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 一乘, 先端科学技術研究科, 博士

氏　　　　　　名	彭　　以　琛
学 位 の 種 類	博士（情報科学）
学 位 記 番 号	博情第 524 号
学位授与年月日	令和 6 年 3 月 22 日
論　文　題　目	AMBIGUOUS-TO-CONCRETE REALISTIC HUMAN IMAGE SYNTHETIC VIA GENERATIVE MODEL
論 文 審 査 委 員	宮田 一乘　北陸先端科学技術大学院大学　教授
	池田 心　同　教授
	小谷 一孔　同　教授
	岡田 将吾　同　准教授
	栗山 繁　豊橋技術科学大学　教授

論文の内容の要旨

In the midst of rapid advancements in data retrieval and large-scale generative model technologies, the paradigms of artistic endeavors, including writing and painting, are constantly evolving. Within the realms of Computer Graphics (CG) and Computer Vision (CV), these technological strides have substantially bolstered the efficiency of image design processes. Historically, designers would hone their artistic prowess through rigorous practice, paired with an immersion in classical artworks to cultivate their aesthetic judgment.

Generative models are able to produce complex, high-quality images from simple inputs. However, it is crucial to demystify the notion that such advancements inherently equate to significant increases in designers' productivity. While the algorithms behind AI generative models are indeed potent and efficient, their value to users is contingent upon their alignment with the designers' needs. From my perspective, the value that current image-generative models provide to designers is not commensurate with the models' performance capabilities. The majority of these models operate as end-to-end 'black boxes', making it challenging for users to achieve desired outcomes through straightforward inputs. To address this issue, this thesis discusses the following approach: 1) From the perspective of developing an algorithm: enhance the algorithm design of the generative model to allow for a variety of input modalities. This would not only improve model performance but also enrich user interaction with the algorithm. Recent advancements in multimodal generative models exemplify the capability to produce images through various inputs such as text prompts, image references, and spatial guidance, including sketches and semantic maps 2) From the perspective of interaction with models: Offer expansive interactive and editing capabilities during the creative process. By doing so, the generative model becomes a tool for exploration and refinement, rather than a one-shot solution. 3) From the perspective of the design process: It is imperative to engage with the designer's workflow, understanding the specific needs at each stage of creation. Generative model algorithms should be tailored to address these needs, ensuring that the tools developed are not just technologically advanced, but also contextually relevant.

As Picasso aptly noted, "A painting is not thought out in advance. While it is being done, it changes as one' s thoughts change. And when it' s finished, it goes on changing, according to the state of mind of whoever is looking at it." Thus, creation is an exploration, with designers iteratively reflecting on and drawing inspiration from intermediate outputs until satisfaction is achieved. Contemporary generative models and data retrieval

techniques have yet to fully encapsulate this ethos. In this dissertation, we conceptualize the creative trajectory as an Ambiguous-to-Concrete continuum. Taking full-body human figure design as a case study, we break down the conventional figure painting workflow into three separate stages. For each stage, we identify the unique requirements of the user and introduce new data retrieval or generative modeling techniques specifically designed to best assist the user in achieving their creative vision. These stages include (In sequential order):

- **Posture Initialization:** At the initial stage where user intent is still forming and exploration is needed, we introduce a 'global-to-local' retrieval scheme for 3D motion data. Instead of traditional skeletal sketching, users draw trajectories for specific joints to retrieve and refine snippets of motion data, choosing keyframes that will guide further design steps. This allows users to view the motion data from different angles within a 3D space, enabling them to select the desired pose, particularly when aiming to depict dynamic actions such as dancing, thus offering users a broader range of referencesand choices.
- **Outfit Selection:** This phase often entails sifting through a plethora of attire references to pinpoint desired designs, coupled with iterative refinements. Considering the intricacies of fabric depiction demand profound sartorial design expertise, our solution offers an 'image-guided' generative model. This approach omits the drawing input stage, allowing users to concentrate on attire's alignment with the posture and the overall character portrayal.
- **Facial and Detail Depiction:** With the overarching attire and posture solidified, users typically possess clearer intentions regarding intricate details, especially facial attributes like hairstyle and expressions. Our solution here is a high-fidelity 'sketch-guided' generative model, ensuring the output closely mirrors the input while maintaining consistency in non-edited areas.

Through rigorous experiments, we validate the efficacy of our phase-specific interactive pipelines, benchmarking them against state-of-the-art (SOTA) counterparts on analogous tasks. The empirical results affirm that our pipeline adeptly navigates each phase of the Ambiguous-to-Concrete spectrum, offering meaningful design support. We posit that our methodologies and concepts are not confined to human figure design but are readily applicable across diverse design scenarios. As such, the frameworks and insights proffered herein can serve as foundational pillars for subsequent inquiries and innovations in the design research landscape.

Keywords: Character Image Generation, Ambiguous to Concrete, Data Retrieval, Generative Models, Artistic Creation.

論文審査の結果の要旨

本博士論文は、深層学習を用いた人物画像生成モデルを提案し、ユーザスタディの結果、提案手法の有効性を客観的数値データに基づいて示したものである。

直近の数年間で生成 AI 技術は飛躍的に進歩し、数行のプロンプト（文章）で高品質な画像を生成可能である。しかしながら、このようなインスタントな制作手法では、制作者の意図を十分に反映できるとは言い難い。本博士論文は、現状の生成 AI 技術の根本的な課題に対し、「曖昧から具体へ」の連続的な制作プロセスを経て、抽象的なモデルから段階的に具現化する新たな画像生成モデルを構築することで、制作者の意図を反映可能な手法を提案している。具体的には、1)人物のポーズ生成、2)服装の生成、3)表情や髪型などの調整、の 3 段階を経て、人物画像を生成する手法を考案した。

まず、ポーズ生成の段階では、ユーザが描画入力した特定の関節の軌跡から、一連の動作データを生成する。3 次元空間内の異なる角度からユーザは生成された動作データを閲覧し、望ましいポーズを選択することができる（写真家がモデルに自由に動いてもらいながら撮影するように）。

次に、服装生成の段階では、ユーザは所望のモデルの参照画像を指定し、第一段階で得た「望ましいポーズ」に適用する。これを実現するために、望みの服装で所望のポーズを取った人物画像を生成する新たな画像生成ネットワークを構築した。

制作の最終段階では、ユーザは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ため、高忠実度の「スケッチガイド」生成モデルを提案した。提案した生成モデルでは、非編集部分の一貫性を維持しながら、出力画像を入力スケッチに忠実に反映することを可能とした。

提案した手法は、人の創作活動と整合性のあるインタフェース上で実装され、評価実験の結果、手法の有効性を確認している。ピカソが残した「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。」という言葉通り、創造は探求であり、人は満足が得られるまで反復して途中の制作物に思いを巡らせ、新たな着想を得る。提案手法はこの哲学に基づいたものである。

以上、本論文は、深層学習を用いた人物画像生成モデルを提案し、制作者の段階的な制作プロセスを支援したものであり、学術的に貢献するところが大きい。よって博士（情報科学）の学位論文として十分価値あるものと認めた。