A Comprehensive Survey on Image Aesthetic Quality Assessment

Hongtao Yang*†, Ping Shi*§,Saike He‡§,Da Pan*,Zefeng Ying*,Ling Lei*

*College of Information Engineering, Communication University of China, Beijing, China

† School of Electrical and Information Engineering, Beijing Polytechnic College, Beijing, China

‡The State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing, China

yhtbyr@163.com, {shiping,pdmeng,yingzf,leiling}@cuc.edu.cn, saike.he@ia.ac.cn

§Corresponding author

***Abstract*—Image aesthetic quality assessment has demonstrated tremendous success in variety of application domains in recent years. This field has been growing so rapidly that various approaches have been proposed trying to solve this challenging problem. This report presents a comprehensive survey on image aesthetic quality assessment, mainly focus on the contributions and novelties of the existing approaches recently. In this work, we firstly illustrate datasets related to image aesthetics and investigate feature extractions. Then five different aesthetic tasks are reviewed, including aesthetic classification, aesthetic regression, aesthetic distribution, aesthetic factors and aesthetic description. In addition, we reviewed recent applications concerning image aesthetics. Finally, different evaluation criterions in different literatures are summarized. We hope the survey could serve as a comprehensive reference and be useful for those who are interested in exploiting image aesthetic for their research.**

***Index Terms*—Image Aesthetic, Aesthetic Quality Assessment, Deep Learning, Convolutional Neural Networks**

I. Introduction

Nowadays, there is a tremendous increase in the use of digital images as a means for representing and communicating information. Therefore, evaluating the quality of aesthetics has become an important task. And image aesthetic quality assessment (IAQA) is becoming a hotspot research direction.

Traditional image quality assessment (IQA) aims at using the computer to simulate the human visual system automatically to evaluate image distortion degrees, which mainly focus on the process of image acquisition, image compression, image processing, image transmission and display. Image quality usually relates to distortions caused by lossy compression, noises, transmission channel attenuation, etc. IQA is to obtain the objective score which is consistent with the subjective assessment. However, IAQA aims at using the computer to simulate human perception and cognition of ”beauty”, automatically evaluate the ”beauty” of the image. The ”beauty” mainly focuses on image aesthetic factors such as interesting content, object emphasis, good lighting, color harmony, vivid color, shallow depth of field, rule of thirds, balancing element, motion blur, etc. Aesthetics is a human intelligent activities. Comparing with tasks like object recognition, it is a big challenge to teach the computer how to assess ”beauty”. Recently, the artificial intelligence technology has made great progress in these areas, the performance is as good as the human's, in some specific situations even better. However, it is far away for computer to perceive ”beauty”, find ”beauty”, process ”beauty” and create ”beauty”. IAQA has become a crossing research direction in computing aesthetics, computer vision, psychology, virtual reality and other research fields.

IAQA has attracted researchers' attention only in recent ten years. From the beginning, it does not follow the rule-based way, but follow the data-driven way. Therefore, the construction of the dataset has become the key prerequisite for the research. In the acquisition of subjective scores of image aesthetics, it can be realized through artificial scoring experiments in the lab [1] , online image sharing and scoring [2], and crowd sourcing [3].

The majority of approaches for IAQA could be classified into feature extraction and decision making. As is summarized in Fig.1.

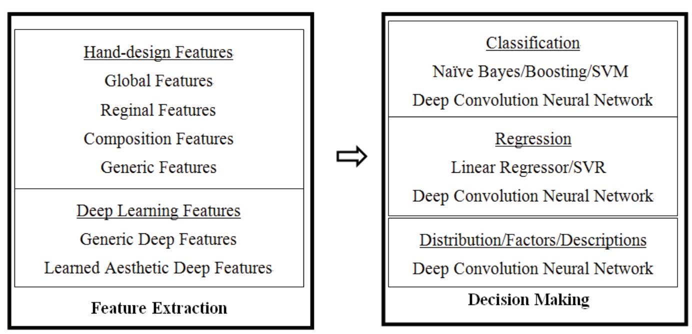

Fig. 1. Feature extraction and decision making in image aesthetic assessment systems.

Feature extraction aims at extracting high-level features and low-level features to describe the aesthetic aspect of an image. Such features are designed to distinguish images depending on the factors of different attributes. Feature types can be classified into hand-design features and deep learning features depending on how features are extracted. Conventional approaches including rule-based approach and traditional machine learning typically adopt hand-design features to design the photographic rules, global image layout and typical objects

978-1-7281-0801-8/19/$31.00 ©2019 IEEE
ICIS 2019, June 17-19, 2019, China

IEEE computer society

in images, while deep learning approach could extract features automatically through deep learning.

Decision making aims at assessing image aesthetics or specific related tasks after feature extraction. The tasks include aesthetics classification, aesthetics regression, aesthetics distribution, aesthetics factors, and aesthetics description. Using machine learning algorithm to train classifiers, which include support vector machine (SVM), boosting, random forest, K-nearest neighbor classification, deep convolution neural network (CNN) for classification, customized classifier used for binary classification. While linear regression, support vector regression (SVR), deep CNN for regression, customized regression are used in ranking or scoring images based on the aesthetic quality. Aesthetic distributions, aesthetic factors and aesthetic description always use deep CNN to assess images.

There are some surveys about image aesthetics [2] [4], however, these papers do not relate to different approaches based on different aesthetic tasks, especially aesthetic distribution, aesthetic factors, aesthetic description. In this paper, we would like to contribute a thorough overview of image aesthetic assessment, and have a comprehensive survey on the important literatures and new literatures recent years. Furthermore, as different input datasets and different evaluation criteria vary in different literatures, we would not focus on comparing the performances of different models, instead, we would focus on the contributions and novelties in literatures, and give a potential target and new insights for future directions in this study. For the whole paper, we would illustrate datasets in IAQA and investigate the feature extraction. Specifically, we would systematically evaluate five different aesthetic tasks with deep learning: aesthetic classification, aesthetic regression, aesthetic distribution, aesthetic factors, aesthetic description. Moreover, we would review aesthetic applications, including image enhancement, image cropping, etc. We would also summarize evaluation criterions in the literatures.

II. DATASETS

In the following, we would introduce these datasets that are most frequently used in IAQA.

- The Photo.Net [2] and DP Challenge [5] could be considered the earliest large-scale database for image aesthetic assessment. The Photo.Net contains 20,278 images with at least 10 ratings for each image and DPChallenge contains 16,509 images in total.
- CUHK-Photo Quality (CUHK-PQ) [1] contains 17,690 images. All the images in CUHK-PQ are given binary aesthetic labels and grouped into 7 scene categories. The ratio of the total number of positive and negative examples in CUHK-PQ is around 1:3.
- Aesthetic Visual Analysis (AVA) [6] contains 255,530 valid images, each of which is rated by artists with 1-10 points. More than 200 papers cite AVA. The annotated data is of high quality, which can support aesthetic classification and aesthetic distribution. AVA also has more than 60 kinds of the semantic annotation in photography style.
- Image Aesthetic Dataset (IAD) [7] contains 1.5 million images collected and annotated from DPChallenge and PHOTO.NET. The ratio of the positive examples and negative examples is around 1.07:1.
- Waterloo IAA [8] contains more than 1000 images. The dataset has continuous aesthetic scores, which could also supply approximately balanced aesthetic distribution from poor to excellent levels.
- Aesthetic Ratings from Online Data (AROD) [9] contains 380,000 images. The score is calculated by the number of image viewers who ”like” or ”favor” the image through online platforms.

In 2016, Kong et al [10] design a new Aesthetics and Attributes Database(AADB); In 2017, Chang et al. [11] design a new aesthetic image captioning dataset called Photo Critique Captioning Dataset (PCCD); In 2018, Wang et al. [12] construct AVA-Reviews dataset based on AVA. We will illustrate them in the following sections.

Besides, some specific-task dataset for particular use is built, such as Chinese Handwriting Aesthetic Evaluation Database (CHAED) [13], which contains 1000 Chinese handwriting images with diverse aesthetic qualities and inviting 33 subjects to evaluate the aesthetic quality for each calligraphic image.

Comparing with ImageNet [14] (15 million images, 22,000 categories), aesthetic classification, regression and distribution dataset (AVA 255,530; AROD 380,000), aesthetic factors dataset (AADB 10,000), aesthetic description dataset (PCCD 4,307) are too small. Therefore, IAQA requires larger databases with richer annotations.

III. FEATURE EXTRACTION

The conventional feature extraction approach for IAQA is to use hand-design feature extractors, which require a considerable amount of engineering skills and domain knowledge.

Global features are first explored to represent the aesthetic aspect of the image. Ke et al. [15] first propose indicators to represent the image, including spatial distribution of edges, color distribution, hue count, blur, and two low-level features including contrast and brightness. Aydn et al. [16] construct 5 image attributes including sharpness, depth, clarity, tone and colorfulness. Improved upon these global features, later studies adopt global saliency to estimate aesthetic attention distribution.

Regional image features are proved to be a good complementary to the general features. Wong et al. [17] compute exposure, sharpness and texture features on salient regions, global image, as well as features depicting the subject background relationship of the image. Nishiyama et al [18] extract bags-of-color-patterns from local image regions with a grid-sampling technique. Lo et al. [19] first extract color and texture features from the images, then build a statistic modeling system with coupled spatial relations.

The composition of the image usually relates to the presence and the position of a salient object. Dhar et al. [20] propose the high-level features including describable attributes of composition, content and illumination, then combine low-level features together to train a SVM classifier. Tang et al. [1] propose content-based photo quality assessment using both regional and global features. Further, image composition is assessed by global hue composition and scene composition. Zhang et al. [21] propose the image descriptors that characterize local and global structural aesthetics through multiple visual channels. To describe the spatial structure of the local regions, they construct graphlets by connecting spatially adjacent atomic regions.

General Features including Bag of Visual Words (BOV) and Fisher Vector (FV) are explored in [6] [22] [23]. Yeh et al. [23] utilize SIFT descriptors and present the relative features. Marchesotti et al [22] first use BOV and FV descriptor, then encode gradient information using SIFT and color information for both descriptors. Usually, SIFT descriptor and color descriptors are used as local descriptors. Gaussian Mixture Model (GMM) is commonly trained upon them where GMM distribution is always encoded using BOV or FV, and each region encoded by FV are concatenated together to represent the final image.

Overall, hand-design features are affected by photography and psychology, which has certain drawbacks. Firstly, the dimensions of hand-design features are limited. Secondly, hand-design features are only approximate values. Besides, general features such as SIFT, FV, or BOV are used to capture general features of natural images, rather than aesthetics of the images.

The hand-design aesthetic features are difficult to be quantified and could not fully depict the image. Deep learning approach could automatically learn features, which do not need to extract image aesthetic features with rich aesthetic knowledge and photography skills. The researchers could alter CNN through adding additional style or content information. Currently, IAQA by deep learning is the mainstream approach, whose performance is much better than the conventional hand-design aesthetic feature attraction with conventional machine learning. We would summarize recent literatures using deep learning to deal with aesthetic assessment in the following sections.

IV. Aesthetic Tasks with Deep Learning

The research about IAQA can be summarized as the following five tasks. 1) Aesthetic classification refers to the output of binary aesthetic quality categories of ”good” and ”bad” for a given image; 2) Aesthetic regression refers to the output of aesthetic quality score of the image, which is expressed as a continuous value; 3) Aesthetic distribution refers to the histogram of aesthetic quality score distribution of images; 4) Study of aesthetic factors refers to the evaluation of interesting content, object emphasis, good lighting, color harmony, vivid color, shallow depth of field, rule of thirds, balancing element, motion blur and other aspects of the image; 5) Aesthetic description refers to linguistic comments on image aesthetics. We would have a survey in the following sections.

A. Aesthetic Classification and Aesthetic Regression

Most of the existing work relate to image aesthetics is a classification or regression task. Here, we would have a brief review of the main aesthetic image assessment approaches in classification or regression by deep learning.

In 2014, Lu et al [24] propose RAPID, which is the first attempt to assess aesthetics by deep learning. A double-column deep convolutional neural network (DCNN) is used to incorporate global view and local view to combine the feature extraction with classifier training. Moreover, style attributes is used to improve the accuracy of aesthetic quality categorization.

Lu et al [25] also design a deep multi-patch aggregation network (DMA-net). The structure contains a group of CNN for extracting features from multiple patches, a statistical aggregation structure and a sorting aggregation structure to aggregate features, which has good performance on image style recognition, aesthetic quality categorization and image quality estimation.

Kao et al. [26] propose multi-task category-specific CNN architectures, including ”scene”, ”object” and ”texture”. Each category has an associated CNN. For each CNN, classification and regression models are developed separately to predict binary aesthetic class and to regress an aesthetic score.

Wang et al. [27] design the Brain-Inspired Deep Networks (BDN). BDN is a multi-column CNN model, which is trained to learn attributes through the parallel pathways on a variety of selected feature dimensions. Wang also use label-preserving transformations in the context of image aesthetics assessment, which facilitates data augmentation.

The input images should always be fixed size before feeding to CNN, however, this will damage the image composition. Multi-Net Adaptive Spatial Pooling ConvNet (MNA-net) [28] is a composition-preserving structure which learns aesthetics features from input photos without any transformations. The architecture contains an adaptive spatial pooling layer to keep the input image in its original size, and a scene-based aggregation layer to combine the predictions from multiple sub-networks.

Ma et al [29] construct an Adaptive Layout-Aware Multi-Patch (A-Lamp) CNN architecture. This framework could accept arbitrary sized image and learn both fine grained details and overall image layout simultaneously. The combination of local and global attributes can effectively capture the layout of the image.

Inception modules with connected Low and Global features (ILG-Net) [30] combines the inception modules with a connected layer of both local and global features together. They fixed the shared inception layers of a pre-trained GoogLeNet model on the ImageNet [14] and fine tune the connected layer.

B. Aesthetic Distribution

A binary classification or a regression of aesthetic score could not describe the aesthetic perception well. The probability distribution of image aesthetic score can describe aesthetic subjectivity to a certain extent. For example, variance could describe the human consensus degree on an image to a certain extent, while kurtosis could describe the popularity degree of an image to a certain extent.

In 2011,Wu et al [31] start the preliminary exploration and prediction on the aesthetic distribution. In the paper, visual quality is represented by a distribution on pre-defined ordinal basic ratings. The label type is structural, not scalar. Besides, a new algorithm called support vector distribution regression (SVDR) is proposed. Murray et al. [32] apply the Huber loss combined with ResNet and SPP-Net to predict the aesthetic score distribution. Jin et al. [33] propose weighted CNNs to use sample weights for aesthetics prediction. The regression model and histogram prediction model demonstrate the effectiveness of the sample weights for reducing the bias in the training set.

Wang et al. [34] explicitly modify the score distribution of the AVA dataset as Gaussian and jointly predict its mean and standard deviation. Hou et al. [35] generate score distribution by mapping the real number labels to 10 aesthetic bins of the AADB dataset [10].

NIMA [36] is proposed to predict the distribution of human's opinion on image assessment from technical perspective and aesthetic perspective. The CNN-based model trained on both aesthetic and pixel-level quality datasets effectively predict the quality ratings distribution.

Jin Xin [37] propose a novel CNN based on the Cumulative distribution with Jensen-Shannon divergence (CJS-CNN) to predict the aesthetic score distribution. CJS-CNN could eliminate the requirement of the original full data of human ratings based on the kurtosis of the score distribution.

C. Aesthetic Factors

Existing computational approaches do not care about which attributes contribute to the quality of the image. Up until now, there are few literatures concerning aesthetic factors. To the best of our study, Kong [10] is nearly the earliest literature to study aesthetic factors.

In 2016, Kong et al. design a new Aesthetics and Attributes database (AADB), which contains a total of 10,000 images, 11 kinds of meaningful aesthetic factors for binary classification evaluation, including interesting content, object emphasis, good lighting, color harmony, vivid color, shallow depth of field, rule of thirds, balancing element, motion blur, repetition, and symmetry. They also construct a deep CNN to rank image aesthetics where the relative ranking of image aesthetics could be modeled in loss function. The method incorporates photographic attributes and content information which could regularize the image aesthetics rating issues.

Malu et al. [38] construct a novel multitask deep convolution neural network (DCNN), which jointly learns eight aesthetic attributes along with the overall aesthetic score. These attributes include balancing elements, content, color harmony, depth of field, light, object emphasis, rule of thirds, vivid colors. To understand the internal representation of these attributes, they also develop the visualization technique using back propagation of gradients, which highlight the key regions for the corresponding attributes.

D. Aesthetic Description

Although aesthetic quality assessment has attracted many researchers in the recent years, few literatures are related with aesthetic description.

In 2017, Chang et al. [11] design a new image aesthetic captioning dataset called Photo Critique Captioning Dataset (PCCD), which contains pairwise image-comment datas from professional photographers. This is the first time to produce the photo aesthetic critiques related to photography skills. Conventional image captioning task always depicts the objects and the relations in the image, while Chang's approach could select a particular aesthetics aspect and generate image captions, moreover, this method could use attention mechanism to generate more aesthetic captions. However, the size of PCCD is small (only 4307), besides, the source of the images (Gurushots.com) has already stopped to release more critiques.

To extend the cognition from rating to reasoning, a deeper understanding of aesthetics should be based on revealing the reason why the image is a high or low aesthetic image. In 2018, Wang et al. [12] propose Neural Aesthetic Image Reviewer, which can not only give an aesthetic score for an image, but also generate a textual description explaining the reason why the image leads to a high or low aesthetic score. Moreover, they construct AVA-Reviews dataset based on AVA, which contains 52,118 images and 312,708 comments in total. However, the size of the dataset is still small, and image reviews do not take aesthetic factors into account.

V. Aesthetic Applications

For an image retrieval system, aesthetics can be used in retrieving high aesthetic quality images; for picture editing software, aesthetics can be used in producing appealing polished photographs; for personal photo management, aesthetics could make it automatically and effectively. Here, we would illustrate the following applications in recent literatures.

A. Image Enhancement

Neural Image Assessment (NIMA) [36] [39] is proposed by Google researchers, who point out that quality and aesthetic scores could help image enhancement. Further, the model can be used as a filter of CNN to better adjust the appearance of the image, such as brightness, highlights and shadows. As is shown in Fig.2.

B. Image Cropping

Image cropping [40] [41] [42] improves the image aesthetic composition by removing undesired regions from an image, making the image to have higher aesthetics. As is shown in Fig.3, the cropping agent [41] starts from the whole image and takes actions to find the best cropping window in the input image.

Fig. 2. Image enhancement by multi-layer Laplacian technique along with aesthetic assessment model NIMA. Left column are the images to be enhanced, right column are the images after enhancement.

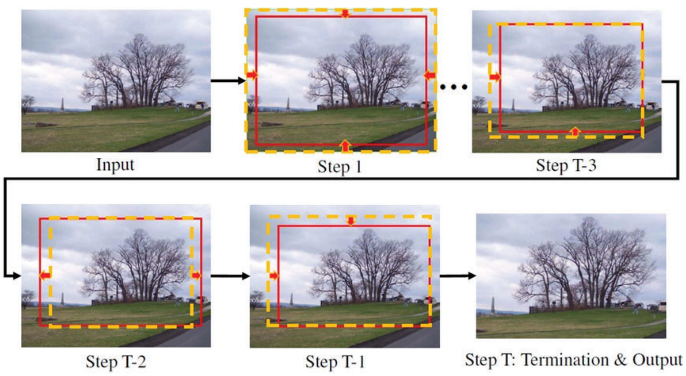

Fig. 3. Illustration of the sequential decision-making based automatic cropping process.

C. Image Retrieval

Collomosse et al. [43] propose an approach of measuring visual similarity for image retrieval, which take both structural and aesthetics into account. The input is the sketch and the images which depict the desired style, the output is the retrieval images which have both representative structure of the sketch and desired aesthetic images.

VI. Evaluation Criterions

Evaluation criterion of IAQA is an important and scientific criterion to measure the performances of the image aesthetic quality. We would illustrate different image aesthetic assessment metrics for different IAQA models in different literatures.

(1) Overall accuracy metric [6] [7] [10] [24] [25] [26] [27] [42] [43] is used most in different models, which depicts the proportion of correctly classified results;

(2) Balanced accuracy metric [44] is used while the IAQA is on an imbalanced testing datasets;

(3) Euclidean distance [13] [45] is residual sum-of-squares error between the groundtruth score and aesthetic rating score;

(4) Correlation ranking [10] [23] depicts the degree of the performance in image regression models;

(5) Earth Movers Distance (EMD) [35] [36] is a measure of similarity between any two weighted distributions.

(6) ROC curve [45] and area under the curve [1], [42] depicts the performance of binary classifier when the discrimination threshold changes;

(7) Precision-and-recall curve [15] [18] [20] depicts the degree of relevance of the retrieved items and the retrieval rate of relevant items;

(8) Mean average precision [7] [24] [25] is the average precision across multiple queries, which is used to summarize the PR-curve for the given samples.

VII. Conclusion

Up to now, although many competitive models have been proposed, the research in this field is far from saturation. Aesthetic subjectivity determines that IAQA is a very challenging task: aesthetic feature extraction is challenging; The ambiguity of learning specific categories of image aesthetics from limited auxiliary information is also challenging. Because of this, a large and diverse dataset should be built to promote the study. Moreover, we should pay more attention on the interpretability of aesthetics, and further explore the mysteries of human aesthetic intelligence.

This paper provides a survey of aesthetic quality assessment and a comprehensive description of different approaches concerning IAQA. It is noted that we have considered most of the papers published after 2014 from when deep learning first used in aesthetics. To sum up, we have investigated different datasets related with IAQA and summarized different approaches of the feature extractions. Further, we have reviewed various deep models related with 5 aesthetic tasks including aesthetic classification, aesthetic regression, aesthetic distribution, aesthetic factors, aesthetic description. Moreover, we have illustrated aesthetic applications in recent literatures. Finally, evaluation criterions are also summed up in different literatures. We wish that this survey could serve as a comprehensive reference and a good summary for future research in image aesthetics.

VIII. Acknowledgments

This work is supported by the National Key R&D Program of China under Grant No. 2017YFC1200302; R&D Program of Beijing Municipal Education Commission under Grant No. KM201910853001; Fundamental Research Funds for the Central Universities under Grant Nos.2018CUCTJ081, 2018CUCTJ080; National Science Foundation of China under Grant Nos. 71602184, 71621002; Beijing Nova Program under Grant No. Z171100001117102; Beijing Youth Talent Support Program under Grant No. CIT&TCD201704032.

References

[1] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 1930–1943, 2013.

[2] D. Joshi, R. Datta, E. Fedorovskaya, and Q. T. Luong, "Aesthetics and emotions in images," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 94–115, 2011.
[3] R. Datta and J. Z. Wang, "Acquine: aesthetic quality inference engine-real-time automatic rating of photo aesthetics," in *Proceedings of the international conference on Multimedia information retrieval*. ACM, 2010, pp. 421–424.
[4] Y. Deng, C. C. Loy, and X. Tang, "Image aesthetic assessment: An experimental survey," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 80–106, 2017.
[5] R. Datta, J. Li, and J. Z. Wang, "Algorithmic inferencing of aesthetics and emotion in natural images: An exposition," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 105–108.
[6] N. Murray, L. Marchesotti, and F. Perronnin, "Ava: A large-scale database for aesthetic visual analysis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2408–2415.
[7] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rating image aesthetics using deep learning," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2021–2034, 2015.
[8] W. Liu and Z. Wang, "A database for perceptual evaluation of image aesthetics," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 1317–1321.
[9] K. Schwarz, P. Wieschollek, and H. P. Lensch, "Will people like your image? learning the aesthetic space," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 2048–2057.
[10] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *European Conference on Computer Vision*. Springer, 2016, pp. 662–679.
[11] K.-Y. Chang, K.-H. Lu, and C.-S. Chen, "Aesthetic critiques generation for photos," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3514–3523.
[12] W. Wang, S. Yang, W. Zhang, and J. Zhang, "Neural aesthetic image reviewer," 2018.
[13] R. Sun, Z. Lian, Y. Tang, and J. Xiao, "Aesthetic visual quality evaluation of chinese handwritings," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[15] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1. IEEE, 2006, pp. 419–426.
[16] T. O. Aydın, A. Smolic, and M. Gross, "Automated aesthetic analysis of photographic images," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 1, pp. 31–42, 2015.
[17] L.-K. Wong and K.-L. Low, "Saliency-enhanced image aesthetics class prediction," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009, pp. 997–1000.
[18] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic quality classification of photographs based on color harmony," in *CVPR 2011*. IEEE, 2011, pp. 33–40.
[19] L.-Y. Lo and J.-C. Chen, "A statistic approach for photo quality assessment," in *2012 International Conference on Information Security and Intelligent Control*. IEEE, 2012, pp. 107–110.
[20] S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *CVPR 2011*. IEEE, 2011, pp. 1657–1664.
[21] L. Zhang, Y. Gao, R. Zimmermann, Q. Tian, and X. Li, "Fusion of multichannel local and global structural cues for photo aesthetics evaluation," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1419–1429, 2014.
[22] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 1784–1791.
[23] M.-C. Yeh and Y.-C. Cheng, "Relative features for photo quality assessment," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 2861–2864.
[24] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rapid: Rating pictorial aesthetics using deep learning," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 457–466.
[25] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 990–998.
[26] Y. Kao, K. Huang, and S. Maybank, "Hierarchical aesthetic quality assessment using deep convolutional neural networks," *Signal Processing: Image Communication*, vol. 47, pp. 500–510, 2016.
[27] Z. Wang, S. Chang, F. Dolcos, D. Beck, D. Liu, and T. S. Huang, "Brain-inspired deep networks for image aesthetics assessment," *arXiv preprint arXiv:1601.04155*, 2016.
[28] L. Mai, H. Jin, and F. Liu, "Composition-preserving deep photo aesthetics assessment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 497–506.
[29] S. Ma, J. Liu, and C. Wen Chen, "A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4535–4544.
[30] X. Jin, L. Wu, X. Li, X. Zhang, J. Chi, S. Peng, S. Ge, G. Zhao, and S. Li, "Ilgnet: inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation," *IET Computer Vision*, 2018.
[31] O. Wu, W. Hu, and J. Gao, "Learning to predict the perceived visual quality of photos," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 225–232.
[32] N. Murray and A. Gordo, "A deep architecture for unified aesthetic prediction," *arXiv preprint arXiv:1708.04890*, 2017.
[33] B. Jin, M. V. O. Segovia, and S. Süsstrunk, "Image aesthetic predictors based on weighted cnns," in *2016 IEEE International Conference on Image Processing (ICIP)*. Ieee, 2016, pp. 2291–2295.
[34] Z. Wang, D. Liu, S. Chang, F. Dolcos, D. Beck, and T. Huang, "Image aesthetics assessment using deep chatterjee's machine," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 941–948.
[35] L. Hou, C.-P. Yu, and D. Samaras, "Squared earth mover's distance-based loss for training deep neural networks," *arXiv preprint arXiv:1611.05916*, 2016.
[36] H. Talebi and P. Milanfar, "Nima: Neural image assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998–4011, 2018.
[37] X. Jin, L. Wu, X. Li, S. Chen, S. Peng, J. Chi, S. Ge, C. Song, and G. Zhao, "Predicting aesthetic score distribution through cumulative jensen-shannon divergence," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
[38] G. Malu, R. S. Bapi, and B. Indurkhya, "Learning photography aesthetics with deep cnns," *arXiv preprint arXiv:1707.03981*, 2017.
[39] H. Talebi and P. Milanfar, "Learned perceptual image enhancement," in *2018 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2018, pp. 1–13.
[40] Y. Kao, R. He, and K. Huang, "Automatic image cropping with aesthetic map and gradient energy map," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1982–1986.
[41] D. Li, H. Wu, J. Zhang, and K. Huang, "A2-rl: aesthetics aware reinforcement learning for image cropping," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8193–8201.
[42] J. Chen, G. Bai, S. Liang, and Z. Li, "Automatic image cropping: A computational complexity study," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 507–515.
[43] J. Collomosse, T. Bui, M. J. Wilber, C. Fang, and H. Jin, "Sketching with style: Visual search with sketches and aesthetic context," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2660–2668.
[44] C. Huang, Y. Li, C. Change Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5375–5384.
[45] A. Lienhard, P. Ladret, and A. Caplier, "Low level features for quality assessment of facial images," in *10th International Conference on Computer Vision Theory and Applications (VISAPP 2015)*, 2015, pp. 545–552.