

# Texterkennung und -detektion in Rastergrafiken

OCR<sup>1</sup>-Pipeline zur Erkennung von Scene Text aus Rastergrafiken im RGB<sup>2</sup> Farbraum am Beispiel einer Kombination aus EAST<sup>3</sup> und CRNN<sup>4</sup> mit CTC<sup>5</sup> Loss

## Ausgangssituation

Bilder können neben einfachen Objekten auch Informationen in Form von Textelementen enthalten. Aufgrund der Bildstruktur sind diese einer Speicherung und Weiterverarbeitung in Form von maschinenlesbaren Texten nicht direkt zugänglich.

Einfache und etablierte Verfahren zur Texterkennung (OCR - Optical Character Recognition) funktionieren in wohldefinierten Umgebungen sehr gut, haben jedoch Probleme, wenn Text in natürlichen Umgebungen (scene text) auftaucht.

Scene text kann in focused scene text für das erwartete und ggf. gesteuerte Vorkommen von Text, sowie incidental scene text, dem unerwarteten und nicht kontrollierbaren Vorkommen von Text, unterteilt werden. Speziell für Rastergrafiken haben sich neuronale Netze als außerordentlich leistungsfähige Lösungen erwiesen.

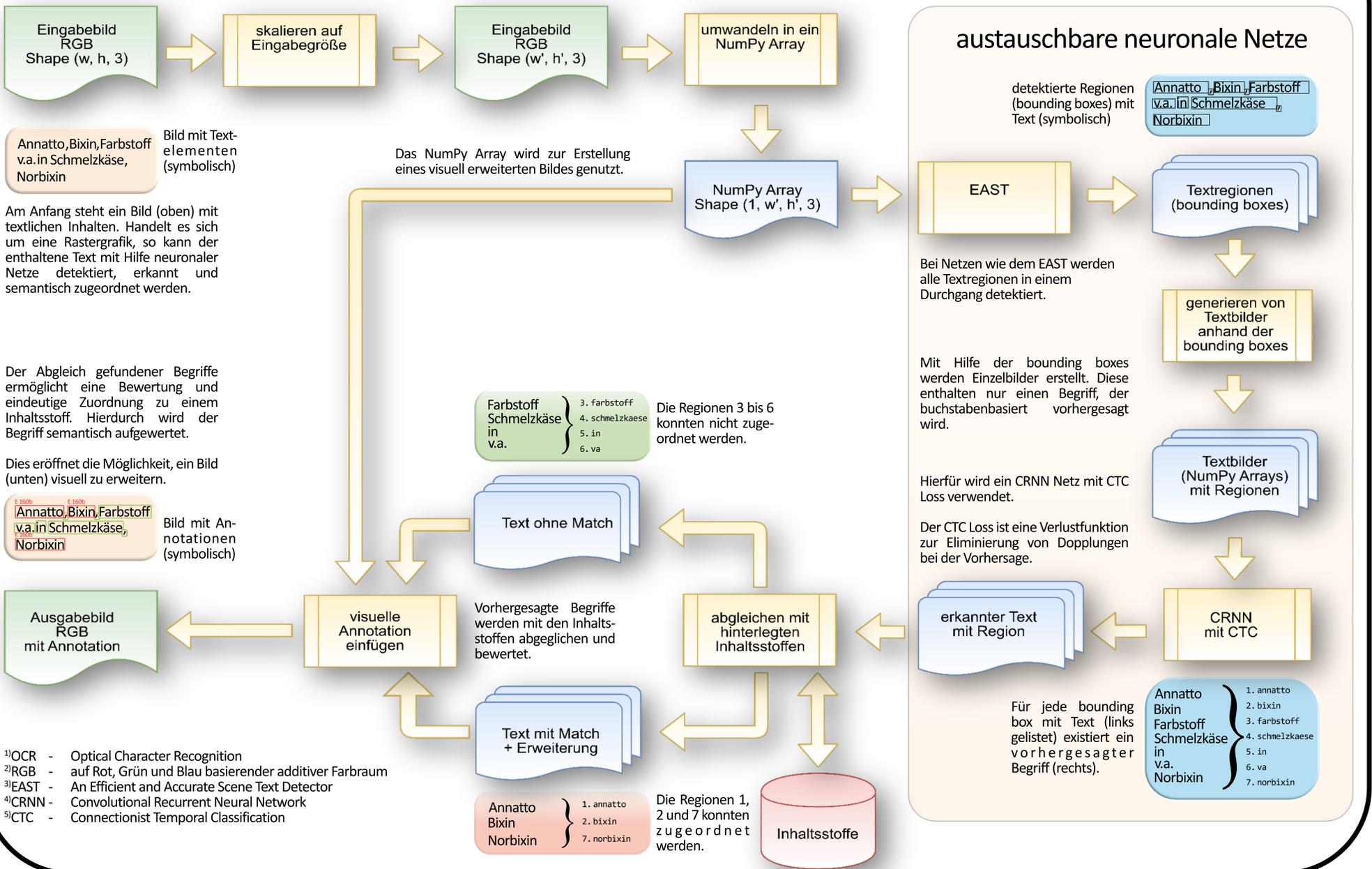
## Zielsituation

Die in einem Bild enthaltenen Regionen mit Text sind identifiziert. Anhand der gefundenen Regionen können Detailbilder erstellt und der jeweils enthaltene Text vorhergesagt werden. Als Ergebnis liegen der enthaltene Text sowie sein Ursprung in einem maschinenlesbaren Format vor.

Die so erfassten Texte können direkt mit einer hinterlegten Datenbasis abgeglichen werden. Dies ermöglicht eine semantische Einordnung der Texte und weitere Aktionen wie ein visuelles Feedback.

Eine einfache Art des visuellen Feedbacks ist die Farbcodierung mit Hilfe eines den Text umgebenden, farbigen Rahmens (bounding box). Das Hinzufügen zusätzlicher Informationen kann den Informationsgehalt eines Bildes zusätzlich steigern. Beides führt zu einer erweiterten Realität des Bildes (augmented reality).

Eingabebilder werden abhängig vom neuronalen Netz skaliert und im Anschluss in ein NumPy Array umgewandelt. Diese Form der Datenübergabe ist häufig anzutreffen.



<sup>1</sup>OCR - Optical Character Recognition  
<sup>2</sup>RGB - auf Rot, Grün und Blau basierender additiver Farbraum  
<sup>3</sup>EAST - An Efficient and Accurate Scene Text Detector  
<sup>4</sup>CRNN - Convolutional Recurrent Neural Network  
<sup>5</sup>CTC - Connectionist Temporal Classification

