Style Normalization and Restitution for Generalizable Person Re-identification

Xin Jin[1]* Cuiling Lan[2]† Wenjun Zeng[2] Zhibo Chen[1]† Li Zhang[3]
[1] University of Science and Technology of China [2] Microsoft Research Asia, Beijing, China [3] University of Oxford
jinxustc@mail.ustc.edu.cn {culan,wezeng}@microsoft.com chenzhibo@ustc.edu.cn lz@robots.ox.ac.uk

Abstract

Existing fully-supervised person re-identification (ReID) methods usually suffer from poor generalization capability caused by domain gaps. The key to solving this problem lies in filtering out ***identity-irrelevant*** *interference and learning domain-invariant person representations. In this paper, we aim to design a generalizable person ReID framework which trains a model on source domains yet is able to generalize/perform well on target domains. To achieve this goal, we propose a simple yet effective* ***S****tyle* ***N****ormalization and* ***R****estitution (SNR) module. Specifically, we filter out style variations (e.g., illumination, color contrast) by Instance Normalization (IN). However, such a process inevitably removes discriminative information. We propose to distill identity-relevant feature from the removed information and restitute it to the network to ensure high discrimination. For better disentanglement, we enforce a dual causality loss constraint in SNR to encourage the separation of identity-relevant features and identity-irrelevant features. Extensive experiments demonstrate the strong generalization capability of our framework. Our models empowered by the SNR modules significantly outperform the state-of-the-art domain generalization approaches on multiple widely-used person ReID benchmarks, and also show superiority on unsupervised domain adaptation.*

1. Introduction

Person re-identification (ReID) aims at matching/identifying a specific person across cameras, times, and locations. It facilitates many applications and has attracted a lot of attention.

Abundant approaches have been proposed for supervised person ReID, where a model is trained and tested on different splits of the same dataset [51, 38, 54, 8, 34, 53, 18, 17]. They typically focus on addressing the challenge of geometric misalignment among images caused by diversity of poses/viewpoints. In general, they perform well on the trained dataset but suffer from signicant performance degradation (poor generalization capability) when testing on a previously unseen dataset. There are usually style discrepancies across domains/datasets which hinder the achievement of high generalization capability. Figure 1 shows some example images[1] from different ReID datasets. The person images are captured by different cameras under different environments (*e.g.*, lighting, seasons). They present a large style discrepancy in terms of illumination, hue, color contrast and saturation, quality/resolution, *etc*. For a ReID

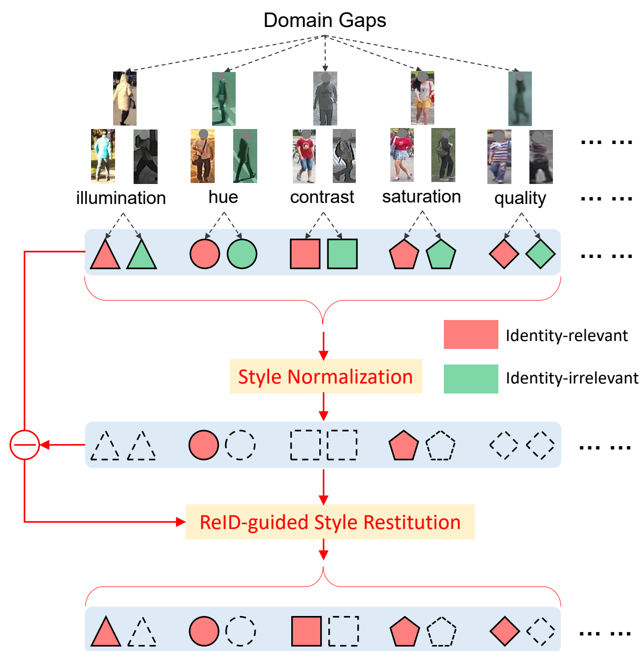

Figure 1: Illustration of motivation and our idea. Person images captured from different cameras and environments present style variations which result in domain gaps. We use style normalization (with Instance Normalization) to alleviate style variations. However, this also results in the loss of some discriminative (identity-relevant) information. We propose to further restitute such information from the residual of the original information and the normalized information for generalizable and discriminative person ReID.

*This work was done when Xin Jin was an intern at Microsoft Research Asia.
†Corresponding Author.

[1] All faces in the images are masked for anonymization.

system, we expect it to be able to identify the same person even captured in different environments, and distinguish between different people even if their appearance are similar. Both generalization and discrimination capabilities, although seemly conflicting with each other, are very important for robust ReID.

Considering the existence of domain gaps and poor generalization capability, fully-supervised approaches or settings are not practical for real-world widespread ReID system deployment, where the onsite manual annotation on the target domain data is expensive and hardly feasible. In recent years, some unsupervised domain adaptation (UDA) methods have been studied to adapt a ReID model from source to target domain [43, 40, 26, 33, 5, 50, 48]. UDA models update using *unlabeled* target domain data, emancipating the labelling efforts. However, data collection and model update are still required, adding additional cost.

We mainly focus on the more economical and practical domain generalizable person ReID. Domain generalization (DG) aims to design models that are generalizable to previously unseen domains [31, 16, 36], without having to access the target domain data and labels, and without requiring model updating. Most DG methods assume that the source and target domains have the same label space [19, 22, 31, 35] and they are not applicable to ReID since the target domains for ReID typically have a different label space from the source domains. Generalizable person ReID is challenging which aims to achieve high discrimination capability on *unseen* target domain that may have large domain discrepancy. The study on domain generalizable ReID is rare [36, 16] and remains an open problem. Jia *et al*. [16] and Zhou *et al*. [61] integrate Instance Normalization (IN) in the networks to alleviate the domain discrepancy due to appearance style variations. However, IN inevitably results in the loss of some discriminative features [14, 32], hindering the achievement of high efficiency ReID.

In this paper, we aim to design a generalizable ReID framework which achieves both high generalization capability and discrimination capability. The key is to find a way to disentangle the identity-relevant features and the identity-irrelevant features (*e.g*., image styles). Figure 1 illustrates our main idea. Considering the domain gaps among image samples, we perform style normalization by means of IN to eliminate style variations. However, the normalization inevitably discards some discriminative information and thus may hamper the ReID performance. From the residual information (which is the difference between the original information and the normalized information), we further distill the identity-relevant information as a compensation to the normalized information. Figure 2 shows our framework with the proposed Style Normalization and Restitution (SNR) modules embedded. To better disentangle the identity-relevant features from the residual, a dual causality loss constraint is added by ensuring the features after restitution of identity-relevant features to be more discriminative, and the features after compensation of identity-irrelevant features to be less discriminative.

We summarize our main contributions as follows:

- We propose a practical domain generalizable person ReID framework that generalizes well on previously unseen domains/datasets. Particularly, we design a Style Normalization and Restitution (SNR) module. SNR is simple yet effective and can be used as a *plug-and-play* module for existing ReID architectures to enhance their generalization capabilities.
- To facilitate the restitution of identity-relevant features from those discarded in the style normalization phase, we introduce a dual causality loss constraint in SNR for better feature disentanglement.

We validate the effectiveness of the proposed SNR module on multiple widely-used benchmarks and settings. Our models significantly outperform the state-of-the-art domain generalizable person ReID approaches and can also boost the performance of unsupervised domain adaptation for ReID.

2. Related Work

Supervised Person ReID. In the last decade, fully-supervised person ReID has achieved great progress, especially for deep learning based approaches [38, 21, 54, 8, 34, 53]. These methods usually perform well on the testing set of the source datasets but generalize poorly to previously unseen domains/datasets due to the style discrepancy across domains. This is problematic especially in practical applications, where the target scenes typically have different styles from the source domains and there is no readily available target domain data or annotation for training.

Unsupervised Domain Adaptation (UDA) for Person ReID. When the target domain data is accessible, even without annotations, it can be explored for the domain adaptation for enhancing the ReID performance. This requires target domain data collection and model updating. UDA-based ReID methods can be roughly divided into three categories: *style transfer* [3, 44, 26], *attribute recognition* [43, 49, 33], and *target-domain pseudo label estimation* [5, 37, 58, 40, 52, 50]. For pseudo label estimation, recently, Yu *et al*. propose a method called **m**ulti**la**bel **r**eference learning (MAR) which evaluates the similarity of a pair of images by comparing them to a set of known reference persons to mine hard negative samples [50].

Our proposed domain generalizable SNR module can also be combined with the UDA methods (*e.g*., by plugging into the UDA backbone) to further enhance the ReID performance. We will demonstrate its effectiveness by combining it with the UDA approach of MAR in Subsection 4.5.

Domain Generalization (DG). Domain Generalization is

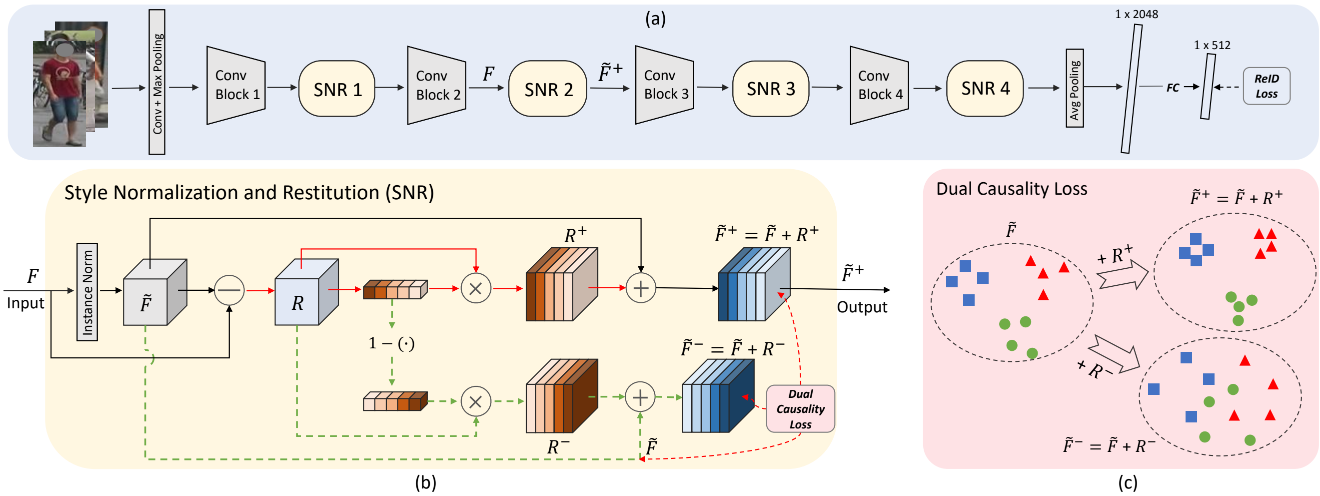

Figure 2: Overall flowchart. (a) Our generalizable person ReID network with the proposed Style Normalization and Restitution (SNR) module being plugged in after some convolutional blocks. Here, we use ResNet-50 as our backbone for illustration. (b) Proposed SNR module. Instance Normalization (IN) is used to eliminate some style discrepancies followed by identity-relevant feature restitution (marked by red solid arrows). Note the branch with dashed green line is only used for enforcing loss constraint and is discarded in inference. (c) Dual causality loss constraint encourages the disentanglement of a residual feature R to identity-relevant one (R^+) and identity-irrelevant one (R^-), which enhances and decreases, respectively, the discrimination by adding them to the style normalized feature $\widetilde{F}$ (see Section 3.1).

a challenging problem of learning models that is generalizable to unseen domains [31, 35]. Muandet *et al.* learn an invariant transformation by minimizing the dissimilarity across source domains [31]. A learning-theoretic analysis shows that reducing dissimilarity improves the generalization ability on new domains. CrossGrad [35] generates pseudo training instances by pertubations in the loss gradients of the domain classifier and category classifier respectively. Most DG methods assume that the source and target domains have the same label space. However, ReID is an open-set problem where the target domains typically have different identities from the source domains, so that the general DG methods could not be directly applied to ReID.

Recently, a strong baseline for domain generalizable person ReID is proposed by simply combing multiple source datasets and training a single CNN [20]. Song *et al.* [36] propose a generalizable person ReID framework by using a meta-learning pipeline to make the model domain invariant. To overcome the inconsistency of label spaces among different datasets, it maintains a training datasets shared memory bank. Instance Normalization (IN) has been widely used in image style transfer [14, 42] and proved that it actually performs a kind of style normalization [32, 14]. Jia *et al.* [16] and Zhou *et al.* [61] apply this idea to ReID to alleviate the domain discrepancy and boost the generalization capability. However, IN inevitably discards some discriminative information. In this paper, we study how to design a generalibale ReID framework that can exploit the merit of IN while avoiding the loss of discriminative information.

3. Proposed Generalizable Person ReID

We aim at designing a generalizable and robust person ReID framework. During the training, we have access to one or several annotated source datasets. The trained model will be deployed directly to unseen domains/datasets and is expected to work well with high generalization capability.

Figure 2 shows the overall flowchat of our framework. Particularly, we propose a Style Normalization and Restitution (SNR) module to boost the generalization and discrimination capability of ReID models especially on unseen domains. SNR can be used as a plug-and-play module for existing ReID networks. Taking the widely used ReID network of ResNet-50 [10, 1, 28] as an example (see Figure 2(a)), SNR module is added after each convolutional block. In the SNR module, we first eliminate style discrepancy among samples by Instance Normalization (IN). Then, a dedicated restitution step is proposed to distill identity-relevant (discriminative) features from those previsouly discarded by IN, and add them to the normalized features. Moreover, for the SNR module, we design a dual causality loss constraint to facilitate the distillation of identity-relevant features from the information discarded by IN.

3.1. Style Normalization and Restitution (SNR)

Person images for ReID could be captured by different cameras under different scenes and environments (*e.g.*, indoor/outdoors, shopping malls, street, sunny/cloudy). As shown in Figure 1, they present style discrepancies (*e.g.*, in illumination, hue, contrast, saturation, quality), especially for samples from two different datasets/domains. Domain discrepancy between the source and target domain generally hinders the generalization capability of ReID models.

A learning-theoretic analysis shows that reducing dissimilarity improves the generalization ability on new domains [31]. Instance Normalization (IN) performs some kinds of style normalization which reduces the discrepancy/dissimilarity among instances/samples [14, 32], so it can enhance the generalization ability of networks [32, 16, 61]. However, IN inevitably removes some discriminative information and results in weaker discrimination capability [32]. To address this problem, we propose to restitute the task-specific discriminative features from the IN removed information, by disentangling it into identity-relevant features and identity-irrelevant features with a dual causality loss constraint (see Figure 2(b)). We elaborate on the designed SNR module hereafter.

For an SNR module, we denote the input (which is a feature map) by $F \in \mathbb{R}^{h\times w\times c}$ and the output by $\widetilde{F}^+ \in \mathbb{R}^{h\times w\times c}$, where h, w, c denote the height, width, and number of channels, respectively.

Style Normalization Phase. In SNR, we first try to reduce the domain discrepancy on the input features by performing Instance Normalization [41, 4, 42, 14] as

$$\widetilde{F} = \mathrm{IN}(F) = \gamma(\frac{F-\mu(F)}{\sigma(F)}) + \beta, \tag{1}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation computed across spatial dimensions independently for each channel and each *sample/instance*, γ, $\beta \in \mathbb{R}^c$ are parameters learned from data. IN could filter out some instance-specific style information from the content. With IN taking place in the feature space, Huang *et al*. [14] have argued and experimentally shown that IN has more profound impacts than a simple contrast normalization and it performs a form of *style normalization* by normalizing feature statistics.

Style Restitution Phase. IN reduces style discrepancy and boosts the generalization capability. However, with the mathematical operations being deterministic and task-irrelevant, it inevitably discards some discriminative (task-relevant) information for ReID. We propose to restitute the identity-relevant feature to the network by distilling it from the residual feature R. R is defined as

$$R = F - \widetilde{F}, \tag{2}$$

which denotes the difference between the original input feature F and the style normalized feature $\widetilde{F}$.

Given R, we further disentangle it into two parts: identity-relevant feature $R^+ \in \mathbb{R}^{h\times w\times c}$ and identity-irrelevant feature $R^- \in \mathbb{R}^{h\times w\times c}$ through masking R by a learned channel attention vector $\mathbf{a} = [a_1, a_2, \cdots, a_c] \in \mathbb{R}^c$:

$$\begin{aligned} R^+(:,:,k) &= a_k R(:,:,k), \\ R^-(:,:,k) &= (1-a_k) R(:,:,k), \end{aligned} \tag{3}$$

where $R(:,:,k) \in \mathbb{R}^{h\times w}$ denotes the k^{th} channel of feature map R, $k = 1, 2, \cdots, c$. We expect the channel attention vector $\mathbf{a}$ to enable the adaptive distillation of the identity-relevant features for restitution, and derive it by SE-like [13] channel attention as

$$\mathbf{a} = g(R) = \sigma(\mathrm{W}_2\delta(\mathrm{W}_1 pool(R))), \tag{4}$$

which consists of a global average pooling layer followed by two FC layers that are parameterized by $\mathrm{W}_2 \in \mathbb{R}^{(c/r)\times c}$ and $\mathrm{W}_1 \in \mathbb{R}^{c\times(c/r)}$ which are followed by ReLU activation function $\delta(\cdot)$ and sigmoid activation function $\sigma(\cdot)$, respectively. To reduce the number of parameters, a dimension reduction ratio r is used and is set to 16.

By adding the distilled identity-relevant feature R^+ to the style normalized feature $\widetilde{F}$, we obtain the output feature $\widetilde{F}^+$ of the SNR module as

$$\widetilde{F}^+ = \widetilde{F} + R^+. \tag{5}$$

Dual Causality Loss Constraint. In order to facilitate the disentanglement of identity-relevant feature and identity-irrelevant feature, we design a dual causality loss constraint by comparing the discrimination capability of features *before* and *after* the restitution. As illustrated in Figure 2(c), the main idea is that: after restituting the identity-relevant feature R^+ to the normalized feature $\widetilde{F}$, the feature becomes more discriminative; On the other hand, after restituting the identity-irrelevant feature R^- to the normalized feature $\widetilde{F}$, the feature should become less discriminative. We achieve this by defining a dual causality loss $\mathcal{L}_{SNR}$ which consists of *clarification loss* $\mathcal{L}^+_{SNR}$ and *destruction loss* $\mathcal{L}^-_{SNR}$, *i.e.*, $\mathcal{L}_{SNR} = \mathcal{L}^+_{SNR} + \mathcal{L}^-_{SNR}$.

Within a mini-batch, we sample three images, *i.e.*, an anchor sample a, a positive sample p that has the same identity as the anchor sample, and a negative sample n that has a different identity from the anchor sample. For simplicity, we differentiate the three samples by subscript. For example, the style normalized feature of sample a is denoted by $\widetilde{F}_a$.

Intuitively, adding the identity-relevant feature R^+ to the normalized feature $\widetilde{F}$, which we refer to as *enhanced feature* $\widetilde{F}^+ = \widetilde{F} + R^+$, results in better discrimination capability — the sample features with same identities are closer and those with different identities are farther apart. We calculate the distances between samples on a spatially average pooled feature to avoid the distraction caused by spatial misalignment among samples (*e.g.*, due to different poses/viewpoints). We denote the spatially average pooled feature of $\widetilde{F}$ and $\widetilde{F}^+$ as $\widetilde{\mathbf{f}} = pool(\widetilde{F})$, $\widetilde{\mathbf{f}}^+ = pool(\widetilde{F}^+)$, respectively. The *clarification loss* is thus defined as

$$\begin{aligned} \mathcal{L}^+_{SNR} = & Softplus(d(\widetilde{\mathbf{f}}^+_a, \widetilde{\mathbf{f}}^+_p) - d(\widetilde{\mathbf{f}}_a, \widetilde{\mathbf{f}}_p)) \\ & + Softplus(d(\widetilde{\mathbf{f}}_a, \widetilde{\mathbf{f}}_n) - d(\widetilde{\mathbf{f}}^+_a, \widetilde{\mathbf{f}}^+_n)), \end{aligned} \tag{6}$$

where $d(\mathbf{x}, \mathbf{y})$ denotes the distance between $\mathbf{x}$ and $\mathbf{y}$ which is defined as $d(\mathbf{x}, \mathbf{y}) = 0.5 - \mathbf{x}^{\mathrm{T}}\mathbf{y}/(2\|\mathbf{x}\|\|\mathbf{y}\|)$. $Softplus(\cdot) = ln(1 + exp(\cdot))$ is a monotonically increasing function that aims to reduce the optimization difficulty by avoiding negative loss values.

On the other hand, we expect that the adding of the identity-irrelevant feature R^- to the normalized feature $\widetilde{F}$, which we refer to as *contaminated feature* $\widetilde{F}^- = \widetilde{F} + R^-$, could decrease the discrimination capability. In comparison with the normalized feature $\widetilde{F}$ before the compensation, we expect that adding R^- would push the sample features with same identities farther apart and pull those with different identities closer. We denote the spatially average pooled feature of $\widetilde{F}^-$ as $\widetilde{\mathbf{f}}^- = pool(\widetilde{F}^-)$. The *destruction loss* is:

$$\begin{aligned} \mathcal{L}_{SNR}^- = & Softplus(d(\widetilde{\mathbf{f}}_a, \widetilde{\mathbf{f}}_p) - d(\widetilde{\mathbf{f}}_a^-, \widetilde{\mathbf{f}}_p^-)) \\ & + Softplus(d(\widetilde{\mathbf{f}}_a^-, \widetilde{\mathbf{f}}_n^-) - d(\widetilde{\mathbf{f}}_a, \widetilde{\mathbf{f}}_n)). \end{aligned} \tag{7}$$

3.2. Joint Training

We use the commonly used ResNet-50 as a base ReID network and insert the proposed SNR module after each convolution block (in total four convolution blocks/stages)(see Figure 2(a)). We train the entire network in an end-to-end manner. The overall loss is

$$\mathcal{L} = \mathcal{L}_{ReID} + \sum_{b=1}^{4} \lambda_b \mathcal{L}_{SNR}^b, \tag{8}$$

where $\mathcal{L}_{SNR}^b$ denotes the dual causality loss for the b^{th} SNR module. $\mathcal{L}_{ReID}$ denotes the widely-used ReID Loss (classification loss [39, 7], and triplet loss with batch hard mining [11]) on the ReID feature vectors. λ_b is a weight which controls the relative importance of the regularization at stage b. In considering that the features of stage 3 and 4 are more relevant to the task (high-level semantics), we experimentally set λ_3, λ_4 to 0.5, and λ_1,λ_2 to 0.1.

4. Experiments

In this section, we first describe the datasets and evaluation metrics in Subsection 4.1. Then, for generalizable ReID, we validate the effectiveness of SNR in Subsection 4.2 and study its design choices in Subsection 4.3. We conduct visualization analysis in Subsection 4.4. Subsection 4.5 shows the comparisons of our schemes with the state-of-the-art approaches for both generalizable person ReID and unsupervised domain adapation ReID, respectively. In Subsection 4.6, we further validate the effectiveness of applying the SNR modules to another backbone network and to cross modality (Infrared-RGB) person ReID.

We use ResNet-50 [10, 1, 53, 28] as our base network for both baselines and our schemes. We build a strong baseline *Baseline* with some commonly used tricks integrated.

4.1. Datasets and Evaluation Metrics

To evaluate the generalization ability of our approach and to be consistent with what were done in prior works for performance comparisons, we conduct extensive experiments on commonly used public ReID datasets, including Market1501 [55], DukeMTMC-reID [57], CUHK03 [24], the large-scale MSMT17 [44], and four small-scale ReID datasets of PRID [12], GRID [27], VIPeR [9], and i-LIDS [45]. We denote Market1501 by M, DukeMTMC-reID by Duke or D, and CUHK03 by C for simplicity.

We follow common practices and use the cumulative matching characteristics (CMC) at Rank-1, and mean average precision (mAP) to evaluate the performance.

4.2. Ablation Study

We perform comprehensive ablation studies to demonstrate the effectiveness of the SNR module and its dual causality loss constraint. We mimic the real-world scenario for generalizable person ReID, where a model is trained on some source dataset(s) A while tested on previously unseen dataset B. We denote this as A→B. We have several experimental settings to evaluate the generalization capability, *e.g.*, Market1501→Duke and others, Duke→Market1501 and others, M+D+C+MSMT17→others. Our settings cover both single source dataset for training and multiple source datasets for training.

Effectiveness of Our SNR. Here we compare several schemes. ***Baseline***: a strong baseline based on ResNet-50. ***Baseline-A-IN***: a naive model where we replace all the Batch Normalization(BN) [15] layers in *Baseline* by Instance Normalization(IN). ***Baseline-IBN***: Similar to IBN-Net (IBN-b) [32] and OSNet [61], we add IN only to the last layers of Conv1 and Conv2 blocks of *Baseline* respectively. ***Baseline-A-SN***: a model where we replace all the BN layers in *Baseline* by Switchable Normalization (SN). SN [29] can be regarded as an adaptive ensemble version of normalization techniques of IN, BN, and LN (Layer Normalization) [2]. ***Baseline-IN***: four IN layers are added after the first four convolutional blocks/stages of *Baseline* respectively. ***Baseline-SNR***: our final scheme where four SNR modules are added after the first four convolutional blocks/stages of *Baseline* respectively (see Figure 2(a)). We also refer to it as ***SNR*** for simplicity. Table 1 shows the results. We have the following observations/conclusions:

1) *Baseline-A-IN* improves *Baseline* by **4.3%** in mAP for Market1501→Duke, and **4.7%** in mAP for Duke→Market1501. Other IN-related baselines also bring gains, which demonstrates the effectiveness of IN for improving the generalization capability for ReID. But, IN also inevitably discards some discriminative (identity-relevant) information and we can see it clearly decreases the performance of *Baseline-A-IN*, *Baseline-IBN* and *Baseline-IN* for the same-domain ReID (*e.g.*, Market1501→Market1501).

Table 1: Performance (%) comparisons of our scheme and others to demonstrate the effectiveness of our SNR module for generalizable person ReID. The rows denote source dataset(s) for training and the columns correspond to different target datasets for testing. We mask the results of supervised ReID by gray where the testing domain has been seen in training. Due to space limitation, we only show a portion of the results here and more comparisons can be found in **Supplementary**.

Source	Method	Target: Market1501		Target: Duke		Target: PRID		Target: GRID		Target: VIPeR		Target: iLIDs	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Market1501 (M)	Baseline	82.8	93.2	19.8	35.3	13.7	6.0	25.8	16.0	37.6	28.5	61.5	53.3
	Baseline-A-IN	75.3	89.8	24.1	42.7	33.9	21.0	35.6	27.2	38.1	29.1	64.2	55.0
	Baseline-IBN	81.1	92.2	21.5	39.2	19.1	12.0	27.5	19.2	32.1	23.4	58.3	48.3
	Baseline-A-SN	83.2	93.9	20.1	38.0	35.4	25.0	29.0	22.0	32.2	23.4	53.4	43.3
	Baseline-IN	79.5	90.9	25.1	44.9	35.0	25.0	35.7	27.8	35.1	27.5	64.0	54.2
	Baseline-SNR (Ours)	**84.7**	**94.4**	**33.6**	**55.1**	**42.2**	**30.0**	**36.7**	**29.0**	**42.3**	**32.3**	**65.6**	**56.7**
Duke (D)	Baseline	21.8	48.3	71.2	83.4	15.7	11.0	14.5	8.8	37.0	26.9	68.3	58.3
	Baseline-A-IN	26.5	56.0	64.5	78.9	38.6	29.0	19.6	13.6	35.1	27.2	67.4	56.7
	Baseline-IBN	24.6	52.5	69.5	81.4	27.4	19.0	19.9	12.0	32.8	23.4	63.5	61.7
	Baseline-A-SN	25.3	55.0	**73.0**	**85.9**	41.4	32.0	18.8	12.8	31.3	24.1	64.8	63.3
	Baseline-IN	27.2	58.5	68.9	80.4	40.5	27.0	20.3	13.2	34.6	26.3	70.6	65.0
	Baseline-SNR (Ours)	**33.9**	**66.7**	72.9	84.4	**45.4**	**35.0**	**35.3**	**26.0**	**41.2**	**32.6**	**79.3**	**68.7**
M + D + CUHK03 + MSMT17	Baseline	72.4	88.7	70.1	83.8	39.0	28.0	29.6	20.8	52.1	41.5	89.0	85.0
	Baseline-SNR (Ours)	**82.3**	**93.4**	**73.2**	**85.5**	**60.0**	**49.0**	**41.3**	**30.4**	**65.0**	**55.1**	**91.9**	**87.0**

Baseline-A-SN learns the combination weights of IN, BN, and LN in the training dataset and thus has superior performance in the same domain, but it does not have dedicated design for boosting the generalization capability.
2) Thanks to the compensation of the identity-relevant information through the proposed *restitution step*, our final scheme *Baseline-SNR* achieves superior generalization capability, which significantly outperforms all the baseline schemes. In particular, *Baseline-SNR* outperforms *Baseline-IN* by **8.5%**, **6.7%**, **15.0%** in mAP for M→D, D→M, and D→GRID, respectively.
3) The generalization performance on previously unseen target domain increases consistently as the number of source datasets increases. When all the four source datasets are used (the large-scale MSMT17 [44] also included), we have a very strong baseline (*i.e.*, 52.1% in mAP on VIPeR dataset vs. 37.6% when Market1501 alone is used as source). Interestingly, our method still significantly outperforms the strong baseline *Baseline*, even by **21.0%** in mAP on PRID dataset, demonstrating SNR's effectiveness.
4) The performance of different schemes with respects to PRID/GRID varies greatly and the mAPs are all relatively low, which is caused by the large style discrepancy between PRID/GRID and other datasets. For such challenging cases, our scheme still outperforms *Baseline-IN* significantly by **7.2%** and **4.9%** in mAP for M→PRID and D→PRID, respectively.
5) For supervised ReID (masked by gray), our scheme also clearly outperforms *Baseline* by **1.9%** and **1.7%** in mAP for M→M and D→D, respectively. That is because there is also style discrepancy within the source domain.

Influence of Dual Causality Loss Constraint. We study the effectiveness of the proposed dual causality loss $\mathcal{L}_{SNR}$ which consists of *clarification loss* $\mathcal{L}_{SNR}^{+}$ and *destruction loss* $\mathcal{L}_{SNR}^{-}$. Table 2a shows the results. Our final scheme *SNR* with the dual causality loss $\mathcal{L}_{SNR}$ outperforms that without such constraints (*i.e.*, scheme *SNR w/o* $\mathcal{L}_{SNR}$) by **7.5%** and **4.7%** in mAP for M→D and D→M, respectively. Such constraints facilitate the disentanglement of identity-relevant/identity-irrelevant features. In addition, both the clarification loss $\mathcal{L}_{SNR}^{+}$ and the destruction loss $\mathcal{L}_{SNR}^{-}$, are vital to SNR and they are complementary and jointly contribute to a superior performance.

Complexity. The model size of our final scheme *SNR* is very similar to that of *Baseline* (24.74 M vs. 24.56 M).

4.3. Design Choices of SNR

Which Stage to Add SNR? We compare the cases of adding a single SNR module to a different convolutional block/stage, and to all the four stages (*i.e.*, stage-1 $\sim$ 4) of the ResNet-50 (see Figure 2(a)). The module is added after the last layer of a convolutional block/stage. As Table 2b shows, in comparison with *Baseline*, the improvement from adding SNR is significant on stage-3 and stage-4 and is a little smaller on stage-1 and stage-2. When SNR is added to all the four stages, we achieve the best performance.

Influence of Disentanglement Design. In our SNR module, as described in (3)(4) of Subsection 3.1, we use $g(\cdot)$, and its complementary one $1 - g(\cdot)$ as masks to extract identity-relevant feature R^{+} and identity-irrelevant feature R^{-} from the residual feature R. Here, we study the influence of different disentanglement designs within SNR. ***SNR***$_{conv}$: we disentangle the residual feature R through 1×1 convolutional layer followed by non-liner ReLU activation, *i.e.*, $R^{+} = ReLU(W^{+}R)$, $R^{-} = ReLU(W^{-}R)$. ***SNR***$_{g(\cdot)^2}$: we use two unshared gates $g(\cdot)^{+}$, $g(\cdot)^{-}$ to obtain R^{+} and R^{-} respectively. Table 2c shows the results. We observe that (1) ours outperforms ***SNR***$_{conv}$ by **3.9%** and **4.5%** in mAP for M→D and D→M, respectively, demonstrating the benefit of content-adaptive design; (2) ours outperforms ***SNR***$_{g(\cdot)^2}$ by **2.4%/2.9%** in mAP on the unseen target Duke/Market1501, demonstrating the benefit of the design which encourages interaction between R^{+} and R^{-}.

Table 2: Effectiveness of dual causality loss constraint (a), and study on design choices of SNR (b) and (c).

(a) Study on the dual causality loss constraint.

Method	M⟶D		D⟶M	
	mAP	Rank-1	mAP	Rank-1
Baseline	19.8	35.3	21.8	48.3
SNR w/o $\mathcal{L}_{SNR}$	26.1	45.0	29.2	57.4
SNR w/o $\mathcal{L}_{SNR}^{+}$	28.8	48.9	30.2	59.8
SNR w/o $\mathcal{L}_{SNR}^{-}$	28.0	48.1	30.3	59.1
SNR	**33.6**	**55.1**	**33.9**	**66.7**

(b) Study on which stage to add SNR.

Method	M⟶D		D⟶M	
	mAP	Rank-1	mAP	Rank-1
Baseline	19.8	35.3	21.8	48.3
stage-1	23.7	42.8	27.6	57.7
stage-2	24.0	44.4	28.6	58.8
stage-3	26.4	46.3	29.5	60.7
stage-4	26.2	45.8	29.4	59.7
stages-all	**33.6**	**55.1**	**33.9**	**66.7**

(c) Disentanglement designs in SNR.

Method	M⟶D		D⟶M	
	mAP	Rank-1	mAP	Rank-1
Baseline	19.8	35.3	21.8	48.3
SNR$_{conv}$	29.7	51.1	29.4	61.7
SNR$_{g(\cdot)^2}$	31.2	52.9	31.0	63.8
SNR	**33.6**	**55.1**	**33.9**	**66.7**

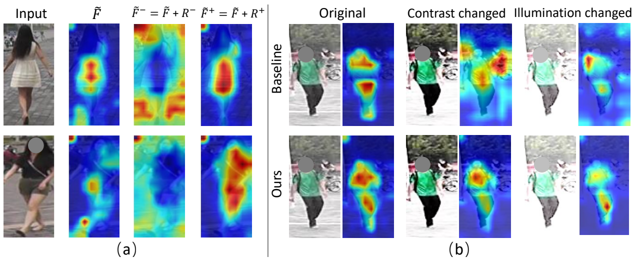

Figure 3: (a) Activation maps of different features within an SNR module (SNR 3). They show SNR can disentangle the identity-relevant/irrelevant features well. (b) Activation maps of our scheme (bottom) and the strong baseline *Baseline* (top) corresponding to images of varied styles. Our maps are more consistent/invariant to style variants.

4.4. Visualization

Feature Map Visualization. To better understand how an SNR module works, we visualize the intermediate feature maps of the third SNR module (SNR 3). Following [61, 56], we get each activation map by summarizing the feature maps along channels followed by a spatial ℓ_2 normalization.

Figure 3(a) shows the activation maps of normalized feature $\widetilde{F}$, enhanced feature $\widetilde{F}^+ = \widetilde{F} + R^+$, and contaminated feature $\widetilde{F}^- = \widetilde{F} + R^-$, respectively. We see that after adding the identity-irrelevant feature R^-, the contaminated feature $\widetilde{F}^-$ has high response mainly on background. In contrast, the enhanced feature $\widetilde{F}^+$ with the restitution of identity-relevant feature R^+ has high responses on regions of the human body, better capturing discriminative regions.

Moreover, in Figure 3(b), we further compare the activation maps $\widetilde{F}^+$ of our scheme and those of the strong baseline scheme *Baseline* by varying the styles of input images (*e.g.*, contrast, illumination, saturation). We can see that, for the images with different styles, the activation maps of our scheme are more consistent/invariant than those of *Baseline*. In contrast, the activation maps of *Baseline* are more disorganized and are easily affected by style variants. These indicate our scheme is more robust to style variations.

Visualization of Feature Distributions. In Figure 4, we visualize the distribution of the features from the 3^{rd} SNR module of our network using t-SNE [30]. They denote the distributions of features for (a) input F, (b) style normalized feature $\widetilde{F}$, and (c) output $\widetilde{F}^+$ of the SNR module. We observe that, (a) before SNR, the extracted features from

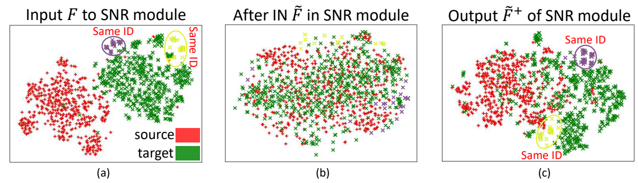

Figure 4: Visualization of distributions of intermediate features before/within/after the SNR module using the tool of t-SNE [30]. 'Red'/'green' nodes: samples from source dataset Market1501/unseen target dataset Duke.

two datasets ('red': source training dataset Market1501; 'green': unseen target dataset Duke) are largely separately distributed and have an obvious *domain gap*. (b) Within the SNR module, after IN, this *domain gap* has been eliminated. But the samples of the same identity ('yellow' and 'purple' colored nodes denote two identities respectively) become dispersive. (c) After the restitution of identity-relevant features, not only has the domain gap of feature distributions been shrunk, but also the feature distribution of samples with same identity become more compact than that in (b).

4.5. Comparison with State-of-the-Arts

Thanks to the capability of reducing style discrepancy and restitution of identity-relevant features, our proposed SNR module can enhance the generalization ability and maintain the discrimintive ability of ReID networks. It can be used for generalizable person ReID, *i.e.*, domain generalization (DG), and can also be used to build the backbone networks for unsupervised domain adaptation (UDA) for person ReID. We evaluate the effectiveness of SNR on both DG-ReID and UDA-ReID by comparing with the state-of-the-art approaches in Table 3.

Domain generalizable person ReID is very attractive in practical applications, which supports "train once and run everywhere". However, there are very few works in this field [36, 16, 61, 20]. Thanks to the exploration of the style normalization and restitution, our scheme *SNR(Ours)* significantly outperforms the second best method *OSNet-IBN* [61] by **6.9%** and **7.8%** for Market1501→Duke and Duke→Market1501 in mAP, respectively. *OSNet-IBN* adds Instance Normalization (IN) to the lower layers of their proposed OSNet following [32]. However, this does not overcome the intrinsic shortcoming of IN and is not optimal.

Song *et al.* [36] also explore domain generalizable

Table 3: Performance (%) comparisons with the state-of-the-art approaches for the Domain Generalizable Person ReID (top rows) and the Unsupervised Domain Adaptation Person ReID (bottom rows), respectively. "(U)" denotes "unlabeled". We mask the schemes that use our *Baseline* and those that use our SNR modules by gray, which provides fair comparison.

	Method	Venue	Source	Target: Duke		Source	Taeget: Market1501	
				mAP	Rank-1		mAP	Rank-1
Domain Generalization (w/o using target data)	OSNet-IBN [61]	ICCV'19	Market1501	26.7	48.5	Duke	26.1	57.7
	Baseline	This work	Market1501	19.8	35.3	Duke	21.8	48.3
	Baseline-IBN [16]	BMVC'19	Market1501	21.5	39.2	Duke	24.6	52.5
	SNR(Ours)	This work	Market1501	**33.6**	**55.1**	Duke	**33.9**	**66.7**
	StrongBaseline [20]	ArXiv'19	MSMT17	43.3	64.5	MSMT17	36.6	64.8
	OSNet-IBN [61]	ICCV'19	MSMT17	45.6	67.4	MSMT17	37.2	66.5
	Baseline	This work	MSMT17	39.1	60.4	MSMT17	33.8	59.9
	SNR(Ours)	This work	MSMT17	**50.0**	**69.2**	MSMT17	**41.4**	**70.1**
Unsupervised Domain Adaptation (using unlabeled target data)	ATNet [26]	CVPR'19	Market1501 + Duke (U)	24.9	45.1	Duke + Market1501 (U)	25.6	55.7
	CamStyle [60]	TIP'19	Market1501 + Duke (U)	25.1	48.4	Duke + Market1501 (U)	27.4	58.8
	ARN [25]	CVPRW'19	Market1501 + Duke (U)	33.4	60.2	Duke + Market1501 (U)	39.4	70.3
	ECN [59]	CVPR'19	Market1501 + Duke (U)	40.4	63.3	Duke + Market1501 (U)	43.0	75.1
	PAST [52]	ICCV'19	Market1501 + Duke (U)	54.3	72.4	Duke + Market1501 (U)	54.6	78.4
	SSG [6]	ICCV'19	Market1501 + Duke (U)	53.4	73.0	Duke + Market1501 (U)	58.3	80.0
	Baseline+MAR [50]	This work	Market1501 + Duke (U)	35.2	56.5	Duke + Market1501 (U)	37.2	62.4
	SNR(Ours)+MAR [50]	This work	Market1501 + Duke (U)	**58.1**	**76.3**	Duke + Market1501 (U)	**61.7**	**82.8**
	MAR [50]	CVPR'19	MSMT17 + Duke (U)	48.0	67.1	MSMT17 + Market1501 (U)	40.0	67.7
	PAUL [48]	CVPR'19	MSMT17 + Duke (U)	53.2	72.0	MSMT17 + Market1501 (U)	40.1	68.5
	Baseline+MAR [50]	This work	MSMT17 + Duke (U)	46.2	66.3	MSMT17 + Market1501 (U)	39.4	66.9
	SNR(Ours) + MAR [50]	This work	MSMT17 + Duke (U)	**61.6**	**78.2**	MSMT17 + Market1501 (U)	**65.9**	**85.5**

person ReID and propose a Domain-Invariant Mapping Network (*DIMN*) to learn the mapping between a person image and its identity classifier with a meta-learning pipeline. We follow [36] and train *SNR* on the same five datasets (M+D+C+CUHK02[23]+CUHK-SYSU[47]). *SNR* outperforms *DIMN* by **14.6%/6.6%/1.2%/11.5%** in mAP and **12.9%/10.9%/1.7%/13.9%** in Rank-1 on the PRID/GRID/VIPeR/i-LIDS.

Unsupervised domain adaptation for ReID has been extensively studied where the unlabeled target data is also used for training. We follow the most commonly-used source→target setting [59, 26, 61, 50, 48] for comparison. We take *SNR* (see Figure 2(a)) as the backbone followed by a domain adaptation strategy MAR [50] for domain adaptation, which we denote as *SNR(Ours)+MAR* [50]. For comparison, we take our strong *Baseline* as the backbone followed by MAR, which we denote as *Baseline+MAR*, to evaluate the effectiveness of the proposed SNR modules. We can see that *SNR(Ours)+MAR* [50] significantly outperforms the second-best UDA ReID method by **3.8%**, **3.4%** in mAP for Market1501+Duke(U)→Duke and Duke+Market1501(U)→Market1501, respectively. In addition, *SNR(Ours)+MAR* outperforms *Baseline+MAR* by **22.9%**, **24.5%** in mAP. Similar trends can be found for MSMT17+Duke(U)→Duke and MSMT17+Market1501(U)→Market1501.

In general, as a plug-and-play module, SNR clearly enhances the generalization capability of ReID networks.

4.6. Extension

Performance on Other Backbone. We add SNR into the recently proposed lightweight ReID network OSNet [61] and observe that by simply inserting SNR modules between the OS-Blocks, the new scheme *OSNet-SNR* outperforms their model *OSNet-IBN* by **5.0%** and **5.5%** in mAP for M→D and D→M, respectively (see **Supplementary**).

RGB-Infrared Cross-Modality Person ReID. To further demonstrate the capability of SNR in handling images with large style variations, we conduct experiment on a more challenging RGB-Infrared cross-modality person ReID task on benchmark dataset SYSU-MM01 [46]. Our scheme which integrates SNR to *Baseline* outperforms *Baseline* significantly by **8.4%**, **8.2%**, **11.0%**, and **11.5%** in mAP under 4 different settings, and also achieves the state-of-the-art performance (see **Supplementary** for more details).

5. Conclusion

In this paper, we propose a generalizable person ReID framework to enable effective ReID. A Style Normalization and Restitution (SNR) module is introduced to exploit the merit of Instance Normalization (IN) that filters out the interference from style variations, and restitute the identity-relevant features that are discarded by IN. To efficiently disentangle the identity-relevant and -irrelevant features, we further design a dual causality loss constraint in SNR. Extensive experiments on several benchmarks/settings demonstrate the effectiveness of SNR. Our framework with SNR embedded achieves the best performance on both domain generalization and unsupervised domain adaptation ReID. Moreover, we have also verified SNR's effectiveness on RGB-Infrared ReID task, and on another backbone.

6. Acknowledgments

This work was supported in part by NSFC under Grant U1908209, 61632001 and the National Key Research and Development Program of China 2018AAA0101400.

References

[1] Jon Almazan, Bojana Gajic, Naila Murray, and Diane Larlus. Re-id done right: towards good practices for person re-identification. *arXiv preprint arXiv:1801.05339*, 2018. 3, 5

[2] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016. 5

[3] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *CVPR*, 2018. 2

[4] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. *ICLR*, 2017. 4

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

[6] Yang Fu, Yunchao Wei, Guanshuo Wang, Xi Zhou, Honghui Shi, and Thomas S. Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. *ICCV*, abs/1811.10144, 2019. 8

[7] Yang Fu, Yunchao Wei, Yuqian Zhou, et al. Horizontal pyramid matching for person re-identification. In *AAAI*, 2019. 5

[8] Yixiao Ge, Zhuowan Li, Haiyu Zhao, et al. Fd-gan: Pose-guided feature distilling gan for robust person re-identification. In *NeurIPS*, 2018. 1, 2

[9] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275. Springer, 2008. 5

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. Deep residual learning for image recognition. In *CVPR*, 2016. 3, 5

[11] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 5

[12] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *SCIA*, pages 91–102. Springer, 2011. 5

[13] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 4

[14] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, pages 1501–1510, 2017. 2, 3, 4

[15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ICML*, 2015. 5

[16] Jieru Jia, Qiuqi Ruan, and Timothy M Hospedales. Frustratingly easy person re-identification: Generalizing person re-id in practice. *BMVC*, 2019. 2, 3, 4, 7, 8

[17] Xin Jin, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. Uncertainty-aware multi-shot knowledge distillation for image-based object re-identification. In *AAAI*, 2020. 1

[18] Xin Jin, Cuiling Lan, Wenjun Zeng, Guoqiang Wei, and Zhibo Chen. Semantics-aligned representation learning for person re-identification. In *AAAI*, 2020. 1

[19] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei A Efros, and Antonio Torralba. Undoing the damage of dataset bias. In *ECCV*, pages 158–171. Springer, 2012. 2

[20] Devinder Kumar, Parthipan Siva, Paul Marchwica, and Alexander Wong. Fairest of them all: Establishing a strong baseline for cross-domain person reid. *arXiv preprint arXiv:1907.12016*, 2019. 3, 7, 8

[21] Dangwei Li, Xiaotang Chen, Zhang Zhang, et al. Learning deep context-aware features over body and latent parts for person re-identification. In *CVPR*, 2017. 2

[22] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018. 2

[23] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *CVPR*, pages 3594–3601, 2013. 8

[24] Wei Li, Rui Zhao, Lu Tian, et al. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 5

[25] Yu-Jhe Li, Fu-En Yang, Yen-Cheng Liu, Yu-Ying Yeh, Xiaofei Du, and Yu-Chiang Frank Wang. Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In *CVPR workshops*, 2018. 8

[26] Jiawei Liu, Zheng-Jun Zha, Di Chen, Richang Hong, and Meng Wang. Adaptive transfer network for cross-domain person re-identification. In *CVPR*, 2019. 2, 8

[27] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *IJCV*, 90(1):106–129, 2010. 5

[28] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *CVPR workshops*, 2019. 3, 5

[29] Ping Luo, Jiamin Ren, Zhanglin Peng, Ruimao Zhang, and Jingyu Li. Differentiable learning-to-normalize via switchable normalization. *ICLR*, 2019. 5

[30] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 2008. 7

[31] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *ICML*, pages 10–18, 2013. 2, 3, 4

[32] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, 2018. 2, 3, 4, 5, 7

[33] Lei Qi, Lei Wang, Jing Huo, Luping Zhou, Yinghuan Shi, and Yang Gao. A novel unsupervised camera-aware domain adaptation framework for person re-identification. *ICCV*, 2019. 2

[34] Xuelin Qian, Yanwei Fu, Wenxuan Wang, et al. Pose-normalized image generation for person re-identification. In *ECCV*, 2018. 1, 2

[35] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. *ICLR*, 2018. 2, 3

[36] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *CVPR*, 2019. 2, 3, 7, 8

[37] Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsupervised domain adaptive re-identification: Theory and practice. *arXiv preprint arXiv:1807.11334*, 2018. 2

[38] Chi Su, Jianing Li, Shiliang Zhang, et al. Pose-driven deep convolutional model for person re-identification. In *ICCV*, 2017. 1, 2

[39] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *ECCV*, pages 480–496, 2018. 5

[40] Haotian Tang, Yiru Zhao, and Hongtao Lu. Unsupervised person re-identification with iterative self-supervised domain adaptation. In *CVPR workshops*, 2019. 2

[41] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 4

[42] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *CVPR*, pages 6924–6932, 2017. 3, 4

[43] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *CVPR*, 2018. 2

[44] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer GAN to bridge domain gap for person re-identification. In *CVPR*, 2018. 2, 5, 6

[45] Zheng Wei-Shi, Gong Shaogang, and Xiang Tao. Associating groups of people. In *BMVC*, pages 23–1, 2009. 5

[46] Ancong Wu, Wei-Shi Zheng, Hong-Xing Yu, Shaogang Gong, and Jianhuang Lai. Rgb-infrared cross-modality person re-identification. In *ICCV*, pages 5380–5389, 2017. 8

[47] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv preprint arXiv:1604.01850*, 2:2, 2016. 8

[48] Qize Yang, Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Patch-based discriminative feature learning for unsupervised person re-identification. In *CVPR*, 2019. 2, 8

[49] Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Cross-view asymmetric metric learning for unsupervised person re-identification. In *ICCV*, 2017. 2

[50] Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jian-Huang Lai. Unsupervised person re-identification by soft multilabel learning. In *CVPR*, 2019. 2, 8

[51] Li Zhang, Tao Xiang, and Shaogang Gong. Learning a discriminative null space for person re-identification. In *CVPR*, 2016. 1

[52] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *ICCV*, 2019. 2, 8

[53] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, et al. Densely semantically aligned person re-identification. In *CVPR*, 2019. 1, 2, 5

[54] Haiyu Zhao, Maoqing Tian, Shuyang Sun, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, 2017. 1, 2

[55] Liang Zheng, Liyue Shen, et al. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 5

[56] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person re-identification by probabilistic relative distance comparison. In *CVPR*, 2011. 7

[57] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *ICCV*, 2017. 5

[58] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero-and homogeneously. In *ECCV*, 2018. 2

[59] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *CVPR*, pages 598–607, 2019. 8

[60] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camstyle: A novel data augmentation method for person re-identification. *TIP*, 2018. 8

[61] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, et al. Omni-scale feature learning for person re-identification. *ICCV*, 2019. 2, 3, 4, 5, 7, 8