

A Study on AI Softwear [Stable Diffusion] ControlNet plug-in Usabilities

Chenghao Wang, Jeanhun Chung*

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

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

*how9951@gmail.com, * evengates@gmail.com*

Abstract

With significant advancements in the field of artificial intelligence, many novel algorithms and technologies have emerged. Currently, AI painting can generate high-quality images based on textual descriptions. However, it is often challenging to control details when generating images, even with complex textual inputs. Therefore, there is a need to implement additional control mechanisms beyond textual descriptions. Based on ControlNet, this passage describes a combined utilization of various local controls (such as edge maps and depth maps) and global control within a single model. It provides a comprehensive exposition of the fundamental concepts of ControlNet, elucidating its theoretical foundation and relevant technological features. Furthermore, combining methods and applications, understanding the technical characteristics involves analyzing distinct advantages and image differences. This further explores insights into the development of image generation patterns.

Keywords: *Stable Diffusion, AI Painting, ControlNet, Open pose, Canny Edge, Depth*

1. Introduction

1.1 Research Background

In recent years, AIGC (Artificial Intelligence Generated Content) has made significant progress and breakthroughs in the field of image generation. Especially mainstream AI painting software such as Stable Diffusion Web UI and Midjourney can generate high-quality images with just a brief descriptive prompt input from the user. However, the instability in controllability during the image generation process can result in the inability to achieve the desired outcomes. Currently, to enhance the controllability of AI-generated images, the application of Control Net in Stable Diffusion Web UI has greatly changed this phenomenon.

ControlNet is a new neural network concept proposed by the authors, which involves using additional inputs to control pretrained large models. While ControlNet has improved the control over images, exploring how to

Manuscript Received: September. 19, 2023 / Revised: September. 27, 2023 / Accepted: September. 30, 2023

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

Tel: +82-2-2260-3, Fax: +82-2-2260-3766

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

further enhance creativity and diversity while maintaining the generation of high-quality images is still worth investigating. This study, based on Stable Diffusion ControlNet, aims to generate more novel and unique images by utilizing various technical parameters and methods. The goal is to enhance the creativity and productivity of ControlNet while maintaining image generation quality, making it a truly creative AI painting tool through practical implementation.

1.2 Research Objectives

Today, utilizing AI for generating artistic paintings has become a novel and effective approach. Text-to-image generation is an emerging field aimed at producing realistic images from textual descriptions. By utilizing machine learning algorithms and neural networks, users can generate complex and artistic pieces of artwork, showcasing a uniqueness and originality that traditional art creation cannot match.

Despite the significant success of AI in painting, it is often challenging to convey detailed control when generating paintings from text descriptions, even with complex prompts. Therefore, this paper introduces ControlNet to help users gain a more detailed understanding of how to apply and generate more stable AI images. For example, Canny, Open Pose, depth information, and so on. And using the AI painting open-source software Stable Diffusion Web UI as a foundation, we analyze the various application advantages and differences in the generated images.

2. Theoretical Background

2.1 AI Painting

With the advancement of AI painting, it has found widespread applications in various domains. AI painting technology can not only generate realistic images, videos, and sound but also simulate the styles of different art movements, such as Impressionism, Surrealism, and more, by learning from a vast collection of artworks created by artists. The application of Stable Diffusion Web UI and Midjourney in commercial projects has significantly boosted production efficiency and driven progress in various industries, greatly enhancing the development of related sectors through AI painting. At the same time, ordinary individuals can easily create images that align with their imagination, enriching the styles and value of painting. This not only elevates AI painting as a new form of artistic expression but also fulfills market demands. As the scope of AI painting applications continues to broaden, it is increasingly being employed in fields such as concept design, game design, and animation production.

AI painting has made remarkable progress but still faces challenges, such as ensuring the originality of works and protecting artists' copyrights. However, with the development of society and the improvement of laws, it is believed that these issues will be gradually addressed, and AI painting will continue to bring more surprises and innovations in the future.

2.2 ControlNet

ControlNet is a plugin based on Stable Diffusion Web UI and has now become an indispensable integral part of AI painting. Its emergence signifies that AI image generation has entered a truly controllable era, which is also a pivotal step in making generated images practical for production.

As it evolves, ControlNet becomes applicable to various industries, including concept design and game design, enabling the generation of artistic images with straightforward parameter adjustments. It offers boundless creative inspiration, thereby enhancing production efficiency. Its core idea is to add extra conditions

beyond the textual description to control diffusion models, thus better controlling information such as character posture and scene structure in generating images. These additional conditions are input in the form of images, and the model can perform tasks like Canny edge detection, depth sensing, and human pose recognition based on this input image. It can also directly fill line drawings with color, generate images with the same underlying structure, and more.

Table 1. ControlNet Classification

Feature	ControlNet Models	Function
Edge	Control_v11p_sd15_canny	Edge detection, generated from line drawings
	Control_v11p_sd15_mlsl	Straight line detection, applicable to architectural design.
	Control_v11p_sd15_softedge	Soft edge detection, preserving more edge details.
	Control_v11p_sd15_scribble	Graffiti generation.
Depth	Control_v11p_sd15s2_lineart_anime	Extraction of animated line drawings.
	Control_v11f1p_sd152_depth	Depth detection, generating images based on depth information.
	Control_v11p_sd15_normal	Normal mapping, extracting normal information.
Picture	Control_v11p_sd15_seg	Different colors represent different object types.
	Control_v11p_sd15_openpose	Character pose detection, generating images based on poses.
Image	Control_v11p_sd15_inpaint	Image extension and restoration.
	Control_v11e_sd15_shuffle	Style transfer fusion.
	Control_v11f1e_sd15_tile	Detail enhancement.

3. ControlNet Image Generation Analysis

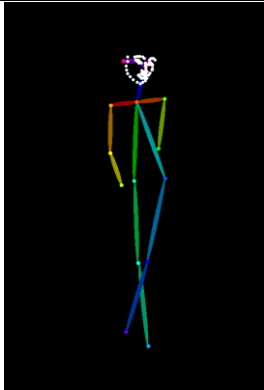

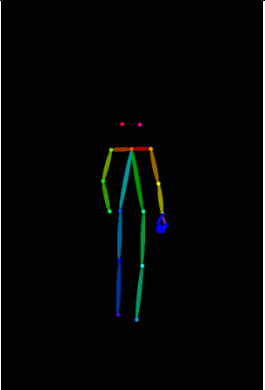

3.1 Open Pose

Open Pose can detect body parts in an image and label key points, including the body, hands, face, and feet, among others. This is used to generate images depicting the poses and actions of characters in the image, facilitating the control of generating specific character movements. This is one of the core features within ControlNet. Users only need to input a pose image (a real person's photo or a skeletal action diagram) as a reference for AI painting. Then, after entering the prompt, AI can generate an image with the same pose based on this reference image. This method effectively addresses the issue of uncontrollable character movements when generating AI paintings.

Therefore, in projects, the images generated by the Open Pose functionality serve various purposes, from creating eye-catching advertisements to generating unique and artistically inspired works. Furthermore, in

game and animation production, the application of Open Pose can generate artistic characters, enhancing visual effects and reducing production costs and time. Through testing and usage, it can be observed that Open Pose exhibits characteristics such as efficiency, versatility, and scalability.

Table 2. Open Pose

Input (Open Pose)	User Prompt	Input (Open Pose)	User Prompt
			



3.2 Canny Edge



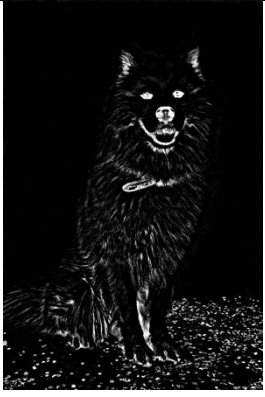

The Canny feature in the Stable Diffusion Web UI ControlNet is an image edge detection algorithm that can accomplish edge detection and generate line drawings. Its effects can be customized according to one's own requirements. For example, if you wish to preserve more edge details, you can opt for a more aggressive edge detection approach than soft edge detection. If it's for architectural design, you can directly utilize straight line detection. Moreover, it also enables further creative image generation based on these selections. The following advantages have been derived through practical experience:

Enhanced Image Quality: ControlNet Canny Edge enables users to generate clearer and more detailed images. This is achieved by enhancing the edges of the image. Clear edges can make the image appear sharper and more realistic.

More Accurate Outlining Control: ControlNet enables Stable Diffusion to adhere more accurately to user-defined outlining preferences, thus generating the desired appearance. This is particularly valuable for creating images in comic and illustrative styles.

In summary, ControlNet Canny Edge is an effective method for enhancing Stable Diffusion image generation quality, control, and efficiency. The Canny algorithm combines smoothing, detection, and other techniques to efficiently identify and detect edges within an image. It is a crucial tool for generating outline effects in AI painting models. It can output soft and coherent edge contours, enhancing the texture and detail of the image.

Table 3. Can Edge

Input (Canny Edge)	User Prompt	Input (Canny Edge)	User Prompt
			







3.3 Depth

By extracting depth information from the image, with white being closer and black being farther away, the depth map is extracted from the user-provided reference image. Based on this, the structural hierarchy of the scene is recreated. This also implies the ability to directly construct a basic scene using 3D modeling software and then feed it to AI painting for generating intricate and artistic compositions. Depth information calculation is used to control spatial distances and generate a depth map. It is commonly employed in scenes with significant depth, effectively portraying the relationship of distance between foreground and background elements.

Therefore, in the early stages of game design, the creation of vast world scenarios is complex, leading to lengthy production cycles and high costs. With the application of Depths, it becomes a tool as well as a productivity enhancer, allowing for the rapid and straightforward generation of scenes with well-defined spatial hierarchies.

Table 4. Depth

Input (Depth)	User Prompt	Input (Depth)	User Prompt
			



4. Conclusion

With the rapid advancement of artificial intelligence, significant achievements have been made in the field of image generation, progressing from generating high-quality images based on simple textual prompts to the current application of ControlNet plugins that enable control over image stability. This study is based on the generation approach of Stable Diffusion Web UI ControlNet and explores various methods' applications in AI-powered art creation, elucidating how to judiciously select appropriate approaches for different domains. Through research, it has been discovered that OpenPose can detect human figures in images and estimate key points to control dynamic poses of characters, making it widely applicable in character design. The Canny Edge feature is an image edge detection algorithm that generates outlines, allowing for the incorporation of additional details during image generation to further enhance image quality. Moreover, compared to methods solely reliant on Stable Diffusion, Canny Edge significantly reduces the required number of iterations for image generation, thereby enhancing efficiency. Lastly, the Depth technique involves extracting depth information from images to generate images, controlling spatial distances. It is commonly employed in scenes that require a strong sense of spatial perspective.

Overall, AI painting technology is rapidly advancing and expanding its application domains, enriching creative methods. However, it also faces challenges related to regulation and copyright issues. The long-term impacts of AI painting technology require ongoing observation.

References

- [1] Chenghao Wang, Jeanhun Chung. "Research on AI Painting Generation Technology Based on the [Stable Diffusion]" *The International Journal of Advanced Smart Convergence* 12.2 pp.90-95 (2023) : 90.
DOI:<http://dx.doi.org/10.7236/IJASC.2023.12.2.903>
- [2] Zhang, Lvmin, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." *arXiv preprint arXiv:2302.05543* (2023).
DOI:<https://doi.org/10.48550/arXiv.2302.05543>
- [3] Jin, Ze, and Zorina Song. "Generating coherent comic with rich story using ChatGPT and Stable Diffusion." *arXiv preprint arXiv:2305.11067* (2023).
DOI:<https://doi.org/10.48550/arXiv.2305.11067>
- [4] Zhao, Shihao, et al. "Uni-ControlNet: All-in-One Control to Text-to-Image Diffusion Models." *arXiv preprint arXiv:2305.16322* (2023).
DOI:<https://doi.org/10.48550/arXiv.2305.16322>
- [5] Qin, Can, et al. "UniControl: A Unified Diffusion Model for Controllable Visual Generation In the Wild." *arXiv preprint arXiv:2305.11147* (2023).
DOI:<https://doi.org/10.48550/arXiv.2305.11147>
- [6] Engadget, ControlNet the magic of controlled diffusion model.
<https://juejin.cn/post/7234554420163493944>
- [7] Hickman Design, AI Generated Art: A Revolutionary Creative Force With A Double Edge
<https://hickmandesign.co.uk/blog/news/ai-generated-art/>
- [8] 51CTO, ControlNet star count exceeds 10,000! In 2023, will AI painting go crazy?
<https://www.51cto.com/article/748002.html>