SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection

Appendix A Implementations details

We trained all models with a batch size of 16 on 8 RTX 4090 GPUs. We present more details as follows.

Waymo Open dataset. We set the voxel size to (0.08m, 0.08m, 0.15m), and the detection range to [-75.52m, 75.52m] in X and Y axes, and [-2m, 4m] in Z axis. We trained SAFDNet for 24 epochs on the training dataset and reported the evaluation results on the validation set to compare with previous methods. For results on the test set, we trained the model on both the training and validation sets. We employed the Adam [**?**] optimizer with a one-cycle learning rate policy, and set the weight-decay to 0.05, and the max learning rate to 0.003. We adopted the faded training strategy in the last epoch. During inference, we applied class-specific NMS with an IoU threshold of 0.75, 0.6 and 0.55 for vehicle, pedestrian, and cyclist, respectively. There are three category groups: group 1 includes vehicle; group 2 includes pedestrian and cyclist; group 3 is the background. The kernel size K for feature diffusion for the three groups are set to 7, 3, and 3, respectively.

nuScenes dataset. We set the voxel size to (0.075m, 0.075m, 0.2m), and the detection range to [-54m, 54m] in X and Y axes, and [-5m, 3m] in Z axis. We trained SAFDNet for 20 epochs on both training and validation sets and reported the evaluation results on the test set to compare with previous methods. We employed the Adam optimizer with a one-cycle learning rate policy, and set the weight-decay to 0.1, the momentum to [0.85, 0.95], and the max learning rate to 0.001. The faded strategy was used during the last 5 epochs. We set the query number of detection head to 300 and did not use any test-time augmentation. There are four category groups: group 1 includes bus and trailer; group 2 includes car, truck, and construction vehicle; group 3 includes motorcycle, bicycle, traffic cone, pedestrian, and barrier; group 4 is the background. The kernel size K for feature diffusion for the four groups are set to 15, 9, 3, and 3, respectively.

Argoverse2 dataset. We set the voxel size to (0.1m, 0.1m, 0.2m), and the detection range to [-200m, 200m] in X and Y axes, and [-4m, 4m] in Z axis. We trained SAFDNet for 24 epochs on the training set and reported the results on the validation set to compare with other methods. We employed the Adam optimizer with a one-cycle learning rate policy, and set the weight-decay to 0.05, and the max learning rate to 0.003. We adopted the faded training strategy in the last epoch. There are four category groups: group 1 includes large vehicle, bus, box truck, truck, truck cab, vehicular trailer, school bus, articulated bus, and message board trailer; group 2 includes regular vehicle; group 3 includes the other object categories; group 4 is the background. The kernel size K for feature diffusion for the four groups are set to 13, 7, 3, and 3, respectively.

3D Backbone	Other Parts	AFD	mAPH	Veh.	Ped.	Cyc.
HEDNet	Dense		73.2	72.1	72.0	75.6
HEDNet	Sparse		71.5	68.8	70.9	74.7
HEDNet	Sparse	✓	**73.3**	71.7	72.3	75.7
VoxelNet	Dense		71.4	69.8	70.9	73.4
VoxelNet	Sparse		69.9	66.1	70.6	72.9
VoxelNet	Sparse	✓	**72.0**	70.2	71.7	74.1
PillarNet	Dense		68.7	69.2	66.5	70.5
PillarNet	Sparse		67.2	65.6	66.8	70.2
PillarNet	Sparse	✓	**69.4**	69.7	67.5	71.4

Table 1. Adaptive feature diffusion(AFD) on different backbones.

Appendix B More experimental results

We conducted experiments with three 3D sparse backbones: HEDNet, VoxelNet, and PillarNet. Table 1 shows that the proposed AFD module worked well on all three backbones.