GeoStyle: Discovering Fashion Trends and Events

Utkarsh Mall[1]
utkarshm@cs.cornell.edu

Kevin Matzen[2]
matzen@fb.com

Bharath Hariharan[1]
bharathh@cs.cornell.edu

Noah Snavely[1]
snavely@cs.cornell.edu

Kavita Bala[1]
kb@cs.cornell.edu

[1]Cornell University, [2]Facebook

Abstract

Understanding fashion styles and trends is of great potential interest to retailers and consumers alike. The photos people upload to social media are a historical and public data source of how people dress across the world and at different times. While we now have tools to automatically recognize the clothing and style attributes of what people are wearing in these photographs, we lack the ability to analyze spatial and temporal trends in these attributes or make predictions about the future. In this paper we address this need by providing an automatic framework that analyzes large corpora of street imagery to (a) discover and forecast long-term trends of various fashion attributes as well as automatically discovered styles, and (b) identify spatio-temporally localized events that affect what people wear. We show that our framework makes long term trend forecasts that are $> 20\%$ more accurate than prior art, and identifies hundreds of socially meaningful events that impact fashion across the globe.

1. Introduction

Each day, we collectively upload to social media platforms billions of photographs that capture a wide range of human life and activities around the world. At the same time, object detection, semantic segmentation, and visual search are seeing rapid advances [13] and are being deployed at scale [22]. With large-scale recognition available as a fundamental tool in our vision toolbox, it is now possible to ask questions about how people dress, eat, and group across the world and over time. In this paper we focus on how people dress. In particular, we ask: can we *detect and predict* fashion trends and styles over space and time?

We answer these questions by designing an automated method to characterize and predict seasonal and year-over-year fashion trends, detect social events (e.g., festivals or sporting events) that impact how people dress, and iden-

tify social-event-specific style elements that epitomize these events. Our approach uses existing recognition algorithms to identify a coarse set of fashion attributes in a large corpus of images. We then fit interpretable parametric models of long-term temporal trends to these fashion attributes. These models capture both seasonal cycles as well as changes in popularity over time. These models not only help in understanding existing trends, but can also make up to 20% more accurate, temporally fine-grained forecasts across long time scales compared to prior methods [1]. For example, we find that year-on-year more people are wearing black, but that they tend to do so more in the winter than in the summer.

Our framework not only models long-term trends, but also identifies sudden, short-term changes in popularity that buck these trends. We find that these outliers often correspond to festivals, sporting events, or other large social gatherings. We provide a methodology to automatically *discover* the events underlying such outliers by looking at associated image tags and captions, thus tying visual analysis to text-based discovery. We find that our framework finds understandable reasons for *all* of the most salient events it discovers, and in so doing surfaces intriguing social events around the world that were unknown to the authors. For example, it discovers an unusual increase in the color yellow in Bangkok in early December, and associates it with the words "father", "day", "king", "live", and "dad". This corresponds to the king's birthday, celebrated as Father's Day in Thailand by wearing yellow [36]. Our framework similarly surfaces events in Ukraine (Vyshyvanka Day), Indonesia (Batik Day), and Japan (Golden Week). Figure 1 shows more of the worldwide events discovered by our framework and the clothes that people wear during those events.

We further show that we can predict trends and events not just at the level of individual fashion attributes (such as "wearing yellow"), but also at the level of *styles* consisting of recurring visual ensembles. These styles are identified by clustering photographs in feature space to reveal *style clusters*: clusters of people dressed in a similar style. Our
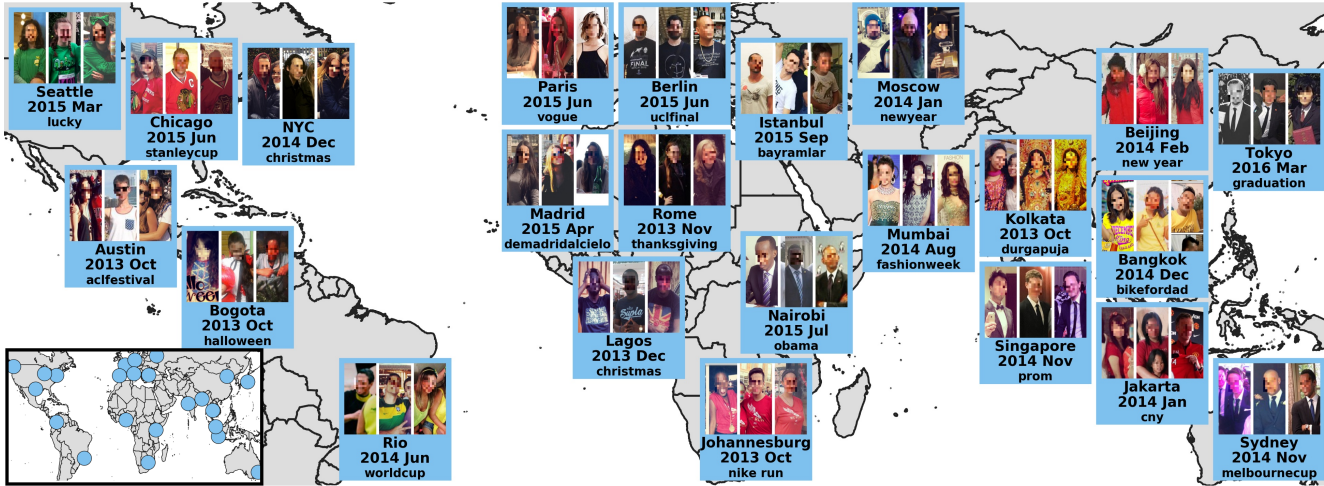
Figure 1: Major events discovered by our framework. For each event, the figure shows the clothing that people typically wear for that event, along with the city, one of the months of occurrence, and the most descriptive word extracted using the images captions. The inset image shows more precise locations of these cities.

forecasts of the future popularity of *styles* are just as accurate as our predictions of individual attributes. Further, we can run the same event detection framework described above on style trends, allowing us to not only automatically detect social events, but also associate each event with its own distinctive style; a *stylistic signature* for each event.

Our contributions, highlighted in Figure 2, include:

- We present an automated framework for analyzing the temporal behavior of fashion elements across the globe. Our framework models and forecasts long-term trends and seasonal behaviors. It also automatically identifies short-term spikes caused by events like festivals and sporting events.
- Our framework automatically discovers the reasons behind these events by leveraging textual descriptions and captions.
- We connect events with signature *styles* by performing this analysis on automatically discovered style clusters.

2. Related work

Visual understanding of clothing. There has been extensive recent work in computer vision on characterizing clothing. Some of this work recognizes attributes of people's clothing, such as whether a shirt has short or long sleeves [6, 5, 4, 42, 19, 23]. Other work goes beyond coarse image-level labels and attempts to segment different clothing items in images [39, 38, 40]. Product identification is an "instance-level classification" task used for detecting specific clothing products in photos [7, 33, 12]. Finally, there is also prior work on classifying the "style": the ensemble of clothing a person is wearing, e.g., "hipster", "goth" etc. [18]. In some cases, these labels might be unknown and require dis-

covery [23, 15], often by leveraging embeddings of images learnt by attribute recognition systems.

Our work borrows from the attribute and style literature. We make use of several human-annotated attributes on a small dataset to form an embedding space for the exploration of a much larger set of images. We use the embedding space to label attributes and styles over a vast internet-scale dataset. However, our goal is not the labeling itself, but the *discovery* of interesting geo-temporal trends and their associated styles.

Visual discovery. Although less common, there has been some prior research into using visual analysis to identify trends. Early work used low-level image features or mined visually distinctive patches [9, 29, 8] to predict geo-spatial properties such as perceived safety of cities [2, 25, 26], or ecological properties such as snow or cloud cover [41, 34, 24]. Advances in visual recognition has enabled more sophisticated analysis, such as the analysis of demographics by recognizing the make and model of cars in Street View [10]. However, while this work is exciting, the focus has been on using vision to predict known geo-spatial trends rather than discover new ones. The notion of using visual recognition to power *discovery* and *prediction of the future* is under-explored. Some initial research in this regard has focused on faces [16, 27, 11] and on human activities in a healthcare setting [21]. However, this prior work has mostly focused on descriptive analytics and manual exploration of the data to discover interesting trends. By contrast, we propose an automated, quantitative framework for both long-term forecasting and discovery. While our work focuses on the fashion domain, our ideas might be adapted to other applications as well.

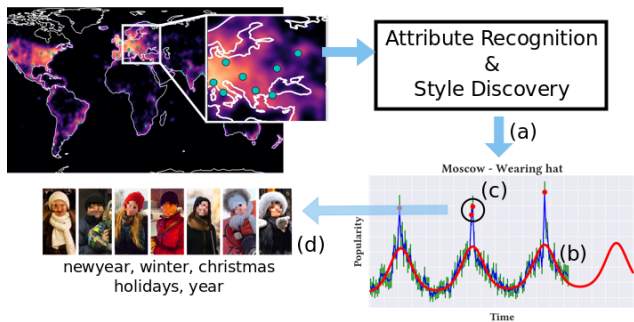**Trend analysis in fashion.** Trend analysis has also been

Figure 2: Approach overview. (a) Attribute recognition and style discovery [23] on internet images from multiple cities gives us temporal trends. (b) We fit interpretable parametric models to these trends to characterize and *forecast* (red curve is the fitted trend used to forecast). (c) Deviations from parametric models are identified as events (red points). (d) We identify text and styles specific to each event.
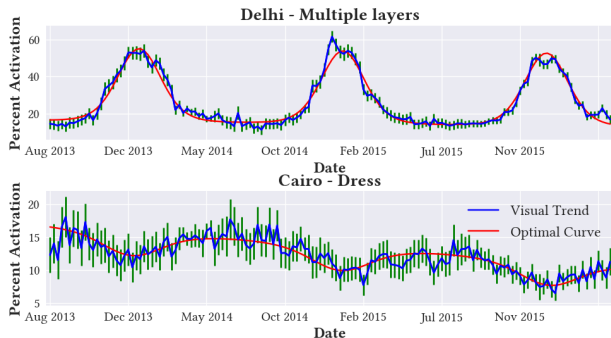

Figure 3: Two examples of observed trends. As can be seen, trends often have seasonal variations, but periodic trends are not necessarily sinusoidal. Trends can also involve a linear component (e.g., the decrease in the incidence of Dresses in Cairo over time). The green bars indicate the 95% confidence interval for each week.

applied to the fashion domain, the focus of our work. Often, prior work has considered small datasets such as catwalk images from NYC fashion shows [14]. Where larger datasets have been analyzed, interesting trends have been discovered, such as a sudden increase in popularity for "heels" in Manila [28] or seasonal trends related to styles such as "floral", "pastel", and "neon" [33]. Matzen *et al.* [23] significantly expand the scope of such trend discovery by leveraging publicly available images uploaded to social media. We build upon the StreetStyle dataset in this work. However, the analysis of the spatial and temporal trends in these papers is often descriptive, and their use for discovery requires significant manual exploration. The first problem is partly addressed by Al-Halah *et al.* [1], who attempt to make quantitative forecasts of fashion trends, but whose temporal models are limited in their expressivity, forcing them to make very coarse yearly predictions for just one year in advance. In contrast, we propose an expressive parametric model for trends that makes much higher quality, fine-grained weekly predictions for as much as 6 months in advance. In addition, we propose a framework that automates discovery by automatically surfacing interesting outlier events for analysis.

3. Method

Our overall pipeline is shown in Figure 2. We first describe our dataset and fashion attribute recognition pipeline, which we adapt from StreetStyle [23] and then describe our trend analysis and event detection pipeline.

3.1. Background: dataset and attribute recognition

Our dataset uses photos from two social media websites, Instagram and Flickr. In particular, we start with the Instagram-based StreetStyle dataset of Matzen *et al.* [23] and

extend it to include photos from the Flickr 100M dataset [32]. The same pre-processing applied to StreetStyle is also applied to Flickr 100M, including categorization of photos into 44 major world cities across 6 continents, person body and face detection, and canonical cropping. Please refer to [23] for details. In total, our dataset includes 7.7 million images of people from around the world.

Matzen *et al.* also collect clothing attribute annotations on a 27k subset of the StreetStyle dataset [23]. As in their work, we use these annotations to train a multi-task CNN (GoogLeNet [31]) where separate heads predict separate attributes, e.g., one head may predict "long-sleeves" whereas another may predict "mostly yellow". This training also has the effect of automatically producing an embedding of images in the penultimate layer of the network that places similar clothing attributes and combinations of these attributes, henceforth refered to as "styles", into the same region of the embedding vector space.

We take these attribute classifiers and apply them to the full unlabeled set of 7.7M of people images. We produce a temporal trend for each attribute in each city by computing, for each week, the mean probability of an attribute across all photos from that week and city. Per-image probabilities are derived from the CNN prediction scores after calibration via isotonic regression on a validation set [23].

3.2. Characterizing trends

Given each weekly clothing attribute trend in each city, we wish to (a) characterize this trend in a human-interpretable manner, and (b) make accurate forecasts about where the trend is headed in the future.

Figure 3 shows two examples of attribute trends over time. We observe several behaviors in these examples. First, there are both coarse-level trends extending over months or years
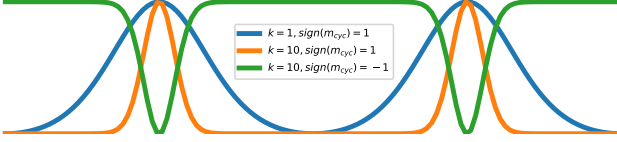
Figure 4: We use a function of the form $m_{\mathsf{cyc}}e^{k\sin(\omega x+\phi)-k}$ as our cyclical component because of its ability to model seasonal spikes. This plot shows this function for three values of k and m_{cyc}. For ease of comparison, all three functions have been centered and rescaled to the same dynamic range.

(e.g., the seasonal cycles in the wearing of multiple layers in Delhi) as well as fine-scale spikes that occur over days or weeks (e.g., the spike in December 2014). Second, the coarse trend often has a strong periodic component usually governed by different seasons. Third, instead of even sinusoidal upswings and downswings, the periodic trend often consists of upward (Figure 3 top) or downward (Figure 3 bottom) *surges* in popularity. Fourth, in some cases this periodic trend is superimposed on a more gradual increase or decrease in popularity, as in Figure 3 (bottom).

We seek to identify both the coarse, slow-changing trends that are governed by seasonal cycles or slow changes in popularity, as well as the fine-grained spikes that might arise from events such as festivals (Christmas, Chinese New Year) or sporting events (FIFA World Cup). The former might tell us how people in a particular place dress in different seasons, while the latter might reveal important social events with many participants. We first fit a parametric model to capture the slow-changing trends (this section), and then identify potential events as large departures from the predicted trends (Section 3.3).

We model slow-changing trends using a parametric model $f_\theta(t)$, which is a convex combination of two components: a linear component and a cyclical component:

$$f_\theta(t) = (1 - r) \cdot L(t) + r \cdot C(t) \qquad (1)$$

where the parameter $r \in [0, 1]$ defines the contribution of each component. The linear component, $L(t)$ is characterized by slope m_{lin} and intercept c_{lin}:

$$L(t) = m_{\mathsf{lin}}t + c_{\mathsf{lin}} \qquad (2)$$

A standard choice for the cyclical component would be a sinusoid. However, we want to capture upward and downward surges, so we instead use a more expressive cyclical component of the form:

$$C(t) = m_{\mathsf{cyc}}e^{k\sin(\omega t+\phi)-k}. \qquad (3)$$

When k is close to 0, this function behaves like a (shifted) sinusoid, but for higher values of k, it has more peaky cycles (Figure 4). ω and ϕ denote period and phase respectively.

Parameter	Intuitive meaning
r	Trade-off between linear and cyclic trend
c_{lin}	Long term bias
m_{lin}	Rate of long-term increase/decrease in popularity
m_{cyc}	Amplitude and sign (upwards/downwards) of cyclical spikes
k	Spikiness of cyclical spikes
ω	Frequency of cyclical spikes
ϕ	Phase of cyclical spikes

Table 1: Intuitive descriptions of all parameters

The full set of parameters in this parametric model is $\theta = \{r, m_{\mathsf{cyc}}, k, \omega, \phi, m_{\mathsf{lin}}, c_{\mathsf{lin}}\}$. Table 1 provides intuitive descriptions of these parameters. Because each parameter is interpretable, our model allows us to not just make predictions about the future but also to *discover* interesting trends and analyze them, as we show in Section 4.1.

We fit the parameters θ of the above model to the weekly trend of each attribute for each city by solving the following non-linear least-squares problem:

$$\theta^* = \underset{\theta}{\arg\min} \sum_t \left(\frac{f_\theta(t) - T(t)}{\sigma(t)} \right)^2 \qquad (4)$$

where $T(t)$ represents the observed average probability of the attribute for week t in the particular city/continent/world and $\sigma(t)$ measures the uncertainty of the measurement (reciprocal of the binomial confidence). We minimize Equation (4) using the Trust Region Reflective algorithm [20]. To prevent overfitting we set an upper bound for ω to keep seasonal variation close to annual variation. We set it to $\frac{2\pi \times 2}{52}$, allowing for a maximum of two complete sinusoidal cycles over a year. We chose 52 because we measure time in weeks.

3.3. Discovering events

Given a fitted model, we now describe how we identify more fine-grained structure in each attribute trend, and correlate these structures with potentially important social gatherings. In particular, we are interested in sharp spikes in popularity of particular kinds of clothing, which often are due to an event. For example, people might wear a particular jersey to support their local team on game night, or wear green on St. Patrick's Day.

To discover such events, we start by identifying weeks with large, positive deviations from the fitted model, or *outliers*, using a binomial hypothesis test. The set of images in week t are considered as a set of trials, with those images classified as positive for the attribute constituting "successes" and others failures. The null hypothesis is that the probability of a success is given by the fitted parametric model, $f_\theta^*(t)$. Because we are interested in *positive* deviations from this expectation, we use a one-tailed hypothesis test, where the alternative hypothesis is that the true probability of success is greater than this expectation. We identify outliers as weeks

with p-value < 0.05. We use the reciprocal of the p-value, denoted by s, as a measure of outlier *saliency*.

We then connect the outliers discovered to the social event that caused them, if any. To do so, we note that some of these events might be repeating, annual affairs (such as festivals), while others might be one-off events (e.g. FIFA World Cup). We therefore formalize an *event* as a *group of outliers* that are either localized on a few weeks (**one-off events**) or are separated by a period of approximately a year (**annual events**, like festivals on a solar or a lunisolar calendar [37]).

To determine if our detected outliers fit some event, we need a way to *score* candidate events. If we have a sequence of outliers $g = \{t_1, \ldots, t_k\}$ for a particular trend in a specific city, how do we say if this group of outliers is likely to be an actual event? There are two main considerations in this determination. First, the outliers involved in the event must be *salient*, that is, they should correspond to significant departures from the background trend. Second, they should have the *temporal signature* described above: the outliers involved should either be localized in time, or separated by approximately a year.

We formalize this intuition by defining a cost function $C(g)$ for each group of outliers $g = \{t_1, \ldots, t_k\}$ such that a smaller cost indicates a higher likelihood of g being an event. $C(g)$ is a product of two terms: a cost incentivizing the use of salient outliers (we use the reciprocal of the average saliency $\bar{s}$ of the outliers involved), and a cost $C_T(g)$ measuring the deviation from the ideal temporal signature:

$$C(g) = \frac{C_T(g)}{\bar{s}} \qquad (5)$$

$C_T(g)$ considers consecutive outliers in g and assigns a low cost if these consecutive outliers occur very close to each other in time, or are very close to following an annual cycle. If consecutive events are neither proximal (they are more than $\Delta_{\max}$ weeks apart) nor part of an annual or multi-year cycle (they miss the cycle by more than $d_{\max}$ weeks), the cost is set to infinity. Concretely, we define C_T as follows:

$$C_T(g) = \frac{\sum_{i=1}^{|g|-1} C_p(t_{i+1} - t_i)}{|g| - 1} \qquad (6)$$

$$C_p(\Delta) = \begin{cases} \frac{\Delta + c}{\Delta_{\max} + c} & \text{if } \Delta < \Delta_{\max} \\ \frac{d(\Delta) + b}{d_{\max} + b} & \text{if } \Delta \geq T - d_{\max} \\ & \text{and } d(\Delta) < d_{\max} \\ \infty & \text{otherwise.} \end{cases} \qquad (7)$$

Here, $|g|$ denotes the cardinality of outlier group g. Δ is the time difference between consecutive outliers, T is the length of a year, and $d(\Delta)$ measures how far Δ is from an annual cycle. In particular, $d(\Delta) = \min(\Delta \mod T, -\Delta$

$\mod T)$. $c = 18, b = 15, \Delta_{\max} = 2$ and $d_{\max} = 5$ are constants. The setting of these is explained in the supplementary. When g contains a single event, $C_T(g)$ is defined to be 1.

$C(g)$ gives us a way of scoring candidate events, but we still need to come up with a set of candidates in the first place from the discovered outliers. There may be multiple events in a city over time (e.g., Christmas and Chinese New Year), and we need to separate these events. We consider this as a *grouping* problem: given a set of outliers occuring at times $t_1, \ldots, t_n$ in the trend of a particular attribute in a particular city, we want to *partition* the set into groups. Each group is then a candidate event. We define the cost of a partition $P = \{g_1, \ldots, g_k\}$ as the average cost $C(g_i)$ of each group g_i in the partition, and choose the partition that minimizes this cost:

$$P^* = \underset{P}{\arg\min} \frac{\sum_i C(g_i)}{|P|} \qquad (8)$$

This is a combinatorial optimization problem. However, we find that there are very few outliers for each trend, so this problem can be solved optimally using simple enumeration.

Running this optimization problem for each trend gives us a set of events, each corresponding to a group of outliers. Each event is then associated with a cost $C(g)$. We define the reciprocal of this cost as the *saliency of the event*, and we rank the events in decreasing order of their saliency.

Mining underlying causes for events. To derive explanations for each event, we analyze image captions that accompany the image dataset. We consider images from the relevant location classified as positive for the relevant attribute across the year, and split them into two subsets: those appearing within the event weeks, and those at other times. Words appearing in captions of the former but not the latter may indicate why the attribute is more popular specifically in that week. To find these words, we do a TF-IDF [30] sorting, considering the captions of the first set as positive documents and the captions of the second set as negatives. Images can contribute to a term at most once in term frequencies. We perform this analysis using the English language captions.

3.4. Style trend analysis

We also wish to identify trends not just in single attributes, but also in combinations of attributes that correspond to looks or styles. However, the number of possible attribute combinations grows exponentially with the number of attributes considered, and most attribute combinations are uninteresting because of their rarity: e.g., pink, short-sleeved, suits. Instead, we want to focus on the limited set of attribute combinations that are actually prevalent in the data. To do so, we follow the work of Matzen *et al.* [23] to discover *style clusters*: popular combinations of attributes. Style clusters are identified using a Gaussian mixture model to cluster images in the feature space learned by the CNN. To ensure coverage of all trends while also maintaining sufficient data

for each style cluster, we separately find a small number of style clusters in each city. In general, we find that the style clusters we discover correspond to intuitive notions of style. As with individual attributes, our trend analysis on these clusters tells us not only which styles are coming into or going out of fashion, but also associates styles with major social events (Section 4.3).

4. Results

We now evaluate our ability to discover and predict style events and trends. In addition, we visualize discovered trends, events, and styles.

4.1. Trend prediction and analysis

We first evaluate our parametric temporal model (Eq. 1) based on its ability to make out-of-sample predictions about the future (in-sample predictions are provided in the supplementary). We compare to models proposed by Al-Halah *et al.* [1], the most relevant prior work. We also compare to four ablations of our model: (a) Linear: $f_{\text{linear}}(t) = m_{\text{lin}}t + c$, (b) Sinusoidal fit: $f_{\text{sin}}(t) = \sin(\omega t + \phi)$, (c) Cyclic fit: $f_{\text{cyclic}}(t) = m_{\text{cyc}}e^{k\sin(\omega t + \phi) - k}$ and (d) a linear combination of f_{linear} and f_{sin}. We use the same metrics as Al-Halah *et al.* [1], namely, MAE and MAPE. The latter looks at the average absolute error relative to the true trend $T(t)$, expressed as a percentage. However, while Al-Halah *et al.* only evaluate prediction accuracy in the extreme short term (the very next data point), we consider prediction accuracy both in the *short term* (next data point, or next week) as well as the long term (next 26 data points, or next 6 months). Note that even though Al-Halah *et al.* only evaluate predictions over the next data-point, that data point corresponds to a full year. Hence they are predicting trends farther in the future, but their prediction is relatively coarser. We also show the results of our prediction for more than one year in supplementary.

We find that our parametric model is significantly better than all baselines at both long-term and short-term predictions (see Table 2). Furthermore, the gap between our model and the best method found by Al-Halah *et al.* (exponential smoothing) *increases* when we move to making long-term predictions. We also observe that our model's out-of-sample performance actually matches in-sample performance (shown in supplementary) very well, indicating strong generalization. This shows that our model generalizes better and can extrapolate significantly further into the future.

Interestingly, our model is also significantly better than the autoregressive baselines. These baselines predict a data point as a weighted linear combination of the previous k data points, where the weights are learned from data and k is cross-validated. Thus, these models have many more parameters than our model (up to $12\times$ more). The fact that

	Attribute-based trends			
Model	**Next week**		**Next 26 weeks**	
	MAE	MAPE	MAE	MAPE
mean	0.0209	19.05	0.0292	25.79
last	0.0153	15.56	0.0226	21.04
AR	0.0147	14.18	0.0207	20.27
VAR	0.0146	16.16	0.0162	18.92
ES	0.0152	14.92	0.0231	20.59
linear	0.0276	18.35	0.0365	24.40
sinusoid	0.0141	13.22	0.0163	16.09
sin+lin	0.0140	13.17	0.0169	16.87
cyclic	0.0129	12.63	0.0165	16.64
Ours	**0.0119**	**12.13**	**0.0145**	**15.73**

	Style-based trends				
Model	**Next 26 weeks**		**Model**	**Next 26 weeks**	
	MAE	MAPE		MAE	MAPE
mean	0.0101	31.82	linear	0.0135	36.05
last	0.0145	44.57	sinusoid	0.0083	23.23
AR	0.0090	37.89	sin+lin	0.0081	23.04
VAR	0.0120	27.97	cyclic	0.0085	24.16
ES	0.0143	43.96	Ours	**0.0077**	**21.78**

Table 2: Comparison of our prediction model against other models from [1]. Mean and Last are naive methods that predict the mean and last of the known time series as the next prediction respectively. AR (autoregression) and VAR (vector-autoregression) are autoregressive methods. ES is exponential smoothing. Lower values are better.

our model still performs better suggests that choosing the right parametric form is more important than merely the size or capacity of the model.

Interpretability: Our model fitting characterizes each attribute trend in terms of a few interpretable parameters, shown in Table 1, which can be used in a straightforward manner to reveal insights. For example, ϕ describes the phase of the cyclical trend. If we look at cities where there is a positive spike in people wearing multiple layers in the winter, then the peaks should occur in winter months, and cities in the northern and southern hemisphere should be exactly out of phase. Figure 5 shows the difference in phase ϕ for the *multiple-layered* clothing attribute between each pair of cities. We find that cities indeed cluster together based on their hemisphere, with cities in the same hemisphere closer to each other in phase. Interestingly, cities closer to the equator seem to be half-way between the two hemispheres and form their own cluster.

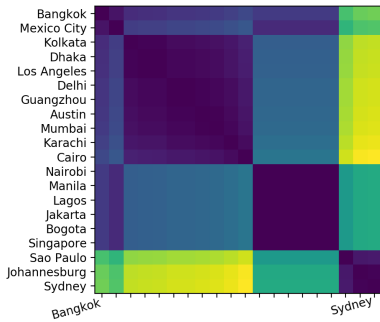As another example, k represents the "spikiness" of the

Figure 5: Phase difference for the *multiple-layered* attribute between 20 cities, using estimated phase parameter ϕ.
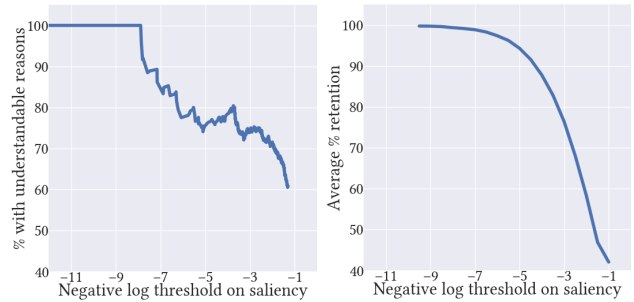

Figure 6: **Left:** The percentage of events with saliency greater than a threshold that are explainable, plotted as the threshold varies. **Right:** The percentage of events retained when another sample with replacement is used for detection.

cyclical trend: a high k corresponds to a very short-duration increase/decrease in popularity. We can search for attribute trends that show the spikiest (i.e., highest k) annual positive spikes. These turn out to be *wearing-scarves* in Bangkok and *clothing-category-dress* in Moscow. This might reveal the fact that Bangkok has a very short winter when people wear scarves, while Moscow has a short summer where people wear dresses.

4.2. Event discovery

After fitting our parametric trend model, we discover events using the method discussed in Section 3.3. Our event discovery pipeline detected hundreds of events, detailed in the supplementary. Table 3 shows the five most salient events along with the corresponding words associated with the event and a set of corresponding images. All five correspond to significant social gatherings that some or all of the authors were unaware of *a priori* :

1. **Father's Day in Bangkok** is celebrated on the King's birthday, and people wear yellow to honor the king.
2. **FreakNight in Seattle** is a dance music event held on or around Halloween. The prevalance of sleeveless clothes is an outlier driven by this event given cool weather at this time of the year.
3. **Songkran in Bangkok** is a festival celebrated in April on the Thai New Year and involves people playing with water in warm weather.
4. The Western Conference Finals of the **Stanley Cup 2014 in Chicago** involved the Chicago Blackhawks and the Los Angeles Kings. People wore their home team's jerseys.
5. The **FIFA World Cup** was held in Brazil in 2014 and featured a prevalence of yellow jerseys in support of Brazil.

Note that events such as Father's Day were further correctly identified as annual events.

Quantitative evaluation: Quantitative evaluation of our discovered events is challenging because there is no dataset or annotations of all the significant social events in the world.

However, we can check if the events we discover do in fact correspond to real social events, which can be construed as a kind of precision.

To do this evaluation, we manually inspect each discovered event and the associated top keywords to see if they reveal an understandable explanation: a real social event. We measure the percentage of events with saliency greater than a threshold for which we found such a reason. Figure 6 shows this percentage as a function of the saliency threshold. We find that **100%** of the most salient events and **60%** of all events have explainable reasons, indicating both the ability of our model to detect events and its ability to identify appropriate keywords for them. Not surprisingly, the percentage of explainable events decreases as event saliency decreases, which validates our model's estimate of saliency as a measure of probability of corresponding to a real-world explainable event.

We also evaluate the robustness of our event detector by measuring the stability of detected events across random subsets of data. We resample the dataset 20 times with replacement, and run both the trend characterization and event detection on each subset. We then measure the fraction of outliers with saliency greater than a threshold in one sampled set that are still salient in a second set. We call this fraction the *retention*, and plot it in Figure 6. Ideally, we want all salient events we detect in one dataset to be detected in all datasets, yielding high retention. Indeed, the high saliency events are retained in other folds. Furthermore, this retention rate increases consistently as the threshold value on saliency increases, indicating that the reciprocal of p-value is indeed a good measure of the saliency of events.

4.3. Style trend analysis

Finally, we run the same trend analysis and event detection pipeline on style clusters. Table 2 shows the prediction error of our parametric trend analysis compared to various baselines when making long-term fine-grained predictions

City	Bangkok	Seattle	Bangkok	Chicago	Rio
Attribute	Yellow color	No sleeves	T-shirt	Red color	Yellow color
Month	2014 Dec, 2015 Dec	2014 Oct	2014 Apr	2014 Jun	2014 Jun, 2014 Jul
Keywords	dad, father	halloween, freaknight	songkran, festival	cup, stanleycup	worldcup, brasil

Table 3: Top five events detected across the world by finding anomalous behaviour in trends using methods from Section 3.2. The words from the captions of the image posts are sorted by their TF-IDF scores in the associated event week (top-2 are shown). Images from each event are sorted based on number of terms in their caption matching the top-5 keywords.

over the next 26 weeks. We find that our approach again significantly outperforms all baselines, and by a larger margin.

Figure 1 shows the most salient style-based events for selected cities. We find that with style clusters, we are able to identify events that involve attribute combinations, e.g., people wearing glasses with sleeveless tops during the ACL festival in Austin. More striking are events such as Durga Puja in Kolkata or Fashion Week in Mumbai which are discovered in spite of the fairly nuanced associated appearance.

4.4. Cross-dataset generalization

We also show that our method generalizes well to cities not seen during CNN training. We collected Flickr images from Barcelona (a city not in [23]) from 2013 to mid-2018 and fed them through the pipeline described in Section 3. We detected a total of 97k people in these photos.

We test the predictability of our trend prediction method on this unseen set of images. We used images from 2013 to mid-2017 to fit trends, then predicted the trend for the final year of data. Our model (MAE=0.043) performs significantly better than the best baseline, Autoregression (MAE=0.047), although fitting a sinusoid with a linear component also gives comparable performance (MAE=0.043). We suspect this is because Barcelona does not see significant variations in weather [35] and hence a smoother sinusoid models the seasonal changes as well as our model.

We also discovered events in Barcelona using the method described in Section 3.3. The top-most event discovered in Barcelona corresponds to people gathering in yellow shirts for the "**Catalan Way**", a long human chain in support of Catalan independence from Spain, in September 2013 (Figure 7). This event is a significant political event, and it validates our framework's ability to identify important social events from raw data across multiple datasets and bring them to the fore.

Figure 7: Images from "Catalan Way" an event discovered from September 2013 in Barcelona.

5. Conclusion and Future Work

This work has established a framework for automatically analyzing temporal trends in fashion attributes and style by examining millions of photos published publicly to the web. We characterized these trends using a new model that is both more interpretable and makes better long-term forecasts. We also presented a methodology to automatically discover social events that impact how people dress. However, this is but a first step and there are many questions still to be answered, such as the identification and mitigation of biases in social media imagery, and the propagation of styles across space. The problem of analyzing trends is also relevant in other visual domains, such as understanding which animals are getting rarer over time in camera trap images [3] or how land-use patterns are changing in satellite imagery [17]. We therefore believe that this is an important problem deserving of future research.

References

[1] Ziad Al-Halah, Rainer Stiefelhagen, and Kristen Grauman. Fashion forward: Forecasting visual style in fashion. In *ICCV*,

2017. 1, 3, 6

[2] Sean M Arietta, Alexei A Efros, Ravi Ramamoorthi, and Maneesh Agrawala. City Forensics: Using visual elements to predict non-visual city attributes. *IEEE Trans. Visualization and Computer Graphics*, 20(12), Dec 2014. 2

[3] Sara Beery, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In *ECCV*, 2018. 8

[4] Lukas Bossard, Matthias Dantone, Christian Leistner, Christian Wengert, Till Quack, and Luc Van Gool. Apparel classification with style. In *Proc. Asian Conf. on Computer Vision*, 2013. 2

[5] Lubomir Bourdev, Subhransu Maji, and Jitendra Malik. Describing people: Poselet-based attribute classification. In *ICCV*, 2011. 2

[6] Huizhong Chen, Andrew Gallagher, and Bernd Girod. Describing clothing by semantic attributes. In *ECCV*, 2012. 2

[7] Wei Di, C. Wah, A. Bhardwaj, R. Piramuthu, and N. Sundaresan. Style finder: Fine-grained clothing style detection and retrieval. In *Proc. CVPR Workshops*, 2013. 2

[8] Carl Doersch, Abhinav Gupta, and Alexei A. Efros. Mid-level visual element discovery as discriminative mode seeking. In *NeurIPS*, 2013. 2

[9] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros. What makes Paris look like Paris? *SIGGRAPH*, 31(4), 2012. 2

[10] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc. National Academy of Sciences*, 2017. 2

[11] Shiry Ginosar, Kate Rakelly, Sarah Sachs, Brian Yin, and Alexei A. Efros. A Century of Portraits: A visual historical record of american high school yearbooks. In *ICCV Workshops*, 2015. 2

[12] M Hadi Kiapour, Xufeng Han, Svetlana Lazebnik, Alexander C Berg, and Tamara L Berg. Where to buy it: Matching street clothing photos in online shops. In *ICCV*, 2015. 2

[13] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *ICCV*, 2017. 1

[14] Shintami C. Hidayati, Kai-Lung Hua, Wen-Huang Cheng, and Shih-Wei Sun. What are the fashion trends in New York? In *Proc. Int. Conf. on Multimedia*, 2014. 3

[15] Wei-Lin Hsiao and Kristen Grauman. Learning the latent "look": Unsupervised discovery of a style-coherent embedding from fashion images. In *ICCV*, 2017. 2

[16] Mohammad T Islam, Connor Greenwell, Richard Souvenir, and Nathan Jacobs. Large-scale geo-facial image analysis. *EURASIP J. on Image and Video Processing*, 2015(1), 2015. 2

[17] Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. Tile2vec: Unsupervised representation learning for spatially distributed data. In *AAAI*, 2019. 8

[18] M. Hadi Kiapour, Kota Yamaguchi, Alexander C. Berg, and Tamara L. Berg. Hipster wars: Discovering elements of fashion styles. In *ECCV*, 2014. 2

[19] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *CVPR*, 2016. 2

[20] Ladislav Lukšan. Inexact trust region method for large sparse nonlinear least squares. volume 29, 1993. 4

[21] Zelun Luo, Jun-Ting Hsieh, Niranjan Balachandar, Serena Yeung, Guido Pusiol, Jay Luxenberg, Grace Li, Li-Jia Li, N Lance Downing, Arnold Milstein, et al. Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring. *Machine Learning for Healthcare (MLHC)*, 2018. 2

[22] Dhruv Mahajan, Ross Girshick, Vignesh Ramanathan, Kaiming He, Manohar Paluri, Yixuan Li, Ashwin Bharambe, and Laurens van der Maaten. Exploring the limits of weakly supervised pretraining. *arXiv preprint arXiv:1805.00932*, 2018. 1

[23] Kevin Matzen, Kavita Bala, and Noah Snavely. Streetstyle: Exploring world-wide clothing styles from millions of photos. *CoRR*, 2017. 2, 3, 5, 8

[24] Calvin Murdock, Nathan Jacobs, and Robert Pless. Building dynamic cloud maps from the ground up. In *ICCV*, 2015. 2

[25] Nikhil Naik, Jade Philipoom, Ramesh Raskar, and César Hidalgo. Streetscore: Predicting the perceived safety of one million streetscapes. In *Proc. CVPR Workshops*, 2014. 2

[26] Vicente Ordonez and Tamara L. Berg. Learning high-level judgments of urban perception. In *ECCV*, 2014. 2

[27] Tawfiq Salem, Scott Workman, Menghua Zhai, and Nathan Jacobs. Analyzing human appearance as a cue for dating images. In *WACV*, 2016. 2

[28] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *CVPR*, 2015. 3

[29] Saurabh Singh, Abhinav Gupta, and Alexei A. Efros. Unsupervised discovery of mid-level discriminative patches. In *ECCV*, 2012. 2

[30] Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972. 5

[31] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *CVPR*, 2015. 3

[32] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. Yfcc100m: The new data in multimedia research. *Commun. ACM*, 59(2):64–73, Jan. 2016. 3

[33] Sirion Vittayakorn, Kota Yamaguchi, Alexander C Berg, and Tamara L Berg. Runway to Realway: Visual analysis of fashion. In *WACV*, 2015. 2, 3

[34] Jingya Wang, Mohammed Korayem, and David J. Crandall. Observing the natural world with flickr. In *ICCV Workshops*, 2013. 2

[35] Wikipedia contributors. Climate of Barcelona. 8

[36] Wikipedia contributors. Father's Day. 1

[37] Wikipedia contributors. Lunisolar Calendar. 5

[38] Kota Yamaguchi, M Hadi Kiapour, and Tamara L Berg. Paper doll parsing: Retrieving similar styles to parse clothing items. In *ICCV*, 2013. 2

[39] Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, and Tamara L Berg. Parsing clothing in fashion photographs. In *CVPR*, 2012. 2

[40] Wei Yang, Ping Luo, and Liang Lin. Clothing co-parsing by joint image segmentation and labeling. In *CVPR*, 2014. 2

[41] Haipeng Zhang, Mohammed Korayem, David J. Crandall, and Gretchen LeBuhn. '. In *WWW*, 2012. 2

[42] Ning Zhang, Manohar Paluri, Marc'Aurelio Rantazo, Trevor Darrell, and Lubomir Bourdev. Panda: Pose aligned networks for deep attribute modeling. In *CVPR*, 2014. 2