

# Skalierbare Verfahren der Text- und Strukturerkennung für die Volltextdigitalisierung historischer Drucke

**Martin Jenckel**  
DFKI German Research Center  
for Artificial Intelligence  
martin.jenckel@dfki.de

**Syed Saqib Bukhari**  
DFKI German Research Center  
for Artificial Intelligence  
saqib.bukhari@dfki.de

**Andreas Dengel**  
DFKI German Research Center  
for Artificial Intelligence  
andreas.dengel@dfki.de

Ziel der Module "Bildoptimierung" und "Layouterkennung" war es, geeignete Methoden für die Vorverarbeitung bzw. Strukturanalyse der Digitalisate zu identifizieren und als Softwarewerkzeuge bereitzustellen. Dafür konnten, neben diversen traditionellen Ansätzen, für einige Werkzeuge auch Deep Learning Ansätze erfolgreich genutzt werden.

Die Performanz der einzel Werkzeuge hängt dabei stark von der Pipeline-architektur und -parametern ab.

## Modul 1: Bildoptimierung

### Binarisierung

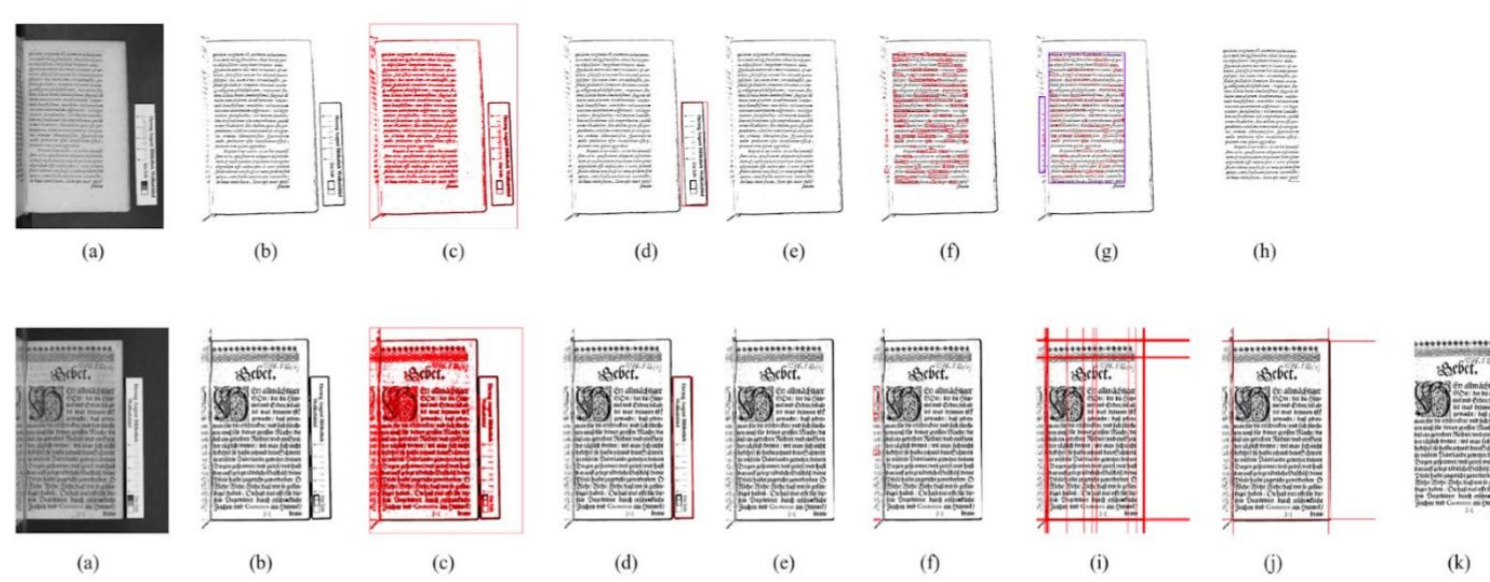
- Lokale Binarisierung mit Prozentfilter
- Bestimmung der Schwellwerte für Vorder- und Hintergrundfilter nur auf Regionen mit hoher Varianz

### Deskewing

- Textzeilenmodell basiert
- Idealer Deskew-Winkel = Maximale Varianz im horizontalen Mittelwert

### Cropping[1]

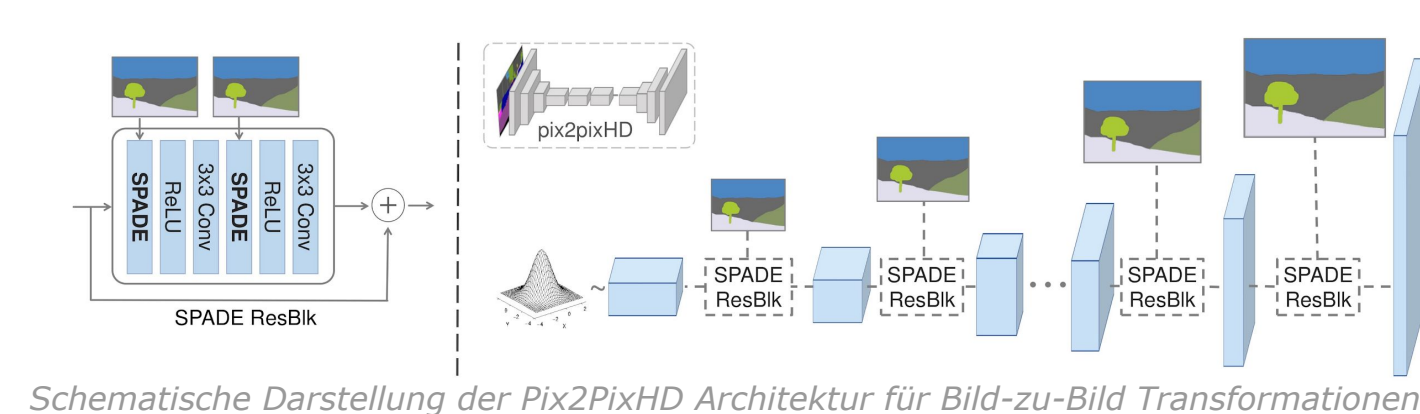
- Entdecken und Entfernen von Skalen und anderen Fremdkörpern
- Erkennen einzelner Dokumentbereiche mittels Gradienten
- Kombination in Gesamtdokument
- Alternativ: Erkennung von Liniensegmenten + Kombination



Bearbeitungsschritte im Cropping-Prozess (oben) sowie Alternativansatz (unten).

### Dewarping[2]

- Dewarping als Bild-zu-Bild Problem
- Pix2PixHD Architektur (GAN)
- Ansatz lernt Transformation aus synthetischem Warping



Schematische Darstellung der Pix2PixHD Architektur für Bild-zu-Bild Transformationen.

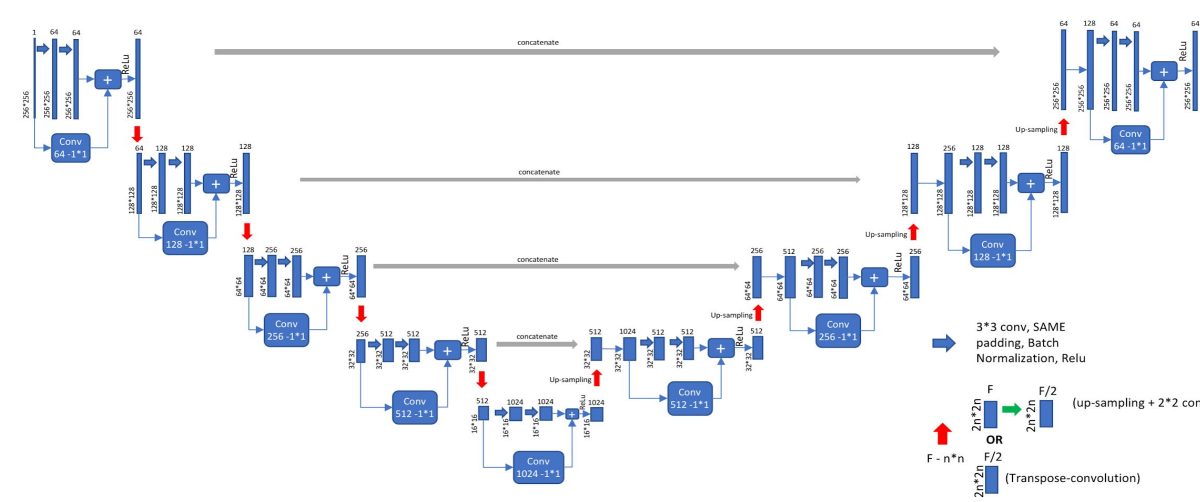
## References:

- [1] "A Robust Page Frame Detection Method for Complex Historical Document Images". Reza M. M., Rakib M. A., Bukhari S. S., Dengel A. ICPRAM2019, Prague, Czech Republic  
[2] "Document Image Dewarping Using Deep Learning". Ramanna V. K. B., Bukhari S. S., Dengel A. ICPRAM2019, Prague, Czech Republic

## Modul 2: Layouterkennung

### Text - Nicht-Text - Segmentierung

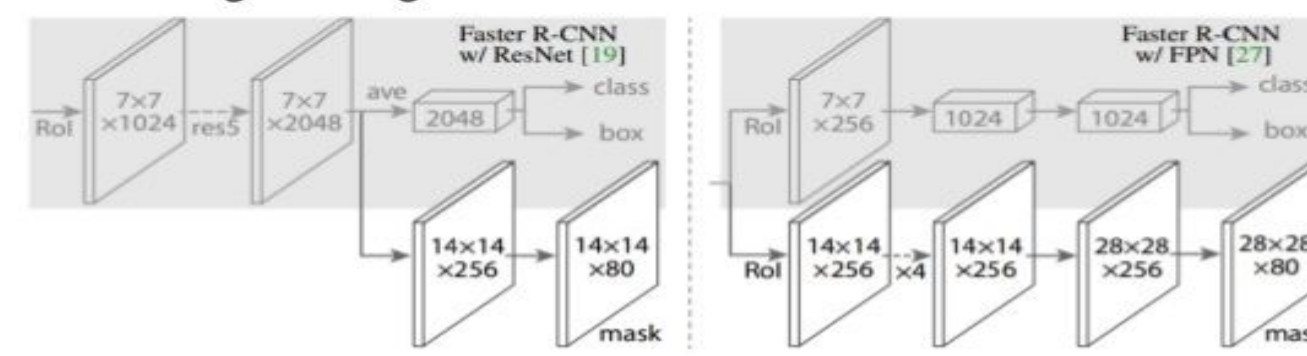
- Pixellevel Klassifizierung mit UNet + morphologische Unterscheidung



Schema eines UNet auf ResNET Basis. Auto-Encoder Architektur mit Skipconnections zwischen Encoder und Decoder.

### Seitensegmentierung & Segmentklassifizierung

- kombiniert Segmentierung (Box + Pixel Maske) und Klassifikation mit Mask-RCNN
- Nachbearbeitung mit Textmaske



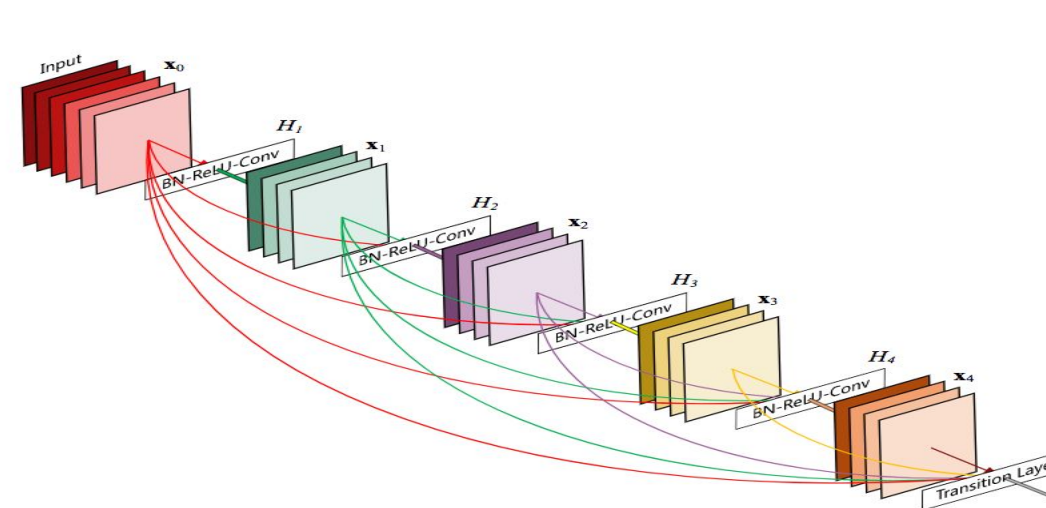
Darstellung der Mask-RCNN Architektur mit verschiedenen Basis-Blöcken. Im Projekt wurde die Linke Architektur genutzt.

### Textzeilenerkennung

- Gaußsche Filter + Kantenerkennung zur Bestimmung der Zeilenbasis
- Erkennen der Zeilenkerne
- Ausweitung der Zeilenkerne auf die komplette Textzeile

### Dokumentanalyse

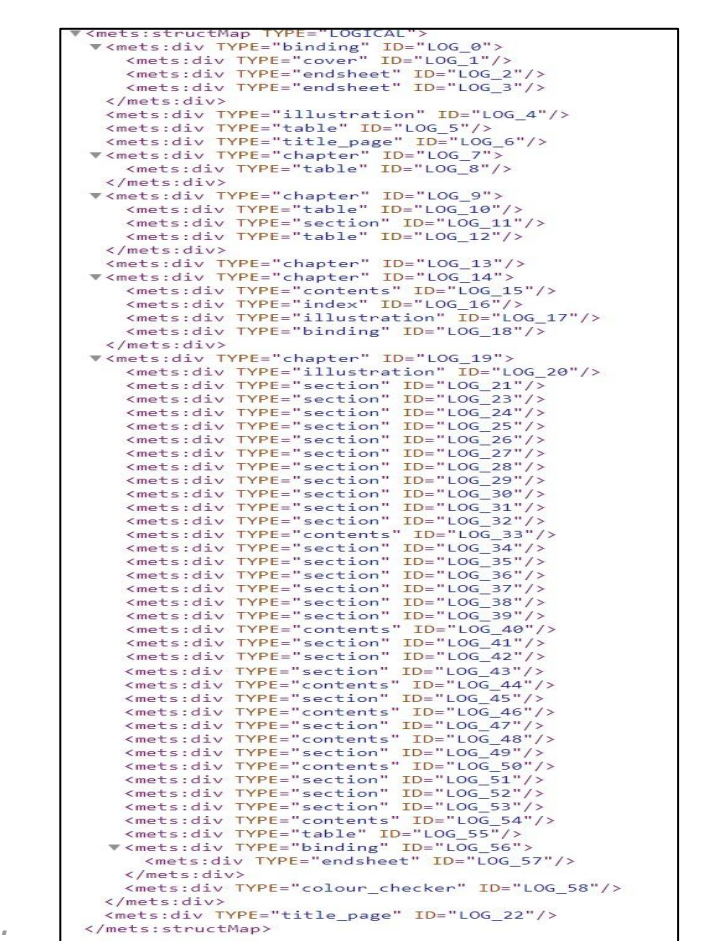
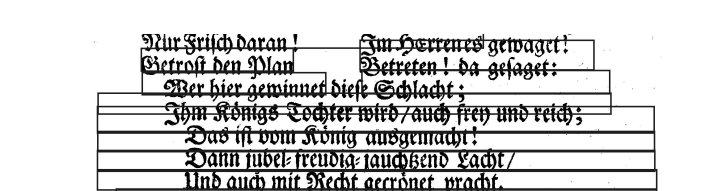
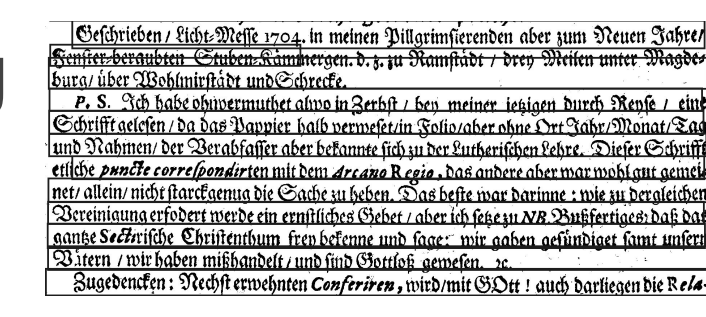
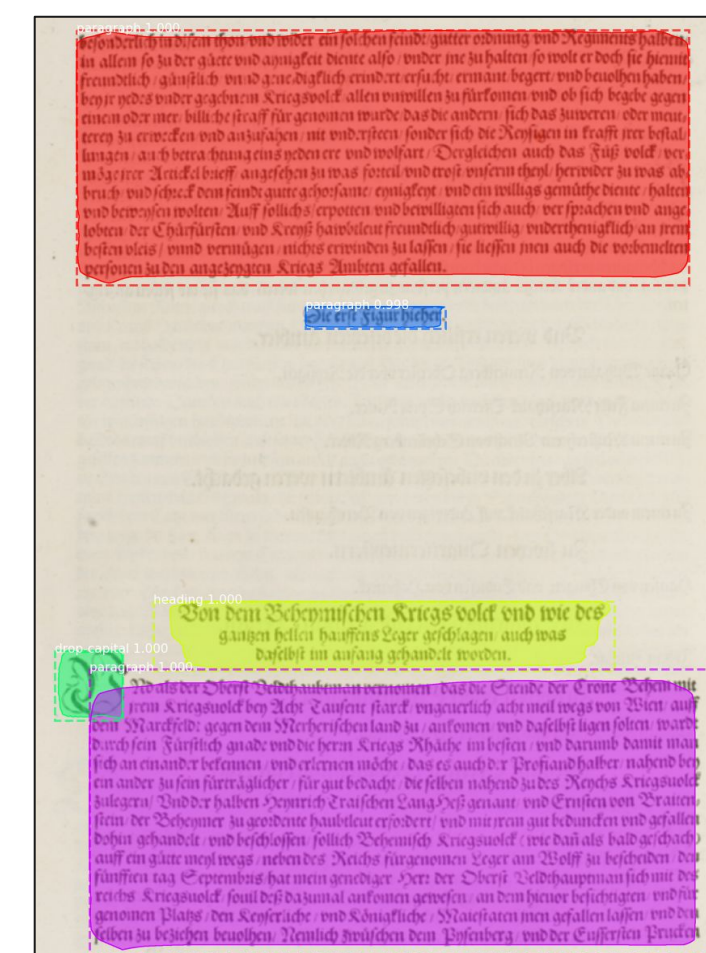
- Klassifikation einzelner Dokumentseiten mit DenseNet101
- heuristische Dokumenthierarchie



Schema der DenseNet101 Architektur mit 5 ResNET-Blöcken und Skip-Connections.

### Weiterentwicklungsvorschläge

- Nachtrainieren des Segmentierungsmodells für neue Datensätze (Bereitstellen der Trainingsinfrastruktur)
- Zusätzlicher RNN Layer für die Dokumentanalyse



## Contact:

DFKI Kaiserslautern / Smart Data Services  
**Martin Jenckel**  
+49 631 20575 3760  
Martin.Jenckel@dfki.de