A Novel Mix-normalization Method for Generalizable Multi-source Person Re-identification

Lei Qi, Lei Wang, Yinghuan Shi, Xin Geng*

Abstract—Person re-identification (Re-ID) has achieved great success in the supervised scenario. However, it is difficult to directly transfer the supervised model to arbitrary unseen domains due to the model overfitting to the seen source domains. In this paper, we aim to tackle the generalizable multi-source person Re-ID task (*i.e.*, there are multiple available source domains, and the testing domain is unseen during training) from the data augmentation perspective, thus we put forward a novel method, termed MixNorm, which consists of domain-aware mix-normalization (DMN) and domain-ware center regularization (DCR). Different from the conventional data augmentation, the proposed domain-aware mix-normalization to enhance the diversity of features during training from the normalization view of the neural network, which can effectively alleviate the model overfitting to the source domains, so as to boost the generalization capability of the model in the unseen domain. To better learn the domain-invariant model, we further develop the domain-aware center regularization to better map the produced diverse features into the same space. Extensive experiments on multiple benchmark datasets validate the effectiveness of the proposed method and show that the proposed method can outperform the state-of-the-art methods. Besides, further analysis also reveals the superiority of the proposed method.

Index Terms—MixNorm, domain-aware mix-normalization, generalizable multi-source person re-identification.

I. INTRODUCTION

IN recent years, person re-identification (Re-ID) has attracted an increasing interest in both academia and industry due to its great potentials in the video surveillance system [1], [2], [3]. Person Re-ID resorts to match images of the same person captured by different cameras with the non-overlapping camera views. This main challenge of person Re-ID is the variations including body pose, viewing angle, illumination, image resolution, occlusion, background and so on across different cameras [4], [5], [6]. In general, person Re-ID can be treated as a special case of the image retrieval problem with the goal of querying from a large-scale gallery set to quickly and accurately find images that match with a query image.

Currently, the typical person Re-ID methods have obtained good performance in the supervised setting due to the power of deep neural network [7], [8], [9], [10], [11], [12]. However, when these models are used to the unseen domain, they will fail on performance because of the model overfitting to the source domains. Generally, when using these methods in a new domain or scenario, we need to collect the data and give them labels, which is expensive and time-consuming, thus these existing supervised methods cannot be utilized on a large scale in the real-world application. Although some unsupervised domains adaptation (UDA) methods are proposed to mitigate the labeling task [13], [14], [15], [16], [17], they still require to collect data and re-train the model for the new scenario.

Domain generalization (DG) methods can address the above problem, which resorts to learning a model in the source domains and testing in the unseen domain [18]. In the person Re-ID community, some DG methods have been developed to produce a robust model in the unseen target domain. For example, the method in [19] treats image matching as finding local correspondences in feature maps and constructs query-adaptive convolution kernels on the fly to achieve local matching. In [20], a meta-learning strategy is introduced to simulate the train-test process of domain generalization for learning more generalizable models. The method [21] adopts an effective voting-based mixture mechanism to dynamically leverage the diverse characteristics of source domains to improve the generalization ability of the model. Differently, we aim to solve the issue from the data-augmentation perspective.

In this paper, we focus on the generalizable multi-source person re-identification task as in [20], where there are multiple available source domains in the training stage, and the testing data is unseen during training. Considering the conventional batch normalization uses the same statistics to normalize all samples in a batch, which could lead to learning the fixed pattern in the training process and further overfitting the trained model to the seen source domains. To address the issue, we propose a domain-aware mix-normalization (DMN) to achieve the feature augmentation, which can alliterative the overfitting problem via yielding the diverse features. To be specific, we randomly chose different domains to generate the mixed statistics (*i.e.*, mean and variance) in a batch when conducting the forward operation during training, and they are employed to normalize the samples in the selected domains. Furthermore, we also propose the domain-aware center regularization (DCR) to better train the model with the diverse features, which can effectively map all features into the same space, so that the domain-invariant feature can be obtained. Together, the above two modules give rise to a novel "MixNorm" method for generalizable multi-source person Re-ID. We conduct the experiments on multiple benchmark person Re-ID datasets to confirm the effectiveness of the proposed method. Besides, the deep analysis by extensive experiments

Lei Qi and Xin Geng are with the School of Computer Science and Engineering, and the Key Lab of Computer Network and Information Integration (Ministry of Education), Southeast University, Nanjing, China, 211189 (e-mail: qilei@seu.edu.cn; xgeng@seu.edu.cn).

Lei Wang is School of Computing and Information Technology, University of Wollongong, Wollongong, Australia (e-mail: leiw@uow.edu.au).

Yinghuan Shi is with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China, 210023 (e-mail: syh@nju.edu.cn).

* Corresponding author: Xin Geng.

reveals the superiority of MixNorm

In MixNorm, our main contributions can be summarized as:

- We propose a simple yet effective domain-aware mix-normalization, which can produce diverse features to enhance the robustness and effectiveness of the model in unseen domains using different statistics for normalizing different samples in a batch.
- We design a domain-aware center regularization to better leverage our proposed mix-normalization during the training course, which can guarantee that all samples from different domains are mapped into the same space.
- We evaluate our approach on multiple standard benchmark datasets, and the results show that our approach outperforms the state-of-the-art accuracy. Moreover, the ablation study and further analysis are provided to validate the efficacy of our method.

The rest of this paper is organized as follows. We review some related work in Section II. The proposed method is introduced in Section III. Experimental results and analysis are presented in Section IV, and Section V is conclusion.

II. RELATED WORK

In this section, we review the most related works to our work, including the generalizable person re-identification and the domain generalization methods. The detailed investigation is presented in the below part.

A. Generalizable Person Re-identification

Currently, several methods have been developed to solve the generalizable (*i.e.*, domain generalization) person R-ID task, including the instance normalization based methods, the meta-learning based methods and the sample-adaptive methods.

Considering that the domain discrepancy in Re-ID is due to style and content variance across datasets, and instance normalization (IN) [22] can effectively remove the style information, Jia *et al.* [23] adopt instance and feature normalization to alleviate much of the resulting domain-shift in Deep Re-ID models. However, IN inevitably removes discriminative information, while it filters out style variations, such as illumination and color contrast. Therefore, Jin *et al.* [24] propose to distill identity-relevant feature from the removed information and restitute it to the network to ensure high discrimination. For better disentanglement, the method enforces a dual causality loss constraint in SNR to encourage the separation of identity-relevant features and identity-irrelevant features.

Besides, some methods employ the meta-learning scheme to enhance the generalization capability to the unseen domains. In [25], a novel deep Re-ID model termed Domain-Invariant Mapping Network (DIMN) is proposed, which utilizes the meta-learning pipeline to sample a subset of source domain training tasks during each training episode, so as to make the model domain-invariant. Choi *et al.* [26] combine learnable batch-instance normalization layers with meta-learning and investigate the challenging cases caused by both batch and instance normalization layers. Moreover, the method diversifies the virtual simulations via our meta-train loss accompanied by a cyclic inner-updating manner to boost the

model's generalization. Zhao *et al.* [20] design the Memory-based Multi-Source MetaLearning (M^{3}L) framework to train a generalizable model for unseen domains. In this framework, a meta-learning strategy is introduced to simulate the train-test process of domain generalization for learning more generalizable models. Moreover, this method also presents a meta batch normalization layer to diversify meta-test features, further establishing the advantage of meta-learning.

Different from the aforementioned methods, some adaptive approaches are also designed to perform well on the unseen target domains, which can adjust the parameters or module according to each testing image in the testing stage. For example, Dai *et al.* [21] introduce a novel method called the relevance-aware mixture of experts (RaMoE), using an effective voting-based mixture mechanism to dynamically leverage the diverse characteristics of source domains to improve the generalization ability of the model. Particularly, in the testing stage, the method can change the weights of each experts according to the current testing image. In [19], the method treats image matching as finding local correspondences in feature maps, and constructs query-adaptive convolution kernels on the fly to achieve local matching. In this way, the matching process and results are interpretable, and this explicit matching is more generalizable than representation features to unseen domains, such as unknown misalignments, pose or viewpoint changes.

In this paper, unlike the above methods for generalizable person Re-ID, we resort to improving the generalization capability of the model from the data augmentation perspective when there are multiple available source domains during training, as the setting in [20].

B. Domain Generalization

Recently, some methods are also proposed to address the domain generalization problem in the classification and semantic segmentation tasks [27], [28], [29], [30], [31], [32], [33], [34]. Particularly, the data augmentation based methods have obtained significant advance in these task, which can be divided into image-level and feature-level data augmentation.

For the image-level augmentation, Zhou *et al.* [35] employ a data generator to synthesize data from pseudo-novel domains to augment the source domains. Moreover, in [36], a novel DG approach based on Deep Domain-Adversarial Image Generation (DDAIG) is proposed, which aims to map the source training data to unseen domains. This is achieved by having a learning objective formulated to ensure that the generated data can be correctly classified by the label classifier while fooling the domain classifier. Particularly, augmenting the source training data with the generated unseen domain data can make the label classifier more robust to unknown domain changes. Differently, considering the Fourier phase information contains high-level semantics and is not easily affected by domain shifts, Xu *et al.* [37] develop a novel Fourier-based data augmentation strategy to force the model to capture phase information, which linearly interpolates between the amplitude spectrums of two images.

For the feature-level augmentation, Huang *et al.* [38] introduce a simple training heuristic, called Representation Self-

Challenging (RSC), that significantly improves the generalization of CNN to the out-of-domain data. RSC iteratively challenges (discards) the dominant features activated on the training data, and forces the network to activate remaining features that correlate with labels. Li *et al.* [39] propose an extremely simple technique of perturbing the feature embedding with Gaussian noise during training leads to a classifier with domain-generalization performance comparable to existing state-of-the-art. Besides, in [40], a novel approach is proposed based on probabilistically mixing instance-level feature statistics of training samples across source domains, termed MixStyle. Mixing styles of training instances results in novel domains being synthesized implicitly, which increases the domain diversity of the source domains, and hence improves the generalization ability of the trained model.

In this paper, we solve the generalizable person Re-ID task from the data augmentation perspective. Particularly, our proposed method belongs to the feature-level augmentation. Differently, we achieve to generate the diverse features from the normalization view of the neural network, which is not well investigated in the existing works.

III. THE PROPOSED METHOD

In this paper, to enhance the generalization capacity of the model to the unseen domain, we put forward a novel mix-moralization method termed MixNorm, as illustrated in Fig. 1. Specifically, in this method, we develop domain-aware mix-normalization (DMN) to conduct the feature augmentation, which randomly combines different domains to compute the statistics during each normalization. Thus, perturbed features can be generated to promote the diversity of the training samples. Furthermore, to better train the model with the rich features, we put forward the domain-aware center regularization (DCR) to map all samples into the same space, which can effectively enforce the model to learn the domain-invariant features. The detailed description of the proposed method will be presented in the following part.

A. Domain-aware Mix-normalization

In this part, we will introduce the domain-aware mix-normalization (DMN) for generalizable multi-source person re-identification. During training, we randomly select the same number of samples from each source domain to form each batch. For example, we have D source domains and choose P samples from each domain, thus the batch size is $N = D \times P$ in the training stage. In general, the conventional batch normalization (BN) [41] is leveraged to moralize all samples in each batch, which can be defined as:

$$\text{BN}(f_d) = \gamma \frac{f_d - \mu}{\sigma} + \beta, \quad d \in \{1, ..., D\}, \quad (1)$$

where $f_d \in \mathbb{R}^{N \times C \times H \times W}$ denotes the feature map from the d-th domain, $\gamma, \beta \in \mathbb{R}^C$ are learnable affine transformation parameters, and C is the number of channel on the corresponding feature maps. $\mu, \sigma \in \mathbb{R}^C$ represent the channel-wise mean and standard deviation of BN of each feature map (*i.e.*,

statistics) as follows:

$$\mu = \frac{1}{DPHW} \sum_{d=1}^{D} \sum_{p=1}^{P} \sum_{h=1}^{H} \sum_{w=1}^{W} f_d[d \times p, :, h, w], \quad (2)$$

$$\sigma = \sqrt{\frac{1}{DPHW} \sum_{d=1}^{D} \sum_{p=1}^{P} \sum_{h=1}^{H} \sum_{w=1}^{W} (f_d[d \times p, :, h, w] - \mu)^2 + \epsilon}, \quad (3)$$

where ϵ is a constant for numerical stability. As seen in the above equation, in each batch, BN utilizes the same statistics (*i.e.*, a pair of μ and σ) for all samples, which could result in that the fixed pattern is learned by neural network, so that the final model overfits to the source domains. Most existing methods address the issue in person Re-ID community by the instance normalization based methods, the meta-learning based methods or the sample-adaptive methods.

Differently, we propose to tackle this problem from the data augmentation view based on the normalization of the neural network. Considering that using the same mean and variance for all samples in each batch could capture the fixed pattern, we propose a domain-aware mix-normalization for performing the normalization, which randomly combines the different domains to yield the diverse statistics for the samples from different domains. Therefore, normalizing the features using diverse statistics can achieve the function of the feature-level data augmentation. The domain-aware mix-normalization (DMN) can be formulated by:

$$\text{DMN}(f_d) = \gamma \frac{f_d - M_\mu(d)}{M_\sigma(d)} + \beta, \quad d \in \{1, ..., D\}, \quad (4)$$

where $M_\mu(d)$ and $M_\sigma(d)$ represent the statistics for the d-th domain. For the DMN, we firstly produce a random integer C, which is constrained between 1 and $D-1$ in our method. Then the C domains are randomly selected to compute the statistics, and they are adopted to perform the moralization for all samples from the C selected domains. Here we use the ϕ as the set of selected domains. Removing the chosen domains and continue to randomly select C domains to calculate the statistics in the residual domains. If the number of the residual domains is less than C, we will leverage all remaining domains to compute the statistics for normalizing themselves. Specifically, $M_\mu(d)$ and $M_\sigma(d)$ can be described as:

$$M_\mu(d) = \frac{1}{CPHW} \sum_{d=1}^{C} \sum_{p=1}^{P} \sum_{h=1}^{H} \sum_{w=1}^{W} f_{d \in \phi}[c \times p, :, h, w], \quad (5)$$

$$M_\sigma(d) =$$
$$\sqrt{\frac{1}{CPHW} \sum_{c=1}^{C} \sum_{p=1}^{P} \sum_{h=1}^{H} \sum_{w=1}^{W} (f_{d \in \phi}[c \times p, :, h, w] - M_\mu(d))^2 + \epsilon}. \quad (6)$$

The forward process of the proposed DMN is given in Algorithm 1. Moreover, it is worth noting that there is only a pair of statistics for evaluation in each DMN during testing, *i.e.*, the statistics of all random combinations are accumulated to a
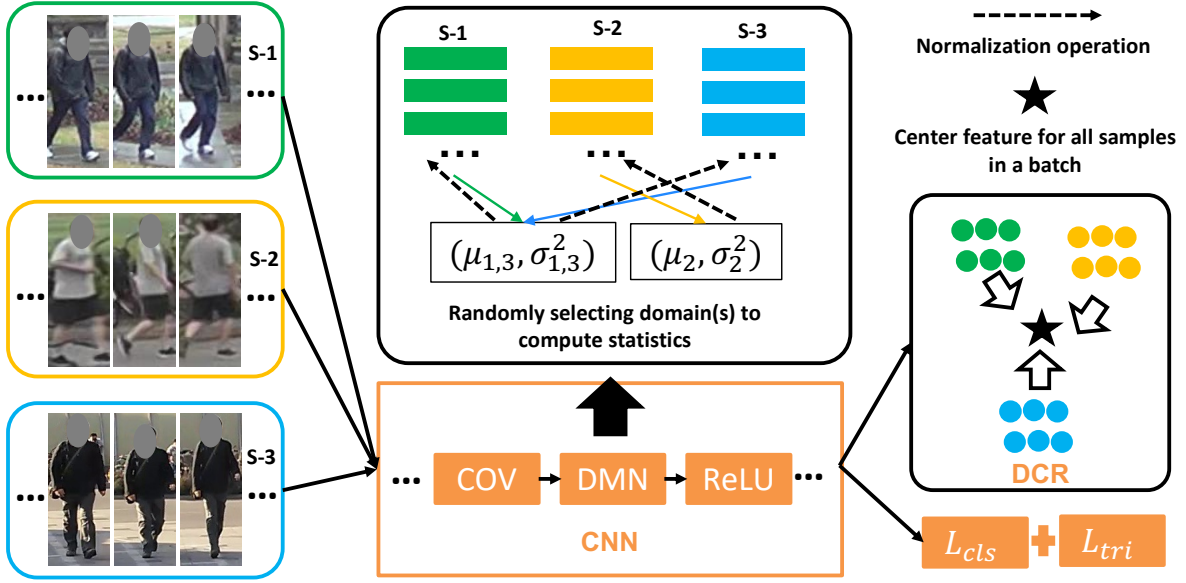
Fig. 1. An illustration of the proposed MixNorm. Here we take three source domains as an example. As seen in this figure, the method is composed of domain-aware mix-normalization (DMN) and domain-aware center regularization (DCR). In this figure, we show that source-1 and source-3 are randomly selected to calculate statistics for normalization together, and the remaining source-2 independently conducts the normalization operation. Best viewed in color.

pair of the mean and variance used for the evaluation, as the operation in BN [41].

Remark. For the proposed DMN, we use it by replacing all the original BN layers in the whole network. Particularly, since each DMN is independent, the randomization appears not only in each DMN but also inter the different DMNs. This way can bring more diversity of features during training, so as to promote the generalization to the unseen domain, as validated in the further analysis of the experimental section.

Besides, when using the DMN to train the model, we set the maximum numbers of the randomly selected domains as $D - 1$. If it is set as D (*i.e.*, using the mean and variance of all domains to normalize features during training), it could cause slight overfitting to the source domains, because the final statistics for testing are the expectation of all domains. The related results are given in the experiment.

Moreover, considering that the DMN requires the uniform sampling for each domain during the training course, *i.e.*, there is the same number of samples for each domain in a batch, which is short of the randomization when compared to the random sampling, *i.e.*, we combine all samples from all domains to a dataset and then randomly sample from the combined dataset to generate each batch. Particularly, although the random sampling scheme can bring the randomization across different batches, it still utilizes the global statistics to normalize all samples in each batch. Differently, our method randomly produces the local statistics in each batch to normalize the different samples, thus it could better prevent the model from learning the fixed pattern, so as to alleviate the overfitting to the seen source domains. In the experimental section, we also verify that our method can produce a more robust model than the random sampling scheme in the generalizable multi-source person Re-ID task.

Algorithm 1 The forward process of mix-normalization (DMN)

1: **Input:** Feature f of all $D \times P$ samples in a batch and the initial domain set **S** including all D source domains.
2: **Output:** The normalized feature $\hat{f}$.
3: Randomly produce the number of mixing domains as C.
4: **while** $|\mathbf{S}| \neq 0$ **do**
5:   **if** $|\mathbf{S}| =< C$ **then**
6:     Set C as $|\mathbf{S}|$.
7:   **end if**
8:   Randomly select C domains ϕ from the **S**.
9:   Compute the statistics (*i.e.*, mean and variance) for the selected C domains as Eqs. 5 and 6.
10:   Utilize the statistics to normalize all samples from the selected C domains as Eq. 4.
11:   Remove the selected C domains from the **S**.
12: **end while**

B. Domain-aware Center Regularization

The proposed domain-aware mix-normalization can enhance the diversity of features during training, thus we further propose domain-aware center regularization (DCR) to better learn the domain-invariant model using the diverse features. Generally, the Maximum Mean Discrepancy (MMD) [42] can be used to reduce domain shift between two domains. In our task, there are available multiple domains in the training process, thus the MMD cannot be directly employed to address our issue. Therefore, we develop the domain-aware center regularization to reduce the domain gap across different domains, which can be defined as:

$$L_{dcr} = \sum_{d=1}^{D} \sum_{p=1}^{P} \|r[d \times p, :] - \bar{r}\|_2^2, \qquad (7)$$

where $r \in \mathbb{R}^{N \times B}$ is the feature representation of all samples in a batch, and B indicates the dimension of the the features. $\bar{r} \in \mathbb{R}^{1 \times B}$ denotes the mean of all features.

Remark. It is worth noting that the domain-aware center regularization is merely effective when the domain-aware mix-normalization is used. This main reason is that the original model is easy to train due to the single pattern (*i.e.*, using the same mean and variance for all samples in a batch) in the training stage. Since using the domain-aware mix-normalization introduces the diverse features, thus it could need a strong regularization to map all features into the same space, so as to learn the domain-invariant model. We will report the results in the experimental section.

C. The Overall Loss

During training, we employ the cross-entropy loss (*i.e.*, L_{cls}), the triplet loss (*i.e.*, L_{tri}) with hard mining sampling [22] and the proposed domain-aware center regularization (*i.e.*, L_{dcr}) to train the model. Particularly, the cross-entropy loss and the triplet loss are the basic loss in the person Re-ID community [43], [44]. The overall loss for training the model can be described as:

$$L_{overall} = L_{cls} + L_{tri} + \lambda L_{dcr}, \tag{8}$$

where λ is the hyper-parameter to trade off the basic loss and the domain-aware center regularization.

IV. EXPERIMENTS

In this part, we firstly introduce the experimental datasets and settings in Section IV-A. Then, we compare the proposed method with the state-of-the-art generalizable Re-ID methods in Section IV-B, respectively. To validate the effectiveness of various components in the proposed framework, we conduct ablation studies in Section IV-C. Lastly, we further analyze the property of the proposed method in Section IV-D.

A. Datasets and Experimental Settings

1) Datasets: We evaluate our approach on four large-scale image datasets: Market1501 [45], DukeMTMC-reID [46], MSMT17 [47] and CUHK03-NP [48], [49]. **Market1501 (Ma)** contains 1,501 persons with 32,668 images from six cameras. Among them, $12,936$ images of 751 identities are used as a training set. For evaluation, there are $3,368$ and $19,732$ images in the query set and the gallery set, respectively. **DukeMTMC-reID (D)** has $1,404$ persons from eight cameras, with $16,522$ training images, $2,228$ queries, and $17,661$ gallery images. **MSMT17 (Ms)** is collected from a 15-camera network deployed on campus. The training set contains $32,621$ images of $1,041$ identities. For evaluation, $11,659$ and $82,161$ images are used as query and gallery images, respectively. **CUHK03-NP (C)** has an average of 4.8 images in each camera. The dataset provides both manually labeled bounding boxes and DPM-detected bounding boxes. On this dataset, there are $7,365$ training images, and $1,400$ images and $5,332$ images in query set and gallery set are used in the testing stage. Particularly, we divide these four datasets

into two parts: three domains as source domains for training and the other one as target domain for testing. We adopt the recommended setting in [20].

In addition, we also train our method on more datasets including Market1501, DukeMTMC-reID, CUHK03, CUHK02 [50] and CUHK-SYSU [51], and test on two small-scale datasets including PRID [52] and VIPeR [53]. Particularly, the performances of these small ReID datasets are evaluated on the average of 10 repeated random splits of gallery and probe sets. For all datasets, we employ CMC (*i.e.*, Cumulative Match Characteristic) accuracy and mAP (*i.e.*, mean Average Precision) for Re-ID evaluation [45].

2) Implementation Details: In this experiment, we use the pre-trained ResNet-50 [54] and IBN-Net50 [55] on ImageNet [56] to initialize the network parameters. In a batch, the number of IDs and the number of images per person are set as 16 and 4 to produce triplets for each domain, respectively. The initial learning rate is 3.5×10^{-4} and divided by 10 at the 30-th and 50-th epochs, respectively. The proposed model is trained with the Adam optimizer in a total of 60 epochs. The size of the input image is 256×128. For data augmentation, we perform random cropping, random flipping and auto-argumentation [57]. Besides, λ in Eq. 8 is set as 0.2. Particularly, we utilize the same setting for all experiments on all datasets in this paper.

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

We compare our proposed method (*i.e.*, MixNorm) with some state-of-the-art methods as reported in Table I, including IBN-Net50 [55], QAConv$_{50}$ [19], and M^{3}L [20]. IBN-Net50 carefully integrates Instance Normalization (IN) and Batch Normalization (BN) as building blocks, where IN learns features that are invariant to appearance changes, such as colors and styles, while BN is essential for preserving content-related information. This method remarkably enhances a CNN's modeling ability on one domain as well as its generalization capacity on another domain without finetuning. QAConv$_{50}$ treats image matching as finding local correspondences in feature maps, and constructs query-adaptive convolution kernels on the fly to achieve local matching. M^{3}L utilizes a meta-learning strategy to simulate the train-test process of domain generalization for learning more generalizable models. As seen in Table I, our MixNorm outperforms all other methods on Rank-1 and mAP. For example, in the "D+C+Ms→Ma" task, our method increases M^{3}L by $+2.2\%$ (54.7 vs. 52.5) and $+3.1\%$ (81.4 vs. 78.3) on mAP and Rank-1 when using IBN-Net50 as the backbone. This mainly owes to the effectiveness of the domain-aware mix-normalization and domain-aware center regularization, which can bring diverse features to prevent the model's overfitting to the source domains and help to better learn the domain-invariant model. Besides, although the performance of our method is slightly poorer than M^{3}L when ResNet-50 is used as the backbone in the "Ma+D+Ms→C" task, our method has better results than M^{3}L based on IBN-Net50. Particularly, in this task, each component of our method is also effective, as shown in the below ablation study.

In addition, our method is also compared with some other methods on small-scale datasets when using more datasets to

TABLE I
EXPERIMENTAL RESULTS ON FOUR LARGE-SCALE DATASETS, INCLUDING
MARKET1501 (MA), DUKEMTMC-REID (D), MSMT17 (MS), AND
CUHK03 (C). "D+C+MS→MA" INDICATES THAT THE MODEL IS
TRAINED ON D, C AND MS, AND TESTED ON MA. THE BOLD DENOTES
THE BEST RESULT.

Method	D+C+Ms→Ma		Ma+C+Ms→D	
	mAP	Rank-1	mAP	Rank-1
IBN-Net50 [55]	43.0	73.4	45.7	64.9
QAConv$_{50}$ [19]	39.5	68.6	43.4	64.9
M^{3}L(ResNet-50) [20]	51.1	76.5	48.2	67.1
M^{3}L(IBN-Net50) [20]	52.5	78.3	48.8	67.2
MixNorm (ResNet-50)	51.4	78.9	49.9	**70.8**
MixNorm (IBN-Net50)	**54.7**	**81.4**	**52.3**	70.6
Method	Ma+D+C→Ms		Ma+D+Ms→C	
	mAP	Rank-1	mAP	Rank-1
IBN-Net50 [55]	17.0	43.9	21.3	21.6
QAConv$_{50}$ [19]	10.0	29.9	19.2	22.9
M^{3}L(ResNet-50) [20]	13.1	32.0	30.9	31.9
M^{3}L(IBN-Net50) [20]	15.4	37.1	31.4	31.6
MixNorm (ResNet-50)	19.4	47.2	29.0	29.6
MixNorm (IBN-Net50)	**23.1**	**52.2**	**32.3**	**32.4**

TABLE II
EXPERIMENTAL RESULTS ON TWO SMALL-SCALE DATASETS. THE MODEL
IS TRAINED ON FIVE DATASETS INCLUDING MARKET1501 (MA),
DUKEMTMC-REID, CUHK03, CUHK02 AND CUHK-SYSU.

Method	PRID		VIPeR	
	mAP	Rank-1	mAP	Rank-1
Agg Align [58]	25.5	17.2	52.9	42.8
Reptile [59]	26.9	17.9	31.3	22.1
CrossGrad [60]	28.2	18.8	30.4	20.9
Agg PCB [61]	32.0	21.5	45.4	38.1
MLDG [62]	35.4	24.0	33.5	23.5
PPA [63]	45.3	31.9	54.5	45.1
DIMN [25]	52.0	39.2	60.1	51.2
SNR [24]	66.5	52.1	61.3	52.9
RaMoE [21]	67.3	57.7	64.6	**56.6**
ResNet-50	47.3	37.2	49.8	40.2
MixNorm (ResNet-50)	59.1	49.2	60.6	50.8
IBN-Net50	52.7	42.6	57.9	47.7
MixNorm (IBN-Net50)	**74.3**	**65.2**	**66.6**	56.4

train the model, and these methods include Agg Align [58], Reptile [59], CrossGrad [60], Agg PCB [61], MLDG [62], PPA [63], DIMN [25], SNR [24] and RaMoE [21]. The experimental results are given in Table II. For example, DIMN [25] learns a mapping between a person image and its identity classifier and follows a meta-learning pipeline and sample a subset of source domain training tasks during each training episode SNR [24] distills identity-relevant features from the removed information conducted by instance normalization and restitute it to the network to ensure high discrimination. RaMoE [21] adopts an effective voting-based mixture mechanism to dynamically leverage the diverse characteristics of source domains to improve the model's generalization. As seen in Table II, our method can basically obtain better performance than other methods when using the IBN-Net50 as the backbone. For example, when compared to SNR, our method improve +7.8% (74.3 vs. 66.5) on mAP. It is worth noting that, SNR uses the IN to remove the style information, thus this is a fair comparison when we use IBN-Net50 as the backbone. Moreover, compared to the baseline, using our method can achieve a great improvement. For example, on PRID, two baselines (*i.e.*, IBN-Net50 and ResNet50) can be gained by +21.6% (74.3 vs. 52.7) and +11.8% (59.1 vs. 47.3) on mAP, thus this validates the efficacy of the proposed method when there are more source domains during the training course.

C. Ablation Studies

In this part, we conduct the ablation study to verify the effectiveness of each component in our method. Our method mainly consists of domain-aware mix-normalization (DMN) and domain-aware center regularization (DCR). Table III shows the experimental results when we employ ResNet-50 as the backbone. As seen, when adding the DMN into the baseline, the baseline can be gained in multiple different tasks. For example, in the "D+C+Ms→Ma" task, using the DMN can increase the baseline by +8.1% (48.5 vs. 40.4) and +7.3% (76.9 vs. 69.6) on mAP and Rank-1. This shows that using the

DMN to enrich the diversity of the feature can indeed alleviate the model's overfitting and boost the generalization capability of the model to the unseen domain. Moreover, using DCR can further improve the model's robustness. For example, in the "D+C+Ms→Ma" and "Ma+D+Ms→C" tasks, adding the DCR can bring the +2.9% (51.4 vs. 48.5) and +2.3% (29.0 vs. 26.7) improvements on mAP, thus it confirms the effectiveness of DCR in our method, which can better utilize the diverse features from the DMN to learn the domain-invariant model.

TABLE III
ABLATION STUDY WHEN RESNET-50 IS UTILIZED AS THE BACKBONE.

Module	mAP	Rank-1	Rank-5	Rank-10
	D+C+Ms→Ma			
Baseline	40.4	69.6	83.8	87.9
Baseline+DMN	48.5	76.9	88.8	92.1
Baseline+DMN+DCR	**51.4**	**78.9**	**90.2**	**93.3**
	Ma+C+Ms→D			
Baseline	44.2	64.0	77.0	81.5
Baseline+DMN	48.4	67.9	80.4	83.7
Baseline+DMN+DCR	**49.9**	**70.8**	**81.7**	**85.0**
	Ma+D+Ms→C			
Baseline	24.0	24.4	41.7	51.4
Baseline+DMN	26.7	26.9	45.5	56.2
Baseline+DMN+DCR	**29.0**	**29.6**	**47.2**	**58.1**

Besides, we also validate the efficacy of each component on IBN-Net50, as reported in Fig. 2. As seen in all reported tasks, "Baseline+DMN" has better results than the "Baseline". Moreover, adding the DCR (*i.e.*, "Baseline+DMN+DCR") can further improve the generalization capability to the unseen domains. For example, in the "Ma+D+Ms→C" task, DMN can increase the baseline by +8.4% (29.7 vs. 21.3), while using the DCR can further gain +2.6% (32.3 vs. 29.7) on mAP. This experiment deeply confirms the value of each component in our method.

D. Further Analysis

Hyper-parameter analysis. We perform the experiment to evaluate the sensitivity of the hyper-parameter λ in Eq. 8. The experimental results are shown in Fig. 3. As seen, when λ is set as 0.2, the best performance can be obtained. If using the smaller λ, the diverse features from the domain-aware
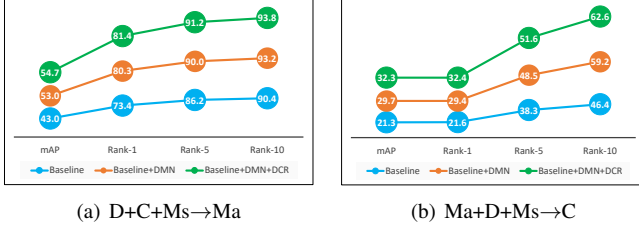
(a) D+C+Ms→Ma   (b) Ma+D+Ms→C

Fig. 2. Ablation study when IBN-Net50 is used as the backbone.
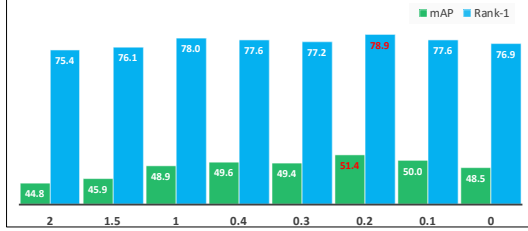
Fig. 3. The impact of λ of Eq. 8 in the "D+C+Ms→Ma" task. Note that the red is the best result.
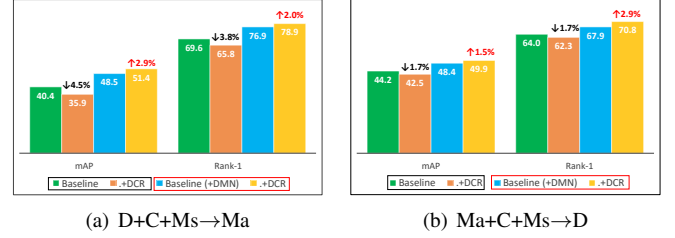

(a) D+C+Ms→Ma   (b) Ma+C+Ms→D

Fig. 4. The experimental results of domain-aware center regularization (DCR) used on different baselines in two tasks. Best viewed in color.

TABLE IV
COMPARISON BETWEEN DCR AND CENTER LOSS (CL) [64].

Method	mAP	Rank-1	Rank-5	Rank-10
	D+C+Ms→Ma			
CL [64]	49.2	77.2	88.8	92.4
DCR (ours)	**51.4**	**78.9**	**90.2**	**93.3**
	Ma+C+Ms→D			
CL [64]	47.7	67.7	79.6	83.7
DCR (ours)	**49.9**	**70.8**	**81.7**	**85.0**
	Ma+D+Ms→C			
CL [64]	26.9	27.1	44.8	55.4
DCR (ours)	**29.0**	**29.6**	**47.2**	**58.1**

mix-normalization could not be sufficiently exploited. If we use the larger λ, the model could more focus on reducing the domain gap across different source domains, thus it might slightly destroy the discriminative information of the learned features. According to the results in Fig. 3, we set λ as 0.2 in all our experiments.

More experiments on DCR. Domain-aware center regularization (DCR) is developed to better train the model when the DMN module produces diverse features. Here we also adopt it to the original baseline (*i.e.*, the raw ResNet50), and the experimental results are displayed in Fig. 4. As seen, if we directly use DCR on the original baseline, the performance will obviously decrease. For example, in the "D+C+Ms→Ma" task, when leveraging the DCR on the baseline, the result will drop by -4.5% (35.9 vs. 40.4) on mAP. Considering the pattern in the original baseline is fixed, which does not need to focus on the domain gap, using the DCR could have a negative on the learning of the discriminative features. On the contrary, using the DCR on the baseline with DMN will obtain an obvious improvement.

Comparison between DCR and center loss. The center loss (CL) in [64] simultaneously learns a center for deep features of each class and penalizes the distances between the deep features and their corresponding class centers, which can be also utilized to reduce the domain gap across different domains in the generalizable person Re-ID. In this experiment, we replace the domain-aware center regularization (DCR) by the center loss in our method. Similar to [21], we set λ in Eq. 8 as 5×10^{-4}. The experimental results are reported in Table IV. As seen, using our DCR can obtain better performance than CL. For example, in the "Ma+C+Ms→D" task, our DCR exceeds CL by $+2.2\%$ (49.9 vs. 47.7) on mAP. This mainly thanks to that , our method aims to pull each domain close to the center of all domains in each batch, thus it can better mitigate the domain gap across different domains.

The effectiveness of the different domain combinations

(DDC) for each DMN at each iteration. During training, each proposed domain-aware mix-normalization (DMN) independently conducts the normalization operation, thus the domain's combination in Eqs. 5 and 6 is different, *i.e.*, the randomly selected domain set ϕ could be different in each DMN. Here we also conduct the experiment when using the same domain combination for all DMN at each iteration, as reported in Table V. As seen, the performance of using the same domain combination for all DMN at each iteration will decrease in general. For example, in the "Ma+D+Ms→C" task, the result will drop by -2.1% (26.9 vs. 29.0) and -1.5% (28.1 vs. 29.6) on mAP and Rank-1. This main reason is that the randomization appears not only in each DMN but also inter the different DMNs when we adopt the independent DMN. Therefore, using the DDC scheme can further boost the diversity of features, so that it can enhance the generalization capability of the model to the unseen domain.

TABLE V
THE EXPERIMENTAL RESULTS OF THE DIFFERENT COMBINATIONS (DDC) AND THE SAME DOMAIN COMBINATION FOR EACH DMN AT EACH ITERATION DURING TRAINING

Using DDC (Y/N)	mAP	Rank-1	Rank-5	Rank-10
	D+C+Ms→Ma			
✗	49.2	77.6	88.4	92.2
✓	**51.4**	**78.9**	**90.2**	**93.3**
	Ma+C+Ms→D			
✗	49.2	68.2	81.0	83.9
✓	**49.9**	**70.8**	**81.7**	**85.0**
	Ma+D+C→Ms			
✗	19.1	46.3	**60.2**	**65.8**
✓	**19.4**	**47.2**	60.0	65.4
	Ma+D+Ms→C			
✗	26.9	28.1	45.1	54.3
✓	**29.0**	**29.6**	**47.2**	**58.1**

Analysis of the number of randomly selected domains in DMN. In our experiment, we observe that, if the maximum

number of randomly selected domains in domain-aware mix-normalization (DMN) is equal to the number of all source domains, the performance will decrease, as shown in Table VI. As observed, when setting the maximum domain number as D, the mAP and Rank-1,5,10 will slightly drop in the "Ma+C+Ms→D" task. This could be because the final statistics for testing accumulate the statistics of all domains, thus if using the statistics of all domains to normalize the features in the training stage, the model could be mildly overfitting to source domains. Therefore, in all experiments, we set the maximum number of randomly selected domains as $D - 1$.

TABLE VI
EXPERIMENTAL RESULTS WHEN THE MAXIMUM NUMBER OF RANDOMLY SELECTED DOMAINS (MNRSD) IN DMN IS EQUAL TO THE NUMBER OF ALL SOURCE DOMAINS. NOTE THAT D IS THE TOTAL NUMBER OF SOURCE DOMAINS.

MNRSD	mAP	Rank-1	Rank-5	Rank-10
	D+C+Ms→Ma			
D	49.9	77.9	89.7	92.0
$D-1$	**51.4**	**78.9**	**90.2**	**93.3**
	Ma+C+Ms→D			
D	49.3	69.1	80.6	84.9
$D-1$	**49.9**	**70.8**	**81.7**	**85.0**

Analysis of different sampling schemes. Our method is based on the uniform sampling (US) scheme, *i.e.*, we randomly select the same number of samples for all domains to generate each batch in the training stage. As mentioned in Section III, it could result in the model's overfitting to source domains due to the fixed pattern of the statistics. To further confirm the effectiveness of our method, we also perform the random sampling (RS) scheme during training, *i.e.*, we first combine all domains into a set, and then randomly select IDs and samples to generate a batch. Particularly, the batch size of RS is the same as US, and all experimental setting is also consistent. The results are reported in Table VII. As seen, compared with "US(Baseline)", "RS(Baseline)" can yield better results due to its randomness across different batches. However, the randomness is limited because the random sampling scheme still utilizes the same statistics to normalize all samples in a batch. On the contrary, our method randomly combines different domains to generate the local statistics to normalize the samples in each batch, which can produce diverse features to reduce the model's overfitting and enhance the generalization capability of the model to the unseen domain. As shown in Table VII, our method (*i.e.*, "US(MixNorm)") significantly outperforms the random sampling scheme in all tasks. Therefore, this experiment further verifies the effectiveness of our method.

Comparison between our MixNorm and MixStyle. MixStyle [40] is developed to tackle the domain generalization issue from the data augmentation in the feature-level, which mixes the style of different images to generate the style to produce the diverse features. We conduct a single experiment to compare it with our method, as displayed in Table VIII. In this experiment, we use the MixStyle with domain information to mix different styles, *i.e.*, mixing samples from different domains to form the new style, and insert the module in the back of Block-1,2,3 of ResNet50. According to the experi-

TABLE VII
EXPERIMENTAL RESULTS OF DIFFERENT SAMPLING METHODS. NOTE THAT RS AND US INDICATE THE RANDOM SAMPLING AND UNIFORM SAMPLING, RESPECTIVELY.

Method	mAP	Rank-1	Rank-5	Rank-10
	D+C+Ms→Ma			
RS (Baseline)	44.5	72.9	85.8	89.6
US (Baseline)	40.4	69.6	83.8	87.9
US (MixNorm)	**51.4**	**78.9**	**90.2**	**93.3**
	Ma+C+Ms→D			
RS (Baseline)	46.1	65.2	79.1	83.0
US (Baseline)	44.2	64.0	77.0	81.5
US (MixNorm)	**49.9**	**70.8**	**81.7**	**85.0**
	Ma+D+C→Ms			
RS (Baseline)	16.1	41.2	54.9	61.2
US (Baseline)	14.7	39.6	53.9	59.5
US (MixNorm)	**19.4**	**47.2**	**60.0**	**65.4**
	Ma+D+Ms→C			
RS (Baseline)	26.0	26.2	43.6	54.1
US (Baseline)	24.0	24.4	41.7	51.4
US (MixNorm)	**29.0**	**29.6**	**47.2**	**58.1**

mental results in Table VIII, we observe that our MixNorm has better performance than MixStyle in all tasks, *e.g.*, in the "D+C+Ms→Ma" task, MixNorm exceeds the MixStyle by $+9.1\%$ (51.4 vs. 42.3) on mAP. This mainly owes to the effectiveness of domain-aware mix-moralization and domain-aware center regularization in our MixNorm.

TABLE VIII
COMPARISON BETWEEN OUR MIXNORM AND MIXSTYLE [40].

Method	mAP	Rank-1	Rank-5	Rank-10
	D+C+Ms→Ma			
MixStyle [40]	42.3	71.5	84.9	89.1
MixNorm (ours)	**51.4**	**78.9**	**90.2**	**93.3**
	Ma+C+Ms→D			
MixStyle [40]	44.8	65.1	78.5	81.6
MixNorm (ours)	**49.9**	**70.8**	**81.7**	**85.0**
	Ma+D+C→Ms			
MixStyle [40]	16.4	42.2	55.9	61.7
MixNorm (ours)	**19.4**	**47.2**	**60.0**	**65.4**
	Ma+D+Ms→C			
MixStyle [40]	24.8	25.1	42.9	53.1
MixNorm (ours)	**29.0**	**29.6**	**47.2**	**58.1**

Evaluation on source domains. In this part, to validate the effectiveness of our method on alleviating the overfitting issue, we compare our method with the baseline on the source domains, as shown in Table IX. As seen, our method has poorer performance than the baseline on all source domains of all tasks. Moreover, we combine the above analysis and experiment (*i.e.*, our method can significantly exceed the baseline on the unseen domain in all tasks), thus this confirms the proposed method can effectively mitigate the overfitting to the source domains. For example, in the "D+Ms+C→Ma" task, the mAp of our method improves the baseline by $+11.0\%$ (51.4 vs. 40.4) on the unseen domain (as shown in Table III), while the mAP of our method decreases by -4.2% (65.4 vs. 69.6) on the source domain "DukeMTMC-reID". Therefore, this validates our method can mitigate the overfitting issue.

Visualization of the feature distribution. In Fig. 5, we visualize the feature distribution on source domains. As observed in this figure, the features of different domains from the baseline are scattered in different regions, while our method

TABLE IX
EXPERIMENTAL RESULTS ON SOURCE DOMAINS.

Method	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
			D+Ms+C→Ma			
	Test: D		Test: Ms		Test: C	
Baseline	69.6	83.7	45.3	73.3	56.1	57.5
MixNorm	65.4	81.0	43.1	73.1	45.5	45.3
			Ma+Ms+C→D			
	Test: Ma		Test: Ms		Test: C	
Baseline	79.6	92.2	45.5	73.5	58.8	60.9
MixNorm	76.5	91.7	42.8	72.4	50.3	51.7
			Ma+D+C→Ms			
	Test: Ma		Test: D		Test: C	
Baseline	79.8	92.2	69.0	83.8	58.2	60.4
MixNorm	76.9	92.0	66.1	82.5	48.3	48.8
			Ma+D+Ms→C			
	Test: Ma		Test: D		Test: Ms	
Baseline	80.4	92.1	71.8	84.9	47.2	74.7
MixNorm	76.6	91.5	67.4	82.4	42.9	72.1

(a) Baseline (D+Ms+C)   (b) MixNorm (D+Ms+C)

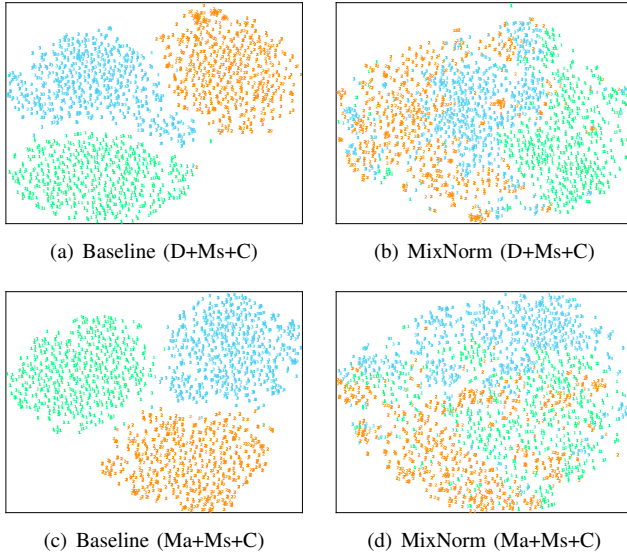(c) Baseline (Ma+Ms+C)   (d) MixNorm (Ma+Ms+C)

Fig. 5. Visualization of the features via t-SNE [65] on source domains. (a) and (b) are the visualization of feature representations from the model trained on DukeMTMC-reID, MSMT17 and CUHK03, and (c) and (d) are the visualization of feature representations from the model trained on Market1501, MSMT17 and CUHK03. Different colors denote different domains.

obviously tends to mix all the source domains into the same region, which validates that our method can well learn the domain-invariant features so as to capture the generalizable model to the unseen domain.

V. CONCLUSION

In this paper, we aim to tackle the generalizable multi-source person Re-ID task from the data augmentation view. Different from the existing methods, we propose a novel domain-aware mix-normalization (DMN) method to achieve data augmentation in the feature level, which can generate diverse features to prevent the overfitting of the model to source domains. Furthermore, we develop a domain-aware center regularization (DCR) for better using the diverse feature from DMN, which can map all features into the same space, thus it can enforce to learn the domain-invariant feature representations. We conduct the experiment on multiple datasets,

which demonstrates the effectiveness of the proposed method.

REFERENCES

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," *arXiv preprint arXiv:1610.02984*, 2016.

[2] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," *arXiv preprint arXiv:2001.04193*, 2020.

[3] Q. Leng, M. Ye, and Q. Tian, "A survey of open-world person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 30, no. 4, pp. 1092–1108, 2020.

[4] M. Li, X. Zhu, and S. Gong, "Unsupervised tracklet person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 42, no. 7, pp. 1770–1782, 2020.

[5] P. Chen, W. Liu, P. Dai, J. Liu, Q. Ye, M. Xu, Q. Chen, and R. Ji, "Occlude them all: Occlusion-aware attention network for occluded person re-id," in *International Conference on Computer Vision (ICCV)*, 2021, pp. 11 833–11 842.

[6] L. Qi, L. Wang, J. Huo, Y. Shi, and Y. Gao, "Progressive cross-camera soft-label learning for semi-supervised person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 30, no. 9, pp. 2815–2829, 2020.

[7] C. Zhao, X. Lv, Z. Zhang, W. Zuo, J. Wu, and D. Miao, "Deep fusion feature representation learning with hard mining center-triplet loss for person re-identification," *IEEE Transactions on Multimedia (TMM)*, vol. 22, no. 12, pp. 3180–3195, 2020.

[8] L. Wei, S. Zhang, H. Yao, W. Gao, and Q. Tian, "GLAD: global-local-alignment descriptor for scalable person re-identification," *IEEE Transactions on Multimedia (TMM)*, vol. 21, no. 4, pp. 986–999, 2019.

[9] A. Wu, W. Zheng, X. Guo, and J. Lai, "Distilled person re-identification: Towards a more scalable system," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1187–1196.

[10] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2285–2294.

[11] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2138–2147.

[12] L. Qi, L. Wang, J. Huo, Y. Shi, and Y. Gao, "Greyreid: A novel two-stream deep framework with rgb-grey information for person re-identification," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 1, pp. 27:1–27:22, 2021.

[13] F. Yang, K. Yan, S. Lu, H. Jia, D. Xie, Z. Yu, X. Guo, F. Huang, and W. Gao, "Part-aware progressive unsupervised domain adaptation for person re-identification," *IEEE Transactions on Multimedia (TMM)*, vol. 23, pp. 1681–1695, 2021.

[14] Y. Zhai, S. Lu, Q. Ye, X. Shan, J. Chen, R. Ji, and Y. Tian, "Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 9018–9027.

[15] A. Wu, W. Zheng, and J. Lai, "Unsupervised person re-identification by camera-aware similarity consistency learning," in *International Conference on Computer Vision (ICCV)*, 2019, pp. 6921–6930.

[16] G. Chen, Y. Lu, J. Lu, and J. Zhou, "Deep credible metric learning for unsupervised domain adaptation person re-identification," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 643–659.

[17] L. Qi, L. Wang, J. Huo, L. Zhou, Y. Shi, and Y. Gao, "A novel unsupervised camera-aware domain adaptation framework for person re-identification," in *International Conference on Computer Vision (ICCV)*, 2019, pp. 8079–8088.

[18] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *arXiv preprint arXiv:2103.02503*, 2021.

[19] S. Liao and L. Shao, "Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 456–474.

[20] Y. Zhao, Z. Zhong, F. Yang, Z. Luo, Y. Lin, S. Li, and N. 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 (CVPR)*, 2021, pp. 6277–6286.

[21] Y. Dai, X. Li, J. Liu, Z. Tong, and L. Duan, "Generalizable person re-identification with relevance-aware mixture of experts," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 16 145–16 154.

[22] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[23] J. Jia, Q. Ruan, and T. M. Hospedales, "Frustratingly easy person re-identification: Generalizing person re-id in practice," in *British Machine Vision Conference (BMVC)*, 2019, p. 117.

[24] X. Jin, C. Lan, W. Zeng, Z. Chen, and L. Zhang, "Style normalization and restitution for generalizable person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3140–3149.

[25] J. Song, Y. Yang, Y. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 719–728.

[26] S. Choi, T. Kim, M. Jeong, H. Park, and C. Kim, "Meta batch-instance normalization for generalizable person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 3425–3435.

[27] H. Nam, H. Lee, J. Park, W. Yoon, and D. Yoo, "Reducing domain gap by reducing style bias," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 8690–8699.

[28] S. Seo, Y. Suh, D. Kim, G. Kim, J. Han, and B. Han, "Learning to optimize domain specific normalization for domain generalization," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 68–83.

[29] X. Yue, Y. Zhang, S. Zhao, A. L. Sangiovanni-Vincentelli, K. Keutzer, and B. Gong, "Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data," in *International Conference on Computer Vision (ICCV)*, 2019, pp. 2100–2110.

[30] F. M. Carlucci, A. D'Innocente, S. Bucci, B. Caputo, and T. Tommasi, "Domain generalization by solving jigsaw puzzles," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2229–2238.

[31] Y. Balaji, S. Sankaranarayanan, and R. Chellappa, "Metareg: Towards domain generalization using meta-regularization," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 1006–1016.

[32] D. Li, J. Zhang, Y. Yang, C. Liu, Y. Song, and T. M. Hospedales, "Episodic training for domain generalization," in *International Conference on Computer Vision (ICCV)*, 2019, pp. 1446–1455.

[33] Y. Li, X. Tian, M. Gong, Y. Liu, T. Liu, K. Zhang, and D. Tao, "Deep domain generalization via conditional invariant adversarial networks," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 647–663.

[34] J. Zhang, L. Qi, Y. Shi, and Y. Gao, "Generalizable model-agnostic semantic segmentation via target-specific normalization," *Pattern Recognition (PR)*, vol. 122, p. 108292, 2022.

[35] Kaiyang Zhou and Yongxin Yang and Timothy M. Hospedales and Tao Xiang, "Learning to generate novel domains for domain generalization," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 561–578.

[36] K. Zhou, Y. Yang, T. M. Hospedales, and T. Xiang, "Deep domain-adversarial image generation for domain generalisation," in *AAAI Conference on Artificial Intelligence (AAAI)*, 2020, pp. 13 025–13 032.

[37] Q. Xu, R. Zhang, Y. Zhang, Y. Wang, and Q. Tian, "A fourier-based framework for domain generalization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14 383–14 392.

[38] Z. Huang, H. Wang, E. P. Xing, and D. Huang, "Self-challenging improves cross-domain generalization," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 124–140.

[39] P. Li, D. Li, W. Li, S. Gong, Y. Fu, and T. M. Hospedales, "A simple feature augmentation for domain generalization," in *International Conference on Computer Vision (ICCV)*, 2021, pp. 8886–8895.

[40] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, "Domain generalization with mixstyle," in *International Conference on Learning Representations (ICLR)*, 2021.

[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning (ICML)*, 2015, pp. 448–456.

[42] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. J. Smola, "A kernel two-sample test," *Journal of Machine Learning Research (JMLR)*, vol. 13, pp. 723–773, 2012.

[43] H. Luo, W. Jiang, Y. Gu, F. Liu, X. Liao, S. Lai, and J. Gu, "A strong baseline and batch normalization neck for deep person re-identification," *IEEE Transactions on Multimedia (TMM)*, vol. 22, no. 10, pp. 2597–2609, 2020.

[44] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, and T. S. Huang, "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification," in *International Conference on Computer Vision (ICCV)*, 2019, pp. 6111–6120.

[45] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *International Conference on Computer Vision (ICCV)*, 2015, pp. 1116–1124.

[46] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," in *International Conference on Computer Vision (ICCV)*, 2017, pp. 3774–3782.

[47] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 79–88.

[48] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 152–159.

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

[50] W. Li and X. Wang, "Locally aligned feature transforms across views," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3594–3601.

[51] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "End-to-end deep learning for person search," *arXiv preprint arXiv:1604.01850*, 2016.

[52] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Scandinavian Conference on Image Analysis (SCIA)*, 2011, pp. 91–102.

[53] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *European Conference on Computer Vision (ECCV)*, 2008, pp. 262–275.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[55] X. Pan, P. Luo, J. Shi, and X. Tang, "Two at once: Enhancing learning and generalization capacities via ibn-net," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 484–500.

[56] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Imagenet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[57] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation policies from data," *arXiv preprint arXiv:1805.09501*, 2018.

[58] X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun, "Alignedreid: Surpassing human-level performance in person re-identification," *arXiv preprint arXiv:1711.08184*, 2017.

[59] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv preprint arXiv:1803.02999*, 2018.

[60] S. Shankar, V. Piratla, S. Chakrabarti, S. Chaudhuri, P. Jyothi, and S. Sarawagi, "Generalizing across domains via cross-gradient training," in *International Conference on Learning Representations (ICLR)*, 2018.

[61] Y. Sun, L. Zheng, Y. Li, Y. Yang, Q. Tian, and S. Wang, "Learning part-based convolutional features for person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 902–917, 2021.

[62] D. Li, Y. Yang, Y. Song, and T. M. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *AAAI Conference on Artificial Intelligence (AAAI)*, 2018, pp. 3490–3497.

[63] S. Qiao, C. Liu, W. Shen, and A. L. Yuille, "Few-shot image recognition by predicting parameters from activations," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7229–7238.

[64] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 499–515.

[65] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research (JMLR)*, vol. 9, no. 11, 2008.