

Supplementary Material: Style Normalization and Restitution for Generalizable Person Re-identification

Xin Jin¹ Cuiling Lan² Wenjun Zeng² Zhibo Chen¹ Li Zhang³

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

1. Implementation Details

Network Details. We use ResNet-50 [9, 1, 43, 25] as our base network for both baselines and our schemes. We build a strong baseline *Baseline* with some commonly used tricks integrated. Similar to [1, 43, 25], the last spatial down-sample operation in the last Conv block is removed. The proposed SNR module is added after the last layer of each convolutional block/stage of the first four stages. The input image resolution is 256×128 .

Data Augmentation. We use the commonly used data augmentation strategies of random cropping [32, 43], horizontal flipping, and label smoothing regularization [30]. To enhance the generalization ability, we further incorporate some useful data augmentation tricks, such as color jittering and disabling random erasing (REA) [25, 49]. REA hurts models in cross-domain ReID task [25, 14], because REA which masks the regions of training images makes the model learn more knowledge in the training source domain. It causes the model to perform worse in the unseen target domain.

Training Details for Domain Generalization. Following [10], a batch is formed by first randomly sampling P identities. For each identity, we sample K images. Then the batch size is $B = P \times K$. We set $P = 24$ and $K = 4$ (*i.e.*, batch size $B = P \times K = 96$).

We use the Adam optimizer [13] for model optimization. Similar to [25, 43], we first warm up the model for 20 epochs with a linear growth learning rate from 8×10^{-6} to 8×10^{-4} . Then we set the initial learning rate as 8×10^{-4} and optimize the Adam optimizer with a weight decay of 5×10^{-4} . The learning rate is decayed by a factor of 0.5 for every 40 epochs. Our model (here we use ResNet-50 as our backbone) with SNR converges well after training of 280 epochs and we use it for evaluating the generalization performance on target datasets. All our models are implemented on PyTorch and trained on a single 32G NVIDIA-V100 GPU.

Training Details for Domain Adaptation. For unsupervised domain adaptation person ReID, we combine our network with the unsupervised ReID approach MAR [41] for fine-tuning on the unlabelled target domain data. MAR [41] plays the role of assigning pseudo labels by hard negative mining, which facilitates the fine-tuning of base network. Similar to [41], during the fine-tuning, both source labeled data and target unlabelled data are jointly used for effective joint training. Specifically, during fine-tuning, a training batch of size 96 is composed of 1) labeled source data (size $B_1 = P \times K = 48$, where $P = 12, K = 4$), and 2) unlabeled target data (size $B_2 = 48$). For the labeled source data, we optimize the network with the ReID loss \mathcal{L}_{ReID} and the proposed dual causality loss \mathcal{L}_{SNR} . For the unlabeled target data, we follow the adaptation strategy of MAR [41] to assign a pseudo soft multilabel for each sample and utilize these pseudo labels to perform soft multilabel-guided hard negative mining for training. We fine-tune the network also with the Adam optimizer [13] with a initial learning rate of 1×10^{-5} for 200 epochs. We optimize the Adam optimizer with a weight decay of 5×10^{-4} . The learning rate is decayed by a factor of 0.5 at 50, 100 and 150 epochs.

Why do we perform disentanglement only on channel level? We perform feature disentanglement only on channel level for **two reasons**: **1)** Those identity-irrelevant style factors (*e.g.*, illumination, contrast, saturation) are typically regarded as spatially consistent, which are hard to disentangle by spatial-attention. **2)** In our SNR, “disentanglement” aims at better “restitution” of the lost discriminative information due to Instance Normalization (IN). IN reduces style discrepancy of input features by performing normalization across spatial dimensions independently for each channel, where the normalization parameters are the same across different spatial positions. To be consistent with IN, we disentangle the features and reconstitute the identity-relevant ones to the normalized features on channel level.

Table 1: Details about the ReID datasets.

Datasets	Identities	Images	Cameras	Scene
Market1501 [44]	1501	32668	6	outdoor
DukeMTMC-reID [45]	1404	32948	8	outdoor
CUHK03 [16]	1467	28192	2	indoor
MSMT17 [34]	4101	126142	15	outdoor, indoor
VIPeR [7]	632	1264	2	outdoor
PRID2011 [11]	385	1134	2	outdoor
GRID [24]	250	500	2	indoor
i-LIDS [35]	119	476	N/A	indoor



Figure 1: Person images from different ReID datasets: Market-1501 [44], DukeMTMC-reID [45], CUHK03 [16], MSMT17 [34], and the four small-scale ReID datasets of PRID [11], GRID [24], VIPeR [7], and i-LIDS [35]. All images have been re-sized to 256×128 for easier comparison. We observe there are obvious domain gaps/style discrepancies across different datasets, especially for PRID [11] and GRID [24].

2. Details of Datasets

In Table 1, we present the detailed information about the related person ReID datasets. Market1501 [44], DukeMTMC-reID [45], CUHK03 [16], and large-scale MSMT17 [34] are the most commonly used datasets for fully supervised ReID [43, 49] and unsupervised domain adaption ReID [41, 42, 6]. VIPeR [7], PRID2011 [11], GRID [24], and i-LIDS [35] are small ReID datasets which could be used for evaluating cross-domain/generalizable person ReID [28, 12, 14]. Market1501 [44] and DukeMTMC-reID [45] have pre-established test probe and test gallery splits which we use for our training and cross-test (*i.e.*, $M \rightarrow D$, $D \rightarrow M$). For the smaller datasets (VIPeR, PRID2011, GRID, and i-LIDS), we use the standard 10 random splits as in [12, 14] for testing (the four small datasets are not involved in training). CUHK03 [16] and MSMT17 [34] are used for training.

We randomly pick up 10 identities from each ReID dataset and show them in Figure 1. We observe that: 1) there is style discrepancy across datasets, which is rather obvious for PRID and GRID; 2) MSMT17 has large style variants within the same dataset.

3. More Ablation Study Results

We show more comparisons of our scheme and others to demonstrate the effectiveness of our SNR module for generalizable person ReID in Table 2.

We have observations consistent with those in our paper. 1) IN-related baselines bring generalization ability improvement but decrease the performance for the same-domain. 2) Our *Baseline-SNR* achieves superior generalization capability thanks to the restitution of identity-relevant information by the SNR modules. 3) The generalization performance on unseen target domain increases consistently as the number of source datasets increases.

In Table 2, we also present the total number of source training images as marked by *data num.* N . For the single source dataset settings, MSMT17 is the largest dataset, which contains 126k images while Market1501 or Duke has about 33K images. For the target testing datasets VIPeR and iLIDS, the performance of *Baseline* trained by this large scale dataset MSMT17 is 3.8% to 12.5% higher than those trained by Market1501 or Duke in mAP. Generally, the increase of training data could improve the performance. However, the performance of *Baseline* trained by MSMT17 has a rather low mAP accuracy of 9.8% on the target dataset GRID, being even poorer than that trained on Market1501 (25.8%) or Duke (14.5%). For the target dataset PRID, similarly, MSMT17 does not provide clear superiority. These indicate that it is not always true that a larger amount of training data results in better performance. The domain gap between MSMT17 and GRID is larger than that between Market1501/Duke and GRID. To validate this, we analyze the feature divergence (FD, detailed descriptions can be found in Section 4 below) between GRID and MSMT17, Market1501, Duke, respectively. We find that the divergence (here we calculate the feature divergence of the third convolutional block/stage within our *Baseline-SNR* trained by combining all the four datasets) of Market1501 vs. GRID, Duke vs. GRID, MSMT17 vs. GRID are 2.17, 3.49, and 4.51, respectively. Note that the larger the FD value, the larger the feature discrepancy between the two domains. The domain gap between MSMT17 and GRID is larger than that between Market1501 (or Duke) and GRID. For the similar reason, we find that additionally adding MSMT17 as the source training data does not bring further performance improvement on GRID and PRID target datasets in our scheme *Baseline-SNR* in comparison with the model trained by $M+D+C$ source datasets.

4. More Visualization Analysis

More Feature Map Visualization. In our paper, we compare the activation maps \tilde{F}^+ of our scheme and those of the strong baseline scheme *Baseline* by varying the styles of input images (*e.g.*, contrast, illumination, saturation). Here,

Table 2: 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. Note that we show the total number of source training images by data num..

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) data num. 32.6k	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) data num. 32.9k	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
	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
Market1501 (M) + Duke (D) data num. 65.5k	Baseline	72.6	88.2	60.0	77.8	14.8	9.0	23.1	15.2	39.4	30.4	74.3	65.0
	Baseline-A-IN	76.5	91.4	62.2	80.1	45.0	30.0	36.7	28.0	37.3	28.2	73.6	65.2
	Baseline-IBN	74.6	90.4	62.3	80.1	43.7	32.0	32.6	24.0	42.8	33.2	73.8	65.0
	Baseline-A-SN	73.1	89.8	61.7	79.0	47.9	37.0	28.0	21.6	38.0	28.8	68.1	61.7
	Baseline-IN	77.5	91.6	63.9	81.5	48.1	36.0	39.2	31.2	43.8	33.9	73.2	64.3
	Baseline-SNR (Ours)	80.3	92.9	67.2	83.1	57.9	50.0	41.3	34.4	46.7	37.7	85.2	80.0
Market1501 (M) + Duke (D) + CUHK03 (C) data num. 93.7k	Baseline	76.4	89.8	63.6	79.0	27.0	19.0	25.7	18.4	46.3	36.4	77.1	66.3
	Baseline-A-IN	76.8	90.7	63.0	81.3	55.6	44.0	40.8	33.6	50.9	41.8	77.7	70.0
	Baseline-IBN	76.2	91.3	62.8	80.5	56.6	48.0	40.9	31.2	48.4	38.9	76.9	68.3
	Baseline-A-SN	71.1	89.3	62.0	78.8	55.4	46.0	34.1	26.4	50.3	39.8	79.6	71.7
	Baseline-IN	77.8	91.3	64.4	81.6	56.4	47.0	41.0	31.8	49.3	39.9	80.9	74.7
	Baseline-SNR (Ours)	81.2	93.3	68.4	84.2	60.9	52.0	45.2	36.8	52.3	42.4	91.0	86.7
MSMT17 (MT) data num. 126k	Baseline	23.1	48.2	29.2	47.6	16.4	11.0	9.8	5.6	40.8	30.1	74.0	66.7
	Baseline-SNR (Ours)	40.9	69.5	49.9	69.2	48.4	39.0	30.3	24.0	57.2	47.5	87.7	81.9
M + D + C + MT data num. 220k	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

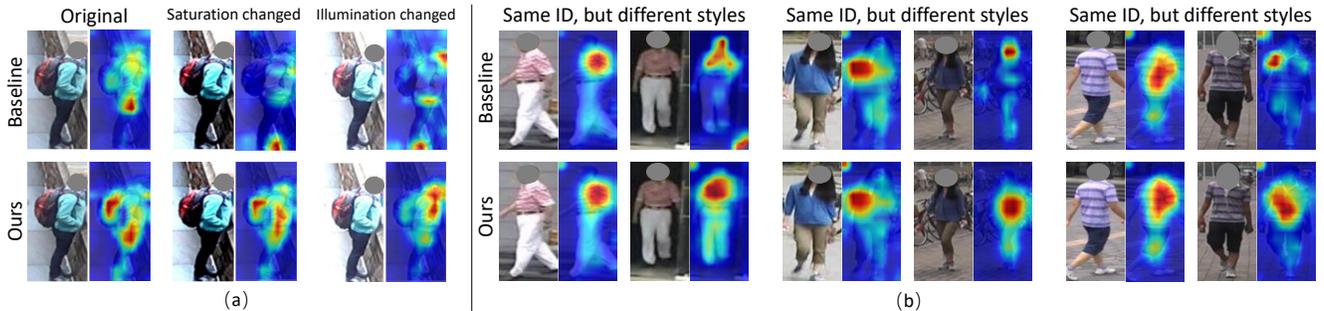


Figure 2: Activation maps of our scheme (bottom) and the strong baseline *Baseline* (top) corresponding to images of varied styles. The maps of our method are more consistent/invariant to style variants.

Figure 2(a) shows more visualization and Figure 2(b) shows visualization results on real images. We have the similar observations that the activation maps of our scheme are more consistent/invariant to style variants.

Feature Divergence Analysis. We analyze the feature divergence between two datasets on three schemes: *Baseline*, *Baseline-IN*, and ours *SNR*, respectively. Following [27, 17], we use the symmetric KL divergence of features between domain A and B as the metric to measure feature divergence of the two domains. We train the models using Market1501 training dataset and evaluate the feature

divergences between the test set of Market1501 and Duke (500 samples are randomly selected from each set). We calculate the feature divergence of the four convolutional blocks/stages respectively and show the results in Figure 3.

We observe that the feature divergence (FD) is large for *Baseline*. The introduction of IN as in scheme *Baseline-IN* significantly reduces the FD on all the four stages. The FD of Stage-4 is higher than that in Stage-3. That is likely because Stage-4 is more related to high-level discriminative semantic features for distinguishing different identities. The discrimination may increase the feature divergence. Due to

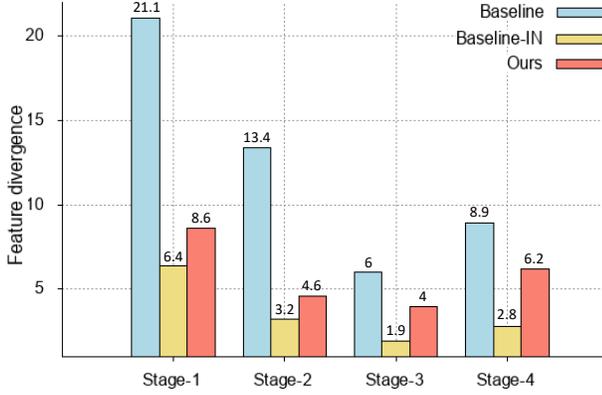


Figure 3: Analysis of the feature divergence between two different domains, Market1501 and Duke.

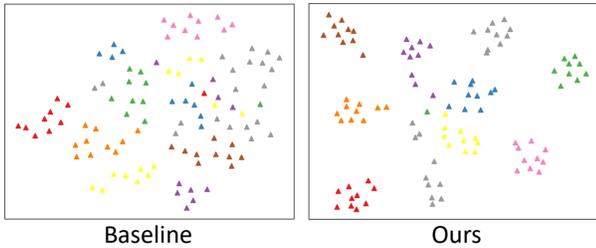


Figure 4: Visualization of the final ReID feature vector distribution for *Baseline* and *Ours* on the unseen target dataset Duke. Different identities are denoted by different colors.

the introduction of the SNR modules, the FD on all convolutional blocks/stages is also significantly reduced in our scheme in comparison with *Baseline*. It is higher than that of the scheme *Baseline-IN* which is probably because the restitution of some identity-relevant features increases the discrimination capability and thus increases the FD.

Visualization of ReID Feature Vector Distributions. In Figure 4, we further visualize the distribution of the final ReID feature vectors using t-SNE [26] for *Baseline* scheme and our final scheme on the unseen target dataset Duke (*i.e.*, Market1501→Duke). In comparison with *Baseline*, the feature distribution of the same identity (same color) becomes more compact while those of the different identities are pushed away in our scheme. It is easier to distinguish between different identities by our method.

5. Performance on Another Backbone

Our SNR is a plug-and-play module which can be added to available ReID networks. We integrate it into the recently proposed lightweight ReID network OSNet [49] and Table 3 shows the results. We can see that by simply inserting SNR modules between the OS-Blocks, the new scheme *OSNet-SNR* outperforms their best model *OSNet-IBN* by 5.0% and 5.5% in mAP for M→D and D→M, respectively. Note that,

for fair comparison, we use the official released weights and codes¹ of OSNet [49] to conduct these experiments.

Table 3: Evaluation of the generalization capability of proposed SNR modules on OSNet [49]. We use the official released weights and codes of OSNet for the experiments.

Method	M→D		D→M	
	mAP	Rank-1	mAP	Rank-1
Baseline (ResNet50)	19.8	35.3	21.8	48.3
OSNet [49]	19.3	35.2	21.7	49.9
OSNet-IBN [49]	26.7	48.5	26.1	57.7
OSNet-SNR	31.7	53.6	31.6	62.7

6. Comparison with State-of-the-Arts (Complete version)

To save space, we only present the latest approaches in the paper and here we show comparisons with more approaches in Table 4. Besides the description in *Introduction* and *Related Work* sections of our paper, we illustrate the difference between domain generalization and domain adaptation for person ReID in Table 6.

Moreover, in Table 5, we further compare our *SNR* with the latest generalizable ReID method Domain-Invariant Mapping Network (*DIMN*) [28] under the same experimental setting, *i.e.*, training on the same five datasets, Market1501 [44] + DukeMTMC-reID [45] + CUHK02 [15] + CUHK03 [16] + CUHK-SYSU [37]. We observe that *SNR* not only outperforms the *Baseline* by a large margin (up to 22.7% in mAP on PRID), but also significantly outperforms *DIMN*[28] by **14.6%/6.6%/1.2%/11.5%** in mAP on PRID/GRID/VIPeR/i-LIDS, respectively.

7. RGB-Infrared Cross-Modality Person ReID

To further demonstrate the generalization capability of the proposed SNR module, we conduct experiment on a more challenging RGB-Infrared cross-modality person ReID task, where there is a large style discrepancy between RGB images and Infrared images.

We evaluate our models on the standard benchmark dataset SYSU-MM01 [36]. Following [36], we conduct evaluation using the released official code based on the average of 10 repeated random split of gallery and probe sets. As shown in Table 7, in comparison with *Baseline*, our scheme which integrates the proposed SNR module on *Baseline* achieves a significant gain of **8.4%, 8.2%, 11.0%**, and **11.5%** in terms of mAP under 4 different experimental settings, and achieves the state-of-the-art performance.

¹<https://github.com/KaiyangZhou/deep-person-reid>

Table 4: Performance (%) comparisons with the state-of-the-art approaches for the Domain Generalizable Person ReID (top rows) and Unsupervised Domain Adaptation for Person ReID (bottom rows), respectively. “(U)” denotes “unlabeled”. We mask the schemes of our *Baseline* and our *Baseline* with SNR modules (*i.e.*, *SNR(Ours)*) by gray, with fair comparison between each pair to validate the effectiveness of SNR modules.

	Method	Venue	Source	Target: Duke		Source	Target: Market1501	
				mAP	Rank-1		mAP	Rank-1
Domain Generalization (w/o using target data)	OSNet-IBN [49]	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 [12]	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 [14]	ArXiv'19	MSMT17	43.3	64.5	MSMT17	36.6	64.8
	OSNet-IBN [49]	ICCV'19	MSMT17	45.6	67.4	MSMT17	37.2	66.5
Unsupervised Domain Adaptation (using unlabeled target data)	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
	PTGAN [34]	CVPR'18	Market1501 + Duke (U)	–	27.4	Duke + Market1501 (U)	–	38.6
	PUL [5]	TOMM'18	Market1501 + Duke (U)	16.4	30.0	Duke + Market1501 (U)	20.5	45.5
	MMFA [22]	BMVC'18	Market1501 + Duke (U)	24.7	45.3	Duke + Market1501 (U)	27.4	56.7
	SPGAN [4]	CVPR'18	Market1501 + Duke (U)	26.2	46.4	Duke + Market1501 (U)	26.7	57.7
	TJ-AIDL [31]	CVPR'18	Market1501 + Duke (U)	23.0	44.3	Duke + Market1501 (U)	26.5	58.2
	ATNet [23]	CVPR'19	Market1501 + Duke (U)	24.9	45.1	Duke + Market1501 (U)	25.6	55.7
	CamStyle [48]	TIP'19	Market1501 + Duke (U)	25.1	48.4	Duke + Market1501 (U)	27.4	58.8
	HHL [46]	ECCV'18	Market1501 + Duke (U)	27.2	46.9	Duke + Market1501 (U)	31.4	62.2
	ARN [18]	CVPRW'19	Market1501 + Duke (U)	33.4	60.2	Duke + Market1501 (U)	39.4	70.3
	ECN [47]	CVPR'19	Market1501 + Duke (U)	40.4	63.3	Duke + Market1501 (U)	43.0	75.1
	UDAP [29]	ArXiv'18	Market1501 + Duke (U)	49.0	68.4	Duke + Market1501 (U)	53.7	75.8
	PAST [42]	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 [41]	This work	Market1501 + Duke (U)	35.2	56.5	Duke + Market1501 (U)	37.2	62.4
	SNR(Ours)+MAR [41]	This work	Market1501 + Duke (U)	58.1	76.3	Duke + Market1501 (U)	61.7	82.8
	MAR [41]	CVPR'19	MSMT17 + Duke (U)	48.0	67.1	MSMT17 + Market1501 (U)	40.0	67.7
	PAUL [38]	CVPR'19	MSMT17 + Duke (U)	53.2	72.0	MSMT17 + Market1501 (U)	40.1	68.5
	Baseline+MAR [41]	This work	MSMT17 + Duke (U)	46.2	66.3	MSMT17 + Market1501 (U)	39.4	66.9
SNR(Ours) + MAR [41]	This work	MSMT17 + Duke (U)	61.6	78.2	MSMT17 + Market1501 (U)	65.9	85.5	

Table 5: Performance (%) comparison with the latest domain generalizable ReID method Domain-Invariant Mapping Network (DIMN) [28] under the same experimental setting (*i.e.*, training on the same five datasets, Market1501[44]+DukeMTMC-reID[45]+CUHK02[15]+CUHK03[16]+CUHK-SYSU[37]).

Source	Method	Target: PRID		Target: GRID		Target: VIPeR		Target: iLIDs	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Market + Duke + CUHK02 + CUHK03 + CUHK-SYSU	DIMN [28] CVPR'19	51.9	39.2	41.1	29.3	60.1	51.2	78.4	70.2
	Baseline	43.8	35.0	37.7	28.0	54.6	45.6	75.3	65.0
	SNR (Ours)	66.5	52.1	47.7	40.2	61.3	52.9	89.9	84.1

Table 6: Differences between settings of supervised, domain adaptive, and domain generalizable ReID.

Setting	Use target domain data?	Use target domain label?
Supervised	✓	✓
Domain adaptation	✓	✗
Domain generalization	✗	✗

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. 1
- [2] Pingyang Dai, Rongrong Ji, Haibin Wang, Qiong Wu, and Yuyu Huang. Cross-modality person re-identification with generative adversarial training. In *IJCAI*, pages 677–683, 2018. 6
- [3] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. 2005. 6
- [4] 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. 5
- [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. 5
- [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. 2, 5

Table 7: Performance (%) comparisons with the state-of-the-art RGB-IR ReID approaches on SYSU-MM01 dataset. R1, R10, R20 denote Rank-1, Rank-10 and Rank-20 accuracy, respectively.

Method	Venue	All Search								Indoor-Search							
		Single-Shot				Multi-shot				Single-Shot				Multi-Shot			
		mAP	R1	R10	R20	mAP	R1	R10	R20	mAP	R1	R10	R20	mAP	R1	R10	R20
HOG [3]	CVPR'05	4.24	2.76	18.3	32.0	2.16	3.82	22.8	37.7	7.25	3.22	24.7	44.6	3.51	4.75	29.1	49.4
MLBP [20]	ICCV'15	3.86	2.12	16.2	28.3	-	-	-	-	-	-	-	-	-	-	-	-
LOMO [19]	CVPR'15	4.53	3.64	23.2	37.3	2.28	4.70	28.3	43.1	10.2	5.75	34.4	54.9	5.64	7.36	40.4	60.4
GSM [21]	TPAMI'17	8.00	5.29	33.7	53.0	-	-	-	-	-	-	-	-	-	-	-	-
One-stream [36]	ICCV'17	13.7	12.1	49.7	66.8	8.59	16.3	58.2	75.1	56.0	17.0	63.6	82.1	15.1	22.7	71.8	87.9
Two-stream [36]	ICCV'17	12.9	11.7	48.0	65.5	8.03	16.4	58.4	74.5	21.5	15.6	61.2	81.1	14.0	22.5	72.3	88.7
Zero-Padding [36]	ICCV'17	16.0	14.8	52.2	71.4	10.9	19.2	61.4	78.5	27.0	20.6	68.4	85.8	18.7	24.5	75.9	91.4
TONE [39]	AAAI'18	14.4	12.5	50.7	68.6	-	-	-	-	-	-	-	-	-	-	-	-
HCML [39]	AAAI'18	16.2	14.3	53.2	69.2	-	-	-	-	-	-	-	-	-	-	-	-
BCTR [40]	IJCAI'18	19.2	16.2	54.9	71.5	-	-	-	-	-	-	-	-	-	-	-	-
BDTR [40]	IJCAI'18	19.7	17.1	55.5	72.0	-	-	-	-	-	-	-	-	-	-	-	-
D-HSME [8]	AAAI'19	23.2	20.7	62.8	78.0	-	-	-	-	-	-	-	-	-	-	-	-
cmGAN [2]	IJCAI'18	27.8	27.0	67.5	80.6	22.3	31.5	72.7	85.0	42.2	31.7	77.2	89.2	32.8	37.0	80.9	92.3
D ² RL [33]	CVPR'19	29.2	28.9	70.6	82.4	-	-	-	-	-	-	-	-	-	-	-	-
Baseline	This work	25.5	26.3	66.7	80.2	19.2	32.7	73.5	86.8	39.4	30.8	75.1	86.8	29.0	40.1	83.1	93.6
Ours	This work	33.9	34.6	75.9	86.6	27.4	41.7	83.3	92.3	50.4	40.9	83.8	91.8	40.5	50.0	91.4	96.1

- [7] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275. Springer, 2008. 2
- [8] Yi Hao, Nannan Wang, Jie Li, and Xinbo Gao. Hsme: Hypersphere manifold embedding for visible thermal person re-identification. In *AAAI*, volume 33, pages 8385–8392, 2019. 6
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. Deep residual learning for image recognition. In *CVPR*, 2016. 1
- [10] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 1
- [11] 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. 2
- [12] Jieru Jia, Qiuqi Ruan, and Timothy M Hospedales. Frustratingly easy person re-identification: Generalizing person re-id in practice. *BMVC*, 2019. 2, 5
- [13] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014. 1
- [14] 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. 1, 2, 5
- [15] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *CVPR*, pages 3594–3601, 2013. 4, 5
- [16] Wei Li, Rui Zhao, Lu Tian, et al. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 2, 4, 5
- [17] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*, 2016. 3
- [18] Yu-Jhe Li, Fu-En Yang, Yen-Cheng Liu, Yu-Ying Yeh, Xiao-fei 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. 5
- [19] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, 2015. 6
- [20] Shengcai Liao and Stan Z Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *ICCV*, pages 3685–3693, 2015. 6
- [21] Liang Lin, Guangrun Wang, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Cross-domain visual matching via generalized similarity measure and feature learning. *TPAMI*, 39(6):1089–1102, 2016. 6
- [22] Shan Lin, Haoliang Li, Chang-Tsun Li, and Alex Chichung Kot. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. *BMVC*, 2018. 5
- [23] Jiawei Liu, Zheng-Jun Zha, Di Chen, Richang Hong, and Meng Wang. Adaptive transfer network for cross-domain person re-identification. In *CVPR*, 2019. 5
- [24] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *IJCV*, 90(1):106–129, 2010. 2
- [25] 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. 1
- [26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 2008. 4
- [27] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, 2018. 3
- [28] 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, 4, 5
- [29] Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsuper-

- vised domain adaptive re-identification: Theory and practice. *arXiv preprint arXiv:1807.11334*, 2018. 5
- [30] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, 2016. 1
- [31] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *CVPR*, 2018. 5
- [32] Yan Wang, Lequn Wang, Yurong You, Xu Zou, Vincent Chen, Serena Li, Gao Huang, Bharath Hariharan, and Kilian Q Weinberger. Resource aware person re-identification across multiple resolutions. In *CVPR*, 2018. 1
- [33] Zhixiang Wang, Zheng Wang, Yinqiang Zheng, Yung-Yu Chuang, and Shin’ichi Satoh. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In *CVPR*, pages 618–626, 2019. 6
- [34] 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
- [35] Zheng Wei-Shi, Gong Shaogang, and Xiang Tao. Associating groups of people. In *BMVC*, pages 23–1, 2009. 2
- [36] 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. 4, 6
- [37] 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. 4, 5
- [38] Qize Yang, Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Patch-based discriminative feature learning for unsupervised person re-identification. In *CVPR*, 2019. 5
- [39] Mang Ye, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Hierarchical discriminative learning for visible thermal person re-identification. In *AAAI*, 2018. 6
- [40] Mang Ye, Zheng Wang, Xiangyuan Lan, and Pong C Yuen. Visible thermal person re-identification via dual-constrained top-ranking. In *IJCAI*, pages 1092–1099, 2018. 6
- [41] 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. 1, 2, 5
- [42] Xinyu Zhang, Jiawei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *ICCV*, 2019. 2, 5
- [43] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, et al. Densely semantically aligned person re-identification. In *CVPR*, 2019. 1, 2
- [44] Liang Zheng, Liyue Shen, et al. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 2, 4, 5
- [45] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *ICCV*, 2017. 2, 4, 5
- [46] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero-and homogeneously. In *ECCV*, 2018. 5
- [47] 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. 5
- [48] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camstyle: A novel data augmentation method for person re-identification. *TIP*, 2018. 5
- [49] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, et al. Omni-scale feature learning for person re-identification. *ICCV*, 2019. 1, 2, 4, 5