in collaboration with:

ICDAR 2015 Robust Reading Competition

D. Karatzas, L. Gomez, A. Nicolaou, S. Ghosh, A. Bagdanov,
M. Iwamura, J. Matas, L. Neumann, V.R. Chandrasekhar,
S. Lu, F. Shafait, S. Uchida, E. Valveny

with the support of:

ICDAR 2015 Robust Reading Challenges

Tasks	**Challenge 1:** Born-Digital	**Challenge 2:** Focused Scene Text	**Challenge 3:** Text in Videos	**Challenge 4:** Incidental Scene Text
1. Localization / Tracking	2011 / 2013	2011 / 2013	2013 / **2015**	**2015**
2. Segmentation	2011 / 2013	2013		
3. Recognition	2011 / 2013	2011 / 2013		**2015**
4. End-to-end	**2015**	**2015**	**2015**	**2015**

The ICDAR RRC Benchmark Datasets

	Challenge 1	Challenge 2	Challenge 3	Challenge 4
Training Set	420	229	25 (13,450)	1,000
Test Set	141	233	24 (14,374)	500
Sequestered Set	-	-	-	170
Annotated Regions	5,003	1,943	184,687	17,548

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

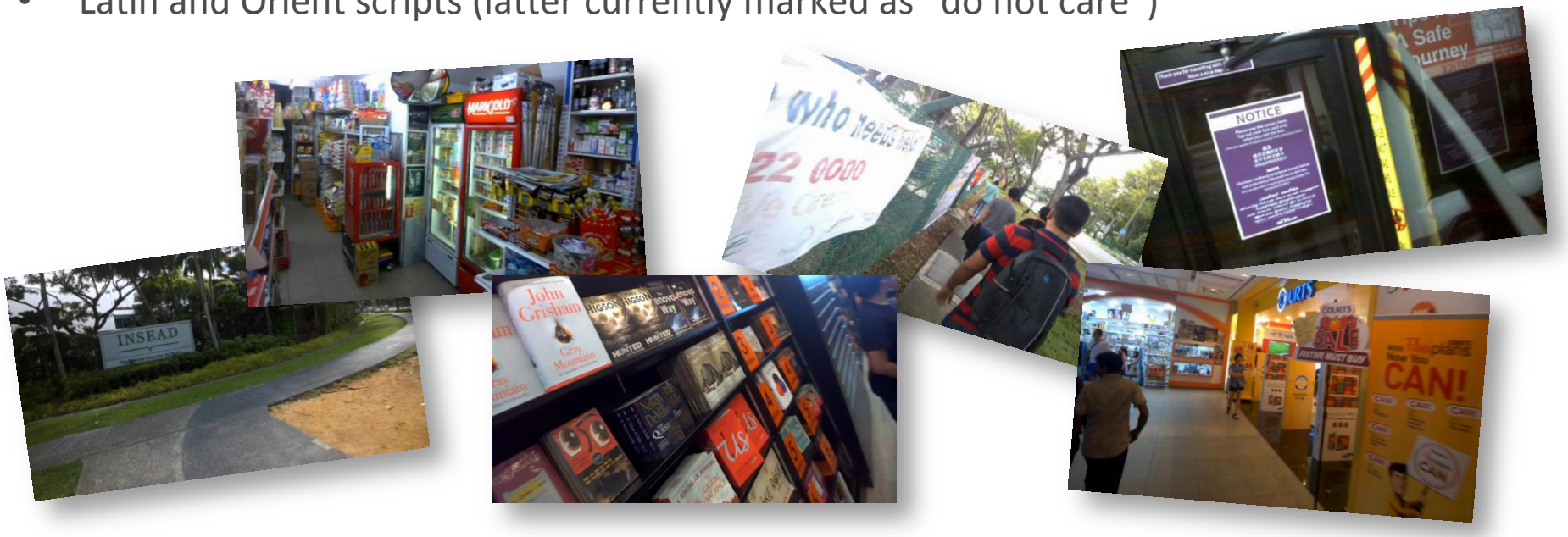
Objective: Localise and recognise all words in the image

- 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”)

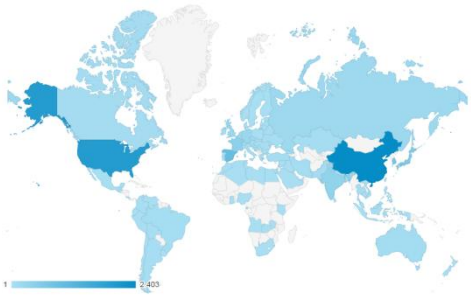


A*STAR Institute for Infocomm Research
Czech Technical University in Prague
The University of Western Australia
大阪府立大学 Osaka Prefecture University
九州大学
CVC Centre de Visió per Computador

Participation in Numbers

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

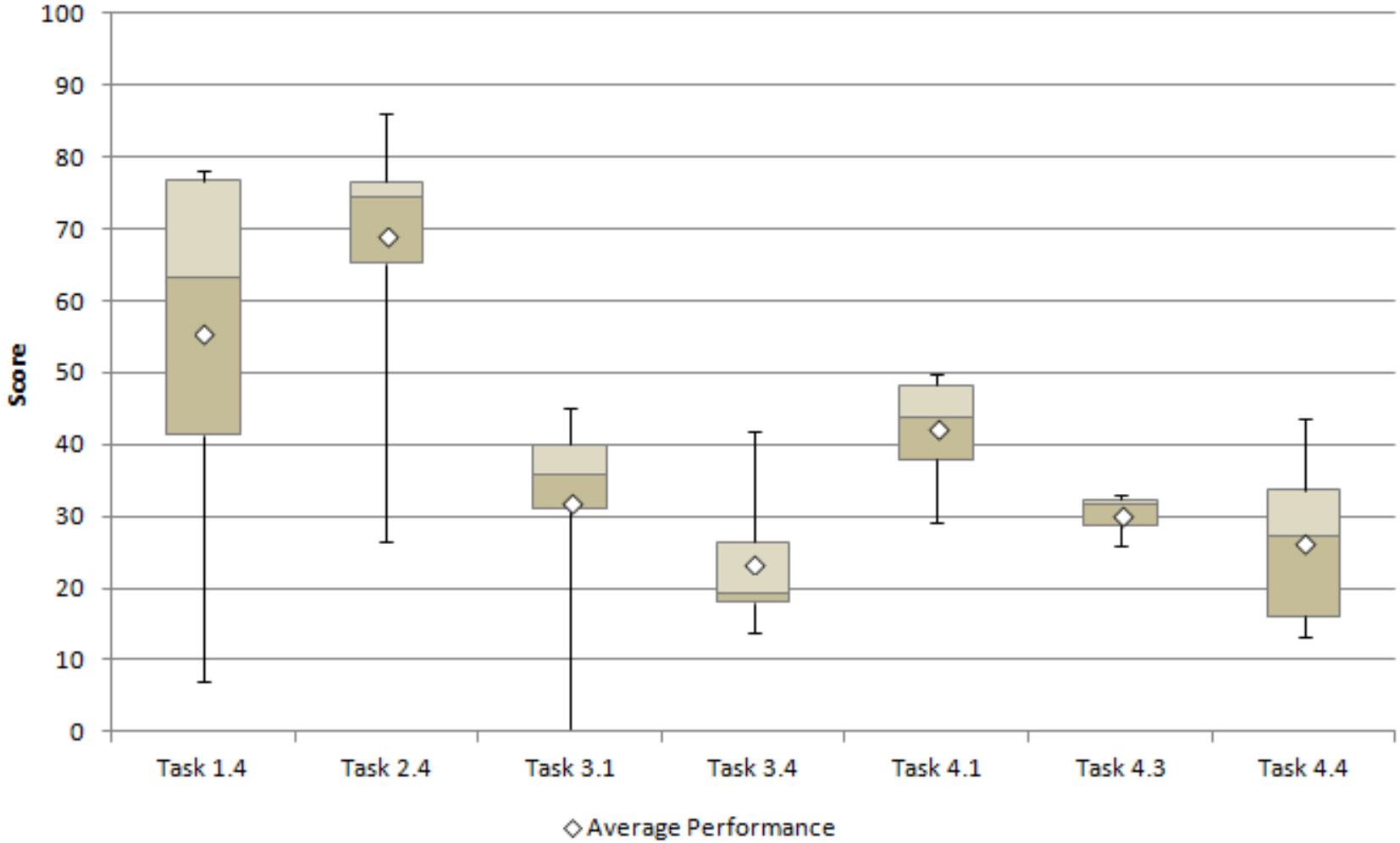
Open mode participation: Participants run their own algorithms and provided results



Data as of 19/8/2015

RESULTS

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]

[1] D. Karatzas, S. Robles, Ll. Gomez, "*An on-line platform for ground truthing and performance evaluation of text extraction systems*", DAS 2014

http://rrc.cvc.uab.es/

Institute for Infocomm Research

Online Evaluation and Ranking Tables

ICDAR 2015
ROBUST READING COMPETITION

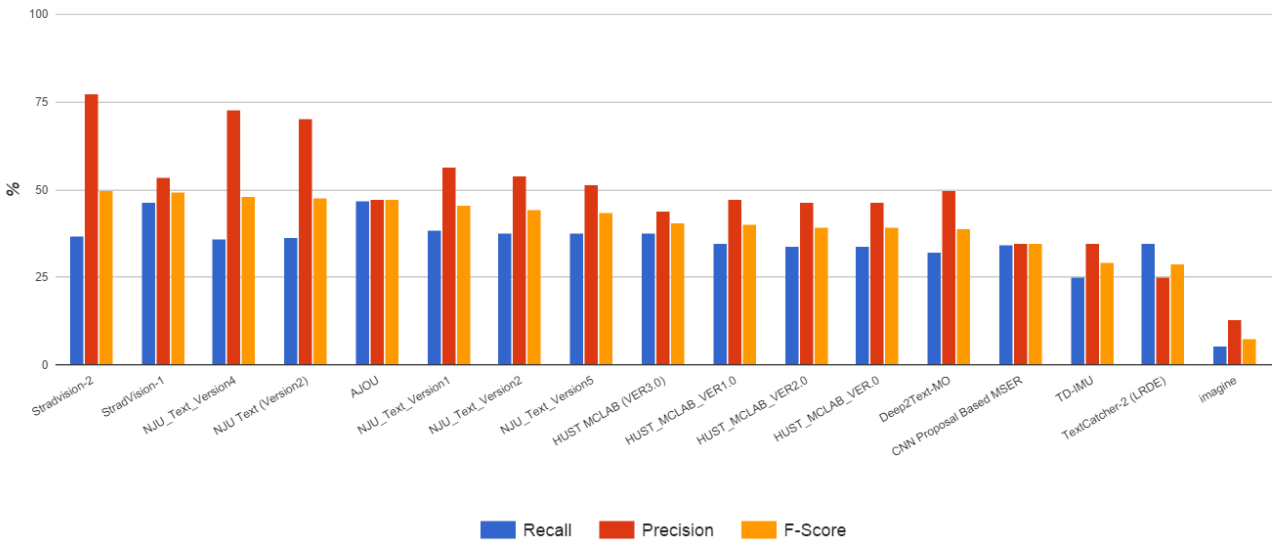
Home Procedure Challenge 1 Challenge 2 Challenge 3 Challenge 4 News F.A.Q. Contact User Users Methods LOG Mailing

Challenge 4: "Incidental Scene Text" Overview Tasks Downloads Results My Methods

Task 1 - Text Localization | Task 3 - Word Recognition | Task 4 - End-to-End

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.

Method	Recall	Precision	Hmean
Stradvision-2	36.74 %	77.46 %	49.84 %
StradVision-1	46.27 %	53.39 %	49.57 %
NJU_Text_Version4	35.82 %	72.73 %	48.00 %
NJU Text (Version2)	36.25 %	70.44 %	47.87 %
AJOU	46.94 %	47.26 %	47.10 %
NJU_Text_Version1	38.32 %	56.33 %	45.62 %
NJU_Text_Version2	37.46 %	54.14 %	44.28 %
NJU_Text_Version5	37.84 %	51.41 %	43.59 %
HUST MCLAB (VER...	37.79 %	44.00 %	40.66 %
HUST_MCLAB_VER...	34.81 %	47.47 %	40.17 %
HUST_MCLAB_VER...	34.09 %	46.49 %	39.33 %
HUST_MCLAB_VER...	34.09 %	46.49 %	39.33 %
Deep2Text-MO	32.11 %	49.59 %	38.98 %
CNN Proposal Bas...	34.42 %	34.71 %	34.57 %
TD-IMU	25.28 %	34.56 %	29.20 %
TextCatcher-2 (LR...	34.81 %	24.91 %	29.04 %
imagine	5.44 %	12.97 %	7.67 %



Per-image Results

ICDAR 2015
ROBUST READING COMPETITION

Home Procedure Challenge 1 Challenge 2 Challenge 3 Challenge 4 News F.A.Q. Contact User Users Methods LOG Mailing

Challenge 4: "Incidental Scene Text" Overview Tasks Downloads Results My Methods

Task 4 - End-to-End - Method: Stradvision-2 | Method Info | Images list | Image details

Images | Table | End-to-End recognition | Word Spotting | Strongly | not 100% F-score | not 100% Precision | not 100% Recall

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

A*STAR Institute for Infocomm Research
Czech Technical University in Prague
The University of Western Australia
大阪府立大学 Osaka Prefecture University
九州大学
CVC Centre de Visió per Computador

Visualisation Tools

ICDAR 2015
ROBUST READING COMPETITION

Home | Procedure | Challenge 1 | Challenge 2 | Challenge 3 | Challenge 4 | News | F.A.Q. | Contact | User | Users | Methods | LOG | Mailing

Challenge 4: "Incidental Scene Text" — Overview | Tasks | Downloads | Results | My Methods

Task 4 - End-to-End - Method: Stradvision-2 — Method Info | Images list | Image details

Previous | Image 2 of 500 | Next | Show image

Ground Truth | Detection

Method comparation

End-to-End recognition | Word SpottingType: Strongly

Method	Recall	Precision	score
4013 Stradvision-2	70	87.5	77.78
3962 StradVision_v1	70	28	40
3854 Beam search C...	30	60	40
3823 Baseline-TextSp...	30	60	40
3790 NJU Text (Versio...	30	42.86	35.29
4288 Baseline OpenC...	20	100	33.33
3883 Deep2Text-MO	20	28.57	23.53
3855 Beam search C...	0	0	0

#4	0	0	0	0	0	28.5	0	0
#5	0	0	0	0	0	23.3	0	0
#6	0	0	0	0	0	0	0	0
#7	74.4	0	0	0	0	0	0	0
#8	0	0	0	0	0	0	81.6	0
#9	0	0	0	0	0	0	0	75
#10	0	0	0	0	0	0	0	0
#11	0	80.4	0	0	0	0	0	0
#12	0	0	16.5	0	0	0	0	0
#13	0	0	0	88.7	0	0	0	0
#14	0	0	0	0	0	0	0	0
#15	0	0	0	0	0	0	0	0
#16	0	0	0	0	0	0	0	0
#17	0	0	0	0	0	0	0	0

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.

Visualisation Tools

ICDAR 2015
ROBUST READING COMPETITION

Home | Procedure | Challenge 1 | Challenge 2 | Challenge 3 | Challenge 4 | News | F.A.Q. | Contact | User | Users | Methods | LOG | Mailing

Challenge 2: "Focused Scene Text" — Overview | Tasks | Downloads | Results | My Methods

Task 2 - Text Segmentation - Method: StradVision — Method Info | Images list | Image details

Previous | Image 11 of 233 | Next | Full Screen

Ground Truth

Image | Atoms GT | Pixel GT | Atoms Classif.

Atoms classification.

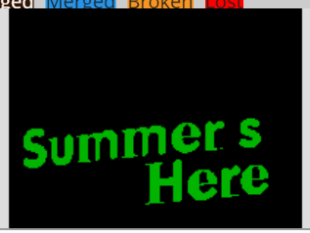
Well segmented | Broken & Merged | Merged | Broken | Lost

Detection

Image | Atoms Det. | Pixels Eval. | Atoms Classif.

Atoms classification.

Well segmented | Broken & Merged | Merged | Broken | Lost



Method comparation

	Pixel Results			Atom Based Results									
Method	Recall	Precision	FScore	Well seg	Merged	Broken	Br. Merg	Lost	False P.	Det. Atoms	Recall	Precision	FScore
USTB_FuStar	84.31	99.89	91.44	13	0	0	0	3	0	13	81.25	100	89.66
Text Detection	82.37	99.95	90.31	11	0	0	0	5	0	12	68.75	91.67	78.57
I2R_NUS_FAR	79.71	100	88.71	11	0	0	0	5	0	11	68.75	100	81.48
I2R_NUS	79.71	100	88.71	11	0	0	0	5	0	11	68.75	100	81.48
BUCT_YST	78.63	99.95	88.02	11	0	0	0	5	0	11	68.75	100	81.48
StradVision	77.17	99.97	87.1	11	0	0	0	5	0	11	68.75	100	81.48
NSTextractor	75.8	99.94	86.21	11	0	0	0	5	0	12	68.75	91.67	78.57
NSTsegmentator	75.8	93.86	83.87	11	0	0	0	5	8	20	68.75	55	61.11
OTCYMIST	0.49	8.37	0.93	0	2	0	0	14	12	13	0	0	0

A*STAR Institute for Infocomm Research | Czech Technical University in Prague | The University of Western Australia | 大阪府立大学 Osaka Prefecture University | 九州大学 | CVC Centre de Visió per Computador

The CVC Annotation and Performance Evaluation Platform

TextExtraction

User: admin Logout

TextExtraction > Ground Truthing > Google Glass > 20150105_130858_130858_1_47923_frames_44.jpg

Classify Words

Google Glass
added by admin
4 points B.B.
PUBLIC

< Prev | Image 1124 of 1670 | Next >

comments: Save Cancel

2015-02-10 eyedea3 19 12
v. 1 l. w.
view GT Words Delete

Write the selected word/line transcription

Tools

Options: Image, Line Rects

Import

CANVAS

Fit to area Real size

ratio: 0.61

LABELED OBJECTS

- l: 1/2
  - w: 1/2
- l: price
  - w: price
- l: sale
  - w: sale
- l:
  - w:
- l:
  - w:
- l:
  - w:

Selection

Components:

Transcription: price

Don't care

Type: Word (only B.B.) Line (only B.B.)

Tags

Ok Cancel

Draws: 83 Canvas: 1042x437 Original image: 1280x720 Current image size: 777x437 Offset: 133x0 Mouse: 615x6 Mouse: 794x9 Image vis.= l:0, t:0, r:1280, b:720, w:1281, h:721

Get the 2013 version from: http://www.cvc.uab.es/apep

Winners

Thanks for participating!

Task 1.4 Born-Digital End-to-End (Strongly contextualised)
"**Stradvision-2**", H. Cho, M. Sung, and B. Jun
Stradvision

Task 1.4 Born-Digital End-to-End (Weakly contextualised)
"**Deep2Text-II**", X. C. Yin, C. Yang, J. B. Hou, W. Y., Pei, Z.-H. Xing, X. Yin and K. Huang
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-1**", H. Cho, M. Sung, and B. Jun
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

Winners

Thanks for participating!

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-I**", H. Cho, M. Sung, and B. Jun
Stradvision

Task 4.1 Incidental Scene Text Localisation
"**Stradvision-2**", H. Cho, M. Sung, and B. Jun
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-2**", H. Cho, M. Sung, and B. Jun
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

GROUND TRUTH AND PERFORMANCE EVALUATION

Ground Truth – Text Localization

Dataset Images

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"
```

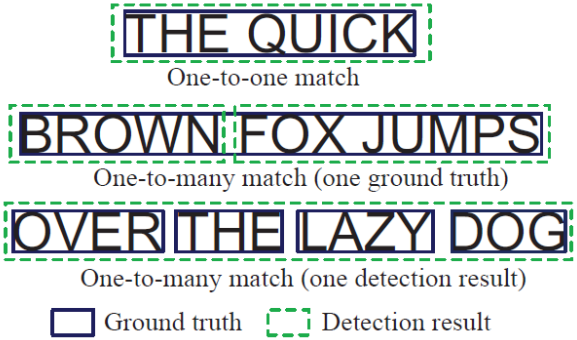
```
22 249 113 286 "The"
142 249 287 286 "Photo"
326 245 620 297 "Specialists"
```

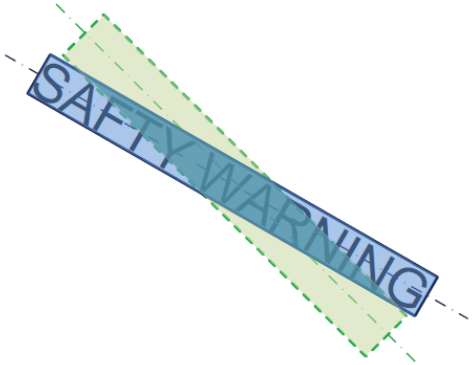
```
0, 1, 177, 32, "\"HIGHER"
11, 35, 182, 63, "SAVINGS"
12, 67, 183, 103, "RATES\""
50, 114, 56, 120, "A"
60, 114, 91, 120, "reward"
96, 114, 108, 120, "for"
112, 116, 126, 120, "our"
130, 114, 164, 120, "current"
...
```

Ground Truth Visualisation

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

C. Wolf and J.M. Jolion, "Object Count / Area Graphs for the Evaluation of Object Detection and Segmentation Algorithms", IJDAR 8(4), pp. 280-296, 2006

M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, (2014). The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 111(1), 98-136.

Ground Truth – Text Segmentation

Dataset Images | *Ground Truth (images)*

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)

(a) (b) (c)

A. Clavelli, D. Karatzas and J. Lladós "A Framework for the Assessment of Text Extraction Algorithms on Complex Colour Images", DAS 2010, pp. 19-28, 2010

Ground Truth – Word Recognition

Dataset Images

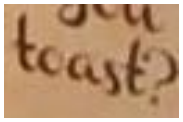
word_1.png

word_2.png

word_3.png

word_4.png

word_5.png

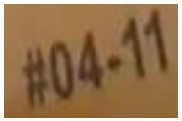


word_6.png

word_7.png

word_8.png

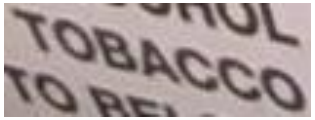
Word_9.png



word_10.png

word_11.png

word_12.png



word_13.png

Ground Truth transcription

```
word_1.png, "$500"
word_2.png, "who"
word_3.png, "SMRT"
word_4.png, "COACH"
word_5.png, "FALL"
word_6.png, "toast?"
word_7.png, "SEASON!"
word_8.png, "HUMP"
word_9.png, "OUT"
word_10.png, "#04-11"
word_11.png, "NEW"
word_12.png, "PLAIN"
word_13.png, "TOBACCO"
...
```

gt.txt

Ground Truth location (ONLY Challenge 4)

```
word_1.png,0,18,88,0,90,50,2,68
word_2.png,23,13,229,0,207,138,0,152
word_3.png,8,22,152,0,146,57,0,90
word_4.png,0,96,153,0,178,40,26,136
word_5.png,0,50,116,0,152,83,3,122
word_6.png,1,0,63,16,62,41,0,26
word_7.png,0,5,82,0,83,24,1,29
word_8.png,9,8,349,0,340,83,0,91
word_9.png,0,41,86,0,101,56,16,97
word_19.png,0,21,70,0,76,29,6,50
word_11.png,0,4,91,0,91,28,0,32
word_12.png,0,90,41,0,72,6,27,96
word_13.png,0,0,100,24,105,39,4,15
...
```

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

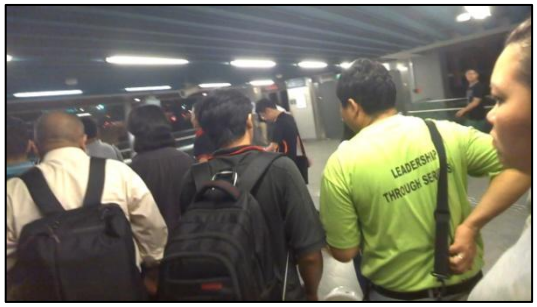
Dataset Images

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
338,248,433,242,423,319,328,325,Way
444,235,539,228,534,296,439,303,out
```

img_1.jpg

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,###
```

img_2.jpg

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 (XMLfile)

```
<?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

D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. Gomez, S. Robles, J. Mas, D. Fernandez, J. Almazan, L.P. de las Heras, ”ICDAR 2013 Robust Reading Competition,” in Proc. IEEE Int’l Conf. Document Analysis and Recognition, pp. 1484-1493, 2013