Point Cloud Compression with Sibling Context and Surface Priors

Supplementary Material

Zhili Chen, Zian Qian, Sukai Wang, and Qifeng Chen

The Hong Kong University of Science and Technology
{zchenei,zqianaa,swangcy,cqf}@ust.hk

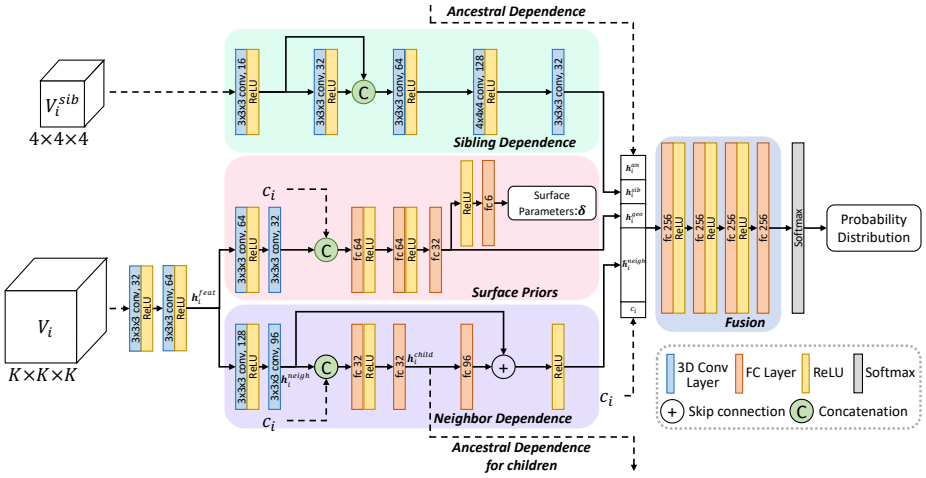

Fig. 1. The detail of our deep entropy model design.

1 Model Design

The detail of our deep entropy model is illustrated in Figure 1. Octant n_i's information $c_i \in \mathbb{R}^4$ consists of two parts, its located octree level and its corresponding spatial coordinates. The sibling and neighbor context of the octant n_i are presented as $V_i^{sib} \in \mathbb{R}^{4\times4\times4}$ and $V_i \in \mathbb{R}^{9\times9\times9}$ in binary voxel representations. The numbers of the channel are one for both V_i and V_i^{sib}. The ancestor information $h_i^{an} \in \mathbb{R}^{32}$ is an extracted feature map passed from the upper level. $h_i^{child} \in \mathbb{R}^{32}$ is the feature map passed to n_i's children. The total number of parameters in our deep entropy model is 1.77M.

2 Additional Quantitative Results

Indoor scene compression results. To show our proposed compression method have consistent performance in both indoor and outdoor scenes, we further eval-

Dataset	Method	BPP↓ Level 5	Level 6	Level 7	Level 8	Level 9
ScanNet [2]	VoxelContext [4]	0.043	0.156	0.647	2.383	5.255
	Ours	**0.036**	**0.126**	**0.538**	**2.164**	**4.948**

Table 1. The quantitative results on ScanNet datset [2] when compared to VoxelContext without any refinement. The reconstructed point clouds of two methods are the same at each level.

Method	BPP↓ Level 5	Level 6	Level 7	Level 8	Level 9
Ours w/o Surface	0.036 (+0.0%)	0.126 (+0.0%)	0.542 (+0.7%)	2.198 (+1.6%)	5.045 (+2.0%)
Ours	**0.036**	**0.126**	**0.538**	**2.164**	**4.948**

Table 2. Ablation study of our entropy model on ScanNet [2] without using context from surface priors.

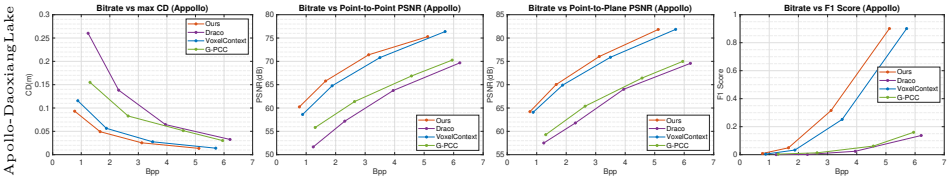

Fig. 2. The cross-dataset quantitative results of our method on the Apollo-DaoxiangLake dataset [5]. The result shows that our model has better generalization ability than baselines.

uate the compression performance on the ScanNet dataset [2]. ScanNet is a large-scale dataset that captures dense point clouds from real-world indoor scenarios. We sample 80,000 points from each scan and use the official training/testing splits [2] for training and testing. The training and the testing sets consist of 1,201 and 312 point clouds, respectively. We construct the octree with a maximum level of 9 in ScanNet dataset. The compression performance of our framework is evaluated by truncating octree levels ranging from 5 to 9 to vary the compression bitrates. The corresponding spatial quantization error ranges from 9.49 cm to 0.59 cm. As illustrated in Table 1, our method saves 6.21% to 24.13% bitrates on the Scannet dataset compared to VoxelContext.

To show that surface priors are not only effective in the outdoor scene, we further train the model of "w/o Surface" to compare with our whole model on the ScanNet dataset [2]. As illustrated in Table 2, the result shows that "Ours w/o Surface" triggers an additional bitrate cost of up to 2.0%, demonstrating the surface priors' effectiveness in indoor and outdoor scenes.

Cross-dataset results. To better show our model's generalization ability, we evaluate the cross-dataset performance on the Apollo-DaoxiangLake dataset [5]

Method	BPP↓				
	Level 8	Level 9	Level 10	Level 11	Level 12
Ours w/o Multi-level	0.158 (+6.0%)	0.435 (+6.4%)	1.067 (+6.8%)	2.290 (+7.2%)	4.189 (+8.0%)
Ours w/o Sibling	0.161 (+8.1%)	0.439 (+7.3%)	1.071 (+7.2%)	2.288 (+7.1%)	4.133 (+6.6%)
Ours w/o Neighbor	0.156 (+4.7%)	0.432 (+5.6%)	1.061 (+6.2%)	2.279 (+6.6%)	4.155 (+7.1%)
Ours	**0.149**	**0.409**	**0.999**	**2.137**	**3.878**

Table 3. Ablation study of our entropy model on KITTI [3]. The first row compares a shared-weight version of our proposed entropy model to our multi-level framework. The second and the third rows are the ablation studies of our entropy model without using sibling context and neighbor context, respectively.

with our model trained on the KITTI Odometry dataset. The Apollo-DaoxiangLake dataset [5] captures point clouds in different driving scenes with the same type of LiDAR as the KITTI Odometry dataset. As shown in Fig. 2, the evaluation results show our models still have competitive compression performance over the baselines with better reconstruction quality. In terms of bitrates, our method saves 11.2-13.7% compared to VoxelContext, 56.91-137.35% to Draco, and 68.83-125.08% to G-PCC, respectively.

3 Additional Ablation Study

To demonstrate the effectiveness of our multi-level framework design, we train a shared-weight entropy model for all octree levels to compare with our multi-level framework. The experiment results are shown in the first row of Table 3. The results show that it is essential to train an independent entropy model for each level to capture resolution-specific context across different levels.

We further ablate over the contextual feature by training the model without exploiting features from neighbors. In this experimental setting, we deleted the neighbor dependence branch as shown in Fig.1 and used the feature map from the Surface Priors branch as the ancestor information passed to the next level. To fairly compare with the other ablation experiments, we increase the channel number of h_i^{sib} from 32 to 96. As shown in the second and the third row of Table 3, incorporating sibling context leads to a more significant performance than the neighbor context. The ablation results in Table 3 demonstrate that our newly proposed sibling context is the main factors contributing to non-trivial compression improvement.

4 Model Complexity

We evaluate the average inference time of our whole model and models without incorporating specific context on a single frame point cloud data. As illustrated in Table 4, the results are calculated by averaging the run time of our entropy

Method	Run Time(s)				
	Level 8	Level 9	Level 10	Level 11	Level 12
Ours w/o Sibling	0.029	0.045	0.053	0.057	0.061
Ours w/o Surface	0.037	0.064	0.092	0.097	0.101
Ours w/o Ancestor	0.070	0.133	0.161	0.166	0.171
Ours	0.068	0.135	0.174	0.180	0.185

Table 4. Model inference time on a single frame point cloud data on different octree levels. The results are the average run time over our training dataset sampled from KITTI Odometry [3].

Method	MACs	#Params
Ours w/o Sibling	16.83M	1.03M
Ours w/o Surface	34.28M	1.59M
Ours w/o Ancestor	37.32M	1.76M
Ours	37.34M	1.77M

Table 5. Ablation study on model complexity in terms of the number of multiply-accumulate operations (MACs) and the number of parameters (#Params).

model on the training dataset sampled from KITTI Odometry [3] with 6000 frames. The model inference time is evaluated across different octree levels.

As illustrated in Table 5, the number of multiply-accumulate operations (number of parameters) of "our model", "w/o siblings", "w/o surface", and "w/o ancestor" are 37.34M (1.77M), 16.83M (1.03M), 34.28M (1.59M), and 37.32M (1.76M), respectively.

5 Additional Qualitative Results

We demonstrate additional qualitative results of our method against VoxelContext [4] cross different octree levels on two datasets KITTI Odometry [3] and nuScenes [1], as illustrated in Figure 3. The first set of the graph indicates the reconstructed point cloud at level 10, the second set of the graph indicates the reconstructed point cloud at level 11, and the third set of the graph indicates the reconstructed point cloud at level 12. The result shows that our model can reconstruct point clouds with a smaller error at lower bitrates compared to VoxelContext.

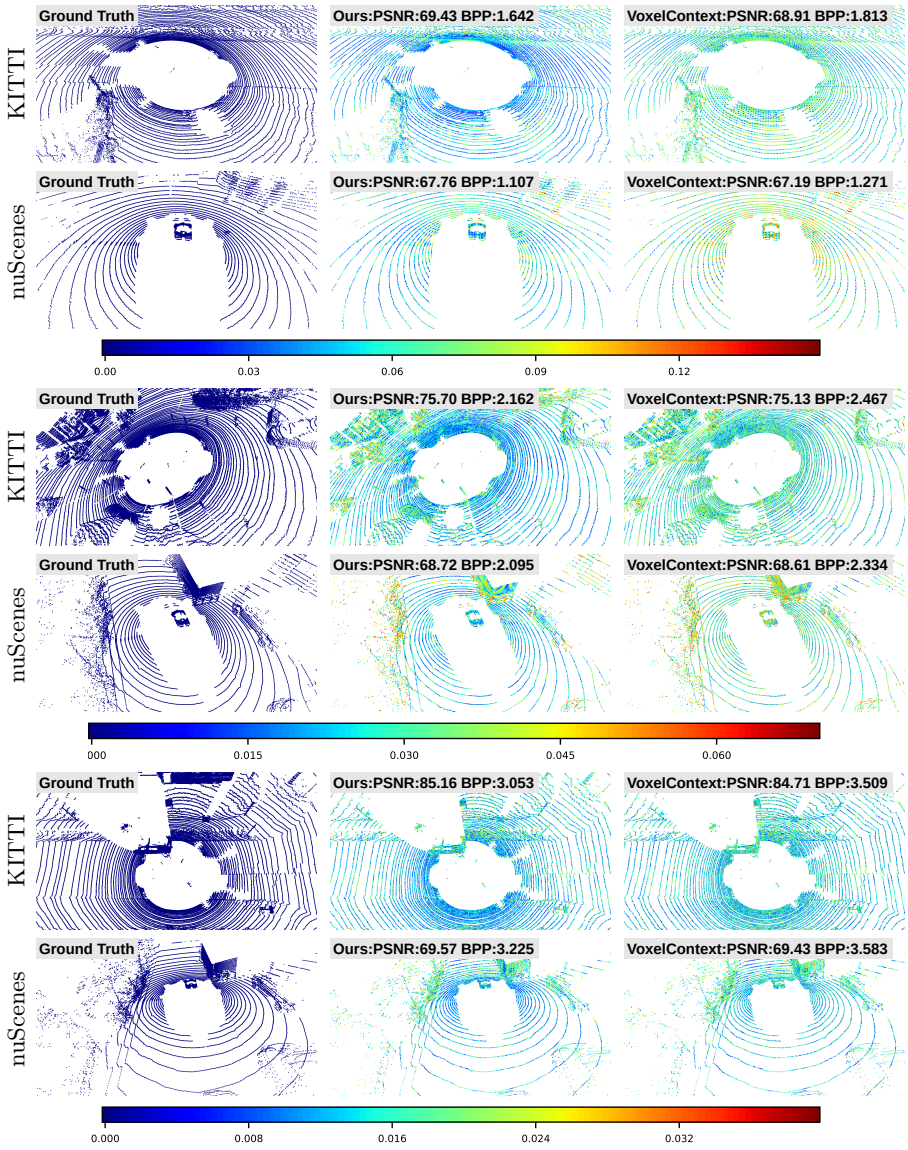

Fig. 3. Additional qualitative results of our method compared with VoxelContext on the KITTI Odometry dataset (first row of each figure) and the nuScenes dataset (second row of each figure) cross different octree levels. The first figure indicates level 10, the second figure indicates level 11, and the third figure indicates level 12.

References

1. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: CVPR (2020)
2. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: CVPR (2017)
3. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: CVPR (2012)
4. Que, Z., Lu, G., Xu, D.: Voxelcontext-net: An octree based framework for point cloud compression. In: CVPR (2021)
5. Zhou, Y., Wan, G., Hou, S., Yu, L., Wang, G., Rui, X., Song, S.: Da4ad: End-to-end deep attention-based visual localization for autonomous driving. In: ECCV (2020)