Goal

Word Recognition

Text-to-Image Retrieval

Query: Brady

Lexicon

Times

Word Images

Our Approach

Focus on large lexicon based recognition

Multiple candidate words generation

Infering diverse solutions

Group edit distance based lexicon re-ranking

Iterative lexicon reduction

Text-to-image retrieval task

- Preprocess images by reducing the lexicon
- Retrieve word images with the query word in the reduced lexicon

Main Idea:

Strengthen pairwise terms by reducing lexicon size using diverse solutions

Lexicon Reduction

Lexicon driven recognition

Wang K et al. [ICCV ’11], Shi et al. [CVPR ’13], Wang T et al. [ICPR ’12], Mishra et al. [CVPR ’12]

- Potential character locations detected using
  - Binarization
  - Sliding window
  - Inference on graphs to recognize text
  - Small lexicons used to correct recognition

Drawbacks:

- Difficult to obtain a single set of true character windows in a graph
- Does not perform well on large lexicon settings due to weak pairwise terms

Proposed Method

1. Given an initial lexicon and diverse solutions, we obtain the next solution by adding the constraint with at least k-hamming distance from best solution.

2. Second solution obtained by modifying unary terms and inferring again.

3. Diverse solutions similar to Batra et al. [ECCV ’12].

4. Process can infer the true label even with a incorrect MAP solution.

Lexicon Reduction Process

Given an initial lexicon and diverse solutions,

I. Re-rank using group edit distances

II. Select the top-K words as reduced lexicon

Multiple Candidate Words

- Connected components from binarization as nodes
- Adjacency graph formation
- Candidate word: nodes on a path from a start to end node

Recognition

I. Given a word image, find multiple candidate words

II. Iteratively reduce the lexicon to size 10

III. Infer diverse solutions with pairwise terms from reduced lexicon

IV. Find a word in the original lexicon with least group edit distance

Word recognition accuracy comparison between various CRF and non-CRF methods

Lexicon Reduction

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>STARS</th>
<th>THIS</th>
<th>TAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diverse solutions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARS</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>TOOLS</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>THIS</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Group edit distance based re-ranking using edit distances between lexicon words and diverse solutions

Retrieval

Pre-processing stage

I. Reduce lexicons for each word image to size 4

II. Compute average edit distance (AED) by averaging the edit distances of all word pairs in the lexicon

III. If AED ≥ 8, then reduce lexicon to size 1 else keep lexicon of size 4

Retrieval stage

I. Retrieve images with the query word in their reduced lexicon

II. Rank the images based on lexicon sizes and query word position

Top-1 precision results on various datasets

Correct retrieval cases using partial reduction and diverse solutions