Fashion World Map: Understanding Cities Through Streetwear Fashion

Conference Paper · October 2017
DOI: 10.1145/3123266.3123268

4 authors, including:

Bo Wu
MIT-IBM Watson AI Lab

Kai-Lung Hua
National Taiwan University of Science and Technology

Fashion World Map: Understanding Cities Through Streetwear Fashion

Yu-Ting Chang
Research Center for Information Technology Innovation, Academia Sinica, Taiwan
juliachang@citi.sinica.edu.tw

Wen-Haung Cheng
Research Center for Information Technology Innovation, Academia Sinica, Taiwan
whcheng@citi.sinica.edu.tw

Bo Wu
Institute of Computing Technology, Chinese Academy of Sciences, China
wubo@ict.ac.cn

Kai-Lung Hua
CSIE, National Taiwan University of Science and Technology, Taiwan
hua@mail.ntust.edu.tw

ABSTRACT

Fashion is an integral part of life. Streets as a social center for people's interaction become the most important public stage to showcase the fashion culture of a metropolitan area. In this paper, therefore, we propose a novel framework based on deep neural networks (DNN) for depicting the street fashion of a city by automatically discovering fashion items (e.g., jackets) in a particular look that are most *iconic* for the city, directly from a large collection of geo-tagged street fashion photos. To obtain a reasonable collection of iconic items, our task is formulated as the prize-collecting Steiner tree (PCST) problem, whereby a visually intuitive summary of the world's iconic street fashion can be created. To the best of our knowledge, this is the first work devoted to investigate the world's fashion landscape in modern times through the visual analytics of big social data. It shows how the visual impression of local fashion cultures across the world can be depicted, modeled, analyzed, compared, and exploited. In the experiments, our approach achieves the best performance (43.19%) on our large collected GS-Fashion dataset (170K photos), with an average of two times higher than all the other algorithms (FII: 20.13%, AP: 18.76%, DC: 17.90%), in terms of the users' agreement ratio on the discovered iconic fashion items of a city. The potential of our proposed framework for advanced sociological understanding is also demonstrated via practical applications.

CCS CONCEPTS

• **Social and professional topics → Cultural characteristics**; • **Computing methodologies → Visual content-based indexing and retrieval**; • **Information systems → Top-k retrieval in databases**;

KEYWORDS

Social media; city profiling; street fashion; visual big data analysis

MM'17, October 23–27, 2017, Mountain View, CA, USA.

DOI: https://doi.org/10.1145/3123266.3123268

Figure 1: Street fashion varies throughout place. From Los Angeles to Cape Town, each city has a distinct style of streetwear that reflects the city's lifestyle (The pictures are searched from the web).

1 INTRODUCTION

Streets have been the social centers of towns and cities where people come together, congregate and interact with one another. According to a recent survey by UN-Habitat [8], streets typically represent the largest area of public space a city has, i.e., making up 30 to 35 percent of a city's land area. Streets are thus known as the 'third place' between home and workplace in a city for people's social gatherings and become the most important public stage to showcase the city's best culture and lifestyle [9].

Like food, landmarks and language, fashion is an integral part of a city's culture and has now become not only an art of self expression but a major influence in business, financial, entertainment,

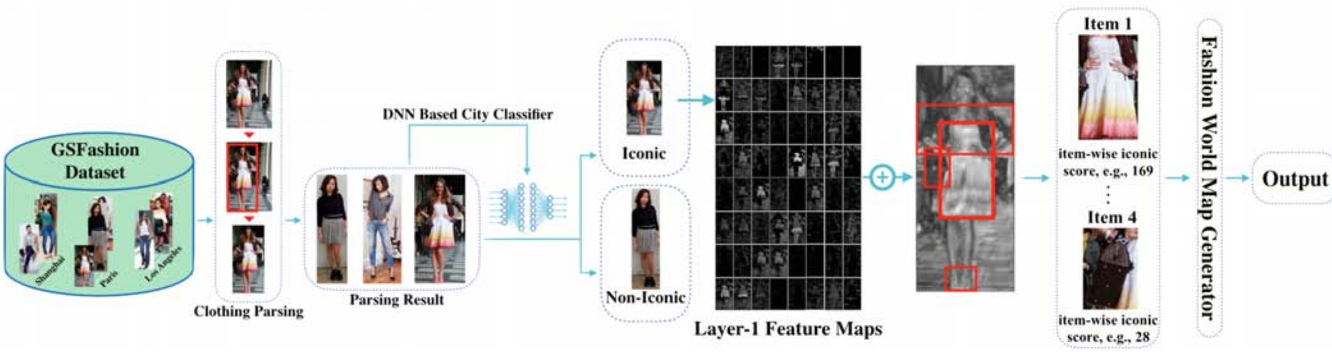

Figure 2: An overview of the proposed framework.

cultural and leisure activities [7, 14]. For example, the global fashion apparel market today has surpassed 1 trillion US dollars since 2013, and accounts for nearly 2 percent of the world's Gross Domestic Product (GDP) [35]. Thus, understanding the fashion life of cities is sociologically significant and can offer new insights into culture-aware community profiling, analysis, and innovation.

As pointed out by Bill Cunningham[1], one of the world's great fashion photographers, streets are the runway of ordinary people. Street fashion, or streetwear fashion, is what everyday people on the street are wearing and is a way of distinguishing one city from another [13]. Figure 1 shows some street fashion photos from six world-famous cities. For example, Milan has a much different fashion scene than Berlin. The Milan style uses vivid patterns and flamboyant colors, while the Berlin style favors dark colors and a more grungy look. When studying the fashion culture of cities, traditional approaches have often relied on free-text comments of fashion experts [14]. A known problem with descriptive comments is that they tend to be described as generic and vague [28], e.g., "the typical Parisian style is all about effortless elegance and character" by a French fashion designer, Julie de Libran. In fact, as suggested by cognitive findings [1], a natural way to explore a concept, especially one of visual nature like fashion, is to learn from seeing [27]. This motivates our research questions: How do we make intuitive fashion profiles of cities? How can the fashion profiles be applied for inferring advanced social knowledge as well as enabling creative applications?

In this paper, therefore, we propose a novel framework (cf. Figure 2) for depicting the street fashion of a place by automatically identifying a number of fashion items in a particular look (e.g., women's mini skirts with black and red stripes) that are most *iconic* for the place, directly from a large database of geo-tagged street fashion photos. Note that "iconic" here does not mean necessarily unique to a place but may be shared with other places, e.g., because of cultural interaction [13]. We first construct a large, complex, and real-world collection of street fashion photos from an image-centered social media site, Lookbook.nu[2], resulting in a total number of over 400K user-contributed photos from 45 cities worldwide. After clothing segmentation [41], there are 170K photos with full-body subjects, and we devise a metric based on deep neural networks (DNN) [31], namely the DNN confidence, to select a number of potentially iconic outfits for each city. For every outfits, we further calculate iconic scores in a finer level based on fashion items (e.g., 'jacket', 'socks', 'hat', etc.). Next, we formulate the detection problem of iconic fashion items of a city as the prize-collecting Steiner tree (PCST) problem [36], whereby the proposed fashion world map can be created. Finally, we demonstrate that how our framework can be applied to infer rich sociological insights for enabling creative applications.

To the best of our knowledge, this is the first work devoted to investigate the world's fashion landscape in modern times through the visual analytics of big social data. It shows how the visual impression of local fashion cultures across the world can be depicted, modeled, analyzed, compared, and exploited. The proposed framework is believed to be generic and can be a demonstrative example to advanced understanding of other social aspects in life, especially for those of visual nature.

In the rest of this paper, section 2 provides the literature review. Section 3 presents the collected dataset. Sections 4 and 5 describe the proposed methods for iconic outfit selection and fashion map generation, respectively. Section 6 gives the experiments and Section 7 concludes this work and discusses future research directions.

2 RELATED WORK

Fashion item extraction is a fundamental task in visual analytics for fashion images, whereby the results can be further exploited for exploring social insights in different societal levels from an individual to a group of individuals. Therefore, we divide the related work into the following three lines of aspects, i.e. fashion item extraction, individual-level sociological understanding, and group-level sociological understanding.

Fashion Item Extraction. Extensive previous research has been focused on object based clothing image analysis, such as clothing modeling, recognition, parsing, retrieval, and recommendations [19]. Yamaguchi *et al.* [41] presented a hybrid method by combining different models to perform pixel-level clothing parsing. Liu *et al.* [20] presented fashion landmark detection. Wang *et al.* [38] developed a personalized garment recommender system based on fuzzy logics. Jia *et al.* [16] explored the relationship between visual features and aesthetic words of clothing. Chen and Luo [5] discovered fine-grained clothing attributes as the representative and discriminative characteristics of popular clothing style

[1]https://en.wikipedia.org/wiki/Bill_Cunningham_(American_photographer)
[2]http://lookbook.nu

(a) (b)

Figure 3: The GSFashion (Global Street Fashion) image dataset constructed in this work, with a total number of 400,822 local street fashion pictures from 45 global major cities: (a) a sample image with the recorded information and (b) the selected cities covering the five continents of the world.

elements. Veit *et al.* [34] learned a visual compatibility across different categories of fashion items to identify item pairs of similar style. Kiapour *et al.* [17] created an application to match a real-world example of a garment item to the same item in an online shop. Simo-Serra and Ishikawa [30] extracted discriminative features of clothing images by joint ranking and classification. Morris *et al.* [24] discussed the crowdsourcing scheme for generating remote shopping advice based on social networking sites.

Individual-Level Sociological Understanding. User profiling helps personalization and has received much attention in the social multimedia research fields. Nguyen *et al.* [26] estimated female attractiveness by using audiovisual cues like dressing, face, and voice. Chen *et al.* [6] exploited convolutional neural networks (CNN) for describing people based on fine-grained clothing attributes. Simo-Serra *et al.* [29] analyzed how fashionable a person looks on a photo whereby advising the user to improve the appeal. Song *et al.* [32] predicted the occupation of a person via the clothing. He *et al.* [12] applied collaborative filtering to build the user's fashion-aware personalized ranking functions of fashion items.

Group-Level Sociological Understanding. A group is a collection of individuals and group-level analysis enables us to capture major trends in the social life of groups of people [15, 39, 40, 43]. Hidayati *et al.* [14], Chen *et al.* [4], Vittayakorn *et al.* [37] and Ma *et al.* [23] analyzed the visual evolution of clothing fashions as a reflection of the society of a period. Murillo *et al.* [25] determined the social subculture or urban tribe for people in group photos. Zhou *et al.* [44] described a city in terms of the spatial form (e.g., green space coverage) and the social functionality (e.g., athletic activity), by using high-level attributes extracted from the corresponding geo-tagged images. Similarly, Ge *et al.* [11] developed a vision-based approach for recognizing the high-level interest category of a city, e.g., sports and politics.

3 DATASET

To evaluate the effectiveness of our approach at capturing local fashion cultures on a global scale, we build a GSFashion (Global Street Fashion) image dataset by crawling a large number of street fashion photos from Lookbook.nu, a social networking website that allows people to post and share their street-style fashion looks. By 2017, Lookbook.nu has over 3 million unique visitors and over 75 million page views per month.

In the GSFashion dataset[3], we collected totally 400,822 local street fashion pictures over 45 global major cities, as summarized in Figure 3(b). For each image, we also recorded six types of the associated information: (1) city name, (2) anonymized user ID, (3) anonymized post ID, (4) posting date, (5) number of likes of the post, and (6) user-provided metadata describing the categories of fashion items included in the image (e.g., 'jeans') along with the brand names (e.g., 'Stella McCartney') and the brand-defined product categories (e.g., 'Stella McCartney Denim Jeans'), cf. Figure 3(a). Note that since our focus is on people who are really residents of a city but not tourists or visitors, we collected pictures for each city only from users who have the same registered location on Lookbook.nu to the current city.

In comparison to existing fashion related image benchmarks, such as Fashion 10000 [22] and Paper Doll [41], our GSFashion dataset is superior by also offering geo-social information for every images, e.g., locations, likes, and brand names of fashion items. For our analysis later in the experiments, we run a clothing parsing tool [41] for each image in the GSFashion dataset as a preprocessing step to exclude those images where the models do not have a full body pose or have too many occluded body parts. This results in 170,180 remaining pictures.

4 CITY BASED SELECTION OF ICONIC OUTFITS

On a city street, people wear all kinds of clothing but not every outfits fit the common dress styles of the local area. In this section, we propose to exploit deep neural networks (DNN) [31] as a selector to distinguish potentially iconic outfits in different cities.

People's outfits are a sort of visual presentation consisted of several basic components which are a variety of fashion items. As suggested by [41], there are 53 defined categories[4] of fashion items. For example, a man who manages to have a casual look outfit may wear T-shirt, shorts and sunglasses; a woman who attends a party may hold a purse and be attired in bodycon dress and high heels.

Let X_{C_i} represent the outfit choice of a person in the city C_i. To model X_{C_i}, we define X_{C_i} as a collection of N random variables $X^1_{C_i},...,X^N_{C_i}$ representing the adopted 53 categories of fashion items

[3] We will make the GSFashion dataset public after our paper was accepted.
[4] In the original taxonomy of [41], there are 56 defined categories but we discard the three ones unrelated to fashion items, i.e. 'background', 'skin', and 'hair'.

($N = 53$), so that the probability of a specific outfit $\mathbf{v} = v_1, ..., v_n$ is

$$P(X_{C_i} = \mathbf{v}|C_i) = P(X^1_{C_i} = v_1, ..., X^n_{C_i} = v_n|C_i), \quad (1)$$

where $X^j_{C_i}$ = **null** if no item of the j-th category is included in the outfit $\mathbf{v}$, e.g., a person does not wear a hat.

Since the iconic outfit of a city can be described as the most common dress style, our goal is then to select the most probable $\mathbf{v}$ as the iconic outfit by learning a model of $P(X_{C_i}|C_i)$ on the training image dataset:

$$\mathbf{v} = \underset{X^1_{C_i},...,X^N_{C_i}}{\arg\max} \; P(X_{C_i}|C_i). \quad (2)$$

Conventionally, a joint probability function will be factorized to simplify the analysis [33], as below:

$$P(X_{C_i}|C_i) = \Pi^n_{j=1} P(X^j_{C_i}|C_i). \quad (3)$$

However, $X^j_{C_i}$ are not independent variables and the factorization is not valid in our case. That is, a person's choice on part of an outfit usually depends on the other parts. For example, it's more often to choose suit pants rather than shorts to match a suit jacket for the outfit.

Therefore, instead of directly computing $P(X_{C_i}|C_i)$, we choose to learn a DNN based city classifier ϕ for all the segmented outfits from the GSFashion dataset (cf. Section 3). The detailed configurations of the DNN architecture are referred to Section 6. For each outfit $\mathbf{v}$ in the city C_i, we then take the prediction score of how likely it belongs to C_i from the softmax layer in the DNN, namely the DNN confidence, as the approximated value of $P(X_{C_i} = \mathbf{v}|C_i)$. The idea here is, if an object is more representative of its class, it would be more confident that we can classify the object. Rigorously speaking, the DNN confidence and $P(X_{C_i}|C_i)$ are not the same. In other words, we do not attempt to replace $P(X_{C_i}|C_i)$ by the DNN confidence. We are most interested in using the DNN confidence as an alternative and efficient metric for the selection of potentially iconic outfits. Finally, for each city, the top-k outfits with highest DNN confidences are returned as the results. In the experiments, we choose $k = 50$, cf. Section 6. Sample outfit images of the top-50 selection are given in Figure 4.

5 FASHION WORLD MAP GENERATION

After a number of potentially iconic outfits were extracted from a city, this section devotes to gain deeper insights into what compositional elements make the outfits to be iconic. To answer this question, we calculate iconic scores for all fashion items parsed from the extracted outfits [41], whereby a number of high-score items can be selected for generating a visual summary of the city based street fashion, known as the proposed fashion world map.

First, for an outfit image, we calculate iconic scores in pixel level by exploiting the DNN visualization technique [42]. Let the DNN classifier ϕ has p filters $o_1, ..., o_p$ in the first layer. When an outfit image is inputted into ϕ, p corresponding feature maps (activations), $q_1, ..., q_p$, can be obtained [31]. Given $\odot$ as the pixel-wise max operator, all feature maps are then combined into a unique scalar "iconic map", q^*, which encodes for the iconic level of a location in the outfit image:

$$q^* = q_1 \odot q_2 \odot ... \odot q_p. \quad (4)$$

Figure 4: Sample outfit images from the GSFashion dataset, where the left and the right groups are respectively images included and excluded in the top-50 selection. Images of the top-50 selection can be observed to give a stronger sense of clothing style than images those are not selected.

The idea here is similar to the use of DNN confidence in Section 4. That is, if a location is more iconic of the city, it would lead to a stronger activation of one or multiple of the filters. We choose to apply the pixel-wise max operator to the feature maps $q_1, ..., q_p$ altogether, because we want to examine the strongest activation associated with a location regardless of which filters this quantity is coming from. Thus, for each fashion item of an outfit, its iconic score can be assigned by the average of all pixel-wise iconic scores over the corresponding spatial extent in q^*.

Based on the item-wise iconic scores, a naive solution to detect the iconic fashion items of a city is to rank the items and return the top-k ones. However, the returned items can be too different from each other, in terms of the item category or the visual style. This is not desired because the results do not really give a sense of the iconically mainstream fashion but simply a collection of diverse items. To address this challenge, we propose to formulate the detection problem of iconic items as the prize-collecting Steiner tree (PCST) problem [36]. Let $G = (V, E)$ denote the *item graph*, where V is a set of nodes and E are the edges. According to the definition of PCST problem [36], given a connected undirected graph G with positive real node weights, $\kappa : V \rightarrow \Re^{\geq 0}$, and positive real edge weights, $\gamma : E \rightarrow \Re^{\geq 0}$, the objective is to find a connected subgraph $R = (V_R \in V, E_R \in E)$ of G whose total profit scores Φ are maximal:

$$\Phi(R) = \max\{\sum_{v \in V_R} \kappa(v) - \sum_{e \in E_R} \gamma(e)\}. \quad (5)$$

Since the PCST problem can be formulated as an integer linear program (ILP), it can be efficiently solved with a branch-and-cut algorithm [21].

Specifically, for each city, we construct an item graph from all the parsed fashion items of the top-50 outfit images, cf. Section 4. Each item is represented by a graph node, and a prize (i.e. the corresponding item-wise iconic score) is associated with the node. An edge is placed between any two nodes and the edge weight is determined by the style descriptor [41]. Therefore, the PCST solution aims to determine which nodes (items) as a collection will give a compact subgraph that gives the best "prize". That is, we hope that the selected items not only have high iconic scores but also similar to each other in style. The collection can then be taken as our resultant iconic fashion items of the corresponding city. The proposed fashion world map can be created, cf. Figure 7.

6 EXPERIMENTS

In this section, we first give the experimental settings in Section 6.1. We then present the experimental results with discussions in Section 6.2. Next, sociological insights from our study are investigated in Section 6.3. Finally, in Section 6.4, extended applications of our proposed framework is demonstrated. For explanation purposes in this section, let $i50$ denote the image collection of top-50 selection over all cities in the GSFashion dataset (c.f. Section 4).

6.1 Experimental Settings

6.1.1 DNN Configurations. We implemented the VGG19 model [31] as the main architecture of our DNN classifier in Section 4, and trained the model on the GSFashion dataset. Briefly, the model has 16 convolutional layers (2 with 64 filters, 2 with 128 filters, 4 with 256 filters, and 8 with 512 filters) and 2 fully connected layers. Since the original VGG19 model aims at working on a larger image collection (i.e. ImageNet, a benchmark of over 14 million images in 1,000 classes), we slightly modified the VGG19 model by reducing the width of the fully connected layers from 4,096 to 1,024 to make the model more fit to the scale of our dataset. Consequently, the loss and the accuracy are 0.3723 and 89.68%, respectively.

6.1.2 Performance Comparison with Different Methods. To validate the effectiveness of our approach, we adopted the following methods for comparison. In these methods, we adopted the clothing features from [41].

Finding Iconic Image (FII) [2]: FII method is used to automatically find representative images of a specified object category. This method is based on k-means and the cluster centers are selected as the iconic examples.

Affinity Propagation (AP) [10]: AP method can cluster the input set automatically and find the members that are representative of the cluster. This approach does not require the number of clusters to be determined or estimated before running the algorithm.

Discriminative Clustering (DC) [9]: DC method can find visual elements that are most distinctive for a certain geo-spatial area. A discriminative clustering approach is used to achieve geographic supervision. The iconic pattern of a city is assumed to satisfy both frequently occurring and geographically discriminative. That is, the characteristic is supposed to appear repeatedly in that locale and seldomly appear elsewhere.

6.1.3 User Study Evaluation Protocol. Since fashion is subjective, we conducted the user study to evaluate our results. After finding the iconic items of each city, we designed a questionnaire and randomly invited 70 people from our campus as the participants (29 males and 41 females, with an average age of 27.6). The whole study process was divided into two phases: pre-study phase and on-study phase.

Pre-study Phase: The pre-study phase is to give the participants some general ideas of the clothing style of each city. First, we collect a number of sample images using the Google Image Search (images.google.com). For each city, we used the city name plus two words "street life" to issue a search query and returned the first 50 images containing people as the image samples of the city. Then a participant is allowed to freely explore the images with no time limit. Note that we do not give any text and oral explanations. After that, the participants can start to fill out the questionnaire.

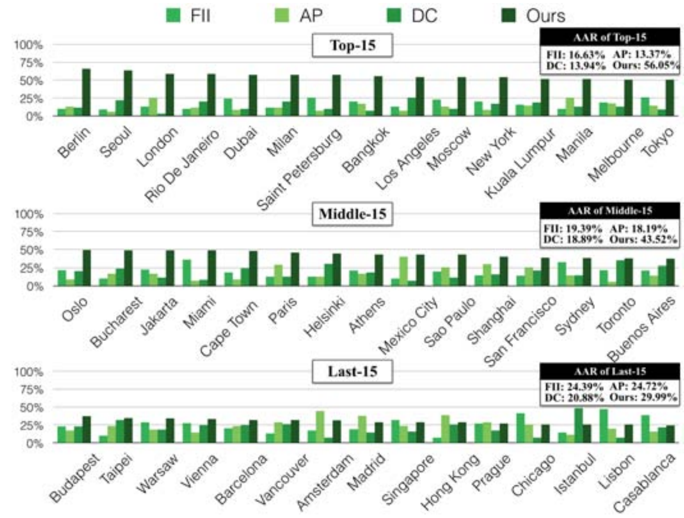

Figure 5: The statistical results of the user study, evaluated by the AR performance. (See Section 6.2.1 for details)

On-study Phase: We focus on 45 cities in this work, so there are altogether 45 questions in the questionnaire and each question is associated with one city. In each question, the city name is prompted and there are four visual options, with each corresponding to the extracted iconic item generated by one of the four comparing methods. The generating method of each option is blind to the participants and the display order of the four options is randomly and dynamically determined. The participants need to choose one option that they think it is most likely the clothing style of the city.

6.1.4 Evaluation Metric: Agreement Ratio. The agreement ratio (AR) is adopted as the evaluaton metric to assess the performance of each method. Suppose there are m participants in the user study, for a method on a city, there are n participants choosing the method. The AR performance of this method on the city can thus be defined as n/m. Given k cities in total, we can derive the average agreement ratio (AAR) of the method as

$$AAR = \frac{1}{k} \sum_{i=1}^{k} AR_i. \tag{6}$$

6.2 Results

6.2.1 AR Performance. From the user study, the AAR of our approach is 43.19% and the number is two times higher than other methods (i.e. FII: 20.13%, AP: 18.76%, DC: 17.90%). FII and AP are methods based on unsupervised clustering, using visual properties to find the representative image. In contrast, DC applies geo-supervision to discover the discriminative object, where the iconic pattern is assumed to be occur much often in the particular city than other cities. However, the AAR performances of these comparing methods are all at a relatively low level, around only 20%.

To get a deeper insight into the results, we sorted the 45 cities into three groups (Top-15, Middle-15, Last-15) by the AR performance

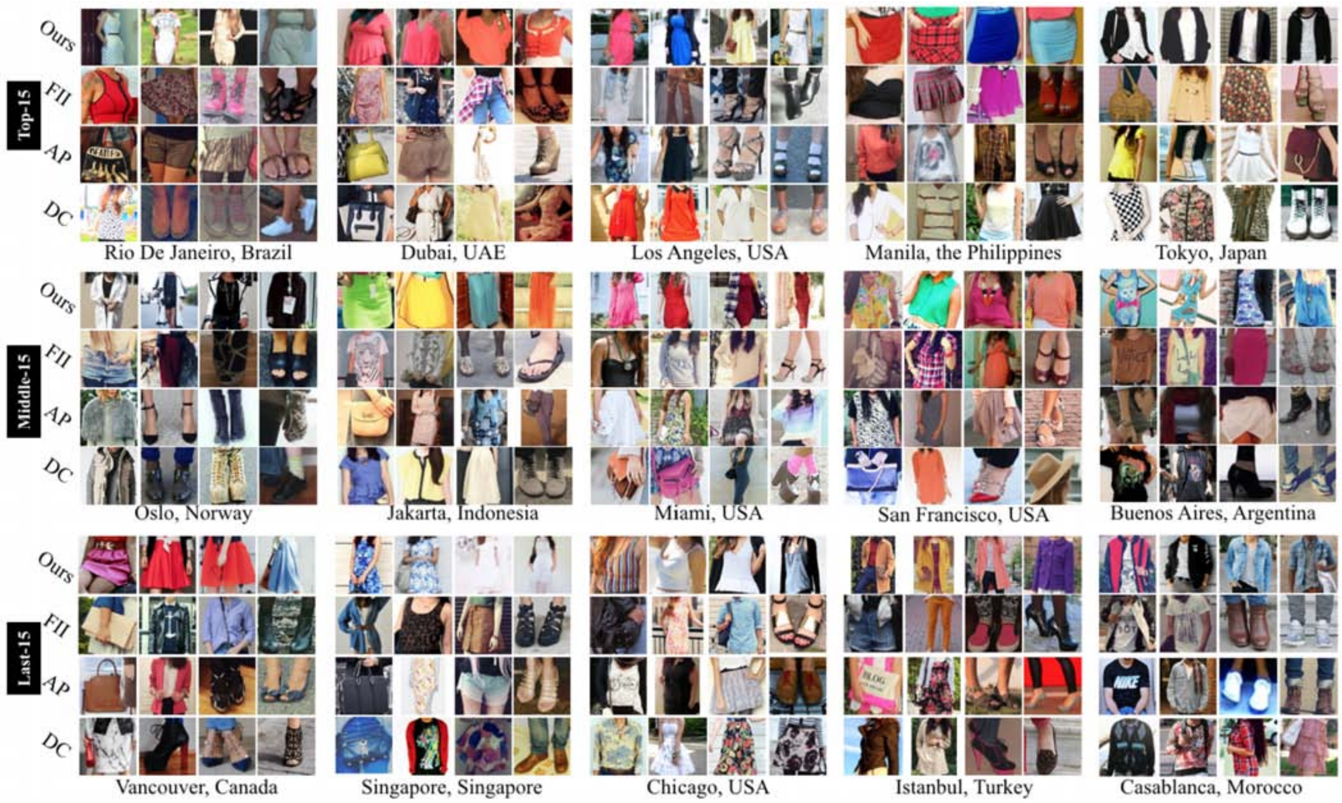

Figure 6: A visual comparison of the iconic items found by four different methods (i.e. Ours, FII [2], AP [10], DC [9]). There are 15 cities randomly selected from Top-15, Middle-15 and Last-15, respectively.

of our approach, as shown in Figure 5. We can find that the Top-15 cities are those with a more distinguishable clothing style like Berlin, Seoul and Dubai, or they are conventionally known as fashionable city like London and Milan. The Top-15 AAR of our method is 56.05%, even higher than our overall AAR (43.19%). It suggests that, with resepect to the other comparing methods, our approach is more effective to discover the fashion elements of the locale. The Middle-15 AAR of our method is 43.52%. We are the majority choice of the participants in the user study. The Last-15 AAR of our method is still the highest among all methods, although decreasing to 29.99%. Based on our observations, the Last-15 cities are mostly multicultural city, such as Hong Kong, Istanbul and Casablanca, with a mix of multiple cultures. Another reason leading to the lower performance might be the lack of clothing characteristics, i.e., the local clothing style is not explicit enough. This reason might cause the outfit style can not be easily distinguished by the user study participants. Overall, the results of our approach are encouraging.

6.2.2 Visual Comparison. The iconic items discovered by different methods for some cities are shown in Figure 6. Generally, the iconic items found by our approach have a more consistent style, e.g., they might have similar colors or they might be the same clothing item such as dress, skirt or coat. In contrast, the discovered iconic items by the other methods are very visually diverse and lacking of consistency. Considering the fact that our approach achieves the best AAR performance, the iconic items discovered by our approach are believed to be an effective reflection of the leading fashion of a city-scale human community.

6.2.3 Fashion World Map. The fashion world map generated by our approach is shown in Figure 7. A number of interesting observations can be made from our fashion world map. Firstly, each of the cities has a different iconic dress style that reflects the local culture and lifestyle [13]. For example, Miami fashion and Rio De Janeiro fashion are similarly form-fitting silhouettes that flatter the body's curves and the lace fabric is usually used, where the both places are largely affected by the Latin American culture. Secondly, climate is another important but implicit effect on the street fashion. The city of Oslo is in high-latitude areas and its fashion is black or white coats which can help to keep out the cold. The plain colors also convey the Scandinavian Minimalist style. Different from Oslo's warm clothing, Tokyo's fashion is much thiner black-and-white blazers and shirts. This style is a type of business attire and would reflect the disciplinary nature of Japanese culture. Besides, religion also plays a role in society. For example, Islam is the state religion of Malaysia and the iconic items of Kuala Lumpur are unsurprisingly found to be the islamic headscarf. In summary, beyond the conventional way of literally naming, the proposed fashion world map is a novel and intuitive tool for empowering the 'seeing is believing' capability of sociological understanding.

Figure 7: A visual overview of the world's iconic street fashion, with 15 selected cities as example.

6.3 Observations and Insights

6.3.1 Finding 1: Correlation of Clothing Color and City Latitude. To develop more insights into sociological understanding of street fashion, we arranged the 45 cities in order of latitude, with each is represented by three dominant colors extracted from all the corresponding outfit images in *i*50. The results are shown in Figure 8. It's revealed that the clothing color and the city latitude are correlated. In higher latitude areas, the main colors of clothing are cool colors and the earth tone[5]. Relatively, in equatorial regions, the clothing color is dominated by bright colors like red, orange, and yellow. In fact, the color conveys meanings in two primary ways – natural associations and psychological (cultural) symbolism.

First, it is known that climate is an important factor affecting the color of the clothing. According to psychological evidence [3], people who live in climates with abundant sunlight prefer warm bright colors, while those from climates with less sunlight prefer cooler and less saturated colors. Our results are in accordance with this fact. Second, color may generate another level of meaning in the mind and this symbolism arises from cultural contexts. For instance, red is the color of joy and celebration in the East. In China, it is often the color worn by brides, and in India, red is the color for purity. We can find that red tone colors are widely distributed in the areas where affected by Eastern culture. Also, in Asian cultures, yellow is considered sacred and imperial, and it is more widely connected to happiness and prosperity in the Middle East. On the contrary, in Latin America, yellow is associated with death and mourning. According to our results, yellow is not common in the cities of Latin America. Globally speaking, the color blue is considered to be the safest and most positive of all the colors. Truly, as shown in Figure 8, blue seems to distribute in each different culture evenly.

6.3.2 Finding 2: Influence of International Clothing Brands on Local Fashion. We utilize the brand names of clothing items in the GSFashion dataset (cf. Section 3) to investigate the effect of globalization on the street fashion across different cities. We focus on four particular brands, i.e., H&M, ZARA, Forever 21 and Topshop,

[5]https://en.wikipedia.org/wiki/Earth_tone

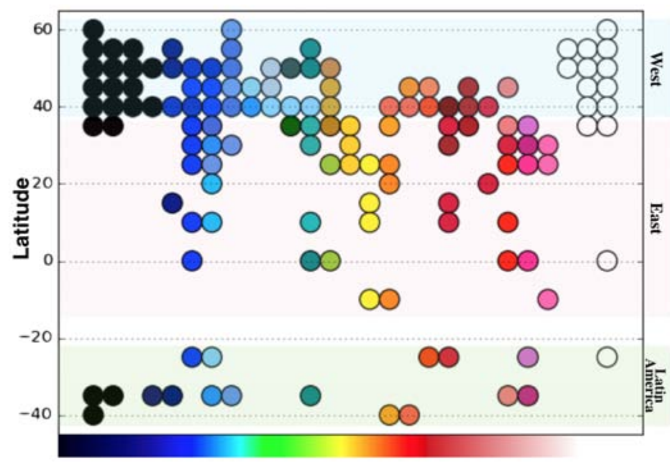

Figure 8: Clothing color versus city latitude. Three dominant colors are extracted from each of the 45 cities in the GSFashion dataset and arranged on the latitude line of the corresponding city. For example, the city of Helsinki is the only one at $+60^o$ latitude and its three dominant colors are black, blue, and white. Note that if one color is shared by two cities, the color circle will be placed twice. In our case, most Eastern cities (i.e. Asia, Southeast Asia and Middle East) are located at between -15^o and $+35^o$ latitude. Western cities (i.e. North America and Europe) are above $+35^o$ latitude and the cities of Latin America lie between -20^o and -40^o latitude.

because our statistics indicate that they are the top-4 biggest brands in *i*50. H&M is from Sweden, ZARA is from Spain, Forever 21 is from USA, and Topshop is from UK. H&M and ZARA are also the two largest apparel retailers in the world[6,7].

For explanation purposes, we define a special format *N:I(R)*, e.g., "ZARA:1(2)", to show the popularity of a brand in a city, where *N* is the brand name, *I* represents the number of occurrence times of the

[6]https://en.wikipedia.org/wiki/Zara_(retailer)
[7]https://en.wikipedia.org/wiki/H&M

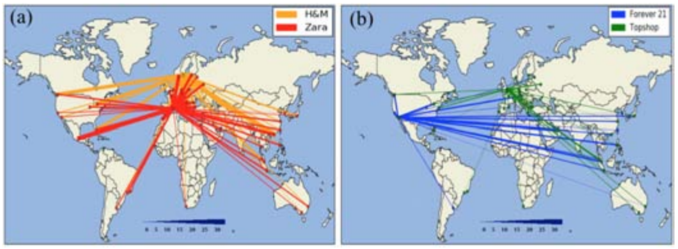

Figure 9: Brand influence in the fashion world. (See Section 6.3.2 for details)

brand's items in the city, and R stands for the rank of the number I from all cities in the same continent. In the world map, for each brand, we connect the brand (represented by its place of origin) to every cities with an edge and the edge width is proportional to the I value of the city connected. Our results are illustrated in Figure 9.

As shown in Figure 9(a), H&M and ZARA distribute extensively over the world, and the average I of them are 10.38 and 10.36, respectively. If the average is calculated by continent, the largest values of the two brands are both in Europe, i.e. 15.17 and 14.00, respectively. On the other hand, in Figure 9(b), the worldwide distribution of Forever 21 and Topshop are not as widespread as like H&M and ZARA. The average I are only 3.01 and 2.33, respectively. Similarly, if the average is calculated by continent, the largest value of Forever 21 is in North America (8.13), while Topshop's is 2.89 in Europe. This fact might imply that these two brands developed more regional and close to the original country, given that Forever 21 is an American brand and Topshop is a British one.

We also observe that, not only the origin of the company will affect the distribution of the brand, but the cultural contexts will do so. For example, Hong Kong has the largest I in Asia for both H&M and ZARA, i.e., H&M:27(1), ZARA:15(1). This may because Hong Kong is the world's largest re-export center and frequently described as a place where "East meets West", reflecting the mix culture of Chinese roots and British colony. Therefore, we can discover the cultural diversity of Hong Kong through the distribution of clothing brands. Another case is Istanbul. Our results show that all European brands here have a relatively larger I than other Asian cities (i.e., H&M:16(2), ZARA:15(1), Topshop:7(1)). Although Istanbul straddles both Europe and Asia, we can infer that the dress style of Istanbul deeply comes under the influence of European culture.

Moreover, we can observe the rise of fast fashion[8] through the statistics. The four brands we focus on are all prominent fast fashion brands in the world. Fast fashion can be defined as low cost clothing collections that mimic current fashion trends. According to our statistics, luxury fashion brands are not common and the garments of fast fashion brands are much more popular. This gives strong support to the fact that street fashion is more representative of a human society than catwalk fashion.

6.4 Applications

In this section, we demonstrate how our research output can be exploited to benefit a wide range of practical applications. For example, user modeling is a powerful technique in business intelligence to achieve market segmentation which can break the market down into smaller, more specific blocks of customers with unique needs. To serve as a complement to traditional visual analytics that utilize intrinsic user attributes such as gender, age or face recognition for analysis purposes [18], we can provide semantic information about a user via extending our framework from city-level style extraction into personal style preference prediction. In particular, we can use the iconic outfit model learnt in Section 4 to estimate a person's wearing style on a city basis. For example, given a street photo taken in New York as shown in Figure 10, we detect five people and make the prediction of them. As a result, it is demonstrated that our proposed framework can help to profile the fashion taste of the customers and offer smart insights embedded within the picture.

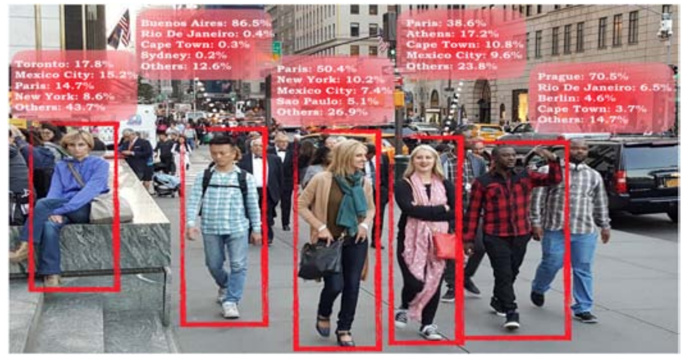

Figure 10: A demonstration of our proposed framework for socio-demographic analysis of wearing style on the street. For each person, we display the four highest possibilities.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a novel framework based on deep neural networks (DNN) for understanding big social data from the fashion perspective and demonstrated its potential for enabling advanced applications. To the best of our knowledge, this is the first work devoted to investigate the world's fashion landscape in modern times through the visual analytics of big social data. It shows how the visual impression of local fashion cultures across the world can be depicted, modeled, analyzed, compared, and exploited.

Several research topics are for future investigation. The first and most obvious limitation is the lack of objective evaluations. Fashion is an inherently subjective and cultural notion and it is essentially difficult to be defined by quantitative measures. Advanced verification techniques, e.g., cultural mapping[9], will be investigated. Currently, our analysis focuses on the clothing outfit, the incorporation of other contextual factors, such as face, gender, age or skin color, might be useful to help improve the analysis quality. Besides, we can extend our approach as an IoT gadget to achieve personal profiling, e.g., being immersed into the daily lives of potential customers to probe personalized information, such as nationality, social status and preference. Creative applications like brand localization can be realized. People's values and behaviors are shaped by the unique culture in which they live. Brand can formulate a localization strategy by understanding the iconic style in a city, and modifying content to suit the tastes and consumption habits of it.

[8]https://en.wikipedia.org/wiki/Fast_fashion

[9]https://en.wikipedia.org/wiki/Cultural_mapping

REFERENCES

[1] 1913. One Look Is Worth A Thousand Words. Piqua Leader-Dispatch. (1913).

[2] Tamara L. Berg and Alexander C. Berg. 2009. Finding Iconic Images. In *CVPR*.

[3] Mario De Bortoli and Jesus Maroto. 2001. Colours across cultures: Translating colours in interactive marketing communications. In *ELICIT*.

[4] K. Chen, K. Chen, P. Cong, W. H. Hsu, and J. Luo. 2015. Who are the devils wearing prada in new york city?. In *ACM Multimedia*.

[5] Kuan-Ting Chen and Jiebo Luo. 2016. When Fashion Meets Big Data: Discriminative Mining of Best Selling Clothing Features. In *WWW*.

[6] Qiang Chen, Junshi Huang, Rogerio Feris, Lisa M Brown, Jian Dong, and Shuicheng Yan. 2015. Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes. In *CVPR*.

[7] Heng-Yu Chi, Chun-Chieh Chen, Wen-Huang Cheng, and Ming-Syan Chen. 2015. UbiShop: Commercial item recommendation using visual part-based object representation. *Multimedia Tools and Applications* doi:10.1007/s11042-015-2916-7 (2015).

[8] Joan Clos. 2013. *Streets as Public Spaces and Drivers of Urban Prosperity*. UN-HABITAT.

[9] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros. 2012. What makes Paris look like Paris?. In *ACM SIGGRAPH*.

[10] Brendan J. Frey and Delbert Dueck. 2007. Clustering by Passing Messages Between Data Points. *Science* 315, 1 (February 2007), 972–976.

[11] Hancheng Ge and James Caverlee. 2016. College Towns, Vacation Spots, and Tech Hubs: Using Geo-Social Media to Model and Compare Locations. In *AAAI*.

[12] Ruining He and Julian McAuley. 2016. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-class Collaborative Filtering. In *WWW*.

[13] Charles A. Heatwole. 2006. Culture: A Geographical Perspective. http://www.p12.nysed.gov/ciai/socst/grade3/geograph.html. (2006).

[14] Shintami Chusnul Hidayati, Kai-Lung Hua, Wen-Huang Cheng, and Shih-Wei Sun. 2014. What are the Fashion Trends in New York?. In *ACM Multimedia*.

[15] Tianran Hu, Eric Bigelow, and Jiebo Luo. 2016. Tales of Two Cities: Using Social Media to Understand Idiosyncratic Lifestyles in Distinctive Metropolitan Areas. *IEEE Transactions on Big Data* (2016).

[16] Jia Jia, Jie Huang, Guangyao Shen, Tao He, Zhiyuan Liu, Huanbo Luan, and Chao Yan. 2015. Learning to Appreciate the Aesthetic Effects of Clothing. In *AAAI*.

[17] M. Hadi Kiapour, Xufeng Han, Svetlana Lazebnik, Alexander C. Berg, and Tamara L. Berg. 2015. Where to Buy It: Matching Street Clothing Photos in Online Shops. In *ICCV*.

[18] Tekoing Lim, Kai-Lung Hua, Hong-Cyuan Wang, Kai-Wen Zhao, Min-Chun Hu, and Wen-Huang Cheng. 2015. VRank: Voting System on Ranking Model for Human Age Estimation. In *MMSP*.

[19] Si Liu, Luoqi Liu, and Shuicheng Yan. 2014. Fashion Analysis: Current Techniques and Future Directions. *IEEE Multimedia* 21, 2 (April-June 2014), 72–79.

[20] Ziwei Liu, Sijie Yan, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2016. Fashion Landmark Detection in the Wild. In *ECCV*.

[21] I. Ljubic, R. Weiskircher, U. Pferschy, G. W. Klau, P. Mutzel, and M. Fischetti. 2006. An algorithmic framework for the exact solution of the prize-collecting Steiner tree problem. *Mathematical Programming* 105, 2-3 (February 2006), 427–449.

[22] Babak Loni, Lei Yen Cheung, Michael Riegler, Alessandro Bozzon, Luke Gottlieb, and Martha Larson. 2014. Fashion 10000: an enriched social image dataset for fashion and clothing. In *MMSys*.

[23] Yihui Ma, Jia Jia, Suping Zhou, Jingtian Fu, Yejun Liu, and Zijian Tong. 2017. Towards Better Understanding the Clothing Fashion Styles: A Multimodal Deep Learning Approach. In *AAAI*.

[24] Meredith Ringel Morris, Kori Inkpen, and Gina Venolia. 2014. Remote Shopping Advice: Enhancing In-Store Shopping with Social Technologies. In *CSCW*.

[25] A.C. Murillo, I.S. Kwak, L. Bourdev, D. Kriegman, and S. Belongie. 2012. Urban Tribes: Analyzing Group Photos from a Social Perspective. In *CVPR Workshop*.

[26] Tam V. Nguyen, Si Liu, Bingbing Ni, Jun Tan, Yong Rui, and Shuicheng Yan. 2012. Sense Beauty via Face, Dressing, and/or Voice. In *ACM Multimedia*.

[27] Rahul Raguram and Svetlana Lazebnik. 2008. Computing Iconic Summaries of General Visual Concepts. In *CVPR Workshop*.

[28] Boon-Siew Seah, Sourav S Bhowmick, and Aixin Sun. 2014. PRISM: Concept-preserving Social Image Search Results Summarization. In *ACM SIGIR*.

[29] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. 2015. Neuroaesthetics in Fashion: Modeling the Perception of Fashionability. In *CVPR*.

[30] Edgar Simo-Serra and Hiroshi Ishikawa. 2016. Fashion Style in 128 Floats: Joint Ranking and Classification using Weak Data for Feature Extraction. In *CVPR*.

[31] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *ICLR*.

[32] Zheng Song, Meng Wang, Xian-Sheng Hua, and Shuicheng Yan. 2011. Predicting occupation via human clothing and contexts. In *ICCV*.

[33] Tsung-Hung Tsai, Wen-Huang Cheng, Chuang-Wen You, Min-Chun Hu, Arvin Wen Tsui, and Heng-Yu Chi. 2014. Learning and Recognition of On-Premise Signs (OPSs) from Weakly Labeled Street View Images. *IEEE Transactions on Image Processing* 23, 3 (March 2014), 1047–1059.

[34] Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, and Serge Belongie. 2015. Learning Visual Clothing Style with Heterogeneous Dyadic Co-occurrences. In *ICCV*.

[35] Vivek K. Verma. 2016. $2.1 trillion Global Apparel Market by 2025: Decoding opportunities for African Apparel Industry. http://educationwithvivek.blogspot.com. (2016).

[36] Sudheendra Vijayanarasimhan and Kristen Grauman. 2011. Efficient Region Search for Object Detection. In *CVPR*.

[37] Sirion Vittayakorn, Kota Yamaguchi, and Alexander C. Berg. 2015. Runway to Realway: Visual Analysis of Fashion. In *WACV*.

[38] L. C. Wang, X. Y. Zeng, L. Koehl, and Y. Chen. 2015. Intelligent Fashion Recommender System: Fuzzy Logic in Personalized Garment Design. *IEEE Transactions on Human-Machine Systems* 45, 1 (February 2015), 95–109.

[39] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, and Tao Mei. 2016. Time Matters: Multi-scale Temporalization of Social Media Popularity. In *ACM MM*.

[40] Bo Wu, Tao Mei, Wen-Huang Cheng, and Yongdong Zhang. 2016. Unfolding Temporal Dynamics: Predicting Social Media Popularity Using Multi-scale Temporal Decomposition. In *AAAI*.

[41] Kota Yamaguchi, M. Hadi Kiapour, and Tamara L. Berg. 2013. Paper Doll Parsing: Retrieving Similar Styles to Parse Clothing Items. In *ICCV*.

[42] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. 2015. Understanding Neural Networks Through Deep Visualization. In *ICML Workshop*.

[43] Dongfei Yu, Xinmei Tian, and Tao Mei. 2015. On the Selection of Trending Image from the Web. In *ICME*.

[44] Bolei Zhou, Liu Liu, Aude Oliva, and Antonio Torralba. 2014. Recognizing City Identity via Attribute Analysis of Geo-tagged Images. In *ECCV*.