International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 9 Issue 10 October - 2025, Pages: 144-151

Applying Artificial Intelligence in Style Transfer to Recreate Assets in Vietnamese History Games

Khai Nghiem Tuan[1] Duc Nguyen Minh[2]

[1] Wellspring International Bilingual School, Viet Nam Email: kaicoder123@gmail.com
[2] AI Engineer, Viet Nam Email: phucanhthien@gmail.com

Abstract: *The digital entertainment sector continues to grow, along with the demand for digital content that captures a slice of history or ancient culture. The development of computer games themed on history, however, faces a challenge: the lack of culturally sensitive, culturally integrated game assets, especially those that portray the goals of various periods of Vietnamese history. The gap undercuts the potential of culturally sensitive creative design and the risk of cultural misappropriation or distortion of the varied cultural legacy of Vietnam. To bridge this gap, the current work presents a proposed AI-based approach that combines a Generative AI model with style transfer methodologies to modify archival materials from other cultures, creating Vietnamese-inspired variants that reflect Vietnamese cultural aesthetics. The proposed approach uses Stable Diffusion, an image-generative AI, which the author has optimized for use on CPU-only machines and for low RAM GPUs. The use of the Gemini Large Language Model (LLM) is an additional novel aspect of this work, where Gemini engages in 'prompt engineering' or converting simple user intentions into rich, contextually appropriate prompts for Stable Diffusion. This significantly enhances the creative work and reduces the game designers' need for technical knowledge and skills. The system demonstrates the production of culturally relevant and visually coherent Vietnamese-style game assets and significant decreases in design time and effort. This research moves beyond a technical contribution by offering a comprehensive and practical approach to the preservation of Vietnamese digital heritage and the creation of culturally-informed bundled video game content. Additionally, this work presents a reproducible prototype and a procedural reference for cross-disciplinary Artificial Intelligence research, contemporary digital art, and cultural heritage preservation.*

Keywords: Style Transfer; Game Assets; Stable Diffusion; Generative AI; Vietnamese History; Digital Heritage.

1. Introduction

With the rapid growth of the digital entertainment industry, the need for digital products containing strong national cultural elements is becoming increasingly important. Creating Vietnamese historical content video games goes beyond the purpose of entertainment; they also serve the function of capturing, preserving, promoting, and educating the nation's cultural heritage. That said, the most significant and most common challenge for game developers, particularly for small and medium-sized independent development teams, is the unavailability of appropriate visual and graphic game assets that match Vietnam's historical context and design. Developing these game assets from scratch, however, would require enormous human and financial resources, along with sophisticated digital artistry, extensive historical knowledge, and cultural understanding. Hence, the growing need for automated processes that can adapt the available reference materials from different cultures and artistic styles to the Vietnamese context.

Motivating Global Research and Its Applications

The introduction and application of artificial intelligence across different creative industries globally, particularly in the production and conception of digital content, has become a phenomenon worthy of discussion. For instance, in East-Asian countries like Japan, South Korea, and China, generative artificial intelligence has been employed in digital media to modernize traditional styles and reinterpret historical traditional art. Projects like Japan's "AI Ukiyo-e" and China's "GAN-based Cultural Restoration" illustrate the capacity of deep learning models to faithfully reproduce and reinterpret classical art, celebrating non- Western cultural traditions. These instances illustrate the growing perspective of artificial intelligence as a digital tool for the intersection of modern tech and age-old culture. For countries like Vietnam, the technology applied in these projects has the potential to be adapted and applied to Vietnam's traditional art styles.

Neural Style Transfer Background and Prior Studies

Gatys, Ecker, and Bethge (2015) [1] were the first to demonstrate the artistic capabilities of convolutional neural networks (CNNs) by showcasing a network's ability to disentangle and recombine the content and style features of images. The generation of output images capturing the artistic features of a reference image or style laid the groundwork for a cascade of research at the confluence of art and artificial intelligence, particularly within the domains of style transfer and image synthesis. However, empirical research within these domains demonstrates that traditionally NST research has suffered the most severe devaluations in visual NST, the coherence of content, the velocity of processing, and overall NST processing delay.

The most significant research leap in NST has stemmed from the development of diffusion models, especially the latent diffusion model (LDM) presented by Rombach et al. (2022) [2]. Rombach et al.'s LDM framework facilitated the generation of high-quality images from text, the text-to-image process, and image-to-image transformation while providing a high degree of control and stylistic consistency. Yang et al. (2024) [3] and other diffusion model studies extended the gaps in NST research by performing stylistic and morphological imposition more seamlessly than traditional NST methods. This model proliferation has created a high demand for unsophisticated diffusion models across cultural civility style reproduction and appropriation.

Aside from digital artwork, the techniques generated by AI have received wider attention in the research and preservation of cultural heritage. Wang et al. (2024) outlines how AI is pivotal in reconstructing, digitizing, and recreating classical artwork in various forms that have lost their original quality over the years. Global initiatives using deep learning techniques in the restoration of styles of artifacts, ancient paintings, and architectural relics have been taken on for preservation and educational purposes. The technology is deemed revolutionary in the digitization of cultural heritage.

The Situation in Vietnam

Vietnam is beginning to get a lot more attention in this regard. There are new initiatives that show how AI is not only used in the processing of archaeological data, but also in the digitization and recreation of cultural heritage tangible and intangible. There has been research and independent AI artwork explorations in reviving Vietnamese classical styles, such as the Đông Sơn bronze motifs and Lý–Trần dynasty architecture, but it has mostly been lackluster and experimental in approach. This suggests that we need a systematic and reproducible approach that combines cultural heritage and AI.

Research Gap and Objectives

Although there have been tremendous advancements on a global scale, the incorporation of foreign artistic materials, particularly Japanese, European, and Korean designs, into the Vietnamese context through AI-driven style transfer has been studied very little. Much of the existing literature in creative industries, particularly in gaming, focuses on restoration rather than the cultural reinterpretation. Hence, the presented research aims to respond to this gap by developing something automated and scalable to address Vietnamese cultural style transfer.

Accordingly, this study addresses three core research questions:

1. How can a diffusion-based generative AI system recreate game assets that reflect Vietnamese feudal-era aesthetics?
2. How can the system be design for low-resource hardware to aid small-scale game developers?
3. What can be automated in the style transfer from foreign to Vietnamese art to ensure cultural fidelity?

2. Theoretical Foundation and Innovation

This research relies on two technologies: diffusion-based image synthesis and reasoning with large language models (LLMs). In this instance, the Gemini LLM and Stable Diffusion are used together under the premise that Gemini will be the "Prompt Engineering Expert." Depending on the recent literature, the semantic richness and coherence of a discourse significantly improves the fidelity and consistency of the images generated with the diffusion models. Incorporating Gemini into the workflow balances the intent of human operators and the control of the models by automatically converting brief user instructions into elaborate, situational prompts that a highly contextual generator model can use. This linguistic and semantic understanding, together with generative visualization capabilities, improves the technical output and the cultural coherence of the generated images.

There are two integrated AI models that make the GPtS work:

- **Google's Gemini LLM** serves as the intelligent assistant for prompt generation, transforming user requests into optimized text descriptions for Stable Diffusion.
- **Stable Diffusion v1.5** assigned the image synthesis task by reproducing the assets with the specified style, high resolution, and artistic fidelity, even on low-end hardware (CPU and low-memory GPU).

This combination of methods is an additional aid to creators and a further easing of the technical choke points for designers and developers. The unification of textual and visual models is a further methodological contribution to the generation of digital content.

Rationale and Expected Contributions

This research makes contributions to methodological and cultural domains. It documents a technique as a contribution to culture and a loss to a culture's digital heritage for the preservation of culturally important content mediated by AI. In the area of technique, it offers a reproducible model that unites diffusion-based generative modeling with LLM-based prompt reasoning for the automated cultural style transfer, and in the area of culture, it helps in the digital heritage preservation.

From a practical standpoint, the system offers a scalable and affordable solution to small studios and independent developers with limited capacity for the manual generation of period accurate assets. It provides a basis for the further use of AI in the advocacy and practice of the digital humanities, pedagogy, and cultural tourism.

This study is intended to democratize the generation of culturally significant digital assets and further offers a model for the digital transformation of creativity in the Vietnamese economy. The study illustrates the interplay of the technologies and a culture's heritage, affording AI newer, dual opportunities for use as a preservation tool and pedagogy for national identity.

Cultural Heritage AI Applications

The usage of AI to automate image generation and synthesis, as well as to assist in the more focused activities of cultural heritage preservation, digital restoration, and creative media production, is skyrocketing. This part highlights the most important works to build on the theoretical and technological advances that inform this study.

Neural Style Transfer (NST)

Neural Style Transfer (NST) is one of the most important and influential subfields of deep learning image processing and was first conceptualized by Gatys, Ecker, and Bethge (2015) [1]. Their most influential contributions involved the use of a Convolutional Neural Network (CNN) to disentangle content from various styles, thus enabling the re-rendering of images in the artistic styles of certain paintings or artists. This creativity sparked massive technological advancements in the fusion of artistry and computing.

Artistic style synthesis and visual transformation are extensive areas of study as a result of the pioneering work of Gatys, Ecker, and Bethge (2015).

Improving inference speed, visual quality, and flexibility in applying multiple styles became the new focus for different NST variants. For example, Johnson et al. (2016) [6] performed real-time style transfer by training a separate network for each style and introducing a feed-forward architecture. Further, Huang and Belongie (2017) [7] introduced Adaptive Instance Normalization (AdaIN) by aligning feature statistics, allowing for multiple styles to be used all within a single model. Such advancements in contrast NST made it efficient and practical.

Nonetheless, cell-embedding CNN NST models still lack global semantic coherence, and as a result, the outputs are rendered inconsistently and are visually distorted. Such drawbacks lead researchers to use architecture-based techniques to improve long-distance dependencies in the image representation and are used in the new generation of Diffusion Models.

Diffusion Models

Diffusion Models have recently been regarded as central to the field of generative AI for their ability to generate stable, detailed, high-resolution, and visually coherent images. Rombach et al. (2022) [2] introduced a new architecture for the Latent Diffusion Model (LDM) allowing for diffusion to occur in a latent feature space as opposed to a tangible pixel space, thus maintaining image quality while drastically cutting down computational cost.

Stable Diffusion has advanced to become one of the premier models for guided image generation and offers text-to-image and image-to-image generation. Furthermore, there are numerous derivatives works which have continued to build on the model's flexibility and controllability. Hertz et al. (2023) [8] introduced Prompt-to-Prompt Editing, which enables users to adjust the semantic relationship between a text and an image and control it through cross-attention. Zhang et al. (2023) [9] designed ControlNet, an adjunct network that uses geometric or edge instructions to maintain spatial control and a structurally accurate edge while the image is created.

These innovations broaden the spectrum of potential uses of Diffusion Models to additional industries that rely on high realism and accuracy, including virtual fashion design, recreation of historical artifacts, and generation of game assets. Diffusion-based models have a distinct advantage over legacy CNN-based NST by offering versatility in diffusion artistic style manipulation while content preservation is maintained through diffusion.

AI in Cultural Heritage Preservation

Artificial intelligence is no longer only used for the generation of art but is also utilized for the preservation and digitization of cultural heritage. Generative and style-transfer techniques have been employed in the reconstruction of lost artworks and the digital preservation of artworks and artistic styles of ancient civilizations (Wang et al., 2024) [3]. Also, the automation of the classification, restoration, and transformation of heritage images assists deep learning in archaeological and museum research (Kim et al., 2023) [10].

AI is valuable for cultural documentation and educational dissemination, as seen in the INCEPTION projects (Llamas et al., 2016) [10]. Machine learning techniques have been shown to support restoration and also facilitate reinterpretation and reimagination of heritage for novel purposes. The intersection of technology and the preservation of cultural identity create new opportunities in the integration of computer vision, the digital humanities, and art history.

The consideration of broad applications of AI to cultural heritage in Vietnam is still in its infancy. While some research groups have looked into AI-aided restorations of ancient texts or other objects, there have been no systematic inquiries into the use of AI to adapt cultural styles, particularly in digital entertainment mediums, video games or animation. This indicates an opportunity for employing generative models to digitally portray Vietnam's heritage in an accessible and authentic manner.

Research Gap and the Direction of this Study

Despite the many promising developments, there is still much to be explored regarding the creative restoration of works prior to the original. Literature does not adequately address the recontextualization of global art into works of Vietnamese culture, particularly those rooted in Vietnamese traditions.

This work aims to build on the theoretical work of NST and Diffusion Models by including the creative re-contextualization of heritage styles. More specifically, it investigates the automatic transformation of international art styles, specifically European, Japanese, and Korean, into Vietnamese feudal styles. As such, it provides the first practical Generative AI framework for the authors to produce culturally authentic, game-ready assets.

This study integrates Gemini, a large language model (LLM), into the Stable Diffusion workflow, setting it apart from previous studies. While earlier works focused on optimizing the architecture or training data to improve the quality of the images, this study proposes semantic prompt engineering to guarantee cultural fidelity and controllable creativity.

3. System Architecture and Methodology

The system in the study incorporates two key components:

1. **Prompt Generation Module** — built on the Gemini large language model, expands user requests into expansive descriptions to improve the inputs to the image generation module.
2. **Image Generation Module** — uilt on the Stable Diffusion model, performs the task of generating images in accordance with the specified instructions for a requested style produced in the prompt generation module.

The overall system outline, with a diagram depicting two modules, the bidirectional flow of data, and the human-in-the-loop interaction to maintain semantic consistency and visual quality across iterations, is presented in Figure 1.

Figure 1. *Overall system workflow for Vietnamese-style image generation.*

3.1. Prompt Generation Module

The initial research sought ways to ease the creative process and lower the required in-depth knowledge threshold for game designers and programmers. This is why, in the first stage of the process, we aimed to transform simple natural language requests like "Change the Japanese sword style to the Vietnamese style during the Trần Dynasty" into a detailed optimized prompt for the image generation model to streamline the efficiency of the following stages concerning image generation. This task is handled directly by the Gemini 2.5 Flash large language model [5], which interacts through API integration by receiving a prompt.

One of the default system directives for the Gemini model has it working as a prompt designer for the Stable Diffusion model. This has been designed in such a way that it contains the required output as a JSON, including:

- prompt: a detailed description to produce and style the images;
- negative_prompt: things to include that should be excluded (e.g., "blurry, low quality, distorted");
- steps, cfg_scale, strength: the composition of the parameters that control the diffusion;
- sampler or seed: to control the reproducibility of the outcome.

User requests are combined with system instructions and submitted to Gemini's large language model via an API. The large language models analyse the request semantically and configure the technical parameters and scope of the prompts with adequate detail and contextual awareness. The JSON output is normalized and sent to the Image Generation Module directly for the diffusion step. This reduces the abstraction gap users have to bridge to work with Stable Diffusion, encapsulating their creative ideas in technically accurate descriptions, and removes the need for in-depth Stable Diffusion model knowledge.

For example, with the user request: "Transform the Japanese sword into the sword used in feudal Vietnam," the Gemini model is likely to respond with the following JSON structure:

Listing 1. *Example of JSON structure generated by Gemini for Stable Diffusion input.*

```
{
  "prompt": "ancient sword, Vietnamese feudal era design, dark aged bronze and steel
blade, intricate gold dragon-phoenix motifs, lotus ornaments, lacquered wood hilt,
handcrafted, traditional patterns, subdued lighting, ultra realistic, highly detailed,
8K, masterpiece.",
  "negative_prompt": "blurry, lowres, distorted, bad anatomy, extra limbs, out of
frame, deformed, unrealistic materials, western armor style, katana, samurai, Japanese
sword, wakizashi, tanto, ninja, kanji, cherry blossoms, minimalist, sleek modern
design.",
  "parameters": {
    "steps": 42,
    "cfg_scale": 9.5,
    "sampler": "DPM++ 2M Karras",
    "seed": -1
  }
}
```

In the interest of reproducibility, each inference run is documented with a timestamp along with a checksum of the configured prompt. This allows for a quantitative assessment and facilitates peer review, which is a crucial step in the transparency of a paper's methods.

3.2. Image Generation Module

Central to the process of image style transfer is the Img2ImgGenerator layer, which implements Hugging Face’s diffusers StableDiffusionImg2ImgPipeline. For the recommendation system, we use the runwayml/stable-diffusion-v1-5 model, which is a version LDM (Latent Diffusion Model) with capabilities of high-quality image generation from a text description plus an input image. Given that this research focuses on deploy ability on resource-constrained hardware (low memory CPUs or GPUs), the pipeline has the following configuration:

- enable_attention_slicing(): splits the attention computation into smaller segments, which diminishes RAM and VRAM usage;
- Disable safety_checker: skips the safety check to lower memory and time costs.

- Using the standard 512x512 RGB image format, which balances speed and image detail.

To illustrate the capabilities of the proposed system, several test cases were run, which are summarized in Table 1. For each case, we show the user's natural-language request, the corresponding prompt generated automatically by the Gemini LLM, and the resulting image rendered by the Stable Diffusion model.

Table 1. *Experimental results of the style transfer system*

No.	User input	Original Image	Prompt generated by Gemini	Result image
1	Transform a Japanese sword into a sword used during the feudal era of Vietnam?		ancient sword, aged steel blade, ornate bronze guard featuring dragon and phoenix motifs, dark lacquered wooden hilt, silk wrapped handle, intricate lotus patterns, Vietnamese feudal era design, handcrafted, traditional patterns, refined, elegant, subtle lighting, muted earthy tones, antique finish, ultra realistic, highly detailed, 8K, masterpiece.	
2	Convert a Japanese samurai helmet into a helmet worn by generals of the Vietnamese feudal army, keeping the structure identical to the original image?		Vietnamese general's helmet, aged bronze, dark polished metal, intricate dragon-phoenix motifs, stylized clouds, lotus ornaments, ancient Vietnam aesthetic, traditional patterns, dark regal tones, warm metallic sheen, soft ambient lighting, ultra realistic, highly detailed, 8K, masterpiece.	
3	Transform a European or American-style shield into a shield used by the Vietnamese army during the Trần Dynasty?		Vietnamese Trần Dynasty shield, ancient bronze, dark carved wood, red and gold lacquer details, intricate dragon and lotus motifs, handcrafted, traditional patterns, aged texture, dramatic lighting, ultra realistic, highly detailed, 8K, masterpiece.	
4	Convert a European-style shield into a shield used by the Vietnamese feudal army in the Trần Dynasty period?		Ancient shield, Vietnamese feudal era style, Tran Dynasty aesthetic, dark polished wood, aged bronze, intricate dragon and lotus engravings, deep red lacquer accents, rustic texture, historical artifact, ultra realistic, highly detailed, 8K, masterpiece.	

Note. The prompts in this table are auto-generated by the Gemini LLM based on user text input, and the images were generated with Stable Diffusion v1.5 in image-to-image mode with no intervening manual editing.

The following outlines how each module operates:

- Step 1: Preprocess input image: resize input image to 512×512 pixels RGB.
- Step 2: Load image generation parameters: parameters prompt, negative_prompt, steps, cfg_scale coming from the Prompt Generation Module are loaded into the pipeline.
- Step 3: Diffusion process: the model applies a series of noises and denoising techniques within the latent space to form a new image, which, at the end of the process, has the distinctive Vietnamese style.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 9 Issue 10 October - 2025, Pages: 144-151

- Step 4: Post-processing and output: output image is generated and, in the end, is saved in the display format.

This module outputs Vietnamized asset images which holds the composition and structure fidelity of the provided source image while integrating genuine Vietnamese artistry — color compositions, material textures, patterns of the motif (dragon-phoenix, lotus) and decorative elements of stylized Vietnamese art.

The users of the module have the outputs to a scientifically grounded process. This was done through the usage of qualitative (Mean Opinion Score) and quantitative measures (FID, LPIPS) which objectively validated the cultural accuracy and realism that the system generates.

4. Experiment and evaluation

To analyze the effectiveness and reliability of the proposed AI-based system, a series of focused research studies with controlled experiments looked at three features: (1) the semantic and stylistic correspondence of the system-generated images, (2) the visual quality in a quantitative sense, and (3) the range of system performances across various computational configurations.

4.1. Experimental Framework

Operational Computing and Storage Facilities

In order to ascertain the system's performance scalability, the system was implemented and evaluated with variance in the constraints and optimization of hardware resources.

- **Environment 1 (CPU-only):** Intel Core i5-6200U (2.30 GHz), 8 GB RAM, no dedicated GPU.
- **Environment 2 (GPU-enabled):** Google Colab Pro+ with NVIDIA T4 GPU (16 GB VRAM), 25 GB RAM, and Python 3.10 runtime.

The entire pipeline was implemented in **Python** using the following libraries and frameworks:

- *Hugging Face Diffusers 0.29* for Stable Diffusion inference;
- *Transformers 4.40* for the Gemini LLM interface via API;
- *PyTorch 2.2* backend for GPU acceleration;
- *Pandas, NumPy, Matplotlib* for logging and visualization;
- *Pillow* for image post-processing and output formatting.

Dataset and Input Samples

There were 30 evaluation images. The dataset was manually collected and classified into 3 main categories focusing on the history of Vietnamese visual culture:

1. **Ancient Weapons**: swords, shields, spears.
2. **Ancient Armors and Costumes:** helmets, military attire, and ceremonial outfits.
3. **Decorative Objects:** artifacts, emblems, and architectural ornaments.

Each image was paired with a short user input (in natural Vietnamese) describing the transformation goal, such as:

"Chuyển thanh kiếm Nhật Bản thành thanh kiếm sử dụng trong thời kỳ phong kiến của Việt Nam."

The Gemini model expanded these prompts into elaborate structured commands in a JSON style for the Stable Diffusion module.

Evaluation Metrics

Qualitative and quantitative evaluation methods were combined to provide a comprehensive assessment:

- **Mean Opinion Score (MOS):** 5 independent evaluators were asked to rate the images based on semantic relevance, stylistic accuracy, and aesthetics, and their scores were averaged on a scale of 1–5.
- **FID (Fréchet Inception Distance):** estimated realism and variance of the AI-generated images, relative to the target cultural dataset.
- **LPIPS (Learned Perceptual Image Patch Similarity):** calculated to measure the perceptual similarity of the original and revised images, thus indicating content fidelity.
- **Processing Efficiency:** average time per image (prompt generation + diffusion) and memory utilization were recorded for both environments.

All results were logged and visualized using Matplotlib for reproducibility and analysis.

4.2. Quantitative and Qualitative Results

Qualitative Evaluation

The qualitative analysis considered visual and cultural coherence. Generated images were matched with historical images documenting the Trần and Lý dynasties.

The evaluators appreciated how the system recreated features considered culturally significant like:

- **Dragon–Phoenix patterns** on armor and sword guards;
- **Lotus ornaments** and **aged bronze textures**;
- The primary colors of the drenched palette were red, gold, and bronze, with earthy tones satisfying the feudal Vietnam aesthetic.

Some images did show slight artifacts, like texture noise and asymmetrically intricate details, but these were the outliers. Most images exhibited strong cultural resonance with high stylistic fidelity.

Quantitative Evaluation

Table 2. *Quantitative evaluation results of the proposed Gemini–Stable Diffusion system.*

Metric	CPU-only Environment	GPU (T4) Environment	Description
MOS (Mean Opinion Score)	3.85 / 5	3.85 / 5	Based on 5 evaluators assessing visual quality, style accuracy, and aesthetics
FID (↓ better)	67.3	58.4	Lower values indicate better realism and diversity
LPIPS (↓ better)	0.273	0.251	Lower scores indicate higher perceptual similarity
Avg. Processing Time per Image	5–6 minutes	1–2 minutes	Including both prompt generation and diffusion
Memory Usage	~7.2 GB RAM	~6.4 GB VRAM	Recorded during peak generation

The results show the system achieved good perceptual quality, and the realism is still acceptable ((MOS = 3.85/5, FID = 58.4).

The CPU-only setup does show an increase in processing time but the visual output quality does remain stable, thus, it is proof that limited output hardware is deployable.

Error Analysis and Discussion

While evaluating the system, some failure cases were noted:

1. **Prompt Over-expansion:** in some cases, very long user directions resulted in repetitive and overstuffed features in the generated image.
2. **Cultural Misdirection:** a few times, Gemini generated artifacts with culturally inappropriate features like Japanese calligraphy and elements of samurai armor.
3. **Visual Artifacts:** excessive contrast in lighting and slightly blurred textures at higher diffusion steps may contribute to unwanted visual artifacts.

These issues indicate the potential of fine-tuning the language model with culturally relevant Vietnamese data sets, and the inclusion of ControlNet-based geometry constraints as the next step in the research.

The system makes a meaningful contribution to the digital heritage preservation and creative industries by providing them with a versatile tool that requires only a brief text to generate Vietnamese-style artifacts with high visual and semantic fidelity.

5. Conclusion and Future Work

This research developed a culturally tailored image generation system that integrates a large language model (Gemini) with a diffusion model (Stable Diffusion). This system enables users to articulate imaginative orders in everyday language and receive generated images in the Vietnamese style, drastically minimizing the time and effort required in traditional handmade designs.

From the conducted experiments, the system delivers excellent outcomes under constrained hardware resources. Furthermore, the system captures the semantics and aesthetics of the images in Vietnamese style. Assigning Gemini to act as a prompt designer illustrates its effectiveness in streamlining the users' imagination and creative workflow.

This research also holds practical value for the video game industry and digital content production, but it further expands the methodological scope for the use of AI for cultural heritage, especially in the preservation and reconstruction phases. The recommender system embodies a creative interdisciplinary prototype that merges art, language, and AI.

In the future, research will expand in the following directions:

- Build and train diffusion models based on the Vietnamese feudal historical style.
- The model will optimize pipelines for low-memory GPUs: Implement techniques, such as LoRA and quantization, that minimize model size while preserving quality and saving resources.
- Extending the system for artifact, costume, and ancient architecture reconstruction based on archaeological documents will aid in digitization and storage of Vietnamese cultural heritage.

This study aims to introduce generative artificial intelligence to indigenous cultural spaces, reiterating the potential language models and diffusion models hold in the reproducing and preservation of cultural heritage.

Broader Impact

The wider implications of this will focus on the deeply embedded responsive system to cultural heritage and the long enduring impact that preserving and revitalizing Vietnamese culture will bring.

In the reconstructions of the aesthetics of the past using generative AI, we will continue the digital preservation of heritage. The system proposes a restorative vision of artifacts, garments, and folk designs which may no longer exist and allows their virtual manifestation to be experienced by the present and future in a meaningful and participatory manner.

Furthermore, the inclusion of Vietnamese stylistic elements in AI-created media aids in the preservation of Vietnamese cultural identity in the economically globalized creative space. As there is more and more dependence on Western or East Asian aesthetics, this system can work as a technological equalizer in the realm of cultural balance and representation by making sure Vietnam's artistic legacy is preserved and transformed in a meaningful way in the digital arts of today.

Future Opportunities

There are exciting opportunities for the framework developed in this study to expand to other fields and disciplines. In the fine arts and design, this can be advanced as an AI creative co-pilot to help designers and artist's sketch out new prototypes for traditional design synthesis that artists can work on, even if they don't have extensive experience with the AI tools. For art students, exciting new project opportunities lie in diffusion-based image generation as a historical form art abstraction and innovative reinterpretation, art grounded in cultural revitalization.

In the history and art education, the system can be infused in digital teaching tools designed to help students learn inner-actively and visually through the lens of history, architecture, and culture. For history and art teachers, the system can be used to visually and accurately generate teaching aids that portray Vietnamese feudal aesthetics, transforming history lessons standardized in textbooks to engage and capture student attention.

This approach could further expand to museum technology, digital exhibitions, and virtual tourism, where automatically created culturally respectful visuals could supplement public engagement and assist the sharing of heritage information in accessible and visually attractive forms.

References

1. Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." arXiv preprint arXiv:1508.06576 (2015).
2. Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
3. Yang, Hyemin, Heekyung Yang, and Kyungha Min. "Artfusion: A Diffusion Model-Based Style Synthesis Framework for Portraits." Electronics 13.3 (2024): 509.
4. Wang, Yunxiang, et al. "Visual style transfer for cultural heritage: A comprehensive review of techniques and applications." Design and Artificial Intelligence (2025): 100027.
5. Comanici, Gheorghe, et al. "Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities." arXiv preprint arXiv:2507.06261 (2025).
6. Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." *European conference on computer vision*. Cham: Springer International Publishing, 2016.
7. Huang, Xun, and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization." Proceedings of the IEEE international conference on computer vision. 2017.
8. Hertz, Amir, et al. "Prompt-to-prompt image editing with cross attention control." *arXiv preprint arXiv:2208.01626* (2022).
9. Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." Proceedings of the IEEE/CVF international conference on computer vision. 2023.
10. Llamas, J., Lerones, P. M., Medina, R., Zalama, E., & Gómez-García-Bermejo, J. (2016). *Applying deep learning techniques to cultural heritage images within the INCEPTION project.* In *Euro-Mediterranean Conference* (pp. 25–32). Springer, Cham.
11. Jing, Y., Yang, Y., Chen, Z., Su, T., Ye, J., & Liu, X. (2019). Neural style transfer: A review. *IEEE Transactions on Visualization and Computer Graphics*, 26(11), 3125-3140.
12. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-Rank Adaptation of Large Language Models. *arXiv preprint arXiv*:2106.09685.
13. Li, Z., & Chen, J. (2024). AI-assisted cultural heritage preservation: Challenges and opportunities. *International Journal of Architectural Heritage*, 18(3), 405-420.