From Culture to Clothing: Discovering the World Events Behind A Century of Fashion Images

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

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

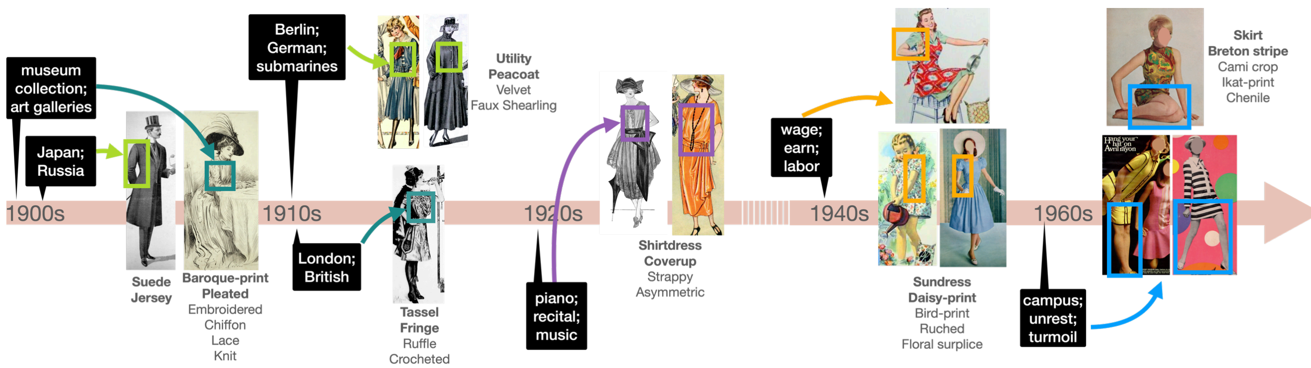


Figure 1: **Fashion history timeline** created by our model. We automatically discover iconic styles in each era, and detect the world events or cultural factors that gave rise to their popularity. For example, during the 1910s, World War I (*i.e.* war-related activities in England and Germany) influenced utility clothing. On the other hand, postwar in the 1920s, people's increasing engagement in music and entertainment gave birth to the flapper style. Words listed in call-out boxes are the top words in the textual cultural topic (topic discovery described in Sec. 3.4). Images and the words below are the iconic styles and detected clothing attributes/categories at that time (details in Sec. 3.3).

Abstract

Fashion is intertwined with external cultural factors, but identifying these links remains a manual process limited to only the most salient phenomena. We propose a data-driven approach to identify specific cultural factors affecting the clothes people wear. Using large-scale datasets of news articles and vintage photos spanning a century, we present a multi-modal statistical model to detect influence relationships between happenings in the world and people's choice of clothing. Furthermore, on two image datasets we apply our model to improve the concrete vision tasks of visual style forecasting and photo timestamping. Our work is a first step towards a computational, scalable, and easily refreshable approach to link culture to clothing.

1. Introduction

Fashion is much more than Paris runways and Instagram darlings. Fashion is what all people wear—how they express their identity, taste, interests, and even mood. At the same time, fashion is a reflection of our society and its cultural influences. Studying clothing trends over long periods of time offers anthropologists a treasure trove of information about how people's clothing is affected by happenings in the world, whether economic, political, or social.

The effects are felt from both localized and broad sources. On the one hand, one-off momentous occasions or iconic individuals can cause ripples of fashion changes, such as the publicity surrounding Titanic's voyage that attracted American designers to European fashions in the 1910s [13], or the oversized sunglasses donned by Jackie Kennedy that continue to inspire women decades later [3]. On the other hand, longer periods of a collective experience can also shape trends, such as U.S. women of the 1950s forgoing boxier wartime styles for softer lines and full skirts [7], or some African-Americans embracing ethnic Afro hairstyles following the civil rights movement [1], or today's shift towards comfortable casual wear in the midst of the COVID pandemic and work-from-home conditions.

While such links are intriguing, today they require people cognizant of both clothing and cultural trends to manually tease them out. This process requires expert knowledge and expensive labor, making it difficult to scale. As a result, such analysis is not updated frequently, and it focuses mostly on a few major salient styles (like those referenced above), ignoring many more subtle or localized ones.

We propose to automatically discover which events and which cultural factors influence the evolution of fashion styles (see Fig. 1). Our idea is to associate changing style trends observed in photos with their preceding world events. To that end, we explore how photos and news articles can together tell the story of why certain clothing elements ebb and flow—or spike and disappear—over time. Specifically, we investigate temporal relations and statistical influences between latent topics discovered from 100M news articles and fashion styles mined from both i) Vintage, a new dataset spanning a century of vintage photos and ii) GeoStyle [44], a massive set of social media photos spanning recent years. Upon discovering influential events, we demonstrate their significance by posing style forecasting and photo timestamping tasks. On two datasets, we show that modeling the culture-clothing link enables more accurate forecasts of visual styles' future popularity and more accurate dating of an unseen clothing photo.

The key novelty of this work is to formalize and automate the process of tying cultural factors to fashion. To meet that goal, we devise technical components on the computer vision and statistical modeling side, such as a robust visual style discovery pipeline based on body regions and a multi-modal Granger causality module. To our knowledge, we are the first to quantify the links from cultural trends to fashion trends in image data, and the first to demonstrate the value for forecasting and timestamping. We show the former improves over existing methods in the literature. In addition, we show examples of the discovered influences, both dominant and subtle. Beyond providing better forecasts and timestamping, our approach is a step towards a computational, scalable, and easily refreshable way to understand how external factors affect the clothes we wear.

2. Related work

Clothing recognition and styles. Computer vision for fashion often starts with retrieving similar garments [34, 37, 40], detecting their characteristics (color, pattern, or shape *attributes*) and categories (dress, blouse) [9, 29, 32, 41], or segmenting individual garments [16, 39, 55, 58], which are all essential for product search [20, 36, 57]. Beyond recognition, recent work infers how well one garment matches another [21, 26, 30, 47, 51, 52], how fashionable an outfit is as a whole [28, 49], and which garments flatter which body shapes [24, 27]. While most prior work learns clothing *styles* with supervision [35, 41, 50], styles can also be mined automatically [5, 25, 33, 44, 45]. Our model begins by recognizing garments' attributes and categories, then automatically discovers localized visual styles. However, unlike any of the methods above, we analyze style changes through time in the context of cultural events.

Visual trend analysis and dating photos. In addition to clothing styles [5, 25, 45], styles are also of interest in other visual phenomena like automobiles [31] and architecture [12]. Earlier work uses hand-engineered features (*e.g.*, HOG descriptors [10]) to mine for localized part patches (*e.g.*, car headlights, building windows, shoulder vs. waist regions) that are visually consistent but temporally discriminative, then tracks styles' transitions through time and/or space [12, 31, 54]. Several prior datasets gather photos annotated by their dates in order to track trends. For example, the car dataset [31] spans cars from year 1920 to 1999; the US yearbook dataset [17] spans faces from 1905 to 2013; a clothing dataset [53] spans 1900 to 2009.[1] The above methods focus on tracking discovered styles' transitions and predicting the year or decade the subject (car or person or clothing) was from.

Our work also entails modeling visual styles over time, but with the unique goal of automatically grounding style trends in world events. While the yearbook project [17] discusses links to social and cultural happenings (e.g., prevalence of glasses, curvature of people's smiles) and points to literature supporting those plausible influences, the links are anecdotal and found manually. In contrast, we *automatically detect* cultural factors that influence the styles, and we use the detected influences to improve two quantitative tasks—forecasting and timestamping.

Influence modeling. While fashions spread to other places and people, only limited prior work explores modeling fashion's *influence*. Hypothesizing a runway to real-way pattern of influence, one method anecdotally tracks the style trend dynamics of three attributes (floral, neon, pastel) [54], while another monitors the correlation coefficient—but not influence or causality—of attribute changes for the NY fashion show and the public within the same year [8]. The GeoStyle project identifies anomalies in an Instagram style's trendline, then looks at text in the corresponding image captions to explain them (e.g., finding a sudden burst of yellow shirts in Thailand for the king's birthday) [44]. Beyond correlations, we discover the influence of a root source that precedes a fashion event. To our knowledge, the only work in computer vision utilizing detected influences are two recent style forecasting methods [4, 43]. They model fashion influences between cities [4] or between multiple styles and their taxonomy [43], with no reference to cultural events. All the above methods only consider influence relations within visual content, whereas we study the influence of external factors (news events) on visual styles.

3. Approach

We first introduce the Vintage clothing image dataset and textual corpus we collected for this problem (Sec. 3.1). We then describe our model for mining clothing styles across a century (Sec. 3.2 and Sec. 3.3), and discovering cultural factors from news articles (Sec. 3.4). Finally, we detect which

[1] This dataset is not publicly available.

cultural factors influence which clothing styles (Sec. 3.5), and apply the discovered influences to forecast future clothing trends (Sec. 3.7) and timestamp photos (Sec. 3.8).

Throughout, we perform parallel experiments using our new Vintage dataset—which offers a long historical window on culture—as well as GeoStyle [44], an existing large-scale dataset of social media photos—which offers a rich record of cultural effects in recent years.

3.1. Collecting a century of data

The twentieth century saw the most rapid evolution in clothing so far. Mass-production techniques were introduced, and people's roles in society changed. Clothing designs became freer in form than in previous centuries. Moreover, photography became popular beyond professionals in the late 19th century [46], generating more records of people's daily outfits. For all these reasons, we select the twentieth century to study clothing's evolution with society. The methodology could be extended to other time frames, provided access to relevant data.

Image data. To construct a large collection of photos spanning the twentieth century, we use the online social platform Flickr, where users share photos of topics/subjects they love. *Vintage* is one of the popular topics on Flickr. Users upload scans of old photos, magazine covers, and posters, with meta-data often describing when and by whom the images were created. We use keywords related to vintage clothing to retrieve publicly available images and their meta-data, and automatically parse the meta-data to obtain date labels for each image. The resulting Vintage image dataset contains 6,292 photos and 11,898 clothing instances in total (details below). It is the largest date-annotated clothing dataset publicly available, and thanks to its community photo sharing origins, it enjoys some organic diversity: anyone can share their photos on Flickr, as opposed to photos curated in museums or textbooks. That said, as with any Internet photo collection, certain biases are possible. In our case, sampling bias exists due to variance in how widespread photography was at different times and locations. For example, the outfits are mostly Western styles, subjects in the images are often fashion models, movie actors, or political characters, and there are more images in later than earlier decades. Fig. 1, Fig. 2, and Fig. 3 show example images.

In Sec. 4, we use Vintage to discover influences across 100 years, and we separately use the 7M-image GeoStyle dataset [44] to discover influences within 2013-2016. The two datasets have complementary strengths for our study: the Vintage photos cover a much longer time period and multi-media sources (personal photos, magazines, ads, etc.), whereas the GeoStyle photos densely cover several recent years with a focus on social media user photos.

Text data. To obtain information on what happened, what people discussed most, and what impacted people's daily lives the most, an ideal source is news articles. We select The New York Times to be our textual corpus. It contains news articles for the entire twentieth century, is regarded as a national "newspaper of record" based on authority and accuracy, and covers a wide range of content (*e.g.*, whereas The Wall Street Journal is business-focused).

We collect all available New York Times news articles from 1900 to now, for a total of 100M articles (details below). Being an American newspaper, its content or perspective is often about the U.S. or Western hemisphere. This reasonably matches the perspective of our collected image dataset, making it suitable for mining potential cultural factors that shaped clothing styles across the same time period. We share the collected datasets at: http://vision.cs.utexas.edu/projects/CultureClothing.

3.2. Clothing features

We first apply person detection [18] on all images to isolate the clothing people wore. This gives us in total $11,898$ clothing instances.

Next we extract clothing styles. A style representation should capture the color, pattern, cuts, *etc.*, of an outfit, while being invariant to the pose or identity of the underlying person or other irrelevant factors. Furthermore, an important trait in style evolution is that changes are often gradual and local. For example, as clothing's function became more practical, in the early 1910s it first relaxed the bust area, and then towards 1920 it raised the hemline above the calves. These factors call for more than a vanilla global image encoding. Global features from an entire outfit are often dominated by larger regions (*e.g.*, entire silhouette of a dress), which would prevent analyzing localized details (subtle patterns, sleeve type, neckline, *etc.*). Similarly, features from a neural network pre-trained on ImageNet [11] could capture an object's overall texture and shape, but are insufficient for the fine-grained and localized details in clothing (*e.g.*, necklines or hemlines).

Thus, we inject two elements to prepare the visual features to be used in style discovery. First, we fine-tune an ImageNet pre-trained ResNet-18 [23] on DeepFashion [41] with the tasks of clothing category and attribute recognition to obtain a clothing-sensitive encoder. Since this network is trained to recognize details in clothing, the styles we obtain will be finer-grained (floral A-line dress vs. A-line dress). Second, we zoom in on an outfit to its *neckline*, *sleeves*, *torso*, and *legs* regions, and analyze the evolution of clothing styles at each region separately. To automatically separate these regions, we use human body joints as anchors. Specifically, we detect joints for neck, arms, waist, hips, and ankles using Mask-RCNN [22]. Clothing-style-based features are then extracted from these localized regions using the clothing-sensitive network. Fig. 2 shows qualitative comparisons using the described two elements above. Note

(a) ImageNet [11] pre-training only (left) vs. DeepFashion [41] fine-tuning (right).

(b) Features extracted from full-body (left) vs. from zoomed-in body-part regions (right).

Figure 2: **Clothing features comparison**. The proposed feature extraction yields more fine-grained and coherent style clusters.

that the face regions in the photos are discarded; our interest is to model clothing, not identity.

3.3. Clothing style discovery

Our goal is to mine for clothing styles and track their trends through time as a function of world events. Each clothing instance I_j will thus be associated with an inferred style label s_j and a year label d_j. The year label is obtained from parsing the tag or description meta-data that comes with the image.

To mine for clothing styles, we run clustering algorithms on the features extracted in Sec. 3.2 for each body region. Since the features are already fine-tuned for clothing and localized for each region, using Euclidean distance as the similarity metric for clothing styles gives reasonable results. We use Affinity Propagation [15] to let the algorithm automatically decide the number of clusters for each body region. Each cluster represents a candidate style.

Due to the entirely automatic process of discovering clothing styles, from person detection to per-part cropping based on detected body joints, as well as the presence of low-level photo statistics orthogonal to the clothing (like scanning artifacts, evolution of photo processing technologies, *etc*.), not all clusters correspond to meaningful styles. To control this risk, we automatically measure the quality of a cluster by how well it corresponds to common [41] clothing attributes or categories. The correspondence score for a cluster c is computed as the entropy of the predicted attribute/category label distribution: $E(c) = -\Sigma_{i \in S} H(i) \log_2 H(i)$, where $H(i)$ is the aggregated output activation for the attribute/category label i across all instances in cluster c, and S is the label set from DeepFashion [41]. Lower entropy means better correspondence to some predicted attributes/categories. We adopt the clusters with entropy values less than 2 standard deviations from the mean for each body region. Tab. 1 lists the final numbers of clusters (styles), which range from 26 to 144 per body region.

To verify the quality of the discovered styles, we conduct a user study: 75% of the time, human judges find the clusters to exhibit coherent clothing styles that they can describe (see Supp. for details).

The popularity of a style i at time step t is then defined as the fraction of occurrences:

$$x_{i,t} := \frac{|\{s_{j'} \mid s_{j'} = i, d_{j'} = t\}|}{\Sigma_{j'} \mathbb{1}(d_{j'} = t)}, \tag{1}$$

and the trend for the i-th visual style over time is the sequence $x_{i,1}, \dots, x_{i,T}$, where T is the most recent time point available. Fig. 3 shows the timeline of the top style trends for the torso region. See Supp. for the other body regions.

Note that attribute and category labels accompanying each style are for interpretability only; styles themselves are discovered bottom-up from clustering on clothing-oriented features, *not* directly adopting DeepFashion [41] attributes. With clothing styles and their trends in hand, we next describe how we obtain cultural factors over time.

3.4. Cultural factor mining

To mine for the latent factors that impacted people's daily lives, we collect all news articles from 1900 to now using the New York Times (NYT) API[2], then discard the extremely short ones (number of words < 15). This gives us 100 million articles in total. See Supp. for the year distribution of news articles. We use the concatenation of the title, abstract, and first paragraph for all articles. To create the vocabulary for this corpus, we use the Natural Language Toolkit [42] to apply stemming and remove stop words.

Since most news focuses on just a few subjects, and most subjects repeatedly appear, *e.g*., baseball matches, presidential elections, *etc*., we use topic modeling, specifically Latent Dirichlet Allocation (LDA) [6], to mine for the latent factors shared among all news articles. LDA assumes that a number of latent K topics account for the distribution of observed words in a document, where each topic is a distribution over words in the vocabulary, and each document can be represented as a distribution over topics. By running LDA (with $K = 400$) over the text corpus, each article M_j is represented as its topic distribution $\boldsymbol{\theta}_j$:[3]

$$\boldsymbol{\theta}_j = [\theta_{j1}, \dots, \theta_{jK}], \tag{2}$$

where $\theta_{jk} \geq 0$, $\Sigma_k \theta_{jk} = 1$. Each article is dated with the date it was published, giving date label d_j for article M_j. The popularity of a topic l at temporal bin t can thus be computed as:

$$y_{l,t} := \frac{\sum\limits_{j:d_j=t} \theta_{jl}}{\sum\limits_{k=0}^{K-1} \sum\limits_{j:d_j=t} \theta_{jk}}, \tag{3}$$

[2] https://developer.nytimes.com/

[3] To set K, we use Google Cloud Natural Language API as reference, where they have hierarchical categories of common topics for news articles. There are 382 categories in the third hierarchy.

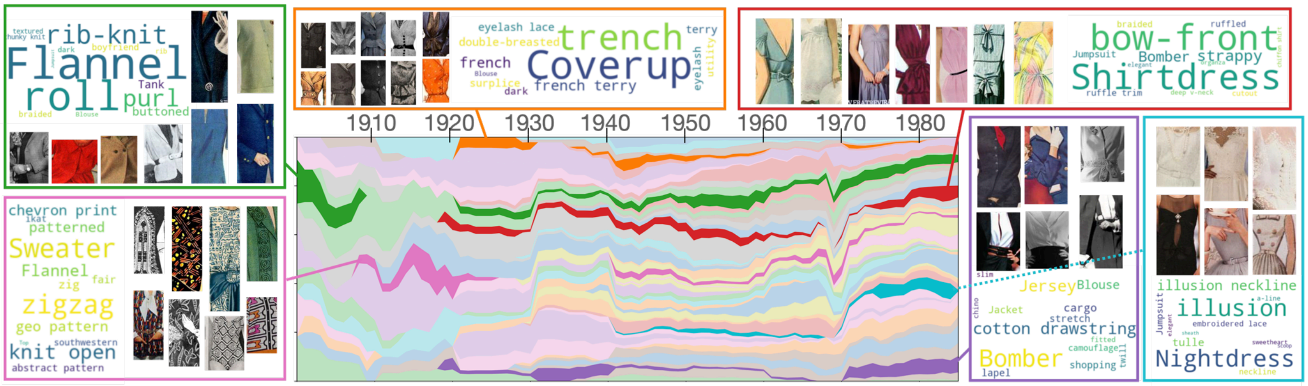


Figure 3: **Timeline of the top styles in the torso region:** Each color represents a style, while the area a style occupies shows the frequency of that style at a time delta. Some example styles are highlighted, showing their centroid images and detected visual attributes (inferred with a classifier on the clothing-sensitive encoder). Interesting trends can be observed, e.g., styles in later times expose more skin regions than in earlier times, and some distinctive busy textures peaked in the 1910s and 1920s. See Supp. for timelines for other body regions.

and the trend for the l-th cultural factor (topic) over time is the sequence $y_{l,1}, \ldots, y_{l,T}$.

3.5. Culture-fashion influence modeling

Having introduced our approach for discovering clothing styles and cultural factors, we next describe our method for modeling how culture shapes what we wear. A natural thing to do would be to find the correlation between style and topic changes: if a style and a topic have similar changes in all adjacent years, it is likely that this topic influences that style. However, this simple method fails to consider other possibilities: the trend of the style and the topic could be positively or negatively correlated, the correlation could happen at arbitrary delays, and local fluctuations in either trend could easily affect the overall correlation metric. In fact, we hypothesize that as long as observing the topic helps improve forecasting the style's trend—no matter what the trends look like—that topic may have influenced the style. This property is essentially the definition of Granger causality [19]:

Definition 1 *Granger-causality.* *A time series $\{y_{l,t}\}$ Granger-causes another time series $\{x_{i,t}\}$ if including the history of y_l improves prediction of x_i over knowledge of the history of x_i alone.*

To determine which topic(s) influences which style(s), we perform a Granger-causality test on all topic-style pairs, where the time series of topic l is $\{y_{l,t}\}$, and time series of style i is $\{x_{i,t}\}$. The test is based on the following regression model:

$$\hat{x_{i,t}} = \sum_{m=1}^{q_1} \alpha_m x_{i,t-m} + \sum_{m=1}^{q_2} \beta_m y_{l,t-m}. \tag{4}$$

Here, α_m, β_m are the regression coefficients for each time series, and q_1, q_2 are their respective regression time windows. The null hypothesis for this test is when $\beta_m = 0, \forall m \in \{1, \ldots, q_2\}$ is optimal. For those topic-style pairs where the null hypothesis is rejected at some significance level α, that topic Granger-causes the style. Fig. 4 overviews our approach.

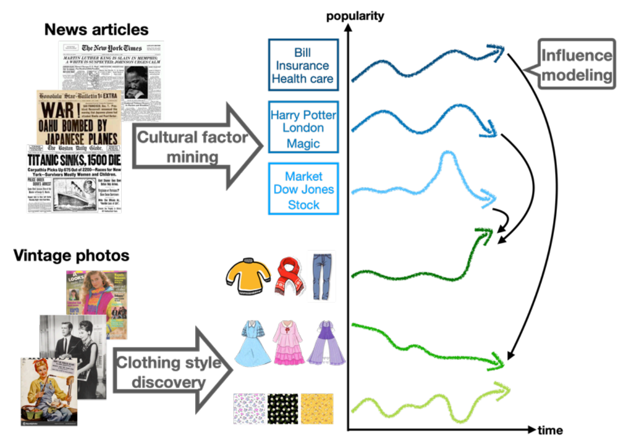


Figure 4: **Approach overview**: Cultural factors are mined from news articles using topic models (Sec. 3.4), and clothing styles are mined from photos by running clustering on clothing-sensitive features (Sec. 3.3). Cultural influences on clothing styles are detected by measuring Granger-causality relations between the respective popularity time series (Sec. 3.5).

In the following sections, we describe how this influence model can be used to automatically create fashion history timelines (Sec. 3.6), then discuss how we verify our discovered influences through 2 quantitative tasks (Sec. 3.7, 3.8.)

3.6. Automatically creating fashion timelines

There are two key factors in creating a fashion history timeline: i) identifying iconic styles in each era, and ii) explaining the social and cultural happenings behind those iconic styles. To identify iconic styles in each decade, we compute the index lift [48] for each style at each decade as $\frac{x_{i,t}}{\Sigma_t x_{i,t}}$, accounting for the uniqueness of observing that style at that time versus the overall style distribution across time. Styles with the largest lift index at time t are identified as iconic styles. To understand which happenings in the world gave rise to those styles, we analyze their Granger-causal topics: top words in a topic explain the coarser cultural factors (*e.g*., German, music, turmoil, *etc*.), and by tracing back the news articles that have the highest probability for that topic at that time, we can also detect the specific events (*e.g*., World War I, The Great Depression, campus protest and unrest, *etc*.) that took place. Fig. 1 is a snippet of a fashion history timeline created with the above procedure on the Vintage data. While some influences shown may be familiar (*e.g*., utility clothing during wartime), others (mini skirts and campus unrest) are potentially newly discovered by our model.

3.7. Influence-based style forecasting

If a topic indeed influences a style, this relation may continue to hold for future time series, thus improving the task of forecasting future style trends. To this end, we use our discovered influences in the *training set* time range to help forecast style trends in the future (disjoint) *test set* time range. Trend forecasting predicts future values for a time series based on its history; autoregressive models are typical for this task. Let C_i be the set of influential topics for style i we obtained from the Granger-causality test. To predict the future trend of clothing style i, we ensemble the predictions from all autoregressive models with the style's Granger-causal topics $l \in C_i$ as exogenous inputs:

$$\hat{x}_{i,t} = \frac{1}{|C_i|} \sum_{l \in C_i} \Big(\sum_{m=1}^{q_1} \alpha_{i,m,l} x_{i,t-m} + \sum_{m=0}^{q_2-1} \beta_{i,m,l} y_{l,t-m} \Big). \tag{5}$$

In our preliminary experiments, we have tried more complex models (*e.g*., neural-network-based), and found them to perform inferior to the simple linear-based one. Similar findings are reported in prior work in forecasting [5].

3.8. Influence-based photo timestamping

Aside from trend forecasting, we also examine how much the cultural factors from the text corpus help to date (timestamp) historic photos. The date label of a test instance's nearest neighbor in the training set is adopted as its predicted timestamp. The similarity metric for the instances can be measured using either the visual feature alone (baseline), or by additionally using the image's *inferred cultural factors* given by our model, as we describe next.

We first train a model to map an image to its inferred cultural factors, *i.e*., latent textual topics. The textual feature v_j for a training photo I_j with date label t is the averaged topic distribution over all news articles with that date label:

$$v_j = \frac{\sum_{i:d_i=t} \boldsymbol{\theta}_i}{\Sigma_i \mathbb{1}(d_i = t)}. \tag{6}$$

At training time, a mapping function (a 3 layer MLP) is learned to transform a photo's visual feature to its textual feature. At test time, given only a photo, the textual feature of a clothing instance is obtained by feeding its visual feature to the learned mapping function, and the output textual feature is used to measure similarity to the training instances. Both visual and textual similarities are measured by Euclidean distance in their respective feature spaces; the resulting similarity is the average of visual and textual distances. In this way, given a photo, we draw on the inferred cultural factors to enrich its encoding for timestamping.

4. Experiments

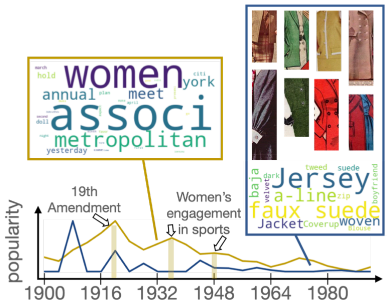
We first show cultural influences on clothing styles discovered by our model. We then evaluate the detected influences by how much they help in trend forecasting and timestamping for two image datasets.

4.1. Discovered trends and influences

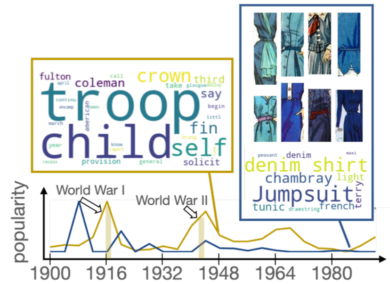
First, to study how cultural factors influence clothing styles in the long run, we use the Vintage data (cf. Sec. 3.1). The granularity of each time point t in the clothing time series $\{x_{i,t}\}$ and textual time series $\{y_{l,t}\}$ is 4 to 5 years. Influences are detected using years 1900 to 1975. The years from 1976 to 1996 are later used for evaluation on trend-forecasting. By using the later years as test data, where the samples are denser, we assure at least hundreds of test samples per time point.

Second, to study influence in modern times in the short term, we use the GeoStyle [44] image data, which contains timestamped and geotagged photos from Instagram spanning a time period from July 2013 to May 2016. Here we use the same styles computed in previous work [44, 45], which aggregates detected visual attributes (at each city) weekly to obtain style trends. This results in $2,024$ style trends, each of length 143 time points. The granularity of each time point t is 1 week for both clothing time series $\{x_{i,t}\}$ and textual time series $\{y_{l,t}\}$. For later evaluation on trend-forecasting, the last 26 points are held out (following previous work [4, 44]), and all previous points are used to detect influences.

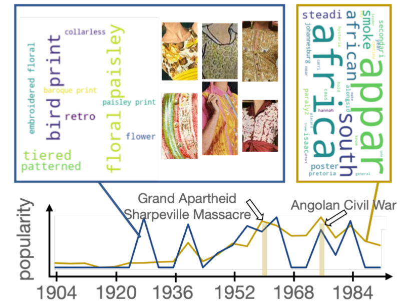
Example influences are in Fig. 1 and Fig. 5. Cultural factors known to influence clothing styles include economic status, political tension, civil rights, wars, ethnic diversity, or new technology. Fig. 5(a-d) show examples of our detected influences from the Vintage photos. Interestingly, they often agree with those reported by experts [7]. Fig. 5a is a topic about women, which has its peak at the time the



(a) Topic '*Women*' influences working attire.

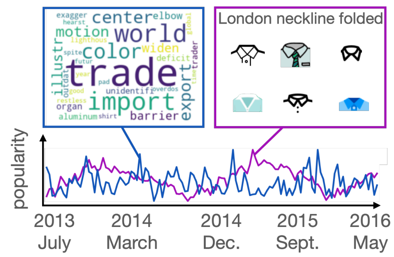


(b) Topic '*War*' influences utility attire.

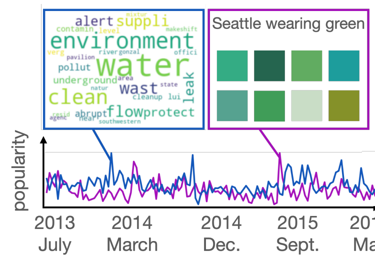


(c) Topic '*African*' influences ethnic attire.

(d) '*Invention*' influences novel clothing details.



(e) '*Finance*' influences necklines in London.



(f) '*Eco*' influences wearing green in Seattle.

Figure 5: **Example detected influences:** curves in each subfigure are popularity trends of visual styles and cultural factors. Corresponding call-out boxes for the curves show centroid images and detected attributes/categories in a style (blue boxes), and also top words in a mined textual topic (yellow boxes). Fig. (a) (b) (c) (d) show discovered influences for the Vintage dataset (1900-1996). They agree with expert knowledge [7]. Fig. (e) (f) show examples for GeoStyle [44] (2013-2016). While space permits displaying images for only a fraction of the discovered influences, our quantitative results report the results over *all* discovered influences (cf. Table 1 and 2).

19th Amendment was passed. The style influenced by this topic depicts working attire, with attributes such as blouse, suede, A-line, *etc*. The second topic in Fig. 5b is about wars, which has its peaks at World War I and World War II. This topic influenced the popularity of utility clothing, with attributes like denim, chambray, peasant, jumpsuit, *etc*. The third topic in Fig. 5c is about (South) Africa, with its peaks at the Grand Aparthied and its civil war. Ethnicity-inspired clothing with colorful prints like floral or paisley along with embroidery are influenced by this topic. Finally, the fourth topic in Fig. 5d is about patents and new inventions, with its peak during the space-race era. Clothing influenced by this topic is mostly made with new techniques, including zipper, bleaching, mineral wash, *etc*. While it is satisfying to find influences that agree with expert insights, our model can also help recover more subtle and previously unexpected influences—a strength of our data-driven approach.

Fig. 5(e-f) show examples of influences discovered by our method on GeoStyle. Our model discovers the seasonal similarity between a finance related topic and the folded necklines in London (Fig. 5e). It also discovers a possible causality relation from a environmentally-conscious topic to the style of wearing green in Seattle (Fig. 5f).

To verify the statistical significance of our results, we identify the top Granger-causality relations, and compare their F-values and F-critical values. F-values higher than F-critical values are considered statistically significant [2]. Vintage has an F-critical value 3.98, and GeoStyle 2.48. The top 20 Granger-causality relations have F-values in the range 9 to 96 and 13 to 17 on each dataset, respectively, indicating they are significant. See Supp. for full lists.

4.2. Forecasting trends

Next, we apply our detected influences on trend forecasting on the held-out time series, i.e., all of 1975-1996 for Vintage, and all of the last half of 2016 for GeoStyle. On both datasets, we evaluate all styles for which the external influences are adopted (*i.e.*, those that rejected the null hypothesis in Granger tests, see Supp.). The numbers of clothing styles are shown in Tab. 1, first row.

Baselines and evaluation. We adopt the baselines from prior work on trend forecasting [5].[4] We use mean-squared-error (MSE) as evaluation metric (on vintage) because it is standard for forecasting tasks [14, 38, 56], whereas we follow mean-absolute-error (MAE) for GeoStyle to be consistent with prior methods on this dataset [4, 44]. Given the ground-truth values from the training set, all methods predict into the future on a long-horizon basis (*i.e.*, they are never given ground-truth values from the test set). Their

[4]The method of [4] requires city labels, so is not applicable on the vintage data. We also tried the method of [44], and it fails by falling back to the linear baseline. It could not converge when its cyclic term is included due to the lack of cyclicity in this data, which that model depends on.

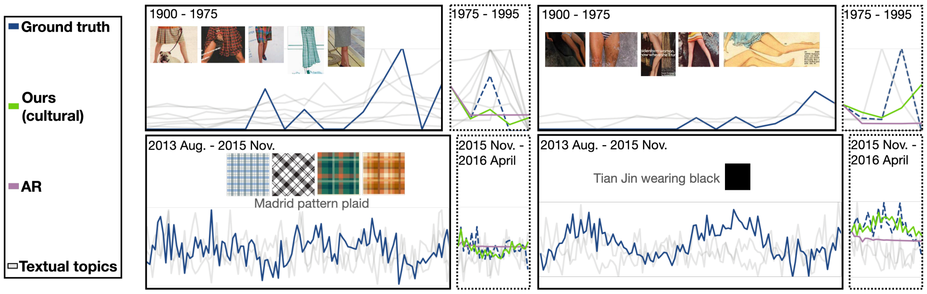


Figure 6: **Examples of trend forecasting** by vanilla AR vs. considering cultural influences (ours). Top row on the Vintage data spanning a century. Bottom row on the GeoStyle [44] data spanning 3 years. The discovered influence relations from topics to visual styles help predict more accurate trends than AR.

	Next 6 months	**Next 20 years**			
	GeoStyle [44]	neck	torso	arms	legs
styles (sig./all)	1607/2024	17/74	86/120	51/144	2/26
last	0.023	0.303	0.209	0.258	0.174
linear	0.025	0.134	0.211	0.182	0.146
mean	0.034	0.089	0.085	0.071	0.149
EXP [5]	0.022	0.134	0.146	0.124	0.122
AR	0.028	0.091	0.084	0.070	0.145
cultural (ours)	**0.019**	**0.081**	**0.083**	**0.068**	**0.088**

Table 1: **Forecasting trends** for styles in the GeoStyle [44] dataset (left) and each body region from the Vintage photos (right).

	prior	visual only	visual + cultural (ours)
Vintage	0.170	0.653	**0.695** (+6.0%)
GeoStyle [44]	0.121	0.124	**0.156** (+25.8%)

Table 2: **Timestamping accuracy**: Given a photo, including its inferred cultural features helps better predict the correct date.

performance is evaluated per style per time-point, and the final error is the average error over the next 20 years and 6 months for Vintage and GeoStyle, respectively.

Trend prediction results are presented in Tab. 1. Our method does best overall. Predicting future trends in the short-term (on GeoStyle) or the long-term (on Vintage) generally requires models with different characterstics. When predicting future trends in the short term (next 6 months), explicitly depending more on recent values (*i.e.* last and EXP) performs better than considering more historical values (*i.e.* mean and linear). Example styles from GeoStyle like those in Fig. 5 and Fig. 6 suggest that even though there are always local fluctuations, the trend of a style generally does not change drastically in a short period. Note how our influence-based model captures not only the overall future trend of a style, but also the local fluctuations (Fig. 6). Prediction in the long term (next 20 years) on Vintage is challenging. Methods that use only 1 or 2 data points (*i.e.* last or linear) to extrapolate future curves often perform poorly. As more historical points are considered (*i.e.* mean, AR, EXP), the model's performance gets better. Our influence-based model performs best by being able to predict the highly dynamic long-term future trends (as in Fig. 6).

In both the long- and short-term settings, including the proposed cultural influences in autoregression (AR) improves the overall performance. On Vintage, 57% of the styles perform better when including influences, with 8% of the styles improving more than 10%. On GeoStyle, 80% of the styles perform better when including influences, with 66% of the styles improving more than 10%.

We stress that these experiments are comprehensive over all styles. While space allows showing qualitative figures for only a sample of discovered influence relationships, the quantitative results (Tab. 1, Tab. 2) are computed over *all* the data. Furthermore, we find the accuracy of our discovered influences is important: if we incorporate all topics, not just the Granger-causal ones, forecasting is 30% worse than vanilla AR.

4.3. Timestamping photos

Finally, we evaluate our method's impact on timestamping unseen photos (cf. Sec 3.8). We randomly split both the Vintage and the GeoStyle[5] datasets to hold out 20% of the clothing instances for evaluation, and use the rest as the training database for retrieval. The set of date labels for Vintage data is every 5-th year, 1900, 1905, 1910, . . . , 1995, totaling 20 labels. The set of labels for GeoStyle is every 4-th month, from July 2013 to May 2016, totaling 10 labels. We evaluate multi-class classification accuracy.

Tab. 2 shows the timestamping results. Our approach outperforms the visual-only baseline significantly. The cultural features offer a better representation for timestamping than the visual features alone, likely because they compress date-specific information in a cleaner manner.

5. Conclusion

Analyzing a century of fashion photos, we explored how world events may have affected the clothes people choose to wear. Our statistical model identifies concrete temporal influence relationships between news events and visual styles, allowing both well-known and more subtle ties to be surfaced in a data-driven manner. To demonstrate the real-world impact, we proposed methods for both forecasting and timestamping that leverage the mined influential cultural factors. Our results on two datasets show how this novel source of context benefits both practical tasks. In future work, we plan to investigate hierarchical models of influence and explore the role of geographic patterns.

[5]Full image data is not publicly available. We use $\sim$ 130 K images.

References

[1] Afro Hairstyle. https://en.wikipedia.org/wiki/Afro, 2021. [Online; accessed 16-August-2021]. 1

[2] F-test Interpretation. https://en.wikipedia.org/wiki/F-test, 2021. [Online; accessed 4-March-2021]. 7

[3] Jacqueline Kennedy Onassis. https://en.wikipedia.org/wiki/Jacqueline_Kennedy_Onassis, 2021. [Online; accessed 16-August-2021]. 1

[4] Ziad Al-Halah and Kristen Grauman. From Paris to Berlin: Discovering fashion style influences around the world. In *CVPR*, 2020. 2, 6, 7

[5] Z. Al-Halah, R. Stiefelhagen, and K. Grauman. Fashion forward: Forecasting visual style in fashion. In *ICCV*, 2017. 2, 6, 7, 8

[6] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *JMLR*, 2003. 4

[7] Susan Brown. Fashion: The definitive history of costume and style. 2012. 1, 6, 7

[8] KuanTing Chen, Kezhen Chen, Peizhong Cong, Winston H Hsu, and Jiebo Luo. Who are the devils wearing prada in new york city? In *ACM MM*, 2015. 2

[9] Qiang Chen, Junshi Huang, Rogerio Feris, Lisa M Brown, Jian Dong, and Shuicheng Yan. Deep domain adaptation for describing people based on fine-grained clothing attributes. In *CVPR*, 2015. 2

[10] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR*, 2005. 2

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009. 3, 4

[12] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros. What makes paris look like paris? *ACM SIGGRAPH*, 2012. 2

[13] Grace Evans. Titanic style: Dress and fashion on the voyage. 2012. 1

[14] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. Temporal relational ranking for stock prediction. *ACM TOIS*, 2019. 7

[15] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 2007. 4

[16] Cheng-Yang Fu, Tamara L. Berg, and Alexander C. Berg. Imp: Instance mask projection for high accuracy semantic segmentation of things. In *ICCV*, 2019. 2

[17] 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 *IEEE Transactions on Computational Imaging*, 2017. 2

[18] Ross Girshick, Ilija Radosavovic, Georgia Gkioxari, Piotr Dollár, and Kaiming He. Detectron. https://github.com/facebookresearch/detectron, 2018. 3

[19] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 1969. 5

[20] X. Guo, H. Wu, Y. Cheng, S. Rennie, and R. Feris. Dialog-based interactive image retrieval. In *NIPS*, 2018. 2

[21] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S. Davis. Learning fashion compatibility with bidirectional lstms. In *ACM MM*, 2017. 2

[22] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *ICCV*, 2017. 3

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 3

[24] Shintami Chusnul Hidayati, Cheng-Chun Hsu, Yu-Ting Chang, Kai-Lung Hua, Jianlong Fu, and Wen-Huang Cheng. What dress fits me best?: Fashion recommendation on the clothing style for personal body shape. In *ACM MM*, 2018. 2

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

[26] Wei-Lin Hsiao and Kristen Grauman. Creating capsule wardrobes from fashion images. In *CVPR*, 2018. 2

[27] Wei-Lin Hsiao and Kristen Grauman. Vibe: Dressing for diverse body shapes. In *CVPR*, 2020. 2

[28] Wei-Lin Hsiao, Isay Katsman, Chao-Yuan Wu, Devi Parikh, and Kristen Grauman. Fashion++: Minimal edits for outfit improvement. In *ICCV*, 2019. 2

[29] Junshi Huang, Rogerio S Feris, Qiang Chen, and Shuicheng Yan. Cross-domain image retrieval with a dual attribute-aware ranking network. In *ICCV*, 2015. 2

[30] C. Huynh, A. Ciptadi, A. Tyagi, and A. Agrawal. Craft: Complementary recommendation by adversarial feature transform. In *ECCV Workshop on Computer Vision For Fashion, Art and Design*, 2018. 2

[31] Yong Jae Lee, Alexei A Efros, and Martial Hebert. Style-aware mid-level representation for discovering visual connections in space and time. In *ICCV*, 2013. 2

[32] Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui, Claire Cardie, Bharath Hariharan, Hartwig Adam, and Serge Belongie. Fashionpedia: Ontology, segmentation, and an attribute localization dataset. In *ECCV*, 2020. 2

[33] Hirokatsu Kataoka, Yutaka Satoh, Kaori Abe, Munetaka Minoguchi, and Akio Nakamura. Ten-million-order human database for world-wide fashion culture analysis. In *CVPR Workshop on FFSS-USAD*, 2019. 2

[34] M. Hadi Kiapour, X. Han, and S. Lazebnik. Where to buy it: Matching street clothing photos in online shops. In *ICCV*, 2015. 2

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

[36] A. Kovashka, D. Parikh, and K. Grauman. WhittleSearch: Interactive image search with relative attribute feedback. *IJCV*, 2015. 2

[37] Zhanghui Kuang, Yiming Gao, Guanbin Li, Ping Luo, Yimin Chen, Liang Lin, and Wayne Zhang. Fashion retrieval via graph reasoning networks on a similarity pyramid. *ICCV*, 2019. 2

[38] Qing Li, Yuanzhu Chen, Li Ling Jiang, Ping Li, and Hsinchun Chen. A tensor-based information framework for predicting the stock market. *ACM TOIS*, 2016. 7

[39] Xiaodan Liang, Si Liu, Xiaohui Shen, Jianchao Yang, Luoqi Liu, Jian Dong, Liang Lin, and Shuicheng Yan. Deep human parsing with active template regression. *TPAMI*, 2015. 2

[40] Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *CVPR*, 2012. 2

[41] 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, 3, 4

[42] Edward Loper and Steven Bird. Nltk: The natural language toolkit. In *ACL Workshop*, 2002. 4

[43] Yunshan Ma, Yujuan Ding, Xun Yang, Lizi Liao, Wai Keung Wong, and Tat-Seng Chua. Knowledge enhanced neural fashion trend forecasting. In *ICMR*, 2020. 2

[44] Utkarsh Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. GeoStyle: Discovering fashion trends and events. In *ICCV*, 2019. 2, 3, 6, 7, 8

[45] K. Matzen, K. Bala, and N. Snavely. Streetstyle: Exploring world-wide clothing styles from millions of photos. *arXiv:1706.01869*, 2017. 2, 6

[46] Christian A Peterson. Home portraiture. *History of photography*, 2011. 3

[47] Yong-Siang Shih, Kai-Yueh Chang, Hsuan-Tien Lin, and Min Sun. Compatibility family learning for item recommendation and generation. In *AAAI*, 2018. 2

[48] Craig Silverstein, Sergey Brin, and Rajeev Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data mining and knowledge discovery*, 1998. 5

[49] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. Neuroaesthetics in Fashion: Modeling the Perception of Fashionability. In *CVPR*, 2015. 2

[50] Moeko Takagi, Edgar Simo-Serra, Satoshi Iizuka, and Hiroshi Ishikawa. What makes a style: Experimental analysis of fashion prediction. In *ICCV Workshop on Computer Vision For Fashion*, 2017. 2

[51] Mariya I. Vasileva, Bryan A. Plummer, Krishna Dusad, Shreya Rajpal, Ranjitha Kumar, and David Forsyth. Learning type-aware embeddings for fashion compatibility. In *ECCV*, 2018. 2

[52] Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, and Serge Belongie. Learning visual clothing style with heterogeneous dyadic co-occurrences. In *ICCV*, 2015. 2

[53] Sirion Vittayakorn, Alexander C Berg, and Tamara L Berg. When was that made? In *WACV*, 2017. 2

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

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

[56] Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. Stock price prediction via discovering multi-frequency trading patterns. In *ACM SIGKDD*, 2017. 7

[57] B. Zhao, J. Feng, X. Wu, and S. Yan. Memory-augmented attribute manipulation networks for interaction fashion search. In *CVPR*, 2017. 2

[58] Shuai Zheng, Fan Yang, M Hadi Kiapour, and Robinson Piramuthu. Modanet: A large-scale street fashion dataset with polygon annotations. In *ACM MM*, 2018. 2