Adaptive Cross-Domain Learning for Generalizable Person Re-Identification

Pengyi Zhang[1], Huanzhang Dou[1], Yunlong Yu[2⋆], and Xi Li[1,3,4⋆]

[1] College of Computer Science & Technology, Zhejiang University
[2] College of Information Science & Electronic Engineering, Zhejiang University
[3] Shanghai Institute for Advanced Study, Zhejiang University
[4] Shanghai AI Laboratory
{pyzhang,hzdou,yuyunlong,xilizju}@zju.edu.cn

This supplementary file consists of:

- Details for datasets
- Details for evaluation settings
- Details of network architecture
- Qualitative analysis for cross-domain meta-weights
- Feature visualization

1 Details for Datasets

There are totally 9 datasets involved during training and testing, the statistics of them are shown in Table 1. It is worth mentioning that, we do not use the DukeMTMC dataset since its privacy issues.

Table 1. The datasets involved in the experiments for training and testing.

Datasets	Train			Valid and Test		
	#IDs	#Imgs	#Cams	#IDs	#Imgs	#Cams
Market1501 (M)	751	12,936	6	750	19,281	6
MSMT17 (MS)	1,041	30,248	15	3,060	96,193	15
CUHK02 (C2)	1,816	7,264	10	-	-	-
CUHK03 (C3)	767	7,368	2	700	6,728	2
CUHK-SYSU (CS)	11,934	34,574	1	-	-	-
PRID	100	100	1	649	749	1
GRID	-	-	-	250	1,275	8
VIPeR	316	732	2	316	732	2
iLIDs	59	241	8	60	120	8

⋆ Co-corresponding authors.

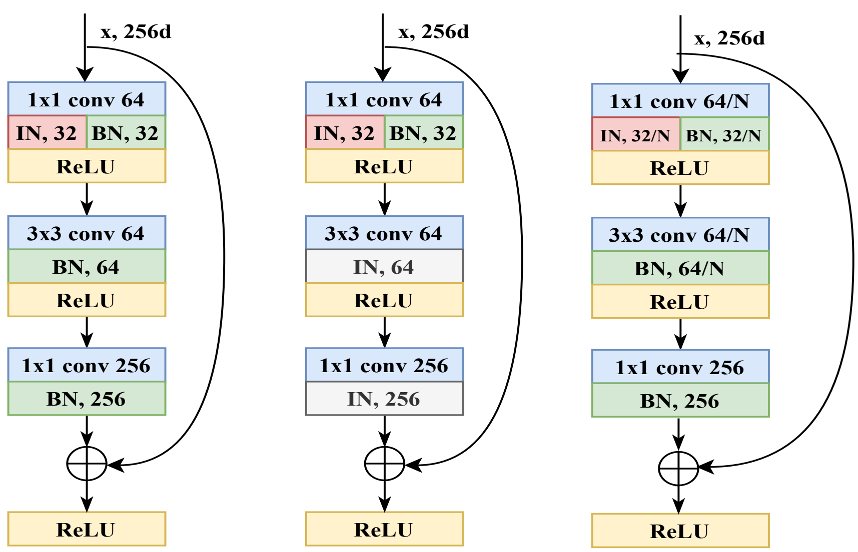
Table 2. Three evaluation protocols in the experiments. 'Com' denotes that both the training and testing data of source domains are used for training.

Setting	Training Data	Testing Data
Protocol-1	Com-(M+C2+C3+CS)	PRID,GRID, VIPeR,iLIDs
Protocol-2	Leave-one-out for M+MS+C3+CS	
Protocol-3	Leave-one-out for Com-(M+MS+C3+CS)	

2 Details for Evaluation Settings

For more clarity, we list the three evaluation protocols proposed in our main paper in Table 2. Under Protocol-1, models are trained on four datasets and tested on another four unseen datasets, which could demonstrate the good generalization performance under different distributions. Under Protocol-2 and Protocol-3, experiments are conducted on four large scale datasets in a leave-one-out mode, which could further verify the effectiveness of our methods.

3 Details of Network Architecture



(a) Standard Block (b) Static Branch (c) Embedding Network

Fig. 1. Network architecture of three blocks involved in our work.

As we have stated in the main paper, our Adaptive Cross-Domain Learning (ACL) framework is implemented by plugging the Cross-Domain Embedding Block (CODE-Block) into each stage of the backbone network.

In our work, we mainly use the ResNet50 with IBN as our backbone network. As shown in Fig. 1, we visualize the conventional bottleneck block, block of our static branch, and block of embedded network for a more clear presentation. As shown in Fig. 1(b), our static branch has the same architecture as a conventional bottleneck block in ResNet-IBN, but we replace the batch normalization (BN) layers in the bottleneck block with the instance normalization (IN) layers. As shown in Fig. 1(c), our embedding network has the same topology of bottleneck block in ResNet-IBN, but we reduce the number of channels in the first two convolutional layers for a light computation cost. Specifically, we reduce the number of channels into $1/N$, where N is the number of embedded networks.

4 Qualitative Analysis for Cross-Domain Meta-Weights

As shown in Fig. 2, we explore the relationship among the samples with similar meta-weights from different domains. In general, the samples with similar meta-weights usually have some common style patterns, which can be regarded as a subdomain. These results further demonstrate the effectiveness of our cross-domain feature embedding and the domain-aware combination strategy.

(a) Illumination (b) Carrying Objects

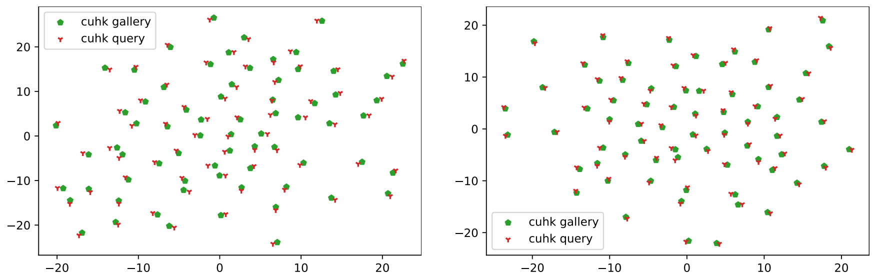
(c) Resolution (d) Clothing Style

Fig. 2. Samples with similar meta-weights from different domains under Protocol-2. The samples with similar meta-weights usually share some common domain style patterns, such as high illumination, similar carrying objects, low resolution, similar clothing style, etc.

5 Feature Visualization

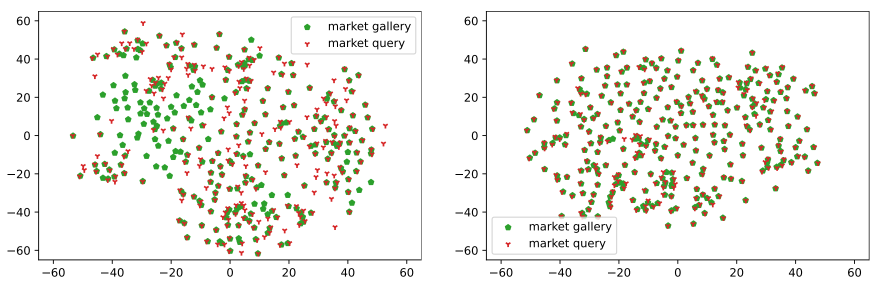
To better understand the effectiveness of our approach, we visualize the final features extracted from different unseen domains under Protocol-2 in Fig. 3 and

Fig. 4, respectively. Specifically, we visualize the feature distributions for each identity in the query set and gallery set by t-SNE, where each point indicates a single identity in the query set or the gallery set. As shown in Fig. 3(b) and Fig. 4(b), under our ACL framework, we could obtain a more compact feature representation for the unseen target domains, which is more discriminative and have better generalization capability.



(a) Baseline (b) Ours

Fig. 3. The t-SNE visualization for the features extracted from target unseen domain (i.e., CUHK03 dataset) under Protocol-2.



(a) Baseline (b) Ours

Fig. 4. The t-SNE visualization for the features extracted from target unseen domain (i.e., Market1501 dataset) under Protocol-2.