Smoothing Adversarial Domain Attack and p-Memory Reconsolidation for Cross-Domain Person Re-Identification

Guangcong Wang[1], Jianhuang Lai[1,2,3]*, Wenqi Liang[1], and Guangrun Wang[1]
[1]School of Data and Computer Science, Sun Yat-sen University, China
[2]Guangdong Key Laboratory of Information Security Technology
[3]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education
{wanggc3,liangwq8,wanggrun}@mail2.sysu.edu.cn, stsljh@mail.sysu.edu.cn

Abstract

Most of the existing person re-identification (re-ID) methods achieve promising accuracy in a supervised manner, but they assume the identity labels of the target domain is available. This greatly limits the scalability of person re-ID in real-world scenarios. Therefore, the current person re-ID community focuses on the cross-domain person re-ID that aims to transfer the knowledge from a labeled source domain to an unlabeled target domain and exploits the specific knowledge from the data distribution of the target domain to further improve the performance. To reduce the gap between the source and target domains, we propose a Smoothing Adversarial Domain Attack (SADA) approach that guides the source domain images to align the target domain images by using a trained camera classifier. To stabilize a memory trace of cross-domain knowledge transfer after its initial acquisition from the source domain, we propose a p-Memory Reconsolidation (pMR) method that reconsolidates the source knowledge with a small probability p during the self-training of the target domain. With both SADA and pMR, the proposed method significantly improves the cross-domain person re-ID. Extensive experiments on Market-1501 and DukeMTMC-reID benchmarks show that our pMR-SADA outperforms all of the state-of-the-arts by a large margin.

1. Introduction

Person re-identification (re-ID) aims at re-targeting person images across non-overlapping camera views given a query image. Recently, most of the existing person re-ID approaches achieve a dramatic improvement using a large number of annotations. However, these person re-ID systems often assume that large-scale identity labels of a target domain are available, which greatly limits their scalability in real-world scenarios. When applied to a new scenario, they suffer from serious performance degradation, e.g., from 92.0% to 47.5% on the Market-1501 dataset [52]. How to transfer the knowledge from a labeled source domain to an unlabeled target domain for person re-ID and how to exploit the specific knowledge from an unlabeled target domain to boost person re-ID methods become hot topics in the current person re-ID community. It is widely called *cross-domain person re-ID*.

Recently, lots of cross-domain person re-ID methods have achieved promising progress. A common approach to solve this problem is to directly apply a pre-trained model of a source domain to a target domain for evaluation. However, there exists a large domain gap between the source and target domains due to different lightings, backgrounds, poses, and camera views. To deal with the large domain gap problem, several GAN based methods [6, 19, 4, 18] are proposed. For example, [6] proposed to directly use a cycle generative adversarial model to reduce the domain gap problem. However, due to the diversity of person images, the person identity information of the generated images is hard to preserve without any identity constraint. Based on [6], [19, 44, 4] used person segmentation tools to extract person masks as extra information to help GAN to preserve the person identity information. Although they can greatly improve the performance of the cross-domain person re-ID, they largely depend on the performance of the person segmentation task that implicitly requires person mask annotations. These extra annotations could also limit their scalability in practice.

Instead of focusing on the image-based knowledge transfer, some researchers attempt to exploit the underlying data distributions of the target domain by person identity clustering for self-training. For example, a progressive unsupervised learning method [8] that iterates k-means clustering and CNN fine-tuning is proposed to learn discriminative features for person re-ID. Based on [8], several clustering

*Corresponding author

methods [31, 23] use DBSCAN and hierarchical clustering methods to mine positive and negative samples of the target domain for pseudo-label self-training. However, these approaches simply use the pre-trained model of the source domain as an initial model for the feature learning of the target domain. After several iterations, the transferred knowledge from the source domain is gradually forgotten due to the decline of memory retention of convolutional neural networks (CNNs).

To address these two problems, we propose a new approach by narrowing the domain gap and strengthening the source transferred knowledge. Specifically, to reduce the domain bias between the source and target domains, we propose a smoothing domain attack approach to align the source domain towards the target domain at the image level. Given a source domain D_s with N_s cameras (labeled as $1, 2, ..., N_s$) and a target domain D_t with N_t cameras (labeled as $N_s+1, N_s+2, ..., N_s+N_t$), a (N_s+N_t)-category camera-based classifier is trained by predicting the labels of these cameras. Given a source image, we randomly generate a target camera label as the new label of the source image to align the target domain. This is achieved by fixing the weights of the classifier and allowing the smoothed gradient to change the pixels of the source images. To avoid the decline of memory retention of the source knowledge, we propose to reconsolidate the source knowledge with a small probability p during the self-training of the target domain. Instead of directly applying the source pre-trained model to the target domain, we propose to select the source dataset to reconsolidate the source memory with a random variable that follows a Bernoulli distribution. With both SA-DA and pMR, the proposed method significantly improves the cross-domain person re-ID.

Overall, the contributions of this paper are:

- We propose a smoothing adversarial domain attack (SADA) approach to force the source images to align the target images at the image-level, so as to reduce the gap between the source and target domains.
- We propose a p-Memory Reconsolidation (pMR) approach that has a small probability p to avoid the decline of memory retention of the source knowledge.
- Extensive experiments on Market-1501 and DukeMTMC-reID benchmarks show that our pMR-SADA outperforms all of the state-of-the-arts by a large margin.

2. Related Work

Supervised person re-ID. Most of the current person re-ID methods focus on deep based models [41, 36, 37, 56] due to the remarkable representation capacity of neural networks [11, 42, 40]. Early works [51, 46, 7, 21, 39, 5, 3] designed different network structures and loss functions to boost the performance of re-ID. Recently, some body-part based methods [33, 50, 49, 32, 35, 10, 38] and attention-based methods [9, 17, 34, 55] have been proposed to further improve accuracy. The part-based models split the last convolutional feature maps into several parts to learn discriminative local feature representations by considering human body structures. The attention-based approaches aim at emphasizing informative regions while depreciating harmful ones. Although these methods achieve high performance, they require large-scale labeled datasets in the target domain, which limits their scalability in real-world scenarios.

Unsupervised cross-domain person re-ID. The unsupervised cross-domain person re-ID aims at transferring the knowledge from a labeled source domain to an unlabeled target domain, which reduces the time-consuming labeling works on the target domain. The current methods can be divided into two groups. In the first group, some methods attempt to reduce the domain gap at the image or feature levels by designing consistent domain losses. For example, Lin et al. [22] proposed to reduce the data distribution discrepancy between the source domain and the target domain by minimizing the maximum mean discrepancy (MMD) of the two domains. Lei et al. [28] extended to the camera-level discrepancy and used a gradient reversal layer to reduce distribution discrepancy. Recent methods focused on GAN-based style transfer methods [6, 44, 4, 18, 25], which can transform images from source domain styles to target domain styles. For example, M2M-GAN [19], PTGAN [44], and CR-GAN [4] translated image styles by adding a mask-based person identity loss to preserve identity information. However, most of these method requires other extra information (e.g., person masks), which implicitly used other expensive annotations and thus could limit their scalability in practice. In the second group, some methods focus on exploiting the data distribution characteristics of the target domain and estimating the labels of the target dataset for self-training. For example, [8, 31, 23] used K-means, DBSCAN or hierarchical clustering methods to mine positive and negative samples of the target domain. Our proposed method belongs to this group and is built upon the clustering-based algorithm. The main differences are that our approach method considers a smoothing domain attack to align the source and target domains and introduces a new p-memory reconsolidation algorithm to solve the decline of memory retention of the source transferred knowledge.

Continual learning. Continual learning (CL) performs learning through a sequence of tasks and it is often assumed CL accesses to one task at once. CL aims to improve previous, current and future learning tasks, especially previous learning tasks due to catastrophic forgetting [26]. To reduce catastrophic forgetting, early works [2, 30] focused on studying linear models to retain knowledge. Recently, lots of deep models [1, 29] used a shared backbone and n specific branches for n learning tasks. Besides, [26, 14, 13]proposed regularization approaches that adjust the learning objective to prevent catastrophic forgetting. Our proposed pMR differs from CL in several aspects. **First,** pMR focuses on one-directional learning (source→target) while CL has to balance both previous and new learning tasks. **Second,** pMR can simultaneously access to all of the tasks while CL accesses to one task at once. Also, pMR focuses on similar learning tasks while CL does not has such a constraint. **Third,** pMR only focuses on restoring shared knowledge while CL attempts to memory all of the knowledge of previous learning tasks, which leads to different optimization algorithms.

3. The Proposed Method

Cross-domain person re-ID is to transfer the knowledge from a labeled source domain to an unlabeled target domain. In this section, we first provide an overview of the proposed method in Section 3.1. Then we focus on two main components of our method in Section 3.2 and Section 3.3. After that, a complete algorithm is provided in Section 3.4.

3.1. Overview of the Proposed Framework

The overview of our proposed method is illustrated in **Figure** 1, which consists of two main components, i.e., a smoothing adversarial domain attack (SADA) and a p-memory Reconsolidation (pMR). Specifically, the SADA module is to force the source images to align the target domain distribution at the image level. After that, the aligned source images are used to pre-train a deep model. The pre-trained model is then transferred to the target domain for Density-Based Spatial Clustering (DBSC). Due to the fact that convolutional neural networks have no memory capacity, one would think if the knowledge transferred from the source domain will be forgotten during the self-training of the target domain. To solve this problem, a pMR module is naturally introduced. The pMR module takes the aligned source images as input with a probability p and takes the target images with probability $1-p$. With the proposed pMR approach, we found that pMR obtains consistent improvement in all of the experiments against direct transfer.

3.2. Smoothing Adversarial Domain Attack

SADA is an iterative adversarial attacker that aims at aligning a camera-based distribution of the source images to be that of the target images. We need to make sure that each person ID in the aligned source domain contains different target camera classes, which enables the cross-camera feature learning against the camera view variances. In this paper, the source images of each person ID are randomly aligned to target camera classes, which satisfies this requirement. Let $D_s = \{(I_i, z_i)\}_{i=1}^{M_s}$ denote a source domain with N_s cameras ($z_i \in [1, N_s]$) and $D_t = \{(I_j, z_j)\}_{j=1}^{M_t}$ denote a target domain with N_t cameras ($z_j \in [N_s+1, N_s+N_t]$), where z represents camera identity, which is available in both the source and target domain. M_s and M_t is the number of the source and target images, respectively. We train a (N_s+N_t)-category classifier $g(\cdot;\Theta)$ that can distinguish (N_s+N_t) camera-based distributions. Given a source image (I_i, z_i) (the subscript i of I is omitted for simplicity below) and a random target camera ID z_j, the alignment of the source domain and the target domain is achieved by an iterative adversarial attacker [16], which can be formulated as

$$\begin{cases} I_0^{adv} & = I \\ I_{k+1}^{adv} & = Clip_{I,\epsilon}\{I_k^{adv} - \nabla I_k^{adv}\} \end{cases} \tag{1}$$

where I_k^{adv} is an adversarial result of the k-th iteration. $Clip_{I,\epsilon}$ represents a function which performs per-pixel clipping of the image. The gradient is computed by

$$\nabla I_k^{adv} = \alpha sign(\nabla_I J(I_k^{adv}, z_j)) \tag{2}$$

z_j is a random camera ID of the target domain. $J(I,z)$ is the cross-entropy function of the neural network and $J(I,z) = -logp(z|g(I;\Theta))$. Note that we fix Θ during the attack process. Eqs. (1) and (2) attempt to push the source domain towards the target domain at the image level until the source domain is guided to be classified as the given target camera z_j. One drawback of Eq. (2) is that the output of the sign function could lead to uneven pixel changes (isolated points) due to the inaccurate location of the gradient. Those isolated points violate the smoothness of the natural images and thus could increase domain attack iterations to reverse the source camera ID z_i as the target camera ID z_j. To this, we propose a smooth iterative adversarial domain attack approach, which can be formulated as

$$\nabla I_k^{adv} = \alpha sign(smooth(\nabla_I J(I_k^{adv}, z_j))) \tag{3}$$

where $smooth(\cdot)$ denotes a smoothing function. In this paper, we use a mean filter as a choice of $smooth(\cdot)$. We will show the fast convergence of the smoothing adversarial domain attack in the experiment section.

3.3. p-Memory Cross-Domain Transfer Learning

3.3.1 Review of Clustering Based Cross-Domain Re-ID

We provide a brief introduction for a conventional clustering based cross-domain person re-ID approach. Let $\mathcal{D}_s = \{(I_i^{'}, y_i)\}_{i=1}^{M_s}$ denote a labeled source domain and $\mathcal{D}_t = \{I_i\}_{i=1}^{M_t}$ denote an unlabeled target domain, where y_i represents a person identity and $I_i^{'}$ denotes an aligned image of I_i. One can use a supervised learning algorithm to pre-train an initial feature extractor $f(\cdot;W_0)$ parameterized by W_0 on $\mathcal{D}_s$. Let $\mathbf{x}_t^i = F(I;w_0)$ and $\mathbf{X}_t =$

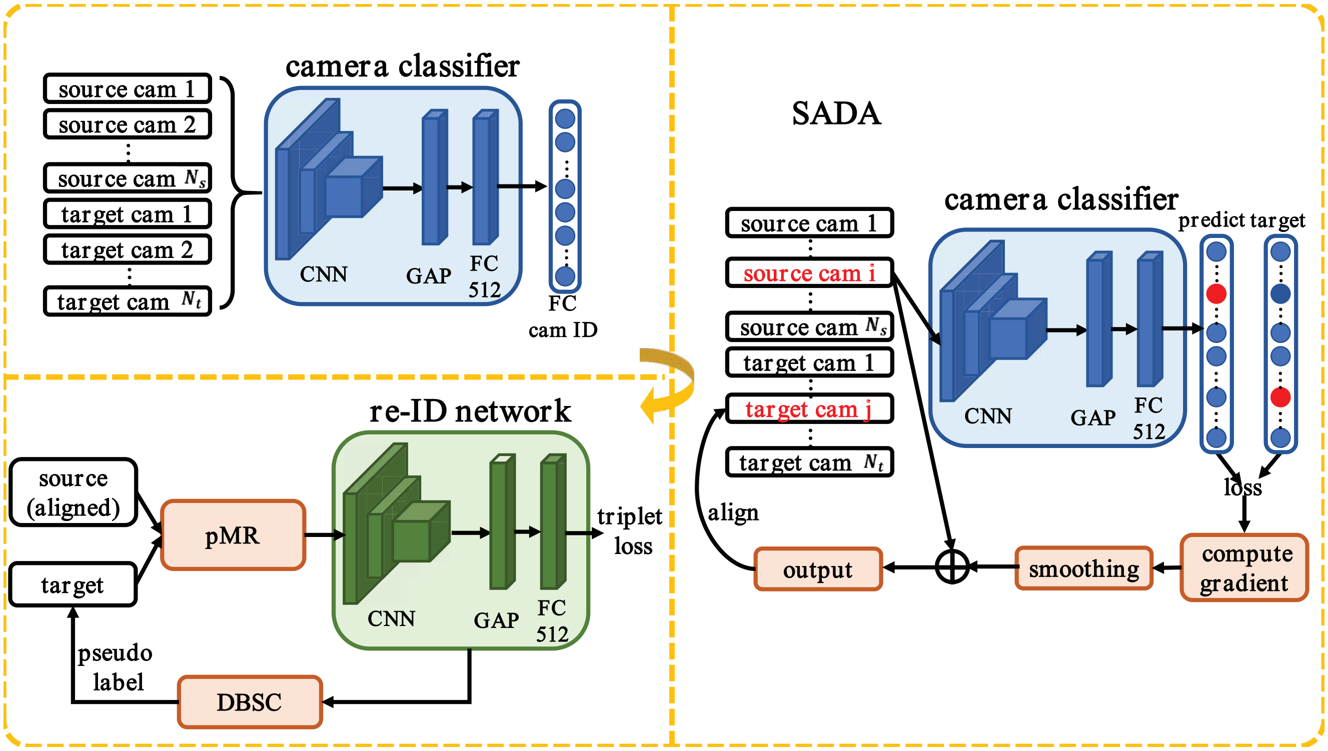

Figure 1. The overview of our proposed method. It consists of two main components, i.e., a smoothing adversarial domain attack (SADA) and a p-memory Reconsolidation (pMR). First, we train a $(N_s + N_t)$-class camera-based classifier using both the source and target images. Second, this camera-based classifier forces the source images to align the target images. Third, the aligned source images are used to pre-train a deep model as a transferred model that is applied to the target domain for Density-Based Spatial Clustering (DBSC). Because the knowledge transferred from the source domain will be forgotten during the target distribution mining, a pMR module is naturally introduced to reconsolidate the transferred knowledge. The pMR module takes the aligned source images as input with probability p and the target images with probability $1 - p$.

$\{\mathbf{x}_t^1, \mathbf{x}_t^2, ..., \mathbf{x}_t^{M_t}\}$ be a set of target feature vectors and C denote a clustering algorithm. We perform the algorithm C and obtain $\{\mathbf{X}_t^*, \bar{\mathbf{X}}_t^*\} = C(\mathbf{X}_t)$, where $\mathbf{X}_t^* \cap \bar{\mathbf{X}}_t^* = \emptyset$ and $\mathbf{X}_t^* \cup \bar{\mathbf{X}}_t^* = \mathbf{X}_t$. $\mathbf{X}_t^*$ denotes confident instances with pseudo-labels y_t^* and $\bar{\mathbf{X}}_t^*$ denotes noisy ones at the current iteration. We construct a pseudo-labeled subset $\mathcal{D}_t^*$ by using the corresponding indexes of $\mathbf{X}_t^*$ and their pseudo-labels y_t^*. Finally, a supervised learning [12] algorithm on $\mathcal{D}_t^*$ is given as follows

$$\mathcal{L}_{triplet}^{t} = \sum_{i=1}^{P}\sum_{a=1}^{K}[m + \overbrace{\max_{p=1...K} dist(\mathbf{x}_a^i, \mathbf{x}_p^i)}^{\text{hardest positve}} - \underbrace{\min_{\substack{j=1...P\\ n=1...K\\ j\neq i}} dist(\mathbf{x}_a^i, \mathbf{x}_n^j)}_{\text{hardest negative}}]_+ \tag{4}$$

where $\mathbf{x}_a$, $\mathbf{x}_p$, and $\mathbf{x}_n$ represent an anchor instance, a positive instance, and a negative instance, respectively. P denotes person classes/identities. For each person identity, K images are sampled. m denotes a margin. $dist(\cdot)$ denotes a distance metric. $\mathbf{x}_j^i$ corresponds to the j-th feature vector of the i-th person in the batch. Eq. (4) aims at the self-training of the target domain. The supervised learning and clustering alteratively iterates.

3.3.2 p-Memory Reconsolidation

Directly applying the pre-trained model that trains on the source images to the target domain often provides a good initialization for further learning of the target domain. We refer to this pipeline as *Direct Transfer* for simplicity. Direct transfer implicitly assumes that knowledge in the source domain can be fully transferred to the target domain and there is no need for another transfer. However, such an assumption does not hold because residual convolutional neural networks have no memory units like recurrent neural networks and thus cannot memory the transferred knowledge after the long-term iterations, especially in self-training that could take too much time for the alternative iterations between feature learning and clustering. Therefore, the memory of the transferred knowledge would gradually

Algorithm 1: Smoothing Adversarial Domain Attack Algorithm

```
Input: $D_s = \{(I_i, z_i)\}_{i=1}^{M_s}$, $D_t = \{(I_j, z_j)\}_{j=1}^{M_t}$,
       $z_i \in \{1, 2, ..., N_s\}$,
       $z_j \in \{N_s + 1, N_s + 2, ..., N_s + N_t\}$,$\mathcal{D}_s = \{(I_i, y_i)\}_{i=1}^{M_s}$
Output: An aligned source dataset
        $\mathcal{D}^{'} = \{(I_k, y_k)\}_{k=1}^{M_s}$
Train a $(N_s + N_t)$-category camera classifier $g(\cdot; \Theta)$ with $\{D_s \cup D_t\}$;
$\mathcal{D}^{'} = \emptyset$;
for $i = 1 : M_s$ do
    Randomly select a target camera label $z \in \{N_s + 1, N_s + 2, ..., N_s + N_t\}$;
    $I = I_i$;
    while Prediction of $I$ by $g$ is not $z$ do
        Compute $\nabla I$ according to Eq. (3);
        $I \leftarrow Clip_{I,\epsilon}\{I - \nabla I\}$;
    end
    $\mathcal{D}^{'} \leftarrow \mathcal{D}^{'} \cup (I, y_i)$
end
```

be forgotten during the self-training on the target domain. We call this the *transfer amnesia* problem.

Figure 2 shows that there exists a transfer amnesia problem when the traditional cluster-based cross-domain model is adopted for the transfer from the DukeMTMC-reID dataset to the Market-1501 dataset. We evaluate the rank-1 accuracy and mean Average Precision (mAP) on the test set of DukeMTMC-reID during the self-training on Market-1501. The self-training model is initialized with a pre-trained model on DukeMTMC-reID. It is observed that without using any constraint, both rank-1 accuracy and mAP gradually decreases as iterations increase, showing that the transferred knowledge is lost over the self-training of the target domain. This finding is consistent with the forgetting curve of human memory.

To solve the transfer amnesia problem, we propose a p-memory reconsolidation approach with a small probability p that stabilizes a memory trace after its initial transfer. We formulate it as

$$\mathcal{L} = (1 - \xi)\mathcal{L}_{triplet}^{t} + \xi\mathcal{L}_{triplet}^{s} \tag{5}$$

where ξ follows a Bernoulli distribution $B(1, p)$ with a small probability p. A small p can reconsolidate the cross-domain transferred knowledge and often accelerates the self-training because the source dataset is often much larger than the confident pseudo-labeled set. $\mathcal{L}_{triplet}^{s}$ is similar to $\mathcal{L}_{triplet}^{t}$, but defined on the source domain. For each epoch, we optimize Eq. (5) by sampling the source dataset or the pseudo-labeled target dataset following a Bernoulli distribution with a probability p to consolidation memory.

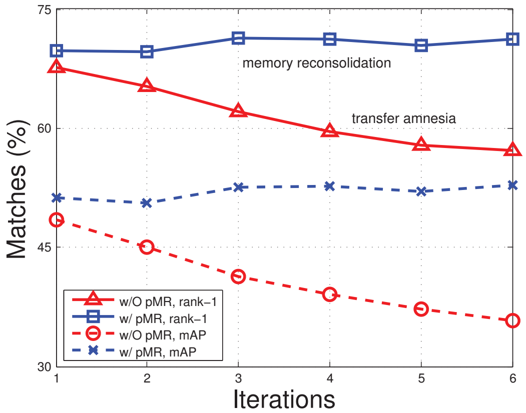

Figure 2. Transfer amnesia problem. We evaluate the rank-1 accuracy and mean Average Precision (mAP) on the test set of the DukeMTMC-reID dataset during the self-training on the Market-1501 dataset. The self-training model is initialized with the pre-trained model on DukeMTMC-reID.

3.4. Algorithm

In this section, we provide an overview of the proposed algorithm, as shown in Algorithm 2. Specifically, we train a $(N_s + N_t)$-category camera-based classifier and use it to guide the source images to align the target images based on the SADA algorithm at the image level. With the aligned source images, we train a feature extractor as the model initialization. When alternatively iterating clustering and feature representation learning during the self-training, the pMR approach is proposed to solve the transfer amnesia problem and further improve the cross-domain person re-ID.

Algorithm 2: Algorithm Overview

```
Input: $D_s = \{(I_i, z_i)\}_{i=1}^{M_s}$, $D_t = \{(I_j, z_j)\}_{j=1}^{M_t}$,
       $\mathcal{D}_s = \{(I_i, y_i)\}_{i=1}^{M_s}$, $\mathcal{D}_t = \{I_i\}_{i=1}^{M_t}$
Output: A feature extractor $f(\cdot; W^*)$
Compute aligned source images based on Algorithm 1 and return $\mathcal{D}^{'}$ ;
Pre-train the model on $\mathcal{D}^{'}$ and obtain $W_0$;
Initialize $f(\cdot; W)$ with $W_0$;
for $i = 1 : iters$ do
    Perform the clustering algorithm $C$ on $\mathcal{D}_t$ and construct a pseudo-labeled dataset $\mathcal{D}_t^*$;
    for $j = 1 : epochs$ do
        Optimize $f(\cdot; W)$ by Eq. (5);
    end
end
```

4. Experiments

In this section, we evaluate our proposed method on two large-scale person re-ID benchmark datasets, i.e., Market-1501 [52] and DukeMTMC-reID [53]. We compare our proposed method with state-of-the-art methods in Section 4.1 and Section 4.2. We then present ablation studies to reveal the importance of each main component/factor of our method in Section 4.3.

Implementation. To align the source and target domains, we train a camera-based classifier using the Adam optimizer with a mini-batch size of 32. The learning rate is set to 0.0003 and is divided by 10 after 30 and 50 epochs. We train the network for 70 epochs. When performing smoothing adversarial domain attack, we normalize the source images and set $\epsilon = 0.02$ for "Market-1501↦DukeMTMC-reID" and $\epsilon = 0.001$ for "DukeMTMC-reID↦Market-1501". The maximum attack iteration is 50. When performing self-training, we use the SGD optimizer. The learning rate is set to $6e - 5$. We train 10 iterations. For each iteration, we train 70 epochs. The mini-batch size is 128. We set $p = 0.3$ for the Bernoulli distribution. We the DBSCAN clustering method for self-training.

Evaluation Metrics. We adopt the standard Cumulative Match Characteristic (CMC) and mean Average Precision (mAP) as evaluation metrics.

Datasets. The Market-1501 dataset with six cameras is collected at Tsinghua University. Overlap exists among different cameras. Overall, this dataset contains 32,668 annotated bounding boxes of 1,501 identities. Among them, 12,936 images from 751 identities are used for training, and 19,732 images from 750 identities plus distractors are used for the gallery. As for query, 3,368 hand-drawn bounding boxes from 750 identities are adopted. Each annotated identity is present in at least two cameras.

DukeMTMC-reID has 8 cameras. There are 1,404 identities appearing in more than two cameras and 408 identities (distractor ID) who appear in only one camera. Specially, 702 IDs are selected as the training set and the remaining 702 IDs are used as the testing set. In the testing set, one query image is picked for each ID in each camera and the remaining images are put in the gallery. In this way, there are 16,522 training images of 702 identities, 2,228 query images of the other 702 identities and 17,661 gallery images (702 ID + 408 distractor ID).

4.1. Comparison with the State-of-the-art on the Market-1501 Dataset

We compare our proposed method with a broad range of existing state-of-the-art approaches on the Market-1501 dataset. The experimental results are reported in **Table 1**. It is encouraging to see that our proposed outperforms all of the state-of-the-art approaches by a large margin. In particular, the competing methods can be classified into two groups. The first group includes five hand-crafted models, i.e., LOMO [20], BoW [52], DIC [15], ISR [24], and UDML [27]. Compared with the best handcrafted feature representation DIC, the proposed model obtains 32.8%, 23.0%, and 37.1% improvement on rank-1, rank-5, and mAP metrics, respectively. The experimental results on the Market-1501 dataset clearly show the superiority of our approach against other hand-crafted feature representations, even though we only use the labels of the source domain for transfer learning and self-training on the unlabeled target domain.

The second group includes ten state-of-the-art deep approaches, i.e., CAMEL [47], PUL [8], TJ-AIDL [43], MAR [48], ATNet [25], UCDA-CCE [28], CR-CAN+TAUDL [4] (denoted as CR-CAN+ in **Table 1**), CASCL [45], ECN [54], and PDA-Net [18]. Among these deep models, UCDA-CCE [28], ATNet [25], CASCL [45] focused on domain adaption while ignored the underlying distribution of the target domain, which leads to moderate performance. TJ-AIDL [43] and PDA-Net [18] exploited extra information to improve the performance, e.g., attribute and pose. CR-CAN+TAUDL (together with feature learning on the target domain) [4] and PDA-Net [18] used generative adversarial nets (GANs) to generated target-style images and obtained good performance. CAMEL [47], PUL [8], MAR, [48], and ECN [54] attempted to explore the distributions of the target domain using self-training approaches. These approaches provide available self-space techniques and produce competitive results. Our proposed method also belongs to this pipeline but outperforms all of them. Compared with the second-best method CR-GAN+TAUDL, our model obtains 5.3%, 2.1%, 1.4% and 5.8% improvement on rank-1, rank-5, rank-10, and mAP metrics. The mAP is the most important metric in multiple camera network surveillance because it can measure the retrieval of all target person images given a query. Note that CR-GAN+TAUDL uses a segmentation method to help the image generation which implicitly used segmentation annotations. Our proposed method does not need extra information. The significant improvement of our model could be attributed to the deployment of the p-memory approach and the alignment of the image-level domains, which can deal with the transfer amnesia problem and narrow the domain gap at the image level.

4.2. Comparison with the State-of-the-art on the DukeMTMC-reID Dataset

We compare our proposed method with thirteen state-of-the-art methods on the DukeMTMC-reID dataset, which are described in Section 4.1. The experimental results are reported in **Table 2**. It is observed that our method outperforms all of the state-of-the-arts. Compared with the best handcrafted feature representation UDML, the proposed model obtains 56.0%, 53.9%, and 51.1% improvement on

Table 3. Ablation studies on the importance of each main component/factor of our proposed method. The main components include a baseline, a LMP/LAP operation, a smoothing adversarial domain attack (denoted as SADA), and a p-memory reconsolidation approach (pMR), please refer to Section 4.3 for detailed analyses.

Components				DukeMTMC-reID↦Market-1501				Market-1501↦DukeMTMC-reID			
Baseline	LMP/LAP	SADA	pMR	rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
✓	✗	✗	✗	71.4	81.6	84.9	46.3	52.2	62.6	67.2	36.5
✓	✓	✗	✗	78.1	88.2	91.6	55.5	71.0	81.6	85.0	50.5
✓	✓	✗	✓	80.3	90.3	93.1	57.7	70.1	80.9	84.1	46.7
✓	✓	✓	✗	80.7	90.1	92.7	57.4	71.9	82.9	86.4	52.3
✓	✓	✓	✓	83.0	91.8	94.1	59.8	74.5	85.3	88.7	55.8

Table 1. Comparison to the state-of-the-art unsupervised results on the Market-1501 dataset. **Red** indicates the best and **Blue** the second best. Measured by %.

Methods	Reference	Market-1501			
		rank-1	rank-5	rank-10	mAP
LOMO [20]	CVPR'15	27.2	41.6	49.1	8.0
BoW [52]	ICCV'15	35.8	52.4	60.3	14.8
DIC [15]	BMVC'15	50.2	68.8	-	22.7
ISR [24]	TPAMI'15	40.4	62.2	-	14.3
UDML [27]	CVPR'16	34.5	52.6	59.6	12.4
CAMEL [47]	ICCV'17	54.5	73.1	-	26.3
PUL [8]	ToMM'18	45.5	60.7	66.7	20.5
TJ-AIDL [43]	CVPR'18	58.2	74.8	81.1	26.5
MAR [48]	CVPR'19	67.7	81.9	-	40.0
ATNet [25]	CVPR'19	55.7	73.2	79.4	25.6
UCDA-CCE [28]	ICCV'19	64.3	-	-	34.5
CASCL [45]	ICCV'19	64.7	80.2	85.6	35.6
ECN [54]	CVPR'19	75.1	87.6	91.6	43.0
PDA-Net [18]	ICCV'19	75.2	86.3	90.2	47.6
CR-CAN+ [4]	ICCV'19	**77.7**	**89.7**	**92.7**	**54.0**
pMR-SADA	This work	**83.0**	**91.8**	**94.1**	**59.8**

Table 2. Comparison to the state-of-the-art unsupervised results on the DukeMTMC-reID dataset. **Red** indicates the best and **Blue** the second best. Measured by %.

Methods	Reference	DukeMTMC-reID			
		rank-1	rank-5	rank-10	mAP
LOMO [20]	CVPR'15	12.3	21.3	26.6	4.8
BoW [52]	ICCV'15	17.1	28.8	34.9	8.3
UDML [27]	CVPR'16	18.5	31.4	37.6	7.3
CAMEL [47]	ICCV'17	40.3	57.6	-	19.8
PUL [8]	ToMM'18	30.0	43.4	48.5	16.4
TJ-AIDL [43]	CVPR'18	44.3	59.6	65.0	23.0
MAR [48]	CVPR'19	67.1	79.8	-	48.0
ATNet [25]	CVPR'19	45.1	59.5	64.2	24.9
UCDA-CCE [28]	ICCV'19	55.4	-	-	36.7
CASCL [45]	ICCV'19	51.5	66.7	71.7	30.5
ECN [54]	CVPR'19	63.3	75.8	80.4	40.4
PDA-Net [18]	ICCV'19	63.2	77.0	82.5	45.1
CR-CAN+ [4]	ICCV'19	**68.9**	**80.2**	**84.7**	**48.6**
pMR-SADA	This work	**74.5**	**85.3**	**88.7**	**55.8**

rank-1, rank-5, and mAP metrics, respectively. Compared with the second best method CR-GAN+TAUDL, our model obtains 5.6%, 5.1%, 4.0% and 7.2% improvement on rank-1, rank-5, rank-10, and mAP metrics. The significant improvement demonstrates the superiority of the proposed method.

4.3. Ablation Studies and Further Analyses

In this section, we further study the importance of each main component/factor of our method by isolating this component/factor. Specifically, the main components of our proposed method includes a baseline, a Local Max Pooling (short for LMP) [6], domain attack (short for Attack), and a p-Memory approach. We extend the local max pooling to a Local Average Pooling (short for LAP), which is similar to LMP. we concatenate the output of LMP/LAP of each part as the final feature representation. The procedure is only used in the testing phase. In our experiment, we empirically found that LAP is the better on Market-1501 and LMP is better on DukeMTMC-reID, so we use these setting on all of our experiments. The experimental results of ablation studies are shown in **Table** 3.

Effectiveness of LMP/LAP. To show the effectiveness of LMP/LAP, we conduct an ablation study by comparing the first and second rows in **Table** 3. It is observed that with LMP/LAP the performance has 6.7%, 6.6%, 6.7%, and 9.2% improvement using four metrics (rank-1, rank-5, rank-10, and mAP) on Market-1501. The improvement on DukeMTMC-reID is 18.8%, 19.0%, 17.8%, and 14.0% with four metrics. The significant improvement could be attributed to the fact that LMP/LAP can reduce the impact of noisy signals by the differences of the source and target domains and thus provide a good initialization for the self-training, leading to significant improvement.

Effectiveness of Smooth Adversarial Domain Attack (SADA). To show the effect of the SADA, we set a baseline by removing this component from the proposed method, as shown in the third and fifth rows of **Table** 3. We can see that with the domain attack, the model has 2.7%, 1.5%, 1.0%, and 2.1% improvement with four metrics on Market-

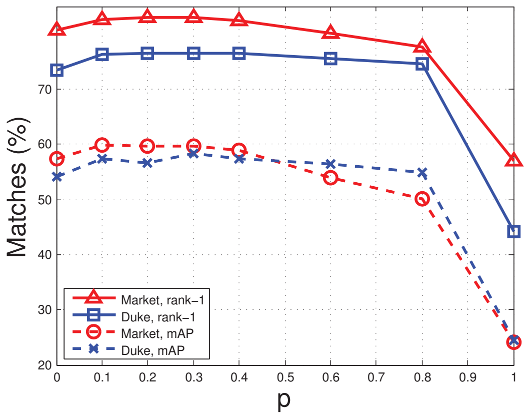

Figure 3. Effect of Parameter Memory Factor p.

1501, and 4.4%, 4.4%, 4.6%, and 9.1% with four metrics on DukeMTMC-reID. This ablation study demonstrates the effectiveness of the domain attack. It indicates that the SADA can guide the source images to align the target images at the image level.

Effectiveness of p-Memory Reconsolidation (pMR). To show the effectiveness of the pMR, we study the importance of this component by removing it from the proposed method. As shown in the fourth and fifth rows of **Table** 3, with the p-memory reconsolidation, the performance improves 2.3%, 1.7%, 1.4%, and 2.4% with four metrics on Market-1501, and 2.6%, 2.4%, 2.3%, and 3.5% with four metrics on DukeMTMC-reID. The experimental analyses show the effectiveness of the pMR. We conclude that pMR can reconsolidation the source transferred knowledge of traditional clustering-based cross-domain re-ID and improve the cross-domain person re-ID.

Effectiveness of both Domain Attack and p-Memory. We also study the good benefits of the combination of both domain attack and p-memory, which is shown in the second and the fifth rows. We observed that with both domain attack and p-memory, the performance improves largely, i.e., 4.9%, 3.6%, 2.5% and 4.3% with four metrics on Market-1501, 3.5%, 3.7%, 3.7%, and 5.3% with four metrics on DukeMTMC-reID.

Effect of Parameter Memory Reconsolidation Factor p. Parameter p is a probability that controls the random variable ξ in Eq. (5), which is used to control the probability of recalling the memory of the source domain. We vary p from 0.0 to 1.0, where $p = 0.0$ denotes there is no memory reconsolidation during the self-training of the target domain, and $p = 1.0$ denotes there is no self-training on the target domain. The rank-1 and mAP accuracies are shown in **Figure** 3. It is observed that when $p = 0.3$, our model obtains the best result.

Table 4. Effect of Smoothing Adversarial Domain Attack.

#Iterations	Duke↦Market	Market↦Duke
w/o smoothing	39.4	4.8
w/ smoothing	18.3	1.9

Effect of Smoothing Adversarial Domain Attack. We investigate the faster convergence of the smoothing adversarial domain attack by comparing the traditional iterative adversarial attack method. We adopt an evaluation metric that computes the average iterative times of the successful domain attack of an image. The experimental results are reported in **Table** 4. It is observed that when aligning the images from DukeMTMC-reID to Market-1501, the smoothing adversarial domain attack method only takes 18.3 attack iterations for each image while traditional method needs 39.4 iterations. When aligning the images from DukeMTMC-reID to Market-1501, our method takes 1.9 iterations while the traditional method takes 4.8 iterations.

Compare SADA with GAN based methods. Compared with GAN based methods, SADA is easy to train. SADA only needs to train a camera classifier (≈2 hours) and requires several domain attacks per image. GAN based methods are hard to converge (≈16 hours for CycleGAN). In **Table** 5, we remove our self-training for fair comparison. SADA achieves a competitive accuracy with two GANs [6].

Table 5. Compare SADA with GANs for Duke→Market (%)

	CycleGAN [6]	SPGAN+LMP [6]	SADA +LMP
rank-1/mAP	48.1/20.7	58.1/26.9	59.6/27.1
training time	≈16 hours	≈16 hours	≈2 hours

5. Conclusion

In this paper, we propose a smoothing adversarial domain attack approach to force the source images to align the target images at the image-level. With the domain attack, the gap between the source and target domains is narrowed at the image level, leading to better cross-domain transfer learning. To avoid the transfer amnesia problem, we also propose a p-memory reconsolidation approach that has a small probability p to reconsolidation the transferred source knowledge. Extensive experiments on Market-1501 and DukeMTMC-reID benchmarks show that our pMR-SADA outperforms all of the state-of-the-arts by a large margin.

Acknowledgments

This project was supported by the National Natural Science Foundation of China (U1611461, 61573387, 61672544) and the Key Field R & D Program of Guangdong Province (2019B010155003).

References

[1] R. Aljundi, P. Chakravarty, and T. Tuytelaars. Expert gate: Lifelong learning with a network of experts. In *CVPR*, pages 3366–3375, 2017. 3

[2] J. Baxter. A model of inductive bias learning. *Journal of artificial intelligence research*, 12:149–198, 2000. 2

[3] W. Chen, X. Chen, J. Zhang, and K. Huang. Beyond triplet loss: a deep quadruplet network for person re-identification. In *CVPR*, pages 403–412, 2017. 2

[4] Y. Chen, X. Zhu, and S. Gong. Instance-guided context rendering for cross-domain person re-identification. In *ICCV*, pages 232–242, 2019. 1, 2, 6, 7

[5] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *CVPR*, pages 1335–1344, 2016. 2

[6] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *CVPR*, pages 994–1003, 2018. 1, 2, 7, 8

[7] S. Ding, L. Lin, G. R. Wang, and H. Chao. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognition*, 48(10):2993–3003, 2015. 2

[8] H. Fan, L. Zheng, C. Yan, and Y. Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):83, 2018. 1, 2, 6, 7

[9] P. Fang, J. Zhou, S. K. Roy, L. Petersson, and M. Harandi. Bilinear attention networks for person retrieval. In *ICCV*, pages 8030–8039, 2019. 2

[10] Y. Fu, Y. Wei, Y. Zhou, H. Shi, G. Huang, X. Wang, Z. Yao, and T. Huang. Horizontal pyramid matching for person re-identification. In *AAAI*, volume 33, pages 8295–8302, 2019. 2

[11] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2

[12] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 4

[13] H. Jung, J. Ju, M. Jung, and J. Kim. Less-forgetting learning in deep neural networks. *arXiv preprint arXiv:1607.00122*, 2016. 3

[14] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 3

[15] E. Kodirov, T. Xiang, and S. Gong. Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. In *BMVC*, volume 3, page 8, 2015. 6, 7

[16] A. Kurakin, I. Goodfellow, and S. Bengio. Adversarial examples in the physical world. *arXiv preprint arXiv:1607.02533*, 2016. 3

[17] W. Li, X. Zhu, and S. Gong. Harmonious attention network for person re-identification. In *CVPR*, pages 2285–2294, 2018. 2

[18] Y. J. Li, C. C. Lin, Y. B. Lin, and Y. F. Wang. Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation. In *ICCV*, pages 7919–7929, 2019. 1, 2, 6, 7

[19] W. Liang, G. Wang, J. Lai, and J. Zhu. M2m-gan: Many-to-many generative adversarial transfer learning for person re-identification. *arXiv preprint arXiv:1811.03768*, 2018. 1, 2

[20] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, pages 2197–2206, 2015. 6, 7

[21] L. Lin, G. R. Wang, W. Zuo, X. Feng, and L. Zhang. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE TPAMI*, 39(6):1089–1102, 2016. 2

[22] S. Lin, H. Li, C. T. Li, and A. C. Kot. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. *arXiv preprint arXiv:1807.01440*, 2018. 2

[23] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI*, volume 33, pages 8738–8745, 2019. 2

[24] G. Lisanti, L. Masi, A. D. Bagdanov, and A. Del Bimbo. Person re-identification by iterative re-weighted sparse ranking. *IEEE TPAMI*, 37(8):1629–1642, 2014. 6, 7

[25] J. Liu, Z. J. Zha, D. Chen, R. Hong, and M. Wang. Adaptive transfer network for cross-domain person re-identification. In *CVPR*, pages 7202–7211, 2019. 2, 6, 7

[26] D. Lopez-Paz and M. Ranzato. Gradient episodic memory for continual learning. In *NeurIPS*, pages 6467–6476, 2017. 2, 3

[27] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian. Unsupervised cross-dataset transfer learning for person re-identification. In *CVPR*, pages 1306–1315, 2016. 6, 7

[28] L. Qi, L. Wang, J. Huo, L. Zhou, Y. Shi, and Y. Gao. A novel unsupervised camera-aware domain adaptation framework for person re-identification. In *ICCV*, pages 8080–8089, 2019. 2, 6, 7

[29] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016. 3

[30] P. Ruvolo and E. Eaton. Ella: An efficient lifelong learning algorithm. In *ICML*, pages 507–515, 2013. 2

[31] L. Song, C. Wang, L. Zhang, B. Du, Q. Zhang, C. Huang, and X. Wang. Unsupervised domain adaptive re-identification: Theory and practice. *arXiv preprint arXiv:1807.11334*, 2018. 2

[32] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose-driven deep convolutional model for person re-identification. In *ICCV*, pages 3960–3969, 2017. 2

[33] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang. Beyond part models: Person retrieval with refined part pooling (and

a strong convolutional baseline). In *ECCV*, pages 480–496, 2018. 2

[34] C. Wang, Q. Zhang, C. Huang, W. Liu, and X. Wang. Mancs: A multi-task attentional network with curriculum sampling for person re-identification. In *ECCV*, pages 365–381, 2018. 2

[35] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou. Learning discriminative features with multiple granularities for person re-identification. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 274–282. ACM, 2018. 2

[36] G. C. Wang, J. Lai, P. Huang, and X. Xie. Spatial-temporal person re-identification. In *AAAI*, pages 8933–8940, 2019. 2

[37] G. C. Wang, J. Lai, and X. Xie. P2snet : Can an image match a video for person re-identification in an end-to-end way? *IEEE TCSVT*, 28:2777–2787, 2018. 2

[38] G. C. Wang, J. Lai, Z. Xie, and X. Xie. Discovering underlying person structure pattern with relative local distance for person re-identification. *arXiv preprint arXiv:1901.10100*, 2019. 2

[39] G. R. Wang, L. Lin, S. Ding, Y. Li, and Q. Wang. Dari: Distance metric and representation integration for person verification. In *AAAI*, 2016. 2

[40] G. R. Wang, P. Luo, X. Wang, L. Lin, and etc. Kalman normalization: Normalizing internal representations across network layers. In *NeurIPS*, pages 21–31, 2018. 2

[41] G. R. Wang, G. C. Wang, X. Zhang, J. Lai, and L. Lin. Weakly supervised person re-identification: Cost-effective learning with a new benchmark. *arXiv preprint arXiv:1904.03845*, 2019. 2

[42] G. R. Wang, K. Wang, and L. Lin. Adaptively connected neural networks. In *CVPR*, pages 1781–1790, 2019. 2

[43] J. Wang, X. Zhu, S. Gong, and W. Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *CVPR*, pages 2275–2284, 2018. 6, 7

[44] L. Wei, S. Zhang, W. Gao, and Q. Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, pages 79–88, 2018. 1, 2

[45] A. Wu, W. S. Zheng, and J. H. Lai. Unsupervised person re-identification by camera-aware similarity consistency learning. In *ICCV*, pages 6922–6931, 2019. 6, 7

[46] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Deep metric learning for person re-identification. In *ICPR*, pages 34–39. IEEE, 2014. 2

[47] H. X. Yu, A. Wu, and W. S. Zheng. Cross-view asymmetric metric learning for unsupervised person re-identification. In *ICCV*, pages 994–1002, 2017. 6, 7

[48] H. X. Yu, W. S. Zheng, A. Wu, X. Guo, S. Gong, and J. H. Lai. Unsupervised person re-identification by soft multilabel learning. In *CVPR*, pages 2148–2157, 2019. 6, 7

[49] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, pages 1077–1085, 2017. 2

[50] L. Zhao, X. Li, Y. Zhuang, and J. Wang. Deeply-learned part-aligned representations for person re-identification. In *ICCV*, pages 3219–3228, 2017. 2

[51] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian. Mars: A video benchmark for large-scale person re-identification. In *ECCV*, pages 868–884. Springer, 2016. 2

[52] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124, 2015. 1, 6, 7

[53] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *ICCV*, 2017. 6

[54] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *CVPR*, pages 598–607, 2019. 6, 7

[55] S. Zhou, F. Wang, Z. Huang, and J. Wang. Discriminative feature learning with consistent attention regularization for person re-identification. In *ICCV*, pages 8040–8049, 2019. 2

[56] J. Zhuo, Z. Chen, J. Lai, and G. C. Wang. Occluded person re-identification. *arXiv preprint arXiv:1804.02792*, 2018. 2