



ARTICLE

# Person Re-Identification with Model-Contrastive Federated Learning in Edge-Cloud Environment

Baixuan Tang<sup>1,2,#</sup>, Xiaolong Xu<sup>1,2,#</sup>, Fei Dai<sup>3</sup> and Song Wang<sup>4,\*</sup>

<sup>1</sup>School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China

<sup>2</sup>Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, Nanjing, 210044, China

<sup>3</sup>School of Big Data and Intelligence Engineering, Southwest Forestry University, Kunming, 650233, China

<sup>4</sup>Nanjing Meteorological Service Center, Nanjing Meteorological Bureau, Nanjing, 210019, China

\*Corresponding Author: Song Wang. Email: wsos2@sina.com

#Baixuan Tang and Xiaolong Xu are co-first authors to this work

Received: 10 October 2022 Accepted: 06 January 2023 Published: 26 January 2024

## ABSTRACT

Person re-identification (ReID) aims to recognize the same person in multiple images from different camera views. Training person ReID models are time-consuming and resource-intensive; thus, cloud computing is an appropriate model training solution. However, the required massive personal data for training contain private information with a significant risk of data leakage in cloud environments, leading to significant communication overheads. This paper proposes a federated person ReID method with model-contrastive learning (MOON) in an edge-cloud environment, named FRM. Specifically, based on federated partial averaging, MOON warmup is added to correct the local training of individual edge servers and improve the model's effectiveness by calculating and back-propagating a model-contrastive loss, which represents the similarity between local and global models. In addition, we propose a lightweight person ReID network, named multi-branch combined depth space network (MB-CDNet), to reduce the computing resource usage of the edge device when training and testing the person ReID model. MB-CDNet is a multi-branch version of combined depth space network (CDNet). We add a part branch and a global branch on the basis of CDNet and introduce an attention pyramid to improve the performance of the model. The experimental results on open-access person ReID datasets demonstrate that FRM achieves better performance than existing baseline.

## KEYWORDS

Person re-identification; federated learning; contrastive learning

## 1 Introduction

With the increasing demands for improved public safety, the number of surveillance cameras has increased greatly, and person re-identification (ReID), a deep learning task involving tracking individuals across multiple cameras [1], has become increasingly important. At present, person ReID approaches are based on images [2], videos, or even text, with an image-based person ReID model presented in this study.



Given that training person ReID models are both time-consuming and resource-intensive, cloud computing is an appropriate solution for model training; in this approach, all the personal data used for training are transmitted to a resource-rich cloud server. However, the transmission of massive volumes of personal data to remote cloud servers for training has numerous issues. In particular, given the risks of data leakage and increasingly stringent data privacy regulations, person ReID models, which require huge amounts of personal data for training, cannot directly use the data stored across many edge servers for training. In addition, the mainstream person ReID model has a large number of parameters, which is a challenge to edge devices with limited computational power.

Edge computing is a form of distributed computing that works in cooperation with cloud computing by migrating the data computation from central cloud servers to edge servers of networks located closer to the end devices. Edge computing has been widely used in fields such as traffic flow prediction [3], service requirement prediction [4], microservice [5], IoV task offloading [6], and data caching optimization [7]. In an edge-cloud environment, federated learning can be used as a distributed training framework for deep learning. By using this approach, personal data privacy can be protected by training the person ReID model locally on edge servers and uploading the person ReID model weights instead of the raw data to the cloud server for federated aggregation [8]. In addition, due to the characteristics of distributed training, each edge server achieves better model performance and avoids the significant communication overhead associated with high-volume data uploads.

Data collected from different edge servers suffer from non-identity and independent distribution (non-IID); the data are collected from different cameras in different locations and at different times, which causes variability in the number of images, the number of identities, and data distributions in the data collection environment. Data with non-IID seriously affect the federated learning performance [9,10], representing the primary issue affecting federated person ReID models. However, to address this issue, the model-contrastive federated learning (MOON) [11] approach aims to correct the local training of individual edge servers and improve the model effectiveness by calculating the similarity between various model representations. In contrast to traditional federation learning, when calculating the person ReID model loss in the local training process, in addition to the cross-entropy loss of the person ReID model itself, the comparison loss is also added; this parameter corrects the local updates by maximizing the representation agreement learned by the current local model and the representation learned by the global model.

Based on these observations, in this study, we propose a federated person ReID with MOON warmup in an edge-cloud environment model, named FRM. Specifically, through applying federated partial averaging (FedPav) [12], the model-contrastive loss is added to correct the local training of individual edge servers and improve model effectiveness by calculating and back-propagating the model-contrastive loss, which represents the similarity between the local model, the global model, and the local model in the previous step. In addition, MOON warmup reduces the number of model-contrastive loss rounds during training, which can enhance the performance of the person ReID model and avoid overfitting. Moreover, considering the lightweight model makes it easier to deploy on edge devices, which is helpful for the wide application of our FRM. We propose a lightweight person ReID network, named multi-branch combined depth space network (MB-CDNet), to replace the ResNet-50 network in FedPav. MB-CDNet is a multi-branch version of combined depth space network (CDNet) [13]. We add two branches on the basis of CDNet and introduce an attention pyramid network (APNet) [14] to improve the performance of the model.

The main contributions of this paper are summarized as follows:

- Propose a federated person ReID method with MOON warmup, named FRM.
- Propose MB-CDNet, a lightweight network for person ReID.

- Conduct extensive experiments based on an edge-cloud scenario to evaluate the performance of FRM.

The rest of this paper is organized as follows: in [Section 2](#), related studies are described. In [Section 3](#), the design of FRM is proposed. In [Section 4](#), the experimental results are presented and analyzed; finally, the study's conclusions are presented in [Section 5](#).

## 2 Related Work

In this section, the existing publications related to our study are reviewed from the perspectives of person ReID and federated learning.

### 2.1 Person ReID

Person ReID is an image retrieval task. With the development of deep learning, the research direction of person ReID has shifted from hand-built body structure [15] and distance metric learning [16] to feature extraction from images by designing and constructing deep neural networks, comparing probe image features with features of gallery images in the database, which forms the basis of tasks such as person searching, and multi-target multi-camera tracking. A considerable gap remains between research-based scenarios and practical applications of these technologies. On the other hand, to reduce the requirement of running environment, some lightweight models are proposed. Howard et al. [17] introduced depthwise separable convolution to reduce the number of parameters significantly, and global hyper-parameters that allow the builder to choose the right sized model for their application based on the constraints of the problem to efficiently trade off between latency and accuracy. Zhang et al. [18] further reduced the number of parameters by introducing group convolution, and channel shuffle. Channel shuffle operation allows group convolution to obtain input data from different groups, which makes the input and output channels fully related and improves the performance of group convolutional layers. Zhou et al. [19] proposed OSNet, a multi-scale network with better performance than the standard networks by designing a residual block composed of multiple convolutional streams. Li et al. [13] proposed a novel search space, and search for an efficient network architecture, named CDNet, via a differentiable architecture search algorithm. These works focus on the extraction of global features and pay less attention to part features. In recent years, researchers have proposed models based on space partitioning [20,21], and most of them are via part pooling after backbone. However, since most of these works are based on a standard network, this is a challenge for the environment of model deployment. This is especially true for multi-branch networks, as branching tends to make the network larger.

### 2.2 Federated Learning

Federated learning is a distributed training framework for deep learning. McMahan et al. [8] proposed a federated averaging (FedAvg) algorithm as a baseline. In recent years, statistical heterogeneity, one of the key challenges in federated learning [22], has been the subject of extensive research [10,23]. Li et al. [10] proposed FedProx, which introduces a proximal term into the objective during local training to limit local model updates. Karimireddy et al. [24] proposed SCAFFOLD, an approach that corrects local updates by introducing control variates. The gradients in local training are adjusted by comparing the differences between local and global control variables. However, researchers [11] pointed out that these methods do not perform well on the computer vision task via deep learning. In addition to federated learning, researchers have proposed other distributed training methods by developing a resource allocation algorithm to jointly minimize training time and energy consumption [25]. On the other hand, federated learning limits the propagation range of the training

data. Sensitive information will not be transmitted outside the edge servers, effectively protecting privacy. Zhuang et al. [12] proposed federated person ReID (FedReID), combining federated learning and person ReID, along with CDW to alleviate the unbalanced impact of huge differences in sizes of datasets and KD to reduce volatility and help the model training to converge.

### 2.3 Contrastive Learning

Contrastive learning, derived from self-supervised learning, achieves excellent performance in computer vision tasks. By reducing the distance between the representations of different augmented views from the same image and increasing the distance between the representations of augmented views from different images, the resulting deep learning model's performance can be improved. Recently, many contrastive learning methods have been proposed. Tian et al. [26] applied contrastive learning to the multiview setting, attempting to maximize the mutual information between representations of different views of the same scene. From a perspective on contrastive learning as dictionary look-up, He et al. [27] built a dynamic dictionary with a queue and a moving-averaged encoder and present momentum contrast for unsupervised visual representation learning. MOON constructs a model-contrastive loss function to minimize the representation distance between differently augmented views from the same image. Since the data needed to calculate model-contrastive loss is recorded locally, MOON does not increase the communication cost (the number of communication rounds).

## 3 Methods

In this section, we first introduce the scenario, followed by our motivation. In addition, we introduce the design of MB-CDNet. Finally, we describe several key technologies involved in FRM and the details of FRM with CDW and KD.

### 3.1 Scenario

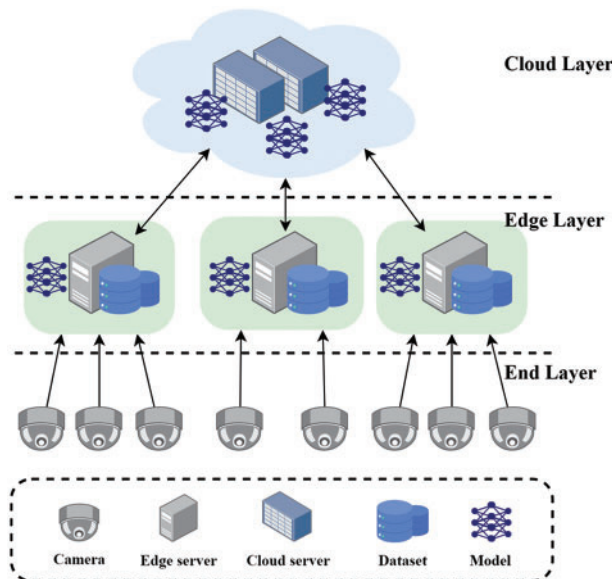
In terms of the application of federal learning and referring to the two scenarios proposed in FedReID [12], a federated-by-dataset approach is selected in this study. The federated-by-dataset scenario is an edge-cloud architecture in which the edge server is regarded as the federated learning "client" that collects and stores images taken from multiple cameras. The cloud server is defined as the "central server" in federated learning, which gathers the person ReID model weights trained and uploaded from each edge server to generate a federated aggregation. This approach is selected to simulate a real-world scenario in which multiple cameras in a community transmit data to a local edge server for storage and training, and the edge server then transmits the person ReID model weights to the cloud server for federated aggregation. As illustrated in Fig. 1, in summary, we consider a three-layer hierarchical federated learning system of person ReID in an edge-cloud environment, consisting of cameras in the end layer, several edge servers in the edge layer near the cameras, and a centralized cloud in the cloud layer.

Furthermore, given that the distance between the edge servers is different, we consider a scenario in which each of the three training datasets is divided into a further three child training datasets with non-overlapping identities to simulate three independent edge servers that store data from the same domain. To test the person ReID model training, each child training dataset is tested with the test dataset of the three parent datasets to verify the model's performance.

### 3.2 Federated Person ReID with MOON Warmup

As described in Section 3.1, the person ReID data are stored on different edge servers and collected from different cameras in different locations and at different times. These data vary in terms of

the number of images and identities, and the data distributions are affected by the data collection environment, i.e., they suffer from non-IID.



**Figure 1:** The architecture of federated person ReID in edge-cloud environment

To solve this problem, we propose FRM. Specifically, based on FedPav, the MOON warmup step is added to correct the local training of the individual edge servers and improve model effectiveness by calculating and back-propagating the model-contrastive loss, which represents the similarity between the local and global models. The framework of FRM is shown in Fig. 2. In addition, CDW is introduced to mitigate differences in data quantity between edge servers by dynamically adjusting the federated aggregation weight of the local model, and KD is introduced to solve the problem of unstable performance during model training by transferring information from the local model to the global model.

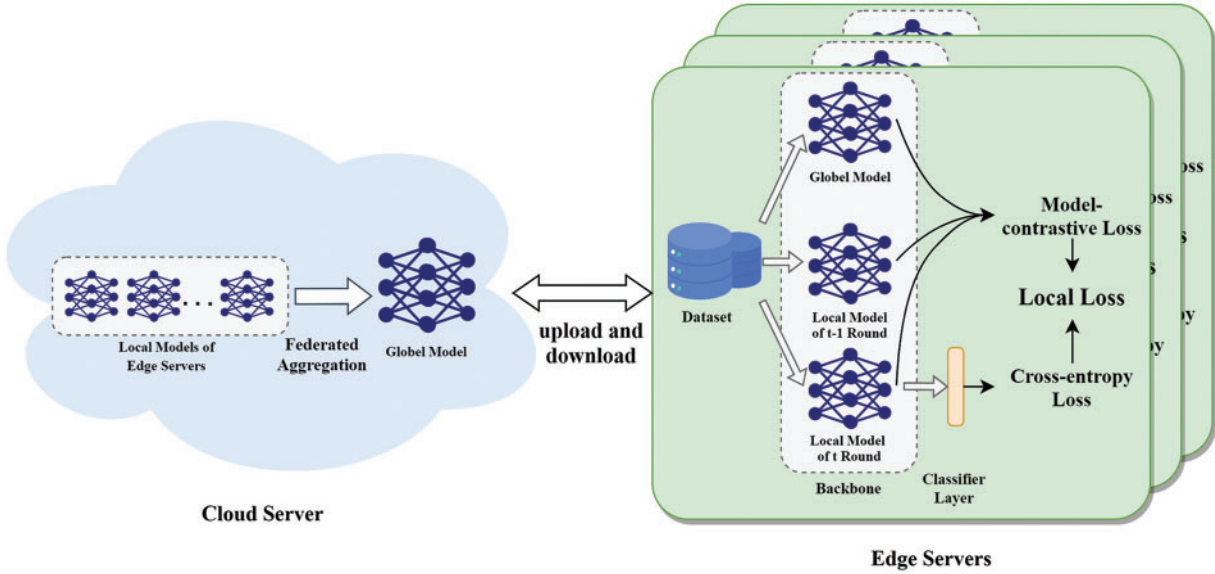
On the other hand, currently, mainstream models have a large number of parameters, which is a challenge for edge devices. The number of parameters may make these models difficult to deploy on edge devices with limited computing power. To solve this problem, we propose MB-CDNet, a lightweight network for person ReID.

### 3.2.1 MB-CDNet

In recent years, mainstream person ReID models have been based on the ID-discriminative embedding (IDE) model [28]. The backbone forms an important part of the IDE model and determines the feature learning capability of the person ReID model. ResNet [29] is a residual learning framework. By reformulating the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions, ResNet achieved phenomenal success at the time. ResNet-50 (i.e., ResNet with 50 layers) is a common choice as the backbone of the IDE model.

In this paper, we propose MB-CDNet, a lightweight network for person ReID which has a smaller number of parameters and better performance than ResNet-50.





**Figure 2:** The framework of FRM

We add two branches on the basis of CDNet and introduce APNet to improve the performance of the model. Specifically, the global branch and part branch improve the ability of the network to extract global and local features of the image, and the APNet makes the network pay attention to the channel features of the image.

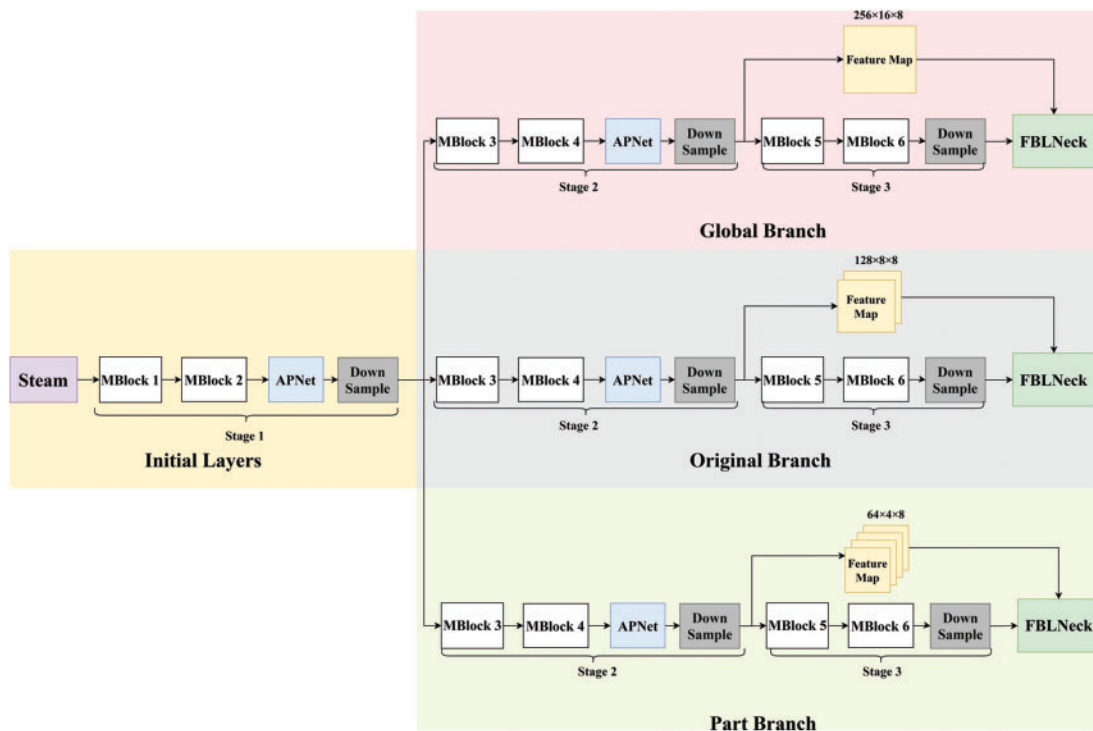
As illustrated in Fig. 3, MB-CDNet consists of initial layers, a global branch, an original branch, and a part branch. Firstly, the input image is passed through initial layers consisting of a stem layer and Stage 1, which stacks two MBlocks, an APNet, and a down sample block. After forwarding the image through the initial layers, the network forms three branches, each of them consisting of Stage 2, Stage 3, and FBLNeck. The construction of Stage 2 is similar to that of Stage 1 but has larger dimensions. The construction of Stage 3 is the same as in CDNet. By this design, the layers in the stem block and Stage 1 are shared by all the branches.

The difference between branches is mainly the way to segment the feature map as the input of FBLNeck. In the global branch, the output of Stage 2 is a complete feature map. In the original branch, the feature map is striped into two parts. In part branch, the feature map is striped into four parts.

We trained MB-CDNet on Market-1501 [2] and DukeMTMC-reID [30] datasets respectively. In this experiment, the number of epoch is 350, batch size is 64. For the optimizer, the weight decay is 0.0005 and the momentum is 0.9. Table 1 shows performance comparisons between MB-CDNet and other advanced models. All of the models listed are pretrained on ImageNet dataset [31]. The number of parameters is counted at inference time. Compared to the performance of SCSN [32], which achieves the best performance on Market-1501 in the table, the mAP and Rank-1 accuracy of MB-CDNet are just lower by 0.1% and 0.3%, respectively, with 19M fewer parameters. On DukeMTMC-reID, the mAP of MB-CDNet are better by 0.6% than SCSN. Compared with network Auto-ReID [33] with the second least parameter, MB-CDNet has advantages in the performance of both datasets.

In contrast to traditional deep learning models for federated learning, the identity classifier of the person ReID model in each client differs due to the different training datasets. To solve the problem

that person ReID models cannot be aggregated due to these different identity classifiers, FedPav is introduced in the FRM approach.



**Figure 3:** The architecture of MB-CDNet

**Table 1:** Performance comparisons between MB-CDNet and other models

Model	Param (M)	Market-1501		DukeMTMC-reID	
		mAP	Rank-1	mAP	Rank-1
VPM [34]	~26	80.8	93.0	72.6	83.6
CtF [35]	~26	84.9	93.7	74.8	87.6
PCB + RPP [36]	~26	81.6	93.8	69.2	83.3
IANet [37]	~26	83.1	94.4	–	–
BagofTricks [38]	~26	85.9	94.5	76.6	86.4
SCSN [32]	~26	88.5	95.7	79.0	90.1
Auto-ReID [33]	11.4	85.1	94.5	75.1	88.5
<b>MB-CDNet</b>	6.9	88.4	95.4	79.6	89.8

### 3.2.2 FedPav

Given the inconsistency in identity classifiers of person ReID models across all clients, FedPav [12] is introduced as an improvement of FedAvg. FedAvg [8] is a standard federated learning algorithm: first, the person ReID model is trained and sent to the server from clients and then aggregated as a weighted average. In contrast to FedAvg, in FedPav, the clients only upload the backbone of each person ReID model to the server for federated aggregation. The number of person IDs in the person

ReID dataset of each edge server is different, meaning that the dimension of the identity classifier of the person ReID model also differs. By uploading only the model backbone, FedPav avoids the issue that the weight of each client's model cannot be directly aggregated due to the differences in their identity classifiers. Same to FedAvg, FedPav aims to obtain models better than the model trained on individual datasets (i.e., local training).

The process of FedPav is as follows: (1) The cloud server initializes the backbone of global model weight and downloads it to all the edge servers. (2) Each edge server initializes a local identity classifier and combined it with the backbone download from the cloud server into a local model. (3) The edge server trains the local model with the local dataset stored in the edge server. (4) The edge server uploads the updated backbone of its local model. (5) The cloud server aggregates the model backbones from all the edge servers, generating an updated global model backbone and downloads it to all the edge servers, and back to step (2). FedPav can also aggregate models from other fields that are only partially identical between models.

### 3.2.3 MOON Warmup

Person ReID datasets with non-IID include not only differences in the number of images between datasets but also differences in image content. Additionally, in federated learning, datasets are not shared, thus it is impossible to directly measure and compare image content differences between datasets. To address the above problems, we introduce MOON warmup, a modified version of MOON for federated person ReID. MOON can indirectly obtain the difference information between datasets and can calculate and back-propagate the model-contrastive loss.

The MOON approach is based on the premise that in federated learning, local training often drifts, whereas the global model may be more representative than the local models. Specifically, MOON alleviates the impact of data with non-IID on the model by forcing the representation learned by the local model to approach that of the global model and diverge from the representation in the previous local model.

The local loss in MOON is defined as:

$$\mathcal{l} = \mathcal{l}_{sup} + \mu \mathcal{l}_{con} \quad (1)$$

where  $\mu$  is the weight of  $\mathcal{l}_{con}$ ,  $\mathcal{l}_{sup}$  represents the loss of supervised learning, and  $\mathcal{l}_{con}$  represents the model-contrastive loss which is defined as:

$$\mathcal{l}_{con} = -\log \frac{\exp\left(\frac{\text{sim}(z, z_{glob})}{\tau}\right)}{\exp\left(\frac{\text{sim}(z, z_{glob})}{\tau}\right) + \exp\left(\frac{\text{sim}(z, z_{prev})}{\tau}\right)} \quad (2)$$

where  $\tau$  represents the temperature parameter.  $z$ ,  $z_{glob}$ , and  $z_{prev}$  are the representation of input  $x$  from the local model  $w_i^t$  the global model  $w^t$ , and the last round's local model  $w_i^{t-1}$ , respectively, which are defined as:

$$\begin{aligned} z &= \mathbf{R}_{w_i^t}(x) \\ z_{glob} &= \mathbf{R}_{w^t}(x) \\ z_{prev} &= \mathbf{R}_{w_i^{t-1}}(x) \end{aligned} \quad (3)$$

MOON was originally proposed to solve the non-IID problem in image classification tasks. Given the similarity between image classification and person ReID tasks, we speculated that this method



would be feasible. However, in our experiment, the performance improvement of the person ReID model is not stable using the MOON method. Initially, the performance of FedPav with MOON exceeds that of FedPav without MOON; subsequently, after a certain epoch (around 150 epochs in this experiment), MOON negatively affects the performance of the person ReID model such that its performance is eventually lower than that of FedPav.

In contrast to image classification, person ReID involves the use of more data with non-IID, and the domain gap between different datasets is larger, leading to frequent performance fluctuations in the process of person ReID model training, i.e., the performance of the person ReID model does not continuously improve. An important function of model-contrastive loss is to reduce the similarity between the current local model and the local model of the previous round, an approach that does not always play a positive role. The fluctuations in person ReID model performance usually occur in the slow growth period after the initial rapid increase of the performance metrics in the early training stages. The negative effects caused by this issue can be avoided by modifying the loss function when the person ReID model's performance fluctuates. Accordingly, to take advantage of the performance improvement effect of MOON on the model at the start of training, the MOON method will only be applied at the onset of the training. This approach is referred to as MOON warmup as its effect is similar to the warmup strategy of learning rate adjusting. In the following experiments, MOON warmup will be used by default, and FedPav with MOON warmup is referred to as FRM.

### 3.3 Performance Optimization Methods

To alleviate the impact of differences between datasets on training in federated person ReID, previous studies [12] introduced CDW and KD approaches. CDW alleviates the non-IID issue caused by quantitative differences between datasets by dynamically adjusting the model weights for federated aggregation. In addition, KD can transfer information from the local model to the global model, which alleviates the problem of unstable model training performance caused by non-IID.

**CDW** is a dynamic weight adjustment method applied according to the changes in the model, which can mitigate the unbalanced effects of large differences in dataset sizes [12]. Given that larger changes should contribute more to model aggregation so that more of the newly learned knowledge can be reflected in the federated model, the algorithm is constructed to calculate each client's model changes by cosine distance.

The CDW process is as follows: (1) The edge server randomly selects a batch of training data  $D$ . (2) The edge server inputs data  $d \in D$  into the local model and the global model, respectively, to obtain logits  $z$  and logits  $z_{glob}$ . (3) The CDW of the local model weight  $m'$  in round  $t$  of training is defined as  $m' \leftarrow \frac{1}{|D|} \sum_{d \in D} 1 - \text{sim}(z, z_{glob})$ . (4) The edge server uploads CDW  $m'$  to the server to replace the weights  $n'$  of each local model in FedPav.

**KD** [39], a method to transfer knowledge between models, is applied to transfer model knowledge from the clients (teacher model) to the model on the server (student model). The KD approach can help to improve the stability of model performance during training. In the experiment, MSMT17 [40] is used for KD as a shared dataset  $D_{shared}$ .

The process of KD is as follows: (1) The cloud server downloads the shared dataset of KD  $D_{shared}$  to all edge servers. (2) The edge server inputs shared dataset  $D_{shared}$  into the local model to generate soft labels  $S_k$ , which are then uploaded to the cloud server. (3) The cloud server averages soft labels from different edge servers with  $S = \frac{1}{K} \sum_{k \in C} S_k$ . (4) The cloud server trains the global model with shared dataset  $D_{shared}$  and soft labels  $S$ .

### 3.3.1 Comparison between CDW and MOON Warmup

As shown in Table 2, both CDW and MOON warmup adjust the model training according to the differences between models to improve the training effect. Specifically, the weights from CDW affect the aggregation of the person ReID model, whereas MOON warmup directly affects the training of the person ReID model. In addition, CDW only compares the differences between the global model and the local model, while MOON warmup also considers the previous round's local model based on comparing the global model and the local model. The result of CDW is a federal aggregation weight parameter, with one number for each local model, while MOON warmup calculates the loss for back-propagation, which contains more information.

**Table 2:** Comparison between CDW and MOON warmup

Method	CDW	MOON warmup
Reference	Global model & local model	Global model, local model & previous local model
Result	Federal aggregation weight	Model-contrastive loss
Influence	Federal aggregation	Local training

### 3.3.2 Performance Optimization of FRM with CDW and KD

The process of FRM with CDW and KD is as follows: (1) The edge servers download the backbone weight  $\omega$  of the global model from the cloud server. (2) The edge server combines the backbone of the global model with the local identity classifier from the previous training to form a new person ReID model. (3) The edge server trains the new person ReID model with the local dataset and stochastic gradient descent for  $E$  local epochs of batch size  $B$  and learning rate  $\eta$ . Local loss  $\mathcal{l}$  in the previous  $e$  epochs is calculated from the supervised learning loss  $\mathcal{l}_{sup}$  (e.g., cross-entropy loss) and model-contrastive loss  $\mathcal{l}_{con}$ . The model-contrastive loss  $\mathcal{l}_{con}$  is calculated from the global model  $z_{glob}$ , the local model from the previous round  $z_{prev}$ , and the local model from this round  $z$ . Model-contrastive loss

is defined as  $\mathcal{l}_{con} = -\log \frac{\exp\left(\frac{\text{sim}(z, z_{glob})}{\tau}\right)}{\exp\left(\frac{\text{sim}(z, z_{glob})}{\tau}\right) + \exp\left(\frac{\text{sim}(z, z_{prev})}{\tau}\right)}$  (4) The edge server inputs the local

dataset into the local and global models; the edge server then calculates the CDW of the local model from the difference between the output logits  $z$  and  $z_{glob}$  from the two models. The CDW is defined as

$m^t \leftarrow \frac{1}{|D|} \sum_{d \in D} 1 - \text{sim}(z, z_{glob})$  (5) The edge server inputs shared dataset  $D_{shared}$  into the local model

to generate the corresponding soft labels  $S$ . (6) The edge server uploads the updated backbone and CDW  $m$  of the local model and the soft labels  $S$  to the cloud server. (7) The cloud server aggregates all the local models to a new global model using the CDW  $m$  of each model. The aggregation is defined

as  $\omega^{t+1} \leftarrow \sum_{k \in C} \frac{m}{M} \omega^t$  (8) The cloud server trains the federated model using the shared dataset  $D_{shared}$  and averaged soft labels  $S$ , which is defined as  $w^{t+1} \leftarrow w^{t+1} - \eta \nabla \text{Loss}(F_{w^{t+1}}(d), s_d^{t+1})$  where  $d \in D_{shared}$  (9) Return to step (1).

The process of FRM with CDW and KD is summarized in Algorithms 1 and 2, where  $T$  represents the number of communication rounds,  $C$  represents the set of clients,  $K$  is the number of the client,  $E$  is the number of the local epoch,  $e$  is the number of the local epoch for MOON warmup,  $\tau$  represents

the temperature of MOON,  $\mu$  is a hyper-parameter for balancing supervised learning loss and model-contrastive learning loss,  $m$  represents the CDW of the local model weight, and  $D_{shared}$  is the dataset for knowledge distillation.  $F_{\omega}()$  represents the whole network with model weight  $\omega$ .  $R_{\omega}()$  represents the network before the identity classifier with model weight  $\omega$ . Table 3 summarizes the parameters used in this paper.

**Table 3:** Introduction of algorithm parameters

Parameters	Introduction
$T$	The number of communication rounds
$C$	The set of clients
$E$	The number of the local epoch
$e$	The number of the local epoch for MOON warmup
$B$	Batch size
$\eta$	Learning rate
$\tau$	The temperature of MOON
$\mu$	The hyper-parameter for balancing supervised learning loss and model-contrastive learning loss
$D$	Dataset
$K$	The number of the client
$\omega$	Model weight
$S$	Soft labels
$m$	The cosine distance weight of the local model
$\mathcal{l}$	Loss
$z$	Logits
$R()$	The network before the identity classifier
$F()$	The whole network

---

**Algorithm 1** FRM + CDW + KD (Server)

---

**Input:**  $T, C, E, e, B, \eta, \tau, \mu, D_{shared}, K$

**Output:**  $w^T$

```

1: initialize  $w^0$ 
2: for  $t = 0, 1, 2, \dots, T - 1$  do
3:   for each  $k \in C$  concurrently do
4:      $\{w_k^t, m_k^t, S_k^t\} \leftarrow \text{Client}_k(w^t)$ 
5:   end for
6:    $M \leftarrow \sum_{k \in C} \frac{1}{m_k}$ 
7:    $w^{t+1} \leftarrow \sum_{k \in C} \frac{m_k}{M} w_k^t$ 
8:    $S^{t+1} \leftarrow \frac{1}{K} \sum_{k \in C} S_k^t$ 
9:   for each  $d \in D_{shared}$  do
10:     $\ell \leftarrow \text{Loss}(F_{w^{t+1}}(d), s_d^{t+1})$ 
11:     $w^{t+1} \leftarrow w^{t+1} - \eta \nabla \ell$ 
12:   end for
13: end for
14:
15: return  $w^T$ 

```

---

**Algorithm 2** FRM + CDW + KD (Client)**Input:**  $T, C, E, e, B, \eta, \tau, \mu, D_{shared}, K$ **Output:**  $w^T$ 


---

```

1:  $w_0^t \leftarrow w^t$ 
2: for  $i = 0, 1, 2, \dots, E$  do
3:   if  $i < e$  then
4:     for each  $b \in B$  do
5:        $\{x, y\} \leftarrow b$ 
6:        $\ell_{sup} \leftarrow \text{Loss}(\mathbb{F}_{w_i^t}(x), y)$ 
7:        $z \leftarrow \mathbb{R}_{w_i^t}(x)$ 
8:        $z_{glob} \leftarrow \mathbb{R}_{w^t}(x)$ 
9:        $z_{prev} \leftarrow \mathbb{R}_{w_{i-1}^t}(x)$ 
10:       $\ell_{con} \leftarrow -\log \frac{\exp(\frac{\text{sim}(z, z_{glob})}{\tau})}{\exp(\frac{\text{sim}(z, z_{glob})}{\tau}) + \exp(\frac{\text{sim}(z, z_{prev})}{\tau})}$ 
11:       $\ell \leftarrow \ell_{sup} + \mu \ell_{con}$ 
12:       $w_i^t \leftarrow w_i^t - \eta \nabla \ell$ 
13:    end for
14:  else
15:    for each  $b \in B$  do
16:       $x, y \leftarrow b$ 
17:       $\ell \leftarrow \text{Loss}(\mathbb{F}_{w_i^t}(x), y)$ 
18:       $w_i^t \leftarrow w_i^t - \eta \nabla \ell$ 
19:    end for
20:  end if
21: end for
22: for each  $d \in D$  do
23:    $x \leftarrow d$ 
24:    $z \leftarrow \mathbb{R}_{w_i^t}(x)$ 
25:    $z_{glob} \leftarrow \mathbb{R}_{w^t}(x)$ 
26:    $m_d \leftarrow 1 - \text{sim}(z, z_{glob})$ 
27: end for
28:  $m^t \leftarrow \frac{1}{|D|} \sum_{d \in D} m_d$ 
29:  $w^t \leftarrow w_E^t$ 
30: for each  $d \in D_{shared}$  do
31:    $x \leftarrow d$ 
32:    $s_d \leftarrow \mathbb{F}_{w_i^t}(x)$ 
33: end for
34:  $S^t \leftarrow \{s_d \mid d \in D_{shared}\}$ 
35:
36: return  $\{w^t, m^t, S^t\}$ 

```

---

## 4 Experiment

This section provides details of extensive experiments that were carried out to evaluate the performance of FRM. The experiments were conducted on Linux 21.2 with an Intel i5-10600KF CPU, 64 GB DRAM, and 2 NVIDIA RTX 3090 GPUs in a single machine. The simulation was performed in a Python 3.8 environment. We experimented with ResNet-50 and MB-CDNet as local models, respectively, and all the local models are pre-trained on ImageNet dataset [31].

## 4.1 Experiment Settings

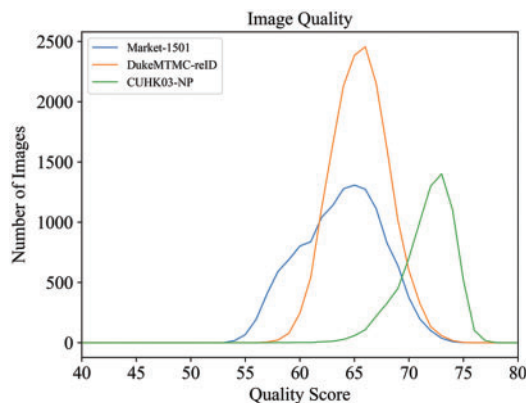
### 4.1.1 Datasets

In the experiments, four open-access datasets, as shown in Table 4, were selected to simulate real-world situations. Among these, DukeMTMC-reID [30], Market-1501 [2], and CUHK03-NP [41] were used to train and test the person ReID models, while the MSMT17 [40] dataset was used to apply knowledge distillation to federated learning. All the above datasets consist of a training set and a test set, where the test set includes a query set and a gallery set.

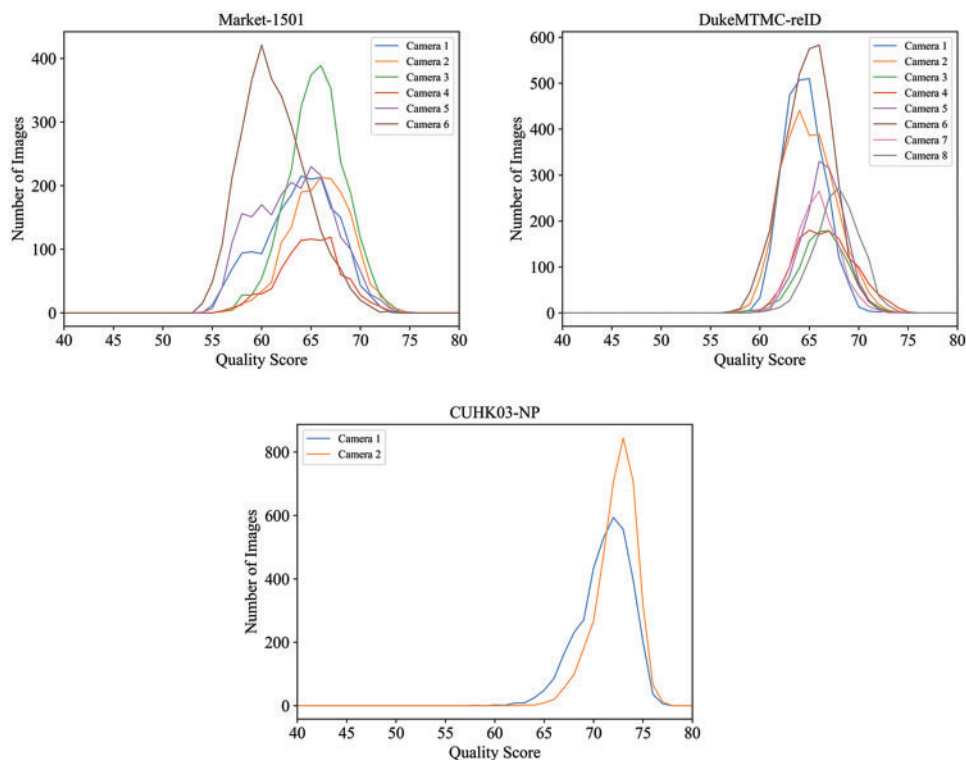
**Table 4:** The characteristics of datasets

Dataset	#Cameras	Train		Test	
		#Identities	#Images	Query #Images	Gallery #Images
DukeMTMC-reID [34]	8	702	16522	2228	17611
Market-1501 [2]	6	751	12936	3368	19732
CUHK03-NP [35]	2	767	7365	1400	5332
MSMT17 [32]	15	1041	32621	11659	82161

Fig. 4 shows the image qualities of each training dataset, and Fig. 5 shows the image quality of each camera in datasets, all quality scores are generated by the official model of PaQ-2-PiQ [42]. It can be found that in addition to the difference in the number of images in these datasets, there are also huge differences in image quality, and camera quantity, leading to a significant domain gap between them. A higher image quality score indicates a lower image quality. As shown in Fig. 4, compared with CUHK03-NP, the image qualities of Market-1501 and DukeMTMC-reID are higher and closer. As shown in Fig. 5, in Market-1501, the qualities of images from camera 1 have a big gap with other images from other cameras. In DukeMTMC-reID, the image quality between cameras 1, 2 and 6 is similar, and the image quality between cameras 3, 4, 5, 7 and 8 is also similar. In CUHK03-NP, the difference in image quality between two cameras is small. This situation is likely to be representative of when the framework is applied to real-world situations. Table 5 shows the characteristics of the datasets after segmentation.



**Figure 4:** Image quality of each dataset



**Figure 5:** Image quality of each camera in datasets

**Table 5:** The characteristics of datasets after segmentation

Dataset	#Identities	#Images
DukeMTMC-reID-1	234	5560
DukeMTMC-reID-2	234	5000
DukeMTMC-reID-3	234	5962
Market-1501-1	250	4345
Market-1501-2	250	4352
Market-1501-3	250	4225
CUHK03-NP-1	255	2459
CUHK03-NP-2	255	2468
CUHK03-NP-3	255	2418

#### 4.1.2 Performance Metrics

Two metrics were chosen to evaluate the performance of person ReID models: the mean average precision (mAP) as measured across all images in a query and the cumulative catching characteristics (CMC) curve, which ranks the query identity similarity to all gallery images. Rank-k represents the probability that the query identity is in the top-k-ranked images of the gallery. The communication



cost is calculated as twice the number of communication rounds multiplied by the person ReID model size. In the evaluation of the person ReID model, only the output features from the model's feature layer are compared and the identity classifier is ignored.

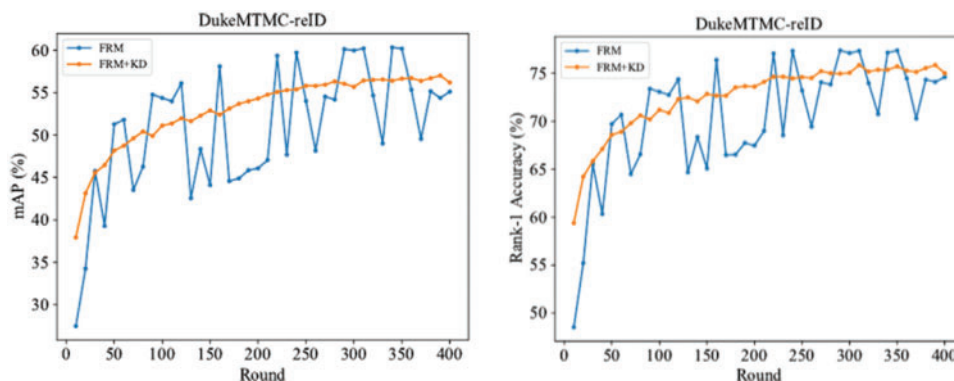
### 4.1.3 Hyper-Parameters Setting

The total number of training rounds  $T$  is 400. The number of interactions between the server and the client is 7,200. The batch size  $B$  is 32. To mitigate the feature discrepancy between the updated backbone and the local identity classifier head, the initialized learning rate of the local identity classifier is higher. Specifically, the initialized learning rate of the backbone is 0.05, and the initialized learning rate of the identity classifier is 0.005. Considering less number of the local epoch allows the local initialized classifier to be updated more timely, which is also conducive to mitigating the feature discrepancy, the number of the local epoch is set to 1. For the optimizer, the weight decay is 0.0005 and the momentum is 0.9. All experiments were performed using these settings unless otherwise specified.

## 4.2 Ablation Studies

### 4.2.1 Impart of KD

As shown in Fig. 6, the person ReID model shows stable performance with KD, however, the model's training performance was unstable without KD. Our experiments further verified that KD can reduce volatility and can help achieve training convergence. Including the MOON warmup step does not reduce the benefits of this approach. As shown in Table 7, for smaller datasets like CUHK03-NP, FRM + CDW + KD is a better choice, whereas for larger datasets, such as Market-1501 and DukeMTMS-reID, excluding CDW can achieve better results. However, notably, the benefit of KD to the stability of person ReID model performance comes at the cost of losing the optimal person ReID model value. Thus, not applying KD may be preferable in scenarios where stable person ReID model performance is not required but optimal model performance is necessary.



**Figure 6:** Performance and convergence comparison of FRM and FRM + KD in DukeMTMC-reID dataset (with ResNet-50)

#### 4.2.2 Impart of CDW

Based on the scenario proposed in this paper, for ResNet-50, the federated framework using CDW for aggregation weighting performs better on the CUHK03-NP datasets; however, on others, CDW performs worse than the framework using the dataset size as the aggregation weight. This may be because person ReID models trained on smaller datasets like CUHK03-NP gain more weight in federated aggregation than the original method of assigning weights according to data volume; thus, the federated models tend to perform better on small datasets and perform worse on other datasets. Generally, the influence of CDW is neutral in our tested scenario, which may indicate that CDW is not required in some cases.

#### 4.2.3 Impart of Cross-Entropy Loss

Table 10 shows the performance of local training, FedPav, and FRM training with triplet loss only. Data shows that the performance of the model that trains with triplet loss only is far worse than the performance of the model that trains with cross-entropy loss and triplet loss. It can be found that cross-entropy loss plays a very important role in the person ReID model. Data also reflects the importance of the identity classifier in person ReID model. The involvement of the identity classifier is necessary for the training of person ReID with federated learning.

#### 4.3 Comparison with FedPav

The results of local training, FedPav, and FRM are shown in Tables 6 and 7. As shown, for the same initial model weights, the person ReID model has higher performance after FRM compared to the person ReID model trained with local datasets only, demonstrating that model obtains more knowledge from different datasets through the use of FRM. For ResNet-50, FRM beats FedPav by 2.60% in terms of mAP, and 2.17% in terms of Rank-1 accuracy on Market-1501. For MB-CDNet, FRM beats FedPav by 1.51% and 1.16% respectively in terms of Rank-1 accuracy and mAP on Market-1501.

**Table 6:** Performance of local training, FedPav, FedPav with MOON, and FRM (with ResNet-50)

Dataset	Local Training		FedPav		FedPav + MOON		FRM	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
DukeMTMC-reID-1	44.18	66.06						
DukeMTMC-reID-2	43.50	64.81	51.33	71.00	51.39	71.00	<b>55.12</b>	<b>74.59</b>
DukeMTMC-reID-3	44.31	65.39						
Market-1501-1	49.31	73.78						
Market-1501-2	52.35	76.18	65.20	85.59	63.98	84.50	<b>67.80</b>	<b>87.76</b>
Market-1501-3	50.55	74.73						
CUHK03-NP-1	21.90	23.71						
CUHK03-NP-2	21.66	24.21	23.58	27.28	19.77	23.00	<b>26.01</b>	<b>29.50</b>
CUHK03-NP-3	23.24	25.28						

**Table 7:** Performance of local training, FedPav, FedPav with MOON, and FRM (with MB-CDNet)

Dataset	Local Training		FedPav		FedPav + MOON		FRM	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
DukeMTMC-reID-1	67.84	82.54						
DukeMTMC-reID-2	68.29	82.85	74.71	84.14	74.53	83.81	<b>76.79</b>	<b>85.86</b>
DukeMTMC-reID-3	68.33	82.58						
Market-1501-1	78.11	91.56						
Market-1501-2	78.15	91.33	82.50	91.30	81.12	90.22	<b>84.01</b>	<b>92.46</b>
Market-1501-3	78.02	91.03						
CUHK03-NP-1	53.38	57.78						
CUHK03-NP-2	52.93	57.35	55.72	61.13	52.74	58.45	<b>57.32</b>	<b>62.44</b>
CUHK03-NP-3	53.38	58.07						

To further verify the effectiveness of MOON warmup, ablation experiments were performed, which are divided into four plans, as shown in [Tables 8](#) and [9](#). Irrespective of the optimization method combination used, the FRM results are always better than those of FedPav. The data show that MOON warmup, which was applied at the beginning of training, had a positive effect that persisted after many epochs of training. We suggest that MOON warmup inhibited overfitting in the person ReID model at the beginning of training, which may help to improve the stability of the model.

**Table 8:** Performance of FRM and FedPav with CDW and KD (with ResNet-50)

Dataset	DukeMTMC-reID		Market-1501		CUHK03-NP	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
FedPav	51.33	71.01	65.21	85.60	23.58	27.29
<b>FRM</b>	<b>55.13</b>	<b>74.60</b>	<b>67.80</b>	<b>87.77</b>	<b>26.01</b>	<b>29.50</b>
FedPav + KD	54.15	74.06	60.36	82.99	27.55	31.86
<b>FRM + KD</b>	<b>56.20</b>	<b>75.00</b>	<b>63.36</b>	<b>84.68</b>	<b>28.47</b>	<b>32.14</b>
FedPav + CDW	50.25	70.38	64.59	85.33	25.26	28.36
<b>FRM + CDW</b>	<b>53.12</b>	<b>72.71</b>	<b>68.07</b>	<b>87.32</b>	<b>29.57</b>	<b>34.00</b>
FedPav + CDW + KD	52.78	72.40	58.44	81.12	28.27	32.21
<b>FRM + CDW + KD</b>	<b>55.16</b>	<b>74.91</b>	<b>63.78</b>	<b>84.86</b>	<b>30.99</b>	<b>34.36</b>

**Table 9:** Performance of FRM and FedPav with CDW and KD (with MB-CDNet)

Dataset	DukeMTMC-reID		Market-1501		CUHK03-NP	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
FedPav	74.71	84.14	82.50	91.30	55.72	61.13
<b>FRM</b>	<b>76.79</b>	<b>85.86</b>	<b>84.01</b>	<b>92.46</b>	<b>57.32</b>	<b>62.44</b>
FedPav + KD	75.55	84.75	82.25	91.12	54.69	59.78
<b>FRM + KD</b>	<b>76.60</b>	<b>85.19</b>	<b>84.13</b>	<b>92.73</b>	<b>55.89</b>	<b>60.68</b>
FedPav + CDW	74.54	83.28	82.58	92.04	61.26	65.11
<b>FRM + CDW</b>	<b>75.77</b>	<b>84.98</b>	<b>84.44</b>	<b>92.90</b>	<b>64.62</b>	<b>68.87</b>
FedPav + CDW + KD	74.90	84.22	82.02	91.40	54.90	60.28
<b>FRM + CDW + KD</b>	<b>76.70</b>	<b>85.37</b>	<b>84.10</b>	<b>92.49</b>	<b>56.93</b>	<b>61.95</b>

Notably, in the proposed scenario, FedPav with optimization methods does not perform as well as FedPav alone on some datasets. FedPav achieves higher performance than all combinations of FedPav with optimization methods on Market-1501. On the one hand, this outcome may be due to performance fluctuations without KD; on the other hand, this result also shows that these optimization methods have certain limitations which are not always effective.

**Table 10:** Performance of local training, FedPav, and FRM training with triplet loss only (with MB-CDNet)

Dataset	Local Training		FedPav		FRM	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
DukeMTMC-reID-1	12.11	27.87				
DukeMTMC-reID-2	11.22	26.79	8.88	21.50	2.25	6.14
DukeMTMC-reID-3	15.18	34.24				
Market-1501-1	11.89	30.34				
Market-1501-2	10.39	26.72	6.60	19.70	2.02	5.81
Market-1501-3	11.69	28.68				
CUHK03-NP-1	47.05	51.99				
CUHK03-NP-2	31.78	35.71	0.74	0.78	0.64	0.91
CUHK03-NP-3	40.69	44.85				

#### 4.4 Comparison of MB-CDNet and ResNet-50

As shown in Tables 6 and 7, MB-CDNet achieves better performances than ResNet-50 in both local training and federated learning. For example, with FRM, MB-CDNet beats ResNet-50 by 16.21% in terms of mAP, and 4.69% in terms of Rank-1 accuracy on Market-1501. With FedPav, MB-CDNet beats ResNet-50 by 17.29% in terms of mAP, and 5.7% in terms of Rank-1 accuracy on Market-1501. It can be found that MB-CDNet achieves better performance by training only on local datasets than ResNet-50, which federated learning with all datasets. Result also shows that federated learning

works for a lightweight model like MB-CDNet. In addition, compared with ResNet-50, federated learning with MB-CDNet reduce the communication overhead associated with model transmission.

## 5 Conclusion

In this paper, we propose FRM, a federated learning framework of person ReID that incorporates MOON, a federal learning method for image classification. Our approach was inspired by the extensive use of image classification backbones in person ReID. However, preliminary experiments showed that the improvements from MOON are not stable. To overcome this issue, we reduced the number of rounds of MOON participation in federal learning, meaning that MOON is used as a warmup method to improve the training effect. Our experiments demonstrate the superiority of FRM compared to local training. On the other hand, the number of parameters may make person ReID models difficult to deploy on edge devices with limited computing power. To solve this problem, we propose MB-CDNet, a lightweight network for person ReID, which has a smaller number of parameters and better performance than ResNet-50. Finally, we conducted extensive experiments on an edge-cloud scenario using nine datasets; a comparison with several optimization methods demonstrates the effectiveness of FRM.

**Acknowledgement:** We thank the anonymous reviewers for their comments and suggestions that improved this paper.

**Funding Statement:** This research is supported by the the Natural Science Foundation of Jiangsu Province of China under Grant No. BK20211284, and the Financial and Science Technology Plan Project of Xinjiang Production and Construction Corps under Grant No. 2020DB005.

**Author Contributions:** The authors confirm contribution to the paper as follows: study conception and design: Baixuan Tang, Xiaolong Xu; analysis and interpretation of results: Baixuan Tang, Fei Dai, Song Wang; draft manuscript preparation: Fei Dai, Song Wang. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The data are not publicly available due to privacy restriction.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] Y. C. Chen, X. Zhu, W. S. Zheng and J. H. Lai, "Person re-identification by camera correlation aware feature augmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 392–408, 2018.
- [2] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang *et al.*, "Scalable person re-identification: A benchmark," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Santiago, Chile, pp. 1116–1124, 2015.
- [3] X. Xu, Z. Fang, L. Qi, X. Zhang, Q. He *et al.*, "TripRes: Traffic flow prediction driven resource reservation for multimedia iov with edge computing," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 2, pp. 1–21, 2021.
- [4] X. Xu, Z. Fang, J. Zhang, Q. He, D. Yu *et al.*, "Edge content caching with deep spatiotemporal residual network for iov in smart city," *ACM Transactions on Sensor Networks*, vol. 17, no. 3, pp. 1–33, 2021.

- [5] H. Tian, X. Xu, T. Lin, Y. Cheng, C. Qian *et al.*, “Dima: Distributed cooperative microservice caching for internet of things in edge computing by deep reinforcement learning,” *World Wide Web-internet and Web Information Systems*, vol. 25, no. 5, pp. 1769–1792, 2022.
- [6] X. Xu, Q. Jiang, P. Zhang, X. Cao, M. R. Khosravi *et al.*, “Game theory for distributed iov task offloading with fuzzy neural network in edge computing,” *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 11, pp. 4593–4604, 2022.
- [7] Y. Liu, Q. He, D. Zheng, X. Xia, F. Chen *et al.*, “Data caching optimization in the edge computing environment,” *IEEE Transactions on Services Computing*, vol. 15, no. 4, pp. 2074–2085, 2022.
- [8] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas *et al.*, “Communication-efficient learning of deep networks from decentralized data,” in *Artificial Intelligence and Statistics*, vol. 54, no. 1, pp. 1273–1282, 2017.
- [9] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin *et al.*, “Federated learning with non-iid data,” arXiv preprint arXiv:1806. 00582, 2018.
- [10] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi and A. Talwalkar, “Federated optimization in heterogeneous networks,” *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.
- [11] Q. Li, B. He and D. Song, “Model-contrastive federated learning,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp. 10713–10722, 2021.
- [12] W. Zhuang, Y. Wen, X. Zhang, X. Gan, D. Yin *et al.*, “Performance optimization of federated person re-identification via benchmark analysis,” in *Proc. of the 28th ACM Int. Conf. on Multimedia*, New York, NY, USA, pp. 955–963, 2020.
- [13] H. Li, G. Wu and W. S. Zheng, “Combined depth space based architecture search for person re-identification,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp. 6729–6738, 2021.
- [14] G. Chen, T. Gu, J. A. Bao and J. Zhou, “Person re-identification via attention pyramid,” *IEEE Transactions on Image Processing*, vol. 30, pp. 7663–7676, 2021.
- [15] T. Matsukawa, T. Okabe, E. Suzuki and Y. Sato, “Hierarchical gaussian descriptor for person re-identification,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, Nevada, USA, pp. 1363–1372, 2016.
- [16] H. X. Yu, A. Wu and W. S. Zheng, “Unsupervised person re-identification by deep asymmetric metric embedding,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 956–973, 2018.
- [17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang *et al.*, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704. 04861, 2017.
- [18] X. Zhang, X. Zhou, M. Lin and J. Sun, “Shufflenet: An extremely efficient convolutional neural network for mobile devices,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 6848–6856, 2018.
- [19] K. Zhou, Y. Yang, A. Cavallaro and T. Xiang, “Omni-scale feature learning for person re-identification,” in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, South Korea, pp. 3702–3712, 2019.
- [20] B. Xie, X. Wu, S. Zhang, S. Zhao and M. Li, “Learning diverse features with part-level resolution for person re-identification,” arXiv preprint arXiv:2001.07442, 2020.
- [21] H. Chen, B. Lagadec and F. Bremond, “Learning discriminative and generalizable representations by spatial channel partition for person re-identification,” in *IEEE Winter Conf. on Applications of Computer Vision, Snowmass Village*, Colorado, USA, pp. 2483–2492, 2020.
- [22] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis *et al.*, “Advances and open problems in federated learning,” *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [23] X. Yao, T. Huang, R. X. Zhang, R. Li and L. Sun, “Federated learning with unbiased gradient aggregation and controllable meta updating,” arXiv preprint arXiv:1910.08234, 2019.
- [24] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich *et al.*, “Scaf-fold: Stochastic controlled averaging for on-device federated learning,” in *Proc. of the 37th Int. Conf. on Machine Learning*, Long Beach Convention Center, Long Beach, USA, 2019.



- [25] X. Xu, H. Tian, X. Zhang, L. Qi, Q. He *et al.*, “DisCOV: Distributed COVID-19 detection on X-ray images with edge-cloud collaboration,” *IEEE Transactions on Services Computing*, vol. 15, no. 3, pp. 1206–1219, 2022.
- [26] Y. Tian, D. Krishnan and P. Isola, “Contrastive multiview coding,” in *European Conf. on Computer Vision*, Glasgow, UK, Springer, pp. 776–794, 2020.
- [27] K. He, H. Fan, Y. Wu, S. Xie and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 9729–9738, 2020.
- [28] L. Zheng, Y. Yang and A. G. Hauptmann, “Person re-identification: Past, present and future,” arXiv preprint arXiv:1610.02984, 2016.
- [29] K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.
- [30] Z. Zheng, L. Zheng and Y. Yang, “Unlabeled samples generated by gan improve the person re-identification baseline in vitro,” in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 3754–3762, 2017.
- [31] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li *et al.*, “Imagenet: A large-scale hierarchical image database,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, Florida, USA, pp. 248–255, 2009.
- [32] X. Chen, C. Fu, Y. Zhao, F. Zheng, J. Song *et al.*, “Salience-guided cascaded suppression network for person re-identification,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 3300–3310, 2020.
- [33] R. Quan, X. Dong, Y. Wu, L. Zhu and Y. Yang, “Auto-reid: Searching for a part-aware convnet for person re-identification,” in *Proc. of the IEEE Int. Conf. on Computer Vision*, Seoul, South Korea, pp. 3750–3759, 2019.
- [34] Y. Sun, Q. Xu, Y. Li, C. Zhang, Y. Li *et al.*, “Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Los Angeles, CA, USA, pp. 393–402, 2019.
- [35] G. Wang, S. Gong, J. Cheng and Z. Hou, “Faster person re-identification,” in *European Conf. on Computer Vision*, Edinburgh, England, UK, Springer, pp. 275–292, 2020.
- [36] Y. Sun, L. Zheng, Y. Yang, Q. Tian and S. Wang, “Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline),” in *Proc. of the European Conf. on Computer Vision*, Munich, Germany, pp. 480–496, 2018.
- [37] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan *et al.*, “Interaction-and-aggregation network for person re-identification,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Los Angeles, CA, USA, pp. 9317–9326, 2019.
- [38] H. Luo, Y. Gu, X. Liao, S. Lai and W. Jiang, “Bag of tricks and a strong baseline for deep person re-identification,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, Los Angeles, CA, USA, 2019.
- [39] G. Hinton, O. Vinyals and J. Dean, “Distilling the knowledge in a neural network,” arXiv preprint arXiv:1503.02531, 2015.
- [40] L. Wei, S. Zhang, W. Gao and Q. Tian, “Person transfer gan to bridge domain gap for person re-identification,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 79–88, 2018.
- [41] W. Li, R. Zhao, T. Xiao and X. Wang, “Deepreid: Deep filter pairing neural network for person re-identification,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 152–159, 2014.
- [42] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram *et al.*, “From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 3575–3585, 2020.