EXPLORING EMOTIONAL EXPRESSION IN DIGITAL ART THROUGH DEEP LEARNING TECHNIQUES

Dr. Prashant Wakhare [1], Dr. Riyazahemed A. Jamadar [2], Dr. Sanjay Bhilegaonkar [3], Pallavi Mulmule [4]

[1] Assistant Professor, All India Shri Shivaji Memorial Society's, Institute of Information Technology, Pune, Maharashtra, India
[2] Assistant Professor, All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune-01, Maharashtra, India
[3] Savitribai Phule Pune University, Pune, Maharashtra, India
[4] Assistant Professor, Department of Electronics and Communication Engineering, DES Pune University, Pune, Maharashtra, India

Received 22 March 2025
Accepted 26 July 2025
Published 20 December 2025

Corresponding Author
Dr. Prashant Wakhare,
pbwakhare@gmail.com

DOI
10.29121/shodhkosh.v6.i3s.2025.6952

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

The expression of emotion is a characteristic but difficult feature of digital art that is most frequently expressed in abstract visual features instead of direct semantics. This paper explores how deep learning methods can be used to learn and analyze emotional expression in digital artwork. The proposed hybrid model that integrates Convolutional Neural Networks and Vision Transformers will be able to capture local visual features, including color and texture as well as global compositional structure. A selected collection of various digital artworks is modeled and cited by a hybrid emotion system based on discrete categories and dimensional valence-arousal models. The experimental findings prove that the proposed hybrid method is more successful than CNN and transformer baselines on both emotion classification and regression problems with a higher F1-score, reduced error in prediction, and increased correlation with human emotional ratings. Embedding-level and qualitative analyses also indicate that the learned representations are able to maintain emotional continuity as well as ambiguity in artistic expression. The results affirm that emotion in digital art is multidimensional and optimal with regard to integrated local-global feature learning. The presented work contributes to the development of affective computing in the world of creativity and offers a premise to the study of emotional art, curating it, and creative collaboration between humans and AI.

Keywords: Digital Art, Emotional Expression, Deep Learning, Affective Computing, Emotion Embeddings, Vision Transformers, Valence–Arousal Model

1. INTRODUCTION

Art has always worked as a potent tool of human expression, it helps the artist to project the internal moods and helps the audience to feel, interpret and relate to the feeling. As digital technologies have evolved, artistic expression has not only been able to move out of the purely physical form of representation but into computationally mediated spaces where images, animations, and interactive forms can be produced, edited and disseminated in digital form Deonna and Teroni (2025). The digital art is not like its traditional ones: it became inherently algorithmic, data-driven, and dynamic,

How to cite this article (APA): Wakhare, P., Jamadar, R. A., Bhilegaonkar, S., and Mulmule, P. (2025). Exploring Emotional Expression in Digital Art through Deep Learning Techniques. *ShodhKosh: Journal of Visual and Performing Arts, 6*(3s), 511–522. doi: 10.29121/shodhkosh.v6.i3s.2025.6952

giving new opportunities to encode and amplify the expression of emotions. Nonetheless, this change also brings about the complexity to the ways of understanding how emotions are engrained, perceived, and expressed in digitally created visual artifacts. Along with the development of digital art, a conceptual breakthrough has taken place in the field of computer vision and artificial intelligence via deep learning Zhou and Lee (2024). Convolutional Neural Networks, Vision Transformers, and generative architectures were models that have been shown to be extremely capable of learning abstract visual features and semantic patterns on large-scale data. These models when used on artistic material present hitherto unheard-of possibilities to study emotional signifiers, like color harmony, texture dynamics, spatial arrangement, and symbolic motif, on a scale, and at a level that surpasses human perceptual consistency Lin et al. (2023). This intersection of deep learning and digital art is what forms a new interdisciplinary space where affective computing is overlapping with creative practice.

Figure 1

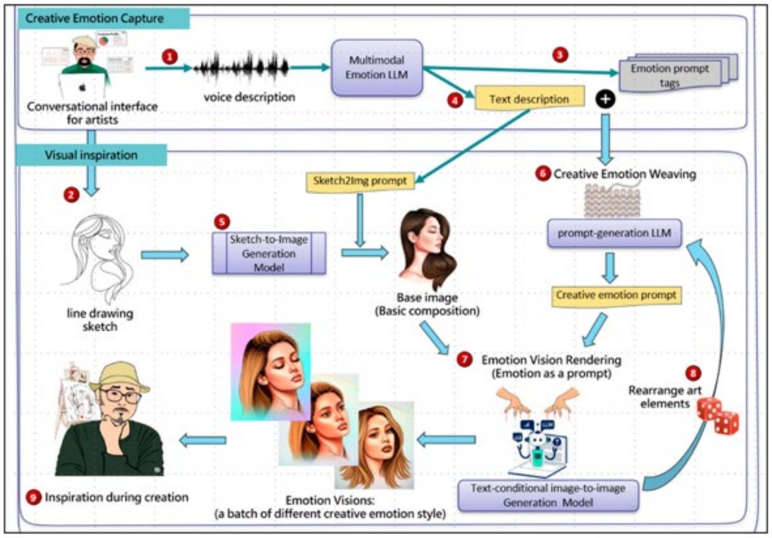
Creative Emotion Capture
Conversational interface for artists
1
voice description
Multimodal Emotion LLM
3
Emotion prompt tags
4
Text description
Visual inspiration
2
Sketch2Img prompt
6 Creative Emotion Weaving
prompt-generation LLM
line drawing sketch
5
Sketch-to-Image Generation Model
Base image (Basic composition)
Creative emotion prompt
7 Emotion Vision Rendering (Emotion as a prompt)
8
Rearrange art elements
9 Inspiration during creation
Emotion Visions: (a batch of different creative emotion style)
Text-conditional image-to-image Generation Model

Figure 1 Conceptual Overview of Emotional Expression in Digital Art Using Deep Learning

Although there has been a considerable advance in the emotional recognition of natural images and facial expressions, emotional analysis in digital art is not investigated at the anticipated level. Artistic feelings are usually hard to define, culturally relative, and abstract willfully, which makes them hard to measure with traditional methods of emotion classification Fisher et al. (2023). Also, the current literature often uses small data sets or categorized emotional labels, which do not reflect the richness and subjectivity of artistic expression as shown in Figure 1. The current study fills these gaps by investigating deep learning methods of interpreting and 512rtifice512 emotional expression in the digital art. The main contributions of this work are as follows: (i) the conceptual connection between theories of emotional aesthetics and deep learning-based visual representation learning; (ii) the creation of a systematic approach to the emotional analysis of digital works; and (iii) the empirical analysis that proves that advanced neural networks are capable of capturing hidden emotional patterns beyond the visual features of the image Yang et al. (2023), Ren et al. (2023). The connection between artistic theory on the one hand and 512rtificeal intelligence on the other hand is expected to foster not only the research on affective computing, but also the discourse on the understanding of emotional creativity as it is relevant to digital art which is currently developing.

2. THEORETICAL BACKGROUND AND RELATED WORK

The expression of emotions in art has long been studied within the framework of aesthetic theory, psychology, and cognitive science where emotions are perceived as part of the meaning of art and the experience of viewers Fu et al. (2023). Classical aesthetic viewpoints focus on the extent to which visual aspects (color, form, balance and symbolism) have affective effects, whereas psychological theories (appraisal theory and dimensional emotion models) describe how viewers perceive these stimuli cognitively. These theoretical principles are still necessary in the framework of digital art but are further prolonged by algorithmic generation Mendes et al. (2023), interactivity, and computational abstraction, which provide new ways of expressing emotion. Computationally, affective computing provides the conceptual framework on how to model the emotions with the help of intelligent systems. The initial study of visual emotion analysis has been based on manual features that were based on the principle of aesthetics such as the harmony of colors, statistics of texture, and compositional heuristics. These methods were interpretable, but could not scale to a larger range of artistic styles Krumhuber et al. (2023). Deep learning introduced a novel paradigm shift in that it allowed one to automatically learn hierarchical visual representations directly on data. CNNs have been widely used to match visual features of paintings and stylized images with the emotional set, whereas more recent transformer based design encodes global information and long-range relationships that are especially important when it comes to complex artistic works Küster et al. (2022). Although these developments have been made, a significant portion of the current literature is on natural pictures, facial expression, or social media material, when emotional indicators are comparatively clear. Digital art, in its turn, can be quite abstract, stylistically deviant, and deliberately ambiguous, which makes the process of emotional interpretation rather subjective and culturally contingent. A number of studies have tried to fill this gap by training deep learning models in data of art like WikiArt, but very often in these studies the models of emotions are simplified or lack theoretical basis in artistic aesthetics Walker et al. (2022). Generative methods, such as Creative Adversarial Networks and diffusion-based models have added more layers to this field since emotion-modulated art creation can be achieved, but evaluation of emotion in these models is mostly implicit.

Table 1

Table 1 Summary of Key Related Works on Emotional Analysis in Art and Visual Media

Domain	Dataset Used	Methodology	Emotion Model	Key Limitations
Paintings	MART, WikiArt (subset)	Handcrafted visual features	Discrete emotions	Limited scalability; manual feature design
Art & Images	WikiArt	CNN-based classification	Discrete emotions	Limited contextual and cultural sensitivity
Affective Images	Flickr, art images	CNN with attention	Valence–Arousal	Primarily focused on natural images
Computational Creativity	WikiArt	Creative Adversarial Networks	Implicit affect	Emotions not explicitly evaluated
Stylized Art	WikiArt Emotions	Multi-label CNN	Discrete emotions	Limited interpretability
Visual Emotion	Mixed datasets	CNN–RNN hybrid	Valence–Arousal	Not tailored for static art
Art & Images	WikiArt, custom datasets	Vision Transformers	Hybrid models	High data dependency
Digital Art	**Curated digital art dataset**	**CNN / ViT with emotion-aware embeddings**	**Hybrid (Discrete + Dimensional)**	**Addresses abstraction and subjectivity**

As shown in Table 1, the current methods show that deep learning can be used to analyze artworks emotionally but also show significant gaps especially when it comes to the abstraction, subjectivity and cultural variability. Additionally, the lack of incorporation of art-theoretical approaches into computer models limits the levels of interpretability and artistic applicability De and Grana (2022). The current research is based on the previous ones as it incorporates deep learning methods into a conceptually grounded model of emotional aesthetics, thus, allowing a more subtle and contextual perception of emotional expressiveness in the field of digital art.

3. EMOTIONAL REPRESENTATION IN DIGITAL ART

The aspect of emotion in digital art seeks to bring out the emotional reaction of the viewer through the integration of visuals that subconsciously influence the emotional reaction of the viewer. Instead of using the explicit narrative or figurative suggestions, digital artworks often use such abstract qualities as color, texture, composition, and stylistic form to represent emotion. These components are used as signals of perception that informs emotional interpretation but permits ambiguity and subjective involvement, which are the primary aspects of digital artistic practice today. Color is a leading factor in emotional communication as it affects the perceived mood and strength. The differences in hue, saturation, brightness, and contrast have a close relationship with affective dimensions of arousal and valence. The palette of warm and highly saturated colors can give the impression of excitement or vitality, whereas colder color pallets and muted contrast can give the impression of calmness, melancholy or introspection. Emotional expression is also supported by texture with surface irregularity, granularity as well as visual noise. Continuous and smooth textures are generally associated with a calm or peaceful mood whereas broken or rough textures may create some level of tension, uneasiness, or emotional nausea. These textual effects are digitally created or controlled in digital art, which provides greater emotional expressivity not limited by the material.

Figure 2

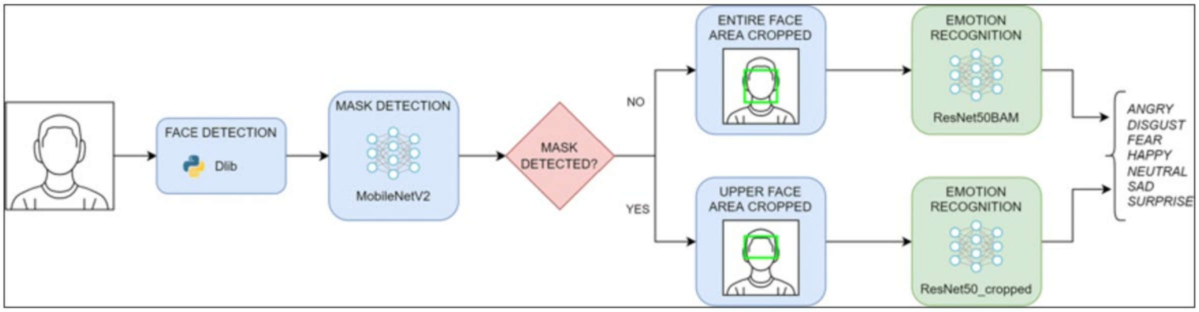

Figure 2 Transformation of Visual Artistic Elements into Emotion Embeddings Using Deep Learning.

Form and composition are also part of emotional perception, organizational elements of visual attention and space. Stability, dominance, vulnerability or unease are effects of balance, symmetry, scale, and distortion. Digital art is often using a non-linear structure, exaggerated proportions and symbolic abstraction, which promotes interpretive richness and emotional richness. Notably, these visual elements do not work alone very often, the emotional meaning is created as a result of their integration, which results in less clear and even contradictory emotional signals. Figure 2 depicts the process through which visual attributes are converted to emotion embeddings to conceptualize the translation of these elements of art into computational representations. The input space, i.e. artistic elements like color, texture and composition, is more and more encoded in deep learning-based layers of feature extraction, as indicated in the figure. The initial perception is captured at low-level and then further, higher-level semantic and stylistic representations are learnt using convolutional and transformer-based models. These representations are then projected into a latent emotion embedding space where affective meaning is manifested in continuous or hybrid emotional space. This conceptual pipeline brings out the model of how deep learning allows holistic modeling of emotional expression through combining several visual dimensions into one affective dimension.

4. DEEP LEARNING TECHNIQUES FOR EMOTIONAL ANALYSIS

Deep learning has become a leading model of the visual expression modeling based on its ability to acquire hierarchical and abstract representations directly out of data. Emotional meaning in digital art can also take forms that are implicit in visual form and explicit in semantic terms, a form that the deep learning can be especially well applied to analyze. Deep neural models learn, based on extensive collections of artworks, to capture complex emotional patterns which result as a consequence of interaction between color, texture, composition, and style. The application of CNNs to analyze emotions in pictures and paintings has been widely applied due to the ability to learn localized visual patterns. The low-level features, including edges, colour transitions, texture primitives, are represented in the early convolutional layers whereas the deeper the convolutional layer, the higher are the mid-level features and compositional fragments. These characteristics ensure that CNNs are very appropriate to understand emotional cues concerning color harmony and surface texture. Nevertheless, they are constrained by the fact that local receptive fields cannot satisfy long-range dependencies and holistic composition (as are commonly the focus of emotional interpretation in abstract or minimalist digital art). Vision Transformers (ViTs) overcome this drawback by having pictures represented as collections of visual patches subjected to self-attention schemes. This enables transformer to grasp the global contextual relationship and compositional structure which allows better interpretation of emotional cues which are generated through spatial arrangement and visual balance. ViTs offer better interpretability as well, in the form of attention maps which reveal areas of a work of art that are emotionally salient. In spite of these virtues, transformer based models usually need large datasets and substantial computational resources and are not generally sensitive to fine-grained texture detail alone.

Table 2

Table 2 Comparison of Deep Learning Architectures for Emotional Analysis in Digital Art

Architecture	Core Characteristics	Strengths for Emotional Analysis	Limitations	Suitability for Digital Art
Convolutional Neural Networks (CNNs)	Hierarchical convolution and pooling; local receptive fields	Effective at capturing color gradients, texture patterns, and local stylistic features	Limited global context modeling	Texture- and color-driven artworks
Vision Transformers (ViTs)	Patch-based representation with self-attention	Strong global context understanding; effective for abstraction and composition	High data and computation requirements	Composition- and structure-driven art
Hybrid Models (CNN + ViT)	CNN feature extraction with transformer attention	Balanced modeling of local detail and global structure; robust emotion embeddings	Increased architectural complexity	Complex and multi-layered digital artworks

In order to overcome the advantages and disadvantages of single architectures, mixed models combining CNNs and transformers have become the leading ones. The CNNs are effective feature extractors of local texture and color patterns, whereas transformer layers combine information of the global context and composition in such structures. The hybrid methods provide stronger emotion embeddings with the capacity to model both local and global affective cues, thus being especially useful with the emotional complexity of digital art. The synthesis of Table 2 results in the fact that there is no single architecture, which can be considered an optimal one in the field of emotional analysis in the digital art. Rather, the model used should capture the artistic properties of data and the emotional granularities needed by the application. These factors are directly reflected in the structure of the designed framework and inspire the implementation of hybrid deep learning practices within the further approach.

5. DATASET CONSTRUCTION AND EMOTION ANNOTATION

The quality, diversity, and emotional reliability of the dataset are critical to the effectiveness of deep learning models to process emotions of digital art. In contrast to natural image collections, digital art collections are highly stylistically diverse, abstract, and culturally specific, which require a strictly developed data construction strategy and annotation approach. This paper, therefore, follows a systematic, theory-based strategy of dataset curation that is neither too artistic nor too computational. The dataset: It is a selected set of about 3,000 digital artworks used in open-access online galleries, digital art collections and portfolios of artists. The selection criteria will focus on stylistic variety and include abstract, figurative, surreal, generative (AI-driven), and mixed-media digital art in order to cover all the modes of expressing emotions. Several artists and cultural backgrounds are used to get a variety of works of art and minimize authorial and regional bias, and low-resolution or derivative images are avoided to avoid visual corruption. To promote the feature extraction and model training across all of the selected artworks, all of them are uniformed in resolution and color space.

Table 3

Table 3 Summary of the Digital Art Emotion Dataset

Attribute	Description
Total Artworks	~3,000 digital artworks
Art Styles	Abstract (32%), Figurative (24%), Surreal (18%), Generative/AI Art (16%), Mixed Media (10%)
Image Resolution	Standardized to 256×256 or 512×512
Color Space	RGB (normalized)
Discrete Emotion Labels	Joy, Sadness, Anger, Fear, Calm, Tension
Dimensional Model	Valence (−1 to +1), Arousal (0 to 1)
Emotion Distribution	Positive (38%), Neutral (27%), Negative (35%)

Annotators per Artwork	3–5 expert annotators
Inter-Annotator Agreement	Cohen's κ ≈ 0.72
Train / Val / Test Split	70% / 15% / 15% (stratified)

Table 3 shows a quantitative summary of the characteristics of the datasets, including the size of the dataset, stylistic composition, emotion models, and statistics of annotations. The level-balanced allocation of emotional extremes and types of art also makes sure that learning algorithms do not favor major dominating affective types. The annotation of emotion is one of the key issues because of the subjective and context-dependent nature of emotion in digital art, as well as its uncertainty. To counter this, a hybrid approach to annotation is utilized which is the integration of discrete emotion categories with dimensional valence-arousal ratings. Each artwork is sensitized to artistic nuance by labeling it by multiple annotators with either an art, design, or visual media background. The quantitative measure of inter-annotator agreement is through the statistical reliability measures and artworks that have significant disagreement are processed through averaged scores or soft-label representations as opposed to being forced into categorical assignments.

Figure 3

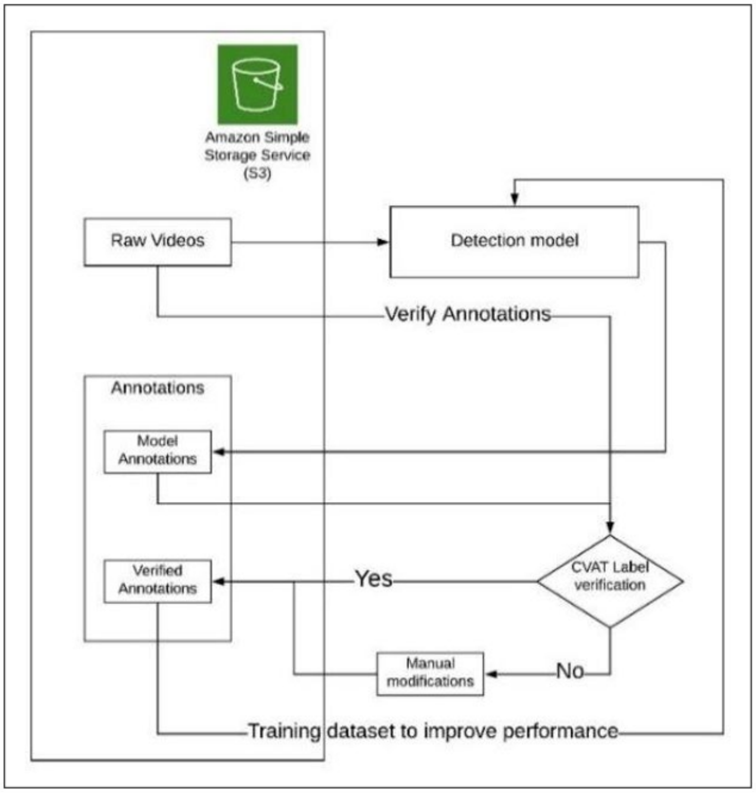

Figure 3 Multi-Rater Emotion Labeling and Embedding Generation Process for Digital Artworks

The annotation procedure adopted in this paper is presented in the form of a flow chart in Figure 3, which describes the entire process of the annotation starting with the selection of the artwork and finishing with creating final emotion embeddings. As illustrated in the figure, multi-rater labeling is performed on artworks, then agreement analysis and label refinement take place and the artworks are then encoded into continuous affective representations. The completed dataset is divided into training, validation, and testing fields using stratified sampling so that there is no loss of emotion and style across splits. The data augmentation methods are used in a conservative way so that they can strengthen the robustness without distorting the emotional salient visual cues. Taken collectively, these data building and labeling processes form a credible empirical basis of the suggested emotion-sensitive deep learning approach of the following section.

6. PROPOSED METHODOLOGY

The suggested emotion-sensitive deep learning model of analyzing emotional expression in digital art. The offered framework is created in the form of a modular and extensible pipeline that sequentially converts raw digital art pieces into semantically significant representations of emotions. It starts by processing artwork by consuming an image, whereby the image is made uniform with resolution normalization, color space matching, and scaling in intensity to maintain consistency between different artistic styles. The preprocessing phase does not only conserve emotionally significant visual stimulation, but also allows stable and efficient learning.

Figure 4

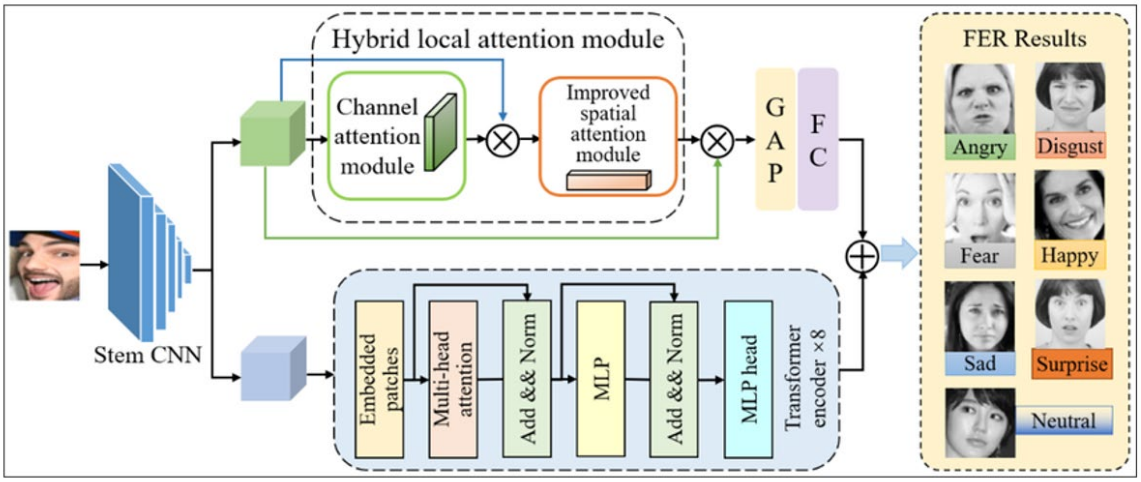

Figure 4 High-Level Architecture of the Proposed Hybrid CNN–Vit Framework for Transforming Digital Artworks into Emotion Embeddings.

After the normalization, the framework uses parallel feature learning to capture the multi-layered nature of the emotion expression in digital arts. In the former, a Convolutional Neural Network (CNN) branch specializes in fine-grained localized visual representations, including color changes, tonal variations, texture discontinuities and density of edges. These are low- to mid-level qualities that are directly related to the affective reactions that are concerned with mood, intensity, and visual tension. Simultaneously, a branch of the Vision Transformer (ViT) uses the artwork in the form of a series of visual patches and runs self-attention mechanisms to learn the concept of global spatial relationships. Such holistic elements of composition as symmetry, balance, spatial depth, abstraction, and stylistic coherence are the specific attributes that are effectively captured by this branch, and a feature fusion stage then combines the local and global representations into a single latent space that then determines the interpretations of digital art in an emotional manner. This combination allows the framework to maintain the fine perceptual details and at the same time carry more general contextual and stylistic details. This projected representation is embedded into a space of emotion modeling that gives a continuous and expressive affective representation that facilitates the modeling of both categorical and dimensional emotions.

7. DETERMINATION AND FINDINGS

The section includes an in-depth assessment of the suggested hybrid CNNViT models in the form of quantitative values, embedding-based assessment, and qualitative visual perception. It aims at evaluating not just the predictive performance, but also the capacity of the framework to approximate the subjective and subtle quality of the emotional expression in digital art. The results of the proposed hybrid model are tested in relation to CNN-only and ViT-only baselines in terms of discrete emotion classification and dimensional emotion estimation. Table 4 is a summary of the macro-averaged classification outcomes of six emotion categories. The hybrid CNNViT model has the best accuracy (78.9%) and F1-score (0.77), which means better performance on the emotion classes. Significant increases in performance are also seen especially in emotionally ambiguous categories like calm and tension where the joint modeling of local texture cues and global compositional structure is beneficial.

Table 4

Table 4 Discrete Emotion Classification Performance (Macro-Averaged)

Model	Accuracy (%)	Precision	Recall	F1-score
CNN-only	71.8	0.70	0.69	0.69
ViT-only	73.6	0.72	0.71	0.71
Hybrid CNN–ViT (Proposed)	**78.9**	**0.78**	**0.77**	**0.77**

Table 5 shows the findings of dimensional emotion prediction with mean squared error and Pearson correlation as the measures of valence and arousal. MSE and correlation coefficients are significantly lower in the hybrid model, indicating that it is more consistent with human affective judgments. Although CNN-only models show good performance on arousal estimation, probably because of the sensitivity to color intensity and texture difference, ViT-only models demonstrate a higher correlation between valence, as they capture the holistic composition. The hybrid architecture is a good way to leverage these complementary strengths.

Table 5

Table 5 Dimensional Emotion Prediction Performance

Model	Valence MSE ↓	Arousal MSE ↓	Valence Corr. ↑	Arousal Corr. ↑
CNN-only	0.083	0.091	0.71	0.68
ViT-only	0.076	0.088	0.74	0.70
Hybrid CNN–ViT (Proposed)	**0.062**	**0.069**	**0.81**	**0.78**

Other than the metric at the label level, the emotion embeddings that were learned were also analyzed to measure affective coherence. The visualizations of dimensionality reduction (t-SNE/UMAP) show distinct cluster formations relating to the positive, neutral, and negative area of emotions, and smooth transitions between them over the valence-arousal dimension. The hybrid model yields more intra-class clustering and more inter-class separation as evidenced by the high cosine similarity consistency and silhouette scores. This implies that the model demonstrates a continuous emotional organization as opposed to discrete labels only. Images with smooth gradients, equilibrium composition and low contrast are always projected into high valence and low arousal regions of the emotional brain related to calmness or contemplation. By comparison, paintings which have broken textures, contrast, and asymmetry are forecasted as low-valence, high-arousal, which would be an equivalent of a state of tension or discomfort. The existing qualitative correspondence between visual properties and the anticipated emotional supports the interpretability of the representations learnt and the use of the representations to human emotional perception. False identifications mainly happen in the works of art that have been constructed in such a way that they feel emotional ambiguity or ambivalent moods. When this happens, then predicted emotion embeddings are likely to be close to class boundaries instead of displaying confident mislabeling. This trend is associated with the subjectivity of artistic feeling as such and confirming the importance of soft-label learning and continuous emotion embeddings. The analysis of robustness also indicates that there is consistency in the performance of most styles of art, although there is slightly more variance in highly symbolic and surreal pieces, where cultural interpretation has a more significant influence.

8. DISCUSSION

The experimental findings, which are backed up by both quantitative and visual methods, show that deep learning may be adequately used to capture emotional expression in digital art provided that the architecture design does not conflict with artistic and affective values. This conclusion is well empirically supported by the comparative plots and tables in the Section 8 that especially focus on the benefits of the suggested hybrid CNN-ViT framework.

Figure 5

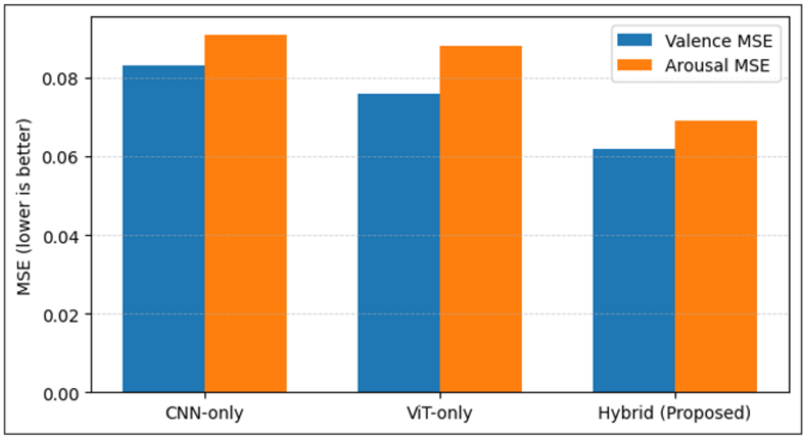

Figure 5 Valence and Arousal Regression Error Analysis

This interpretation is further supported by the regression error analysis found in Figure 5. The smaller value of the lower mean squared error of both the valence and arousal obtained with the hybrid model means that the model is more similar to the human-rated emotional intensities. The CNN-only models demonstrate comparatively competitive results in the arousal prediction, which can be explained by their sensitivity to the saturation of colors, contrast, and the density of textures as the characteristics that are often related to the intensity of emotions. ViT-only models, in contrast, have a higher valence estimation, which entails the ability to estimate holistic balance, symmetry, and spatial harmony. The hybrid architecture combines these complementary advantages, and the error in prediction has always been lower in both dimensions of affectivity.

Figure 6

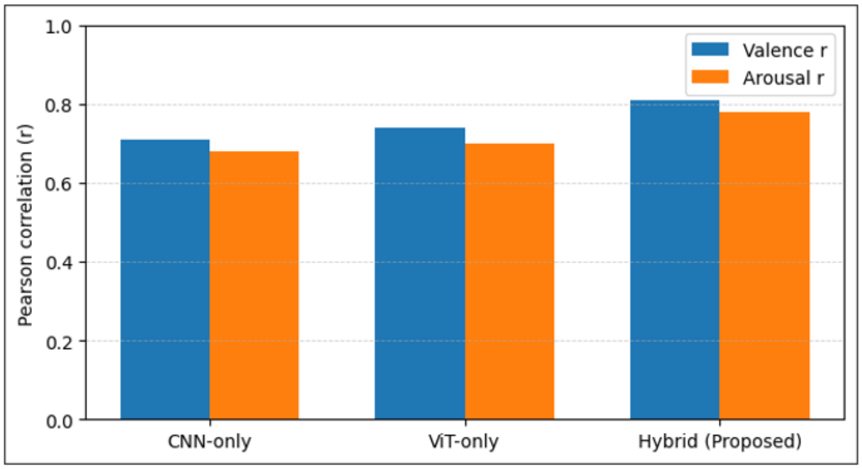

Figure 6 Correlation with Human Emotion Annotations

Concensus with human emotional perception is also confirmed by the analysis of correlation illustrated in Figure 6. The increased Pearson correlation coefficients of the hybrid model shows better correspondence with subjective human ratings of emotion. This finding is especially important within the framework of digital art, in which the emotional meaning is interpretive and always conditioned on the viewer. Large correlation coefficients are an indication that the learned emotion embeddings represent perceptual regularities that are appealing to human affective experience and not a fit to noise on the annotation.

Figure 7

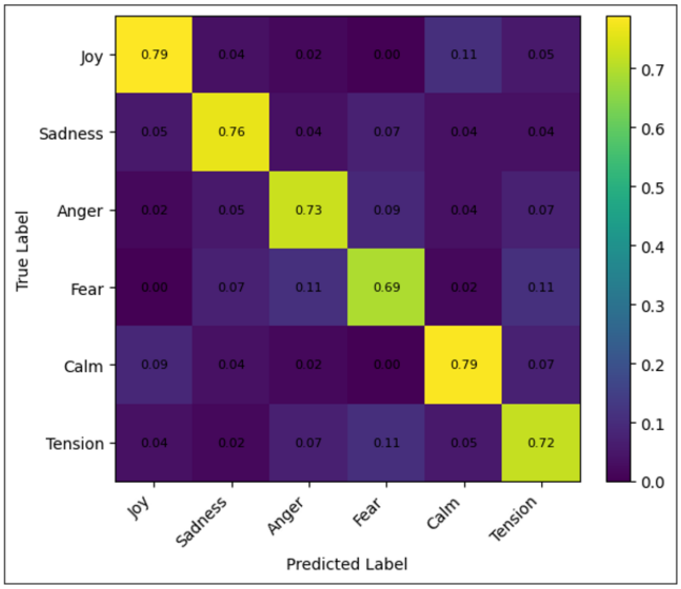

Figure 7 Normalized Confusion Matrix for Discrete Emotions

Figure 7, the analysis of the normalized confusion matrix at the class level, provides the information on model interpretability and limitation. High diagonal dominance implies good recognition of emotion whereas mix-up between closely similar classes like calm and tension are evidence of overlapping visual clues as opposed to the weakness of the model. This action justifies soft-label learning and perpetual emotion embeddings on artistic information. In general, the plots and tables prove that the expression of emotions in digital arts cannot be explained by only local or global representations. Rather, emotion is a result of color, texture, composition, and style interaction, which can be successfully represented by the hybrid CNN ViT framework. Gradual emotional transitions are also portrayed by smooth clustering in the embedding space, which is in line with the affect theory and visual aesthetics. But the fact that there is some variability in performance between styles and more confusion between very abstract or symbolic works means that the visual features do not encode the cultural context or individual interpretation in their entirety. The calculation cost of transformer based components is also a constraint. Nevertheless, the combined quantitative and qualitative obstacles have demonstrated that the suggested framework progresses affective computing in digital art and provides an interpretable and well-balanced basis of emotion-sensitive creation, curation, and human-AI interaction.

9. CONCLUSION AND FUTURE WORK

This paper has investigated the opportunities of emotional expression in digital art through deep learning on the problem of the modeling of affective meaning in visually abstract and stylistically varied artworks. To simultaneously learn local visual features, including color and texture, and global composition, a hybrid CNN ViT was suggested. These findings show that even the emotional representation of digital art is inherently a multidimensional concept that can be better represented by utilizing an integrated local-global feature learning as compared to the single-architecture implementations. The experimental analysis proved that the hybrid model is better than CNN-only and ViT-only baselines in discrete emotion classification as well as dimensional valence-arousal prediction. Greater F1-scores, smaller regression error and better correlation with human annotations show a better fit to subjective emotional perception. Integration of analysis also indicated consistent affective organization and unbroken emotional flow, which indicates the ambiguous and continuous quality of artistic emotion. Although these strengths are present, there are weaknesses. Cultural context and personal experience lead to emotional interpretation, and can only be described in part by the visual features. Components based on transformers raise computational complexity, and subjectivity of annotation cannot be completely removed. Further research will include including multimodal contextual information like textual descriptions, viewer feedback, and better work on model efficiency, and multimedia datasets within cultural contexts. It is also a promising line of extension of the framework to emotion-sensitive art generation, curation, or educational, or therapeutic uses. All in all, the work brings a compact and efficient strategy to affective computing in digital art, which interposes artistic theory with the practice of deep learning.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

De Lope, J., and Grana, M. (2022). A Hybrid Time-Distributed Deep Neural Architecture for Speech Emotion Recognition. International Journal of Neural Systems, 32(5), 2250024. https://doi.org/10.1142/S0129065722500241

Deonna, J., and Teroni, F. (2025). The Creativity of Emotions. Philosophical Explorations, 28(2), 165–179. https://doi.org/10.1080/13869795.2025.2471824

Dupré, D., Krumhuber, E. G., Küster, D., and McKeown, G. J. (2020). A Performance Comparison of Eight Commercially Available Automatic Classifiers for Facial Affect Recognition. PLOS ONE, 15(4), e0231968. https://doi.org/10.1371/journal.pone.0231968

Fisher, H., Reiss, P. T., Atias, D., Malka, M., Shahar, B., Shamay-Tsoory, S., and Zilcha-Mano, S. (2023). Facing Emotions: Between- and Within-Sessions Changes in Facial Expression During Psychological Treatment for Depression. Clinical Psychological Science. Advance online publication. https://doi.org/10.1177/21677026231195793

Fu, H., et al. (2023). Cross-Corpus Speech Emotion Recognition Based on Multi-Task Learning and Subdomain Adaptation. Entropy, 25(1), 124. https://doi.org/10.3390/e25010124

Galanos, T., Liapis, A., and Yannakakis, G. N. (2021). Affectgan: Affect-Based Generative Art Driven by Semantics. In Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction Workshops (ACIIW) (1–7). IEEE. https://doi.org/10.1109/ACIIW52867.2021.9666317

Ghanem, B., Rosso, P., and Rangel, F. (2020). An Emotional Analysis of False Information in Social Media and News Articles. ACM Transactions on Internet Technology, 20(1), 1–18. https://doi.org/10.1145/3381750

Krumhuber, E. G., Küster, D., Namba, S., and Skora, L. (2021). Human and Machine Validation of 14 Databases of Dynamic Facial Expressions. Behavior Research Methods, 53(2), 686–701. https://doi.org/10.3758/s13428-020-01443-y

Krumhuber, E. G., Skora, L. I., Hill, H. C. H., and Lander, K. (2023). The Role of Facial Movements in Emotion Recognition. Nature Reviews Psychology, 2(5), 283–296. https://doi.org/10.1038/s44159-023-00172-1

Küster, D., Steinert, L., Baker, M., Bhardwaj, N., and Krumhuber, E. G. (2022). Teardrops on My Face: Automatic Weeping Detection from Nonverbal Behavior. IEEE Transactions on Affective Computing. Advance Online Publication. https://doi.org/10.1109/TAFFC.2022.3228749

Lin, C., Bulls, L. S., Tepfer, L. J., Vyas, A. D., and Thornton, M. A. (2023). Advancing Naturalistic Affective Science with Deep Learning. Affective Science, 4(4), 550–562. https://doi.org/10.1007/s42761-023-00215-z

Lu, C., et al. (2022). Progressively Discriminative Transfer Network for Cross-Corpus Speech Emotion Recognition. Entropy, 24(8), 1046. https://doi.org/10.3390/e24081046

Mendes, C., Pereira, R., Ribeiro, J., Rodrigues, N., and Pereira, A. (2023). Chatto: An Emotionally Intelligent Avatar for Elderly Care in Ambient Assisted Living. In Proceedings of the International Symposium on Ambient Intelligence (Lecture Notes in Networks and Systems, Vol. 770, pp. 93–102). Springer. https://doi.org/10.1007/978-3-031-43461-7_10

Radlak, K., and Smolka, B. (2012). A Novel Approach to the Eye Movement Analysis Using a High Speed Camera. In Proceedings of the 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA) (145–150). IEEE. https://doi.org/10.1109/ICTEA.2012.6462854

Ren, Z., et al. (2023). VEATIC: Video-Based Emotion and Affect Tracking in Context Dataset. Arxiv Preprint arXiv:2309.06745. https://doi.org/10.1109/WACV57701.2024.00441

Siddiqui, M. F. H., Dhakal, P., Yang, X., and Javaid, A. Y. (2022). A Survey on Databases for Multimodal Emotion Recognition and an Introduction to the VIRI Database. Multimodal Technologies and Interaction, 6(6), 47. https://doi.org/10.3390/mti6060047

Swain, M., Routray, A., and Kabisatpathy, P. (2018). Databases, Features and Classifiers for Speech Emotion Recognition: A Review. International Journal of Speech Technology, 21(1), 93–120. https://doi.org/10.1007/s10772-018-9491-z

Walker, S. A., Double, K. S., Kunst, H., Zhang, M., and MacCann, C. (2022). Emotional Intelligence and Attachment in Adulthood: A Meta-Analysis. Personality and Individual Differences, 184, 111174. https://doi.org/10.1016/j.paid.2021.111174

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., and Movellan, J. R. (2014). The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. IEEE Transactions on Affective Computing, 5(1), 86–98. https://doi.org/10.1109/TAFFC.2014.2316163

Wu, C.-H., Chuang, Z.-J., and Lin, Y.-C. (2006). Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models. ACM Transactions on Asian Language Information Processing, 5(2), 165–183. https://doi.org/10.1145/1165255.1165259

Yang, Y., et al. (2023). Facial Expression Recognition with Contrastive Learning and Uncertainty-Guided Relabeling. International Journal of Neural Systems, 33(5), 2350032. https://doi.org/10.1142/S0129065723500326

Zhou, E., and Lee, D. (2024). Generative Artificial Intelligence, Human Creativity, and Art. PNAS Nexus, 3(5), pgae052. https://doi.org/10.1093/pnasnexus/pgae052

Zong, Y., et al. (2022). Adapting Multiple Distributions for Bridging Emotions from Different Speech Corpora. Entropy, 24(9), 1250. https://doi.org/10.3390/e24091250