Rethinking Normalization Layers for Domain Generalizable Person Re-identification

Ren Nie[1], Jin Ding[1], Xue Zhou[2,1]✉, and Xi Li[3]

[1] University of Electronic Science and Technology of China, Chengdu, China
nieren@std.uestc.edu.cn, dingjin1998@outlook.com, zhouxue@uestc.edu.cn
[2] Shenzhen Institute for Advanced Study, UESTC
[3] College of Computer Science & Technology, Zhejiang University
xilizju@zju.edu.cn

Abstract. Domain Generalizable Person Re-Identification (DG-ReID) strives to transfer learned feature representation from source domains to unseen target domains, despite significant distribution shifts. While most existing methods enhance model generalization and discriminative feature extraction capability by introducing Instance Normalization (IN) in combination with Batch Normalization (BN), these approaches still struggle with the overfitting of normalization layers to the source domains, posing challenges in domain generalization. To address this issue, we propose **ReNorm**, a purely normalization-based framework that integrates two complementary normalization layers through two forward propagations for the same weight matrix. In the first forward propagation, **Remix Normalization** (**RN**) combines IN and BN in a concise manner to ensure the feature extraction capability. As an effective complement to **RN**, **Emulation Normalization** (**EN**) simulates the testing process of **RN**, implicitly mitigating the domain shifts caused by the absence of target domain information and actively guiding the model in learning how to generalize the feature extraction capability to unseen target domains. Meanwhile, we propose **Domain Frozen** (**DF**), freezing updates to affine parameters to reduce the impact of these less robust parameters on overfitting to the source domains. Extensive experiments show that our framework achieves state-of-the-art performance on the popular benchmarks. The code is available at https://github.com/3699nr/ReNorm.

Keywords: Normalization · Domain generalization · Re-identification

1 Introduction

Person Re-identification (ReID) is a practical task that has been widely applied in video surveillance and security. It aims at retrieving identical persons across a multi-camera network with non-overlapping fields of view. While current approaches [6,10,12,28,35,43,49] have achieved impressive performance within the same source domains during both training and testing phases. However, when

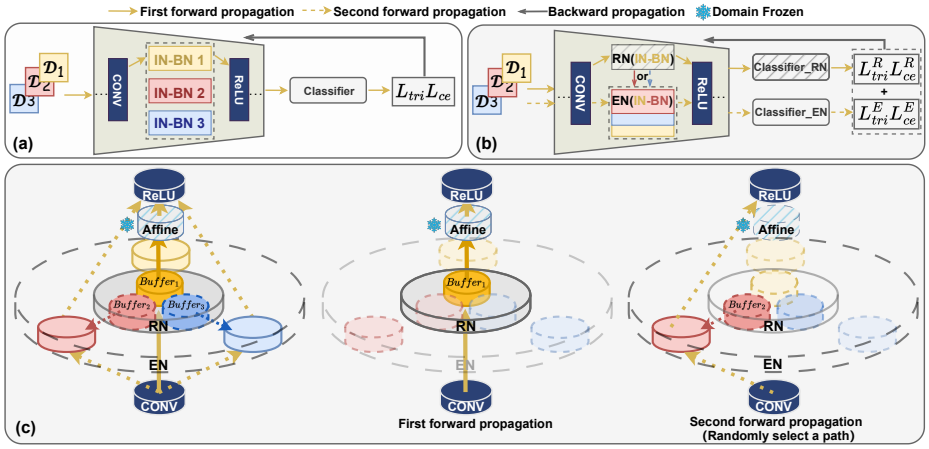


Fig. 1: Differences between existing normalization-based DG-ReID methods and ReNorm. To illustrate more clearly, we use the example of training with three source domains. (a) Existing normalization-based methods combine IN and BN in a specific way to ensure both the generalization and discrimination capability of the weight matrix. (b) ReNorm integrates two complementary normalization layers through two forward propagations for the same weight matrix. RN enables the weight matrix to extract discriminative features within the source domains, while EN is employed to actively guide the weight matrix to generalize the feature extraction capability to unseen target domains. (c) Some details of ReNorm. Detailed explanations can be found in Sec 3.

directly deploying these well-trained models on an unseen target domain, an issue known as "domain shift" [39] arises, significantly impacting the performance of the models. To alleviate this problem, recent efforts have been devoted to two approaches: unsupervised domain adaptive (DA) ReID [5,17,20,26], and domain generalizable (DG) ReID [2,3,19,37,44,45]. In contrast to DA ReID, DG ReID is more practical and challenging as it trains on the multiple source domains and then directly transfers the learned representation to the unseen target domain without any target data for fine-tuning.

Among numerous DG ReID methods, normalization-based approaches [2,3,8, 19,37] stand out for their simplicity and versatility, often integrating seamlessly with other advanced techniques such as meta-learning, distribution alignment, and ensemble learning. Illustrated in Fig.1(a), the combination of Batch Normalization (BN) [16] for capturing discriminative features and Instance Normalization (IN) [38] for eliminating domain-specific information facilitates normalization-based DG-ReID methods in achieving both high generalization and discrimination capabilities. During training, the statistical parameters of both IN and BN within the normalization layers are derived from the same data, ensuring a consistent distribution. However, this consistency is disrupted when the model is deployed on the target domain, as IN statistical parameters originate from the target domain itself, while BN statistical parameters are derived from population

statistics accumulated from source domains during training. The distribution of the target domain significantly differs from that of the source domains, resulting in a marked decline in the model's performance. Furthermore, the vulnerable and sensitive learnable affine parameters are also prone to overfitting to the source domains, leading to performance degeneration in both training and testing phase.

Moreover, emerging researches [24, 40] have shown that the identity and domain information are separately embedded in the weight matrix (learnable parameters in convolutional layers) and normalization layers. Consequently, we rethink the key challenges of DG ReID, which are rooted in normalization layers. We believe that the capability of the weight matrix to extract discriminative features sets the upper limit of the model's performance. Simultaneously, the normalization layers, responsible for fitting and scaling the input from target domains, establish the lower limit for the model's generalization performance.

To address the issue of normalization layers being prone to overfitting to the source domains, we restructure the normalization layers, proposing **ReNorm**, a concise yet effective DG ReID framework entirely based on the design of normalization layers, as shown in Fig.1(b). In each iteration, the network undergoes two forward propagations and then one backpropagation after aggregating two forward losses. The two forward propagations apply different processing to the normalization layers, utilizing two complementary normalization layers to facilitate the weight matrix in extracting discriminative features from familiar source domains, while simultaneously guiding the same weight matrix to learn how to transfer this acquired feature extraction capability to unfamiliar target domains.

Specifically, in our **ReNorm** framework, the two forward propagations correspond to two different but complementary normalization strategies: Remix Normalization (RN) and Emulation Normalization (EN). In the first forward propagation, RN introduces Instance Normalization (IN) in combination with Batch Normalization (BN) by directly summing their statistical parameters for standardization. Meanwhile, the BN statistical parameters of the current source domain are stored in the corresponding buffer, as shown in Fig.1(c). In the second forward propagation, EN adopts a similar summing approach but emulates the inference process of RN. It treats the current input as a simulated target domain, and calculates the IN statistical parameters accordingly. Simultaneously, the required BN statistical parameters are obtained from a remaining source domain's buffer stored in RN. This pairwise normalization strategy enables the network to actively learn how to transfer knowledge from source domains to the target domain with a distinct distribution. Additionally, we introduce Domain Frozen (DF), where the vulnerable settings of learnable affine parameters within both normalization layers are frozen and exempt from updates. By doing so, **ReNorm** greatly raises the lower limit of the model's generalization ability.

We summarize our contributions as follows:

- We introduce **ReNorm**, a simple yet effective framework for DG ReID, entirely and solely based on the design of normalization layers. This framework significantly mitigates performance degradation resulting from distribution shifts between source and target domains in a concise yet effective way.

- We introduce two complementary normalization strategies (*i.e.*, Remix Normalization and Emulation Normalization) equipped with Domain Frozen, actively enabling the network to learn how to transfer the feature extraction capability acquired from source domains to unseen target domains.
- Extensive experiments demonstrate the simplicity yet state-of-the-art performance of our **ReNorm** framework under multiple testing protocols.

2 Generalizable Person Re-Identification.

Deep learning-based Person ReID approaches [6, 10, 12, 28, 35, 43, 49] have made great progress in recent years, achieving impressive accuracy in Person ReID. However, when these well-trained models are applied to unseen target domains that have distribution shifts with source domains, the performance decreases significantly. To alleviate this problem, recent efforts are devoted to two approaches: unsupervised domain adaptive (DA) ReID [5,17,20,26] and domain generalizable (DG) ReID. Diverging from the DA ReID, which needs unlabeled target domain datasets for unsupervised fine-tuning. DG ReID trains on the source domains and then directly tests on the unseen target domain. As a result, DG ReID has attracted attention for its robustness and generalization.

Existing DG ReID methods can be briefly divided into four categories:

Normalization strategies [2, 33, 34]. Batch Normalization (BN) can preserve diversity among samples, capturing more discriminative features and accelerating convergence. However, it tends to be prone to overfitting to source domains. Instance Normalization (IN) can eliminate domain-specific information to enhance generalization but inevitably removes discriminative information. Thus, normalization-based methods typically combine IN and BN in various ways to ensure both high generalization and discrimination capability of the network. Jiao et al. [18] proposed Dynamically Transformed Instance Normalization (DTIN) to alleviate the drawback of IN by transforming normalized features into appropriate representations in a learnable and adaptive manner. Chen et al. [2] designed a Cluster-Instance Normalization (CINorm) to eliminate the effect of outliers in original BN that causes severe overfitting.

Domain alignment [19, 32]. These methods try to minimize the difference among source domains for learning domain-invariant representations. They assume that the invariant features from the source domains should also be robust to any unseen target domains. Jin et al. [19] proposed a dual causality loss to align the outputs and a Style Normalization and Restitution (SNR) module to distill identity-relevant features via a restitution module for IN compensation.

Meta-learning [3, 37, 46]. These methods improve generalized performance through virtual simulations of real train-test domain shifts. Following the concept of 'learning to lean ', Choi et al. [3]proposed MetaBIN which utilized a meta-learning training strategy with sets of learnable balancing parameters to integrate BN and IN.

Ensemble learning [4,25,44]. These methods exploit the relationship among multiple source domains by using diverse experts to extract domain-specific features from the corresponding domain. To avoid the large model size, Xu et al. [44]

proposed a lightweight model META, which leverages IN and introduces it into a global branch to pursue invariant features across domains.

By introducing IN in combination with BN, normalization-based methods are simple and effective and can be combined with other techniques. However, when tested on unknown target domains, the distribution shifts between source and target domains significantly affect the fitting and scaling ability of normalization layers. Thus, leading to a decline in the model's generalization performance.

3 Methodology

Inspired by the observation that the weight matrix tends to encapsulate identity information while domain-specific knowledge is embedded within the normalization layers [24, 40], we propose a novel **ReNorm** framework for tackling the domain shifts problem for DG-ReID which is rooted in normalization layers. As shown in Fig.2, **ReNorm** integrates two complementary normalization layers through two forward propagations for the same weight matrix. In each iteration, we sample a mini-batch from a source domain. In the first forward propagation, Remix Normalization (RN) combines IN and BN to ensure the discriminative feature extraction capability of the weight matrix. At the same time, the BN statistical parameters of the current source domain are stored in the corresponding buffer by Exponential Moving Average (EMA). During the second forward propagation, we treat the current input as a simulated target domain of the remaining source domains. Thus, Emulation Normalization (EN) can emulate the inference process of RN to guide the weight matrix to learn how to transfer the acquired feature extraction capability to an unseen domain. Meanwhile, we introduce Domain Frozen (DF) to stabilize the weight matrix by freezing the updates of affine parameters in normalization layers.

3.1 Preliminary

In the DG ReID task, we have access to K source domains (datasets) for training. We denote the k-th source domain as $D_k, k \in 1, 2, ..., K$, which have completely non-overlapping label spaces. Given a feature map $X_k \in \mathbb{R}^{N\times C\times H\times W}$ of a mini-batch extracted from a certain domain D_k, where N, C, H, and W denote the batch size, the number of channels, the height, and the width, respectively. Following the definition from [41], we decompose the normalization layers into two steps as follows:

$$\hat{X}_k = \frac{X_k - \mu_k}{\sqrt{\sigma_k^2 + \epsilon}}, \quad \tilde{X}_k = \gamma \hat{X}_k + \beta, \tag{1}$$

The first step involves standardization using the mean μ_k and variance σ_k^2 calculated from the input feature map. After that, channel-wise learnable affine parameters γ and β are defined for scaling and shifting during end-to-end training. $\epsilon > 0$ is a constant to avoid numerical problems. The main distinction among various normalization techniques primarily lies in the first step.

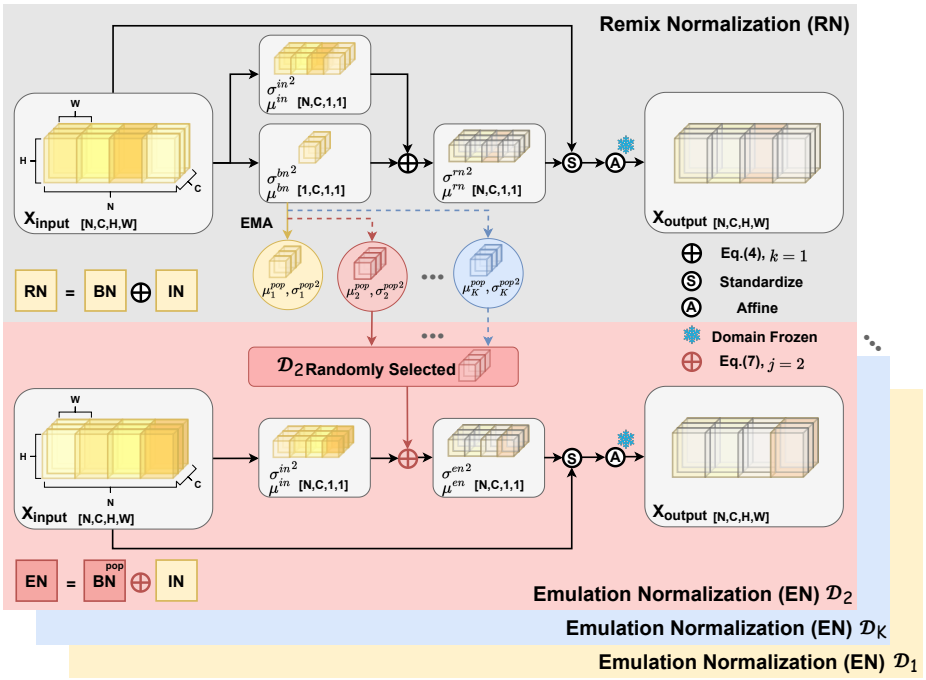


Fig. 2: The intra-layer structure of our ReNorm framework. The upper part corresponds to the first forward propagation with Remix Normalization, while the lower part corresponds to the second forward propagation with Emulation Normalization. After standardization, we employ Domain Frozen strategy for the affine parameters. Best viewed in color, detailed explanations can be found in Sec 3.2 and Sec 3.3.

Instance Normalization [38] can remove style information from input features. So it has been widely used for the DG-ReID task to extract domain-invariant representations by removing their statistical contrast across feature channels. IN calculates mean $\mu_k^{in} \in \mathbb{R}^{N\times C}$ and variance ${\sigma_k^{in}}^2 \in \mathbb{R}^{N\times C}$ with respect to each sample and each channel for standardization:

$$\mu_k^{in} = \frac{\sum_{h,w}^{HW} X_k}{HW} \quad \text{and} \quad {\sigma_k^{in}}^2 = \frac{\sum_{h,w}^{HW} \left(X_k - \mu_k^{in}\right)^2}{HW}. \tag{2}$$

Batch Normalization [16] is widely used in convolutional neural networks to capture discriminative features and mitigate gradient vanishing or exploding. It calculates mean $\mu_k^{bn} \in \mathbb{R}^C$ and variance ${\sigma_k^{bn}}^2 \in \mathbb{R}^C$ with respect each channel in a mini-batch:

$$\mu_k^{bn} = \frac{\sum_{n,h,w}^{NHW} X_k}{NHW} \quad \text{and} \quad {\sigma_k^{bn}}^2 = \frac{\sum_{n,h,w}^{NHW} \left(X_k - \mu_k^{bn}\right)^2}{NHW}. \tag{3}$$

It is worth noting that different from IN where the mean μ^{in} and variance ${\sigma^{in}}^2$ are computed from the input feature map X during both training and test-

ing. During testing, BN standardizes the input feature map X by population statistics which are estimated using Exponential Moving Average (EMA) during training. This special mechanism of BN enriched with domain-specific information of source domains poses a significant risk to the generalization performance of the model.

3.2 Remix Normalization

Previous studies [15,31] have shown that Batch Normalization (BN) can extract domain-specific discriminative information, but it can easily overfit to a particular domain data in a mini-batch, while Instance Normalization (IN) can reduce the domain-specific style information, but it results in the loss of some discriminative information. Therefore, we combine these two normalization methods for a trade-off. Inspired by DSON [36] and SN [29], we design Remix Normalization (RN) which combines the statistical parameters of BN and IN in a simple way. Given an input feature map X_k, RN calculates mean $\mu_k^{rn} \in \mathbb{R}^{N\times C}$ and variance $\sigma_k^{rn\,2} \in \mathbb{R}^{N\times C}$ for standardization as follows:

$$\begin{aligned} \mu_k^{rn} &= (\mu_k^{\text{in}} + \mu_k^{\text{bn}})/2, \\ \sigma_k^{rn\,2} &= (\sigma_k^{in\,2} + \sigma_k^{bn\,2})/2. \end{aligned} \tag{4}$$

Meanwhile, unlike previous approaches where each normalization layer corresponds to one BN buffer, we set up corresponding $Buffer_k : \mu_k^{pop}, \sigma_k^{pop\,2} \in \mathbb{R}^C, k \in 1, 2, ..., K$ for each RN for K source domains to store their domain-specific information as follows:

$$\begin{aligned} \mu_k^{\text{pop}} &\leftarrow \lambda\mu_k^{\text{pop}} + (1-\lambda)\mu_k^{bn}, \\ \sigma_k^{pop\,2} &\leftarrow \lambda\sigma_k^{pop\,2} + (1-\lambda)\sigma_k^{bn\,2}. \end{aligned} \tag{5}$$

As shown in Fig.2, we use three primary colors (yellow, red, blue) to represent the corresponding source domains D_1, D_2, D_K. The input mini-batch is from D_1. After calculating the required IN and BN statistical parameters, we follow Eq.5 to update the corresponding BN population statistics (*i.e.*, $Buffer_1 : \mu_1^{pop}, \sigma_1^{pop\,2}$). In Fig.2, we represent this update process with solid lines.

The efficacy of combining BN and IN has been verified in many previous DG ReID works [2, 33, 34]. Nevertheless, a common challenge persists across these approaches: overfitting of the normalization layers to the source domains. Specifically, during the training process, following Eq.4, the statistical parameters of IN and BN within the normalization layer are derived from the same source domain, resulting in a consistent distribution. However, when the model is directly tested on the target domain, there is no concept of mini-batch anymore [41]. Thus, RN calculates the parameters for standardization as follows:

$$\begin{aligned} \mu_k^{rn} &= (\mu_{tar}^{\text{in}} + \mu_k^{\text{pop}})/2, \\ \sigma_k^{rn\,2} &= (\sigma_{tar}^{in\ 2} + \sigma_k^{pop\,2})/2, \end{aligned} \tag{6}$$

the IN statistical parameters (*i.e.*, $\mu_{tar}^{\text{in}}, {\sigma_{tar}^{in}}^2$) are derived from the input target domain itself. Conversely, the BN statistical parameters are based on population statistics (*i.e.*, $\mu_k^{pop}, \sigma_k^{pop2}$) of source domains. The actual distribution between the target domain and the source domain is completely different, which leads to the inability of the statistics parameters (*i.e.*, $\mu_k^{rn}, \sigma_k^{rn2}$) to fit the input X_{tar} from the target domain. Concurrently, the model lacks the adaptability to address this mismatch, resulting in a performance decline. Furthermore, the learnable affine parameters also tend to overfit to the source domains, leading to a decrease in performance during both training and testing.

3.3 Emulation Normalization

Remix Normalization enables the weight matrix in learning to extract discriminative features while also possessing some degree of domain invariance. However, a crucial limitation emerges when the model encounters the distribution shifts introduced by the unseen target domain, as it lacks the knowledge of how to adapt. Hence, a natural consideration arises: given the entirely distinct distributions among different source domains, we can treat the current source domain as a simulated target domain of a remaining source domain.

Thus, we introduce a novel normalization operation named Emulation Normalization (EN), which emulates the inference procedure of Remix Normalization during training. By doing so, EN actively guides the weight matrix to acquire the capability of handling issues arising from the inability of RN to fit the input from unseen target domains. EN calculates mean $\mu_k^{en} \in \mathbb{R}^{N\times C}$ and variance $\sigma_k^{en2} \in \mathbb{R}^{N\times C}$ as follows:

$$\begin{aligned} \mu_k^{en} &= \left(1 - w_j^m\right) \mu_k^{\text{in}} + w_j^m \mu_j^{\text{pop}} \\ \sigma_k^{en2} &= \left(1 - w_j^v\right) {\sigma_k^{in}}^2 + w_j^v \sigma_j^{pop2}, \end{aligned} \tag{7}$$

where we treat the current input source domain D_k as the simulated target domain of D_j, which is randomly selected from the remaining. By combining IN statistical parameters of X_k from this simulated target domain D_k and BN population statistics from a remaining source domain D_j, our EN can actively guide the weight matrix to transfer discriminative feature extraction from familiar source domains to unfamiliar target domain with different distribution. Notably, the learnable balancing ratios (*i.e.*, w_j^m, w_j^v) correspond individually to the randomly selected source domain D_j.

As shown in Fig.2, We use the current input D_1 as the simulated target domain of D_2 which supports the required BN statistical parameters stored in $Buffer_2$ (*i.e.*, $\mu_2^{pop}, \sigma_2^{pop2}$), as specified in Eq.7. It is worth noting that $Buffer_2$ could combine IN statistical parameters from D_1 and D_K, this pattern holds true for all the buffers.

3.4 Domain Frozen

Another factor that may lead to a decrease in the model's generalization capability lies in the learnable affine parameters of the normalization layers. Compared

to the weight matrix, the number of learnable affine parameters is significantly smaller, implying a weaker robustness when handling the target domain with a distribution divergence from the source domains. Moreover, ReID can be approximated as a channel-wise classification task, indicating that channel-wise affine parameters exert a more profound influence.

Futhermore, as shown in Fig.3, we visualize the affine parameters for each source domain, approximating the domain-specific information within. It is evident that significant disparities exist among different source domains. These channel-wise divergences will accumulate into the weight matrix through backward propagation, leading to confusion in feature learning.

Consequently, we propose Domain Frozen (DF), a strategy that stabilizes the weight matrix by freezing the updates of affine parameters in every normalization layer. For Remix Normalization, we maintain the affine parameters in a frozen state following their initialization with pre-trained weights. In the case of Emulation Normalization, the input feature map X_k is standardized by statistical parameters from two different domains. The output $\hat{X}_k$ inherently deviates from a normal distribution. Consequently, with the integration of Domain Frozen, we have opted to directly eliminate the original affine parameters, which is tantamount to setting the affine parameters γ and β to 1 and 0, respectively.

Moreover, a further apparent advantage of DF is that it indirectly shares the affine parameters between the source domains (during training) and the target domain (during testing). This sharing further mitigates the domain shifts caused by the absence of target domain information, enhancing the model's ability to generalize to unseen domains.

3.5 Training and testing

As illustrated in Fig.1(b), we propose a novel **ReNorm** framework to integrate the aforementioned normalization strategies. For clarification purposes, we refer to the weight matrix equipped with Remix Normalization (RN) as $Network-R$ and the same weight matrix equipped with Emulation Normalization (EN) as $Network-E$. Meanwhile, Domain Frozen (DF) is applied uniformly in both RN and EN.

Model Training. In our methodology, there is no need for additional design of loss functions. The employment of two fundamental ReID loss functions, triplet loss (L_{tri}) [13] and cross-entropy loss (L_{ce}) suffices for achieving optimal performance. In each iteration, we randomly select one source domain D_k from a set of K source domains and randomly sample a mini-batch samples of size N_b from this domain. As illustrated in Fig.1, this set of data is then fed through the $Network-R$ for first forward propagation and calculate corresponding triplet loss L^R_{tri} and cross-entropy loss L^R_{ce}, which is represented by the solid line. Subsequently, we treat the current source domain as a simulated target domain and randomly select one D_j from the remaining source domains to serve as a simulated source domain, providing Batch Normalization (BN) statistical parameters (*i.e.*, $\mu^{pop}_j, \sigma^{pop2}_j$) for the second forward propagation, which is indicated by the dashed line. The same mini-batch will be fed into $Network-E$ to compute

the corresponding triplet loss L_{tri}^{E} and cross-entropy loss L_{ce}^{E}. Combing these two networks mentioned above, we have the following total loss for backward propagation:

$$L_{total} = L_{tri}^{R} + L_{ce}^{R} + L_{tri}^{E} + L_{ce}^{E}. \tag{8}$$

Testing. During testing, we feed the target domain to both networks K times each. Subsequently, we compute the average of the $2 * K$ features obtained from two networks to obtain the final inference result.

4 Experiments

4.1 Datasets and Settings

Datesets. Our approach is evaluated on 9 person ReID datasets, including 5 large-scale datasets: Market1501 (M) [47], MSMT17 (MS) [39], CUHK02 (C2) [22], CUHK03 (C3) [23], and CUHK-SYSU (CS) [42], and 4 small-scale datasets: PRID [14], GRID [27], VIPeR [7], and iLIDs [48]. We do not use DukeMTMC-reID (D) [50] since it has been retracted. We adopt Cumulative Matching Characteristics (CMC) and mean Average Precision (mAP) as evaluation metrics.

Evaluation Protocols. We set three evaluation protocols following [37,44]. For Protocol-1, all images including both training and testing sets from four large-scale datasets (*i.e.*, M, CS, C2, C3) are used for training, and the four small-scale datasets (*i.e.*, PRID, GRID, VIPeR, iLIDs) are used for testing. The final results for each dataset are evaluated on the average of 10 repeated random splits of gallery and query sets. For Protocol-2 and Protocol-3, following the leave-one-out setting in [44], we select one domain from M+MS+CS+C3 for testing, while the remaining three domains serve as the training sets. CS is excluded as a target domain due to its images all come from one camera view. Additionally, under Protocol-2, only the training sets of the source domains are used for training, while all images are exploited under Protocol-3.

Implementation Details. We employ ResNet50 [9] pre-trained on ImageNet as our backbone. The size of a mini-batch is set to 64, including 16 identities and 4 images per identity. Each image is resized to 256×128. We use random horizontal flipping, random cropping, zero padding, color jittering, and auto-augmentation [11] as data augmentation. We also adopt a warm-up strategy in the first 10 epochs. The learning rate is initialized as 3.5×10^{-4} and divided by 10 at the 30th, 60th and, 90th epochs, respectively. All experiments are conducted using Pytorch on a single RTX 4090 GPU.

4.2 Comparison with the State-of-the-Arts

Comparison with DG-ReID Methods Under Protocol-1. As shown in Table 1, we compare our method with other state-of-the-arts under Protocol-1. Since DukeMTMC-reID (D) [50] has been retracted, we only use other large-scale datasets for training. Notably, our method demonstrates superior performance on PRID, GRID, and VIPeR. Overall, **ReNorm** significantly outperforms other

Table 1: Comparison with the SOTA methods in DG ReID under the setting of Protocol-1. All results are the average of 10 random sampling. The best and the second best results are highlighted in **bold** and underline.

Method	Reference	Source domains	Target domain								**Average**	
			PRID		GRID		VIPeR		iLIDs			
			mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
SNR [19]	CVPR'20	M+D+C2 +C3+CS	66.5	52.1	47.7	40.2	61.3	52.9	89.9	84.1	66.4	57.3
RaMoE [4]	CVPR'21		67.3	57.7	54.2	46.8	64.6	56.6	**90.2**	**85.0**	69.1	61.5
DMG-Net [1]	CVPR'21		68.4	60.6	56.6	51.0	60.4	53.9	83.9	79.3	67.3	61.2
DTIN-Net [18]	ECCV'22		79.7	71.0	60.6	51.8	70.7	62.9	87.2	81.8	74.6	66.9
MetaBIN [3]	CVPR'21	M+C2 +C3+CS	70.8	61.2	57.9	50.2	64.3	55.9	82.7	74.7	68.9	60.5
META [44]	ECCV'22		71.7	61.9	60.1	52.4	68.4	61.5	83.5	79.2	68.9	60.5
ACL [45]	ECCV'22		73.4	63.0	65.7	55.2	75.1	66.4	86.5	81.8	75.2	66.6
CINorm [2]	TMM'23		80.5	72.4	59.9	49.4	76.7	67.6	88.0	82.2	76.2	67.9
PAOA [21]	WACV'24		74.0	65.6	67.2	56.3	76.6	66.7	87.1	83.1	76.2	67.9
ReNorm	This paper		**81.3**	**73.9**	**67.5**	**59.4**	**77.0**	**69.1**	87.0	82.2	**78.2**	**71.2**

methods by at least 2.0% and 3.3% in average mAP and Rank-1, respectively. The experimental findings substantiate the efficacy of our **ReNorm**, which combines simplicity with substantial impact.

Comparison Under Protocol-2 and Protocol-3. We compare our method with previous state-of-the-arts under protocol-2 and protocol-3 to evaluate our method on large-scale datasets. As shown in Table 2, we can find that our **ReNorm** achieves superior performance compared with previous methods on almost all datasets under the two protocols. Overall, **ReNorm** outperforms other approaches by an average of at least 2.0% mAP and 4.1% Rank-1 under Protocol-2, and by 1.7% mAP and 2.9% Rank-1 under Protocol-3, respectively. Notably, when MSMT17 (MS) serves as the target domain, our method surpasses the second best by 3.1% mAP and 5.7% Rank-1 under Protocol-2, and by 2.7% mAP and 5.7% Rank-1 under Protocol-3. Collectively, these results further illustrate our model's superiority in domain generalization.

4.3 Ablation Study

Effectiveness of different modules. We investigate the effects of different components in **ReNorm** design to study their individual contributions, as shown in Table 3. The first row corresponds to a basic ResNet50 framework. In the table, every even row compared to the previous row incorporates the Domain Frozen (DF) strategy, which straightforwardly boosts the network by at least 1.3% mAP and 1.5% Rank-1, on average. Comparing first and second rows with the third and fourth rows, even when Remix Normalization (RN) simply combines IN and BN, it significantly enhances the network's generalization performance. Furthermore, with the addition of Emulation Normalization (EN), the two complementary normalization layers further improve the model's generalization ability, resulting in an average improvement of at least 2.0% mAP and 2.5% Rank-1. The experiments thoroughly demonstrate the simplicity and effectiveness of our approach.

Table 2: Comparison with the SOTA methods in DG ReID under the setting of Protocol-2 and Protocol-3. All results are the average of 10 random sampling. The best and the second best results are highlighted in **bold** and underline.

Method	Reference	Setting	M+MS+CS →C3		M+CS+C3 →MS		MS+CS+C3 →M		**Average**	
			mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
SNR [19]	CVPR'20	Protocol-2	8.9	8.9	6.8	19.9	34.6	62.7	16.8	30.5
M^3L [46]	CVPR'21		20.9	31.9	15.9	36.9	58.4	79.9	31.7	49.6
MetaBIN [3]	CVPR'21		28.8	28.1	17.8	40.2	57.9	80.1	34.8	49.5
META [44]	ECCV'22		36.3	35.1	22.5	49.9	67.5	86.1	42.1	57.0
ACL [45]	ECCV'22		41.2	41.8	20.4	45.9	**74.3**	**89.3**	45.3	59.0
ReNorm	This Paper		**43.6**	**44.7**	**25.6**	**55.6**	72.7	89.1	**47.3**	**63.1**
SNR [19]	CVPR'20	Protocol-3	17.5	17.1	7.7	22.0	52.4	77.8	25.9	39.0
M^3L [46]	CVPR'21		35.7	36.5	17.4	38.6	62.4	82.7	38.5	52.6
MetaBIN [3]	CVPR'21		43.0	43.1	18.8	41.2	67.2	84.5	43.0	56.3
META [44]	ECCV'22		47.1	46.2	24.4	52.1	76.5	90.5	49.3	62.9
ACL [45]	ECCV'22		49.9	50.1	21.7	47.3	76.8	90.6	49.3	62.7
PAOA [21]	WACV'24		49.8	50.5	25.1	51.5	77.1	90.8	50.7	64.3
ReNorm	This Paper		**51.9**	**52.5**	**27.8**	**57.8**	**77.4**	**91.2**	**52.4**	**67.2**

Table 3: Ablation studies for different components of our framework. The experiment is conducted under protocol-2. The best results are highlighted in **bold**.

RN	EN	DF	Target: C3		Target: MS		Target: M		**Average**	
			mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
			35.3	34.9	19.9	44.9	66.3	85.1	40.5	55.0
		✓	36.5	36.9	22.1	48.5	68.2	86.6	42.3	57.3
✓			40.5	40.5	22.1	49.4	69.0	87.1	43.9	59.0
✓		✓	40.4	40.2	23.7	52.7	71.9	88.9	45.3	60.6
✓	✓		42.5	43.6	24.4	53.4	71.1	87.9	46.0	61.6
✓	✓	✓	**43.6**	**44.7**	**25.6**	**55.6**	**72.7**	**89.1**	**47.3**	**63.1**

Visualization of the domain-specific information within the affine parameters. In this experiment, we establish corresponding RN for each source domain and reinstate the updates of all affine parameters. After training, we sum the affine parameters γ of each source domain at each RN layer to approximate the domain-specific information within. As shown in Fig.3, The horizontal axis represents the layer position of RN in the backbone, while the vertical axis represents the corresponding proportions. It is evident that significant disparities exist among source domains across all the normalization layers. During training, the disparities accumulate in the weight matrix, potentially leading to misguided learning.

Effectiveness on where to plug the modules. As normalization-based modules, the proposed ReNorm and DF can be plugged into any stage of the network by replacing the corresponding BN layers. In our experiments, we employ ResNet50 as our backbone, which consists of four groups of residual blocks and is organized into four stages. We replace our ReNorm and DF with diverse stages'

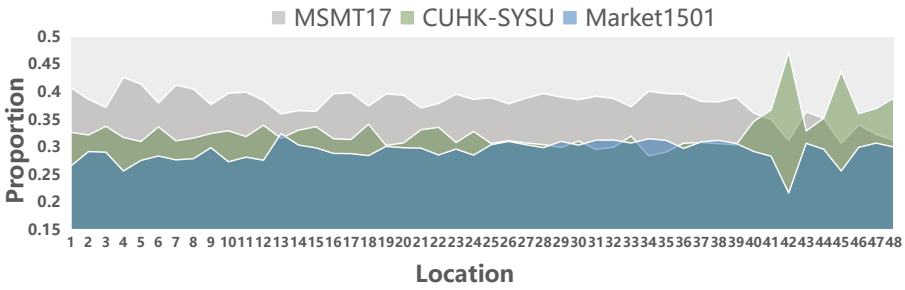


Fig. 3: Visualization of the domain-specific information within the affine parameters under Protocol-2, the source domains used for training being MSMT17, CUHK-SYSU and Market1501.

Table 4: Ablation study on where to plug the modules into ResNet50.

Stage	Target: C3			
	Module: ReNorm		Module: DF	
	mAP	Rank-1	mAP	Rank-1
None	34.9	35.3	42.5	43.6
4	39.4	39.8	43.0	43.3
3 - 4	41.9	42.0	43.3	43.5
2 - 4	42.2	43.3	43.1	44.0
1 - 4	**43.6**	**44.7**	**43.6**	**44.7**

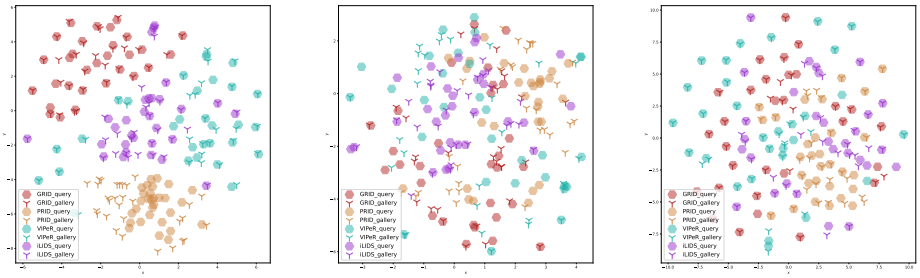
Table 5: Ablation study on different combination of outputs into the result.

Method	Average	
	mAP	Rank-1
concat(R)	46.2	61.8
concat(E)	45.8	60.8
mean(R)	46.3	61.9
mean(E)	45.8	60.9
concat(R,E)	47.0	62.8
mean(R,E)	**47.3**	**63.1**

BN layers to explore the effects of different plugging stages. As shown in Table 4, the results demonstrate that employing ReNorm and DF across all normalization layers performs better than others. Additionally, incorporating ReNorm and DF from any stage could improve overall performance, further demonstrating the effectiveness of our modules.

Combination of outputs into the final result. The experiment is conducted under Protocol-2. During testing, each image will go through $Network-R$ (R) and $Network-E$ (E) separately K times, resulting in a total of $2*K$ outputs. As shown in Table 5, from the first four rows, it is evident that network-R generally outperforms network-E, and both have surpassed the previous SOTA with 45.3% mAP and 59.0% Rank-1 on average. After combining the outputs of the two networks, there is a significant improvement in the final results. Ultimately, taking the mean of the $2*K$ outputs directly yields the best result with 47.3% mAP and 63.1% Rank-1. Experimental results confirm the enhanced generalization capability of our framework in domain-invariant feature extraction.

Visualization of the target domain distribution. To further demonstrate the effectiveness of our ReNorm framework, we visualize the feature distribution of unseen target domains under Protocol-1. We conduct t-SNE [30] experiments on the same query and gallery image pairs from four target domains by implementing BN, IN, and our ReNorm, respectively. In Fig.4 (a), the



(a) BN **(b)** IN **(c)** ReNorm (ours)

Fig. 4: The t-SNE results of features under Protocol-1 on three models with different normalization layers. (a) indicates ResNet50 with BN for all normalization layers, (b) denotes using instance normalization, and (c) utilizes our ReNorm for all normalization layers. Various colors illustrate samples from different target domains, and different shapes denote sample pairs from query or gallery sets.

"BN" model exhibits smaller distances between query and gallery sets of the same identity but is heavily influenced by domain factors, leading to clear domain boundaries. In contrast, Fig.4 (b) shows that IN removes these boundaries but increases the distance between query and gallery sample pairs. The observations prove that BN can capture more discriminative features but tends to overfit to source domains, while IN can eliminate domain-specific information to enhance generalization but inevitably removes discriminative information. Fig.4 (c) shows that the distances between query and gallery sample pairs are much smaller, demonstrating that our approach achieves better performance. Simultaneously, there are no apparent boundaries among target domains, indicating that our method effectively eliminates domain-specific information.

5 Conclusion

In this paper, we present a purely normalization-based framework called **ReNorm** for tackling the problem of domain generalizable person re-identification (DG-ReID). Our framework integrates two complementary normalization layers through two forward propagations for the same weight matrix. **Remix Normalization (RN)** enables the weight matrix to extract discriminative features within the source domains, while **Emulation Normalization (EN)** is employed to actively guide the weight matrix to learn how to generalize the feature extraction capability to unseen target domains. Meanwhile, we introduce **Domain Frozen (DF)**, a strategy that stabilizes the weight matrix by freezing the updates of affine parameters in normalization layers. Extensive experiments demonstrate the simplicity yet state-of-the-art performance of **ReNorm** among all the popular benchmarks.

Acknowledgements

This work was supported by the Natural Science Foundation of China (No. 62372082, 61972071, U20A20184), Natural Science Foundation of Sichuan Province (No.2023NSFSC0485), National Science Foundation for Distinguished Young Scholars under Grant 62225605 and Zhejiang Key Research and Development Program under Grant 2023C03196.

References

1. Bai, Y., Jiao, J., Ce, W., Liu, J., Lou, Y., Feng, X., Duan, L.Y.: Person30k: A dual-meta generalization network for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2123–2132 (2021)
2. Chen, Z., Wang, W., Zhao, Z., Su, F., Men, A., Dong, Y.: Cluster-instance normalization: A statistical relation-aware normalization for generalizable person re-identification. IEEE Transactions on Multimedia (2023)
3. Choi, S., Kim, T., Jeong, M., Park, H., Kim, C.: Meta batch-instance normalization for generalizable person re-identification. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 3425–3435 (2021)
4. Dai, Y., Li, X., Liu, J., Tong, Z., Duan, L.Y.: Generalizable person re-identification with relevance-aware mixture of experts. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16145–16154 (2021)
5. Dai, Y., Liu, J., Sun, Y., Tong, Z., Zhang, C., Duan, L.Y.: Idm: An intermediate domain module for domain adaptive person re-id. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11864–11874 (2021)
6. Fu, D., Chen, D., Bao, J., Yang, H., Yuan, L., Zhang, L., Li, H., Chen, D.: Unsupervised pre-training for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14750–14759 (2021)
7. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. ECCV (1) **2008**, 262–275 (2008)
8. Han, K., Si, C., Huang, Y., Wang, L., Tan, T.: Generalizable person re-identification via self-supervised batch norm test-time adaption. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 817–825 (2022)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. He, L., Liang, J., Li, H., Sun, Z.: Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7073–7082 (2018)
11. He, L., Liao, X., Liu, W., Liu, X., Cheng, P., Mei, T.: Fastreid: A pytorch toolbox for general instance re-identification. arXiv preprint arXiv:2006.02631 (2020)
12. He, S., Luo, H., Chen, W., Zhang, M., Zhang, Y., Wang, F., Li, H., Jiang, W.: Multi-domain learning and identity mining for vehicle re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 582–583 (2020)
13. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)

14. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17. pp. 91–102. Springer (2011)
15. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE international conference on computer vision. pp. 1501–1510 (2017)
16. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pp. 448–456. pmlr (2015)
17. Jiang, K., Zhang, T., Zhang, Y., Wu, F., Rui, Y.: Self-supervised agent learning for unsupervised cross-domain person re-identification. IEEE Transactions on Image Processing **29**, 8549–8560 (2020)
18. Jiao, B., Liu, L., Gao, L., Lin, G., Yang, L., Zhang, S., Wang, P., Zhang, Y.: Dynamically transformed instance normalization network for generalizable person re-identification. In: European Conference on Computer Vision. pp. 285–301. Springer (2022)
19. Jin, X., Lan, C., Zeng, W., Chen, Z., Zhang, L.: Style normalization and restitution for generalizable person re-identification. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3143–3152 (2020)
20. Lee, G., Lee, S., Kim, D., Shin, Y., Yoon, Y., Ham, B.: Camera-driven representation learning for unsupervised domain adaptive person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11453–11462 (2023)
21. Li, Q., Gong, S.: Mitigate domain shift by primary-auxiliary objectives association for generalizing person reid. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 394–403 (2024)
22. Li, W., Wang, X.: Locally aligned feature transforms across views. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3594–3601 (2013)
23. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 152–159 (2014)
24. Li, Y., Wang, N., Shi, J., Hou, X., Liu, J.: Adaptive batch normalization for practical domain adaptation. Pattern Recognition **80**, 109–117 (2018)
25. Lin, C.S., Cheng, Y.C., Wang, Y.C.F.: Domain generalized person re-identification via cross-domain episodic learning. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 6758–6763 (2021). https://doi.org/10.1109/ICPR48806.2021.9413013
26. Liu, J., Zha, Z.J., Chen, D., Hong, R., Wang, M.: Adaptive transfer network for cross-domain person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7202–7211 (2019)
27. Loy, C.C., Xiang, T., Gong, S.: Time-delayed correlation analysis for multi-camera activity understanding. International Journal of Computer Vision **90**, 106–129 (2010)
28. Luo, H., Jiang, W., Gu, Y., Liu, F., Liao, X., Lai, S., Gu, J.: A strong baseline and batch normalization neck for deep person re-identification. IEEE Transactions on Multimedia **22**(10), 2597–2609 (2019)
29. Luo, P., Ren, J., Peng, Z., Zhang, R., Li, J.: Differentiable learning-to-normalize via switchable normalization. arXiv preprint arXiv:1806.10779 (2018)

30. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
31. Nam, H., Kim, H.E.: Batch-instance normalization for adaptively style-invariant neural networks. Advances in Neural Information Processing Systems **31** (2018)
32. Ni, H., Song, J., Luo, X., Zheng, F., Li, W., Shen, H.T.: Meta distribution alignment for generalizable person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2487–2496 (2022)
33. Pan, X., Luo, P., Shi, J., Tang, X.: Two at once: Enhancing learning and generalization capacities via ibn-net. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 464–479 (2018)
34. Qi, L., Wang, L., Shi, Y., Geng, X.: A novel mix-normalization method for generalizable multi-source person re-identification. IEEE Transactions on Multimedia (2022)
35. Ren, M., He, L., Liao, X., Liu, W., Wang, Y., Tan, T.: Learning instance-level spatial-temporal patterns for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14930–14939 (2021)
36. Seo, S., Suh, Y., Kim, D., Kim, G., Han, J., Han, B.: Learning to optimize domain specific normalization for domain generalization. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16. pp. 68–83. Springer (2020)
37. Song, J., Yang, Y., Song, Y.Z., Xiang, T., Hospedales, T.M.: Generalizable person re-identification by domain-invariant mapping network. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 719–728 (2019)
38. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
39. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 79–88 (2018)
40. Wu, Y., Chi, Z., Wang, Y., Plataniotis, K.N., Feng, S.: Test-time domain adaptation by learning domain-aware batch normalization. arXiv preprint arXiv:2312.10165 (2023)
41. Wu, Y., Johnson, J.: Rethinking" batch" in batchnorm. arXiv preprint arXiv:2105.07576 (2021)
42. Xiao, T., Li, S., Wang, B., Lin, L., Wang, X.: End-to-end deep learning for person search. arXiv preprint arXiv:1604.01850 **2**(2), 4 (2016)
43. Xu, B., He, L., Liao, X., Liu, W., Sun, Z., Mei, T.: Black re-id: A head-shoulder descriptor for the challenging problem of person re-identification. In: Proceedings of the 28th ACM International Conference on Multimedia. pp. 673–681 (2020)
44. Xu, B., Liang, J., He, L., Sun, Z.: Mimic embedding via adaptive aggregation: Learning generalizable person re-identification. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIV. pp. 372–388. Springer (2022)
45. Zhang, P., Dou, H., Yu, Y., Li, X.: Adaptive cross-domain learning for generalizable person re-identification. In: European Conference on Computer Vision. pp. 215–232. Springer (2022)
46. Zhao, Y., Zhong, Z., Yang, F., Luo, Z., Lin, Y., Li, S., Sebe, N.: Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6277–6286 (2021)

47. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proceedings of the IEEE international conference on computer vision. pp. 1116–1124 (2015)
48. Zheng, W., Gong, S., Xiang, T.: Associating groups of people. In: British Machine Vision Conference (2009)
49. Zheng, Z., Ruan, T., Wei, Y., Yang, Y., Mei, T.: Vehiclenet: Learning robust visual representation for vehicle re-identification. IEEE Transactions on Multimedia **23**, 2683–2693 (2020)
50. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE international conference on computer vision. pp. 3754–3762 (2017)