

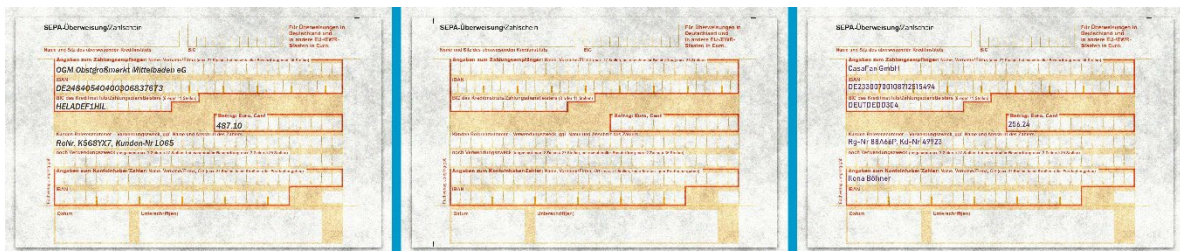
Vorschlag für eine Bachelor-/Masterarbeit

Fotorealistische Anonymisierung von Dokumenten

Die automatische Extraktion und Interpretation von Informationen aus Dokumenten ist in vielen Bereichen, von der Finanzwelt bis zur Medizin, von entscheidender Bedeutung. Dadurch können Aufgaben optimiert, die Entscheidungsfindung verbessert und auch die Zugänglichkeit für Menschen mit Sehbehinderungen verbessert werden. Visual Document Understanding (VDU) bietet eine Lösung, indem es diesen Prozess automatisiert, effizienter macht und ein schnelles Auffinden relevanter Daten ermöglicht.

Für den Aufbau einer robusten und zuverlässigen VDU ist eine große Menge an Daten erforderlich. Die Datenschutz-Grundverordnung (DSGVO) stellt jedoch aufgrund ihrer strengen Vorschriften zum Datenschutz und zur Datensicherheit eine Herausforderung für die Erfassung großer Datenmengen für VDU dar. Daher ist die Anonymisierung von Dokumenten unter Beibehaltung der Nützlichkeit der Daten ein Schlüsselfaktor bei der Erstellung von Datensätzen für die Entwicklung von VDU.

Das Ziel dieser Arbeit ist es, eine Methode zu entwickeln, mit der bestimmte Textfelder aus einem fotografierten oder gescannten Dokument realistisch gelöscht werden können. Darüber hinaus sollen diese Felder durch künstliche, aber realistische Daten ersetzt werden. Die folgende Abbildung zeigt ein Beispiel für die Anonymisierungspipeline.



Optional kann der Schritt des Inpainting auch übersprungen werden, so dass die Ersetzung der Informationen als Ende-zu-Ende-Prozess durchgeführt wird.

Die Implementierung und Auswertung soll mit Python erfolgen. Am Ende der Arbeit ist ein gut strukturierter und gut kommentierter Code abzugeben.

Die Diplomarbeit wird in Zusammenarbeit mit der Gini GmbH angeboten, einem marktführenden Unternehmen, das den Zahlungsverkehr für Millionen von Nutzern von Banken, Versicherungen und E-Commerce-Unternehmen in Deutschland vereinfacht.

Kontakt: Jürgen Seiler, 06.034, juergen.seiler@fau.de
Professor: Prof. Dr.-Ing. André Kaup
Voraussetzungen: Erfahrung mit Python-Programmierung, Computer Vision und maschinellem Lernen.
Verfügbar: Unmittelbar