FedCAPR:Federated Camera-Aware Unsupervised Person Re-Identification with Identity-Distributed Equalization for Decentralized Data Clustering

Supplementary Material

6. Computation and Communication Costs

Table 6 compares the computational and communication costs of our proposed method with state-of-the-art methods in unsupervised federated person re-identification. The computational cost is defined as the total number of training epochs, calculated by the product of local training epochs per round and the total number of communication rounds, derived as:

$$\text{Computational Cost} = \text{num of local epochs} \times \text{num of rounds} \tag{14}$$

This metric provides a comprehensive measure of resource consumption, particularly valuable for real-world deployments where computational resources are often limited.

The communication cost refers to the total number of communication rounds during federated learning training.

FedCAPR demonstrates significantly lower costs compared to recent works [21, 32, 33, 37]. Specifically, it reduces computation costs by 3.9x, 3.4x, and 7.6x compared to FedUReID, FedUCA, and FedUCC(Weng et al., 2022), respectively. Additionally, FedCAPR achieves a 15x and 4x reduction in communication rounds compared to FedUCC(Weng et al., 2022) and FedUCC(Weng et al., 2023). While the communication cost of our method is higher than that of FedUCA, we achieve significant reductions in computational cost. Moreover, FedCAPR outperforms in accuracy, as shown in Table 2.

Table 6. Comparison of computational and communication costs across different methods.

Methods	Computation Cost (Epoch)	Communication Cost (Round)
FedUReID [21]	566	20
FedUCC [37]	1100	300
FedUCC [32]	160	80
FedUCA [33]	500	10
FedCAPR(Ours)	145	20

7. Hyperparameter Sensitivity Analysis

In Table 7, we illustrate the impact of different coefficient values (γ) on the camera-aware loss with respect to Rank-1 accuracy. We experimented with γ values of 0.25, 0.5, 0.6, 0.7, and 0.75. It is shown that, for most datasets, the accuracy drop tends to be smaller as γ increases, which is particularly evident in datasets such as Duke, Market, CUHK03, VIPeR, and CUHK01. Notably, when $\gamma = 0.7$, the accuracy is 7% higher than at $\gamma = 0.75$ in the PRID dataset. Therefore, we chose $\gamma = 0.7$ as a balanced coefficient in our experiments, offering a good trade-off between regularization strength and model performance.

Table 7. Impact of Coefficients in Camera-Aware Loss.

γ	0.25	0.5	0.6	0.7	0.75
Duke	83.3	84.0	83.8	83.6	83.9
Market	91.6	93.1	92.7	92.7	**92.8**
iLIDS	78.6	78.6	79.6	**81.6**	78.6
CUHK03	72.5	72.9	72.3	73.4	**73.6**
Prid	77.0	82.0	84.0	**85.0**	78.0
VIPeR	62.3	67.7	66.1	**68.7**	**68.7**
CUHK01	95.1	95.0	95.2	**95.6**	95.4
3DPeS	83.7	85.8	83.7	**87.4**	85.8

Table 8. Impact of Coefficients in Cosine Similarity Loss for Regularization.

δ	0.1	0.25	0.5	0.75
Duke	**83.7**	83.6	83.3	83.6
Market	93.1	92.7	93.2	**93.5**
iLIDS	77.6	**81.6**	79.6	**81.6**
CUHK03	**74.2**	73.4	73.1	73.1
Prid	82.0	**85.0**	80.0	82.0
VIPeR	65.5	**68.7**	66.5	65.2
CUHK01	95.7	95.6	95.5	**95.8**
3DPeS	86.6	**87.4**	83.7	83.7

In Table 8, we illustrate the impact of different coefficient values (δ) on the cosine-similarity regularization loss with respect to Rank-1 accuracy. The evaluated coefficients include 0.1, 0.25, 0.5, and 0.75. It can be observed that most datasets experience the largest accuracy drop when $\delta = 0.5$, with PRID and 3DPeS showing this trend most clearly. Notably, when $\delta = 0.25$, the PRID dataset achieves a 5% higher accuracy compared to $\delta = 0.5$. Based on the relatively stable performance on 3DPeS, CUHK03, and CUHK01 when $\delta = 0.25$, we selected 0.25 as the final coefficient value.

8. Robustness of IDE Mechanism

The purpose of this experiment is to demonstrate the robustness of the IDE mechanism. We evaluate the performance impact of varying the number of images per person identity(pid) in the IDE mechanism, as shown in Tab. 10. The

Table 9. Comparison of the accuracy(%) with Existed Unsupervised Federated Person Re-ID.

Methods	MSMT17		Market		iLIDS		CUHK03		PRID		VIPeR		CUHK01		3DPeS	
	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP
FedUReID [21]	-	-	65.2	-	73.5	-	8.9	-	38.0	-	26.6	-	43.6	-	65.5	-
FedUCC [37]	-	-	86.5	65.5	74.7	59.7	9.6	9.7	58.9	63.1	31.3	36.7	78.3	75.3	68.9	50.9
FedUCC [32]	**60.9**	30.4	90.3	75.2	**82.8**	72.0	38.7	35.5	69.0	72.0	43.0	48.6	80.1	76.6	73.2	57.8
FedUCA [33]	-	-	**92.5**	**79.4**	80.5	-	50.0	-	75.0	-	51.0	-	86.0	-	**85.0**	-
FedCAPR(Ours)	59.56	**30.5**	90.5	77.3	76.5	**76.5**	**65.5**	**62.1**	**77.0**	**82.2**	**61.7**	**68.1**	**94.0**	**93.5**	80.9	**73.6**

Table 10. Performance evaluation(%) for different number of image/pid.

Number of images per pid	Duke		Market		iLIDS		CUHK03		Prid		VIPeR		CUHK01		3DPeS	
	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP
Initial dataset	83.4	69.8	92.7	81.6	80.6	76.6	66.9	62.7	**86.0**	**89.7**	53.8	62.7	89.2	88.0	86.5	80.0
10	83.4	69.8	92.8	**82.5**	79.6	77.9	**74.9**	**70.8**	84.0	88.2	66.8	73.0	95.4	94.7	86.6	**81.2**
15	83.1	69.8	92.7	81.8	79.6	75.5	73.4	69.7	84.0	87.4	65.3	73.5	95.3	94.7	**87.8**	79.9
20	**83.6**	69.6	92.7	**82.5**	81.6	77.8	73.4	69.7	85.0	88.4	**68.7**	**75.8**	**95.6**	**95.1**	87.4	79.2
25	83.2	**70.0**	92.5	81.6	**82.7**	75.4	73.4	69.5	84.0	87.3	64.2	71.4	93.7	92.7	83.3	78.0
30	83.0	69.5	**93.0**	82.0	81.6	**79.3**	68.6	65.1	83.0	87.1	63.9	70.6	93.7	92.7	84.2	76.7

Table 11. Comparison of Traditional Federated Learning Methods with FedProx.

Methods	FedProx		FedCAPR(Ours)	
	mAP	R1	mAP	R1
Duke	65.9	79.7	69.6	83.6
Market	77.4	90.6	82.5	92.7
iLIDS	72.4	75.5	77.8	81.6
CUHK03	60.7	63.2	69.7	73.4
Prid	77.4	73.0	88.4	85.0
VIPeR	62.7	53.5	75.6	68.7
CUHK01	93.0	93.3	95.1	95.6
3DPeS	76.3	82.1	79.2	87.4

results indicate that aligning the number of images to 20 achieves the best performance, but slight deviations (e.g., between 10 and 30) do not significantly degrade accuracy. This demonstrates that the IDE mechanism is resilient to minor variations in alignment while still providing sufficient standardization to improve clustering quality.

9. Comparison with FedProx [19]

As shown in Table 11, our proposed FedCAPR consistently surpasses FedProx across all benchmark datasets in both mAP and Rank-1 accuracy. Unlike FedProx, which merely adopts an L_2-norm based regularization by minimizing the difference between global and local models, FedCAPR introduces a camera-aware cosine similarity loss to enhance inter-client generalization under heterogeneous conditions. Notably, our method achieves significant improvements on challenging datasets such as VIPeR (mAP: $+12.9\%$, R1: $+15.2\%$), CUHK03 (mAP: $+9.0\%$, R1: $+10.2\%$), and PRID (mAP: $+11.0\%$, R1: $+12.0\%$). Even on relatively saturated datasets like CUHK01 and 3DPeS, FedCAPR still yields noticeable gains, highlighting its robustness and generalizability across diverse domains. These results demonstrate the effectiveness of our cosine-based regularization strategy in addressing real-world data heterogeneity.

10. Extended Analysis with MSMT17 [52]

To further evaluate the generalizability of our proposed method, we replace the commonly used DukeMTMC-reID dataset with MSMT17, which contains a larger number of cameras and greater scene diversity. As shown in Table 9, FedCAPR consistently achieves superior performance across various person re-identification benchmarks when compared with recent federated methods including FedUReID [21], FedUCC(Weng et al., 2022) [37], FedUCC(Weng et al., 2023) [32], and FedUCA [33]. Specifically, FedCAPR obtains the highest mAP on MSMT17 (30.5%) and achieves state-of-the-art accuracy on most small-scale datasets such as iLIDS (R1: 76.5%, mAP: 76.5%), CUHK03 (R1: 65.5%, mAP: 62.1%), PRID (R1: 77.0%, mAP: 82.2%), and VIPeR (R1: 61.7%, mAP: 68.1%). These results demonstrate the robustness of our camera-aware regularization design in handling data heterogeneity across both large and small domains. Despite the slight Rank-1 difference on MSMT17 compared with FedUCC (59.56% vs. 60.9%), FedCAPR still achieves the best trade-off between accuracy and consistency across diverse datasets.