Learning the Latent "Look": Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images (Supplementary material)

Wei-Lin Hsiao
UT-Austin
kimhsiao@cs.utexas.edu

Kristen Grauman
UT-Austin
grauman@cs.utexas.edu

This document consists of:

- Vocabulary for predicted attributes in Section 3.3 of the main paper
- Examples of HipsterWars baseline clustering and DeepFashion top images in Section 4.1 of the main paper
- Qualitative example images for style retrieval experiment in Section 4.2 of the main paper
- Procedure of how we collect data for mix-style experiment in Section 4.3 of the main paper
- Procedure of how we align each discovered topic/cluster to a human pre-defined style in Section 4.3 of the main paper
- More examples of traversing style, as shown in Section 4.3 of the main paper
- List of style labels in DeepFashion, and its collapsed result referenced in Section 4 of the main paper

1. Vocabulary for predicted attributes

Table 1 shows the list of predicted attribute names, organized by types. We do not write out the resulting attributes that align to each body part here. After pairing *pattern*, *material*, *color* with body parts, we get 195 attributes in total.

pattern	material	shape	collar	article	color
crochet	translucent	skirt drape pleated	scoop	T-shirt	white
camouflage	leather	skirt drape prairie	vneck	blouse	black
floral	denim	skirt drape flat	square	jacket	red
geo		skirt length long	off-shoulder	blazer	pink
horizontal striped		skirt length medium	sweetheart	cardigan	orange
lace		skirt length short	turtle-neck	coat	yellow
leopard		skirt shape tight	shirt collar	vest	green
plaid		skirt shape loose		dress	blue
paisley		skirt shape full		skirt	purple
plain		pants loose		pants	brown
polka dot		pants flared		jeans	gray
tribal		pants peg-leg		leggings	beige
vertical striped		pants skinny		stocking	
zebra		pants short		boots	
				shoes	
				sunglasses	
				hat	
				belt	
				scarf	
				bag	
				socks	
				sweater	

Table 1: Predicted attributes organized by types.

2. More examples of top images

In our main paper in Section 4.1, we show top images for the five discovered topics on the HipsterWars dataset [1] using our method; here, we show the corresponding top images for the StyleNet CNN [3] and attribute clusters. Figure 1a are results for CNN clusters: Goth, Bohemian, and Preppy styles are succesfully discovered; however, there are 2 variants of Goth, one for pants and one for skirt, thus the baseline sacrifices discovering Pinup style. Figure 1b shows results for attribute clusters: images within each cluster are indeed visually similar, but superficial similarity does not lead to consistent styles.

Figure 2 shows top images for some sampled five topics on the DeepFashion dataset [2]. Notice that topic D is a style that could be discovered solely by PolyLDA: plain, scoop-neck T-shirt or tribal short skirt will each be an independent discovered style using MonoLDA, but with PolyLDA, it is possible to capture such compatibility of a whole outfit.

3. Qualitative example images for style retrieval experiment

In Section 4.2 in our main paper, we show quantitatively that our proposed PolyLDA method achieves the best combination of coherent style with diversity/novelty. In Figure 3 we show example retrieved images for the baseline methods and our method. Baseline methods often retrieve outfits that share the same colors, patterns, or clothing article composi-

(a) Top images for the five clusters with StyleNet [3]: map 2 variants of Goth into two clusters, and no cluster corresponds to the Pinup.

(b) Top images for the five clusters with attributes: discovered clusters are superficially visually similar, but inconsistent with style labels.

Figure 1: Top images for the baselines' discovered styles. Compare to ours in Figure 7 in the main paper. Labels indicate human-assigned styles from the HipsterWars datset [1].

Figure 2: Top images for five sampled topics discovered by our method for the DeepFashion dataset [2]. Topic D is a style that accounts for attribute composition across all body parts: plain, scoop-neck T-shirt or tribal short skirt will each be an independent discovered style using MonoLDA, but with PolyLDA, it is possible to capture such compatibility of a whole outfit.

tion as the query. However, outfits similar in these aspects do not necessarily share the same style, which could be seen in Figure 3a, Figure 3b, Figure 3d's attribute cases. Even when baseline methods retrieve style coherent outfits (Figure 3a's StyleNet, Figure 3c's attribute), the results are often too similar, which become near duplicates.

4. Data collection for mix-style experiment

In order to demonstrate the capability of our learnt representation in mixing styles, we collect 177 Web images for this experiment: we first collect 20 images for Hipster, Preppy, Goth, Bohemain styles, where these images have exclusively one of these four style labels (we

Figure 3: Examples of qualitative result for style retrieval experiment. On the left are query images, for each of which we show its nearest neigbors retrieved by Attr., Stylenet [3] and *PolyLDA*. Retrieved images with incorrect style labels are shown with their true labels.

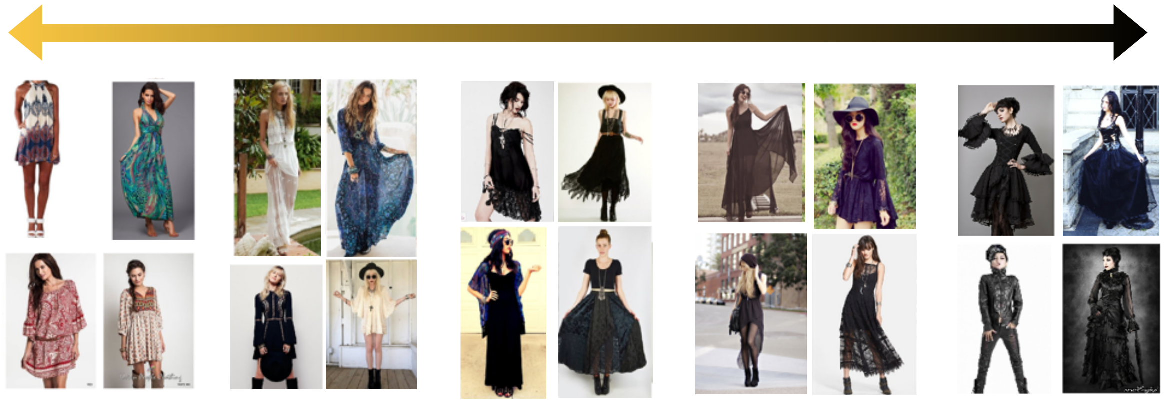
found that although all HipsterWars images have only one style label, many of them actually should have more than one style labels), giving us 80 images. Next, we collect about 20 images for each of the combination Hipster×Goth, Hipster×Bohemian, Bohemian×Goth, Preppy×Hipster, Preppy×Goth, resulting in 97 more images. The above collection process guaranteed the labels to be exclusive by using Google advanced image search to specify which styles to include and exclude.

5. Procedure of aligning discovered topics with human-labeled styles

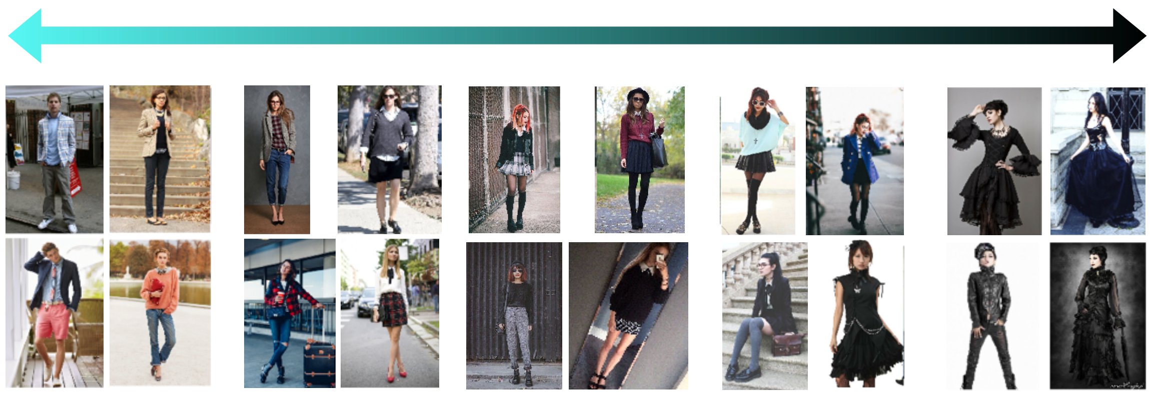
While in practice our approach can mix any selection of discovered topics, for sake of evaluation, we focus on blending GT-labeled styles. That way we can verify whether the results agree with GT labels. We do the alignment similar to how we did in Section 4.1: we use each topic's probability as a relevance score for a style to sort all images, then record the average precision (AP) per topic per style. The best AP a topic has in all styles is that topic's aligned style label.

6. More examples of traversing styles

Our learnt representation allows users to mix styles, and even traverse through styles. We have shown in Section 4.3 how we move from Bohemian to Hipster; here we show more examples like moving from Bohemian to Goth in Figure 4a and from Preppy to Goth in Figure 4b.



(a) Gradually traversing from Bohemian to Goth.



(b) Gradually traversing from Preppy to Goth.

Figure 4: More examples of mixing styles using our proposed method. These examples augment the ones shown in Figure 9 of the main paper. Note how the same Goth style individually mixing with 2 different styles, Bohemian and Preppy, demonstrates different effects: tribal pattern in Bohemian is substituted by translucent lace when mixed with Goth, and the warm color tone in Preppy is overridden by Goth's dark color tone.

7. List of style labels in DeepFashion

Each image in the DeepFashion dataset [2] comes with 1000 attributes, and each attribute belongs to one of the five categories: texture, fabric, shape, part and style. We use the attributes that belong to *style* as our style labels. However, since these styles are mined from the Internet instead of labeled by humans, we can see in Figure 5a that it is a little noisy (giving identical images different labels), and in the full list of styles in Table 2 we can see that many styles are ambiguous (i.e. tacos, pizza, pineapple; ny, nyc, new york are 3 different styles.) As a result, we cluster the style labels into higher level styles by their co-occurrence in images. As noted in Section 4.1 in the main paper, we use affinity propagation with cosine similarity. The resulting set of style labels is shown in Table 2. To demonstrate that our resulting collapsed style labels capture higher level styles, we show example images for three groups of styles in Figure 6.

References

[1] M. H. Kiapour, K. Yamaguchi, A. Berg, and T. Berg. Hipster wars: Discovering elements of fashion styles. In *ECCV*, 2014. 1, 2

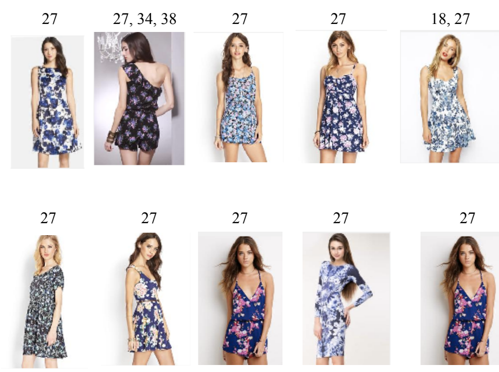
[2] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *CVPR*, 2016. 1, 2, 4

[3] E. Simo-Serra and H. Ishikawa. Fashion Style in 128 Floats: Joint Ranking and Classification using Weak Data for Feature Extraction. In *CVPR*, 2016. 1, 2, 3

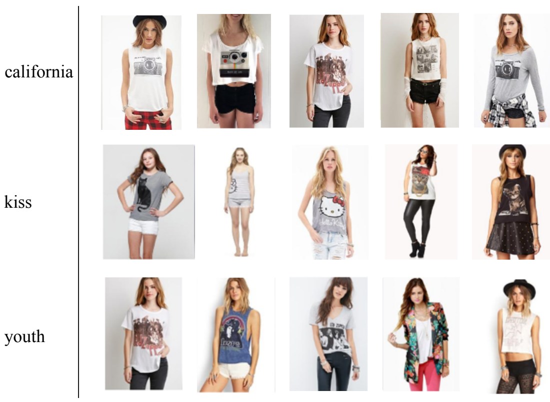
1	americana	lady	smart	sporty	taco
2	athletic	dream	garden party		
3	california	kiss	youth		
4	cardio	classic	lightning		
5	cities	ny	yoke		
6	coffee	night	reversible	roses	shark
7	beatles	eagle	france	rainbow	
8	bike	fancy	loyal	studio	
9	candy	enchanted	festive	heroes	red
10	baseball	field	island		
11	dark isle	fisherman regime	flawless safari	free spirit texas	 tokyo
12	fox	mandarin	paradise		
13	fresh	joie			
14	audrey	basquiat	frida	laser	tasmanian
15	galaxy	garden	paris	wild	
16	coast	guns	tupac		
17	inset	nyc	retro		
18	beach everyday mob	blurred fan springs	darling kitty sweet	dreamer la utility	dynamite logo yoga
19	babydoll	edge	love	wave	
20	marilyn	modernist			
21	bed matelot shopping	boyfriend new york soft	chic notorious star	ethereal quirky summer	luxe rad swiss
22	dreamcatcher	meow	pan		
23	miami	mickey	minnie		
24	kid	mod			
25	moon	oxford	rebel	spirit	
26	cat	popcorn	seaside		
27	angeles boho dainty elegant internet minnie mouse pizza roman standout thermal	art bold devil flirty kahlo mirrored please rustic sunburst trench	babe brooklyn nets doodle genuine killin monroe power smile surfer triangle	barbie camera eiffel girls lounge morning relaxed snap sweetheart workout	biker civil halen map pineapple rolling stones snoopy swim zeppelin
28	grunge voyager	hepburn	marilyn monroe	performance	raga
29	city	pink	refined	tropical	
30	brooklyn	reverse	rugby	venice	
31	roll	rolling			
32	pj	running	weekend		
33	floyd	light	sea		
34	shore	sky	solid	west	
35	girl	muse	sunflower		
36	tower	track	trouble		
37	basic	training			
38	bella	doll	lakers	tree	wildflower
39	spongebob	van	varsity		
40	blah life stars	cute lover wake	daring mickey mouse	defyant mina	desert party
41	destroyed	rose	wifey	york	
42	heat	posh	run	sun	woke

Table 2: Collapsed label groups for DeepFashion styles: each row corresponds to a group of styles. *Minnie* and *mickey* are in the same group; *cities* and *ny* are also in the same group.

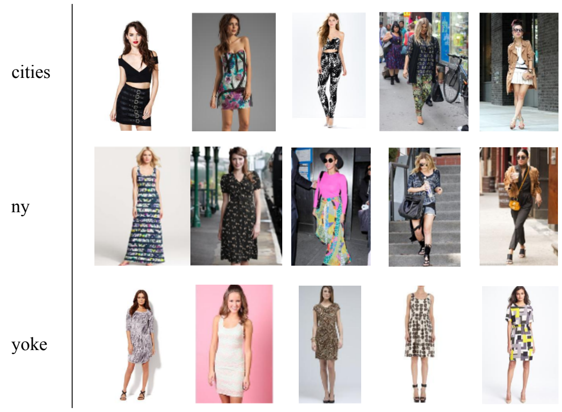
(a) Example of noisy labels: these are nearest neighbor images with their raw style labels as supplied with DeepFashion. Red are apparently incorrect labels; purple are different labels given to duplicate images.



(b) Example of automatically grouping noisy labels with affinity propagation: after grouping, these nearest neighbor images share more styles in common than before.



(a) Example images for collapsed style label 3: Images in this group are mostly wearing scoop neck, graphic T-shirt.



(b) Example images for collapsed style label 5: Images in this group are mostly wearing dress.

(c) Example images for collapsed style label 42: Images in this group are mostly wearing exercising outfit.

Figure 6: Example images for the collapsed label groups. Group 3 and group 42 are cleaner groups. Due to the noise in the original labels, it is still a little difficult to tell what style group 5 is exactly trying to capture.