Table 6: The ablation of mIoU improvement of each component on the nuScenes and WOD `val` split when trained only for the segmentation task. XSF and STD stand for cross-space transformer and segmentation transformer decoder.

Baseline (3.1)	STD	Multi-frame	XSF	nuScenes	WOD
✓				76.6	70.3
✓	✓			78.3 (+1.7)	70.6 (+0.3)
✓	✓	✓		80.8 (+4.2)	71.2 (+0.9)
✓	✓	✓	✓	**81.7** (+5.1)	**71.3** (+1.0)

Table 7: The ablation of the improvement of shared transformer decoder on the nuScenes `val` split when jointly trained with detection task.

Baseline [66]	XTF		XSF	mIoU	mAP	NDS
	Seg	Det				
✓				81.8	65.2	70.0
✓	✓			82.1 (+0.3)	65.4 (+0.2)	70.2 (+0.2)
✓		✓		82.4 (+0.6)	65.9 (+0.7)	70.3 (+0.3)
✓	✓	✓		82.6 (+0.8)	66.0 (+0.8)	70.2 (+0.2)
✓	✓	✓	✓	**82.7** (+0.9)	**66.6** (+1.4)	**70.8** (+0.8)