Generative AI-Enabled Vehicular Networks: Fundamentals, Framework, and Case Study

Ruichen Zhang (iD), Ke Xiong (iD), Hongyang Du (iD), Dusit Niyato (iD), Jiawen Kang (iD), Xuemin Shen (iD), and H. Vincent Poor (iD)

ABSTRACT

Recognizing the tremendous improvements that the integration of generative artificial intelligence (AI) can bring to intelligent transportation systems, this article explores the integration of generative AI technologies in vehicular networks, focusing on their potential applications and challenges. Generative AI, with its capabilities of generating realistic data and facilitating advanced decision-making processes, enhances various applications when combined with vehicular networks, such as navigation optimization, traffic prediction, data generation, and evaluation. Despite these promising applications, the integration of generative AI with vehicular networks faces several challenges, such as real-time data processing and decision-making, adapting to dynamic and unpredictable environments, as well as privacy and security concerns. To address these challenges, we propose a multi-modality semantic-aware framework to enhance the service quality of generative AI. By leveraging multi-modal and semantic communication technologies, the framework enables the use of text and image data for creating multi-modal content, providing more reliable guidance to receiving vehicles and ultimately improving system usability and efficiency. To further improve the reliability and efficiency of information transmission and reconstruction within the framework, taking generative AI-enabled vehicle-to-vehicle (V2V) as a case study, a deep reinforcement learning (DRL)-based approach is proposed for resource allocation. Finally, we discuss potential research directions and anticipated advancements in the field of generative AI-enabled vehicular networks.

INTRODUCTION

The rapid development of the Internet of Things (IoT) and connected devices has facilitated the development of vehicle networks, making them a critical component of modern transportation systems. In-vehicle networks can significantly improve the efficiency of transportation by facilitating seamless communication between vehicles, infrastructure and other connected devices. So far, existing works have investigated the vehicular networks. For example, in [1], a brief overview was proposed to address key challenges in securing 5G V2X by identifying its concerns on trust, security, and privacy.

On the other hand, with the development of artificial intelligence (AI) technology, generative AI, as a branch of AI, has been receiving increasing attention. In particular, generative AI focuses on creating new content or data, such as images, texts, audios, videos, or even system designs, by learning patterns and structures from existing data and autonomously generating novel outputs. For example, ChatGPT,[1] a generative AI language model developed by OpenAI, is capable of generating human-like text based on a given context. Also, some generative AI techniques such as generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models can be utilized to generate realistic traffic simulation scenarios, create alternative routes for connected vehicles, or minimize energy consumption in electric vehicles. For example, in [2], a DriveGAN was proposed to generate high-resolution and diverse simulations based on user-defined driving conditions, such as weather and positioning. Also, in [3], a realistic sensor data synthesis framework was proposed for enhancing autonomous driving, where the data-driven camera generation approach not only produces visually high-quality data but also serves as training datasets that enhance the performance of AI algorithms.

The potential integration of generative AI into vehicular networks systems can lead to substantial advancements in intelligent transportation systems. By automating the content creation process and customizing it to users' personalized and proactive AI requirements, the effectiveness of vehicular networks-based services can be significantly improved [4]. For instance, the generative AI could automatically generate pertinent and timely information for drivers, such as real-time traffic updates, weather forecasts, and nearby amenities. Furthermore, it could also be utilized to generate specialized driving data, such as accident simulations, to enhance real-world operational efficiency. Companies like Siemens Mobility Corp.

[1] https://chat.openai.com/

Ruichen Zhang is with the School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China, and also with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798; Ke Xiong (corresponding author) is with the Engineering Research Center of Network Management Technology for High Speed Railway of Ministry of Education, the School of Computer Science and Technology, the Innovation Center of Railway Traffic Safety, and the National Engineering Research Center of Advanced Network Technologies, Beijing Jiaotong University, Beijing 100044, China; Hongyang Du and Dusit Niyato are with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798; Jiawen Kang is with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China; Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada; H. Vincent Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA.

0890-8044/24©2024 IEEE

have leveraged AI techniques to improve traffic management by generating reliable real-time traffic predictions, facilitating better routing decisions, and minimizing congestion.[2] Similarly, Waymo Corp. released a VectorNet model to predict the behavior of traffic agents and their interactions with the surrounding environment in autonomous driving scenarios.[3] Thus, generative AI can ultimately improve the efficiency and convenience of vehicular networks-based services by automating specific aspects of the content creation process and tailoring it to users' requirements.

Despite the numerous advantages offered by generative AI-enabled vehicular networks, it encounters several challenges and limitations during implementation. The most prominent issue is limited communication bandwidth, which may lead to unreliable data transmission and communication security concerns within the vehicular networks ecosystem. Additionally, latency in communication between vehicles and infrastructure, coupled with the intricacy of data processing, poses significant barriers to the widespread adoption of vehicular networks. Motivated by these, this article is an attempt to

[2] https://press.siemens.com/global/en/pressrelease/siemens-mobilityshowcases-solutions-reduce-traffic-congestion

[3] https://blog.waymo.com/2020/05/vectornet.html

provide a forward-looking research for leveraging generative AI in vehicular networks systems. **First**, we introduce different types of generative AI techniques and then describe the potential applications and challenges of these techniques in vehicular networks scenarios. *To the best of the authors' knowledge, this is the first article that presents synergy between generative AI and vehicular technologies.* **Second**, to address these challenges, a novel generative AI-enabled vehicular networks framework that integrates a multi-modal architecture is proposed, which is capable of accommodating both text and image data, ultimately offering more reliable guidance to receiving vehicles. **Third**, to further improve the reliability and efficiency of information transmission with the framework, a deep reinforcement learning (DRL)-based approach is proposed to maximize the definite system quality of experience (QoE) within the constraints of the transmission power budget and the probability of successful transmission for each vehicle. Our approach jointly optimizes transmission communication resources and generative AI resources and simulation results validate the effectiveness of the proposed approach.

GENERATIVE AI-ENABLED VEHICULAR NETWORKS

In this section, the various types of generative AI technologies is presented. Subsequently, we explore the key applications of these techniques in vehicular networks scenarios. Lastly, we discuss the prevailing challenges associated with implementing generative AI-enabled vehicular networks.

GENERATIVE AI TECHNOLOGIES

As an emerging subfield of AI, generative AI technology focuses on autonomously generating new content or data, including images, text, audio, video, and even system designs. Similar to traditional AI technologies, such as discriminative AI, generative AI technologies can also be classified into two categories, i.e., the model-based technologies and the data-based technologies. Specifically, data-based generative AI employs learning algorithms to create novel content from existing data, while model-based generative AI leverages predefined models and simulations to generate new outputs by manipulating model parameters. These techniques excel in learning patterns and structures from existing data, producing outputs that closely resemble real-world samples, and enabling machines to mimic human creativity, imagination, and problem-solving abilities.

Developing a deep understanding of generative AI techniques suitable for specific applications in vehicular networks is crucial for optimizing performance and user experience. For instance, GANs can be utilized for data augmentation, enhancing the performance of AI models in tasks such as object detection or traffic prediction systems [2]. VAEs are valuable for dimensionality reduction and feature extraction in sensor data, enabling more efficient processing and analysis [9]. Recognizing the strengths and limitations of each generative AI technique is essential for developers when implementing the most suitable strategy for

Data-based Generative AI technologies		
Technologies	**Advantages**	**Disadvantages**
Neural style transfer [5]	· Ability to combine style and content from different images · Real-time style transfer with optimized algorithms · Applicable to various forms of media	· Dependence on high-quality style and content images · Difficulty in preserving fine details · Difficulty in controlling the level of stylization
Deep belief networks [6]	· Layer-wise unsupervised learning · Effective in learning hierarchies of features · Applicable to be fine-tuned using backpropagation	· Unstable performance · Require careful pre-training · Complexity in training deep structures
Transformer-based model [7]	· Effective modeling of long-range dependencies · Scalability with parallelization techniques · Superior performance on various tasks	· Memory consumption in large-scale applications · Complexity increases with sequence length · Dependence on large training datasets
Example-based synthesis [8]	· Leveraging existing data for generation · Ability to produce high-quality results · Flexibility in incorporating domain-specific constraints	· Dependence on the quality and variety of input examples · Difficulty in generalizing to unseen data · Computational complexity
Model-based Generative AI Technologies		
Technologies	**Advantages**	**Disadvantages**
Generative adversarial network [2]	· Exceptional quality of generated images · Can incorporate conditions to control generated output · Ability to address mode collapse with advanced techniques	· Training instability · Difficulty in evaluating model quality · Susceptibility to mode collapse
Variational autoencoder [9]	· Effective representation of data in a probabilistic latent space · Achieving disentangled representations · Unified framework for both inference and generation	· Blurry image generation · Limited expressive power of the prior distribution · Difficulty in choosing likelihood functions
Autoregressive model [10]	· Direct modeling of temporal dependencies · Scalability achieved through parallelization techniques · Compatibility with both discrete and continuous data	· Sequential nature limits speed · Complexity increases with sequence length · Difficulty in modeling long-range dependencies
Diffusion-based model [4]	· Generative process based on denoising score matching · Flexible noise schedule selection · Resistance to mode collapse and overfitting	· Slower generation process · Complexity in training and selecting hyperparameters

TABLE 1. The advantages and disadvantages for different generative AI technologies.

specific applications within vehicular networks, ultimately improving overall system performance and user satisfaction. To provide a clear overview, the advantages and disadvantages of various Generative AI technologies are summarized in Table 1, which can better harness the potential of generative AI-enabled vehicular networks and related applications.

APPLICATIONS

As shown in Fig. 1, by incorporating these advanced generative AI techniques, generative AI-enabled vehicular networks are capable of analyzing traffic patterns and providing drivers with real-time suggestions for alternative routes to circumvent congestion or accidents. For clarity, some major applications of generative AI-enabled vehicular networks are summarized as follows:

- **Navigation and Route Optimization:** Generative AI can be used to optimize vehicle routes by creating paths based on real-time traffic data, road conditions, and weather. By effectively analyzing data and making quick decisions, generative AI can offer optimal routes for vehicles, improving overall driving efficiency and safety. Furthermore, generative AI provides vehicles with real-time information about alternative routes and suggested driving speeds, helping them make informed decisions. In case of unexpected events like road closures, generative AI can quickly recalculate the best route for vehicles and adjust the path accordingly. For instance, Tomtom, a navigation system that uses generative AI,[4] adjusts routes dynamically by using user-generated data and employing GANs to predict traffic patterns, which allows drivers to avoid congested areas and reduce travel time. Thus, the potential of generative AI technologies for navigation and route optimization could lead to smarter and more efficient transportation systems for vehicular networks.

- **Insurance and Risk Assessment:** Generative AI models can be employed to analyze driving behavior, accident risks, and other factors, allowing for more reliable and personalized insurance pricing and risk assessments. These models can examine large amounts of historical driving data and generate synthesized instances that represent the underlying patterns and structures, which helps insurers better understand the factors contributing to accidents and other risks and adjust their pricing accordingly. For instance, Metromile, a pay-per-mile insurance provider, utilizes generative AI to analyze driver data and create personalized premiums based on individual driving habits and accident risk profiles.[5] By harnessing generative AI in insurance and risk assessment, companies can develop more precise and equitable pricing models, ultimately fostering a more efficient and fair insurance market.

- **Traffic Simulation and Prediction:** Generative AI can be used for simulating and predicting traffic scenarios, helping transportation planners and traffic management systems optimize traffic flow and decrease congestion. Specifically, generative AI
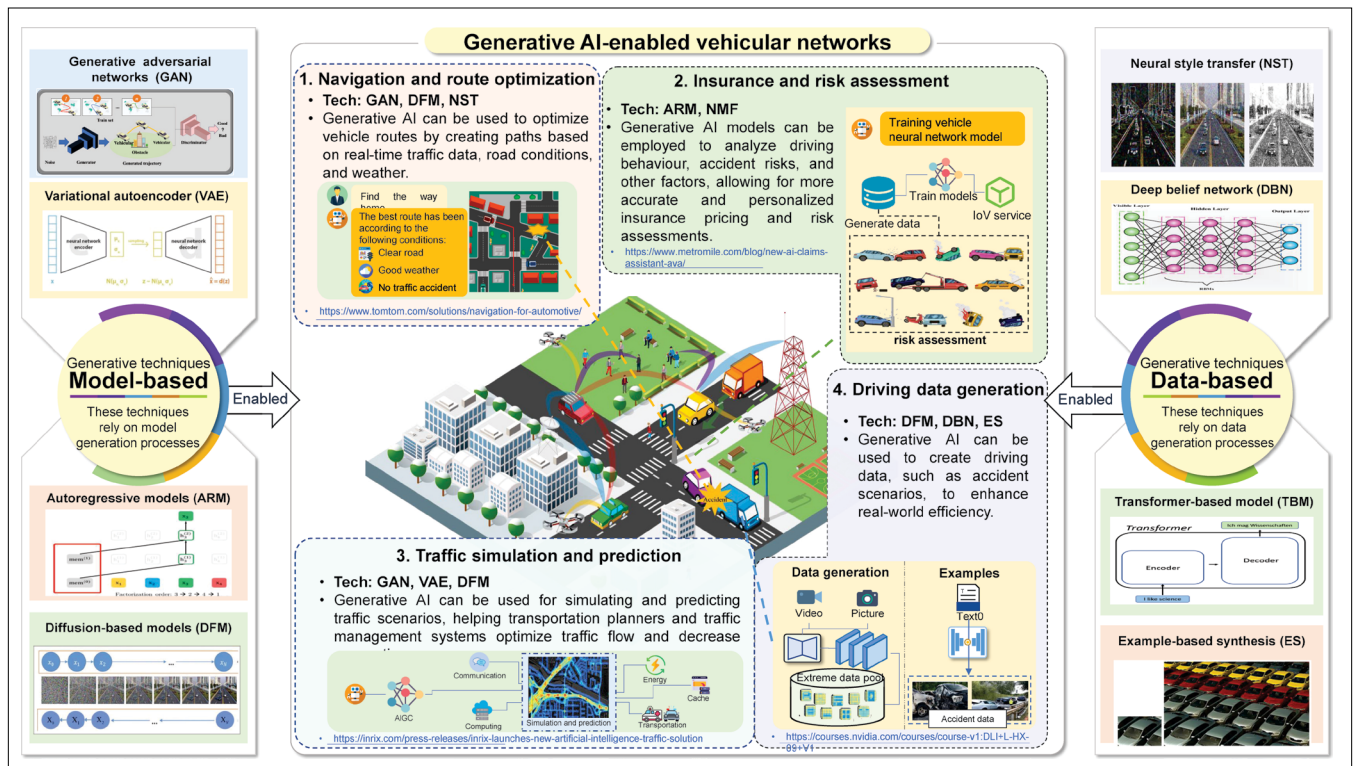
FIGURE 1. The schematic of generative AI-enabled vehicular networks. Some classical generative AI techniques (i.e., the model-based generative AI technologies and the data-based generative AI technologies) can be used to perform navigation and route optimization, traffic simulation and prediction, insurance and risk assessment, and driving data generation in vehicular networks.

models can examine historically traffic data to create reliable simulations of various traffic conditions, considering factors like time of day and road network configurations. Additionally, generative AI can be applied to forecast future traffic patterns and demands, enabling proactive adjustments in traffic signal timings, dynamic rerouting, and effective resource allocation. For instance, INRIX, a leading provider of real-time traffic and transportation data,[6] adopts transformer-based generative models to predict traffic congestion levels in real-time, allowing city planners to make informed decisions about traffic management. By using generative AI technology in traffic simulation and prediction, transportation systems can become more responsive and adaptive to changing conditions, ultimately leading to improved traffic flow and reduced congestion.

- **Driving Data Generation:** Generative AI can be used to create driving data, such as accident scenarios, to enhance road safety. In particular, this driving data is employed to train generative machine learning models, improving their reliable in predicting and preventing accidents during real-world driving. For example, NVIDIA has developed a method called "data augmentation through GANs" to generate synthetic driving situations, including rare or dangerous conditions, which can be used to train and refine autonomous vehicular systems.[7] By generating a large amount of diverse and representative data, generative AI helps close the gap between simulated and realworld data, making machine learning models more robust and adaptable. Moreover, generative AI can address the challenge of data scarcity in vehicular networks, as collecting real-world driving data is usually costly and time-consuming. By leveraging generative AI technology in driving data generation, researchers and engineers can develop more efficient and safer vehicular systems, ultimately benefiting users and the transportation industry.

Existing Challenges

While the application of generative AI in vehicular networks has exhibited substantial potential, it also presents certain challenges that can be organized into three primary domains as follows.

- **Real-Time Data Processing and Decision Making:** Generative AI models in vehicular networks necessitate the ability to process and analyze vast amounts of data in real-time, consuming considerable bandwidth in the process. Especially for AI applications that involve high-resolution images or large data streams, both the upload and download processes require significant network resources to ensure low-latency services. For instance, highdefinition maps used by autonomous vehicles, such as those provided by HERE Technologies,[8] may consist of detailed data layers with file sizes ranging from several to tens of megabytes. Moreover, due to the dynamic nature of vehicular networks and the need for up-to-date information, vehicles may make multiple repeated requests to specific edge servers to obtain reliable data.

- **Adapting to Dynamic and Unpredictable Environments:** Vehicular networks are exposed to dynamic and often unpredictable environments, which necessitate generative AI models to rapidly adapt and respond to varying conditions. Specifically, the channel quality of vehicular networks can be severely impacted by the mobility of the vehicles themselves or changes in the surrounding environment, which affects the QoE of AI-enabled vehicular networks and further lead to decreased user satisfaction.

- **Privacy and Security:** Addressing data privacy and security challenges in generative AI models for vehicular networks is critical, as these systems deal with sensitive user information. The large-scale data collection and processing by connected vehicular service providers pose concerns regarding personal information protection, as mandated by data protection laws. Specifically, vehicle-to-everything (V2X) communication systems, enabling data exchange between vehicles and connected devices, may inadvertently expose user information (e.g., license plate numbers, facial features, and vehicle details) to potential attackers, resulting in data breaches or privacy violations. Additionally, the presence of multiple parties in vehicular networks complicates interest alignment, with enhanced security potentially diminishing privacy.

Semantic-Aware Framework for Generative AI-Enabled Vehicular Networks

In this section, we propose a semantic-aware framework for generative AI-enabled vehicular networks to address current challenges and enhance road safety.

Conventional Semantic-Aware Framework for Generative AI-Enabled Vehicular Networks

Vehicle-to-vehicle (V2V) communication within vehicular networks has emerged as a promising approach to improve road safety. One potential involves using multiple cameras in vehicles to capture real-time images of road conditions. For example, during a car accident, another vehicle can capture an image of the incident and transmit it to the driver through a smart assistant, alerting other drivers and preventing further accidents. However, transmitting high-resolution images in real-time is not practical due to low latency requirements and limited bandwidth resources in vehicular networks.

As a remedy, a semantic communication image compression method is proposed. The semantic-based image compression methods can extract key information from images for transmission, thereby greatly decreasing the amount of transmitted data and reducing communication delays, which is crucial for the vehicular networks with strict delay
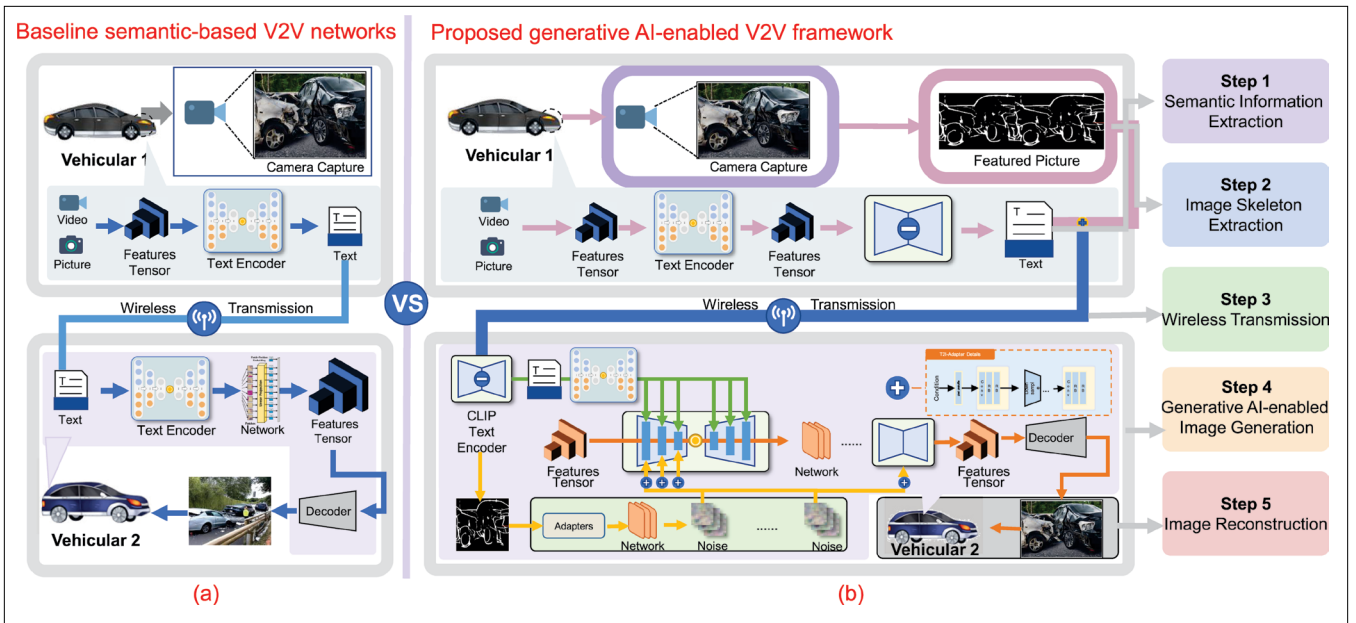
FIGURE 2. The framework of generative AI-enabled V2V framework. a) the baseline semantic-based V2V networks. b) multi-modality Semantic-aware Framework for generative AI-enabled vehicular networks. Text encoder is responsible for encoding the text information into a feature tensor. Image semantic encoder is responsible for extracting the image skeleton from the captured image and encoding it into a feature tensor. Adapter is responsible for adapting the feature tensors from the text encoder and image semantic encoder to the input format required by the generative AI algorithm. CLIP encoder is responsible for encoding the text information into a feature tensor by using the contrastive language-image pre-training method, enabling the model to understand and associate text with corresponding images.

requirements [11]. Moreover, since only the key information is extracted, the semantic-based image compression methods are able to protect and preserve the meaningful details, such as human information or license plate information in the image. As illustrated in Fig. 2(a), semantic-based V2V networks can use semantic communication technology to extract information from images at the transmitter end and restore it using generative AI technology at the receiver end.

However, a challenge remains with existing generative AI models that rely only on essential text information: They suffer from uncertainty and unreliability, making them unable to provide precise road information to smart assistants in other vehicles. For instance, when a receiving vehicle receives a message: "Two vehicles have been involved in a traffic accident near a green belt," even if a generative AI generates a high-quality image of the accident, there may be significant discrepancies between the generated image and the actual accident scene. This unreliability could result in vehicles receiving incorrect road information, which is unacceptable for safety-focused vehicular networks. Therefore, a more reliable and efficient generative AI framework is needed in vehicular networks to address these challenges and provide precise road information for better decision-making and enhanced safety.

MULTI-MODALITY SEMANTIC-AWARE FRAMEWORK FOR GENERATIVE AI-ENABLED VEHICULAR NETWORKS

In this section, we introduce a multi-modality semanticaware framework for generative AI-enabled vehicular networks. Multi-modal

approach refers to systems or solutions that combine and process information from multiple types of data inputs or sources. The advantage of multi-modal approach is that it can provide a more complete and accurate understanding of a situation by leveraging multiple types of data. If the data in one modality is unclear or ambiguous, it is also possible to leverage the data in other modalities to achieve better clarity and higher accuracy of tasks. Fig. 2(b) shows the proposed framework, which uses both text data and image data to give more reliable guidance to receiving vehicles. By using the semantic skeleton of the transmitted image and text information, the generative AI-generated images of road conditions closely match the actual images.

Step 1. Semantic Information Extraction: Similar to the baseline semantic-based V2V networks, the first step is to extract semantic text information from real-time road images. Note that in both the baseline framework and the proposed framework, the text information is obtained from the text encoder. This process uses computer vision techniques to identify and classify objects in the images, such as vehicles, pedestrians, road signs, and traffic lights.Once the objects are identified, their attributes and characteristics, such as direction, and location, are extracted to better understand the current road environment. Note that this extraction process can be optimized based on factors such as wireless transmission resources and the image reconstruction algorithms.

Step 2. Image Skeleton Extraction: Extracting the image skeleton is important because it provides a basic structure of the image

and serves as the basis for extracting more advanced semantic information. This process involves identifying the edges and contours of the road condition image and creating a simplified representation of these features. The resulting image skeleton is a streamlined version of the original image that captures its essential features without unnecessary details. The image skeleton is then used for more advanced semantic analysis, including object recognition, lane detection, and traffic sign identification, making its extraction a key step in generating reliable and useful semantic information from road condition images.

Step 3. Wireless Transmission: Upon extracting the image skeleton and text information, they are combined into a compact data package. This data package is then wirelessly transmitted to the receiving vehicle through V2V communication. Note that the data volume of the skeleton information (i.e., approximately 0.5 megabytes) is relatively small compared with that of the original captured image (i.e., approximately 6.7 megabytes), and the less bandwidth is required for transmitting only the skeleton and text prompt. Consequently, the wireless transmission effectively ensures that the receiving vehicle acquires the fundamental information needed to reliably reconstruct the road condition image, while conserving network resources.

Step 4. Generative AI-Enabled Image Generation: The generative AI model uses the extracted image skeleton and semantic information to generate a road condition image that closely resembles the actual image. The image skeleton provides the basic structure, while the semantic text information adds extra details such as road signs, vehicles, and pedestrians. By using both the image skeleton and semantic text information, the generative AI model can create a more reliable road condition image that helps vehicles make informed decisions on the road. Additionally, the model can generate images for various scenarios, such as different weather conditions, times of day, or traffic situations, giving a more complete understanding of road conditions. Note that this generation process can be optimized based on factors such as diffusion steps.

Step 5. Image Reconstruction: The generated road condition image is reconstructed on the receiving side and provided to the intelligent assistant, which alerts the driver about potential hazards. The intelligent assistant analyzes the reconstructed image and compares it with the current road condition to identify potential hazards, such as accidents or roadblocks. If hazards are detected, the assistant alerts the driver through audio or visual cues, prompting them to take necessary precautions. This process allows the vehicle to stay informed about road conditions and take appropriate actions

to ensure the safety of both its occupants and other vehicles on the road.

Therefore, compared with the baseline semantic-based V2V framework, since the proposed framework is a multi-modality semantic-aware framework, it can simultaneously use text information and image information as input to the framework to improve the reliability and accuracy of generative AI-generated images. Moreover, unlike baseline frameworks that use text encoders, the proposed framework uses a U-Net decoder and an image semantic encoder to denoise and compress the image semantic information, resulting in more precise road information for better decision-making in vehicular networks.

Case Study: Generative AI-Enabled V2V Resource Allocation

In this section, we propose a DRL-based approach to address the generative AI-enabled V2V resource allocation problem.

System Model

Following 3GPP V2X standards [12], we consider a generative AI-enabled vehicular networks consisting of multiple V2V links. The primary objective of this network is to ensure reliable and timely safety information sharing for each vehicle. To achieve this, orthogonal frequency division multiplexing technology is adopted and each V2V link has varying achievable transmission rates at different sub-channels.

In addition to the transmission rates, the image (i.e., the skeleton) and text (i.e., the prompt) information transmission frequencies also vary and dynamically change for each V2V link. We define the probability of successful transmission as a constraint, considering the advanced driving services of vehicle mobility. This constraint takes into account the achievable transmission rate, image similarity metric, channel coherence time, and generated image payload.

Since transmission rate and image similarity are two critical indicators in generative AI-enabled V2V systems, we adopt a novel QoE metric based on the Weber-Fechner law to integrate these two indicators into the system optimization goal [13]. Specifically, the Weber-Fechner law describes the relationship between the magnitude of a physical stimulus and the perceived intensity, which in our case refers to the transmitted image quality. By incorporating this law into our QoE metric, we can account for both the transmission rate and the image similarity in a unified manner. To optimize generative AI-enabled V2V resource allocation, we formulate the problem as follows: Our objective is to maximize the system QoE within the constraints of the transmission power budget and the probability of successful transmission for each vehicle. To achieve this, we need to optimize several parameters, including channel selection, transmission power for each vehicle, and the diffusion steps for inserting the skeleton. By doing so, we can effectively allocate resources to each V2V link in a way that maximizes system performance while ensuring reliable and regular safety information sharing among vehicles in the vehicular networks.

Proposed DRL-Based Approach and Simulations

For the considered problem, we propose a double deep Q-network (DDQN)-based approach to address the resource allocation problem in generative AI-enabled V2V communication. DDQN is a deep reinforcement learning technique that employs two neural networks [14], namely the online network and the target network, to approximate the Q-value function. Specifically, the online network selects actions based on the current state, while the target network is utilized to estimate the Q-value of the subsequent state. The DDQN-based approach enhances the stability of the training process and prevents the overestimation of Q-values. Moreover, the action space, state space, and reward function are designed as follows.

1) Actions: The action space is composed of the optimizing parameters, i.e., the selectable channel, the transmit power, and the diffusion steps. Specifically, for the selectable subchannel, binary variables indicate whether to select the subchannel or not. For the transmit power and diffusion steps, feasible values are quantized into multiple values to accommodate practical circuit limitations and facilitate learning.

2) States: In defining the state space, our aim is to incorporate as much relevant environmental information as possible for the considered problem [15]. In the proposed DDQN-based approach, the state space consists of two parts: the current information and the previously selected action. Specifically, the current information depends on the channel information of each V2V link, the transmission rate of each V2V link, and the generated image payload. The previously selected action depends on the action taken during the previous time step.

3) Rewards: For the proposed RL-based method, the constraints are ensured by constructing an efficient reward function. Specifically, the reward function takes into account the objective function and the constraints of the problem under consideration. The objective function is regarded as an instant reward term, indicating the unconstrained system QoE achieved at each time step. The constraint under consideration is regarded as a penalty term, indicating that a larger penalty is assigned if the constraint is not satisfied. By combining these two components, the reward function enables the agent to achieve high goals while satisfying the corresponding constraints. Such a reward function design has been widely adopted in existing works on DRL, see e.g., [15].

Fig. 3 shows an example of the multi-modality semantic-aware generative AI-enabled V2V, where the received image skeleton is extracted from the captured road condition image. It shows that in the received vehicles with the semantic encoder, the semantic information for the keywords of the prompt that generate the image is presented, where the image representation reflects the importance of each keyword through different brightness levels, with higher brightness indicating greater weight assigned to the corresponding keywords. The reason is that semantic information is designed to intelligently emphasize specific keywords (i.e., two vehicles, a green belt, and a traffic accident) during the image generation process, which facilitates the creation of more reliable traffic images based on the image skeletons and text prompts, leading to a clearer
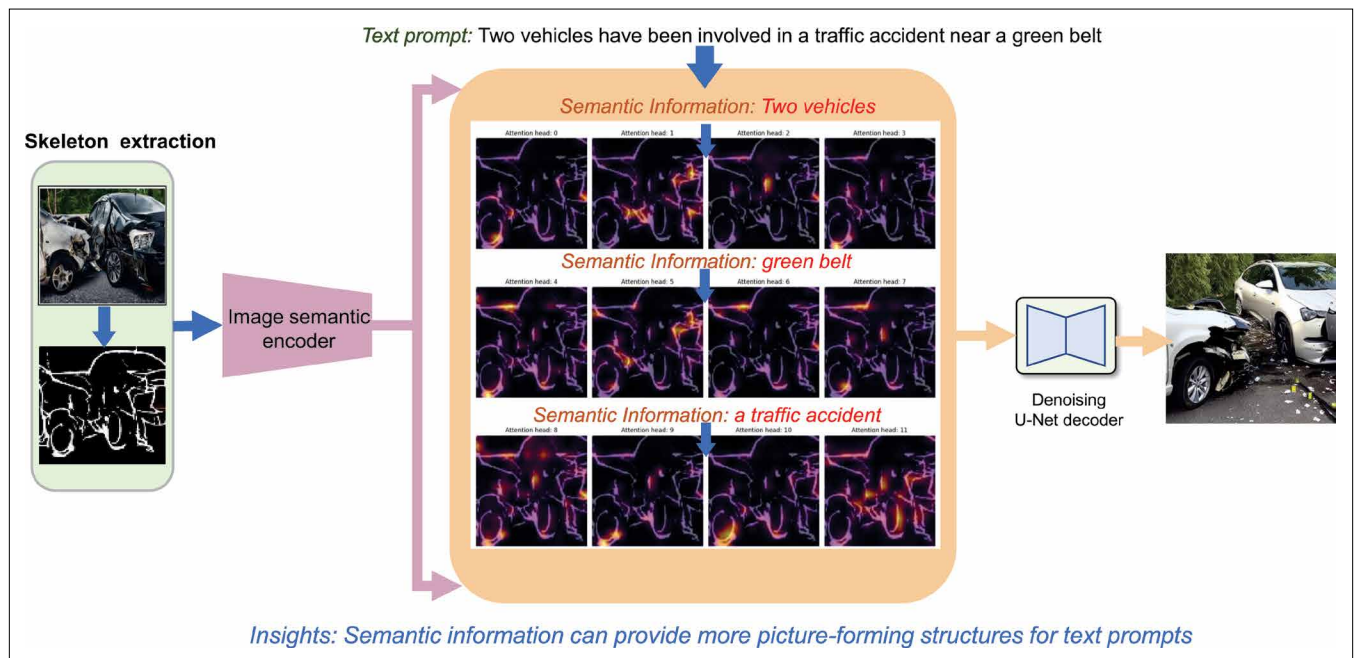

FIGURE 3. Compressing the image semantic information in multi-modality generative AI. First, the skeleton of the captured image is extracted through the image skeleton extraction algorithm, and then the key information of the skeleton image is extracted through the Image semantic encoder. Next, it is combined with textual information through a multi-modal approach. Finally, the received image is generated through denoising U-Net decoder.
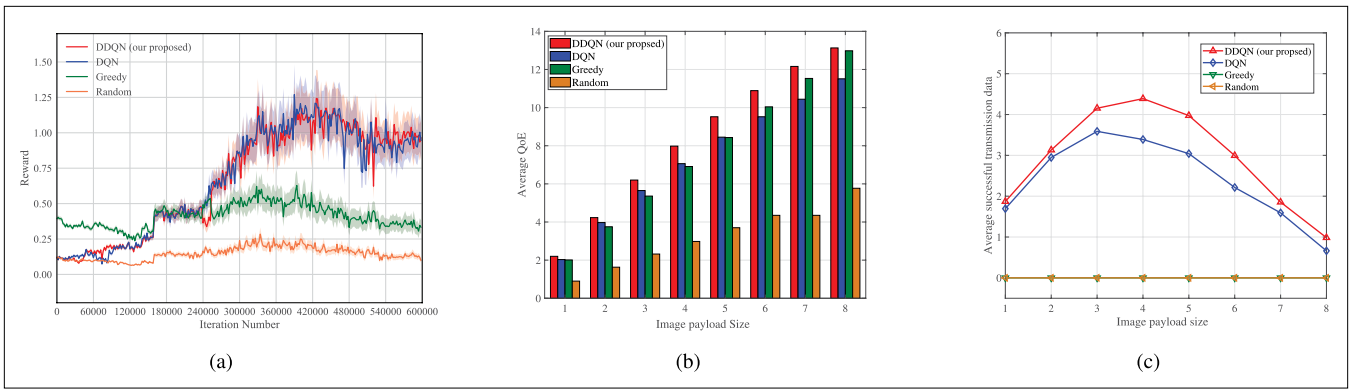
FIGURE 4. Simulation results. a) The obtained reward versus the number of iteration number with the different approaches. b) The average QoE versus the size of image payload with the different approaches. c) The average successful transmission data with the different approaches.

and more comprehensive understanding of the road conditions.

In the training stage, Fig. 4(a) compares of the average rewards obtained by four types of resource allocation strategies as the iteration number increases, where the corresponding curves are smoothed via a sliding window to provide a clearer overall trend of the raw data. It shows that the proposed DDQN-based approach consistently achieves higher rewards and converges more rapidly compared to the greedy-based and random-based approaches. Furthermore, although the rewards acquired by the proposed DDQN and DQN during the training phase are roughly similar, the reward value of the proposed DDQN surpasses that of DQN after convergence, and the fluctuations in the proposed DDQN-based approach are smaller than that in DQN. This observation can be attributed to the proposed DDQN-based approach's ability to prevent overestimation by decoupling the Q-target.

Next, in the testing stage, Fig. 4(b) shows the QoE achieved by four types of resource allocation strategies as the size of image payload. Note that the QoE is calculated based on the unconstrained system at each time step, which is adopted to reflect the quality of the generated image. It is seen that the proposed DDQN-based approach is superior to the other benchmarks. It also shows that the increment of the size of image payload is beneficial to the system QoE, as a larger payload implies a more substantial amount of transmitted semantic information.

To further examine the impact of image payload size on received QoE, Fig. 4(c) is plotted. Note that the QoE is calculated under the premise of ensuring the successful transmission rate and satisfying all other constraints. It is observed that the proposed DDQN-based approach is superior to other benchmarks, where the amount of data successfully transmitted under random and greedy are always 0. This is because the two methods do not guarantee the successful transmission probability constraint for each vehicle. Moreover, it is interesting to find that as the image payload increases, the average amount of data for successful image transmission initially rises and then declines. The phenomenon may be attributed to vehicular networks congestion and elevated probability of packet loss (i.e.,

decreased transmission success rate) under high payload conditions. Consequently, the number of retransmissions may increase, thereby reducing the overall transmission success rate.

FUTURE DIRECTIONS

SUSTAINABLE AND ENERGY-EFFICIENT SOLUTIONS

As the transportation sector moves towards electrification and sustainability, generative AI-enabled vehicular networks can play a crucial role in promoting energy-efficient and environmentally friendly solutions. Future research could explore how generative AI can optimize routes, speeds, and driving behaviors to reduce energy consumption and emissions. Additionally, generative AI-enabled vehicular networks could facilitate the integration of electric vehicles (EVs) and renewable energy sources by intelligently managing charging schedules and energy consumption based on the availability of renewable energy and the demands of the grid.

CROSS-DOMAIN COLLABORATION AND STANDARDIZATION

As generative AI-enabled vehicular networks applications continue to expand, it will become increasingly important to establish collaboration and standardization across various domains. This includes the development of unified communication protocols and data formats to facilitate seamless interaction between vehicles, infrastructure, and other connected devices. Additionally, the collaboration between researchers, industry stakeholders, and policymakers will be essential to create comprehensive regulatory frameworks that ensure safety, privacy, and equitable access to generative AI-enabled vehicular networks services.

COMBINED WITH ADVANCED MULTI-MODAL TECHNOLOGIES

Modern vehicular networks increasingly rely on advanced multi-modal technologies that integrate multiple data inputs such as text, images, audio and sensor feedback to optimize responsiveness and accuracy. This integration enables vehicles to gain a richer understanding of their environment, which is critical especially for safety and other applications. As vehicle communications transition from V2V to V2X frameworks, coordinating these different data patterns becomes critical.

Future research is needed not only to develop algorithms that seamlessly merge these different data types, but also to address challenges related to data privacy, storage, and efficient transmission.

Conclusion

In this article, we have introduced generative AI-enabled vehicular networks, emphasizing generative AI technologies and their application scenarios. Following this, we have proposed a multi-modality semantic-aware framework designed to enhance the service quality of generative AI, which encompassed both semantic and multi-modal technologies. To further improve system transmission efficiency, we have proposed a DRL-based approach to tackle the generative AI-enabled V2V resource allocation problem. Our simulation results have validated the effectiveness of the proposed approach. Finally, we have summarized potential research directions in the realm of generative AI-enabled vehicular networks.

References

[1] R. Lu et al., "5G vehicle-to-everything services: Gearing up for security and privacy," *Proc. IEEE*, vol. 108, no. 2, pp. 373–389, Feb. 2020.
[2] S. W. Kim et al., "DriveGAN: Towards a controllable high-quality neural simulation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5816–5825.
[3] Z. Yang et al., "SurfelGAN: Synthesizing realistic sensor data for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11115–11124.
[4] C. Mou et al., "T2I-Adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models," 2023, *arXiv:2302.08453*.
[5] Y. Jing et al., "Neural style transfer: A review," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 11, pp. 3365–3385, Nov. 2020.
[6] Y. Hua, J. Guo, and H. Zhao, "Deep belief networks and deep learning," in *Proc. Int. Conf. Intell. Comput. Internet Things*, Jan. 2015, pp. 1–4.
[7] E. A. M. van Dis et al., "ChatGPT: Five priorities for research," *Nature*, vol. 614, no. 7947, pp. 224–226, Feb. 2023.
[8] M. Fisher et al., "Example-based synthesis of 3D object arrangements," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–11, 2012, doi: 10.1145/2366145.2366154.
[9] W. Jin, R. Barzilay, and T. Jaakkola, "Junction tree variational autoencoder for molecular graph generation," in *Proc. Mach. Learn. Res.*, 2018, pp. 1–10. [Online]. Available: https://proceedings.mlr.press/v80/jin18a.html
[10] H. Du et al., "Beyond deep reinforcement learning: A tutorial on generative diffusion models in network optimization," 2023, *arXiv:2308.05384*.
[11] H. Xie et al., "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.
[12] M. H. C. Garcia et al., "A tutorial on 5G NR V2X communications," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1972–2026, 3rd Quart., 2021.
[13] H. Du et al., "Attention-aware resource allocation and QoE analysis for metaverse xURLLC services," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 7, pp. 2158–2175, Jul. 2023.
[14] R. Zhang et al., "Joint coordinated beamforming and power splitting ratio optimization in MU-MISO SWIPT-enabled HetNets: A multi-agent DDQN-based approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 2, pp. 677–693, Feb. 2022.
[15] R. Zhang et al., "Energy efficiency maximization in RIS-assisted SWIPT networks with RSMA: A PPO-based approach," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1413–1430, May 2023.

Biographies

RUICHEN ZHANG (ruichen.zhang@ntu.edu.sg) received the B.E. degree from Henan University (HENU), China, in 2018, and the Ph.D. degree from Beijing Jiaotong University (BJTU), China, in 2023. He is currently working as a Post-Doctoral Research Fellow with the College of Computing and Data Science, Nanyang Technological University (NTU), Singapore. His research interests include reinforcement learning enabled wireless networks, the Internet of Things, and generative models.

KE XIONG (kxiong@bjtu.edu.cn) received the B.S. and Ph.D. degrees from Beijing Jiaotong University (BJTU), Beijing, China, in 2004 and 2010, respectively. From April 2010 to February 2013, he was a Post-Doctoral Research Fellow at the Department of Electronics Engineering, Tsinghua University, Beijing. He is currently a Full Professor and the Vice Dean of the School of Computer Science and Technology, BJTU. His current research interests include wireless cooperative networks, wireless powered networks, and network information theory.

HONGYANG DU (hongyang001@e.ntu.edu.sg) received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree with the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests include semantic communications, generative AI, integrated sensing and communications, and metaverse.

DUSIT NIYATO (Fellow, IEEE) (dniyato@ntu.edu.sg) received the B.Eng. degree from the King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a Professor with the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests include the Internet of Things (IoT), machine learning, and incentive mechanism design.

JIAWEN KANG (kavinkang@gdut.edu.cn) received the Ph.D. degree from the Guangdong University of Technology, China, in 2018. He was a Post-Doctoral at Nanyang Technological University, Singapore, from 2018 to 2021. He is currently a Professor with the Guangdong University of Technology, China. His research interests include blockchain, security, and privacy protection in wireless communications and networking.

XUEMIN (SHERMAN) SHEN (Fellow, IEEE) (sshen@uwaterloo.ca) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor of electrical and computer engineering, University of Waterloo, Canada. His research interests include wireless communication networks, including capacity analysis, mobility, and radio resource management.

H. VINCENT POOR (Life Fellow, IEEE) (poor@princeton.edu) received the Ph.D. degree in EECS from Princeton University, Princeton, NJ, USA, in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana–Champaign. Since 1990 he has been on the faculty at Princeton, where he is currently the Michael Henry Strater University Professor. His research interests include information theory and signal processing, and their applications in wireless networks, energy systems, and related fields.