A smart roadside unit with integrated vision and 24 GHz IOT radar for real-time behavior prediction and V2X communication

Le Manh Tuan[1*], Nguyen Tien Viet[1], Nguyen Van Long[1],
Nguyen Van Nghia[2], Tran Anh Dung[3]

[1]Institute of Information Technology and Electronics, Academy of Military Science and Technology, 17 Hoang Sam, Nghia Do, Hanoi, Vietnam;
[2]Posts and Telecommunications Institute of Technology, 96A Tran Phu, Ha Dong, Hanoi, Vietnam;
[3]X-Force Design Solutions Viet Nam JSC, To Huu, Dai Mo, Hanoi, Vietnam.
[*]Corresponding author: lemanhtuan.le3@gmail.com
Received 5 Aug. 2025; Revised 17 Sep. 2025; Accepted 10 Oct. 2025; Published 30 Oct. 2025.
DOI: https://doi.org/10.54939/1859-1043.j.mst.IITE.2025.35-44

ABSTRACT

This paper proposes a Smart Roadside Unit (RSU) architecture designed for intelligent transportation systems in high-density urban environments with heterogeneous traffic. The system integrates V2X communication, 24 GHz radar, and AI-based vision to enable autonomous, context-aware decision-making for traffic regulation and incident response. Leveraging a machine learning model trained on localized traffic behavior, the RSU not only detects violations but also predicts potentially hazardous actions specific to each area. Real-time sensor fusion between radar and camera enhances the accuracy of object detection, tracking, and behavioral analysis under various environmental conditions. The proposed system improves traffic flow management, congestion mitigation, and proactive control capabilities. The solution is well-suited for deployment in smart cities, particularly in the context of increasing urbanization and the growing demand for data-driven mobility infrastructure.

Keywords: Smart RSU; V2X communication; 24 GHz radar; AI camera; Sensor fusion; Traffic behavior prediction; Intelligent transportation systems; Smart city.

1. INTRODUCTION

In today's digital age, the V2X OBU (On-Board Unit) and RSU (Roadside Unit) designs are two important components in the intelligent transportation system using V2X (Vehicle-to-Everything) technology - a means of communicating with everything around. Many studies have shown that V2X is very effective in regulating traffic, reducing traffic congestion, minimizing traffic accidents and increasing traffic efficiency, and is especially effective in its ability to regulate traffic flow with the central traffic management system [1, 2].

Normally, the Roadside Unit often plays an active role in connecting or communicating with On-Board Unit devices (on the vehicle), so in addition to the communication role, the RSU also plays an important role as a data transfer point between traffic participating devices and the remote central management system. This opens up a new direction for researchers to apply the combination of AI and V2X communication on current RSU devices. Helping RSU blocks become intelligent and able to predict and navigate based on panoramic data and detect and respond to abnormal behavior of vehicles [3, 4].

The old RSU designs are usually only used for the main purpose of transmitting data to support the remote monitoring and control process with many different protocols divided into 2 main types: 3GPP LTE/5G, Ethernet, Lora, Wifi when communicating with the central management system and OBU, CAN, RS-485, SPI, USB… for communicating with sensor devices [5, 6]. Recently, with the development of generations of computers and microprocessors, RSUs combined with AI have become a new development direction in the high-tech era. Based on the diversity of sensors on the RSU as well as the information continuously provided from the OBU, the AI-integrated processing system on the RSU is becoming increasingly smarter with accurate warnings and predictions, helping the central management system to control vehicles and traffic participants more closely and optimally.

This paper presents an RSU design with many different communication protocol standards, and also provides an AI solution for observing and detecting vehicles and traffic participants, combining a camera with a 24 GHz IoT radar system to help improve accuracy and predict the behavior of the subject, thereby improving congestion or proactively intervening, minimizing accidents and improving the efficiency of urban traffic coordination.

2. SYSTEM DESIGN

2.1. Block diagram

Figure 1 proposes the block diagram of an AI RSU module with detailed design and technology options, including an infineon 24 GHz Radar and Mobile AI camera with NVIDIA Jetson AI processing module.

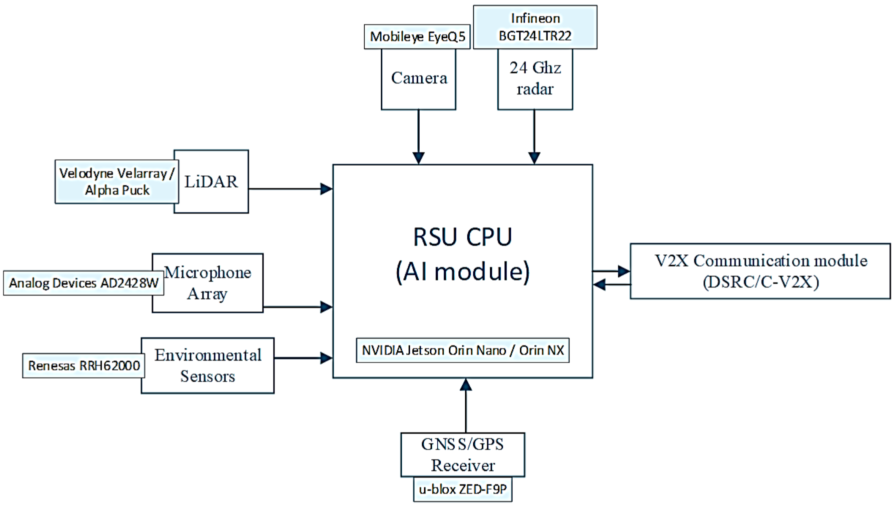

Figure 1. *Block diagram of the proposed AI RSU module.*

In this design, integrating a powerful processing platform such as NVIDIA Jetson Orin NX into RSU (Roadside Unit) devices brings many breakthroughs in performance and responsiveness. With a processing capacity of up to 100 TOPS, Jetson Orin NX allows simultaneous deployment of multiple deep learning models such as YOLOv8, ResNet or MobileNet to perform tasks such as pedestrian detection, vehicle classification, dangerous traffic behavior recognition, and real-time collision risk prediction.

In addition, Orin NX supports communication with a variety of sensors such as AI cameras, 24 GHz radars, microphone arrays, environmental sensors, and GNSS, helping the RSU system effectively perform data fusion to build accurate situation maps under complex weather and lighting conditions. With a powerful software platform including Ubuntu Linux, JetPack SDK and the ability to integrate V2X protocols, RSUs can perform coordinated actions such as controlling traffic lights or early warnings to vehicles via V2I communication [7].

Jetson Orin NX also meets industrial deployment requirements with reasonable power consumption (~10–20W) [8], wide temperature range and remote update capabilities. Therefore, this platform is an ideal choice for building smart RSU systems with high performance, good scalability and meeting the trend of AI in urban traffic infrastructure.

Choosing NVIDIA Jetson also makes it more convenient and easier to deploy AL algorithms with cameras and Radar at the same time.

2.2. 24 GHz radar algorithm on Infineon radar platform

24 GHz radar is becoming a core technology in intelligent transportation systems, providing high-precision urban traffic monitoring and management. With the ability to measure vehicle speed, distance and direction in all weather and light conditions, 24 GHz radar provides real-time data to optimize traffic flow, reduce congestion and improve road safety. The practical application process can be carried out in the following directions:

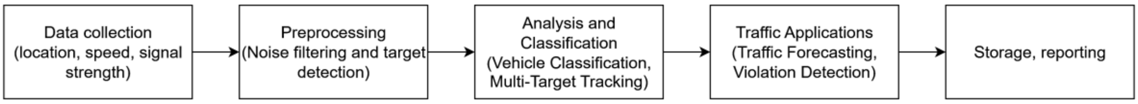

Figure 2. *Block diagram of radar 24 GHz sensor signal processing [20].*

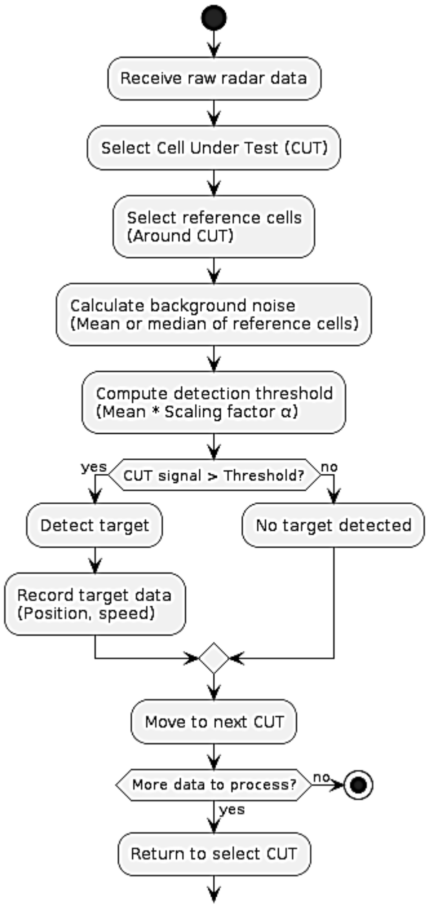

Figure 3. *CFAR (Constant False Alarm Rate) algorithm flowchart.*

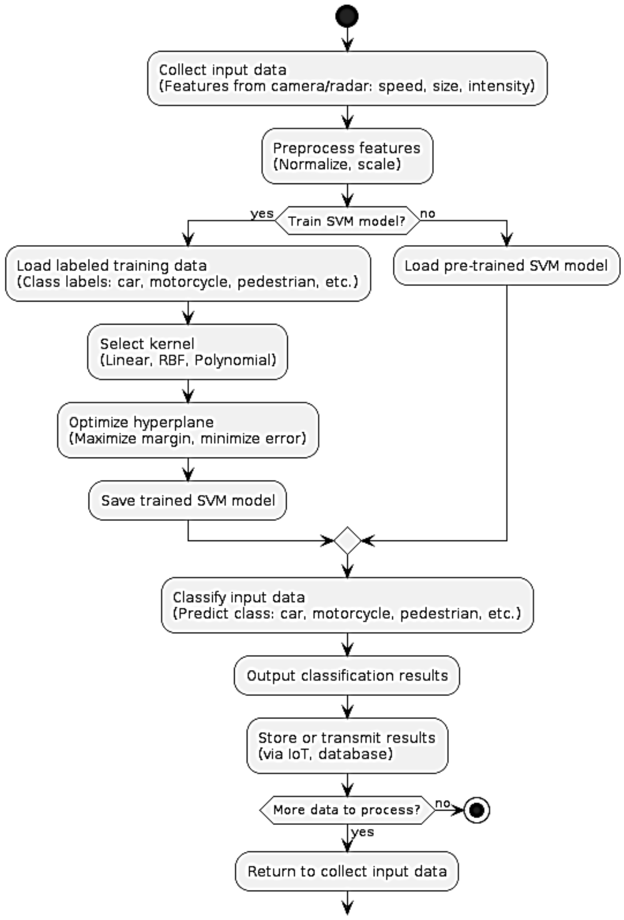

Figure 4. *SVM (Support Vector Machine) algorithm flowchart.*

The signal received from the 24 GHz radar, including (location, speed and signal strength), is passed through the preprocessor to filter noise and detect targets. Some algorithms are applied, such as Kalman Filter noise filtering, CFAR target detection [10]. The data after the preprocessor is sent to the data analysis and data classification unit to classify vehicles, and perform multi-target tracking (following). Using algorithms such as SVM [11, 12], CNN, the vehicles are classified in detail. Combined with tracking algorithms such as JPDA, Particle Filter, the software is capable of continuously tracking vehicles, providing full information to the system.

Based on the collected and processed information, the smart traffic application can build a picture of the traffic situation, perform traffic flow forecasting, traffic flow separation, traffic signal optimization, or detect traffic violations.

At its core, the CFAR algorithm estimates the power of the background noise or clutter surrounding a "cell under test" (CUT). This noise estimate is then multiplied by a scaling factor (often denoted as α or T) to establish a detection threshold. If the signal strength in the CUT surpasses this threshold, it is declared a target.

The fundamental purpose of the scaling factor is to maintain a constant probability of false alarm (P_{fa}). In a predictable, homogeneous environment where the noise characteristics are well-understood (typically assumed to follow a Rayleigh distribution), the scaling factor can be a predetermined constant. This constant is calculated based on the desired (P_{fa}) and the number of training cells used to estimate the noise. A higher scaling factor results in a higher threshold, leading to fewer false alarms but also potentially missing weaker targets. Conversely, a lower scaling factor increases the detection probability at the cost of more frequent false alarms.

The assumption of a homogeneous noise environment often breaks down in real-world scenarios. Radar systems frequently encounter non-homogeneous clutter, which can include:

- **Clutter edges:** Abrupt transitions in the background noise level, such as moving from sea to land.
- **Multiple targets:** The presence of other targets within the training window can artificially inflate the noise estimate.
- **Interference:** Signals from other sources can contaminate the background noise.

In these situations, a fixed scaling factor becomes inadequate. For instance, at a clutter edge, a standard CFAR algorithm might incorrectly estimate the noise level, leading to an inappropriate threshold and either missed detections or excessive false alarms. The statistical distribution of the clutter may also deviate from the assumed model, rendering the pre-calculated scaling factor suboptimal.

To address these shortcomings, two adaptive strategies are commonly employed. The first is Local Clutter Variability Adaptation (LCVA), which adjusts the scaling factor α according to the statistical variability of the reference cells. In this approach, both the mean and the variance of the reference cells are computed. A high variance indicates non-homogeneous clutter, prompting an increase in α to suppress false alarms, whereas a low variance corresponds to homogeneous clutter, allowing α to be reduced to improve sensitivity.

The second method is SNR-dependent Scaling, in which α is adapted based on the estimated local signal-to-noise ratio (SNR). In high-SNR conditions, α is reduced to enhance the detection of weak targets near strong reflectors. Conversely, in low-SNR environments, α is increased to mitigate clutter breakthrough and reduce false alarms.

Together, these adaptive mechanisms make CFAR more robust in complex and dynamic environments, improving detection performance across a wide range of operating conditions.

2.3. AI camera algorithm

Cameras provide the ability to collect and analyze images to improve the efficiency of urban traffic management. With the ability to record high-resolution images, cameras support vehicle identification, license plates, traffic behavior, and detect incidents such as accidents or law violations. The combination of cameras and 24 GHz radars diversifies data and increases the accuracy required for the system.

To exploit data from cameras, image processing and deep DL algorithms are used to analyze, identify, and make decisions in real time. The practical application process includes collecting images from cameras, processing them with computer vision algorithms such as object recognition (YOLO, SSD), feature extraction, and behavior classification (using CNN, SVM) [13, 14].

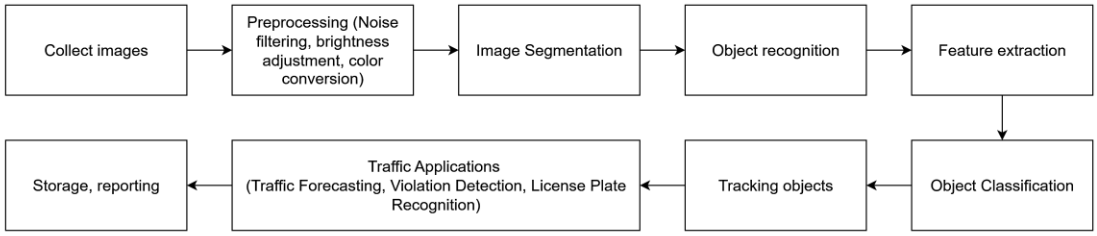

Figure 5. *Camera image processing flow chart.*

Images captured from the camera are passed through a preprocessor to improve quality and reduce noise. Some algorithms are applied, such as Gaussian Blur noise filtering, Histogram Equalization contrast adjustment, and Edge Detection. The images are then passed through an image classifier or Image Segmentation to determine vehicle boundaries, lanes, and areas. Algorithms such as Thresholding, Watershed Algorithm can be applied to enhance the ability to distinguish and separate close vehicles in complex conditions [15, 16].

The object will then be recognized thanks to algorithms such as Deep Learning, YOLO or Faster R-CNN [17, 18]. After successful recognition, the object is extracted with features, classified and tracked. Using traffic application data to build functions such as traffic forecasting, violation detection, and license plate recognition. Data is then stored and reported visually on the system.

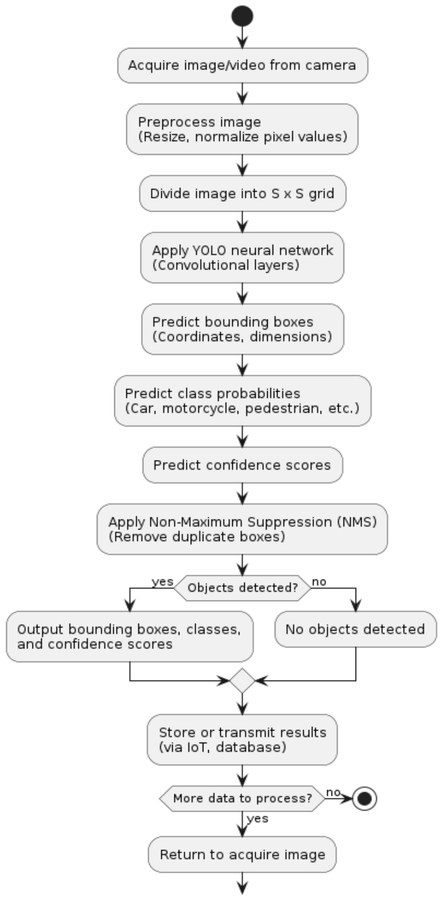

Figure 6. *YOLO algorithm flowchart.*

3. MODEL COMBINING PROCESSING BETWEEN AI CAMERA AND 24 GHz RADAR

When combining AI cameras and 24 GHz radars in intelligent transportation systems, especially at RSUs (Roadside Units), we can take advantage of the advantages of both types of sensors to achieve higher and more stable detection of people and vehicles in different conditions, especially operating environments.

At the same time, this combination also helps improve the ability to monitor, analyze and handle complex traffic situations. Each type of sensor has its own strengths: 24 GHz radar allows accurate measurement of the speed, distance and direction of movement of objects in all weather and lighting conditions; meanwhile, AI-integrated cameras have the ability to accurately identify and classify objects such as motorbikes, cars, pedestrians or trucks through machine learning. When integrated synchronously, these two technologies create a multi-modal sensing system, which helps detect not only "there is an object" but also "who it is", "what it is doing", and "what risks are occurring" [19].

This combination also significantly reduces false positive/negative warnings that are likely to occur when using only a single type of sensor. At the same time, the ability to analyze real-time behavior from AI cameras, combined with kinematic data from radar, allows the system to predict dangerous behaviors early, such as people crossing the street unexpectedly or vehicles running red lights. Thanks to that, the system not only detects events but also proactively warns and intervenes, contributing significantly to reducing accidents, reducing congestion and improving the efficiency of urban traffic coordination [20].

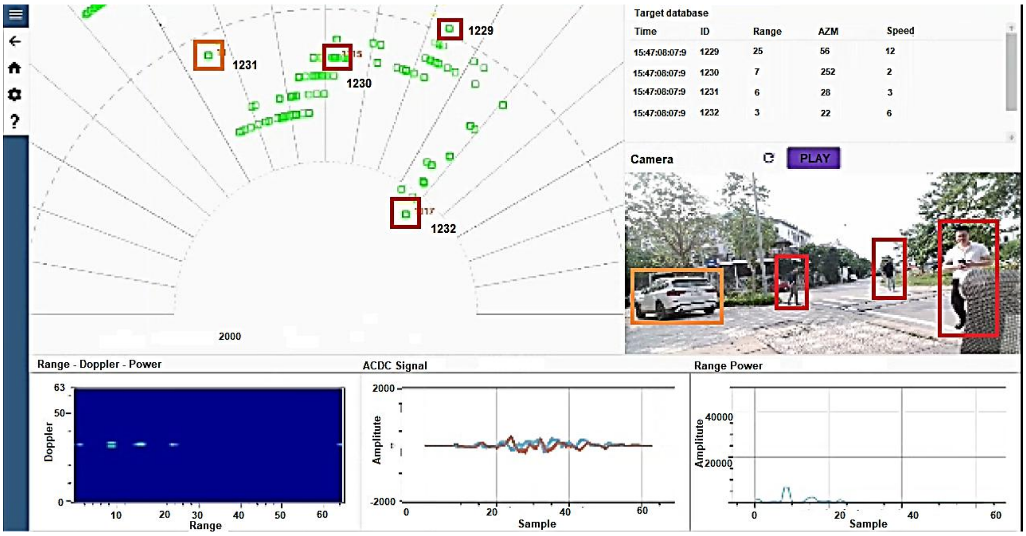

Figure 7. *Comparison results between radar and camera in identifying people and vehicles participating in traffic.*

3.1. Algorithm for combining radar and vision data (sensor fusion)

To fuse radar (providing speed, distance, sparse points, then the direction of movement) and vision (providing classification, dense bounding boxes) data for enhanced detection/tracking, use Object-Level Fusion with Extended Kalman Filter (aka EKF). This is a *mid-to-high level fusion* approach that aligns detections temporally and spatially before merging.

Data alignment:

- Temporal synchronization: Timestamp both sensor streams and interpolate/extrapolate to a common rate.
- Spatial calibration: Project radar points into camera coordinates using extrinsic parameters

(rotation/translation matrix from calibration). Here we can apply Homography or Perspective-n-Point (aka PnP) solver to project radar points to the image. Align the radar's polar (range, angle) to the camera's pixel space.

- Association: Match radar detections to vision bounding boxes using metrics like Mahalanobis distance after projection. The Hungarian algorithm is used for optimal assignment in multi-target scenarios.

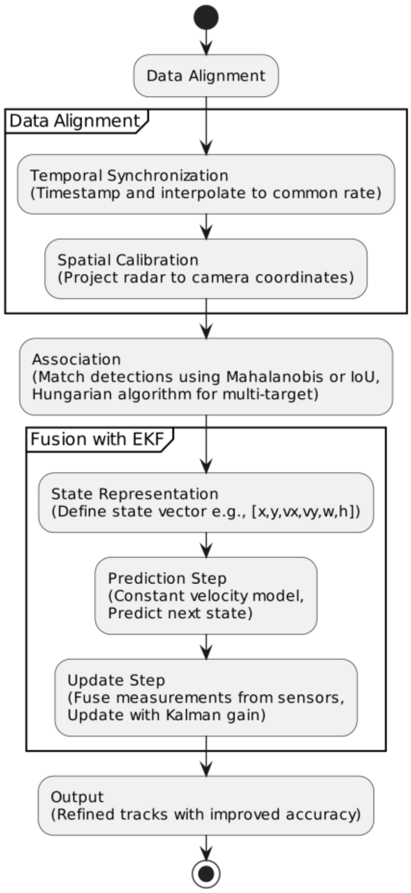

Figure 8. *Algorithm for combining radar and vision data.*

Fusion with EKF

- State Representation: Define a state vector for each tracked object

$$[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{v_x}, \boldsymbol{v_y}, \boldsymbol{w}, \boldsymbol{h}]$$

where $(\boldsymbol{x}, \boldsymbol{y})$ is position, $(\boldsymbol{v_x}, \boldsymbol{v_y})$ is velocity and $(\boldsymbol{w}, \boldsymbol{h})$ is the bounding box size.

- Prediction Step: Use a constant velocity motion model to predict the next state:

$$\widehat{\boldsymbol{x}}_{\boldsymbol{k}|\boldsymbol{k}-\mathbf{1}} = \boldsymbol{F}\widehat{\boldsymbol{x}}_{\boldsymbol{k}-\mathbf{1}|\boldsymbol{k}-\mathbf{1}}$$

where $\boldsymbol{F}$ is the state transition matrix.

- Update Step: Fuse measurements from both sensors. Radar provides accurate $[x, y, v_x, v_y]$; vision provides $[x, y, w, h]$ and class. Using this equation to update object state:

$$K = P_{k|k-1}H^T\left(HP_{k|k-1}H^T + R\right)^{-1}$$
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K\left(z_k - H\hat{x}_{k|k-1}\right)$$

where H is the measurement matrix, R is the measurement noise covariance (lower for radar velocity, higher for vision in low light).

- Handle sensor uncertainties: Weight fusion by inverse covariance (e.g., trust radar more in fog, vision in clear conditions).

Output:

Refined tracks with improved accuracy, reduced false positives (filter uncertain radar data using vision confirmation) and enriched attributes (vision's class and radar's speed). This EKF-based method is computationally efficient and handles non-linearities via linearization.

3.2. Behavior prediction

Combining a 24 GHz radar with a camera can support behavior prediction (e.g., "pedestrian will cross," "vehicle will change lane," "object will cut-in"). Radar gives range/velocity and robustness in bad weather, vision gives fine semantics and shape. Together, they're great for both tracking and future-state prediction.

The pipeline of this process is described as below:

Objects detection → A tracker extends object into a trajectory → Enrich trajectory with kinematic → A prediction head (SVM) turns those trajectories into future behavior predictions. With the data from the tracker, behaviors like:

- Pedestrian behaviors: Crossing vs not crossing (intent to cross the road); Walking along the road (parallel movement); Stopping/standing still; Running / sudden crossing

- Vehicle behaviors: Lane keeping (normal driving); Lane change (left/right); Turning at an intersection (left/right turn); Stopping (traffic light, congestion, hazard).

4. CONCLUSIONS

The paper presents a prototype design for RSU AI with the combination of an AI Camera and a 24 GHz Radar sensor (Infineon chipset). By implementing the real model with the combined model and comparing the new model and single models using camera or radar, we have the following conclusions:

Number of test subjects: 10.

Feature	AI Camera	24 GHz Radar	Combined System
Normal conditions	High accuracy	Good for presence detection	Very high, complementary sensing improves reliability
Light rain	10/10	10/10	10/10
Heavy rain or darkness	6/10	9/10	10/10
Object Classification	Excellent with trained models	Basic (cannot distinguish pedestrian vs. cyclist clearly)	Improved classification via sensor fusion

5. LIMITATIONS AND FUTURE WORK

Currently, the article has demonstrated the advantages and basic algorithms when applying radar sensors combined with AI cameras in detecting and identifying people. However, the results are only at the experimental level, with a small number of tests, and need to be further developed with actual testing on roads and weather in different regions. In addition, the development direction combined with intelligent traffic regulation will also need to be developed in the future.

REFERENCES

[1]. Hesham A. Rakha, Kyoungho Ahn, Jianhe Du, and Mohamed Farag. "*Quantifying the Impact of Cellular Vehicle-to-Everything (C-V2X) on Transportation System Efficiency, Energy and Environment*", IEEE Transactions on Intelligent Transportation Systems, (2020).

[2]. Tom Rebbeck, Janette Stewart, Hugues-Antoine, and Andrew. "*Final report for 5GAA on cellular V2X socio-economic benefits*", Ref. 2011027-492, (2017).

[3]. Lejun Jiang, Tamás G. Molnár, and Gábor Orosz. "*On the deployment of V2X roadside units for traffic prediction*", Transportation Research Part C, 129, 103238, (2021).

[4]. Ali R. Abdellah, Ahmed Abdelmoaty, Abdelhamied A. Ateya, Ammar Muthanna, and Andrey Koucheryavy. "*Accurate V2X traffic prediction with deep learning architectures*", IEEE Access, PMCID: PMC11962783, (2018).

A. Ghosal and M. Conti. "*Security issues and challenges in V2X: A survey*", Computer Networks, 169, 107093, (2020). doi:10.1016/j.comnet.2019.107093

[5]. Y. Zhou, S. Liu, M. Xiao, S. Deng, and X. Wang. "*An efficient V2I authentication scheme for VANETs*", Mobile Information Systems, 2018, 1–11, (2018). doi:10.1155/2018/4070283

[6]. Shaoheng Fang, Rui Ye, Wenhao Wang, Zuhong Liu, Yuxiao Wang, Yafei Wang, Siheng Chen, and Yanfeng Wang. "*Federated Learning for Scene Flow Estimation on Roadside Units*", *arXiv preprint*, arXiv:2401.12862, (2024).

[7]. Widhi Yahya, Ying-Dar Lin, Faysal Marzuk, and Piotr Chołda. "*Offloading in V2X with road side units: Deep reinforcement learning*", Computer Communications, 51, 100862, (2025).

[8]. NVIDIA Corporation. "*Jetson Orin NX Module Datasheet*", (2023). [Online]. Available: https://developer.nvidia.com/embedded/jetson-orin

[9]. S. Liu, Y. Zhao, and H. Wang. "*Edge AI computing for intelligent transportation systems: A review and case study*", IEEE Transactions on Intelligent Transportation Systems, 24, 5, 4835–4848, (2023). doi:10.1109/TITS.2022.3149783

A. Ding, L. Xie, and H. Zhang. "*Obstacle detection method based on RSU and vehicle camera fusion using YOLOv5*", Sensors, 23, 10, 4920, (2023). doi:10.3390/s23104920

[10]. S. Gupta et al. "*DeepStream: A system for real-time AI-based video analytics*", NVIDIA Developer Blog, (2022). [Online]. Available: https://developer.nvidia.com/blog/deepstream-real-time-video-analytics/

[11]. R. Sundar and V. Ramachandran. "*Edge AI acceleration in intelligent transportation: A case study using NVIDIA Jetson platform*", Proc. IEEE Int. Conf. Intelligent Transportation Engineering (ICITE), Singapore, pp. 125–130, (2023). doi:10.1109/ICITE56765.2023.10234029

[12]. Infineon Technologies AG. "*BGT24LTR22 XENSIV™ 24 GHz Radar Sensor Datasheet*", Rev. 1.0, (2020). [Online]. Available: Infineon BGT24LTR22 DataSheet

[13]. Rohling. "*Radar CFAR thresholding in clutter and multiple target situations*", IEEE Transactions on Aerospace and Electronic Systems, 19, 4, 608–621, (1983). doi:10.1109/TAES.1983.309393

[14]. Scharf. "*Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*", Addison-Wesley, Boston, MA, ch. 6, (1991).

[15]. S. Wang, J. Zhang, and L. Sun. "*A constant false alarm rate algorithm based on the difference of box filter for radar pulse detection*", IET Radar, Sonar & Navigation, 18, 2, 150–157, (2024). doi:10.1049/rsn2.12573

[16]. R. M. Elnadree and A. El-Sisi. "*Flowchart of SVM algorithm*", in Performance Investigation of Features Extraction and Classification Approaches for Sentiment Analysis Systems, (2021). [Online]. Available: researchgate.net/figure/Flowchart-of-SVM-algorithm_fig3_354303930

[17]. D. Mukherjee, S. Chakraborty, and S. Ghosh. "*Flowchart of support vector machine (SVM)*", in Power

System State Forecasting Using Machine Learning Techniques, (2021). [Online]. Available: researchgate.net/figure/Flowchart-of-support-vector-machine-SVM_fig4_353341688

[18].J. M. Moguerza and A. Muñoz. "*Support Vector Machines with Applications*", *arXiv preprint*, arXiv:math/0612817, (2006).

[19].B. Benjdira et al. "*Car detection using unmanned aerial vehicles: Comparison between Faster R-CNN and YOLOv3*", *arXiv preprint*, arXiv:1812.10968, (2018).

[20].Kumaran, D. P. Dogra, and P. P. Roy. "*Anomaly detection in road traffic using visual surveillance: A survey*", *arXiv preprint*, arXiv:1901.08292, (2019).

TÓM TẮT

Thiết bị RSU thông minh tích hợp trí tuệ nhân tạo với camera và radar 24 GHz để dự đoán và giám sát giao thông theo chuẩn V2X

Bài báo đưa ra một thiết kế hệ thống mới về Smart RSU phù hợp với ứng dụng trong giao thông thông minh trong các thành phố đông đúc kèm mật độ phương tiện dày đặc và đa dạng chủng loại. Dựa trên các công nghệ có sẵn bao gồm V2X, Radar và AI camera, hệ thống smart RSU thông minh có thể tự đưa ra các quyết định riêng biệt để xử lý trong các tình huống giao thông phức tạp cũng như điều tiết giao thông. Với mô hình học máy kết hợp với các sensor có độ chính xác cao, hệ thống có khả năng học từ "hành vi giao thông địa phương" để không chỉ phát hiện vi phạm mà còn dự đoán trước các hành vi nguy hiểm đặc trưng của từng khu vực và chủ động can thiệp vào hệ thống điều khiển giao thông. Việc kết hợp giữa Radar và Camera cũng giúp tăng độ chính xác hơn trong việc giám sát và thu nhập dữ liệu trong thực tiễn. Điều này giúp cho nâng cao hiệu suất quản lý giao thông, cải thiện tình trạng tắc nghẽn và điều tiết giao thông khi có sự cố xảy ra. Nghiên cứu phù hợp với ứng dụng trong các thành phố thông minh, đặc biệt là khi kỷ nguyên số và thông tin đang ngày một phát triển.

Từ khoá: RSU thông minh; Truyền thông V2X; Ra đa sensor 24 GHz IOT; AI camera; Dự đoán lưu lượng giao thông; Hệ thống giao thông thông minh; Thành phố thông minh.