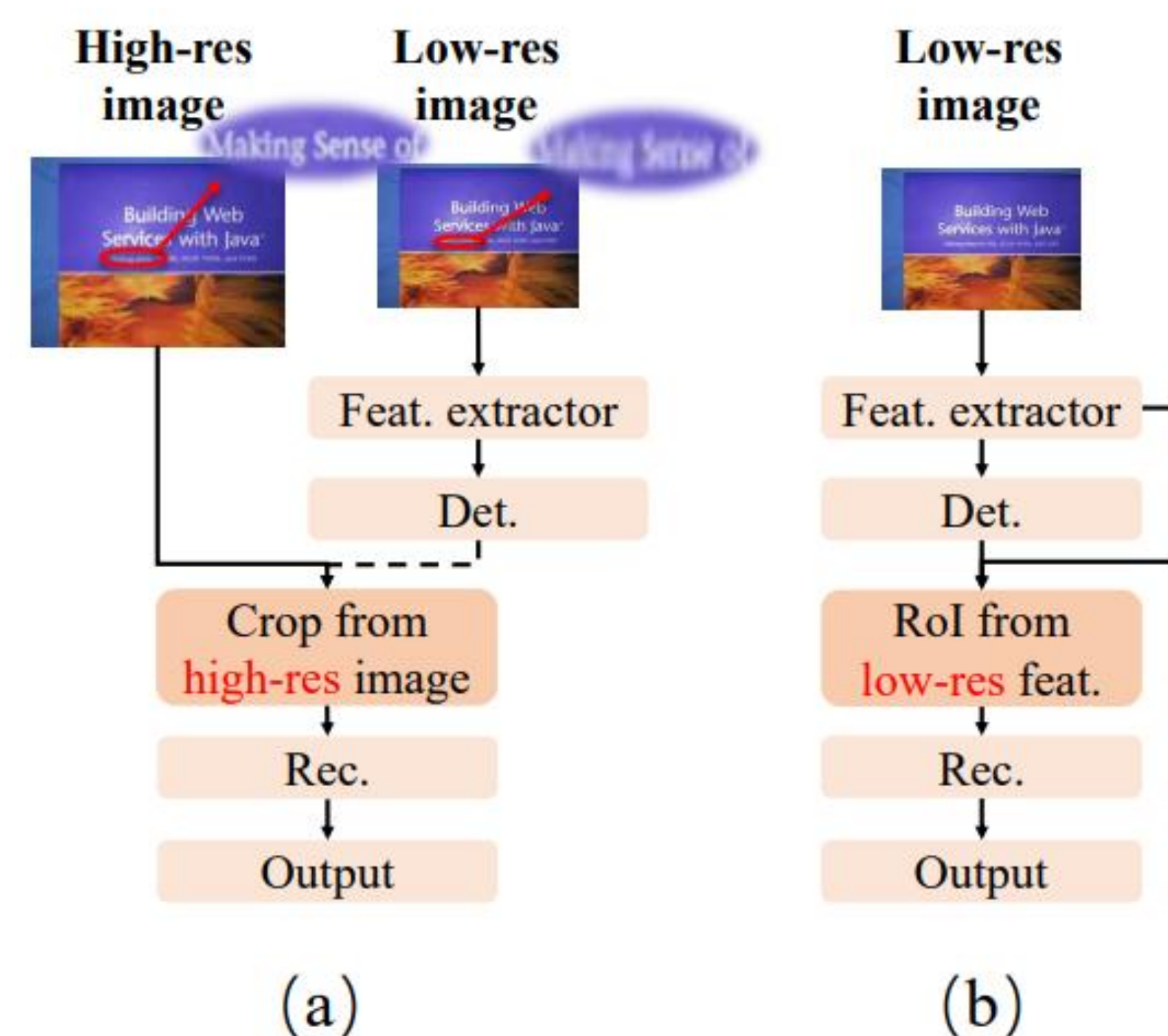


Dynamic Low-Resolution Distillation for Cost-Efficient End-to-End Text Spotting

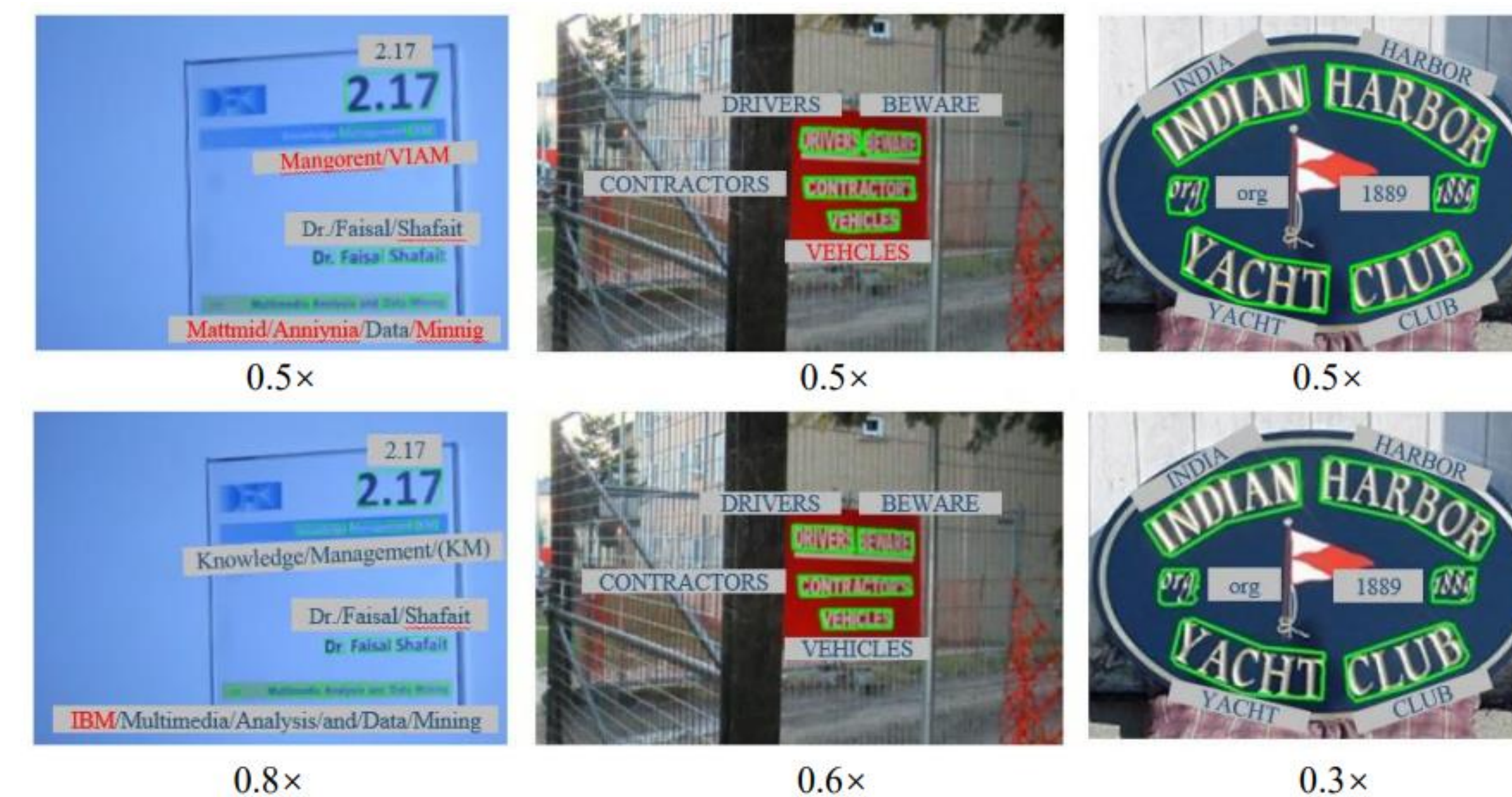
Ying Chen; Liang Qiao; Zhanzhan Cheng; Shiliang Pu; Yi Niu; Xi Li
Hikvision Research Institute; Zhejiang University

Motivations

- Existing text spotting methods always set a **fixed resolution**, which not **suitable** for all images.
- The input scale has always been a **tough trade-off** since recognizing a small text instance usually requires **enlarging the whole image**, which brings **high computational costs**.



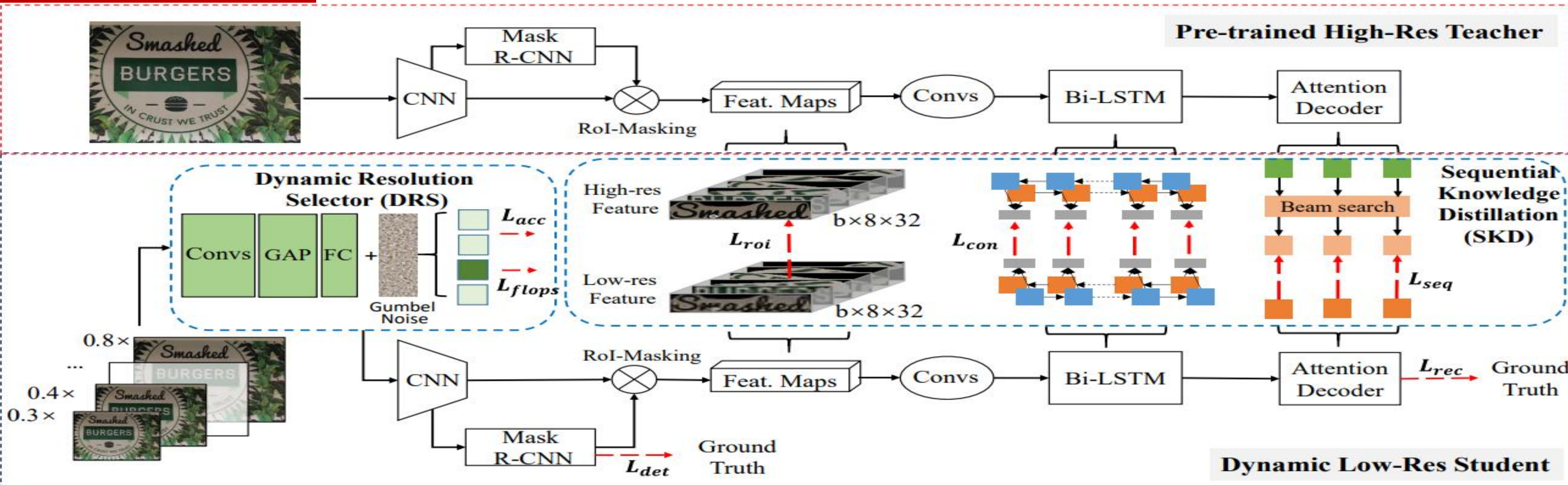
Visualize



When the text is clear, our proposed method tends to select a small scale.

When the text is blur, our proposed method tends to select a large scale.

Method

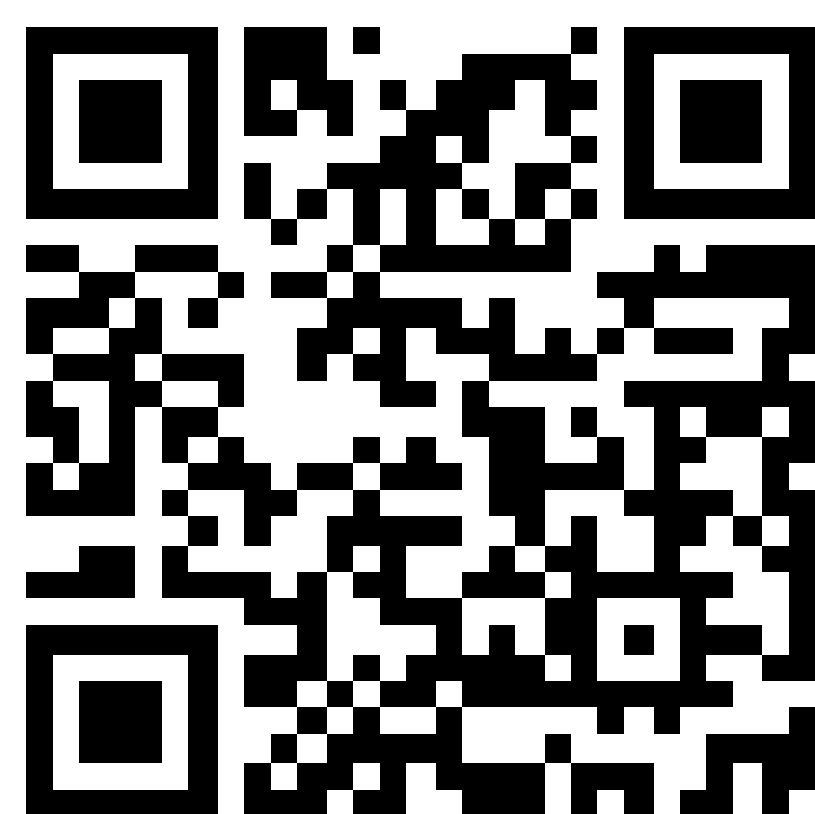


Performance

Dataset	Training Method	Type	Input Size	End-to-End (%)					Word Spotting (%)					FLOPS
				S	W	G	None	Full	S	W	G	None	Full	
IC13	Vanilla Multi-Scale	H	S-768	86.9	86.6	82.9	-	-	91.4	91.0	86.3	-	-	142.9G
	Vanilla Multi-Scale	L	S-384	80.9	78.9	74.4	-	-	85.2	82.7	77.3	-	-	35.8G
	SKD-only	L	S-384	84.1	82.8	78.8	-	-	88.0	86.5	81.7	-	-	35.8G
	DRS-only ($\gamma=0.1$)	L	Dynamic	85.7	84.8	80.7	-	-	90.1	88.9	84.0	-	-	80.7G
	DRS-only ($\gamma=0.3$)	L	Dynamic	83.7	82.0	77.6	-	-	87.8	85.8	80.5	-	-	48.8G
	DLD ($\gamma=0.1$)	L	Dynamic	86.5	85.7	82.7	-	-	90.9	89.9	86.1	-	-	71.5G
IC15	Vanilla Multi-Scale	H	S-1280	78.0	74.4	69.5	-	-	81.4	77.2	71.7	-	-	517.2G
	Vanilla Multi-Scale	L	S-640	72.2	67.8	62.9	-	-	75.7	70.8	65.3	-	-	129.3G
	SKD-only	L	S-640	75.4	71.7	67.1	-	-	78.9	74.6	69.6	-	-	129.3G
	DRS-only ($\gamma=0.1$)	L	Dynamic	76.2	72.1	66.8	-	-	79.8	75.2	69.3	-	-	298.8G
	DRS-only ($\gamma=0.3$)	L	Dynamic	73.6	68.9	63.7	-	-	76.4	71.5	66.3	-	-	163.6G
	DLD ($\gamma=0.1$)	L	Dynamic	79.0	75.7	70.9	-	-	82.4	78.6	73.3	-	-	261.8G
TT	Vanilla Multi-Scale	H	S-896	-	-	-	62.3	71.4	-	-	-	65.2	75.9	206.7G
	Vanilla Multi-Scale	L	S-448	-	-	-	55.4	66.5	-	-	-	58.1	71.1	52.0G
	SKD-only	L	S-448	-	-	-	59.6	68.9	-	-	-	62.6	73.5	52.0G
	DRS-only ($\gamma=0.1$)	L	Dynamic	-	-	-	60.9	70.4	-	-	-	63.5	75.0	119.2G
	DRS-only ($\gamma=0.3$)	L	Dynamic	-	-	-	58.8	68.9	-	-	-	61.6	73.6	75.0G
	DLD ($\gamma=0.1$)	L	Dynamic	-	-	-	63.9	73.7	-	-	-	66.4	77.8	103.0G
DLD ($\gamma=0.3$)	L	Dynamic	-	-	-	61.9	71.9	-	-	-	64.0	75.9	62.1G	

Contribution

- We first study the input resolution problem on end-to-end text spotting tasks and propose a Dynamic Low-resolution Distillation text spotting framework that can effectively enhance the performance and reduce the computational cost.
- We propose a sequential KD strategy with a dynamic resolution selector that allows the model to choose a small but recognizable input scale.
- Extensive experiments and ablation studies demonstrate the effectiveness of our method.



Scan Me for Paper (ECCV-2022)