoT industrial park," *IEEE Sensors J.*, vol. 24, no. 2, pp. 2209–2215, Jan. 2024.



Keyin Wang (Graduate Student Member, IEEE) received the master's degree in control engineering from Hubei University of Automotive Technology, Shiyan, China, in 2022. He is currently pursuing the Ph.D. degree with the College of Communication Engineering, Xidian University, Xi'an, China.

His research interest is vehicle localization based on multisource information fusion.

Guoqiang Mao (Fellow, IEEE) is a Chair Professor, Director of the Center for Smart Driving and Intelligent Transportation Systems, Southeast University, Nanjing, China. Between 2014 and 2019, he was a Leading Professor, Founding Director of the Research Institute of Smart Transportation, and a Vice-Director of the ISN State Key Laboratory at Xidian University. Before that he was with the University of Technology Sydney and the University of Sydney. He has published 300 papers in international conferences and journals that have been cited more than 15,000 times. His H-index is 57 and is in the list of Top 2% most-cited Scientists Worldwide by Stanford University in 2022, 2023, and 2024 both by Single Year and by Career Impact. He serves as a Vice-Director of Smart Transportation Information Engineering Society, Chinese Institute of Electronics (2022). His research interest includes intelligent transport systems, Internet of Things, wireless localization techniques, mobile communication systems, and applied graph theory and its applications in telecommunications.

Dr. Mao is a Fellow of AAIA and IET. He was a co-chair IEEE ITS Technical Committee on Communication Networks (2014-2017). He is an editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (since 2018), IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2014-2019), IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2010-2020) and received "Top Editor" award for outstanding contributions to the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2011, 2014, and 2015. He has served as a chair, co-chair and TPC member in a number of international conferences.

Xiaojiang Ren (Member, IEEE) received the Ph.D. degree in computer science from The Australian National University, Canberra, ACT, Australia, in 2016.

He is an Associate Professor with Xidian University, Xi'an, China. His research interests include intelligent transport systems, the Internet of Things, wireless sensor networks, routing protocol design for wireless networks, and optimization problems.

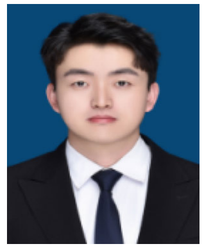
Haoyuan Du received the bachelor's degree in communication engineering from Dongguan University of Technology, Dongguan, China, in 2021. He is currently pursuing the master's degree with Guangzhou Institute of Technology, Xidian University, Guangzhou, China.

His research interest is new-generation electronic information technology.

Baoqi Huang (Senior Member, IEEE) received the B.E. degree in computer science from Inner Mongolia University (IMU), Hohhot, China, in 2002, the M.S. degree in computer science from Peking University, Beijing, China, in 2005, and the Ph.D. degree in information engineering from The Australian National University, Canberra, ACT, Australia, in 2012, respectively.

He is with the College of Computer Science, IMU, where he is currently a Professor. His research interests include indoor localization and navigation, wireless sensor networks, and mobile computing.

Dr. Huang was a recipient of the Chinese Government Award for Outstanding Chinese Students Abroad in 2011.



Tianxuan Fu received the master's degree in transport information engineering and control from Lanzhou Jiaotong University, Lanzhou, China, in 2019. He is currently pursuing the Ph.D. degree with the College of Communication Engineering, Xidian University, Xi'an, China.

His research interests include target tracking, information fusion, and intelligent transportation systems.

Zhaozhong Zhang received the B.Eng. degree from the University of Nottingham, Nottingham, U.K., in 2013, and the M.Sc. and Ph.D. degrees in automotive engineering and transport systems from Cranfield University, Cranfield, U.K., in 2015 and 2020, respectively.

He is currently a Lecturer with the School of Information Engineering, Nanchang Hangkong University, Nanchang, China. His research interests include vehicle localization and target tracking.