

IJACT 23-6-46

Comparative Analysis of AI Painting Using [Midjourney] and [Stable Diffusion] - A Case Study on Character Drawing -

Pingjian Jie, Xinyi Shan, Jeanhun Chung*

*Master's Course, Dept. of Multimedia, Graduate School of Digital Image and Contents,
Dongguk University, Seoul, Korea*

*Doctor's Course, Dept. of Multimedia, Graduate School of Digital Image and Contents,
Dongguk University, Seoul, Korea*

**Professor, Dept. of Multimedia, Graduate School of Digital Image and Contents,
Dongguk University, Seoul, Korea*

*jjp952523045@gmail.com, yarina.0122@gmail.com, *evengates@gmail.com*

Abstract

The widespread discussion of AI-generated content, fueled by the emergence of consumer applications like ChatGPT and Midjourney, has attracted significant attention. Among various AI applications, AI painting has gained popularity due to its mature technology, user-friendly nature, and excellent output quality, resulting in a rapid growth in user numbers. Midjourney and Stable Diffusion are two of the most widely used AI painting tools by users. In this study, the author adopts a perspective that represents the general public and utilizes case studies and comparative analysis to summarize the distinctive features and differences between Midjourney and Stable Diffusion in the context of AI character illustration. The aim is to provide informative material for those interested in AI painting and lay a solid foundation for further in-depth research on AI-generated content. The research findings indicate that both software can generate excellent character images but with distinct features.

Keywords: *Midjourney, Stable Diffusion, AI Painting, Character Design, AI-generated Content, Prompt, ControlNet*

1. Introduction

Midjourney and stable diffusion AI are both deep learning-based generative models that can automatically generate excellent images. As two of the most popular and outstanding AI drawing tools currently available, many people struggle with choosing between them. In order to understand the differences and advantages of these two tools, this paper will introduce their respective features. Taking character illustration as an example, the paper will compare and analyze the differences between them in terms of character appearance, facial expressions, poses, art style, and other aspects. This will provide convenience for individuals to make an informed choice between the two tools.

Manuscript Received: May 20, 2023 / revised: June 5, 2023 / accepted: June 10, 2023

Corresponding Author: evengates@gmail.com (Jeanhun Chung)

Tel: +82-2-2260-3767, Fax: +82-2-2260-3766

Professor, Dept. of Multimedia, Graduate School of Digital Image and Contents, Dongguk University, Seoul, Korea

2. Theoretical Background

Stable Diffusion is a deep learning model that was released in 2022. It is designed to generate detailed images conditioned on text descriptions and can be applied to tasks such as restoration, repair, and image translation. Developed in collaboration between Stability AI, academic researchers, and nonprofit organizations, Stable Diffusion is an open-source model. Unlike proprietary models like DALL-E and Midjourney that were previously accessible only through cloud services, Stable Diffusion can run on regular consumer-grade GPUs with a minimum of 8GB VRAM.

Midjourney is a generative artificial intelligence program and service created and hosted by Midjourney, Inc. It is capable of generating images based on natural language descriptions and can be accessed through Discord bot commands. Similar to OpenAI's DALL-E and Stable Diffusion, Midjourney is currently in the public testing phase and is led by David Holz, co-founder of Leap Motion.

3. Comparison of Basic Characteristics

Table 1. Comparison of Basic Characteristics

Characteristics	Stable Diffusion	Midjourney
Customization Level	High	Low
The Difficulty of Generating Quality Images	Hard	Easy
Inpainting	Yes	No
Out painting	Yes	No
Adjust aspect ratio	Yes	No
Number of models	1000+	10+
Negative prompt	Yes	No
Control composition and pose	Hard	Easy
Train your own model	Yes	No
Price	Free	10~60\$/month
Open source	Yes	NO
Content Filtering (Restriction)	No	Yes
Img2img	Yes	Yes
Hardware requirements	High	Low
Scalability	Yes	No

Stable Diffusion offers more options for customizing images, including size, prompt strength, number of generations, seed values, and samplers. In contrast, Midjourney has fewer options and allows only modifications of aspect ratio, seed, and stop criteria. Installation of Stable Diffusion is more complex, requiring the additional search and installation of required style models. On the other hand, Midjourney is user-friendly and easily accessible through Discord. It simplifies the process of generating artistic images with rich details, requiring less effort. Using Stable Diffusion requires more effort to create well-crafted prompts and experiment with models and parameter adjustments to generate similar quality images. It is open-source and offers over 1000 downloadable models in various styles, with further customization options. Midjourney has relatively limited models but allows stylization of images through additional parameters. Overall, Stable Diffusion provides more choices. Stable Diffusion offers various ways to control composition and poses, such as image-to-image, depth-to-image, guided pix2pix, and controlNet. In contrast, Midjourney can only control image generation through prompts, img2img, and seeds, albeit with some limitations. One of the appealing aspects

of Stable Diffusion is that users can create their own models and train them to match their personal preferences in style. Midjourney lacks this capability. Stable Diffusion operates offline, allowing users to generate content without restrictions or surveillance on their prompts. On the other hand, Midjourney restricts the production of certain themes by blocking explicit or violent keywords. Stable Diffusion requires local installation and utilizes the user's GPU for computation, making higher hardware demands. In contrast, Midjourney operates through cloud services with lower hardware requirements. Stable Diffusion supports the installation of many third-party plugins, such as ControlNet, while Midjourney is limited to official functionalities.

4. Character Production Comparison

4.1 Appearance (facial features, hair style, clothes)

“ Prompt: a cute girl, almond-shaped eyes, bob cut, button nose, oval face ”



Figure 1. Midjourney Facial Features

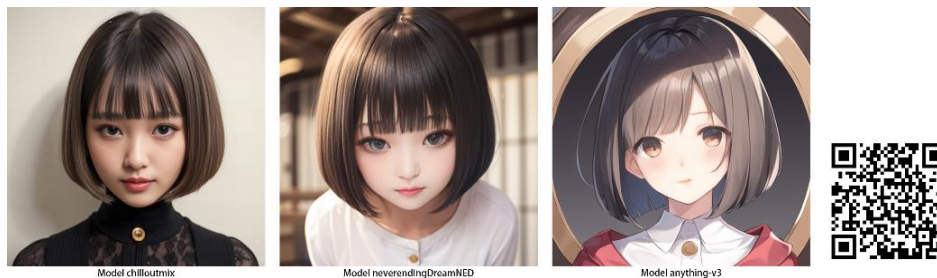


Figure 2. Stable Diffusion Version: 5ab7f21 Facial Features

“ Add Lora: a cute girl, almond-shaped eyes, bob cut, button nose, oval face <lora: japaneseDoll Likeness_v10:0.3> <lora: koreanDollLikeness_v15:0.3> <lora: taiwanDollLikeness_v10:0.2> ”

From the generated results, it is evident that both software can effectively depict the visual descriptions. Additionally, both software allows for the creation of images in different art styles by modifying the models or adding keywords. Under similar conditions, Midjourney produces higher-quality images and tends to yield outstanding results more easily. On the other hand, Stable Diffusion offers a broader range of styles by utilizing multiple models and Lora. As shown in Figure 3, it allows for fine-tuning to achieve realistic character depictions that align with personal aesthetics, resulting in higher consistency among the generated characters.



Figure 3. Stable Diffusion Model Chilloutmix Add Lora

When no specific age description is given for a girl, except in the case of 2D animation style, Midjourney's "a cute girl" tends to portray a child-like appearance, while Stable Diffusion leans towards a teenage representation. Overall, both software can achieve good results in visual depiction, each with its own advantages. Midjourney requires only prompt words and is more straightforward to obtain excellent results. On the other hand, Stable Diffusion offers greater flexibility with a wide range of models, plugins, and a greater number of adjustable parameters. There are slight differences in the artistic styles within the same genre between the two software applications.

4.2 Expression

“ Prompt: a cute girl, almond-shaped eyes, bob cut, button nose, oval face, crying ”



Figure 4. Midjourney Expression



Figure 5. Stable Diffusion Expression

When it comes to depicting a crying expression, both Midjourney and Stable Diffusion faced significant challenges. Midjourney's V4 model can draw a crying expression, but the color of tears is consistently incorrect. The V5 model portrays a more accurate "sad" expression. Similarly, Stable Diffusion's realistic style characters encountered similar issues in depicting crying expressions. The expressions lacked sufficient sadness and either did not depict tears or inaccurately rendered them. Surprisingly, both software applications

performed exceptionally well in drawing crying expressions in an anime style.

4.3 Pose

To obtain more accurate poses for character images, the following techniques will be utilized: Midjourney’s “image to image” and “describe image to text” functions, as well as Stable Diffusion’s “ControlNet open pose” and “Tagger image to text” functions, will be employed to depict the poses of the characters in the illustration below.

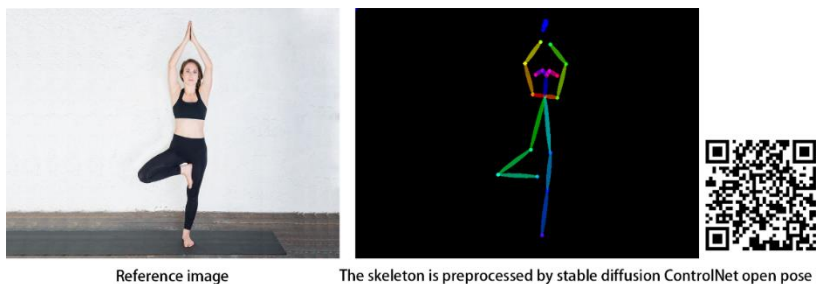


Figure 6. The Demonstrate Pose

In Midjourney, the following description can be obtained by using the "description" function of the reference image. And after importing the reference image, the result output is shown in Figure 7.

“ Prompt: <https://s.mj.run/gfYxAZNgiv4>, a cute girl, almond-shaped eyes, bob cut, button nose, oval face, doing her yoga, in the style of light maroon and dark black, precisionist lines, asymmetrical balance, smile core, Samikshavad, full body, clean-lined --ar 4:3 ”



Figure 7. The Posture Output in Midjourney

After multiple attempts, it was found that achieving a pose similar to the one in the reference image is quite challenging, often resulting in errors such as drawing three legs or four legs.

On the other hand, in Stable Diffusion, using its ControlNet open pose and Tagger image to text functionalities, we can obtain the results shown in Figure 8.



Figure 8. The Posture Output in Stable Diffusion

“ Prompt: a cute girl, almond-shaped eyes, bob cut, button nose, oval face, solo, brown hair, barefoot, pants, sports bra, arms up, black pants, midriff, navel, standing, long hair, full body, standing on one leg, braid, breasts ”

Using ControlNet open pose in Stable Diffusion yielded mostly accurate results, although there were some inaccuracies in the recognition of hand details. However, these can be addressed by utilizing the local repaint feature for refinement. Overall, in terms of pose drawing, Stable Diffusion is more convenient and accurate compared to Midjourney.

5. Conclusion

Based on the case studies and comparative analysis of Midjourney and Stable Diffusion in AI character drawing, both software can generate excellent character images with distinct features. From a practical perspective, Midjourney's simpler operation and user-friendly interface make it suitable for beginners, enabling effortless creation of visually appealing images. Stable Diffusion offers advanced features like character detail control, parameter personalization, and more customization options, making it preferable for those seeking greater flexibility and control. From an academic standpoint, the stylistic differences in character images produced by Midjourney and Stable Diffusion provide opportunities for exploring various artistic styles and techniques in AI-generated character illustration, contributing to the advancement of AI-generated content.

Overall, considering individual needs and preferences is crucial when choosing between Midjourney and Stable Diffusion, enhancing the AI drawing experience by selecting the software that aligns best with desired outcomes.

References

- [1] Xinyi Shan, Jeanhun Chung, “A Study on the Development Direction of Traditional Cultural Contents in the Age of Convergence Media”, *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 22, No. 3, pp. 99-104, 2022. DOI: <https://doi.org/10.7236/IJIBC.2022.22.3.99>
- [2] Pingjian Jie, Xinyi Shan, “<Game Scorn> Concept Design Research Using AI Generation Program [Midjourney]”, *Journal of Convergence Contents*, Vol.5 No.1, pp.44-47, Spring 2023.
- [3] Soyoun Park, Jin-hee Kim, Jongsu Kwak, Hyun-jing Jung, Aeyoung Kim, “Analysis of AI·SW education tool: Analysis of Changes in Generated Images According to Changes in text for Prompt of Midjourney”, *The Journal of Korean Association of Computer Education*, Vol. 27, No. 1, pp. 119-122, 2023.
- [4] Soo-Hwan Lee, Ki-Sang Song, “Exploring the possibility of using ChatGPT and Stable Diffusion as a tool to recommend picture materials for teaching and learning”, *Journal of The Korea Society of Computer and Information*, Vol.28, No.4, pp.209-216, 2023.
- [5] Xiaodi Cui, Xinyi Shan, “A Color Study of the Sky Area Focused on the Van Gogh's Paintings”, *The International Journal of Internet, Broadcasting and Communication*, Vol.15, No.1, pp. 113-119, 2023. DOI: <http://dx.doi.org/10.7236/IJIBC.2023.15.1.113>
- [6] Yi Qi, Xinyi Shan, Jeanhun Chung, “Study on the Influence of the Language Symbols on Side-Scrolling Puzzle Games -- Focused on the Players' Empathy”, *The International Journal of Internet, Broadcasting and Communication*, vol.15, No.1, pp.120-125, 2023. DOI: <http://dx.doi.org/10.7236/IJIBC.2023.15.1.120>
- [7] Xinyi Shan, Jeanhun Chung, “Comparison of the Characteristics of Green Screen and LED Wall in Virtual Production System”, *The International Journal of Advanced Smart Convergence*, Vol.11, No.2, pp.64-70, 2022. DOI: <https://doi.org/10.7236/IJASC.2022.11.2.64>