amazon

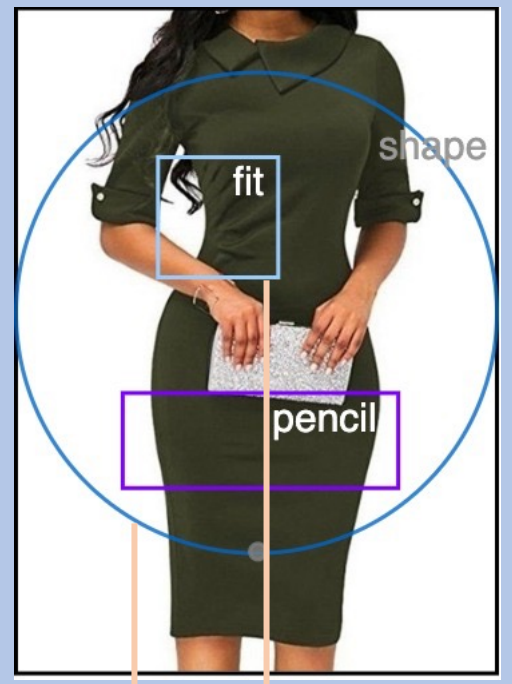
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

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

Amazon

Yang Jiao, Yan Gao, Jingjing Meng, Jin Shang, Yi Sun

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



Introduction

Attribute-specific representation learning

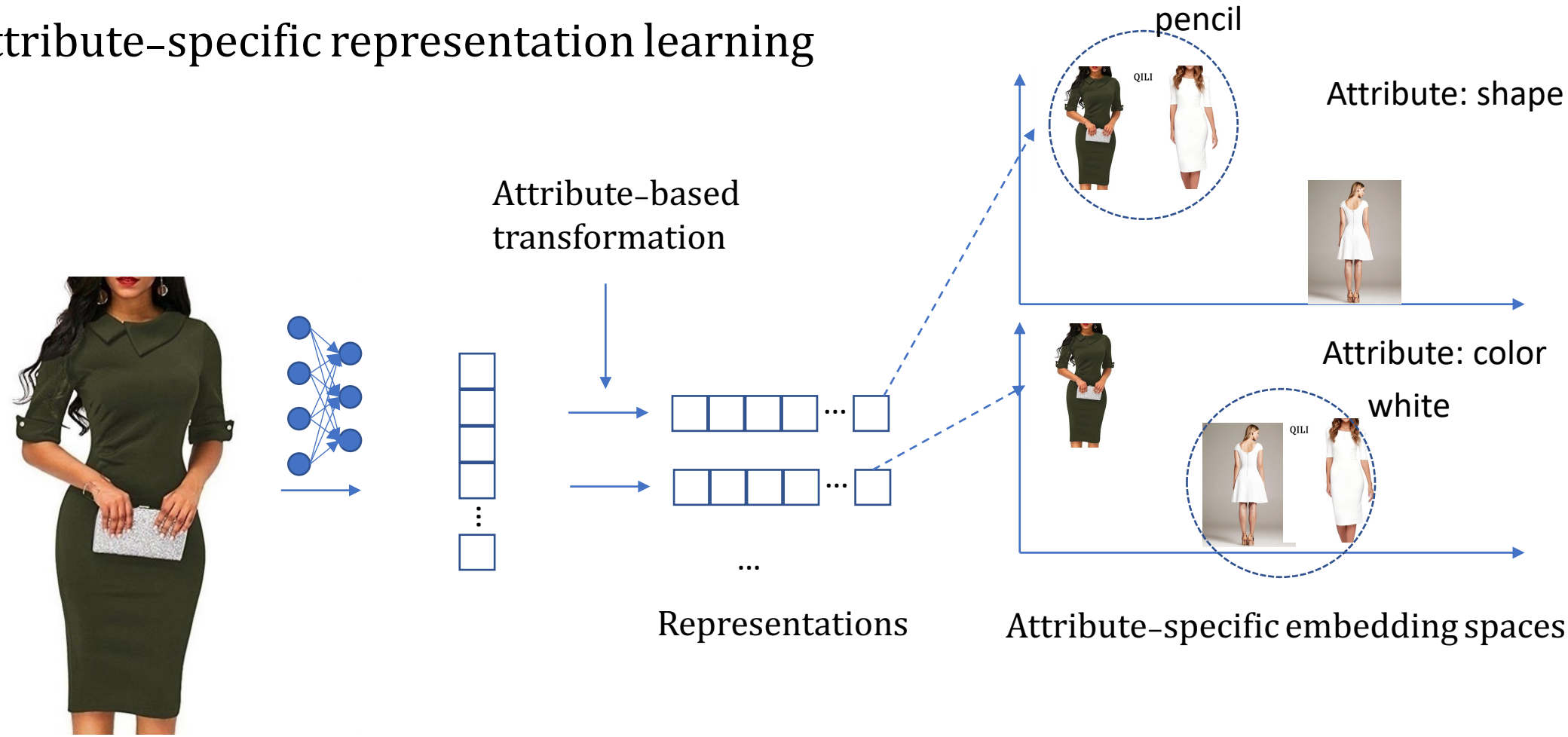


image → model → global representation → ***N*** attribute-specific representations → tasks

Problem 1: attributes are not fine-grained enough to discriminate fashion

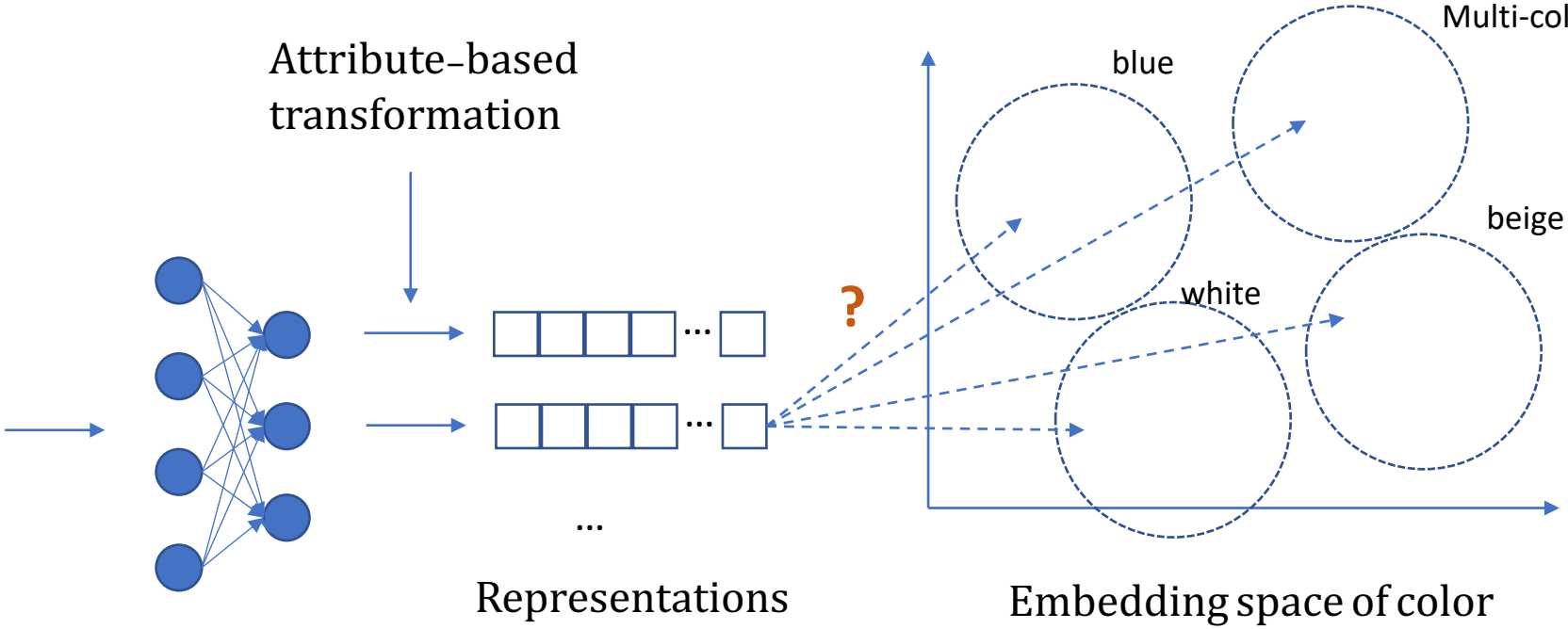


image → model → *N* attribute-specific representations → tasks

Problem 1

Solution：more fine-grained class-specific representation learning

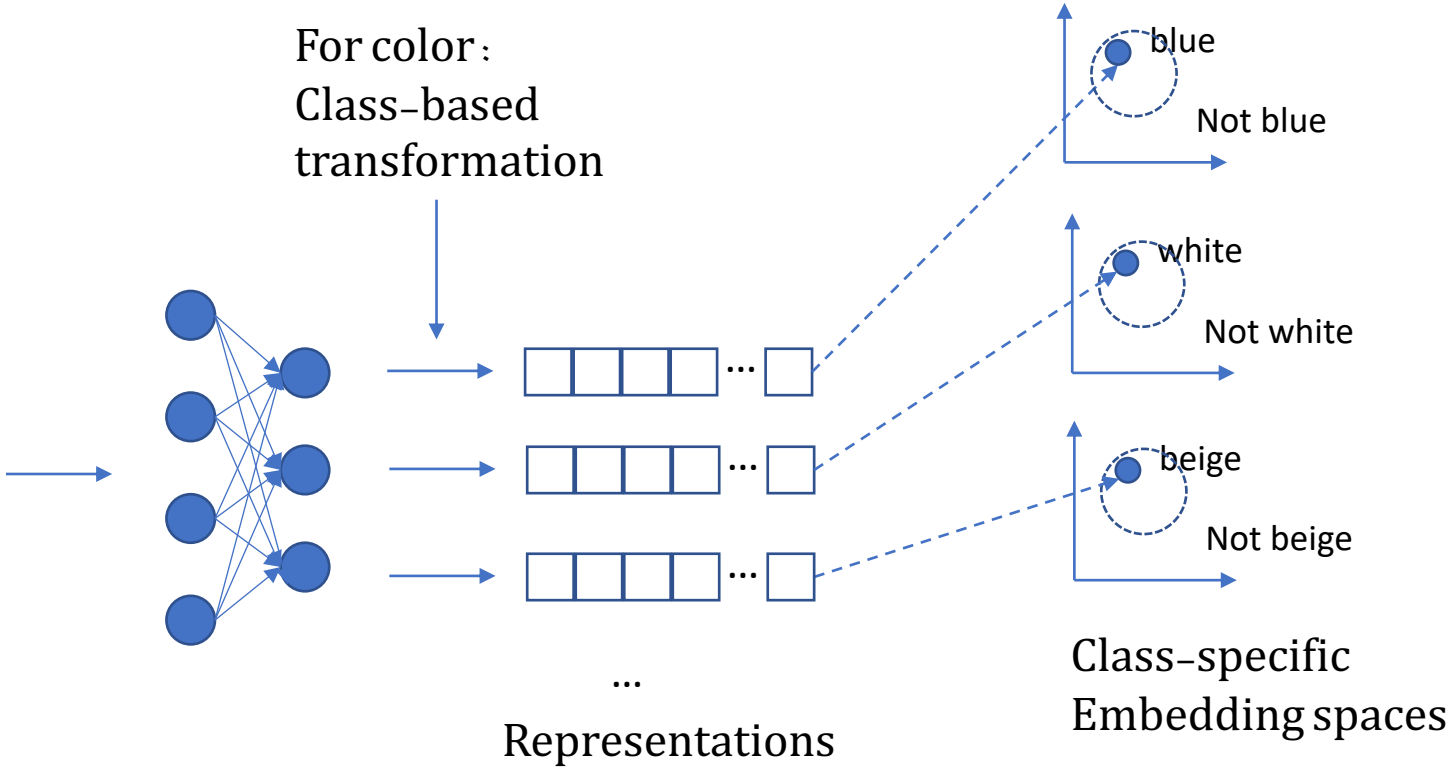
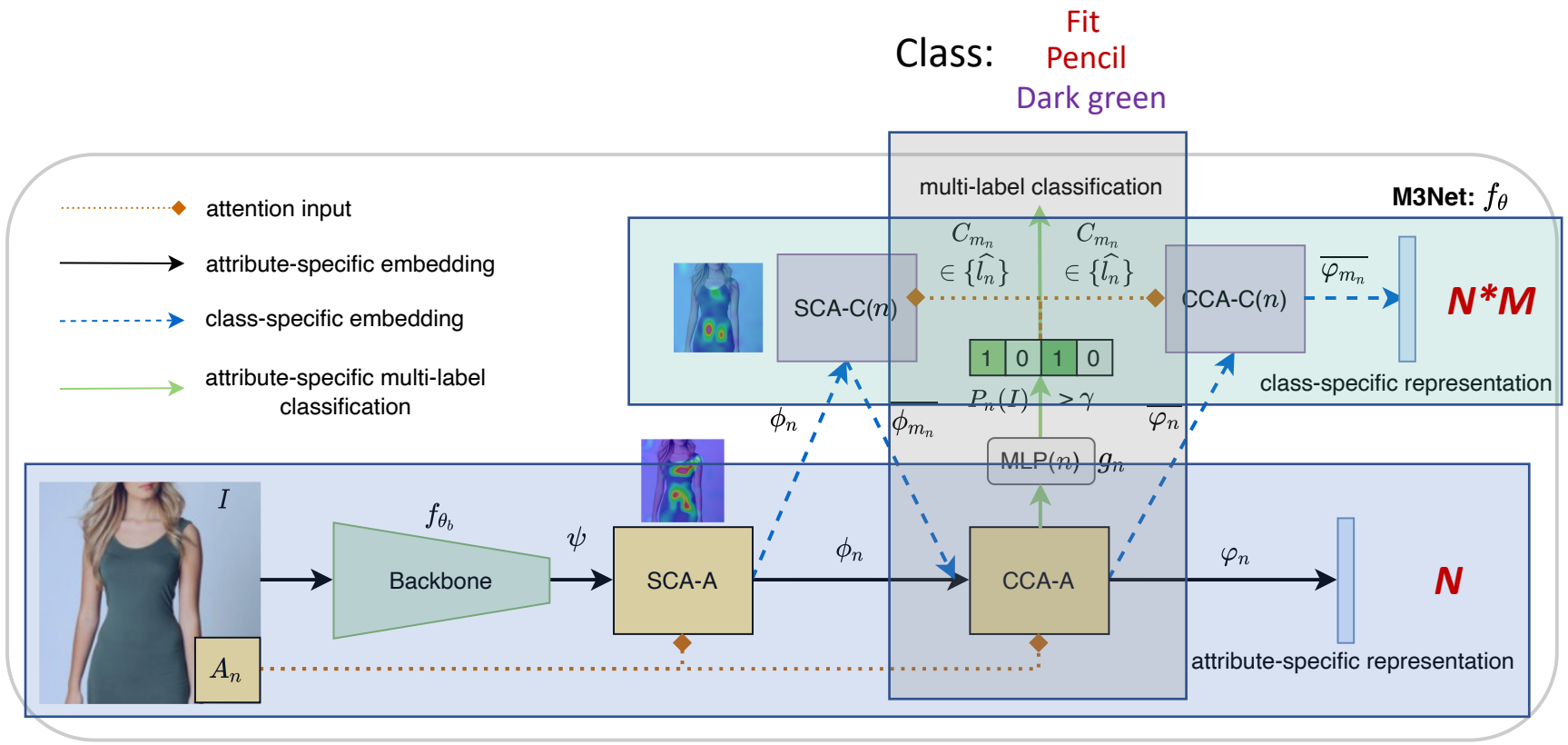
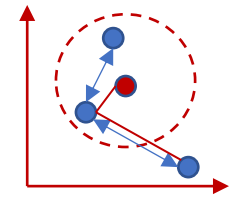


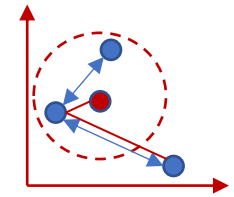
image → model → *M* class-specific representations → tasks

M3-Net: Multi-attribute multi-granularity multi- label network

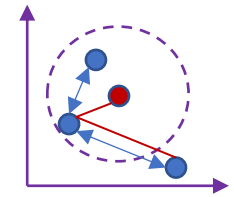




Fit-specific representation



Pencil-specific representation

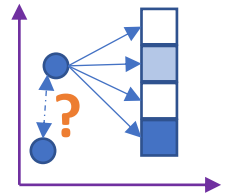


Dark-green-specific representation

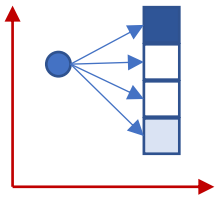
A_n: known attribute

Three key components:

- Attribute-specific representation learning
- Class-specific representation learning
- Multi-label classification



Color-specific representation



Shape-specific representation

Multi-granularity training objectives & loss functions

- ## Attribute-level objective

Loss 1: $$\mathcal{L}_{\mathcal{M}}(I, A_n|y_{m_n}) = \frac{1}{M_n}(\sum_{m=1}^{M_n}[-w_p y_{m_n} \cdot \log x_{m_n} + (1 - y_{m_n}) \cdot \log(1 - x_{m_n})]), \forall x_{m_n} \in P_n(I),$$

- I, A_n : inputs, image *I*, attribute *n*
- x_{m_n} : predicted probability of attribute *n*, class *m*
- y_{m_n} *: GT label of* attribute *n*, class *m*
- w_p*: weight of pos class for class balancing*

- ## Class-level objective

Loss 2: $$\mathcal{L}_{\mathcal{CI}}(I, A_n, C_{m_n}) = \mathcal{L}_{\triangle}(\overline{\varphi_{m_n}(I)}, \overline{\varphi_{m_n}(I^+)}, \overline{\varphi_{m_n}(I^-)}),$$ optimizes in a local view (instance-level)

Loss 3: $$\mathcal{L}_{\mathcal{CC}}(I, A_n, C_{m_n}) = \mathcal{L}_{\triangle}(\overline{\varphi_{m_n}(I)}, \overline{\varphi_{C^+}}, \overline{\varphi_{m_n}(I^-)}),$$ optimizes in a global view (cluster-level)

- I ,A_n, C_{m_n} : inputs, image *I*, attribute *n, class m_n*
- φ_{m_n}: image representation on attribute *n*, class *m_n*
- I, I+, I-: anchor, pos, and neg in a triplet
- φ_{C^+}: class center representation of class m_n

Experiments

- Datasets

Dataset	Attr type	# Attr	# Class per attr	Train/val/test
DeepFashion [18]	multi-label	5	156-230	220K/28K/28K
FashionAI [28]	single-label	8	5-10	144k/18k/18k
DARN [13]	single-label	9	7-55	163k/20k/20k

- Task: Fashion retrieval
- Metric: Mean Average Precision & Recall

- Baselines:
  - ASEN V1/V2 (AAAI 2020): attribute-based fashion representation learning
  - ASEN ++ (IEEE TIP 2021): cascade ASEN with multi-scale learning
  - MODC (ECCV 2022): multi-granularity fashion representation learning

Results

- DeepFashion (multi-label attributes)

Model	DeepFashion							
	MAP@all					MAP@all	MAP@100	Recall@100
Attribute	texture	fabric	shape	part	style	overall	overall	overall
ASEN [19]	21.03	11.61	14.68	7.81	4.66	12.33	20.60	5.10
ASEN$_{v2}$ [19]	21.86	11.67	14.58	7.93	4.68	12.51	20.55	4.99
ASEN++ [6]	22.20	11.71	14.70	8.15	4.72	12.74	20.79	5.21
MODC [14]	22.26	11.98	14.68	7.96	5.23	12.78	22.21	6.06
M3-Net$_a$	27.37	18.34	22.96	14.66	9.74	19.03	28.14	10.59
M3-Net$_c$	30.09	19.60	25.24	16.43	11.32	20.92	30.58	12.10
M3-Net	**30.79**	**20.20**	**27.11**	**17.28**	**11.61**	**21.78**	**35.08**	**12.88**
						+70.42%	+57.95%	+112.5%

M3-Net_a: only attribute-conditioning attentions
M3-Net_c: only class-conditioning attentions

Results

- DARN (single-label attributes)

Model	DARN											
	MAP@all									MAP@all	MAP@100	Recall@100
Attribute	clothes category	clothes button	clothes color	clothes length	clothes pattern	clothes shape	collar shape	sleeve length	sleeve shape	overall	overall	overall
ASEN [19]	36.62	46.01	52.76	56.85	54.89	56.85	34.40	79.95	58.08	52.75	58.72	20.26
ASEN$_{v2}$ [19]	37.97	49.24	52.26	59.13	55.32	59.06	36.86	81.54	58.82	54.29	59.66	20.88
ASEN++ [6]	40.21	50.04	53.14	59.83	57.41	59.70	37.45	83.70	60.41	55.78	61.09	21.51
MODC [14]	49.94	60.75	58.79	66.34	62.24	68.41	45.14	87.41	65.32	62.56	72.16	26.76
M3-Net$_a$	59.16	69.94	68.18	72.58	72.87	76.21	60.22	89.36	71.58	71.06	78.12	30.77
M3-Net$_c$	**60.98**	**70.91**	68.52	**74.95**	**74.48**	**79.15**	**63.09**	**90.31**	73.00	**72.79**	81.46	**31.64**
M3-Net	60.54	70.39	**69.52**	73.97	74.40	77.83	61.63	90.03	**73.94**	72.42	**82.69**	31.34

- FashionAI (single-label attributes)

Model	FashionAI										
	MAP@all								MAP@all	MAP@100	Recall@100
Attribute	skirt length	sleeve length	coat length	pant length	collar design	lapel design	neckline design	neck design	overall	overall	overall
ASEN [19]	64.61	49.98	49.75	65.76	70.30	62.86	52.14	56.42	57.37	64.70	22.77
ASEN$_{v2}$ [19]	65.58	54.42	52.03	67.41	71.36	66.76	60.91	59.58	61.13	67.85	24.14
ASEN++ [6]	66.31	57.51	55.43	68.83	72.79	66.85	66.78	67.02	64.27	70.62	25.30
MODC [14]	74.54	67.48	**68.25**	77.69	81.11	76.90	**77.46**	**77.10**	74.32	80.29	30.26
M3-Net$_a$	73.21	69.58	65.27	78.79	80.80	78.05	77.04	73.40	73.93	81.57	30.48
M3-Net$_c$	74.33	65.91	64.27	78.26	82.00	**78.85**	74.80	72.39	72.88	82.58	30.77
M3-Net	**75.27**	**70.04**	67.90	**79.31**	**82.82**	78.58	76.81	75.51	**75.04**	**87.04**	**32.01**

Results

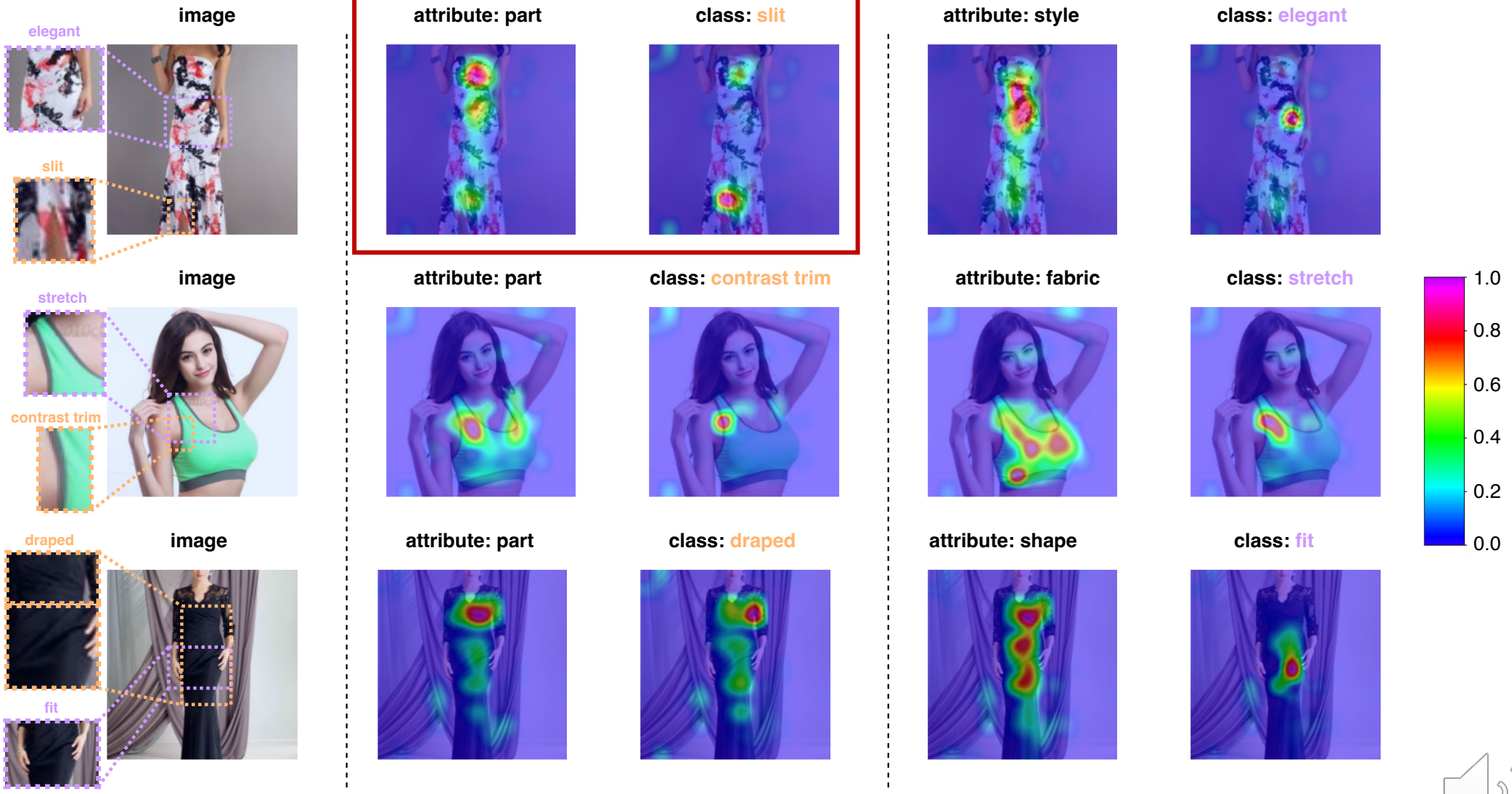
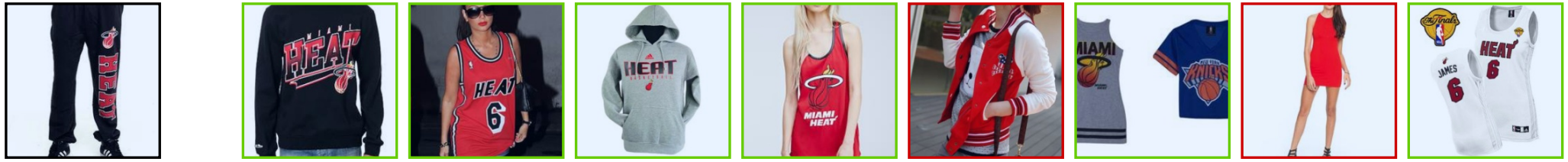


image
elegant
slit
attribute: part
class: slit
attribute: style
class: elegant
image
stretch
contrast trim
attribute: part
class: contrast trim
attribute: fabric
class: stretch
1.0
0.8
0.6
0.4
0.2
0.0
draped
image
fit
attribute: part
class: draped
attribute: shape
class: fit

Results

attribute: style
label: miami, heat

M3-Net



MODC

attribute: texture
label: print, tribal

M3-Net

MODC

amazon

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

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

Amazon

Yang Jiao

jaoyan@amazon.com