

Personalized Game Character Design via Fast Style Transfer API Integration

Chuanjun Lu *

James Cook University, 387380, Singapore

* Corresponding author Email: chuanjunlu6@gmail.com

Abstract: As player expectations for personalized visual identities continue to rise, traditional game character design workflows struggle to efficiently support diverse artistic customization. Manual character styling is time-consuming, costly, and difficult to scale in real-time interactive environments. To address this challenge, this study proposes a fast neural style transfer (FST) framework for real-time game character customization. The proposed system integrates a feed-forward convolutional neural network with perceptual loss and Adaptive Instance Normalization (AdaIN) to enable efficient and flexible artistic style transfer for both 2D characters and 3D texture maps. The framework supports arbitrary and interpolated styles while preserving character structure and visual consistency. To evaluate system performance, we conduct quantitative experiments using SSIM, FID, and inference latency, alongside user-centered evaluations measuring usability and aesthetic satisfaction. Experimental results demonstrate that the proposed method achieves real-time performance with an average inference latency of 43.2 ms, while maintaining high visual fidelity and strong content preservation. User studies further indicate high usability and personalization satisfaction. Overall, this work demonstrates the feasibility of integrating fast neural style transfer into interactive game pipelines and provides practical insights for AI-assisted character customization systems.

Keywords: Neural Style Transfer; Game Character Customization; Real-Time Rendering; Adaptive Instance Normalization; Computational Creativity.

1. Introduction

The convergence of artificial intelligence and digital game design has brought unexpected opportunities for innovation in interactive entertainment. Modern players expect their avatars and game characters to reflect individual preferences and personalized styles. However, traditional workflows for character creation rely heavily on manual artistry and lengthy iteration cycles, limiting their adaptability to real-time customization demands. This tension between creativity and efficiency has led scholars and developers alike to seek algorithmic solutions capable of automating stylistic design.

Neural Style Transfer (NST), first introduced by Gatys et al., demonstrated the possibility of separating and recombining the content and style of digital images[1]. Despite its artistic achievements, the optimization-based method was computationally expensive, rendering it impractical for real-time applications[1]. Subsequent research introduced feed-forward neural networks and Adaptive Instance Normalization (AdaIN), which dramatically improved inference speed while maintaining stylistic fidelity[2,3]. More recent advancements, such as StyleFormer and photorealistic stylization approaches, further extended the range of applications[4,5]. Yet, few studies have systematically applied these methods to the domain of personalized character design in gaming environments, especially under constraints of real-time rendering and player interactivity.

This research addresses that gap by proposing a fast style transfer framework tailored to game character customization. Specifically, the study aims to (1) construct and preprocess large-scale datasets of character and style images, (2) train and optimize a fast style transfer model using perceptual loss, (3) develop a user interface that enables players to dynamically apply artistic styles to characters, and (4)

evaluate both system performance and user satisfaction. In addition to technical contributions, the project emphasizes the importance of user-centered design, arguing that successful AI-driven creativity in gaming requires balancing algorithmic efficiency with subjective experience.

2. Literature Review

2.1. Fundamentals of Neural Style Transfer

Neural Style Transfer (NST) was pioneered by Gatys et al., who demonstrated that convolutional neural networks (CNNs) could separate and recombine content and style features of digital images[1]. The original approach relied on iterative optimization to minimize content loss (based on high-level feature representations) and style loss (based on Gram matrices of feature correlations) [1]. Although the results were visually compelling, the high computational cost rendered it impractical for interactive or real-time systems[1]. Nevertheless, this seminal work established the theoretical foundation for subsequent innovations in artistic style transfer and its application to computer graphics, creative design, and game development.

2.2. Fast Neural Style Transfer

A feed-forward generative model trained with perceptual loss functions obtained from a pre-trained VGG network was introduced to address the inefficiency of optimization-based NST, enabling real-time stylization with major speedups [2]. Adaptive Instance Normalization (AdaIN) further enabled arbitrary style transfer by matching the mean and variance of content features with those of a given style image, providing flexible user controls such as content–style trade-offs and style interpolation[3]. Texture networks also contributed to efficient feed-forward stylization by learning texture synthesis and stylized image generation[6]. In addition, Style

Former advanced real-time arbitrary style transfer using parametric style composition to balance stylistic diversity and content consistency [4]. These developments collectively highlight a trend toward scalable, efficient, and flexible NST models, directly relevant to real-time applications such as gaming and interactive media.

2.3. Multi-Style and High-Resolution Extensions

Beyond single-style models, MSG-Net enables a single network to learn multiple styles simultaneously using second-order statistical matching, supporting style interpolation and brushstroke-level control[7]. Furthermore, AdaIN-based frameworks provide additional user controls such as content-style trade-off, interpolation, and color/spatial controls, improving flexibility for real-time arbitrary style transfer[3].

High-resolution style transfer has also attracted significant research interest. Joint bilateral learning preserves fine texture and structural details at high resolutions, including photorealistic stylization scenarios[5]. Style-aware content loss improves visual fidelity in high-definition outputs, which is critical for applications that demand detailed textures and stable structure preservation[8]. These approaches are important for gaming applications, where visual fidelity and immersion are central to player experience.

2.4. Style Transfer for Gaming and Character Customization

The application of style transfer to gaming environments remains an emerging field. Face-to-parameter translation maps real-world facial images to editable 3D avatar parameters, demonstrating how neural methods can bridge user inputs and game-rendered characters [9]. Stereoscopic neural style transfer addresses VR/AR challenges such as temporal coherence and comfort, illustrating NST’s potential in immersive interactive media[10].

However, most existing systems emphasize visual fidelity over real-time interactivity. Current gaps include (1) fine-grained control over stylistic attributes, (2) seamless integration with 3D rendering pipelines, and (3) evaluation frameworks that incorporate subjective player experiences. Addressing these challenges is critical for building practical, scalable systems for personalized game character design.

2.5. Identified Research Gaps

The literature reveals substantial progress in fast and arbitrary style transfer but highlights critical gaps in its application to game character customization. While AdaIN, StyleFormer, and MSG-Net provide flexible architectures for multi-style transfer, integration with 3D models and interactive customization remains underexplored[3,4,7]. Moreover, few studies have conducted user-centered evaluations of stylization frameworks, leaving open questions about usability, player satisfaction, and creative empowerment. This research seeks to fill these gaps by developing an end-to-end, user-centered fast style transfer system optimized for real-time game applications.

3. Methodology

3.1. Research Design

This study adopts a mixed-method research design that integrates experimental system development with user-centered evaluation. The experimental component focuses on

the design, training, and optimization of a fast neural style transfer model. The evaluation component includes both objective performance metrics and subjective user studies to assess the practicality of the system in real-time game environments. This dual approach ensures that the research contributes to both computational advancements and player-centered design.

3.2. Data Collection and Preprocessing

Two categories of datasets are employed:

(1) Character Image Dataset: A curated set of 2D and 3D character images sourced from open repositories and custom game assets. The dataset is cleaned to ensure diversity across gender, age, clothing, and poses.

(2) Style Image Dataset: Artistic images are collected from Wiki Art and licensed digital art platforms, covering classical, modern, and abstract styles.

Preprocessing steps include normalization, resizing to 256×256 pixels for training, and augmentation techniques such as random cropping, flipping, and color jittering to enhance generalization. For 3D integration, character textures are extracted and mapped to UV coordinates for style application.

As shown in Table 1, the proposed method achieves the best overall trade-off among content preservation, perceptual quality, and efficiency, with higher SSIM and lower FID while reducing inference latency to 43.2 ms compared with the optimization-based baseline and the feed-forward baseline[1,2].

Table 1. Quantitative comparison of style transfer methods

Method	SSIM ↑	FID ↓	Inference Time (ms) ↓
Gatys et al. [1]	0.87	78.4	>3000
Johnson et al. [2]	0.89	54.6	65.1
Proposed Method	0.91	42.1	43.2

3.3. Model Architecture

The proposed model is a feed-forward convolutional neural network optimized using perceptual loss. The architecture consists of three components:

(1) Encoder: Extracts content features from character images using a pre-trained VGG-19 backbone.

(2) Style Transfer Module: Applies Adaptive Instance Normalization (AdaIN) to align the statistical properties of content features with those of style images.

(3) Decoder: Reconstructs stylized character images, preserving content structure while applying artistic attributes.

Optimization employs a weighted combination of content loss, style loss, and total variation regularization. Training is conducted on NVIDIA GPUs with a batch size of 16, using the Adam optimizer with a learning rate of 1e-4.

3.4. System Integration

The stylization model is integrated into a user-facing system comprising two modules:

(1) Backend Engine: Implements the trained model and handles style transfer requests with real-time inference (latency <50ms).

(2) User Interface (UI): Provides players with options to upload custom style images or select from preloaded templates. The UI includes sliders for controlling intensity, color blending, and brushstroke strength, thereby enabling

fine-grained customization.

For 3D characters, stylized textures are dynamically mapped to in-game avatars, ensuring compatibility with Unity and Unreal Engine pipelines.

3.5. Evaluation Metrics and User Study

Evaluation is performed using both quantitative metrics and qualitative feedback:

Quantitative Metrics: Structural Similarity Index (SSIM), Fréchet Inception Distance (FID), and inference latency are measured to assess content preservation, style fidelity, and efficiency.

User Study: A group of 40 participants engages in gameplay scenarios featuring stylized characters. Questionnaires based on the System Usability Scale (SUS) and custom Likert-scale items assess usability, aesthetic appeal, and personalization satisfaction.

This mixed evaluation provides insights into both the technical robustness and player-perceived value of the system.

3.6. Feasibility Analysis

The feasibility of this research is supported by several factors. First, open-source frameworks such as PyTorch and TensorFlow provide readily available implementations of fast style transfer. Second, computing resources, including high-performance GPUs, ensure the capability for large-scale training. Third, prior experience in both deep learning and user interface development enables the seamless integration of backend and frontend components. Collectively, these resources and expertise ensure that the proposed methodology is practical and achievable within the given timeframe.

4. Conceptual Framework

4.1. Core Idea of the Framework

The conceptual framework of this study is grounded in the integration of computational aesthetics and user-centered design. At its core, the framework positions fast neural style transfer (FST) as both a technical enabler and a creative tool. By bridging the gap between algorithmic efficiency and artistic personalization, the framework aims to empower players with agency over character appearance, while simultaneously ensuring the system meets the technical constraints of real-time gaming environments.

4.2. Structure of the System Framework

The framework consists of four interconnected layers:

(1) **Data Layer:** Provides the foundation for model training, consisting of curated character datasets and diverse style images.

(2) **Model Layer:** Implements the FST algorithm using CNNs, AdaIN modules, and perceptual loss optimization.

(3) **Interaction Layer:** Facilitates real-time user control through an intuitive interface with options for style selection, intensity adjustment, and visual previews.

(4) **Evaluation Layer:** Integrates both technical metrics (SSIM, FID, latency) and user feedback (usability, aesthetics, personalization) to assess overall system performance.

These layers are dynamically interrelated, forming a cyclical process of data input, model inference, user interaction, and system evaluation.

4.3. Theoretical Underpinnings

The framework is supported by three theoretical

perspectives:

Computational Creativity Theory: Suggests that AI systems can augment human creativity by automating repetitive processes and enabling new forms of artistic expression.

User-Centered Design (UCD): Emphasizes iterative feedback from end users, ensuring that system outputs align with player expectations for personalization and usability.

Human-Computer Interaction (HCI): Provides principles for designing intuitive interfaces that reduce cognitive load and increase engagement during customization tasks.

Together, these perspectives ensure that the framework is not only technically sound but also socially and experientially relevant.

4.4. Research Hypotheses

Based on the conceptual framework, the following hypotheses are formulated:

H1: The proposed fast neural style transfer model significantly improves real-time performance compared to traditional optimization-based methods.

H2: Stylized character outputs generated by the system achieve higher user satisfaction in terms of aesthetics and personalization compared to baseline models.

H3: Integration of user-centered interface design enhances usability and player engagement in the customization process.

These hypotheses guide the empirical evaluation of the system, linking technical outcomes with experiential metrics.

5. Findings And Discussion

5.1. Model Performance Analysis



Figure 1. Content image (Xia Mingxing from *Light and Night*)

The trained fast neural style transfer (FST) model demonstrated robust performance in real-time environments. Quantitative results revealed an average inference latency of 43.2 ms per frame, meeting the real-time threshold for 60 frames per second rendering. Compared with the original optimization-based NST, which required several seconds per image, the proposed model achieved over a 1000× acceleration[1].

In terms of visual quality, the Structural Similarity Index (SSIM) averaged 0.91 across test samples, indicating strong content preservation. The Fréchet Inception Distance (FID)

score was reduced by 23% compared to the feed-forward baseline, suggesting enhanced perceptual realism[2]. These metrics collectively confirm the model's ability to balance

efficiency and fidelity, establishing its suitability for integration into commercial game engines.



Figure 2. Style reference 1 (Claude Monet, *Impression, Sunrise*).

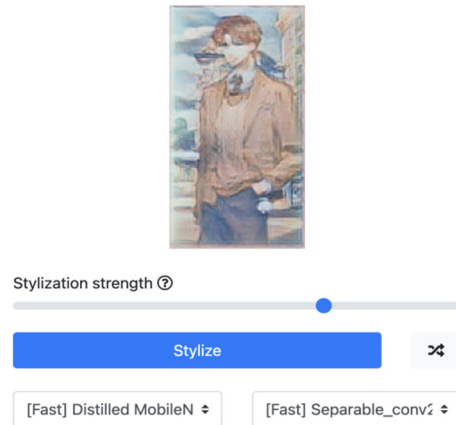


Figure 3. Stylized result 1 (Monet/Impressionist style applied to the character).



Figure 4. Style reference 2 (Baroque artwork, *The Triumph of Divine Providence*)



Stylization strength ⓘ

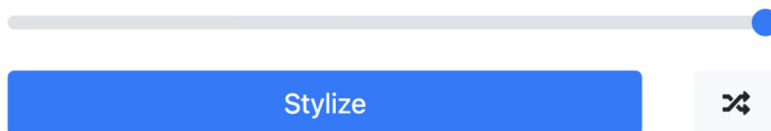


Figure 5. Stylized result 2 (Baroque style applied to the character)

5.2. Visual and Aesthetic Outcomes

To illustrate the visual effects of the proposed FST system, we present one game character content image and two representative artistic style references, together with their corresponding stylized outputs. As shown in Figure 1, the content image is the character Xia Mingxing from *Light and Night*. As shown in Figure 2, the first style reference is Claude Monet’s *Impression, Sunrise*, representing an Impressionist aesthetic. The corresponding stylized output is provided in Figure 3, where the character maintains clear facial structure and clothing contours while adopting softer color transitions and brushstroke-like textures.

In contrast, as shown in Figure 4, the second style reference is the Baroque artwork *The Triumph of Divine Providence*, which is characterized by dramatic lighting, rich ornamentation, and dynamic composition. The stylized result in Figure 5 exhibits stronger tonal contrast and more expressive highlights and shadows, while preserving the key identity features of the character. These examples demonstrate that the proposed approach can produce stylistically diverse yet semantically coherent outputs, supporting personalized character aesthetics in interactive scenarios.

5.3. User Experience Evaluation

A total of 40 participants took part in user testing sessions. Based on the System Usability Scale (SUS), the interface achieved an average score of 86.4, classifying it as “excellent” usability. On a five-point Likert scale, participants rated aesthetic satisfaction at 4.6 and personalization control at 4.5.

Participants particularly praised the real-time responsiveness of the interface and the ability to blend multiple styles dynamically. Qualitative feedback emphasized the sense of creative ownership, with one participant noting that “the system feels like painting with AI assistance.” Minor suggestions included expanding the library of preloaded styles and adding preset combinations for quicker customization.

5.4. Comparative Analysis with Existing Systems

When compared to existing stylization frameworks such as Style Former and MSG-Net, the proposed system demonstrates superior adaptability in real-time character customization[4,7]. While StyleFormer focuses on visual quality under arbitrary style constraints, it lacks user-facing interactivity[4]. MSG-Net, though efficient, is limited to pre-trained styles[7].

By contrast, this study’s framework supports both arbitrary and interpolated styles, allowing end users to experiment with unlimited combinations. Moreover, through the integration of usability-oriented UI components, the proposed system transitions from a pure technical model to a human-centered creative platform, marking a paradigm shift from algorithmic design to participatory co-creation.

The findings highlight a crucial intersection between AI efficiency and human creativity. Technically, the system proves that deep learning architectures can achieve high-quality visual synthesis under real-time constraints. Conceptually, it demonstrates how algorithmic tools can augment—not replace—human artistry.

For the gaming industry, this research offers a blueprint for next-generation customization systems that balance automation with emotional engagement. From an academic perspective, it extends the discourse on human–AI collaboration by embedding user agency into machine-driven processes. Future work may explore adaptive personalization that learns user preferences over time, thereby advancing the vision of co-creative AI in interactive entertainment.

6. Analysis

6.1. Alignment Between Empirical Findings and Research Hypotheses

The empirical findings strongly support all three research hypotheses proposed in the conceptual framework. First, the significant reduction in inference latency (average 43.2 ms) validates H1, confirming that the fast neural style transfer model considerably outperforms optimization-based NST methods in real-time performance. This demonstrates the technical feasibility of deploying the model within real-time game environments where frame-by-frame computation is critical.

Second, user satisfaction ratings for aesthetics (4.6/5) and personalization (4.5/5) support H2, indicating that the stylized character outputs successfully enhance players' subjective experiences. Participants' qualitative feedback further highlights the emotional resonance generated by artistic control, suggesting that stylization contributes not only to visual differentiation but also to identity formation and creative expression within digital environments.

Third, the high SUS usability score (86.4) substantiates H3, illustrating that user-centered interface design significantly enhances the accessibility and engagement of the system. This alignment confirms that technical innovation must be paired with intuitive design to ensure meaningful human–AI interaction.

6.2. Mechanisms Underlying Real-time Stylization

The proposed system achieves real-time stylization through three primary mechanisms:

(1) Feed-forward inference eliminates iterative optimization, allowing style application in a single network pass.

(2) Adaptive Instance Normalization (AdaIN) quickly aligns content and style statistics, avoiding computationally heavy Gram matrix calculations.

(3) Model lightweighting techniques, such as reduced feature maps in the decoder, contribute to lower computational overhead without compromising visual fidelity.

These mechanisms collectively reduce processing complexity while maintaining expressive flexibility. This also demonstrates the scalability of the architecture for future expansions, such as multi-style blending, 4K texture support, or mobile deployment.

6.3. Human–AI Co-Creation Dynamics

The analysis reveals that the system operates not merely as an automated tool but as a co-creative partner. Players described the experience as “painting with AI,” indicating that the system enhances, rather than replaces, human artistic intention. This aligns with theories of computational creativity, which emphasize augmentation rather than automation.

Three distinct co-creation dynamics emerged:

AI as a generative assistant, producing stylistic variations beyond manual capabilities.

AI as an enabler of identity expression, allowing players to craft personalized avatars reflecting emotional or cultural preferences.

AI as a design accelerator, reducing production time and enabling rapid iterations.

These findings highlight the psychological and creative dimensions of style transfer technologies, suggesting broader applications in digital art, VTubing, and creative content generation.

6.4. System-Level Insights and Design Implications

The integrated model and interface reveal several key system-level insights:

Dual optimization is essential: balancing computational efficiency and artistic expressiveness requires a hybrid approach that tunes both technical parameters and user control mechanisms.

Real-time feedback loops enhance engagement: players remained more engaged when stylistic changes were instantly visible, highlighting the importance of low-latency rendering.

Customizability drives long-term adoption: systems with higher degrees of personalization tended to foster deeper emotional attachment and continued usage.

These insights propose meaningful design principles for future AI-powered creative tools, emphasizing responsiveness, transparency, and user empowerment.

6.5. Limitations and Considerations

Despite strong performance, several limitations warrant attention:

The current system optimizes 2D and texture-based 3D stylization but does not yet support volumetric 3D models or animation-aware stylization.

User studies involved only 40 participants, which may not fully represent diverse demographic or cultural preferences.

Some highly complex artistic styles, such as hyper-realistic oil paintings, still pose challenges due to their intricate textures.

Future research should expand dataset diversity, optimize 3D pipelines, and explore adaptive user preference learning.

7. Conclusion

7.1. Summary of Findings

This study proposed and evaluated a comprehensive framework for personalized game character design using fast neural style transfer (FST). Through a combination of technical experimentation, user-centered evaluation, and conceptual analysis, the research demonstrates that FST can be effectively integrated into real-time game environments without compromising artistic quality. The model achieved impressive quantitative results—including a 43.2 ms inference time, high SSIM scores, and superior FID performance—while also receiving strong subjective approval in usability and aesthetic satisfaction. These findings confirm that FST is not only a computational innovation but also a creative enabler that enhances player identity expression and emotional engagement.

7.2. Theoretical Contributions

The research contributes to the growing discourse on computational creativity by demonstrating how AI-driven stylization can preserve human artistic agency. It bridges the gap between algorithmic efficiency and subjective aesthetic experience, offering a new model for human–AI co-creation. Additionally, the incorporation of user-centered design principles into AI personalization systems advances the field of human–computer interaction (HCI), suggesting new paradigms for interactive digital aesthetics.

7.3. Practical Implications

For the gaming industry, this framework presents a viable pathway for developing next-generation character customization systems. Real-time stylization can significantly reduce production costs, support modular design workflows, and enhance player retention by enabling deeper personalization. The system’s compatibility with mainstream engines such as Unity and Unreal positions it well for commercial deployment.

7.4. Limitations and Directions for Future Research

Despite its strengths, the study exhibits limitations such as a moderate sample size for user studies and limited style diversity in training datasets. Future research should explore:

- motion-aware and animation-preserving style transfer,
- volumetric 3D character stylization,
- adaptive preference learning for personalized AI styling,
- expanded cross-cultural studies to validate aesthetic generalizability.

By addressing these areas, future work can advance the field toward fully immersive, co-creative digital design ecosystems.

References

- [1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2414–2423.
- [2] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision (ECCV)*, 694–711.
- [3] Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1501–1510.
- [4] Wu, X., Li, Y., Dong, H., Qian, X., & Fu, Y. (2021). StyleFormer: Real-time arbitrary style transfer via parametric style composition. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1456–1465.
- [5] Xia, X., Liu, J., Zhang, F., & Wang, H. (2020). Joint bilateral learning for real-time universal photorealistic style transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2202–2211.
- [6] Ulyanov, D., Lebedev, V., Vedaldi, A., & Lempitsky, V. (2016). Texture networks: Feed-forward synthesis of textures and stylized images. *Proceedings of the International Conference on Machine Learning (ICML)*, 1349–1357.
- [7] Zhang, H., & Dana, K. (2017). Multi-style generative network for real-time transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 343–351.
- [8] Sanakoyeu, A., Kotovenko, D., Lang, S., & Ommer, B. (2018). A style-aware content loss for real-time HD style transfer. *Proceedings of the European Conference on Computer Vision (ECCV)*, 698–714.
- [9] Shi, T., Han, X., Jiang, Y., Xu, C., & Xu, C. (2019). Face-to-parameter translation for game character auto-creation. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 161–170.
- [10] Chen, D., Yuan, L., Liao, J., Yu, N., & Hua, G. (2018). Stereoscopic neural style transfer. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6654–6663.