

ARTICLE

Exploring the use of artificial intelligence painting technology to restore the portraits of ancient poets

Yao Song[†], Zhuolin Wu^{†*} , Lingfang Lan , Wenjing Sun , Yunjian Fan ,
and Zhuan Liu 

Department of Advertising, School of Literature and Journalism, Sichuan University, Chengdu, Sichuan, China

Abstract

Present mainstream artificial intelligence painting models and platforms exhibit a fundamental difficulty in comprehending the defining characteristics of classical Chinese painting: The emphasis on capturing the spirit (*Xie Shen*) or evocative essence over literal depiction, as well as its characteristic flattened esthetic paradigm. This paper reviews the evolutionary trajectory of artificial intelligence painting technology and delineates its fundamental principles. By applying this technology to restore ancient poet portraits into three-dimensional lifelike representations, it provides technical pathways and parameter references for poet portrait restoration. This exploration charts novel application paths for artificial intelligence technology, enabling it to contribute more effectively to Chinese cultural studies. Furthermore, utilizing Stable Diffusion to restore ancient poet portraits bridges the psychological distance between historical poets and contemporary audiences, thus advancing the dissemination of ancient Chinese poetry culture.

Keywords: Artificial intelligence painting; Ancient poet; Chinese culture

[†]These authors contributed equally to this work.

***Corresponding author:**Zhuolin Wu
(2653666832@qq.com)

Citation: Song Y, Wu Z, Lan L, Sun W, Fan Y, Liu Z. Exploring the use of artificial intelligence painting technology to restore the portraits of ancient poets. *Arts & Communication*. 2026;4(1):025170026. doi: 10.36922/AC025170026

Received: April 22, 2025**1st revised:** August 5, 2025**2nd revised:** August 22, 2025**Accepted:** August 25, 2025**Published online:** September 18, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons AttributionNoncommercial License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The development of the internet presents both challenges and opportunities. While pervasive mainstream cultures are eroding traditional cultures, digital technologies and artificial intelligence (AI) offer novel pathways for the innovative development and creative inheritance of traditional culture. As an emerging yet contentious technology, AI painting plays a significant role in advancing the preservation of cultural heritage and its innovation.

The evolution of the internet and AI painting has catalyzed new approaches for cultural innovation and traditional cultural transmission. However, the integration of AI technology and culture still faces multifaceted challenges. This study examines the transmission and dissemination dilemmas, aiming to leverage AI painting technology to restore a comprehensive collection of ancient Chinese poet portraits. Utilizing three-dimensional (3D) poetic imagery, this study bridges the psychological distance between historical poets and contemporary audiences, thereby facilitating cultural inheritance and development.

Within this section, the evolutionary trajectory of the internet and AI painting, as well as the opportunities and challenges brought by AI painting technology in China, will be clearly delineated. All in all, social life is constantly adapting to technological progress, and the inheritance and dissemination of traditional culture and cultural undertakings necessitate adaptation to the internet medium.

1.1. The advent of the internet era

The rapid development of the internet has paved the way for the advancement of AI technology. The widespread adoption of internet technology has fundamentally transformed communication paradigms in the digital age, creating systemic impacts across socioeconomic domains. As documented in the 45th Statistical Report on China's Internet Development (March 2020), China's internet user base reached 904 million, reflecting an annual growth rate of 4.9% since 2018. Notably, mobile internet users accounted for 99.2% of this total (897 million), representing a 0.7% increase in the penetration rate during the same period.

The ubiquitous integration of smart devices into daily life has established digital platforms as the predominant medium for knowledge dissemination. The pervasive advancement of mobile internet ecosystems, coupled with the multifunctional capacities of smartphone technologies and optimized utilization of interstitial time segments, has precipitated an unprecedented level of technological dependency in contemporary society.¹ While propelling the continuous advancement and iterative innovation of internet technologies, these developments have created unprecedented pathways for disseminating China's traditional cultural heritage, simultaneously catalyzing tripartite innovation in cultural transmission through enhanced digital formats, redefined content frameworks, and optimized dissemination methodologies.

1.2. Development of artificial intelligence painting

In recent years, the development of AI painting technology has brought both opportunities and challenges to cultural heritage and the art industry. This section will examine the present application challenges and prospects of AI painting technology by reviewing its development trajectory.

As early as 1943, American neuroscientist Warren McCulloch and logician Walter Pitts proposed the mathematical model of neurons, which became one of the cornerstones of the modern AI discipline. Since the 1956 Dartmouth Conference formally launched AI research, a global proliferation of AI advancements has permeated every aspect of human life.

In literature, the 1962 debut of "Auto-beatnik" poetry software by California-based researchers, Wosy *et al.*,

marked a milestone with the production of machine-generated poems, such as "Rose" and "Child," for magazine publication.² Cinematic history was rewritten in 2001 when Columbia Pictures' *Final Fantasy* became the first fully digital film, with 85% of visuals created through computer-generated imagery rather than traditional filming.

From 2011 to today, the growth of big data, cloud computing, the internet, the internet of things, and similar technologies has helped AI bridge the gap between science and real-world applications. In 2015, Google made its TensorFlow framework publicly available. This system, built on data flow programming, became a key tool for implementing various machine learning algorithms. That same year, intelligent image recognition emerged as a major AI field, paving the way for AI painting technology. It is worth noting that the real pioneer of AI painting was Cohen,³ an art professor at UC San Diego. He created the Aaron software in 1972, which could produce still-life and portrait paintings with unique artistic styles.

The diffusion model (DM) framework, introduced in 2016, has garnered significant academic interest through its innovative stochastic denoising mechanism. This paradigm progressively synthesizes stochastically perturbed images by learning latent feature representations through iterative noise injection, ultimately generating Gaussian-distributed outputs that effectively mitigate mode collapse and gradient instability prevalent in conventional generative adversarial networks.

Subsequent advancements in AI-generated art systems, particularly Midjourney and Stable Diffusion, have catalyzed sociocultural transformations across industries. Engineers use AI painting to generate bridge rendering, outperforming the original rendering production method, which had long cycles and high costs.

In the context of media convergence, AI painting technology plays a crucial role in optimizing multimodal news production workflows and advancing the integration of traditional media into converged media ecosystems. In the Yellow Crane Poetry Virtual Reality Experience System at Huazhong University of Science and Technology, AI painting technology fulfills three core functions: Generating original visual representations of poetic scenarios, reconstructing contextual environments described in classical verses, and artistically rendering poetic atmospheres. This integration creates an omnidirectional immersive experience transitioning from concrete spatial constructs to macro-contextual backgrounds.

However, AI painting has also sparked controversies and debates across various dimensions, including legal, ethical, professional, and artistic. In the internet era, leveraging AI

painting technology to promote cultural inheritance and dissemination appropriately has become crucial.

1.3. The present status of the integration of artificial intelligence technology in Chinese culture

In recent years, numerous innovative endeavors have emerged to integrate China's traditional cultural heritage with modern media platforms. Chinese Poetry Competition employs a studio-based competition format to reinvigorate classical Chinese poetry culture, thereby engaging nationwide audiences in appreciating poetic esthetics. The documentary film *Like Drops of Water Reflecting the Moon* chronicles the life and poetic oeuvre of Ye Jiaying, utilizing the film's narrative framework to exemplify the esthetic essence of Chinese poetry.

Recent advancements in AI and digital technologies, particularly virtual reality (VR) and AI applications, have pioneered novel approaches to innovatively transmit classical Chinese poetry. For instance, the 6th season of the Chinese Poetry Competition pioneered the integration of AI-powered VR with naked-eye 3D studio technology. By establishing real-time 3D visualization zones during recordings, the program delivered immersive experiences to audiences and participants, demonstrating how emerging technologies can inform the preservation of poetic cultural heritage.

Nevertheless, the transmission of classical Chinese poetry culture continues to confront systemic challenges, including outmoded dissemination methodologies and a deficit in immersive experiences of poetic-pictorial synesthesia.⁴ These challenges are primarily manifested in appreciating poetry and experiencing its allure, both of which require savoring the contextual imagery embedded within verses. However, conventional poetry dissemination predominantly focuses on literal interpretations, lacking immersive visual and esthetic engagement. This experiential deficit obscures the full charm of classical poetry, thereby diminishing public interest in its study.

In addition, present explorations in poetic innovation predominantly emphasize textual digitization while neglecting the reconstruction and embodiment of poets' personas. Poets are increasingly reduced to symbolic abstractions within cultural transmission, becoming visible yet unapproachable entities. The predominance of mainstream AI painting software and models is largely held by Western developed nations, with China yet to develop an AI painting software comparable to Stable Diffusion. Both AI painting models and software exhibit strong cultural biases, rendering them ill-equipped to comprehend Chinese cultural contexts and scenarios. Consequently, this poses significant challenges to integrating AI technology for Chinese culture.

As an application of AI technology in the field of painting, AI painting technology can be employed to restore and reconstruct the appearances of ancient poets, transforming portraits of poets into lifelike representations. This approach bridges the psychological distance between poets and contemporary audiences, thereby enhancing interest in poetry studies. This paper provides technical pathways and parameter references for restoring portraits of ancient poets, enabling the present AI painting technology to contribute more effectively to Chinese cultural studies. Simultaneously, this study provides corresponding recommendations for integrating AI painting technology into Chinese culture.

2. Literature review

This study constitutes interdisciplinary and cross-domain research. Consequently, the relevant literature requires reference spans multiple fields and orientations. To enhance the readability of this paper and strengthen the systematicity of the study, the literature review is organized into the following thematic sections: Literature on AI applications, literature on AI painting technology applications, literature on the cultural inheritance of classical poetry, and literature on the symbolization of poet imagery.

2.1. Literature on the application of artificial intelligence

At present, AI applications span multiple domains, including education, healthcare, transportation, arts, communication, culture, finance, and agriculture, encompassing various facets of societal life, with particularly notable applications observed in the arts, cultural industries, education, and cultural heritage preservation.

First, it is essential to explore the application of AI in the art domain. Tao⁵ explores the application of AI in the artistic domain by employing esthetic theory as the foundational methodology. Focusing on AI-generated visual art—particularly painting—as a case study, the research combines the latest advancements in AI design to analyze the creative mechanisms, intentionality, esthetic agency, and evaluation frameworks of AI art. The study concludes by proposing two potential developmental trajectories: Either continuing to emulate human artistic styles in both content and the spiritual essence of form, or exploring innovative approaches to redefine computational esthetic principles through technological evolution.

In addition, the application of AI technology in the cultural industry has been extensively discussed. In exploring the cultural applications of AI, immersive

experience has emerged as a prominent research domain. Hua and Chen⁶ investigate this phenomenon as a novel cultural industry paradigm. Their analysis adopts a dual framework examining both innovative integration mechanisms between cultural production and emerging technologies (AI/VR/augmented reality), as well as the operational characteristics of this sector through four critical dimensions: Developmental patterns, technological foundations, creative content production, and cross-domain innovation mechanisms. Furthermore, from the perspective of global cultural soft power competition, the study draws comparative insights from international best practices to propose four context-specific strategic recommendations for accelerating China's immersive experience industry development.

The role of AI technology in optimizing educational approaches should not be underestimated. Xi⁷ employs a rigorous literature analysis methodology to systematically investigate the application of AI in educational contexts. The study first delineates the conceptual dimensions of AI-driven educational technologies through critical theoretical examination. Subsequently, it conducts a technical taxonomy by dissecting core AI technologies into their constituent elements and architectural frameworks. The research further examines paradigmatic application cases, including but not limited to: Automated grading systems with real-time feedback mechanisms; computer vision algorithms for educational content recognition and analytical processing; natural language processing interfaces enabling intelligent question and answer capabilities; adaptive instructional models supporting personalized teaching-learning interactions; and intelligent learning analytics platforms for behavioral pattern tracking. These case studies are methodically analyzed in terms of their operational scenarios and functional implementations within an authentic educational ecosystem. In addition, Hou and Xiao⁸ explored the feasibility and effectiveness of AI painting technology in optimizing pre-school education.

Finally, present research on AI applications in cultural heritage preservation predominantly focuses on two dimensions: Macro-level discourse regarding AI's functional role in cultural continuity, and its domain-specific implementations in various fields, such as traditional Chinese medicine and the conservation of intangible cultural heritage (ICH).

A representative study by Huang and Yu⁹ elucidates multimodal AI-driven approaches for ICH transmission. This includes leveraging internet and internet of things infrastructures to expand dissemination stakeholders, deploying multifunctional intelligent technologies to

enhance spatiotemporal dissemination capabilities, employing algorithmic systems with deep learning architectures to optimize transmission efficacy, and implementing granular digitization alongside VR interfaces to activate multisensory engagement—collectively constituting a systematic framework for revitalizing heritage practices through technological mediation.

Song and Liu¹⁰ propose a conceptual framework for the digital safeguarding and intergenerational transmission of ICH. The study systematically explores the feasibility of leveraging digital interaction design technologies to facilitate the effective transformation of ICH resources. However, the research exhibits limited engagement with AI applications in the cultural inheritance and innovative reinterpretation of classical Chinese poetry, particularly in terms of semantic analysis of ancient texts, AI-assisted poetic creation, and multimodal dissemination mechanisms.

Tang¹¹ implemented refined model training on painting models through AI technology, enabling AI painting models to learn the regional styles and techniques of Chinese Tianjin painting. The study employed methods, such as a questionnaire survey, to test the effectiveness of the model's application and public acceptance, thereby providing a reference for the similarity investigation in this research.

2.2. Literature on the application of artificial intelligence in painting

Present domestic research on AI painting technology applications primarily focuses on architectural visualization, artistic design (encompassing fashion, cultural innovation, as well as film and television sectors), medical imaging, new media content creation, the entertainment industry, and literary domains, such as classical Chinese poetry hermeneutics and stylistic regeneration.

Yang¹² demonstrated a technical workflow where preliminary bridge visualizations generated through Stable Diffusion undergo iterative refinement through realistic model processing and structural optimization using ControlNet plugins. This methodology enables precise compliance with pre-determined engineering parameters, achieving pixel-level control over generative outputs. The empirical validation confirms that integrating cutting-edge AI systems with civil engineering expertise can establish an efficient bridge rendering production pipeline, thereby unlocking novel possibilities for infrastructure design innovation through human-AI collaborative frameworks.

In addition, Liu¹³ conducts a tripartite analysis through user-centric operational dynamics, rapid iteration cycles of AI models, and generative output efficacy to delineate

the characteristic features and application scenarios of AI painting technology. The study systematically synthesizes the professional ecosystem disruptions and ethical implications arising from this technological paradigm, culminating in forward-looking propositions regarding advancements in algorithmic architectures and model optimization techniques. Yang¹⁴ investigated the technology's transformative potential within the new media sector. Across examining its role in innovating news product formats, restructuring production workflows, catalyzing educational system reforms, enhancing digital political communication, and fortifying next-generation mainstream media institutions, the analysis underscores concomitant ethical dilemmas while proposing multistakeholder governance frameworks to ensure sustainable technological deployment. Another study examines online users in the cultural and creative consumption sector as its research subjects. It explores critical factors influencing users' behavioral intentions toward AI-generated painting content, revealing that the behavioral intentions of consumer-end online users are primarily affected by the response efficacy of individual cognition.¹⁵

In summary, contemporary research on AI painting technology applications predominantly focuses on deconstructing technical mechanisms, evaluating practical advantages and sociocultural implications, and identifying inherent constraints. These studies methodologically investigate the technological feasibility and developmental trajectories within established application domains. However, scholarly attention remains conspicuously absent regarding critical investigation of AI painting's potential in pedagogical contexts, particularly its capacity to innovate cultural transmission paradigms through algorithmic innovation in disseminating classical Chinese poetry heritage.

2.3. Literature on the cultural inheritance of ancient Chinese poetry

Present research on disseminating classical Chinese poetry culture predominantly employs multimodal approaches encompassing cinematic programming, gaming interfaces, VR technologies, computer vision, and digital architecture systems. Implementation strategies frequently prioritize poetic content reconstitution through scenographic environment simulation and artistic conception construction.

This operational paradigm emphasizes maximizing the fidelity of reconstructing textually described settings and esthetic atmospheres to enhance hermeneutic processes, thereby achieving the objectives of cultural transmission.

Nevertheless, extant methodological frameworks and empirical findings provide critical referential value for this research initiative. Cai *et al.*⁴ critically addresses two challenges in cultural transmission: Antiquated dissemination methodologies and deficiencies in immersive conception experience. The proposed solution involves architecting VR ecosystems to develop hyper-realistic environmental simulations that facilitate profound cognitive engagement, thereby transforming abstract poetic imagery into embodied multisensory immersion through real-time haptic feedback and 360° spatial audio rendering.

2.4. Literature on the symbolization of poetic figures

At present, research on the symbolic studies of poet imagery is limited and primarily concentrated on renowned poets. Gong¹⁶ elucidates how Lu You's poetic corpus, imbued with a patriotic ethos, underwent state-sanctioned canonization as a nationalist icon since the post-Southern Dynasties era. This historical process transformed the poet into a symbolic receptacle for successive generations' projection of socio-political aspirations. Such codification compromises the scholarly articulation of Lu You's multidimensional, anthropomorphized literary persona. In summary, scholarly investigation into the symbolization of poetic personas remains limited.

However, as extrapolated from Gong's seminal analysis, the symbolic reduction of poets' identities constitutes a pervasive phenomenon within literary historiography. The persistent emphasis and foregrounding of poets' virtuous attributes through such semiotic constructs engender cognitive estrangement and cultural alienation among general audiences during their engagement with poetic legacies, thereby impeding authentic historical comprehension of literary figures as multidimensional social actors.

3. Poets' portraits screening

Screening of poet portraits enables effective technical workflows and credible outcomes. Numerous scattered individual portraits and portrait collections of poets from China's Song Dynasty have been transmitted to the present day, with significant variations in the depicted appearances of the poets. Therefore, defining clear screening pathways and delineating the scope for these poet portraits is of critical importance.

This research project curated portrait paintings from classical Chinese art compendia, online-to-offline public art resource platforms, and the Palace Museum's digital repository. Primary reference materials include *An Illustrated Compendium of Chinese Historical Figures*

Vol.2 (compiled by the Conservation Department of the National Museum of Chinese History) and *Half-Portraits of Ancient Sages Through the Ages*. The former showcases authenticated portraits of eminent historical figures preserved in the National Museum's collection, except for Lin Zexu's portrait, comprising predominantly Yuan (1271–1368), Ming (1368–1644), and Qing (1644–1912) dynasty original ink-on-silk masterworks, supplemented by pre-modern woodblock prints, stone rubbings, and sculptural reliefs. Three selection criteria were prioritized:

- (i) Frontally oriented portraits with unobstructed facial physiognomy;
- (ii) Verifiable provenance documentation;
- (iii) Portraits painted in the Northern Song Dynasty and later.

4. Research methods

4.1. The screening of artificial intelligence painting software

Different AI painting software tools apply to distinct scenarios and contexts. The selection of AI painting software significantly impacts the efficiency and quality of generated images. Across investigation and practical experimentation, this study has selected Stable Diffusion for the restoration of ancient poet portraits. The rationale for choosing Stable Diffusion is primarily due to its superior operational speed, reliable safety mechanisms, and advanced generative models.

In terms of operational speed, under identical hardware configurations, Stable Diffusion demonstrates relative operational efficiency advantages compared to other AI models, primarily manifested in: Simpler required parameter settings; optimized image rendering effects specifically for landscape photography and portrait close-ups; incorporation of diverse artist style keywords within the model library, with enhanced display quality for recognized celebrity facial templates.

In addition, the model integrates a Not Safe For Work safety filter that automatically converts generated images to solid black when the AI system detects inappropriate content.¹⁷ Li and Xu¹⁸ conduct a thorough analysis of AI applications in garment design, pioneering novel pathways for AI within the garment design domain. In addition, Yao and Wei¹⁹ investigated the technical pathway for generating interior design renderings using Stable Diffusion, integrating the ControlNet plugin to achieve precise control over the rendering effects. Their research process is relatively comprehensive, encompassing a complete workflow of sketch input, information input, intelligent generation, and manual refinement, thereby providing a reference for the technical route design of this study. The

forementioned applications and practices offer valuable insights and enhance the feasibility of this research.

Finally, Stable Diffusion is a model based on the Latent Diffusion Model. The architecture comprises three core components: A Vector Quantized Variational Autoencoder, a DM, and a conditioning mechanism. The initial image undergoes iterative denoising cycles within the latent space of a U-shaped convolutional neural network (U-Net). Users can input prompt texts to intervene and guide the process of generating desired images. This procedure requires the contrastive language-image pre-training tokenizer to parse and arrange lexical relationships from user-input texts, converting them into token embeddings that can be injected into the latent space. This enables the latent space to correctly comprehend phrase relationships, thereby ensuring effective intervention in the denoising and image generation process.²⁰

When these embedded numerical values enter the latent space, the prompt information is continuously consumed during each iterative denoising step to achieve conditional constraint-guided noise reduction. Once the set number of iterative steps is completed, the latent space data are converted back to pixel space information through the variational autoencoder (VAE), ultimately generating images that meet the specified conditions. In summary, images are first compressed into a low-dimensional latent space through VAE, cooperating with the conditioning controller's DM to generate new variables, and then use VAE to convert these variables back into images.

4.2. Selection of function and model

This research primarily utilizes three core functionalities of Stable Diffusion: The image-to-image generation mode, iterative denoising steps, and the ControlNet conditioning plugin.

4.2.1. The image-to-image generation mode

The image-to-image generation mode primarily focuses on two parameters: The guidance scale and denoising strength. The guidance scale determines the correlation between generated image content and textual prompts, where higher values (6–15 in this study) increase prompt influence. Denoising strength reflects the degree of image transformation relative to the source input, configured as a user-defined parameter between 0 and 1. Lower denoising values (<0.5) enable facial refinement with preserved likeness; mid-range values (0.5–0.7) induce substantial alterations in facial features, attire, and overall composition; values exceeding 0.7 result in complete visual reinterpretation. Therefore, the denoising strength range

adopted in this research is maintained between 0.1 and 0.4 to balance fidelity and creative variation.

4.2.2. Iterative denoising steps

Sampling steps refer to the number of iterations Stable Diffusion employs to generate images from input text prompts. Each incremental step provides the AI system with additional computational cycles to cross-validate semantic alignment between textual prompts and evolving latent representations, thereby enabling progressive refinement. This research utilizes step counts typically ranging from 20 to 40 iterations. Suboptimal step counts below this threshold risk incomplete denoising convergence resulting in compromised image integrity. In contrast, higher step counts (>50) enhance textural granularity at exponentially increasing computational costs, with excessive steps (>80) potentially inducing generative adversarial artifacts through over-optimization of latent space trajectories.

4.2.3. The ControlNet conditioning plugin

Control Net is a plugin based on the Stable Diffusion framework that utilizes edge features, depth features, or human pose skeletal features from input images to precisely guide Stable Diffusion in image generation. It introduces new control mechanisms into the image generation process, enabling more stable control over visual compositions and effectively solving the challenge of uncontrollable details in text-to-image generation. The Control Net system incorporates multiple pre-processors; each specialized for efficiently extracting different image features, which are then used to control the image generation process.

The most commonly used Canny pre-processor extracts precise image edges based on set threshold values. The Depth pre-processor converts images into depth maps and can process authentic depth maps from rendering engines. The OpenPose pre-processor generates skeletal pose maps to fix human poses in images, enabling the generation of different characters with identical poses. The Line art pre-processor produces either coarse or refined line drawings from images, and its corresponding model can receive hand-drawn line sketches as input.²¹ ControlNet demonstrates robust functionality by requiring an input image containing specific informational cues as guidance, enabling it to perform operations, including but not limited to composition stabilization, pose definition, and contour delineation.

4.3. Technical roadmap

The complete technical workflow of this study comprises four steps: Initial generation, detail refinement, detail preservation, and final compilation. Initial generation involves adjusting relevant parameters in Stable Diffusion

to restore portraits and generate preliminary results; detail refinement entails utilizing various ControlNet plugins to refine subtle facial features and expressions in the portraits; detail preservation employs corresponding plugins to enhance details and improve image clarity. The subsequent sections elaborate on these four steps.

4.3.1. Initial generation

Utilizing the image-to-image mode of Stable Diffusion to restore the portraits of poets. The image-to-image technology can be effectively applied to the modern restoration of ancient poets, preserving the character features of the original paintings. The selection of realistic human models can alter the stylistic presentation of the images, and the input of keywords ensures that the images do not deviate excessively from the original artwork.

4.3.2. Detail refinement

Utilizing the OpenPose plugin in the ControlNet extension can accurately identify limb movements, facial contours, and other information in images, precisely preserving the facial details of characters to enhance the resemblance of generated images to the poets. The Lineart plugin can extract linear features from the artwork, retaining facial details, natural expressions, and facial features of the original painting. The Depth plugin assists the U-net in understanding spatial relationships within the image by calculating its depth, preventing spatial depth errors in generated images while partially preserving the original contour outlines. By comprehensively utilizing these three plugins, the restoration accuracy of poet portraits can be significantly improved. Photoshop can be used to fine-tune imperfections and illogical elements in generated images.

4.3.3. Detail preservation

The After Detailer plugin can effectively address facial distortion issues. Tiled Diffusion is employed for latent space tiling with single batch processing. The tile count varies depending on the image dimensions and the video random access memory capacity of the computer's graphics card. In this study, the latent space tile height and width are set to half of the image dimensions. By enabling noise inversion with inversion steps set around 25, checking the "Enable Tiled VAE" option to maintain parameter consistency, and adjusting the redraw strength to between 0.6 and 0.75, the detail density of the output can be significantly enriched.

4.3.4. Filtering of generated results

To evaluate the similarity of the restored outputs, we use each poet's portrait to generate five images of the poet with varying similarity levels by configuring different

parameters, then assign numerical identifiers and archive them. Utilizing the Credamo survey software, questionnaires are distributed to participants, requiring them to select the 3D realistic human images exhibiting the highest similarity to the original paintings. Finally, the filtered images are aggregated and compiled into an image database.

5. Research process and results

This study sets the sampling steps to 30 and primarily employs three realistic large models: majicMIXrealistic (v6, Merjic), MoyouArtificial (v1060, Moyou, China), and realistic vision. Different combinations of these realistic large models with varied denoising strength and ControlNet plugins will produce distinct generative effects.

5.1. Denoising strength and model bias

The adjustment of denoising strength: This study sets the denoising strength between 0.2 and 0.4. By modulating denoising strength parameters and utilizing the ControlNet plugin, five distinct images are generated for each poet. The denoising strength is incremented in steps of 0.02 or 0.05 to ensure perceptible variations between iterations. Empirical findings demonstrate that under identical ControlNet configurations using the majicMIXrealistic large model with denoising strength exceeding 0.4, the generated portraits of different poets exhibit similar facial features (Figures 1-3). Comparative analysis reveals that higher denoising strength values correlate with homogenized nasal, ocular, and labial characteristics in reconstructed poet images, predominantly reflecting the model's learned East Asian facial archetypes while obscuring original facial distinctions, resulting in suboptimal restoration outcomes.

Under identical realistic large model conditions (majicMIXrealistic), when the source image exhibits low resolution with indistinct facial features, setting the denoising strength above 0.4 results in generated images displaying Caucasian facial characteristics, as demonstrated in Figures 4 and 5. The source image of Zhang Sanfeng suffers from an ambiguous definition, compounded by low tonal contrast between the figure and background, with minimal chromatic differentiation between facial features and skin tone. These conditions cause the generated image to exhibit elongated and enlarged nasal morphology in the reconstructed portrait, resembling Caucasian facial features.

Furthermore, even when the denoising strength is set within 0.2–0.4, similar phenomena persist. When employing the realistic vision large model on source images with ambiguous facial feature boundaries, setting the denoising strength as low as 0.27 still results in generated



Figure 1. Portrait of Zhao Pu. Image source: 1-and-1 Learning Network. *Chinese Historical Figures Image Library*, URL: <https://www.121xx.cn/cnmr/index> (retrieval date: March 3, 2024). Original image derived from the Qing Dynasty Palace collection. AI processing: Generated by the author using Stable Diffusion (model: majicMIXrealistic_v6; prompt: “an ancient Chinese poet, high-definition image quality, realism,” redraw index 0.5) on March 22, 2024, followed by manual color correction.



Figure 2. Portrait of Zhou Dunyi. Image source: 1-and-1 Learning Network. *Chinese Historical Figures Image Library*, URL: <https://www.121xx.cn/cnmr/index> (retrieval date: March 7, 2024). Original image provenance: Collection of the National Palace Museum, Taipei. AI processing: Generated by the author using Stable Diffusion (model: majicMIXrealistic_v6; prompt: “an ancient Chinese poet, high-definition image quality, realism,” redraw index 0.45) on March 22, 2024, followed by manual color correction.

images exhibiting Caucasian facial characteristics. As shown in Figures 6 and 7, despite the low denoising strength, the reconstructed portraits display light green irises (a trait associated with Caucasian phenotypes) and distinctly feminine facial features. In conclusion, under identical model conditions, higher denoising strength values combined with low-definition source images amplify the model's inherent biases, introducing



Figure 3. Portrait of Zhang Shi. Image source: 1-and-1 Learning Network. *Chinese Historical Figures Image Library*, URL: <https://www.121xx.cn/cnmr/index> (retrieval date: March 6, 2024). Original image provenance: Qing Dynasty Palace Collection. AI processing: Generated by the author using Stable Diffusion (model: majicMIXrealistic_v6; prompt: “an ancient Chinese poet, high-definition image quality, realism,” redraw index 0.43) on March 20, 2024, followed by manual color correction.



Figure 5. Portrait of Zhang Sanfeng. AI processing: Generated by the author using Stable Diffusion (model: realistic vision; prompt: “an ancient Chinese poet, high-definition image quality, realism,” redraw index 0.40) on March 20, 2024, followed by manual color correction.



Figure 4. Original portrait of Zhang Sanfeng. Image source: 1-and-1 Learning Network. *Chinese Historical Figures Image Library*, URL: <https://www.121xx.cn/cnmr/index> (retrieval date: March 5, 2024). Original image provenance: *Portraits and Cultural Relics of the Qiyang Lineage* (1937).

significant challenges to achieving historically accurate portrait reconstructions.

When the denoising strength is set to 0.2, the generated images retain the painting style characteristics of the source material but exhibit low textural dimensionality in clothing and skin rendering, resulting in suboptimal restoration quality. Therefore, configuring the denoising strength within the 0.2–0.4 range demonstrates both methodological rigor and empirical validity for this study.



Figure 6. Original portrait of Yuan Mei. Image source: 1-and-1 Learning Network. *Chinese Historical Figures Image Library*, URL: <https://www.121xx.cn/cnmr/index> (retrieval date: March 5, 2024). Original image provenance: *Portraits of Qing Dynasty Scholars* (1989).

In addition, it is critical to note that in most classical Chinese poet portraits—which prioritize expressionistic over realistic representation—subjects frequently exhibit slightly upturned eye corners. At lower denoising strength values (0.2–0.3), Stable Diffusion fails to reconstruct these subtle epicanthic morphological features. While further reducing denoising strength could theoretically preserve ocular characteristics, such adjustments would concurrently exacerbate loss of facial detail and incomplete stylistic transitions, ultimately compromising visual coherence.

5.2. Utilization of ControlNet plugins

The ControlNet plugins demonstrate operational flexibility, primarily through Depth, Line art, and



Figure 7. Portrait of Yuan Mei. AI processing: Generated by the author using Stable Diffusion (model: realistic vision; prompt: “an ancient Chinese poet, high-definition image quality, realism,” redraw index 0.27) on March 20, 2024, followed by manual color correction.

OpenPose functionalities. The OpenPose plugin enables the precise identification of body posture and facial contours, thereby preserving facial details to enhance the resemblance between generated images and historical poets. The Line art plugin extracts linear features from artworks to retain original facial details, natural expressions, and distinct facial attributes. The Depth plugin calculates spatial depth to support the U-net architecture in accurately interpreting spatial relationships, thereby mitigating depth-related inaccuracies in generated outputs. The application of these three plugins is dynamically adjusted based on empirical reconstruction outcomes, constituting a critical phase of detail refinement. The integrated deployment of all three ControlNet variants enables the systematic modulation of facial feature retention and granular detail preservation.

5.3. Similarity survey and conclusion

This study generated 210 poet images (42 poets in total, 5 images per poet) by modulating denoising strength parameters, ControlNet plugins, and large models. A total of 86 validated questionnaires were collected from participants across diverse genders, age groups, and domestic Chinese cities. Respondents selected the image with the highest similarity to the source image from each set of five generated images. Across statistical analysis of these data, key findings emerged: The majority of optimal images exhibited denoising strength values below 0.3 (concentrated within the 0.23–0.30 range), indicating enhanced facial detail preservation and heightened realism in generated portraits under these parameter conditions. As exemplified in Figures 8 and 9—the latter being a filtered portrait with a denoising strength of 0.25—these



Figure 8. Original portrait of Zhou Dunyi. Image source: 1-and-1 Learning Network. *Chinese Historical Figures Image Library*, URL: <https://www.121xx.cn/cnmr/index> (retrieval date: March 5, 2024). Original image provenance: Collection of the National Palace Museum, Taipei.



Figure 9. Restored Zhou Dunyi Portrait. AI processing: Generated by the author using Stable Diffusion (model: majicMIXrealistic_v6; prompt: “an ancient Chinese poet, high-definition image quality, realism,” redraw index 0.25) on March 21, 2024, followed by manual color correction.

settings yielded the maximum perceptual congruence between input and output images.

5.4. Results evaluation

5.4.1. Clarity assessment

The *Half-Portraits of Ancient Sages Through the Ages* comprehensively covers Song Dynasty poet portraits required for this research. These half-portraits exhibit clearly defined facial details, expressions, and contours, resulting in reconstructed images of Song Dynasty poets with superior facial clarity. Yuan Dynasty poet portraits, though fewer due to the abolition of the imperial examination system (which diminished literati status and elevated popular culture,

leading to the decline of literati culture), are also sourced from the same compendium and retain high-definition characteristics. Portraits of Ming and Qing Dynasty poets primarily derive from *An Illustrated Compendium of Chinese Historical Figures Vol.2*. However, as some entries in this compendium feature full-body portraits with minimal facial details and indistinct features, their reconstructed counterparts exhibit comparatively lower fidelity.

Beyond source image clarity, pixel resolution settings significantly impact generative outcomes: A 1:1 aspect ratio produces squarer facial contours, whereas a 1.5:1 ratio elongates facial features. Nevertheless, even with fixed aspect ratios and increased pixel resolution, low-definition source portraits still yield suboptimal reconstructions. Consequently, the image-to-image workflow demands exceptionally high standards for source portrait clarity and iconographic accuracy to achieve viable restoration results.

5.4.2. Instability assessment

In practical reconstruction workflows, several recurrent failure modes manifest: Stylistic incongruence, facial feature distortion, unintended Caucasian morphological prominence, and model bias amplification.

The first category, stylistic incongruence, primarily stems from suboptimal model selection; employing stable, realistic large models significantly mitigates such errors.

The second failure mode, facial feature distortion, arises from inaccuracies in facial feature interpretation and the use of low-definition source images. AI models fail to decode the nuanced ocular and expressive conventions characteristic of classical Chinese portraiture, resulting in generated images with anomalous ocular expressions (e.g., asymmetrical eye proportions, discordant gaze directions), which can only be rectified by reducing denoising strength.

The third failure mode involves unintended Caucasian morphological prominence. Even when utilizing Asian-biased 3D realistic models, generated images may exhibit distinct Caucasian traits (e.g., azure ocular pigmentation). This phenomenon predominantly occurs when source images lack definitive ocular features and correlates strongly with the Western-centric anthropometric priors embedded during model training, particularly given that most realistic models prioritize esthetically idealized large eyes as a facial archetype. Consequently, the sporadic emergence of Caucasian characteristics remains an irreducible artifact within present reconstruction paradigms.

6. Conclusion

Artificial intelligence-generated painting technology, a novel AI application in the contemporary era, exerts

profound impacts on societal economics, political discourse, and cultural evolution. There is persistent skepticism surrounding AI-generated painting, particularly regarding unresolved legal ambiguities and ethical dilemmas inherent in its implementation. Therefore, establishing frameworks for the ethical deployment of AI-generated painting technology is imperative to harness its potential for advancing sociocultural progress.

In summary, present state-of-the-art AI painting software demonstrates significant difficulty in comprehending the portrayal of figures within the Chinese artistic tradition. Characterized by a flattened perspective and an emphasis on “capturing the spirit” (*Xie Shen*) rather than realism, classical Chinese painting presents substantial challenges in accurately recreating the likeness of historical poets. Predominant AI image generation models, aligning primarily with mainstream esthetic norms, exhibit deficiencies in their grasp of classical anatomical structures and the distinctive features inherent in traditional Chinese painting.

This study proposes specific technical parameters and implementation pathways tailored for the reconstruction of portraits depicting ancient Chinese poets. Furthermore, a digital museum platform was developed to exhibit these AI-reconstructed portraits. The work aims to provide novel methodologies for poet portrait reconstruction, utilizing the potential of 3D representations to bridge the psychological distance between historical figures and contemporary viewers, thereby facilitating the advancement and appreciation of classical Chinese poetry culture. The preservation and transmission of China’s exceptional traditional culture remains a significant journey ahead.

The critical question of how emerging AI technologies can be effectively harnessed to empower the transmission of cultural heritage warrants sustained discussion. Throughout the extensive history of technological development, the advent of new technologies has consistently been accompanied by professional anxiety among practitioners, akin to the transition from the analog to the digital era, where technological innovation directly impacted industry evolution and the psychological states of creators. Unlike other emerging technologies that merely sparked discursive controversy, the resistance against AI painting technology has extended beyond rhetorical opposition, evolving into collective mobilization.²²

Crucially, mainstream AI painting models and software currently lack the capacity to interpret the nuances of Chinese painting styles and poet iconography. Consequently, the development of AI art tools and

foundational models specifically adapted to the tenets and esthetics of Chinese culture is of paramount importance.

Acknowledgments

None.

Funding

None.

Conflict of interest

Yao Song is an Editorial Board Member of this journal, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author contributions

Conceptualization: Yao Song, Zhuolin Wu

Formal analysis: Zhuolin Wu

Investigation: Lingfang Lan, Wenjing Sun

Methodology: Zhuolin Wu, Yunjian Fan, Zhuan Liu

Writing – original draft: Lingfang Lan, Wenjing Sun

Writing – review & editing: Zhuolin Wu

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

No datasets were generated or analyzed during the present study.

References

1. Dong TT. *Research on Digital Art Communication of Chinese Elements in the Internet Era*. Art Theory of Southeast University. [Doctoral Dissertation]; 2020.
doi: 10.27014/d.cnki.gdnau.2019.000287
2. Deng WL. The impact of media evolution and intelligent media on traditional literature-one of the challenges of intelligent media to literature and traditional literary theory. *J Shantou Univ*. 2019;35:1-55.
3. Wang Q. *Reflections on the Phenomenon of Artificial Intelligence Literary Creation*. Hebei Normal University; 2019. Available from: https://kns.cnki.net/kcms2/article/abstract?v=bugi047jrximnrfphbbmxk8ygx2ojfxyhtfb_zf7df5tfv2fdw-raefkpcfsvrlnhv3e26cbg29iwgidqficsdzve
4. Cai XY, Chen QC, Zhang J. Design and realization of VR experience system of Chinese ancient poetry culture with artificial intelligence co-creation. *Publishing Sci*. 2023;31(4):80-88.
doi: 10.13363/j.publishingjournal.20230712.004
5. Tao F. Research on artificial intelligence visual art. *Lit Art Controversy*. 2019;(7):73-81.
6. Hua J, Chen QH. Immersive experience: A new business form for the integration of culture and technology. *J Shanghai Univ Finan Econ*. 2019;21(05):18-32.
doi: 10.16538/j.cnki.jsufe.2019.05.002
7. Xi MRA, Tian SMB, Wu WXC, Yao XD. A review of research on the application of artificial intelligence in education. *Sci Educ Guide*. 2021;(10):70-73.
doi: 10.16400/j.cnki.kjdk.2021.10.024
8. Hou WJ, Xiao YX. Innovative application and deep empowerment of generative AI painting technology in preschool education settings. *Sci Educ Article Collects*. 2025;(12):153-156.
doi: 10.16871/j.cnki.kjwh.2025.12.035
9. Huang YL, Yu H. The use of smart media technology in the communication of intangible cultural heritage. *J Cent China Norm Univ*. 2019;58(6):122-129.
10. Song FH, Liu Y. Digital protection and inheritance strategy of intangible cultural heritage under the perspective of cultural Industry. *Shandong Soc Sci*. 2015;(2):83-87.
doi: 10.14112/j.cnki.37-1053/c.2015.02.015
11. Tang X. *Digital Transformation of Tianjin Regional Chinese Painting: Impact of AI Model Training on Cultural Heritage*. China: Tianjin Vocational and Technical Normal University; 2024.
doi: 10.27711/d.cnki.gtjgc.2024.000218
12. Yang LH. Application exploration of AI painting technology in bridge rendering generation. *Railway Constr Technol*. 2023;(9):24-27.
13. Liu SL. On the impact of artificial intelligence painting on cultural and creative fields. *Contemp Anim*. 2023;(2):91-95.
14. Yang A. Application of artificial intelligence painting technology in the context of media convergence. *News Outpost*. 2023;(14):9-11.
15. Shen J. Influencing factors of consumer behavior towards AI painting in digital transformation of cultural and creative products. *Sci Technol Commun*. 2025;17(4):141-148.
doi: 10.16607/j.cnki.1674-6708.2025.04.022

16. Gong G. English translation of Lu You's poems and the symbolization of Lu You's image. *Chin Comp Lit.* 2013;(4):18-29.
17. Li XLY, Xu WY. Artificial intelligence in garment design: Practical implementation based on stable diffusion. *West Leather.* 2025;47(13):107-110.
doi: 10.20143/j.1671-1602.2025.13.107
18. Zheng K, Wang G. Application of artificial intelligence in image generation - an example of stable diffusion and ERNIE-ViLG. *Sci Technol Perspect.* 2022;(35):50-54.
19. Yao SS, Wei ZL. Precise control of interior design renderings based on stable diffusion. *Tomorrows Style.* 2025;(13):99-101.
20. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models. In: *Computer Vision and Pattern Recognition*. [Preprint]; 2021.
doi: 10.48550/arxiv.2112.10752
21. Wen YX. Analysis of control net plug-in control application in artificial intelligence-generated images. *Film Telev Prod.* 2024;30(2):57-62.
22. Li YJ. *Space, Discourse and Action: Research on Netizens' Resistance Practices to AI Painting Technology*. China: Xi'an International Studies University; 2024.
doi: 10.27815/d.cnki.gxawd.2024.000219