# Photo Quality Assessment with DCNN that Understands Image Well

Zhe Dong, Xu Shen, Houqiang Li, and Xinmei Tian

University of Science and Technology of China, Hefei, Anhui, 230027, China
{ustcdz,shenxu}@mail.ustc.edu.cn, {lihq,xinmei}@ustc.edu.cn

**Abstract.** Photo quality assessment from the view of human aesthetics, which tries to classify images into the categories of good and bad, has drawn a lot of attention in computer vision field. Up to now, experts have proposed many methods to deal with this problem. Most of those methods are based on the design of hand-crafted features. However, due to the complexity and subjectivity of human's aesthetic activities, it is difficult to describe and model all the factors that affect the photo aesthetic quality. Therefore those methods just obtain limited success. On the other hand, deep convolutional neural network has been proved to be effective in many computer vision problems and it does not need human efforts in the design of features. In this paper, we try to adopt a deep convolutional neural network that "understands" images well to conduct the photo aesthetic quality assessment. Firstly, we implement a deep convolutional neural network which has eight layers and millions of parameters. Then to "teach" this network enough knowledge about images, we train it on the ImageNet which is one of the largest available image database. Next, for each given image, we take the activations of the last layer of the neural network as its aesthetic feature. The experimental results on two large and reliable image aesthetic quality assessment datasets prove the effectiveness of our method.

**Keywords:** Image aesthetics, photo quality assessment, deep convolutional neural network.

## 1 Introduction

With the fast development of Internet, the amount of online images has gained an explosive growth that easily breaks through the magnitude of billions. What's more, even general consumers can also easily take a lot of digital photos as the high-performance capture devices become cheaper and more popular. According to this, new methods are needed to manage photos online or offline better and more intelligently.

Photo quality assessment from the perspective of human aesthetics attempts to classify images into good and bad. Fig. 1 shows an example. Most people will prefer the left four images as they are more beautiful. With effective photo aesthetic quality assessment method, image retrieval system can return images not only related to the given queries but also with high quality. This will surely provide better

X. He et al. (Eds.): MMM 2015, Part II, LNCS 8936, pp. 524–535, 2015.

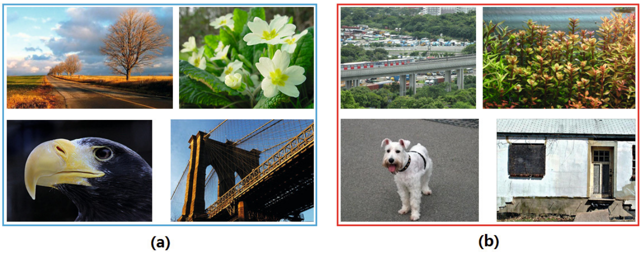


**Fig. 1.** It may be easy for most people to agree on that the left four images in (a) look more beautiful than the right four images in (b). In other words, they have higher aesthetic quality.

user experience according to the user investigation conducted in [1]. Offline users who have a large image collection can also benefit from this method as it can help to select a small number of representative and beautiful photos automatically [2]. In addition, photo aesthetic quality assessment also helps to develop new image enhancement tools to make images look better [3, 4].

Researchers have made a lot of efforts to conduct photo quality assessment [3, 5, 6, 7, 8, 9, 10, 14, 15, 16, 24, 25]. They try to find out image attributes which are related to image aesthetic quality and describe them with mathematical model that can be handled by computers to extract features. The way by which they select the image attributes is mainly based on the user intuition and photography knowledge, such as the rule of thirds [3, 5, 6], colorfulness [5, 6, 7, 8], simplicity [7, 8] and so on.

But this kind of method suffers from several drawbacks. Firstly, we cannot point out all attributes that are related to image aesthetic quality assessment. In fact, it is hard to discover attributes that can affect image aesthetic quality. Researchers just adopt a small number of them which are well-known and easy to be implemented. Secondly, it is hard to explain how those image attributes affect the image aesthetic quality. For example, we may agree on that image color will affect the image aesthetic quality a lot. Professional photographers carefully set the image color to gain better visual effect. But there is no fixed discipline that the color scheme of beautiful images must follow. Both images with simple color palette and images which are colorful can have high aesthetic quality. Thirdly, it is not easy to describe the image attribute accurately with mathematical model even we are sure that the attribute can affect the image aesthetic quality. For example, sharpness is one of the most popular image attributes used in the research of image aesthetic quality assessment [6, 7, 8, 9, 14]. But the measure of sharpness itself is a hard problem which has already been studied for a long time and there still exists no method that can assess sharpness properly on all kinds of photos. When turning to more abstract image attributes such as simplicity and image composition, the problem becomes more complex and challenging.

To avoid those problems, some experts propose to adopt general image descriptors which are used to describe image content to do image aesthetic quality assessment [10, 25]. Those general image descriptors have gained remarkable success in the general image classification problem which tries to classify images into different categories according to their content [11]. As images look alike may also have similar image aesthetic quality, general descriptors such as dense SIFT also gain acceptable performance in this higher level computer vision problem reported by Marchesotti et al. in [10].

Our method is similar to Marchesotti as we also want to avoid the problems listed above that exist in the hand-crafted features. We apply the deep convolutional neural networks [13], which achieve the best performance in general image classification problem, to do the image aesthetic quality assessment. The neural network is firstly trained on part of the ImageNet image database, which contains millions of images with various categories, to "understand" images well. Then this deep network can directly compress the image into a relatively lower dimensional feature vector and meanwhile reserve most information in the image. We speculate that those information may implicitly reflect the aesthetic quality of the image. The experimental results obtained on two large and reliable datasets further confirm our assumption. Our method achieves much better performance compared with state-of-the-art methods.

The rest of this paper is organized as follows. In Section 2, we review related methods. In Section 3, we describe the deep convolutional neural networks and our methods in details. Then the experimental settings, results and comparisons with state-of-the-art methods are presented in Section 4. Finally, we give a conclusion in Section 5.

## 2 Related Work

Photo aesthetic quality assessment has been studied for a long time. In [9], Tong et al. tried to classify images taken by photographers and home users. This was actually the same with photo aesthetic quality assessment. They adopted many low-level features such as color histogram, image energy which were proposed for previous content-based image retrieval system. Those features did not have close relationship with aesthetic quality and just achieved limited success [8].

Subsequently, Datta et al. and Ke et al. proposed different ways from Tong et al. to do photo quality assessment [5, 8]. Both of them tried to firstly analyze which attributes can affect the image aesthetic quality. After this, Datta et al. used many low-level features which were related to those attributes. Ke et al. designed seven new features which all had high-level semantics such as simplicity, colorfulness and so on. Both of their methods performed better in effectiveness and efficiency.

Luo et al. and Wong et al. adopted other image attributes [6, 14]. They found out that most of professional photos carefully set the contrast between the subject and background. Due to this, there existed big differences between the subject that the photographers wanted to capture and the background they chose. So Luo et al. and Wong et al. determined to separate the subject and background first and then dealt with them respectively. They extracted features describing those two regions and the relationship between them.

As we have mentioned above, hand-crafted features suffer from several problems. Though there exist other works which adopt new attributes or refine mathematical models to describe previous attributes [7, 15, 16], limited improvement is gotten. Most of them lack robustness thus they perform badly when handling with different kinds of images or the dataset changes.

Different from those works which try to analyze the image aesthetic quality in the high semantic level, Marchesotti et al. choose another innovative way that they adopt the generic image descriptors such as BOV to represent the image [10]. This kind of descriptors is used to describe the image content previously. We can infer that it is related to the image aesthetic quality in an implicit way.

In recent years, deep learning methods have brought breakthroughs in many traditional computer vision problems. They need no human ingenuity but enough training data. Among them, Krizhevsky et al. used the deep convolutional neural networks (DCNN) to conduct image classification on ImageNet [13]. Their method performed better than previous ones in which hand-crafted features and generic image descriptors were used to represent the image. This deep neural network could describe and encode the image content more effectively. Inspired by this, in this paper, we propose to extract image features with the help of the deep convolutional neural network well trained on the ImageNet database and apply those features to do image aesthetic quality assessment.

## 3 Features Extracted by the Deep Convolutional Neural Network

In general, when facing a new computer vision problem, experts are accustomed to analyze it in a high semantic level and design features to represent the data. Then those features are transferred into the learning algorithm with or without preprocessing such as PCA and so on. Though this scheme can make use of human intelligence inside, it is hard to figure out how the brain perceives the outside world sometimes especially when the given question is very abstract such as photo aesthetic quality assessment we try to deal with here.

Convolutional neural networks, firstly proposed by LeCun et al. and applied in handwritten digit recognition [17], are quite different from other machine learning methods. Fig. 2 shows the architecture of their neural network model. It can directly handle with the raw image instead of feature vector. The structure of convolutional neural network imitates the biological vision system and therefore obtains strong ability to understand the nature of images.

General convolutional neural network applies several architectural thoughts, including the local receptive fields, shared weights and subsampling [18]. Before fed to the network, the given image should be scaled to a fixed size. Then pixels are translated to the range -1$\sim$1 in order to decrease the effect of the image exposure. But the resized image still has many pixels. If a traditional full-connected network with enough discriminative power is applied, too many parameters should be learned. For convolutional network, it applies the local receptive field to avoid this problem. This idea is originated from the survey on

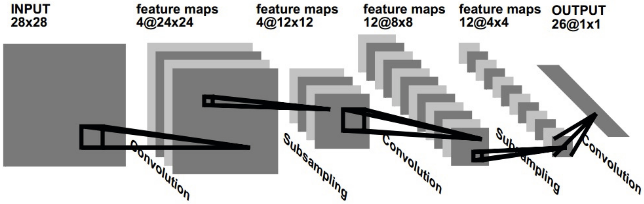


**Fig. 2.** The convolutional neural network designed for hand-written digit recognition proposed in [17]. It can directly deal with the raw image.

the visual system of cats [19]. Hubel et al. finds that visual neurons are locally-sensitive and orientation-selective. The local receptive field in convolutional neural network means for each unit in elementary layers, it only receives inputs from a small set of units in the neighborhood in the previous layer. This mathematical model tactfully imitates the locally-sensitive character of visual neurons and avoids to train too many parameters which may lead to over-fitting.

The setting of shared weights means receptive fields in different places of the image will have the identical weight vectors. This idea is proposed based on the knowledge that feature detector which is useful in one part of the image maybe also work across the whole image. It is equivalent to scan all over the image with a small convolutional kernel followed by a non-linear function. This is also why we call the method convolutional neural network.

Outputs of neurons in the same layer form a plane which we call feature map. We can obtain different feature maps by setting different weights of the local receptive field. Then local operations, such as averaging or maximum selection, and subsampling are conducted all over the feature map. Those operations form the pooling layer. In general, the convolutional layer and pooling layer alternately appear in the convolutional neural network. The output layer is fully connected with its previous layer and produces a feature vector which can further transferred to a logistic regression layer to accomplish the recognition task. All the weights in the network are learned with the back-propagation method.

Due to the application of those innovative and outstanding ideas, convolutional neural networks are adopted in many pattern recognition problems and achieve good performances [20, 21, 22]. But in the past, it is almost impossible to train an artificial neural network with many hidden layers and millions of neurons due to the limitation of computing power and lack of methods to avoid falling into the local optimization and over-fitting. Luckily, recently deep learning methods have gained a speedy development. Those methods improve the traditional artificial neural network in several aspects, such as the unsupervised and layer-wised pre-training, better activation functions and new training methods. Besides, large scale available datasets help to train more powerful models without over-fitting. The application of cheap and high-performance GPU and multi-core computers also greatly reduce the training time.

Due to this, Krizhevsky et al. designs a deep convolutional neural network with millions of parameters and applied it on the ImageNet classification [13]. Their

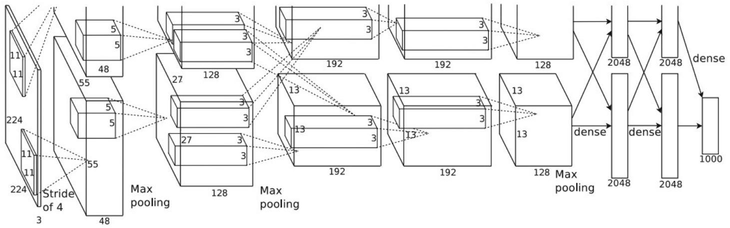


**Fig. 3.** The architecture of deep convolutional neural network used in [13]. It is similar with the neural network in [17], but has more layers and neurons. The first five layers of it are convolutional and the remaining three layers are full connected. It can take the 224*224*3 image patches as input.

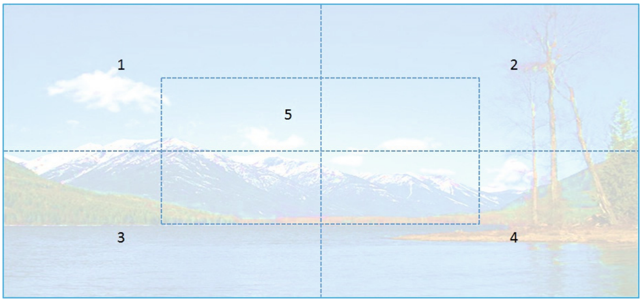


**Fig. 4.** In order to reduce the impact of image scaling and encode the information of image composition, we adopt the idea of spatial pyramid. We segment the image into five regions and each patch is fed into the deep convolutional neural network. All of their activations are concatenated with the activation gotten by the whole image.

model contains eight learned layers and adopts the Rectified Linear Units as the activation function [23]. The overall architecture of their network is shown in Fig. 3. It can handle with the RGB images instead of gray images only. Their neural network consists of five convolutional layers and three full connected layers. The output of the last layer is followed by a 1000-way soft-max that produces a distribution over the given 1000 image class labels. Then supervised learning process is conducted on the training set of the ImageNet database. Finally it achieves a breakthrough on this challenging dataset which proves its descriptive power on all kinds of images.

In view of this, we here adopt the deep convolutional neural network, trained on the ImageNet dataset to understand images well, to represent the image and challenge the problem of photo quality assessment. This task requires to classify images into high quality and low quality which seems to be easier in the question format but require more intelligence inside. But we can still trust that a model which know the image content well may also be powerful in aesthetic quality assessment.

Firstly, we implement a deep convolutional neural network that has the same architecture with [13]. Then this neural network is carefully trained on the ILSVRC-

2012 training set that extracts 1.2 million images covering 1000 categories from the whole ImageNet database. In this way, the convolutional neural network accumulates enough knowledge to understand various images well. We remove the external 1000-way soft-max in the network. For a given image $I_i$, the last hidden layer of the convolutional neural network $C$ produces 4096-dimensional activations. We normalize those activations as DCNN_Aesth features $f_i$ for $I_i$. The image aesthetic quality training set $\mathcal{L}$ will be represented as $\{(f_1, y_1),(f_2, y_2),...,(f_N, y_N)\}$, where the aesthetic quality label $y \in \{-1, 1\}$. Here $y = 1$ denotes high aesthetic quality and $y = -1$ denotes low aesthetic quality. Finally, we train a new classifier $\mathcal{S}$ on $\mathcal{L}$ to predict the aesthetic quality of new images.

As the convolutional neural network requires to scale the image into a fixed size which may lose some information, we further adopt the idea of spatial pyramid on the image. This improvement can also implicitly encode the image composition information inside. To do this, we segment the image $I_i$ into five regions as shown in Fig. 4. Each region $R_j$ is fed into the neural network independently to produce 4096-dimensional feature vector $f_{ij}$. Then those vectors are directly concatenated to form a long feature vector $F_i = \{f_i, f_{i1}, f_{i2}, f_{i3}, f_{i4}, f_{i5}\}$, which is 24576-dimensional, to represent the image $I_i$. We will test the DCNN_Aesth features talked above and those DCNN_Aesth_SP features in following experiments.

## 4 Experiments

In this section, we report the experimental results to verify the effectiveness of deep convolutional neural network when dealing with photo quality assessment. It will be compared with several state-of-the-art methods. Most of them are based on hand-crafted features [5, 6, 7, 8, 10]. All the experiments are conducted on two large scale and reliable public datasets, CUHKPQ and AVA, which are specifically designed for the research of photo quality assessment [24, 25].

### 4.1 Datasets

**CUHKPQ.** To explore the image aesthetics better, a large dataset which is carefully labeled and contains various kinds of images is needed. It is impossible for individual to form a dataset like this. Most experts turn to Internet for help. There exist many photo sharing websites which allow users to upload photos freely and score photos for each other. For example, DPChallenge.com contains hundreds of thousands images, most of which are scored by tens of different users. This makes it reliable to form a dataset used for the survey of image aesthetic quality assessment.

CUHKPQ is a public dataset contains 17690 images collected from the university students and professional photo sharing websites [24]. Each image is labeled as "good" or "bad" by ten independent volunteers and is reserved only when more than eight of them have the same opinions. Therefore, the image labels in CUHKPQ are more reliable and have less noise. Those make it easier to classify images in CUHKPQ into high quality and low quality. Another character of CUHKPQ is that the categories of images in this dataset can be divided into “animal”, “architecture”, “human”, “landscape”, “night”, “static” and “plant”. Those categories are common in natural images that occupy a large percentage of images we see in daily life. In our experiment, we randomly and evenly separate “good” and “bad” images in each category into training and test set. Finally both sets contain 8845 images.

**AVA.** Compared with CUHKPQ, Aesthetic Visual Analysis (AVA) is a larger dataset formed by more than 250 thousands of images [25]. This database is specifically constructed for the purpose of learning more about image aesthetics. All those images are directly downloaded from the DPChallenge.com. Though many other datasets make use of this website such as [6], they only contain a far less number of images. Besides, image categories in this dataset are more abundant including abstract images, advertisement images and so on.

For each image in AVA, there is an associated distribution of scores voted by different viewers. As reported in [25], the number of votes that per image gets is ranged in 78∼549. Besides, users in the DPChallenge.com consist of all kinds of people, without restriction of age, gender and profession. Both professional photographers and amateur enthusiasts can enjoy images shared in this website. All of them have viewed a large amount of photos and can make the independent judgments. So it is reliable to take the average score of the image as its photo aesthetic quality.

Actually, the providers of AVA database only release the web links of those images. We successfully download 193077 images and other image links are invalid due to the update of the website. In our experiments, to increase the gap between images with high aesthetic quality and images with low aesthetic quality, we firstly sort all images by their mean scores. Then we pick out the top 10% images as good and the bottom 10% images as bad as done in other works [5, 8]. Then both good images and bad images are randomly and evenly separated into training set and test set. Due to this, each set contains 19308 images.

### 4.2 Experimental Settings

Here we implement several state-of-the-art methods as comparisons, including the 56-d features proposed by Datta et al. [5], 7-d features proposed by Ke et al. [8], 5-d features proposed by Luo et al. [6], 17-d original features proposed by Lo et al. [7], and bag-of-visual-words features of dense SIFT image descriptor with 1024 visual words proposed by Marchesotti et al. [10]. Among them, the method of Marchesotti is similar with ours which also try to apply features used to describe image content instead of hand-crafted features in aesthetic quality assessment.

Besides, the deep convolutional neural network used in our method is implemented with the help of the open source deep leaning framework called Caffe [27]. We inherit the architecture of DCNN model proposed by Krizhevsky et al. in [13]. The overall structure of this network has already been shown in Fig. 3.

**Table 1.** Classification accuracy comparisons between state-of-the-art methods and our proposed DCNN aesthetic features on CUHKPQ dataset

| Aesthetic Features | Accuracy (%) |
|---|---|
| Luo[6] | 76.91 |
| Lo[7] | 81.76 |
| Datta[5] | 85.27 |
| Ke[8] | 81.70 |
| Marchesotti[10] | 79.53 |
| DCNN_Aesth | 90.76 |
| DCNN_Aesth_SP | **91.93** |

ImageNet is a very large dataset which collects more than 14 million high-resolution images from the Internet [26]. This dataset is meaningful for the researches on image understanding algorithms and helpful to train powerful machine learning models. ILSVRC is an annual competition on computer vision problems, especially the image classification. The dataset provided by ILSVRC is a part of the ImageNet. We train our implemented deep convolutional neural network on the dataset used in ILSVRC-2012 [27]. It contains more than 1.2 million images come from 1000 categories. Each category provides 732 to 1300 images. With the help of those image data, this DCNN obtains enough power to describe various image content.

As reported in [13], the 4096-dimensional output of the last full connected layer reveals the networks visual knowledge. The images look alike may also have a small Euclidean distance between their 4096-d activations. We conduct a further normalization on this output and take it as DCNN_Aesth features.

Before fed to the network, the image have to be scaled into a fixed resolution of 224*224. This may lead to the loss of information and DCNN may overlook the whole layout of the image. Due to this, we also adopt the spatial pyramid method that has been introduced in the last section. Its performance is also shown below denoted as DCNN_Aesth_SP.

For each kind of features, we train a two-class SVM classifier and take the classification accuracy as the performance evaluation. We get the parameters of all those SVM classifiers by conducting the cross-validation processes on the training set.

### 4.3 Experimental Results on the CUHKPQ

The experimental results on the CUHKPQ dataset are summarized in Table. 1. It can be observed that the DCNN_Aesth_SP features achieve the best performance of 91.93%. Then the DCNN_Aesth features follow with the accuracy of 90.76%. Among above methods which are mainly based on the hand-crafted features, Datta et al. performs the best with the accuracy of 85.27%. Other methods just obtain accuracies around 80%. Though Marchesotti et al. hold the similar idea

**Table 2.** Classification accuracy comparisons between state-of-the-art methods and our proposed DCNN_Aesth features on AVA dataset

| Aesthetic Features | Accuracy (%) |
|---|---|
| Luo[6] | 61.49 |
| Lo[7] | 68.13 |
| Datta[5] | 68.67 |
| Ke[8] | 71.06 |
| Marchesotti[10] | 68.55 |
| DCNN_Aesth | 78.92 |
| DCNN_Aesth_SP | **83.52** |

with us, the image descriptive power of dense SIFT is not as good as the deep convolutional neural network. Finally, their method just submits an ordinary grade with the accuracy of 79.53%.

In conclusion, features extracted by the deep convolutional neural network beat the best hand-crafted features on CUHKPQ dataset that they improve the accuracy by almost 6%. This phenomenon proves the effectiveness of DCNN when applied to deal with the problem of photo aesthetic quality assessment.

### 4.4 Experimental Results on the AVA

We present the experimental results obtained on the AVA dataset in Table. 2. Again, it can be observed that DCNN_Aesth_SP features get the best performance with the accuracy of 83.52%. DCNN_Aesth also performs very well that it obtains an accuracy of 78.92%. The best performance obtained by the methods based on hand-crafted features is 71.06%, which is still worse than the DCNN features. Other hand-crafted features just get accuracies under 70%. In addition, the method of Marchesotti et al. only gets a performance of 68.55% which is closed with most hand-crafted features.

Another thing we have to explain here is that all the methods perform better on the CUHKPQ dataset than on the AVA dataset. This is firstly due to that there are more various image categories existing in the AVA dataset. Secondly, as images in the AVA dataset are directly downloaded from the website without the careful selection that conducted on the CUHKPQ, their labels inevitably have more noise inside. On such a challenging dataset, DCNN_Aesth and DCNN_Aesth_SP features still perform very well that they improve state-of-the-art by more than 6%. This result is quite promising.

## 5 Conclusion

In this paper, we adopt the deep convolutional neural network to conduct the photo quality assessment. The network is firstly trained on the ImageNet to gain strong image descriptive power. Then several experiments are conducted on two public dataset specifically designed for the research of image aesthetic quality assessment. The experimental results show that our method is considerably better than state-of-the-art methods which are mostly based on the hand-crafted features. To our best knowledge, this paper is the first to apply deep learning methods in image aesthetic quality assessment. The success of it further proves the potential power of those kinds of deep neural networks with millions of neurons. In the future, we will further apply multi-view methods to combine our DCNN features and other hand-crafted aesthetic features [28].

**Acknowledgment.** This work is supported by the 973 project under the contract No.2015CB351803, the NSFC under the contract No.61390514 and No.61201413, the Fundamental Research Funds for the Central Universities No. WK2100060007 and No. WK2100060011, the Specialized Research Fund for the Doctoral Program of Higher Education No. WJ2100060003.

## References

1. Geng, B., Yang, L., Xu, C., Hua, X.-S., Li, S.: The role of attractiveness in web image search. In: ACM MM, pp. 63–72 (2011)
2. Li, C., Loui, A.C., Chen, T.: Towards aesthetics: A photo quality assessment and photo selection system. In: ACM MM, pp. 827–830 (2010)
3. Bhattacharya, S., Sukthankar, R., Shah, M.: A Framework for Photo-Quality Assessment and Enhancement based on Visual Aesthetics. In: MM, pp. 271–280 (2010)
4. Zhang, F.-L., Wang, M., Hu, S.-.M.: Aesthetic image enhancement by dependence-aware object recomposition. IEEE Trans. on Multimedia, 1480–1490 (2013)
5. 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)
6. 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)
7. Lo, K.Y., Liu, K.H., Chen, C.S.: Assessment of photo aesthetics with efficiency. In: ICPR, pp. 2186–2189 (2012)
8. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: CVPR, pp. 419–426 (2006)
9. 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)
10. Marchesotti, L., Perronnin, F., Larlus, D., Csurka, G.: Assessing the aesthetic quality of photographs using generic image descriptors. In: ICCV, pp. 1784–1791 (2011)
11. Dance, C., Willamowski, J., Fan, L., Bray, C., Csurka, G.: Visual categorization with bags of keypoints. In: ECCV International Workshop on Statistical Learning in Computer Vision, Prague (2004)
12. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: CVPR (2006)

13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS (2012)
14. Wong, L.K., Low, K.L.: Saliency-enhanced image aesthetics class prediction. In: ICPR, pp. 997–1000 (2009)
15. Yin, W., Mei, T., Chen, C.W.: Assessing photo quality with geo-context and crowdsourced photos. In: Visual Communications and Image Processing, pp. 1–6 (2012)
16. Chu, W.T., Chen, Y.K., Chen, K.T.: Size does matter: how image size affects aesthetic perception? In: ACM MM, pp. 53–62 (2013)
17. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., et al.: Handwritten digit recognition with a back-propagation network. In: Advances in Neural Information Processing Systems (1990)
18. LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time series. In: The Handbook of Brain Science and Neural Networks. MIT Press (1995)
19. Hubel, D.H., Wiesel, T.: Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. Journal of Physiology (London) 160, 106–154 (1962)
20. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: A convolutional neural network approach. IEEE Trans. Neural Networks 8, 98–113 (1997)
21. Le Callet, P., Viard-Gaudin, C., Barba, D.: A convolutional neural network approach for objective video quality assessment. IEEE Trans. Neural Netw. 17(5), 1316–1327 (2006)
22. Szarvas, M., Yoshizawa, A., Yamamoto, M., Ogata, J.: Pedestrian detection with convolutional neural networks. In: Proc. IEEE Intell. Veh. Symp., pp. 224–229 (2005)
23. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proc. 27th International Conference on Machine Learning (2010)
24. Luo, W., Wang, X., Tang, X.: Content-based photo quality assessment. In: ICCV, pp. 2206–2213 (2011)
25. Murray, N., Marchesotti, L., Perronnin, F.: AVA: A largescale database for aesthetic visual analysis. In: CVPR, pp. 2408–2415 (2012)
26. Deng, J., Berg, A., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: CVPR (2009)
27. Jia, Y.: Caffe: An open source convolutional architecture for fast feature embedding (2013), http://cafe.berkeleyvision.org/
28. Liu, W.F., Tao, D.: Multiview hessian regularization for image annotation. IEEE Trans. on Image Processing, 2676–2687 (2013)