生成AI時代におけるDeepFakeのリアルタイム検出基盤の研究

Research into real-time deepfake detection infrastructure in the era of generative AI

清水優・ネットワーク分科会・情報セキュリティ大学院大学

研究背景

生成AIの進歩により、高品質なディープフェイクが容易に作成され、偽動画・偽画像がオンライン上で急速に拡散する問題が深刻化している。特に拡散速度の速い現代のオンライン環境では、フェイクが短時間で広範囲に影響を及ぼすため、拡散後に対応する従来の手法では被害を十分に抑えられない。一方で、既存の検出技術は高精度であるものの、処理負荷や実環境への適応性に課題があり、リアルタイムに動作可能な軽量・高速な検出手法の必要性が高まっている。

目的

本研究の目的は、オンライン上で拡散されるディープフェイクによる被害を最小化するために偽動画・偽画像を早期に検出できる技術を確立することである。近年増加している個人を対象とした性的ディープフェイクや誤解・風評被害を引き起こす偽コンテンツに対し、拡散前で迅速に検知し、被害の抑制や問題解消に寄与することを目指す。

これまでの取り組み

- 論文精読による現状の技術動向と課題の把握、基礎的な背景などを理解
- 研究テーマの方向性を整理し、主要な問題領域を特定
- 既存手法の調査を行い、自身のアイディアと重複がないかを調査

~3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月
研究方向性の確定	・関連研究の整理 ・輪講発表		・環境構築 ・実験手法の提案とベースライン実験 ・中間発表				・内容の改善と修正 ・提案手法の試作実装 ・本格的な実験、評価 ・修論準備				修士発表