

Technical Analysis of Autonomous Driving in Complex Road Conditions Based on Multimodal Perception Fusion

Yiming Xie

{4202350301428@stu.tyu.edu.cn}

Department of Electromechanical and Vehicle Engineering, Taiyuan University, Taiyuan, 030032, China

Abstract: With the continuous development of autonomous driving technology, single-modality perception has been unable to meet current development needs. Therefore, multi-modality perception fusion has become an important technology to ensure the safe driving of vehicles. This paper mainly analyzes the defects of single-modality perception and the feasibility of multi-modality perception in complex environments. In addition, this paper discusses the fusion methods and novel algorithms based on the data layer, feature layer, and decision-making layer of the auto drive system in recent years, and raises some questions. Utilizing these studies, this paper explains how multi-modality perception fusion overcomes the impact of complex environments on vehicle driving. Finally, this paper summarizes the current technological bottlenecks and limitations of autonomous driving and proposes future development directions and challenges.

Keywords: Intelligent driving, multimodal perception, data fusion, feature fusion, decision fusion

1 Introduction

Autonomous driving (AD) can increase safety by avoiding issues such as fatigue driving and drunk driving, while also providing passengers with the most comfortable driving experience, especially during long-distance travel. Nowadays, this field has become an important indicator of the digital age, with broad application prospects and commercial value. Many countries regard it as a key research direction. However, it is worth noting that AD faces many risks and challenges in its development process. Due to the obvious defects of a single sensor, AD fails to function in many harsh environments or extreme scenarios: driving in rainy, snowy, or hazy environments, temperature and external noise often cause confusion in LiDAR perception, making it difficult for cameras to capture signals and resulting in a serious decline in image quality, greatly increasing the probability of accidents.

In contrast, the advantages of Multimodal Perception Fusion (MPF) are very obvious[1]. MPF is the fusion of data from multiple sensors and different characteristics to achieve accurate perception of the environment by AD. The advantage of MPF lies in compensating for the shortcomings of a single sensor, improving system robustness, perception accuracy, and safety. Wang's team[2], proposed a framework that combines features and dynamics to address the issue of low pedestrian recognition accuracy in complex scenarios. After comparing the data obtained

using this framework with the benchmark dataset, the high timeliness of multimodal fusion compared to single-modal perception was verified. With the rapid development of artificial intelligence algorithms such as deep learning, the efficient fusion of multimodal data has strong technical support[3]. Yin team proposed a 3D object detection algorithm based on multi-scale feature weighted point-by-point fusion by corresponding camera images and LiDAR point clouds point-by-point[4]. The experimental results on the KITTI object detection benchmark dataset ultimately proved that the system can better solve the problems faced by individual perception through intelligent algorithms, and improve the robustness and safety factor of the system.

This article analyzes the application of multimodal perception fusion in complex road conditions for intelligent driving. Firstly, this article outlines the core characteristics and perception requirements of different complex road conditions. Meanwhile, this article analyzes the technical characteristics and fusion logic of important sensors, and summarizes the key technologies and algorithm frameworks for data fusion. Finally, this article also explores the challenges and future development trends faced by current research, providing a reference for further optimization of complex road condition perception technology for intelligent driving.

2 Core characteristics and perception requirements of complex road conditions

For intelligent driving, Complex Road Condition (CRC) refers to driving scenarios where there are multiple interfering factors on non-ideal roads that hinder the perception and judgment of intelligent driving. The scenarios are diverse, such as rainy, foggy, snowy weather, high unknowns, such as emergency overtaking or pedestrians crossing the highway, and are greatly affected by the environment, such as blurred cameras. Therefore, AD requires higher real-time response and accuracy to accurately capture targets and cope with the joint influence of multiple targets. At the same time, improving robustness is also crucial. Fig. 1 shows the imaging differences of the same location under different degrees of illumination and the influence of rain or fog. Therefore, developing effective image restoration techniques to address the pain points currently faced by sensors is necessary to provide stable support for AD[5].



Fig.1 Images in different weather [6]

In CRC road environments, a single sensor often loses accuracy due to factors such as lighting, weather, and occlusion. On the contrary, MPF fusion can provide a more comprehensive and robust environmental perception result by combining data from multiple sensors such as cameras, LiDAR, millimeter wave radar, ultrasonic radar, GPS/IMU, etc., thereby improving

the safety and reliability of AD. From Fig.2, it can be seen that MPF integrates environmental data from multiple sensors for feature fusion, data fusion, and decision fusion. The system complements missing or differential information in the fusion of three levels, making the AD perception system more reliable and secure.

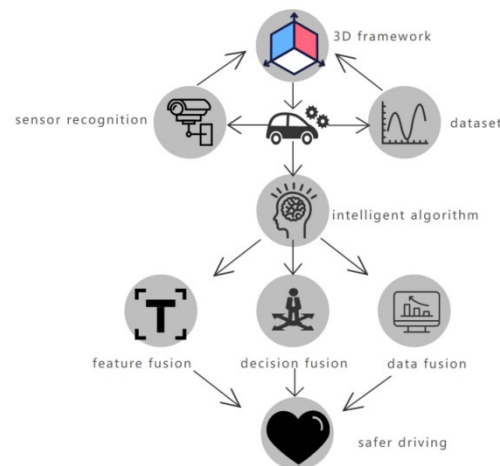


Fig.2 Operation Logic of Autonomous Driving with Multi-modal Perception Fusion (Photo/Picture credit: Original).

3 Analysis of the Technical Characteristics of the Autonomous Driving MPF Sensor

3.1 Mainstream types of perception sensors

Perception sensors are the environmental perception organs of vehicles. Vehicles recognize objects and collect data through sensors, which are processed into an important basis for vehicle decision-making during the driving process. The mainstream perception sensors are mainly ultrasonic radar, laser radar, cameras, etc., and sometimes infrared sensors and thermal imagers are also used. Each sensor, as shown in Table 1, has its own function to jointly ensure driving safety.

Table 1. Main sensors for AD

Types	Applications and Advantages of Sensors
Ultrasonic radar	Low cost, mainly suitable for scenarios such as reversing and obstacle warning
radar	Measurement accuracy reaches the centimeter level, which helps to construct three-dimensional spatial structures and is not affected by lighting
camera	Strong visual extraction ability, able to quickly identify and capture

	targets
Infrared sensor	Clear detection of targets at night, compensating for the shortcomings of cameras in nighttime imaging
GPS	Low cost, wide range, providing global positioning for vehicles, and supporting path planning
Thermal imaging instrument	Strong resistance to adverse weather interference, capable of detecting obstacles and pedestrians in any climate

3.2 Limitations of a single sensor

The positioning and map construction system using a single sensor has significant drawbacks and limitations. Light can easily affect the sensing of visual sensors, especially cameras, causing them to become disordered. Lidar sensors are prone to imaging drift and cannot obtain environmental color, texture, and other information in environments with relatively smooth ground, such as squares, due to the lack of a reflection point[7]. At the same time, the infrared images obtained by infrared sensors have defects such as low resolution, poor contrast, and blurred edges[8]. Secondly, public datasets based on a single sensor lack remote annotation, making it difficult to effectively construct a 3D system framework and serve as the foundational neural network for AD in scenarios such as highways[9]. It is obvious that a simple stack of single sensors cannot guarantee driving safety in terms of accuracy, real-time performance, and robustness, while the complementary advantage of MPF can overcome the problems encountered by a single device.

3.3 Analyzing the complementarity of composite sensors

The complementarity of composite sensors refers to the synergistic effect of different types of sensors to compensate for the shortcomings of a single sensor, ensure complete and accurate data, clear imaging, and enhance the robustness of the system. Wei team[10]analyzed the characteristics of different sensors, combined the millimeter wave radar with the vision sensor, calibrated different data and information, significantly improved the perception ability of AD, and made the vehicle safer to drive in a complex environment. The Kruttdipta team[11]proposed a spatiotemporal sampling algorithm based on the problem of high power consumption during the independent operation of LiDAR. This method utilizes a combination of vision LiDAR to achieve target locking and path planning, greatly reducing the system's operational load while ensuring operational quality. The Liu team's system[12]is mainly based on radar and supplemented by cameras in terms of hardware, using a joint probability function algorithm to provide more accurate environmental information for AD and ensure that the system makes better decisions. The DeMas-Giménez team[13]integrates various sensors and positioning systems such as high-resolution solid-state LiDAR, visible light, and thermal sensors. By providing complementary and redundant information through each component, they quickly obtain data that is closest to the real environment, verifying the complementarity of composite sensors from multiple perspectives. In short, the complementarity of composite sensors provides feasibility for MPF, and how to achieve maximum information coordination through the key technologies and fusion algorithms of MPF has become the core issue to break through the bottleneck of AD development today.

4 Key Technologies and Fusion Algorithms of MPF

4.1 Type of fusion

The underlying logic of MPF is to integrate data from multiple sensors and perform leak detection to overcome problems that cannot be solved by a single sensor, and the key to solving these problems lies in optimizing the fusion results. Therefore, ensuring the rationality of fusion-type classification is very important. Based on the process of information processing, the system needs to integrate data, features, and decisions from three levels to ensure the security of AD while ensuring both efficiency and accuracy of information.

Data fusion.•When a single sensor encounters extreme weather conditions such as sandstorms, data collection will face significant obstacles. For example, camera imaging will appear blurry and noisy, while laser pulses emitted by LiDAR will be absorbed or dispersed by atmospheric particles, resulting in the inability of vehicles to effectively receive reflected information and tracking failure[14].On the contrary, multimodal data fusion (MDF) uses the data collected by different sensors from multiple angles to provide more accurate target location and information for AD [15], which can not only improve target detection accuracy but also enhance system robustness[16].In order to ensure the integrity of features and the real-time performance of algorithms, the Lv team integrates data from various image information acquisition terminals[17], such as radar, and uses memory converters for efficient data encoding, reducing the difference between 3D imaging and the real environment. The Chen team combined visual and laser data and developed the PSPNet_CBAM network framework based on the performance of the PSPNet network in accurately labeling each pixel[18]. This new framework can automatically focus on key areas of the image and reduce the interference of irrelevant information. The team compared the recognition performance of different targets, for example, the PSPNet_CBAM framework increased the recognition rate of riders by about 3 percentage points and the recognition rate of trains by 7 percentage points. These data validate that the PSPNet_CBAM framework can improve the performance of real-time positioning and map construction in the system, enabling the system to better interpret complex environments. However, it is worth noting that the recognition rate of the framework for walls has decreased by 3 percentage points. The reason for this result is not yet known, and no relevant solutions have been proposed, which threatens the safety of autonomous driving.

Feature fusion.•Feature fusion refers to the integration of feature information extracted from different feature extractors into high-precision features using a specific algorithm. Due to the limitations of a single type of feature, the system requires multiple complementary features to enhance robustness. Therefore, the accuracy of target recognition is crucial for AD, but on the other hand, although feature accumulation is important for recognition, the system is often affected by redundant and similar features and loses its effectiveness. Therefore, how to collect and fuse effective features appropriately and efficiently is the core issue of feature layer fusion[19].The Wang team proposed a multi feature progressive interactive network for the task of predicting multimodal trajectories[20]. Through a phased progressive feature interaction mechanism, it achieves deep coupling between traffic participants and road constraints in dynamic scenes. By comparing the use of trajectory multimodal decoders as shown in Fig.3, the predicted trajectories are more closely aligned with real road conditions. In order to ensure the safe operation of the AD system in complex environments and sufficient response speed, Xu

team established a multi-sensor feature fusion model, which uses a sparse channel pooling mechanism (deep learning sparsifies feature channels[21], pooling processing only selects messages from key channels) to avoid the influence of irrelevant information. At the same time, the residual triple preservation mechanism is used to preserve the original features of the target and obtain key features from other perspectives, solving the problem of feature misalignment and improving the computational efficiency of the system.

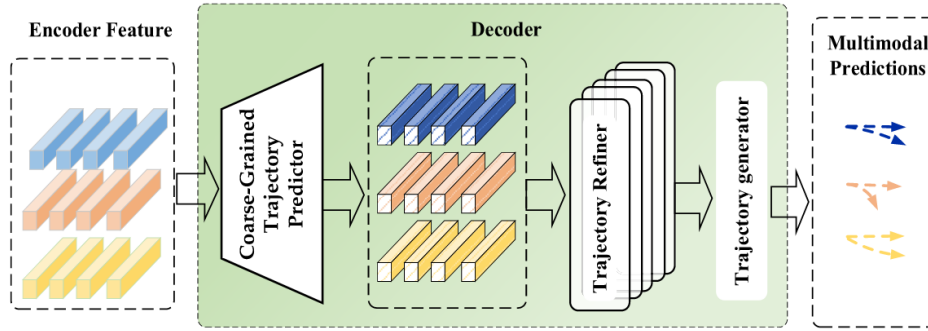


Fig.3 Structural diagram of trajectory multimodal decoder[20]

Decision fusion. • The decision layer fusion of the AD system is very important, and the decisions made by the system are closely related to the personal safety of passengers. Therefore, ensuring the unity and reliability of the system's conclusion output while improving output efficiency cannot be ignored[22]. The Wang team[23], proposed the Edge Enhanced Graph Attention Reinforcement Learning (EGARL) framework to enhance the interaction between driving participants, including computing devices, data processing units, and traffic scenes, and make the most reasonable driving decisions. By comprehensively modeling the interaction between participants, accidents are minimized, and the safety of AD driving is improved. The Zhao team[24], proposed a Bayesian network decision model based on calculating the probability of danger to address the potential threat of extreme weather conditions to the system. For example, when the system is in a strong wind environment, even with sparse pedestrian flow and no other adverse environmental factors such as sandstorms, the probability of accidents occurring in autonomous driving is still as high as 75%. This model can extract safety elements of dangerous scenarios for 3D modeling and determine the safest driving behavior based on actual situations. This model significantly reduces the driving safety factor, but its reference variables are not comprehensive, as they include only ten different environments. In the future, researchers need to add more variables to further validate the performance of Bayesian network decision models.

4.2 Fusion algorithm

Executing emergency braking when encountering special situations during AD is the last resort. Optimizing system algorithms and solving algorithm uncertainties are the most important measures to ensure the lives of passengers[25]. In order to solve the problem of excessive burden on traditional algorithms in processing 3D LiDAR point cloud data, the Wang team[26] proposed

an instance segmentation fusion (ISF) - PointPillars algorithm, which uses a 2D object detection network RGB (red, green, blue) to obtain the contours of objects in the scene. Subsequently, the contour is extended into three-dimensional space to further extract target features and remove redundant data to improve the accuracy and efficiency of the model. The Yuan team proposed a fusion tracking algorithm for visible light and infrared images based on the Visible Light and Infrared Image Fusion Network (VIF net) and SiamRPN to improve the perception accuracy of the AD system[27]. This algorithm divides the object tracking process into two parts: image fusion and target tracking, which is of great value in consolidating the AD security defense line. The Pan team proposed using the YOLOv5 algorithm, which can recognize the location and type of targets in a single scan, to solve the problem of low efficiency in MPF[28]. By using the YOLOv5 algorithm, the detection task can be transformed into an end-to-end regression problem, which simply involves inputting raw data and directly outputting the final result without the need for manual intermediate steps, increasing the operational efficiency of MPF and reducing system recognition errors.

5 Prospect

With the development and research of MPF, the environmental cognition of AD will become more precise, and decision-making will be more secure, but it still faces many limitations in the application of complex environments. Firstly, the development of general artificial intelligence has been stagnant for a long time. Currently, artificial intelligence cannot think independently and automatically adapt to different tasks like humans. At the same time, the logic of current artificial intelligence is poor, and it cannot combine knowledge from different fields. Another huge obstacle is the problem of handling combination scenarios. AD systems face slow processing speeds when dealing with single complex scenarios, such as rainy days. At the same time, the system's ability to respond to unexpected situations, such as overtaking by other vehicles on the road, is insufficient. When the two are combined, the load on the system is too large, and the system's decisions and calculations will be affected, seriously threatening driving safety. These technological bottlenecks will hinder the further development of AD, and security risks will increase dramatically.

The development of AD faces not only technical issues, but also significant problems at the application level that cannot be ignored. Although the existing laser radar camera (LC) fusion methods have high accuracy, the sensors themselves are easily affected by extreme weather, resulting in data problems in the input stage, and the cost of laser radar is higher compared to microwave radar. At the same time, the existing AD database also has many problems. The most obvious issue is that the number of databases is too small and mainly concentrated in more developed regions, with almost none in low-income areas. Secondly, the purchase and research costs of the database are also very high, and there are often cases of missing bids in the database. Due to the different data in different environments, the adaptability of the existing database is insufficient, and it cannot be fixed once and for all.

Even if existing technological bottlenecks and application limitations are overcome, AD still has significant issues that cannot be ignored. In emergency situations, the system will face many decision-making difficulties, such as whether to prioritize protecting the occupants of the vehicle or pedestrians on the road, protecting the majority or the minority. There is no unified

answer to these questions, and due to the different values of different countries and populations, decision-making will also be affected. At the same time, there is no clear evaluation standard for whether responsibility should be attributed to passengers in the event of an accident, and it is difficult to have a standard that can achieve absolute fairness. At the same time, passengers and even pedestrians also face the problem of privacy breaches. When various sensors collect data, it inevitably infringes on personal privacy. Important information, such as passengers' physiological activities or commercial secrets, can easily be recorded by the system and even used for illegal purposes. So in the future development path, it is necessary to balance these issues and improve corresponding laws and regulations, otherwise system defects will lead to incalculable consequences.

6 Conclusion

This article mainly summarizes the basic framework of MPF and the rationality of multimodal compared to single modal. Through research, it has been found that multimodal fusion has high robustness, real-time performance, and accuracy, proving that MPF meets the current development needs. At the same time, regarding the issue of the influence of redundant features on feature fusion, this paper analyzes the sparse channel pool mechanism and the multi-feature progressive interaction network to avoid the interference of redundant information. This article also analyzes the operational logic of the PSPNet_CBAM network framework, demonstrating its improvement in robustness. On the other hand, this article found that the recognition rate of this framework for certain objects will decrease. At the same time, this article also discovered the deficiency of insufficient variables in Bayesian network decision models, providing a development direction for future decision fusion. In addition, this article also raises issues such as the slow development of artificial intelligence and difficulties in assigning responsibility for future autonomous driving. Finally, the development directions regarding technology and law were provided.

References

- [1] Yuan, W.: Graph neural network-based multimodal sensor fusion for robust autonomous driving perception. In: Proceedings of the 2nd International Conference on Intelligent Transportation and Smart Cities (ICITSC) (2025) DOI:10.1117/12.3073578
- [2] Wang, C.M., Xie, H.: FCD-Net: Feature decorrelation and confidence-driven dynamic fusion for robust pedestrian recognition in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*. 26(11), 19468–19480 (2025) DOI: 10.1109/tits.2025.3596166
- [3] Xing, J.M., Wei, D.W., Zhou, S.H., et al.: A comprehensive study on self-learning methods and implications to autonomous driving. *IEEE Transactions on Neural Networks and Learning Systems*. 36(5), 7786–7805 (2025) DOI: 10.1109/TNNLS.2024.3440498
- [4] Yin, C.F., Qu, H.C., Li, Y.C.: 3D object detection algorithm for autonomous driving based on multi-scale feature weighted point-by-point fusion. *Journal of Intelligent & Fuzzy Systems*. 49(4), 933–947 (2025) DOI:10.1177/18758967251356865
- [5] Yin, S., Liu, H.: A multimodal contrastive and transfer learning-based image restoration model for multiple adverse weather driving scenes. *IEEE Signal Processing Letters*. 32, 3745–3749 (2025) DOI:10.1109/lsp.2025.3611705

- [6] Wu, W.S., Deng, X.H., Jiang, P., et al.: CrossFuser: Multi-modal feature fusion for end-to-end autonomous driving under unseen weather conditions. *IEEE Transactions on Intelligent Transportation Systems*. 24(12), 14378–14392 (2023) DOI: 10.1109/TITS.2023.3307589
- [7] Wang, W.Q.: Intelligent vehicle localization and mapping based on multi-sensor fusion. Ph.D. thesis, Dalian University of Technology (2023) DOI:10.26991/d.cnki.gdllu.2023.001458
- [8] Wang, J.Y.: Research on multi-scale deep learning fusion methods for infrared and visible light images. Master's thesis, Taiyuan University of Science and Technology (2022) DOI:10.27721/d.cnki.gyzjc.2022.000194
- [9] Matuszka, T., Barton, I., Butykai, Á., et al.: aiMotive dataset: A multimodal dataset for robust autonomous driving with long-range perception. *arXiv preprint* (2022) DOI:10.48550/arXiv.2211.09445
- [10] Wei, Z.Q., Zhang, F.K., Chang, S., et al.: MmWave radar and vision fusion for object detection in autonomous driving: A review. *Sensors*. 22(7), 2542 (2022) DOI:10.3390/s22072542
- [11] Samal, K., Kumawat, H., Saha, P., et al.: Task-driven RGB–LiDAR fusion for object tracking in resource-efficient autonomous systems. *IEEE Transactions on Intelligent Vehicles*. 7(1), 102–112 (2022) DOI:10.1109/TIV.2021.3087664
- [12] Liu, Z., Cai, Y.F., Wang, H., et al.: Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions. *IEEE Transactions on Intelligent Transportation Systems*. 23(7), 6640–6653 (2022) DOI:10.1109/TITS.2021.3059674
- [13] deMas-Giménez, G., Subirana, A., García-Gómez, P., et al.: Multimodal sensing prototype for robust autonomous driving under adverse weather conditions. In: *Proceedings of SPIE Optical Metrology* (2025) DOI:10.1117/12.3062388
- [14] Han, L.Y., Wang, J., Li, C., et al.: A novel multi-object tracking framework based on multisensor data fusion for autonomous driving in adverse weather environments. *IEEE Sensors Journal*. 25(9), 16068–16079 (2025) DOI:10.1109/JSEN.2025.3550506
- [15] Wang, Y.Q., Xu, L.F., Chen, W.: A review of target detection and recognition in multimodal data. *Industrial Control Computer*. 38(10), 90–91, 94 (2025)
URL: <http://dianda.cqvip.com/Qikan/Article/Detail?id=7202057097>
- [16] Wu, J., Zhang, H.D.: The rise of intelligence, artificial intelligence, and information paradigms. *Research on Natural Dialectics*. 40(8), 56–65 (2024) DOI:10.19484/j.cnki.1000-8934.2024.08.008
- [17] Li, Z., Chen, G., Lv, K.F.: Real-time localization and navigation method for autonomous vehicles based on multi-modal data fusion by integrating memory transformer and DDQN. *Image and Vision Computing*. 156 (2025) DOI:10.1016/j.imavis.2025.105484
- [18] Chen, N., Wei, D., Lin, D.S., et al.: Semantic SLAM using laser–vision data fusion: Enhancing autonomous navigation in unstructured environments. *Alexandria Engineering Journal*. 127, 606–618 (2025) DOI:10.1016/j.aej.2025.05.015
- [19] Xu, W.F., Meng, X.W., Hu, Y., et al.: RCFusionNet: A radar-guided two-stage fusion approach to effective multimodal perception. *IEEE Sensors Journal*. 25(17), 33890–33900 (2025) DOI:10.1109/JSEN.2025.3592236
- [20] Wang, H.H., Jin, L.S., Sun, X.Y., et al.: MFPINet: Multi-feature fusion with progressive interaction for trajectory prediction in autonomous driving. *Engineering Research Express*. 7(3), 035243 (2025) DOI:10.1088/2631-8695/adf59e
- [21] Xu, Q.Q., Guo, J.H., Ju, Y.N.: Research on vehicle safety based on multi-sensor feature fusion for autonomous driving tasks. *Journal of Intelligent Transportation Systems* (2025) DOI:10.1080/15472450.2025.2543825

- [22] Yang, Z.Y., Zhang, L.L., Bian, Y.G., et al.: Decision making for highway autonomous driving using hybrid reinforcement learning. *Journal of Control and Decision* (2025) DOI:10.1080/23307706.2025.2521782
- [23] Qiang, Y.C., Wang, X.L., Liu, X.T., et al.: Edge-enhanced graph attention network for driving decision-making of autonomous vehicles via deep reinforcement learning. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*. 239(4), 1168–1180 (2025) DOI:10.1177/09544070231217762
- [24] Zhao, Y., Zhu, Y., Zhao, L., et al.: Behavioral decision-making and safety verification approaches for autonomous driving systems in extreme scenarios. *Journal of Systems and Software*. 226 (2025) DOI:10.1016/j.jss.2025.112385
- [25] Zhang, R.H., Sun, C., Valiollahimehrizi, R., et al.: An uncertainty-aware, dual-tiered decision-making method for safe autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*. 26(1), 691–702 (2025) DOI:10.1109/TITS.2024.3492794
- [26] Wang, Y.S., Han, X.B., Wei, X.X., et al.: Instance segmentation frustum-pointpillars: A lightweight fusion algorithm for camera – LiDAR perception in autonomous driving. *Mathematics*. 12(1), 153 (2024) DOI:10.3390/math12010153
- [27] Yuan, Q., Shi, H.X., Tan, Y.X.A., et al.: Enhanced target tracking algorithm for autonomous driving based on visible and infrared image fusion. *Journal of Intelligent and Connected Vehicles*. 6(4), 237 – 249 (2023) DOI:10.26599/JICV.2023.9210018
- [28] Pan, H.H., Wang, X.P.: A YOLOv5 algorithm for intelligent autonomous driving improvement with multi-sensor fusion. In: *Proceedings of an International Conference* (pp. 369–379) (2025) DOI:10.1007/978-981-96-7332-2_34