**知能プログラミング演習I 第８回レポート（課題１）**

2024年 7月 29 日 学籍番号 34714037 氏名 加藤薫

調べた深層学習モデル：Stable diffusion

１．モデルの原理

ダイアグラム

自動的に生成された説明

図 1**「The Illustrated Stable Diffusion」より引用**

Stable DiffusionはVAE，U-Net，Text Encoder CLIPの3つによって構成される．学習時は，ピクセル空間に画像を与えてVAEエンコーダで潜在表現に変換し，Text Encoder CLIPにテキストを与えて潜在空間で拡散モデルの処理を行い，ノイズが付与され，U-Netのパラメータが更新される．新しくデータを生成する時は，Text Encoder CLIPにテキストを与えて，潜在空間で順方向拡散過程の出力をデノイズする逆拡散過程を行い，その後VAEのデコーダを使ってピクセル空間に変換する．CLIPは画像とそれに対応するテキストを使って学習したニューラルネットワークであり，画像とテキストをそれぞれベクトルに変換し，その2つのベクトルから類似度を計算して学習を行う．U-Netでは畳み込み結果と，テキストを変換した「条件」をAttention層に入力する．

２．モデルの実応用上の利点

テキストから画像を生成することができ，創造的なプロジェクトやビジネスソリューションで誰でも簡単に画像を生成することができる．ユーザーはサンプリング方式、出力画像の解像度、シード値などのパラメータを調整することができる．テキストプロンプトで指定された要素を含む、オリジナルの画像に基づいた新しい画像を出力することもでき，匿名化する，画像を元の解像度以上に拡張し、指定されたプロンプトに基づいて生成されたコンテンツで拡張された空白の部分を埋めるなどができる。

３．モデルの実用上の（あるい社会的倫理的観点からの）問題点

Stable Diffusionは著作権のある画像から権利者の同意なしに学習を行っているため、画像の使用に関してユーザーに提供される自由について、所有権の倫理をめぐる論争が引き起こされている．画像生成AI開発者は画像生成AIがディープフェイク・ポルノに悪用されることを想定する必要があり，「Stable Diffusion」は露骨な画像の生成を防ぐガードレールや悪意ある使用を禁じるポリシーを設けている．

４．考察および感想

Stable diffusionは学習時は拡散過程でノイズを加え，生成時は逆拡散過程でノイズを取り除いていくという拡散モデルの考え方をもとに，U-NetとVAEを組み合わせることで高解像度の画像を効率よく学習・生成できると考えられる．画像生成モデルは学習に大量の画像が必要なため，著作権のある画像も使わざるを得ず，著作権の問題がある．生成された画像を著作権侵害となるかどうかは難しく問題であり，技術の発展のためにある程度は許されてよいと思うが，クリエイターに対して協調的な態度も必要である．

モデルの原理は概要は分かったが，具体的な数式は難しく詳細は理解できていない．時間があるときにプログラムを書いて実行し，理解を深めようと思う．

参考文献（Webページはすべて７月３０日アクセス）

「ゼロから作るDeep Learning⑤―生成モデル編」斎藤康毅著

オライリージャパン　2024年４月８日発行

「The Illustrated Stable Diffusion」Jay Alammar著

[The Illustrated Stable Diffusion – Jay Alammar – Visualizing machine learning one concept at a time. (jalammar.github.io)](https://jalammar.github.io/illustrated-stable-diffusion/)

Wikipedia「Stable Diffusion」

[Stable Diffusion - Wikipedia](https://ja.wikipedia.org/wiki/Stable_Diffusion#cite_note-bijapan-69)

「生成AIの悪用について」

[氾濫するディープフェイク・ポルノ、生成AIの悪用に高まる懸念 | WIRED.jp](https://wired.jp/article/dark-side-open-source-ai-image-generators/)