Supplemental Document of Scalable Autoregressive Monocular Depth Estimation

Jinhong Wang[1] Jian Liu[2] Dongqi Tang[2] Weiqiang Wang[2] Wentong Li[1]
Danny Chen[3] Jintai Chen[4†] Jian Wu[1†]

[1]ZJU [2]Ant Group [3]University of Notre Dame [4]HKUST(Guangzhou)
† Corresponding Authors

1. More Ablation Study

1.1. Bin strategy

We conduct additional experiments to show comprehensive performance comparisons for (1) MTBin compared to other bin strategies and (2) the length of the expanded to adjacent bins. The results are given below.

Table 1. Ablation study of different bin strategies on the NYU-Depth-v2 dataset.

Bin strategy	RMSE↓	δ_1↑
Space Increasing Bins [14]	0.233	0.975
Adaptive Bins [4]	0.229	0.977
Elastic Bins [49]	0.223	0.978
Multiway Tree Bins (no expand)	0.225	0.977
Multiway Tree Bins (expand 1 bin)	**0.217**	**0.979**
Multiway Tree Bins (expand 2 bins)	**0.217**	**0.979**
Multiway Tree Bins (expand 3 bins)	0.219	0.978

For (1), comparisons between existing bin strategies in the first block and our Multiway Tree Bins(MTBin) show that the optimal MTbin greatly exceeds the previous strategies, further demonstrating the superiority of our MTBin strategy. (2), by compareing MTbins with different expanded length in the second block, we can find that when expanding to one or two adjacent bins, MTbin achieves the best performance which validates the effectiveness of expanding strategy to maintain the model‘s error tolerance and expanding too long will affect the performance.

1.2. Bin number

To investigate the best setting of bin number N, we conduct ablation experiments on `NYU Depth V2` and `KITTI` with different bin numbers N. Table. 2 reports the results. It can be observed that the model achieves the best performance with a bin number equal to 16, thus we choose 16 as our final bin number.

Table 2. Ablation study of bin numbers on the NYU-Depth-v2 dataset and KITTI dataset.

Bin number N	NYU Depth V2			KITTI		
	Abs Rel ↓	RMSE ↓	δ_1 ↑	Abs Rel ↓	RMSE ↓	δ_1 ↑
8	0.061	0.219	0.978	0.048	1.840	0.982
16	**0.059**	**0.217**	**0.979**	**0.046**	**1.839**	**0.984**
32	0.060	0.219	0.977	0.047	1.842	0.982

2. Qualitative Results of Each Step

To further illustrate the model prediction details of each step, we visualize the depth map of each step, as shown in Fig. 1. For better comparison, we scale the depth map to the same size as the input RGB image. We can observe that as the step progresses, the depth prediction becomes more accurate and smooth. This further illustrates the effectiveness of our granularity autoregressive objective.

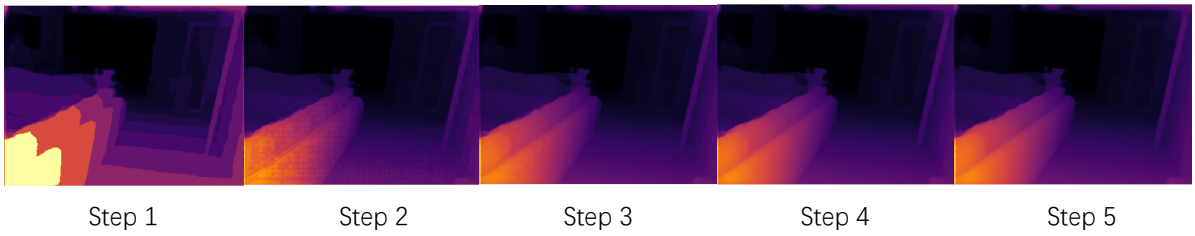

Figure 1. Depth map visualization of each step on `NYU Depth V2`.