AIGC-Driven Real-Time Interactive 4-D Traffic Scene Generation in Vehicular Networks

Xiaolong Li, Ruiting Deng, Jianhao Wei, Xin Wu, Jiayuan Chen, Changyan Yi, Jun Cai, Dusit Niyato, and Xuemin Shen

Abstract

Real-time, interactive 4D traffic scene generation enables rapid digital twinning of traffic scenarios, improving management and decision-making in intelligent transportation systems. However, current text-to-video models, such as Sora, struggle to maintain the temporal coherence of traffic elements and interact with dynamic environments and users when generating 4D scenes. This article introduces a novel cloud-edge-terminal collaborative framework that leverages Artificial Intelligence-Generated Content (AIGC) in vehicular networks to tackle these challenges, ensuring long-term coherence and improved interactivity. The framework presents a comprehensive architecture for real-time interactive 4D scene generation, encompassing data collection, management, model pre-training, fine-tuning, and inference. We examine key design requirements and challenges, demonstrating that our microservice-based framework enables the system to generate and update 4D traffic scenes in real time, effectively responding to traffic data and user inputs. To the best of our knowledge, this is the first successful implementation of real-time, interactive 4D traffic scene generation. Performance evaluations show the superiority of our framework, powered by microservice-based code fine-tuning, over traditional frameworks. Finally, we discuss future research directions to enhance AIGC-driven 4D traffic scene generation.

Introduction

Background

4D traffic scene generation creates realistic three-dimensional representations of transportation elements, such as roads, vehicles, pedestrians, and infrastructure, while incorporating the fourth dimension of time to capture ongoing changes and movements. Real-time processing and interactivity with dynamic environments are critical for managing complex transportation systems, wherein numerous traffic participants operate independently under traffic rules and frequently interact. Offline and non-interactive frameworks [1] cannot adapt to real-time changes, provide dynamic feedback, or integrate with intelligent transportation systems (ITS). Besides, traditional AI models are restricted to processing specific data types and performing predefined tasks in specific environments. As a result, they cannot meet the dynamic and adaptive requirements of 4 D traffic scene generation, which demands multimodal data processing and continuous adaptation to evolving traffic conditions. The development of Artificial Intelligence-Generated Content (AIGC) offers a promising solution [2], [3]. AIGC uses advanced generative models to create traffic scenes and enhances interactivity by generating new content in response to user inputs. AIGC is thus essential for realizing the full potential of real-time interactive 4D traffic scene generation.

AIGC-driven real-time interactive 4D traffic scene generation offers vital benefits to ITS, including greater adaptability to dynamic environments, improved decision-making in complex situations, and increased interactivity with natural user instructions. For example, an AIGC-driven real-time interactive 4D traffic scene generation system in autonomous driving allows vehicles to interact effectively with their surroundings and perform safe driving operations by continuously updating traffic scenes. These advantages make the AIGC-driven system essential for advancing ITS, leading to safer, more efficient, and more responsive traffic management. colorTo fully harness AIGC's potential, we leverage vehicular networks, which facilitate continuous data exchange between vehicles and infrastructure elements like traffic lights and roadside sensors. However, using AIGC to create real-time interactive 4D traffic scenes over vehicular networks with limited computational resources is very challenging. Although text-to-video models, such as Sora, can generate high-fidelity, coherent videos from user-provided descriptions, they are not suitable for real-time interactive 4D traffic scene generation because i) they require significant computational power and time to produce high-quality videos and can only generate static outputs based on text inputs, preventing real-time interactivity, and ii) due to technical limitations, existing text-to-video models cannot maintain the long-term temporal coherence of dynamic traffic elements. To address these severe limitations of existing AIGC models, this article proposes a novel framework for

Xiaolong Li, Ruiting Deng, Jianhao Wei (corresponding author), and Xin Wu (corresponding author) are with the School of Computer Science, Hunan University of Technology and Business, Changsha 410205, China, and also with the Xiangjiang Laboratory, Changsha 410205, China; Jiayuan Chen and Changyan Yi are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; Jun Cai is with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1M8, Canada; Dusit Niyato is with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798; Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Digital Object Identifier: 10.1109/MNET.2025.3545609
Date of Current Version: 18 November 2025
Date of Publication: 25 February 2025

AIGC-driven real-time interactive 4D traffic scene generation over vehicular networks, designed to deliver low-latency and interactive traffic services.

Related Work and Contributions

Recent advancements in 4D scene generation and AIGC have significantly impacted ITS. In 4D scene generation, models that combine image, multiview, and video diffusion techniques have been developed to create dynamic scenes [4]. Neural Radiance Fields (NeRF) generate photorealistic scenes as continuous volumetric fields, with Dynamic NeRF incorporating temporal elements to capture object deformations and evolving interactions [5]. Similarly, D-3DGS employs deformable 3D Gaussian Splats to create dynamic scenes with improved photorealism and structural integrity. Savkin et al. [6] introduced an unsupervised neural network architecture for realistic traffic scene generation using synthetic 3D scene graphs. As aforementioned, these text-to-video and text-to-image models take significant computational time and cannot maintain continuous temporal coherence. Thus, they cannot be efficiently applied to generating real-time interactive 4D traffic scenes in ITS.

In AIGC-based ITS applications, Zhang et al. [7] proposed a cloud-edge-terminal collaborative architecture for autonomous driving, leveraging AIGC for system design and resource management. Masri et al. [8] employed GPT-4o-mini to optimize traffic flow and reduce wait times at urban intersections. Wang et al. [9] integrated large language models (LLMs) into traffic signal control systems, while Manas et al. [10] developed a framework to translate natural language traffic rules into metric temporal logic for autonomous vehicles. However, the potential of AIGC to enable real-time, interactive 4D traffic scene generation over vehicular networks remains unexplored.

This article proposes a novel AIGC-driven real-time interactive 4D traffic scene generation over vehicular networks by addressing this gap. Our contributions include the following:

- We introduce a novel cloud-edge-terminal collaborative system architecture that supports text-to-code model-driven real-time 4D traffic scenes over vehicular networks. This architecture covers the entire AIGC lifecycle, from data collection and management to model pretraining, fine-tuning, and inference, providing valuable insights for ITS implementation.
- We conduct an in-depth analysis of the critical design requirements and challenges for deploying real-time, interactive 4D traffic scene generation systems using text-to-code models. This analysis offers a roadmap for addressing these challenges and ensuring effective implementation in dynamic transportation environments.
- We detail the practical implementation of a real-time, interactive 4D traffic scene generation system based on text-to-code models. This system demonstrates its ability to generate interactive 4D scenes in real time. To our best knowledge, this is the first instance of successfully implementing real-time interactive 4D traffic scene generation. We evaluate the framework's performance by comparing traditional and microservice-based frameworks in terms of inference time and execution accuracy.
- We outline future research directions by highlighting key open questions to inspire further exploration and innovation in this field.

The Architecture Framework of AIGC-Driven Real-Time Interactive 4D Traffic Scene Generation System

The proposed AIGC-driven 4D traffic scene generation system for vehicular networks consists of three layers: the terminal layer, the edge layer, and the cloud layer, as shown in Fig. 1. The terminal layer focuses on collecting real-time data from sensors and vehicles; the edge layer processes this data locally to ensure quick response and handles latency-sensitive tasks; and the cloud layer manages resource-intensive tasks, such as model pretraining and large-scale data storage, providing refined model updates to the edge for real-time adaptability. The framework encompasses five key computational processes: data collection, management, pretraining, fine-tuning, and inference, which are described as follows. Unlike existing architectures, this framework leverages a cloud-edge-terminal collaboration and text-to-code models to ensure real-time interactivity, temporal coherence, and adaptability, addressing the dynamic demands of intelligent transportation systems.

1) Data Collection: Data collection is the first step for enabling real-time, interactive 4D traffic scene generation within an AIGC-driven system. This process, primarily handled by terminal devices, involves gathering data comprehensively from various sources. The quality and scope of the collected data are critical for generating accurate and responsive traffic scenes. Key data sources include real-time traffic information from road infrastructure sensors, and GPS systems, as well as environmental data from meteorological stations and road surface sensors, vehicle-mounted sensors and smart devices, which are vital for capturing the dynamic nature of transportation systems. Additionally, integrating non-traditional data sources, such as social media and online platforms, provides valuable insights into human behavior and urban activities, thereby enhancing the realism and interactivity of the generated scenes. By combining these diverse data collection strategies, the system ensures that AIGC-driven 4D traffic scenes are accurate and capable of adapting in real-time to evolving conditions.

2) Data Management: Effective data management is crucial for enabling efficient information sharing and integration among stakeholders while protecting sensitive information throughout its lifecycle. This process involves handling diverse data types, including real-time traffic information, simulated data, AIGC-created information, and historical traffic records. It begins with comprehensive data preprocessing, which includes real-time cleaning, filtering, synchronization, spatial processing, aggregation, feature extraction, and image and video processing. These steps ensure the data are accurate, consistent, and ready for immediate use. The preprocessed data is

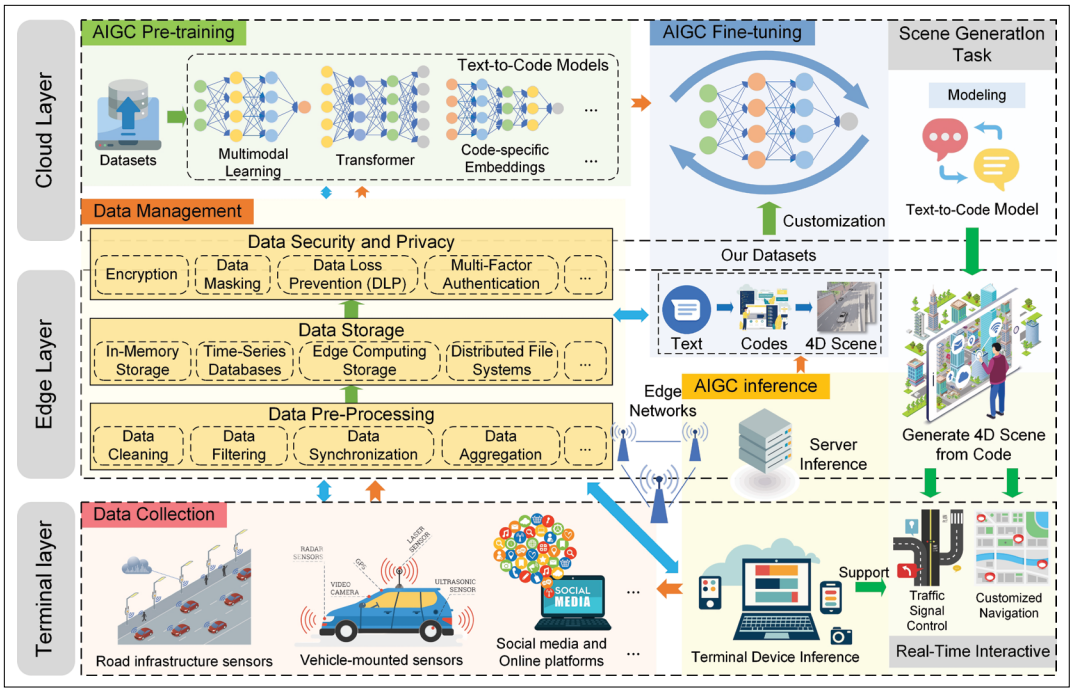

FIGURE 1. The proposed vehicular network framework for AIGC-driven real-time interactive 4 D traffic scene generation system consists of the terminal layer, edge layer, and cloud layer. All these support the 4D traffic scene generation with AIGC lifecycle, i.e., data collection, management, pre-training, fine-tuning, and inference.

then stored using high-performance technologies, such as in-memory storage, time-series databases, edge computing storage, and distributed file systems. This robust storage infrastructure supports the continuous training and updating of AIGC models, enabling the seamless generation of real-time, interactive 4D traffic scenes. Given the data sensitivity involved, strict measures are necessary to prevent privacy violations, unauthorized surveillance, and public security risks.

3) AIGC Pre-Training: Selecting suitable AIGC models is essential for achieving real-time, interactive 4D traffic scene generation. Text-to-code models are favored over text-to-video models because they produce adaptable, executable codes that dynamically responds to real-time data, making them ideal for complex, evolving traffic environments. These models are computationally efficient and integrate seamlessly with existing ITS, enhancing overall system responsiveness. Pretraining for these models usually occurs on cloud servers with significant computational resources. Advanced pretraining techniques can further optimize text-to-code models for real-time 4D traffic scene generation. For instance, transformer models capture the sequential and temporal dynamics of traffic data, enabling the generation of contextually accurate and responsive codes. Additionally, code-specific embeddings refer to representations of code (such as source code), which ensure that the generated codes are contextually appropriate by accurately representing the syntax and semantics of the relevant programming languages. Large-scale pretraining on diverse code corpora equips the model with the versatility to handle various programming challenges in dynamic, real-time interactive traffic management systems.

4) AIGC Fine-Tuning: AIGC fine-tuning customizes pretrained models for specific tasks. Using curated datasets, fine-tuning can occur either in cloud environments or at the edge. Advanced fine-tuning techniques are employed to optimize text-to-code models for real-time, interactive 4D scene generation, allowing for dynamic, user-driven creation and modification. Given the ever-changing nature of transportation scenes, which involve moving vehicles and shifting environments, techniques like multimodal learning and alignment are essential for converting text inputs into temporally coherent sequences that accurately represent object trajectories and interactions. Instruction tuning is critical for text-to-code models because it allows the model to understand better how users might describe their coding needs. Additionally, data augmentation can expand the variety of natural language descriptions and code examples available for fine-tuning. It is essential when a large amount of task-specific training data is impossible. Fine-tuning for scene composition enables the model to generate modular components (e.g., roads, buildings, vehicles) that can be assembled and modified in real time based on text inputs.

5) AIGC Inference: The inference phase utilizes pre-trained and fine-tuned AIGC models to generate content from specific inputs while minimizing computational overhead efficiently. This optimization enables AIGC inference to be deployed on edge servers or devices, such as in-vehicle systems and roadside units, facilitating real-time interactive services. The content generated is essential for creating real-time interactive 4D transportation scenes, which support comprehensive analysis and predictive modeling for effective traffic management and optimization.

By deploying these models directly within the transportation infrastructure at the edge, the system can dynamically generate and update 4 D scenes in real time, accurately reflecting current traffic conditions and environmental changes. This approach significantly enhances the responsiveness and precision of traffic management systems, supporting real-time interactive ITS applications.

Key Design Requirements and Challenges

With the architecture established, implementing this framework requires addressing several key design challenges that impact real-time performance and interactivity. The key design requirements and major challenges of deploying AIGC-driven real-time interactive 4D traffic scene generation in vehicular networks are discussed below.

Adaptive Self-Evolution of AIGC-Driven Real-Time Interactive 4D Traffic Scene Generation

To ensure synchronization between real-world and generated traffic scenes, AIGC-driven systems must autonomously adapt and evolve. This adaptive self-evolution relies on continuous learning from real-time data, reflecting the ever-changing nature of traffic environments. The system must continuously process and integrate data from various sources to stay aligned with current traffic conditions, environmental changes, and emerging trends. Our framework addresses this need through a cloud-edge-terminal collaborative structure, which optimally distributes processing tasks. By dynamically updating its models in response to new training data, the AIGC-driven system ensures that the generated 4D traffic scenes accurately reflect the current state of the transportation network. Additionally, this capability allows the system to recognize and adapt to long-term trends, ensuring its predictive algorithms remain effective as traffic patterns evolve. The self-evolution process must be scalable and flexible, allowing the system to seamlessly transition across different geographic regions and respond effectively to diverse traffic conditions. Key aspects of this process are highlighted below.

Generating High-Precision and Fast-Responsive Contextual Text-to-Code: Text-to-code models must produce executable code that captures the complexity and dynamics of traffic scenes while adapting to environmental changes in real time. It requires a pretraining process that thoroughly understands traffic contexts, captures nuances in traffic patterns, and generates code that seamlessly adapts to new inputs and evolving conditions. The challenge is ensuring the model consistently generates contextually relevant, efficient, and adaptable code for real-time execution, while preserving the fidelity and accuracy of the 4D traffic scenes. Enhancing pretraining with diverse traffic datasets can improve the model's ability to generate contextually appropriate code. Additionally, incorporating a reward mechanism can encourage the model to produce code that meets real-time execution requirements while maintaining high fidelity and accuracy.

Designing Efficient Data Acquisition Schemes for Dynamic Fine-Tuning: Dynamic fine-tuning enables models to continuously learn and adapt to new traffic data, evolving conditions, and specific domain requirements. Efficient data acquisition is essential for maintaining temporal and spatial coherence in 4D scene generation, ensuring outputs align with domain-specific needs, user demands, and application contexts. Aligning fine-tuning with transfer learning enhances this coherence. Techniques like model distillation and pruning optimize performance in resource-constrained environments, such as vehicular networks, without compromising efficiency. However, the availability of high-quality and diverse data remains a significant challenge. Limited access to diverse traffic data impairs the model's ability to generalize across different scenarios, while low-quality or noisy data can lead to inaccurate code and reduce the effectiveness of fine-tuning. To address these issues, robust preprocessing methods such as anomaly detection and data normalization could be employed to reduce noise and inconsistencies. Furthermore, a hierarchical data fusion approach integrating heterogeneous data sources like vehicle sensors, infrastructure devices, and weather feeds, can ensure a reliable foundation. Another critical challenge is managing the latency introduced by dynamic fine-tuning. In real-time 4D traffic scene generation, even slight delays in processing can result in outdated or incorrect outputs, severely impacting traffic management and safety. Edge computing is leveraged for localized fine-tuning tasks, and predictive algorithms are further adopted to address data quality issues proactively, enhancing system resilience.

Optimizing Networking Resources for Real-Time 4D Traffic Scene Generation: Real-time 4D traffic scene generation over vehicular networks demands high computational efficiency and minimal latency. Vehicular edge computing (VEC) addresses these challenges by offloading computational tasks to roadside units (RSUs), reducing reliance on cloud-based processing, and thereby minimizing communication latency [11]. By processing data locally, edge computing further reduces latency, enhancing system performance [12]. However, managing resources in heterogeneous environments with varying computational capacities is challenging. Optimization strategies include using neural architecture search to deploy lightweight AIGC models tailored to edge-specific constraints. Techniques such as pruning and quantization reduce model size while preserving accuracy, enabling real-time inference on resource-limited devices. AI-driven resource scheduling dynamically allocates computational power based on traffic complexity and predictive analytics, further improving performance. These approaches ensure that 4D traffic scene generation remains responsive and scalable within dynamic vehicular networks.

Personalized Interactions in AIGC-Driven Real-Time Interactive 4D Traffic Scene Generation

In real-world traffic, each participant-whether a vehicle, pedestrian, or cyclist-exhibits unique characteristics, preferences, and decision-making processes. By treating these participants as independent agents and personalizing their interactions, the system can more accurately capture the complexity of traffic dynamics. Personalized

interactions enable the system to adapt to individual driving behaviors, such as route preferences, driving habits, and the way the driver interacts with other vehicles, leading to more precise predictions of traffic flow, congestion, and potential hazards. This personalization also accounts for external factors such as weather and road conditions, further refining driver behavior models. Ultimately, personalized interactions make AIGC-driven 4D traffic scene generation systems realistic and applicable to real-world traffic scenarios.

Producing Tailored Interaction in Uncertain and Diverse Traffic Conditions: In real-world traffic, unpredictable conditions, such as dense fog on specific road sections, require the 4D simulation system to dynamically adapt interactions among traffic participants. Tailoring these interactions to environmental factors enhances realism and improves the system's ability to predict and manage risks. For instance, when vehicles enter or exit foggy areas, to account for reduced visibility, the system must adjust vehicle operations, such as increasing safe following distances and lowering driving speeds. The system must continuously fine-tune these interactions to maintain realism, generate accurate traffic flow predictions, and identify potential hazards. Any deviation from realistic interaction behaviors can compromise the simulation's accuracy, particularly in real-time scenarios. The simulation can deliver more effective traffic management and risk assessment by seamlessly adapting to these uncertain and diverse conditions.

Synthesizing Customized Interaction for Various Traffic Participants: Customized 4D traffic scene generation is essential for developing realistic traffic simulations. An AIGC-driven system personalizes traffic scenes to match each participant's characteristics and objectives, ensuring that the scenarios accurately replicate diverse real-world behaviors. To achieve this, the system must gather and integrate detailed data on individual users, such as driving habits, preferred routes, speed preferences, and typical responses to traffic signals and road conditions. Furthermore, it should model human driver behaviors precisely, accounting for personalized characteristics, reaction times, and responses to external factors like weather and road conditions.

Guaranteeing Immersive Quality of Experience in High-Fidelity Interactions: Ensuring immersive Quality of Experience (QoE) is a key focus in high-fidelity interactions in complex systems like 4D traffic scene simulations. Immersive QoE ensures that users perceive these simulations as realistic and engaging. Achieving this requires evaluation metrics that address both subjective and objective aspects of the user experience. Subjective metrics assess qualitative elements, such as the smoothness of movements and overall realism, while objective metrics measure quantifiable factors like traffic flow consistency and response times. To maintain high-fidelity interactions, the system must integrate these metrics and make real-time adjustments, such as smoothing vehicle trajectories or enhancing responsiveness to sudden environmental changes. By continuously optimizing based on these metrics, the system can deliver a more engaging and realistic user experience.

Applications of AIGC-Driven Real-Time Interactive 4D Traffic Scene Generation

AIGC-Driven Real-Time Interactive 4D Traffic Scene Generation Empowers Customized Navigation

Customized navigation is a critical application in intelligent transportation systems. AIGC-driven navigation systems address significant challenges, showcasing their potential in customized navigation. For example, Kuribayashi et al. [13] proposed a vision-and-language model that uses LLMs to interpret user instructions and generate code for robotic navigation. However, these AIGC methods cannot effectively replicate the complex interactions between driving vehicles and surrounding environments, resulting in poor navigation experience.

AIGC-driven real-time interactive 4D scene generation offers high-fidelity navigation by continuously updating driving scenes in real time. As shown in Fig. 2(a), the AIGC-driven system enables vehicle drivers to visualize driving paths, traffic participants, and surrounding environments. The system enhances safety and efficiency by offering a personalized, immersive navigation experience, meanwhile improving the accuracy and dependability of customized navigation while expanding their applicability to diverse, complex environments. As a result, AIGC-driven real-time interactive 4D traffic scene generation systems represent a significant advancement in intelligent transportation, providing vehicle drivers with adaptive guidance tailored to their needs and ensuring real-time responsiveness and interactivity.

AIGC-Driven Real-Time Interactive 4D Traffic Scene Generation Assists Smart Transportation Management

Smart transportation management optimizes traffic signal control timings, reduces congestion, and

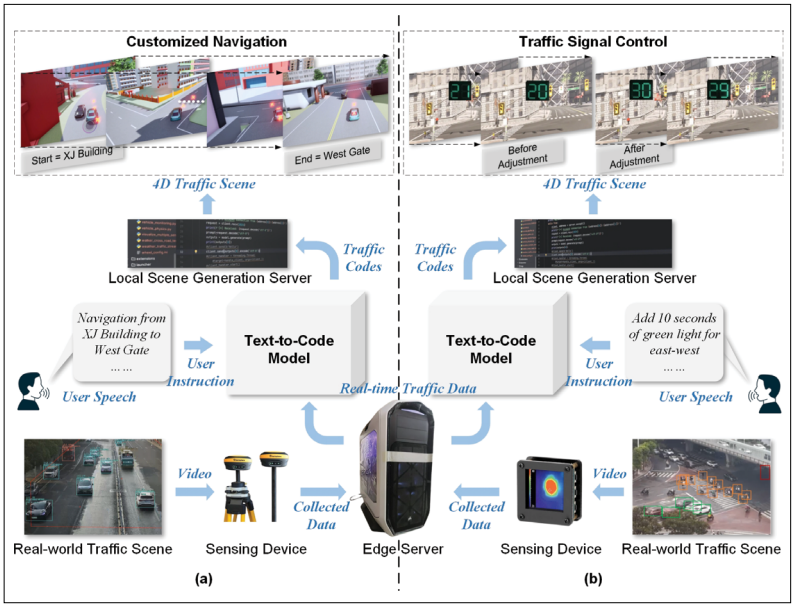

FIGURE 2. AIGC-driven real-time interactive 4D traffic scene generation for customized navigation and traffic signal control applications, with the detailed data processing workflow.

enhances overall traffic flow efficiency. AIGC captures vehicle and pedestrian movements, enabling dynamic real-time adjustments to traffic signal timings. A prominent example is the cross-city Meta multi-agent traffic signal control model [14]. This model uses a dual-layer Transformer architecture to improve adaptability across urban environments. Wei et al. [15] introduced a generative adversarial network to enable more accurate signal adjustments. However, these AIGC methods lack the ability to simulate complex traffic scenarios, and operating in solos limits their effectiveness in adjusting traffic signals.

AIGC-driven real-time interactive 4D scene generation systems digitally twin vehicle movements, traffic light sequences, and environmental conditions in real time. Thus, the AIGC-driven system improves traffic signal timing, congestion management, and efficiency. As shown in Fig. 2(b), traffic police can adjust traffic signal controls within the system, which autonomously generates corresponding 4D scene simulations. These simulations allow detailed analysis of traffic congestion and pedestrian-vehicle interactions, enabling dynamic optimization of signal settings. Integrating real-time interactive 4D scenes allows cities to adopt more intelligent, responsive systems, paving the way for innovative urban transportation management.

Implementation of Real-Time Interactive AIGC-Driven 4D Traffic Scene Generation System

Based on the overall architecture of two applications in the section "Applications of AIGC-Driven Real-Time Interactive 4D Traffic Scene Generation," this section presents the practical implementation of a AIGC-driven real-time interactive 4D traffic scene generation framework, addressing the section "Key Design Requirements and Challenges"'s design requirements and challenges.

System Implementation

Fig. 3 illustrates the practical implementation of the proposed real-time interactive AIGC-driven 4D traffic scene generation framework, which employs a cloud-edge-terminal collaborative architecture. This system integrates several key components: laser-vision equipment, microphones, speech recognition servers, local text-to-code models, local scene generation servers, and cloud servers.

- **Laser-Vision Equipment:** Tracks the real-time positions of pedestrians and vehicles.
- **Microphones:** Facilitate interaction by receiving spoken commands, such as navigation requests or weather updates.
- **Speech Recognition Servers:** Accurately convert spoken instructions into text, ensuring a seamless user experience.
- **Local Text-to-Code Models:** Deployed on edge servers equipped with NVIDIA A100 GPUs, these models handle AIGC fine-tuning and inference. They process user instructions and real-time traffic data from the speech recognition servers and laser-vision equipment to generate corresponding traffic codes.
- **Local Scene Generation Servers:** Running CARLA, an open-source simulator powered by Unreal Engine assets, these servers use the traffic code generated by the text-to-code models to create 4D traffic scenes. With the CARLA Map Editor, local scene generation servers digitally replicate real-world environments, creating immersive scenarios with vehicles, pedestrians, and detailed surroundings. The system continuously updates the CARLA model with real-time traffic data and user interactions.
- **Cloud Servers:** Manage pretraining of text-to-code models and transmit trained model parameters to edge servers for further fine-tuning.

In the system, atomic operations are basic and indivisible API functions, e.g., adding a number of vehicles and pedestrians, or changing the weather from cloudy to sunny. These operations are modeled as microservices, each handling a specific task to efficiently manage real-time traffic data and user instructions. The microservices

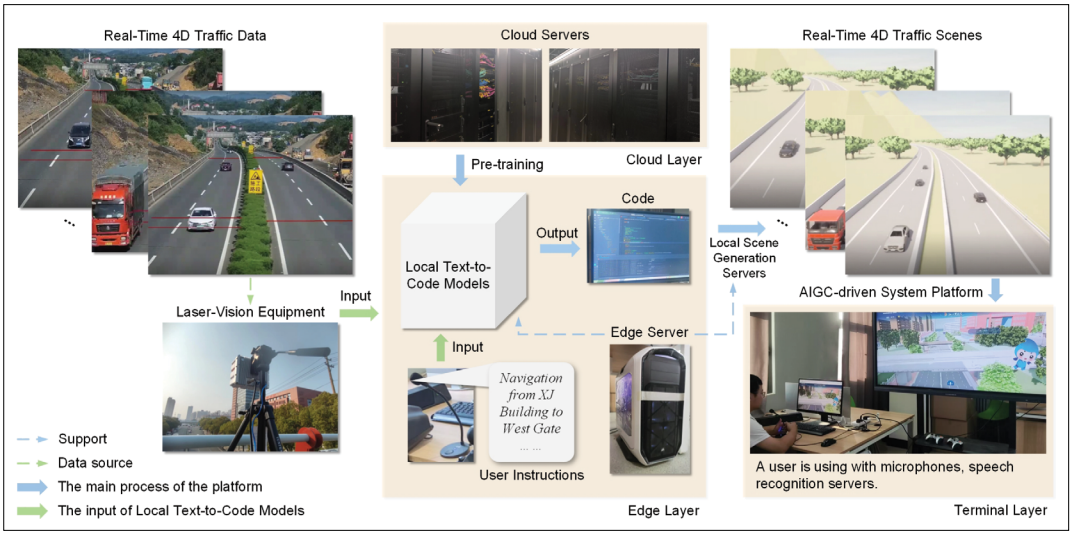

FIGURE 3. Practical implementation of the AIGC-driven real-time interactive 4D traffic scene generation framework.

interact via APIs with predefined parameters. Each API call triggers a corresponding microservice action, enabling modular control over scene elements. Communication among different components is managed using RESTful APIs and WebSocket for real-time updates, with Protocol Buffers (ProtoBuf) ensuring efficient message serialization. Local text-to-code models generate the API codes, while scene generation servers execute the corresponding functions based on these codes[1]. WizardCoder was chosen as the pretraining model, as its extensive training can be conducted on large programming corpora, including microservices, enabling it to understand, generate, and orchestrate atomic operations effectively.

The fine-tuning process customizes pre-trained models for real-time 4D traffic scene generation, specifically targeting microservice orchestration. This fine-tuning uses specialized datasets that represent the system's microservices. Standard prompts are employed to build the fine-tuning datasets. We offer a microservice-based training dataset[2] for instruction tuning, including user instructions, prompts, and corresponding outputs. Every user instruction executable via an API call generates a minimum of 20 training samples comprising input-prompt-output pairs. As illustrated in Fig. 4, user instructions guide the system in executing specific tasks. After fine-tuning with vehicle-adding instructions, the model translates commands like "adding *j* vehicles" into the corresponding code "Add_Vehicles(*j*)". This atomic operation ensures the accurate execution of simple tasks. Additional training samples ensure precise code generation for complex instructions involving multiple API calls. Fig. 5 illustrates the AIGC-driven testbed system, which generates real-time interactive 4D traffic scenes. The system accomplishes this by continuously monitoring and detecting traffic flow through laser-vision equipment. User interactions further enable tasks like adding vehicles to the road or changing weather conditions, highlighting the system's interactive capabilities.

Prompt Template	Input-Prompt-Out Examples		
	Input	Prompt	Out
Task-level Instruction	Add Vehicles	Generating codes of adding vehicles	Add_Vehicles()
Example 1	Add 10 Vehicles	Generating codes of adding 10 vehicles	Add_Vehicles (10)
Example 2	Add 15 Vehicles	Generating codes of adding 15 vehicles	Add_Vehicles (15)
...	...	...	...
Example 19	Add 90 Vehicles	Generating codes of adding 90 vehicles	Add_Vehicles (90)
Example 20	Add 92 Vehicles	Generating codes of adding 92 vehicles	Add_Vehicles (92)

FIGURE 4. User instruction and corresponding input-prompt-output pairs for fine-tuning training.

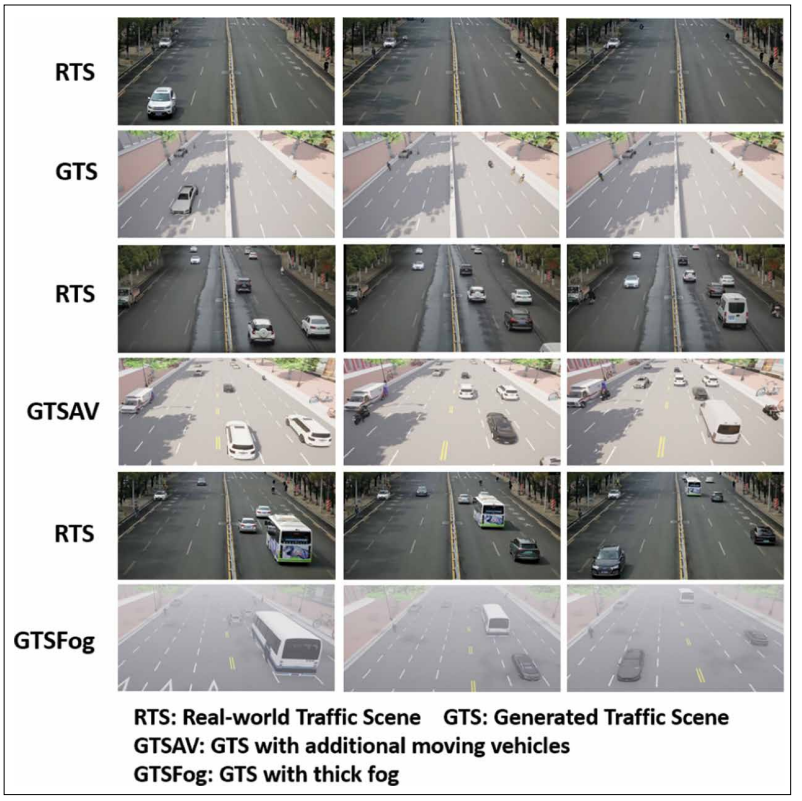

FIGURE 5. Real-time interactive 4D traffic scenes generated by the AIGC-driven testbed system.

Performance Evaluation

To validate the proposed system, we evaluated its performance using two fine-tuned text-to-code frameworks: (i) a traditional framework that directly generates implementation code based on user instructions, and (ii) the proposed microservice-based framework, which generates and orchestrates API code from user instructions. We assessed their performance using four key metrics: 1) **Inference Time**, the time required for the model to generate code from a text input. 2) **Execution Accuracy**, the percentage of generated codes that produce the correct output when executed in the CARLA simulation environment, is measured by the percentage of correct outputs. Inference Time measures system responsiveness, while Execution Accuracy ensures scene reliability. 3) **Computing Cost**, the proportion of system resources (CPU and GPU) consumed during the execution of a process, expressed as a percentage of the total available resources. 4) **Delay**, the time required to generate 4D traffic scene from speech. We constructed 80 user instructions for evaluation, covering tasks including adding/removing traffic participants, route navigation, traffic signal control, and weather changes. Fig. 6(a) compares the inference time. The microservice-based framework achieves an inference time of less than 0.5 seconds, meeting real-time interaction requirements and significantly outperforming the traditional approach. Fig. 6(b) illustrates the execution accuracy results. The microservice-based method generates accurate traffic codes across all instruction types, while the traditional framework struggles with certain instructions, leading to incorrect or non-executable codes. Fig. 6(c) shows the computing cost, where the microservice-based framework exhibits comparable costs to the traditional approach, indicating similar resource efficiency. Fig. 6(d) compares the delay between the traditional approach and the microservice-based framework.

[1] The code is available at https://github.com/Open-HUTB/4D_SCENE_GEN.

[2] The dataset is available at https://github.com/Open-HUTB/4D_SCENE_GEN/training_data.

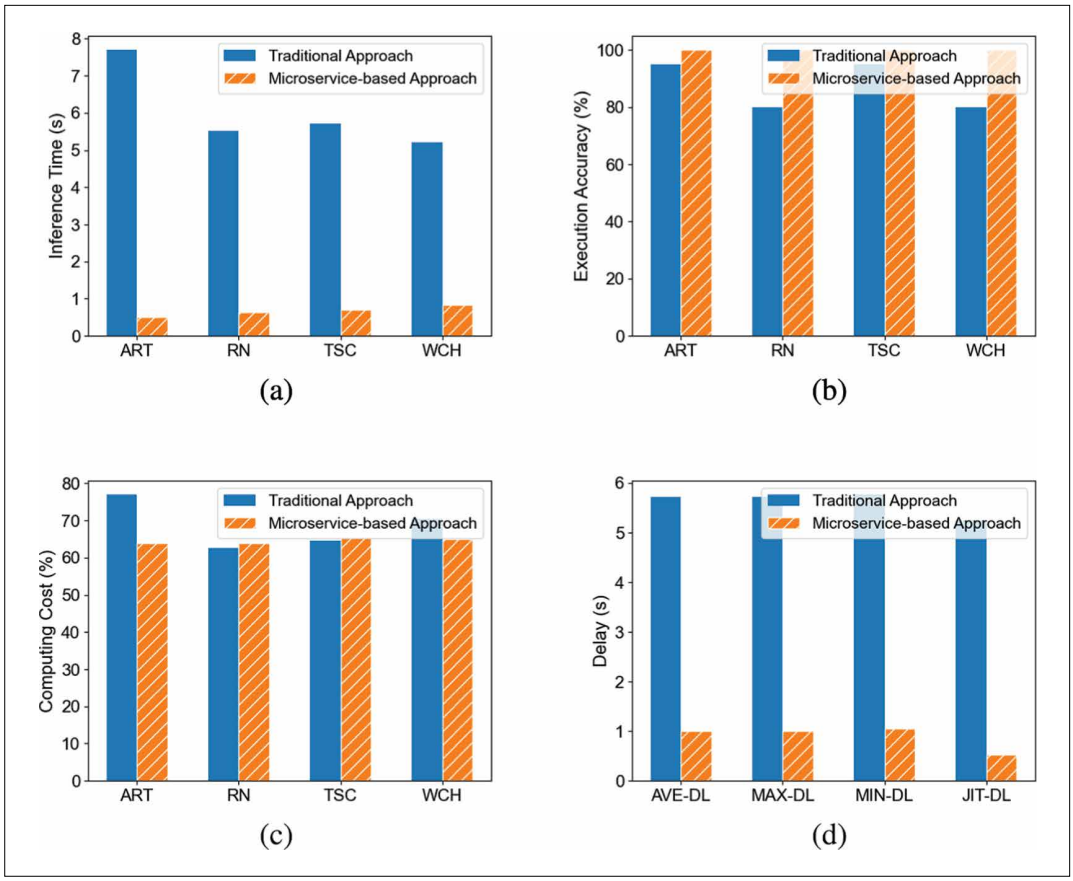

FIGURE 6. Performance comparison of traditional versus microservice-based frameworks: a) Inferencing Time, b) Execution Accuracy, c) Computing Cost and d) Delay. ART = adding/removing traffic participants, RN = route navigation, TSC = traffic signal control, WC = weather changes. AVE-DL = lower average delay, MAX-DL = maximum delay, MIN-DL = minimum delay, and JIT-DL = jitter delay.

The microservice-based approach consistently outperforms the traditional method, achieving significantly lower AVE-DL, MAX-DL, MIN-DL, and JIT-DL. Compared to recent edge computing schemes (without enabling cloud computing and terminal's assistance) in intelligent transportation, our approach leverages a cloud-edge-terminal setup to balance processing loads, enhancing scalability for high-density traffic. This architecture enables efficient large-scale data management, complementing existing purely edge computing based edge-focused solutions.

Conclusion and Future Research Directions

In this article, we have introduced an innovative AIGC-driven framework for real-time, interactive 4D traffic scene generation over vehicular networks. The proposed framework addresses the limitations of traditional text-to-video models by utilizing text-to-code models to create dynamic, temporally coherent scenes that enhance the capabilities of ITS. We have demonstrated the practical applications of the framework and proposed a cloud-edge-terminal collaborative architecture, highlighting the key design requirements and challenges for real-time 4D traffic scene generation. A case study has validated the effectiveness of our microservice-based framework, demonstrating its ability in effectively responding to traffic data and user inputs, and in generating and updating 4D scenes in real time, effectively responding to traffic data and user inputs. However, several challenges remain in fully harnessing AIGC's potential for 4D traffic scene generation.

Optimizing Data Fusion and Processing. Future research can focus on advanced techniques for integrating and processing the vast, heterogeneous data generated by ITS. Approaches such as deep learning, edge computing, and distributed ledger technology could enhance real-time data processing while ensuring security and privacy.

Enhancing Interactive Experiences. To improve user engagement in virtual transportation platforms, future work should explore advanced interfaces, virtual reality (VR), augmented reality (AR), and natural language processing (NLP) to create more immersive and intuitive environments.

Balancing Real-Time Performance With Scalability. As ITS scale, maintaining real-time performance and stability becomes challenging. Future research should optimize algorithms and architectures, such as microservices and edge computing, to support large-scale, high-concurrency applications.

Acknowledgment

This work was supported in part by the Major Program Project of Xiangjiang Laboratory under Grant XJ2023001 and Grant XJ2022001, in part by the Innovation Fund of QiYuan Laboratory under Grant 2022-JCJQ-LA-001-088, in part by

the IoT Intelligent Sensing Support Project for Science and Technology Innovation Teams in Hunan Province, in part by the Key Research and Development Program of Hunan Province under Grant 2024JK2007, in part by the Hunan Provincial Natural Science Foundation of China under Grant 2023JJ40237, in part by the Hunan Provincial Graduate Research and Innovation Project under Grant CX20231172, and in part by the Guangxi Science and Technology Major Program under Grant 2024AA15007.

References

[1] R. Wang et al., "End-to-end multitarget flexible job shop scheduling with deep reinforcement learning," *IEEE Internet Things J.*, vol. 12, no. 4, pp. 4420–4434, Feb. 2025.
[2] J. Chen et al., "A revolution of personalized healthcare: Enabling human digital twin with mobile AIGC," *IEEE Netw.*, vol. 38, no. 6, pp. 234–242, Nov. 2024, doi: 10.1109/MNET.2024.3366560.
[3] S. D. Okegbile et al., "FLeS: A federated learning-enhanced semantic communication framework for mobile AIGC-driven human digital twins," *IEEE Netw.*, early access, Jan. 6, 2025, doi: 10.1109/MNET.2025.3526556.
[4] Y. Zheng et al., "A unified approach for text-and image-guided 4D scene generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 7300–7309.
[5] C. Wang et al., "Flow supervision for deformable NeRF," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 21128–21137.
[6] A. Savkin et al., "Unsupervised traffic scene generation with synthetic 3D scene graphs," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 1229–1235.
[7] J. Zhang et al., "Cloud-edge-terminal collaborative AIGC for autonomous driving," *IEEE Wireless Commun.*, vol. 31, no. 4, pp. 40–47, Aug. 2024.
[8] S. Masri, H. I. Ashqar, and M. Elhenawy, "Leveraging large language models (LLMs) for traffic management at urban intersections: The case of mixed traffic scenarios," 2024, *arXiv:2408.00948*.
[9] M. Wang et al., "LLM-assisted light: Leveraging large language model capabilities for human-mimetic traffic signal control in complex urban environments," 2024, *arXiv:2403.08337*.
[10] K. Manas, S. Zwicklbauer, and A. Paschke, "TR2MTL: LLM based framework for metric temporal logic formalization of traffic rules," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2024, pp. 1206–1213.
[11] Q. Luo et al., "Minimizing the delay and cost of computation offloading for vehicular edge computing," *IEEE Trans. Services Comput.*, vol. 15, no. 5, pp. 2897–2909, Sep. 2022.
[12] J. Chen et al., "Networking architecture and key supporting technologies for human digital twin in personalized healthcare: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 1, pp. 706–746, 1st Quart., 2024.
[13] M. Kuribayashi et al., "Memory-maze: Scenario driven benchmark and visual language navigation model for guiding blind people," 2024, *arXiv:2405.07060*.
[14] H. Jiang et al., "X-light: Cross-city traffic signal control using transformer on transformer as meta multi-agent reinforcement learner," in *Proc. 33rd Int. Joint Conf. Artif. Intell.*, Aug. 2024, pp. 94–102.
[15] C. Wei et al., "KI-GAN: Knowledge-informed generative adversarial networks for enhanced multi-vehicle trajectory forecasting at signalized intersections," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 7115–7124.

Biographies

Xiaolong Li (Member, IEEE) (1x1@hutb.edu.cn) is currently a Professor with the Hunan University of Technology and Business, Changsha. His research interests include deep learning, intelligent transportation systems, and the Internet of Things.

Ruiting Deng (deng362500549@163.com) is currently pursuing the master's degree with the School of Computer Science, Hunan University of Technology and Business. Her research interests include digital twin and deep learning.

Jianhao Wei (jianhao@hnu.edu.cn) is currently an Associate Professor with the Hunan University of Technology and Business, Changsha. His research interests include deep learning, intelligent transportation systems, and the Internet of Things.

Xin Wu (xinwu@hnu.edu.cn) is currently a Lecturer with the Hunan University of Technology and Business, Changsha. His research interests include deep learning, intelligent transportation systems, and digital twin.

Jiayuan Chen (jiayuan.chen@nuaa.edu.cn) is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. His research interests include edge computing and digital twin.

Changyan Yi (Senior Member, IEEE) (changyan.yi@nuaa.edu.cn) is currently a Professor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. His research interests include edge computing, the industrial IoT, digital twin, and 5G and beyond.

Jun Cai (Senior Member, IEEE) (jun.cai@concordia.ca) is currently a Professor and the PERFORM Centre Research Chair with the Department of Electrical and Computer Engineering, Concordia University, Canada. His research interests include edge/fog computing and eHealth.

Dusit Niyato (Fellow, IEEE) (dniyato@ntu.edu.sg) is currently a President's Chair Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include edge intelligence, machine learning, and incentive mechanism design.

Xuemin (Sherman) Shen (Fellow, IEEE) (sshen@uwaterloo.ca) is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, wireless network security, social networks, and vehicular ad hoc networks. He is a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Chinese Academy of Engineering Foreign Fellow.