Submitted to NicoInt 2024 (accepted as a poster, updated on 2024-6-29)

Emergent Text-to-Image Generation Using Short Neologism Prompts and Negative Prompts

Yasusi Kanada
Dasyn.com
Tokyo, Japan
yasusi@kanadas.com

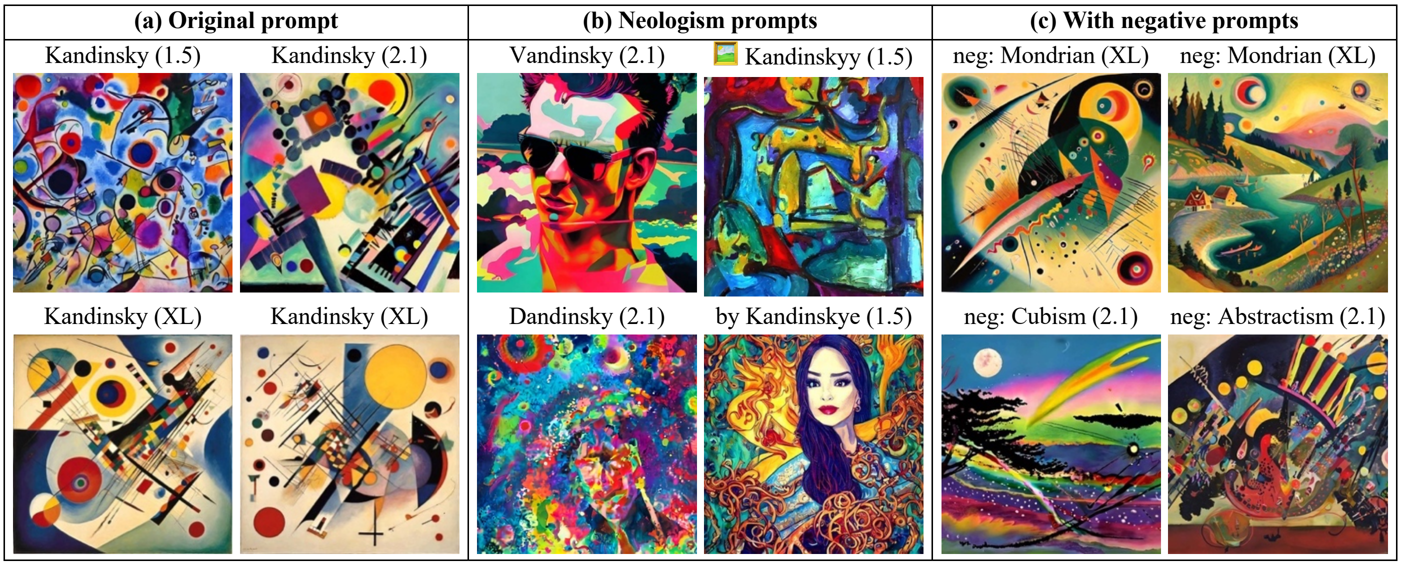


Fig. 1. Summarized results: More divergent images are generated by using neologism prompts (coined or modified prompts) (b), and specific negative prompts (c), than by the original prompt (a) (Stable Diffusion 1.5, 2.1, and XL were used).

Abstract — Recent advances in text-to-image generation models such as Stable Diffusion and DALL·E have made it possible to generate a wide variety of images from text prompts. While AI artists often use long prompts to generate desired images, this paper proposes a method for generating diverse painterly images using Stable Diffusion and short prompts of only 1-2 words. This method first employs neologisms (or coined words) to diversify the generated images. Feeding Stable Diffusion with prompts containing randomly spelled words or neologisms generated by Markov chains resulted in a variety of unpredictable images, such as birds and distorted people. Images were also generated by varying the spelling of existing words, such as artist names. This resulted in a diverse range of images, from those similar to the images generated with the original spelling to those that were significantly different. This method secondly employs negative prompts to drastically transform the images. We found that this not only diversifies the images but also significantly changes the image style, such as when "Mondrian" is specified. We found that specifying the same word as the positive and negative prompt is often effective.

Keywords — *AI art; Text-to-image generation; Stable Diffusion; Neologism prompt; Coined-word prompt; Random prompt; Markov-chain; Negative prompt; Prompt engineering.*

I. Introduction

Based on diffusion models [2], such as Stable Diffusion [6], DALL·E [1, 5], and Midjourney, image-generating AI (i.e., text-to-image generation models) can generate various types of images from text called prompts. So-called AI artists often use long prompts of around several tens of words to create preferred images with image-generating AI. This is probably because they want image-generating AI to produce pictures resembling those drawn by humans and struggle to make the obtained images closer to them. While the method of organizing such long prompts is called Prompt engineering [3], many AI artists rather determine prompts through 'brute-force trial and error' [3]. It is not particularly creative for AI artists to hack prompts and see numerous similar images in the process of obtaining desired images.

The purpose of this research is to develop a method to encounter a diverse range of unknown images generated by image-generating AI without relying on the above image creation methods, namely, without taking time to describe prompts or view similar images, and to discover interesting images that meet the objectives from among them.

To encounter unknown interesting images, it is preferable to create an environment where AI can draw images emergently, in other words, "freely", from short prompts, allowing diverse images to be generated and producing numerous images. Short prompts are preferred because using long prompts, similar to those used by many AI artists, guides the generated images relatively detailed, making it difficult to discover unknown elements. Additionally, since generating only interesting images is difficult, it is necessary to automatically generate large number of images that humans need to review and to ensure that diverse images are generated by using random numbers. If relatively interesting images are included in those images at a relatively high probability, it is more efficient to obtain desired images than to provide detailed instructions.

There are two more reasons for aiming to enable AI to draw creatively. First, it is hypothesized that this will allow the information contained in the large-scale image data that AI has learned to be utilized more effectively in the newly generated images, leading to new discoveries. Second, it is hypothesized that there must be a way to create interesting images that humans cannot draw or find difficult to draw by transforming and combining the information in these image data by AI, in other words, to create works that are unique to AI art.

As a first step toward making AI draw emergently, this paper proposes a method to generate a variety of images simply by using Stable Diffusion and a short prompt, typically one or two words. Figure 1 shows the overview, and the caption describes it in detail. The method involves using neologisms and negative prompts to induce AI to generate a large number of images in diverse styles, from which we can select the most interesting ones. Section II presents experiments and results using neologisms, Section III presents experiments using negative prompts, and Section IV concludes the paper.

This method has two features. First, neologisms are used to generate unknown and diverse images. Stable Diffusion can accept neologisms as prompts. While unpredictable unknown images are generated from random neologisms, varying the spellings of existing words such as artist names can generate a wide range of images, from images similar to those given the original spelling to images significantly different from them.

Second, negative prompts [4] are used to drastically transform the images. In Stable Diffusion, negative prompts are used to prevent specific types of objects from appearing in images [4]. However, negative prompts affect the entire image rather than just removing parts of it, and in some cases, they can significantly change the style of the image. Therefore, using various negative prompts can diversify the generated images.

In the experiments, mainly painterly images, especially abstract images, were generated. The reasons for this are as follows: Many images generated by image-generating AI so far have pursued realism. Many AI art creators have been creating realistic images like photographs or anime images, and models, such as Midjourney, have been adjusted to produce more realistic images of human bodies accordingly. However, in fact, image-generating AI lacks knowledge about the structure of real-world objects such as the human body, so sometimes, improbable images, such as hands with six fingers, are generated. Therefore, these models are not necessarily suitable for generating realistic images. Hence, the focus here is on abstract paintings. Additionally, since the number of images that can be included in the paper is limited, other more than 40,000 image examples are published on https://dasyn.com/aiart/.

II. Image Variation by Neologisms

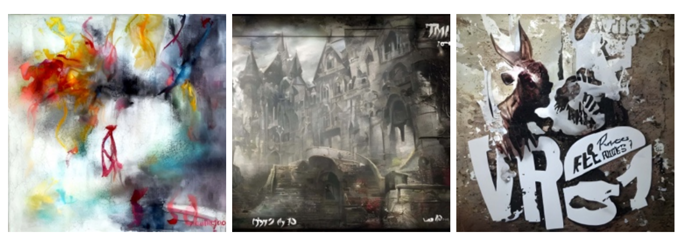
This section presents the results of image generation experiments using neologisms in the proposed image creation method. Neologisms should not be effective in models that operate on word-level units such as ChatGPT, but they can be effectively used in image generation models such as Stable Diffusion. This experiment revealed that a diverse range of images, unpredicted from the used neologisms, were generated. While numerous neologisms were tested in the experiment, only a subset is shown in this paper. Other images can be found at https://dasyn.com/aiart/.

In this experiment, images of size 512 to 1024 were generated using Stable Diffusion on the OpenArt.ai, which is an image generation website, and then enlarged by four times by super-resolution to before being downloaded. Although various additional trained models are available on this site, it is presumed that the model used in this experiment did not undergo any additional training. As the generated images often had significantly low contrast, automatic adjustments were first made to the contrast and brightness, followed by manual adjustments. The obtained images are resized and presented here again.

A. Unpredictable image generation by random spellings

This subsection describes examples of random spellings of neologisms and neologisms generated by Markov chains, as shown in Figure 2, along with examples of images generated from them as prompts.

(a) apqacnwxmeepq gleuspk (1.5)   (b) l-kciuovaflnnytwcnirxactdou ounsjbhyzkdrndwzpgmodvakwv yjwaubh-td-oisbyzrvutqcflsvl-h gzhsxtpsa-otrc-uppi-x-k-zdopd-ouoejjinlxrfez (1.5)   (c) ustamjqimdlcdk osbeev (1.5)



(d) inias-orys-tansol agi (1.5)   (e) re-4-mend-rs-142-anghed-ea n-pous-fonalgly-try-ithanigod-anat-tept-200-s-inin-isond-36-e-at-te-f-uran (1.5)   (f) ralfoos-rwes-ey-t w-e (1.5)

Fig. 2. Example images generated by using random- and Markov-chain-based coined words (a-c: https://dasyn.com/aiart/random_photo-e.html, d-f: https://dasyn.com/aiart/markov_photo-e.html).

Figures 2a-c are examples of images generated by the random neologisms as captioned. These words are generated using randomly selected characters based on the occurrence probabilities of English letters and hyphens. The "1.5" in the caption indicates that the images were generated using Stable Diffusion 1.5 [7]. Although 36 images were generated in this experiment, only three images are shown here (the URL in the caption of Figure 2 shows all images, as is the case for the following figures).

As meaningful spellings are not included in the prompts, all images are unpredictably generated from the prompts. Of these, nine images (25%) were bird-like photographs. Figure 2a is an example of this, with several images including numerous birds (including one where it is unclear if they are birds). It is speculated that images of birds are generated more frequently from prompts with weak associations to specific training images. Additionally, there are five images (14%) including humans, some of which feature distorted faces. However, publishing grotesque images was avoided. Figure 2b appears to be a collage of multiple photos, but the details are unclear. Figure 2c is speculated to be an image composed of cows, other animals, and humans, which are unique to image-generating AI. However, since most of these images are unclear or grotesque, they may not suit the preferences of most people.

Figures 2d-f are examples generated by neologisms using Markov chains. The spellings are generated based on the occurrence probabilities of these characters following English letters and hyphens. Again, as meaningful spellings are scarce, unpredictably generated images result from the prompts. Figure 2e shows European-style buildings, and there are other images of buildings. Some of the images not shown here depict insects, plants, as well as animals such as tigers and pandas. While some human faces are distorted, many of these images are relatively naturally drawn, clearer than those generated by random neologisms, and may appeal to a wider audience's preferences.

B. *Unpredictable image generation by coined words*

This subsection explains various example images obtained by providing man-made neologisms as prompts, as shown in Figure 3. Similar to the previous subsection, it is completely unknown what kind of images will be obtained from these neologisms, but these images seem to be much more preferable than those in the previous subsection.

Figures 3a-c show images obtained by providing the neologism "Mondalian" to Stable Diffusion 1.5. Although this term is a variation of "Mondrian," which will be discussed in the following subsection, the resulting images are completely unrelated and irrelevant to that discussion. These images depict European-style cityscapes with seas or waterways. It is speculated that cities are depicted because the term ends with "dalian" (a city name in China), which somewhat allows for predictions about the generated images, but the depicted scenery does not resemble Dalian.

Slightly altering the spelling to "Mondalium" results in images resembling plants, as shown in Figures 3d-f. This might be because several botanical names such as Aquifolium and Trifolium end with "lium." Furthermore, Figures 3g-i show images generated by providing prompts with numbers attached to "unfin" to Stable Diffusion 2.1 [8]. Although various types of images can be obtained from "unfin," these three images are animal-like. The unpredictability of obtaining such animal-like images from this spelling is unclear, but these images are particularly intriguing.

(a) Mondalian (1.5) (b) Mondalian (1.5) (c) Mondalian (1.5)

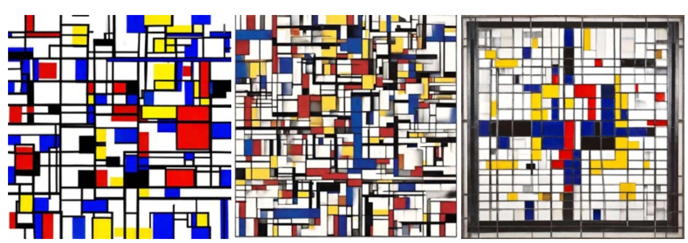
(d) Mondalium (1.5) (e) Mondalium (1.5) (f) Mondalium (1.5)

(g) Unfin3 (2.1) (h) Unfin6 (2.1) (i) Unfin5 (2.1)

Fig. 3. Example images generated by using man-made coined words (a-c: https://dasyn.com/aiart/broken_city-e.html, d-i: https://dasyn.com/aiart/broken_creature-e.html).

C. *Image variation by modification of "Mondrian"*

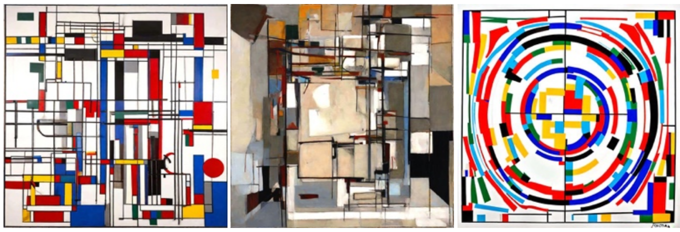
This subsection presents images generated by providing variations of the painter's name "Mondrian" as prompts. To serve as a benchmark, the painter's name "Mondrian" is provided directly to Stable Diffusion XL [9] to generate images, and an example of this is shown in Figure 4. Typically, images resembling Mondrian patterns like those in Figures 4a and 4b are obtained. However, images with frames, as in Figure 4c, or images of rooms decorated with Mondrian patterns were also obtained.



(a) Mondrian (XL) (b) Mondrian (XL) (c) Mondrian (XL)

Fig. 4. Example images generated by using "Mondrian".

Examples of neologisms derived from "Mondrian" are classified into three categories below. As the first example, the neologisms obtained by changing the beginning of "Mondrian" and the images obtained from them are explained, as shown in Figure 5. Changing the beginning of the word can sometimes result in images that appear entirely unrelated to Mondrian patterns. However, in most cases, images looking like variations of Mondrian patterns are obtained.



(a) fondrian (XL) (b) fondrian (XL) (c) fondrian (XL)

(d) iondrian (XL) (e) iondrian (XL) (f) iondrian (XL)

(g) drian (XL) (h) drian (XL) (i) drian (XL)

Fig. 5. Example images generated by coined words that changes the beginning of "Mondrian".

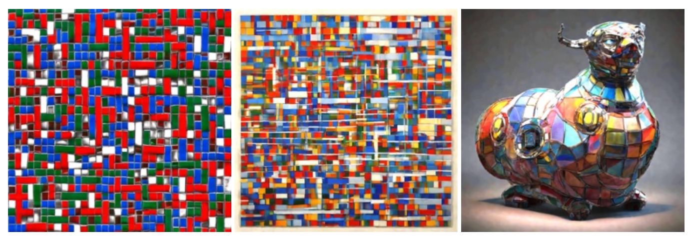
Figures 5a-c show images obtained by replacing the first letter with "F." Figure 5a resembles Mondrian patterns but also includes circles and arcs. Figure 5b shares linear patterns, although the colors are different. Figure 5c features colors arranged concentrically, which can be regarded as a Mondrian pattern transformed from Cartesian coordinates (x, y) to polar coordinates (r, θ) (where $r = x$, $\theta = cy$, c is a constant, and $0 \le cy < 2\pi$).

Similarly, Figures 5d-f show images obtained by replacing the first letter with "I." In this case, images that can be regarded as Mondrian patterns were not obtained. Figure 5d bears some resemblance, but with an oval shape at the center. In Figure 5e, although it is a linear pattern like Mondrian patterns, the interconnected small circles form a different shape. Figure 5f also shows interconnected shapes like circles and squares, but with a background that adds depth to the image. The significant difference between "iondrian" and "fondrian" may be attributed to the prefix "ion," which is probably associated with many training images.

Furthermore, Figures 5g-i show images obtained by "drian," which is the result of deleting the first three letters of "Mondrian." Despite the deletion, images similar to Mondrian patterns, like in Figure 5g, were still obtained. However, the colors are more subdued. Additionally, 9 out of 12 images (75%) were considerably different. In Figure 5h, a relation to Mondrian patterns can be observed, as if Mondrian patterns were applied to buildings with rounded colored areas. Conversely, Figure 5i shows squares for windows, resembling Mondrian patterns, but lacks color and has little resemblance to it. At least 4 out of the remaining 7 images similarly do not show a clear relationship with Mondrian patterns.

As a second example, images obtained by altering the end of "Mondrian," as shown in Figure 6 are explained. In many cases here as well, images resembling variations of Mondrian patterns were obtained.

(a) mondrial (XL) (b) mondrial (XL) (c) mondrial (XL)



(d) mondri (XL) (e) mondri (XL) (f) mondri (XL)

Fig. 6. Example images generated by coined words that changes the end of "Mondrian".

Figures 6a-c show images obtained by replacing the last letter with "L," resulting in "mondrial." Figure 6a is relatively close to Mondrian patterns but includes overlapping circles. In Figure 6b, while dominated by straight lines, they are diagonal, with circles and squares overlapping on the surface. Figure 6c features a human face, with colored squares not only forming the background but also overlaying the face, yet still exhibiting a relation to Mondrian patterns.

Figures 6-d illustrate images generated from the word "mondri" with the last two characters removed. Figures 6d and 6e exhibit linear patterns, arranged in a grid-like fashion, with most squares colored. Figure 6d resembles tiled flooring, while Figure 6f appears to overlay a tile-like pattern on the surface of an animal sculpture. In Figures 6-f, Mondrian patterns seem to have transformed into a tiled pattern.

Removing one more character to form "mondr" resulted in no images related to Mondrian patterns being generated.

As a third example, images obtained by altering the middle of "Mondrian" are explained using Figure 7. In this case, images resembling variations of Mondrian patterns can sometimes be obtained, but often, completely different images are generated. For instance, the word “Mandrian” produced images resembling statues or emblems.

(a) Mondriian (XL) (b) Mondriian (XL) (c) Mondriian (XL)

(d) Mondriean (XL) (e) Mondriean (XL) (f) Mondriean (XL)

Fig. 7. Example images generated by neologisms that changes the middle of “Mondrian”.

Figures 7a-c show images generated from the word “mondriian,” with an extra “i” inserted in the middle. In this case, images resembling variations of Mondrian patterns were obtained, with no images appearing unrelated to Mondrian patterns among the 12 generated. Figure 7a closely resembles Mondrian patterns, while Figure 7b shows a similar pattern attached to a sphere. Figure 7c features a complex shape with more white areas resembling the Mondrian pattern adhered to it.

Figures 7d-f show images generated from the word “mondriean,” formed by changing the “i” to “ie” in the middle. Figure 7d displays a fine Mondrian-style pattern overlaid with circles, while Figure 7e presents a diagonal “Mondrian pattern.” In contrast, Figure 7f shows a woman’s face. However, among the 12 images, this was the only one considered unrelated to Mondrian patterns.

D. Image variation by modification of “Kandinsky”

This subsection presents images generated by varying the painter’s name “Kandinsky” as a prompt. I focus on images that significantly differ from those obtained from the original spelling because, in the case of “Kandinsky,” even when images differ significantly from those generated by the original spelling, the relationship with Kandinsky can often be observed.

Firstly, images obtained by directly inputting the painter’s name “Kandinsky” into Stable Diffusion are mentioned. Figure 1a shows examples of these images, which resemble Kandinsky’s typical abstract paintings.

Then, examples of images obtained by varying the end or beginning of “Kandinsky” are shown in Figure 8. When the spelling is altered, images of people or less abstract objects are often generated instead of highly abstract paintings by Kandinsky.

(a) by Kandinskye (1.5) (b) 🖼 Kandinskye (1.5) (c) 🖼 Kandinskyy (1.5)

(d) Mandinsky (2.1) (e) Vandinsky (2.1) (f) Dandinsky (2.1)

Fig. 8. Example images generated by coined words that changes the end or beginning of “Kandinsky” (a-f: https://dasyn.com/aiart/kandinsky-e.html, a-c: https://dasyn.com/aiart/kandinsky_positive-e.html).

Figures 8a and 8b show images obtained from words with an added “e” at the end. However, to obtain images in the style of paintings, “by” is prefixed in Figure 8a and “🖼” (a Unicode character representing a painting) is prefixed in Figure 8b. While “🖼” is effective in obtaining painting-style images in Stable Diffusion 1.5, it is less effective in other versions of Stable Diffusion. Figure 8c shows an image obtained from a word with a “y” added at the end. Most of Kandinsky’s paintings are abstract. However, the image in Figures 8b and 8c are abstract paintings with different styles. Figure 8a includes figures surrounded by abstract patterns.

Figures 8d-f display images obtained by replacing the initial “K” with “M,” “V,” and “D.” Except for the presence of figures in Figures 8e and 8f, the images are abstract. Although they differ in style from the famous Kandinsky paintings, a similar painting by Kandinsky was found in a Google image search for Figure 8d.

E. Image variation by coined art-movement names

By using art movement names such as “Expressionism,” “Impressionism,” and “Cubism” as prompts, it is possible to obtain images mimicking the styles based on those art movements. Even by providing neologisms derived from these terms, images with a relatively consistent style can be obtained. An example of this is shown in Figure 9. This suggests that it is possible to generate images with styles that differ from existing art movements and to create a variety of images.

“Expressionism” and “Impressionism” both have “pressionism” at the end. By specifying this spelling or variations like “ressionism” or further truncated forms such as “essionism” or augmented versions like “kressionism” as prompts, images with painting styles resembling Expressionism can be obtained. Figure 9a shows an image generated from “essionism,” while Figure 9b is derived from “kressionism.” However, in the case of “kressionism,” photorealistic images can also be generated, so the word “painting” was added to the prompt. Furthermore, Figure 9c shows an image generated by "expretionism," obtained by replacing the letters in the middle of the word "Expressionism," Figures 9d and 9e are images generated by "ussionism" and "iressionism," obtained by changing the beginning of the word. Finally, Figure 9f shows an image generated by further modifying the prompt to "conalism."

(a) Essionism (XL) (b) painting Kressionism (2.1) (c) Expretionism (2.1)

(d) Ussionism (2.1) (e) Iressionism (2.1) (f) Conalism (2.1)

Fig. 9. Example images generated by coined words that represent art movements (https://dasyn.com/aiart/art-style-e.html).

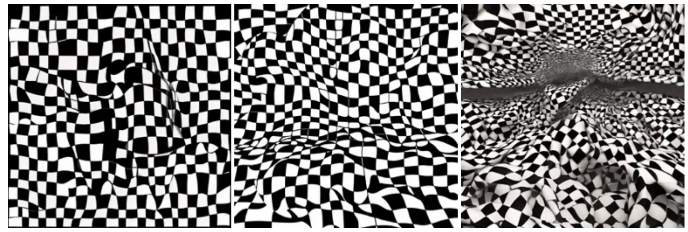
III. Style Variation by Negative Prompts

This section presents the results of experiments using negative prompts in the proposed image creation method. The goal of these experiments is to investigate whether negative prompts can significantly change the image style and diversify the generated images. Negative prompts are typically used to prevent specific types of objects from appearing in images. However, it was found that negative prompts can also be used to change the overall image style in some cases. The methods for image generation, post-processing, and publication are the same as in Section II. A number of negative prompts were tested in the experiment, but only a small subset is shown here. The other images are available at https://dasyn.com/aiart/.

A. *Negative prompt and "self-negation" examples*

This subsection examines images generated using the negative prompts "checker" and "Mondrian." It compares the results of using these prompts only as positive prompts versus using them as both positive and negative prompts, which means the prompt is self-negated.

First, Figure 10 shows examples of images generated by using "checker." Figures 10a-c show images generated using "checker" only as a positive prompt. Twelve images were generated, all of which exhibited wrinkles and twists, as shown in these images. Figures 10d-f show examples of images generated using "checker" as both a positive and negative prompt. Only three images are shown here, but a wide variety of images were generated in terms of patterns, objects, and colors. It is hypothesized that this is because the negative prompt suppresses the main features of "checker" while allowing various peripheral features to remain.



(a) checker (XL) (b) checker (XL) (c) checker (XL)

(d) checker, neg: checker (XL) (e) checker, neg: checker (XL) (f) checker, neg: checker (XL)

Fig. 10. Example images generated by using "checker" as a positive (and negative) prompt.

Next, Figure 11 shows examples of images generated using "Mondrian" as both a positive and negative prompt. Figure 4a shows an image generated using "Mondrian" as a positive prompt only. Here again, a wide variety of images are generated, with diverse patterns, objects, and colors. Figure 11c includes a human face, but some of the 12 generated images have more clearer faces.

Only "checker" and "Mondrian" are used as examples here, but specifying other spellings in combination with positive and negative prompts can often generate a variety of images.

(a) Mondrian, neg: Mondrian (XL) (b) Mondrian, neg: Mondrian (XL) (c) Mondrian, neg: Mondrian (XL)

Fig. 11. Example images generated by using "Mondrian" as a positive and negative prompt.

B. *Image variation by adding negative prompt*

This subsection compares images generated using Stable Diffusion 2.1 with "Klee" as the positive prompt and no negative prompt versus "Mondrian" as the negative prompt, based on Figure 12. Figures 12a-c show examples of images without a negative prompt. They are relatively diverse, but all are Klee-style paintings. Figures 12d-f show examples of images with "checker" as the negative prompt. The difference between these images and those without a negative prompt is not always clear. On the other hand, Figures 12g-i show examples of images with "Mondrian" as the negative prompt. These images are in a completely different style. They are not only unlike Mondrian's paintings, but also unlike Klee's paintings. The reason for this is unknown.

(a) Klee (2.1) (b) Klee (2.1) (c) Klee (2.1)

(d) Klee, neg: checker (2.1) (e) Klee, neg: checker (2.1) (f) Klee, neg: checker (2.1)

(g) Klee, neg: Mondrian (2.1) (h) Klee, neg: Mondrian (2.1) (i) Klee, neg: Mondrian (2.1)

Fig. 12. Example images generated by using "Klee" as a positive prompt and "Mondrian" as a negative prompt in Stable Diffusion 2.1 (g-i: https://dasyn.com/aiart/klee_positive-e.html).

Next, a similar comparison for Stable Diffusion XL were performed, based on Figure 13. Figures 13a-c show examples of images without a negative prompt. They are relatively diverse, but all are Klee-style paintings.

Figures 13d-f show examples of images with "checker" as the negative prompt. These images have trees, moons, or suns, and are less abstract than those without a negative prompt.

Furthermore, Figures 13g-i show examples of images obtained when specifying "Mondrian" as the negative prompt, showcasing a style that differs from Mondrian, Klee, or Figures 12d-f. These images lack clear outlines as seen in Figures 13a, c, d, f. While these images are overall red and lacking color diversity, this "color fading" phenomenon often occurs when generating colorful images with negative prompts specified as positive prompts. It occurs frequently when "Mondrian" is designated as the negative prompt. This phenomenon can sometimes be solved by specifying color names or words that delete colors together with the negative prompt. Typically, terms like "grayscale" are effective, but in this case, specifying "red" resulted in colorful images like Figures 13j-l.

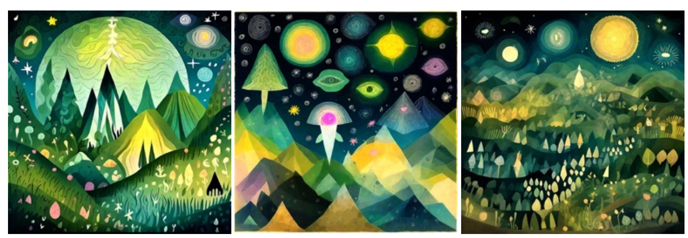
C. *Art-movement names as negative prompts*

This section compares images generated with Stable Diffusion 2.1 using "Cezanne" and "Kandinsky" as positive prompts, with and without negative prompts. The negative prompts used are the names of real and imaginary art movements.

(a) Klee (XL) (b) Klee (XL) (c) Klee (XL)

(d) Klee, neg: checker (XL) (e) Klee, neg: checker (XL) (f) Klee, neg: checker (XL)

(g) Klee, neg: Mondrian (XL) (h) Klee, neg: Mondrian (XL) (i) Klee, neg: Mondrian (XL)



(j) Klee, neg: red_Mondrian (XL) (k) Klee, neg: red_Mondrian (XL) (l) Klee, neg: red_Mondrian (XL)

Fig. 13. Example images generated by using "Klee" as a positive prompt and "checker" as a negative prompt in Stable Diffusion XL.

Figure 14 shows examples of images generated by simply feeding the painter's name "Cezanne" to Stable Diffusion 2.1 for comparison. All 24 images generated were of landscapes, and were all in the same Cezanne style in terms of color scheme. Another example of a positive prompt used is "Kandinsky." However, as mentioned before, examples of images generated without a negative prompt are shown in Figure 1a.

Figure 15 shows examples of images generated with art movement names as negative prompts. First, Figures 15a-c show three of the 32 images generated with "Cezanne" as the positive prompt and "Pressionism," a fictional art movement name, as the negative prompt. Some of the 32 images are similar to those in Figure 14, but the images in Figure 15 are quite different in terms of both color scheme and subject matter.

Fig. 14. Example images generated by using "Cezanne" as a positive prompt.

(a) Cezanne, neg: pressionism (2.1) (b) Cezanne, neg: pressionism (2.1) (c) Cezanne, neg: pressionism (2.1)

(d) Kandinsky, neg: cubism (2.1) (e) Kandinsky, neg: cubism (2.1) (f) Kandinsky, neg: cubism (2.1)

Fig. 15. Example images generated by using art-movement names as a positive prompt.

Second, Figures 15d-f show three images generated with "Kandinsky" as the positive prompt and "cubism" as the negative prompt. These images are significantly different from the images generated without a negative prompt. Figure 15d is figurative but has strong colors. Figures 15e and 15f are dramatic in both style and color, and are abstract but reminiscent of celestial bodies.

This section has only shown results for positive-negative prompt pairs that produce significant changes when a negative prompt is specified. However, there are cases where specifying a negative prompt does not produce a clear change. Therefore, trial and error is necessary to find an appropriate negative prompt from a given positive prompt.

IV. Conclusion

This paper proposes a method for generating diverse images using Stable Diffusion and short prompts of only one or two words. The method involves using neologisms and negative prompts to induce AI to generate a large number of images in diverse styles, from which humans can select the most interesting ones.

This method has two features. First, neologisms are used to generate unknown and diverse images in this method. Prompts with random spellings generated unclear birds or people with distorted faces, while prompts with spellings generated by Markov chains generated buildings and animals. Prompts generated artificially generated images of city photographs and fictional plants and animals. We believe that the generated images become clearer and more attractive in this order. In addition, we found that changing the spelling of existing words, such as the names of painters, can generate a variety of images, from images similar to those obtained with the original spelling to images that are very different.

Second, negative prompts are used to diversify the images in this method. Negative prompts can not only generate diverse images but also drastically change the image style in some cases. In particular, giving the same spelling as the positive and negative prompts often results in a variety of images. In addition, specifying a negative prompt can sometimes cause "color fading," which can sometimes be solved by specifying a color name or a word that erases color. However, when using different prompts for positive and negative, it is not easy to find a negative prompt that will drastically change the image style.

Future work for prompts using neologisms is to find a way to generate more interesting images than the Markov chain used in this paper and to find a way to effectively change existing words. Future work for negative prompts is to find a way to find negative prompts that draw out diversity from a given positive prompt.

References

[1] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, Wesam Manassra, Prafulla Dhariwal, Casey Chu, Yunxin Jiao, and Aditya Ramesh. Improving Image Generation with Better Captions, 2023, https://cdn.openai.com/papers/dall-e-3.pdf

[2] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

[3] Liu, Vivian, and Lydia B Chilton., Design Guidelines for Prompt Engineering Text-to-Image Generative Models. *2022 CHI Conference on Human Factors in Computing Systems* (*CHI'22*). Association for Computing Machinery, New York, NY, Article 384, 23 pages. https://doi.org/10.1145/3491102.3501825, 2022.

[4] Negative prompt, 2022, https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Negative-prompt

[5] Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. "Zero-Shot Text-to-Image Generation." *38th International Conference on Machine Learning* (*Proceedings of Machine Learning Research, Vol. 139)*, ed. Marina Meila, and Tong Zhang, 8821–8831. PMLR.

[6] Rombach, R., et al. High-resolution image synthesis with latent diffusion models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

[7] runwayml, stable-diffusion-v1-5, https://huggingface.co/runwayml/stable-diffusion-v1-5, 2022.

[8] Stability AI, Stable Diffusion 2.1, https://stability.ai/news/stablediffusion2-1-release7-dec-2022, 2022.

[9] Stability AI, generative-models, https://github.com/Stability-AI/generative-models, 2023.