ScaDS
DRESDEN LEIPZIG

TEXT RECOGNITION IN HISTORICAL MAPS

Eric Schölzel[1], Niklas Werner[1],
Hendrik Herold[2], Peter Winkler[1]
[1] *TU Dresden*, [2] *IÖR Dresden*

ZIH
Zentrum für Informationsdienste und Hochleistungsrechnen

Leibniz Institute of Ecological Urban and Regional Development

www.scads.de

HISTORICAL MAPS – EXAMPLE 1

ScaDS DRESDEN LEIPZIG

HISTORICAL MAPS – EXAMPLE 2

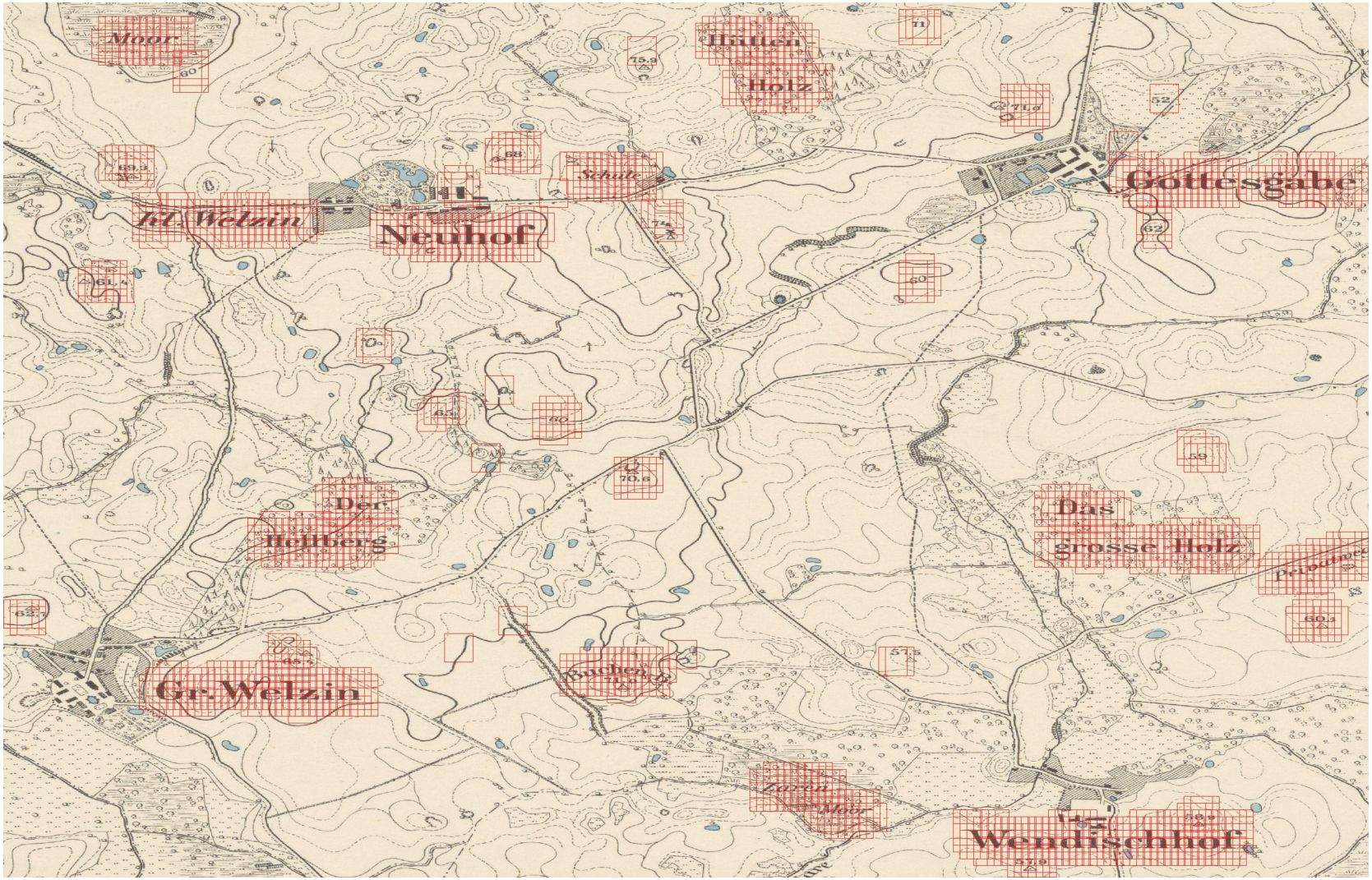
OBJECTIVES

- Large-scale analysis of historical maps
  - Historical maps contain precious information for urban and regional planning and ecological research
  - Development of settlements
  - Changes of land-use

- Mapset “Messtischblätter“
  - Middle-European maps from 1850-1943
  - Scale 1:25 000
  - Available from the SLUB Dresden

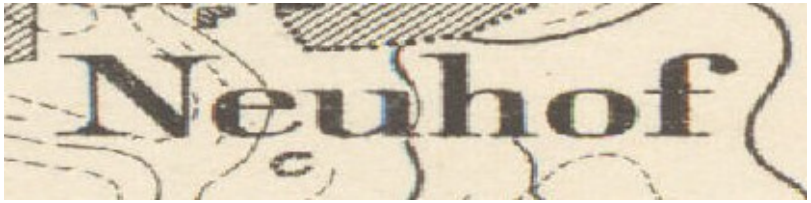
- Text recognition in two steps
  - Text detection (finding the location of the Text)
  - Optical character recognition (OCR)

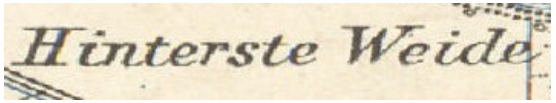
TEXT DETECTION

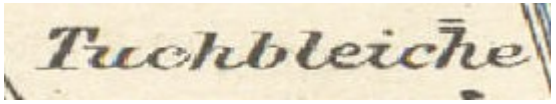


Source: Nathanael Philipp, ASV Uni Leipzig

EASY AND TRICKY CASES...







TEXT RECOGNITION: ONGOING WORK

- Convolutional neural network
  - One-hot encoding
- Artificial training data
  - Ca 120 fonts
  - 40 000 words
  - Background randomly chosen from different maps
  - Choice of colours automatised
  - Use of dictionary possible
  - Score: less than 40% (validation)
- Test case:
  - Counting letters (without recognition of letters)
  - Score: 55% (validation)

ScaDS DRESDEN LEIPZIG

ARTIFICIAL TRAINING DATA

ScaDS
DRESDEN LEIPZIG

THANKS FOR YOUR ATTENTION!