



Title	Text-to-Image Generation for Art and Society
Author(s)	Wu, Yankun
Citation	大阪大学, 2025, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/103166
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Abstract of Thesis

Name (WU YANKUN)	
Title	Text-to-Image Generation for Art and Society (アートと社会のためのテキストから画像生成)
<p>Abstract of Thesis</p> <p>Text-to-image generation models have attracted increasing attention due to their remarkable capabilities in producing high-quality images given natural language prompts. From a positive perspective, these models provide novel opportunities to address challenges that were difficult to solve. However, they also inherit and amplify societal biases, reproducing stereotypes related to gender, race, and age in the generated images.</p> <p>To examine both their potential for downstream applications and their ethical risks, this thesis investigates the impact of text-to-image generation models on both art and society. First, we explore how generative models can be used to improve digital art analysis. We propose GOYA, which leverages synthetic images generated by Stable Diffusion and employs contrastive learning to disentangle content and style in art paintings. GOYA demonstrates strong performance in downstream tasks such as classification and retrieval, highlighting the potential of generative models in the digital humanities.</p> <p>Despite their promising applications in representation learning, the inherent bias of generative models cannot to overlooked. To address these ethical concerns, we introduce an automatic protocol to evaluate gender bias in text-to-image generation. Through a systematic analysis of representational disparities, object co-occurrence similarity, and prompt-image dependencies, we reveal that gender bias originates from text embeddings, propagates through the generation process, and is reflected in the entire output image. Based on these findings, we present a training-free method for mitigating gender bias. Our approach interpolates between feminine and masculine embeddings within both the text embeddings and the attention module of the model to generate fairer and diverse neutral outputs. This method does not require model fine-tuning or additional data, offering a lightweight and practical solution for improving fairness in image generation.</p> <p>With these contributions, this thesis offers a comprehensive view of the capabilities and risks of text-to-image models, shedding light on their potential to enhance representation learning while underscoring the importance of addressing their social impact.</p>	

論文審査の結果の要旨及び担当者

氏 名 (Yankun Wu)			
	(職)	氏 名	
論文審査担当者	主 査	教授	中島悠太
	副 査	教授	長原一
	副 査	教授	浦西友樹
	副 査	准教授	大倉史生

論文審査の結果の要旨

本学位論文は、テキストから画像を生成する画像生成モデルの安全性についての検討から応用に関するものであり、画像生成モデルの芸術分野への応用として絵画の内容と画風に関する特徴の分離（第2章）、画像生成モデルの社会的バイアスの定量的評価（第3章）、さらに社会的バイアスの低減手法に関する研究成果に基づく。

第2章では、画像生成モデルの新しい応用可能性の探究を目的として、絵画における描画内容と画風に関する特徴を分離する方法について論じている。カメラ等を利用して撮影された画像と異なり、絵画は同一の内容（人物、物体、抽象的概念など）を異なる表現方法、いわゆる画風で表すことができる。絵画の分析では、絵画に何が描かれているかに注目することがあり、内容と画風の分離が前処理として必要となる。画像の特徴量空間でこの分離を実現しようとする場合、絵画の内容と画風に関するアノテーションが必要となるが、これには多くのコストがかかる。そこで、本学位論文では画像生成モデルを利用して絵画内容と画風を指定することで絵画風の画像を生成し、それらを利用して内容と画風を分離するためのモデルを学習するアプローチを提案している。提案手法の既存手法に対する優位性が実験により定量的に示されるとともに、提案手法により得られた絵画内容と画風のそれぞれの特徴量が、実際に絵画内容と画風に対応することが定性的に示されている。

第3章では、画像生成モデルが持つ社会的バイアスの問題について、よく利用されるLatent Diffusionモデル（や類似の構造を持つモデル）のためのバイアスの定量化手法を議論している。本学位論文において、画像生成モデルが持つバイアスとは性別を含意しない単語（Personなど）によって生成された画像中の人物の属性や、その人物とともに生成される物体の統計的な偏りを指すものとしている。提案手法では、Latent Diffusion における言語エンコーダの出力として得られる言語特徴量、拡散モデルの出力として得られる潜在特徴量、さらに潜在特徴量を画像デコーダに入力して得られる画像の3つ状態に着目し、それぞれで社会的バイアスを定量評価している。加えて、画像生成では入力される自然言語テキストに記述されていない詳細が含まれることになることから、生成されたオブジェクト等の概念と自然言語テキストに含まれる単語の対応関係から、「記述の有無」と「テキスト-画像の対応の有無」の二つの観点で画像中の概念を分類し、それぞれについて定量評価し、バイアスの原因に関して考察している。

第4章では、Latent Diffusionモデルの社会的バイアス低減手法について詳述している。第3章と同様に性別を含意しない単語を含む場合を検討しており、バイアス低減のアプローチとして各属性の頻度を均等にする等が考えられる。その中でも、本学位論文では属性間の連続性を考慮し、言語として離散化された属性の間の属性をランダムにサンプリングすることによりバイアスを低減している。生成される画像中の人物がサンプリングされた属性を強く反映させるため、サンプリングされた属性の補間は拡散モデルのデノイジングプロセスで利用されている注意機構の中で行われる。実験では、第3章で提案した定量化手法を利用して提案手法を評価し、効果を客観的に確認している。

以上のとおり、本学位論文は画像生成モデルに関して、その利活用のための方法論から内在的問題である社会的バイアスの問題への取り組みについて論じており、情報科学分野のみならず社会的にも大きく寄与するものである。このことから、博士（情報科学）の学位論文として価値のあるものと認める。