A Study on Character Consistency Generated in [Midjourney V6] Technology

Xi Chen , Jeanhun Chung*

Doctor, Dept. of Multimedia, Gradute School of Digital Image and Contents
Dongguk University, Korea
*Professor, Dept. of Multimedia, Gradute School of Digital Image and Contents
Dongguk University, Korea
chen_chen1211@hotmail.com, *evengates@gmail.com

Abstract

The emergence of programs like Midjourney, particularly known for its text-to-image capability, has significantly impacted design and creative industries. Midjourney continually updates its database and algorithms to enhance user experience, with a focus on character consistency. This paper's examination of the latest V6 version of Midjourney reveals notable advancements in its characteristics and design principles, especially in the realm of character generation. By comparing V6 with its predecessors, this study underscores the significant strides made in ensuring consistent character portrayal across different plots and timelines.Such improvements in AI-driven character consistency are pivotal for storytelling. They ensure coherent and reliable character representation, which is essential for narrative clarity, emotional resonance, and overall effectiveness. This coherence supports a more immersive and engaging storytelling experience, fostering deeper audience connection and enhancing creative expression.The findings of this study encourage further exploration of Midjourney's capabilities for artistic innovation. By leveraging its advanced character consistency, creators can push the boundaries of storytelling, leading to new and exciting developments in the fusion of technology and art.

1. Introduction

With the rapid development of artificial intelligence image generation technology, artificial intelligence-driven Text Text-to-Image Generation (T2I) technology is a rapidly developing field, which can automatically generate corresponding images according to a given text description. The technology has shown great potential in image editing, virtual reality, game development, and multimedia content production. In this technique,

consistency studies focus on solving the problem of high matching between text descriptions and the generated image, ensuring that the image is not only visually appealing but also accurately reflects the intentions and details of the text content. It plays a crucial role in enhancing the user experience and designing the application time. Through further understanding of the text description to more accurately convey the user's design intention and express emotion. As technology advances, users, especially those engaged in design work, can use advanced consistency technologies to create more diverse and vivid plot-consistent designs. This helps to strengthen the narrative effect and user experience of the design.

1.1 Overview of Development

AIGC (Artificial Intelligence Generated Content) is a rapidly developing field in AI technology. It involves using machine learning and deep learning models to generate text, images, music, and video content. These technologies can be applied in various fields, such as automated news reporting, artistic creation, and advertising creativity. Among them, taking image generation as an example, models like DeepDream and StyleGAN are already capable of producing high-quality images. These models excel in style transfer, facial generation, and even creating entire fictional scenes. Among the user groups, Midjourney has the broadest base; after the release of its V6 update, it offers users more powerful character consistency. Midjourney can generate images that more closely match character traits, such as facial expressions, hairstyles, and clothing. It also offers more flexible character applications, where character consistency can be applied across multiple characters, facilitating users in creating more complex character scenes.

1.2 Field of Investigation

The research content covers the impact of artificial intelligence image generation technology and character consistency design on narrative coherence. Through model comparison experiments, the practical significance of existing AIGC technology for character generation training is discussed.

1.3 The Importance of Conducting the Research Task

Improve accuracy: Ensure that the generated image matches the input text description as much as possible, improving the accuracy and controllability of image generation.Enhanced user experience: Taking Midjourney as an example, improving consistency can significantly enhance users' recognition of narrative coherence and consistency.Promote technological innovation: By solving the consistency problem, we can promote the continuous innovation and optimization of related technologies, opening up new possibilities for the application of artificial intelligence in the field of artistic creativity, especially script narrative creation.

2. Purpose of Study Execution

Figure 1. Generate 2D Character Multi-angle Reference Materials Through V6 Model

First, the technical features of the Midjourney graphic consistency system are analyzed in detail and compared with the advantages and disadvantages of previous versions (such as the results of other system versions, such as the V5V4 engine). Next, examine how the Midjourney system provides a new means of expression for role generation by providing more consistent and coherent effects, especially in the design of role theme consistency, attracting the attention of users to design and creation, and forming a unified style. Finally, based on the characteristics of Midjourney technology, explore and propose new means of design creation, and discuss how to maximize the use of consistency characteristics to enhance the consistency of design effect and design style.

2.1 Improve Accuracy and Efficiency of Character Generation

By introducing deep learning and machine learning algorithms, this research aims to train models for character consistency, especially expression accuracy in terms of complex and subtle expressions and movements. In addition, this study aims to analyze the improvement of narrative fluency through character consistency generation training to reduce the shortcomings of insufficient immersion and coherence caused by character inconsistency, thereby improving the overall viewing effect and production efficiency.
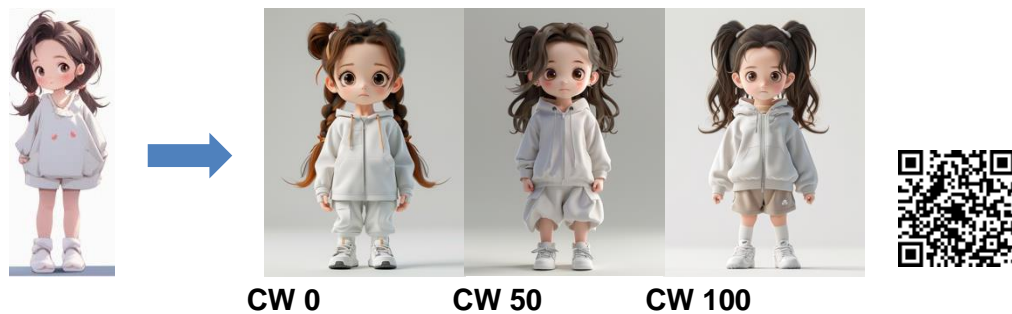
3. Problem Points and Solutions

CW 0   CW 50   CW 100

Figure 2. Generate Detailed and Differentiated Characters Through Different Reference Values --CW

3.1 Adverse Effects of Insufficient Role Consistency

Role consistency plays a crucial role in storytelling as it significantly impacts the coherence and logic of a story.

Fragmented Logic: Inconsistencies in a character's behavior, appearance, or personality throughout a story can lead to fragmented logic within the narrative. Readers or viewers might become confused or lose their motivation to follow the story, affecting their engagement and understanding of the storyline.

Incoherent Information: If a character displays inconsistent traits or actions across different scenes, it can result in information inconsistencies. Such inconsistencies can hinder the audience's understanding of the story's plot, making the overall narrative confusing or less believable.

Weakened Emotional Connection: Role consistency is crucial for establishing an emotional connection between readers/viewers and characters. Inconsistent character portrayals across different contexts can weaken the emotional bond between the audience and the characters, reducing emotional investment and empathy.

Impact on Brand Image: In marketing or advertising, a lack of role consistency can harm a brand's image. If the representative character of a brand exhibits inconsistencies across various ads or platforms, it can lead to a confused and unstable brand image, affecting consumer trust and loyalty.

In conclusion, role consistency plays a vital role in storytelling. Maintaining role consistency helps ensure the coherence, logic, and emotional appeal of a story, while inconsistencies can have negative effects on the story, impacting the audience's experience and comprehension.

3.2 Experiment Design and Implementation

The training uses the v6 version to generate a 2D character model. After modifying the relevant prompt words to generate a 3D image of the character, and obtaining a stable 3D effect character, the character's expressions and movements are enriched through partial redrawing. Evaluate the accuracy, consistency and stable output capabilities of the selected AI model through cross-validation and other methods. Assessing the performance of the character congruence maintenance effect under narrative scene conditions.

3.3 Model Training

As shown in the figure, the first step is to edit and determine the basic description of the protagonist's appearance in the story, and set the descriptor prompt to meet the requirements for generating Midjourney Vincentian diagrams. After multiple character screenings and adjustments, a basic reference character model is selected, and then through consistent reference operations, the required multi-angle, rich expression and action character model materials are generated. Combining the character material with the story and narrative scenes, the required relatively complete picture is finally generated.

The current upgraded V6 version of Midjourney utilizes CREF values to adjust role consistency. CREF stands for "Conditional Role Embedding Feature," which is a vector containing information about a character's identity and features, such as name, age, gender, and personality.

How CREF Values Work:

Creating a Character: First, you need to use the /imagine command to create a character. The prompt will include relevant information about the character, such as name, age, gender, and personality.

Generating CREF Values: Midjourney will generate a CREF value based on your prompt, which will encompass information about the character.

Adjusting Role Consistency: In subsequent generations, you can use the /imagine command and add the cref parameter to pass the generated CREF value to the model. This ensures that the generated image remains consistent with the character you have created.

4. Results and Analysis

4.1 Analysis of Application Effect of Role Consistency in Narrative Scenes

By training the updated Midjourney model for character consistency, notable differences arise when compared to the previous version that relied on placeholder seed values and other constraints for character features. The new version adopts a targeted approach in analyzing and sampling reference images, aiming for improved role consistency and the portrayal of intricate expressions and details through flexible redraw and keyword adjustment settings. While the new version has encountered minor bugs, such as inconsistencies in depicting fingers during training sessions, there have been significant enhancements in overall image quality, rendering speed, and the interpretation of keywords and references.For users involved in picture book creation or character design, these enhancements boost work efficiency, streamline operational procedures, and elevate the user experience by ensuring storytelling completeness and coherence.

4.2 Current Technical Problems

The Midjourney V6 model still has many problems in generating character details, such as the number of fingers and uncoordinated body movements in the old version. This is still a technical problem that needs to be overcome in the next upgrade of the system version. In addition, the new version has prompt words A large number of updates were made to the game, requiring adjustments to descriptors to achieve consistency with Vincentian characters.

Figure 3. Stable Output of Character Images and Actions Through Reference Images

5. Midjourney's Strengths in Role Consistency

Midjourney's strengths in role consistency, particularly evident since the release of its V6 version, lie in several key areas that enhance the user experience in generating consistent and detailed character-driven images.

Enhanced Detail Fidelity: Midjourney V6 significantly improves the generation of consistent details across various images of the same character. This means that facial features, hairstyles, and clothing remain consistent when the character is depicted in different scenes or from different angles. This attention to detail is crucial for users who require cohesive visual storytelling or branding.

Improved Character Recognition: With the latest update, Midjourney has improved its AI's ability to recognize and maintain specific character traits over multiple sessions. This capability is particularly useful in projects involving long-term character development, such as graphic novels or continuous marketing campaigns.

Customizable Traits: Users have better control over defining and customizing specific traits of the characters. This allows for more precise adjustments and refinements in character portrayal, which can be critical for creative professionals who need characters to convey specific emotions or messages.

Application Across Multiple Characters: The role consistency technology is not limited to individual characters but can be extended to multiple characters in the same scene. This feature is invaluable for creating complex narratives or scenes where interaction between characters must be believable and consistent.

6. Conclusion

Enhancing user experience and creative freedom remains a central focus for the Midjourney team, as evident in their commitment to consistently reduce learning costs through intuitive operation improvements. By offering more robust features, adjustable options, and expanding character consistency, Midjourney aims to widen its application range and provide users with a seamless experience. The integration of new technologies enhances realism in generated images, ensuring a closer alignment with character characteristics and overall creation quality.These progressive enhancements underscore Midjourney's ongoing dedication to advancing text-to-image generation tools. It is anticipated that Midjourney will continue to refine its character consistency function, offering users increased convenience, enhanced capabilities, and greater flexibility for creative expression in the future.

References

[1] CAO Yin, QIN Junping, MA Qianli, SUN Hao, YAN Kai, WANG Lei1,,REN Jiaqi,Survey of text-to-image synthesis.Journal of Zhejiang University (Engineering Science), Vol.58 ,No.2 pp.219-238, Feb. 2024

[2] JlANG Shangsong SANG Ruijuan, Application of Text Semantic lmage Generation Technology in Furniture Design, Furniture,Vol.44 No.4, pp.36-41, 2023

[3] Yang chenlu, Wan wanggen,Wang xuzhi，Image Semantic Understanding Based on Text-Aware and Non repetitive Word Generation，Industrial control computer, Vol. 36,No. 1,pp.105-109,2023.

[4] YANG lili, AI Image Generation: Ending the Art Contract,Art China,vol.22(03).pp54-57,2023

[5] Jihan Kim, A Study on the Usability of Digital Humans in New Media Contents,The International Journal of Advanced Smart Convergence, Vol.12 No.4,pp300-305,Dec,2023

[6] DOI: http://dx.doi.org/10.7236/IJASC.2023.12.4.293

[7] Ke Ma, Jeanhun Chung，A Research on 3D Texture Production Using Artificial Intelligence Softwear International Journal of Internet, Broadcasting and Communication Vol.15 No.4 pp178-184 (2023) DOI:http://dx.doi.org/10.7236/IJIBC.2023.15.4.178

[8] SHI tianyuan,Research on the Theory and Method of Dialogue Response Diversity and Persona-Consistency Based on Neural Network,Chongqing University,Oct,2022