IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Using Qualifiers In Prompts For Stable Diffusion

An Experimental Study

[1]Dr. Saba Hilal, [2]Ms. Ishbah Hilal, [3]Ms. Yusma Hilal

[1]Professor (CSE), [2]Finance Manager, [3]BTech Student

[1]Department of Engineering and Technology,

[1]Echelon Institute of Technology, Faridabad, HR, India

Abstract: Stable Diffusion uses a text-to-image machine learning model to generate images matching the prompt specification. This paper presents the experimental study of using different qualifiers with four different types of text prompts as a method of creating variations in the generated images while preserving the semantics. It was found that all qualifiers helped in creating some variation in the images. However, depending on the type of variation created in the images, this paper presents an insight into enhancement or insertion of relevant or irrelevant features in the generated images based on one- or two-word qualifiers used.

***Index Terms* - Stable Diffusion, text-to-image, Generative AI, neural networks**

I. Introduction

The Generative AI is creating wonders by successfully generating the intelligent text, audio and images. Stable Diffusion is a text to image model created in 2022 that can produce photograph level quality images based on the text specifications. The text given at the prompt gives the semantics to the model. Although, the type of image generated can be controlled by different factors, however the text specification at the prompt is important for the type of image that is generated.

II. Text Specification for Stable Diffusion

Stable Diffusion 2.1 is an important online text-to-image model from Stability AI [3], that can be used to check out the type of images generated by providing the text and other image control specifications.

Figure 1: Images generated using the text “room” in the prompt

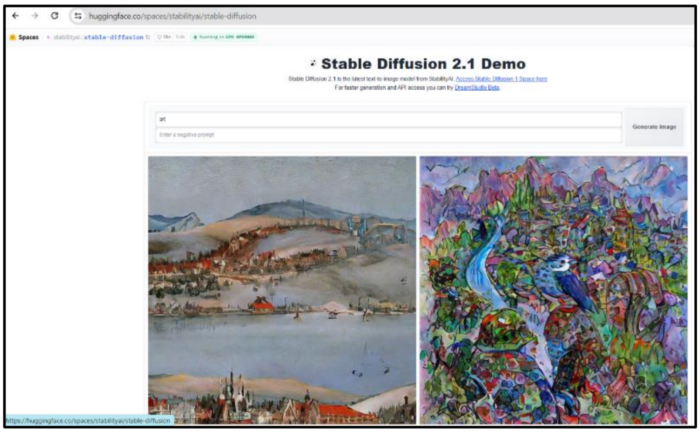


Figure 2: Images generated using the text “art” in the prompt

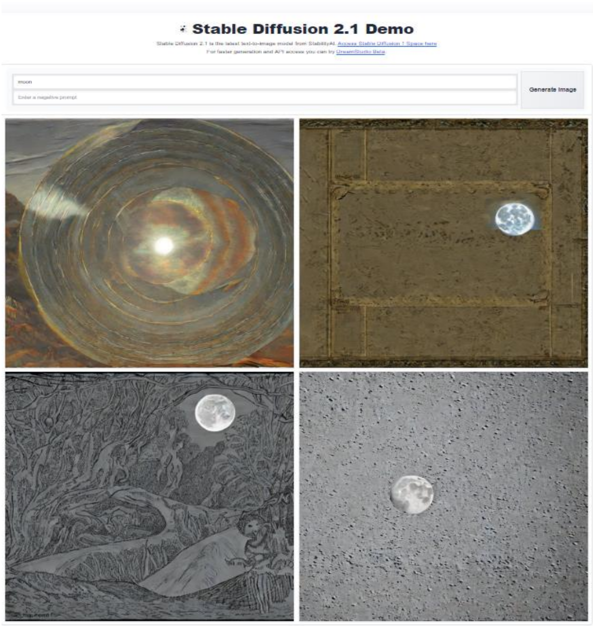


Figure 3: Images generated using the text “moon” in the prompt

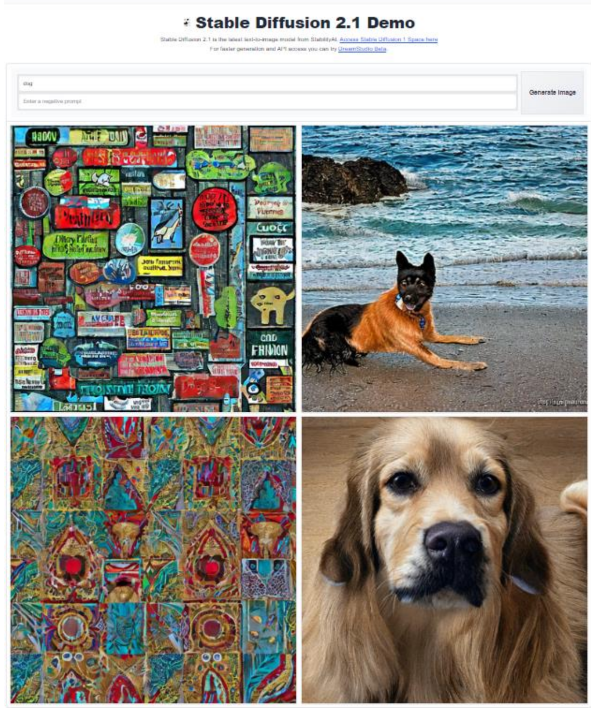


Figure 4: Images generated using the text “dog” in the prompt

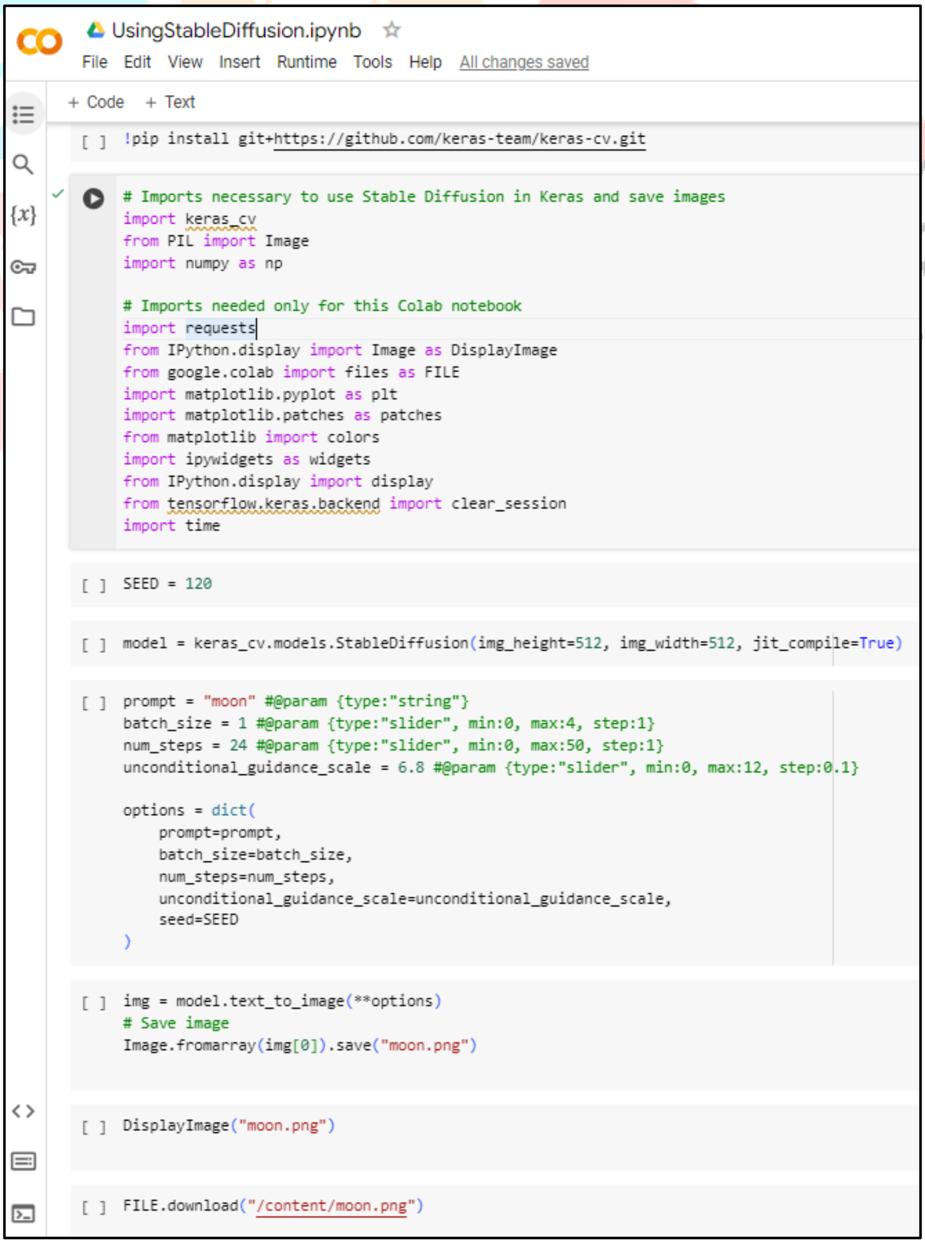
It was found that on keeping the other factors same and just changing the text specification, different types of results are produced. The types of results produced can be labelled as realistic, mixed and artistic.

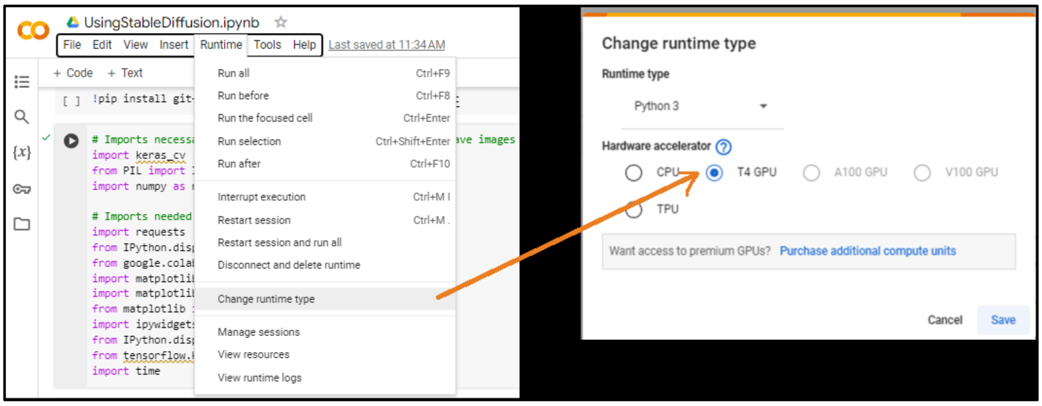
Text Used	Type of Image Generated	Remarks
room	Mixed images	It was found that simple use of "room" at the prompt produced 2 out of 4 images where room was seen in the image while the other two images that were generated appeared as artworks not related to the specified text.
art	Artistic images	The model produced beautiful artworks.
moon	Artistic images	The model produced images that can be called artworks and not the realistic moon images.
dog	Mixed images	It was found that simple use of "dog" at the prompt produced 2 out of 4 images where dog was seen in the image while the other two images that were generated appeared as artworks not related to the specified text.

Table 1: The Different types of Text Prompt Specifications with Type of Images Generated

III. CODE USED

The same type of text specifications for image generation can be experimented with if the following Python code is used in Google Colab. To use this TPU setting is also needed which is given below.





Images Generated by using the code given above. The prompt used was “dog”.

Figure 5: Images generated using the text “dog” in the prompt by using the code given above

IV. Text Specification with Different Qualifiers for Stable Diffusion

The same text terms were used with different qualifiers. It was found that using qualifiers with the text makes the model introduce better semantics into the generated image, increases the realistic types and reduces the generation of non-related images or artworks.

V. Result Variations on Using Qualifiers with Prompts

The results obtained are listed here for the different prompts:

Figure 6: Prompt - room (With and without Qualifiers)

Figure 7: Prompt - art (With and without Qualifiers)

Figure 8: Prompt - moon (With and without Qualifiers)

Figure 9: Prompt - dog (With Qualifiers)

VI. ABOUT USING TWO-WORD QUALIFIERS WITH PROMPTS

The use of two words qualifiers with text was seen creating mixed results. For example, the term "Bull headed dog" created the dogs with horns or as the bull. The term "cat friendly dog" produced the dog images with cat type hair.

Figure 10: Prompt - dog (With Two Word Qualifier - Bull-headed)

Figure 11: Prompt - dog (With Two Word Qualifier – Cat Friendly)

VII. Some Qualifiers That Introduced incorrect features in images

The examples show here are specific to the term “dog”. Some of the qualifiers such as “naughty dog” and “spotted dog” added irrelevant features in the generated images. For example, in spotted dog (2), springs were put into the image. Also, in spotted dog (4) and naughty dog (3), the images contain two tails in the dog images.

spotted dog (1) spotted dog (2) spotted dog (3) spotted dog (4)

Figure 12: Prompt - dog (With One Word Qualifier – Spotted)

naughty dog (1) naughty dog (2) naughty dog (3) naughty dog (4)

Figure 13: Prompt - dog (With One Word Qualifier – Naughty)

IV. RELATED STUDIES

Authors in [1], [2], [3], [4], [5] and [6] present different ways to provide the control in the image generation using the Stable Diffusion Model. Even in the Python code given above, the type of image generation can be controlled by providing different values for seed, num_steps, or unconditional_guidance_scale. In the current study, it was seen that although the changes in these options led to creation of variations in images, however, the use of one-word qualifiers increased the meaningful features and semantic clarity for the generated images, in comparison to two-word qualifiers.

IV. CONCLUSION

In this paper, the authors present their experimental results on testing the use of one- or two-words Qualifiers with Stable Diffusion prompts for the different types of text specifications. It was found that the meaningful one-word qualifiers enhance the meaningful features in the generated images while two-word qualifiers added some irrelevant or mixed features in the generated images.

References

[1]Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B., 2022, High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684–10695). https://openaccess.thecvf.com/content/CVPR2022/papers/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.pdf

[2]Lvmin Zhang, Anyi Rao, and Maneesh Agrawala 2023. Adding conditional control to text-to-image diffusion models. https://arxiv.org/pdf/2302.05543.pdf

[3]Hugging Face, 2024. Stable Diffusion 2.1 Demo, 2024, https://huggingface.co/spaces/stabilityai/stable-diffusion, retrieved on 02-01-2024.

[4]Mohammed, 2023. Stable Diffusion and Control-Net -A Beginners Guide, https://medium.com/@techlatest.net/stable-diffusion-and-control-net-a-beginners-guide-9efefe2790f, retrieved on 03/01/2024

[5]Hugging Face. 2024. Controlled generation. https://huggingface.co/docs/diffusers/using-diffusers/controlling_generation, retrieved on 03/01/2024.
[6]Stanislav Frolov, 2021. Adversarial text-to-image synthesis: A review, Neural Networks, Volume 144, December 2021, Pages 187-209