Supplementary Material for Conditional Cross Attention Network for Multi-Space Embedding without Entanglement in Only a SINGLE Network

Figure A1: Examples of Our TrainSets. The order of each row is FashionAI, DARN, DeepFashion and Zappos50K.

A. Datasets

**FashionAI** [6] The data published in the FashionAI Global Challenge 2018 has 180,335 apparel images. This comprises 8 fashion attributes containing 55 classes each.

**DARN** [1] An open dataset for attribute classification and street-to-shop image retrieval, comprising 253,983 images and 9 attributes. Each attribute contains 185 classes. The data is provided as image URLs; excluding broken URLs that cannot be downloaded, we used 195,771 URLs.


**Zappos50K** [5] This dataset comprises 50,025 shoe images collected from Zappos.com. It consists of 4 attributes containing 34 classes each.

---

**Algorithm 1** Pseudo-Code for CCA Training

1: **input:** Image $I$, Condition $c$
2: batch $B$, training epochs $K$, triplet set $T$
3: Self Attention Block $SA$, Conditional Cross Attention $CCA$
4: for epoch $= 1, ..., K$ do
5:   for $B = 1, ..., M \in T$ do
6:     Triplet $(A_c, P_c, N_c) \leftarrow B$
7:     $I, c \leftarrow A_c, P_c, N_c$
8:     $Q_i, K_i, V_i \leftarrow \text{Token Embedding}(I)$
9:     for $l = 1, ..., (L - 1)$ do
10:        $Q_i, K_i, V_i \leftarrow SA(Q_i, K_i, V_i)$
11:     end for
12:     Last iteration $l = L$ do
13:        $Q_c \leftarrow \text{Conditional Token Embedding}(c)$
14:        $\leftarrow CCA(Q_c, K_i, V_i)$
15:        $f \leftarrow f(FC())$
16:        calculate $f_a, f_p, f_n \leftarrow \text{Triplet}(A_c, P_c, N_c)$
17:        calculate triplet loss $L(f_a, f_p, f_n | c)$
18:        calculate gradients of $\nabla L(\theta)$
19:        $\theta \leftarrow \text{Adam}(\nabla L(\theta))$
20:     end for
21:   end for

B. More Visualization

**B.1. Ranking and Attention Heat map**

Figure A2 shows the Top 3 results along with each actual attention map. Each part of each attribute is considered, interpreted as the result of disentanglement multi-space modeling. The order in the figure is lapel design (notched), neckline design (round), skirt length (floor), pant length (midi), sleeve length (short), neck design (low turtle), coat length (midi), and collar design (peter pan).

**B.2. Ours vs. Previous Works : Multi-Space Embedding**

Figure 6 comparatively analyzed the embedding results of our study and previous studies. Of the 8 categories, the
Figure A2: Examples of our top 3 ranking pair (image, attention heat map) results for FashionAI of 8 attributes. Red rectangle is query images. The order of each line is lapel design (notched), neckline design (round), skirt length (floor), pant length (midi), sleeve length (short), neck design (low turtle), coat length (midi) and collar design (peter pan).

results for Neck Design, Sleeve Length, and Coat Length were presented. Figure A3 shows the expanded results for all 8 categories. Our method solves the entanglement problem much better than ASEN [3] and CAMNet [4].

References

Figure A3: Ours vs. Previous Works (ASEN, CAMNet): Multi-space embedding’s visualization using t-SNE about FashionAI.


