

顔認識システムにおけるプロンプトベース画像符号化と検出器の組み合わせによる精度評価

Evaluating the effectiveness of face impersonation detection for prompt-generated images

藤浪 優里佳

Yurika Fujinami

渡辺 裕

Hiroshi Watanabe

早稲田大学基幹理工学部情報通信学科

School of Fundamental and Science and
Engineering, Waseda University

Abstract: 近年、顔認識システムはさまざまな場面で応用されている。また、拡散モデルを用いた高品質な顔画像生成も進展している。本研究では、プロンプトベースの画像符号化により生成された顔画像が、顔認識システムでどのように真偽判定されるかを評価した。Stable Diffusion を用いて生成された顔画像について、複数のアルゴリズムによる類似度測定を行い、顔認識モデルと検出器の組み合わせが判定精度に与える影響を検討した。結果として、組み合わせより認識精度が変動し、認識精度の安定性にはモデルと検出器の選択が重要であることが示唆された。また、Anti-Spoofing の技術が生成画像の判別に有効ではなく、新たなトレーニングが必要であることも確認した。

1 はじめに

近年、拡散モデルを用いた画像生成の研究が急速に進んでいる。テキストから画像、あるいはプロンプトから画像への変換が主な処理である。中でも顔画像の生成については、多くのモデルが研究され、高品質な顔画像のバリエーションが得られている。特に、画像からプロンプト、エッジ、色情報を抽出するプロンプトベースの画像符号化方式は、顔画像の再現性に優れている。

CNN を用いた顔認識システムは高い認識精度を達成しており、社会インフラとして様々な応用が提案されている。実用化段階では、外部からの攻撃に対して脆弱であってはならない[1]。そのため、顔認識システムが、拡散モデルに基づく画像生成手法によって生成された顔画像を認識する際にどのような性質を持つかが重要な検討課題である。

本研究では、顔画像認識システムのモデルと検出器の組み合わせを変え、プロンプトベースの画像符号化により得られた顔画像が、どの程度の類似度で真偽判定されるかを検討する。

Adversarial Level of Face Images Generated by Prompt-based Image Coding in Face Recognition System[2]では、生成された画像と元画像の類似度を DeepFace によって計測している。認証には VGG-Face、検出には OpenCV を用いている。生成画像は Stable Diffusion を用いて、元画像から品質の異なる 5 つを生成している。ControlNet によって色と輪郭を指定し、元画像に近いものが生成される。

エッジ量が多いほど類似度も高くなる傾向があるが、真偽判定の精度は低く、顔画像認識のモデルと検出器について、再検討の余地がある。

3 顔画像認識モデルと検出器

顔画像認識モデルは、検出、アライメント、表現、検証の 4 段階から構成されている[3]。まず初めに検出器を用いて、画像中の顔領域以外の領域を破棄する。次に、目の位置に 2 つの角を持つ直角三角形を構成するアライメントを実行する。これによって、目の距離を測ることができ、cos 類似度が測定可能である。

4 実験および結果

2 関連研究

表 1: 認識アルゴリズムによる Dissimilarity 値の違い

Model	Threshold	Dissimilarity
VGG-Face	0.68	0.77963
FaceNet	0.40	0.46237
FaceNet512	0.30	0.53682
Dlib	0.070	0.054545
ArcFace	0.68	0.59263
SFace	0.59	0.63446
GhostFaceNet	0.65	0.66363
DeepID	0.015	0.0064626
OpenFace	0.10	0.15491

表 2: 顔検出アルゴリズムによる Dissimilarity 値の違い(Threshold=0.68)

Detector	Dissimilarity
OpenCV	0.77964
MtCnn	0.74739
FastMtCnn	0.75953
RetinaFace	0.74125
YuNet	0.74307
CenterFace	0.75730
Yolov8	0.75677
Mediapipe	0.68774
Dlib	0.65440
SSD	0.73713

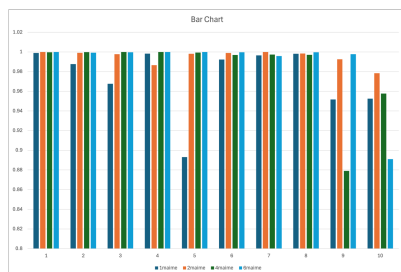


図 1: Face Anti Spoofing による判定結果

顔画像認識システムのモデルと検出器の性能を評価する。使用するのは DeepFace モデルである。画像は Face Research Lab London Set から無作為に選んだ 1 枚を用いる[4]。

まず Stable Diffusion を用いて、元画像からフェイク画像を生成する。画像生成の際には、エッジ抽出に Canny エッジ検出器を用いた[5]。

生成された画像と元画像の類似度は DeepFace を用いて計測される。顔検出アルゴリズムに OpenCV を固定で使い、認証アルゴリズムを変化させた。また、同様に顔認証アルゴリズムを VGG-Face に固定し、検出

アルゴリズムを変化させた。それぞれ結果を表 1、2 に示す。非類似度は(1.0 - cosine similarity)で表され、値が 0 に近いほど画像が類似していることを示す。

Face Anti-Spoofing を用いて顔を認識させた場合には、図 1 のように数値にばらつきが見られた。

5 まとめ

本稿では、顔画像認識システムのモデルと検出器の適切な組み合わせを検討した。実験により、それらを変化させると、判定にぶれが生じていることを確認した。したがって生成された画像を正しく判断したい場合、DeepFace の設定は、デフォルトのもの以外も検討する必要がある。また、Anti-Spoofing は Stable Diffusion によって生成された画像を判別することはできず、別のトレーニングが必要である。

参考文献

- [1] F. Vakhshiteh, A. Nickabadi, R. Ramchandra, “Adversarial Attacks against Face Recognition: A Comprehensive Study,” arXiv: 2007.11709v3, Feb. 2021.
- [2] Y.Fujinami, H.Watanabe, “Adversarial Level of Face Images Generated by Prompt-based Image Coding in Face Recognition System”, GCCE, Oct. 2024
- [3] Sefik Serengil, Alper Özpınar, “A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules”, Journal of Information Technologies, Volume 17 Issue2, P95-107, Apr. 2024
- [4] Lisa DeBruine, Benedict Jones, Face Research Lab Londonset, https://figshare.com/articles/dataset/Face_Research_Lab_London_Set/5047666
- [5] H. Watanabe, L Jin, T. Hayami, T. Chujoh, Y. Yasugi, S. Hong, Z. Fan, T. Ikai, “Prompt-based Image Coding with Edge and Color Information,” IEICE General Conference, D-11A-27, Mar. 2024.