

Title	生成モデルによる曖昧から具体的なリアルな人物画像の合成
Author(s)	彭, 以琛
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19065
Rights	
Description	Supervisor: 宮田 一乗, 先端科学技術研究科, 博士

概要

データ取得と大規模生成モデル技術の急速な進展に伴い、文章表現や絵画を含む芸術活動のパラダイムは絶えず進化している。コンピュータグラフィックス（CG）とコンピュータビジョン（CV）の領域において、これらの技術的進歩はイメージデザインプロセスの効率を大幅に向上させてきた。歴史的に、デザイナーは厳しい練習を通じて芸術的能力を磨き、古典的な芸術作品に没頭することで審美眼を養ってきた。

現在、粗いスケッチやテキストプロンプトだけで、生成モデルやデータベースは魅力的な画像を生成することができ、基礎的な芸術性がほとんどまたは全くない初心者でさえも、グラフィカルなデザイン能力を発揮できるようになった。

生成モデルは、単純な入力から複雑で高品質な画像を生成することができる。しかし、このような進歩がデザイナーの生産性の顕著な向上に直接つながるかどうかを解明することは重要である。AI生成モデルの背後にあるアルゴリズムは確かに強力かつ効率的であるが、その価値はデザイナーのニーズとの整合性にかかっている。筆者は、現在の画像生成モデルがデザイナーに提供する価値は、モデルの性能能力と一致していないと考える。これらのモデルの大多数はエンドツーエンドの「ブラックボックス」として機能しており、ユーザーの直接的な入力から所望の結果を得ることを困難にしている。この問題に対処するため、本論文では以下のアプローチを議論する：1) アルゴリズムの開発の観点から：入力モダリティの多様性を許容するように生成モデルのアルゴリズム設計を強化する。これにより、モデルの性能が向上するだけでなく、アルゴリズムとのインタラクションも豊かになる。最近の多様な入力、例えばテキストプロンプト、画像参照、スケッチやセマンティックマップを含む空間的ガイダンスを通じて画像を生成する能力を例証する。2) モデルとの対話の観点から：創造的プロセス中に広範なインタラクティブおよび編集機能を提供する。これにより、生成モデルはワンショットソリューションではなく、探求と洗練のためのツールとなり得る。3) デザインプロセスの観点から：創造の各段階で特定のニーズを理解し、デザイナーのワークフローに関与することが不可欠である。生成モデルのアルゴリズムはこれらのニーズに対応するように調整されるべきであり、開発したツールは技術的に優れているだけでなく、文脈的にも関連していることを確認する必要がある。

ピカソは次のような指摘をしている、「絵は事前に考えられたものではない。描かれる過程で、それは人の考えが変わるにつれて変化する。そして完成した時、それは見る人の心の状態に応じて変わり続ける。」したがって、創造は探求であり、デザイナーは満足が得られるまで中間成果に反復して思い

を巡らせ、着想を得ている。現代の生成モデルやデータ取得技術は、この理念を完全に包含しているとは言えない。

本論文では、創造的軌跡を「曖昧から具体へ」という連続体として概念化する。フルボディヒューマンイメージデザインをケーススタディとして、従来のデザインのワークフローを三つの別々の段階に分解する。各段階において、ユーザのユニークな要求を特定し、設計意図を実現するために最も適したデータ取得または生成モデリング技術を導入する。これらの段階は以下の通りである（順序に従って）：

- **ポーズ初期化**：ユーザの意図がまだ形成されておらず、探索が必要な初期段階では、「グローバルからローカルへ」という 3D モーションデータの取得スキームを導入する。伝統的な骨格スケッチの代わりに、ユーザは特定の関節の軌跡を描き、モーションデータの断片を取得する。これにより、ユーザは 3D 空間内の異なる角度からモーションデータを見ることができ、特にダンスなどのダイナミックなアクションを描写することを目指している場合に望ましいポーズを選択することができ、ユーザにより広範な参照と選択肢を提供する。
- **服装選択**：この段階では、多数の服装サンプルをふるいにかけて望ましいデザインを特定し、反復的な洗練を行うことがよくある。衣装デザインには布地に対する専門知識が必要であることを考慮して、筆者は、そのような専門知識を必要としない「イメージガイド」の生成モデルを提供する。このアプローチにより描画入力段階を省略し、ユーザが姿勢と全体的なキャラクターの描写との服装の整合に集中できるようにする。
- **顔とディテールの描写**：全体的な服装とポーズが確立されると、ユーザは通常、特に髪型や表情のような顔の属性を含む複雑な細部のデザインへと意識を移す。ここで、高忠実度の「スケッチガイド」生成モデルを提案し、出力が入力を密接に反映しつつ、編集されていない領域の一貫性を保つことを保証する。

検証実験を通じて、本研究はフェーズごとのインタラクティブパイプラインの有効性を検証し、類似のタスクにおける最先端（SOTA）の成果物と比較してベンチマークする。実証結果は、私たちのパイプラインが曖昧から具体への各フェーズを巧みにナビゲートし、意義深いデザインサポートを提供することを確認する。筆者は、提案した方法論や概念が人物像デザインに限定されるものではなく、多様なデザインシナリオに容易に適用可能であると考えている。このようにして、ここで提供されるフレームワークと洞察は、今後のデザイン研究における探求と革新のための基礎的な柱として機能することができると思う。

Keywords: キャラクター画像生成、曖昧から具体へ、データ取得、生成モデル、芸術的創造。