Generative Edge Intelligence for IoT-Assisted Vehicle Accident Detection: Challenges and Prospects

Jiahui Liu, Yang Liu, Kun Gao, and Liang Wang

Abstract

With the emergence of generative intelligence at the edge of modern Internet of Things (IoT) networks, promising solutions are proposed to further improve road safety. As a crucial component of proactive traffic safety management, vehicle accident detection (VAD) encounters multiple existing challenges in terms of data accuracy, accident classification, communication latency, etc. Thus, generative edge intelligence (GEI) can be introduced to VAD systems and contribute to improving performance by augmenting data, learning underlying patterns, and so on. In this article, we investigate the integration of GEI technology in VAD systems, focusing on its applications, challenges, and prospects. We begin by reviewing conventional VAD methods and highlighting their limitations. Following this, we explore the potential of GEI in IoT-assisted VAD and then propose a novel architecture for the GEI-VAD system that is based on an end-edge-cloud framework. We delve into the details of each component and layer within the system. Finally, we conclude this article by suggesting avenues for future research.

Introduction

Road traffic incidents are major contributors to global injury and economic loss. As per the World Health Organization's statistics, vehicle accidents result in over 1.2 million fatalities annually [1]. Consequently, it has become imperative to address the issue of mitigating the detrimental effects of vehicle accidents.To tackle this problem, vehicle accident detection (VAD) is developed to in an instant transmit the location of an accident and the key information on incident casualties to first aid centers, thus decreasing death rates. As VAD has emerged as a crucial component of proactive traffic safety management, the advent of cutting-edge technologies like the Internet of Things (IoT) catalyze the improvement of its performance [2]. Furthermore, edge computing that enables massive data processing and analysis for edge devices has been introduced, which delivers intelligent services at the edge of IoT networks, specifically near data sources like mobile devices or sensors. Edge servers carry out computations and process raw data, subsequently transmitting the refined data to the centralized server [3]. This approach significantly reduces the amount of data sent to the cloud, thereby alleviating pressure on the cloud and addressing network latency and bandwidth limitations. Since the primary challenges faced by VAD are late accident reporting, inaccurate geographic location, and lack of injured medical information, edge computing can leverage the following benefits:

- **Faster Response:** edge computing processes data closer to the source of generation, reducing the time taken to transmit data to a centralized cloud for processing. Furthermore, decisions can be made in real time based on local data processing, which allows for faster response time from emergency services.
- **Enhanced Accuracy:** by deploying advanced algorithms and adaptive machine learning methods, edge computing provides enhanced data collection and processing capabilities for VAD, including information on the severity of the accident, the incident location, and other relevant details that can be used for post-incident analysis.
- **Better Resilience:** edge computing provides better resilience to network failures and other disruptions, as it continues to function even if the centralized computing resources are unavailable. This guarantees that any sudden occurrence of vehicle accidents can be covered by the detection system.
- **Improved Privacy and Security:** by processing sensitive data locally, edge computing decreases the risk of data breaches and unauthorized access to private information involved in vehicle accidents.
- **Reduced Resources and Costs:** edge computing reduces data transmission over IoT networks, conserving bandwidth and preventing congestion. Additionally, it reduces the reliance on costly centralized computing resources, leading to cost savings in infrastructure and maintenance for VAD systems.

In the edge computing environment, deep learning generative models have exhibited creativity, scalability, and a better understanding of the underlying mechanisms and dynamics of complex systems. The scarcity of vehicle accident data

This work was supported by the National Key Research and Development Program of China (2023YFB4301800), National Natural Science Foundation of China (52220105001, 52221005, 72322002, 72331001, 72361137001), State Key Laboratory of Intelligent Green Vehicle and Mobility under Project No. KFY2421, JPI Urban Europe and Energimyndigheten (e-MATS, P2023-00029) and the Area of Advance Transport at Chalmers University of Technology.

Digital Object Identifier: 10.1109/IOTM.001.2300282

Jiahui Liu, Yang Liu (corresponding author), and Liang Wang are with the School of Vehicle and Mobility, Tsinghua University, China; Kun Gao is with the Department of Architecture and Civil Engineering, Chalmers University of Technology, Sweden.

and the inadequacy of conventional detection algorithms to accurately analyze and categorize accidents present significant challenges for VAD. To address this, generative models excel in uncovering the inherent patterns and characteristics of the data distribution, enabling the generation of numerous similar samples from a limited amount of data [4]. By continuously training neural network models using deep learning techniques, these models can effectively overcome the limitations posed by the lack of data and enhance the accuracy of VAD. Thus, generative edge intelligence (GEI), which implements generative models at the edge of IoT networks, is expected to combine with VAD systems to provide emerging solutions to existing obstacles. In VAD systems, the essence of GEI is adeptly generating and enhancing data. In contrast, the centralized cloud server can utilize a variety of algorithms for accident classification and decision-making, exceeding the purview of edge intelligence.

Fueled by the mentioned advantages, we first give an overview of IoT-assisted VAD methods. Subsequently, we discuss the potential deployments of GEI technologies for VAD. Next, we outline the fundamental architecture of GEI-VAD systems, elucidating the system components and operational mechanisms in detail. Finally, we conclude our article and propose future research directions.

VAD Methods in IoT Networks

The incorporation of IoT technologies substantially improves VAD methods' capability in terms of detection, precise localization, reporting, modeling, and analysis of vehicle accidents [5]. This section gives an overview of current IoT-assisted VAD methods and the prevailing challenges encountered in this field.

Overview

Conventional VAD typically necessitates manual investigation at the accident scene, and the data collection with single devices like surveillance cameras often results in substantial gaps in post-accident analysis. In contrast, IoT-assisted VAD utilizes sensors, embedded computers, and communication modules installed on vehicles to enable precise feedback of crucial data before and after the accident, achieving real-time monitoring and processing of vehicle status and behavior. In the IoT environment, the cloud can store a vast amount of real-time data, including vehicle motion parameters, driver status, and road conditions. By deploying edge computing, VAD not only saves computational resources of the cloud but also enables timely identification of potential risks and proactive measures at the edge to prevent accidents. Some VAD systems directly intervene after detecting anomalies, like activating the Airbags during the time of accidents.

Specifically (Fig. 1), a conventional IoT-assisted VAD system consists of three phases. Vehicles on the road are regarded as detection objectives of which state parameters will be standardized as input to the accident detection and classification models. In the first phase, the system employs a group of sensors equipped onboard to acquire real-time state data and send it to the cloud. The next phase relies on cloud servers with computing and storage capabilities. The obtained data is then processed and analyzed by decision and classification algorithms to intelligently determine whether vehicle incidents have occurred. Finally in the last phase, upon the detection of accidents, an alert message emerges and is immediately transmitted to the traffic control departments, emergency services providers, and relatives of the vehicle owner. As an example, an IoT-based Smart Accident Detection and Insurance Claiming System (ISADICS) is proposed to quickly notify nearby hospitals, police departments, and insurance companies, enabling them to promptly reach the accident site and carry out their responsibilities [6]. Some VAD sensing modules, in addition to collecting data, also provide vehicle positions with a Global Positioning System (GPS) and utilize in-vehicle cameras to monitor the inside conditions and assist in assessing the severity of the accident [7].

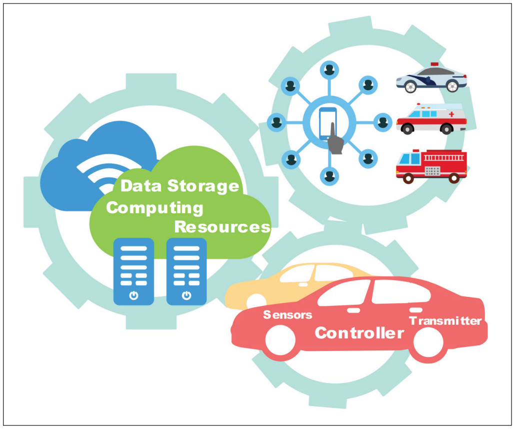

FIGURE 1. IoT-assisted VAD system framework.

Challenges

In addition to the benefits that come with the integration of IoT networks, there are certain limitations associated with previous approaches.

Detection accuracy is the primary challenge to consider, as it determines the subsequent series of decisions. Some VAD systems rely on one-modal data, such as video or audio, which is easily affected by noise and disturbances. Moreover, communication signals from technologies like GPS can be lost or interfered with in certain environments. In addition, some VAD systems identify collisions or crashes based on conventional vehicle operating data (e.g., speed, acceleration, change-in-altitude, pitch, and roll). This type of data is highly susceptible to road conditions and individual driver behavior, leading to misjudgments and omissions.

Another challenge arises from the fact that certain existing VAD systems frequently struggle to accurately determine accident types, which impedes both the establishment of liability and the execution of medical rescue operations. As illustrated in Fig. 2, vehicle accident types are mainly categorized into side collision, frontal collision, vehicle fall, roll-over, etc [8]. The major reasons VADs fail to accurately distinguish the aforementioned types of accidents are data and algorithm barriers. Data-related obstacles include noise and errors during collection, insufficient varieties used for classification (e.g., misclassification

Conventional VAD typically necessitates manual investigation at the accident scene, and the data collection with single devices like surveillance cameras often results in substantial gaps in post-accident analysis.

Communication performance also urgently needs to be addressed. Conventional VAD communication modules typically transmit data through GPS and Global System for Mobile Communications, which can be limited by restricted communication ranges, slower speeds, higher costs, and unreliable network stability.

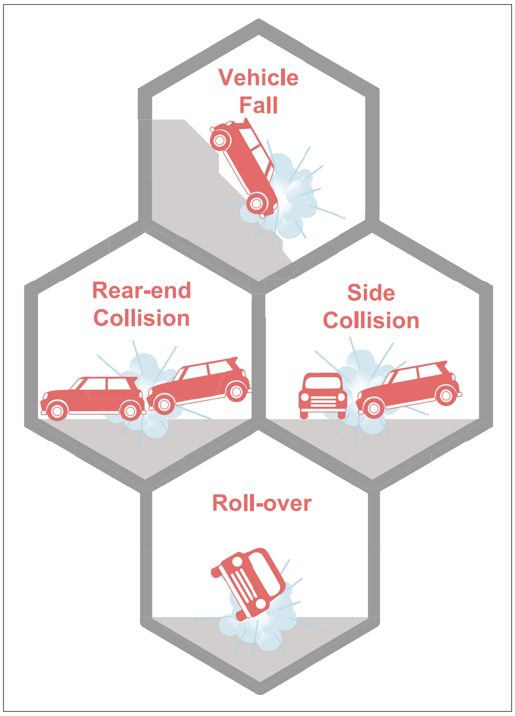

FIGURE 2. Typical types of vehicle accidents.

may occur if kinematic data is used alone without visual data), and the need for large amounts of labeled data for training. As for algorithms, different accident types may necessitate distinct recognition and classification models, with the demand to compute within a short time [9]. For instance, complex collisions involving multiple vehicles cannot be recognized by a single-vehicle accident classification model.

Communication performance also urgently needs to be addressed. Conventional VAD communication modules typically transmit data through GPS and Global System for Mobile Communications (GSM), which can be limited by restricted communication ranges, slower speeds, higher costs, and unreliable network stability. With the emergence of more efficient IoT platforms, Vehicle Ad-hoc Networks (VANETs) have become a common mode of communication. However, VANETs face significant challenges in terms of security and privacy, as well as routing due to their highly dynamic topology.

Applications of GEI Technology For VAD

The application of generative models in edge computing is recently gaining attention. The key advantage of edge computing lies in its ability to perform fast and low-latency data processing at the source of data generation, which helps reduce the latency and bandwidth consumption associated with data transmission [10]. This is particularly important for generative models, which typically require significant computational resources for both model training and data generation. In recent years, generative models have achieved remarkable success and have been broadly utilized to expand original datasets and balance label distribution through data augmentation in various fields, including image text, and sound data generation and processing [11].

In the following, we outline five prominent categories of generative models. For clustering and classification problems, the Gaussian Mixture Model (GMM) is widely applied, which is a probabilistic model that uses multiple Gaussian distributions to describe the data distribution. The Hidden Markov Model (HMM) is a statistical sequence model that is primarily used for time series data modeling and prediction, which is usually deployed in speech recognition and translation. As previously noted, generative models play a prominent role in data generation. The Autoencoder (AE) is a neural network-based generative model that learns an efficient representation of input data, allowing it to generate new data similar to the input data. Therefore, AE is commonly applied in image generation, denoising, and feature extraction. As a variant of AE, the Variational Autoencoder (VAE) introduces randomness, making the generated data more diverse. VAEs are mainly used for generating new data instances. The Generative Adversarial Network (GAN) is extensively recognized. It is a network structure composed of a generator and a discriminator, where the generator generates data and the discriminator determines whether the generated data is real or fake. GANs have wide applications in image generation, style transfer, and super-resolution.

Leveraging the respective strengths of multi-class generative models and the advantages of edge-cloud services, GEI technologies for VAD systems encompass three primary applications: data augmentation, accident classification, and active safety control.

- **Data Augmentation:** providing data to meet the requirements including diversity, multi-modality, and large volumes, is challenging with sensor networks [12]. GEI can generate synthetic data to expand existing datasets, addressing the issue of limited or imbalanced real-world accident data. Moreover, GEI technology provides VAD with unseen categories and data variations, aiding in hazardous vehicle condition detection and prediction. This is particularly crucial as simulating vehicle limit states and accident scenarios, as well as collecting diverse data are both challenging and costly.
- **Accident Classification:** GEI outperforms conventional classification methods in scene understanding. Through extensive training on diverse data, GEI models can discern intricate patterns and relationships that may not be immediately apparent. This enables VAD to make precise decisions even in scenarios that are not part of the training sets. Consequently, accident classification becomes more accurate and nuanced, thereby enhancing the development of rational rescue plans.
- **Active Safety Control:** beyond traditional VAD functions, the integration of GEI enables the pre-accident analysis and the realization of active safety control for vehicles [13]. To facilitate decision-making, GEI establishes the interconnections and functional responsibilities among its components to ascertain a secure driving pattern for the vehicle. In terms of planning, GEI generates real-time, collision-free trajectories involving critical driving parameters such as longitudinal distance, speed control, and lateral vehicle positioning.

GEI-VAD System Architecture

System Components

Considering the potential applications and benefits of integrating GEI technology into VAD systems, this section delves into the fundamental components of the GEI-VAD system.

- **Database:** various categories of data related to the accident are required [14]. In addition to vehicle and driver data, a database of hospitals is crucial as the system needs to determine the nearest hospital using mapping services and then provide first aid instructions.
- **Detection:** detection is the core component of the system and can be divided into two phases. Firstly, by applying a network of on-board sensors, real-time states of vehicles are collected. Secondly, detection methods like deep learning algorithms are used to analyze data and determine the occurrence of incidents.
- **Notification:** the ability to notify rescue and other authorities promptly determines whether the system can directly reduce fatalities and minimize other damages caused by accidents. Thus, notification is another essential component in support of most post-incident services.

End-Edge-Cloud Framework

To incorporate the system components mentioned above, we propose a three-layer GEI-VAD system framework. From bottom to top are the end layer, edge layer, and cloud layer, where each layer is responsible for different tasks (see Fig. 3).

Consisting of devices equipped with computing and storage capabilities, the end layer encompasses a vast network of sensors installed on vehicles, mobile devices, and infrastructures on the road. On one hand, onboard sensors are responsible for monitoring the vehicle's real-time state parameters and obtaining information about nearby vehicles through vehicle-to-vehicle communication, aiding in assessing the likelihood of a vehicle accident. On the other hand, road infrastructure, such as surveillance cameras and microwave vehicle detectors, also provide on-site data to the cloud layer for decision-making. In the three-layer framework, devices at the end not only possess basic functions like perception and data collection but also perform intelligent operations like data pre-processing. By processing data locally in the end layer, only compacted processed data is transmitted to upper layers (edge or cloud layers), thereby minimizing latency and network costs.

At the core of enhancing the efficiency of VAD services is the edge layer, comprising numerous edge servers and deploying generative models. These servers act as a pivotal connection between the end layer and the cloud layer, incorporating components such as routers, IoT gateways, and access points. The primary responsibilities of the edge layer include functions such as authentication, edge inference, model selection, and request forwarding. Specifically, it authenticates and authorizes trusted devices for secure collaboration. Edge nodes, sited on mobile devices and road infrastructures, can receive and store data from end devices, perform edge training or inference, and deliver final results to end devices or offload intermediate results to cloud servers through communication networks. Due to their proximity to data sources compared to cloud data centers, edge processing can reduce service latency and provide supplementary computing resources for computationally intensive intelligent applications.

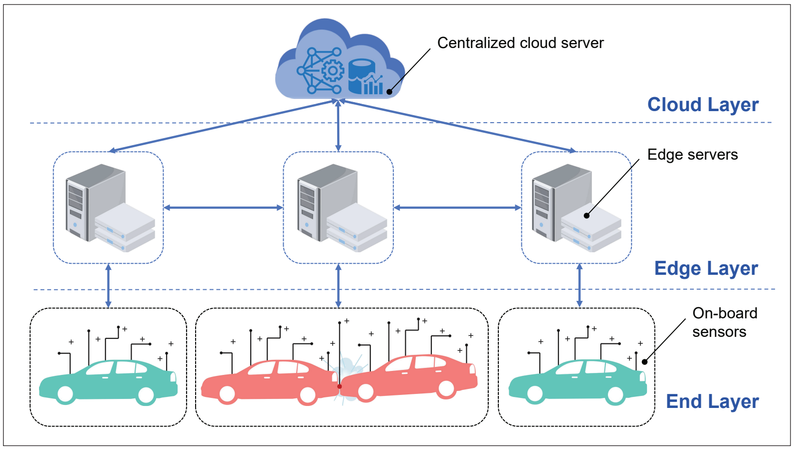

FIGURE 3. GEI-VAD system architecture based on end-edge-cloud networks.

The cloud layer is endowed with robust computational prowess and expansive storage reserves, facilitating cloud data centers to provide on-demand services that bolster a diverse spectrum of generative model applications. Specifically, the cloud layer implements tasks including data storage, device management, control, and complex model training. Notably, the cloud layer leverages its vast data and resources to achieve high-accuracy model training through the use of generative intelligence. In addition, the cloud layer serves as a coordinator between the end layer and edge layer by receiving data or intermediate results from both layers and providing robust resource support.

Deployment and Application

We present a practical application example as well as a detailed deployment of the GEI-VAD system. Regarding the sensor layout, we can integrate data from smartphones and in-vehicle sensors to establish a comprehensive sensing platform. The feature parameters for the detection and classification models include longitudinal speed, absolute linear acceleration, change-in-altitude, pitch, and roll. Consequently, vehicles must be equipped with an array of sensors, including but not limited to accelerometers, magnetometers, gyroscopes, and others. As the vehicle's speed, direction, rotation, orientation, and external forces are recorded, the real-time data can be supplemented using generative algorithms like VAE at the edge. As the pre-processed data is transmitted to the centralized cloud server, GMM which surpasses other GEI methods in both classification and clustering tasks, can serve as an invaluable tool for aiding the decision-making process.

Future Research Directions

While we provide insights into the integration of GEI and VAD systems, it is valuable to conduct further studies addressing the following issues:

- **Device Compatibility and Interoperability:** to establish the end-edge-cloud network architecture, it is necessary to investigate how to achieve cross-platform and cross-device compatibility. Further research should involve developing relevant standard specifications and common collaboration protocols to enable different equipment vendors to realize interoperability between heterogeneous edge devices and systems. Meanwhile, the diversity of hardware, software, and access methods in the IoT networks poses challenges to the data access function, which also affects the deployment of GEI applications.

- **Data Encryption:** VAD systems not only involve road traffic data and vehicle status parameters but also utilize sensitive health data of numerous patients. Therefore, encryption is necessary in the architecture to prevent data leakage. On the one hand, strong encryption algorithms should be employed during data transmission to ensure data integrity and confidentiality. On the other hand, encrypted storage is required for data processing on the end layer to prevent unauthorized access or misuse. In addition to encryption, authentication, privacy protection, and access control are also essential components in creating a secure GEI environment.
- **Generative Model Interpretability:** the decision-making and classification results of the GEI-VAD system will be referenced by multiple parties (e.g., traffic control departments, emergency services providers, and other parties involved in accidents). However, the interpretability of current deep learning algorithms needs improvement [15]. Enhancing the transparency and interpretability of generative models can increase user trust in the GEI-VAD system and aid in its supervision and management.
- **Communication Framework:** vehicle accident analysis requires timeliness. As GEI significantly enhances computational efficiency and plays the core role of the VAD systems, the construction of a corresponding information transmission architecture needs to be proposed. Therefore, the key research directions in the future include building a V2X communication architecture suitable for VAD and improving performance in terms of bandwidth, network consumption, packet transmission delay, clustering performance, and packet loss.

Concluding Remarks

This article highlights the advantages of integrating GEI technologies with VAD systems. By providing an overview of current IoT-assisted VAD methods and exploring the potential application of GEI, we propose a novel architecture for the GEI-VAD system that follows an end-edge-cloud framework. The integration of GEI technology can significantly enhance the performance of VAD in terms of timeliness, accuracy, and stability, thereby contributing to improved road traffic safety. Furthermore, there is substantial value in extensive research on both generative modeling and edge computing technology, as well as their integration with VAD.

References

[1] T. Anjuman *et al.*, "Road Traffic Accident: A Leading Cause of the Global Burden of Public Health Injuries and Fatalities," *Proc. Int'l. Conf. Mech. Eng. Dhaka Bangladesh*, 2020, pp. 29–31.

[2] C. Dashora, P. E. Sudhagar, and J. Marietta, "IoT Based Framework for the Detection of Vehicle Accident," *Cluster Computing*, vol. 23, 2020, pp. 1235–50.

[3] Z. Zhou *et al.*, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing," *Proc. IEEE*, vol. 107, no. 8, 2019, pp. 1738–62.

[4] A. Oussidi and A. Elhassouny, "Deep Generative Models: Survey," *2018 Int'l. Conf. Intelligent Systems and computer vision (ISCV)*, 2018, pp. 1–8.

[5] M. Balfaqih *et al.*, "An Accident Detection and Classification System Using Internet of Things and Machine Learning Towards Smart City," *Sustainability*, vol. 14, no. 1, 2021, p. 210.

[6] K. L. Narayanan *et al.*, "IoT Based Smart Accident Detection & Insurance Claiming System," *2021 3rd Int'l. Conf. Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2021, pp. 306–11.

[7] E. P. Ijjina *et al.*, "Computer Visionbased Accident Detection in Traffic Surveillance," *2019 10th Int'l. Conf. Computing, Commun. and Networking Tech.* (ICCCNT), 2019, pp. 1–6.

[8] Đ. Petrović, R. Mijailović, and D. Pešić, "Traffic Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers," *Transportation research procedia*, vol. 45, 2020, pp. 161–68.

[9] J. Xiao, "SVM and KNN Ensemble Learning for Traffic Incident Detection," *Physica A: Statistical Mechanics and its Applications*, vol. 517, 2019, pp. 29–35.

[10] D. Xu *et al.*, "Edge Intelligence: Architectures, Challenges, and Applications," *arXiv preprint arXiv:2003.12172*, 2020.

[11] Q. Cai *et al.*, "Real-Time Crash Prediction on Expressways Using Deep Generative Models," *Transportation Research part C: Emerging Technologies*, vol. 117, 2020, p. 102697.

[12] A. B. Parsa *et al.*, "Real-Time Accident Detection: Coping with Imbalanced Data," *Accident Analysis & Prevention*, vol. 129, 2019, pp. 202–10.

[13] S. Aggarwal and N. Kumar, "Path Planning Techniques for Unmanned Aerial Vehicles: A Review, Solutions, and Challenges," *Computer Commun.*, vol. 149, 2020, pp. 270–99.

[14] A. Vangala *et al.*, "Blockchain-Enabled Certificate-Based Authentication for Vehicle Accident Detection and Notification in Intelligent Transportation Systems," *IEEE Sensors Journal*, vol. 21, no. 14, 2020, pp. 15,824–38.

[15] A. Ross *et al.*, "Evaluating the Interpretability of Generative Models by Interactive Reconstruction," *Proc. 2021 CHI Conf. Human Factors in Computing Systems*, 2021, pp. 1–15.

Biographies

Jiahui Liu (jh-liu22@mails.tsinghua.edu.cn) received the B.S. degree in automotive engineering from the School of Automobile Engineering, Jilin University in 2022. Currently, she is a Ph.D. student in mechanical engineering at the School of Vehicle and Mobility, Tsinghua University, under the guidance of Professor Xiaobo Qu. Her research interests include intelligent vehicle control and transportation digital-twin technology.

Yang Liu (thu_ets_lab@tsinghua.edu.cn) received the Ph.D. degree at Southeast University, China, in 2021. Currently, he is a research associate at the School of Vehicle and Mobility, Tsinghua University. His research interests include intelligent transportation systems, AI techniques, and data mining, with applications in complex real-world problems such as autonomous vehicles and urban mobility. He serves as the Associate Editor of *IEEE Transactions on Intelligent Vehicles* and *Journal of Intelligent and Connected Vehicles*. In addition, he is a Young Editor for two top-tier academic journals, including the Innovation and IEEE/CAA Journal of Automatica Sinica.

Kun Gao (gkun@chalmers.se) is currently Assistant Professor in the Department of Architecture and Civil Engineering, Chalmers University of Technology (CTH). He received his Bachelor and Ph.D. degrees from Tongji University. His research works on promoting sustainable mobility with focuses on electrification, shared mobility, and connected automation. Special interests are attached to establishing new approaches and tools for system planning, optimization and evaluation of emerging transport systems leveraging big data and machine learning. The overall goal is to facilitate the development of a safer, more sustainable and equitable transportation system.

Liang Wang (wangliang0117@mail.tsinghua.edu.cn) received his B.Eng. degree and M.Eng. degree from Tsinghua University, Beijing, China, and his Ph.D. degree from Columbia University, New York, USA. He is currently an Assistant Research Professor with the School of Vehicle and Mobility at Tsinghua University. His research interests include transportation digital-twin technology, building information modeling, ground-air cooperation, and UAV applications. Additionally, he serves as the associate editor of the *Journal of Intelligent and Connected Vehicles*.