Instagrammers, Fashionistas, and Me: Recurrent Fashion Recommendation with Implicit Visual Influence

Yin Zhang and James Caverlee
Department of Computer Science and Engineering, Texas A&M University
zhan13679@tamu.edu,caverlee@cse.tamu.edu

ABSTRACT

Fashion-focused key opinion bloggers on Instagram, Facebook, and other social media platforms are fast becoming critical influencers. They can inspire consumer clothing purchases by linking high fashion visual evolution with daily street style. In this paper, we build the *first* visual influence-aware fashion recommender (FIRN) with leveraging fashion bloggers and their dynamic visual posts. Specifically, we extract the *dynamic fashion features* highlighted by these bloggers via a BiLSTM that integrates a large corpus of visual posts and community influence. We then learn the *implicit visual influence funnel* from bloggers to individual users via a personalized attention layer. Finally, we incorporate user personal style and her preferred fashion features across time in a recurrent recommendation network for dynamic fashion-updated clothing recommendation. Experiments show that FIRN outperforms state-of-the-art fashion recommenders, especially for users who are most impacted by fashion influencers, and utilizing fashion bloggers can bring greater improvements in recommendation compared with using other potential sources of visual information. We also release a large *time-aware high-quality visual* dataset of fashion influencers that can be exploited for future research.

CCS CONCEPTS

• **Information systems** → *Social recommendation.*

KEYWORDS

Fashion recommendation; Instagrammers; Bloggers; Recurrent neural network

ACM Reference Format:
Yin Zhang and James Caverlee. 2019. Instagrammers, Fashionistas, and Me: Recurrent Fashion Recommendation with Implicit Visual Influence. In *The 28th ACM International Conference on Information and Knowledge Management (CIKM '19), November 3–7, 2019, Beijing, China.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3357384.3358042

1 INTRODUCTION

Opinion leaders can impact consumer purchase and consumption behaviors in a variety of different markets [7, 27, 31]. Among these, *fashion* opinion leaders wield outsize influence on fashion trends [26, 29, 32, 38]. And with the rise of visual media platforms like Instagram and Pinterest, influential fashion leaders are not just celebrities and famous designers, but also *fashion bloggers* who have built a name and reputation within the platform itself [6, 23, 25]. These fashion bloggers link high fashion with daily wear through their appealing posts. For example, many bloggers attend high profile fashion shows, such as New York Fashion Week, to keep up with the frontiers of fashion trends (like dress design and fashionable colors) [25]. At the same time, connecting these fashion trends with our daily clothing choices through visual social media, fashion bloggers can directly disseminate their fashion choices to consumers, as illustrated in Figure 1. Typical example posts on Instagram include "#10 pieces every woman should have in her wardrobe", "#OOTD" (outfit of the day), and "#Top Trends of the season", which are very useful for clothing choices.

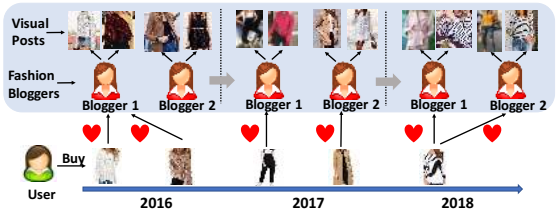

Figure 1: Fashion bloggers and their implicit influence funnel. The top two rows show bloggers and their posts. The bottom row shows a user and her purchases.

Since those influencers can play a significant role in fashion adoption [32] and consumer aesthetic evaluation is largely based on current fashion trends [8, 9, 35], we explore in this paper the potential of enhancing fashion recommendation by carefully modeling the visual influence of these fashion bloggers. We collect more than 130,000 Instagram posts by influential female fashion bloggers, and connect this visual style to Amazon item purchases over time. This recommender can extract the current hottest fashion clothing based on a user's visual taste, as well as capture trends reflected in the choices of these fashion bloggers. While incorporating influencers into recommendation has great potential value, there are a number of key challenges.

First, the fashion tastes on platforms like Instagram is diffused across millions of posts, and these tastes vary across bloggers. Moreover, their styles are always in flux, since fashion bloggers adapt to new trends. How can we extract each fashion blogger's unique dynamic fashion features from a large corpora of posts? Second, in practice, it is extremely difficult to directly capture the explicit connections/influence from a fashion blogger to a user's purchase [27]. This influence can also be complicated: users can be directly influenced by a blogger's posts or indirectly influenced by the blogger through their friends or communities. Besides, each user's visual preference is personal and some users may be strongly influenced

CIKM '19, November 3–7, 2019, Beijing, China

ACM ISBN 978-1-4503-6976-3/19/11…$15.00
https://doi.org/10.1145/3357384.3358042

by fashion bloggers while others may not be at all. Hence, how can we learn such personal *implicit visual influence* funnel from fashion bloggers to users for fashion recommendation? Third, the visual influence from bloggers to users could change over time, as an example shown in Figure 1. How can we effectively *learn these temporal dynamics* for visual influence-aware fashion recommendation?

In this paper, our main goal is to address these three challenges to build a personalized visual influence-aware fashion recommender that can learn both fashion trends and user visual preference evolution across time. Specifically, we propose a Fashion visual Influence-aware Recurrent Network (FIRN) that is characterized by three unique features:

- FIRN uncovers fashion features in each time period through a bidirectional LSTM that captures each fashion blogger's style over time as well as the trends in the overall fashion community;
- FIRN naturally models each user's personal visual taste towards these fashion features by learning an implicit visual influence funnel from the extracted fashion features to individual users;
- FIRN builds a novel visual influence-aware recurrent neural network that effectively models temporal dynamics of fashion features from bloggers, users, and their visual preferences.

To our knowledge, this is *the first work to leverage influential fashion bloggers and their visual posts as a dynamic visual signal for user clothing recommendation.* Through experiments over bloggers sampled from Instagram and purchases on Amazon, we quantitatively and qualitatively evaluate the performance of FIRN. We find that FIRN significantly outperforms the state-of-the-art fashion recommendation FSVD [11] by 8.38% on average in RMSE with an even greater improvement (14.05%) for users who have previously consistently purchased items that are similar to posts by fashion bloggers. Furthermore, compared with using other potential sources of visual fashion influence – i.e. the images of a user's previous purchases [24] and a dataset of static aesthetic images (AVA) [28] – fashion bloggers can bring larger improvements in recommendation. These results further confirm that fashion bloggers can provide strong fashion visual signals across time and important dynamic influence towards user clothing purchase decisions.

2 RELATED WORK

Fashion Recommendation. With the rapid expansion of online shopping for fashion, recommending personalized fashion items has gained increasing attention [2, 8, 11, 13, 15, 17]. Different from traditional item-based recommendation, visual information plays a significant role in fashion recommendation [11, 40]. For example, He *et al.* [11] extracted fashion trends from user's purchase history and built a visual-time aware matrix factorization to recommend clothing. Jagadeesh *et al.* [16] built a visual-aware complementary recommender to find items of a similar style based on the user's purchase history. Recently, Yu *et al.* [40] used aesthetic visual features extracted from the AVA activities dataset [28] to improve Amazon clothing recommendation. Gabale *et al.* [8] explored community influence on fashion trends and identified the importance of social media on fashion evolution. Our work exploits trends revealed through fashion bloggers, in contrast to most existing approaches that use purchase history [11, 16] or static visual datasets [40].

Figure 2: An example of a post by a fashion blogger on Instagram and users comments.

Time-Aware Recommendation. Since fashion evolves over time, time-aware recommendation can be used to model user and item temporal dynamics [1, 19, 34, 39]. Considerable prior works focus on RNN-based models in these cases. For example, Wu *et al.* [39] built a recurrent recommender network that can achieve high performance with fewer parameters for rating recommendation. Beutel *et al.* [1] used a latent cross recurrent neural model to effectively model contextual information in neural recommender systems. Hidasi *et al.* [12] used a session-based RNN recommender to achieve an improvement for implicit recommendation. Ko *et al.* [18] proposed a collaborative sequence model based on RNNs to capture a user's contextual state as a personalized hidden vector. Sun *et al.* analyzed the importance of user social dynamic influence and built a recurrent recommender with utilizing user explicit static social network. Different from these works, considering the significant importance of key fashion opinion leaders (fashion bloggers) in the fashion area [6, 25], we focus on fashion bloggers and use the implicit influence of their visual posts as a dynamic fashion signal for user clothing recommendation.

Fashion Bloggers and Instagram. Many previous research has shown that fashion bloggers can influence fashion preferences, and even directly influence user purchase preference, especially for young women [6, 23, 25]. For example, Vineyard [36] examined the relations between fashion bloggers and consumer purchase (e.g. "I buy one or more products which I have browsed on a blog") and the results show they are strongly positively connected (Cronbach's $\alpha = 0.931$). Zain [41] interviewed consumers who had commented on fashion blogs, finding that their purchase preferences are strongly influenced by fashion bloggers and their posts. Marwick [23] interviewed fashion bloggers to show the high aesthetic quality of their posts and their commercial value. McQuarrie *et al.* [25] highlighted the influence of fashion bloggers on consumption. Among those work, Instagram is regarded as the platform with the largest number of influential fashion bloggers with a large reach [4]. Many brands specifically utilize Instagram to promote their clothing [33], with around £1 billion spent per year to sponsor Instagram posts [6]. Figure 2 shows an example of one of our crawled fashion blogger's visual posts and the corresponding comments by others. We see many commenters show their strong willingness to buy similar clothing as the blogger posted.

3 VISUAL INFLUENCE-AWARE FASHION RECOMMENDATION

Inspired by these works of fashion bloggers and observations, we explore in this paper the potential of integrating fashion bloggers for user clothing recommendation.

Table 1: Notations.

Notation	Explanation
$\mathcal{U}, \mathcal{P}$	user set, item set
$r_{u,p}(t)$	ratings of user u to item p at time t
$\boldsymbol{\theta}_u(t)$	visual influence-aware hidden state of user u at time t
$\boldsymbol{\theta}_p(t)$	hidden state of item p at time t
$\mathbf{m}(p)$	image vector of item p
$P_u(t)$	users bought item at time t
Π	bloggers set
$\mathcal{B}_k$	blogger k, $\Pi = \{\mathcal{B}_1, \mathcal{B}_2, \dots\}$
$\mathbf{m}_i(\mathcal{B}_k\|t)$	image vector of i_{th} post for blogger k at time t
$\mathcal{I}(\mathcal{B}_k\|t)$	post set of blogger k at time t
$\mathbf{v}(\mathcal{B}_k\|t)$	visual vector delivered by blogger k at time t

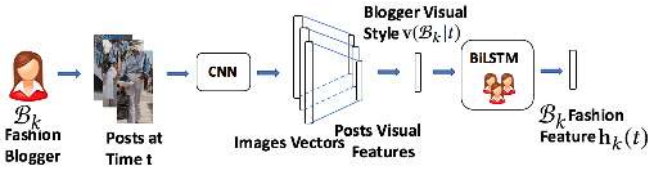

Figure 3: Extracting fashion features $\mathbf{h}_k(t)$ at time t.

Problem Statement. Formally, we assume a set of users $\mathcal{U}$, a set of fashion items $\mathcal{P}$, and their ratings in time period T. Specifically, $r_{u,p}(t)$ is the rating that user $u \in \mathcal{U}$ rates $p \in \mathcal{P}$ at time $t \in T$. We further assume a set of fashion influential bloggers $\Pi = \{\mathcal{B}_1, \mathcal{B}_2, \dots.\}$ and a set of their visual posts $\mathcal{I}(\mathcal{B}_k|t)$ for each blogger $\mathcal{B}_k$, which contains fashion features at time t. Notice here that $\mathcal{U}$ and Π are two different groups of people. By leveraging visual posts in $\mathcal{I}(\mathcal{B}_k|t)$ for each $\mathcal{B}_k \in \Pi$, we aim to recommend for each user $u \in \mathcal{U}$ a visual influence-aware and time-dependant list of items from the set $\mathcal{P}$ that considers both user visual preference and fashion features. Notations are summarized in Table 1.

In the following three sections, we present the design of our proposed visual influence-aware recurrent fashion recommendation FIRN in detail.

3.1 Extracting Fashion Features

We begin by extracting *fashion features* $\mathbf{h}_k(t)$ from each blogger $\mathcal{B}_k$. These fashion features represent the blogger's personal preferred fashion style smoothed by the common popular fashion trends in the overall fashion community, as shown in Figure 3.

Individual Visual Style. We first represent each blogger's individual visual style by a vector $\mathbf{v}(\mathcal{B}_k|t)$ derived from their visual posts. Since raw image vectors are noisy and low-level representations [11, 24], we use an embedding to obtain high-level visual features of each post. Specifically, a post's visual features $\mathbf{v}_i(\mathcal{B}_k|t)$ from blogger $\mathcal{B}_k$ is represented by:

$$\mathbf{v}_i(\mathcal{B}_k|t) = \mathbf{E}_m \mathbf{m}_i(\mathcal{B}_k|t), \tag{1}$$

where $\mathbf{E}_m \in R^{K_v \times K_m}$ is the embedding matrix. K_v is the dimension of the embedded visual features $\mathbf{v}_i(\mathcal{B}_k|t)$ and $K_v < K_m$.

Then based on $\mathbf{v}_i(\mathcal{B}_k|t)$, similar to [16, 22], we define the fashion blogger individual-level visual style $\mathbf{v}(\mathcal{B}_k|t)$ at time t as:

$$\mathbf{v}(\mathcal{B}_k|t) = \frac{\sum_{i \in \mathcal{I}(\mathcal{B}_k|t)} \mathbf{v}_i(\mathcal{B}_k|t)}{|\mathcal{I}(\mathcal{B}_k|t)|}. \tag{2}$$

Since we observe that posts can be highly visually similar in a short time period (such as one month), we adopt an average here to convey these similar visual features. In this way, the dynamic visual vector $\mathbf{v}(\mathcal{B}_k|t)$ can strengthen the common visual features that the blogger wants to deliver in the time t across her posts. Furthermore, this approach can effectively deal with the distinct different number of posts from bloggers in different time periods.

Incorporating Community Trends. This individual-level blogger visual style vector $\mathbf{v}(\mathcal{B}_k|t)$ is only a partial view of the current fashion features and maybe is noisy for fashion, since the vector is barely based on the current posts of a blogger. In practice, for any time period, a blogger may deliver some visual features that are not closely connected to current fashion trends but only some randomly posts, which could have an influence on the final fashion recommendation. However, the overall fashion community may adopt certain fashion trends that can help re-inforce which aspects of $\mathbf{v}(\mathcal{B}_k|t)$ are representative of fashion features, rather than quirks of this particular collection of posts.

Hence, we propose to smooth the individual blogger style vector $\mathbf{v}(\mathcal{B}_k|t)$ with this community influence to arrive at our goal of *fashion features* $\mathbf{h}_k(t)$. Considering other fashion bloggers can directly (or indirectly) connect to each other by fashion, we model this flow of fashion ideas at time t through a Bidirectional Long Short-Term Memory (BiLSTM) [10] among bloggers.[1]

Specifically, our blogger BiLSTM is based on the traditional LSTM [10, 14] which has been widely adopted, to capture visual features among bloggers. Formally, we first sort bloggers by the average number of likes of each post to general modulate the flow of fashion information among bloggers. Since the LSTM contains the gating units to bridge very long lags and effectively utilize input information, it is able to capture different blogger visual features to re-inforce the fashion features for each blogger. Concretely, the fashion features across bloggers at time t is built by:

$$\begin{aligned}
[\mathbf{g}_k(t), \mathbf{i}_k(t), \mathbf{o}_k(t)] &= sigmoid(\mathbf{W}[\mathbf{h}'_{k-1}(t), \mathbf{v}(\mathcal{B}_k|t)] + \mathbf{b}), \\
\mathbf{q}_k(t) &= tanh(\mathbf{W}'[\mathbf{h}_{k-1}(t), \mathbf{v}(\mathcal{B}_k|t)] + \mathbf{b}'), \\
\mathbf{c}_k(t) &= \mathbf{g}_k(t) \circ \mathbf{c}_{k-1}(t) + \mathbf{i}_k(t) \circ \mathbf{q}_k(t), \\
\mathbf{h}'_k(t) &= \mathbf{o}_k(t) \circ tanh(\mathbf{c}_k(t)),
\end{aligned} \tag{3}$$

where the input gate $\mathbf{i}_k$, output gate $\mathbf{o}_k$ and forget gate $\mathbf{g}_k$ are used to control how each fashion blogger influences other bloggers. $\circ$ denotes the element-wise product. For simplicity, we use $\mathbf{h}'_k(t) = LSTM(\mathbf{h}'_{k-1}(t), \mathbf{v}(\mathcal{B}_k|t))$ to denote these operations.

Based on Equation 3, the activations of the forward LSTM and backward LSTM for influence $\mathcal{B}_k$ is denoted as $\overrightarrow{\mathbf{h}}_k(t)$ and $\overleftarrow{\mathbf{h}}_k(t)$ which represent the fashion flow from other bloggers to blogger $\mathcal{B}_k$. So the final fashion feature based on blogger $\mathcal{B}_k$ at t is denoted as $\mathbf{h}_k(t) = [\overrightarrow{\mathbf{h}}_k(t), \overleftarrow{\mathbf{h}}_k(t)]$, which considers both a blogger's individual visual posts and their community interactions. This formulation has the benefit of smoothing each blogger's fashion features with the overall trends in the community, so that the extracted fashion

[1] Compared with directly using $\mathbf{v}(\mathcal{B}_k|t)$, experiments in Section 4 further show that the BiLSTM can achieve higher accuracy and also has a good efficiency. Furthermore, note that LSTM is not the only choice. For example, the bidirectional gated recurrent unit (BiGRU) can also be used. We use LSTMs here since they are slightly more general as [39] indicated.

features from each blogger are more representative and with more emphasis on common popular fashion trends.

3.2 Implicit Personal Visual Funnel

Given these fashion features $h_k(t)$, how can we model the influence from $h_k(t)$ of bloggers to each user u? In practice, it is hard (if not impossible) to get the explicit mapping from fashion bloggers to users and their purchases in many situations considering privacy constraints (e.g. from Instagram posts to Amazon purchases). Recently, Mukherjee *et al.* [27] used review data to build a user social latent influence without requiring explicit social network and showed great improvements in recommendation. Inspired by their success, we aim to leverage the *visual* signals from bloggers to reveal the implicit influence funnel from fashion bloggers to users. We also consider that the influence from extracted fashion features to users may be personalized in both visual aspects and degrees. For example, users who like brighter colors would prefer bloggers who post fashionable clothes in similar colors. And this influence could be highly-receptive for some users or not at all for others. Hence, we propose to model this heterogeneity in influence of user preference towards each blogger through a visual personalized attention layer. Concretely, for each user, we hypothesize that if the user's purchased products are visually similar to fashion features across time, the user is more likely to be influenced by the fashion features. Thus, we use *the attention weights [3] capture the fashion aspects that a user prefers and the visual distance models how deep the user is influenced by the extracted fashion features.*

Specifically, given the learned blogger fashion style vectors $(h_1(t), h_2(t), ..., h_K(t))$ at time t, the attention module first transforms each blogger's learned fashion vector through a single perceptron to a lower space:

$$\mathbf{s}_k(t) = sigmoid(\mathbf{E}_s\mathbf{h}_k(t) + \mathbf{b}_s), \tag{4}$$

where $\mathbf{E}_s$ and $\mathbf{b}_s$ is the corresponding embedding matrix and bias. Then the attention module compares the similarities between a user u latent influenced visual vector $\mathbf{w}(u)$ and the blogger $\mathcal{B}_k$'s transformed style $\mathbf{s}_k(t)$ by computing the dot products. So the attention weights $\alpha_k(u,t)$ of user u to blogger $\mathcal{B}_k$ is calculated by the softmax of the similarity:

$$\alpha_k(u,t) = \frac{exp(\mathbf{w}(u)^T\mathbf{s}_k(t))}{\sum_{k'}(exp(\mathbf{w}(u)^T\mathbf{s}_{k'}(t)))}. \tag{5}$$

Then the attention module computes each user's *influence-aware visual style* vector $\hat{\mathbf{h}}(u,t)$ as the weighted sum of the blogger's fashion style:

$$\hat{\mathbf{h}}(u,t) = \sum_k \alpha_k(u,t)\mathbf{h}_k(t), \tag{6}$$

where the user u latent influenced visual vector $\mathbf{w}(u)$ is a trained parameter to minimize the distance between user's influence-aware visual style and user's previously purchased items by:

$$min \sum_t \sum_u ||\mathbf{v}(u,t) - \hat{\mathbf{h}}(u,t)||_F, \tag{7}$$

where $\mathbf{v}(u,t)$ is the user visual vector based on the visual features of the user's purchase history. Specifically, for unity and to capture the same visual features, $\mathbf{v}(u,t)$ is calculated by

$$\mathbf{v}(p) = \mathbf{E}_m\mathbf{m}(p), \quad \mathbf{v}(u,t) = \frac{\sum_{p\in P_u(t)}\mathbf{v}(p)}{|P_u(t)|},$$

which is similar to the blogger's visual vector calculation (Equation 2). $\mathbf{E}_m$ is the same embedding in Equation 1. As a result, each user u has a learned influence-aware visual style $\hat{\mathbf{h}}(u,t)$ that captures the user *personal* preferred visual fashion features at time t. In the next section, we discuss how to integrate this personal influence into time-dependent fashion recommendation.

3.3 Visual Influence-aware Recurrent Network

Fashion acts as a strong influence factor for user purchase preference across different time periods. In addition to the fashion influence previously described, other static and dynamic factors may also impact user purchase preferences. Examples include the brand of an item (a static feature), the personal taste drift in clothing materials (a dynamic factor), and so on. In this section, we incorporate these additional factors with the extracted personal fashion influence to build a joint visual influence-aware recurrent network for user clothing recommendation.

To capture these dynamic and stationary states, especially the visual influence from fashion bloggers, we extend recurrent recommendation networks (RRN) [39] with visual influence-aware personalized fashion factors. Specifically, an RRN can capture temporal dependencies for both users and items with a dynamic user state and a dynamic item state. Besides modeling user internal dynamic and stationary states, our proposed FIRN also considers the external fashion drift and its influence for users. Concretely, for the user state, suppose $\mathbf{r}_u(t)$ is the rating vector for user u. That is, $r_{u,p}(t) = r$ is the user rates item p with score r at time t otherwise $r_{u,p}(t) = 0$. We denote τ_t as the wallclock at time step t and $\mathbf{1}_{newbie}$ as the indicator of whether the user is new. So the constructed input for user at time t in FIRN is [39]:

$$\mathbf{x}_u(t) := [\mathbf{r}_u(t), \mathbf{1}_{newbie}, \tau_t, \tau_{t-1}]. \tag{8}$$

Then the personalized fashion-aware user vector is:

$$\mathbf{f}_u(t) = \mathbf{E}_u\mathbf{x}_u(t) + \mathbf{E}_f\hat{\mathbf{h}}(u,t), \tag{9}$$

where $\mathbf{E}_u$ and $\mathbf{E}_f$ are transformations to be learned to project source and fashion information into the joint user embedding space. Specifically, $\mathbf{E}_u x_u(t)$ represents the latent factor of user personal preference and $\mathbf{E}_f\hat{\mathbf{h}}(u,t)$ is the extracted popular fashion features that the user prefers at time t. The state of u at time t is decided by the user's previous hidden state $\boldsymbol{\theta}_u(t-1)$ and $\mathbf{f}_u(t)$. So:

$$\boldsymbol{\theta}_u(t) := LSTM(\boldsymbol{\theta}_u(t-1), \mathbf{f}_u(t)). \tag{10}$$

For an item's time dependent state, similarly, the item vector are calculated by $\mathbf{f}_p(t) = \mathbf{E}_p\mathbf{x}_p(t)$ where $\mathbf{x}_p(t) := [\mathbf{r}_p(t), \mathbf{1}_{newbie}, \tau_t, \tau_{t-1}]$ is the constructed item input. The overall framework for FIRN is shown in Figure 4.

Rating Prediction. Besides dynamic states, users/items also contain stationary components across time. So we incorporate the time-dependent user states $u(t)$ and item states $p(t)$ with the stationary state $\tilde{\gamma}_u$ and $\tilde{\gamma}_p$ respectively. Similar to [39], the rating prediction is calculated by:

$$\hat{r}_{u,p}(t) :=< \boldsymbol{\theta}_u(t), \boldsymbol{\theta}_p(t) > + < \tilde{\boldsymbol{\gamma}}_u, \tilde{\boldsymbol{\gamma}}_p > . \tag{11}$$

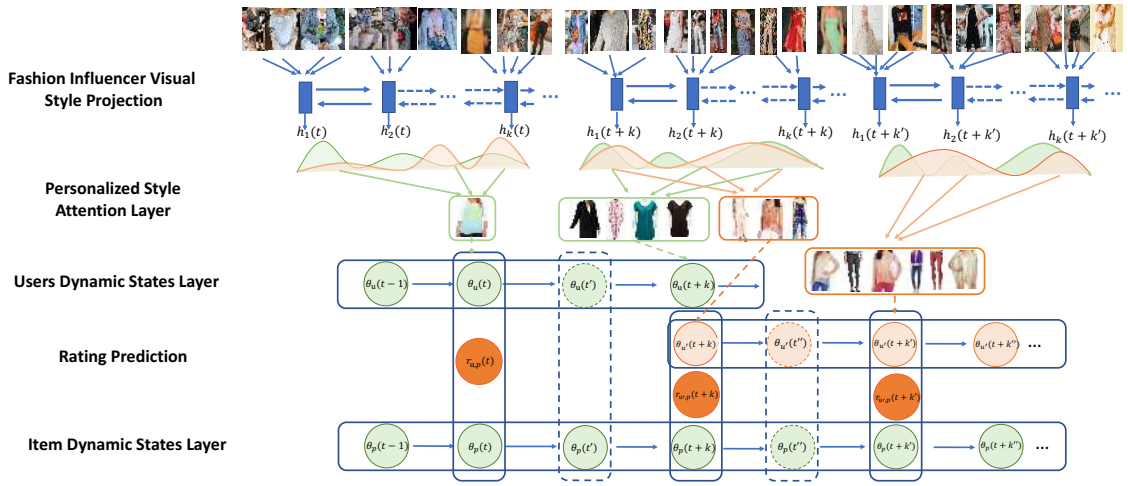

Figure 4: Visual Influence-Aware Recurrent Fashion Recommendation Framework. The vertical solid/dashed boxes represent the user dynamic states with/without ratings and the item states at the same time.

The $\tilde{\gamma}_u$ and $\tilde{\gamma}_p$ are affine functions of γ_u and γ_p which are $\tilde{\boldsymbol{\gamma}}_u = \mathbf{E}'_u\boldsymbol{\gamma}_u + \mathbf{b}_u, \tilde{\boldsymbol{\gamma}}_p = \mathbf{E}'_p\boldsymbol{\gamma}_p + \mathbf{b}_p$, where $\mathbf{E}'_u$ ($\mathbf{E}'_p$) is the transformations of user (item) stationary states, and $\mathbf{b}_u$ ($\mathbf{b}_p$) is the corresponding bias term. The stationary part of $\boldsymbol{\gamma}_u$ and $\boldsymbol{\gamma}_p$ is similarly calculated based on the standard factorization. In sum, we build a recurrent recommender FIRN that incorporates both dynamic fashion features from fashion bloggers and user's purchase history to give users a personalized fashion recommendation.

3.4 Optimization

In order to predict user ratings that are close to the actual ratings as well as capturing the fashion blogger's visual influence for each user across time, we propose an objective function that jointly learns user's ratings and their visual preference:

$$\underset{\Omega}{\text{minimize}} \sum_{(u,p,t)\in O_{train}} (r_{u,p}(t) - \hat{r}_{u,p}(t|\Omega))^2 + \lambda_1||\mathbf{v}(u,t) - \hat{\mathbf{h}}(u,t)||_F^2) + \lambda_2 R(\Omega), \tag{12}$$

where $r_{u,p}(t) - \hat{r}_{u,p}(t|\Omega)$ is used to yield predictions that are close to the actual ratings. $||v(u,t) - \hat{h}_k(u,t)||_F^2$ is to ensure we effectively adapt blogger's various visual information for different users in BiLSMT. Here $||\cdot||_F$ is the Frobenius Norm. λ_1/λ_2 is a hyper-parameter which is used to balance the visual/regularizer and rating information. O_{train} are the observed (user, item, timestep) tuples in the training set. Ω is the set of model parameters, and $R(\Omega)$ is the regularization function. Here we use the Frobenius Norm for each model parameter. The optimization method is the same as [39] (subspace gradient descent).

4 EXPERIMENTS

In this section, we conduct experiments on real-world datasets to evaluate the proposed FIRN recommender. Specifically, there are two key research questions: (1) How well does FIRN perform for fashion recommendation; (2) Whether our modeled fashion bloggers implicit visual influence is really helpful for the recommendation, especially compared with other widely used sources of visual information? Besides focusing on answering these two questions, we also conduct an ablation study, as well as explore FIRN performance on different users and their corresponding recommended items to further evaluate the FIRN architecture and the influence of fashion bloggers.

4.1 Experimental Setup

Datasets. We require datasets that contain both time-aware visual posts of influential fashion bloggers and user fashion purchases in the same time periods to model the dynamic influence. It is extremely difficult to find available public datasets that satisfy these requirements since few (if any) previous works consider the influence of fashion bloggers in recommendation. So for fashion bloggers, we crawled bloggers and their time-aware visual posts from Instagram which contains many influential and representative fashion bloggers and their rich visual posts [6, 33]. Since other sources of visual fashion could potentially be just as useful as these bloggers, we also consider a collection of images from user purchase histories on Amazon [11] and the AVA dataset of static aesthetic images [40]. For user purchases, we follow previous fashion works that use a public Amazon dataset [11, 40].

- *Instagram.* As an illustration of fashion bloggers and their visual influence in fashion recommendation, we use a list of 100 influential US female fashion bloggers as a seed set of fashion bloggers.[2] While this list is partial and reflects one view on who is influential, it gives us a starting point of many popular fashion bloggers (we further discuss the limitations of this dataset for fashion recommendation in Section 5). We crawl each of their Instagram [5] accounts, collecting all posts that overlap in time with the Amazon dataset. Concretely, we crawl the images associated with each post and the posted time, the number of likes for each post, the blogger's comments on the post, and five

[2]https://www.aransweatersdirect.com/blogs/blog/46644481-the-top-100-us-female-fashion-bloggers-to-follow-on-instagram

user comments featured by Instagram (where the comments are from accounts with large followings), resulting in 131,883 *time-aware visual* posts. The time-aware visual dataset is available at http://people.tamu.edu/~zhan13679/ for reproducibility and further research.

- *Amazon.* We use a public large real-world Amazon dataset [24] focusing on the Women's Clothing category that has been widely used for fashion recommendation [11, 40]. The dataset includes both rating history and item images. Concretely, we forcus on women's skirts, dresses, pants and so on (removing Intimates and Socks & Hosiery since they are typically not featured by fashion bloggers). We select items with an image that are rated between Jun. 2011 and Jul. 2014 (overlapping with our Instagram dataset) and their corresponding users. Keeping users with at least two ratings, we finally arrive at a ratings matrix with 22,217 users, 27,244 items and 59,866 ratings.
- *AVA Dataset*: This is a well-known public Aesthetic Visual Analysis (AVA) dataset [28]. It contains over 250,000 images with aesthetic ratings from 1 to 10 and we use the images rated 6-10 as aesthetic visual information for fashion recommendation.

Dataset Inspection. Particularly, we investigate the crawled fashion bloggers from different aspects to confirm its quality. We first examine the user comments towards those bloggers' posts. We find the most frequent unigrams express strong personal affinity – *love* is the most popular followed by *beauty*, *cute*, *like*. Further, unigrams *want*, *get*, and *need* also appear in the top-20 most popular unigrams. Those top frequent words prove bloggers'posts contain user favored visual features and even influence their purchase preference. We then explore the number of users who directly express their likes towards these posts. Figure 5 shows the growth in number of posts (blue) and the average number of likes per post (red) from 2011 to 2015. For example, one of our bloggers has a post with 675,000 likes in 2014. The huge amount of likes per post along with the top frequent words in user comments further confirms our intuition of the widely influence of fashion blogger's posts towards user aesthetic preference. Additionally, the fast increases in both number of posts and average likes per post shows our crawled bloggers are active across time. All those observations ensure the high-quality of our crawled data. In sum, the Instagram dataset naturally contains both *dynamic* and *high aesthetic quality* properties, which makes it potentially valuable for fashion recommendation.

In FIRN, the time variable is used to link Amazon dataset and Instagram dataset, and fashion information is transferred to capture user visual drift. The time intersection for the two datasets is from Jun. 2011 to Jul. 2014. We select the corresponding posts and user records in that time period, and discretize time by month, resulting in 38 time intervals. The choice of granularity, as an important hyper-parameter, is revisited on model performance in Section 4.2. We also compare our blogger visual features versus visual images from users time-aware purchased products used in [11] and those derived from the AVA dataset [28] used in [40].

Visual Features. For the visual features in Amazon items ($\mathbf{m}(p)$), Instagram posts ($\mathbf{m}_i(\mathcal{B}_k|t)$) and the AVA images, following previous work [11, 24], we use a convolutional neural network (CNN) proposed by [21] to *unify* the extracted visual features in the three

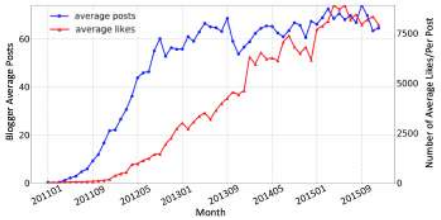

Figure 5: Posts and likes for Instagram fashion bloggers in our dataset grew rapidly.

Table 2: Model Comparison: FIRN is personalized, temporally-aware, visually-aware, and considers the impact of fashion bloggers.

Model	Personalized	Temporally-aware	Visually-aware	Influence-aware
SVD	✓	✗	✗	✗
AutoRecIU	✓	✗	✗	✗
TimeSVD++	✓	✓	✗	✗
SGRU	✓	✓	✗	✗
RRN	✓	✓	✗	✗
NSCR	✓	✗	✓	✓
FSVD	✓	✓	✓	✗
FIRN	✓	✓	✓	✓

datasets for fair comparison. The CNN is pre-trained by Caffe 1.2 million ImageNet. Particularly, the features that we use are the output of the second fully connected layer in CNN based on their strong performance in previous work [24]. The visual feature vector length is 4, 096.

Baselines. We compare FIRN against the following baselines:

SVD [20]: This is a widely used method which achieves robust and strong results in rating prediction. It uses user ratings without considering temporal dynamics. The regularization parameter is 0.01 by cross-validation.

AutoRecIU [30]: This is a recent autoencoder-based method for rating prediction. We use both item-based and user-based AutoRec methods and report the best performing one. The regularization parameter is 1 by cross validation.

TimeSVD++ [19]: One of the most successful models for time-aware recommendation based on matrix factorization, showing strong results across different datasets [39]. It considers temporal dynamics for both users and items. The regularization parameter is 0.01 by cross validation.

SGRU [12]: This method uses a session-based recurrent neural network (RNN) method to capture dynamics in recommendation and has strong results in prediction based on implicit feedback. Here we adapt it to predict ratings for each user. The loss function is mean square error. The drop out rate is 0.5.

RRN [39]: This is a recent state-of-the-art method for time-aware rating prediction. It uses a new RNN method to model long-range dynamics and stationary effects for users and items. The regularization parameter is 16 by cross-validation.

NSCR [37]: This is a recent state-of-the-art method for cross-domain recommendation. It utilizes both user-item attributes and a social network to give an item recommendation. We adapt this method for fashion recommendation, where the social network part is used

to model the influence from fashion bloggers to user purchase behavior. Item attributes are represented by item visual vectors and user attributes are denoted by the average visual vectors of their bought items. The social network between users and bloggers is built by their visual similarity between bought items and posts. If the similarity is larger than average, then there is a connect.

FSVD [11]: This is a recent state-of-the-art fashion-specific recommender, where fashion trends are modeled from user purchase history. The method is based on matrix factorization that considers temporal dynamics and item visual information. The regularization parameter is 0.001 by cross-validation.

Ultimately these methods are designed to evaluate the impact of temporal dynamics, visual factors, and FIRN framework for fashion recommendation as shown in Table 2.

Metrics. For fashion recommendation, following [39], we split the dataset by time into a training set (Jun. 2011 to Jun. 2013, with 21,112 ratings and 43,870 posts), validation set (Jul. 2013 to Jan. 2014, with 18,089 ratings and 16,452 posts) and test set (Feb. 2014 to Jul. 2014, with 20,665 ratings and 14,390 posts). Our evaluation consists of calculating how bloggers influence each user's ratings for items. So similar to [39], we use the user's average standard root-mean-square error (RMSE) to evaluate rating prediction:

$$RMSE = \frac{1}{|\mathcal{U}_{test}|}\Big(\sum_{u\in\mathcal{U}_{test}} \frac{1}{|P_u(t|\mathcal{U}_{test})|}||r_{u,p}(t) - \hat{r}_{u,p}(t)||_F^2\Big)^{1/2},$$

where $|\mathcal{U}_{test}|$ is the number of users in the test dataset. Here $|P_u(t|\mathcal{U}_{test})|$ represents the number of items that the user u rated at time t in the test dataset.

Reproducibility. All code, the Instagram data, and experimental results are available at http://people.tamu.edu/~zhan13679/. All results are reported over the same test set. For a fair comparison, the hidden dimension for all approaches is set to be 100 empirically for a trade-off between performance and computation complexity. Specifically, for RRN and FIRN, the dimension of stationary factors is 90 and the embedding dimension for temporal dynamics is 10. We set the visual dimension to 100 for all methods that use visual information. The hidden dimension of BiLSTM is 100 and $\mathbf{E}_s \in R^{100\times 10}$. The embedding dimension for FIRN is 10 which is the same as embedding dimension for temporal dynamics. Other hyperparameters are tuned based on the best performance on the same validation dataset. The regularization hyperparameters are tuned by grid search from 0.001 to 30 for different time-step granularity. Specifically, $\lambda_1 = 0.1$ and $\lambda_2 = 12$ for FIRN. Model parameters are first randomly initialized according to truncated normal distribution with mean 0 and standard deviation 0.01. For the optimization, we use mini-batch gradient descent where the max batch size is 100,000, and the corresponding learning rate is determined by grid search in the range of {0.00001, 0.0001, ...,0.1}.

4.2 Recommendation Performance of FIRN

We begin by investigating the overall performance of FIRN versus alternatives as shown in Table 3. The rows of the table capture different time-step granularaties. Overall, FIRN results in the best RMSE, with a 6% to 10% improvement versus the state-of-the-art fashion recommender FSVD and a 2% to 3% improvement versus the next-best approach (RRN in this case) across rows.

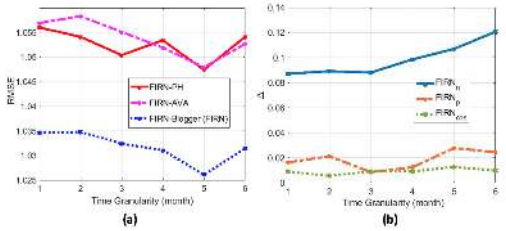

Figure 6: (a) RMSE of FIRN with different visual information; (b) Differences between FIRN and its variations.

Comparing in different time step granularities, FIRN consistently outperforms other methods, with relatively little change indicating the stability of FIRN in time-step granularaties. The best (lowest) RMSE of FIRN is gained when the time interval is five months, which suggests that fashion trends do not typically change abruptly in a short time period. We also observe that all of the time-aware methods result in significantly lower RMSEs than static methods (SVD and AutoRecUI), verifying the importance of modeling temporal dynamics of fashion preferences. Of the time-aware methods, SGRU does not perform very well, most likely since it is designed for implicit recommendation rather than rating prediction. Furthermore, both RRN and FIRN outperform the other methods, which indicates RRN is effective to capture dynamic changes compared with the other time-aware methods. More importantly, while FIRN and RRN have similar architecture, FIRN consistently performs better than RRN which highlights that incorporation of dynamic visual features from fashion bloggers gives FIRN its edge versus the temporal methods. Specifically, FIRN outperforms TimeSVD++ by 5.52% on average, SGRU by 26.35% on average, and RRN by 2.49% on average. We also observe that FIRN outperforms the state-of-the-art fashion recommendation method FSVD by 8.38% on average. It further highlights the efficacy of our model framework and the importance of incorporating fashion bloggers.

4.3 Influence of Fashion Bloggers

An important question is does the modeled blogger's implicit visual influence really help for the recommendation? Or put differently: does this give better performance in fashion recommendation compared with using other sources of visual information? In Section 4.2, we showed that FIRN consistently outperforms the corresponding alternative without the blogger's visual fashion information (RRN). This indicates that our modeled bloggers implicit visual influence improves the recommendation performance. However, is the improvement based on the blogger's implicit visual influence or could another visual source achieve similar performance? Here, we compare FIRN versus two alternatives that incorporate two widely used sources of visual information in fashion recommendation:

◦ *FIRN-PH*: The first approach replaces the posts of fashion bloggers with visual features derived from the users' Purchase History. Specifically, we use the average visual features of each user's purchase items in each time period.

◦ *FIRN-AVA*: The second approach uses the AVA dataset [28] as the indicator for user aesthetic preference [40]. Since AVA is static, we use the highest rated images (with ratings of 6-10) and cluster those images by their ratings. Each cluster acts as a virtual blogger in our FIRN framework.

Table 3: FIRN outperforms state-of-the-art methods in terms of RMSE for different time-step granularity. Δ_{RRN} shows the RSME improvement versus the next-best alternative, while Δ_{FSVD} shows the RMSE improvement versus the state-of-the-art fashion recommender FSVD. SVD, AutoRecUI and NSCR have the same performance cross rows since they are not time-aware.

	No time		Time aware			Visual Time & Blogger			Δ	
	SVD	AutoRecUI	TimeSVD++	SGRU	RRN	NSCR	FSVD	FIRN	Δ_{RRN}	Δ_{FSVD}
1 month	2.1392	1.6723	1.0775	1.4146	1.0655	1.1306	1.1146	1.0346	3.09%	7.99%
2 months	2.1392	1.6723	1.0805	1.3371	1.0627	1.1306	1.1126	1.0348	2.79%	7.78%
3 months	2.1392	1.6723	1.0774	1.3125	1.0581	1.1306	1.0983	1.0324	2.56%	6.59%
4 months	2.1392	1.6723	1.0884	1.2263	1.0521	1.1306	1.1187	1.0311	2.10%	8.76%
5 months	2.1392	1.6723	1.0964	1.2108	1.0478	1.1306	1.1197	1.0261	2.17%	9.36%
6 months	2.1392	1.6723	1.1015	1.2708	1.0535	1.1306	1.1293	1.0314	2.21%	9.79%

Figure 6(a) shows RMSEs for the two models and FIRN-Blogger (which is FIRN) – one using purchase history, one using AVA, and our original FIRN approach with fashion bloggers. FIRN consistently results in the best performance, illustrating that our modeled blogger's implicit visual influence brings the largest improvement in fashion recommendation. Particularly, the outperformance of FIRN compared with using purchase history indicates the high aesthetics quality of fashion blogger's posts. FIRN performs better than AVA. One likely reason is that as user's visual interest changes over time with fashion trends, the fashion blogger are able to reflect this evolution versus the static visual information in AVA. Interestingly, though purchase history and AVA outperform the baseline RRN which does not use visual information, they do not perform the best separately, which shows the dynamic and aesthetic visual properties are *mutually correlated* and enhance each other for fashion recommendation. This further shows that the learned implicit visual influence from bloggers, with the unique properties of containing both high-quality and dynamic visual features, indeed captures more visual information for user clothing recommendation than the other two visual datasets.

4.4 Ablation Study

This section evaluates the key design choices of FIRN: the impact of personalized attention layer, the impacts of visual distance between items and blogger's posts, and the visual distance choice in loss function. Concretely, method $FIRN_n$ uses a non-personalized attention layer (i.e. $\alpha_k(t) = softmax(w_k^T s_k(t))$) and the loss function is to minimize $\sum_{(u,p,t)}((r_{u,p}(t) - \hat{r}_{u,p}(t|\Omega))^2 + \lambda_2 R(\Omega)$ which does not consider visual distance between items and blogger's posts. The second method $FIRN_p$ uses a personalized attention layer but its loss function also does not consider the visual distance. The third method $FIRN_{cos}$ uses a cosine similarity of user and blogger visual vectors rather than Frobenius Norm (i.e. minimize $\sum_{(u,p,t)}((r_{u,p}(t) - \hat{r}_{u,p}(t|\Omega))^2 - \lambda_1 cos(v(u,t), \hat{h}_k(u,t)) + \lambda_2 R(\Omega))$.

Figure 6(b) shows the RMSE differences between FIRN and its variations (Δ is RMSE of FIRN variations minus RMSE of FIRN) by different time-step granularity. We observe that FIRN gives an average improvement of 9.8% in RMSE compared with $FIRN_n$. $FIRN_p$ performs better than $FIRN_n$ with a average improvement of 8.0%. This confirms the importance of utilizing personal attention in FIRN for fashion recommendation. Moreover, *FIRN* offers an 1.8% improvement compared with $FIRN_p$. It illustrates the importance of measuring visual distance between user bought items and blogger's posts. Interestingly, we find FIRN just slightly performs better than $FIRN_{cos}$, which indicates cosine and Frobenius Norm have similar effects for capturing personalized visual preference. It is reasonable since both cosine and Frobenius norm measure linear distance.

4.5 Fashion-Sensitive Users

Although FIRN improves the recommendation performance across users, a concern is that users may be differently influenced by fashion and thus a general good prediction can not ensure the recommended items are *fashionable* (e.g., a good prediction for a non-fashion influenced user can not show the recommended items are fashionable). In this section, we examine the FIRN performance for the fashion-sensitive users to further evaluate the fashion recommendation quality. Since it is hard to directly measure the impact of fashion to individual users without knowing user personal information, according to the experimental results from [36] and Section 2, as a proxy, we assume that in a small time-step granularity t, if a user purchased an item which is visually similar to posts that a blogger shares in same time period t, and such similarity is consistent for a long time period, then there is a higher probability that the user is more influenced by fashion bloggers/fashion.

Hence, based on the assumption, we sort users by the visual distance between the user purchased items and bloggers' posts across a long time period T:

$$d_v(u, \cup\mathcal{B}_k|T) = \frac{1}{|T|}\sum_{t\in T}\min_{p\in P_u(t), B_k\in\cup\mathcal{B}_k}(||\mathbf{m}(p) - \mathbf{m}(\mathcal{B}_k|t)||_F), \tag{13}$$

where $m(\mathcal{B}_k|t) = \frac{\sum_{i\in\mathcal{I}(\mathcal{B}_k|t)}\mathbf{m}_i(\mathcal{B}_k|t)}{|\mathcal{I}(\mathcal{B}_k|t)|}$ is the blogger's visual features at time t. Since the influenced users could buy one or more items that are not similar to the influenced blogger, here we use *min* across user bought items $p \in P_u(t)$ to find the smallest distance at time t. Similarly, the *min* is also calculated across bloggers $B_k \in \Pi$. The sum for the totel time T (38 time intervals in our case) can drop the probability that we include uninfluenced users who coincidentally bought visually similar items.

Table 4 reports the performance of FIRN for the most fashion-sensitive users according to Equation 13 for different thresholds. For these users who bought items that are most similar to bloggers (top 100 users), we find an RMSE of 0.9716 which is 9.68% better than the next-best alternative and 14.05% better than the fashion-aware FSVD. While as more users are considered, FIRN still maintains its superiority versus the next-best alternative. Furthermore, We observe that most other methods show approximately flat performance for users with different $d_v(u, \cup\mathcal{B}_k|T)$ while FIRN shows approximate monotonous relationship, indicating FIRN could gain a better performance for users with small $d_v(u, \cup\mathcal{B}_k|T)$. One likely

Table 4: Performance of different methods for the most fashion-sensitive users.

Most Fashion-Sensitive Users	No time		Time aware			Visual Time & Blogger			Δ	
	SVD	AutoRecUI	TimeSVD++	SGRU	RRN	NSCR	FSVD	FIRN	Δ_{RRN}	Δ_{FSVD}
100	2.0278	1.6964	1.0078	1.4666	1.0684	1.3828	1.1121	0.9716	9.68%	14.05%
500	2.0558	1.6799	1.0828	1.4666	1.0631	1.2552	1.1198	0.9852	7.79%	13.46%
1,000	2.0763	1.6589	1.0642	1.4398	1.0487	1.3156	1.1289	0.9940	5.47%	13.49%
5,000	2.1291	1.6762	1.0850	1.3279	1.0587	1.0744	1.1661	1.0225	3.63%	14.36%

Figure 7: (a) A selection of items purchased by three users. FIRN makes the best predictions for User 1 (low RMSE); User 2 and User 3 have higher RMSEs; (b) Corresponding recommendations for those users in Feb. 2014 (201402); (c) The least and most influential bloggers for User 1; (d) Examples of bloggers posts in 201402.

reason is that FIRN considers the visual distance between bloggers and user purchased products to capture the visual similarity in Equation 13. Those results demonstrate FIRN achieves a significant better performance to these fashion-sensitive users.

4.6 Case Study

To further investigate the implicit visual influence for FIRN recommendation, we also look at the FIRN recommended items. In particular, we focus on three specific users as shown in Figure 7, for whom FIRN provides varying recommendation quality: user 1 (good, RMSE=0.0778), user 2 (medium, RMSE=0.9767), and user 3 (poor, RMSE=2.7621). Each row of Figure 7(a) shows the purchase history for one user, and Figure 7(b) shows the recommendations for the corresponding user in the same row made by FIRN for Feb. 2014 (which is the earliest time in the test set). Figure 7(c) shows the predicted most influential blogger (first row) and least influential blogger (second row) for user 1 (by the value of the attention weight $\sum_t \alpha_k(u,t)$). Figure 7(d) shows the posts by the most influence blogger and a popular blogger ($\sum_u \alpha_k(u,t)$ is the largest) in Feb. 2014 for comparison.

Comparing Figure 7(a) and (c), we observe that the most influential blogger for user 1 (as learned by the attention weights $\sum_t \alpha_k(u,t)$) shares similar style/colors posts to user 1's purchased items. For example, the blogger posted a paisley dress (in the second post) and the blue/black color pairing in the second post. It confirms FIRN *does* learn visual features from bloggers that are similar to users. Furthermore, for the recommended items in Figure 7(b), we observe FIRN recommends items that reflect both fashion trends revealed by bloggers and the user's purchase history. For example, for user 1, the recommended black suit and black jacket are similar as bloggers, and blue/black color pairing is similar her purchase history. For user 3 whose RMSE by FIRN is poor, we observe the recommended items are visually diverse (stylish according to the fashion bloggers and visually related to the user's purchase history). This shows the potential of FIRN to recommend stylish clothing based on the current posts by fashion bloggers.

5 CONCLUSION AND DISCUSSION

In this paper, we have focused on the fashion blogger's implicit visual influence towards user's purchase preferences. We propose a novel recurrent neural fashion recommender – FIRN – which utilizes fashion bloggers dynamic visual information to extract fashion features and gives users *personalized visual influence-aware fashion recommendations*. The experimental results show the potential of incorporating dynamic visual fashion trends from fashion bloggers into a recommender, particularly for those users who are

most fashion-sensitive, and utilizing fashion bloggers can bring greater improvements in fashion recommendation. We also release a large *time-aware high-quality visual dataset* for reproducibility and further research. In our continuing work, we focus on two major issues:

New Fashion-Related Suggestions. We are interested to explore how to best create new recommenders tailored for different kinds of users – the fashion-sensitive, the fashion-neutral, and the unaffected (such as user 3 in Figure 7). For example, are unaffected users interested in seeing trendy recommendations (derived from fashion bloggers) alongside traditional purchase history recommendations (that are more suited to their current preferences)? Can we build recommenders that subtly move users from one group to another, say by tuning the attention weights over time? Furthermore, since fashion as a cultural phenomenon has a huge social influence, we can not declare that the unaffected users are not influenced by fashion at all. The fashion influence for these users could merely by invisible when the time span is relatively short. Since FIRN learns personal influence for each user, FIRN can benefit from the training data which has a relative long time span.

Expanding the Source of Fashion Trends. One key limitation of the current work is the reliance on a single source of fashion trends – Instagram fashion bloggers. We have seen how the fashion features highlighted by these popular and influential bloggers can improve recommendation quality, but we are interested to explore a wider range of fashion bloggers in our future work. As the growth of visual social media continues, we anticipate the influence of these unique fashion personalities will grow even more in comparison to traditional retailers, designers, and celebrities [36], enabling even more powerful recommenders based on their fashion leadership.

ACKNOWLEDGMENTS

This work is partially supported by NSF Grant IIS-1841138.

REFERENCES

[1] Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Jia Li, Vince Gatto, and Ed H Chi. 2018. Latent Cross: Making Use of Context in Recurrent Recommender Systems. In *WSDM*. ACM.

[2] Xiaofei Chao, Mark J Huiskes, Tommaso Gritti, and Calina Ciuhu. 2009. A framework for robust feature selection for real-time fashion style recommendation. In *IMCE*. ACM.

[3] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In *SIGIR*. ACM.

[4] Marijke De Veirman, Veroline Cauberghe, and Liselot Hudders. 2017. Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. *International Journal of Advertising* (2017).

[5] Nathaniel J Evans, Joe Phua, Jay Lim, and Hyoyeun Jun. 2017. Disclosing instagram influencer advertising: The effects of disclosure language on advertising recognition, attitudes, and behavioral intent. *Journal of Interactive Advertising* (2017).

[6] Fashion influencer. 2018. Fashion influencer — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Fashion_influencer [Online; accessed 5-November-2018].

[7] Leisa Reinecke Flynn, Ronald E Goldsmith, and Jacqueline K Eastman. 1996. Opinion leaders and opinion seekers: Two new measurement scales. *Journal of the academy of marketing science* (1996).

[8] Vijay Gabale and Anand Prabhu Subramanian. 2018. How To Extract Fashion Trends From Social Media? A Robust Object Detector With Support For Unsupervised Learning. *arXiv preprint arXiv:1806.10787* (2018).

[9] Pamela Church Gibson. 2012. *Fashion and celebrity culture*. Berg.

[10] Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* (2005).

[11] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *WWW*. International World Wide Web Conferences Steering Committee.

[12] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).

[13] Shintami Chusnul Hidayati, Cheng-Chun Hsu, Yu-Ting Chang, Kai-Lung Hua, Jianlong Fu, and Wen-Huang Cheng. 2018. What Dress Fits Me Best?: Fashion Recommendation on the Clothing Style for Personal Body Shape. In *2018 ACM Multimedia Conference on Multimedia Conference*. ACM, 438–446.

[14] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* (1997).

[15] Yang Hu, Xi Yi, and Larry S Davis. 2015. Collaborative fashion recommendation: A functional tensor factorization approach. In *MULTIMEDIA*. ACM.

[16] Vignesh Jagadeesh, Robinson Piramuthu, Anurag Bhardwaj, Wei Di, and Neel Sundaresan. 2014. Large scale visual recommendations from street fashion images. In *SIGKDD*. ACM.

[17] Wang-Cheng Kang, Eric Kim, Jure Leskovec, Charles Rosenberg, and Julian McAuley. 2019. Complete the Look: Scene-based Complementary Product Recommendation. In *CVPR*.

[18] Young Jun Ko, Lucas Maystre, and Matthias Grossglauser. 2016. Collaborative recurrent neural networks for dynamic recommender systems. In *Journal of Machine Learning Research: Workshop and Conference Proceedings*.

[19] Yehuda Koren. 2009. Collaborative filtering with temporal dynamics. In *SIGKDD*. ACM.

[20] Yehuda Koren and Robert Bell. 2011. Recommender systems handbook. *F. Ricci, L. Rokach, B. Shapira, & BP Kantor (Eds.)* (2011).

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS*.

[22] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: short-term attention/memory priority model for session-based recommendation. In *SIGKDD*. ACM.

[23] Alice Marwick. 2013. âĂŸTheyâĂŹre Really Profound Women, TheyâĂŹre EntrepreneursâĂŹ: Conceptions of Authenticity in Fashion Blogging. In *ICWSM*.

[24] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *SIGIR*. ACM.

[25] Edward F McQuarrie, Jessica Miller, and Barbara J Phillips. 2012. The megaphone effect: Taste and audience in fashion blogging. *Journal of Consumer Research* (2012).

[26] Miren Mendoza. 2010. I Blog. You Buy. How bloggers are creating a new generation of product endorsers. *Digital Research & Publishing* (2010).

[27] Subhabrata Mukherjee and Stephan Günnemann. 2019. GhostLink: Latent Network Inference for Influence-aware Recommendation. *Work* 1145 (2019), 3308558–3313449.

[28] Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *CVPR*. IEEE.

[29] Holly L Schrank and D Lois Gilmore. 1973. Correlates of fashion leadership: Implications for fashion process theory. *Sociological Quarterly* (1973).

[30] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. 2015. Autorec: Autoencoders meet collaborative filtering. In *WWW*. ACM.

[31] Aviv Shoham and Ayalla Ruvio. 2008. Opinion leaders and followers: A replication and extension. *Psychology & Marketing* (2008).

[32] Tiana Stefanic. 2010. Outsiders looking in: how everyday bloggers are gaining access to the elite fashion world. *Journal of Digital Research and Publishing* (2010).

[33] M Sudha and K Sheena. 2017. Impact of influencers in consumer decision process: the fashion industry. *SCMS Journal of Indian Management* (2017).

[34] Peijie Sun, Le Wu, and Meng Wang. 2018. Attentive recurrent social recommendation. In *SIGIR*. ACM.

[35] Paige Thornley. 2014. Examining the role of bloggers in the fashion industry: A public relations strategy for new designers. (2014).

[36] Cassidy L Vineyard. 2014. The relationship between fashion blogs and intention to purchase and word of mouth behavior. (2014).

[37] Xiang Wang, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. 2017. Item silk road: Recommending items from information domains to social users. In *SIGIR*. ACM.

[38] Jane E Workman and Kim KP Johnson. 1993. Fashion opinion leadership, fashion innovativeness, and need for variety. *Clothing and Textiles Research Journal* (1993).

[39] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J Smola, and How Jing. 2017. Recurrent recommender networks. In *WSDM*. ACM.

[40] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, and Zheng Qin. 2018. Aesthetic-based clothing recommendation. In *WWW*. ACM.

[41] Mohd Zaimmudin Mohd Zain, Patsy Perry, and Lee Quinn. 2018. The Influence of Fashion Bloggers on the Pre-Purchase Decision for Online Fashion Products among Generation Y Female Malaysian Consumers. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* (2018).