

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

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

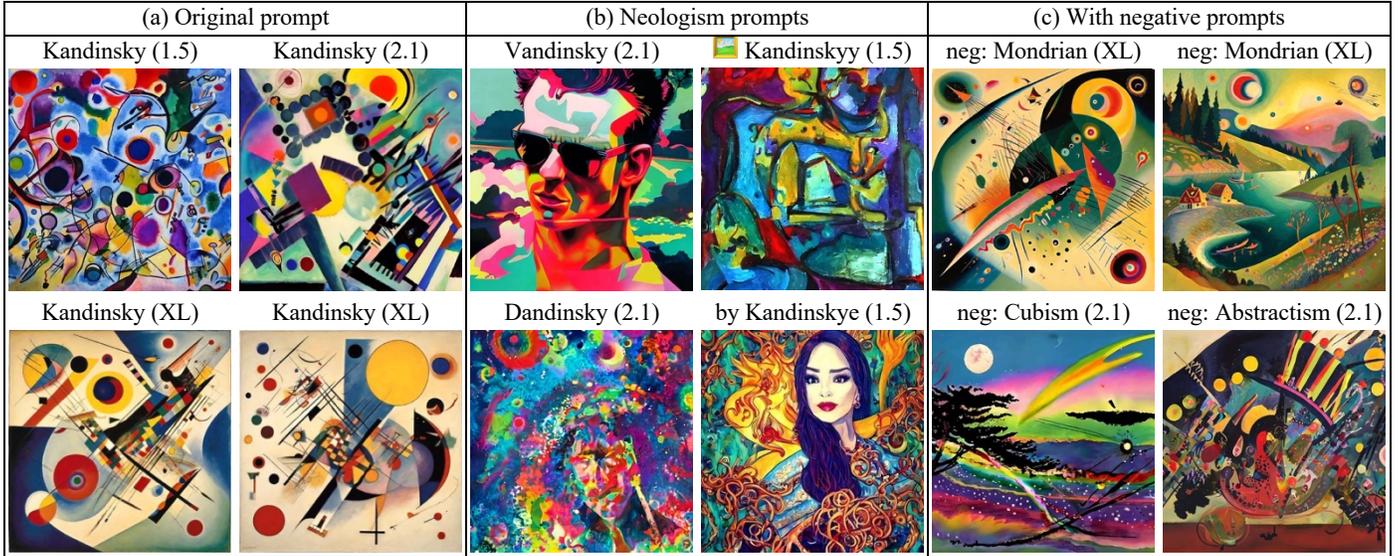


Fig. 1. Summarized results: More divergent images are generated by using neologism prompts (coined or modified prompts) (b), and specific negative prompts (c), than by the original prompt (a) (Stable Diffusion 1.5, 2.1, and XL were used).

Abstract — Stable Diffusion や DALL·E のような text-to-image generation model によって、プロンプトとよばれるテキストから様々な種類の画像を生成することができる。AI アーティストは好みの画像を生成するために長いプロンプトを使用することが多いが、この論文では Stable Diffusion と 1~2 語の短いプロンプトを使用するだけで多様な絵画的な画像を生成する方法を提案する。この方法では第 1 に、生成画像を多様化させるために造語を使用する。ランダムなつづりのプロンプトやマルコフ連鎖によって生成した造語をプロンプトとして Stable Diffusion にあたえたところ、鳥やくずれた人などさまざまな予測困難な画像が生成された。また、画家名などの既存の語のつづりを変化させながら画像を生成したところ、もとのつづりをあたえた場合と似た画像から、それとはおおきくことなる画像まで、多様な画像が生成された。また、この方法では第 2 に、おおきく画像を変容させるために負のプロンプトをあたえる。それによって画像が多様化するだけでなく、“Mondrian” を指定したときなどに画像スタイルがおおきく変化することがわかった。正負のプロンプトとして同一のことばを指定することが多くの場合に効果的である。

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

I. INTRODUCTION

Diffusion models [2] にもとづく Stable Diffusion [6], DALL·E [1, 5], Midjourney などの画像生成 AI (text-to-image generation model) によって、プロンプトとよばれるテキストから様々な種類の画像を生成することができる。いわゆる AI アーティストは画像生成 AI によって好みの画像を生成するために数 10 語程度という長いプロンプトを使用することが多い。そうするのはたぶん、AI アートによって人が描いたような絵を描かせたいとかんがえ、得られる絵を人が描くものにちかづけようとしているからである。このような長いプロンプトを組織する方法は Prompt engineering [3] とよばれているが、AI アーティストの多くはむしろ ‘brute-force trial and error’ [3] によってプロンプトをきめている。望みの画像を得るまでの過程で AI アーティストがプロンプトでハックし、多数の似たような画像を見るのは非常に創造的とはいえない。

この研究の目的は、このような画像生成法によらずに、すなわちプロンプトの記述や似たような画像を見ることで時間を

とられることなしに、画像生成 AI によって生成した多様な画像と出会い、そのなかから目的に合う興味深い画像を発見する方法を開発することである。

未知の興味深い画像に出会うためには、AI に創発的に、いいかえれば「自由に」画像を描かせる、すなわち短いプロンプトから多様な画像が生成されるような環境をととのえて、多数の画像を生成させることがもとめられる。プロンプトが短いほうがよりよく機能するのは、多くの AI アーティストが使用しているような長いプロンプトを使用すると生成画像は比較的細部までガイドされるため、未知のものを発見しにくくなるとかんがえられるからである。また、興味深い画像だけを生成させるのは困難なので、それに出会うためには自動的に多数の画像を生成させて、それを人が見る必要があり、そのためには乱数によって多様な画像が生成されるようにする必要がある。それらの画像のなかに比較的高い確率で興味深い画像がふくまれるようにできれば、こまかい指示をあたえるより効率的に望む画像をえることができる。

AI に創発的に描かせることをめざすことには、つぎのような理由もある。すなわち、そうすることで第 1 に AI が学習した大規模な画像データにふくまれる情報があらたに生成される画像により多く活かされ、そこからあらたな発見ができるとかんがえられる。また、第 2 にそれらの画像データの情報が AI によって変容されくみあわせられて、人間には描けないまたは描くことが困難な興味深い画像を描く、つまり AI アートならではの作品を創造する方法があるはずだとかんがえられる。

この論文では、AI に創発的に描かせるための第 1 歩として、Stable Diffusion と通常は 1~2 語という短いプロンプトから多様な画像を生成する方法を提案する。Figure 1 はその概要を示す図であり、その説明はキャプションに書かれている。この方法では造語や負のプロンプトを使用して多様なスタイルの画像を AI に多数生成させ、そのなかから人が興味深い画像を選択する。II では造語、III では負のプロンプトを使用した実験の内容と結果を示し、IV で結論をのべる。

この方法では第 1 に、未知の画像、多様な画像を生成するために造語を使用する。Stable Diffusion はプロンプトとして造語を使用できる。ランダムな造語からは予測できない未知の画像が生成される一方で、画家名などの既存の語のつづりを変化させていくと、もとのつづりをあたえた場合と似た画像から、それとはおおきくことなる画像まで、多様な画像を生成することができる。

この方法では第 2 に画像のスタイルを様々に変容させるために負のプロンプト [4] を使用する。Stable Diffusion においては特定の種類のオブジェクトなどが画像にあらわれないようにするために負のプロンプトを使用する [4]。しかし、負のプロンプトは画像からその一部をとりぞくのではなく画像全体に作用し、場合によっては画像のスタイルをおおきく変化させることができる。したがって、様々な負のプロンプトを使用することによって生成画像を多様化することができる。

実験においてはおもに絵画的な画像とくに抽象的な画像を生成させた。その理由はつぎのとおりである。これまでに画像生成 AI によってつくられた画像にはリアリティを追求し

たものが多い。AI アート作家の多くは写真やアニメ画像のようなリアルな画像を作っていて、Midjourney などのモデルもそれにあわせてより人体などのリアルな画像がつかれるようにしている。しかし、実は画像生成 AI は人体など実世界のものの構造に関する知識もたないため、しばしば、ありえない画像、たとえば指が 6 本ある手、が生成される。したがって、これらのモデルはリアルな画像の生成にはかならずしも適さないとかんがえられる。そのため、ここでは抽象画を中心とする。なお、論文に掲載できる画像の数はかぎられているので、他の 4 万枚以上の画像例を <https://dasyn.com/aiart/> に掲載している。

II. IMAGE VARIATION BY NEOLOGISMS

この章においては、提案する画像作成法における造語の使用法を、画像生成実験の結果を記述するかたちで示す。造語は ChatGPT のような単語を単位とするモデルでは使用できないはずだが、Stable Diffusion などの画像生成モデルにおいては効果的に使用できる。この実験によって、使用した造語からは予測できない多様な画像が生成されることがわかった。実験においては多数の造語をためしたが、この論文ではその一部だけを示す。他の画像は <https://dasyn.com/aiart/> にある。

この実験では OpenArt.ai という画像生成サイトにおいて、Stable Diffusion を使用して画像を 512~1024 のサイズで生成し、超解像によって 4 倍に拡大したうえでダウンロードした。このサイトにはさまざまな追加学習したモデルも用意されているが、使用したモデルにおいては追加学習をしていないとかんがえられる。生成画像はコントラストがかなり低いことが多いため、コントラストとあかさをまず自動で調整し、さらに手動で調整した。えられた画像をここでは再度縮小して掲載する。

A. Unpredictable image generation by random spellings

この節においては、Figure 2 に示すランダムなつづりの造語およびマルコフ連鎖によって生成したつづりの造語の例とそれをプロンプトとして生成される画像例について説明する。

Figures 2a-c がそれらのキャプションに示したランダムな造語によって生成された画像の例である。これらの語には、英語における英字とハイフンの出現確率にしたがってランダムに選択した文字からなるつづりを使用している。キャプションにある“1.5”は画像を Stable Diffusion 1.5 によって生成したことを示す。この実験では 36 枚の画像を生成したが、ここには 3 枚の画像だけを示す (Figure 2 のキャプションにある URL には全画像を示した。以下の図に関しても同様である)。

プロンプトには意味のあるつづりは含まれていないとかんがえられるから、いずれの画像もプロンプトからは予測困難である。そのうち 9 枚 (25%) が鳥の写真風の画像だった。Figure 2a がその例だが、このように多数の鳥をふくむ画像が他にも 1~2 枚 (うち 1 枚は鳥かどうか不明確) ある。理由はわからないが、特定の学習画像とのむすびつきが弱いプロンプトからは鳥の画像が多く生成されるのだと推定される。また人をふくむ画像は 5 枚 (14%) あり、人を多数ふくむ画像もある。ここにグロテスクな画像を掲載することは避けるが、大半の人の顔はくずれている。Figure 2b は多数の写真を貼り合わせた

コラージュであるように見えるが、詳細は不明である。Figure 2c は牛と他の動物と人間が合成された画像だとかんがえられるが、このような画像は生成 AI 固有のものだとかんがえられる。しかしながら、これらの画像の大半は不明確だったりグロテスクだったりするので、多くの人の嗜好には合わないとかんがえられる。



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 2d-f がマルコフ連鎖によって生成された造語による例である。英語において英字とハイフンの直後にこれらの文字が出現する確率にもとづいて生成したつづりを使用している。この場合も意味のあるつづりはほとんど含まれていないため、プロンプトからは予測困難な画像が生成される。Figure 2e にはヨーロッパ風の建物が描かれているし、他にも建物の絵がある。ここに示していない画像のなかには昆虫、植物、虎やパンダなどの動物が描かれたものがある。人の顔のなかにはくずれているものもあるが、これらの画像のなかには比較的自然的に描かれているものが多く、ランダムな造語による画像より明確であり、それらより多くの人の嗜好に合い興味をひくものとかんがえられる。

B. Unpredictable image generation by coined words

この節においては、Figure 3 に示す人為的な造語の例とそれをプロンプトとしてあたえることによってえられる様々な画像について説明する。前節と同様に造語からどのような画像がえられるかはまったく未知だが、これらの画像は前節の画像よりはるかに魅力的にみえる。

Figures 3a-c は Stable Diffusion 1.5 に “Mondalian” という造語をあたえてえられた画像である。この語は以下の節・章においてとりあげる “Mondrian” を変形したものだが、それとはまったくことなる画像がえられていて、そこでの議論とは無関係である。これらの画像には海や水路があ

るヨーロッパ風の都市の風景が描かれている。都市が描かれるのは語尾が “dalian” (中国の都市名) であるためではないかとかんがえられる (そのため生成画像をある程度、予測できる) が、描かれた風景は Dalian とは似ていない。

つづりをわずかに変えて “Mondalium” とすると、Figures 3d-f のような植物らしき画像が描かれる。その理由は、植物の学名である Aquifolium, Trifolium などが “lium” という語尾をもっているからではないかとかんがえられる。また Figures 3g-i には “unfin” に数字がついたプロンプトを Stable Diffusion 2.1 にあたえて生成された画像である。“Unfin” からは他の種類の画像もえられるが、これらの 3 つの画像は動物的である。このつづりからこのような動物的な画像がえられることは予測不能であり、その理由は不明だが、これらの画像はとくに興味深い。



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”

この節においては、“Mondrian” という画家名を変化させたつづりをプロンプトとしてあたえてえられる画像を示す。比較基準とするため “Mondrian” という画家名をそのまま Stable Diffusion XL にあたえて画像を生成したが、その例を Figure 4 に示す。通常は Figures 4a, 4b のようなモンドリアン・パターン風の画像がえられるが、Figure 4c のように枠がついた画像やモンドリアン・パターンで装飾された部屋の画像もえられた。

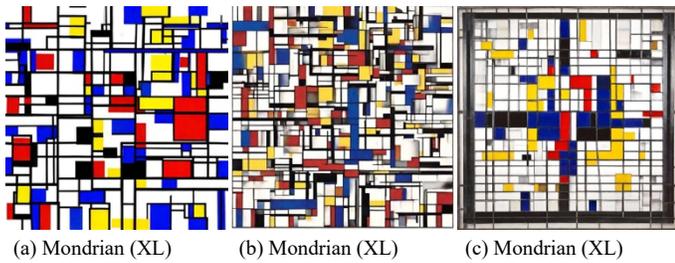


Fig. 4. Example images generated by using "Mondrian".

以下, "Mondrian" を変形した造語の例を 3 つに分類してしめす. 第 1 の例として, Figure 5 に示す "Mondrian" の語頭を変化させた造語とそれによってえられる画像について説明する. 語頭を変化させると, モンドリアン・パターンとはまったく異なる, それとは無関係に見える画像がえられる場合もあるが, 多くの場合にそれを変化させたような画像がえられる.

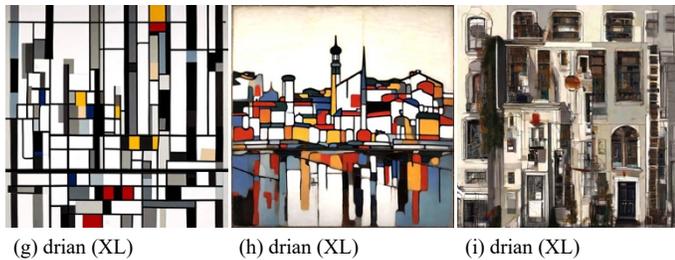
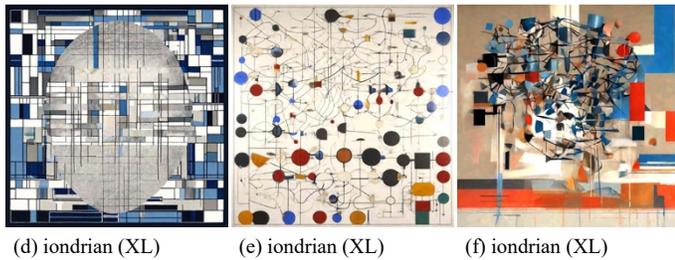
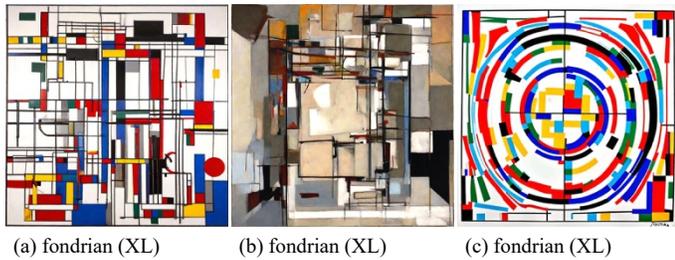


Fig. 5. Example images generated by coined words that changes the beginning of "Mondrian".

Figures 5a-c は先頭 1 文字を "F" に置換して得られた画像である. Figure 5a はモンドリアン・パターンにちかいが, 円や弧もふくまれる. Figure 5b も直線的なパターンだという点でそれにちかいが, とくに着色のしかたがちがう. Figure 5c は色が同心円状に配置されているが, モンドリアン・パターンを直交座標 (x, y) から極座標 (r, θ) ($r=x, \theta=cy, c$ は定数, $0 \leq cy < 2\pi$) に変換したもののみなせる.

同様に, Figures 5-f は先頭 1 文字を "I" に置換して得られた画像である. この場合はモンドリアン・パターンそのものとかんがえられる画像はえられなかった. Figure 5d は比較的モ

ンドリアン・パターンにちかいが, 中心に卵形がある点が多くなる. Figure 5e においては, モンドリアン・パターンと同様に平面的に描かれているが, 小さな円が線につながった形状はそれとは異なる. モンドリアン・パターンにおけるような直線や四角形はわずかし描かれていない. Figure 5f においても円や四角形などが線でむすばれているが, 背景があり立体感がある点は異なる. "fondrian" とくらべて "iondrian" のほうが "Mondrian" からの相違がおおきいのは, 語頭に "ion" というつづりがあり, それが多くの学習画像とむすびつけられているためだと推定される.

また, Figures 5g-i は "Mondrian" の先頭 3 文字を削除した "drian" というつづりから得られた画像である. 3 文字削除してもなお Figure 5g のようなモンドリアン・パターンにちかい画像がえられるが, それより色は地味である. また, そこから遠い画像が 12 枚中 9 枚 (75%) あった. それらの画像のうち Figure 5h にはモンドリアン・パターンとの関係が認められる. すなわち, Figure 5h はモンドリアン・パターンを建物にはりつけたものと解釈できるが, 着色された四角い領域には丸みがある. 一方, Figure 5i には窓が四角いので四角形があらわれる点でがモンドリアン・パターンと共通しているが, 色はほとんどなく, それとの関係はほぼ認められない. のこりの 7 枚の画像のうちのすくなくとも 4 枚は同様にモンドリアン・パターンとの関係がみとめられない.

第 2 の例として, Figure 6 に示す "Mondrian" の語尾を変化させた造語によってえられる画像について説明する. この場合にも, 多くの場合にモンドリアン・パターンを変化させたような画像がえられる.

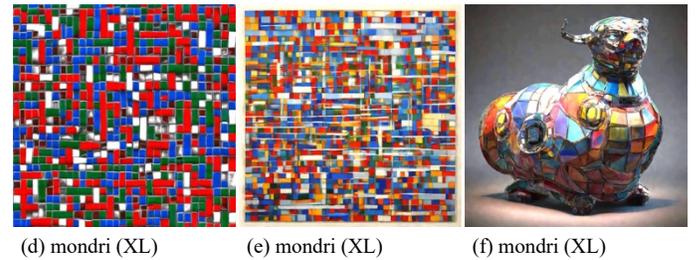
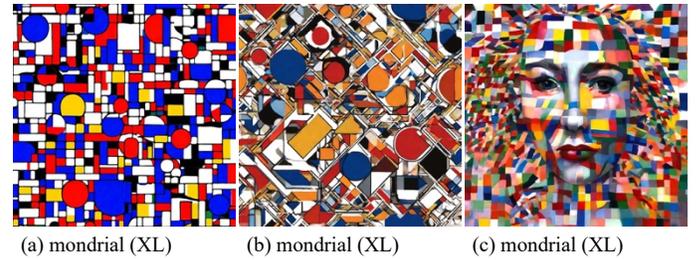


Fig. 6. Example images generated by coined words that changes the end of "Mondrian".

Figures 6a-c は末尾 1 文字を "L" に置換した "mondrial" から得られた画像である. Figure 6a は比較的モンドリアン・パターンに近いが, それに円がかさねて描かれている. Figure 6b においても直線が支配的だが, 直線はななめになり, 円や四角形が表面にかさなっている. Figure 6c には人の顔が描かれている. 色のついた四角形が背景だけでなく顔の

うえにも描かれているが、ここにはモンドリアン・パターンとの関係が認められる。

Figures 6-f は末尾 2 文字を削除した “mondri” から生成された画像である。Figures 6d, 6e は直線的なパターンだが、マス状になっていて、ほとんどのマスは着色されている。Figure 6d はタイルが敷きつめられているようにも見える。Figure 6f は動物の彫刻の表面にタイルのようなものを貼っているように見える。Figures 6-f ではモンドリアン・パターンがタイル模様に変化したのだとかがえられる。

さらに 1 文字を削除して “mondri” にすると、もはやモンドリアン・パターンと関係するような画像は生成されなくなった。

第 3 の例として、Figure 7 に示す “Mondrian” の語中を変化させた造語によってえられる画像について説明する。この場合もモンドリアン・パターンを変化させたような画像がえられることがあるが、それとはまったくことなる画像がえられることが多い。たとえば “Mandrian” という造語からは銅像やエンブレムのような画像が生成された。

Figures 7a-c は語中の “i” をかきかえて “Mondriian” とした造語からえられた画像を示す。この場合はモンドリアン・パターンを変化させたような画像がえられた。生成した 12 枚のなかにモンドリアン・パターンと無関係にみえるものはなかった。Figure 7a はモンドリアン・パターンに比較的ちかい画像である。Figure 7b は同様のパターンが球にはりつけられている。Figure 7c はよりモンドリアン・パターンにちかい白色部分が多いパターンが複雑な形状の立体に貼りつけられている。

Figures 7d-f は語中の “i” を “ie” に変化させて “Mondreian” とした造語からえられた画像を示す。Figure 7d はこまかいモンドリアン風パターンのうえに円をかきかえた画像であり、Figure 7e は斜めの “モンドリアン・パターン” だとかがえられる。それに対して Figure 7f には女の顔が描かれている。ただし、モンドリアン・パターンと無関係とかがえられる画像は 12 枚中この 1 枚だけだった。

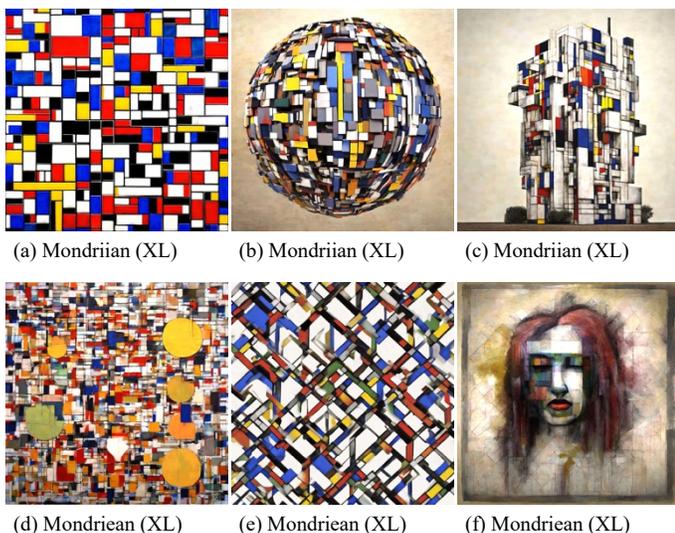


Fig. 7. Example images generated by coined words that changes the middle of “Mondrian”.

D. Image variation by modification of “Kandinsky”

この節においては、“Kandinsky” という画家名を変化させたものをプロンプトとしてあたえることによってえられる画像を示す。もとのつづりからえられる画像とおおきことなるものを中心にとりあげるが、その理由は “Kandinsky” の場合はもとのつづりによる画像とおおきことなる画像が生成されても、それと Kandinsky との関係が把握できる場合が多いからである。

まず、“Kandinsky” という画家名をそのまま Stable Diffusion にあたえてえられる画像についてのべる。Figure 1a に例示したそれらの画像は Kandinsky の典型的な抽象画に似ている。

そして、“Kandinsky” の末尾か語頭を変化させた造語によってえられる画像の例を Figure 8 に示す。つづりを変化させると、Kandinsky が描いた抽象度の高い絵ではなく、人や具体的なものなどの画像がえられる場合が多い。

Figures 8a, 8b は語の最後に “e” をつけた造語からえられた画像である。ただし、絵画風の画像をえるために Figure 8a においては “by”, Figure 8b においては “” を先頭につけている。“” は絵画を意味するユニコード文字であり、Stable Diffusion 1.5 においては絵画風の画像をえるために有効にはたらくが、Stable Diffusion の他の版においてはあまり有効でない。Figure 8c は語の最後に “y” をつけた造語からえられた画像である。Kandinsky が描いた絵の大半は抽象画だが、Figure 8b, 8c はそれらとはスタイルがちがう抽象画である。Figure 8a には人物がふくまれているが、その周囲は抽象的な模様でうめられている。

Figures 8d-f は先頭の文字 “K” を “M”, “V”, “D” に置換したつづりからえられた画像である。Figures 8e, 8f に人物がふくまれていることをのぞけば抽象的な画像である。有名な Kandinsky の絵とはスタイルがことなっているが、Google の画像検索で似た絵をさがすと Figure 8d については Kandinsky の絵が 1 枚みつかった。



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).

E. Image variation by coined art-movement names

“Expressionism”, “Impressionism”, “Cubism”などの芸術運動名をプロンプトとすることによって、それらの芸術運動にもとづくスタイルを模した絵画風の画像をえることができるが、これらの語をもとにした造語をあたえてもある程度一定したスタイルの画像がえられる。その例を Figure 9 に示す。これにより、既存の芸術運動とはことなるスタイルの画像を生成し、多様な画像を生成することができるとかんがえられる。

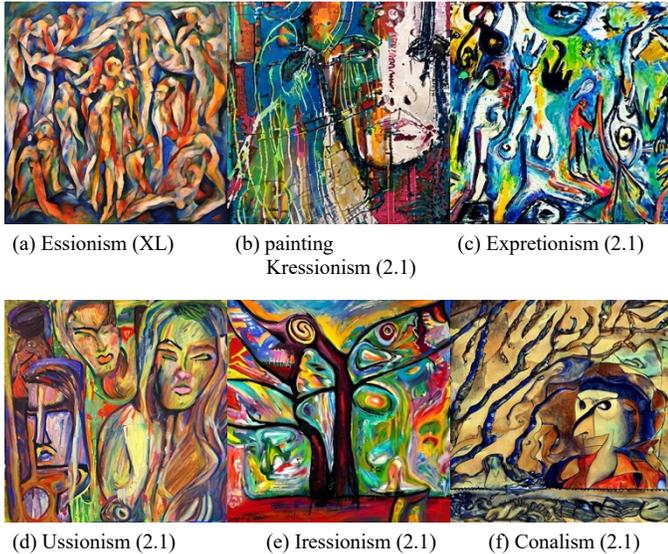


Fig. 9. Example images generated by coined words that represent art movements (<https://dasyn.com/aiart/art-style-e.html>).

“Expressionism”と“Impressionism”にはともに末尾に“pressionism”があるが、このつづりやそれをさらに削った“ressionism”や“essionism”,あるいはそれに文字を足した“kressionism”などをプロンプトとして指定すると、Expressionism にちかいスタイルの絵画風の画像がえられる。Figure 9a は“essionism”, Figure 9b は“kressionism”から生成された画像である。ただし, “kressionism”の場合には写真風の画像がえられる場合もあるため, “painting”をプロンプトにくわえている。また, Figure 9c は“Expressionism”の語中の文字を置換してえられた“expretrionism”, Figure 9d, 9e は語頭を変化させた“ussionism”, “iressionism”から得られた画像, Figure 9f はプロンプトをさらに変化させた“conalism”から得られた画像である。

III. STYLE VARIATION BY NEGATIVE PROMPTS

この章においては, 提案する画像作成法における負のプロンプトの使用法を, さまざまな負のプロンプトによって画像スタイルがおおきく変化し生成画像を多様化させることができるかどうかの実験結果を記述するかたちで示す。通常, 負のプロンプトは特定の種類のオブジェクトなどが画像に出現しないようにするために使用される。しかし, 負のプロンプトは画像全体に作用するので, 場合によっては画像スタイルをおおきく変化させられることがわかった。画像の生成, 後処理, 掲載の方法は II とおなじである。実験においては多数の負のプロ

ンプトをためしたが, ここではその一部だけを示す。他の画像は <https://dasyn.com/aiart/> に掲載している。

A. Negative prompt and “self-negation” examples

ここでは負のプロンプトの例として“checker”と“Mondrian”をとりあげて, それらを正のプロンプトだけに指定した場合と, それを正負のプロンプトにあわせて指定した場合, いわばプロンプトを自己否定した場合にえられる画像についてのべる。

まず“checker”からえられる画像の例を Figures 10 にしめす。Figure 10a-c は正のプロンプトとしてだけ指定してえられた画像である。12 枚の画像を生成したが, そのすべてに, これらの画像におけるように, しわやねじれがみられた。また, Figures 10d-f は正および負のプロンプトとして“checker”を指定したときにえられる画像の例である。ここには 3 枚だけを掲載するが, 模様, 出現するオブジェクト, 色彩, いずれにおいても非常に多様な画像が生成される。これは, 負のプロンプトに指定されることによって“checker”の主要な特徴が否定され, 多様な周辺的な特徴がのこるためではないかとかんがえられる。

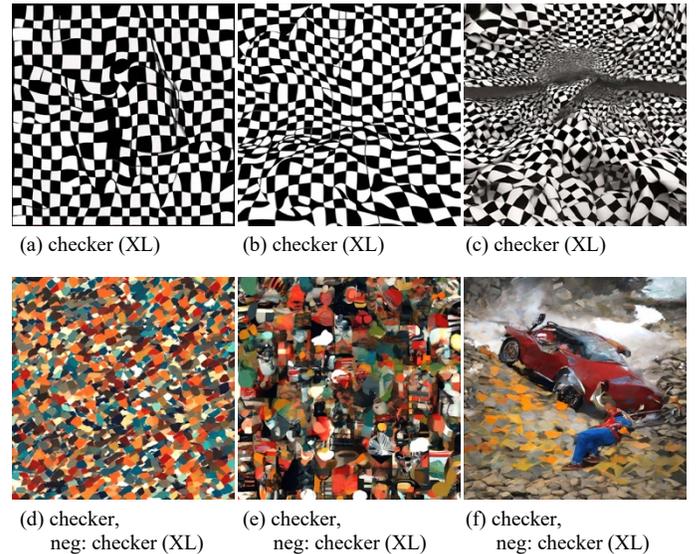


Fig. 10. Example images generated by using “checker” as a positive (and negative) prompt.

つぎに, “Mondrian”を正負のプロンプトとして指定してえられる画像の例を Figure 11 に示す。正のプロンプトとしてだけ“Mondrian”を指定してえられる画像は Figure 4a にしめした。ここでも非常に多様な画像がえられている。模様, 出現するオブジェクト, 色彩, いずれにおいても非常に多様である。Figure 11c には人の顔がふくまれているが, 生成した 12 枚の画像のなかには, もっとはっきりした顔がふくまれているものもある。

ここでは例として“checker”と“Mondrian”だけをとりあげたが, 他のつづりに関しても, 正負のプロンプトにあわせて指定することによって多様な画像がえられることが多い。

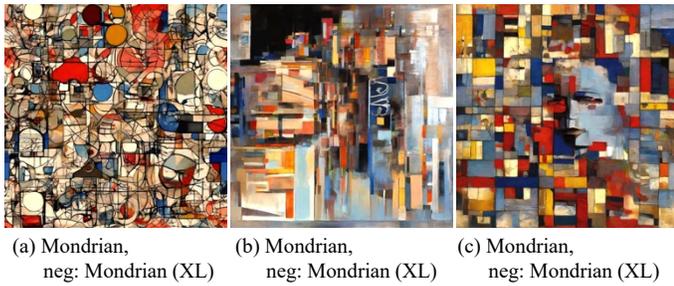


Fig. 11. Example images generated by using “Mondrian” as a positive and negative prompt.

B. Image variation without positive prompt

この節では, Stable Diffusion 2.1 において, 正のプロンプトとして “Klee” を使用し, 負のプロンプトを指定しない場合と負のプロンプトとして “Mondrian” を指定した場合の画像を Figure 12 にもとづいて比較する.

Figures 12a-c は負のプロンプトを指定しない場合の画像例である. 比較的多様性があるが, いずれも Klee 風の絵である.



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).

Figures 12d-f は負のプロンプトとして “checker” を指定した場合の画像例であるが, この場合は負のプロンプトを指定しない場合との差はかならずしも明確でない. それに対して

Figures 12g-i は負のプロンプトとして “Mondrian” を指定した場合の画像例であるが, まったくことなるスタイルの画像が出現している. これらは Mondrian の絵に似ていないだけでなく, Klee の絵にも似ていない. なぜこのような画像があらわれるのか, 理由はわからない.

つぎに, Figure 13 にもとづいて, 同様の比較を Stable Diffusion XL に関しておこなう. Figures 13a-c は負のプロンプトを指定しない場合の画像例である. 比較的多様性があるが, いずれも Klee 風の絵である.

Figures 13d-f は負のプロンプトとして “checker” を指定した場合の画像例であるが, 木々や月か太陽のようなものや描かれていて, 負のプロンプトを指定しない場合より絵の抽象度が低下している.



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

また, Figures 13g-i は負のプロンプトとして “Mondrian” を指定した場合の画像例であるが, Mondrian, Klee, Figures

12d-f のいずれともことなるスタイルの画像が出現している。Figures 13a, c, d, f にあるような、はっきりした輪郭線は描かれていない。これらの画像は全体が赤くて色に多様性がないが、このような「色落ち」現象は負のプロンプトを正のプロンプトとして描かせたときにカラフルな画像を生成する場合に発生しやすい。とくに負のプロンプトに“Mondrian”を指定する場合にはしばしば発生することがわかっている。このような現象は色名や色を消すことばを負のプロンプトにあわせて指定することで解消できる場合がある。色を消す“grayscale”などが通常は効果的だが、この場合は“red”を指定した結果 Figures 13j-l のようなカラフルな画像がえられた。

C. Art-movement names as negative prompts

この節では、Stable Diffusion 2.1 において、正のプロンプトとして“Cezanne”および“Kandinsky”を使用し、負のプロンプトを指定しない場合と負のプロンプトとして実在または仮想の芸術運動名を指定した場合の画像を比較する。

まず、比較のために“Cezanne”という画家名をそのまま Stable Diffusion 2.1 にあててえられる画像の例を Figure 14 に示す。えられた 24 枚の画像はすべて田園風景であり、色調もほぼおなじ Cezanne 風の絵だった。正のプロンプトのもうひとつの例として使用する“Kandinsky”も使用するが、負のプロンプトを指定しない場合の画像は、すでにのべたように Figure 1a に例示した。



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

Figure 15 は芸術運動名を負のプロンプトとして指定してえられた画像の例である。まず Figures 15a-c には正のプロンプトに“Cezanne”，負のプロンプトに仮想的な芸術運動名である“Pressionism”を指定してえられた 32 枚の画像のうち 3 枚を示す。32 枚のなかには Figure 14 の画像にちかい画像もあるが、Figure 15 に示した画像は色調も描かれたものもそれらとはかなりちがう。

Figures 15d-f には正のプロンプトに“Kandinsky”，負のプロンプトに“cubism”を指定してえられた 3 枚の画像を示す。これらは負のプロンプトを指定しない場合とはおおきくことなる。Figure 15d は具象的だが色彩は強烈である。Figures 15e, 15f はスタイルも色彩も劇的であり、抽象的だが天体をおもわせる。

ここでは負のプロンプトを指定することによっておおきな変化がえられる正負のプロンプトの組にかぎって結果を示してきたが、負のプロンプトを指定しても変化が明確でないこともあり、あてられた正のプロンプトから適切な負のプロンプトを見つけるには試行錯誤が必要である。



(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.

IV. CONCLUSION

この論文では Stable Diffusion と 1~2 語という短いプロンプトを使用するだけで、多様な画像を生成する方法を提案した。この方法では第 1 に未知あるいは多様な画像を生成するために造語を使用するが、ランダムなつづりのプロンプトからは不明確な鳥やくずれた顔の人などが生成され、マルコフ連鎖によって生成したつづりのプロンプトからは建物や動物などが生成された。また人為的に生成したプロンプトからは都市の写真風画像や架空の動植物の画像などが生成された。生成画像はこの順により明確かつ魅力的になっているとかがえられる。また、画家名などの既存の語のつづりを変化させると、もとのつづりからえられるのと似た画像から、それとはおおきくことなる画像まで、さまざまな画像を生成することができた。

また、この方法では第 2 に画像を多様化するために負のプロンプトをあてえるが、それにより多様な画像が生成できるだけでなく、画像スタイルをおおきく変容させられる場合があることがわかった。とくに、正負のプロンプトとしておなじつづりをあてえることによって多様な画像がえられることが多いことがわかった。また、負のプロンプトの指定により「色落ち」が発生する場合があるが、色名や色を消すことばの指定で解決できる場合があることがわかった。しかし、正負にことなるプロンプトを使用するときには、画像スタイルをおおきく変化させる負のプロンプトを見つけるのは容易でない。

今後、造語によるプロンプトに関しては、この論文で使用したマルコフ連鎖より興味深い画像がえられる方法をみつけることと、既存の語を効果的に変化させる方法をみつけることが課題である。また、負のプロンプトに関しては、あてられた正のプロンプトから多様性をひきだす負のプロンプトをみつける方法をみつけることが課題である。

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>
- [8] Stability AI, Stable Diffusion 2.1, <https://stability.ai/news/stablediffusion2-1-release7-dec-2022>
- [9] Stability AI, generative-models, <https://github.com/Stability-AI/generative-models>