Style-Based Outfit Recommendation

Lavinia De Divitiis
University of Florence
Florence, Italy
lavinia.dedivitiis@unifi.it

Federico Becattini
University of Florence
Florence, Italy
federico.becattini@unifi.it

Claudio Baecchi
University of Florence
Florence, Italy
claudio.baecchi@unifi.it

Alberto Del Bimbo
University of Florence
Florence, Italy
alberto.delbimbo@unifi.it

2021 International Conference on Content-Based Multimedia Indexing (CBMI) | 978-1-6654-4220-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/CBMI50038.2021.9461912

***Abstract*—In this paper we propose a garment recommendation system that leverages emotive color information to give recommendations that adhere to a desired style. We leverage previous work by Shigenobu Kobayashi on how specific color combinations, that pertain to certain pre-defined styles, are able to convey specific emotions in human beings. Leveraging this information, we extend the classic general garment recommendation to a style-driven one, where the user can adapt the suggestions to a specific style that may be more appropriate for a specific social event. Here, first we train a generalized style classifier based on Kobayashi's color triplets, then we leverage a recent memory network-based garment recommendation system to perform suggestions of bottom garments (e.g. skirts, trousers, etc.) given a user-defined top (e.g. a shirt, T-shirt, etc.). Suggestions are then processed to maintain only the ones that, according to our classifier, are coherent with the user defined style. Experiments show that our system is able to generalise on Kobayashi's color styles and that the recommendation system is able to propose garments that are in line with the user desire while also introducing diversity in the proposed garments.**

***Index Terms*—garment recommendation, memory augmented networks, fashion**

I. Introduction

Individuals have several ways of expressing moods, emotions and feelings. Fashion for instance 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. However, general clothing rules appear to be well established and grounded into how people live within the society. For instance, formal events will require a formal dressing as well as schools and offices might require a clothing conduct. Such rules mostly reflect on which category of garments an individual will wear to certain occasions and do not consider any additional feeling that one may want to communicate. For this reason, people are likely to play with color combinations to express an emotional statement or to stand out.

This sort of behavior has been well known for years, since the seminal work by Shigenobu Kobayashi [1] who 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.

Diversity is also 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.

In this paper we explore the integration of such emotions into a garment recommendation system, which we condition to generate outfits that satisfy a certain mood or style, following Kobayashi's color scale.

II. Related work

Recently, garment recommendation systems have received an increasing interest given their key role in suggesting meaningful fashion items to accommodate user personal styles and emotions. Recent examples are given by [3]–[5], where both user preferences and interest are considered to produce appropriate suggestions.

[6] focus on complementary clothing matching and propose a compatibility modeling scheme with attentive knowledge distillation also exploiting the teacher-student network scheme. [7] model 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.

To promote diversity, [8] also try to consider context and address the compatibility prediction problem using a graph neural network that learns to generate product embeddings. As for interpreting and diagnosing the proposed outfit compatibility and suggestions, [9] learn type-specified pairwise similarities between items, and use the backpropagation gradients to diagnose incompatible factors.

[4] study personalized compatibility modeling leveraging both general and subjective aesthetic preferences by using a personalized compatibility modeling scheme GP-BPR. Also on user-item interaction and with general item-item interaction, [10] propose an attribute-wise interpretable compatibility scheme with personal preference modelling.

Exploiting Bayesian Personalized Ranking, [11] make use of multiple autoencoder neural networks to leverage the multi-modalities of fashion items and their inter-compatibility. Towards interpretable and customized fashion outfit compositions, [12] train a partitioned embedding network to favor interpretability of intermediate representations.

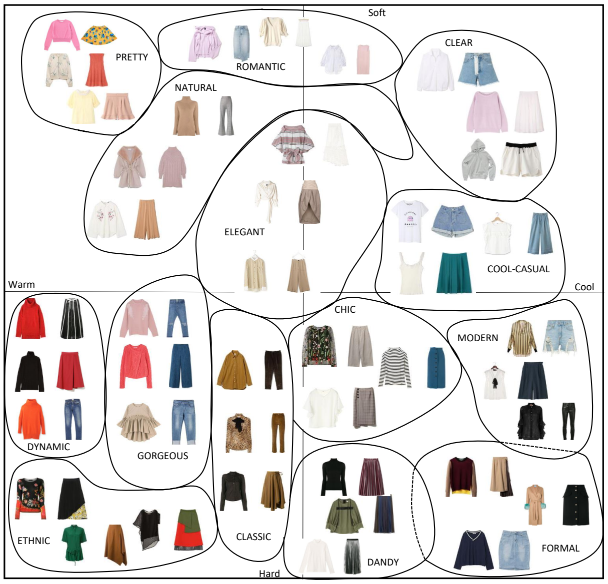

Fig. 1. Kobayashi's color image scale [1] applied to outfits.

Finally, [13], propose to leverage Memory Augmented Neural Networks (MANNs) [14]–[17] for garment recommendation by pairing different clothing items and training a memory writing controller to store a non-redundant subset of samples, which is then used to retrieve a ranked list of suitable bottoms to complement a given top.

III. Method

In this paper we address the task of Style-based Outfit Recommendation, i.e. the task of recommending fashion items to complement an outfit, conditioned by a given style. Styles are derived from Kobayashi's Color Image Scale [1] by training a style classifier to be combined with a recommendation system. We carry out this study taking as reference GR-MANN [13], a recent state of the art garment recommendation system based on the usage of Memory Augmented Neural Networks.

A. Kobayashi's Color Image Scale

Originally presented in the early 90s', Kobayashi's Color Image Scale (CIS) [1] connected images and colors from a psychological viewpoint, investigating associations between colors and their underlying semantics. The study carried out by Kobayashi involved a decade of color-based psychophysical experiments, asking humans to annotate triplets of colors with an adjective describing perceived emotions, moods or styles. The research eventually identified a list of 1170 triplets made of 130 unique colors and associated with 180 adjectives referred to as color images. These labels have been grouped into 15 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).

B. Outfit Style Classifier

We exploit a CNN model to infer the style of an outfit. An outfit o is a concatenation of two images depicting a top and a bottom garment. 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:

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

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 $d^* < \theta$. 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 [18] 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.

C. Style-Based Outfit Recommendation

We integrate our outfit style classifier in the state of the art garment recommendation system GR-MANN [13]. Note that in principle the classifier could be applied to any recommendation system capable of generating an outfit composed of a top and a bottom garment.

GR-MANN is a neural network based on a persistent external memory in which non redundant samples are stored to guide recommendations. The samples in memory express different modalities to combine tops and bottoms. At inference time, the query top is presented to the model and encoded with a convolutional encoder. Its feature is then used as key to retrieve similar tops in memory and access the correspondent stored bottoms.

We extend GR-MANN by using the outfit style classifier to check whether the recommended outfits comply to a certain style requested by the user. Once a bottom is proposed, it is concatenated to the input top and fed to the style classifier. We perform different fusion methods, in order to characterize the behaviour of the system, as detailed in Section IV.

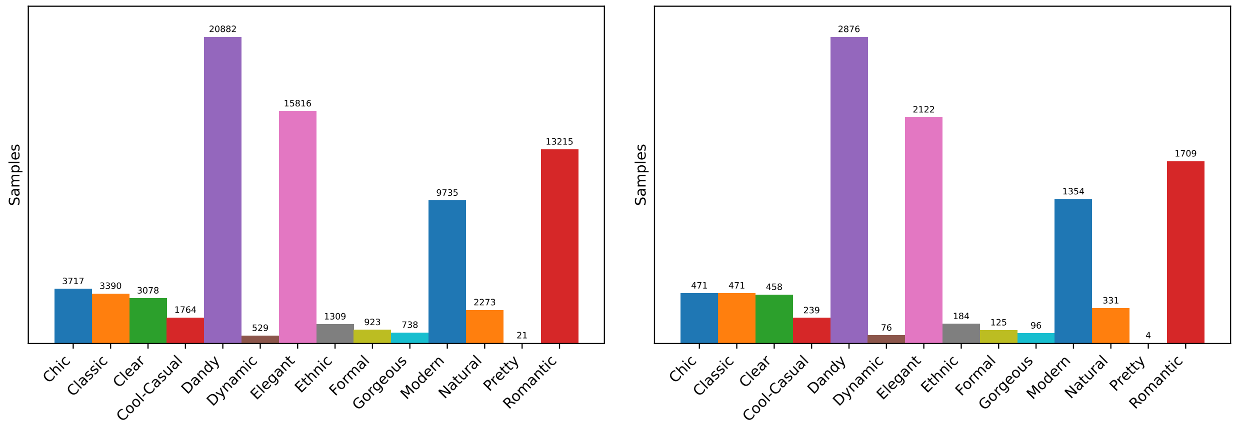

Fig. 2. Sample distribution for the training (left) and test (right) sets of the collected dataset. Categories correspond to Kobayashi's color patterns.

TABLE I
ACCURACY AND MAP OBTAINED BY GR-MANN [13] ON THE IQON3000 [4] DATASET. THE METRICS ARE COMPUTED IN ORDER TO RETRIEVE AN OUTFIT WITH THE SAME STYLE OF THE GROUND TRUTH.

Num Items	5	10	20	30	40	50	60
Accuracy	73.42	84.17	91.47	94.22	95.73	96.70	97.39
Random Acc.	36.50	53.12	80.37	91.14	96.51	99.02	99.89
mAP	48.17	46.35	43.46	41.78	40.70	39.96	39.40
Random mAP	18.14	17.21	17.04	16.59	14.39	13.41	11.46

TABLE II
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 $\times$ 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

The choice of exploiting GR-MANN stems from the fact that its memory is populated by a controller, trained to store only relevant samples. Classifying such samples based on style will therefore allow us to quantify the diversity that GR-MANN strives to achieve in its recommendations.

IV. Experiments

We demonstrate our method on the IQON3000 dataset [4], performing several evaluations. At first we discuss the accuracy of the style classifier and then we combine it with GR-MANN to analyze its capability to recommend a diverse set of bottom garments covering multiple styles.

A. Style Classifier Evaluation

In Fig. 3 the confusion matrix for the style classifier is shown. As a reference, a Nearest Neighbor (NN) baseline is also reported. Here we simply extract features from a standard ResNet, pretrained on Imagenet, and transfer the style category from the closest sample in the training set onto a given test outfit. Our model achieves an accuracy of 78.56% while the NN baseline simply 49.11%, highlighting how the task is not as straightforward as comparing visual features.

Interestingly, most of the errors committed by the model tend to confuse similar categories, such as *elegant* and *chic* or *classic* and *dandy*.

B. Style-Based Outfit Recommendation Evaluation

We now assess the capabilities of the GR-MANN recommendation system with reference to style categories. First, we measure how the recommendation system is able to suggest garments that comply with the ground truth style, without adding any prior knowledge about Kobayashi's categories to the model. In Tab. I we compare the results 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.

Additionally, following the evaluation protocol of [13], we also measure accuracy 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 style using the style classifier. Therefore, in this experiment we are relaxing the formulation of the task, assuming that the desired style is known a priori. In Tab. II we report results for both accuracy and mAP.

C. 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 is first proposed in [19] to perform unsupervised evaluation of a generic classifier with unsupervised data. Given a probability distribution $X = \{x_1, x_2, \dots, x_n\}$ 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

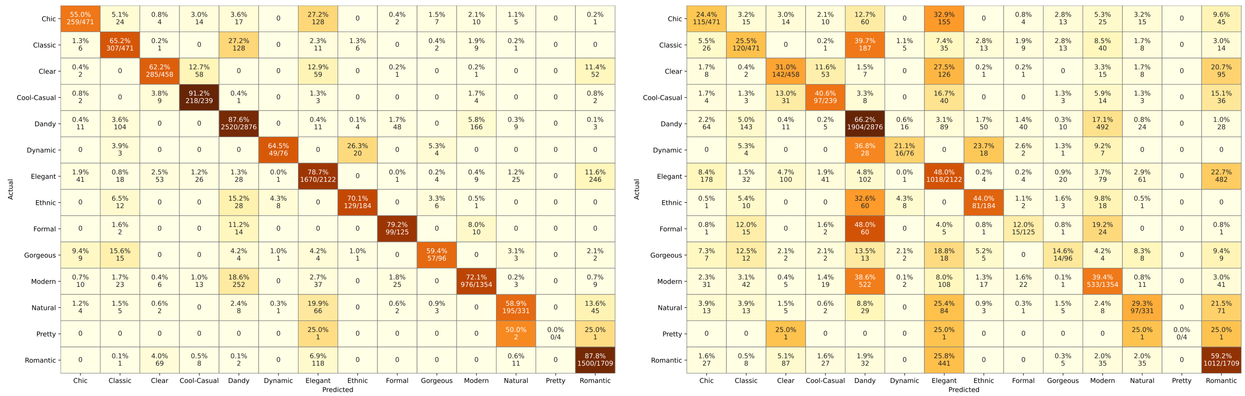

Fig. 3. Confusion matrix for the outfit style classifier. Left: our ResNet18-based style classifer. Right: nearest neighbor

TABLE III
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
Ours	0.849	1.061	1.152	1.265	1.267	1.288	1.317

possible to the entropy of any random label permutations, but preserving good recommendation results. Results, shown in Tab. III, show that our method is able to maintain a reasonable amount of entropy in the predictions while performing significantly better than random, as shown in Tab. I.

V. Conclusions

In this paper we presented a style based approach to suggest bottom outfits according to a user style preference. 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. Experiments show that our system is able to generalise on Kobayashi's color styles and that the recommendation system is able to propose a variety of outfit styles compatible with the query garment.

Acknowledgments

This work was partially supported by the Italian MIUR within PRIN 2017, Project Grant 20172BH297: I-MALL - improving the customer experience in stores by intelligent computer vision. The authors thank Federico Nocentini for his contribution to the development of the system.

References

[1] S. Kobayashi, "Color image scale," *http://www. ncd-ri. co. jp/english/main_0104. html*, 2009.
[2] J. Z. Wang, N. Boujemaa, A. Del Bimbo, D. Geman, A. G. Hauptmann, and J. Tesić, "Diversity in multimedia information retrieval research," in *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, ser. MIR '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 5–12. [Online]. Available: https://doi.org/10.1145/1178677.1178681
[3] W.-H. Cheng, S. Song, C.-Y. Chen, S. C. Hidayati, and J. Liu, "Fashion meets computer vision: A survey," *arXiv preprint arXiv:2003.13988*, 2020.
[4] X. Song, X. Han, Y. Li, J. Chen, X.-S. Xu, and L. Nie, "Gp-bpr: Personalized compatibility modeling for clothing matching," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 320–328.
[5] W. Bigi, C. Baecchi, and A. Del Bimbo, "Automatic interest recognition from posture and behaviour," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2472–2480.
[6] X. Song, F. Feng, X. Han, X. Yang, W. Liu, and L. Nie, "Neural compatibility modeling with attentive knowledge distillation," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 5–14.
[7] X. Han, Z. Wu, Y.-G. Jiang, and L. S. Davis, "Learning fashion compatibility with bidirectional lstms," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 1078–1086.
[8] G. Cucurull, P. Taslakian, and D. Vazquez, "Context-aware visual compatibility prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 617–12 626.
[9] X. Wang, B. Wu, and Y. Zhong, "Outfit compatibility prediction and diagnosis with multi-layered comparison network," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 329–337.
[10] D. Sagar, J. Garg, P. Kansal, S. Bhalla, R. R. Shah, and Y. Yu, "Pai-bpr: Personalized outfit recommendation scheme with attribute-wise interpretability," in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2020, pp. 221–230.
[11] J. Liu, X. Song, Z. Chen, and J. Ma, "Neural fashion experts: I know how to make the complementary clothing matching," *Neurocomputing*, vol. 359, pp. 249–263, 2019.
[12] Z. Feng, Z. Yu, Y. Yang, Y. Jing, J. Jiang, and M. Song, "Interpretable partitioned embedding for customized fashion outfit composition," *arXiv preprint arXiv:1806.04845*, 2018.
[13] L. D. Divitiis, F. Becattini, C. Baecchi, and A. D. Bimbo, "Garment recommendation with memory augmented neural networks," in *Pattern Recognition. ICPR International Workshops and Challenges - Virtual Event, January 10-15, 2021, Proceedings, Part II*, ser. Lecture Notes in Computer Science, vol. 12662. Springer, 2020, pp. 282–295. [Online]. Available: https://doi.org/10.1007/978-3-030-68790-8_23
[14] A. Graves, G. Wayne, and I. Danihelka, "Neural turing machines," *arXiv preprint arXiv:1410.5401*, 2014.
[15] J. Weston, S. Chopra, and A. Bordes, "Memory networks," *arXiv preprint arXiv:1410.3916*, 2014.
[16] F. Marchetti, F. Becattini, L. Seidenari, and A. Del Bimbo, "Mantra: Memory augmented networks for multiple trajectory prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
[17] F. Marchetti, F. Becattini, L. Seidenari, and A. Del Bimbo, "Multiple trajectory prediction of moving agents with memory augmented networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[19] G. Cuffaro, F. Becattini, C. Baecchi, L. Seidenari, and A. Del Bimbo, "Segmentation free object discovery in video," in *European Conference on Computer Vision*. Springer, 2016, pp. 25–31.