

# FedARKS: Federated Aggregation via Robust and Discriminative Knowledge Selection and Integration for Person Re-identification

Xin Xu<sup>1</sup>, Binchang Ma<sup>1</sup>, Zhixi Yu<sup>1\*</sup>, Wei Liu<sup>1\*</sup>

<sup>1</sup> School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065 China  
xuxin@wust.edu.cn, 202213407218@wust.edu.cn, yuzhixi@wust.edu.cn, liuwei@wust.edu.cn

## Abstract

The application of federated domain generalization in person re-identification (FedDG-ReID) aims to enhance model's generalization ability in unseen domains while protecting client data privacy. However, existing methods typically rely on global feature representation using simple averaging operations for model aggregation, leading to two limitations in domain generalization: (1) relying only on global features makes it difficult to capture domain-invariant local details (such as accessories or features); (2) Uniform parameter averaging treats all clients as equivalent, ignoring their differences in robust feature extraction capabilities, thereby diluting the contributions of quality clients. To address these issues, we propose a federated learning framework—**Federated Aggregating Robust and Discriminative Knowledge Selection and Integration (FedARKS)**—comprising two mechanisms: **RK (Robust Knowledge)** and **KS (Knowledge Selection)**. In this design, each client employs a dual-branch network of RK: the **Global Feature Processing Branch** serves as the primary component, extracting overall representations for model aggregation and server-side updates; while the **Body Part Processing Branch** acts as an auxiliary component, focusing on extracting domain-invariant local details to supplement and guide the local training process during global feature learning. Additionally, our **KS** mechanism adaptively assigns corresponding aggregation weights to clients based on their ability to extract domain-invariant knowledge, enabling the server to better integrate cross-domain invariant knowledge extracted by clients. Extensive experiments validate that FedARKS achieves state-of-the-art generalization results on the FedDG-ReID benchmark, demonstrating that learning subtle body part features can effectively assist and reinforce global representations, thereby enabling robust cross-domain person ReID capabilities.

## Introduction

Person Re-identification (ReID) aims to retrieve target pedestrians across non-overlapping camera views, playing a crucial role in smart city applications such as public security and missing person searches. Although deep learning methods (He et al. 2016; Huang et al. 2022; Wang et al. 2018; Ye et al. 2022; Sun et al. 2018) have significantly improved

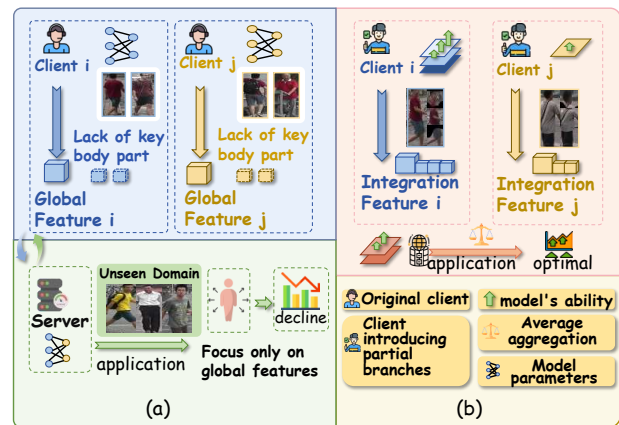


Figure 1: Core challenges in FedDG-ReID: (a) Ignoring key local details that are invariant across domains; (b) Aggregation methods based on averaging dilute the ability of clients that are good at extracting cross-domain invariant features, thereby suppressing overall generalization ability.

ReID performance, they remain highly susceptible to domain shift, resulting in substantial performance degradation and reduced reliability when deployed in new environments with different camera characteristics, lighting conditions, or pedestrian distributions.

Recent efforts (Zhao et al. 2021; Song et al. 2019; Dai et al. 2021) address this challenge through Domain Generalization (DG) techniques, which train models on multiple labeled source domains to learn domain-invariant representations and mitigate domain shift. However, growing concerns about personal data protection restrict the development of DG-ReID, as its training relies on centralizing large volumes of personal image data, posing potential privacy risks to individuals. Therefore, it is necessary to develop privacy-preserving DG-ReID methods.

To address privacy preservation concerns, recent research has applied Federated Learning (FL) to Domain Generalization in Person Re-identification (FedDG-ReID). While existing FedDG-ReID approaches (Wu and Gong 2021) have made progress in mitigating local overfitting and adapting generic FL techniques, they crucially overlook two funda-

\*Corresponding author.

mental challenges:

**First**, as visualized in Fig. 1(a), cross-client **feature misalignment** arises when key local body part details—crucial for identity recognition and often possessing stronger cross-domain invariance—are neglected in feature representations. Specifically, due to data heterogeneity and limited coverage, clients often miss key body-part details critical for identity recognition and cross-domain invariance. Consequently, local models miss these important discriminative cues, leading to feature misalignment during aggregation, degrading generalization on unseen domains.

**Second**, as shown in Fig. 1(b), conventional averaging-based aggregation methods indiscriminately merge client updates, overlooking differences in each client’s capacity to capture robust, domain-invariant representations. As a result, the distinct advantages of clients adept at extracting such transferable features are diminished, leading to suboptimal integration of discriminative knowledge at the server. This ultimately suppresses the global model’s ability to generalize effectively across diverse domains.

To address these challenges, we propose FedARKS (as shown in Fig.2), which introduces two complementary mechanisms for enhanced domain generalization. The first mechanism, FedARKS-RK (Robust Knowledge), is deployed on each client and consists of two specialized branches: a body part processing branch for learning fine-grained local features, and a global feature processing branch for capturing holistic representations. To alleviate cross-client pedestrian feature misalignment, we fuse global features (from the Global Feature Processing Branch) and part-combined features (from the Body Part Processing Branch) via weighted averaging. Crucially, only the global branch parameters are aggregated, allowing the server to model domain-invariant features for better generalization. The part branch parameters remain local, where they can effectively capture client-specific, fine-grained cues based on local data. Aggregating these heterogeneous part parameters would undermine their discriminative power, adaptability and capacity for personalized representation

Building upon this, the second mechanism, FedARKS-KS (Knowledge Selection), is introduced to make the most of the robust local knowledge developed by FedARKS-RK. Specifically, the KS module adaptively assigns greater aggregation weights to clients with stronger domain-invariant feature extraction capabilities. This targeted aggregation mechanism ensures that the server can fully leverage the domain-invariant knowledge from all clients, thus amplifying the overall generalization capacity of the global model on both seen and unseen domains. Comprehensive experiments validate the effectiveness of our framework, showing robust generalization across person ReID datasets. Our contributions are summarized as follows:

- We analyze the key factors impacting the generalization ability of federated person ReID models, revealing that traditional aggregation often overlooks subtle, domain-invariant local features, and fails to utilize the domain-generalization advancements contributed by individual clients, resulting in limited cross-domain robustness of the aggregated model.

- We propose FedARKS, a federated learning framework specifically designed for person ReID, whose core lies in the combination of two complementary mechanisms. FedARKS combines robust local feature extraction with adaptive knowledge aggregation: at each client, it separates and retains fine-grained local feature representations with critical local details, while the server selectively integrates these cross-domain invariant features from all clients. This design enables the server-side model to fully integrate domain-invariant features, thereby enhancing its generalization capabilities.
- We introduce the Representation Selection (RS) and Knowledge Selection (KS) modules. The RS module enables each client to capture subtle and highly discriminative local features with strong cross-domain invariance, whereas the KS module adaptively assigns higher aggregation weights to clients with superior domain-invariant representation learning capabilities. This collaborative mechanism ensures that the global model fully leverages transferable knowledge across diverse domains and mitigates the adverse effects of feature misalignment and indiscriminate aggregation.

## Related Work

### Domain Generalization

Domain Generalization (DG), due to its significant practical value, has become an emerging hotspot in the field of AI. Current mainstream DG methods, such as domain alignment techniques(Li et al. 2018b,d,c), meta-learning frameworks(Li et al. 2019; Du et al. 2020; Zhao et al. 2021; Li et al. 2018a), and style/data augmentation strategies(Zhou et al. 2020; Zhang et al. 2022), have demonstrated excellent performance under centralized training settings. Existing domain generalization (DG) methods achieve strong performance in centralized settings, yet their frequent dependence on aggregated data collection poses substantial privacy concerns. Within federated learning (FL), DG encounters critical obstacles including client-level data scarcity – characterized by insufficient sample diversity – and the absence of domain-specific labels. These limitations severely hinder the deployment of meta-learning frameworks and domain alignment mechanisms. Notably, for challenging ReID applications, the ability to extract and distill fine-grained local details—such as body-part features—has proven crucial for domain-invariant representation, but remains challenging to leverage under federated constraints.

### Federated Learning

Federated Learning (FL)(Konečný et al. 2016; McMahan et al. 2017; Mohassel and Zhang 2017; Ma et al. 2025; Huang et al. 2024, 2023a,b; Huang, Ye, and Du 2022; Liu et al. 2025; Xu et al. 2022) enables multiple clients to collaboratively train shared or personalized models while preserving data privacy. The early representative FedAvg(McMahan et al. 2017)adopts a mechanism of local model averaging and redistribution. Subsequent studies like FedProx(Li

et al. 2020), MOON(Li, He, and Song 2021), and SCAF-FOLD(Karimireddy et al. 2020) aim to mitigate local overfitting issues, though they primarily target closed-set settings (e.g., image classification with fixed classes). To address open-set(Panareda Busto and Gall 2017) tasks, Zhuang et al.(Zhuang et al. 2020)proposed FedPav, which adapts to re-ID tasks by exchanging feature extractors. Wu et al.(Wu and Gong 2021)pioneered the definition of the FedDG-ReID problem (where each source domain corresponds to a client) and introduced model distillation as a solution. The core of these methods revolves around maintaining consistency between local and global model.However, federated settings for person ReID must also contend with severe occlusions and intra-client variation, motivating the need for structured, body part-aware representations. Building upon this motivation, recent methods(Yang et al. 2022; Gao et al. 2020; Miao et al. 2019; Wang, Chen, and Liu 2021; Xu, Zhao, and Qin 2021; Zhang et al. 2021; Kalayeh et al. 2018; Song et al. 2018; Yang et al. 2020) suggest reconstructing structured sets of body parts within the source domain, enabling networks to focus on visible, informative regions, even without universal part annotations.

Specifically, we design a dual-branch network architecture, where one branch specializes in learning and distilling feature knowledge from these predefined body parts, and the other branch serves as the primary model learning from source domain data for re-identification tasks. Through our RK mechanism, the fine-grained features from the body-part branch are effectively distilled into the primary model, thus greatly enhancing its discriminative power on general pedestrian datasets and significantly improving model generalization across diverse unseen domains, even when facing substantial domain shifts or novel environments during deployment. This work draws on the stylization method proposed by(Yang et al. 2024), while further demonstrating its effectiveness in real-world cross-domain scenarios.

## Methodology

### Preliminaries

**Problem Definition.** Given  $N$  labeled source domains  $S = \{D_1, D_2, \dots, D_K\}$ , where the  $i$ -th domain  $D_i = \{X_i, Y_i\}$  is comprised of  $M_i$  training images  $X_i$  and their corresponding identity labels  $Y_i$ . FedDG-ReID constructs a federated architecture where each data domain serves as an independent node, enforcing strict data isolation constraints. Its core principle lies in leveraging a central server to coordinate distributed clients. Through a Client-Server Collaborative Learning (CSCL) mechanism, it jointly optimizes the person ReID model to enhance its generalization performance on unseen target domains. This framework adheres to the cross-silo federated learning paradigm(Kairouz et al. 2021). Under the strict prohibition of data centralization, FedDG-ReID presents greater challenges compared to classical domain generalization methods for person ReID.

### Extraction of Robust Knowledge

**Motivation.** In federated person re-identification, data collected from distributed clients exhibit significant heterogeneity due to variations in viewpoint, pose and background, presenting a non-IID challenge.Traditional global feature learning paradigms primarily emphasize holistic appearance cues derived from the entire person bounding box, often systematically overlooking subtle yet domain-invariant local discriminative features—such as distinctive accessories, unique fabric textures, or semantically key body part—which are crucial for robust matching under cross-client variations. This critical oversight inevitably results in suboptimal feature alignment during aggregation and severely hampers cross-client generalization. To decisively overcome these limitations, we propose integrating a dedicated RK module within each client’s local learning process; this module is meticulously designed to extract specific body parts (e.g., head/shoulders, torso, lower body/legs) .This module enables (1) robust and adaptive learning despite extreme data heterogeneity, (2) effective preservation, weighting, and fusion of discriminative local features during server-side aggregation, and (3) richer, more comprehensive, and structurally informed identity modeling. Consequently, the RK module empowers the global model to integrate diverse and salient identity cues from multiple clients, significantly enhancing re-identification performance in challenging federated learning environments—while strictly preserving data privacy.

**Dual-Branch Network.** RK employs a dual-branch design: a **Body Part Processing Branch** and a **Global Feature Processing Branch**. First, the PifPaf pose estimation model(Kreiss, Bertoni, and Alahi 2019)is used to segment pedestrian images into three main body regions—the head, the torso (including the left and right shoulders and hips), and the lower limbs (including the left and right knees and legs). The Body Part Processing Branch is used in the client-side model to process these body part images. Since the body part features learned by the Body Part Processing Branch vary depending on the local data distribution of the client, forcing the aggregation of the Body Part Processing Branch parameters would compromise the integrity of client-specific body part details. Therefore, this branch only runs during the training phase. During federated aggregation, only the parameters of the Global Feature Processing Branch are aggregated, and the server only distributes the parameters of the Global Feature Processing Branch. During training, the Body Part Processing Branch extracts domain-invariant knowledge into the main model (Global Feature Processing Branch), enhancing its generalization ability. The Global Feature Processing Branch processes raw full-image data. During the local training phase on the client side, we fuse the combined body part feature vectors extracted from the Body Part Processing Branch with the overall features output by the Global Feature Processing Branch, which constitutes the final output features of the client model. The final characteristics of each client output are defined as follows:

$$\theta_{\text{final}} = \partial \cdot \theta_{\text{global}} + (1 - \partial) \cdot \theta_{\text{part}}, \quad (1)$$

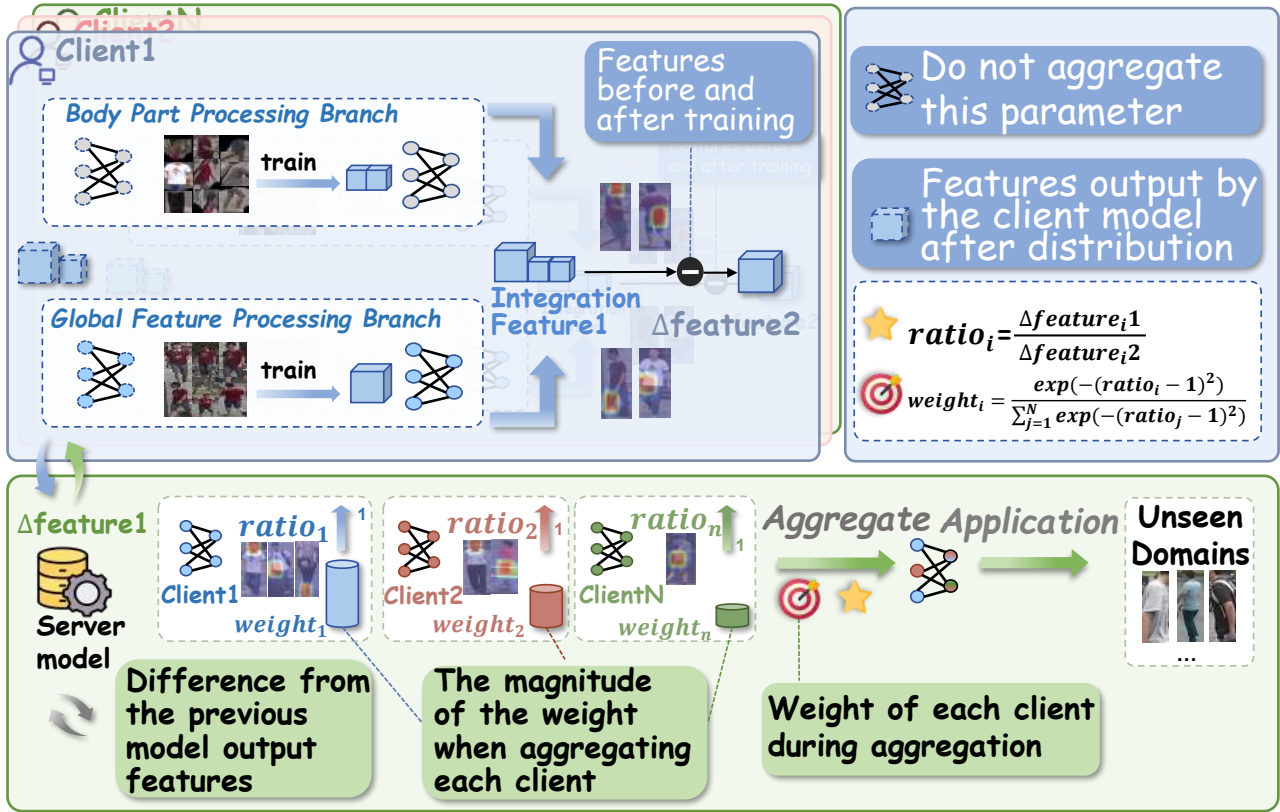


Figure 2: Architectural diagram of FedARKS. Our method employs local branch training, in which each client fuses features in a dual-branch network. This figure illustrates the detailed process of local feature fusion and parameter aggregation, using client 1 as an example (other clients are shown in the background). After local training is complete, the server dynamically calibrates the aggregation weights of each client based on feature difference vectors and updates the global model in the next round of communication to improve cross-domain generalization capabilities.

where  $\theta_{\text{global}}$  denotes the Global feature vector output by Global Feature Processing Branch,  $\theta_{\text{part}}$  denotes the Combined local features output by the Body Part Processing Branch,  $\vartheta$  denotes the weights assigned to the global features (default  $\vartheta = 0.5$ ). The optimization objective of each client model is defined as follows:

$$L_{\text{loc}}(x, y; f_L) = L_{\text{tri}}(\theta_L(x), y) + L_{\text{ce}}(f_L(x), y), \quad (2)$$

where  $\theta_L$  is the feature extractor part of local model  $f_L$ ,  $L_{\text{CE}}$  denotes the cross-entropy loss for part-level identification,  $L_{\text{tri}}$  represents the triplet loss for metric learning.

### Federated Aggregation with Robust Knowledge Selection

**Motivation.** Although introducing Body Part Processing Branch enables each client to learn more refined and discriminative local features, there are still significant differences in the effectiveness of different clients in extracting such domain-invariant body-part-based knowledge. If the server simply aggregates all client models by averaging unified parameters, the unique contributions of clients that excel in capturing high-quality local details will be diluted by clients with poorer learning performance. This approach

limits the generalization ability of the global model. To fully leverage the cross-domain invariant knowledge possessed by each client, we introduce KS(knowledge selection). This mechanism objectively evaluates and assigns higher aggregation weights to clients that excel in learning domain-invariant local features. The collaboration between RS and KS ensures that the global model benefits more from cross-domain invariant knowledge, ultimately enhancing overall performance and adaptability across diverse domains.

**Directional Consistency Metric.** The direction consistency metric aims to assess the consistency between the update direction of each client’s local model and the update direction of the global model through a bidirectional vector comparison framework. Specifically, this metric is calculated based on two types of feature differences: (1) the difference between the current global model and the aggregated features of the previous version of the global model, which reflects the update direction of the global model; (2) the difference between the features output by the client model before and after local training, which captures the local update direction of that client. The ratio of the norms of these two feature differences serves as a quantitative measure of directional consistency. The closer the ratio is to 1, the higher the align-

ment between the client’s update direction and the global model’s update direction. This ratio directly influences the dynamic weight adjustment mechanism during the aggregation process: clients with higher directional consistency receive higher aggregation weights, while those with lower directional consistency receive lower aggregation weights.

The directional consistency metric and the aggregation weight are mathematically defined as follows:

$$\mathcal{R}_k^{(t)} = \frac{\left\| \theta_g^{(t)} - \theta_g^{(t-1)} \right\|_2}{\left\| \theta_k^{(t+\Delta t)} - \theta_k^{(t)} \right\|_2}, \quad (3)$$

where  $\theta_g^{(t)}$  denotes the global model parameters after the  $t$ -th round of aggregation, serving as the server-side global state to guide client initialization. The initial parameters after receiving the new global model are represented as  $\theta_k^{(t)}$ . Subsequently, the client performs  $\Delta t$  rounds of optimization based on local data, yielding the updated  $\theta_k^{(t+\Delta t)}$ , which reflect its personalized adjustment results.

**Dynamic Weight Allocation.** Based on the direction consistency metric, our dynamic aggregation strategy can adaptively determine the contribution weight of each client in federated aggregation. Specifically, the weight assigned to each client is an exponentially decaying function of the deviation between its direction consistency ratio and the ideal value of 1, controlled by a temperature coefficient. This ensures that clients whose local updates are highly consistent with the global target receive higher aggregation weights, while clients with larger deviations are correspondingly assigned lower weights. Before aggregation, the obtained weights are normalized to ensure that the sum is 1. To further enhance stability, a protection mechanism freezes the weights of clients with very small local updates to prevent noise amplification, and applies historical smoothing to prevent the aggregated weights from undergoing drastic changes between different rounds.

The adaptive weighting mechanism and associated enhancement techniques are mathematically defined as follows:

$$\alpha_k^{(t)} = \frac{\exp\left(-\beta \left| 1 - \mathcal{R}_k^{(t)} \right| \right)}{\sum_{j=1}^K \exp\left(-\beta \left| 1 - \mathcal{R}_j^{(t)} \right| \right)}, \quad (4)$$

where  $\alpha_k^{(t)}$  denotes the aggregation weight assigned to client  $k$  at communication round  $t$ ,  $\beta$  is the temperature coefficient (default  $\beta = 5.0$ ) controlling weight decay intensity.

$$\theta_g^{(t+1)} = \sum_{k=1}^K \alpha_k^{(t)} \cdot \theta_k^{(t+\Delta t)}, \quad (5)$$

where  $(\theta_g^{(t+1)})$  denotes the global model parameters after the  $(t + 1)$ th round of aggregation.

$$\tilde{\mathcal{R}}_k^{(t)} = \gamma \mathcal{R}_k^{(t)} + (1 - \gamma) \tilde{\mathcal{R}}_k^{(t-1)}, \quad (6)$$

where  $\gamma \in [0, 1]$  denotes the smoothing factor (default  $\gamma = 0.7$ ) for historical consistency regularization.

$$\alpha_k^{(t)} = \begin{cases} 0 & \text{if } \left\| \theta_k^{(t+\Delta t)} - \theta_k^{(t)} \right\|_2 < \varepsilon, \\ \alpha_k^{(t)} & \text{otherwise,} \end{cases} \quad (7)$$

---

Algorithm 1: **FedARKS**: Federated Aggregation via Robust and Discriminative Knowledge Selection

---

**Require:** Total rounds  $T$ , client set  $\mathcal{C}$ , local epochs  $E$ , temperature  $\beta$ , threshold  $\varepsilon$   
**Ensure:** Optimized global model  $\theta_g$

- 1: Initialize global model parameters  $\theta_g^{(0)}$
- 2: **for** each round  $t = 1, 2, \dots, T$  **do**
- 3:   *// Server broadcasts the global model*
- 4:   Send  $\theta_g^{(t)}$  to all clients  $k \in \mathcal{C}$
- 5:   **for** each client  $k \in \mathcal{C}$  **in parallel do**
- 6:     Initialize local model  $\theta_k \leftarrow \theta_g^{(t)}$
- 7:     **for** epoch = 1 to  $E$  **do**
- 8:       *// Local dual-branch RK training:*
- 9:       Update  $\theta_k$  using RK network with
- 10:       • **Global branch:** learns holistic features for aggregation
- 11:       • **Part branch:** extracts domain-invariant features to regularize the global branch
- 12:     **end for**
- 13:     Compute reliability score  $\mathcal{R}_k$  for the global branch
- 14:     Upload  $(\theta_k, \mathcal{R}_k)$  to server
- 15:     **end for**
- 16:   *// Adaptive knowledge selection and aggregation*
- 17:   **for** each client  $k \in \mathcal{C}$  **do**
- 18:      $w_k \leftarrow \exp(-\beta |1 - \mathcal{R}_k|)$  if  $\mathcal{R}_k > \varepsilon$ ; else  $w_k \leftarrow 0$
- 19:   **end for**
- 20:   Normalize weights:  $\alpha_k = \frac{w_k}{\sum_j w_j}$
- 21:   Update  $\theta_g^{(t+1)} \leftarrow \sum_{k \in \mathcal{C}} \alpha_k \cdot \theta_k$
- 22: **end for**
- 23: **Return**  $\theta_g^{(T)}$

---

where  $\varepsilon$  represents the update magnitude threshold (default  $\varepsilon = 10^{-8}$ ) for confidence gating.

## Experiments

### Experimental Setup

**Datasets.** We evaluate our method on four person re-identification (ReID) datasets:

- **CUHK02** (Li and Wang 2013): Contains 1,816 identities captured by 2 camera views, with 7,264 annotated bounding boxes.
- **CUHK03** (Li et al. 2014): Includes 1,467 identities from 2 camera views, providing 14,096 detected or manually labeled bounding boxes.
- **Market1501** (Zheng et al. 2015): Comprises 32,668 bounding boxes of 1,501 identities captured by 6 disjoint cameras.
- **MSMT17** (Wei et al. 2018): A large-scale dataset with 126,441 bounding boxes of 4,101 identities from 15 cameras, featuring complex lighting and weather variations.

**Model.** For the person re-identification tasks on all datasets, we employ a ResNet50 backbone. During federated training, each client appends a client-specific classifier head where the output dimension equals the number of identities in the

client’s local dataset. For the ViT backbone, we used a batch size of 16 during training.

### Comparison to State-of-the-Arts

As shown in Tab.1, our method consistently outperforms all baselines across different datasets and backbones. Under both ResNet50 and ViT, it achieves the best results, surpassing DACS (Yang et al. 2024) in every evaluation scenario. The gains are particularly evident on challenging datasets such as MSMT17 and CUHK03, highlighting our approach’s strong robustness to both backbone variations and unseen domains. For example, in the M+C2+C3→MS configuration with ViT, our method achieves 18.2% mAP and 40.6% R1, demonstrating excellent scalability to large datasets. Similarly, in MS+C2+M→C3, it maintains 31.5% mAP and 34.7% R1, reflecting high adaptability even in small-scale target domains. Our approach performs strongly with both powerful and lightweight backbone architectures, ensuring practical versatility. Results on source domains (see Tab.2) further validate the effectiveness of our method. FedARKS achieves state-of-the-art performance on all benchmark datasets while adhering to federated learning constraints. Specifically, on Market1501, it sets new SOTA results of 73.5% mAP and 89.4% Rank-1 accuracy (R1), exceeding DACS by  $\uparrow 1.4\%$  mAP and  $\uparrow 1.2\%$  R1. Gains are observed on CUHK02, with 86.8% mAP and 86.5% R1 – improvements of  $\uparrow 2.3\%$  mAP and  $\uparrow 3.1\%$  R1 over DACS. For the challenging CUHK03 dataset, FedARKS attains 54.5% mAP and 56.8% R1, surpassing DACS by  $\uparrow 7.1\%$  mAP and  $\uparrow 6.7\%$  R1. It outperforms the best prior federated method (FedProx) by over 30% in mAP on CUHK03. These results demonstrate FedARKS’s superior cross-domain generalization, architectural adaptability, and scalability for real-world person ReID tasks. Moreover, SAvg results clearly further demonstrate superiority of our method across all evaluated settings.

### Ablation Study

In order to better understand the impact of **RK** and **KS** on the FedDG-ReID results, we gradually incorporated them into the training processes of MS+C2+C3 → M, MS+C2+M → C3, and M+C2+C3 → MS for ablation studies. We compare four different training schemes in Tab.3. and summarize the following conclusions:

**Synergistic Effect of Dual Components:** Integrating both RK and KS yields the most substantial improvements across all tasks, demonstrating strong complementarity. For example, in the MS+C2+C3→M setting, their combination boosts mAP by  $\uparrow 3.1\%$  and R1 by  $\uparrow 3.5\%$ , while in MS+C2+M→C3, the R1 gain reaches  $\uparrow 10.3\%$ .

**RK’s Feature Discrimination Contribution:** RK is most effective in transfers between similar domains, notably improving R1 in MS+C2+C3→M by  $\uparrow 3.0\%$ . Its combination with KS provides additional benefits, as seen in the incremental mAP gain in the M+C2+C3→MS task.

**KS’s Cross-Domain Generalization Strength:** KS excels in challenging domain shifts, achieving larger improvements than RK, especially in MS+C2+M→C3 ( $\uparrow 5.6\%$  mAP and  $\uparrow 7.2\%$  R1 over baseline), establishing KS as the key to

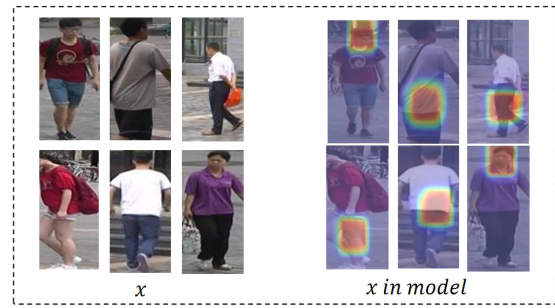


Figure 3: Dynamic Attention Distribution Heatmap Comparison Across Diverse Scenarios.

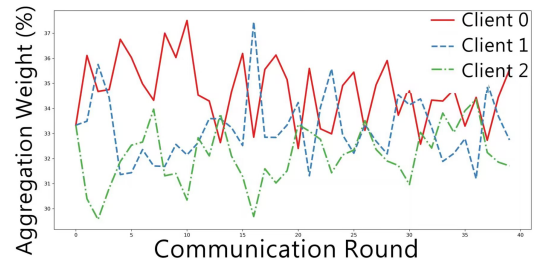


Figure 4: Dynamic client weight distribution across 40 training epochs in FedARKS.

cross-domain adaptation.

**Complementary Properties:** Tab.3 shows that using both RK and KS together delivers greater mAP and R1 improvements than either alone—e.g., in MS+C2+M→C3, mAP increases from 28.5% (RK only) or 30.2% (KS only) to 31.5% (both), and R1 from 28.4% (RK) or 31.6% (KS) to 34.7% (both). This clearly confirms their complementarity and the effectiveness of joint use for enhancing federated cross-domain ReID performance.

### Visualization

To gain a deeper understanding of our RK mechanism’s practical effects, we present comprehensive attention heatmaps comparing network focus patterns in a variety of real-world and challenging scenarios. In standard viewpoints, the RK module adaptively highlights the head and face, which are the most identifiable features for person ReID. In complex conditions—such as rare camera perspectives or cluttered backgrounds—the heatmaps show the model shifting focus to other discriminative regions, such as the torso or lower limbs. This dynamic adaptation proves to be crucial: for example, when the face is occluded or indistinct, unique clothing textures or accessories on the lower body become primary sources of cross-client invariance. Thus, the RK mechanism greatly enhances clients’ ability to mine and use contextually relevant identity features, while promoting aggregation of representations that are both semantically meaningful and robust to environmental or camera-view changes. This effectively addresses poor generalization under non-IID conditions, ensuring the

Category	Method	Venue	MS+C2+C3→M		M+C2+C3→MS		MS+C2+M→C3		SAvg	
			mAP	R1	mAP	R1	mAP	R1	mAP	R1
<b>Backbone: ResNet50</b>										
Federated Learning	SCAFFOLD	ICML'20	26.0	50.5	5.3	15.8	22.9	26.0	18.1	30.8
	MOON	CVPR'21	26.8	51.1	4.8	14.5	20.9	22.5	17.5	29.4
	FedProx	MLSys'20	29.3	53.8	5.8	17.4	19.1	17.7	18.1	29.6
Domain Generalization	SNR	CVPR'20	32.7	59.4	5.1	15.3	28.5	30.0	22.1	34.9
	MixStyle	ICLR'21	31.2	53.5	5.5	16.0	28.6	31.5	21.8	33.7
	CrossStyle	ICCV'21	35.5	59.6	4.6	14.0	27.8	28.0	22.6	33.9
Fed ReID	FedPav	MM'20	25.4	49.4	5.2	15.5	22.5	24.3	17.7	29.7
	FedReID	AAAI'21	30.1	53.7	4.5	13.7	26.4	26.5	20.3	31.3
FedDG-ReID	DACS	AAAI'24	35.5	60.3	9.0	26.7	24.6	24.4	23.0	37.1
	SSCU	MM'25	39.5	66.4	11.9	32.3	32.8	34.1	28.1	44.3
	FedSupWA	AAAI'26	42.0	68.9	12.4	33.4	36.1	36.6	30.2	46.3
	<b>FedARKS</b>	-	<b>42.5</b>	<b>69.2</b>	<b>12.6</b>	<b>33.7</b>	<b>36.8</b>	<b>36.9</b>	<b>30.6</b>	<b>46.6</b>
<b>Backbone: ViT</b>										
Fed ReID	FedPav	MM'20	32.5	56.0	13.6	32.6	25.1	24.0	23.7	37.5
DG	CrossStyle	ICCV'21	30.9	55.1	16.8	39.9	19.2	18.2	22.3	37.7
FedDG-ReID	DACS	AAAI'24	38.3	63.7	17.3	39.9	29.1	28.8	28.2	44.1
	<b>FedARKS</b>	-	<b>43.6</b>	<b>68.4</b>	<b>18.2</b>	<b>40.6</b>	<b>33.0</b>	<b>33.3</b>	<b>31.6</b>	<b>47.4</b>

Table 1: Comparison of methods with different backbones and target unseen domains. M:Market1501, C2:CUHK02, C3:CUHK03, MS:MSMT17,R1: Rank-1. SAvg columns report the average mAP and average R1 across the three transfer tasks,highlighting model generalization ability.

Method	Market1501		CUHK02		CUHK03	
	mAP	R1	mAP	R1	mAP	R1
FedProx	61.0	80.4	66.8	65.5	24.2	23.9
FedPav	53.9	76.0	59.7	56.3	19.6	19.6
FedReID	71.8	87.6	82.9	82.8	44.0	44.9
DACS	72.1	88.2	84.5	83.4	47.4	50.1
<b>FedARKS</b>	<b>73.5</b>	<b>89.4</b>	<b>86.8</b>	<b>86.5</b>	<b>54.5</b>	<b>56.8</b>

Table 2: Source domain performance comparison. Training domains: Market1501 + CUHK02 + CUHK03; Test domains: Market1501, CUHK02, and CUHK03.

learned global model captures ensembles of locally optimal, scene-adaptive features critical for robust cross-domain ReID. Additionally, Fig.4 vividly shows the dynamic evolution of aggregated weights for each client over 40 federated training epochs. The curves indicate every client's contribution is continuously adjusted based on its local updates and, as training converges, gradually stabilizes. This means the KS mechanism robustly integrates clients with different key detail learning capabilities, even with severely non-IID data. Dynamic weighting effectively prevents low-quality or mismatched client updates from causing negative transfer and helps the global model better absorb valuable cross-domain invariant knowledge. Overall, these visualization results clearly demonstrate our RK and KS mechanisms work synergistically for reliable feature learning in challenging joint ReID tasks.

Attributes		MS+C2		C3+C2		MS+C2	
		+C3→M		+M→MS		+M→C3	
RK	KS	mAP	R1	mAP	R1	mAP	R1
×	×	35.5	60.3	9.0	26.7	24.6	24.4
✓	×	37.3	63.3	9.6	28.2	28.5	28.4
×	✓	37.7	62.6	9.7	27.6	30.2	31.6
✓	✓	<b>38.6</b>	<b>63.8</b>	<b>10.5</b>	<b>28.8</b>	<b>31.5</b>	<b>34.7</b>

Table 3: Ablation study on Robust Knowledge(RK) and Knowledge Selection(KS).

## Conclusion

This paper proposes a novel federated aggregation via robust and discriminative knowledge selection and integration (FedARKS) to address the inherent challenges of FedDG-ReID. Specifically, our RK architecture combines local body part feature learning with global features, enabling client models to learn key local details and richer context. KS dynamically adjusts aggregation weights by evaluating directional consistency between client updates and global optimization trajectories, ensuring robust semantic consistency and fine-grained alignment in the federated feature space. Extensive experiments show that our method demonstrates significant robustness on popular backbones like ResNet50 and ViT, further validating its scalability and effectiveness for privacy-preserving person ReID in distributed learning scenarios with heterogeneous clients.

## Acknowledgments

This work was supported by the Natural Science Foundation of China (62376201), Hubei Provincial Science & Technology Talent Enterprise Services Program (2025DJB059), and Hubei Provincial Special Fund for Central-Guided Local S&T Development (2025CSA017).

## References

- Dai, Y.; Li, X.; Liu, J.; Tong, Z.; and Duan, L.-Y. 2021. Generalizable Person Re-Identification With Relevance-Aware Mixture of Experts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16145–16154.
- Du, Y.; Xu, J.; Xiong, H.; Qiu, Q.; Zhen, X.; Snoek, C. G. M.; and Shao, L. 2020. Learning to Learn with Variational Information Bottleneck for Domain Generalization. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J.-M., eds., *Computer Vision – ECCV 2020*, 200–216. Cham: Springer International Publishing. ISBN 978-3-030-58607-2.
- Gao, S.; Wang, J.; Lu, H.; and Liu, Z. 2020. Pose-Guided Visible Part Matching for Occluded Person ReID. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Huang, G.; Liu, Z.; Pleiss, G.; Maaten, L. v. d.; and Weinberger, K. Q. 2022. Convolutional Networks with Dense Connectivity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 8704–8716.
- Huang, W.; Ye, M.; and Du, B. 2022. Learn from others and be yourself in heterogeneous federated learning. In *CVPR*.
- Huang, W.; Ye, M.; Shi, Z.; and Du, B. 2023a. Generalizable Heterogeneous Federated Cross-Correlation and Instance Similarity Learning. *TPAMI*.
- Huang, W.; Ye, M.; Shi, Z.; Li, H.; and Du, B. 2023b. Rethinking Federated Learning with Domain Shift: A Prototype View. In *CVPR*.
- Huang, W.; Ye, M.; Shi, Z.; Wan, G.; Li, H.; Du, B.; and Yang, Q. 2024. A Federated Learning for Generalization, Robustness, Fairness: A Survey and Benchmark. *TPAMI*.
- Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; D’Oliveira, R. G. L.; Eichner, H.; Rouayheb, S. E.; Evans, D.; Gardner, J.; Garrett, Z.; Gascón, A.; Ghazi, B.; Gibbons, P. B.; Gruteser, M.; Harchaoui, Z.; He, C.; He, L.; Huo, Z.; Hutchinson, B.; Hsu, J.; Jaggi, M.; Javidi, T.; Joshi, G.; Khodak, M.; Konečný, J.; Korolova, A.; Koushanfar, F.; Koyejo, S.; Lepoint, T.; Liu, Y.; Mittal, P.; Mohri, M.; Nock, R.; Özgür, A.; Pagh, R.; Qi, H.; Ramage, D.; Raskar, R.; Raykova, M.; Song, D.; Song, W.; Stich, S. U.; Sun, Z.; Suresh, A. T.; Tramèr, F.; Vepakomma, P.; Wang, J.; Xiong, L.; Xu, Z.; Yang, Q.; Yu, F. X.; Yu, H.; and Zhao, S. 2021. Advances and Open Problems in Federated Learning. *Foundations and Trends® in Machine Learning*, 14(1–2): 1–210.
- Kalayeh, M. M.; Basaran, E.; Gökmen, M.; Kamasak, M. E.; and Shah, M. 2018. Human Semantic Parsing for Person Re-Identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; and Suresh, A. T. 2020. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In III, H. D.; and Singh, A., eds., *Proceedings of the 37th International Conference on Machine Learning Research*, 5132–5143. PMLR.
- Konečný, J.; McMahan, H. B.; Yu, F. X.; Richtárik, P.; Suresh, A. T.; and Bacon, D. 2016. Federated Learning: Strategies for Improving Communication Efficiency. *arXiv preprint arXiv:1610.05492*. Original arXiv submission (v1: 2016-10-18).
- Kreiss, S.; Bertoni, L.; and Alahi, A. 2019. PifPaf: Composite Fields for Human Pose Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. 2018a. Learning to Generalize: Meta-Learning for Domain Generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Li, D.; Zhang, J.; Yang, Y.; Liu, C.; Song, Y.-Z.; and Hospedales, T. M. 2019. Episodic Training for Domain Generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Li, H.; Pan, S. J.; Wang, S.; and Kot, A. C. 2018b. Domain Generalization With Adversarial Feature Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Li, Q.; He, B.; and Song, D. 2021. Model-Contrastive Federated Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10713–10722.
- Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated Optimization in Heterogeneous Networks. In Dhillon, I.; Papailiopoulos, D.; and Sze, V., eds., *Proceedings of Machine Learning and Systems*, volume 2, 429–450.
- Li, W.; and Wang, X. 2013. Locally aligned feature transforms across views. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3594–3601.
- Li, W.; Zhao, R.; Xiao, T.; and Wang, X. 2014. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 152–159.
- Li, Y.; Gong, M.; Tian, X.; Liu, T.; and Tao, D. 2018c. Domain Generalization via Conditional Invariant Representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Li, Y.; Tian, X.; Gong, M.; Liu, Y.; Liu, T.; Zhang, K.; and Tao, D. 2018d. Deep Domain Generalization via Conditional Invariant Adversarial Networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

- Liu, W.; Xu, X.; Chang, H.; Yuan, X.; and Wang, Z. 2025. Mix-modality person re-identification: A new and practical paradigm. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(4): 1–21.
- Ma, Y.; Dai, W.; Huang, W.; and Chen, J. 2025. Geometric Knowledge-Guided Localized Global Distribution Alignment for Federated Learning. In *CVPR*.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and Arcas, B. A. y. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Singh, A.; and Zhu, J., eds., *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, 1273–1282. PMLR.
- Miao, J.; Wu, Y.; Liu, P.; Ding, Y.; and Yang, Y. 2019. Pose-Guided Feature Alignment for Occluded Person Re-Identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Mohassel, P.; and Zhang, Y. 2017. SecureML: A System for Scalable Privacy-Preserving Machine Learning. In *2017 IEEE Symposium on Security and Privacy (SP)*, 19–38.
- Panareda Busto, P.; and Gall, J. 2017. Open Set Domain Adaptation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
- Song, C.; Huang, Y.; Ouyang, W.; and Wang, L. 2018. Mask-Guided Contrastive Attention Model for Person Re-Identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Song, J.; Yang, Y.; Song, Y.-Z.; Xiang, T.; and Hospedales, T. M. 2019. Generalizable Person Re-Identification by Domain-Invariant Mapping Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; and Wang, S. 2018. Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Wang, G.; Yuan, Y.; Chen, X.; Li, J.; and Zhou, X. 2018. Learning Discriminative Features with Multiple Granularities for Person Re-Identification. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM '18, 274–282. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356657.
- Wang, H.; Chen, X.; and Liu, C. 2021. Pose-guided part matching network via shrinking and reweighting for occluded person re-identification. *Image and Vision Computing*, 111: 104186.
- Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 79–88.
- Wu, G.; and Gong, S. 2021. Decentralised Learning from Independent Multi-Domain Labels for Person Re-Identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2898–2906. Virtual Event, USA.
- Xu, X.; Liu, W.; Wang, Z.; Hu, R.; and Tian, Q. 2022. Towards generalizable person re-identification with a bi-stream generative model. *Pattern Recognition*, 132: 108954.
- Xu, Y.; Zhao, L.; and Qin, F. 2021. Dual attention-based method for occluded person re-identification. *Knowledge-Based Systems*, 212: 106554.
- Yang, F.; Zhong, Z.; Luo, Z.; He, Y.; Li, S.; and Sebe, N. 2024. Diversity-Authenticity Co-constrained Stylization for Federated Domain Generalization in Person Re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI-24, 6477–6485. Vancouver, Canada: Association for the Advancement of Artificial Intelligence.
- Yang, J.; Zhang, C.; Tang, Y.; et al. 2022. PAFM: Pose-Driven Attention Fusion Mechanism for Occluded Person Re-Identification. *Neural Computing and Applications*, 34: 8241–8252.
- Yang, Q.; Wang, P.; Fang, Z.; and Lu, Q. 2020. Focus on the Visible Regions: Semantic-Guided Alignment Model for Occluded Person Re-Identification. *Sensors*, 20(16).
- Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Shao, L.; and Hoi, S. C. H. 2022. Deep Learning for Person Re-Identification: A Survey and Outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6): 2872–2893.
- Zhang, X.; Yan, Y.; Xue, J.-H.; Hua, Y.; and Wang, H. 2021. Semantic-Aware Occlusion-Robust Network for Occluded Person Re-Identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(7): 2764–2778.
- Zhang, Y.; Li, M.; Li, R.; Jia, K.; and Zhang, L. 2022. Exact Feature Distribution Matching for Arbitrary Style Transfer and Domain Generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8035–8045.
- Zhao, Y.; Zhong, Z.; Yang, F.; Luo, Z.; Lin, Y.; Li, S.; and Sebe, N. 2021. Learning to Generalize Unseen Domains via Memory-based Multi-Source Meta-Learning for Person Re-Identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6277–6286.
- Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, 1116–1124.
- Zhou, K.; Yang, Y.; Hospedales, T.; and Xiang, T. 2020. Learning to Generate Novel Domains for Domain Generalization. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J.-M., eds., *Computer Vision – ECCV 2020*, 561–578. Cham: Springer International Publishing. ISBN 978-3-030-58517-4.
- Zhuang, W.; Wen, Y.; Zhang, X.; Gan, X.; Yin, D.; Zhou, D.; Zhang, S.; and Yi, S. 2020. Performance Optimization of Federated Person Re-identification via Benchmark Analysis. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, 955–963. New York, NY, USA: Association for Computing Machinery. ISBN 9781450379885.