

Stable Diffusion を利用したシンボルマーク画像の生成

藤本 竜也¹ 竹内 孔一²

¹ 岡山大学工学部 情報系学科

² 岡山大学大学院 環境生命自然科学研究科

pzla9iuw@s.okayama-u.ac.jp

takeuc-k@okayama-u.ac.jp

概要

デザイン会社などで多くのデザイン画像が蓄積されている一方で、近年の Stable Diffusion¹⁾をはじめとする Diffusion モデル [1] を利用した言語から画像を生成する研究が発展している。本研究では既存のデザインデータを利用することで、目的に応じた新たなデザイン画像を生成する手法について研究する。第一段階として本稿では、デザイン画像として、シンボルマークに注目し、Diffusion モデルにシンボルマークと関連するテキストを学習させた場合の画像の出力についてどの程度反映されているかを明らかにする。約 770 件程度のシンボルマークデータを学習させた結果、学習したデータの画像に近い絵を出力するようになったが、一方で、現段階ではデザイン画像として利用できない点があることを明らかにする。

1 はじめに

近年 Stable Diffusion (以降 SD) をはじめとした画像生成モデルが提案されて、言葉を入力することでそれに合った様々な画像を出力することが可能となっている²⁾。さらに、手持ち画像データを SD に LoRA を用いて学習させることで目的とした絵を出力できるようになった。これにより、データがあれば出力する絵をコントロールできるようになってきた³⁾ [2]。

一方で、デザイン会社などではすでに作成したデザイン画像が蓄積されており、これらのデータを学習させることで新たなデザイン画像を生成する際に、過去にデザインした内容を踏まえつつ新たなデザインを生成することができれば、人がデザイン画

像を作成する際の作業を補助できる可能性がある。

本研究ではシンボルマークなどある組織や団体のためのマークのデザインを生成することを目的とした画像生成モデル構築の方法について議論する。学習データとしてデザイン会社にご提供いただいた画像とテキストのペアデータ (約 770 件) を利用する。

画像生成実験の結果から、シンボルマークとして言葉で指定して学習したものに沿った画像は出るが、単純な適用ではテーマの欠如および文字データの不足からデザインとしては利用が難しいことを明らかにする。

以下では、シンボルマーク画像生成の開発枠組みについて述べた後、実験環境およびモデル、実験と結果と考察について述べる⁴⁾。

2 シンボルマーク画像生成システムの開発枠組み

シンボルマーク画像生成システムを開発するためにどのような方法による画像生成があるか整理する。大きく分けて、画像生成の入力方法、キーワードの選択、LoRA による学習におけるキーワードがあげられる。

まず、画像生成の入力方法として、テキスト (キーワード列) から生成する場合と画像とテキストから生成する場合がある。画像を入力する場合は参考となる画像が存在して、その画像を元に特徴を加えることになる。例えば ControlNet[3] では、線画像でポーズを指定するなど画像に対する指定を行うことができるが、現段階では発想を得るためのシステムを想定しているため、元とする画像は想定しない。一方で、テキストで入力することで生成したいシンボルマークをモデルの持っている自由度のある画像が生成されることが期待できる。

1) <https://ja.stability.ai/stable-diffusion>

2) <https://ja.stability.ai/stable-diffusion>

3) <https://huggingface.co/blog/lora>.

4) 本稿ではデザイン会社にご提供いただいた画像を提示せずかわりに、市町村章などの絵で処理画像などを説明する。一方で 4.3 節の結果はご提供いただいた画像による学習結果である。

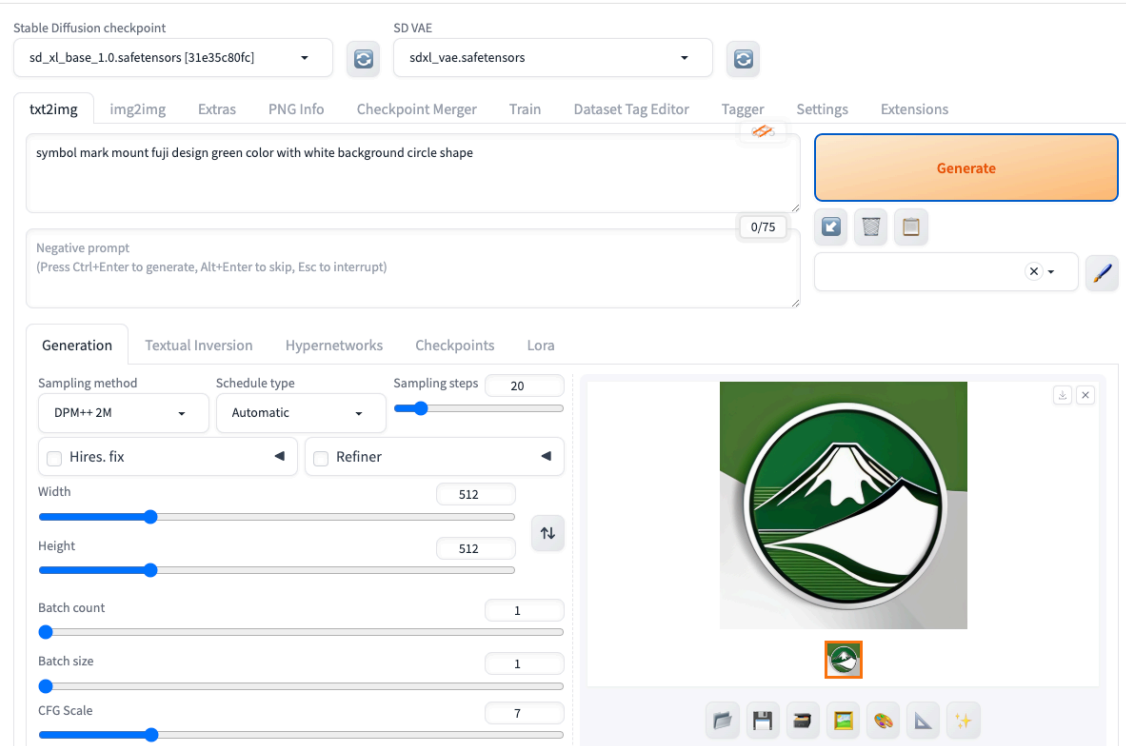


図 1 SDXL によるシンボルマーク開発環境と生成例

次にキーワードの選択であるが、SD は 75 文字までが画像生成に有効であるため、画像データに付随したテキストデータが説明調で長いテキストの場合は重要な語を選択する必要がある。また、SD の基本学習データのほとんどが英語のキャプションと画像なので、キーワードは英語の方が正確に画像を生成する。これらの点については実験後の考察であらためて結果と共に記述する。

最後に LoRA による学習であるが、画像の学習の際に入力するキーワードが重要となる。キーワードを入れた際に SD が持つ元々の画像データよりも学習したデータをより画像に反映させることを目的として、タスクを指定したキーワードを学習させておき、生成でも利用する。また上記と同様に生成された画像の優劣はキーワードに依存する部分が大いことを考察で記述する。

以上の方法の組み合わせを実行して絵を生成して、生成したデザインについてデザイナーの方に評価していただく。いただいた改善点をもとに、上記のシステムを改善することでシステムを開発する。

上記の開発システムはよく利用される SD を利用した開発環境で利用されている範囲であるため、実装のためのツールが整っている。下記に利用した環境について説明する。

3 シンボルマーク画像生成開発環境

画像生成モデルとして Stable Diffusion XL (以降 SDXL) を利用する。また画像生成タスクでは言葉を入力した後、すぐに画像を確認できることが重要であるため Stable Diffusion web UI⁵⁾ (以降、webUI) を利用する。webUI では、元々の SDXL の画像生成だけではなく、複数の学習した LoRA を登録して画像生成の際に選択して利用できる。これにより、異なる LoRA を利用した場合の画像生成において、目視での確認が容易となる。

図 2 にキーワード “symbol mark mount fuji design green color with white background circle shape” を入力した際の SDXL の画像生成結果を示している。プロンプトの入力画面として 2 段あり、上段が通常のキーワードであり、下段がネガティブプロンプト(ここに入力したキーワードにあたる絵を表示しないこと)を入力する部分がある。中央には LoRA を選択するタブがあり、右下に生成された画像が表示される。

こうした開発環境によって個別の生成が確認できる。一方で、下記の実験で行うように複数の生成画像に対して機械的に評価する場合は python による

5) <https://github.com/AUTOMATIC111/stable-diffusion-webui>

評価システムを別途作成した。

4 シンボルマーク画像生成実験

まず、1) 元々の SDXL でシンボルマークがどの程度の画像を生成できるか複数のキーワードを入力して目視で確認する。

次に、2) LoRA 学習の効果を測定するために、学習なしの元々のモデル SDXL が生成する画像と LoRA 学習した場合の SDXL が生成する画像のちがいのついて量的に比較する。具体的には 4.1 節で準備するシンボルマークデータを利用して、SDXL にテストデータに基づくテキストを入力して、画像を生成させてどの程度近い画像が生成されるかを評価する。その際、LoRA を利用して学習データを反映させた SDXL と比較する。

最後に 3) LoRA 学習したモデルが生成した画像を目視で確認して定性的な評価を行う。

4.1 実験データの準備

デザイン会社から提供された画像とテキストのペアは約 770 件であり、テキストにはデザインに対する要望などが記載されている。テキストは日本語で記述されているため、Llama-3-ELYZA-JP-8B を利用して英語に翻訳した。学習データとテストデータは 8:2 の割合に分割した。

4.2 評価方法

上記の評価実験 1) と 3) は人によるコメントによる評価とする。一方で、実験 2) では生成されたシンボルマークとテストデータのシンボルマークとの差分を比較するため二乗誤差で評価を行う。

4.3 実験結果と考察

評価実験 1): 例として図 2 で示したテキストによるシンボルマーク画像生成を複数回実行して、表示された画像を評価した。キーワードは静岡県富士市の市章⁶⁾を意識して作成している。そのためキーワードに green や circle shape など色や形まで指定している。富士市を選んだ理由は富士山ならばすでに SDXL が学習している可能性が高いと仮定したためである。図 2 に生成された画像の 1 例を示しているが、富士山の形を表した画像が生成されている。しかしながら市章と比較するとシンボルデザインとしてより抽象度を高める必要があるように見受けら

6) <https://www.city.fuji.shizuoka.jp/>

れる。

また、他の市章として岡山市の市章の生成⁷⁾は「岡」という漢字を図案化しているため英語のテキスト (キャプション) と画像を学習している SDXL では生成が期待できない⁸⁾。このようにシンボルマークの場合に日本語の文字や漢字をモチーフにしたシンボルマーク得たい場合には SDXL は利用が難しいことがわかる。

評価実験 2): SDXL とシンボルマークデータを LoRA を利用して学習した SDXL に対してテストデータのテキストを入力して生成させた画像とテストデータの画像との二乗誤差について表 1 に示す。

表 1 生成したテスト画像に対する二乗誤差

モデル	テストデータ
stable-diffusion-xl-base-1.0 (base)	107.7
stable-diffusion-xl-base-1.0 (LoRA)	102.7

表 1 から LoRA を適用した学習モデルの方が、テストデータに対しても人が作成した画像に近いことを示している。

評価実験 3): 上記の評価実験 2) の LoRA を学習したモデルに対してキーワードを入れて生成した画像についてデザイン会社の方に評価していただいた。その結果、生成された画像にはシンボルマークが持つべきテーマが見当たらないこと、さらに、文字に基づくデザインが生成できない点などの指摘を受けた。

このことから LoRA を利用して学習データに基づく画像を表示することは可能であるが、現段階ではシンボルマークがもつべきテーマが画像に現れておらず人がデザイン作成をする際の補助にはまだ距離があることが明らかになった。

4.4 まとめ

本研究ではデザイン会社などで蓄積されているデザインデータを利用して、人のデザイン作業を補助できる画像生成システムの構成法について議論した。シンボルマークを対象に約 770 件のデータから学習およびテストデータを作成して、SDXL にシンボルマークを生成させた。その結果、生成されたシンボルマークは学習データの画像をある程度再現しているが、デザインとして重要なテーマの欠如、および、漢字を含む文字のデザイン情報が入っておら

7) <https://www.city.okayama.jp/shisei/0000020691.html>

8) 日本語のテキストを受け付けるので、「シンボルマーク 中央に「岡」の字、周辺に山を配置」と入力すると画像は出力されたがよくわからない画像が出力された A 節参照。

ず生成されないことが明らかになった。

今後、テーマを結びつけた学習データの作成や文字情報の生成方法について検討する。

謝辞

本研究はデザイン会社からデータの提供およびデータの評価をいただきました。デザイン会社の皆様に深く感謝の意を表します。

参考文献

- [1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 6840–6851. Curran Associates, Inc., 2020.
- [2] Simian Luo, Yiqin Tan, Suraj Patil, Daniel Gu, Patrick von Platen, Apolinário Passos, Longbo Huang, Jian Li, and Hang Zhao. Lcm-lora: A universal stable-diffusion acceleration module, 2023.
- [3] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In **Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)**, pp. 3836–3847, October 2023.

A 評価実験 1) で日本語による画像生成例

日本語のキーワードで漢字を指示する「シンボルマーク 中央に「岡」の字，周辺に山を配置」に対して生成した画像例．両方の図とも「岡」の字に対す

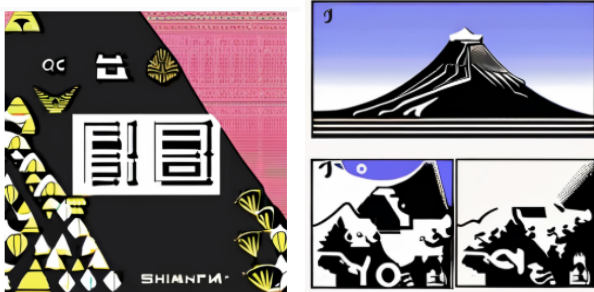


図2 SDXL によるシンボルマーク生成例 (岡山市)

る絵は見当たらない．一方で「周辺に山を配置」の表現に対応する山が左の図の三角や右の図の山に見受けられる．しかしながら，日本語での指示は期待できないことがわかる．