www.cad-journal.net Computer-Aided Design AND APPLICATIONS

Multi-Modal Art Design Style Conversion Algorithm Based on Deep Learning

Meiling Mao[1] and Ming Li[2]

[1]College of Digital Creativity, Luzhou Vocational and Technical College, Luzhou, Sichuan 646000, China, maomeiling@lzy.edu.cn
[2]Xinglu Sewage Treatment Co., Ltd., Luzhou, Sichuan 646000, China, 13795996646@163.com

Corresponding author: Meiling Mao, maomeiling@lzy.edu.cn

Abstract. This article conducts a needs analysis on the diversified transformation of contemporary multimodal design styles through personalized design styles. In the process of feature extraction and integration in deep learning, this paper performs feature style transformation and extraction by transforming the style model between images and texts. After the key process of artistic style extraction, the original content was digitized and reflected. The research results indicate that the algorithm used in this article has a certain degree of foresight in terms of artistic style data content. In the process of designing works in the field of art and design, this article effectively captures the essence of the target style and the artistic design construction of data style in the multimodal art and design style process.

Keywords: Multimodal; Art Design; Style Conversion; Deep Learning; Neural Network
DOI: https://doi.org/10.14733/cadaps.2025.S3.175-187

1 INTRODUCTION

Multimodal design is the process of conveying specific information to the recipient through visible art forms with a certain purpose as the guide and influencing the recipient. Its basic theory can be applied to advertising design, packaging design, printing design, clothing design, product design, environmental design, and even all design fields related to visual perception. The reason behind this is that multimodal communication in art and design serves as a bridge between the artwork (product) and the audience (consumer) [1]. Art design with a multimodal communication function is a design that uses multimodal symbols to convey information and communicate. It is also a formal rule for exploring and explaining the functions, purposes, and aesthetics of art design. Art and design can effectively showcase designers' creativity in a more vivid and imaginative manner during multimodal communication, making the transmission of information more artistic. Multimodal propagation is a non-verbal form of communication [2]. Nonverbal communication is more suitable for the expression and manifestation of subjective emotions. Its expression methods include body language, pictorial language, object language, etc. Among them, pictorial language is the most widely used in multimodal symbols and plays an important role in public life, culture, and technology, such as painting text, logos, icons, etc [3].

Multimodal communication is essentially the use of multimodal language symbols, rational multimodal language structures, and optimal multimodal programs to convey information. Multimodal propagation can endow design works in two-dimensional space with complex and unique spiritual content and multimodal language [4]. Integrating the forms of multimodal symbols into material forms to create rich and colourful multimodal works. Images are an important way and means of describing the world, as they can capture information that cannot be expressed in language. Prior to photography, people used painting to record the scenes they experienced. Artists combine their inner subjective aesthetic feelings with objective reality through painting [5]. Painting refers to the use of artistic language such as lines, colours, and forms to create static visual images in two-dimensional space. Images are an important way and means of describing the world, as they can capture information that cannot be expressed in language. Prior to photography, people used painting to record the scenes they experienced. Through art style transfer technology, it is possible to simulate the texture, colour, lines, and other stylistic features of artistic works, generate images with artistic styles, and ensure that the final output composite image is consistent with the content of the input image. Many scholars in the field of computer vision have also developed an interest in studying the personal artistic style behind the paintings pursued by artists. Through art style transfer technology, it is possible to simulate the texture, colour, lines, and other stylistic features of artistic works, generate images with artistic styles, and ensure that the final output composite image is consistent with the content of the input image [6]. However, in the digital age, although we can conveniently appreciate these precious paintings, the limitations of traditional media make it difficult for audiences to have a deep interactive experience with the works. This blank not only limits the diversity of art forms but also hinders the audience's deeper understanding and experience of traditional art. Chinese ink painting is renowned for its unique brushwork charm and artistic expression, but its blurred boundaries and simple texture also pose challenges for transforming into realistic images [7]. Traditional research often focuses on transforming modern photos or images into various artistic styles, such as Chinese ink painting, sketching, etc. However, research on the reverse operation of converting Chinese painting into realistic images or modern art design styles is relatively scarce. This process not only enhances the realism of the converted images but also provides the possibility for the integration of Chinese painting and modern art design styles, allowing the converted works to display richer visual layers and details while maintaining the ink-wash charm [8]. Through the application of GAN technology, we can not only achieve this transformation but also integrate modern art and design elements into the transformation process, creating new works that retain the essence of Chinese painting while possessing modern aesthetics. To this end, we innovatively combined the cyclically consistent GAN with the pix2pix model and introduced boundary enhancement functionality, aiming to enhance the clarity and detail representation of image boundaries [9].

This study aims to explore a novel multi-modal art design style conversion method, integrating DL technology to achieve style conversion between images and text. Ultimately, this research aspires to provide innovative technical support and creative inspiration to the art design field, fostering innovation and development within the industry. At present, the research on the transformation of artistic design style has made some progress. In recent years, the ascendancy of DL technology has led to the gradual emergence of DL-based style conversion algorithms. These algorithms accomplish a more natural and realistic style conversion effect by learning from extensive style sample characteristics. Nonetheless, current research encounters challenges like inadequate multi-modal data processing capabilities, inconsistent style conversion effects, and high algorithmic complexity, which hinder the further application of DL in artistic design style conversion.

2 RELATED WORKS

Liu and Yang [10] designed an innovative pencil drawing learning system based on Generative Adversarial Networks (GANs). These datasets not only cover image pairs of natural images and their corresponding pencil drawings but also extensively include pencil drawing works of various art and design styles, such as Impressionism, Realism, Surrealism, etc. The system utilizes the powerful generation capability of GAN to train on a large dataset containing rich artistic design-style pencil drawings. This process not only helps students visually see the pencil drawing effects in different styles but also stimulates their curiosity and desire to explore art and design styles, promoting the cultivation of innovative thinking. This system not only focuses on teaching pencil sketching skills but also deeply integrates exploration and practice of artistic design styles, aiming to provide students with a comprehensive art education platform. In addition, the system greatly facilitates the teaching work of teachers. Teachers can flexibly use the system to assign challenging design tasks and encourage students to apply their learned pencil drawing skills to the creation of artistic design styles. Through this system, students can not only learn how to draw realistic pencil drawings for natural images but also gain a deeper understanding and practice pencil expression techniques for different art and design styles. Students can upload images or scenes they want to try drawing at any time, and the system quickly generates pencil drawing examples of various art and design styles for students' reference. In terms of expanding art and design style analysis, this system not only promotes students' mastery of traditional pencil drawing techniques but more importantly, opens a door for students to enter the world of art and design. Le et al. [11] evaluated the diverse assignments submitted by students to gain a more comprehensive understanding of their learning progress, style preferences, and potential needs. Provide students with more personalized and targeted guidance. Students can experience the collision and fusion of different styles in practice, and try to integrate their personal emotions and creativity into pencil drawings to form unique artistic expressions.

As a researcher in subject teaching, after paying attention to the teaching modes of the entire subject, I have discovered that the common multimodal teaching theories in subject English research can be applied across disciplines to high school art appreciation activities. Enable students to use high school art appreciation classes as a starting point for their love of art. Through systematic learning and practice, their ability to appreciate and discern art will be improved, and they will have a correct and personalized understanding and choice of current art and trends. The multimodal teaching mode focuses on cultivating students' diverse abilities. Various appreciation activities can stimulate a positive and enjoyable classroom atmosphere and stimulate students' learning motivation. By utilizing a combination of multimodal composite resources, multimodal network tools, diverse interactive learning, and various evaluation methods, we aim to enrich the high school art appreciation classroom. Taking students as the main body of activities, enhancing their interpersonal communication skills and class unity and cohesion through self-directed learning, group discussions, and cooperation. The characteristic of multimodal teaching is to allocate and integrate teaching resources and methods based on specific course content and student situations and stimulate students' multiple senses to improve their ability to perceive beauty. Shugrina et al. [12] borrowed the popular multimodal perspective to explore the teaching and practice of high school art appreciation and are committed to innovating the classroom teaching form of ordinary high school art appreciation. Teachers should also guide students in understanding national culture and establishing cultural confidence deeply. Let students personally experience that the sense of national pride and honour spontaneously generated is more solid and stable than the feelings given by teachers. More importantly, artists can directly control the shape, colour, and thickness of brushstrokes, achieving precise adjustment and personalized expression of artistic design styles. By optimizing brush parameters, they can automatically adapt and reproduce the unique style of existing artworks, achieving automatic line painting stylization. In addition, brush interpolation technology enables artists to smoothly transition between different styles, creating unique blended styles. Sun et al. [13] conducted in-depth research on the extensive application potential of continuous latent spaces in the fields of art and design styles. The natural language search function has also been implemented in the continuous painting tool space, allowing artists to quickly locate and apply the art design style they want through simple language descriptions.

Artificial intelligence (AI) has demonstrated significant technological advantages in large-scale design fields such as e-commerce, particularly in automated banner production. However, in the pursuit of creative uniqueness and collaborative artistic design output, its limitations are becoming increasingly prominent. Vo and Sugimoto [14] use carefully selected triplet data: high-quality cartoon images, corresponding accurate semantic label maps, and detailed edge detection maps, to jointly train a GAN model. However, transforming these sketches into finished products that maintain their original intent while also possessing artistic beauty often requires a high level of professional skills and aesthetic judgment, which is currently difficult for many artificial intelligence systems to achieve. Traditionally, human designers tend to conceptualize and express their creativity through sketches. To fill this gap and drive artificial intelligence to explore new heights in art and design styles, the system focuses on the collaborative creation of cartoon landscape paintings, aiming to achieve seamless collaboration between machine and human designers in art and design styles. In practical applications, SmartPaint allows users to input their ideas in sketch form, which the system views as a prototype for semantic label mapping. This process not only enables machines to deeply understand the essence of cartoon style but also to learn to capture and parse the semantic information and spatial relationships of complex objects in landscape images. Zhang et al. [15] further unleash the creativity of users by creating cartoon styles that match their sketches and are expressive, resulting in semantically relevant and aesthetically pleasing finished paintings. The system utilizes its well-trained neural network to automatically convert these sketches into structured and detailed edge maps, which is crucial for stabilizing the uncertainty and differences of various hand-drawn sketches.

3 DL-BASED MULTIMODAL ART DESIGN STYLE CONVERSION ALGORITHM

3.1 Feature Extraction and Fusion

Feature extraction and fusion are pivotal steps in transforming multi-modal art design styles, with complexity and challenges arising from the diversity and heterogeneity of multimodal data. To effectively process this data, this article selects a DL model for feature extraction, specifically utilizing CNN and RNN models to extract features from various data types in the feature extraction stage. CNN is widely used in image feature extraction because of its powerful image-processing ability. Through the deepening of the convolution layer, pooling layer, and other structures, CNN can effectively capture the low-level features such as edges, textures, and shapes, as well as more abstract high-level features in the image, which provides strong support for the subsequent style conversion. The CNN formula is as follows:

$$CNN(x) = \mathrm{Re}LU(Conv2D(x,\theta) + b) \tag{1}$$

$$\mathrm{Features} = \mathrm{MaxPooling}\left[CNN(x_1), CNN(x_2), \ldots, CNN(x_n)\right] \tag{2}$$

Where x is the input data, θ is the parameter of convolution kernel, b is the offset term, $\mathrm{Re}LU$ is the activation function, $Conv2D$ is the convolution operation, and $\mathrm{MaxPooling}$ is the pooling operation. CNN extracts image features through multi-layer convolution and pooling.

For sequence data such as text or audio, this article chooses RNN for feature extraction. RNN, through its unique cyclic structure, can deal with the temporal dependence in sequence data well and capture the semantic information in text or the rhythm and tone in audio.

The RNN formula is as follows:

$$h_t = RNN(h_{t-1}, x_t) \tag{3}$$

Here, h_t denotes the hidden state at time step t, h_{t-1} represents the hidden state at time step $t-1$, x_t signifies the input data at time step t, and RNN is the recurrent function of the recurrent neural network. By employing its circular structure, RNN processes sequential data and captures temporal dependencies.

In terms of feature fusion, this article proposes a novel fusion mechanism aimed at effectively integrating features extracted from different modal data. The core idea of this fusion mechanism is to utilize the correlation and complementarity between different modal data fully and, through certain fusion strategies, organically fuse the feature representations of each modality to obtain a more comprehensive and rich feature representation. This fusion not only preserves the uniqueness of various modal data but also achieves cross-modal information sharing and complementarity, providing a more accurate and rich feature foundation for style conversion. Through this new fusion mechanism, this article has achieved significant improvements in the transformation of multimodal art and design styles.

3.2 Style Conversion Model

In this article, GAN is used as the basic framework of the style conversion model. GAN consists of two parts, a generator and a discriminator, and generates realistic data samples through game learning. This process function is expressed as follows:

$$\underset{G}{Min}\,\underset{D}{Max}\,(D,G) = E_{x\sim p_{data}(x)}\left[\log D(x)\right] + E_{z\sim p_g(z)}\left[\log\left(1 - D(G(z))\right)\right] \tag{4}$$

Among the components:

The distribution of the real sample x is denoted as $p_{data}(x)$.

The distribution of input noise z is represented by $p_g(z)$, typically adhering to a Gaussian distribution.

G And D symbolize the generator and discriminator, respectively.

Maximizing D necessitates ensuring $D(x)$ is maximized and $D(D(z))$ is minimized.

In the task of style conversion, the generator converts input multimodal data into the target style's output data while the discriminator assesses the output's consistency with the target style. Through continuous optimization of the generator and discriminator's parameters, the model learns the mapping between styles and achieves style conversion. Key aspects of the style conversion model include network structure and loss function design. This article employs a CNN-based generator structure, realizing input-to-output mapping through stacked convolution and deconvolution layers, as illustrated in Figure 1.

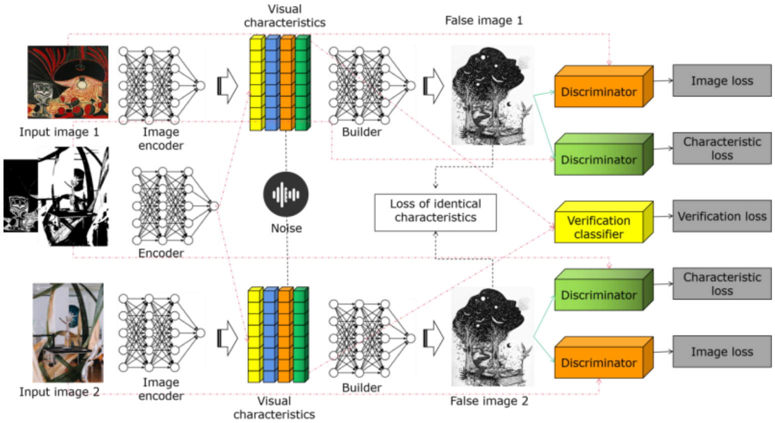

Figure 1: Network structure diagram.

Furthermore, this article also designs a new loss function, including style loss, content loss and confrontation loss, which is used to measure the similarity between output data and target style, content consistency between output data and input data, and the fidelity of output data.

Let $G(x)$ be the generator, $D(x)$ be the discriminator, $C(x)$ be the content loss function, $S(x)$ be the style loss function, and $A(x)$ be the anti-loss function.

Style loss: measure the similarity between the output data and the target style. The style loss function is defined as:

$$S(x) = \frac{1}{N}\sum_{i=1}^{N}\left\|G(x) - S_{target}\right\|_1 \tag{5}$$

Where N is the number of feature maps in the output data, S_{target} is the target style and $\|\cdot\|_1$ represents L_1 norm.

Content loss: measure the content consistency between output data and input data. Norms can be used to define the content loss function:

$$C(x) = \left\|G(x) - x\right\|_1 \tag{6}$$

Where x is the input data?

Countering loss: measure the fidelity of output data. Binary cross-entropy loss can be used to define the anti-loss function:

$$A(x) = \frac{1}{M}\sum_{i=1}^{M}\left\|\log(D(G(x)))\right\| \tag{7}$$

Where M is the number of samples in the output data?

Finally, these three loss functions can be combined into a comprehensive loss function:

$$L(x) = \lambda_s S(x) + \lambda_c C(x) + \lambda_a A(x) \tag{8}$$

Among them, $\lambda_s, \lambda_c, \lambda_a$ is a superparameter, which is used to balance the importance of different loss items. By optimizing these loss functions, the model can achieve better results in the task of style transformation.

3.3 Algorithm Optimization and Innovation

In order to improve the effect and quality of the multi-modal art design style conversion algorithm, this article also carried out a series of optimizations and innovations. Firstly, in the aspect of feature extraction, this article adopts a more advanced DL model and technology-residual network, which improves the effect and accuracy of feature extraction. Secondly, in the aspect of style conversion model, this article improves and optimizes GAN, introduces a new network structure, and adjusts the weight of loss function, which makes the model more stable and efficient in style conversion tasks.

4 EXPERIMENTAL DESIGN AND IMPLEMENTATION

4.1 Data Set Construction

Although there is a good generation effect on the style transfer dataset, the generated images can still be further optimized in terms of evaluation metrics. The loss function of the model is the same as that of the multimodal model, including generative adversarial loss, cyclic consistency loss, identity mapping loss, multi-scale structural similarity loss, and related weight values. Therefore, this chapter proposes a multimodal art style image transfer model. The optimization mainly includes the following three aspects: firstly, inserting the position normalization and moment shortcut (PONO-MS) module into the encoder-decoder. While effectively preserving feature information extracted from input images, it can also improve network optimization and convergence characteristics. The main purpose is to enable the discriminator to better assist the model in distinguishing the differences between input images and generated images, and to improve the style expression and sensitivity of network model training. The results show that multimodal models not only generate images with higher visual quality than multimodal models, but also improve content consistency, style consistency, and learning semantic relationships between natural and artistic style images compared to other models. Introducing a channel-based attention mechanism into the discriminator to focus on more important content during the transfer process.

4.2 Experimental Environment

The hardware environment used in the experiment includes a server equipped with a high-performance GPU and enough memory and storage space to support large-scale data processing and model training. In the software environment, Python programming language is used, and the algorithm is realized with the help of the TensorFlowDL framework. In addition, NumPy, Pandas, and other libraries were used in the experiment for data processing and analysis.

4.3 Experimental Procedure

The implementation process of the experiment mainly includes the following steps:

Data preparation: Prepare experimental data according to the above data set construction and pretreatment process.

Model construction: according to the algorithm framework proposed in the third section, a multi-modal art design style conversion model is constructed.

Model training: use the prepared data to train the model and improve the performance of the model by adjusting parameters and optimizing algorithms.

4.4 Experimental Results and Analysis

Figures 2 to 5 show the conversion accuracy of the multi-modal art design style conversion algorithm proposed in this article in various artistic styles.

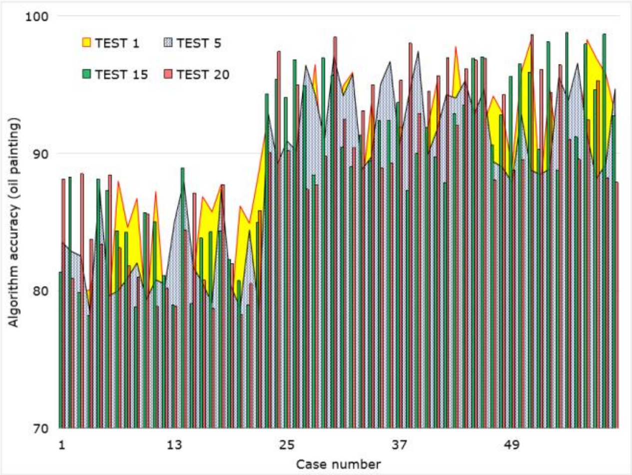

Figure 2: Accuracy of art design style conversion algorithm (oil painting).

As shown in Figure 2, this algorithm demonstrates remarkable accuracy and artistry in converting ordinary input data into oil painting style. Specifically, the algorithm successfully integrates typical features of oil paintings, such as the differences in brushstrokes between thick and thin strokes, subtle transitions and contrasts between colours, and the three-dimensional sense and atmosphere created by light and shadow on the canvas, into the conversion process. Its core advantage lies in its ability to delicately and profoundly capture and reproduce the essence of oil painting - unique brushstrokes, rich and varied colour layers, and intricate texture details. This achievement not only reflects the algorithm's profound expertise in the field of image processing but also demonstrates its outstanding ability to handle complex visual elements and lighting effects. It can not only simulate the unique texture and temperature of oil painting, but also give it new artistic vitality while maintaining the recognition of the original content, so that the final work retains the emotions and storytelling of the original image, and adds a strong oil painting artistic style. For the reproduction of light and shadow effects, algorithms achieve the natural flow and variation of light in the picture through precise calculations and simulations. In addition, the algorithm demonstrates flexibility in handling complex colour palettes, allowing it to adapt to different styles and periods of oil painting. Whether it is the delicacy and harmony of classicism or the boldness and freedom of impressionism, they can be expressed appropriately. Enhanced the three-dimensional and spatial sense of the work, allowing viewers to feel the dance of light and shadow on the canvas.

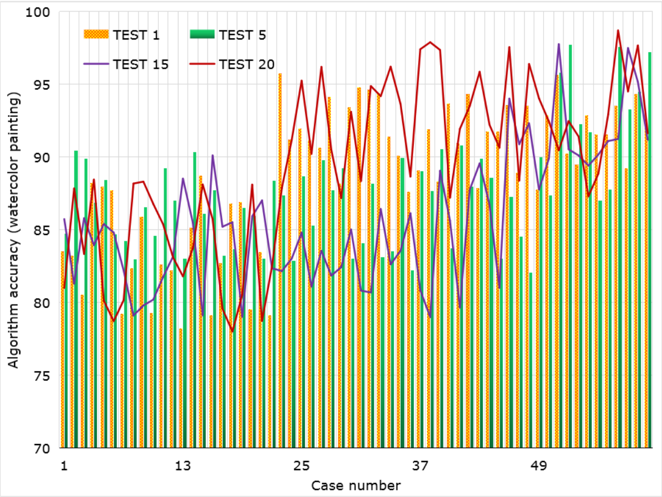

Figure 3: Accuracy of artistic design style conversion algorithm (watercolour painting).

Figure 3 shows the accuracy of the artistic design style conversion algorithm (watercolour painting). The performance of this algorithm in the field of watercolour style conversion is also remarkable, as it accurately captures and simulates the unique artistic charm of watercolour painting. In addition, the algorithm also demonstrates excellent colour gradient processing capabilities. The colour transitions in the watercolour painting are often soft and natural, and this gradient effect is an important component of the unique beauty of the watercolour painting. Specifically, algorithms can not only delicately reproduce the wetting effect in watercolour paintings, but also vividly present the subtle texture of pigment and water blending and spreading on paper in the transformed works. This precise simulation of the wetting effect is a major challenge in watercolour style conversion, but this algorithm can handle it with ease. This algorithm achieves accurate simulation of watercolour transparency through complex transparency calculation and fusion techniques. This algorithm successfully simulates the gradient process of colours from dark to light and from dark to light in watercolour painting through fine colour analysis and calculation, making the converted works more rich and vivid in colour expression. More noteworthy is the outstanding performance of the algorithm in transparency processing. The colours in watercolour painting often have a certain degree of transparency, and when different colours are stacked together, they create a unique visual effect. The converted works have achieved extremely high levels of colour hierarchy and transparency performance.

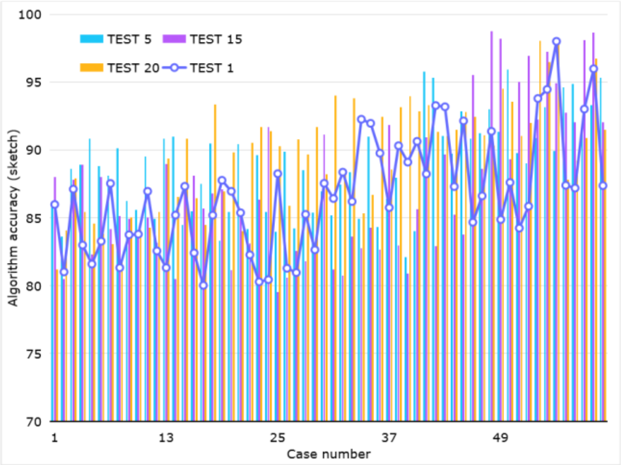

Figure 4: Accuracy of art design style conversion algorithm (sketch).

There is insufficient stylization around the stylized effect, resulting in texture noise. Therefore, proposing a multimodal model can produce higher-quality images. Strengthen constraints on image brightness, colour contrast, and structure while preserving the original image details.

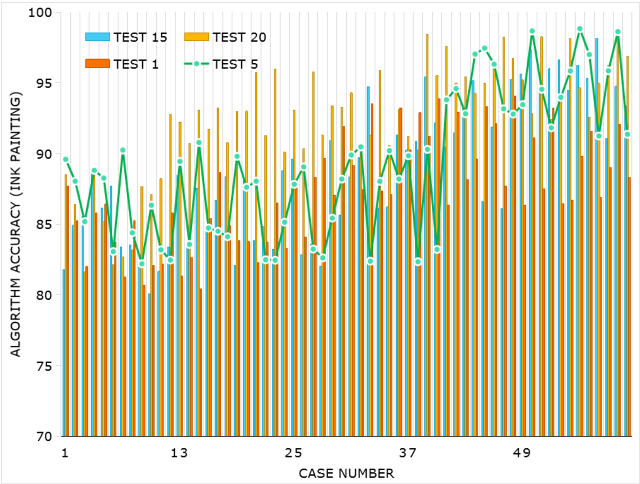

Figure 5: Accuracy of art design style conversion algorithm (ink painting).

From the perspective of artistic style, the ink wash style is unique in the world. However, in the era of traditional film shooting, the ink wash style was subject to many technical limitations and could not freely use the language of the lens. Most could only use simple camera movements such as pushing, pulling, and translating, and the style of characters could not be portrayed delicately and deeply to a certain extent. As analyzed earlier, the commonly used lenses in traditional ink painting styles include fixed lenses, sliding lenses, and motion lenses that move up, down, left, and right. Due to the limitations of two-dimensional plane space, the lens movement is relatively gentle and lacks variation, making it difficult to depict more complex scenes and spaces.

Fortunately, these issues have been easily resolved in the multimodal 3D ink style, making it effortless to use a rotating lens in the scene. The 3D ink painting style not only allows for the free installation of machine positions in the scene and the expression of spatial movements of characters and objects from multiple angles, creating a sense of depth and hierarchy, but also allows for easy scheduling, movement, and addition of various multimodal elements in the image. Compared to traditional ink painting styles, multimodal 3D ink painting styles have richer language and more diverse narrative and expressive techniques. The arbitrary selection of the angle of expression, combined with the comprehensive use of follow-up, shaking, stretching, lifting, and rotation, allows the audience to freely walk in the visual space freely, bringing a strong sense of space and immersion and a sense of immersion.

5 APPLICATION CASE AND EFFECT DISPLAY

5.1 Effect Display

On the one hand, digital technology provides a new tool for ink animation expression. The application of 3D animation technology, motion capture systems, dynamic simulation, digital special effects, and digital synthesis technology has expanded the expressive space of ink animation, enriched its forms of expression, and enhanced its artistic appeal. The ink wash style is very suitable for making modern animations, and 3D ink wash animation is an innovation that combines technology and art. The purpose is to find a point of integration between Chinese ink painting and computer technology and to apply the style of ink painting to modern animation creation. Digital printing technology, graphic and image processing technology, non-linear editing systems, and digital process technology have saved time, reduced production costs and difficulties, and improved production efficiency for ink animation production. This is the fusion of digital and ink painting, as well as the integration of modern technology and traditional Chinese culture. As shown in Figure 6 and Figure 7 respectively.

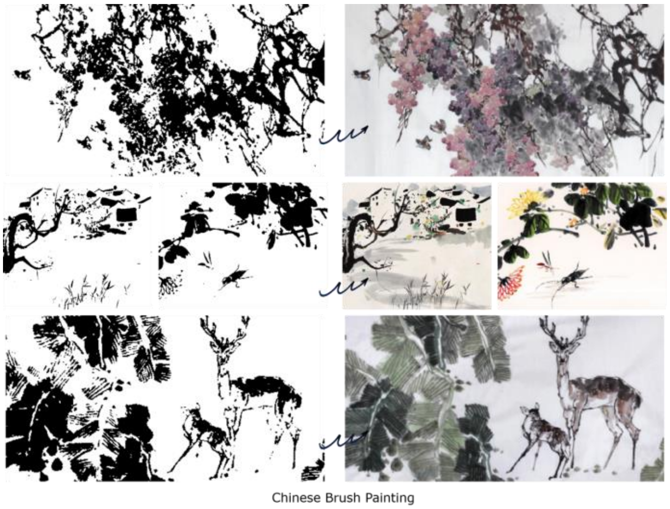

Figure 6: Application Case 1.

As depicted in Figure 6 and Figure 7, the algorithm precisely captures and replicates the colour gradient and brushwork texture of the original painting. This meticulous attention to detail and accurate representation result in transformed works that are highly consistent with the original paintings in style, nearly indistinguishable from the real ones.

In conclusion, the algorithm presented in this article has demonstrated outstanding capabilities in maintaining the original image content and capturing the stylistic characteristics of the original paintings. This not only validates the effectiveness and practicality of the algorithm but also underscores its unique advantages and extensive application potential in the field of style conversion.

Figure 7: Application case 2.

5.2 User Feedback and Assessment

To evaluate the practical application effectiveness of the algorithm, this section gathers and analyzes user feedback and assessments of the algorithm's performance, as illustrated in Figure 8 and Table 1.

User ID	*Assessment of Style Transfer Naturalness*	*Assessment of Style Transfer Fidelity*	*Satisfaction with Generated Works*	*Specific Suggestions and Comments*
1	Excellent	Excellent	High	The algorithm handles details very well, and the generated works are almost indistinguishable from the real ones.
2	Good	Excellent	High	It is hoped that the algorithm can further speed up the processing, especially when dealing with high-resolution images.
3	Excellent	Good	Medium	The overall style transfer effect is good, but some styles appear slightly unnatural during the transfer. Optimization is suggested.
4	Outstanding	Outstanding	Extremely High	Very satisfied. The algorithm generates works that are not only unique in style but also blend naturally with the original image.
5	Good	Good	Medium	It is suggested to add more style options to meet the personalized needs of different users.
6	Excellent	Excellent	High	The algorithm maintains a high level even when dealing with complex scenes, which is impressive.
7	Average	Good	Medium	Under certain specific styles, the algorithm-generated works differ significantly from the original image. Improvement is hoped.
8	Outstanding	Excellent	High	The algorithm processes colours very precisely, and the generated works are rich and layered in colour.

9	Excellent	Good	High	It is hoped that more types of file inputs, such as videos, will be supported in the future.
10	Good	Excellent	Medium	The algorithm performs well in the fidelity of style transfer, but the user interface can be more user-friendly.

Table 1: User feedback assessment table for algorithm effectiveness.

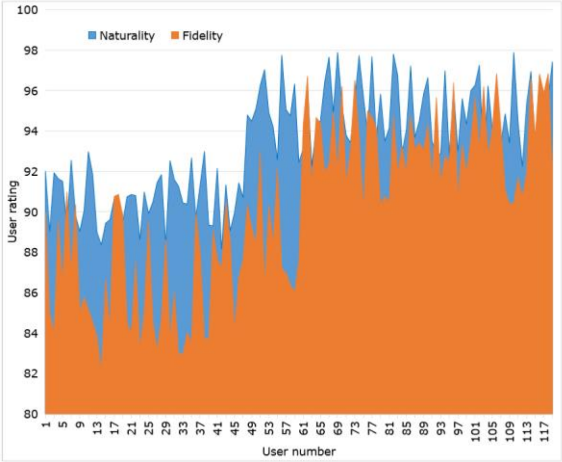

Figure 8: User's assessment of algorithm effect.

A majority of users believe that the algorithm presented in this article excels in the naturalness and fidelity of style transformation, capable of generating artwork that aligns with the target style. Additionally, users have provided valuable suggestions, including enhancing the algorithm's real-time performance and reducing computational complexity. This feedback and assessments offer insightful references and guidance for our future research endeavors.

6 CONCLUSIONS

This article centres on the pivotal issue of multi-modal art design style transformation, conducting extensive research and exploration. Leveraging the relevant theoretical foundation, it introduces a DL-based algorithm for multi-modal art design style conversion. The algorithm facilitates style conversion between multi-modal data through crucial steps, including feature extraction and fusion, as well as the style conversion model. This article comprehensively describes the algorithm's overall framework, the method of feature extraction and fusion, the network structure, and the loss function of the style conversion model, highlighting the algorithm's innovation and practicality. The proposed algorithm demonstrates good results in practical applications, proving its feasibility and effectiveness in real-world scenarios and offering practical value. Additionally, it serves as a valuable reference for future research.

7 ACKNOWLEDGEMENT

Luzhou Institute of Vocational Technology Scientific Research and Innovation Team Construction Project "Sichuan Characteristic Culture Industrialisation Development Research Team" (2021YJTD13).

Meiling Mao, https://orcid.org/0000-0002-4100-6017
Ming Li, https://orcid.org/0009-0003-9169-8562

REFERENCES

[1] Carmen, J.: Makerspaces: combining information literacy with pattern design for fiber art through digital images, Library Trends, 69(3), 2021, 585-611. https://doi.org/10.1353/lib.2021.0005

[2] Chung, C.-Y.; Huang, S.-H.: Interactively transforming Chinese ink paintings into realistic images using a border enhance generative adversarial network, Multimedia Tools and Applications, 82(8), 2023, 11663-11696. https://doi.org/10.1007/s11042-022-13684-4

[3] Gan, J.; Wang, W.; Leng, J.; Gao, X.: HiGAN+: Handwriting Imitation GAN with Disentangled Representations, ACM Transactions on Graphics (TOG), 42(1), 2022, 1-17. https://doi.org/10.1145/3550070

[4] Guo, S.; Wang, B.: Application of computer-aided modeling design in the expression techniques of sculpture art space, Computer-Aided Design and Applications, 19(S3), 2021, 1-12. https://doi.org/10.14733/cadaps.2022.S3.1-12

[5] He, C.; Sun, B.: Application of artificial intelligence technology in computer-aided art teaching, Computer-Aided Design and Applications, 18(S4), 2021, 118-129. https://doi.org/10.14733/cadaps.2021.S4.118-129

[6] Hsu, C.-Y.; Ou, S.-J.: Innovative practice of sustainable landscape architecture education—parametric-aided design and application, Sustainability, 14(8), 2022, 4627. https://doi.org/10.3390/su14084627

[7] Jiang, W.; Luo, X.: Research on unsupervised coloring method of Chinese painting based on an improved generative adversarial network, World Scientific Research Journal, 5(11), 2019, 168-176. https://doi.org/P20190709001-201911-201911060004-201911060004-168-176

[8] Jin, H.; Yang, J.: Using computer-aided design software in teaching environmental art design, Computer-Aided Design and Applications, 19(S1), 2021, 173-183. https://doi.org/10.14733/cadaps.2022.S1.173-183

[9] Jin, Y.; Li, P.; Wang, W.; Zhang, S.; Lin, D.; Yin, C.: GAN-based pencil drawing learning system for art education on large-scale image datasets with learning analytics, Interactive Learning Environments, 31(5), 2023, 2544-2561. https://doi.org/10.1080/10494820.2019.1636827

[10] Liu, F.; Yang, K.: Exploration on the teaching mode of contemporary art computer aided design centered on creativity, Computer-Aided Design and Applications, 19(S1), 2021, 105-116. https://doi.org/10.14733/cadaps.2022.S1.105-116

[11] Le, Q.-T.; Ladret, P.; Nguyen, H.-T.; Caplier, A.: Image aesthetic assessment based on image classification and region segmentation, Journal of Imaging, 7(1), 2020, 3. https://doi.org/10.3390/jimaging7010003

[12] Shugrina, M.; Li, C.-Y.; Fidler, S.: Neural brushstroke engine: learning a latent style space of interactive drawing tools, ACM Transactions on Graphics (TOG), 41(6), 2022, 1-18. https://doi.org/10.1145/3550454.3555472

[13] Sun, L.; Chen, P.; Xiang, W.; Chen, P.; Gao, W.-Y.; Zhang, K.-J.: SmartPaint: a co-creative drawing system based on generative adversarial networks, Frontiers of Information Technology & Electronic Engineering, 20(12), 2019, 1644-1656. https://doi.org/10.1631/FITEE.1900386

[14] Vo, D.-M.; Sugimoto, A.: Paired-D++ GAN for image manipulation with text, Machine Vision and Applications, 33(3), 2022, 45. https://doi.org/10.1007/s00138-022-01298-7

[15] Zhang, G.; Qu, M.; Jin, Y.; Song, Q.: Colorization for anime sketches with cycle-consistent adversarial network, International Journal of Performability Engineering, 15(3), 2019, 910. https://doi.org/10.23940/ijpe.19.03.p20.910918