The Robust Reading Competition Annotation and Evaluation Platform

Dimosthenis Karatzas, Lluis Gómez Marçal Rusiñol

Computer Vision Centre, Universitat Autonoma de Barcelona, Barcelona, Spain;
{dimos, lgomez, marcal}@cvc.uab.es

***Abstract*—The ICDAR Robust Reading Competition (RRC), initiated in 2003 and re-established in 2011, has become the de-facto evaluation standard for the international community.**

Concurrent with its second incarnation in 2011, a continuous effort started to develop an online framework to facilitate the hosting and management of competitions.

This short paper briefly outlines the Robust Reading Competition Annotation and Evaluation Platform, the backbone of the Robust Reading Competition, comprising a collection of tools and processes that aim to simplify the management and annotation of data, and to provide online and offline performance evaluation and analysis services.

I. Introduction

The Robust Reading Competition (RRC) series[1] addresses the need to quantify and track progress in the domain of text extraction from a variety of text containers like born-digital images, real scenes, and videos. The competition was initiated in 2003 by Simon Lucas et al. [1] initially focusing only on scene text detection and recognition, and extended to include challenges on born-digital images [2], video sequences [3], and incidental scene text [4]. The 2017 edition of the Competition, under way at the time of writing, introduces five new challenges: a challenge on scene text detection and recognition based on the COCO-Text dataset, the largest scene text dataset currently available [5]; a challenge on text extraction from biomedical literature figures based on the DeText dataset [6]; a challenge on video scene text localization and recognition on the Downtown Osaka Scene Text (DOST) dataset [7]; a challenge on constrained real world end-to-end scene-text understanding based on the $> 1M$ images French Street Name Signs (FSNS) dataset [8]; and a challenge on Multi-lingual scene text detection and script identification [9].

Over the past six years, the competition has grown steadily, reaching more than 3,000 registered users from more than 80 countries by mid-2017, who have submitted more than 10,000 results that have been automatically evaluated online. Out of these, 424 have been made public by their authors. A summary of the submissions made is given in Table I. The portal receives and evaluates on average 10-20 new submissions per day. In terms of visibility, the RRC Web portal has received 360K page views from 21K users over the past four years.

To manage all the above Challenges and respond to the increasing demand, the Computer Vision Centre has invested significant resources to the development of the RRC Annotation and Evaluation Platform, which is the backbone of the competition and is briefly introduced next.

[1]http://rrc.cvc.uab.es/

II. The RRC Annotation and Evaluation Platform

The RRC Annotation and Evaluation Platform, is a collection of tools and processes that aim to facilitate the generation of data, the definition of performance evaluation metrics for different research tasks and the visualisation and analysis of results. The interface has evolved over time to support image annotation at different levels, provide version control and coordination mechanisms between ground-truthers and facilitate the verification of the final annotations. All online software tools are implemented as HTML5 interfaces, while specialised processing (e.g. the calculation of performance evaluation metrics) takes place on the server side and is principally coded in python. Key features of the platform include:

- A comprehensive range of ground truthing tools
- Centralised management of the annotation process
- Quality control and versioning
- Streamlined definition of evaluation scenarios
- Calculation of performance evaluation metrics
- In-depth results visualisation including intermediate evaluation steps

An earlier version of the annotation platform was made public in 2013, and is described in detail in [10]. In 2015, key updates were made to support the definition of non-axis oriented, quadrilateral boxes for words and text lines, as required for the Incidental Scene Text dataset. In the following section we briefly highlight some of the key aspects of the RRC Annotation and Evaluation Platform.

A. Dataset Management

Datasets of images can be created and managed through online interfaces, supporting the direct uploading of images, but also offering tools to harvest images online. As an example, a Google Street View crawler is integrated in the interface and can be used to automatically harvest images from Street View as seen in Figure 1.

Figure 2 shows a screenshot of the ground truthing management tool. The platform presents a searchable list to the ground truth manager that allows one to keep track of the overall progress, respond to specific comments that ground truthers

TABLE I
NUMBER OF SUBMISSIONS TO THE DIFFERENT DATASETS OF THE RRC.

	Public Submissions	Private Submissions	Years Active
Born Digital	63	1,403	2011 - 2017
Focused Scene Text	142	6,445	2003 - 2017 [a]
Text in Video	18	435	2013 - 2017
Incidental Scene Text	93	1,734	2015 - 2017
New 2017 Challenges[b]	108	44	2017

[a] Activity shown is for period from 2013 to August 2017.
[b] Preliminary figures, as the competition is still under way at the time of writing.

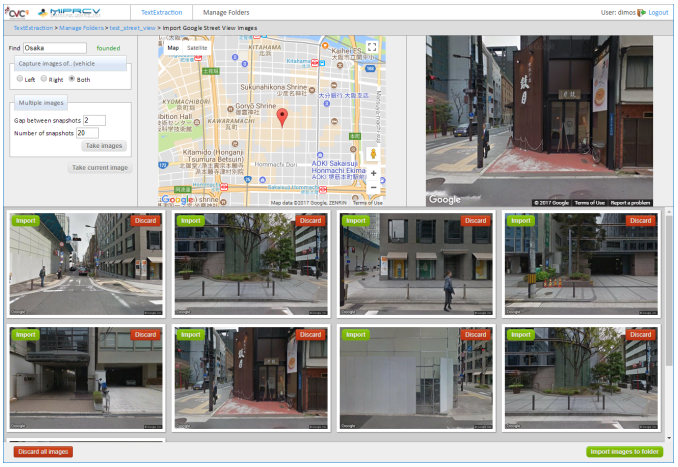

Fig. 1. The integrated Street View crawler.

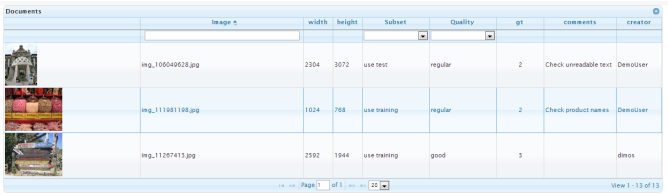

Fig. 2. A reduced screenshot of the ground truth management tool.

make and assign a quality rating to each image. The Quality Manager can make use of this information to provide feedback and ensure consistency of the ground truthing process. The same interface allows assigning images to the different subsets (training, validation, public and sequestered test) that are subsequently used for defining evaluation scenarios.

B. Image Annotation

Using the RRC Annotation and Evaluation platform it is possible to generate ground truth that represents everything from the pixel level to text lines in an image. behind the scenes, the RRC Annotation and Evaluation Platform stores such ground truth in a hierarchical tree using a combination of XML files for all metadata and transcription information and image files for pixel level annotations.

A screenshot of one of the Web-based annotation tools can be seen in Figure 5. The hierarchy of textual content and the

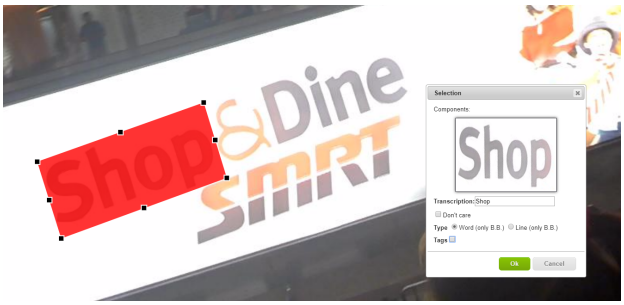

Fig. 3. Annotators are shown a real time preview of a rectified version of the word region being defined.

defined text parts is displayed on the left-hand side of the interface. In the example shown, text atoms are defined at the pixel level in terms of their area and skeleton, and grouped together to form words and text lines. Alternatively, annotations directly at the bounding box level (axis oriented or 4-point quadrilaterals), and at different granularities (characters, words, text blocks) are supported.

A number of tools are provided to ensure consistency and quality. For example, in the particular case of 4-point quadrilateral bounding boxes, when text with perspective distortion is annotated, it is often difficult for annotators to agree on what is a good annotation. To ensure consistency in the ground truth definition, a real time preview of a rectified view of the region is provided, and annotators are required to adjust the quadrilateral so that the *rectified* word appears correct (see Figure 3). This process improves substantially the consistency between different annotators.

All annotated elements, apart from their transcription, can have any number of custom defined associated metadata like script information, quality metrics etc. A special element type is text that should be excluded from the competitive process, and is thus marked as *do not care*. Depending on the challenge, such cases can include text which is partially cut, low resolution text, text in scripts other than the ones the challenge focuses on, or indeed any other text that the annotator deems as unreadable text.

Judging whether a text should be marked as *do not care* is challenging and in some cases similar text might be treated differently by individual annotators. At the same time, there are many cases where text can be read because the context

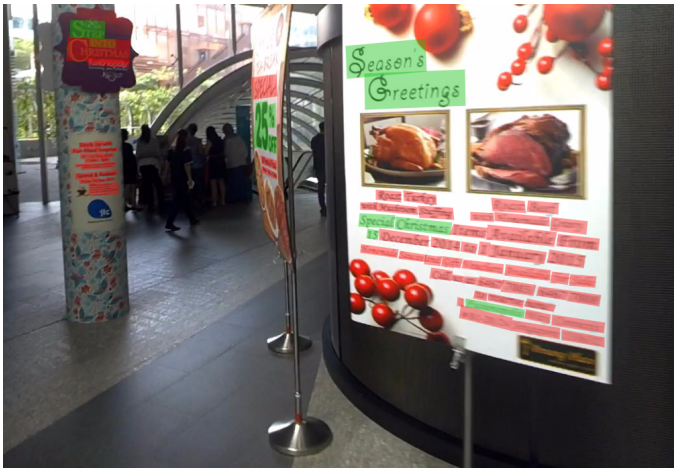

Fig. 4. *Do not care* regions appear in red, normal regions appear in green. *Do not care* regions do not have to respect the granularity of the rest of the ground truth. In the example, words have gone through a second-stage verifications where their readability was judged individually to eliminate any annotation bias introduced by contextual information (e.g. words that can be guessed to say "roast chicken" due to the visual context were judged as unreadable when seen individually).

is clear (e.g. if the words on the left and right are readable the middle word can be easily guessed), and annotators have trouble deciding whether such text should be actually marked as *do not care* or not. To reduce such subjective judgements different verification processes are available through the RRC platform, including interfaces for verifying words on their own, out the context, which has been shown to eliminate the inherent bias of annotators to use textual context to guess the transcription (see Figure 4).

The Web-based annotation functionality is available to use for research purposes by contacting with the RRC organisers.

C. Evaluation and Visualisation of Results

The online portal permits users to upload results of their methods against a public validation or test dataset and obtain evaluation results online. Apart from ranked tables of quantitative results of submitted methods, users can see per sample visualisations of their results along with insights about the intermediate evaluation steps, as seen in Figure 6. Through the same interface users can hot-swap between different methods to easily compare behaviours.

The python evaluation scripts used by the RRC platform are publicly available for each of the tasks. In addition, a standalone pack integrating the evaluation and visualisation interface is available to download. The pack can run offline on the user's machine and provides a Web-based graphical interface similar to the RRC portal's.

III. Conclusion

The RRC Annotation and Evaluation Platform is the backbone of the Robust Reading Competition's online portal. Many of the functionalities are exposed to the public (e.g. evaluation and visualisation of results), while others are accessible through contacting with the authors (e.g. annotation tools and dataset management). We strive to provide code when possible, although this is not always feasible due to the tight integration of certain functionalities with the Web portal. Nevertheless, a full version of the RRC Web portal was made public in the past [10], while more recently standalone interfaces for evaluation and visualisation were also made available to download.

Acknowledgements

This work is supported by Spanish project TIN2014-52072-P and the CERCA Programme / Generalitat de Catalunya.

References

[1] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "ICDAR 2003 robust reading competitions," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2003, pp. 682–687.

[2] D. Karatzas, S. R. Mestre, J. Mas, F. Nourbakhsh, and P. P. Roy, "ICDAR 2011 robust reading competition-challenge 1: reading text in born-digital images (web and email)," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2011, pp. 1485–1490.

[3] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazan, and L. P. de las Heras, "ICDAR 2013 robust reading competition," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2013, pp. 1484–1493.

[4] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu *et al.*, "ICDAR 2015 competition on robust reading," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 1156–1160.

[5] R. Gomez, B. Shi, L. Gomez, L. Neumann, A. Veit, J. Matas, S. Belongie, and D. Karatzas, "ICDAR2017 robust reading challenge on COCO-Text," in *Document Analysis and Recognition (ICDAR), 2017 14th International Conference on*. IEEE, 2017.

[6] C. Yang, X.-C. Yin, H. Yuz, D. Karatzas, and Y. Cao, "ICDAR2017 robust reading challenge on text extraction from biomedical literature figures (DeTEXT)," in *Document Analysis and Recognition (ICDAR), 2017 14th International Conference on*. IEEE, 2017.

[7] M. Iwamura, N. Morimoto, K. Tainaka, D. Bazazian, L. Gomez, and D. Karatzas, "ICDAR2017 robust reading challenge on omnidirectional video," in *Document Analysis and Recognition (ICDAR), 2017 14th International Conference on*. IEEE, 2017.

[8] R. Smith, C. Gu, D.-S. Lee, H. Hu, R. Unnikrishnan, J. Ibarz, S. Arnoud, and S. Lin, "End-to-end interpretation of the french street name signs dataset," in *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 411–426.

[9] N. Nayef, F. Yin, I. Bizid, H. Choi, Y. Feng, D. Karatzas, Z. Luo, U. Pal, C. Rigaud, J. Chazalon, W. Khlif, M. L. Muzzamil, J.-C. Burie, C.-l. Liu, and J.-M. Ogier, "ICDAR2017 robust reading challenge on multi-lingual scene text detection and script identification – RRC-MLT," in *Document Analysis and Recognition (ICDAR), 2017 14th International Conference on*. IEEE, 2017.

[10] D. Karatzas, S. Robles, and L. Gomez, "An on-line platform for ground truthing and performance evaluation of text extraction systems," in *International Workshop on Document Analysis Systems (DAS)*, 2014, pp. 242–246.

Fig. 5. A screenshot of one of the Web-based annotation tools.

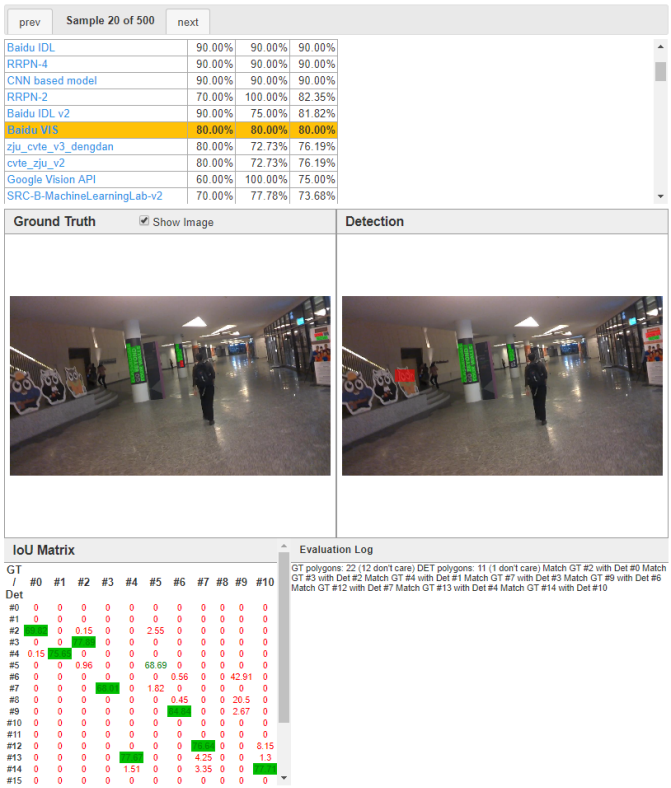

Fig. 6. Per-image results interface for text localisation.