M-037

LoRA を活用した個人スタイルに対応したファッション画像生成方式
A Method for Personalized Fashion Image Generation Using LoRA

王 子涵† 阿倍 博信†
Zihan Wang Hironobu Abe

1. はじめに

従来の EC サイトは，顧客一回に買物した後，その購買履歴を基づいてその人に服を勧める仕組みとなっている．ただし，新規ユーザーへの対応に限界がある．

ファッション業界の個々のユーザーの好みやファッション系統に応じた個別化のニーズに応えるため，敵対的生成ネットワーク（GAN）や視覚認識技術を活用したシステムが既に提案されてきたが，それらにはいくつかの課題が存在する．特に，既存のシステムは過去の購入履歴や閲覧データに依存しているため，新規ユーザーへの対応や，ユーザーの個々の意図に即した画像生成が難しいという制約が見られていた．以上を踏まえて,本研究では，研究目的として，新規ユーザーでも利用可能なカスタマイズ性の高いファッション画像生成システムの構築を設定した．

2. 関連研究

2.1 LoRA と Stable Diffusion

LoRA（Low-Rank Adaptation）[1]は，大規模な AI モデルを効率的に微調整する手法であり，特に画像生成モデルにおいて注目されている．Stable Diffusion[2]は，テキストから高品質な画像を生成する拡散モデルであり，オープンソースとして広く利用されている．LoRA を Stable Diffusion に適用することで，モデル全体を再訓練することなく，特定のスタイルや特徴を持つ画像生成が可能となる．この手法は，計算コストやメモリ消費を抑えつつ，高精度な学習を実現するため，個別のファッションスタイルに対応した画像生成において有効であると考えられる．

2.2 DeepFashion

DeepFashion[3]は，ファッション画像解析のための大規模なデータセットであり，80 万枚以上の多様な衣類画像を含んでいる．各画像には，50 のカテゴリ，1,000 の属性，バウンディングボックス,ランドマーク情報などの豊富なアノテーションが付与されている．このデータセットは，衣類の検出，認識,画像検索などのコンピュータビジョンタスクにおけるトレーニングおよび評価に広く利用されている．特に，消費者から店舗への衣類検索や，店内の衣類検索，ランドマーク検出などのベンチマークが開発されており，ファッション関連の研究において重要なリソースとなっている．

3. 提案手法

本研究では，Stable Diffusion をベースモデルとして採用し，そのモデルに LoRA を導入することで，個人ユーザーのファッションスタイルを忠実に再現した画像生成を目指した．具体的には，Realistic Vision V6[4]モデルを用いて，個人の服装画像を用いたモデルのファインチューニングを行った．さらに，生成された画像が実際の市場で提供されている衣服とどの程度類似するかを評価するため，DeepFashion データベースを活用した画像検索システムを導入し，生成画像とデータベース内の類似画像との比較を行った．

4. 画像生成と評価実験

実験者が所有する 10 枚の上着の写真（うち，ジャケットの写真 6 枚）を用いて，Realistic Vision V6 モデルに LoRA を適用した．チューニング後のモデルに対して，ユーザーの希望を反映したプロンプトを与え，画像を生成した．今回の実験では，「jacket」などのプロンプトを用いて，複数回の画像生成を行い，その中から実験者のファッションスタイルに最も合った画像 2 枚を評価実験に使用する．

今回の実験では，生成されたファッション画像がユーザーのファッションスタイルにどの程度適合しているかを検証するため，定量評価と定性評価の二つの手法を組み合わせた評価実験を実施した．定量評価では，生成画像とユーザーの私服画像とのコサイン類似度を算出し，画像の近似度を数値的に測定した.一方，定性評価では，DeepFashion データセット全体を対象とした画像検索システムを構築し，生成画像に最も視覚的に近い 5 枚の画像を検索することで，生成画像の視覚的一貫性を評価した．

4.1 定量評価：コサイン類似度による類似度測定

定量評価では，LoRA によりファインチューニングされた生成画像と，ユーザーが実際に所有する 10 枚の私服画像との間で，画像の類似度をコサイン類似度を用いて算出した．全画像に対して事前に MobileNetV3 による特徴抽出処理を行い，それぞれの画像を 128 次元の特徴ベクトルに変換した．

その後，生成画像と 10 枚の私服画像との間で 1 対 1 の類似度スコアを計算し，類似度が高かった画像を特定した．

図 1　類似度が高いマッチング例(1)

† 東京電機大学　Tokyo Denki University

図 2　類似度が高いマッチング(2)

評価の結果，複数の生成画像においてコサイン類似度が 0.7〜0.8 程度と高いスコアを記録し，質感や色調の再現性において，生成画像がユーザーの私服スタイルをよく反映していることが確認された．図 1 と図 2 では，左は生成された画像で，右が類似度の高い私服画像である.図 1 のコサイン類似度は 0.806，図 2 のコサイン類似度は 0.842 である.

4.2 定性評価：DeepFashion 全体を対象とした検索エンジンによる類似画像の抽出

定性評価では，生成画像が実在するファッションアイテムにどの程度視覚的に近いかを評価するため，DeepFashion 全体を対象とする画像検索エンジンを用いた．これは，あらかじめ DeepFashion の全画像を MobileNetV3[5]を使ってベクトル化した情報を,検索エンジンである Elasticsearch[6]上に登録し，特徴ベクトル空間上で高速検索を可能にしたものである.

今回の実験では，画像特徴量の抽出にあたり，PyTorch[7]の TorchVision[8]ライブラリで提供されている事前学習済みモデルである MobileNetV3-Large を使用した．生成されたファッション画像をクエリ画像として検索エンジンに入力し，DeepFashion 内から最も類似する上位 5 枚を抽出した．検索結果には，色・質感・スタイルにおいて高い視覚的類似性を持つ画像が含まれているが，生成された画像と類似性を持たない画像も含まれている.

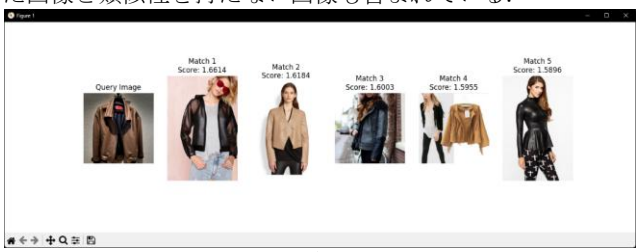

図 3　検索エンジンの検索結果例(1)

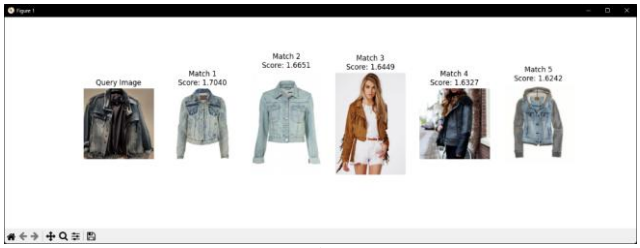

図 4　検索エンジンの検索結果例(2)

図 3 と図 4 は，生成された 2 枚の画像それぞれに対する検索結果である．図 4 では，検索結果と生成画像が類似していることが分かるが，図 3 では，検索結果と生成画像があまり類似していないことが分かる．背景処理の違いやポーズのズレにより，意図とは異なる衣服種が上位に含まれるケースも見られた．これは，視覚的意味に対するベクトル表現の限界を示すものである.

5. 考察

評価実験を通じて，LoRA による少量データでのモデルチューニングが，ユーザーのファッションスタイルに即した画像生成に有効であることが確認された．特に質感や色調の再現には一定の成果が見られた一方で，ポーズや背景の差異に起因する視覚的一貫性の乱れや，意図しない衣服カテゴリの出現といった課題も顕在化した．これらは，視覚特徴の抽出・表現の限界や，検索・比較アルゴリズムの精度に起因する可能性がある．また，生成結果が特定のサンプルに依存する傾向もあり，スタイルの汎化にはさらなる工夫が必要といえる．今後は，ランドマーク情報や衣服構造の特徴量の導入，ユーザーとの対話的フィードバックを活用した生成の最適化などが，より高精度なカスタマイズに寄与すると考えられる.

6. まとめ

本研究では，Stable Diffusion に LoRA を適用することで，新規ユーザーにも対応可能なファッション画像生成システムを構築した．ユーザーの私服画像 10 枚を用いたモデルのファインチューニングにより，個々のスタイルを反映した画像の生成が可能となった．評価では，生成画像と私服画像のコサイン類似度が高く，定性評価でも視覚的な一致が確認された．一方で，一部の画像では意図しない衣服種が生成されるなどの課題も見られた．今後は，データの多様化や特徴量の追加によって，さらなる再現性と安定性の向上が期待される.

参考文献

[1] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, “LoRA: Low-Rank Adaptation of Large Language Models”, arXiv:2106.09685 (2021).
[2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models”, arXiv:2112.10752 (2022).
[3] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, Xiaoou Tang, “DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations”, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1096-1104 (2016).
[4] SG 161222, “Realistic Vision V6.0 B1”, Stable Diffusion model card (2023).
[5] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, Hartwig Adam, “Searching for MobileNetV3”, arXiv:1905.02244 (2019).
[6] Elastic, “Elasticsearch”, https://www.elastic.co/jp/elasticsearch
[7] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, Soumith Chintala, “PyTorch: An Imperative Style, High-Performance Deep Learning Library”, arXiv:1912.01703 (2019).
[8] TorchVision maintainers and contributors, “TorchVision: PyTorch’s Computer Vision Library”, GitHub repository (2016).