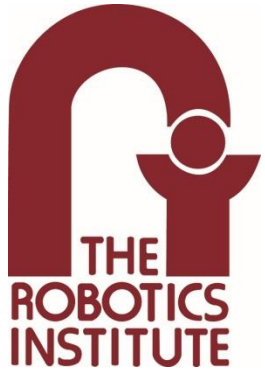
ICDAR2019 Robust Reading Challenge on Multi-lingual Scene Text Detection and Recognition – RRC-MLT-2019

Nibal Nayef, Yash Patel, Michal Busta, Pinaki Nath Chowdhury, Dimosthenis Karatzas, Wafa Khlif, Jiri Matas, Umapada Pal, Jean-Christophe Burie, Cheng-lin Liu *and* Jean-Marc Ogier



Outline

- Introduction: Multi-lingual Text Detection & Recognition in Scene Images
- The RRC-MLT Dataset
- The MLT Challenge Tasks
- RRC Framework
  - MLT Challenge Organization and Participants
  - Evaluation Protocol
- Results & Discussion (poster session)
- Conclusions

Introduction

- Text detection and recognition in a natural environment is essential to many applications
  - Tourist guidance, helping the visually impaired, data mining and autonomous driving, ….

Introduction: RRC-MLT Objectives

- Build a large benchmarking dataset for scene text detection & recognition
  - Multi-lingual text, multi-oriented text, content variety, complex layout etc.

The RRC-MLT-2019 Dataset

- **Real set: 20,000 scene images containing:**
  - Text of 10 languages, 2,000 images per language
    - Arabic, Bangla, Chinese, Devanagari, English, French, German, Italian, Japanese, Korean
    - An image usually contains text of more than one language
  - **7 different scripts:** Arabic, Bangla, Chinese, Hindi, Japanese, Korean and Latin, **+2 defined scripts: "**Symbols" and "Mixed"
  - **Dataset Division:** 50% for training and 50% for testing

The RRC-MLT-2019 Dataset

- **Synthetic set: 277,000 images**
  - Same set of 10 scripts as in the real set, rendered over natural scene images selected from the 8,000 background images collected by [Gupta et al. 2016]

(a) Arabic Scene Text (b) Bangla Scene Text (c) Chinese Scene Text

(d) Japanese Scene Text (e) Korean Scene Text (f) Latin Scene Text

MLT Challenge Tasks

- Detecting multi-lingual text at word level
  - Except in Chinese and Japanese: text is labeled at line level
- Script classification of cropped word images
  - Valid scripts for this task are: “Arabic”, “Bangla”, “Chinese”, “Hindi”, “Japanese”, “Korean”, “Latin” and “Symbols”
- Joint text detection and script identification
- End-to-End text detection and recognition
  - The synthetic dataset is provided to help with training

MLT Challenge Tasks – Ground Truth

- Tasks 1, 3 and 4
  - 10,000 training images, 10,000 test images
  - Each image has a corresponding GT file
    - A list of the coordinates of the bounding boxes of all the words inside an image (including “don’t care” words), the script id, and the transcription for each text box
- Task 2
  - 89,177 training word images and 102,462 test word images
  - Single script name (*ID*) per image

RRC-Framework: Evaluation Metrics

- The following metrics have been used for ranking participants methods:
  - **Detection:** f-measure (based on the overlap between detected word bounding box and the GT box)
  - **Cropped word script identification:** accuracy of the detected script *IDs* of all the word images versus ground-truth script *IDs*
  - **Joint detection & script id:** a cascade of correct detection of a text box ***and*** correct script classification
  - **End-to-End recognition:** cascade of correct localization of a text box ***and*** its correct transcription

RRC-Framework: Challenge Organization

- We have used the web portal of the RRC platform
  - Interacting with participants, downloads and online submissions

- Overall, we had **60** different submissions:
  - 25 in Task-1, 15 in Task-2, 10 in Task-3 and 10 in Task-4

Results – Winners: in other sessions

- More details & Winners certificates:

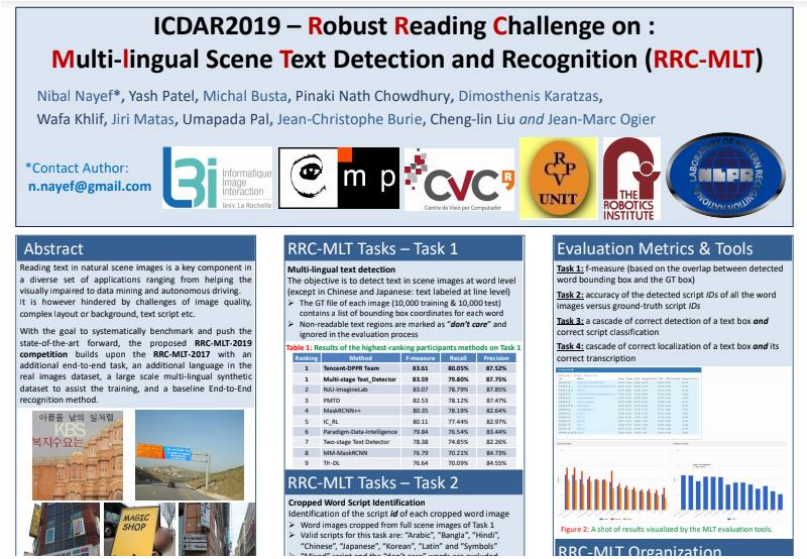
  Oral competition session

  Monday 23rd Sep. 16:20 – 17:40

- Results & discussion

  Poster session

  Tuesday 24th Sep. 15:40 – 17:40



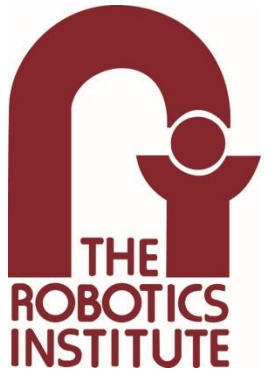
Conclusions

- Novel aspects of our work
  - Size of the dataset (20000 scene images)
  - Multi-lingual text
    - 10 languages, 7 Scripts plus *Symbols* and *Mixed* scripts
  - Multi-oriented text, variety of scenes content and image resolution
  - A new synthetic dataset that matches the real set
  - A baseline method for the new End-to-End multi-lingual recognition task
- Results show that the dataset is very challenging

Thank you

Contact

n.nayef@gmail.com



Baidu VIS on ICDAR 2019 Robust Reading Challenge on MLT Task I

Pengfei Wang~, Mengyi En*, Xiaoqiang Zhang*, Chengquan Zhang*

Affiliation: VIS-VAR Team, Baidu Inc.*; Xidian University~

Speaker: Xiameng Qin*

ICDAR 2019

Results on MLT19 Test Set

TABLE I. RESULTS OF THE RRC-MLT-2019 CHALLENGE FOR TASK-1: MULTI-LINGUAL TEXT DETECTION

Rank	Method	Hmean	Precision	Recall
1	Tencent-DPPR Team	83.61%	87.52%	80.05%
1	Multi-stage_Text_Detector	83.59%	87.75%	79.80%
2	NJU-ImagineLab	83.07%	87.85%	78.79%
3	PMTD [21]	82.53%	87.47%	78.12%
4	MaskRCNN++	80.35%	82.64%	78.19%
5	IC_RL	80.11%	82.97%	77.44%
6	4Paradigm-Data-Intelligence	79.84%	83.44%	76.54%
7	Two-stage Text Detector —based on Cascade-RCNN	78.38%	82.26%	74.85%
8	MM-MaskRCNN	76.79%	84.73%	70.21%
9	TH-DL	76.64%	84.55%	70.09%
10	SOT	74.24%	79.96%	69.28%

Nayef, Nibal, et al. "ICDAR2019 Robust Reading Challenge on Multi-lingual Scene Text Detection and Recognition--RRC-MLT-2019." arXiv preprint arXiv:1907.00945 (2019).

Overview

The main characteristics of the task I of RRC-MLT-2019 challenge are:

- Multi-oriented

  Unfocused text with various orientations.

- Multilingual

  10 languages with different principles of sentence.

- Multi-scale

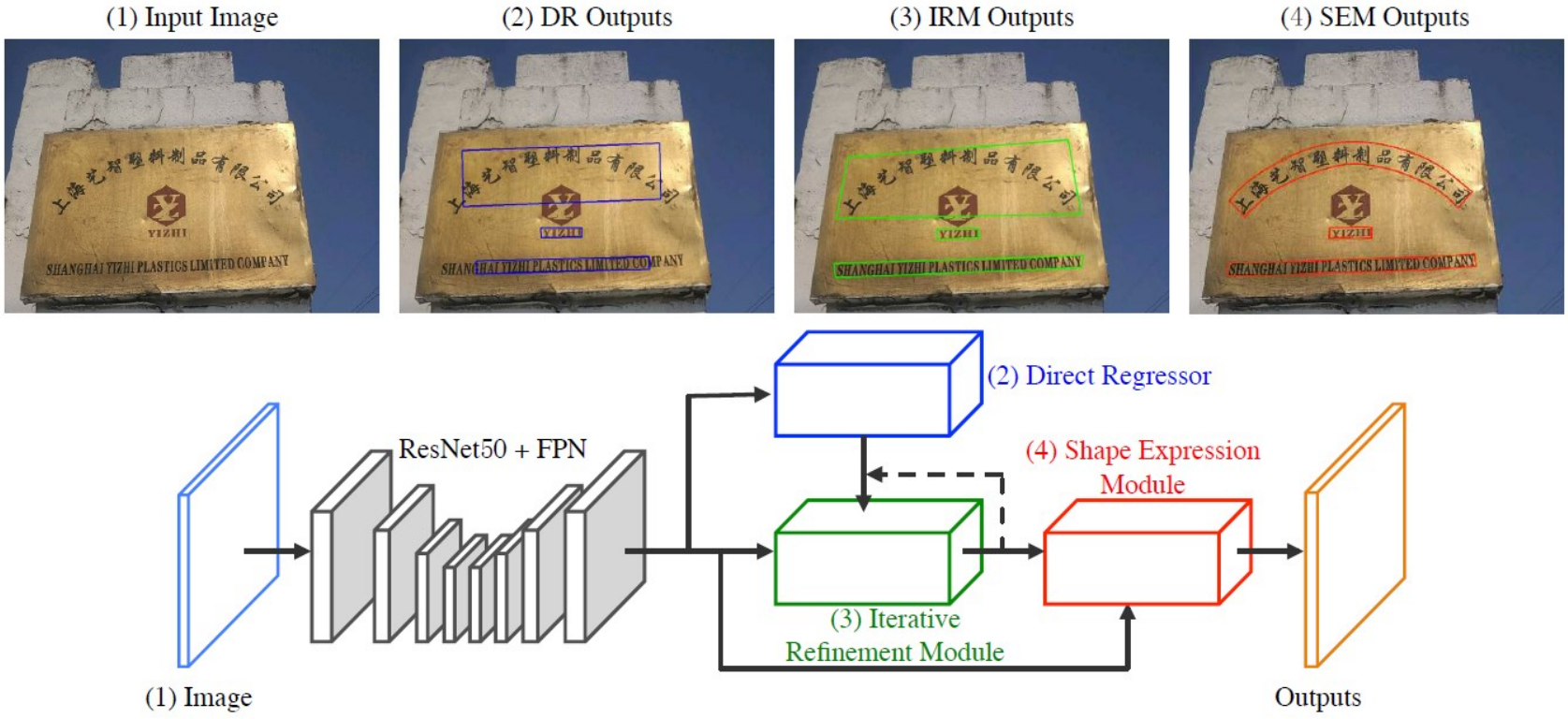
  Extremely large or small text appears at the same time.

Top Down Method + Bottom Up Method + Ensemble

Extra partial KAIST (Korean data) / No private data used.

Top Down Method (LOMO)

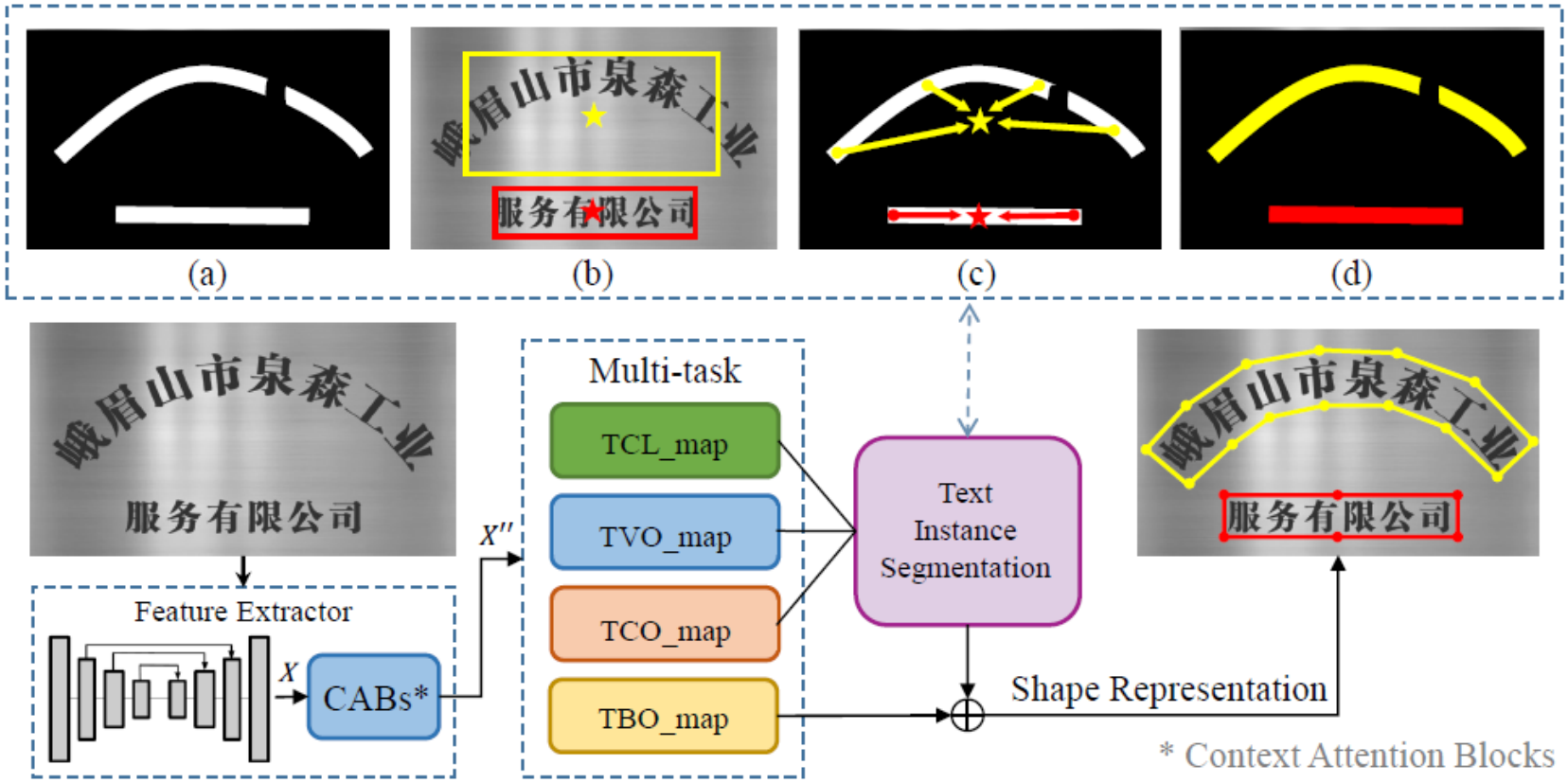
Look More Than Once: An Accurate Detector for Text of Arbitrary Shapes



Chengquan Zhang, Borong Liang, Zuming Huang, Mengyi En, Junyu Han, Errui Ding, Xinghao Ding; Look More Than Once: An Accurate Detector for Text of Arbitrary Shapes. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 10552-10561

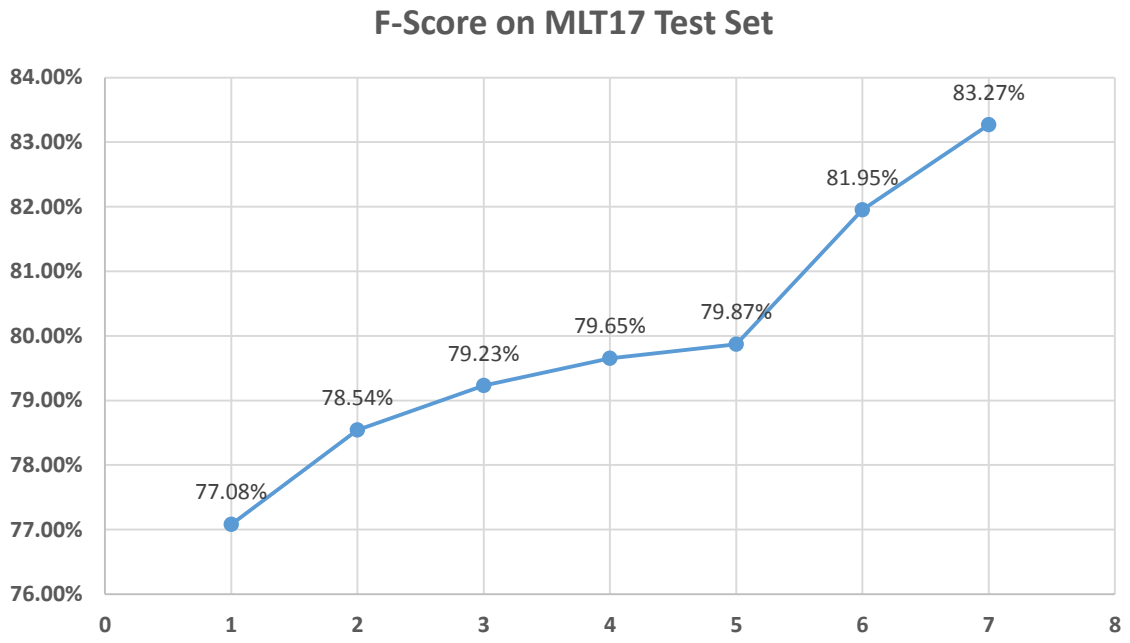
Bottom Up Method (SAST)

A Single-Shot Arbitrarily-Shaped Text Detector based on Context Attended Multi-Task Learning



Pengfei Wang, Chengquan Zhang, Fei Qi, Zuming Huang, Mengyi En, JunyuHan, Jingtuo Liu, Errui Ding, and Guangming Shi. 2019. A Single-Shot Arbitrarily-Shaped Text Detector based on Context Attended Multi-Task Learning. In Proceedings of the 27th ACM International Conference on Multi-media (MM'19).

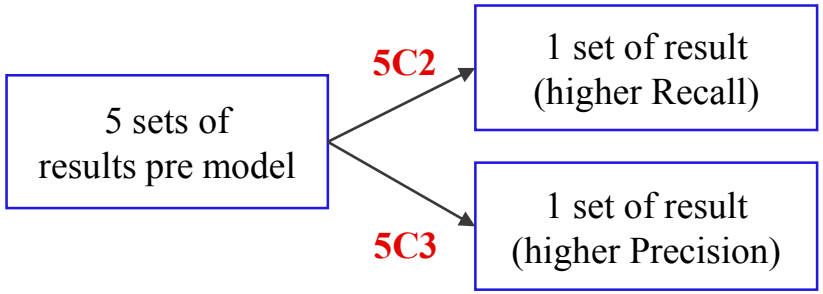
Results on MLT17 Test Set



1. LOMO_baseline_mlt17 (baseline, 77.08%,)
2. Random reshape and crop (+1.46%, 78.54%)
3. Pretrain on synth data and train on MLT19 train data (+0.69%, 79.23%)
4. OHEM (+0.42%, 79.64%)
5. Extra data(IC15, partial KAIST)(+0.22%, 79.87%)
6. Multi-scale testing (1024, 1280, 1536, 2048, 2560) (+2.08%, 81.95%)
7. Multi-mode ensemble (Six models) (+1.32%, 82.27%)

Ensemble Strategy

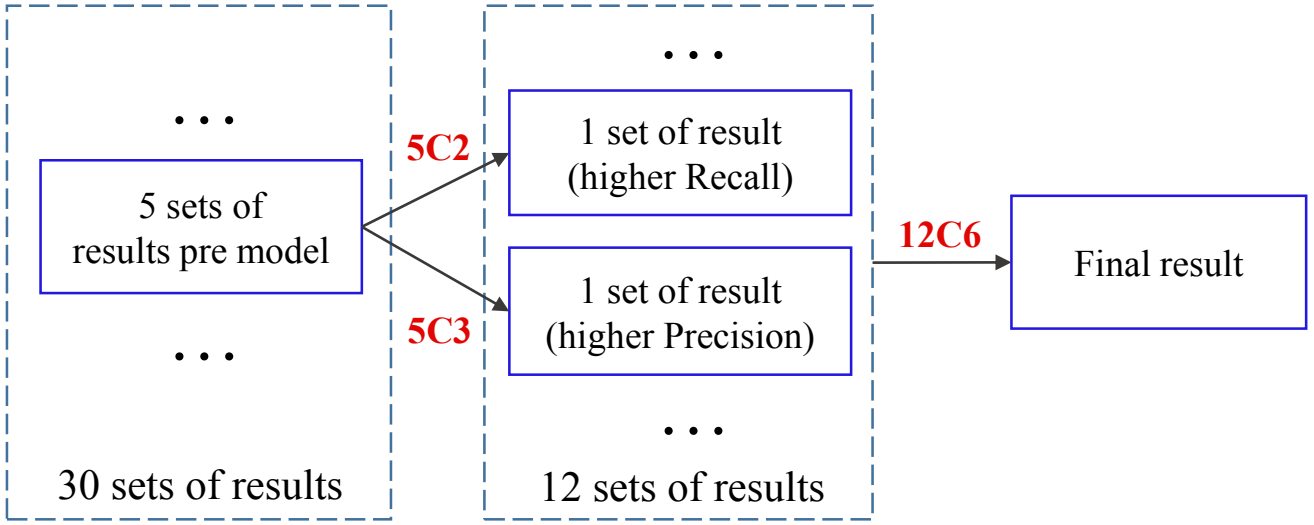
xCy voting strategy: Keep the detection results that appear at least y times in the x sets of detection results. For example, **5C2** strategy means we only keep the quadrilaterals detected at least in two different sets of results.



Five sets of results pre model:
1024, 1280, 1536, 2048, 2560

Ensemble Strategy

xCy voting strategy: Keep the detection results that appear at least y times in the x sets of detection results. For example, **5C2** strategy means we only keep the quadrilaterals detected at least in two different sets of results.



Five sets of results pre model:

1024, 1280, 1536, 2048, 2560

Six models:

LOMO_Resnet_50, LOMO_Inception_v4, LOMO_w_OHEM

SAST_Resnet_50, SAST_Inception_v4, SAST_w_OHEM

To Be Explored

There are several unsolved problems we encountered in the competition, which may be open questions in the field of scene text detection and could be explored in the future:

1. The detection of scene text with **variety of sizes**, including some both extremely large and small text.

2. The detection of scene text with **mixed horizontal and vertical layout**, which is be more common in Chinese.

3. The detectors of scene text are often optimized for specific scene, and some techniques, such as domain adaptation, may be used for **more general text detector**.

4. And so on…

Baidu AI Cloud Platform

Our OCR service is available on Baidu AI Cloud Platform.

https://ai.baidu.com/tech/ocr/

Thank you!

Looking for Intern, Research Developer.

hanjunyu@baidu.com

ICDAR 2019

ICDAR 2019 Robust Reading Challenge on MLT

Tencent-DPPR Team

Tencent-DPPR Team *(Chunchao Guo, Hongfa Wang, et al.)*

Data Platform Department Precision Recommendation Team, Tencent

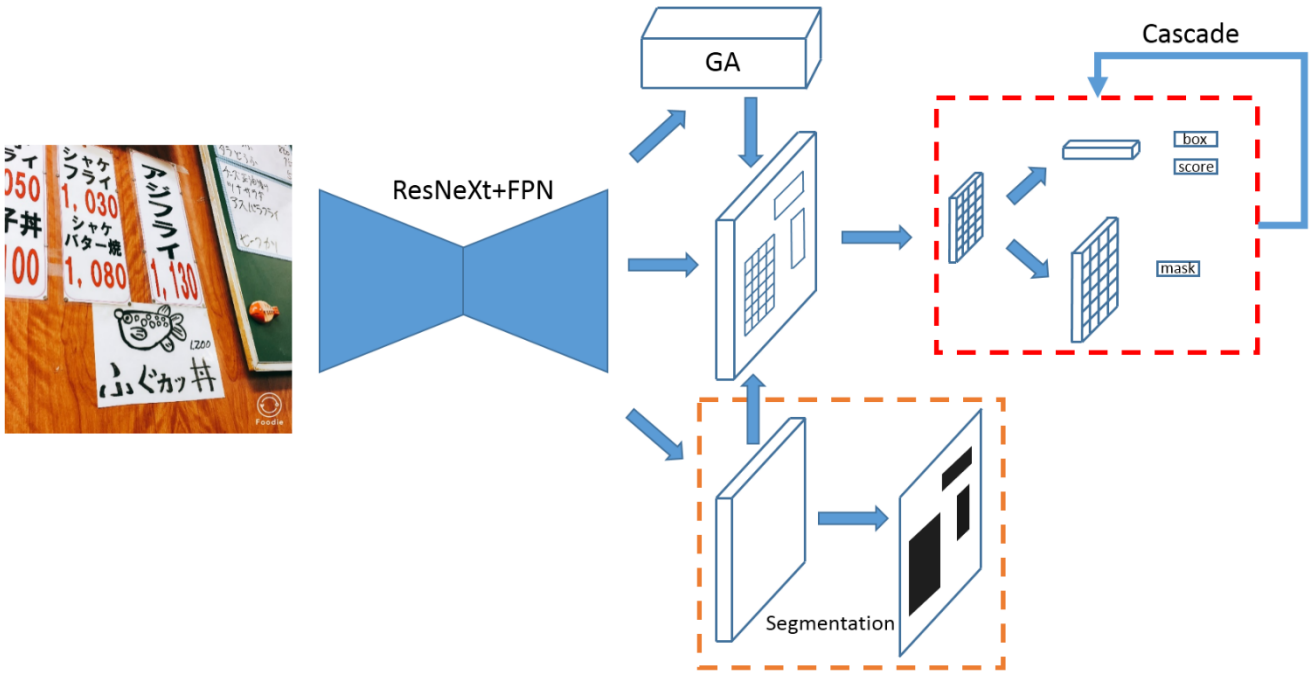
Tencent 腾讯

Overview

- **Participate Four Tasks of MLT-19**
  - T1: Multi-script text detection
  - T2: Cropped Word Script identification
  - T3: Joint text detection and script identification
  - T4: End-to-End text detection and recognition

Task 1: Text Detection

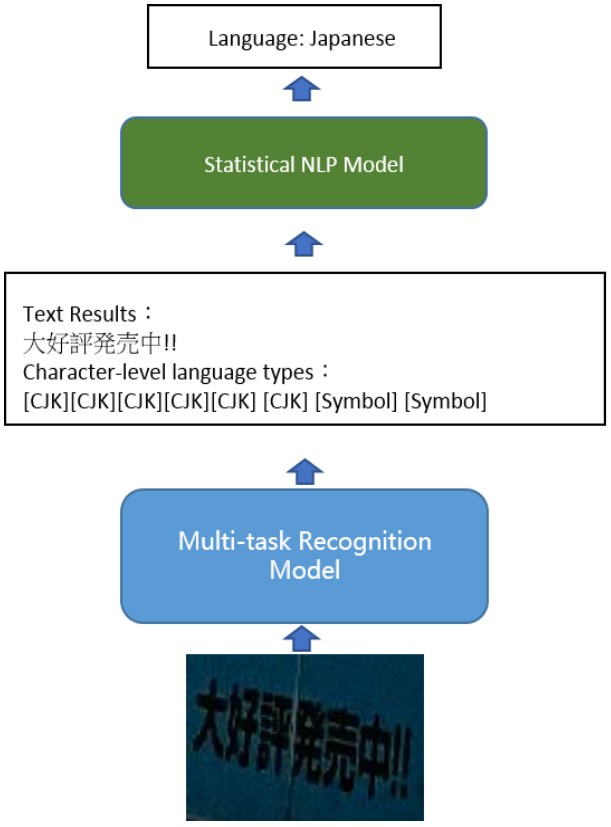
- **Approach**
  - In a Mask R-CNN style
  - Incorporate Mask R-CNN and instance segmentation
  - Introduce Guided Anchor (GA)



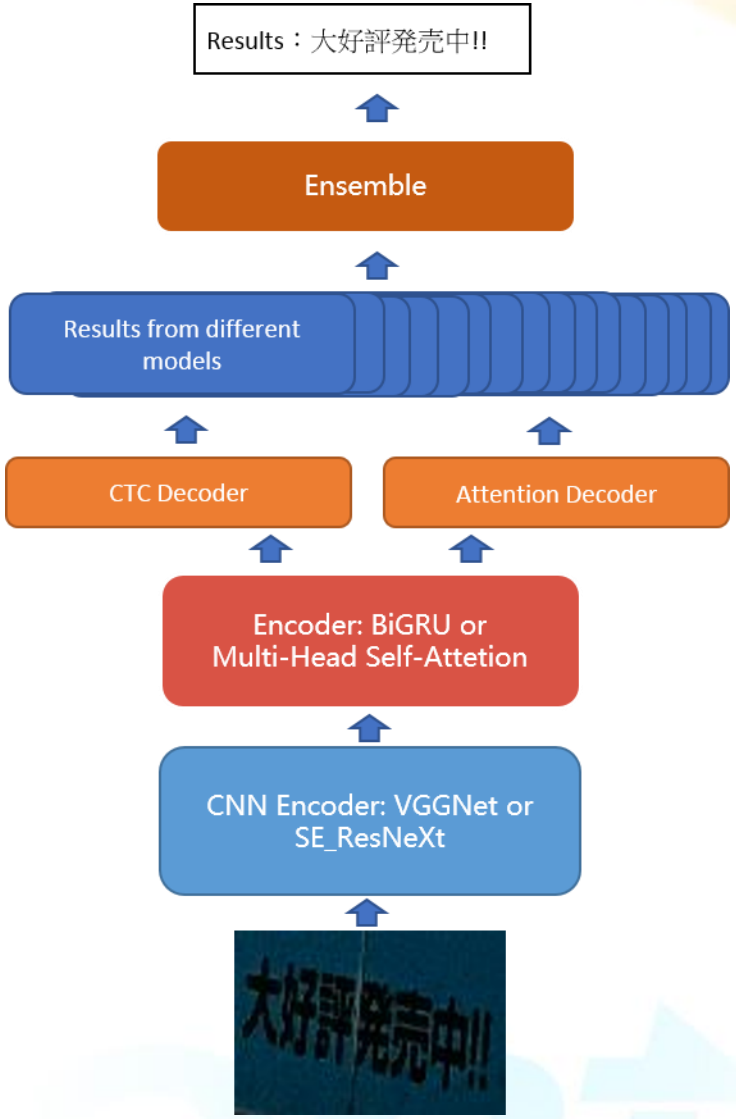
Task 2: Cropped Word Script Identification

Approach

- First recognition
- Then language identification



Script Identification



Line Recognition

Tencent 腾讯

Task 3: Joint Text Detection and Script Identification

- **Approach**
  - Combination of Task 1 and Task 2
  - Text detection in an image
  - Cropped word script identification
  - Using ensemble models

Task 4: End-to-End Text Detection and Recognition

- **Approach**
  - Combination of detection and recogntion
  - First text detection
  - Then cropped word recognition
  - Using ensemble models

Results

- **T1. Text Detection:** Rank 1st
- **T2. Script identification:** Rank 1st
- **T3. E2E Script identification:** Rank 1st
- **T4. E2E Recognition:** Rank 1st

Thank You!