

StableDiffusion を用いた文学作品の挿絵生成における プロンプトの改良手順

松本龍翔 岩井将行†

東京電機大学未来科学部情報メディア学科†

1. はじめに

昨今話題となっている画像生成 AI で今や絵を描いたことのない人でも「簡単」に、絵を生成できるようになった。そこで昔の挿絵の無い文学作品において、この技術を用いて挿絵の生成をすることで昔の文学を、若者を中心に知ってもらう機会になるであろう。同時に、ファンとして自身の想像を絵に起こして楽しむこともできるであろう。しかし、「簡単」とはいえ、「プロンプト」に対する理解は必須となるだろう。「プロンプト」とは、AI に対する指示文である。この「プロンプト」の理解度が、絵の完成度に繋がるだろう。「プロンプト」は知識であるため、テンプレートを真似することで誰でも簡単に絵の生成を行えると考える。

そのため本稿では、「江戸川乱歩」の青空文庫文学作品を対象に挿絵生成を行い、時代の雰囲気を出し、現代でも受け入れやすい絵にするための「プロンプト」の改良手順を提案する。

2. 関連研究

画像生成 AI について既存の研究を 2 つ紹介する。

1 つ目は、StableDiffusion を用いたアートを美術教育と結びつける研究である [1]。よく使用されているプロンプトについて、それらを元に美術的視点からプロンプトの特徴について言及している。2 つ目は、AI アートを建築に応用できるかの研究である [2]。この研究では、StableDiffusion だけでなく Midjourney, DALL-E2 を用いて、構想、スケッチ、モデリング等のタスクにどのように適用できるかを調査している。これらは「美術」「建築」の視点で研究しているのに対して、本稿では、「文学」の視点からプロンプトについて整理を行う。

2. 提案手法

使用する環境は、NMKD StableDiffusion GUI [3]

Improved prompting procedure for generating illustrations of literary works using Stable Diffusion.

† Ryuto Matsumoto, Masayuki Iwai, Department of Information and Media Engineering, Tokyo Denki University

で、モデルは Hugging Face の提供している「hakurei/waifu-diffusion-v1-3」 [4] である。

事前準備として、図 1 のように、作品を選び、その中から挿絵とする内容の文章を抜き出す。その文章を区切り、英語に変換する。そして、変換したものをプロンプトとして生成を行う。

本章で提案するのは、図 2 のように事前準備で得た画像を元に、画像から画像を生成する image2image を行う。プロンプトで細部の指定を行い、時代の雰囲気を出し、現代的な絵に改良する手法である。一度で目的の絵を生成できない可能性もあるため、その場合は、修正したい点をプロンプトに反映して再度生成を行う。

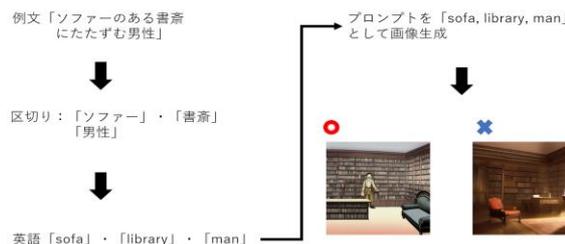


図 1 事前準備の流れ



図 2 提案手法

4. 改良手順

今回使用したプロンプトは 2 種類ある。1 つ目は、雰囲気や印象を変えるプロンプト。2 つ目は、絵の品質を変えるプロンプト。この 2 つの組み合わせで生成していく。図 3 のように、2 種類のそれぞれを最低一つ選択して、プロンプトを決める。表現次第で任意の数入力しても良い。このテンプレートに沿うことで、おおよそ目的に沿った絵の生成を行えるであろう。

ここでは分かりやすく生成する画像を 2 種類に分けて、それぞれにおいて、この手法における変化の例を挙げていく。

表 1. プロンプトの種類と具体例

種類\	プロンプトの具体例
雰囲気・印象	old fashioned, Japanese style ~, retro, showa era ~, vintage, etc.
品質	masterpiece, best quality, beautiful detailed ~, etc.

テンプレート

プロンプト「雰囲気・印象に関するプロンプト（任意の数）、品質に関するプロンプト（任意の数）」

例「masterpiece, best quality, old fashioned, Japanese style city, showa era」

図 3 プロンプトのテンプレート

4.1 背景のみの場合

作品「悪童」 「道の両側に壊れかかった高い石垣が聳え、その上に森のような樹木が空を覆っていたり……」



区切り「道、壊れかかった高い石垣、樹木、空を覆う」



英語「road, broken high stone wall, trees and shrubs, cover up the sky」



(a) 改良前

(b) 改良後

図 4 改良結果の例

ここで主に用いたプロンプトは以下である。

「masterpiece, old fashioned, japanese style, showa era」

改良により、品質の向上はハッキリと見て取れる。時代の雰囲気は、微々たるものではあるが変化は見取れる。元の描画内容に影響される可能性はあると考えられる。

4.2 キャラクターのみの場合

作品「接吻」 「お花はちゃんと長火鉢の前に座っている。」



区切り「女性（お花）、座っている」



英語「woman, sitting down」



(a) 改良前

(b) 改良後

図 5 改良結果の例

ここで主に用いたプロンプトは以下である。

「masterpiece, best quality, old fashioned, japanese style clothes, retro, beautiful detailed eyes」

改良により、他の例と比べてかなり変化している。これは、NMKD StableDiffusion GUI にあるパラメーター設定に「Int Image Strength」があり、この数値によって元の絵の影響をどれほど受けるのかを指定することができる。微調整を行いたい場合は高い値に。大きく調整をした場合には低い値にすると良い。

5. 考察

現状では、雰囲気を表現できてはいるが挿絵とするには惹かれるものが無いと感じた。時代の雰囲気、服装の変化など改良できている部分はあるものの、現代的なものとは程遠いであろう。今回できなかった事として考えられるのは、1. 脈絡の反映、2. 提案したプロンプトの種類以外の内容を反映である。1 では、文章の一部を切り取り、それを元にしたため文章の前後から読み取れる情報を無視した状態となってしまった。2 では、雰囲気・印象、品質にのみフォーカスを当てた生成を行なったが、他に、陰影やライティング、環境、アングルなど様々な要素を指定できるため、これらを使用することでさらに良い改良ができたと考えられる。

6. まとめ

本論文では、「江戸川乱歩」の青空文庫文学作品を対象に、挿絵の生成を AI で行い、時代感を出しつつ現代でも受け入れやすい絵を生成するためのプロンプトの改良手順を提案した。時代の雰囲気を出すことは可能であったが、現代的なものにはならず、作品を知ってもらわなければならないと感じた。

参考文献

- [1] Nassim Dehouche, Kullathida Dehouche: What is in a Text-to-Image Prompt: The Potential of Stable Diffusion in Visual Arts Education, <https://arxiv.org/abs/2301.01902>
- [2] Joern Ploennigs, Markus Berger: AI Art in Architecture, <https://arxiv.org/abs/2212.09399>
- [3] <https://nmkd.itch.io/t2i-gui>
- [4] <https://huggingface.co/hakurei/waifu-diffusion-v1-3>