画像圧縮耐性をもつAI生成画像の識別

Identification of AI-Generated Images with Image Compression Resistance

大沢　美優
指導教員　青木　輝勝

東京工科大学　コンピュータサイエンス学部　コンピュータサイエンス学科
コンピュータビジョン研究室

現在、世界中で生成系AIを使用したフェイクニュースが増えている。そのため、実写画像とAI生成画像の正確な識別が重要である。また、実際に識別したい画像は主にSNS上にあがっており、画像圧縮がかかっていることが多い。このことから、本研究は画像圧縮耐性のある識別器作成を目的としている。

AI生成画像識別，深層学習，ウェーブレット変換，JPEG圧縮

１．はじめに

昨年、アメリカの国防総省付近で爆発が起きたとされる偽画像がSNS(Social Networking Service)上で拡散され、実際にアメリカの株価にまで影響を与えた事件があった[1]。

現在、生成系AIの技術の進化により、本物の画像(写真)と全く見分けのつかない画像を簡単に生成することが可能になっている。これにより、上記のような事件が世界中で発生している。そのため、私たちが悪意のあるフェイクニュースに騙されないために、実写画像とAI生成画像を正確に識別することが重要である。

実際に識別したい画像は主にSNS上にあがっており、これには画像圧縮がかかっていることが多い。しかし、既存の識別手法のほとんどは圧縮前の画像に対する識別精度は高いが、圧縮後の画像に対しては低い結果となってしまっている。

この問題を解決するために、本研究ではウェーブレット変換を用いることで、圧縮耐性のある識別器を作成することを目標としている。また、画像圧縮技術には多くの種類があるが，それらの原理は類似しているため今回は最も普及しているJPEG圧縮を対象として研究を進めていく。

２．既存研究

本章では、JPEG圧縮の特性、AI生成画像識別に関する既存研究とその問題点について述べる。既存研究には、Tong QiaoらのCSC-Net[2]やMingjian ZhuらのGenDet[3]など多くの識別手法があるが、本研究では多くの生成手法で生成された画像に対して実験を行っており、かつ精度が高いPatchCraft[4]に着目する。

２．１．JPEG圧縮の特性

現在、JPEG圧縮技術は広く利用されていて、インターネット上のほとんどの画像は圧縮をして表示されている。例えば、携帯電話で写真を撮った場合、写真は初めから圧縮された状態で保存するように設定されている。

この圧縮技術は、人間の目ではあまり認識でない細かい情報を削除することで、画像品質に大きな影響を与えることなく、ファイルサイズを小さくしている。そのため、画像の低周波成分は保持され、高周波成分は削除されるという特性がある。

２．２．PatchCraftによるAI生成画像識別

Nan Zhongらはテクスチャの複雑さに着目し、それぞれのピクセル間相関の違いを分析することでAI生成画像を識別する手法を提案した。この手法では、入力画像を複数のパッチに分割し、テクスチャが豊富な領域と乏しい領域に分けて特徴を抽出することで、識別の際に画像の高周波成分にだけ

着目するようにしている。これにより，従来困難であった学習時に用いなかった生成手法で作られた偽画像にも対応できる。

しかし、識別をする際に高周波成分だけを見ているため、入力画像に圧縮などの後処理がかかっている場合、識別精度が大幅に下がってしまうという問題がある。この問題はほとんどの既存手法が抱えている問題であり、この問題を根本的に解決できる手法は私の知る限り存在しない。

３．提案概要

本研究では、2.2 節で述べた問題を解決するために、ウェーブレット変換を用いた手法を提案する。

ウェーブレット変換とは、画像や音声などの信号を周波数ごとに分解し、異なる周波数帯でその特徴を捉える手法である。これを画像に適応させると、画像として主要な情報を含んでいる低周波成分だけで構成された LL とエッジや細かい表現を捉えることができる高周波成分を含んだ HL、LH、HH の４つの成分に分解することができる。ウェーブレット変換後の画像を図 1 に示す。

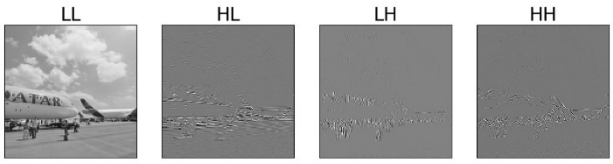

図１　ウェーブレット変換後の画像

処理の流れは、まず入力画像をウェーブレット変換し、上記の４つに分解する。その後、それぞれを ResNet50 に与え、出力結果を合わせることで実写画像か AI 生成画像であるかを識別する。ウェーブレット変換をして出てくる４つの画像をすべて使うのではなく、組み合わせることで識別精度を上げる。特に、LL は JPEG 圧縮のときに保持される低周波成分のみを含んでいるため、これを利用することで圧縮耐性のある識別器になると考える。

４．実験

今回は実験として、ウェーブレット変換が JPEG 圧縮された画像の識別に有効であるかを調べた。使用したウェーブレット変換後の画像は HL、LH、HH の３つであり、それぞれを ResNet50 に入力し、最終的に出力結果を合わせて識別を行った。また、比較対象は 2.2 節で述べた既存研究である PatchCraft である。

実験結果を表 1 に示す。全体的な識別精度は低いが、CycleGAN や ProGAN、StarGAN は PatchCraf よりも精度が高くなった。これより、ウェーブレット変換は JPEG 圧縮された画像の識別に有効であると言える。

	CycleGAN	ProGAN	StarGAN	StyleGAN	StyleGAN2
PatchCraft	48.0%	49.8%	49.1%	**49.9%**	**49.3%**
wavelet	**48.2%**	**50.4%**	**49.9%**	48.5%	48.7%

表 1　JPEG 圧縮画像に対する識別精度

５．おわりに

本研究では、ウェーブレット変換を用いることで、JPEG 圧縮に耐性のある AI 生成画像識別器の作成を目的としている。その結果、圧縮されていない画像で高い識別精度がでるだけではなく、高い圧縮率の画像に対しても正確な識別ができるようになることを目指す。Stable Diffusion や MidJourney などの生成系 AI が普及し、SNS の利用が一般的になっている現在において、悪意のある AI 生成画像は簡単に世の中に広がってしまう。これを防ぐために、本研究のような JPEG 圧縮された画像に対する識別器の作成というテーマは社会にとって意義の高いことである。

参考文献

[1] テレ朝 news，”「国防総省近くで爆発」“偽画像”で混乱　AI 画像が金融市場を動かした初の例か”，https://news.tv-asahi.co.jp/news_international/articles/000300405.html, (2024 年 10 月 29 日閲覧)
[2] Qiao, Tong, et al. "Csc-net: Cross-color spatial co-occurrence matrix network for detecting synthesized fake images." IEEE Transactions on Cognitive and Developmental Systems 16.1 (2023): 369-379.
[3] Zhu, Mingjian, et al. "Gendet: Towards good generalizations for ai-generated image detection." arXiv preprint arXiv:2312.08880 (2023).
[4] Zhong, Nan, et al. "PatchCraft: Exploring Texture Patch for Efficient AI-generated Image Detection." arXiv preprint arXiv:2311.12397 (2024).