Motivation

1. The feature vectors in the background region cannot be well supervised due to their very low attention weights when training.
2. The noise points in the feature map can confuse the attention module and lead to attention drift.
3. The quality of the feature map determines the correctness of prediction results when attention drift happens.

Method

1. Denoise feature map with SEC Loss on 2D feature map.
2. Both the Transformer branch and the feature refinement branch share the same character decode layer.
3. With the supervision of Spatial Existence Classification (SEC) loss, the attention module (RNN or Transformer) can better align the character area.
SEC Loss

1. For each character in the character set, calculate its probability of appearing in the decoded feature map.
2. Encode the annotation to binary code for each character in the character set according to its appearance in the annotation.
3. Calculate the cross-entropy between value from 1. and 2.

Results

1. By supervising the feature map with SEC Loss, the model based on attention mechanism (i.e. RNN attention and Transformer attention) have constant improvement.
2. The improvements on irregular datasets are larger than that on regular datasets because irregular samples have more disturbing factors.
3. Our method achieve comparable performance with other SOTA methods.

<table>
<thead>
<tr>
<th>method</th>
<th>IIIT5K</th>
<th>SVT</th>
<th>IC03</th>
<th>IC13</th>
<th>IC15</th>
<th>SVTP</th>
<th>CT80</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNN baseline</td>
<td>89.2</td>
<td>85.4</td>
<td>92.6</td>
<td>90.1</td>
<td>71.9</td>
<td>73.6</td>
<td>72.0</td>
</tr>
<tr>
<td>RNN SEC</td>
<td>89.9 (+0.7)</td>
<td>87.5 (+2.1)</td>
<td>92.6 (+0.0)</td>
<td>90.5 (+0.4)</td>
<td>72.6 (+0.7)</td>
<td>75.0 (+1.4)</td>
<td>78.5 (+6.5)</td>
</tr>
<tr>
<td>Transformer baseline</td>
<td>91.4</td>
<td>87.4</td>
<td>93.6</td>
<td>91.2</td>
<td>75.6</td>
<td>77.6</td>
<td>79.2</td>
</tr>
<tr>
<td>Transformer SEC</td>
<td>92. (+1.5)</td>
<td>89.6 (+2.2)</td>
<td>95.3 (+1.7)</td>
<td>93.6 (+2.4)</td>
<td>79.9 (+4.3)</td>
<td>82.2 (+4.6)</td>
<td>84.3 (+5.1)</td>
</tr>
</tbody>
</table>