ICDAR 2015 Robust Reading Competition


with the support of:
ICDAR 2015 Robust Reading Challenges

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Challenge 1: Born-Digital</th>
<th>Challenge 2: Focused Scene Text</th>
<th>Challenge 3: Text in Videos</th>
<th>Challenge 4: Incidental Scene Text</th>
</tr>
</thead>
</table>
The ICDAR RRC Benchmark Datasets

<table>
<thead>
<tr>
<th></th>
<th>Challenge 1</th>
<th>Challenge 2</th>
<th>Challenge 3</th>
<th>Challenge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>420</td>
<td>229</td>
<td>25 (13,450)</td>
<td>1,000</td>
</tr>
<tr>
<td>Test Set</td>
<td>141</td>
<td>233</td>
<td>24 (14,374)</td>
<td>500</td>
</tr>
<tr>
<td>Sequestered Set</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>170</td>
</tr>
<tr>
<td>Annotated Regions</td>
<td>5,003</td>
<td>1,943</td>
<td>184,687</td>
<td>17,548</td>
</tr>
</tbody>
</table>
Research tasks

Task 1 – Text Localization

**Objective:** To obtain a rough estimation of text areas in the image, in terms of bounding boxes corresponding to parts of text (words or text lines)

Task 2 – Text Segmentation

**Objective:** Pixel-level separation of text from the background

Task 3 – Word Recognition

**Objective:** Assuming known word bounding boxes, to obtain the correct text transcriptions

Task 4 – End-to-end

**Objective:** Localise and recognise all words in the image
End-to-End tasks - Specifics

Objective: Localise and recognise all words in the image

- Vocabularies were given, including only proper words (3 characters or longer, only letters)
  - Strongly Contextualised: per-image vocabularies of 100 words including all words that appear in the image + distractors from the same subset
  - Weakly Contextualised: all proper words that appear in the entire test set, and
  - Generic: any vocabulary can be used, a 90k word vocabulary was provided

- Authors were able to submit results in any variant

- Evaluated in two modalities:
  - End-to-End recognition: the ability of a method to detect and recognise all words in the image
  - Word Spotting: the ability of a method to detect and recognise the words provided in the vocabulary (ignoring the rest)
The ICDAR 2015 Incidental Text Dataset

**Incidental text** refers to text that appears in the scene without the user having taken any prior action to cause its appearance in the field of view, or improve its positioning or quality in the frame

- 1,600 images acquired using Google glass
- Collected over a period of a few months in Singapore
- Annotated collectively by 6 institutions using the RRC annotation framework
- Latin and Orient scripts (latter currently marked as “do not care”)
## Participation in Numbers

<table>
<thead>
<tr>
<th></th>
<th>2011</th>
<th>2013</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web page views (Unique Visitors)</td>
<td>64,041 (4,575)</td>
<td>115,778 (7,342)</td>
<td>53,341 (4,773)</td>
</tr>
<tr>
<td>Registered Users</td>
<td></td>
<td>717</td>
<td>1,056</td>
</tr>
<tr>
<td>Private Submissions</td>
<td></td>
<td>2,633</td>
<td>5,292</td>
</tr>
<tr>
<td>Public Submissions</td>
<td>30</td>
<td>63</td>
<td>157</td>
</tr>
<tr>
<td>Challenge 1</td>
<td>30</td>
<td>25</td>
<td>49</td>
</tr>
<tr>
<td>Challenge 2</td>
<td>-</td>
<td>36</td>
<td>67</td>
</tr>
<tr>
<td>Challenge 3</td>
<td>-</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Challenge 4</td>
<td>-</td>
<td>-</td>
<td>27</td>
</tr>
</tbody>
</table>

**Open mode participation:** Participants run their own algorithms and provided results

*Data as of 19/8/2015*
State of the art - 2015
Trends

• Almost all submitted methods employ an initial segmentation step and text detection is obtained by classifying connected components or their groupings.
  • Almost all methods make use of the MSER segmentation algorithm.
• Top performing (pre-located) word recognition methods make use of commercial OCRs.
WHAT’S NEXT?
Robust Reading portal remains open

- The competition is open in a **continuous mode**
- **Datasets** are freely available
- Upload new results to obtain **real-time on-line evaluation**
  - **Ranking charts and tables**
  - **Visualisation** of the results of all tasks, and all challenges
- Keep your results **private** or make them **public** to the community
- The robust reading annotation platform is also available [1]


http://rrc.cvc.uab.es/
Online Evaluation and Ranking Tables

Challenge 4: "Incidental Scene Text"

You can see below the latest results, updated in real time when new submissions are made to the system. Methods published in ICDAR 2015 in grey, plus other public methods in white and your methods in yellow.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>Hmean</th>
</tr>
</thead>
<tbody>
<tr>
<td>StradVision-2</td>
<td>36.74%</td>
<td>77.46%</td>
<td>49.84%</td>
</tr>
<tr>
<td>StradVision-1</td>
<td>46.27%</td>
<td>53.39%</td>
<td>49.57%</td>
</tr>
<tr>
<td>NJU_Text_Version4</td>
<td>35.82%</td>
<td>72.73%</td>
<td>48.00%</td>
</tr>
<tr>
<td>NJU Text (Version2)</td>
<td>36.25%</td>
<td>70.44%</td>
<td>47.87%</td>
</tr>
<tr>
<td>AJOU</td>
<td>46.94%</td>
<td>47.26%</td>
<td>47.10%</td>
</tr>
<tr>
<td>NJU_Text_Version1</td>
<td>38.32%</td>
<td>56.33%</td>
<td>45.62%</td>
</tr>
<tr>
<td>NJU_Text_Version2</td>
<td>37.46%</td>
<td>54.14%</td>
<td>44.28%</td>
</tr>
<tr>
<td>NJU_Text_Version5</td>
<td>37.84%</td>
<td>51.41%</td>
<td>43.59%</td>
</tr>
<tr>
<td>HUST MCLAB (VER...)</td>
<td>37.79%</td>
<td>44.00%</td>
<td>40.66%</td>
</tr>
<tr>
<td>HUST_MCLAB_VER1_0</td>
<td>34.81%</td>
<td>47.47%</td>
<td>40.17%</td>
</tr>
<tr>
<td>HUST_MCLAB_VER1_2</td>
<td>34.09%</td>
<td>46.49%</td>
<td>39.33%</td>
</tr>
<tr>
<td>Deep2Text-MO</td>
<td>32.11%</td>
<td>49.59%</td>
<td>38.98%</td>
</tr>
<tr>
<td>CNN Proposal Bas...</td>
<td>34.42%</td>
<td>34.71%</td>
<td>34.57%</td>
</tr>
<tr>
<td>TD-MU</td>
<td>25.28%</td>
<td>34.56%</td>
<td>29.20%</td>
</tr>
<tr>
<td>TextCatcher-2 (L...)</td>
<td>34.81%</td>
<td>24.91%</td>
<td>29.04%</td>
</tr>
<tr>
<td>imagine</td>
<td>5.44%</td>
<td>12.97%</td>
<td>7.67%</td>
</tr>
</tbody>
</table>
Per-image Results

Task 4 - End-to-End - Method: Stradvision-2

Images:
- image 1: recall 100%, precision 0%, hmean 0%
- image 2: recall 70%, precision 87.5%, hmean 77.78%
- image 3: recall 0%, precision 0%, hmean 0%
- image 4: recall 0%, precision 0%, hmean 0%
- image 5: recall 100%, precision 100%, hmean 100%
- image 6: recall 0%, precision 0%, hmean 0%
- image 7: recall 100%, precision 100%, hmean 100%
- image 8: recall 0%, precision 0%, hmean 0%
- image 9: recall 100%, precision 100%, hmean 100%
- image 10: recall 55.56%, precision 71.43%, hmean 62.5%
- image 11: recall 100%, precision 100%, hmean 100%
- image 12: recall 0%, precision 0%, hmean 0%
- image 13: recall 25%, precision 100%, hmean 40%
- image 14: recall 60%, precision 100%, hmean 75%
- image 15
- image 16
- image 17
- image 18
- image 19
- image 20
- image 21
Visualisation Tools

ICDAR 2015
ROBUST READING COMPETITION

Challenge 4: "Incidental Scene Text"

Task 4 - End-to-End - Method: Stradvision-2

Ground Truth

Detection

Method comparation

End-to-End recognition

Word Spotting Type: Strongly

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>4013 Stradvision-2</td>
<td>70</td>
<td>87.5</td>
<td>77.78</td>
</tr>
<tr>
<td>3962 StradVision_v1</td>
<td>70</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>3854 Beam search C...</td>
<td>30</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>3823 Baseline.TextSp...</td>
<td>30</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>3790 NJU Text (Version...)</td>
<td>30</td>
<td>42.86</td>
<td>35.29</td>
</tr>
<tr>
<td>4288 Baseline.OpenC...</td>
<td>20</td>
<td>100</td>
<td>33.33</td>
</tr>
<tr>
<td>3883 Deep2Text-MO</td>
<td>20</td>
<td>28.57</td>
<td>23.53</td>
</tr>
<tr>
<td>3855 Beam search C...</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Evaluation Log

#GT polygons: 22
#DET polygons: 8
Match found D#1 G#7 74.4.
Match found D#2 G#11 80.4.
Match found D#3 G#2 65.3.
Match found D#4 G#13 88.7.
Match found D#5 G#1 94.1.
Discarded D#6 G#3 18.7.
Discarded D#6 G#4 28.5.
Discarded D#6 G#5 23.3.
NO Match for D#6.
Match found D#7 G#8 81.6.

Institute for Infocomm Research
Visualisation Tools

ICDAR 2015
ROBUST READING COMPETITION

Challenge 2: "Focused Scene Text"

Task 2 - Text Segmentation - Method: StradVision

**Ground Truth**
- ○ Image
- ○ Atoms GT
- ○ Pixel GT
- ○ Atoms Classification

**Detection**
- ○ Image
- ○ Atoms Det.
- ○ Pixels Eval.
- ○ Atoms Classification

Method Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>Pixel Results</th>
<th>Atom Based Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>USTB_FuStar</td>
<td>84.31</td>
<td>99.89</td>
</tr>
<tr>
<td>Text Detection</td>
<td>82.37</td>
<td>99.95</td>
</tr>
<tr>
<td>I2R_NUS_FAR</td>
<td>79.71</td>
<td>100</td>
</tr>
<tr>
<td>I2R_NUS</td>
<td>79.71</td>
<td>100</td>
</tr>
<tr>
<td>BUCT_VST</td>
<td>78.63</td>
<td>99.95</td>
</tr>
<tr>
<td>StradVision</td>
<td>77.17</td>
<td>99.97</td>
</tr>
<tr>
<td>NTextExtractor</td>
<td>75.8</td>
<td>99.94</td>
</tr>
<tr>
<td>NSTSegmentator</td>
<td>75.8</td>
<td>93.86</td>
</tr>
<tr>
<td>OTCYMIST</td>
<td>0.49</td>
<td>8.37</td>
</tr>
</tbody>
</table>

Previous | Image 11 of 233 | Next | Full Screen
The CVC Annotation and Performance Evaluation Platform

Get the 2013 version from: http://www.cvc.uab.es/apep
Task 1.4 Born-Digital End-to-End (Strongly contextualised)
Stradvision

Task 1.4 Born-Digital End-to-End (Weakly contextualised)
University of Science and Technology Beijing / Xi’an Jiaotong - Liverpool University

Task 1.4 Born-Digital End-to-End (Generic)
“Deep2Text-II”, X. C. Yin, C. Yang, J. B. Hou, W. Y., Pei, X. Yin, and K. Huang
University of Science and Technology Beijing / Xi’an Jiaotong - Liverpool University

Task 2.4 Focused Text End-to-End (Strongly contextualised)
“VGGMaxBBNet”, A. Gupta, M. Jaderberg, A. Zisserman,
Visual Geometry Group, University of Oxford

Task 2.4 Focused Text End-to-End (Weakly contextualised)
Stradvision

Task 2.4 Focused Text End-to-End (Generic)
“Deep2Text-II”, X. C. Yin, C. Yang, J. B. Hou, W. Y., Pei, X. Yin, and K. Huang
University of Science and Technology Beijing / Xi’an Jiaotong - Liverpool University

Thanks for participating!
Winners

Task 3.1 Video Text Localisation
“Deep2Text-I”, X. C. Yin, C. Yang, J. B. Hou, W. Y., Pei, X. Yin, and K. Huang
*University of Science and Technology Beijing / Xi’an Jiaotong - Liverpool University*

Task 3.4 Video Text End-to-End (honourable mention)
*Stradvision*

Task 4.1 Incidental Scene Text Localisation
*Stradvision*

Task 4.3 Incidental Text Recognition
“MAPS”, D. Kumar and A. G. Ramakrishnan
*Dayananda Sagar Institutions and Indian Institute of Science*

Task 4.4 Incidental Text End-to-End (Strongly contextualised)
*Stradvision*

Task 4.4 Incidental Text End-to-End (Weakly contextualised) (honourable mention)
“Beam Search CUNI”, J. Libovický, and P. Pecina,
*Charles University in Prague*

Task 4.4 Incidental Text End-to-End (Generic)
“Beam Search CUNI”, J. Libovický, and P. Pecina,
*Charles University in Prague*

Thanks for participating!
GROUND TRUTH AND PERFORMANCE EVALUATION
Ground Truth – Text Localization

**Dataset Images**

- How to Find the Perfect HDTV
- The Photo Specialists
- HIGHER SAVINGS RATES

**Ground Truth (text files)**

```
11, 15, 42, 28, "How"
41, 16, 61, 28, "to"
8, 33, 36, 46, "Find"
41, 32, 64, 46, "the"
11, 50, 61, 65, "Perfect"
16, 69, 56, 82, "HDTV"
```

```
22 249 113 286 "The"
142 249 287 286 "Photo"
326 245 620 297 "Specialists"
```

```
0, 1, 177, 32, "\"HIGHER\"
11, 35, 182, 63, "SAVINGS"
12, 67, 183, 103, "RATES\"
50, 114, 56, 120, "A"
60, 114, 91, 120, "reward"
96, 114, 108, 120, "for"
112, 116, 126, 120, "our"
130, 114, 164, 120, "current"
...```

**Ground Truth Visualisation**

- Visual representation of the text localization results
Evaluation Protocol – Text Localization

• Challenges 1 and 2: **DetEval** Methodology
  • Takes into account both
    • Bounding Box area overlapping and
    • Precision at the level of detection counts
  • Deals with **one-to-many** and **many-to-one** cases
  • Set up to penalise **over-segmentation** (words split to parts), but no under-segmentation (group of words detected as text line)

• Challenge 4: **Intersection over Union** above 0.5
  • Word granularity, penalises any other behaviour

• Areas marked as “**Do Not Care**” are not taken into account during evaluation

---


Ground Truth – Text Segmentation

Dataset Images

Ground Truth (images)

okcupid

okcupid

How to Find the Perfect HDTV

How to Find the Perfect HDTV

The Photo Specialists

The Photo Specialists
Evaluation Protocol – Text Segmentation

Performance Evaluation Methodology

• Framework proposed by Clavelli et al [2]
• Measures the degree to which morphological properties of the text are preserved, not simply the number of misclassified pixels
• As a secondary evaluation scheme we implemented standard pixel level precision and recall (compatibility with other results)

Ground Truth – Word Recognition

**Dataset Images**

<table>
<thead>
<tr>
<th>Image</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>word_1.png</td>
<td>&quot;$500&quot;</td>
</tr>
<tr>
<td>word_2.png</td>
<td>&quot;who&quot;</td>
</tr>
<tr>
<td>word_3.png</td>
<td>&quot;SMRT&quot;</td>
</tr>
<tr>
<td>word_4.png</td>
<td>&quot;COACH&quot;</td>
</tr>
<tr>
<td>word_5.png</td>
<td>&quot;FALL&quot;</td>
</tr>
<tr>
<td>word_6.png</td>
<td>&quot;toast?&quot;</td>
</tr>
<tr>
<td>word_7.png</td>
<td>&quot;SEASON!&quot;</td>
</tr>
<tr>
<td>word_8.png</td>
<td>&quot;HUMP&quot;</td>
</tr>
<tr>
<td>word_9.png</td>
<td>&quot;OUT&quot;</td>
</tr>
<tr>
<td>word_10.png</td>
<td>&quot;#04-11&quot;</td>
</tr>
<tr>
<td>word_11.png</td>
<td>&quot;NEW&quot;</td>
</tr>
<tr>
<td>word_12.png</td>
<td>&quot;PLAIN&quot;</td>
</tr>
<tr>
<td>word_13.png</td>
<td>&quot;TOBACCO&quot;</td>
</tr>
</tbody>
</table>

**Ground Truth transcription**

<table>
<thead>
<tr>
<th>Image</th>
<th>Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>word_1.png</td>
<td>&quot;$500&quot;</td>
</tr>
<tr>
<td>word_2.png</td>
<td>&quot;who&quot;</td>
</tr>
<tr>
<td>word_3.png</td>
<td>&quot;SMRT&quot;</td>
</tr>
<tr>
<td>word_4.png</td>
<td>&quot;COACH&quot;</td>
</tr>
<tr>
<td>word_5.png</td>
<td>&quot;FALL&quot;</td>
</tr>
<tr>
<td>word_6.png</td>
<td>&quot;toast?&quot;</td>
</tr>
<tr>
<td>word_7.png</td>
<td>&quot;SEASON!&quot;</td>
</tr>
<tr>
<td>word_8.png</td>
<td>&quot;HUMP&quot;</td>
</tr>
<tr>
<td>word_9.png</td>
<td>&quot;OUT&quot;</td>
</tr>
<tr>
<td>word_10.png</td>
<td>&quot;#04-11&quot;</td>
</tr>
<tr>
<td>word_11.png</td>
<td>&quot;NEW&quot;</td>
</tr>
<tr>
<td>word_12.png</td>
<td>&quot;PLAIN&quot;</td>
</tr>
<tr>
<td>word_13.png</td>
<td>&quot;TOBACCO&quot;</td>
</tr>
</tbody>
</table>

**Ground Truth location (ONLY Challenge 4)**

<table>
<thead>
<tr>
<th>Image</th>
<th>X</th>
<th>Y</th>
<th>Width</th>
<th>Height</th>
<th>X</th>
<th>Y</th>
<th>Width</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>word_1.png</td>
<td>0,18,88</td>
<td>0,90,50</td>
<td>2,68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_2.png</td>
<td>23,13,229</td>
<td>0,207,138</td>
<td>0,152</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_3.png</td>
<td>8,22,152</td>
<td>0,146,57</td>
<td>0,90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_4.png</td>
<td>0,96,153</td>
<td>0,178,40,26,136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_5.png</td>
<td>0,50,116</td>
<td>0,152,83,3,122</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_6.png</td>
<td>1,0,63,16,62,41,0,26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_7.png</td>
<td>0,5,82,0,83,24,1,29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_8.png</td>
<td>9,8,349,0,340,83,0,91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_9.png</td>
<td>0,41,86,0,101,56,16,97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_10.png</td>
<td>0,21,70,0,76,29,6,50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_11.png</td>
<td>0,4,91,0,91,28,0,32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_12.png</td>
<td>0,90,41,0,72,6,27,96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>word_13.png</td>
<td>0,0,100,24,105,39,4,15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**gt.txt**

**coords.txt**
Evaluation Protocol – Word Recognition

Performance Evaluation Methodology

- **Edit distance** (normalised to the length of the ground truth transcription)
- Equal weights for deletions, additions, substitutions
- We also report statistics on correctly recognised words (case sensitive)

Baseline method

- Commercial OCR package (ABBYY OCR SDK v.10)
Ground truth – End-to-end

Dataset Images

Ground Truth (text files)

341, 129, 399, 117, 395, 153, 337, 165, East
405, 113, 479, 97, 479, 133, 403, 147, West
491, 92, 561, 75, 559, 118, 487, 132, Line
444, 235, 539, 228, 534, 296, 439, 303, out

gt_img_1.txt

931, 400, 1040, 348, 1052, 374, 943, 426, LEADERSHIP
913, 460, 1005, 417, 1016, 444, 924, 487, THROUGH
1009, 416, 1076, 382, 1087, 407, 1020, 442, ###

gt_img_2.txt
Evaluation Protocols – End to End

Combination of text localisation and word recognition

- Localisation: Intersection over Union over 0.5
- Recognition: words match (100%) ignoring the case
- Detecting or missing words marked as “do not care” will not affect (positively or negatively) the results
Ground truth - Video Challenge

Frame Image

Ground Truth Visualisation

Ground Truth (XML file)

<?xml version="1.0" encoding="us-ascii"?>
<Frames>
  <frame ID="1" />
  <frame ID="2" />
  <frame ID="3" />
  ...
  <frame ID="434">
    <object Transcription="Badia" ID="377001" Language="Catalan" Quality="MODERATE">
      <Point x="446" y="58" />
      <Point x="447" y="69" />
      <Point x="484" y="67" />
      <Point x="484" y="55" />
    </object>
    <object Transcription="Barbera" ID="377002" Language="Catalan" Quality="MODERATE">
      <Point x="502" y="81" />
      <Point x="447" y="85" />
      <Point x="446" y="73" />
      <Point x="502" y="70" />
    </object>
    ...
  </frame>
  ...
  <frame ID="1859" />
  <frame ID="1860" />
</Frames>
Evaluation Protocols – Video Challenge

- Evaluation based on CLEAR-MOT [3, 4], and VACE [5] metrics
  - Multiple Object Tracking Precision (MOTP), expresses how well locations of words are estimated
  - Multiple Object Tracking Accuracy (MOTA), how many mistakes the tracker has made (false positives, ID mismatches, etc), can be negative
  - Average Tracking Accuracy (ATA), comprehensive spatio-temporal measure.

Baseline Method

- Detection stage performed using the ABBYY OCR SDK, followed by a tracking stage where each detected word is assigned the identifier of the previously detected word with the best overlapping ratio