Learning Relative Aesthetic Quality with a Pairwise Approach

Hao Lv and Xinmei Tian(✉)

University of Science and Technology of China, Hefei 230027, Anhui, China
haolv@mail.ustc.edu.cn, xinmei@ustc.edu.cn

Abstract. Image aesthetic quality assessment is very useful in many multimedia applications. However, most existing researchers restrict quality assessment to a binary classification problem, which is to classify the aesthetic quality of images into "high" or "low" category. The strategy they applied is to learn the mapping from the aesthetic features to the absolute binary labels of images. The binary label description is restrictive and fails to capture the general relative relationship between images. We propose a pairwise-based ranking framework that takes image pairs as input to address this challenge. The main idea is to generate and select image pairs to utilize the relative ordering information between images rather than the absolute binary label information. We test our approach on two large scale and public datasets. The experimental results show our clear advantages over traditional binary classification-based approach.

Keywords: Aesthetic quality · Binary classification · Relative ranking

1 Introduction

Image aesthetic quality assessment is a hot research topic, and has drawn much attention recent years. It is a useful technique in many real-word applications. For example, image search engine can incorporate aesthetic quality to refine its search results. Photo management system should consider aesthetic quality as an important factor when ranking photos for users. Hence, users can more easily select the photos with better aesthetic quality.

Most researchers focus their attention on aesthetic quality classification problem, which is to predict whether an image is of "high" or "low" aesthetic quality [1,2,4,5,8,11–14,17,18]. They have spent a lot of efforts on extracting effective aesthetic features, from low-level features [2,18], high-level features [4,8,12,13] to generic features [5,11,14,17]. Despite different ideas and approaches to extract aesthetic features, they share the same thought on training the binary classification model, which is to learn the mapping from aesthetic features to binary aesthetic labels of images. They utilize the absolute binary label information of images, but ignore the relative ordering information between different images.

However, the binary aesthetic labels they predicted are restrictive and unnatural. As shown in Fig. 1, it is hard to decide whether Fig. 1(b) is of "high" or

Q. Tian et al. (Eds.): MMM 2016, Part I, LNCS 9516, pp. 493–504, 2016.
DOI: 10.1007/978-3-319-27671-7_41

(a) "low" (2.51) (b) "?" (5.0) (c) "high" (8.0)

Fig. 1. The aesthetic scores (collected from popular photo sharing website DPChallenge.com [6] and scored by many different users) and labels of images. It is unnatural and restrictive to describe image quality with binary label, since it is hard to decide whether the quality of (b) is "high" or "low". However, we can express quality of (b) in a more informative and natural way: (b) is more aesthetically pleasing than (a) while less beautiful than (c).

"low" quality. However, we can describe the quality of Fig. 1(b) in a more general and natural way: Fig. 1(b) is more beautiful than Fig. 1(a) but less beautiful than Fig. 1(c). In this work, we propose to model the relative aesthetic quality, which is to focus on relative aesthetic quality ranking. It is of great practical significance, since relative comparison is a more natural way for people to describe and compare objects in real life.

To address the relative aesthetic quality ranking problem, existing methods can also estimate a probability of the learned binary classifiers prediction, which indicating the absolute aesthetic quality of an image. However, they suffer from same limitation during training. The aesthetic quality of training images is restricted to be binary, "high" or "low", which is not precise or natural. For example, it is not so reasonable to assign Fig. 1(b) with an aesthetic label of "high" or "low". Thus, this binary label description of aesthetic quality may introduce "noisy" information, while the relative supervision is more precise. For example, it is easier to define and agree on, "Is this image more beautiful than the other?" than "What the absolute aesthetic quality does this image has?". Thus, we expect the relative supervision to be more natural and precise.

How do we learn relative aesthetic quality? We propose a ranking framework based on a pairwise approach to address this problem. Traditional binary classification models learn the classifiers by utilizing the absolute binary label information. In contrast, our goal is to learn the relative ordering relationship of images with different aesthetic quality. The main idea of our approach is to capture the relative relationship of training images by generating and selecting training image pairs. We generate training image pairs which consist of images with different aesthetic quality. Furthermore, considering that not all pairs generated are useful, we select certain pairs based on proposed rules. The selecting process acts as a filter and filters out "noisy" pairs. The selected pairs contain more useful and precise relative information, which are important for improving the performance of our ranking framework. The way we generating pairs and selecting pairs is not only easy to understand but also very effective, which is verified in our experiments. We then adopt a ranking model that takes image pairs as input to learn a ranking function. It will estimate ranking scores for testing images. The ranking scores are used for ordering images only, which have no meaning in absolute sense.

In summary, in this paper we focus on relative aesthetic quality while most existing works are committed to traditional binary classification problem. To address relative aesthetic quality problem, we propose a pairwise-based ranking framework. We generate and select image pairs that contain relative order information between training images, which is essential for improving the performance of proposed ranking framework. The experiments on two large scale and public datasets show that our pairwise approach significantly outperforms the binary classification-based approaches.

The remainder of this paper is organized as follows. We review related works in Sect. 2 and describe the details of our proposed pairwise-based approach in Sect. 3. Then we evaluate the performance of our approach in Sect. 4. Finally, we conclude and discuss the future work in Sect. 5.

2 Related Work

In this section, we first review related works on aesthetic quality classification, and then discuss works that concern about the relative ranking problem.

Aesthetic Quality Classification. Many image aesthetic quality assessment approaches have been proposed in recent years. However, most existing approaches focus on aesthetic quality classification. They share the same thought on training binary classification model and they spend a lot of efforts on designing different aesthetic features. Roughly, these methods can be divided into three categories: low-level feature-based approaches, high-level feature-based approaches and generic feature-based approaches.

Low-level feature-based approaches extract a set of low-level features that are commonly used in computer vision tasks [2,18]. Tong et al. extracted blurriness, colorfulness, saliency value and so on [18]. They achieved limited success because of these features are not specially designed for the aesthetic quality of images. Datta et al. designed a set of low-level features, which are related to user intuition and some photography literature, i.e. "rule of thirds", "simplicity" and "interestingness" [2]. After carefully designed, they extracted 56-dim features and obtained a better performance.

High-level feature-based approaches focus on designing high-level features based on photography and psychology literature [4,8,12,13]. Dhar et al. proposed a set of attribute-based predictors to conduct aesthetic quality evaluation [4]. Luo et al. extracted different features for different categories of photos and then generated category-specific classifiers [12]. Luo and Wang designed features mainly describing the image composition and relationship between subject region and background region [13].

Generic feature-based approaches extracted a large set of image features, which are used to describe image content [14,17]. Marchesotti et al. extracted generic image content descriptors to conduct aesthetic quality classification and gained certain improvement [14]. Lu et al. applied three schemes to incorporate deep learning with aesthetic quality assessment, and obtained improved performance [11]. Dong et al. directly adopted the deep neural network trained on ImageNet [3] and extracted the 4096-dim output activations of the seventh layer as aesthetic features, and achieved remarkable success [5].

Relative Ranking. Many researchers focus their attention on relative ranking problems of images recent years. Kumar et al. proposed comparative facial attributes for face verification [10]. The attributes they explored are similarity-based. Wang et al. learned fine-grained image similarity with deep ranking model [19]. They proposed a deep ranking network and an efficient triplet sampling algorithm to address the fine-grained image similarity. Parikh and Grauman devised a ranking framework to learn ranking functions for image attributes, given relative similarity constraints on pairs of examples [16]. Based on relative attributes, a novel form of zero-shot learning and image describing experiments were conducted. Significant improvement was obtained by relative attributes-based approach compared with traditional binary classification-based approach. Inspired by the work related to relative ranking, we propose a pairwise-based ranking framework to address relative aesthetic quality ranking problem.

3 Relative Aesthetic Quality Ranking

Existing binary classifier-based approaches utilize absolute binary label information of training images. Unlike existing approaches, the intention of our pairwise approach is that we want to utilize the relative ordering information between training images. The architecture of proposed pairwise-based ranking framework is shown in Fig. 2. Given a set of training images, we first generate and select image pairs based on certain rules. Then we feed these selected pairs into the ranking model to learn a ranking function. During testing stage, the ranking function will estimate real-value ranking scores for images, which are used for ordering examples. In this section, we present the pairwise-based ranking model (Sect. 3.1) and explain the details of our training image pairs generation (Sect. 3.2) and selection (Sect. 3.3).

3.1 Pairwise-Based Ranking Model

We are given a set of training images $\mathcal{I} = \{I_i\}, i = 1, 2, \ldots, m$, represented in $\mathbb{R}^n$ by feature-vectors $\{x_i\}$, a set of aesthetic quality labels $\mathcal{A} = \{a_i\}, a_i \in \{0, 1\}$, and a set of image class labels $C = \{c_i\}, c_i \in \{1, 2, \ldots, C\}$. The aesthetic label "0" is for "low quality" and "1" is for "high quality". The image class labels describe the semantic content of images. Based on certain rules (described in Sects. 3.2 and 3.3), we generate a set of ordered image pairs denoted as $O = \{(I_i, I_j)\}$,

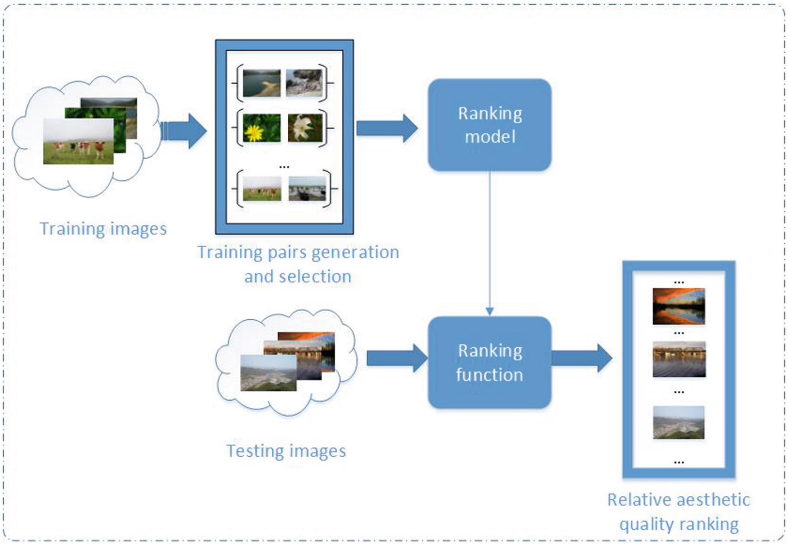

Fig. 2. The architecture of proposed pairwise-based ranking framework. We capture the relative information by training the ranking model with generated and selected image pairs.

pairs in which satisfied $(I_i, I_j) \in O \Rightarrow a_i > a_j$. Our goal is to learn a ranking function:

$$r(x_i) = w^t x_i, \tag{1}$$

by maximizing the number of the following constraints satisfied:

$$\forall (I_i, I_j) \in O : w^t x_i > w^t x_j. \tag{2}$$

This leads to an optimization problem:

$$\begin{aligned} \min \quad & (\frac{1}{2} \|w\|_2^2 + C \sum \xi_{ij}^2) \\ s.t. \quad & \forall (I_i, I_j) \in O : w^t x_i \geq w^t x_j + 1 - \xi_{ij}, \\ & \xi_{ij} \geq 0, \end{aligned} \tag{3}$$

where C is the trade-off constant.

Problem (3) can be reformulated as:

$$\begin{aligned} \min \quad & (\frac{1}{2} \|w\|_2^2 + C \sum \xi_{ij}^2) \\ s.t. \quad & \forall (I_i, I_j) \in O : w^t (x_i - x_j) \geq 1 - \xi_{ij}, \\ & \xi_{ij} \geq 0. \end{aligned} \tag{4}$$

We solve this problem using SVM^{rank} [7] with linear kernel. This learning-to-rank model explicitly enforces a desired ordering on training examples. When given a set of testing images, we apply learned ranking function to estimate real-value ranking scores for images. The ranking scores are used to order the testing examples only. The absolute values of which have no practical significance.

The aesthetic features we extracted is 4096-dim normalized output of the seventh layer of deep convolutional neural network (DCNN) designed by Krizhevsky et al. [9]. The DCNN has achieved great success in many computer vision tasks, e.g. image classification, due to its strong ability to describe the content of images. It also obtained remarkable success on aesthetic quality classification task as reported in [5,11]. It consists of eight layers in total, the first five layers are convolution layers and the last three layers are fully connected layers. The details can be referred in [9]. We extract the DCNN features as our aesthetic features, which is well-suited to the task as hand. We feed the raw RGB image to the DCNN framework, and take normalized 4096-dim output activation of the seventh layer as our aesthetic features.

3.2 Training Pairs Generation

When facing the challenge of relative aesthetic quality ranking, it is not enough to train just a binary classifier, despite that the learned classifier can also estimate a score indicating the absolute strength of images aesthetic quality. The main limitation is that the binary classifier-based approaches ignore the relative ordering information between images during training process.

To overcome the limitation shared by existing approaches, we focus on generating effective and informative image pairs to capture the relative ordering information. Considering that images with different aesthetic quality are potential to contain relative information, we generate all the possible image pairs in the training set. We generate all the possible pairs that consist of images with different aesthetic quality, which means that an image from class of "high" quality will form pairs with all images from class of "low" quality. Then the generated image pairs are denoted as:

$$O = \{(I_i, I_j) | a_i > a_j\} \tag{5}$$

Our strategy of generating pairs is easy to understand and effective, which is verified in our experiments. We feed the generated pairs to the ranking model to learn the ranking function.

3.3 Informative Training Pairs Selection

Images in the datasets are from different image classes. For example, the CUHKPQ dataset consists of seven categories (please refer Sect. 4 for detail). It has to be noticed that not all images are comparable, i.e. comparison between an image on animal and an image on architecture does not make much sense. The method described above generates all the possible pairs, which means that they contain a large set of image pairs consisting of images from different categories.

Image pairs consisting of images with similar content are much more reasonable and comparable. It is more natural to compare the aesthetic quality of images both on landscape. The ranking framework can benefit a lot from the image pairs that contain useful and comparable relative information. Therefore, it is important to keep the comparable pairs while wipe out the others. Based on this consideration, we take a selection step to reserve the comparable pairs, and the selected pairs are denoted as:

$$O_s = \{(I_i, I_j)|a_i > a_j, c_i = c_j\} \tag{6}$$

Compared with the method in Sect. 3.2, we put constraint on the image class labels during the selection step. Images from the same category are more likely to have similar content, which are more reasonable to be compared. Only image pairs consisting of images from the same category are reserved, while others are considered as "noisy" pairs and wiped out. After selection, the number of pairs are largely reduced, and the selected pairs contain less noisy information. Then we feed the selected pairs into the ranking model.

4 Experiments

In this section, we test our proposed approaches on two public datasets, CUHKPQ and a subset of AVA. Two datasets are widely used in aesthetic quality assessment field, and both contain considerable quantity of photos. We implement the-state-of-art approach as baselines [5]. We compare the performance of our approach with the baselines to verify the effectiveness of our proposed pairwise-based ranking framework.

4.1 Datasets

CUHKPQ. CUHKPQ dataset is released by Luo et al. [12]. It consists of photos downloaded from professional photography sharing websites and photos contributed by amateurs. It is divided into seven categories according to photo content. Each photo in the dataset is evaluated by ten independent viewers. Each photo is assigned with an aesthetic label "high" or "low" under the condition that eight out of ten viewers share same opinion on its assessment. It contains a total of 17690 photos from seven categories. We randomly and evenly divide it into training set and testing set, each with 8845 photos. We focus on the problem of relative aesthetic ranking. Images with similar content are more reasonable to be compared as explained in Sect. 3. Based on this consideration, during testing, we restrict relative comparison between images within the same category. Therefore, we compare the predicted relative order with original relative order within the same category to obtain test performance on seven categories.

AVA. DPChallenge.com [6] is one of the most active and popular photo sharing communities on the Internet. Users in this community are from different levels of photography enthusiasts. There are a variety of photographic challenges in the community, each defined by a title and short description. Users can upload their photos according to a specific photographic challenge. Other users can score and comment on the uploaded images. AVA is collected by Murray et al. [15] from DPChallenge.com. It consists of more than 250000 images with different tags indicating the semantic content of images. The number of aesthetic scores each image received is range from 78 to 549, with an average of 210. Average aesthetic score of the image is taken as the ground-truth value.

Murray et al. only offer the web links of images and some of them are invalid because of the update of their website. We successfully downloaded 193077 images. As aforementioned, we restrict the relative aesthetic quality comparison within the same category during testing. We extract nine categories with largest number of images in the 193077 images. We randomly and evenly divide it into training set and testing set. We adopt the same criteria to assign each image an aesthetic label "high" or "low" as reported in [15]. Images with mean score larger than or equal to $5 + \delta$ are defined as "high" quality images while those with mean score smaller than or equal to $5 - \delta$ are "low" quality images. Others are discarded. In this paper, we set $\delta = 1$. Under this setting, we have 21116 images for training and 21117 images for testing.

4.2 Experimental Settings

We implement the approach based on binary classification as baseline [5]. The aesthetic feature extraction method of baseline is the same as reported in [5]. We adopt the DCNN framework trained on ILSVRC-2012 and take the 4096-dim output activation value of seventh layer as aesthetic features. The widely used machine learning algorithm SVM is trained with the extracted features and aesthetic labels to generate the binary classifier. During testing, baseline approach estimates the probability of the binary classifiers prediction rather than a binary label for each image. We calculate the AP value within same category as the evaluation indicator. The AP value is often used in information retrieval field. We use it here to measure how well the predicted ranking is consistent with the ground truth ranking within the same category. The model parameters for all methods are determined via five-fold cross-validation on training set.

In our proposed pairwise approach, we also use extracted 4096-dim DCNN features as aesthetic features. To implement the method described in Sect. 3.2, we generate all the possible image pairs. We denote this method as "pairwise_gen". Then, we feed all these pairs to SVM^{rank}. We implement the method described in Sect. 3.3, and denote it as "pairwise_sel". We take a selecting step to wipe out some of the pairs generated in method "pairwise_gen". Then we feed the selected pairs to SVM^{rank} to learn the ranking function. Linear kernel is adopted for SVM^{rank} and the model parameters are also determined via five-fold cross-validation on training set.

Table 1. Mean AP value on CUHKPQ dataset with different approaches.

Approach	Mean AP
Baseline	0.811
Pairwise_gen	0.828
Pairwise_sel	**0.879**

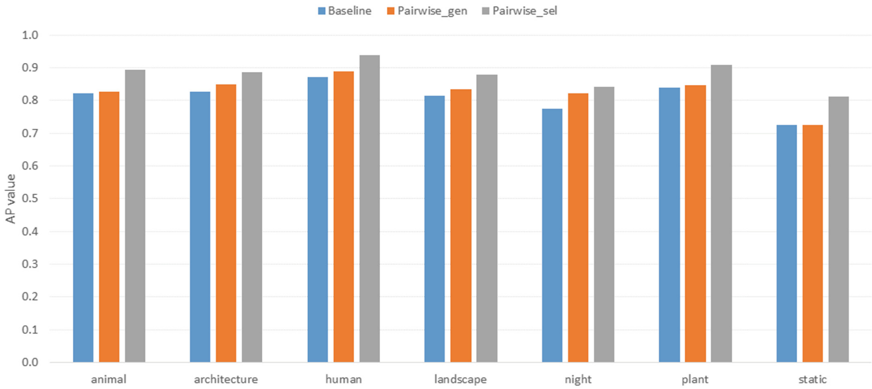

Fig. 3. The AP values on seven categories of CUHKPQ dataset with different approaches.

4.3 Experimental Results on CUHKPQ

In this experiment, we evaluate the performance of binary classification-based approach and our pairwise approaches on CUHKPQ dataset. We present the comparison of mean AP value on seven categories in Table 1 and the details of each category in Fig. 3. Among three methods, our proposed pairwise approach "pairwise_sel" achieves the best performance with mean AP value at 0.879. Method "pairwise_gen" obtains mean AP value at 0.828, while the baseline method reaches mean AP at 0.811. Although we generated image pairs in an easy way in method "pairwise_gen", we still obtain a better performance than the binary classification-based method. As shown in Fig. 3, method "pairwise_gen" performs better than baseline method on all seven categories, which indicates the robustness of our pairwise approach. Moreover, the method "pairwise_sel" significantly outperforms the method "pairwise_gen", which demonstrates that our pairwise-based ranking framework can benefit a lot from the training image pairs selection step.

The experimental results on CUHKPQ dataset show the advantage of our proposed pairwise approaches on relative aesthetic quality ranking. Although the binary classification-based method achieves an acceptable result, our proposed method "pairwise_gen" improves the performance by using a pairwise approach. Whats more, with a selecting step, proposed method "pairwise_sel" outperforms the baseline with a larger margin.

Table 2. Mean AP value on AVA dataset with different approaches.

Approach	Mean AP
Baseline	0.470
Pairwise_gen	0.531
Pairwise_sel	**0.611**

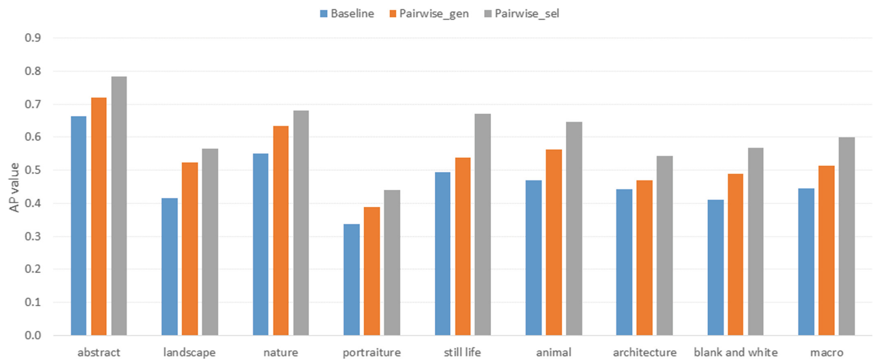

Fig. 4. The AP values on nine categories of the subset of AVA dataset with different approaches.

4.4 Experimental Results on AVA

We present the experimental results of three methods on the subset of AVA dataset in Table 2 and Fig. 3. The mean AP value over nine categories achieved by our method "pairwise_sel" is 0.611, which is the best result. The method "pairwise_gen" achieves a better result than baseline. Our proposed method "pairwise_sel" improves the performance of method "pairwise_gen" and baseline with a large margin, which indicates the effectiveness of proposed selecting step. The details on each category are shown in Fig. 4. Method "pairwise_gen" consistently outperforms baseline on all nine categories. Method "pairwise_sel" obtains better performances on nine categories over other two methods. The results on nine categories show the clear advantages of our proposed pairwise approaches.

Compared with baseline, we improves the performance by using a pairwise approach. The improvement shows the advantage of pairwise-based ranking framework at capturing the relative ranking information of images. We obtained an even larger improvement when taking a selecting step on image pairs generated in method "pairwise_gen". This verifies the contribution of proposed selecting step, which is to wipe out "noisy" pairs.

5 Conclusion and Future Work

Inspired by that it is more natural to model the relative aesthetic quality than absolute binary labels, we aim to study the aesthetic quality ranking rather than traditional aesthetic quality classification. In particular, we have proposed a pairwise-based ranking framework, which takes image pairs as input. In order to better capture the relative ordering information, we have proposed certain rules to generate and select training image pairs. We took the DCNN features as aesthetic features and SVM^{rank} as our rank model. The experimental results revealed that the proposed pairwise approach could capture the relative information of images better than traditional binary classification approach. The proposed selection step helped to wipe out "noisy" pairs and improved the performance. Despite the encouraging results achieved, this is just an attempt to study the relative aesthetic ranking problem, and there are still many open challenges. In the future, we will investigate more effective ways and more powerful models to utilize the relative order information.

Acknowledgement. This work was supported by the 973 project under Contract 2015CB351803, by the NSFC under Contracts 61390514 and 61201413, by the Youth Innovation Promotion Association CAS No. CX2100060016, by the Fundamental Research Funds for the Central Universities No. WK2100060011 and No. WK2100100021, and by the Specialized Research Fund for the Doctoral Program of Higher Education No. WJ2100060003.

References

1. Bhattacharya, S., Sukthankar, R., Shah, M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In: Proceedings of the international conference on Multimedia, pp. 271–280. ACM (2010)
2. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
4. Dhar, S., Ordonez, V., Berg, T.L.: High level describable attributes for predicting aesthetics and interestingness. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1657–1664. IEEE (2011)
5. Dong, Z., Shen, X., Li, H., Tian, X.: Photo quality assessment with DCNN that understands image well. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) MMM 2015, Part II. LNCS, vol. 8936, pp. 524–535. Springer, Heidelberg (2015)
6. DPChallenge: Dpchallenge. http://www.dpchallenge.com/
7. Joachims, T.: Training linear svms in linear time. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 217–226. ACM (2006)
8. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 419–426. IEEE (2006)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

10. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 365–372. IEEE (2009)
11. Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J.Z.: Rapid: rating pictorial aesthetics using deep learning. In: Proceedings of the ACM International Conference on Multimedia, pp. 457–466. ACM (2014)
12. Luo, W., Wang, X., Tang, X.: Content-based photo quality assessment. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2206–2213. IEEE (2011)
13. Luo, Y., Tang, X.: Photo and video quality evaluation: focusing on the subject. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 386–399. Springer, Heidelberg (2008)
14. Marchesotti, L., Perronnin, F., Larlus, D., Csurka, G.: Assessing the aesthetic quality of photographs using generic image descriptors. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1784–1791. IEEE (2011)
15. Murray, N., Marchesotti, L., Perronnin, F.: Ava: a large-scale database for aesthetic visual analysis. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2408–2415. IEEE (2012)
16. Parikh, D., Grauman, K.: Relative attributes. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 503–510. IEEE (2011)
17. Su, H.H., Chen, T.W., Kao, C.C., Hsu, W.H., Chien, S.Y.: Scenic photo quality assessment with bag of aesthetics-preserving features. In: Proceedings of the 19th ACM International Conference on Multimedia, pp. 1213–1216. ACM (2011)
18. Tong, H., Li, M., Zhang, H.-J., He, J., Zhang, C.: Classification of digital photos taken by photographers or home users. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3331, pp. 198–205. Springer, Heidelberg (2004)
19. Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., Chen, B., Wu, Y.: Learning fine-grained image similarity with deep ranking. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1386–1393. IEEE (2014)