

Aesthetic Assessment of Packaging Design Based on Con-Transformer

Wei Li, Hefei Normal University, China*

ABSTRACT

Different from the traditional natural images' aesthetic assessment task, the aesthetic assessment of packaging design should not only pay attention to artistic beauty, but also pay attention to functional beauty, that is, the attraction of the packaging design to consumers. In this paper, the authors propose a con-transformer packaging design aesthetic assessment method, which takes advantage of convolutional operations and self-attention mechanisms for enhanced representation learning, resulting in an effective aesthetic assessment of the packaging design images. Specifically, con-transformer integrates convolution network branch and transformer network branch to extract local representation features and global representation features of the packaging design images respectively. Finally, the fused representation features are used for aesthetic assessment. Experimental results show that the proposed method can not only effectively assess the aesthetic of packaging design images, but also be applied to the aesthetic assessment of natural images.

KEYWORDS

Attention Mechanism, CNN, Image Aesthetic Quality Assessment, Packaging Design Images, Transformer

1. INTRODUCTION

Aesthetics is people's innate ability. Studying the artificial intelligence technology to make computers perceive and discover "beauty" will help computers understand and learn the thinking process of professional photographers or professional designers, and provide professional aesthetic suggestions for people to take photos or make designs, which is a very challenging task. At present, with the rapid development and wide popularization of camera, video camera, smartphone and other photographing devices, more and more people like to capture and record the bits of life and upload them to the Internet. Especially since the beginning of 2020, due to the impact of COVID-19, face-to-face communication between people becomes decreased, which further increasing the sharing and dissemination of images, videos and other visual content on the Internet. For example, according to statistics, the total number of images and videos published on Instagram every day exceeds 100000000 (Lan 2021). Users on Flickr have shared more than 50 million photos (Smith 2021). In the face of such complex data, how to quickly and automatically screen out high-quality aesthetic images and rank the results with high visual attraction in the top of the search results has important practical significance. In addition, for album management software, automatically and quickly filtering out good-looking photos and deleting photos with low aesthetic quality can save a lot of time for manual screening of photos. In the photo editing software, it is very difficult for ordinary users to learn photography knowledge in limited time and energy (Zhu 2016). They do not know how to improve the aesthetic attributes of

DOI: 10.4018/IJeC.316873

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

the image and which aspect makes the image more attractive. In this case, the software that can give professional and clear image cutting suggestions becomes more and more important. Therefore, the research on the intelligent assessment technology of image aesthetic quality with artificial intelligence as the core not only has great economic benefits, but also can promote the development of artificial intelligence technology to simulate people's aesthetic and thinking process.

Packaging design refers to the deep integration of visual form and product function, highlighting product characteristics and accurately conveying product information through the use of reasonable organization and layout methods and novel graphic expression methods, so as to form a unique packaging (Zhu 2017). The packaging can protect the products and should be easy to carry, and it also has the sales function, that is, the visual aesthetics of product packaging attracts the attention of consumers as a silent salesperson, therefore, visual element design occupies an important position in the process of packaging design. There are many kinds of packaging design aesthetics, such as functional beauty and artistic beauty. Artistic beauty is a perfect visual image composed of modeling, color and composition, which is endowed with decorative characteristics. Decorative products can convey the cultural connotation and emotional experience given by designers to consumers, so as to make the products have functional beauty and pleasing purpose. Artistic beauty should not only satisfy the development of society, economy, culture, science and technology, but also satisfy the consumer psychological and spiritual needs of consumers. Besides, artistic beauty originates from natural beauty. Artistic beauty is a product with cultural connotation, aesthetic value and spiritual value created by people and combined with their own thinking about things that existing in nature. Artistic beauty brings certain positive changes to people's aesthetic feelings and aesthetic concepts, and plays a guiding, propaganda and educational role in people. Functional beauty mainly refers to the attraction to consumers, including visual impact and findability. Visual impact mainly measures the ability of packaging to "stand out from the crowd" on the shelf, while findability determines how consumers can focus on the target packaging in many products. In addition, the visual element design usually contains text information. The unique text is vivid and has strong recognition and aesthetic functions, which visually brings consumers visual enjoyment of beauty, so as to stimulate consumers' desire to buy. Therefore, researching the aesthetic assessment algorithm of packaging design based on computer vision technology can help enterprises quickly design product packaging images with both artistic beauty and functional beauty.

Image aesthetic quality assessment can automatically and quickly help people rate and sort massive images, help people select high-quality and beautiful images, and greatly reduce the time spent by people in managing personal albums, editing photos (Wang 2019), image retrieval (Baraldi 2016). Therefore, image aesthetic quality assessment technology has attracted the attention of a large number of researchers. Traditional image aesthetic quality assessment methods are usually developed based on manual design features (that is, the low-level visual features (Zhu 2022), high-level aesthetic features and composition aesthetic features are manually designed by experts) to evaluate image aesthetics (Datta 2010, Tong 2005, Nishiyama 2011, Dhar 2011). However, this type of methods requires researchers to have professional knowledge of photographic aesthetics such as composition and color, and the representation ability of manual design features is usually limited. In recent years, with the rise of deep learning technology, researchers have introduced convolution neural network (CNN) to solve related problems in image aesthetic assessment tasks. Due to the strong automatic learning ability of CNN, researchers can automatically extract high-level semantic features from a large number of image data without professional aesthetic knowledge, which makes CNN becoming the mainstream method to solve the problem of image aesthetic assessment (Lu 2015, Dong 2015, Kao 2016, Kong 2016, Jin 2017, Wang 2019). However, the above methods are designed based on the aesthetic quality assessment task of natural images, which does not consider the unique functional beauty attribute of packaging design images. In addition, CNN is designed based on convolution operation and is good at extracting local features of images. However, it is difficult to capture global

representations, e.g., long-distance relationships among visual elements, which are often critical for high-level computer visual tasks.

In view of the above problems, in this paper, we propose a two branches network architecture, termed Con-Transformer, with the aim of packaging design image aesthetic assessment. Specifically, we first extend the data label to include both artistic beauty label and functional beauty label. Then, in order to fuse the local information and global information of packaging design images, the proposed Con-Transformer develops a CNN branch and a Transformer branch which respectively follow the design of ResNet (Kaiming 2016) and ViT (Alexey 2020) networks. A large number of experiments demonstrate that our proposed method can not only effectively assess packaging design images, but also be applied to the task of natural images aesthetic assessment.

The following part of this paper is organized as follows: the related works are reviewed in Section 2; the architecture of the proposed Con-Transformer model is proposed in Section 3; the experiments are provided in Section 4; Section 5 is the conclusion.

2. RELATED WORKS AND ANALYSIS

With the continuous development of computer vision and digital image processing technology and the continuous progress of computer aesthetics, the research on image aesthetic quality evaluation has attracted more and more attention, and plays a more and more important role in image retrieval and image editing. The existing image aesthetic quality assessment methods can be grouped into traditional manual design features-based aesthetic assessment methods and popular deep learning-based aesthetic assessment methods.

Manual design features-based aesthetic assessment methods mainly assess image aesthetics by manually designing low-level visual features, high-level aesthetic features and composition aesthetic features. For example, Datta et al. (Datta 2010), as pioneers, first proposed the relationship between computer vision features and image aesthetics. Based on the basic aesthetic principles such as color matching and contrast of pictures, pictures are divided into two categories: high aesthetic feeling and low aesthetic feeling through support vector machine and regression. Tong et al. (Tong 2005) used the low-level feature to learn classification model to distinguish the photographic pictures of professional photographers and ordinary users. Nishiyama et al. (Nishiyama 2011) developed a method to evaluate the aesthetic quality of pictures based on color coordination. Sagnik et al. (Dhar 2011) selected high-quality images based on image layout, scene and natural lighting conditions. Literature (Obrador 2010) focused on the impact of image structural features on its aesthetic quality, and 55 features related to structural information are used, and the classification accuracy is very close to the accuracy of aesthetic benchmark. Literatures (Wang1 2015, Wang2 2015) used 41 global and local features such as composition, color and brightness to classify the aesthetic quality of images. Through experiments, these 41 features have good accuracy in assessing and classifying the aesthetic quality of images. However, these manually extracted features have limited representation ability, and cannot fully meet the needs of image aesthetic quality assessment.

Deep learning based aesthetic assessment methods introduce deep neural network into the task of image aesthetic assessment, and the assessment results are generally better than the traditional methods. For instance, by modifying the convolutional neural network, researchers (Kong 2016, Jin 2017, Wang 2019) can make it suitable for solving different image aesthetic assessment problems. Jin et al. (Jin 2017) proposed the deep convolution neural network RS-CJS. Wang et al. (Wang 2019) proposed an aesthetic image reviewer model NAIR based on CNN and recurrent neural network (RNN), which can not only predict the aesthetic score, but also generate semantic assessment. Different from the traditional aesthetic classification methods, which classify aesthetics into good and bad two classes, the NIMA (Talebi 2018) method proposed by Google predicts the probability distribution of human aesthetic assessment of an image through convolution neural network. The obtained probability distribution map can more accurately understand the centralized trend of user evaluation of an image,

and can more accurately guide how many people in the crowd feel good-looking. In addition, recently some scholars have combined the attention mechanism with the aesthetic assessment task to extract the global representation of the image (Zhang 2019). After analyzing the impact of image fine-grained information on aesthetic assessment task, Zhang et al. (Wu 2017) proposed a gated concave visual convolution neural network (GPF-CNN) for predicting the distribution of image aesthetic score. After that, Zhang et al. (Zhang 2021) considered that the aesthetic quality assessment task is a highly subjective and complex task. Aiming at the existing aesthetic assessment algorithms that rely heavily on the visual features extracted by convolution features without considering the relationship between visual elements in image composition, a multi-modal self-cooperative attention network was proposed.

3. AESTHETIC ASSESSMENT OF PACKAGING DESIGN BASED ON CON-TRANSFORMER

3.1 Packaging Design Image Data Acquisition

Since there is no public dataset suitable for this scenario, through web crawlers and other ways, we have collected a total of 1200 packaging design images, including 100 award-winning packaging design images over the years downloaded from professional websites, 1100 ordinary design images, and all images are resized of 400×200 pixels. And most of the collected packaging design images contain text information (some sample examples are shown in Figure 1). In order to label the data, participants were recruited in the school, taking into account the participants' age, gender, educational background and design experience. A total of 10 participants were recruited, including five women, five men. In addition, among these participants, 8 are ordinary users and 2 are experienced designers, aged from 20 to 35. Each participant scored independently, and the 5-point system is used to score the artistic beauty and functional beauty of packaging design images respectively (i.e., for artistic beauty: 5 points for very good-looking, 4 points for good-looking, 3 points for general, 2 points for not good-looking, and 1 point for very bad-looking; for functional beauty: 5 points for very attractive, 4 points for attractive, 3 points for general, 2 points for not attractive, and 1 point for very unattractive). In addition, for the award-winning images, the lowest score is assumed to be 3 points and the highest score is 5 points (some sample examples are shown in Table 1).

The data distribution of the collected packaging design images is shown in Figure 2. As can be seen from Figure 2, the data samples of each grade are uneven. For the artistic beauty score, most of the marked data fall between grade 3 and grade 3.6. For functional beauty, most of the marked data fall between grade 3 and grade 3.3. Moreover, there are few high-level and low-level data, the overall distribution is Gaussian. Finally, 90% of the sample data in each grade are randomly sampled as the training dataset, and the remaining 10% are used as the testing dataset. Finally, 1080 training image samples and 120 testing image samples are obtained.

Figure 1. Some sample examples of the collected packaging design datasets



Table 1. Data of participants' aesthetic assessment on packaging design, where AB denotes artistic beauty, FB denotes functional beauty. Besides, ID 5 is the award-winning image.

ID	Types	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
1	AB	2	5	4	4	2	2	5	3	3	3	3.4
	FB	4	2	3	3	1	4	5	3	2	3	3.0
2	AB	3	3	5	4	2	4	2	3	3	4	3.2
	FB	3	3	2	3	1	4	5	2	3	4	3.0
3	AB	5	3	1	3	3	4	4	3	5	4	3.5
	FB	4	4	3	2	5	1	3	2	5	3	3.2
4	AB	4	3	1	3	4	2	3	5	3	2	3.0
	FB	3	1	3	2	3	3	5	4	4	3	3.1
5	AB	4	5	4	4	4	3	5	4	4	5	4.2
	FB	3	5	2	3	3	4	4	2	3	3	3.2

3.2 Aesthetic Assessment of Packaging Design Based on Con-Transformer

The key to aesthetic assessment is to learn more representative features. Traditional convolutional neural network is good at extracting local features of images, however, it is difficult to capture global representations, e.g., long-distance relationships among visual elements, which are often critical for high-level computer visual tasks. Fortunately, the recently proposed visual Transformer (ViT) (Alexey 2020) method can effectively capture global representations because it makes full use of the attention mechanism. In view of this, we propose a new method called Con-Transformer to integrate convolution and attention mechanism. The flow chart of the proposed method is shown in Figure 3. In Figure 3, our Con-Transformer architectural includes two branches, the upper branch is global feature extraction networks, which using ViT as the basic network, and the feature vector

output by MLP layer is finally obtained. The lower branch is the local feature extraction network, which using ResNet-50 as the basic network, and the output feature vector is also obtained after passing through the FC layer. Finally, by concatenating the two feature vectors, we can get a new feature representation which fused the local and global features, then the new feature representation is used for prediction. Moreover, the two branches are pre-trained separately on ImageNet data to obtain the pre-training models.

Figure 2. Data distribution of packaging design images aesthetic assessment. (a) Artistic beauty score data distribution, (b) functional beauty score data distribution.

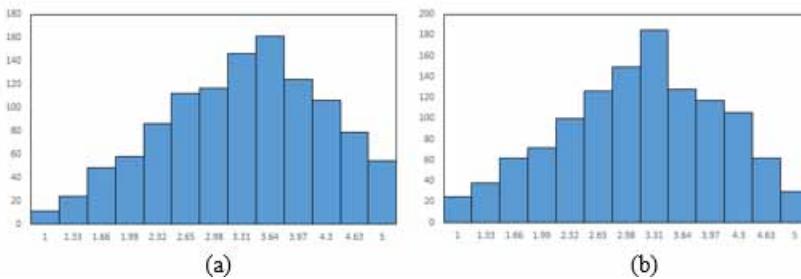
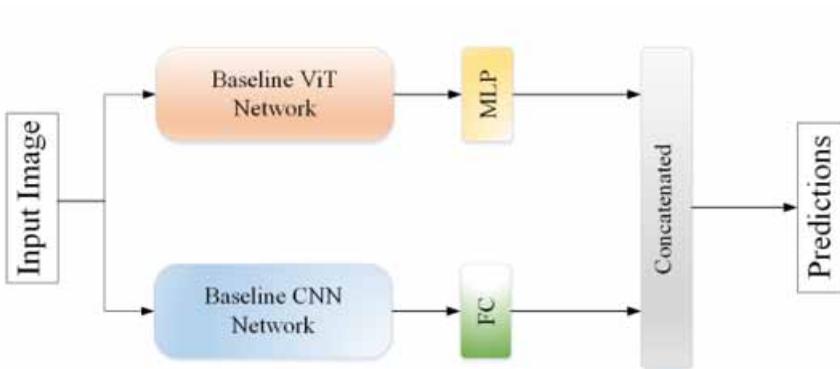


Figure 3. Flowchart of the proposed Con-Transformer method



3.3 Loss Function

The goal of our proposed method is to predict the distribution of ratings for a given image. In other words, we want to predict the probability that the input image belongs to each score. However, the output class of the traditional assessment task is ordered-classes. For example, for AVA data, the output is 1, 2, ..., 10, which satisfy $1 < 2 < \dots < 10$. But in our task, the output includes two parts: the output belonging to artistic beauty 1, 2, ..., 5 and the output belonging to functional beauty 1, 2, ..., 5. In order to unify with traditional tasks, we change the label of functional beauty from 1, 2, ..., 5 to 6, 7, ..., 10. Such that, the ground truth distribution of our task of human ratings of a given image can be expressed as an empirical probability mass function $p = [p_{s_1}, \dots, p_{s_{10}}]$ with $s_1 \leq s_i \leq s_{10}$, where s_i denotes the i^{th} score bucket.

In addition, the traditional aesthetic assessment task requires that $\sum_{i=1}^{10} p_{s_i} = 1$, where p_{s_i} represents the probability of a quality score falling in the i^{th} bucket. Yet, in our task we modify the rule to the sum of the first five probabilities is 1, and the sum of the last five probabilities is 1, that is, artistic beauty is required to be a group and functional beauty is required to be a group. In other words, we can get the probability distributions of artistic beauty and functional beauty at the same time.

In our collected dataset, each example consists of an image and its round truth ratings p . Our objective is to find the probability mass function \hat{p} that is an accurate estimate of p . Next, our training loss function is discussed.

NIMA has proved that for the ordered-classes task like image aesthetic assessment, the classification loss can outperform regression loss. Therefore, drawing on NIMA, we also use the EMD-based losses as our loss function. Specifically, for image quality ratings task, our classes are inherently ordered as $s_1 \leq s_i \leq s_{10}$, and based on this denotes, we can define the r -norm distance between classes as $s_i - s_{j_r}$, where $1 \leq i, j \leq 10$. EMD is defined as the minimum cost to move the mass of one distribution to another. For example, given the ground truth and estimated probability mass functions p and \hat{p} , the EMD can be formulated as:

$$EMD(p, \hat{p}) = \left(\frac{1}{N} \sum_{k=1}^{10} \sum_{i=1}^k p_{s_i} - \sum_{i=1}^k \hat{p}_{s_i}^r \right)^{1/r} \quad (1)$$

where r is set as 2 in our experiments.

4. EXPERIMENTAL RESULTS AND ANALYSIS

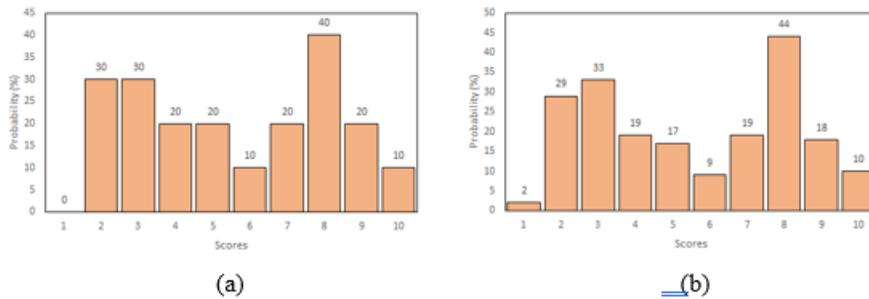
The effectiveness of our proposed method is verified on our collected packaging design dataset and the AVA dataset. For AVA dataset, we use 80% as training data and use the remaining 20% as testing data. For our proposed dataset, 90% are used as the training data and 10% are used as the testing data. When using AVA dataset, the weights of our baselines ResNet-50 and ViT are initialized by respectively training on ImageNet dataset, and the last FC layer and MLP layer are randomly initialized. When using our proposed dataset, the best weights training on AVA are used as our initialized weights.

Linear correlation coefficient (LCC), Spearman's rank correlation coefficient (SRCC) and the EMD value are used as our evaluation criteria. In order to output the rating probabilities of artistic beauty and functional beauty at the same time, the neurons in the last full connection layer are grouped, that is, the first five neurons form a group, and the last five neurons form another group, which are activated by SoftMax function respectively.

4.1 Performance on Our Proposed Dataset

First, in order to visualize the difference between the outputs of our network and the ground truth rating, the comparison results on ID1 sample are listed in Figure 4. As can be observed from Figure 4, for the artistic beauty of ID1 sample, 30% of the people think that the aesthetic grade of this packaging design image is 2 or 3, and 20% think that the grade is 4 or 5, while our Con-Transformer network thinks that the probability of grade 2 is 29%, and the probability of grade 3 is 33%. For functional beauty, 40% of people think its grade is 3 (i.e., $3 = 8 - 5$), and the network thinks that the probability of grade 3 is 44%. Therefore, it can be considered that the network's prediction results are basically consistent with people's scoring distribution.

Figure 4. Aesthetic assessment distribution on ID1 sample, where 6-10 scoring represents the 1-5 scoring probability of functional beauty. (a) Ground truth assessment distribution, (b) network's prediction assessment distribution.



Then, the comparison results on our proposed dataset are shown in Table 2. From Table 2, we can observe that in all cases, our proposed method achieves the best results, which means that our proposed method is effective. Specifically, the linear correlation between ground truth and results of our Con-Transformer is 0.61, which means that our predict results have strong correlation with the ground truth. While the linear correlation between ground truth and results of NIMA(Inception-v2), ResNet-50, ViT are 0.59, 0.54, 0.58, respectively. This means that their results have moderate correlation with the ground truth. In addition, the comparison between Con-Transformer and the two backbones, ResNet-50 and ViT, demonstrate that our fusion network can integrate the discrimination ability from local features and global features.

Table 2. Performance of the proposed method in predicting packaging design quality ratings compared to the state-of-the-art

Methods	LCC (mean)	SRCC (mean)	LCC (std.dev)	SRCC (std.dev)	EMD
NIMA(Inception-v2)	0.59	0.57	0.19	0.17	0.05
ResNet-50	0.54	0.54	0.14	0.16	0.07
ViT	0.58	0.55	0.17	0.18	0.06
Con-Transformer	0.61	0.60	0.24	0.23	0.04

4.2 Performance on AVA Dataset

We also carried out experiments on the natural image aesthetic assessment dataset AVA to verify the effect of the method proposed in this paper. Referring to the setting in NIMA, the specific experimental results are shown in Table 3, where the Reg+Rank+Att+Cont method is proposed in literature (Kong 2016), A-Lamp CNN method is proposed in literature (Ma 2017).

Table 3. Performance of the proposed method in predicting AVA quality ratings compared to the state-of-the-art

Methods	Accuracy (2 classes)	LCC (mean)	SRCC (mean)	LCC (std.dev)	SRCC (std. dev)	EMD
Reg+Rank+Att+Cont	77.33%	--	0.558	--	--	--
Mai et al. 2016	77.10%	--	--	--	--	--
A-Lamp CNN	81.70%	--	--	--	--	--
NIMA(Inception-v2)	81.51%	0.64	0.61	0.23	0.22	0.05
ResNet-50	78.25%	0.60	0.58	0.21	0.22	0.06
ViT	80.31%	0.61	0.59	0.22	0.23	0.06
Con-Transformer	82.35%	0.67	0.63	0.23	0.26	0.05

From Table 3, it can be seen that our Con-Transformer is superior to NIMA method in almost all attributes. Specifically, in two-class aesthetic categorization task, our Con-Transformer shows the highest accuracy, while the results from A-Lamp CNN and NIMA(Inception-v2) show the comparable accuracies. In multi-classes aesthetic categorization task, Con-Transformer achieves the best results on LCC (mean), SRCC (mean) and SRCC (std.dev) indicators. This means that our proposed method can not only effectively deal with the packaging design aesthetic assessment task, but also be effectively applied to the natural image aesthetic assessment task. In addition, we also compare our method with the two baselines: ResNet-50 and ViT. And the experimental results again show that our proposed Con-Transformer can inherit the advantages of ResNet-50 and ViT.

5. CONCLUSION

In this paper, aiming at the intelligent packaging design aesthetic assessment task, we first construct a new dataset that containing various packaging design images and corresponding scores. Then, a new local feature and global feature fusion network, i.e., Con-Transformer, is proposed, which can not only be used on the packaging design image aesthetic assessment task, but also be used on the natural image aesthetic assessment task. The experimental results on both packaging design aesthetic assessment dataset and natural image aesthetic assessment dataset show the superiority to some state-of-the-art methods. Future research work includes continuing to expand the packaging design image aesthetic assessment dataset and exploring more intelligent aesthetic assessment methods.

ACKNOWLEDGMENT

This work is supported by the Anhui Social Science Innovation and Development Research Project (2021CX136) and the 2022 Anhui Provincial Scientific Research Preparation Plan Project (2022AH052089).

REFERENCES

- Baraldi, L. (2016). Scene-driven Retrieval in Edited Videos using Aesthetic and Semantic Deep Features. *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, 23-29. doi:10.1145/2911996.2912012
- Craig, S. (2021). *Fun Easter Facts and Statistics*. <https://expandedramblings.com/index.php/easter-statistics/>
- Datta, R., & Wang, J. Z. (2010). ACQUINE: Aesthetic Quality Inference Engine: Real-time Automatic Rating of Photo Aesthetics. *Proceedings of the 11th ACM SIGMM International Conference on Multimedia Information Retrieval*, 421-424. doi:10.1145/1743384.1743457
- Dhar, S., Ordonez, V., & Berg, T. L. (2011). High Level Describable Attributes for Predicting Aesthetics and Interestingness. *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition*, 1657-1664. doi:10.1109/CVPR.2011.5995467
- Dong, Z., & Tian, X. (2015). Multi-level Photo Quality Assessment with Multi-View Features. *Neurocomputing*, 168(3), 308-319. doi:10.1016/j.neucom.2015.05.095
- Dosovitskiy, A., & Beyer, L. (2020). *An image is worth 16x16 words: Transformers for image recognition at scale*. arXiv preprint arXiv:2010.11929.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE CVPR*, 770-778. doi:10.1109/CVPR.2016.90
- Jin, X., Wu, L., & Li, X. (2017). Predicting Aesthetic Score Distribution through Cumulative Jensen-Shannon Divergence. *Menlo Park, 2017(3)*, 77-84.
- Kao, Y., Huang, K., & Maybank, S. (2016). Hierarchical Aesthetic Quality Assessment Using Deep Convolutional Neural Networks. *Signal Processing Image Communication*, 47(4), 500-510. doi:10.1016/j.image.2016.05.004
- Kong, S., Shen, X., & Lin, Z. (2016). Photo Aesthetics Ranking Network with Attributes and Content Adaptation. *Proceedings of 14th European Conference on Computer Vision*. doi:10.1007/978-3-319-46448-0_40
- Kong, S., Shen, X., & Lin, Z. (2016). Photo aesthetics ranking network with attributes and content adaptation. In *European Conference on Computer Vision*. Springer. doi:10.1007/978-3-319-46448-0_40
- LanM. (2021). <https://new.qq.com/omn/20210128/20210128A05D9IOO.html>
- Lu, X., Lin, Z., & Shen, X. (2015). Deep Multi-patch Aggregation Net Work for Image Style, Aesthetics, and Quality Estimation. *Proceedings of 2015 IEEE International Conference on Computer Vision*.
- Ma, S., Liu, J., & Chen, C. (2017). A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Mai, L., Jin, H., & Liu, F. (2016). Composition-preserving deep photo aesthetics assessment. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 497-506.
- Nishiyama, M., Okabe, T., Sato, I., & Sato, Y. (2011). Aesthetic Quality Classification of Photographs Based on Color Harmony. *Proceedings of Computer Vision & Pattern Recognition, 2011*, 33-40. doi:10.1109/CVPR.2011.5995539
- Obrador, P., Schmidt, C.-H., & Oliver, R. N. (2010). The role of image composition in image aesthetics. *IEEE International Conference on Image Processing*, 3185-3188. doi:10.1109/ICIP.2010.5654231
- Talebi, H., & Milanfar, P. (2018). NIMA: Neural Image Assessment. *IEEE Transactions on Image Processing*, 27(8), 3998-4011. doi:10.1109/TIP.2018.2831899 PMID:29994025
- Tong, H., Li, M., & Zhang, H. J. (2005). Classification of Digital Photos Taken by Photographers or Home Users. *Proceedings of Pacific-Rim Conference on Multimedia*.
- Wang, W., Shen, J., & Ling, H. (2019). A Deep Network Solution for Attention and Aesthetics Aware Photo Cropping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7), 1531-1544. doi:10.1109/TPAMI.2018.2840724 PMID:29993710

- Wang, W. S., Yang, S., Zhang, W. S., & Zhang, J. (2019). Neural Aesthetic Image Reviewer. *IET Computer Vision*, 13(8), 749–758. doi:10.1049/iet-cvi.2019.0361
- Wang, C., Pu, Y., & Xu, D. (2015a). Evaluating aesthetics quality in scenery photo. *Proceeding of National Conference on Multimedia Technology*, 141-149.
- Wang, C., Pu, Y., & Xu, D. (2015b). Evaluating aesthetics qualities in portrait images. *Journal of Software*, 26(S2), 20-28.
- Wu, J., Liu, Y., Shi, G., & Lin, W. (2017). Saliency change based reduced reference image quality assessment. *2017 IEEE Visual Communications and Image Processing (VCIP)*, 1-4.
- Zhang, X., Gao, X., He, L., & Lu, W. (2021). MSCAN: Multimodal Self-and-Collaborative Attention Network for image aesthetic prediction tasks. *Neurocomputing*, 2021(430), 14–23. doi:10.1016/j.neucom.2020.10.046
- Zhang, X., Gao, X., Lu, W., & He, L. (2019). A Gated Peripheral-Foveal Convolutional Neural Network for Unified Image Aesthetic Prediction. *IEEE Transactions on Multimedia*, 21(11), 2815–2826. doi:10.1109/TMM.2019.2911428
- Zhu, F., Gao, J., Xu, C., Yang, J., & Tao, D. (2017). On selecting effective patterns for fast support vector regression training. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8), 3610–3622. PMID:28841559
- Zhu, F., Gao, J., Yang, J., & Ye, N. (2022). Neighborhood linear discriminant analysis. *Pattern Recognition*, 123, 108422. doi:10.1016/j.patcog.2021.108422
- Zhu, F., Yang, J., Gao, J., & Xu, C. (2016). Extended nearest neighbor chain induced instance-weights for SVMs. *Pattern Recognition*, 60, 863–874. doi:10.1016/j.patcog.2016.07.012