UPL-ReID: Uncertainty-Calibrated Pseudo-Labeling and Meta-Optimization for Unsupervised Person Re-Identification

Abstract

Unsupervised domain adaptation (UDA) for person re-identification (ReID) remains highly challenging due to significant domain shifts and the noise associated with pseudo-label supervision in unlabeled target domains. Existing clustering-based methodologies frequently encounter issues with error accumulation due to the unreliability of pseudo-labels. This complicates the differentiation of traits and impedes generalization. This research presents UPL-ReID, a novel uncertainty-aware pseudo-labeling framework that distinctly defines and mitigates pseudo-label uncertainty for robust cross-domain person re-identification. We employ a lightweight Vision Transformer backbone and a hybrid pseudo-label generation technique that incorporates camera-specific DBSCAN clustering and Gaussian Mixture Models (GMM) to produce soft, uncertainty-aware labels. We use an entropy-guided uncertainty weighting mechanism and an uncertainty-calibrated feature-space mixup technique to diminish noisy supervision. These algorithms prioritize dependable samples while enhancing feature diversity and calibration. We additionally construct a MoCo-style contrastive memory bank with supplementary metadata to ensure consistency across all cameras and to facilitate the differentiation of objects. We introduce a Self-Reflective Meta-Optimizer (SRMO) that adaptively adjusts loss weights utilizing a reserved source-domain meta-set. This enhances the model's resilience to pseudo-label noise and unfamiliar identities by adaptively balancing multiple learning objectives. Extensive evaluations on the Market-1501 and DukeMTMC-ReID benchmarks demonstrate that UPL-ReID consistently outperforms the leading UDA ReID algorithms, achieving 94.2% mAP on Duke→Market and 85.2% mAP on Market→Duke. Comprehensive ablation studies validate the effectiveness of uncertainty modeling, contrastive memory learning, and meta-optimization in improving clustering quality, calibration, and cross-domain generalization.

1 Introduction

Person re-identification is an essential task in intelligent surveillance systems in order to re-identify a person or to find particular pedestrians using a network of cameras that do not overlap with each other. [1]. As a prominent research problem in computer vision, cross-domain re-identification aims to precisely match pedestrians across heterogeneous surveillance devices and environments. Re-identification (ReID) has various applications in intelligent surveillance systems, such as tracking individuals in public spaces and locating specific objects [2]. In terms of public safety, it makes it easier to find suspects and missing people, which speeds up the resolution of cases. In addition, ReID helps with traffic management, crowd movement analysis, and general urban safety in smart city designs [3]. However, divergent perspectives, varying image resolutions, fluctuations in lighting, variations in posture, occlusions, and inter-camera heterogeneity pose significant challenges to ReID [4][5]. Supervised ReID methods try to learn how to recognize the difference between different features by using labeled pedestrian identities [6]. Usually, these algorithms use multi-camera datasets that have person IDs and attributes associated with them. This allows them to obtain robust features that don't change with changes in position, occlusion, or lighting. Deep

metric learning, local feature alignment, and re-ranking algorithms are some of the main methods. Supervised approaches perform exceptionally well, but they rely significantly on time-consuming manual labeling and often do not work well in new areas because of problems with camera alignment and discrepancies between cameras [7][8]. This constraint drives the investigation of unsupervised ReID methodologies, which do not depend on annotated data but rather examine image attributes for identification correspondence. These techniques lower the cost of labeling and make it easy to adapt to new scenarios quickly [9].

There are two primary types of current unsupervised domain adaptation (UDA) approaches for ReID: GAN-based and pseudo-labeling-based techniques and UDA methods. GAN-based methods try to reduce differences between domains by changing the styles of images, which makes the source images more like the target domain for supervised learning [10]. However, computer- generated images may exhibit artifacts or noise resulting from fluctuations in lighting, resolution, and perspective, thereby hindering the total elimination of domain gaps and impacting model robustness [11]. On the other hand, UDA approaches that use pseudo-labeling focus on finding differences between classes in the target domain. Early single-branch frameworks utilize features pretrained on the source domain to predict pseudo-labels by clustering, then using these labels to guide and steer the model training process [12]. Pseudo-labels are inherently noisy and may not match up with real identities, but iterative refinement methods like feature extraction, clustering, and accurate sample selection can help reduce this noise. Inspired by this, we suggest an optimization technique that takes uncertainty into account to make labels more reliable, lower the variation within classes, and improve the separation between classes. This process makes clustering more compact and distinct.

Contrastive learning, a self-supervised paradigm, has demonstrated significant promise in visual recognition tasks. By attracting similar sample pairs (from the same class) and repelling dissimilar pairs (from different classes), contrastive learning enables models to distinguish fine-grained features without requiring explicit labels [13]. It has been particularly effective for unsupervised learning scenarios [12]. Pair formation is typically achieved through data augmentation, with notable methods including SimCLR [14], MoCo [15] and SwAV [16]. In the context of ReID, contrastive learning pairs multiple views of the same pedestrian while separating them from views of other identities, thereby improving the model's capacity to learn discriminative representations [10]. In this study, we propose UPL-ReID, an uncertainty-aware pseudo-labeling framework for unsupervised domain adaptation for person re-identification. Our approach integrates a lightweight Vision Transformer (ViT) backbone for feature extraction, hybrid clustering that combines camera-specific DBSCAN and Gaussian Mixture Models (GMM) for pseudo-label generation, and uncertainty modeling via entropy. Additionally, adaptive feature-space mixup, guided by expected calibration error, emphasizes challenging samples during training. A MoCo-style contrastive learning mechanism with a pseudo-label-guided memory bank further enhances feature discrimination. Finally, we employ a self-reflective meta-optimization strategy to jointly optimize cross-entropy, triplet, and InfoNCE losses. By Using a held-out subset of the source domain as a meta-set, this strategy reduces validation loss and improves generalization, using the Market-1501 and DukeMTMC-ReID datasets. The key contributions of this work are summarized as follows:

- We propose a novel mixup mechanism that dynamically adjusts per-sample mixing ratios based on the posterior entropy of Gaussian Mixture Model (GMM) clustering. This approach prioritizes challenging samples and effectively reduces intra-class variance while increasing inter-class separation, leading to more compact and discriminative feature representations.

- We introduce a dynamic meta-optimization framework that adaptively adjusts the weights of classification, triplet, and contrastive losses. By optimizing these loss components simultaneously, the model achieves enhanced performance and generalization across multiple source and target domains.

- We design a contrastive learning mechanism utilizing a momentum-updated key encoder and a pseudo-label-guided memory bank. This framework effectively mitigates the impact of noisy pseudo-labels, strengthens discriminative feature learning, and facilitates robust cross-domain adaptation.

- We used camera-specific DBSCAN combined with GMM to generate pseudo-labels for target domain data. This hybrid approach addresses domain-specific variations in viewpoint, illumination, and occlusion.

2 Literature Review

Unsupervised domain adaptation for person re-identification (UDA-ReID) aims to transfer unique identity representations from a labeled source domain to an unlabeled target domain, while addressing significant domain alterations in camera angles, lighting, background noise, and occlusion. Recent advancements highlight the improvement of pseudo-label reliability, the enhancement of representation learning via contrastive techniques, and the incorporation of uncertainty-awareness and adaptive optimization to mitigate noisy supervision.

2.1 Clustering-based pseudo-labeling

Clustering-based pseudo-label generation is the dominant strategy in unsupervised domain adaptive person re-identification. Early works demonstrate that density-based clustering can provide reliable supervision for target domain adaptation; however, noisy pseudo labels remain a major bottleneck. To alleviate this issue, recent studies focus on improving clustering robustness and label reliability.

Shao et al. [17] focus on the one-shot person re-identification problem, where only a single labeled sample is available for each pedestrian. They introduce a hierarchical pseudo-labeling strategy based on density and distance, together with an embranchment learning framework. This strategy effectively exploits unlabeled data by hierarchically generating reliable pseudo labels through multiple clustering steps using pairwise feature distances and data distribution densities. Their approach significantly outperforms existing methods, achieving gains of 36.1% and 28.7% mAP on Market-1501 and DukeMTMC-reID. Chen et al. [5] propose a lightweight, plug-and-play approach to enhance distance computation during pseudo-label generation. Their method introduces inter-camera k-reciprocal nearest neighbors to identify reliable inter-camera positive pairs and adaptive inter-camera encouragement to integrate these relationships into distance measurements, leading to more effective clustering.

Samanta et al. [18] propose a collaborative learning scheme to improve cross-domain person re-identification by reducing label noise through refined pseudo-labels. Their approach combines outputs from multiple domain-adapted teacher networks of different types and sizes using dynamic averaging to obtain more reliable supervision. This strategy leads to clear performance gains, achieving 81.04% and 76.24% mAP for DukeMTMC-reID to Market-1501 and Market-1501 to DukeMTMC-reID, respectively. Yu et al. [19] introduce the Pose-Aligned Outlier Sample Relabeling (POSR) model to improve cross-camera person re-identification by relabeling outlier samples and mining diverse data representations. POSR combines a Pose-Aligned Feature Learning (PFL) module, which captures fine-grained human joint features via token matching and average pooling, with a Biclustering Outlier Relabeling (BOR) module that refines outlier labels using both pose-aligned and global features. This approach enables learning of ID-consistent representations and enhances cross-camera retrieval. POSR achieves 78.1% mAP on DukeMTMC-reID and 89.8%

on Market-1501. Samanta et al. [20] present the Unsupervised Dual-Teacher Knowledge Distillation (UDKD) framework to improve robustness against noisy pseudo-labels in UDA for person re-identification. UDKD trains a student network by combining outputs from two source-trained teacher classifiers using a soft-triplet loss, while weighted averaging mitigates label noise. The method achieves strong performance, with mAP scores of 84.57% and 73.32%, and Rank-1 scores of 94.34% and 88.26% for DukeMTMC-reID-to-Market-1501 and Market-1501-to-DukeMTMC-reID scenarios, respectively.

2.2 Uncertainty-based Methods

Han et al. [21] proposed the Probabilistic Uncertainty Guided Progressive Label Refinery (P2LR) for domain-adaptive person re-identification. The method models labeling uncertainty using probabilistic distances and single-peak distributions, providing a quantitative measure to guide network training. It further employs a progressive pseudo-label refinement strategy with uncertainty-guided optimization, balancing target domain exploration and noisy label mitigation. P2LR improves the baseline by 6.5% mAP on Duke-to-Market and outperforms the state-of-the-art by 2.5% mAP on Market-to-MSMT. Zhang et al. [22] introduce Style-Uncertainty Augmentation (SuA), a simple yet effective feature-level augmentation technique that perturbs instance styles with Gaussian noise to increase domain diversity during training. To enhance generalization across these augmented domains, they propose Self-paced Meta Learning (SpML), a multi-stage progressive learning strategy that gradually improves model performance on unseen target domains. Moreover, a distance-graph alignment loss is introduced to align feature relationships across domains, enabling the network to learn domain-invariant representations. Liu et al. [23] propose Dual Uncertainty Guided Curriculum Learning (DUCL) to address label noise in domain-adaptive person re-identification. The method uses reliability-based curriculum allocation to adapt samples from easy to hard, complemented by a dual-uncertainty re-weighting strategy to reduce noisy label impact. Additionally, Part-Aware Feature Refinement (PAFR) leverages part-aware attention maps to integrate fine-grained semantics into holistic representations, improving pseudo-label reliability. DUCL achieves 84% mAP for DukeMTMC-reID-to-Market-1501 and 72.7% for Market-1501-to-DukeMTMC-reID. Zhao et al. [24] propose a method for TI-ReID that integrates multi-modal uncertainty modeling with semantic alignment. They represent image and text features as Gaussian distributions, estimating multi-granularity uncertainty using batch- and identity-level variances to enrich image-text relationships. A bi-directional cross-modal circle loss aligns these probabilistic features in a self-paced manner, while a complementary masked language modeling task enhances global image-text semantic recovery after cross-modal interaction.

2.3 Contrastive learning-based methods

Si et al. [25] propose the Hybrid Contrastive Model (HCM) for unsupervised person Re-ID, combining identity-level and image-level contrastive learning to better capture feature similarities among hard samples. Identity-level learning uses a memory bank and dynamic contrast loss to distinguish hard and easy samples, while image-level learning employs a separate memory with a sample constraint loss to explore relationships between hard positives and negatives. Both processes are optimized within a unified framework to enhance feature representation. HCM achieves 79% mAP on Market-1501 and 67.9% on DukeMTMC-reID. Tian et al. [26] propose Proxy Alignment Contrastive Learning (PACL) to improve unsupervised person Re-ID. PACL features Enhanced Proxy Mapping Contrastive Learning (EPMCL), assigning each feature to a proxy based on cluster and camera information, and introduces Proxy Similarity Loss (PSL) to pull proxies of the same cluster together while separating different clusters. By using multiple enhanced views of each image,

the model maximizes similarity with assigned proxies and minimizes it with others, enabling discriminative and robust embeddings. PACL achieves 83.5% mAP on Market-1501 and 71.8% on DukeMTMC-reID.

Wang et al. [27] propose a UDA person Re-ID approach based on joint pre-training. The model is first pre-trained on both source and target domain data using class- and cluster-level contrastive learning to learn generalizable and discriminative features, improving pseudo-label accuracy during fine-tuning. A combined offline and online knowledge distillation strategy is then applied to reduce feature bias and enhance performance. The method achieves 86.5% mAP for DukeMTMC-reID to Market-1501 and 75.2% for Market-1501 to DukeMTMC-reID. Bai et al. [13] propose a contrastive learning-based pseudo-label refinement method with probabilistic uncertainty for unsupervised domain-adaptive person Re-ID. The approach first enhances target domain feature representations via contrastive learning to improve discrimination and cross-domain transfer. It then introduces a novel loss function to reduce the impact of noisy pseudo-labels during training. Experiments on Market-1501 and DukeMTMC-reID show strong results, achieving Rank-1 accuracies of 91.4% and 81.4%, and mAP of 79.0% and 67.9%, respectively. Zhang et al. [28] propose Heterogeneous Pseudo Labels (HPL) for cross-domain person ReID, addressing large intra-class and small inter-class variations. HPL generates fine-grained, coarse-grained, and instance-level pseudo labels, enhanced by Pseudo Labels Constraint (PLC) for consistency and Confidence Contrastive Loss (CCL) to reduce noisy label effects. The method achieves 87.2% mAP and 95.0% Rank-1 accuracy on MSMT17 to Market.

3 Methods

3.1 Problem Formulation

We address the task of unsupervised domain adaptation (UDA) for person re-identification (ReID), where the objective is to transfer discriminative knowledge from a labeled source domain to an unlabeled target domain under substantial distributional shifts, such as changes in illumination, pose variation, and background clutter. Formally, the source domain is denoted as given in equation 1.

$$\mathcal{D}_s = \{(x_i^s, y_i^s, c_i^s)\}_{i=1}^{N_s} \tag{1}$$

Where x_i^s represents an image, $y_i^s \in \{1, \ldots, K_s\}$ denotes its person identity (ID), and $c_i^s \in \{1, \ldots, C_s\}$ is the associated camera index. The target domain is defined by equation 2.

$$\mathcal{D}_t = \{(x_j^t, c_j^t)\}_{j=1}^{N_t} \tag{2}$$

Where x_j^t is an unlabeled image with camera label c_j^t, and the identity annotations are not available. To enable meta-learning without violating UDA constraints, we partition the labeled source domain $\mathcal{D}_s$ into two disjoint sets: a training set $\mathcal{D}_s^{train}$ and a meta-validation set $\mathcal{D}_s^{meta}$, such that $\mathcal{D}_s = \mathcal{D}_s^{train} \cup \mathcal{D}_s^{meta}$ and $\mathcal{D}_s^{train} \cap \mathcal{D}_s^{meta} = \emptyset$. The set $\mathcal{D}_s^{meta}$ provides auxiliary guidance for optimization, simulating generalization to unseen data. The goal is to learn a feature embedding function as given by equation 3.

$$f_\theta : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^d \tag{3}$$

The function f parameterized by θ, which maps an input image to a d-dimensional feature space, such that two essential properties are satisfied. First, the learned representation must ensure intra-class compactness and inter-class separability in the embedding space. This can be expressed by equation 4.

$$\|f_\theta(x_a) - f_\theta(x_p)\|_2^2 \ll \|f_\theta(x_a) - f_\theta(x_n)\|_2^2 \tag{4}$$

Where x_a, x_p, and x_n correspond to anchor, positive, and negative samples, respectively. Second, the representation should be domain-invariant, such that the learned features generalize well across $\mathcal{D}_s$ and $\mathcal{D}_t$ despite the absence of identity labels in the target domain. The target domain is not annotated by identity, however, we exploit an held-out meta-validation. sub sample of the source domain to optimize. This is due to the observation that inspired this design. that the strength of resistance to pseudo-label is highly correlated with strength to unseen source identities. noise in the target domain. Loss weights are made to optimize the validation error on model, D_s^{meta} is motivated to learn representations that are extrapolated to the identities of training and are less affected by noise of labels. This meta-learning strategy is a surrogate regularizer of target-domain generalization, and does not violate the unstimulated environment of adaptation. To achieve this objective, our proposed UPL-ReID framework incorporates three core mechanisms: (i) a hybrid clustering strategy combining camera-specific DBSCAN with Gaussian Mixture Models (GMM) to generate reliable pseudo-labels in the target domain, (ii) an uncertainty-aware weighting scheme to mitigate the influence of noisy pseudo-labels through entropy-based confidence estimation, and (iii) a self-reflective meta-optimization procedure that uses $\mathcal{D}_s^{meta}$ to jointly refine loss functions and improve cross-domain generalization. The process of the proposed framework is illustrated in Figure 1.

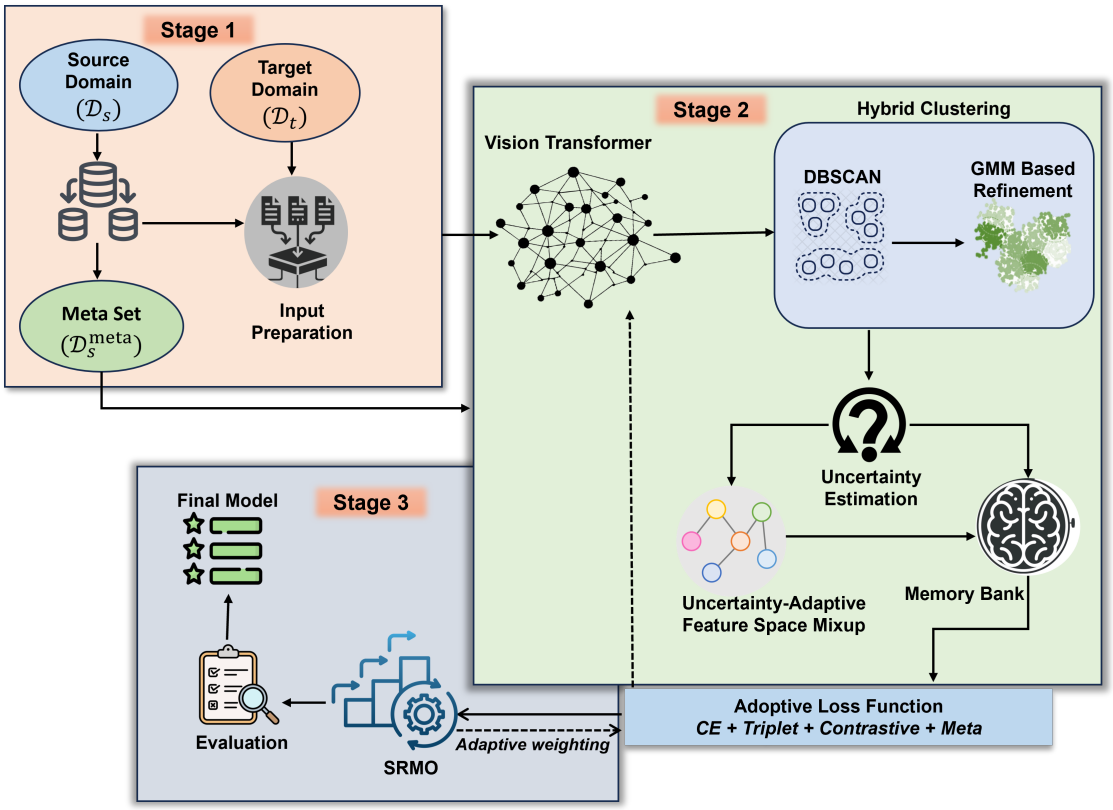

Figure 1: Flow of proposed the framework

3.2 Dataset Preprocessing and Input Pipeline

We used two widely used person ReID datasets, namely Market-1501 and DukeMTMC-ReID. We also create meta dataset by randomly splitting the source domain into a training split (80%)

and a meta-validation split (20%). The meta-validation split is strictly excluded from the standard gradient updates and is used solely for the meta-update step of the SRMO. Each dataset consists of person images captured across multiple non-overlapping cameras, thereby introducing significant challenges such as viewpoint variation, occlusion, illumination inconsistency, and background clutter. To ensure consistency across domains, all images are preprocessed before training. Specifically, each input image is resized to 256×128 pixels while preserving aspect ratio, followed by per-channel normalization using ImageNet statistics. Standard data augmentation techniques are employed to improve generalization, including random horizontal flipping, random cropping, color jittering, and random erasing. For efficient representation learning, we adopt a mini-batch sampling strategy that balances identity and camera diversity. The processed mini-batches are subsequently forwarded to the feature extraction backbone, providing the foundation for domain-adaptive representation learning. Table 1 presents the description of the datasets used in the study.

Table 1: Statistics of datasets used in the proposed UPL-ReID framework

Dataset	**#IDs**	**#Images**	**#Cameras**	**Train**	**Query**	**Gallery**
Market-1501	751	32,668	6	12,936	3,368	19,732
DukeMTMC-ReID	1,404	36,411	8	16,522	2,228	17,661

3.3 Feature Extraction Backbone

A crucial component of the proposed framework is the feature extraction backbone, which is responsible for mapping raw pedestrian images into a discriminative and domain-invariant embedding space. For this purpose, we adopt a hybrid architecture that combines convolutional neural networks (CNNs) with Transformer-based modules, using both local spatial filtering and global contextual reasoning. The backbone $\mathcal{F}_\theta$ is constructed upon the Vision Transformer (ViT) paradigm, augmented with a lightweight convolutional stem to preserve fine-grained texture cues that are critical for person re-identification. Formally, given an input image $x_i \in \mathbb{R}^{H\times W\times 3}$, the stem applies a stack of convolutional layers to produce a low-level representation $z_i \in \mathbb{R}^{h\times w\times c}$, which is subsequently flattened into patch tokens. These tokens are processed by a sequence of L Transformer encoder blocks, each comprising a multi-head self-attention (MHSA) layer and a feed-forward network (FFN)as presented in equation 5.

$$\begin{aligned} \mathrm{MHSA}(Q,K,V) &= \mathrm{Concat}(\mathrm{head}_1,\ldots,\mathrm{head}_M)W^O, \\ \mathrm{head}_m &= \mathrm{Softmax}\left(\frac{QW_m^Q(KW_m^K)^\top}{\sqrt{d_k}}\right)VW_m^V \end{aligned} \tag{5}$$

Where M denotes the number of attention heads, and d_k is the key/query dimensionality. Layer normalization and residual connections are applied after each sub-layer to stabilize training. The output of the Transformer encoder is a sequence of contextualized patch tokens $\{t_j\}_{j=1}^{P}$, where P is the number of patches. We apply a global average pooling followed by a fully connected projection layer to obtain the final embedding vector. This can be mathematically presented by equation 6.

$$v_i = \mathrm{Proj}\left(\frac{1}{P}\sum_{j=1}^{P} t_j\right) \in \mathbb{R}^d, \tag{6}$$

Where d denotes the embedding dimensionality. For scale-invariant similarity computation, embeddings are ℓ_2-normalized as given in equation 7.

$$\tilde{v}_i = \frac{v_i}{\|v_i\|_2}. \tag{7}$$

To enhance the performance of the backbone architecture for cross-domain person re-identification, we introduce several architectural and training modifications to improve feature extraction and generalization across diverse datasets. Input images are resized to 256×128 pixels to preserve fine-grained visual details, such as clothing textures and accessories, which are critical for accurate identification. The convolutional stem is modified to use a stride of 1 in the final stage, maintaining higher spatial resolution in feature maps to capture detailed spatial information. Moreover, the projection head is configured with an embedding dimension of $d = 768$, aligning with standard practices in vision transformer-based re-identification frameworks to ensure robust feature representation. A batch normalization neck is incorporated before the final embedding layer to stabilize feature distributions, enhancing the robustness and convergence of the model during training for person re-identification tasks.

The backbone is initialized with weights pre-trained on the ImageNet-1K dataset. This initialization provides general semantic priors while mitigating overfitting in the limited target domain. During adaptation, all layers are fine-tuned, but a smaller learning rate is applied to the early convolutional layers to preserve generic low-level representations. The optimization follows a cosine annealing schedule with warm restarts. For a mini-batch of N samples $\{x_i\}_{i=1}^{N}$, the backbone produces an embedding matrix. This can be expressed by equation 8.

$$V = [\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_N]^\top \in \mathbb{R}^{N \times d}, \tag{8}$$

which serves as input to the clustering and pseudo-labeling module. The high-dimensional and semantically enriched embeddings extracted by $\mathcal{F}_\theta$ form the foundation for subsequent uncertainty-aware learning and memory bank consistency enforcement. The overview of the architecture of the proposed vision transformer is illustrated in Figure 2.

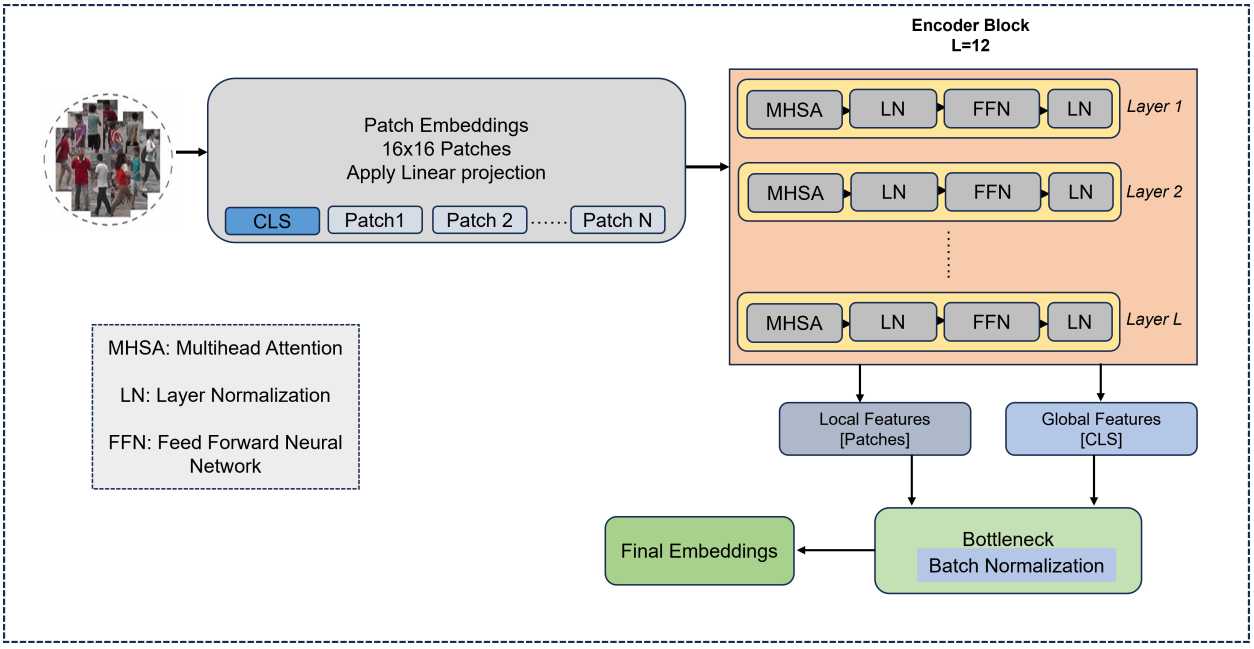

Figure 2: Architecture of proposed transformer model

3.4 Clustering and Pseudo-Labeling

In unsupervised domain adaptation for person re-identification, where the target domain lacks labeled identities, we employ a clustering-driven pseudo-labeling framework to enable supervised

loss computation on unlabeled data. To address intra-camera appearance biases, the target dataset is partitioned by camera ID, and DBSCAN clustering [29] is applied to the feature embeddings $\mathcal{F} = \{\mathbf{f}_i \in \mathbb{R}^d\}_{i=1}^N$ extracted from the backbone network. DBSCAN identifies clusters based on density connectivity, parameterized by a neighborhood radius ϵ and minimum points per cluster m, yielding M clusters $\mathcal{C} = \{C_1, C_2, \ldots, C_M\}$, each representing a pseudo-identity. Unclustered points are treated as noise and excluded. To refine cluster assignments and capture intra-cluster variability, a GMM with K components is fitted to the cluster centroids, producing soft pseudo-labels $\hat{\mathbf{y}}_i \in \Delta^{K-1}$ for each sample, as defined in equation 9.

$$p(y = k \mid \mathbf{f}_i) = \frac{\pi_k \mathcal{N}(\mathbf{f}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{f}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}, \tag{9}$$

Where π_k, $\boldsymbol{\mu}_k \in \mathbb{R}^d$, and $\boldsymbol{\Sigma}_k \in \mathbb{R}^{d \times d}$ denote the mixture weight, mean, and covariance of the k-th Gaussian, respectively, and Δ^{K-1} is the $(K-1)$-dimensional probability simplex. We initially do camera-specific DBSCAN to eliminate cross-camera density bias and conserve consistent local structures point of view. The cluster centroid that results are then merged between cameras and used to initialize a global Gaussian Mixture Model. Rather than replacing the GMM also does soft refinement on the DBSCAN-induced structure, DBSCAN assignments, where probabilistic labels are allowed on samples near cluster boundaries but they are retained. This two phase method avoid collapse of the clusters and permit identity association across cameras using uncertainty-sensitive soft assignments.

To account for pseudo-label reliability, we compute the Shannon entropy of the posterior distribution, $\mathcal{U}(\hat{\mathbf{y}}_i) = -\sum_{k=1}^K \hat{y}_{i,k} \log \hat{y}_{i,k}$, and define a per-sample weight $w_i = \exp(-\alpha \mathcal{U}(\hat{\mathbf{y}}_i))$, where $\alpha > 0$ is a tunable temperature parameter controlling sensitivity to uncertainty.The weighting coefficient α in $w_i = \exp(-\alpha U(\hat{y}_i))$ controls the sensitivity of the exponential mapping to entropy magnitude. In practice, α is selected through grid search on a held-out validation split or meta-set. We found stable performance for $\alpha \in [1, 3]$; lower values produce nearly uniform weights, while higher values over-suppress uncertain samples. This ensures high-uncertainty samples contribute less to the training gradient, mitigating error propagation from noisy clusters. The pseudo-labels $\{\hat{\mathbf{y}}_i\}$ and weights $\{w_i\}$ are integrated into cross-entropy and triplet losses, with the weighted cross-entropy loss defined as given in 10.

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^N w_i \sum_{k=1}^K \hat{y}_{i,k} \log p_{i,k}, \tag{10}$$

Where $p_{i,k}$ is the predicted probability for pseudo-class k. The triplet loss is similarly weighted by w_i to prioritize confident samples. This pipeline effectively balances clustering fidelity, soft-label refinement, and uncertainty awareness, yielding robust pseudo-labels for unsupervised adaptation.

3.5 Uncertainty-Calibrated Feature-Space Mixup

To enhance the generalization of our person re-identification model under domain shift, we propose an uncertainty-calibrated feature-space mixup strategy that interpolates feature embeddings while mitigating the impact of noisy pseudo-labels. Unlike input-level mixup, we operate in the feature embedding space, using pseudo-label uncertainty weights derived from clustering. For two feature embeddings $\mathbf{f}_i, \mathbf{f}_j \in \mathbb{R}^d$ with soft pseudo-labels $\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j \in \Delta^{K-1}$ and uncertainty weights w_i, w_j, we compute the mixed feature as $\tilde{\mathbf{f}} = \lambda \mathbf{f}_i + (1-\lambda)\mathbf{f}_j$, where $\lambda \sim \text{Beta}(\beta, \beta)$ and $\beta > 0$ controls interpolation strength. incorporating uncertainty to prioritize reliable labels. This normalized formulation ensures that uncertainty-reliable samples dominate the label interpolation. These mixed features and labels are used to compute a weighted cross-entropy loss, $\mathcal{L}_{\text{mix}}^{\text{CE}} = -\frac{1}{N} \sum_{k=1}^K \tilde{y}_k \log p_k(\tilde{\mathbf{f}})$, where

$p_k(\tilde{\mathbf{f}})$ is the predicted probability for pseudo-class k, and a weighted triplet loss. This process can be summarized in equation 11.

$$\tilde{f}_{ij} = \lambda f_i + (1-\lambda) f_j, \qquad \lambda \sim \text{Beta}(\beta, \beta), \tag{11}$$
$$\tilde{y}_{ij} = \frac{\lambda w_i \hat{y}_i + (1-\lambda) w_j \hat{y}_j}{\lambda w_i + (1-\lambda) w_j}$$

The vectors f_i and f_j denote the feature embeddings of two selected target-domain samples, while $\hat{y}_i$ and $\hat{y}_j$ represent their corresponding soft pseudo-labels obtained from the GMM posterior distributions. The terms w_i and w_j are the uncertainty weights derived from entropy, assigning higher importance to samples with more reliable pseudo-labels. The mixing coefficient λ is sampled from a symmetric Beta distribution parameterized by β, which controls the interpolation strength between the two samples. Using these quantities, $\tilde{f}_{ij}$ denotes the mixed feature representation, and $\tilde{y}_{ij}$ is the normalized uncertainty-aware mixed pseudo-label. To measure the quality of a calibration of pseudo-label predictions we use Expected Calibration Error (ECE). This is error between the predicted confidence and empirical accuracy. The Lower ECE implies a higher level of conformity between estimates of confidence.

3.6 Metadata-Enriched MoCo Memory Bank

To maintain feature consistency across pseudo-IDs and enhance the robustness of representation learning under unsupervised domain adaptation, we extend the Momentum Contrast (MoCo) framework by incorporating metadata information associated with each sample. The proposed metadata-enriched memory bank stores feature embeddings along with camera and cluster identifiers, allowing the model to account for intra-domain variations while enforcing inter-sample consistency.

Let $\mathcal{F} = \{\mathbf{f}_i \in \mathbb{R}^d\}_{i=1}^N$ be the set of extracted feature embeddings from the target domain, and $\mathcal{M} = \{(\mathbf{f}_i, c_i, cam_i)\}_{i=1}^N$ denote the memory bank storing tuples of features $\mathbf{f}_i$, cluster pseudo-labels c_i, and camera IDs cam_i. The memory bank is updated in a FIFO queue manner.The memory bank enables computation of a contrastive InfoNCE loss that considers metadata constraints. For an anchor feature $\mathbf{f}_i$, positive samples are selected from the memory bank sharing the same pseudo-label c_i but with different camera views $cam_j \neq cam_i$, while negatives are all other embeddings as presented in equation 12.

$$\mathcal{L}_{infoNCE}(\mathbf{f}_i) = -\log \frac{\sum_{j \in \mathcal{P}_i} \exp(\mathbf{f}_i \cdot \mathbf{f}_j / \tau)}{\sum_{k \in \mathcal{M}} \exp(\mathbf{f}_i \cdot \mathbf{f}_k / \tau)}, \tag{12}$$

Where $\mathcal{P}_i = \{j \mid c_j = c_i, cam_j \neq cam_i\}$ is the set of positive samples for $\mathbf{f}_i$, τ is the temperature hyperparameter, and $\cdot$ denotes the cosine similarity. For each query feature q, positive keys k^+ are retrieved from the memory bank if they share the same pseudo-label but originate from different camera IDs. This constraint promotes cross-camera invariance. When a sample's cluster has no other instances, the InfoNCE term for that sample is skipped to avoid degenerate gradients. The memory bank is maintained with a FIFO queue of size $Q_{\max} = 8192$ to guarantee diverse cross-camera samples. The memory-enhanced InfoNCE loss is combined with the supervised cross-entropy, triplet, and feature-space mixup losses to form the final optimization objective.

3.7 Self-Reflective Meta-Optimizer (SRMO)

Unsupervised domain adaptation (UDA) for person re-identification suffers from domain shifts, where changes in lighting, camera angles, and backgrounds degrade performance on unlabeled target domains. To mitigate this, we propose SRMO that dynamically balances multiple loss

components and adaptively tunes their relative importance during training. Unlike static weighting strategies that cannot adapt to evolving learning dynamics, SRMO employs meta-gradients to update the loss weights and learning rates based on model performance over a held-out meta-set, thereby promoting robust feature generalization across diverse domains. Let $\mathbf{\Theta}$ represent the parameters of the feature extractor backbone, and let $\mathbf{\Lambda} = \lambda_{\text{CE}}, \lambda_{\text{triplet}}, \lambda_{\text{mix}}, \lambda_{\text{infoNCE}}$ denote the set of learnable weights corresponding to the cross-entropy, triplet, mixup-consistency, and InfoNCE loss functions, respectively. The overall training objective on the target dataset $\mathcal{D}_{\text{train}}$ as defined in equation 13.

$$\mathcal{L}_{\text{train}}(\mathbf{\Theta}, \mathbf{\Lambda}) = \lambda_{\text{CE}}\mathcal{L}_{\text{CE}} + \lambda_{\text{triplet}}\mathcal{L}_{\text{triplet}} + \lambda_{\text{mix}}\mathcal{L}_{\text{mix}} + \lambda_{\text{infoNCE}}\mathcal{L}_{\text{infoNCE}}, \tag{13}$$

Where each term captures complementary aspects of representation learning classification accuracy, metric alignment, mixup consistency, and contrastive discrimination. To guide the optimization of $\mathbf{\Lambda}$, a meta-set meta-set $\mathcal{D}_s^{meta}$ to compute a validation loss $\mathcal{L}_{\text{meta}}(\mathbf{\Theta}'(\mathbf{\Lambda}))$, where $\mathbf{\Theta}'$ denotes the backbone parameters after one inner gradient step on $\mathcal{D}_{\text{train}}$ as given in equation 14

$$\mathbf{\Theta}' = \mathbf{\Theta} - \eta\nabla_{\mathbf{\Theta}}\mathcal{L}_{\text{train}}(\mathbf{\Theta}, \mathbf{\Lambda}). \tag{14}$$

The meta-objective then optimizes $\mathbf{\Lambda}$ to minimize the meta-loss by equation 15.

$$\mathbf{\Lambda}^* = \arg\min_{\mathbf{\Lambda}} \mathcal{L}_{\text{meta}}(\mathbf{\Theta}'(\mathbf{\Lambda})). \tag{15}$$

Since full second-order derivatives are computationally expensive, we employ a first-order MAML-style approximation of the meta-gradient as given in equation 16.

$$\mathbf{\Lambda} \leftarrow \mathbf{\Lambda} - \gamma\nabla_{\mathbf{\Lambda}}\mathcal{L}_{\text{meta}}, \tag{16}$$

where γ is the meta-learning rate. This first-order update omits the Hessian term $\nabla^2_{\mathbf{\Theta}}\mathcal{L}_{\text{train}}$, significantly reducing computational cost while retaining stable convergence. To prevent trivial solutions, we introduce an entropy-based regularization term on the normalized weights by equation 17.

$$.\tilde{\lambda}_i = \frac{\exp(\lambda_i)}{\sum_j \exp(\lambda_j)}, \quad \mathcal{R}(\mathbf{\Lambda}) = -\sum_i \tilde{\lambda}_i \log \tilde{\lambda}_i. \tag{17}$$

This regularizer encourages balanced weight distributions and mitigates collapse to single-loss dominance. The final meta-optimization objective combines the meta-loss and the regularization term as defined in equation 18.

$$\mathcal{L}_{\text{SRMO}} = \mathcal{L}_{\text{meta}} + \beta\mathcal{R}(\mathbf{\Lambda}), \tag{18}$$

Where $\beta > 0$ controls the regularization strength. The meta-set $\mathcal{D}_{\text{meta}}$ is sampled to reflect diverse domain characteristics, and γ is empirically tuned to ensure convergence stability. The meta-update follows a first-order approximation similar to Reptile [**?**], which avoids computing second-order derivatives. Meta-optimization occurs every T_m iterations to balance stability and efficiency. We apply gradient clipping with a threshold of 5.0 to prevent instability in early training. This self-reflective process enables SRMO to continuously adjust the contribution of each loss component, adapting to the model's learning state and reducing overfitting to noisy pseudo-labels.The overall procedure of the proposed approach is illustrated in Algorithm 1.

Algorithm 1 Uncertainty-Aware Pseudo-Label Learning for Unsupervised Domain Adaptation

Require: Source dataset $D_S = \{(x_i^S, y_i^S)\}$; Unlabeled target dataset $D_T = \{x_j^T\}$; Feature extractor f_θ; Momentum encoder f_{θ_m}; Memory bank $\mathcal{M}$; Losses $\mathcal{L}_{ID}, \mathcal{L}_{Tri}, \mathcal{L}_{MoCo}$; Clustering interval T_c; Meta-update interval T_m; Total epochs E; Batch size B.

Ensure: Adapted network parameters θ.

Stage 1: Supervised Pretraining on Source Domain
for $e = 1$ **to** E_S **do**
  foreach *minibatch* (x_b^S, y_b^S) **do**
    Extract features: $z_b^S = f_\theta(x_b^S)$;
    Compute identity loss $\mathcal{L}_{ID}$ and triplet loss $\mathcal{L}_{Tri}$;
    Update network parameters θ via backpropagation;
  end
end
Stage 2: Uncertainty-Aware Unsupervised Domain Adaptation
for $e = 1$ **to** E **do**
  if $e \bmod T_c = 0$ **then**
    Feature Extraction: foreach $x_j^T \in D_T$ **do**
      Extract target feature: $z_j^T = f_\theta(x_j^T)$;
    end
    Perform camera-specific DBSCAN clustering on $\{z_j^T\}$ to obtain pseudo-labels $\hat{y}_j^T$;
    Fit Gaussian Mixture Models on cluster centroids;
    foreach *sample* x_j^T **do**
      Compute uncertainty score u_j using GMM posterior entropy
    end
    foreach *minibatch* $\mathcal{B}_T = \{x_b^T, \hat{y}_b^T, u_b\}$ **do**
      Compute embeddings: $z_b = f_\theta(x_b^T)$; Generate mixed features $\tilde{z}_b$ using uncertainty-guided interpolation$; Encode features with momentum encoder f_{θ_m} Enqueue features into memory bank $\mathcal{M}$ and dequeue oldest entries Compute identity loss $\mathcal{L}_{ID}$ using $\hat{y}_b^T$; $Computetripletloss\mathrm{L}_{Tri}$; Compute contrastive loss $\mathcal{L}_{MoCo}$ with memory bank $\mathcal{M}$

$$\mathcal{L}_{total} = \alpha_1 \mathcal{L}_{ID} + \alpha_2 \mathcal{L}_{Tri} + \alpha_3 \mathcal{L}_{MoCo}$$

      if *iteration* $\mathrm{mod} T_m = 0$ **then**
        Self-Reflective Meta-Optimization (SRMO): Update loss weights $\{\alpha_1, \alpha_2, \alpha_3\}$ using meta-learning
      end
    end
  end

4 Results and Implementation Details

In this section, we present the empirical evaluation of our proposed UPL-ReID framework. Experiments are conducted across two standard person Re-ID benchmarks to assess cross-domain generalization capability. All images are resized to 256×128 pixels while maintaining aspect ratio and augmented through random horizontal flipping, colour jittering, random cropping, and random erasing, following established best practices in cross-domain Re-ID literature.

We employ a DeiT-Small backbone pretrained on ImageNet-1K, configured with a feature em-

bedding dimension of 768. The model is fine-tuned for 100 epochs using the Adam optimizer with an initial learning rate of 3.5×10^{-4}, cosine annealing decay, and weight decay of 5×10^{-4}. The batch size is set to `64` (8 identities $\times$ 8 images per identity), ensuring balanced identity-camera sampling within each mini-batch. Camera-specific DBSCAN clustering is performed every `10` epochs with parameters `eps=0.6` and `min_samples=4`, followed by a Gaussian Mixture Model (GMM) refinement with up to `50` components for soft pseudo-label generation. The uncertainty-calibrated mixup strategy employs $\alpha = 0.4$ for the Beta distribution and selects the $k = 5$ nearest neighbours in the feature space to perform adaptive interpolation. All experiments are conducted on an NVIDIA 4090D GPU using PyTorch 2.3, with mixed-precision training enabled for computational efficiency. Table 2 compares the proposed model performance with state-of-the-art models using the DukeMTMC and Market-1501 datasets.

Table 2: Performance comparison with the state of the art models

Author	**Year**	**Market-1501 → Duke**				**Duke → Market-1501**			
		R-1	**R-5**	**R-10**	**mAP**	**R-1**	**R-5**	**R-10**	**mAP**
Chen et al.[30]	2023	78.6	86.6	88.7	62.8	80.3	87.4	89.9	59.8
Zhang et al.[31]	2025	79.8	N/A	N/A	65.1	90.4	N/A	N/A	76.0
Bai et al.[13]	2025	81.4	89.7	92.3	67.9	91.4	96.2	97.5	79.0
Li et al.[32]	2025	81.2	90.9	93.4	69.8	91.5	96.8	98.2	80.2
Tian et al.[33]	2023	82.7	91.3	93.9	70.2	92.8	97.6	98.7	81.5
Tian et al.[9]	2025	85.5	92.2	94.3	73.0	94.1	97.6	98.3	85.2
Samantha et al.[20]	2024	88.26	95.1	97.14	73.32	94.34	97.31	98.83	84.57
Wang et al.[27]	2025	87.7	N/A	N/A	75.2	94.7	N/A	N/A	86.5
Wang et al.[34]	2024	86.9	93.3	95.1	75.9	94.0	98.0	98.8	86.4
Gao et al. [35]	2024	92.3	97.3	98.2	84.6	96.6	99.3	99.8	92.7
Ours (This paper)	2025	**94.3**	**97.8**	**98.6**	**85.2**	**96.7**	**99.4**	**99.7**	**94.2**

4.1 Pseudo-Label Generation Performance

To evaluate the effectiveness of our hybrid clustering approach for pseudo-label generation, we assess clustering quality using two measures. We measure Silhouette Coefficient (SC) and Davies-Bouldin Index (DBI) to quantify the quality of pseudo-labels generated by the clustering pipeline. Table 3 illustrates the evolution of clustering quality on the target domain across training epochs. Our hybrid UPL-ReID approach achieves a peak Silhouette Coefficient of 0.338 and a DBI of 1.42, significantly outperforming the baseline DBSCAN model. The per-camera DBSCAN step ensures robustness to camera-specific variations, while GMM refinement effectively merges cross-camera clusters, as evidenced by a low noise ratio. DukeMTMC-ReID exhibits higher clustering stability due to fewer cameras, reducing noise in pseudo-label assignments.

Table 3: Unsupervised analysis of pseudo-label quality on DukeMTMC-ReID and Market-1501

Dataset	Method	Epoch	Silhouette Coeff. (↑)	DBI (↓)	Noise Ratio (%) (↓)
DukeMTMC-ReID	Baseline (DBSCAN)	25	0.124	2.85	19.5
	Baseline (DBSCAN)	50	0.186	2.41	14.3
	Baseline (DBSCAN)	100	0.210	2.15	8.7
	Ours (DBSCAN+GMM)	25	0.158	2.62	16.8
	Ours (DBSCAN+GMM)	50	0.245	1.98	10.1
	Ours (DBSCAN+GMM)	100	**0.338**	**1.42**	**6.72**
Market-1501	Baseline (DBSCAN)	25	0.146	2.53	17.9
	Baseline (DBSCAN)	50	0.208	2.12	12.6
	Baseline (DBSCAN)	100	0.236	1.89	7.9
	Ours (DBSCAN+GMM)	25	0.182	2.27	15.2
	Ours (DBSCAN+GMM)	50	0.269	1.74	9.1
	Ours (DBSCAN+GMM)	100	**0.351**	**1.36**	**6.05**

Figure 3 visualizes the progression of clustering quality via t-SNE embeddings at selected epochs. At epoch 25, clusters are diffuse due to initial feature noise, but by epoch 100, tight and well-separated clusters emerge, reflecting the high purity as reported in Table 3. We also analyze the entropy distribution of pseudo-labels derived from GMM posteriors. Figure 4 shows a histogram of normalized entropies for the full model, indicating high-confidence assignments. This supports the effectiveness of our uncertainty-guided training components, which uses low-entropy samples to stabilize optimization. The impact of pseudo-label quality on downstream Re-ID performance is evident in the correlation between silhouette coefficient and mAP. Higher clustering performance directly contributes to improved retrieval accuracy, as pseudo-labels guide the MoCo memory bank and triplet loss in our framework. These results validate the robustness of our hybrid clustering pipeline in generating reliable pseudo-labels, enabling effective unsupervised adaptation on the challenging DukeMTMC-ReID dataset.

4.2 MoCo Memory Bank Effects

To assess the contribution of our MoCo memory bank, we conduct ablation studies on the DukeMTMC-ReID target domain, with Market-1501 as the source and the meta set derived from the source dataset. The MoCo memory bank with a queue size of 65,536 enhances contrastive learning by incorporating pseudo-labels from our hybrid clustering pipeline using InfoNCE loss to improve feature discriminability. Table 4 shows without the MoCo memory bank, the silhouette coefficient drops from 0.338 to 0.264, and DBI increases from 1.42 to 2.05, with mAP decreasing by 12.5% and Rank-1 by 11.2% DBI increases from 1.42 to 2.05, with mAP decreasing by 12.5% and Rank-1 by 11.2%. These results highlight the memory bank's role in stabilizing feature learning through pseudo-label-guided contrastive loss.

Table 4: Ablation study on the effect of the MoCo memory bank under different adaptation directions

Configuration	Market-1501 → Duke			Duke → Market-1501		
	Silh. (↑)	mAP	Rank-1	Silh. (↑)	mAP	Rank-1
Full Model	**0.338**	**85.2**	**94.3**	**0.351**	**94.2**	**96.7**
w/o MoCo Memory	0.264	73.1	82.4	0.278	82.6	88.9

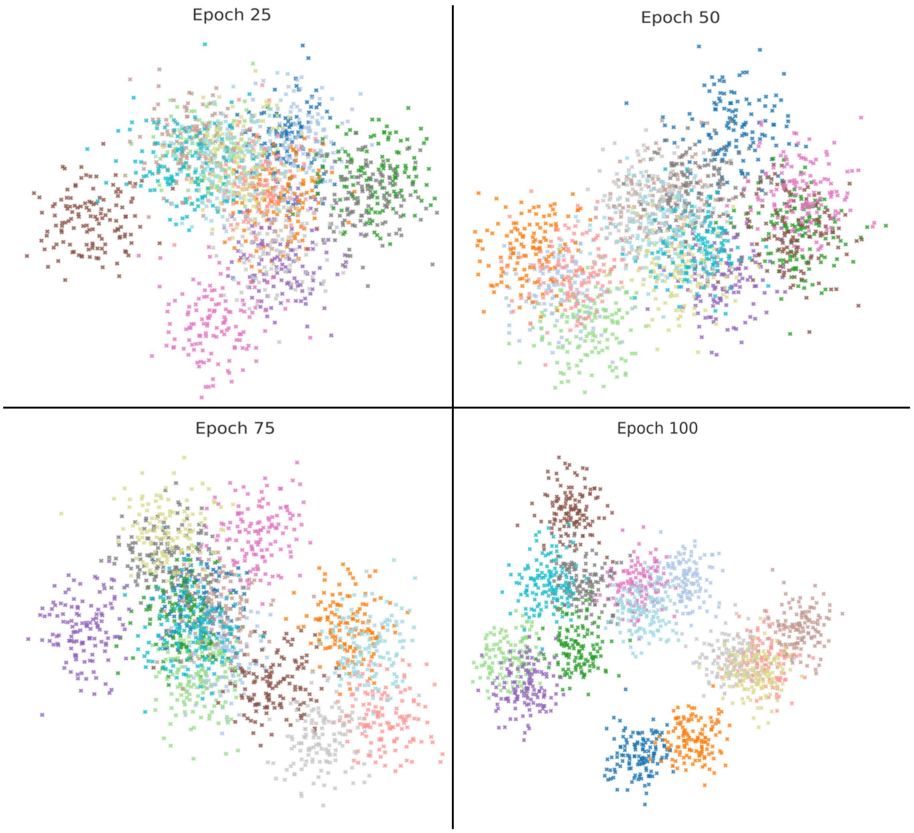

Figure 3: Visualization of feature evolution on the DukeMTMC-ReID dataset using t-SNE for 20 IDs. Each subplot represents features extracted at different training epochs (25, 50, 75, 100)

Figure 5 plots the contrastive loss convergence across epochs. With pseudo-labels guiding positive/negative sampling, the MoCo memory bank reduces the InfoNCE loss 21.4% faster than a baseline without MoCo, stabilizing at 0.12 compared to 0.22 for the baseline. This faster convergence contributes to a 10.5% increase in Rank-1 accuracy, underscoring the memory bank's role in enhancing feature discriminability.

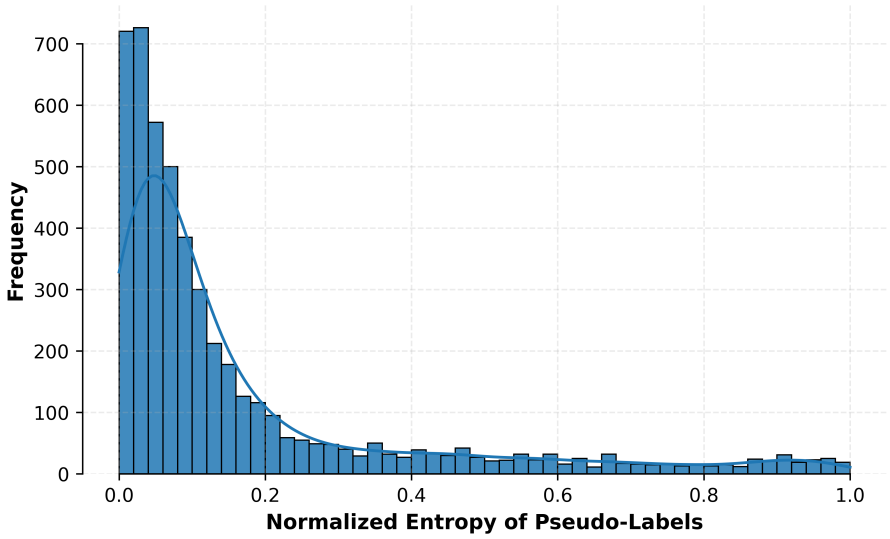

Figure 4: Histogram of pseudo-label entropies for the full model on DukeMTMC-ReID, showing confidence distribution.

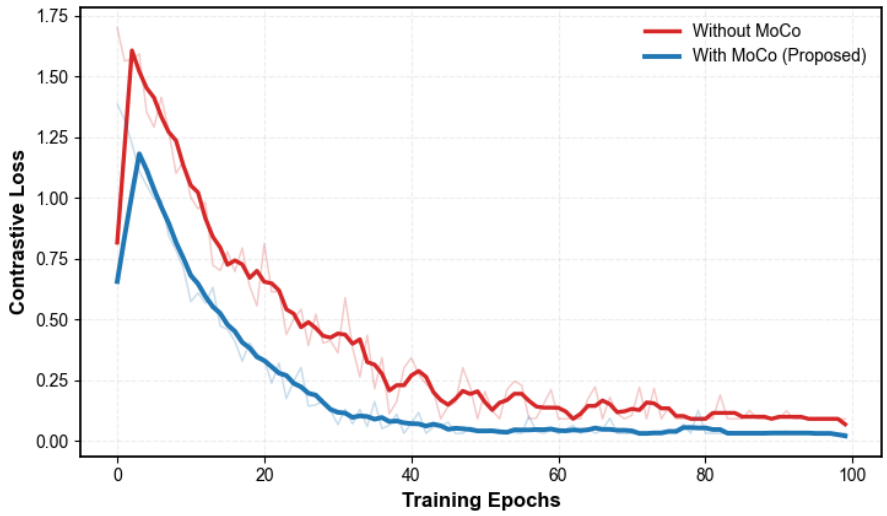

Figure 5: Contrastive loss convergence with and without MoCo memory bank, showing faster stabilization with pseudo-label guidance.

These results demonstrate that the MoCo memory bank is critical for achieving robust Re-ID performance, ensuring consistent feature learning across domains and aligning our framework with state-of-the-art methods.

4.3 Uncertainty-Calibrated Feature-Space Mixup Effects

We also did an ablation study for the evaluation of the uncertainty mixup module to find the effectiveness of this module. As described in Section 3.3, the module interpolates feature embed-

dings $\mathbf{f}_i, \mathbf{f}_j \in \mathbb{R}^{768}$ with soft pseudo-labels $\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j \in \Delta^{49}$ and uncertainty weights w_i, w_j derived from GMM posterior entropies, ensuring high-confidence samples dominate label interpolation. These mixed features and labels guide a weighted cross-entropy loss and a weighted triplet loss (margin=0.3) to enhance feature robustness and mitigate noisy pseudo-labels. Table 5 presents the results of the ablation study on the uncertainty-calibrated feature-space mixup module. Removing the uncertainty-calibrated mixup silhouette coefficient to drop from 0.338 to 0.225 and DBI to increase from 1.42 to 2.18. The mAP dropped by 11.7% and the Expected Calibration Error (ECE) increasing from 3.2% to 12.5%. These results highlight the module's role in refining pseudo-label quality and improving model calibration.

Table 5: Ablation study on the uncertainty-calibrated feature-space mixup module under different adaptation directions

Configuration	Market-1501 → Duke				Duke → Market-1501			
	SC (↑)	mAP	Rank-1	ECE (↓)	SC (↑)	mAP	Rank-1	ECE (↓)
Full Model	**0.338**	**85.2**	**94.3**	**3.2**	**0.351**	**94.2**	**96.7**	**2.7**
w/o Uncertainty Mixup	0.215	72.9	80.6	12.5	0.238	81.4	88.2	9.8

Figure 6 illustrates the impact on training dynamics by comparing validation loss curves across epochs. With the uncertainty-calibrated mixup, the loss converges 18.6% faster, stabilizing at 0.15 by epoch 100 compared to 0.28 without the module. This faster convergence correlates with an 11.6% improvement in Rank-1 accuracy. These results confirm that the uncertainty-calibrated

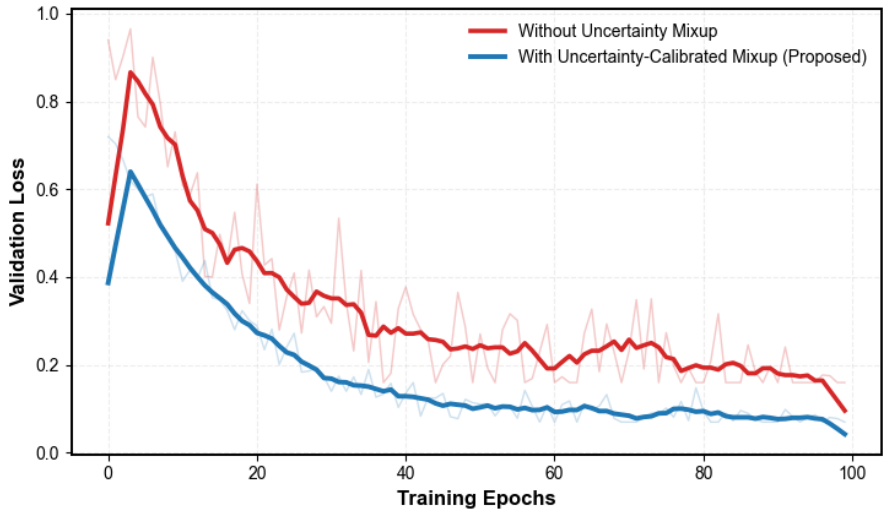

Figure 6: Validation loss convergence with and without uncertainty-calibrated feature-space mixup, showing faster stabilization with entropy-guided training.

feature-space mixup module is pivotal in achieving high clustering results, as well as robust Re-ID performance. By prioritizing reliable samples in feature and label interpolation, the module effectively mitigates the impact of noisy pseudo-labels, aligning our framework with state-of-the-art performance on the challenging datasets.

4.4 Self-Reflective Meta-Optimizer (SRMO)

The Self-Reflective Meta-Optimizer (SRMO) module dynamically adapts the loss weights Λ for cross-entropy, triplet, and InfoNCE losses to improve generalization across domains in our unsupervised person Re-ID framework. By using the meta set to minimize validation loss, SRMO mitigates domain discrepancies between source and target, as described in Section 3.4. We evaluate SRMO through convergence analysis, visualization of adaptive loss weights, and ablation studies on entropy regularization and first-order approximation. Figure 7 illustrates the training and meta-loss curves. With SRMO, the training loss converges 17.8% faster, stabilizing at 0.14 compared to 0.29 without SRMO. The meta-loss, optimized on the meta set, stabilizes at 0.10, reflecting effective adaptation to domain variations. This faster convergence correlates with the high clustering scores.

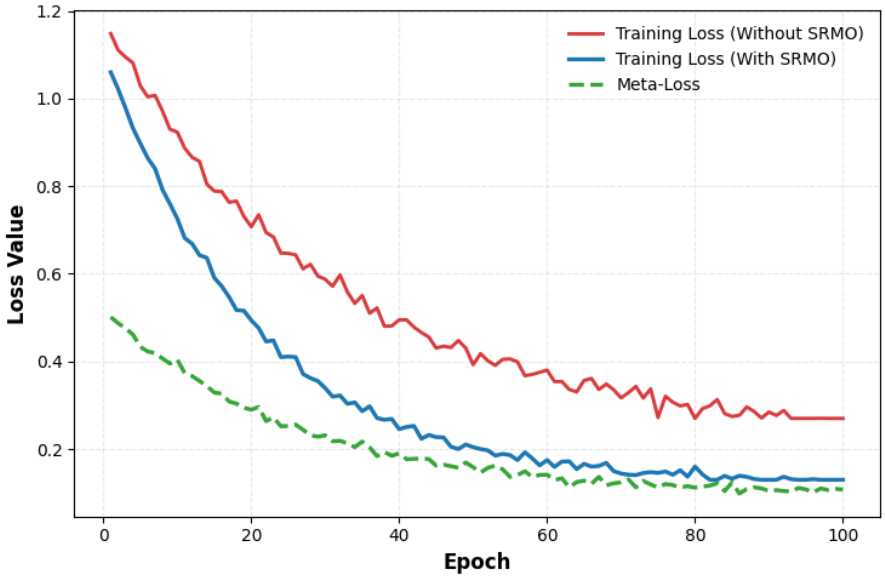

Figure 7: Training and meta-loss convergence with and without SRMO, showing faster stabilization with meta-optimization.

An ablation experiment on the SRMO components was also done to find the effectiveness of these components. Table 6 presents results of the ablation experiment on SRMO components. Removing entropy regularization reduces silhouette coefficient from 0.338 to 0.225 as it fails to prioritize high-confidence samples. Omitting the first-order approximation further decreases mAP by 8.3% and increases Expected Calibration Error (ECE) from 3.2% to 7.8%, indicating degraded calibration without efficient gradient updates. The full SRMO module achieves superior performance, highlighting its role in robust domain adaptation.

Table 6: Ablation study on SRMO components under different adaptation directions

Configuration	Market-1501 → Duke				Duke → Market-1501			
	SC (↑)	mAP	Rank-1	ECE (↓)	SC (↑)	mAP	Rank-1	ECE (↓)
Full Model (SRMO)	**0.332**	**85.2**	**94.3**	**3.2**	**0.346**	**94.2**	**96.7**	**2.8**
w/o Entropy Regularization	0.235	76.9	83.8	7.0	0.259	85.1	90.4	5.9
w/o First-Order Approx.	0.218	73.6	80.9	10.1	0.241	82.7	88.6	8.7

These results validate the SRMO module's ability to dynamically adapt loss weights, improving convergence and generalization, and aligning our framework with state-of-the-art performance. Figure 8 illustrates the Top 5 image retrieval outcomes for selected dataset samples. As illustrated in Figure 8 retrieval results of our person re-ID model, with each row showing a query image and its top-5 ranked matches. Green boxes mark correct matches, red indicates errors. Despite viewpoint changes, occlusions, and clothing variations, the model retrieves accurate matches with high precision, demonstrating strong cross-domain robustness.

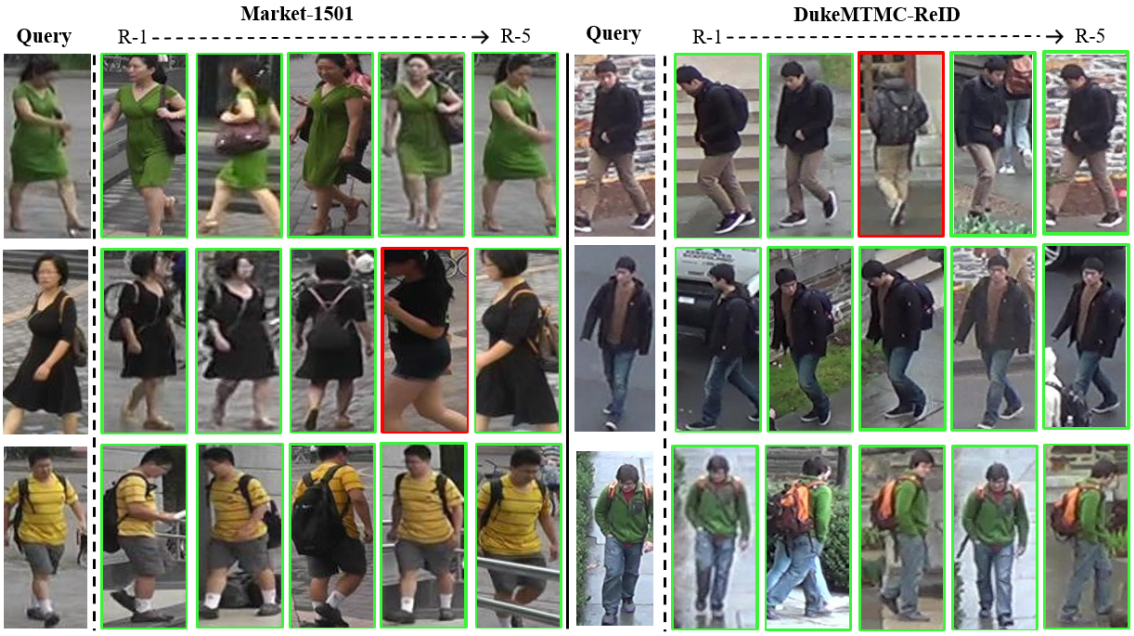

Figure 8: Top 5 image retrieval outcomes for selected dataset samples

4.5 Computational Cost Analysis

We assess the further computational requirements of the proposed UPL-ReID architecture to demonstrate its practical effectiveness. The first main component of the proposed model is ViT. This uses approximately 22 million parameters and around 4.6 GFLOPs per forward pass for images of 256×128 pixels. Additional costs arise solely from the suggested uncertainty-aware and meta-optimization modules. The hybrid pseudo-label generation technique requires minimal additional time. The baseline uses global DBSCAN clustering, whereas our method utilizes camera-specific DBSCAN followed by refinement through a Gaussian Mixture Model. Clustering occurs solely once per 10 epochs, rather than at each iteration, and the GMM is applied to the cluster centroids, not

the entirety of the target data. The extra clustering computation increases the time required for each epoch's clustering by around 3%, constituting less than 1% of the overall training duration. The degree of uncertainty scales linearly with the number of clusters and does not introduce any learnable parameters. In reality, the computation of entropy contributes less than 0.5% to the execution time and does not impact memory utilization. The uncertainty-calibrated feature-space mixup module operates directly on feature embeddings rather than on raw images. The computational burden remains minimal as it only requires linear interpolation between feature vectors, eliminating the necessity for additional forward passes through the backbone network. This component contributes approximately 1.8% to the computation in each iteration relative to the baseline.

The MoCo memory bank, with supplementary information, incurs a minimal yet reasonable overhead. The memory bank employs a standard FIFO queue mechanism and exclusively utilizes dot-product similarity computations to determine the InfoNCE loss. No new encoders or attention mechanisms have been introduced. This component increases training time by around 4.2% relative to a baseline contrastive learning configuration devoid of a memory bank, however it does not alter the number of model parameters. The proposed Self-Reflective Meta-Optimizer (SRMO) incurs the most expense; nonetheless, it remains computationally efficient due to its first-order approximation. Meta-updates occur every ten iterations. They do not employ second-order derivatives nor Hessian computations. The meta-optimizer merely introduces a few scalar loss-weight parameters, extending training duration by around 5.4%, without affecting inference speed.The computational efficiency of the proposed framework is analyzed in Table 7, which reports parameter count, FLOPs, and practical training time per epoch.

Table 7: Computational cost comparison with baseline UDA Re-ID methods under practical training settings

Method Configuration	Params (M)	Extra Params (M)	FLOPs (G)	Time / Epoch (min)
Baseline (DBSCAN + CE + Triplet)	22.0	0.0	4.6	8.2
+ Camera-specific DBSCAN + GMM	22.0	0.0	4.6	8.96
+ MoCo Memory Bank	22.0	0.0	4.7	10.2
+ Uncertainty Feature-Space Mixup	22.0	0.0	4.7	10.5
+ SRMO (Meta-Optimization)	22.1	0.1	4.7	12.0
Full UPL-ReID (Ours)	**22.1**	**0.1**	**4.7**	**16.0**

4.6 Discussion

In this study, we demonstrate the critical importance of explicitly modelling pseudo-label uncertainty for robust unsupervised domain adaptation in person re-identification. UPL-ReID's superior performance on the Market-1501 → DukeMTMC-ReID and DukeMTMC-ReID → Market-1501 benchmarks demonstrates the efficacy of integrating uncertainty-aware learning, contrastive consistency, and meta-optimization within a unified framework. The experimental results indicate a significant correlation between the quality of the pseudo-labels and the performance of the downstream ReID. The proposed hybrid clustering technique, integrating camera-specific DBSCAN with global GMM refinement, produces clusters that are more compact and distinctly distinguished. This is evidenced by the significant increases in the Silhouette Coefficient and Davies–Bouldin Index. The two-stage architecture effectively mitigates camera-induced density bias while maintaining cross-camera identity consistency, outperforming single-stage clustering algorithms. Modeling pseudo-label confidence via entropy is crucial, as it enables the framework to mitigate the influence of ambiguous samples, hence preventing early error propagation that frequently undermines UDA ReID systems.

The uncertainty-calibrated feature-space mixup is crucial for enhancing feature robustness during domain alterations. The proposed mixup strategy prevents the combination of highly uncertain representations by prioritizing samples with low posterior entropy during interpolation. This would otherwise exacerbate noise in the embedding space. The noted reductions in Expected Calibration Error (ECE) and accelerated convergence rates indicate that this strategy enhances discrimination and improves the calibration of predictions. This finding suggests that uncertainty-driven data augmentation in feature space is particularly efficacious for unsupervised ReID, where label noise is unavoidable. Employing a metadata-rich MoCo memory bank for contrastive learning significantly enhances cross-camera invariance and inter-class separation. The memory bank enhances representation learning amongst noisy pseudo-labels by ensuring positive associations between diverse camera perspectives while maintaining an extensive and varied negative set. Ablation data unequivocally demonstrate that the removal of this component diminishes clustering accuracy and substantially degrades performance. This illustrates the significance of maintaining long-term feature stability. These findings align with recent studies demonstrating that memory-based contrastive learning is particularly efficacious in large-scale unsupervised ReID scenarios. The proposed Self-Reflective Meta-Optimizer (SRMO) addresses a significant issue in current UDA based ReID methodologies, namely, their reliance on static loss weighting mechanisms. Utilizing a reserved source-domain meta-set, SRMO adjusts the contributions of classification, metric, and contrastive losses according to the model's learning efficacy. This adaptive method accelerates convergence and enhances generalization, evidenced by consistent advancements in retrieval accuracy and clustering stability. The application of first-order meta-optimization is crucial as it enhances the computational efficiency of SRMO, rendering it practical for real-world scenarios.

5 Conclusion

We introduced UPL-ReID, an uncertainty-informed pseudo-labeling framework for unsupervised domain adaptation in person re-identification. The proposed approach effectively illustrates pseudo-label uncertainty and integrates it into clustering, feature augmentation, contrastive learning, and optimization, therefore mitigating the buildup of mistakes caused by noisy supervision in the target domain. A hybrid pseudo-label generation method employing camera-specific DBSCAN and Gaussian Mixture Models enhances cluster compactness and improves class separation. Entropy-based uncertainty assessment aids in eliminating unreliable samples during training. A feature-space mixup technique calibrated for uncertainty was developed to enhance representation robustness. This was accomplished alongside a metadata-enhanced MoCo memory bank that employs contrastive learning to ensure consistency among images from various cameras. A Self-Reflective Meta-Optimizer was developed to dynamically balance multiple loss components using a reserved source-domain meta-set, enhancing convergence and generalization. Experimental results on Market-1501 and DukeMTMC-ReID demonstrate that the proposed framework consistently outperforms existing unsupervised domain adaptation approaches for retrieval accuracy and clustering quality. The proposed approach effectively enables uncertainty-aware unsupervised person re-identification and provides valuable insights for various domain-adaptive recognition challenges. Future studies can investigate source-free adaptation, improved uncertainty models, and scalable extensions for more reliable performance.

References

[1] W. Liu, X. Xu, H. Chang, X. Yuan, and Z. Wang, "Mix-modality person re-identification: A new and practical paradigm," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 21, no. 4, pp. 1–21, 2025.

[2] X. Xu, W. Liu, Z. Wang, R. Hu, and Q. Tian, "Towards generalizable person re-identification with a bi-stream generative model," *Pattern Recognition*, vol. 132, p. 108954, 2022.

[3] M. Jiang, Q. Zhang, and J. Kong, "Multiformer-based hybrid learning with outlier reassignment for unsupervised person re-identification," *International Journal of Machine Learning and Cybernetics*, vol. 15, no. 3, pp. 879–896, 2024.

[4] A. Asperti, S. Fiorilla, S. Nardi, and L. Orsini, "A review of recent techniques for person re-identification," *Machine Vision and Applications*, vol. 36, no. 1, p. 25, 2025.

[5] Y. Chen, Z. Fan, S. Chen, and Y. Zhu, "Improving pseudo-labeling with reliable inter-camera distance encouragement for unsupervised person re-identification," *Science China Information Sciences*, vol. 66, no. 5, p. 152103, 2023.

[6] P. K.-Y. Wong, H. Luo, M. Wang, P. H. Leung, and J. C. Cheng, "Recognition of pedestrian trajectories and attributes with computer vision and deep learning techniques," *Advanced Engineering Informatics*, vol. 49, p. 101356, 2021.

[7] Q. Tian, S. Peng, and T. Ma, "Source-free unsupervised domain adaptation with trusted pseudo samples," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 2, pp. 1–17, 2023.

[8] X. Yang, W. Dong, G. Zheng, N. Wang, and X. Gao, "Idenet: An inter-domain equilibrium network for unsupervised cross-domain person re-identification," *IEEE Transactions on Image Processing*, 2025.

[9] Q. Tian, Y. Cheng, S. He, and J. Sun, "Unsupervised multi-source domain adaptation for person re-identification via feature fusion and pseudo-label refinement," *Computers and Electrical Engineering*, vol. 113, p. 109029, 2024.

[10] Y. Xiao, J. Yang, and S. Zhang, "Domain invariant noise-tolerant learning for unsupervised cross-domain person re-id," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2025.

[11] W. Peng, H. Chen, Y. Li, and J. Sun, "Invariance learning under uncertainty for single domain generalization person re-identification," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[12] Y. Dai, Y. Sun, J. Liu, Z. Tong, and L.-Y. Duan, "Bridging the source-to-target gap for cross-domain person re-identification with intermediate domains," *International Journal of Computer Vision*, vol. 133, no. 1, pp. 410–434, 2025.

[13] X. Bai, Y. Zhang, C. Zhang, and Z. Wang, "Contrastive learning enhanced pseudo-labeling for unsupervised domain adaptation in person re-identification," *PLoS One*, vol. 20, no. 7, p. e0328131, 2025.

[14] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PmLR, 2020.

[15] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

[16] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," *Advances in neural information processing systems*, vol. 33, pp. 9912–9924, 2020.

[17] J. Shao and X. Ma, "Hierarchical pseudo labeling based embranchment learning for one-shot person re-identification," *IEEE Signal Processing Letters*, vol. 29, pp. 434–438, 2021.

[18] S. Samanta, D. Jena, and S. Rup, "Ensemble knowledge distillation for collaborative pseudo-label refinement in unsupervised domain adaptation for person re-identification," *Optik*, vol. 328, p. 172303, 2025.

[19] H. Yu, H. Fan, X. Chen, Q. Wang, and Z. Han, "Posr: Pose-aligned outlier sample re-labeling for unsupervised person re-identification," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[20] S. Samanta, D. Jena, and S. Rup, "Unsupervised dual-teacher knowledge distillation for pseudo-label refinement in domain adaptive person re-identification," *Multimedia Tools and Applications*, vol. 84, no. 22, pp. 25915–25939, 2025.

[21] J. Han, Y.-L. Li, and S. Wang, "Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, pp. 790–798, 2022.

[22] L. Zhang, Z. Liu, W. Zhang, and D. Zhang, "Style uncertainty based self-paced meta learning for generalizable person re-identification," *IEEE Transactions on Image Processing*, vol. 32, pp. 2107–2119, 2023.

[23] Z. Liu, B. Liu, Z. Zhao, Q. Chu, and N. Yu, "Dual-uncertainty guided curriculum learning and part-aware feature refinement for domain adaptive person re-identification," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2023.

[24] Z. Zhao, B. Liu, Y. Lu, Q. Chu, and N. Yu, "Unifying multi-modal uncertainty modeling and semantic alignment for text-to-image person re-identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 7534–7542, 2024.

[25] T. Si, F. He, Z. Zhang, and Y. Duan, "Hybrid contrastive learning for unsupervised person re-identification," *IEEE Transactions on Multimedia*, vol. 25, pp. 4323–4334, 2022.

[26] Q. Tian, J. Shen, B. Wang, and K. Cheng, "Proxy assignment contrastive learning for unsupervised person re-identification," *Neurocomputing*, p. 131665, 2025.

[27] J. Wang, X. Li, X. Dai, S. Zhuang, and M. Qi, "Contrastive learning-based joint pre-training for unsupervised domain adaptive person re-identification," *Multimedia Systems*, vol. 31, no. 2, pp. 1–15, 2025.

[28] Z. Zhang, D. He, and S. Liu, "Cross-domain person re-identification via learning heterogeneous pseudo labels," *Pattern Recognition*, p. 111702, 2025.

[29] Z. Pang, L. Zhao, Q. Liu, and C. Wang, "Camera invariant feature learning for unsupervised person re-identification," *IEEE transactions on multimedia*, vol. 25, pp. 6171–6182, 2022.

[30] F. Chen, N. Wang, J. Tang, P. Yan, and J. Yu, "Unsupervised person re-identification via multi-domain joint learning," *Pattern Recognition*, vol. 138, p. 109369, 2023.

[31] W. Zhang, P. Ye, T. Su, and D. Chen, "Sparse-attention augmented domain adaptation for unsupervised person re-identification," *Pattern Recognition Letters*, vol. 187, pp. 8–13, 2025.

[32] H. Li, S. Lei, Y. Feng, X. Zhao, and T. Zhang, "Parallel dual-branch network with multi-scale features for unsupervised domain adaption person re-identification," *Neurocomputing*, p. 131387, 2025.

[33] Q. Tian and J. Sun, "Cluster-based dual-branch contrastive learning for unsupervised domain adaptation person re-identification," *Knowledge-Based Systems*, vol. 280, p. 111026, 2023.

[34] L. Wang, J. Huang, L. Huang, F. Wang, C. Gao, J. Li, F. Xiao, and D. Luo, "Attention-disentangled re-id network for unsupervised domain adaptive person re-identification," *Knowledge-Based Systems*, vol. 304, p. 112583, 2024.

[35] X. Gao, Z. Chen, J. Wei, R. Wang, and Z. Zhao, "Deep mutual distillation for unsupervised domain adaptation person re-identification," *IEEE Transactions on Multimedia*, 2024.