LSTM APPLICATIONS

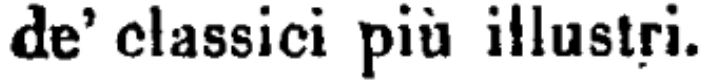
(other than handwriting)

Thomas Breuel
University of Kaiserslautern

OCR WITH LSTM

text line recognition

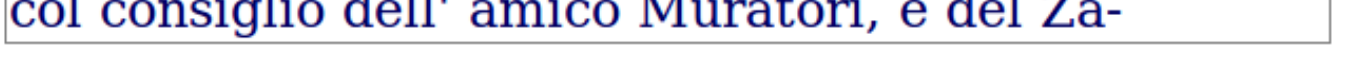
984-22/010007.bin.png



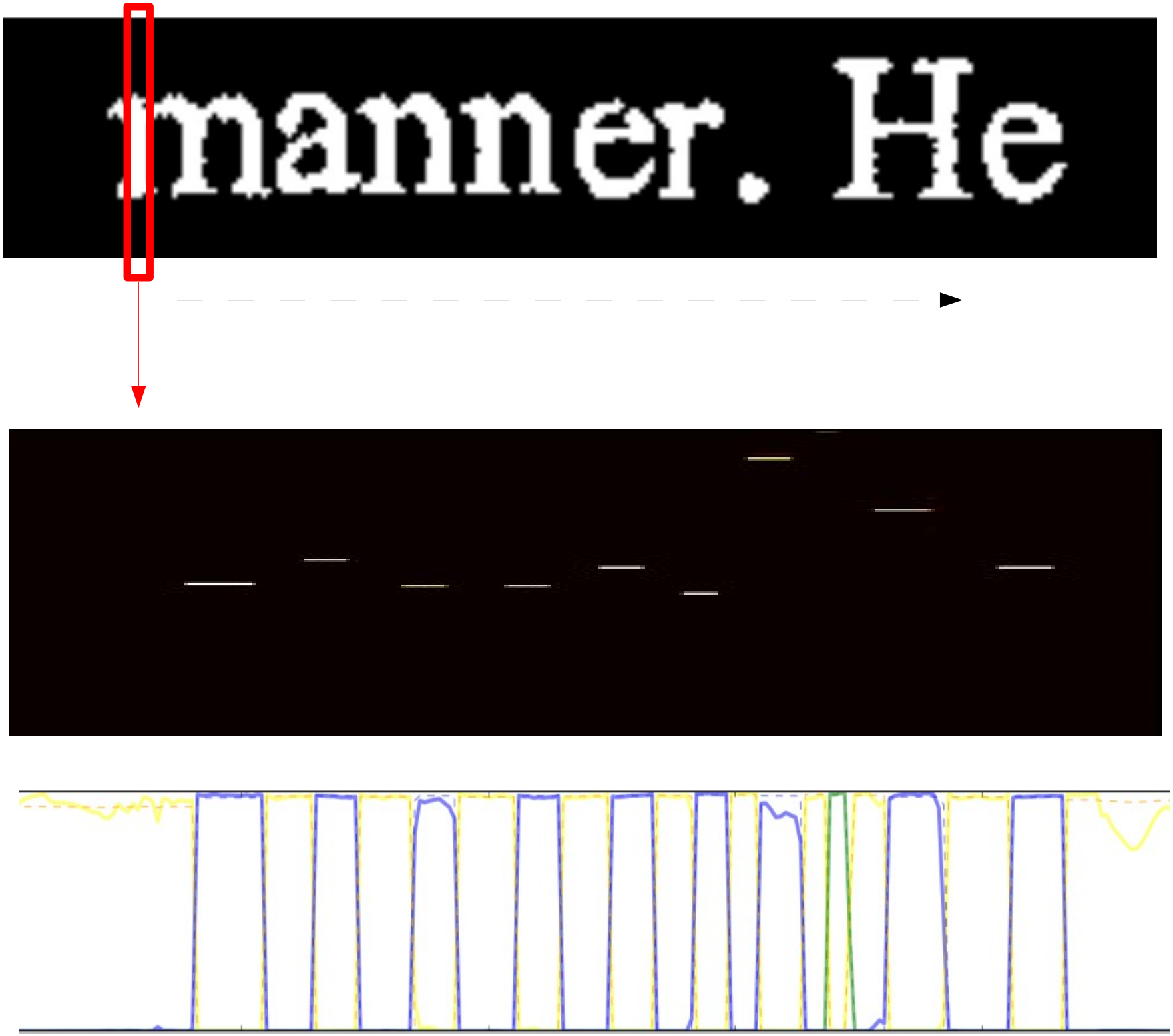
984-22/010008.bin.png

984-22/010009.bin.png

984-22/01000a.bin.png

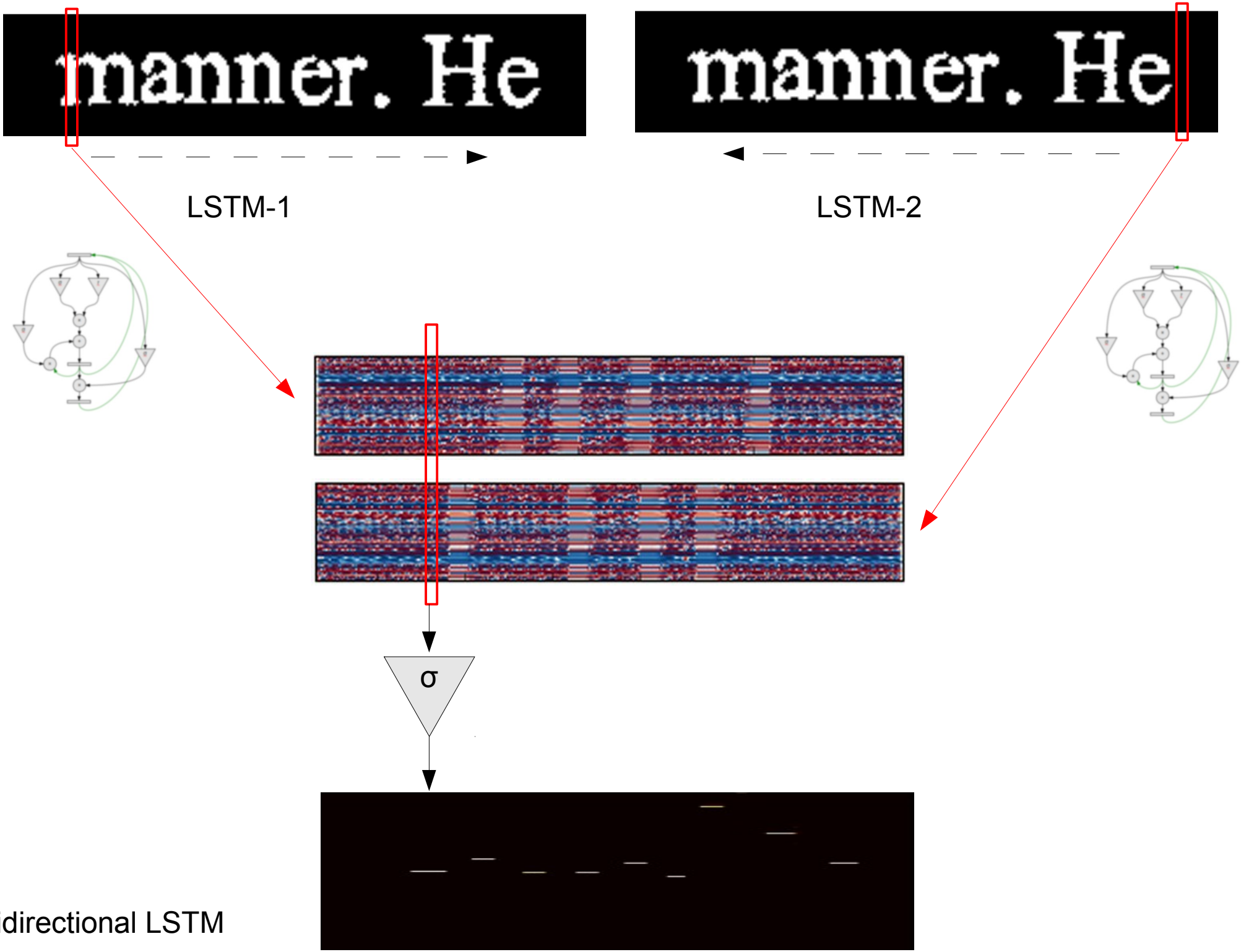


984-22/01000b.bin.png


manner. He
class

OCR with LSTM

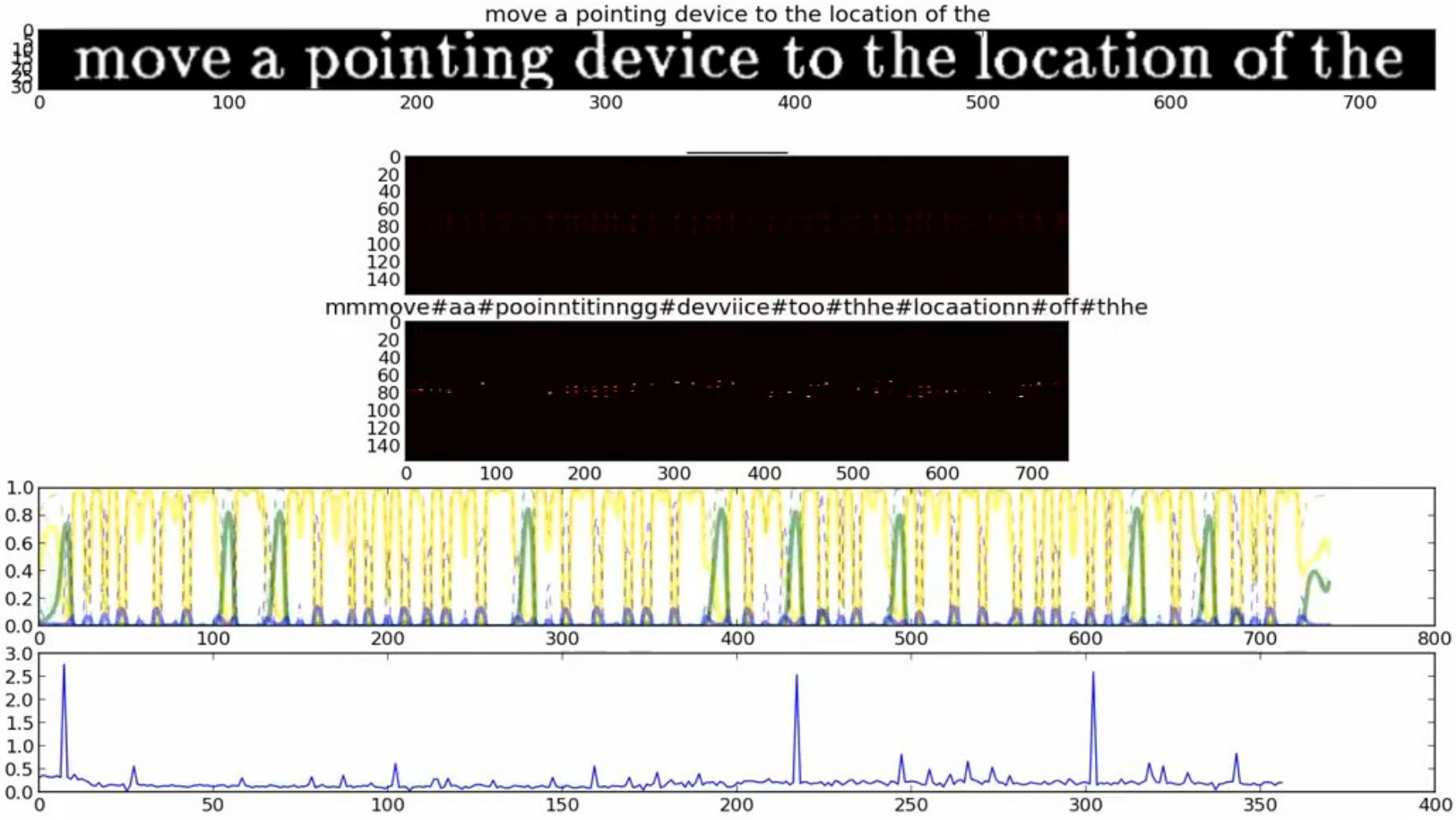
- **recognition operates on text lines**
- **1D LSTM, not 2D LSTM**
- **input text line needs to be carefully normalized (baseline, size)**
- **similar to HMM-based OCR**
- **output is a sequence of vectors**
  - same length as width of input image
  - each vector represents posterior probability
  - requires “decoding” into symbol sequence



bidirectional LSTM

training LSTM

- **parameters are adjusted using backpropagation, just as for simple MLPs**
- **problem: input is in pixels, output is in characters**
- **solution: use forward-backward alignment (from HMMs), called “CTC” by Schmidhuber**



http://www.youtube.com/watch?v=czG5Jk9iC7c

CTC, Viterbi, Forward-Backward (sorry about the informality)

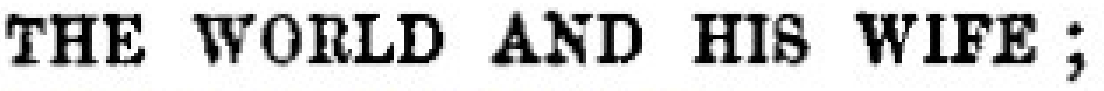
- **CTC is basically forward-backward**
- **training does not work with Viterbi**
  - odd... it usually does for HMMs
- **what's the difference?**
  - initially, CTC spreads multiple classes
  - Viterbi just assigns one class, which is often wrong
- **what makes CTC work is the combination of**
  - learning posterior probabilities (instead of classifications)
  - spreading possible classifications, instead of guessing the best

OCR System	Train	Lang Mod?	English (UW3)	Fraktur Fontane	Fraktur E-G
OCRopus-LSTM	UW3 / artificial	-	0.60	0.15	1.37
Tesseract + dict	many	YES	1.3	0.90	1.47
OCRopus nnet + HMM	UW3	-	1.6		
OCRopus-lattice	many	YES	1.6		
OCRopus-lattice + ngraphs	UW3	-	2.14		
ABBYY 3.0 + dict	many	YES	0.85	(high)	(high)

(these results are a few months old; for more recent results and a detailed explanation, see our paper)

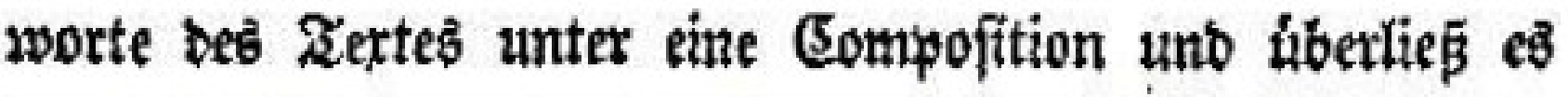
(a) stand in the window of his chamber in the morning,

(b) Travers on Constitutional Irritation.

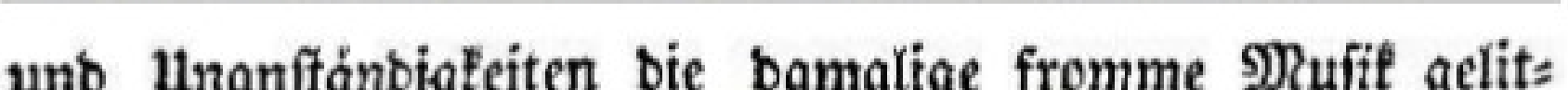


(c) THE WORLD AND HI8 WIFE ;

(d) Adjustments in OECD Countries." Economic Policy 21: 205-248.



(e) worte des Textes unter eine Eomposition und überließ es



(f) und Unanständijgkeiten die damalige fromme Musik gelit-

OCR with LSTM

- Latin, Devanagari, etc. solved well with 1D LSTM, other scripts may require MDLSTM
- training only requires text line images and transcriptions (no characters, segmentation)
- almost no script-specific assumptions
- ligatures are learned automatically
- much less training data required (a few thousand lines)
- artificially generated training data suffices

OTHER LSTM APPLICATIONS

language modeling

- **recurrent neural network language models**
  - simple RNN
  - LSTM-based
- **similar to Reber grammar examples**
  - usually predict probabilty distribution of next word
  - other possibilities
- **examples**
  - Soutner and Müller, 2012
  - Sundermeyer, Schlüter, Ney, 2012
  - Azawi, Afzal, Breuel, 2013

speech recognition

- **no full speech recognition system yet, more work on deep neural networks**
- **special purpose, preliminary uses**
  - Graves and Schmidhuber: “Framewise phoneme classification with bidirectional LSTM and other neural network architectures” (2009)
  - Wöllner et al: “Robust discriminative keyword spotting for emotionally colored spontaneous speech using bidirectional LSTM networks (2012)

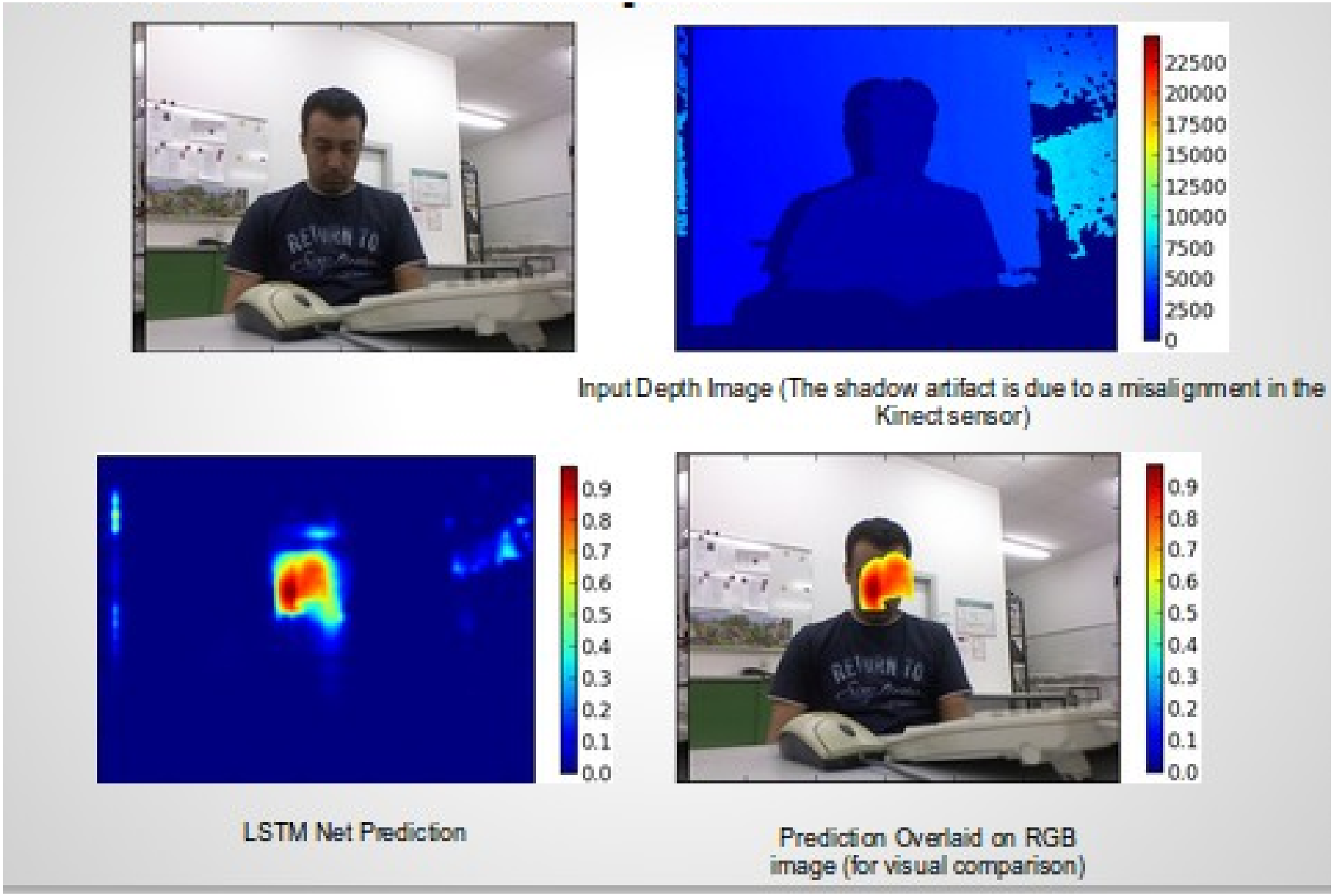
other applications

- **computer vision**
  - object recognition in very specific contexts, like meeting segmentation, multimodal interaction etc.
- **time series prediction**
  - chaotic time series prediction as test case, no significant applications to financial time series or forecast known

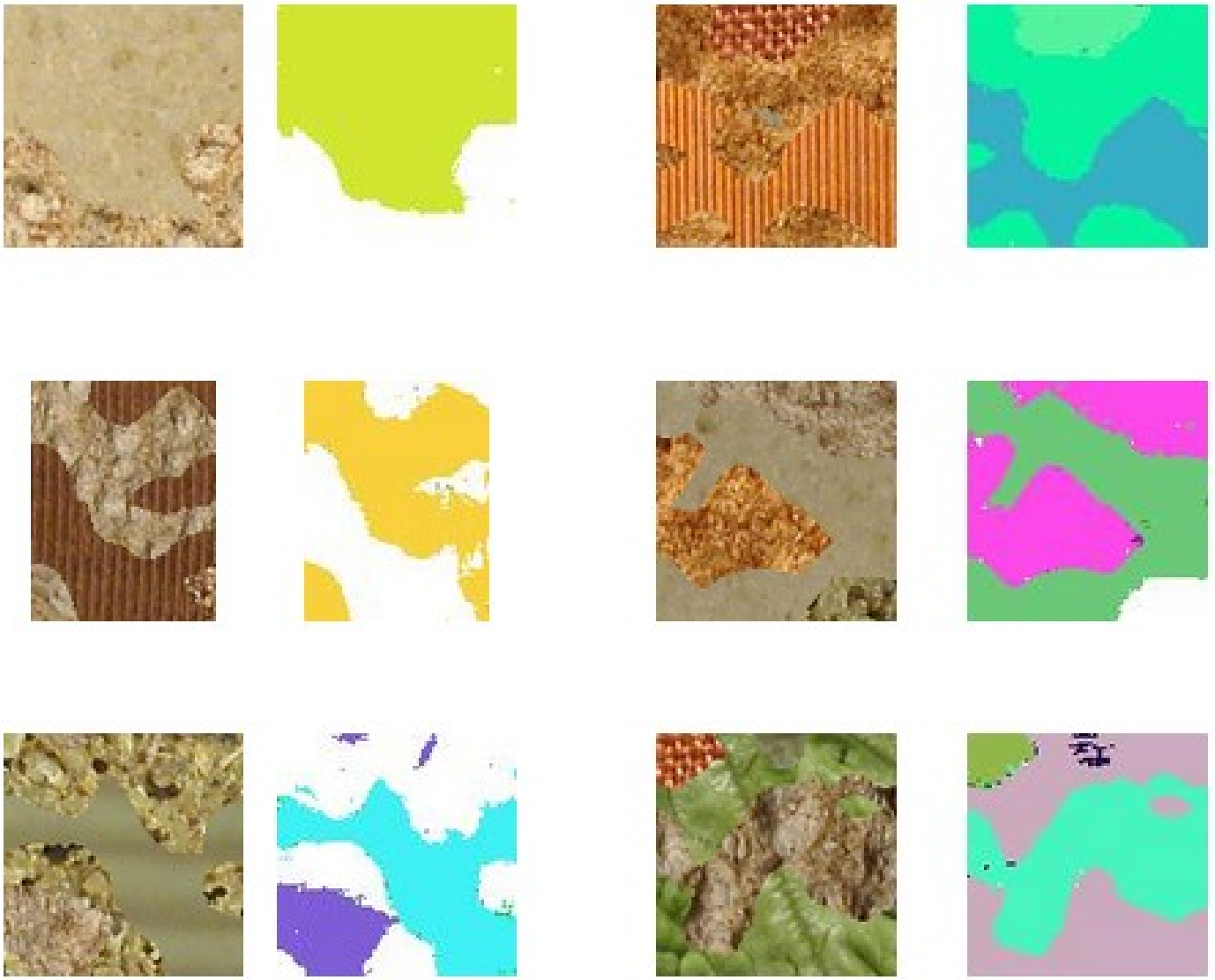
current work in IUPR

- **classic computer vision problems**
  - texture segmentation
  - object recognition
  - interest point detection
  - face detection / recognition
- **document analysis problems**
  - binarization
  - layout analysis
  - language modeling
- **general sequence prediction problems**

face detection in depth images

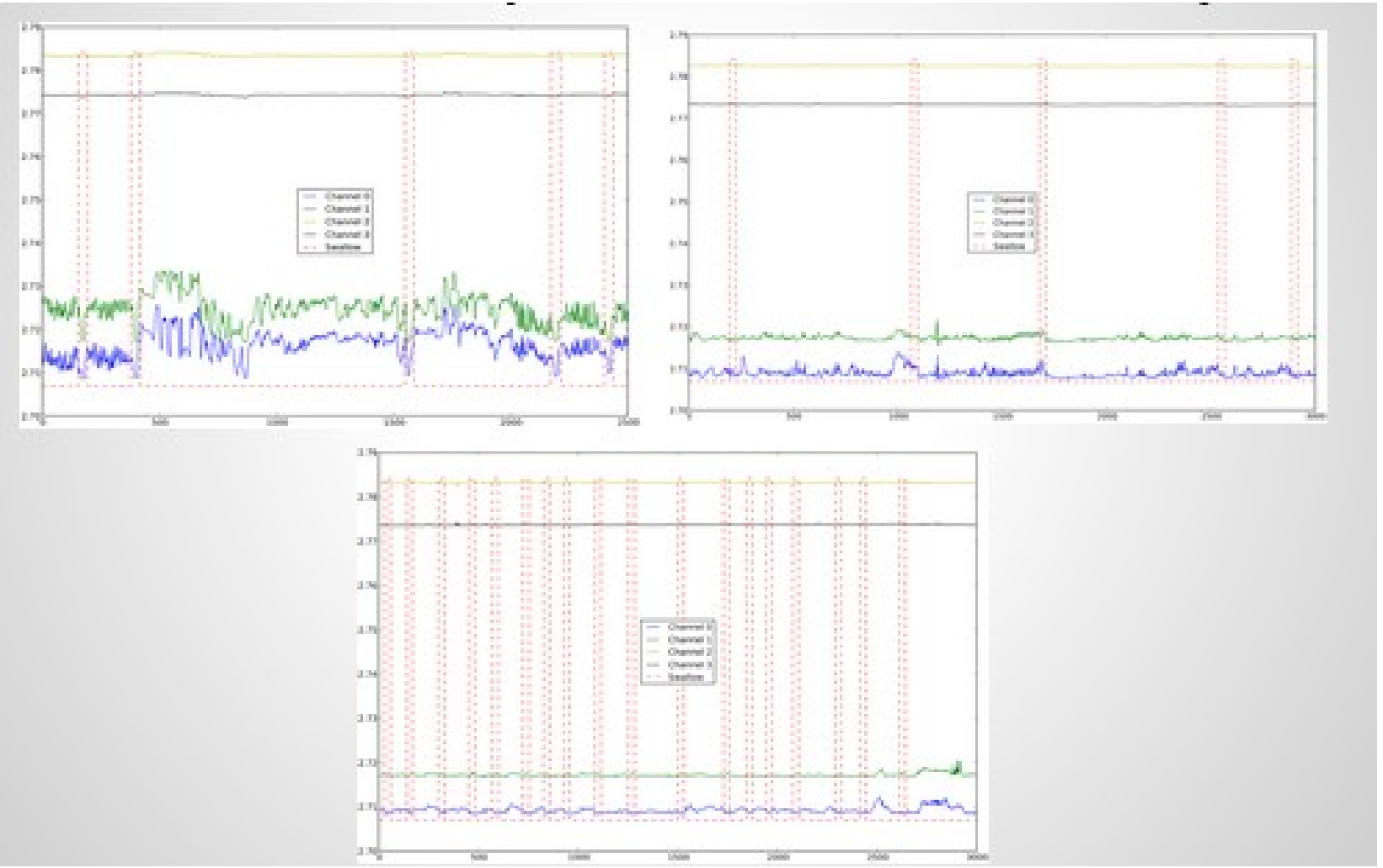


texture segmentation



layout analysis

event detection in time series



SUMMARY

summary

- **many potential applications for LSTM**
- **most important so far:**
  handwriting recognition and OCR
- **very active:**
  language modeling, speech recognition
- **lots of room for work and new domains**