TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 05/18 pp270–287
DOI: 10.26599/TST.2020.9010025
Volume 27, Number 2, April 2022

Artificial Intelligence for Edge Service Optimization in Internet of Vehicles: A Survey

Xiaolong Xu, Haoyuan Li, Weijie Xu, Zhongjian Liu, Liang Yao, and Fei Dai*

Abstract: The Internet of Vehicles (IoV) plays a crucial role in providing diversified services because of its powerful capability of collecting real-time information. Generally, collected information is transmitted to a centralized resource-intensive cloud platform for service implementation. Edge Computing (EC) that deploys physical resources near road-side units is involved in IoV to support real-time services for vehicular users. Additionally, many measures are adopted to optimize the performance of EC-enabled IoV, but they hardly help make dynamic decisions according to real-time requests. Artificial Intelligence (AI) is capable of enhancing the learning capacity of edge devices and thus assists in allocating resources dynamically. Although extensive research has employed AI to optimize EC performance, summaries with relative concepts or prospects are quite few. To address this gap, we conduct an exhaustive survey about utilizing AI in edge service optimization in IoV. Firstly, we establish the general condition and relative concepts about IoV, EC, and AI. Secondly, we review the edge service frameworks for IoV and explore the use of AI in edge server placement and service offloading. Finally, we discuss a number of open issues in optimizing edge services with AI.

Key words: edge service; internet of vehicles; artificial intelligence

1 Introduction

In 2017, the cumulative production and sales of Chinese cars were 25.721 million and 25.769 million, respectively[1]. The increasing number of vehicles has caused a number of problems, including traffic accidents, traffic congestion, and so on. Hence, great attention has recently been paid to improve vehicle safety and efficiency. An emerging paradigm called the Internet of Vehicles (IoV) has been proposed. IoV has played a crucial role in helping avoid traffic accidents, easing traffic congestion, and providing diversified services because of its powerful capability of collecting real-time vehicle information[2–5]. However, the latency of transferring big traffic data collected by IoV to a centralized resource-intensive cloud platform hardly meets the requirement of providing real-time IoV services to vehicular users[6, 7]. Therefore, Edge Computing (EC) that deploys physical resources near Road-Side Units (RSUs) is involved in IoV. However, EC is interfered by several factors, including changes in device positions, users' variable requirements, and so on. Traditional optimization methods (e.g., game theory and convex optimization) do not meet the requirements that change dynamically in IoV. The three main factors explaining this gap are as follows:

(1) Continuous changes in the position of

• Xiaolong Xu, Haoyuan Li, Weijie Xu, Zhongjian Liu, and Liang Yao are with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China. E-mail: njuxlxu@gmail.com.
• Xiaolong Xu is also with the Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science and Technology and Engineering, Nanjing 210044, China, and the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China.
• Fei Dai is with the School of Big Data and Intelligence Engineering, Southwest Forestry University, Kunming 650233, China. E-mail: flydai.cn@gmail.com.
∗ To whom correspondence should be addressed.
Manuscript received: 2020-06-28; revised: 2020-07-14; accepted: 2020-08-02

connected vehicles. The position of Base Stations (BSes) and RSUs is stationary, but the connected vehicles are continuously moving. Traditional methods are unable to forecast users' positions, and thus fail to plot suitable resource allocation strategies.

(2) Variable application requests. Various requests in different application scenarios are sent to meet users' requirements (e.g., weather forecast, destination navigation, and entertainment). However, with traditional methods, only one type of service for service is generally considered.

(3) Changing network and time-varying channels. In IoV, information is transmitted via a wireless network, which is not always stable. Additionally, the channels that transmit mobile information vary with time. In general, traditional methods consider optimization problems in an ideal network environment.

These factors hinder traditional methods from addressing the problems caused by the complexity and dynamic feature of information transmission. Hence, the emerging Artificial Intelligence (AI) technology has been widely considered in addressing these problems in three aspects:

(1) Edge server placement. Connected vehicles keep moving and changing their positions; thus, dynamically predicting the placement of edge servers according to service demands will improve the efficiency of information transmission[8, 9]. AI technology is capable of learning from massive data generated by IoV to establish the relationship between edge server placement and information transmission delay. Therefore, AI has been used to predict optimal deployment. The prediction of coarse-grained service demands is conducted to form a network topology with a low-latency information transmission. In summary, the overall latency of the whole IoV is reduced by edge computing with the assistance of AI.

(2) Computation Offloading (CO). The computation latency of EC is high due to insufficient computing and radio resources that are shared by multiple vehicles[10]. Therefore, CO is utilized to meet various requirements with limited resources. The key step in CO is plotting the offloading strategy because the offloading process is affected by several factors, such as user habits, quality of backhaul connections, and vehicle performance[11]. AI technology is utilized to ease computation pressure and reduce feedback latency because it is capable of performing adaptive, efficient, self-organizing, and model-free learning to plot an efficient CO strategy[12].

(3) Service offloading. When offloading complex services to EC nodes, their destinations should be determined according to various factors, such as transmission latency, energy consumption, etc.[13]. However, traditional measures (e.g., game theory[14] and convex optimization) are not capable of dealing with dynamic problems affected by numerous factors because they do not consider the simultaneous influence of multiple factors on service offloading[15, 16]. AI technology compensates for this deficiency because it can work out the optimal service offloading strategy directly from a complex environment[17].

Despite the extensive researches and numerous experiments on the utilization of AI in optimizing IoV[18–22], the systematic summaries that identify the basic concepts and development strategies in AI for edge service optimization in IoV remain limited. The existing surveys on IoV mainly tackle the aspects of architectures and applications[23–25], scalability perspectives and quality of services[26], privacy and security[27], and the challenges and opportunities[28]. In this work, we address such gap, by presenting a detailed and complete survey about relative achievements in this area.

In this paper, Section 2 introduces the related concepts and definitions. Section 3 presents a survey on edge service frameworks for IoV, and Section 4 discusses a survey on AI for edge server placement. Then, Section 5 expounds AI for CO in EC-enabled IoV. Section 6 describes AI for service offloading across edge servers in IoV. Section 7 discusses several open issues. Finally, Section 8 concludes the paper.

2 Preliminary

In this section, we review the definitions of IoV, AI, EC, and edge server.

2.1 Internet of vehicles

IoV is a kind of mobile network consisting of vehicles which support the internet of things. It helps maintain traffic flow through the use of modern electronic equipment, such as sensors, global positioning systems, brakes, and throttles, along with information integration. It also performs effective fleet management and aids accident avoidance. IoV can be seen as a new paradigm that stresses the interaction of information with vehicles and humans; in such an environment, vehicles are linked to devices, such as intelligent cameras and actuators[29]. IoV is also the integration of the internet of things and mobile internet. It consists of vehicles that come into contact with one another, RSUs, handheld devices of pedestrians, and public networks. It is a rising field in the automotive industry and a vital part of smart cities[30]. IoV is an integrated and open network system which has high controllability, operationalization, and reliability, and consists of multiple users, multiple vehicles, and multiple networks[31]. Its existence is based on the architecture comprising the network layer, application layer, and perception layer[32]. IoV is also perceived to have storage, intelligence, and the ability to communicate along with learning capabilities that meet the customers' expectations[33].

In Ref. [34], IoV was viewed as a superset of VANETs[35], which extends the scale, applications, and structure of VANETs. It emphasizes the interaction of information from vehicles, RSUs, and human users. It also aims to provide people with road traffic information without latency to ensure their travel comfort. In contrast to VANETs, IoV involves vehicles as smart objects that possess multiple sensors that are connected to the internet and feature computational abilities[36].

2.2 Artificial intelligence

AI is a new branch of computer science that primarily involves computers that can simulate the thinking processes and intelligent behaviors of human beings[37]. AI mainly includes the theory of realizing intelligence by computers and manufacturing computers which are as intelligent as a human brain and thus makes computers realize high-level applications. AI manages to understand the essence of human intelligence and responds like human beings[38]. As the concept concluded, AI involves the study of the imitation and understanding of human intelligence along with the law of human behavior. In the field of AI, its research involves many other fields, such as robots, language recognition, image recognition, and expert systems. AI has greatly developed since it was proposed.

2.3 Edge computing

EC is recognized as an open platform that integrates computation, network, storage, and the key capabilities of applications. Edge devices are located on the side close to the data source and thus EC is able to produce a fast response to network service, and perform timely computation at the network edge[39].

As defined in Ref. [40], EC refers to a kind of computing paradigm in which the capabilities of processing, communication, and intelligence are pushed to the edge of network systems from which data originate. EC is also a kind of distributed computing mode[41] which can solve some of the main issues by transferring storage resources and computation from centralized points, which is, in other words, pushing data, all kinds of services and applications closer to the requests geographically[42]. In theory, EC is an extension of the content delivery network architecture. EC is also being promoted as a kind of strategy which realizes highly available and scalable network services, such as the example in Ref. [43]. It pushes the processing of data and business logic to proxy servers in the network edge[44].

In Ref. [45], "edge" was defined as any computation and network resources located between cloud data centers and all kinds of data sources. EC is viewed as a technology which allows computation to be executed at the network "edge" of networks. As a fundamental method of the IoT network, EC is a platform which is able to eliminate the burden of processing data at a centralized infrastructure and the issues related to personal privacy[46]; in this way, it differs from cloud computing, which is restricted in terms of system efficiency and data transfer[47].

2.4 Edge server

An edge server is a kind of edge device placed in internet exchange points to allow different networks to link and share transmission[48] and thus provides entry points into networks. Generally, edge servers are arranged at the network edge so as to ensure that the computation is performed near data sources. Hence, numerous edge servers ought to be deployed to reduce transmission latency[49]. Deploying a number of edge servers optimally at the basic physical base stations, such as network edges, where data are generated, can help avoid the redundant utilization of bandwidth[11]. With the deployment of edge servers, end users can enjoy edge service with low-latency anytime and anywhere. Such level of access can increase the influence of the abundant data on central data centers and backbone networks.

3 Edge Service Framework for IoV

In this section, we sort out several proposed edge service frameworks for IoV and then put forward ours. The relative architectures and their features are listed in Table 1.

Table 1 Edge service frameworks and their features.

Architecture	Feature	Related work
Edge computing-enabled software-defined IoV	Edge computing with entirety which consists of vehicles and road infrastructures.	Ref. [50]
RSUs helped by edge servers	Some edge servers arranged to be collocated with a number of specific RSUs to deal with massive data.	Ref. [51]
Edge servers, vehicles, and user devices with vehicle-to-everything (V2X) communication	Two scenarios based on the effects of vehicles.	Ref. [52]
A structure with one centralized cloud server along with multiple single RSUs	Can be equipped with mobile edge computing servers and act as an edge node.	Ref. [53]
A multi-access vehicular network	Consisting of one macro-base station and multiple RSUs.	Ref. [54]
A kind of centralized structure which makes use of a Software-Defined Network (SDN) controller	Guarantee not only efficiency but also flexibility to manage edge infrastructures such as RSUs.	Ref. [55]
Introducing an information-centered network architecture into IoV	Has two kinds of nodes with computation resources to deal with computing tasks and can analyze big data.	Ref. [56]
Three-layer architecture	(1) Physical-world layer; (2) edge computing layer; and (3) social network layer.	Ref. [57]
	(1) Mobile user layer; (2) mobile edge computing layer; and (3) cloud layer.	Ref. [58]
	(1) Infrastructure layer; (2) edge computing layer; and (3) core computing layer.	Ref. [59]
	(1) Data collection layer; (2) network layer; and (3) data processing and knowledge discovery layer.	Ref. [60]

3.1 Survey on proposed framework

As stated in Ref. [50], the authors proposed a kind of EC with entirety which consists of vehicles and road infrastructure to realize services which are sensitive to latency and thereby ensure that advanced IoV applications could be supported well. They introduced an SDN as an orchestrator which facilitates EC nodes and later developed an EC-enabled software-sefined IoV as a result.

In Ref. [51], RSUs' lack of data processing ability prompted the collocation of edge servers with a number of specific RSUs to deal with massive social media data collected by RSUs, and facilitate communication with cloud access points along with other RSUs and edge servers. The framework proposed in Ref. [52] comprises edge servers, vehicles, and user devices, along with V2X communication. In Ref. [53], the authors considered a kind of structure with one centralized cloud server along with multiple single RSUs equipped with mobile EC servers; this structure can be seen as an edge node with a mobile EC server that has finite resources of data, cache, and storage. In Ref. [54], the authors considered a multi-access vehicular network consisting of one macro-base station and multiple RSUs.

In Ref. [55], the authors proposed a kind of centralized structure which makes use of an SDN controller to program, deploy, and operate the networks in a logic-centralized way. This structure consists of mobile EC servers and Wide Area Network (WAN). This centralized structure guarantees not only efficiency but also flexibility to manage edge infrastructure such as RSUs. In Ref. [56], the authors aimed to analyze big data at the network layer and thus introduced an information-centered networking architecture called Information-Centered Network (ICN) to IoV and proposed an architecture called edge-MapReduce, which has Mapper nodes and Reducer nodes with computation resources to deal with computing tasks.

The model proposed in Ref. [57] comprises three layers, namely, the physical-world layer, EC layer, and social network layer. Physical objects, such as vehicles, drivers, and intelligent devices, constitute vehicular networks in the physical-world layer. Unmanned Aerial Vehicles (UAVs) act as flying RSUs to support mobile EC services by executing computation tasks offloaded from vehicles in the EC layer. In the social network layer, a social relation structure is constructed by the social ties of vehicles. Ji et al.[58] proposed a framework for vehicle EC with three layers, namely, mobile user layer, mobile EC layer, and cloud layer. By offloading mobile data to nodes with computing and storage capabilities such as RSUs, computational tasks are offloaded to mobile EC servers or even remote cloud servers. In Ref. [59], the authors proposed collaborative Vehicular Edge Computing (VEC), which also comprises an infrastructure layer, an edge computing layer, and a core computing layer. Each layer connects with a form of computation, including local computing, edge computing, and remote cloud computing; collaborating is possible between any two layers. Similarly, in Ref. [60], a thress-layer architecture was proposed. The three layers are the data collection layer; network layer with infrastructure devices, vehicles, and mobile devices; and the data processing and knowledge discovery layer, which presents various aspects of knowledge discovery which can be applied to the data collected from physical domains and cyberspaces[61].

3.2 Edge server in IoV

As illustrated in Fig. 1, we give the introduction of the integration of RSUs, edge servers, vehicles, and cloud data center. In the IoV structure, all the raw data and requests of vehicles are collected by RSUs, which lack data processing capabilities. To deal with this deficiency, we arrange edge servers close to RSUs to help carry out the processing of data. This arrangement enables the edge servers to connect directly to RSUs, access the data collected by them, and provide service and processing with low latency relative to RSUs. We also place one macro-base station as a kind of auxiliary, which can reduce RSUs' burden to a certain extent. Edge servers and RSUs can also link to cloud data centers, which have strong data processing and real-time analysis capabilities.

4 AI for Edge Server Placement

Recently, the utilization of AI technology in edge

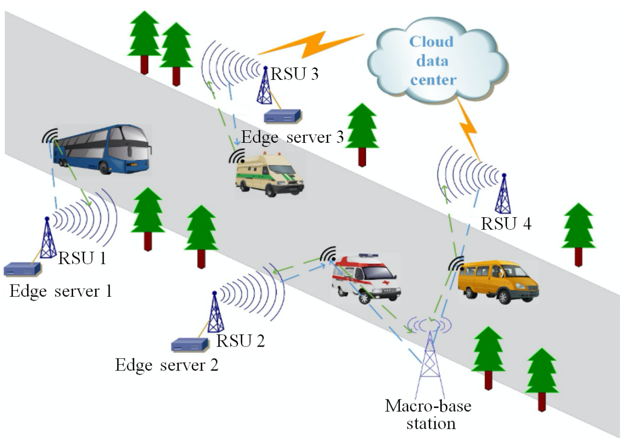


Fig. 1 Introduction of integration of RSUs, edge servers, vehicles, and cloud data center.

server placement has attracted considerable attention from numerous researchers due to its importance in formulating reasonable edge server placement strategies. Therefore, many improved methods based on deep learning algorithms have been proposed. The typical optimization algorithms are as follows.

Saputra et al.[62] proposed two new methods for designing the placement of edge servers. These two methods include deep learning algorithms and distributed deep learning algorithms for edge server placement. For the deep learning algorithms, the authors designed a centralized node-set to collect information from mobile edge nodes. Then, the collected information was used in deep learning algorithms to predict the needs of users in the entire network. The placement model of the edge server was derived on the basis of the predicted user requirements. Considering the information privacy and communication overhead in deep learning algorithms, the authors designed a framework based on distributed deep learning. In the new framework, the task of centralized nodes is to collect trained models from mobile edge nodes and update the global edge server model on the basis of the trained models. After completing this task, the global edge server model is sent to the mobile edge nodes again for the next update. In such a cyclical way, the global edge server model is continuously updated to ensure that the strategy edge server placement is optimal.

Bensalem et al.[63] laid out an edge server based on a deep neural network and proposed the best layout model. Different kinds of deep neural networks have the ability to extract various features from data because they have their own distinct structures, which can help with model optimization[64]. In the process of designing the optimal layout model, the authors[63] put forward the formula of deep neural network parameter selection after comprehensively considering the communication delay between nodes and the cost of EC nodes. They then determined the most suitable parameters of the deep neural network model through this formula. After the training process, the deep neural network model automatically provides recommendations for the placement of edge servers based on the current network conditions. After experimental verification, the authors confirmed that the edge service placement scheme given by the model under low and high network loads exerts a considerable effect on reducing the average delay of each request.

As the widely used group convolution is not suitable for devices with low memory efficiency, Yang et al.[65] proposed the Edge Convolution Neural Network (EdgeCNN) model to deal with resource scarcity and enhance memory access efficiency. The authors also improved the EdgeCNN model to obtain the Edge Convolution Neural Network-G (EdgeCNN-G) model. The similarity of the two models lies in the direct execution of the Convolutional Neural Network (CNN) algorithm on the EC server, resulting in the placement scheme of the edge server. EdgeCNN-G consists of packet computers, whereas EdgeCNN does not. As many EC servers, such as the famous Raspberry[66], have relatively few computer resources, their memory access speed is only one-tenth of that of mobile phones. Therefore, one needs to choose whether to use the EdgeCNN-G model or the EdgeCNN model to place the edge server according to the actual network environment.

Tian et al.[67] proposed the Lightweight Edge device assisted Private CNN (LEP-CNN) model. The LEP-CNN was uniquely designed with a lightweight encryption scheme to enable resource-constrained IoT devices to execute CNN requests with privacy protection to derive an edge server placement scheme. The design idea of LEP-CNN is to improve network latency and service availability by integrating local edge devices. LEP-CNN has two advantages in addition to the placement strategy of the edge server. Firstly, it can make the privacy protection on the edge device as effective as the privacy protection on the unencrypted data. Secondly, it can ensure that the privacy protection does not cause accuracy loss on the CNN. The experiment results showed that LEP-CNN has higher accuracy and better efficiency than the traditional CNN edge server placement scheme. The summary of the strategies related to AI for edge server placement is shown in Table 2.

5 AI for Computation Offloading in EC-Enabled IoV

CO is regarded as an essential technology in EC. Through the offloading of computing tasks to edge servers, the consumption of time and energy in data computations and transmissions are reduced. In IoV, RSUs deployed along roads act as edge servers. Therefore, with the assistance of RSUs, Vehicle-to-Vehicle (V2V) communication is performed efficiently.

5.1 Offloading requirement

In this sub-section, we introduce certain requirements for CO strategies for users and service providers according to the research results. These requirements mainly include energy consumption and intentions for users as well as the delay and reliability for service providers.

(1) Energy consumption for users. The emergence of equipment, such as smart phones, has enriched the lives of users. With such equipment, we can communicate expediently with others and alleviate the pressure involved in working or learning. As these convenient features entail extensive computation, the energy consumption of user equipment in computing is huge. Therefore, the battery life of equipment is not always sufficient for user calculation. In Ref. [68], the concept of Mobile Edge Computing (MEC), which involves offloading some appropriate computing tasks from the user equipment to the network edge, was introduced to decrease energy consumption.

(2) Intentions for users. In the process of offloading, the feasibility of one strategy is determined by the intentions of users.

(3) Delay for service providers. Pursuing low energy consumption excessively while ignoring acceptable limited delays is not feasible. In other words, low energy consumption should not be compromised high delay. Therefore, in Ref. [69], the authors paid attention to certain delay limitations and designed a resource allocation strategy for computing and network resources to satisfy predefined delay constraints.

(4) Reliability for service providers. The high

Table 2 Summary of strategies about AI for edge server placement.

Reference	Strategy	Actual utility
[62]	Distributed deep learning	In the distributed deep learning algorithm, the edge server placement model is continuously updated through the model transmission of centralized and edge nodes.
[63]	Deep neural network	According to the communication delay between nodes and the calculation cost, the edge server placement model of the deep neural network is obtained and the model is trained in the actual network environment to improve its performance.
[66]	Edge convolutional neural network-G	The EdgeCNN-G model is different from the EdgeCNN model as it uses a group computer to place the edge server.
[67]	Lightweight edge device assisted privacy-preserving convolutional neural network	The LEP-CNN model reduces network latency and provides privacy protection for the edge server by integrating the local edge device while placing the edge server.

reliability in data transmission is also a major challenge in CO. In Ref. [70], the authors proposed the concept of reliability factor, which determines the reliability of each edge node by periodically transferring the computing tasks and returning computing results. With the usage of the reliability factor, the offloading process achieves superior performance in terms of reliability.

5.2 Offloading strategy

In this sub-section, we introduce several studies about offloading strategies from four aspects.

(1) Strategies related to delay. A summary of these strategies is shown in Table 3.

In Ref. [71], the authors proposed a computation task offloading framework in IoV. For this framework, the authors considered the limitations of resources in vehicles and delays in CO and designed a valid resource allocation strategy-based contract theory to accomplish every service. A large amount of experimental data show that this strategy obviously reduces delay.

The development of cellular network has greatly enriched the daily lives of users. With the increasing demands of users for real-time information, IoV with cellular network is facing huge challenges brought about by delays in data transmissions. In Ref. [72], the vehicles were allowed to build wireless connections to RSUs directly so as to decrease the delay in V2V information interactions.

The high mobility of vehicles has increased the requirement for low delay in data transmission in IoV. In Ref. [73], the authors investigated a publish/subscribe message scheme starting with reducing the delay of information interaction in IoV. The edge service center deployed in the edge server provides real-time information forwarding service. Therefore, the data center in the cloud guarantees that information is spread in a larger scope. The feasibility of the scheme is verified by simulation experiments.

In motorways, numerous requirements have been established to limit delay in CO. The fast speed of vehicles leads to their high mobility, hence the need for highly accurate computations for the road conduction information. In Ref. [74], the authors established the vulnerable road user scenario of a motorway and evaluated the experiment data, which then revealed that supplementing MEC into cellular V2X reduces the delays in data transmissions.

The conventional EC cannot be easily applied to IoV directly due to vehicle mobility. In Ref. [75], a collaborative task offloading strategy was designed with the capability of guaranteeing low delay in the processes of offloading. Then, the 3D construction of the scenario was carried out to analyze the strategy scientifically. Experimental data demonstrated that the strategy not only ensures a comfortable driving experience, but also reduces the perception response time of drivers under special circumstances.

In Ref. [76], a strategy involving short packet delays in CO with ultra-reliable and low-latency communication service was designed. In the edge of the network, a processor-sharing server was set up to assist in computing, analyzing message latency, and optimizing task offloading. The computing capacity of this server was divided into all the packets. Considering the impact

Table 3 Summary of strategies about delay.

Reference	Core method	Actual utility
[71]	A valid resource allocation strategy-based contract theory	Reduce delays, augment the utilization rate of edge servers, and improve the utility of vehicles.
[72]	Vehicles building connections to RSUs	Accomplish V2V information interactions conveniently and reduce delay.
[73]	Low delay of information interaction	Provide real-time information forwarding service and guarantee the spread of information in a larger scope.
[74]	A strategy based on a motorway scenario	Reduce delay in offloading processes.
[75]	A collaborative task offloading strategy	Ensure a comfortable driving experience and reduce the perception response time of drivers under special circumstances.
[76]	Message latency and task offloading optimization	Propose a latency balance algorithm to minimize end-to-end delay.
[77]	Fight RSUs, a flight algorithm based on swarm intelligence	Flexibly implement information interaction in IoV and meet the high response requirements of applications.
[78]	An adaptive learning-based task offloading algorithm	Reduce the delay in offloading processes.
[79]	Integrating load balancing with offloading, a low-complexity algorithm	Reduce delays and maximize system utility.

of packet loss rate, the authors designed a latency balance algorithm to minimize end-to-end delay.

The emergence of RSUs has simplified vehicle communications. However, if we would like to decrease the inevitable delays in data transmissions, the number of RSUs with fixed characteristics that should be deployed along roads needs to be increased. In Ref. [77], the authors studied fight RSUs carried by UAVs and designed a flight algorithm on the basis of swarm intelligence. Through intelligence calculation for road conduction information, the authors were able to dynamically formulate the processes of task offloading and resource allocation.

In Ref. [78], the authors studied task offloading in IoV and designed a method that could make vehicles realize the offloading delay performance of adjacent vehicles while offloading computation tasks. The Adaptive Learning based on Task Offloading (ALTO) algorithm was proposed to minimize the average offloading delay. The essential characteristic of the algorithm is that the states in computing are not exchanged frequently during ALTO. Extensive simulation experiments demonstrated that the algorithm decreases delays in offloading processes.

The emergence of delay-sensitive in-vehicle applications is a new challenge for IoV. In Ref. [79], the concept of VEC was proposed. In VEC, the computing tasks are offloaded from the resource-constrained vehicles to the VEC servers. The performance of VEC becomes limited if all vehicular tasks are offloaded into only one server. Therefore, the authors proposed integrating load balancing with offloading to optimize the offloading ratio and the usage of computing resources. Then, a low-complexity algorithm was designed. Numerical experiment data demonstrated the superior performance of the algorithm in terms of the low delay in offloading.

(2) Strategies about energy consumption. A summary of these strategies is shown in Table 4.

MEC has the potential to enhance vehicular services by CO. In Ref. [80], the computation constraints of MEC servers possibly decreasing the quality of offloading in traffic-dense roads were considered, and an offloading framework based on VEC was proposed to tackle this problem. An optimization offloading scheme with game theory was designed to maximize the utility of vehicles and edge servers.

In Ref. [81], the authors studied the energy consumption in offloading processes and built a model for visual analysis based on queue theory. Then, an energy efficiency optimization problem was formulated based on the theoretical analysis. Finally, a distributed algorithm was designed to minimize the energy cost under certain delay constraints.

In Ref. [82], the authors studied the CO strategies in IoV under the 5G environment. In the study, a small-cell network architecture was designed. The authors also formulated an energy optimization problem, which was discussed in two aspects of task computation and communication. To solve this optimal problem, the authors designed an artificial fish swarm algorithm. The numerical experimental data demonstrated the superior performance of the architecture.

In Ref. [83], a collaborative offloading method based on MEC and cloud computing was proposed, in which services are offloaded into the vehicles in IoV. By optimizing the offloading decision and effectively allocating resources, the authors formulated a collaborative offloading problem based on cloud-MEC. To solve this problem, a distributed CO and resource allocation algorithm was designed. The simulation results demonstrated that the algorithm improves system

Table 4 Summary of strategies about energy consumption.

Reference	Core method	Actual utility
[80]	An offloading framework based on VEC	Maximize the utility of vehicles and edge servers.
[81]	A distributed algorithm to solve the energy efficiency optimization problem	Minimize the energy cost under the certain delay constraints.
[82]	Artificial fish swarm algorithm to solve the energy optimization problem	Improve energy efficiency and solve the problem of energy minimization.
[83]	A distributed resource allocation algorithm	Improve the system utility.
[84]	An algorithm to solve the problem of energy efficiency maximization	Proved efficient by numerical experimental results.
[85]	Partial offloading strategy, an iterative algorithm	Minimize the system energy consumption.

utility.

In Ref. [84], the authors studied an energy transfer-based MEC system. In this system, the mobile devices are charged with the radio frequency signals. The authors then formulated the energy efficiency maximization problem by computing the time distribution, energy consumption and ability of local computing in computing offloading. This optimization problem was tackled with the design of a specific algorithm. Finally, the feasibility of the algorithm was evaluated via simulation experiments.

In Ref. [85], the authors considered the partial offloading strategy based on multi-access EC network, that the data are calculated in cloud or user devices. Then an iterative algorithm was designed. With the algorithm, users could select the partial data to offload. The simulation results demonstrated that the algorithm minimizes system energy consumption and achieves superior performance.

(3) Strategies about comprehensive aspects. A summary of these strategies is shown in Table 5.

In Ref. [5], the authors firstly considered the high mobility of vehicles and the energy cost in data computation. A predictive task-file transfer strategy for V2V communication was then designed. The authors also proposed an offloading scheme with predictive combination mode. The extensive experimental data showed that, the scheme can minimize the cost of offloading and satisfy the delay constraint of users.

In Ref. [86], the power delay balance in the CO environment was studied in the scenario of multiple servers. The authors set a probability constraint for the length of the offloading tasks queue and analyzed this queue with extreme value theory. Then, a strategy for dynamic resources allocation under complicated task calculations was designed. The energy consumption of the data computations and transmissions was minimized under the constraints of delay and reliability. The simulation results demonstrated that the strategy breaks the conflict between power and delay.

In Ref. [87], the authors described a novel scenario based on MEC architecture, in which the local servers need to accomplish the real-time services of mobile vehicles in their own service range. The problem of distributed real-time service scheduling was proposed by considering the balance of service workloads and real-time services in IoV. To tackle this problem, the authors designed a utility-based learning algorithm. With the assistance of a simulation model, the authors extensively evaluated the performance of the algorithm and verified its superiority.

Smart vehicles are becoming widespread. In Ref. [88], the authors formulated the problem that the mobility of Vehicles Terminals (VTs) increases the consumption of time and energy in offloading. To tackle this problem, the authors designed a heuristic offloading decision method. Furthermore, the problem was divided into two sub-problems which were then solved accordingly. The excellent performance of the method was verified by the simulation data.

(4) Strategies about intentions of users. A summary of these strategies is shown in Table 6.

In Ref. [90], the authors aimed to reduce the communication delays in IoV and enhance the experience of users, and thus proposed a network architecture assisted by MEC based on SDN, and then studied its rapid response and scalability. The reliability of the architecture was verified by the simulation results, which met the high response and real-time requirements as well as the requirements of different applications. Software management and update were simply performed with the architecture.

In Ref. [89], the authors took into account of the limited resources in MEC servers and formulated the problem of optimally utilizing the resources while enhancing the Quality of Equipment (QoE) in VTs.

Table 5 Summary of strategies about comprehensive aspects.

Reference	Core method	Actual utility
[5]	A predictive task file transfer strategy for V2V communication	Minimize the cost of offloading processes and satisfy the users to accept limited delays.
[86]	Power delay balance in CO in the scenario involving multiple servers; analysis with extreme value theory	Optimize the energy consumption of computations and data transmissions under the limits of delay and reliability.
[87]	A utility-based learning algorithm to solve the problem of distributed real-time service scheduling	Balance the service workloads and real-time services in IoV and achieve superior performance.
[88]	A heuristic offloading decision method	Reduce the consumption of time and energy in vehicle terminals.

Table 6 Summary of strategies about intentions of users.

Reference	Core method	Actual utility
[74]	A network architecture assisted by MEC based on SDN	Meet the high response and real-time requirements as well as the requirements of different applications; easily perform software management and update.
[89]	A scheme to tackle the problem of optimally utilizing the resources while enhancing QoE in VTs	Maximize the QoE in VTs and have higher utility than conventional schemes.
[76]	An offloading scheme using TVWS	Reduce the cost in VTs and edge servers.
[77]	A Deep Reinforcement Learning (DRL) method	Maximize the utility of the vehicle edge computing network and achieve high performance.

To tackle this problem, the authors designed a novel offloading scheme, in which the QoE of VTs is maximized. The simulation data demonstrated the higher utility of the proposed scheme in comparison with those of conventional schemes.

In-vehicle applications entailed higher requirements in terms of the delay and bandwidth of information interaction. In Ref. [76], the authors regarded the TV White Space (TVWS), which compensates for the shortage of bandwidth in CO. Therefore, a cognitive vehicular network using the TVWS band was considered. A dual-side optimization cost problem was also formulated. To tackle this optimization problem, the authors designed a specific algorithm, which showed superior performance in minimizing the costs on the basis of the simulation data.

In Ref. [77], the authors designed a vehicle EC network architecture, in which the vehicles could act as the EC servers to provide adjacent vehicles with services. Then, considering the delay in computations, the authors proposed a vehicle-assisted offloading scheme. To maximize the utility of the vehicle EC network, they then designed a DRL method. The numerical results revealed the superior performance of the scheme.

IoV based on CO technology has made people's lives especially convenient. Therefore, the aforementioned CO strategies will be further optimized, and other novel strategies will be designed. In the near future, the IoV technology is expected to mature further and thus add intelligence to our lives.

6 AI for Service Offloading Across Edge Servers in IoV

In this section, we review the research on service offloading. A summary of the literatures related to the methods of service offloading across edge servers in IoV is shown in Table 7, which include each article's key points. We review the studies on the common methods of service offloading across edge servers in Section 6.1, identify the service offloading methods assisted by AI in Section 6.2, and propose an AI structure for service offloading across edge servers in IoV in Section 6.3.

6.1 Method of service offloading

As a method to minimize the service latency and maximize the optimal revenue, a kind of adaptive service offloading program, which offers service utilization and revenue to the maximum extent to the mobile EC platform, was proposed by Samanta and Chang[17]. In their program, services are offloaded to multiple edge servers according to the rate of flow of the computational services, which are organized by their priorities. Similarly, in Ref. [91], the authors proposed an adaptable offloading mechanism, which took into consideration the quality of service requirements of the executing application, especially the real-time requirements. The authors put forward an approach to code offloading, which allows applications and devices to offload some of the services to other nodes, with the aim of supporting adaptable applications.

As some mobile security suites require extensive resources, an ad-hoc (cooperative access of a set of mobile nodes, that do not require any centralized points or interaction of existing infrastructure) mobile edge cloud was presented in Ref. [79], the proposed cloud uses Wi-Fi Direct to achieve connectivity, integrate security services, and share resources with surrounding mobile devices. This scheme embeds a multi-objective resource-aware optimization model along with genetic-based algorithms to supply smart offloading decisions, which are on the basis of dynamic processing of statistical and contextual data from ad-hoc mobile devices. With the similar idea, Deng et al.[93] took advantage of multi-hop vehicular ad-hoc networks to help with the computational tasks offloading of vehicles. They also built a liability model of a multihop routing

Table 7 A summary of the literature about methods of service offloading across edge servers in IoV.

Method	Feature	Related work
Adaptive service offloading program	(1) Offload services by the rate of flow of the computational services; (2) take into consideration the quality of service requirements of the executing application.	[78]
An approach to code offloading	Allow applications and devices to offload some services to other nodes adaptably.	[91]
Ad-hoc networks	(1) An ad-hoc mobile edge cloud that makes use of Wi-Fi Direct; (2) multi-hop vehicular ad-hoc networks.	[80]
Architecture with features of multi-X	(1) A high-level edge computing orchestrator with multiple access; (2) an auction mechanism featured as multi-round-sealed sequential and combinatorial; and (3) a vehicular edge network with multiple access.	[81]
A heuristic algorithm	Focus on migrating user-generated data to the edge servers.	[82]
An SDN-based cross-layer network slicing architecture	Comprise two parts: transport layer and control layer.	[83]
Cooperative method for parallel computing and transmission in Virtual Reality (VR)	Dividing one VR task into two small tasks, both of which can be dealt with at mobile edge servers in IoV simultaneously and separately.	[84]
An offloading scheme of distributed game theory	Work in a wireless environment with multiple channels.	[85]
Visual odometry with intensive offloading	Combine visual odometry with the intensive offloading to near infrastructure and utilize a method consisting of 3 dimension to 2 dimension correspondences.	[92]
Framework with deep reinforcement learning algorithm	(1) A framework of knowledge-driven service offloading decision for IoV; (2) a system of intelligent offloading with the DRL, the finite-state Markov chain along with vehicle edges; (3) a framework based on reinforcement learning for a service migration system of single-user EC; and (4) an offloading scheme with the DRL, the knowledge of vehicular network and the mobile edge computing.	[86]
An adaptive task offloading algorithm	(1) Give vehicles the ability to learn an adjacent vehicle's offloading delay performance; (2) increase the occurrence awareness for adaptation in dynamic environment.	[87]
Active offload balancing with the deep convolutional neural networks	Have the ability to solve the optimization problem of Natural Language Processing (NLP) and implement efficient collaborative scheduling of data.	[88]
A moving edge algorithm with non-orthogonal multiple access	Consider task offloading and the selection of user between edge devices and macro units.	[90]

path on the basis of the theory of link correlation in VANETs in real-time traffic environments.

Samanta and Li[13] presented a novel latency-oblivious incentive service offloading design to maximize offloading performance and users' profit along with the estimation of the latency requirements of different user-specific applications. As the actual latency requirements of those mobile devices are hard to determine and change dynamically, regardless of EC platform, this scheme estimates the total service latency run up against by mobile devices and then takes different services' unique priorities into account.

Gilly et al.[82] presented a high-level EC orchestrator with multiple access which arranges vehicular edge services that are on the basis of location in the way of the management of hierarchical dynamic resource. With the help of this orchestrator, low latency responses are guaranteed owing to the geo-aware service dynamic migration and allocation. Tran and Pompili[83] considered a multi-cell wireless network enabled by mobile EC; in this network, each BS is equipped with an MEC server that helps mobile users execute various tasks by task offloading. In Ref. [84], an auction mechanism characterized as multi-round-sealed sequential and combinatorial was proposed to connect mobile terminals along with the mobile EC servers in order to offload tasks to the best mobile EC server. In Ref. [85], a vehicular edge network with multiple access was introduced; in this network, the vehicles are treated as edge computation resources and a mechanism of collaborative task offloading along with the output transmission guarantees low latency and the performance of applications.

By focusing on the migration of user-generated data to the edge servers and with consideration of these tasks' characteristics and the connection between nodes, the authors in Ref. [5] designed a heuristic algorithm for reducing the transmission costs by using the real-time information.

In Ref. [86], an SDN-based cross-layer network slicing architecture was proposed for the service migration in mobile EC infrastructure. This kind of structure is divided into two parts, namely, the transport layer and the control layer. The transport layer, comprises three kinds of domain controllers based on the SDN, that are built for the access layer, the convergence layer, and core layer so that service migration among mobile EC servers in each layer can be achieved. The control layer is composed of orchestrator layer along with the management layer and involves three kinds of mobile EC server placement scenarios.

To deal with the long completion time of the VR applications used in IoV, the authors in Ref. [87] proposed a kind of cooperative method for parallel computing and transmission in VR that divides one VR task into two small tasks. One of the two sub-tasks is offloaded to vehicles through wireless transmission, and the two sub-tasks are dealt with at the mobile EC servers and vehicles in IoV simultaneously and separately. Therefore, the offloading proportion and communication resource allocation are optimized jointly in vehicular networks with mobile EC and the latency is minimized greatly. In Ref. [88], an offloading scheme of distributed game theory was designed on the basis of the study of a service offloading problem, which related to multi-user mobile EC in a wireless environment with multiple channels.

6.2 Method with AI

Given the difficulty of the offloading multiple data dependency tasks in a complex service, which depends on data, researchers are inspired to use AI and have thus proposed several novel methods for the service offloading across edge servers in IoV.

In Ref. [90], the authors proposed a framework of knowledge-driven service offloading decision for IoV, which supplies the optimal strategy from the environment directly. This framework includes a decision model with the DRL algorithm to help acquire knowledge about offloading decisions along with the observation function. The model is responsible for obtaining data from EC nodes and vehicular mobility. The DRL algorithm proposed by the team DeepMind is a learning method which combines the neural networks for the storage of state with the decision-making methods of reinforcement learning. The algorithm can solve problems which are hard for traditional reinforcement learning to deal with, such as the representation of an endless number of states for the image input[94]. The combination of the two parts provides a unique platform, which can offload various vehicular services to three types of EC nodes. In Ref. [89], the authors made use of the DRL, the finite-state Markov chain along with the integration of vehicle EC to construct an intelligent offloading system. They also raised the problem of resource allocation and the joint optimization of task scheduling. They solved the problem with a DRL method integrated with a bilateral matching algorithm. In Ref. [95], the authors proposed an offloading scheme involving the knowledge of vehicular networks, mobile EC along with the DRL; the scheme improves mobile edge offloading. Similarly, the authors in Ref. [92] designed a framework based on the reinforcement learning for a service migration system of single-user EC. This kind of model can take long-term goal into account and thus facilitate communication decision making and service migration.

On the basis of the theory of multi-armed bandit, the authors in Ref. [96] proposed an adaptive task offloading algorithm with distributed characteristics. This method gives vehicles the ability to learn an adjacent vehicle's offloading delay performance while the calculation tasks are being offloaded. It also increases the occurrence awareness for adaptation in a dynamic environment and reduces the request on frequent exchanges of states.

In Ref. [97], the authors proposed a method of active offloading balancing, which makes use of the deep CNN in order to learn the spatiotemporal correlations and predict road traffic conditions. CNN is a kind of artificial neural network that utilizes shared weights to extend across space, and is thus suitable for tasks related to computer vision[98]. This method has the ability to deal with the optimization problem of NLP and implements the efficient collaborative scheduling of data which are cached between mobile edge servers.

In Ref. [69], the authors considered task offloading along with the selection of users between edge devices and macro units; on the basis of non-orthogonal multiple access, the authors proposed a moving edge algorithm along with a heuristic algorithm designed from three aspects, namely, offloading decision, channel allocation, and power control, with the aim of improving the rate gain of transmission and the efficiency of discharge offloading.

6.3 AI for service offloading across edge servers in IoV

As shown in the Fig. 2, we present an architecture of service offloading across edge servers with the help of AI in IoV. The architecture includes a decision model with DRL algorithm. As introduced in Section 3.2, edge servers are arranged close to RSUs to help transmit data because of their powerful computational capacity. We maximize such powerful capacity and train the decision model at RSUs with EC nodes. The result of training, which is distributed across service offloading decisions in the real world, is cached in edge servers or uploaded to the cloud server. Vehicles then receive the model from RSUs or download it from the cloud. As the DRL model puts forward rewards for each decision, the model is trained as long as the services are running and its parameters are sent to EC nodes to update the previous model. In this way, the model makes intelligent decisions for service offloading.

7 Open Issue

The existing research and experiments highlight a few crucial open issues on AI for edge service optimization in IoV.

(1) Optimize EC with joint measures

AI plays a crucial role in optimizing edge service, but the optimization effect of a single AI method is restricted[99]. Moreover, solving the problem of overfitting or underfitting in the model is problematic[100]. Therefore, a framework for integrating the DRL techniques with federated learning is required to promote the joint optimization of edge service in IoV.

(2) Accelerate AI tasks through EC

The task of AI model training is also computationally intensive because of the long time needed to construct an efficient model[101]. At present, the dynamic and adaptive splitting of AI tasks has become an open issue involving the use of edge service to complete AI tasks efficiently[102]. EC could provide a high-quality computing architecture for AI or a practical and feasible operation solution for some AI applications that are computationally complex.

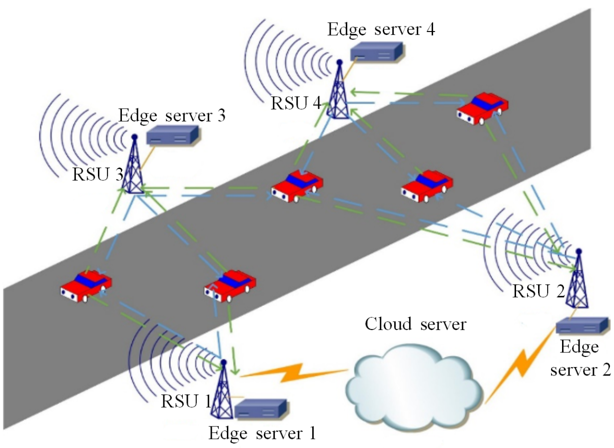


Fig. 2 An architecture of service offloading across edge servers with the help of AI in IoV.

(3) Improve security and privacy protection

Vehicles collect information mainly from communications between vehicles and vehicles and between vehicles and road-side infrastructure. These communications are often interrupted due to vehicles' high mobility. The interference causes communication links to fail frequently. Moreover, hackers' attacks on sensors and communication channels lead to severe problems in privacy. Potential solutions include the access control of servers and communication authentication[103].

(4) Deal with high mobility

With the rapidly increasing density of road traffic, the main factors that affect the network topology dynamics are frequent and high speed vehicle movements. As an important feature of vehicular networks, intelligent vehicles' high mobility not only blocks the supply of stable wireless communication, but also increases the complexity of the collaborative optimization of the computing and cache resources' allocation[104]. Studying the distribution protocols of data routing along with predicting vehicles' movement could improve this situation.

8 Conclusion

The emerging communication technology accelerates digital transformation and provides benefits for many industries, including education, energy, smart cities, and smart transportation. Given the vast development prospects of smart transportation, IoV attracts increasing attention from numerous researchers. Edge service is considered in addressing problems that involve intensive computation and high latency. AI is utilized to optimize edge service in IoV in various aspects including edge server placement, CO, and service offloading.

In this work, a comprehensive and detailed survey on AI for edge service optimization in IoV is presented. Firstly, the basic driving forces of this survey are reviewed. Secondly, the related concepts and definitions are introduced. Thirdly, an overview of frameworks and crucial techniques are provided. Finally, several open issues are enumerated to guide our future research directions. In sum, this survey is presented to promote the further progress of AI in optimizing edge service.

Acknowledgment

This research was supported by the Financial and Science Technology Plan Project of Xinjiang Production and Construction Corps (No. 2020DB005), the National Key R&D Program of China (No. 2019YFE0190500), the National Natural Science Foundation of China (Nos. 61702442, 61862065, and 61702277), the Application Basic Research Project in Yunnan Province (No. 2018FB105), the Major Project of Science and Technology of Yunnan Province (Nos. 202002AD080002 and 2019ZE005), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund.

References

[1] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Y. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, 5G: A tutorial overview of standards, trials, challenges, deployment, and practice, *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[2] W. Y. Zhong, X. C. Yin, X. Y. Zhang, S. C. Li, W. C. Dou, R. L. Wang, and L. Y. Qi, Multi-dimensional quality-driven service recommendation with privacy-preservation in mobile edge environment, *Computer Communications*, vol. 157, pp. 116–123, 2020.

[3] X. L. Xu, X. Zhang, X. H. Liu, J. L. Jiang, L. Y. Qi, and M. Z. A. Bhuiyan, Adaptive computation offloading with edge for 5G-envisioned internet of connected vehicles, *IEEE Transactions on Intelligent Transportation Systems*, doi: 10.1109/tits.2020.2982186.

[4] C. J. Zhou, A. L. Li, A. H. Hou, Z. W. Zhang, Z. X. Zhang, P. F. Dai, and F. S. Wang, Modeling methodology for early warning of chronic heart failure based on real medical big data, *Expert Systems with Applications*, vol. 151, p. 113361, 2020.

[5] X. L. Xu, Q. Wu, L. Y. Qi, W. C. Dou, S. B Tsai, and M. Z. A. Bhuiyan, Trust-aware service offloading for video surveillance in edge computing enabled internet of vehicles, *IEEE Transactions on Intelligent Transportation Systems*, doi: 10.1109/tits.2020.2995622.

[6] H. Yang, F. Li, D. X. Yu, Y. F. Zou, and J. G. Yu, Reliable data storage in heterogeneous wireless sensor networks by jointly optimizing routing and storage node deployment, *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 230–238, 2020.

[7] P. Mach and Z. Becvar, Mobile edge computing: A survey on architecture and computation offloading, *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[8] S. G. Wang, Y. L. Zhao, J. Xu, J. Yuan, and C. H. Hsu, Edge server placement in mobile edge computing, *Journal of Parallel and Distributed Computing*, vol. 127, pp. 160–168, 2019.

[9] H. Yin, X. Zhang, H. H. Liu, Y. Luo, C. Tian, S. Y. Zhao, and F. Li, Edge provisioning with flexible server placement, *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 4, pp. 1031–1045, 2017.

[10] Y. Y. Mao, J. Zhang, and K. B. Letaief, Dynamic computation offloading for mobile-edge computing with energy harvesting devices, *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3590–3605, 2016.

[11] X. L. Xu, B. W. Shen, X. C. Yin, M. R. Khosravi, H. M. Wu, L. Y. Qi, and S. H. Wan, Edge server quantification and placement for offloading social media services in industrial cognitive IoV, *IEEE Transactions on Industrial Informatics*, doi: 10.1109/tii.2020.2987994.

[12] C. M. Wang, C. C. Liang, F. R. Yu, Q. B. Chen, and L. Tang, Computation offloading and resource allocation in wireless cellular networks with mobile edge computing, *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 4924–4938, 2017.

[13] A. Samanta and Y. Li, Latency-oblivious incentive service offloading in mobile edge computing, presented at 2018 IEEE/ACM Symp. Edge Computing (SEC), Seattle, WA, USA, 2018, pp. 351–353.

[14] Q. He, G. M. Cui, X. Y. Zhang, F. F. Chen, S. G. Deng, H. Jin, Y. H. Li, and Y. Yang, A game-theoretical approach for user allocation in edge computing environment, *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 3, pp. 515–529, 2020.

[15] X. Z. Chen, H. M. Ma, J. Wan, B. Li, and T. Xia, Multi-view 3D object detection network for autonomous driving, presented at 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6526–6534.

[16] X. Y. Huang, X. J. Cheng, Q. C. Geng, B. B. Cao, D. F. Zhou, P. Wang, Y. Q. Lin, and R. G. Yang, The ApolloScape dataset for autonomous driving, presented at 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 1067–1076.

[17] A. Samanta and Z. Chang, Adaptive service offloading for revenue maximization in mobile edge computing with delay-constraint, *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3864–3872, 2019.

[18] Q. C. Cao, W. L. Zhang, and Y. H. Zhu, Deep learning-based classification of the polar emotions of “moe”-style cartoon pictures, *Tsinghua Science and Technology*, vol. 26, no. 3, pp. 275–286, 2020.

[19] E. Li, L. K. Zeng, Z. Zhou, and X. Chen, Edge AI: On-demand accelerating deep neural network inference via edge computing, *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 447–457, 2020.

[20] Y. N. Malek, M. Najib, M. Bakhouya, and M. Essaaidi, Multivariate deep learning approach for electric vehicle speed forecasting, *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 56–64, 2021.

[21] Z. Zou, Y. Jin, P. Nevalainen, Y. X. Huan, J. Heikkonen, and T. Westerlund, Edge and fog computing enabled AI

for IoT—An overview, presented at 2019 IEEE Int. Conf. Artificial Intelligence Circuits and Systems (AICAS), Hsinchu, China, 2019, pp. 51–56.

[22] A. H. Sodhro, S. Pirbhulal, and V. H. C. de Albuquerque, Artificial intelligence-driven mechanism for edge computing-based industrial applications, *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4235–4243, 2019.

[23] D. W. Wei, H. S. Ning, F. F. Shi, Y. L. Wan, J. B. Xu, S. K. Yang, and L. Zhu, Dataflow management in the Internet of Things: Sensing, control, and security, *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 918–930, 2021.

[24] L. M. Ang, K. P. Seng, G. K. Ijemaru, and A. M. Zungeru, Deployment of IOV for smart cities: Applications, architecture, and challenges, *IEEE Access*, vol. 7, pp. 6473–6492, 2019.

[25] B. F. Ji, X. R. Zhang, S. Mumtaz, C. Z. Han, C. G. Li, H. Wen, and D. Wang, Survey on the internet of vehicles: Network architectures and applications, *IEEE Communications Standards Magazine*, vol. 4, no. 1, pp. 34–41, 2020.

[26] K. Smida, H. Tounsi, M. Frikha, and Y. Q. Song, Software defined internet of vehicles: A survey from QoS and scalability perspectives, presented at 2019 15th Int. Wireless Communications & Mobile Computing Conf. (IWCMC), Tangier, Morocco, 2019, pp. 1349–1354.

[27] H. Talat, T. Nomani, M. Mohsin, and S. Sattar, A survey on location privacy techniques deployed in vehicular networks, presented at 2019 16th Int. Bhurban Conf. Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 2019, pp. 604–613.

[28] H. B. Zhou, W. C. Xu, J. C. Chen, and W. Wang, Evolutionary v2x technologies toward the internet of vehicles: Challenges and opportunities, *Proceedings of the IEEE*, vol. 108, no. 2, pp. 308–323, 2020.

[29] X. L. Xu, Y. Xue, L. Y. Qi, Y. Yuan, X. Y. Zhang, T. Umer, and S. H. Wan, An edge computing-enabled computation offloading method with privacy preservation for internet of connected vehicles, *Future Generation Computer Systems*, vol. 96, pp. 89–100, 2019.

[30] Z. Tong, F. Ye, M. Yan, H. Liu, and S. Basodi, A survey on algorithms for intelligent computing and smart city applications, *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 155–172, 2021.

[31] F. C. Yang, S. G. Wang, J. L. Li, Z. H. Liu, and Q. B. Sun, An overview of internet of vehicles, *China Communications*, vol. 11, no. 10, pp. 1–15, 2014.

[32] M. Azrour, J. Mabrouki, A. Guezzaz, and Y. Farhaoui, New enhanced authentication protocol for Internet of Things, *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 1–9, 2021.

[33] X. K. Wang, L. T. Yang, L. W. Song, H. H. Wang, L. Ren, and J. Deen, A tensor-based multi-attributes visual feature recognition method for industrial intelligence, *IEEE Transactions on Industrial Informatics*, doi: 10.1109/tii.2020.2999901.

[34] D. R. Lawrence, C. Palacios-González, and J. Harris, Artificial intelligence: The shylock syndrome, *Cambridge Quarterly of Healthcare Ethics*, vol. 25, no. 2, pp. 250–261, 2016.

[35] J. X. Li, T. T. Cai, K. Deng, X. J. Wang, T. Sellis, and F. Xia, Community-diversified influence maximization in social networks, *Information Systems*, vol. 92, p. 101522, 2020.

[36] M. Caprolu, R. Di Pietro, F. Lombardi, and S. Raponi, Edge computing perspectives: Architectures, technologies, and open security issues, presented at 2019 IEEE Int. Conf. Edge Computing (EDGE), Milan, Italy, 2019, pp. 116–123.

[37] H. H. Pang and K. L. Tan, Authenticating query results in edge computing, in *Proc. 20th Int. Conf. Data Engineering*, Boston, MA, USA, 2004, pp. 560–571 .

[38] S. Al-Janabi and A. H. Salman, Sensitive integration of multilevel optimization model in human activity recognition for smartphone and smartwatch applications, *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 124–138, 2021.

[39] M. Alrowaily and Z. Lu, Secure edge computing in IoT systems: Review and case studies, presented at 2018 IEEE/ACM Symp. Edge Computing (SEC), Seattle, WA, USA, 2018, pp. 440–444.

[40] X. X. Chi, C. Yan, H. Wang, W. Rafique, and L. Y. Qi, Amplified locality-sensitive hashing-based recommender systems with privacy protection, *Concurrency and Computation*: *Practice and Experience*, doi: 10.1002/cpe.5681.

[41] R. Bi, Q. Liu, J. K. Ren, and G. Z. Tan, Utility aware offloading for mobile-edge computing, *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 239–250, 2020.

[42] G. Manasvi, A. Chakraborty, and B. S. Manoj, Social network aware dynamic edge server placement for next-generation cellular networks, presented at 2020 Int. Conf. Communication Systems & Networks (COMSNETS), Bengaluru, India, 2020, pp. 499–502.

[43] Y. Liang, J. D. Ge, S. Zhang, J. Wu, Z. Tang, and B. Luo, A utility-based optimization framework for edge service entity caching, *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 11, pp. 2384–2395, 2019.

[44] X. L. Xu, X. H. Liu, Z. Y. Xu, F. Dai, X. Y. Zhang, and L. Y. Qi, Trust-oriented IoT service placement for smart cities in edge computing, *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4084–4091, 2020.

[45] D. Kombate and L. N. Wang, The internet of vehicles based on 5G communications, presented at 2016 IEEE Int. Conf. Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Chengdu, China, 2016, pp. 445–448.

[46] N. Sharma, N. Chauhan, and N. Chand, Smart logistics vehicle management system based on internet of vehicles, presented at 2016 4th Int Conf. Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 2016, pp. 495–499.

[47] J. X. Li, T. T. Cai, A. Mian, R. H. Li, T. Sellis, and J. X. Yu, Holistic influence maximization for targeted

advertisements in spatial social networks, presented at 2018 IEEE 34th Int. Conf. Data Engineering (ICDE), Paris, France, 2018, pp. 1340–1343.

[48] P. Lai, Q. He, M. Abdelrazek, F. F. Chen, J. Hosking, J. Grundy, and Y. Yang, Optimal edge user allocation in edge computing with variable sized vector bin packing, in *Proc. 16th Int. Conf. Service-Oriented Computing*, Hangzhou, China, 2018, pp. 230–245.

[49] X. W. Hou, Z. Y. Ren, J. J. Wang, W. C. Cheng, Y. Ren, K. C. Chen, and H. L. Zhang, Reliable computation offloading for edge-computing-enabled software-defined IoV, *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7097–7111, 2020.

[50] S. Khorsandroo and A. S. Tosun, An experimental investigation of SDN controller live migration in virtual data centers, presented at 2017 IEEE Conf. Network Function Virtualization and Software Defined Networks (NFV-SDN), Berlin, Germany, 2017, pp. 309–314.

[51] Z. J. Hu, D. Y. Wang, Z. Li, M. Sun, and W. Z. Wang, Differential compression for mobile edge computing in internet of vehicles, presented at 2019 Int. Conf. Wireless and Mobile Computing, Networking and Communications (WiMob), Barcelona, Spain, 2019, pp. 336–341.

[52] K. Wang, H. Yin, W. Quan, and G. Y. Min, Enabling collaborative edge computing for software defined vehicular networks, *IEEE Network*, vol. 32, no. 5, pp. 112–117, 2018.

[53] Y. Y. Dai, D. Xu, Y. L. Lu, S. Maharjan, and Y. Zhang, Deep reinforcement learning for edge caching and content delivery in internet of vehicles, presented at 2019 IEEE/CIC Int. Conf. Communications in China (ICCC), Changchun, China, 2019, pp. 134–139.

[54] H. W. Liu, H. Z. Kou, C. Yan, and L. Y. Qi, Keywords-driven and popularity-aware paper recommendation based on undirected paper citation graph, *Complexity*, vol. 2020, p. 2085638, 2020.

[55] X. K. Wang, L. T. Yang, X. Xie, J. R. Jin, and M. J. Deen, A cloud-edge computing framework for cyber-physical-social services, *IEEE Communications Magazine*, vol. 55, no. 11, pp. 80–85, 2017.

[56] M. Körner, T. M. Runge, A. Panda, S. Ratnasamy, and S. Shenker, Open carrier interface: An open source edge computing framework, in *Proc. 2018 Workshop on Networking for Emerging Applications and Technologies*, Budapest, Hungary, 2018, pp. 27–32.

[57] Y. M. Saputra, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, D. Niyato, and D. I. Kim, Distributed deep learning at the edge: A novel proactive and cooperative caching framework for mobile edge networks, *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1220–1223, 2019.

[58] H. Ji, O. Alfarraj, and A. Tolba, Artificial intelligence-empowered edge of vehicles: Architecture, enabling technologies, and applications, *IEEE Access*, vol. 8, pp. 61 020–61 034, 2020.

[59] S. Park and N. Kwak, Cultural event recognition by subregion classification with convolutional neural network, presented at 2015 IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 2015, pp. 45–50.

[60] Z. Y. Jiao, Y. Yang, H. R. Zhu, and F. J. Ren, Realization and improvement of object recognition system on raspberry Pi 3B+, presented at 2018 5th IEEE Int. Conf. Cloud Computing and Intelligence Systems (CCIS), Nanjing, China, 2018, pp. 465–469.

[61] Y. F. Tian, J. W. Yuan, and H. B. Song, Efficient privacy-preserving authentication framework for edge-assisted internet of drones, *Journal of Information Security and Applications*, vol. 48, p. 102354, 2019.

[62] Y. M. Saputra, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, D. Niyato, and I. K. Dong, Distributed deep learning at the edge: A novel proactive and cooperative caching framework for mobile edge networks, *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1220–1223, 2019.

[63] M. Bensalem, J. Dizdarevic, and A. Jukan, Modeling of Deep Neural Network (DNN) placement and inference in edge computing, arXiv preprint arXiv: 2001.06901, 2020.

[64] H. Sadr, M. M. Pedram, and M. Teshnehlab, Multi-view deep network: A deep model based on learning features from heterogeneous neural networks for sentiment analysis, *IEEE Access*, vol. 8, pp. 86 984–86 997, 2020.

[65] S. Z. Yang, Z Gong, K Ye, Y. G. Wei, Z. Huang, and Z. H. Huang, EdgeCNN: Convolutional neural network classification model with small inputs for edge computing, arXiv preprint arXiv: Computer Vision and Pattern Recognition, 2019.

[66] A. Jonathan, M. Uluyol, A. Chandra, and J. Weissman, Ensuring reliability in geo-distributed edge cloud, presented at 2017 Resilience Week (RWS), Wilmington, DE, USA, 2017, pp. 127–132.

[67] Y. Tian, J. Yuan, S. Yu, and Y. Hou, LEP-CNN: A lightweight edge device assisted privacy-preserving CNN inference solution for IoT, arXiv preprint arXiv: 1901. 04100v1, 2019.

[68] G. H. Qiao, S. P. Leng, K. Zhang, and Y. J. He, Collaborative task offloading in vehicular edge multi-access networks, *IEEE Communications Magazine*, vol. 56, no. 8, pp. 48–54, 2018.

[69] H. El-Sayed and M. Chaqfeh, The deployment of mobile edges in vehicular environments, presented at 2018 Int. Conf. Information Networking (ICOIN), Chiang Mai, Thailand, 2018, pp. 322–324.

[70] Y. X. Sun, X. Y. Guo, J. H. Song, S. Zhou, Z. Y. Jiang, X. Liu, and Z. S. Niu, Adaptive learning-based task offloading for vehicular edge computing systems, *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3061–3074, 2019.

[71] Y. Y. Dai, D. Xu, S. Maharjan, and Y. Zhang, Joint load balancing and offloading in vehicular edge computing and networks, *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4377–4387, 2019.

[72] K. Zhang, Y. M. Mao, S. P. Leng, S. Maharjan, and Y. Zhang, Optimal delay constrained offloading for vehicular edge computing networks, presented at 2017 IEEE Int. Conf. Communications (ICC), Paris, France, 2017, pp. 1–6.

[73] K. Zhang, Y. M. Mao, S. P. Leng, Y. J. He, and Y. Zhang,

Mobile-edge computing for vehicular networks: A promising network paradigm with predictive off-loading, *IEEE Vehicular Technology Magazine*, vol. 12, no. 2, pp. 36–44, 2017.

[74] Z. Chang, Z. Y. Zhou, T. Ristaniemi, and Z. S. Niu, Energy efficient optimization for computation offloading in fog computing system, presented at GLOBECOM 2017–2017 IEEE Global Communications Conf., Singapore, 2017, pp. 1–6.

[75] L. C. Yang, H. L. Zhang, M. Li, J. Guo, and H. Ji, Mobile edge computing empowered energy efficient task offloading in 5G, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6398–6409, 2018.

[76] J. H. Zhao, Q. P. Li, Y. Gong, and K. Zhang, Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks, *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7944–7956, 2019.

[77] J. J. Yu, M. X. Zhao, W. T. Li, D. Liu, S. W. Yao, and W. Feng, Joint offloading and resource allocation for time-sensitive multi-access edge computing network, presented at 2020 IEEE Wireless Communications and Networking Conf. (WCNC), Seoul, Korea, 2020, pp. 1–6.

[78] C. L. Li, M. Y. Song, L. Zhang, W. N. Chen, and Y. L. Luo, Offloading optimization and time allocation for multiuser wireless energy transfer based mobile edge computing system, *Mobile Networks and Applications*, doi: 10.1007/s11036-019-01396-3.

[79] C. F. Liu, M. Bennis, and H. V. Poor, Latency and reliability-aware task offloading and resource allocation for mobile edge computing, presented at 2017 IEEE Globecom Workshops (GC Wkshps), Singapore, 2017, pp. 1–7.

[80] P. L. Dai, K. Liu, X. Wu, H. L. Xing, Z. F. Yu, and V. C. S. Lee, A learning algorithm for real-time service in vehicular networks with mobile-edge computing, presented at ICC 2019–2019 IEEE Int. Conf. Communications (ICC), Shanghai, China, 2019, pp. 1–6.

[81] W. H. Zhan, H. C. Duan, and Q. X. Zhu, Multi-user offloading and resource allocation for vehicular multi-access edge computing, presented at 2019 IEEE Int. Conf. Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), Shenyang, China, 2019, pp. 50–57.

[82] K. Gilly, A. Mishev, S. Filiposka and S. Alcaraz, Offloading edge vehicular services in realistic urban environments, *IEEE Access*, vol. 8, pp. 11 491–11 502, 2020.

[83] T. X. Tran and D. Pompili, Joint task offloading and resource allocation for multi-server mobile-edge computing networks, *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 856–868, 2019.

[84] J. F. Lv, J. Y. Xiong, H. Z. Guo, and J. J. Liu, Joint computation offloading and resource configuration in ultra-dense edge computing networks: A deep reinforcement learning solution, presented at 2019 IEEE 90th Vehicular Technology Conf. (VTC2019-Fall), Honolulu, HI, USA, 2019, pp. 1–5.

[85] J. B. Du, F. R. Yu, X. L. Chu, J. Feng, and G. Y. Lu, Computation offloading and resource allocation in vehicular networks based on dual-side cost minimization, *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1079–1092, 2019.

[86] T. Dbouk, A. Mourad, H. Otrok, H. Tout, and C. Talhi, A novel ad-hoc mobile edge cloud offering security services through intelligent resource-aware offloading, *IEEE Transactions on Network and Service Management*, vol. 16, no. 4, pp. 1665–1680, 2019.

[87] Z. P. Gao, J. Meng, Q. Wang, and Y. Yang, Service migration for deadline-varying user-generated data in mobile edge-clouds, presented at 2018 IEEE World Congress on Services (SERVICES), San Francisco, CA, USA, 2018, pp. 51–52.

[88] Z. H. Wang, R. T. Gu, G. Zhang, T. Y. Zhao, Y. N. Wang, Y. Wang, and Y. F. Ji, Demonstration of network slicing in mobile edge computing service migration, presented at 2018 Asia Communications and Photonics Conf. (ACP), Hangzhou, China, 2018, pp. 1–3.

[89] J. Zhou, F. Wu, K. Zhang, Y. M. Mao, and S. P. Leng, Joint optimization of offloading and resource allocation in vehicular networks with mobile edge computing, presented at 2018 10th Int. Conf. Wireless Communications and Signal Processing (WCSP), Hangzhou, China, 2018, pp. 1–6.

[90] T. X. Tran and D. Pompili, Joint task offloading and resource allocation for multi-server mobile-edge computing networks, *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 856–868, 2019.

[91] X. L. Xu, Q. H. Huang, X. C. Yin, M. Abbasi, M. R. Khosravi, and L. Y. Qi, Intelligent offloading for collaborative smart city services in edge computing, *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 7919–7927, 2020.

[92] L. Qingqing, J. Peñia Queralta, T. N. Gia, H. Tenhunen, Z. Zou, and T. Westerlund, Visual odometry offloading in internet of vehicles with compression at the edge of the network, presented at 2019 12th Int. Conf. Mobile Computing and Ubiquitous Network (ICMU), Kathmandu, Nepal, 2019, pp. 1–2.

[93] Z. Deng, Z. Cai, and M. Liang, A multi-hop VANETs-assisted offloading strategy in vehicular mobile edge computing, *IEEE Access*, doi: 10.1109/ACCESS.2020.2981501.

[94] R. J. Tan, J. Zhou, H. B. Du, S. C. Shang, and L. Dai, An modeling processing method for video games based on deep reinforcement learning, presented at 2019 IEEE 8th Joint Int. Information Technology and Artificial Intelligence Conf. (ITAIC), Chongqing, China, 2019, pp. 939–942.

[95] X. Chen, L. Jiao, W. Z. Li, and X. M. Fu, Efficient multi-user computation offloading for mobile-edge cloud computing, *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, 2016.

[96] Q. Qi, J. Y. Wang, Z. Y. Ma, H. F. Sun, Y. F. Cao, L. X. Zhang, and J. X. Liao, Knowledge-driven service offloading decision for vehicular edge computing: A deep reinforcement learning approach, *IEEE Transactions on*

Vehicular Technology, vol. 68, no. 5, pp. 4192–4203, 2019.

[97] Z. L. Ning, P. R. Dong, X. J. Wang, J. J. P. C. Rodrigues, and F. Xia, Deep reinforcement learning for vehicular edge computing: An intelligent offloading system, *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 6, p. 60, 2019.

[98] E. Nishani and B. Çiço, Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation, presented at 2017 6th Mediterranean Conf. Embedded Computing (MECO), Bar, Montenegro, 2017, pp. 1–4.

[99] J. L. Li, G. Y. Luo, N. Cheng, Q. Yuan, Z. H. Wu, S. Gao, and Z. H. Liu, An end-to-end load balancer based on deep learning for vehicular network traffic control, *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 953–966, 2019.

[100] Z. L. Ning, X. J. Wang, and J. Huang, Mobile edge computing-enabled 5G vehicular networks: Toward the integration of communication and computing, *IEEE Vehicular Technology Magazine*, vol. 14, no. 1, pp. 54–61, 2019.

[101] Z. P. Gao, Q. D. Jiao, K. L. Xiao, Q. Wang, Z. J. Mo, and Y. Yang, Deep reinforcement learning based service migration strategy for edge computing, presented at 2019 IEEE Int. Conf. Service-Oriented System Engineering (SOSE), San Francisco, CA, USA, 2019, pp. 116–121.

[102] Y. X. Guo, Z. L. Ning, and R. Kwok, Deep reinforcement learning based traffic offloading scheme for vehicular networks, presented at 2019 IEEE 5th Int. Conf. Computer and Communications (ICCC), Chengdu, China, 2019, pp. 81–85.

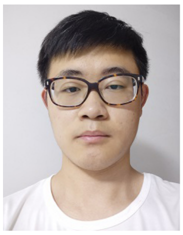
[103] J. A. Onieva, R. Rios, R. Roman, and J. Lopez, Edge-assisted vehicular networks security, *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8038–8045, 2019.

[104] H. J. Ji, O. Alfarraj, and A. Tolba, Artificial intelligence-empowered edge of vehicles: Architecture, enabling technologies, and applications, *IEEE Access*, vol. 8, pp. 61 020–61 034, 2020.

Xiaolong Xu received the PhD degree from Nanjing University, China in 2016. He worked as a research scholar at Michigan State University, USA, from April 2017 to May 2018. He is currently an assistant professor at the School of Computer and Software, Nanjing University of Information Science and Technology. He has published more than 80 peer review papers in the international journals and conferences, including *IEEE TITS*, *ACM TOIT*, *ACM TOMM*, *IEEE TII*, *IEEE TCC*, *IEEE TBD*, *IEEE IOT*, *IEEE TETCI*, *IEEE TCSS*, *JNCA*, *SPE*, *WWWj*, *IEEE ICWS*, *ICSOC*, etc. His research interests include fog computing, edge computing, IoT, cloud computing, and big data.

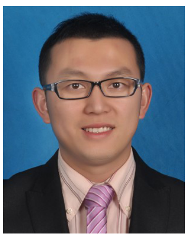
Haoyuan Li is currently studying for the BS degree in software engineering at Nanjing University of Information Science and Technology, Nanjing, China. His areas of interest are edge computing, cloud computing, and machine learning.

Weijie Xu received the BS degree in computer science and technology engineering from Nanjing University of Information Science and Technology in 2020. He is currently studying for the master degree in software engineering at Nanjing University of Information Science and Technology. His areas of interest are edge computing, big data, and cloud computing.



Zhongjian Liu received the BS degree in information and computational science from Shandong University of Science and Technology in 2020. Now, he is pursuing the master degree in computer science and technology at Nanjing University of Information Science and Technology. His research interests include edge computing, cloud computing, and big data.

Liang Yao received the BS degree in software engineering from Nanjing University of Information Science and Technology in 2020. He is currently pursuing the master degree in software engineering at Nanjing University of Information Science and Technology. His areas of interest are big data, cloud computing, and edge computing.



Fei Dai received the BE and PhD degrees from Yunnan University in 2008 and 2011, respectively. He is currently a professor at the School of Big Data and Intelligence Engineering, Southwest Forestry University, Kunming, China. He has authored and co-authored more than 50 scientific articles. Over the past years, his research interests include machine learning, business process management, and software engineering.