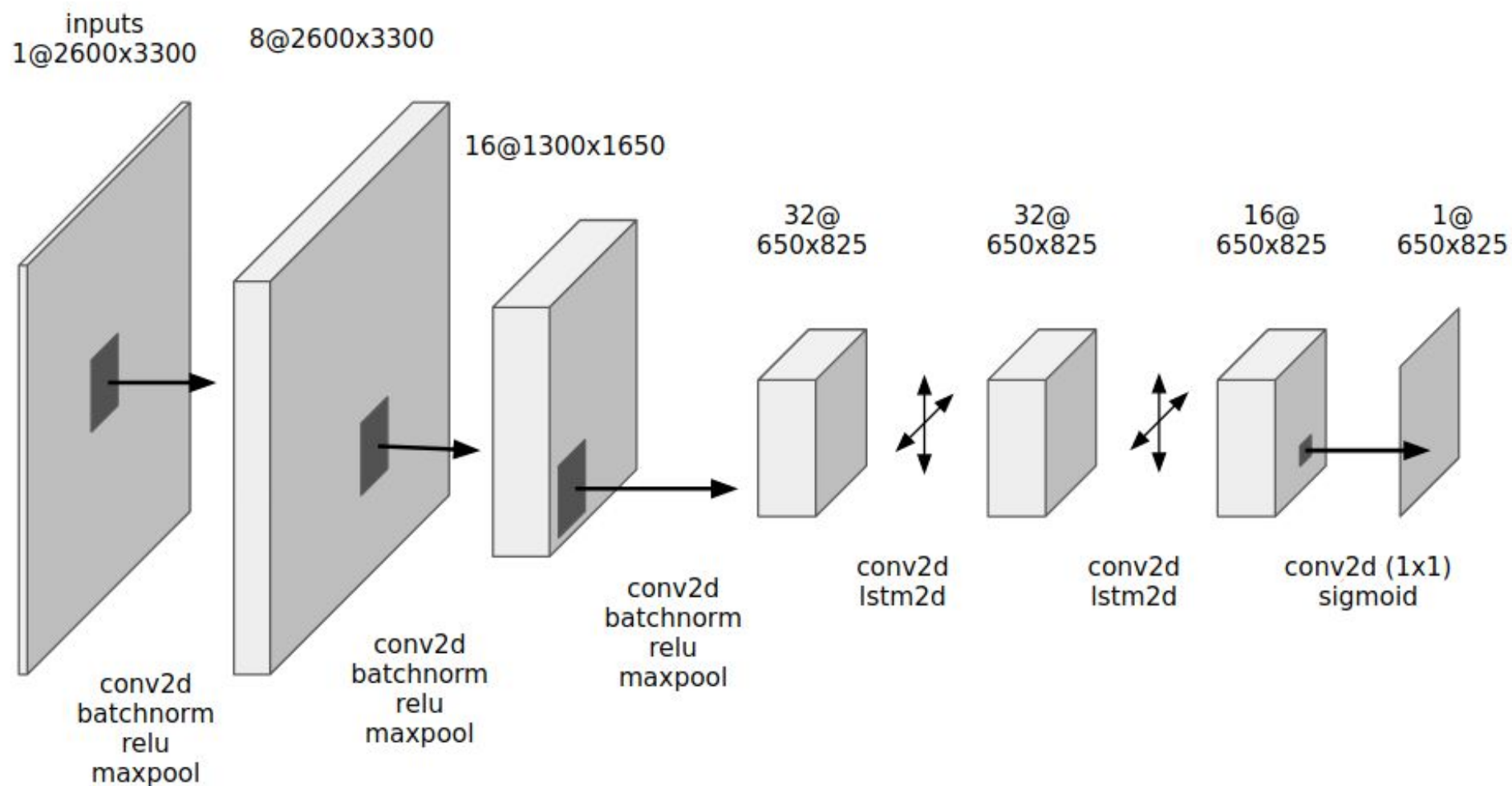# Robust, Simple Page Segmentation using Hybrid Convolutional MDLSTM Networks

Thomas Breuel

NVIDIA Research

Single Step Deep Learning Layout Analysis

- simple deep learning approach
- convolutional layers + MDLSTM layers
- trainable, adapts to document databases
- simultaneously performs: text/image segmentation, border noise removal, text/math segmentation, text line detection
- robust to noise
- can be applied directly to grayscale, photographic images
- runs on GPU (1 sec / page)
- comparable performance to existing layout analysis systems using just 1600 training pages

Deep Learning Architecture for Layout Analysis

inputs
1@2600x3300

8@2600x3300

16@1300x1650

32@
650x825

32@
650x825

16@
650x825

1@
650x825

conv2d
batchnorm
relu
maxpool

conv2d
batchnorm
relu
maxpool

conv2d
batchnorm
relu
maxpool

conv2d
lstm2d

conv2d
lstm2d

conv2d (1x1)
sigmoid

Single Step Border Removal, Multicolumn Textline Finding

Single Step Border Removal, Math Removal, Figure Removal, Text Line Detection
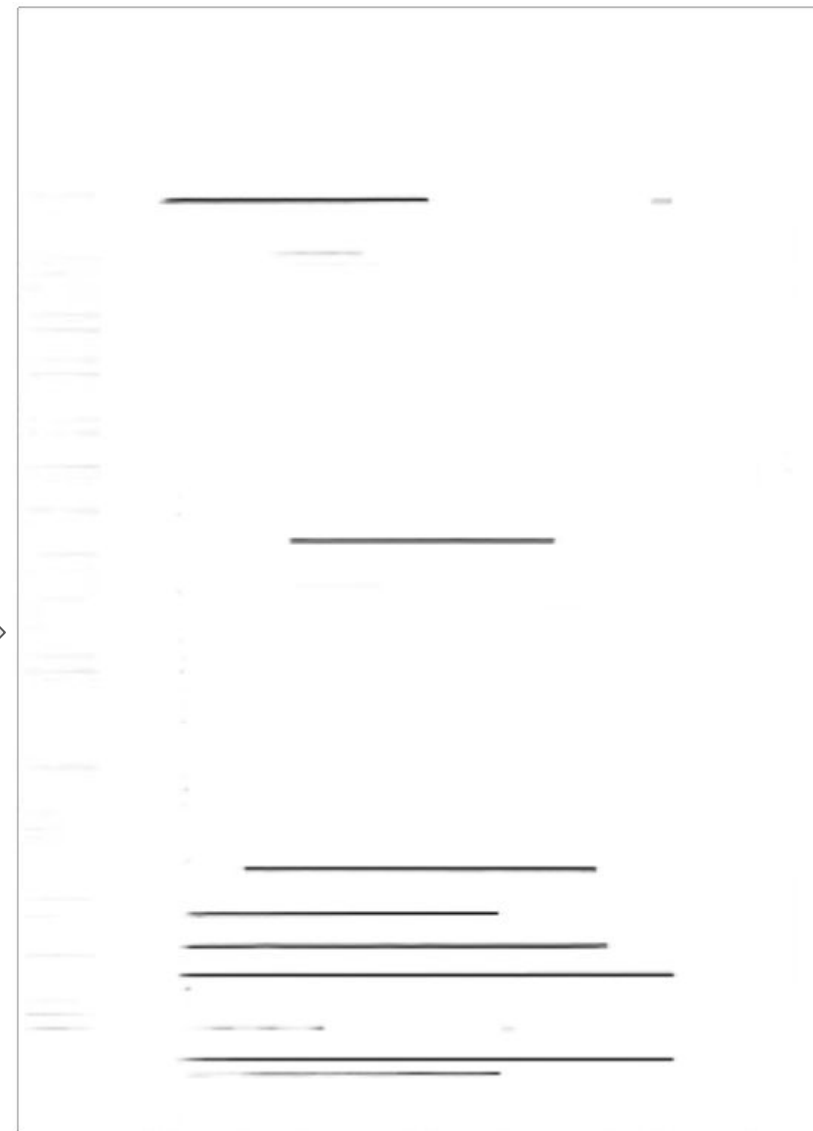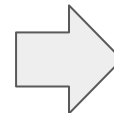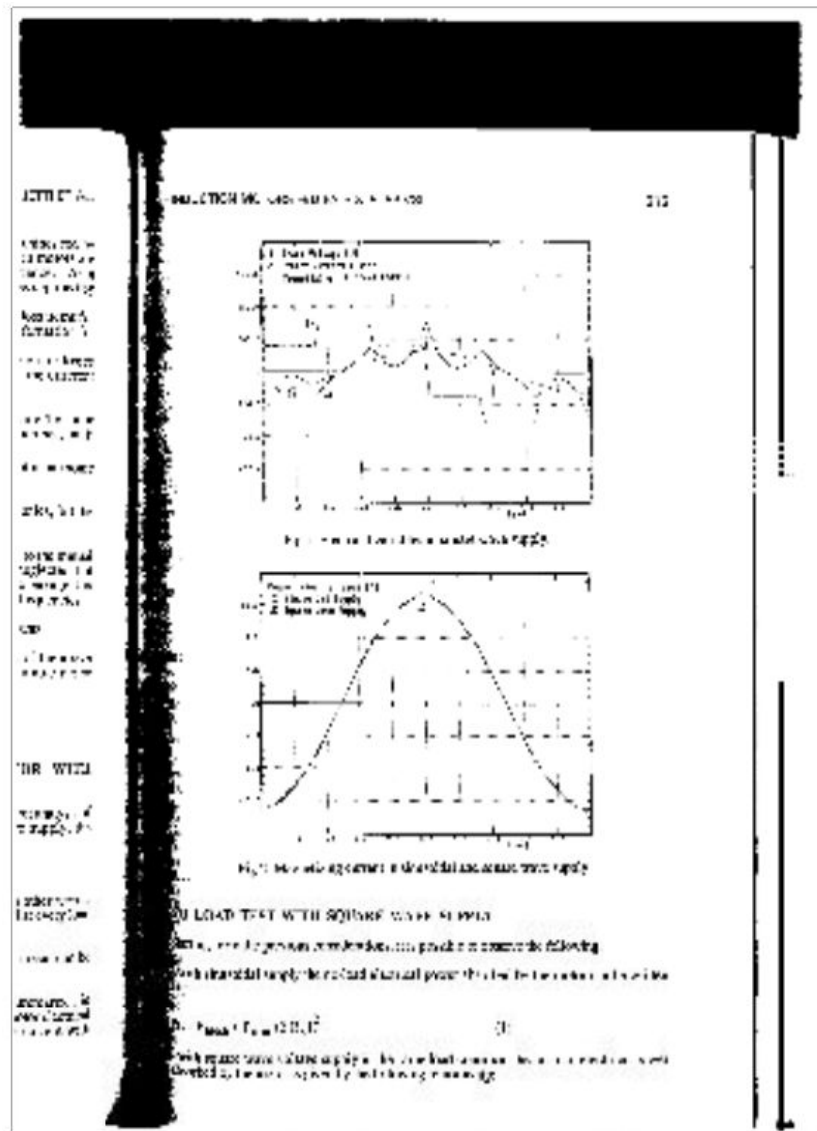
Direct Application to Grayscale Input
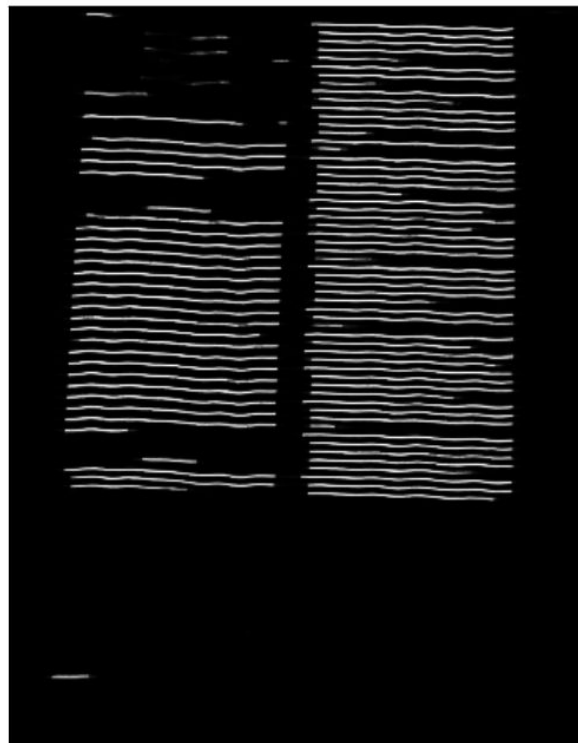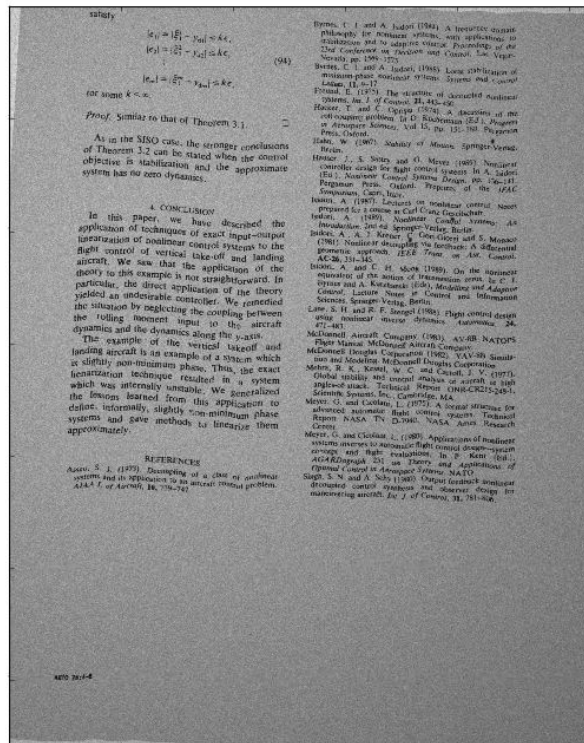
Deep Segmenter vs Tab Segmenter
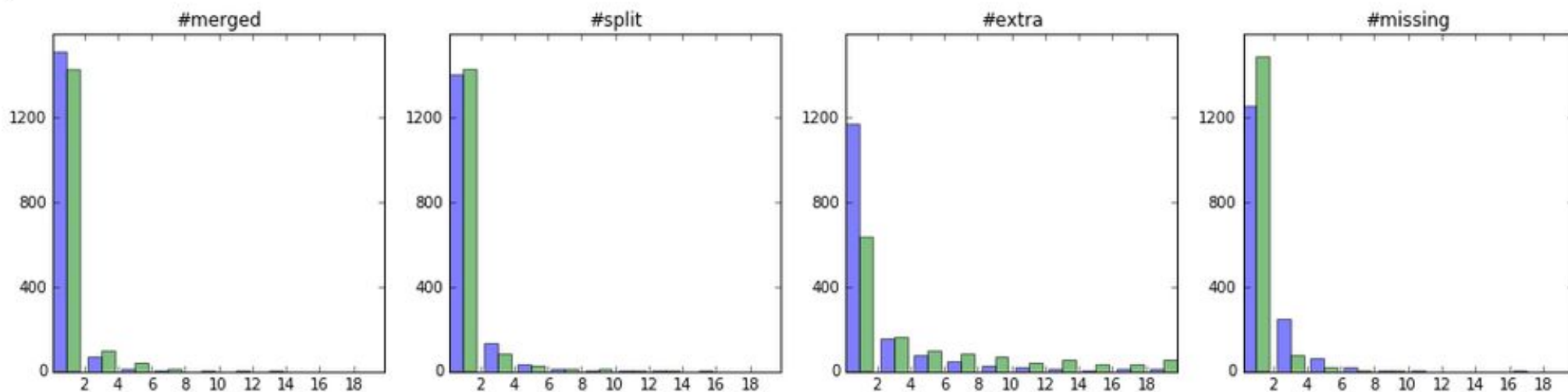
The Deep Segmenter performs comparable to, or slightly better than, the Tab Segmenter.

Results on page frame only (border noise removed).

Ground Truth Prep vs Layout Analysis Training

learns both text line detection
+ border noise removal

learns only text line detection

border noise removal

error masking

training input

training target
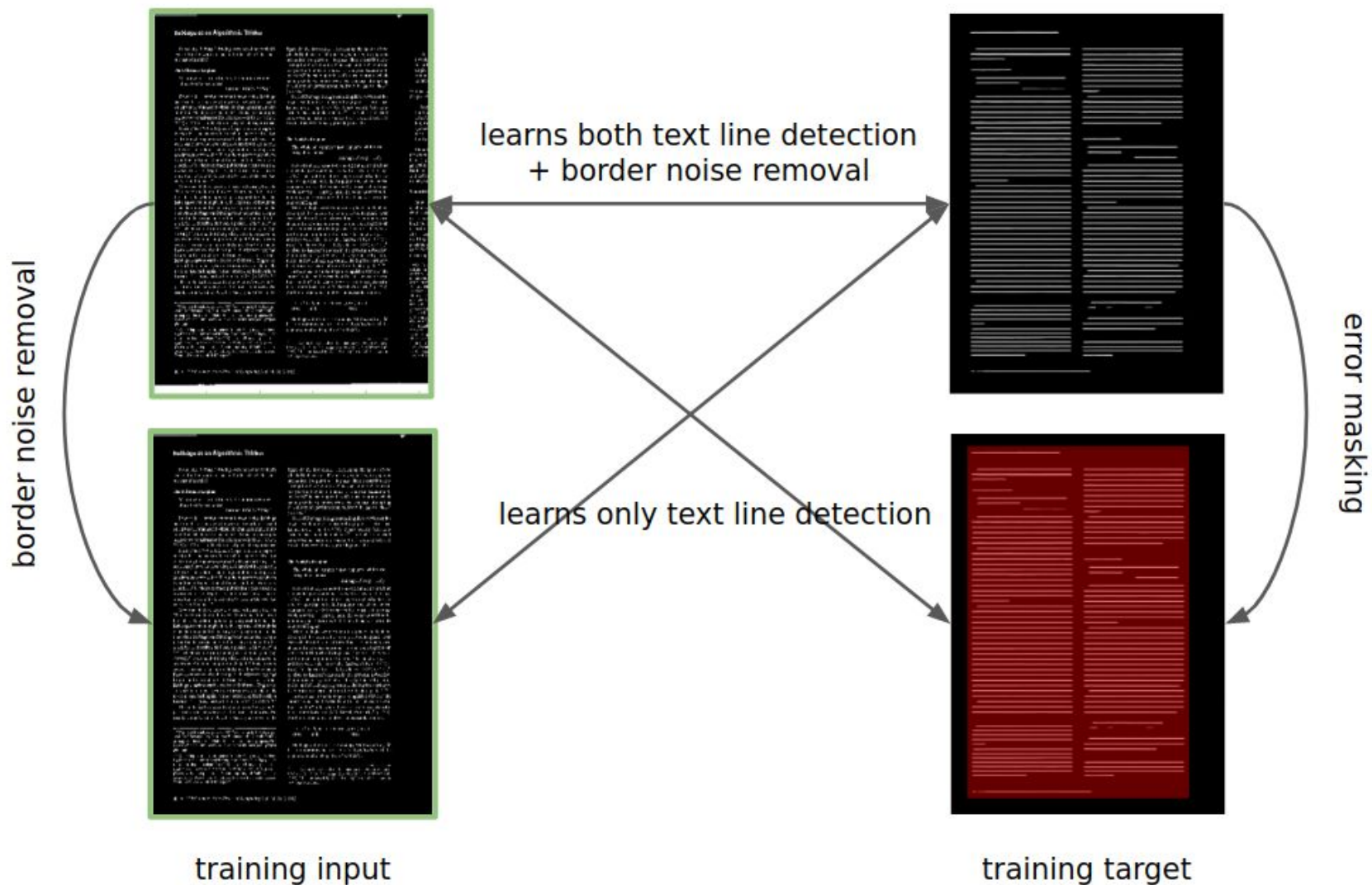
Training for Different Detection Tasks

original page   body text detection   text detection   text frame detection   border noise detection

Line Extraction from Text Center Lines

- *The text center lines identified by the DNN are used as seeds for identifying the connected components belonging to the actual text line.*
- *Each connected component within a halo around the text center line seed is assigned to the closest seed.*

Layout Analysis with Deep Learning

processor simulator and a detailed mem-
ory simulator for the Dash prototype.
Tango allows a parallel application to

vers
Wa
the

Deep Neural Network computes text line centers from binary or grayscale document image. Images, math, etc. are implicitly removed.

processor simulator and a detailed mem-
ory simulator for the Dash prototype.
Tango allows a parallel application to

vers
Wa
the

For each identified text line center, all the closest connected components within a halo are assigned to the line.