Two-Layer Federated Learning With Heterogeneous Model Aggregation for 6G Supported Internet of Vehicles

Xiaokang Zhou, *Member, IEEE*, Wei Liang, *Member, IEEE*, Jinhua She, *Fellow, IEEE*, Zheng Yan, *Senior Member, IEEE*, and Kevin I-Kai Wang, *Member, IEEE*

***Abstract*—The vision of the upcoming 6G technologies that have fast data rate, low latency, and ultra-dense network, draws great attentions to the Internet of Vehicles (IoV) and Vehicle-to-Everything (V2X) communication for intelligent transportation systems. There is an urgent need for distributed machine learning techniques that can take advantages of massive interconnected networks with explosive amount of heterogeneous data generated at the network edge. In this study, a two-layer federated learning model is proposed to take advantages of the distributed end-edge-cloud architecture typical in 6G environment, and to achieve a more efficient and more accurate learning while ensuring data privacy protection and reducing communication overheads. A novel multi-layer heterogeneous model selection and aggregation scheme is designed as a part of the federated learning process to better utilize the local and global contexts of individual vehicles and road side units (RSUs) in 6G supported vehicular networks. This context-aware distributed learning mechanism is then developed and applied to address intelligent object detection, which is one of the most critical challenges in modern intelligent transportation systems with autonomous vehicles. Evaluation results showed that the proposed method, which demonstrates a higher learning accuracy with better precision, recall and F1 score, outperforms other state-of-the-art methods under 6G network configuration by achieving faster convergence, and scales better with larger numbers of RSUs involved in the learning process.**

Manuscript received August 1, 2020; revised January 13, 2021 and March 27, 2021; accepted April 5, 2021. Date of publication May 6, 2021; date of current version July 8, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 72088101, 72091515, and 62072171, in part by the National Key R&D Program of China under Grants 2017YFE0117500, 2019YFE0190500 and 2019GK1010, and in part by the Natural Science Foundation of Hunan Province of China under Grant 2019JJ40150. The review of this article was coordinated by the Guest Editors of the Special Section on Vehicular Networks in the Era of 6G: End-Edge-Cloud Orchestrated Intelligence. *(Corresponding author W. Liang.)*

Xiaokang Zhou is with the Faculty of Data Science, Shiga University, Hikone 522-8522, Japan, and also with the RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan (e-mail: zhou@biwako.shiga-u.ac.jp).

Wei Liang is with the Base of International Science and Technology Innovation and Cooperation on Big Data Technology and Management, Hunan University of Technology and Business, Changsha 410205, China (e-mail: weiliang@csu.edu.cn).

Jinhua She is with the School of Engineering, Tokyo University of Technology, Tokyo 192-0982, Japan (e-mail: she@stf.teu.ac.jp).

Zheng Yan is with the State Key Laboratory on Integrated Services Networks and the School of Cyber Engineering, Xidian University, Xi'an 710071, China and also with the Department of Communications and Networking, Aalto University, 02150 Espoo, Finland (e-mail: zyan@xidian.edu.cn).

Kevin I-Kai Wang is with the Department of Electrical, Computer and Software Engineering, The University of Auckland, Auckland 1010, New Zealand (e-mail: kevin.wang@auckland.ac.nz).

Digital Object Identifier 10.1109/TVT.2021.3077893

***Index Terms*—Federated learning, End-edge-cloud computing, Internet of vehicles, Heterogeneous data, 6G technology.**

I. INTRODUCTION

SEAMLESS and ubiquitous communication infrastructure has been a key enabler of modern intelligent services. While some existing 5G technologies are able to provide a maximum data rate of up to 20 GB per second, such performance may still not meet the requirement of vehicular network applications of less than 1 ms latency [1], [2], especially in end-edge-cloud environments [3]. One of the main objectives and vision of beyond 5G (B5G) or 6G networks is to support future intelligent applications and services, such as autonomous vehicles, with ultra-low latency and high reliability [4]. Those applications are expected to be composed of large number of end user equipment (i.e., large scale, high density networks) that generates large quantities of heterogeneous and potentially sensitive data in real-time [5]. It is of critical importance to leverage the communication capabilities offered by 6G networks and the distributed data generated by user equipment in offering intelligent services.

The emergence of 6G technology brings unique advantages to vehicular network applications. The large-scale and mobile ad-hoc nature, with high data volume, stringent reliability and security requirement, makes vehicular networks a perfect use case for 6G technology [6]. A typical high-density vehicular network consists of end (e.g., vehicles), edge (e.g., road side units (RSUs)), and cloud components which constantly communicate and exchange information with one another, providing intelligent and adaptive services [7], [8]. Many smart applications and services can rise from vehicular networks [9]. For example, based on driver assistance (or autonomous driving) with local information sharing and dissemination, road side information, such as temporary traffic sign, adaptive speed signs, parking information, or petrol station information, can be recognized, searched, and shared with drivers in a city. Such services may enhance safer driving behavior (automatic and adaptive cruise control), more efficient traffic management (intelligent intersection control), and more intelligent route planning and navigation.

Considering a vehicular network application scenario such as local information sharing and dissemination, visual or lidar data can be generated from individual vehicles, where local training and model generation can be performed to achieve object detection and sign recognition. With small quantity of data (or low data quality depending on the camera resolution) and varying computational power of individual vehicles, local training is usually limited by its accuracy. In addition, video data may contain sensitive information. Uploading such data to a central location for centralized training and model generation, not only waste communication bandwidth, but also may leak potentially sensitive information coming from one's camera [10]. Such practice is also very likely to be prevented by data protection regulations such as GDPR (General Data Protection Regulation). In this scenario, the challenges are threefold: 1) intelligent applications and services such as navigation and object avoidance need higher learning accuracy that requires efficient information sharing among multiple vehicles; 2) information shared across multiple vehicles should preserve data privacy; 3) learning latency needs to satisfy the application requirement.

In this study, we propose a two-layer federated learning model based on convolutional neural network (TFL-CNN), which makes use of the local and global contexts of individual vehicles and RSUs to perform hierarchical and heterogeneous model selection and aggregation at the edge and cloud level. According to the aforementioned application scenario, individual vehicles with onboard camera can collect real-time data, which allows, for example, object detection or sign recognition. With the limited data, locally trained model is expected to be less accurate unless data from multiple vehicles are aggregated at a higher level for developing a global model. Benefit from federated learning, individual vehicles will perform training locally and share the trained parameters instead of raw data with the connected RSUs. Therefore, data privacy is protected for individual users. Depending on the local context information between an RSU and individual vehicles, for example distance between an RSU and vehicle, the quality of the corresponding model parameters can be estimated for the purpose of local model selection and aggregation. Once the local model is aggregated, multiple RSUs will upload the aggregated parameters to the cloud where the global model selection and aggregation will happen. Different to the RSU layer, the cloud infers the model quality based on the location of a specific RSU (i.e., whether it is at a busy junction with large amount of data) and its computational power (i.e., whether it is capable of processing a large quantity of data), which are known during system setup time. The cloud layer will perform a global model aggregation based on the selected RSU models. The multi-layer heterogeneous model selection and aggregation enable achieving more superior learning accuracy, and at the same time reducing the learning latency. Specifically, the contributions of this paper are summarized as follows.

i) A two-layer federated learning architecture is newly designed, targeting 6G supported vehicular networks with end-edge-cloud components, which preserves data privacy (i.e., no sharing of raw data) while achieving higher learning accuracy.
ii) A novel multi-layer heterogeneous model selection and aggregation scheme is developed to make use of the local (RSU) and global (cloud) context information, based on 6G supported vehicular networks.
iii) A context-aware learning mechanism is developed for intelligent object detection, targeting lower learning latency and better training efficiency in 6G supported Internet of Vehicles (IoV) environments.

The rest of this article is organized as follows. Section II addresses an overview of related works. In Section III, we present the fundamental framework of the two-layer federated learning in 6G support vehicular networks. The detailed CNN-based federated learning model, heterogeneous model selection and aggregation scheme, and context-aware learning mechanism for object detection, are introduced in Section IV. Experiment and evaluation results are discussed in Section V, followed by the conclusions and future research perspectives in Section VI.

II. Related Work

The emergence of advanced embedded, communication, and sensory technologies brought forward the possibility of intelligent vehicular network applications such as autonomous vehicles, cooperative driving, and intelligent traffic management. The safety-critical and data-rich nature makes vehicular network applications hot research topics. In this section, targeted vehicular network applications are summarized and introduced, along with the state-of-the-art distributed artificial intelligence (AI) approaches developed for these applications.

A. Vehicular Network Applications

With the proliferation of advanced technologies, one of the main goals for intelligent vehicular applications is to provide driver assistance and to enable safer driving behavior [11]. For example, displaying warning signs on user dashboard which proactively provide surrounding information beyond driver's visual sight could offer enormous help to avoid dangerous driving or accidents caused by unknown conditions [12, 13]. Other techniques based on peer-to-peer communications are also developed targeting collision avoidance and cooperative driving to enhance safety [14]–[16]. In addition to vehicle-to-vehicle (V2V) communication, similar applications using vehicle-to-infrastructure (V2I) communication (e.g., RSU or traffic signs), are also available to support intelligent applications such as intersection management and adaptive cruise control [6], [17], [18]. The wide use of communication networks for sharing user and traffic information also brings the research attention on information integrity, security and privacy [19], [20]. Alghamdi et al. [21] investigated the large intelligent surfaces technologies for 6G wireless platforms. They discussed their working principles and performance analysis frameworks to demonstrate their impact and positions in 6G enabled wireless network applications from the technical aspects. Envisaging 6G as a massively connected complex network, which might give feedback to users' service calls based on the rapid learning of network states in end-edge-cloud structures, Nawaz et al. [22] reviewed the emerging computing paradigms including machine learning, quantum computing, etc., for 6G based communication networks, and introduced their applications with some case studies in B5G networks. To figure out a roadmap for the future vision of intelligent services supported by 6G networks, Bariah et al. [23] compared enabling technologies of 5G and 6G, and summarized 6G empowered network services with their potential applications in terms of their main challenges and open research issues. Yuan et al. [24] presented estimation framework for interest variables in intelligent vehicular networks, in which the factor graph was employed with consensus algorithms to realize estimations in each vehicle and facilitate local processing and communications at RSUs.

B. Federated Learning With Vehicular Networks

Vehicular Internet of Things (IoT) or IoV are well-known for their 3V characteristics, i.e., data volume, variety, and velocity. Typically, a vehicle can generate up to 30 TB of data each day, and the vision of 6G supported vehicular networks are likely to promote these characteristics [25]. It is of critical importance for applications to make use of these distributed data (i.e., per vehicle) wisely and efficiently. Tang et al. [26] surveyed and classified several machine learning techniques applied in vehicular networks for networking, communication, and security issues, to meet requirements and challenges in next-generation (6G) networks. Traditional centralized AI approaches, such as deep neural networks, require distributed data to be transferred and located at a central location (i.e., RSU or cloud) to enable model training and generation. Such approaches have several disadvantages such as high network bandwidth usage, high learning latency, and potential threat of leaking sensitive user information. In contrast, federated learning, as a distributed AI approach, allows local models to be trained on individual vehicles and then aggregated in the cloud (or RSU) to enhance learning accuracy, communication efficiency and privacy preservation [27]. With these benefits, federated learning has been considered as a promising approach and recently applied to various different IoV applications, which might be developed targeting the challenge of object detection in vehicular network applications. Du et al. [28] discussed the advances when applying federated learning in vehicular IoT systems, and pointed out several technical challenges and the necessity in integrating vehicular IoT with federated learning framework. Ye et al. [29] proposed an aggregation approach with a selective model to pursue the better model accuracy and aggregation efficiency for 3D object detection in vehicular edge computing. They employed a two-dimension contract theory and built a distributed framework to improve the interaction between the server and vehicular clients. To achieving data integrity and privacy purposes, Lu et al. [30] introduced an intelligent architecture with a federated learning mechanism, in which a two-phase mitigating scheme was used to improve the smart data transformation and collaborative data leakage detection. Brik et al. [31] focused on federated deep learning applications in unmanned aerial vehicle related wireless networks. Following a series of use cases of federated learning strategies in 5G or B5G network environments, they summarized the key challenges with open issues for the future research direction of application of federated learning in vehicular networks. He et al. [32] designed a federated edge learning system, in which a so-called importance-aware joint data selection and resource allocation algorithm was developed to improve the learning speed, and further solve the learning efficiency maximization problem in mobile computing. Yu et al. [33] utilized the content caching scheme for vehicular application development in edge networks. They applied the federated learning technique to build a global model to predict the content popularity while protecting the privacy of training data in local vehicles.

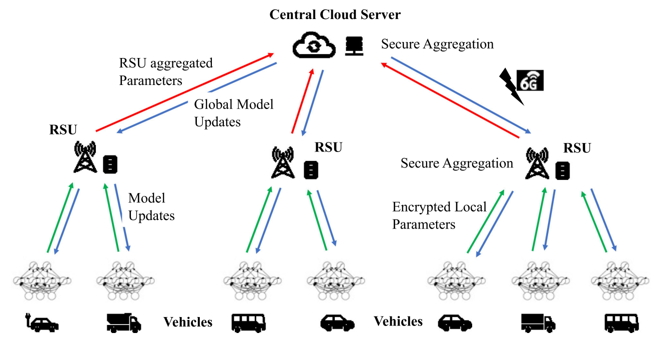


Fig. 1. Two-Layer Federated Learning Framework in 6G Supported Vehicular Networks.

III. Two-Layer Federated Learning for 6G Supported Vehicular Networks

A. Basic Model Architecture

In this study, we aim to explore typical object detection applications in 6G supported vehicular network scenarios, such as traffic sign recognition, object avoidance, and pedestrian detection. The proposed framework architecture of the two-layer federated learning in 6G supported vehicular networks is illustrated in Fig. 1. It should be noticed that one of the main characteristics of 6G is the use of THz communication frequency range, which leads to the short-range communication (e.g., tens of meters coverage) and ultra-high-density network [34], [35]. Specifically, 6G will be designed to support large-scale mMTC (massive Machine Type Communication) via enabling technologies such as NOMA (Non-Orthogonal Multiple Access) [36] and cell-free (or cell-less) communication architecture [37]. With these enabling technologies, individual vehicles will connect to the entire 6G supported vehicular network as a whole without the burden differentiating individual cell connection (or in this case the RSUs as relaying nodes). These technologies will also greatly reduce the overhead of handover. In addition, the ultra-low latency characteristic of 6G (1ms, end-to-end) will allow individual vehicles to complete their learning cycle (i.e., reaching the preconfigured error rate via one physical RSU). Therefore, every individual learning cycle will be considered as a static snapshot of the vehicular network at a specific time instance.

As shown in Fig. 1, a typical 6G supported vehicular network is composed of one central cloud server (i.e., a macro base station (MBS)), a few RSUs, and a large number of vehicles. A two-layer federated learning framework incorporating end-edge-cloud computing scheme for real-time object detection is investigated in this urban vehicular network, which is introduced in detail as follows.

In the top layer, the central cloud server offers capabilities of high-performance computing and higher-level knowledge sharing, which enable it to conduct a global caching and aggregation computing task. In the middle layer, each RSU has limited caching and computing capabilities, and is responsible for supervising all the interconnected vehicles in its coverage. Empowered by the 6G technologies, RSUs are the middle brokers to collect and aggregate not only learning parameters but also contextual information, such as vehicle locations and navigation direction, from all the connected vehicles within the coverage to facilitate parameter aggregation. RSUs then communicate with the central cloud server to upload aggregated model parameters and its own contextual information [37]. In return, the updated model parameters will be dispatched from the central cloud server to RSUs and then to individual vehicles.

In the bottom layer, each vehicle generates raw data including both the captured photos/videos by built-in camera and the contextual information (e.g., GPS locality data, driving information, etc.). The computation capability of individual vehicle is able to support a relatively light computing task (e.g., training a learning model for object detection or road sign recognition). All the vehicles share a unified deep learning model. Moreover, in a 6G supported vehicular network, more vehicles will be connected to an RSU and share the same network resource. It is critical to ensure privacy protection and efficient knowledge sharing within the high-speed, high-density 6G network. The proposed two-layer federated learning framework only requires transmission of model parameters, which ensures both data privacy and efficient knowledge sharing.

Datasets generated by individual vehicles (i.e., data owners) share identical features with different samples. For example, two vehicles in different regions acquire their own piece of data respectively, which can be represented by a common set of features. The intersection of samples collected from different vehicles (e.g., captured photos of road signs, cars, pedestrians) may be very small since two vehicles in different regions can seldom capture photos of an identical object. However, the purpose of the learning model and targeted intelligent services (e.g., for various applications of object detection) in these vehicles are likely to be very similar. Therefore, the basic model and the feature space of these datasets can be consolidated to a unique framework.

B. Problem Definition

Assuming a typical vehicular network is composed of n vehicles, m RSUs and a central cloud server. Those vehicles are a set of data owners $V = \{v_i\}_{i=1}^{n}$, which generate real-time photos or videos as raw data by themselves and can train a deep learning model from their respective data $D = \{d_i\}_{i=1}^{n}$. Each local dataset d_i consists of raw data x_i and the corresponding label y_i as $d_i = (x_i, y_i)$. Conventional supervised learning process is to consolidate all the data together from the vehicles $D_{total} = d_1 \cup d_2 \cup \ldots \cup d_n$ in a central cloud server. For the purpose of data privacy protection and network traffic alleviation, a horizontal federated learning framework is proposed for real-time object detection across vehicles. Empowered by 6G technologies, any individual data owner v_i in this framework can keep its own data d_i and train the object detection model locally rather than transmitting data d_i to an RSU and other owners. In addition, a set of RSUs is denoted as $R = \{r_i\}_{i=1}^{m}$ and the central cloud server is denoted as CS.

Accordingly, a global deep learning model $M = h(x, \omega)$ is trained on the distributed dataset D_{total} across the 6G supported vehicular network. The goal of the proposed framework is to recognize the objects or road signs during the real-time autonomous driving scenario based on samples in D_{total}. As a typical horizontal federated learning framework, we assume data owners are secure against a curious RSUs, which means an RSU cannot compromise the privacy of data owners [38]. To achieve this, parameters computed via local training by v_i need to be encrypted and then sent to the corresponding RSU which conducts secure aggregation without exposing privacy information about any data owners.

IV. Integrated Learning Mechanism for Intelligent Object Detection in End-Edge-Cloud Environments

A. CNN Based Two-Layer Federated Learning

Considering the aforementioned object detection applications in a 6G supported vehicular network, a horizontal TFL-CNN is designed to tackle these applications, while ensuring data privacy and communication efficiency. The detailed structure with the concrete workflow of our model is illustrated in Fig. 2.

Assuming x_i is the input samples of a data owner v_i, the CNN model, which is introduced to perform an object detection task (e.g., traffic sign recognition, pedestrian detection, or object avoidance), can be represented as the hypothesis $h(x_i, \omega)$, and trained locally by the data owner v_i.

$$h(x_i, \omega) = \mathrm{FC}\left(\mathrm{Pool}\left(\mathrm{Conv}\left(x_i, \omega_{Conv}\right), \omega_{Pool}\right), \omega_{FC}\right) \quad (1)$$

where $\mathrm{Conv}(*)$, $\mathrm{Pool}(*)$, and $\mathrm{FC}(*)$ stand for the convolution layer, pooling layer, and the fully connected layer respectively. ω_* is the parameter of the corresponding layer in the model.

The result of $h(x_i, \omega)$ is predicted by a SoftMax classifier as follows.

$$y^{predict} = \mathrm{SoftMax}\left(h(x_i, \omega)\right) \quad (2)$$

Furthermore, to achieve the goal of the proposed model, we define the cost function as follows.

$$J_i(\omega) = -\frac{1}{k}\left[\sum_{q=1}^{k}\sum_{p=1}^{l}\{y_i = p\}\log\left(\frac{e^{\omega_p^T x_i}}{\sum_{q=1}^{k} e^{\omega_q^T x_i}}\right)\right] \quad (3)$$

where we assume there are k samples for data owner v_i.

In each iteration, the local parameters acquired by individual data owners will be encrypted and uploaded to the corresponding RSUs. It should be noticed that only encrypted parameters are uploaded instead of raw data, which ensures data privacy protection and greatly reduces the communication overhead required during the learning process.

Besides, the vehicular contextual information, such as location and navigation information, can be directly acquired by RSUs via 6G technologies [37] to conduct the proposed weighted aggregation in RSU. Similarly, the aggregated parameters and the corresponding RSU contextual information will be encrypted and sent to the central cloud server for global aggregation.

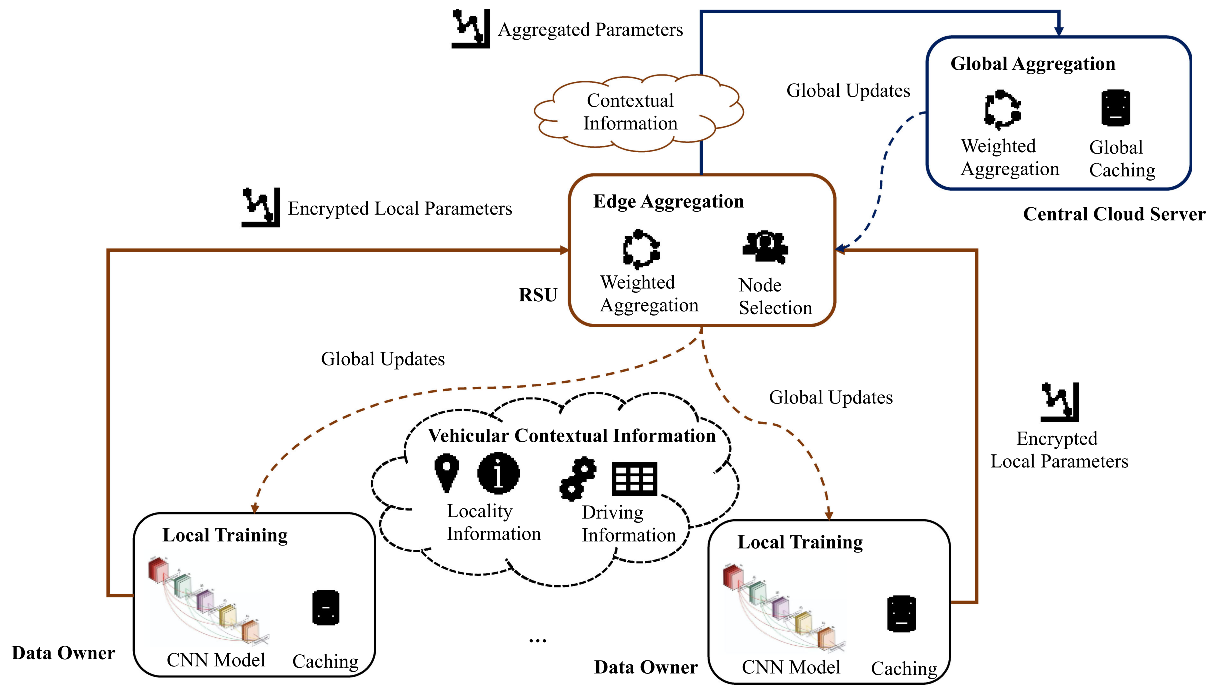


Fig. 2. Architecture of TFL-CNN.

B. Multi-Layer Heterogeneous Selection and Aggregation Scheme

To enhance the efficiency and the accuracy of the global federated learning process, both the vehicular and RSU contextual information are taken into account in the aggregation process.

The local parameters gained by data owner v_i at iteration $t \in T$ is $\omega_i(t)$. The total parameter $\omega_j^{RSU}(t)$ by an RSU r_j is aggregated by Eq. (4) as follows.

$$\omega_j^{RSU}(t) = \frac{1}{J}\sum_{i=1}^{J} \delta_i * \omega_i(t) \tag{4}$$

where J is the total number of the vehicles supervised by an RSU r_j in a specific region. δ_i is the weighted coefficient of vehicle v_i, which is applied to measure how the local training result can be influenced by its locality and driving contextual information.

For simplicity, a distance-based measurement between an RSU r_j and vehicle v_i is defined to represent the locality attribute of v_i since the amount and quality of data from a vehicle near an RSU may be more valuable to the object detection. The distance is calculated by the longitude and latitude as follows.

$$dis(v_i, r_j) = \sqrt{(lat_i - lat_j)^2 + (lot_i - lot_j)^2} \tag{5}$$

In addition, the quality of captured raw data by vehicles may be influenced by the direction and velocity of the moving vehicles. Thus, the driving attribute can be simplified based on Eq. (6).

$$drv\ (v_i) = \frac{\theta_i}{90} * \log(vlc_i) \tag{6}$$

where $\theta_i \in [090]$ represents the direction angle of the moving vehicle, and vlc_i is the instant velocity when capturing the raw data.

According to Eq. (5) and Eq. (6), the weighted coefficient δ_i can be computed by the vehicular contextual information as follows.

$$\delta_i = \frac{dis(v_i, r_j)}{\sum_{j=1}^{J} dis(v_i, r_j)} + drv(v_i) \tag{7}$$

Once the local models are aggregated by individual RSUs, multiple RSUs will upload their aggregated parameters to the central cloud server CS where the global model selection and aggregation will happen. Different to the RSU layer, the cloud infers the model quality based on the computational power of a specific RSU, and whether it is at a busy junction with large amount of data, which are determined at system setup time and in real-time respectively. The throughput measurement for an RSU r_j is defined by the number of received local training results from the uploading data owners. The computational power can be measured by its hardware configuration (e.g., CPU/GPU configuration) for simplicity. Thus, we can deduce the weighted coefficient ξ_j for an RSU r_j as follows.

$$\xi_j = \frac{J}{n} * \log\left(\frac{\mathrm{P}(r_j)}{\mathrm{P}(CS)}\right) \tag{8}$$

where $\mathrm{P}(*)$ represent the computational power for the RSU or cloud server. Each vehicle will generate one set of parameters, thus J indicates the received number of sets of uploading local parameters by r_j. n stands for the whole vehicular network capacity.

C. TFL-CNN Based Intelligent Object Detection Algorithm

According to the discussions above, different aggregation strategies have been developed for the edge and cloud respectively. Based on the applied horizontal federated learning framework, we can share an identical global model and update

```
Input: A set of raw data provided by data owners D = {d_i}_{i=1}^n
       Participated data owners V = {v_i}_{i=1}^n
       Participated RSUs R = {r_j}_{j=1}^m
Output: A trained global object detection model M

Initialize model parameter ω, maximum iterations T
for t = 1 to T do
  for each RSU r_j ∈ R do
    for each data owner v_i ∈ V do
      if v_i supervised by r_j do
        Conduct local training on dataset d_i
        Calculate the contextual information for v_i by Eq.
        (5), Eq. (6)
        Submit the local training parameter ω_i(t) to r_j
    end for
    Aggregate the parameters in r_j by Eq. (4)-(7)
    Upload the parameters to server CS
  end for
  Calculate the global parameters ω(t + 1) for model M by
    Eq. (9)
  if ∇(ω) reach the convergency threshold goto 17
  Broadcast ω(t + 1) to the network
end for
return M
```

Fig. 3. Intelligent Object Detection Based on TFL-CNN.

the training parameters within the whole 6G supported vehicular network. Assuming there are total T iterations required during the global learning process, we define a gradient descent calculation for parameters in each iteration as follows.

$$\omega\ (t+1) = \ \omega\,(t) + \nabla\,(\omega) \tag{9}$$

where ω indicates the aggregated parameter in the global model. $\nabla(\omega)$ stands for the gradient descent calculated by the central cloud server CS.

The concrete algorithm based on the two-layer federated learning for object detection is illustrated in Fig. 3.

As described in the algorithm, the training process via the TFL-CNN is divided into two steps: aggregation in RSUs and aggregation in the cloud server. Both the vehicular contextual information and RSU contextual information are considered in these aggregations. The global model will be trained and distributed without exposing any private data to the network. It is noted that the proposed TFL-CNN is designed following the consideration of 6G supported vehicular network architecture (as exemplified in Fig. 1), and the high-density and low-latency communication characteristics envisioned by 6G technologies as well. Thus, it ensures data privacy and efficient communication during the whole learning process, which may tackle the key challenges faced by emerging 6G technologies [34], [35].

V. Experiment And Analysis

In particular, traffic sign recognition is used as a case study in the end-edge-cloud computing scenario, to evaluate the proposed TFL-CNN in 6G supported vehicular networks. Experiments are conducted and discussed to demonstrate the usefulness and effectiveness of the proposed method comparing with several baseline methods.

TABLE I.
PARAMETER CONFIGURATION OF 5G AND 6G

Issue	5G	6G
Per device peak data rate	10Gbps	1Tbps
E2E latency	10ms	1ms
Maximum spectral efficiency	30bps/Hz	100bps/Hz
Mobility support	Up to 500km/hr	Up to 1000km/hr
Maximum frequency	90GHz	10THz

A. Dataset

To investigate the effectiveness of the proposed TFL-CNN in 6G supported vehicular networks, a traffic sign image dataset named BelgiumTSC is considered. It includes more than 7000 images and provides a total of 11219 bounding boxes, which correspond to more than 2000 traffic sign images [39]. Both the labeled class and 3D location view were recorded for each sign. Although the videos and images are captured at less than 50 meters in at least one view, issues such as within-class variability, between-class similarity, and bad standardization may affect the performance of conventional methods. It needs be noticed that the traffic sign recognition task has the locality characteristic (i.e., 50 meters) similar to the 6G RSU coverage, where the task is initiated and needs to be completed within a short time period. In the following subsections, this is analyzed to give indications that why 6G is necessary for future intelligent network applications.

To effectively evaluate the detection performance for 6G supported vehicular networks, data pre-processing is necessary to refine the training and testing data to fit the 6G environment, which consists of three main steps in our experiment:

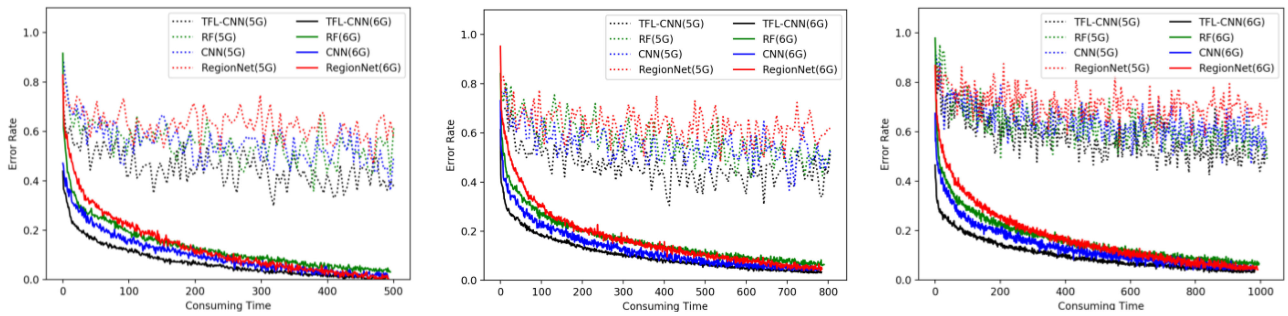
1) Remove the badly formatted and missing features records.
2) Enlarge the size of dataset to 10 times bigger than the original size to simulate the 6G supported IoV environment.
3) Split the whole enlarged dataset into 500 shards, and assign these shards to 25 RSUs for federated learning scenarios.

B. Experiment Design and Evaluation Metrics

We adopt several widely used and classical machine learning methods in this field as the baseline methods, including classical methods: Random Forest (RF) and CNN, and one state-of-the-art method: RegionNet [40], which is capable of achieving both clear detection boundary and multi-scale contextual robustness for traffic sign detection, to conduct the comparison evaluation. Furthermore, we refer to the existing 6G wireless communication requirements [35], to simulate the comparison between 6G and 5G network. The parameter configurations set for our experiments are listed in Table I. It can be observed that the E2E (End-to-End) latency of 6G supported network is 10 times smaller than 5G. We investigate the time consumption of all the methods by varying the E2E latency from 0.5ms to 5ms.

Evaluation indicators used in performance comparisons are introduced and defined as follows.

1) True positive (TP): A target traffic sign has been predicted to a correct type.



(a) 0.5ms E2E Latency (b) 1ms E2E Latency (c) 5ms E2E Latency

Fig. 4. Error Rate Observation Under 5G/6G Configuration in Federated Learning Process.

2) True negative (TN): A non-target traffic sign has been predicted to a correct type.
3) False positive (FP): A target traffic sign has been predicted to a wrong type.
4) False negative (FN): A non-target traffic sign has been predicted to a wrong type.

Following the above definition, three widely used metrics: Precision, Recall, and F1 score, can be calculated as follows.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = \frac{2{*}Precision * Recall}{Precision + Recall} \tag{12}$$

Precision indicates the model's ability to distinguish traffic signs. The higher the precision value is, the better the model's ability to distinguish the similar traffic signs will be. Recall reflects the ratio of the model to tackle the intra-class similarity issue. The higher the recall value is, the better the model's ability to detect the traffic signs with intra-class similarity will be. F1 score describes the balance between Precision and Recall. The higher the F1 score is, the more robust performance of the model can achieve.

C. Detection Performance Evaluation

We first evaluate the learning efficiency of the proposed method. The training process of the TFL-CNN is investigated by observing how the gradient shifts and loss declines during the aggregation process. The learning rate is set to 0.005 and the maximum iteration is set to 1000. Three different potential E2E latencies (i.e., 0.5ms, 1ms, and 5ms) of 6G supported networks are studied. The aggregated error rates for each method across a total of 25 RSUs during the training process are illustrated in Fig. 4.

As shown in Fig. 4, our proposed TFL-CNN outperforms all the other three methods in achieving faster convergence to reach a reasonable error rate at 0.1. Furthermore, it can be clearly observed that with lower latency, all the methods can accomplish faster convergence in terms of error rate for the traffic sign recognition task in 6G. Whereas in 5G, the four methods could not converge to a reasonable error rate within the similar period of time. For vehicular networks or IoV applications, where real time response is a critical factor, 5G is obviously not sufficient. In addition, it is noticed that with higher latency (i.e., from 0.5 ms to 5 ms), these methods may take longer time to reach convergence. This result can be viewed as a good indication of the performance of the proposed method under the expected 6G configuration.

TABLE II.
COMPARISONS ON TRAFFIC SIGN RECOGNITION PERFORMANCE

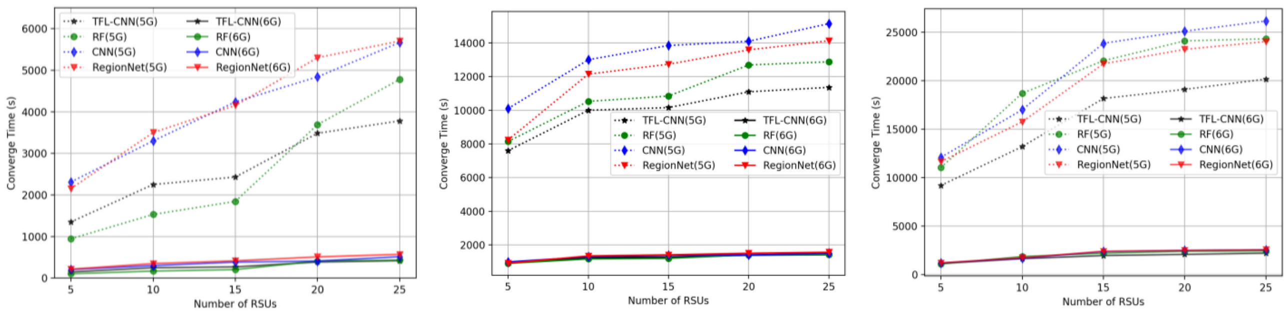
Methods	Precision	Recall	F1 score
CNN	0.889	0.821	0.854
RegionNet	0.899	0.910	0.904
RF	0.746	0.852	0.795
TFL-CNN	**0.906**	**0.968**	**0.936**

We then analyze the impact of different numbers of RSUs involved in the training process. As shown in Fig. 5, generally, all the four methods can reach the preconfigured threshold of error rate within a much shorter time period in 6G. However, when the number of RSUs increases, a significant increase in convergence time can be observed in 5G configuration. This clearly indicates that 5G configuration is still not sufficient to support large-scale IoV applications. It is obvious that the convergence time in all methods are increasing relatively linearly as the number of involved RSU increases, and the proposed TFL-CNN outperforms all the other methods with a faster convergence. Besides, it is interesting to note that the RF method requires less training data and could converge slightly faster. However, its scalability is not very good compared with our TFL-CNN, especially when dealing with larger number of RSUs in the possible future 6G environment.

We go further to investigate the four methods in traffic sign recognition using the enlarged dataset based on metrics Precision, Recall, and F1 score. Particularly, to demonstrate the strength of the proposed federated learning method in the end-edge-cloud computing scenario, one shard of dataset is used for CNN and RF, and the whole dataset is used for RegionNet and our TFL-CNN, so as to simulate a realistic distributed learning in a 6G supported vehicular network. The results are listed and compared in Table II.

Based on Table II, obviously, the proposed TFL-CNN achieves the best results in F1 score at 0.936. This indicates that the proposed method can result in higher accuracy in the traffic sign recognition scenario compared with other baseline methods.

Finally, we evaluate the performances of all the methods in terms of detection accuracy using the whole dataset (500 shards).



(a) 0.5ms E2E Latency (b) 1ms E2E Latency (c) 5ms E2E Latency

Fig. 5. Total Convergence Time(s) Under 5G/6G Configuration to Meet the Threshold of Error Rate.

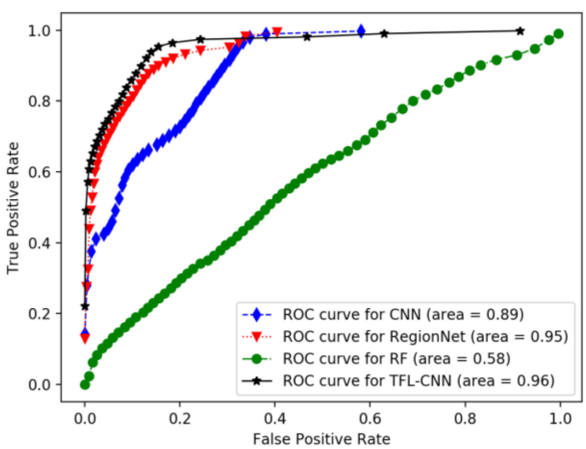


Fig. 6. Performance Evaluation Based on ROC.

The ROC (Receiver Operating Characteristic) curve is utilized to demonstrate the evaluation performance, the results of which are illustrated in Fig. 6.

As shown in Fig. 6, all the curves appeared in the upper left corner above the diagonal line, which indicates that all the methods can predict the traffic sign correctly to a certain extent. This can exemplify that the modern machine learning schemes are practical to facilitate autonomous driving. Apart from RF, the other three learning methods achieve a relatively good result (nearly 1.0) when FPR and TPR increase to 1.0 at the right upper corner in the figure. According to the ROC curves, the TFL-CNN achieves the best result at average 0.96. This indicates that the proposed two-layer federated learning model can perform efficient training with higher accuracy, so as to recognize the traffic sign effectively, comparing with the baseline methods in the 6G supported vehicular network scenario. The proposed TFL-CNN can efficiently identify the features of traffic signs and optimize the local training parameters in terms of the contextual information calculated from both the vehicles and RSUs based on the distributed dataset. It is noticed that the RegionNet method shows a result of ROC at 0.95 compared with the proposed TFL-CNN, which can be explained as the federated learning scheme can benefit a lot when addressing problems with distributed big dataset in a 6G supported IoV environment.

VI. Conclusion

In this paper, we proposed a two-layer federated learning model for intelligent object detection, which could achieve more efficient and more accurate learning result in 6G supported IoV environments.

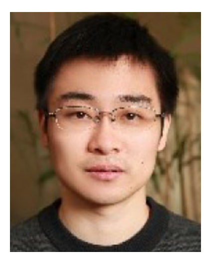
In particular, a two-layer federated learning framework was designed and built to enhance the typical end-edge-cloud computing architecture in 6G supported vehicular networks. An integrated TFL-CNN was constructed to train using the local data and only share the parameters. Both the local and global contexts of individual vehicles and RSUs were taken into account in a multi-layer heterogeneous model selection and aggregation scheme, which could effectively improve the training efficiency. A context-aware learning mechanism was then developed and applied in the intelligent object detection. Experiment and evaluation results demonstrated the outstanding performance of our proposed method in faster convergence and better learning accuracy for 6G supported IoV applications, comparing with other similar methods.

In the future studies, we will go further to study the federated learning architecture with more deep learning techniques, and conduct more evaluations to improve the algorithm with better accuracy and efficiency in different practical scenarios.

References

[1] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1146–1159, Feb. 2020.
[2] Z. Zhang, *et al.*, "6G Wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sept. 2019.
[3] J. Ren, D. Zhang, S. He, Y. Zhang, and T. Li. "A survey on end-edge-cloud orchestrated network computing paradigms: Transparent computing, mobile edge computing, fog computing, and cloudlet," *ACM Comput. Surv.*, vol. 52, no. 6, 2019, Art. no. 125.
[4] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
[5] F. Lyu, *et al.*, "LEAD: Large-scale edge cache deployment based on spatio-temporal wifi traffic statistics," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2020.2984261.
[6] H. Moustafa, and Y. Zhang, "*Vehicular Networks: Techniques, Standards, and Applications*," Boca Raton, FL, USA: CRC Press, 2009.
[7] M. Sivasakthi, and S. Suresh, "Research on vehicular ad hoc networks (VANETs): An overview," *J. Appl. Sci. Eng. Res.*, vol. 2, no. 1, pp. 23–27, 2013.

[8] W. Liang, Z. Li, H. Zhang, S. Wang, and R. Bie, "Vehicular ad hoc networks: Architectures, research issues, methodologies, challenges, and trends," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 8, 2015.
[9] Y. Zhang, X. Xiong, J. Ren, and L. Cai, "Efficient computing resource sharing for mobile edge-cloud computing networks," *IEEE/ACM Trans. Netw.*, vol. 28, no. 3, pp. 1227–1240, Jun. 2020.
[10] J. Liu, *et al.*, "Online multi-workflow scheduling under uncertain task execution time in IaaS clouds," *IEEE Trans. Cloud Comput.*, to be published, doi: 10.1109/TCC.2019.2906300.
[11] X. Mosquet, M. Andersen, and A. Arora, "A roadmap to safer driving through advanced driver assistance systems," *Auto Tech Rev.*, vol. 5, no. 7, pp. 20–25, 2016.
[12] C. Maihofer, and R. Eberhardt, "Time-stable geocast for ad hoc networks and its application with virtual warning signs," *Comput. Commun.*, vol. 27, no. 11, pp. 1065–1075, 2004.
[13] T. Nadeem, P. Shankar, and L. Iftode, "A comparative study of data dissemination models for VANETs,"in *Proc. 3rd Annu. Int. Conf. Mobile Ubiquitous Syst.: Netw. Serv.*, San Jose, CA, 2006 pp. 1–10.
[14] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 74–82, Jan. 2006.
[15] X. Mosquet, M. Andersen, and A. Arora, "A roadmap to safer driving through advanced driver assistance systems," *Auto Tech. Rev.*, vol. 5, no. 7, pp. 20–25, 2016.
[16] U. Z. A. Hamid, K. Pushkin, H. Zamzuri, D. Gueraiche, and M. A. A. Rahman, "Current collision mitigation technologies for advanced driver assistance systems–A survey," *PERINTIS eJournal*, vol. 6, no. 2, pp. 78–90, 2016.
[17] K. Dresner, and P. Stone, "Multiagent traffic management: An improved intersection control mechanism," in *Proc. 4th Int. Joint Conf. Auton. Agents Multiagent Syst.*, Utrecht, The Netherlands, 2005 pp. 471–477.
[18] H. Hartenstein, and K. Laberteaux, "VANET: Vehicular applications and inter-networking technologies," *Wiley Online Library*, 2010.
[19] F. Qu, Z. Wu, F.-Y. Wang, and W. Cho, "A security and privacy review of vanets," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 2985–2996, Dec. 2015.
[20] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4298–4311, April 2020.
[21] R. Alghamdi, *et al.*, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," *IEEE Access*, vol. 8, pp. 202795–202818, 2020.
[22] S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary, and M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-Art and vision for the future," *IEEE Access*, vol. 7, pp. 46317–46350, 2019.
[23] L. Bariah, *et al.*, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, 2020.
[24] W. Yuan, S. Li, L. Xiang, and D. W. K. Ng, "Distributed estimation framework for beyond 5G intelligent vehicular networks," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 190–214, 2020.
[25] M. S. H. Sassi and L. C. Fourati, "Investigation on deep learning methods for privacy and security challenges of cognitive ioV," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, pp. 714–720, 2020.
[26] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6G: Machine-learning approaches," in *Proc. IEEE*, vol. 108, no. 2, pp. 292–307, Feb. 2020.
[27] K. Tan, D. Bremner, J. L. Kernec, and M. Imran, "Federated machine learning in vehicular networks: A summary of recent applications," in *Proc. Int. Conf. U.K.-China Emerg. Technol. (UCET)*, Glasgow, United Kingdom, 2020.
[28] Z. Du, C. Wu, T. Yoshinaga, K. -L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular internet of things: Recent advances and open issues," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 45–61, 2020.
[29] D. Ye, R. Yu, M. Pan, and Z. Han, "Federated learning in vehicular edge computing: A selective model aggregation approach," *IEEE Access*, vol. 8, pp. 23920–23935, 2020.
[30] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Federated learning for data privacy preservation in vehicular cyber-physical systems," *IEEE Netw.*, vol. 34, no. 3, pp. 50–56, 2020.
[31] B. Brik, A. Ksentini, and M. Bouaziz, "Federated learning for UAVs-Enabled wireless networks: Use cases, challenges, and open problems," *IEEE Access*, vol. 8, pp. 53841–53849, 2020.
[32] Y. He, J. Ren, G. Yu, and J. Yuan, "Importance-aware data selection and resource allocation in federated edge learning system," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13593–13605, 2020.
[33] Z. Yu, J. Hu, G. Min, Z. Zhao, W. Miao, and M. S. Hossain, "Mobility-aware proactive edge caching for connected vehicles using federated learning," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi: 10.1109/TITS.2020.3017474.
[34] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G Wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, 2020.
[35] J. He, K. Yang, and H. -H. Chen, "6G Cellular networks and connected autonomous vehicles," *IEEE Netw.*, to be published, doi: 10.1109/MNET.011.2000541.
[36] L. Zhu, Z. Xiao, X. Xia, and D. O. Wu, "Millimeter-wave communications with non-orthogonal multiple access for B5G/6G," *IEEE Access*, vol. 7, pp. 116123–116132, 2019.
[37] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, 2020.
[38] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, 2019.
[39] R. Timofte, K. Zimmermann, and L. V. Gool. "Multi-view traffic sign detection, recognition, and 3D localization," *ACM J. Mach. Vis. Appl.*, vol. 25, no.3 pp. 633–647, 2014.
[40] X. Wang, H. Ma, X. Chen, and S. You, "Edge preserving and multi-scale contextual neural network for salient object detection," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 121–134, 2018.



Xiaokang Zhou (Member, IEEE) received the Ph.D. degree in human sciences from Waseda University, Tokyo, Japan, in 2014. He is currently an Associate Professor with the Faculty of Data Science, Shiga University, Hikone, Japan. From 2012 to 2015, he was a Research Associate with the Faculty of Human Sciences, Waseda University. Since 2017, he has also been a Visiting Researcher with the RIKEN Center for Advanced Intelligence Project, RIKEN, Wako, Japan. He is involved in interdisciplinary research works in the fields of computer science and engineering, information systems, and social and human informatics. His recent research interests include ubiquitous computing, big data, machine learning, behavior and cognitive informatics, cyber-physical-social systems, cyber intelligence, and cyber-enabled applications. He is a Member of the IEEE CS, and ACM, USA, IPSJ, and JSAI, Japan, and CCF, China.

Wei Liang (Member, IEEE) received the M.S. and Ph.D. degrees in computer science from Central South University, Changsha, China, in 2005 and 2016, respectively. From 2005 to 2012, he was with Microsoft, China, for soft engineering. From 2014 to 2015, he was an Exchange Researcher with the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Tokyo, Japan. He is currently with the Key Laboratory of Hunan Province for New Retail Virtual Reality Technology, Hunan University of Technology and Business, Changsha, China. He has authored or coauthored more than 20 papers at various conferences and journals, including FGCS, JCSS, and PUC. His research interests include information retrieval, data mining, and artificial intelligence.

Jinhua She (Fellow, IEEE) received the B.S. degree in engineering from Central South University, Changsha, China, in 1983, and the M.S. and Ph.D. degrees in engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1990 and 1993, respectively. In 1993, he joined the School of Engineering, Tokyo University of Technology, Tokyo, Japan, where he is currently a Professor with the Department of Mechanical Engineering. His research interests include the application of control theory, repetitive control, process control, Internet-based engineering education, and assistive robotics. He is a Member of the Society of Instrument and Control Engineers, the Institute of Electrical Engineers of Japan, the Japan Society of Mechanical Engineers, and the Asian Control Association. He was the recipient of the International Federation of Automatic Control Control Engineering Practice Paper Prize in 1999 jointly with M. Wu and M. Nakano.

Zheng Yan (Senior Member, IEEE) received the B.Eng. degree in electrical engineering and the M.Eng. degree in computer science and engineering from the Xi'an Jiaotong University, Xi'an, China, in 1994 and 1997, respectively, the second M.Eng. degree in information security from the National University of Singapore, Singapore, in 2000, and the Licentiate of Science and the Doctor of Science in Technology in electrical engineering from Helsinki University of Technology, Helsinki, Finland. She is currently a Professor with Xidian University, Xi'an, China and a Visiting Professor with Aalto University, Espoo, Finland. Her research interests include trust, security, privacy, and security-related data analytics. She is a General or Program Chair for more than 30 international conferences and workshops. She is a steering committee co-Chair of IEEE Blockchain international conference. She is also an Associate Editor of many reputable journals, including the IEEE INTERNET OF THINGS JOURNAL, *Information Sciences*, *Information Fusion*, *Journal of Network and Computer Applications*, IEEE ACCESS, and *Security and Communication Networks*.

Kevin I-Kai Wang (Member, IEEE) received the B.Eng. (Hons.) degree in computer systems engineering and the Ph.D. degree in electrical and electronics engineering from the Department of Electrical and Computer Engineering, University of Auckland, Auckland, New Zealand, in 2004 and 2009, respectively. He is currently a Senior Lecturer with the Department of Electrical and Computer Engineering, University of Auckland. From 2009 to 2011, he was also a Research Engineer designing commercial home automation systems and traffic sensing systems. His current research interests include wireless sensor network based ambient intelligence, pervasive healthcare systems, human activity recognition, behaviour data analytics, and bio-cybernetic systems.