Unsupervised Meta-Domain Adaptation for Fashion Retrieval Supplementary Material

Vivek Sharma[1,3,4], Naila Murray[2], Diane Larlus[2],
M. Saquib Sarfraz[1], Rainer Stiefelhagen[1], Gabriela Csurka[2]
[1] Karlsruhe Institute of Technology, [2] NAVER LABS Europe
[3] Massachusetts Institute of Technology, [4] Harvard Medical School

This supplementary material is composed of additional details about our method including mapping between DF subcategories and S2S categories (Section 1); a statistical and qualitative analysis of the clustering performed on shop images (Section 2, which complements Section 4.2 from the main paper); and additional experimental results with per-category analysis of the `UMDA` approach for both seen and unseen categories (Section 3, which extends Section 5.3 from the main paper). In addition in Section 4 we show a high resolution version of Figure 3 from the paper which illustrates the different pseudo labeling strategies we compared.

1. Category alignment between DF and S2S

As discussed in Section 5.1 of the main paper, there is no one-to-one correspondence between the fashion categories/sub-categories of the two datasets. To overcome this additional challenge, we establish a mapping between DF subcategories and S2S categories —(as shown in Table 1). We refer to the 6 categories from S2S that we were able to match with DF as *seen categories*. The other five S2S categories, referred as *unseen*, do not match any category in DF and they were not used at train time. Note that not all sub-categories in DF were mapped to the meta-categories.

2. Shop image clustering

As discussed in Section 5.2 of the main paper, here we provide more details on the statistics of the joint DF and S2S shop image clustering. The FINCH algorithm produces multiple partitions. The first one corresponds to linking samples through the first neighbor relations, while the second one links clusters created in the first step. We use clusters from the first partition to mine positive and negative pairs as this partition increases diversity, without compromising on quality of the labels. Statistics of these clusters are shown in Table 2. Qualitative results are shown in Figure 1.

3. Additional experimental results

Table 3 compares, for each fashion category, the baseline `DF-BL`, `UMDA-MLP`, and `UMDA-E2E` for both seen and unseen categories. For this evaluation, we report per category mean average precision (mAP). It is interesting to note that both of our methods improve over the baseline both for seen and unseen categories.

4. Different labeling strategy

Figure 2 is a higher resolution version of Figure 3 from the main paper. It enumerates and illustrates all the different pseudo-labeling strategies defined and discussed in Section 4.2 from the main paper.

References

[1] Vivek Sharma, M Saquib Sarfraz, and Rainer Stiefelhagen. A simple and effective technique for face clustering in tv series. In *CVPR: Brave New Motion Representations Workshop*, 2017.

[2] Vivek Sharma, Makarand Tapaswi, M Saquib Sarfraz, and Rainer Stiefelhagen. Self-supervised learning of face representations for video face clustering. In *IEEE Automatic Face and Gesture Recognition*, 2019.

[3] Vivek Sharma, Makarand Tapaswi, M Saquib Sarfraz, and Rainer Stiefelhagen. Clustering based contrastive learning for improving face representations. In *IEEE Automatic Face and Gesture Recognition*, 2020.

Table 1. Category-level mapping between S2S and DF. We build 6 *seen* meta-categories from DF and map them with 6 existing S2S categories. The remaining 5 *unseen* categories in S2S have no match in DF and were not used for training.

	Street2Shop (S2S)	DeepFashion (DF)
seen	Dresses	Clothing: {Dress, Lace_Dress, Sleeveless_Dress}
	Leggings	Trousers: {Leggings}
	Outerwear	Clothing: {Coat}, Tops: {Coat}
	Pants	Clothing: {Pants}, Trousers:{Pants}
	Skirts	Dresses: {Skirt}, Dresses: {Suspenders_Skirt}
	Tops	Tops: {Blouse, Chiffon, Lace_Shirt, Summer_Wear, Tank_Top, T_Shirt}
unseen	{Belts, Bags, Eyewear, Footwear, Hats }	

Table 2. Statistical analysis of our clusters (summary after merging the 6 individual clustering per category). We show i) the number of samples available (and the corresponding number of different product IDs) and ii) the actual number of samples (and number of pIDs) used for clustering. The last row shows weighted clustering purity (WCP) [1, 2, 3] for each dataset computed individually on the joint clustering.

	All-#Samples	All-#PIDs	#Samples-Used	#PIDs-Used	WCP (%)
DF	29240	21676	2996	1544	40.92
S2S	22327	5422	4925	2243	69.32

Table 3. Per-category mAP for cross-domain retrieval on S2S using ResNet-101.

mAP	Dresses	Leggings	Outerwear	Pants	Skirts	Tops	Belts	Bags	Eyewear	Footwear	Hats
	Seen (6 Categories)						Unseen (5 Categories)				
`DF-BL`	27.61	10.55	15.03	21.06	38.75	24.82	01.78	09.77	07.77	02.25	17.53
`UMDA-MLP`	29.56	10.92	15.53	23.32	41.85	25.91	02.24	10.52	08.42	02.36	18.71
`UMDA-E2E`	31.45	11.69	16.45	24.29	42.19	26.68	02.28	10.35	09.32	02.58	19.14

Figure 1. Images from several clusters (DF and S2S shop images) and a corresponding DF consumer-DF/S2S shop image pairing (red arrow) example obtained using one of the dominant PID (images with dotted blue frame). Best viewed in color.

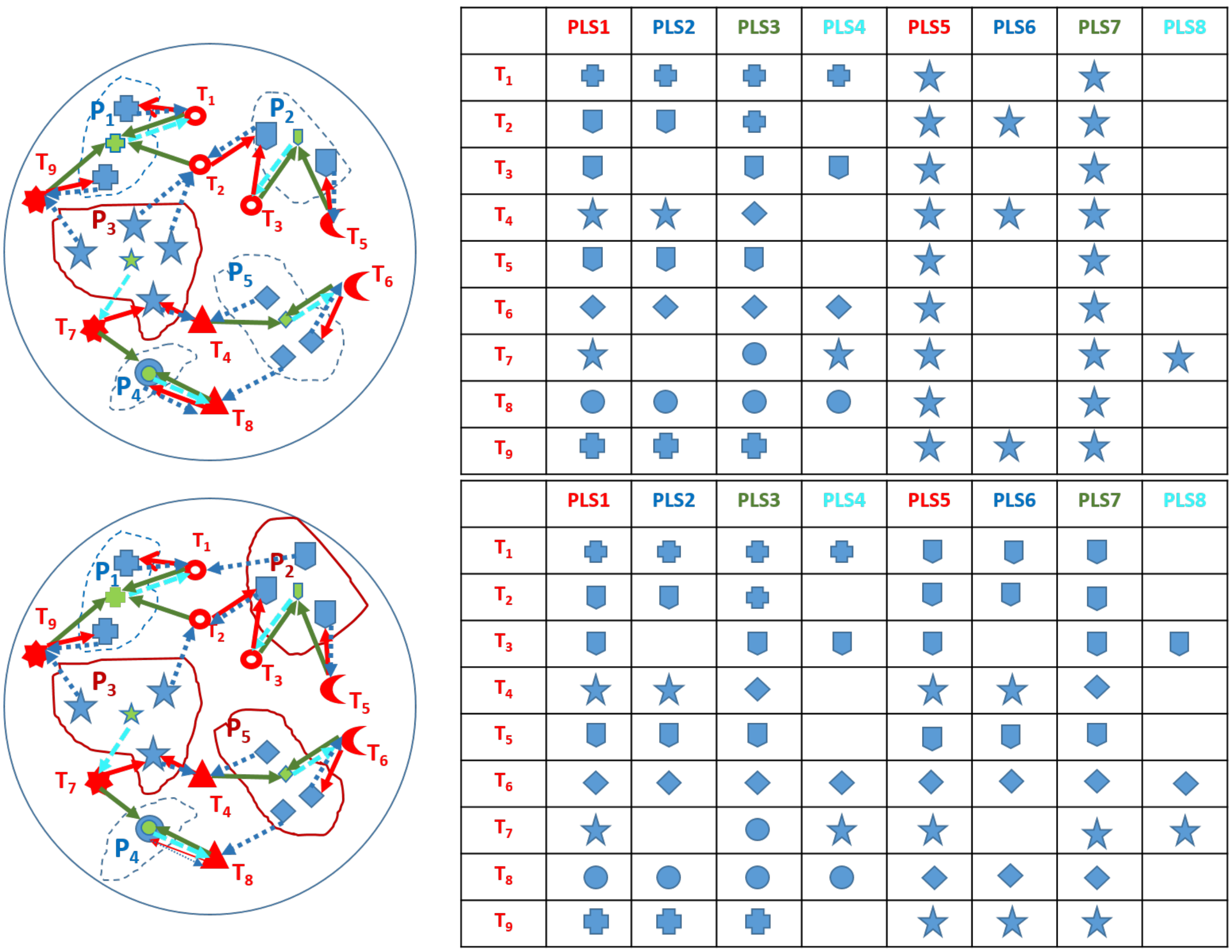

Figure 2. Illustration of the different pseudo-labeling strategies. Top: single dominant PID. Bottom: several dominant PIDs. This figure is a higher resolution version of Figure 3 from the main paper.