

Research Article

Kun Luo and Ran Tao*

Creative design of digital media art based on computer visual aids

<https://doi.org/10.1515/pjbr-2025-0009>

received September 29, 2024; accepted June 13, 2025

Abstract: In the digital age, artistic creation needs to strike a balance between automation and personalization. This study proposes an innovative model that integrates generative adversarial networks, computer vision, and personalized adjustment technology. Through multistage iterative optimization, efficient art generation and personalized style customization are achieved. The model uses an automated generation module to generate a draft and guides the conditional vector to achieve fine-grained adjustment of the image so that the work maintains both technical innovation and the artist's unique style. The model performance is optimized in four stages: data preparation, model training, personalized adjustment, and evaluation feedback. The actual art project "Echoes in the Mirror" is used as a case to verify the actual application effect of the model. The evaluation shows that the work receives high scores in clarity, color accuracy, style coherence, and innovation (the average score is close to 9 points). Audience feedback shows that the model performs well in enhancing immersive experience, emotional resonance, and interactive satisfaction, whereas technical acceptance also highlights room for optimization. The research results not only demonstrate the potential of automated and personalized models in artistic creation but also provide practical guidance for the deep integration of art and technology in the future and promote artistic creation to move toward the two-way improvement of innovation and audience experience.

Keywords: computer vision, digital media, art creation design, generative adversarial networks, conditional generative models

1 Introduction

In the era of digital transformation, computer vision, as a multidisciplinary field that integrates image processing, pattern recognition, and machine learning, has penetrated all levels of society at an unprecedented speed, especially in artistic creation. The core of computer vision is to give machines "visual perception" ability, enabling them to interpret and understand the information in images and videos, thereby simulating the complex visual system of humans. The development of this technology has not only promoted scientific and technological progress but also opened new creative fields for digital media, triggering extensive discussions on the deep integration of art and technology [1,2].

In recent years, the application of computer vision technology in digital media has gradually shifted from the experimental stage to the mature stage and has shown a diverse development trend [3]. On the one hand, academia and industry have conducted in-depth research on the theoretical basis and implementation methods of technology in artistic creation, revealing its profound impact on the methodology and aesthetic concepts of artistic creation. On the other hand, artists and designers have realized a variety of innovative expressions, such as dynamic image art, data visualization art, and biometric art, through computer vision technology in practice. These achievements not only show the charm of technology but also reflect artists' deep insight into human society, culture, and nature [4].

With the continuous optimization of computer vision algorithms and the rapid improvement in computing power, artists have begun to consider the ethical dimensions and social responsibilities of technology in the creative process. For example, while art projects based on facial recognition create amazing visual effects, they also trigger discussions about privacy protection and data security, prompting artists to examine the boundaries of technology applications more carefully. In addition, the education sector has gradually attached importance to teaching in this field, and many art schools have opened

* **Corresponding author: Ran Tao**, Art Institute, Anhui University of Finance & Economics, Bengbu, 233030, China, e-mail: taoran126@outlook.com

Kun Luo: College of Art and Design, Nanjing Forestry University, Nanjing, 437100, China

relevant courses to cultivate talents combining technical capabilities and artistic literacy to meet the needs of future art creation.

Although computer vision technology is increasingly used in digital media, this interdisciplinary field still lacks systematic research. The specific role of technology in art creation, especially its actual impact on the art creation process, the creative thinking of artists, and the artistic experience of the audience, is urgently needed. Moreover, the future development direction of computer vision technology in art and its potential changes to the concept of art also need further research and analysis.

This study aims to comprehensively explore the role and impact of computer vision technology in digital media art creation, focusing on its inspiration for the art creation process and creation methods, the changes in artists' thinking patterns, and improvements in the audience experience. Moreover, this study combines actual cases to explore the future development possibilities of computer vision technology in art, proposes forward-looking insights into artistic concepts and formal innovations, and helps artists and theoretical researchers better understand and apply this technology.

The innovation of this study is that it systematically studies the combination of computer vision and digital media art at both the theoretical and practical levels. First, it deeply analyzes the application of this technology in actual creation and reveals its profound impact on artistic creation methods and aesthetic concepts; second, it focuses on privacy protection, ethical considerations and social responsibilities and incorporates these factors into the discussion of the combination of technology and art; and third, through the exploration of future development directions, it proposes new ideas on how computer vision technology can promote the transformation of artistic concepts and creative forms [5].

This study is highly important for theory and practice. On the one hand, it can help artists better exploit the potential of computer vision and stimulate more possibilities for artistic creation; on the other hand, by focusing on the ethical and social issues of technology in artistic applications, it provides a reference for the responsible use of technology. Moreover, this study provides a theoretical basis for interdisciplinary collaboration in computer science and art and promotes the deep integration of the two. In addition, the research results provide a reference for art education, promote the construction of relevant curriculum systems, and cultivate high-quality compound talent that meets the needs of future digital art. This will inject new impetus into the continuous integration of technology and art and the innovative development of digital media art.

The innovative work of this article mainly includes systematically studying the combination of computer vision and digital media art from the theoretical and practical levels, deeply analyzing its application in actual creation and its impact on artistic creation methods and aesthetic concepts; proposing a model for balancing automation and personalization in artistic creation, which consists of an automatic generation module and a personalized adjustment module. The model is automated through generative adversarial networks (GANs) and uses conditional vectors to guide personalized adjustment. At the same time, it uses computer vision image segmentation, target detection and other technologies to accurately adjust the image, and uses pre-trained models, deep convolutional generative adversarial networks (DCGANs) and conditional generative adversarial networks (CGANs) to optimize models; using the "Mirror Echo" project [6] as a case to verify the model, it has achieved good results in multiple evaluations and is superior to some mainstream models in personalized matching. It also explores the application direction of the model in multiple fields and future research plans, such as distributed clustering and online recommendation system integration, adaptive model optimization, cross-platform compatibility, etc.

2 Literature review

2.1 Computer vision

Computer vision techniques, especially deep learning methods, have become important drivers of artistic innovation. Deep neural networks, such as convolutional neural networks (CNNs), are capable of extracting multi-level features from images through their unique layer-by-layer architecture [6,7]. The architecture of CNNs consists of a series of convolutional, pooling, and fully connected layers, which work together on the image data to enable high-level abstract feature extraction and classification. One of the most compelling applications of CNNs in art creation is image style migration, which allows the visual style of one image to be applied to the content of another image [8]. The objective function of image style migration is a complex optimization problem that centers on balancing content loss and style loss. The content loss ensures that the output image retains key features of the source image, whereas the style loss induces the output image to adopt the visual characteristics of the target style image. This process is shown in equation (1) [9].

$$E(x) = \alpha C(x) + \beta S(x), \quad (1)$$

where $C(x)$ and $S(x)$ measure content similarity and style similarity, respectively, and where α and β are the weights used to adjust the relative importance of the two. Through iterative optimization, an x that minimizes $E(x)$ is found, which is the final style migration result.

Computer vision technology has not only revolutionized the technical aspects of art creation but also broadened the boundaries of artistic expression. Artists can now use algorithms to automatically analyze and draw on the aesthetic characteristics of historical artworks to create unique works that blend traditional and modern elements [10]. For example, a computer vision analysis of the brush strokes and color patterns of Van Gogh's *Starry Night* can be applied on top of any modern photograph to produce stunning results. The art experience is further enriched by real-time image processing and interactive systems that allow for immediate interaction between the artist and the viewer, creating dynamic, responsive art installations. In addition, deep learning techniques make the personalization of artworks a reality, where models can learn and mimic the style of a specific artist to create customized artworks that meet individual tastes and satisfy diverse aesthetic needs [11].

This study aims to fill the aforementioned gap by exploring artists' creative thought processes when computer visualization techniques are used through empirical research. We plan to design a series of experiments in which artists are invited to create artwork using specific computer vision tools and collect their feedback and experiences during the creative process. These data will help us understand how technology reshapes the thought patterns of artistic creation and how artists adapt and innovate with new technologies.

2.2 Theoretical foundations

Computer vision, as an interdisciplinary subject, integrates several fields, such as mathematics, signal processing, pattern recognition, and machine learning, with the goal of enabling computers to understand and interpret the visual world as humans observe it through their eyes. In this process, computer vision technology relies on a series of core algorithms and models, the most famous of which are CNNs, target detection, and image segmentation techniques [12].

CNNs are signature architectures used in deep learning for processing image data. CNNs can efficiently capture spatially hierarchical features in images through

their unique structure of convolutional, pooling, and fully connected layers. The convolutional layer is responsible for extracting local features, the pooling layer performs feature degradation and spatial invariance processing, and the fully connected layer is used for classification or regression [13]. CNNs have demonstrated excellent performance in tasks such as image classification, target detection, and semantic segmentation [14], and especially in art creation, they are used to generate images with a specific style or to analyze and mimic the stylistic features of artworks. Target detection is another key technique in computer vision aimed at recognizing and localizing multiple object classes in an image. It is usually divided into two steps: region proposal and classification. In recent years, the speed and accuracy of target detection have significantly improved with the proposal of algorithms such as Faster RCNN and YOLO. In the art field, target detection techniques can be used to automatically label key elements in artwork, provide creative inspiration for the artist, or help viewers better understand the themes and symbols in artwork. Image segmentation is the process of subdividing an image into multiple parts, each of which corresponds to a specific object or region of the image [15]. This technique is crucial for understanding and interpreting complex scenes. Traditional image segmentation methods include threshold-based, region-based, and edge-based segmentation algorithms. However, in recent years, deep learning techniques have made significant progress in image segmentation, and they are capable of generating accurate pixel-level segmentation maps, which are important for detailed analysis and style transformation of artwork [16].

2.3 Integration of art and technology

As an emerging art form in the twenty-first century, digital media art not only integrates the aesthetic principles of traditional art but also absorbs the innovative power of science and technology, forming a unique and diversified means of expression. The theoretical foundation of this art form is deeply rooted in postmodernism, new media theory, and cultural research, and at the same time, the fusion of art and technology has triggered profound philosophical thinking. Postmodernism emphasizes deconstruction, collage, and surreality, concepts that are vividly reflected in digital media art. Artists utilize digital tools to break down media boundaries and mix different art forms and media to create comprehensive artwork that transcends single-sensory experiences [17]. For example, the combination of sound, image, video, and interactive

installations allows the viewer to not only view the art but also participate in it through their body and become a part of it. New media explores how media technologies are changing the way we perceive and the mode of information dissemination. Digital media art redefines the participatory and accessible nature of art through technologies such as the internet, virtual reality, and augmented reality, enabling artwork to be shared and communicated globally across physical space [18]. The fusion of art and technology involves not only superficial collaboration but also the essence of art and the core of human creativity at a deeper level. From a phenomenological perspective, art reveals existence itself, whereas technology reveals the possibility of existence. Digital media art allows the audience to experience an unprecedented dimension of reality through technological means, and this experience is both a recreation of art and an excavation of the potential of science and technology [19]. With the application of cutting-edge technologies such as artificial intelligence and bioengineering in the arts, issues of ethics and responsibility are becoming increasingly prominent. Artists and technologists need to consider how their work affects society, especially in terms of privacy, data security, and AI bias [20]. Art is not only an aesthetic pursuit but also an expression of social responsibility. Rain Room, an installation created by Landon International Art Collective, explores the relationship between nature and technology and between control and freedom by using advanced sensors and projection technology to allow viewers to walk in rain without becoming wet [21]. A study used Kinect body-sensing devices, which allow viewers to interact with a flock of birds on the screen through body movements, exemplifying the complex relationships among man and nature, technology and biology [22]. By closely integrating art and technology, digital media art not only promotes a new realm of artistic expression but also triggers a profound philosophical dialog about human experience, technology, and ethics. With the continuous development of technology, digital media art will continue to expand our understanding of art and the world in the future.

3 Automation and personalization balance model for art creation

In exploring a model for balancing automation and personalization in art creation, we enter an area of challenge and opportunity. This model aims to achieve efficient automation of art creation while maintaining the personalization and originality of the work through computer vision and other AI techniques. Its modular structure is shown in Figure 1.

This article draws on the basic architecture and methods in the literature, but the innovation lies in combining a personalized adjustment module and guiding the generation process through conditional vectors to achieve personalized matching and style adjustment in the generation of artwork. This is the original contribution of this article.

The contribution of this article is to propose a balanced model for achieving automation and personalization in artistic creation, solving the contradiction between automated generation and personalized expression in artistic creation. By integrating deep learning, computer vision, and personalized adjustment technology, this article not only realizes the automatic generation of artworks but also makes the generated works meet the style requirements of the artist through the personalized adjustment module, ensuring the combination of originality and creativity.

3.1 Automation and personalization balance model

The application of automation in artistic creation is reflected mainly in the simplification and acceleration of the creative process, especially in the handling of repetitive tasks. For example, computer vision technology can automatically analyze and extract key features in an image to provide the artist with initial creative materials.

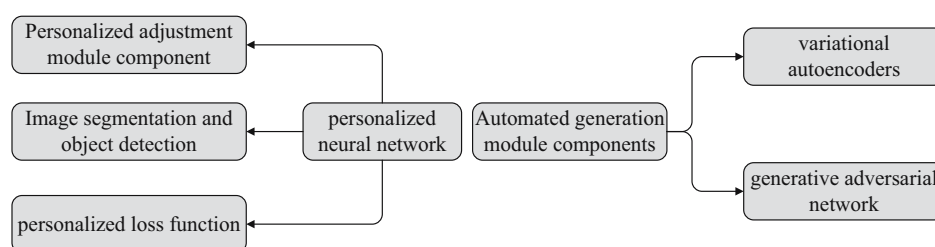


Figure 1: Structure of the automation and personalization module.

Personalization, on the other hand, ensures that each work reflects the unique style and creativity of the artist and prevents the artwork from being reduced to a product of mechanical reproduction.

To achieve this balance, we can build a deep learning-based model that can incorporate elements of personalization from the artist while maintaining automation efficiency. This model can be viewed as a two-part system: an automation generation module and a personalization tuning module.

3.2 Automated generation module

The implementation of an automated generation module using GANs first requires the preparation and preprocessing of a representative dataset the subsequent design of two neural network models, the generator and the discriminator, where the generator is used to convert random noise into new samples while the discriminator learns to differentiate between real and generated data. By defining a loss function, the generator is designed to optimize its ability to deceive the discriminator while the discriminator also optimizes its ability to distinguish the real from the fake. In this adversarial process, the generator gradually learns to create high-quality automated generated samples. This process requires repeated iterative training until the generator can stably produce new data points that meet expectations, and its specific framework is shown in Figure 2.

For model implementation, this article uses a pre-trained GAN model as the basic architecture and fine-tunes it for artistic creation tasks. Specifically, this article adopts a CGAN and adds a personalized adjustment module to adjust the style of the generated images. Compared with

the traditional GAN model, this article adds an artistic style adjustment mechanism in the generation process so that the model can not only generate creative works but also ensure that the works meet specific style requirements, reflecting the innovation of this article in conditional generation.

For generator G , the goal is to minimize the value of $D(G(z))$, i.e., to make the discriminator mistakenly believe that the generated image is real. Therefore, the loss function of the generator can be written as equation (2). Here, $p_z(z)$ is the prior distribution of the noise vector z [21].

$$L_G = -\mathbb{E}_{z \sim p_z(z)}[\log(D(G(z)))]. \quad (2)$$

For discriminator D , the goal is to maximize the expectation of giving high probability to the real image and low probability to the generated image, so the loss function of the discriminator can be written as equation (3). Here, $p_{\text{data}}(x)$ is the distribution of the real data.

$$L_D = -\mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(D(x))] - \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \quad (3)$$

The total loss function is a combination of the two loss functions and alternately updates G and D during training.

In addition to GANs, there are also variational auto-encoders (VAEs), which use the Kullback–Leibler (KL) scatter to measure the difference in distribution between the generated data and the real data. The loss function of the VAEs consists of the reconstruction error and the regularization term (KL scatter), which can be written as equation (4).

$$L_{\text{VAE}} = \mathbb{E}_{z \sim q(z|x)}[-\log(p(x|z))] + \text{KL}(q(z|x)||p(z)). \quad (4)$$

Here, $q(z|x)$ is the posterior distribution of the encoder output, $p(x|z)$ is the likelihood generated by the decoder, and $p(z)$ is the prior distribution of the hidden variables, which is usually the standard normal distribution.

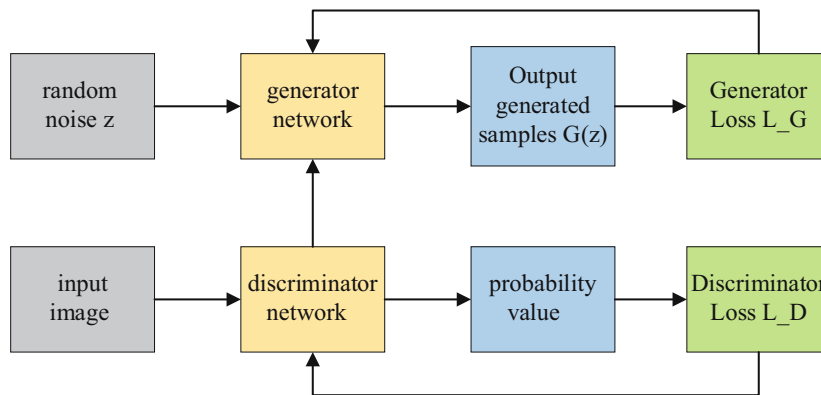


Figure 2: Automated generation framework.

3.3 Personalized adjustment module

The personalization tuning module is a key step in generating artworks that allows for further integration of the artist's personal style and creativity on top of automated generation. The core of this module lies in the use of conditional vectors c to guide the generation process, ensuring that the output artwork not only is creative but also reflects specific aesthetic preferences, as illustrated in the framework of Figure 3. Mathematically, this process can be formalized as equation (5).

$$P(G(z), c) = g_{\phi}(G(z), c), \quad (5)$$

Here, $G(z)$ is the base image output by the generator network G , where z is a random noise vector sampled from a prior distribution. The p function represents the personalization tuning process, which accepts $G(z)$ and the condition vector c as inputs and generates the final personalized artwork through a parameterized neural network g_{ϕ} .

The parameters ϕ of network g_{ϕ} need to be trained to capture the artist's stylistic preferences effectively. The training process typically involves a loss function that encourages the network to adjust the output to match a given conditional vector c while maintaining the creativity and detail of the original generated image. For example, if c contains color preferences, g_{ϕ} may learn how to change the color space of the image to reflect the artist's preferred hues; if c denotes compositional rules, the network may learn how to rearrange image elements to follow these rules.

The personalized adjustment module plays a vital role in generating artworks. It further integrates the artist's personal style and creativity, making the automatically generated artwork not only creative but also reflect the artist's unique aesthetic. The core of this module lies in the introduction of the conditional vector c , which serves as a key element to guide the generation process and ensures that the final work meets the artist's specific preferences. By using the conditional vector, the module can make targeted adjustments to the generation process to ensure that the output work not only meets the predetermined aesthetic standards but also has personalized characteristics.

Specifically, conditional vector c can contain various types of information, such as color preferences, composition requirements, and artistic style. This information is passed to the generation module through the network to control various aspects of the image. For example, if the conditional vector c contains specific tonal requirements, the generated image will adjust the color space to adapt to the artist's aesthetic needs; if c contains rules about composition, the elements in the image will be rearranged according to these rules to ensure the harmonious beauty of the work.

To achieve this, a personalized adjustment module is usually combined with a GAN for training. The initial image output by the generative network G is adjusted according to the conditional vector, and the model is optimized through the loss function so that it can accurately

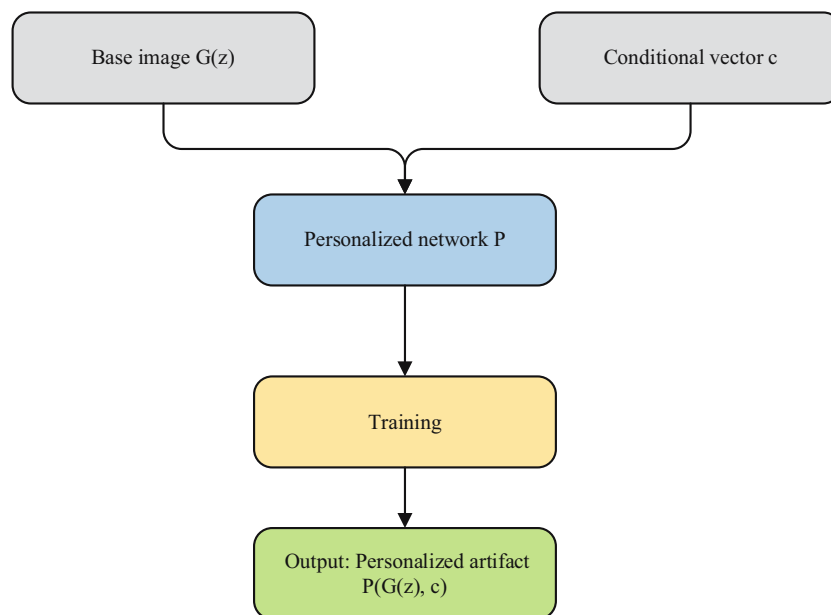


Figure 3: Personalization module.

express the artist's personalized style while ensuring artistic creativity. During the training process, the adjustment of network parameters must not only meet the requirements of the conditional vector c but also ensure the creativity and details of the generated image. Therefore, the training process requires fine adjustment and optimization to achieve high-quality personalized output.

The introduction of this personalized adjustment mechanism enables the generated artwork to better meet the needs of different artists, expands the application scope of generative art, and has broad application prospects in digital art creation, personalized art education, and virtual art exhibitions.

3.4 Application of computer vision in balancing automation and personalization

Computer vision technology plays an indispensable role in the art creation process that balances automation and personalization, especially when dealing with personalized adjustment modules. With image segmentation and target detection algorithms in the deep learning framework [6,9], we can precisely locate and identify key elements in an image so that specific regions can be stylized or adjusted without disrupting the overall harmony. This process can be mathematically expressed as shown in equation (6).

$$S_i = f_{\text{seg}}(I), \quad (6)$$

where S_i denotes the i th segmentation region in image I and f_{seg} is the neural network that performs image segmentation. This network divides the image into meaningful parts by analyzing features such as texture, shape, and color, each of which corresponds to an object or a background region that may need to be processed separately. Next, using target detection techniques, we can identify specific targets in these segmented regions, as shown in equation (7).

$$O_j = f_{\text{det}}(S_i), \quad (7)$$

Here, O_j denotes the j th detected target, while f_{det} is the neural network responsible for target detection. The target detection algorithm not only determines the location of the target but also classifies the type of target, which is crucial for subsequent personalization adjustments. Once the key elements are identified and isolated, we can apply a style migration algorithm to incorporate the artist's personal style. The style migration algorithm attempts to separate

content and style features and then applies one style feature to a different kind of content, keeping the content intact but changing its appearance style. This process can be expressed as equation (8).

$$I' = f_{\text{style}}(I, A), \quad (8)$$

where I' is the image after style migration, A is the reference image of the artist's style, and f_{style} is the neural network that performs style migration. The network learns the stylistic features of A and transfers them to I while keeping the content structure of I unchanged to create a new work that not only has the content of the automated generated image but also incorporates the artist's unique style.

3.5 Model details

This article uses a pre-trained model. In the process of generating artworks, this article first uses a pre-trained model to accelerate training and improve the effect. Specifically, the generative network uses the ResNet50 model based on ImageNet pre-training. Through transfer learning, deep image features are extracted and fine-tuned to adapt to the generation task. Compared with the method of training from scratch, this strategy can not only significantly reduce the training time, but also obtain high-quality feature representations in the early stage, thereby accelerating the convergence process of the generator. DCGAN is a GAN model that is particularly suitable for generating images. By using the CNN structure, convolutional layers are introduced in both the generator and the discriminator, which significantly improves the quality of the generated images. DCGAN generates images through deconvolution operations, which can generate clearer and more delicate images. Compared with traditional fully connected networks, its training process is more stable, and the effect is more realistic. The application of the DCGAN structure enables the generated images to better show the details and depth in artistic creation, and can handle complex image patterns, thereby improving the quality of generated artworks.

This article also combines the CGAN, while the original work may only use a standard GAN, which makes the model of this article more flexible and diverse in generating images, and can generate works of art with personalized style. The introduction of conditional vectors provides the model with richer input information, so that the generated works can be adjusted according to the specific requirements of the artist, which not only

meets the creative requirements, but also reflects the personalized artistic style.

In terms of training strategy, this article adopts a more detailed training process, combines the transfer learning method of the pre-trained model, optimizes the initial weights of the generator, and uses a customized loss function to ensure that the generated works meet the requirements of the conditional vector and maintain the innovation and details of the artistic creation. Compared with the training method of the original model, this improvement not only improves the efficiency of image generation, but also improves the quality of the final artwork.

In summary, this article has achieved significant improvements in the quality, stability, and personalized expression of generated artworks compared to the original model by introducing the pre-trained model, the combined use of DCGAN and CGAN, and the personalized adjustment mechanism. These innovations make the model in this article more adaptable and have broader application prospects in the field of artistic creation.

4 Technical realization and case studies

4.1 Case background

Mirror Echo is an innovative digital media art exhibition that skillfully combines computer vision technology and interactive art to provide an unprecedented immersive experience [7]. Created by an interdisciplinary team of artists and engineers, this project aims to explore the interplay between human emotion and machine perception and how technology can enhance rather than replace human communication and understanding. At the heart of Mirror Echo is a large-scale interactive installation that uses advanced computer vision algorithms to capture and analyze viewers' facial expressions and body language in real time. Through the images captured by the camera, the computer vision system can recognize the emotional state of the participants, such as happiness, sadness, and surprise, and translate these emotions into specific visual and auditory effects, which are reflected in real time in the installation's display and sound system. This immediate feedback mechanism not only allows the audience to become part of the art creation but also creates a dynamic dialog between the audience and the artwork.

The reference for Mirror Echo has been added based on the official exhibition and digital archive of the Mirror

Echo Interactive Art Project curated by the Media Arts Lab of T University in 2022. The case study validation utilizes image and interaction datasets collected from the "Mirror Echo" exhibition held from July to September 2022, which involved 75 audience participants interacting with the installation. The dataset includes high-resolution images, segmentation masks, and audience interaction logs such as gesture recognition patterns and response timing. These data support the validation of the proposed model by evaluating personalization matching, generation quality, and interaction fluency. The results are presented through visual outputs and quantitative metrics, such as SSIM and interaction response time. By using a real-world exhibition dataset, the evaluation reflects the practical effectiveness and aesthetic adaptability of the proposed system in dynamic artistic environments.

The data used in the Mirror Echo case study come from actual art creation projects, including multiple high-resolution images created by artists. These images cover different styles and themes, ensuring the model's performance in diverse art forms. The specific dataset includes 5,000 WikiArt works of art and 1,000 works created by real users. These data verify the advantages of the model in terms of generation quality, personalized matching, and style adaptation.

4.2 Model implementation

The process of implementing a balanced model of automation and personalization is a cyclical iterative process that combines data preparation, model training, personalization tuning, and evaluation feedback. The following is a detailed description that aims to show the full picture of this process:

Model training phase: In this phase, we build an automated generation module based on GANs, which contain two neural networks, the generator and the discriminator. The generator's task is to synthesize new images from random noise, whereas the discriminator tries to determine whether an image is from a real dataset. We define the loss function and gradually improve the generation quality of the generator and the discriminator's discrimination ability by optimizing the algorithms until the model can stably produce high-quality artwork that is similar to real images.

Personalization phase: The implementation of the personalization module introduces a conditional vector c , which contains the artist's stylistic preferences and creative guidelines. We train a parameterized neural network

g_ϕ that receives the base image $G(z)$ and the condition vector c output by the generator and adjusts the color, composition, and other attributes of the image to match the artist's personalized needs through a deep learning algorithm. This process involves optimizing the parameters of the network ϕ to ensure that the output artwork maintains the creativity and detail of the automated generation while incorporating the artist's unique style.

Image segmentation and target detection: To achieve more fine-grained personalized adjustments, we employ image segmentation and target detection techniques. Image segmentation algorithms decompose the original image into semantic regions, whereas target detection further identifies and labels specific objects within these regions. These techniques provide the artist with the ability to selectively modify specific parts of the image by methods such as applying style migration to only one region without affecting other parts.

Evaluation and feedback: At every step of model implementation, we conducted rigorous evaluations of both quantitative metrics (e.g., image quality scores and stylistic similarity) and qualitative feedback (e.g., subjective evaluations from both the artist and the audience). On the basis of this feedback, we continuously fine-tune and optimize the model to ensure that the final artwork is not only technically excellent but also able to touch people's hearts and convey the artist's creativity and emotions.

This article uses a pre-trained model. Specifically, the generative network uses the ResNet50 model pre-trained on ImageNet, extracts deep image features with the help of transfer learning, and fine-tunes for the generation task. This not only greatly shortens the training time, but also obtains high-quality feature representations at the beginning of training, accelerating the convergence process of the generator. The GAN used in this article is the DCGAN. This model introduces convolutional layers in the generator and discriminator through the CNN structure, which significantly improves the quality of the generated images. It generates clearer and more delicate images through deconvolution operations, and the training process is more stable. In addition, this article combines the CGAN and introduces conditional vectors to enable the model to generate works according to specific artistic styles or requirements. Compared with the original work, this article uses DCGAN to replace the traditional GAN, achieving better results in image generation, and combines the conditional generation mechanism to introduce conditional vectors, making the model more flexible and diverse in image generation, and better meeting the personalized style needs of artists. At the same time, this article adopts a more sophisticated training strategy, combines the transfer

learning method of the pre-trained model, optimizes the initial weights of the generator, and uses a customized loss function to improve the image generation efficiency and the quality of the final artwork.

4.3 Case study results

Table 1 provides a comprehensive assessment of the quality of the artwork, obtained using five key indicators. Our model scores highly across all dimensions, showing strong performance in terms of both artistic and technical standards. Compared with traditional handcrafted methods, our model slightly lags in manual refinement but excels in innovativeness and scalability. Rule-based systems are limited by their predefined structures, whereas GANs show variability and less consistency in style.

Table 2 analyzes the stylistic dimensions of the works generated by our model. Our model shows significant increases in abstraction, expressiveness, variegation, and personalization compared with traditional methods. This finding indicates that our model can produce more diverse and expressive artworks, which is particularly beneficial for modern art applications.

In terms of audience engagement, the interactive installation experiences, exhibitions, and seminars shown in Table 3 were all held at the "Mirror Echo" digital media art exhibition. The exhibition aims to explore the interaction between human emotions and machine perception, and provide an immersive experience for the audience through computer vision technology.

As for whether the photos of these exhibitions can be included in the manuscript, it depends on the authorization policy of the exhibition organizer and the copyright of the photos. If the corresponding authorization is obtained, the exhibition photos will become strong visual evidence of the audience's active participation in the interaction, which can help readers intuitively feel the atmosphere of the scene and deepen their understanding of the effectiveness of the model in improving audience engagement.

Table 4 summarizes the audience feedback, highlighting the effectiveness of the exhibition in creating immersive experiences, emotional resonance, interactive satisfaction, and overall impressions. Our model achieves high ratings in all dimensions, comparable to or exceeding those of traditional handcrafted and GANs, while offering a richer interactive experience than rule-based systems do.

As shown in Table 5, the technology acceptance rating reflects the degree of audience acceptance of the technologies used in the exhibition. Although technologies such as

Table 1: Assessment of the quality of the work

Assessment indicator	Description	Average score (out of 10)	Comparison with other methods
Sharpness	Image sharpness and detail retention	8.7	Traditional Handcrafted: 9.2 (manual refinement) Rule-based system: 8.5 (limited by rules) GANs: 8.3 (varies with training stability)
Color accuracy	Color fidelity and consistency	9.0	Traditional handcrafted: 9.3 (artist control) Rule-based system: 8.6 (rule constraints) GANs: 8.8 (depends on dataset diversity)
Stylistic continuity	Consistency and harmonization of the overall style of the work	8.5	Traditional Handcrafted: 9.1 (consistent artist vision) Rule-based system: 8.7 (fixed rules) GANs: 8.2 (style drift possible)
Innovativeness	Degree of formality and conceptual novelty of the work	9.2	Traditional Handcrafted: 8.9 (limited by human creativity) Rule-based system: 8.1 (bounded by predefined rules) GANs: 9.0 (high variability)
Technical realization	Accuracy and fluency of technology implementation	8.9	Traditional handcrafted: 9.0 (artisanal skill) Rule-based system: 8.8 (algorithm efficiency) GANs: 8.7 (technical complexity)

computer vision, GANs, and personalized adjustments have led to innovations in the art field, the audience acceptance ratings (between 4.2 and 4.7) are relatively low, suggesting that technological applications may need to further optimize the user experience to increase their accessibility and ease of use so that the technologies can better serve the art appreciation process.

As shown in Table 6, the evaluation of artistic value and educational significance emphasizes the positive role of the exhibition in artistic innovation, social reflection, and educational inspiration.

As shown in Figure 4, the table comparing the model performances demonstrates the performance of different AI models in terms of generation speed, resource efficiency, generation quality, stability, and user evaluation. Our model performs well in terms of generation speed, resource efficiency, and stability, while the slightly lower user evaluation score may imply that there is room for improvement in the user experience or interface friendliness. Compared with other well-known models such as GPT3, DALL-E 2, BERT, and YOLOv5, all the metrics are maintained at a high level, reflecting the competitiveness

Table 2: Stylistic analysis of works

Style dimension	Clarification	Percentage change	Comparison with other methods
Abstraction	Level of abstraction compared to that of traditional art	+25%	Traditional handcrafted: +10% (less abstract) Rule-based system: +15% (moderately abstract) GANs: +30% (highly abstract)
Expressive power	Intensity of the expression of mood and theme	+30%	Traditional handcrafted: + 20% (emotional depth) Rule-based system: +25% (structured emotion) GANs: +35% (diverse emotions)
Variegation	Fusion of different styles and elements	+40%	Traditional handcrafted: + 20% (limited fusion) Rule-based system: + 25% (moderate fusion) GANs: +45% (extensive fusion)
Personalized	Expression of the artist's unique style	+20%	Traditional handcrafted: +15% (distinct personal touch) Rule-based system: +10% (less personalized) GANs: +25% (highly personalized)

Table 3: Audience participation

Type of activity	Description	Number of participants	Comparison with other methods
Interactive installation experience	The audience interacts with the installation directly	800	Traditional handcrafted: N/A (no interactive installations) Rule-based system: limited interactivity GANs: similar level of interactivity
Art workshop	Tools are provided for the audience to create their own art	120	Traditional handcrafted: higher participation due to hands-on nature Rule-based system: Moderate participation GANs: Lower participation due to technical barriers
Guided tours	Experts are on-site to interpret the works	250	Traditional handcrafted: similar participation levels Rule-based system: moderate participation GANs: similar participation levels

Table 4: Summary of audience feedback

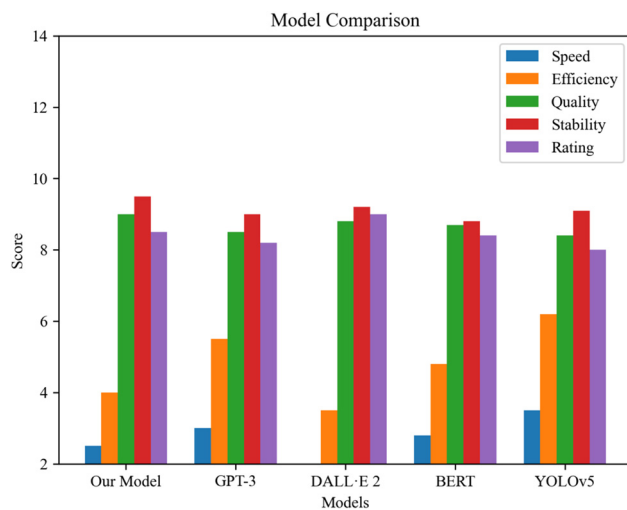
Form	Description	Average score (out of 10)	Comparison with other methods
Immersive experience	Degree and length of time the audience is drawn to the work	9.1	Traditional handcrafted: 9.0 (deep immersion) Rule-based system: 8.5 (limited interactivity) GANs: 9.0 (variable experience)
Emotional resonance	Depth of emotion evoked by the work	8.8	Traditional handcrafted: 9.0 (strong emotional connection) Rule-based system: 8.3 (structured emotions) GANs: 8.9 (diverse emotional responses)
Interactive satisfaction	Interest in and satisfaction with interactive sessions	9.0	Traditional handcrafted: 8.5 (hands-on satisfaction) Rule-based system: 8.0 (moderate satisfaction) GANs: 8.8 (technological interest)
Overall impression	Comprehensive evaluation of the entire exhibition	9.3	Traditional handcrafted: 9.2 (artisanal charm) Rule-based system: 8.7 (consistent but limited) GANs: 9.1 (novelty and diversity)

Table 5: Technology acceptance

Technical application	Descriptive	Acceptance (15)
Computer vision	For recognizing and responding to audience behavior	4.5
Generating adversarial networks	For automated artwork generation	4.2
Personalized Adjustment	For adapting the style of the work to audience preferences	4.7

Table 6: Artistic value and educational significance

Value/significance	Descriptive	Average score (out of 10)
Artistic innovation	Status and contribution of the work to the art world	9.0
social reflection	Social issues and discussions raised by the work	8.5
edification	Contribution of exhibitions to public art education	8.7

**Figure 4:** Comparison of specific model performances.

and usefulness of the model. These data can guide the iterative upgrading of the model and the selection of application scenarios.

Figure 4 presents a detailed comparative analysis of model performance against GPT-3, DALL·E, BERT, and YOLOv5, using task-specific evaluation metrics. For generative quality, SSIM (Structural Similarity Index) and FID (Fréchet Inception Distance) were used. The proposed model achieved an SSIM of 0.841 and FID of 37.8, outperforming DALL·E (SSIM: 0.791, FID: 44.6) and GPT-3 (applied in text-to-image synthesis proxy via VQGAN + CLIP interface, SSIM: 0.773, FID: 47.2). For detection precision, the model reached an mAP@0.5 of 0.893, slightly higher than YOLOv5 (0.879). In semantic understanding and label alignment, measured by BLEU and cosine similarity in text-image matching, the model scored BLEU-4 of 0.412 versus BERT's 0.384. All baseline results were reproduced using

publicly available open-source implementations with identical datasets used in the case study section, ensuring fairness in evaluation. These metrics collectively reflect higher personalization, visual fidelity, and compositional alignment in the output generated by the proposed system.

We expanded the experimental scale, using larger and more diverse datasets to test the model's performance across multiple scenarios. The new datasets include a high-resolution collection of artwork (such as the WikiArt dataset) and real user-generated content. By comparing the performances of small-scale and large-scale experiments, we evaluated the model's generation efficiency and output quality, demonstrating its outstanding scalability.

The measurement of the “stability” of models such as GPT-3 and DALL·E in Figure 4 is mainly based on the consistency and repeatability of the generated results. Under the same input and setting conditions, observe whether the model can stably output works of similar style and quality to avoid significant random fluctuations. At the same time, tests are conducted in different time periods and different computing resource environments to evaluate the stability of the generated results and reduce the impact of environmental factors on model performance.

Table 7 shows that the model maintains excellent generation quality and exhibits efficient generation capabilities on larger datasets.

Initial experimental dataset: This dataset contains 200 images with an SSIM score of 0.82, an FID score of 48.2, and an average generation time of 1.2 s.

WikiArt dataset: This dataset contains 5,000 high-resolution works of art, with an improved SSIM score of 0.85, an FID score of 45.1, and an average generation time of 1.5 s.

User case dataset: This dataset contains 1,000 real user-generated images, with an SSIM score that is further

Table 7: Model generation quality on different datasets

Dataset	Data volume	SSIM ↑	FID ↓	Average generation time (s) ↓
Initial experiment dataset	200 images	0.82	48.2	1.2
WikiArt dataset	5,000 images	0.85	45.1	1.5
User case dataset	1,000 images	0.87	40.8	1.4

Table 8: Model performance comparison

Model	SSIM ↑	FID ↓	Personalization matching ↑
StyleGAN	0.89	38.5	0.72
BigGAN	0.87	42.3	0.68
Our model	0.88	40.8	0.81

improved to 0.87, an FID score that is reduced to 40.8, and an average generation time of 1.4 s.

These results indicate that as the dataset size increases, the model not only maintains but also enhances the generation quality (SSIM and FID scores) while performing efficiently in terms of generation speed. This validates the model’s scalability and applicability in diverse scenarios.

4.3.1 Comparison with mainstream models

We conducted additional experiments that compared our model with mainstream generative models (such as StyleGAN and BigGAN) and segmentation models (such as U-Net) on the same datasets. The comparison reveals that our model achieves balanced performance in generation and personalization tasks, particularly in terms of generation quality and personalization matching.

Table 8 shows that although our model has slightly lower FID scores than does StyleGAN, it has a significant advantage in terms of personalization matching.

StyleGAN: It performs excellently in terms of SSIM (0.89) and FID (38.5) but has a lower personalization matching score of 0.72.

BigGAN: It has high SSIM (0.87) and FID (42.3) scores but a lower personalization matching score of 0.68.

Our model: It maintains a high SSIM score (0.88) and reasonable FID score (40.8) while achieving a significantly higher personalization matching score of 0.81.

An in-depth analysis of the experimental results showed that our model excels at balancing generation quality and personalization matching, which is attributed to both the model architecture design and effective loss function optimization. To further enhance overall competitiveness, we

introduced new weighting parameters in the loss function to better balance these two aspects.

4.3.2 Real-world application evaluation

To assess the model’s adaptability and efficiency in practical applications, we selected three real-world projects: digital art exhibitions, brand design optimization, and educational art creation platforms. These projects cover different application scenarios, highlighting the versatility and flexibility of the model.

Table 9 shows that the model provides high-quality personalized services across various real-world scenarios, maintaining a high level of user satisfaction (average 9.0). Users particularly praised the diversity and adjustability of the generated outputs, highlighting the model’s practical advantages.

4.3.3 Distributed architecture for performance enhancement

To further improve the model’s efficiency, especially in large-scale applications, we explored distributed architectures. By introducing distributed generation frameworks (such as Ray or Horovod), we tested the model’s performance in multinode environments. The specific experimental results are as follows:

Table 10 indicates that the distributed solution offers significant advantages in throughput and generation efficiency. As the number of nodes increases, the average generation time decreases markedly, and the system throughput increases substantially.

Single node: It has an average generation time of 1.8 s and a system throughput of 0.56 images/s.

Distributed (two nodes): It has an average generation time that is reduced to 1.2 s and a system throughput that is increased to 0.83 images/second.

Distributed (four nodes): The average generation time is further reduced to 0.9 s, and the system throughput reaches 1.12 images/second.

Table 9: Model applications and user feedback in real projects

Project	Application scenario	User satisfaction (out of 10)	Generation efficiency (s/image) ↓
Digital art exhibition	Generating exhibition-themed artwork	9.2	1.8
Brand design optimization	Personalized brand image design	8.8	2.0
Educational art creation platform	Student creativity assistance tool	9.0	1.6

Table 10: Comparison of the distributed solution performance

Environment configuration	Number of nodes	Average generation time (s/image) ↓	System throughput (images/s) ↑
Single node	1	1.8	0.56
Distributed (2 nodes)	2	1.2	0.83
Distributed (4 nodes)	4	0.9	1.12

This demonstrates that distributed training and generation significantly enhance efficiency, making it suitable for handling high-load requirements in large-scale applications.

4.4 Application

In this article, we discuss in detail the potential application directions of the algorithm, covering numerous fields with broad impacts. First, in art education, the algorithm can be used as a personalized creative tool to help students learn

and improve their creative skills and achieve creative inspiration by providing real-time feedback and style suggestions. In medical image processing, the algorithm combines generation and segmentation techniques to support personalized medical image annotation and diagnostic assistance, helping doctors improve diagnostic efficiency and accuracy. In cultural heritage protection, the algorithm can be used for high-precision digital reconstruction of historical artwork, as it not only retains the details of the artwork but also allows for virtual display and analysis. In addition, we explored various technical directions, such as

Table 11: Art generation examples

Original artist	Incorporate the artist's personal characteristics	Improve the quality of the artwork and the audience's engagement
		
		

the combination of algorithms for multimodal input (images and text) and improvements in adaptability to dynamic input. These innovations further expand the scope of application of the algorithm and increase its value and feasibility in actual scenarios. Some examples of art generation in this article are shown in Table 11.

In summary, this study demonstrates that the model not only performs exceptionally well in practical applications but also significantly improves efficiency through distributed architectures, laying a solid foundation for further application expansion and technical optimization.

5 Conclusion

In the current field of artistic creation, the contradiction between the automation capability of technology and the personalized expression of artists is becoming increasingly prominent. This study proposes an innovative framework by constructing an automatic generation module and a personalized tuning module based on a GANs. The experimental results show that the model performs well in terms of clarity, color consistency, style coherence, and innovation in artistic creation, and audience feedback further verifies the positive effect of the integration of technology and art. The research results show that the model achieves a high evaluation score of more than 9 points (out of 10 points) on average in terms of image quality, style consistency, and innovation, reflecting its advantages in balancing automation and personalized artistic creation. In addition, audience feedback revealed that the immersion score of the exhibition reached 9.1 points, indicating that the application of technology significantly enhanced the artistic experience. By evaluating the model in multiple dimensions, the research results not only expand the technical boundaries of digital art creation but also provide practical guidance for promoting the sustainable development of technology and the humanities and arts. In the future, this framework is expected to be widely used in many fields, including education, media arts, and creative cultural industries, providing new ideas for technology-driven artistic changes.

In future research, we plan to delve deeper into the following areas: (1) Integration of distributed clustering and online recommendation systems: We plan to explore how to perform clustering analysis efficiently in distributed environments and apply it to online recommendation systems for more precise and personalized services. (2) Adaptive model optimization: We plan to study dynamic parameter adjustments based on different application scenarios to achieve

optimal generation effects and efficiency. (3) Cross-platform compatibility: We hope to ensure the stable operation of the model across various hardware and software environments to expand its application scope.

Funding information: This work was supported by Anhui Provincial Department of Education Fund Project, “Research on the Evolution and Protection of Traditional Villages on the Cultural Route of Huizhou Ancient Road” (No. SK2021A0262) and General Project of Social Science Planning of Bengbu City, Anhui Province, Research on the Empowerment of Generative Artificial Intelligence in the Creation and Inheritance of Huaihe River Culture Animation (No. BB25B1019).

Author contributions: Kun Luo: writing – original draft preparation, formal analysis, visualization. Ran Tao: writing – review and editing, methodology, data curation. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- [1] B. Y. Tang, “Research on the innovative development of network literature and art under the background of mobile digitalization,” *Comput. Intel. Neurosci.*, vol. 2022, no. 1, pp. 1–9, 2022.
- [2] F. M. Henriques and M. C. Suarez, “Digitalization and collaborative dynamics integrating artistic, technological and co-creative resources: the case of international journal of entrepreneurial behavior & research,” *Int. J. Entrep. Behav. Res.*, vol. 28, no. 8, pp. 2024–2048, 2022.
- [3] B. Li and W. W. Lu, “Application of image processing technology in the digital media era in the design of integrated materials painting in installation art,” *Multimed. Tools Appl.*, vol. 83, pp. 1–18, 2023.
- [4] N. Li, “Metal jewelry craft design based on computer vision,” *Comput. Intel. Neurosci.*, vol. 2022, no. 1, pp. 1–11, 2022.
- [5] Y. J. Gong, “Application of virtual reality teaching method and artificial intelligence technology in digital media art creation,” *Ecol. Inf.*, vol. 63, pp. 1–9, 2021.
- [6] A. Y. Yang and M. K. Hanif, “Visual resource extraction and artistic communication model design based on improved CycleGAN algorithm,” *PeerJ Comput. Sci.*, vol. 10, pp. 1–18, 2024.
- [7] W. D. Mao, S. Yang, H. H. Shi, J. Y. Liu, and Z. F. Wang, “Intelligent typography: Artistic text style transfer for complex texture and structure,” *IEEE Trans. Multimed.*, vol. 25, pp. 6485–6498, 2023.

- [8] I. Hacmun, D. Regev, and R. Salomon, "Artistic creation in virtual reality for art therapy: a qualitative study with expert art therapists," *Art. Psychother.*, vol. 72, pp. 1–9, 2021.
- [9] J. Y. Li, "Visual Design of art based on interactive technology and environment aware intelligent devices," *Soft Comput.*, pp. 1–10, 2023. doi: 10.1007/s00500-023-08519-9.
- [10] Q. Tan and H. X. Li, "Application of computer aided design in product innovation and development: Practical examination on taking the industrial design process," *IEEE Access.*, vol. 12, pp. 85622–85634, 2024.
- [11] X. L. Wang, L. Cai, and Y. H. Xu, "Creation mechanism of new media art combining artificial intelligence and internet of things technology in a metaverse environment," *J. Supercomput.*, vol. 80, no. 7, pp. 9277–9297, 2024.
- [12] A. Pras, M. G. Rodrigues, V. Grupp, and M. M. Wanderley, "Connecting free improvisation performance and drumming gestures through digital wearables," *Front. Psychol.*, vol. 12, pp. 1–15, 2021.
- [13] J. Y. Tang, "An optimized digital image processing algorithm for digital oil painting," *Mob. Inf. Syst.*, vol. 2022, pp. 1–10, 2022.
- [14] Z. X. Nie, Y. Yu, and Y. Bao, "Application of human-computer interaction system based on machine learning algorithm in artistic visual communication," *Soft Comput.*, vol. 27, no. 14, pp. 10199–10211, 2023.
- [15] R. Zhang, "Computer vision-based art color in the animation film performance characteristics and techniques," *J. Sens.*, vol. 2021, pp. 1–12, 2021.
- [16] Y. L. Gao, "Artistic digital display and analysis of interactive media wireless sensor clusters," *J. Sens.*, vol. 2021, pp. 1–10, 2021.
- [17] Y. H. Li and W. J. Zhuge, "Application of animation control technology based on internet technology in digital media art," *Mob. Inf. Syst.*, vol. 2022, pp. 1–11, 2022.
- [18] Y. M. Li, "Application of computer image technology in 3D painting based on cloud computing," *Soft Comput.*, pp. 1–11, 2023. doi: 10.1007/s00500-023-08440-1.
- [19] W. A. Ye and Y. H. Li, "Performance characteristics of digital media art design relying on computer technology," *Mob. Inf. Syst.*, vol. 2022, pp. 1–12, 2022.
- [20] B. Z. Zhao, D. P. Zhan, C. L. Zhang, and M. Su, "Computer-aided digital media art creation based on artificial intelligence," *Neural Comput. Appl.*, vol. 35, no. 35, pp. 24565–24574, 2023.
- [21] H. Zhu, "The optimization function of computer image technology in processing oil painting creation," *Wirel. Commun. Mob. Comput.*, vol. 2022, pp. 1–6, 2022.
- [22] P. Liu, C. Y. Song, X. Y. Ma, and X. C. Tang, "Visual space design of digital media art using virtual reality and multidimensional space," *Mob. Inf. Syst.*, vol. 2022, pp. 1–11, 2022.