Computer Vision–Centric Digital Twins for Intelligent Mobility: Simulation of Object Detection and Scene Segmentation

Reamy Womodowe Achirobe[1], Derrick Appiah Osei[2]

[1]Department of Mechanical Engineering, Kwame Nkrumah University of Science and Technology, Ghana
rachirobe@gmail.com
[2]Drexel University, U.S.A, derrickappiahosei@gmail.com

Abstract:
Digital twins (DTs) are rapidly emerging as a foundational paradigm for intelligent mobility systems, enabling high-fidelity virtual representations of physical transportation environments. In parallel, computer vision, particularly object detection and scene/semantic segmentation remain central to perception, decision-making, and safety in autonomous and connected vehicles. This review synthesizes recent advances at the intersection of digital twins and computer vision, with a focus on how DT-based simulation frameworks support the development, validation, and deployment of vision algorithms for intelligent mobility. Drawing on recent studies in DT-assisted autonomous driving, cooperative perception, simulation-based perception evaluation, and data-driven DT architectures, the paper reviews methodologies for generating synthetic visual data, closing the sim-to-real gap, and embedding vision models within cyber–physical feedback loops.

Keywords: digital twins, deep learning, intelligent transport systems, CARLA

1.0 Introduction

Intelligent mobility systems increasingly depend on visual perception to interpret complex, dynamic, and safety-critical traffic environments. Core computer vision tasks, most notably object detection and scene/semantic segmentation, enable autonomous and connected vehicles to identify road users, recognize infrastructure elements, estimate drivable space, and reason about contextual cues such as traffic signals, signage, and lane markings (Osei and Aliu, 2026). These perception capabilities underpin higher-level functions, including trajectory planning, collision avoidance, and cooperative driving, making vision a central pillar of intelligent transportation systems (ITS). Recent advances in deep learning have substantially improved perception accuracy; however, real-world deployment continues to expose vulnerabilities related to environmental variability, rare events, and domain shifts across locations and operating conditions (Muhammad et al., 2022).

A fundamental setback in vision-based ITS is data availability and validation. Training and testing robust object detection and segmentation models require large volumes of labeled data that capture the full diversity of traffic scenarios, weather conditions, lighting variations, and sensor configurations. Collecting such data in real environments is costly and time-consuming, while manual annotation remains labor-intensive and error-prone. More critically, many safety-relevant scenarios, such as near-collisions, abnormal vehicle behavior, or extreme weather, are rare by nature and ethically difficult to capture at scale. Privacy regulations and safety constraints further limit the deployment of large-scale sensing infrastructures in public spaces, constraining the availability of representative datasets for perception research (Hernández-Martínez et al., 2021).

Digital twins (DTs) have emerged as a powerful paradigm to address these challenges. In the context of intelligent mobility, a digital twin is commonly defined as a synchronized virtual replica of a physical transportation system, integrating real-time data streams, high-fidelity simulation, and analytics to mirror and predict system behavior. Recent studies emphasize a shift from static, model-based twins toward data-driven and AI-enabled digital twin architectures, where machine learning continuously updates the virtual representation based on observed physical-world dynamics (Babaei et al., 2025). This evolution has positioned DTs not only as tools for monitoring and control, but as active experimentation and development environments for intelligent mobility systems.

Crucially, recent research demonstrates that digital twins can function as computer vision development platforms, enabling large-scale simulation of visual scenes under controlled yet realistic conditions. By leveraging advanced rendering engines, physics-based modeling, and sensor simulation, DTs can generate synthetic camera data that reflects diverse traffic densities, road geometries, weather patterns, and lighting conditions. Studies using simulation frameworks such as CARLA and Unreal Engine show that DT-generated imagery can be systematically varied to probe the robustness of object detection and scene segmentation models in ways that are infeasible in real-world testing (Liu et al., 2023). These capabilities are particularly valuable for stress-testing perception systems against corner cases such as occlusions, glare, night-time driving, or complex multi-agent interactions.

Beyond data generation, vision-centric digital twins enable closed-loop integration between perception, simulation, and decision-making. For example, abnormal vehicle behavior detected through vision-based tracking can be projected into a digital twin to reconstruct trajectories, evaluate risk, and test alternative control strategies in silico (Zhang et al., 2021). Similarly, DT-assisted intelligent transportation systems have been shown to alleviate the reliance on extensive real-field data by combining limited real observations with large-scale synthetic datasets, while preserving performance accuracy (Hernández-Martínez et al., 2021). These approaches highlight the potential of DTs to act as bridges between perception algorithms and system-level mobility intelligence. The relevance of vision-centric DTs is further amplified in connected and cooperative mobility scenarios. Recent work on local digital twins at smart intersections illustrates how roadside perception, infrastructure sensing, and vehicle-mounted cameras can be fused within a digital twin to enhance object detection and scene understanding beyond the line of sight of individual vehicles (Wang et al., 2024). Such systems support hybrid autonomous driving modes and cooperative perception, where digital twins mediate information sharing and predictive reasoning across agents. These developments align with broader visions of parallel intelligence, in which physical and digital systems co-evolve through continuous feedback (Wang et al., 2023).

2.0 Digital Twins in Intelligent Mobility

Digital twins (DTs) in intelligent mobility are commonly defined as cyber–physical systems composed of three tightly coupled layers: the physical layer (vehicles, roads, sensors, and traffic participants), the digital layer (virtual models, simulations, and analytics), and the communication layer that enables bidirectional data exchange between the physical and digital domains. This layered architecture allows real-time sensor data to continuously update the digital representation while insights generated in the virtual space inform physical-world operations. Recent work emphasizes a transition from static digital replicas to data-driven and AI-enabled digital twins, particularly for autonomous driving applications, where machine learning models enable continuous adaptation of the digital twin to real-world dynamics (Babaei et al., 2025).

In intelligent transportation systems, DTs have been applied to a wide range of use cases, including traffic monitoring, vehicle–infrastructure interaction, cooperative perception, and autonomous driving. For example, DT-based traffic analysis frameworks replicate vehicle trajectories and interactions to detect abnormal behavior and evaluate safety risks (Zhang et al., 2021). Similarly, local digital twins deployed at smart intersections integrate roadside sensors and vehicle data to enhance situational awareness beyond the sensing capabilities of individual vehicles (Wang et al., 2024). A key enabler of these applications is the development of multi-modal simulation frameworks that leverage game engines, ray tracing, and physics-based modeling to generate realistic traffic environments. Platforms such as CARLA and Unreal Engine–based simulators support accurate rendering of road geometry, traffic actors, lighting, and weather, while modeling camera optics and sensor noise (Liu et al., 2023). These capabilities provide the foundation for embedding computer vision pipelines directly into DT workflows, enabling scalable development, evaluation, and deployment of perception-driven intelligent mobility systems.

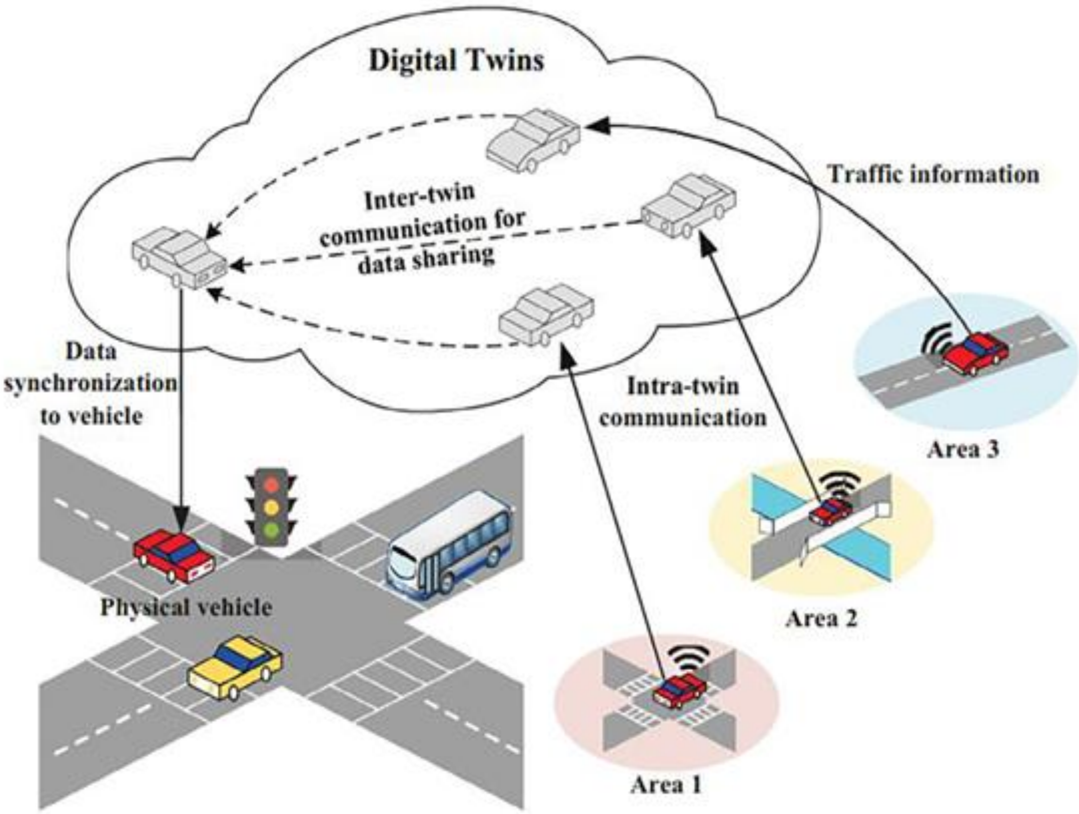

Figure 1: Digital Twins in Intelligent Mobility (Liu et al., 2023)

3.0 Computer Vision for Intelligent Transportation

3.1 Object Detection

Object detection is a fundamental perception task in intelligent transportation systems, enabling vehicles and infrastructure to identify and localize traffic participants and obstacles such as vehicles, pedestrians, cyclists, and roadside objects. Deep learning–based detectors, including YOLO, Faster R-CNN, and transformer-based architectures, have become the dominant approaches due to their strong performance in complex visual scenes. However, detection accuracy in real-world traffic environments remains highly sensitive to camera configuration, resolution, lighting conditions, weather, and traffic density (Muhammad et al., 2022).

One of the main challenges in deploying robust object detection systems is the limited availability of representative training and validation data. Real-world data collection is expensive and constrained by safety, privacy, and the rarity of critical edge cases. Digital twins address these limitations by enabling simulation-based evaluation and data generation under controlled yet realistic conditions. DT-generated imagery allows systematic variation of environmental factors such as day–night transitions, adverse weather, and sensor placement, facilitating performance analysis that is difficult to achieve through field testing alone (Liu et al., 2023).

Several studies demonstrate that synthetic visual data generated by digital twins can augment real-world datasets and improve object detection robustness. Hernández-Martínez et al. (2021) show that DT-based data generation can significantly reduce the reliance on extensive field data while maintaining detection accuracy. Beyond offline training, object detection outputs can be integrated into closed-loop DT frameworks, where detected objects update the digital twin's traffic state and support predictive simulation and decision-making. As a result, object detection increasingly functions as a core component of perception-aware digital twins for intelligent mobility.

3.2 Scene and Semantic Segmentation

Scene and semantic segmentation provide pixel-level understanding of traffic environments, supporting lane detection, drivable-area estimation, and contextual reasoning in intelligent transportation systems. Unlike object detection, segmentation captures fine-grained spatial structure essential for safe navigation and planning (Muhammad et al., 2022). Recent advances in deep learning, including convolutional and attention-based networks, have significantly improved segmentation performance in urban driving scenarios (Muhammad et al., 2022). However, achieving robust generalization remains challenging due to sensitivity to lighting, weather, road surface appearance, and camera configurations. Segmentation models often perform well on benchmark datasets but struggle with rare events and domain shifts encountered in real-world deployment (Muhammad et al., 2022). Safety-critical scenarios such as construction zones, heavy rain, or snow-covered roads are typically underrepresented in training data. Digital twins address these limitations by enabling controlled generation of pixel-level labeled data under diverse conditions (Liu et al., 2023). Simulation environments allow systematic evaluation under adverse and uncommon scenarios that are difficult to capture in real-world data collection. Digital twins further support iterative validation and refinement of segmentation models through closed-loop testing (Babaei et al., 2025). By embedding segmentation pipelines within digital twin workflows, researchers can identify failure modes and improve robustness. These capabilities position vision-centric digital twins as essential tools for advancing reliable scene understanding in intelligent mobility systems.

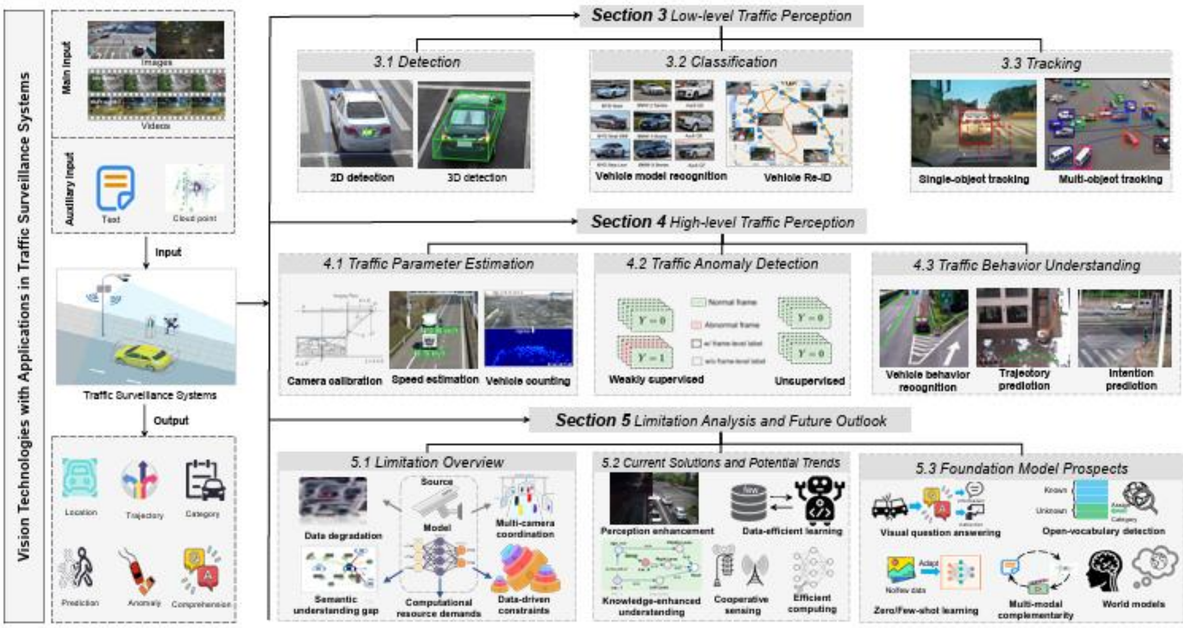

Figure 2: Computer Vision for Intelligent Transportation (Liu et al., 2023)

4.0 Digital Twins as Vision Simulation Platforms

Digital twins (DTs) are increasingly used as vision simulation platforms that integrate high-fidelity rendering, physics-based modeling, and real-time data synchronization to support perception development in intelligent mobility systems. By coupling virtual environments with sensor and traffic dynamics, DTs provide controlled yet realistic settings for testing vision algorithms. These platforms enable systematic variation of environmental factors that are difficult to isolate in real-world experiments. DT-based simulation supports scalable experimentation without safety or privacy constraints. Recent studies highlight DTs as essential for perception benchmarking and stress testing. The integration of vision pipelines within DTs enables end-to-end evaluation of perception-driven decision-making. Collectively, DTs shift perception development from data-limited field testing to reproducible, simulation-centered workflows (Babaei et al., 2025; Liu et al., 2023).

4.1 Synthetic Data Generation

Digital twins enable large-scale generation of labeled synthetic data for computer vision tasks, reducing dependence on costly real-world annotation. Simulation environments such as CARLA and Unreal Engine–based frameworks accurately replicate camera calibration, vehicle dynamics, weather conditions, and lighting variations. These platforms provide precise ground-truth annotations at pixel, object, and instance levels. Studies show that DT-generated datasets can augment real-world data and improve model robustness. Synthetic data supports coverage of rare and safety-critical scenarios. When carefully designed, DT datasets yield competitive performance on real-world perception tasks. As a result, synthetic data generation has become a core application of vision-centric digital twins (Hernández-Martínez et al., 2021; Liu et al., 2023).

4.2 Sim-to-Real Transfer

A central challenge in DT-based vision is ensuring that models trained in simulation generalize effectively to real environments. Performance gaps arise due to differences in texture realism, sensor noise, and environmental complexity. Research shows that reducing this domain gap through accurate camera modeling, realistic rendering, and scenario-specific tuning significantly improves sim-to-real transfer. Data-driven digital twins calibrated with real sensor data further enhance generalization. Site-specific DTs tailored to intersections or corridors outperform generic simulators. Hybrid training using real and synthetic data is particularly effective. These findings position DTs as key enablers of robust sim-to-real vision pipelines (Babaei et al., 2025; Liu et al., 2023).

4.3 Vision-Driven Feedback Loops

Beyond data generation, digital twins support vision-driven closed-loop feedback systems in intelligent mobility. In these systems, vision outputs continuously update the digital twin's state representation. For example, abnormal vehicle behavior detected via vision can be mapped into a DT for trajectory reconstruction and risk analysis. This enables secondary evaluation of perception and control strategies under alternative scenarios. Vision-informed DTs can simulate "what-if" outcomes without physical-world risk. Such feedback loops transform DTs into active perception–decision platforms. This paradigm supports adaptive, resilient, and explainable intelligent transportation systems (Zhang et al., 2021; Wang et al., 2024).

TABLE I: Systematic Review of Digital Twin–Based Vision Studies in Intelligent Mobility

Study	Digital Twin Type	Vision Task	Simulator / Platform	Dataset	Application
Babaei et al., 2025	Data-driven DT	Object detection, segmentation	CARLA + custom AI twin	Synthetic + real	Autonomous driving

Wang et al., 2024	Local intersection DT	Object detection	Unreal Engine + edge DT	Camera streams	Smart intersections
Liu et al., 2023	Simulation-based DT	Detection benchmarking	CARLA	Synthetic	Perception evaluation
Hernández-Martínez et al., 2021	Vision-based DT	Vehicle speed detection	Unity	Synthetic	Traffic monitoring
Zhang et al., 2021	Traffic DT	Abnormal behavior detection	Custom simulator	Video data	Traffic safety
Babaei et al., 2025 (IET ITS)	AI-based DT	Multi-task perception	Hybrid DT	Mixed	Self-driving systems

5.0 Challenges and Research Gaps

Despite growing progress, several challenges limit the widespread adoption of vision-centric digital twins in intelligent mobility systems. Achieving sufficient simulation realism remains difficult, particularly for accurately modeling lighting, weather, sensor noise, and complex urban textures that strongly affect vision performance. High-fidelity simulation often comes at the cost of computational scalability, making large-scale or city-wide digital twins resource intensive. Synchronization latency between the physical system and its digital counterpart poses additional challenges, especially for real-time perception and control. These issues are amplified in edge-assisted deployments, where limited computational and communication resources require careful trade-offs between fidelity and efficiency. Moreover, vision-centric DTs must support diverse perception tasks without excessive reconfiguration. A further gap lies in the lack of standardized benchmarks and evaluation protocols, which hinders reproducibility and cross-study comparison. Existing datasets and simulators vary widely in assumptions and fidelity. This fragmentation makes it difficult to assess progress systematically. Addressing these challenges is essential for advancing reliable, scalable, and deployable vision-centric digital twin frameworks.

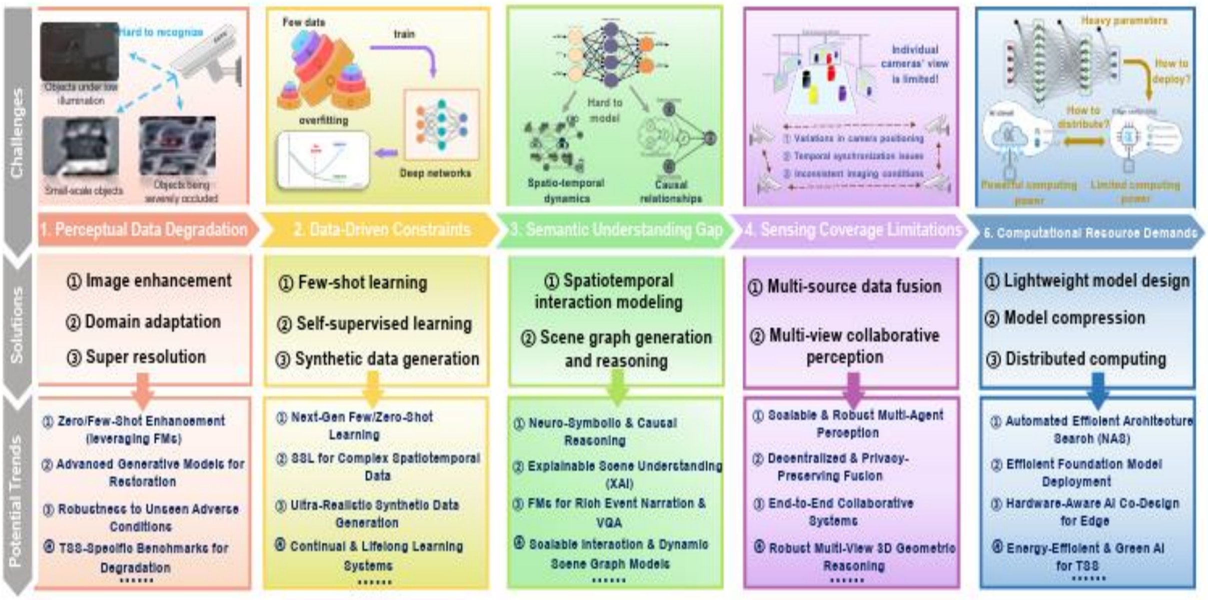

Figure 3: Challenges, solutions, and potential trends for computer vision technologies

REFERENCES:

1. Babaei, P., Li, X., Shah, V. K., & Chowdhury, K. R. (2025). Towards a data-driven digital twin AI-based architecture for self-driving vehicles. IET Intelligent Transport Systems. Advance online publication.
2. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848.
3. Dosovitskiy, A., Fischer, P., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., … Brox, T. (2015). FlowNet: Learning optical flow with convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 2758–2766).
4. Hernández-Martínez, A., García, F., de la Escalera, A., & Armingol, J. M. (2021). Digital twins to alleviate the need for real field data in vision-based vehicle speed detection systems. IEEE Transactions on Intelligent Transportation Systems, 22(4), 2383–2396.
5. Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A. A., Yogamani, S., & Pérez, P. (2021). Deep reinforcement learning for autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems, 23(6), 4909–4926.
6. Liu, Z., Zhang, Y., Chen, L., & Wang, J. (2023). Using simulation to quantify the performance of automotive perception systems.
7. Muhammad, K., Khan, S., Ser, J. D., & Albuquerque, V. H. C. (2022). Vision-based semantic segmentation in scene understanding for autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems, 23(7), 5767–5793.
8. Osei, D. A. & Aliu, A. “Intelligent Control Strategies for Robust and Adaptive Autonomy: A Comparative Analysis” *Sarcouncil Journal of Engineering and Computer Sciences* 5.1 (2026): pp 26-32.
9. Rosique, F., Navarro, P. J., Fernández, C., & Padilla, A. (2019). A systematic review of perception systems and simulators for autonomous vehicles research. Sensors, 19(3), 648.
10. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., … Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo Open Dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2446–2454).
11. Wang, F.-Y., Qin, R., Xu, X., & Zhang, L. (2023). Parallel intelligence and digital twins for intelligent transportation systems. IEEE Transactions on Intelligent Vehicles, 8(1), 1–14.
12. Wang, K., Liu, Y., Zhang, H., & Chen, D. (2024). Augmented intelligence in smart intersections: Local digital twins–assisted hybrid autonomous driving. IEEE Transactions on Intelligent Vehicles. Advance online publication.
13. Zhang, Y., Li, H., Sun, Q., & Wang, X. (2021). Intelligent analysis of abnormal vehicle behavior based on a digital twin. IEEE Transactions on Intelligent Transportation Systems, 22(9), 5603–5614.