Entwicklung einer modularen Analysepipeline zur forensischen Untersuchung digitaler Dokumente

Niklas Weiand

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Motivation

In forensischen Untersuchungen werden oft Dokumente wie Rechnungen und Urkunden untersucht, um Straftaten im Bereich der Wirstschaftskriminalität aufzudecken. Da bei einer solchen Untersuchung oft tausende von Dokumenten manuell gesichtet werden müssen, kann eine Methode zur Erkennung von potenziellen Fälschungen und Anomalien die Arbeit von Forensikern erleichtern. Durch die fortschreitende Entwicklung von Technologien, die Bildmanipulationen erlauben, wird es auch immer einfacher, Dokumente zu fälschen. Hinsichtlich dieses Problems wurde in dieser Arbeit eine Vielzahl von Methoden der Computer Vision genutzt, um über visuelle Merkmale Fälschungen zu erkennen.

Ziel der Arbeit

Ziel dieser Masterarbeit ist die Entwicklung einer modularen Analyse-Pipeline zur automatisierten Erkennung visueller Manipulationen in digitalen Dokumenten. Die Pipeline soll forensische Fachkräfte bei der systematischen Bewertung großer Dokumentenmengen unterstützen und insbesondere bei der Markierung auffälliger Bereiche zum Einsatz kommen.

Im Fokus steht dabei nicht die vollständige Automatisierung einer Fälschungserkennung, sondern die Schaffung eines nachvollziehbaren, visuell interpretierbaren Werkzeugs, das Hinweise auf Manipulationen liefert und die manuelle Analyse gezielt leitet. Die praktische Umsetzung sollte durch die Implementierung großteils voneinander unabhängiger Module, die verschiedene Arten von Fälschungen erkennen sollten, erfolgen.

Daten

Ein zentrales Problem dieser Arbeit war ein Mangel an geeigneten Daten. Ein eigener Datensatz musste erstellt werden, um eine Grundlage zur Erstellung der Pipeline zu schaffen. Bei diesem Datensatz wurden Imitationsfälschungen und Copy-Move Fälschungen genutzt. Bei ersteren werden digital neue Informationen und Zeichen eingefügt. Bei Copy-Move Fälschungen werden Inhalte von anderer Stelle, aus demselben oder einem anderen Dokument, kopiert und an den gewünschten Ort eingefügt. Zusätzlich wurde durch das Modell TextCrtl [4] weitere Fälschungen mittels Deep Learning-Modellen durchgeführt und als zusätzliche Testdaten eingebaut.

Die Qualität der verschiedenen Fälschungen wurde zusätzlich betrachtet, sodass zwischen guten und schlechten Fälschungen unterschieden werden kann. Diese verschiedenen Einteilungen erlaubten ein strukturiertes Vorgehen bei der Analyse des Problems und der Entwicklung von Methoden, um verschiedene Manipulationen zu erkennen.

Überblick über die Pipeline

Die Pipeline besteht aus mehreren eigenständigen Modulen, die jeweils unterschiedliche Aspekte potenzieller Fälschungen untersuchen. Die Module sind kombinierbar und bilden gemeinsam ein robustes System zur Detektion visueller Anomalien. Module innerhalb der Pipeline sind die folgenden:

1. Vorsortierung mittels Feature-Vektoren auf Basis neuronaler Netze (ViT, ResNet, VGG)
2. Nutzung der durch eine Binarisierung hervorgehobenen visuellen Artefakte
3. Erkennung falscher horizontaler Ausrichtung
4. Fourier-Analyse zur Detektion von Frequenzartefakten
5. Anomalieerkennung mittels Local Binary Patterns und Varianz von Pixelumgebungen

Die entwickelte Pipeline basiert auf einem modularen Konzept, bei dem verschiedene Analyseverfahren kombiniert werden können. Ziel ist es, Dokumente sowohl einzeln als auch im Vergleich auf visuelle Auffälligkeiten zu untersuchen.

Jeder Bestandteil der Pipeline deckt unterschiedliche Merkmalsbereiche ab, wobei alle Module unabhängig voneinander ausführbar und erklärbar sind. Die Pipeline wurde iterativ aufgebaut, sodass zunächst schlechte und damit offensichtliche Fälschungen betrachtet wurden. Nach der ersten Implementierung einzelner Module folgte eine systematische Überprüfung ihrer Wirksamkeit mittels sich zu der jeweiligen Methode eignenden Tests. Somit konnten Stärken und Schwächen der jeweiligen Methoden gezielt analysiert werden und insbesondere auf Grundlage der nicht erkannten Fälschungen weitere Methoden erarbeitet werden.

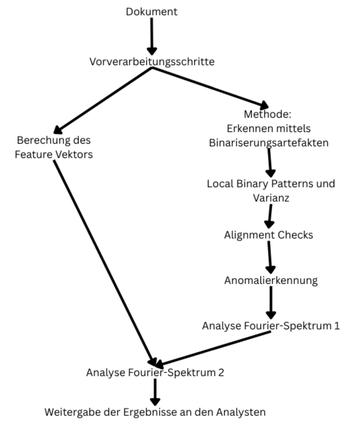

Figure 1. Die Pipeline in ihrer Anwendung mit der Reihenfolge der verschiedenen Module.

Ergebnisse der Experimente

Die Feature-Vektor-Vorsortierung basiert auf der Extraktion visueller Repräsentationen mittels vortrainierter neuronaler Netze. Jedes Dokument kam dabei in mindestens drei verschiedenen Resolutionen vor, zusätzlich wurden einige andere gefälscht. Alle Dokumente des Datensatzes wurden miteinander mittels der Kosinus-Ähnlichkeit verglichen, zur absoluten Einschätzung wurden final die durchschnittlichen Ähnlichkeits-Werte von gleichen Dokumenten und paarweise unterschiedlichen Dokumenten subtrahiert. Hierbei konnten mit allen genutzten Modellen eine Differenzierung zwischen paarweise verschiedenen Dokumenten festgestellt werden. Der festgestellt Wert der Differenz liegt zwischen 0,135 bei ResNet50 und 0,370 bei DenseNet121. Allgemein konnten kleinere und weniger komplexe Netze eine klarere Trennung zwischen den Dokumentenklassen durchführen.

Ein weiteres Modul analysierte die Artefakte, die nach einer Binarisierung zu erkennen sind. Zunächst wurden manuell Features definiert, hierbei konnte jedoch keine großen Erfolge erzielt werden. Als weiterer Ansatz wurden daraufhin verschiedene neuronale Netze trainiert, die zwischen digitalen Buchstaben und gescannten Buchstaben differenzieren sollen. Mit ViT [1] konnte ein Netz geschaffen werden, in dem die True-Positive Rate bei circa 90% liegt, während die Precision bei 31% liegt. Da meist nur ein kleiner Teil eines Dokuments gefälscht ist, bietet diese Methode einen guten Anhaltspunkt zur weiteren Betrachtung eines Dokuments. Ein AUC-Wert der berechnetet ROC-Kurve von 0,99 deutet zusätzlich eine hohe Modellgüte an.

Bei der Analyse der Textausrichtung wurde das Dokument zunächst mittels EasyOCR [2] in Abschnitte und Zeilen geteilt. Daraufhin wurden innerhalb jeden Bereichs betrachtet, ob die Ausrichtung einzelner Zeichen signifikant abweicht, wobei auf den Textsatz geachtet wurde, also Buchstaben wie „g" speziell betrachtet wurden. Hierbei konnten einzelne Beispiele erkannt werden, jedoch waren für eine alleinstehende Nutzung und Einordnung zu viele False-Positive Ergebnisse vorhanden.
Weiterhin wurde das Frequenzspektrum sowohl einzelner Teile als auch des gesamten Dokuments erkannt. Es konnten Muster nach Imitationsfälschungen und KI-unterstütztem Fälschungen erkannt werden, jedoch führte die Suche einer allgemeingültigen Trennungslinie nicht zu einem eindeutigen Ergebnis, weswegen die Frequenzspektren insbesondere als unterstützende Mittel eingesetzt werden.
Ein weiteres Modul nutzte Local Binary Patterns in Verbindung mit Varianzanalysen um homogene und retuschierte Bildregionen zu identifizieren. Hierbei sollen unnatürlich glatte oder zu raue Stellen, insbesondere im Hintergrund des Dokuments, erkannt werden. Bei Tests wurde festgestellt, dass bei nur circa 1,3% der gescannten Dokumente solche glatten Regionen vorkommen. Im Gegensatz hierzu wurden in einem anderen Test gezeigt, dass 100% der modifizierten Regionen erkannt werden konnten.

Diskussion und Fazit

Die Ergebnisse der entwickelten Module zeigten, dass sich Dokumentenfälschungen vor allem bei Imitationsfälschungen erkennen lassen, während Copy-Move Fälschungen mit den aktuellen Methoden nicht zuverlässig erkannt werden können. Weiterhin konnte erkannt werden, dass einzelne Module unterschiedlich sensitiv auf die Qualität der Fälschung reagieren. Methoden wie die Überprüfung der Zeilenausrichtung können bei Copy-Move Fälschungen anschlagen, zeigen jedoch noch hohe False-Positive Raten.

Mit der entwickelten Pipeline, wurde gezeigt, dass eine modulare, visuelle Analyse digitaler Dokumente eine wirksame Unterstützung zur manuellen Analyse sein kann, jedoch diese aktuell nicht ersetzen kann. Die Vorsortierung und Bereitstellung der Dokumente mit der höchsten Wahrscheinlichkeit einer Manipulation, sowie die Unterstützung der Analyse mittels der bisher entwickelten Module können durch die Pipeline ausgeführt werden.

Aussicht

In dieser Arbeit wurde nur ein kleiner Teil der Möglichkeiten zur Erkennung von Manipulationen mit einem Fokus auf visuelle Artefakte betrachtet. Semantische Texterkennung, Analyse von Unterschriften oder die Erkennung von Metadatenveränderungen können in Zukunft Aufschluss über die Echtheit von Dokumenten geben und weitere Arten der Fälschung aufdecken. Die Erschaffung von zusätzlichen Datensätzen und einer Diversifizierung der daran Beteiligten und der eingesetzten Software, sowie die Nutzung kommender Fortschritte bei der Nutzung von KI-gestützten Methoden, könnte zusätzlich zu diversen Datensätzen beitragen, die einen unumgänglichen Beitrag in der Erkennung von neuen Manipulationen und deren Aufdeckung darstellen.

References

[1] Alexey Dosovitskiy et al.
An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[2] JaidedAI.
Easyocr - ready-to-use ocr with 80+ languages supported.
`https://github.com/JaidedAI/EasyOCR`, 2020.

[3] Claude E. Shannon.
A mathematical theory of communication.
Bell System Technical Journal, 27(3):379–423, 1948.

[4] Weichao et al. Zeng.
Textctrl: Diffusion-based scene text editing with prior guidance control.
Advances in Neural Information Processing Systems, 37:138569–138594, 2024.

Referentin: Prof. Dr. Elke Hergenröther

Koreferent: Prof. Dr. Thomas März

niklas.weiand@stud.h-da.de