Meta Distribution Alignment for Generalizable Person Re-Identification

Hao Ni[1] Jingkuan Song[1]* Xiaopeng Luo[1] Feng Zheng[2] Wen Li[1] Heng Tao Shen[1]

[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China
[2]Southern University of Science and Technology

haoni0812@gmail.com, jingkuan.song@gmail.com, xpluo1949@gmail.com

Abstract

Domain Generalizable (DG) person ReID is a challenging task which trains a model on source domains yet generalizes well on target domains. Existing methods use source domains to learn domain-invariant features, and assume those features are also irrelevant with target domains. However, they do not consider the target domain information which is unavailable in the training phrase of DG. To address this issue, we propose a novel Meta Distribution Alignment (MDA) method to enable them to share similar distribution in a test-time-training fashion. Specifically, since high-dimensional features are difficult to constrain with a known simple distribution, we first introduce an intermediate latent space constrained to a known prior distribution. The source domain data is mapped to this latent space and then reconstructed back. A meta-learning strategy is introduced to facilitate generalization and support fast adaption. To reduce their discrepancy, we further propose a test-time adaptive updating strategy based on the latent space which efficiently adapts model to unseen domains with a few samples. Extensive experimental results show that our model outperforms the state-of-the-art methods by up to 5.2% R-1 on average on the large-scale and 4.7% R-1 on the single-source domain generalization ReID benchmark. Source code is publicly available at https://github.com/haoni0812/MDA.git.

1. Introduction

Person Re-identification (ReID) aims to match persons with the same ID across different camera views. Thanks to the development of deep convolutional neural networks (CNNs) [14], supervised ReID and unsupervised domain adaptation (UDA) [42] have achieved remarkable performance. However, they both need data of target domain for training. In real-world applications, the ReID system will inevitably search persons in unseen domains. Therefore, domain generalization (DG) ReID has attracted extensive research attention in a practical setting.

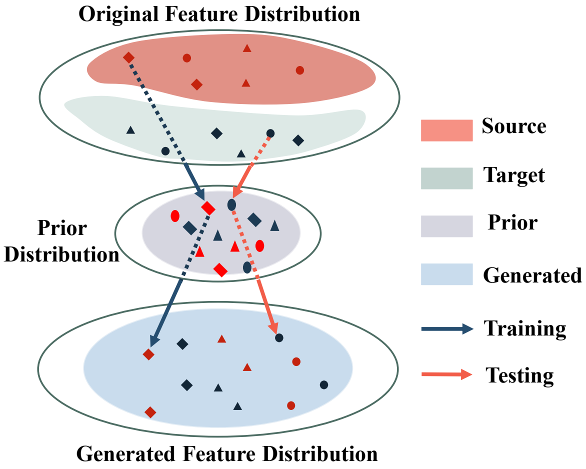


Figure 1. Illustration of our idea. Since target domain data (green) is unavailable during training, we cannot directly align source and target distributions. To address this issue, we align them to a prior distribution (purple) during training (source domains) and testing (target domains). Considering that high-dimensional ID features are difficult to constrain to a prior distribution, ID features are encoded into latent embedding space. The same prior distribution and decoder guarantee the same generated feature distribution.

Compared with supervised ReID and UDA setting, DG ReID does not use target domain data for training. Only one or more labeled source datasets are available. Thus, most existing DG methods aim to learn domain invariant feature through multiple source domains to generalize unseen domains. These methods explore generalization at feature-level based on disentanglement [13] or meta-learning [2, 3, 43]. However, a typical and effective cross-domain solution has been ignored in DG, that is, to align feature distributions across source and target domains. Methods based on distribution alignment have not been researched because only source data is available for DG ReID. These method usually

[*]Corresponding author.

need both source and target domain data for training. So we cannot align distributions based on the previous method.

Such observations reveal that it is challenging to align distributions directly. So we take a known prior distribution as the aligned goal, and ask: can we align source and target distributions to the same known prior distribution during training and testing, respectively? However, constraining the distribution of ID features to a known prior distribution is difficult. Because ID feature is high-dimensional and contains ID information, its distribution is so complex that a known simple prior distribution cannot constrain it.

Thus, we adopt an encoding-decoding structure to encode ID feature into latent space, whose latent embedding is low-dimensional and constrained to a known prior distribution. Then we decode latent embedding through a decoder. If the latent embedding distribution can also be close to the prior distribution during testing, we can treat latent embedding distributions during training and testing as the same. Note that the premise of the above conclusion is that the distance metric satisfies the triangle inequality, so we use Wasserstein distance instead of KL divergence to measure the distance between distributions. Eventually, similar posterior distributions and the same decoder enable us to obtain aligned distributions. The main idea of our method is shown in Figure 1.

To further enhance the model generalization on unseen domains, we introduce a meta-learning strategy to simulate the real train-test process. Specifically, we dynamically divide the source domain into a meta-train domain and a meta-test domain in each batch. The meta-train process is regarded as the training process of the source domain, and the meta-test simulates the situation of testing on the unseen domain. In the meta-train stage, we use the meta-train domain to inner-update parameters. In the meta-test stage, we examine various generalization scenarios depending on the movement of inner loop, and perform a second-order updating on the original parameters. These two processes are performed alternately to improve generalization ability of the model on various unseen domains. During testing, we can further pull in latent embedding distributions across source and target domains by a test-train strategy at batch-level. We will only update the encoder in this process, to ensure the same decoder for source and target domains.

In summary, our contributions are three-fold:

- We propose a novel Meta Distribution Alignment (MDA) for DG ReID, which is a pioneering work on aligning distributions across source and target domains for DG ReID task.

- We design a meta-learning strategy to simulate the real train-test process, which improves the generalization of the model. A test-time adaptive updating strategy is further proposed to efficiently adapt the model to unseen domains with a few samples.

- We perform extensive experiments and achieve state-of-the-art performance on the large-scale DG ReID and single-source DG ReID.

2. Related Work

Person Re-identification. In the last decade, person ReID has achieved great progress. Among them, supervised person ReID has achieved impressive performance. Relying on labeled data, supervised ReID can performs supervised training and testing in the same domain. But current supervised models degrade dramatically when deployed to unseen domains. The main reason is the gap across source and target domain. To improve the performance on target domain, many Unsupervised Domain Adaptation (UDA) Person ReID methods have been proposed. They can be roughly grouped into four categories: a) clustering in the target domain to generate pseudo-labels [6, 8, 20, 31, 36, 40], b) self-supervised training on target domain [37, 49], c)generating images with target domain style and source domain labels for data augmentation [5,33,48], d) aligning feature distributions across target and source domains [19, 32, 34]. In fact, the latter two can also be classified as distribution alignment methods, which perform at input level and feature level, respectively. However, the methods based on distribution alignment in UDA all need to use the unlabeled target dataset for training, which is not available in DG ReID. Therefore, all previous distribution alignment methods will fail in DG ReID.

Domain Generalization Person Re-identification. Although supervised and UDA person ReID have achieved good performance, they all require target domain data for training. However, in practical applications, the ReID system often needs to be deployed directly without training. DG ReID was first proposed under this background [38]. After that, [27] applied meta-learning to learn domain-invariant feature. [13] proposed SNR disentangle identity-irrelevant information. [24] proposed IBN-net explored the effect of combining instance and batch normalization. And [12] combined IBN and meta leaning to further improve performance. More recently, meta-leaning has also been studied more deeply and played an important role in DG ReID. The core idea of these methods is to learn domain-invariant features, while ignoring the gap between the source and target domain feature distributions.

Meta Learning. The concept of meta-learning [29] is learning to learn, and has been initially proposed in the machine learning community. It has been applied to tasks such as few-shot learning [41], domain generalization and model optimization. Model-Agnostic Meta-learning (MAML) [7] was proposed to learn good initialization parameters for fast adapting the model to new tasks. [15] extends MAML to domain generalization. Inspired by those methods, we design a meta-learning strategy to simulate real train-test domain shifts, which makes the model more generalized and can quickly adapt the model to unseen domain.

3. Methodology

Our goal is to align distributions across source and target domains. To this end, we propose a novel Meta Distribution Alignment algorithm. In the following, we describe the main components of our method. First, we introduce the DG problem setting and overview of the entire method (Sec. 3.1). Next, we describe the Meta Distribution Alignment algorithm in detail, which includes the following parts: ID Feature Learning module (Sec. 3.2.1), Prior Distribution Alignment module(Sec. 3.2.2) and Distribution Guided Refining (Sec. 3.2.3). To facilitate generalization and support fast adaption, we design a Meta-learning Based Optimization strategy to simulate train-test process (Sec. 3.2.4). Finally, we discuss how we fine-tune our model by Test-time Adaptive Updating (Sec. 3.3). Figure 2 shows an overview of our method.

3.1. Problem Setting and Overview

Problem setting For DG ReID, we use one or several labeled datasets for training. Each source dataset $D_S = \{(X_s^i, y_s^i)\}_{i=1}^{N_s}$ is composed of a training set and a testing set. After training, we directly deploy the model into unseen target domains. Each target dataset $D_T = \{X_t^j\}_{j=1}^{N_t}$ only contains one testing set. In other words, we cannot access unseen target training samples during training. In the training stage, the goal of ReID is to learn a mapping function: $f_\theta : X \to x$, which maps X to a feature space with parameters θ, so that features meet the following condition:

$$\begin{aligned} &\forall y^i = y^j \neq y_k \quad i, j, k \in (1, 2, \cdot, N_s) \\ &\text{s.t.} \qquad d(x^i, x^j) < d(x^i, x^k) \end{aligned} \tag{1}$$

where d is a distance metric.

Overview In supervised ReID, a training set and a testing set usually come from a same domain, so they basically share the same distribution. As a result, a distance between testing set features is not susceptible to interference from different environments. It can still meet requirements of Condition, defined in Eq. 1. However, when a model is deployed to a new domain, where the source domain distribution $p(\boldsymbol{x_s})$ and the target domain distribution $p(\boldsymbol{x_t})$ are quite different, the distance between the features is affected by the scene. In this case, Condition 1 can no longer be met.

In order to solve this problem, UDA uses target training set to align $p(\boldsymbol{x_s})$ and $p(\boldsymbol{x_t})$. However, target domain data is unavailable for DG ReID during training. Therefore, we propose meta distribution alignment (MDA) to align $p(\boldsymbol{x_s})$ and $p(\boldsymbol{x_t})$ to a prior known distribution during training and testing.

3.2. Meta Distribution Alignment

3.2.1 ID Feature Learning

In this stage, we randomly sample from different source domains to get a mini-batch X_b of size N_b. We take the entire mini-batch data as input and use two common loss functions to constrain the distance between these features so that they meet the Condition 1 as much as possible. The first one is the cross entropy loss function:

$$\mathcal{L}_{ce}(X_b|\theta) = -\frac{1}{N_b}\sum_{i=1}^{N_b} \log p\left(y_i | f_\theta(X_i)\right) \tag{2}$$

Compared with CE loss, a triplet loss directly constrains the distance between positive and negative sample pairs. The specific formula is as follows:

$$\mathcal{L}_{tri}(X_b|\theta) = \frac{1}{N_b}\sum_{i=1}^{N_b} [d_p - d_n + \alpha]_+ \tag{3}$$

where d_p and d_n are Euclidean distance of positive and negative feature pairs, respectively. α is the margin of triplet loss, $[s]_+$ is $max(s, 0)$. The overall loss of ID feature learning is formulated as follows:

$$\mathcal{L}_{ID}(X_b|\theta) = \mathcal{L}_{ce}(X_b|\theta) + \mathcal{L}_{tri}(X_b|\theta) \tag{4}$$

3.2.2 Prior Distribution Alignment

Distribution encoding As we mentioned in Sec. 1, ID features are high-dimensional and contain discriminative ID information. Its distribution is difficult to constrain with a known simple distribution, such as standard Gaussian distribution. Thus, given an ID feature $\boldsymbol{x_s}$ based on current f_θ and X_s, our encoder E_ϕ parameterizes a multivariate Gaussian distribution with a diagonal covariance as follows:

$$\begin{gathered} \boldsymbol{\mu}_n, \boldsymbol{\sigma}_n = E_\phi(\boldsymbol{x_s}) \\ \boldsymbol{z} \sim q(\boldsymbol{z}|\boldsymbol{x_s}) = \mathcal{N}\left(\boldsymbol{\mu}_n, diag(\boldsymbol{\sigma}_n)\right) \end{gathered} \tag{5}$$

where n is the dimension of latent embedding z. μ_n and σ_n are mean and variance of Gaussian distribution, respectively.

In general, KL divergence is used to measure the similarity of distributions. However, it does not meet triangle inequality. That is, the sum of $KL[p(\boldsymbol{z}|\boldsymbol{x_s})||q(\boldsymbol{z})]$ and $KL[p(\boldsymbol{z}|\boldsymbol{x_t})||q(\boldsymbol{z})]$ is irrelevant to $KL[p(\boldsymbol{z}|\boldsymbol{x_s})||p(\boldsymbol{z}|\boldsymbol{x_t})]$.

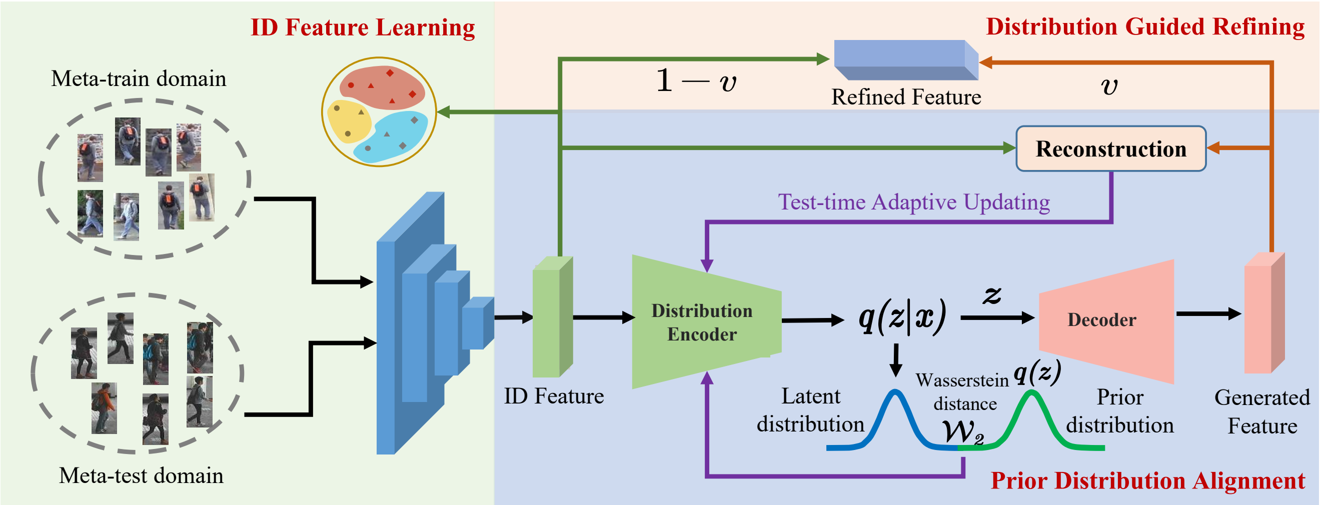


Figure 2. Illustration of our proposed Meta Distribution Alignment. It includes three modules and a meta-learning strategy. In the ID feature learning, $\mathcal{L}_{ID}$ is used to enhanced the ID-discrimination ability. In prior distribution alignment, ID features are encoded into latent space. The distribution encoder E_ϕ parameterizes a multivariate Gaussian distribution for each sample. We align this distribution to a standard Gaussian distribution by minimizing Wasserstein distance. Then we sample from the distribution and decode latent embedding to generated ID feature space through distribution guided decoder. The reconstruction loss $\mathcal{L}_{rec}$ is used to preserve ID label. In distribution guided refining, we use refine net R_φ to retain the ID information as much as possible while constraining the distribution. Combining the $\mathcal{W}_2$, $\mathcal{L}_{rec}$ and $\mathcal{L}_{ref}$, we perform our meta-learning based optimization by separating multiple source domains. During testing, we can fast adapt model to unseen domains with a few samples by test-time adaptive updating (purple arrow).

Even if $p(\boldsymbol{z}|\boldsymbol{x_s})$ and $p(\boldsymbol{z}|\boldsymbol{x_t})$ are close to $q(\boldsymbol{z})$ under the measure of KL divergence, it cannot guarantee $p(\boldsymbol{z}|\boldsymbol{x_s})$ and $p(\boldsymbol{z}|\boldsymbol{x_t})$ are close. Thus, we chose Wasserstein distance to measure the similarity, which can meet triangle inequality:

$$\mathcal{W}(P;Q) <= \mathcal{W}(P;O) + \mathcal{W}(Q;O) \tag{6}$$

where O, P and Q are three arbitrary distributions. The second order Wasserstein distance between the embedding distribution $p(\boldsymbol{z}|\boldsymbol{x_s})$ and the standard Gaussian distribution $q(\boldsymbol{z}) = \mathcal{N}\left(\boldsymbol{\mu}_n^0, diag(\boldsymbol{\sigma}_n^0)\right)$ can be formulated as:

$$\begin{aligned}\mathcal{W}_2(p(\boldsymbol{z}|\boldsymbol{x_s}); q(\boldsymbol{z})) = (\|\boldsymbol{\mu}_n - \boldsymbol{\mu}_n^0\|_2^2 \\ + \|diag((\boldsymbol{\sigma}_n)^{\frac{1}{2}} - (\boldsymbol{\sigma}_n^0)^{\frac{1}{2}})\|_F^2)^{\frac{1}{2}}\end{aligned} \tag{7}$$

where $\|\cdot\|_2$ is the second norm and $\|\cdot\|_F$ is *Frobenius* norm.

Distribution based decoding Although the embedding is constrained to a known prior distribution, but it tends to lose the discriminative information about ID due to significant feature dimension decreasing. To preserve the ID information as much as possible, we decode the latent embedding $\boldsymbol{z}$ to a high-dimensional ID space. Given $\boldsymbol{z} \sim p(\boldsymbol{z}|\boldsymbol{x_s})$, the decoder $G_\psi : \boldsymbol{z} \to x'$ is used to map z back to the ID feature space. The reconstruction loss is introduced to ensure that the generated features can maintain the original ID, which can be formulated as:

$$\mathcal{L}_{rec}(\boldsymbol{x_s}|\phi, \psi) = \|\boldsymbol{x_s} - G_\psi(\boldsymbol{z}))\|_2 \tag{8}$$

3.2.3 Distribution Guided Refining

Although we use reconstruction loss to preserve ID information, it is still inevitably lost during the encoding-decoding process. The loss of ID information can be considered as the price we paid to constrain the distribution. To make up for this loss, we introduce a refine network R_φ to integrate the original ID features $\boldsymbol{x_s}$ to complement the generated features $G_\psi(\boldsymbol{z})$). The Refined feature and its corresponding ID loss function can be expressed as:

$$\begin{aligned}\boldsymbol{x_r}{=}(1-\boldsymbol{v}) \times \boldsymbol{x_s} + \boldsymbol{v} \times G_\psi(\boldsymbol{z}) \\ \mathcal{L}_{ref}(\boldsymbol{x_s}|\phi, \psi, \varphi){=}\mathcal{L}_{ce}(\boldsymbol{x_r}|\phi, \psi, \varphi){+}\mathcal{L}_{tri}(\boldsymbol{x_r}|\phi, \psi, \varphi)\end{aligned} \tag{9}$$

where $v = R(x_s - G_\psi(\boldsymbol{z}))$. Intuitively, the refine network learns a set of weights, allowing the model to make a trade-off between ID discrimination and distribution constraints. In this way, we are capable of aligning a feature to a specific distribution while preserving its ID information as much as possible.

3.2.4 Meta-learning Based Optimization

Motivated by the success of meta-learning for cross-domain image classification [15] in terms of generalization and fast adaption, in this paper we propose a meta-learning strategy for DG ReID to update our distribution encoder E_ϕ, distribution based decoder G_ψ and refined net R_φ during training. The whole learning strategy includes meta-train and meta-test.

Domain separation. To simulate the train-test process during training, we divide multiple source domains in Sec. 3.2.1 into meta-train domain D^{mtr} and meta-test domain D^{mte}. In the meta-train stage, we use D^{mtr} as input to calculate the meta-train loss, and then inner-update E_ϕ, G_ψ and R_φ. In the meta-test stage, we use the updated parameters and D^{mte} to calculate the meta-test loss. For final optimization, we use the meta-train loss to update the entire encoder-decoder and use the meta-test loss to update the encoder.

Meta-train. The overall loss for meta-train is as follows:

$$\begin{aligned}\mathcal{L}_{mt}(D^{mtr}|\phi,\psi,\varphi) &= \mathcal{W}_2^2(p(\boldsymbol{z}|\boldsymbol{x_s});q(\boldsymbol{z})) \\ &+ \mathcal{L}_{rec}(\boldsymbol{x_s}|\phi,\psi) + \mathcal{L}_{ref}(\boldsymbol{x_r}|\phi,\psi,\varphi)\end{aligned} \tag{10}$$

where x_s is sampled from D^{mtr}. In inner loop, we update the parameters of meta distribution align from (ϕ,ψ,φ) to (ϕ',ψ',φ') as follows:

$$(\phi',\psi',\varphi')=(\phi,\psi,\varphi)-\alpha\bigtriangledown_{\phi,\psi,\varphi}\mathcal{L}_{mtr}(D^{mtr}|\phi,\psi,\varphi) \tag{11}$$

where α is a learning rate for inner loop optimization.

Meta-test In order to simulate an unseen domain, we evaluate our model at unseen-like samples from meta-test domain D^{mte}. With the meta-test domain, we can examine various generalization scenarios depending on the movement of the meta distribution align parameters. In the outer loop, the same loss as the meta-train is used to update the parameters, which can be formulated as:

$$(\phi,\psi,\varphi)=(\phi,\psi,\varphi)-\beta\bigtriangledown_{\phi,\psi,\varphi}\mathcal{L}_{mt}(D^{mte}|\phi',\psi',\varphi') \tag{12}$$

where β is a learning rate for outer loop optimization. Intuitively, meta distribution align simulates such a process: In meta-train, model is trained on the source domain. In meta-test, model is validated and fine-tuned on unseen domains. These two processes are performed alternately. Finally, the model can obtain better generalization ability in various unseen domains. The better initialization parameters also give the model the ability to quickly transfer, which provides convenience for our adaptive fine-tuning in the test.

3.3. Test-time Adaptive Updating

Through prior distribution alignment, we can get good initialization parameters. However, when transferring to unseen domains, the distribution of latent embedding may change slightly. Therefore, we propose a test-time adaptive updating to further constrain the distribution on unseen domain. Specifically, we fine-tune the encoder E_ϕ according to the current batch-test samples to make latent embedding distribution closer to $q(\boldsymbol{z})$. Since test samples are unlabeled, our purpose is only to align the distribution of hidden variables. Therefore, we use the Wasserstein distance to constrain the distribution, and reconstruction loss prevents the generated feature $G_\psi(\boldsymbol{z})$ from being too far away from the original feature and completely losing ID discrimination. The update process is as follows:

$$\phi=\phi-\beta\bigtriangledown_\phi\left(\left(\mathcal{L}_{rec}(x_t|\phi)+\mathcal{W}_2^2\left(p(\boldsymbol{z}|\boldsymbol{x_t});q(\boldsymbol{z})\right)\right) \tag{13}$$

where x_t is a test sample. Note that only distribution encoder is fine-tuned during test, decoder and refine net will not be changed. Because only the same input distribution and decoder can be guaranteed to the same $G_\psi(\boldsymbol{z})$ distribution during training and testing. After fine-tuning the decoder, we use the refined feature for retrieval.

4. Experiment

4.1. Datasets

To evaluate the generalization of the model, we conduct extensive experiments on large-scale domain generalization (DG) ReID benchmark [27] and single-source DG problem. For Large-scale DG ReID, we use CUHK02 [16], CUHK03 [17], Market-1501 [44], DukeMTMC-ReID [46], and CUHK-SYSU PersonSearch [35] for training, and evaluate on VIPeR [9], GRID [22], and QMUL i-LIDS [45]. For single-source DG ReID, we choose one of DukeMTMC-ReID, Market-1501, MSMT17 [33] as the training set and directly test on the remaining datasets. In a single-source DG ReID, only one source domain is available, so we divide the data of different cameras as different domains for domain separation in meta-learning. In particular, for MSMT17 we use the entire test set and training set for training [18]. For simplicity, we denote Market-1501, DukeMTMC-reID, and MSMT17 as Market, Duke and MSMT tables.

4.2. Implementation details

We implement our method with two common backbones, i.e., ResNet-50 [11] and MobileNetV2 [26]. The models are pre-trained on ImageNet [4]. For training, images are resized to 256 × 128. The training batch size is set to 160. In ID feature learning, all images in a mini-batch are used. Half of them are meta-train domains, and the other half are meta-test domains. Both the encoder and decoder are composed of three fully connected layers, and the dimension of the latent embedding is set to 64. The refine net consists of a fully connected layer and a sigmoid activation function. The label-smoothing parameter is 0.1, and the margin in the triplet loss is 0.3. For backbone, we use the SGD optimizer with a momentum of 0.9 and a weight decay of 5×10^{-4}.

Table 1. Quantitative comparisons of large-scale domain generalization ReID. CUHK02, CUHK03, Market-1501, DukeMTMC-ReID, and CUHK-SYSU PersonSearch are used for training. All results are the average of 10 random sampling. Our results are highlighted in bold and others' best results are underlined. Results with '*' is based on ResNet-50, otherwise is MobileNetV2.

Method	Large-scale domain generalization ReID (multi-source DG)											
	Target: VIPeR				Target: GRID				Target: i-LIDS			
	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
DIMN [27]	51.2	70.2	76.0	60.1	29.3	53.3	65.8	41.1	70.2	89.7	94.5	78.4
AugMining [28]	49.8	70.8	77.0	-	46.6	67.5	76.1	-	76.3	93.0	95.3	-
Switchable (BN+IN) [23]	51.6	72.9	80.8	61.4	39.3	58.8	68.1	48.1	77.3	91.2	94.8	83.5
DualNorm [12]	53.9	62.5	75.3	58.0	41.4	47.4	64.7	45.7	74.8	82.0	91.5	78.5
DDAN [1]	52.3	60.6	71.8	56.4	50.6	62.1	73.8	55.7	78.5	85.3	92.5	81.5
DDAN w/ [12]	56.5	65.6	76.3	60.8	46.2	55.4	68.0	50.9	78.0	85.7	93.2	81.2
MetaBIN [2]	56.9	76.7	82.0	66.0	49.7	67.5	76.8	58.1	79.7	93.3	97.3	85.5
MDA(Ours)	**61.9**	**80.73**	**85.5**	**70.4**	**53.8**	**76.1**	**83.3**	**63.8**	**81.0**	**92.8**	**95.5**	**86.1**
SNR* [13]	52.9	-	-	61.3	40.2	-	-	47.7	84.1	-	-	89.9
DualNorm* [12]	59.4	-	-	-	43.7	-	-	-	78.2	-	-	-
MetaBIN* [2]	59.9	78.4	82.8	68.6	48.4	70.3	77.2	57.9	81.3	95.0	97.0	87.0
MDA*(Ours)	**63.5**	**80.6**	**84.2**	**71.7**	**61.2**	**83.4**	**88.9**	**62.9**	**80.4**	**92.2**	**95.0**	**84.4**

The initial learning rate of backbone and encoder-decoder is 0.01, which is warmed up for 10 epochs [10].

4.3. Comparison with State-of-the-art Methods

Large-scale DG ReID We evaluate our meta distribution alignment framework on the large-scale domain generalization ReID benchmark [27]. For a fair comparison, we conduct our experiments on both MobileNetV2 and ResNet-50. We follow the single-shot setting [12] with the number of query/gallery images set as: VIPeR: 316/316; GRID: 125/900; i-LIDS: 60/60 respectively. Following [12], all results are the average of 10 random splits on the target dataset.

The experimental results are shown in Table 1. From it, we can observe that with MobileNetV2 as a backbone, our method outperforms all competing methods by a significant margin on all three datasets across all evaluation metrics.For example, MDA beats state-of-the-art by 5.0% and 4.4% in R-1 and mAP, respectively. In general, the R-1 and mAP of MDA are 3.5% and 3.6% higher than state-of-the-art on average. Even with ResNet-50 as a backbone, our method still achieves the best results on both VIPeR and GRID. Especially on GRID, our method outperforms MetaBIN by 12.8% in R-1. By contrast, SNR with ResNet-50 obtains the best R-1 and mAP on i-LIDS, but it performs worst in both VIPeR and GRID. Our method is on average higher than state-of-the-art 5.2% and 1.8% in R-1 and mAP. This demonstrates the effectiveness of our proposed methods. In addition, our method obtains higher gains on GRID with 8 cameras than VIPeR and i-LIDS with two cameras. This demonstrates that our MDA has a good potential in dealing with complex distributions.

Single-source DG ReID In order to further validate the performance of our method, we also conduct the experiments on a single source dataset. Specifically, we use Market, Duke, and MSMT as the training sets, and use Market and Duke as the test sets. Moreover, we divide the cross-domain experimental settings into UDA, fully unsupervised and DG. In addition, we also respectively enumerate three experimental settings and compare their differences.

The experimental results are shown in Table 2. When training with Market or Duke, our method outperforms the state-of-the-art DG ReID method. In particular, our method's mAP is 2.1% higher than the current best DG ReID method (MetaBIN) when testing on Market. It shows our MDA is effective with different training datasets.

It is worth noting that our method trained with MSMT achieves comparable or even better results than unsupervised method, reaching 52.4% mAP and 71.7% R-1 on Duke. Our MDA beats state-of-the-art DG ReID method by 4.7% and 4.8% on average in R-1 and mAP, respectively. Also compared with training on Market, the performance of our method trained on MSMT is greatly improved, with an increase of 5.0% in mAP and 8.0% in Rank-1 on Duke. This demonstrates MDA can obtain better generalization and exceed the performance of fully supervised method as the quality of the source training set improves.

4.4. Ablation Study

Ablation study of main components of our method. To further demonstrate the effectiveness of the proposed

Table 2. Quantitative comparisons between ours and the state-of-the-arts in single-source DG ReID. Our results are highlighted in bold and others' best results are underlined.

Methods	Reference	Setting	Training		Test:Duke		Training		Test:Market	
			Source	Target	R-1	mAP	Source	Target	R-1	mAP
TJ-AIDL [32]	CVPR18	UDA	Market	Duke	44.3	23.0	Duke	Market	58.2	26.5
PAUL [37]	CVPR19		Market	Duke	56.1	35.7	Duke	Market	66.7	36.8
ECN [47]	CVPR19		Market	Duke	63.3	40.4	Duke	Market	75.1	40.0
ECN baseline [47]	CVPR19	DG	Market	-	28.9	14.8	Duke	-	43.1	17.7
PN-GAN [25]	ECCV18		Market	-	29.9	15.8	-	-	-	-
$QAConv_{50}$ [18]	ECCV20		Market	-	48.8	28.7	Duke	-	58.6	27.2
SNR [13]	CVPR20		Market	-	55.1	33.6	Duke	-	66.7	33.9
MetaBIN [2]	CVPR21		Market	-	55.2	33.1	Duke	-	69.2	35.9
MDA	This paper		Market	-	**56.7**	**34.4**	Duke	-	**70.3**	**38.0**
MAR [39]	CVPR19	TJ-AIDL	-	Duke	67.1	48.0	-	Market	67.7	40.0
SSL [21]	CVPR20		-	Duke	52.5	28.6	-	Market	71.7	37.8
MMCL [31]	CVPR20		-	Duke	65.2	40.2	-	Market	80.3	45.5
MAR baseline [39]	CVPR19	DG	MSMT	-	43.1	28.8	MSMT	-	46.2	24.6
PAUL baseline [37]	CVPR19		MSMT	-	65.7	45.6	MSMT	-	59.3	31.0
$QAConv_{50}$ [18]	ECCV20		MSMT	-	69.4	52.6	MSMT	-	72.6	43.1
SNR [13]	CVPR20		MSMT	-	69.2	49.9	MSMT	-	69.5	40.9
MDA	This paper		MSMT	-	**71.7**	**52.4**	MSMT	-	**79.7**	**53.0**

Table 3. Ablation studies on effectiveness of main components. Trained on Duke and Tested on Market.

Mthod	Duke→Market			
	Rank 1	Rank 5	Rank 10	mAP
Baseline	61.1	77.5	83.4	29.9
MDA w/o Meta	66.7	82.1	86.4	34.9
MDA w/o Refined net	69.0	84.3	88.7	35.8
MDA	**70.3**	**85.2**	**89.6**	**38.0**

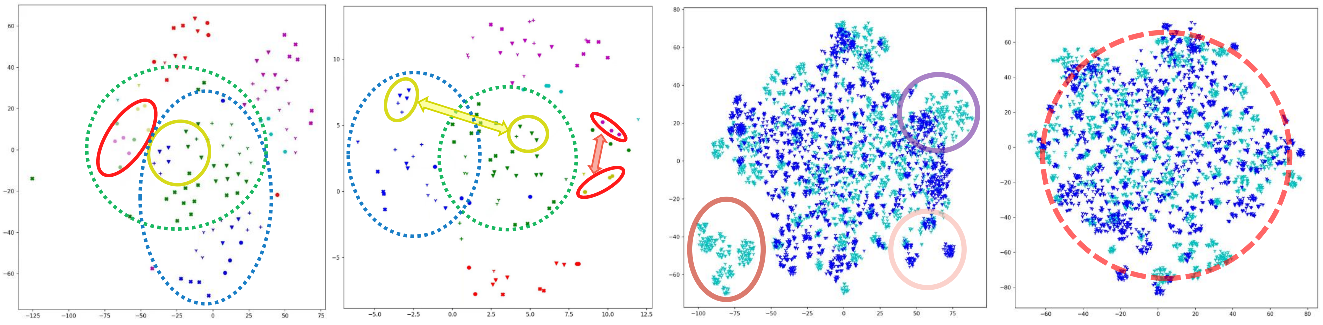
method and analyze the impact of different major components on the DG ReID task, we conduct an ablation study. Each model is trained on the Duke dataset and trained on the Market dataset. We test the following models: 1) baseline, which removes the refine net as well as the prior distribution alignment; 2) MDA w/o meta, which removes the meta-learning based optimization strategy; 3) MDA w/o refined net, which directly combines the ID feature and the generated feature without considering their weights; and 4) MDA with all components.

The experimental results are demonstrated in Table 3. These results clearly show the advantage of our contributions. Firstly, our meta-learning based optimization is useful for improving generalization. Secondly, refine net is effective for keep ID information. In general, MDA effectively weakens the influence of domain by aligning the distribution.

Ablation study on sample size for test-time adaptive updating. To investigate the effect of sample size for test-time adaptive updating, we conduct experiments on Market and Duke datasets by choosing one as the source domain and the other as the target domain. We report both mAP and R-1. Moreover, we set different ration of test-train ranging from % 0.3 to 12%. Note that 0.1% equals to 20 samples.

The experimental results are shown in Figure 4. When Duke and Market are used as training and testing respectively, both mAP and R-1 increase as the the number of samples increases and almost keeps steady at 6%. It is worth noting that with an increase of 0.3% test samples (i.e., 0.3% to 0.6%), the mAP and R-1 increased by 2.5% and 3.3%, respectively. Furthermore, when trained with Market and tested on Duke, the performance in mAP and R-1 in general keeps increasing as the number of ration increases. These results clearly prove that our test-time adaptive updating strategy can quickly adapt a model to unseen domains with a few samples.

A qualitative comparison in terms of IDs distribution. We conduct illustrative experiments by comparing our method with baseline in terms visualizing features with t-SNE, where color and shape respectively represents the person ID and camera (domain). For simplicity, we randomly select 6 Ids. Each model is trained on Market and test on Duke. The visualization results are shown in Figure 3(a) and (b). By observing the yellow and red circle of Figure 3(a), we can see that samples in the same shape (domain)



(a) Baseline (b) MDA (ours) (c) Baseline (d) MDA (ours)

Figure 3. t-SNE [30] visualization. For (a) and (b), models are trained on Market and tested on Duke. 6 person IDs are randomly selected. The color and shape represents different Person ID and Camera Id. For (c) and (d), models are trained on MSML and tested on Duke and Market. 150 person IDs are randomly selected from the test dataset. One color indicates one dataset.

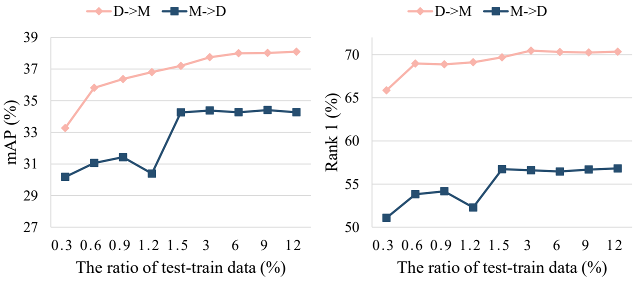


Figure 4. Ablation study on sample size of test-time adaptive updating. D→M indicates that trained on Duke and tested on Market, vice verse.

tends to be close due to the domain similarity, which influences ID discrimination. MDA makes features more close to those features with same ID instead of the same camera. As shown in Figure 3(b), Features with the same color are more concentrated.

A qualitative comparison in terms of distribution on domains We also conduct illustrative experiments on the MSMT, Duke and Market, where MSMT is a used for training, while the rest two are for testing. For each testing dataset, we randomly select 150 IDs. The t-SNE embedding are visualized in Figure 3(c) and (d), where different color represents different dataset (domain). The visualization results clearly show that distributions of different domains are closer. Specifically, isolated color blocks are reduced, which means the two distributions are closer. It shows our MDA can align multiple target domains to a similar distribution, which enhances model generalization on multiple target domains.

5. Conclusion and Discussion

In this paper, we propose a novel Meta Distribution Alignment (MDA) framework that aligns distributions across source and target domains. Previous work can not align distributions because target dataset is not available. To this end, we propose to align source and target feature distributions to a prior distribution in latent space. Furthermore, we design a meta-learning strategy to mimic the train-test process, which help the model learn a good initialization and fast adapt to unseen domains. Experimental results on large-scale DG ReID benchmark and single-source DG ReID problem show that our approach outperforms the state-of-the-art DG ReID model.

Broader impacts The most significant contribution of ReID is to improve the accuracy of automatic person recognition, autonomous driving, and other fields. Our work effectively improves the accuracy of DG ReID, which makes ReID system more practicable in security. However, ReID may bring privacy issues to our society. Firstly, ReID uses images that involve the privacy of pedestrians. These datasets should be carefully distributed and not used in illegal ways. Secondly, ReID system may intentionally or unintentionally cause an invasion of privacy, so the deployment and application of the systems should be strictly controlled. To avoid privacy breaches due to face images, we only use the back and side views of pedestrians for display.

Acknowledgements

This work is supported by National Key Research and Development Program of China (No. 2018AAA0102200), the National Natural Science Foundation of China (Grant No. 62122018, No. 61772116, No. 61872064), Sichuan Science and Technology Program (Grant No.2019JDTD0005).

References

[1] Peixian Chen, Pingyang Dai, Jianzhuang Liu, Feng Zheng, Qi Tian, and Rongrong Ji. Dual distribution alignment network for generalizable person re-identification. *arXiv preprint arXiv:2007.13249*, 2020. 6

[2] Seokeon Choi, Taekyung Kim, Minki Jeong, Hyoungseob Park, and Changick Kim. Meta batch-instance normalization for generalizable person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1, 6, 7

[3] Yongxing Dai, Xiaotong Li, Jun Liu, Zekun Tong, and Ling-Yu Duan. Generalizable person re-identification with relevance-aware mixture of experts. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, 2009. 5

[5] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2

[6] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018. 2

[7] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*. PMLR, 2017. 2

[8] Yang Fu, Yunchao Wei, Guanshuo Wang, Yuqian Zhou, Honghui Shi, and Thomas S Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *International Conference on Computer Vision*, 2019. 2

[9] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, 2008. 5

[10] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*. Springer, 2008. 6

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 5

[12] Jieru Jia, Qiuqi Ruan, and Timothy M Hospedales. Frustratingly easy person re-identification: Generalizing person re-id in practice. *arXiv preprint arXiv:1905.03422*, 2019. 2, 6

[13] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 1, 2, 6, 7

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012. 1

[15] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI Conference on Artificial Intelligence*, 2018. 3, 4

[16] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 5

[17] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 5

[18] Shengcai Liao and Ling Shao. Interpretable and Generalizable Person Re-Identification with Query-Adaptive Convolution and Temporal Lifting. In *European Conference on Computer Vision*, 2020. 5, 7

[19] Shan Lin, Haoliang Li, Chang-Tsun Li, and Alex Chichung Kot. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. *arXiv preprint arXiv:1807.01440*, 2018. 2

[20] Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI Conference on Artificial Intelligence*, 2019. 2

[21] Yutian Lin, Lingxi Xie, Yu Wu, Chenggang Yan, and Qi Tian. Unsupervised person re-identification via softened similarity learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 7

[22] Chen Change Loy, Tao Xiang, and Shaogang Gong. Multi-camera activity correlation analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 5

[23] Ping Luo, Ruimao Zhang, Jiamin Ren, Zhanglin Peng, and Jingyu Li. Switchable normalization for learning-to-normalize deep representation. *IEEE transactions on pattern analysis and machine intelligence*, 2019. 6

[24] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Asian Conference on Computer Vision*, 2018. 2

[25] Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, and Xiangyang Xue. Pose-normalized image generation for person re-identification. In *Asian Conference on Computer Vision*, 2018. 7

[26] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 5

[27] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2, 5, 6

[28] Masato Tamura and Tomokazu Murakami. Augmented hard example mining for generalizable person re-identification. *arXiv preprint arXiv:1910.05280*, 2019. 6

[29] Sebastian Thrun and Lorien Pratt. Learning to learn: Introduction and overview. In *Learning to learn*. Springer, 1998. 2

[30] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 8

[31] Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2, 7

[32] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2, 7

[33] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2, 5

[34] Ancong Wu, Wei-Shi Zheng, and Jian-Huang Lai. Unsupervised person re-identification by camera-aware similarity consistency learning. In *International Conference on Computer Vision*, 2019. 2

[35] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv preprint arXiv:1604.01850*, 2016. 5

[36] Shiyu Xuan and Shiliang Zhang. Intra-inter camera similarity for unsupervised person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 2

[37] Qize Yang, Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Patch-based discriminative feature learning for unsupervised person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2, 7

[38] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Deep metric learning for person re-identification. In *ICPR*. 2

[39] Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jian-Huang Lai. Unsupervised person re-identification by soft multilabel learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 7

[40] Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, and Yonghong Tian. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2

[41] Ji Zhang, Jingkuan Song, Yazhou Yao, and Lianli Gao. Curriculum-based meta-learning. In *ACM on Multimedia Conference*, 2021. 2

[42] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *International Conference on Computer Vision*, 2019. 1

[43] Yuyang Zhao, Zhun Zhong, Fengxiang Yang, Zhiming Luo, Yaojin Lin, Shaozi Li, and Nicu Sebe. Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1

[44] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015. 5

[45] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Associating groups of people. In *British Machine Vision Conference*, 2009. 5

[46] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 5

[47] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 7

[48] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camera style adaptation for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2

[49] Yang Zou, Xiaodong Yang, Zhiding Yu, BVK Vijaya Kumar, and Jan Kautz. Joint disentangling and adaptation for cross-domain person re-identification. In *European Conference on Computer Vision*. Springer, 2020. 2