

Stable Diffusion を用いた構図を指定する AI 画像生成に関する研究

A Study on AI Image Generation Using Stable Diffusion to Specify Composition

鄭 璟
指導教員 小川 毅彦

拓殖大学 工学研究科 機械・電子システム工学専攻 小川研究室

アブストラクト： Stable Diffusion は拡散モデルを使用した画像生成 AI である。プロンプトに基づく画像生成では、構図やポーズの指定が難しいという問題がある。本研究では、ControlNet を用いてその問題を解決する方法を検討する。

キーワード： Stable Diffusion、AI 画像生成

1. はじめに

2022 年 8 月にオープンソースとしてリリースされた Stable Diffusion は「拡散モデル(Diffusion Model)」を使用した画像生成 AI である。「プロンプト」という短いテキストで制御することで、そのイメージと近い画像を生成してくれる。しかし、このモデルでは、生成される画像のランダム性が高く実用化は難しい。

画像生成の応用として、指定した構図やポーズに合わせて人物画像を生成する問題が考えられる。Stable Diffusion の拡張機能として、ControlNet が発表されている。この拡張機能には、入力画像の姿勢を解析し、その姿勢に沿った画像を生成する機能がある。本研究では、この拡張機能を使い、構図やポーズなどを指定する AI 画像生成について検討を行う。

2. 拡散モデル (Diffusion Model) について

拡散モデルは 2020 年にカリフォルニア大学バークレー校の研究者らが発表した論文に基づいた仕組みである。画像にノイズを加われば完全なノイズになる。ノイズを取り除けば画像になる。ノイズを取り除く過程を AI に学習させ、その過程をテキストなどで制御して新たな画像を生成できる。図 1 に拡散モデルの原理と画像の生成過程を示す。

3. ControlNet について

ControlNet はハンドルネーム llyasviel、チャン・リュミン氏が 2023 年 2 月に発表したものであり、その後 2 週間程度で Stable Diffusion の拡張機能に組み込まれた。最新版の ControlNet は 14 種類の機能がある。本研究ではその中の「OpenPose」という機能を主に研究する。これは入力画像の情報を読み取らせ、その画像の姿勢を解析して、棒人間状のデータに変換して、そのポーズに沿った画像を生成する機能である。

4. 研究内容と開発環境

ControlNet の OpenPose (オープンポーズ) 機能を使用し、ポーズを指定する AI 画像生成について、生成された画像の構図とポーズについて評価を行う。さらに、Depth 機能で求めた深度情報によって、体の前後関係を指定する方法についても検討を行う。開発環境については、ローカル版「Stable Diffusion web UI」を用いる。

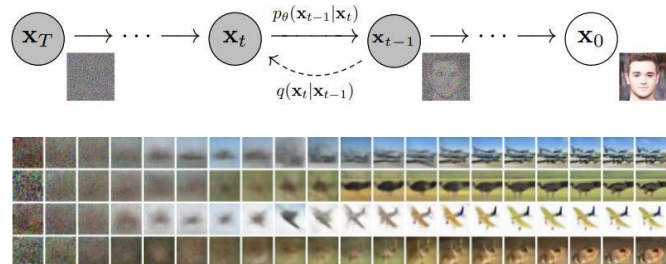


図 1 拡散モデルの原理と画像の生成過程

5. 実験結果と考察

(1) ヨガをする女性の画像から新たな画像を生成

ヨガをする女性から、プロンプトと OpenPose を用い、同じヨガの姿勢をとる警察官の画像を生成する。入力画像と姿勢、生成された画像の 1 つを図 2 に示す。結果は全体的によく生成できている。



図 2 ヨガをする女性と姿勢データ、生成したヨガをする警察官の画像

(2) 腕が重なる姿勢データから新たな画像を生成
運動する男性から、プロンプトと OpenPose を使用し、同じ運動の姿勢をとる警察官の画像を生成する。図 3 に示すように、この男性は腕が交差しており、読み取った姿勢データでは、その部分の読み取りが誤っていた。生成結果では警察官の左腕が後ろに回り、生成は失敗だった。

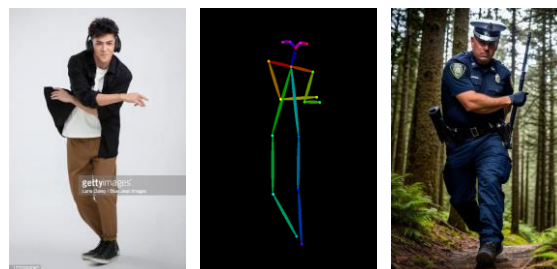


図 3 運動する男性と姿勢データ、生成した運動する警察官の画像

(3) 姿勢データを手動で調整して行った実験

(2) で読み取った姿勢データを手動で調整した後で、警察官の画像を再度生成した。結果の 1 つを図 4 に示す。結果から、調整後のデータがうまく適用され画像が生成されたことが分かった。

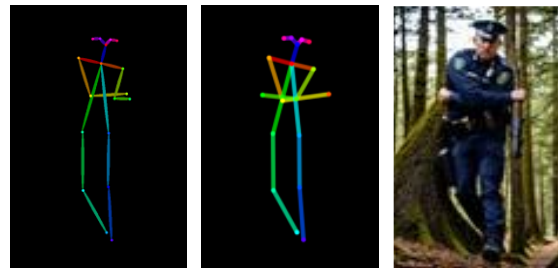


図 4 調整前後の姿勢データと、生成した運動する警察官の画像

(4) 深度情報を用いて行った実験

Depth 機能を用いて深度情報を求め、(2) を再実験する。Depth は画像の前後関係を読み取ることができる。これを用いて行った結果を図 5 に示す。結果から元の画像における体の部分の前後関係が生成結果に適用され、正しく生成されている。



図 5 運動する男性と深度情報データ、生成した警察官の画像

6. まとめ

本研究では、Stable Diffusion を用いた画像生成について、ControlNet のツールである OpenPose と Depth で構図を指定することを検討した。実験の結果、プロンプトだけでは構図やポーズを十分に指定できない場合であっても、ControlNet のツールを用いることで、その問題を解決できることを示した。

参考文献

- [1] 今村勇輔, Stable Diffusion AI 画像生成ガイドブック, ソシム, 2023.
- [2] 次世代生成モデル, Diffusion Model の紹介, https://recruit.gmo.jp/engineer/jisedai/blog/diffusion_model
- [3] 画像生成 AI に 2 度目の革命を起こした「Control Net」, <https://ascii.jp/element/000/004/137/4137393/>