TSC-Net: Theme-Style-Color Guided Artistic Image Aesthetics Assessment Network

Yin Wang, Wenjing Cao, Nan Sheng, Huiying Shi, Congwei Guo and Yongzhen Ke

September 13, 2023

TSC-Net: Theme-Style-Color guided Artistic Image Aesthetics Assessment Network

Yin Wang[1], Wenjing Cao[1], Nan Sheng[1], Huiying Shi[1], Congwei Guo[1] and Yongzhen Ke[1,2,*]

[1] School of Computer Science and Technology,Tiangong University, No. 399 Binshui Road, Tianjin 300387, China
[2] Tianjin Key Laboratory of Autonomous Intelligence Technology and Systems, Tianjin,300387, China
keyongzhen@tiangong.edu.cn

Abstract. Image aesthetic assessment is a hot issue in current research, but less research has been done in the art image aesthetic assessment field, mainly due to the lack of large-scale artwork datasets. The recently proposed BAID dataset fills this gap and allows us to delve into the aesthetic assessment methods of artworks, and this research will contribute to the study of artworks and can also be applied to real-life scenarios, such as art exams, to assist in judging. In this paper, we propose a new method, TSC-Net (Theme-Style-Color guided Artistic Image Aesthetics Assessment Network), which extracts image theme information, image style information, and color information and fuses general aesthetic information to assess art images. Experiments show that our proposed method outperforms existing methods using the BAID dataset.

Keywords: Artistic Image Aesthetics Assessment, BAID, Theme, Style, Color.

1 Introduction

Image aesthetic assessment aims to automatically determine aesthetic quality qualitatively or quantitatively and can be widely used in many downstream applications such as assisted photo editing, web-scale image retrieval, and smart album management. Image aesthetic assessment is a particular challenge because of the highly subjective and complex nature of human aesthetic preferences. In recent years assisted scoring of artworks and AI painting have become hot research areas, and the study of computational aesthetics of art has important applied and scientific values. BAID [1] compensates for the lack of a dataset of art images, which collects 60,408 artworks with different styles, themes, and contents and contributes to the automated analysis of large-scale art images to provide auxiliary information for art appreciation.

In fact, art images have diverse themes, and most of the aesthetic scores given by the human subconscious are related to themes, which are different for different art images, but existing methods do not take this into account or utilize this point; moreover, art images are extremely sensitive to colors, and in the art world, color is a unique artistic language, as artists will express the style and ideas of their work through color. We add an information-entropy method to perceive the color of art images in our model. Art images are significantly different from common images in terms of visual features such as color, texture, and composition; compared with real photographs, art images also have extremely strong artistic features and unique artistic styles, which include pastel, watercolor, ink, and others. Observing the style of an artwork is crucial in assessing its aesthetic value. In the aesthetic assessment of art images, our identification and judgment of artistic style will effectively improve the aesthetic assessment of art images.

In order to solve the above problems, this paper takes art images as the research object. Combined with the unique attributes of art images, we propose an art image assessment network TSC-Net based on theme, style, and color. It can effectively utilize the theme, style, color, and general aesthetic features of a given artwork. Our model consists of four parts: 1) Theme feature extraction sub-network: theme is an extremely important factor in appreciating and assessing art images. Art images with different themes need to extract different theme-related features. We propose a theme feature extraction sub-network incorporating theme information into the model to deal with the distraction problem effectively. 2) Style feature extraction sub-network: inspired by the study [1], observing the style of an artwork is crucial in assessing its aesthetic value, and different styles require the extraction of different style-related aesthetic features. We propose a style-specific aesthetic branch that incorporates the style information into the aesthetic features and extracts the style-specific aesthetic features via an adaptive instance normalization module. 3) Color feature extraction sub-network: art images are extremely sensitive to color. We incorporate an information-entropy guided color distributed feature extraction module to extract color features in art images. 4) General aesthetic feature extraction sub-network: we train the general aesthetic feature extraction sub-network to extract aesthetic perceptual features, which allows the model to learn the aesthetic quality of different artworks better.

The main contributions of our work:

1. exploring the problem of aesthetic assessment of art images from the dimension of theme and proposing a theme-specific art image assessment network that incorporates theme features into the aesthetic model.

2. based on the unique attributes of art images: color, and style, we design the style feature extraction module and information-entropy guided color extraction module in the model, which enables the aesthetic model to extract more fine-grained art features.

2 Related Work

Generic Image Aesthetic Assessment (IAA). Since 2014, research on aesthetic classification and scoring of images has fully entered the deep learning era. Lu *et al*. [2] proposed RAPID for binary aesthetic classification using an architecture similar to AlexNet [3]. In 2017, NIMA [4], based on MobileNet [5], VGG [6] and Inception [7], performed image preprocessing for better accuracy. NIMA also scores the image for errors and other problems and optimizes the image by selecting the best parameters based on the given scores, which outperforms the existing methods. In 2016, Kong *et* *al*. [8] created an AADB dataset with multiple aesthetic labels where each image has an aesthetic score and an attribute score. Gao *et al*. [9] converted a classifier into an SVM and used ResNet [10] to categorize the attributes and extract the image feature layer. The model achieved on the AVA dataset [11] with advanced performance.

Multiple-theme Image Aesthetic Assessment. Generalized aesthetic assessment studies have also pointed out that the subject matter or style of an image can also have a direct impact on aesthetic scores. However, early studies did not explicitly propose a solution for the impact of subject matter variation on aesthetic assessment models. Earlier researchers assisted the aesthetic assessment task through image classification and scene recognition methods, Cui *et al*. [12] proposed an improvement based on VGG-16 to develop a hybrid fully convolutional network that utilizes semantic cues of objects and image scenes to predict their perceived aesthetic quality; in the most recent study, He *et al*. [13] pointed out that the existing labels of the IAA dataset generally do not consider that different themes have different scoring criteria, created a theme-oriented aesthetic dataset TAD66K, and established a baseline model TANet, which can effectively extract theme information and adaptively build perceptual rules to assess images with different themes. The article demonstrates that TANet achieves state-of-the-art performance through large-scale testing.

Artistic Image Aesthetic Assessment (AIAA). Computational aesthetic assessment of artworks has become popular research in recent years. Hosu *et al*. [14] proposed the first AIAA method, which efficiently supports full-resolution images as inputs and can be trained on variable input sizes. Yi *et al*. [1] proposed the first large-scale art image dataset BAID and designed the style-specific Network SAAN to assess art images. The AIAA method has yet to be fully investigated. Theme and style are the most important features in a collection of artworks, and different works have different themes and styles, so we combine stylistic features, thematic features, color features, and generalized aesthetic features to evaluate artworks.

3 Methodology

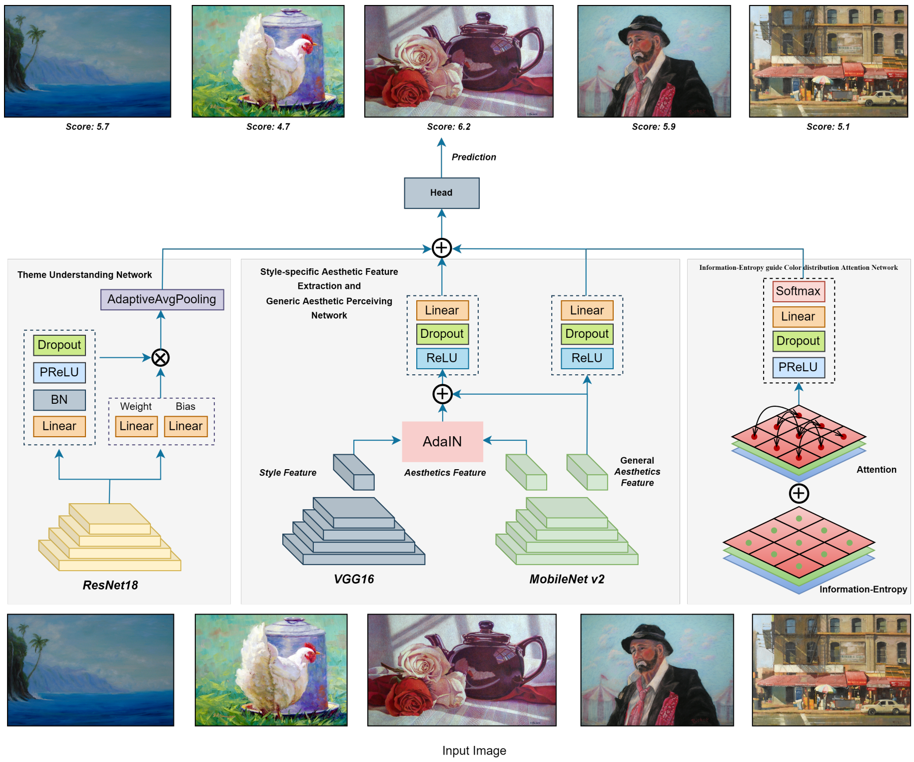


Fig. 1. The overall pipeline of our TSC-Net for artistic image aesthetic assessment.

3.1 Theme Understanding Network

The BAID dataset contains various themes, such as people, animals, and landscapes. Learning aesthetics directly from images will ignore the effect of theme changes on human visual perception. As shown in Fig. 1., we add a theme information extraction module to the model, which can effectively extract theme information and adaptively build perceptual rules to assess images with different themes.

We use ResNet-18 [10], trained on the scene dataset [24], as the backbone network. The scene dataset contains 10 million images, labeled with more than 400 unique topic semantic categories and environments, thus the backbone network can learn thematic information. In addition, in order to enable topic features to build perceptual rules adaptively, we refer to TANet [13] and process the output of the backbone into two streams, one to adaptively generate weights and biases via a parameter generator, and the other stream to pass through a feature preprocessor to reduce spatial redundancy in the potential representations. The final output is obtained by multiplying the two streams with a linear layer. This module's output features thus contain basic topic and perceptual rule information. For the artwork dataset BAID, most of the artworks have scene themes that are basically the same as the daily environments in the scene dataset [24], but some artworks may have scenes that are different from the normal daily scenes, such as some abstract works, which should be taken into account in future research. For this paper, the model can perceive the thematic scenes of most art pictures and enhance the results with other aesthetic features.

3.2 Style Feature Extraction

The art paintings in the BAID dataset contain many painting styles, such as Expressionism, Cubism, Fauvism, and Surrealism. From the perspective of human perception, different styles bring different feelings to people. For example, the lines of Expressionism are more twisted, and Cubism contains many geometric shapes, and these features determine the differences in the underlying characteristics of the images. Therefore, the goal of the style feature extraction branch is to extract the aesthetic characteristics of artworks that are compatible with their artistic style. Drawing on mainstream style migration methods [16-19], we use VGG-19 [6], which has been pre-trained by ImageNet [15], as the backbone network to extract the style feature f_{sty}, while obtaining the generic aesthetic feature (Section 3.4) f_{aes}, and then fusing the style feature into the generic aesthetic feature via the AdaIN [16] layer to obtain f_{style}. Given the content feature mapping f_{aes} and the style feature mapping f_{sty}, AdaIN encodes the content and style information in the feature space by aligning the mean and variance of the channel direction of f_{aes} to match the mean and variance of the channel direction of f_{sty}. The AdaIN layer encodes the content and style information in the feature space:

$$\begin{aligned} f_{style} &= AdaIN\left(f_{aes}, f_{sty}\right) \\ &= \sigma\left(f_{sty}\right) * \frac{f_{aes} - \mu\left(f_{aes}\right)}{\sigma\left(f_{aes}\right)} + \mu\left(f_{sty}\right) \end{aligned} \tag{1}$$

3.3 Information-Entropy guided Color distribution Attention Network

Fig. 2. Information-entropy of different kinds of images.

The color perception module extracts high-level color features from RGB space to perceive features such as color distribution and harmony of art images to improve the understanding of the artistic aesthetics of images. The color distribution is important information in aesthetics, and it is especially important in art images. From the point of view of human aesthetic perception, the single or strong color of an image and whether the hue is harmonious and rhythmic will visually leave different impressions, affecting the aesthetic rating of the whole image. In order to explore the characteristics of the color in the image, we introduce the information-entropy [20] into the image. The information-entropy can quantify the amount of information contained in a signal, the entropy is large to indicate that the system is more chaotic and has more information, and the entropy is small to contain less information. For the image, the entropy is large to indicate that the image is rich in color. As shown in Fig. 2, compared to the second image, the first image has only one object, so the information-entropy is low; while the second image contains more content, so the information-entropy is high; compared to the fourth image, the third image has monotonous colors, so the information-entropy is low; while the fourth image has rich colors and high information-entropy.

Specifically, computing the information-entropy of a region in an image can be expressed as：

$$H(C) = -\sum_{c\epsilon C} P(c) log_2 P(c) \tag{2}$$

where $P(C) = S_c/S$, S_c denotes the number of pixels of color c, and S denotes the total number of pixels in the image. We partition the input into non-overlapping patches, and for each patch, we compute the information-entropy values of the three channels and the average of the original pixel RGB values. For an image of size 224×224, the patch space consists of k*k (k=8) centroids, and the information-entropy value of each patch is summed with the average RGB value and input to the Attention module. f_{ab}, f_{uv} represent the two centroids, and the output of the color perception module f_{color}, can be described as:

$$f_{color} = ||_{l=1}^{N} \left(Softmax\left(\frac{(Q^l f_{ab})^T (K^l f_{uv})}{\sqrt{d}} \right) \right) \tag{3}$$

Where $||_{l=1}^{N}$ is the splicing of RGB channels, Q^l, K^l, and d are the query, key, and dimension generated by the standard self-attentive input, respectively. After extracting such relational features, they are sent to the final Softmax layer of the color perception module to get the result f_{color}.

3.4 Generic Aesthetic Perceiving Network

In addition to the theme feature understanding network, the style feature extraction network, and the color perception module, we used a generic aesthetic branch to extract aesthetic features common to artworks. Given an image, we use MobileNetV2 [21] as the backbone to extract the aesthetic feature map f_{aes} of shape $W \times H \times C$, where W and H denote width and height, and C denotes channel-wise. Finally, four features are fused, and ReLu, Dropout, Linear, and Sigmiod are added to the outputs of the four feature fusions. Get the predicted aesthetic score, and the whole process is described as:

$$p = head\left(f_{theme} \oplus f_{style} \oplus f_{color} \oplus f_{aes}\right) \tag{4}$$

4 Experiments and discussions

4.1 Datasets

We use the art image dataset BAID [1], which aims to fill a research gap in the aesthetic assessment of art images. The source of the BAID dataset is the Boldbrush website, which hosts a monthly artwork contest in which certified artists upload their artwork and receive public votes from online users. Voters can click on the detail page of the artwork and vote, with the number of votes reflecting the aesthetics and popularity of the artwork.

The selection of the BAID dataset has the following advantages: first, the contest does not restrict the subject matter, style, or medium of the works, so the dataset contains works in a wide range of artistic styles and content. Second, the voters consisted mainly of artists and art collectors, so the results of the dataset have a high degree of credibility and authority. The BAID dataset employs a sigmoid-like method to generate scores. The number of votes was converted into an image score, where the higher the number of votes, the higher the aesthetic value of the image. The scoring range was the common [0, 10] interval, where 0 means the worst and 10 means the best. The introduction of the BAID dataset provides insights for further research in the field of AIAA and improves the accuracy and universality of assessing the aesthetic quality of art images.

4.2 Experimental Setup

Evaluation Metric. We employ three commonly used metrics to evaluate IAA tasks. Accuracy is reported for binary aesthetic quality classification. For the regression task of aesthetic scores, we used Spearman's rank correlation coefficient (SRCC) and linear correlation coefficient (PCC). PCC and SRCC are computed between predicted scores and ground truth mean scores and standard deviation of scores. These metrics can verify the gap between the model prediction and ground truth in various aspects.

The PCC linear correlation coefficient describes the linear correlation between the subjective and objective assessments and is defined as follows：

$$PCC = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}} \tag{5}$$

Where N represents the number of distorted images, y、$\hat{y}_i$ denote the true aesthetic value and the predicted aesthetic score of the ith image, respectively, and $\bar{y}$、$\check{y}$ represent the aesthetic true mean value and the algorithm predicted mean value, respectively.

SRCC is used to measure the monotonicity and order correlation of the algorithm predictions and is the best nonlinear correlation metric in aesthetic assessment, calculated as：

$$SRCC = 1 - \frac{6\sum_{i=1}^{N}(v_i - p_i)^2}{N(N^2 - 1)} \tag{6}$$

Where v_i、 p_i denote the ranking positions of y_i、 $\hat{y}_i$ in the sequence of true and predicted values, respectively.

Loss Function. In BAID, since each image has only one score label s, s $\in$ [0,10]. Therefore, the mean square error (MSE) loss between the predicted and true aesthetic scores is used to train the whole network. The loss function formula is as follows:

$$\mathcal{L}(p, \hat{p}) = \frac{1}{n} \times \sum_{i=1}^{n} (p_i - \hat{p}_i)^2 \quad (7)$$

where n is the size of a batch during training, p is the true value, and p is the predicted value.

Implementation Details. We use ResNet-18 and VGG-19 pre-trained on ImageNet as the backbone of our feature extraction module. During training, we scaled the images to 224×224 size. We train the network on an NVIDIA RTX A5000 machine, set the batch_size to 64, and run it for 20 epochs. The initial learning rate is set to 0.0003. We implemented our proposed framework using the deep learning platform PyTorch.

Table 1. Performance of different methods on the BAID dataset.

Methods	**Backbone Network**	**Image size**	**Classification**	**Score Regression**	
			Accuracy↑	SRCC↑	PCC↑
MP$_{ada}$ [22]	ResNet-18	224×224	74.33%	0.437	0.425
MLSP [14]	InceptionResNet	Full resolution	74.92%	0.441	0.430
NIMA [4]	Inception-v2	299×299	71.01%	0.393	0.382
BIAA [23]	Inception-v3	299×299	71.61%	0.389	0.376
TANet [13]	MobileNet-V2 ResNet-18	224×224	75.45%	0.453	0.437
SAAN [1]	VGG-19 ResNet-50	224×224	76.80%	0.473	0.467
ours	MobileNet-V2 VGG-19	224×224	**76.97%**	**0.480**	**0.479**

4.3 Comparison with Generic Aesthetic Models

Table 1 shows the performance of MP$_{ada}$ [22], MLSP [14], NIMA [4], BIAA [23], TANet [13], SAAN [1] and our method on the BAID dataset. Compared to these methods, our model uses MobileNet-V2 and VGG-19 as the backbone network with input images scaled to 224 × 224 and achieves the best performance on all metrics. Fig. 3 shows some of the prediction results of our model, TANet, and NIMA on the BAID dataset.

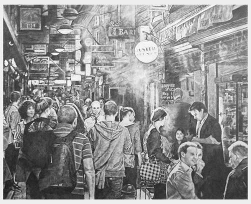
Test images			
NIMA	**4.60 (4.78)**	**5.72 (8.74)**	**4.57 (4.46)**
TANet	**4.61 (4.78)**	**4.54 (8.74)**	**4.64 (4.46)**
Ours	**4.56 (4.78)**	**6.57 (8.74)**	**4.53 (4.46)**

Fig. 3. Selected predictions on the BAID dataset. Black font indicates the predicted value, green font indicates the true value.

4.4 Ablation Study

Table 2 demonstrates the results of the ablation experiments: 1) We first verified the effectiveness of the theme-aware module. When the theme-aware module was removed, SRCC decreased by 7.3%, PCC decreased by 6.5%, and Accuracy decreased by 0.36%, and this discrepancy suggests that the theme-aware module has a greater impact on the evaluation of artworks. 2) We then verified the effectiveness of the style-specific feature extraction module. When this module was removed, SRCC decreased by 3.5%, PCC decreased by 6.7%, and Accuracy decreased by 1.52%. 3) When the Information-Entropy guided Color distribution Attention Network was removed, SRCC decreased by 3.1%, PCC by 1.5%, and Accuracy by 0.91%. 4) When the Generic Aesthetic Prediction Module was removed, SRCC decreased by 10.8%, PCC decreased by 18.2%, and Accuracy decreased by 0.35%，It can be seen that the general aesthetic perception network improves the model results a lot, indicating that ordinary photographic images and artworks have some commonalities, but the general aesthetic model also has potential limitations. For example, the general aesthetic perception network may not obtain the correct aesthetic information for artworks with abstract styles, so this should be considered in future research to further improve the assessment of artworks. This discrepancy suggests that the combination of the image's subject matter, style, color information, and generic aesthetic information can impact the most in assessing the artwork.

Table 2. Ablation studies of different components in our model.

Method	SRCC ↑	PCC ↑	Accuracy ↑
w/o Theme Understanding Network	0.445	0.448	76.61%
w/o Style-specific Aesthetic Feature Extraction	0.463	0.447	75.45%
w/o Information-Entropy guided Color distribution Attention Network	0.465	0.472	76.06%
w/o Generic Aesthetic Perceiving Network	0.428	0.392	76.62%
Ours	**0.480**	**0.479**	**76.97%**

5 Conclusion

In this paper, we focus on the challenging problem of AIAA. Benefiting from the emergence of the BAID large-scale art image assessment dataset, we construct a model that fuses image subject-aware and stylistic feature-aware networks to achieve state-of-the-art performance on the BAID dataset. The emergence of the BAID has allowed researchers to explore the characteristics of art images in depth, which is valuable for practical applications such as fine art examination assistance in judging and generating art images. In the future, different network models can be developed for the characteristics of artworks.

References

1. Yi, R., Tian, H., Gu, Z., Lai, Y.-K., Rosin, P.: Towards Artistic Image Aesthetics Assessment: a Large-scale Dataset and a New Method. ArXiv. (2023).
2. Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J.Z.: RAPID: Rating Pictorial Aesthetics using Deep Learning. Proceedings of the 22nd ACM international conference on Multimedia. 457–466 (2014). https://doi.org/10.1145/2647868.2654927.
3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM. 60, 84–90 (2017). https://doi.org/10.1145/3065386.
4. Talebi, H., Milanfar, P.: NIMA: Neural Image Assessment. IEEE Trans. on Image Process. 27, 3998–4011 (2018). https://doi.org/10.1109/TIP.2018.2831899.
5. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv. (2017).
6. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR. (2014).
7. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. 2015 IEEE Conference

on Computer Vision and Pattern Recognition (CVPR). 1–9 (2015). https://doi.org/10.1109/CVPR.2015.7298594.
8. Kong, S., Shen, X., Lin, Z., Mech, R., Fowlkes, C.: Photo Aesthetics Ranking Network with Attributes and Content Adaptation. 9905, 662–679 (2016). https://doi.org/10.1007/978-3-319-46448-0_40.
9. Gao, F., Li, Z., Yu, J., Yu, J., Huang, Q., Tian, Q.: Style-adaptive photo aesthetic rating via convolutional neural networks and multi-task learning. Neurocomputing. 395, 247–254 (2020). https://doi.org/10.1016/j.neucom.2018.06.099.
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90.
11. Murray, N., Marchesotti, L., Perronnin, F.: AVA: A large-scale database for aesthetic visual analysis. 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2408–2415 (2012). https://doi.org/10.1109/CVPR.2012.6247954.
12. Cui, C., Liu, H., Lian, T., Nie, L., Zhu, L., Yin, Y.: Distribution-Oriented Aesthetics Assessment With Semantic-Aware Hybrid Network. IEEE Trans. Multimedia. 21, 1209–1220 (2019). https://doi.org/10.1109/TMM.2018.2875357.
13. He, S., Zhang, Y., Xie, R., Jiang, D., Ming, A.: Rethinking Image Aesthetics Assessment: Models, Datasets and Benchmarks. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. pp. 942–948. International Joint Conferences on Artificial Intelligence Organization, Vienna, Austria (2022). https://doi.org/10.24963/ijcai.2022/132.
14. Hosu, V., Goldlucke, B., Saupe, D.: Effective Aesthetics Prediction With Multi-Level Spatially Pooled Features. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 9367–9375 (2019). https://doi.org/10.1109/CVPR.2019.00960.
15. Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, Li Fei-Fei: ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848.
16. Huang, X., Belongie, S.: Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. 2017 IEEE International Conference on Computer Vision (ICCV). 1510–1519 (2017). https://doi.org/10.1109/ICCV.2017.167.
17. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.-H.: Universal Style Transfer via Feature Transforms. Presented at the NIPS May 23 (2017).
18. Liu, S., Lin, T., He, D., Li, F., Wang, M., Li, X., Sun, Z., Li, Q., Ding, E.: AdaAttN: Revisit Attention Mechanism in Arbitrary Neural Style Transfer. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 6629–6638 (2021). https://doi.org/10.1109/ICCV48922.2021.00658.
19. Park, D.Y., Lee, K.H.: Arbitrary Style Transfer With Style-Attentional Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 5873–5881 (2019). https://doi.org/10.1109/CVPR.2019.00603.
20. Lee, B., Seo, M.K., Kim, D., Shin, I., Schich, M., Jeong, H., Han, S.K.: Dissecting landscape art history with information theory. Proc. Natl. Acad. Sci. U.S.A. 117, 26580–26590 (2020). https://doi.org/10.1073/pnas.2011927117.
21. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 4510–4520 (2018). https://doi.org/10.1109/CVPR.2018.00474.
22. Sheng, K., Dong, W., Ma, C., Mei, X., Huang, F., Hu, B.-G.: Attention-based Multi-Patch Aggregation for Image Aesthetic Assessment. Proceedings of the 26th ACM international conference on Multimedia. 879–886 (2018). https://doi.org/10.1145/3240508.3240554.

23. Zhu, H., Li, L., Wu, J., Zhao, S., Ding, G., Shi, G.: Personalized Image Aesthetics Assessment via Meta-Learning With Bilevel Gradient Optimization. IEEE Trans. Cybern. 52, 1798–1811 (2022). https://doi.org/10.1109/TCYB.2020.2984670.
24. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 Million Image Database for Scene Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 40, 1452–1464 (2018). https://doi.org/10.1109/TPAMI.2017.2723009.