A practical usage: creation of training material

To train recognition systems, we need annotated images without line positions. We have to map the transcript to the image. In the literature, line positions are assumed to be known or reliably obtained with automatic methods.

- Consider several segmentation hypotheses
- Jointly find the segmentation and transcript mapping
- Reject lines in the segmentation, which content is not in the transcript
- Perform the mapping with a recognition system, constrained by the transcript

Results

**Analysis**
- Evaluate the method, we have to measure the quality of the segmentation and of the mapping
- We applied the method on public databases for which we know the line positions and transcript

**Segmentation error** in ZoneMap
- ZoneMap aligns bounding boxes from a reference and an hypothesis in terms of Match, Misses, Missing, Spikes and False Alarms
- The error counts black pixels that are missed or falsely included in an hypothesis segmentation w.r.t. the reference segmentation.

**Mapping error** in **Edit Distance**
- We use the bounding box matching found with ZoneMap.
- For each configuration, count the number of word substitutions, deletions, and insertions.

**Optical Model**
- GMM: 0.00 / 0.22
- BLSTM-RNN (Kaldi): 1.00 / 0.50

**Results on IAM (dev)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Seg.Err</th>
<th>Map.Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLSTM-RNN</td>
<td>1.56</td>
<td>1.24</td>
</tr>
<tr>
<td>GMM</td>
<td>3.60</td>
<td>4.48</td>
</tr>
<tr>
<td>BLSTM-RNN</td>
<td>0.75</td>
<td>3.23</td>
</tr>
<tr>
<td>No recognition (no G in hypo)</td>
<td>98.05</td>
<td>92.26</td>
</tr>
<tr>
<td>Known-less break symbols</td>
<td>0.82</td>
<td>0.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLSTM-RNN</td>
<td>words</td>
<td>words</td>
</tr>
<tr>
<td>RNNA</td>
<td>lines</td>
<td>lines</td>
</tr>
<tr>
<td>ZoneMap</td>
<td>lines</td>
<td>words</td>
</tr>
</tbody>
</table>

**Limitations - Future Work**
- The current segmentation FST can only handle simple inputs
- The segmentation FST could be improved if the segmentation algorithms returned positions with confidence scores
- The recognition is very constrained, and allows to only recognize transcript words
- An implementation of fine regression at this level could be beneficial
- The method cannot cope with transcript errors, as in other publications
- It could be implemented in the FST

Conclusions

We implemented several trivial constraints derived from the knowledge of the transcript,
- The recognition in the decoding graph enables a quick recognition and is crucial for a good mapping even with a recognition system which has not been adapted
- The manuscript FST is important for a mapping that is consistent at the document level (i.e., the same part of the transcript is not mapped to several lines)
- Finding a good mapping with this method generally improves the segmentation (less lines are falsely accepted, but some are wrongly discarded)
- Keeping several segmentation hypotheses is not always better than the best segmentation, but good since we do not know a priori which segmentation algorithm will be better

We applied this method to retrieve more training material for recognition systems.
- We applied this method to retrieve more training material for handwriting recognition systems. For example, we retrieved training material for handwriting recognition systems in other projects, this helped to quickly create annotated databases for handwriting recognition systems.
- In other publications, this helped to quickly create annotated databases for handwriting recognition systems.