Joint Disentangling and Adaptation for Cross-Domain Person Re-Identification
Supplementary Material

Yang Zou[1]* Xiaodong Yang[2] Zhiding Yu[2]
B.V.K. Vijaya Kumar[1] Jan Kautz[2]

[1]Carnegie Mellon University [2]NVIDIA

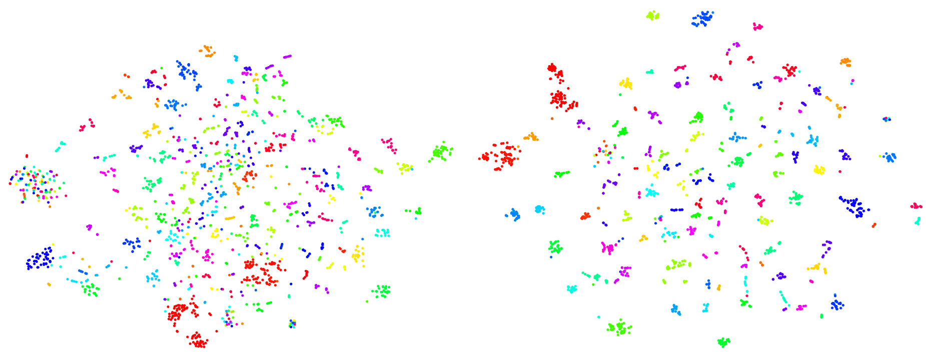
A Additional Implementation Details

DG-Net++ consists of an appearance encoder E_{app}, two source and target structure encoders E^s_{str}, E^t_{str}, two source and target decoders G^s, G^t, an image discriminator D_{img}, and a domain discriminator D_{dom}. As described in the main paper, E_{app} is modified from ResNet50, and E^s_{str}, E^t_{str} and G^s, G^t follow the within-domain architecture designs as DG-Net [4]. D_{dom} is a multi-layer perceptron containing four fully-connected layer, where the input dimension is 2048, output dimension is 1, and the dimensions of hidden layers are 1024, 512 and 256. Note after each fully connected layer, we apply a batch normalization layer [1] and a LReLU [3] (negative slope set to 0.2). In all experiments, the input images are resized to 256×128. SGD is used to train E_{app} with learning rate 0.0006 and momentum 0.9, and Adam is applied to optimize E^s_{str}, E^t_{str}, G^s, G^t, D_{img} with learning rate 0.000001 and $(\beta_1, \beta_2) = (0, 0.999)$, and D_{dom} with learning rate 0.00001. To warm up E_{app}, E^s_{str}, G^s and D_{img}, we follow the configuration as [4]. We use an iterative self-training approach to generate pseudo-labels every two epochs. We utilize labeled source and pseudo-labeled target data in self-training with softmax loss. DBSCAN is used for clustering with k-reciprocal encoding to compute pairwise distances. Every experiment is conducted on a single NVIDIA TITAN V100 GPU. Our full model takes 15.8 GPU memory and runs for 460K iterations. Our source code with all implementation details is available at https://github.com/NVlabs/DG-Net-PP.

B Feature Distribution Visualization

DG-Net++ is a joint learning framework that disentangles id-related/unrelated factors such that adaptation can be more effectively conducted on id-related space to prevent id-unrelated interference. Figure 7 illustrates the feature distributions of the images in target domain visualized by t-SNE [2]. It can be apparently observed that by using DG-Net++ the features of different identities are more separable and the features of the same identity are more clustered.

* Work done during an internship at NVIDIA Research.



(a) (b)

Fig. 7: Visualization by t-SNE of the feature distributions of the images in target domain. Features are extracted by (a) the baseline method and (b) DG-Net++ on the cross-domain pair Market → Duke.

Market-1501 → DukeMTMC-reID									
ε	0.35	0.40	0.45	0.50	`MinPts`	5	6	7	8
mAP	62.4	61,7	63.8	62.2	mAP	61.4	61.6	63.8	62.8

Table 3: Sensitivity analysis of the hyper-parameters ε (the maximum distance between two samples to be treated as neighbors) and `MinPts` (the minimal number of neighboring samples of a point to be selected as a core point) of DBSCAN.

To further quantitatively evaluate the target domain feature distributions, we compute the purity scores for the features produced by the baseline method and DG-Net++ on the cross-domain pair Market → Duke. To compute the purity score, each cluster is assigned to the identity that is most frequent in the cluster, then the purity score is measured by the number of correctly assigned images divided by the total number of images. The purity score is 51.9% for baseline and 76.3% for DG-Net++, clearly indicating that the intra-class similarity and inter-class difference are more encouraged in DG-Net++.

C Additional Ablation Study

Sensitivity analysis of DBSCAN. We adopt DBSCAN to produce the pseudo-labels of images in target domain. Experiments show that our model is not sensitive to the hyper-parameters of DBSCAN. Specifically, we conduct sensitivity analysis for (1) ε which is the maximum distance between two samples to be considered as neighbors, and (2) `MinPts` which is the minimal number of neighbouring samples for a point to be considered as a core point. Table 3 shows the experimental results for sensitivity analysis of ε (fixing `MinPts` to 7) and `MinPts` (fixing ε to 0.45) on the benchmark pair Market → Duke. It can be found that DG-Net++ is overall not sensitive to ε and `MinPts`.

Method	Metric	Market→ Duke	Duke→ Market	MSMT→ Market	Market→ MSMT	MSMT→ Duke	Duke→ MSMT
DG-Net [4]	Rank@1	42.6	56.1	61.8	17.1	61.9	20.6
	mAP	24.3	26.8	33.6	5.4	40.7	6.4
DG-Net++	Rank@1	**78.9** (+36.3)	**82.1** (+26.0)	**83.1** (+21.3)	**48.4** (+31.3)	**75.2** (+13.3)	**48.8** (+28.2)
	mAP	**63.8** (+39.5)	**61.7** (+34.9)	**64.6** (+31.0)	**22.1** (+16.7)	**58.2** (+17.5)	**22.1** (+15.7)

Table 4: Comparison between DG-Net and DG-Net++ for unsupervised cross-domain person re-id on the six benchmark pairs.

DG-Net++ vs. DG-Net. To illustrate the cross-domain performance difference between DG-Net [4] and DG-Net++, we show their comparisons over the six cross-domain pairs in Table 4. DG-Net++ is found to substantially and consistently outperform DG-Net over all benchmarks. This is evident to validate the efficacy of the proposed learning framework in coupling cross-domain disentanglement and adaptation, backing the necessity of such combination for unsupervised cross-domain re-id.

References

1. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: ICML (2015)
2. van der Maaten, L., Hinton, G.: Visualizing high-dimensional data using t-sne. JMLR (2008)
3. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. In: ICML Workshop (2015)
4. Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., Kautz, J.: Joint discriminative and generative learning for person re-identification. In: CVPR (2019)