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

Yasusi Kanada

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

Text-to-image generation models such as Stable Diffusion (SD) and DALL-E can produce a wide variety of images from text prompts. While so-called AI artists often use long prompts to generate desired images, this poster proposes a method for generating diverse painterly images emergently by SD with short prompts of 1-2 words. The reason for such attempts is that the author believes that AI, drawing from learned image data, extracts representations of color, shape, and other aspects, as well as the underlying intentions and motivations of the original images' authors, and reflects them in the generated images, and that he also believes that the process of the AI user encountering and selecting the generated images is a creative one.

I. Image Variation by Neologisms

To diversify the generated images, our proposed method utilizes neologisms. When neologisms generated by random spelling or Markov chains were fed as prompts into SD 1.5, diverse and unpredictable images were generated, including birds and distorted figures. Figure 1 a) shows an example of an image generated from random spelling, and b) shows an example of an image generated from Markov chains. The prompts are shown in the captions.

a) apqacnwxmeepqgleuspk   b) ralfoos-rwes-ey-tw-e

Fig. 1. Example images generated by using random- and Markov-chain-based coined words (a: https://bit.ly/3PkjOhe, b: https://bit.ly/3TiCEGI).

a) Mondalian   b) Mondalium   c) Unfin3

Fig. 2. Example images generated by using man-made coined words (a: https://bit.ly/3IFnq9H, b-c: https://bit.ly/3Tst1p0).

Figure 2 illustrates examples of images generated from SD using human-crafted spellings a) and b) for SD1.5, c) for SD2.1. While these images are clearer than those generated from randomized spellings, the random numbers used in the prompts produces a wide variety of images from the same prompt. The AI user (the image producer) encounters a variety of images that cannot be predicted from the given prompts.

Similarly, Figure 3 shows examples of images generated by varying the spellings of existing words, such as artist names.

While some images are similar to the original spelling of "Mondrian," others exhibit significant deviations. The AI user encounters images with changes that cannot be predicted.
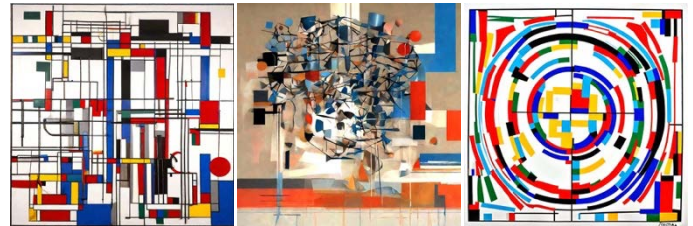
a) fondrian   b) iondrian   c) fdrian

d) mondrial   e) mondrial   f) mondriian

Fig. 3. Example images generated by coined words that changes the beginning, end, or middle of "Mondrian" (https://bit.ly/3TjxGcA).

II. Style Variation by Negative Prompts

Our proposed method utilizes negative prompts to generate diverse images from a single positive prompt. This approach can lead to significant changes in image style, beyond mere diversification. Figure 4 illustrates an example of using "Klee" as the positive prompt for SD XL. Figure 4 a) shows the result of the positive prompt alone, while b) and c) demonstrate the effect of negative prompts: "checker" in b) and "red_Mondrian" in c) ("red" is primarily added for color adjustment). Figure 4 c) is the least predictable. While the impact of negative prompts can be subtle in some cases, as seen in b), specifying the same word as both positive and negative prompts is often effective.

a) Klee   b) Klee, neg: checker   c) Klee, neg: red_Mondrian

Fig. 4. Example images generated by using "Klee" as a positive prompt and "checker" as a negative prompt in Stable Diffusion XL (https://bit.ly/3Vl93z1).

While the use of neologisms and negative prompts can lead to more creative and diverse images, there is yet to be an established method for consistently achieving this outcome. This poster has outlined the findings gathered thus far.