Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.055157

ARTICLE

Enhancing Unsupervised Domain Adaptation for Person Re-Identification with the Minimal Transfer Cost Framework

Sheng Xu[1], Shixiong Xiang[2], Feiyu Meng[1] and Qiang Wu[1,*]

[1]Research Institute of Electronic Science and Technology, University of Electronic Science and Technology of China, Chengdu, 610000, China

[2]National Key Laboratory of Optical Field Manipulation Science and Technology, Chinese Academy of Sciences, Chengdu, 610209, China

*Corresponding Author: Qiang Wu. Email: qiang.wu@uestc.edu.cn

Received: 19 June 2024 Accepted: 05 August 2024 Published: 12 September 2024

ABSTRACT

In Unsupervised Domain Adaptation (UDA) for person re-identification (re-ID), the primary challenge is reducing the distribution discrepancy between the source and target domains. This can be achieved by implicitly or explicitly constructing an appropriate intermediate domain to enhance recognition capability on the target domain. Implicit construction is difficult due to the absence of intermediate state supervision, making smooth knowledge transfer from the source to the target domain a challenge. To explicitly construct the most suitable intermediate domain for the model to gradually adapt to the feature distribution changes from the source to the target domain, we propose the Minimal Transfer Cost Framework (MTCF). MTCF considers all scenarios of the intermediate domain during the transfer process, ensuring smoother and more efficient domain alignment. Our framework mainly includes three modules: Intermediate Domain Generator (IDG), Cross-domain Feature Constraint Module (CFCM), and Residual Channel Space Module (RCSM). First, the IDG Module is introduced to generate all possible intermediate domains, ensuring a smooth transition of knowledge from the source to the target domain. To reduce the cross-domain feature distribution discrepancy, we propose the CFCM Module, which quantifies the difficulty of knowledge transfer and ensures the diversity of intermediate domain features and their semantic relevance, achieving alignment between the source and target domains by incorporating mutual information and maximum mean discrepancy. We also design the RCSM, which utilizes attention mechanism to enhance the model's focus on personnel features in low-resolution images, improving the accuracy and efficiency of person re-ID. Our proposed method outperforms existing technologies in all common UDA re-ID tasks and improves the Mean Average Precision (mAP) by 2.3% in the Market to Duke task compared to the state-of-the-art (SOTA) methods.

KEYWORDS

Person re-identification; unsupervised domain adaptation; attention mechanism; mutual information; maximum mean discrepancy

1 Introduction

Person re-identification (re-ID) [1–3] is a computer vision task at matching images of individuals from different camera perspectives. It is widely used in surveillance and security, enabling identification across different times and spaces, with potential value in social analysis and behavior understanding. Traditional re-ID algorithms often rely on handcrafted features such as SIFT [4] and HOG [5], followed by Support Vector Machine (SVM) classifiers [6]. However, these methods face challenges like lighting changes and background occlusions. With advancements in deep learning (DL), DL-based re-ID methods have emerged, categorized into supervised [1,7,8] and unsupervised [9–11] methods. Supervised methods require extensive labeling and often face performance limitations due to domain discrepancies between labeled source and unlabeled target datasets. To address cross-domain differences, researchers focus on unsupervised cross-domain re-ID methods [12,13], which use both source domain labels and target domain data to improve model generalization (see Fig. 1).

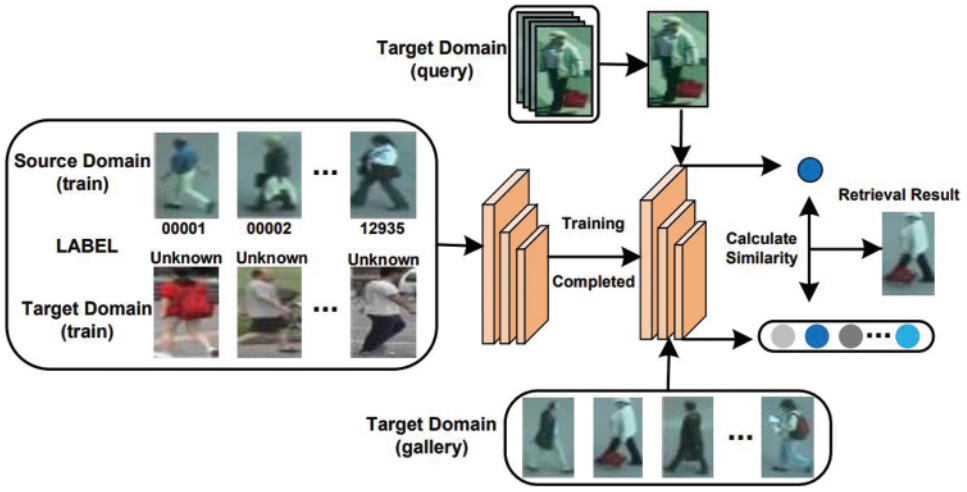

Figure 1: The concept of training and testing models for unsupervised cross-domain person re-ID

Recent studies in unsupervised cross-domain person re-ID have adopted Domain Adaptation (DA) methods to directly align the distributions of the source and target domains. These methods implicitly construct an intermediate domain [14,15], aiming to reduce cross-domain discrepancies by mapping the source and target domains to this intermediate domain. However, due to the lack of supervision for the intermediate state, these methods struggle to provide a stable and smooth transition state, resulting in limited generalization capability. Additionally, guided by the labels of the source domain, direct alignment methods sometimes overfit to the feature distribution of the source domain rather than achieving a more generalized feature representation that encompasses both domains.

In this paper, we propose the Minimal Transfer Cost Framework (MTCF), which adaptively considers all possible intermediate domains during alignment, ensuring feature diversity and smooth knowledge transfer. It comprises three modules: the Intermediate Domain Generator (IDG), the Cross-domain Feature Constraint Module (CFCM), and the Residual Channel Space Module (RCSM). The IDG considers all intermediate domains for knowledge transfer, ensuring computational simplicity. The CFCM uses Maximum Mean Discrepancy (MMD) to balance feature transfer costs, ensuring key feature sharing between domains and utilizing Mutual Information (MI) to maintain intra-domain diversity. The RCSM enhances feature extraction accuracy, improving the model's focus on key personnel features.

Our contributions are summarized as follows:

- We introduce a novel approach that considers all potential intermediate domains to ensure a smooth transition of knowledge from the source domain to the target domain. This strategy effectively bridges the gap between different domains, facilitating more efficient and accurate person re-identification.
- We design a key feature approach that shares between the source and target domains by accurately quantifying the knowledge transfer cost. This approach maintains the integrity and relevance of features, achieving precise alignment and improving the model's generalization capabilities.
- We focus on enhancing feature extraction accuracy, particularly in low-resolution images. This improvement boosts the overall effectiveness and efficiency of the person re-identification process, ensuring that key personnel features are accurately captured.
- Extensive experiments show our method achieves state-of-the-art (SOTA) results across various datasets, with a 2.3% improvement in mean Average Precision (mAP) over advanced techniques, enhancing transfer learning performance.

We compare our method with several state-of-the-art approaches to provide a clear context for our contributions. For instance, SPGAN [14] uses GANs for image-to-image translation between source and target domains but struggles with preserving identity information. ECN [15] employs exemplar and camera-invariance constraints to enhance re-ID performance but faces challenges with large domain gaps. Our method differs by explicitly generating intermediate feature representations, avoiding identity mismatch issues. Moreover, methods like MMT [12] focus on mutual mean teaching but lack mechanisms for handling intermediate domain diversity, which our CFCM addresses through mutual information and maximum mean discrepancy.

By addressing these limitations and introducing a comprehensive framework that incorporates intermediate domain generation, feature constraint, and enhanced feature extraction, our method provides a robust solution for cross-domain person re-identification.

2 Related Work

In this section, we discuss two key technologies: person re-identification (re-ID) and Domain Adaptation (DA).

2.1 Person Re-Identification

Person re-identification (re-ID) aims to match images of individuals captured from different camera perspectives. Traditional techniques relied on manually extracting features [16–18]. However, these methods could only extract shallow features, such as color and texture, and failed to capture high-level semantic information. The advent of deep learning methods has effectively addressed this limitation.

Deep learning methods encompass both supervised and unsupervised approaches. Supervised methods [19,20] are suitable for scenarios where all images are labeled and have similar styles. These methods often employ global/local feature representation learning, attention mechanisms, and semantic feature extraction. Nevertheless, they do not meet the diverse needs of real-world scenarios. Consequently, unsupervised re-ID methods, which do not require identity labels, have gained considerable attention.

Unsupervised methods cluster target domain data to generate pseudo-labels for fine-tuning or training. Fan et al. [9] introduced clustering algorithms within the progressive unsupervised learning method PUL. Additionally, Fu et al. [21] exploited similarities between global and local features to create multiple independent clustering pseudo-labels, thereby improving robustness. Zhao et al. [22] combined pseudo-label clustering with the selection of reliable instances to mitigate label noise. Further advancements were made by Zhai et al. [23], who implemented iterative density clustering, adaptive sample augmentation, and discriminative learning.

Hybrid clustering and sample selection methods have also been investigated. For example, Sun et al. [24] and Jin et al. [25] explored various techniques, while Li et al. [26] introduced a confidence-adaptive method for sample separation. To address pose variations and occlusions, Zhang et al. [27] and Raj et al. [11] developed end-to-end networks, which significantly improved accuracy in complex scenarios.

2.2 Domain Adaptation

Domain Adaptation (DA) aims to mitigate the impact of domain shifts on cross-domain re-identification (re-ID) performance by effectively transferring knowledge and constructing implicit or explicit intermediate domains to bridge the gap between the source (labeled) and target (unlabeled) domains. As shown in Fig. 2, (a) represents the typical method of using Generative Adversarial Networks (GANs) to construct implicit intermediate domains (I-Inter), transforming the labeled source domain into an intermediate domain styled after the target domain. In contrast, (b) illustrates our method, which directly constructs explicit intermediate domains (E-Inter) using the source and target domains, serving as a bridge connecting the two.

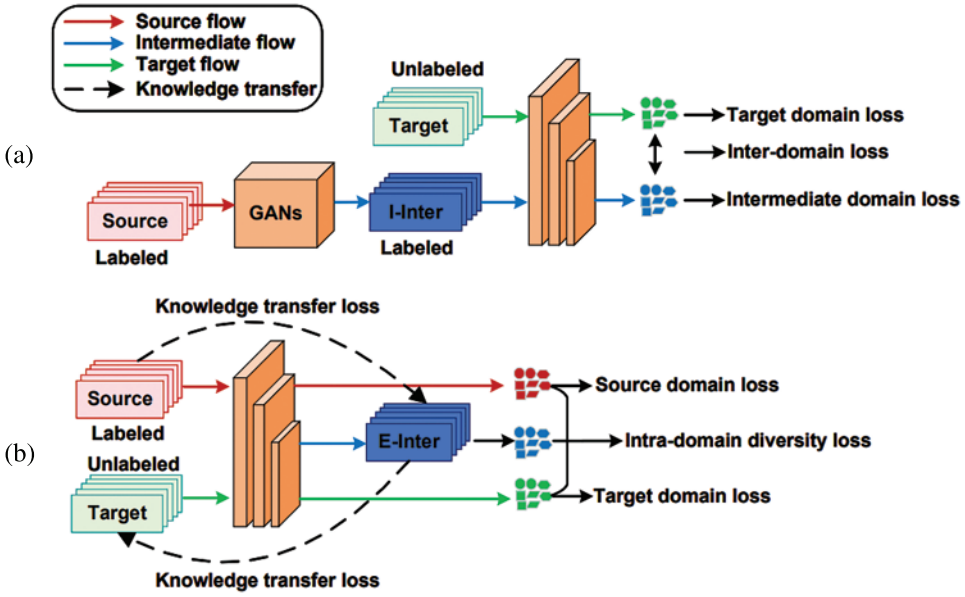

Figure 2: Comparing GAN-based implicit intermediate domains and our explicit intermediate domains for bridging source and target domains

GAN-based methods generate mappings from the source to the target domain, creating implicit intermediate domains. For example, Zheng et al. [28] used GANs to synthesize images with smooth labels. Wei et al. [29] and Deng et al. [30] mapped source images to the target domain style, thereby

narrowing the domain gap. Chen et al. [31] combined GANs with contrastive learning modules, enhancing viewpoint invariance and performance. Additionally, Dai et al. [32] introduced a cyclic GAN to select valuable source images for transferring discriminative information to the target domain. Yang et al. [33] employed DPG-GAN [34] and StarGAN [35] to generate and convert images into various camera styles. Zhong et al. [36] aimed to augment datasets by learning invariant features across domains, thus mitigating image style changes caused by camera transformations.

Explicit intermediate domain construction effectively utilizes intermediate representations to bridge domain gaps. For instance, Dai et al. [37] introduced an intermediate domain module to blend hidden representations, reducing disparities. This approach was further advanced by Dai et al. [38], who generated multiple intermediate domains to minimize feature differences while preserving identity information. Moreover, Na et al. [39] addressed the issue of source domain label dominance by segregating the intermediate space into contrastive and consensus areas, thereby enhancing adaptive model performance. Finally, DFDSN-Net [40] diminished style discrepancies through feature fusion and normalization, implicitly introducing intermediate domains within the feature space.

3 Method

In our section, we propose the Minimal Transfer Cost Framework (MTCF), see Fig. 3, designed for unsupervised domain-adaptive person re-ID. The core of MTCF is identifying an appropriate intermediate domain among all possible intermediate domains, serving as a bridge between the source and target domains. This framework facilitates smooth knowledge transfer, aligning feature distributions and enhancing re-ID performance in the target domain.

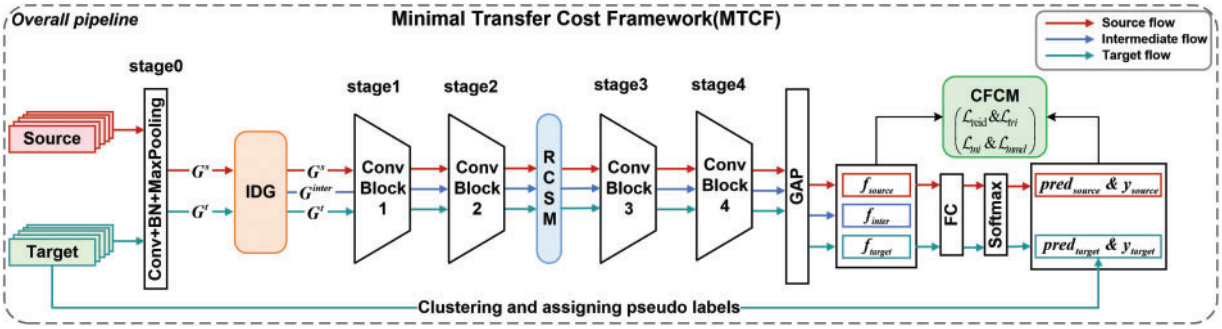

Figure 3: Overall pipeline of our MTCF method

3.1 Overview

Person re-identification involves two key types of data: source domain data with clear labels and target domain data lacking labels. The backbone network in our study is built upon the IBN-ResNet50 [41], comprising five distinct stages. Initially, data from both the source and target domains pass through stage 0, consisting of Conv+BN+MaxPooling layers, resulting in initial feature representations G_s and G_t for the source and target domains, respectively. These initial features are then processed by our proposed Intermediate Domain Generator (IDG) module, which ensures smooth knowledge transition while considering all intermediate domains, producing further feature representations G_s, G_{inter}, and G_t for the source, intermediate, and target domains, respectively.

Subsequently, these features undergo deeper feature extraction through stage 1 to 4, which consist of ConvBlock1 to ConvBlock4 layers, alongside our specially designed Residual Channel Space

Module (RCSM) to enhance feature focus. The feature maps from the three domains then pass through a Global Average Pooling (GAP) layer to generate feature vectors f_{source}, f_{inter}, and f_{target}, which create compact representations for easier comparison and alignment across domains. Finally, these feature vectors are input into the Cross-domain Feature Constraint Module (CFCM) for the calculation of the total loss, aligning the source and target domains while maintaining feature diversity and relevance by incorporating mutual information and maximum mean discrepancy.

3.2 Minimal Transfer Cost Framework

The Minimal Transfer Cost Framework (MTCF) is designed to facilitate efficient and effective cross-domain person re-identification by leveraging three synergistic modules: the Residual Channel Space Module (RCSM), the Intermediate Domain Generator (IDG), and the Cross-domain Feature Constraint Module (CFCM). The RCSM enhances feature extraction by integrating spatial and channel attention mechanisms, ensuring key regions are accurately captured. The IDG generates intermediate domain features that blend source and target domain characteristics, ensuring smooth knowledge transfer and minimizing feature distribution discrepancies. The CFCM employs Maximum Mean Discrepancy (MMD) and Mutual Information Neural Estimation (MINE) to align source and target domains while maintaining feature diversity and relevance. Together, these modules optimize the overall performance and generalization capability of the model by balancing classification, triplet, transfer, and intra-domain diversity losses.

3.2.1 Residual Channel Space Module

To more accurately capture key regions within images and enhance the efficiency of capturing essential information across different channels, we propose the Residual Channel Space Module (RCSM), shown in Fig. 4. It integrates spatial attention [42], channel attention, and a residual structure [43].

(1) Channel Attention

Given a feature map G (which can be G_s, G_{inter}, or G_t), maximum and average pooling operations produce outputs C_{max_out} and C_{avg_out}. These are input into a Multi-Layer Perceptron (MLP) with a ReLU layer, and summed outputs are passed through a Sigmoid function to obtain channel attention weights:

$$CA_{weight} = \sigma(MLP(C_{max\_out}) + MLP(C_{avg\_out})) \tag{1}$$

The original feature map G is channel-weighted by element-wise multiplication with CA_{weight}:$G_{ca} = G \times CA_{weight}$.

(2) Spatial Attention and Residual Structure

For spatial attention, the average S_{avg_out} and maximum S_{max_out} of G_{ca} along the channel dimension are concatenated, transformed through a convolution layer and batch normalization, and passed through a Sigmoid function to obtain spatial attention weights:

$$SA_{weight} = \sigma(BN(Conv([S_{avg\_out},\ S_{max\_out}]))) \tag{2}$$

The spatially adjusted feature map G_{sa} is obtained: $G_{sa} = G_{ca} \times SA_{weight}$. The final output is $G_{out} = G + G_{sa}$.

During the model training process, the attention mechanism dynamically adjusts the channel and spatial responses of the feature map, while the residual structure ensures minimal loss of original information.

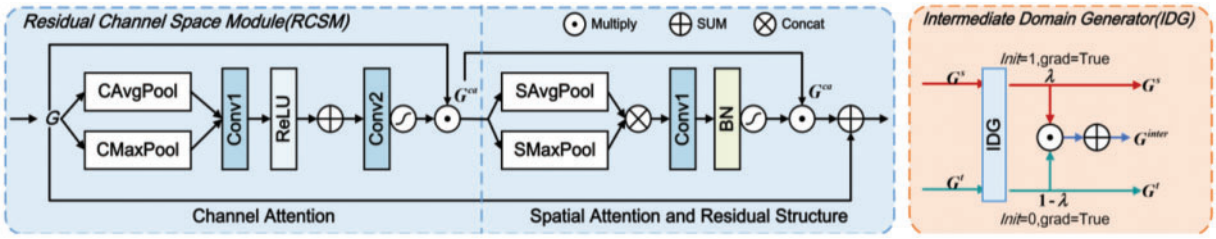

Figure 4: The internal structure of RCSM and IDG

3.2.2 Intermediate Domain Generator

For criterion (1), the Intermediate Domain Generator (IDG) facilitates smooth knowledge transfer. In MTCF, as illustrated in Fig. 4, an intermediate domain feature representation G_{inter} is constructed by blending feature maps G_s and G_t:

$$G_{inter} = \lambda G_s + (1 - \lambda) G_t \tag{3}$$

In Eq. (3), the Intermediate Domain Generator (IDG) generates a gradient-based, learnable parameter λ, initially set to 1, which is optimized through backpropagation and gradient descent to minimize the feature distribution discrepancy between the source and target domains. λ starts at 1, indicating initial dominance by source domain features, and gradually decreases during training, increasing the weight of the target domain features. This dynamic adjustment ensures a smooth transition of feature representation, preventing abrupt changes in feature distribution.

By progressively adjusting λ, this linear blending mechanism creates a continuous spectrum of intermediate domain features, ranging from purely source domain features to purely target domain features. Consequently, the model adapts better to the feature distributions of different domains during training, enhancing its generalization capability and robustness to unseen target domain data. This approach effectively serves as both a computational unit and a facilitator of smooth knowledge transfer, thereby improving the model's predictive accuracy and robustness.

We randomly selected 100 pairs of source and target domain images (with each pair having the same person ID) and generated intermediate domain images. The figure presents the PCA projection of the source domain (blue), intermediate domain (green), and target domain (red). As illustrated in Fig. 5, the intermediate domain data points are positioned between the source and target domains in the feature space, indicating that the intermediate domain effectively integrates the characteristics of both the source and target domains. This generation method of the intermediate domain can effectively mitigate the feature distribution discrepancy between the source and target domains, thereby enhancing the model's generalization capability in cross-domain tasks.

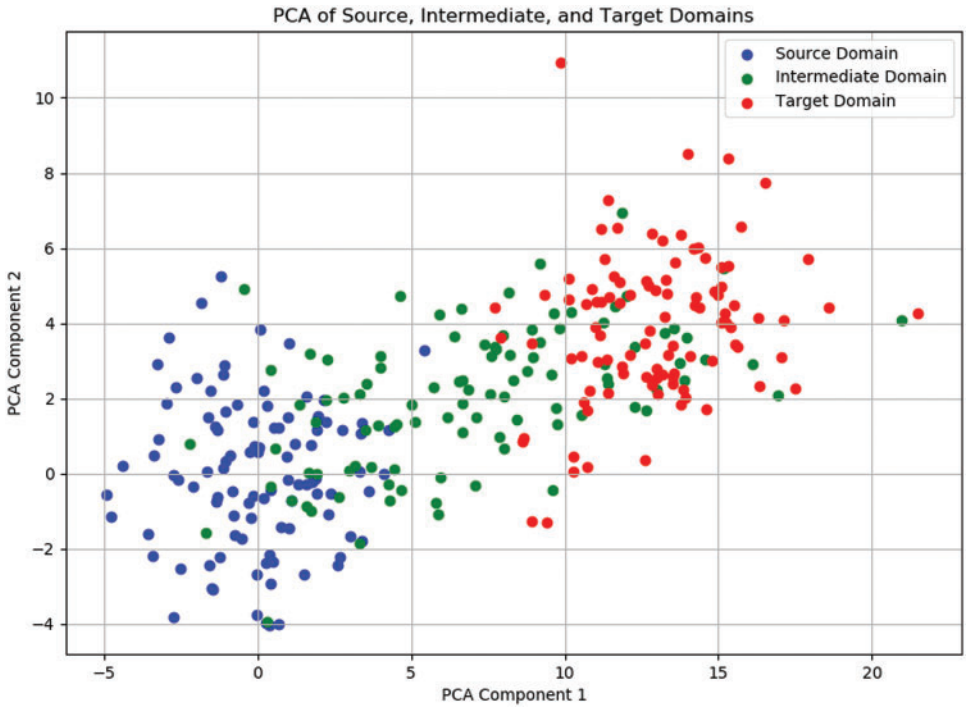

Figure 5: PCA of source, intermediate, and target domains, demonstrating that the intermediate domain (green) effectively integrates characteristics from both the source (blue) and target (red) domains. PCA Component 1 and PCA Component 2 represent the first and second principal components extracted through PCA, capturing the most significant variance in the data

3.2.3 Cross-Domain Feature Constraint Module

The Cross-domain Feature Constraint Module (CFCM) ensures knowledge transfer simplicity using Maximum Mean Discrepancy (MMD) distance loss [44] and maintains feature diversity with Mutual Information (MI) [45].

(1) Quantify Transfer Cost

Maximum Mean Discrepancy (MMD) reduces the distributional discrepancy between domains by minimizing the distance between the mean feature mappings in the RKHS of the source and target domains. Using MMD distance loss, we quantify and minimize knowledge transfer costs, aligning source and target domains. The MMD loss function is formulated as Eq. (4):

$$MMD^2(X_s,\ X_t) = \| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(x_j^t) \|_H^2 \tag{4}$$

where X_s and X_t are source and target domain data, ϕ is a nonlinear mapping to RKHS, and $\| \cdot \|_H$ is the RKHS norm.

Minimizing this discrepancy allows the intermediate domain to transition easily to both extremes, enhancing target domain performance. The objective function is formulated as Eq. (5):

$$minL_{mmd} = min(MMD(f_s,\ f_m) + MMD(f_m,\ f_t)) \tag{5}$$

(2) Diversification of Intra-Domain Features

Addressing criterion (2), We employ Mutual Information Neural Estimation (MINE) to estimate the mutual information (MI) between two random variables x and y. Initially, x and y are concatenated as samples of the joint distribution, approximated as $P(f^s, f^t)$, while randomly permutated x and y are concatenated as samples of the marginal distribution, approximated as $P(f^s)\,P(f^t)$. The MINE module contains two fully connected layers $FC1$ and $FC2$, utilizing the ReLU activation function to introduce non-linear transformations for approximating the logarithmic term within the integral. Subsequently, the empirical mean of the samples is used to compute the approximate value of mutual information, represented as Eqs. (6) and (7):

$$Output_{joint} = FC2(ReLU(FC1(joint))) \tag{6}$$

$$Output_{marginal} = FC2(ReLU(FC1(marginal))) \tag{7}$$

The estimation of mutual information is based on the following formula:

$$MI = E(Output_{joint}) - log(E(Output_{marginal})) \tag{8}$$

The objective function to strengthen similarity is Eq. (9):

$$minL_{mi} = min - (MI(f_s, f_m) + MI(f_t, f_m)) \tag{9}$$

Here, $MI(\cdot)$ represents the mutual information function, and f^s, f^{m} and f^t denote the features of the source, intermediate, and target domains, respectively. Minimizing this function aims to ensure maximum similarity between the intermediate domain and both the source and target domains, thereby promoting diversity and semantic relevance of features within the intermediate domain.

3.2.4 Optimization Objective

In the Minimal Transfer Cost Framework (MTCF), the total loss function L_{total} is derived from a comprehensive consideration of various aspects of the person re-ID task. Specifically, the total loss is a weighted sum of classification loss, triplet loss, transfer loss, and intra-domain diversity loss, as shown in Eq. (10):

$$L_{total} = L_{reid} + L_{tri} + \mu_3(L_{mmd} + L_{mi}) \tag{10}$$

The classification loss L_{reid} for the person re-identification task is the weighted sum of classification losses L_{st} on the source and target domains and L_m on the intermediate domain:

$$L_{reid} = \mu_1 L_{st} + (1 - \mu_1) L_m L_{tri} = \mu_2 L_{tri\_ori} + \mu_2 L_{tri\_xbm} \tag{11}$$

The triplet loss L_{tri} integrates the traditional triplet loss L_{tri_ori} with the triplet loss incorporating the XBM module L_{tri_xbm}:

$$L_{tri} = \mu_2 L_{tri\_ori} + \mu_2 L_{tri\_xbm} \tag{12}$$

L_{mmd} and L_{mi} represent the knowledge transfer loss and intra-domain diversity loss, respectively. Within this framework, our objective is to minimize the total loss value L_{total}.

4 Experiment

4.1 Experimental Settings

4.1.1 Datasets and Evaluation Protocol

To evaluate the effectiveness of the MTCF, we conducted experiments on five datasets: Market-1501 [46], DukeMTMC [47], MSMT17 [32], PersonX [48], and Unreal [49]. Detailed dataset information is presented in Table 1.

Table 1: Detailed information about the five datasets

Dataset	Train		Test			
	Images	IDs	Query	IDs	Gallery	IDs
Market-1501	12936	751	3368	750	15913	751
DukeMTMC	16522	702	2228	702	17661	1110
MSMT17	32621	1041	11659	3060	82161	3060
PersonX	9840	410	5136	856	30816	856
Unreal	120000	2998	–	–	–	–

The "Train" subset is used for model training, while the "Query" and "Gallery" subsets are used for model testing. In person re-ID tasks, the trained model matches each "Query" image with the most similar images in the "Gallery" subset. The model's performance is evaluated using mean Average Precision (mAP) and Cumulative Matching Characteristics at Ranks 1/5/10 (R1/5/10). The mAP measures average retrieval performance across all queries, while R1, R5, and R10 indicate the probability of a correct match in the top 1, 5, and 10 results, respectively. These metrics reflect the precision and recall range of the retrieval.

4.1.2 Implementation Details

The proposed MTCF network is implemented using the PyTorch framework and runs on an NVIDIA 4090 GPU. In our experiments, the batch size was set to 16, with image dimensions of 256 $\times$ 128 and a feature size of 2048. The Adam optimizer was employed with an initial learning rate of 0.00025 and a weight decay of 0.0005. Training was conducted over 60 epochs, with evaluations performed every 1600 iterations (eval_step = 1). The learning rate was adjusted at the end of each epoch using a step size of 20 and a decay factor of 0.1. To perform unsupervised clustering in the target domain, we utilized the DBSCAN algorithm with eps set to 0.6 and min_samples set to 4. Additionally, we incorporated the XBM module to enhance feature learning, setting the memory size to 8192 and the usage ratio to 1. The hyperparameters for the loss function were configured as follows: mu1 = 0.7, mu2 = 1.0, and mu3 = 0.1. To optimize the selection of convolution algorithms, we enabled cuDNN's auto-tuner. The initial learning rate of 0.00025 was chosen to balance stability and efficiency, with higher rates causing instability and lower rates slowing convergence. A batch size of 16 optimizes computational resources and training effectiveness, avoiding memory issues from larger sizes while ensuring frequent parameter updates for better generalization. These hyperparameters were tuned to achieve optimal performance.

4.2 Results

4.2.1 Comparison with State-of-the-Art Methods

In this section, we compare our MTCF method with recent state-of-the-art person re-ID methods. Based on their training schemes, we categorize UDA person re-ID methods into four types: GAN transferring methods, joint training methods, fine-tuning methods, and intermediate domain methods.

Most methods listed in Table 2 largely overlook the importance of intermediate domains, which can serve as a bridge in domain adaptation between the source and target domains to better transfer the source knowledge to the target domain. While some methods consider explicit intermediate domains, they do not account for all possible intermediate domains. However, our MTCF is capable of identifying the most suitable intermediate domain among all possible options to better improve the performance of UDA re-ID. As shown in Table 2, our method significantly outperforms the best UDA re-ID methods in the Duke to Market task in terms of mAP, as well as in the Market to Duke task in terms of mAP, R1/5/10 accuracy, across all these benchmarks. Notably, in the Market to Duke task, our method shows a 2.3% increase in mAP compared to the best performing methods, CCL+PDA+FA and P^{2}LR.

Table 2: Performance comparison of MTCF with state-of-the-art re-ID methods on DukeMTMC→Market-1501 and Market-1501→DukeMTMC tasks

Methods	Reference	Duke→Market				Market→Duke			
		mAP	R1	R5	R10	mAP	R1	R5	R10
SPGAN+LMP [30]	CVPR 2018	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
ECN [50]	CVPR 2018	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
CAIL [51]	ECCV 2020	71.5	88.1	94.4	96.2	65.2	79.5	88.3	91.4
MMT [52]	ICLR 2020	71.2	87.7	94.9	96.9	65.1	78	88.8	92.5
SpCL [53]	NeurIPS 2020	76.7	90.3	96.2	97.7	68.8	82.9	90.1	92.5
LNL [54]	Neurocomputing 2021	75.2	88.9	95.7	97.6	62.5	77.4	88.1	90.6
HC w LP [55]	ICCV 2021	80.0	91.5	–	–	70.1	82.2	–	–
CCL+PDA+FA [56]	ICCV 2021	83.4	**94.2**	–	–	70.8	83.5	–	–
UNRN [57]	AAAI 2021	78.1	91.9	96.1	97.8	69.1	82.0	90.7	93.5
GLT [58]	CVPR 2021	79.5	92.2	96.5	97.8	69.2	82.0	90.2	92.8
IDM [37]	ICCV 2021	82.8	93.2	97.5	98.1	70.5	83.6	91.5	93.7
ICMiF [59]	Information sciences 2022	80.2	92.3	–	–	69.4	83.7	–	–
HDS [60]	Pattern recognition 2022	81.3	92.5	97.4	98.1	69.1	82.0	90.7	93.5
P^{2}LR [61]	AAAI 2022	81.0	92.6	97.4	98.3	70.8	82.6	90.8	93.7
SECRET-Joint (MT) [62]	AAAI 2022	83.0	93.3	–	–	69.2	82.0	–	–
CDCL [63]	Knowledge-based systems 2023	81.5	92.8	**97.6**	**98.7**	70.2	82.7	91.3	93.9

(Continued)

Table 2 (continued)

Methods	Reference	Duke→Market				Market→Duke			
		mAP	R1	R5	R10	mAP	R1	R5	R10
DFDSN-Net [40]	Digital signal rocessing 2023	81.4	92.9	97.0	98.0	68.9	81.9	91.3	93.4
M BDA [64]	JVCI 2024	81.2	92.2	97.0	98.0	66.0	78.8	88.2	91.5
S2ADAP [65]	Knowledge-based systems 2024	82.0	93.5	98.0	98.8	71.8	83.7	91.3	94.1
MTCF (ours)	CMC 2024	**83.5**	92.7	97.4	98.4	**73.1**	**85.1**	**92.5**	**94.7**

In Table 3, MTCF achieves comparable performance in R5/10 accuracy to leading methods in the Market to MSMT task, ranking third in mAP. In the Duke to MSMT task, it demonstrates superior performance, surpassing the most advanced techniques in mAP and R1/5 accuracy.

Table 3: Performance comparison of MTCF with state-of-the-art re-ID methods on the Market-1501→MSMT17 and DukeMTMC→MSMT17 tasks

Methods	Reference	Market→MSMT				Duke→MSMT			
		mAP	R1	R5	R10	mAP	R1	R5	R10
ECN [50]	CVPR 2018	8.5	25.3	36.3	42.1	10.2	30.2	41.5	46.8
CAIL [51]	CVPR 2018	20.4	43.7	56.1	61.9	24.3	51.7	64.0	68.9
MMT [52]	ICLR 2020	22.9	49.2	63.1	68.8	23.3	50.1	63.9	69.8
SpCL [53]	NeurIPS 2020	26.8	53.7	65.0	69.8	26.5	53.1	65.8	70.5
UNRN [57]	AAAI 2021	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
GLT [58]	CVPR 2021	26.5	56.6	67.5	72.0	27.7	59.5	70.1	74.2
IDM [37]	ICCV 2021	33.5	**61.3**	**73.9**	78.4	35.4	63.6	75.5	80.2
ICMiF [59]	Information sciences 2022	25.8	52.4	–	–	25.7	52.5	–	–
HC w LP [55]	ICCV 2021	28.4	54.9	–	–	29.3	56.1	–	–
HDS [60]	Pattern recognition 2022	27.1	52.8	65.2	70.6	29.4	56.8	69.7	74.7
P^2LR [61]	AAAI 2022	29.0	58.8	71.2	76.0	29.9	60.9	73.1	77.9
SECRET-Joint(MT) [62]	AAAI 2022	31.7	60.0	–	–	–	–	–	–
CDCL [63]	Knowledge-based systems 2023	29.5	59.7	71.4	75.9	30.3	61.4	74.8	77.1
DFDSN-Net [40]	Digital signal processing 2023	**33.9**	**61.3**	73.9	78.3	33.0	60.7	73.2	78.0
M BDA [64]	JVCI 2024	26.7	51.4	64.3	68.7	23.4	47.3	59.5	64.5
S2ADAP [65]	Knowledge-based systems 2024	27.0	53.8	64.5	70.2	29.7	58.5	70.6	75.8
MTCF (ours)	CMC 2024	33.0	61.2	**73.9**	**78.8**	**35.5**	**63.7**	**75.6**	80.0

Results from Tables 2 and 3 highlight the exceptional capabilities of our approach in real-to-real re-ID tasks. When compared to recent methods like MBDA and S2ADAP, MTCF excels in multiple metrics, indicating its breakthrough performance in UDA person re-ID tasks.

As shown in Table 4, for the PersonX to Market and PersonX to Duke tasks, our method produced the best results to date in terms of mAP and R1/10 accuracy. In the PersonX to MSMT task, it achieved a mAP of 30.2%, outperforming DFDSN-Net by 1.2%.

Table 4: Performance comparison of MTCF with state-of-the-art re-ID methods on the PersonX→Market-1501, PersonX→DukeMTMC, and PersonX→MSMT17 tasks

Methods	Reference	PersonX→Market				PersonX→Duke				PersonX→MSMT			
		mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
MMT [52]	ICLR 2020	71.0	86.5	94.8	97.0	60.1	74.3	86.5	90.5	17.7	39.1	52.6	58.5
SpCL [53]	NeurIPS 2020	73.8	88.0	95.3	96.9	67.2	81.8	90.2	92.6	22.7	47.7	60.0	65.5
IDM [37]	ICCV 2021	81.3	92.0	97.4	98.2	68.5	82.6	91.2	93.4	**30.3**	**58.4**	**70.7**	**75.5**
CCL+PDA+FA [56]	ICCV 2021	79.6	92.5	–	–	–	–	28.9	53.2	–	–	–	–
CDCL [63]	Knowledge-based systems 2023	77.9	92.4	97.2	98.0	69.5	83.1	91.6	93.3	–	–	–	–
DFDSN-Net [40]	DSP 2023	79.0	91.1	96.7	97.7	67.1	80.8	90.3	93.3	29.0	55.3	68.2	73.3
MTCF (ours)	CMC 2024	**82.7**	**92.7**	97.3	**98.4**	**70.7**	**83.2**	**91.6**	**94.3**	30.2	56.9	69.9	74.7

Table 5 reveals that in Unreal to Market and Unreal to Duke tasks, MTCF method reached the highest current results in mAP and R1/5/10 accuracy. In the Unreal to MSMT task, MTCF achieved results nearly equal to the top-performing IDM method in various metrics.

Table 5: Performance comparison of MTCF with state-of-the-art re-ID methods on the Unreal→Market-1501, Unreal→DukeMTMC, and Unreal→MSMT17 tasks

Methods	Reference	Unreal→Market				Unreal→Duke				Unreal→MSMT			
		mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
JVTC [66]	ECCV 2020	78.3	90.8	–	–	66.1	81.2	–	–	25.0	53.7	–	–
IDM [37]	ICCV 2021	83.2	92.8	97.3	98.2	72.4	84.6	92.0	94.0	38.3	67.3	**78.4**	**82.6**
MTCF (ours)	CMC 2024	**85.2**	**93.8**	**97.7**	**98.5**	**74.6**	**85.5**	**92.9**	**95.0**	**38.4**	**67.3**	78.3	82.4

Overall, MTCF stands out in synthetic-to-real UDA re-ID tasks. Across benchmarks on both PersonX and Unreal datasets, our approach surpasses front-runners like IDM and DFDSN-Net, particularly in mAP and R1 accuracy. These comparisons emphasize the effectiveness and forefront status of MTCF.

4.2.2 Ablation Study

Analyzing Table 6, we integrated the IDG module at different stages of IBN-ResNet50. For Duke to Market, the highest mAP (83.8%) was observed when IDG was inserted after the 1st stage, followed by 83.5% after the 0th stage and 83.4% after the 4th stage. Inserting IDG in the middle stages (2nd and 3rd stages) led to a slight decrease in performance, indicating that integration in the early or late stages is more beneficial. For Market to Duke, early integration of IDG showed better performance, as the mAP value decreased with deeper integration. Our default configuration is to insert IDG after the 0th stage of IBN-ResNet50, which generally produces superior results. We also integrated the RCSM

module at different stages. For Duke to Market, inserting RCSM in the early or late stages achieved better results. For Market to Duke, mid-stage integration of RCSM showed better performance. Our default configuration is to insert RCSM after the 2nd stage of IBN-ResNet50.

Table 6: Performance comparison of IDG and RCSM inserted at different stages of IBN-ResNet50 in the Duke→Market and Market→Duke tasks

IDG	Duke→Market		Market→Duke		RCSM	Duke→Market		Market→Duke	
	mAP	R1	mAP	R1		mAP	R1	mAP	R1
Stage0	83.5	92.7	**73.1**	**85.1**	stage0	84.08	93.0	72.5	83.9
Stage1	**83.8**	**93.1**	73.0	84.6	stage1	83.3	92.8	**73.1**	**85.1**
Stage2	83.1	93.0	71.3	83.3	stage2	83.5	92.7	**73.1**	**85.1**
Stage3	82.3	91.7	71.2	83.8	stage3	**84.2**	**93.3**	72.9	84.5
Stage4	83.4	93.0	71.7	83.1	stage4	84.1	93.2	72.4	84.3

Table 7 compares our method using ResNet50 and IBN-ResNet50 architectures. The performance is generally better on the Duke to Market task than on Market to Duke. This may be attributed to Duke dataset features being more adaptable to Market conditions. IBN-ResNet50 consistently outperforms ResNet50, underscoring its enhanced domain adaptation capabilities.

Table 7: Performance comparison between IBN-ResNet50 and ResNet50 in our method

Method	Target	ResNet50		IBN-ResNet50	
		mAP	R1	mAP	R1
Duke	Market	81.9	92.2	**83.5**	**92.7**
Market	Duke	68.7	81.5	73.1	85.1

Our ablation study in Table 8 evaluates the performance of our model on the Market to Duke re-ID task. The Baseline method, IBN-ResNet50+XBM with only L_{reid}, achieves an mAP of 68.5% and R1 of 82.4%. Incorporating IDG (denoted as 'a') improves mAP to 70.3%, demonstrating IDG's efficacy in intermediate domain and training. This result demonstrates that the generation of intermediate domains effectively mitigates the feature distribution discrepancy between the source and target domains. By gradually adjusting feature representations, IDG achieves a smooth transition from source domain features to target domain features, avoiding abrupt changes in feature distribution. This enhances the model's generalization capability on the target domain, validating the importance of intermediate domain generation in cross-domain person re-identification tasks. Adding the RCSM (denoted as 'b') further enhances performance, with mAP reaching 70.5% and R1 83.2%, confirming the RCS module's positive impact. The subsequent addition of the CFCM (denoted as 'c') brought a significant improvement to the model's performance, with the mAP rising from 70.5% to 73.1% (an increase of 2.6%), and R1 accuracy soaring to 85.1%. This underscores the importance of considering the cost of knowledge transfer and the diversity of features within the intermediate domain, thereby validating the effectiveness of the two criteria we proposed.

Table 8: Ablation studies on different components of our method

Methods	IDG	RCSM	CFCM	mAP	R1
Baseline				68.5	82.4
a	√			70.3	83.0
b	√	√		70.3	83.2
c	√	√	√	**73.1**	**85.1**

4.2.3 Hyperparameter Sensitivity Analysis

In our person re-ID framework MTCF, the three hyperparameters μ_1, μ_2, and μ_3 respectively control the classification loss, triplet loss, and the MMD and MI losses. In Table 9, comparing E1, E5, and E6, it is evident that when the hyperparameters are simultaneously reduced, the performance of the model significantly declines, indicating that lower parameter values negatively impact the model. When we decrease the value of μ_1 (E1, E2), the performance of the model decreases, suggesting that the classification loss has a certain positive effect on model performance. When μ_2 and μ_3 are reduced separately, the performance of the model improves. This result indicates that high weights of triplet loss and MMD and MI losses might impose excessive constraints on the model. Therefore, appropriately reducing the weights of these losses can liberate the model's learning capability, thereby enhancing model performance.

Table 9: Hyperparameter sensitivity experiment results

Experiment	μ_1	μ_2	μ_3	mAP	R1
E1	0.1	0.1	0.1	58.1%	74.8%
E2	0.01	0.1	0.1	8.4%	15.6%
E3	0.1	0.01	0.1	58.4%	74.6%
E4	0.1	0.1	0.01	64%	79.0%
E5	0.01	0.01	0.01	6.7%	13.4%
E6	0.001	0.001	0.01	4.0%	8.3%

4.2.4 Visualization of Model Performance

In the Market to Duke task, our method notably surpasses other models. Fig. 6 illustrates our method's clear advantage in both mAP and R1 accuracy compared to competing technologies. In the left panel, our method achieves the highest mAP of 73.1%, while the closest competitor (CCL + PDA + FA) reaches 70.8%. In the right panel, our method achieves a Rank-1 accuracy of 85.1%, compared to 83.5% by the same competitor. This significant performance improvement highlights the effectiveness of our approach in cross-domain person re-identification tasks.

Furthermore, as shown in Fig. 7, our method consistently outperforms the Baseline in each training iteration in terms of mAP and Rank-1 accuracy, further validating our criteria for intermediate domains. Our model consists of several modules, including the Intermediate Domain Generator (IDG, approximately 0.1 K parameters), the Residual Channel Space Module (RCSM, approximately 0.6 K parameters), and the Cross-domain Feature Constraint Module (CFCM, approximately 13.0 K

parameters), with a total parameter count of around 13.7 K. The synergistic operation and parameter optimization of these modules contribute to the superior performance of our method in feature extraction and cross-domain recognition tasks.

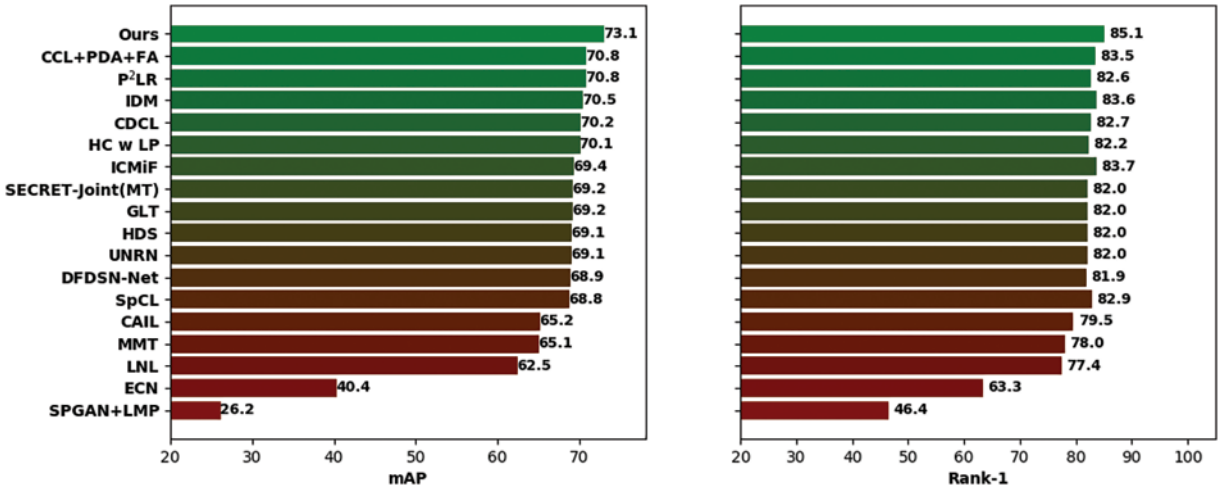

Figure 6: In the Market→Duke task, comparison of our model's mAP and Rank-1 with baseline models

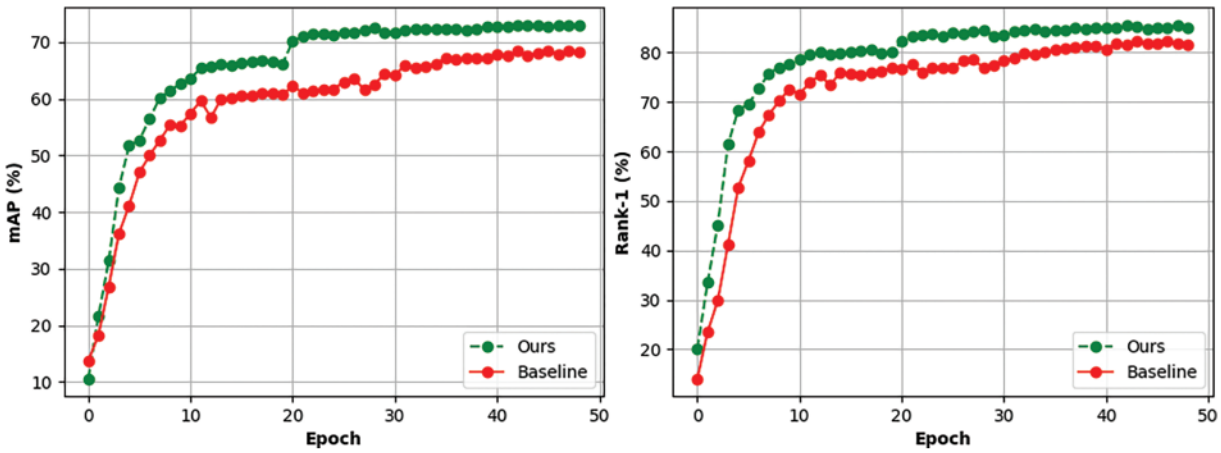

Figure 7: In the Market→Duke task, comparison of each round's mAP and Rank-1 with baseline method

4.2.5 Visualization of Retrieval Results

For the MTCF trained on adapting from Market to Duke, we tested it using three randomly selected images and compared the retrieval results with the Baseline method, see Fig. 8. In each group, the first image is the Query image, with backbone results on the left and MTCF results on the right. Green marks indicate successful retrieval, while red marks indicate failure. Our method shows significant improvement over the Baseline in Rank-10 accuracy, with increases of 60%, 90%, and 100% in each group, respectively. Particularly, observing the third image, when encountering situations with

partial occlusions and low-resolution images, our MTCF method achieves flawless person re-ID, in contrast to the Baseline method.

Figure 8: Examples of randomly selected queries in the Market-1501→DukeMTMC task and their top ten retrieval results (Green and red respectively indicate successful and failed retrieval)

4.2.6 Discussion

Our method, the Minimal Transfer Cost Framework (MTCF), demonstrates significant advantages over baseline methods, particularly in terms of accuracy and generalization capability. The primary strengths of MTCF include its ability to generate a continuous spectrum of intermediate domains, facilitating smoother knowledge transfer between source and target domains. This is achieved through our Intermediate Domain Generator (IDG), which ensures a stable transition and minimizes feature distribution discrepancies. Additionally, the Residual Channel Spatial Module (RCSM) enhances feature extraction, especially in low-resolution images, while the Cross-domain Feature Constraint Module (CFCM) maintains feature diversity and relevance across domains.

However, MTCF is not without its limitations. One notable challenge is the training stability of GAN-based models, which can affect the overall robustness of the framework. Additionally, while our method integrates existing technologies effectively, this integration might be perceived as lacking novelty.

Our innovation lies in the strategic design of the MTCF architecture, which builds upon existing methods to surpass state-of-the-art (SOTA) results. By designing the MTCF framework, we effectively combine simple methods to achieve superior results. This strategic combination ensures that our method not only enhances performance but also provides a robust and efficient solution for unsupervised domain adaptation in person re-identification.

5 Conclusion

In our study, we define two key criteria for an appropriate intermediate domain: (1) The ability to smoothly transition to the source or target domain with minimal cost criterion. (2) Diversity within the intermediate domain's features. We introduce the Minimal Transfer Cost Framework (MTCF) to generate all intermediate domains, ensuring smooth knowledge transfer from the source to the target domain. Within this framework, we firstly propose the Intermediate Domain Generator (IDG),

which considers the full spectrum of intermediate domains, serving as both a computational unit and a facilitator of smooth knowledge transfer between the source and target domains. We also introduce the Residual Channel Spatial Module (RCSM), aimed at enhancing the model's feature extraction capabilities and improving attention to person features in low-resolution images. Subsequently, to reduce the disparity in feature distribution across domains, we propose the Cross-domain Feature Constraint Module (CFCM), which aligns the source and target domains while maintaining feature diversity and semantic relevance. Overall, our method not only aligns with traditional transfer learning theories but also demonstrates its superior performance through extensive experiments on five datasets.

Acknowledgement: We would like to express our gratitude to the creators of the Market-1501, DukeMTMC, MSMT17, PersonX, and Unreal datasets for making these valuable resources available to the research community. These datasets have been crucial for the evaluation of our Minimal Transfer Cost Framework (MTCF). Additionally, we would like to thank the National Key Laboratory of Optical Field Manipulation Science and Technology for their hardware support, which has been instrumental in conducting our experiments.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Sheng Xu and Shixiong Xiang; data collection: Sheng Xu and Feiyu Meng; analysis and interpretation of results: Feiyu Meng and Qiang Wu; draft manuscript preparation: Sheng Xu and Qiang Wu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The additional datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Z. Shi, W. Song, J. Shan, and F. Liu, "Augmented deep multi-granularity pose-aware feature fusion network for visible-infrared person re-identification," *Comput. Mater. Contin.*, vol. 77, no. 3, pp. 3467–3488, 2023. doi: 10.32604/cmc.2023.045849.

[2] C. Feng, D. Han, and C. Chen, "DTHN: Dual-transformer head end-to-end person search network," *Comput. Mater. Contin.*, vol. 77, no. 1, pp. 245–261, 2023. doi: 10.32604/cmc.2023.042765.

[3] B. Tang, X. Xu, F. Dai, and S. Wang, "Person re-identification with model-contrastive federated learning in edge-cloud environment," *Intell. Autom. Soft Comput.*, vol. 38, no. 1, pp. 35–55, 2023. doi: 10.32604/iasc.2023.036715.

[4] L. Wu, C. Shen, and A. Van Den Hengel, "Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification," *Pattern Recognit.*, vol. 65, pp. 238–250, 2017. doi: 10.1016/j.patcog.2016.12.022.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," presented at the 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., San Diego, CA, USA, 2005, vol. 1, pp. 886–893.

[6] X. Liu, H. Wang, Y. Wu, J. Yang, and M. -H. Yang, "An ensemble color model for human re-identification," presented at the 2015 IEEE Winter Conf. Appl. Comput. Vis., Waikoloa, HI, USA, 2015, pp. 868–875.
[7] X. Zhu, X. Zhu, M. Li, V. Murino, and S. Gong, "Intra-camera supervised person re-identification: A new benchmark," presented at the IEEE/CVF Int. Conf. Comput. Vis. Workshops, Seoul, Republic of Korea, 2019.
[8] Y. Chen, X. Zhu, and S. Gong, "Person re-identification by deep learning multi-scale representations," presented at the IEEE Int. Conf. Comput. Vis., Venice, Italy, 2017, pp. 2590–2600.
[9] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 14, no. 4, pp. 1–18, 2018. doi: 10.1145/3243316.
[10] H. -X. Yu, W. -S. Zheng, A. Wu, X. Guo, S. Gong and J. -H. Lai, "Unsupervised person re-identification by soft multilabel learning," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Long Beach, CA, USA, 2019, pp. 2148–2157.
[11] S. S. Raj, M. V. Prasad, and R. Balakrishnan, "Spatial segment-aware clustering based dynamic reliability threshold determination (SSC-DRTD) for unsupervised person re-identification," *Expert Syst. Appl.*, vol. 170, 2021, Art. no. 114502. doi: 10.1016/j.eswa.2020.114502.
[12] X. Lin, P. Ren, C. -H. Yeh, L. Yao, A. Song, and X. Chang, "Unsupervised person re-identification: A systematic survey of challenges and solutions," arXiv preprint arXiv:2109.06057, 2021.
[13] R. Delussu, L. Putzu, and G. Fumera, "Human-in-the-loop cross-domain person re-identification," *Expert Syst. Appl.*, vol. 226, 2023, Art. no. 120216. doi: 10.1016/j.eswa.2023.120216.
[14] L. He and W. Liu, "Guided saliency feature learning for person re-identification in crowded scenes," presented at the Eur. Conf. Comput. Vis., Glasgow, UK, 2020, vol. 16, pp. 357–373.
[15] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-scale feature learning for person re-identification," presented at the IEEE/CVF Int. Conf. Comput. Vis., Seoul, Republic of Korea, 2019, pp. 3702–3712.
[16] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Portland, OR, USA, 2013, pp. 3586–3593.
[17] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi and S. Z. Li, "Salient color names for person re-identification," presented at the Eur. Conf. Comput. Vis., Zurich, Switzerland, 2014, vol. 13, pp. 536–551.
[18] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Boston, MA, USA, 2015, pp. 2197–2206.
[19] G. Wu, X. Zhu, and S. Gong, "Learning hybrid ranking representation for person re-identification," *Pattern Recognit.*, vol. 121, 2022, Art. no. 108239. doi: 10.1016/j.patcog.2021.108239.
[20] H. Fu, K. Zhang, and J. Wang, "An adaptive self-correction joint training framework for person re-identification with noisy labels," *Expert Syst. Appl.*, vol. 238, 2024, Art. no. 121771. doi: 10.1016/j.eswa.2023.121771.
[21] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi and T. S. Huang, "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification," presented at the IEEE/CVF Int. Conf. Comput. Vis., Seoul, Republic of Korea, 2019, pp. 6112–6121.
[22] F. Zhao, S. Liao, G. -S. Xie, J. Zhao, K. Zhang and L. Shao, "Unsupervised domain adaptation with noise resistible mutual-training for person re-identification," presented at the Eur. Conf. Comput. Vis., Glasgow, UK, 2020, vol. 16, pp. 526–544.
[23] Y. Zhai *et al.*, "Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Seattle, WA, USA, 2020, pp. 9021–9030.
[24] H. Sun, M. Li, and C. -G. Li, "Hybrid contrastive learning with cluster ensemble for unsupervised person re-identification," presented at the Asian Conf. Pattern Recognit., New Delhi, India, 2021, pp. 532–546.
[25] X. Jin *et al.*, "Meta clustering learning for large-scale unsupervised person re-identification," presented at the 30th ACM Int. Conf. Multimed., Nice, France, 2022, pp. 2163–2172.

[26] X. Li, Q. Li, W. Xue, Y. Liu, F. Liang and W. Wang, "Confidence-adapted meta-interaction for unsupervised person re-identification," *Appl. Intell.*, vol. 53, no. 21, pp. 25525–25542, 2023. doi: 10.1007/s10489-023-04863-3.
[27] W. Zhang, L. Huang, Z. Wei, and J. Nie, "Appearance feature enhancement for person re-identification," *Expert Syst. Appl.*, vol. 163, 2021, Art. no. 113771. doi: 10.1016/j.eswa.2020.113771.
[28] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," presented at the IEEE Int. Conf. Comput. Vis., Venice, Italy, 2017, pp. 3754–3762.
[29] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, 2018, pp. 79–88.
[30] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, 2018, pp. 994–1003.
[31] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, and F. Bremond, "Joint generative and contrastive learning for unsupervised person re-identification," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Nashville, TN, USA, 2021, pp. 2004–2013.
[32] C. Dai, C. Peng, and M. Chen, "Selective transfer cycle GAN for unsupervised person re-identification," *Multimed. Tools Appl.*, vol. 79, no. 17, pp. 12597–12613, 2020. doi: 10.1007/s11042-019-08604-y.
[33] F. Yang, Z. Zhong, Z. Luo, S. Lian, and S. Li, "Leveraging virtual and real person for unsupervised person re-identification," *IEEE Trans. Multimed.*, vol. 22, no. 9, pp. 2444–2453, 2019. doi: 10.1109/TMM.2019.2957928.
[34] L. Ma, Q. Sun, S. Georgoulis, L. Van Gool, B. Schiele and M. Fritz, "Disentangled person image generation," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, 2018, pp. 99–108.
[35] Y. Choi, M. Choi, M. Kim, J. -W. Ha, S. Kim and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, 2018, pp. 8789–8797.
[36] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "CamStyle: A novel data augmentation method for person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1176–1190, 2018. doi: 10.1109/TIP.2018.2874313.
[37] Y. Dai, J. Liu, Y. Sun, Z. Tong, C. Zhang and L. -Y. Duan, "IDM: An intermediate domain module for domain adaptive person re-ID," presented at the IEEE/CVF Int. Conf. Comput. Vis., Montreal, QC, Canada, 2021, pp. 11864–11874.
[38] Y. Dai, Y. Sun, J. Liu, Z. Tong, Y. Yang and L. -Y. Duan, "Bridging the source-to-target gap for cross-domain person re-identification with intermediate domains," arXiv preprint arXiv:2203.01682, 2022.
[39] J. Na, D. Han, H. J. Chang, and W. Hwang, "Contrastive vicinal space for unsupervised domain adaptation," presented at the Eur. Conf. Comput. Vis., Munich, Germany, 2022, pp. 92–110.
[40] H. Du, L. He, P. Liu, and X. Hao, "Inter-domain fusion and intra-domain style normalization network for unsupervised domain adaptive person re-identification," *Digit. Signal Process.*, vol. 133, 2023, Art. no. 103848. doi: 10.1016/j.dsp.2022.103848.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Las Vegas, NV, USA, 2016, pp. 770–778.
[42] P. Wei, C. Zhang, Y. Tang, Z. Li, and Z. Wang, "Reinforced domain adaptation with attention and adversarial learning for unsupervised person re-ID," *Appl. Intell.*, vol. 53, no. 4, pp. 4109–4123, 2023. doi: 10.1007/s10489-022-03640-y.
[43] T. Shen and H. Xu, "Facial expression recognition based on multi-channel attention residual network," *Comput. Model. Eng. Sci.*, vol. 135, no. 1, pp. 539–560, 2023. doi: 10.32604/cmes.2022.022312.
[44] P. Chen, T. Jia, P. Wu, J. Wu, and D. Chen, "Learning deep representations by mutual information for person re-identification," arXiv preprint arXiv:1908.05860, 2019.

[45] C. Jambigi, R. Rawal, and A. Chakraborty, "MMD-reID: A simple but effective solution for visible-thermal person reid," arXiv preprint arXiv:2111.05059, 2021.
[46] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1116–1124.
[47] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 17–35.
[48] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of viewpoint," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 608–617.
[49] T. Zhang *et al.*, "UnrealPerson: An adaptive pipeline towards costless person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 11506–11515.
[50] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: Exemplar memory for domain adaptive person re-identification," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Long Beach, CA, USA, 2019, pp. 598–607.
[51] C. Luo, C. Song, and Z. Zhang, "Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup," presented at the Eur. Conf. Comput. Vis., Glasgow, UK, 2020, vol. 16, pp. 224–241.
[52] Y. Ge, D. Chen, and H. Li, "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification," arXiv preprint arXiv:2001.01526, 2020.
[53] Y. Ge *et al.*, "Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 11309–11321, 2020.
[54] X. Zhu, Y. Li, J. Sun, H. Chen, and J. Zhu, "Learning with noisy labels method for unsupervised domain adaptive person re-identification," *Neurocomputing*, vol. 452, pp. 78–88, 2021. doi: 10.1016/j.neucom.2021.04.120.
[55] Y. Zheng *et al.*, "Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification," presented at the IEEE/CVF Int. Conf. Comput. Vis., Montreal, QC, Canada, 2021, pp. 8371–8381.
[56] T. Isobe, D. Li, L. Tian, W. Chen, Y. Shan and S. Wang, "Towards discriminative representation learning for unsupervised person re-identification," presented at the IEEE/CVF Int. Conf. Comput. Vis., Montreal, QC, Canada, 2021, pp. 8526–8536.
[57] K. Zheng, C. Lan, W. Zeng, Z. Zhang, and Z. -J. Zha, "Exploiting sample uncertainty for domain adaptive person re-identification," presented at the AAAI Conf. Artif. Intell., Vancouver, BC, Canada, 2021, vol. 35, pp. 3538–3546.
[58] K. Zheng, W. Liu, L. He, T. Mei, J. Luo and Z. -J. Zha, "Group-aware label transfer for domain adaptive person re-identification," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Nashville, TN, USA, 2021, pp. 5310–5319.
[59] J. Huang, H. Ge, L. Sun, Y. Hou, and X. Wang, "ICMiF: Interactive cascade microformers for cross-domain person re-identification," *Inf. Sci.*, vol. 617, pp. 177–192, 2022. doi: 10.1016/j.ins.2022.10.106.
[60] D. Zheng, J. Xiao, Y. Wei, Q. Wang, K. Huang and Y. Zhao, "Unsupervised domain adaptation in homogeneous distance space for person re-identification," *Pattern Recognit.*, vol. 132, 2022, Art. no. 108941. doi: 10.1016/j.patcog.2022.108941.
[61] J. Han, Y. -L. Li, and S. Wang, "Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification," presented at the AAAI Conf. Artif. Intell., Vancouver, BC, Canada, 2022, vol. 36, pp. 790–798.
[62] T. He, L. Shen, Y. Guo, G. Ding, and Z. Guo, "Secret: Self-consistent pseudo label refinement for unsupervised domain adaptive person re-identification," presented at the AAAI Conf. Artif. Intell., Vancouver, BC, Canada, 2022, vol. 36, pp. 879–887.
[63] Q. Tian and J. Sun, "Cluster-based dual-branch contrastive learning for unsupervised domain adaptation person re-identification," *Knowl.-Based Syst.*, vol. 280, 2023, Art. no. 111026. doi: 10.1016/j.knosys.2023.111026.

[64] B. Zhang *et al.*, "A domain generalized person re-identification algorithm based on meta-bond domain alignment," *J. Vis. Commun. Image Rep.*, vol. 98, 2024, Art. no. 104054. doi: 10.1016/j.jvcir.2024.104054.

[65] X. Qu, L. Liu, L. Zhu, L. Nie, and H. Zhang, "Source-free style-diversity adversarial domain adaptation with privacy-preservation for person re-identification," *Knowl.-Based Syst.*, vol. 283, 2024, Art. no. 111150. doi: 10.1016/j.knosys.2023.111150.

[66] J. Li and S. Zhang, "Joint visual and temporal consistency for unsupervised domain adaptive person re-identification," presented at the Eur. Conf. Comput. Vis., Glasgow, UK, 2020, vol. 16, pp. 483–499.