S09　電子透かしによるデータ保護の試み

林昌純（帝京平成大学）

研究概要

オリジナル画像で編集履歴の簡易な判断を行う

背景

画像生成AIで作成されたものか判断したい
→AIでの作成を示す「電子透かし」の動き
　例えば，ロゴマークが入ることで確認が可能

[1] 編集履歴を示すために取り組む標準化団体
Coalition for Content Provenance and. Authenticity
（略称 C2PA）

目的

①オリジナル画像にも編集履歴を判断したい

②数字やアルファベットについて
　スマートフォンでテキスト認識させたい

問題①

業界によっては以前から類似のシステムは存在

[2] 履歴を残すラボラトリー情報管理システム
LIMS（Laboratory Information Management System）

問題②

・スマートフォンで二次元コード認識への攻撃

[3] QRコードへレーザー照射による誤認識の報告

・画像認識への攻撃

[4] 生成AIによる誤認識へ誘導するノイズの報告

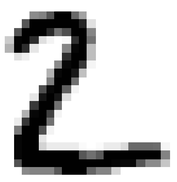
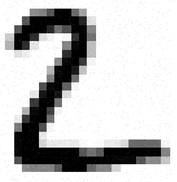
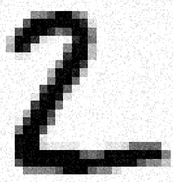
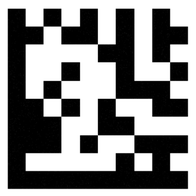
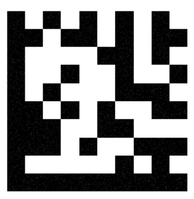
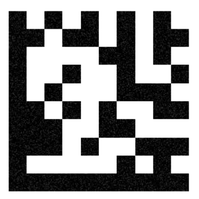
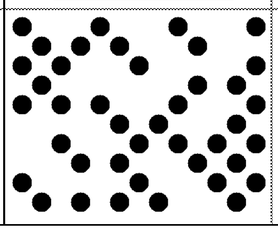
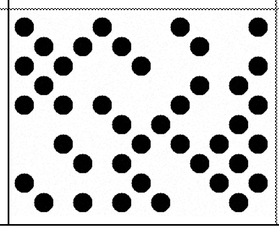
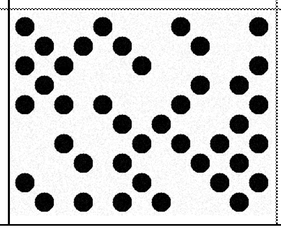
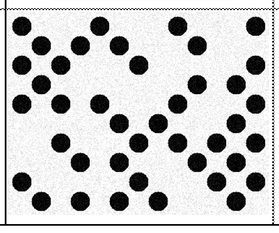
Adversarial Robustness Toolbox （ART）による例
手書き文字（Mnisit）

画像からのテキスト認識
（iOS 17）

左：○ 可能　　右：× 不可

現状報告　誤認識とは無関係なノイズ付加ではテキスト認識に影響低い

[5] 画像からのテキスト認識（iOS 17）[6] アプリケーションからの認識（ iOS 17 ）

	元の画像		元の画像 ＋25%ノイズ		元の画像 ＋50%ノイズ		元の画像 ＋75%ノイズ	
手書き文字 （Mnisit）		※1 ○ 可能		※1 ○ 可能		※1 ○ 可能		※1 ○ 可能
印刷文字 （Sansan）		※1 ○ 可能		※1 ○ 可能		※1 ○ 可能		※1 ○ 可能
DataMatrix		※2 ○ 可能		※2 ○ 可能		※2 ○ 可能		※2 ○ 可能
DotCode		※2 ○ 可能		※2 ○ 可能		※2 ○ 可能		※2 ○ 可能

提案・課題

・目には見えない「電子透かし（Digimarc Barcode ）」や「コード（スクリーンコード）」の可能性

・手元のスマートフォンを用いた簡易的な編集履歴の有無だけでもチェックできる仕組みの検討など

本研究では，国立情報学研究所のIDRデータセット提供サービスによりSansan株式会社様から提供を受けた「Sansanデータセット」を利用しました.