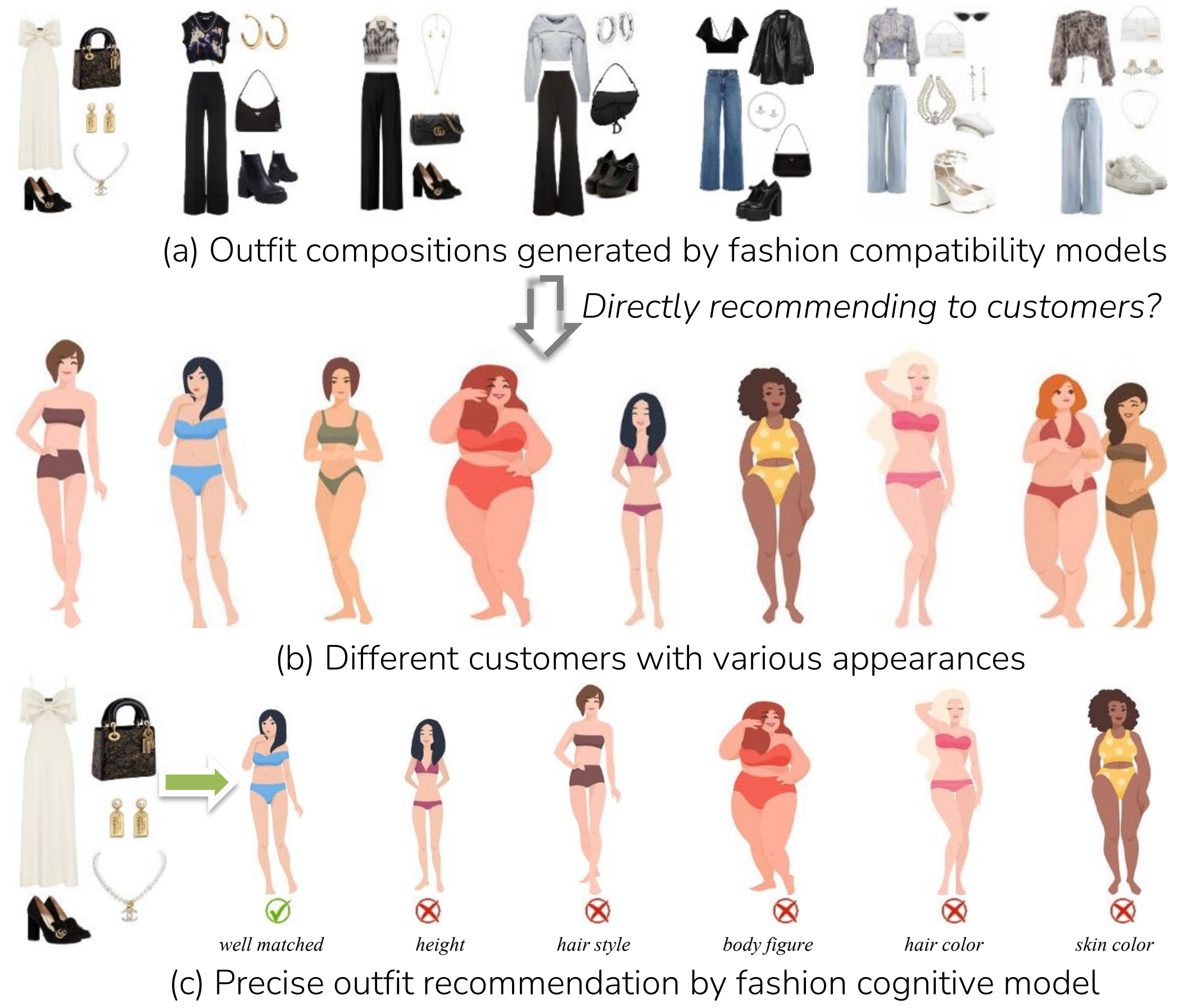
Dress Well via Fashion Cognitive Learning

Kaicheng Pang[1,2], Xingxing Zou[1,2], Waikeung Wong[1,2*]
[1]Laboratory for Artificial Intelligence in Design, [2]School of Fashion and Textiles, The Hong Kong Polytechnic University

Overview

Unlike previous research [4,10] focused on learning the fashion compatibility of an outfit, we introduce a new task, namely, fashion cognitive learning, which targets learning the compatibility relationships between outfits and personal physical information. A new outfit dataset with tremendous personal physical information and a new end-to-end framework called Fashion Convolutional Network is proposed to tackle this task. Through extensive experiments, our network outperforms several alternative methods with clear margins.


(a) Outfit compositions generated by fashion compatibility models

Directly recommending to customers?

(b) Different customers with various appearances

(c) Precise outfit recommendation by fashion cognitive model

Defined physical information

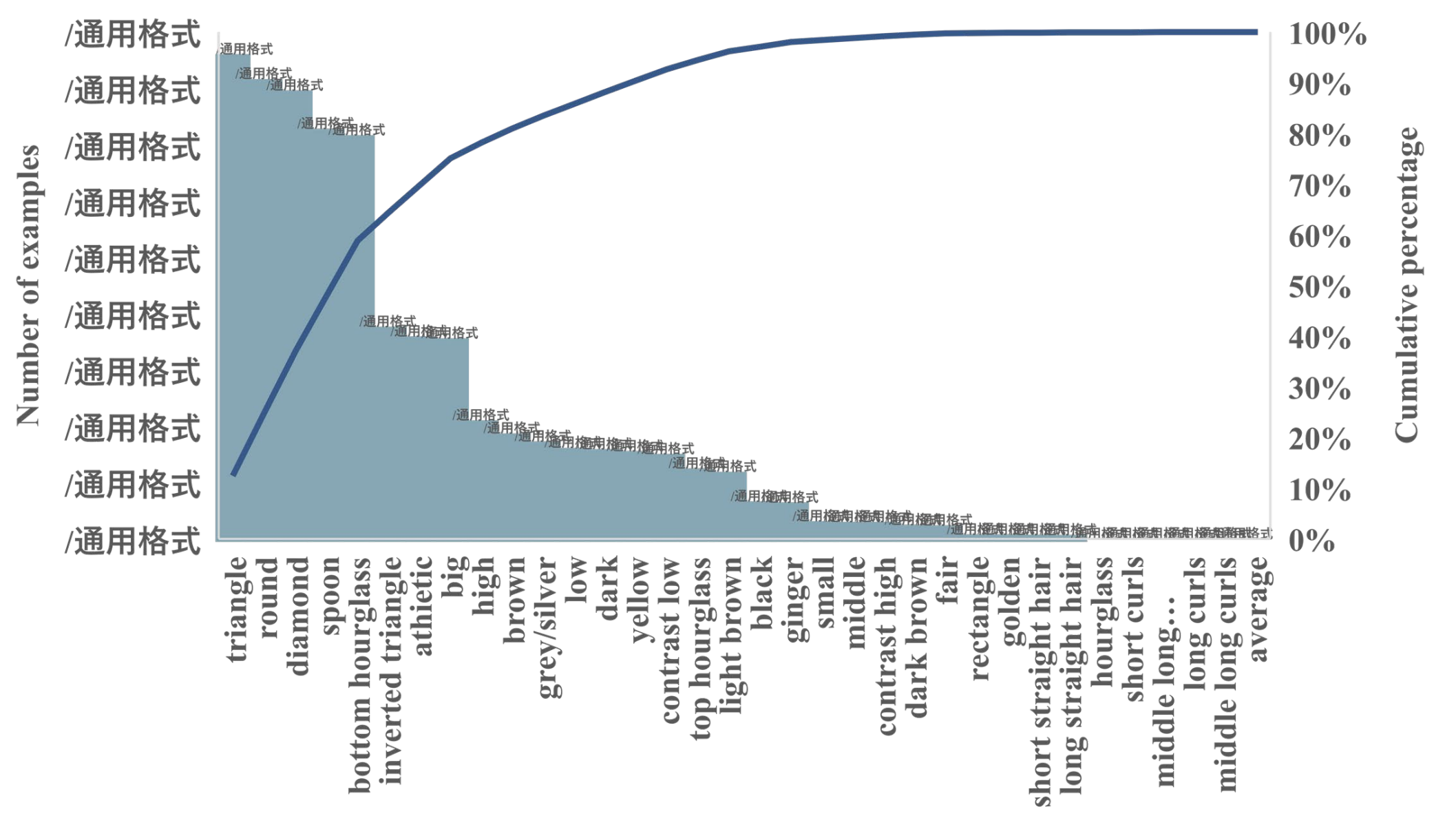
Features	Sub-features (N - numbers of sub-features)
Body Shape	rectangle, top hourglass, athletics... (10)
Skin Color	yellow, dark, fair, brown (4)
Hair Style	long curls, long straight hair... (6)
Hair Color	ginger,black, dark brown, light brown... (6)
Height	high, middle, low (3)
Breasts Size	big, average, small (3)
Color-contrast	high, low (2)

All data can be found at:
https://github.com/AemikaChow/AiDLab-fAshIon-Data

Approach

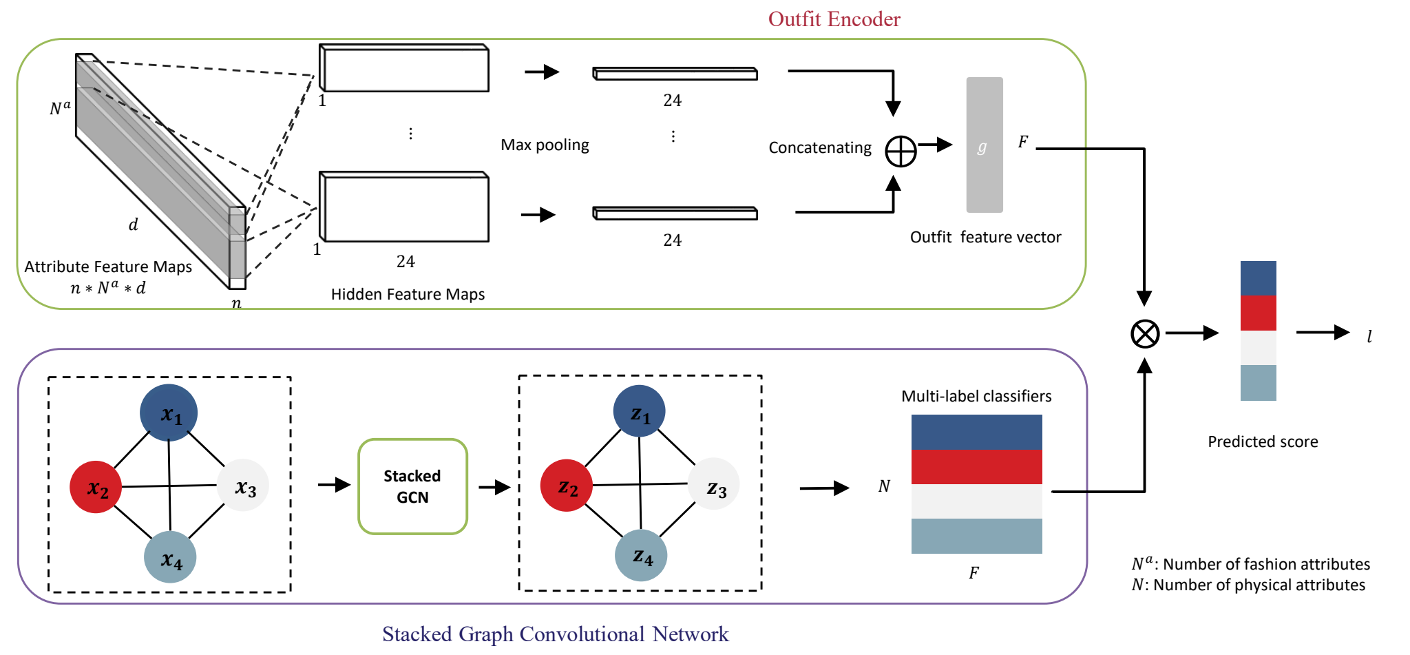
Outfit for You (O4U) Dataset

- Labeling system is carefully designed
- 15,748 outfits with 5.25 incompatible physical labels on average



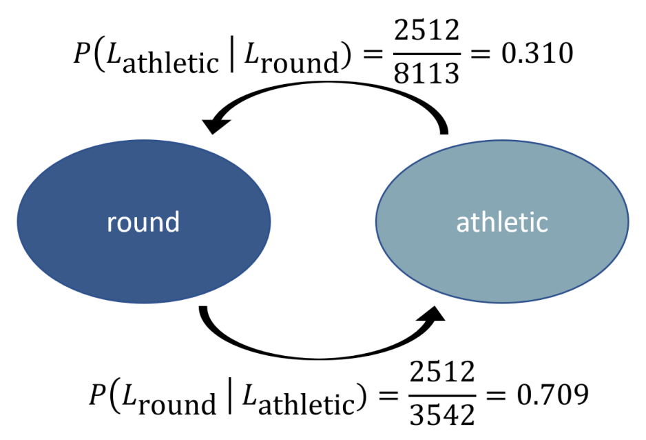
Fashion Convolutional Network

- Outfit Encoder: A convolutional layer to encode input outfit into outfit embedding
- Multi-label Graph Convolutional Networks: Learn label classifier based on label correlations
- Input space: Extracted attribute features from item images and physical label correlations represented by conditional probability.



Conditional Probability

If ‘athletic’ is not compatible with an outfit, there is a high probability (0.709) that ‘round’ is also not compatible with this outfit.


$P(L_{\text{athletic}} \mid L_{\text{round}}) = \frac{2512}{8113} = 0.310$
round
athletic
$P(L_{\text{round}} \mid L_{\text{athletic}}) = \frac{2512}{3542} = 0.709$

Experiments

Quantitative Results

We compared FCN with four baselines to show its effectiveness:

1. SVM [16]: The support vector machine.
2. Linear: A network consisting of multiple fully connected layers and ReLU activation functions.
3. ResNet [5]: ResNet with the input of the mean value of all item images.
4. Attention [23]: Several stacked multi-head attention layers.

The mAP results for 17 physical labels. Our proposed method FCN achieves the best performance over 14 out of 17 labels compared with other baseline methods. Especially on labels belonging to the body shape category, our method achieves a huge improvement compared to other methods.

	Body shape							
Methods	top hourglass	hourglass	athletics	inverted triangle	triangle	spoon	round	dimension
Linear	**15.47**	63.90	66.47	76.14	63.41	63.09	71.96	70.90
ResNet [5]	10.73	31.76	33.37	79.22	67.86	67.15	66.44	65.26
Attn [23]	9.48	30.20	31.69	69.68	59.07	57.46	61.36	61.52
FCN	15.39	**66.53**	**70.15**	**83.48**	**70.29**	**69.82**	**77.52**	**76.35**

	Skin			Hair color		Height		Breasts	Contrast
Methods	yellow	dark	brown	light brown	grey	high	low	big	low
Linear	11.59	24.35	14.35	9.17	13.31	17.38	13.94	31.23	12.27
ResNet [5]	12.05	41.57	14.29	**9.83**	**13.68**	18.08	11.98	26.50	12.06
Attn [23]	12.31	11.68	14.27	8.42	12.24	14.30	12.43	27.51	12.29
FCN	**13.24**	**46.84**	**15.11**	9.31	13.00	**21.57**	**23.23**	**31.91**	**12.72**

Model performances covering all 17 labels. FCN outperforms other baselines on most all metrics. SVM shows good performance on average overall metrics, while FCN surpasses SVM by 4.22, 0.74, and 2.66 on the CP, CR, and OP. The linear method works best in the recall indexes, indicating that this method may have a high sensitivity to the labels. The performance of ResNet is not good on mAP indicating that treating outfits as the mean value of item images is not a good idea for this task.

		All						Top-3					
Methods	mAP	CP	CR	CF1	OP	OR	OF1	CP	CR	CF1	OP	OR	OF1
SVM [16]	-	28.07	33.10	30.38	68.70	61.54	64.90	-	-	-	-	-	-
Linear	37.59	26.59	**33.93**	29.81	63.23	**65.14**	64.17	28.96	20.57	24.06	68.25	41.29	51.46
ResNet [5]	34.22	22.83	27.55	24.97	64.29	57.18	60.53	23.98	18.80	21.08	67.52	40.06	50.29
Attn [23]	29.76	18.18	29.41	22.47	61.82	62.33	62.07	11.44	17.65	13.88	64.82	39.22	48.87
FCN	**42.14**	**32.29**	33.84	**33.04**	**68.89**	62.17	**65.36**	**34.16**	**21.06**	**26.06**	**73.25**	**41.32**	**52.83**

Reference

[4]: Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S Davis. Learning fashion compatibility with bidirectional lstms. In Proceedings of the 2017 ACM on Multimedia Conference, pages 1078–1086. ACM, 2017.

[10]: Yen-Liang Lin, Son Tran, and Larry S Davis. Fashion outfit complementary item retrieval. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3311–3319, 2020.

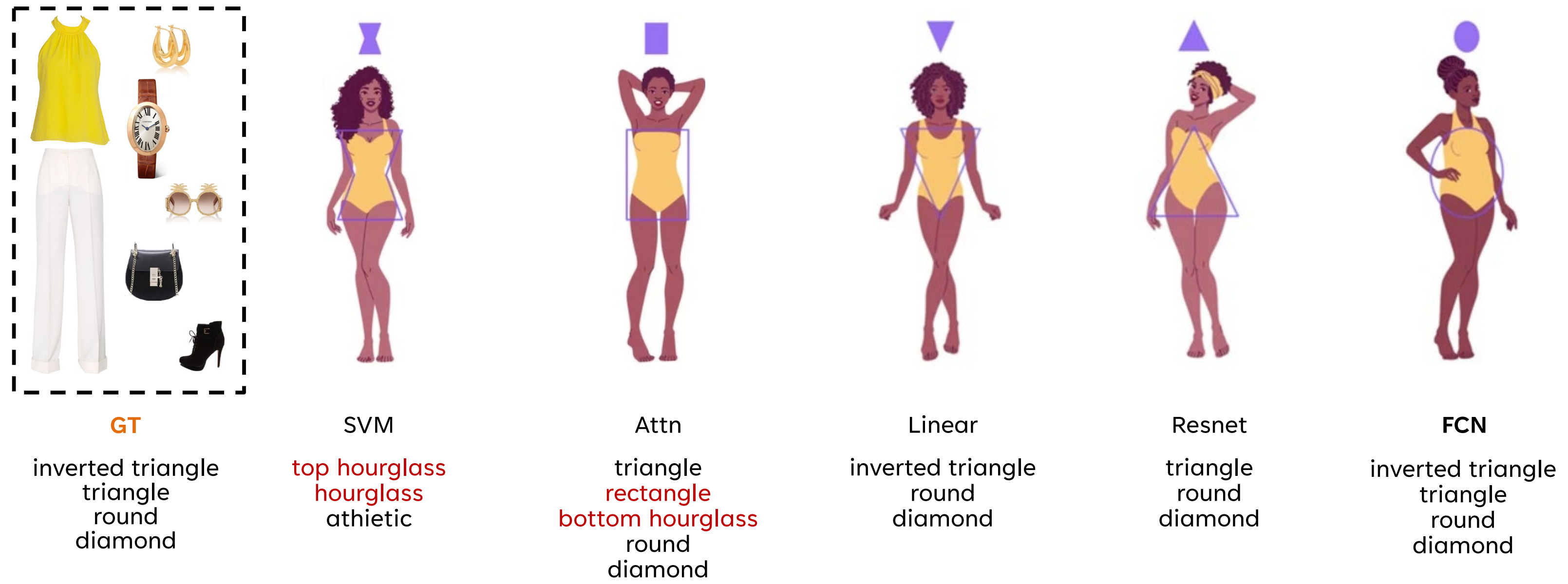
[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008, 2017.

Qualitative Results

Qualitative results of all compared methods and the proposed FCN. People with body figures including "inverted triangle", "triangle", "round", and "diamond" are not suitable for the outfit on the left side. The main reason is that the silhouette of the tank top and the straight-line pants are not matched these types of body shapes. The text in red is the wrong prediction. FCN precisely predicts all incompatible body shapes for the given outfit.



Ablation Study

1. Effect of filter region size. Only using one kind of convolutional filter size shows the worst performance. Using filters with a big region size has a negative effect on model performance. Using multiple filters with the same size achieves lower than FCN on the top-3 labels. The combination (1, 2, 4, 6, 8) shows the best performance on CF1 and Top-3 metrics.

		All						Top-3					
Region size	mAP	CP	CR	CF1	OP	OR	OF1	CP	CR	CF1	OP	OR	OF1
(1)	40.68	32.55	32.99	32.77	67.68	62.50	64.99	29.56	20.27	24.05	71.29	40.53	51.67
(2)	38.93	28.36	32.42	30.25	68.67	61.28	64.77	30.64	20.44	24.52	73.01	40.24	51.89
(4,4,4,4,4)	**43.11**	**32.82**	33.46	**33.13**	68.70	62.02	65.19	32.30	20.82	25.32	72.30	40.87	52.22
(8,9,10)	41.38	28.29	32.72	30.35	68.83	61.29	64.85	30.29	21.19	24.93	73.12	40.96	52.51
(1,2,4,6,8)	42.14	32.29	**33.84**	33.04	**68.89**	**62.17**	**65.36**	**34.16**	**21.06**	**26.06**	**73.25**	**41.32**	**52.83**

2. Effect of numbers of kernels for each filter. The performance achieves the best results when the number of kernels is 24. Using too few convolutional kernels will deteriorate performance significantly. Using too many kernels cannot dramatically improve performance, and it may cause a negative impact on recall metrics.

		All						Top-3					
No. Kernels	mAP	CP	CR	CF1	OP	OR	OF1	CP	CR	CF1	OP	OR	OF1
2	35.46	26.54	35.58	30.40	62.52	68.25	65.26	27.94	19.72	23.12	66.03	39.95	49.78
12	41.67	32.19	33.91	33.03	67.96	63.09	65.44	**36.09**	20.77	**26.37**	72.31	40.95	52.29
24	42.14	32.29	**33.84**	**33.04**	68.89	**62.17**	**65.36**	34.16	**21.06**	26.06	73.25	**41.32**	**52.83**
48	**42.65**	**32.55**	33.10	32.82	**69.17**	60.90	64.77	34.75	21.02	26.20	73.75	40.78	52.52

3. Effect of numbers of GCN layers. Deeper multi-layer GCNs degrade the performance on almost all metrics and two-layer GCN achieves the best performance.

		All						Top-3					
No. GCN	mAP	CP	CR	CF1	OP	OR	OF1	CP	CR	CF1	OP	OR	OF1
1	40.85	**34.11**	32.09	**33.07**	68.19	61.38	64.61	25.01	18.93	21.55	71.63	39.67	51.06
2	**42.14**	32.29	**33.84**	33.04	68.89	**62.17**	**65.36**	**34.16**	**21.06**	**26.06**	**73.25**	**41.32**	**52.83**
4	40.73	28.59	32.24	30.31	**69.05**	60.46	64.47	30.49	20.83	24.75	73.02	40.40	52.02
8	39.45	28.00	32.29	29.99	67.73	60.19	63.74	21.25	19.55	20.36	71.18	35.54	47.41