

# Efficient Segmentation-aided Text Detection for Intelligent Robots

Junting Zhang, Yuewei Na, Siyang Li and C.-C. Jay Kuo

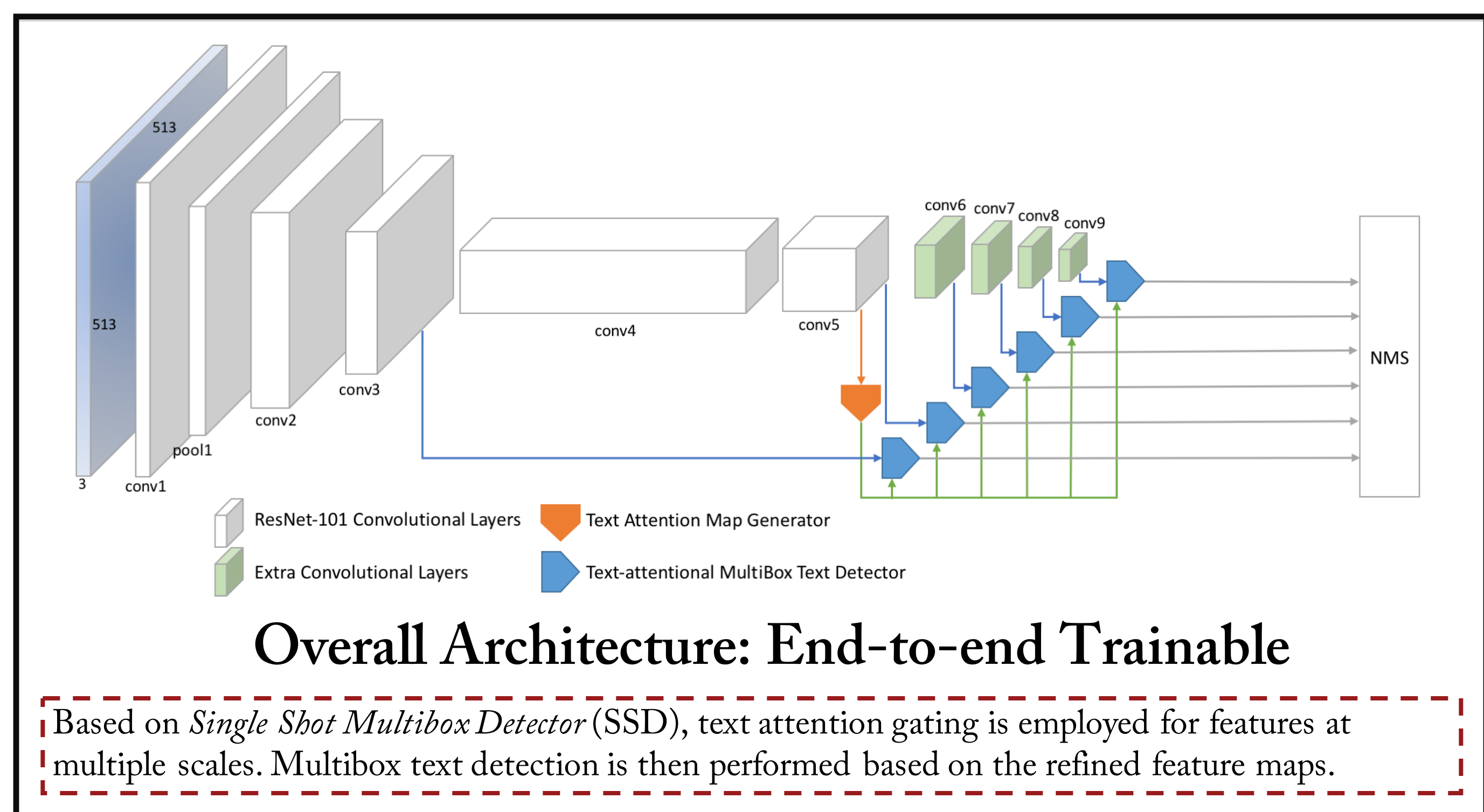
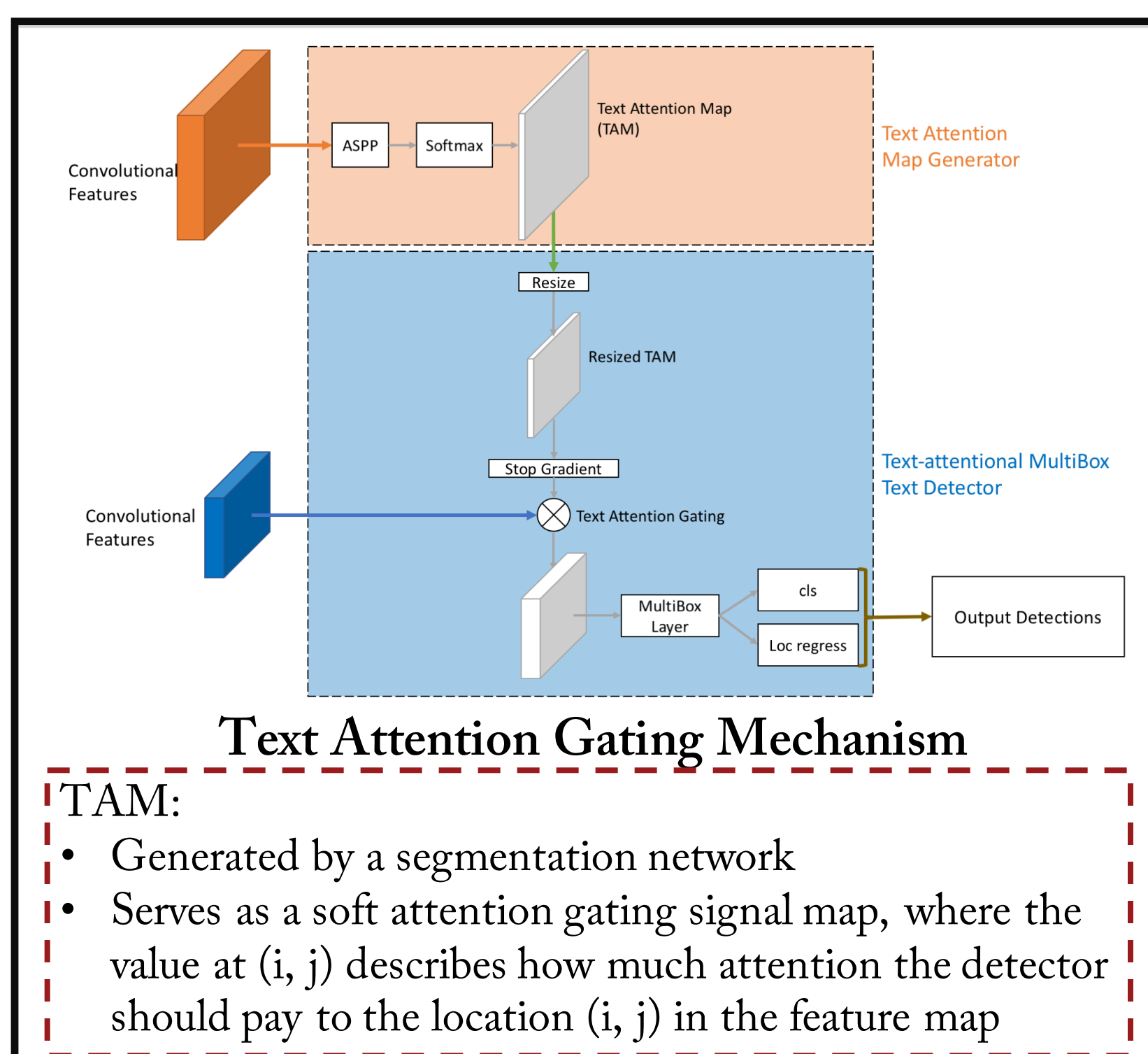
## Introduction

- Problem:
  - Detecting text in natural images (OCR in the wild)
  - Desired output: **word-level** bounding boxes
- Challenges: Variabilities in fonts, scales and layout as well as complex backgrounds and perspective distortion
- Previous Deep-Learning-based methods:
  - Detection-based: prone to text-like patterns due to the small receptive field
  - Segmentation-based: fail to produce fine-scale word-level bounding boxes due to the large receptive field

- Dataset:
  - COCO-Text Dataset
  - Largest publicly available dataset for text detection and recognition
  - Images are harvested from the Microsoft COCO dataset
  - Training/validation : 43,686/10,000



## Our Method



## Experiments

### Quantitative Results

Evaluations on COCO-Text-Legible validation set			
Models	Recall	Precision	F-Score
VGG-SSD	30.38	42.01	35.26
ResNet-SSD	34.42	46.14	39.43
<b>Ours</b>	<b>47.99</b>	<b>39.93</b>	<b>43.59</b>
Evaluations on COCO-Text-Full validation set			
Models	Recall	Precision	F-Score
Yao et al.	23.1	43.23	33.31
ResNet-SSD	35.4	31.03	27.17
<b>Ours</b>	<b>40.7</b>	<b>28.59</b>	<b>33.57</b>
Anonymous Baselines			
A	23.3	83.78	36.48
B	10.7	89.73	19.14
C	4.7	18.56	7.47

### Qualitative Results



Note: TAMs clearly highlight the text regions, and thus yield higher detection accuracy with reduced false positives.