Unsupervised Domain Adaptive Vehicle Re-Identification: A Federated Learning Scheme

Xiao Xiao, *Member, IEEE*, Xinyue Yang, Yucheng Wang, Yilong Hui, *Member, IEEE*, Jianchuan Zhou, and Guoqiang Mao, *Fellow, IEEE*

***Abstract*— While traditional domain adaptation methods have yielded satisfactory outcomes in vehicle re-identification on established datasets, they typically operate within a centralized framework. This approach falls short in addressing the challenges of large-scale cross-domain adaptation encountered in real-world traffic networks, and often overlooks the computational constraints of roadside units. In response, we introduce a federated learning-based distributed framework for unsupervised domain adaptive vehicle re-identification, structured in three stages: local model processing, dynamic model aggregation, and aggregated model downloading. Initially, diverse local re-identification models are trained independently within edge computing units. For unlabeled target data, we employ a novel clustering technique that averages features from multiple local models to generate pseudo labels. The framework's second stage incorporates a historical accumulation strategy to retain and utilize the knowledge from local models, ensuring effective adaptation. Additionally, we propose a weight regularization approach for dynamic model aggregation, significantly boosting the framework's discriminative capabilities. Our methodology, tested on the VeRi-776 and VehicleID datasets, showcases marked improvements in the discriminative performance of domain adaptive vehicle re-identification models, affirming the efficacy of our approach.**

***Index Terms*— Federated learning, domain adaptation, deep learning, vehicle re-identification, smart city.**

I. Introduction

VEHICLE re-identification (re-ID), which is crucial for intelligent transportation system (ITS) and smart city applications [1], aims to match vehicles across different camera views. Although license plate recognition is a common approach [2], [3], [4], visual features offer a more tamper-resistant identifier. Deep learning has significantly advanced vehicle re-ID, with a primary focus on supervised, unsupervised, and domain adaptation techniques.

Manuscript received 6 April 2023; revised 5 November 2023 and 6 February 2024; accepted 9 February 2024. Date of publication 11 September 2025; date of current version 21 October 2025. This work was supported by the Key Project of the Joint Fund of the National Natural Science Foundation of China under Grant U21A20446. The Associate Editor for this article was S. S. Nedevschi. *(Corresponding author: Xinyue Yang.)*

Xiao Xiao, Xinyue Yang, Jianchuan Zhou, and Guoqiang Mao are with the School of Telecommunications Engineering, Xidian University, Xi'an 710068, China (e-mail: xinyueyang0212@163.com).

Yucheng Wang is with the Perception Team, Horizon Robotics, Beijing 100094, China.

Yilong Hui is with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China.

Digital Object Identifier 10.1109/TITS.2024.3368515

The majority of supervised learning based vehicle re-ID methods focus on robust and effective feature learning, such as fusing multiple features extracted from different networks [5], [6], [7], [8], exploring discriminative information in part features [9], [10], [11], [12]. These methods have shown excellent performance by leveraging vehicle identity annotations, but are limited by the need for large amounts of annotated data. Hence, some works adopt unsupervised learning instead of supervised learning to tackle the vehicle re-ID task. The main idea of the unsupervised learning method is to utilize pseudo-label generation techniques for unlabeled vehicle samples to train a re-ID model, such as clustering [13], [14], [15]. Although these methods can overcome the difficult circumstances where the number of annotations is limited, they suffers from poor generalization ability. That is, an unsupervised learning-based model performs well on a particular dataset, but its accuracy inevitably drops when applied to a more general dataset.

Unsupervised domain adaptive (UDA) re-ID methods, efficiently addressing the aforementioned problem, have been widely proposed in recent years [16], [17], [18], [19]. These methods represent a key approach in transfer learning, utilizing samples from the source domain to enhance performance on unlabeled target-domain datasets. Through domain adaptation, the model's generalization capability can be significantly improved.

However, there still remains two challenges for UDA vehicle re-ID:

- **Multiple large cross-domain gaps:** The significant differences in feature distribution across datasets from various traffic cameras create multiple large domain gaps. These gaps, exacerbated by different camera resolutions and locations, make it challenging for models to learn a feature representation that performs well across both source and target domains. Handling multiple such gaps can lead to underperformance, even in trained models.
- **Centralized setting issues:** Traditional vehicle re-ID methods operate in a centralized setting, where vast amounts of image data from numerous cameras are sent to a central server. This approach results in substantial bandwidth usage, delays, and raises privacy concerns due to all data being stored in one location. Additionally, it is not well-suited for multi-domain adaptation, a necessity

for effective vehicle re-ID in real-world scenarios, and struggles to scale as more cameras are added, leading to performance drops. These issues have been largely overlooked in domain adaptive vehicle re-ID research.

Federated learning, an emerging and promising paradigm in distributed machine learning, has garnered significant interest [20]. The basic idea of federated learning is to train a model on decentralized data from multiple devices or locations without requiring data to be centralized on a single server. Federated learning reduces data transfer, protects local data privacy, and improves model generalization by learning from diverse data sources, offering a scalable and adaptable solution for large traffic networks.

However, integrating the federated learning framework into UDA for re-ID is far from straightforward:

- **Label dependency**: Federated learning typically relies on labeled data for training, but in UDA tasks, target-domain data often lacks labels. This discrepancy poses a significant challenge for applying federated learning to UDA re-ID tasks.
- **Historical knowledge ignoration**: In federated learning, the computational workload is predominantly shifted to the main server, and local data remains at the edge computing units to maintain privacy. This setup implies that the domain adaptation process in the main server relies solely on the local models received in the current iteration, without access to any historical model data. This lack of access to temporal information could potentially hinder the adaptation process, as the server might miss out on valuable insights from the models' evolution over time, which are crucial for accurately bridging the domain gaps.
- **Simple model aggregation**: Common federated learning aggregation strategies, such as FedAvg [21], treat all client models equally, regardless of their performance. This approach can be problematic in UDA re-ID, where different source-domain datasets may lead to varying levels of recognition performance across re-ID models.

To address these challenges, we introduce a novel three-stage unsupervised domain adaptive vehicle re-ID framework that encompasses local model processing, dynamic model aggregation, and aggregated model deployment. Our approach is particularly innovative in the dynamic model aggregation stage, where domain adaptation and local model integration are concurrently executed. Firstly, to counter the absence of labels in the target dataset, we generate pseudo labels by predicting the average clustering features derived from multiple pre-trained source-domain re-ID models. Secondly, we incorporate a historical accumulation strategy during the domain adaptation phase. This strategy aims to retain the knowledge accumulated by local models from previous iterations, facilitating predictions that guide the learning of other models. Thirdly, we introduce a weight regularization strategy for dynamic aggregation within our federated vehicle re-ID framework. This strategy assigns differential weights to models based on their feature clustering performance, thereby bolstering the discriminative prowess of the overall framework.

The contributions of this paper can be summarized as follows:

- We firstly propose a three-stage federated learning framework for unsupervised domain adaptation vehicle re-ID. This approach enables re-ID models to be trained locally on data from individual cameras, and then combines them in a central server while achieving domain adaptation across cameras.
- Our work advances federated learning by introducing a unique pseudo labeling technique through average feature clustering, incorporating historical accumulation strategy to preserve and utilize historical knowledge, and implementing a weight regularization strategy for dynamic aggregation, all of which collectively enhance the discriminative capability of the system.
- Our proposed framework, rigorously tested on the VeRi-776 and VehicleID benchmark datasets, showcases competitive performance, achieving notable gains over existing domain adaptation-based vehicle re-ID methods.

II. Related Work

A. Vehicle Re-Identification

To obtain an optimal performance on the existing datasets, many deep learning based vehicle re-ID algorithms adopt supervised learning to train models. For instance, Liu et al. proposed FACT [5] combines traditional features [6] with GoogleNet [22] features, BOW-SIFT [23] and BOW-CN [24] features to form more robust features that can enhance the recognition capability of the re-ID model. The progressive vehicle re-ID method proposed in [7] integrates Null Foley-Sammon Transform (NFST) [25] into FACT algorithm for deeper feature extraction and fusion. To address viewpoint variations that can affect the appearance of vehicles, Zhu et al. [26] proposed a quadruple directional deep learning networks to represent features of the vehicle from multiple directions. In order to better explore latent viewpoints, Chu et al. [27] proposed a metric learning based viewpoint-aware method for vehicle re-ID to respectively learn similar viewpoint metrics and different viewpoint metrics. Their method can greatly improve the re-ID accuracy by training with constraints for within-space and cross-space.

However, supervised learning requires a large number of annotations, making it both labor-intensive and time-consuming. As an alternative, unsupervised learning has emerged as a better choice for vehicle re-ID. These methods are typically clustering-based iterative learning approaches that use pseudo-labels generated by clustering extracted features to train re-ID models. For instance, the VR-PROUD method proposed in [14] uses K-Means clustering, while the UDAR method proposed in [16] uses DBSCAN clustering to obtain reliable pseudo-labels. Apart from the above methods, a viewpoint-aware clustering algorithm was proposed in [15] for unsupervised vehicle re-ID. This algorithm uses predicted viewpoints to cluster training samples in a progressive manner, thereby facilitating the exploration of relationships between samples. A triplet network based unsupervised vehicle re-ID method was proposed in [28], which can close the distance of

samples with the same identity while maintaining the distance of different identities in the feature space by utilizing pairwise and triplet constraints.

To improve the performance of unsupervised methods on cross-datasets, researchers have been studying UDA re-ID methods. Wang and Zeng [17] firstly introduced domain adaption for unsupervised vehicle re-ID and used Maximum Mean Discrepancy (MMD) to narrow down the bias between source domain and target domain. Peng et al. [18] proposed a domain adaptive vehicle re-ID method named PAL, which can progressively adapt to target domain from source domain. In PAL, clustering based pseudo target samples and real unlabeled samples selected in dynamic sampling are combined together to make the training faster. Huang et al. [19] proposed dual domain multi-task model (DDM) to handle the problem of trivial vehicle appearance differences. The method can divide the dataset into two domains according to the frequency of each training sample and use a progressive strategy to accelerate the training speed. In some studies, Generative Adversarial Networks (GAN) [29] are used to transfer the style of an image to other domains, as seen in [30] and [31]. In such work, vehicle re-ID is achieved by comparing the similarities and differences in the performance of image samples across different domains. An embedding adversarial learning network called EALN is proposed in [32] to enhance the capability of distinguishing similar vehicles of vehicle re-ID model, which can generate hard negative samples in an embedding space. However, most of aforementioned methods utilize a centralized training mode which are inappropriate for large datasets in real traffic networks with significant cross-domain differences. Meanwhile, the issue of computing efficiency is largely ignored by these methods, which should be considered when facing real-world scenarios.

B. Federated Learning

Federated Learning was first proposed by Google [33], which was originally designed to use data distributed across multiple mobile devices to train machine learning models while preserving user privacy. Since then, federated learning has aroused widespread interest among researchers.

Yang et al. [20] introduced comprehensive secure federated learning frameworks, including horizontal federated learning, vertical federated learning and federated transfer learning. In order to apply federated learning to some specific scenarios, Bonawitz et al. [34] constructed the federated learning framework on a platform of mobile phones, and proposed some open issues and future development directions of this federated learning framework.

Apart from the studies mentioned above, federated learning has been successfully applied in a variety of tasks, such as image segmentation [35], [36], [37] and object detection [38], [39], [40], [41]. Chang et al. [35] proposed the Distributed Asynchronized Discriminator GAN (AsynDGAN) for health entity segmentation, which combines federated learning with GAN. A two-layer federated learning framework for object detection tasks was introduced in [42]. These frameworks have been shown to achieve more efficient and accurate learning, while ensuring data privacy protection and reducing communication overhead.

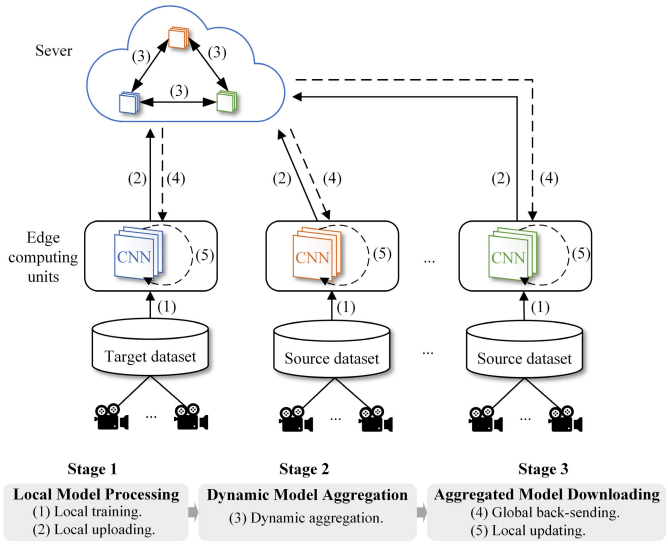

Fig. 1. The proposed federated learning framework for unsupervised domain adaptive vehicle re-ID.

However, federated learning based vehicle re-ID has not been studied yet. Due to the intra-class difference and inter-class similarity of vehicles, training of a vehicle re-ID model usually requires the collection of a large number of images to a cloud central server in order to achieve a high recognition accuracy. This usually costs a lot of computation and communication resources. In addition, because the cameras are installed in different blocks and at different angles, the amount of data they collect, the number of vehicles, lighting and other environmental factors are quite different. It brings about the problem of data heterogeneity between multiple image sets.

The aforementioned problems can be effectively solved by applying federated learning to vehicle re-ID. In federated learning framework, model training is usually performed on edge devices which collect data from local cameras and model ensemble is performed in a central server. In this way, computation can be greatly reduced due to the distributed training mode, and heterogeneous local models can adapt to each other by model ensemble. Thus, it is important to explore the application of federated learning in the field of vehicle re-ID.

III. The Proposed Approach

A. Overall Framework

The proposed vehicle re-ID framework is demonstrated in Fig. 1, which consists of a central cloud server, several edge computing units, and a large number of traffic cameras.

In our framework, multiple sets of cameras in the bottom layer generate raw images as source-domain datasets for training local re-ID models. Each local source dataset can be represented as $SD = \left\{\left(x_i^s, y_i^s\right)|_{i=1}^{N_s}\right\}$, where x_i^s and y_i^s respectively denotes the i-th sample image and the corresponding vehicle identity label in the source domain, N_s is the number of sample images of the source domain dataset. In particular, we consider one of the local image datasets as

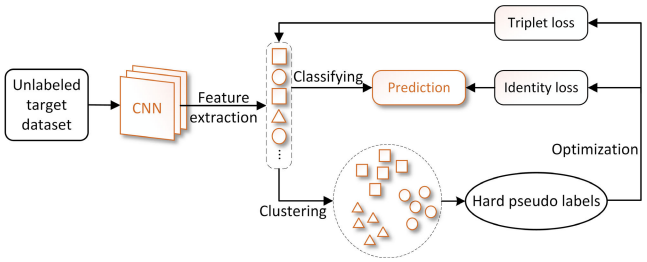

Fig. 2. Illustration of clustering-based pseudo label generation scheme. By performing clustering on the extracted average target features, we can take the cluster IDs as the pseudo target labels to refine the CNN model by optimizing the identity loss and the triplet loss.

the unlabeled target-domain dataset for other re-ID models to learn. The target domain dataset is represented as $TD = \left\{x_i^t|_{i=1}^{N_t}\right\}$, where each target sample image x_i^t has no identity label. In the middle layer of the framework, multiple low-cost edge computing units with limited computation ability are utilized to process vehicle images collected by cameras in order to train local models. At the top of the framework, a central server with strong computing power is used to perform global aggregation for local models to achieve mutual adaptation.

The pipeline of our proposed vehicle re-ID framework consists of three stages:

- **Local model processing**. This stage consists of two steps: local training and local uploading. Initially, local training is performed in the edge computing units to reduce computation. Specifically, heterogeneous local models are respectively pre-trained by corresponding source-domain datasets, and retrained by the target-domain dataset. After local training, each edge computing unit will upload current local model parameters, feature maps and predictions on target-domain samples to the server.
- **Dynamic model aggregation**. In this stage, local models transmitted from these edge computing units are dynamically aggregated in the central server to improve the generalization ability of the model. In order to enhance the overall recognition accuracy during the model aggregation, multiple heterogeneous re-ID models in different domains collaboratively learn from each other and are ensembled in a weighted way.
- **Aggregated model downloading**. This stage consists of two steps: global back-sending and local updating. After the dynamic aggregation, global model parameters will be sent back to edge computing units for augmenting local retraining. Then, each edge computing unit will update its local model parameters according to the received updates from the server, and conduct retraining in the target domain for further adaptation and generalization.

By training according to this pipeline, multiple vehicle re-ID models from different domains will learn from each other and the recognition capability can be improved comprehensively. The algorithm of our proposed method is given in Algorithm 1.

B. Local Training

In our framework, edge computing units handle local training and uploading. This local training encompasses two phases: supervised learning in source domains and unsupervised learning in the target domain. Supervised learning plays a crucial role, utilizing labeled data to train our model and establish a fundamental grasp of data distribution. It equips the model with a precise mapping from input features to expected outputs. Conversely, unsupervised learning, particularly through clustering methods, enables our model to discern the inherent structure in the unlabeled data of the target domain. By integrating these two learning approaches, our methodology benefits from the accuracy of supervised learning while adapting to the target domain's intricacies via unsupervised learning.

(1) Supervised learning in source domains. In this phase of local training, multiple CNN models with different network architectures are pre-trained on the corresponding source datasets in edge computing units. For each local model M^k, a deep neural network is initially trained by the local source-domain dataset. θ^k represents the parameters of model M^k. The pre-trained networks can capture the training data distribution and their class predictions. Thus, each input sample image x_i^s can be transformed into a feature representation $f(x_i^s|\theta^k)$. And the output is the classification probability $p_j(x_i^s|\theta^k)$ which indicates the probability of image x_i^s belonging to the identity j.

In order to suppress unimportant channel features and pay more attention to the channel features with the most information, we add Squeeze-and-Excitation (SE) module [43] to the original CNNs. By controlling the size of the scale, the features of vehicles with higher availability can be retained, and the background features of the image can be weakened, thereby making the extracted features more directional.

A classification loss $\mathcal{L}_{id}^s\left(\theta^k\right)$ and a triplet loss $\mathcal{L}_{tri}^s\left(\theta^k\right)$ are used to optimize the neural network. For classification loss, we adopt cross entropy loss with label smoothing to alleviate the impact brought by wrong annotations, which is defined as

$$L_{id}^s(\theta^k) = \frac{1}{N_s}\sum_{i=1}^{N_s}\sum_{j=1}^{M_s} q_j \log p_j\left(x_i^s \mid \theta^k\right), \tag{1}$$

where $q_j = 1 - \sigma + \frac{\sigma}{M_s}$ if $j = y_i^s$, otherwise $q_j = \frac{\sigma}{M_s}$. M_s is the number of source-domain vehicle identities. σ is a small constant. In this paper, σ is set as 0.1.

Considering that traditional triplet loss cannot support soft-label training of multiple networks, we utilize the softmax-triplet loss [44] which is defined as

$$L_{tri}^s\left(\theta^k\right) = -\frac{1}{N_s}\sum_{i=1}^{N_s}\log \mathcal{P}_i^s\left(\theta^k\right), \tag{2}$$

where

$$\mathcal{P}_i^s\left(\theta^k\right) = \frac{e^{\|f\left(x_i^s|\theta^k\right) - f\left(x_{i-}^s|\theta^k\right)\|}}{e^{\|f\left(x_i^s|\theta^k\right) - f\left(x_{i+}^s|\theta^k\right)\|} + e^{\|f\left(x_i^s|\theta^k\right) - f\left(x_{i-}^s|\theta^k\right)\|}}. \tag{3}$$

Here $\mathcal{P}_i^s(\cdot)$ represents the softmax-distance between negative sample pairs in the source domain. x_{i+}^s and x_{i-}^s respectively denote the hardest positive sample and negative sample of the anchor sample x_i^s. $\|\cdot\|$ represents the L_2 distance. Therefore,

the overall loss can be calculated as

$$\mathcal{L}^s\left(\theta^k\right) = \mathcal{L}^s_{id}\left(\theta^k\right) + \mathcal{L}^s_{tri}\left(\theta^k\right). \tag{4}$$

Through local model pre-training, each local model can well perform feature extraction and vehicle identity recognition on sample images in the source domain.

(2) Unsupervised learning in the target domain In this phase of local training, an unsupervised learning of the local model on the unlabeled target dataset is performed to transfer the model from the source domain to the target domain, thus improving the generalization ability of the local model. We can obtain the feature representation $f(x_i^t|\theta^k)$ and the classification probability $p_j(x_i^t|\theta^k)$ of each target sample image x_i^t by inputting it into the model M^k.

As depicted in Fig. 2, our method utilizes a K-means clustering-based pseudo label generation scheme, chosen for its computational efficiency and simplicity, crucial for large-scale datasets. K-means efficiently handles large data volumes without sacrificing feature representation quality. Note that, we set the number of clusters M_t as 500 based on empirical observations and preliminary tests, optimizing the balance between granularity and computational manageability. The process involves four detailed steps:

- Each local model extracts convolutional features, denoted as $f(x_i^t|\theta^k)$, from each target-domain sample image x_i^t.
- The mean features are computed by averaging the features extracted by all local models. This step ensures a consolidated representation of the features.
- Using the K-means clustering algorithm, each local model clusters the mean features in a mini-batch manner, effectively grouping the target-domain samples into M_t distinct clusters or classes.
- Pseudo-labels, represented as $\widetilde{Y}_t$, are then generated based on the cluster IDs assigned by the K-means algorithm.

C. Dynamic Aggregation

In our framework, the central server is responsible for dynamic aggregation and updates back-sending. Aiming to improve generalization ability of the whole framework, we adopt collaborative learning among multiple models to transfer knowledge from one to another in the target domain. Meanwhile, we introduce two strategies for conducting collaborative learning including historical accumulation and weight regularization, which can respectively produce more reliable pseudo labels and enhance the recognition ability of the whole framework.

1) Historical Accumulation: In order to transfer knowledge from one local model to another, the current class predictions can serve as soft pseudo labels to supervise other models [45]. However, directly using the current class predictions as soft pseudo labels will cause the predictions of different networks to converge to equal each other, which might lead to a reduction of output independences between different networks. Also, it will additionally cause an error amplification if using wrong classification predictions as pseudo labels [44]. To mitigate these issues, we introduce the historical accumulated model for each network, which preserves more of the model's initial knowledge. This approach is used to generate more reliable soft pseudo labels for mutual supervision, as illustrated in Fig. 3.

More specifically, the parameters of the historical accumulated model of local model M^k at the current iteration T are denoted as Θ_T^k, which can be calculated as

$$\Theta_T^k = \lambda\Theta_{T-1}^k + (1-\lambda)\,\theta^k, \tag{5}$$

where Θ_{T-1}^k denotes the historical accumulated parameters of M^k in the previous iteration $(T-1)$, the initial historical accumulated parameters are $\Theta_0^k = \theta^k$ and λ is the scale factor in $[0, 1)$. The predictions obtained by historical accumulated models are more independent, and utilize the historical accumulated model to generate soft pseudo labels can better avoid error amplification during training.

a) Mutual identity loss: For each local model M^k, the soft classification loss for optimizing θ^k with the soft pseudo labels generated by model M^e can be defined as

$$L^t_{mid}\left(\theta^k|\theta^e\right) = \frac{1}{N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{M_t} p_j(x_i^t|\Theta_T^e)\log p_j(x_i^s|\theta^k), \tag{6}$$

where M_t denotes the number of vehicle identities in the target domain, and N_t denotes the total number of target images. The mutual identity loss for the local model M^k can then be defined as the average of all soft classification losses, which can be calculated as

$$\mathcal{L}^t_{mid}\left(\theta^k\right) = \frac{1}{m-1}\sum_{e\neq k}^{m}\mathcal{L}^t_{mid}\left(\theta^k|\theta^e\right), \tag{7}$$

where m is the number of local models.

b) Mutual triplet loss: For each local model M^k, the soft softmax-triplet loss for optimizing θ^k with the soft triplet labels generated by model M^e can be defined as

$$\mathcal{L}^t_{stri}\left(\theta^k|\theta^e\right) = \frac{1}{N_t}\sum_{i=1}^{N_t} L_{bce}(\mathcal{P}_i^t\left(\theta^k\right), \mathcal{P}_i^t\left(\Theta^k\right)), \tag{8}$$

where $\mathcal{P}_i^t(\cdot)$ represents the softmax-distance between negative sample pairs in the target domain and $L_{bce}(\cdot)$ represents the binary cross entropy function.

The mutual triplet loss for the local model M^k also can be defined as the average of all soft softmax-triplet losses, which can be calculated as

$$\mathcal{L}^t_{mtri}\left(\theta^k\right) = \frac{1}{m-1}\sum_{e\neq k}^{m}\mathcal{L}^t_{mtri}\left(\theta^k|\theta^e\right). \tag{9}$$

c) Voting loss: We additionally introduce voting loss including the identity loss and the triplet loss, which can be respectively calculated as

$$\mathcal{L}^t_{id}(\theta^k) = \frac{1}{N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{M_t} q_j\log p_j\left(x_i^t \mid \theta^k\right), \tag{10}$$

$$\mathcal{L}^t_{tri}\left(\theta^k\right) = -\frac{1}{N_t}\sum_{i=1}^{N_t}\log\mathcal{P}_i^t\left(\theta^k\right), \tag{11}$$

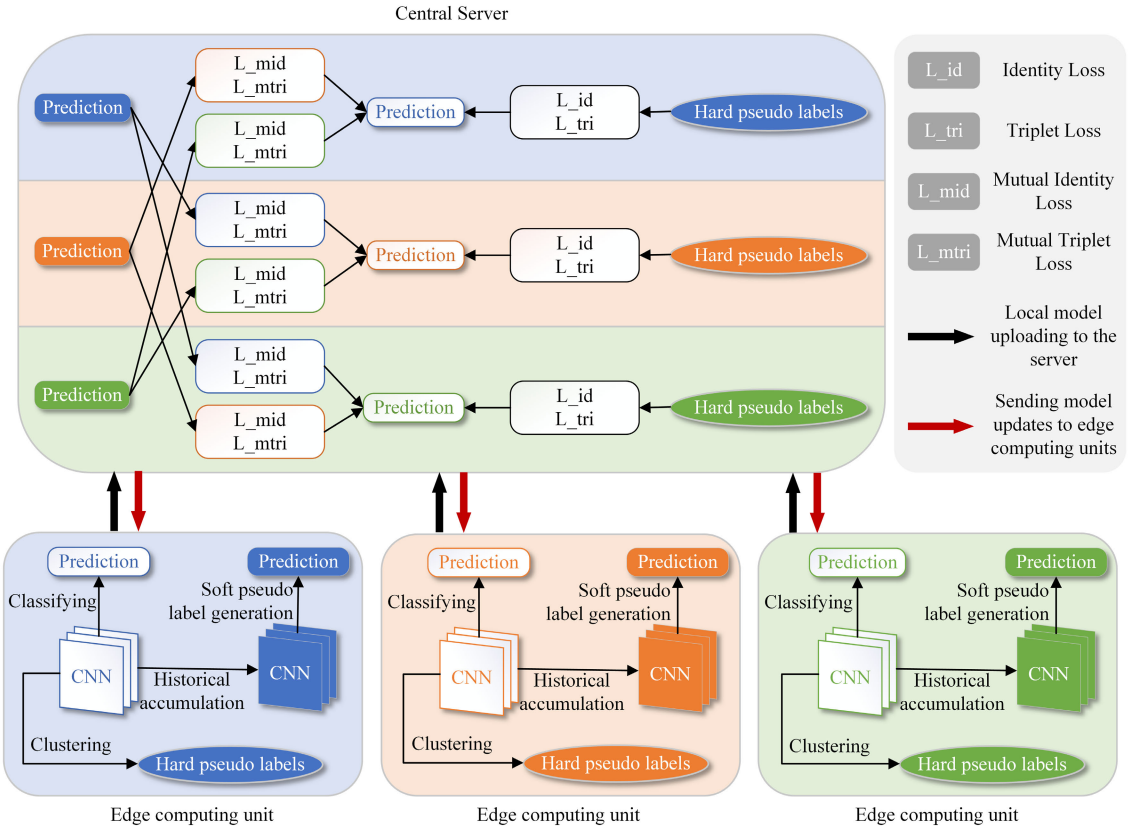

Fig. 3. An illustration of local uploading and dynamic aggregation. Edge computing units upload current model parameters, feature maps and predictions on target domain samples to the server. Multiple local models collaboratively learn from each other in the server to enhance the generalization ability. Meanwhile, soft pseudo labels generated by the historical accumulated model and hard pseudo labels generated by clustering are combined for joint optimization.

where $q_j = 1 - \sigma + \frac{\sigma}{M_t}$ if $j = \widetilde{y}_i^t$, otherwise $q_j = \frac{\sigma}{M_t}$. And $\mathcal{P}_i^t(\cdot)$ denotes the softmax-distance between negative sample pairs in the target domain, which is represented as

$$\mathcal{P}_i^t\left(\theta^k\right) = \frac{e^{\|f(x_i^t|\theta^k) - f(x_{i\_}^t|\theta^k)\|}}{e^{\|f(x_i^t|\theta^k) - f(x_{i+}^t|\theta^k)\|} + e^{\|f(x_i^t|\theta^k) - f(x_{i\_}^t|\theta^k)\|}}. \quad (12)$$

The voting loss can be defined as the sum of the identity loss and the triplet loss, we have

$$\mathcal{L}_{vot}^t(\theta^k) = \mathcal{L}_{id}^t(\theta^k) + \mathcal{L}_{tri}^t(\theta^k). \quad (13)$$

d) Overall loss: The individual loss for each local model can be defined as

$$\mathcal{L}_{one}^t(\theta^k) = \mathcal{L}_{mid}^t(\theta^k) + \mathcal{L}_{mtri}^t(\theta^k) + \mathcal{L}_{vot}^t(\theta^k). \quad (14)$$

Therefore, the overall loss is defined as

$$\mathcal{L}^t = \sum_{k=1}^{m} \mathcal{L}_{one}^t(\theta^k). \quad (15)$$

After obtaining the overall loss of the whole framework, the server sends the updated model parameters back to edge computing units for local updating.

2) Weight Regularization: we propose a weight regularization strategy to accommodate the heterogeneity of multiple local models. In each iteration, individual model extracts target-domain sample features $f(x|\Theta_T)$ and performs a clustering algorithm based on K-means to group all training samples into M_t clusters as $\mathbb{C}$. Since the discrimination capability can be indicated by the clustering results, the weight of each local model can be defined as the ratio of the inter-cluster scatter and the sum of intra-cluster scatters.

The intra-cluster scatter of the i-th cluster $\mathbb{C}_i$ can be calculated as

$$S_{intra}^i = \sum_{x \in \mathbb{C}_i} \| f(x|\Theta_T) - \mu_i \|^2, \quad (16)$$

where $\mu_i = \sum_{x \in \mathbb{C}_i} f(x|\Theta_T)/n_t^i$ indicates the mean feature of the cluster $\mathbb{C}_i$ in the target domain, and n_t^i denotes the number of features in $\mathbb{C}_i$. The inter-cluster scatter can be calculated as

$$S_{inter} = \sum_{i=1}^{M_t} n_t^i \| \mu_i - \mu \|^2, \quad (17)$$

where $\mu = \sum_{i=1}^{N_t} f(x_{t,i}|\Theta_T)/N_t$ indicates the mean feature of all training target-domain samples. The ratio of the inter-cluster scatter and the sum of intra-cluster scatters R then can be defined as

$$R = \frac{S_{inter}}{\sum_{i=1}^{M_t} S_{intra}^i}. \quad (18)$$

The model with a better discrimination capability has a larger scalar ratio R due to the larger inter-cluster scatter or the smaller intra-cluster scatter compared with other models.

For dynamic aggregation in the main server, the weight w^k of local model M^k can be defined as the mean normalization of R, shown as

$$w^k = \frac{MR^k}{\sum_{i=1}^{M} R^i}. \tag{19}$$

Then, we can respectively re-define the mutual identity loss in Eq. (7) and the mutual triplet loss in Eq. (9) as

$$\mathcal{L}_{mid}^t\left(\theta^k\right) = \frac{1}{m-1}\sum_{e\neq k}^{m} w^k \mathcal{L}_{mid}^t\left(\theta^k|\theta^e\right), \tag{20}$$

and

$$\mathcal{L}_{mtri}^t\left(\theta^k\right) = \frac{1}{m-1}\sum_{e\neq k}^{m} w^k \mathcal{L}_{mtri}^t\left(\theta^k|\theta^e\right). \tag{21}$$

In this way, models with a better discrimination capability will play more important roles in the current iteration, which enhances the aggregation and the adaptation among multiple local models.

Algorithm 1 Federated-Learning-Based Unsupervised Domain Adaptive Vehicle Re-ID

Require: The number of local models, K; source domain datasets, $SD = \{X_s, Y_s\}$; target domain dataset, $TD = \{X_t\}$; maximum iterations T;
Ensure: Optimized vehicle re-ID models $\{M^k\}$.

```
Pre-train K local models on source datasets.
for t = 1 to T do
    for each edge computing unit r_j ∈ R do
        Extract target features f_t by model M^k with θ^k.
        Generate hard pseudo labels Ỹ_t by clustering.
        Calculate classifying predictions p(X_t|θ^k) by local model M^k with θ^k.
        Generate soft labels p(X_t|Θ^k) from historical accumulated model with Θ^k_T.
        Upload model parameters θ^k, predictions p(X_t|θ^k) and features f_t to the server.
    end for
    Evaluate the weight {w^k} of each local model via Eq. (19) in the server.
    Update parameters {θ^k} by optimizing Eq. (15) with {w^k}.
    Send back updated parameters {θ^k} to edge computing units.
end for
return Optimized vehicle re-ID models {M^k}.
```

IV. Experiment

A. Datasets

In order to validate the effectiveness of our proposed framework, we use two vehicle re-ID benchmark datasets:

1) VeRi-776 [46]: This dataset comprises 50,000 images of 776 vehicles captured by 20 surveillance cameras in various directions and angles in a real traffic network. Each image is thoroughly annotated with details like vehicle identity, type, color, location, time, and camera tag. For training, 37,781 images of 576 vehicles are used, while testing involves 11,579 images of 200 different vehicles. Additionally, 1,678 images from the test set are randomly selected as the query set.

2) VehicleID [47]: This dataset, taken from a city surveillance system during the day, contains 221,763 images of 26,267 vehicles, with 90,196 images of 10,319 vehicles labeled with model information. The training set includes 11,585 images of 13,134 vehicles, and the testing set has 111,585 images of 13,133 vehicles. To manage the large test data volume, four subsets are created: "small" (800 vehicles), "medium" (1,600 vehicles), "large" (2,400 vehicles), and "huge" (3,200 vehicles). Each subset's test comprises one randomly chosen gallery image per vehicle and the remaining images as the query set.

TABLE I
Results on VeRi-776

Methods	mAP	Rank-1(%)	Rank-5(%)
VehicleNet [48]	83.4	96.7	-
FastReID [49]	81.9	97.0	99.0
TBE-Net [50]	79.5	96.0	98.5
FACT [5]	18.7	52.1	72.6
Mixed Diff+CCL [47]	27.7	60.8	77.7
Cycle GAN [30]	21.8	55.4	66.6
OIFE+STR [51]	51.4	68.3	89.7
PUL [52]	17.0	55.2	66.2
UDAR [16]	35.8	76.9	85.8
PROVID [7]	53.4	81.5	95.1
VAMI [53]	50.1	77.0	90.8
NuFACT [7]	48.4	76.7	91.4
SDC-CNN [54]	53.4	83.4	92.5
STP [55]	53.3	82.0	92.3
DHMVI [56]	25.1	60.8	78.5
VRIC [57]	49.3	88.56	-
VR-PROUD [14]	22.1	55.7	70.0
FDA-Net [58]	55.4	84.2	92.4
View-EALN [32]	50.3	81.3	90.8
PAL [18]	41.5	68.0	79.8
Direct Transfer	18.9	36.0	48.1
Baseline System	31.4	58.2	73.1
Ours	**56.1**	**70.0**	**81.1**

B. Experimental Details

We respectively take two datasets as source and target domains for training: In Table I, we use VehicleID as the source dataset and VeRi-776 as the target dataset; in Table II, we use VeRi-776 as the source dataset and VehicleID as the domain dataset.

During the pre-training phase, we strategically selected three distinct backbone architectures for our branch networks based on their unique strengths and capabilities. Specifically, we chose DenseNet-121 for its efficient parameter utilization and feature reuse capabilities, ResNet-50 for its proven robustness and ability to address the vanishing gradient problem through residual connections, and Inception-v3 for its multi-scale feature extraction and computational efficiency. Initially, image features are extracted by backbone networks fused with the attention mechanism. In the training duration, we resize the input image to 256×128, and perform traditional image augmentation including random flipping and random erasing. For each identity from the training set, a mini-batch is sampled with selected identities and randomly sampled images to compute the hard batch triplet loss. To optimize

parameters, the Adam with a weight decay of 0.0005 is used. We set the initial learning rate as 0.00035. In the total 80 epochs, the learning rate is decreased to 1/10 of its previous value on the 40th and 70th epoch.

In the training of the target domain, there are 100 iterative training epochs in total and the learning rate is consistently set as 0.00035. Each epoch consists of 800 training iterations, and in each epoch the algorithm conducts a mini-batch clustering based on 500 clusters. In addition, we use ResNet-50 network to extract features for testing.

For the performance evaluation on our proposed method, three metrics are adopted: **mean Average Precision** (mAP), **rank-1 accuracy** and **rank-5 accuracy**. MAP evaluates the ability of the model to both correctly identify vehicles (precision) and to detect as many true instances of the vehicle as possible (recall) across varying confidence thresholds, providing an aggregated measure of precision and recall over all possible levels of confidence. Rank-1 indicates the proportion of times the true match is ranked first in the list of candidate matches, providing a focused measure on the most relevant match. Rank-5 assesses the proportion of times the true match is within the top 5 ranked candidate matches, providing a broader perspective on the model's performance.

C. Results and Analysis

In this subsection, the proposed algorithm is compared with existing vehicle re-ID methods including: VehicleNet [48], FastReID [49], TBE-Net [50], FACT [5], Mixed Diff+CCL [47], Cycle GAN [30], OIFE+STR [51], PUL [52], UDAR [16], PROVID [7], VAMI [53], NuFACT [7], SDC-CNN [54], STP [55], DHMVI [56], VRIC [57], VR-PROUD [14], FDA-Net [58], View-EALN [32], PAL [18], Direct Transfer and Baseline System. Specifically, Direct Transfer denotes the method that adopts directly testing in the target domain after training in source domains, and Baseline System denotes the method that merely generates pseudo labels by clustering after feature extraction to train the re-ID model. For unsupervised domain adaptive vehicle re-ID, most algorithms need samples with special annotations during the training, such as coordinates of special marks in images or split lines of different targets, etc. Therefore, we only compare the proposed method with some representative methods.

1) Results on VeRi-776: The test results of all compared algorithms on the VeRi-776 dataset are listed in Table I. Note that, VehicleNet, FastReID and TBE-Net are supervised methods. These methods are expected to perform better since they are trained with ground truth labels. Our approach notably excels in mAP, reaching 56.1%, which surpasses all other unsupervised and domain adaptation methods evaluated. This superior mAP underscores our method's effectiveness in accurately identifying the correct vehicles throughout the dataset. However, our Rank-1 and Rank-5 accuracies are modest in comparison to certain methods like PROVID, SDC-CNN, and VRIC. These methods may benefit from advanced feature extraction techniques that finely capture vehicle appearance nuances, essential for achieving high accuracy in top-ranked identifications. In contrast, our method utilizes a streamlined CNN architecture for feature extraction. Despite the marginally lower Rank-1 and Rank-5 results, this choice highlights our method's operational efficiency and practicality. The robust mAP performance of our framework attests to its reliability and the balanced precision-recall trade-off it offers, reinforcing its potential for real-world application where computational simplicity is as valued as accuracy.

Fig. 4. Examples of re-identification ranking results on VeRi-776.

2) Results on VehicleID: The comparative performance of various algorithms on the VehicleID dataset across different test sizes is detailed in Table II. Our method demonstrates superior performance, achieving an mAP of 60.0%, 56.6%, 53.4%, and 50.3%, and Rank-1 accuracies of 53.8%, 50.0%, 45.6%, and 41.4% for test sizes of 800, 1600, 2400, and 3200, respectively. The PUL approach employs k-means clustering to assign pseudo-labels to unlabeled samples [52]. However, this method grapples with the challenge of determining an optimal cluster count, which can result in misclassifications or unassigned samples. Our method circumvents this issue by setting a predetermined cluster number, thereby reducing the likelihood of such errors and enhancing pseudo-label accuracy. While the PAL method also leverages a pseudo-labeling strategy for unsupervised training, it relies solely on a single backbone network, ResNet-50 [18]. In contrast, our approach harnesses the collective strength of multiple heterogeneous networks. By utilizing the mean features from unlabeled samples extracted through these diverse networks as pseudo-labels, and by implementing inter-network mutual supervision, our method significantly refines the accuracy of pseudo-labels, leading to improved overall performance.

D. Ablation Studies

In order to verify the effectiveness of each module in the proposed algorithm, we conduct extensive ablation studies on VeRi-776 and VehicleID. The results are summarized in Tables III and IV.

TABLE II
RESULTS ON VEHICLEID

Methods	size=800(%)			size=1600(%)			size=2400(%)			size=3200(%)		
	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5
FACT [5]	-	49.53	67.96	-	44.63	64.18	-	39.91	60.49	-	-	-
Mixed Diff+CCL [47]	-	49.00	73.50	-	42.80	66.80	-	38.20	61.60	-	-	-
PUL [52]	43.90	40.03	56.03	37.68	33.83	49.72	34.71	30.90	47.18	32.44	28.86	43.41
Cycle GAN [30]	42.32	37.29	58.56	34.92	30.00	49.72	31.89	27.15	46.52	29.17	24.83	42.17
PAL [18]	53.50	50.25	64.91	48.05	44.25	60.95	45.14	41.08	59.12	42.13	38.19	55.32
Direct Transfer	40.58	35.48	57.26	33.59	28.86	48.34	30.50	26.08	44.02	27.90	23.85	39.76
Baseline System	42.96	39.11	55.24	38.03	34.04	50.91	34.04	30.10	48.41	31.98	28.24	43.77
Ours	**60.0**	**53.8**	**66.4**	**56.6**	**50.0**	**65.9**	**53.4**	**45.6**	**63.0**	**50.3**	**41.4**	**60.5**

Fig. 5. Examples of re-identification ranking results on VehicleID.

TABLE III
RESULTS OF ABLATION STUDY ON VERI-776

Methods	mAP	Rank-1(%)	Rank-5(%)
Supervised Method	67.3	71.4	85.2
Direct Transfer	18.9	36.0	48.1
Baseline	40.2	45.9	49.9
Baseline + MIL + MTL + WR	40.5	46.3	50.1
Baseline + HA + MTL + WR	44.2	49.4	52.1
Baseline + HA + MIL + WR	43.2	47.7	51.7
Baseline + HA + MIL + MTL	47.8	50.6	53.3
Baseline + HA + CL	54.3	71.4	80.6
The Whole Network	**56.1**	**69.3**	**81.1**

Supervised Method is to perform supervised learning in the target domain and then test on target samples. **Direct Transfer** is to directly utilize the model trained in the source domain to test on target samples. The above mentioned methods respectively define the upper performance limit and the lower performance limit. As we can see from the tables, there is a large performance gap between the direct transfer model and supervised models.

We take the proposed model using only voting loss as a baseline model. Specifically, the initial training of the model is firstly performed on each edge computing unit. And then the obtained model parameters and the target predictions are sent to the central server for dynamic aggregation. After receiving the integrated parameters from the server, each local model conducts self-optimization by voting loss. As shown in Tables III and IV, the baseline model outperforms the direct transfer model by a large margin. This shows that the decision loss can effectively utilize the ensemble model to predict more accurate pseudo-labels and fine-tune each network.

TABLE IV
RESULTS OF ABLATION STUDY ON VEHICLEID

Methods	mAP	Rank-1(%)	Rank-5(%)
Supervised Method	71.7	68.3	80.9
Direct Transfer	30.5	26.1	44.0
Baseline	41.9	35.4	50.1
Baseline + MIL + MTL + WR	42.1	35.6	50.4
Baseline + HA + MTL + WR	43.2	36.1	51.6
Baseline + HA + MIL + WR	42.9	35.9	51.4
Baseline + HA + MIL + MTL	48.3	37.2	53.1
Baseline + HA + CL	58.1	54.1	65.5
The Whole Network	**59.7**	**53.8**	**66.3**

HA represents historical accumulation. **MIL** and **MTL** respectively represent the mutual identity loss and the mutual triplet loss. **WR** represents the weight regularization. **CL** represents circle loss [59]. In the proposed method, we use historical accumulated models to produce soft pseudo labels for supervising other models, so as to avoid the network error amplification and to preserve the historical knowledge of local models. In ablation studies, we directly use the current model to produce predictions as soft pseudo labels. As shown in Tables III and IV, without using the historical accumulation strategy, the model has a certain degree of degradation in mAP, Rank-1 and Rank-5 on both datasets. It can be concluded that without using the historical accumulated model, the network tends to degenerate into uniformity, which greatly reduces the learning ability. Furthermore, we have investigated the application of circle loss in our experiments, attracted by its flexible and direct approach to optimizing feature space distribution. However, our empirical findings indicate that the loss functions originally designed for our framework yield a marginally superior performance compared to circle loss. Additionally, we observed that the convergence rate when employing circle loss was significantly slower than that achieved with our original loss functions. This suggests that while circle loss has its merits, our tailored loss functions are better suited to our specific federated learning framework for vehicle re-identification, both in terms of performance and computational efficiency.

To validate the effectiveness of dynamic aggregation in the server, we respectively remove the mutual identity loss and

the mutual triplet loss. The results show that the mAP on Veri776 dataset drops from 76.0% to 70.2%, and the mAP on VehicleID dataset drops from 66.1% to 60.4%. A similar drop can also be observed when studying the mutual triplet loss. Therefore, we argue that the above two mutual losses can improve the effectiveness in mutual learning. In addition, we dismiss the weight regularization strategy and set the weights of all models to 1 during dynamic aggregation. The results show that the model performances on VehicleID and VeRi-776 both drop significantly, which indicates that the aggregated parameters sent by the central server affect the identification of features to a certain extent.

V. Conclusion

In this paper, we presented an federated unsupervised domain adaptive vehicle re-ID framework utilizing a three-stage process: local model processing, dynamic model aggregation and aggregated model downloading. Through the innovative use of pseudo labels generated from averaged features across multiple local models, we effectively utilize unlabeled target data, a common obstacle in domain adaptation tasks. The introduction of historical accumulation ensures that valuable knowledge is not discarded but rather harnessed to guide the adaptation process. Furthermore, our weight regularization strategy for model aggregation refines the discriminative power of the system, leading to a more robust and accurate vehicle re-ID framework. Our experimental results affirm the efficacy of our framework in amplifying the discriminative capacity of vehicle re-ID models. Notably, given the framework's generic design, it holds the potential to be applied to other domains, such as pedestrian [60], [61] or object re-ID [62], [63], a prospect we find intriguing and worth exploring in future studies. Building on this foundation, our future endeavors will focus on exploring more intricate network architectures and aggregation methods to propel our framework into practical, real-world applications.

References

[1] Y. Hui, Z. Su, and T. H. Luan, "Unmanned era: A service response framework in smart city," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5791–5805, Jun. 2022, doi: 10.1109/TITS.2021.3058385.

[2] S. Zherzdev and A. Gruzdev, "LPRNet: License plate recognition via deep neural networks," 2018, *arXiv:1806.10447*.

[3] R. Laroca et al., "A robust real-time automatic license plate recognition based on the YOLO detector," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–10, doi: 10.1109/IJCNN.2018.8489629.

[4] S. M. Silva and C. R. Jung, "License plate detection and recognition in unconstrained scenarios," in *Computer Vision–(ECCV)*. Cham, Switzerland: Springer, 2018, pp. 593–609, doi: 10.1007/978-3-030-01258-8_36.

[5] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2016, pp. 1–6, doi: 10.1109/ICME.2016.7553002.

[6] D. Zapletal and A. Herout, "Vehicle re-identification for automatic video traffic surveillance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 1568–1574, doi: 10.1109/CVPRW.2016.195.

[7] X. Liu, W. Liu, T. Mei, and H. Ma, "PROVID: Progressive and multimodal vehicle reidentification for large-scale urban surveillance," *IEEE Trans. Multimedia*, vol. 20, no. 3, pp. 645–658, Mar. 2018, doi: 10.1109/TMM.2017.2751966.

[8] H. Guo, C. Zhao, Z. Liu, J. Wang, and H. Lu, "Learning coarse-to-fine structured feature embedding for vehicle re-identification," in *Proc. 32nd AAAI Conf. Artif. Intell. 13th Innov. Appl. Artif. Intell. Conf. 8th AAAI Symp. Educ. Adv. Artif. Intell.*, 2018, doi: 10.1609/aaai.v32i1.12237.

[9] H. Chen, B. Lagadec, and F. Brémond, "Partition and reunion: A two-branch neural network for vehicle re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Long Beach, CA, USA, Jun. 2019, pp. 184–192.

[10] A. Suprem and C. Pu, "Looking GLAMORous: Vehicle re-id in heterogeneous cameras networks with global and local attention," 2020, *arXiv:2002.02256*.

[11] X. Lin et al., "Aggregating global and local visual representation for vehicle re-IDentification," *IEEE Trans. Multimedia*, vol. 23, pp. 3968–3977, 2021, doi: 10.1109/TMM.2020.3035279.

[12] X. Chen, H. Sui, J. Fang, W. Feng, and M. Zhou, "Vehicle re-identification using distance-based global and partial multi-regional feature learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1276–1286, Feb. 2021, doi: 10.1109/TITS.2020.2968517.

[13] R. M. S. Bashir, M. Shahzad, and M. Fraz, "DUPL-VR: Deep unsupervised progressive learning for vehicle re-identification," in *Proc. 13th Int. Symp. Visual Comput. (ISVC)*, Las Vegas, NV, USA, 2018, pp. 286–295, doi: 10.1007/978-3-030-03801-4_26.

[14] R. M. S. Bashir, M. Shahzad, and M. M. Fraz, "VR-PROUD: Vehicle re-identification using progressive unsupervised deep architecture," *Pattern Recognit.*, vol. 90, pp. 52–65, Jun. 2019, doi: 10.1016/j.patcog.2019.01.008.

[15] A. Zheng, X. Sun, C. Li, and J. Tang, "Viewpoint-aware progressive clustering for unsupervised vehicle re-identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11422–11435, Aug. 2022, doi: 10.1109/TITS.2021.3103961.

[16] L. Song et al., "Unsupervised domain adaptive re-identification: Theory and practice," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107173, doi: 10.1016/j.patcog.2019.107173.

[17] Y. Wang and D. Zeng, "Deep domain adaptation on vehicle re-identification," in *Proc. IEEE 5th Int. Conf. Multimedia Big Data (BigMM)*, Sep. 2019, pp. 416–420, doi: 10.1109/BIGMM.2019.00024.

[18] J. Peng, Y. Wang, H. Wang, Z. Zhang, X. Fu, and M. Wang, "Unsupervised vehicle re-identification with progressive adaptation," in *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2020, pp. 913–919.

[19] Y. Huang et al., "Dual domain multi-task model for vehicle re-identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 2991–2999, Apr. 2022, doi: 10.1109/TITS.2020.3027578.

[20] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[21] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2017, pp. 1175–1191, doi: 10.1145/3133956.3133982.

[22] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[23] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1116–1124, doi: 10.1109/ICCV.2015.133.

[24] L. Zheng, S. Wang, W. Zhou, and Q. Tian, "Bayes merging of multiple vocabularies for scalable image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1963–1970, doi: 10.1109/CVPR.2014.252.

[25] Y. F. Guo, L. Wu, L. Hong, F. Zhe, and X. Xue, "Null foley-sammon transform," *Pattern Recognit.*, vol. 39, no. 11, pp. 2248–2251, 2006, doi: 10.1016/j.patcog.2006.05.009.

[26] J. Zhu et al., "Vehicle re-identification using quadruple directional deep learning features," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 410–420, Jan. 2020, doi: 10.1109/TITS.2019.2901312.

[27] R. Chu, Y. Sun, Y. Li, Z. Liu, C. Zhang, and Y. Wei, "Vehicle re-identification with viewpoint-aware metric learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8281–8290, doi: 10.1109/ICCV.2019.00837.

[28] P. A. Marín-Reyes, L. Bergamini, J. Lorenzo-Navarro, A. Palazzi, S. Calderara, and R. Cucchiara, "Unsupervised vehicle re-identification using triplet networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 166–171.

[29] I. Goodfellow et al., "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, Jun. 2014, doi: 10.1145/3422622.
[30] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.
[31] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. C. Courville, "Augmented cycleGAN: Learning many-to-many mappings from unpaired data," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, vol. 80, pp. 195–204, 2018.
[32] Y. Lou, Y. Bai, J. Liu, S. Wang, and L. Duan, "Embedding adversarial learning for vehicle re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3794–3807, Aug. 2019, doi: 10.1109/TIP.2019.2902112.
[33] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," 2016, *arXiv:1602.05629*.
[34] K. Bonawitz et al., "Towards federated learning at scale: System design," 2019, *arXiv:1902.01046*.
[35] Q. Chang et al., "Synthetic learning: Learn from distributed asynchronized discriminator GAN without sharing medical image data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13853–13863, doi: 10.1109/CVPR42600.2020.01387.
[36] D. Li, A. Kar, N. Ravikumar, A. F. Frangi, and S. Fidler, "Fedsim: Federated simulation for medical imaging," 2020, *arXiv:2009.00668*.
[37] W. Li et al., "Privacy-preserving federated brain tumour segmentation," in *Proc. 10th Int. Workshop Mach. Learn. Med. Imag.*, Shenzhen, China, Berlin, Germany: Springer-Verlag, 2019, pp. 133–141.
[38] P. Yu and Y. Liu, "Federated object detection: Optimizing object detection model with federated learning," in *Proc. 3rd Int. Conf. Vis., Image Signal Process.*, New York, NY, USA, 2020, doi: 10.1145/3387168.3387181.
[39] X. Kong et al., "A federated learning-based license plate recognition scheme for 5G-enabled Internet of Vehicles," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 8523–8530, Dec. 2021, doi: 10.1109/TII.2021.3067324.
[40] X. Huang, P. Li, R. Yu, Y. Wu, K. Xie, and S. Xie, "FedParking: A federated learning based parking space estimation with parked vehicle assisted edge computing," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9355–9368, Sep. 2021, doi: 10.1109/TVT.2021.3098170.
[41] C. Ma, J. Zhu, M. Liu, H. Zhao, N. Liu, and X. Zou, "Parking edge computing: Parked-vehicle-assisted task offloading for urban VANETs," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9344–9358, Jun. 2021, doi: 10.1109/JIOT.2021.3056396.
[42] X. Zhou, W. Liang, J. She, Z. Yan, and K. I.-K. Wang, "Two-layer federated learning with heterogeneous model aggregation for 6G supported Internet of Vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5308–5317, Jun. 2021, doi: 10.1109/TVT.2021.3077893.
[43] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.
[44] Y. Ge, D. Chen, and H. Li, "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification," 2020, *arXiv:2001.01526*, doi: 10.48550/arXiv.2001.01526.
[45] Y. Zhai, Q. Ye, S. Lu, M. Jia, R. Ji, and Y. Tian, "Multiple expert brainstorming for domain adaptive person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 594–611, doi: 10.1007/978-3-030-58571-6_35.
[46] X. Liu, W. Liu, T. Mei, and H. Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *Proc. Eur. Conf. Comput. Vis.*, vol. 9906, Nov. 2016, pp. 869–884, doi: 10.1007/978-3-319-46475-6_53.
[47] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2167–2175, doi: 10.1109/CVPR.2016.238.
[48] Z. Zheng, T. Ruan, Y. Wei, Y. Yang, and T. Mei, "VehicleNet: Learning robust visual representation for vehicle re-identification," *IEEE Trans. Multimedia*, vol. 23, pp. 2683–2693, 2021, doi: 10.1109/TMM.2020.3014488.
[49] L. He, X. Liao, W. Liu, X. Liu, P. Cheng, and T. Mei, "FastReID: A PyTorch toolbox for general instance re-identification," 2020, *arXiv:2006.02631*.
[50] W. Sun, G. Dai, X. Zhang, X. He, and X. Chen, "TBE-Net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 14557–14569, Sep. 2022, doi: 10.1109/TITS.2021.3130403.
[51] Z. Wang et al., "Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 379–387, doi: 10.1109/ICCV.2017.49.
[52] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 4, p. 83, Oct. 2018, doi: 10.1145/3243316.
[53] Y. Zhouy and L. Shao, "Viewpoint-aware attentive multi-view inference for vehicle re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6489–6498, doi: 10.1109/CVPR.2018.00679.
[54] J. Zhu et al., "A shortly and densely connected convolutional neural network for vehicle re-identification," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3285–3290, doi: 10.1109/ICPR.2018.8545514.
[55] C.-W. Wu, C.-T. Liu, C.-E. Chiang, W.-C. Tu, and S.-Y. Chien, "Vehicle re-identification with the space-time prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, pp. 121–128, 2018, doi: 10.1109/CVPRW.2018.00024.
[56] Y. Zhou, L. Liu, and L. Shao, "Vehicle re-identification by deep hidden multi-view inference," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3275–3287, Jul. 2018, doi: 10.1109/TIP.2018.2819820.
[57] A. Kanacı, X. Zhu, and S. Gong, "Vehicle re-identification in context," in *Proc. 40th German Conf. Pattern Recognit. (GCPR)*, Stuttgart, Germany: Springer, 2018, pp. 377–390.
[58] Y. Lou, Y. Bai, J. Liu, S. Wang, and L. Duan, "VERI-Wild: A large dataset and a new method for vehicle re-identification in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3235–3243, doi: 10.1109/CVPR.2019.00335.
[59] Y. Sun et al., "Circle loss: A unified perspective of pair similarity optimization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 6398–6407, doi: 10.1109/CVPR42600.2020.00643.
[60] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2872–2893, Jun. 2022, doi: 10.1109/TPAMI.2021.3054775.
[61] W. Zhuang et al., "Performance optimization of federated person re-identification via benchmark analysis," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 955–963, doi: 10.1145/3394171.3413814.
[62] H. Zhu, W. Ke, D. Li, J. Liu, L. Tian, and Y. Shan, "Dual cross-attention learning for fine-grained visual categorization and object re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4692–4702.
[63] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "TransReID: Transformer-based object re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 15013–15022.

Xiao Xiao (Member, IEEE) received the bachelor's degree in control technology and instrumentation from Xidian University, China, in 2004, and the Ph.D. degree in measuring and testing technologies and instruments from Zhejiang University, Hangzhou, China, in 2009. He joined the Min H. Kao Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, as a Research Fellow, in 2019. He is currently an Associate Professor with the School of Telecommunication Engineering, Xidian University. His research interests include wireless network architecture, intelligent transportation systems, computer vision, and deep learning.

Xinyue Yang received the bachelor's degree in communication engineering from the School of Computer and Information Engineering, Hubei University. She is currently pursuing the master's degree in electronic information with the School of Electronic and Communication Engineering, Xidian University. Her research interests include computer vision.

Yucheng Wang received the B.Sc. degree from Nanchang University in 2009 and the dual Ph.D. degree from Beijing Institute of Technology in 2016 and the University of Technology Sydney in 2017. His research interests include computer vision applications in autonomous driving.

Yilong Hui (Member, IEEE) received the Ph.D. degree in control theory and control engineering from Shanghai University, Shanghai, China, in 2018. He is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China. He is also a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He has authored or coauthored more than 70 scientific papers with leading journals and international conferences. His research interests include mobile edge computing, digital twins, space-air-ground integrated networks, and autonomous driving. He was a recipient of the Best Paper Award of International Conference WiCon2016 and the IEEE Cyber-SciTech2017.

Jianchuan Zhou received the bachelor's degree in communication engineering from the School of Communication Engineering, Jilin University, and the master's degree in electronic and communication engineering from the School of Electronic and Communication Engineering, Xidian University. His research interests include vehicle recognition and re-identification.

Guoqiang Mao (Fellow, IEEE) has authored or coauthored more than 200 papers in international conferences and journals, which have been cited more than 9000 times. His research interests include intelligent transport systems, applied graph theory and its applications in telecommunications, the Internet of Things, wireless sensor networks, wireless localization techniques, and network modeling and performance analysis. He received the Top Editor Award for outstanding contributions to IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2011, 2014, and 2015. He was the Co-Chair of the IEEE Intelligent Transport Systems Society Technical Committee on Communication Networks. He has been an Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS since 2018, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2014–2019), and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY since 2010.