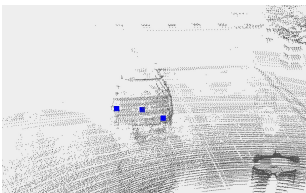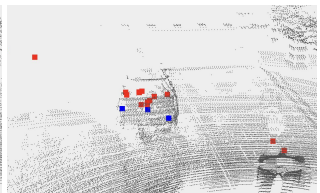Table 10: The ablation of XSF on the nuScenes `val` split. S→D and D→S denote sparse-to-dense (3b) and dense-to-sparse (3a) XSFs.

S→D	D→S	Add Convs	mIoU	mAP	NDS
			Segmentation Only		
✓	✓		81.7	-	-
	✓		80.9 (-0.8)	-	-
			Multi-task		
✓	✓		82.7	66.6	70.8
	✓	✓	82.8 (+0.1)	66.0 (-0.6)	70.5 (-0.3)

Direct Mapping   Cross-space Transformer

Figure 5: **Visualization of the learned offsets.** We showcase the features of a car's 3D voxels (blue) and their corresponding deformable offsets (red) that were learned in our XSF module. For a better visual representation, we only highlight the offsets with high attention scores.