

ARTICLE

Studies on the Extraction and Recognition of Aesthetic Features Using Comprehensive material painting Images

Jincheng Yang^{1,*} and Haiyun Yang¹

¹ School of Fine Arts and Design, Hunan First Normal University

Abstract

To develop the visual language of animated films, this study examines the process of identifying and extracting aesthetic elements from comprehensive material painting images. The study demonstrates that, due to the diversity of materials and the episodic nature of its development, comprehensive material painting offers a distinctive visual impact and artistic expression. A graph convolutional network-based layout perception method is proposed to address the classic convolutional neural network's limitation in capturing an image's overall layout information when assessing aesthetics. A novel adjacency matrix is constructed by incorporating dependencies between spatial distance information and regional features, enabling comprehensive perception of image layout. Experimental results show that the proposed LARC-CNN model significantly improves the accuracy of aesthetic quality binary classification, aesthetic score regression, and aesthetic score distribution prediction. It also performs well in evaluating aesthetic quality on the AVA dataset. Furthermore, by analyzing the works of artists like Gu Wenda and Hu Wei, the research highlights the diversity and creativity of ink fused with other materials, exploring the use of ink elements in modern mixed-media paintings. Through the appropriation and recombination of materials, these mixed-media paintings not only retain the essence of traditional ink and wash painting but also enhance the visual and artistic impact of the artwork.

Submitted: 09 August 2023

Accepted: 10 December 2023

Published: 30 December 2023

Vol. 2023, No. 1, 2023.

*Corresponding author:

✉ Jincheng Yang

yangjincheng202308@163.com

Keywords: comprehensive material paintings, aesthetic characteristics, graph convolution networks, animated films, ink elements, visual art, layout perception

Citation

Jincheng Yang and Haiyun Yang (2023). Studies on the Extraction and Recognition of Aesthetic Features Using Comprehensive material painting Images . Mari Papel Y Corrugado, 2023(1), 88–95.

© The authors. <https://creativecommons.org/licenses/by/4.0/>.

1 Introduction

Due to its distinctive artistic expression and rich material vocabulary, mixed-media painting has steadily gained prominence as a contemporary art genre [1]. Unlike traditional painting, mixed-media painting expands the possibilities for artistic expression and offers the viewer a new visual experience by integrating multiple media and using innovative techniques. Mixed-media painting holds a unique place within the framework of modern art because of its originality and diversity, especially in today's world where multiculturalism and globalization are intertwined [2]. However, a new challenge and area of research lies in extracting the aesthetic elements of this art form and applying them to other artistic mediums, such as the visual language of animated films.

Mixed-media paintings are characterized by the fusion of diverse materials, including acrylic, emulsion, wormwood, metal, plaster, and more, resulting in artworks with layered expression and a powerful visual impact [3, 4]. For example, as shown in Figures 1 and 2, Kiefer's "Serpent of Paradise" and "Ancient Woman" use materials like sand, clay, and steel to create a harsh and cold visual effect, while simultaneously incorporating warm, soft elements like raindrops and a little girl's skirt to form a striking visual contrast. This versatile creative approach has informed the development of visual language in animated films.



Figure 1. Kiefer, Serpent of Paradise, cloth panel, 190 x 280 x 42 cm, emulsion, shellac, metal, gesso, lead, and gold leaf



Figure 3. Peter Watts' animated short film, "The Rabbit and the Deer"



Figure 4. Peter Watts's animated short film, The Rabbit and the Deer



Figure 2. Kiefer, Ancient Women

The visual impact and experience of animated films are largely shaped by their image language, a key component of audiovisual art. A thorough investigation into incorporating the artistic qualities of mixed-media painting into the visual language of animated films is essential to preserve the work's theme and content while enhancing its visual quality [5, 6]. For instance, Hungarian animation director Péter Vác's 2013 animated short film *Rabbit and Deer* (shown in Figures 3 and 4) utilizes techniques such as "insertion," "collage," and "combination" from mixed-media painting to alternate between two- and three-dimensional worlds. These techniques effectively convey the characters' struggles and conflicts with their own identities and the outside world, contributing to the animation's artistic depth and visual appeal.

While mixed-media painting has been widely used in the visual arts, significant challenges remain in successfully extracting its aesthetic qualities for application in animated films. First, the material language of mixed-media paintings is highly variable and episodic, requiring further research to determine how these qualities can be represented in animated films while achieving a balanced integration. Secondly, the visual elements of mixed-media paintings must

harmonize closely with the plot and composition of animated films [7, 8]. Additionally, current deep learning-based techniques for evaluating image aesthetics primarily rely on convolutional operations, which often struggle to capture an image’s overall layout information, impacting the accuracy of aesthetic quality evaluation.

Recent advances have been made in deep learning-based aesthetic evaluation techniques. For example, the A-Lamp model proposed in [9] encodes the global compositional information of an image by selecting multiple image blocks from the original image using object detection and creating attribute-relationship maps of these regions. However, this approach has limitations in layout perception, particularly for images lacking prominent objects. Furthermore, many systems apply stretching or cropping when processing images, which can distort the original layout and reduce aesthetic quality [10].

To address these limitations, this study proposes a graph convolutional network-based layout perception technique that achieves a comprehensive perception of image layout by modeling the interdependencies among local parts of an image. Specifically, this study uses a graph convolutional network for learning and reasoning to represent an image’s layout information as dependencies between its local regions, enhancing the model’s capacity for aesthetic learning [11]. Based on this, a new model called LARC-CNN is developed for aesthetic quality evaluation and validated using the AVA dataset. Experimental results show that the model performs exceptionally well in binary aesthetic quality classification, aesthetic score regression, and aesthetic score distribution prediction tasks.

2 Graph Convolution-Based Perception of Layout

An image’s overall layout significantly impacts its aesthetic quality. The approach presented in this chapter aims to enhance aesthetic learning and improve the performance of aesthetic quality evaluation by enabling the perception of an image’s layout information. Specifically, this chapter models an image’s layout information as dependencies among its local regions [12]. Graph convolutional networks, which are well-suited for learning complex relationships between image regions, are employed to perform dependency inference, facilitating layout perception. Conventional convolution operations in convolutional neural networks are often inefficient in capturing relationships between regions that are



Figure 5. Graph-convolution layout-aware subnetwork’s specific structure

distant in coordinate space.

Figure 5 illustrates the detailed architecture of the graph-convolution layout-aware subnetwork. The fully convolutional network (FCN) is used as the backbone network to encode the regional features of an input image, which is stretched and preprocessed to a fixed size. This process results in a feature map with dimensions $W \times H \times D$. Because the FCN encoding process preserves the spatial information of the original image, the feature vectors at each spatial location can be stacked into a matrix based on the obtained feature map to produce the region feature representation $F = [f_1, f_2, \dots, f_N] \in \mathbb{R}^{N \times D}$, where D denotes the dimensionality of the features and $N = W \times H$ indicates the number of region features.

Next, using this region feature representation, a fully connected graph $\text{Graph} = (V, E)$ is created to model the dependencies between regions. The nodes in the graph, representing the respective region features, are numbered N , with $V = \{f_1, f_2, \dots, f_N\}$. Then, using the similarity function $s(f_i, f_j)$, the semantic content similarity between any two nodes is computed, assigning a weighted edge between nodes to characterize the dependencies between regions, where $1 \leq i, j \leq N$. This chapter measures feature similarity $s(f_i, f_j)$ using the vector inner product, and defines the similarity function as follows:

$$s(f_i, f_j) = \varphi(f_i)^T \psi(f_j)^I, \quad (1)$$

where $\varphi(f_i) = Af_i, \psi(f_j) = Bf_j, A, B \in \square^{(D \times D)}$ is a learnable parameter that may be enhanced via back propagation, and the model’s generalization can be strengthened by linearly transforming the eigenvectors [13, 14]. The matrix $\text{Sim} \in \square^{(N \times N)}$ is produced once the weights between each pair of nodes have

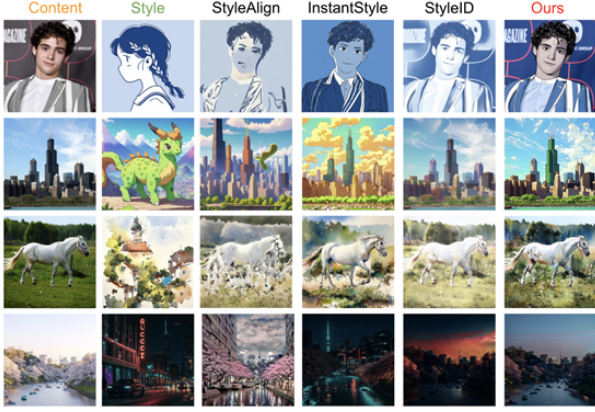


Figure 6. Schematic representation showing how the original image's spatial composition information was reduced

been calculated. The Softmax function is then used to normalize each column element in the Sim matrix in accordance with the work in the literature [15]:

$$Sim_{i,j} = \frac{\exp(s(f_i, f_j))}{\sum_{k=1}^N \exp(s(f_k, f_j))}. \quad (2)$$

The exponential function is represented by $\exp(\cdot)$. After normalization using the Softmax function, the total weight of the edges connecting to each node in the network Graph is 1. This normalized graph is represented by the matrix Sim . The dependencies among the nodes in the graph are described by the adjacency matrix of Graph.

To capture semantic content similarity across different areas, each element in the Sim matrix calculates similarity between region features using an inner product, thus completing the modeling of inter-region relationships. However, beyond dependencies based on semantic content similarity, spatial composition information of visual elements is also a critical feature that affects the layout and aesthetics of an image [16]. As previously mentioned, the method proposed in this chapter, like most current methods, preprocesses the input image to a fixed size before the FCN feature encoding operation. This preprocessing uses a stretching operation, which helps preserve the image layout to some extent but inevitably alters the spatial composition of the image's visual elements and changes the spatial distances between different regions, thereby affecting the image's original layout.

After applying the FCN convolution operation to the preprocessed image, Figure 6 illustrates the spatial relationships between different regions in the regional feature representation F . The spatial distances

between the nodes, which define the local regions, are as follows:

$$d'(f_i, f_j) = \sqrt{x_{i,j}^2 + y_{i,j}^2}, \quad (3)$$

where $x_{i,j} \in \{0, 1, \dots, W-1\}$ and $y_{i,j} \in \{0, 1, \dots, H-1\}$ represent the horizontal and vertical distances, respectively, between nodes f_i and f_j . Since the CNN model typically accepts square data as input, $W = H$ in general. The graphic clearly shows that the stretching preprocessing introduces a discrepancy between this spatial relationship and the spatial relationship in the original image. To address this issue, this chapter adjusts the spatial distances between regions using the aspect ratio information from the original image. This spatial information is then embedded into the region relation map Graph to preserve as much of the image's original structure as possible. Letting h and w represent the height and width of the original image, respectively, and $Aspect_ratio = \frac{w}{h}$, the original spatial distance between the nodes is:

$$d(f_i, f_j) = \begin{cases} \sqrt{(x_{i,j} \times Aspect_r)^2 + y_{i,j}^2}, & Aspect_r \leq 1 \\ \sqrt{x_{i,j}^2 + \left(\frac{y_{i,j}}{Aspect_r}\right)^2}, & Aspect_r > 1 \end{cases}. \quad (4)$$

The matrix $Dis \in \square^{N \times N}$ characterizing the original spatial distance relationship between nodes can be obtained through Eq. 4. Similarly, each column element in the D matrix is normalized using the Softmax function:

$$Dis_{i,j} = \frac{\exp(d(f_i, f_j))}{\sum_{k=1}^N \exp(d(f_k, f_j))}. \quad (5)$$

The original aspect ratio is preserved by integrating the spatial composition information into the region dependency map $Graph$ to create the new adjacency matrix $A \in \square^{N \times N}$.

$$A = Sim + Dis. \quad (6)$$

At this stage, the weights of the edges connecting nodes in the graph Graph incorporate both the spatial distance relationships between them and the semantic content similarity between areas, aligning with the original image structure. This process completes the construction of the dependency graph. As previously mentioned, Graph is a fully connected graph, and

Method	ACC	SRCC	PLCC	RMSE	EMD
RAPID	74.45	-	-	-	-
DMA-Net	75.52	-	-	-	-
MNA-CNN	78.89	-	-	-	-
MTRLCNN	78.85	-	-	-	-
A-Lamp	82.24	-	-	-	-
NIMA	81.16	0.625	0.634	-	0.05
MP _{ada}	82.32	-	-	-	-
GPF-CNN	80.56	0.674	0.532	0.532	0.045
ILGNet	82.55	-	-	-	-
PA_IAA	82.88	0.067	-	-	0.045
AFDC	83.24	0.65	-	0.521	-
LARC-CNN	83.26	0.712	0.725	0.508	0.045

Table 1. A comparison of the AVA dataset with various current methodologies

Figure 6 illustrates only partial connections to highlight the strong relationships between nodes for clarity.

3 Experimental Results and Analysis

To further verify the effectiveness and superiority of the proposed LARC-CNN model, this subsection quantitatively compares it with several current state-of-the-art approaches. The model includes three graph convolutional layers, with the number of loop connections T set to two, and the neural layers with additional loop connections in the set R are set to $\{\text{conv5}_2, \text{conv4}_2, \text{conv3}_2\}$. The results of the comparison experiments are shown in Table 1. These parameter choices are informed by the analysis and experimental findings presented in subsection 2.

The LARC-CNN model presented in this chapter achieves competitive results in three key areas: binary classification of aesthetic quality, regression of aesthetic scores, and prediction of aesthetic score distributions, as shown by comparing the aesthetic quality evaluation models in Table 1. Unlike image block input-based methods such as RAPID, DMA-Net, MNA-CNN, A-Lamp, MP_{ada}, and GPF-CNN, the proposed network framework directly learns discriminative features from the entire image, bypassing the need for labor-intensive image block sampling and result aggregation. This approach leads to higher prediction accuracies.

Specifically, the A-Lamp method encodes the spatial layout of an image using a manually designed aggregation structure and a complex image block sampling strategy. Its layout perception relies on learning the relationships between prominent regions, which must be specified in advance and affects

performance based on the extraction of these regions. In contrast, the LARC-CNN proposed in this chapter models the interdependence of visual elements at the feature level, achieving automatic end-to-end layout perception and aesthetic prediction via graph convolution. This approach improves the accuracy of binary classification of aesthetic quality by 1.18% over the A-Lamp method.

One MNA-CNN approach uses Spatial Pyramid Pooling (SPP) [17] to process images at their original size and aspect ratio, recognizing that preprocessing images through stretching, cropping, or padding distorts their composition. While effective, this approach does not explicitly model visual elements or their interdependencies as thoroughly as the approach in this chapter, which improves the LARC-CNN model’s accuracy by 6.28%. This method aims to preserve compositional information in an aesthetic model.

For the aesthetic score regression task, Pool-3FC method performed best on SRCC and PLCC by directly regressing on the mean aesthetic score, although this limits prediction accuracy by omitting variance and maximum values in the score distribution. However, Pool-3FC’s use of raw resolution images with preserved aspect ratios and layouts significantly slows model training and inference.

The LARC-CNN model outperforms other compatible batch-trained methods, including MTRLCNN, NIMA, ILGNet, PA_IAA, and ADFC, across all three aspects of the aesthetic quality evaluation task. It achieves improvements in binary prediction of aesthetic quality by 4.60%, 2.17%, 1.02%, and 0.78%, respectively, and further improves accuracy by 0.44%, 0.85%, and 2.78% in other metrics. In conclusion, the proposed LARC-CNN model effectively perceives image layout and performs efficient aesthetic learning, yielding more accurate assessments of aesthetic quality across three dimensions. Sample images from the AVA test set, along with their model-predicted results, are displayed in Figure 7.

4 Expression of Ink Elements in Composite Material Painting

Chinese ink and wash elements are frequently incorporated into contemporary composite material paintings, where the fusion of ink with various materials serves as a primary means of expression. Gu Wenda’s *The Lost Dynasty - The Modern Meaning of Totems and Taboos* is a notable example of

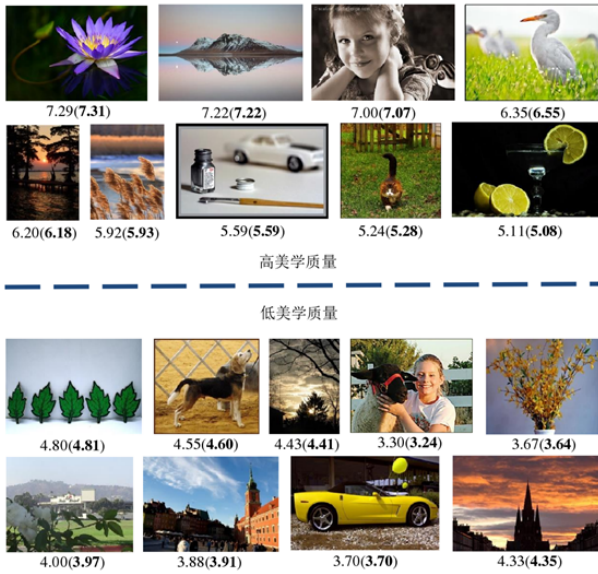


Figure 7. Shows a few sample photos taken from the test set together with the matching aesthetic quality forecasts. The genuine aesthetic score averages of the photographs are indicated by the values in parenthesis



Figure 8. The Lost Dynasty by Ku Wenda: The Contemporary Importance of Taboos and Totems

mixed-media painting, combining multiple materials with innovative ink and wash techniques as its central feature. Instead of using traditional Chinese ink techniques such as "hook, strangulation, chafing, and dotting" to convey ideas, this work utilizes ink as the primary medium, supplemented by acrylic and gouache. Additionally, the piece interprets historical divergence from reality through the use of intentionally misspelled words on a large-format newspaper (Figure 8).

For example, to enhance the artistic appeal of his paintings, Mr. Hu Wei incorporated additional materials alongside ink and wash as the primary medium in his composite material artworks. This approach resulted in the piece titled *Impression of Sima* (Figure 9).



Figure 9. Hu Wei, Impression of Sima



Figure 10. Chen Shouyi, Landscape

The *Late Tang* ink painting technique is applied using various types of paper and mineral pigments to create the piece *Sima Impressions*, lending the artwork a sense of "harmony" through this blend of materials. Additionally, the natural bond between ink and mineral pigments in these paintings provides viewers with an atmospheric experience, enhancing their appreciation of the unique qualities of ink. For example, in Chen Shouyi's piece *Landscape*, rolling river waves are accentuated using rice paper and ink as primary media, combined with additional materials. The ink collage, with various textures from inked rice paper, clay paper, and watercolor paper, enriches the natural expression of the scene, better reflecting the image's texture (Figure 10).

One of the primary ways Chinese ink elements are expressed in mixed-media paintings is through the appropriation of ink imagery. For instance, the use of traditional motifs, such as fish, water, pavilions, and insects, can enhance the artistic quality of mixed-media paintings. For example, in Qiu Deshu's *Fission - Soul Series*, the visual appeal is elevated by incorporating Chinese ink and wash elements (Figure 11).

This mixed-media artwork was created by tearing rice paper collage to depict landscapes such as mountains and rivers. Unlike traditional works, the outline of the mountains is formed using shredded paper rather than ink and brush. At the same time, the dots and washes in the pictorial style are crafted with acrylic paint, a relatively modern technique that draws the



Figure 11. Chou Teh-shu "Fission - Soul Series"



Figure 12. Sang Yang The Dong Qichang Project 2

viewer's attention [18, 19]. For example, Shang Yang's mixed-media painting *Dong Qichang Project 2* directly appropriates traditional landscape imagery without using ink and wash materials (Figure 12).

This piece appropriates the traditional Chinese ink motif of the "mountain" image. However, in terms of materials, the airbrush technique combined with oil and acrylic paints enhances the depth of the scene, allowing for a more fluid representation of the near, middle, and distant views of the rolling hills and rivers. An example of ink and watercolor appropriation can be seen in Chen Junhao's replica of Fan Kuan's *Traveling in the Mountains* from the Northern Song Dynasty, created with mosquito nails.

5 Conclusion

This study presents a method for assessing the aesthetic quality of images based on graph convolution networks and thoroughly discusses the aesthetic aspects of composite material painting and its use in the visual language of animated films. We summarize its distinctive aesthetic qualities in visual art—material diversity, visual contrast, texture richness, etc.—through the analysis and case study of material language in comprehensive material painting, offering a crucial resource for the development of

animation film picture language. In order to obtain a thorough understanding of the image layout, we present a novel approach to layout perception in this paper that uses graph convolutional networks to describe the local inter-region dependencies of an image. We developed the LARC-CNN model, which assesses aesthetic quality, using this methodology, and it performed exceptionally well in experimental validation using the AVA dataset.

Fund support

This paper is a phased research achievement of the project "Construction and Practice of Contemporary Comprehensive Material Painting Aesthetic System," funded by the Humanities and Social Sciences Research Planning Fund of the Ministry of Education [project approval number: 23YJA760104].

References

- [1] Li, B., & Lu, W. (2024). Application of image processing technology in the digital media era in the design of integrated materials painting in installation art. *Multimedia Tools and Applications*, 83(18), 54211-54228.
- [2] Wu, B., Dong, Q., & Sun, W. (2023). Automatic colorization of Chinese ink painting combining multi-level features and generative adversarial networks. *Fractals*, 31(06), 2340144.
- [3] El Kadiri Boutchich, D. (2024). Painting art and sustainability: relationship from composite indices and a neural network. *International Journal of Social Economics*, 51(1), 46-61.
- [4] Huang, T., Zhou, C., Wang, X., & Kaner, J. (2023). A Study of Visual Perception Based on Colour and Texture of Reconstituted Decorative Veneer. *Coatings*, 14(1), 57.
- [5] Liu, Y., Wang, F., Jiang, Z., Sfarra, S., Liu, K., & Yao, Y. (2023). Generative deep learning-based thermographic inspection of artwork. *Sensors*, 23(14), 6362.
- [6] Hu, Q., Zhou, W., Peng, X., Zhang, X., Xie, P., Liu, Y., ... & Fan, J. (2024). DRANet: A semantic segmentation network for Chinese landscape paintings. *Digital Signal Processing*, 147, 104427.
- [7] Zhang, R. (2023). Art Portrait Design Based on Mobile Internet in Digital Media. *Journal of Multimedia Information System*, 10(1), 25-34.
- [8] Tian, W. (2023). AI-based artist style appreciation: folk art in the central plains oriented implementation platform. *International Journal of Computer Applications in Technology*, 71(3), 181-191.
- [9] Liang, T., Xiao, Z., & Guo, L. (2024). A Deep Learning Model-Based Feature Extraction Method for Architectural Space and Color Connection in Interior Design. *Scalable Computing: Practice and Experience*, 25(4), 2948-2959.
- [10] Calovi, M., & Rossi, S. (2024). Exploring polyamide 11 as a novel renewable resource-based filler in wood paint: Investigating aesthetic aspects and durability impact of the composite coating. *Progress in Organic Coatings*, 188, 108262.
- [11] Cui, Y. (2024). Application of cultural elements of dunhuang murals in landscape design based on mean shift algorithm extraction. *Journal of Computational Methods in Sciences and Engineering*, 24(1), 473-487.
- [12] Guo, M., Xiong, M., Huang, J., Hu, X., & Peng, T. (2024). Face photo-sketch portraits transformation via generation pipeline. *The Visual Computer*, 1-14.
- [13] Sarikahya, E. (2023). The structural formation and expression of morphological and biomorphic composition in geometric abstraction. *Premium e-Journal of Social Sciences (PEJOSS)*, 7(28), 210-221.
- [14] Peng, B., & Sirisuk, M. (2024). Exploration of CAD and Neural Network Integration in Art Design and Cultural Heritage Protection. *Computer-Aided Design*, 21(18), 128-144.
- [15] Menon, S. (2024). Research on Various Plastic Arts in Interior Decoration Design Based on Big Data under the Environment of Sustainable Development. In *Big Data Computing* (pp. 298-310). CRC Press.
- [16] Zhao, M., Qian, X., & Song, W. (2024). MA-GAN: the style transfer model based on multi-adaptive generative adversarial networks. *Journal of Electronic Imaging*, 33(3), 033017-033017.
- [17] Li, P., & Khan, J. (2023). Feature extraction and analysis of landscape imaging using drones and machine vision. *Soft Computing*, 27(24), 18529-18547.
- [18] Dong, Q., & Liu, X. (2023). Optimization practice of university innovation and entrepreneurship education based on the perspective of OBE. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 118, 181-189.
- [19] Liao, Q. (2023). English Teaching Project Quality Evaluation Based on Deep Decision-Making and Rule Association Analysis. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 118, 119-127.