Aesthetic Features for Personalized Photo Recommendation

Yu Qing Zhou
Shopify
ivan.zhou@shopify.com

Ga Wu
University of Toronto
wuga@mie.utoronto.ca

Scott Sanner
University of Toronto
ssanner@mie.utoronto.ca

Putra Manggala
Shopify
putra.manggala@shopify.com

ABSTRACT

Many photography websites such as Flickr, 500px, Unsplash, and Adobe Behance are used by amateur and professional photography enthusiasts. Unlike content-based image search, such users of photography websites are not just looking for photos with certain content, but more generally for photos with a certain photographic "aesthetic". In this context, we explore personalized photo recommendation and propose two aesthetic feature extraction methods based on (i) color space and (ii) deep style transfer embeddings. Using a dataset from 500px, we evaluate how these features can be best leveraged by collaborative filtering methods and show that (ii) provides a significant boost in photo recommendation performance.

KEYWORDS

Photo Recommendation, Image Style, HSV, Aesthetic Features

ACM Reference Format:
Yu Qing Zhou, Ga Wu, Scott Sanner, and Putra Manggala. 2018. Aesthetic Features for Personalized Photo Recommendation. In *Proceedings of the Late-Breaking Results track part of the Twelfth ACM Conference on Recommender Systems (RecSys 2018).* Vancouver, BC, Canada, October 2-7, 2018, 2 pages.

1 INTRODUCTION

Personalized aesthetic-based photo recommendations can help photography websites better serve the needs of amateur and professional photographers. Content-based image search does not fully satisfy the needs of such users since they are usually not interested in content alone. Instead, they are often looking for photos with certain photographic aesthetics, which may include monochromaticity, light contrast, and style. While there has been research on automatic assessment of image aesthetics (e.g., [2, 3]) and clothing recommendation leveraging aesthetic features (e.g., most recently [8]), this work is not appropriate for personalized photo recommendation for photography enthusiasts as we detail in Section 2.

In this paper, we conduct experiments on a dataset from 500px. We propose two feature extraction approaches to obtain aesthetic features from photos without manual annotation. The first approach extracts features from the Hue-Saturation-Value (HSV) channels of a photo since these align with human perception of brightness and whiteness [1]. The second approach extracts deep style embeddings from photos [4]. Empirically, we show that the use of deep style embeddings as side information outperforms a variety of baselines.

2 RELATED WORK

Many existing works in the literature on image aesthetic assessment are based on Photo.Net [2] or datasets from the DPChallenge, like AVA [3]. These datasets were annotated with semantic and aesthetic labels and rated by users unidentifiable to researchers. For the purposes of this paper, it is not clear that these annotators align with the photography enthusiast community. Further, these works explored (non-personalized) image classification tasks.

Some previous work has leveraged stylistic and aesthetic image features for personalized recommendation of fashion products. For example, [6] used style-based features derived from deep embeddings for personalized clothing recommendation. Alternatively, [8] incorporated simple binary aesthetic features (e.g., *if the images are aesthetically pleasing to the public or not*). In contrast, we focus on extracting rich aesthetic photo features such as color composition and texture-based deep style embeddings of [4] that we conjecture may better relate to photography enthusiast preferences.

3 APPROACH

Photo recommendation can be formalized in the usual matrix view: given a set of m users, a set of n photos, and the observed interaction of users with photos $R \in \{0, 1\}$ (1 indicates a positive interaction, and 0 indicates no observed interaction) with shape of $m \times n$, we want to rank the top-k images that a user may positively interact with in the future.

One variant[1] of item-item Nearest Neighbor Collaborative Filtering (I-NN) [5] predicts user i's interaction with photo j as

$$\hat{r}_{i,j} = \sum_{k \in \{1 \dots n, k \neq j\}} Sim(\phi(j), \phi(k)) \cdot r_{i,k}, \tag{1}$$

where $\phi(j)$ is a vector of information for item j, and $Sim(\phi(j), \phi(k))$ defines similarity between photos j and k; multiple similarity functions Sim are available such as Cosine, Pearson, and Euclidean. If there is side information $\mathbf{p}_j$ available, in addition to the rating column $\mathbf{r}_{:,j}$ in matrix R, it is expressed as

$$Sim(\phi(j), \phi(k)) = \theta \cdot Sim(\mathbf{p}_j, \mathbf{p}_k) + (1 - \theta) \cdot Sim(\mathbf{r}_{:,j}, \mathbf{r}_{:,k}), \tag{2}$$

where θ is a relative weighting hyperparameter that can be tuned through cross-validation.

Next we describe two aesthetic feature extraction methods.

HSV Color-Embedding. The HSV color space was designed to capture the human perception of color [1]. It has been used in (non-personalized) photo aesthetic assessment [2]. In contrast with the traditional RGB color space, HSV separates out luminance from color information. We represent the HSV (or RGB) color-embedding $\mathbf{p}_j$ of photo j as a concatenated histogram vector of the three HSV (or RGB) channels illustrated in Figure 1.

RecSys 2018, October 2018, Vancouver, BC Canada

[1] In another variant of I-NN, the predicted interaction in (1) is normalized by $\sum_{k \in \{1 \dots n, k \neq j\}} Sim(\phi(j), \phi(k))$. We observe better ranking performance without this.

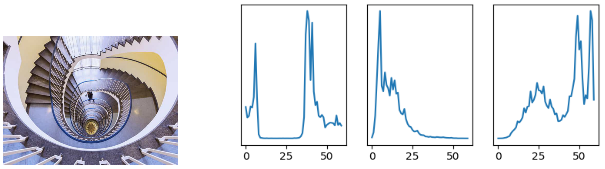

Figure 1: (left) An example photo. (right) Density plot of Hue, Saturation, and Value channels for the photo.

Style Embedding. Histograms extracted from HSV or RGB channels are intuitive, but fail to capture more complex photo properties such as style (e.g., textures). Fortunately, recent research in neural style transfer [4] shows that stylistic information of an image can be extracted with a deep convolutional neural network. Specifically, the style embedding $\mathbf{p}_j$ is the vectorized form of the Gram matrix of a specific layer (determined through experimentation) from an inference performed on the photo j using a VGG-19 network [7].

4 RESULTS

We conduct our experiments on a dataset from 500px, an online photography website. The dataset contains 225,922 users and 300,000 photos. We prepared five temporally-split triples of train, validation, and test sets from this dataset. The validation set is used to tune hyperparameters, which are θ in (2) and the similarity measure (either Cosine, Euclidean, or Pearson). We evaluate three different ranking metrics: *Precision@10*, *R-Precision*, and *Average Precision*.

The baseline models used in the experiment are the following:

(1) *Random*: photos are ranked in a random order.
(2) *Popular*: photos are ranked by popularity (most likes).
(3) *I-NN*: item-based nearest neighbour as covered in Section 3.
(4) *I-NN-Meta*: *I-NN* with photo metadata ($\mathbf{p}_j$), which includes categories, keywords, and *"Editor's Choice"* labels.

The novel photo recommendation models (including typical hyperparameter settings from validation-based tuning) that we compare against the baselines are *I-NN-HSV* ($\theta = 0.01$, Cosine), *I-NN-RGB* ($\theta = 0.04$, Cosine), and *I-NN-Style* ($\theta = 0.20$ weights the sum of Euclidean on style vectors and Cosine on interaction vectors), which use aesthetic features from the HSV and RGB color channels, and deep style embeddings [4], respectively, as defined in Section 3.

Next, we experimentally answer a few research questions.

Is HSV better than RGB? In Table 1, we note that *I-NN-HSV* outperforms *I-NN-RGB* indicating that HSV color features may provide better similarity information for photo recommendation than RGB as we conjectured. However, color information may not help at all since neither outperformed the best baseline *I-NN*.

What is the best layer of VGG and best distance metric for determining style-based similarity? We set up an experiment with multiple sets of photos that are aesthetically similar to each other based on human annotation. We extract the style vectors from all photos with one specific layer of VGG-19 and then apply one of the similarity measures (Cosine, Euclidean, and Pearson) to compute the similarity between style vectors. We select the top k pairs of photos based on similarity scores and determine how many of these pairs were in the same aesthetic set (denoted as the *Precision@k* for $k = 10$ and $k = 15$). The results are provided in Figure 2. The Euclidean similarity metric applied to style embeddings extracted from layer 8 of VGG-19 gives us the highest *Precision@k* for both $k = 10$ and $k = 15$. Thus, we use this configuration for *I-NN-Style*.

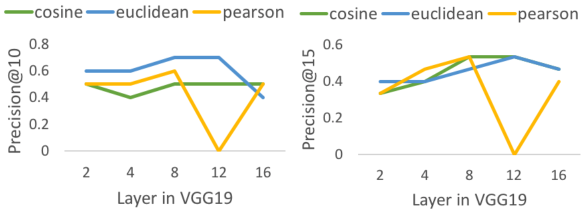

Figure 2: Evaluation of VGG-19 layer and metric for style extraction that maximizes recommendation performance.

Table 1: Recommender performance comparison (±95% CIs).

Model	Precision@10	R-Precision	Avg Precision
Random	0.006 ± 0.002	0.006 ± 0.002	0.006 ± 0.002
Popular	0.021 ± 0.004	0.018 ± 0.004	0.021 ± 0.005
I-NN	0.048 ± 0.009	0.041 ± 0.008	0.047 ± 0.010
I-NN-Meta	0.044 ± 0.005	0.037 ± 0.004	0.042 ± 0.006
I-NN-HSV	0.041 ± 0.006	0.033 ± 0.006	0.040 ± 0.008
I-NN-RGB	0.038 ± 0.003	0.030 ± 0.004	0.037 ± 0.006
I-NN-Style	**0.059** ± **0.012**	**0.050** ± **0.011**	**0.057** ± **0.014**

Which photo recommender performs best overall? Table 1 provides comparative results. As noted earlier, *I-NN* performs best among baselines and outperforms three methods that use photo-related information: *I-NN-Meta*, *I-NN-HSV*, and *I-NN-RGB*. However, the best performer overall by a substantial margin is *I-NN-Style* that uses photo aesthetic information derived from style embeddings.

5 CONCLUSION

Our results show that color and explicit metadata side information for photos do not help photo recommendation performance. However, style embeddings derived from layer 8 of VGG-19 provide a significant boost, demonstrating the importance of aesthetic style features in recommendation for photography enthusiasts.

REFERENCES

[1] H.D. Cheng, X.H. Jiang, Y. Sun, and Jingli Wang. 2001. Color image segmentation: advances and prospects. *Pattern Recognition* 34, 12 (12 2001), 2259–2281.
[2] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. 2006. Studying Aesthetics in Photographic Images Using a Computational Approach. Springer, 288–301.
[3] Yubin Deng, Chen Change Loy, and Xiaoou Tang. 2017. Image Aesthetic Assessment: An experimental survey. *IEEE Signal Processing Magazine* 34, 4 (7 2017), 80–106. http://ieeexplore.ieee.org/document/7974874/
[4] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *2016 CVPR*. IEEE, 2414–2423.
[5] G. Linden, B. Smith, and J. York. 2003. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing* 7, 1 (1 2003), 76–80.
[6] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. 2015. Image-Based Recommendations on Styles and Substitutes. In *Proceedings of SIGIR '15*. 43–52. http://dl.acm.org/citation.cfm?doid=2766462.2767755
[7] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. (9 2014). http://arxiv.org/abs/1409.1556
[8] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, and Zheng Qin. 2018. Aesthetic-based Clothing Recommendation. In *Proceedings of WWW '18*. 649–658.