Our Method

- Introducing a dictionary method [8] to increase the probability of success based on the CDistNet structure.
- Adding special parking-related words and punctuations into the dictionary to improve specificity (e.g., “9 AM”, “Mon.”, “9AM-4:30PM”)

Step 1: Input a text image x to CDistNet and obtain preliminary result y, then generate k candidates according to y from the dictionary.

Step 2: Compute compatibility scores (y, y’) = \log \text{softmax}(\text{logit}(y)) and edit distances d(y, y’). P is the probability matrix of size s \times m, where m is the max length of a word and s is the size of alphabet. y’ represents the index of the jth character of y. y’ \in \{1, 2, ..., s\}.

Step 3: Compute KL-divergence between two probability distributions D and K. LKL(D || K) = \sum_{i} p(i) \log \frac{p(i)}{q(i)}.

Step 4: Compute training loss is CDistNet training error plus KKL(D || K), and output most compatible candidate y.

Experiments

Datasets: SynthText [9] is a synthetic image dataset where the word image contains punctuations. The second dataset is from our curbside parking image dataset. The sizes of these two datasets are approximately 100,000 and 5%. The training set is SynthText plus 80% of our parking set, and the test set is the remaining 20% parking set.

Baselines:
- CDistNet (O) is an established deep neural network that has demonstrated competitive results in recognizing sequence-like images.
- CRNN: a recently proposed state-of-the-art architecture specifically designed for text recognition tasks.
- E-CNN: an edited model based on CDistNet as a variant backbone to address the limitation to handle punctuations marks.
- TOCR: a recently proposed OCR method that has gained recognition as an exceptional model for text recognition.

Table 1: Parking Sign Recognition Performance Comparison (ER: Error Rate).

<table>
<thead>
<tr>
<th>Model</th>
<th>CRNN</th>
<th>CDistNet</th>
<th>E-CNNDistNet</th>
<th>TOCR</th>
<th>Our Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER(%)</td>
<td>13.3</td>
<td>16.3</td>
<td>10.0</td>
<td>7.4</td>
<td>4.3</td>
</tr>
<tr>
<td>Reparable ER(%)</td>
<td>4.8</td>
<td>5.3</td>
<td>3.7</td>
<td>5.4</td>
<td>3.2</td>
</tr>
<tr>
<td>Irreparabler ER(%)</td>
<td>8.5</td>
<td>11.4</td>
<td>6.3</td>
<td>2.0</td>
<td>1.1</td>
</tr>
<tr>
<td>CER(%)</td>
<td>10.1</td>
<td>13.2</td>
<td>7.3</td>
<td>3.8</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Conclusions

- We propose an innovative model for recognizing parking text that outperforms other methods, particularly in the following areas:
  - Collecting a dataset for street parking sign text recognition.
  - Overcoming punctuation recognition challenges through well-designed strategies and building a parking-specific dictionary.
  - Significantly increasing accuracy for text recognition, especially for parking-specific words, like abbreviations of time words.

References