**DUET: Detection Utilizing Enhancement for Text in Scanned or Captured Documents**

Eun-Soo Jung, HyeongGwan Son, Kyusam Oh, Yongkeun Yun, Soonhwan Kwon, and Min Soo Kim

### Background and Motivation

- Most of the previous studies on text detection focus on text in the wild[1-6]
- **Text in the wild** (or scene text): (Relatively) More labeled data
- **Text in documents**: Very few labeled data → insufficient to train DNN
- Text detection models trained with scene text: *Limited* to cover features of document images

### Contributions

- **Text detector** with improved accuracies for **document images**
  - Synthesizing training data
  - Multi-task learning
  - Overcome various types of noise in document images

### Training

- **Main task**: text detection
- **Auxiliary task**: text enhancement
- **Weak-supervision** to train enhancement task for real data
  - Binarized detection GT (ground truth) → false positive loss
  - Using interim trained detector → detection loss for enhanced output

### Results

- **Test data**: **FUNSD** dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster R-CNN</td>
<td>70.4</td>
<td>84.8</td>
<td>76.9</td>
</tr>
<tr>
<td>EAST</td>
<td>51.6</td>
<td>84.0</td>
<td>63.9</td>
</tr>
<tr>
<td>CRAFT</td>
<td>91.2</td>
<td>84.2</td>
<td>87.6</td>
</tr>
<tr>
<td>CharNet</td>
<td>95.1</td>
<td>57.4</td>
<td>71.6</td>
</tr>
<tr>
<td><strong>DUET (proposed)</strong></td>
<td><strong>93.1</strong></td>
<td><strong>92.2</strong></td>
<td><strong>92.6</strong></td>
</tr>
</tbody>
</table>

### Output examples

- Detection
- Enhancement