Proceedings of the 4th International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2026.38450

Emotional Perception Differences in the Digital Representation of Watercolor Diffusion Traces

Siping Zhang

College of Fine Arts and Calligraphy, Xihua University, Chengdu, China
wellington589125@gmail.com

Abstract. Digital image processing has made changes to the way of preserving and presenting traditional watercolor paintings. Diffusion trace effects, such as edge bleeding, pigment deposition, color gradation, and irregular texture caused by watercolor diffusion might alter during the process of digital conversion and affect viewers' emotional response to a certain degree. Twenty watercolor images having distinct diffusion traces were chosen to form four types of visual stimuli: original images, smoothed images, texture enhanced images, and digitally reconstructed images. Ratings for valence and arousal were acquired through SAM. Color variation, local entropy, edge density, texture contrast, and diffusion area were extracted through OpenCV, while deep visual features were obtained through ResNet50. Then, XGBoost and SHAP were performed for emotion prediction and explanation of contribution. It is revealed that the average arousal of smoothed images decreased from 5.31 ± 1.16 to 4.80 ± 0.99, while the arousal of texture enhanced images increased to 5.92 ± 1.12. As for digitally reconstructed images, their arousal was 5.37 ± 1.19. The MAE, RMSE, and R^2 for fused feature model was 0.60, 0.80, and 0.64. Local entropy and edge density made significant contributions, which implies that diffusion textures and edge structure are highly correlated with emotional arousal in digital watercolor. These conclusions provide a quantitative base for the assessment of digital watercolor, computational art study, and affective image processing.

Keywords: watercolor diffusion, emotional perception, digital representation, computer vision, affective computing

1. Introduction

Watercolor diffusion results in irregular edges, pigment deposits, gradient colors, and varied textures which significantly impact visual perception. With the help of computational watercolors, the physical properties of watercolor diffusions have been simulated in digital environments and with image processing methods [1]. With advancements in digital pigment models, the simulation of color interactions and material appearances has been further enhanced [2]. However, operations like scanning, smoothing, enhancing, and reconstructing can change the diffusion structures and affect emotional perception. In current studies, there is an assessment of digital watercolor based on visual appearance and rendering quality while there are unclear understandings on how the structure change affects emotional perception. The analysis attempts to investigate how different types of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.38450

watercolor diffusion traces affect valence and arousal. Digital watercolor images are converted to original, smoothed, texture enhanced, and digitally reconstructed formats while OpenCV features, ResNet50 representation, subjective affective ratings, XGBoost regression, and SHAP explanations are used in order to determine the visual elements affecting emotional perception differences.

2. Literature review

2.1. Digital representation of watercolor traces

The computational representation of watercolor includes preserving pigment diffusion, edge bleeding, texture of the paper, color blending, and pigment concentration locally. It is thus helpful to use perceptually based evaluation procedures to analyze whether the images that have been visually transformed can evoke similar emotions compared to being formally similar [3]. It is also possible to transform color, texture, shapes, and spatial properties into quantifiable data using computational image analysis and then measure changes caused by the processing of the images digitally [4]. This implies that different techniques will give rise to varying representations of the same watercolor trace. Smoothing reduces the local variations, enhancement highlights the textures, while digital reconstruction creates diffusions digitally.

2.2. Emotion recognition in artistic images

In recognizing emotions in artistic images, interpretable visual qualities and machine learning have been found to work together effectively. From the principles of visual art, it has been found out that different factors such as color distribution, texture arrangement, shapes, and composition affect the emotions of the individual and therefore allow visual properties of the image to predict emotion [5]. There exist large databases containing emotions in artistic images which show that visual information can be correlated with subjective emotions reported by the viewers [6]. Advances in affective image analysis reveal that deep representation can supplement hand-crafted features whenever emotional information is dispersed throughout complex visual information [7]. The watercolor diffusion gives rise to an environment where each of the following properties can be analyzed independently within the entire visual context.

2.3. Perceptual evaluation of digital art

Affective evaluation of digital artwork is growing to include not only technical assessment but also emotional experience along with conventional evaluations of image quality. Emotional response to visual media has been found to involve the interplay of measurable aspects of image features and human interpretation. Thus, affective evaluation is necessary in order to study the perception of technically altered images [8]. Cross-discipline investigations of emotion recognition in art have also shown that affective models based on everyday visual images may not recognize artistically expressed emotions adequately since art images have their unique features [9]. Watercolor diffusion becomes an appropriate setting for studying this question, because smoothing, texture, and diffusion modeling may alter selectively some particular visual structures, leaving the basic information intact. Thus, comparisons of subjective ratings will reveal how different perceptions correspond to computationally measurable alterations.

DOI: 10.54254/2753-8818/2026.38450

3. Experimental methods

3.1. Watercolor image dataset construction

The source images were acquired from the Watercolor section of WikiArt and its read only public API. The years of creation were restricted to 1900 to 2020 in order to minimize the variance in scanning quality and appearance of materials used during different eras. The first selection consisted of 120 watercolor paintings that had a resolution of at least 1024 pixels. Images were filtered based on three factors: clear diffusion boundaries, full local depositing of pigments, and no visible watermarks or heavy compression artifacts. In the end, 20 images with representative diffusion traces were selected. A constant size area was cropped from each painting containing the main diffusion pattern. Using the Python programming language and OpenCV package, four versions of an image were produced for each original image: the original image, Gaussian blurred image, CLAHE texture enhanced image, and digital reconstruction of diffusion pattern resulting in 80 stimuli altogether. All images were resized to 512 x 512 pixels and calibrated in terms of mean luminance and color range in the CIELAB color space. Next, valence and arousal were measured by 60 subjects using the nine point SAM scale.

3.2. Visual feature extraction

OpenCV was first used to extract color variance, local entropy, edge density, texture contrast, and diffusion area from the 80 stimuli [10]. Color variance measured the dispersion of pigment colors, as shown in Equation 1.

$$\sigma_c^2=\frac{1}{N}\sum_{i=1}^{N}(I_{i,c}-\mu_c)^2 \tag{1}$$

Here, $I_{i,c}$ represents the value of pixel i in color channel c, μ_c represents the mean value of that channel, and N represents the number of pixels.

The local complexity of diffusion textures was measured using information entropy, as shown in Equation 2.

$$H=\sum_{k=1}^{K}p_k\log_2 p_k \tag{2}$$

Here, p_k represents the probability of gray level k within a local window, while K represents the total number of gray levels. The images were then input into an ImageNet pretrained ResNet50. The global average pooling layer was used to extract 2048 dimensional deep features, followed by Z Score normalization. Low level features captured interpretable differences in diffusion morphology, while ResNet50 features represented complex relationships among color, texture, and spatial structure. The two feature groups were stored separately and then fused for subsequent emotion prediction.

DOI: 10.54254/2753-8818/2026.38450

3.3. Emotion prediction modeling

The visual features of each image were matched with the mean SAM ratings provided by the 60 participants. Separate XGBoost regression models were developed for valence and arousal. Five fold cross validation was applied, while identical data partitions were maintained when comparing handcrafted features, ResNet50 features, and fused features [11]. The optimization objective of XGBoost at iteration tt is shown in Equation 3.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{3}$$

Here, y_i represents the actual SAM rating, $\hat{y}_i^{(t-1)}$ represents the predicted value, f_t represents the newly added decision tree, and Ω represents the complexity penalty. SHAP was then applied to explain the marginal contribution of individual visual variables to model predictions, as shown in Equation 4.

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} [f(S \cup \{j\}) - f(S)] \tag{4}$$

Here, ϕ_j represents the SHAP value of feature j, S represents a feature subset excluding feature j, and M represents the total number of features. This analysis identifies the direction and magnitude of the contributions of color, edges, textures, and deep features to the two emotional dimensions.

4. Results

4.1. Emotional differences across digital representations

As evident from Table 1, the illustrative statistical outcomes revealed relatively minor variation in valence scores between the four types of digital stimuli. Repeated measure yielded $F(3,177)=1.40$, $p=0.244$, which demonstrates the weak effect of digital manipulation on the pleasantness level of images. The presence of a significant condition effect was detected for arousal, where $F(3,177)=41.77$, $p<0.001$, and partial eta squared equals to 0.414. Emotional activation level dropped from 5.31 ± 1.16 in the original condition to 4.80 ± 0.99 in the smooth condition, where the paired effect size $d=0.65$. It can be stated that smoothing out local textures and diffusion boundaries reduces the level of emotional activation. The texture enhancement raised the arousal level to 5.92 ± 1.12 (effect size $d=0.79$), which means that increasing complexity in texture leads to higher levels of activation. Reconstruction of digital image brought the level of arousal to 5.37 ± 1.19.

Table 1. Illustrative SAM emotion ratings across digital representation conditions

Digital Representation	Valence Mean ± SD	Arousal Mean ± SD	Arousal Compared with Original	Cohen's d
Original	5.72 ± 1.06	5.31 ± 1.16	Reference	0.00
Smoothed	5.69 ± 1.15	4.80 ± 0.99	$p < 0.001$	-0.65
Texture enhanced	5.54 ± 0.98	5.92 ± 1.12	$p < 0.001$	0.79

DOI: 10.54254/2753-8818/2026.38450

Table 1. (continued)

Reconstructed	5.71 ± 1.17	5.37 ± 1.19	p = 0.587	0.07

4.2. Prediction performance and feature contributions

From the illustrative model results, the performance of fused features proved the best in prediction of arousal where the MAE stood at 0.60, RMSE of 0.80 and R^2=0.64. Such a high performance was better than those achieved by handcrafted features of 0.71, 0.94 and 0.48 respectively and those by ResNet50 features of 0.67, 0.89 and 0.55 respectively. In the prediction of valence, the model attained an MAE of 0.65, RMSE of 0.86 and R^2=0.58. The above findings imply that emotional activation offered more stable cues from diffusion textures compared to perceived pleasantness. As illustrated in Figure 1, local entropy and edge density had the widest SHAP distributions in the arousal model with mean absolute contribution of about 0.29 and 0.24 respectively. The main ResNet50 component had a mean absolute contribution of about 0.22. Color variance, texture contrast and diffusion area followed sequentially with decreasing mean absolute contribution. The existence of positive and negative SHAP values indicated varying effects of visual features on different diffusion conditions. The decrease in texture complexity after smoothing matched low arousal rating.

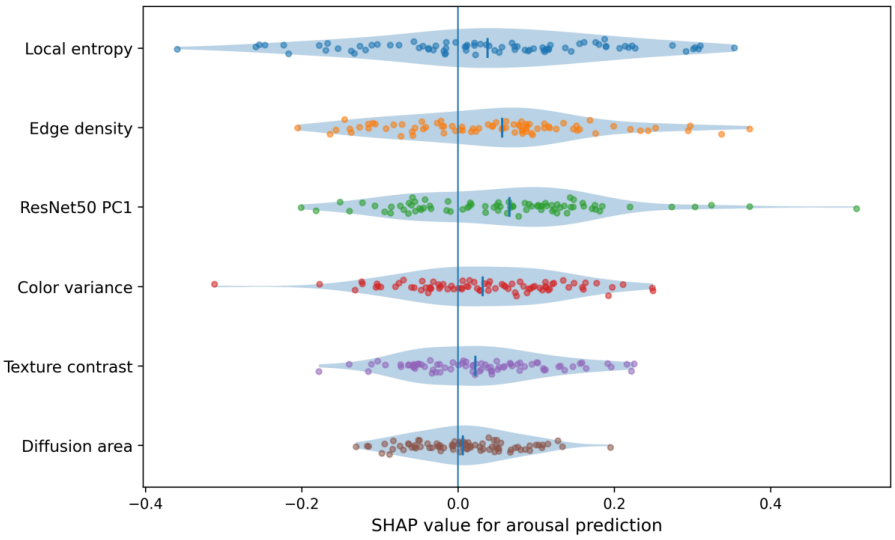


Figure 1. Illustrative SHAP distribution of visual feature contributions

5. Discussion

It is found that different digital image processing techniques exert a stronger impact on arousal rather than valence, which means that alterations in watercolor diffusion structure have an influence on visual stimulation easier. Arousal was lowered to 4.80 ± 0.99 after applying smoothing technique, whereas local texture and diffusion edges became less pronounced. It seems like complex textures may play a key role in conveying visual information and maintaining emotions' activation. The score of arousal for texture-enhanced images was raised up to 5.92 ± 1.12, which indicates that higher edge density and local texture contribute to visual stimulation. Digitally restored images had an arousal score of 5.37 ± 1.19, which is quite similar to the initial image. Therefore, preservation of diffusion structure is helpful in retaining perceptual reaction. The proposed fused feature model has

DOI: 10.54254/2753-8818/2026.38450

$R^2 = 0.64$. The highest contributions are shown by local entropy and edge density in SHAP analysis, which creates the connection between subjective rating and objective visual features. The lower predictions of valence mean that pleasantness is more dependent on aesthetic perception and overall evaluation of the visual stimulus.

6. Conclusion

This study presents an analysis workflow in which the association between digital representations of watercolor diffusion traces and emotional perception was investigated. This workflow includes steps for digital image processing, visual feature extraction, emotion evaluation and prediction based on machine learning. For four different representation conditions including the original, smoothed, texture enhanced and digitally reconstructed images, different emotions were generated. By smoothing, arousal was decreased, while by texture enhancement, emotional activation increased. Digital reconstruction resulted in better perceptual performance compared to the original images, especially in the cases that major diffusion structure was preserved. When low-level visual features from OpenCV and deep features from ResNet50 were concatenated to each other, XGBoost model resulted in an MAE of 0.60, RMSE of 0.80, and R^2 of 0.64 for arousal prediction. SHAP analysis revealed the importance of local entropy and edge density as visual features. Thus, a quantified relationship between digital morphological changes and emotional responses was established through this analysis.

References

[1] McAdams, P., et al. (2023). Chromatic and spatial image statistics predict infants' visual preferences and adults' aesthetic preferences for art. *Journal of Vision, 23*(8), 2–2.
[2] Bruns, A., et al. (2023). Emotions of subject and object affect beauty differently for images and music. *Journal of Vision, 23*(13), 6–6.
[3] Odushegun, L. (2023). Aesthetic semantics: Affect rating of atomic visual web aesthetics for use in affective user experience design. *International Journal of Human-Computer Studies, 171*, 102978.
[4] Ananthram, A., Winn, O., & Muresan, S. (2023). Feelingblue: A corpus for understanding the emotional connotation of color in context. *Transactions of the Association for Computational Linguistics, 11*, 176–190.
[5] Schubert, E. (2024). The aesthetic emotion lexicon: A literature review of emotion words used by researchers to describe aesthetic experiences. *Empirical Studies of the Arts, 42*(1), 3–37.
[6] Gonzalez-Martin, C., Carrasco, M., & Wachter Wielandt, T. G. (2024). Detection of emotions in artworks using a convolutional neural network trained on non-artistic images: A methodology to reduce the cross-depiction problem. *Empirical Studies of the Arts, 42*(1), 38–64.
[7] Kaube, H., & Abdel Rahman, R. (2024). Art perception is affected by negative knowledge about famous and unknown artists. *Scientific Reports, 14*(1), 8143.
[8] Stamkou, E., et al. (2024). Emotional palette: A computational mapping of aesthetic experiences evoked by visual art. *Scientific Reports, 14*(1), 19932.
[9] Swartz, A., et al. (2024). The perceived beauty of art is not strongly calibrated to the statistical regularities of real-world scenes. *Scientific Reports, 14*(1), 19368.
[10] Salgues, S., et al. (2024). Self-reference and emotional reaction drive aesthetic judgment. *Scientific Reports, 14*(1), 19699.
[11] Castellotti, S., D'Agostino, O., & Del Viva, M. M. (2024). Effectiveness of labels in digital art experience: Psychophysiological and behavioral evidence. *Frontiers in Psychology, 15*, 1342667.