HOKKAIDO UNIVERSITY

Title	A study on latent representation learning for understanding, generation, and copyright protection of artificially created works
Author(s)	蘭, 子文
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16997号
Issue Date	2026-03-25
DOI	https://doi.org/10.14943/doctoral.k16997
Doc URL	https://hdl.handle.net/2115/99841
Type	doctoral thesis
File Information	Ziwen_Lan_review.pdf, 審査の要旨 https://creativecommons.org/licenses/by/4.0/

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　氏名　蘭 子文

審査担当者　主　査　教　授　小川 貴弘
副　査　教　授　土橋 宜典
副　査　教　授　長谷山 美紀

学位論文題名

A study on latent representation learning for understanding, generation, and copyright protection of artificially created works
（人工創作物の理解・生成および著作権保護のための潜在表現学習に関する研究）

本論文は、人工創作物の理解・生成および著作権保護のための潜在表現学習に関する研究の成果をまとめたものである。

近年、生成 AI は飛躍的に発展し、イラストレーション、グラフィックデザインをはじめ多様な創作分野で高い表現力を獲得している。その結果、人工創作物は急速に普及し、プロフェッショナルからアマチュアに至るまで、幅広い創作プロセスに AI が組み込まれ始めている。しかし、AI が創作や実務に広く浸透するにつれ、以下の三つの課題が生じている。(1) 意味理解の向上、(2) 生成モデルの柔軟性とロバスト性の強化、(3) 人工創作物の安全性と保護である。これらは相互に関連し、意味理解の精度向上は信頼できる生成を支え、双方の能力は安全で倫理的なデータ利用の仕組みによって初めて成立する。これらの課題に取り組むことは、今後の AI において、信頼性と包括性を備えた枠組みを構築するために不可欠である。

これらの問題に対処するため、本論文では、意味理解の向上、制御可能かつロバストな生成、そして創作物の不正利用防止に向けた統一的基盤として、潜在表現学習による技術を構築する。潜在空間は現代の生成モデルにおいて中心的役割を果たしており、意味概念の形成・変換・分離・統合に利用される。本論文で提案する枠組みは、画素レベルの情報ではなくこの中間表現に注目することで、上記の課題に対処可能な基盤を構築する。

第一の課題は、創作コンテンツに含まれる意味構造の複雑さに起因する意味理解性能の問題である。人工的創作物には、抽象表現や階層的属性などの強い構造性を持つセマンティクスが含まれているが、従来の画像認識モデルではコンテンツに内在するマルチラベル関係や階層的な意味構造を十分に捉えられず、精度の低下を引き起こす。これに対して、潜在表現学習を導入することで、画素レベル特徴では表現が困難な高次の意味構造や文脈依存性を捉えることを可能とする。

第二の課題は、生成モデルの柔軟性とロバスト性に関するものである。実際の利用場面、特にユーザーが新しい概念を継続的にモデルへパーソナライズする場面では、テキスト―画像拡散モデルにおいて、追加された概念が既存の概念と干渉する「概念干渉」が発生する。ユーザーが複数のスタイルを作成する状況では、こうした概念干渉の影響が特に大きい。従来の継続学習手法はモデルパラメータを適応し対処するが、潜在空間における意味表現の明確な分離を維持することが難しく、新旧概念の境界が曖昧となる。そこで、学習過程において意味表現を揺さぶる敵対的摂動手法を提案する。これにより、モデル構造を変更することなく概念間の分離性を維持し、時間とともに

追加される概念に対しても柔軟かつロバストな学習を実現する。

第三の課題は、人工創作物の安全性と著作権保護に関する問題である。従来の保護技術では、意味表現を学習する仕組みに直接影響を与えることができず、十分な保護性能を有していない。これを克服するため、潜在空間に着目した保護手法を導入し、学習に用いられる画像の潜在分布を視覚的には不可視な形で操作する。この潜在レベルでの操作により、モデルの意味抽出プロセスが意図的に妨害され、拡散モデルが保護対象となるスタイルや属性を学習・再現する能力を大幅に低下させることが可能となる。これにより、創作者にとって実用的かつ効果的な保護手段を提供可能とする。

本論文の構成は以下の通りである。第 1 章では、人工創作物の理解・生成・保護において潜在表現ベースの手法の必要性について述べる。第 2 章では関連研究を整理し、既存手法の課題を明らかにする。第 3 章では、意味理解のための潜在表現学習手法を提示する。第 4 章では、拡散モデルにおける概念干渉を防ぎ、生成挙動を安定化する潜在セマンティクスに基づく生成制御手法を紹介する。第 5 章では、生成モデルが無許可の芸術的概念を学習することを防ぐための、敵対的摂動に基づく著作権保護手法を提案する。第 6 章では、本論文を総括し、今後の課題とともに本論文全体のまとめを行う。

以上を要約すると、本論文は、人工的に生成された作品を「理解し、生成し、保護する」という多面的課題の解決に向けて、潜在表現学習を活用する統一的枠組みを構築する。多様なタスクを対象とした広範な評価により、提案技術は既存手法を大きく上回る性能改善を示している。これらの成果は、革新的な創作技術の発展を支援するものである。この貢献は、情報科学分野の発展に寄与するものと認められる。したがって、本論文における著者は、北海道大学博士 (情報科学) の学位を授与される資格を有するものと認める。