Meta Batch-Instance Normalization for Generalizable Person Re-Identification

Seokeon Choi Taekyung Kim Minki Jeong Hyoungseob Park Changick Kim
Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
{seokeon, tkkim93, rhm033, hyoungseob, changick}@kaist.ac.kr

Abstract

Although supervised person re-identification (Re-ID) methods have shown impressive performance, they suffer from a poor generalization capability on unseen domains. Therefore, generalizable Re-ID has recently attracted growing attention. Many existing methods have employed an instance normalization technique to reduce style variations, but the loss of discriminative information could not be avoided. In this paper, we propose a novel generalizable Re-ID framework, named Meta Batch-Instance Normalization (MetaBIN). Our main idea is to generalize normalization layers by simulating unsuccessful generalization scenarios beforehand in the meta-learning pipeline. To this end, we combine learnable batch-instance normalization layers with meta-learning and investigate the challenging cases caused by both batch and instance normalization layers. Moreover, we diversify the virtual simulations via our meta-train loss accompanied by a cyclic inner-updating manner to boost generalization capability. After all, the MetaBIN framework prevents our model from overfitting to the given source styles and improves the generalization capability to unseen domains without additional data augmentation or complicated network design. Extensive experimental results show that our model outperforms the state-of-the-art methods on the large-scale domain generalization Re-ID benchmark and the cross-domain Re-ID problem. The source code is available at: https://github.com/bismex/MetaBIN.

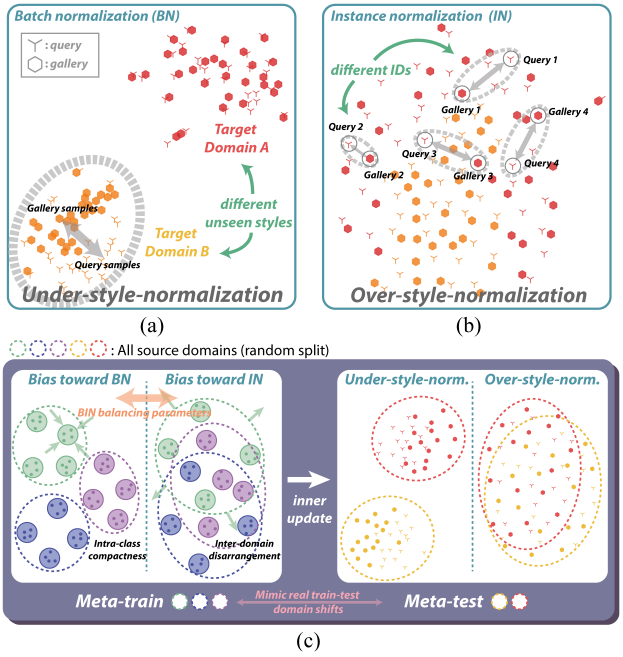


Figure 1. Illustration of unsuccessful generalization scenarios and our framework. (a) *Under-style-normalization* happens when the trained BN model fails to distinguish identities on unseen domains. (b) *Over-style-normalization* happens when the trained IN model removes even ID-discriminative information. (c) Our key idea is to generalize BIN layers by simulating the preceding cases in a meta-learning pipeline. By overcoming the harsh situations, our model learns to avoid overfitting to source styles.

1. Introduction

Person re-identification (Re-ID) aims to identify a specific person across non-overlapping cameras under various viewpoints and locations. Re-ID has attracted extensive research attention thanks to its practical importance in surveillance systems. With the development of deep Convolution Neural Networks (CNNs), person Re-ID methods [51, 54, 48, 4, 23] have achieved remarkable performance in a supervised manner, where a model is trained and tested on separated splits of the same dataset. However, this supervised approach is hardly applicable in practice due to expensive labeling costs and also suffers from severe performance degradation on an unseen target domain. For resolving this problem, unsupervised domain adaptation (UDA) methods [46, 41, 5, 11, 47, 24] have been introduced, which adapt a Re-ID model from a labeled source domain to an unlabeled target domain. The UDA approach is more practical than the supervised approach, but data collection is still required for updating the model on the target domain.

Beyond the concept of UDA, the task of domain generalization (DG) is more plausible for real-world applications since it does not require any target images to train a model. Since Finn *et al*. [10] proposed the Model-Agnostic Meta-

Learning (MAML) scheme for few-shot learning and reinforcement learning, several MAML-based methods [1, 19] have been investigated to solve the DG problem. This approach enables a DG model to achieve a good generalization capability by dividing multiple source domains into meta-train and meta-test domains to mimic real train-test domain shifts. However, most DG methods [1, 19] assume a homogeneous environment, where the source and target domains share the same label space, and they are designed for a classification task. In contrast, the task of domain generalization for person re-identification (DG Re-ID) deals with different label spaces between source and target domains for a retrieval task. Thus, it is difficult to obtain good performance when the existing DG methods are directly applied to DG Re-ID.

To this end, the recent DG Re-ID methods [17, 18, 53] have focused on a combination of batch normalization (BN) [16] and instance normalization (IN) [40]. Jia *et al*. [17] adopted this idea to Re-ID by inserting IN after the residual connection in the specific layers, inspired by [31]. However, this naïve approach not only leads to the loss of discriminative information but also requires a careful selection of the locations for applying IN. For another instance, Jin *et al*. [18] designed the Style Normalization and Restitution (SNR) module based on instance normalization and feature distillation. Since this method aims at removing style discrepancy only from the given source domains, it lacks the ability to attenuate the new style of unseen domains sufficiently.

So, how do we design normalization layers to be well generalized for DG Re-ID? To find out the answer, we conducted simple experiments that explore the properties and limitations of BN and IN layers. After training each of the BN and IN models on multiple source domains, we observed the retrieval results on unseen target domains. The BN model strives to learn discriminative information based on the style variations within each mini-batch. However, when samples of unexpected styles are given from unseen target domains, the trained model does not have sufficient ability to distinguish their IDs. We call it *under-style-normalization* in Fig. 1 (a). On the contrary, IN eliminates instance-specific style information using its own statistics. Even though IN can be helpful to remove unseen styles on target domains, it filters out even some discriminative information, as shown in Fig. 1 (b). We call it *over-style-normalization*. Since both normalization methods have limitations in the DG setting, the combination of BN and IN has to be handled carefully.

To deal with the above issues, we propose a novel generalizable Re-ID framework, named Meta Batch-Instance Normalization (MetaBIN), for learning to generalize normalization layers. The key idea is to simulate the unsuccessful generalization scenarios mentioned earlier within a meta-learning pipeline and learn more generalized representation from the virtual simulations, as shown in Fig. 1 (c). For this purpose, we design a batch-instance normalization with learnable balancing parameters between BN and IN. Depending on the balancing parameter's bias toward BN or IN, the DG model suffers from both *under-style-normalization* and *over-style-normalization* scenarios in meta-learning. By overcoming these challenging cases, the normalization layer becomes generalized. Moreover, we intentionally diversify the virtual simulations via our meta-train loss and a cyclic inner-updating manner to effectively boost the generalization capability. Our MetaBIN framework enables to train the DG model equipped with a sufficient generalization capability to novel domains.

Our main contributions can be summarized as follows:

- We propose a novel generalizable Re-ID framework called MetaBIN. This approach prevents our model from overfitting to the given source styles by generalizing normalization layers via the simulation of unsuccessful generalization scenarios in meta-learning.
- We diversify the virtual simulations through our meta-train loss and a cyclic inner-updating manner. Both ideas effectively boost generalization capability.
- We make comprehensive comparisons and achieve state-of-the-art performance on the large-scale domain generalization Re-ID benchmark and the cross-domain Re-ID problem.

2. Related Work

Generalizable person re-identification: Domain generalizable person re-identification (DG Re-ID) aims to learn a robust model for obtaining good performance on an unseen target domain without additional updates. In recent years, DG Re-ID has attracted a rapidly growing interest owing to its practical applications. Existing methods are categorized into two groups depending on the configuration of source datasets. One is to learn a robust model from a single dataset [53, 18, 22, 37] and the other is to utilize multiple large-scale datasets [36, 39, 17]. In the context of DG Re-ID based on multiple source datasets, Song *et al*. [36] constructed the large-scale domain generalization Re-ID benchmark to validate the generalization capability of DG models. They also designed Domain-Invariant Mapping Network (DIMN) to learn a mapping between a person image and its ID classifier. However, the additional mapping network slows the inference speed. Tamura *et al*. [39] suggested a simple selection strategy for data augmentation. While this method is relatively lightweight and easily applicable to other models, it lacks consistency in maintaining generalization capabilities across various unseen styles. For another instance, Jia *et al*. [17] inserted instance normalization (IN) [40] into each bottleneck in shallow layers

to eliminate useless style information, but IN brings about the loss of discriminative information unavoidably. Unlike these methods, we propose an effective framework to generalize a DG model via the simulation of poorly generalized scenarios without an additional test module and data augmentation.

Batch-instance normalization: Normalization techniques are widely used in most of the deep neural network architectures for better optimization and regularization. In particular, a combination of batch normalization (BN) [16] and instance normalization (IN) [40] has recently gained attention as a technique for improving generalization capability to novel domains, which can be divided into non-parametric and parametric methods. The non-parametric methods [17, 31, 53] focus on strategies to replace BN with IN for some layers or add IN to specific locations. However, since their performances depend on where BN and IN are mixed, it is difficult to keep task-independent consistency and requires a lot of trial and error. In contrast, the parametric methods concentrate on how to combine BN and IN in a learnable manner. For example, Nam *et al*. [30] introduced an effective batch-instance normalization layer through a simple training strategy in which BN and IN are balanced with learnable parameters. Although it has improved recognition performance compared to BN, its learnable manner sometimes leads to overfitting the style of given source domains in the DG setting. In addition, some parametric methods have been studied to handle the following tasks: few-shot learning [2] and homogeneous DG [33]. However, these methods are rarely applicable for DG Re-ID, because evaluation procedures are different from that of DG Re-ID. In this paper, we propose a novel framework that generalizes normalization layers more effectively to solve the challenging DG Re-ID problem.

Model-agnostic meta-learning for DG: Model-Agnostic Meta-Learning (MAML) [10] is a meta-learning framework for few-shot learning and reinforcement learning. MAML helps find a good initialization of parameters that are sensitive to changes in novel tasks. Recently, the MAML-based approach has been widely applied to various tasks, such as domain generalization [19, 1], object tracking [42], and frame interpolation [6]. From the perspective of DG, MLDG [19] is a representative method based on MAML, which enables to train models with a good generalization capability to novel domains via the simulation of real train-test domain-shifts. Moreover, Balaji *et al*. [1] trained regularizers in a meta-learning pipeline to achieve good cross-domain generalization. However, since these methods are designed for homogeneous DG, the performance inevitably decreases when these methods are applied directly to DG Re-ID. Inspired by those methods, we propose a novel generalizable Re-ID framework that combines normalization layers with meta-learning.

3. Proposed Method

3.1. Problem Formulation

We begin with a formal description of the domain generalizable person re-identification (DG Re-ID) problem. We assume that we are given K source domains $\mathcal{D} = \{\mathcal{D}_k\}_{k=1}^{K}$. Each source domain contains its own image-label pairs $\mathcal{D}_k = \{(\boldsymbol{x}_i^k, y_i^k)\}_{i=1}^{N_k}$, where N_k is the number of images in the source domain $\mathcal{D}_k$. Each sample $\boldsymbol{x}_i^k \in \mathcal{X}_k$ is associated with an identity label $y_i^k \in \mathcal{Y}_k = \{1, 2, \ldots, M_k\}$, where M_k is the number of identities in the source domain $\mathcal{D}_k$. While all domains share the label spaces in the general homogeneous DG setting as $\mathcal{Y}_i = \mathcal{Y}_j = \mathcal{Y}_*, \forall i,j, 1 \leq i,j \leq K$, source and target domains have completely disjoint label spaces in the DG Re-ID setting as $\mathcal{Y}_i \neq \mathcal{Y}_j \neq \mathcal{Y}_*$. In other words, this task is an application of heterogeneous DG, such that the number of identities in all source domains can be expressed as $M = \sum_{k=1}^{K} M_k$. In the training phase, we train a DG model using the aggregated image-label pairs of all source domains. In the testing phase, we perform a retrieval task on unseen target domains without additional model updating.

3.2. Batch-Instance Normalization for DG-ReID

In line with our goal of DG, we employ an instance normalization technique [40] to generalize well on unseen domains. Similar to [30], we design a batch-instance normalization module as a mixture of BN and IN with learnable balancing parameters for each channel of all normalization layers. Let $\mathbf{x} \in \mathbb{R}^{N \times C \times H \times W}$ be an input mini-batch of a certain layer, where N, C, H, and W denote the batch, channel, height, and width dimensions, respectively. We combine batch normalization (BN) with instance normalization (IN) as follows:

$$\mathbf{y} = \rho\left(\gamma_B \cdot \hat{\mathbf{x}}_B + \beta_B\right) + (1-\rho)\left(\gamma_I \cdot \hat{\mathbf{x}}_I + \beta_I\right), \quad (1)$$

where the subscripts B and I denote variables with respect to BN and IN, $\hat{\mathbf{x}}$ is the normalized response by mean and variance, $\gamma, \beta \in \mathbb{R}^C$ are affine transformation parameters, and $\rho \in [0,1]^C$ is an additional learnable parameter to balance BN and IN. Note that ρ is a channel-wise parameter for each normalization layer in a feature extractor. Different from [30], we apply each affine transformation to normalized responses for BN and IN, which allows batch-instance normalization layers to learn various representations.

However, the training mechanism of balancing BN and IN through learnable parameters is a critical issue that can easily overfit the source domain's styles. In other words, the learnable manner forces the balancing parameters to be optimized depending only on styles within source domains. As a result, the normalization layers are unsuccessfully generalized at given unseen target domains, as shown in Fig. 1 (a) and (b). To deal with this problem, we develop our normalization model by using a meta-learning pipeline.

3.3. Meta Batch-Instance Normalization

MetaBIN framework: MLDG [19] is a representative DG method based on Model-Agnostic Meta-Learning (MAML) [10], which enables to train models with a good generalization capability to novel domains through the virtual simulation of real train-test domain-shifts. Inspired by this, we apply the MAML scheme to the updating process of the balancing parameters to prevent our model from overfitting to the source style. To this end, we separate an episode that updates the balancing parameters from another episode that updates the rest of the parameters, and then perform both episodes alternately at each training iteration. The separation of the base model updating process not only ensures baseline performance but also enables effective learning compared to updating all parameters in MLDG [19].

Formally, we denote a classifier $g_\phi(\cdot)$ parameterized by ϕ and a feature extractor $f_\theta(\cdot)$ parameterized by $\theta = (\theta_f, \theta_\rho)$, where θ_ρ and θ_f are the balancing parameters and the remaining parameters of the feature extractor, respectively. In the base model updating process, we update the classifier parameters ϕ and the feature extractor parameters θ_f except the balancing parameters θ_ρ. In the meta-learning process, only the balancing parameters θ_ρ are updated. The overall methodological flow is summarized in Algorithm 1.

Base model updates: In this stage, we update all parameters without the balancing parameters θ_ρ. To this end, we sample a mini-batch $\mathcal{X}_B$ by aggregating the labeled images from all source domains $\mathcal{D}$. We adopt two loss functions for updating a base model. First, we use the cross-entropy loss $\mathcal{L}_{\text{ce}}$ for ID-discriminative learning as follows:

$$\mathcal{L}_{\text{ce}}(\mathcal{X}_B;\theta,\phi) = \frac{1}{N_B}\sum_{i=1}^{N_B} l_{\text{ce}}\Big(g_\phi\big(f_\theta(\boldsymbol{x}_i)\big), y_i\Big), \quad (2)$$

where N_B denotes the number of samples in a mini-batch $\mathcal{X}_B$. In addition, we apply the label-smoothing method [38] since there are too many identities in aggregated source domains. It helps prevent our model from overfitting to training IDs.

Most of the Re-ID methods [7, 43, 18] combine the cross-entropy loss with the triplet loss together for similarity learning. The second loss is expressed as follows:

$$\mathcal{L}_{\text{tr}}(\mathcal{X}_B;\theta) = \frac{1}{N_B}\sum_{i=1}^{N_B} [d(\boldsymbol{f}_i^a, \boldsymbol{f}_i^p) - d(\boldsymbol{f}_i^a, \boldsymbol{f}_i^n) + m]_+, \quad (3)$$

where $\boldsymbol{f}_i^a = f_\theta(x_i^a)$ indicates the feature vector of an anchor sample x_i^a, $d(\cdot,\cdot)$ is the Euclidean distance, m is a margin parameter, and $[z]_+ = \max(z, 0)$. For each sample $\boldsymbol{f}_i^a$, we select the hardest positive sample $\boldsymbol{f}_i^p$ and the hardest negative sample $\boldsymbol{f}_i^n$ within a mini-batch in the same way as [14]. The triplet loss helps enhance the intra-class compactness and inter-class separability in the Euclidean space. To maximize the synergy between $\mathcal{L}_{\text{ce}}$ and $\mathcal{L}_{\text{tr}}$, we use the BNNeck structure as proposed in [26]. The overall loss is formulated as follows:

Algorithm 1 MetaBIN

Input: Source domains $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K\}$, pre-trained parameters θ_f, hyperparameters α, β, γ.
Output: Feature extractor $f_\theta(\cdot)$, classifier $g_\phi(\cdot)$

Initialize parameters θ_ρ, ϕ
for ite **in** iterations **do**
  Base model update: // Eq. (2)-Eq. (5)
  Sample a mini-batch $\mathcal{X}_B$ from $\mathcal{D}$.
  $\mathcal{L}_{\text{base}}(\mathcal{X}_B;\theta,\phi) = \mathcal{L}_{\text{ce}}(\mathcal{X}_B;\theta,\phi) + \mathcal{L}_{\text{tr}}(\mathcal{X}_B;\theta)$
  $(\theta_f, \phi) \leftarrow \big(\theta_f - \alpha\nabla_{\theta_f}\mathcal{L}_{\text{base}}(\mathcal{X}_B;\theta_f,\theta_\rho,\phi),$
    $\phi - \alpha\nabla_{\phi}\mathcal{L}_{\text{base}}(\mathcal{X}_B;\theta_f,\theta_\rho,\phi)\big)$
  Domain-level sampling:
  Split $\mathcal{D}$ as $(\mathcal{D}_{\text{mtr}} \cap \mathcal{D}_{\text{mte}} = \emptyset,\ \mathcal{D}_{\text{mtr}} \cup \mathcal{D}_{\text{mte}} = \mathcal{D})$
  Meta-train: // Eq. (6)-Eq. (9)
  Sample a mini-batch $\mathcal{X}_S$ from $\mathcal{D}_{\text{mtr}}$.
  $\mathcal{L}_{\text{mtr}}(\mathcal{X}_S;\theta) = \mathcal{L}_{\text{scat}}(\mathcal{X}_S;\theta) + \mathcal{L}_{\text{shuf}}(\mathcal{X}_S;\theta) + \mathcal{L}_{\text{tr}}(\mathcal{X}_S;\theta)$
  $\theta'_\rho = \theta_\rho - \beta\nabla_{\theta_\rho}\mathcal{L}_{\text{mtr}}(\mathcal{X}_S;\theta_f,\theta_\rho)$
  Meta-test: // Eq. (10)
  Sample a mini-batch $\mathcal{X}_T$ from $\mathcal{D}_{\text{mte}}$.
  $\theta_\rho \leftarrow \theta_\rho - \gamma\nabla_{\theta_\rho}\mathcal{L}_{\text{tr}}(\mathcal{X}_T;\theta_f,\theta'_\rho)$

$$\mathcal{L}_{\text{base}}(\mathcal{X}_B;\theta,\phi) = \mathcal{L}_{\text{ce}}(\mathcal{X}_B;\theta,\phi) + \mathcal{L}_{\text{tr}}(\mathcal{X}_B;\theta). \quad (4)$$

Then we update our base model as follows:

$$(\theta_f, \phi) \leftarrow \big(\theta_f - \alpha\nabla_{\theta_f}\mathcal{L}_{\text{base}}(\mathcal{X}_B;\theta_f,\theta_\rho,\phi), \quad \phi - \alpha\nabla_{\phi}\mathcal{L}_{\text{base}}(\mathcal{X}_B;\theta_f,\theta_\rho,\phi)\big). \quad (5)$$

We note that the balancing parameters θ_ρ are not updated.

Domain-level sampling and meta-learning: To achieve domain generalization, we split source domains $\mathcal{D}$ into meta-train domains $\mathcal{D}_{\text{mtr}}$ and meta-test domains $\mathcal{D}_{\text{mte}}$ randomly at each iteration. This separation is to mimic real train-test domain-shifts to generalize our normalization layers. In this way, we inner-update the balancing parameters via the meta-train loss, and then validate the updated model at unseen-like meta-test domains. Next, we meta-update the balancing parameters via the meta-test loss. After all, the balancing parameters and base model parameters are alternately generalized in the whole process.

Meta-train: Compared to the previous DG methods [19, 1], we introduce a novel concept to significantly improve generalization capability. We start with an explanation of virtual simulations, as shown in Fig. 1 (c). We attempt to simulate unsuccessful generalization scenarios via our meta-train loss by inner-updating the balancing parameters to any space that cannot be explored with the given source domains and general classification loss functions. More specifically, we induce the balancing parameters to be biased toward IN for investigating the virtual *over-style-normalization* situation where our model becomes much more normalized beyond the styles of the source domain. On the contrary, we also lead to the parameters to be

biased toward BN for exploring the virtual *under-style-normalization* situation in which our model fails to distinguish identities at the unexpected styles. In this way, we design a complementary objective function that can trigger both challenging cases, which provides an opportunity to escape from the local-minima by overcoming the harsh situations at the unseen-like meta-test domains.

To implement this, we first suggest the losses from the perspective of *over-style-normalization*. Our main point is to enhance intra-domain diversity and disarrange inter-domain distributions like confusing multiple styles. To spread the feature distribution for each domain, we introduce an intra-domain scatter loss as follows:

$$\mathcal{L}_{\text{scat}}(\mathcal{X}_S;\theta) = \frac{1}{N_S}\sum_{k=1}^{K_S}\sum_{i=1}^{N_S^k} cos\big(\boldsymbol{f}_i^k, \bar{\boldsymbol{f}}^k\big), \tag{6}$$

where $\bar{\boldsymbol{f}}^k$ denotes the mean feature vector (centroid) of domain k in a mini-batch, K_S is the number of meta-train domains, N_S^k is the number of meta-train samples for domain k, N_S is the number of all meta-train samples, and $cos(\boldsymbol{a},\boldsymbol{b}) = \boldsymbol{a}\cdot\boldsymbol{b}/\|\boldsymbol{a}\|\|\boldsymbol{b}\|$.

In addition to this, we propose an inter-domain shuffle loss for supporting the virtual effect of style normalization further. This loss pulls the negative sample of the inter-domain and pushes the negative sample of the intra-domain, so that the inter-domain distributions are shuffled. It is expressed as follows:

$$\mathcal{L}_{\text{shuf}}(\mathcal{X}_S;\theta) = \frac{1}{N_S}\sum_{i=1}^{N_S} l_s\big(d(\boldsymbol{f}_i^a, \boldsymbol{f}_i^{n-}) - d(\boldsymbol{f}_i^a, \boldsymbol{f}_i^{n+})\big), \tag{7}$$

where $\boldsymbol{f}_i^a$, $\boldsymbol{f}_i^{n-}$, and $\boldsymbol{f}_i^{n+}$ indicate the feature representations of an anchor sample, an inter-domain negative sample, and an intra-domain negative sample. $d(\cdot,\cdot)$ is the Euclidean distance, and l_s is the softplus function [9].

At the same time, we add the triplet loss $\mathcal{L}_{\text{tr}}(\mathcal{X}_S;\theta)$ from perspective of *under-style-normalization*. It enhances intra-class compactness regardless style differences. The overall loss for meta-train is as follows:

$$\mathcal{L}_{\text{mtr}}(\mathcal{X}_S;\theta) = \mathcal{L}_{\text{scat}}(\mathcal{X}_S;\theta) + \mathcal{L}_{\text{shuf}}(\mathcal{X}_S;\theta) + \mathcal{L}_{\text{tr}}(\mathcal{X}_S;\theta). \tag{8}$$

The combination of the triplet loss alleviates the excessive discrimination reduction problem caused by $\mathcal{L}_{\text{scat}}(\mathcal{X}_S;\theta)$ and $\mathcal{L}_{\text{shuf}}(\mathcal{X}_S;\theta)$.

Based on the meta-train loss $\mathcal{L}_{\text{mtr}}$, we inner-update the balancing parameters from θ_ρ to θ'_ρ as follows:

$$\theta'_\rho = \theta_\rho - \beta\nabla_{\theta_\rho}\mathcal{L}_{\text{mtr}}(\mathcal{X}_S;\theta_f,\theta_\rho), \tag{9}$$

where β is a learning rate for the inner-level optimization.

Cyclic inner-updates: In the general MAML-based DG methods [19, 1], the learning rate β is pre-defined as a constant. To promote the diversity of virtual simulations, we adopt the cyclical learning rate [35] for the inner-level optimization, which we call it a cyclical inner-updating method. More specifically, this method affects how much the balancing parameters are updated. Therefore, it diversifies generalization scenarios from the easy case by a small β to the difficult case by a large β.

Meta-test: After moving the balancing parameters in the inner-level optimization step, we evaluate our model at the unseen-like samples $\mathcal{X}_T$ from meta-test domains $\mathcal{D}_{\text{mte}}$. In this step, we can examine various generalization scenarios depending on the movement of the balancing parameters. To validate a retrieval task effectively, we employ the triplet loss with the updated balancing parameters θ'_ρ as follows:

$$\theta_\rho \leftarrow \theta_\rho - \gamma\nabla_{\theta_\rho}\mathcal{L}_{\text{tr}}(\mathcal{X}_T;\theta_f,\theta'_\rho). \tag{10}$$

From the equation, we meta-update the balancing parameters to overcome the virtual simulations. Eventually, our model learns to generalize the balancing parameters by the final optimization process. As these meta-learning and base model updating steps are alternately trained, the generalization capability to unseen domains is improved effectively.

4. Experiments

4.1. Datasets and Settings

Datasets: To evaluate the generalization capability of our method, we employ the large-scale domain generalization (DG) Re-ID benchmark [36] and the cross-domain Re-ID problem. In the large-scale DG Re-ID setting, source datasets contain CUHK02 [20], CUHK03 [21], Market-1501 [49], DukeMTMC-ReID [50], and CUHK-SYSU PersonSearch [45], and target datasets include VIPeR [12], PRID [15], GRID [25], and QMUL i-LIDS [44]. All images in the source datasets are used for training regardless of train/test splits, and the total number of identities is M = 18,530 with N = 121,765 images. In the cross-domain Re-ID setting, we employ Market-1501 [49] and DukeMTMC-ReID [50]. We alternately construct the two datasets into source and target domains. At this time, we regard each camera within a dataset as an individual domain.

Implementation details: For a fair comparison, we adopted MobileNetV2 [32] with a multiplier of 1.4 and ResNet-50 [13] as backbone networks. The weights are pre-trained on ImageNet [8]. All balancing parameters are initialized to 1. For training, each image is resized to 256 $\times$ 128. We select 16 samples for each domain in a mini-batch. In the base model updating step, all images in a mini-batch are used. However, in a meta-learning pipeline, domains are separated into meta-train and meta-test from another mini-batch. The label-smoothing parameter is 0.1 and the margin in the triplet loss is 0.3. For updating our base model, we use the SGD optimizer with a momentum of 0.9 and a weight decay of 5×10^{-4}. Its initial learning rate α is 0.01, which is warmed up for 10 epochs as [26] and decayed to its 0.1$\times$ and 0.01$\times$ at 40 and 70 epochs. In the meta-learning step, the balancing parameters are updated by

Table 1. Performance (%) comparison with the state-of-the-arts on the large-scale DG Re-ID benchmark, where '†' is based on ResNet-50.

Method	Large-scale domain generalization Re-ID (multi-source DG)																	
	Average		Target: VIPeR (V) [12]				Target: PRID (P) [15]				Target: GRID (G) [25]				Target: i-LIDS (I) [44]			
	R-1	*m*AP	R-1	R-5	R-10	*m*AP	R-1	R-5	R-10	*m*AP	R-1	R-5	R-10	*m*AP	R-1	R-5	R-10	*m*AP
DIMN [36]	47.5	57.9	51.2	70.2	76.0	60.1	39.2	67.0	76.7	52.0	29.3	53.3	65.8	41.1	70.2	89.7	94.5	78.4
AugMining [39]	51.8	-	49.8	70.8	77.0	-	34.3	56.2	65.7	-	46.6	**67.5**	76.1	-	76.3	93.0	95.3	-
Switchable (BN+IN) [27]	57.0	65.6	51.6	72.9	80.8	61.4	59.6	78.6	90.1	69.4	39.3	58.8	68.1	48.1	77.3	91.2	94.8	83.5
DualNorm [17]	57.6	61.8	53.9	62.5	75.3	58.0	60.4	73.6	84.8	64.9	41.4	47.4	64.7	45.7	74.8	82.0	91.5	78.5
DDAN [3]	59.0	63.1	52.3	60.6	71.8	56.4	54.5	62.7	74.9	58.9	**50.6**	62.1	73.8	55.7	78.5	85.3	92.5	81.5
DDAN [3] w/ [17]	60.9	65.1	56.5	65.6	76.3	60.8	62.9	74.2	85.3	67.5	46.2	55.4	68.0	50.9	78.0	85.7	93.2	81.2
MetaBIN (Ours)	**64.7**	**72.3**	**56.9**	**76.7**	**82.0**	**66.0**	**72.5**	**88.2**	**91.3**	**79.8**	49.7	**67.5**	**76.8**	**58.1**	**79.7**	**93.3**	**97.3**	**85.5**
SNR† [18]	57.3	66.4	52.9	-	-	61.3	52.1	-	-	66.5	40.2	-	-	47.7	**84.1**	-	-	**89.9**
DualNorm† [17]	62.7	-	59.4	-	-	-	69.6	-	-	-	43.7	-	-	-	78.2	-	-	-
MetaBIN† (Ours)	**66.0**	**73.6**	**59.9**	**78.4**	**82.8**	**68.6**	**74.2**	**89.7**	**92.2**	**81.0**	**48.4**	**70.3**	**77.2**	**57.9**	81.3	**95.0**	**97.0**	87.0

another SGD optimizer without momentum and weight decay. The meta-train step-size β oscillates back and forth in the range [0.001, 0.1] with the triangular policy [35]. The meta-test step-size γ is fixed to 0.1. The training stops at 100 epochs. To speed up the training process and increase memory efficiency, we use the automatic mixed-precision training [29] in the entire process and the first-order approximations [10] in meta-optimization. All experiments are conducted on an NVIDIA Titan Xp GPU using Pytorch.

Evaluation metrics: We follow the common evaluation metrics for Re-ID as mean Average Precision (mAP) and Cumulative Matching Characteristic (CMC) at Rank-k.

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

Large-scale DG Re-ID benchmark: We evaluate our MetaBIN framework on the large-scale domain generalization Re-ID benchmark [36], which is shown in Table 1. For a fair comparison, we provide both results based on MobileNetV2 and ResNet-50. All Re-ID models have been evaluated individually on each target dataset. Among the competitors, DualNorm [17] achieved relatively good performance on VIPeR and PRID, and DDAN [3] also obtained comparable performance on GRID. In addition, SNR [18] achieved impressive results on the i-LIDS dataset in the experiment based on ResNet-50. Nevertheless, our MetaBIN method outperformed all competing methods by a significant margin in the average performance. The main reasons come from two aspects: 1) Our MetaBIN framework, which simulates the virtual scenarios in a meta-learning pipeline, effectively enhances the generalization capability to novel domains; 2) With the addition of simple balancing parameters and the separation of learning episodes, we can solve the challenging heterogeneous DG problem without synthetic data augmentation or complicated network design.

Cross-domain Re-ID problem: To demonstrate the superiority of our method, we additionally compare our MetaBIN framework with a variety of state-of-the-art methods in the cross-domain Re-ID problem, which is shown in Table 2. 'Market1501 → DukeMTMC' indicates that Market-1501 [49] is a labeled source domain and DukeMTMC-ReID [50] is an unseen target domain. In other words, we need to generalize a Re-ID model so that it works well on the unseen target dataset using only the images on a single source dataset. The setting of cross-domain Re-ID is somewhat different from that of large-scale domain generalization. First, the style variation within a single dataset is relatively small. Therefore, we increased the meta-optimization step size to 0.2 on Market-1501 [49] and 0.5 on DukeMTMC-ReID [50]. Second, some identities are simultaneously distributed on different camera domains. Thus, we added the cross-entropy loss to the meta-train and meta-test losses for enhancing the ID-discrimination ability. Third, since the number of domains is different from the previous one, we divided the entire camera domains in half for each dataset. Finally, for a fair comparison with SNR [18], we employed ResNet-50 [13] with color jittering. Table 2 show that our MetaBIN framework achieved comparable performance on both settings. In the end, we prove that our framework has the potential to improve the generalization capability even on the single-source dataset with relatively small domain discrepancy.

Table 2. Performance (%) comparison with the state-of-the-arts on the cross-domain Re-ID problem.

Method	Cross-domain Re-ID (single-source DG)							
	Market1501 → DukeMTMC				DukeMTMC → Market1501			
	R-1	R-5	R-10	*m*AP	R-1	R-5	R-10	*m*AP
IBN-Net [31]	43.7	59.1	65.2	24.3	50.7	69.1	76.3	23.5
OSNet [53]	44.7	59.6	65.4	25.9	52.2	67.5	74.7	24.0
OSNet-IBN [53]	47.9	62.7	68.2	27.6	57.8	74.0	79.5	27.4
CrossGrad [34]	48.5	63.5	69.5	27.1	56.7	73.5	79.5	26.3
QAConv [22]	48.8	-	-	28.7	58.6	-	-	27.2
L2A-OT [52]	50.1	64.5	70.1	29.2	63.8	80.2	84.6	30.2
OSNet-AIN [53]	52.4	66.1	71.2	30.5	61.0	77.0	82.5	30.6
SNR [18]	55.1	-	-	**33.6**	66.7	-	-	33.9
MetaBIN (Ours)	**55.2**	**69.0**	**74.4**	33.1	**69.2**	**83.1**	**87.8**	**35.9**

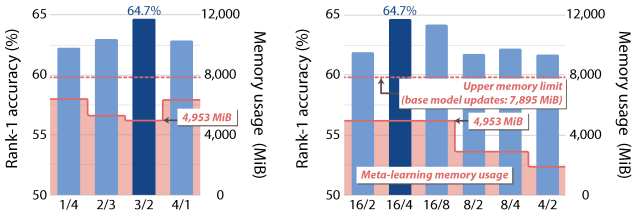
4.3. Ablation study

We perform comprehensive ablation studies to show the effectiveness of our MetaBIN framework and detailed components through the average performance on the large-scale DG Re-ID benchmark.

Influence of model components: Table 3 reports the ablation results of our novel framework including meta-train losses and the cyclic inner-updating method. We employed the cross-entropy and triplet losses as a baseline. We ob-

Table 3. Ablation studies of our MetaBIN framework in the average performance on the large-scale DG Re-ID benchmark.

Method	$\mathcal{L}_{\text{mtr}}$	$\mathcal{L}_{\text{mte}}$	β	R-1	*m*AP
BN	-	-	-	50.9	59.5
MetaBIN	$\mathcal{L}_{\text{ce}}$	$\mathcal{L}_{\text{ce}}$	fixed	60.6	69.4
MetaBIN	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	fixed	62.0	69.9
MetaBIN	$\mathcal{L}_{\text{tr}}$	$\mathcal{L}_{\text{tr}}$	fixed	62.8	70.8
MetaBIN	$\mathcal{L}_{\text{tr}}, \mathcal{L}_{\text{scat}}$	$\mathcal{L}_{\text{tr}}$	fixed	63.0	71.0
MetaBIN	$\mathcal{L}_{\text{tr}}, \mathcal{L}_{\text{shuf}}$	$\mathcal{L}_{\text{tr}}$	fixed	63.1	71.0
MetaBIN	$\mathcal{L}_{\text{tr}}, \mathcal{L}_{\text{scat}}, \mathcal{L}_{\text{shuf}}$	$\mathcal{L}_{\text{tr}}$	fixed	63.5	71.3
MetaBIN	$\mathcal{L}_{\textbf{tr}}, \mathcal{L}_{\textbf{scat}}, \mathcal{L}_{\textbf{shuf}}$	$\mathcal{L}_{\textbf{tr}}$	**cyclic**	**64.7**	**72.3**



(a) Number of domains (meta-train / meta-test)

(b) Number of images (for each domain / for each ID)

Figure 2. Performance (%) and memory usage (MiB) analysis according to the sampling manner of domains and images.

served the following aspects: 1) Our novel MetaBIN framework that updates balancing parameters in meta-learning improves the generalization capability with a large performance gap by alleviating the overfitting problem at given source domains; 2) In simulating real train-test domain shifts, the triplet loss is more suitable than the cross-entropy loss for DG Re-ID, which is different from the general homogeneous DG methods [19, 1]; 3) Our meta-train losses are complementary to each other and have a synergistic effect with the cyclic inner-updating method.

Influence of domain-level sampling: In this part, we compare the differences depending on the sampling manner of domains and images in the meta-learning step. In Fig. 2 (a), whereas the simulation quality deteriorates if the number of meta-train domains is too small, the generalization capability to overcome the overfitting issues becomes insufficient if the number of meta-test domains is too small. Therefore, the domain sampling between meta-train and meta-test should be adequately balanced. Our method achieved the highest performance and the lowest memory usage when the number of meta-train and meta-test domains is 3 and 2, respectively. Figure 2 (b) shows the performance and memory usage according to the number of images per domain in a mini-batch. It is noteworthy that we chose the highest performance case under the condition that does not exceed the memory usage (7,895MiB) of updating a base model. In other words, the proposed meta-learning pipeline does not increase the memory usage, since the meta-learning and base model updating steps are divided. Thus, it has an advantage in terms of memory consumption over the MLDG method [19] of meta-updating the entire parameters.

Table 4. Performance (%) comparison in a meta-learning pipeline.

Method	$\mathcal{L}_{\text{base}}$	MLDG [19]	cyclic β	R-1	*m*AP
BN	$\mathcal{L}_{\text{ce}}$	✗	✗	50.2	59.6
	$\mathcal{L}_{\text{ce}}$	✓	✗	50.5	59.2
	$\mathcal{L}_{\text{ce}}$	✓	✓	52.3	60.9
	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	✗	✗	50.9	59.5
	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	✓	✗	52.2	61.2
	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	✓	✓	53.6	61.8
BIN [30]	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	✗	✗	54.8	63.1
	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	✓	✗	57.9	65.7
	$\mathcal{L}_{\text{ce}}, \mathcal{L}_{\text{tr}}$	✓	✓	58.4	66.3
MetaBIN (replace with BIN [30])				60.6	68.8
MetaBIN (w/o episode separation)				60.9	69.1
MetaBIN				**64.7**	**72.3**

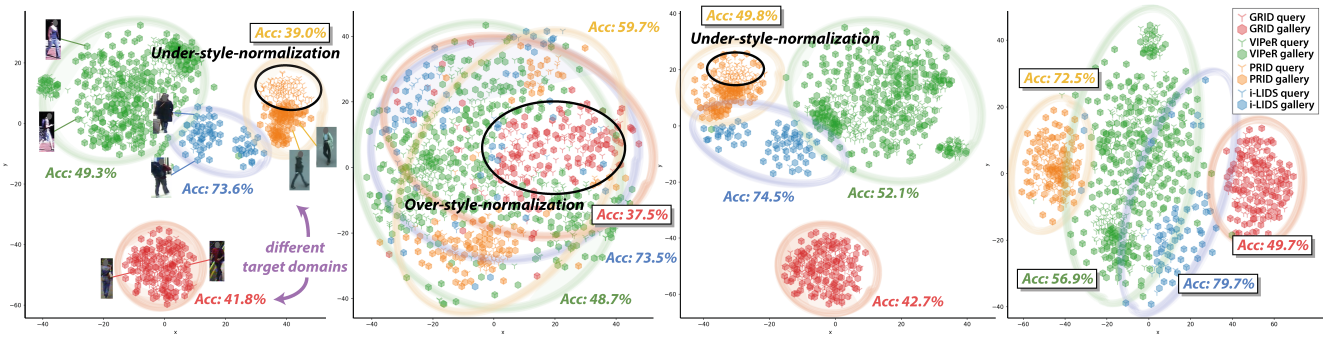
Table 5. Performance (%) comparison with normalization methods in DG and supervised settings, where 'S' is single normalization, 'N' is non-parametric normalization, 'P' is parametric normalization, 'BN+IN half' is a channel-wise combination of BN and IN.

	Method	Large-scale DG		Supervised (Market1501)	
		R-1	*m*AP	R-1	*m*AP
S	BN	50.9	59.5	87.2	67.9
	IN	54.9	63.3	71.9	46.1
N	DualNorm [17]	57.6	61.8	82.6	57.2
	BN+IN half	56.5	65.3	79.5	53.9
P	BIN [30]	54.8	63.1	87.5	67.8
	MetaBIN (Ours)	**64.7**	**72.3**	**87.9**	**68.5**

4.4. Further Analysis

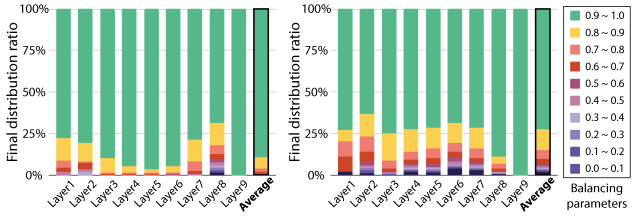
Analysis from the perspective of MAML: Table 4 presents the experimental results in a meta-learning pipeline. MLDG [19] is a representative MAML-based method designed for homogeneous DG. We observed that MLDG does not improve performance much, which means that it is difficult to directly apply the conventional homogeneous DG method to the challenging DG Re-ID task. On the other hand, when the meta-learning approach is combined with the triplet loss or a batch-instance normalization layer, its performance has been meaningfully improved. Thus, these methods are valid for solving the DG Re-ID problem. However, we note that the generalization potential of the existing BIN model [30] was insufficient compared to that of our model. The proposed cyclic inner-updating method has made significant improvements in all cases, indicating an effective model-free method. Finally, the separation of training episodes contributed to improving performance considerably. It demonstrates that selectively updating only the balancing parameters in the complicated meta-learning pipeline brings training stability and improves the generalization capacity.

Analysis from the perspective of normalization: We analyze our MetaBIN framework and other normalization methods through the performance and t-SNE visualization [28]. First, we compare BN with IN as single normalization techniques. We observed *under-style-normalization* in Fig. 3 (a) and *over-style-normalization* in Fig. 3 (b), and it turned out that the performance plummeted when these scenarios occurred as reported in Table 5. Furthermore, since IN filtered out discriminative information too much,



(a) Batch Normalization (BN) (b) Instance Normalization (IN) (c) Batch-Instance Normalization (BIN) (d) MetaBIN (Ours)

Figure 3. The t-SNE visualization of the embedding vectors on four unseen target datasets (VIPeR, PRID, GRID, and i-LIDS). Trained models are evaluated individually for each dataset, and query and gallery samples are expressed in different shapes. Best viewed in color.



(a) Batch-Instance Normalization (BIN) (b) MetaBIN (Ours)

Figure 4. Analysis on the final distribution ratios of balancing parameters for each layer. Best viewed in color.

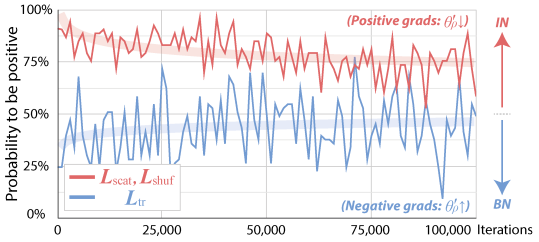


Figure 5. Analysis on the gradients in the inner-optimization step according to the meta-train loss ($\mathcal{L}_{tr}$, $\mathcal{L}_{scat}$, $\mathcal{L}_{shuf}$).

its performance had been greatly degraded in the supervised setting. Meanwhile, the non-parametric or parametric combination of BN and IN alleviated these problems. The non-parametric methods in the DG experiment and the parametric methods in the supervised setting showed relatively high performance. However, a poorly generalized case was also observed at the BIN method [30], as expressed in Fig. 3 (c). On the other hand, our MetaBIN method overcame the overfitting issue by simulating unsuccessful generalization in the meta-learning pipeline, eventually surpassing all normalization methods in the DG experiment. Figure 3 (d) shows that our method has shorter query-gallery distances than those of other methods. Besides, our method improves performance even in the supervised setting. Thus, we demonstrate that our MetaBIN method is generalizable and practical for real-world situations.

Analysis on the balancing parameters: To understand how normalization layers are generalized, we investigate the final distribution ratio of balancing parameters, as shown in Fig. 4. Both experiments have been conducted based on MobileNetV2 [32] trained on the large-scale DG Re-ID dataset. Note that all balancing parameters have been initialized to 1 (toward BN). We observed that most of the parameters tend to maintain the BN properties, but only some parameters are biased toward IN. Since the performance improved through this process, we demonstrate that removing useless instance-specific style information contributes to improving generalization capability. In particular, the distribution difference between BIN [30] and MetaBIN shows that deflecting some channels toward IN in the middle layer as well as the shallow layer alleviates the overfitting issue and promotes the generalization capability.

Analysis on the meta-train loss: We analyze the virtual simulation scenarios through visualization of gradients calculated in the inner-optimization step, as illustrated in Fig. 5. First, the intra-domain scatter loss $\mathcal{L}_{scat}$ and inter-domain shuffle loss $\mathcal{L}_{shuf}$ generate positive gradients, which act to move the balancing parameters in the IN direction. In addition, the triplet loss $\mathcal{L}_{tr}$ promotes negative gradients, which cause the parameters to shift in the BN direction. Eventually, we prove that our meta-train losses induce the balancing parameters to the specific directions, which means that unsuccessful generalization scenarios can be deliberately simulated. We also emphasize that combining all these losses helps improve performance, as shown in Table 3.

5. Conclusion

In this work, we have proposed a novel generalizable person re-identification framework, called Meta Batch-Instance Normalization (MetaBIN). Compared to previous studies, the proposed method alleviates the overfitting problem by investigating unsuccessful generalization scenarios based on our observation. Furthermore, our novel meta-train loss and cyclic inner-updating method diversify virtual simulations and eventually boost generalization capability to unseen domains. Extensive experiments and comprehensive analysis in the single- and multi-source DG settings demonstrate its superiority over state-of-the-art methods.

References

[1] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 998–1008, 2018.

[2] John Bronskill, Jonathan Gordon, James Requeima, Sebastian Nowozin, and Richard Turner. Tasknorm: Rethinking batch normalization for meta-learning. In *International Conference on Machine Learning (ICML)*, pages 1153–1164. PMLR, 2020.

[3] Peixian Chen, Pingyang Dai, Jianzhuang Liu, Feng Zheng, Qi Tian, and Rongrong Ji. Dual distribution alignment network for generalizable person re-identification. *Proceedings of AAAI Conference on Artificial Intelligence*, 2021.

[4] Xuesong Chen, Canmiao Fu, Yong Zhao, Feng Zheng, Jingkuan Song, Rongrong Ji, and Yi Yang. Salience-guided cascaded suppression network for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3300–3310, 2020.

[5] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Instance-guided context rendering for cross-domain person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 232–242, 2019.

[6] Myungsub Choi, Janghoon Choi, Sungyong Baik, Tae Hyun Kim, and Kyoung Mu Lee. Scene-adaptive video frame interpolation via meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9444–9453, 2020.

[7] Seokeon Choi, Sumin Lee, Youngeun Kim, Taekyung Kim, and Changick Kim. Hi-cmd: Hierarchical cross-modality disentanglement for visible-infrared person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10257–10266, 2020.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. Ieee, 2009.

[9] Charles Dugas, Yoshua Bengio, François Bélisle, Claude Nadeau, and René Garcia. Incorporating second-order functional knowledge for better option pricing. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 472–478, 2001.

[10] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *arXiv preprint arXiv:1703.03400*, 2017.

[11] 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 *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 6112–6121, 2019.

[12] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision (ECCV)*, pages 262–275. Springer, 2008.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[14] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[15] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *Scandinavian Conference on Image Analysis*, pages 91–102. Springer, 2011.

[16] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[17] 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.

[18] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3143–3152, 2020.

[19] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. *arXiv preprint arXiv:1710.03463*, 2017.

[20] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3594–3601, 2013.

[21] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 152–159, 2014.

[22] Shengcai Liao and Ling Shao. Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting. *arXiv preprint arXiv:1904.10424*, 2019.

[23] Chong Liu, Xiaojun Chang, and Yi-Dong Shen. Unity style transfer for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6887–6896, 2020.

[24] Jiawei Liu, Zheng-Jun Zha, Di Chen, Richang Hong, and Meng Wang. Adaptive transfer network for cross-domain person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7202–7211, 2019.

[25] Chen Change Loy, Tao Xiang, and Shaogang Gong. Multi-camera activity correlation analysis. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1988–1995. IEEE, 2009.

[26] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE Confer-*

ence on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 0–0, 2019.

[27] 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 (TPAMI)*, 2019.

[28] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[29] Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, et al. Mixed precision training. *arXiv preprint arXiv:1710.03740*, 2017.

[30] Hyeonseob Nam and Hyo-Eun Kim. Batch-instance normalization for adaptively style-invariant neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2558–2567, 2018.

[31] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 464–479, 2018.

[32] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, 2018.

[33] Seonguk Seo, Yumin Suh, Dongwan Kim, Jongwoo Han, and Bohyung Han. Learning to optimize domain specific normalization for domain generalization. *arXiv preprint arXiv:1907.04275*, 2019.

[34] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. In *International Conference on Learning Representations (ICLR)*, 2018.

[35] Leslie N Smith. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 464–472. IEEE, 2017.

[36] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 719–728, 2019.

[37] Vladislav Sovrasov and Dmitry Sidnev. Building computationally efficient and well-generalizing person re-identification models with metric learning. *arXiv preprint arXiv:2003.07618*, 2020.

[38] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

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

[40] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[41] Guangcong Wang, Jian-Huang Lai, Wenqi Liang, and Guangrun Wang. Smoothing adversarial domain attack and p-memory reconsolidation for cross-domain person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10568–10577, 2020.

[42] Guangting Wang, Chong Luo, Xiaoyan Sun, Zhiwei Xiong, and Wenjun Zeng. Tracking by instance detection: A meta-learning approach. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6288–6297, 2020.

[43] Zhixiang Wang, Zheng Wang, Yinqiang Zheng, Yung-Yu Chuang, and Shin'ichi Satoh. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 618–626, 2019.

[44] Zheng Wei-Shi, Gong Shaogang, and Xiang Tao. Associating groups of people. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 23–1, 2009.

[45] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv preprint arXiv:1604.01850*, 2(2), 2016.

[46] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9021–9030, 2020.

[47] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 8222–8231, 2019.

[48] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Relation-aware global attention for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3186–3195, 2020.

[49] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124, 2015.

[50] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3754–3762, 2017.

[51] Jiahuan Zhou, Bing Su, and Ying Wu. Online joint multi-metric adaptation from frequent sharing-subset mining for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2909–2918, 2020.

[52] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *European Conference on Computer Vision (ECCV)*, pages 561–578. Springer, 2020.

[53] Kaiyang Zhou, Xiatian Zhu, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Learning generalisable omni-scale representations for person re-identification. *arXiv preprint arXiv:1910.06827*, 2019.

[54] Zijie Zhuang, Longhui Wei, Lingxi Xie, Tianyu Zhang, Hengheng Zhang, Haozhe Wu, Haizhou Ai, and Qi Tian. Rethinking the distribution gap of person re-identification with camera-based batch normalization. In *European Conference on Computer Vision (ECCV)*, pages 140–157. Springer, 2020.