Fashion Recommendation Based on Style

Federico Becattini, Lavinia De Divitiis, Claudio Baecchi and Alberto Del Bimbo

[1*]University of Florence, Media Integration and Communication Center (MICC), Viale Morgagni 65, Florence, Italy.

*Corresponding author(s). E-mail(s): federico.becattini@unifi.it; Contributing authors: lavinia.dedivitiis@unifi.it; claudio.baecchi@unifi.it; alberto.delbimbo@unifi.it;

Abstract

Fashion recommendation is often declined as the task of finding complementary items given a query garment or retrieving outfits that are suitable for a given user. In this work we address the problem by adding an additional semantic layer based on the style of the proposed dressing. We model style according to two important aspects: the mood and the emotion concealed behind color combination patterns and the appropriateness of the retrieved garments for a given type of social event. To address the former we rely on Shigenobu Kobayashi's color image scale, which associated emotional patterns and moods to color triples. The latter instead is analyzed by extracting gar- ments from images of social events. Overall, we integrate in a state of the art garment recommendation framework a style classifier and an event classifier in order to condition recommendation on a given query.

Keywords: Style, Social Events, Garment Recommendation, Fashion

1 Introduction

Fashion is a way to express personal style and to communicate emotional states that reflect personality or mood. At the same time, certain social events follow a dress code which can be strictly required or implicitly followed by participants. The rules of such social requirements however are not written and can be hard to comply with, especially for first-time attendants. Even formal events, where dressing options are more constrained, might leave space for attendants to decline an outfit according to specific preferences or personalities. Even at ceremonies such as weddings, one can adhere to the most formal requirements or play with details to give a casual or sporty touch the outfit and thus to his or her appearance. Such customization of an outfit is a mean to communicate to others a certain intent, to establish a friendly or formal environment and even to be provocative or remark a social status. Such moods and tastes are conveyed by both the shape and color of any fashion garment and may be subjected to trends and personal interpretations, yet color combinations and outfit making appear to be grounded on general clothing rules that are well established within the society.

Patterns behind color-combinations have been extensively studied, in particular with reference to different styles one may want to communicate. The seminal work by Shigenobu Kobayashi [1] introduced a scale to express emotions or attitudes based on color combinations. According to Kobayashi, just considering triplets of colors, one can identify a wide variety of lifestyles which can then be expressed by personal spaces such as interiors or offices or personal items such as outfits. On the other hand, deriving precise rules for specific events is harder since it can vary in time and follow different trends in different communities of people.

In this work we propose to follow a data driven approach to learn to model these implicit rules, both following color-combination styles and social event requirements. For color-combinations we rely on the Kobayashi's color image scale, whereas for social events we extract from online available data information about garment appearance and the respective social event. We intend to develop a complete recommendation system capable of taking into account the conveyed emotion and the compliance to societal standards for different types of events. We believe that this can come to the aid of people that seek help in shop assistants, either physically in shops or virtually on an online marketplace.

Since different modalities can convey different emotions or be suitable for different events, we also analyze the capabilities of a recommendation system to take into account also diversity. In general, diversity is an important factor to consider while tackling any information retrieval task. As suggested in [2], this also applies to benchmark datasets and the way they are exploited to produce any recommendation. The introduction of stylish aspects based on visual cues such as color combinations is also a step towards this direction.

A preliminary version of our approach was described in [3]. The system presented in this work differs substantially from [3] in several ways: (i) instead of focusing only on Kobayashi's color image scale, we present a novel analysis concerning social events by gathering the *Fashion4Events* dataset comprising approximately 400k garment images with social event labels; (ii) using the collected dataset, we train an outfit based event classifier; (iii) we integrate both the color classifier and the event classifier in our recommendation system in order to provide a more fine-grained filtering of the recommendations. We also provide an improved analysis of the state of the art.

The paper is organized as follows. In Sec. 2 we first provide an overview on the state of the art for fashion recommendation. On overview of our presented method is presented in Sec. 3. Here we introduce: (i) an outfit emotion classifier based on color combinations, capable of mapping a generic outfit onto Kobayashi's color scale; (ii) a garment based social event classifier, used to infer the event category suitable for a given outfit; (iii) the integration of such modules with a state of the art garment recommendation system. The three aspects are then detailed respectively in Sec. 4, Sec. 5 and Sec. 6. The results of our experiments are reported in Sec. 7. Finally, we draw conclusions in Sec. 8

2 Related work

Fashion recommendation must pursuit the goals of recommending pertinent items [4–7], providing a set of different dressing modalities [6, 8, 9] and possibly complying with some user query or desired criteria [3, 10].

Several methods have focused on proposing garments in order to comple- ment a given query item [4, 8, 11]. In particular, [12] focused on complementary clothing matching, starting from a top garment and recommending a bot- tom item. The proposed method devised a compatibility modeling scheme with attentive knowledge distillation also exploiting a teacher-student network scheme. The approach has then been improved by studying a personalized compatibility modeling, leveraging both general and subjective aesthetic preferences with a personalized compatibility modeling scheme named GP-BPR [4]. Still exploiting Bayesian Personalized Ranking (BPR), [13] used multi- ple autoencoder neural networks to leverage the multi-modalities of fashion items and their inter-compatibility. Following up on this line of research, PAI-BPR [14] proposed an attribute-wise interpretable compatibility scheme with personal preference modelling.

Top-bottom recommendation has also been addressed exploiting Memory Augmented Neural Networks (MANN) [3, 6, 8]. This type of architecture exploits an external memory [15–19] by pairing different clothing items and training a memory writing controller to store a non-redundant subset of samples. This is then used to retrieve a ranked list of suitable bottoms to complement a given top. External memories have also been used to store disen- tangled features for separate attributes [6, 10]. In our work we rely on a memory based approach, namely GR-MANN [8], building on top of it a style-based and event-based recommendation system.

A similar, yet more complex task is the one of generating a whole outfit from a given garment seed. One of the first approaches [20] modeled sequences of suggestions by jointly learning a visual-semantic embedding and training a bidirectional LSTM (Bi-LSTM) model to sequentially predict the next item conditioned on previous ones. *Vasileva et al.* [11] instead learned pairwise embeddings to obtain separate representation for different pairs of fashion categories. Recent approaches tend to represent an outfit as a graph, linking fashion items among themselves if they are compatible with each other [21–23].

As for interpreting and diagnosing the proposed outfit compatibility and suggestions, [24] learns type-specified pairwise similarities between items and uses the backpropagation gradients to diagnose incompatible factors. Towards interpretable and customized fashion outfit compositions, [25] train a partitioned embedding network to favor interpretability of intermediate representations.

Finally, an interesting emerging topic is the one of understanding user reactions in order to integrate implicit user feedback into an iterative recommendation system, e.g. looking at body movements [26] or facial expressions [27].

3 Overview

In this paper we propose a system for compatible outfit recommendation, meaning that given a garment of a specific category (e.g., a top), we are able to propose suitable garments of a complementary category (e.g., bottoms) in order to compose an outfit. Our system pays attention to three important aspects. First, *style-based recommendations* take into account the desired style and emotions the user would like to express. We base the recommendations on Kobayashi's Color Image Scale (CIS) [1] by training a style classifier, which we integrate into our system to condition garment proposals. Second, *event-based recommendations* focus on garments that can complement the given input and at the same time can be suitable for attending a certain kind of social event, such as a work meeting, a hike or a wedding. To this end, we train an event classifier that, given an outfit, provides a probability distribution over different event categories. Similarly to the style-based recommendation, we integrate the event classifier into our system to condition the recommendations. Third, we focus on a fundamental, yet often neglected, aspect for recommendation systems, i.e. *recommendation variety*. We rely on a model specifically designed to recommend diverse garments with few redundancies and repetitions. Rather than proposing several variations of the same outfit/style, we aim at proposing different modalities that the user can choose. To this end we introduce an entropy-based evaluation to quantify such variety.

In the following we are going to provide detailed explanations of the style and event classifiers and we are then going to illustrate how these can be integrated into a recommendation system thought for recommending variegate outfits. We carry out this study taking as reference GR-MANN [8], a recent state of the art garment recommendation system based on the usage of Memory Augmented Neural Networks.

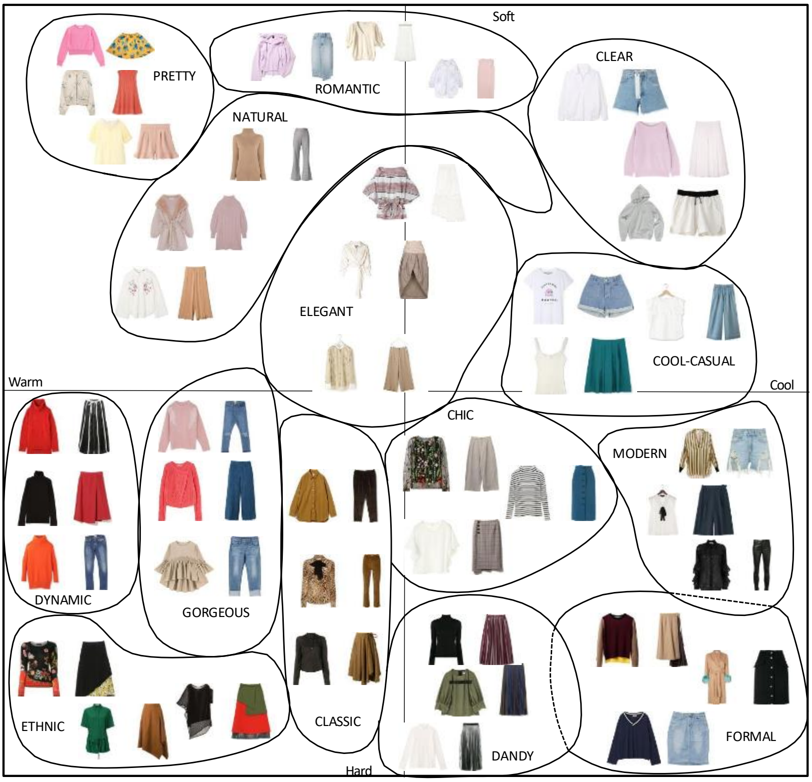

Fig. 1 Kobayashi's color image scale [1] applied to outfits.

4 Style-based Outfit Recommendation

We refer to the task of Style-based Outfit Recommendation as the task of recommending fashion items to complement an outfit, conditioned by a given style. Styles can convey moods or emotions and with this in mind we derived an interpretation of color patterns from Kobayashi's Color Image Scale [1]. We first train a style classifier which can then be combined with a recommendation system.

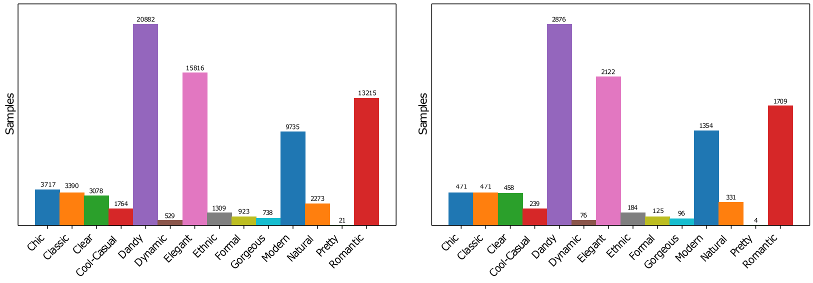

Fig. 2 Sample distribution for the training (left) and test (right) sets of the collected dataset. Categories correspond to Kobayashi's color patterns.

representing selected terms in fashion and lifestyle and identifying clusters in a color space spanned by two orthogonal warm-cool and short-hard axes (Fig. 1).

4.1 Outfit Style Classifier

We exploit a CNN model to infer the style of an outfit. In order to analyze outfits starting from a top and a bottom, we use a concatenation of the two images depicting the two garments. Styles are instead identified by Kobayashi's patterns in the Color Image Scale. We consider each pattern as a semantic description for the style of an outfit (e.g., *casual*, *elegant*, *dandy*). Such styles indicate the feelings that an individual may want to communicate rather than describing outfit characteristics such as shape.

In order to train such a model, we collected a set of ground truth annotations following a semi-automatic procedure. Outfit images are preprocessed by removing the background and color-quantizing foreground pixels to the palette of 130 tonalities used in CIS. We then take the three most frequent colors and compare them to the Kobayashi's triplets using an euclidean distance in order to find the closest style characterizing the outfit:

$$d* = \min_{p,j} \sqrt{\left\| P\left(c_o, p\right) - c_j \right\|_2} \tag{1}$$

where $P(c, p)$ is the p-th permutation of colors in the c triplet, c_o is the triplet for outfit o and c_j the j-th of the 1170 triplets identified by Kobayashi. We retain only outfits with a clear style, i.e. if $d* < \vartheta$. The resulting category is then mapped to one of Kobayashi's 15 style patterns. Finally, we asked human annotators to validate the final labeling, discarding or correcting erroneous assignments.

The procedure yielded a dataset of 77390 labeled outfits for training and 10516 for testing. All outfits are taken from the IQON3000 dataset [4]. In our experiments we considered all styles except the *casual* pattern, for which enough samples are not present in the dataset, yielding to a total of 14 style categories (Fig. 2). In order to recognize outfit styles, we trained a ResNet18 [28] classifier that takes as input a concatenation of the top and bottom images. We trained the network for 50 epochs using an Adam optimizer with a learning rate of 0.001.

Fig. 3 Confusion matrix for the outfit style classifier. Left: our ResNet18-based style classifer. Right: nearest neighbor

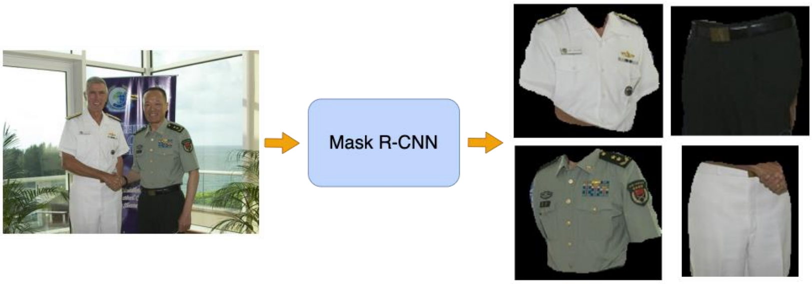

Fig. 4 Garments are detected and segmented using Mask-RCNN. For each detected gar- ment, a new image is created. Event labels associated with the original images are associated with the detected garments.

Personal Stylist (Mood-based)

A personal stylist based on mood suggests outfits according to how a person feels. For example, if someone is happy, it may recommend bright and colorful clothes, while calm moods may get soft pastel suggestions. This makes dressing more fun and expressive, as clothes reflect emotions. It can also be expanded by adding weather and wardrobe preferences for smarter suggestions.

4.2 Dataset creation

To collect Fashion4Events, a dataset of garment images paired with social event labels, we exploited two different sources of data: the DeepFashion2 dataset [29] and the USED dataset [30].

DeepFashion2 is a dataset that proposed an unified benchmark for clothes detection, segmentation, retrieval and landmark prediction. It contains approximately 491K images of clothes, divided into train (391K), validation (34K) and test (67K) belonging to 13 different classes. Garments exhibit large variations in style, pose, scale, color, occlusion and viewing angle.

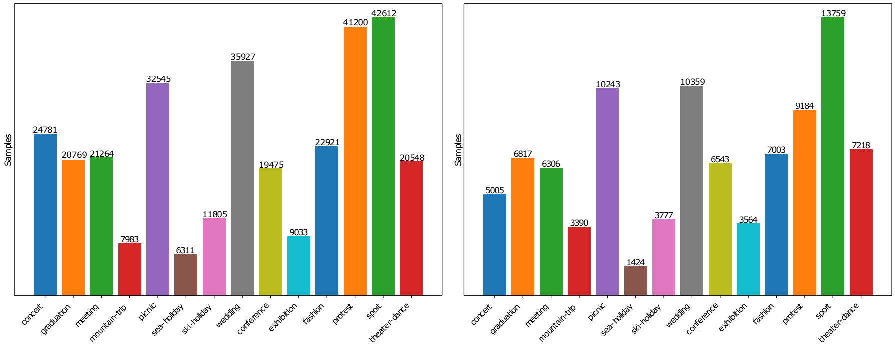

Fig. 5 Sample distribution for the training (left) and test (right) sets of the collected Fashion4Events dataset. Categories correspond to social events.

We use DeepFashion2 to train a garment detection and segmentation network by fine-tuning a Mask-RCNN model [31]. Once we have a model capable of detecting and segmenting individual garments in images, we apply it on the USED dataset [30]. USED is a dataset comprising 525K images of people attending 14 different types of social events. Images have been downloaded from Flickr and event types have been selected among the most common ones on social media. The categories are the following: concert, graduation, meeting, mountain-trip, picnic, sea-holiday, ski-holiday, wedding, conference, exhibition, fashion, protest, sport and theater-dance. Images represent both indoor and outdoor scenes and the dataset accounts for a variable number of people in each image.

By applying Mask-RCNN on the images from USED, we obtain a set of detected garments paired with pixelwise segmentations and a social event label. We retain only detections with a confidence score higher than 0.8 in order to remove noise. For each detected garment, we generate a separate image by taking the segmented clothing item on a black background, as shown in Fig. 4. The final set of detections yields a dataset of 317,174 images for training and 94,592 for testing. In Fig. 5 we show the distribution of samples over the 14 event categories in the Fashion4Events dataset.

5 Style-Based and weather Outfit Recommendation

6 A weather-based outfit recommendation system suggests clothes according to the weather conditions. For example, on a hot sunny day it may recommend light cotton clothes, while on rainy days it suggests waterproof jackets or umbrellas. In cold weather, it recommends sweaters, coats, or warm layers. This helps people dress comfortably and practically while also matching the climate. It can even combine with mood for smarter styling.

7 Experiments

We demonstrate our method on the IQON3000 dataset [4], performing several evaluations. At first we discuss the accuracy of the style classifier and the event classifier and then we combine them with GR-MANN to analyze their capability to recommend a diverse set of bottom garments covering multiple styles. In order to evaluate the performance of the event classifier we rely on the event-garment dataset based on USED, obtained as explained in Sec. 5.1.

7.1 Style-Based and Event-Based Outfit Recommendation Evaluation

We now assess the capabilities of the GR-MANN recommendation system with reference to styles and social events. First, we measure how the recommenda- tion system is able to suggest garments that comply with the ground truth style and event category, without adding any prior knowledge to the model. We report Accuracy, measuring if at least one of the recommended outfits adheres with the requested category, and mAP, which takes the ranking of the correct bottoms into account.

In Tab. 1 we compare the results for style-based recommendation obtained by the model against a baseline in which style categories are drawn at random. It can be seen that, both for Accuracy and for mAP, the results improve considerably and that the model is able to provide at least an outfit with the desired style most of the times even with only 5 recommendations.

Table 2 Accuracy and mAP obtained by GR-MANN [8] on the IQON3000 [4] dataset. The metrics are computed in order to retrieve an outfit with the same social event category of the ground truth.

Num Items	5	10	20	30	40	50	60
Accuracy	75.83	88.50	94.96	95.06	96.72	97.24	97.70
Random Acc.	37.12	53.03	79.87	91.21	96.73	98.99	99.16
mAP	43.05	40.00	36.66	34.25	33.07	32.30	31.42
Random mAP	18.31	16.87	17.34	16.44	14.12	13.36	11.07

Table 3 Accuracy and mAP varying the number or retrieved items. All proposals are filtered by the style classifier in order to share the desired one.

Num Items	5	10	20	30	40	50	60
Cat × Col Acc.	57.66	59.22	61.92	64.37	66.42	68.36	69.97
Category Accuracy	83.92	84.81	86.27	87.60	88.68	89.67	90.39
Color Accuracy	81.41	82.73	84.71	86.37	87.69	88.82	89.77
mAP	18.50	18.48	18.37	18.23	18.11	17.98	17.88

Table 4 Accuracy and mAP varying the number or retrieved items. All proposals are filtered by the event classifier in order to share the desired category.

Num Items	5	10	20	30	40	50	60
Cat × Col Acc.	52.00	53.70	55.46	57.92	60.86	62.74	64.68
Category Accuracy	86.14	86.36	88.24	88.78	89.98	90.76	90.40
Color Accuracy	75.92	77.70	79.96	81.58	83.78	84.74	86.46
mAP	16.02	15.92	15.47	15.33	15.57	15.42	14.98

As for event-based recommendation, we report a similar analysis in Tab. 2. Also in this case, the model generates a set of recommendations comprising a suitable bottom to comply with the social event of interest.

Additionally, following the evaluation protocol of [8], we also measure accu- racy color-wise, category-wise and combining both together. However, we filter the output of GR-MANN in order to provide a ranked list of bottoms with the correct category using the style classifier or the event classifier. Therefore, in this experiment we are relaxing the formulation of the task, assuming that the desired style is known a-priori. The rationale behind this evaluation is to see if the proposed garments share similar visual traits with the ground truth when performing a category-conditioned (either style-based or event-based) recommendation. In Tab. 3 and Tab. 4 we report results for both accuracy and mAP.

7.2 Recommendation Diversity Evaluation

As studied in [2], diversity is an important aspect of information retrieval systems. To this end, we also perform an evaluation of the entropy of the proposed labels to establish the variation degree of our proposals. This evaluation was first proposed in [32] to perform an unsupervised assessment of a generic classifier with unsupervised data. Given a probability distribution

Table 5 Entropy of the recommendations with reference to outfit styles. A sufficiently high entropy indicates variety in the proposed outfits.

Method \Num Items	5	10	20	30	40	50	60
Random	1.419	1.905	2.255	2.391	2.458	2.499	2.523
Style-Entropy	0.849	1.061	1.152	1.265	1.267	1.288	1.317

Table 6 Entropy of the recommendations with reference to outfit event categories. A sufficiently high entropy indicates variety in the proposed outfits.

Method \Num Items	5	10	20	30	40	50	60
Random	1.353	1.839	2.189	2.325	2.392	2.433	2.457
Event-Entropy	0.671	0.819	0.908	0.941	0.958	0.965	0.970

$X = \{x_1, x_2, \ldots, x_n\}$ over N different classes, we can compute the Shannon entropy H for the probability vector X as $H(X) = -\sum_{i=1}^{N} x_i log(x_i)$. The entropy will be 0 when all samples are labeled with the same class, and will increase as more information and diversity are introduced in the predictions. Ideally we would like to stay as close as possible to the entropy of any ran- dom label permutations, but preserving good recommendation results. Results, shown in Tab. 5 and Tab. 6, show that our method is able to maintain a rea- sonable amount of entropy in the predictions while performing significantly better than random, as shown in Tab. 1 and Tab. 2.

Interestingly, the entropy for Kobayashi's style categories is higher than the one relative to event categories. We attribute this to the distribution of garments in IQON3000, for which certain categories such as ski-holiday and wedding are underrepresented.

8 Conclusions

In this paper we presented an approach to take into account a style based filtering and an event based filtering for fashion recommendation. Styles can convey a color-based mood according to Kobayashi's Color Image Scale or can reflect dress codes for specicic social event categories. We leveraged the work of Kobayashi to train a style classifier that we used to filter the results of a memory network based garment recommender. Similarly, we trained a garment-based event classifier to be combined with the recommender by exploiting a garment detector and Fashion4Events, an image dataset of anno- tated social events. Experiments show that our system is able to generalise on color styles and social events and that the recommendation system is able to propose a variety of outfit styles compatible with the query garment.