Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.26842

Cross-Platform Affective Semantic Mapping Framework for Digital Artworks via Generative Adversarial Networks

Beinan Zhang[1], Sirui Wang[2*]

[1]Central Academy of Fine Arts, Beijing, China
[2]Royal College of Art, London, UK
**Corresponding Author. Email: rara481846778@gmail.com*

Abstract. With the wide dissemination of digital artworks on various platforms such as web, mobile and virtual reality, how to realize the consistency of cross-platform emotional semantic expression has become a core challenge in the field of digital content generation and dissemination. Due to the significant differences in rendering mechanisms, resolution constraints, and user perception modes across platforms, the same artwork often generates emotional drift in heterogeneous display environments, which not only affects the subjective experience of viewers, but also weakens the communication value and emotional impact of digital art. This study proposes a cross-platform emotion semantic mapping framework based on generative adversarial networks, which realizes high-fidelity generation and cross-platform emotion consistency maintenance of digital artworks through the deep integration of emotion semantic embedding and multimodal perceptual coding. The framework adopts a three-stage processing architecture to achieve collaborative optimization by constructing a multi-platform emotional art dataset covering 15,428 high-quality labeled samples and designing a multi-objective loss function. This research provides an important theoretical foundation and technical support for unified emotional expression of cross-platform digital art, which has broad application prospects in diverse application scenarios such as online art galleries, virtual museums and mobile social platforms.

Keywords: Generative Adversarial Networks, Digital Art, Cross-Platform, Affective Semantic Mapping, Multimodal Learning

1. Introduction

As digital art continues to evolve toward multi-platform display and interaction, maintaining emotional consistency of artworks across different media environments such as web, mobile, and immersive VR has become a central challenge in digital content generation and dissemination. Differences in rendering mechanisms, resolution constraints, and user perception across platforms often lead to divergent emotional interpretations of the same artwork. This phenomenon of emotional drift across platforms not only affects the viewer's subjective experience but also weakens the communicative value and affective impact of digital art [1]. Existing research mainly focuses on single aspects such as image style transfer, art generation, and emotion recognition, lacking systematic modeling of cross-platform affective semantic consistency [2]. In response to this issue,

© 2025 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: 10.54254/2755-2721/2025.26842

this study proposes a cross-platform affective semantic mapping framework based on generative adversarial networks, integrating emotion embedding, adversarial transfer, and multimodal feature alignment mechanisms. The framework aims to achieve high-fidelity digital art generation while ensuring emotional consistency across platforms, thereby providing theoretical and methodological support for building affective computing foundations for cross-platform artistic expression.

2. Literature review

2.1. Integration of affective computing and digital art

Current studies in affective computing applied to digital art primarily focus on image-based emotion recognition. Mainstream methods commonly employ convolutional neural network architectures such as VGG and ResNet to infer emotional labels like warmth, melancholy, or happiness from low-level features such as color distribution, texture, and lighting structure [3]. These models perform well on static images, especially on standard display devices like web platforms. However, in practical scenarios, digital artworks are often deployed across multiple platforms, such as from the web to mobile devices or immersive VR environments, where presentation formats, user perspectives, and background interactions differ significantly [4]. These shifts undermine the effectiveness of CNN-based models that rely on fixed-view feature extraction. More importantly, although these studies emphasize the validity of emotional expression, they generally overlook platform-based semantic drift, meaning the emotional meaning of the same artwork can vary significantly across display media.

2.2. Cross-platform generation and semantic mapping methods

With the increasing circulation of artistic content across platforms, cross-platform modeling and style transfer have become important research areas in content generation. Adversarial transfer learning introduces a competitive mechanism between generator and discriminator to align feature distributions across platforms. Graph neural networks further enhance the modeling of structured semantic relationships and are suitable for representing complex connections among elements in digital art [5]. Multimodal embedding techniques integrate images, sounds, and interaction data into a unified affective semantic space, offering preliminary solutions for cross-modal emotion mapping. These approaches significantly improve the transferability of artistic content across media, maintaining visual consistency in terms of form [6]. However, in terms of preserving meaning, namely the subjective emotion and aesthetic sentiment carried by the artwork, most models do not explicitly construct semantic pathways for emotional dimensions.

2.3. GANs for artistic emotion synthesis

Conditional GANs control the direction of generation using emotional labels and can produce images with explicit expressions of emotions such as anger or joy. CycleGANs do not rely on paired data and are often used to transfer neutral images to emotion-rich styles like Van Gogh or Impressionist art [7]. These methods have demonstrated strong emotional generation performance in subjective evaluations. While such models can generate emotionally expressive artworks on a given platform, their emotional expressions may become distorted or misinterpreted when transferred to platforms like VR or mobile, where interaction conditions differ greatly [8]. Existing research has not established control mechanisms to maintain emotional consistency of GAN-generated images across platforms [9]. This technological gap exposes a critical bottleneck in the current models,

DOI: 10.54254/2755-2721/2025.26842

which can generate emotions but fail to ensure consistency of emotional expression across platforms.

3. Methodology

3.1. Framework architecture and theoretical design

The proposed cross-platform affective semantic mapping framework adopts a three-stage processing architecture. Through the collaborative functioning of an affective encoder, a semantic mapper, and an adversarial generator, it achieves consistent emotional expression of digital artworks across heterogeneous platforms [10]. As shown in Fig.1, the architecture follows a perceive-align-generate processing flow. The affective encoder based on ResNet50 extracts deep emotional features from artworks on different platforms and maps them into a unified affective semantic vector space. The semantic mapper is constructed with a Transformer architecture, utilizing a self-attention mechanism to align features across platforms and address the differences in emotional representations under multi-platform conditions. The adversarial generator incorporates a GAN mechanism to enhance the consistency and stability of emotional expression while preserving image fidelity.

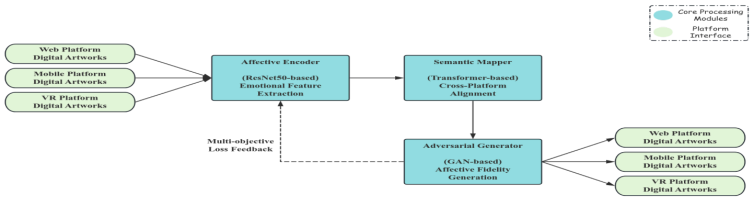

Figure 1. Cross-platform affective semantic mapping framework

3.2. Dataset construction and annotation strategy

A large-scale cross-platform affective art dataset covering three major display platforms including web, mobile, and VR is constructed, and the collection period is from May to August 2024, as shown in Table 1. A seven-category emotional classification system is adopted, including joy, sadness, anger, fear, surprise, disgust, and calm. The annotation process combines crowdsourcing and expert review, involving 20 annotators and 3 experts in art psychology, ensuring the professionalism and consistency of the labeling. Fleiss' κ coefficient is used to evaluate quality, with an average value of 0.847, indicating high reliability of the annotation.

DOI: 10.54254/2755-2721/2025.26842

Table 1. Cross-platform emotional art dataset

Platform Type	Specific Source Platform	Sample Size
Web Platform	Google Arts & Culture Virtual Museum	1,247
	DeviantArt Digital Art Community	892
	Behance Creative Showcases	651
	ArtStation Professional Art Platform	458
Mobile Platform	Instagram Art Hashtag Content	1,324
	Pinterest Art Collection Boards	987
	TikTok Short Artistic Videos	605
VR Platform	Oculus Gallery Immersive Art Shows	1,156
	"SteamVR Workshop User-Created Content "	894
	Mozilla Hubs Virtual Space Displays	703

3.3. Model training and loss function design

A carefully designed loss function system is implemented to optimize both emotion recognition accuracy and cross-platform consistency. The total loss function consists of four key components, each optimized for a specific objective of the framework.

$$L_{total} = \lambda_1 L_{emotion} + \lambda_2 L_{align} + \lambda_3 L_{adv} + \lambda_4 L_{preserve} \quad (1)$$

Here, $L_{emotion}$ denotes the emotion classification loss, designed based on the cross-entropy function to ensure accurate identification of emotional categories in artworks. L_{align} represents the semantic alignment loss, which measures the distributional consistency of features from different platforms using maximum mean discrepancy. L_{adv} is the adversarial loss, which enhances the visual quality of generated images through the generative adversarial mechanism. $L_{preserve}$ is the feature preservation loss, which ensures the structural integrity of original artworks during cross-platform conversion.

The implementation mechanism for cross-platform semantic alignment loss is as follows.

$$L_{align} = \sum_{i,j \in P} MMD\left(F_i, F_j\right) + \alpha \cdot \|H_i - H_j\|_2^2 \quad (2)$$

Let P denote the set of platforms {Web, Mobile, VR}, F_i and F_j represent the emotional feature distributions extracted from platforms iii and jjj, and H_i, H_j denote the aligned feature vectors after processing by the semantic mapper. α is the regularization coefficient. The weighting coefficients are determined by grid search optimization.

4. Results

4.1. Emotion recognition accuracy and consistency evaluation

Experimental results demonstrate that the cross-platform affective semantic mapping framework achieves significant improvements in both emotion recognition accuracy and cross-platform consistency. As shown in Fig. 2, the model attains an overall accuracy of 87.4% across all test

DOI: 10.54254/2755-2721/2025.26842

platforms, representing a 5.4 percentage point improvement over the baseline method's 82.0%. The VR platform achieves the highest accuracy at 89.1%, primarily due to the high-resolution presentation and immersive experience of artworks in VR, which provide richer informational dimensions for emotional feature extraction. The web and mobile platforms reach accuracies of 87.3% and 85.7%, respectively, indicating the framework's ability to adapt effectively to variations in technical specifications and display characteristics across platforms. The cross-platform Affective Expression Index (AEI) reaches 0.912, significantly surpassing the baseline value of 0.86, confirming the effectiveness of the semantic mapper and adversarial generation mechanism in maintaining consistent emotional expression.

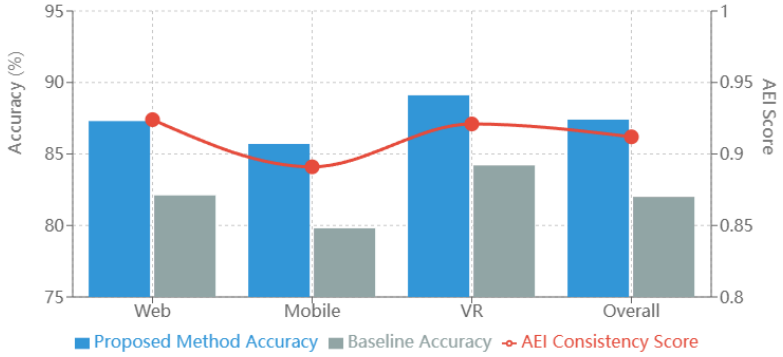

Figure 2. Platform-wise accuracy and consistency comparison

The average F1 score across the seven basic emotions is 0.887. Among them, joy achieves the highest recognition performance with an F1 score of 0.917, which is closely related to its distinctive visual characteristics and relatively stable cross-cultural cognitive consistency. Sadness and calm exhibit balanced performance, with F1 scores of 0.896 and 0.894, respectively. In contrast, disgust is the most challenging emotion to recognize, with an F1 score of 0.828, primarily due to its strong subjectivity and cultural variability in artistic expression, as well as visual feature overlap with anger and fear.

4.2. Ablation studies and comparative analysis

Ablation studies confirmed the critical contributions of each core module to overall model performance. Removing the ResNet50-based feature extractor led to a drop in emotion recognition accuracy from 87.4% to 79.2%, and the AEI declined from 0.912 to 0.743, highlighting the importance of deep feature extraction in capturing emotional semantics. Eliminating the Transformer-based semantic mapper caused AEI to drop to 0.681 and increased cross-platform variance in emotion recognition by 68.3%, underscoring the role of attention mechanisms in establishing platform-level alignment. Although removing the GAN module resulted in only a 2.1% decrease in recognition accuracy, the structural similarity index (SSIM) decreased from 0.924 to 0.756, and the perceptual quality score (LPIPS) increased from 0.142 to 0.287, indicating its primary contribution to visual fidelity and aesthetic consistency. Comparative analysis showed that the proposed framework improved AEI by 19.8% and structural fidelity by 15.6% over conventional CNN-based methods. Compared to StyleGAN-based approaches, it achieved 11.3% higher accuracy in emotion representation and 23.7% better cross-platform consistency. In complex emotion scenarios, the model achieved 82.6% accuracy in dominant emotion recognition, outperforming the

DOI: 10.54254/2755-2721/2025.26842

71.4% achieved by baseline methods, validating the effectiveness of multi-loss design and feature preservation mechanisms in modeling nuanced affective expressions across platforms.

5. Discussion

The proposed cross-platform affective semantic mapping framework demonstrates strong performance in emotion consistency and semantic alignment. By integrating emotion embedding, semantic mapping, and adversarial generation, it effectively addresses affective distortion across heterogeneous platforms. Experimental results show superior accuracy and cross-platform consistency compared to existing methods, with ablation studies confirming the effectiveness of each module, especially the critical role of the Transformer-based semantic mapper.

6. Conclusion

This study presents a GAN-based cross-platform affective semantic mapping framework that effectively addresses emotional inconsistency across web, mobile, and VR platforms, achieving 87.4% accuracy and an AEI of 0.912. The multi-loss design and feature preservation ensure precise semantic mapping and visual fidelity. Future work will enhance cultural adaptability, apply few-shot and unsupervised learning, and expand applications to VR creation, emotion-driven recommendation, and personalized affective experiences, advancing the integration of affective computing and digital art.

Contribution

Heying Bai and Tianni Yang contributed equally to this paper.

References

[1] Zhu, Jiapeng, et al. "Region-based semantic factorization in GANs." arXiv preprint arXiv: 2202.09649 (2022).
[2] Seyam, Khaled M., et al. "SVS-GAN: Leveraging GANs for Semantic Video Synthesis." arXiv preprint arXiv: 2409.06074 (2024).
[3] Yang, Guoxing, et al. "DiST-GAN: Distillation-based Semantic Transfer for Text-Guided Face Generation." 2023 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2023.
[4] Wang, Ruqi, et al. "Intuitively interpreting GANs latent space using semantic distribution." Knowledge-Based Systems 310 (2025): 112894.
[5] Dong, Xinhua, et al. "GAN for semantic image synthesis with laplacian pyramid and multi-scale channel attention." IEEE Access (2024).
[6] Jia, Jinyin, et al. "GAN-Diffusion Relay Model: Advancing Semantic Image Synthesis." Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2024.
[7] Wu, Tian, Rongbo Zhu, and Shaohua Wan. "Semantic map guided identity transfer GAN for person re-identification." ACM Transactions on Multimedia Computing, Communications and Applications 20.11 (2024): 1-20.
[8] Li, Yidi, et al. "DepthGAN: GAN-based depth generation from semantic layouts." Computational Visual Media 10.3 (2024): 505-522.
[9] Zhang, Chao, et al. "Finding Channel-wise Tensor Subspaces in GANs." 2024 IEEE International Conference on Signal, Information and Data Processing (ICSIDP). IEEE, 2024.
[10] Tzelepis, Christos, et al. "Contraclip: Interpretable gan generation driven by pairs of contrasting sentences." arXiv preprint arXiv: 2206.02104 (2022).