

Learning Attribute and Class-Specific Representation Duet for Fine-grained Fashion Analysis

Yang Jiao
Amazon

jaoyan@amazon.com

Yan Gao, Jingjing Meng, Jin Shang, Yi Sun
Amazon

{yanngao, ajmeng, imjshang, yisun}@amazon.com

Supplementary Materials

1. Hyperparameter Sensitivity Analysis

(1) We have conducted a sensitivity study on the threshold γ (Table 1). As can be seen, our model is not sensitive to γ and consistently outperforms the baselines. Although we present results for $\gamma=0.8$ in the paper, in fact $\gamma=0.925$ gives the best MAP@all. It also shows that our results are not handpicked. (2) For w_p , i.e., the weight on the positive labels, we use the average ratio of negative and positive labels among samples in the dataset.

models	γ	MAP@all	MAP@100	Recall@100
<i>ASEN</i> [19]	-	12.33	20.60	5.10
<i>ASEN_{v2}</i> [19]	-	12.51	20.55	4.99
<i>ASEN++</i> [6]	-	12.74	20.79	5.21
<i>MODC</i> [14]	-	12.78	22.21	6.06
<i>M3-Net</i>	0.500	20.78	32.62	12.03
<i>M3-Net</i>	0.600	21.11	33.26	12.25
<i>M3-Net</i>	0.700	21.42	34.01	12.50
<i>M3-Net</i>	0.800	21.78	35.08	12.88
<i>M3-Net</i>	0.900	22.14	37.07	13.42
<i>M3-Net</i>	0.925	22.17	38.01	13.60
<i>M3-Net</i>	0.950	22.13	39.45	13.82
<i>M3-Net</i>	0.975	21.58	42.01	13.83

Table 1. Sensitivity study of γ for M3-Net. Study is on DeepFashion with comparison with baselines.

2. More Experimental results

Table 2, Table 3, Table 4 presents MAP@100 on each attribute in the three benchmark datasets: DeepFashion, FashionAI, and DARN. According to the tables, M3-Net consistently outperforms other methods in the top 100 retrieval performance on most attributes. The best performing method on each dataset are underlined in **bold**.

3. More Results of Fine-grained Fashion Retrieval with M3-Net

In Figure 1, we present more results of multi-label fine-grained fashion retrieval on DeepFashion.

Model	DeepFashion, MAP@100					
	Attribute	texture	fabric	shape	part	style
<i>ASEN</i> [3]	29.70	21.16	25.06	15.38	10.20	20.60
<i>ASEN_{v2}</i> [3]	30.24	20.00	24.62	15.29	10.13	20.55
<i>ASEN++</i> [1]	30.25	20.21	25.22	15.41	10.34	20.79
<i>MODC</i> [2]	32.42	21.96	26.55	15.81	11.50	22.21
<i>M3-Net_a</i>	35.71	27.94	32.63	23.82	18.22	28.14
<i>M3-Net_c</i>	39.15	29.61	34.67	26.07	21.41	30.58
<i>M3-Net</i>	43.47	33.27	41.71	31.15	23.87	35.08

Table 2. Overall performance comparison on all attributes in the multi-label dataset, DeepFashion.

References

- [1] Jianfeng Dong, Zhe Ma, Xiaofeng Mao, Xun Yang, Yuan He, Richang Hong, and Shouling Ji. Fine-grained fashion similarity prediction by attribute-specific embedding learning. *arXiv preprint arXiv:2104.02429*, 2021.
- [2] Yang Jiao, Ning Xie, Yan Gao, Chien-Chih Wang, and Yi Sun. Fine-grained fashion representation learning by online deep clustering. In *European Conference on Computer Vision*, pages 19–35. Springer, 2022.
- [3] Zhe Ma, Jianfeng Dong, Zhongzi Long, Yao Zhang, Yuan He, Hui Xue, and Shouling Ji. Fine-grained fashion similarity learning by attribute-specific embedding network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11741–11748, 2020.

Model	FashionAI, MAP@100								
	skirt length	sleeve length	coat length	pant length	collar design	lapel design	neckline design	neck design	overall
<i>ASEN</i> [3]	72.20	56.74	58.87	70.48	75.43	68.57	61.82	63.73	64.70
<i>ASEN_{v2}</i> [3]	72.57	61.33	60.14	71.67	75.70	71.82	69.42	66.48	67.85
<i>ASEN++</i> [1]	73.65	63.90	63.90	74	76.81	71.29	73.82	72.53	70.62
<i>MODC</i> [2]	81.53	73.88	76.23	82.10	84.52	80.86	84.18	82.39	80.29
<i>M3-Net_a</i>	80.89	78.54	74.25	84.00	85.73	84.70	85.84	80.52	81.57
<i>M3-Net_c</i>	84.19	77.05	78.35	84.94	87.35	86.46	84.80	81.55	82.58
<i>M3-Net</i>	87.57	84.72	82.77	88.45	89.48	90.49	89.13	85.62	87.04

Table 3. Overall performance comparison on all attributes in the single-label dataset, FashionAI.

Model	DARN, MAP@100									
	clothes category	clothes button	clothes color	clothes length	clothes pattern	clothes shape	collar shape	sleeve length	sleeve shape	overall
<i>ASEN</i> [3]	42.57	53.58	58.47	62.62	61.91	65.04	42.43	80.71	62.25	58.72
<i>ASEN_{v2}</i> [3]	43.16	55.87	56.97	65.23	62.33	64.78	44.03	83.08	62.86	59.66
<i>ASEN++</i> [1]	45.12	56.40	57.71	65.22	65.08	65.81	45.47	85.45	64.93	61.09
<i>MODC</i> [2]	59.20	72.08	67.96	74.56	74.43	4.59	61.66	91.14	75.07	72.16
<i>M3-Net_a</i>	67.96	78.52	76.35	79.09	79.94	83.16	71.47	90.94	75.64	78.12
<i>M3-Net_c</i>	70.46	81.00	76.57	84.24	83.22	87.03	76.87	94.32	79.38	81.46
<i>M3-Net</i>	70.15	83.01	79.84	86.10	83.84	87.83	78.67	93.81	80.85	82.69

Table 4. Overall performance comparison on all attributes in the single-label dataset, DARN.

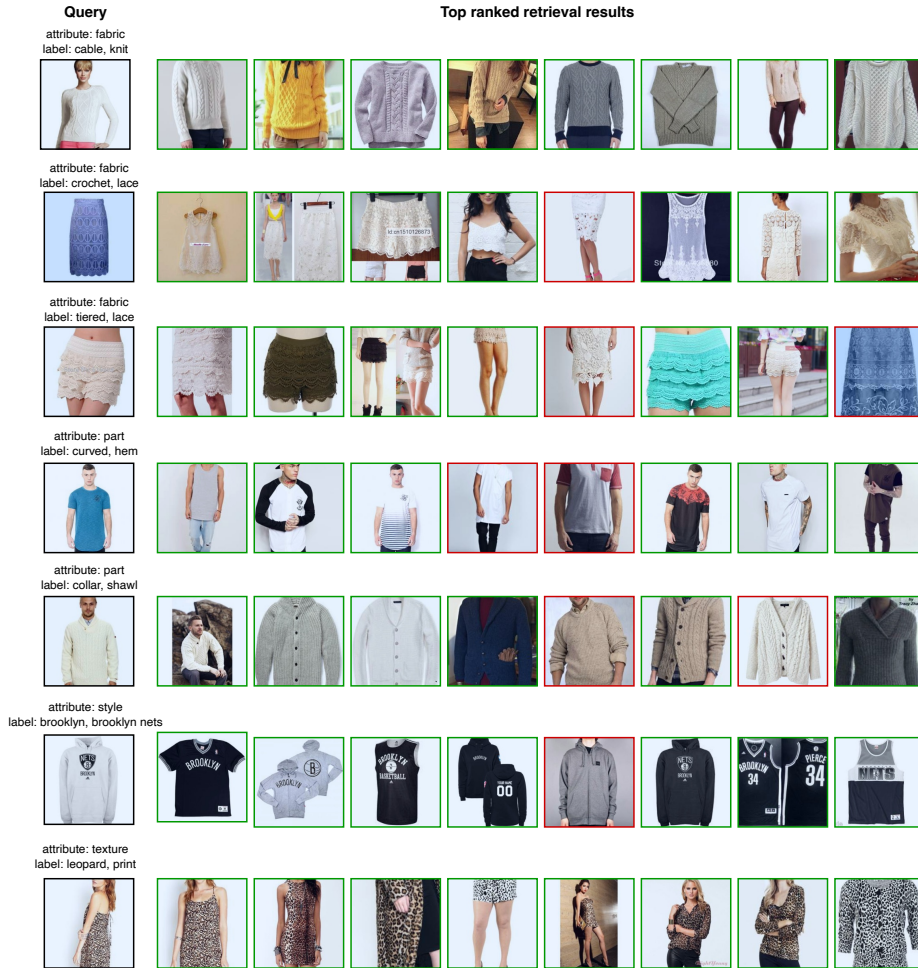


Figure 1. Examples of multi-label fashion retrieval. The retrieval is conducted by M3-Net on DeepFashion test split. Green shows true positives retrieved, while red shows false positives retrieved. Best viewed in color on a computer screen.