Applied Soft Computing Journal 174 (2025) 112932

Contents lists available at ScienceDirect

Applied Soft Computing

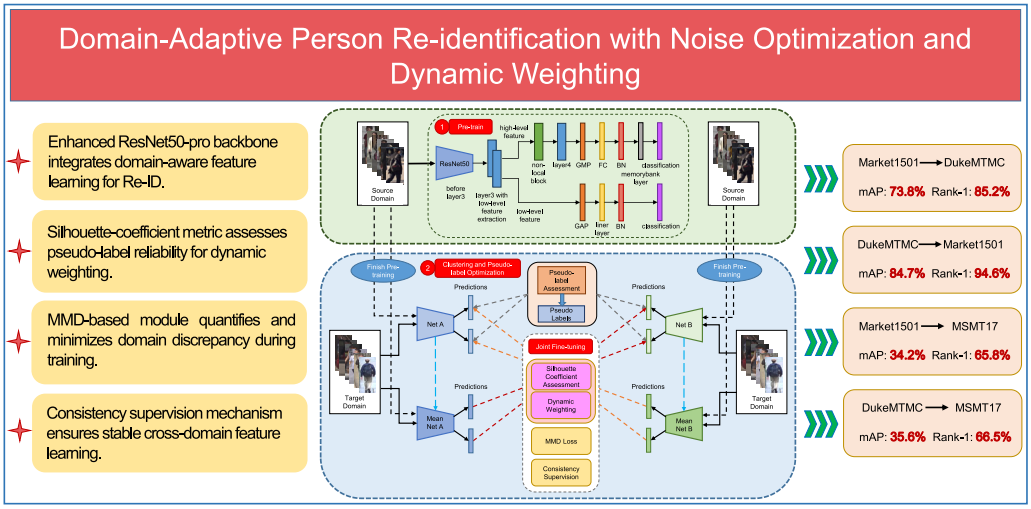
journal homepage: www.elsevier.com/locate/asoc

Domain adaptive person re-identification with noise optimization and dynamic weighting

Zhengyang Wang [a], Xiufen Ye [a,*], Xue Shang [a,b], Shuxiang Guo [c]

[a] College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, 150001, China
[b] Department of Electrical, Computer and Software Engineering, University of Auckland, Auckland, 1142, New Zealand
[c] Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen, 518055, China

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Domain adaptive
Person re-identification
Pseudo-label refinement
Dynamic weighting

ABSTRACT

Domain adaptive person re-identification (Re-ID) faces challenges due to inherent noise from limited domain transferability and the uncertainty in pseudo-label generation. To address this, we propose NODW (Noise Optimization and Dynamic Weighting), a comprehensive domain adaptive person Re-ID framework that systematically tackles these issues through quantitative noise assessment and dynamic optimization. Our method proposes: (1) an enhanced ResNet50-pro backbone specifically designed for cross-domain feature extraction, (2) a silhouette coefficient-based module for pseudo-label quality assessment with dynamic weighting, (3) a Maximum Mean Discrepancy (MMD)-based module for minimizing domain transferability limitations, and (4) a robust consistency supervision mechanism to ensure stable feature learning. Extensive experiments demonstrate state-of-the-art performance across multiple domain transfer tasks, achieving mAP scores of 73.8% (Market to Duke), 84.7% (Duke to Market), 34.2% (Market to MSMT), and 35.6% (Duke to MSMT). These results represent significant improvements over existing methods, particularly in challenging scenarios with large domain gaps, validating the effectiveness of our noise-aware adaptation strategy.

* Corresponding author.
E-mail address: yexiufen@hrbeu.edu.cn (X. Ye).

https://doi.org/10.1016/j.asoc.2025.112932
Received 5 August 2024; Received in revised form 23 January 2025; Accepted 21 February 2025
Available online 10 March 2025

1. Introduction

Domain Adaptive Person Re-identification (Re-ID) refers to the task of identifying and matching individuals across different camera views in surveillance systems, where the data distribution (appearance of individuals, lighting, background, etc.) differs between the source domain and the target domain (unseen). The main goal is to adapt the model trained on the source domain to perform well on the target domain without requiring extensive labeled data [1,2]. A fundamental challenge in domain adaptive Re-ID deployment is the significant performance degradation when models trained on one domain (source) are applied to another (target) due to domain shift [3,4]. This challenge is particularly acute in real-world scenarios, where variations in camera characteristics, lighting conditions, and environmental factors create substantial distributional discrepancies between domains.

Traditional supervised Re-ID methods require extensive manual annotations and struggle with cross-domain generalization [5]. The primary challenges of traditional Re-ID include handling cross-camera variations, intra-class variability, inter-class similarity, occlusions, and scalability issues, which make it difficult to consistently identify individuals across different views. In contrast, domain adaptive Re-ID specifically addresses the domain shift problem, focusing on adapting models to perform well in new, unseen domains without extensive labeled data [6]. Unlike traditional Re-ID, domain adaptive Re-ID leverages unsupervised learning, semi-supervised learning, or self-supervised learning and aims to learn domain-invariant features that generalize across diverse environments, making it more robust for real-world applications [7,8].

Thus, the primary challenges of domain adaptive person Re-ID include Domain Shift and Label Scarcity. Images of different individuals from the same domain often appear more similar than images of the same individual across different domains (Fig. 1), creating core challenges for feature learning. The absence of target domain labels necessitates pseudo-label generation, introducing additional noise and uncertainty into the learning process.

Current pseudo-label-based methods [9–11] typically employ a two-stage framework (Fig. 2): source domain pre-training followed by iterative pseudo-label generation and refinement. Despite their effectiveness, these methods suffer from three major limitations. The first is pseudo-label noise. Existing clustering-based methods [12,13] often employ DBSCAN [14] for its robustness to outliers, making it more effective than K-means in handling complex data distributions and noise. However, they lack systematic assessment of pseudo-label quality, leading to error propagation during training. While MMT [9] introduced teacher–student mutual learning and [15] proposed self-consistency refinement, these methods address symptoms rather than underlying causes of noise. The second is limited transferability. Domain disparities introduce inherent noise that fundamentally limits feature transferability. Previous works like [16,17] employed self-training and triplet loss but failed to quantitatively assess and address this transfer-related noise. Recent attempts [18,19] focus on pseudo-label optimization without explicitly considering domain transfer limitations. The third challenge is consistency maintenance. Existing frameworks struggle to maintain prediction consistency across different views of the same identity, particularly when domain shift is significant. While some methods [10] employ mutual refinement, they lack mechanisms to ensure robust cross-view consistency.

Building upon the challenges identified in existing methods and inspired by MMT [9], we propose NODW (Noise Optimization and Dynamic Weighting), a comprehensive domain adaptive person Re-ID framework that makes several key contributions:

- An enhanced feature extraction backbone (ResNet50-pro) specifically designed for domain adaptive Re-ID, incorporating domain-aware feature learning mechanisms.
- A novel silhouette-coefficient-based noise assessment module, providing quantitative metrics for pseudo-label reliability and enabling dynamic weight adjustment.
- A Maximum Mean Discrepancy (MMD)-based transferability assessment module that explicitly quantifies and minimizes domain discrepancy during training.
- A robust consistency supervision mechanism ensures stable feature learning across domain shifts.

Together, these modules form a cohesive framework that systematically addresses noise, transferability, and consistency challenges in domain adaptive Re-ID.

Our framework fundamentally differs from previous methods. It introduces reliable noise metrics to enable dynamic weighting for pseudo-label refinement, explicitly addresses domain transferability limitations through MMD-based modules, and maintains robust cross-view consistency-supervised alignment. Extensive experiments on Market1501, DukeMTMC-reID, and MSMT17 datasets validate our method. The results demonstrate state-of-the-art performance with significant improvements in both accuracy and robustness. Our approach achieves particular advantages in challenging scenarios with large domain gaps, validating the effectiveness of our noise-aware adaptation strategy.

2. Background

Domain adaptive person Re-ID is a crucial computer vision task aimed at matching individuals across non-overlapping cameras in surveillance networks. A significant challenge in its deployment is the performance degradation when models are applied across diverse domains and conditions due to domain shifts. This often leads to performance degradation, necessitating the development of robust domain adaptive methodologies to bridge the gap between training and deployment distributions. Addressing this challenge requires developing robust domain adaptive methodologies to mitigate the discrepancies between training and deployment distributions.

2.1. Domain adaptive person re-identification

Person Re-ID systems identify individuals across multiple camera views by extracting discriminative features robust to environmental and viewpoint variations [20,21]. Traditional methods primarily relied on supervised learning, where models were trained on extensively annotated datasets with identity labels [22]. While achieving high accuracy, these methods suffered from limited generalization to unseen scenarios and the prohibitive cost of large-scale annotation.

Combining these two learning paradigms, domain adaptive person Re-ID addresses the crucial challenge of transferring knowledge from a labeled source domain to an unlabeled target domain, focusing on learning identity-discriminative features without explicit identity labels [23]. This framework combines the advantages of supervised and unsupervised learning, offering a practical solution for real-world deployment scenarios [5]. Early approaches, such as [24], employed cross-dataset transfer learning through multi-task dictionary learning. [25] advanced this concept by disentangling identity-related and identity-unrelated features, facilitating more focused domain adaptation. Adversarial learning has proven effective in domain adaptive Re-ID. Notable works include PTGAN [26] and ATNet [27], which employ generative adversarial networks for style transfer between domains. Recent advances in self-supervised learning have significantly enhanced pseudo-label refinement for domain adaptive Re-ID, such as [28,29], have focused on clustering re-training, joint loss learning, and data augmentation methods. Clustering and pseudo-label optimization methods generate pseudo-labels for target domain data, facilitating self-supervised learning. Notable improvements have been made in pseudo-label reliability assessment, as demonstrated in [9,11,30].

Fig. 1. Illustration of the domain shift challenge in person Re-ID. Images from different identities within the same camera domain (intra-domain) often show higher feature similarity than images of the same identity across different camera domains (inter-domain), emphasizing the core challenge of cross-domain feature learning.

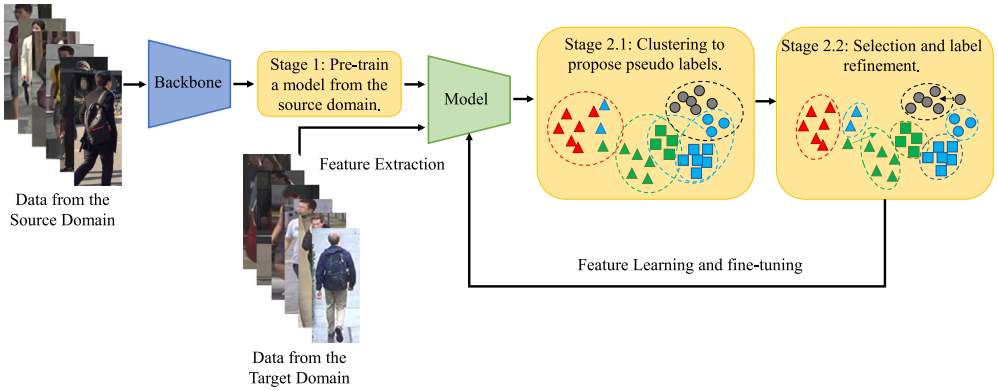


Fig. 2. Overview of conventional pseudo-label-based domain adaptive Re-ID frameworks. The process consists of two main stages: (1) source domain pre-training for initial feature extraction, and (2) iterative pseudo-label generation and optimization in the target domain through clustering-based methods.

2.2. Pseudo-label-based domain adaptation

Pseudo-label-based methods have emerged as a promising strategy for domain adaptive Re-ID. These approaches typically involve source domain pre-training and target domain adaptation through pseudo-label generation and refinement. However, the effectiveness of these methods is often limited by two critical challenges. The first is pseudo-label noise. Current methods usually suffer from noisy pseudo-labels generated through clustering, leading to error propagation during training. The second is noise from limited transferability. The inherent domain gap creates noise in feature representations, affecting the quality of pseudo-labels and subsequent adaptation [25].

While existing works have proposed various strategies for pseudo-label refinement, systematic assessment and optimization of noise sources remain understudied. [31] introduced an enhanced discriminative clustering (AD-Cluster) approach that refines clusters in the

target domain to improve the model's discriminative performance. [32] proposed a multi-label classification loss (MMCL) for label prediction, which includes similarity calculation and cycle consistency to ensure high-quality pseudo-labels, thereby boosting re-ID performance. [9] developed the Mutual Mean-Teaching (MMT) framework, which refines pseudo-labels by offline and online processes to learn better features from the target domain. Similarly, [33] employed a mean teacher approach to assess pseudo-label reliability through uncertainty (measured by consistency level), optimizing pseudo-label quality to improve model performance. [15] introduced the self-consistency refinement method (SECRET), which mutually refines pseudo-labels generated from different feature spaces to enhance cross-domain re-ID performance progressively. However, they lack quantitative measures for noise assessment. The impact of inherent noise due to limited transferability between domains presents another significant challenge. While some methods [26,27] employ adversarial learning to address domain gaps, they often overlook the systematic assessment and mitigation of transfer-related noise.

Current domain adaptive Re-ID methods face several limitations, including the absence of systematic assessment standards for pseudo-label noise, insufficient attention to the inherent noise from limited domain transferability, and the lack of comprehensive frameworks that simultaneously address both types of noise. Our work tackles these issues by introducing a novel framework for the quantitative assessment of pseudo-label noise and a dedicated module for analyzing and optimizing inherent transfer-related noise. Furthermore, we propose a comprehensive framework for domain adaptive person Re-ID that effectively addresses those challenges. This systematic approach to noise assessment and optimization marks a notable advancement in the domain adaptive person re-identification, providing both theoretical insights and practical performance improvements.

3. Methodology

3.1. Problem definition

The primary challenge in domain adaptive person re-identification (Re-ID) is domain shift, wherein the underlying data distributions exhibit significant divergence between source domain (training) and target (unseen application) domains. This distributional discrepancy invariably leads to substantial performance degradation. Formally, we define the problem as follows:

Given a labeled source domain $\mathcal{D}_s = \{F_s, Y_s\} = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$, where $F_s \subset \mathbb{R}^d$ is the feature space and Y_s the label space, and an unlabeled (unseen) target domain $\mathcal{D}_t = \{F_t, Y_t\}$. The source domain comprises samples $\{x_i^s \in X_s\}$ and the target domain $\{x_i^t \in X_t\}$. Since Y_t is unknown and unseen, it is approximated through clustering-generated pseudo-labels, where target samples are partitioned into C_t clusters, yielding $\{y_i^t \in Y_t\}$ for $\{x_i^t \in X_t\}$. Our objective is to learn a mapping function $f : F_s \to F_t$ that minimizes domain shift while preserving discriminative features essential for Re-ID.

Two types of noise affect this process: (1) Inherent Noise from limited cross-domain transferability, and (2) Clustering Noise from imperfect pseudo-label assignments. To address these challenges, we propose a two-stage methodology. The initial stage focuses on source domain pre-training, where a deep neural network architecture is trained on the source domain to learn robust feature representations. The subsequent stage encompasses pseudo-label generation and optimization, wherein the model undergoes fine-tuning on the target domain. This work primarily focuses on the second stage, with particular emphasis on addressing the two types of noise through a dual-step approach: first, developing formal definitions and identification mechanisms for different noise types, and subsequently, implementing targeted strategies for noise mitigation.

3.2. Network architecture

To extract more discriminative features for the domain adaptive tasks, we propose ResNet50-pro based on ResNet50 [34]. We modify the last convolutional block by setting its stride to 1 and adding a downsampling layer to increase spatial resolution. In layer 3, we implement dual feature extraction paths for high-level and low-level features, thereby enhancing the spatial resolution of feature maps and preserving fine-grained spatial information critical for person Re-ID.

In the high-level feature branch, we integrate a non-local block post layer 3:

$$Y_i = \frac{1}{C(X)} \sum_{\forall j} f(X_i, X_j) \cdot g(X_j) \tag{1}$$

where Y_i enhances position i using information from the entire feature map. This operation effectively captures global structural information and long-range dependencies in the feature space. We replace the Global Average Pooling (GAP) with Generalized Mean Pooling (GMP), followed by a fully connected (FC) layer, batch normalization (BN), and a MemoryBank module [35].

The low-level branch applies GAP, a linear layer, and BN to match the high-level feature dimension. We implement independent classifiers for each feature type, enabling multi-task learning that captures complementary feature characteristics.

Building on [9], we implement a Mutual Mean-Teaching (MMT) framework with dual student–teacher networks. Both student networks ($f(x;\theta_s)$) and teacher networks ($f(x;\theta_t)$) share the ResNet50-pro architecture but with different initialization. Feature extraction occurs in both domains:

$$\begin{aligned} z_s^{\mathrm{A,B}} &= f(x;\theta_s^{\mathrm{A,B}}) \in \mathbb{R}^d, \quad x \in \mathcal{D}_s \\ z_t^{\mathrm{A,B}} &= f(x;\theta_s^{\mathrm{A,B}}) \in \mathbb{R}^d, \quad x \in \mathcal{D}_t \end{aligned} \tag{2}$$

where d denotes the feature dimension.

We employ DBSCAN clustering for target domain pseudo-labels and update student networks through cross-network learning:

$$\begin{aligned} \theta_s^{\mathrm{A,new}} &= \mathrm{Update}(\theta_s^{\mathrm{A}}, \mathrm{PseudoLabels}^{\mathrm{B}}) \\ \theta_s^{\mathrm{B,new}} &= \mathrm{Update}(\theta_s^{\mathrm{B}}, \mathrm{PseudoLabels}^{\mathrm{A}}) \end{aligned} \tag{3}$$

where PseudoLabels is derived from temporal ensemble models. Teacher network parameters evolve via Exponential Moving Average (EMA):

$$\theta_t^{\mathrm{A,B}} \leftarrow \alpha_m \theta_t^{\mathrm{A,B}} + (1-\alpha_m)\theta_s^{\mathrm{A,B}} \tag{4}$$

where α_m is the ensembling momentum, set to 0.999 in this paper.

Fig. 3 provides a comprehensive visualization of our framework. The architecture operates in two distinct stages: initial source domain pre-training (green section) and subsequent domain adaptation (blue section). The yellow section delineates our proposed noise optimization and dynamic weighting modules, detailed in subsequent sections.

3.3. Silhouette-coefficient-based pseudo-label reliability assessment

While traditional clustering evaluation metrics such as purity and completeness [36] effectively assess cluster quality, they require ground truth labels unavailable for the target domain (unseen test set). To address this, we propose a label quality assessment framework based on the silhouette coefficient (SC) [37] combined with a DBSCAN clustering algorithm.

The SC metric simultaneously evaluates two critical aspects of clustering quality: intra-cluster variance (ICV) and inter-cluster separation (ICS). For any point x, we define SC as:

$$SC(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \tag{5}$$

where $a(x)$ quantifies ICV through the mean intra-cluster distance:

$$a(x_{i,j}) = \frac{1}{|C_x| - 1} \sum_{x_{i,j} \in C_x, i \neq j} d(x_i, x_j) \tag{6}$$

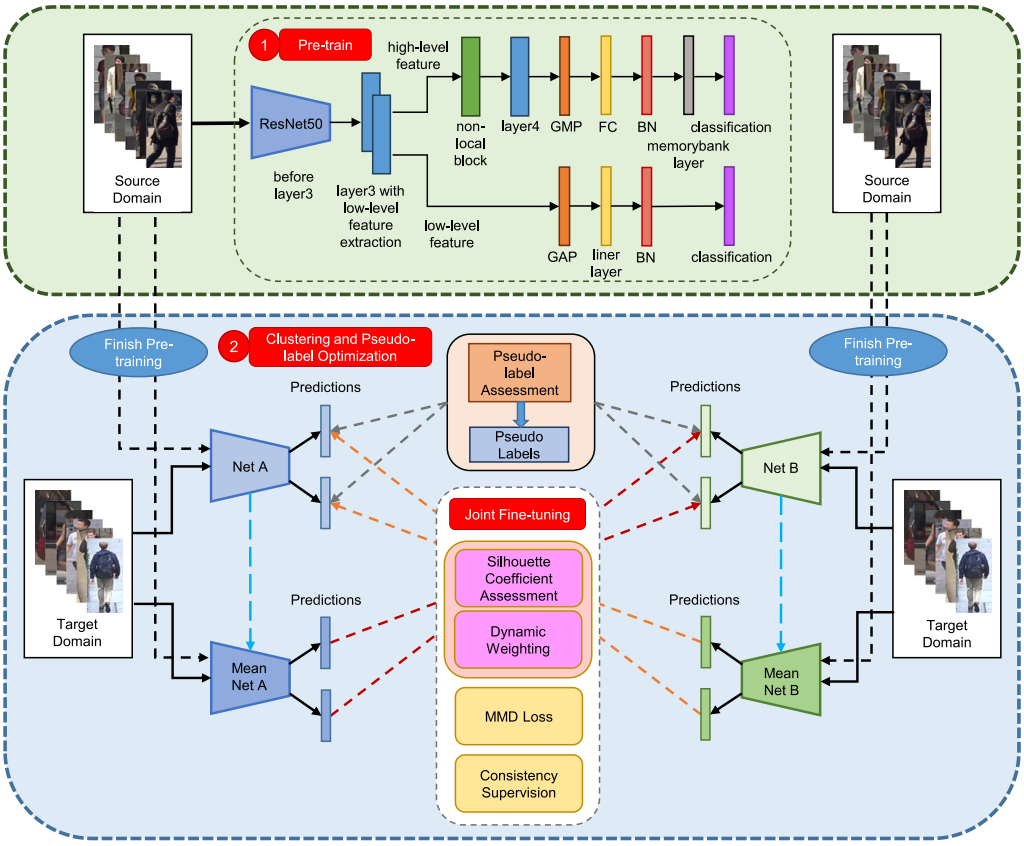


Fig. 3. The system architecture of the proposed domain adaptive person Re-ID framework. Stage 1 (green) shows the enhanced ResNet50-pro backbone with dual-path feature extraction. Stage 2 (blue) depicts the MMT framework with dual student–teacher networks, where proposed modules handle noise optimization and dynamic weighting for robust domain adaptation.

Here, $x_{i,j}$ represents points in cluster C_x, and $d(x_i, x_j)$ denotes the Euclidean distance. In our DBSCAN implementation, we specifically consider core points to minimize outlier influence, as clusters are formed based on density characteristics.

Conversely, $b(x)$ quantifies ICS, representing dissimilarity between clusters:

$$b(x) = \min_{C \neq C_x} \left(\frac{1}{|C|} \sum_{y \in C} d(x, y) \right) \tag{7}$$

where x and y are core points in clusters C_x and C respectively. $b(x)$ is the minimum average distance from x to points in the nearest distinct cluster. This formulation ensures robust measurement of cluster separation by considering only density-connected core points.

For noise points identified by DBSCAN, we assign $SC(x) = -1$, explicitly marking their outlier status. The resulting SC scores range from -1 to 1. This approach ensures that the $SC(x)$ accurately reflects the density and separation in DBSCAN, offering a robust measure of ICV and ICS. A score close to 1 indicates well-clustered data points, with $a(x)$ significantly smaller than $b(x)$, suggesting proximity to their cluster rather than neighboring clusters. A score near 0 implies closeness to the decision boundary between clusters, while a score near -1 suggests possible misassignment to the wrong cluster. By calculating $SC(x)$, we can refine pseudo-labels more accurately. Data points with $SC(x) > 0$ are considered reliable, while those with $SC(x) < 0$ are deemed unreliable.

3.4. Dynamic weighting strategy

Building upon our Silhouette-Coefficient-based assessment proposed in Section 3.3, to further manage noise, we propose a dynamic weighting strategy to modulate pseudo-label influence throughout training. Unlike conventional methods that discard unreliable pseudo-labels, our method preserves all samples while adjusting their training impact based on reliability scores and training progression.

For a pseudo-label y_i associated with samples $X_i = \{x_i^1, x_i^2, \ldots, x_i^m\}$, we define the dynamic weight $w(t)_i$ as:

$$w(t)_i = \begin{cases} \frac{1-e^{-\alpha t}}{1-e^{-\alpha t_m}}, & \text{if } SC(x_i) < 0, \\ \frac{SC(x_i)+1}{2}, & \text{if } SC(x_i) > 0 \end{cases} \tag{8}$$

where t denotes the current epoch, t_m is the maximum epoch and α controls the weight adjustment rate. In our paper, optimal α values are task-dependent: $\alpha = 1.0$ for Market to Duke, $\alpha = 2.0$ for Duke to Market, and $\alpha = 3.0$ for tasks involving MSMT17. This part is demonstrated in our ablation studies.

The algorithmic flow is illustrated in Algorithm 1.

Algorithm 1 presents the complete optimization procedure. The key innovation lies in the progressive adaptation of sample weights: unreliable samples initially contribute minimally to model updates, with their influence gradually increasing as the model's feature representation improves. This approach ensures robust learning while maintaining comprehensive sample coverage, as validated by our experimental results showing consistent performance improvements across all evaluation metrics.

3.5. MMD-based assessment and optimization module

Beyond the clustering noise discussed in Section 3.3, another significant source of noise in domain adaptive Re-ID is the inherent noise due to limited transferability between domains. To address this problem, we propose a specialized assessment and optimization module that quantifies and minimizes this domain-specific noise through an adapted

Algorithm 1: Progressive Pseudo-Label Optimization with Reliability-based Dynamic Weighting

Input: Source domain data $\mathcal{D}_s$, Target domain data $\mathcal{D}_t$, Maximum Epoch t_m, Sharpness factor α
Output: Trained Model $\mathcal{M}$

Initialize model $\mathcal{M}$ with source domain data $\mathcal{D}_s$
for $t = 1$ **to** t_m **do**
  Step 1: Feature Learning
  Extract features for target domain: $\{f(x_i^t)\}_{i=1}^{N_t}$ where $x_i^t \in \mathcal{D}_t$
  Step 2: Pseudo-Label Generation
  Cluster the target domain features $\{f(x_i^t)\}$ to generate pseudo-labels $\{y_i\}_{i=1}^{N_t}$
  Step 3: Pseudo-Label Classification and Weight Assignment
  for *each pseudo-label* y_i **do**
    Compute credibility score $SC(x_i)$ for instances X_i corresponding to y_i
    if $SC(x_i) < 0$ **then**
      Assign weight: $w(t)_i = \frac{1-e^{-\alpha t}}{1-e^{-\alpha t_m}}$
    else
      Assign weight: $w(t)_i = \frac{SC(x_i)+1}{2}$
    end
  end
  Step 4: Model Training Update
  Update model $\mathcal{M}$ using weighted pseudo-labels:

$$\mathcal{L} = \sum_{i=1} w(t)_i \cdot \mathcal{L}(\mathcal{M}(x_i), y_i)$$

end

Maximum Mean Discrepancy (MMD) [38] framework. Our adaptation focuses specifically on person Re-ID feature spaces, where the domain shift manifests in identity-relevant characteristics.

We quantify domain shift through statistical distribution analysis in the feature embedding space. Given feature extractors $F_s(\cdot)$ and $F_t(\cdot)$ mapping to identity-sensitive representations, we initially formulate the MMD as:

$$MMD(D_s, D_t) = \left\| \frac{1}{N_s}\sum_{i=1}^{N_s} F_s(x_i^s) - \frac{1}{N_t}\sum_{j=1}^{N_t} F_t(x_j^t) \right\| \tag{9}$$

This equation can measure distribution dissimilarity by calculating the difference between the means of feature representations in the two domains, providing a robust estimate of domain shift. A smaller MMD value suggests less domain shift and greater transferability. To further analyze domain discrepancies, inspired by [38], we extend Eq. (9) to a reproducing kernel Hilbert space (RKHS) representation through the Kullback–Leibler (KL) divergence:

$$MMD(p_s(x^s), p_t(x^t)) = \left\| \mathbb{E}_{x^s \sim p_s(x^s)}[\phi(x^s)] - \mathbb{E}_{x^t \sim p_t(x^t)}[\phi(x^t)] \right\|_{\mathcal{H}}^2 \tag{10}$$

For now, Eq. (10) is calculated from samples, and as the sample size approaches infinity, the empirical estimate converges to the theoretical MMD. Expanding Eq. (10), we have:

$$\begin{aligned} \widehat{MMD}^2 &= \left(\frac{1}{N_s}\sum_{i=1}^{N_s} \phi(x_i^s) - \frac{1}{N_t}\sum_{j=1}^{N_t} \phi(x_j^t) \right)^\top \left(\frac{1}{N_s}\sum_{i=1}^{N_s} \phi(x_i^s) - \frac{1}{N_t}\sum_{j=1}^{N_t} \phi(x_j^t) \right) \\ &= \left\| \frac{1}{N_s}\sum_{i=1}^{N_s} \phi(x_i^s) \right\|_{\mathcal{H}}^2 + \left\| \frac{1}{N_t}\sum_{j=1}^{N_t} \phi(x_j^t) \right\|_{\mathcal{H}}^2 \\ &\quad - 2\left(\frac{1}{N_s}\sum_{i=1}^{N_s} \phi(x_i^s) \right)^\top \left(\frac{1}{N_t}\sum_{j=1}^{N_t} \phi(x_j^t) \right) \end{aligned} \tag{11}$$

For computational efficiency in high-dimensional Re-ID feature spaces, we develop a kernel-based formulation:

$$\begin{aligned} \widehat{MMD}^2 &= \frac{1}{N_s^2}\sum_{i=1}^{N_s}\sum_{i'=1}^{N_s} k(x_i^s, x_{i'}^s) \\ &\quad + \frac{1}{N_t^2}\sum_{j=1}^{N_t}\sum_{j'=1}^{N_t} k(x_j^t, x_{j'}^t) \\ &\quad - \frac{2}{N_s N_t}\sum_{i=1}^{N_s}\sum_{j=1}^{N_t} k(x_i^s, x_j^t) \end{aligned} \tag{12}$$

where $k(x, y) = \exp(-\|x - y\|^2/2\sigma^2)$ is our chosen Gaussian kernel. This choice ensures stability via Mercer's theorem [39] while capturing nonlinear relationships crucial for person Re-ID. And based on the Law of Large Numbers [40] and Hoeffding's inequality [41], our tailored $\widehat{MMD}^2$ from Eq. (12) converges to $MMD(p_s(x^s), p_t(x^t))$ from Eq. (10).

The MMD estimate between $\mathcal{D}_s$ and $\mathcal{D}_t$ is theoretically bounded by a probabilistic inequality. For finite samples of size n drawn from each domain, with probability at least $1 - \delta$, the following inequality holds:

$$|\mathrm{MMD}^2(\mathcal{D}_s, \mathcal{D}_t) - \widehat{\mathrm{MMD}}^2| \le 2\sqrt{\frac{\log(2/\delta)}{n}} \tag{13}$$

This bound has significant implications for batch size selection in domain adaptation. Firstly, the estimation error decreases at a rate of $O(1/\sqrt{n})$. Secondly, larger batch sizes result in tighter bounds on the true MMD. For a Gaussian kernel defined as $k(x, y) = \exp(-\|x - y\|^2/2\sigma^2)$, the MMD's discriminative power is further bounded by:

$$\mathrm{MMD}(P_s, P_t) \le \frac{1}{\sigma}\sqrt{\mathbb{E}_{x \sim P_s, y \sim P_t}[\|x - y\|^2]} \tag{14}$$

Through extensive experiments (Section 4.6.2), the optimal σ is 1.0 and the recommended batch size is 128.

During training, we compute gradients and update model parameters using limited samples. Thus, the Eq. (12) transforms to a form of MMD loss function:

$$\begin{aligned} \mathcal{L}_{\mathrm{MMD}} = \widehat{MMD}^2 &= \frac{1}{N_s^2}\sum_{i=1}^{N_s}\sum_{i'=1}^{N_s} k(x_i^s, x_{i'}^s) \\ &\quad + \frac{1}{N_t^2}\sum_{j=1}^{N_t}\sum_{j'=1}^{N_t} k(x_j^t, x_{j'}^t) \\ &\quad - \frac{2}{N_s N_t}\sum_{i=1}^{N_s}\sum_{j=1}^{N_t} k(x_i^s, x_j^t) \end{aligned} \tag{15}$$

This MMD loss ensures theoretical and practical consistency, effectively reducing distribution discrepancies between source and target domains. By minimizing the intrinsic noise due to limited transferability, it enhances domain adaptive Re-ID performance. The proposed Eq. (15) is employed during the MMT label optimization and training process in the second stage.

3.6. Consistency-supervised module

In MMT frameworks for domain adaptive learning, prediction discrepancies between the student and teacher models occur when processing perturbed versions of the same input. To minimize these discrepancies, the student model's predictions must align with the teacher model's more stable outputs across various augmentations of the same input during training. We address this by introducing a class center P_g, which leverages data distribution from both the source and target domains to generate more reliable pseudo-labels. The class center P_g is defined as:

$$P_g = C_s + C_t \tag{16}$$

where C_s and C_t are the class centers for the source and target domains, respectively. Combining these centers enhances the pseudo-label

generation process. By calculating the similarity between target domain samples and P_g, we generate soft pseudo-labels to improve label quality. The soft pseudo-label $\tilde{p}_i$ is generated:

$$\tilde{p}_i = \frac{P_g \cdot f(x_i)}{\|P_g\| \|f(x_i)\|} \tag{17}$$

where $\tilde{p}_i$ indicates the model's confidence in assigning sample x_i to a particular class. A higher similarity between $\tilde{p}_i$ and P_g results in increased confidence.

To further enforce prediction consistency, we utilize KL divergence to measure discrepancies between the student and teacher models' outputs on perturbed inputs:

$$\mathcal{L}_c = \frac{1}{N} \sum_{i=1}^{N} D_{KL}(f(x_i'; \theta_s) \parallel f(x_i'; \theta_t)) \tag{18}$$

where x_i' represents the perturbed input, and $f(x_i'; \theta_s)$ and $f(x_i'; \theta_t)$ denote the output probability distributions of the student and teacher models, respectively. The soft pseudo-label $\tilde{p}_i$ guides the student model's training by ensuring alignment with the pseudo-labels. We define the pseudo-label loss as follows:

$$\mathcal{L}_p = \frac{1}{N} \sum_{i=1}^{N} D_{KL}(\hat{p}_i \parallel \tilde{p}_i) \tag{19}$$

where $\hat{p}_i$ is the prediction by the student model. This loss measures the divergence between the student model's predictions and the soft pseudo-labels. This soft labeling approach allows for uncertainty in the pseudo-label assignments, making the system more robust to noise.

During the optimization stage, we introduce uncertainty into the triplet loss for consistency supervision. First, we calculate the uncertainty of each sample:

$$u_c = D_{KL}(\tilde{p}_i \parallel p_i) = \sum_{k=1}^{K_t} \tilde{p}_{i,k} \log \frac{\tilde{p}_{i,k}}{p_{i,k}} \tag{20}$$

Following [42], the uncertainty modulation for the triplet loss is defined as:

$$\begin{aligned} u_c^{ap_i} &= \frac{1}{e^{u_c^a}} + \frac{1}{e^{u_c^{p_i}}} \\ u_c^{an_i} &= \frac{1}{e^{u_c^a}} + \frac{1}{e^{u_c^{n_i}}} \end{aligned} \tag{21}$$

where $u_c^{ap_i}$ and $u_c^{an_i}$ represent the uncertainty associated with positive and negative samples relative to the anchor. The uncertainty-modulated triplet loss is given by:

$$\mathcal{L}_{\text{tri}} = -\frac{1}{N} \sum_{i=1}^{N} \log \left(\frac{e^{-u_c^{ap_i} \cdot D(a,p)}}{e^{-u_c^{ap_i} \cdot D(a,p)} + e^{-u_c^{an_i} \cdot D(a,n)}} \right) \tag{22}$$

In this equation, softmax normalization is applied to the Euclidean distances $D(a, n)$ and $D(a, p)$ of negative and positive sample pairs, respectively. This process smooths the loss function, ensuring that each sample contributes a gradient, thereby effectively transmitting loss information.

3.7. Loss functions

In our method, each module plays a critical role in improving the model's performance to adapt to noisy and domain-shifted data. This Section rigorously examines the relationships between modules from Section 3.3, 3.4, 3.5, and 3.6, and how they collectively contribute to the overall loss function optimization.

Firstly, it is the core losses in our method: the cross-entropy loss $\mathcal{L}_{\text{ce}}^t$ for classification and the triplet loss $\mathcal{L}_{\text{tri}}^t$ for metric learning. The cross-entropy loss $\mathcal{L}_{\text{ce}}^t$ is defined as:

$$\mathcal{L}_{\text{ce}}^t = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} \mathbf{y}_{\text{smooth},i,j} \cdot \log \left(\frac{\exp(x_{i,j})}{\sum_{k=1}^{C} \exp(x_{i,k})} \right) \tag{23}$$

where N is the batch size, $\mathbf{y}_{\text{smooth},i,j}$ represents smoothed labels, corresponding to true labels for source domain data and pseudo-labels for target domain data, and x denotes the sample instances. The triplet loss $\mathcal{L}_{\text{tri}}^t$ is expressed as:

$$\mathcal{L}_{\text{tri}}^t = -\frac{1}{N_t} \sum_{i=1}^{N_t} \log \left(\frac{e^{-D(a,p)}}{e^{-D(a,p)} + e^{-D(a,n)}} \right) \tag{24}$$

where $D(a, n)$ and $D(a, p)$ are the Euclidean distances between the anchor-negative and anchor-positive pairs, respectively. These two losses, $\mathcal{L}_{\text{ce}}^t$ and $\mathcal{L}_{\text{tri}}^t$, also constitute the complete loss function for Stage One (source domain pre-training).

Except for the two core losses, based on Section 3.6, we also employ an uncertainty-modulated triplet loss $\mathcal{L}_{\text{tri}}^t$, which is defined as:

$$\mathcal{L}_{\text{tri}_{uc}}^t = -\frac{1}{N} \sum_{i=1}^{N} \log \left(\frac{e^{-u_c^{ap_i} \cdot D(a,p)}}{e^{-u_c^{ap_i} \cdot D(a,p)} + e^{-u_c^{an_i} \cdot D(a,n)}} \right) \tag{25}$$

where $u_c^{ap_i}$ and $u_c^{an_i}$ are uncertainty factors from Eq. (21) for positive and negative samples, respectively. The uncertainty modulation factor adjusts the triplet loss based on the confidence of pseudo-label assignments, paying more attention to more reliable triplets while down-weighting uncertain ones.

The Dynamic Weighting adjusts the influence of pseudo-labels throughout the training process based on the Eq. (8). During early training, when pseudo-labels are less reliable, their weights are reduced, allowing the model to focus more on the reliable labels. The MMD loss from Eq. (15) directly addresses the domain shift by minimizing the distribution discrepancy between the source and target domains. This loss ensures that the feature representations of both domains are aligned in a shared space, which is essential for the model to generalize well on the target domain. The consistency-supervised loss operates on two levels: the consistency loss from Eq. (18) and the pseudo-label consistency loss from Eq. (19). Consistency supervision ensures robust feature learning and stable pseudo-label generation.

The final loss function $\mathcal{L}_{\text{total}}$ integrates all components with carefully chosen weights:

$$\mathcal{L}_{\text{total}} = \sum_{i=1} w(t)_i \left(\gamma_1 \mathcal{L}_{\text{ce}}^t + \gamma_2 \mathcal{L}_{\text{tri}}^t + \gamma_3 \mathcal{L}_{\text{tri}_{uc}}^t + \gamma_4 (\mathcal{L}_c + \mathcal{L}_p) + \gamma_5 \mathcal{L}_{\text{MMD}} \right) \tag{26}$$

where γ_1, γ_2, γ_3, γ_4, and γ_5 are weights balancing these loss functions, and $w(t)_i$ denotes dynamic weight adjustment during training.

4. Experiments

4.1. Datasets

To rigorously evaluate our proposed method, we conducted extensive experiments on three widely-adopted person Re-ID benchmarks: Market1501 [20], DukeMTMC-reID [43], and MSMT17 [26]. These datasets represent increasing levels of complexity and scale in person Re-ID challenges, offering diverse evaluation scenarios that closely approximate real-world surveillance applications. Table 1 presents a comprehensive overview of these benchmarks' characteristics.

These datasets collectively present a comprehensive evaluation framework, encompassing variations in illumination, viewpoint, background complexity, occlusion patterns, and environmental conditions. Such diversity is essential for validating the generalization capability and practical applicability of person Re-ID algorithms in real-world deployment scenarios.

4.2. Metrics

Followed by existing work [9,10,12], Cumulative Matching Characteristics (CMC) [44] and mean Average Precision (mAP) [20,45] are employed as objective evaluation metrics to evaluate the performance

Table 1
Summary of Three Person Re-identification Benchmarks.

Dataset	Market1501	DukeMTMC-reID	MSMT17
Number of Cameras	6	8	15 (12 outdoor, 3 indoor)
Total Images	32,668	36,411	126,441
Number of Persons	1,501	1,110	4,101
Training Images	12,936	16,522	32,621
Training Persons	751	702	1,041
Testing Images	19,732	19,889	93,820
Testing Persons	750	702	3,060
Query Images	3,368	2,228	11,659
Gallery Images	16,364	17,661	82,161
Additional Challenges	Multiple camera perspectives	Camera view and background diversity	Camera intra- and inter-variations

of our proposed methods. mAP measures the mean of the Average Precision (AP) across all queries. For a single query i, the AP is calculated as:

$$\mathrm{AP}_i = \frac{1}{m_i} \sum_{j=1}^{n_i} P(j) \cdot \mathbb{1}(\mathrm{relevant}_j) \tag{27}$$

where m_i is the number of relevant instances for query i, n_i is the total number of instances retrieved, and $P(j)$ is the precision at rank j. $\mathbb{1}(\cdot)$ is a indicator function. Then the mAP is then given by:

$$\mathrm{mAP} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{AP}_i \tag{28}$$

where N is the total number of queries. In our paper, the CMC is represented as the probability that a correct match appears within the top k ranks, where $k \in \{1, 5, 10\}$. If R_i is the rank of the first correct match for the ith query, the Rank-k accuracy is defined as:

$$\text{Rank-}k = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(R_i \leq k) \tag{29}$$

where $\mathbb{1}(\cdot)$ is the indicator function, which equals 1 if $R_i \leq k$, and 0 otherwise, and N is the total number of queries. These metrics are widely recognized and utilized within the field of Re-ID.

4.3. Implementation details

Our implementation framework consists of two distinct stages: source domain pre-training and domain adaptive learning. We detail the technical specifications, hyperparameters, and computational requirements for reproducibility.

All experiments were conducted using the PyTorch framework on a single NVIDIA RTX 4090 GPU with CUDA 12.0. We employed ResNet50-pro architecture initialized with ImageNet pre-trained weights (detailed in Section 3.2). Input images underwent standardized preprocessing: resizing to 256 × 128 pixels followed by data augmentation comprising random horizontal flipping, cropping with 10-pixel padding, and random erasing. We maintained a consistent batch size of 128 across all experiments, with each identity represented by 4 instances to ensure stable mini-batch sampling.

In the first stage of source domain pre-training, we employed the Adam optimizer with an initial learning rate of 1×10^{-3}. The learning rate schedule incorporated a 10-epoch warm-up period, with decay points implemented at epochs 40 and 70 to ensure optimal convergence. For the second stage of domain adaptive learning, we utilized the DBSCAN clustering algorithm with parameters eps = 0.6 and min_samples = 4. The optimization process continued with the Adam optimizer, maintaining the initial learning rate of 1×10^{-3} and a momentum of 0.9. The loss function weights were empirically determined as $\gamma_1 = 1.0$, $\gamma_2 = 1.0$, $\gamma_3 = 0.3$, $\gamma_4 = 0.3$, and $\gamma_5 = 0.5$, with training extending over 80 epochs.

The implementation follows a systematic two-stage approach. During the initial source domain training phase, we employ supervised learning on the source domain data $\mathcal{D}_s = \{F_s, Y_s\}$ to obtain a pre-trained model $F(\cdot \mid \theta)$ capable of discriminative feature extraction. The subsequent domain adaptive learning phase begins with feature extraction from the target domain $\mathcal{D}_t = \{F_t\}$, followed by DBSCAN clustering for hard pseudo-label generation ($\tilde{y}_i^t$). These pseudo-labels undergo quality assessment using Eq. (5), with dynamic weights assigned according to Eq. (8). The process concludes with the implementation of the MMT framework, utilizing the loss function defined in Eq. (26).

Table 2
Model Complexity and Resource Requirements Analysis.

Architecture	Parameters	Memory	FLOPs
Base (ResNet-50)	25.69M	102.76 MB	6.01G
Enhanced (ResNet-50-pro)	38.53M	154.12 MB	7.71G

Table 2 provides a detailed analysis of model complexity and computational requirements. The teacher–student framework employs identical architectures for both networks, with the teacher network maintained as an EMA of the student network. This configuration results in a total parameter count of 77.06M and approximately 308.24 MB memory requirement, effectively balancing computational efficiency with model performance.

The whole implementation process is shown in Algorithm 2:

4.4. Comparison with the state-of-the-art methods

To rigorously evaluate our proposed method, we conducted comprehensive comparisons with state-of-the-art (SOTA) methods across four challenging cross-domain person Re-ID tasks: Market to Duke, Duke to Market, Market to MSMT, and Duke to MSMT. Tables 3 and 4 present detailed performance comparisons using standard evaluation metrics.

Table 3 demonstrates our method's superior performance on bidirectional transfer tasks between Market1501 and DukeMTMC datasets. In the Market to Duke task, our method achieves SOTA performance with mAP of 73.8% and Rank-1/5/10 accuracies of 85.2%/93.1%/95.3%, respectively. These results represent significant improvements over the previous best method (AWB [48]), with gains of 2.9% in mAP, 1.4% in Rank-1, 1.8% in Rank-5, and 1.3% in Rank-10 accuracy.

For the Duke to Market task, our method demonstrates even more compelling performance, achieving 84.7% mAP and 94.6%/97.9%/99.1% in Rank-1/5/10 metrics. This represents substantial improvements over previous SOTA results across multiple metrics: surpassing SECRET [15] by 1.8% in mAP, AWB [48] by 1.1% in Rank-1, and DREAMT [51] by 1.9% and 0.4% in Rank-5 and Rank-10, respectively.

To further validate our method's generalization capability, we conducted experiments on more challenging transfer tasks involving the MSMT17 dataset, which presents additional complexities due to its larger scale and greater environmental variations. As shown in Table 4, our method maintains its superior performance in these more demanding scenarios.

Algorithm 2: The implementation process of our proposed method

Input: source domain set $D_s = \{F_s, Y_s\}$, target domain set $D_t = \{F_t\}$
Output: $F(\cdot \mid \theta)$

The first stage: pre-training on source domain
Input : Source domain set $D_s = \{F_s, Y_s\}$
Output: Pre-trained model $F(\cdot \mid \theta)$
1. Initialize the model;
2. Define the loss function according to Eqs. (23) and (24);
3. Train the initialized model on the source domain through supervised learning;
4. Finish training and obtain the pre-trained model with discriminative features $F(\cdot \mid \theta)$;

The second stage: domain-adaptive learning
Input : Pre-trained model $F(\cdot \mid \theta)$, target domain set $D_t = \{F_t\}$
Input : High robust domain-adaptive person re-identification model
foreach *Iterate* $k = 1, 2, \ldots, K$ **do**
1. Extract features F_t from target domain $D_t = \{F_t\}$ through $F(\cdot \mid \theta)$;
2. Cluster the extracted features F_t and generate hard pseudo-labels $\tilde{y}_i^t$ through DBSCAN;
3. Build new dataset $D_t' = \{F_t, \tilde{y}_i^t\}$;
4. Assess the generated pseudo-labels $\tilde{y}_i^t$ according to Eq. (5);
5. Set the dynamic weighting ω_i according to the assessment result in step 4 and Eq. (8);
6. Generate soft pseudo-labels $\hat{y}_i^t$ and optimize them based on MMT;
7. Define loss functions according to Eq. (26);
8. Update and fine-tune the model $F(\cdot \mid \theta)$ by epochs;

Finally, a highly robust domain adaptive person re-identification model is trained $F(\cdot \mid \theta)$;

For the Market to MSMT task, our method achieves 34.2% mAP and 65.8%/75.5%/79.3% in Rank-1/5/10 accuracy, substantially outperforming the previous best method (P2LR [47]) by margins of 4.3%, 4.8%, 2.4%, and 1.4% across respective metrics. Similarly, in the Duke to MSMT task, our method demonstrates remarkable results with 35.6% mAP and 66.5%/77.8%/80.6% in Rank-1/5/10 metrics. These results represent significant improvements over previous SOTA methods: surpassing DREAMT by 5.3% in mAP and AWB [48] by 5.5%, 4.3%, and 2.7% in Rank-1/5/10 accuracy, respectively.

The consistent performance advantages across all transfer scenarios, particularly in the more challenging MSMT17-based tasks, validate our method's robust generalization capability and practical applicability in complex cross-domain person Re-ID scenarios.

4.5. Ablation study

4.5.1. Ablation study for proposed modules

To evaluate the contribution of each proposed component, we conducted comprehensive ablation studies. Using MMT with ResNet50 as our baseline [9], we progressively incorporated our proposed modules:

- **B (w/ ResNet50)**: Baseline implementation with standard ResNet50 backbone.
- **B (w/ ResNet50-pro)**: Optimized backbone architecture (Section 3.2)
- **SC&DW**: Silhouette-coefficient-based assessment and dynamic weighting module (Eqs. (5), (8))
- **MMD**: MMD-based module (Section 3.5 and Eq. (15))
- **C**: Consistency-supervised module (Eqs. (18), (19), (22))

Tables 5 and 6 present the results on Market to Duke and Market/Duke to MSMT tasks, respectively.

From Table 5, the ResNet50-pro backbone demonstrates clear advantages over the baseline ResNet50. For Market to Duke, it improves mAP from 64.1% to 66.2% (+2.1%) and Rank-1 from 77.8% to 78.8% (+1.0%). Similar gains are seen in Duke to Market, with mAP increasing from 71.6% to 73.8% (+2.2%) and Rank-1 from 87.3% to 88.9% (+1.6%). Adding the SC&DW module further enhances performance. On Market to Duke, mAP improves to 68.6% (+2.4%) and Rank-1 to 81.5% (+2.7%). The Duke to Market task shows similar improvements with mAP reaching 76.8% (+3.0%) and Rank-1 achieving 90.2% (+1.3%). The incorporation of MMD yields substantial gains across both tasks. For Market to Duke, mAP increases to 72.6% (+4.0%) and Rank-1 to 84.7% (+3.2%). In Duke to Market, we observe improvements to 82.1% (+5.3%) in mAP and 93.3% (+3.1%) in Rank-1. Finally, the consistency supervision module completes our framework, achieving the best performance with mAP/Rank-1 scores of 73.8%/85.2% for Market to Duke and 84.7%/94.6% for Duke to Market, representing additional improvements of 1.2%/0.5% and 2.6%/1.3% respectively.

Table 6 further validates these findings on more challenging Market/Duke to MSMT tasks. The complete model achieves substantial improvements over the baseline, with final mAP/Rank-1 scores of 34.2%/65.8% for Market to MSMT17 and 35.6%/66.5% for Duke to MSMT, demonstrating the effectiveness of our proposed components across diverse cross-domain scenarios.

4.5.2. Ablation study on improvement of the backbone

The backbone network is pivotal due to its enhanced feature extraction capabilities, particularly beneficial for person re-identification. Our ResNet50-pro's architecture (Section 3.2) facilitates the representation of complex patterns across domains, increasing the model's robustness and adaptability. This strategic selection leverages advanced convolutional features to capture domain-specific nuances, directly contributing to improved performance. We analyze the impact of three main components within the enhanced backbone on performance:

- **B(w/ ResNet50)**: Uses the original ResNet50 as in [9].
- **LL (Low-Level Layer)**: Adds a low-level feature extraction layer at layer 3 of ResNet50, introducing a dual feature extraction mechanism.
- **non-local**: Implements a non-local block to aggregate information from distant image regions, enhancing contextual understanding.
- **memorybank**: Adopts a memory bank from [35] to store and update feature representations during training.

As shown in Table 7, each component contributes incrementally to performance improvement. For the Market to Duke task, the low-level feature extraction layer provides a 1.2% improvement in mAP (64.1% to 65.3%), demonstrating the importance of multi-scale feature learning. The addition of the non-local block further enhances mAP by 0.7% (65.3% to 66.0%), validating its effectiveness in capturing global contextual information. The memory bank mechanism provides an additional 0.2% improvement (66.0% to 66.2%), resulting in the optimal configuration. Similar patterns are observed in the Duke to Market direction, where the components yield mAP improvements of 0.6%, 1.1%, and 0.5% respectively. The consistent improvements across both transfer directions validate the generalizability of our architectural enhancements.

For the more challenging MSMT17 transfer tasks (Table 8), the improvements become more pronounced. In the Market to MSMT task, the low-level feature extraction layer contributes a substantial 1.1% mAP improvement, while the non-local block and memory bank add 1.0% and 0.3% respectively. The Duke to MSMT results show even larger gains: 0.9%, 1.4%, and 0.3% in mAP for each component.

Notably, the non-local block demonstrates particularly strong performance improvements in MSMT17 tasks, suggesting its effectiveness

Table 3
Comparisons with the SOTA Methods Between Market1501 and DukeMTMC datasets.

Methods	Market to Duke				Duke to Market			
	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
MMCL (CVPR, 2020) [32]	51.4	72.4	82.9	85.0	60.4	84.4	92.8	95.0
AD-Cluster (CVPR, 2020) [31]	54.1	72.6	82.5	85.5	68.3	86.7	94.4	96.5
MMT (ICLR, 2020) [9]	65.1	78.0	88.8	92.5	71.2	87.7	94.9	96.9
MEN-Net(ECCV, 2020) [11]	66.1	79.6	88.3	92.2	76.0	89.9	96.0	97.5
SpCL (NeurIPS,2020) [10]	68.8	82.9	90.1	92.5	76.7	90.3	96.2	97.7
HGA (AAAI, 2021) [46]	67.1	79.4	88.7	90.3	70.3	89.5	93.6	95.5
GCMT (IJCAI, 2021) [30]	67.8	81.1	91.1	93.9	77.1	90.6	96.3	97.7
UNRN (AAAI, 2021) [33]	69.1	82.0	90.7	93.5	78.1	91.9	96.1	97.8
P2LR (AAAI, 2022) [47]	70.8	82.6	90.8	93.7	81.0	92.6	97.4	98.3
SECRET (AAAI, 2022) [15]	68.8	81.7	–	–	82.9	93.1	–	–
AWB (TIP, 2022) [48]	70.9	83.8	92.3	94.0	81.0	93.5	97.4	98.3
MDJL (PR, 2023) [49]	62.8	78.6	86.6	88.7	59.8	80.3	87.4	89.9
FastReID (MM, 2023) [50]	69.2	82.7	–	–	80.5	92.7	–	–
DREAMT (TIM, 2023) [51]	69.8	82.3	90.9	93.6	81.4	93.3	98.0	98.7
Ours	73.8	85.2	93.1	95.3	84.7	94.6	97.9	99.1

Best results are underlined and bold, second-best are underlined only.

Table 4
Comparisons with the SOTA Methods on Market to MSMT and Duke to MSMT Tasks.

Method	Market to MSMT				Duke to MSMT			
	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
MMCL (CVPR, 2020) [32]	15.1	40.8	51.8	56.7	16.2	43.6	54.3	58.9
AD-Cluster (CVPR, 2020) [31]	–	–	–	–	–	–	–	–
MMT (ICLR, 2020) [9]	22.9	49.2	63.1	68.8	23.5	50.0	63.6	69.2
MEN-Net(ECCV, 2020) [11]	–	–	–	–	–	–	–	–
SpCL (NeurIPS,2020) [10]	26.8	53.7	65.0	69.8	26.5	53.1	65.8	70.5
HGA (AAAI, 2021) [46]	25.5	55.1	61.2	65.5	26.8	58.6	64.7	69.2
GCMT (IJCAI, 2021) [30]	24.9	54.8	–	–	26.6	57.9	–	–
UNRN (AAAI, 2021) [33]	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
P2LR (AAAI, 2022) [47]	29.9	60.9	73.1	77.9	29.0	58.8	71.2	76.0
SECRET (AAAI, 2022) [15]	–	–	–	–	–	–	–	–
AWB (TIP, 2022) [48]	29.0	57.3	70.7	75.9	29.5	61.0	73.5	77.9
MDJL (PR, 2023) [49]	13.4	34.3	44.5	50.6	17.1	40.3	51.2	56.3
FastReID (MM, 2023) [50]	26.5	56.6	–	–	27.7	59.5	–	–
DREAMT (TIM, 2023) [51]	25.3	51.6	64.3	69.7	30.3	58.0	70.5	75.3
Ours	34.2	65.8	75.5	79.3	35.6	66.5	77.8	80.6

Best results are underlined and bold, second-best are underlined only.

Table 5
Ablation Study Results on Market to Duke and Duke to Market Tasks.

Components					Market to Duke		Duke to Market	
B(w/ ResNet50)	B(w/ ResNet50-pro)	SC&DW	MMD	C	mAP(%)	Rank-1(%)	mAP(%)	Rank-1(%)
✓	–	–	–	–	64.1	77.8	71.6	87.3
–	✓	–	–	–	66.2	78.8	73.8	88.9
–	✓	✓	–	–	68.6	81.5	76.8	90.2
–	✓	✓	✓	–	72.6	84.7	82.1	93.3
–	✓	✓	✓	✓	**73.8**	**85.2**	**84.7**	**94.6**

The best results are marked in bold.

in handling more complex domain shifts. This aligns with our intuition that long-range dependencies become increasingly important as domain complexity grows. The consistent, albeit smaller, contributions from the memory bank mechanism indicate its role in stabilizing feature learning across all transfer scenarios.

4.5.3. Computational efficiency analysis

Building upon the architectural analysis discussed in Section 3.2, we analyze the computational requirements of our ResNet50-pro model in comparison with contemporary architectures. Table 9 presents a quantitative comparison of model complexity across key metrics.

As shown in Table 9, our model effectively balances model capacity, computational efficiency, and memory usage, making it a versatile solution for practical applications. By strategically enhancing the base ResNet50, our model achieves a 50% increase in parameters with only a moderate 28.3% increase in FLOPs, demonstrating competitive efficiency compared to more resource-intensive architectures like MGN, ViT-Base, and DINO-v2. While maintaining a peak memory consumption of approximately 3-4 GB, ResNet50-pro significantly outperforms lightweight models like OSNet in feature representation and remains comparable to HRNet-W32 in complexity and performance. This efficiency-performance balance is particularly crucial for real-world applications requiring robust feature extraction and domain adaptation capabilities while operating within typical hardware constraints.

4.5.4. Training time analysis

We conducted comprehensive training efficiency analyses using a single NVIDIA RTX 4090 GPU, with consistent batch size 128 and training epochs 80 across all experiments. Tables 10, 11, 12, and 13 present the computational requirements for different model configurations on four cross-domain tasks.

The integration of each module led to incremental increases in training time. As shown in Table 10, for the Market to Duke task,

Table 6
Ablation Study Results on Market/Duke to MSMT Tasks.

Components					Market to MSMT		Duke to MSMT	
B(w/ ResNet50)	B(w/ ResNet50-pro)	SC+DW	MMD	C	mAP(%)	Rank-1(%)	mAP(%)	Rank-1(%)
✓	–	–	–	–	21.9	48.8	23.6	49.7
–	✓	–	–	–	24.3	54.5	26.2	59.4
–	✓	✓	–	–	27.9	58.1	28.6	60.2
–	✓	✓	✓	–	32.3	63.0	34.2	64.9
–	✓	✓	✓	✓	**34.2**	**65.8**	**35.6**	**66.5**

The best results are marked in bold.

Table 7
Ablation Study Results for Backbone on Market to Duke and Duke to Market Tasks.

components				Market to Duke		Duke to Market	
B(w/ ResNet50)	+LL	+non-local	+memorybank	mAP(%)	Rank-1(%)	mAP(%)	Rank-1(%)
✓	–	–	–	64.1	77.8	71.6	87.3
✓	✓	–	–	65.3	78.1	72.2	87.5
✓	✓	✓	–	66.0	78.5	73.3	88.6
✓	✓	✓	✓	**66.2**	**78.8**	**73.8**	**88.9**

The best results are marked in bold.

Table 8
Ablation Study Results for Backbone on Market/Duke to MSMT Tasks.

components				Market to MSMT		Duke to MSMT17	
B(w/ ResNet50)	+LL	+non-local	+memorybank	mAP(%)	Rank-1(%)	mAP(%)	Rank-1(%)
✓	–	–	–	21.9	48.8	23.6	49.7
✓	✓	–	–	23.0	52.2	24.5	53.9
✓	✓	✓	–	24.0	53.9	25.9	58.4
✓	✓	✓	✓	**24.3**	**54.5**	**26.2**	**59.4**

The best Results are marked in bold.

Table 9
Computational Complexity Comparison.

Model	Parameters	FLOPs	Peak Memory[a]
ResNet50 (Base) [34]	25.69M	6.01G	≈ 2 GB
ResNet-IBN-a [52]	28.12M	6.33G	≈ 2.5 GB
OSNet [53]	2.2M	2.0G	≈ 1.5 GB
MGN [54]	70.24M	14.2G	≈ 6 GB
HRNet-W32 [55]	41.23M	8.31G	≈ 4 GB
ViT-Base [56]	86M	17.6G	≈ 8 GB
DINO-v2 [57]	304M	55.4G	≈ 16 GB
ResNet50-pro (Ours)	38.53M	7.71G	≈ 3–4 GB

[a] Peak runtime memory including all computations and buffer.

the SC&DW module added 23.52 s per epoch for silhouette scores and weight adjustments, MMD-based module contributed 13.08 s for kernel computations and distribution alignment, and the consistency-supervised module demanded 62.21 s for maintaining teacher–student networks, generating soft pseudo-labels, computing Kullback–Leibler divergence, and updating global centers.

As shown in Table 11, the Duke to Market showed similar percentages but lower absolute times due to Market's smaller dataset, which is 18.42 s for SC&DW, 10.25 s for MMD-based module, and 48.72 s for Consistency-supervised module.

As shown in Table 12, the Market to MSMT task exhibited larger absolute times due to MSMT17's 2.5-fold larger dataset. SC&DW required 58.80 s, MMD added 32.70 s, and Consistency-supervised module demanded 155.53 s.

Similarly, as shown in Table 13, Duke to MSMT showed intermediate times: SC&DW adding 46.10 s, MMD contributing 25.61 s, and Consistency-supervised module requiring 121.87 s.

Across all four tasks, while absolute times scaled with dataset sizes, percentage increases remained consistent: SC&DW at ≈ 88.4%, MMD-based module at ≈ 26.1%, and Consistency-supervised module at ≈ 98.4%. This pattern indicates linear scaling of computational complexity with dataset size, with the Consistency module consistently being the most computationally intensive addition.

The relative computational overhead (take Market to Duke as an example) can be expressed as:

$$R_{module} = \frac{T_{module}}{T_{baseline}} = \begin{cases} 1.88 & \text{for SC\&DW} \\ 2.38 & \text{for SC\&DW+MMD} \\ 4.71 & \text{for Full Model (SC\&DW+MMD+C)} \end{cases} \quad (30)$$

The epoch time variation also increased significantly from the baseline's 9-s range (24.09 – 33.15 s) to 41 s (107.27 – 148.05 s) in the full model. This variation is primarily due to the overhead from pseudo-label updates, clustering operations, and memory bank updates. While the full model exhibits increased computational requirements (167.26 min total training time), the overhead is justified by significant performance improvements in pseudo-label quality, domain alignment, and overall Re-ID accuracy. The training time remains practical for modern GPU architectures while achieving SOTA performance.

4.5.5. Convergence analysis

We analyze the convergence behavior of our framework across four domain adaptation tasks. The analysis encompasses both the evolution of evaluation metrics and the dynamics of different loss components during training. Fig. 4 shows the convergence patterns of mAP, Rank-1, Rank-5, and Rank-10 accuracy.

For Market to Duke and Duke to Market tasks, we observe rapid initial improvements (epochs 1–15), with mAP increasing from 42.3% to 67.4% (Market to Duke) and 45.6% to 74.5% (Duke to Market). The performance stabilizes during epochs 15–40, followed by fine-tuning until convergence at 73.8% for Market to Duke and 84.7% mAP for Duke to Market respectively. For Market/Duke to MSMT tasks, Fig. 4 demonstrates more gradual convergence due to increased domain complexity, with mAP improving from 15.8% to 34.2% (Market to MSMT) and 16.4% to 35.6% (Duke to MSMT). These tasks exhibit periodic plateaus followed by performance improvements, reflecting the challenges of adapting to MSMT17's diverse conditions.

Table 10
Training Time Analysis on Market to Duke Task.

Model Configuration	Average Time per Epoch (s)	Fastest Epoch (s)	Slowest Epoch (s)	Total Time (min)
Baseline	26.62	24.03	33.15	35.58
+ SC&DW	50.14	47.21	62.36	66.83
+ MMD	63.22	56.08	79.16	84.39
Full Model (SC&DW+MMD+C)	125.43	107.27	148.05	167.26

Table 11
Training Time Analysis on Duke to Market Task.

Model Configuration	Average Time per Epoch (s)	Fastest Epoch (s)	Slowest Epoch (s)	Total Time (min)
Baseline	20.84	18.82	25.96	27.85
+ SC&DW	39.26	36.96	48.82	52.35
+ MMD	49.51	43.92	61.98	66.01
Full Model (SC&DW+MMD+C)	98.23	84.02	115.94	130.97

Table 12
Training Time Analysis on Market to MSMT Task.

Model Configuration	Average Time per Epoch (s)	Fastest Epoch (s)	Slowest Epoch (s)	Total Time (min)
Baseline	66.55	60.08	82.88	88.95
+ SC&DW	125.35	118.03	155.90	167.13
+ MMD	158.05	140.20	197.90	210.73
Full Model (SC&DW+MMD+C)	313.58	268.18	370.13	418.11

Table 13
Training Time Analysis on Duke to MSMT Task.

Model Configuration	Average Time per Epoch (s)	Fastest Epoch (s)	Slowest Epoch (s)	Total Time (min)
Baseline	52.13	47.06	64.91	69.71
+ SC&DW	98.23	92.48	122.13	130.97
+ MMD	123.84	109.86	155.05	165.12
Full Model (SC&DW+MMD+C)	245.71	210.15	290.06	327.61

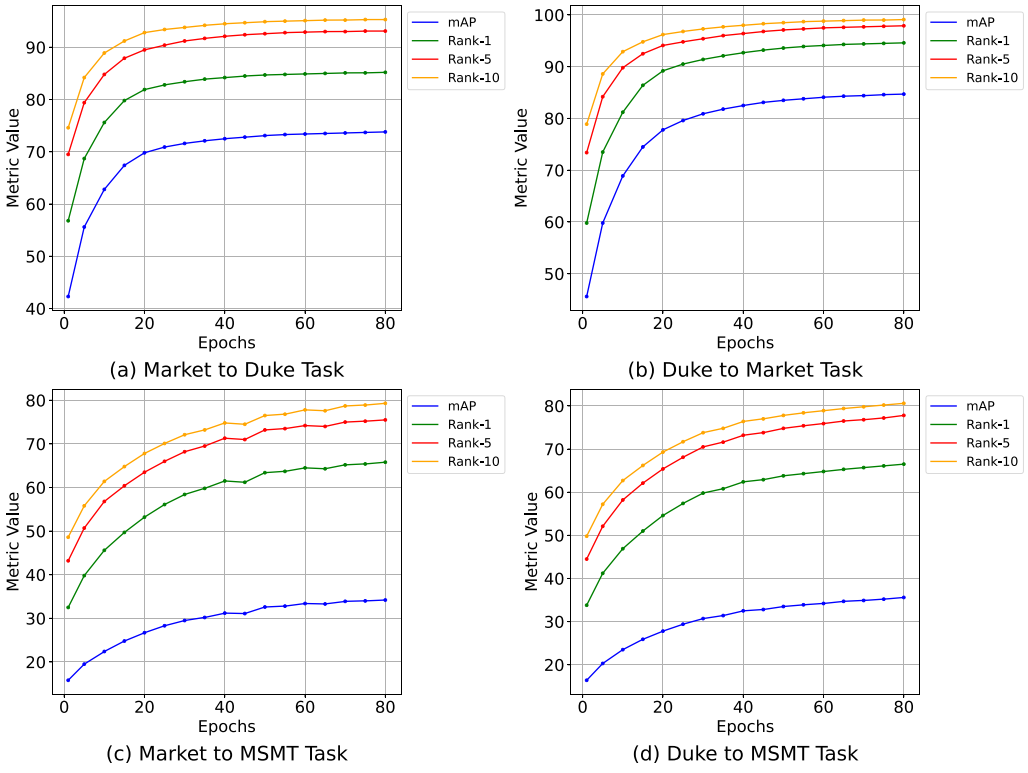


Fig. 4. Metrics Tested on Target Domain during Training.

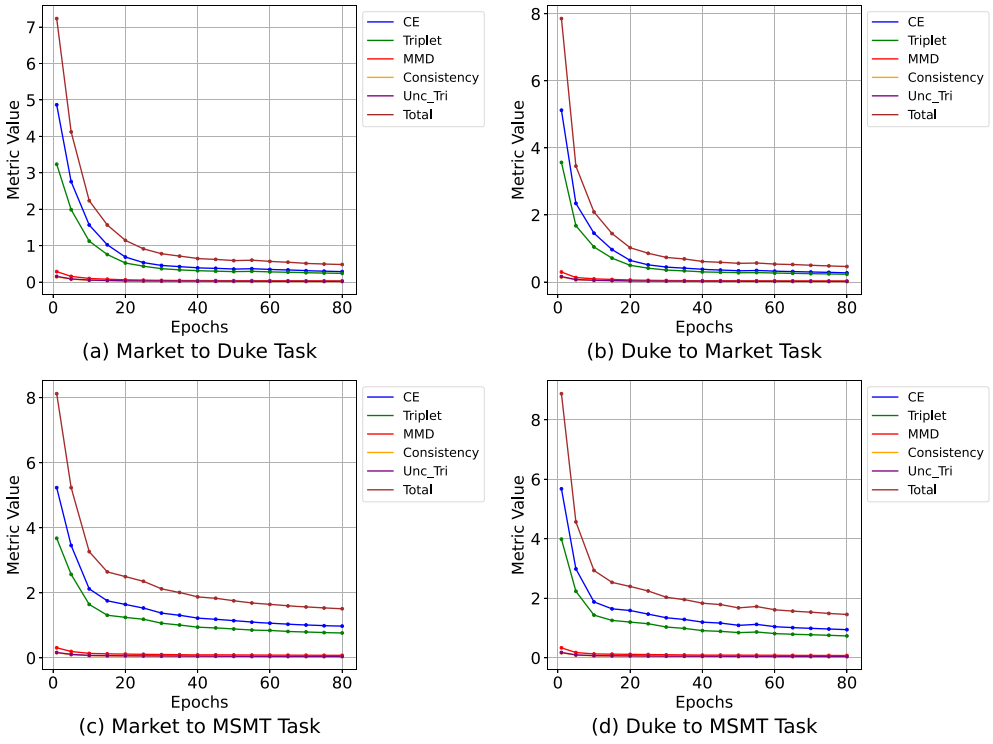


Fig. 5. Losses during Training.

Fig. 5 shows the convergence patterns of different losses involved in our framework. The cross-entropy loss exhibits the fastest decay, reducing by approximately 80% within the first 20 epochs. Similarly, the triplet loss follows a comparable pattern but stabilizes more gradually. Both losses converge after 40 epochs, showing minor fluctuations thereafter. The MMD loss starts at values ranging between 0.28 and 0.33 and reduces to a stable range of 0.03 to 0.08. The consistency loss maintains a similar trend with the MMD loss. The uncertainty loss, although smaller in magnitude, demonstrates stable convergence across tasks.

The convergence behavior can be divided into three distinct phases. The first is the rapid adaptation phase (epochs 1–15). In this phase, a steep reduction in loss ($\approx 70\%$) is observed, accompanied by significant performance gains. This phase achieves the most substantial domain alignment. The second is the refinement phase (epochs 15–40). In this phase, moderate loss reductions occur, leading to steady improvements in evaluation metrics. Feature representations are fine-tuned during this phase. The last is the stabilization phase (epochs 40–80). In this phase, minor and consistent improvements are observed, with loss fluctuations remaining below 5%. This phase consolidates the final performance.

Overall, our framework demonstrates stable convergence patterns across various domain adaptation scenarios. While more challenging tasks, such as those involving MSMT17, exhibit higher final loss values, the convergence trends remain consistent, validating the robustness of the proposed approach.

4.6. Parameter sensitivity analysis

4.6.1. Variations for α in $w(t)$

The hyperparameter α in Eq. (8) governs the dynamic weighting adjustment rate during training. For unreliable samples ($SC(x) < 0$), α controls the rate of weight decay, while for reliable samples ($SC(x) > 0$), the weight adjustment primarily depends on the sample's silhouette coefficient from Eq. (5).

As shown in Fig. 6, optimal α values vary significantly across different tasks. In the Market to Duke scenario, the model achieves peak performance at $\alpha = 1.0$, with notable performance degradation observed when α exceeds 2.0. For the Duke to Market task, optimal results are obtained at $\alpha = 2.0$, where lower values result in slower convergence due to insufficient weight adjustment rates. In both MSMT17-related tasks, the model demonstrates the best performance at $\alpha = 3.0$, enabling rapid adaptation to sample reliability variations in this more complex domain.

This observed variation in optimal α values exhibits a clear correlation with underlying dataset characteristics. Target domains with lower noise levels benefit from smaller α values, which maintain stable weight adjustments and prevent overly aggressive adaptations. Conversely, more challenging domains with higher noise levels require larger α values to facilitate rapid reliability-based weight recalibration, enabling efficient adaptation to increased domain complexity. This relationship between dataset complexity and optimal α values underscores the importance of appropriate hyperparameter selection in cross-domain person Re-ID tasks.

4.6.2. Analysis of MMD-based optimization module

To further validate our theoretical analysis, we explore the relationship between MMD estimation accuracy and computational efficiency. We investigate the effects of batch size and kernel bandwidth σ on MMD estimation and Re-ID performance across four domain adaptation tasks. Our analysis focuses on the theoretical bounds and empirical validation of MMD-based optimization.

The theoretical bound in Eq. (13) suggests that the estimation error decreases at a rate of $O(1/\sqrt{n})$, where n is the batch size. This relationship is empirically verified across all four domain adaptation tasks, as shown in Tables 14 and 15.

As shown in Table 14, for the Market to Duke task, increasing the batch size from 32 to 128 leads to substantial improvements in all metrics, with mAP rising significantly from 62.4% to 73.8% (an 11.4% absolute improvement). Similar trends are observed in the Duke to Market direction, where mAP improves from 74.3% to 84.7% (a 10.4% gain). Notably, the performance gains begin to plateau when the

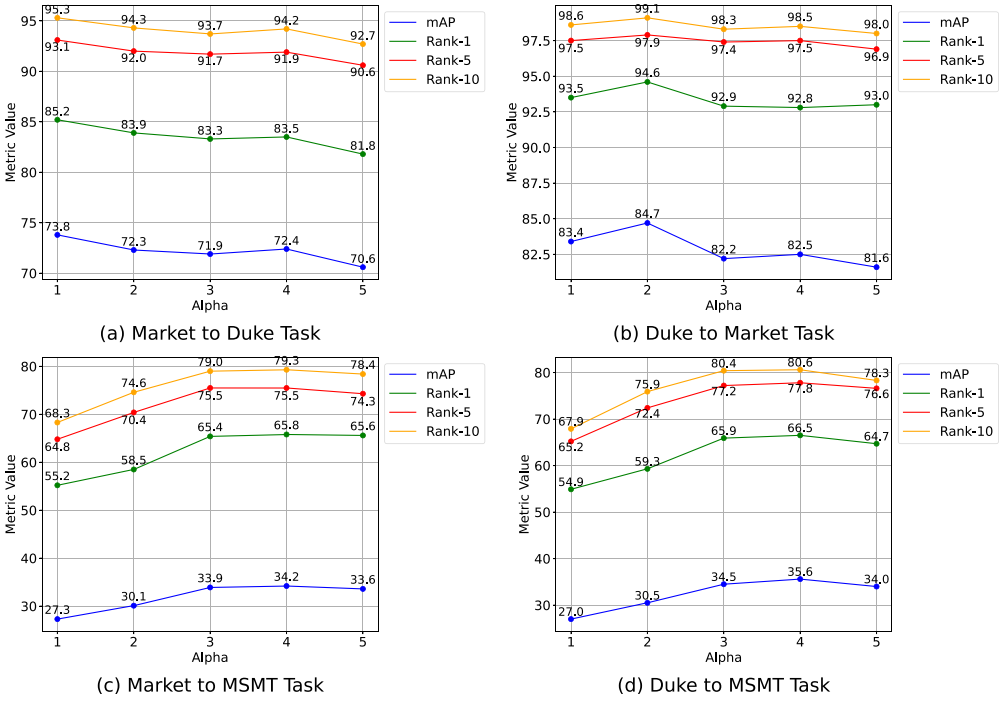


Fig. 6. Performance analysis of different α values in the dynamic weighting strategy. The evaluation spans four cross-domain Re-ID tasks with varying complexity levels. We report mAP and Rank-1/5/10 accuracies for: (a) Market to Duke, showing optimal performance at $\alpha = 1.0$; (b) Duke to Market, peaking at $\alpha = 2.0$; (c) Market to MSMT and (d) Duke to MSMT, both achieving best results at $\alpha = 3.0$.

Table 14
Effect of Batch Size on Market and Duke Tasks Performance.

	Market to Duke				Duke to Market			
Batch Size	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
32	62.4	75.2	84.6	87.8	74.3	87.8	92.4	94.5
64	65.8	78.9	87.5	90.6	78.5	90.2	94.8	96.7
128	**73.8**	85.2	**93.1**	**95.3**	84.7	**94.6**	**97.9**	**99.1**
256	73.6	**85.3**	92.8	**95.3**	**84.8**	94.2	97.8	**99.1**

The best Results are marked in bold.

Table 15
Effect of Batch Size on MSMT17 Transfer Performance.

	Market to MSMT				Duke to MSMT			
Batch Size	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
32	28.4	52.1	63.8	68.5	27.8	51.4	64.5	69.2
64	30.2	54.8	67.2	71.4	29.5	53.9	68.4	72.8
128	**34.2**	**65.8**	75.5	79.3	**35.6**	**66.5**	77.8	80.6
256	33.9	65.5	**75.6**	**79.4**	35.5	66.4	**77.9**	**80.7**

The best Results are marked in Bold.

batch size increases from 128 to 256, with minimal or no improvements across all metrics, suggesting that a batch size of 128 achieves an optimal balance between estimation accuracy and model performance. The impact of batch size is even more pronounced in the more challenging MSMT17 transfer tasks, as illustrated in Table 15. These results align with our theoretical analysis, demonstrating that larger batch sizes of up to 128 provide more reliable MMD estimates, leading to better domain adaptation performance.

Regarding the kernel bandwidth σ, Fig. 7 reveal its crucial role in model performance. For the Market to Duke task, we observe that $\sigma = 1.0$ achieves the best performance with mAP of 73.8% and Rank-1 accuracy of 85.2%. As σ increases, performance deteriorates consistently, with mAP dropping to 68.7% at $\sigma = 5.0$, a significant 5.1% decrease. This pattern is even more pronounced in the Duke to Market task, where mAP declines from 84.7% to 80.6% as σ increases from 1.0 to 5.0. The sensitivity to σ becomes more evident in the more challenging MSMT17 tasks.

These experimental findings strongly support our theoretical analysis in several aspects. First, the diminishing returns observed when increasing batch size beyond 128 align with the $O(1/\sqrt{n})$ convergence rate predicted by Eq. (13). Second, the optimal performance at $\sigma = 1.0$ and subsequent degradation with larger σ values validate the theoretical relationship established in Eq. (14), demonstrating that larger bandwidth values compromise the discriminative power of the MMD metric. Furthermore, the more pronounced performance variations in MSMT17 transfers corroborate our analysis of domain complexity.

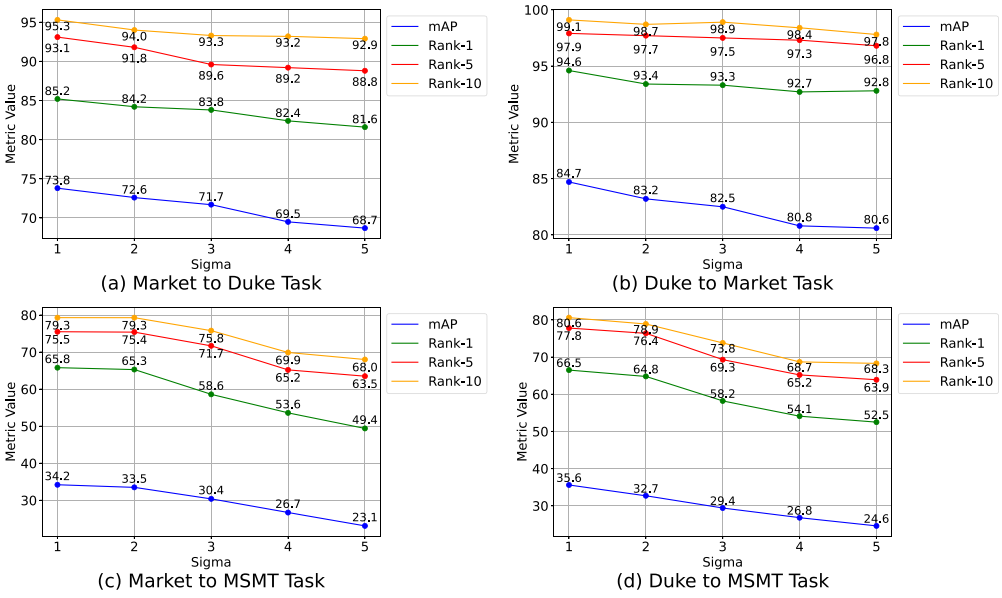


Fig. 7. Performance variation with different kernel bandwidth (σ) values across four cross-domain Re-ID tasks. (a) Market to Duke, (b) Duke to Market, (c) Market to MSMT, and (d) Duke to MSMT. The plots show how mAP and Rank-1/5/10 accuracies vary as σ increases from 1.0 to 5.0, demonstrating optimal performance at $\sigma = 1.0$ across all transfer scenarios.

The consistency between theoretical bounds and experimental results validates the robustness of our MMD-based optimization module for addressing domain adaptation challenges in person Re-ID.

4.6.3. Analysis of memory requirements

We provide a comprehensive analysis of memory requirements from two aspects: memory bank storage and runtime memory dynamics. Our framework employs two distinct memory structures, each tailored to specific aspects of the domain adaptation process. The first is the online memory bank, which is designed for contrastive learning and maintains a consistent structure across tasks. The memory requirement M_{online} is calculated as:

$$M_{online} = (N_s + K) \times d \times s \tag{31}$$

where N_s denotes the number of source domain training identities (751 for Market1501, 702 for DukeMTMC), K is the memory queue size (set to 8192 in our implementation) and d is the feature dimension (set to 2048 in our implementation) [9,33]. s is the size of each feature value (4 bytes). This setup results in a constant memory footprint of approximately 73.5 MB according to Eq. (31), irrespective of the target domain's complexity.

The second is the cluster memory $M_{cluster}$, which adapts to the complexity of the dataset and is empirically optimized for performance:

$$M_{cluster} = k \times d \times s \tag{32}$$

where k is the number of clusters. The selection of the number of clusters k is a critical factor in domain adaptive person Re-ID, influenced by the dataset's complexity, category count, distribution characteristics, and the clustering algorithm's performance. The domain adaptation tasks we address involve significant domain-specific differences, meaning that the complexity and size of datasets directly impact the required granularity of clustering. As such, the choice of k must be tailored to each specific task.

For the tasks of Market to Duke and Duke to Market, we tested k values of 500, 700, and 900 based on prevailing research insights [9,33,58], which involve similarly scaled datasets (12,936 and 16,522 training images) and comparable identity counts (751 and 702 training persons). As shown in Fig. 8(a) and (b), the optimal k value is 700, leading to a memory size of approximately 5.74 MB according to Eq. (32). This suggests that this k value achieves a balance between capturing sufficient intra-domain variance and avoiding over-segmentation, which can lead to noise. When k is 500, which leads to insufficient clustering, while memory size is 4.10 MB. When k is 900, which leads to diminishing returns, while memory size is 7.37 MB.

For the more complex Market/Duke to MSMT tasks, which involve a larger target dataset (32,621 training images) and more identities (1041 training persons). As shown in Fig. 8(c) and (d), the optimal k value is 1500, yielding a memory size of approximately 12.29 MB. k = 1500 provided the best results, demonstrating that this cluster density is sufficient to encapsulate detailed variances without incurring excessive fragmentation. We explored k values of 500, 1000, 1500, and 2000 [9, 33,58]. When k is 500, which leads to inadequate representation (while memory size is 4.10 MB), a lower number of clusters failed to capture the nuanced feature differences within such complex datasets, resulting in suboptimal cross-domain recognition performance. When k is 1000, it leads to suboptimal clustering, while memory size is 8.19 MB. When k is 1500, which leads to excessive fragmentation, while memory size is 16.38 MB.

For the runtime memory analysis, there are two stages involved in our method. During Stage One (Pre-training), static memory includes 154.12 MB for the ResNet50-pro model and an equal amount for the Adam optimizer state, totaling approximately 308.24 MB. The batch processing of 128 images (256 × 128 × 3) requires about 50 MB, with the peak memory usage reaching around 2 GB due to feature maps and intermediate computations. During Stage Two (Domain adaptive learning), the total static memory consumption includes 387.5 MB to 394 MB for the teacher–student networks, 73.5 MB for the online memory bank, and dataset-dependent cluster memory, in which 5.74 MB for Market to Duke and Duke to Market tasks, while 12.29 MB for Market/Duke to MSMT tasks. For the dynamic memory requirements, primarily due to feature extraction, clustering operations, and EMA updates, lead to peak usage of approximately 3 GB for Market to Duke and Duke to Market tasks, and 4 GB for Market1501/Duke to MSMT tasks.

Our ablation study demonstrates that our memory requirements are carefully tuned to the characteristics of each dataset, ensuring optimal performance. This design facilitates effective domain adaptation across varying dataset complexities while maintaining practical memory usage for modern GPUs.

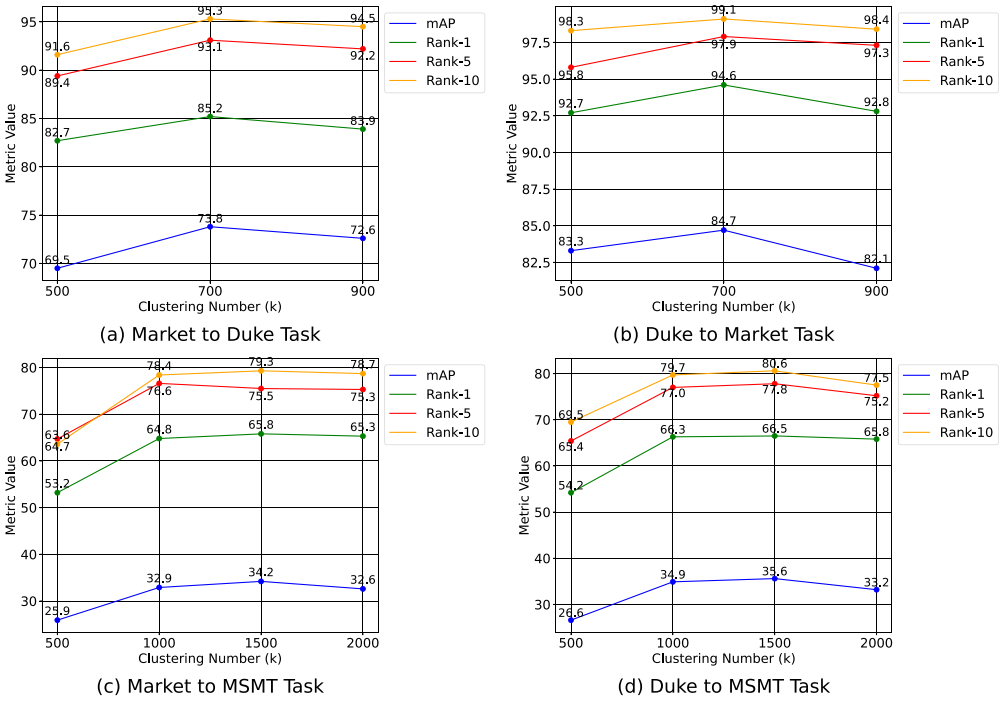


Fig. 8. Performance analysis of varying cluster numbers (k) across different domain adaptive Re-ID tasks. Results are shown for: (a) Market to Duke, demonstrating optimal performance at $k = 700$; (b) Duke to Market, showing similar optimal characteristics; (c) Market to MSMT and (d) Duke to MSMT, both exhibiting best performance at $k = 1500$ due to increased dataset complexity.

Table 16
Impact of Weights on Market to Duke and Duke to Market Performance.

Weight Setting	Market to Duke				Duke to Market			
	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
$\gamma_5 = 0.3$	70.2	82.4	90.8	93.1	81.5	91.8	95.6	97.2
$\gamma_5 = 0.5$	**73.8**	**85.2**	**93.1**	**95.3**	**84.7**	**94.6**	**97.9**	**99.1**
$\gamma_5 = 0.7$	71.5	83.1	91.2	93.8	82.3	92.5	96.1	97.8
$\gamma_3 = 0.2$	71.6	83.5	91.4	93.9	82.6	92.7	96.3	97.9
$\gamma_3 = 0.3$	**73.8**	**85.2**	**93.1**	**95.3**	**84.7**	**94.6**	**97.9**	**99.1**
$\gamma_3 = 0.4$	72.2	84.0	91.9	94.4	83.4	93.4	96.7	98.3
$\gamma_4 = 0.2$	71.9	83.7	91.6	94.1	82.9	92.9	96.5	98.1
$\gamma_4 = 0.3$	**73.8**	**85.2**	**93.1**	**95.3**	**84.7**	**94.6**	**97.9**	**99.1**
$\gamma_4 = 0.4$	72.5	84.2	92.1	94.6	83.7	93.6	96.9	98.5

The best results are marked in bold.

4.6.4. Variations for weight coefficients γ_i

The relationship between different loss components is governed by weight coefficients γ_i according to Eq. (26), which are determined through both theoretical analysis and comprehensive empirical validation across all cross-domain tasks. The cross-entropy and standard triplet losses, serving as primary supervision components, receive unit weights ($\gamma_1, \gamma_2 = 1.0$) to maintain fundamental identity discrimination capabilities.

Tables 16 and 17 present comprehensive ablation studies validating these weight selections, revealing several crucial findings. For the optimization modules, the MMD-based module plays a domain adaptation role, the loss weight is set to 0.5 to prevent over-alignment between domains. This value is determined based on the observation that larger weights (> 0.5) can lead to negative transfer by forcing excessive domain alignment, while smaller weights (< 0.5) provide insufficient adaptation. The 0.5 weight achieves a balance between domain adaptation and preservation of discriminative features. The Consistency supervised module is the auxiliary component. The uncertainty-modulated triplet loss (γ_3) and consistency losses (γ_4) receive smaller weights to serve as regularizers. The 0.3 weight is chosen because it is large enough to influence model behavior (weights < 0.2 showed minimal impact), and small enough to not overshadow primary supervision (weights > 0.4 led to training instability. Thus, 0.3 provides the best trade-off between regularization and stability. Experimental results show that this configuration outperforms other combinations across various tasks, with Market to Duke tasks tolerating greater weight variation compared to MSMT17 tasks, which are more sensitive.

These fixed coefficients work in conjunction with the dynamic weight $w(t)_i$, which modulates sample contributions based on pseudo-label reliability throughout training. This dual-weighting mechanism ensures both structural stability through fixed coefficients and adaptive learning through dynamic sample weighting, contributing to the robust performance of our approach across diverse cross-domain scenarios.

4.7. Visualization and analysis

4.7.1. Analysis of feature distribution

To vividly illustrate the feature learning capabilities of our proposed method, we visualize the learned feature distributions using t-SNE [59] on the Market1501 dataset, as shown in Fig. 9.

Table 17
Impact of Weights on Market to MSMT and Duke to MSMT Performance.

Weight Setting	Market to MSMT				Duke to MSMT			
	mAP (%)	Rank-1(%)	Rank-5 (%)	Rank-10 (%)	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
$\gamma_5 = 0.3$	31.5	62.4	72.8	76.5	32.8	63.2	74.5	77.4
$\gamma_5 = 0.5$	**34.2**	**65.8**	**75.5**	**79.3**	**35.6**	**66.5**	**77.8**	**80.6**
$\gamma_5 = 0.7$	32.1	63.2	73.4	77.1	33.5	64.1	75.2	78.1
$\gamma_3 = 0.2$	32.3	63.4	73.5	77.2	33.6	64.2	75.3	78.2
$\gamma_3 = 0.3$	**34.2**	**65.8**	**75.5**	**79.3**	**35.6**	**66.5**	**77.8**	**80.6**
$\gamma_3 = 0.4$	32.7	64.1	74.0	77.7	34.1	65.0	76.1	79.0
$\gamma_4 = 0.2$	32.5	63.6	73.7	77.5	33.8	64.4	75.5	78.4
$\gamma_4 = 0.3$	**34.2**	**65.8**	**75.5**	**79.3**	**35.6**	**66.5**	**77.8**	**80.6**
$\gamma_4 = 0.4$	32.9	64.3	74.2	77.9	34.3	65.2	76.3	79.2

The best results are marked in bold.

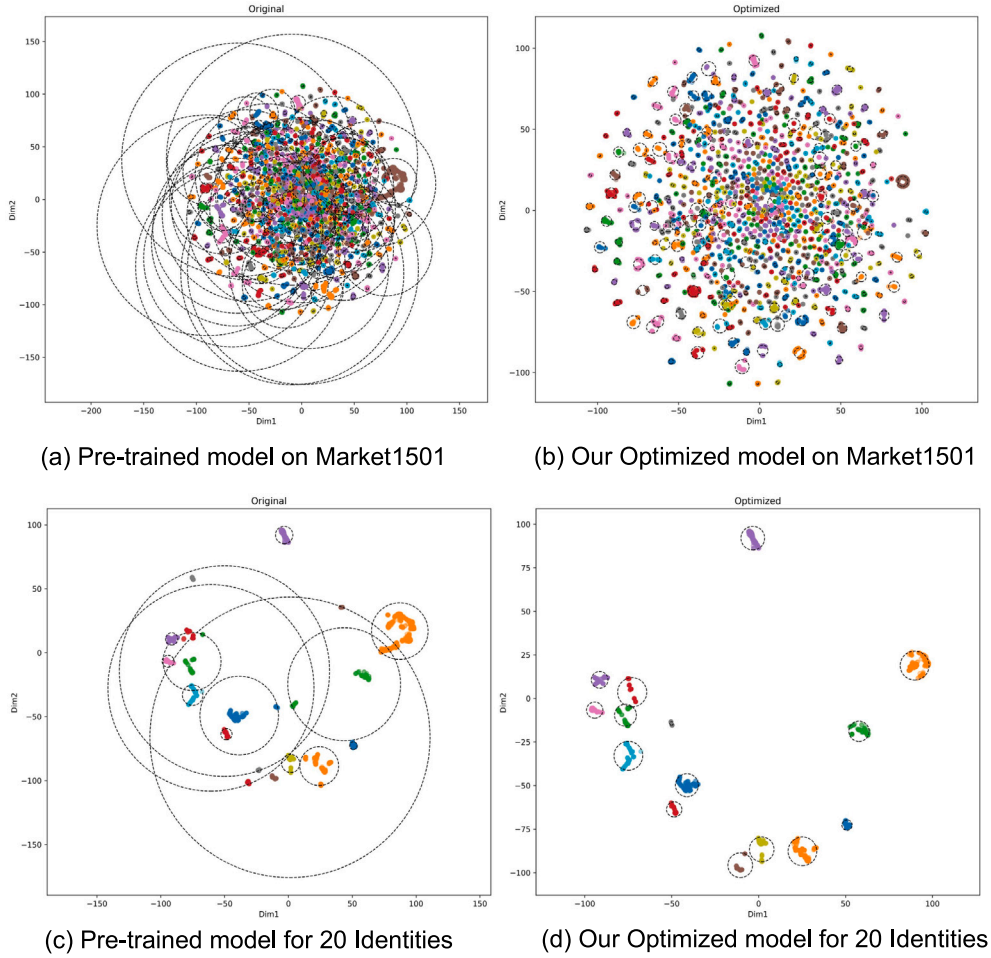


Fig. 9. t-SNE visualization of feature distributions on Market1501. (a) Pre-trained model on full gallery set; (b) Our optimized model on full gallery set; (c) Pre-trained model on 20 selected identities; (d) Our optimized model on 20 selected identities.

Figs. 9(a) and (b) compare feature distributions across the complete gallery set. The pre-trained model exhibits significant feature overlap between different identities, indicating suboptimal discriminative power. In contrast, our optimized model demonstrates clear inter-class boundaries while maintaining intra-class compactness.

This improvement is further evidenced in Figs. 9(c) and (d), which focus on 20 randomly selected identities. The pre-trained model shows scattered intra-class features (highlighted by dashed circles), while our model achieves well-defined, compact clusters for each identity. This enhanced feature separation directly contributes to improved Re-ID performance, particularly in challenging cross-domain scenarios.

4.7.2. Feature visualization analysis

To visualize the model's attention patterns, we employ Grad-CAM [60] to generate activation heatmaps that highlight regions contributing most significantly to feature extraction. Fig. 10 compares the activation patterns between the baseline and our optimized model on the Market to Duke task.

The baseline model exhibits diffuse activation patterns, with attention scattered across both relevant and irrelevant image regions. In contrast, our optimized model demonstrates more focused attention on identity-discriminative features, particularly anatomical regions

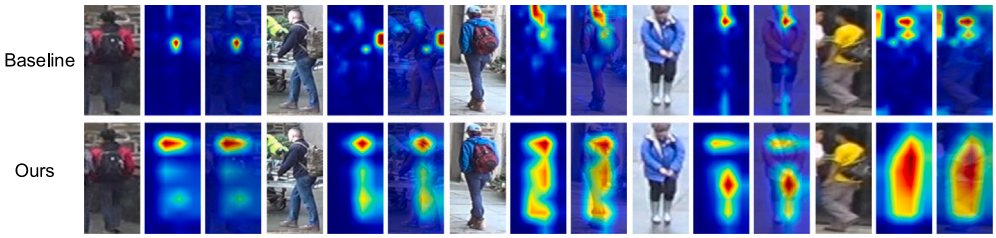


Fig. 10. Grad-CAM visualization comparing baseline and optimized models on Market to Duke task. The heatmaps indicate regions of high activation in the feature extraction process.

and distinctive accessories. This enhanced feature localization capability directly contributes to the model's improved domain adaptation performance.

5. Future discussion

While our work advances the field of domain adaptive person Re-ID, numerous opportunities remain for further exploration and optimization. The first is enhancing algorithmic efficiency, particularly in dynamic architecture adjustments and computational resource management, which could further streamline the deployment of robust Re-ID models. Similarly, integrating advanced feature learning techniques holds promise for improving discrimination and generalization, which is also our focus. Expanding the application of Re-ID to multi-domain adaptation, real-world scenarios, and diverse fields such as vehicle tracking, wildlife conservation, and industrial automation could broaden the impact and utility of these systems. These future directions highlight the potential for continuous innovation in this rapidly evolving field.

6. Conclusion

In this paper, we propose NODW, a comprehensive framework for domain adaptive person re-identification that effectively addresses cross-domain feature learning and pseudo-label noise challenges. Our framework introduces three key components: a silhouette coefficient-based noise assessment module with dynamic weighting, an MMD-based domain alignment mechanism, and a consistency-supervised learning strategy. Extensive experiments on standard benchmarks demonstrate the effectiveness of our method, achieving 73.8% mAP for Market to Duke task, 84.7% mAP for Duke to Market, and maintaining robust performance on the more challenging MSMT17 dataset (34.2% mAP for Market to MSMT, and 35.6% mAP for Duke to MSMT). The ablation studies validate each component's contribution, throughout all tasks, with the noise assessment providing about 3.8% mAP gain, the MMD module adding about 4.8%, and consistency supervision contributing about 1.8% improvement while maintaining practical computational requirements. These results, combined with the framework's demonstrated scalability and efficiency, establish NODW as an effective solution for real-world cross-domain person re-identification challenges.

CRediT authorship contribution statement

Zhengyang Wang: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Conceptualization. **Xiufen Ye:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition. **Xue Shang:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis. **Shuxiang Guo:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xiufen Ye reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 42276187).

Data availability

Data will be made available on request.

References

[1] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, S.C. Hoi, Deep learning for person re-identification: A survey and outlook, IEEE Trans. Pattern Anal. Mach. Intell. 44 (6) (2021) 2872–2893, http://dx.doi.org/10.1109/TPAMI.2021.3054775.

[2] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, Y. Xu, Deep-person: Learning discriminative deep features for person re-identification, Pattern Recognit. 98 (2020) 107036, http://dx.doi.org/10.1016/j.patcog.2019.107036.

[3] Z. Wang, X. Ye, X. Shang, S.S. Ge, S. Guo, Person re-identification method with Mahalanobis TRM triplet on multi-branch network, Appl. Intell. 53 (23) (2023) 29183–29204, http://dx.doi.org/10.1007/s10489-023-05039-9.

[4] W. Wu, D. Tao, H. Li, Z. Yang, J. Cheng, Deep features for person re-identification on metric learning, Pattern Recognit. 110 (2021) 107424, http://dx.doi.org/10.1016/j.patcog.2020.107424.

[5] Z. Zhong, L. Zheng, Z. Luo, S. Li, Y. Yang, Invariance matters: Exemplar memory for domain adaptive person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 598–607, http://dx.doi.org/10.1109/CVPR.2019.00069.

[6] A. Zahra, N. Perwaiz, M. Shahzad, M.M. Fraz, Person re-identification: A retrospective on domain specific open challenges and future trends, Pattern Recognit. (2023) 109669, http://dx.doi.org/10.1016/j.patcog.2023.109669.

[7] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, T.S. Huang, Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6112–6121, http://dx.doi.org/10.1109/ICCV.2019.00621.

[8] X. Jin, C. Lan, W. Zeng, Z. Chen, Global distance-distributions separation for unsupervised person re-identification, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, 2020, pp. 735–751, http://dx.doi.org/10.1007/978-3-030-58571-6_43.

[9] Y. Ge, D. Chen, H. Li, Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification, in: International Conference on Learning Representations, 2020, URL https://openreview.net/forum?id=rJlnOhVYPS.

[10] Y. Ge, F. Zhu, D. Chen, R. Zhao, et al., Self-paced contrastive learning with hybrid memory for domain adaptive object re-id, Adv. Neural Inf. Process. Syst. 33 (2020) 11309–11321, http://dx.doi.org/10.5555/3495724.3496673.

[11] Y. Zhai, Q. Ye, S. Lu, M. Jia, R. Ji, Y. Tian, Multiple expert brainstorming for domain adaptive person re-identification, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, Springer, 2020, pp. 594–611, http://dx.doi.org/10.1007/978-3-030-58571-6_35.

[12] T. Isobe, D. Li, L. Tian, W. Chen, Y. Shan, S. Wang, Towards discriminative representation learning for unsupervised person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 8526–8536, http://dx.doi.org/10.1109/ICCV48922.2021.00841.

[13] H. Rami, M. Ospici, S. Lathuilière, Online unsupervised domain adaptation for person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 3830–3839, http://dx.doi.org/10.1109/CVPRW56347.2022.00428.

[14] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in: Kdd, vol. 96, (no. 34) 1996, pp. 226–231, http://dx.doi.org/10.5555/3001460.3001507.

[15] T. He, L. Shen, Y. Guo, G. Ding, Z. Guo, Secret: Self-consistent pseudo label refinement for unsupervised domain adaptive person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, No. 1, 2022, pp. 879–887, http://dx.doi.org/10.1609/aaai.v36i1.19970.

[16] H. Fan, L. Zheng, C. Yan, Y. Yang, Unsupervised person re-identification: Clustering and fine-tuning, ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) 14 (4) (2018) 1–18, http://dx.doi.org/10.1145/3243316.

[17] X. Zhang, J. Cao, C. Shen, M. You, Self-training with progressive augmentation for unsupervised cross-domain person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8222–8231, http://dx.doi.org/10.1109/ICCV.2019.00831.

[18] D. Zheng, J. Xiao, K. Chen, X. Huang, L. Chen, Y. Zhao, Soft pseudo-label shrinkage for unsupervised domain adaptive person re-identification, Pattern Recognit. 127 (2022) 108615, http://dx.doi.org/10.1016/j.patcog.2022.108615.

[19] Z. Wang, S. Guo, X. Shang, X. Ye, Pseudo-label assisted optimization of multi-branch network for cross-domain person re-identification, in: 2023 IEEE International Conference on Mechatronics and Automation, ICMA, IEEE, 2023, pp. 13–18, http://dx.doi.org/10.1109/ICMA57826.2023.10215862.

[20] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1116–1124, http://dx.doi.org/10.1109/ICCV.2015.133.

[21] H. Park, B. Ham, Relation network for person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 07, 2020, pp. 11839–11847, http://dx.doi.org/10.1609/aaai.v34i07.6857.

[22] D. Wu, S.-J. Zheng, X.-P. Zhang, C.-A. Yuan, F. Cheng, Y. Zhao, Y.-J. Lin, Z.-Q. Zhao, Y.-L. Jiang, D.-S. Huang, Deep learning-based methods for person re-identification: A comprehensive review, Neurocomputing 337 (2019) 354–371, http://dx.doi.org/10.1016/j.neucom.2019.01.079.

[23] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, Q. Tian, Person re-identification in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1367–1376, http://dx.doi.org/10.1109/CVPR.2017.357.

[24] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, Y. Tian, Unsupervised cross-dataset transfer learning for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1306–1315, http://dx.doi.org/10.1109/CVPR.2016.146.

[25] Y. Zou, X. Yang, Z. Yu, B.V. Kumar, J. Kautz, Joint disentangling and adaptation for cross-domain person re-identification, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, Springer, 2020, pp. 87–104, http://dx.doi.org/10.1007/978-3-030-58536-5_6.

[26] L. Wei, S. Zhang, W. Gao, Q. Tian, Person transfer gan to bridge domain gap for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 79–88, http://dx.doi.org/10.1109/CVPR.2018.00016.

[27] J. Liu, Z.-J. Zha, D. Chen, R. Hong, M. Wang, Adaptive transfer network for cross-domain person re-identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7202–7211, http://dx.doi.org/10.1109/CVPR.2019.00737.

[28] Y. Ge, L. Liu, H. Zhang, A three-stage learning approach to cross-domain person re-identification, Appl. Soft Comput. 112 (2021) 107793, http://dx.doi.org/10.1016/j.asoc.2021.107793.

[29] F. Chen, N. Wang, J. Tang, F. Zhu, A feature disentangling approach for person re-identification via self-supervised data augmentation, Appl. Soft Comput. 100 (2021) 106939.

[30] X. Liu, S. Zhang, Graph consistency based mean-teaching for unsupervised domain adaptive person re-identification, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, 2021, pp. 874–880, http://dx.doi.org/10.24963/ijcai.2021/121.

[31] Y. Zhai, S. Lu, Q. Ye, X. Shan, J. Chen, R. Ji, Y. Tian, AD-cluster: Augmented discriminative clustering for domain adaptive person re-identification, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 9018–9027, http://dx.doi.org/10.1109/CVPR42600.2020.00904.

[32] D. Wang, S. Zhang, Unsupervised person re-identification via multi-label classification, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 10978–10987, http://dx.doi.org/10.1109/CVPR42600.2020.01099.

[33] K. Zheng, C. Lan, W. Zeng, Z. Zhang, Z.-J. Zha, Exploiting sample uncertainty for domain adaptive person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, No. 4, 2021, pp. 3538–3546, http://dx.doi.org/10.1609/aaai.v35i4.16468.

[34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778, http://dx.doi.org/10.1109/CVPR.2016.90.

[35] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 9726–9735, http://dx.doi.org/10.1109/CVPR42600.2020.00975.

[36] C.D. Manning, Introduction to Information Retrieval, Cambridge University Press, 2008.

[37] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, http://dx.doi.org/10.1016/0377-0427(87)90125-7.

[38] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (1) (2012) 723–773, http://dx.doi.org/10.5555/2188385.2188410.

[39] J. Mercer, Xvi. functions of positive and negative type, and their connection the theory of integral equations, Philos. Trans. R. Soc. Lond. Ser. A, Contain. Pap. A Math. Or Phys. Character 209 (441–458) (1909) 415–446, http://dx.doi.org/10.1098/rsta.1909.0016.

[40] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, L.E. Meester, A Modern Introduction to Probability and Statistics: Understanding Why and How, Springer Science & Business Media, 2006, http://dx.doi.org/10.1007/1-84628-168-7.

[41] W. Hoeffding, Probability inequalities for sums of bounded random variables, Collect. Work. Wassily Hoeffding (1994) 409–426, http://dx.doi.org/10.1080/01621459.1963.10500830.

[42] A. Kendall, Y. Gal, What uncertainties do we need in bayesian deep learning for computer vision? Adv. Neural Inf. Process. Syst. 30 (11) (2017) 5580–5590, http://dx.doi.org/10.5555/3295222.3295309.

[43] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: European Conference on Computer Vision, Springer, 2016, pp. 17–35, http://dx.doi.org/10.1007/978-3-319-48881-3_2.

[44] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, Shape and appearance context modeling, in: 2007 IEEE 11th International Conference on Computer Vision, IEEE, 2007, pp. 1–8, http://dx.doi.org/10.1109/ICCV.2007.4409019.

[45] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2010) 303–338, http://dx.doi.org/10.1007/s11263-009-0275-4.

[46] M. Zhang, K. Liu, Y. Li, S. Guo, H. Duan, Y. Long, Y. Jin, Unsupervised domain adaptation for person re-identification via heterogeneous graph alignment, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, No. 4, 2021, pp. 3360–3368, http://dx.doi.org/10.1609/aaai.v35i4.16448.

[47] J. Han, Y.-L. Li, S. Wang, Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, No. 1, 2022, pp. 790–798, http://dx.doi.org/10.1609/aaai.v36i1.19960.

[48] W. Wang, F. Zhao, S. Liao, L. Shao, Attentive WaveBlock: Complementarity-enhanced mutual networks for unsupervised domain adaptation in person re-identification and beyond, IEEE Trans. Image Process. 31 (2022) 1532–1544, http://dx.doi.org/10.1109/TIP.2022.3140614.

[49] F. Chen, N. Wang, J. Tang, P. Yan, J. Yu, Unsupervised person re-identification via multi-domain joint learning, Pattern Recognit. 138 (2023) 109369, http://dx.doi.org/10.1016/j.patcog.2023.109369.

[50] L. He, X. Liao, W. Liu, X. Liu, P. Cheng, T. Mei, Fastreid: A pytorch toolbox for general instance re-identification, in: Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 9664–9667, http://dx.doi.org/10.1145/3581783.3613460.

[51] Y. Tao, J. Zhang, J. Hong, Y. Zhu, DREAMT: Diversity enlarged mutual teaching for unsupervised domain adaptive person re-identifcation, IEEE Trans. Multimed. (2022) http://dx.doi.org/10.1109/TMM.2022.3178599.

[52] X. Pan, P. Luo, J. Shi, X. Tang, Two at once: Enhancing learning and generalization capacities via ibn-net, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 464–479, http://dx.doi.org/10.1007/978-3-030-01225-0_29.

[53] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Omni-scale feature learning for person re-identification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3702–3712, http://dx.doi.org/10.1109/ICCV.2019.00380.

[54] G. Wang, Y. Yuan, X. Chen, J. Li, X. Zhou, Learning discriminative features with multiple granularities for person re-identification, in: Proceedings of the 26th ACM International Conference on Multimedia, 2018, pp. 274–282, http://dx.doi.org/10.1145/3240508.3240552.

[55] B. Cheng, B. Xiao, J. Wang, H. Shi, T.S. Huang, L. Zhang, Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5386–5395, http://dx.doi.org/10.1109/CVPR42600.2020.00543.

[56] A. Dosovitskiy, An image is worth 16x16 words: Transformers for image recognition at scale, 2020, http://dx.doi.org/10.48550/arXiv.2010.11929, arXiv preprint arXiv:2010.11929.

[57] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al., Dinov2: Learning robust visual features without supervision, 2023, http://dx.doi.org/10.48550/arXiv.2304.07193, arXiv preprint arXiv:2304.07193.

[58] Y. Li, H. Yao, C. Xu, TEST: Triplet ensemble student-teacher model for unsupervised person re-identification, IEEE Trans. Image Process. 30 (2021) 7952–7963, http://dx.doi.org/10.1109/TIP.2021.3112039.

[59] L. Van der Maaten, G. Hinton, Visualizing data using t-sne., J. Mach. Learn. Res. 9 (11) (2008) URL http://jmlr.org/papers/v9/vandermaaten08a.html.

[60] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626, http://dx.doi.org/10.1109/ICCV.2017.74.