Runway to Realway: Visual Analysis of Fashion

Sirion Vittayakorn[1] Kota Yamaguchi[2] Alexander C. Berg[1] Tamara L. Berg[1]
University of North Carolina at Chapel Hill[1] Tohoku University[2]
{sirionv,acberg,tlberg}@cs.unc.edu kyamagu@vision.is.tohoku.ac.jp

Abstract

Clothing and fashion are an integral part of our everyday lives. In this paper we present an approach to studying fashion both on the runway and in more real-world settings, computationally, and at large scale, using computer vision. Our contributions include collecting a new runway dataset, designing features suitable for capturing outfit appearance, collecting human judgments of outfit similarity, and learning similarity functions on the features to mimic those judgments. We provide both intrinsic and extrinsic evaluations of our learned models to assess performance on outfit similarity prediction as well as season, year, and brand estimation. An example application tracks visual trends as runway fashions filter down to "realway" street fashions.

1. Introduction

Clothing conveys one aspect of an individual's choices about how they represent themselves to the world. While some choices are truly unique, fashion is also heavily driven by trends, similar aesthetics followed by groups of people that appear and reappear cyclically over time. Seasonal trends like floral for spring or reds and oranges for fall recur without fail each year, while other trends pop up more sporadically, often appearing first on the runway and then filtering quickly into the real world. For example, we saw a burst of neon colors reminiscent of the 90s in Spring 2012 runway collections from Rodarte, Peter Som, Jason Wu, and Nanette Lepore (among others), and have been observing these color palletes ever since in fashion followers' closets.

Fashion is an excellent domain for applying computer vision. Since fashion is not only fundamentally visual, but also enormous – the fashion industry employs roughly 4.2 million people, with an annual global revenue of 1.2 trillion dollars[1] – computational techniques could make a valuable contribution. With over 2 million online blogs and stores devoted to fashion [7], millions of images are readily available for computer vision applications.

In this paper, we present the first attempt to provide a quantitative analysis of fashion both on the runway and in the real world, computationally, and at large scale, using computer vision. To enable this study, we collect a large new dataset called the *Runway Dataset*, containing 348,598 runway fashion photos, representing 9,328 fashion show collections over 15 years. Using this dataset we develop computer vision techniques to measure the similarity between images of clothing outfits. Additionally, we make use of the Paper Doll dataset [26] collected from the Chictopia social network, where people interested in fashion post and annotate pictures of their daily outfits. The combination of these two datasets allows us to make initial steps toward analyzing how fashion trends transfer from runway collections to the clothing people wear in real life.

To produce quantitative analyses for fashion and trends we first need methods for representing clothing appearance and for measuring similarity between outfits. Therefore, we develop a feature representation that can usefully capture the appearance of clothing items in outfits. This representation first estimates the pose of a person [27] and what they are wearing in the form of a clothing parse [26], and then computes pose-dependent features related to clothing style. To understand visual relationships between different outfits, we collect human judgments of outfit similarity using Amazon Mechanical Turk. Using these labels we train multiple models of fashion similarity to: a) predict similarity between pairs of runway outfits and b) predict similarity between runway and street fashion outfits. Evaluating our learned models to estimate human judgments of similarity, we find that we can predict similar outfits quite well.

Since fashion similarity may be considered potentially nebulous and subjective, we also evaluate our learned similarity measures on several extrinsic, objective, tasks: predicting the season, year, and brand depicted in a runway photograph. Though these are tremendously challenging tasks, we find that nearest neighbor techniques based on our learned similarity models perform surprisingly well, even *outperforming humans* on the same tasks. We also find that even though our models were trained for a different task (similarity prediction), we can achieve comparable performance to classifiers trained directly for these extrinsic tasks.

[1] http://www.statisticbrain.com/fashion-industry-statistics/

Finally, we apply our measures of similarity to our broader goals of improving our understanding of fashion with preliminary investigations of predicting fashion trends.

1.1. Related Work

There has been increasing interest in clothing recognition from computer vision researchers. Recent work addresses attributes of shopping images of clothing items [2, 20], adding some clothing attributes to the person detection pipeline [4], and detecting clothing items and annotating attributes for an image or collection of images [6, 3]. More closely related to the parsing approach used in this paper, there has been work on predicting semantic segmentations of clothing images [24, 25, 26], and we use the open source implementation from [26] as part of our pipeline. There is also related work on recognizing attributes or concepts that are strongly related to clothing, for example, occupation [22, 21] or social identity [15]. In this paper, we take a closer look at learning human based judgments of outfit similarity and also look at detecting visual trends in fashion over time.

A key aspect of this problem is identifying subtle differences in fashion styles that are not fully captured by listing items of clothing or recognizing a small number of attributes or styles. This challenge could be viewed as a form of fine-grained categorization, studied in other types of data e.g., for birds and faces [10, 28, 1, 11]. Typically fine-grained categorization approaches require coarse correspondence of keypoints in order to reliably extract subtle differences in appearance. In this paper, we make use of pose estimation [27] and clothing parsing [26] to automatically localize important regions in a fashion outfit. Using these as an input parse, our approach learns to combine features computed over parsed regions in order to mimic human judgments of fashion similarity.

We then use the learned similarities for a variety of evaluations – predicting human judgments of similarity, comparing runway fashions to "realway" fashions worn in the real world, and identifying clothing by season, designer, and year. These later tasks, predicting year and season for clothing, dovetail with recent work trying to predict when a historical photo was taken based on its distribution of colors [19], or identifying the year or make of a car from discriminative image patches [16]. Clothing style is especially difficult to analyze due to pose variation and deformation, and the nuanced variation from year to year.

Clothing applications

Our work is related to the street-to-shop clothing retrieval by Liu et al. [18] in that both attempt to connect street fashion pictures with a different domain. While [18] retrieves images of clothing items from a shopping-image database given a street snapshot, one of our applications is to find everyday street *outfits* that are similar to outfits in runway collections. Two key differences are the direction of the matching and the focus on outfits rather that clothing items. Our trend analyses could also enable new fashion applications, for example, by improving market analysis of fashion brands [13], interactive search [14] or recommendation [12, 17] of trending styles.

1.2. Approach

Our overarching goals are to use computer vision to quantify and learn measures for similarity of outfits and to discover trends in fashion. In this paper we consider both runway fashion – where we collect and process a new dataset of runway photos from a runway fashion accumulator (style.com) – and "realway" pictures of outfits worn in the real world sampled from the Paper Doll dataset [26] (Sec. 2).

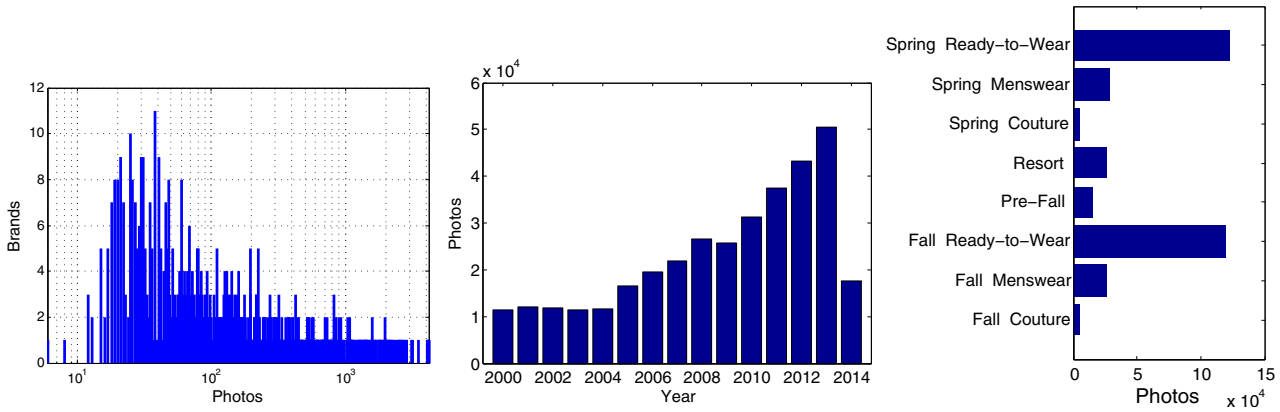
The subroutine in our approach is pairwise comparison of clothing outfits based on a learned combination of features aggregated over semantic parses of clothing. Some of our innovations are in defining features over the semantic parse of clothing, and in learning a comparison between features that replicates human ratings of similarity (Sec. 3). We use Mechanical Turk to collect human judgment of similarity and train discriminative models to approximate these judgments.

To evaluate the learned similarity models, we first check how well the learned similarity replicates human judgments of similar outfits. Then, we also show that the learned similarity can be used in a voting based technique for predicting the season, brand, and even year for a collection (Sec. 5.1). Finally we show that the learned similarity can be used to study how runway fashions relate to realway fashions on the street (Sec. 6).

2. Data

Runway Dataset We introduce a new dataset, the *Runway Dataset*, which consists of runway images from thousands fashion shows of a wide variety of brands over several years. These 348,598 images of 9,328 fashion shows were collected from *style.com*, and cover 15 years, from 2000 to 2014 together with their meta data which includes season (e.g., Spring 2014), category (e.g., Ready-to-Wear, Couture), brand name, date of the event, and a short text description. Examples of the images can be seen in Fig. 2 along with predicted semantic parses of the outfits.

In the dataset, we observe 852 distinctive brand names, ranging from haute couture designers like Chanel or Fendi, to more common consumer brands like JCrew or Topshop. Fig. 1a shows a histogram of the number of photos collected per brand. Most brands have between 10 to 100 photos, while a few brands have significantly more. Fig. 1b shows



(a) Photos per brand (b) Photos per year (c) Photos per season

Figure 1: Style dataset statistics.

Figure 2: Example images from our new *Runway Dataset* with estimated clothing parses. 1^{st} row shows samples from Betsey Johnson Spring 2014 Ready-to-Wear, 2^{nd} row from Carolina Herrera Fall 2011 Ready-to-Wear, 3^{rd} row from BURBERRY Fall 2011 Ready-to-Wear, and 4^{th} row from Alexander McQueen Pre-fall 2012 collection.

the number of photos in the dataset over time, and Fig. 1c shows number of photos per season. Here, season refers to a specific fashion event (e.g., fashion week). The Ready-to-Wear shows in spring and fall are the most common events.

Paper Doll dataset The Paper Doll dataset [26] is used to sample realway photos of outfits that people wear in their everyday lives. This dataset contains 339,797 images collected from a social network focused on fashion called Chictopia. We use outfit photographs from the Paper Doll dataset to test retrieval of realway outfits that are similar to runway outfits, and to study how realway fashions correlate with runway fashions.

3. Visual representation

Our visual representation consists of low-level features aggregated in various ways over a semantic parse using the clothing parser of Yamaguchi *et al.* [26], which in turn uses Yang & Ramanan's pose estimation [27] to detect people and localize body parts. As a pre-processing step, we first resize each person detection window to 320×160 pixels and then automatically extract 9 sub-regions defined relative to the pose estimate. Sub-regions are defined for the head, chest, torso, left/right arm, hip, left/right/between legs, as illustrated in Fig. 4. Features measure: color, texture, shape, and the clothing parse's estimates of clothing items worn,

(a) Runway to Runway (b) Runway to Realway

Figure 3: Retrieved similar outfits for example query runway outfits (red boxes) using our learned similarity. On the right we show retrieved outfits from everyday, realway outfits, on the left we show outfits retrieved from other runway collections.

	Runway to Runway		Runway to Realway	
Method	our feature	Style descriptor	our feature	Style descriptor
Majority	0.76 ± 0.11	0.66 ± 0.11	0.54 ± 0.03	0.45 ± 0.02
Unanimity	0.73 ± 0.08	0.62 ± 0.07	0.53 ± 0.02	0.42 ± 0.01
Some	0.73 ± 0.14	0.63 ± 0.12	0.55 ± 0.01	0.43 ± 0.03

Table 1: Intrinsic Evaluation: AUC for predicting outfit similarity from Runway images to Runway images or from Runway images to Realway (street-style) images.

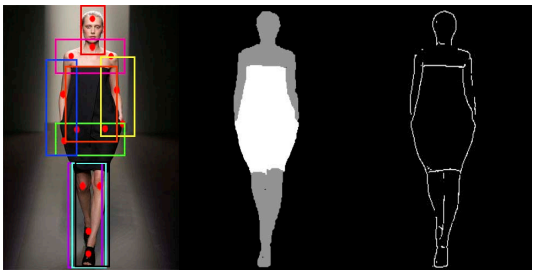


Figure 4: Visual representation from left to right: (a) Pose estimation and corresponding bounding boxes (b) Foreground binary mask and (c) Binary edge maps

along with an existing global style descriptor [26].

Color For each sub-region, we extract two 512 dimensional histograms in RGB and Lab color spaces, from only pixels parsed as foreground (clothing items, hair and skin).

Texture For each sub-region, we extract two bag-of-words histograms from the pixels parsed as foreground. The first is a histogram of MR8 responses [23] quantized into 256 visual words. The second is a histogram of HOG descriptors [8] (8x8 blocks, 4 pixel step size, 9 orientations) quantized into 1000 words.

Shape We resize each sub-region to 32×16 pixels, and extract 2 features: a binary mask of the foreground pixels estimated by the clothing parsing algorithm, and an edge map of the foreground region estimated using structured forest efficient edge detection [9].

We select the threshold, t, to binarize the edge map, in an image-dependent manner. Using the image's parse as guidance, we find a threshold that provides information about shapes of the parsed regions without too many extraneous pixels by minimizing the following cost function:

$$c(t) = \sum_{i \in x} d\left(x_i, \bar{x}_j\right) + \sum_{j \in \bar{x}} d\left(\bar{x}_j, x_i\right) \qquad (1)$$

where $d(x_i, \bar{x}_j)$ is the Euclidean distance transform of pixel i of binary edge map x binarized at threshold equal to t to the nearest pixel j of the clothing contour $\bar{x}$. This provides an edge map that has a similar level of detail to the predicted clothing parse boundaries (Fig 4).

Parse For each sub-region, we extract individual item-masks for each of the 56 different clothing categories (e.g. dress, shirt, shoes, etc) then form a 56-dimensional descriptor of the percentage of each item present in the sub-region.

Style descriptor We also include as a global feature, the style descriptor of [26] that is computed over the entire body and includes RGB, Lab, MR8, Gradients, HOG, Boundary Distance and Pose Distance. PCA is used to reduce dimensionality from 39,168 to 441 dimensions.

These five features (**c**olor, **t**exture, **s**hape, **p**arse and **style** descriptor) are concatenated to form a vector for each sub-region used for learning outfit similarity models (Sec 5).

4. Human judgments

Determining "similarity" between images of clothing outfits may be difficult, even for humans. To address this challenge, we collect multiple human judgments of similarity between fashion pictures using crowdsourcing. Here, agreement between people, where present, provides a wisdom-of-crowds definition for similarity between styles.

Data collection Given a query outfit, we ask labelers on Amazon Mechanical Turk (MTurk) to pick the most similar outfit from a set of 5 pictures, or *none* in case of no similar outfits. We select these 5 outfits based on cosine similarity using each individual feature in isolation (e.g., color, texture), or to an equally weighted combination of all features.

We collect human judgments for 2000 random query images from the runway dataset. We choose the 5 similar candidates for each query under two scenarios; selecting the candidates from the runway dataset, and selecting the candidates from the Paper Doll, realway dataset. We collect similarity annotations from 5 labelers for each query image and scenario.

Results Overall, we observed good agreement between labelers for the within-runway scenario. For 20.4% of queries, all five labelers agreed on the most similar outfit. For an additional 29.8% and 24.6% of queries three and four out of five labelers agree on the best match. In total, we observed the majority of labelers agreeing for 74.8% of the queries in the within-runway scenario.

The agreement is a little lower in the runway-to-realway scenario where all five labelers agreed in 10.9% of queries, and 39.3% and 23.7% of queries three and four out of five labelers agree on the best match. This yields majority agreement for 73.9% of the runway-to-realway queries.

5. Learning to compare outfits

Using the human judgments, we learn a model of similarity for fashion outfits. To do this, we train a linear SVM [5] to classify a pair of pictures as similar or dissimilar. We consider the following strategies for converting human judgments to positive/negative labels for training:

Majority: An image-pair is marked as positive when the pair gets the *majority* of labeler clicks. Any query for which all the labelers clicked "none" is used to form 5 negative pairs with each of its 5 potentially similar images.

Unanimity: Query-image pairs for which all-five labelers agree on the best match are treated as positive. Any query for which all the labelers clicked "none" is used to form negative pairs with each of its 5 potentially similar images.

Some: Image pairs marked by any of the five labelers are treated as positive. And any query for which all the labelers clicked "none" is used to form 5 negative pairs with each of its 5 potentially similar images.

Qualitative Evaluation Example results are shown in Fig. 3. On the left are examples where the query is a runway outfit and the retrieved results are also sampled from runway collections. On the right are example results where the query is a runway image and the retrieved outfits are from the realway Chictopia dataset. In both cases the model used is the majority training approach. Outfits retrieved both from the runway and from the realway images look quite promising, colors are matched well, and shape and pattern also tend to be similar. We are even able to match challenging visual characteristics quite well such as neckline shape (e.g. white strapless dress, bottom row left), and overall garment shape (e.g. red a-line dress, 2nd row right).

Quantitative Evaluation We evaluate the performance of our model by area under the precision recall curve (AUC), using 6-fold cross validation. Results for the runway-to-runway and runway-to-realway scenarios are shown in Table 1. We also perform baseline experiments for comparison using the previously proposed style descriptor [26] as an outfit appearance descriptor. Results show that our learned similarity models agree with human similarity judgments quite well. For the runway-to-runway task, learned similarity using our proposed features and framework achieves $73-76\%$ AUC compared to the style descriptor baseline of $62-66\%$, giving an increase in performance of about 10% in these experiments. We see similar improvements for the runway-to-realway task ($53-55\%$ vs $42-45\%$).

5.1. Similarity for extrinsic tasks

So far we have evaluated how well we can predict human judgments of similarity on held out data – an intrinsic evaluation of the learning process. The usefulness of the learned similarity can also be measured through extrinsic evaluations, where the similarity is used for other prediction tasks.

				10-nearest neighbor		Classifiers	
	Random	Most common	Human	Style descriptor	Our feature	Style descriptor	Our feature
Years	0.067	0.151	0.240	0.234	0.258	0.244	0.278
Seasons	0.333	0.478	0.520	0.534	0.572	0.554	0.578
Brands	0.003	0.012	-	0.106	0.122	0.098	0.129

Table 2: Extrinsic Evaluation: AUC for season, year and brand prediction for humans compared to our proposed features and the global style descriptor [26]. *K-nearest neighbor* indicates using our learned similarity models and k-NN voting for prediction while *Classifiers* indicates learning classifiers for season, year, and brand directly. Note that k-NN voting using learned similarity is on-par with classifiers trained directly for season, year, and brand directly even though k-NN was not trained for these tasks.

(a) Neon Retrieval (b) Floral Retrieval

Figure 5: Example retrieval results for neon (left) and floral (right) trends. Query outfits from the runway are shown in red with Chictopia streetstyle images retrieved using the learned similarity.

5.1.1 Season, year and brand prediction

Here we use our similarity to retrieve outfits for k-nearest neighbor classification of season, year, and brand of a query image. We attempt this because we believe that clothing from the same season, year, or brand should share aspects of visual appearance that could be captured by our similarity learning framework.

Given a query image, we retrieve similar images according to the learned similarity and output the majority classification label. The predicted season, year, or brand is thereby estimated as the majority vote of the top k retrieved runway images from other collections, excluding images from the same collection as the query image. Since the number of possible years and brands is large, sometimes there is no candidate with majority vote. In that case, we randomly predict year or brand from the candidate pool. To evaluate the difficulty of these prediction tasks, we also ask labelers on MTurk to perform the same predictions. For season prediction, the definitions and 5 example images of each season are shown, and labelers on MTurk are asked to identify the season for a query image. Unlike season prediction, we do not show example images for year prediction, but simply ask the labelers to choose one of the years.

In each prediction, 500 runway images are randomly selected as query images. Comparison of our automatic prediction vs human prediction is shown in Table 2. Results indicate that humans are better than random chance, and better than the "Most common" baseline that predicts the most commonly occurring label. However, we see that our k-nearest neighbor voting based prediction ($k = 10$) is bet-

(a) Query runway image, and 5 nearest realway outfits.

	Most similar	TFIDF	Probability
Style	Sexy	Chic	Chic
Trend	Lace dress	Calvin-Klein	Vintage
Item	Dress	Dress	Dress
Color	-	Black	Black

Figure 6: The transferred labels from 10-nearest neighbors of query image in Figure 6a

ter than both *humans* and the baseline style descriptor [26] for all 3 of these challenging tasks.

To further evaluate the effectiveness of our learned similarity models we also train classifiers to directly predict the year, season, and brand of outfits using our features or the style descriptor [26]. Results indicate (Table 2) that even though these k-NN based predictions for season, year, and brand were trained to predict outfit similarity (rather than for predicting these labels directly), performance is on-par with classifiers directly trained to predict these labels.

5.1.2 Label Transfer

Since the realway images are collected from a social network where users upload photos of their outfits and provide meta-information in the form of textual tags we can experiment with methods for predicting textual labels for runway outfits using label transfer. For a query image, we retrieve the k-nearest realway images according to our learned similarity and explore 3 different solutions for predicting style, trend, clothing item, and color tags.

Most similar image transfers labels directly from the most similar realway image to the runway query image. *Highest probability* retrieves the 10 nearest realway outfits, and selects candidate tags according to their frequency within the set of tags contained in the retrieval set. *TFIDF weighting* also retrieves the 10 nearest realway neighbors, but selects candidate tags weighted according to term frequency (tf) and inverse document frequency (idf). Fig 6 shows an example query image, 5 retrieved realway outfits, and tags predicted by each method.

To evaluate the predicted labels, we randomly select 100 query images and ask 5 labelers on MTurk to verify whether each label is relevant to the query image or not. To simplify the task for the style label, we also provide the definition of each style and some example images to the labeler. Labels that receive majority agreement, greater than or equal to 3 agreements, are counted as relevant. The average accuracies of the label transfer task are shown in Table 3.

	Most similar	TFIDF	Probability
Style	0.52	0.68	0.65
Trend	0.07	0.30	0.11
Item	0.75	0.80	0.95
Color	0.72	0.75	0.76

Table 3: The average accuracy of label transfer from 3 different approaches.

6. Influence of runways on street fashion

In addition to the intrinsic and extrinsic evaluations above, we also provide preliminary experiments examining how runway styles influence street style dressing. In particular, we select images from the runway collections that illustrate three potential visual trends: "floral" print, "pastel" colors and "neon" colors. To study these trends we manually select 110 example images for each trend from the *Runway dataset*. Using each of these runway images as a query, all street fashion images from the Paper Doll Dataset are retrieved with similarity above a fixed threshold, the retrieval results are shown in Fig 5. The percentage of similar images (normalized for increasing dataset size) for each trend is plotted in Figs. 7a - 7c. By studying the density of retrieved images over time, we can see temporal trends at the resolution of months in the street fashion. The seasonal variation is clear for all three styles, but neon and pastel show a clear increasing trend over time, unlike the floral style. Moreover, we also observe that even if the similarity threshold is varied, the trend pattern remains the same, as shown in Fig. 7c where we plot the density of neon images over time using different thresholds.

7. Conclusion

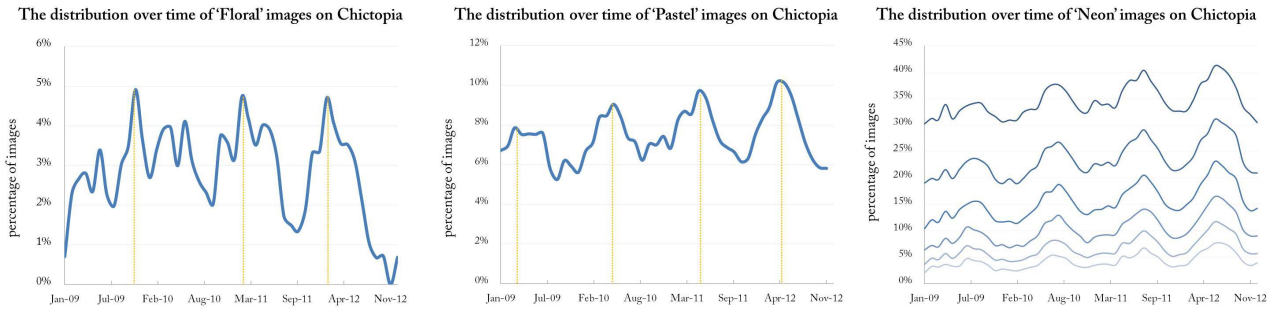
This paper presents a new approach for learning human judgments of outfit similarity. The approach is extensively analyzed using both intrinsic and extrinsic evaluation tasks. We find that the learned similarities match well with human judgments of clothing style similarity and find preliminary indications that the learned similarity could be useful for identifying and analyzing visual trends in fashion. Future work includes using learned similarity measures to mine large datasets for similar styles, trends, and studying further how street fashion trends are influenced by runway styles.

8. Acknowledgements

This work was funded by Google Award "Seeing Social: Exploiting Computer Vision in Online Communities" and NSF Award #1444234.

References

[1] T. Berg and P. N. Belhumeur. Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and



(a) Floral images (b) Pastel images (c) Neon images with different thresholds

Figure 7: Street fashion trends for 'Floral', 'Pastel' and 'Neon' in the Chictopia dataset from 2009-2012. Plot shows the density of images similar to example images for the trends. Number of images is expressed as a fraction of all images posted for that month. See Sec. 6

attribute estimation. In *CVPR*, pages 955–962, 2013. 2

[2] T. L. Berg, A. C. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *ECCV*, pages 663–676. Springer-Verlag, 2010. 2

[3] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel classification with style. *ACCV*, pages 1–14, 2012. 2

[4] L. Bourdev, S. Maji, and J. Malik. Describing people: A poselet-based approach to attribute classification. In *ICCV*, pages 1543–1550, 2011. 2

[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2:27:1–27:27, 2011. 5

[6] H. Chen, A. Gallagher, and B. Girod. Describing clothing by semantic attributes. In *ECCV*, pages 609–623. Springer-Verlag, 2012. 2

[7] C. T. Corcoran. The blogs that took over the tents. In *Women's Wear Daily*, Feb. 2006. 1

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 4

[9] P. Dollar and C. L. Zitnick. Structured forests for fast edge detection. In *ICCV*, 2013. 4

[10] R. Farrell, O. Oza, N. Zhang, V. I. Morariu, T. Darrell, and L. S. Davis. Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance. In *ICCV*, 2011. 2

[11] E. Gavves, B. Fernando, C. Snoek, A. Smeulders, and T. Tuytelaars. Fine-grained categorization by alignments. *ICCV*, pages 1–8, 2013. 2

[12] Y. Kalantidis, L. Kennedy, and L.-J. Li. Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In *ICMR*, pages 105–112. ACM, 2013. 2

[13] G. Kim and E. P. Xing. Discovering pictorial brand associations from large-scale online image data. *ICCV Workshops*, 2013. 2

[14] A. Kovashka, D. Parikh, and K. Grauman. Whittlesearch: Image search with relative attribute feedback. In *CVPR*, pages 2973–2980. IEEE, 2012. 2

[15] I. S. Kwak, A. C. Murillo, P. N. Belhumeur, D. Kriegman, and S. Belongie. From bikers to surfers: Visual recognition of urban tribes. In *BMVC*, 2013. 2

[16] Y. J. Lee, A. A. Efros, and M. Hebert. Style-aware mid-level representation for discovering visual connections in space and time. In *ICCV*, 2013. 2

[17] S. Liu, J. Feng, Z. Song, T. Zhang, H. Lu, C. Xu, and S. Yan. Hi, magic closet, tell me what to wear! In *ACM MM*, pages 619–628. ACM, 2012. 2

[18] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *CVPR*, pages 3330–3337, 2012. 2

[19] F. Palermo, J. Hays, and A. A. Efros. Dating historical color images. In *ECCV*, pages 499–512. Springer, 2012. 2

[20] D. Parikh and K. Grauman. Relative attributes. In *ICCV*, pages 503–510. IEEE, 2011. 2

[21] M. Shao, L. Li, and Y. Fu. What do you do? occupation recognition in a photo via social context. *ICCV*, 2013. 2

[22] Z. Song, M. Wang, X.-s. Hua, and S. Yan. Predicting occupation via human clothing and contexts. In *ICCV*, pages 1084–1091, 2011. 2

[23] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *Int. J. Comput. Vision*, 62(1-2):61–81, Apr. 2005. 4

[24] N. Wang and H. Ai. Who blocks who: Simultaneous clothing segmentation for grouping images. In *ICCV*, pages 1535–1542, 2011. 2

[25] K. Yamaguchi, M. H. Kiapour, and T. L. Berg. Parsing clothing in fashion photographs. In *CVPR*, pages 3570–3577, 2012. 2

[26] K. Yamaguchi, M. H. Kiapour, and T. L. Berg. Paper doll parsing: Retrieving similar styles to parse clothing items. In *ICCV*, 2013. 1, 2, 3, 4, 5, 6, 7

[27] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *CVPR*, pages 1385–1392, 2011. 1, 2, 3

[28] B. Yao, G. Bradski, and L. Fei-Fei. A codebook-free and annotation-free approach for fine-grained image categorization. In *CVPR*, pages 3466–3473. IEEE, 2012. 2