ITM Web of Conferences **70**, 03004 (2025) https://doi.org/10.1051/itmconf/20257003004
DAI 2024

Perspectives on Image Aesthetic Evaluation Techniques

Kaiwei Yang

Beijing New Channel - Dongfanghong School, Beijing, China

Abstract. With the advancement of technology, support of datasets, cross-fertilization of disciplines, and increasing demand for reference scenarios, image aesthetics has become an active research field and has been widely applied in real life. This paper discusses the application of deep learning in image aesthetics evaluation, including the selection of datasets and the process of image aesthetics evaluation (including aesthetics feature extraction and aesthetics evaluation decision-making), analyzes the application of aesthetics evaluation techniques, and looks forward to the technological prospect. Although deep learning technology can improve the accuracy of evaluation and has achieved some results in aesthetic evaluation tasks, the technology still faces challenges such as high subjectivity and poor diversity, and it is believed that with the progress of science and technology, these problems will be solved. This paper also provides basic information on image aesthetic evaluation, analyzes the research trends in image aesthetic evaluation, and provides insights into future research directions.

1 Introduction

Image aesthetic evaluation refers to the use of computers instead of humans to analyze and judge the visual aesthetics of an image. This task involves not only the analysis of the basic features in an image but also the comprehensive consideration of a variety of complex factors, such as composition, color, light, and shadow effects, details, emotional expression, and so on [1]. Since each person's understanding and feeling of beauty may vary significantly, aesthetic evaluation is inherently highly subjective and complex, which makes image aesthetic evaluation a challenging research topic in the field of computer vision and artificial intelligence.

Deep learning is a machine learning method based on artificial neural networks, which builds and trains multi-level neural networks to achieve a learning understanding of data. In recent years, deep learning has made significant progress in computer vision fields such as image classification, object recognition, face recognition, etc., which has greatly promoted the development and application of related technologies. With the advancement of deep learning, automated image aesthetics evaluation procedures have also become possible [2].

This paper will introduce the aesthetic evaluation in detail in the following three aspects:

Corresponding author: 1812010802@stu.hrbust.edu.cn

(1) Generalized summary of the evaluation process of image aesthetics and operational methods. The image aesthetics evaluation process consists of two main steps: feature extraction and the decision-making phase. This paper will introduce the basic concepts of aesthetic feature extraction and aesthetic evaluation decision-making, describe the innovations and specific operational methods of the evaluation process, and compare them with traditional methods.

(2) Common datasets for image aesthetic evaluation tasks. A dataset is a collection of data that is collected and organized together, which is crucial for research and practice in various fields. A high-quality dataset can significantly improve the performance of machine learning algorithms because it provides the basis for the algorithm to learn patterns and regularities. To summarize, dataset selection is an integral part of image aesthetic evaluation.

(3) General description of the development status of image aesthetics evaluation. By citing existing research results, the development process of image aesthetics evaluation is visually described, and the field and direction of the existing results are outlined and summarized to facilitate the understanding of the research progress, technological trends, and future directions in this field.

2 Image aesthetic evaluation

Image aesthetic evaluation is mainly divided into two parts: aesthetic feature extraction and aesthetic evaluation decision. Aesthetic feature extraction refers to extracting features from an image that can reflect its aesthetic quality. These features are usually high-dimensional and describe the visual content and structural information of the image. Aesthetic evaluation decision-making refers to evaluating or categorizing the aesthetic quality of an image based on the extracted features and giving a specific aesthetic score or category label to the evaluated target.

2.1 Aesthetic feature extraction

Aesthetic feature extraction is classified into three types: full reference, half reference, and no reference. Full reference requires a one-to-one comparison between the image to be evaluated and the pixel points of the reference image. Semi-reference only requires the image to be evaluated to compare some of its features with those of the reference image. No reference does not require a reference image and directly evaluates the aesthetic quality of the target image. This is currently a research focus in the field. Aesthetic feature extraction technology can automatically extract features from images that can characterize their beauty by simulating human cognition and perception of beauty. Aesthetic feature extraction is the foundation of image aesthetic evaluation. It is the first step in both traditional methods and deep learning-based methods. Aesthetic feature extraction technology can identify and extract image aesthetic features, improve the accuracy of aesthetic evaluation, and achieve overall aesthetic evaluation of images that are closer to human perception.

Before the popularity of deep learning, most research on image aesthetic quality assessment focused on extracting the aesthetic features of an image. For example, the RGB format image is converted to the HSV format, which is a color model representing hue, saturation, and lightness designed to align with human perception of color. This format is commonly used in image processing and computer vision tasks. Researchers also designed underlying features such as shape, color, and texture, as well as high-level features like regional contrast and the three-part composition method. Effective features were then selected to train the classifier for aesthetic quality classification [3]. These methods have achieved some results, but they are complex and have limited effects. With the development of deep learning, an increasing number of researchers have started using deep neural networks, specifically Convolutional Neural Networks (CNN), to automatically extract image features. CNNs are trained on a large amount of labeled image data with aesthetic scores. The CNN model can learn the relationship between different aesthetic features and aesthetic scores, enabling it to predict the aesthetic scores for new images [4].

The input layer of the CNN is used to receive raw image data consisting of three-color channels to form a two-dimensional matrix to represent the intensity values of the pixels. The convolutional layer is used to perform the convolutional operation, that is, to extract the aesthetic features from the image, using different convolutional kernels with different features, such as composition, color matching, and lighting effects. After convolution, nonlinearities are introduced through the activation function to provide the network with the ability to learn complex features. Finally, a dimensionality reduction operation is performed, and a pooling layer is used to reduce the spatial size of the feature map, thereby reducing the computational complexity. After multiple convolution and pooling layers, the fully connected layer maps the extracted high-level features to regression values or classification labels [5]. By constructing different CNN architectures and adopting integrated methods, the performance of image aesthetic score prediction can be further improved [6].

2.2 Aesthetic evaluation decision making

Aesthetic evaluation is the final and decisive step in deep learning-based image aesthetic decision-making. This is because a good evaluation strategy determines whether the model can effectively learn information from the data and whether it can have a positive evaluation effect on new data. Based on the purpose of the evaluation, the overall classification of aesthetic evaluations can be categorized into three types. The first type is classification decisions, which classify images into different aesthetic categories such as high aesthetic quality, medium aesthetic quality, and low aesthetic quality. This type of evaluation realizes the process of aesthetic decision-making on images through a classification task. The second type is regression decisions, which assign a continuous aesthetic score to an image to reflect its aesthetic quality. The third type is sorting decisions, which sort a series of images to reflect the relative degree of aesthetic quality through a specified decision-making process. Sorting decisions are used to demonstrate the relationship between high- and low-quality sample images [7]. Each of these three types of decisions has its own specific application scenario and evaluation index for deep learning. By selecting and optimizing the corresponding algorithms, practical problems can be effectively solved and the prediction performance can be improved.

3 Image aesthetics dataset

The evaluation of image aesthetics is not solely a technical matter, but also encompasses human perception. Consequently, the annotations within the dataset should strive to closely align with the subjective evaluations made by humans. Human aesthetic evaluation is influenced by factors such as culture, experience, and personal preference. Therefore, the dataset should comprehensively account for these factors in order to enable the model to better approximate real-world scenarios.

Commonly used datasets are the AVA datasets and others, and the specific information is shown in Table 1.

Table 1. Dataset Comparison Chart

Dataset	data volume	marking scheme	Type of labeling
AVA	250000+	1-10	66+14
AADB	10000	High-low	11
AROD	380000+	0-1	-
CUHKP	30000	High-low	7
PN	20278	0-7	-

The AVA dataset contains more than 250,000 images, each of which was scored by a different number of evaluators, ranging from 78 to 539, for an average of approximately 210 ratings. The scoring system used a scale of 1 to 10, resulting in more comprehensive and reliable aesthetic ratings for each image in the dataset. The dataset contains not only aesthetic ratings but also 66 categories of semantic annotations and 14 categories of photographic style annotations, which provide rich metadata for the study and help explore the multiple factors that affect aesthetic ratings [8]. The AVA dataset has significant advantages in terms of the size, diversity, and heterogeneity of annotations, which makes it ideal for studying computational models of aesthetic preferences. It is worth noting that some of the images have heavy traces during post-processing, which may have an impact on the learning of the machine learning model. In addition, because the vast majority of images originate from professional photographers, this introduces some bias into the dataset. The AADB dataset was designed in 2016 by a team of researchers from the University of California, Irvine. The dataset focuses on the aesthetic factors of an image and aims to improve the understanding and assessment of the aesthetics of an image through an in-depth analysis of these factors. The AADB dataset consists of eight aesthetic factors: balance, color harmony, interest, depth of field, lighting, subject, third, and color richness. Each aesthetic factor is evaluated using a binary classification, labeled as good or bad. However, this type of evaluation has also been noted to be oversimplified and may not adequately reflect the subjectivity and diversity of aesthetic evaluations. The AROD dataset is a large image aesthetics dataset constructed by Schwarz et al. at the University of Tübingen, Germany, which aims to derive aesthetically scored annotation data for images by calculating the number of times images are viewed versus the number of times they are liked on the online image-sharing site Flickr. The dataset contains 380,000 images and is primarily used for image aesthetic score estimation. The AROD dataset is significant within the field of image aesthetic quality assessment because it not only provides a large-scale database for research use but also utilizes for the first time social interaction data (e.g., number of views and likes) of images to evaluate the aesthetic quality of images. This approach brings a new perspective to image aesthetics evaluation, that is, understanding the popularity of an image through user interaction behavior, which indirectly reflects its aesthetic value.

The CUHKPQ dataset covers 17,309 images, each of which received an aesthetic rating given by real users through a comparative evaluation method, which ensures the consistency and reliability of the ratings.

https://doi.org/10.1051/itmconf/20257003004

The PN dataset contains 20,278 images, where each image received an average of 12 ratings ranging from 0 to 7, with higher scores indicating higher quality images. More than 30% of the images in the PN dataset had a frame added by their photographer to enhance aesthetics, which may result in high ratings for these images.

In summary, the choice of the dataset is a decision that needs to be carefully considered because it determines the quality of the model training and the generalization ability, which can usually affect the experimental results. To ensure the success of the experiment and the reliability of the model, the researcher should consider the size, quality, diversity, division, random seed setting, representativeness, and timeliness of the dataset.

4 Application of image aesthetics evaluation

Image aesthetic assessment has a wide range of applications in many fields. For example, in the field of photography and videography, image aesthetics assessment technology helps photographers and videographers assess the aesthetic effect of an image in real time during the shooting process to adjust shooting parameters and composition [9]. It can also provide personalized creative suggestions based on photographer-style preferences. Simultaneously, image aesthetic evaluation technology can combine user behavior data and image aesthetic evaluation to recommend content that meets user aesthetic preferences. Social media platforms need to review uploaded images to ensure that the content meets specifications. Image aesthetic assessment technology can help identify low-quality or non-compliant images to minimize the distribution of undesirable content. On the other hand, in the field of graphic design, image aesthetics evaluation techniques can be used in advertising and poster design to help designers create more creative and attractive works. In the field of clothing design, image aesthetic evaluation technology can provide style design and matching suggestions to help designers create more attractive clothing. In the field of film and television production, image aesthetic evaluation technology can be used in video editing and the production of special effects to help producers choose the best combination of shots and scenes to enhance the audience's visual experience.

Despite significant advances in image aesthetic evaluation techniques in recent years, several limitations and challenges remain. One of the core challenges of image aesthetic evaluation is its highly subjective nature. Each individual has different definitions and feelings of beauty, and such differences make it difficult for automated evaluation algorithms to model aesthetic preferences accurately. In addition, cultural differences bring additional complexity to image aesthetic evaluation, requiring algorithms to be able to generalize across cultures. At the same time, the generalization ability of the usual models is flawed. This is because image aesthetics are influenced by a variety of factors, including composition, color, light and shadow, depth of field, reality, and falsehood. These factors interact with each other and determine the overall aesthetics of an image. Therefore, ignoring any of these factors may affect the accuracy of the evaluation results. Finally, the internal workings and decision-making processes of deep learning models are difficult to understand and explain [10]. The poor interpretability of the model makes it difficult to trust and adopt its evaluation results. For example, the application of image aesthetic evaluation techniques in the medical field is questionable because people cannot fully trust robots.

5 Conclusion

The utilization of deep learning in image aesthetics not only enhances the efficiency of image processing but also enhances individuals' perception and expression of beauty. Specifically, the deep learning model can autonomously acquire the various levels of features present in an image, such as color, texture, and composition, and subsequently conduct aesthetic scoring based on these features. Despite the existence of certain limitations in the application of deep learning in image aesthetics, these studies will not only contribute to the advancement of scientific and standardized evaluation of image aesthetics but also offer novel ideas and methodologies for the integration of artificial intelligence in the realm of art and design. In the future, as technology continues to progress and interdisciplinary collaboration deepens, image aesthetic evaluation will demonstrate its distinctive value and potential in a wider range of fields.

References

1. P. Obrador, L. Schmidt-Hackenberg, N. Oliver, The role of image composition in image aesthetics, in Proceedings of the 2010 IEEE International Conference on Image Processing, (2010).
2. H. Jang, Y. Lee, J.-S. Lee, Modeling, Quantifying, and Predicting Subjectivity of Image Aesthetics. arXiv preprint arXiv:2208.09666 (2022).
3. C.-H. Su, H.-S. Chiu, J.-H. Hung, et al. Color space comparison between RGB and HSV based images retrieval. Adv. Mat. Res. **989**, 4123-4126 (2014).
4. A. Krizhevsky, I. Sutskever, G.-E. Hinton, ImageNet classification with deep convolutional neural networks. Commun. ACM. **60**, 84-90 (2017).
5. S. Kanwal, M. Uzair H, A Survey of Hand Crafted and Deep Learning Methods for Image Aesthetic Assessment. arXiv preprint arXiv:2103.11616 (2021).
6. Y. Dai. Exploring CNN-based models for image's aesthetic score prediction with using ensemble. arXiv preprint arXiv:2210.05119 (2022).
7. J. Zhou, Q. Zhang, et al. Joint regression and learning from pairwise rankings for personalized image aesthetic assessment. Com.Vis.Med. 241-252 (2021).
8. N. Murray, L. Marchesotti, F. Perronninx, AVA: A large-scale database for aesthetic visual analysis, in Proceedings of the 2012 IEEE conference on computer vision and pattern recognition, (2012).
9. T.-O. Aydın, A. Smolic, M. Gross. Automated aesthetic analysis of photographic images. IEEE Trans. Vis. Comput. Graph. **21**, 31-42 (2014).
10. Q. Zhang, S.-C. Zhu. Visual interpretability for deep learning: a survey. Front. Inf. Technol. Electron. Eng. **19**, 27-39 (2018).