FEDHRA: FEDERATED HARD-RELIABLE RECOVERY AND FAIR AGGREGATION FOR PERSON RE-IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Federated learning (FL) enables privacy-preserving person re-identification (ReID) without sharing raw pedestrian images. However, severe ReID domain skew causes client-wise performance unfairness: global models often favor easy clients with clear pedestrian appearances, while hard clients affected by occlusion, blur, poor illumination, and complex backgrounds remain under-optimized. Existing fair FL methods typically compensate under-performing clients by increasing their aggregation weights or suppressing conflicting updates. Such strategies are unreliable for ReID because hard-client updates contain both useful domain-specific identity cues and noisy visual components. Blindly amplifying them may strengthen noise, whereas aggressively removing conflicts may discard valuable hard-domain information. Fair federated ReID therefore requires reliability-aware protection of hard-client updates rather than uniform compensation. We propose **FedHRA**, a **F**ederated **H**ard-**R**eliable Recovery and Fair **A**ggregation framework. FedHRA identifies reliable hard clients by jointly measuring client hardness and update reliability, aligns their updates with the easy-client consensus, and selectively recovers reliable hard-domain components. Reliability-guided aggregation further prevents the corrected updates from being diluted. Experiments on four ReID benchmarks with ResNet50 and ViT demonstrate that FedHRA improves hard-client retrieval, overall performance, and cross-client fairness.

1 INTRODUCTION

Federated learning (FL) aims to collaboratively train a shared model while preserving data privacy (McMahan et al., 2017). Instead of collecting raw data from different clients, FL allows each client to train locally and upload model parameters or updates to the server for aggregation. This privacy-preserving property makes FL suitable for person re-identification (ReID), where pedestrian images are usually collected from different camera networks and may contain sensitive personal information. Existing ReID benchmarks, such as Market-1501 (Zheng et al., 2015), MSMT17 (Wei et al., 2018), CUHK02 (Li & Wang, 2013), and CUHK03 (Li et al., 2014), show large variations in illumination, viewpoints, backgrounds, resolutions, occlusions, and pedestrian poses. When these datasets or camera systems are treated as different clients, federated ReID naturally suffers from severe domain skew across clients.

Under domain skew in federated ReID, different clients often present inconsistent identity matching difficulty, resulting in a client-wise performance fairness problem. Existing federated ReID methods mainly pursue high average retrieval accuracy, but high average performance does not necessarily indicate that all ReID clients are fairly optimized. In practical ReID systems, each client may correspond to an independent camera network, and poor performance on hard clients can lead to unreliable retrieval under occlusion, blur, low illumination, and complex backgrounds. Therefore, federated ReID should not only improve the average retrieval performance, but also ensure that hard clients are not consistently under-served by the global model. As illustrated in Figure 1, easy clients usually contain clearer pedestrian appearances and more stable identity boundaries, so their updates are more concentrated and can form dominant aggregation directions. In contrast, hard clients affected by challenging visual conditions often contain ambiguous identity cues and unstable hard samples, making their update directions more scattered and inconsistent. During global aggrega-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tion, these hard-client updates are easily counteracted by dominant easy-client updates, and hard-domain-specific ReID information is insufficiently preserved. As a result, the global model tends to favor easy clients while under-optimizing hard clients, leading to obvious performance disparity among different ReID domains. This issue is closely related to performance fairness in federated learning (Chen et al., 2023; Li et al., 2019; Mohri et al., 2019; Cui et al., 2021; Li et al., 2021b).

Although fair federated learning has been studied to reduce client-wise performance disparity in general FL scenarios, existing fair FL methods cannot be applied to federated ReID. Most of them compensate under-performing clients according to losses, performance gaps, or update conflicts. However, these task-agnostic signals only indicate which clients perform poorly or which updates are inconsistent, but they cannot distinguish the internal composition of hard-client ReID updates. As shown in Figure 1, hard-client updates are not simply weak or harmful updates. On the one hand, they may contain useful local identity cues for hard domains, such as partial body regions, low-resolution details, and occlusion-related appearance patterns. On the other hand, they may also contain noisy components caused by background bias, detection errors, and unstable hard samples. Therefore, simply increasing the contribution of under-performing clients may amplify unreliable visual noise, while suppressing conflicting updates may mistakenly remove hard-domain-specific identity information. This motivates us to formulate **Fair Federated ReID** as a reliability-aware hard-client update protection problem: reliable hard-client updates should be identified, corrected, and preserved during aggregation, instead of being blindly amplified or completely suppressed.

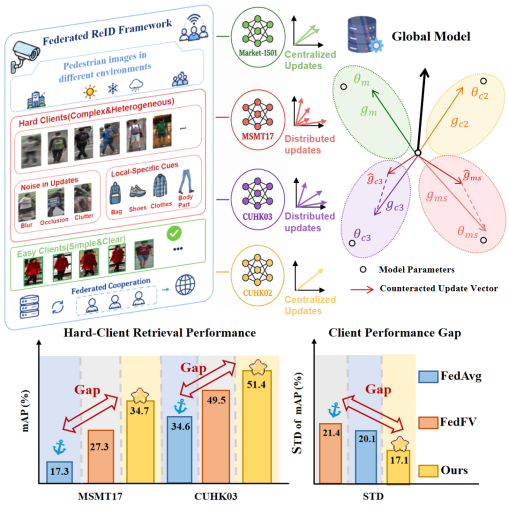


Figure 1: Illustration of update conflict and aggregation bias in federated ReID. Dominant easy-client updates pull the global update toward stable directions, while scattered hard-client updates are counteracted and under-preserved during aggregation, leading to degraded hard-client performance and larger client-wise performance gaps.

To address these issues, we propose **FedHRA**, a **F**ederated **H**ard-**R**eliable Recovery and Fair **A**ggregation framework for person re-identification. FedHRA aims to protect reliable hard-client updates while maintaining the stable identity-discriminative update learned from easy clients. Specifically, we first design a hard-reliable client scoring strategy, which jointly considers client difficulty and update reliability to identify hard clients that truly deserve compensation. Then, selected hard-client updates are aligned with the easy-client consensus update, and part of their original updates is recovered according to the hard-reliable score. Finally, reliability-guided fair aggregation assigns suitable weights to reliable hard clients, preventing their corrected updates from being diluted again during global aggregation. Different from conventional conflict-based correction or performance-based reweighting, FedHRA explicitly considers whether hard-client updates contain reliable ReID information worth preserving.

The main contributions of this paper are summarized as follows:

- We formulate **Fair Federated ReID** as a new task under ReID domain skew, and reveal that reliable hard-domain update dilution and unreliable fair compensation are key causes of client-wise performance disparity.
- We propose **FedHRA**, a **F**ederated **H**ard-**R**eliable Recovery and Fair **A**ggregation framework, which identifies reliable hard clients, corrects their updates through consensus-guided projection and recovery, and protects them with reliability-guided fair aggregation.
- Extensive experiments on multiple ReID benchmarks demonstrate that FedHRA outperforms existing fair federated learning methods under both CNN-based and Transformer-based backbones, achieving better retrieval accuracy and client-wise fairness.

2 Related Work

2.1 Federated Person Re-Identification

Person re-identification (ReID) aims to match the same pedestrian across different cameras. Although centralized ReID methods have achieved strong performance, collecting pedestrian images from different camera networks may lead to privacy risks. Federated learning provides a privacy-preserving solution by collaboratively training a shared model without sharing raw data. General federated learning methods, such as SCAFFOLD (Karimireddy et al., 2020), MOON (Li et al., 2021a), and FedProx (Li et al., 2020), are often adopted to improve optimization stability under heterogeneous client data. Existing federated ReID methods have explored benchmark construction (Zhuang et al., 2020), decentralized multi-domain learning (Wu & Gong, 2021), data augmentation (Song et al., 2024), pose-guided feature learning (Oh et al., 2026), style utilization (Xu et al., 2025), domain-aware aggregation (Yu et al., 2026), and robust knowledge selection (Xu et al., 2026). However, these methods mainly focus on improving average performance or cross-domain generalization. Under domain heterogeneity, applying FedAvg (McMahan et al., 2017) to federated ReID may cause the global model to be dominated by easy clients, while reliable hard-client updates are insufficiently preserved during global aggregation. In contrast, FedHRA focuses on reliable fairness for federated ReID by protecting reliable hard-client updates while maintaining the stable consensus learned from easy clients throughout heterogeneous federated optimization rounds.

2.2 Fair Federated Learning

Fairness in federated learning has attracted increasing attention (Chen et al., 2023; Horváth et al., 2021; Qi et al., 2022). Existing studies mainly focus on Performance Fairness (Jiang et al., 2023; Li et al., 2019; 2021b; Mohri et al., 2019), Individual/Group Fairness (Cui et al., 2021; Ezzeldin et al., 2023; Zeng et al., 2021), and Collaborative Fairness (Jiang et al., 2023; Lyu et al., 2020; Ray Chaudhury et al., 2022; Xu et al., 2021; Zhou et al., 2021). This paper focuses on Performance Fairness, which aims to prevent certain clients from being consistently under-served by the global model. AFL (Mohri et al., 2019) improves the worst-performing clients with a min-max objective. q-FedAvg (Li et al., 2019) assigns larger weights to high-loss clients. FedFV (Wang et al., 2021) modifies conflicting updates to reduce unfair degradation. Ditto (Li et al., 2021b) improves client performance through personalized models, while FedCE (Jiang et al., 2023) considers both performance fairness and collaborative fairness. Recent methods further study heterogeneous fairness correction and personalized fairness weighting, such as FedHEAL (Chen et al., 2024) and FedPW (Guan et al., 2025). However, these methods are mainly designed for general FL scenarios and usually ignore the reliability of hard-client updates in federated ReID. In federated ReID, hard-client updates may contain both useful identity-related components and noisy components. Directly increasing their weights may amplify noise, while only relying on stable easy-client updates may overwrite hard-domain-specific information. Therefore, FedHRA introduces hard-client reliability into update correction and fair aggregation to protect reliable hard-client information.

3 Preliminaries

Following the standard federated learning objective (Li et al., 2019; McMahan et al., 2017; Miao et al., 2023), we consider M clients, where client m owns a private ReID dataset $\mathcal{D}_m$. The global model w is optimized by

$$\min_{w} F(w) = \sum_{m=1}^{M} p_m f_m(w), \tag{1}$$

where $f_m(w)$ is the local objective and p_m is the aggregation weight. At communication round t, client m starts from w^t, performs local ReID training, and returns the shared-parameter update g_m^t to the server without sharing private identity labels.

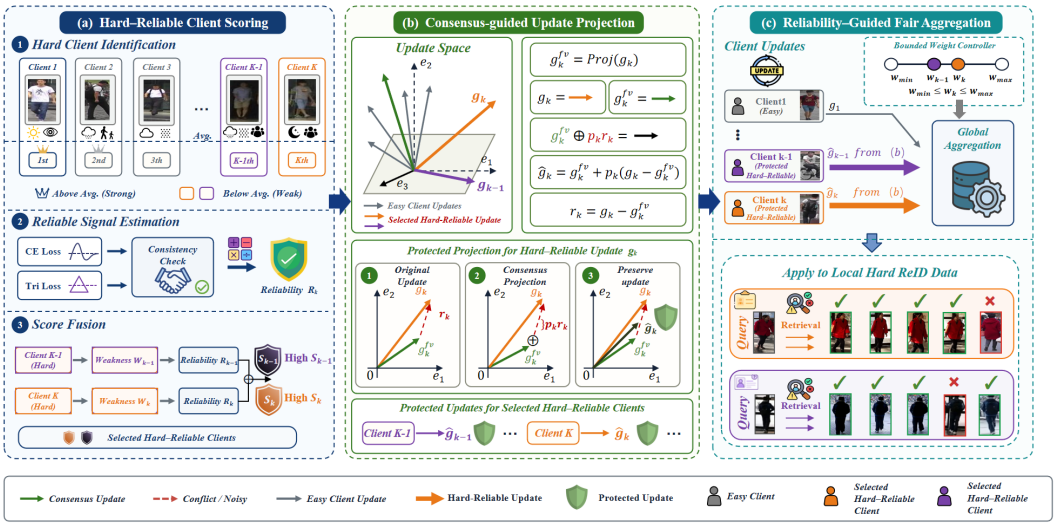


Figure 2: **Architectural diagram of FedHRA.** FedHRA first scores hard-reliable clients by considering both client hardness and update reliability. Then, selected hard-client updates are projected to the easy-client consensus update and partially recovered to preserve reliable hard-domain ReID cues. Finally, reliability-guided fair aggregation protects corrected hard-client updates during global aggregation while avoiding excessive amplification of unreliable visual noise.

4 Methodology

4.1 Motivation

Federated ReID suffers from severe domain skew caused by diverse camera conditions, pedestrian appearances, occlusions, resolutions, and backgrounds. Easy clients usually generate stable identity-discriminative updates from clear pedestrian images, while hard clients under challenging visual conditions produce more dispersed updates. These hard-client updates are not uniformly harmful: they may include useful hard-domain identity cues, such as partial body regions, low-resolution details, and occlusion-related appearance patterns, but they may also contain noisy components caused by background bias, detection errors, and unstable hard samples. Therefore, simply increasing hard-client contributions may amplify noise, whereas directly suppressing hard-client updates may remove useful hard-domain-specific ReID information. FedHRA is designed to identify reliable hard clients, align their updates with the easy-client consensus update, recover part of their original updates according to reliability, and finally protect them through reliability-guided aggregation more effectively throughout subsequent rounds of global model optimization.

4.2 Proposed Method

Based on the above motivation, we propose **FedHRA**, a **F**ederated **H**ard-**R**eliable Recovery and Fair **A**ggregation framework for federated ReID. As shown in Figure 2, FedHRA contains three modules: *(a) Hard-Reliable Client Scoring*, *(b) Consensus-Guided Update Projection*, and *(c) Reliability-Guided Fair Aggregation*. The first module identifies hard clients that are both under-optimized and reliable, avoiding blindly increasing all weak clients. The second module corrects selected hard-client updates by aligning them with the easy-client consensus update and recovering part of their original updates. The third module assigns reliability-guided weights to reliable hard clients, preventing corrected updates from being diluted. Together, these modules preserve useful hard-domain cues while limiting the influence of unreliable components on global optimization. In the ablation study, modules (a) and (b) form **HRUP**, while modules (a) and (c) form **RGA**. The complete client–server optimization procedure is summarized in Algorithm 1.

Hard-Reliable Client Scoring. This module identifies under-optimized clients whose updates contain reliable ReID information and therefore deserve fairness compensation. At each round, the server broadcasts the shared encoder, while the client-specific identity classifiers remain local be-

Algorithm 1 Overall Training Procedure of FedHRA

Input: clients $\mathcal{C} = \{1, \ldots, M\}$, splits $\{\mathcal{D}_k^{\text{tr}}, \mathcal{D}_k^{\text{val}}\}_{k=1}^{M}$, base weights $\{q_k\}_{k=1}^{M}$, rounds T, local epochs E, and w^0
Output: final global embedding parameters w^T
for $t = 0, \ldots, T-1$ **do**
  for all $k \in \mathcal{C}$ **in parallel do**
    $P_k^{\text{pre}} \leftarrow \text{mAP}(w^t, \mathcal{D}_k^{\text{val}})$ // Validation before update
    Train on $\mathcal{D}_k^{\text{tr}}$ for E epochs; obtain g_k and normalized loss progress $(\delta_{ce}^k, \delta_{tri}^k)$ // Local ReID update
    $P_k^{\text{post}} \leftarrow \text{mAP}(w^t - g_k, \mathcal{D}_k^{\text{val}})$ // Held-out validation
  end for
  Compute W_k, R_k, and S_k; set $\mathcal{H}_s = \{k \mid S_k > 0\}$ and $\mathcal{E} = \{k \mid W_k = 0\}$ // Hard-reliable scoring
  $g^c \leftarrow \sum_{e \in \mathcal{E}} \tilde{q}_e g_e$ // Easy-client consensus
  for $k \in \mathcal{H}_s$ **do**
    Project g_k onto g^c and recover p_k of its residual to obtain $\hat{g}_k$ // Correct reliable hard update
  end for
  Set $\tilde{g}_k \leftarrow \hat{g}_k$ for $k \in \mathcal{H}_s$, otherwise $\tilde{g}_k \leftarrow g_k$ // Choose corrected updates
  Compute bounded weights w_k from (q_k, S_k) and update $w^{t+1} \leftarrow w^t - \sum_k w_k \tilde{g}_k$ // Fair aggregation
end for
return w^T

cause different ReID domains have disjoint identity spaces. After local training, client k uploads only its shared-encoder update g_k and the scalar statistics required for scoring.

A hard client may represent a domain with severe occlusion, blur, poor illumination, or complex backgrounds, but low performance alone does not imply that its update should receive a larger contribution. Some hard-client updates may contain useful domain-specific identity cues, whereas others may be dominated by unstable visual noise. Directly compensating every low-performing client may therefore amplify unreliable update directions.

We measure client hardness using the incoming global model on a held-out local validation split. Specifically, let $\bar{P}^{\text{pre}} = M^{-1} \sum_{j=1}^{M} P_j^{\text{pre}}$ denote the average validation mAP before local optimization. The hardness of client k is defined as $W_k = \max(0, \bar{P}^{\text{pre}} - P_k^{\text{pre}})$, so only clients performing below the current cross-client average are treated as hard. Recomputing W_k at each round allows the hard-client set to change as the global model improves. The validation split is drawn exclusively from the official training data, and the official test split is never accessed during training.

Hardness alone cannot determine whether a local update is reliable. Classification and triplet losses have different numerical scales, so their raw difference is not directly meaningful. We instead compute the normalized relative progress $\delta_u^k = (L_{u,k}^{\text{pre}} - L_{u,k}^{\text{post}})/(|L_{u,k}^{\text{pre}}| + \epsilon)$ for $u \in \{ce, tri\}$. Their progress agreement is summarized by $R_k^{\text{loss}} = [1 + |\delta_{ce}^k - \delta_{tri}^k|]^{-1}$. This normalization compares relative optimization changes rather than the absolute magnitudes of the two losses.

Because similar loss changes do not necessarily imply better retrieval, we add independent held-out evidence through $R_k^{\text{val}} = \text{clip}((P_k^{\text{post}} - P_k^{\text{pre}})/(1 - P_k^{\text{pre}} + \epsilon), 0, 1)$. Negative validation progress is clipped to zero, preventing an overfitted local update from receiving compensation solely because its training losses decrease. The final hard-reliable score is

$$S_k = W_k R_k, \qquad R_k = R_k^{\text{loss}} R_k^{\text{val}}. \tag{2}$$

Thus, a large S_k requires client k to be under-optimized, exhibit consistent progress across the two ReID objectives, and achieve positive held-out retrieval improvement. This score is an operational reliability proxy rather than a claim that loss agreement alone proves semantic correctness.

We select $\mathcal{H}_s = \{k \mid S_k > 0\}$ as the hard-reliable client set and define $\mathcal{E} = \{k \mid W_k = 0\}$ as the easy-client set. Hard clients with $S_k = 0$ retain original updates without compensation. Clients transmit two validation-mAP scalars and two normalized loss-progress scalars; validation samples, identity predictions, and client-specific classifier parameters remain local.

Consensus-Guided Update Projection and Recovery. This module corrects selected hard-client updates while preserving reliable hard-domain identity information. Clients in $\mathcal{E}$ provide relatively stable optimization directions under the current global model. Their updates are combined into the consensus direction $g^c = \sum_{e \in \mathcal{E}} \tilde{q}_e g_e$, where $\tilde{q}_e = q_e / \sum_{j \in \mathcal{E}} q_j$ is the normalized base weight

within the easy-client set. The consensus is used only as a correction reference and does not assign permanent priority to these clients.

For each selected hard-reliable client, FedHRA retains the component aligned with this consensus:

$$g_k^{fv} = \frac{\max(0, \langle g_k, g^c \rangle)}{\|g^c\|_2^2 + \epsilon} g^c, \qquad k \in \mathcal{H}_s. \tag{3}$$

A positive inner product preserves the component supporting the consensus, whereas a non-positive inner product produces a zero projection. This prevents a conflicting hard-client update from reversing the stable direction. However, using only the projected component may remove useful cues associated with occlusion, low resolution, camera style, or poor illumination.

To retain reliable client-specific information, we normalize S_k within the selected hard-client set and define $\tilde{S}_k = (S_k - S_{\min})/(S_{\max} - S_{\min} + \epsilon)$. The recovery coefficient is $p_k = \mathrm{clip}(\lambda_p \tilde{S}_k, 0, p_{\max})$, where λ_p controls the recovery strength and $p_{\max}$ limits the maximum recovered proportion. The corrected hard-client update is

$$\hat{g}_k = g_k^{fv} + p_k \big(g_k - g_k^{fv}\big). \tag{4}$$

The first term preserves the stable consensus-aligned direction, while the second restores a bounded part of the client-specific residual. More reliable hard clients retain more of their original ReID information, whereas less reliable clients remain closer to the conservative projection. FedHRA therefore avoids both unrestricted noise amplification and complete suppression of hard-domain cues during successive rounds of global federated representation learning.

Reliability-Guided Fair Aggregation. After correction, clients in $\mathcal{H}_s$ use $\hat{g}_k$, whereas all other clients retain their original updates g_k; we denote the resulting update by $\tilde{g}_k$. Reliable hard clients receive the compensation factor $\alpha_k = 1 + \lambda_w S_k$, while $\alpha_k = 1$ for all remaining clients. Here, λ_w controls the compensation strength. This additional weighting prevents corrected hard-client updates from being immediately diluted during global aggregation.

To prevent any client from dominating global optimization, the compensation factor is clipped to $[w_{\min}, w_{\max}]$ and normalized together with the declared base weight q_k. The final aggregation is

$$\begin{aligned} w_k &= \frac{q_k \ \mathrm{clip}(\alpha_k, w_{\min}, w_{\max})}{\sum_{j=1}^{M} q_j \ \mathrm{clip}(\alpha_j, w_{\min}, w_{\max})}, \\ g^t &= \sum_{k=1}^{M} w_k \tilde{g}_k, \qquad w^{t+1} = w^t - g^t. \end{aligned} \tag{5}$$

We use $q_k = 1/M$ under the equal-domain protocol. Retaining q_k explicitly distinguishes the base aggregation rule from FedHRA's bounded reliability-aware compensation. If no client obtains a positive S_k, no update is corrected or upweighted, and the method reduces to the underlying base aggregation rule. Overall, FedHRA identifies reliable hard clients, corrects their updates through consensus-guided projection and controlled recovery, and protects their contributions during global aggregation without directly altering the client-specific identity classification heads.

5 Experiments

5.1 Experimental Setup

Datasets and client scope. We evaluate FedHRA on Market-1501 (Zheng et al., 2015), MSMT17 (Wei et al., 2018), CUHK03 (Li et al., 2014), and CUHK02 (Li & Wang, 2013). Each dataset represents an independent camera-network domain and cross-silo client. Local validation data are drawn only from the official training split, which remains disjoint from the test split. This four-client setting evaluates domain-level rather than large-scale cross-device fairness.

Model and federated protocol. We use ResNet-50 and ViT. With disjoint identity spaces, classifiers and batch-normalization statistics remain local; only shared parameters are aggregated. All clients participate with equal weight $q_k = 1/M$. FedHRA uploads P_k^{pre}, P_k^{post}, δ_{ce}^k, and δ_{tri}^k. Test data are used only for final reporting, and all methods follow the same protocol.

Table 1: **Source-domain federated ReID comparison.** M, MS, C3, and C2 denote Market-1501, MSMT17, CUHK03, and CUHK02; R1 denotes Rank-1. AVG and STD denote the unweighted mean and population standard deviation across the four client domains.

Category	Methods	Venue	M		MS		C3		C2		mAP		rank-1	
			mAP	R1	mAP	R1	mAP	R1	mAP	R1	AVG ↑	STD ↓	AVG ↑	STD ↓
Backbone: ResNet50														
Federated Learning	SCAFFOLD	ICML'20	32.4	57.6	24.1	46.4	17.6	17.8	56.2	55.7	32.6	**14.6**	44.4	15.9
	MOON	CVPR'21	34.0	58.4	24.3	48.0	18.1	17.4	58.9	56.5	33.8	15.5	45.1	16.5
	FedProx	MLSys'20	30.4	56.8	24.7	48.8	18.0	17.5	58.5	57.2	32.9	15.4	45.1	16.3
Fed ReID	FedPav	MM'20	30.7	58.4	17.3	43.3	34.6	27.1	71.6	73.8	38.6	20.1	50.7	17.4
	FedReID	AAAI'21	61.9	82.2	30.9	60.1	37.7	38.1	76.6	76.4	51.8	18.4	64.2	17.1
	MEDA	ICASSP'24	65.4	82.9	23.2	46.7	49.0	50.6	81.9	79.9	54.9	21.7	65.0	16.5
	PEFL	CVPR'26	66.9	85.0	25.4	50.4	49.2	51.1	81.0	81.7	55.6	20.8	67.1	16.3
	SSCU	MM'25	57.9	79.1	23.8	51.8	38.6	42.2	75.7	77.5	49.0	19.6	62.7	16.0
	FedSupWA	AAAI'26	59.0	79.9	24.6	52.8	39.8	43.6	76.8	79.0	50.1	19.7	63.8	16.0
	FedARKS	AAAI'26	59.8	80.4	25.2	53.5	40.9	44.7	77.9	79.9	51.0	19.8	64.6	15.8
Fair Federated Learning	q-FFL	ICLR'20	56.4	67.3	22.5	47.0	25.4	28.4	62.8	61.2	41.8	18.0	51.0	15.0
	FedFV	IJCAI'21	66.0	84.7	28.3	52.4	49.8	51.0	81.2	82.0	56.3	19.6	67.5	15.9
	FedHEAL	CVPR'24	67.8	84.8	22.0	45.0	37.5	39.2	74.7	74.8	50.5	21.6	61.0	19.3
	FedPW	NeurIPS'25	64.3	76.3	28.6	54.2	32.4	32.9	72.9	72.6	49.6	19.3	59.0	17.2
Fair Federated ReID	**FedHRA**	**Ours**	**70.2**	**87.3**	**34.7**	**59.1**	**51.4**	**53.4**	**84.2**	**84.5**	**60.1**	18.7	**71.1**	**15.0**
Backbone: ViT														
Fed ReID	FedPav	MM'20	47.8	70.0	27.1	53.9	37.1	36.2	68.7	66.1	45.2	**15.4**	56.6	13.2
	FedReID	AAAI'21	64.6	79.2	38.4	64.2	43.2	46.1	76.8	73.2	55.8	15.7	65.7	**12.5**
	MEDA	ICASSP'24	70.2	87.8	27.1	53.0	51.3	54.1	84.6	85.6	58.3	21.5	70.1	16.6
	PEFL	CVPR'26	68.4	86.6	26.6	52.7	50.5	52.5	85.3	86.2	57.7	21.8	69.5	16.9
	SSCU	MM'25	65.9	84.1	30.8	57.8	47.6	49.2	83.2	83.0	56.9	19.6	68.5	15.3
	FedSupWA	AAAI'26	67.0	84.9	31.6	58.8	48.8	50.6	84.3	84.5	57.9	19.7	69.7	15.3
	FedARKS	AAAI'26	67.8	85.4	32.2	59.5	49.9	51.7	85.4	85.4	58.8	19.8	70.5	15.2
Fair Federated Learning	q-FFL	ICLR'20	59.7	77.4	31.2	55.0	38.0	43.6	71.4	72.7	50.1	16.2	62.2	13.6
	FedFV	IJCAI'21	70.0	84.4	34.2	58.8	48.9	48.7	83.4	82.3	59.1	18.9	68.6	15.2
	FedHEAL	CVPR'24	65.5	83.3	37.3	62.1	39.5	39.1	76.3	76.4	54.7	16.7	65.2	16.9
	FedPW	NeurIPS'25	66.5	84.7	35.6	60.4	38.6	38.5	75.8	73.7	54.1	17.4	64.3	17.2
Fair Federated ReID	**FedHRA**	**Ours**	**75.8**	**89.7**	**44.4**	**68.4**	**57.4**	**58.1**	**87.6**	**86.4**	**66.3**	16.6	**75.7**	13.0

Evaluation metrics. We report domain-level mAP and Rank-1, their unweighted mean (AVG), population standard deviation (STD), worst-domain score, and maximum domain gap. AVG measures retrieval utility, STD and Gap describe dispersion, and Worst records the least-served domain. Since a small STD can accompany uniformly poor performance, these statistics are interpreted jointly. Table 1 recomputes AVG and STD from the displayed one-decimal scores. Training curves are averaged over the final five communication rounds.

5.2 Comparison to State-of-the-Arts

Comparison with fair federated learning methods. Table 1 compares FedHRA with general federated optimizers, federated ReID methods, performance reweighting, and conflict-correction baselines. Under ResNet50, FedHRA achieves 60.1 mAP AVG and 71.1 Rank-1 AVG, improving over FedFV by 3.8 and 3.6 points, respectively. The largest gains occur on the difficult MSMT17 client, where mAP increases from 28.3 to 34.7 and Rank-1 increases from 52.4 to 59.1. On CUHK03, mAP and Rank-1 also improve from 49.8 and 51.0 to 51.4 and 53.4. Meanwhile, the mAP STD decreases from 19.6 to 18.7 and the Rank-1 STD decreases from 15.9 to 15.0. These results show that the average improvements are accompanied by better hard-domain retrieval and lower cross-domain dispersion under the same federated ReID training protocol.

Under ViT, FedHRA reaches 66.3 mAP AVG and 75.7 Rank-1 AVG, exceeding FedFV by 7.2 and 7.1 points. The largest improvements again appear on the challenging domains: mAP increases by 10.2 points on MSMT17 and 8.5 points on CUHK03, while Rank-1 increases by 9.6 and 9.4 points, respectively. FedHRA also improves both metrics on Market-1501 and CUHK02, indicating that hard-client compensation does not sacrifice performance on easier domains. The mAP STD decreases from 18.9 to 16.6, while the Rank-1 STD decreases from 15.2 to 13.0.

Table 2: **Complementary diagnostics on mAP and Rank-1.** Worst denotes the minimum domain score, while Gap is the difference between the maximum and minimum domain scores.

Backbone	Method	mAP				Rank-1			
		AVG↑	STD↓	Worst↑	Gap↓	AVG↑	STD↓	Worst↑	Gap↓
ResNet50	FedFV	56.3	19.6	28.3	52.9	67.5	15.9	51.0	**33.7**
	FedHRA	**60.1**	**18.7**	**34.7**	**49.5**	**71.1**	**15.0**	**53.4**	33.9
ViT	FedFV	59.1	18.9	34.2	49.2	68.6	15.2	48.7	35.7
	FedHRA	**66.3**	**16.6**	**44.4**	**43.2**	**75.7**	**13.0**	**58.1**	**31.6**

FedHRA does not obtain the smallest STD against every baseline because it aims to improve reliable hard clients without suppressing well-performing clients, rather than directly minimizing raw dispersion. STD measures deviations from each method's own mean but ignores the absolute performance level and whether similarity is achieved by improving hard clients or lowering strong clients. Thus, uniformly weak domain scores can still produce a small STD. For example, FedPav obtains a smaller ViT mAP STD of 15.4, but its AVG and worst-domain mAP are only 45.2 and 27.1, compared with 66.3 and 44.4 for FedHRA. Similarly, FedReID achieves a slightly smaller ViT Rank-1 STD of 12.5, whereas FedHRA increases its AVG from 65.7 to 75.7 and worst-domain score from 46.1 to 58.1. Because the gains differ across domains, improving every domain does not necessarily minimize STD. We therefore interpret STD jointly with AVG, Worst, Gap, and domain-level results. Together, these results show that FedHRA improves retrieval performance and cross-domain balance without obtaining apparent fairness by weakening strong clients.

Accuracy–fairness interpretation. Table 2 complements AVG and STD with the worst-domain score and maximum domain gap. Under ResNet50, FedHRA raises the worst-domain mAP from 28.3 to 34.7, increases the average mAP from 56.3 to 60.1, and reduces the mAP STD and Gap by 0.9 and 3.4 points. For Rank-1, AVG, STD, and Worst improve from 67.5, 15.9, and 51.0 to 71.1, 15.0, and 53.4, while the Gap changes slightly from 33.7 to 33.9. Under ViT, the worst-domain mAP and Rank-1 increase by 10.2 and 9.4 points, while their Gaps decrease by 6.0 and 4.1 points. We therefore interpret the domain scores jointly with AVG, STD, Worst, and Gap rather than treating one dispersion statistic as a complete fairness measure.

5.3 Ablation Study

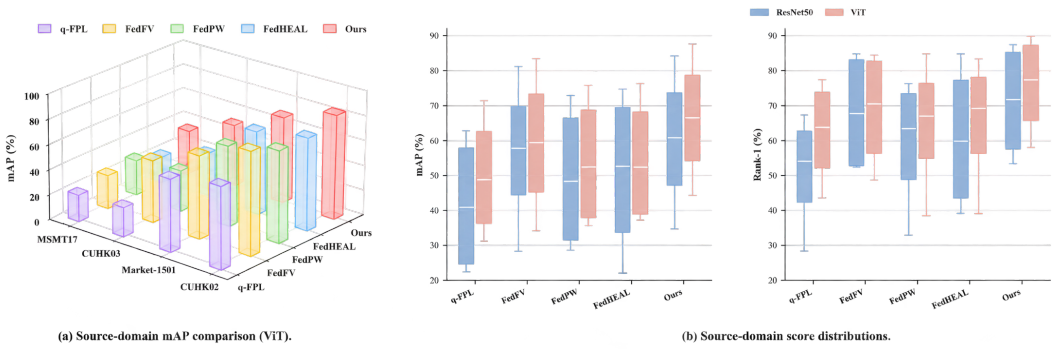
Table 3: **Ablation of hard-reliable update projection (HRUP) and reliability-guided aggregation (RGA).** AVG and STD summarize the four source-domain scores. Red arrows report changes relative to FedFV for the same backbone.

Backbone	Variant	HRUP	RGA	M		MS		C3		C2		mAP		Rank-1	
				mAP	R1	mAP	R1	mAP	R1	mAP	R1	AVG↑	STD↓	AVG↑	STD↓
ResNet50	FedFV	×	×	66.0	84.7	28.3	52.4	49.8	51.0	81.2	82.0	56.3	19.6	67.5	15.9
	+HRUP	✓	×	67.5 ↑+1.5	85.0 ↑+0.3	30.6 ↑+2.3	56.4 ↑+4.0	50.4 ↑+0.6	51.4 ↑+0.4	82.8 ↑+1.6	82.5 ↑+0.5	57.8 ↑+1.5	19.5 ↓0.1	68.8 ↑+1.3	15.1 ↓0.8
	+RGA	×	✓	68.4 ↑+2.4	86.5 ↑+1.8	32.7 ↑+4.4	58.2 ↑+5.8	51.5 ↑+1.7	52.8 ↑+1.8	83.6 ↑+2.4	83.8 ↑+1.8	59.0 ↑+2.7	19.0 ↓0.6	70.3 ↑+2.8	15.0 ↓0.9
	FedHRA	✓	✓	**70.2** ↑+4.2	**87.3** ↑+2.6	**34.7** ↑+6.4	**59.1** ↑+6.7	**51.4** ↑+1.6	**53.4** ↑+2.4	**84.2** ↑+3.0	**84.5** ↑+2.5	**60.1** ↑+3.8	**18.7** ↓0.9	**71.1** ↑+3.6	**15.0** ↓0.9
ViT	FedFV	×	×	70.0	84.4	34.2	58.8	48.9	48.7	83.4	82.3	59.1	18.9	68.6	15.2
	+HRUP	✓	×	72.6 ↑+2.6	87.2 ↑+2.8	39.8 ↑+5.6	63.5 ↑+4.7	53.4 ↑+4.5	55.6 ↑+6.9	85.1 ↑+1.7	85.7 ↑+3.4	62.7 ↑+3.6	17.4 ↓1.5	73.0 ↑+4.4	13.7 ↓1.5
	+RGA	×	✓	73.5 ↑+3.5	88.0 ↑+3.6	41.2 ↑+7.0	65.4 ↑+6.6	53.6 ↑+4.7	54.7 ↑+6.0	86.8 ↑+3.4	85.9 ↑+3.6	63.8 ↑+4.7	17.6 ↓1.3	73.5 ↑+4.9	14.0 ↓1.2
	FedHRA	✓	✓	**75.8** ↑+5.8	**89.7** ↑+5.3	**44.4** ↑+10.2	**68.4** ↑+9.6	**57.4** ↑+8.5	**58.1** ↑+9.4	**87.6** ↑+4.2	**86.4** ↑+4.1	**66.3** ↑+7.2	**16.6** ↓2.3	**75.7** ↑+7.1	**13.0** ↓2.2

FedFV is adopted as the baseline because it is the closest update-conflict correction method to FedHRA. HRUP and RGA are integrated without changing the optimizer, data splits, communication schedule, or backbone, allowing Table 3 to isolate their contributions. Under ResNet50, HRUP improves the mAP and Rank-1 AVGs by 1.5 and 1.3 points and reduces their STDs by 0.1 and 0.8 points. RGA produces larger AVG gains of 2.7 and 2.8 points, with STD reductions of 0.6 and 0.9 points. Combining both components achieves 60.1 mAP AVG and 71.1 Rank-1 AVG, improvements of 3.8 and 3.6 points over FedFV, while reducing both STDs by 0.9 points.

Under ViT, HRUP improves the mAP and Rank-1 AVGs by 3.6 and 4.4 points, whereas RGA obtains larger gains of 4.7 and 4.9 points. HRUP produces slightly lower STDs than RGA, but their combination achieves the strongest overall results, increasing the AVGs from 59.1 and 68.6 to 66.3 and 75.7 while reducing the STDs from 18.9 and 15.2 to 16.6 and 13.0. Thus, HRUP and RGA provide complementary update correction and contribution protection, yielding the highest AVG and the lowest or tied-lowest STD for both backbones.

5.4 VISUALIZATION



(a) Source-domain mAP comparison (ViT). (b) Source-domain score distributions.

Figure 3: Source-domain performance analysis. Panel (a) compares ViT mAP across fair federated methods, while panel (b) summarizes the four-domain mAP and Rank-1 distributions under ResNet50 and ViT. Distribution width is interpreted together with the numerical STD results.

As shown in Figure 3(a), the 3D bar chart compares ViT source-domain mAP among fair federated learning methods. FedHRA outperforms q-FFL, FedFV, FedPW, and FedHEAL across all source domains, especially on hard domains such as MSMT17 and CUHK03. Compared with FedFV, FedHRA improves mAP by 10.2 points on MSMT17 and 8.5 points on CUHK03, while also achieving gains of 5.8 and 4.2 points on Market-1501 and CUHK02, respectively. This indicates that FedHRA better preserves reliable hard-client updates without sacrificing retrieval performance on easier domains, producing improvements across different levels of domain difficulty.

Figure 3(b) visualizes the source-domain score distributions. Each box summarizes the four domain scores. FedHRA shifts both mAP and Rank-1 distributions toward higher performance ranges under ResNet50 and ViT. Its elevated lower whiskers indicate stronger performance on the challenging domains, while higher medians show that these hard-domain gains are accompanied by improved overall retrieval quality. The consistent upward movement across both metrics and backbones confirms that the improvements are distributed across heterogeneous clients rather than concentrated on a single easy domain. These visual trends agree with the domain-level, AVG, STD, Worst, and Gap results reported in Tables 1 and 2 under the same evaluation protocol.

6 CONCLUSION

In this paper, we address Performance Fairness in federated ReID under domain skew by tackling hard-client update dilution and unreliable fair compensation. We observe that hard-client updates are often dispersed and may contain useful client-specific components and noisy components, making direct compensation unreliable. To address this, we propose FedHRA, which identifies reliable hard clients, aligns their updates with the easy-client consensus update, and recovers part of their original updates according to reliability. Moreover, reliability-guided fair aggregation prevents corrected hard-client updates from being diluted during global aggregation. Experiments on ReID benchmarks demonstrate that FedHRA improves performance and client-wise fairness. These results verify the effectiveness of reliability-aware hard-client update protection. FedHRA brings benefits to hard domains under severe visual variations. This work provides a reliable update-protection perspective for Fair Federated ReID under practical cross-domain federated training conditions.

AI USE STATEMENT

Generative AI tools were used during the preparation of this work to assist with language editing, LaTeX formatting, citation style consistency, and the clarity and presentation of the metric formulation. In particular, these tools were used to improve grammar, readability, and organization, as well as to suggest clearer ways of expressing selected technical descriptions and mathematical notation. AI-generated outputs were treated only as editorial and drafting suggestions and were not accepted without review.

The authors carefully examined, revised, and verified all AI-assisted material. Responsibility for the study design, analytical reasoning, mathematical derivations, interpretation of results, factual claims, citation accuracy, and final presentation remains entirely with the authors. The authors also take full responsibility for any errors or omissions in the final manuscript and associated artifacts.

REFERENCES

Huiqiang Chen, Tianqing Zhu, Tao Zhang, Wanlei Zhou, and Philip S. Yu. Privacy and fairness in federated learning: On the perspective of tradeoff. *ACM Comput. Surv.*, 56(2), September 2023. ISSN 0360-0300. doi: 10.1145/3606017. URL https://doi.org/10.1145/3606017.

Yuhang Chen, Wenke Huang, and Mang Ye. Fair federated learning under domain skew with local consistency and domain diversity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Sen Cui, Weishen Pan, Jian Liang, Changshui Zhang, and Fei Wang. Addressing algorithmic disparity and performance inconsistency in federated learning. *Advances in Neural Information Processing Systems*, 34:26091–26102, 2021.

Yahya H Ezzeldin, Shen Yan, Chaoyang He, Emilio Ferrara, and A Salman Avestimehr. Fairfed: Enabling group fairness in federated learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 7494–7502, 2023.

Kaiqi Guan, Wenke Huang, Xianda Guo, Yueyang Yuan, Bin Yang, and Mang Ye. Rethinking fair federated learning from parameter and client view. In *Advances in Neural Information Processing Systems*, 2025.

Samuel Horváth, Stefanos Laskaridis, Mario Almeida, Ilias Leontiadis, Stylianos Venieris, and Nicholas Lane. Fjord: Fair and accurate federated learning under heterogeneous targets with ordered dropout. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 12876–12889. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/6aed000af86a084f9cb0264161e29dd3-Paper.pdf.

Meirui Jiang, Holger R Roth, Wenqi Li, Dong Yang, Can Zhao, Vishwesh Nath, Daguang Xu, Qi Dou, and Ziyue Xu. Fair federated medical image segmentation via client contribution estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16302–16311, 2023.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank J Reddi, Sebastian U Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*, pp. 5132–5143. PMLR, 2020.

Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10713–10722, 2021a.

Tian Li, Maziar Sanjabi, Ahmad Beirami, and Virginia Smith. Fair resource allocation in federated learning. *arXiv preprint arXiv:1905.10497*, 2019.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems*, volume 2, pp. 429–450, 2020.

Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In *International conference on machine learning*, pp. 6357–6368. PMLR, 2021b.

Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3594–3601, 2013.

Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 152–159, 2014.

Lingjuan Lyu, Xinyi Xu, Qian Wang, and Han Yu. Collaborative fairness in federated learning. In *Federated Learning: Privacy and Incentive*, pp. 189–204. Springer, 2020.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Aarti Singh and Jerry Zhu (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

Jiaxu Miao, Zongxin Yang, Leilei Fan, and Yi Yang. Fedseg: Class-heterogeneous federated learning for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8042–8052, 2023.

Mehryar Mohri, Gary Sivek, and Ananda Theertha Suresh. Agnostic federated learning. In *International conference on machine learning*, pp. 4615–4625. PMLR, 2019.

JooHyung Oh, Minyoung Oh, Sung Whan Yoon, and Jae-Young Sim. Pose-guided enriched feature learning for federated-by-camera person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Tao Qi, Fangzhao Wu, Chuhan Wu, Lingjuan Lyu, Tong Xu, Hao Liao, Zhongliang Yang, Yongfeng Huang, and Xing Xie. Fairvfl: A fair vertical federated learning framework with contrastive adversarial learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 7852–7865. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/333a7697dbb67f09249337f81c27d749-Paper-Conference.pdf.

Bhaskar Ray Chaudhury, Linyi Li, Mintong Kang, Bo Li, and Ruta Mehta. Fairness in federated learning via core-stability. *Advances in neural information processing systems*, 35:5738–5750, 2022.

Chunli Song, Xiaohua Chen, Wenqiu Zhu, Yucan Zhou, Xiaoyan Gu, and Bo Li. Meta-knowledge enhanced data augmentation for federated person re-identification. In *ICASSP 2024–2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8901–8905. IEEE, 2024. doi: 10.1109/ICASSP48485.2024.10447543.

Zheng Wang, Xiaoliang Fan, Jianzhong Qi, Chenglu Wen, Cheng Wang, and Rongshan Yu. Federated learning with fair averaging. *arXiv preprint arXiv:2104.14937*, 2021.

Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 79–88, 2018.

Guile Wu and Shaogang Gong. Decentralised learning from independent multi-domain labels for person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 2898–2906, 2021.

Xin Xu, Chaoyue Ren, Wei Liu, Wenke Huang, Bin Yang, Zhixi Yu, and Kui Jiang. Positive style accumulation: A style screening and continuous utilization framework for federated dg-reid. *arXiv preprint arXiv:2507.16238*, 2025.

Xin Xu, Binchang Ma, Zhixi Yu, and Wei Liu. Fedarks: Federated aggregation via robust and discriminative knowledge selection and integration for person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11424–11432, 2026. doi: 10.1609/aaai.v40i14.38124.

Xinyi Xu, Lingjuan Lyu, Xingjun Ma, Chenglin Miao, Chuan Sheng Foo, and Bryan Kian Hsiang Low. Gradient driven rewards to guarantee fairness in collaborative machine learning. *Advances in Neural Information Processing Systems*, 34:16104–16117, 2021.

Zhixi Yu, Wei Liu, Wenke Huang, Bin Yang, Qian Bie, Guancheng Wan, and Xin Xu. Domain-aware suppression and aggregation for federated dg reid. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12231–12239, 2026. doi: 10.1609/aaai.v40i14.38214.

Yuchen Zeng, Hongxu Chen, and Kangwook Lee. Improving fairness via federated learning. *arXiv preprint arXiv:2110.15545*, 2021.

Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pp. 1116–1124, 2015.

Zirui Zhou, Lingyang Chu, Changxin Liu, Lanjun Wang, Jian Pei, and Yong Zhang. Towards fair federated learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 4100–4101, 2021.

Weiming Zhuang, Yonggang Wen, Xuesen Zhang, Xin Gan, Daiying Yin, Dongzhan Zhou, Shuai Zhang, and Shuai Yi. Performance optimization of federated person re-identification via benchmark analysis. In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 955–963, 2020.