An Inference Similarity-based Federated Learning Framework for Enhancing Collaborative Perception in Autonomous Driving

Zilong Jin[1], Chi Zhang[1], and Lejun Zhang[2*]
[1] School of Software, Nanjing University of Information Science and Technology
Nanjing 210044, China
[e-mail: zljin@outlook.com, zc071099@163.com]
[2] Cyberspace Institute Advanced Technology, Guangzhou University
Guangzhou 510006, China
[e-mail: zhanglejun@gzhu.edu.cn]
[*]Corresponding author: Lejun Zhang

Received November 20, 2023; revised Feburary 21, 2024; accepted April 26, 2024; published May 31, 2024

Abstract

Autonomous vehicles use onboard sensors to sense the surrounding environment. In complex autonomous driving scenarios, the detection and recognition capabilities are constrained, which may result in serious accidents. An efficient way to enhance the detection and recognition capabilities is establishing collaborations with the neighbor vehicles. However, the collaborations introduce additional challenges in terms of the data heterogeneity, communication cost, and data privacy. In this paper, a novel personalized federated learning framework is proposed for addressing the challenges and enabling efficient collaborations in autonomous driving environment. For obtaining a global model, vehicles perform local training and transmit logits to a central unit instead of the entire model, and thus the communication cost is minimized, and the data privacy is protected. Then, the inference similarity is derived for capturing the characteristics of data heterogeneity. The vehicles are divided into clusters based on the inference similarity and a weighted aggregation is performed within a cluster. Finally, the vehicles download the corresponding aggregated global model and train a personalized model which is personalized for the cluster that has similar data distribution, so that accuracy is not affected by heterogeneous data. Experimental results demonstrate significant advantages of our proposed method in improving the efficiency of collaborative perception and reducing communication cost.

Keywords: Autonomous Vehicles, Clustering, Collaborative Perception, Personalized Federated Learning

This work was sponsored by the National Natural Science Foundation of China (Grant Nos. 62271264, 42175194, and 61972207).

http://doi.org/10.3837/tiis.2024.05.005 **ISSN : 1976-7277**

1. Introduction

With the rapid advancements of Information and Communications Technology (ICT), in terms of sensor technology, wireless communications, Artificial Intelligence (AI), and computing resources, smart vehicles can gather data from their surrounding environment [1]. These vehicles can engage in local learning processes and leverage cloud or edge cloud assistants to perform learning tasks. As a result, a multitude of smart applications become possible, including autonomous driving, advanced driver assistance, intelligent navigation, etc.

In the application of autonomous driving, vehicles fuse and process the data obtained from onboard sensors [2], i.e., RGB cameras, ultrasonic radar, lidar, and millimeter wave radar, to generate a comprehensive understanding of the surrounding environment. By utilizing machine learning algorithms, vehicles can compare sensor data with previously acquired knowledge, allowing them to make safe and intelligent driving decisions.

However, insufficient availability of local computing resources and limited sensor data pose a significant obstacle to real-time data learning and surrounding perception in autonomous driving. A promising solution lies in collaboration with other vehicles and edge clouds which can enhance the sensing capabilities of connected vehicles and guarantee real-time data learning [3].

To facilitate effective collaboration, it is crucial to adopt inter-vehicle communication and information-sharing strategies. Sharing information involves the exchange of surrounding perception data between interconnected vehicles and edge clouds through the Internet of Vehicles (IoV). The data exchanges aim to enhance the perception and data learning capabilities of autonomous vehicles, consequently leading to improved driving safety [4]. However, this endeavor gives rise to an additional safety concern wherein data becomes vulnerable to interception during wireless transmission or acquisition by neighboring vehicles acting with malicious intent. Such private data comprises a multitude of sensitive vehicle information, encompassing surrounding perception data, driving control data, driving routes, GPS coordinates, driving preferences, etc. These data elements hold significant importance in ensuring the overall safety of the driving experience.

Given the above concerns, Federated Learning (FL) emerges as a promising learning framework for autonomous driving. FL employs a distributed learning manner where vehicles train their local models using their own data and subsequently update the parameters of these local models to an edge cloud for the purpose of training a global model. Acting as a central server, the edge cloud then multicasts the global model to neighboring vehicles, effectively safeguarding data privacy while simultaneously reducing communication overhead.

However, the traditional FL method assumed i.i.d. (independent and identically distributed) data [5] which is not a common case in autonomous driving due to the data heterogeneity, and it may result in a global model trained by the i.i.d. data being efficient for specific driving cases and unable to handle complex driving scenarios.

In this paper, considering the issue of data being maliciously obtained and the limitation of the i.i.d. data in IoV, a personalized federated learning framework based on inference similarity is proposed. A clustering method is proposed that the vehicles are clustered with similarity of data, and then train the model for each cluster without accessing the private data. This eliminates the risk of data being maliciously intercepted and obtained during transmission and reduces the impact of data heterogeneity on perception effects. The number of clusters does not need to be set in advance, which means that the algorithm can better adjust according to the actual distribution of data. Additionally, logits are transmitted instead of model parameters between clusters and RoadSide Units (RSUs) to reduce communication cost. The contributions of our research are summarized as follows:

- In order to solve the issue of accuracy degradation caused by data heterogeneity in autonomous driving scenarios, a clustering method based on inference similarity is proposed in the framework of federated learning which enables vehicles to learn personalized knowledge in a cluster.
- Some nodes are selected in each round of training. Logits are transmitted instead of model parameters for each round to participate in clustering and updating the local model with logits, further reducing communication cost between vehicles and RSUs, and enhancing the security of data transmission.
- Extensive experiments are conducted on a real-world traffic sign dataset. The simulation results show that the proposed method outperforms state-of-the-art algorithms in terms of perception efficiency and communication cost.

The rest of this paper is organized as follows. The related work on autonomous driving perception, and clustering-based federated learning are reviewed in Section 2. Then, we introduce the analytical model for data distribution and the formation mechanism of clusters in Section 3. The specific process of our proposed clustering-based federated learning method is provided in Section 4. Our proposed method is evaluated through simulation experiments in Section 5. Finally, we conclude our work with remarks in Section 6.

2. Related Work

2.1 Local Perception

Autonomous vehicles require real-time environmental information collection and processing during the driving process. The main tasks include lane and road detection, traffic sign recognition, vehicle tracking, behavior analysis, prediction, and scene understanding [6]. Many studies have been devoted to developing robust and efficient environment perception methods. For example, perception methods based on lidar generated detailed 3D point clouds, capturing the geometric and depth information of surrounding objects [7]. Vision-based perception methods utilize cameras and image processing techniques to extract valuable information from visual data [8], e.g., lane markings, traffic signs, and traffic signals. In addition, recurrent neural networks and long short-term memory networks [9] help model the temporal dependencies in dynamic traffic scenes, enabling better prediction and decision-making. However, there are still some challenges, such as handling occlusions, adverse weather conditions, and real-time processing. Further research and development are needed to achieve fully autonomous and safe driving systems.

2.2 Collaborative Perception

Collaborative perception leverages wireless communication technology to interactively integrate environmental information obtained from distributed vehicles with local perception information. Through cooperation, the perception accuracy of vehicles is improved, and perception blind spots are eliminated. Than. *et al.* [10] investigated the information that should be included in collaborative perception to enhance the perceptual reliability of cars. Gabb. *et al.* [11] proposed a hybrid vehicle perception system that combines local onboard sensor data with received global sensor data. Chen *et al.* [12] conducted research on the collaborative perception of raw data to enhance the detection capability of autonomous driving systems. This approach integrates sensor data from vehicles in IoV at different locations and directions. Arnold *et al.* [13] proposed two novel collaborative 3D object detection schemes named postfusion and pre-fusion, depending on whether the fusion occurs after or before the object detection stage. However, these methods did not consider data privacy and security, as well as the problem of increased data transmission latency due to limited communication resources.

2.3 Clustering Based FL

Federated learning (FL) is a decentralized learning technique where training data is distributed between work nodes, rather than sending raw data to the server for centralized training. To address the issue of heterogeneous data affecting the efficiency of collaborative perception, researchers have proposed several personalized federated learning methods [14]. These methods aim to train one or more different models corresponding to the feature of client perception data [15], such as assigning global data by categories [16] and introducing meta-learning to capture fine-grained information [17].

In contrast, personalized federated learning based on clustering considers cost issues by aggregating clients with similar data distributions into a cluster. As is shown in **Fig. 1**, nodes within each cluster can share knowledge. Each cluster trains different models to adapt to other clients, thereby improving perception efficiency independently. Clustering methods [18] can be categorized as one-shot clustering and iterative clustering. Based on the underlying assumptions about the cluster structure, it can also be divided into inter-cluster knowledge sharing and non-sharing. Ghosh *et al.* [19] studied one-shot clustering and non-sharing methods, where the number of clusters is initially specified. It is similar to *K*-means clustering. However, comparing the Euclidean or cosine distances between neural networks to determine model similarity does not yield satisfactory results. Sattler *et al.* [20] proposed that the number of clustering groups can change during iterations without pre-specifying the number of clusters. Ghosh *et al.* [21] suggested that the server sends *N* models to clients during each round-trip communication. Then clients use their local data to train these models and select the most suitable one. However, this method increases the payload of the downlink *N* times.

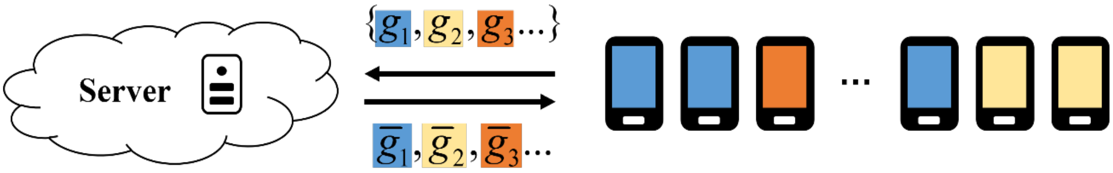

Server
$\{g_1, g_2, g_3\cdots\}$
$\bar{g}_1, \bar{g}_2, \bar{g}_3\cdots$
···

Fig. 1. Clustering Federated Learning

To overcome the above problems, this paper proposes a novel adaptive clustering-based FL method that does not require the prior determination of the number of clusters. It is consistent with the uncertainty of the number of vehicles and the constantly changing traffic conditions in autonomous driving scenarios. The vehicle clustering method that does not require setting the number of clusters in advance has better adaptability and universality, and can better adapt to different types and distributions of data, thereby better discovering similar structures in the data and making corresponding adjustments. All onboard clients are assigned to the most relevant clusters. Nodes within each cluster share a set of averaged parameters. The algorithm reassigns nodes to clusters in each iteration by minimizing the loss function.

3. Problem Definition

This paper considers a road scene composed of multiple vehicles and RSUs, each autonomous vehicle utilizes the sensing data to understand the surrounding environment, such as lane detection, traffic signs, cars, cyclists, and pedestrians. In the communication range, it is assumed that some vehicles are located in the blind spot of other vehicles. Obstacles or severe weather conditions may obstruct their sensors and cause difficulty in recognizing images captured by cameras. Under the coordination of the RSUs, multiple vehicles collaborate to train a global model. By learning from the global model sent by the RSU, they enhance their perception efficiency and improve their ability to understand the surrounding environment and recognize traffic signs. The proposed clustering-based FL framework is illustrated in **Fig. 2**.

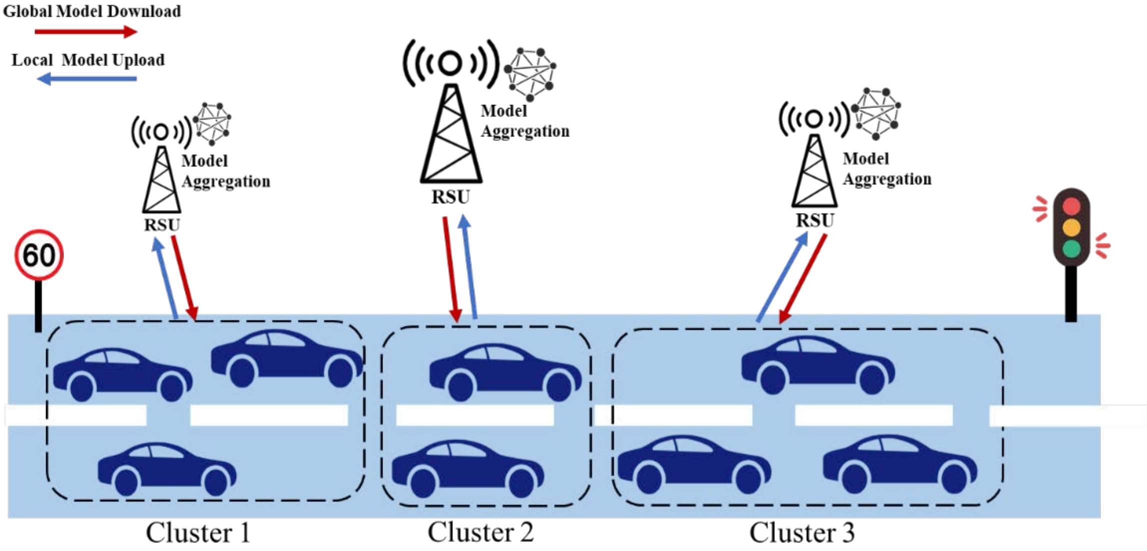


Fig. 2. Collaborative Perception of Vehicles through RSU

3.1 Data Distribution

Given the complexity of autonomous driving scenarios, this paper aims to use a federated learning framework to collaborate and train personalized learning models in traffic sign recognition for vehicle clients with different perceived data distributions. Assuming there are K vehicles, each vehicle can only access its private dataset D_k. Each data sample ξ is represented by (x, y), where $x \in \mathbb{R}^d$ denotes the input data, and $y \in [1, N]$ represents the corresponding label. $D_k = \{D_1, \ldots, D_k\}$ denotes the collection of datasets of all vehicles. This paper assumes that these datasets follow various non-IID distributions.

- Feature distribution skew: The marginal distribution of $P_i(x)$ varies among different clients, while the conditional distribution of $P(y|x)$ is the same. In other words, the marginal distributions of data samples in D_i and D_j are different. For example, different countries may have different representations for the same traffic sign features in a dataset of vehicle identification.
- Label distribution skew: The marginal distribution of $P_i(y)$ varies among different clients, while the conditional distribution of $P(x|y)$ is the same. In other words, the marginal distributions of labels in D_i and D_j are different.
- Same label, different features: The conditional distribution of $P(x|y)$ varies among various

clients, and the marginal distribution of $P_i(y)$ is the same. In other words, different clients may have different feature representations for the same label.

- Same features, different label: The conditional distribution of $P(x|y)$ varies among different clients, and the marginal distribution of $P_i(x)$ is the same. In other words, various labels may be assigned to the exact feature representations in different client data.

V2V data exhibits highly non-IID characteristics. For instance, in a given scenario, different vehicles may capture slightly different features in the same perceptual data due to varying viewpoints. Therefore, the data labels obtained may be different. When data is distributed in a highly non-IID manner, more than one model is required to meet the requirements of all vehicles. Therefore, it becomes necessary to establish multiple models tailored to vehicle clusters with similar data distributions.

3.2 Clustering Model

The vehicles are managed in clusters based on the inference similarity which is utilized to identify vehicles with similar data distributions without prior knowledge about the data distribution and the number of clusters is not known in advance. An adjacency matrix is taken as input and group similar vehicles are into clusters. Based on the specific features of the model trained from different user data, RSU constructs the adjacency matrix $A_{i,j}$, i, j=1, …, $|S_t|$, i and j are from set {1, 2, …, $|S_t|$}, using the computation results from each vehicle. This step facilitates the creation of subsets of similar data within each cluster. We define a hard thresholding operator Γ which is applied on $\widetilde{A}_{i,j}=\Gamma(A_{i,j})=Sign(A_{i,j}-\beta)$. The distance threshold is called the clustering threshold and is shown by β. Then, the values are grouped in each row of $\widetilde{A}_{i,j}$ into the same cluster.

4. The Proposed Scheme

In standard federated learning, the vehicles aim to collaboratively find the parameter vector of the model $w \in R^n$ that minimizes the empirical loss, mapping the input data x to the label y.

$$F(w)=\sum_{k=1}^{K}\frac{|D_k|}{\sum_{k=1}^{K}|D_k|}F_k(w) \tag{1}$$

where the function $F_k(w)$ represents the local objective of the user vehicle k, defined as

$$F_k(w)=\frac{1}{|D_k|}\sum_{\xi\in D_k} f(w,\xi) \tag{2}$$

where $f(w,\xi)$ is a composite loss function.

Due to the data heterogeneity across vehicles, the optimal model parameters w^* that minimize $F(w)$ can generalize poorly to vehicles whose local objective $F_k(w)$ significantly differs from $F(w)$. Additionally, data transmission during vehicles and RSU leads to an increase in communication cost. A clustering method is proposed in this paper which utilizes the data similarity of different vehicles. This method proves that the vehicles can benefit from other users in a cluster with the improved generalization ability of the learning model and decrease the amount of data transmission to reduce communication cost.

In the proposed clustering method, the classification model $w_k, k \in [K]$ outputs logits on a predefined number N of classes, representing a probability vector over the N classes. The logits of the model w_k on the data x from the private dataset are defined as $g(w_k, x)$, $x \in D$, stacked into rows for each x. Each vehicle seeks to find the model parameters w_k that minimize the empirical risk $L\left(w_k; \bar{g}_k\right)$, which is the sum of the empirical risk of its own local training data $F_k(w_k)$ and the regularization term as follows:

$$L\left(w_k; \bar{g}_k\right) = F_k\left(w_k\right) + \frac{\lambda}{|D_k|} \sum_{x \in D} \left\| \bar{g}_k\left(x\right) - g\left(w_k, x\right) \right\|_2^2 \tag{3}$$

In (3), $\bar{g}_k\left(x\right) = \sum_{i=1}^{K} \alpha_{k,i}\, g\left(w_i, x\right)$ denotes the weighted average of the logits from all vehicles for an arbitrary set of weights for the vehicle, i.e., $\{\alpha_{k,i}\}_{i \in [K]}$ and $\sum_{i=1}^{K} \alpha_{k,i}=1$. The weight of the regularization term is modulated by λ. Therefore, vehicles with similar logits have higher weights towards each other and they are merged into a cluster. The RSU computes the weighted average of logits for each cluster and sends it to the corresponding vehicles. The client updates the model parameters with the corresponding logits, and then sends the updated parameters back to RSU. The overview of the proposed scheme is shown in **Fig. 3**.

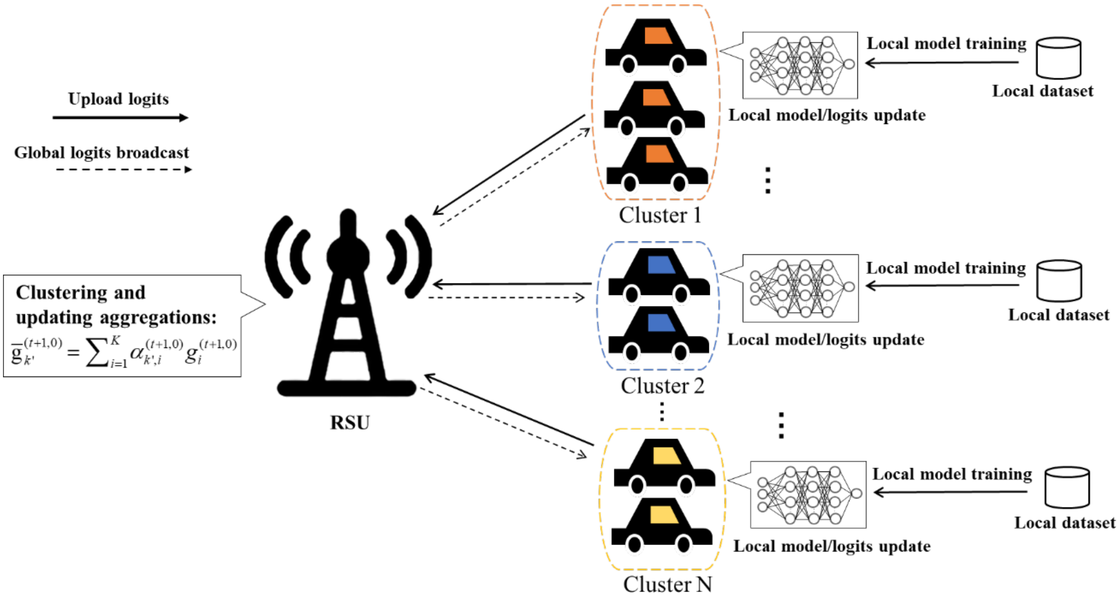

Upload logits
Global logits broadcast
Clustering and updating aggregations:
$\bar{g}_{k'}^{(t+1,0)} = \sum_{i=1}^{K} \alpha_{k',i}^{(t+1,0)} g_i^{(t+1,0)}$
RSU
Local model training
Local dataset
Local model/logits update
Cluster 1
Local model training
Local dataset
Local model/logits update
Cluster 2
Local model training
Local dataset
Local model/logits update
Cluster N

Fig. 3. Overview of the Proposed Scheme

```
Algorithm
Input: Number of available vehicles $N$, sampling rate $R \in \{0, 1\}$, clustering threshold $\beta$
Output: $\{w_k\}_{k \in [K]}$

Initialize: $\{w_k^{(0,0)}\}_{k \in S_t}$, selected set of $n$ vehicles $S_t$
For $t$=0,1,2 ... ,$T$-2, $T$-1 communication rounds do:
    Vehicle $k$ do:
        Get $g_k^{(t,0)}$ for current local model
        For $r$=0, ... ,$\tau$-1 local iterations do:
```

```
        Update $w_k^{(t,r+1)} \leftarrow w_k^{(t,r)} - \eta \cdot h_k\left(w_k^{(t,r)}; \overline{g}_k^{(t,0)}\right)$
        Send $s_k^{(t+1,0)} = s_k^{(t,\tau)}$ for the updated local model
        $w_k^{(t,\tau)}$ back to the RSU
    RSU in parallel do:
        Get $g_k^{(t,r)}$ from each vehicle
        $A_{i,j} = \frac{\| g_i \odot g_j \|_F}{\| g_i \|_F \| g_j \|_F}; i, j = 1, \ldots, n$
        // RSU constructs the adjacency matrix
        $\widetilde{A}_{i,j} = \Gamma(A_{i,j}) = Sign(A_{i,j} - \beta)$
        // RSU applies hard thresholding and does clustering
        Return $\{C_{j_{t+1}}\}_{j_{t+1}=1}^{T_{t+1}}$
        $\overline{g}_{k'}^{(t+1,0)} = \sum_{i=1}^{K} \alpha_{k',i}^{(t+1,0)} g_i^{(t+1,0)}$
```

In the first round, the RSU broadcasts initialized model parameters $w_k^{(0,0)}$. Each vehicle has its private dataset D_k, which contains multiple samples. The classification model $w_k, k \in [K]$ outputs logits in a predetermined number of categories N, a probability vector for N categories. The logits of the model w_k on the input data x from the private dataset are defined as $g(w_k, x)$.

Each vehicle trains its own data using the initial model and sends the updated model logits to the RSU for similarity clustering. No private information needs to be received about the data for inference similarity clustering. An adjacency matrix is constructed based on $\{g_k^t\}_{k \in S_t}$ obtained from the vehicle clients.

$$A_{i,j} = \frac{\| g_i \odot g_j \|_F}{\| g_i \|_F \| g_j \|_F}; i, j = 1, \ldots, n \tag{4}$$

The symbol $\odot$ represents Hadamard product. The hard thresholding operator is defined as Γ which is applied on $A_{i,j}$, and yields as a threshold value. The RSU groups the values of each row into the same cluster with similar information to form clusters $\{C_{j_{t+1}}\}_{j_{t+1}=1}^{T_{t+1}}, j_t = 1, \ldots, T_t$. Finally, the RSU computes the weighted average of logits in the same cluster and sends them back to the corresponding vehicle. The vehicles use the obtained new weighted model for the next round of local model updates and communication. The logits for the updates within each cluster are defined as:

$$\overline{g}_k(x) = \sum_{i=1}^{K} \alpha_{k,i} g(w_i, x) \tag{5}$$

In the t-th iteration, the RSU samples data from the vehicles and broadcasts the current parameters from the model parameters $g_k^{(t,0)}$ to the onboard client. The empirical loss over local data typically defines the local objective L_k. Each vehicle then estimates its cluster identity and obtains the corresponding training model by finding the model parameter that yields minimum loss on its test data,

$$g_{k,j_t}^{(t+1,0)} = \arg\min L_k\left(D_k^{test}; g_{k,j_t}^{(t,0)}\right) \tag{6}$$

Afterward, the vehicles perform τ-step stochastic gradient descent update, and these updated parameters $g_k^{(t+1,0)}$ are then sent back to the RSU. The vehicles and RSU only communicate logits instead of model parameters. The formula for the local model update of each vehicle in each round is as follows:

$$w_k^{(t,r+1)} = w_k^{(t,r)} - \eta \cdot h_k\left(w_k^{(t,r)}; \bar{s}_k^{(t,0)}\right) \tag{7}$$

$$\begin{aligned} h_k\left(w_k^{(t,r)}; \bar{s}_k^{(t,0)}\right) = {} & \frac{2\lambda}{\left|D_k\right|} \sum_{x \in D_k} \nabla g\left(w_k^{(t,r)}, x\right)^T \left(g\left(w_k^{(t,r)}, x\right) - \bar{g}_k^{(t,0)}\left(x\right)\right) \\ & + \frac{1}{\left|\xi_k^{(t,r)}\right|} \sum_{\xi \in \left|\xi_k^{(t,r)}\right|} \nabla f\left(w_k^{(t,r)}, \xi\right) \end{aligned} \tag{8}$$

The term $w_k^{(t,r)}$ denotes the local model parameters of the vehicle k, η is the learning rate, $\xi_k^{(t,r)}$ is the mini-batch randomly sampled from the dataset of local vehicle k, and D_k represents the data samples of vehicle k. The number of local iterations r is fixed for each round. We use $w_k^{(t+1,0)} = w_k^{(t,\tau)}$ to represent the local model updated by vehicle k after all the local iterations in iteration t. The RSU reuses updated parameters to form a dynamic vehicle cluster with similar data and computes the average parameter for each cluster.

5. Experiments

5.1 Experimental Settings

5.1.1 Datasets and Models

This paper evaluated the performance of the proposed model on two popular datasets: MNIST [22] and CIFAR-10 [23]. The dataset MNIST contains 60,000 training images and 10,000 testing images. MNIST has become a standard benchmark for evaluating the performance of new machine learning algorithms. The dataset CIFAR-10 consists of 60,000 32x32 color images in 10 classes which is divided into 50,000 training images and 10,000 testing images. It contains complex objects such as animals, ships, cars, and airplanes. Additionally, Belgium TSC [24] and GTSRB [25] datasets were used to assess the performance of environment perception in the context of autonomous driving. They include 62 classes of traffic signals and 42 classes of traffic signs. Each image in the Belgium TSC dataset represents a traffic sign captured under various environmental conditions. The dataset is designed to support research on traffic sign recognition and classification tasks. The dataset GTSRB covers a wide range of traffic sign classes, including speed limits, no-entry signs, yield signs, and various other regulatory, warning, and information signs commonly found on roadways.

A convolutional neural network was constructed as the training model. The model comprises two convolutional layers, two pooling layers, and three fully connected layers (with the last fully connected layer as the output layer).

5.1.2 Baselines

To demonstrate the effectiveness of our proposed method, we compared it against the following approaches: 1) Methods that aim to learn a single global model: FedAvg [26] and FedProx [27]. FedAvg can be considered a form of collaborative learning among distributed devices. FedProx, on the other hand, can be seen as a generalization and refinement of FedAvg; 2) Personalized federated learning methods: PerFedAvg [28] involves transferring and fine-tuning the initial model parameters. And FedFomo [29] determines that which models should be computed by the RSU and sent to which vehicles.

5.1.3 Training Settings

We use SGD as the local optimizer for all methods and set the batch size to 32 with a learning rate of 0.001. One hundred vehicular clients are considered and ten vehicles are randomly selected for training each round. The purpose of this setting is to fully utilize a wider and more diverse range of data samples and protect privacy while selecting a subset of vehicles to reduce communication and computing costs. Three hundred rounds of communication training were conducted for each dataset, which is sufficient for the model to converge.

Table 1. Parameter Settings for Simulation Experiments

Parameter	**Value**
Bandwidth	5 MHz
Vehicle Computing Power	4×10^6~2×10^7 cycles/s
Distance between Vehicle and RSU	30~50 m
Vehicle Transmission Power	1.3 W
Batch Size	32
Learning Rate	0.001
Number of Vehicles	100
Weight of Clients Selected for Training Each Round	0.1

5.2 Results

5.2.1 Model Performance

In our experiments, we ran multiple times and recorded the average results.

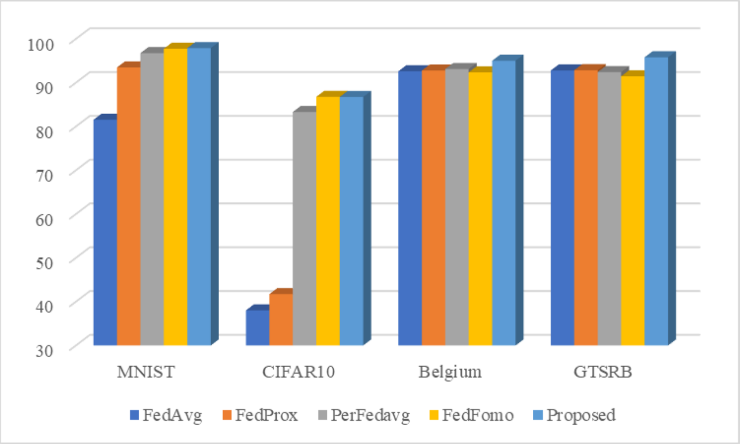


Fig. 4. Accuracy of Five Algorithms on Different Datasets

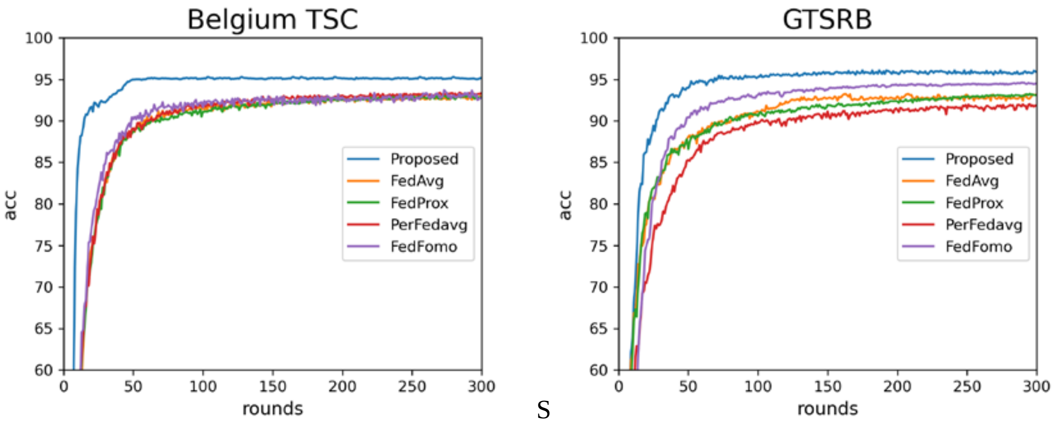
Table 2. For a homogeneous scenario with a total number of vehicles *K*=100, the average testing accuracy of the entire vehicle and the whole communication cost (number of parameters per round of communication) using the GTSRB dataset as an example.

Algorithm	**Test Acc.**				**Com.-Cost**
	MNIST	CIFAR-10	Belgium TSC	GTSRB	
FedAvg	81.53±0.58	38.01±1.91	92.61±0.22	92.79±0.38	6.458E +09
FedProx	93.47±1.85	41.72±1.05	92.77±0.15	92.84±0.24	6.458E +09
PerFedAvg	97.75±0.23	83.31±0.89	93.12±0.16	92.41±1.53	6.458E+09
FedFomo	97.75±0.11	86.77±0.53	92.36±0.42	91.45±0.21	1.162E+10
Proposed	**97.89±0.21**	**86.73±1.21**	**95.01±0.12**	**95.79±0.68**	**3.5E+07**

In **Table 2**, we show the performance of our proposed method along with the performance of comparison algorithms regarding the highest test accuracy and communicated number of parameters between RSU and vehicles. For the MNIST and CIFAR-10 datasets, our proposed algorithm achieves significantly higher accuracy than FedAvg and FedProx, and achieves comparable results to the personalized method PerFedAvg and FedFomo. The bar chart in **Fig. 4** clearly illustrates the comparative effects of model performance, showing that our proposed method achieves the highest accuracy across different datasets.

5.2.2 Environmental Perception Performance

The proposed algorithm achieves higher accuracy for the Belgium TSC and GTSRB datasets than the four comparison algorithms. **Fig. 5** shows the training accuracy curves and the number of communication rounds for various methods on the Belgium TSC and GTSRB datasets. It can be observed that our proposed algorithm reaches convergence around 50 to 70 rounds. The four comparison algorithms show a slow increase in accuracy even after 200 rounds. They go convergence only after nearly 300 rounds. The personalized federated learning method, FedFomo, demonstrates convergence performance that is second only to our method. This indicates that our proposed method significantly improves perception efficiency in general cooperative perception tasks.



S

Fig. 5. Relationship between Testing Accuracy and Communication Rounds

From the Com.-Cost values in Table 2, it can be observed that PerFedAvg, FedFomo, and our proposed method achieve similar performance. The total number of model parameters communicated between the uplink and downlink is 6.458E+10 and 1.162E+10, respectively. Our proposed algorithm requires a communication cost of only 3.5E+07 parameters, resulting in a maximum saving of up to 332 times.

Fig. 6 displays the latency generated by different vehicles during a single training round, comparing it with two baseline methods, FedAvg and FedFomo. The results demonstrate that the latency of these two baseline methods is significantly higher than the proposed method in this paper. Due to their inability to adapt to the constantly changing computing power of vehicular clients, the training latency of baseline methods exhibits a wide range of fluctuate ons. On the other hand, the proposed method shows a stable latency range throughout the training process. The latency is significantly lower than that of the two baseline methods. This indicates that the proposed method is better suited for autonomous driving scenarios and provides more efficient and reliable communication latency performance in federated learning.

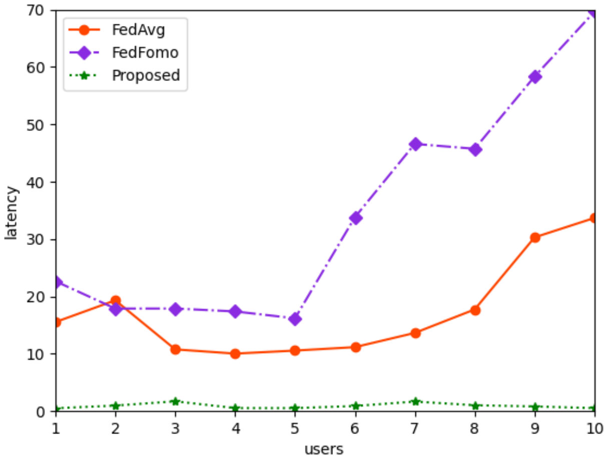


Fig. 6. Latency Generated by Vehicles.

6. Conclusion

This paper has proposed a personalized federated learning framework based on similarity inference clustering for enhancing collaborative perception in autonomous driving scenarios. The proposed scheme can achieve more accurate perception results by introducing the inference similarity of models and grouping vehicles with similar data. Additionally, transmitting a more miniature parameter instead of the entire model can reduce communication overhead and improve communication efficiency. We have validated the proposed algorithm on real-world datasets which has exhibited superior performance compared to existing collaborative perception methods. Therefore, this approach can effectively be applied to environment perception in autonomous vehicles, improving the quality and efficiency of real-time vehicle services.

The future work of this study mainly lies in two aspects. One is to combine the hierarchical idea of triplets [30] to further optimize and find higher precision solutions. Another is to combine federated learning and resource allocation to better fit real-world scenarios.

References

[1] Abdel-Aziz, Mohamed K., et al., "Vehicular cooperative perception through action branching and federated reinforcement learning," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 891-903, Feb.2022. Article (CrossRef Link)

[2] Lee, Gyu Ho, Ki Hoon Kwon, and Min Young Kim, "Ambient environment recognition algorithm fusing vision and LiDAR sensors for robust multi-channel V2X system," in *Proc. of International Conference on Ubiquitous and Future Networks*, Zagreb, Croatia, pp.98-101, 2019. Article (CrossRef Link)

[3] Aoki, Shunsuke, Takamasa Higuchi, and Onur Altintas, "Cooperative perception with deep reinforcement learning for connected vehicles," in *Proc. of IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, pp. 328-334, Oct. 2020. Article (CrossRef Link)

[4] Das, Subasish, "Autonomous vehicle safety: Understanding perceptions of pedestrians and bicyclists," *Transportation research part F: traffic psychology and behaviour*, vol. 81, pp. 41-54, Aug. 2021. Article (CrossRef Link)

[5] Kopparapu K, Lin E, "Fedfmc: Sequential efficient federated learning on non-iid data," *arXiv preprint arXiv:2006.10937*, 2020. Article (CrossRef Link)

[6] Chen Q, Xie Y, Guo S, et al., "Sensing system of environmental perception technologies for driverless vehicle: A review of the state of the art and challenges," *Sensors and Actuators A: Physical*, vol.319, pp.112566, Mar. 2021. Article (CrossRef Link)

[7] Chen, Siheng, et al., "3d point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 68-86, Jan. 2021. Article (CrossRef Link)

[8] Azfar T, Li J, Yu H, et al., "Deep Learning based Computer Vision Methods for Complex Traffic Environments Perception: A Review," *arXiv preprint arXiv:2211.05120*, 2022. Article (CrossRef Link)

[9] Zou Q, Hou Y, Wang Z, "Predicting vehicle lane-changing behavior with awareness of surrounding vehicles using LSTM network," in *Proc. of International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Singapore, pp. 79-83, Dec. 2019. Article (CrossRef Link)

[10] G. Thandavarayan, M. Sepulcre and J. Gozalvez, "Generation of cooperative perception messages for connected and automated vehicles," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16336-16341, Dec. 2020. Article (CrossRef Link)

[11] M. Gabb, H. Digel, T. Müller and R.-W. Henn, "Infrastructure-supported perception and track-level fusion using edge computing," in *Proc. of IEEE Intelligent Vehicles Symposium (IV)*, Paris, France, pp. 1739-1745, Jun. 2019. Article (CrossRef Link)

[12] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative perception for connected autonomous vehicles based on 3D point clouds," in *Proc. of International Conference on Distributed Computing Systems (ICDCS)*, Dallas, TX, USA, pp. 514-524, Jul. 2019. Article (CrossRef Link)

[13] E. Arnold, M. Dianati, R. de Temple and S. Fallah, "Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol.23, no.3, pp.1852-1864, Mar. 2022. Article (CrossRef Link)

[14] Tan, Alysa Ziying, et al., "Towards personalized federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 12, pp. 9587-9603, 2023. Article (CrossRef Link)

[15] Marfoq, Othmane, et al., "Federated multi-task learning under a mixture of distributions," *Advances in Neural Information Processing Systems*, vol. 34, pp. 15434-15447, 2021. Article (CrossRef Link)

[16] Zhao Y, Li M, Lai L, et al., "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018. Article (CrossRef Link)

[17] Jiang Y, Konečný J, Rush K, et al., "Improving federated learning personalization via model agnostic meta-learning," *arXiv preprint arXiv:1909.12488*, 2019. Article (CrossRef Link)

[18] Kim Y, Al-Hakim E, Haraldson J, et al., "Dynamic clustering in federated learning," in *Proc. of International Conference on Communications*, Montreal, QC, Canada, pp.1-6, Aug.2021. Article (CrossRef Link)

[19] Ghosh A, Hong J, Yin D, et al., "Robust federated learning in a heterogeneous environment," *arXiv preprint arXiv:1906.06629*, 2019. Article (CrossRef Link)

[20] F. Sattler, K.R. Muller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3710–3722, Aug. 2021. Article (CrossRef Link)

[21] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An Efficient Framework for Clustered Federated Learning," *IEEE Transactions on Information Theory*, vol. 68, no.12, pp. 8076-8091, Jul. 2022. Article (CrossRef Link)

[22] Y. LeCun and C. Cortes, "MNIST Handwritten Digit Database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[23] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton, "Learning multiple layers of features from tiny images," *CIFAR-10* (*Canadian Institute for Advanced Research*), 2009. Article (CrossRef Link)

[24] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localization," in *Proc. of 2009 Workshop on Applications of Computer Vision (WACV)*, pp. 1-8, 2009. Article (CrossRef Link)

[25] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," in *Proc. of* The *2011 International Joint Conference on Neural Networks*, pp. 1453-1460, San Jose, CA, USA, 2011. Article (CrossRef Link)

[26] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. of the 20th International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, Florida USA, pp. 1273-1282, 2017. Article (CrossRef Link)

[27] LI T, SAHU A K, ZAHEER M, et al., "Federated optimization in heterogeneous networks," in *Proc. of Machine Learning and Systems*, Austin, TX, USA, pp.429-450, 2020. Article (CrossRef Link)

[28] Fallah A, Mokhtari A, Ozdaglar A, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," in *Proc. of 34th Conference on Neural Information Processing Systems*, Vancouver, Canada, 2020. Article (CrossRef Link)

[29] Zhang M, Sapra K, Fidler S, et al., "Personalized federated learning with first order model optimization," *arXiv preprint arXiv:2012.08565*, 2020. Article (CrossRef Link)

[30] Yang Q, Qiao Z Y, Xu P, et al., "Triple competitive differential evolution for global numerical optimization," *Swarm and Evolutionary Computation*, vol.84, 101450, 2024. Article (CrossRef Link)

Zilong Jin received the B.E. degree in computer engineering from Harbin University of Science and Technology, China, in 2009, and the M.S. and Ph.D. degrees in computer engineering from Kyung Hee University, Korea, in 2011 and 2016, respectively. He is currently an associate professor at School of Software at Nanjing University of Information Science and Technology, China. His research interests include mobile wireless networks, cognitive radio networks, and mobile edge networks.

Chi Zhang received her B.E. degree in information security from Nanjing University of Information Science and Technology, China, in 2021. Now she is a master student in School of Software, Nanjing University of Information Science and Technology. Her main research interests include internet of vehicles and federated learning.

Lejun Zhang received his M.S. degree in computer science and technology in Harbin Institute of Technology and the Ph.D. degrees in computer science and technology at Harbin Engineering University. Now he is currently a professor and Ph.D. Supervisor of the Cyberspace Institute of Advanced Technology, Guangzhou University. He was a Visiting Scholar with Carnegie Mellon University. His research interests include Cyberspace Security, blockchain and information security.