Global Distance-distributions Separation for Unsupervised Person Re-identification (Supplementary Material)

Xin Jin[1,2⋆] Cuiling Lan[2⋆⋆] Wenjun Zeng[2] Zhibo Chen[1⋆⋆]

[1] University of Science and Technology of China
[2] Microsoft Research Asia, Beijing, China
jinxustc@mail.ustc.edu.cn {culan,wezeng}@microsoft.com
chenzhibo@ustc.edu.cn

1 Mathematical Analysis of the Rationality of Momentum Update Mechanism

In order to mathematically prove the rationality of our momentum update design as described in Section 3.1 of the main manuscript, we design a toy game to present the momentum update process of the distance-distributions for different sample pair sets (*i.e.*, mean μ, variance σ^2) with analysis/derivation.

Assume two random sets $\mathcal{A}, \mathcal{B}$ with N and M sample pairs, respectively, and both sets exhibit Gaussian distribution. The mean and variance of set $\mathcal{A} = \{d_i | i = 1, \cdots, N\}$ with N sample pairs are represented as $\mu_A = \frac{1}{N}\sum_{i=1}^{N} d_i$, $\sigma_A^2 = \frac{1}{N}\sum_{i=1}^{N}(d_i - \mu_A)^2$ while those of set $\mathcal{B} = \{d_j^{'} | j = 1, \cdots, M\}$ with M sample pairs are estimated by $\mu_B = \frac{1}{M}\sum_{j=1}^{M} d_j^{'}$, $\sigma_B^2 = \frac{1}{M}\sum_{j=1}^{M}\left(d_j^{'} - \mu_B\right)^2$. We represent the set $\mathcal{C}$ as the combination of set $\mathcal{A}$ and set $\mathcal{B}$, the mean of the combined set $\mathcal{C}$ can be formulated as:

$$\mu_C = \frac{\sum_{i=1}^{N} d_i + \sum_{j=1}^{M} d_j^{'}}{N+M} = \frac{N}{N+M}\mu_A + \frac{M}{N+M}\mu_B = \beta\mu_A + (1-\beta)\mu_B, \tag{1}$$

where $\beta = \frac{N}{N+M}$. Similarly, the variance of the combined set $\mathcal{C}$ can be obtained:

$$\sigma_C^2 = \frac{\sum_{i=1}^{N}(d_i - \mu_C)^2 + \sum_{j=1}^{M}(d_j^{'} - \mu_C)^2}{N+M}, \tag{2}$$

when N is much larger than M (just like the situation in our training where the number of the previously "seen" mini-batches/samples is much larger than the number of samples in the current mini-batch), we could use μ_A to approximate

⋆ This work was done when Xin Jin was an intern at MSRA.
⋆⋆ Corresponding Author.

Table 1: Details about the ReID datasets.

Datasets	Abbreviation	Identities	Images	Cameras	Scene
Market1501 [25]	M	1501	32668	6	outdoor
DukeMTMC-reID [26]	D	1404	32948	8	outdoor
CUHK03 [9]	C	1467	28192	2	indoor
MSMT17 [18]	MSMT17	4101	126142	15	outdoor, indoor

μ_C, *i.e.*, $\mu_C \approx \mu_A$, thus we can have:

$$\begin{aligned}\sigma_C^2 &\approx \frac{\sum_{i=1}^{N}(d_i-\mu_A)^2+\sum_{j=1}^{M}(d_j^{'}-\mu_A)^2}{N+M} \\ &= \frac{N}{N+M}\sigma_A^2+\frac{M}{N+M}\frac{\sum_{j=1}^{M}(d_j^{'}-\mu_A)^2}{M} \\ &= \beta\sigma_A^2+(1-\beta)\frac{\sum_{j=1}^{M}(d_j^{'}-\mu_A)^2}{M}.\end{aligned} \tag{3}$$

By taking the sample pairs within a min-batch as the sample pairs of set $\mathcal{B}$, we can see that our momentum update design in Eq. (1) of our main manuscript is consistent with the above analysis/derivation.

2 Details of Datasets

In Table 1, we present the detailed information about the related person ReID datasets. Market1501 [25], DukeMTMC-reID [26], CUHK03 [9], and large-scale MSMT17 [18] are the most commonly used datasets for unsupervised domain adaptive person ReID [22, 23, 4] and fully supervised person ReID [24, 30]. Market1501, DukeMTMC-reID, CUHK03, and MSMT17 all have commonly used pre-established train and test splits, which we use for our training and cross dataset test (*e.g.*, M→D, D→M).

3 Implementation Details

Data Augmentation and Training. In the first stage of *model pre-training*, just as in [13], we use the commonly used data augmentation strategies of random cropping [17, 24], horizontal flipping, random erasing (REA) [12, 30], and the label smoothing regularization [15] to train the network for obtaining the capability of extracting discriminative features for person ReID on the labeled source dataset. The training is supervised by classification loss [14, 5] and triplet loss with batch hard mining [6]. In the second stage of *clustering*, we discard all the previous data augmentation operations and just simply extract features for the images of the target datasets for clustering. For the third stage of *adaptation*, consistent with the operations in the first stage, we leverage all these data augmentations to fine-tune the network.

Table 2: Performance (%) comparisons with the state-of-the-art approaches for unsupervised person ReID on the target dataset MSMT17 [18].

Unsupervised ReID	Venue	M→MSMT17		D→MSMT17	
		mAP	Rank-1	mAP	Rank-1
PTGAN [18]	CVPR'18	2.9	10.2	3.3	11.8
SSG [4]	ICCV'19	13.2	31.6	13.3	32.2
Baseline	This work	7.2	18.9	9.2	25.3
Baseline+**GDS-H**	This work	**14.9**	**34.3**	**14.2**	**33.9**

In the first and third stages, following [6], 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 $P = 32$ and $K = 4$ (*i.e.*, batch size $B = P \times K = 128$). We use Adam optimizer [8] for both stages.

For the first stage of *model pre-training*, we set the initial learning rate to 3×10^{-4} and regularize the network with a weight decay of 5×10^{-4}. The learning rate is decayed by a factor of 0.1 for every 50 epochs. We train the model on the source dataset for a total of 150 epochs. For the third stage of *adaptation*, we set the learning rate to 6×10^{-5} and keep it unchanged. The second stage and the third stage are executed alternatively for 30 iterations. For each iteration, we train our model for 70 epochs (that means, traverse all the target training samples for 70 times). For our proposed schemes, on top of *Baseline*, we add the proposed GDS constraint in the third stage.

All our models are implemented on PyTorch and trained on a single 16G NVIDIA-P100 GPU. We will release our code upon acceptance.

4 Influence of the Hyper-parameters λ_h and λ_σ

The hyper-parameter λ_h is used to balance the importance between the basic GDS loss $\mathcal{L}_{GDS}$ and the distribution-based hard mining loss $\mathcal{L}_H$. λ_σ aims to balance the mean and variance constraints within $\mathcal{L}_{GDS}$. For λ_h and λ_σ, we initially set them to 1, and then coarsely determine each one based on the corresponding loss values and their gradients observed during the training. The decision principle is to set their values to make the loss values/gradients lie in a similar range. Grid search within a small range of the derived λ_h/λ_σ is further employed to get better parameters. Actually, we observed the final performance is not very sensitive to the two hyper-parameters, we experimentally set $\lambda_h = 0.5, \lambda_\sigma = 1.0$ in the end.

5 Comparison with State-of-the-Arts (Complete Version)

More comparison results with state-of-the-art methods on the target dataset MSMT17 can be found in Table 2. We observe that in comparison with *Baseline*, our GDS constraint brings gains of **7.7%/15.4%** and **5.0%/8.6%** in

Table 3: Performance (%) comparisons with the state-of-the-art approaches for unsupervised person ReID. * means applying a re-ranking method of k-reciprocal encoding [27]. Note that *Baseline* is built following [13] with ResNet-50 backbone and thus has nearly the same performance as *Theory*[13]. To save space, we only present the latest approaches in the main manuscripts and here we show comparisons with more approaches.

		M→D		D→M		M→C		D→C		C→M		C→D	
Unsupervised ReID	Venue	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
CAMEL [21]	ICCV'17	–	–	26.3	54.5	–	–	–	–	–	–	–	–
PUL [3]	TOMM'18	–	–	20.5	45.5	–	–	–	–	–	–	–	–
PTGAN [18]	CVPR'18	–	27.4	–	38.6	–	–	–	–	–	31.5	–	17.6
SPGAN [2]	CVPR'18	22.3	41.1	22.8	51.5	–	–	–	–	19.0	42.8	–	–
TJ-AIDL [16]	CVPR'18	23.0	44.3	26.5	58.2	–	–	–	–	–	–	–	–
ARN [10]	CVPRW'18	33.4	60.2	39.4	70.3	–	–	–	–	–	–	–	–
MMFA [11]	BMVC'18	24.7	45.3	27.4	56.7	–	–	–	–	–	–	–	–
HHL [28]	ECCV'18	27.2	46.9	31.4	62.2	–	–	–	–	29.8	56.8	23.4	42.7
CFSM [1]	AAAI'19	27.3	49.8	28.3	61.2	–	–	–	–	–	–	–	–
MAR [22]	CVPR'19	48.0	67.1	40.0	67.7	–	–	–	–	–	–	–	–
ECN [29]	CVPR'19	40.4	63.3	43.0	75.1	–	–	–	–	–	–	–	–
PAUL [20]	CVPR'19	53.2	72.0	40.1	68.5	–	–	–	–	–	–	–	–
SSG [4]	ICCV'19	53.4	73.0	58.3	80.0	–	–	–	–	–	–	–	–
PCB-R-PAST* [23]	ICCV'19	54.3	72.4	54.6	78.4	–	–	–	–	57.3	79.5	**51.8**	**69.9**
Theory [13]	PR'2020	48.4	67.0	52.0	74.1	46.4	47.0	28.8	28.5	51.2	71.4	32.2	49.4
ACT [19]	AAAI'20	54.5	72.4	60.6	80.5	48.9	49.5	30.0	30.6	64.1	81.2	35.4	52.8
Baseline	This work	48.4	67.1	52.1	74.3	46.2	47.0	28.8	28.4	51.2	71.4	32.0	49.4
Baseline + GDS	This work	52.9	71.4	57.1	78.5	48.0	48.9	30.7	32.5	63.6	81.6	44.1	64.0
Baseline + GDS-H	This work	**55.1**	**73.1**	**61.2**	**81.1**	**49.7**	**50.2**	**34.6**	**36.0**	**66.1**	**84.2**	45.3	64.9
B-SNR[7]	CVPR'20	54.3	72.4	66.1	82.2	47.6	47.5	31.5	33.5	62.4	80.6	45.7	66.7
B-SNR[7]+GDS	This work	57.2	74.6	68.6	84.9	49.8	50.5	36.7	38.8	67.2	85.1	49.4	69.9
B-SNR[7]+GDS-H	This work	**59.7**	**76.7**	**72.5**	**89.3**	**50.7**	**51.4**	**38.9**	**41.0**	**68.3**	**86.7**	**51.0**	**71.5**

mAP/Rank-1 for M→MSMT17 and D→MSMT17, respectively, which demonstrates the effectiveness of our proposed GDS constraint. SSG [4] also belongs to clustering-based approach. It exploits the potential similarity from the global body to local parts to build multiple clusters at different granularities. As a comparison, our *Baseline* and *Baseline+GDS-H* only consider the similarity at global body. Being simple in design, our final scheme *Baseline+GDS-H* outperforms the second best method SSG [4] by **2.7%** and **1.7%** in Rank-1 accuracy for M→MSMT17 and D→MSMT17, respectively.

In addition, to save space, we only present the latest approaches in the Section 4.6 "Comparison with State-of-the-Arts" in the main manuscripts and here we show comparisons with more approaches in Table 3.

6 More Visualization Results

Visualization of Dataset-wise (Global) Distance Distributions. To better understand how well our GDS constraint works, in Fig. 1, we not only visualize the dataset-wise Pos-distr and Neg-distr on the test set of target dataset (as shown in Fig. 6 in the main manuscripts), but also visualize the counterpart on the training set of target dataset. We have the following observations. 1) Thanks to the adaptation on the unlabeled target dataset and our GDS constraint, the distance distributions of our final scheme *Baseline+GDS-H* present a much better separability than that of other schemes. This trend can be observed on both the training set and test set. 2) On the training set, each scheme

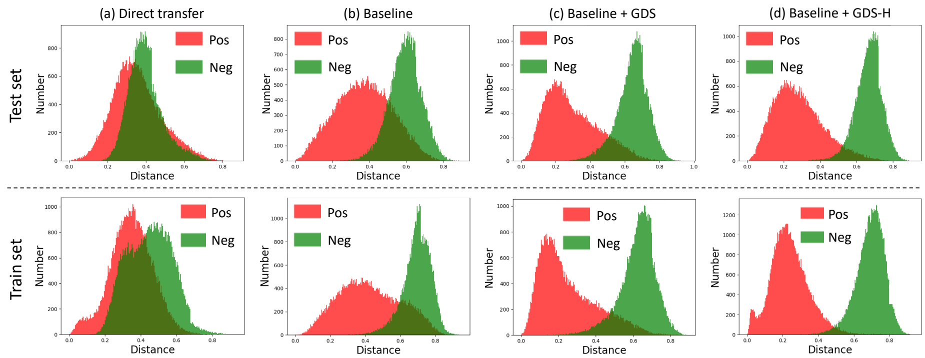

Fig. 1: Histograms of the distances of the positive sample pairs (red) and negative sample pairs (green) on the **test set (top)** and **train set (bottom)** of the target dataset Duke (Market1501→Duke) for schemes of (a) *Direct transfer*, (b) *Baseline*, (c) *Baseline+GDS*, and (d) *Baseline+GDS-H*.

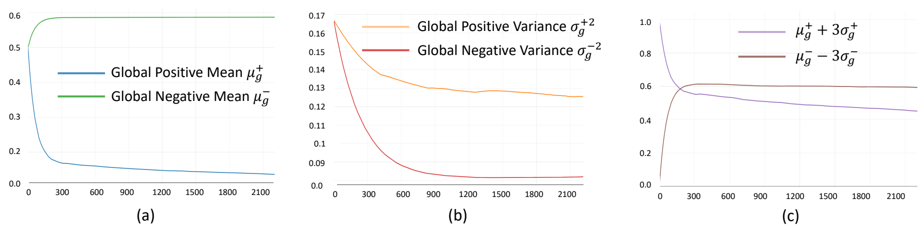

Fig. 2: Trend analysis of the learned dataset-wise (global) statistics in the training.

presents better separability than that on the test set, especially for our final scheme*Baseline+GDS-H*, which suggests that our GDS constraint is actually very helpful in promoting the separation after the optimization.

Trend Analysis of the Learned Dataset-wise (Global) Statistics. We observe the changing trend of the global statistics of distance distributions (including the mean μ_g^+ of global Pos-distr, the mean μ_g^- of global Neg-distr, the variance σ_g^{+2} of global Pos-distr, and the variance σ_g^{-2} of global Neg-distr) in the training process and show the curves in Fig. 2[3]. The horizontal axis denotes the identities of the epochs (30 iterations $\times$ 70 epochs = 2100 epochs). We observe that 1) as we expected, the centers/means of two distributions (μ_g^+, μ_g^-) and their hard tails $(\mu_g^+ + 3\sigma_g^+, \mu_g^- - 3\sigma_g^-)$ become further apart as the training goes; 2) the two distributions variance $(\sigma_g^{+2}, \sigma_g^{-2})$ become sharper since the variances become smaller as the training progresses.

[3] We initialize the two distributions with mean of 0.5 and variance of 1/6 for the observation. Actually, we found the performance is not sensitive to the initialization values of the statistics.

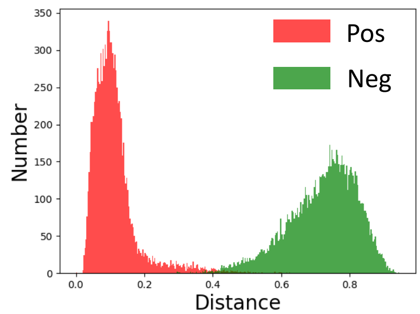

Fig. 3: Histograms of the distances of the positive sample pairs (red) and negative sample pairs (green) on the *training* set of the labeled dataset CUHK03 for fully supervised person ReID.

Table 4: Effectiveness of the proposed GDS loss and the distribution-based hard-mining loss (H) for the fully supervised Person ReID.

Supervised ReID	CUHK03 (L)		MSMT17	
	mAP	Rank-1	mAP	Rank-1
Baseline	69.8	73.7	47.2	73.8
Baseline+**GDS**	70.7	74.3	48.3	74.4
Baseline+**GDS-H**	**71.4**	**75.5**	**49.1**	**74.9**

7 GDS Constraint Applied to Supervised Person ReID

We design the GDS constraint for addressing the inseparability of distance distributions in unsupervised person ReID, where there is no groudtruth labels for the target dataset. The use of either the pseudo labels or style transferred images results in noises and overlapping of the two distributions. For fully supervised person ReID, the proposed GDS is also expected to enhance the performance. However, on the benchmark datasets, due to the use of reliable labels and the over-fitting problem, we found the distance distributions on the training set are already well separated (see Fig. 3) and thus there left small optimization space for us. Quantitatively, as shown in Table 5, although our GDS brings some performance improvement (1.6% and 1.9% in mAP for CUHK03(L) and MSMT17, respectively), it is not significant in comparison with the unsupervised ReID setting.

8 Training Complexity Analysis

The increase of training time of our design in comparison with *Baseline* [13] is negligible. We build *Baseline* with the representative clustering-based method [13], and add the proposed GDS constraint in the training. Both our loss calculation and momentum update have very low computation complexity in compar-

Table 5: Performance w.r.t different clustering algorithms for the M→D setting.

M→D	K-means		DBSCAN		HDBSCAN	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Baseline	40.2	57.9	48.4	67.1	49.6	67.9
Baseline+**GDS-H**	49.0	67.2	55.1	73.1	55.7	73.6

ison with the convolutional operations of the network. Take the setting of using DukeMTMC-reID as source dataset and Market1501 as target dataset as an example, the training time of *Baseline* [13] and our scheme *Baseline+GDS-H* is 17.9 hours and 18.2 hours, respectively (*i.e.*, about 1.7% increase). The training time is comparable to that of the existing STOA methods (PAUL [20] with 16.3 hours, SSG [4] with 20.8 hours, MAR [22] with 25.6 hours). Note that all these training courses are conducted on a single 16G NVIDIA-P100 GPU.

9 Performance w.r.t Different Clustering Algorithms

The performance of the *Baseline* scheme with the cluttering approach of DBSCAN is similar to that with hierarchical DBSCAN (HDBSCAN), 48.4% vs. 49.6% in mAP for M→D, and both outperforms the *Baseline* scheme with K-means (40.2%). Our GDS constraint consistently brings improvement of 8.8%, 6.7%, and 6.1% for that with K-means, DBSCAN, and HBSCAN, respectively. For simplicity, we use DBSCAN by default in our experiments.

References

1. Chang, X., Yang, Y., Xiang, T., Hospedales, T.M.: Disjoint label space transfer learning with common factorised space. In: AAAI. vol. 33, pp. 3288–3295 (2019)
2. Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person reidentification. In: CVPR (2018)
3. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: Clustering and fine-tuning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) (2018)
4. Fu, Y., Wei, Y., Wang, G., Zhou, X., Shi, H., Huang, T.S.: Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person reidentification. ICCV (2019)
5. Fu, Y., Wei, Y., Zhou, Y., et al.: Horizontal pyramid matching for person reidentification. In: AAAI (2019)
6. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person reidentification. arXiv preprint arXiv:1703.07737 (2017)
7. Jin, X., Lan, C., Zeng, W., Chen, Z., Zhang, L.: Style normalization and restitution for generalizable person re-identification. In: CVPR (2020)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2014)

9. Li, W., Zhao, R., Tian, L., et al.: Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR (2014)
10. Li, Y.J., Yang, F.E., Liu, Y.C., Yeh, Y.Y., Du, X., Frank Wang, Y.C.: Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In: CVPR workshops (2018)
11. Lin, S., Li, H., Li, C.T., Kot, A.C.: Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. BMVC (2018)
12. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: CVPR workshops (2019)
13. Song, L., Wang, C., Zhang, L., Du, B., Zhang, Q., Huang, C., Wang, X.: Unsupervised domain adaptive re-identification: Theory and practice. Pattern Recognition p. 107173 (2020)
14. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: ECCV. pp. 480–496 (2018)
15. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR (2016)
16. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: CVPR (2018)
17. Wang, Y., Wang, L., You, Y., Zou, X., Chen, V., Li, S., Huang, G., Hariharan, B., Weinberger, K.Q.: Resource aware person re-identification across multiple resolutions. In: CVPR (2018)
18. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: CVPR (2018)
19. Yang, F., Li, K., Zhong, Z., Luo, Z., Sun, X., Cheng, H., Guo, X., Huang, F., Ji, R., Li, S.: Asymmetric co-teaching for unsupervised cross domain person re-identification. AAAI (2020)
20. Yang, Q., Yu, H.X., Wu, A., Zheng, W.S.: Patch-based discriminative feature learning for unsupervised person re-identification. In: CVPR (2019)
21. Yu, H.X., Wu, A., Zheng, W.S.: Cross-view asymmetric metric learning for unsupervised person re-identification. In: ICCV (2017)
22. Yu, H.X., Zheng, W.S., Wu, A., Guo, X., Gong, S., Lai, J.H.: Unsupervised person re-identification by soft multilabel learning. In: CVPR (2019)
23. Zhang, X., Cao, J., Shen, C., You, M.: Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: ICCV (2019)
24. Zhang, Z., Lan, C., Zeng, W., et al.: Densely semantically aligned person re-identification. In: CVPR (2019)
25. Zheng, L., Shen, L., et al.: Scalable person re-identification: A benchmark. In: ICCV (2015)
26. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: ICCV (2017)
27. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR (2017)
28. Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero- and homogeneously. In: ECCV (2018)
29. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: Exemplar memory for domain adaptive person re-identification. In: CVPR. pp. 598–607 (2019)
30. Zhou, K., Yang, Y., Cavallaro, A., et al.: Omni-scale feature learning for person re-identification. ICCV (2019)