



REVIEW OF THE RESEARCH STATUS OF IMAGE AESTHETIC QUALITY

Shasha Lv

College of Computer Science and Technology, Tianjin Polytechnic University, Tianjin, China.

ABSTRACT

As the popularity of photo taking devices leads to more and more pictures being stored by everyone, various ways of describing pictures, especially in terms of aesthetic quality evaluation, have attracted great attention in the past decade or so. The purpose of the research on image aesthetic quality is to enable the computer to simulate human thinking and aesthetics to judge the aesthetic value of a picture, so as to output score or text description. From the level of art, image aesthetics is a subject that shows natural aesthetics through the integration of various photographic skills such as color collocation, depth of field processing and composition design. The research involves computer vision, psychology, text description, cognitive science and other aspects, there is a phenomenon of interdisciplinary research, has important research significance. Therefore, we will systematically analyze and summarize the research status of image aesthetic quality, and give Suggestions on the development of image aesthetic quality research.

KEYWORDS: aesthetic quality evaluation; Research on image aesthetic quality; Computer vision.

1. INTRODUCTION:

Aesthetics, a branch of philosophy, is a subject that studies the essence and significance of beauty. The main object of the study is art, but it does not study the specific expression of art, but the philosophy of art, so it is called "the philosophy of art of beauty". The basic problems of aesthetics include the essence of beauty, the relationship between aesthetic consciousness and aesthetic object, etc.

Nowadays, with the wide popularization of photographic equipment, the perception and understanding of image content has become the key research direction in computer vision, computational photography and other disciplines. In addition, more and more aesthetic works are created, preserved and disseminated digitally as images, including paintings, films, video and computer games. Meanwhile, people are exposed to various human-computer interaction interfaces every day, hoping to create beautiful enjoyment and pleasure. In the face of the increasingly surging mass data, human beings are no longer competent for complex processing work, so it is hoped that computers can help humans to conduct auxiliary aesthetic processing. The content processed by computer is no longer limited to formulaic computing requirements, but also hopes that the computer can help humans to carry out subjective work, such as intelligent robot, computer image aesthetic processing, image content recognition, image aesthetic rating, image aesthetic comment and so on.

Therefore, in this digital age, whether the computer can help people in aesthetic aspects of what to do, to what extent, is a very worthy of discussion. At the same time, because of aesthetics itself is a discipline of hard to quantify, for evaluation of beauty is different also, so to be able to use a computer to system of aesthetics, for the intelligent computer field is also very big breakthroughs, as well as artificial intelligence, computer simulation of other human behavior of the subsequent research laid the foundation. Therefore, this paper focuses on the current situation of image aesthetics research, summarizes the existing image aesthetics research, summarizes the latest international research results, and systematically summarizes and analyzes the research methods and existing problems.

2. RESEARCH STATUS OF IMAGE AESTHETICS:

In this section, we mainly analyze two aspects: the first is the classification of image aesthetics, and the other is the description of generative aesthetics.

2.1 Aesthetic Classification of Images:

The typical classification process of image aesthetics classification is to classify the concerned pictures into high or low aesthetics under the supervised machine learning background. There are many factors influencing image aesthetics, including light and shadow, color, composition, depth of field, clarity, etc., which have different rules in different aspects. There are many studies on image aesthetic quality from these aspects [1-3]. Datta[1] considers many aspects of image aesthetics, such as bottom image color, texture, shape, image size, top image depth, rule of thirds, and regional contrast (image is divided into several parts by segmentation algorithm to calculate the contrast between regions, etc.). However, the characteristics of its structure are generally very simple and low-level, such as using only the HSV average component of the central region to represent the characteristics of the rule of thirds.

The initial application of deep learning in image aesthetic classification was proposed by A. Krizhevsky *et al.* [4]. They trained a large deep convolutional neural network and divided 1.2 million high-definition images in the Image Net Isvrc-

2010 data set into 1,000 different categories. The classification network has 60 million parameters and 650,000 neurons, with a total of five convolution layers, some of which are followed by the maximum pooling layer, and the network USES softmax function to realize the last three full connection layers of 1000 categories. Meanwhile, in order to make the training speed faster, efficient GPU is used to realize convolution operation. In order to reduce over fitting, Dropout regularization method is adopted. This inspires the application of deep learning in image aesthetic classification. Subsequently, Kang L, Ye P, Li Y *et al.* [5] considered that the effect of using CNN directly for image quality evaluation is not good, because CNN extracts local features and these features will not change. Therefore, Kang L, Ye P, Li Y *et al.* modified the network structure of CNN, and improved the classification performance by extracting global and local features from images, and could evaluate the quality of local images.

2.2 Aesthetic Description of Images:

Before the discussion according to the description of the image generation aesthetic, we should start with the image caption began, and then the natural transition to the image of aesthetics caption.

Image Caption is a multi-disciplinary synthesis problem, which combines three aspects of computer vision, language processing and machine learning. Image Caption problem can be defined as binary group (I,S), where I represents the graph, S is the target word sequence, where $S = \{S_1, S_2, \dots\}$, where S_t is the word extracted from the data set. The goal of training is to maximize the maximum likelihood $p(S|I)$, and even if the generated statement is more compatible with the target statement, it can be expressed to describe the image in terms as accurate as possible.

Image Caption's main research is divided into the following directions:

1. Use separate CNN to obtain image features, and then use these features to generate sentences (sort, retrieve, generate);
 2. Embed the features obtained by CNN and sentence features into a space, and select the optimal description from them;
 3. Combining CNN and RNN with some new mechanisms, the purpose is to use the global feature or local feature of CNN to guide the generation of description.
- The above three are similar to neural network translation
4. Fill in some objects in the image by using the template method;
 5. Search for similar descriptions by using the retrieval method. Both methods use a generalization that makes the description close to the picture, but not quite accurate.

There are many methods [6-15] that achieve good results by using natural language to describe images and entities in video. Vinyals O, Toshev A, Bengio S, *et al.* [6] proposed NIC model, which is also the simplest encoder-decoder model. This model USES CNN to extract features, connects to LSTM after the full connection layer of CNN, and USES LSTM to generate target language. The model mainly includes encoder-decoder. Encoder part is a convolutional neural net-

work used to extract image features, which can adopt VGG16, VGG19, GoogleNet and other models. Decoder is a classical LSTM recursive neural network, in which the input of the first step is the image features extracted by the convolutional neural network, followed by the input of the word vector expression of each word. For each word, it is first represented by the one-hot vector, and then embedded into the model by the word, which becomes the same dimension as image features. Its model diagram is shown in figure 1[6]:

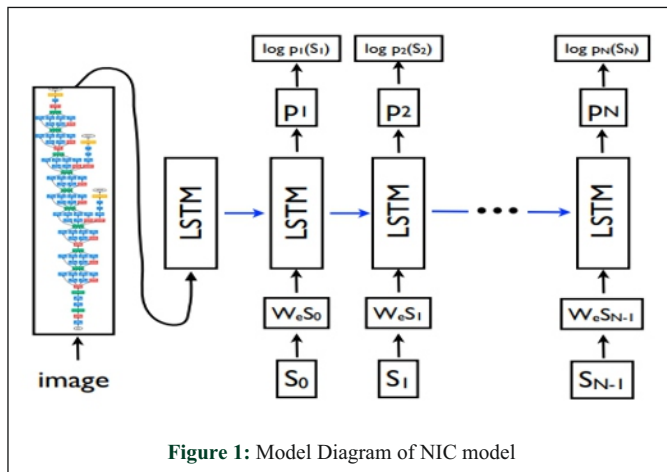


Figure 1: Model Diagram of NIC model

Fang H, Gupta S, Iandola F, *et al.* [7] to pass target detection + object recognition, the image of entities in the word are identified, the connection between the terms related to the real word is the core of constructing complete sentences, this paper use the Multiple Instance Learning (MIL) weak supervision methods to make a sentence with the process steps that entity identification word first, then generate sentences, the reforming sentence structure then Xu K, Ba J, Kiros R, *et al.* [8] proposed to add the attention mechanism to the cnn-lstm model. In this paper, two attention mechanisms were proposed. In encoder phase, and directly through the link layer before feature extraction is different, the authors use the lower-level convolution layer for image features, which retained the convolution layer image spatial information, and then combined with the attention mechanism, to the choice of dynamic image spatial characteristics used in the decoder stages in the decoder, increased input image context vector, the vector is the feature of the current moment images significantly.

Although there are many researches about image caption, there are few descriptions about aesthetics. Existing researches mainly describe the relationship between the content in images and the entities in images, and there are only two researches on aesthetics. Chang K Y, Lu K H, Chen C S[16] was the first study to transform the description of images from the perspective of aesthetics, and also to output aesthetic subtitles rather than scoring in the evaluation of aesthetic quality. Chang K Y *et al.* proposed two approaches to solve the problem of aesthetic evaluation. One is baseline-based approach: the aspect oriented (AO) approach; The other is an improved version of the first approach: the aspect fusion (AF) approach. AO method is cnn-lstm training for the data of one aspect of the image (composition, color, light and shadow), which results in the number of data training far less than the whole data set, resulting in too single sentence output. In this regard, AF proposed an improvement. It used the whole data set to learn the cnn-lstm model and generated hidden annotations h_l ($l=1...L$) further used to learn cnn-lstm, and added soft-attention to LSTM. Finally, AF method was implemented, which could use hidden annotations of different aspects to dynamically select appropriate aspects to output aesthetic subtitles. The AO and AF model diagrams are shown in figure 2[16]. In addition, the authors created a new aesthetic database named Photo processing Dataset (PCCD). The aesthetic annotation was comprehensive and included one overall and six aspects, but the amount of data was too small (4307) to meet the training requirements of the large-scale deep neural network.

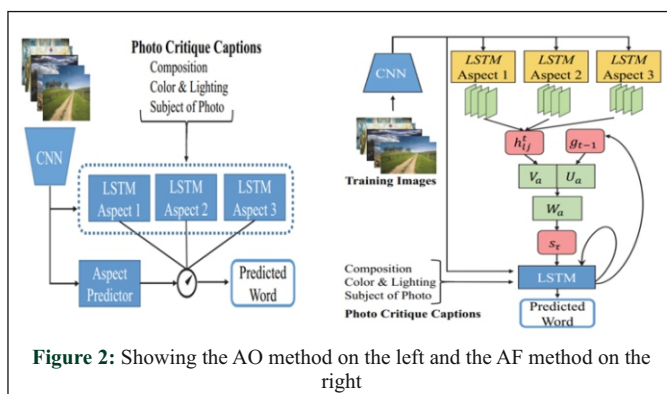


Figure 2: Showing the AO method on the left and the AF method on the right

In this paper, Wang W S[17] adds a Shared aesthetic semantic layer and a task-specific embedding layer in the cnn-lstm layer, which not only classifies the quality of pictures, but also outputs the aesthetic description of pictures, which is too simple. Meanwhile, AVA-Reviews data set was collected, but the aesthetic annotation was not comprehensive enough and the data volume was not rich enough. Because there is still a lot of room for development in the field of output aesthetic criticism.

3. CONCLUSIONS AND SUGGESTIONS:

Currently, most researches focus on the aesthetic classification of images, including the binary classification of images, high-quality or low-quality image classification; Aesthetic rating refers to the rating of different scores of pictures. And the study of aesthetic distribution. However, there is little research on aesthetic factors and aesthetic description, and the data set in this regard is small. Therefore, in the future research of aesthetic quality evaluation, the research on high-level figures of image aesthetic quality evaluation can be strengthened, so as to make new progress in the automatic evaluation of image aesthetic quality, which will also lay a foundation for machines to better imitate human behaviors.

Finally, as the currently popular method to study image aesthetic quality is deep neural network learning, and deep learning is black box learning, it is difficult to explain the specific connection with aesthetics. Therefore, in order to better study the evaluation of image aesthetic quality, we need to open the black box of deep learning, and correspond the aesthetic features with deep learning, so as to further explore the mystery of human aesthetic intelligence.

REFERENCES:

- Datta R, Joshi D, Li J, *et al.* Studying Aesthetics in Photographic Images Using a Computational Approach [C]// Computer Vision - ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part III. Springer-Verlag, 2006.
- 王伟凝, 刘剑聪, 徐向民, 姜怡孜, 王励, "基于构图规则的图像美学优化," 华南理工大学学报(自然科学版), pp. 51-58, 2015-05-15 2015., pp. 51-58, 2015-05-15 2015.
- Lu X, Lin Z, Jin H, *et al.* Rating Pictorial Aesthetics Using Deep Learning[J]. IEEE Transactions on Multimedia, 2014, 17(11):457-466.
- Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.
- Kang L, Ye P, Li Y, *et al.* Convolutional Neural Networks for No-Reference Image Quality Assessment[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014.
- Vinyals O, Toshev A, Bengio S, *et al.* Show and tell: A neural image caption generator[J]. Computer Science, 2015:3156-3164.
- Fang H, Gupta S, Iandola F, *et al.* From captions to visual concepts and back[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015:1473-1482.
- Xu K, Ba J, Kiros R, *et al.* Show, Attend and Tell: Neural Image Caption Generation with Visual Attention[J]. Computer Science, 2016:2048-2057.
- J. Johnson, A. Karpathy, and L. Fei-Fei. Densecap: Fully convolutional localization networks for dense captioning. In CVPR, 2016. 1, 2
- A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In CVPR, 2015.1, 2, 3
- J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, and A. Yuille. Deep captioning with multimodal recurrent neural networks (m-rnn). In ICLR, 2015. 1, 2
- L. Anne Hendricks, S. Venugopalan, M. Rohrbach, R. Mooney, K. Saenko, and T. Darrell. Deep compositional captioning: Describing novel object categories without paired training data. In CVPR, 2016. 1, 2
- J. Mao, X. Wei, Y. Yang, J. Wang, Z. Huang, and A. L. Yuille. Learning like a child: Fast novel visual concept learning from sentence descriptions of images. In ICCV, 2015. 1, 2
- J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In CVPR, 2015. 1, 2
- H. Fang, S. Gupta, F. Iandola, R. K. Srivastava, L. Deng, P. Dollar, J. Gao, X. He, M. Mitchell, J. C. Platt, *et al.* From captions to visual concepts and back. In CVPR, 2015. 1, 2
- Chang K Y, Lu K H, Chen C S. Aesthetic critiques generation for photos [C]// Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway NJ: IEEE, 2017: 3534-3543
- Wang W S, Yang S, Zhang W S, *et al.* Neural aesthetic image reviewer[J]. arXiv.org, 2018, arXiv:1802.10240.