International Journal of Engineering Research And Management (IJERM)
ISSN: 2349- 2058, Volume-07, Issue-09, September 2020

Image Aesthetic Assessment Based on Deep Learning: An Survey

Yaoting Wang

***Abstract*—With the rise of deep learning in recent years, the field of image aesthetics quality evaluation has developed from traditional machine learning methods to end-to-end convolutional neural network evaluation methods, which has achieved a qualitative leap in the results of image aesthetics quality evaluation. This article mainly summarizes and introduces the research of convolutional neural network methods in image aesthetics evaluation in recent years. It aims to solve the problems of incomplete generalization and insufficient understanding of the existing review literature. Explains in detail the development from manual feature extraction to deep learning, image aesthetics related data sets, and various application directions of image aesthetics evaluation, including automatic image cropping based on image aesthetics, image semantic line detection, image composition classification, and image aesthetic attributes Analysis etc. Finally, the future work in the direction of image aesthetics is prospected.**

***Index Terms*—Convolutional Neural Network, deep learning, feature extraction, image aesthetic evaluation, Summary.**

I. INTRODUCTION

With the rapid popularization of smartphones and the substantial increase in network speed and bandwidth, the amount of visual data such as images and videos has exploded. The screening of high-quality data has also become more arduous. With the pursuit of beautiful things by humankind, researchers have also begun to study how to make computers automatically evaluate and screen images. Therefore, a series of research directions have been born, image quality evaluation and Image aesthetics quality evaluation falls into this category. Different from the image quality evaluation, although the purpose is to allow the computer to obtain an objective evaluation value consistent with the subjective evaluation result, the image aesthetic quality evaluation aims to make the computer simulate the human perception of the beauty of the image and make an automatic evaluation of the beauty of the image. This requires the computer model to analyze not only the low-level features of the image, but also the high-level semantic features of the image. In the evaluation of image aesthetics quality, not only the image quality itself (noise, distortion, exposure, etc.) must be considered, but also aesthetic attributes such as image composition, depth of field, color, light and shadow, etc., combined with psychology and photography theory, to gradually achieve aesthetic quality Evaluate this requirement. The image aesthetics quality evaluation has undergone a transition from traditional manual feature extraction combined with machine learning methods to end-to-end deep learning methods, and the evaluation results have also been qualitatively improved. At the same time, the goal of image aesthetics evaluation has evolved from only distinguishing good (beautiful) images from bad (ugly) to regression (predicting an aesthetic score) to predicting the aesthetic evaluation distribution of images. This article will introduce this evolution in detail.

With the gradual expansion of image aesthetics quality evaluation scale and the rise of deep learning, small-scale data sets used for manual feature extraction and traditional machine learning methods are no longer sufficient to support research, so some large-scale image aesthetics data sets are constructed, AVA[1], AADB[2], etc., we will introduce below. At the same time, the research of aesthetic images has also stepped out of the laboratory and stepped into all aspects of people's lives, resulting in a large number of applications, such as automatic cropping of image composition, automatic image filtering, automatic image restoration, etc. We will briefly introduce them below. Finally, we will make a summary of the research on the direction of image aesthetics and look forward to future trends.

II. AESTHETIC QUALITY EVALUATION METHOD

A. Traditional image aesthetics evaluation method

Traditional aesthetic evaluation methods are mainly based on purely manual feature design and extraction combined with machine learning algorithms to evaluate images. The evaluation of image aesthetic quality by computer fitting humans is mainly divided into two aspects, classification (good and bad) and regression (value). It is mainly divided into three steps, collecting data sets, designing and extracting features, and using machine learning algorithms to train the model based on the extracted features to achieve the evaluation of the aesthetic quality of the image.

The earliest research on image aesthetics evaluation was put forward by Microsoft Research Asia[3] in 2004. This work constructed the initial data set of aesthetic evaluation with nearly 30,000 images, including 16,643 professional photographic photos and 12,897 snapshots, but it was not open. At the same time, 846-dimensional features are used and the classification model is trained using an ensemble algorithm to classify images into two categories, good and bad. Table I shows the experimental results of this work.

Table I. Testing error of different mothed.

Ada-Boost	Real-AdaBoost	SVM	Bayesia

Manuscript received September 03, 2020

Yaoting Wang, School of Computer Science & Technology, Tiangong University, Tianjin, China

				n
error	8.9%	6.6%	6.1 %	4.9%

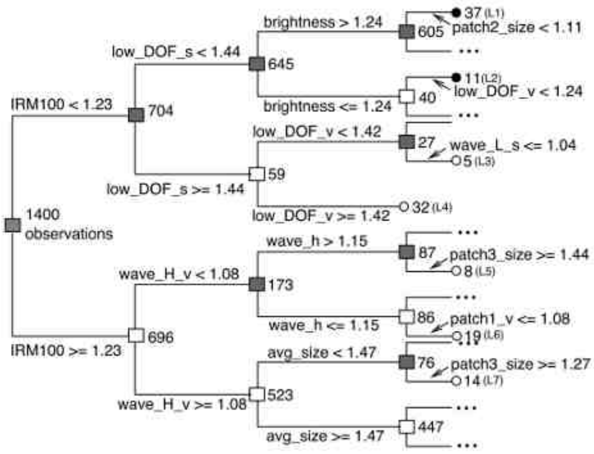

Fig 1. Decision tree of feature

So far, traditional image aesthetics quality evaluation has been divided into two categories, one is based on the low-level features of the image and has nothing to do with the content, and the other is based on the content of the image.

Content-independent image aesthetics methods are mainly focused on the design of visual features. In 2006, Datta et al.[4]designed 56-dimensional image features, including global and local features, and studied the effects of these features on aesthetic evaluation through random forest algorithms. Influence, as shown in Fig 1. Wong et al.[5] proposed image salient features for image aesthetic evaluation and discussed the relationship between image local features. Then some studies put forward a series of features (color, mood, composition, scene, etc.).

However, none of the aforementioned works considers the influence of image content on aesthetic evaluation, so the research direction of researchers gradually shifts to content-based aesthetic evaluation. The more famous one is the content-based image quality evaluation method proposed by Tang et al.[6] in 2011. This work divides the image into 7 different scenes (architecture, animal, night scene, portrait, still life, plant, landscape), as shown in the Fig 2 and extract different features according to different scenes, analyze them, and use SVM model to evaluate the aesthetic quality of images. Since then, a large number of aesthetic evaluation studies on images with different content have been derived, mainly focusing on the study of portraits and landscape images[7][8][9][10].

B. Evaluation method of image aesthetics based on deep learning

With the introduction of Alexnet[11] in 2012 and the release of the AVA dataset[1], the evaluation of image aesthetics quality has entered a new stage-the era of deep learning, abandoning the complicated manual feature design, and extracting features through end-to-end convolutional neural networks. Realizing the classification or scoring of image aesthetics greatly improves the accuracy of evaluation. Researchers have developed many high-performance image aesthetics evaluation models by transforming the neural network used for image classification, which also makes computer image aesthetics evaluation out of the laboratory and truly applied to people's lives. Thanks to the greatly improved evaluation accuracy, many new tasks have evolved, such as aesthetic description, aesthetic attributes, aesthetic distribution prediction, and aesthetic evaluation based on the attention mechanism.

This article introduces several landmark or representative articles. In 2014, Xin Lu et al.[12] proposed an evaluation method called RAPID. For the first time, the deep learning method was applied to the evaluation of image aesthetics. The two-column convolutional neural network was used to extract features and classify them. The data set used was the AVA dataset. In 2016, Kong et al. proposed a new dataset AADB[2], which contains the aesthetic attributes of the image (BalancElement, ColorHarmony, InterestContent, ShallowDOF, Good lighting, OBjectEmphsis, RuleOfThirds, VividColor). This dataset plays a vital role for future research and analysis. At the same time, this work no longer simply divides images into two categories, good or bad, but can score images (1 to 5 points), and then can rank the aesthetic quality of a group of images. Since the label of the image in the AVA dataset is not a score or category, but the number of people who rated 1 to 10 points, in 2018, Google[13] proposed a new loss function that can directly predict the distribution of image scores, or through the predicted distribution is used to calculate the aesthetic score and category of the image, and it achieves the best effect at the time. Tested on the AVA dataset, the accuracy rate reached 81.51%. At the same time, you can also change the image parameters and then score to achieve the purpose of image enhancement. As the image shows, gradually increase the noise of the image, and the score will gradually decrease. In 2019, Xin Jin et al. proposed ILGNet[14], which is the state of art of the current public data set AVA, and the accuracy of the two classifications reached 85.53%. They use the inceptionNet[15] proposed by Google as the network skeleton, using multi-scale feature map information, making the model richer in the low-level features of the image.

Figure 2. Images with different content, the first line is high aesthetic quality images, the second line is low aesthetic quality images

III. DATASETS OF IMAGE AESTHETICS

Image aesthetics evaluation is a research direction that has emerged in the past ten years, so the data-driven construction method was followed in the construction of the data set from the beginning, rather than the traditional rule-based construction method. At present, most aesthetic data sets are constructed according to the following steps: first download images on open source image websites, and secondly collect artificial subjective ratings (outsourced, collected from existing online rating websites). As the types of tasks increase, the annotation content of the data set becomes more and more abundant.

In 2012, with the deepening of research in the field of computational aesthetics, Murray et al.[1] constructed a large-scale aesthetic image dataset AVA, which is still widely used and has been used as the standard data set in this field. It contains 250,000 images, all of which are downloaded from an online sharing site (dpchallenge.com), and each image is rated by 78 to 539 people, with a score ranging from 0 to 10. In addition to the score distribution labels, the data set also contains 60 kinds of labels such as image theme and style. However, this data set is mainly used for the task of image aesthetics evaluation, and cannot satisfy the detailed analysis of aesthetic images.

In 2016, Kong et al.[2] proposed an analysis of the aesthetic properties of images, thereby constructing a large data set AADB. In addition to the necessary score tags, the data set also contains 8 aesthetic properties (BalancElement, ColorHarmony, InterestContent, ShallowDOF, Good lighting, OBjectEmphsis, RuleOfThirds, VividColor), making the research on the attributes of image aesthetics one step closer. However, these eight aesthetic attributes only have binary classification evaluation, which is too simple to be quantified.

Due to the rapid development of NLP, researchers are no longer satisfied with simple image scoring, and have evolved to hope that the computer can make a true linguistic evaluation of images. In 2017, Chang et al.[16] constructed a brand new image aesthetics data set, added language-based comment information to the tags, and also included scores of 6 aesthetic factors and aesthetic scores. However, there are too few images in this dataset, only 4307 images.

IV. APPLICATIONS OF COMPUTATIONAL AESTHETICS

The dramatic increase in the accuracy of image aesthetics evaluation has spawned many related applications. Applications such as image cropping, automatic image composition, image enhancement, and style conversion are also evolving.

The purpose of cropping based on image aesthetics evaluation is to crop out the part with the highest aesthetic evaluation in the image, so as to improve the image composition and the aesthetic quality of the image. In 2017, Wang et al.[17] proposed to use the attention mechanism-based cropping method. First, a series of cropping frames were generated through the attention mechanism, and then the edge regression of the cropping frames was performed using an aesthetic evaluation model to select the best cropping method. In 2018, adobe[18] proposed a new network structure. This model can select multiple sub-images that meet the composition rules and have high aesthetic evaluation in the panoramic image.

The research of image aesthetics evaluation directly affects the research of image enhancement. The NIMA[13] model proposed by Google in 2018 can sort a group of photos with different shooting parameters to achieve the purpose of image enhancement.

Image aesthetic evaluation is also applied to all aspects of life, clothing aesthetic evaluation, ink painting aesthetic evaluation, oil painting aesthetic evaluation and so on. It is also widely used in business. It can recommend good-looking images for businesses, automatically select the most beautiful frame from the video as the cover, and so on.

V. CONCLUSION AND POTENTIAL DIRECTIONS

This article summarizes the existing image aesthetics evaluation methods, aesthetics-related data sets, and the applications that have spawned with the development of aesthetics evaluation research. Although the evaluation of image aesthetics has developed rapidly, there are still shortcomings. Images have different scenes and different contents, and their composition rules and aesthetic elements should be very different. How to divide them carefully or find a common method is a problem that needs to be solved in the future.

REFERENCES

[1] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis." pp. 2408–2415, 2012.

[2] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo Aesthetics Ranking Network with Attributes and Content Adaptation," vol. 9905, 2016, pp. 662–679.

[3] H. Tong, M. Li, H.-J. Zhang, J. He, and C. Zhang, "Classification of digital photos taken by photographers or home users," 2004, pp. 198–205.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," vol. 3953, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 288–301.

[5] L. K. Wong and K.-L. Low, "Saliency-enhanced image aesthetics class prediction," in *Proceedings - International Conference on Image Processing, ICIP*, 2009, pp. 997–1000, doi: 10.1109/ICIP.2009.5413825.

[6] W. Luo, X. Wang, and X. Tang, "Content-based photo quality assessment," 2011, pp. 2206–2213.

[7] A. Lienhard, P. Ladret, and A. Caplier, "Low level features for quality assessment of facial images," *VISAPP 2015 - 10th Int. Conf. Comput. Vis. Theory Appl. VISIGRAPP, Proc.*, vol. 1, pp. 545–552, 2015, doi: 10.5220/0005308805450552.

[8] M. Redi, N. Rasiwasia, G. Aggarwal, and A. Jaimes, "The beauty of capturing faces: Rating the quality of digital portraits," *2015 11th IEEE Int. Conf. Work. Autom. Face Gesture Recognition, FG 2015*, 2015, doi: 10.1109/FG.2015.7163086.

[9] C. Li, A. Gallagher, A. C. Loui, and T. Chen, "Aesthetic quality assessment of consumer photos with faces." pp. 3221–3224, 2010.

[10] R. Maity and S. Bhattacharya, "Is My Interface Beautiful? - A Computational Model-Based Approach," *IEEE Trans. Comput.*

Soc. Syst., vol. 6, no. 1, pp. 149–161, 2019, doi: 10.1109/TCSS.2019.2891126.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.*, vol. 2, pp. 1097–1105, 2012.
[12] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, “Rapid: Rating pictorial aesthetics using deep learning,” *MM 2014 - Proc. 2014 ACM Conf. Multimed.*, pp. 457–466, 2014, doi: 10.1145/2647868.2654927.
[13] H. Talebi and P. Milanfar, “NIMA: Neural Image Assessment,” *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, 2018, doi: 10.1109/TIP.2018.2831899.
[14] X. Jin *et al.*, “ILGNet: inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation,” *IET Comput. Vis.*, vol. 13, no. 2, pp. 206–212, 2019, doi: 10.1049/iet-cvi.2018.5249.
[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.
[16] Kuang-Yu Chang, K. Lu, and C. Chen, “Aesthetic Critiques Generation for Photos,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3534–3543, doi: 10.1109/ICCV.2017.380.
[17] W. Wang and J. Shen, “Deep Cropping via Attention Box Prediction and Aesthetics Assessment,” in *international conference on computer vision*, 2017, pp. 2205–2213.
[18] Z. Wei *et al.*, “Good View Hunting: Learning Photo Composition from Dense View Pairs,” in *computer vision and pattern recognition*, 2018, pp. 2225–2231.

Yaoting Wang graduated from Tianjin Polytechnic University with a bachelor's degree, currently studying for postgraduate in Tiangong University.