Fine-Grained Fashion Similarity Learning by Attribute-Specific Embedding Network

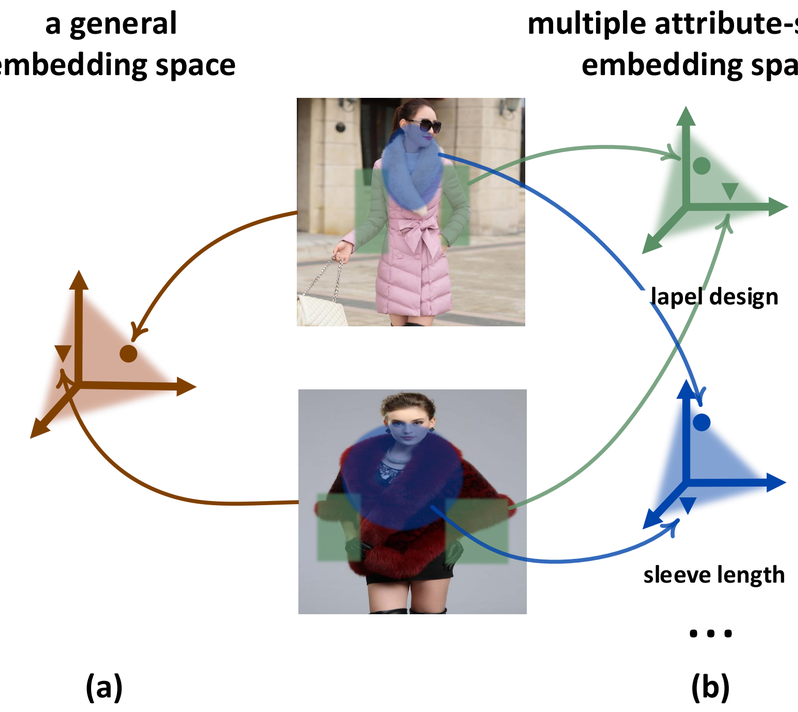
Zhe Ma[1], Jianfeng Dong[2,3*], Yao Zhang[1], Zhongzi Long[1], Yuan He[4], Hui Xue[4] and Shouling Ji[1,3*]

[1]Zhejiang University [2]Zhejiang Gongshang University

[3]Alibaba-Zhejiang University Joint Institute of Frontier Technologies [4]Alibaba Group

Introduction

Attribute-Specific Fashion Retrieval: Search for fashion items in terms of certain fine-grained similar designs instead of identical or overall similar items.

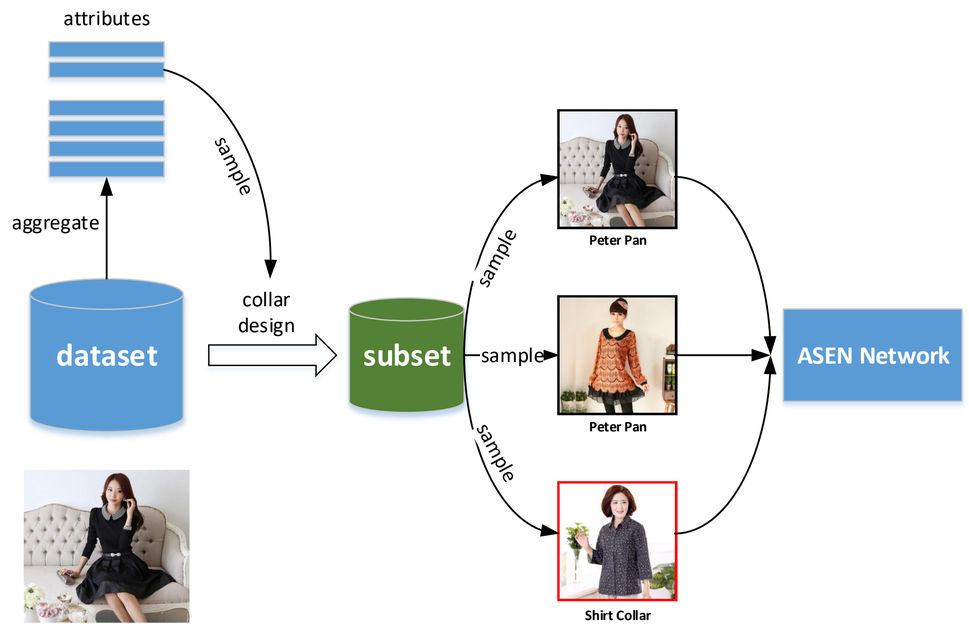


Left: Existing methods tend to focus on overall similarity.
Right: Two items appearing to be irrelevant overall actually present similar characteristics over some attributes.

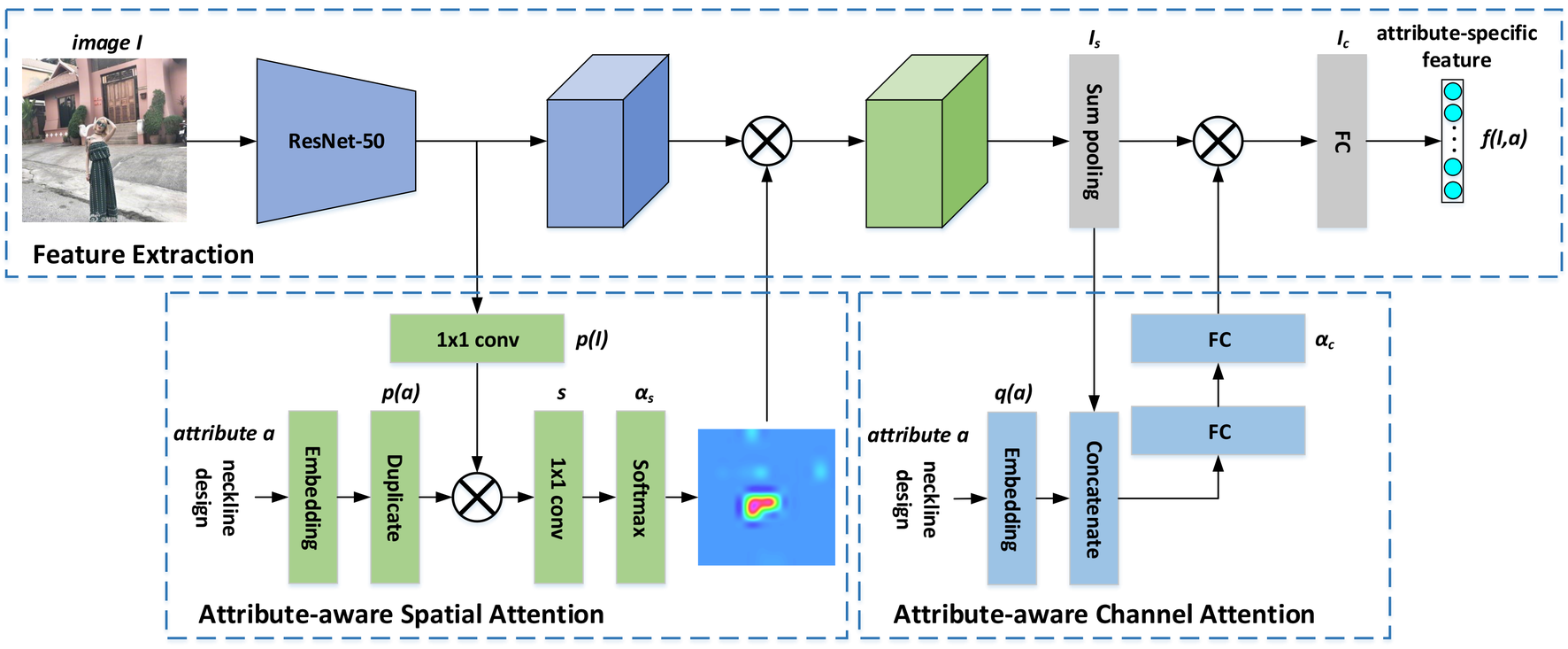
Training Data Generation

We utilize large-scale fashion recognition datasets to conduct metric learning target.

- Aggregate appropriate fashion attributes and construct different subsets according to certain attributes.
- Random sample triplets to train our proposed Attribute-Specific Embedding Network(ASEN).



Our proposal: Attribute-Specific Embedding Network



Attribute-aware Spatial Attention:
Fashion attributes are typically related to certain regions. We first use an attribute guided spatial attention to attend to relevant parts of clothes.

Attribute-aware Channel Attention:
The same regions may still corresponds to multiple attributes. We further emply an attribute-guided channel attention to select discriminative dimensions.

Experiments

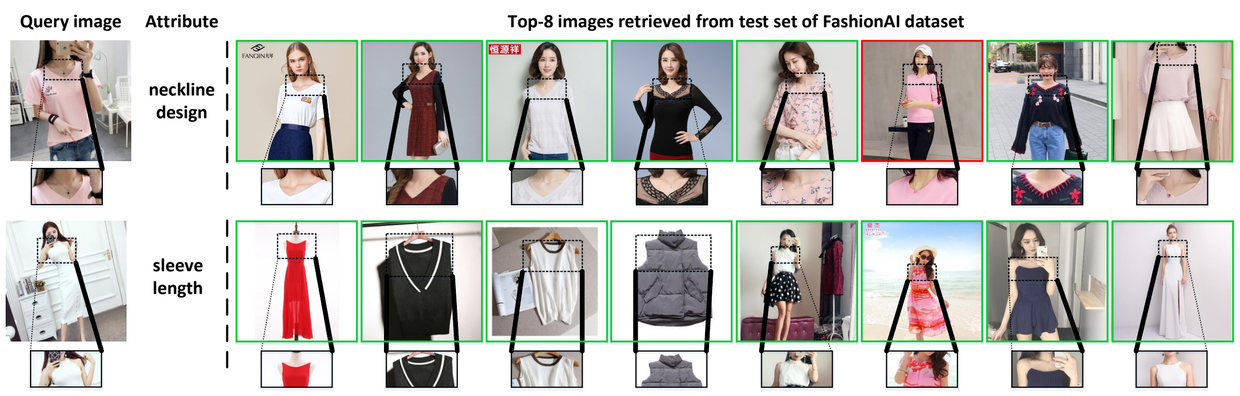
- Attribute-specific fashion retrieval on FashionAI Dataset

Method	MAP for each attribute								MAP
	skirt length	sleeve length	coat length	pant length	collar design	lapel design	neckline design	neck design	
Random baseline	17.20	12.50	13.35	17.45	22.36	21.63	11.09	21.19	15.79
Triplet network	48.38	28.14	29.82	54.56	62.58	38.31	26.64	40.02	38.52
CSN	61.97	45.06	47.30	62.85	69.83	54.14	46.56	54.47	53.52
ASEN w/o ASA	62.65	49.98	49.02	63.48	69.10	61.65	50.88	57.10	56.35
ASEN w/o ACA	58.12	43.30	42.30	60.03	65.98	49.95	46.86	52.06	50.87
ASEN	**64.44**	**54.63**	**51.27**	**63.53**	**70.79**	**65.36**	**59.50**	**58.67**	**61.02**

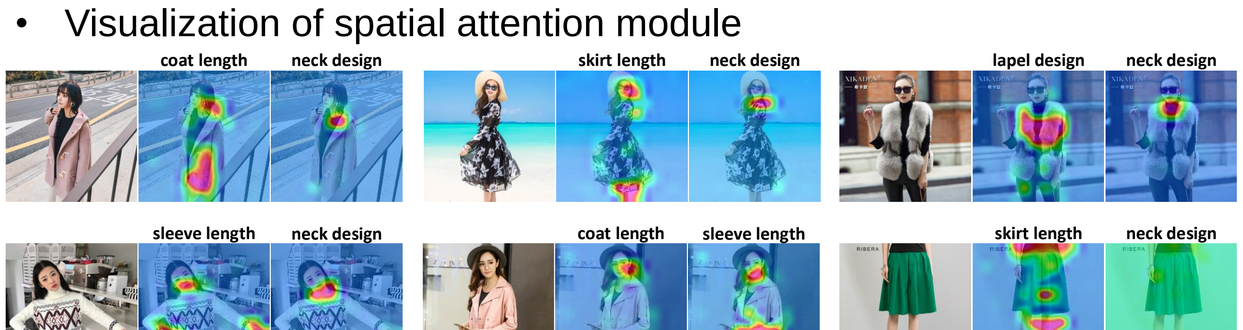
- Attribute-specific fashion retrieval on DARN Dataset

Method	MAP for each attribute									MAP
	clothes category	clothes button	clothes color	clothes length	clothes pattern	clothes shape	collar shape	sleeve length	sleeve shape	
Random baseline	8.49	24.45	12.54	29.90	43.26	39.76	15.22	63.03	55.54	32.26
Triplet network	23.59	38.07	16.83	39.77	49.56	47.00	23.43	68.49	56.48	40.14
CSN	34.10	44.32	47.38	53.68	54.09	56.32	31.82	78.05	58.76	50.86
ASEN w/o ASA	33.94	45.37	48.56	54.36	53.83	57.33	32.78	77.77	59.32	51.39
ASEN w/o ACA	30.39	42.37	49.14	50.18	53.63	48.84	26.03	75.28	57.99	48.02
ASEN	**36.69**	**46.96**	**51.35**	**56.47**	**54.49**	**60.02**	**34.18**	**80.11**	**60.04**	**53.31**

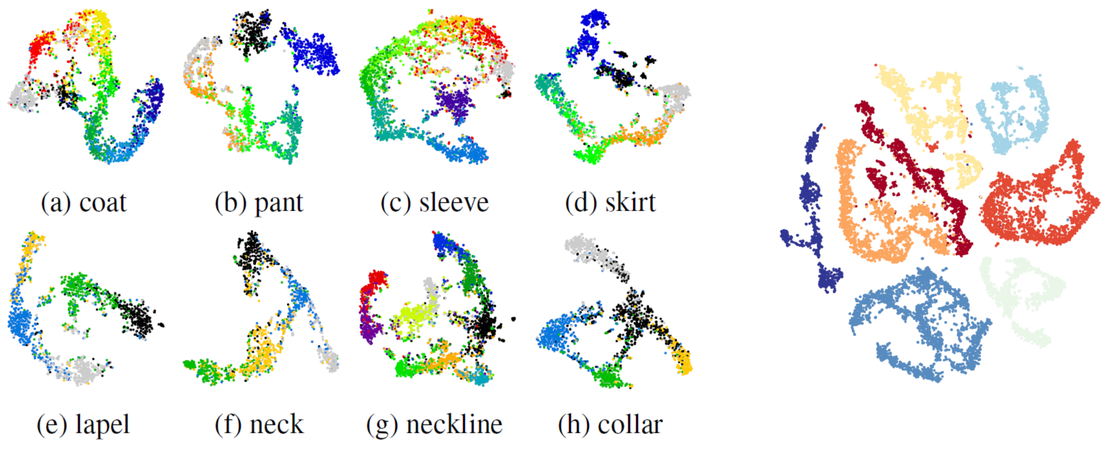
- Attribute-specific fashion retrieval



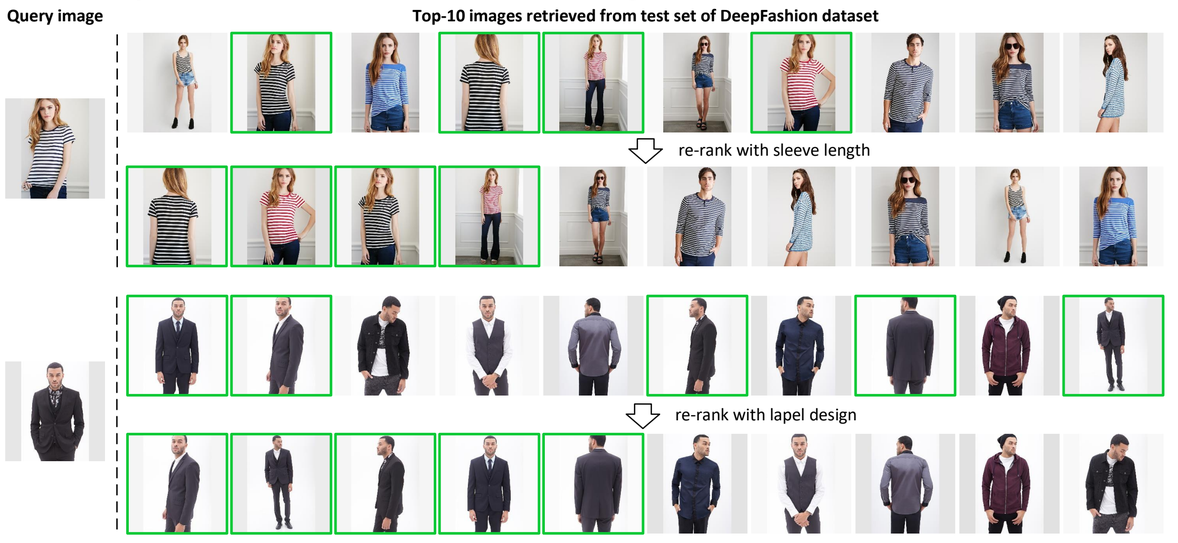
- Visualization of spatial attention module



- The learned attribute-specific embedding space



- The potential for fashion reranking



Take-home Message

- Our ASEN Network for learning multiple fine-grained similarities.
- Attention modules are beneficial for fashion variance.
- For fine-grained similairy consideration, learning multiple attribute-specific embedding spaces is better than a single general embedding space.

https://github.com/maryeon/asen maryeon@zju.edu.cn dongjf24@gmail.com