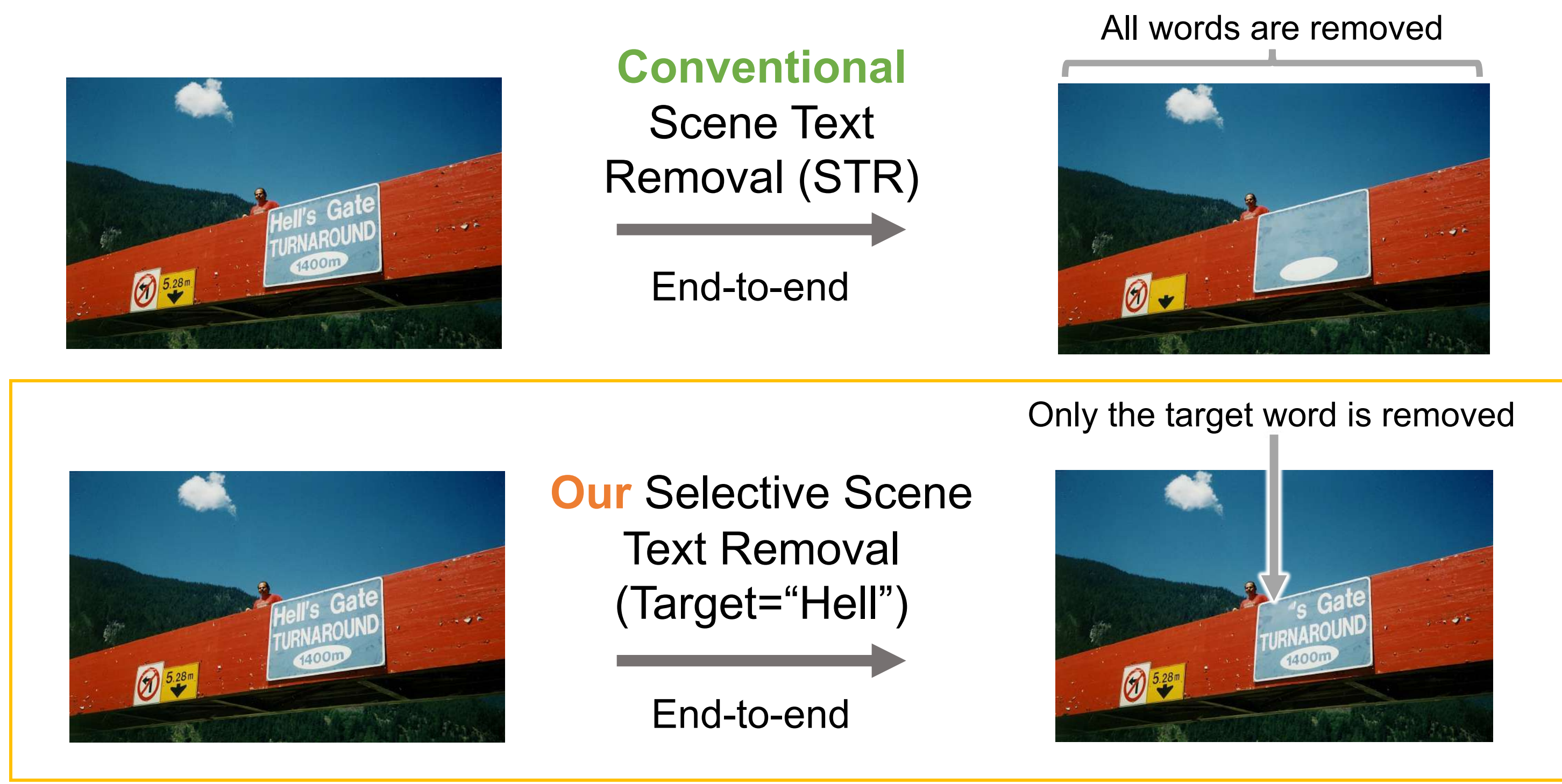
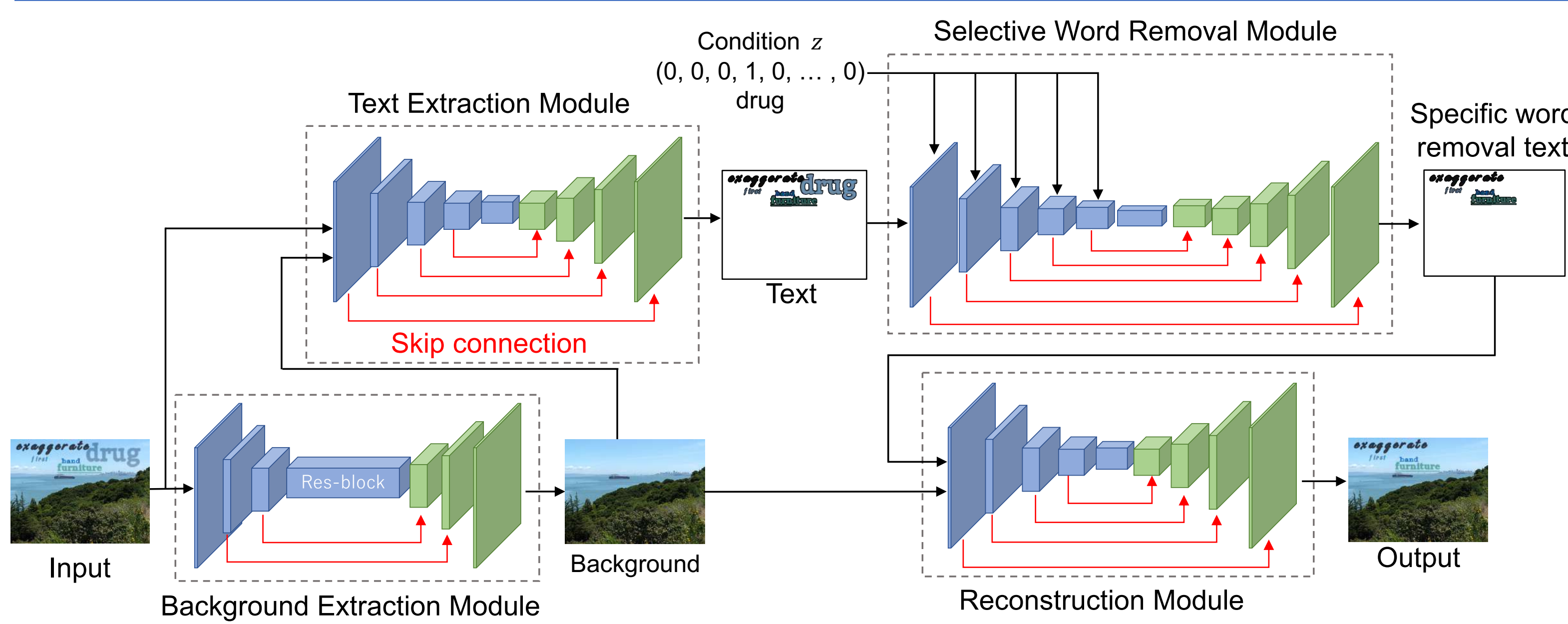


Hayato Mitani¹, Akisato Kimura², Seiichi Uchida¹
 (1Kyushu University, 2NTT Corporation) Email: hayato.mitani@human.ait.kyushu-u.ac.jp

Purpose

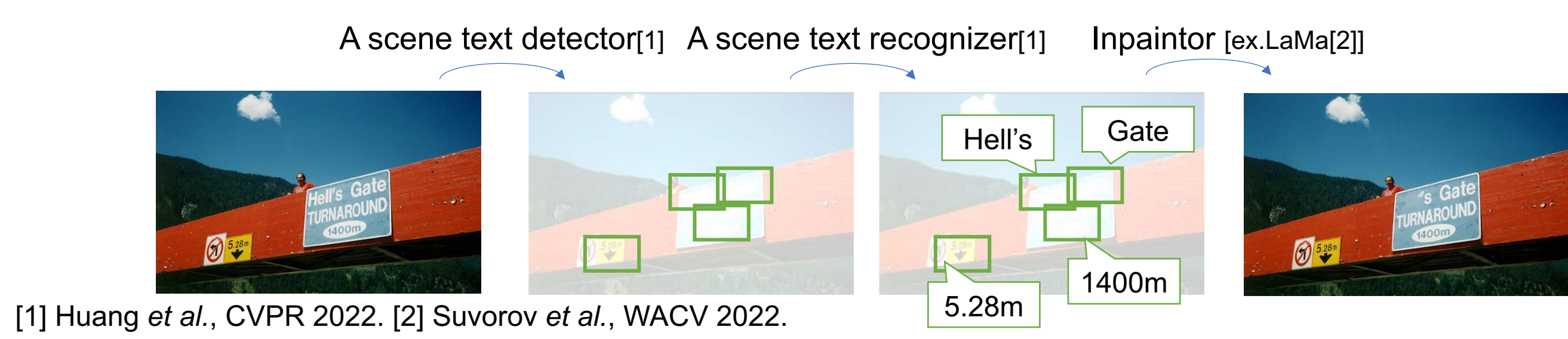


Our end-to-end SSTR model

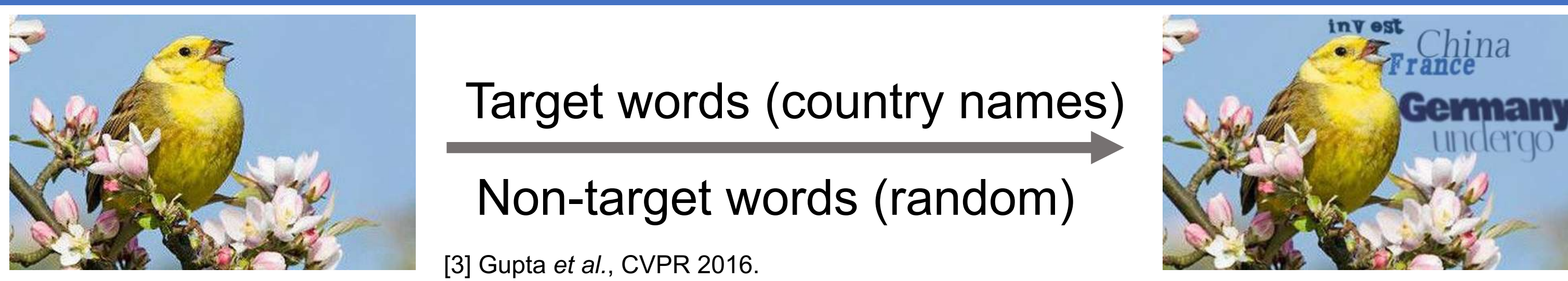


Comparative model (Multi-step model)

Naive SSTR realization by non-"end-to-end" multi-step processing



Synthetic Image Dataset by SynthText[3]

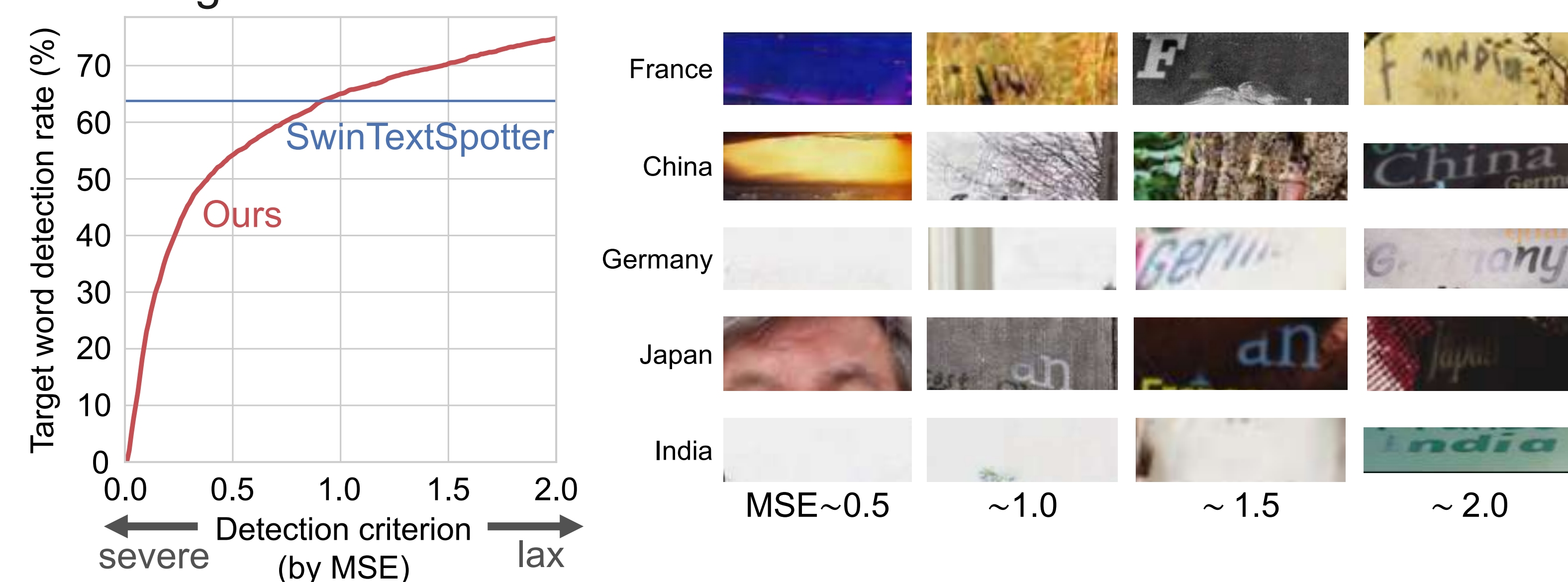


Quantitative Results (Synthetic images)

• R: Recall(%), SN: PSNR, M: MSE(%)

	Target word			Non-target word			Background	
	R↓	SN↑	M↓	R↑	SN↑	M↓	SN↑	M↓
Original Image	88.15	—	—	88.15	—	—	—	—
Ours	24.82	23.32	1.78	86.88	30.13	0.19	32.09	0.08

• Target word detection rate



• Computation time

	Ours	Multi-step
Computation time(s)↓	0.104	0.697

Qualitative Results (Synthetic images)



Qualitative Results (Real images)

