FEDCROSSID: RECOVERING FRAGMENTED CROSS-CAMERA SUPERVISION IN FEDERATED PERSON RE-IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Person re-identification (ReID) relies on recognizing the same person across cameras. In federated-by-camera training, however, each camera uses an independent identity namespace, leaving cross-camera correspondences unobserved. Existing federated ReID methods improve aggregation, domain generalization, or local feature diversity, but generally leave unknown correspondences between partially overlapping client identity sets unused. Two difficulties arise: client identity sets overlap only partially, and uncertain inferred relations can exert increasing influence relative to the main-task gradient. We introduce FedCrossID, a discover-then-calibrate framework. Protected Partial Identity Association (PPIA) discovers soft correspondences with an unmatched option, while Reliability-Calibrated Relation Learning (RCRL) weights detached relation targets by confidence and attenuates their influence over communication rounds. Under a protocol that withholds global identity mappings from training, FedCrossID improves a local FedPav-style baseline by 16.81 and 4.17 mAP points on Market1501 and MSMT17, reaching 49.22 ± 0.52 and 17.33 ± 0.14, respectively. Across three paired seeds, PPIA provides most of the improvement. Loss-formulation controls attribute most of the additional MSMT17 gain to the revised relation-loss formulation. Gradient diagnostics motivate phase attenuation to control the changing relative influence of relation supervision. In the tested settings, selected relations outperform degree-preserving randomized relations, including at zero identity overlap; this does not establish that the gains come from recovering true identities. Tuned constants achieve similar retrieval, and partial OT does not consistently outperform forced matching. Summary perturbation provides no formal privacy guarantee.

1 INTRODUCTION

Person re-identification retrieves a person observed by one camera from images captured by other cameras. The task is therefore built around cross-camera evidence: appearance changes caused by viewpoint, illumination, occlusion, and background must be separated from identity. Centralizing camera streams simplifies supervision but is often undesirable because the images are sensitive, the data owners differ, or data movement is restricted. Federated learning offers a practical alternative by exchanging learned information rather than pooling images (McMahan et al., 2017). Federated ReID systems commonly aggregate a shared representation while retaining client-specific classifiers (Zhuang et al., 2020; Wu & Gong, 2021). Subsequent work has addressed domain shift, camera imbalance, and limited local variation (Yang et al., 2024; Song et al., 2024; Tseng et al., 2025; Xu et al., 2026a; Oh et al., 2026).

Federated-by-camera ReID exposes a more specific conflict. When each camera is a client, its labels name identities only within that camera. The person labelled 17 by one client may be the person labelled 8 by another, but neither number carries cross-client meaning. Standard model aggregation can transfer generic visual features, but it cannot tell the local objectives that these two classes should be close. This fragmentation is substantial in standard benchmarks. The 751 Market1501 training identities yield 3,262 camera-local identity instances, while the 1,041 MSMT17 training identities yield 4,821. Thus, an identity appears in 4.34 Market1501 cameras and 4.63 MSMT17

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.

cameras on average. ***Treating every local namespace as unrelated discards a substantial source of task-specific supervision.***

Existing methods enrich local features or improve cross-client knowledge transfer (Song et al., 2024; Yang et al., 2024; Xu et al., 2026b;a; Oh et al., 2026), but do not make unknown identity correspondences explicit training objects. Prototype sharing is related (Tan et al., 2022b; Rami et al., 2024); aligned classes or source-domain prototypes differ from partially overlapping camera-local label spaces with unavailable mappings.

Recovering that structure presents two challenges. First, client identity sets overlap only partially. A forced one-to-one alignment must assign genuinely unmatched identities and can therefore manufacture false positives. The association input is also imperfect: some identities have few images, camera shift alters the feature distribution, and perturbing exchanged summaries further reduces their fidelity. Second, an inferred link should not be treated as an equally strong positive throughout training. The relative scale of relation and local ReID gradients can change during training even when the correspondence remains uncertain. Confidence weights describe individual relations, while a global coefficient controls their overall influence at each training stage. The questions are distinct: how should candidate relations be discovered, and how strongly should they be used?

We answer them with FedCrossID, a discover-then-calibrate framework. Protected Partial Identity Association (PPIA) forms a summary for each local identity, clips and projects it, and adds independent noise. It then solves an uncertainty-adjusted partial optimal-transport problem with an unmatched bin. It returns mutual assignments that pass probability and margin tests. Reliability-Calibrated Relation Learning (RCRL) builds detached cross-client target memories from the accepted relations. The reported configuration uses relation confidence as an outer sample weight and linearly reduces the global relation coefficient from 1.6 at training index 25 to 0.2 at index 125. The term *protected* is deliberately narrow: perturbation obfuscates the input to matching, but our research implementation still constructs detached remote-prototype targets for relation learning. We claim neither differential privacy nor end-to-end prototype confidentiality.

We make three contributions.

1. **A cross-client relation learning framework.** We propose FedCrossID, a cross-client relation learning framework for federated-by-camera person ReID. It turns candidate correspondences between independent local label spaces into training supervision, enabling cross-camera relation learning without training access to global identity mappings.

2. **Two complementary components, PPIA and RCRL.** PPIA infers candidate associations between partially overlapping identity sets using perturbed summaries, uncertainty-adjusted matching, and an unmatched option; RCRL incorporates these associations into local training through detached targets, confidence weighting, and a round-dependent coefficient.

3. **Experimental validation on two benchmarks.** Experiments on Market1501 and MSMT17 demonstrate improved retrieval over the local baseline across three paired seeds. Component ablations identify the association-based relation objective as the main source of improvement, while randomized forced-matching controls with fixed graph size and node degrees demonstrate the benefit of relation selection.

2 Related Work

Federated person ReID. Federated ReID commonly separates a shared feature extractor from private identity classifiers because client label spaces are independent (Zhuang et al., 2020; Wu & Gong, 2021). This body/head decomposition is also studied more broadly in personalized federated representation learning (Collins et al., 2021). Unsupervised variants combine local clustering with federated optimization (Zhuang et al., 2021; Liu et al., 2022; Weng et al., 2023), while personalized approaches retain client-specific knowledge. These methods establish that a transferable representation can be learned without a shared classifier. Independent heads, however, leave cross-client positive relationships unobserved. FedCrossID retains the shared-body/private-head architecture and changes the supervision available to the body.

Intra-camera supervision. Intra-camera supervised ReID likewise assigns independent identity labels within each camera and withholds inter-camera correspondences (Zhu et al., 2019; Wang et al., 2021). Prior methods learned camera-specific classifiers and mined cross-camera associations, establishing that the missing label structure can be recovered from visual evidence. Their training procedure assumes joint access to the camera datasets, whereas our setting keeps raw camera data in separate federated clients and exposes only encoder updates and the explicitly described relation objects. The shared label-fragmentation problem therefore provides the closest non-federated precedent; our contribution concerns its leakage-controlled federated realization and the calibrated use of recovered relations.

Domain, camera, and feature heterogeneity. A second line of work transfers distribution-level knowledge. Domain and feature hallucination, stylization, and meta-knowledge augmentation enlarge the range of appearances seen locally (Yang et al., 2022; 2024; Song et al., 2024), while parameter suppression, aggregation, and body-aware visual prompts target domain-specific interference (Yu et al., 2026; Xu et al., 2026a). Camera-aware clustering and equalization address local pseudo-label or identity imbalance (Liu et al., 2022; Tseng et al., 2025). Most closely, Pose-guided Enriched Feature Learning synthesizes pose-swapped features for federated-by-camera ReID (Oh et al., 2026). Such feature enrichment is complementary to our objective: it increases variation within a local class, whereas FedCrossID determines whether two local classes should be related.

Prototype sharing and partial correspondence. Federated prototype methods communicate class summaries to coordinate heterogeneous clients (Tan et al., 2022b;a); Fed-Protoid distributes source-domain prototypes for decentralized adaptation (Rami et al., 2024). Prototype alignment through optimal transport has also been studied in personalized federated learning (Ek et al., 2024). Our matching construction inherits the dustbin-augmented optimal-transport idea used for partial feature matching (Sarlin et al., 2020). Neither prototypes nor dustbin matching are new in isolation. Our setting instead requires recovering unknown correspondences between partially overlapping camera-local identity sets without exposing their global mapping during training.

Reliability-aware learning. Confidence-based pseudo-label learning and dynamic regularization are established ideas (Zhang et al., 2021; Acar et al., 2021). We study their use for inferred cross-client relations; the tuned-constant controls delimit the schedule's contribution.

3 METHOD

Figure 1 summarizes the information flow from camera-local training through protected association and relation learning. The evaluator-only global mapping is outside the training path.

3.1 PROBLEM SETTING AND EXPOSURE MODEL

Let client $c \in \{1, \dots, C\}$ denote one camera. Its dataset is $\mathcal{D}_c = \{(x_{cn}, y_{cn})\}_{n=1}^{N_c}$, where $y_{cn} \in \mathcal{Y}_c$ is a contiguous camera-local label. Although two labels in $\mathcal{Y}_c$ and $\mathcal{Y}_{c'}$ may refer to the same physical identity, the cross-client mapping is absent from normal training inputs. We retain it in private tools for offline protocol construction, development-stage parameter selection, and evaluation. The association-temperature selection that uses this mapping is described in Section 4.1. Public training manifests contain opaque image names, local labels, and camera identifiers; original filenames and global identities are excluded.

The model contains a shared encoder f_θ and one private linear classifier h_c per client, following the body/head split used by FedPav-style ReID. Training indices start at $t = 0$. At index t, every client initializes f_θ from the current global state, trains for one local epoch, and returns the encoder update; the reported round-r checkpoint follows update $t = r - 1$. The server performs sample-count-weighted averaging. The classifier h_c remains local. Without cross-client relations, client c minimizes the objective in Equation 1:

$$\mathcal{L}_{\text{reid}}^{c} = \mathcal{L}_{\text{ce}}(h_c(f_\theta(x)), y) + \mathcal{L}_{\text{tri}}(f_\theta(x), y), \tag{1}$$

where the second term is batch-hard cosine triplet loss.

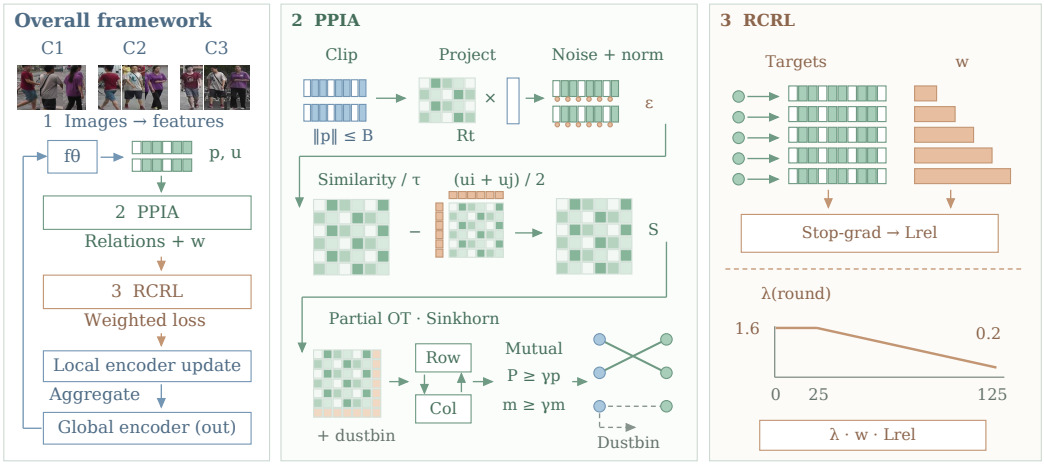

Figure 1: FedCrossID discovers relations with PPIA and calibrates their use with RCRL (expanded at right). Clients retain private classifiers; only encoder updates are aggregated. PPIA perturbs summaries before uncertainty-adjusted partial OT and mutual filtering with an unmatched option. RCRL uses detached targets, confidence weights, and a round-dependent coefficient. Photos are Market1501 training crops; graphs are schematic. Global mappings remain evaluator-only; obfuscation provides no formal privacy guarantee.

3.2 Protected Partial Identity Association

Local identity summaries. For identity $i \in \mathcal{Y}_c$, the client computes a normalized mean embedding p_{ci}. Its uncertainty $u_{ci} \in [0, 1]$ is the mean cosine dispersion of the corresponding image embeddings around p_{ci}. Before pairwise matching, PPIA transforms the summary as

$$\tilde{p}_{ci}^{t} = \mathrm{norm}\big(\mathrm{clip}(p_{ci}, B)R_t + \epsilon_{ci}^{t}\big), \qquad \epsilon_{ci}^{t} \sim \mathcal{N}(0, \sigma^2 I). \tag{2}$$

The projection $R_t \in \mathbb{R}^{D\times d}$ has entries sampled with variance $1/d$ and is shared by clients within round t so that summaries remain comparable. The projection changes at every association refresh. Noise streams are client-specific to avoid row-wise correlations. Our default uses $B = 1$, $d = 256$, and $\sigma = 0.02$.

Equation 2 is an obfuscation mechanism, not a differential-privacy mechanism. We do not define neighboring datasets, an accountant, or an (ϵ, δ) budget. Moreover, the final relation loss uses detached remote prototype targets. The experiments therefore evaluate matching utility under perturbed summaries, not end-to-end privacy.

Uncertainty-adjusted partial matching. For clients c and c', Equation 3 defines the score of local identities i and j:

$$s_{ij} = \frac{(\tilde{p}_{ci}^{t})^{\top}\tilde{p}_{c'j}^{t}}{\tau_a} - \frac{u_{ci} + u_{c'j}}{2}. \tag{3}$$

For a camera pair with N and M local identities, we augment the score matrix with one dustbin row and column, assigning every dustbin entry a score of zero. Let $\bar{S}$ denote the augmented matrix. The transport marginals are

$$a = \frac{(\mathbf{1}_N^{\top}, M)^{\top}}{N + M}, \qquad b = \frac{(\mathbf{1}_M^{\top}, N)^{\top}}{N + M}. \tag{4}$$

We solve

$$T^{*} = \arg\max_{T\geq 0}\{\langle T, \bar{S}\rangle + H(T)\}, \quad T\mathbf{1}_{M+1} = a, \quad T^{\top}\mathbf{1}_{N+1} = b, \tag{5}$$

where $H(T) = -\sum_{ij} T_{ij}\log T_{ij}$. Log-space Sinkhorn iterations approximate this entropy-regularized transport solution (Cuturi, 2013; Sarlin et al., 2020). We rescale the coupling as $P = (N + M)T^{*}$. At convergence, each real-identity row and column of P has unit mass, including its dustbin entry; all acceptance thresholds use this rescaled matrix.

Mutual nearest assignments are computed within the real block $P_{1:N,1:M}$. For candidate (i, j), the margin is

$$m_{ij} = P_{ij} - \max \left\{ \max_{\substack{k \le M \\ k \ne j}} P_{ik},\ \max_{\substack{\ell \le N \\ \ell \ne i}} P_{\ell j},\ P_{i,M+1},\ P_{N+1,j} \right\}. \tag{6}$$

Empty sets of real alternatives contribute zero. Thus, the margin compares the candidate with other real entries in its row or column and with both associated dustbin entries. We accept a relation when

$$(i, j) \in \mathcal{E}_t \iff j = \arg\max_{k \le M} P_{ik},\ i = \arg\max_{\ell \le N} P_{\ell j},\ P_{ij} \ge \gamma_p,\ m_{ij} \ge \gamma_m. \tag{7}$$

The accepted matching score is the relation confidence, $q_{ij} = P_{ij}$; it is used as a weight, not assumed to be a calibrated probability of identity agreement. We use 50 Sinkhorn iterations, $\gamma_p = 0.20$, and $\gamma_m = 0.02$. Relations are refreshed every five communication rounds and are not promoted to permanent global identity labels.

3.3 Reliability-Calibrated Relation Learning

PPIA identifies remote identities that may correspond to a local class and supplies confidence $q_{ij} = P_{ij}$ for each accepted relation. The confidence is detached and clipped to $[0, 1]$ before use; its exponent is one. For client c, the accepted remote prototypes form a detached memory bank $M_c = [m_1, \ldots, m_K]$. Entries are deduplicated by remote client and local identity, and are normalized in the original 512-dimensional embedding space. Matching uses projected summaries, whereas relation learning uses these original-space prototypes; each accepted edge supplies a target to both endpoint clients. For a normalized sample feature z_n with local label y_n, let $\mathcal{P}(y_n)$ be the memory entries linked to that label. Equation 8 defines the multi-positive contrastive target:

$$\ell_{\mathrm{rel}}(z_n, y_n) = -\sum_{k \in \mathcal{P}(y_n)} \bar{q}_{nk} \log \frac{\exp(z_n^\top m_k / \tau_r)}{\sum_{\ell=1}^{K} \exp(z_n^\top m_\ell / \tau_r)}, \tag{8}$$

where q_{nk} is the clipped confidence of the relation linking sample n's local identity to memory entry k, and $\bar{q}_{nk} = q_{nk} / \sum_{j \in \mathcal{P}(y_n)} q_{nj}$ normalizes confidence within the positive set when its sum is nonzero. The denominator includes all K entries in the client memory bank. If a sample has no accepted positive, its relation loss and outer weight are zero; if the bank is empty, the client relation loss is zero. For samples with an accepted positive, the maximum detached confidence $a_n = \max_{k \in \mathcal{P}(y_n)} q_{nk}$ controls the outer contribution:

$$\mathcal{L}_{\mathrm{rel}}^{c} = \frac{1}{|\mathcal{B}|} \sum_{n \in \mathcal{B}} a_n \mathbb{1}[\mathcal{P}(y_n) \ne \emptyset]\, \ell_{\mathrm{rel}}(z_n, y_n). \tag{9}$$

Normalization by the full batch, rather than by the confidence sum, reduces the contribution of uncertain or low-coverage batches. Both q and M_c are detached, so relation gradients update the sample encoder but not the confidence scores or target memory. For a single positive, the normalized target weight is one regardless of confidence; the outer weight preserves the confidence-dependent scale.

Confidence weighting and phase attenuation address two distinct aspects of relation supervision. Confidence determines how strongly to use each inferred link, but does not constrain the total auxiliary gradient relative to the local ReID gradient. A fixed coefficient can therefore produce different relative influences across training stages. We use a larger coefficient early to promote cross-client learning and reduce it later to limit continued reliance on uncertain targets. This motivates the following phase schedule:

$$\lambda_t = \begin{cases} 1.6, & t \le 25, \\ 1.6 + \dfrac{t - 25}{100}(0.2 - 1.6), & 25 < t < 125, \\ 0.2, & t \ge 125, \end{cases} \qquad \mathcal{L}_c^t = \mathcal{L}_{\mathrm{reid}}^{c} + \lambda_t \mathcal{L}_{\mathrm{rel}}^{c}. \tag{10}$$

The reported round-125 checkpoint follows update index $t = 124$ and therefore uses $\lambda_{124} = 0.214$; the schedule reaches 0.2 at index $t = 125$ (the 126th update). We refer to this combination as RCRL.

3.4 FEDERATED PROCEDURE AND INFORMATION FLOW

At an association refresh, clients extract identity summaries from the current global encoder. The simulator applies Equation 2, matches each camera pair, and constructs detached target memories for each client from the accepted edges. Clients optimize Equation 10 for five rounds before relations are refreshed. Only the encoder is averaged; local classifiers and local labels remain client-specific. The global identity mapping is never imported by the normal training path. Only separate offline protocol-construction, OracleLink, and evaluation tools read this mapping.

The current code is a federated simulator rather than a deployed secure protocol. Raw images remain in camera-local data loaders, but prototype construction and detached target routing occur in one process. Our central methodological claim therefore concerns learning from inferred cross-client relations. Secure aggregation, client authentication, access control, and protection of routed prototype targets remain outside the present scope.

4 EXPERIMENTS

The experiments evaluate whether inferred cross-camera relations improve retrieval, factorize association and utilization choices, and test controlled-overlap and degree-preserving randomized alternatives. We interpret these results within the stated limits of the hidden-mapping protocol.

4.1 DATASETS, PROTOCOL, AND IMPLEMENTATION

We use Market1501 (Zheng et al., 2015) and MSMT17 (Wei et al., 2018). Market1501 contains 12,936 training images of 751 identities from six cameras. For MSMT17, we combine the official training and validation partitions, yielding 32,621 images of 1,041 identities from fifteen cameras. Each camera becomes one client. Within a client, identities are re-indexed to contiguous local labels, image filenames are replaced by opaque identifiers, and the global mapping is isolated from training. Evaluation uses the official query and gallery splits with the final global encoder.

All locally controlled methods use an ImageNet-pretrained ResNet-50, a 512-dimensional bottleneck, input size 256×128, cross-entropy plus batch-hard triplet loss, and PK sampling with four images per identity. Each client trains for one local epoch per communication round using SGD with a learning rate of 10^{-3} and a batch size of 32. Association is refreshed every five rounds. The common reporting point is round 125. We run seeds 42, 123, and 3407 and report mean $\pm$ sample standard deviation. Controlled comparisons pair variants by seed. With three pairs, we treat the summaries as descriptive and do not report hypothesis tests. The evaluator measures mAP and CMC Rank-1. The association temperature $\tau_a = 0.01$ was selected by link F1 in a one-round Market1501 development sweep using the evaluator-only identity mapping, then fixed for MSMT17. Thus, global mappings are excluded from normal training inputs but were used for this development-stage hyperparameter choice.

We distinguish two evidence tiers. The published-method results in Table 1 are taken from the unified camera-client comparison of Oh et al. (2026); they provide context but are not locally aligned reproductions. Our local controlled experiments share the same codebase, data split, and training budget, with association and relation-loss choices varied as specified. Because the local FedPav result differs from its author-reported value, especially on MSMT17, cross-tier numbers must not be interpreted as a strict ranking.

4.2 MAIN COMPARISON

Table 1 reports 49.22 and 17.33 mAP for FedCrossID on Market1501 and MSMT17. Relative to the local baseline in Table 2, the gains are 16.81/14.91 mAP/Rank-1 points on Market1501 and 4.17/8.00 on MSMT17. Published rows provide context; differing implementations and hidden-mapping protocols preclude a controlled state-of-the-art claim.

Table 1: Main retrieval results (mAP/R1, %). Published values are reported by Oh et al. (2026); ours are three-seed means at round 125. Comparisons are contextual owing to differing implementations and label-visibility protocols.

Method	Market1501		MSMT17	
	mAP	R1	mAP	R1
BDB (Dai et al., 2019)	32.5	54.7	10.5	24.2
ISE (Zhang et al., 2022)	34.8	56.6	10.0	22.5
SCAFFOLD (Karimireddy et al., 2020)	28.6	51.1	7.5	19.1
FedProx (Li et al., 2020)	31.7	55.3	10.4	25.4
MOON (Li et al., 2021a)	33.2	55.7	13.2	31.2
FedRCL (Seo et al., 2024)	29.6	53.4	7.3	17.9
FedPav (Zhuang et al., 2020)	35.0	61.6	9.1	23.8
DACS (Yang et al., 2024)	40.0	63.1	11.4	29.3
MEDA (Song et al., 2024)	41.4	66.1	9.7	24.8
Pose-guided (Oh et al., 2026)	45.9	66.4	14.5	32.7
FedCrossID (local)	**49.2**	**70.7**	**17.3**	**40.9**

Table 2: Association, weighting, and schedule ablations (round 125; mAP/R1, %; three-seed mean $\pm$ sample SD). Raw RNN and PPIA share the legacy loss; the full row is FedCrossID. The historical fixed-confidence control covers MSMT17; Table 3 adds both-dataset reruns.

Association	Weighting	Schedule	Market1501		MSMT17	
			mAP	R1	mAP	R1
None (baseline)	None	None	32.41 ± 0.45	55.82 ± 0.40	13.16 ± 0.07	32.90 ± 0.26
Raw RNN	Legacy	Fixed	44.56 ± 0.65	66.91 ± 0.76	16.08 ± 0.12	39.37 ± 0.07
PPIA	Legacy	Fixed	47.64 ± 0.55	69.63 ± 0.79	16.51 ± 0.18	40.01 ± 0.21
PPIA	Legacy	Phase	48.56 ± 0.34	70.15 ± 0.37	15.92 ± 0.22	38.93 ± 0.46
PPIA	Confidence	Fixed	–	–	17.29 ± 0.05	40.59 ± 0.38
PPIA (full)	Confidence	Phase	$\mathbf{49.22\pm0.52}$	$\mathbf{70.73\pm0.27}$	$\mathbf{17.33\pm0.14}$	$\mathbf{40.90\pm0.16}$

4.3 Contribution ablation

Table 2 compares association, relation-loss formulation, and phase attenuation under a shared training recipe. PPIA produced most of the improvement, adding 15.23 mAP on Market1501 and 3.35 on MSMT17 over the baseline. Replacing PPIA with reciprocal nearest-neighbor matching on normalized, unperturbed identity summaries (Raw RNN) under the same fixed legacy relation loss reduced mAP by 3.08 points on Market1501 and 0.43 on MSMT17. This shows that the two matchers produce different downstream retrieval outcomes under the same legacy loss. The utilization controls compare relation-loss formulations and phase attenuation. Adding the phase schedule to legacy weighting changed mAP by $+0.92$ on Market1501 but -0.59 on MSMT17. On MSMT17, replacing fixed legacy weighting with fixed outer confidence weighting added 0.78 points; adding phase attenuation then contributed only 0.04 points. The full method exceeded fixed legacy PPIA for every paired seed on both datasets. The Legacy-to-confidence comparison jointly changes positive-target aggregation, outer sample weighting, and the averaging denominator (Appendix A); the gain therefore belongs to the combined loss formulation. The three-seed results are descriptive.

4.4 Does the schedule outperform a tuned constant?

We complete the Market1501 fixed-coefficient control and rerun both datasets with the same association and outer confidence weighting. An identity-disjoint holdout of 20% of official training identities selects among coefficients $\{0.2, 0.5, 0.8, 1.2, 1.6\}$ by mean final validation mAP over seeds 42 and 123. Both datasets select 0.8. The holdout mAP uses the evaluator-only validation partition of those held-out training identities and is a coefficient-selection score; it is not directly comparable with official test mAP. Full-data confirmation uses seeds 42, 123, and 3407, the unchanged official test sets, and the same 125-round endpoint. The historical schedule is frozen, not retuned in this comparison; Appendix B gives selection details.

Table 3: Full-data coefficient controls with outer confidence weighting (round 125; mAP/R1, %; three-seed mean ± sample SD). Fixed 0.8 is holdout-selected; these reruns are separate from Table 2.

Coefficient	Market1501		MSMT17	
	mAP	R1	mAP	R1
Fixed 0.5	47.59 ± 0.56	69.19 ± 0.95	17.29 ± 0.07	40.55 ± 0.42
Selected fixed 0.8	48.47 ± 0.71	70.15 ± 1.21	17.02 ± 0.11	40.41 ± 0.43
Schedule	48.71 ± 0.20	70.11 ± 0.55	17.28 ± 0.16	40.82 ± 0.28

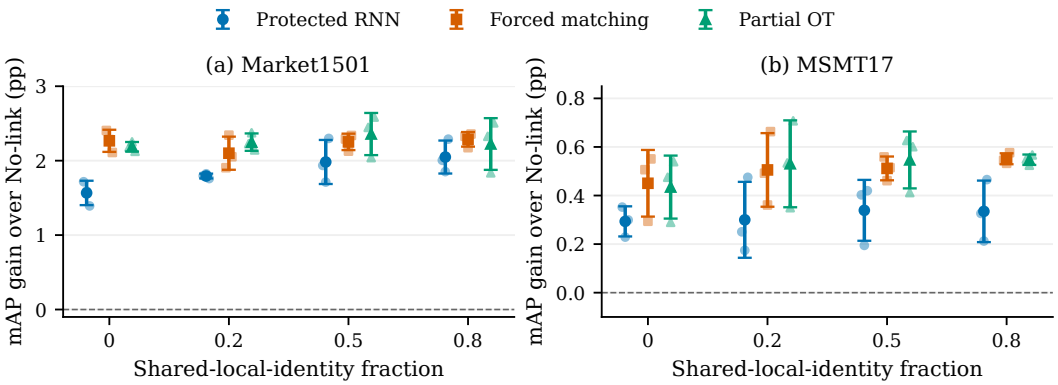

Figure 2: Round-125 mAP gains over No-link across controlled overlap levels; zero-overlap gains do not establish identity recovery, and partial OT does not consistently beat forced matching. Small markers show three construction/training seeds; large markers and bars show mean ± sample SD. RNN is reciprocal nearest neighbors. Matchers share protected inputs and uniform weights. Vertical scales differ; identity composition varies across overlap levels.

In Table 3, the schedule exceeds fixed 0.5 by 1.12 mAP points on Market1501 but changes MSMT17 mAP by -0.01. Against the validation-selected constant, paired differences are only $+0.23 \pm 0.54$ and $+0.26 \pm 0.24$ points, respectively; these three-seed comparisons do not establish a final-accuracy advantage for the schedule.

The gradient diagnostics motivate this control of training-stage influence. At fixed MSMT17 checkpoints, the relation-to-ReID gradient-norm ratio rose from 1.007 ± 0.008 at round 25 to 2.300 ± 0.059 at round 125, while the corresponding mean gradient cosines remained positive (0.446 and 0.351). These observations show a change in relative gradient scale. They motivate attenuation to limit the overall contribution of uncertain relations as training progresses. The fixed-coefficient controls assess the resulting retrieval benefit separately: the schedule serves this optimization objective, while its necessity for final retrieval accuracy is not established by the present comparison.

4.5 Retrieval under controlled identity overlap

We construct four reduced-data protocols with shared-local-identity fractions 0, 0.2, 0.5, and 0.8. Each uses six cameras, 80 local identities per camera, and four images per local identity (1,920 images). Shared identities occur in exactly two cameras; unmatched identities occur in one. Global mappings are used only for offline construction and evaluation, then excluded from public training inputs. Within each subset, RNN, forced Hungarian matching, and partial OT receive the same protected summaries, unit accepted-relation weights, and fixed coefficient 0.5. Thus, this tests matching under uniform confidence, not the full scheduled method. All methods train for 125 rounds and use the unchanged official query/gallery.

Figure 2 visualizes the paired gains; Tables 6 and 7 retain the complete retrieval values. Partial OT exceeds forced matching at low and medium overlap by 0.15 and 0.11 mAP points on Market1501 and 0.03 and 0.04 on MSMT17. It has no consistent advantage at zero or high overlap. All three matchers exceed No-link even at zero overlap, where accepted identity correspondences must be false. This means the gains cannot be interpreted as evidence of correct identity recovery; effects of

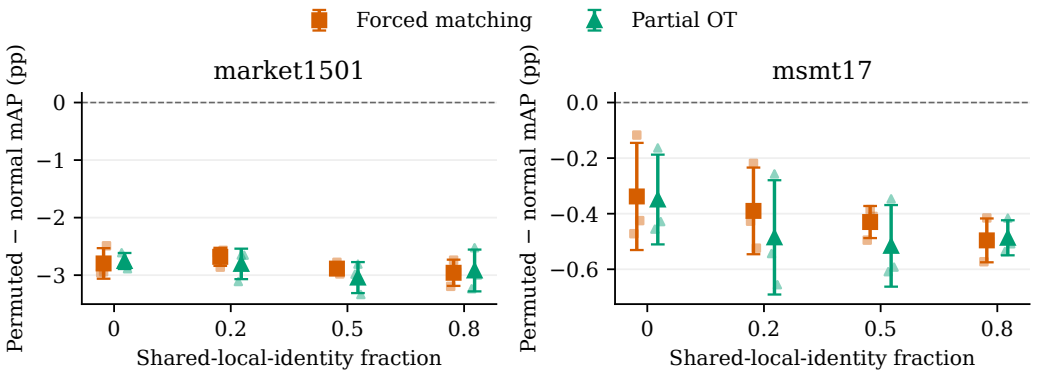

Figure 3: Randomizing endpoints lowers retrieval across all tested settings. Round-125 permuted-minus-normal mAP: small points show three paired seeds; large markers and bars show mean ± sample SD. Weights are uniform. Permutation preserves each refresh graph's degrees; only forced matching fixes degrees across trajectories. Vertical scales differ.

the relation objective unrelated to true cross-camera identity links remain possible. The experiment also does not isolate a causal benefit of unmatched rejection. These small-data protocols also change identity diversity across overlap levels; comparisons are paired within each subset, not a pure causal intervention on overlap (Appendix B).

4.6 Does relation selection matter beyond graph size?

We add 48 runs that uniformly permute right endpoints within each camera-pair matching, retaining its edge count, endpoint degrees, unit weights, and coefficient 0.5. Permutations use no global labels and allow chance unchanged links. Forced matching preserves 1,200 edges and degree five across both conditions; partial OT preserves degrees within each refresh, but its normal and randomized training trajectories can discover different graphs (Appendix D). Figure 3 shows mean mAP reductions of 2.68–3.04 points on Market1501 and 0.34–0.52 on MSMT17 across both matchers and all overlap levels. In the forced-matching control, this comparison supports an effect of relation selection beyond graph size and degree for the tested settings. The effect persists at zero overlap, where correct identity links are absent, so it cannot be attributed solely to identity recovery. Table 8 reports the complete comparison.

5 Discussion and Limitations

Selected relations outperform degree-preserving randomization in the tested settings, although zero-overlap gains limit an identity-recovery explanation. Gradient diagnostics motivate phase attenuation; tuned constants achieve similar retrieval, and partial OT has no consistent advantage over forced matching. The evidence is limited by three seeds, changing subset composition, and loss-formulation controls that vary several components together. Published baselines are not fully aligned reproductions. Dense matching has quadratic cost, and summary perturbation provides no formal privacy guarantee for the exchanged information.

6 Conclusion

FedCrossID combines inference of cross-client relations with confidence-weighted supervision and phase attenuation under a protocol that hides global identity mappings from normal training. It improves the local baseline by 16.81 and 4.17 mAP points on Market1501 and MSMT17. Controlled randomization supports the utility of relation selection in the tested settings. The gradient diagnostics explain the motivation for regulating relation influence during training, and the coefficient controls delimit its final-accuracy contribution. Together, these results support learning from inferred cross-client relations while leaving the contribution of correct identity recovery open.

AI USE STATEMENT

AI tools assisted with manuscript refinement, including language, translation. The authors reviewed and take responsibility for the final content.

ETHICS STATEMENT

Person ReID and cross-camera identity association can be used in high-impact surveillance contexts. This work studies an offline research protocol on established benchmarks and does not introduce new identities or images. Summary obfuscation should not justify deployment as a privacy guarantee. Real-world use would require legal authorization, purpose limitation, retention controls, access auditing, demographic performance assessment, and an explicit threat model. The method should not be deployed for indiscriminate tracking.

REPRODUCIBILITY STATEMENT

Section 4.1 states the shared protocol, while the appendix records dataset statistics and hyperparameters. The global identity mapping is structurally separated from normal training and is used for offline protocol construction, development-stage parameter selection, and evaluation. All primary results use seeds 42, 123, and 3407 and report mean plus sample standard deviation.

REFERENCES

Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N. Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. In *International Conference on Learning Representations*, 2021.

Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 2089–2099, 2021.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, pp. 2292–2300, 2013.

Zuozhuo Dai, Mingqiang Chen, Xiaodong Gu, Siyu Zhu, and Ping Tan. Batch dropblock network for person re-identification and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3691–3701, 2019.

Sannara Ek, Kaile Wang, François Portet, Philippe Lalanda, and Jiannong Cao. Fedali: Personalized federated learning with aligned prototypes through optimal transport. *arXiv preprint arXiv:2411.10595*, 2024.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 5132–5143, 2020.

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

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 *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 6357–6368, 2021b.

Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. In *International Conference on Learning Representations*, 2021c.

Jiabei Liu, Weiming Zhuang, Yonggang Wen, Jun Huang, and Wei Lin. Optimizing federated unsupervised person re-identification via camera-aware clustering. In *IEEE International Workshop on Multimedia Signal Processing*, pp. 1–6, 2022.

Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282, 2017.

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*, pp. 40458–40467, 2026.

Hamza Rami, Jhony H. Giraldo, Nicolas Winckler, and Stéphane Lathuilière. Privacy-preserving adaptive re-identification without image transfer. In *European Conference on Computer Vision*, pp. 95–111, 2024.

Sashank J. Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečný, Sanjiv Kumar, and H. Brendan McMahan. Adaptive federated optimization. In *International Conference on Learning Representations*, 2021.

Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4938–4947, 2020.

Seonguk Seo, Jinkyu Kim, Geeho Kim, and Bohyung Han. Relaxed contrastive learning for federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12279–12288, 2024.

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

Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision*, pp. 480–496, 2018.

Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated prototype learning across heterogeneous clients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8432–8440, 2022a.

Yue Tan, Guodong Long, Jie Ma, Lu Liu, Tianyi Zhou, and Jing Jiang. Federated learning from pre-trained models: A contrastive learning approach. In *Advances in Neural Information Processing Systems*, 2022b.

Yu-Syuan Tseng, Tzu-Chin Hsu, Chih-Ting Liu, and Shao-Yi Chien. Fedcapr: Federated camera-aware unsupervised person re-identification with identity-distributed equalization for decentralized data clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1787–1796, 2025.

Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H. Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7610–7619, 2020.

Menglin Wang, Baisheng Lai, Haokun Chen, Jianqiang Huang, Xiaojin Gong, and Xian-Sheng Hua. Towards precise intra-camera supervised person re-identification. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3229–3238, 2021.

Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 79–88, 2018.

Jianfeng Weng, Kun Hu, Tingting Yao, Jingya Wang, and Zhiyong Wang. Federated unsupervised cluster-contrastive learning for person re-identification: A coarse-to-fine approach. *Computer Vision and Image Understanding*, 237:103831, 2023.

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, Weilong Li, Wei Liu, Wenke Huang, Zhixi Yu, Bin Yang, Xiaoying Liao, and Kui Jiang. Fedbprompt: Federated domain generalization person re-identification via body distribution aware visual prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 40448–40457, 2026a.

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, 2026b.

Fengxiang Yang, Zhun Zhong, Zhiming Luo, Shaozi Li, and Nicu Sebe. Federated and generalized person re-identification through domain and feature hallucinating. *arXiv preprint arXiv:2203.02689*, 2022.

Fengxiang Yang, Zhun Zhong, Zhiming Luo, Yifan He, Shaozi Li, and Nicu Sebe. Diversity-authenticity co-constrained stylization for federated domain generalization in person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 6477–6485, 2024.

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.

Xiao Zhang, Yixiao Ge, Yu Qiao, and Hongsheng Li. Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3436–3445, 2021.

Xinyu Zhang, Dongdong Li, Zhigang Wang, Jian Wang, Errui Ding, Javen Qinfeng Shi, Zhaoxiang Zhang, and Jingdong Wang. Implicit sample extension for unsupervised person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7369–7378, 2022.

Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *IEEE International Conference on Computer Vision*, pp. 1116–1124, 2015.

Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1318–1327, 2017.

Xiangping Zhu, Xiatian Zhu, Minxian Li, Vittorio Murino, and Shaogang Gong. Intra-camera supervised person re-identification: A new benchmark. In *IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1079–1087, 2019.

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 ACM International Conference on Multimedia*, pp. 955–963, 2020.

Weiming Zhuang, Yonggang Wen, and Shuai Zhang. Joint optimization in edge-cloud continuum for federated unsupervised person re-identification. In *Proceedings of the ACM International Conference on Multimedia*, pp. 433–441, 2021.

Table 4: Camera-client partitions. A local ID instance counts a physical identity separately in each camera where it appears.

Dataset	Clients	Images	Global IDs	Local ID instances
Market1501	6	12,936	751	3,262
MSMT17	15	32,621	1,041	4,821

A Protocol and implementation details

Table 4 summarizes the full-data camera-client partitions; the reduced-data follow-up protocols are specified in Appendix B.

For Market1501, the six clients contain 2,017, 1,709, 2,707, 920, 2,338, and 3,245 training images. MSMT17 client sizes range from 203 images and 14 local identities to 4,910 images and 669 local identities. This imbalance follows the original camera distribution rather than a synthetic Dirichlet partition.

The default association configuration uses a clip norm of 1.0, projection dimension of 256, noise standard deviation of 0.02, association temperature of 0.01, dustbin score of 0, 50 Sinkhorn iterations, minimum probability of 0.20, and minimum margin of 0.02. The relation contrastive temperature is 0.07. All clients participate in every round. Checkpoints and evaluation are produced every 25 rounds.

Relation-loss configurations. The Legacy control uses a pooled-positive objective. Let $v_{nk} = z_n^\top m_k/\tau_r$, let w_{nk} be the stored matching weight, and let $\mathcal{B}_+ = \{n \in \mathcal{B} : \mathcal{P}(y_n) \neq \emptyset\}$. Its loss is

$$\mathcal{L}^c_{\text{legacy}} = -\frac{1}{|\mathcal{B}_+|} \sum_{n\in\mathcal{B}_+} \log \frac{\sum_{k\in\mathcal{P}(y_n)} w_{nk} \exp(v_{nk})}{\sum_{\ell=1}^{K} \exp(v_{n\ell})}, \tag{11}$$

with loss zero when $\mathcal{B}_+$ is empty. PPIA uses the accepted matching score for w_{nk}, with a floor of 10^{-6}. Raw RNN uses normalized, unperturbed prototypes, accepts mutual cosine nearest neighbors, and maps their cosine similarity s to $w = \max(10^{-6}, \min(1, (s+1)/2))$. Relative to Equations 8–9, Legacy changes positive-target aggregation, sample weighting, and the averaging denominator. The ablation compares these combined changes.

B Fixed-coefficient selection and controlled-overlap protocols

The follow-up comprises 134 completed runs: 20 development runs, 12 full-data fixed-0.5/schedule runs, six selected-constant confirmations, and 96 overlap runs. All use the final round-125 checkpoint, rather than a best test checkpoint. These fresh runs form their own paired comparison and are not pooled with historical runs in Tables 1 and 2. Fixed confidence in this comparison means a fixed global coefficient with per-relation confidence weighting retained.

For coefficient selection, split seed 20260909 assigns 20% of official training identities to validation and the remainder to development training. The split is fixed across candidate coefficients and development seeds. Validation mAP is computed on the evaluator-only validation partition of the held-out identities; it selects the coefficient and is not comparable with official test mAP. We select the highest mean final holdout mAP, breaking ties toward the smaller coefficient. Official test data are not used for this choice. Table 5 reports the complete selection screen. The pre-existing linear schedule remains fixed at 1.6 through training index 25 and decreases to 0.2 at index 125; indices start at zero, so the round-125 checkpoint (index 124) uses 0.214.

For overlap construction, we select the six cameras with the most local identities having at least four images, using training availability only. This retains all Market1501 cameras and MSMT17 cameras 1, 5, 7, 11, 13, and 14. Integer-program constraints enforce exactly 80 local classes per camera and at most one singleton or camera-pair assignment per physical identity. At shared-local fractions 0, 0.2, 0.5, and 0.8, the respective numbers of shared physical identities are 0, 48, 120, and

Table 5: Holdout mAP (%), mean ± sample SD over development seeds 42 and 123. Both datasets select 0.8. These held-out training-identity scores are used for coefficient selection and are not directly comparable with official test mAP.

Fixed coefficient	Market1501	MSMT17
0.2	60.74 ± 0.23	31.46 ± 0.36
0.5	67.19 ± 0.06	34.69 ± 0.43
0.8	67.38 ± 0.26	35.00 ± 0.07
1.2	64.68 ± 0.28	33.77 ± 0.24
1.6	62.25 ± 0.62	31.33 ± 0.57

192; singleton counts are 480, 384, 240, and 96. The fraction counts local classes with a counterpart, not pairwise camera Jaccard overlap. Total physical-identity counts therefore vary from 480 to 288 despite fixed local-class and image budgets.

Each seed (42, 123, 3407) defines both a construction and training replicate. Every matcher shares the same subset within a seed and level. Reported variation consequently combines subset and training randomness. MSMT17 conclusions concern six well-populated cameras, not the complete 15-camera distribution. Identity composition may change across overlap levels, so cross-level differences do not isolate overlap alone.

All matchers use the same clipping, projection, noise, and refresh seed. Accepted relation weights are set to one with the outer loss normalized by the full batch and coefficient fixed at 0.5. RNN is reciprocal nearest-neighbor matching on protected summaries; forced matching uses per-camera-pair Hungarian assignment. No-link disables relation learning. These settings separate matching choices from confidence calibration and scheduling. Tables 6 and 7 report final retrieval and paired differences from No-link.

Global mappings and original identity-bearing filenames remain in private construction/evaluation artifacts. Public manifests expose opaque image paths, independently shuffled camera-local labels, and camera identifiers. Leakage and symlink checks verify the constructed public inputs; normal training configurations omit oracle roots and oracle matching. This separation prevents accidental label consumption, not malicious filesystem access.

Table 6: Controlled-overlap retrieval on Market1501 (round 125; %; three paired construction/training seeds, mean ± sample SD). ΔmAP: paired gain over No-link, in percentage points.

Shared fraction	Matcher	mAP	R1	ΔmAP
0.0	No-link	9.39 ± 0.05	23.47 ± 0.41	—
0.0	Protected RNN	10.96 ± 0.19	26.45 ± 0.08	1.57 ± 0.16
0.0	Forced matching	11.66 ± 0.18	27.72 ± 0.30	2.27 ± 0.15
0.0	Partial OT	11.58 ± 0.06	27.74 ± 0.38	2.19 ± 0.06
0.2	No-link	9.33 ± 0.23	23.74 ± 0.74	—
0.2	Protected RNN	11.12 ± 0.22	26.89 ± 0.48	1.79 ± 0.03
0.2	Forced matching	11.43 ± 0.06	27.34 ± 0.31	2.10 ± 0.22
0.2	Partial OT	11.57 ± 0.19	27.66 ± 0.09	2.25 ± 0.12
0.5	No-link	9.35 ± 0.12	23.48 ± 0.58	—
0.5	Protected RNN	11.34 ± 0.40	27.17 ± 1.11	1.98 ± 0.30
0.5	Forced matching	11.61 ± 0.20	27.30 ± 0.52	2.25 ± 0.11
0.5	Partial OT	11.71 ± 0.37	27.80 ± 0.84	2.36 ± 0.28
0.8	No-link	9.36 ± 0.15	23.34 ± 0.26	—
0.8	Protected RNN	11.41 ± 0.19	27.55 ± 0.58	2.05 ± 0.22
0.8	Forced matching	11.65 ± 0.18	27.86 ± 0.87	2.29 ± 0.10
0.8	Partial OT	11.58 ± 0.36	27.44 ± 1.00	2.22 ± 0.35

Table 7: Controlled-overlap retrieval on MSMT17 (round 125; %; three paired construction/training seeds, mean ± sample SD). ΔmAP: paired gain over No-link, in percentage points.

Shared fraction	Matcher	mAP	R1	ΔmAP
0.0	No-link	2.85 ± 0.08	10.61 ± 0.22	—
0.0	Protected RNN	3.14 ± 0.14	11.20 ± 0.40	0.29 ± 0.06
0.0	Forced matching	3.30 ± 0.21	11.78 ± 0.53	0.45 ± 0.14
0.0	Partial OT	3.28 ± 0.20	11.61 ± 0.44	0.43 ± 0.13
0.2	No-link	2.84 ± 0.10	10.49 ± 0.19	—
0.2	Protected RNN	3.14 ± 0.22	11.13 ± 0.38	0.30 ± 0.16
0.2	Forced matching	3.34 ± 0.22	11.82 ± 0.50	0.51 ± 0.15
0.2	Partial OT	3.37 ± 0.25	11.75 ± 0.59	0.53 ± 0.18
0.5	No-link	2.88 ± 0.05	10.71 ± 0.15	—
0.5	Protected RNN	3.22 ± 0.17	11.47 ± 0.38	0.34 ± 0.13
0.5	Forced matching	3.39 ± 0.10	11.99 ± 0.06	0.51 ± 0.05
0.5	Partial OT	3.43 ± 0.17	12.03 ± 0.22	0.55 ± 0.12
0.8	No-link	2.87 ± 0.06	10.55 ± 0.22	—
0.8	Protected RNN	3.20 ± 0.18	11.38 ± 0.44	0.33 ± 0.13
0.8	Forced matching	3.42 ± 0.08	12.15 ± 0.06	0.55 ± 0.02
0.8	Partial OT	3.41 ± 0.08	12.03 ± 0.19	0.55 ± 0.02

C Additional related context

More general federated optimizers address heterogeneous local progress through proximal regularization, control variates, normalized aggregation, or adaptive server updates (Li et al., 2020; Karimireddy et al., 2020; Wang et al., 2020; Reddi et al., 2021). Personalized objectives and relaxed contrastive learning provide complementary ways to accommodate client-specific distributions or representation mismatch (Li et al., 2021b; Seo et al., 2024).

Person ReID representation and retrieval. Strong centralized ReID systems have improved training recipes, part-based representations, and spatially distributed features (Luo et al., 2019; Sun et al., 2018; Dai et al., 2019). Neighborhood structure has also been used for retrieval reranking (Zhong et al., 2017), while implicit sample extension targets noisy clustering in unsupervised ReID (Zhang et al., 2022). These works improve the image representation or retrieval stage under centralized access; our problem is the recovery of training relations across separated camera clients.

Dynamic and reliability-aware learning. Client drift, feature shift, and local/global inconsistency have motivated control variates, local normalization, contrastive objectives, and dynamic regularization in general federated learning (Karimireddy et al., 2020; Li et al., 2020; 2021a;c; Acar et al., 2021). Noisy pseudo-label learning in ReID likewise benefits from confidence and cross-generation consistency (Zhang et al., 2021). Dynamic weighting and confidence calibration are therefore established ideas. We apply them to the changing relative influence of inferred cross-client identity relations.

D Degree-preserving randomized-relation controls

The 48 new runs cover two datasets, four overlap settings, two matchers, and seeds 42, 123, and 3407. They reuse the corresponding public subsets and normal-matcher results from Appendix B. At each refresh, a local seeded random-number generator samples a uniform permutation of right endpoints independently within each camera-pair matching. It reads no global identity labels. One-to-one input matchings guarantee distinct edges and preserve both endpoint multisets; chance fixed points and chance true links are allowed. Accepted weights remain one, the batch-normalized outer loss coefficient remains 0.5, and initialization, optimization, evaluation sets, and the final round-125 endpoint are held fixed.

The preservation guarantee applies to each graph before and after permutation. Independently trained partial-OT models can subsequently produce different graph sizes, so that comparison does not fix degrees across trajectories. Forced matching always yields 15 camera pairs times 80 edges, with degree five for each of 480 local classes, providing the sharper graph-size control. The three seeds combine construction and training variation; paired differences are descriptive, without a significance claim.

Table 8 includes all outcomes. At zero overlap, normal relations cannot be correct identity matches, so their advantage over randomized relations does not verify identity recovery. On Market1501 the randomized means fall below No-link in every setting, while on MSMT17 they remain above it. This dataset dependence limits the conclusion to the tested comparisons and cautions against assigning all relation-loss effects to one mechanism.

Table 8: Randomized controls at round 125 (mAP, %; three-seed mean $\pm$ sample SD). Δ: paired permuted-minus-normal mAP, in percentage points.

Dataset	Fraction	Matcher	Normal	Permuted	Δ
Market1501	0	Forced	11.66 ± 0.18	8.86 ± 0.17	-2.80 ± 0.27
Market1501	0	Partial OT	11.58 ± 0.06	8.83 ± 0.19	-2.75 ± 0.14
Market1501	0.2	Forced	11.43 ± 0.06	8.74 ± 0.14	-2.68 ± 0.16
Market1501	0.2	Partial OT	11.57 ± 0.19	8.77 ± 0.08	-2.80 ± 0.26
Market1501	0.5	Forced	11.61 ± 0.20	8.72 ± 0.11	-2.89 ± 0.10
Market1501	0.5	Partial OT	11.71 ± 0.37	8.67 ± 0.11	-3.04 ± 0.27
Market1501	0.8	Forced	11.65 ± 0.18	8.69 ± 0.07	-2.96 ± 0.23
Market1501	0.8	Partial OT	11.58 ± 0.36	8.67 ± 0.06	-2.92 ± 0.36
MSMT17	0	Forced	3.30 ± 0.21	2.96 ± 0.04	-0.34 ± 0.19
MSMT17	0	Partial OT	3.28 ± 0.20	2.93 ± 0.05	-0.35 ± 0.16
MSMT17	0.2	Forced	3.34 ± 0.22	2.95 ± 0.07	-0.39 ± 0.16
MSMT17	0.2	Partial OT	3.37 ± 0.25	2.88 ± 0.05	-0.49 ± 0.21
MSMT17	0.5	Forced	3.39 ± 0.10	2.96 ± 0.06	-0.43 ± 0.06
MSMT17	0.5	Partial OT	3.43 ± 0.17	2.91 ± 0.02	-0.52 ± 0.15
MSMT17	0.8	Forced	3.42 ± 0.08	2.92 ± 0.06	-0.50 ± 0.08
MSMT17	0.8	Partial OT	3.41 ± 0.08	2.93 ± 0.03	-0.49 ± 0.06

E CLAIM AND EXPOSURE BOUNDARIES

The public training protocol prevents accidental consumption of benchmark global labels. It does not prevent a malicious process with filesystem access from opening evaluator files. Random projection and Gaussian noise alter summaries used by matching but have no formal privacy interpretation here. The simulator observes intermediate prototype objects, and accepted relations route detached remote-prototype targets. Accordingly, *protected summary* and *leakage-controlled protocol* refer to matching-input obfuscation and restrictions on label visibility. They do not imply secure multiparty computation, differential privacy, resistance to reconstruction, or resistance to identity linkage.