Article

Unsupervised Domain Adaptive Person Re-Identification via Intermediate Domains

Haonan Xie, Hao Luo, Jianyang Gu and Wei Jiang *

Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, China; hn_xie@zju.edu.cn (H.X.); haoluocsc@zju.edu.cn (H.L.); jianyang_gu@zju.edu.cn (J.G.)

* Correspondence: jiangwei_zju@zju.edu.cn

Abstract: Recent years have witnessed outstanding success in supervised domain adaptive person re-identification (ReID). However, the model often suffers serious performance drops when transferring to another domain in real-world applications. To address the domain gap situations, many unsupervised domain adaptive (UDA) methods have been proposed to adapt the model trained on the source domain to a target domain. Such methods are typically based on clustering algorithms to generate pseudo labels. Noisy labels, which often exist due to the instability of clustering algorithms, will substantially affect the performance of UDA methods. In this study, we focused on intermediate domains that can be regarded as a bridge that connects source and target domains. We added a domainness factor in the loss function of SPGAN that can decide the style of the image generated by the GAN model. We obtained a series of intermediate domains by changing the value of the domainness factor. Pseudo labels are more reliable because intermediate domains are closer to the source domain compared with the target domain. We then fine-tuned the model pre-trained with source data on these intermediate domains. The fine-tuning process was conducted repeatedly because intermediate domains are composed of more than one dataset. Finally, the model fine-tuned on intermediate domains was adapted to the target domain. The model can easily adapt to changes in image style as we gradually transfer the model to the target domain along the bridge consisting of several intermediate domains. To the best of our knowledge, we are the first to apply intermediate domains to UDA problems. We evaluated our method on Market1501, DukeMTMC-reID and MSMT17 datasets. Experimental results proved that our method brings a significant improvement and achieves a state-of-the-art performance.

Keywords: person re-identification; domain adaptation; intermediate domains; semi-supervised learning

1. Introduction

Person re-identification (ReID), a key component of pedestrian-tracking pipelines that aims to track the same pedestrian across different camera devices, has drawn considerable research attention. Numerous CNN-based methods for ReID have been proposed to achieve a human-level performance or even outperform it [1,2]. These methods utilize the identities of pedestrians as labels with train and test sets belong to the same domain.

Although supervised ReID methods have achieved outstanding success, they often suffer catastrophic performance drops when applied to another domain. As shown in Figure 1, an obvious domain gap exists between the feature distributions of the Market1501 [3] and DukeMTMC-reID [4] datasets. If we directly adapt the model trained on one domain to another, then it fails to achieve a good performance on the target dataset. As shown in Table 1, the model trained on the Market1501 dataset and then directly transferred to the DukeMTMC-reID dataset dropped by 47% and 57.4% in rank-1 and mAP accuracies, respectively.

Figure 1. Illustration of the features of Market1501 and DukeMTMC-reID datasets. The green dots represent the feature distributions of Market1501 dataset, whereas the red ones represent the feature distributions of DukeMTMC-reID dataset.

Thus, unsupervised domain adaptation (UDA) methods are proposed to adapt the model trained on one domain to a new domain without identity labels. Recent UDA methods [5,6] for ReID aim to learn models that appropriately fit on the target domain with clustering algorithms. Clustering-based methods first train a model on the source domain and then fine-tune the model based on pesudo labels. The process of model training and clustering is carried out iteratively until the model converges. Although pesudo labels generated through clustering algorithms can improve the performance of the model on the target domain to some extent, they often contain noisy (wrong) labels due to the gap between source and target domains that can mislead the training of the model.

Table 1. The performance of different models evaluated on cross-domain datasets. M→D means that we train the model on the Market1501 and evaluate it on the DukeMTMC-reID.

Dataset	Rank-1	mAP
Market1501	94.5	91.0
M→D	47.5	37.6
DukeMTMC-reID	86.7	84.1
D→M	61.3	43.5

In this work, we focus on the intermediate domains between source and target domains. Many studies have proven the effectiveness of intermediate domains in domain adaptation tasks [7]. Images from intermediate domains contain both the style of the source and target domains. Intermediate domains will be utilized as a bridge to narrow the domain gap between source and target domains. Pseudo labels generated on intermediate domains are reliable given the small domain divergence between source and intermediate domains.

Our work is based on DLOW [8] and SPGAN [9], which is one of the state-of-the-art (SOTA) image-to-image translation methods for ReID. We added a domainness factor in the loss function of SPGAN on the basis of DLOW to determine the style of output images. We fixed the domainness factor during the training of GAN to generate a specific image style. Then, the value of the domainness factor was set up manually to obtain a series of intermediate domains.

We followed the UDA strong baseline in [6], which consists of three main stages, namely, model pre-training, clustering and fine-tuning. The baseline of our proposed method is shown in Figure 2. Typical UDA methods directly transfer the model trained on the source domain to the target domain. We first pre-trained the model using source labeled data and then fine-tuned the model on unlabeled intermediate domains and generated pseudo labels on the basis of clustering results. Finally, we conducted clustering on the unlabeled target domain and fine-tuned the model with pseudo labels. Notably, the intermediate domains consist of a series of domains instead of one dataset. That is, the fine-tuning process is repeated between one intermediate domain and another.
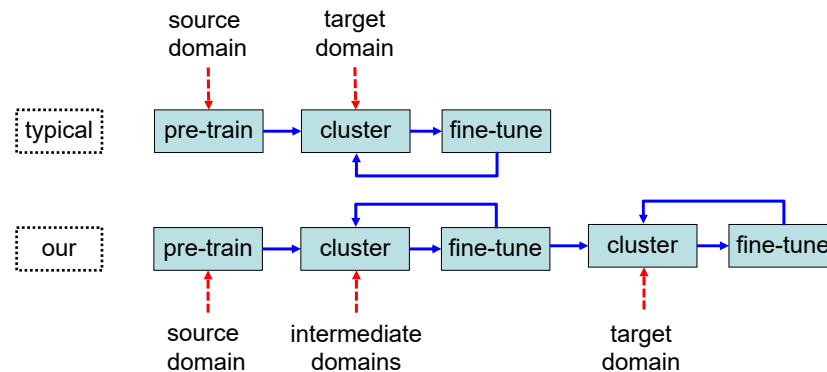

Figure 2. The comparison of UDA ReID pipeline between our proposed method and typical method. We utilize the intermediate domains as a bridge to narrow the domain gap between source and target domain.

The contributions of this study are presented as follows:

- We set a domainness factor in the loss function of SPGAN [9] on the basis of DLOW [8] to affect the style of the output images and obtain a series of intermediate style datasets.
- Intermediate domains were utilized as a bridge to narrow the gap between source and target domains. We first adopted the model pre-trained on the source domain to unlabeled intermediate domains and then fine-tuned it on the unlabeled target domain on the basis of the UDA strong baseline in [6].
- We conducted extensive experiments on Market1501, DukeMTMC-reID and MSMT17 datasets. The results demonstrate the effectiveness of our framework. Our scheme achieves a state-of-the-art performance on all of the benchmark datasets.

Moreover, the proposed framework can be applied to other SOTA methods to further improve their performance.

2. Related Works

2.1. Unsupervised Domain Adaptive ReID

Unsupervised domain adaptation (UDA) for person ReID methods aim to adapt the model trained on a labeled source domain to an unlabeled target domain. UDA methods have drawn considerable research attention due to their capability of saving manual annotation costs in recent years. UDA can be classified into three categories: feature alignment, image translation and semi-supervised learning.

2.1.1. Feature Alignment

Lin et al. [10] proposed a model that can be jointly optimised under a person's identity classification and the attribute-learning task with a cross-dataset mid-level feature alignment regularization term. The researchers used the maximum mean discrepancy (MMD) distance to align the mid-level features of the source and target domain. Ref. [11] introduced a model that can simultaneously learn an attribute-semantic and identity-discriminative

feature representation space, which can be transferred to the target domain. However, due to the differences in pedestrian classes between the two domains, it is hard for the model to learn a common feature in the representation space.

2.1.2. Image Translation

Image translation methods aim to transfer the image style of the source domain into the target domain. Many methods based on generative adversarial networks (GANs) [12] have been proposed to minimize the domain gap. SPGAN [9], PTGAN [13] and PDA-Net [14] transferred images with identity labels from the source domain to the target domain and then trained the model on target domains. HHL [15] enforced camera invariance and domain connectedness simultaneously to improve the generalization ability of ReID models on the target domain.

Although GAN-based methods can improve the performance of the model on the target domain to some extent, the lack of pre-training on source domain hinders the advancement of these approaches.

2.1.3. Semi-Supervised Learning

Recent studies regard UDA as a semi-supervised problem. These methods first pre-train the model on the source domain and then fine-tune it on the target domain based on pseudo labels generated by clustering algorithms. Ge et al. [5] proposed a method named mutual mean-teaching (MMT) that learned enhanced features from the target domain via off-line refined hard pseudo labels and on-line refined soft pseudo labels in an alternative training manner. The researchers also proposed a novel self-paced contrastive learning framework [16] with hybrid memory that can generate source-domain class-level, target-domain cluster-level and un-clustered instance-level supervisory signals for learning feature representations to optimize all valuable information. Zhao et al. [17] proposed a noise resistible mutual-training method that maintains two networks during training to perform collaborative clustering and mutual instance selection. Zhai et al. [18] put forward a multiple expert brainstorming network that trained multiple base-level networks on the source domain. The model is adapted to the target domain through brainstorming (mutual learning) among expert models. Zheng et al. [6] established a network named uncertainty-guided noise resilient network (UNRN) to explore the credibility of the predicted pseudo labels of target domain samples.

However, noisy labels in these approaches due to the divergence between source and target data will harm the performance of domain adaptation.

2.2. Domain Adaptation and Generalization

Domain adaptation aims to train a model on a labeled source domain that performs well on an unlabeled target domain [19–21], while the goal of domain generalization [8,22,23] is to enable the trained model to generalize unseen domains appropriately. Domain generalization is more robust than domain adaptation because multiple source domains containing the same visual classes are available for model training.

The aforementioned works only focus on the source and target domains. However, many studies have proven that the intermediate domains between source and target domains are useful for addressing the domain adaptation problems. Cui et al. [7] characterized samples from each domain as one covariance matrix and interpolated some intermediate points between source and target domains to bridge them. Wang et al. [24] proposed a method for continuously indexed domain adaptation by combining traditional adversarial adaptation with a novel discriminator that models the encoding-conditioned domain index distribution. Gong et al. [8] designed a model that can generate a continuous sequence of intermediate domains flowing from one domain to the other for semantic segmentation.

Unlike the studies mentioned above, we generated intermediate domains based on GAN and conducted the UDA baseline on the source domain and intermediate domains. Then, the model fine-tuned on intermediate domains was transferred to the target domain.

3. Proposed Method

Our proposed method aims to generate intermediate domains to bridge the source and target domains. In this section, we will first introduce the means for obtaining intermediate domain images based on SPGAN and then briefly discuss the UDA baseline.

3.1. Intermediate Domains

3.1.1. SPGAN Model

The proposed method is based on SPGAN [9], which consists of Siamese network (SiaNet) and CycleGAN [25]. SPGAN can not only transfer the style of the image from source domain to target domain but also preserve its underlying identity information.

CycleGAN includes two generators and discriminators between source domain $\mathcal{S}$ and target domain $\mathcal{T}$. $\{G, D_\mathcal{T}\}$ transfer the images in $\mathcal{S}$ into the style of $\mathcal{T}$, whereas $\{F, D_\mathcal{S}\}$ act in the inverse direction. The following cycle consistency loss is proposed to preserve the semantic content of an image after translating it to another style:

$$
\begin{aligned}
\mathcal{L}_{\text{rec}}(G, F) =& \mathbb{E}_{x \sim p_{\text{data}(\mathcal{S})}} \|F(G(x)) - x\|_1 \\
&+ \mathbb{E}_{y \sim p_{\text{data}(\mathcal{T})}} \|G(F(y)) - y\|_1
\end{aligned}
\tag{1}
$$

where $p_{\text{data}(\mathcal{S})}$ and $p_{\text{data}(\mathcal{T})}$ denote the sample distributions in the source and target domain, respectively.

The loss function of cycleGAN can be written as:

$$
\begin{aligned}
\mathcal{L}_{cyc}(G, F, D_\mathcal{T}, D_\mathcal{S}) =& \mathcal{L}_{adv}(G, D_\mathcal{T}) + \mathcal{L}_{adv}(F, D_\mathcal{S}) \\
&+ \alpha \mathcal{L}_{rec}(G, F)
\end{aligned}
\tag{2}
$$

$\mathcal{L}_{adv}(G, D_\mathcal{T})$ and $\mathcal{L}_{adv}(F, D_\mathcal{S})$ represent the adversarial loss for GAN model:

$$
\begin{aligned}
\mathcal{L}_{adv}(G, D_\mathcal{T}) =& \mathbb{E}_{y \sim p_{\text{data}(\mathcal{T})}} [\log D_\mathcal{T}(y)] \\
&+ \mathbb{E}_{x \sim p_{\text{data}(\mathcal{S})}} [\log(1 - D_\mathcal{T}(G(x)))]
\end{aligned}
\tag{3}
$$

Based on cycleGAN, SPGAN introduced the inside-domain identity constraint [26] as an auxiliary for image translation. For each image x from domain $\mathcal{S}$, the generator F should translated it into source domain.

$$
\begin{aligned}
\mathcal{L}_{ide}(G, F) =& \mathbb{E}_{x \sim p_{\text{data}(\mathcal{S})}} \|F(x) - x\|_1 \\
&+ \mathbb{E}_{y \sim p_{\text{data}(\mathcal{T})}} \|G(y) - y\|_1
\end{aligned}
\tag{4}
$$

The SiaNet M_e was introduced to preserve the information of a pedestrian, and was trained by similarity-preserving loss [27] as follows:

$$
\mathcal{L}_{\text{con}}(i, x_1, x_2) = (1 - i)\{\max(0, m - d)\}^2 + id^2
\tag{5}
$$

where x_1 and x_2 are two input vectors and d denotes the Euclidean distance between the normalized embeddings of the two inputs. $i = 1$ if x_1 and x_2 belong to the same pedestrian but different domains, whereas $i = 0$ if x_1 and x_2 belong to different pedestrians.

The overall loss function of SPGAN is as follows:

$$
\begin{aligned}
\mathcal{L}_{\text{sp}}(G, F, D_\mathcal{T}, D_\mathcal{S}, M) =& \mathcal{L}_{cyc}(G, F, D_\mathcal{T}, D_\mathcal{S}) \\
&+ \beta \mathcal{L}_{ide}(G, F) + \gamma \mathcal{L}_{\text{con}}(G, F, M_e)
\end{aligned}
\tag{6}
$$

where the parameter β and γ control the importance of identity constraint and similarity-preserving loss. We would like to refer the readers to the SPGAN [9] for more details about these loss functions.

3.1.2. Generating Intermediate Domains

Intermediate domains have been shown to be helpful for addressing domain adaptation problems that can connect source and target domains. In this way, the model pre-trained on source domain can be fine-tuned on intermediate domains first and then adopted to target domain. We focused on image-level translating based on DLOW [8] and directly translated source images into different styles. In this way, the intermediate domains generated by the GAN model can be utilized as a bridge to connect source and target domains. The baseline of the GAN model is shown in Figure 3. Our method can be integrated with other UDA approaches to improve the cross-domain ability of the model.
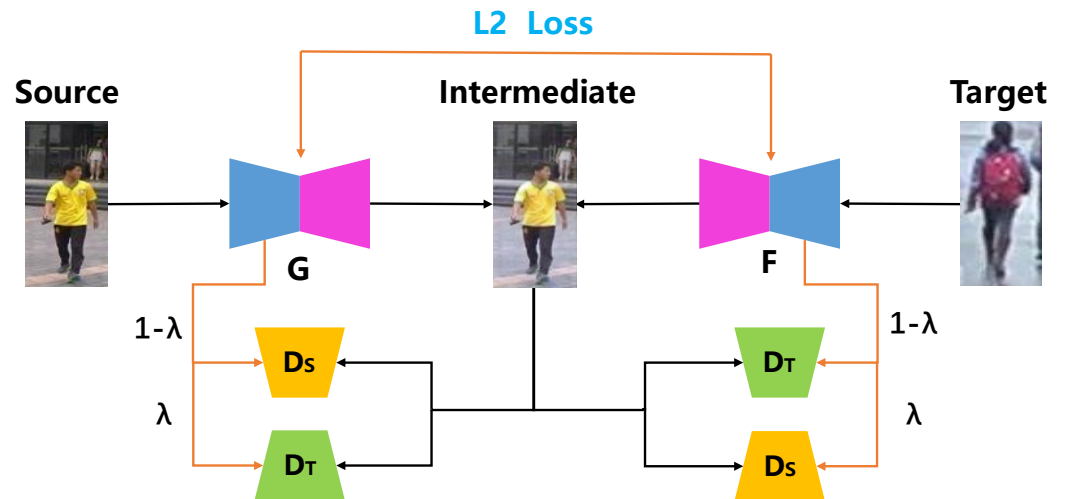

Figure 3. The overview of our proposed GAN model based on SPGAN to generate intermediate domains. λ represents the domainness factor to control the style of the images generated by GAN. The loss function is declared in Equation (14).

There are samples $x \in \mathcal{S}$ and $y \in \mathcal{T}$ that belong to data distributions P_S and P_T in the domain adaptation problem. We denoted an intermediate domain as $\mathcal{M}^{(\lambda)}$ and the data distribution of $\mathcal{M}^{(\lambda)}$ as P_M according to DLOW, where $\lambda \in [0,1]$ is the domainness factor that decides the relatedness between source and intermediate domains. A smaller λ indicates that P_S and P_M are closer to each other. For example, we have $\mathcal{M}^{(0)} = \mathcal{S}$ when $\lambda = 0$ and $\mathcal{M}^{(1)} = \mathcal{T}$ when $\lambda = 1$.

The DLOW model regarded a domain with given data distribution as a point residing at the manifold connecting source and target domain:

$$\mathrm{JSD}(P_S, P_M) + \mathrm{JSD}(P_M, P_T) = \mathrm{JSD}(P_S, P_T) \tag{7}$$

where $\mathrm{JSD}(\cdot, \cdot)$ is the Jessen–Shannon (JS) divergence for measuring differences between two distributions.

Therefore, the following distance from $\mathcal{S}$ to $\mathcal{M}^{(\lambda)}$ should also be proportional to the distance between $\mathcal{S}$ and $\mathcal{T}$ for a specific λ:

$$\frac{\mathrm{JSD}(P_S, P_M)}{\mathrm{JSD}(P_S, P_T)} = \lambda \qquad \frac{\mathrm{JSD}(P_M, P_T)}{\mathrm{JSD}(P_S, P_T)} = 1 - \lambda \tag{8}$$

Thus, finding the data distribution P_M corresponding to a specific λ becomes finding a point satisfying Equation (8). The loss function can be expressed as follows:

$$\mathcal{L} = (1 - \lambda) \cdot \mathrm{JSD}(P_S, P_M) + \lambda \cdot \mathrm{JSD}(P_M, P_T) \tag{9}$$

By using the expression of JS divergence, we have:

$$
\begin{aligned}
\text{JSD}(P_S, P_M) = {} & \frac{1}{2}\mathbb{E}_{x \sim P_S}\left[\log \frac{P_S(x)}{P_S(x) + P_M(x)}\right] \\
& + \frac{1}{2}\mathbb{E}_{x \sim P_M}\left[\log \frac{P_M(x)}{P_S(x) + P_M(x)}\right] + \log 2
\end{aligned}
\tag{10}
$$

Specifically, we denoted D_S to discriminate the images in S and M and D_T to discriminate the images in M and T. The optimal discriminator D_S for a fixed G is expressed as follows:

$$
D_S = \frac{P_S(x)}{P_S(x) + P_M(x)}
\tag{11}
$$

Ignoring the coefficients and constant terms in Equation (10), we have:

$$
\begin{aligned}
\mathcal{L}^* = {} & (1 - \lambda) \cdot [\mathbb{E}_{x \sim P_S}[\log D_S(x)] + \mathbb{E}_{x \sim P_M}[\log(1 - D_S(x))]] \\
& + \lambda \cdot [\mathbb{E}_{x \sim P_S}[\log D_T(x)] + \mathbb{E}_{x \sim P_M}[\log(1 - D_T(x))]] \\
= {} & (1 - \lambda) \cdot \mathcal{L}_{adv}(G, D_S) + \lambda \cdot \mathcal{L}_{adv}(G, D_T)
\end{aligned}
\tag{12}
$$

Equation (12) shows the adversarial loss of $S \to T$. We can similarly apply it to another direction as follows:

$$
\begin{aligned}
\mathcal{L}_{adv} = {} & (1 - \lambda) \cdot \mathcal{L}_{adv}(G, D_S) + \lambda \cdot \mathcal{L}_{adv}(G, D_T) \\
& + \lambda \cdot \mathcal{L}_{adv}(F, D_S) + (1 - \lambda) \cdot \mathcal{L}_{adv}(F, D_T)
\end{aligned}
\tag{13}
$$

We also applied identity constraint and similarity-preserving loss according to SPGAN to ensure that the underlying identity information was appropriately preserved after translating the style of the image. We could then obtain the total loss function:

$$
\mathcal{L} = \mathcal{L}_{adv} + \alpha \mathcal{L}_{rec}(G, F) + \beta \mathcal{L}_{ide}(G, F) + \gamma \mathcal{L}_{con}(G, F, M_e)
\tag{14}
$$

The training procedure is similar to SPGAN. The three parts, namely, generators $\{G, F\}$, discriminators $\{D_S, D_T\}$ and SiaNet M_e, are optimized alternately during training. Meanwhile, the remaining two parts are fixed when any one part is updated. We trained the model until convergence or reaching maximum iterations.

3.2. UDA Baseline via Intermediate Domains

We followed the strong baseline used in [6], which exploits contrastive loss and MMT. We pre-trained the network using source domain labeled data in the pre-training stage. We performed clustering on the unlabeled intermediate domain data using the more accurate features from the mean teacher model and generated pseudo labels based on the clustering results in the clustering stage. Finally, the clustering stage was repeated on the unlabeled target domain. Take the Market1501 to Duke-MTMC-reID for example; the process of training via intermediate domains is shown in Figure 4. m2d_0.2 denotes the unlabeled intermediate domain generated by GAN with Market1501 and DukeMTMC-reID datasets as source and target domains, respectively, for a domainness factor of $\lambda = 0.2$.
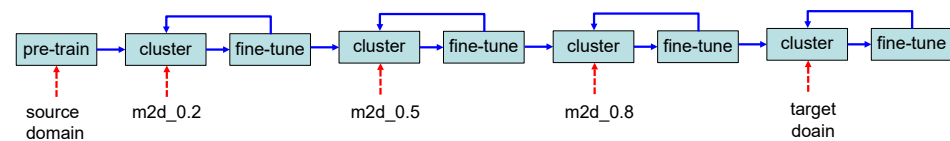

Figure 4. The training pipeline of our method. m2d_0.2 denotes the features of images generated by GAN, which transfers the style from Market1501 to DukeMTMC-reID for $\lambda = 0.2$.

3.2.1. Fine-Tuning with Source Data

Many works have proven that the source domain data can boost the performance of cross domain model [16]. We also used the source data with reliable groundtruth labels to fine-tune the model. The ID loss with label smoothing [28] in a mini-batch of n_t for a sample x_i with pseudo label $\tilde{y}_i$ can be written as:

$$\mathcal{L}_{\text{ID}} = -\frac{1}{n_t} \sum_{i=1}^{n_t} \tilde{y}_i \log p(\tilde{y}_i \mid x_i) \tag{15}$$

where $p(\tilde{y}_i \mid x_i)$ is the probability of being class $\tilde{y}_i$.

3.2.2. Mean Teacher Method

The mean teacher strategy proposed in [5] can temporally average model weights over training stage. We used the temporally average model of each network to generate reliable soft pseudo labels for supervising the other network. Features from the teacher model were utilized for clustering and the final inference.

3.2.3. Contrastive Loss

Zheng et al. [6] proposed a contrastive loss to maximize the within-class similarity and minimize the between-class similarity across the memory bank. The memory bank consisted of N samples from target domain and the central feature of the C_s source classes. The similarity was optimized with respect to the query sample.

Given a query sample q_k, we have N^+ positive samples and N^- negative samples in the memory. The contrastive loss can be expressed as follows:

$$\mathcal{L}_{\text{C}} = \frac{1}{n_t} \sum_{k=1}^{n_t} \log \left[1 + \sum_{j=1}^{N^-} \exp\left(s_{kj}^-\right) \sum_{i=1}^{N^+} \exp\left(-s_{ki}^+\right) \right] \tag{16}$$

where s_{kj}^- denotes the similarity between q_k and jth negative sample and s_{ki}^+ denotes the similarity between q_k and ith positive sample.

3.2.4. Triplet Loss

The triplet loss can minimize the similarity between the same pedestrian while maximizing the similarity between different pedestrians. Finally, the images belonging to the same pedestrian were an independent class.

$$\mathcal{L}_{TRI} = \frac{1}{n_t} \sum_{a \in n_t} \left(\max_p \left(\|f_a - f_p\|_2 \right) - \min_n \left(\|f_a - f_n\|_2 \right) + \alpha \right) \tag{17}$$

where the first item denotes the distance between the probe image a and the positive image p, whereas the second item is the distance between the probe image a and the negative image n.

The aforementioned loss functions were applied jointly to fine-tune the model on intermediate and target domains. Each of these loss functions are equally important, following previous work [6]:

$$\mathcal{L} = \mathcal{L}_{\text{ID}} + \mathcal{L}_{\text{C}} + \mathcal{L}_{TRI} \tag{18}$$

4. Experimental Results and Discussion

4.1. Datasets

We evaluated our method on the Market1501, DukeMTMC-reID and MSMT17 datasets. Following previous works, we used rank-1 accuracy of the cumulative matching characteristics and the mean average precision (mAP) accuracy for evaluation on all datasets.

Market1501 [3], a large public benchmark dataset on the ReID task, was collected using six different outdoor cameras in the campus of Tsinghua University. The dataset

includes training and test sets. The training set contains 12,936 images and 751 pedestrian IDs, whereas the test set contains 19,732 gallery images and 3386 query images from 750 other pedestrian IDs. Images of each pedestrian were captured by at least two cameras.

DukeMTMC-reID [4], a ReID subset of the Duke-MTMC dataset, was collected simultaneously using eight cameras at Duke University. The training set contains 16,522 images from 702 people, and the test set contains 2228 query images and 17,661 gallery images from 702 other pedestrians.

MSMT17 [13] was collected with 15 cameras. The dataset was randomly divided according to the training-test ratio of 1:3. Finally, the training set contains 1,041 pedestrians with a total of 32,621 bounding boxes, whereas the test set includes 3060 pedestrians with a total of 93,820 bounding boxes. From the test set, 11,659 bounding boxes were randomly selected as query images, and the other 82,161 bounding boxes were used as gallery images.

4.2. Implementation Details

4.2.1. Training of the GAN Model

We adopted the architecture and hyper-parameter released in [9]. We set $m = 2$ in Equation (5) and $\alpha = 10, \beta = 5, \gamma = 2$ in Equation (14). Each image was resized into 256×128 pixels, and data augmentation included random horizontal flipping and cropping [2]. We used the Adam optimizer [29] with an initial learning rate of 2×10^{-4}. β_1 and β_2 were set to 0.5 and 0.999, respectively. Different intermediate domains were generated by changing the value of the domainness factor. Specifically, we set $\lambda = 0.2, 0.5, 0.8$ and obtained three intermediate domains.

4.2.2. Pre-Training on Source Domain

We used ResNet50 [30] pre-trained on ImageNet as the backbone architecture on the source domain for the pre-training stage. For each mini-batch, we utilized the PK strategy used in [2]. We set P as 8 and K as 4. We used the Adam optimizer [29] with an initial learning rate of 3.5×10^{-4} and shrunk this learning rate by a factor of 0.1 at 40 and 70 epochs until convergence was achieved. Different from the training of GAN, we set $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

4.2.3. Fine-Tuning on Intermediate and Target Domains

Following [6], source data were also used in the fine-tuning stage. Each mini-batch contains 64 source-domain images of four identities and 64 target-domain images of four pseudo identities, with 16 images for each identity. We used the clustering algorithm of DBSCAN [31]. The maximum distance between neighbors was set to $eps = 0.6$ and the minimum number of neighbors for a dense point was set to four for DBSCAN. The hyper-parameter of the optimizer was the same as that in the pre-training stage. All experiments were conducted on NVIDIA GeForce RTX 2080 Ti with 12GB VRAM.

4.3. Analysis of Intermediate Domains

4.3.1. Qualitative Evaluation

Intermediate domains are proposed to bridge the source and target domains. The style of the images is gradually changed in the training process of GAN. A large domainness factor λ indicates that the image style is closer to the target domain. Figure 5 shows the effect of the loss function with different λ values in Equation (13) on the image generated by GAN. Illumination is a very important factor that affects the image style for ReID datasets. The gradual enhancement of the illumination of the image with the increase in λ proves that the domainness factor λ can control the style of generated images (Figure 5).

The data distribution of intermediate domains is illustrated in Figure 6. DukeMTMC-reID and Market1501 are the source and target domains for GAN, respectively. Features are extracted by the model trained on the Market1501 dataset. Red and blue stars denotes the features of Market1501 and DukeMTMC-reID datasets, respectively. The d2m_0.5 denotes features of the intermediate domain with $\lambda = 0.5$; that is, the images' own half of the style

of source data. The result of visualization showed that the data distribution for $\lambda = 0.8$ is closer to the Market1501 (target) dataset compared to that for $\lambda = 0.5$. This finding is consistent with the conclusion that the intermediate domain with a larger domainness factor λ is closer to the target domain.

Figure 5. The images of intermediate domains. λ denotes the domainness factor, which decides the style of the images generated by GAN.
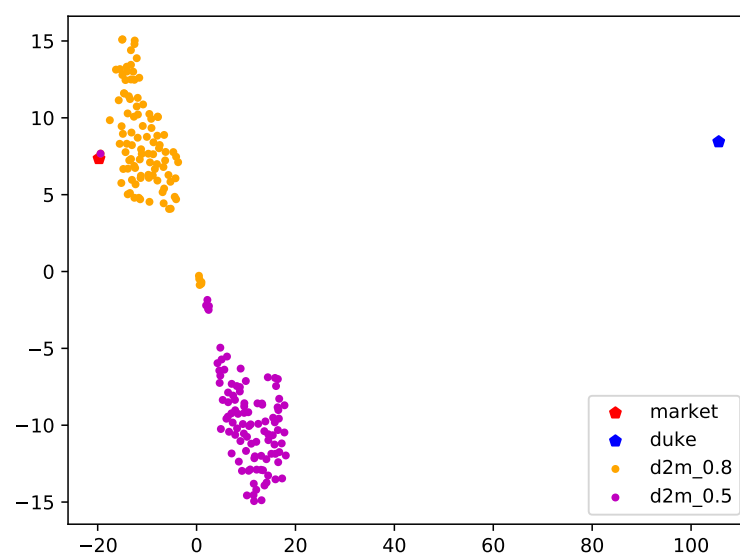

Figure 6. Illustration of data distribution of intermediate domains. The red and blue stars represent the data distribution of Market1501 and DukeMTMC-reID dataset, respectively. d2m_0.5 denotes the features of images generated by GAN, which transfers the style from DukeMTMC-reID to Market1501 for $\lambda = 0.5$.

4.3.2. Quantitative Evaluation

We conducted ablation experiments on several intermediate domains to evaluate the effectiveness of them. We first pre-trained the model on the source domain and then fine-tuned the model on several intermediate domains. The experiment results are shown in Figures 7 and 8.
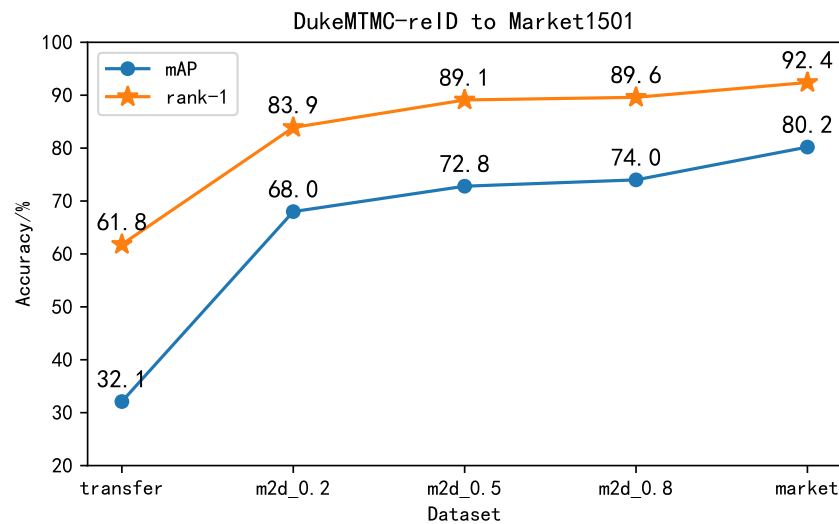

Figure 7. The results of ablation experiments for DukeMTMC-reID to Market1501 dataset. Transfer represents the model trained on source domain and then directly evaluated on target domain. m2d_0.2 denotes that we conducted UDA baseline on intermediate domain with $\lambda = 0.2$.
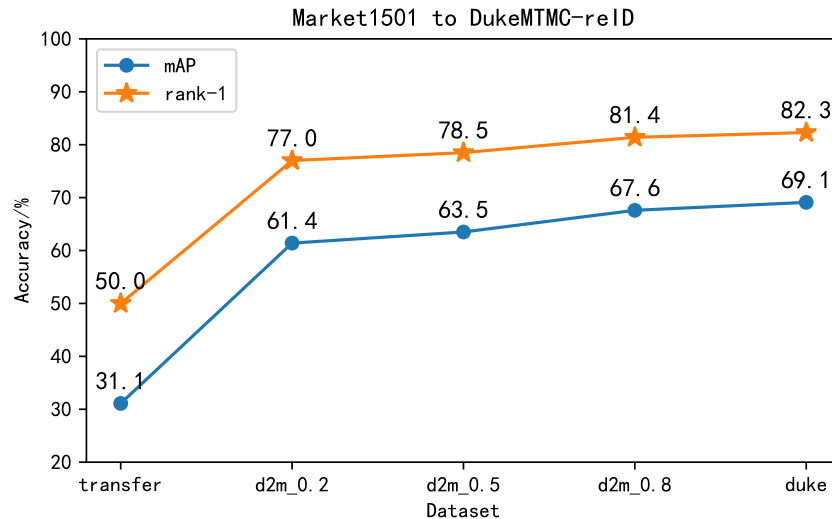
Figure 8. The results of ablation experiments for Market1501 to DukeMTMC-reID dataset.

The results suggested that the model suffers a serious reduction in accuracy when we directly transfer the model trained on the source domain to the target domain. For example, the model pre-trained on the DukeMTMC-reID dataset only achieved 61.8% of rank-1 accuracy on the Market1501 dataset. The model achieved 83.9% and 68.9% rank-1 and mAP accuracy, respectively, after fine-tuning on the m2d_0.2 dataset. We then regarded the m2d_0.2 dataset as the source domain and the m2d_0.5 dataset as the target domain, and the rank-1 accuracy improved by 5.2%, which further improved by 0.5% from m2d_0.5 to m2d_0.8. Finally, the model was fine-tuned on the Market1501 dataset and achieved rank-1 and mAP accuracy of 92.4% and 80.2%, respectively. The results of Market1501 to DukeMTMC-reID are consistent with those of DukeMTMC-reID to Market1501.

Furthermore, the performance of the model gradually improves with the increase in the domainness factor λ. A large λ indicates that the image style is closer to the target domain. Hence, the intermediate domains do serve as a bridge between the source and target domain. The model smoothly adapts between the source and target domains and transfers the source knowledge more effectively in order to improve the model's performance on the target domain. That is, fine-tuning along the bridge can benefit the cross-domain performance of the model.

The results of the Market1501 to MSMT17 dataset and DukeMTMC-reID to MSMT17 datasets are listed in Tables 2 and 3, respectively. All of the above results of the ablation experiments further verify the effectiveness of intermediate domains and the domainness factor in SPGAN.

Table 2. The results of ablation experiments for Market1501 to MSMT17 dataset.

Dataset	mAP	Rank-1	Rank-5	Rank-10
market(pre-train)	83.2	93.9	97.7	98.5
directly transfer	6.9	19.4	28.7	33.9
msmt2m_0.2	12.3	30.1	40.7	46.1
msmt2m_0.5	19.0	42.6	53.5	59.0
msmt2m_0.8	22.8	48.3	59.8	65.0
msmt17	27.6	55.5	66.8	71.7

Table 3. The results of ablation experiments for DukeMTMC-reID to MSMT17 dataset.

Dataset	mAP	Rank-1	Rank-5	Rank-10
duke(pre-train)	73.7	87.4	93.4	94.9
directly transfer	9.4	27.4	38.4	43.5
msmt2d_0.2	19.0	41.9	53.3	58.7
msmt2d_0.5	23.6	49.1	61.6	66.8
msmt2d_0.8	25.3	52.7	65.1	69.9
msmt17	29.5	58.3	69.7	74.3

4.4. Comparison with Other Methods

The results of the proposed method are compared with those of other SOTA methods of UDA in this section. The performance is very competitive, as illustrated in Table 4.

Table 4. The performance of different models evaluated on cross-domain datasets. Market1501→DukeMTMC-reID means that we trained the model on the Market1501 and evaluated it on the DukeMTMC-reID.

Methods		DukeMTMC-reID→Market1501				Market1501→DukeMTMC-reID			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
ECN [32]	CVPR'19	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
CR-GAN [33]	ICCV'19	54.0	77.7	89.7	92.7	48.6	68.9	80.2	84.7
SSG [34]	ICCV'19	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
MMCL [35]	CVPR'20	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.0
ACT [36]	AAAI'20	60.6	80.5	-	-	54.5	72.4	-	-
SNR [37]	CVPR'20	61.7	82.8	-	-	58.1	76.3	-	-
ECN++ [38]	TPAMI'20	63.8	84.1	92.8	95.4	54.4	74.0	83.7	87.4
AD-Cluster [39]	CVPR'20	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
DAAM [40]	AAAI'20	67.8	86.4	-	-	63.9	77.6	-	-
NRMT [17]	ECCV'20	71.7	87.8	94.6	96.5	62.2	77.8	86.9	98.5
MMT [5]	ICLR'20	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5
B-SNR+GDS-H [41]	ECCV'20	72.5	89.3	-	-	59.7	76.7	-	-
MEB-Net [18]	ECCV'20	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
SpCL [16]	NeurIPS'20	76.7	90.3	_96.2_	97.7	68.8	**82.9**	90.1	92.5
HSR [42]	ICIP'21	65.2	85.3	-	-	58.1	76.1	-	-
UNRN [6]	AAAI'21	_78.1_	_91.9_	96.1	_97.8_	**69.1**	82.0	_90.7_	**93.5**
Baseline		78.7	92.0	96.5	97.9	68.5	82.1	90.1	92.3
Ours		**80.2**	**92.4**	**97.1**	**98.1**	**69.1**	_82.3_	**90.8**	_92.8_

Table 4. *Cont.*

Methods		Market1501→MSMT17				DukeMTMC-reID→MSMT17			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
ECN [32]	CVPR'19	8.5	25.3	36.3	42.1	10.2	30.2	41.5	46.8
SSG [34]	ICCV'19	13.2	31.6	-	49.6	13.3	32.3	-	51.2
MMCL [35]	CVPR'20	15.1	40.8	51.8	56.7	16.2	43.6	54.3	58.9
ECN++ [38]	TPAMI'20	15.2	40.4	53.1	58.7	16.0	42.5	55.9	61.5
NRMT [17]	ECCV'20	19.8	43.7	56.5	62.2	20.6	45.2	57.8	63.3
MMT [5]	ICLR'20	22.9	49.2	63.1	68.8	23.3	50.1	63.9	68.8
SpCL [16]	NeurIPS'20	25.4	51.6	64.3	69.7	26.5	53.1	65.8	70.5
UNRN [6]	AAAI'21	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
Baseline		25.1	51.1	62.4	67.6	27.6	55.4	68.3	72.7
Ours		**27.6**	**55.5**	**66.8**	**71.7**	**29.5**	**58.3**	**69.7**	**74.3**

We used the strong UDA baseline mentioned in Section 3. The performance of our baseline approaches even exceeds that of many SOTA methods. The accuracy of our proposed method is further improved with the help of intermediate domains. Our method significantly outperforms the second best UDA method by 2.1%, 2.3% and 3.0%, in mAP accuracy, for Duke→Market, Market→MSMT and Duke→MSMT, respectively. Although the mAP accuracy of Duke→Market is the same as that of UNRN [6], the rank-1 accuracy outperforms it by 0.3% with a plain ResNet-50 backbone. DAAM [40] introduces an attention module and incorporates domain alignment constraints. MMT [5] uses two networks (four models) and MEB-Net [18] utilizes three networks (six models) to perform mutual mean teacher training, which presents a high computation complexity in training. Our proposed method only utilizes three intermediate domains without requiring extra computing and hardware resources, but still outperforms other approaches.

4.5. Further Analysis

4.5.1. Differences between GAN-Based Approaches and our Proposed Method

The proposed method is completely different from GAN-based approaches despite its use of GAN to translate images. Market1501 to DukeMTMC-reID is used as an example: SPGAN [9] translated the labeled images from Market1501 to DukeMTMC-reID in an unsupervised manner. The author then trained the ReID model with the translated images using supervised feature learning methods. Our method only utilizes GAN as a style transfer network. We translated unlabeled images from DukeMTMC-reID to Market1501 and regarded them as intermediate domains. We subsequently conducted an unsupervised domain adaptation baseline on these unlabeled intermediate domains on the basis of clustering algorithms. The experimental results proved that the performance of our method significantly outperforms that of GAN-based methods.

4.5.2. Number of Intermediate Domains

We set the domainness factor to $\lambda = 0.2, 0.5, 0.8$ and obtained three intermediate domains in this study. We also chose two and four intermediate domains for the experiments. The results are listed in Table 5.

Table 5. The effect of numbers of intermediate domains on the performance of the model.

Numbers of Intermediate Domains	Duke→Market		Market→Duke	
	mAP	Rank-1	mAP	Rank-1
2	78.8	92.1	67.6	81.1
3	**80.2**	**92.4**	**69.1**	**82.3**
4	79.7	92.0	67.8	81.4

The experimental results demonstrated that three intermediate domains achieve the best performance. If we only select two intermediate domains, the domain gap among

these domains is relatively great. We obtained additional wrong labels during the process of clustering.

However, the number of intermediate domains is not as large as possible. Style differences among different datasets are relatively small compared with style transfer tasks for the ReID dataset. Additional domainness factors will lead to smaller differences among intermediate domains. On one hand, it is difficult for the GAN model to generate such images. On the other hand, fine-tuning on such datasets will lead to overfitting on intermediate domains. Hence, the model achieves a poor performance on the target domain.

The combination of different intermediate domains will also exert a serious influence on the experimental results. For example, the results of m2d_0.2 and m2d_0.8 are different from those of m2d_0.5 and m2d_0.8, although both findings are worse than when three intermediate domains are used.

Furthermore, the quality of intermediate domain images plays an important role in improving the performance. We further used cycleGAN to generate intermediate datasets. The performance of the UDA baseline is presented in Table 6. The experimental result of using cycleGAN to generate the intermediate domains is approximately 2% lower than that of SPGAN because SPGAN is an improvement that focuses on the ReID problem based on cycleGAN. Hence, the quality of the image generated by SPGAN is better than that of cycleGAN. The experimental results can be improved by using an enhanced GAN model.

Table 6. The comparison of UDA performance between cycleGAN and SPGAN.

Methods	Duke→Market		Market→Duke	
	mAP	Rank-1	mAP	Rank-1
cycleGAN	77.8	90.4	68.2	81.5
SPGAN	**80.2**	**92.4**	**69.1**	**82.3**

Intermediate domains demonstrate a high application potential in many computer vision tasks, and obtaining high-quality intermediate domain images can be the focus of a future investigation.

5. Conclusions

In this paper, we focused on intermediate domains to solve the problems of noisy labels in other UDA methods for ReID. We first added a domainness factor λ in the loss function of SPGAN that can control the process of image translation. A high λ can generate images close to the target domain. The target domain of UDA is the source domain of GAN. Images generated by GAN will be regarded as intermediate domains. Pseudo labels produced by clustering algorithms present a high reliability given that intermediate domains own part of the style of the source domain.

We then conducted a clustering-based UDA baseline along a bridge consisting of a series of intermediate domains. Strategies that included mean teacher, joint fine-tuning with source data and memory bank were utilized in the baseline. The model pre-trained on the source domain was fine-tuned on intermediate domains firstly. The UDA process was conducted repeatedly given that intermediate domains contain several datasets. We adapted the model to the target domain after fine-tuning on the last intermediate domain.

Moreover, the proposed method can be easily applied to any other SOTA methods to further improve the performance of these algorithms. The results of this study can provide a reference for both the academia and industry.

Author Contributions: Conceptualization and methodology, H.X. and H.L.; software, validation, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization and project administration, H.X.; formal analysis and supervision, H.X. and J.G.; funding acquisition, W.J. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ReID	Person Re-identification
UDA	Unsupervised Domain Adaptive
GAN	Generative Adversarial Network
SPGAN	Similarity Preserving Generative Adversarial Network
CNN	Convolutional Neural Network
DLOW	Domain Flow for Adaptation and Generalization
SOTA	State-Of-The-Art
MMD	Maximum Mean Discrepancy
PTGAN	Person Transfer Generative Adversarial Network
PDA-Net	Pose Disentanglement and Adaptation Network
HHL	Hetero and Homogeneously Learning
UNRN	Uncertainty-Guided Noise Resilient Network
CycleGAN	Cycle-Consistent Adversarial Networks
MMT	Mutual Mean-Teaching
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
mAP	mean Average Precision
MEB-Net	Multiple Expert Brainstorming Network

References

1. Wang, G.; Yang, S.; Liu, H.; Wang, Z.; Yang, Y.; Wang, S.; Yu, G.; Zhou, E.; Sun, J. High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 6449–6458.
2. Luo, H.; Jiang, W.; Gu, Y.; Liu, F.; Liao, X.; Lai, S.; Gu, J. A Strong Baseline and Batch Normalization Neck for Deep Person Re-Identification. *IEEE Trans. Multimed.* **2020**, *22*, 2597–2609. [CrossRef]
3. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable Person Re-identification: A Benchmark. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1116–1124.
4. Ristani, E.; Solera, F.; Zou, R.S.; Cucchiara, R.; Tomasi, C. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2016; pp. 17–35.
5. Ge, Y.; Chen, D.; Li, H. Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification. In Proceedings of the ICLR 2020: The Eighth International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
6. Zheng, K.; Lan, C.; Zeng, W.; Zhang, Z.; Zha, Z.J. Exploiting Sample Uncertainty for Domain Adaptive Person Re-Identification. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 3538–3546.
7. Cui, Z.; Li, W.; Xu, D.; Shan, S.; Chen, X.; Li, X. Flowing on Riemannian manifold: Domain adaptation by shifting covariance. *IEEE Trans. Syst. Man Cybern.* **2014**, *44*, 2264–2273.
8. Gong, R.; Li, W.; Chen, Y.; Gool, L.V. DLOW: Domain Flow for Adaptation and Generalization. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2477–2486.
9. Deng, W.; Zheng, L.; Ye, Q.; Kang, G.; Yang, Y.; Jiao, J. Image-Image Domain Adaptation with Preserved Self-Similarity and Domain-Dissimilarity for Person Re-identification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 994–1003.
10. Lin, S.; Li, H.; Li, C.T.; Kot, A.C. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. In Proceedings of the BMVC 2018: Proceedings of the 29th British Machine Vision Conference, Newcastle, UK, 3–6 September 2018; p. 9.

11. Wang, J.; Zhu, X.; Gong, S.; Li, W. Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-identification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2275–2284.

12. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.

13. Wei, L.; Zhang, S.; Gao, W.; Tian, Q. Person Transfer GAN to Bridge Domain Gap for Person Re-identification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 79–88.

14. Lv, J.; Wang, X. Cross-Dataset Person Re-identification Using Similarity Preserved Generative Adversarial Networks. In *International Conference on Knowledge Science, Engineering and Management*; Springer: Cham, Switzerland, 2018; pp. 171–183.

15. Zhong, Z.; Zheng, L.; Li, S.; Yang, Y. Generalizing A Person Retrieval Model Hetero- and Homogeneously. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2018; pp. 172–188.

16. Ge, Y.; Zhu, F.; Chen, D.; Zhao, R.; Li, H. Self-paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 11309–11321.

17. Zhao, F.; Liao, S.; Xie, G.S.; Zhao, J.; Zhang, K.; Shao, L. Unsupervised Domain Adaptation with Noise Resistible Mutual-Training for Person Re-identification. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2020; pp. 526–544.

18. Zhai, Y.; Ye, Q.; Lu, S.; Jia, M.; Ji, R.; Tian, Y. Multiple Expert Brainstorming for Domain Adaptive Person Re-Identification. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2020; pp. 594–611.

19. Li, R.; Jiao, Q.; Cao, W.; Wong, H.S.; Wu, S. Model Adaptation: Unsupervised Domain Adaptation Without Source Data. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9641–9650.

20. Zhou, B.; Kalra, N.; Krahenbuhl, P. Domain Adaptation Through Task Distillation. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2020; pp. 664–680.

21. Huang, F.; Zhang, L.; Yang, Y.; Zhou, X. Probability Weighted Compact Feature for Domain Adaptive Retrieval. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9582–9591.

22. Zhou, K.; Yang, Y.; Hospedales, T.M.; Xiang, T. Learning to Generate Novel Domains for Domain Generalization. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2020; pp. 561–578.

23. Zhou, K.; Yang, Y.; Qiao, Y.; Xiang, T. Domain Generalization with MixStyle. In Proceedings of the ICLR 2021: The Ninth International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021.

24. Wang, H.; He, H.; Katabi, D. Continuously Indexed Domain Adaptation. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; Volume 1, pp. 9898–9907.

25. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251.

26. Taigman, Y.; Polyak, A.; Wolf, L. Unsupervised Cross-Domain Image Generation. In Proceedings of the ICLR 2016: The Fourth International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.

27. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality Reduction by Learning an Invariant Mapping. In Proceedings of the 2006 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1735–1742.

28. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

29. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the ICLR 2015: The Third International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.

30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

31. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial Databases with Noise. In Proceedings of the 2nd International Conference on Knowledg Discovery and Data Mining (KDD-96), Portland, OR, USA, 2–4 August 1996; pp. 226–231.

32. Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; Yang, Y. Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-Identification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 598–607.

33. Chen, Y.; Zhu, X.; Gong, S. Instance-Guided Context Rendering for Cross-Domain Person Re-Identification. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 232–242.

34. Fu, Y.; Wei, Y.; Wang, G.; Zhou, Y.; Shi, H.; Uiuc, U.; Huang, T. Self-Similarity Grouping: A Simple Unsupervised Cross Domain Adaptation Approach for Person Re-Identification. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 6111–6120.

35. Wang, D.; Zhang, S. Unsupervised Person Re-Identification via Multi-Label Classification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10981–10990.

36. Yang, F.; Li, K.; Zhong, Z.; Luo, Z.; Sun, X.; Cheng, H.; Guo, X.; Huang, F.; Ji, R.; Li, S. Asymmetric Co-Teaching for Unsupervised Cross-Domain Person Re-Identification. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 12597–12604. [CrossRef]

37. Jin, X.; Lan, C.; Zeng, W.; Chen, Z.; Zhang, L. Style Normalization and Restitution for Generalizable Person Re-Identification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3143–3152.

38. Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; Yang, Y. Learning to Adapt Invariance in Memory for Person Re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2723–2738. [CrossRef] [PubMed]

39. Zhai, Y.; Lu, S.; Ye, Q.; Shan, X.; Chen, J.; Ji, R.; Tian, Y. AD-Cluster: Augmented Discriminative Clustering for Domain Adaptive Person Re-Identification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9021–9030.

40. Huang, Y.; Peng, P.; Jin, Y.; Xing, J.; Lang, C.; Feng, S. Domain Adaptive Attention Model for Unsupervised Cross-Domain Person Re-Identification. *arXiv* **2019**, arXiv:1905.10529.

41. Jin, X.; Lan, C.; Zeng, W.; Chen, Z. Global Distance-Distributions Separation for Unsupervised Person Re-identification. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2020; pp. 735–751.

42. Liu, C.T.; Lee, M.Y.; Chen, T.S.; Chien, S.Y. Hard samples rectification for unsupervised cross-domain person re-identification. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 429–433.