Joint Optimization in Edge-Cloud Continuum for Federated Unsupervised Person Re-identification

Weiming Zhuang1,3  Yonggang Wen2  Shuai Zhang3
1 S-Lab, Nanyang Technological University  2 Nanyang Technological University  3 SenseTime Research
weiming001@e.ntu.edu.sg, ygwen@ntu.edu.sg, zhangshuai@sensetime.com

ABSTRACT

Person re-identification (ReID) aims to re-identify a person from non-overlapping camera views. Since person ReID data contains sensitive personal information, researchers have adopted federated learning, an emerging distributed training method, to mitigate the privacy leakage risks. However, existing studies rely on data labels that are laborious and time-consuming to obtain. We present FedUReID, a federated unsupervised person ReID system to learn person ReID models without any labels while preserving privacy. FedUReID enables in-situ model training on edges with unlabeled data. A cloud server aggregates models from edges instead of centralizing raw data to preserve data privacy. Moreover, to tackle the problem that edges vary in data volumes and distributions, we personalize training in edges with joint optimization of cloud and edge. Specifically, we propose personalized epoch to reassign computation throughout training, personalized clustering to iteratively predict suitable labels for unlabeled data, and personalized update to adapt the server aggregated model to each edge. Extensive experiments on eight person ReID datasets demonstrate that FedUReID not only achieves higher accuracy but also reduces computation cost by 29%. Our FedUReID system with the joint optimization will shed light on implementing federated learning to more multimedia tasks without data labels.

CCS CONCEPTS
• Computing methodologies → Distributed algorithms; Object identification; Matching; • Information systems → Top-k retrieval in databases.

KEYWORDS
federated learning, person re-identification, unsupervised learning, unsupervised person re-identification

1 INTRODUCTION

Person re-identification (ReID) is an important computer vision task that has attracted considerable research interests in the multimedia community due to its wide applications on public safety and public security [14]. Person ReID aims to match the same person in non-overlapping camera views. Over the years, researchers have proposed many innovative approaches, based on either supervised learning or unsupervised learning, to excel the performance of person ReID [20, 34, 35]. The majority of these methods require centralizing plenty of images into a central server.

However, the increasingly stringent data privacy regulations limit the development of person ReID [3]. The data collected from cameras contain sensitive personal information such as the identity and location of individuals. Due to data privacy concerns, centralizing the data would not be feasible because it would impose potential privacy leakage risks.

A recent study [44] proposes federated person re-identification (FedReID) to train person ReID models while preserving data privacy. Federated learning (FL) is a distributed learning technique that allows multiple parties to train models collectively without centralizing data [23]. FedReID implements FL to person ReID. It achieves outstanding performance and effectively preserves data privacy by transmitting model updates instead of centralizing data. However, FedReID heavily relies on the assumption that data has labels in clients. In real-world applications, annotating data is expensive, laborious, and time-consuming. Relying on labels is also not scalable for large-scale deployment. It takes three annotators two months to produce 126,441 bounding boxes of 4,101 identities from 180 hours of videos [30].

In this paper, we present a new federated unsupervised person ReID system, FedUReID, to learn person ReID models from unlabeled data while preserving data privacy. FedUReID stores data collected from multiple cameras in edges. Edges are coordinated by a cloud server to train models with unsupervised person ReID methods like Bottom-Up Clustering (BUC) [20]. FedUReID preserves data privacy because edges only access their local raw data. However, the training in edges suffers from statistical heterogeneity. Since data are collected from cameras in various locations, they would vary in the number of images, the number of identities, and data distributions affected by the data collection environment. These variances cause statistical heterogeneity [16, 44] among edges, affecting the performance of the system.

To tackle the statistical heterogeneity among edges, we propose joint optimization of cloud and edge to personalize models for each edge. In particular, we introduce three optimizations, personalized epoch (PE) and personalized clustering (PC) in the edges and personalized update (PU) in the cloud. (1) PE: unlike traditional FL that trains the same epochs in all training rounds, PE dynamically adjusts the
number of training epochs by early stopping a training round when conditions are satisfied. (2) PC: for unlabeled data, edges adopt a hierarchical clustering method to iteratively predict labels by merging similar clusters in each training round. It regards each data point as a cluster at the start. Unlike simple implementation that all edges merge the same percentage of clusters, PC customizes the merge percent for edges such that the numbers of final clusters are similar to their actual labeled number of identities. (3) PU: the cloud server aggregates model updates from edges to obtain a new model. Unlike traditional FL that directly updates models in edges with the aggregated model, PU updates these models with exponential moving average (EMA) to better adapt the aggregated model to each edge. We calculate the weighting of EMA using the similarity between the cloud aggregated model and the edge local models, measured by normalized Euclidean distance.

Extensive experiments and ablations on eight datasets demonstrate the effectiveness of FedUReID with joint optimization. We first construct the baseline with a simple combination of FL [44] and unsupervised person ReID training [20]. It does not outperform standalone training in all datasets. Standalone training means that each edge conducts unsupervised person ReID training [20] with its local dataset. In contrast, any single optimization method (PE, PC, or PU) outperforms standalone training and the baseline. Joint optimization of all optimization methods achieves the best performance. Compared with standalone training, it improves rank-1 accuracy by over 18% on the two smallest datasets [1, 29]. Moreover, compared with the baseline, it not only improves 4% on the two largest datasets [38, 40], but also reduces computation cost by 29%.

In this paper, we make the following contributions:

- We propose the first federated unsupervised person ReID system. It learns person ReID models without any labels while preserving data privacy.
- We propose joint optimization of cloud and edge to address the statistical heterogeneity among edges via personalization. In particular, we introduce personalized epoch, personalized clustering, and personalized update.
- We demonstrate the effectiveness of our proposed optimizations via extensive experiments and ablation studies.

The rest of the paper is organized as follows. In Section 2, we review related work about unsupervised person ReID and federated learning. Section 3 introduces our proposed FedUReID with joint optimization of cloud and edge. We present the experimental results and analyze the optimization methods via ablations in Section 4. In Section 5, we summarize the paper and provide future directions.

2 RELATED WORK

2.1 Unsupervised Person Re-identification

Person ReID aims to match a person in non-overlapping camera views. Supervised person ReID has achieved outstanding performance over the years of development [13, 22, 27, 34, 39]. Recently, unsupervised person ReID is receiving increased attention [7, 20, 35, 41]. Most unsupervised person ReID methods fall into two categories: unsupervised domain adaptation (UDA) and purely unsupervised.

Unsupervised Domain Adaptation for Person ReID UDA aims to learn a model that performs well for unlabeled data in a target domain, given labeled data in a source domain. These two domains differ in data distributions. On the one hand, some studies improve target domain performance by transferring image styles from the source to target domain [4, 21, 31] based on generative adversarial networks (GAN) [8]. On the other hand, some studies use clustering-based approach [7, 43] or graph matching [33] to generate pseudo labels for the unlabeled data. Besides, HHL [41] leverages camera invariance and domain connectedness to obtain a generalized model for the target domain. However, all these methods would impose potential privacy leakage because they require co-locating data from both domains.

Purely Unsupervised Person ReID Unlike UDA methods that assume some data has labels, purely unsupervised person ReID does not rely on any labels, which is even more challenging. Researchers mainly leverage bottom-up clustering methods [2, 20] to predict pseudo labels for the unlabeled data. These methods iteratively generate new pseudo labels and update the classifiers. However, these methods are not satisfactory for datasets with small data volumes. They require centralizing a large amount of data, which imposes potential privacy leakage risks. In this paper, built on the bottom-up clustering (BUC) [20] method for unsupervised person ReID training in each client, we propose joint optimization of cloud and edge to elevate performance while preserving data privacy.

2.2 Federated Learning

Federated learning (FL) is an emerging technique for training with decentralized data without privacy leakage [23]. FL trains models collectively from distributed clients under the collaboration of a central server. Federated person re-identification (FedReID) implements federated learning to person ReID [44]. It proposes Federated Partial Averaging (FedPav) to aggregate part of the models from clients. We integrate FedPav with BUC as the baseline.

Unsupervised Federated Learning The majority of studies on FL are based on supervised learning [23, 28, 44]. Recently, several studies investigate unsupervised FL [26, 36]. But these methods are inapplicable to person ReID because they mainly focus on learning generic representations.

Statistical Heterogeneity in FL Statistical heterogeneity is one of the key challenges of FL [12, 16]. It has attracted extensive research interests in recent years [17, 32, 37, 44]. Among them, some studies propose to personalize models for clients. These personalized federated learning methods are based on meta-learning [5, 6], multi-task learning [25], knowledge distillation [15, 44], etc. However, these methods assume that data has labels. Our proposed FedUReID introduces joint optimization of cloud and edge to personalize clients without any labels.

3 METHODOLOGY

In this section, we present FedUReID, a new federated unsupervised person ReID system with joint optimization of cloud and edge, to learn models without any labels while preserving privacy. It effectively tackles the statistical heterogeneity among edges.

3.1 Overview

We first provide an overview of FedUReID. Figure 1 depicts the system architecture of FedUReID. We embrace a hierarchical system
We design with a cloud server, multiple edges, and several cameras connected to each edge. Each edge is regarded as a client. FedUREID preserves data privacy because only the edge that collects data accesses it.

We design three flows for the training process: training flow, control flow, and clustering flow. At the start of the training flow, the server initializes a model and distributes it to clients. For each training round $r$, the server and clients collaborate to train models with four steps: (1) Local training: each client $k$ trains the concatenation of the model $\theta_k^r$ and a locally initialized classifier $\xi_k^r$; (2) Model upload: each client $k$ uploads the model $\theta_k^r$ to the server; (3) Model aggregation: the server aggregates these models to obtain a new global model $\theta^{r+1}$; (4) Model update: the server updates clients’ models for the next training round. Control flow and clustering flow interact with the training flow in the first step, local training. Before local training of the first training round, the control flow profiles the clients for clustering configurations. During local training, the control flow controls training according to training feedback. After local training, the clustering flow predicts new pseudo labels for the next round via hierarchical clustering with clustering configurations. Then, it updates the dimensions of the classifier.

Within these three flows, we propose joint optimization of cloud and edge: Personalized Epoch (PE) and Personalized Clustering (PC), and one optimization in the cloud—

![Figure 1: Architecture of our federated unsupervised person ReID system, FedUREID. FedUREID consists of three flows: training flow to learn models, control flow to control training, and clustering flow to predict pseudo labels for unlabeled data. We propose joint optimization of cloud and edge: personalized epoch and personalized clustering in the edges, and personalized update in the cloud.](image)

**Personalized Update** (PU). These optimizations address the statistical heterogeneity among clients. We summarize FedUREID with joint optimization in Algorithm 1. Next, we explain these optimizations in detail.

### 3.2 Client Design

Clients are responsible to perform unsupervised person ReID training. We adopt the hierarchical clustering algorithm [20] to train models with unlabeled data. The unlabeled data is collected from multiple connected cameras and stored in Data Store. As the cameras could be deployed in various places, clients could have large variations in the number of images, the number of identities, and data distributions, leading to statistical heterogeneity among these clients. To address it, we propose two components to personalize training in each edge: (1) a Profiler that generates customized clustering configurations; (2) a Controller that personalizes clustering with the configurations and reassigns computation throughout training according to training feedback.

#### 3.2.1 Hierarchical Clustering

At the end of each training round, each client uses hierarchical clustering [20] to predict labels for the next training round. The clustering flow is illustrated in Figure 1. Firstly, client $k$ extracts features using the trained model $\theta_k^r$. Then, these features are merged based on similarities to form new clusters, where each cluster is regarded as a label. Lastly, the classifier is updated with the new dimension equaling to the number of clusters.

![Figure 2: Illustration of hierarchical clustering. Each circle represents a sample. It starts with $M = 8$ samples (clusters) and merges $m = 2$ clusters each round.](image)

**Baseline Method and Problems** We use the direct combination of hierarchical clustering and federated learning as the baseline method. In particular, all clients first train $E$ local epochs in all rounds, and then cluster with the same merge percent $mp$. However, due to statistical heterogeneity of clients, such combination is problematic: (1) Using the same local epoch causes inappropriate computation assignment in different training rounds (Section 3.2.2). (2) Using the same merge percent causes inappropriate clustering paces among clients (Section 3.2.3). We analyze these two problems in detail and propose two optimization methods for them.
3.2.2 Personalized Epoch. We propose personalized epoch (PE) to reassign computations throughout training by dynamically adjusting the number of trained epochs each round. Training with larger local epochs consumes larger computation.

The majority of federated learning algorithms train for the same local epoch in all rounds [17, 23, 44]. Researchers design these algorithms for supervised learning. Since data is labeled in supervised learning, using the same computation drives training to achieve better performance, regardless of training rounds. However, data is unlabeled in our scenario, and the pseudo labels are predicted by hierarchical clustering every round. The amount of computation on different rounds would have various impacts on the performance.

Hence, we propose to reassign computation throughout training: using a large local epoch $E$ for the first training round and dynamically adjusting it according to training feedback for the remaining rounds. We first hypothesize that more computation in the first round is desirable. In the first round, each image is a cluster, which is regarded as a unique identity. Although training with these labels ignores intra-camera and inter-camera relationship of the same identity, it is helpful for learning representations of person ReID images. Compared with labels of the first round, labels predicted in later rounds could be incorrect in visually similar images [20]. Since the first training round is not affected by such incorrectness, we propose to train for larger epochs.

For the remaining rounds, we assign enough computation for clients to reach good training precision in each round. As the data statistics of clients and the clustering results of rounds are different, we propose a Controller to dynamically adjust the number of local epochs according to real-time training feedback. Specifically, we collect the training precision of each batch and send these feedbacks to the Controller after each training epoch. The Controller early stops a training round if any of the following conditions are satisfied: (1) the precision of any batch equals 100%; (2) the cumulative average precision of batches is larger than 95% (line 25-29 in Algorithm 1). This early-stop mechanism enables dynamic computation assignment among rounds, even among clients.

3.2.3 Personalized Clustering. We propose personalized clustering (PC) to enable clients to customize the merge percent $mp$ by approximating the number of identities via profiling. The merge percent $mp$ determines the number of merged clusters in each round ($m = n_k \times mp$), which controls the pace of clustering.

Clients should not use the same merge percent because their data vary in the number of images and identities. As the clusters are regarded as pseudo labels (identities), when the number of predicted identities falls below the actual number of identities, some labels are certainly wrongly predicted. It would cause performance drops. Since clients have different numbers of images and identities, they reach the number of clusters below the number of identities at different training rounds. Thus, they suffer from performance drop at different rounds. As such, the model obtained in the server by aggregating clients’ models would not be optimal because of the degraded performance of some clients’ models.

We propose to enable clients to personalize clustering paces to fit their characteristics of datasets. To mitigate the performance drop caused when the number of clusters is smaller than the actual number of identities $I_k$, a natural idea is to control client $k$ to finish training with $I_k$ clusters. In this way, by fixing the number of training rounds $R$, client $k$ obtains customized merged clusters each round $m = \frac{n_k - I_k}{p}$ However, this solution is not feasible in real-world scenarios — data is unlabeled, so the actual number of identities is unknown.

To this end, we design a Profiler to estimate the number of identities of clients. We profile clients before they start the first training round (line 20 in Algorithm 1). Profiling produces the number of merged clusters per round $m_k$ and the merge percent $mp_k$ for client $k$. To minimize the computation overhead caused by profiling, we conduct unsupervised person ReID for each client using larger merge percent $mp_{profile}$ (i.e., fewer rounds) and smaller
local epochs $E_{profile}$. After profiling, we analyze the results and choose the round $r_{profile}$ that achieves the best accuracy. For each client, we select the number of clusters $M_{profile}$ in round $r_{profile}$ as the estimated number of identities. As a result, we calculate the merged clusters $m_k$ and the merge percent $mp_k$ for client $k$ with the following formula:

$$m_k = \frac{n_k - M_{profile}}{R},$$

$$mp_k = \frac{m_k}{n_k}$$

where $R$ is the total training rounds and $n_k$ is the data volume of client $k$. The Profiler sends these three configurations to the Controller to control the training flow and clustering flow.

### 3.3 Server Design

The cloud server is responsible for coordinating clients to conduct training. It aggregates models trained in clients and updates clients with a new aggregated model for the next training round. To address the statistical heterogeneity of clients, we propose an optimization method, Personalized Update (PU), to adapt the aggregated model for clients.

#### 3.3.1 Model Aggregation

At the end of each training round $r$, the server aggregates models uploaded from clients with weighted averages. The weightage of client $k$ depends on its data volume $n_k$. For $K$ participated clients, the model aggregation formula is as follows:

$$\theta^{r+1} = \sum_{k=1}^{K} \frac{n_k}{n} \theta^{r}_k$$

where $n = \sum_{k=1}^{K} n_k$ is the total data volume. The global model $\theta^{r+1}$ is the generalized model, which is deployable to other scenarios. We also evaluate $\theta^{r+1}$ on test sets of all clients.

#### 3.3.2 Personalized Update

We propose Personalized Update (PU) to adapt the aggregated model to client $k$ by updating the local model in client $k$ with exponential moving average (EMA) of the global model. The weighting of EMA is measured by the similarity between the global and local two models.

Standard federated algorithms like FedAvg [23] simply updates the models in clients by replacing them with the global model. However, the global model may not fit all clients [17] because of statistical heterogeneity among clients — the scenes (indoor or outdoor) and illumination could be different. To mitigate statistical heterogeneity, we incorporate the local model $\theta^{r}_k$ of client $k$ in round $r$ by updating it with an exponential moving average. We formulate the personalized update as follows:

$$\theta^{r+1}_k = \mu \theta^{r}_k + (1 - \mu) \theta^{r+1}$$

where $\mu$ is the weighting ranging from [0, 1]. $\mu$ determines the importance of the local model $\theta^{r}_k$ and global model $\theta^{r+1}$ in update. Instead of setting $\mu$ as constant in all training rounds in all clients, We propose to calculate $\mu$ based on the similarity of the global and local model (line 9-12 in Algorithm 1): (1) Calculate the Euclidean distance of each layer of these two models; (2) Normalize distances of layers to [0, 1]; (3) Average these distances as $\mu$. The intuition of calculating $\mu$ is to increase the importance of the local model when the local model is not similar to the global model. As such, the updated model $\theta^{r+1}$ retains more historical information of the local model, which is more personalized to the data in the client.

### 4 EXPERIMENTAL RESULTS

In this section, we start by describing the experimental setup. We then present the overall performance of FedUReID. We end by analyzing the optimization methods via ablation studies.

#### 4.1 Experimental Setup

We first present the experimental setups, including datasets, evaluation metrics, implementation details, and experimental settings.

**Datasets** We evaluate our experiments with eight person ReID datasets adopted from the benchmark, FedReIDBench [44]. Table 1 presents the statistics of the datasets. These datasets vary in the number of images and identities. As the datasets are collected from various locations, the visual appearances of them are also quite different, as shown in Figure 3. These characteristics of datasets simulate the statistical heterogeneity in real-world scenarios.

**Evaluation Metrics** We evaluate the performances with standard ReID evaluation metrics and computation cost. To evaluate the performances of ReID, we use the two most common evaluation metrics: Cumulative Match Characteristic (CMC) curve and mean Average Precision (mAP) [39]. Given an image as a query, CMC first ranks gallery images by similarity (from most similar to least similar). It then compares whether the ranked top-k images match the query image. The probability of such matching is denoted as rank-k accuracy. We report the results of rank-1, rank-5, and rank-10 accuracy. We also evaluate the performance with mAP, which measures the mean average precision of all queries.

We measure the computation cost by the number of local epochs. Although the computation cost of the classifier varies in rounds, it is
Table 2: Performance comparison of FedUReID with the baseline and unsupervised domain adaptation (UDA) methods on the two largest datasets. Despite that FedUReID does not need any labels like UDA methods, it outperforms all other methods by 4% and 3% on rank-1 accuracy of DukeMTMC-reID and Market-1501 datasets, respectively.

| Methods            | Types                      | Market-1501 Dataset [38] (%) | DukeMTMC-reID Dataset [40] (%) |
|--------------------|----------------------------|------------------------------|--------------------------------|
|                    |                            | Rank-1 | Rank-5 | Rank-10 | mAP  | Rank-1 | Rank-5 | Rank-10 | mAP  |
| PUL [7]            | Domain Adaptation          | 44.7   | 59.1   | 65.6    | 20.1 | 30.4   | 46.4   | 50.7    | 16.4 |
| SPGAN [4]          | Domain Adaptation          | 58.1   | 76.0   | 82.7    | 26.7 | 46.9   | 62.6   | 68.5    | 26.4 |
| HHL [41]           | Domain Adaptation          | 62.2   | 78.8   | 84.0    | 31.4 | 46.9   | 61.0   | 66.7    | 27.2 |
| BUC [20] (Standalone) | Purely Unsupervised      | 61.9   | 73.5   | 78.2    | 29.6 | 40.4   | 52.5   | 58.2    | 22.1 |
| Baseline           | Purely Unsupervised        | 60.5   | 73.3   | 77.9    | 27.4 | 47.0   | 58.3   | 64.1    | 25.2 |
| FedUReID (Ours)    | Purely Unsupervised        | 65.2   | 77.8   | 82.2    | 34.2 | 51.0   | 62.4   | 67.6    | 29.5 |

Figure 4: Performance (rank-1 accuracy) comparison of standalone training, the baseline, and FedUReID. FedUReID achieves the best performance on all datasets.

Table 3: Ablation studies of our proposed optimization methods: personalized clustering (PC) and personalized epoch (PE) in the edges; personalized update (PU) in the cloud; joint optimization of cloud and edge.

| Datasets         | Baseline | Edge | Cloud | Joint |
|------------------|----------|------|-------|-------|
|                  | PC       | PE   | Both  | PU    | All  |
| DukeMTMC-reID[40]| 47.0     | 48.3 | 49.5  | 50.4  | 49.2 | 51.0 |
| Market-1501[38]  | 60.5     | 62.5 | 64.0  | 65.1  | 62.2 | 65.2 |
| CUHK03-NP[19]    | 7.8      | 8.4  | 7.9   | 8.1   | 8.8  | 8.9  |
| PRID2011[11]     | 31.0     | 34.0 | 35.0  | 37.0  | 36.0 | 38.0 |
| CUHK01[18]       | 34.8     | 39.3 | 39.2  | 42.6  | 35.4 | 43.6 |
| VIPeR[9]         | 21.8     | 24.4 | 24.4  | 24.7  | 22.5 | 26.6 |
| 3DPeS[1]         | 63.8     | 65.5 | 64.6  | 67.5  | 65.0 | 65.5 |
| iLIDS-VID[29]    | 71.4     | 73.5 | 70.4  | 70.4  | 72.5 | 73.5 |

4.2 Performance Comparison

We demonstrate the effectiveness of joint optimization by comparing FedUReID with standalone training, the baseline, and several unsupervised domain adaptation (UDA) methods. Standalone training means that each client performs unsupervised person ReID training [20] with its dataset — not collaborating with other clients. It is only meaningful for a client to participate in federated learning (FL) if the performance is better than its standalone training. The baseline method is the simple combination of FL [44] and unsupervised person ReID [20], as described in Section 3.2.1.

Figure 4 compares the rank-1 accuracy of FedUReID with standalone training and the baseline. The standalone training is better than the baseline on Market-1501 [38] and PRID2011 [11] datasets. These results indicate that combining FL and unsupervised person ReID is not trivial. It requires deep understanding and analysis to optimize the performance. Our proposed FedUReID, with joint optimization of cloud and edge, outperforms both standalone training and the baseline in all datasets. Compared with standalone training, another insight is that FL-based methods significantly improve performance on smaller datasets (less than 2,000 training images). For example, the improvement is over 26% on the 3DPeS [1] dataset.

In addition, Table 2 compares FedUReID with three UDA methods: PUL [7], SPGAN [4], and HHL [41], on the two largest datasets. UDA methods improve the performance for unlabeled data in a target domain by heavily relying on plenty of labeled data in a source domain. Despite that FedUReID does not need any labels, it effectively improves the performance by at least 3% and 4% on rank-1 accuracy of the Market-1501 [38] and DukeMTMC-reID [40] datasets.
4.3 Ablation Studies

We conduct ablation studies on the baseline, three proposed optimization methods (PE, PC, and PU), and the combinations of these methods. These ablation studies demonstrate the effectiveness of our proposed optimizations.

We first present the rank-1 accuracy comparison of edge optimizations, the cloud optimization, and joint optimization of cloud and edge. Table 3 shows that joint optimization achieves the best overall performance. Also, the client optimization with both PC and PE outperforms either one alone. Besides, any single optimization outperforms the baseline almost on all datasets. Although the performance of PE on iLIDS-VID [29] is slightly lower than the baseline, it is still much better than standalone training. These results demonstrate that our optimization methods effectively elevate the performance. Next, we analyze the baseline and these optimizations in detail.

Baseline Method We select the best setting for the baseline method by comparing performances of different local epochs $E = \{1, 5, 10, 15, 20\}$. Larger local epoch requires higher computation. Following the setting in BUC [20], we fix the merge percent $mp = 0.05$. Figure 5 shows that $E = 5$ performs best on rank-1 accuracy. The performance of $E = 10$ is also comparable, but it costs 2x more computation than $E = 5$. Thus, we choose $E = 5$ as our baseline. Besides, the accuracy decreases as increasing computation from $E = 5$ to $E = 20$. This insight indicates that simply increasing computation in all rounds harms the performance. Based on this insight, we propose personalized epoch to reassign computation across different training rounds.

Effectiveness of Personalized Epoch We evaluate personalized epoch (PE) in two folds. Firstly, we evaluate reassigning larger computation to the first training round. For a fair comparison, we maintain the computation similar to the baseline $E = 5$. We train for 20 epochs for the first round and $4 = \lfloor \frac{5 \times 20 - 20}{19} \rfloor$ epochs for the remaining 19 rounds, denoted as $E = 20, 4$. Secondly, we evaluate dynamically adjusting epochs with an early stop in the remaining rounds, as described in Section 3.2.2. We use 20 local epochs for the first round and dynamic local epochs for the remaining 19 rounds.

Table 4 compares performances of these three settings of local epochs. A larger epoch on the first round ($E = 20, 4$) has better performance than the baseline, except for a small gap on CUHK03 datasets, respectively. FedUReID is also superior to other methods on rank-5 accuracy, rank-10 accuracy, and mAP.

Figure 5: Performance (rank-1 accuracy) comparison of different values of local epochs $E$. We select $E = 5$ as the baseline as it performs better with lower computation costs.

| Datasets    | duke | market | cuhk03 | prid | cuhk01 | viper | 3dpes | iLIDS | VID |
|-------------|------|--------|--------|------|--------|-------|-------|-------|-----|
| Baseline    | 47.0 | 60.5   | 7.8    | 31.0 | 34.8   | 21.8  | 63.8  | 71.4  |     |
| $E = 20, 4$ | 47.9 | 63.6   | 7.4    | 32.0 | 35.4   | 22.8  | **66.3** | **71.4** |     |
| PE          | 49.5 | 64.0   | **7.9** | 35.0 | 39.2   | 24.4  | 64.6  | 70.4  |     |

Table 4: Performance (rank-1 accuracy) comparison of different local epochs. $E = 20, 4$ means training 20 epochs for the first round and 4 epochs for the rest rounds. Personalized epoch (PE) achieves the best overall performance.

Table 5: Validation of personalized clustering (PC): (1) The motivation of PC is that clients achieve the best performance at various rounds (Best Round); (2) The profiled numbers of identities are close to the labeled ones; (3) The performance of PC using profiled number of identities is comparable to using the labeled ones.

| Datasets    | duke | market | cuhk03 | prid | cuhk01 | viper | 3dpes | iLIDS | VID |
|-------------|------|--------|--------|------|--------|-------|-------|-------|-----|
| Best Round  | 18   | 17     | 14     | 17   | 9      | 0     | 7     | 17    |     |
| Number of clusters/identities: | | | | | | | | |
| Labeled     | 702  | 751    | 767    | 285  | 485    | 316   | 93    | 59    |     |
| PC          | 670  | 528    | 886    | 156  | 855    | 432   | 90    | 13    |     |
| Rank-1 accuracy using the number of identities above: | | | | | | | | |
| Baseline    | 47.0 | 60.5   | 7.8    | 31.0 | 34.8   | 21.8  | 63.8  | 71.4  |     |
| Labeled     | 47.7 | **63.9** | 8.3    | 32.0 | 36.8   | 22.8  | **66.3** | **74.5** |     |
| PC          | **48.3** | 62.5   | **8.4** | 34.0 | **39.3** | **24.4** | 65.5  | 73.5  |     |

Figure 6: Analysis of personalized update (PU). Measuring the similarity of global and local models with $\mu$. Larger $\mu$ means less similar, i.e., stronger personalization. Models are more personalized at the end of training.

[19] dataset. PE (with early stop mechanism) generally outperforms $E = 20, 4$. Moreover, PE needs much less computation than $E = 20, 4$; $E = 20, 4$ needs in total 768 epochs; while PE costs only 479 epochs (Table 6), around 38% lower than. These results indicate that PE achieves better performance with lower computation costs.

Effectiveness of Personalized Clustering We evaluate personalized clustering (PC) in three folds. Firstly, we validate our motivation that clients should personalize clustering pace. Best round in Table 5 represents the round that standalone training achieves the best performance when the merge percent is the same for all clients. The best round of clients are different, so we propose PC
to customize their merge percent. Secondly, we present that the profiled number of identities is similar to the actual labeled ones (Table 5). To minimize computation and ensure performance at the same time, we profile with using following settings: merge percent \( m_{profile} = 0.08 \), total rounds \( r_{profile} = 12 \), and 5 epochs for the first round and 1 epoch for the remaining rounds \( E_{profile} = 5 \). Profiling incurs 16 epochs of extra computation in each client, 128 epochs in total. Thirdly, Table 5 also shows that PC (using the profiled number of identities) is comparable to using the actual labeled number of identities, both outperforming the baseline. It demonstrates the effectiveness of profiling and PC.

**Effectiveness of Personalized Update** we analyze personalized update (PU) by understanding changes in similarities between the global model and local models. The similarity is measured by \( \mu \) as described in Section 3.3.2. Two models are less similar if the value of \( \mu \) is larger, so larger \( \mu \) means stronger personalization. Figure 6 shows the changes of similarity of four datasets throughout the training process. All these datasets share the same trend in two stages: (1) \( \mu \) decreases in the first few training rounds — the similarity increases as clients are learning knowledge from the others; (2) \( \mu \) increases in the second stage — the similarity decreases as clients personalize models for their local datasets. Models are more personalized at the end of training. It validates the effectiveness of PU to personalize the global model to models in edges. Besides, PU effectively improves the performance, as compared in Table 3. We provide the results of other datasets in the supplementary.

**5 CONCLUSION**

In this paper, we present FedUReID, a new federated unsupervised ReID system to learn models without any labels while preserving privacy. To address the statistical heterogeneity among edges, we propose joint optimization of cloud and edge to personalize models for each edge. For optimizations in edge, we design a Controller to support personalized epoch and a Profiler to facilitate personalized clustering. For the optimization in the cloud, we introduce personalized update to adapt the cloud aggregated models to edges. Extensive empirical studies demonstrate that FedUReID effectively elevates performance on all datasets and reduces computation cost by 29%. For future work, we will consider the system heterogeneity among edges.

**ACKNOWLEDGMENTS**

This study is supported by 1) supported under the RIE2020 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from the industry partner(s); 2) the National Research Foundation, Singapore, and the Energy Market Authority, under its Energy Programme (EP Award <NRF2017EW7-EP003-023>); 3) Singapore MOE under its Tier 1 grant call, Reference number R096/20.
