User-Specific Ranking Method for Photographic Images (Postprint)

Authors: Su Shimei, Wang Meng, Xu Yongbo

Date: 2018-05-20T00:00:00+00:00

Abstract

Personalized photo ranking holds significant importance in image quality assessment and image retrieval. To address the shortcomings of existing methods that neglect user preferences and suffer from low accuracy, we propose a novel user-specific aesthetic ranking model based on ranking support vector machines. First, images preferred by the user are input, after which features are extracted via deep convolutional neural networks and compared with the dataset to create a user-specific aesthetic training set. Subsequently, ranking support vector machines are employed to learn a customized hyperplane and generate personalized aesthetic rankings exclusive to the user. In subsequent experiments, the first group of experiments invited users to evaluate the algorithm' s personalized predictions, while the second group tested the accuracy of high versus low image quality classification. Experimental results demonstrate that the algorithm' s predictions align well with user preferences, while simultaneously achieving high accuracy in classifying image quality levels. Therefore, this algorithm is an effective personalized ranking method.

Full Text

Preamble

A User-Specific Ranking Method for Photographic Images

Su Shimei, Wang Meng, Xu Yongbo

School of Electrical Engineering, Zhengzhou University, Zhengzhou 450000, China

Abstract: Personalized aesthetic ranking plays an important role in image quality assessment and image retrieval. To address the drawbacks of existing methods that neglect user preferences and suffer from low accuracy, this paper proposes a novel user-specific aesthetic ranking model based on ranking support vector machines. The framework first takes as input a series of photos preferred by the user, then employs a deep convolutional neural network to extract features and compare them with a dataset to create a user-specific aesthetic training set. Subsequently, ranking support vector machines are used to learn a customized hyperplane and generate a personalized aesthetic ranking for the user. In subsequent experiments, the first group invited users to evaluate the algorithm' s personalized predictions, while the second group tested the accuracy of high-low image quality classification. Experimental results demonstrate that the algorithm' s predictions align well with user preferences while achieving high accuracy in high-low quality image classification. Therefore, the proposed algorithm is an effective personalized ranking method.

Key Words: user-specific; aesthetic assessment; ranking model; deep learning

0 Introduction

With the flourishing development of social networks and digital cameras, the number of photos captured, transmitted, and shared online has exploded, making personal photo collections sufficiently large for users to select and preserve valuable images. However, manually curating ideal and personalized albums or commemorative collections from massive image repositories is not only time-consuming and tedious but also labor-intensive—a challenge yet to be fully explored. The underlying problem lies in accurately identifying individual user preferences. This paper focuses on how to automatically evaluate the aesthetic characteristics of input images while accounting for user preferences.

In recent years, universal descriptors derived from high-quality images have gained popularity among researchers due to their adherence to photographic rules and computational efficiency. Applying simple aesthetic and photographic principles to image classification can effectively enhance automatic computational performance in distinguishing high-quality from low-quality images. Researchers have conducted extensive experiments ranging from low-level visual features to high-level compositional characteristics. Tong et al. [1] proposed a set of low-level visual features to differentiate professional from amateur photography. Datta et al. [2] focused on understanding images through computational methods using a 56-dimensional feature vector. Based on the assumption that professional photographers more prominently highlight photo subjects, Luo et al. [3] first introduced subject region extraction. Mavridaki et al. [4] pioneered pattern aesthetic attributes and achieved promising results when combined with simplicity and composition. Aydin et al. [5] used five aesthetic features (sharpness, color, hue, and clarity) to automatically evaluate and edit images. However, low-level visual features struggle to fully capture the deeper traits of aesthetics, prompting some researchers to turn toward high-level compositional features. Dhar et al. [6] proposed describable features based on content, composition, and lighting to predict image interestingness. Lo et al. [7] introduced a set of discriminative yet efficiently computable features. Luo et al. [8] extracted corresponding features from different categories and trained respective classifiers. Although handcrafted aesthetic features still hold a place in image quality assessment, they mostly represent approximations of abstract aesthetic principles and cannot fully capture the diversity and beauty of photographic images.

With the rise of deep learning, deep feature extraction has gradually replaced handcrafted features in image quality assessment. Lu et al. [9] proposed a double-column deep convolutional neural network to simultaneously capture global and local features. Dong et al. [10] achieved favorable classification results by directly adopting the ImageNet network [11] to extract 4096-dimensional feature vectors. Focusing on similarity within the same image category, Tian et al. [12] combined deep and semantic features to create a query-dependent model that outperformed previous methods. Wang et al. [13] innovatively fed both original images and HSV channel images into a double-column neural network, demonstrating significant performance improvements. Despite deep neural networks achieving high accuracy in high-low quality image classification, aesthetics lacks absolute meaning when personal preferences are absent. Due to the highly subjective and complex nature of human aesthetic perception, finding universal methods for the masses is unrealistic. As social networks have evolved, users tend to attach tags or descriptive language to uploaded images. Researchers have consequently shifted their focus from personalized ranking to customized search and online recommendation. Vijendran et al. [14] proposed a comprehensive tag-annotation system spanning image upload, tag matching, semantic prediction, and user-preference-based retrieval. Lu et al. [15] defined an online image as an event and proposed a tag-based ranking system that extracts tag, description, and comment information during event browsing and integrates them into semantic sentences to predict customized rankings. Nwana et al. [16] introduced a novel personalized tagging method that adjusts tag order and automatically eliminates irrelevant tags based on visual content. While researchers strive to improve personalized online search performance, online images may suffer from mismatched topics or ambiguity. For an image awaiting evaluation, the absence of descriptive information or tags is fatal.

The user-specific aesthetic ranking method proposed in this paper can efficiently and comprehensively solve the aforementioned problems. To address the shortcomings of low-level or high-level visual features, this paper adopts deep convolutional neural networks (DCNN) to extract image features, ensuring full capture of intrinsic image diversity. For personalized ranking, rather than using tag-based semantic search, this paper first constructs a user-specific training set through similarity retrieval from the entire training set, then employs SVMrank [17] to learn a user-specific model. The contributions of this paper are twofold: (a) a novel method for constructing user-specific training sets that combines user preferences with deep neural networks; (b) a user-specific aesthetic ranking model based on the constructed training set to predict the aesthetic ranking of input images.

1 Traditional Aesthetic Ranking Model

For existing input images, traditional aesthetic evaluation models aim to automatically predict image aesthetic characteristics using machine learning methods on training sets. In these models, lowercase letters denote aesthetic scores ranging between -1 and 1. The traditional model represents a special case of typical binary classification problems, where +1 indicates high quality and -1 indicates low quality. Currently, aesthetic scores are typically obtained by maximizing posterior probability as shown in Equation (1).

SVMrankuI11{(,),···,(,),···,(,)}uunnIyIyIy uIuy[1,1] 11 uyarg min p(|,) (1,1)uuyyIy ()SVMSVMrank

2.1 Personalized Training Set Generation and Correction

When addressing the challenge of personalized ranking, using traditional training sets with binary labels alone is insufficient. The limitation of binary labels lies in their neglect of user preferences during the selection process. The most common solution involves adding personalized features [19], such as color preferences and aspect ratios. Rather than using traditional training sets, this paper establishes user-specific training sets to correctly rank user preferences.

Given an existing user preference image set $U = \{I_{u1}, I_{u2}, ..., I_{un}\}$, where I_{ui} represents the user' s most preferred photo set and $U \subseteq \Gamma$. For image set U, the algorithm first extracts deep features and then performs image similarity retrieval. This paper argues that learning a personalized ranking model requires a corresponding exclusive training set. Therefore, we first search the shared visual space for images visually similar to user-selected specific images to form the exclusive training set Ψ as shown in Equation (2).

$$\Psi = \{I_i, S(I_i)\} \quad \text{where} \quad S(I_i) = \{I_j | I_j \in \Gamma, I_j \in \mathcal{N}(I_i)\}$$

where Γ is the training set and $\mathcal{N}(I_i)$ is the neighboring space set of user-specific image I_i. The features extracted in this paper are based on DCNN, designed and proposed by Krizhevsky et al. [11]. DCNN has achieved tremendous success in various computer vision domains such as classification and object detection. Similarly, DCNN has yielded remarkable results in photographic image aesthetic quality assessment [10,20]. The DCNN used in this paper consists of eight layers: five convolutional layers followed by three fully connected layers. After testing the performance of the last three layers, this paper uses fully connected layer 7 for feature extraction.

During the similarity retrieval process, Equation (3) is employed to reduce training phase errors.

$$c(e, I_i, I_j) = \begin{cases} 0, & \text{if } I_i = I_j \\ \text{ceiling}(f(I_i, I_j)), & \text{if } I_i \neq I_j \end{cases}$$

The function aims to push retrieval results to higher priority levels, i.e., if the same image repeatedly appears at different ranks in personalized retrieval, lower-priority images (ranked lower in retrieval) should defer to higher-priority images (ranked higher in retrieval).

The size selection for the exclusive training set is crucial. Due to variations in image quantity and quality across different categories, selecting a moderate number of highly relevant images is necessary. In retrieval results for 'landscape' and 'object' categories, similar images are abundant (more than 10), while for 'night scene' categories, only 10-15 relevant images are retrieved. To ensure the training set adequately captures input image characteristics, we select 10 images for each user-specific photo. After constructing the user-specific training set, its feature vectors are input into the ranking function to learn a customized ranking model. The DCNN-based aesthetic learning model has achieved favorable results. [Figure 2: see original paper] shows partial retrieval results for user-specific images.

2.2 User-Specific Aesthetic Ranking Model

As described in Section 2.1, a qualified ranking model should firmly grasp deep aesthetic abstractions without neglecting user preferences. After generating the personalized training set, the model' s objective is to rank and predict user preferences through a new set of input test images. The data extracted from Ψ is represented by feature vectors $x \in \mathbb{R}^d$, where d is the feature vector dimension. This generates the personalized training set described in Section 2.1.

The goal of the user-specific aesthetic ranking model is to learn the ranking function shown in Equation (4) by finding a maximizing hyperplane w while simultaneously reducing generalization error to satisfy the constraints shown in Equation (5).

$$f(x) = w^T x$$

$$w^T x_i > w^T x_j \quad \forall (i,j) \in \Gamma$$

This hyperplane is similar to SVM classification, with the objective of generating image pairs consistent with the internal query. This leads to the optimization problem shown in Equation (6).

$$\min_{w,\xi} \frac{1}{2}||w||^2 + C \sum_{(i,j)\in\Gamma} \xi_{ij}$$

$$\text{s.t. } w^T x_i - w^T x_j \geq 1 - \xi_{ij}; \quad \xi_{ij} \geq 0 \quad \forall (i,j) \in \Gamma$$

where w is the weight vector of the ranking function, C is a parameter balancing training error and margin, and ξ_{ij} are slack variables across different images. By using SVMrank, the model learns a customized hyperplane from the training set and enforces an ideal order for ranking test images. This paper is inspired by the customized hyperplane and employs it to learn user-specific aesthetic rankings.

3 Experiments and User Study

This chapter introduces two large public datasets used in experiments: CUHKPQ and a subset of the AVA dataset. Both datasets are widely used in aesthetic evaluation research. This paper compares several state-of-the-art methods [12,13,21,22] to demonstrate the effectiveness of user-specific aesthetic ranking.

The CUHKPQ dataset, collected by Tang et al. [8], contains 17,673 images provided by amateur photographers to professional photography websites with manually labeled ground truth quality. The entire dataset is divided into seven categories: 'people' , 'plants' , 'night' , 'static' , 'architecture' , and 'animals' . Additionally, each image receives ratings (high or low quality) from ten independent viewers. Considering user subjective preferences, this paper assumes comparisons within the same category are more detailed, and therefore conducts experiments both within individual categories and across the entire dataset.

AVA, collected by Murray et al. [23], is commonly used for aesthetic visual analysis experiments. The website contains over 250,000 images with detailed ratings from dpchallenge.com [25]. This paper uses a subset of 76,005 images (due to the large size of AVA, we randomly downloaded from dpchallenge.com) and adopts the method of Lv et al. [20], treating images with average scores above 5 as high quality and below 5 as low quality. For each dataset, we randomly split it ten times, using half for training and half for testing.

3.2 Experimental Setup

Given several test images, this paper conducts two experiments. For the first experiment, the system randomly selects several images from the entire test set and presents them to users, who are then asked to browse and rank the images according to preference from highest to lowest. Simultaneously, the system inputs the same batch of images for ranking. Finally, we compare the user's actual ranking with the system' s predicted ranking to validate algorithm effectiveness. For the AVA dataset, we compare with the current best-performing method [22]. For the second experiment, we evaluate the algorithm from the perspective of traditional binary classification and compare classification accuracy with binary classification methods from literature [2-4,12,13,17,19].

3.3 Experiment 1: User-Specific Aesthetic Ranking

3.3.1 User Study This section focuses on obtaining subjective user evaluations through organized user surveys to assess algorithm effectiveness. The authors invited 20 users aged 23-53 to participate. During the experiment, users were asked to select several preferred images from system-randomized presentations—these images served as user-specific images. Under the premise of ensuring efficient selection and avoiding visual fatigue, extensive experiments were conducted with different quantities of user-specific images (5, 10, 15, 20, 25). This section discusses the average precision values for these five groups. Results for different categories and quantities are shown in .

Average precision for different quantities of specific images on CHHKPQ dataset

To explore the relationship between user subjective preferences and their external manifestation (i.e., the number of user-specific images selected), we conducted user surveys as described in Section 3.3.1. Related average precision values are shown in . For each category in the CUHKPQ database, the system randomly presented 20 images for user selection. When users selected only 5 user-specific images for learning, average precision was highest and prediction performance optimal. During selection, we observed that when user-specific images were limited to a small range (e.g., 5), users tended to choose visually expressive images that were easy to rank. We infer this phenomenon can be explained by two factors: (a) the CUHKPQ dataset used in experiments is large (training and test set sizes: 'animals, 1621 [Figure 1621: see original paper], 1624' , 'architecture, 990, 895' , 'people, 1568, 1570' , 'landscape, 1397, 1381' , 'night, 854, 854' , 'plants, 1198, 1199' , 'static, 1265, 1267'), providing rich and diverse options for user-specific training set generation and enabling learning of reliable customized hyperplanes; (b) the aforementioned selection constraints also affect prediction results. When user-specific image quantity is limited, users tend to select images that best match their preferences, making the batch more representative of user preferences. Using the 'animals' category as an example, [Figure 3: see original paper] shows different individuals selecting 5 and 15 query images. As illustrated, User 1 even selected five animal categories within 5 images. The significant difference in user-specific images makes selecting 5 queries yield higher average precision. Additionally, we observed that when allowed to select more user-specific images (15 or 20), subjects tended to choose images with similar visual content. In [Figure 5: see original paper], both User 1 and User 2 selected images containing sky, water, and similar composition. The similarity among user-specific query images also enables better retrieval performance.

Notably, results when users selected 15 specific images also warrant attention. During selection, we observed that most users focused more on the quality of higher-ranked images and less on lower-ranked images (post-selection user-ranked high-quality images were very high quality, while lower-ranked images were noticeably inferior). The key question becomes: which set of user-specific images better reflects user preferences? As shown in , 15 user-specific images achieved the highest average precision in 'people' and 'landscape' categories (0.9075 and 0.7382 respectively), while 5 specific images performed best in 'animals' , 'architecture' , 'night' , and 'plants' categories with average precisions of 0.7143, 0.7169, 0.7483, and 0.7439 respectively. We believe different quantities of user-specific images should be selected for different categories to achieve optimal results.

3.3.3 AVA Database Results and Analysis shows the predicted average precision of models learned with different quantities of user-specific images on the AVA dataset and comparison with state-of-the-art [22]. Since AVA lacks categories and user-selected images are mostly mixed across categories, this experiment can be viewed as a multi-category high-low quality classification task incorporating user subjective preferences—images users like are considered high quality, disliked images low quality. As shown in , similar to CUHKPQ results, selecting 5 user-specific images yields the most accurate prediction (70.15%), followed by 15 images (58.58%), and finally 10 images (58.26%). Because AVA images are not categorized but rather scored by random viewers, selecting personal preference images without a baseline is ambiguous for users—this represents future work. While learning user preferences within the same category has value, cross-category preference learning better reflects real-world scenarios due to diverse user preferences. The fifth row of shows Lee et al.' s [22] work, which achieved favorable average precision (77.03%) using DCNN feature encoding for high-low quality classification. Although our average precision is lower than [20], we still achieved a relatively high precision of 70.15% while incorporating user preferences—a consideration absent in prior work.

3.4 Experiment 2: Image Quality High-Low Classification Accuracy

To validate the reliability of the proposed user-specific ranking model, we conducted traditional high-low quality classification experiments. Since we focus more on personalized user ranking, user preferences must be considered in high-low quality classification. During Experiment 2, user-specific rankings were divided into four levels: "very good" , "good" , "bad" , and "very bad" . Images marked "very good" or "good" were treated as high quality, while "bad" and "very bad" as low quality. Classification accuracy for each category in the CUHKPQ database is shown in . compares the accuracy of our user-specific ranking model with previous work [2-4,17] and current state-of-the-art methods [12,13,21]. As shown in , compared with cutting-edge work [21] (82.41%), [16] (80.38%), and [13] (80.28%), our accuracy reaches 70.9%. Although not specifically designed for high-low quality classification, our user-specific ranking model still achieves commendable results.

4 Conclusion

This paper proposes a novel user-specific personalized preference ranking model. By inputting user-preferred photos and constructing user-specific training sets, the model learns personalized ranking functions. By adjusting the quantity of user-specific photos and training set size, the model can better capture user subjective preferences while avoiding aesthetic fatigue.

References

[1] Hanghang T, Mingjing L, Hong J Z, et al. Classification of digital photos taken by photographers or home users [C]// Proc of Pacific Rim Conference on Advances in Multimedia Information Processing. 2004: 198-205.

[2] Ritendra D, Dhiraj J, Jia L, et al. Studying aesthetics in photographic images using a computational approach [C]// Proc of European Conference on Computer Vision. 2006: 288-301.

[3] Yiwen L, Xiaoou T. Photo and video quality evaluation: focusing on the subject [C]// Proc of European Conference on Computer Vision. 2008: 386-399.

[4] Luca M, Florent P, Diane L, Gabriela C. Assessing the aesthetic quality of photographs using generic image descriptors [C]// Proc of International Conference on Computer Vision. 2011: 1784–1791.

[5] Tunc O, Aljoscha S, Markus G. Automated Aesthetic Analysis of Photographic Images [J]. IEEE Trans on Visualization and Computer Graphics, 2015, 21 (1): 31-42.

[6] Dhar S, Ordonez V, Berg T L. High level describable attributes for predicting aesthetics and interestingness [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2011: 1657-1664.

[7] Kuo Yen L, Keng Hao L, Chu Song C. Assessment of photo aesthetics with efficiency [C]// Proc of International Conference on Pattern Recognition. 2012: 2186-2189.

[8] Luo W, Wang X, Tang X. Content-based photo quality assessment [C]// Proc of International Conference on Computer Vision. 2011: 2206-2213.

[9] Xin L, Zhe L, Hailin J, et al. RAPID: rating pictorial aesthetics using deep learning [C]// Proc of ACM International Conference on Multimedia. 2014: 457-466.

[10] Zhe D, Xu S, Houqiang L, et al. Photo quality assessment with DCNN that understands image well [C]// Proc of International Conference on MultiMedia Modeling. 2015: 524-535.

[11] Alex K, Ilya S, Geoffrey H. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information

Processing Systems. 2012: 1097-1105.

[12] Xinmei T, Zhe D, Kuiyuan Y, et al. Query-dependent aesthetic model with deep learning for photo quality assessment [J]. IEEE Trans on Multimedia, 2015, 17 (11): 2035–2048.

[13] 王伟凝，王励，赵明权，等．基于并行深度卷积神经网络的图像美感分类 [J]．自动化学报，2016, 42 (6): 904-914.

[14] Deepa C, Vijendran A S. TAD-RR: tag-annotation based demand reranking approach for personalized image retrieval [C]// Proc of International Conference on Intelligent Computing Applications, Coimbatore. 2014: 440–445.

[15] Yeqi L, Yao S, Minyi G. Tag-based personalized image ranking in event browsing [J]. Peer-to-Peer Networking and Applications, 2013, 6 (4): 434-447.

[16] Amandianeze N, Tsuhan C. QUOTE: querying users as oracles in tag engines—a semi-supervised learning approach to personalized image tagging [C]// Proc of IEEE International Symposium on Multimedia. 2016: 30–37.

[17] Joachims T. Training linear SVMs in linear time [C]// Proc of International Conference on Knowledge Discovery and Data Mining. 2006: 217-226.

[18] Yan K, Xiaoou T, Feng J. The design of high-level features for photo quality assessment [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006: 419-426.

[19] Che H Y, Yuan C H, Brian A B, et al. Personalized photograph ranking and selection system [C]// Proc of International Conference on Multimedia. 2010: 211-220.

[20] Lyu Hao, Tian Xinmei. Learning relative aesthetic quality with a pairwise approach [C]// Proc of International Conference on Multimedia Modeling. 2016: 493-504.

[21] Mavridaki E, Mezaris V. A comprehensive aesthetic quality assessment method for natural images using basic rules of photography [C]// Proc of IEEE International Conference on Image Processing. 2015: 887–891.

[22] Lee H J, Hong K S, Kang H, et al. Photo aesthetics analysis via DCNN feature encoding [J]. IEEE Trans on Multimedia, 2017, PP (99): 1-1.

[23] Luca M, Naila M, Florent P. AVA: A large-scale database for aesthetic visual analysis [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2012: 2408–2415.

[24] Hsiao H S, Tse W C, Chieh C K, et al. Scenic photo quality assessment with bag of aesthetics-preserving features [C]// Proc of International Conference on Multimedia. 2016: 1213–1216.

[25] http://www.dpchallenge.com [EB/OL].

Note: Figure translations are in progress. See original paper for figures.