Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7, 1235-1248
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i7.8876

Algorithmic art meets computational creativity: Reinventing interactive multimedia experiences through convergent innovation

Bowen Peng[1]*
[1]Jingdezhen Ceramic University, Jingdezhen, 333000, China; pbwjcu@163.com (B.P.).

Abstract: Traditional multimedia systems often lack real-time emotional responsiveness and visual richness. This research addresses this by proposing the Creative Interaction through Multimedia Algorithmic Convergence (CIMAC) framework, integrating algorithmic art and computational creativity to generate dynamic, user-driven visual art. CIMAC operates in four stages: 1) Kinect sensors and face recognition capture user gestures and expressions; 2) A Convolutional Neural Network (CNN) analyzes facial data to determine emotional state; 3) A Generative Adversarial Network (GAN) generates artistic visuals based on the recognized emotion; 4) The Coyote Optimization Algorithm (COA) dynamically generates and optimizes mathematical patterns (e.g., fractals, grids) for structural foundation and aesthetic harmony. The GAN-generated visuals are then fused with the COA-driven patterns to create cohesive artwork. This integrated output is displayed on a large interactive screen and continuously adapted in real-time to user movements and emotional shifts. The synergistic CNN+GAN+COA integration within CIMAC demonstrably improves output quality scores by up to 0.3 and reduces content creation and engagement time by up to 6 hours compared to traditional methods. This yields a powerful platform for highly personalized and immersive interactive multimedia experiences.

Keywords: *Computational Creativity, Emotional, Face Recognition, Interactive Multimedia, User Engagement, User Gestures, Visual Art.*

1. Introduction

Interactive multimedia communication and applications have become an essential component of how people interact, learn, and entertain themselves in today's digital age. They employ a variety of content types, including text, images, video, music, and animations, to create rich, engaging experiences with which users can interact in real time [1]. Unlike traditional multimedia, which is often linear and passive, interactive multimedia promotes active engagement by allowing viewers to control the flow and end of the information [2]. This change reflects the increased need for customized, immersive, and responsive digital environments that are tailored to the preferences and needs of individual users.

Convergent innovation advances interactive multimedia communication and applications by combining diverse disciplines and technologies to produce new, hybrid solutions [3]. It combines artificial intelligence, human-computer interaction, sensor technology, digital art, and data analytics to create intelligent, adaptive, and emotionally aware systems [4]. This multidisciplinary approach enables interactive multimedia communication to expand beyond basic input-output interactions and evolve into dynamic platforms that adapt in real-time to user emotions, gestures, voice commands, and environmental conditions [5]. This convergence enables richer, more natural multimedia communication between humans and machines, as illustrated in Figure 1.

History: Received: 8 May 2025; Revised: 23 June 2025; Accepted: 26 June 2025; Published: 15 July 2025
* Correspondence: pbwjcu@163.com

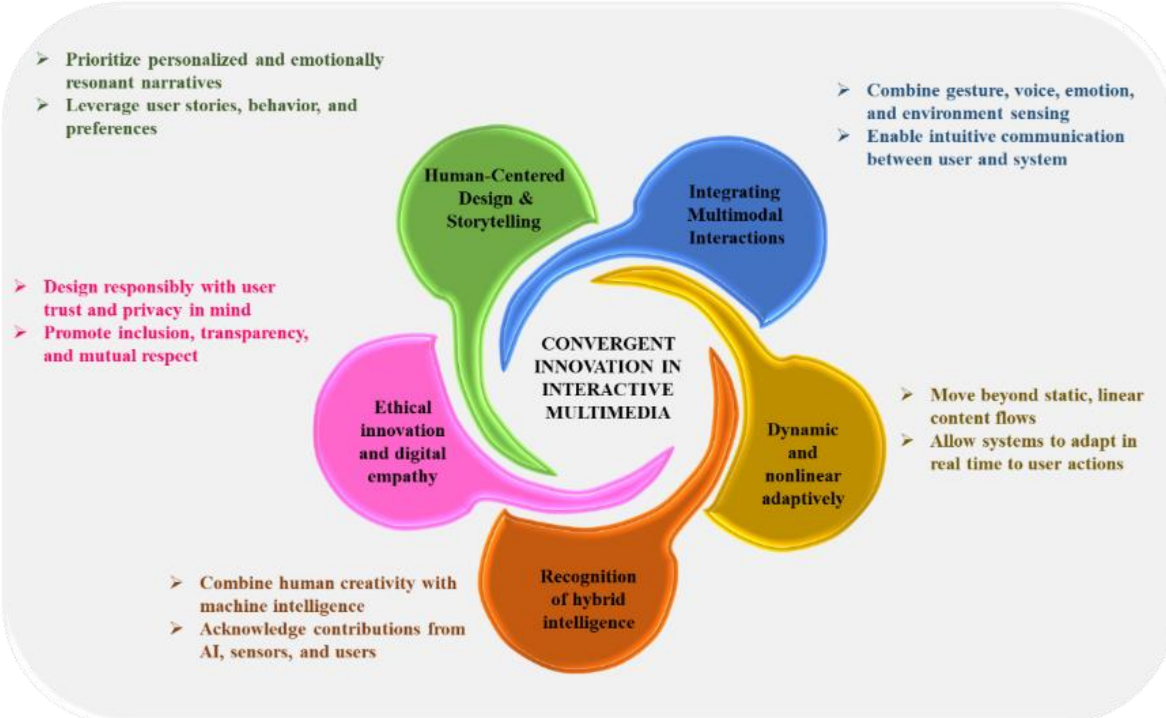

Figure 1.
Convergent innovation in interactive multimedia.

By redefining interactive multimedia with convergent innovation, Creators can develop applications that provide more meaningful, engaging, and accessible experiences [6]. Virtual reality settings, for example, can adapt to user feedback, educational platforms can tailor lessons to specific learning styles, and digital art displays can react emotionally to audience behavior [7]. Convergent innovation is converting interactive multimedia into a smarter, more human-centered technology capable of meeting the diverse and ever-changing needs of the digital age [8].

Traditional algorithmic art methods, such as SVM and Decision Tree techniques, rely on simple procedural generating processes that do not fully incorporate user feedback or sophisticated creative models. This results in less engaging and tailored experiences. This investigation proposes a CNN+GAN+COA that combines algorithmic art and computational creativity through the use of advanced optimization strategies. The platform enables the production of dynamic, real-time visual art that evolves in response to user inputs and environmental data.

This review begins with a literature analysis highlighting gaps in existing work, followed by an introduction to the dataset, the CNN model, CNN for emotion recognition, and GAN for artistic image generation. Next, the proposed COA approach is presented, along with experimental design and evaluation metrics. Finally, a comparative analysis is presented, concluding with key findings and future research directions.

2. Related Works

The rapid advancement of modern science and technology is hastening the pace of digital media communication and diversifying communication formats. To create and improve digital media art Chen [9] designers must consider intelligence, networking, and content diversity. Digital media art would encounter problems in the age of artificial intelligence (AI), a growing number of artists continue to explore, and designers are available for continuous inquiry and investigation.

DOI: 10.55214/25768484.v9i7.8876

Emotional design is increasingly appreciated by interface designers as a way to increase user comfort and happiness [10]. Cluster analysis algorithms could be utilized to handle complex cultural information as well as innovative cultural item creation. The modified method achieved the maximum data clustering effect of 90%, allowing designers to identify appropriate creative information for better products. Artificial intelligence has accelerated the art design process by retrieving multidimensional conceptual data that was examined [11]. CNNs were used to extract expressive characteristics from posters, which were subsequently grouped using a variational autoencoder (VAE).

The use of artificial neural networks (ANN) and the Internet of Things (IoT) in computer-aided design (CAD) creation and experimental design was investigated by Shi and Yuan [12]. By analyzing massive amounts of artwork data, ANN discovered creative patterns and characteristics, offering artists innovative recommendations and optimization solutions. Their technology offered a more efficient and sophisticated auxiliary tool to artists and designers.

A strategy for improving emotional expressiveness in advertising graphic design by combining deep learning algorithms and CAD modeling optimization was described [13]. The technology employed recurrent and CNN to extract emotional patterns and themes from commercials. This innovative data-driven strategy improves advertising efficacy and increases customer satisfaction while also providing useful references for designers.

The four aspects of media convergence- content consumption, product/service form convergence, content distribution channel convergence, and content creative convergence were outlined in Li and Liang [14]. It highlights the need for precise digital content services, a multifaceted user experience, three-dimensional marketing integration, and varied scene-based services.

The design of visual communication in digital technology, with an emphasis on network media and how it affects people's work and lives, was examined by Cai and Su [15]. It introduced a multi-layer network-based clustering recommendation system that incorporated qualities, time, and ratings. It explored the potential that modern media technology presents for visual communication design, such as information transmission, visual language, and aesthetics.

Existing approaches, such as the cluster analysis algorithm described in Gao and Huang [10] have limitations in capturing complex emotional and cultural contexts, resulting in diminished flexibility and creative relevance. Similarly, the ANN-based CAD system Shi and Yuan [12] provides faster processing but lacks emotional expressiveness and difficulties with delicate creative optimization. To solve these issues, the proposed CNN+GAN+COA model combines CNN's powerful capability to extract expressive features with GAN's generative creativity and COA adaptive parameter tweaking. This integrative technique increases emotional depth, grouping accuracy, and creative flexibility.

3. CIMAC Framework

The CIMAC framework is an innovative technology that combines algorithmic art and artificial intelligence to provide dynamic, emotionally responsive multimedia communication and applications. It functions in four major stages:

Stage 1 – Data Capture: Kinect sensors and facial recognition technologies record live user actions and expressions, forming a real-time dataset of gestures and emotions.

Stage 2 – Emotion Recognition: A CNN processes the captured data to accurately assess the user's emotional state.

Stage 3 – Visual Generation: A GAN creates individualized creative visuals based on the recognized emotional input.

Stage 4 – Pattern Optimization and Display: Simultaneously, the COA generates and refines mathematically structured patterns like fractals and grids, serving as the harmonious foundation for the artwork. The outputs from GAN and COA are then merged and displayed on a large interactive screen, updating in real-time based on user movements and emotional shifts. The framework for interactive multimedia experiences is shown in Figure 2.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 1235-1248, 2025
DOI: 10.55214/25768484.v9i7.8876

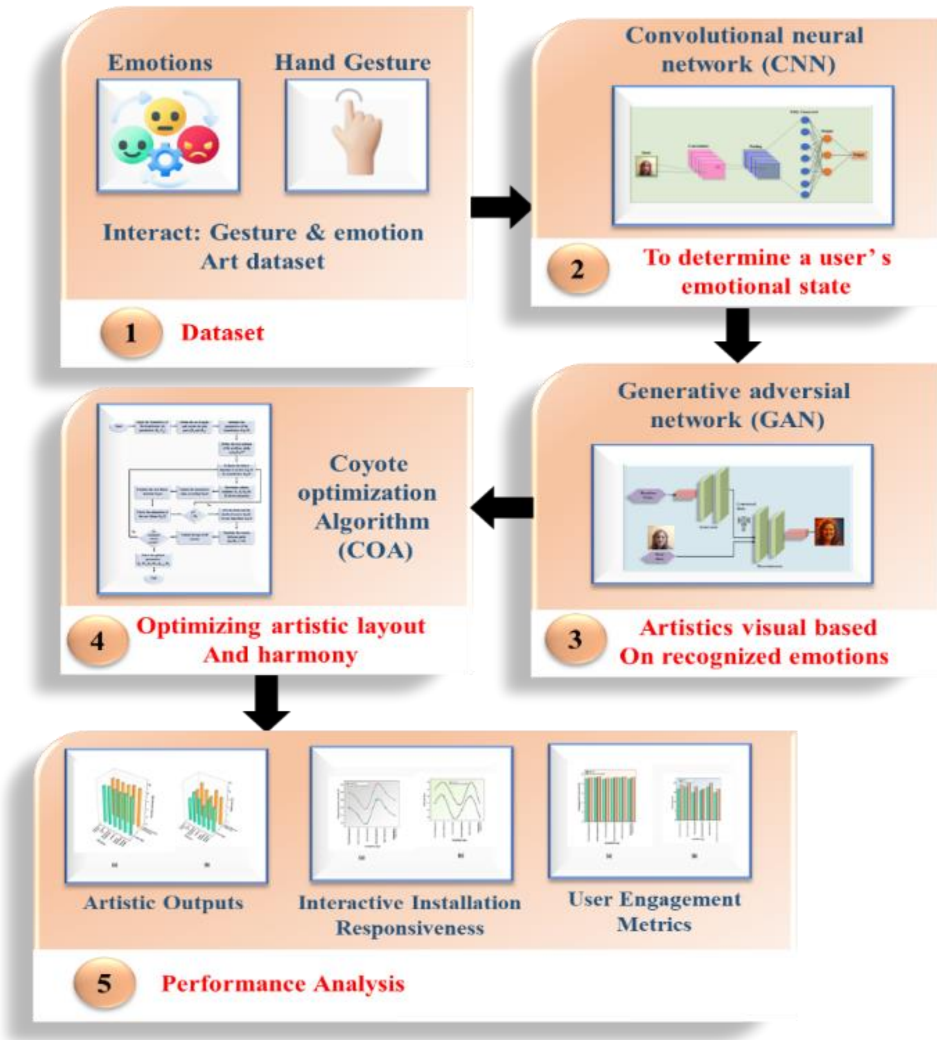

Figure 2.
CIMAC framework for interactive multimedia experiences.
Source: https://www.kaggle.com/datasets/programmer3/interact-gesture-and-emotion-art-dataset

3.1. Dataset

Interact: Gesture & expression Art Dataset is intended to help in the creation of emotionally responsive, real-time interactive multimedia art by combining dynamic hand gestures with face expression recognition. It comprises around 119,000 depth images recorded with Kinect-like sensors that record various hand motions, allowing for accurate gesture identification. In addition, the dataset contains 152 high-quality facial images tagged with emotional states, enabling classification of emotions in real time, and the dataset has been preprocessed

3.2. CNN+GAN+COA

Combining CNN, GAN, and the COA provides a ground-breaking strategy for reinventing interactive multimedia communication and applications using convergent innovation. CNNs excel in extracting rich, hierarchical characteristics from complicated multimedia data, allowing for the exact detection and interpretation of user interactions in real time. Meanwhile, GANs improve the experience by creating high-quality, dynamic multimedia material that responds smoothly to user input, resulting in a more immersive and customized environment. Integrating the COA, a nature-inspired metaheuristic optimization approach, refines this system by effectively optimizing CNN and GAN parameters, hence enhancing learning accuracy and content quality. The combination of these technologies allows for the development of intelligent multimedia systems that dynamically adapt, forecast user preferences, and deliver compelling content on demand.

3.3. CNN for Emotion Recognition

CNN enables emotion recognition in convergent innovation for interactive multimedia applications by evaluating face characteristics in real-time, hence increasing user engagement through adaptive responses that customize multimedia experiences to emotional states and expressions. Raw pixels are transformed into a scoring function that evaluates each image. Usually, the design of CNNs adds an input and then places a CONV layer, a POOL layer, a flatten layer, a dense layer, and a dropout layer in the order described in Figure 3.

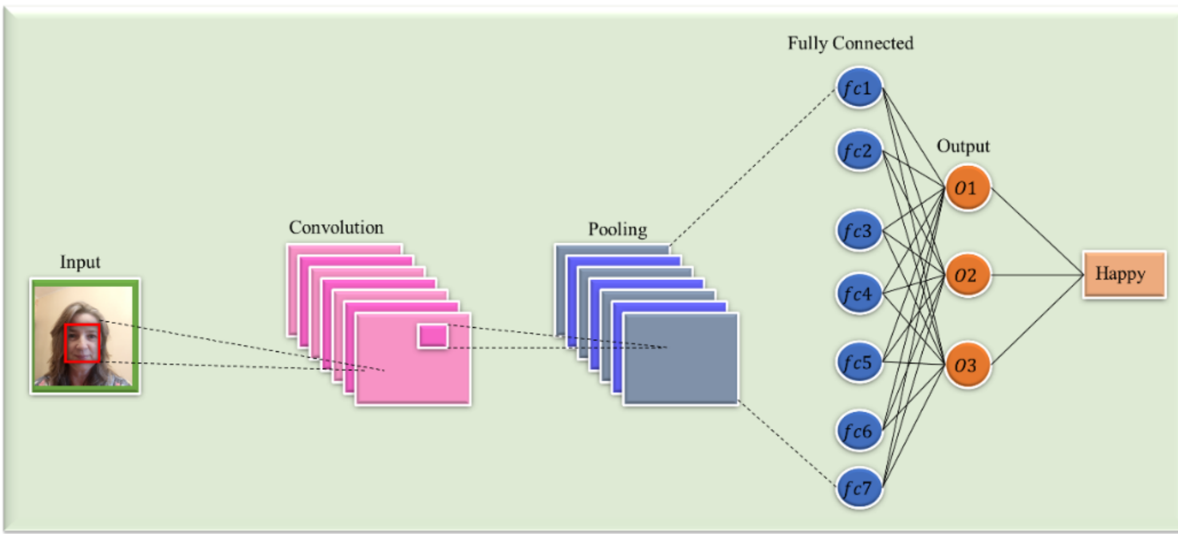

Figure 3.
CNN structure for emotion recognition.

Convolutional layer: Raw images are filtered using many sets of filters in the convolutional layers of the CNN. Most user options are included in this section. The main thing to consider is the shape and quantity of the kernels. The data is then sent to the next step.

Max pooling layer: After convolutional layers are applied, pooling layers follow. There are two main kinds of pooling algorithms: max pooling, which selects the highest value in each group of data, and average pooling, which averages the numbers in that region. Usually, these operations are carried out to lower the dimensionality of the system.

Flatten layer: Data is flattened by converting it into a one-dimensional array to be passed on to the next layer. Merging all data flattens it into a single dimension for the next layer. It connects to the fully connected layer, which acts as the final classification model.

Dense layer: The dense layer is a straightforward layer that receives input from the previous layer. Dense layers process images using the output produced by the convolutional layers. The main layers of a network are based on the use of the sigmoid and rectified linear unit (ReLU) functions. All values returned by the sigmoid function are within the range [0,1]. The function is monotonic, but the derivative is not monotonic. The logistic sigmoid function sometimes causes neural networks to stop improving during training. The most popular nonlinear activation function is the ReLU, and it is followed by convolutional layers.

Dropout layer: Dropout is used to prevent overfitting in models. During training, the outgoing edges of hidden units (neurons in hidden layers) are set to 0 with each update.

3.4. GAN For Generating Emotion-Conditioned Visuals

A GAN creates emotion-conditioned graphics to promote convergent innovation in interactive multimedia communication and applications, allowing for dynamic, user-specific creative reactions. It learns emotional traits and combines emotive images to improve real-time interactivity and aesthetic customization across a variety of multimedia events. A GAN consists of two parts: the generator model and the discriminator model. The discrimination model acts as a discriminator, attempting to distinguish between actual and simulated data generated by the model. The discriminator should indicate the output of the input image, as illustrated in Figure 4.

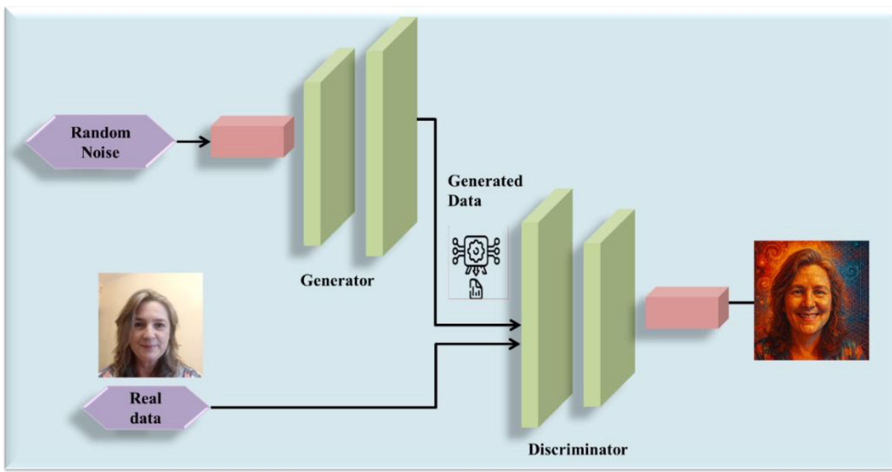

Figure 4.
GAN structure for generating emotion-conditioned visuals.

To define a generative adversarial problem, assign the probability density function of the target high-dimensional data to $o_{data(W)}$ and the probability density function of the generator to $o_{H(W;\theta)}$, where θ represents the probability density function parameters. To calculate $\theta o_{H(W;\theta)}$, which moves closer to $o_{data(W)}$. Extract n data points from $o_{data(W)}$, such as $w_{1,}w_{2,}w_{3,\ldots}w_{n}$. Finding θ Involves maximizing the greatest likelihood function $K = \prod_{j=1}^{n} o_{H(W^{j};\theta)}$. The derivation procedure is outlined below in (Eq.1).

$$\begin{aligned}\theta^* &= arg\,\underset{\theta}{max}\prod_{j=1}^{n} o_{H(W^{j};\theta)} = arg\,\underset{\theta}{max}\,log\prod_{j=1}^{n} o_{H(W^{j};\theta)}\\ &= arg\,\underset{\theta}{max}\sum_{j=1}^{n} log o_{H(W^{j};\theta)}\{w_{1,}w_{2,}w_{3,\ldots}w_{n}\} from o_{data(W)}\\ &= arg\,\underset{\theta}{max}\int_{W} o_{data(W)} log o_{H(W;\theta)} cw - \int_{W} o_{data(W)} log o_{data(W)} cw\\ &\approx arg\,\underset{\theta}{max} F_{W\sim o_{data(W)}}\left[log o_{H(W;\theta)}\right] \qquad (1)\end{aligned}$$

The task involves determining the smallest KL distance between $0_{(data)(w)}$ $and\ o(h)(w;\theta)$. Eq. (2) summarizes the complete optimization objective function.

$$\underset{H}{min}\,\underset{C}{\max}\, U(C,H) = \mathbb{F}_{W\sim o_{data(W)}}[logC(w)] + \mathbb{F}_{y\sim o_{y(y)}}\left[\log\left(1 - C\left(H(y)\right)\right)\right] \qquad (2)$$

Where H is the generator, C is the discriminator, and $U(C, H)$ is the value function. It can train H by minimizing $\log\left(1 - C\left(H(y)\right)\right)$. The discriminator C and the generator engages in a minimax contest on the value function $U(C, H)$.

3.5. Coyote Optimization for Optimizing Artistic Layout and Harmony

Coyote Optimization improves creative layout and harmony in convergent innovation for interactive multimedia communication and application by mimicking adaptive social behavior, which allows for dynamic balance, aesthetic coherence, and user-driven design. It effectively investigates visual possibilities for streamlined, responsive, and emotionally engaging multimedia experiences. COA is a novel nature-inspired metaheuristic algorithm. The algorithm is population-based and influenced by the coyote's social structure and environment. COA is classed as both swarm intelligence and evolutionary heuristic. Coyotes maximize their function by forming social groups and sharing experiences. The coyote population is made up of MO packs, with MD coyotes in each one. The total number of coyotes in all packs is the population for the optimization issue.

The response to the optimization issue ensures optimal adaptability to all social contexts. The social circumstances of coyotes represent c-space choice variables in the optimization problem and the flowchart are shown in Figure 5.

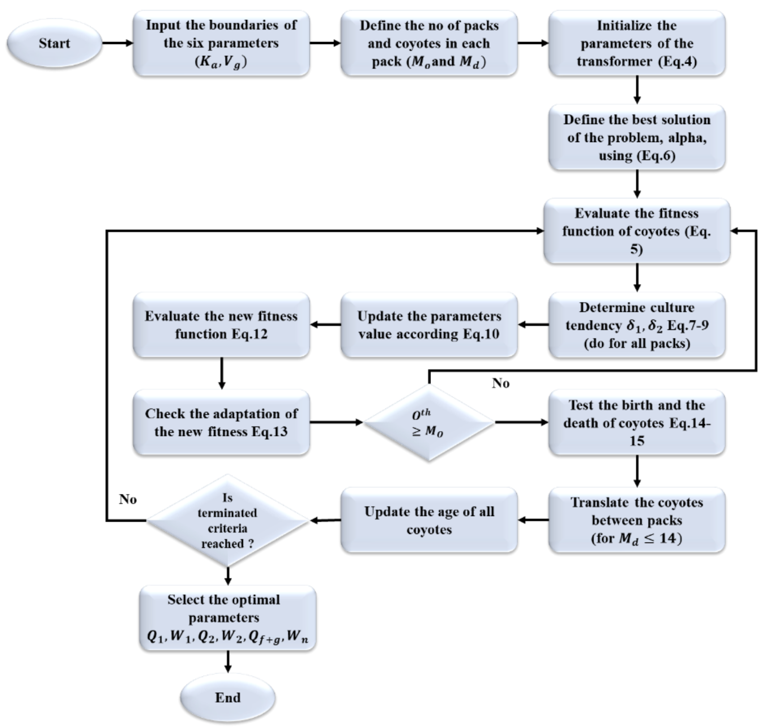

Figure 5.
COA flowchart.

The social condition ' soc' of the coyote in the o^{th} pack at the s^{th} moment is $soc_d^{o,s}$. The coyote's conditions reflect choice variables W in a global optimization problem. It's provided as in (Equation 3).

$$soc_d^{o,s} = \bar{W} = \left(w_1, w_{2,} \ldots \ldots . w_c\right) \tag{3}$$

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 1235-1248, 2025
DOI: 10.55214/25768484.v9i7.8876

The social conditions for each coyote d^{th} of o^{th} are randomly initiated at an instant s^{th} and i^{th} dimensions within the range of the decision variable's lower and upper bounds, LB_i *and* UB_i, as follows in (Eq. 4).

$$soc_{d,i}^{o,s} = LB_i + q_i \times (UB_i - LB_i) \qquad (4)$$

Where q_i is a real random number in the $[0-1]$ range created with uniform probability.

The desired function is determined by assessing the coyote's conditions for the present choice variables, as shown below in (Eq. 5)

$$fit_d^{o,s} = f\left(soc_d^{o,s}\right) \qquad (5)$$

Coyotes' social system allows them to quit or join another pack based on the present coyote, M_d (up to 14 coyotes). For the global population, 'alpha' is the optimal solution to the optimization issue at the time s^{th} of packet o^{th}. This is determined as follows in (Equation 6).

$$alpha^{o,s} = \left\{soc_d^{o,s} | arg_{d=(1,2,\ldots M_d)} mine\left(soc_d^{o,s}\right)\right\} \qquad (6)$$

The COA facilitates the exchange of social circumstances and information across the global population. The COA calculates the pack's cultural tendency, which represents the average social status of all coyotes in that pack (Eq.7).

$$cult_i^{o,s} = \left\{\begin{matrix} P_{\frac{(M_d+1)}{2},}^{o,s}, i, \ M_d \text{ is odd} \\ \dfrac{P_{\frac{M_d}{2}i}^{o,s} + + P_{\frac{(M_d+1)}{2}i}^{o,s}}{2}, \ otherwise \end{matrix}\right\} \qquad (7)$$

The ranking decision variables (i.e., social conditions) of all coyotes inside the o^{th} pack at the s^{th} instant for every i in the space of decision variables, c, are represented by $p^{o,s}$.

Coyotes' social circumstances (new_soc) are influenced by two factors: alpha influence (δ_1) and cultural tendency influence (δ_2), as shown below in (Equation 8-10).

The influence δ_1 is calculated as the difference between a random coyote (C_{q1}) within the pack and the alpha coyote. Pack influence (δ_2) refers to the difference between a random coyote (C_{q2}) and the pack's cultural tendency.

$$\delta_1 = alpha^{o,s} - soc_{dq1}^{o,s} \qquad (8)$$

$$\delta_2 = cult^{o,s} - soc_{dq2}^{o,s} \qquad (9)$$

$$new\ soc_d^{o,s} = soc_d^{o,s} + q_1\delta_1 + q_2\delta_2 \qquad (10)$$

Where q_1 and q_2 are uniformly random values in the range of 0 *to* 1.

The objective function's new value is calculated by evaluating new social situations, as shown below in (Eq. 11).

$$new\ fit_d^{o,s} = e\left(new\ soc_d^{o,s}\right) \qquad (11)$$

At the next $(s+1)^{th}$ time instant, the choice about new social conditions is made based on the value of the objective function, as shown below in (Equation 12):

$$soc_d^{o,s} = \left\{\begin{matrix} new\ soc_d^{o,s}, new\ fit_d^{o,s} \le fit_d^{o,s} \\ soc_d^{o,s} \qquad otherwise \end{matrix}\right\} \qquad (12)$$

The optimal social conditions for a coyote to adapt to its surroundings are the worldwide response to the problem. To maintain pack size, COA calculates the ages of all coyotes (in years) as $age_d^{o,s} \in M$. The birth of a young coyote is determined by the social circumstances of two randomly selected parents within a pack in (Equation 13).

$$soc_i^{o,s} = \begin{Bmatrix} soc_{q_1,i.}^{o,s}\ q\ and\ i\ \leq o_t or\ i = i_1 \\ soc_{q_1,i.}^{o,s}\ q\ and\ \geq (o_t + o_b) or\ i = i_2 \\ Q_i\ otherwise \end{Bmatrix} \quad (13)$$

Therefore, q_1 and q_2 are random coyotes inside the o^{th} pack. i_1 and i_2 indicate two randomly chosen dimensions of the optimization problem. Equation (14-15) provides the scatter and association probabilities, denoted as o_t and o_b, respectively. Q_i is a random number that falls inside the decision variable bound of the i^{th} dimension. The true random number generated by c_i ranges from $0\ to\ 1$ and follows a uniform probability.

$$o_t = \frac{1}{c} \quad (14)$$

$$o_b = \frac{1 - o_t}{2} \quad (15)$$

4. Results and Discussion

Python was used to implement the proposed CNN+GAN+COA algorithm and compare it against the Variational Autoencoder and Particle Swarm Optimization (VAE+PSO) model [16]. The findings indicate increased artistic performance, responsiveness in interactive displays, and increased user engagement. CNN+GAN+COA provides richer visuals and real-time adaptation, making it more effective for convergent innovation in interactive multimedia communication and applications.

The training and testing performance of the CNN+GAN+COA model created for emotionally responsive multimedia systems is shown in Figure 6. Figure 6a illustrates that training and testing accuracy increase significantly with each epoch, reaching over 99% by the 50th epoch. This shows that the model effectively learns from the data and accurately predicts user inputs such as gestures or emotions. Figure 6b shows that the training and testing loss decrease continuously throughout epochs, indicating that the model is minimizing mistakes while learning. The close alignment of training and testing curves in both graphs indicates strong generalization with no significant overfitting.

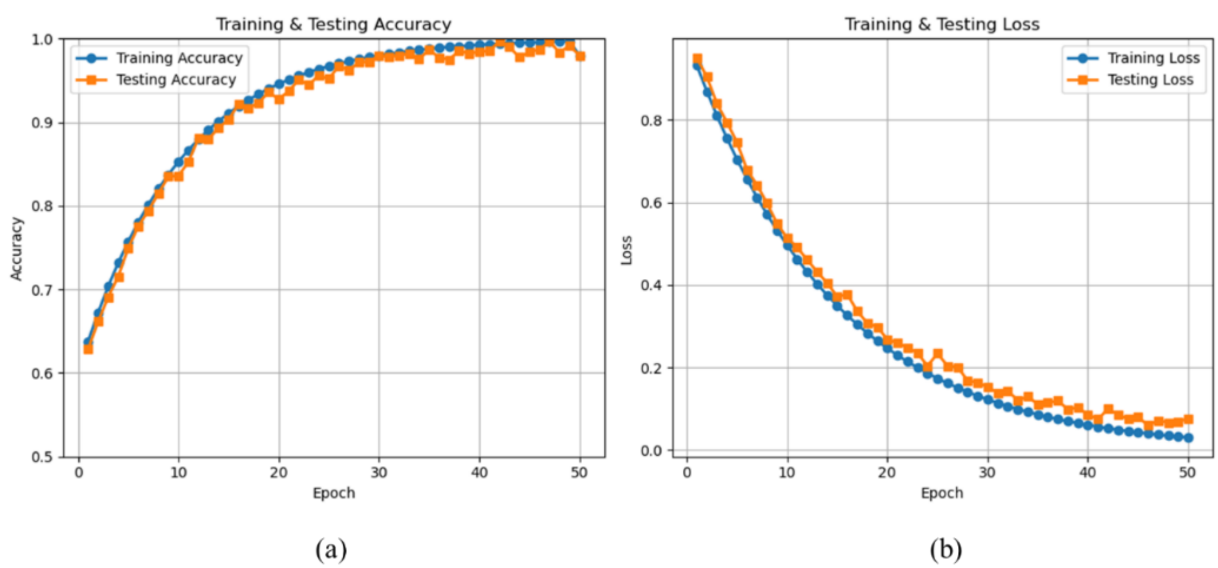

Figure 6.
Training and testing for interactive multimedia experiences (a) accuracy (b) loss.

The effects of several generative model combinations, CNN + GAN + COA, CNN + GAN, GAN + COA, and GAN Only, on visual complexity scores across user emotional states and interactive art sessions are presented in Figure 7. The suggested CNN + GAN + COA model produces consistently decreased visual complexity error scores, indicating reliable and refined results. This demonstrates the strength of convergent innovation, where integrating deep learning and optimization techniques enhances emotional reactivity and visual richness in interactive multimedia experiences.

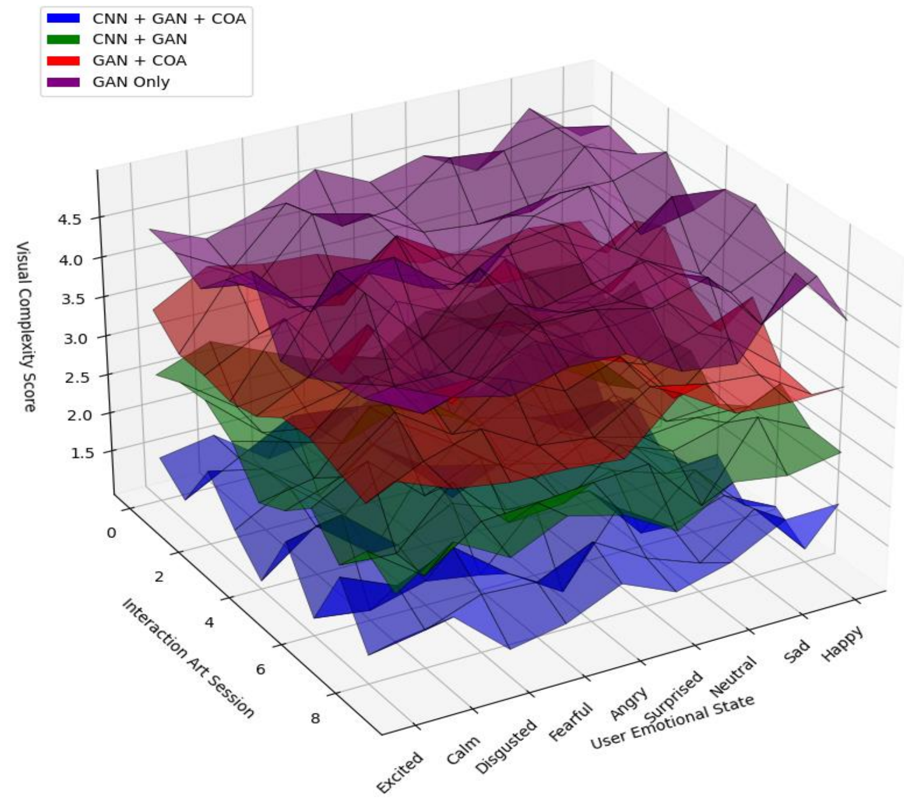

Figure 7.
Visual complexity scores across user emotional states and interaction sessions.

4.1. Comparison With Existing Methods

Reinventing interactive multimedia experiences involves integrating modern technology, such as AI, emotion detection, and gesture control, to create compelling, responsive environments. VAEs often generate fuzzy or less detailed outputs due to approximation inference. PSO could encounter early convergence and get entangled in local optima, lowering optimization quality. Combining them increases computing complexity and training time, which could limit real-time responsiveness and scalability in dynamic multimedia environments. The integration of CNN, GAN, and the COA improves interactive multimedia by increasing recognition of features, creating realistic content, and optimizing efficiency. CNN handles deep visual understanding, GAN generates dynamic and creative material, and COA fine-tunes system settings for real-time efficiency and flexibility. This collaboration provides engaging, intelligent, and responsive user experiences across a variety of multimedia platforms.

The artistic performance of the proposed and existing methods is shown in Figure 8 and Table 1. The comparison demonstrates that the suggested CNN+GAN+COA method outperforms the VAE+PSO strategy in all art forms in Figure 8a. Scores increased from 9.1 to 9.4 for abstract visuals, 9.3 to 9.6 for digital paintings, 9.0 to 9.2 for patterns, 9.2 to 9.3 for generative sculptures, 9.4 to 9.6 for kinetic art, and 9.1 to 9.4 for interactive murals, indicating that the proposed model improved creative quality and aesthetic appeal. Figure 8b compares the time required to create different art forms using two methods, VAE+PSO and the proposed CNN+GAN+COA. The timeframes for abstract visuals are 15 and 12 hours, digital paintings are 18 and 14 hours, patterns are 13 and 11 hours, generative sculptures are 16 and 13 hours, kinetic art is 19 and 15 hours, and interactive murals are 14 and 10 hours. These findings demonstrate that the proposed method not only improves artistic quality but also reduces generation time, likely due to its more efficient and advanced model architecture.

Table 1.
Values for artistic performance comparison with existing and proposed methods.

Art form	**(Quality score/10)**		**time (hours)**	
	VAE+PSO (Quality score/10)	**CNN+GAN+COA [Proposed]**	**VAE+PSO time (hours)**	**CNN+GAN+COA [Proposed]**
Abstract visuals	9.1	9.4	15	12
Digital paintings	9.3	9.6	18	14
patterns	9	9.2	13	11
Generative sculptures	9.2	9.3	16	13
kinetic art	9.4	9.6	19	15
Interactive murals	9.1	9.4	14	10

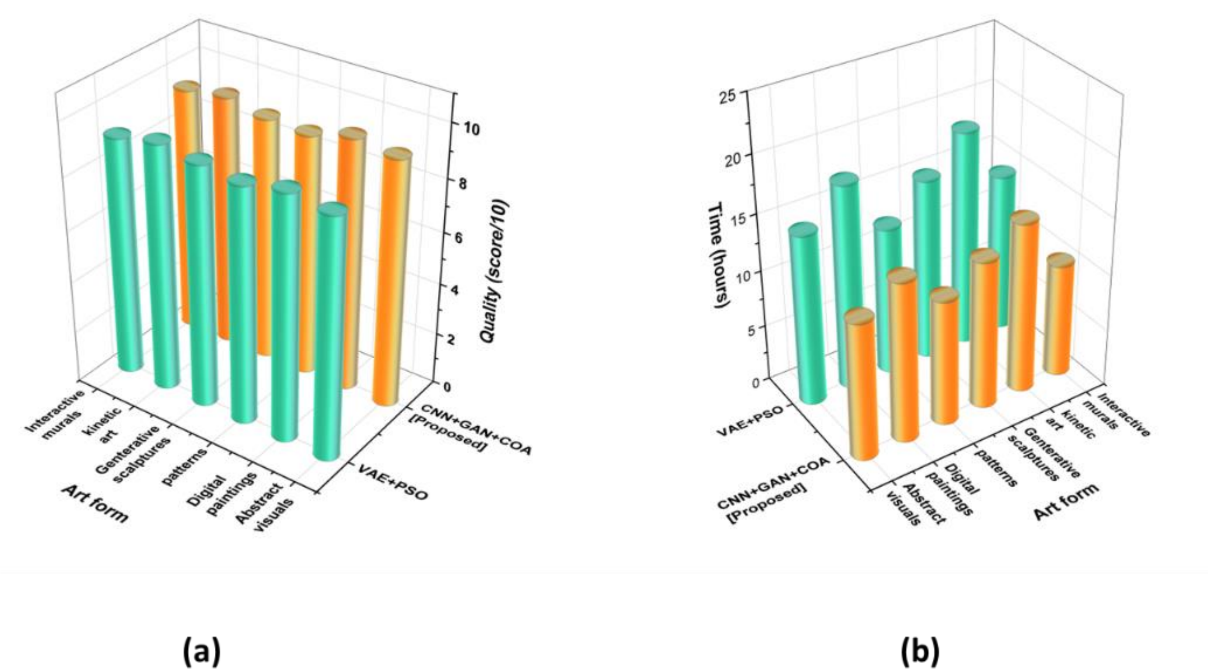

Figure 8.
Artistic performance comparison for existing and the suggested technique (a) Quality score (b) time.

The interactive installation's responsiveness performance is shown in Figure 9. In Figure 9a, the proposed CNN+GAN+COA model outperforms the VAE+PSO method across all installation types, with touch-responsive increasing from 9.2 to 9.4, motion-responsive from 9.1 to 9.3, environmental responsive from 9.0 to 9.2, sound-responsive from 9.3 to 9.5, light-responsive from 9.2 to 9.4, and temperature-responsive from 9.1 to 9.3, demonstrating increased interactivity and responsiveness.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 1235-1248, 2025
DOI: 10.55214/25768484.v9i7.8876

Figure 9b compares the installation timeframes in hours for two methods: VAE+PSO and the proposed CNN+GAN+COA, across the six types of responsiveness. VAE+PSO takes 17 hours to be touch-responsive, whereas the proposed takes 13 hours. Motion-responsive takes 18 hours vs 14 hours; environmental responsiveness takes 15 hours versus 13 hours; sound-responsive takes 16 hours versus 12 hours; light-responsive takes 18 hours versus 15 hours; and temperature-responsive takes 15 hours for VAE+PSO and 11 hours for CNN+GAN+COA. Overall, the proposed solution significantly reduces installation time in all categories while improving performance.

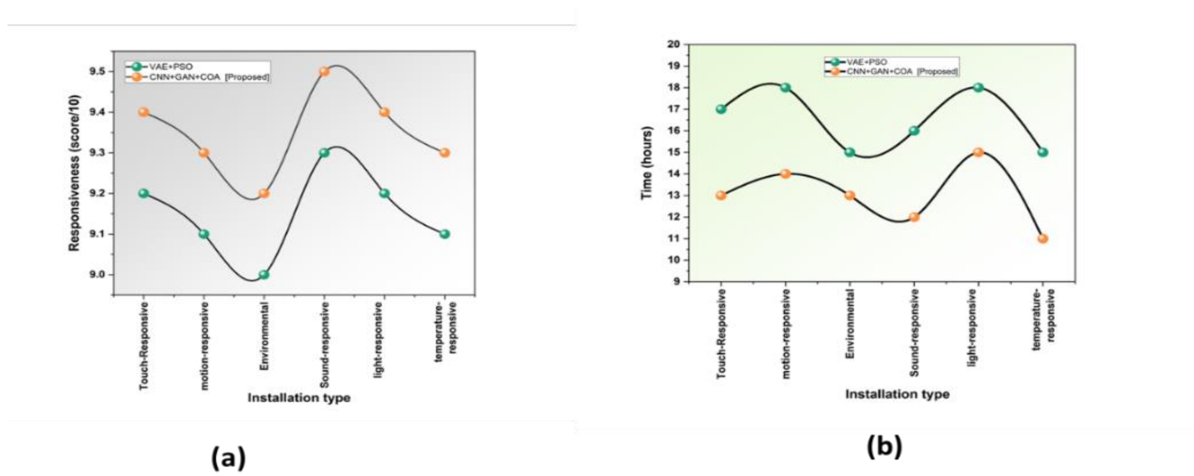

Figure 9.
Comparative performance for proposed and traditional methods in interactive installation responsiveness (a) responsiveness score, (b) time.

The user engagement performance for the suggested and existing methods is shown in Figure 10. The suggested CNN+GAN+COA approach outperforms VAE+PSO across all installation types. In Figure 10a, CNN+GAN+COA scores 9.3 and 9.1 for touch, 9.5 and 9.3 for motion, 9.2 and 9.0 for environmental, 9.4 and 9.2 for sound, 9.5 and 9.3 for light, and 9.4 and 9.1 for temperature. These results show that CNN+GAN+COA consistently increases responsiveness across all test categories. Figure 10b compares user engagement in hours for two installation types, VAE+PSO and the proposed CNN+GAN+COA, across several response categories. Engagement in touch-responsive installations is 17 hours with VAE+PSO and 12 hours with CNN+GAN+COA. Motion-responsive engagement increases from 18 to 12 hours, environmental from 15 to 10 hours, sound-responsive from 16 to 11 hours, light-responsive from 18 to 12 hours, and temperature-responsive from 15 to 11 hours. Thus, the CNN+GAN+COA approach enhances responsiveness and interaction quality, even with reduced engagement duration.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 7: 1235-1248, 2025
DOI: 10.55214/25768484.v9i7.8876

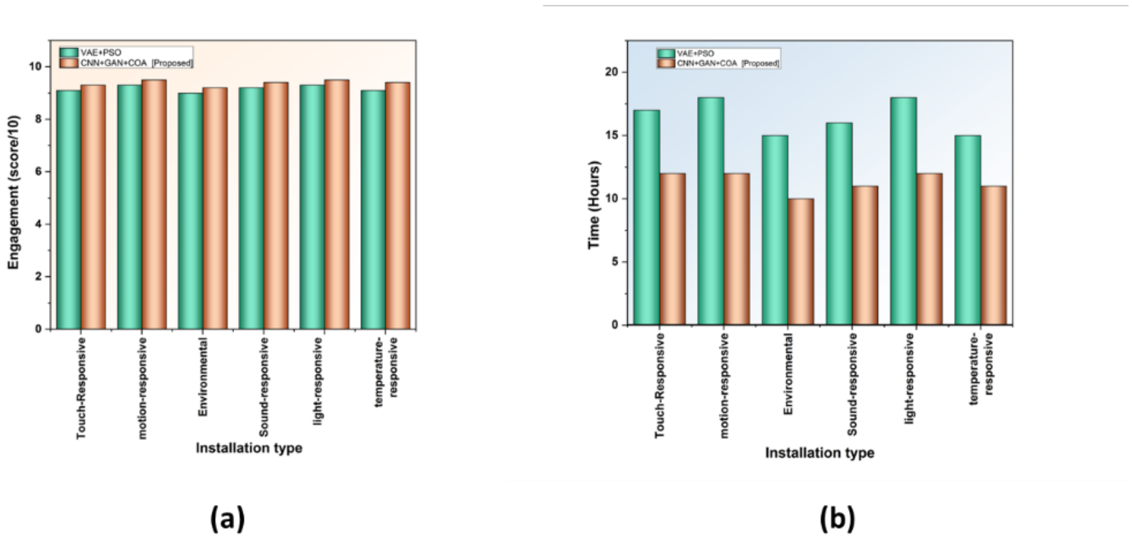

Figure 10.
Comparison of performance for traditional and suggested techniques in user engagement (a) engagement score, (b) time.

5. Conclusion

Algorithmic art and computational creativity are revolutionizing the design of interactive multimedia communication and application experiences, allowing artworks to adapt dynamically to user input. This research presents the CIMAC framework, which consists of four stages. Kinect and facial recognition technologies capture live user gestures and reactions. A CNN uses facial data to predict a user's emotional state. A GAN generates creative graphics based on detected emotions. The COA is used to create dynamic, mathematically driven patterns like fractals and grids, which serve as the foundation for the artwork. The proposed CNN+GAN+COA model outperforms VAE+PSO in terms of artistic quality (scores up to 9.6) while also reducing production and installation times by up to 5 hours. It also improves user engagement and response, reducing total interaction duration by up to 6 hours across all installation types. A drawback of this investigation is the additional processing and installation time caused by sophisticated model structures. Future goals include enhancing model efficiency and increasing real-time adaptive capabilities to provide more immersive interactive multimedia experiences.

Transparency:

The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] G. Zhang, "Visual communication design based on collaborative CAD and visual importance under the impact of new media art," *Computer-Aided Design and Applications*, vol. 15, pp. 15–30, 2025.

[2] Z. Ma, J. Guan, and R. Li, "Research on innovative teaching mode of art education in the age of convergence of media," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 2, pp. 272-284, 2021. https://doi.org/10.3991/ijet.v16i02.18235

[3] H. Tian, "Optimization of hybrid multimedia art and design teaching mode in the era of big data," *Scientific Programming*, vol. 2021, no. 1, p. 8266436, 2021. https://doi.org/10.1155/2021/8266436

[4] Z. Jie, S. Dalal, M. S. A. Bakar, and W. A. W. Yahaya, "Designing for sustainability: Exploring the intersection of creative arts in digital sustainable ecology by interactive animation," *Idealogy Journal*, vol. 9, no. 1, 2024.

[5] C. Yang and S. S. B. Singh, "User experience in information system platforms: A study on learning styles and academic challenges," *Journal of Internet Services and Information Security*, vol. 14, no. 4, pp. 209-223, 2024. https://doi.org/10.58346/JISIS.2024.I4.012

[6] R. A. Sánchez-Ancajima, S. M. Peres, J. A. López-Céspedes, J. L. Saly-Rosas-solano, R. M. Hernández, and M. A. Saavedra-López, "Gesture phase segmentation dataset: An extension for development of gesture analysis models," *Journal of Internet Services and Information Security*, vol. 12, no. 4, pp. 139-155, 2022. https://doi.org/10.58346/JISIS.2022.I4.010

[7] Q. Jiang and J. Chung, "A research of user experience on multi-modal interactive digital art," *International Journal of Internet, Broadcasting and Communication*, vol. 16, no. 1, pp. 80-85, 2024. https://doi.org/10.7236/IJIBC.2024.16.1.80

[8] G. B. Takala, "The interactive creativity of the digital era: Exploring how media art redefines the relationship between audience and artwork," *Studies in Art and Architecture*, vol. 2, no. 3, pp. 28-44, 2023. https://doi.org/10.56397/SAA.2023.09.04

[9] C. Chen, "Study on the innovative development of digital media art in the context of artificial intelligence," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 1004204, 2022. https://doi.org/10.1155/2022/1004204

[10] Z. Gao and J. Huang, "Human-computer interaction emotional design and innovative cultural and creative product design," *Frontiers in Psychology*, vol. 13, p. 982303, 2022. https://doi.org/10.3389/fpsyg.2022.982303

[11] X. Rui, "A convolutional neural networks based approach for clustering of emotional elements in art design," *PeerJ Computer Science*, vol. 9, p. e1548, 2023. https://doi.org/10.7717/peerj-cs.1548

[12] H. Shi and T. Yuan, "Creative and experimental design of artwork based on artificial neural network algorithm and internet of things," *Computer-Aided Design and Applications*, vol. 21, pp. 92-104, 2023.

[13] W. Guo and H. Ye, "Emotional analysis and expression in advertising art design based on deep learning algorithms," *Computer-Aided Design and Applications*, vol. 15, pp. 147–162, 2024.

[14] J. Li and Y. Liang, "Research on the application of digital media technology in media convergence," presented at the IEEE Sixth International Conference on Data Science in Cyberspace (DSC) (pp. 462–465). IEEE. https://doi.org/10.1109/DSC53577.2021.00074, 2021.

[15] J. Cai and J. Su, "Application characteristics and innovation of digital technology in visual communication design," *Advances in Multimedia*, vol. 2022, no. 1, p. 8806770, 2022. https://doi.org/10.1155/2022/8806770

[16] J. Xie, M. Yu, and G. Liu, "Fusing algorithms for intersection of computer science and art: Innovations in generative art and interactive digital installations," *IEEE Access*, vol. 12, pp. 173255-173267, 2024. https://doi.org/10.1109/ACCESS.2024.3488398