IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 22, 2020, accepted December 6, 2020, date of publication December 14, 2020,
date of current version December 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044573

Inkthetics: A Comprehensive Computational Model for Aesthetic Evaluation of Chinese Ink Paintings

JIAJING ZHANG[1], YONGWEI MIAO[1], (Member, IEEE), JUNSONG ZHANG[2], AND JINHUI YU[3]

[1]Department of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2]Mind, Art and Computation Group, Cognitive Science Department, Xiamen University, Xiamen 361005, China
[3]State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058, China

Corresponding author: Jiajing Zhang (zhangjj@zstu.edu.cn)

This work was supported in part by the Natural Science Foundation of Zhejiang Province under Grant LQ20F020022, in part by the Fundamental Research Funds for Zhejiang Sci-Tech Universities under Grant 18032115-Y, and in part by the National Natural Science Foundation of China under Grant 61772463, Grant 61972458, and Grant 61772440.

ABSTRACT Assessing the aesthetic appeal of artworks has become an active research direction recently. However, previous works mainly focus on photographs and oil paintings, there have been few attempts in predicting aesthetics of Chinese ink paintings, due to their significant differences in visual features, semantic features, and aesthetic principles. Aiming at this problem, we propose a comprehensive framework, named *Inkthetics*, to quantify aesthetics of Chinese ink paintings based on deep learning. Firstly, an aesthetic assessment dataset is built for Chinese ink painting images. Secondly, a deep multi-view parallel convolutional neural network (DMVCNN) is designed by extracting global attribute images and multi-patches as inputs to jointly learn aesthetic features. Finally, we build a comprehensive aesthetic evaluation model by fusing the deeply-learned features with handcrafted features that rely on art expert knowledge. Experimental results show that our proposed deep network significantly outperforms existing methods on the dataset, and our proposed model can predict human aesthetic judgment with Pearson highly significant correlation of 0.843, which indicates an improvement up to 5.7% than the DMVCNN model when the handcrafted features are fused with activation from DMVCNN. Our work not only provides a deep-learning-based reference framework for computational aesthetic evaluation of Chinese paintings, but also explores to what extent can handcrafted features aid learning-based features in predicting human aesthetic perceptions.

INDEX TERMS Computational aesthetics, Chinese ink paintings evaluation, deep convolutional neural networks, feature fusion, handcrafted aesthetic features.

I. INTRODUCTION

Chinese ink painting is a typical representative of traditional brush painting in Chinese origin, which creates a variation of tonality, shading, and dry-wet of the ink color, achieved by differential grinding of the ink stick in the water to vary the ink density, ink load, and pressure within a single brushstroke, as shown in Fig. 1. For centuries, this most prestigious form of Chinese art has the extremely high artistic effect and aesthetics value, therefore it occupies an important position in the world fine arts domain. Traditional aesthetic evaluation of Chinese ink paintings can only be done by experts qualitatively in words with a high degree of abstraction, thus evaluation results have a high degree of subjective uncertainty. Additionally, it is practically impossible to invite high-level experts to manually evaluate massive images stored on the internet. The efficient, automated, and quantitative evaluation capability of Chinese ink paintings would have great guiding significance to the teaching of Chinese ink paintings. Moreover, it has a profound impact on the applications of advanced search for Chinese paintings, identification of drawing styles, computer-aided creation of Chinese paintings, and promotional exhibition of digital Chinese painting art gallery.

Computational aesthetics aims to simulate the human visual system and perception to make applicable aesthetic

The associate editor coordinating the review of this manuscript and approving it for publication was G. R. Sinha.

FIGURE 1. Examples of typical Chinese ink paintings.

judgment on images automatically [1], [2]. which is valuable in managing large collection of artworks with respect to appreciating and judging their beauties for not only amateurish users but also professional artists, thus freeing them from the tedious painting management work. It has recently attracted a lot of interest and has become an active research direction in the computer vision field [3]–[5]. Previous works mostly focused on aesthetic evaluation of photographs [6]–[13] and western paintings [14]–[16], however, they are not fully applicable to quantitative aesthetic evaluation of Chinese ink paintings. The challenges include the following aspects:

- While there exist several aesthetic assessment benchmark datasets such as CUHKPQ [17], Aesthetic Visual Analysis [18], MART [15], and Jenaesthetics [19], which are aiming at photographs or western paintings, there is still no standard manual calibration scheme designed for the aesthetic evaluation of Chinese ink painting. We need to establish a professional dataset according to aesthetic criteria of Chinese ink paintings.
- In terms of aesthetic criteria, the Chinese ink painting stresses the spirit vividness, verve of brush and ink, bone with brush, color adhere to type, and position management [20]–[22]. These standards correspond to high-level aesthetic semantics with high abstraction, which are difficult to be quantified by traditional handcrafted features and may result in a semantic gap.
- The characteristics in ink color, brushstroke, and composition with white space distribution [23] differ significantly from those in photographs and western paintings. Previous frameworks mostly manually designed features based on expert knowledge, e.g. established photographic techniques, or western artistic rules, or automatically extracted deeply-learned features from convolutional neural networks (CNNs) that rely on photo scene contents or style attributes, which are not entirely appropriate to comprehensively depict aesthetic properties of Chinese ink paintings.

To tackle these challenges, in this study we propose a comprehensive framework, named as *Inkthetics*, to quantify aesthetics of Chinese ink paintings, based on a fusion of deep learning abstract features and art professional appreciation knowledge. More specifically, we make the following three contributions:

- We collect the artworks from representative professional Chinese artists in modern times and amateur students, and build a benchmark dataset for aesthetic evaluation of Chinese ink paintings with subjective ratings from art professionals, thus providing training data for aesthetic features learning.
- We transform the basic VGG16 network by designing a multi-subject layer based on the aesthetic characteristics of Chinese ink paintings, and design a deep multi-view parallel deep convolutional neural network by extracting global attribute images and adaptive local patches as heterogeneous inputs to jointly learn deep aesthetic features, which significantly outperforms existing methods on the Chinese ink painting dataset.
- We build a comprehensive aesthetic evaluation model by fusing the deeply-learned features with handcrafted features that rely on art expert knowledge, which can better capture perceptual aesthetic information of ink color, brushstroke, and layout in Chinese ink paintings. Our proposed model can predict human aesthetic judgment with Pearson highly significant correlation of 0.843.

The rest of this paper is organized as follows. Section II summarizes previous works on computational aesthetics. Section III introduces the framework of our work. We then introduce the aesthetic assessment benchmark dataset of Chinese ink paintings in Section IV. A detailed description of our method in multi-view deep aesthetic feature learning is presented in Section V. Section VI provides the fusion of deeply-learned and handcrafted aesthetic features. The experiments are presented in Section VII. Finally, we conclude and discuss our future work in Section VIII.

II. RELATED WORK

Computational aesthetics, as a sub-discipline of computer vision, has become an active research field in recent years. In this section, we introduce some research work on computational aesthetic evaluation of images, the features that the algorithms compute can be described to be either handcrafted or deep learning features.

A. HAND-CRAFTED FEATURES

By using simplicity, realism, and other basic photography techniques as guidelines, early works extracted spatial distribution of edges, color, blur, brightness to differentiate between professional photographs versus snapshots. Dhar *et al.* [24] used low-level features to estimate high-level human-describable attributes, such as composition, illumination, and scene contents, which were used to predict aesthetic quality and interesting of photos. Wang *et al.* [25] evaluated the beauty of portrait photos by extracting visual structure, dark channel, and facial region descriptors. A preference-aware view recommendation system for scenic photos was proposed by Su *et al.* [26] to construct

bag-of-aesthetics-preserving features. Obrador *et al.* [27] divided photos into ''animal'', ''plant'', ''static'', ''architecture'', ''landscape'', ''human'', and ''night'' based on scene contents, then regional features were extracted from subject and background to assess photo aesthetics in different categories. Encapsulated aesthetic signatures including sharpness, exposure, colorfulness, tone, clarity, and depth were computed from images in [28], which comprised calibrated ratings of meaningful attributes. However, most extracted features are related to basic photography techniques and camera parameters, which are not important aesthetic criteria in Chinese ink paintings. Besides, there exist significant differences in the expression of scene semantics. For example, the facial descriptors in portraiture, tone combination, indoor-outdoor, and sky-illumination attributes in the landscape do not existed in Chinese ink paintings. Thus the above methods are not entirely suitable for aesthetic evaluation of Chinese ink paintings.

In terms of western paintings, Li *et al.* [14] evaluated the aesthetic quality of oil paintings by detecting some characteristics related to artistic knowledge, such as color distribution, brightness, blur effect, and edge distribution globally, together with local features such as the shape of segments and contrast between different regions. Sartori *et al.* [15] analyzed the difference between positive and negative abstract paintings by extracting LAB based color and SIFT based texture visual words. Since the colorfulness, lightness/shadow in nature towards objects, and focus perspective are not emphasized in Chinese ink paintings, the hue and saturation-lightness models for oil paintings are also not suitable for Chinese ink paintings. For abstract paintings, they mainly use lines, shapes, and colors to express aesthetics in strong forms, which are significantly different from aesthetic criterion of Chinese ink paintings.

For aesthetic assessment of Chinese ink paintings, Zhang *et al.* [29] developed a linear aesthetic model with 7 low-level handcrafted features. Unfortunately, the variation effect of brush strokes, line rules, white space distribution, as important aesthetic factors in Chinese ink paintings, were ignored in this model. Besides, the dataset was built by collecting only 60 flower and bird paintings from Qi Baishi, with a single subject and fewer styles, which resulted in lacking generalization in the aesthetic model. Moreover, the low-level handcrafted features are difficult to accurately quantify the high-level semantics in ink painting.

B. DEEP LEARNING FEATURES

Due to the vagueness of certain photographic or psychologic rules, the handcrafted features are often difficult in approximating them computationally. Beginning with the strong performance of Krizhevsky *et al.* [30] in the image classification, deep convolutional neural networks [31] have been applied to aesthetic quality assessment of photographs, which can automatically learn effective aesthetic features without manual involvement and extensive expert knowledge. The RAPID model by Lu *et al.* [6] used an AlexNet-like architecture where the last fully-connected layer was set to output 2-dim probability for aesthetic binary classification, the best model was obtained by stacking a global warped image and a local random cropped patch as inputs to form a double-column CNN. Moreover, they further boosted the performance of the network by incorporating image style information using a style-column CNN. Wang *et al.* [7] presented a multi-scene deep learning model, which consisted of multi-group descriptors in the network elaborately so that it had a comprehensive learning capacity for photo aesthetic. Ma *et al.* [8] developed an A-Lamp CNN architecture, which extracted features from both fined grained details and holistic image layout simultaneously. Kao *et al.* [9] proposed three category-specific CNN architectures for aesthetic classification, one for object, one for scene, and one for texture. Talebi *et al.* [10] adopted the distance between the histogram of network and real output as the loss function for training, and proposed a NIMA model to predict the distribution of human opinion scores on image aesthetic. Li *et al.* [16] applied deep convolutional features in the classification and evaluation of sketch works, which were collected from the sketch teaching scenario. Zhang *et al.* [12] presented a double-subnet Gated Peripheral-Foveal Convolutional Neural Network, which simulated the peripheral vision to encode the holistic information, and the foveal vision was used to extract fine-grained features on the attended regions.

Although the performance of the above deep networks are significantly higher than those using handcrafted features, it is difficult to explain which essential aesthetic knowledge the networks are extracted. Michal *et al.* [11] investigated the possibility of improving image aesthetic inference of convolutional neural networks with hand-designed features that rely on domain expert feature knowledge in photography. However, most of the network structures above are designed and trained based on the style and content semantics of photo aesthetics datasets, the models have certain scene limitations. In addition, the inputs of the traditional deep neural networks are mostly randomly cropped according to the layout of photos, which cannot accurately quantify the aesthetic attributes of Chinese ink paintings. In this study, we will design a new network architecture with heterogeneous inputs according to the characteristics of Chinese ink paintings to retain both global and local details, and leverage expert knowledge in designing handcrafted features for predicting aesthetics.

III. OVERVIEW OF OUR COMPUTATIONAL EVALUATION FRAMEWORK

Figure 2 shows our comprehensive framework of computational aesthetic evaluation of Chinese ink paintings based on deep learning. We first collect 1200 Chinese ink paintings and build an aesthetic benchmark dataset with subjective ratings from art specialty subjects. Then we transform the basic VGG16 network by designing a multi-subject layer based on the aesthetic characteristics of Chinese ink paintings, and a deep multi-view parallel convolutional neural network is then designed by extracting global attribute images and adaptive local patches as heterogeneous inputs to jointly learn deep

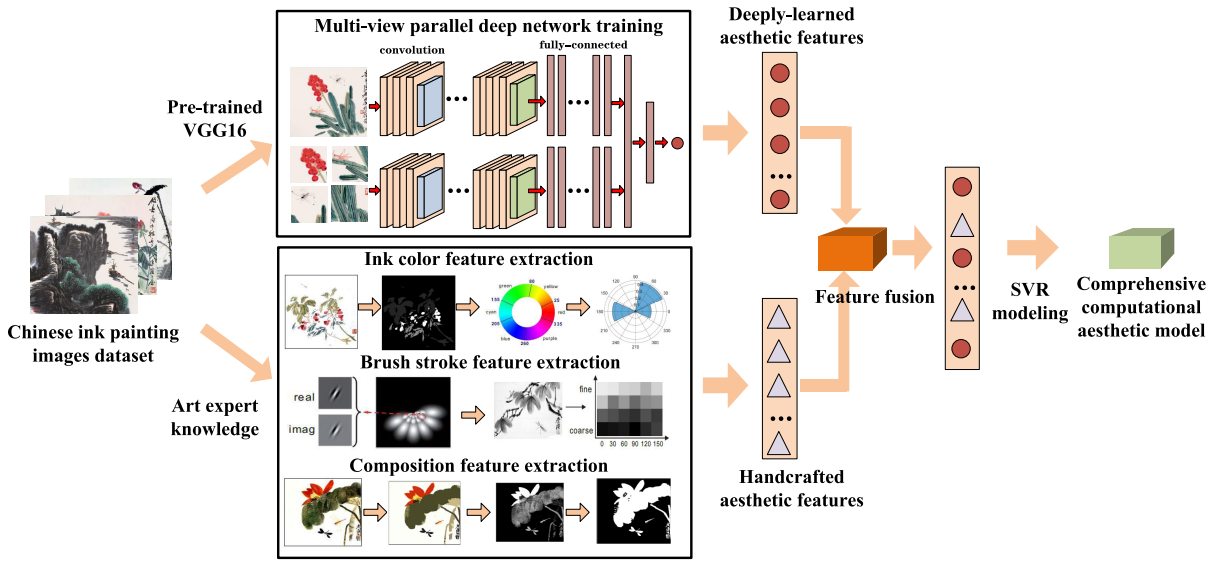


FIGURE 2. Overview of the framework of our computational aesthetic evaluation of Chinese ink paintings.

aesthetic features. Finally, we design a set of ink color, brush stroke, and composition-related feature variables from the point of art expert knowledge in Chinese painting, and build a comprehensive aesthetic evaluation model by support vector regression (SVR) training with a fusion of the deeply-learned features and handcrafted features in Chinese ink paintings.

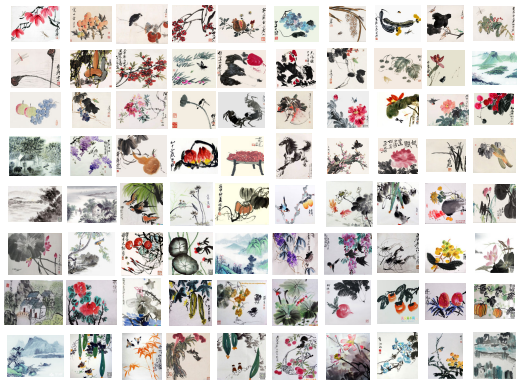


FIGURE 3. Samples of experimental image dataset.

IV. AESTHETIC ASSESSMENT BENCHMARK DATASET

Being treated as a deep learning problem, the dataset construction becomes the key precondition in aesthetic evaluation of Chinese ink paintings. We build a dataset by selecting 1200 Chinese ink painting images, which cover 6 themes including flower and bird, grass and insect, shrimp and crab, melon fruit, beast, and landscape, with 600 images created by 78 famous Chinese painting artists in modern times such as Qi Bashi, Wu Changshuo, Xu Zhu, Pan Tianshou, Zhang Xinjia, and Wang Xuetao, and 600 images students' artworks, as shown in Figure 3. Some of the images are downloaded in the digital online library of China academy of arts, which

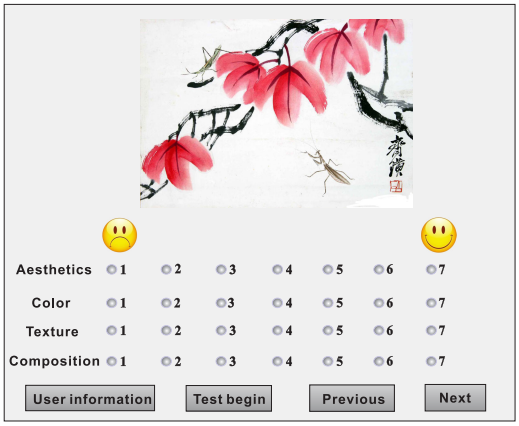


FIGURE 4. An example page of the survey.

contains more than 100,000 professional Chinese paintings, digital cultural heritage works in Taibei Palace Museum's digital culture database, and an online social network devoted to the creation and sharing of amateurs' artworks. Others are obtained by scanning the artwork collections of famous painters [32], [33].

In the phase of human evaluation, we recruit a total of 180 subjects (aged 18-65) including art experts, teachers in art specialty colleges, visitors of the museum, undergraduate and graduate students from the Faculties of Art. According to the educational background and artistic attainment, we assign a different weight for each type of subject so as to calculate the final weighted average scores, among which the weight of the expert is 3, the teacher is 2, the visitor and student are 1. According to the aesthetic criteria introduced in [22], we design an evaluation rating page, as shown in Figure 4, where the painting image is displayed in a window above. Each subject is required to give 4 ratings for evaluating the

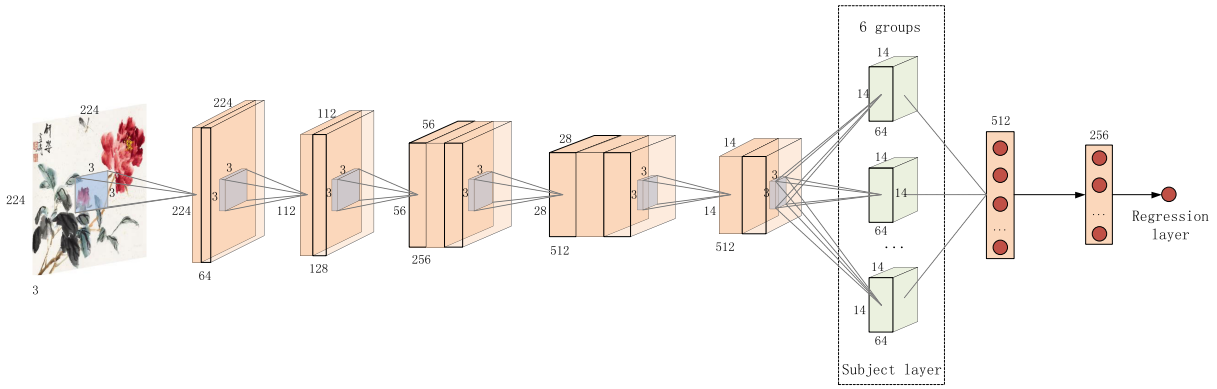


FIGURE 5. The basic architecture of our deep learning network.

following 4 aspects of each painting image: ''Aesthetics'' describes the whole feeling of current painting at the first sight, ''Color'' describes the blending naturalness of ink color and shading level changes, ''Texture'' describes the contrast of dry-wet ink and variation of brush ability, and ''Composition'' describes the feelings towards the spatial organization of objects in the Chinese ink painting. Here we adopt the most widely used and well-established seven-point Likert scale (good reliability and validity) for rating, ranging from 1 to 7, where 7 (rightmost) means the most positive score and 1 (leftmost) means the most negative score. The paintings were randomly divided into 6 groups of 200 paintings each, and only one group was presented to a single subject. Each artwork received 30 different human judgments, we calculated the weighted average scores as final scores for model training.

V. MULTI-VIEW DEEP AESTHETIC FEATURE LEARNING

CNN has an outstanding ability in extracting high-level semantic features, and its high-level output can be regarded as a high aesthetic attribute descriptor. Therefore, we transform the basic VGG16 network by designing a multi-subject layer based on the aesthetic characteristics of Chinese ink paintings, and design a deep multi-view parallel CNN to automatically extract deep aesthetic features, as detailed next.

A. BASIC NETWORK ARCHITECTURE

We propose to formulate the aesthetics rating task as a regression modeling problem of predicting continuous aesthetics rating scores. Denoting a Chinese ink painting by x_n, we aim to construct a regression mapping $f : R^2 \rightarrow R$ which estimates the aesthetics rating prediction as $\hat{y}_n = f(x_n)$. Given the user rating pairs of Chinese ink paintings $(x_n, y_n), n \in [1, N]$, where N is the size of training data, y_n is the human rating. Our method is to exploit a deep CNN to automatically learn the mapping function f, as shown in Figure 5. Denoting the parameters of the network by W, the overall optimization objective is to minimize the following euclidian regression loss function:

$$l(W) = \frac{1}{2N} \sum_{n=1}^{N} ||\hat{y}_n - y_n||_2^2 + \lambda ||W||_2^2 \quad (1)$$

where the second term is the weighted decay regularization that drives the weights closer to the origin, and λ is the trade-off parameter. To combat the over-fitting problem, we choose the VGG16 model pre-trained on the ImageNet dataset [30] as our base network for transfer learning [34].

The network contains 13 convolution layers and 3 fully-connected layers, we adjust the network structure by (1) inspired by the multi-scene layer in [7], we design a multi-subject convolutional layer at the original 13th layer. This layer consists of 6 parallel convolutional network groups (with $14 \times 14 \times 64$ convolutional kernels), which are used to learn image aesthetic descriptors discriminatingly according to different subjects in Chinese ink paintings. The descriptors separately correspond to 6 subjects which are namely ''flower and bird'', ''grass and insect'', ''shrimp and crab'', ''melon fruit'', ''beast'', and ''landscape''. Each of the 6 independent branches links one convolutional network group in the 12th layer to the 13th layer. The first fully connected layer has 512 neurons connected to the outputs of 6 convolutional groups via mean-pooling; (2) we modify the last classification layer into the regression layer as we have formulated before, by replacing the Softmax loss with the Euclidean loss. (3) We also reduce the number of neurons in the last two fully-connected layers from 4096 to 512 and 256, for reducing the number of parameters and further alleviate the risk of over-fitting. As a result, the network contains 13 convolutional layers, 4 max-pooling layers, 2 fully-connected layers, and a regression layer.

In the pre-training stage, we initialize the weights of the first 12 convolutional layers by the VGG16 model. Then for the 13th subject layer, we train each group independently by using painting images of one subject class. During the training, the weight of the corresponding subject's group and the fully connected layers are updated in the network. After all the groups are trained, 6 groups are parallel linked

to the previous layer with their weights initialized. And the initial weights of the fully-connected layers are set randomly. Subsequently, the overall basic network is further fine-tuned end-to-end.

B. DEEP MULTI-VIEW PARALLEL CNN

Aesthetic perception of Chinese ink painting relies on a combination of local and global visual cues. For example, the ''three farness rule'' composition is a global cue in landscape painting, while the distribution of white space and shading contrast are local visual characteristics. To support network training on heterogeneous inputs, we extend the single-column basic network by developing a deep multi-view parallel CNN structure (DMVCNN), as shown in Figure 6. Given an arbitrary sized image, one subnet extracts several global attribute images from different representations according to aesthetic principles of Chinese ink paintings, and multiple local patches are adaptively selected by the Patch Selection module, which is then fed into a multi-patch subnet. A statistic aggregation structure [35] is then followed to effectively combine the extracted features from each view of these multiple channels. Finally, we concatenate the global and local features from the two subnets and produce aesthetic prediction. More details will be illustrated in this section.

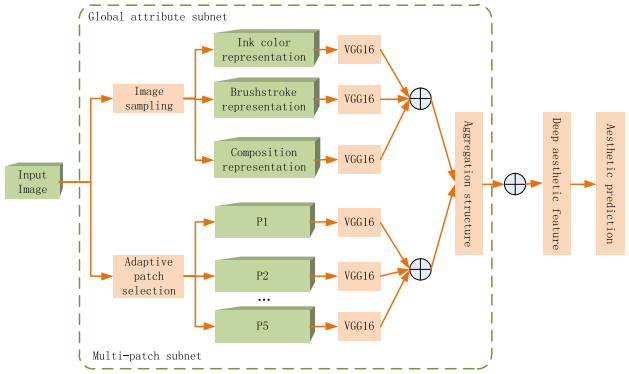


FIGURE 6. Deep multi-view parallel CNN structure.

1) GLOBAL ATTRIBUTE IMAGE EXTRACTION

For ink color representation, the Chinese ink painting stresses the contrast between color, ink brush stroke, and white space, as well as the shading variation. Therefore, we choose the original warped image, HSV components, and grayscale map as inputs. For the texture representation, we extract brushstroke image of each ink painting [36], which exploits an integration of edge detection and clustering-based segmentation, as shown in Figure 7. Since the wavelet coefficients of the image contain abundant edge energy information and capture the local details of wielding the brush, we also use the first layer wavelet coefficient matrix of Daubechies as another texture input.

For the composition representation, we analyze the spatial distribution by detecting salient map inside the image.

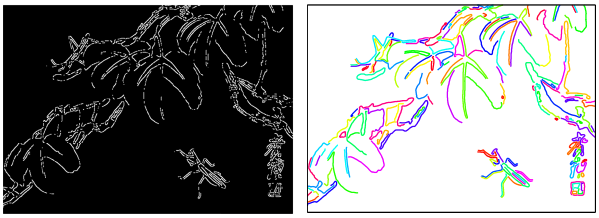


FIGURE 7. Procedure of strokes segments extraction. (a) Detection of strokes boundaries using canny edge operator; (b) Selection of continuous segments on detected boundaries in extracted color map.

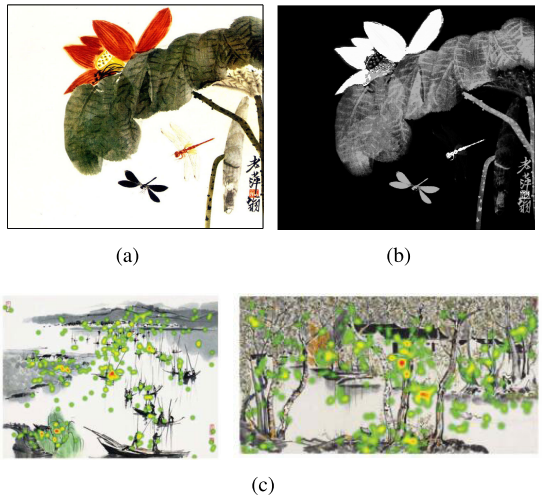


FIGURE 8. Composition representations in Chinese ink painting. (a) Original image; (b) Grayscale saliency map (lighter regions have larger saliency values); (c) Heat maps of eye movements [23].

Here we use the histogram-based contrast method proposed in [37] to calculate the grayscale saliency map of the original image, as shown in Figure 8(a)-8(b). Also, the eye movements are recorded by a Tobii eye-tracker of model X120 to record viewers' eye movements. The heat map can clearly illustrate the attended regions in Chinese ink paintings, as shown in Figure 8(c) [23].

2) ADAPTIVE LOCAL PATCH SELECTION

For local view inputs, we represent each image with a set of carefully cropped patches and associate the set with the image's aesthetic score. Different from the random-cropping method in [35], we adopt the adaptive local patch selection strategy proposed in [8] to carefully select the most discriminative and aesthetic informative patches in Chinese ink paintings, as shown in Figure 9. Several criteria have been developed to perform patch selection:

a: SALIENCY AND HEAT MAP

The tasks of saliency detection and eye movement of the participants are to identify the most important and attended region of a painting. Therefore, it is natural to adopt saliency

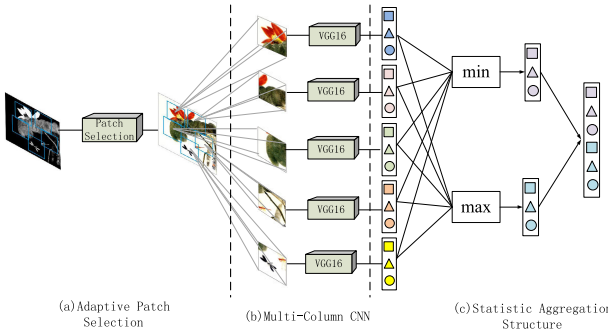


FIGURE 9. The architecture of Multi-Patch subnet [8]: (a) adaptive patch selection, (b) a set of multi-column CNNs that are used for extracting deep features from each patch, (c) aggregation module which incorporates the extracted features from the paralleled VGG16.

and heat maps for selecting visually significant regions that human usually pays more attention to.

b: PATTERN DIVERSITY

Different from image classification that often focuses on the foreground objects, some important aesthetic characteristics, e.g. shading contrast, sparse and dense contrast, and visual balance can only be perceived by analyzing the relationships between different subjects and white space as a whole, so we should focus on diversification within a set of patches.

c: OVERLAPPING CONSTRAINT

The overlapped ratio of any selected patch pairs should be restrained based on the spatial distance. Therefore, we can formulate the patch selection as an objective function, which can be defined to find the optimal combination of patches [8]:

$$c^* = argmax F(S, H, D_p, D_s) \tag{2}$$

$$F(\cdot) = \sum_{i=1}^{M}(S_i \cdot H_i) + \sum_{i\neq j}^{M} D_p(\widetilde{N}_i, \widetilde{N}_j) + \sum_{i\neq j}^{M} D_s(c_i, c_j) \tag{3}$$

where $\{c_m^*\}_{m\in[1,M]}$ is the centers of the M selected optimal patches. $S_i \cdot H_i = \frac{sal(p_i)*gray(p_i)}{area(p_i)}$ is the normalized saliency value of each patch p_i. $D_p(\cdot)$ is the pattern distance function which evaluates the patches' pattern difference. Here we use a multivariant Gaussian to model the pattern of a patch p_m [8]:

$$\widetilde{N_m} = \{\{N_e(\mu_e, \sigma_e)\}_m, \{N_c(\mu_c, \sigma_c)\}_m\}_{m\in[1,M]} \tag{4}$$

where $\{N_e(\mu_e, \sigma_e)\}_m$ and $\{N_c(\mu_c, \sigma_c)\}_m$ denote brushstroke distribution and shading distribution of patch p_m, respectively. $D_s(\cdot)$ is the Euclidean spatial distance function. Figure 10 shows some results of patches extracted by the adaptive selection strategy. We can see that, the proposed method not only is effective in selecting the most prominent regions (e.g. flowers, birds, and loquat), but also is capable of capturing the pattern diversity (e.g. the branches and mountains, grass and insects).

FIGURE 10. Examples of selected patches by the proposed adaptive patch-selection scheme. The size of all the patches are 224 × 224.

3) TRAINING DETAILS

Training a DMVCNN consists of three steps. First, the basic CNN (from the input layer to the fc256 layer) is trained, then the weights of CNN in the DMVCNN are initialized by the weights of the learned CNN. Then we train the two subnets respectively, in which the global input is 9 channels and the number of local patches is 5. Finally, we combine the two subnets to fine-tune the overall DMVCNN, and concatenate the two 256 × 1 vectors from the output of each subnet to jointly train the weights of the final regression layer. We use the final output of the 512 × 1 vector as the deep aesthetic features of Chinese ink paintings.

VI. COMBINING THE DEEPLY-LEARNED AND HANDCRAFTED FEATURES

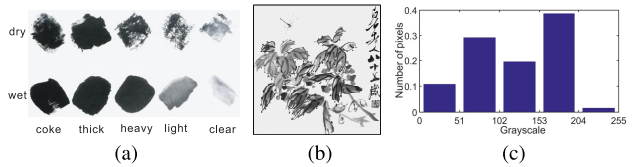
The handcrafted features are based on visual psychology and art aesthetic rules in the domain expertise of Chinese ink paintings, while the deep aesthetic features are automatically learned from the original image pixels through DMVCNN, which lacks interpretability. By combining these two types of complementary features, the aesthetic information can be described more comprehensively. We construct a total of 77 numerical features in the sets denoted by $X_{HC} = \{f_i | 1 \leq i \leq 77\}$, which are derived from authoritative appreciation books and professional Chinese painters empirical recommendations.

A. INK COLOR-RELATED FEATURES

In Chinese ink paintings, the objects are expressed with ink and colored brush strokes, and white space is left for imagination. A high aesthetics ink painting should have a natural blending of color and ink. Here we calculate the area ratio of the colored brushstroke region R_c to the whole brushstroke region R_b as f_1, the hue count as f_2, the maximum and average hue contrast between different regions in the R_c as f_3–f_4, and the weighted hue as f_5 [29].

1) INK SHADING

In Chinese ink paintings, various shading effects as shown in Figure 11(a) are produced by varying ink density including coke, thick, heavy, light, and clear, which are used as "colors" and traditionally called as "five-color ink" principle. Since those effects can be characterized by the grayscale



(a) (b) (c)

FIGURE 11. Ink shading. (a) Five-color ink; (b) Grayscale image; (c) The 5-bin grayscale histogram.

of the ink strokes in the painting, we divide the grayscale image I_g (Figure 11(b)) into 5 bins and calculate a 5-bin grayscale histogram h_g on I_g, as indicated in Figure 11(c). We characterize the shading effects globally by calculating the standard deviation, skewness, and kurtosis of h_g, denoted as f_6, f_7, and f_8, respectively.

From a local perspective, we define $I_S(i)$ and $I_g(ink)$ as the average saturation of ith colored brushstroke region $R_c(i) \in R_c$ and the average grayscale of ink brushstroke region R_{ink}, respectively, the area of R_{ink} is denoted as A_{ink}. We define the weighted saturation to reflect the degree of a color impact compared with white space. The weighted saturation is expressed as:

$$f_9 = \frac{\sum_i I_S(i) \cdot A_c(i) + (255 - I_g(ink)) \cdot A_{ink}}{\sum_i A_c(i) + A_{ink}} \quad (5)$$

B. BRUSHWORK-RELATED FEATURES

In Chinese ink paintings, textures are produced by brushstrokes, due to interactions between the paper and brush controlled with different direction, strength, touching time, mark thickness, etc. It has become an important aesthetic feeling in Chinese ink paintings that the brush strokes with moisture flow out on the paper with fluent traces.

(a) (b)

FIGURE 12. Dry-wet contrast. (a) The maple leaf rendered with wet ink and the leaf stalk outlined with coke ink; (b) Extracted dry ink regions (red boxes) using the energy feature in GLCM.

The **dry-wet contrast** creates gradation variations over brush strokes by grading different amounts of water in Chinese ink paintings, as shown in Figure 12(a) where the maple leaf is rendered with wet ink, and the leaf stalk is outlined with coke ink. An aesthetically pleasing Chinese ink painting should have a good combination of dry-wet inks to produce "spirit" in visualization. Since the wet ink brush strokes usually produce smooth textures and the dry strokes produce rough textures, which can be characterized by the energy in gray-level co-occurrence matrix (GLCM) [38], we adopt GLCM to detect regions associated with dry and wet ink brush strokes. We divide the input image into 40×40 equal sub-regions (the region with such size can capture enough variation in texture) and calculate the energy of each region. If the value is smaller than 0.3, the region is regarded as a dry ink region, as shown by red boxes in Figure 12(b). All pixels in the sub-regions belong to dry ink are summed to obtain the dry ink stroke region, the rest pixels form the wet ink stroke region, and their area proportion is denoted as f_{10}.

1) VARIABILITY OF STROKES

The strokes techniques stress the richness of variations in thickness, transition, pause, and strength, which create rhythm and dynamic aesthetics in Chinese ink paintings. Here we calculate the statistical measures of 0.25 quantile, 0.75 quantile, median and mean for the curving degrees of all the contour segments [39]. Besides, the number of brushstrokes with similar orientations in the neighborhood, orientation standard deviation in the neighborhood, and the standard deviation, skewness and kurtosis of the histograms in edge length, straightness, stroke intensity, broadness homogeneity [36], [40], respectively. Accordingly, features $f_{11} \sim f_{28}$ are obtained.

2) PERCEPTUAL FLUENCY OF STROKES

A natural fluency of wielding the brush strokes, with relaxation, tension, pause, and transition, also contributes to the aesthetics of paintings. We calculate the frequently-used Tamura features of coarseness, the Neighbourhood Gray Tone Difference Matrix features of strength, and the mean, standard deviation, entropy of the Gabor energy map entries [38] to capture perceptual properties of visual textures in Chinese ink paintings. Accordingly, features $f_{29} \sim f_{33}$ are obtained.

(a) (b)

FIGURE 13. Procedure of extracting the salient regions. (a) SLIC superpixel segmentation; (b) Salient regions binary map (salient regions are white and the rest black).

C. COMPOSITION-RELATED FEATURES

A Chinese ink painting is aesthetically pleasing to the eye if the objects within the work are arranged in a balanced compositional way. Here we first segment the original image into perceptually homogeneous regions, using an adaptive SLIC superpixels segmentation technique [41], as shown in Figure 13(a). Then, according to the saliency map, we select the salient regions whose average saliency values are greater than 2 of the overall average saliency value, as shown in Figure 13(b). The prominent lines, denoted as P_i, where

$i = 1, 2, \ldots, n_p$ and n_p is the number of the prominent lines, are also detected in the image. We define $V(P_i)$ as average saliency value of P_i. On this basis, we quantize the following several composition rules commonly-used in Chinese ink painting.

1) RULE OF THIRDS

The rule considers the image to be equally divided into thirds both horizontally and vertically, as shown in Figure 14(a), and it specifies that the main subjects should lie around these intersections or concentrated distribute around the third lines, which produces a sense of pleasant and vivid feeling visually. Here we utilize the method in [42] to measure the degree of closeness between the salient regions and the power points as f_{34}.

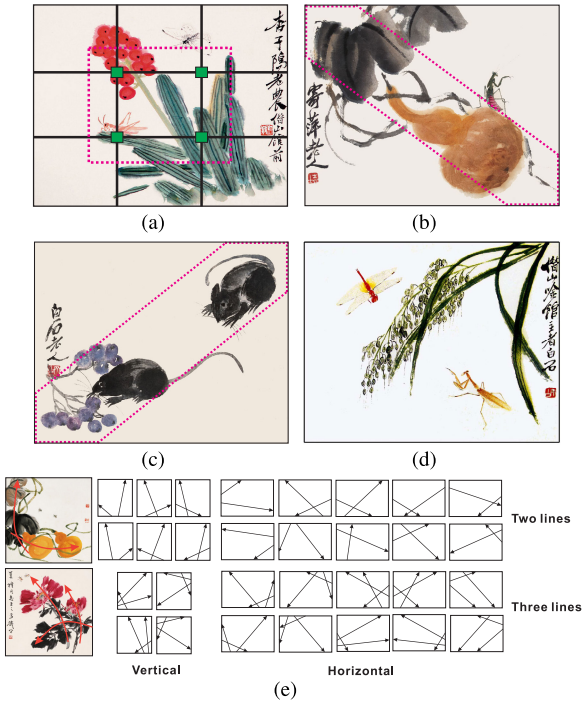


FIGURE 14. Basic composition rules and examples. (a) Rule of thirds. The third lines (black) are overlayed with 4 power points (green); (b, c) Diagonal and back diagonal dominance and associated masks (pink); (d) Visual balance. Objects are evenly distributed around the center; (e) Two-line and three-line rules and associated different line dragrams.

2) DIAGONAL DOMINANCE

The diagonals of the painting image are also aesthetically significant, which create a dynamic emphasizing effect, as shown in Figure 14(b) $\sim$ 14(c). We compute the maximum of RFA based on diagonal and back diagonal mask [42] as f_{35}.

3) VISUAL BALANCE

The visual balance refers to the equal distribution of visual weight in a Chinese ink painting, as shown in Figure 14(d). In the painting arts, it is a more aesthetic and lively expression form rather than symmetry. Here we use the visual balance score [42] to measure the degree of closeness between the "center of mass" which incorporates all salient regions and the image center as f_{36}.

4) TWO-LINE AND THREE-LINE RULES

The Chinese ink painting uses two-line or three-line combinations as composition skeletons to organize the objects, thus creating a potential power of movement in the painting, as shown in Figure 14(e). Different line changes produce various line diagrams (vertical or horizontal direction), each one corresponds to a picture, with intersections at the golden section ratio [22]. The k^{th} line in the two-line or three-line diagram is denoted as $L2_k$ or $L3_k$, respectively. Here we measure how close the prominent lines lie to the two-line or three-line diagram by calculating the line score as:

$$E_{l2} = \max\{\frac{\sum_{i=1}^{n_p} V(P_i) \cdot e^{-\frac{D_2^2(P_i,k)}{2\sigma}}}{\sum_{i=1}^{n_p} V(P_i)}\} \tag{6}$$

$$E_{l3} = \max\{\frac{\sum_{i=1}^{n_p} V(P_i) \cdot e^{-\frac{D_3^2(P_i,k)}{2\sigma}}}{\sum_{i=1}^{n_p} V(P_i)}\} \tag{7}$$

$$f_{37} = \max\{E_{l2}, E_{l3}\} \tag{8}$$

where $\sigma = 0.17$, $D_2(P_i, k)$ and $D_3(P_i, k)$ are the minimal distances from the prominent line P_i to the k^{th} lines in two-line and three-line diagrams, respecitvely:

$$D_2(P_i, k) = \min\{d_L(P_i, L2_k)\}, \quad k = 1, 2 \tag{9}$$

$$D_3(P_i, k) = \min\{d_L(P_i, L3_k)\}, \quad k = 1, 2, 3 \tag{10}$$

where d_L represents distance between two line segments.

5) DISTRIBUTION OF WHITE SPACE

Leaving white space is a typical artistic treatment in Chinese ink painting. Here the white space is not really empty, it creates the balance between ink and white space, which in turn better contributes to the composition of the entire painting, and the ink occupation in the painting should be strictly controlled [23]. Here we first calculate the area ratio of the brush stroke region to the whole image as f_{38}, to reflects a global relationship between the emptiness and solidness in the painting.

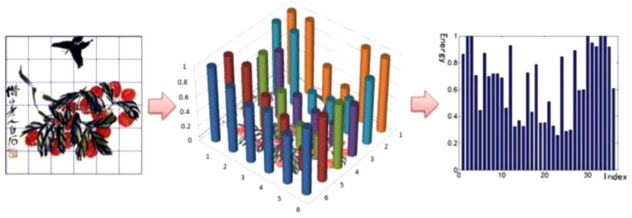


FIGURE 15. The ink intensity distribution measured by GLCM [40].

Besides, a well-composed Chinese ink painting should have a good arrangement of ink intensities in a given region. We divide a whole painting area into 6×6 equal size sub-regions, as shown in Figure 15. Then the GLCM is used to measure the energy of each sub-region. Finally we combine each sub-regions' energy together in their index order,

denoted as $\{f_{39}, f_{40}, \ldots, f_{74}\}$, as shown in Figure 15. We also characterize the distribution of sparse and dense by calculating the mean, standard deviation, and entropy of the array, denoted as f_{75}, f_{76} and f_{77}, respectively.

D. FEATURE FUSION

The handcrafted features X_{HC} are concatenated with the learned deep aesthetic features X_{CNN} (the DMVCNN output of 512×1 vector) into a single features representation $X_{aes} = [X_{CNN}; X_{HC}]$, which are then used to learn a function $f : X_{aes} \rightarrow Y$ through SVR modeling. Then a comprehensive SVR model is trained on this fused feature set to obtain the final aesthetic evaluation model.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

In the implementation, all the network training and testing are done on the PCs with Intel i5-7500k CPUs and NVIDIA 1080Ti GPUs, by using the Tensorflow deep learning framework [43]. The base learning rate is 0.001, and is annealed by 0.1 every time the training loss plateaus. The weight decay is 1e-5 and momentum is 0.9. The optimizer uses stochastic gradient descent. We make a horizontal transformation on the ink painting dataset from 1200 to 2400, and evaluate the prediction performance of the models using 5-fold cross-validation. Accordingly, three objective evaluation metrics were used to judge the experimental results. Compared with other existing methods, our DMVCNN architecture and associated comprehensive model achieve the most advanced performance, as detailed in the rest of this section.

A. EVALUATION METRICS

To evaluate the prediction performance of our proposed method, we calculate the average Pearson correlation coefficient with 2-tailed Significance (denoted as ave $R_P/Sig.$) and average mean squared error (denoted by ave MSE) between model evaluated scores and manually evaluated scores on all test folds. Here the R_P is used to measure the linear correlation between two variables, ranging between -1 and 1, which can be calculated as:

$$R_{P(a,p)} = \frac{Cov(a,p)}{\sqrt{D(a)}\sqrt{D(p)}} \tag{11}$$

where $Cov(a, p)$ can be expressed as

$$Cov(a,p) = \frac{\sum_{i=1}^{N}(a_i - \overline{a})(p_i - \overline{p})}{N} \tag{12}$$

where $\overline{a}$ and $\overline{p}$ are the mean of the human evaluated actual aesthetic score and model predicted score, respectively, and N is the number of images in test sets. Then, $\sqrt{D(a)}$ and $\sqrt{D(p)}$ can be written as:

$$\sqrt{D(a)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(a_i - \overline{a})^2} \tag{13}$$

$$\sqrt{D(p)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \overline{p})^2} \tag{14}$$

TABLE 1. Predictive performance comparisons of CNN models in different basic architectures.

Architecture	ave $R_P/Sig.$	ave MSE	ave Acc(%)
AlexNet	0.627/0.000	0.415	72.213
VGG16-arc1	0.674/0.000	0.336	76.851
VGG16-arc2	0.661/0.000	0.348	75.285
VGG16-arc3	**0.712/0.000**	**0.285**	**81.155**

Statistically, any regressor with R_P larger than 0.8 is regarded as strong, and less than 0.5 as weak. The $Sig.$ indicates the statistical significance of the regression model. If the Sig-value is less than or equal to 0.05, meaning that the regression model statistically significantly predicts the outcome variable. Also, the MSE is a measure of the degree of difference between the human rating and model prediction, which can be formulated as:

$$MSE_(g,p) = \frac{1}{N}\sum_{i=1}^{N}(a_i - p_i)^2 \tag{15}$$

By simply thresholding the predicted aesthetic score, we can get a binary classification average accuracy over all test folds (denoted as ave Acc). When the aesthetic score of an image is higher than 4, it is a highly aesthetic image. Aesthetic scores below 4 are considered to be low aesthetic images. Suppose that TH_p is the correct number of predicted highly aesthetic image, and TL_p is the correct number of predicted low-aesthetic image. Thus, the Acc can be calculated as:

$$Acc = \frac{TH_p + TL_p}{N} \tag{16}$$

B. PERFORMANCE ANALYSIS OF BACKBONE CNN NETWORK

By taking the original painting image as input, we first compare the aesthetic evaluation performances of single-column CNN models under different basic network architectures, as shown in Table 1. We first fine-tuned the AlexNet model on our dataset (set the output of the last fully connected layer to 1), then VGG16-arc1 denotes the original VGG16 structure, VGG16-arc2 denotes the VGG16 network of reduced neurons in the last two fully-connected layers without multi-subject layer, and VGG16-arc3 denotes the VGG16 network with 6 parallel convolutional groups in subject layer and reduced fully-connected neurons, which is the backbone CNN network in our method. The statistics in the table show that the model in VGG16-arc3 yields a higher significant average $R_P/Sig.$ ($p < 0.05$), higher average Acc, and lower average MSE than the other two architectures, which verify the effectiveness of the paralleled multi-subject layer design and pre-training strategy. This indicates that by using fewer network parameters, our network architecture and transfer learning strategy can not only make full use of relevant general aesthetics features (e.g. edge filters and color blobs in earlier layers) in the pre-trained VGG16 model, but also effectively capture aesthetic descriptors for different

categories in Chinese ink paintings by using the adaptive subject convolutional layer.

C. PERFORMANCE ANALYSIS OF DMVCNN MODEL

On this basis, we compare the predictive performances on the aesthetic perception in Chinese ink painting of our proposed DMVCNN models with inputs from different perspectives, and evaluate it with some other existing methods, as shown in Table 2, as detailed next:

TABLE 2. Predictive performance comparison of different models on the Chinese ink painting dataset.

Method	ave $R_P/Sig.$	ave MSE	ave Acc(%)
AVA [18](2012)	0.462/0.000	0.532	56.125
Zhang *et al.* [29](2017)	0.556/0.000	0.446	64.013
RAPID [6](2014)	0.645/0.000	0.371	73.155
DMA-Net [35](2015)	0.654/0.000	0.352	74.202
Wang *et al.* [7](2016)	0.669/0.000	0.344	76.060
STCNN [9](2017)	0.649/0.000	0.364	74.861
A-lamp [8](2017)	0.768/0.000	0.229	84.533
Kucer *et al.* [11](2018)	0.686/0.000	0.336	76.451
ILGNet [13](2019)	0.759/0.000	0.244	84.134
VGG16-arc3	0.712/0.000	0.285	81.155
DMVCNN-global	0.742/0.000	0.254	84.128
DMVCNN-local	0.786/0.000	0.206	86.211
DMVCNN	**0.797/0.000**	**0.185**	87.055
SVR$_{Deep+HC_{color}}$	**0.815/0.000**	**0.177**	88.410
SVR$_{Deep+HC_{stroke}}$	**0.820/0.000**	**0.168**	88.713
SVR$_{Deep+HC_{comp}}$	**0.827/0.000**	**0.162**	89.233
SVR$_{Deep+HC_{hybrid}}$	**0.843/0.000**	**0.146**	91.075

1) DMVCNN VS. BASELINE

To examine the effectiveness of the proposed scheme, we compare DMVCNN with some baseline methods using handcrafted features, i.e. AVA [18] and Zhang *et al.* [29]. We can see that, all recently developed deep CNN schemes, especially DMVCNN, outperform conventional handcrafted feature-based approaches by a significant margin, thus verifying the effectiveness of deep learning in the quantitative aesthetic evaluation of Chinese ink painting. This superior performance comes from the fact that the deep network has effectively gained knowledge and is capable to extract highly discriminative representations from Chinese ink painting.

2) DMVCNN VS. CNNs IN PHOTO AESTHETICS

We also compare the proposed scheme with some latest CNN models in aesthetic assessment of photographs, i.e. RAPID [6], DMA-Net [35], Wang *et al.* [7], STCNN [9], A-lamp [8], Kucer *et al.* [11], and ILGNet [13]. The results show that the DMVCNN model significantly outperforms all these existing methods in predicting the aesthetics of Chinese ink paintings. Such results further validate the effectiveness of our proposed network in comprehensively depicting aesthetic properties of Chinese ink paintings, especially the adaptive selection strategy can effectively extract the most aesthetically significant regions, and capture the pattern diversity between different subjects and background in the Chinese ink painting, such as the density contrast between brushstroke and white space, the contrast between emptiness and solidness, dynamic-static comparison.

3) DMVCNN VS. SINGLE-VIEW CNNs

Besides, we experiment on the proposed scheme with three views of inputs. VGG16-arc3 denotes the single-view VGG16 network that takes only original warping as input, DMVCNN-global and DMVCNN-local denote the multi-column VGG16 that takes global attributes images and adaptive local patches as inputs, respectively. The statistics in the table indicate that our DMVCNN model that takes hybrid inputs yields a higher significant average $R_P/Sig.$ ($p < 0.05$), higher average Acc, and lower values of average MSE than other sing-view CNNs. This proves that our multi-view parallel deep CNN can effectively extract global attribute information and local fine-grained details as heterogeneous inputs to jointly learn deep aesthetic features in Chinese ink paintings.

D. EFFECTIVENESS OF COMPREHENSIVE AESTHETIC MODEL

We further investigate the effectiveness of our comprehensive aesthetic evaluation model by fusing DMVCNN activations with handcrafted features. The obtained model is denoted as SVR$_{Deep+HC_{hybird}}$. We compare it with the DMVCNN model, and SVR model trained with a fusion of deep aesthetic features and handcrafted ink color-related features (denoted as SVR$_{Deep+HC_{color}}$), brushwork-related features (denoted as SVR$_{Deep+HC_{stroke}}$), and composition-related features (denoted as SVR$_{Deep+HC_{comp}}$), respectively, as shown in Table 2. The results show that the SVR$_{Deep+HC}$ model yields a higher significant average $R_P/Sig.$ ($p < 0.05$), higher average Acc, and lower average MSE than other models. By combining the deep aesthetic and hybrid handcrafted features, our proposed model can predict human aesthetic judgment with Pearson highly significant correlation of 0.843, which indicates an improvement up to 5.7% than the DMVCNN model, and lower values of average MSE indicate that our model can predict manually evaluated results with a lower regression loss. We can conclude that the fusion of deeply-learned aesthetic features and handcrafted expert knowledge in the SVR model have high predictive power in aesthetic evaluation of Chinese ink paintings.

We show some examples of Chinese ink paintings associated with scores evaluated in opposite positions for predicted aesthetics by subjects and comprehensive aesthetic model in Figure 16. The Chinese ink paintings on the top row are very high aesthetic and those on the bottom row are very low aesthetic among the experimental dataset. The numerical figures indicate that our model can describe the predicted aesthetics of Chinese ink paintings in accord with human perception. Obviously, the images in the top row have more nature gradation in ink shading, more smooth variation in brushstroke, and more harmonious distribution than in the bottom row. Also, we show some examples that are misjudged by the model, as shown in Figure 17. The main reason is that these famous paintings with abstract styles have a large

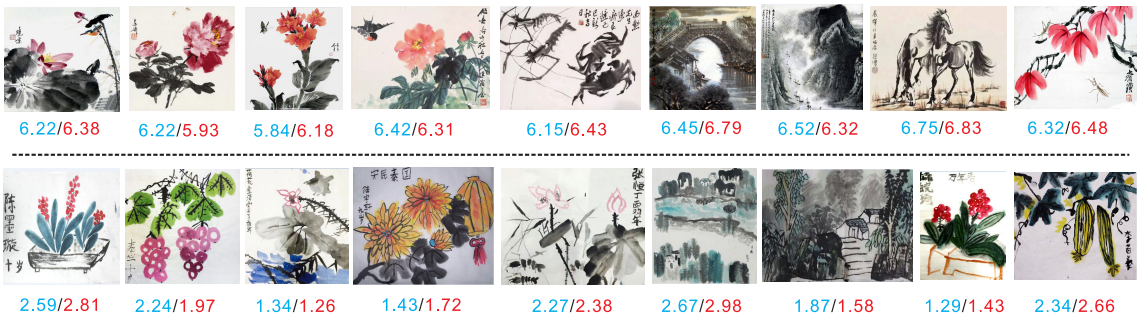


FIGURE 16. Examples of Chinese ink paintings and associated scores evaluated in opposite positions by subjects (in blue) and comprehensive aesthetic models (in red).

FIGURE 17. Examples of Chinese ink paintings that are misjudged by our comprehensive aesthetic model.

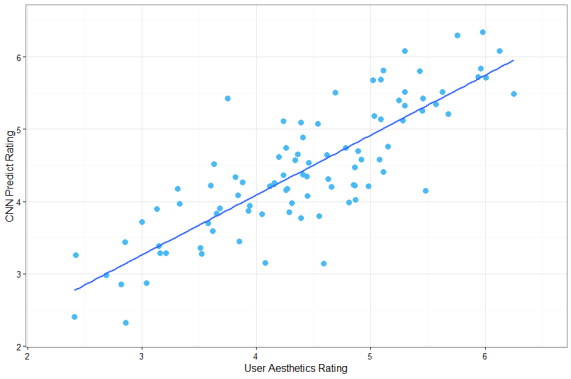


FIGURE 18. Plot diagram of aesthetic rating predictions with comprehensive model and human judgment.

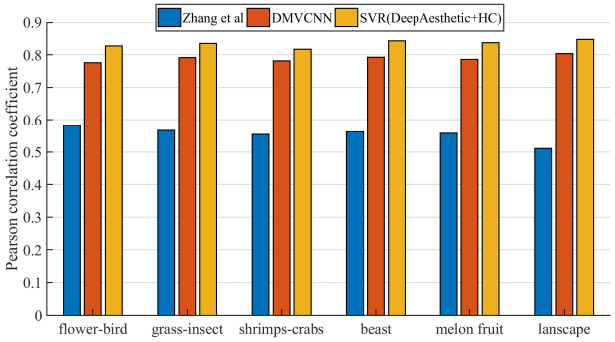


FIGURE 19. Comparison of aesthetic prediction performance in different categories.

deviation in the general aesthetic criteria of the similar subjects in the dataset. Besides, we show some prediction results in Figure 18. We can observe that the model estimated ratings are highly correlated with the user ratings.

We further investigate the predictive performance of different models on each of the 6 subjects, as shown in Figure 19. This figure shows that the comprehensive aesthetic model significantly and consistently outperforms Zhang *et al.* [29] with low-level visual features and DMVCNN models across all 6 categories. The results further demonstrate the effectiveness and robustness of our fusion of deeply-learned and handcrafted aesthetic features.

E. MULTI-DIMENSIONAL FEATURE IMPORTANCE ANAYSIS

Since the DMVCNN model achieves a significant improvement after feature fusion, we can examine the top-performing HC features that providing the most discriminatory power after recursive feature elimination methods in the comprehensive model, and understand, for example, the type of high-level art knowledge that is approximated by the HC features and missing from DMVCNN. Table 3 shows the descriptions of 18 most important handcrafted features ranking by gain based ranking criterion in descending order. We try to give a meaningful interpretation of importance in these feature variables based on professional Chinese painting art knowledge and psychology theory:

f_9, f_6, f_7, f_5, f_1: An affordable countless shades of ink color would result in rich gradation of colors and enhance the aesthetic feeling in Chinese ink paintings. f_9 and f_5 measure the degree of average saturation and hue impact, which indicate the degree of harmonious blending and contrast between color, ink, and white space. A harmonious grayscale distribution characterized by f_6 and f_7 indicates a natural flexible transition and rich gradation in ink color. Besides, a good

TABLE 3. List and description of the the top performing HC features that improve the performance of DMVCNN model.

Feature	Description	Category
f_9: saturation	Weighted saturation	ink color
f_6: sd-gray	Standard deviation of the grayscale histogram	ink color
f_7: skew-gray	Skewness of the grayscale histogram	ink color
f_5: hue	Weighted hue	ink color
f_1: area-c	Area proportion of colored and the whole brush strokes	ink color
f_{10}: dry-wet	Area proportion of dry ink stroke and wet ink stroke region	brush stroke
f_{32}: sd-gabor	Standard deviation of the Gabor energy map	brush stroke
f_{24}: skew-int	Skewness of stroke intensity	brush stroke
f_{29}: coarseness	Tamura-based coarseness	brush stroke
f_{30}: strength	NGTDM-based strength	brush stroke
f_{14}: curving	Mean curving degree of stroke boundaries	brush stroke
f_{76}: sd-sparse	Standard deviation of sparse and dense distribution in the GLCM energy of 6x6 sub-regions	composition
f_{36}: balance	Visual balance score	composition
f_{34}: RT	Closeness between the salient regions and the power points	composition
f_{35}: RFA-d	Maximum of RFA based on diagonal and back diagonal masks	composition
f_{37}: line score	Maximum of line score based on two-line and three-line diagrams	composition
f_{77}: ent-sparse	Entropy of sparse and dense distribution in the GLCM energy of 6x6 sub-regions	composition
f_{38}: area-b	Area proportion of brush strokes in the whole image frame	composition

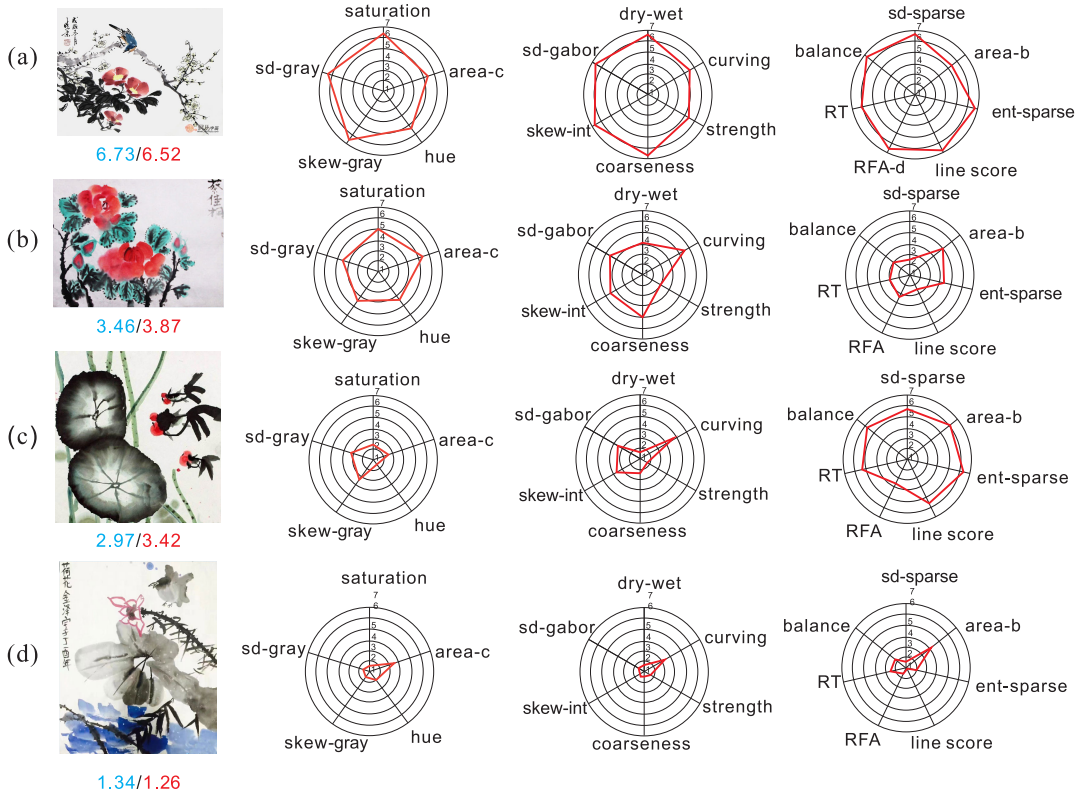


FIGURE 20. Comparison of Chinese ink paintings with different ranking aesthetic scores, and selected feature variables are displayed in polar diagrams.

proportion of color and ink reflected by f_1 causes a pleasing visual effect in Chinese ink paintings.

$f_{10}, f_{32}, f_{24}, f_{29}, f_{30}, f_{14}$: A freely flowed brushstroke should have a harmonious dry-wet contrast measured by f_{10}, an ordered and flexible variation measured by f_{32}, and smooth and powerful line movement measured by f_{24}, f_{29}, and f_{30}. Also, a Chinese ink painting with more sharp transitions in strokes reflected by f_{14} may bring negative and insecure feelings. This kind of feelings may cause slight anxieties instead of pleasing emotions.

f_{76}, f_{77}, f_{38}: Highly aesthetic Chinese ink painting stresses the contrast equilibrium of emptiness, solidness, sparseness, dense between brush strokes and white space. A balanced and harmonious ink occupation must be neither too crowded nor too loose. Therefore, if the degree of dispersion for the brush stroke distribution measured by f_{76} and f_{77} is too large, and

the proportion of brush strokes and white space measured by f_{38} is too large or too small, the Chinese ink painting is likely to be low aesthetic quality.

$f_{36}, f_{34}, f_{35}, f_{37}$: A visually balanced painting characterized by f_{36} must have a harmonious distribution of white space and brush strokes, thick and light, wet and dry ink brush strokes. On this basis, the rule of third measured by f_{34} plays the role of organizing the subjects in the picture to strengthen the sense of coordination in surroundings. The diagonal dominance measured by f_{35} in the painting can create a dynamic effect which could cause aesthetically pleasing. Besides, the two-line and three-line rules reflected by f_{37} create a rhythmic and ordered variation of visual effect in the painting.

Furthermore, the magnitudes of all 18 top-performing HC feature provide some insight into which kind of attribute contribute more on overall aesthetic evaluation of Chinese ink painting, as demonstrated by 4 examples in Figure 20, where comparative ranking aesthetic scores and selected feature variables are displayed with polar diagrams. In Figure 20(a) all features are obtained with higher values, thus achieving a high ranking for the overall perceived aesthetics of this professional painting. For the amateur paintings, a lower score is given to composition in Figure 20(b) due to the visual imbalance of objects and white space distribution, unfitness for rule of third, diagonal dominance, and two-line rules. Also, lower quality in perceived color and texture for Figure 20(c) is caused by the unnatural grayscale distribution, redundant wet ink brush strokes, and disordered variation of brushwork, the overall aesthetics scores are slightly smaller than that in Figure 20(a). While all values of features in Figure 20(d) are very low, consequently the overall aesthetics score is the lowest.

VIII. CONCLUSION AND FUTURE WORK

Building a connection between human aesthetic perception and computational visual features is a challenging multidisciplinary problem. We propose a comprehensive aesthetic evaluation framework of Chinese ink paintings based on deep learning. Experimental results show that, when fusing the deeply-learned features with handcrafted features that rely on art expert knowledge, our proposed model can predict human aesthetic judgment with Pearson highly significant correlation of 0.843, with an improvement up to 5.7% than the purely deep learning model. Our work provides a deep-learning scheme for the quantitative aesthetic evaluation of Chinese paintings, and also an innovative exploration of Chinese traditional art resources in the current era of artificial intelligence. Besides, it investigates the possibility of fusing handcrafted features with learned features for improving aesthetic inference, which enhances our understanding on the interpretability of the computational aesthetic model of Chinese ink paintings.

Currently the images selected in our experiment are mostly Xieyi brush works, in the future we will include the Gongbi ink paintings in our dataset and improve our comprehensive aesthetic model. Besides, we will extend the deep neural network to the multi-task deep learning of Chinese paintings, and research on how to better assist the aesthetic evaluation task through the prediction of other high-level semantic information, such as emotional artistic conception. Moreover, we will perform a more in-depth study on quantitative analysis of philosophical meaning in Chinese paintings.

REFERENCES

[1] F. Hoenig, "Defining computational aesthetics," in *Proc. Euro. Conf. Comput. Aesthet.* Cham, Switzerland: Eurographics Association Aire-la-Ville, 2005, pp. 13–18.
[2] P. A. Fishwick, *Aesthetic Computing*. Cambridge, MA, USA: MIT Press, 2008.
[3] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1531–1544, Jul. 2019.
[4] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[5] Y. Zhang and D. M. Chandler, "Opinion-unaware blind quality assessment of multiply and singly distorted images via distortion parameter estimation," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5433–5448, Nov. 2018.
[6] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "RAPID: Rating pictorial aesthetics using deep learning," in *Proc. ACM Int. Conf. Multimedia*, New York, NY, USA, 2014, pp. 457–466.
[7] W. Wang, M. Zhao, L. Wang, J. Huang, C. Cai, and X. Xu, "A multi-scene deep learning model for image aesthetic evaluation," *Signal Process., Image Commun.*, vol. 47, pp. 511–518, Sep. 2016.
[8] S. Ma, J. Liu, and C. W. Chen, "A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4535–4544.
[9] Y. Kao, R. He, and K. Huang, "Deep aesthetic quality assessment with semantic information," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1482–1495, Mar. 2017.
[10] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.
[11] K. Michal, C. L. Alexander, and W. M. David, "Leveraging expert feature knowledge for predicting image aesthetics," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5100–5122, Dec. 2018.
[12] X. Zhang, X. Gao, W. Lu, and L. He, "A gated peripheral-foveal convolutional neural network for unified image aesthetic prediction," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2815–2826, Nov. 2019.
[13] X. Jin, L. Wu, X. Li, X. Zhang, J. Chi, S. Peng, S. Ge, G. Zhao, and S. Li, "ILGNet: Inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation," *IET Comput. Vis.*, vol. 13, no. 2, pp. 206–212, Mar. 2019.
[14] C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 236–252, Apr. 2009.
[15] A. Sartori, V. Yanulevskaya, A. A. Salah, J. Uijlings, E. Bruni, and N. Sebe, "Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 2, pp. 1–27, Jul. 2015.
[16] C. Li, S. Q. Sun, X. Min, W. X. Wang, and Z. C. Tang, "Application of deep convolutional features in sketch works classification and evaluation," *J. Comput. Aided Des. Comput. Graph.*, vol. 27, no. 10, pp. 1898–1904, 2017.
[17] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1930–1943, Dec. 2013.
[18] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2408–2415.
[19] S. A. Amirshahi, G. U. Hayn-Leichsenring, J. Denzler, and C. Redies, "Jenaesthetics subjective dataset: Analyzing paintings by subjective scores," in *Computer Vision* (Lecture Notes in Computer Science). Springer, 2015, pp. 3–19.
[20] S. Yokochi and T. Okada, "Creative cognitive process of art making: A field study of a traditional chinese ink painter," *Creativity Res. J.*, vol. 17, nos. 2–3, pp. 241–255, Jul. 2005.

[21] J. Rawson, *The British Museum Book of Chinese Art*, 2nd ed. London, U.K.: British Museum Press, 2007.

[22] C. Hannian, "Charm, elegance, charm and vividness-interpretation of the first of the six law in Xiehe's ancient painting records," *Beauty age*, vol. 649, no. 04, pp. 38–39, 2016.

[23] Z. Fan, X. S. Zheng, and K. Zhang, "Computational analysis and eye movement experiments of white space in chinese paintings," in *Proc. IEEE Int. Conf. Prog. Informat. Comput. (PIC)*, Washington, DC, USA, Dec. 2015, pp. 301–306.

[24] S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *Proc. CVPR*, Washington, DC, USA, Jun. 2011, pp. 1657–1664.

[25] C. H. Wang, Y. Y. Pu, D. Xu, and J. Zhu, "Evaluating aesthetics quality in portrait photos," *J. Softw.*, vol. 26, no. 2, pp. 20–28, 2015.

[26] H.-H. Su, T.-W. Chen, C.-C. Kao, W. H. Hsu, and S.-Y. Chien, "Preference-aware view recommendation system for scenic photos based on Bag-of-Aesthetics-Preserving features," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 833–843, Jun. 2012.

[27] P. Obrador, M. A. Saad, P. Suryanarayan, and N. Oliver, "Towards category-based aesthetic models of photographs," in *Proc. Conf. Multimedia Model. (MMM)*. Cham, Switzerland: Springer, 2012, pp. 63–76.

[28] T. O. Aydin, A. Smolic, and M. Gross, "Automated aesthetic analysis of photographic images," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 1, pp. 31–42, Jan. 2015.

[29] J. J. Zhang, R. Peng, J. Wang, and J. H. Yu, "Computational aesthetic evaluation of chinese wash paintings," *J. Softw.*, vol. 27, no. S2, pp. 220–233, 2017.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Proces. Syst.*, 2012, pp. 1097–1105.

[31] O. V. J. Donahue and Y. G. Jia, "Decaf: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn. (ICML)*. New York, NY, USA, 2014, pp. 647–655.

[32] M. L. Shi, *Modern Chinese Ink Paintings*. Beijing, China: China Friendship Publishing Company, 2018.

[33] S. M. Wu, Z. F. Li, and J. L. Pan, *A Careful Study of Famous Works by Chinese Painters: Selection Works by Modern Masters*, 1st ed. Zhejiang, China: Zhejiang Photography Press, 2018.

[34] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Proces. Syst.*, 2014, pp. 3320–3328.

[35] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Washington, DC, USA, Dec. 2015, pp. 990–998.

[36] J. Li, L. Yao, E. Hendriks, and J. Z. Wang, "Rhythmic brushstrokes distinguish van gogh from his contemporaries: Findings via automated brushstroke extraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1159–1176, Jun. 2012.

[37] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[38] S. Thumfart, R. H. A. H. Jacobs, E. Lughofer, C. Eitzinger, F. W. Cornelissen, W. Groissboeck, and R. Richter, "Modeling human aesthetic perception of visual textures," *ACM Trans. Appl. Perception*, vol. 8, no. 4, pp. 1–29, Nov. 2011.

[39] X. Lu, P. Suryanarayan, R. B. Adams, J. Li, M. G. Newman, and J. Z. Wang, "On shape and the computability of emotions," in *Proc. 20th ACM Int. Conf. Multimedia*, New York, NY, USA, 2012, pp. 229–238.

[40] M. Sun, D. Zhang, Z. Wang, J. Ren, and J. S. Jin, "Monte Carlo convex hull model for classification of traditional chinese paintings," *Neurocomputing*, vol. 171, pp. 788–797, Jan. 2016.

[41] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, USA, Jun. 2012, pp. 733–740.

[42] Y. Jin, Q. Wu, and L. Liu, "Aesthetic photo composition by optimal crop-and-warp," *Comput. Graph.*, vol. 36, no. 8, pp. 955–965, Dec. 2012.

[43] J. C. M. Abadi and P. Barham, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Oper. Syst. Des. Implement.*, 2016, pp. 265–283.

JIAJING ZHANG was born in Hebei, China, in 1991. She received the B.S. degree in computer science and technology from the South China University of Technology, Guangzhou, Guangdong, China, in 2012, and the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, Zhejiang, China, in 2017.

Since 2018, she has been a Lecturer with the Information Science and Technology Department, Zhejiang Sci-Tech University, Hangzhou, Zhejiang, China. Her current research interests include image computational aesthetic assessment, visual media computing, image processing, and deep learning.

YONGWEI MIAO (Member, IEEE) received the master's degree in mathematics from the Institute of Mathematics, Chinese Academy of Sciences, Beijing, in July 1996, and the Ph.D. degree in computer graphics from the State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, in March 2007. From February 2008 to February 2009, he worked as a Visiting Scholar at the University of Zurich, Switzerland. From November 2011 to May 2012, he worked as a Visiting Scholar at the University of Maryland, USA. From July 2015 to August 2015, he worked as a Visiting Professor at the University of Tokyo, Japan. He is currently a Professor with the College of Information Science and Technology, Zhejiang Sci-Tech University, China. His research interests include computer graphics, 3D computer vision, 3D reconstruction, visual media computing, and deep learning. He is the author or coauthor of more than 130 technical papers published in scientific journals or presented at conferences.

JUNSONG ZHANG received the Ph.D. degree from the Department of Computer Science and Technology, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, in 2008.

Since 2010, he has been an Associate Professor with the Mind, Art and Computation Group, Cognitive Science Department, Xiamen University, Xiamen, Fujian. His main research interests include expressive rendering, human–computer interaction, Chinese information processing, and brain and cognitive science.

Dr. Zhang is a Member of the Professional Committee in Image Analysis and Recognition of CSIG, the China Artificial Intelligence Society in Intelligent Creativity and Digital Art Major, and the Mobile Media and Cultural Computing of China Communication Society. Besides, he was a recipient of the China CAD&CG Best Paper Award, in 2016.

JINHUI YU received the B.S. and M.S. degrees from the Department of Electronic Engineering, Harbin Institute of Marine Engineering, Heilongjiang, in 1987, and the Ph.D. degree from the Department of Computer Science and Technology, University of Glasgow, U.K., in 1999.

Since 2001, he has been a Professor with State Key Lab of CAD&CG, and a Doctoral Supervisor of Computer Application Technology and Digital Art with Zhejiang University, Hangzhou, Zhejiang. His research interests include computational aesthetics of traditional Chinese art, non-photorealistic rendering, and computer animation.

Dr. Yu is a Member of ACM SIGGRAPH, the Professional Committee of Computer Animation and Digital Entertainment in China Image Graphic Society, Digital Art, and China Animation Society.