

## Stable Diffusion を用いた Text-to-Image の生成過程可視化に関する検討

### A study on Visualization of the Generation Process on Text-to-Image using Stable Diffusion

○篠崎俊<sup>1</sup>, 関弘翔<sup>2</sup>, 古田土祐樹<sup>3</sup>, 細野裕行<sup>2</sup>\*Shinozaki Shun<sup>1</sup>, Seki Hiroto<sup>2</sup>, Yuki Kotato<sup>3</sup>, Hosono Hiroyuki<sup>2</sup>

**Abstract:** This study examined the relationship between images generated by Stable Diffusion and text prompts. The objective is to visualize what the image is generated by gazing at by means of a heatmap with DAAM and a differencing process.

#### 1. 研究背景

近年, 生成 AI によるイラスト生成が非常に盛んに行われている. 代表的な生成 AI である Stable Diffusion は, Diffusion モデルを応用した Latent Diffusion Model(潜在拡散モデル:LDM)<sup>[1]</sup>を利用して. Diffusion モデルは, データにノイズを繰り返し付与する拡散過程と, 繰り返しデノイズする逆拡散過程からなり, 逆拡散過程を応用して画像生成などを行うことができる. LDM は, ピクセル空間ではなく, 圧縮された潜在空間で逆拡散過程を学習する点が特徴である. また, Stable Diffusion では Text-to-Image による生成が可能であり, 大きくテキストエンコーダと画像生成器の二つの部分で考えられる. テキストで示されたプロンプトがテキストエンコーダによりベクトルに変換され, これを潜在空間でデノイズする際に情報として付加することで, プロンプトに沿った画像を生成している.

本研究では, Text-to-Image におけるプロンプトと生成画像の関係の解明を目的に, 画像生成過程の差分処理およびヒートマップ可視化に取り組んだ.

#### 2. 画像生成過程差分処理およびヒートマップ可視化

本研究では, Stability AI が公開している Stable Diffusion v2.1<sup>[3]</sup>を学習済みモデルとして採用した. テキストは”a dog runs across the field”をベースとし, 単語や形容詞, 数名詞といったものを付加することで画像にどのような影響を与えるかを調査した. 生成過程に利用するデノイズのループ回数は 50 として固定し, 実験を行う. テキストと画像の関係を表すヒートマップ可視化には Diffusion Attentive Attribution Maps(DAAM)<sup>[2]</sup>を用いた. DAAM で得たヒートマップと, 画像生成過程におけるステップごとの差分処理を行い, プロンプトと画像生成, ヒートマップの関連性を調査した.

Fig. 1 に”a dog runs across the field”で画像を生成する過程および DAAM で得られたヒートマップを示す. また, Fig.2 に 25 ステップと 30 ステップ間の差分処理の例を示す. 生成過程の結果より, 初めに犬を中心に生成が行われ, デノイズが進むと犬の周囲の部分に変化が大きく見られた. この際, ”A dog”としていたテキストを”Three dogs”と変化させ画像を生成すると, 複数の犬を同時に注視し生成する様子が確認できた. Fig.1, 2 より, ヒートマップで最も強く注視されている部分を中心に, メインとなるオブジェクトや生物から, 徐々に背景などを生成していくことが明らかとなった.



Figure 1. Image generation process and heat map



Figure 2. Difference processing during image generation

#### 3. まとめ

本研究では, Stable Diffusion と DAAM を利用し, 生成に利用するテキストと生成された画像の関連性を検証した.

#### 参考文献

- [1] R. Rombach, et al., In Proc. of CVPR, pp.10684-10695, 2022.
- [2] R, Tang, et al, In Proc. of the 61st Annual Meeting of the Association for Computational Linguistics, 1, 2022.
- [3] Hugging Face, <https://huggingface.co/stabilityai/stable-diffusion-2-1> (2023/09/30)