



Figure 1: Image generation with pretrained Stable Diffusion 1.5 at two output resolutions: (a) 512x512 and (b) 256x256. Images in each column are generated with the same prompt and random seed. Since Stable Diffusion 1.5 is optimized for 512×512 , **reducing the resolution to 256×256 noticeably degrades image quality.**

Table 1: Unsupervised object segmentation results on real-world datasets. We report CODA under two image resolutions 256×256 and 512×512 . On VOC, CODA still achieves the best performance among slot-based diffusion methods. On COCO, it attains the highest FG-ARI and is only worse than SlotAdapt on the remaining metrics.

VOC	FG-ARI \uparrow	mBO $^i\uparrow$	mBO $^c\uparrow$	mIoU $^i\uparrow$	mIoU $^c\uparrow$
SlotDiffusion †	17.8	50.4	55.3	44.9	49.3
SlotAdapt	29.6	51.5	51.9	-	-
CODA (256 \times 256)	33.23	54.28	56.81	50.00	52.84
CODA (512 \times 512)	32.23	55.38	61.32	50.77	56.30

COCO	FG-ARI \uparrow	mBO $^i\uparrow$	mBO $^c\uparrow$	mIoU $^i\uparrow$	mIoU $^c\uparrow$
SlotDiffusion †	37.2	31.0	35.0	31.2	36.5
Stable-LSD	35.0	30.4	-	-	-
SlotAdapt	41.4	35.1	39.2	36.1	41.4
CODA (256 \times 256)	46.33	32.42	35.81	32.16	36.79
CODA (512 \times 512)	47.54	36.61	41.43	36.41	42.60