Stable Diffusionによる画像自動生成AIの実装と設計レファレンスへの応用
Implementation of Automatic Image Generation AI by Stable Diffusion and Application to Design Reference

○稲田　浩也[*1]
Kouya INADA[*1]

*1 京都大学大学院工学研究科建築学専攻　博士後期課程　修士(工学)
Grad Student, Graduate School of Engineering, Kyoto University, M. Eng.

キーワード：Stable Diffusion; 拡散モデル; 画像自動生成AI; Text-to-Image; 設計レファレンス
Keywords: Stable Diffusion; Diffusion model; Automatic image generation AI; Text-to-Image; Design reference.

1. 開発背景と目的

近年，事前学習画像分類モデルや，拡散モデル（Diffusion Models）の精度向上により，自然言語を入力とし画像を出力する画像自動生成AIの開発が世界各国で進んでいる．

建築物の基本設計や企画段階において，施主との成果物イメージの共有や参考資料として，複数の建築物の画像・ドローイング・イメージパース（以下イメージと称する）などを参照することがある．その際に，適当なイメージの探索に多くの時間を費やすことがある．

こうした状況を踏まえ，本研究では，画像自動生成AIをクラウド上に実装し，設計者のレファレンスツールとして用いる有用性を検証する．

2. Stable Diffusionによる画像自動生成AI

多くの画像自動生成AIは，入力した語と画像を紐づけるText encoderと画像を生成するImage generatorからなる．Text encoderにはCLIPというモデルが，Image generatorには拡散モデル(Diffusion Models)が主に用いられている．

CLIPは2021年にOpenAI社のRadfordらが発表した「CLIP(Contrastive Language-Image Pre-training)」という事前学習画像分類モデルのことである[1]．画像とテキストの組み合わせを学習データとしているため、カテゴリー設定の自由度が向上していることが特徴である．

拡散モデル（Diffusion Models）は2015年にSohl-Dicksteinらが提唱した論文[2]が元となっている．その後，2020年にHoらがDenoising Diffusion Probabilistic Models（DDPM）として改良したモデル[3]を発表している．Diffusion Modelsは，画像にノイズを加えて最終的に全ての情報が失われノイズのみになる確率課程を逆向きにたどることでモデルを学習させるというものである．このモデルはトレーニングや推論に膨大なGPUリソースを必要であったが，2021年にRombachらが,潜在拡散モデル(Latent Diffusion Models)を発表[4]し計算量を大幅に削減しながら、従来に匹敵するパフォーマンスを実現することに成功している．

Stable DiffusionはText encoderにCLIPを，Image generatorにLatent Diffusion Modelsをとして活用しているオープンソースのモデルであり，Stability AI社によって，オープンソースコミュニティーHuggingFace内で提供され，クラウドサービスを利用することで誰でも自前で画像自動生成AIを実装できるようになった．

3. 実装環境と具体的な実装内容

開発環境は無料で一定のGPUリソースが活用できるGoogle colaboratoryを利用する．また，スクリプト言語はPython（Python 3.7.13）とする．まず，Google colaboratory上で「!pip install diffusers==0.3.0 transformers scipy ftfy」としStable Diffusionのインストールする．その後HuggingFace Hubのトークンを割り当て，パイプラインを構築する．

4. 画像生成

Stable Diffusionに対して英語で出力したい画像の文章（プロンプト）を英語で入力する．縦横比は512pt×512pt，guidance_scaleは7.5，ステップ数は50とする．この際チェリーピッキングはせず，初めに生成された画像を掲載する．

本報では大きく①既存の事例に近い建築物，②今までない建築物の生成を試みる．

4.1. 既存の事例に近い建築物

プロンプトは以下の通り．

Fig.1：「a photo showing the whole picture of High-rise building built by a major Japanese general contractor, with commercial facilities on the lower floors, offices on the upper floors, and a large park attached, high quality, Architectural design competition, high-definition rendering, beautiful」

Fig:2「a photo of Interior view of state-of-the-art office designed by a renowned designer with atrium, award-winning, high quality,　Architectural design competition, high-definition rendering, beautiful」

Fig:3「Image of a large beautiful park near a High-rise building built by a major Japanese general contractor, with a café attached, award-winning, high quality, Architectural design competition, high-definition rendering, beautiful」

4.2. 今までない建築物

プロンプトは以下の通り.

Fig.4「a photo of futuristic art museum on the sea, high quality, Architectural design competition, high-definition rendering, beautiful」

Fig.5「Photo of Stone house made of hollowed out large tree , High quality , Architectural design competition , High-definition rendering , Beautiful」

Fig.6「Watercolor of a station building made of sweets, by an impressionist painter , High quality , Architectural design competition , High definition , Beautiful」

5. 生成結果の考察と今後の展望

細部ではやや破綻があるが，プロンプトの抽象度や建材・立地・画風を問わず参考画像として十分なクオリティの画像が生成できた．一方で，参考画像としての使用に際しては，生成された画像内の建築物が建築として成立しているか把握できる知識が求められる．また，施主が明確にイメージを持っている場合は実空間のデータベースで代替できることから，今までにないビルディングタイプを施主に示す場合や、施主自身のイメージが漠然としている場合により有用である可能性が高い.

こうした画像が数十秒でほぼ無限に生成できること，2022年9月13日現在の日本の著作権法上では著作権フリーであり資料が一般的に公開されるような場合にも活用できることから，設計者のレファレンスツールとして十分に活用可能性があると考えられる.

今後の展望としては，例えば「house」を含むプロンプトで生成した画像に含まれる建築物は欧米でよく見られる住宅の形式であるなど，学習に使用されたデータセットLAION-5Bに多少の偏りがあると思われるため，特定の建築物や景観に特化させてファインチューニングを行い，より実用性の高いモデルを構築することなどが考えられる.

[参考文献]

1) Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever : Learning Transferable Visual Models From Natural Language Supervision, arXiv:2103.00020, ODI : https://doi.org/8550/arXiv.2103.00020
2) Jonathan Ho, Ajay Jain, Pieter Abbeel : Denoising Diffusion Probabilistic Models, arXiv:2006.11239 ODI : https://doi.org/10.48550/arXiv.2006.11239
3) Jonathan Ho, Ajay Jain, Pieter Abbeel : Denoising Diffusion Probabilistic Models, arXiv:2006.11239 ODI : https://doi.org/10.48550/arXiv.2006.11239
4) Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer : High-Resolution Image Synthesis with Latent Diffusion Models, arXiv:2112.10752 ODI : https://doi.org/10.48550/arXiv.2112.10752

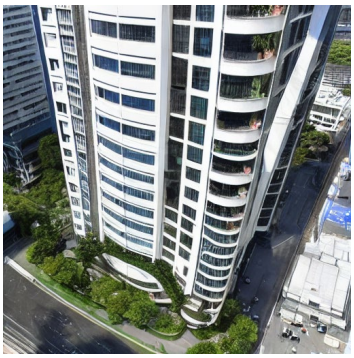

Fig.1 事務所ビル外観の生成画像

Fig.2 事務所ビル内観の生成画像

Fig.3ランドスケープの生成画像

Fig.4 美術館外観の生成画像

Fig.5 家外観の生成画像

Fig.6 駅舎外観の生成画像