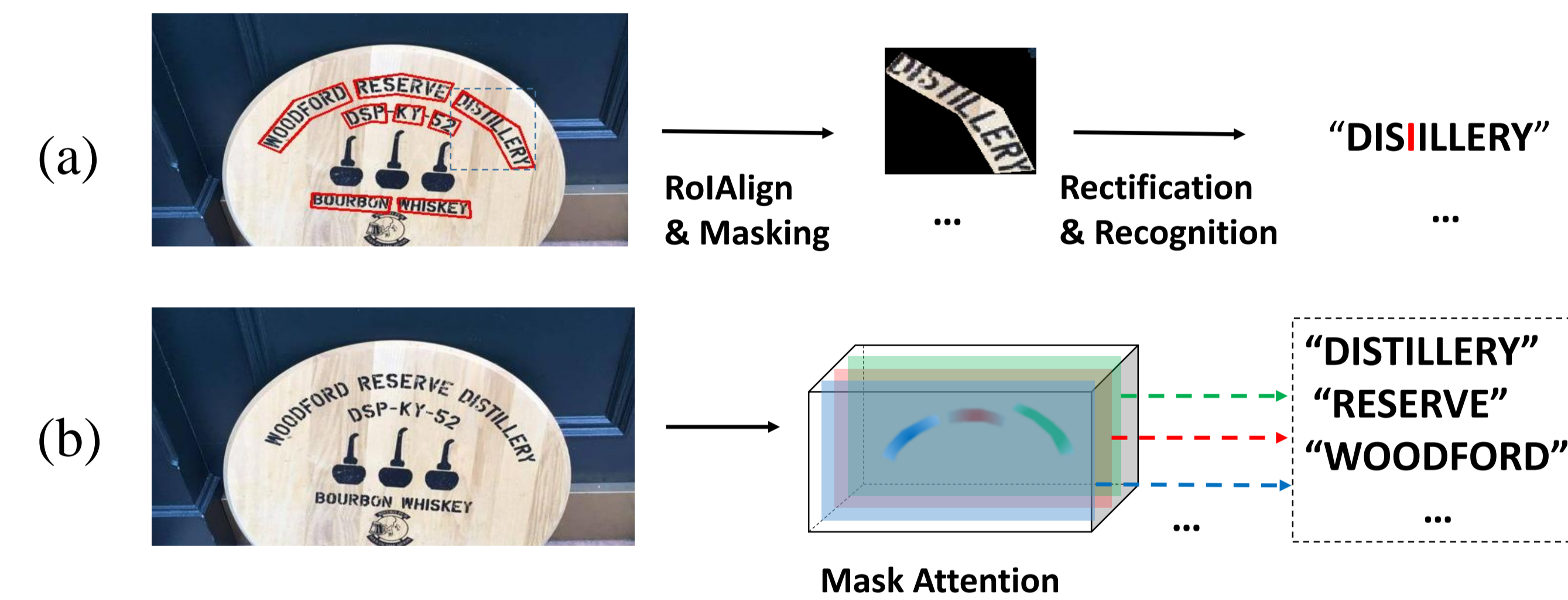




Liang Qiao¹ Ying Cheng² Zhanzhan Cheng³¹ Yunlu Xu¹ Yi Niu¹ Shiliang Pu¹ Fei Wu³
¹Hikvision Research Institute ²Tongji University ³Zhejiang University

MOTIVATION & CONTRIBUTIONS

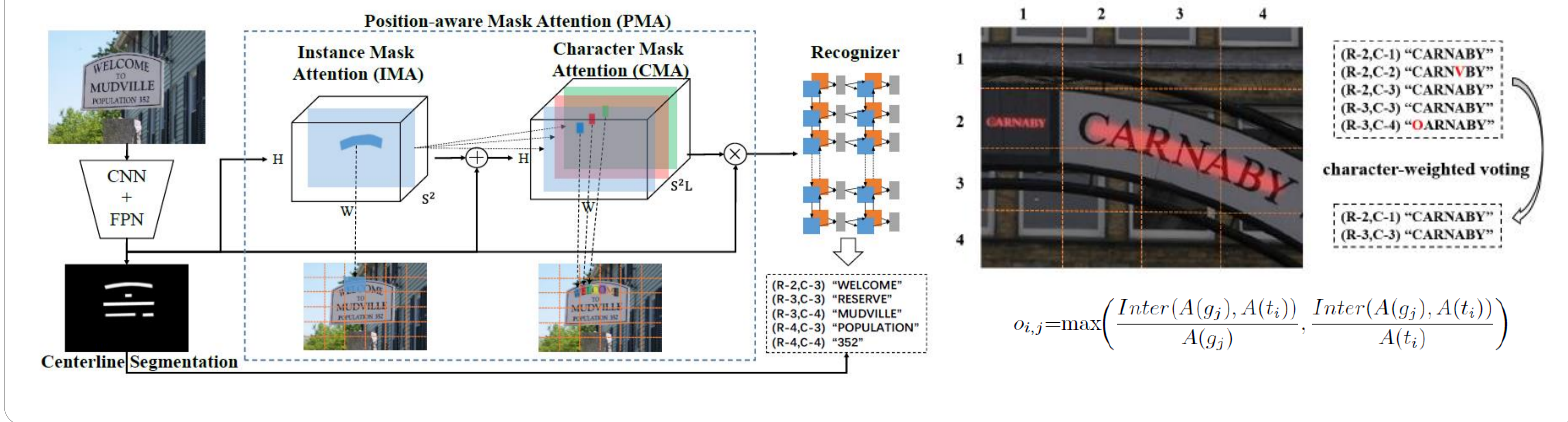
- Existing **two-staged** end-to-end text spotting methods adopt complicated **RoI operations** with **time-consuming NMS**.
- Recognition highly relies on detection results** and its annotations. Accurate boundaries need to be depicted during detection.



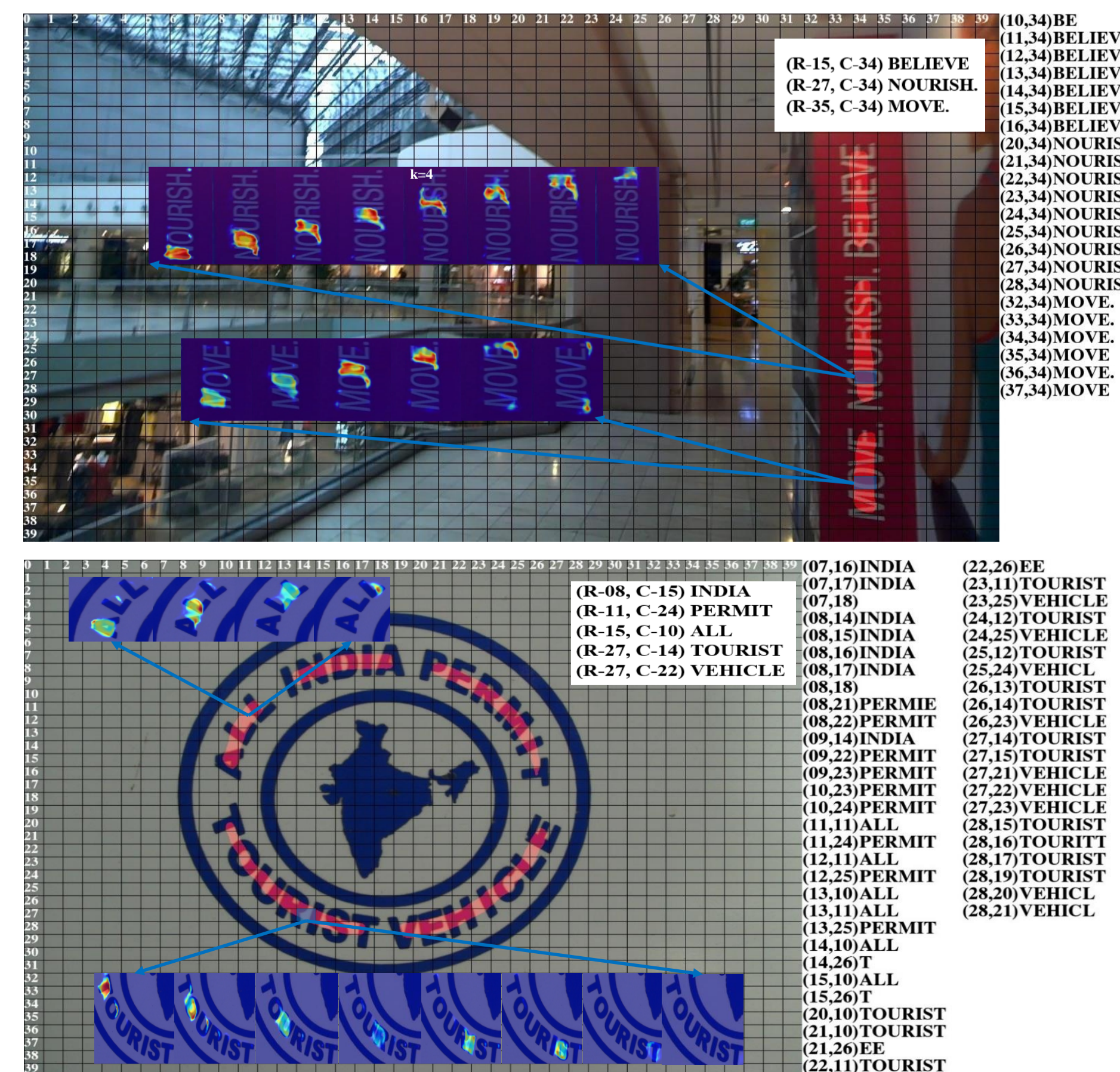
- We propose a compact and robust **one-stage text spotting framework** named MANGO that can be trained in an end-to-end manner.

- We develop the position-aware **mask attention module** to generate the text instance features into a batch, and build the one-to-one mapping with final character sequences. The module can be trained with **only rough text position information** and text annotations.

METHOD



EVALUATION & VISUALIZATION



Method	End-to-End		FPS
	None	Full	
Mask TextSpotter (Liao et al. 2019)	65.3	77.4	2.0
CharNet R-50 (Xing et al. 2019)	66.2	-	1.2
TextDragon (Feng et al. 2019)	48.8	74.8	-
Unconstrained (Qin et al. 2019)	67.8	-	-
Boundary (Wang et al. 2020a)	65.0	76.1	-
Text Perceptron (Qiao et al. 2020)	69.7	78.3	-
ABCNet (Liu et al. 2020)	64.2	75.7	17.9
MANGO (1280)	71.7	82.6	8.9
MANGO (1600)	72.9	83.6	4.3

Method	End-to-End		FPS
	None	Full	
Text Perceptron (Qiao et al. 2020)	57.0	-	-
ABCNet (Liu et al. 2020)	45.2	74.1	-
MANGO (1080)	58.9	78.7	8.4

Results on Total-Text

Method	Base(100k)	DB	FN	Rotate	Tilt	Weather	Challenge	AP
SSD300 + HC	98.3	96.6	95.9	88.4	91.5	87.3	83.8	95.2
RPnet(Xu et al. 2018)	98.5	96.9	94.3	90.8	92.5	87.9	85.1	95.5
MANGO	99.0	97.1	95.5	95.0	96.5	95.9	83.1	96.9

Results on SCUT-CTW1500

Results on CCPD without using detection annotation