Method	Vehicle	Pedestrian	Cyclist	Latency
FSHNet$_{light}$	72.5/72.0	77.9/72.6	77.2/76.1	81
- *SF*	70.5/70.1	77.1/71.6	75.6/74.5	65
- *SU*	71.4/70.9	75.9/70.4	76.6/75.4	71
- *DSLA*	70.8/70.3	77.8/72.2	76.1/75.0	81

Table 9. Comparison of different FSHNet variants. The latency (*ms*) is tested on a single RTX 3090 GPU. The LEVEL 2 AP/APH results (20% training data) are reported. '-' denotes FSHNet$_{light}$ without using corresponding module.

	Vehicle	Pedestrian	Cyclist	Latency
VoxelNext [4]	69.9/69.4	73.5/68.6	73.3/72.2	56
SAFDNet [46]	72.7/72.3	77.3/73.1	77.2.76.2	94
FSHNet$_{light}$	73.0/72.5	78.6/73.7	77.4/76.4	81
FSHNet$_{base}$	74.5/74.0	78.9/73.9	78.0/76.9	123

Table 10. Comparison of different sparse detectors. The latency (*ms*) is tested on a single RTX 3090 GPU. The LEVEL 2 AP/APH results (100% training data) are reported.

w	mAP/mAPH	Vehicle	Pedestrian	Cyclist
6	75.8/73.5	72.1/71.7	78.1/72.8	77.1/76.0
12	75.9/73.6	72.5/72.0	77.9/72.6	77.2/76.1
24	75.7/73.4	72.2/71.8	77.8/72.5	77.1/76.0
36	75.8/73.5	72.1/71.7	77.8/72.4	77.4/76.3

Table 11. Effect of different w settings. The LEVEL 2 AP/APH results (20% training data) are reported.

A. Runtime Analysis

In this section, we first discuss the latency of different components of FSHNet and then compare the inference latency of various sparse detectors. Models are trained on the Waymo Open dataset. All latency measurements were conducted on a single RTX 3090 GPU. As shown in the 1^{st} and 2^{nd} rows of Table 9, the SlotFormer (*SF*) block adds 16 *ms* of latency to FSHNet$_{light}$, while significantly enhancing detection performance, particularly for large objects. As demonstrated in the 1^{st} and 3^{rd} rows of Table 9, the sparse upsampling (*SU*) module introduces an additional 10 *ms* latency to the detector, yet it markedly improves performance on small objects. As illustrated in the 1^{st} and 4^{th} rows of Table 9, our dynamic sparse label assignment (*DSLA*) significantly boosts detection performance without adding latency.

We further compare the inference latency of our FSHNet with existing sparse detectors. As shown in Table 10, compared to the current state-of-the-art sparse detector SAFDNet, our FSHNet$_{light}$ exhibits lower inference latency and superior detection performance. Regarding our FSHNet$_{base}$, although it has relatively high inference latency, it greatly extends detection accuracy compared to existing sparse detectors.

B. Hyper-parameter Analysis

In this section, we determine the optimal value for the slot width w in Eq.1 through experiments on the Waymo Open dataset. The performance for different w settings is shown in Table 11, indicating minimal variations. This is due to our SlotFormer having a global receptive field that is independent of slot width. When $w = 12$, there is a slightly better performance compared to other settings. Thus, we adopt $w = 12$ as the default setting.

C. Visualizations

To provide an intuitive understanding of the slot partitioning manner and the sparse upsampling strategy, we present visual demonstrations of each. As shown in Figure 4, the slot partitioning process first scatters sparse voxels into grids and then groups them into different slots along the X- and Y-axes, respectively. Figure 5 illustrates the sparse upsampling strategy, where coarse voxels are initially compressed into smaller grids, followed by the application of sparse convolution to diffuse and refine them. Additionally, qualitative results are presented in Figure 6, demonstrating our method's ability to handle diverse and complex traffic scenes. The predicted boxes closely match the ground-truth boxes within an extensive detection range. However, we observe some missed or false detections in cases where objects are heavily occluded or located at extreme distances.

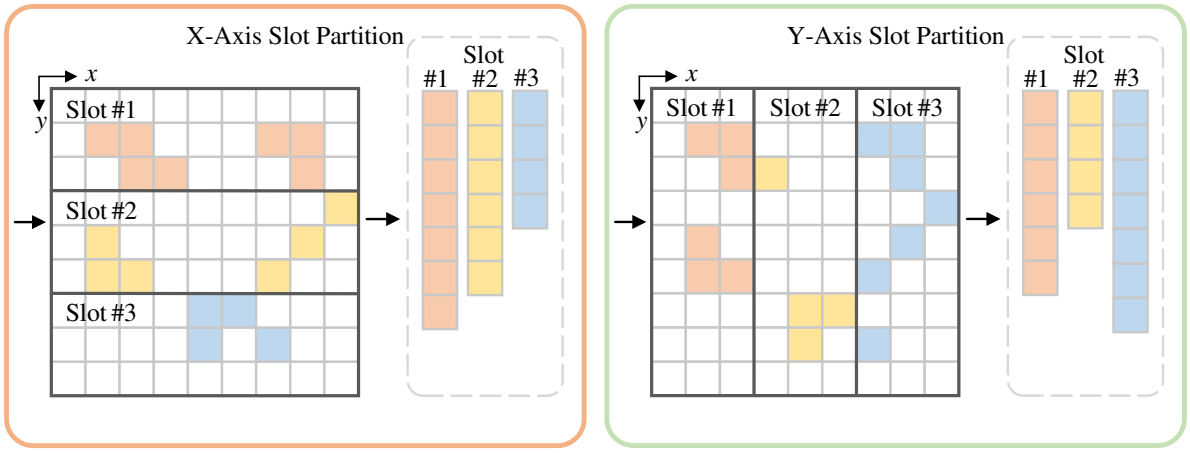

Figure 4. A demonstration of slot partition manner for sparse voxels.

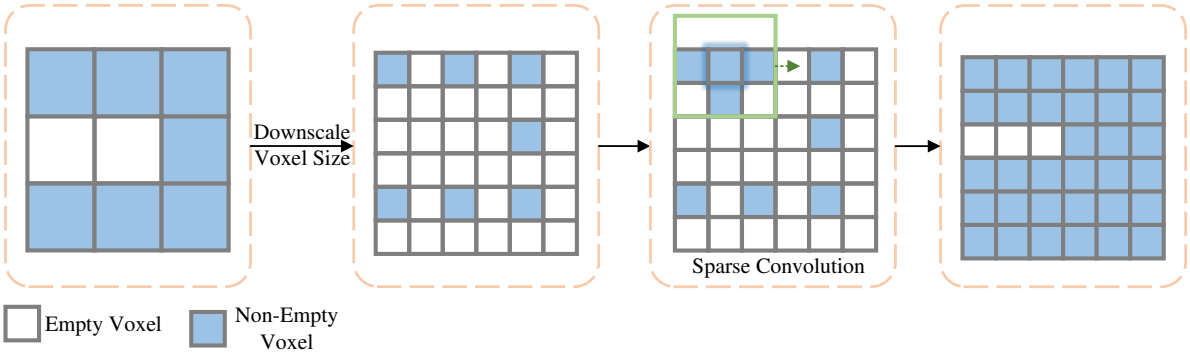

Figure 5. A demonstration of sparse voxel upsampling.

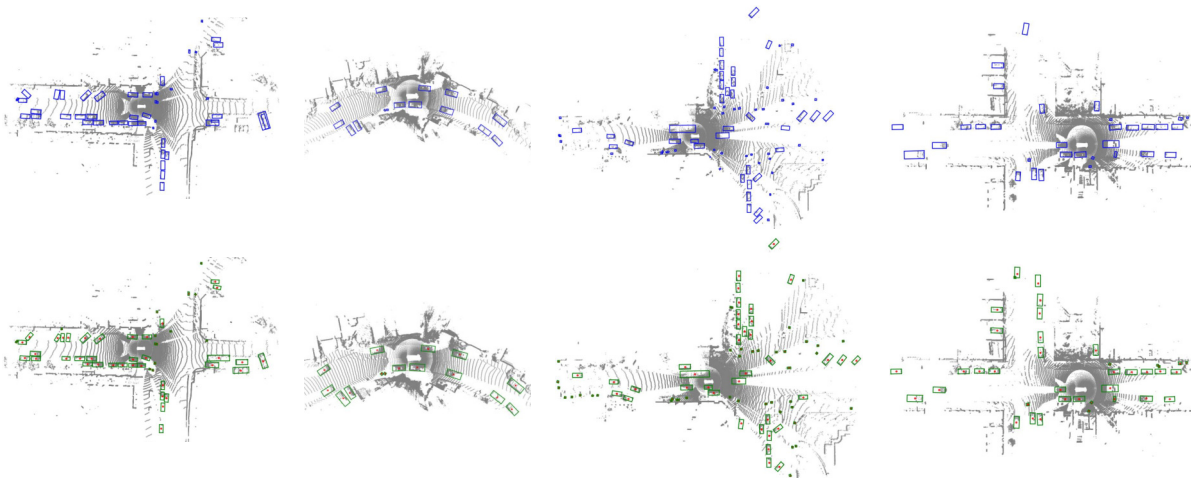

Figure 6. Qualitative results of FSHNet. The 1^{st} row demonstrates the ground-truth boxes with blue color, and the 2^{nd} row illustrates the predicted boxes with green color.