ScienceDirect
www.sciencedirect.com

Knowledge-Based Systems 00 (2023) 1–14

GAN-based Data Augmentation and Pseudo-Label Refinement with Holistic Features for Unsupervised Domain Adaptation Person Re-Identification

Dang H. Pham[a,b], Anh D. Nguyen[a], Hoa N. Nguyen[a,*]

[a] *Department of Information Systems, VNU University of Engineering and Technology, Hanoi 100000, Vietnam*
[b] *Department of Information Technology, University of Khanh Hoa, Khanh Hoa 650000, Vietnam*

Abstract

Person re-identification (re-ID) by using an unsupervised domain adaptation (UDA) approach has drawn considerable attention in contemporary security research. Thus, UDA person re-ID usually employs a model learned from a labeled source domain, adjusted by pseudo-labels, for an unlabeled target domain. However, this method still needs to overcome three main challenges: the significant gap between the source and target domain data, the accuracy of pseudo-labels generated by a clustering algorithm, and the disregarding of important information during the learning process. To tackle these problems, we propose a novel method to improve UDA person re-ID performance by combining GAN-based Data Augmentation and Unsupervised Pseudo-Label refinement methods for training on Target Domain, named DAPRH. In particular, we first use a generative adversarial network (GAN) method to bridge the distribution of the source and target domains. Then we propose a supervised learning approach to maximize the benefits of the virtual dataset. Finally, we utilize pseudo-label refinement and holistic features to enhance the unsupervised learning process. Extensive experiments on three well-known datasets, Market-1501, DukeMTMC-reID, and MSMT17, demonstrate that our DAPRH method can significantly surpass the state-of-the-art performance of the UDA person re-ID.

Keywords: Unsupervised Person Re-Identification, Unsupervised Domain Adaptation, GAN-based Data Augmentation, Pseudo-Label Refinement.

1. Introduction

The increased demand for social security has resulted in the installation of numerous cameras in various positions and angles. It results in an influx of requests for solutions to the person re-ID challenge, which seeks to determine the same person from many perspectives in a distributed multi-camera system. Deep Convolutional Neural Network (DCNN) accomplishments in the presentation of visual characteristics have lately inspired academics to investigate new person re-ID models based on deep learning. Researchers concentrated on building innovative network structures and practical loss functions based on the supervised learning methodology, considerably enhancing the performance of the model and producing exceptional results. However, these supervised learning approaches are data-driven and require a vast amount of labeled data, which incurs substantial labeling expenses. Moreover, person re-ID models are frequently trained on datasets that do not adequately represent real-world scenarios, such as variations in illumination, weather conditions, camera perspectives, etc. Therefore, poor outcomes are unavoidable when these models are utilized with genuine personal information. To solve these challenges, researchers have concentrated on unsupervised learning methods for person re-ID tasks, which enable the model to be trained without the use of labeled datasets.

Most currently person re-ID unsupervised learning approaches rely on clustering algorithms to generate pseudo-labels that can replace truth labels. They can be classified into two main lines: *"Pure Unsupervised Learning"* (USL) and *"Unsupervised Domain Adaptation"* (UDA). In particular, USL is a traditional approach for directly training models on unlabeled sets. However, almost all state-of-the-art (SOTA) methods in this approach, such as pseudo-label refinery [1], primarily rely on multi-GPU

*Corresponding author
Email addresses: dangph@vnu.edu.vn (Dang H. Pham), 19021208@vnu.edu.vn (Anh D. Nguyen), hoa.nguyen@vnu.edu.vn (Hoa N. Nguyen)

machines and large batch sizes to reduce label noise when training from scratch. On the other hand, UDA first trains on labeled datasets to learn about human characteristics before learning on the unlabeled set. This procedure, known as "knowledge transformation", improves the reliability of initial pseudo-labels and enhances performance[2]. Numerous UDA techniques attempt to lessen the feature or image-level distribution disparity between the source and target domains [3, 4]. At the feature-level, we try to gain knowledge of domain-invariant information by analyzing the feature statistics, for example, the mean, covariance, and standard deviation of feature distribution [5]. Regarding the image-level, GAN is used to convert pixel-space samples from the source to the target domain [6, 7]. The initial pseudo-label in early epochs is more reliable and rapidly improves the performance of the person re-ID model in learning on the unlabeled domain, according to the robust pre-trained model. Despite UDA's recent progress, its performance gap compared to supervised learning remains significant, falling short of expectations in the research community.

Research challenges: While researching a novel method, in comparison to earlier works on UDA person re-ID, we face three main challenges, which are as follows:

1. There is a substantial disparity between the source and target domain due to the different surroundings, such as models trained on one domain may perform much worse when applied to another. Thus, there are noticeable differences in lighting and background clutter between photos from various datasets.
2. The dependence on pseudo-labels assigned by the clustering algorithm is a crucial reason for the poor performance in several unsupervised learning methods for person re-ID. These algorithms sometimes generate incorrect labels corresponding to the samples. That results in the fact that several identities allocated the same pseudo-labels, as well as ones that may be suggested to more than one label in the unlabeled dataset. The learning could become ineffective as a result.
3. Normally, we only concentrate on global information encoded in feature vectors extracted from the feature map by the global average pooling layer (GAP) [8]. Although this can make the model reach a relatively high discriminative performance, this ignores some valuable information that local regions can reveal. Therefore, we seek methods to utilize information from global and local regions to enhance training effectiveness.

Contribution highlights: From the challenges above, the primary objective of this research is to overcome the inevitable limitations of the clustering algorithm and find a way to alleviate the gap between the two data domains. Our key contributions are summarized below:

- We suggest utilizing a GAN method in order to make a virtual dataset that bridges the distribution of two data domains. Then we can utilize this dataset to make the pre-trained model more generalized.
- We propose a supervised learning approach to maximize the benefits of the virtual dataset while minimizing the impact of low-quality samples generated by GAN models.
- We elaborate a simple pseudo-label refinement mechanism to enhance the unsupervised learning process.
- We provide an approach to utilize information from partial regions of the image to intensify learning from the dataset.
- We carry out extensive experiments on three well-known benchmarks, Market-1501, DukeMTMC-reID, and MSMT17, to evaluate and confirm our proposed method's performance. The result demonstrates the practical potential of the research direction. Our method outperforms the performance of some SOTA achievements.

The remainder of this work is structured as follows. Section 2 provides the literature reviews in UDA person re-ID. Section 3 introduces our proposed method. In Section 4, we present our experiments, evaluate our proposed work, and compare the results with recent publications. Finally, Section 5 summarizes our contributions and identifies potential future works.

2. Literature Reviews

In order to justify clearly three challenges of UDA, this section is mainly focused on recent works of (i) unsupervised domain adaptation, (ii) GAN-based data augmentation, (iii) pseudo-label noise reduction, and (iv) part-based learning for person re-ID.

2.1. Unsupervised Domain Adaptation for Person Re-ID

Almost all existing UDA methods for person re-ID use two steps: pre-training with a labeled source domain and training on an unlabeled dataset and pseudo-labels developed via clustering algorithms such as DBSCAN [9] and Kmeans [10]. While the first step allows the re-ID model to take advantage of the labeled dataset to build a strong base, the second step aims to learn a new environment where the model would be deployed. According to that, to bridge the performance gap between UDA and supervised learning, recent studies can be categorized into two branches: (1.) distribution aligning methods and (2.) clustering-based methods. Because UDA usually works on the idea that the source and the target domain are not very different from each other, the distribution aligning approach believes that a model that learns well enough on the labeled dataset, which is similar to the target domain (environments, view angles, etc.) will be able to perform well on it. To do that, [11] proposes two

RDSBN and MDIF modules to reduce the source-target domain gaps. Currently, methods based on Generative Adversarial Network (GAN) [12] to minimize the distance between the source and target domains also draw much attention. Similarly, recent studies on clustering-based methods aim to enhance the quality of the pseudo-label, which allows replacing traditional labels completely and achieves remarkable results. We will discuss more about both approaches in the next sections.

2.2. GAN-based Data Augmentation

According to [13], data augmentation methods based on GAN were often conditionally employed on some factors: *pose, illumination, camera style, background* and *genetic structure*.

(1.) Pose: to provide representations for single-domain supervised ReID that is pose-independent, FD-GAN [14] and PN-GAN [15] produced the new poses of a target person under the supervision of 2D poses. It was later suggested to handle unsupervised domain adaptive (UDA) ReID via similar pose transfer. **(2.) Dataset style (illumination)**: Because datasets are often recorded in uniform illumination, [16] utilized CycleGAN [17] to reduce the gap between two domain datasets by generating human pictures in the target domain's style. **(3.) Camera style**: Rather than utilizing a generic dataset style, CMFC [7] and CamStyle [18] transformed photos from one camera are transformed into those from another in order to close the stylistic gaps between different cameras. While CMFC utilizes StarGAN, CamStyle is built based on CycleGAN. **(4.) Background**: SBSGAN [19] was designed to remove and switch the backdrop of a human picture in order to reduce the background effect on UDA ReID. **(5.) Generic structure**: Through the process of applying recoloring gray-scaled human photographs with other images' color distribution, DGNet [20] and DG-Net++ [21] trained disentangled identity representations invariant to structural variables. Inspired by the achievement in [7], we provide a cross-domain camera style transformation module that takes pictures from the source domain to various cameras in the target domain while explicitly considering camera-level discrepancies. In addition, we present a powerful method for supervising the DCNN model by combining real and fake datasets.

2.3. Pseudo-Label Noise Reduction for Unsupervised Person re-ID

In clustering-based algorithms, due to the imperfect quality of pseudo-labels, handling noisy pseudo-labels has recently been an appealing research topic. MMT [22], MEB-Net [23] suggest leveraging numerous predictions from auxiliary networks to improve the pseudo-labels. UNRN [24] proposes an uncertainty estimation module to investigate the accuracy of each sample's pseudo-label. The part-based pseudo-label refining proposed by both PPLR [1] and SECRET [25] employs additional trustworthy supplementary information to increase the quality of the pseudo-labels. While MCRN [26] uses numerous centroids to proactively identify and represent multiple subclasses that may exist in a cluster., GLT [27] uses a group-aware label transfer technique to reference the pseudo-labels online. CooRL framework [28] employs a multi-branches structure with a refinement method to prevent noisy labels and leverage the potential benefits of outliers. RMCL [29] defines the idea of probabilistic stability and offers a technique for estimating stability to enhance the model of pseudo-labels' dependability. In addition, RESEL [30] uses ensemble learning to estimate the reliability of samples, select those that are deemed reliable, and apply a weight to the re-ID loss function to mitigate the impact of noisy labels that may arise from clustering.

In our work, we base on the CGL method [31], which merely utilizes the relationship of feature vectors to alleviate harms brought on by label noise in training.

2.4. Part-based Learning for person re-ID

Cross-entropy loss and batch-hard triplet loss [32] are two well-known losses used to train re-ID models employed in part-based learning [33, 7, 34, 35, 36] for learning local representations. Specifically, besides GAN based-data augmentation, CMFC [7] splits two partial feature branches and learns partial features using only Triplet loss. On the other hand, [33] utilizes these losses on a fused embedding, which combines global and local features through a learnable Fusion Module to avoid confusing learning of several false labels. BPBreID [34] suggests using triplet loss and identity loss on the part-concatenated feature as a holistic embedding; however, it is applied in supervised learning. In our works, we also utilize holistic embedding to learn local representations but only use the identity loss function on it. Combining learning global representation, which is trained by both identity and triplet losses, shows significant improvement in performance compared to other strategies.

Table 1 summarizes related studies in the following factors: method, approach summary, learning method category, benchmarks, and corresponding results. In this table, the names of benchmarks are abbreviated, consisting of Market-1501 [37] (**M**), DukeMTMC-reID [38] (**D**), MSMT17 [39] (**MT**) and some domain adaption tasks (**D2M, M2D, D2MT, M2MT**) based on these datasets. Results, which those methods attained, also are presented in this table. Note that, we just report the results using ResNet50 and equivalent DCNN models on mAP metrics [40] for short. We will discuss those benchmarks and metrics in the section 4.

3. Proposed Method: DAPRH

This section is dedicated to defining the UDA problem and outlining the two-stage approach we employ to tackle it: Supervised training in Source Domain and Unsupervised Training in Target Domain. In particular, we describe

Table 1. Summary of some mentioned methods in the Related Works section. SL, PUSL, UDA denote Supervised Learning, Pure Unsupervised Learning, and Unsupervised domain adaptation, respectively.

Method	Summary	Type	Benchmarks	Results (mAP %)
RDSBN [11]	Apply two modules: RDSBN and MDIF to alleviate the domain gaps.	UDA	D2M, M2D, D2MT, M2MT	81.5, 71.5, 33.6, 30.9
FD-GAN [14]	Produce images of a target person in various poses to diversify the source domain.	SL	M, D	77.7, 64.5
PN-GAN [15]		SL	M, D	72.6, 53.2
Cycle-GAN [17]	Provide the illumination translation between two different domains.	UDA	D2M, M2D	77.8, 68.2
CamStyle [18]	Focus on transferring camera styles from the source to the target domains.	SL	M, D, MT	71.6, 57.6, 52.3
SBSGAN [19]	Change background of image from the source domain to lessen the influence of the background variant.	UDA	D2M, M2D	27.3, 30.8
DG-NET [20]	Recolor the human images with the other images' color distribution.	SL	M, D	86.0, 74.8
DG-NET++ [20]		UDA	D2M, M2D, D2MT, M2MT	61.7, 63.8, 22.1, 22.1
MMT [22]	Utilize the predictions from several networks to generate pseudo-labels.	UDA	D2M, M2D, D2MT, M2MT	71.2, 65.1, 23.3, 22.9
MEBNET [23]		UDA	D2M, M2D	76.0,66.1
UNRN [24]	Propose an uncertainty estimation module that investigates the accuracy of each sample's pseudo-label.	UDA	D2M, M2D, D2MT, M2MT	78.1, 69.1, 26.2, 25.3
GLT [27]	Propose a group-aware label transfer technique to improve pseudo-labels' accuracy.	UDA	D2M, M2D, D2MT, M2MT	79.5, 69.2, 27.7, 26.5
PPLR [1]	Combine the information from local regions of an image to refine the pseudo-labels.	PUSL	M, MT	81.5, 31.4
SECRET[25]		UDA	D2M, M2D, M2MT	83.0, 69.2, 31.7
MCRN [26]	Introduce multiple centroids learning to improve the learned representations to label noise.	UDA	D2M, M2D, D2MT, M2MT	83.8, 71.5, 35.7, 32.8
CooRL [28]	Develop a new pseudo-label refinement mechanism with a multi-branches architecture.	UDA	D2M, M2D, MT2D, MT2M	74.3, 65.2, 65.7, 77.1
RESEL [30]	Propose a new ensemble learning framework utilizing the consistency of different predictions to select reliable samples.	UDA	D2M, M2D, D2MT, M2MT	83.1, 72.3, 33.6, 34.2
BPBreID [34]	Supervised learning based on global features, local features, and holistic features, which are local features-concatenation	SL	M, D	87.0, 78.3
LF_2 [33]	Combine global and local features by a learnable Fusion Module.	UDA	D2M, M2D	83.2, 73.5
CMFC [7]	Utilize both GAN-based augmentation and partial learning to improve the re-ID model.	UDA	D2M, M2D	81.0, 71.2

in detail our proposed method, DAPRH, based mainly on GAN-based data augmentation and pseudo-label refinement methods with holistic features for unsupervised person re-ID. Consequently, all important ideas shaped by DAPRH are specified in the subsequent subsections.

3.1. Problem Statement

The goal is to use a model $\mathbf{\Omega}$ trained on a labeled dataset (source domain) to transfer knowledge and learn new features on an unlabeled dataset (target domain). Specifically, we assume that the labeled dataset is $\mathcal{D}_s = \{x_i^s, y_i^s\}_{i=1}^{\mathcal{N}_s}$, where (x_i^s, y_i^s) denote the i-th labeled sample, $\mathcal{N}_s$ denotes the quantity of source-domain samples. Similarly, we also denote $\mathcal{D}_t = \{x_i^t\}_{i=1}^{\mathcal{N}_t}$ as the target domain, where x_i^t and $\mathcal{N}_t$ correspond the i-th unlabeled sample and the length of this dataset. Clearly, the target domain dataset's photos are not attached with identification labels. In light of this, we can formulate the re-ID problem as follows:

$$\mathbf{\Omega} = \mathcal{F}_{USL}\left(\mathcal{F}_{SL}(\mathbf{\Omega}, \mathcal{D}_s), \mathcal{D}_t\right) \tag{1}$$

3.2. Approach Direction

To solve the challenges described in Section 1, we approach a novel method with the following ideas: (i) mitigating the enormous disparity between the source and target domains by utilizing the GAN-based data augmentation, (ii) adding a local branch to extract local features to improve the learning process and (iii) applying the unsupervised pseudo-label editation methods for training on the target domain. We call the proposed

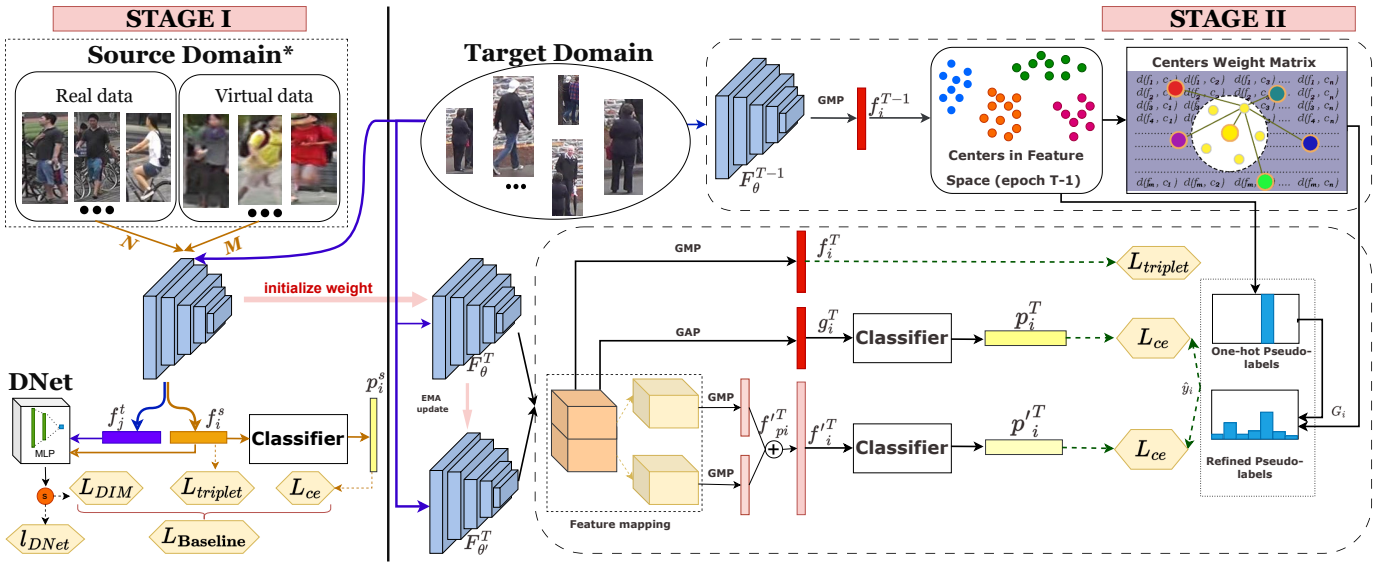

Figure 1. Overall architecture of our proposed DAPRH.

method DAPRH standing for integrating GAN-based **D**ata **A**ugmentation and **P**seudo-Label **R**efinement methods with **H**olistic features for UDA . The DAPRH method is illustrated in Fig. 1 by two stages described in detail as follows.

Stage I: Supervised Training in Source Domain. Because our goal is to transfer the knowledge of the DCNN model trained with labeled data to build the fundamental basis for learning on the unlabeled domain, we first trained our model on the labeled source dataset as **baseline**. Traditionally, we use simultaneously cross-entropy loss [41] for id loss and batch-hard triplet loss [42] for feature loss. Specifically, we denote $y_i \in \mathcal{R}^{\mathbf{N}}$ as the label for the image x_i, where $\mathbf{N}$ being the number of grouped clusters, then the cross-entropy loss is written as:

$$\mathcal{L}_{ce} = -\frac{1}{N_c}\sum_{i=1}^{N_c}\sum_{k=1}^{K}\mathbf{y}_{i,k}\log\left(\mathbf{p}_{i,k}\right) \tag{2}$$

where N_c is the number of clustered examples, $\mathbf{p}_i = H_\phi\left(\mathbf{f}_i\right) \in \mathbb{R}^K$ is the identity prediction vector, H_ϕ is the classifier head implemented by a fully connected layer, $\mathbf{y}_{i,k}$ and $\mathbf{p}_{i,k}$ are the k-th elements in $\mathbf{y}_i$ and $\mathbf{p}_i$, respectively. Similarly, the triplet loss can be formulated as follows:

$$\mathcal{L}_{tri} = \sum_{f_a,f_p,f_n}\left[m + D(f_a,f_p) - D(f_a,f_n)\right]_+ \tag{3}$$

where f_a denotes the feature anchor vector. f_p and f_n represent the hardest positive and hardest negative samples of the *i-th* image in the same batch respectively. $D(.)$ is a function computing the Euclidean distance of two feature vectors, and m is a margin hyperparameter.

Stage II: Unsupervised Training in Target Domain. In the same way as previous works, we utilize clustering algorithms for generating pseudo-labels corresponding to each sample in the dataset. It will group the training image features $\mathcal{F} = \{\mathbf{f}\}_{i=1}^{N}$ into clustered inliers and un-clustered outliers, where the outliers are discarded in the following training epochs. Then, we use these cluster assignments as labels for the samples and train the person re-ID model in a supervised manner with the cross-entropy loss and the triplet loss, which is similar to stage I. Besides, because of the small number of samples chosen by DBSCAN in early epochs, which easily makes the model become over-fitting, we follow the mean teacher architecture [43] as illustrated in stage II of Fig. 1.

Note that our method differs from the previous UDA person re-ID studies mainly in three aspects:

1. Camera Style Adaption: Inspired by [7, 6, 44, 6], we utilize StarGAN [45] to lessen the difference in sample distribution between the source domain and the target domain as described in Section 3.3. Based on that, we can train robust pre-trained models in stage I.
2. Pseudo-Label Refinement: Instead of directly using pseudo-labels generated by DBSCAN[9], we conduct a score based on the distance relationship among samples and all cluster centers in feature space to refine the pseudo-label. Details of this stage are described in Section 3.5.
3. Holistic Features: Besides global features, we also utilize information from partial regions of images to improve learning performance. Details of this stage are described in Section 3.4.

The combination of these two above stages is coordinated by a mean teacher framework to optimize the person re-ID models. This framework is regulated by an overall loss function for the entire training process.

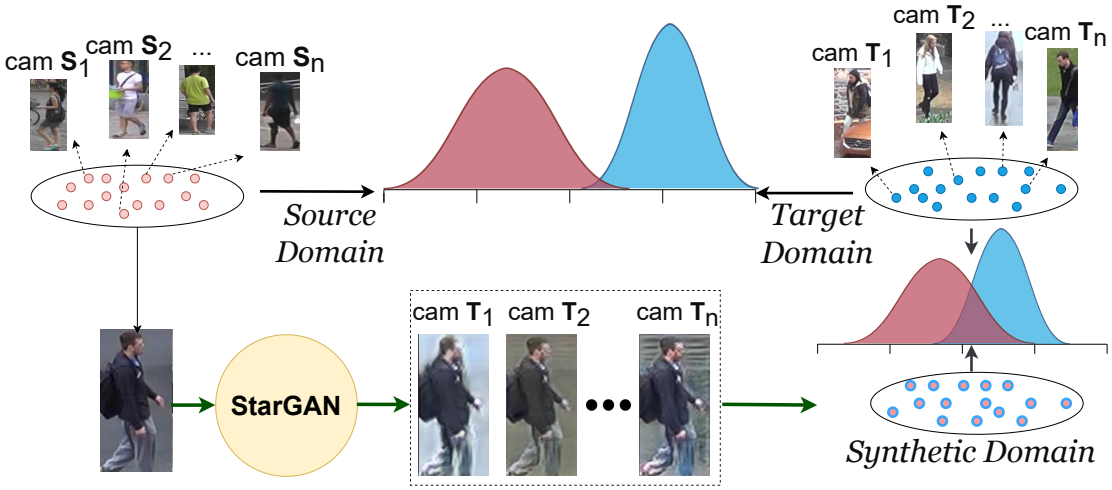

Figure 2. GAN-based Data Augmentation.

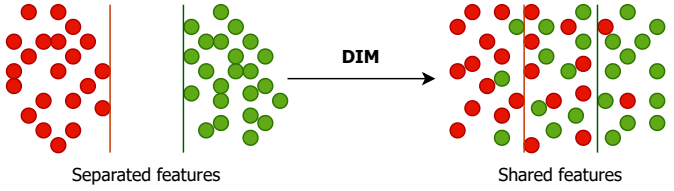

Figure 3. DIM: Feature distribution discrepancy reduction.

Figure 4. Low-quality samples generated by StarGAN.

3.3. GAN-Based Training Data Augmentation

3.3.1. Virtual Dataset Generation

Our goal is to minimize the distribution discrepancy between the source and target domains caused by the different styles of cameras. We consider a camera as an individual domain, resulting in C+1 different style-domains for an unlabeled dataset with images collected from C camera views and a labeled dataset. In order to accomplish this objective, we train image-image translation models using StarGAN [45] to learn the style mapping between those domains, as shown in Fig. 2. Specifically, we want each image (x_i^s, y_i^s) from $\mathcal{D}_s^*$ to be converted to a different camera style in C camera styles of the target domains, while keeping the label information during the translation.

StarGAN has a substantial advantage of merely training a single Generator G to transform an sample (x_i^s, y_i^s) into a synthesized sample (x_i^t, y_i^s) corresponding to the camera label c in the target dataset, $G((x_i^s, y_i^s), c) \rightarrow (x_i^t, y_i^s)$. Besides, StarGAN uses an auxiliary classifier to improve Generator G, which allows a single discriminator to control multiple domains $D : x \rightarrow D_{src}(x), D_{cls}(x)$. The overall objectives to optimize Generator G and Discriminator D are written, respectively, as below:

$$\begin{aligned} \mathcal{L}_D &= -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^r \\ \mathcal{L}_G &= \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^f + \lambda_{rec}\mathcal{L}_{rec} \end{aligned} \tag{4}$$

where λ_{cls} and λ_{rec} are hyperparameters that regulate the relative impact of domain classification and reconstruction losses in comparison to adversarial loss. In our work, we use $\lambda_{cls} = 1$ and $\lambda_{\text{rec}} = 10$.

3.3.2. Training Coarse Deep Re-ID Model

The style-transferred dataset $\mathcal{D}_{st}$ can be used in supervised learning alongside the labeled dataset $\mathcal{D}_s$ because the Generator produces samples with the same labels as the original images. However, we notice the noise in the style-transferred dataset, for example, low-quality images as shown in Fig. 4. These images lost almost all discriminative characteristics of identities; hence, the re-ID model can confuse the model in learning feature representation. To reduce the noise's effect, batches are composed of real and fake (style-transferred) images with a corresponding ratio N:M. Note that, from now when we refer to the source domain, it means the mixed dataset between the real images and fake images $\mathcal{D}_s^*$.

In addition, we employed the Domain-Invariant Mapping (DIM) method [46] to alleviate the feature distribution discrepancy between the source and target domains at the feature-level. DIM comprises a feature extractor, which is a DCNN model, and a domain classifier called DNet. The feature extractor learns to present domain-invariant features, while the domain classifier ensures the production of discriminative features for person re-ID. This is illustrated in Fig. 3. DNet uses both source and target domain information as input, and produces domain recognition scores in the range [0,1]. The target value of the domain

recognition scores is 1 for f_i^s from the source domain $\mathcal{D}_s^*$ while 0 for f_j^t from the target domain $\mathcal{D}_t$. By using Mean Square Error, DNet will be optimized by the loss function written as:

$$l_{DNet} = \frac{1}{N_s}\sum_{i=0}^{N_s}(DNet(f_i^s)-1)^2 + \frac{1}{N_t}\sum_{j=0}^{N_t}(DNet(f_j^t))^2 \quad (5)$$

Then, using the DIM loss to extract features with 0.5 domain recognition scores, we use DNet to monitor *Omega* DCNN models in an effort to deceive DNet with the DIM objective loss as follows:

$$\begin{aligned}\mathcal{L}_{DIM}(\Omega) = &\frac{1}{N_s}\sum_{i=0}^{N_s}(DNet(f_i^s) - 0.5)^2 \\ &+ \frac{1}{N_t}\sum_{j=0}^{N_t}(DNet(f_j^t - 0.5))^2\end{aligned} \quad (6)$$

Now, the final objective loss for supervising the DCNN model is formulated as:

$$\mathcal{L}_{SL} = \mathcal{L}_{ce} + \mathcal{L}_{tr} + \lambda * \mathcal{L}_{DIM} \quad (7)$$

where λ is loss weight. In every iteration, DNet will first be trained with Eq. 5. Then, we train re-ID models Ω with Eq. 7.

3.4. Global and Part-Learning for Person re-ID

From the achievements of previous works in 2.4, we not only concern information from the global features but also utilize information from local parts of images to obtain the target images' similarities to improve pseudo-labels. As illustrated in stage II of Fig. 1, our DCNN model includes two branches sharing the same backbone:

Global feature branch. To get global features of an image x_i at epoch T, first, we need to extract the shared feature map $F_\theta^T(x_i) \in \mathcal{R}^{C\times H\times W}$, where C, H and W are indicative of the channel, height, and width, respectively. Given this shared representation, unlike some previous works, to get the global feature f_i^T, we add an extra-global max pooling layer (**GMP**) [8] parallelly with a global average pooling layer (**GAP**) over the feature map. The GMP vector will be utilized for clustering and learning through the triplet loss, while the GAP feature will be used for the identity classifier loss Then, the extra-global feature g_i^T outputted by that branch is then fed into the classifier and utilized for computing the identity loss as presented in the bottom part of stage II of Fig. 1. Since there is more discriminative information about personal photos in the regions with higher response values, this method enables our model to exploit valuable information collected by the max pool layer.

Partial feature branch. Besides the global information, we also consider benefits coming from the local information extracted from partial regions of images. Thus, like [1, 33, 7], we can obtain the partial features by horizontally dividing the feature map into K uniformly partitioned regions and then applying GMP layers. Then, we have a set containing K partial features $\{f'_{p0}, f'_{p1}, ..., f'_{pk} | f'_{pi} \in \mathcal{R}^{C\times \frac{H}{K}\times W}\}$. Due to the non-discriminative local appearance in many samples, we do not directly apply the aforementioned loss functions across these component features in contrast to some prior research. Thus, we concatenate K part feature to a local holistic embedding f'_i. This feature is then exploited by merely the identity loss and the same label used for the global branch. We introduce this approach for two targets: (i) to furnish the model with comprehensive local information; (ii) to reduce the number of tuning steps for partial branch optimization. In this work, K is set to 2. On the other hand, we can train local and global branches simultaneously by the identity loss:

$$\mathcal{L}_{id}(p_i, p'_i, y_i) = \mathcal{L}_{ce}(p_i, y_i) + \mathcal{L}_{ce}(p'_i, y_i) \quad (8)$$

Figure 5. Example for the label noise problem.

3.5. Centroid-based Pseudo-Label Refinement

Centroid-based Pseudo-Label Refinement. One of the inevitable problems when using a clustering algorithm for generating pseudo-labels is noise labels. As presented in Fig. 5, when individuals wear the same clothing colors, the clustering algorithm, namely DBSCAN, can occasionally be fooled. Then, it makes inconsistent clusters, for instance, the Red cluster consists of four different persons with yellow skirts, as well as the problem in the Blue cluster. As a result, samples assigned incorrect labels limit the model's training. Therefore, to reduce side-effect of the noise labels, we will modify the basic one-hot pseudo-label y_i as follows:

$$\hat{\mathbf{y}}_i = (1-\alpha)\mathbf{y}_i + \alpha\mathbf{G}_i \quad (9)$$

with $\hat{\mathbf{y}} \in \mathbb{R}^K$ defined as the refined pseudo-label, it is a soft target label (one-soft) instead of the original fixed label (one-hot). Furthermore, $\mathbf{G}_i$ is a score vector indicating a distance relationship between the $i-th$ feature vector and all centroids in the feature space. The goals labeling encourages samples to approach not only the initially assigned centroid but their neighbor centroids where their identity information is potentially embedded.

In order to calculate $\mathbf{G_i}$, we built an overall graph of the initial cluster centers and feature vectors. Let the matrix

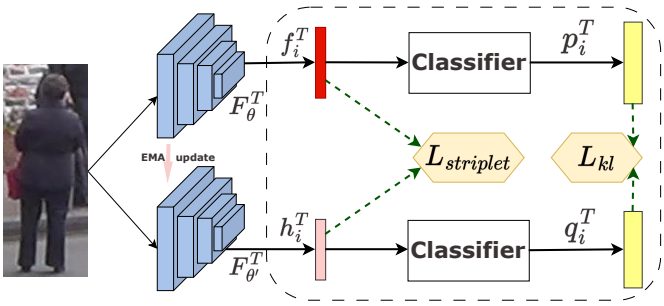

Figure 6. Mean Teacher Architecture.

G represent that diagram, with

$$G_{i,j} = \frac{exp(d_E(f_i, m_j)/\tau)}{\sum_{j=1}^{|C|} exp(d_E(f_i, m_j)/\tau)} \tag{10}$$

is the Euclidean distance score between the feature vector f_i corresponding the *i-th* sample from $\mathcal{D}_t$ and the cluster center vector m_j corresponding the *j-th* cluster from the initial centroid set $\mathcal{C}$, which calculated by the formula:

$$m_j = \frac{1}{|\mathcal{C}_j|} \sum_{f_t \in \mathcal{C}_j} f_t \tag{11}$$

with $\mathcal{C}_j$ denotes the clustered image set corresponding to the $j-th$ cluster. $|.|$ denotes the operation to get the number of items in the set and τ is the temperature hyperparameter empirically set to 0.05. This refinement strategy helps to improve the diversity of the refined pseudo-label $\hat{y}$. For a sample in the frontier of its cluster, our strategy will reduce the confidence of learning from its one-hot label since it has a high probability of incorrect labels. Moreover, the model will be learned to make the diversity of features by enlarging the difference between this sample and the others belonging to its cluster, which are close to the centroid.

In addition, because of the inter-noise samples in the cluster, as mentioned, the center vectors calculated directly with the expression in Eq. 11 may be unreliable. Therefore, [31] proposes to utilize Silhouette score[47] to evaluate the correlations between samples and cluster centers, which is formulated as follows:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \in [-1, 1] \tag{12}$$

where a_i denotes the average distance between the *i-th* feature vector and the others in the same cluster initialized by DBSCAN and b_i is the distance of that vector to the corresponding nearest neighboring cluster in the feature space. A sample will be considered to be close to points in the same cluster and far away from points in other clusters if the silhouette value is high. In other words, its label is high reliability, and Eq. 11 can be rewritten as:

$$m_j = \frac{1}{|S_j|} \sum_{f_i \in S_j} f_i \text{ with } S_j = \{f_t \in \mathcal{C}_j \mid s_t > \sigma\} \tag{13}$$

Here, a confidence subset S_j is selected from the original cluster $\mathcal{C}_j$ according to the threshold σ. This filtering will alleviate the impact of low-quality images or early background clutter, which may belong to other identities. With the natural intuition that samples having negative Silhouette scores are typically found at the frontier of their cluster and have a high likelihood of being labeled incorrectly, σ is set to 0.

Note that we do compute the cluster centers and the center weight matrix **G** by the beginning of each epoch T as shown in Fig. 1.

3.6. Mean Teacher Architecture

As illustrated in Fig. 1, in stage II, we train two models simultaneously, which assumes a dual role as a teacher $F_{\theta'}$ and a student F_θ. With each image batch fed, the student will learn new knowledge from inputs, but the teacher is updated less often based on the student's weights. The main goal of this design is to educate the student model to consistently give results that match those of the instructor model, even when the input data is skewed or noisy.

Furthermore, we use two loss functions to allow the student to learn from the teacher are: Kullback–Leibler divergence loss [48] $\mathcal{L}_{kl}(p_i, q_i)$ and Soft-Triplet loss $\mathcal{L}_{stri}(f_i, h_i)$. In these losses, p_i, q_i, f_i, and h_i represent the output of the classification class and feature vectors extracted from the student and EMA teacher model as presented in Fig. 6. Thus, they enable to evaluate the disparity in the distribution of predictions between the teacher and the student. Then, the teacher model θ' is learnt by the student network θ at training step t by the *exponential moving average strategy*:

$$\theta'_t = w * \theta'_{t-1} + (1 - w) * \theta_{t-1} \tag{14}$$

This ensures that the teacher model remains stable even during the training process.

3.7. Overall Loss Function

Finally, to optimize the re-ID models (student model), we utilize the identity loss (Eq. 8), triplet loss (Eq. 3) to learn new information from the dataset, and two loss functions to learn from teacher model. With a refined pseudo-label, the formulation of the objective can be written as:

$$\begin{aligned}\mathcal{L}_{\text{USL}} = (1 - w_1)\mathcal{L}_{id}(p_i, p'_i, \hat{\mathbf{y}}_i) + w_1\mathcal{L}_{kl}(p_i, q'_i) \\ + (1 - w_2)\mathcal{L}_{tri}(f_i, y_i) + w_2\mathcal{L}_{str}(f_i, h_i)\end{aligned} \tag{15}$$

The two weights w_1, w_2, which are adjusted to 0.4 and 0.6, are used to manage learning from the mean teacher and learning new information. With p_i, p'_i and q_i represent the classification class output from the global, local, and global branches of the teacher, respectively. f_i and h_i are global features extracted from student and teacher models.

4. Experiments & Evaluation

In order to prove the effectiveness of DAPRH, we carry out deep experiments to respond to the following research questions:

- RQ1: Does using style-transferred images and DIM improve the knowledge transformation?
- RQ2: Do partial learning and GMP improve the performance of models based on unsupervised learning?
- RQ3: Does the proposed pseudo-label refinement reduce the label noise's impact and enhance the re-ID model's discriminative performance?
- RQ4: What are the optimal hyperparameters for our proposed method?

The following sections will describe our experimental results and evaluation.

4.1. Implementation Details

Datasets and Evaluation Metrics. We conduct experiments on three well-known benchmarks in person re-ID:

1. *Market-1501* [37] contains 32,668 photos taken by 6 distinct cameras and 1,501 IDs. 12,936 photos representing 750 individuals make up the training set. The query set and gallery set, which have a total of 19,732 photos, contain 3,368 and 751 identities, respectively, in the test set.
2. *DukeMTMC-reID* [38] comprises the 8 cameras, which produced 36,411 photos with 1,812 individuals. This dataset includes 17,661 gallery photos, 2,228 query images, and 16,522 training images in full.
3. *MSMT17* [39] comprises a total of 126,441 pictures of 4,101 people taken using 15 cameras. It is partitioned into 93,820 test photos with 3,060 identities and 32,621 training images with 1,041 identities.

Hereafter, in this work, we use the term "Market" to denote "Market-1501", "Duke" to denote "DukeMTMC-reID" and "MSMT" to denote "MSMT17". We set up four adaptation tasks based on Market-to-Duke, Duke-to-Market, Market-to-MSMT, and Duke-to-MSMT.

Following previous works, mAP (mean Average Precision) and CMC Rank-1/5/10 (Cumulative Matching Characteristic) are metrics used for the evaluation in our work as well as no post-processing operation, like re-ranking [49]. Specifically, mAP is computed by finding each class's average precision (AP) and then taking the mean across all classes. The region below the precision-recall curve (AP) represents the trade-off between precision (accuracy of positive predictions) and recall (ability to discover all relevant examples). On the contrary, CMC is calculated for each query by analyzing the prioritized list of retrieved samples. It calculates the proportion of accurate matches among the top-k retrieved samples, where k is commonly predefined (1, 5, or 10). In evaluation, we will prioritize results with higher mAP scores and Rank-1 accuracy in evaluation.

Table 2. Baseline model performance (Acc %).

Method	Duke → Market				Market → Duke			
	mAP	R1	R5	R10	mAP	R1	R5	R10
Base.	26.2	55.3	67.8	75.2	25.8	43.7	57.9	66.2
+GAN	36.2	66.4	81.1	86.1	34.3	57.6	70.3	76.0
+DIM	30.1	60.4	75.1	80.8	28.8	51.1	66.0	71.2
Our base.	36.4	68.9	83.2	87.9	35.4	58.2	72.4	77.5

GAN Training Setting. We use the StarGAN [45] architecture with 256×128 resized images as input and Adam optimizer with the learning rate of 3.5×10^{-5} and the batch size is 16. After GAN training, every source domain image is converted into a C target camera style. For example, in the "Market-to-Duke" case, we generate 8 virtual images corresponding to 8 cameras in Duke for each image in Market, and conversely, with each image in Duke, we will create 6 new images corresponding to 6 cameras in Market.

Person Re-ID Model Training Setting. We use random cropping and flipping for data augmentation. We use triplet loss with a margin value of 0.3, Adam optimizer with a learning rate is 3.5×10^{-4} , and DNet architecture with an inter-batch normalization layer. The total epoch number is 80, and the learning rate is divided by 10 at the 40th and 70th epochs. The architecture of DNet consists of two full connection layers and an inter-batch normalization layer. Finally, the value of λ in Eq. 7 is set 0.05, similar to the original paper, and the ratio $N : M$ is elected to $4 : 1$.

For stage II, we train the model in 40 epochs. We utilize the DBSCAN to produce a pseudo-label at the start of every epoch. For the DBSCAN clustering algorithm, the minimal number of neighbors is set to 8, 16, 16 on the Market, Duke, and MSMT, respectively, while the maximum distance d is set to 0.6 in both cases. In addition to the methods used in stage I, we also used Random Erasing[50] to augment the dataset. Finally, each mini-batch is randomly sampled with 16 identities and 8 images, with a batch size of 128 in both stages.

Experiment Environment. All of our experiments deploy on an NVIDIA Tesla T4 (16GB) GPU, 2 x Xeon Platinum 8160, and 256GB RAM. Besides that, we implement our method using the PyTorch framework, which robustly supports deep learning.

4.2. Results & Ablation Study

Training Baseline. For question RQ1, we first discuss the effectiveness of our proposal in stage I. Table 2 shows that when we use the supervised baseline models (*Base.*), which are trained by the labeled domain, to test on another domain directly, the performance of these models only reaches 26.2%, 55.3%, 25.8%, and 43.7% on mAP and

Table 3. Comparison of SOTA unsupervised domain adaptive methods for person re-ID (Acc %). **Bold** denotes the best while Underline indicates the second best.

Method	Venue	Duke → Market				Market → Duke				Market → MSMT				Duke → MSMT			
		mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
PGPPM[44]	IEEE.TMM22	33.9	63.9	81.1	86.4	17.9	36.3	54.0	61.6	–	–	–	–	–	–	–	–
CDCSA[7]	ACM.MM22	34.0	66.8	82.1	87.9	31.4	53.8	67.8	72.2	–	–	–	–	–	–	–	–
MEB-Net[23]	ECCV20	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2	–	–	–	–	–	–	–	–
MMT[22]	ICLR20	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5	22.9	49.2	63.1	68.8	23.3	50.1	63.9	69.8
UNRN[24]	AAAI21	78.1	91.9	96.1	97.8	69.1	82.0	90.7	93.5	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
GLT[27]	CVPR21	79.5	92.2	96.5	97.8	69.2	82.0	90.2	92.8	26.5	56.6	67.5	72.0	27.7	59.5	70.1	74.2
RDSBN[11]	CVPR21	81.5	92.9	97.6	98.4	66.6	80.3	89.1	92.6	30.9	61.2	73.1	77.4	33.6	64.0	75.6	79.6
MCRN[26]	AAAI22	83.8	93.8	97.5	98.5	71.5	**84.5**	91.7	93.8	32.8	64.4	75.1	79.2	35.7	**67.5**	**77.9**	**81.6**
RESL[30]	AAAI22	83.1	93.2	96.8	98.0	72.3	83.9	91.7	93.6	33.6	**64.8**	74.6	79.6	34.2	65.2	74.6	80.1
SECRET[25]	AAAI22	83.0	93.3	–	–	69.2	82.0	–	–	31.7	60.0	–	–	–	–	–	–
CMFC [7]	ACM.MM22	81.0	94.0	97.1	98.3	71.2	83.2	91.6	94.0	–	–	–	–	–	–	–	–
LF_2[33]	ICPR22	83.2	92.8	**97.8**	98.4	**73.5**	83.7	91.9	94.3	–	–	–	–	–	–	–	–
DAPRH		**85.9**	**94.4**	**97.8**	**98.8**	72.0	83.7	**92.1**	**94.5**	**35.8**	**64.8**	**77.0**	**81.1**	**36.0**	65.5	77.0	80.9

Table 4. Influence of pooling operations and part-learning (Acc %).

Method	Duke → Market				Market → Duke			
	mAP	R1	R5	10	mAP	R1	R5	R10
Base.	81.5	92.0	96.9	98.0	58.8	74.5	82.7	85.5
+GMP	84.5	93.0	97.2	98.0	69.0	81.6	90.4	92.5
+GMP+PL	85.6	93.9	98.0	98.7	70.3	82.8	91.7	94.2

Table 5. Results of label refinement methods (Acc %).

Method	Duke → Market				Market → Duke			
	mAP	R1	R5	R10	mAP	R1	R5	R10
One-hot	85.6	93.9	98.0	98.7	70.3	82.8	91.7	94.2
CRL	85.9	94.4	97.8	98.8	72.0	83.7	92.1	94.5

Rank-1 in the case of Duke-to-Market and Market-to-Duke, respectively. The table also shows the effectiveness of our proposed approach in Section 3.3.2. Specifically, training with the style-transferred dataset improves the person re-ID baselines by about 10-13% performance on all metrics in both cases. We also examined the effectiveness of the DIM method, which improved the baseline performance from 26.2%, 55.3% to 30.1% and 60.4% on mAP and Rank-1, respectively, in the Duke-to-Market case. Finally, we achieved the best results in all cases by combining both methods in training models. Our method surpasses some SOTA methods such as [44, 7] in Table 3 with the same target.

Different combinations of the components. To answer the RQ2, we also evaluate the effectiveness of combinations on unsupervised learning in Stage II, consisting of using GMP and partial learning. Results are presented in Table 4. First, we train models having only a global branch without label refinement and GMP in DCNN architecture as baseline *(Base.)*. These models reach 81.5%, 92.0% on mAP and Rank-1 in the Market benchmark as well as 58.8%, 74.5% on mAP and Rank-1 in the case of Duke.

Secondly, the feature maps are subjected to integration with the GMP layer to extract image feature vectors (*Base. + GMP*). This architecture attains rank-1 accuracy of 93.0% and mAP of 84.5% on the Market dataset, substantially surpassing the performances of the baseline architecture. When testing on Duke, we also reach an analogous enhancement. The advantage of GMP stems from the fact that max-pooling only concentrates on feature maps with high response values, which could be finding strong discriminative features of pedestrian pictures.

Finally, we utilize the partial branch (*Base. + GMP + PL*). Compared to *Base. + GMP*, the utilization of partial learning (PL) can yield a significant enhancement in the performance of models merely learned on the global feature by $1-2\%$ and $0.5-1\%$ in mAP and Rank-1 on both Market and Duke, respectively. The outcome demonstrates that our suggested local-branch learning can supplement valuable information that the global branch ignores to increase the overall effectiveness of USL for person re-ID.

Effectiveness of pseudo-label refinement. To evaluate the efficacy of pseudo-label refinement methods mentioned in section 3.5 as well as answer the RQ3, we compare the best one-hot labels-guided models in Table 4 ("One-hot") with models trained by CRL. The performances are shown in Table 5. As one can see, CRL improves mAP and rank-1 by 0.3% and 0.5% on Market, by 1.7% and 1.1% on Duke benchmark, compared to the baseline "One-hot". It displays the usefulness of CRL, which contributes to the creation of refined labels that are more potent.

4.3. Analysis of Hyperparameters

Effectiveness of the ratio N:M. The ratio N:M is an essential parameter in training a solid baseline, where N and M denote the ratio between real and fake samples in each input batch. We show the experimental results in

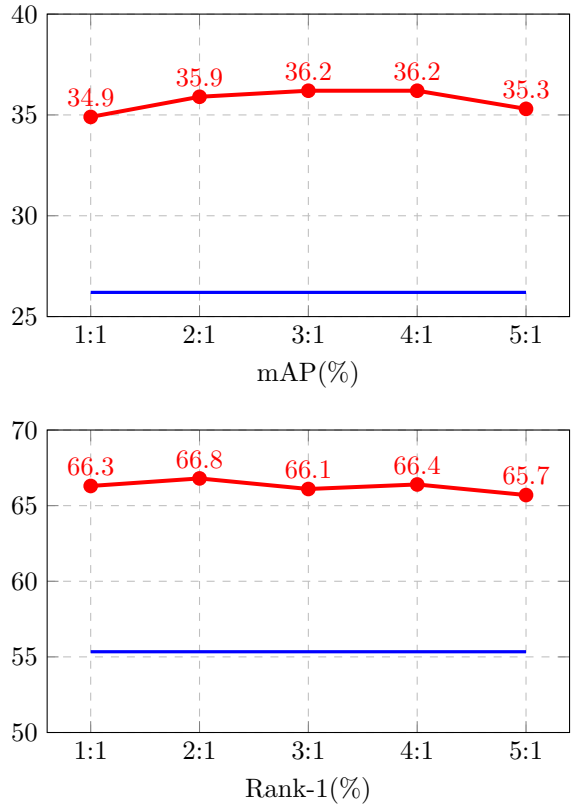

Figure 7. Evaluation of Market-1501 training mini-batch with varying $N : M$ ratios. The blue-line indicates the baseline while the red-line is Baseline+GAN.

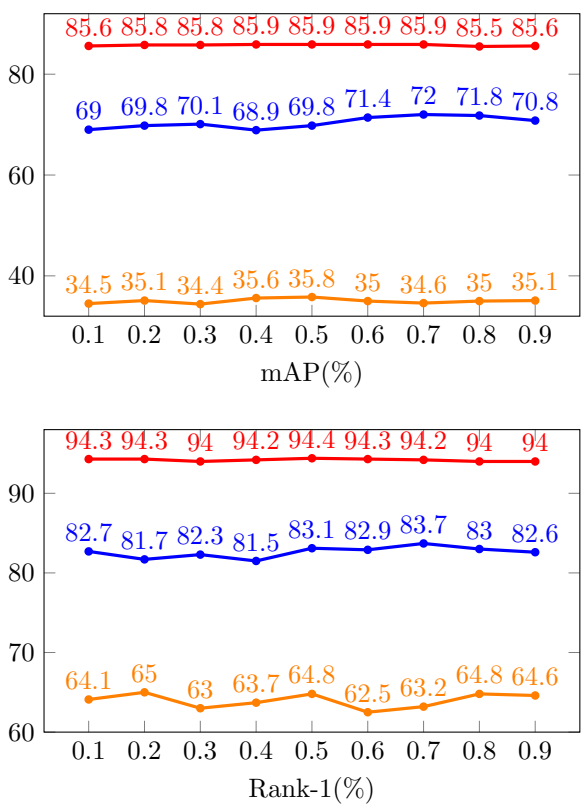

Figure 8. Impact of hyperparameter α on Market-1501 (red-line), DukeMTMC-reID (blue-line) and MSMT17 (orange-line).

Fig. 7 by varying this ratio. Apparently, with various ratios $N : M$, the synthetic dataset consistently improves over the baseline and achieved the highest results when $N : M = 4 : 1$.

Analysis of Coefficient α. We analyze the impact of the coefficient α in Eq. 9 by tuning α from 0.1 to 0.9. The results on the three benchmarks in Fig. 8 show that α significantly impacts the performance of the pseudo-label refinement method and needs to be carefully selected for each data domain. For Market and MSMT (in the case of Market-to-MSMT), setting $\alpha = 0.5$ leads to achieving the best performance while $\alpha = 0.7$ is the best optimal value on Duke. We also finetuned the hyperparameter α in case Duke-to-MSMT by the same way. We found that $\alpha = 0.6$ brings the highest performance in this case while $\alpha = 0.5$ gives a slightly lower performance at Rank-1.

According to all prior analyses, we demonstrate the advantages of our suggestions to improve the person re-ID model in unsupervised learning and the best collection of hyperparameters, which answers RQ4.

4.4. Comparison with SOTA methods

Finally, we compare DAPRH with recent UDA person re-ID methods in Table 3. Experimental results show that our method outperforms some SOTA methods, achieving 85.9% mAP and 94.4% Rank-1 accuracy on Duke-to-Market, 72.0% mAP, and 83.7% Rank-1 on Market-to-Duke. For the other two more difficult scenarios, our method also gets remarkable results, including 35.8% mAP and 64.8% Rank-1 on Market-to-MSMT as well as 35.8% mAP and 65.2% Rank-1 on Duke-to-MSMT.

Compared to well-known UDA methods such as UNRN [24], GLT [27], and RDSBN [11], our method achieves greater mAP and rank-1 accuracy by higher 1% in all benchmarks. When compared to CMFC [7], which utilizes both StarGAN-based data augmentation and partial learning, DAPRH outperforms it by 4.9%, 0.8% on mAP and 0.4%, 0.5% on Rank-1 accuracy in both scenarios of Duke-to-Market and Market-to-Duke, respectively. DAPRH also utterly outperforms the newest solution such as SECRET [25], MCRN [26], RESL [30], LF$_2$[33] on Duke-to-Market, Market-to-MSMT, and Duke-to-MSMT in the term of mAP.

Although in two scenarios Market-to-Duke and Duke-to-MSMT, our model only reaches the newest SOTA performance, with its efficient performance on limited hardware, DAPRH shows potential for real-world applications and further improvements. To clarify this statement, Table 6 summarizes the hardware information of several methods appeared in comparison with DAPRH as shown in Table 3. Apparently, in terms of single-precision computation performance (SP), our GPU is the lowest computation capability with 8.1 TFLOPS, compared to almost current

Table 6. Hardware comparison with some SOTA UDA methods. RAM measured in GB; *SP* denotes the single-precision performance of GPU; No is the number of GPU; BS stands for "batch size"; "-" means that information is not published.

Method	Experiment Environment				
	GPU	RAM	SP	No	BS
GLT[27]	RTX-2080Ti	11	13.4	4	64
RDSBN[11]	P100	16	9.3	2	256
MCRN[26]	RTX-2080Ti	11	13.4	4	64
RESL[30]	-	-	-	-	128
SECRET[25]	GTX-1080Ti	11	11.3	4	64
LF_2[33]	RTX-2080Ti	11	13.4	2	64
DAPRH	Tesla T4	16	8.1	1	128

SOTA methods. Furthermore, our system only has one GPU, whereas the other methods (excluding RESL) use multi-GPU systems. These systems may enable them to train re-ID models with greater batch sizes. In other words, it can reduce noise in loss value computation and convergence faster than our method. Therefore, our method can be considered as being more efficient than the others in its achievements.

5. Conclusions

In the subject of UDA person re-ID, this work emphasizes three primary challenges: (i) domain-intensive training models, (ii) pseudo-labels supplied by the clustering process, and (iii) precise information skipping. Thus, we propose the DAPRH method in order to solve these challenges. In the proposed method, we first use StarGAN to bridge the distribution of two data domains brought on by environmental factors. Secondly, we use the virtual dataset with supervised learning to build a robust pre-trained model. Then, we add a local branch as well as combine both GMP and GAP to catch up on significant features, which can be utilized to enhance the re-ID model's performance. Lastly, we suggest using the mean instructor architecture and a simple method to improve one-hot labels to enhance UDA task performance. Our experiments on three challenging benchmarks consisting of Market-1501, DukeMTMC-reID, and MSMT17 datasets prove that the DAPRH method outperforms SOTA methods in the UDA person re-ID.

There are still some directions for improving our DAPRH method. Indeed, the current GAN method needs to be optimized due to leading to many low-quality samples. In addition, the pseudo-label refining method assumes that the proportion of inaccurate labels in the pseudo-label collection is relatively low. Thus, such a pseudo-label refining method would lose effectiveness if the clustering findings were highly subpar, as shown in the case of Market-to-Duke. In the near future, we will therefore concentrate on refining and optimizing the DAPRH method for perspective works.

Data and Code Availability

The DAPRH codes, datasets, and experiment results used in this manuscript are available at https://github.com/ewigspace1910/DAPRH/.

References

[1] Y. Cho, W. J. Kim, S. Hong, S.-E. Yoon, Part-based pseudo label refinement for unsupervised person re-identification, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 7298–7308. doi:10.1109/CVPR52688.2022.00716.

[2] C.-X. Ren, Y.-H. Liu, X.-W. Zhang, K.-K. Huang, Multi-source unsupervised domain adaptation via pseudo target domain, IEEE Transactions on Image Processing 31 (2022) 2122–2135. doi:10.1109/TIP.2022.3152052.

[3] J. Hu, J. Lu, Y.-P. Tan, Deep transfer metric learning, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 325–333. doi:10.1109/CVPR.2015.7298629.

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

[5] B. Sun, J. Feng, K. Saenko, Return of frustratingly easy domain adaptation, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, AAAI Press, 2016, p. 2058–2065.

[6] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 3774–3782. doi:10.1109/ICCV.2017.405.

[7] Y. Tu, Domain camera adaptation and collaborative multiple feature clustering for unsupervised person re-id, in: Proceedings of the 3rd International Workshop on Human-Centric Multimedia Analysis, ACM, 2022, p. 51–59. doi:10.1145/3552458.3556446.

[8] J. Raitoharju, Chapter 3 - convolutional neural networks, in: A. Iosifidis, A. Tefas (Eds.), Deep Learning for Robot Perception and Cognition, Academic Press, 2022, pp. 35–69. doi:10.1016/B978-0-32-385787-1.00008-7.

[9] D. Deng, Dbscan clustering algorithm based on density, in: 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), 2020, pp. 949–953. doi:10.1109/IFEEA51475.2020.00199.

[10] A. Likas, N. Vlassis, J. J. Verbeek, The global k-means clustering algorithm, Pattern Recognition 36 (2) (2003) 451–461, biometrics. doi:https://doi.org/10.1016/S0031-3203(02)00060-2.

[11] Z. Bai, Z. Wang, J. Wang, D. Hu, E. Ding, Unsupervised multi-source domain adaptation for person re-identification, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 12909–12918. doi:10.1109/CVPR46437.2021.01272.

[12] J. M. Wolterink, K. Kamnitsas, C. Ledig, I. Išgum, Chapter 23 - deep learning: Generative adversarial networks and adversarial methods, in: S. K. Zhou, D. Rueckert, G. Fichtinger (Eds.), Handbook of Medical Image Computing and Computer Assisted Intervention, The Elsevier and MICCAI Society Book Series, Academic Press, 2020, pp. 547–574. doi:10.1016/B978-0-12-816176-0.00028-4.

[13] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, F. Bremond, Learning invariance from generated variance for unsupervised person re-identification, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) 1–15doi:10.1109/TPAMI.2022.3226866.

[14] Y. Ge, Z. Li, H. Zhao, G. Yin, S. Yi, X. Wang, H. Li, Fd-gan: Pose-guided feature distilling gan for robust person re-identification, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, Curran Associates Inc., Red Hook, NY, USA, 2018, p. 1230–1241.

[15] X. Qian, Y. Fu, T. Xiang, W. Wang, J. Qiu, Y. Wu, Y.-G. Jiang, X. Xue, Pose-normalized image generation for person re-identification, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018, pp. 661–678. doi:10.1007/978-3-030-01240-3_40.

[16] H. Xie, H. Luo, J. Gu, W. Jiang, Unsupervised domain adaptive person re-identification via intermediate domains, Applied Sciences 12 (14) (2022). doi:10.3390/app12146990.

[17] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242–2251. doi:10.1109/ICCV.2017.244.

[18] Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, Camera style adaptation for person re-identification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5157–5166. doi:10.1109/CVPR.2018.00541.

[19] Y. Huang, Q. Wu, J. Xu, Y. Zhong, Sbsgan: Suppression of inter-domain background shift for person re-identification, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9526–9535. doi:10.1109/ICCV.2019.00962.

[20] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, J. Kautz, Joint discriminative and generative learning for person re-identification, 2019, pp. 2133–2142. doi:10.1109/CVPR.2019.00224.

[21] M. Liao, Z. Wan, C. Yao, K. Chen, X. Bai, Real-time scene text detection with differentiable binarization, Proceedings of the AAAI Conference on Artificial Intelligence 34 (07) (2020) 11474–11481. doi:10.1609/aaai.v34i07.6812.

[22] 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.

[23] 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, Springer-Verlag, 2020, p. 594–611. doi:10.1007/978-3-030-58571-6_35.

[24] K. Zheng, C. Lan, W. Zeng, Z. Zhang, Z.-J. Zha, Exploiting sample uncertainty for domain adaptive person re-identification, Vol. 35, 2021, pp. 3538–3546. doi:10.1609/aaai.v35i4.16468.

[25] T. He, L. Shen, Y. Guo, G. Ding, Z. Guo, Secret: Self-consistent pseudo label refinement for unsupervised domain adaptive person re-identification, Vol. 36, 2022, pp. 879–887. doi:10.1609/aaai.v36i1.19970.

[26] Y. Wu, T. Huang, H. Yao, C. Zhang, Y. Shao, C. Han, C. Gao, N. Sang, Multi-centroid representation network for domain adaptive person re-id, Proceedings of the AAAI Conference on Artificial Intelligence 36 (3) (2022) 2750–2758. doi:10.1609/aaai.v36i3.20178.

[27] K. Zheng, W. Liu, L. He, T. Mei, J. Luo, Z.-J. Zha, Group-aware label transfer for domain adaptive person re-identification, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 5306–5315. doi:10.1109/CVPR46437.2021.00527.

[28] J. Peng, G. Jiang, H. Wang, Cooperative refinement learning for domain adaptive person re-identification, Knowledge-Based Systems 242 (2022) 108349. doi:10.1016/j.knosys.2022.108349.

[29] Z. Pang, C. Wang, J. Wang, L. Zhao, Reliability modeling and contrastive learning for unsupervised person re-identification, Knowledge-Based Systems 263 (2023) 110263. doi:10.1016/j.knosys.2023.110263.

[30] Z. Li, Y. Shi, H. Ling, J. Chen, Q. Wang, F. Zhou, Reliability exploration with self-ensemble learning for domain adaptive person re-identification, Proceedings of the AAAI Conference on Artificial Intelligence 36 (2) (2022) 1527–1535. doi:10.1609/aaai.v36i2.20043.

[31] Y. Miao, J. Deng, G. Ding, J. Han, Confidence-guided centroids for unsupervised person re-identification (2022). doi:10.48550/ARXIV.2211.11921.

[32] T. Si, Z. Zhang, S. Liu, Compact triplet loss for person re-identification in camera sensor networks, Ad Hoc Networks 95 (2019) 101984. doi:10.1016/j.adhoc.2019.101984.

[33] J. Ding, X. Zhou, Learning feature fusion for unsupervised domain adaptive person re-identification, in: 2022 26th International Conference on Pattern Recognition (ICPR), 2022, pp. 2613–2619. doi:10.1109/ICPR56361.2022.9956264.

[34] V. Somers, C. De Vleeschouwer, A. Alahi, Body Part-Based Representation Learning for Occluded Person Re-Identification, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV23) (nov 2023). arXiv:2211.03679, doi:10.48550/arxiv.2211.03679.

[35] H. Tan, X. Liu, B. Yin, X. Li, Mhsa-net: Multihead self-attention network for occluded person re-identification, IEEE Transactions on Neural Networks and Learning Systems (2022) 1–15doi:10.1109/TNNLS.2022.3144163.

[36] K. Zhu, H. Guo, Z. Liu, M. Tang, J. Wang, Identity-guided human semantic parsing for person re-identification, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham, 2020, pp. 346–363. doi:10.1007/978-3-030-58580-8_21.

[37] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1116–1124. doi:10.1109/ICCV.2015.133.

[38] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: G. Hua, H. Jégou (Eds.), Computer Vision – ECCV 2016 Workshops, Springer International Publishing, Cham, 2016, pp. 17–35. doi:10.1007/978-3-319-48881-3_2.

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

[40] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, S. H. Hoi, Deep learning for person re-identification: A survey and outlook, IEEE Transactions on Pattern Analysis & Machine Intelligence 44 (06) (2022) 2872–2893. doi:10.1109/TPAMI.2021.3054775.

[41] H. Gu, J. Li, G. Fu, M. Yue, J. Zhu, Loss function search for person re-identification, Pattern Recognition 124 (2022) 108432. doi:10.1016/j.patcog.2021.108432.

[42] Z. Ming, J. Chazalon, M. M. Luqman, M. Visani, J.-C. Burie, Simple triplet loss based on intra/inter-class metric learning for face verification, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017, pp. 1656–1664. doi:10.1109/ICCVW.2017.194.

[43] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, p. 1195–1204. doi:10.5555/3294771.3294885.

[44] F. Yang, Z. Zhong, Z. Luo, S. Lian, S. Li, Leveraging virtual and real person for unsupervised person re-identification, IEEE Transactions on Multimedia 22 (9) (2020) 2444–2453. doi:10.1109/TMM.2019.2957928.

[45] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, J. Choo, Stargan: Unified generative adversarial networks for multi-domain image-to-image translation, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8789–8797. doi:10.1109/CVPR.2018.00916.

[46] X. Liu, S. Zhang, Domain adaptive person re-identification via coupling optimization, in: Proceedings of the 28th ACM International Conference on Multimedia, MM '20, ACM, 2020, p. 547–555. doi:10.1145/3394171.3413904.

[47] K. R. Shahapure, C. Nicholas, Cluster quality analysis using silhouette score, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 747–748. doi:10.1109/DSAA49011.2020.00096.

[48] W. Chen, Y. Liu, E. M. Bakker, M. S. Lew, Integrating information theory and adversarial learning for cross-modal retrieval, Pattern Recognition 117 (2021) 107983. doi:10.1016/j.patcog.2021.107983.

[49] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3652–3661. doi:10.1109/CVPR.2017.389.

[50] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, Proceedings of the AAAI Conference on Artificial Intelligence 34 (07) (2020) 13001–13008. doi:10.1609/aaai.v34i07.7000.