コンテンツ保護特論2月7日締切課題

43M22403　古谷康平

**フォレンジクス技術の研究動向**

トウェンテ大学のクインテン・リファーゲン（Quinten Riphagen）氏は、オランダ法医学研究所でディープフェイクに関するプロジェクトに携わっていた。クインテン・リファーゲン氏による文献レビューから紹介する。次の分野は歩行解析に基づくバイオメトリクス、そしてこれからの分野はローリングシャッター効果によるENF(EndNote Filter Format)ビデオで、今後はPRNU(Photo-response Non-uniformity、光に対するカメラの応答の均一性を表す)でレビューを終える。

ディープフェイクは、一般的にGenerative Adversarial Network (以下GAN)を使用して作成されるネットワークである[16]。[16]は現在ほとんどのDeepfake 技術が使用しているGANの最初の数学的モデルを提供している。これらのネットワークは、画像、ビデオ、オーディオなどのほぼ検出不可能な生成メディアを作成することが可能である。これらのネットワークが検出不可能なメディアを生成する方法は，生成モデルと，与えられたサンプルが学習モデルによるものか生成モデルによるものかを判断する敵対的な判別モデルとを戦わせることである。生成モデルは、学習モデルから学習する。判別モデルから、どのサンプルが検出を通過し、どのサンプルが検出されたかを学習する。次の繰り返しでは、検出されなかったサンプルの特徴を保持し、拡張していく。このプロセスを何世代も繰り返し、あるサンプルが本物か生成されたものかを識別モデルが正確に予測できなくなるまで続ける。ネットワーク構造の一例をFig.1に示す[5]。

ダイアグラム

自動的に生成された説明

ディープフェイクの検出の問題は、二項分類問題として捉えられている。動画は本物か偽物かのどちらかに分類される。分類器には入力データを本物か偽物か判別するための特徴が必要であるため、深層学習手法がよく使われる。法医学では、ある映像の分類を説明する必要があるため、このような二項対立的な方法で問題に取り組むことはできない。証拠の真偽について評決を下すのではなく、尤度比やモデルからの分析を用いてそのビデオが本物である可能性を説明する。そして、その証拠が本物であるかどうかの判断は、裁判官の判断に委ねられる。完全な検出パイプラインは、多くの場合、 ビデオフレームから顔を抽出し、その顔をディープニューラルネットワークにかけ特徴を抽出し、ディープニューラルネットワークを学習させる。次にバイナリ分類器を使用して抽出された特徴に基づいてフレームを分類する。ネットワーク構造の一例をFig.2に示す[9]。

ダイアグラム

自動的に生成された説明

**操作分類ネットワーク(MCNet) (デジタルフォレンジクス技術の紹介)**

[53]は、画像に対して行われた様々な種類の操作を分類する手法を提案している。操作分類ネットワーク(MCNet)は、周波数、空間、圧縮から複数の特徴を抽出するマルチドメイン特徴量を用いる、ディープフェイク検出モデルである[53]。このネットワークは、マルチストリーム構造によって実現されており、以下のことが可能である。複数のドメインの融合した特徴を分析することによって、画像操作を検出することができる。また、画像操作を分類することができ、圧縮と操作の両方を考慮することができる。 [53]では、一般的に使用される画像操作として、画像ぼかし、ノイズ付加、コントラスト変更、画像モーフィング、画像リサンプリングを行う。このネットワークは、視覚的アーティファクトネットワーク（VANet）と圧縮アーティファクトネットワーク（CANet）の2種類の学習ネットワークで構成される。これらのネットワークから学習された特徴量は、転送学習によりアンサンブル（ニューラル）ネットワークに転送され、単一の分類器に融合される。

**出典**

Zeno Geradts, Quinten Riphagen, Interpol review of forensic video analysis, 2019–2022.

[16]Ian Goodfellow, et al., Generative adversarial nets, in: Advances in Neural Information Processing Systems 27, 2014.

[5] Jason Brownlee, A Gentle Introduction to Generative Adversarial Net- Works (GANs)

[9] Polychronis Charitidis, et al., Investigating the Impact of Preprocessing and Prediction Aggregation on the DeepFake Detection Task”, 2020.

[53] I.-J. Yu, et al., Manipulation classification for JPEG images using multi- domain features, in: IEEE Access 8, 2020.