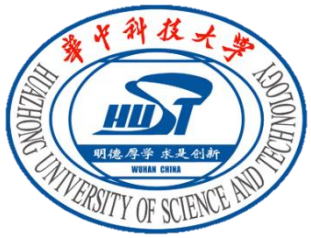


華中科技大學

Huazhong University of Science & Technology

Photo OCR (2017-2018)

Xiang Bai

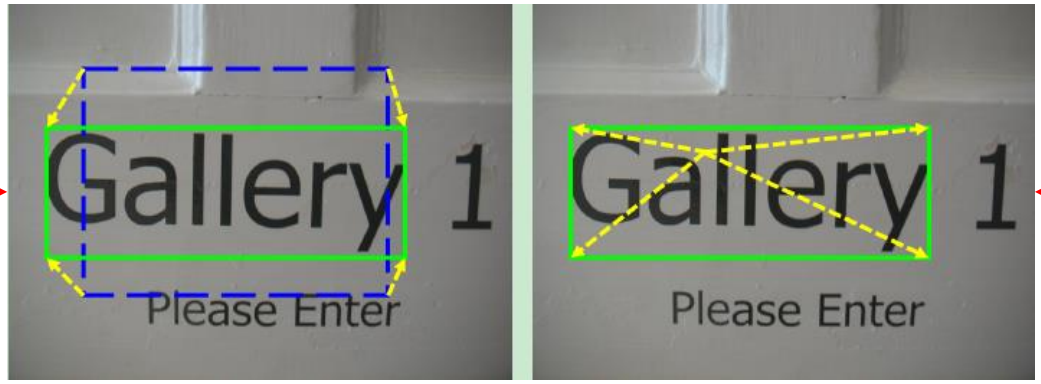
Huazhong University of Science and Technology

Outline

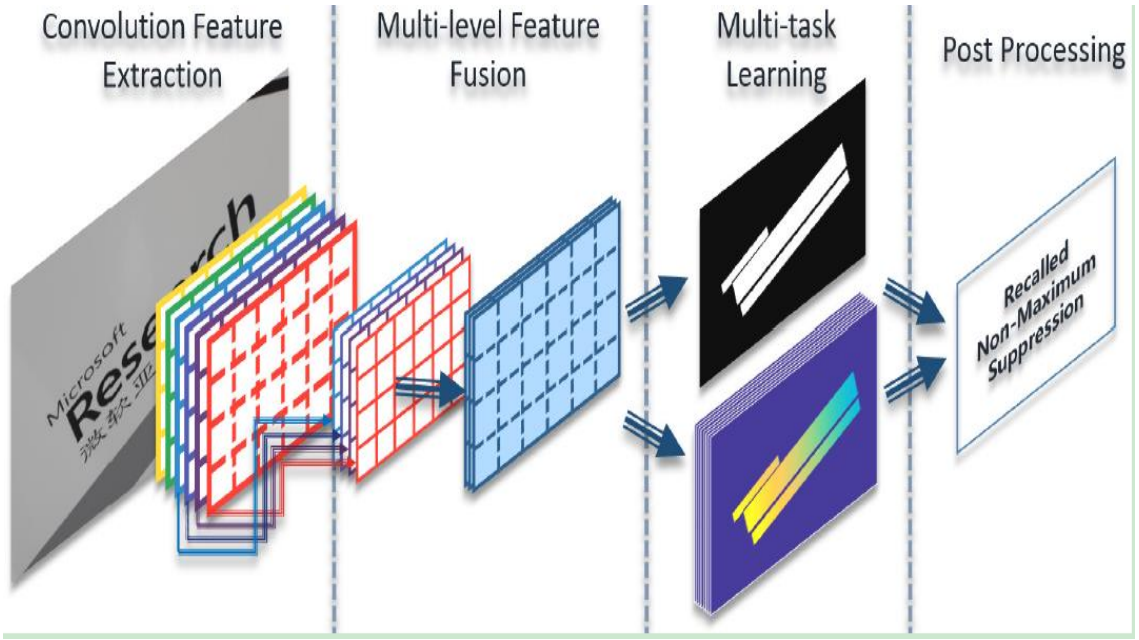
- Detection
- Recognition
- End-to-end recognition
- New datasets

Deep Direct Regression for Multi-Oriented Scene Text Detection [He et al., ICCV, 2017.]

Indirect regression



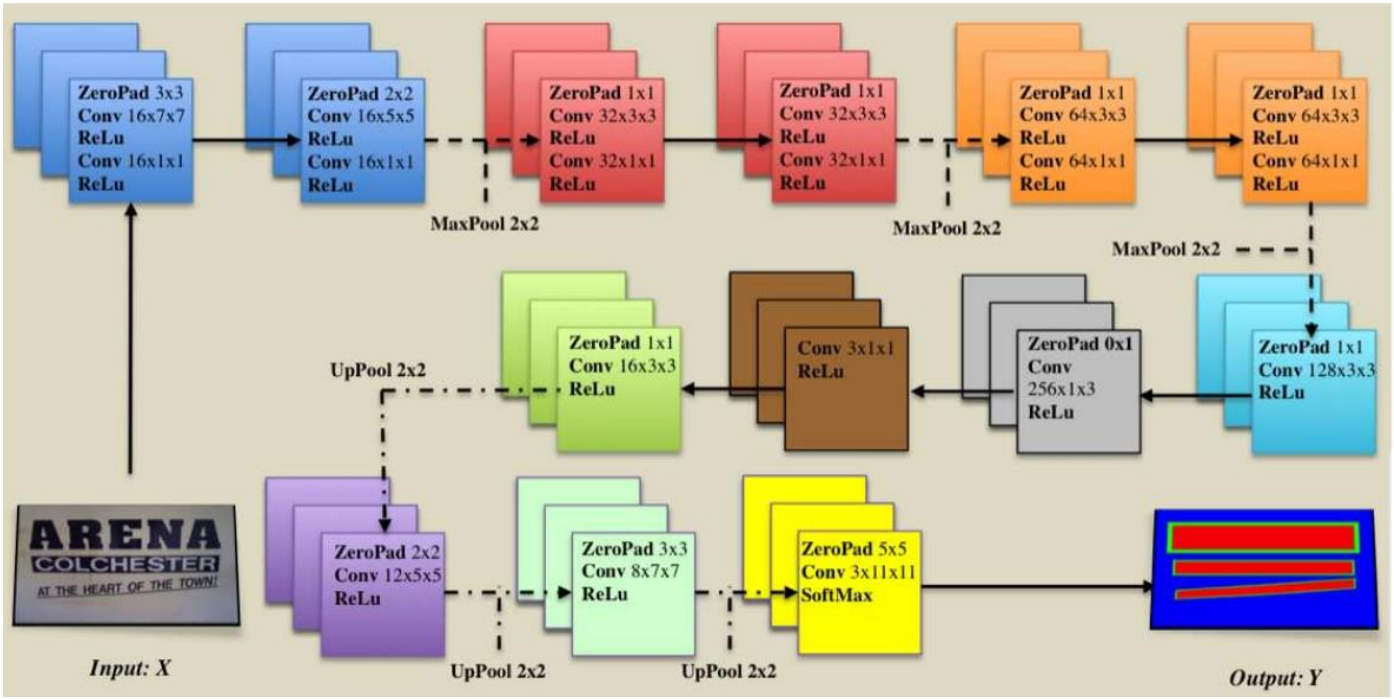
Direct regression

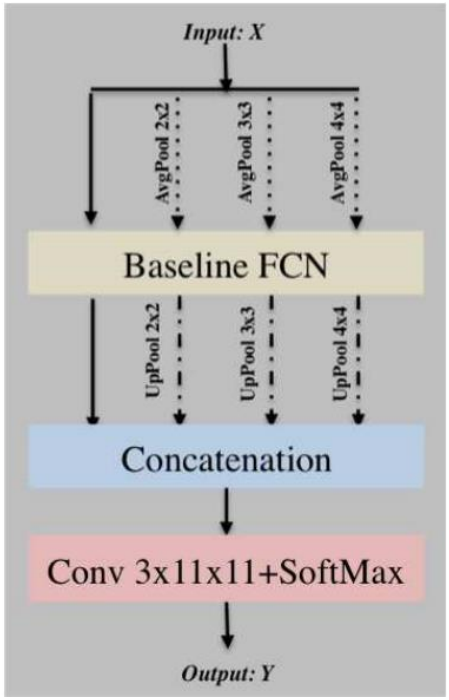


Architecture

- Multi-level feature fusion
- Up-sample to quarter size of the input image
- Multi-task learning for classification and regression
- Post Processing:
  - Refined NMS

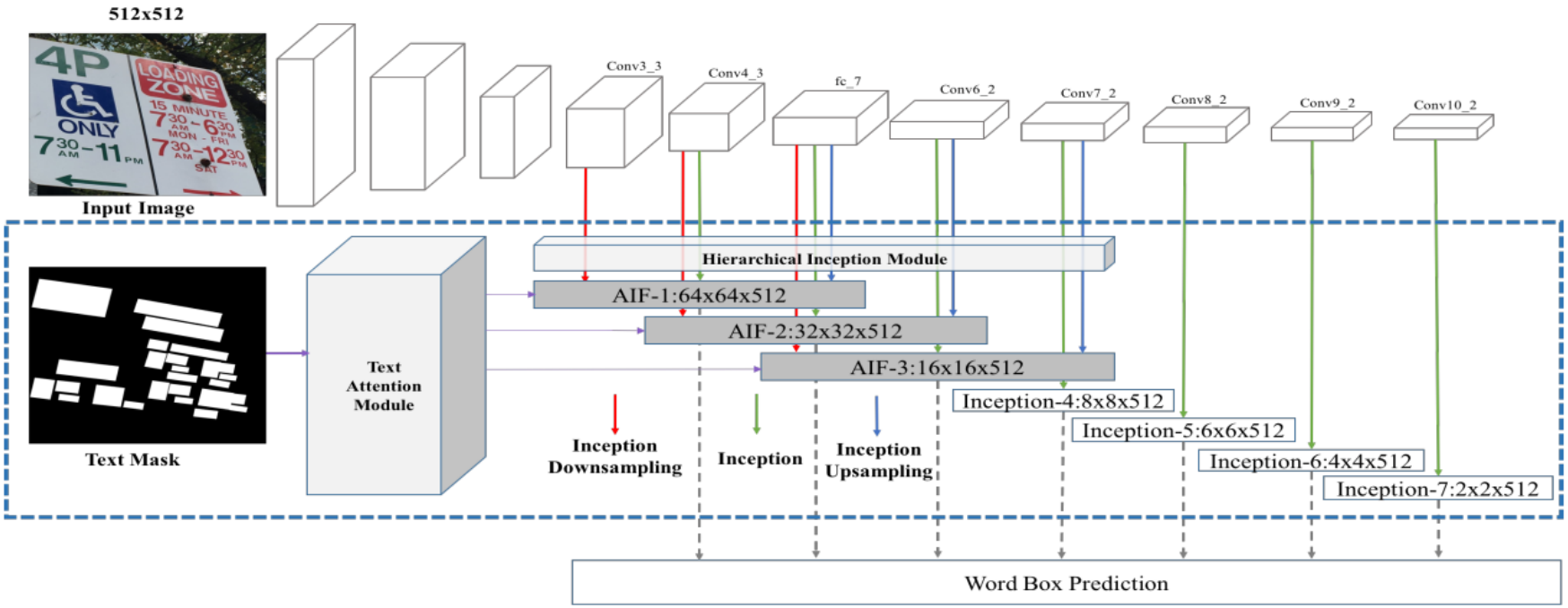
Self-Organized Text Detection With Minimal Post-Processing via Border Learning [Wu et al., ICCV, 2017.]





- Multi-scale FCN
- 3 classes: text, no text, and border

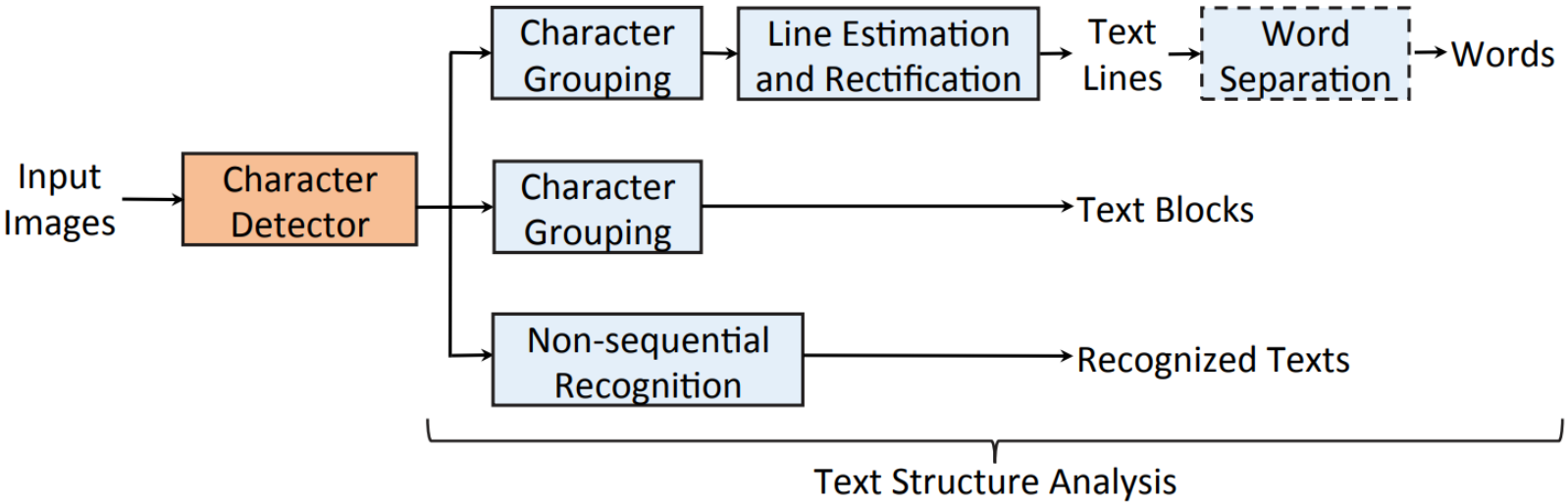
Single Shot Text Detector with Regional Attention [He et al., ICCV, 2017.]



- SSD [1] backbone
- Using text mask as attention information

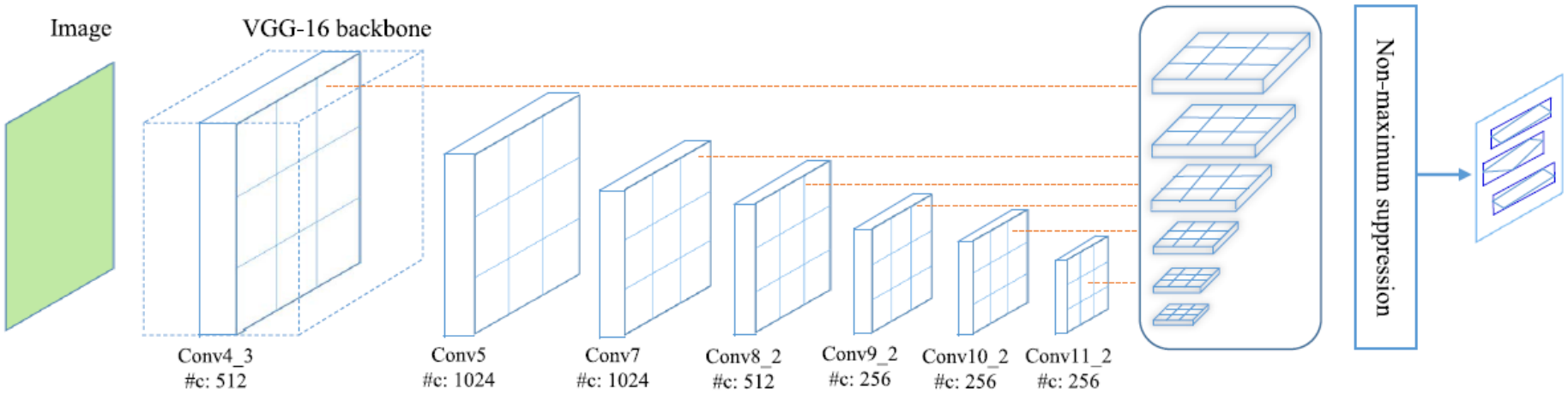
[1] W. Liu, et al. SSD: single shot multibox detector. ECCV, 2016

WordSup: Exploiting Word Annotations for Character based Text Detection [Hu et al., ICCV, 2017.]



- A weakly supervised framework that can utilize word annotations for character detector training (inspired by [1])
- Text structure analysis to group the characters

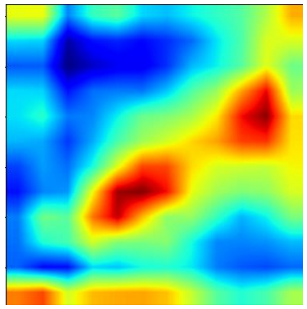
[1] J. Dai, et al. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. ICCV, 2015



- TextBoxes++ is an extension of TextBoxes [1]
- SSD backbone
- From horizontal text detector (TextBoxes) to oriented text detector
  - Regress four vertexes of a quadrilateral
- Better combination with recognition

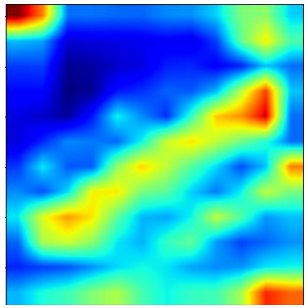
[1] M. Liao et al. TextBoxes: A Fast Text Detector with a Single Deep Neural Network. AAAI, 2017

Rotation-sensitive Regression for Oriented Scene Text Detection [Liao et al., CVPR, 2018.]



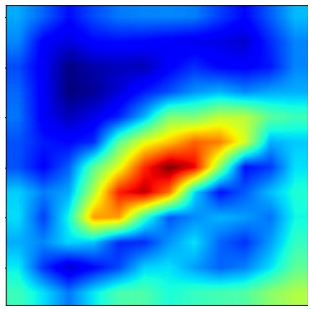
Shared feature for cls. & reg.

Input image



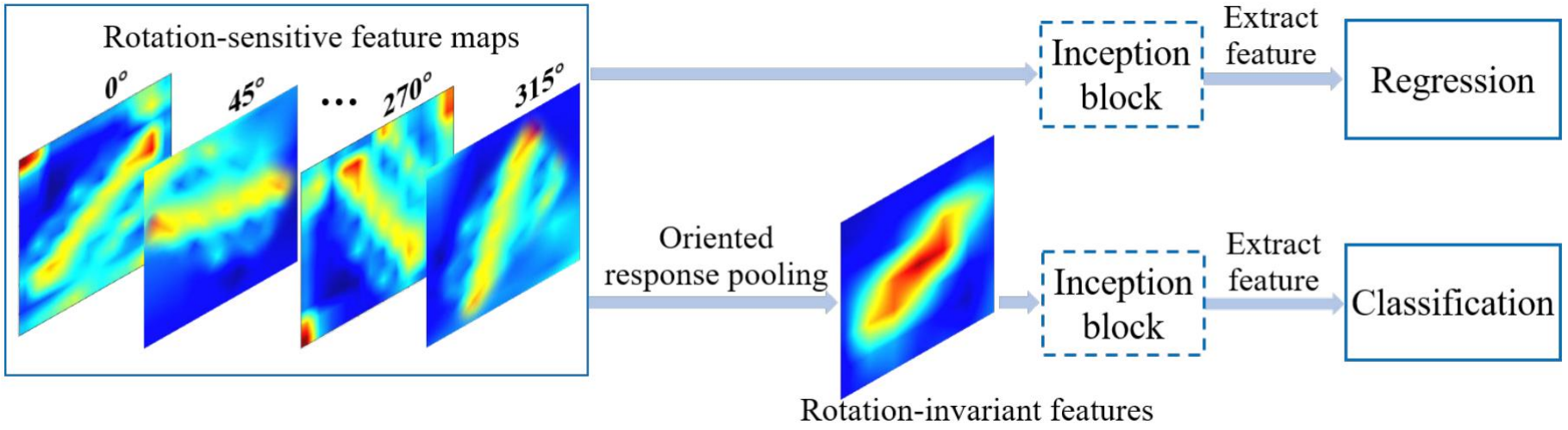
Rotation-sensitive feature for reg.

Predicted bounding box and score



Rotation-invariant feature for cls.

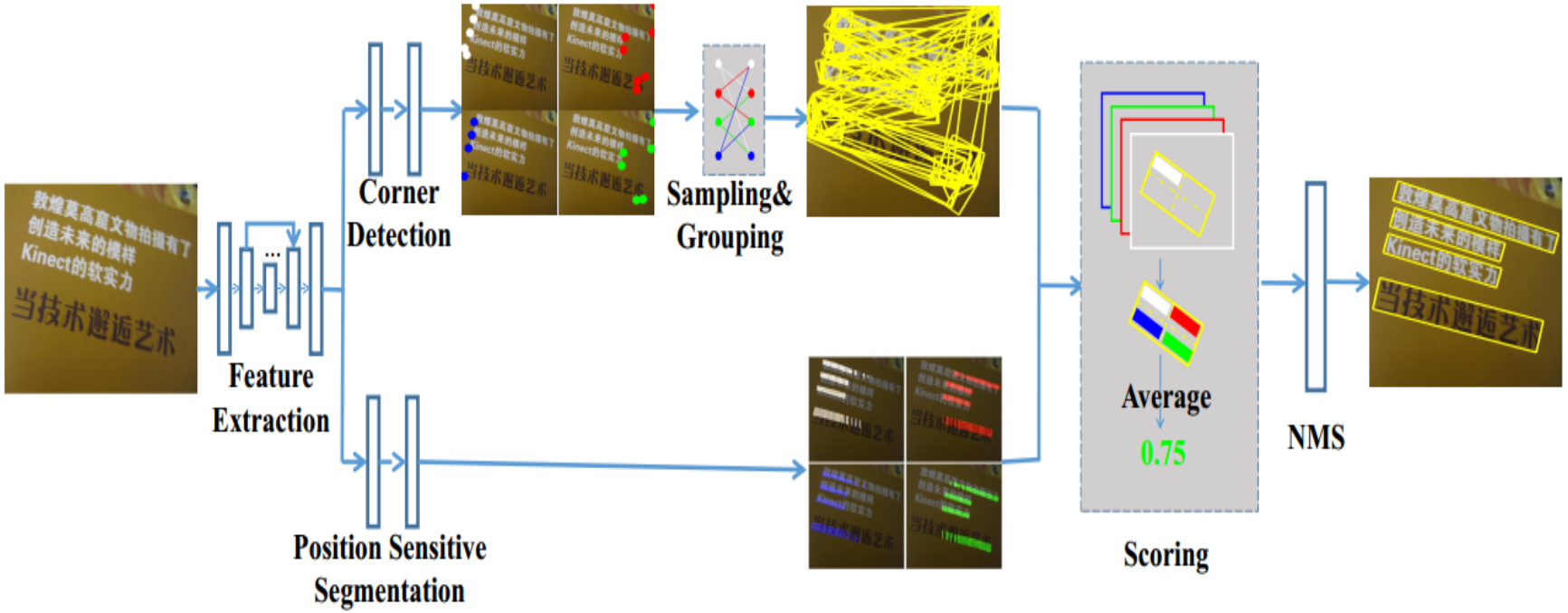
Rotation-sensitive Regression for Oriented Scene Text Detection [Liao et al., CVPR, 2018.]



- Rotation-sensitive feature maps for regression
  - using oriented response convolution [1]
- Rotation-invariant features for classification
  - pooling all rotation-sensitive feature maps

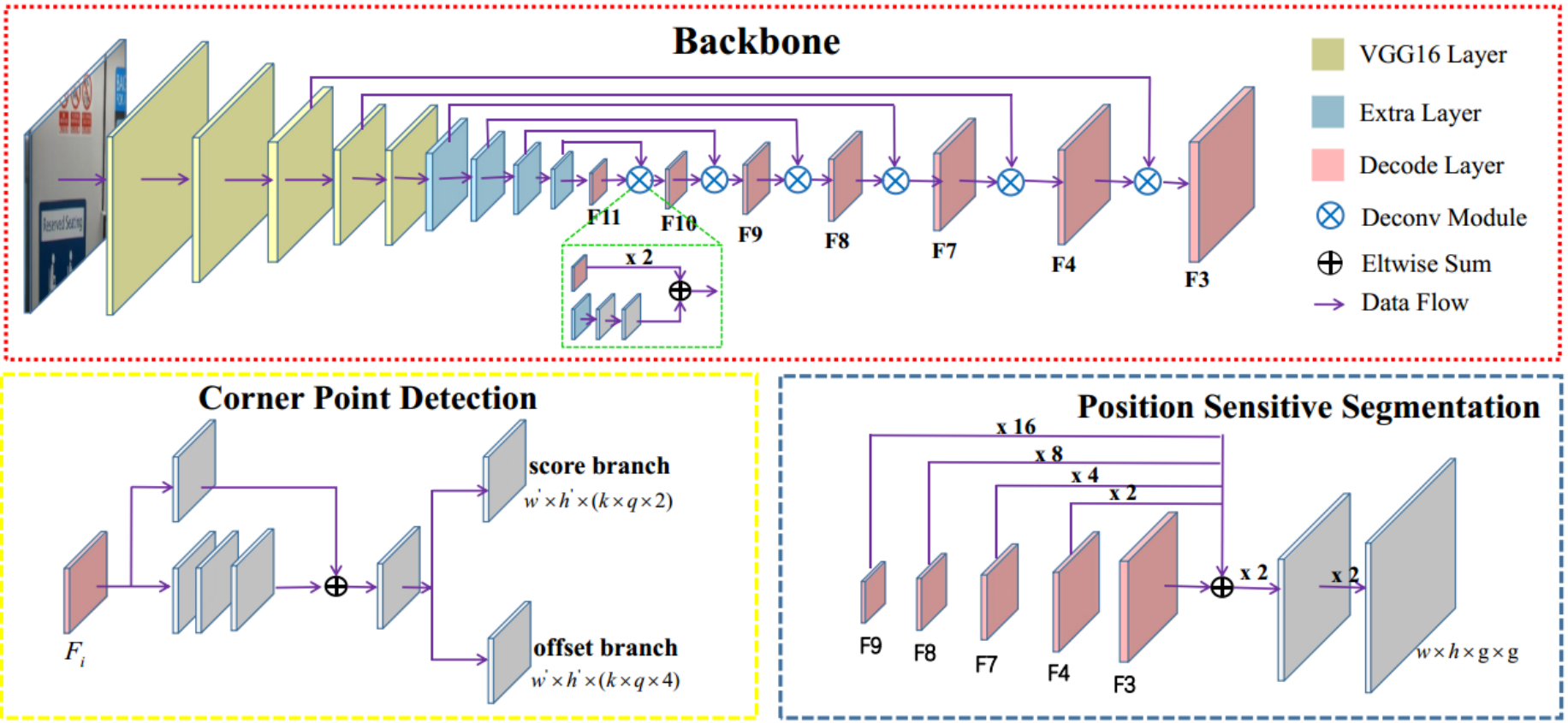
[1] Y, Zhou, et al. Oriented response networks. CVPR, 2017

Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation [Lyu et al., CVPR 2018]



- A compound text detection method: corner localization and region segmentation.
- Proposal generation via sampling and grouping corner points.
- Scoring proposals with position sensitive segmentation maps.

Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation [Lyu et al., CVPR 2018]



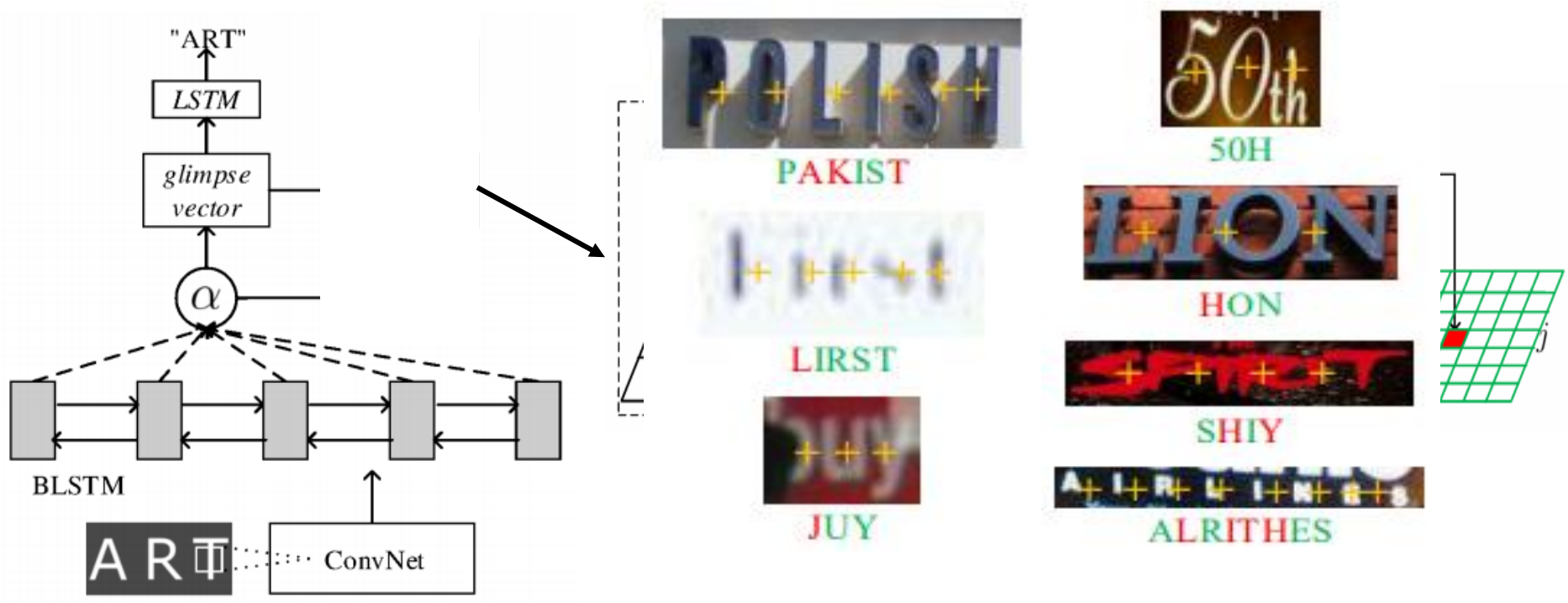
- Corner localization: corner detection with DSSD[1].
- Region segmentation: position-sensitive segmentation

[1] C. Fu, et al. DSSD: Deconvolutional Single Shot Detector. Arxiv, 2017

Outline

- Detection
- **Recognition**
- End-to-end recognition
- New datasets

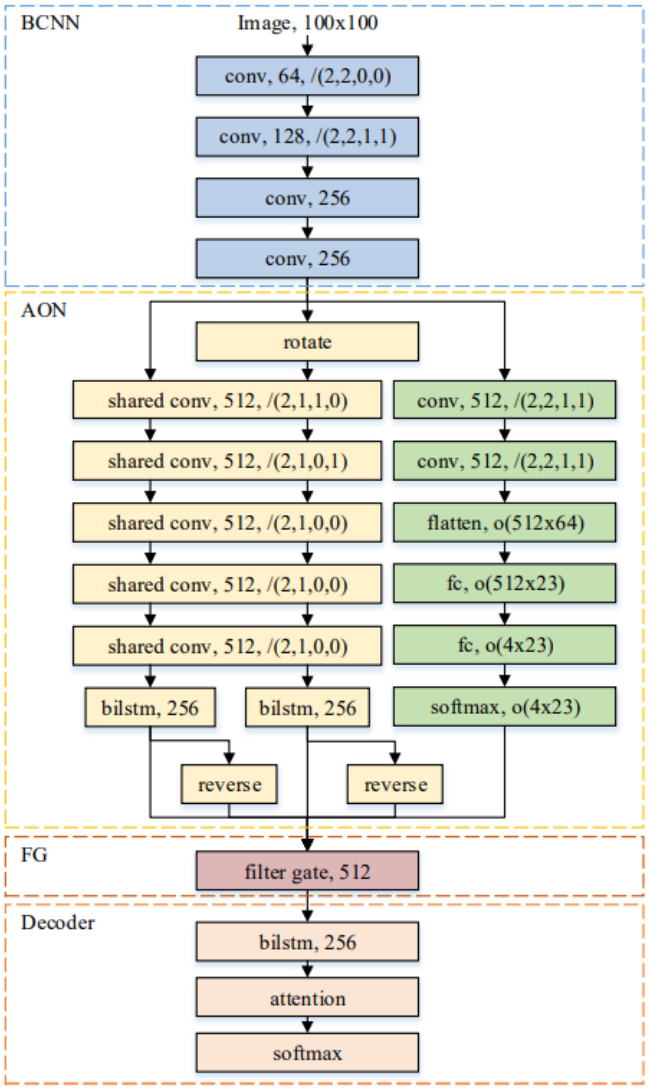
Focusing Attention: Towards Accurate Text Recognition in Natural Images [Cheng et al., ICCV, 2017.]

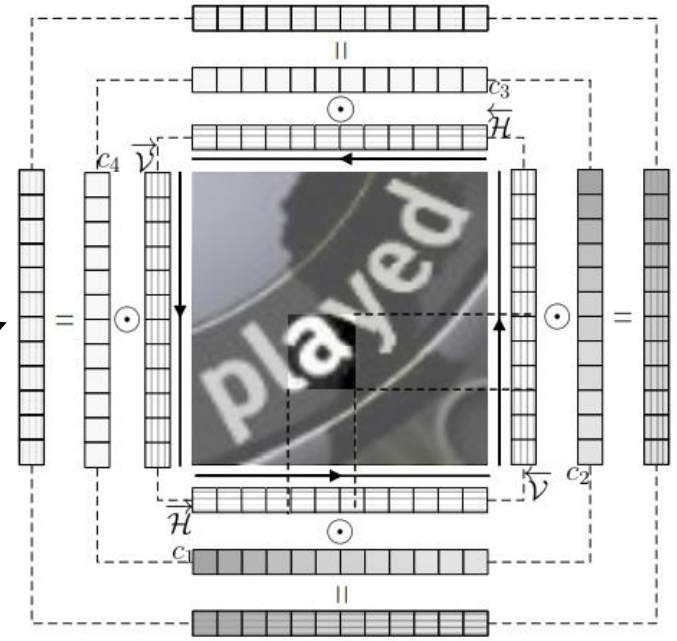


- Attention-based text recognition, like [1].
- Focusing Network to handle the 'attention drift'.

[1] B. Shi, et al. Robust Scene Text Recognition with Automatic Rectification. CVPR, 2016

AON: Towards Arbitrarily-Oriented Text Recognition [Cheng et al., CVPR, 2018.]



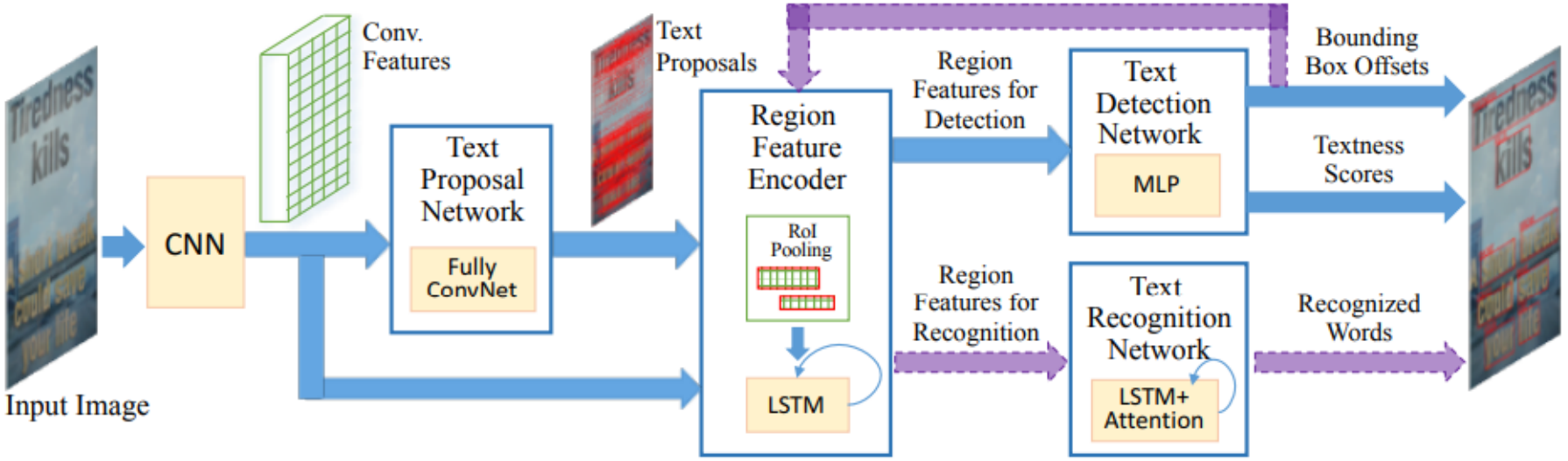


- Each pixel is represented by four weighted sequences of features.
- Attention-based text recognition.

Outline

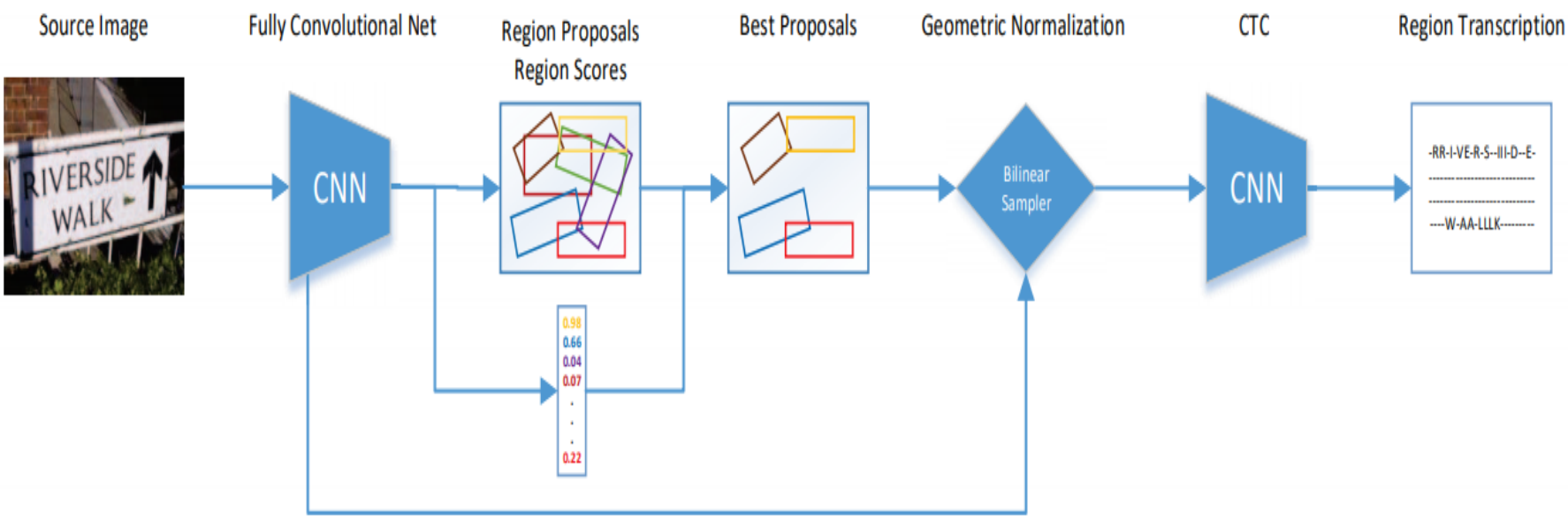
- Detection
- Recognition
- **End-to-end recognition**
- New datasets

Towards End-to-end Text Spotting with Convolutional Recurrent Neural Networks [Li et al., ICCV 2017]



- An end-to-end trainable network for end-to-end scene text recognition.
- A RPN subnet is used to detect scene text, an attention based network is used to recognize the detected text.
- Designed for horizontal scene text.

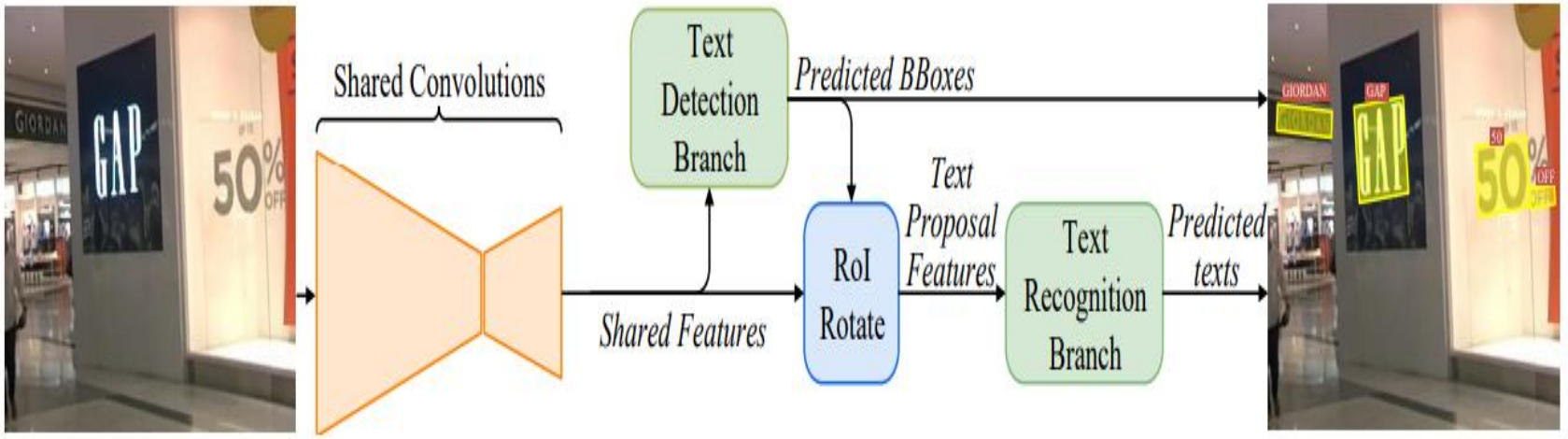
Deep TextSpotter: An End-To-End Trainable Scene Text Localization and Recognition Framework [Michal et al., ICCV 2017]



- An end-to-end trainable network which consists of a RPN for text proposal generation and a CRNN[1] for region transcription.
- Having the ability to detect and recognize horizontal and multi-oriented scene text.

[1] B. Shi, et al. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. TPAMI, 2017

FOTS: Fast Oriented Text Spotting with a Unified Network [Liu et al., CVPR 2018]



- Using EAST[1] as text detector.
- Using CRNN as text recognizer.
- Rotating rotated proposal to horizontal proposal with RoIRotate.
- End-to-end training and evaluation.

[1] X. Zhou, et al. EAST: An Efficient and Accurate Scene Text Detector. CVPR, 2017

Outline

- Detection
- Recognition
- End-to-end recognition
- **New datasets**

RCTW-17 dataset

- Chinese Text in the Wild(12,034 images, 8034 images for training and 4000 images for testing)
- The text annotated in RCTW-17 consists of Chinese characters, digits, and English characters, with Chinese characters taking the largest portion.
- Link: http://mclab.eic.hust.edu.cn/icdar2017chinese/
- State of the art: Detection 0.703 f-measure; Recognition 0.37 normalized edit distance

MLT: Multi-lingual scene text detection and script identification

- Multi-lingual text: 18,000 images, 9 different languages representing 6 different scripts
- Tasks: Text Detection, Script identification, Joint text detection and script identification
- Link: http://rrc.cvc.uab.es/?ch=8
- State of the art: Detection: AP=0.643, Hmean=0.735; Script identification: 88.1%; Joint detection and script identification



SCUT-CTW1500

- Curved text in the wild
- 1000 train images, 500 test images. Each contains at least one instance of curved text
- Link: https://github.com/Yuliang-Liu/Curve-Text-Detector
- State of the art: Hmean 0.734

Total-Text

- Horizontal, multi-oriented, and curved text
- 1555 images
- Link: https://github.com/cs-chan/Total-Text-Dataset
- State of the art: Hmean 0.36

Thank you!