

TSUBAME 共同利用 令和7年度 学術利用 成果報告書

利用課題名 対話型画像生成 AI の開発
英文: Development of interactive visual generation AI

利用課題責任者 HAO Guoqing
First name Surname Guoqing HAO

所属 青山学院大学
Affiliation Aoyama Gakuin University

邦文抄録(300 字程度)

本研究では、画像生成 AI における生成後再編集の困難性を解消し、ユーザー意図をより正確に反映可能な対話型画像生成 AI の実現を目指した。TSUBAME 上で大規模データセットを用いたモデル学習、条件制御手法の検討、比較評価実験を実施した。具体的には、意味マスクを用いて指定領域に指定カテゴリの物体をより忠実に生成する手法を検討した。さらに、実画像編集において編集強度を連続的に制御する学習不要の手法 FlowSlider を提案した。これらの成果により、ユーザー指定領域への忠実な生成と、連続的かつ安定した画像編集の両面から、対話型画像生成 AI の基盤技術を前進させた。

英文抄録(100 words 程度)

This project aims to develop interactive image generation AI that enables accurate reflection of user intent and easy post-generation editing. Using TSUBAME, we conducted large-scale training, fine-tuning, and evaluation of image generation and editing models. We studied a semantic-mask-based object generation method faithful to user-specified mask shapes. In addition, we proposed FlowSlider, a training-free continuous image editing method that enables stable slider-style control of edit strength while preserving source fidelity. These results advance the core technologies for interactive and controllable image generation.

Keywords: Interactive image generation, image editing, diffusion model, semantic mask, continuous control

背景と目的

近年、Stable Diffusion をはじめとする画像生成 AI は著しく高品質化しているが、生成結果の内部表現が複雑であるため、生成後に一部だけを自然に修正したり、ユーザー意図に応じて段階的に編集したりすることは容易ではない。特に、どこに何を生成するかという空間的制御と、どの程度編集するかという連続的制御の両立は重要な課題である。

本研究の目的は、対話型画像生成 AI の基盤として、ユーザーの指示や操作を反映しながら、画像中の対象領域や編集強度を柔軟に制御可能な生成・編集技術を開発することである。そのため、TSUBAME の計算資源を活用し、大規模データセットを用いたモデル微調整、条件制御アルゴリズムの学習、反復的な評価実験を実施した。

概要

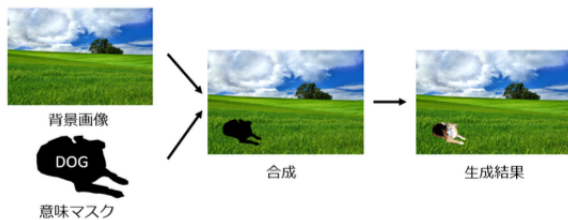
研究は大きく二つのテーマから構成される。第一に、指定領域に対して指定カテゴリの物体をマスク形状に忠実に生成する手法の検討である。背景画像と意味マスクを条件として利用し、ControlNet によって空間条件と意味条件を生成過程に導入することで、既存の inpainting 手法では難しかった形状追従性の改善を目指した。PRMU 発表では、MSCOCO のセグメンテーションデータから 10 カテゴリのデータセットを構築し、再学習した SD1.5 Inpainting と比較評価を行っている。

第二に、実画像編集における連続制御の実現である。FlowSlider では、Rectified Flow に基づく編集過程を fidelity term と steering term に分解し、steering term のみをスケールリングすることで、追加学

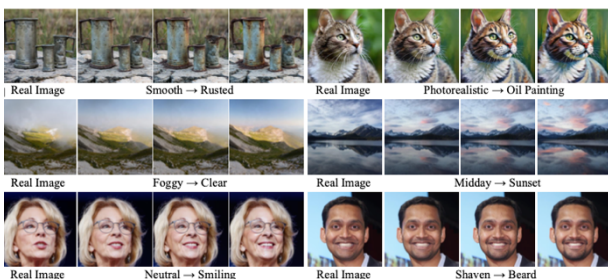
習なしにスライダ形式の連続編集を可能にした。この仕組みにより、元画像の忠実性を保ちながら、編集強度を滑らかに変化させることを狙った。

結果および考察

意味マスクを用いた物体生成では、提案手法は再学習した SD1.5 Inpainting と比較して、FID を 4.570 から 3.765 に改善し、IoU を 0.373 から 0.491 に向上させた。これにより、指定されたマスク形状への追従性が高まり、ユーザーが意図する領域に物体を生成しやすくなることが確認された。一方で、複雑背景や細長いマスクでは欠損や境界破綻が残り、カテゴリ整合と形状保持の両立が今後の課題であることも明らかとなった。



また、FlowSlider では、連続画像編集ベンチマークにおいて、FlowSlider (FLUX.1) が Mono 0.833、Smooth 0.01、DreamSim 0.090 を示し、安定したスライダ挙動と高い元画像忠実性を両立した。これは、編集強度を増してもソース画像の構造や見た目を保ちつつ、意図した方向に滑らかに編集を進められることを示している。従来の単純なスケーリングではアーティファクトや不安定性が生じやすかったのに対し、本手法は対話的編集に適した実用的な制御方法であると考えられる。



まとめ、今後の課題

本研究では、TSUBAMEを活用して、対話型画像生成 AI の基盤技術として、指定領域に忠実な物体生成手法と、追加学習を必要としない連続画像編集手法を検討した。その結果、国内発表および国際会議投稿とい

う形で研究成果を発信することができた。

今後は、構図・色彩・形状など複数要素をより独立に制御できる生成モデルへ拡張するとともに、テキスト指示、参照画像、マウス操作など多様なユーザーフィードバックを統合する対話的枠組みへ発展させる予定である。また、評価実験の規模拡大とユーザー評価の導入により、実運用を見据えた使いやすさと解釈性の向上を目指す。