Computational Aesthetics of Photos Quality Assessment and Classification Based on Artificial Neural Network with Deep Learning Methods

Yimin Zhou[1], Guangyao Li[1] and *Yunlan Tan[2]

1. College of Electronics and Information, Tongji University, Shanghai, 201804, China,
2. School of Electronics and Information, Jinggangshan University, Jiangxi, 343009, China
Yimin Zhou: 1010080053@tongji.edu.cn
*Yunlan Tan: Corresponding author, tanyunlan@163.com

Abstract

Photograph aesthetical evaluation has been widely investigated in these decades. The most used assessing methods are mainly classical data mining methods such as SVM, ANN(Artificial Neural Network), linear programming and so on. In this paper, we presented a method based on artificial neural network and deep learning methods which is also a hot research topic recently. We downloaded a medium and a large dataset from a well-known online photograph portal and trained on them. Results showed that the accuracy of classification was above 82.1%, which was better than all state-of-the-art methods as well as a moderate result from those methods never adopted up to now.

Keywords: Computational aesthetics; Quality assessment and classification aesthetics of photography; Artificial Neural Network; Deep learning methods

1. Introduction

In this epoch of information technology, digital photographs are not difficult to access as well as share among friends or even strangers through the Internet. This convenience brings us new industries like network sharing center and portals attracting people of the same interest, i.e. professional or amateur photographers around the world. Inevitably, works are compared and assessed by the insiders of the community. Thus it proposes a challenging problem to differentiate high quality photographs from low ones due to the vast amount of images produced and submitted to the internet. According to Flickr's statistics [1], users upload about 6.5 million photographs every day. Therefore, computational aesthetics of photo quality assessment has been investigated through these decades. Most of the methods adopted in consist of two phases, feature extraction and computational classification. Although common machine learning methods, such as Support Vector Machine, Linear Programming and Neural Network, have been introduced to evaluate the aesthetic values of photographs by a great number of researchers, most of them were applied to a relatively small set of images and just attained a passable accuracy for classification, and the feature set was artificially selected for extraction. Machine learning has become a hot research topic since the middle of last century and even more appealing after the emergence of deep learning methods.

In this work, we presented an image processing method related to features detection, training, classification and prediction by exploiting Dattra's work [2]. In addition, we applied artificial neural network [3] and autoencoder [4], a deep learning method, which is firstly adopted for aesthetic value evaluation, to Dattra's method to achieve a more accurate result. We also applied other deep learning methods such as convolutional neural

network [5, 6] and deep belief network [7] to the image datasets, but only obtained moderate results. The remainder of this article is organized as follows. In the next section, the related works is reviewed. In section 3, the method of image feature extraction, artificial neural network, autoencoder and other deep learning methods adopted in this work are illustrated. In section 4, the experimental results are presented. In section 5, a conclusion is drawn.

2. Related work

Measurement of the aesthetic value can be retraced to the American mathematician Birkhoff [8], who wrote the "Aesthetic Measure" in 1933. He took the lead in proposing a mathematical explanation of aesthetic value, M=O/C, where M means measurement, O represents order and C is complexity. Many successors of his theory improved the representation. For example, Machado *et al.* [9] and Rigau *et al.* [10-11] combined it with modern information theories such as compression ratio or Shannon's Entropy and Kolmogorov complexity.

Although the theoretical representation of aesthetic value acts as a standard way of measurement, classical assessments of features such as intensity, combination, hue, saturation and contrast, still dominate the senses of most observers and connoisseurs. Wang and Datta *et al.* [2] were the first to realize the quantization of image features, including the brightness, color distribution, wavelet, region composition and depth of field, and applied SVM or linear regression to classify the high from the low quality photographs and achieved an accuracy of 70.12%. The ACQUINE [12] aesthetics value measurement system is a typical aesthetic evaluation and search engine developed by them. However, this method is impractical for large datasets due to the limit of SVM. Wu *et al.* [13] extended their SVM classification method to predict aesthetic measurement values and Ke *et al.* [14]extended their work by adding visual features of images, but both methods remained the same drawbacks. Wong *et al.* [15] presented a saliency-enhanced image classification method but this method should be preprocessed to detect the saliency region. Moreover, all of the above researchers did not go further into the classification methods on the machine learning side. Sidra *et al.* [16] provided a classification method based on machine learning, but their dataset seemed too small to produce a highly convincing result.

3. Proposed Method

Unlike the above methods, we proposed an aesthetic value evaluation method which combined the common feature extraction method with the artificial neural network and deep learning methods. This is the first time to train Datta's dataset with an artificial neural network using autoencoder top reprocess the input. The whole process could be demonstrated as three phases, an each took a lot ofwork. The flow chart of the proposed method is illustrated as in Figure 1. The first phase is data collection. We downloaded tens of thousands of images from internet and excluded those with few or no evaluation records. The second phase is building autoencoder, with which different amounts of encoded sizes was utilized. The last phase is training and classifying as much as possible combination of configurations. Then we took a much larger dataset as input and trained with deep learning methods such as convolutional neural network and deep belief network. Both methods are configured to classify the photographs into high and low aesthetic values. The former achieved a considerably good result, while the latter only a moderate one.
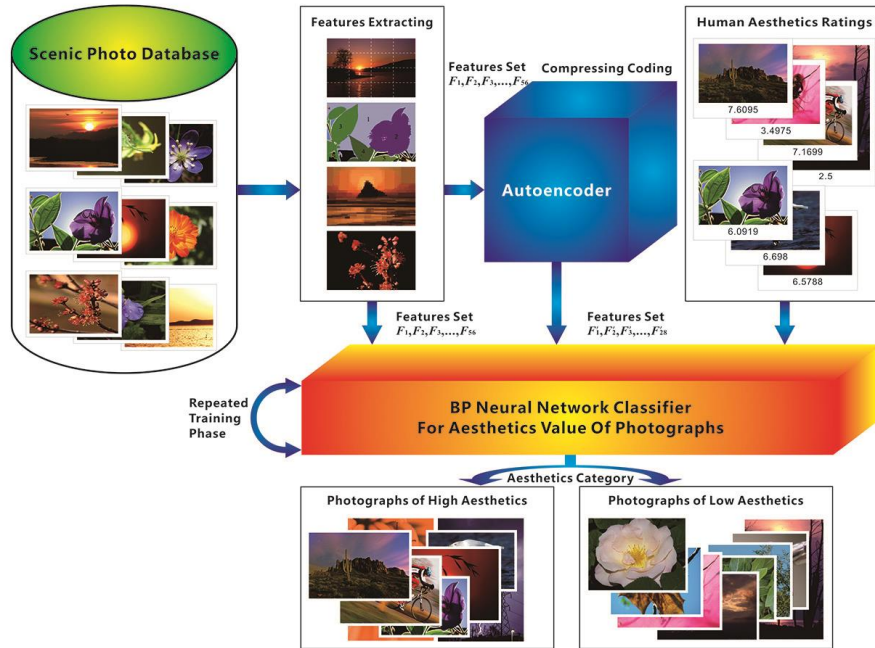
Figure 1. The Flow Chart of the Proposed Method

3.1. Feature extraction

Since we firstly exploited Datta's work of extraction features, it's better to demonstrate them in advance. They treated each image separately and extracted its features. RGB values of each image were converted to HSV image space.

They extracted 56 visual features for each image. Feature f_1 denotes the intensity of the image exposed to the light, or called brightness. To compute the relative color distribution and distinguish high contrast images from the low ones, earthmover's distance (EMD)[17] is extracted as f_2.

$f_2 = \text{emd}(D_1, D_2, \{d(a,b) \mid 0 \leq a,b \leq 63\})$, where $d(a,b) = \|\text{rgb2luv}(c_a) - \text{rgb2luv}(c_b)\|$.

The RGB color space is divided into 64 cubic blocks with four equal partitions along each dimension, and each cube represents a sample point. Distribution D1 is generated as the color distribution of a hypothetical image, for each of the 64 sample points, the frequency is 1/64. Distribution D_2 is computed from the given image by finding the frequency of color occurrence within each cube, and then the pairwise Euclidean distance between the geometric centers ci of each cube i is computed after converting to LUV space.

|(a)|(b)|(c)|(d)|

Figure 2. The proposed colorfulness measure. (a) and (b) are the two photographs with high values. (c) and (d) are the two photographs with low values

The common hue and saturation channels of HSV are extracted as f_3 and f_4. The popular rules of the thirds measurement of a photograph is considered as a sloppy

approximation to the 'golden ratio'(about 0.618) which is adopted to create f_5-f_7 corresponding to the HSV channels of that crucial area.

$$f_5 = \frac{9}{XY} \sum_{x=X/3}^{2X/3} \sum_{y=Y/3}^{2Y/3} I_H(x, y)$$

| (a) | (b) | (c) | (d) |

Figure 3. Two famous aesthetic rules in the photography literature. (a) The rule of thirds in photography: imaginary lines cut the image horizontally and vertically into three parts at each direction. Intersection points by the four crossing lines are chosen to place important parts of the composition instead of the center. (b)-(d) are Daubechies wavelet transform.(b) is the original image; (c) is a three-level transformation, in which the levels are separated by borders; (d) is an arrangement of the three bands LH, HL and HH of the coefficients.

The IRM [18] distances invented by Datta *et al*. are calculated as features f_8 and f_9. There are different ways to measure the graininess and smoothness of a photograph, and one of them is to use Daubechies wavelet transform [19]. It performs a three-level wavelet transformation on all three color bands IH, IS and IV. An example of such transformation on the intensity band is illustrated in Figure 3(b)-(c). The three levels of wavelet bands are arranged from top left to bottom right in the transformed image, and the four coefficients per level, LL, LH, HL, and HH are arranged as shown in Figure 3(d). Denoting the coefficients (except LL) in level i of the wavelet transform on hue image IH as w_i^{hh}, w_i^{hl} and w_i^{lh}, i = {1, 2, 3}, the features f_{10}, f_{11} and f_{12} are defined as follows:

$$f_{i+9} = \frac{1}{S_i} \left\{ \sum_x \sum_y w_i^{hh}(x, y) + \sum_x \sum_y w_i^{hl}(x, y) + \sum_x \sum_y w_i^{lh}(x, y) \right\}$$

Where i = 1,2,3. A total of nine such features, f_{10}-f_{18}, are calculated; three of them in one dimension corresponding to the combination of three levels of wavelet coefficient bands, while three in another dimension to the channels of HSV. The sum of the average wavelet coefficients of all three band levels for each of H, S and V are used as three additional features f_{19}-f_{21}. The average size and aspect-ratio of a photograph are assigned to f_{22} and f_{23}, respectively. Segmentation results from rough grouping of similar pixels often represent objects in the scene, so the largest five connected components or patches are formed by the segmentation algorithm. The number of patches, t $\leq$ 5, which are covered at least one percent of the whole photograph area is denoted as f_{24}. The number of color-based clusters formed by K-Means in the LUV space is feature f_{25}, followed by the average H, S and V values of each of the top 5 patches, f_{26}-f_{40}. Features f_{41}-f_{45} store the relative size of each segment of the image. Two new features, f_{46} and f_{47}, respectively correspond to the average color spread around the wheel of the HSV hue component and the average complimentary colors of the hues of the top 5 patches. The rough positions of each segment are stored as features f_{48}-f_{52}. Professional photographers often reduce the depth of field (DOF)[20] for a single object shooting by using larger aperture settings, macro lenses, or telephoto lenses. We divided an image into 16 equal rectangular blocks

$\{M_1, ...M_{16}\}$, numbered in row-major order. Let $w_3 = \{ w_3^{lh}, w_3^{hl}, w_3^{hh} \}$ denotes the set of wavelet coefficients in the high-frequency of the hue image IH. The low depth of field indicator feature f_{53} for hue is computed as follows, while f_{54} and f_{55} are computed similarly for I_S and I_V, respectively:

$$f_{53} = \frac{\sum_{(x,y) \in M_6 \cup M_7 \cup M_{10} \cup M_{11}} w_3(x, y)}{\sum_{i=1}^{16} \sum_{(x,y) \in M_i} w_3(x, y)}$$

The last feature f_{56} is shape complexity. Therefore, a total of 56 features are extracted from each original image.

3.2. Artificial Neural Network

In machine learning and related fields, artificial neural networks (ANNs)[3] are computational models inspired by animal central nervous systems (in particular the brain), which is capable of machine learning as well as pattern recognition. Artificial neural networks are typically presented as systems of inter connected "neurons" which compute values from inputs.

Generally speaking, an artificial neural network comprises multiple layers of neurons. Each neuron accepts one or several inputs from the neurons of a precedent layer and generates an output conveyed to the neurons of a subsequent layer, working like the synapses. Every input into a neuron is associated with a weight whose products are summed up onto the target neuron whose activation function is always sigmoid or tanh.

A common multilayer feed forward network consists of three parts: an input layer accepts a multitude of nonlinear input information or vector; an output layer produces an output vector in most cases for decision; one hidden layer or several hidden layers contain undefined amount of neurons which help improving the nonlinearity and robustness of the network.

3.3. Autoencoder

Deep learning is based on distributed representations, a concept in machine learning. The underlying assumption of the distributed representations is that the observed data is generated by the interactions of many different factors on different levels. Varying numbers of layers and layer sizes can be used to provide different amounts of abstraction.

An autoencoder is an artificial neural network used for learning efficient codings. It aims to learn a compressed, distributed representation (encoding) of a set of data, typically for the purpose of dimensionality reduction.

Architecturally, the simplest form of the autoencoder is a feed forward, non-recurrent neural network that is very similar to the multilayer perceptron (MLP) [21, 22], which contains an input layer, an output layer and one or more hidden layers connecting them. The difference of an autoencoder from the MLP is that its output layer equally takes many nodes as the input layer, and is trained to reconstruct its own input x instead of a given input x to predict some target value y.

3.4 Quality Assessment and Classification

3.4.1 ANN with Autoencoder: According to Datta's work, the raw inputs of the artificial neural network should be those features which were extracted and collected from the given dataset, *i.e.* the 56 features. Therefore these 56 features are part of the input that was applied to the ANN. Autoencoder, as aforementioned, is a compressed representation of its input, usually the raw data. We trained autoencoders to compress the raw features into new features with 1/2 or 1/4 of its original amount. In addition to the original 56 features, we used the new compressed features as a second part of the input of ANN to

obtain a remarkably good result. That means ANN with Autoencoder method consists of two phases: 1) training raw input features into compressed representative new ones; 2) applying the combined features to ANN and training. The whole process is demonstrated in Figure 4.

(a) The structure of the autoencoder

(b) The structure of the artificial neural network
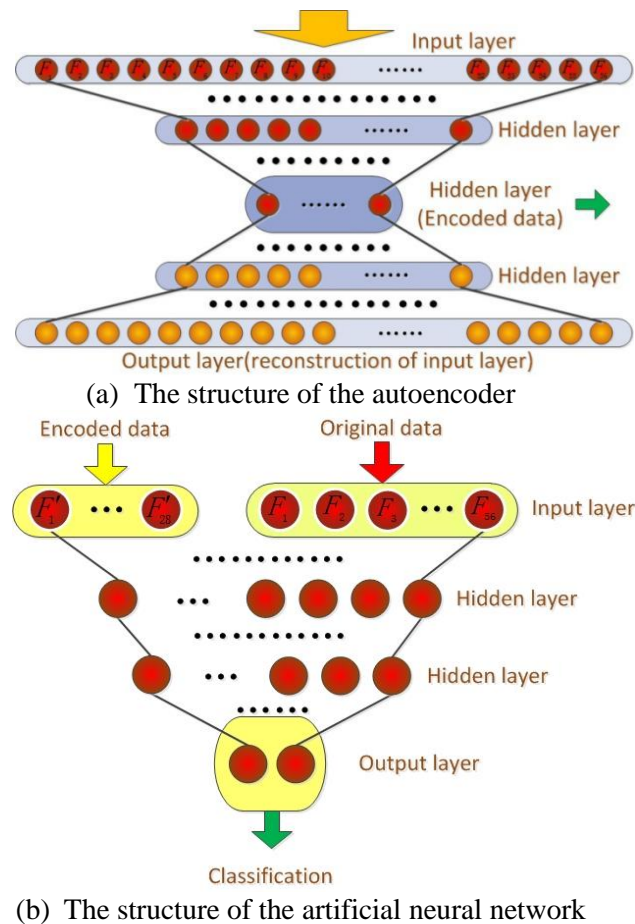
Figure 4. Autoencoder with ANN Method

3.4.2 Other Deep Learning Methods

1) Convolutional neural network

Convolutional neural networks (CNN) [5, 6] consist of multiple layers of small neuron collections which focus on a small portion of the input images, i.e. receptive fields, whose results are then tiled to obtain a more complicated and thorough representation of the image. Pooling layers are commonly included to refine the features of lower layers for advanced learning in deeper layers. CNN concurrently extracts and classifies features and then shares weights to reduce size and improve performance.

(2) Deep Belief Network

A restricted Boltzmann machine (RBM) [7, 23] is a generative stochastic artificial neural network that can learn a probability distribution from a set of inputs. As their name implies, RBMs are a variant of Boltzmann machines [24], with the restriction that their neurons must form a bipartite graph: 1) the input units, corresponding to the features of their inputs; and 2) the hidden units that are trained, and each connection in an RBM must connect a visible unit to a hidden unit. Deep belief network (DBN) [7] is a type of deep neural network, composed of multiple layers of latent variables ("hidden units"),

with connections between the layers but not the units within each layer. DBN can learn to probabilistically reconstruct its inputs from training examples in an unsupervised way. The layers then act as feature detectors on the inputs. After this learning step, a DBN can be further trained in a supervised way to perform classification. A DBN can be viewed as a composition of simple, unsupervised networks of stacked RBMs.

Current existing computational photo aesthetic assessment and classification methods can be divided into two parts, features extraction and classification using machine learning methods. Although this fixed process does provide satisfying results, it takes up plenty of time and tremendous effort especially on designing how to extract the features and determining the final ones. This is still a painful manual work without substitutions. Therefore, in this paper, we presented a method without manually designing and selecting the features based on deep learning methods.

The dataset applied is much larger than that in the previous part. We firstly built a CNN for this round of training. Every photograph was first clipped as less as possible to retain the most central portion with an aspect ratio of 4:3. Then each of them was scaled to the same size of 160 x 120 and converted to HSV color space. Such size and configuration should meet human vision characteristics, preserve the most essential part of the image and still within a feasible computation amount for a modern computer. The CNN consisted of 2 convolutional layers with pooling layers, and a final classification layer. The convolutional kernel sizes of the two layers were 49x49, 9x9 respectively and the first layer had a pool size of 4x4. Figure 5 illustrated this structure. Secondly, we also constructed a DBN for the same dataset as the CNN. The neurons of hidden layers within this network from top to bottom were 2000, 1000, 500, 200, respectively. A fine tuning was executed after the DBN were trained and converted to an ANN.
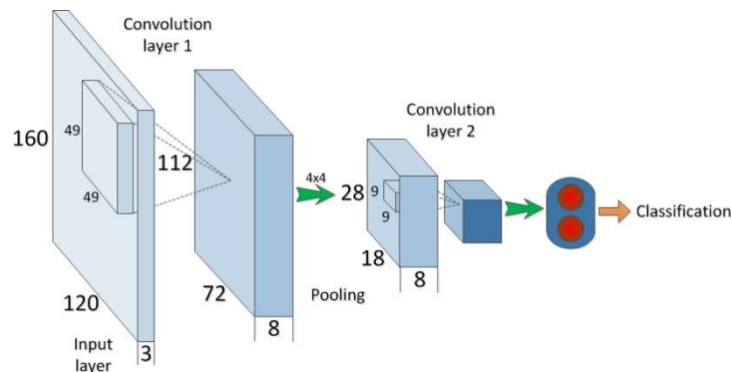
Figure 5. The Structure of the Convolutional Neural Network

4. Experiment Result

We tested the formerly proposed algorithms on tens of thousands of color images depicting typical scenes of natural landscapes by a PC with AMD Athlon II CPU 2.7GHz. The experiments were scheduled in Python [25] and MATLAB [26].

4.1 Database Collection

One of our data source is the large online photo sharing community, Photo.net, which is the same as Datta's report [27]. The dataset contained a collection of 3581 images. Another professional digital photography community we exploited is dpchallenge.com [28]. An advantage of the collection is that each photo had been evaluated by at least one hundred users. The collection contains the number of aesthetics ratings of each image, the means of rating, and a distribution of quality ratings on a1-10 scale. The aesthetic values accepted in the experiments are from all users. While amateurs represented the general population, the professional photographers tend to pay more attention to technical factors

for evaluation. Totally, we have downloaded 28,896 photographs with miscellaneous contents from the portal.

4.2 Experimental Results

We first applied the ANN with autoencoder training method to Datta's dataset. It seemed to work well and obtain a good result. We made up from Datta's dataset at a 5:1 proportion of training set to test set, with the training set of more discriminative ones and the test set distributed almost uniformly through all the ranges of aesthetic values. Each configuration was trained at least 10 times. The gained results are listed in Table 1. Both of the results of precedent researchers like Datta and Wong etc. and ours are listed, including the average accuracy and precision. It shows even with the same configuration as Datta's dataset, the accuracy obtained from our experiment was higher. To demonstrate our neural network training methods surpass the previous ones, we used the same training and test set as Datta, which applied SVM [29] to achieve a much higher accuracy. Table 2 illustrates the comparison of methods that we applied and mentioned. Instead of treating the extracted features as input, we also listed different amount of encoded features as additional inputs for training and testing. As expected, when the more encoded representation of features is added, the more promising result is gained. The best accuracy obtained in our method is even above 82%.

Table 1. Comparison between our Methods and Previous Ones

Classification results/%	Datta's [4]	Wong's [3]	Wang's [30]	Our method
	(15 dimensions by SVM)	(44 dimensions by SVM)	(261 dimensions by SVM)	(56 dimensions by ANN+Autoencoder)
High aesthetic value	73.20	75.47	77.60	84.5
Low aesthetic value	62.27	70.81	73.14	79.7
Average accuracy	67.74	73.14	75.37	82.1

Table 2. Comparison among our ANN with Autoencoder Method under Different Dimensions

Classification accuracy/%	56 dimensions only	56 dimensions with 7 encoded	56 dimensions with 14 encoded	56 dimensions with 28 encoded
Average accuracy	75.13	76.21	79.39	82.14

For other deep learning methods we used, the training and test settings were as follows: according to the distribution, we treated > 5.35 as high quality aesthetic value and ≤ 5.35 as low value. Then we chose 2,460 images from the data set including all ranges of aesthetic values and almost uniformly distributed to make up the test set. After that, we selected 12200 images with aesthetic values greater than 5.9 or less than 4.7 from the remaining dataset to form the training set for they were more discriminative than those with medium values. Thus the proportion between training and test set is 5:1. As for the method of training by a convolutional neural network described in section 3.4.2, we obtained a modest result with an accuracy of 61.8% considering the nature of the task and the scale of the data. For the method of deep belief network, a modest accuracy of 58.7% is gained.

These results are all generated on an ordinarily configured PC. Each of the above experiment took no more than several hours to obtain the final result. The CNN and DBN methods can save the work of feature extraction.

5. Conclusion

We presented an approach of machine determining the aesthetic value of photographs. We applied artificial neural network with autoencoder as the first model of training. We first separated dataset into training set and test set with a common proportion and obtained a surprisingly good result. Such accuracy and stability was above all the currently proposed methods and applicable in most practical situation. Then we made another arrangement under the same circumstances as others and achieved a much advantageous result over the cited precedent methods. We also downloaded tens of thousands images from a professional photo portal and applied other deep learning methods as convolutional neural network and deep belief network to a carefully chosen set from all the previously downloaded images to obtain a modest accuracy considering the nature of the task and the scale of the dataset.

Acknowledgements

References

[1] Flickr, http://www.flickr.com
[2] R. Datta, D. Joshi, J. Li and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach", Computer Vision–ECCV **(2006)**, pp. 288-301.
[3] C. M. Bishop, Neural networks for pattern recognition, Oxford University **(1995)**.
[4] Y. Bengio, "Learning deep architectures for AI", Foundations and trends in Machine Learning, vol.2, no.1, **(2009)**, pp. 1-127.
[5] D. Ciresan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification", Computer Vision and Pattern Recognition (CVPR), IEEE Conference on **(2012)**, pp. 3642-3649.
[6] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", Biological cybernetics, vol.36, no.4, **(1980)**, pp. 193-202.
[7] G. E. Hinton, R. R. Salakhutdinov. "Reducing the dimensionality of data with neural networks", Science, vol.313, no.5786, **(2006)**, pp. 504-507.
[8] G. D. Birkhoff, Aesthetic measure, Cambridge, Mass., **(1933)**.
[9] P. Machado and A. Cardoso, "Computing aesthetics", Advances in Artificial Intelligence, Springer Berlin Heidelberg, **(1998)**, pp. 219-228.
[10] J. Rigau, M. Feixas and M. Sbert, "Conceptualizing Birkhoff's aesthetic measure using Shannon entropy and Kolmogorov complexity", Proceedings of the Third Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging, Eurographics Association, **(2007)**, pp. 105-112.
[11] J. Rigau, M. Feixas and M. Sbert, "Informational aesthetics measures", IEEE Computer Graphics and Applications, vol.28, no.2, **(2008)**, pp. 24-34.
[12] R. Datta and J. Z. Wang, "ACQUINE: aesthetic quality inference engine-real-time automatic rating of photo aesthetics", Proceedings of the international conference on Multimedia information retrieval. ACM, **(2010)**, pp. 421-424.
[13] Y. Wu, C. Bauckhage and C. Thurau, "The good, the bad, and the ugly: Predicting aesthetic image labels", Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, **(2010)**, pp. 1586-1589.
[14] Y. Ke, X. Tang and F. Jing, "The design of high-level features for photo quality assessment", Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, vol.1, **(2006)**, pp. 419-426.
[15] L. K. Wong and K. L. Low, "Saliency-enhanced image aesthetics class prediction", Image Processing (ICIP), 2009 15th IEEE International Conference on. IEEE, **(2009)**, pp. 997-1000.
[16] S. Riaz, K. H. Lee and S. W. Lee, "Aesthetic score assessment based on generic features in digital photography", 5th AUN/SEED-Net Regional Conference on Information and Communicaation Technology, Manila, Philippine, **(2012)**, pp. 76-79.
[17] Y. Rubner, C. Tomasi and L. J. Guibas, "The earth mover's distance as a metric for image retrieval", International Journal of Computer Vision, vol.40, no.2, **(2000)**, pp. 99-121.

[18] J. Z. Wang, J. Li and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.23, no.9, **(2001)**, pp. 947-963.

[19] I. Daubechies, Ten Lectures on Wavelets, Philadelphia, SIAM, **(1992)**.

[20] Couzin and Dennis, "Depths of Field", SMPTE Journal, November 1982, **(1982)**, pp. 1096–1098.

[21] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms", CORNELL AERONAUTICAL LAB INC BUFFALONY, **(1961)**.

[22] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation", CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR COGNITIVE SCIENCE, **(1985)**.

[23] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines", Proceedings of the 25th international conference on Machine learning. ACM, **(2008)**, pp. 536-543.

[24] G. E. Hinton and T. Sejnowski, "Analyzing cooperative computation", Proceedings of the Fifth, **(1983)**.

[25] I. J. Goodfellow, D. Warde-Farley and P. Lamblin, V. Domoulin, M. Mirza, R. Pascanu, J. Bergstra, F. Bastien and Y. Bengio, "Pylearn2: a machine learning research library", arXiv preprint arXiv, **(2013)**, pp. 1308.4214.

[26] R. B. Palm, "Prediction as a candidate for learning deep hierarchical models of data", Technical University of Denmark, Palm, **(2012)**.

[27] Photo.Net, http://www.photo.net.

[28] DPChallenge.com, http://www.dpchallenge.com.

[29] A. Aizerman, E. M. Braverman and L. I. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning", Automation and remote control, vol.25, **(1964)**, pp. 821-837.

[30] W. N. Wang, J. J. Yi, X. M. Xu and Li. Wang, "Computational aesthetics of image classification and evaluation", Journal of Computer-Aided Design & Computer Graghics, vol.26, no.7, **(2014)**, pp. 1075,1083.

Authors

Yimin Zhou, he was born in Shanghai, China in 1980. He received his B.S. degree in computer science and technology from Shanghai Jiaotong University, Shanghai, China, in 2002. He received his M.S. degree in Tongji University, Shanghai, China. Now he is pursuing his Ph.D. degree. His research interests concentrate on image processing, machine learning and computer graphics.

Guangyao Li, he was born in Shanghai, China in 1980. He received his B.S. and M.S. degree in Nanjing University of Aeronautics and Astronautics in 1986 and 1989 respectively. He received his Ph.D. degree in 1997. Now he is a professor in the College of Electrical and Information Engineering, Tongji University, Shanghai, China. His current research interests include virtual reality and simulation, computer graphics and image processing.
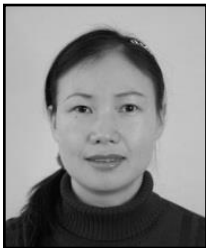
Yunlan Tan, she was born in Jiangxi province, China in 1972. She received her B.S. degree in Computer Science from Jiangxi Normal University and her M.E. degree in computer application technology from East China Normal University, China, in 1996 and 2004 respectively. Now she is an associate professor in the School of Electrical and Information Engineering, Jinggangshan University, Jiangxi, China and she is pursuing his Ph.D. degree in the College of Electrical and Information Engineering, Tongji University, Shanghai, China. Her current research interests include image processing and machine learning.