Learnable Skeleton-Aware 3D Point Cloud Sampling (Supplementary Materials)

Cheng Wen Baosheng Yu Dacheng Tao
School of Computer Science, The University of Sydney, NSW 2008, Australia
cwen6671@uni.sydney.edu.au; baosheng.yu@sydney.edu.au; dacheng.tao@gmail.com

1. Network Architecture

We show the detailed architecture of the skeleton estimation network in Fig. 1 and the sampling network in Fig. 2, respectively. Both two networks adopt the DGCNN [8] as the backbone, which is composed of four EdgeConv layers. The outputs of the EdgeConv layers are concatenated together and go through the MLP layer, followed by a max pooling to get the global features. Then we repeat the concatenation operation again to get the contextual features. The following MLPs are for processing the contextual features to predict the weight matrix. In our network, each MLP layer is followed by a batch normalization and a LeakyReLU non-linearity. The output of the skeleton estimation network is a set of skeletal spheres and the output of the sampling network is a subset of points.

2. Temperature τ

In our main paper, we have given our different annealing strategies when training the sampling network with Gumbel-softmax operation. Here we visualize the annealing curves in Fig. 3, including step, linear and exponential. In practice, we rectify τ to avoid the gradient blowing up.

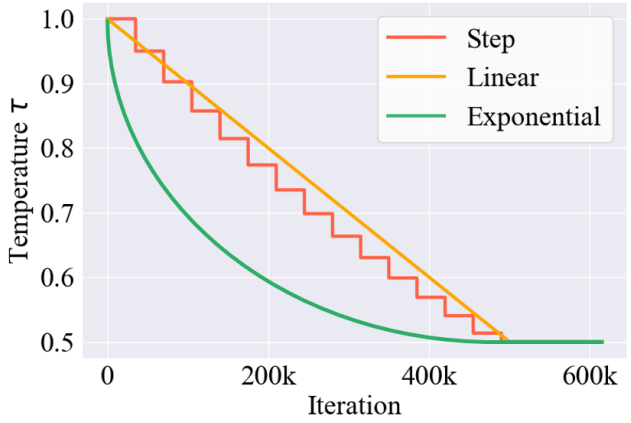


Figure 3. Different annealing strategies, including step, linear and exponential. We rectify τ at 0.5 in case of gradient blowing up.

3. Visualization of Sampled Points

In this section, we visualize the sampling examples from the classification task. As shown in Fig. 4, we show the results of four different sampling ratios.

4. Skeleton Estimation

As we estimate the skeleton in an unsupervised manner, a byproduct of our work is the predicted skeletal spheres. Following [4], we evaluate our method on the ModelNet40 [10] dataset from two aspects. We first measure the difference between the shapes reconstructed from the skeletons and the ground truth shapes using the Chamfer distance (CD) [1] and Hausdorff distance (HD), which are denoted as CD-Recon and HD-Recon, respectively. Then, we again use the CD and HD to measure the difference between the output skeletal representations and the ground truth skeleton, denoted as CD-Skel and HD-Skel. In this paper, we use the skeletons computed by DPC [9] as the ground truth. These distances are calculated by randomly sampled points from the respective geometries.

We compare with three works, two traditional ones, i.e., mean curvature skeleton [7] and L_1-medial skeleton [2], and one learning-based, i.e., Point2Skeleton [4]. We use the code released by the authors for comparison. As shown in Table 1, we present the quantitative results with these three methods using the four metrics aforementioned. In addition, we also report the results of our method with different numbers of predicted skeletal spheres. In Fig. 5, we visualize the examples of these methods. The L_1-medial skeleton [2] and mean curvature skeleton [7] can only produce 1D curves, thus resulting in large errors when used to abstract non-tubular shapes. In contrast, our method can generate more compact and structurally meaningful skeletal representations for various geometries. Overall, the results validate that our method not only more accurately encodes the information from the original input, but also produces more reasonable skeletonization results that are geometrically meaningful.

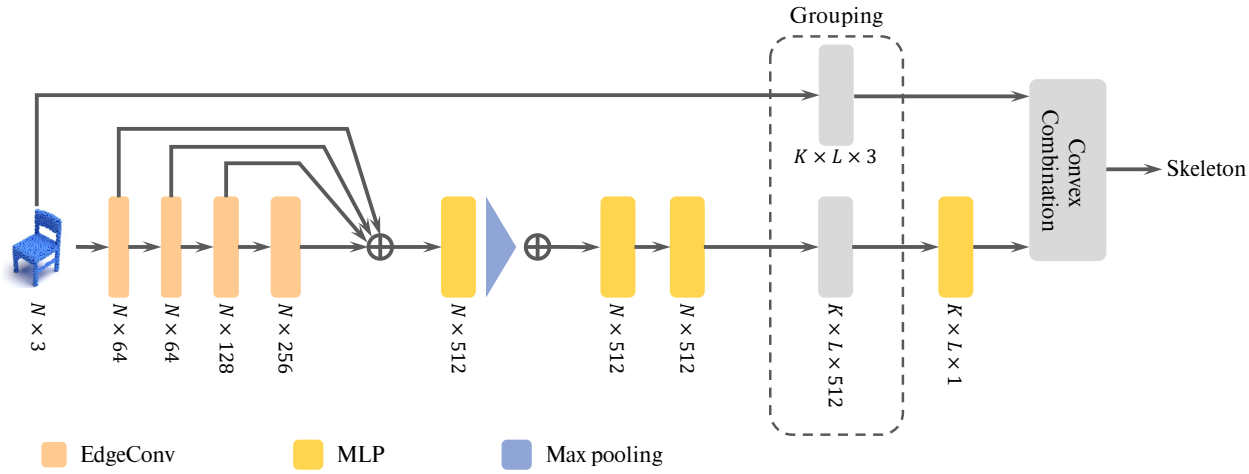


Figure 1. The detailed architecture of the proposed skeleton estimation network.

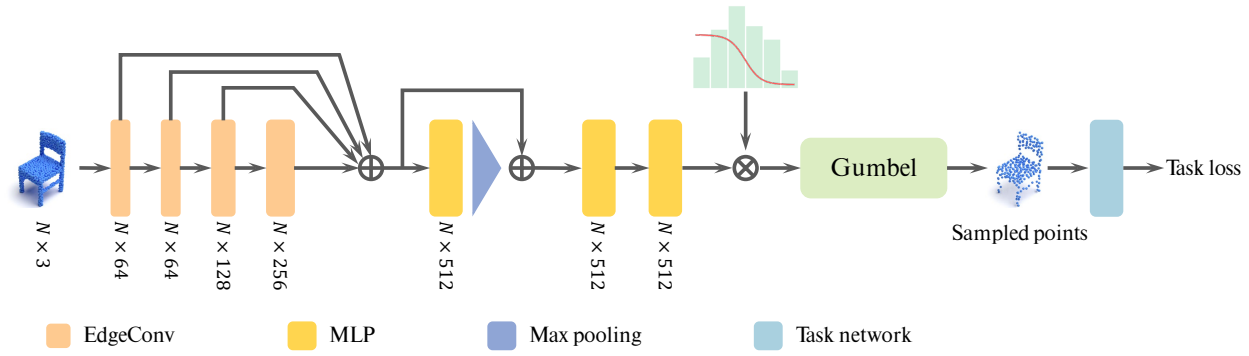


Figure 2. The detailed architecture of the proposed sampling network.

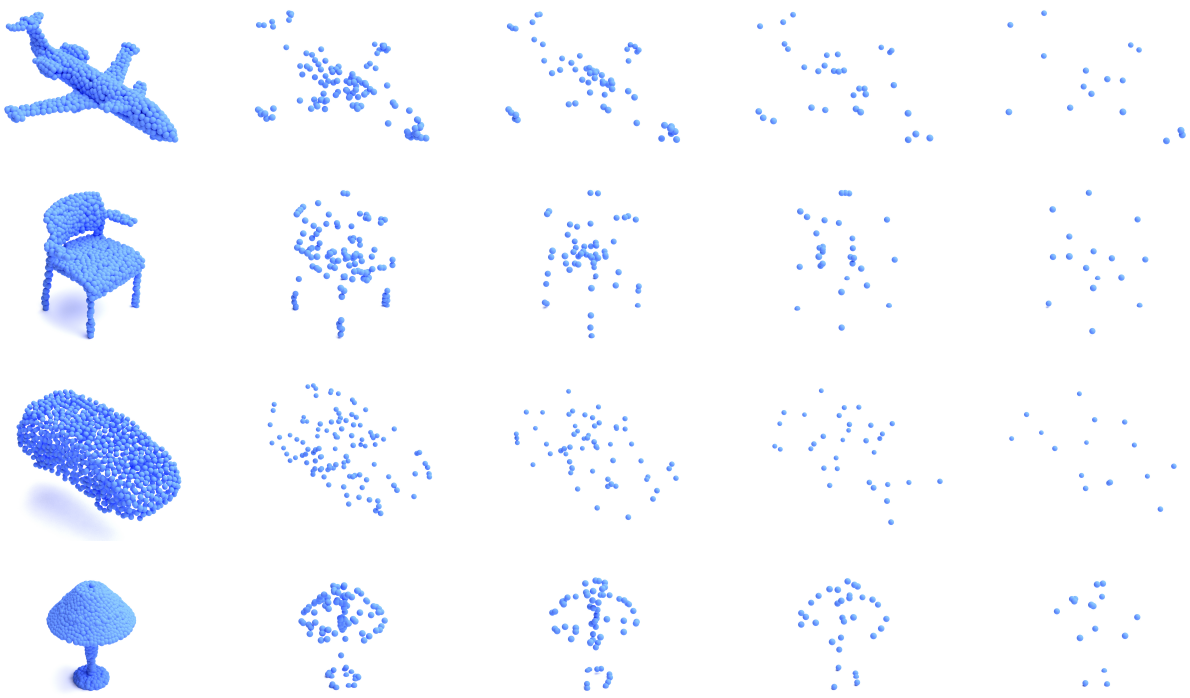
Method	CD-Recon	HD-Recon	CD-Skel	HD-Skel
L1 [2]	0.0651	0.0476	0.1632	0.1287
MCS [7]	0.0517	0.0308	0.1145	0.0865
Point2Skeleton [4]	0.0334	0.0164	0.0887	0.0551
Ours-64	0.0352	0.0265	0.0916	0.6729
Ours-128	0.0303	0.0159	0.0860	0.0561
Ours-256	0.0314	0.0167	0.0854	0.0559

Table 1. Quantitative comparisons with other point cloud skeletonization methods, including L_1-medial skeleton (L1), mean curvature skeleton (MCS) and Point2Skeleton. In addition, we also report the results of our method with different numbers of predicted skeletal spheres, namely 64, 128 and 256 spheres.

5. Point Cloud Registration

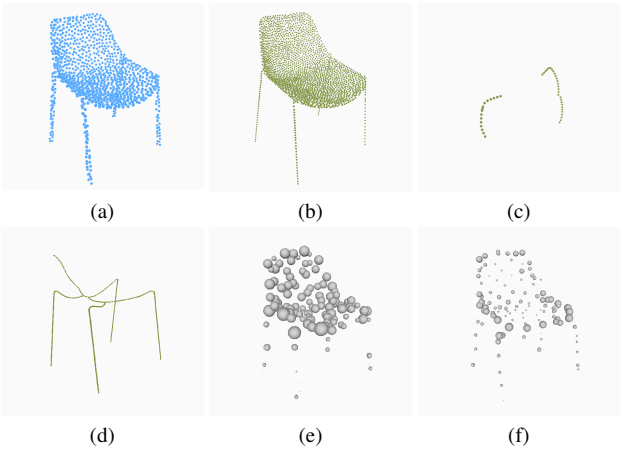
In our main paper, we have shown the performance of different point cloud tasks, including classification, retrieval and reconstruction. In this section, we will report our performance on point cloud registration. For a fair comparison, we follow the setting of SampleNet [3] to conduct this experiment. Specifically, we employ the PCRNet [6] to construct a point cloud registration network and use the car category with 1024 points in ModelNet40 [10] dataset to train this network. We use 4925 pairs of source and template point clouds from the training split and 100 source-template pairs from the testing split for performance evaluation. The template is rotated by three random Euler angles in the range of [-45, 45] to obtain the source.

We first train the PCRNet on complete point clouds with



(a) 1024 points (b) 128 points (c) 64 points (d) 32 points (e) 16 points

Figure 4. Sampled points from the classification experiment. From left to right: input with 1024 points, 128 points, 64 points, 32 points and 16 points.



(a) (b) (c) (d) (e) (f)

Figure 5. Visualization of skeletons of different works. (a) Input point cloud; (b) Ground truth skeleton (DPC [9]); (c) L_1-medial skeleton [2]; (d) Mean curvature skeleton [7]; (e) Point2Skeleton [4]; (f) Ours. In (e) and (f), we show the skeletal spheres which consist of the coordinates of the sphere centers and their radii.

two losses: the ground truth rotation and the Chamfer distance [1] between the registered source and template point clouds. Then, we freeze PCRNet and apply the same sampling strategy to both the source and template. The evaluation metric of registration performance is the mean rotation error (MRE) between the estimated and the ground truth rotation in angle-axis representation. We compare with three methods, RS, FPS and SampleNet, and report the performance in Table 2. As shown, our method outperforms others by a large margin, and the MRE of the proposed sampling remains low with the increasing sampling ratio. For example, at the sampling ratio of 32, the MRE with our method is 3.73°, while SampleNet results in an MRE of 5.94°. The performance validates our approach to be an efficient sampling method for the registration task, overcoming the challenge of sampling two different point clouds.

Sampling Ratio	MRE			
	RS	FPS [5]	SN [3]	Ours
8	10.42	5.66	4.88	**2.99**
16	16.69	8.07	5.48	**3.30**
32	24.52	13.46	5.94	**3.73**
64	36.99	21.69	7.95	**4.54**

Table 2. Point cloud registration results on ModelNet40. RS: random sampling; FPS: farthest point sampling; SN: SampleNet.

References

[1] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 605–613, 2017. 1, 3

[2] Hui Huang, Shihao Wu, Daniel Cohen-Or, Minglun Gong, Hao Zhang, Guiqing Li, and Baoquan Chen. L1-medial skeleton of point cloud. *ACM Transactions on Graphics (TOG)*, 32(4):65–1, 2013. 1, 2, 3

[3] Itai Lang, Asaf Manor, and Shai Avidan. Samplenet: Differentiable point cloud sampling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7578–7588, 2020. 2, 3

[4] Cheng Lin, Changjian Li, Yuan Liu, Nenglun Chen, Yi-King Choi, and Wenping Wang. Point2skeleton: Learning skeletal representations from point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4277–4286, 2021. 1, 2, 3

[5] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5099–5108, 2017. 3

[6] Vinit Sarode, Xueqian Li, Hunter Goforth, Yasuhiro Aoki, Rangaprasad Arun Srivatsan, Simon Lucey, and Howie Choset. Pcrnet: point cloud registration network using pointnet encoding. *CoRR*, 2019. 2

[7] Andrea Tagliasacchi, Ibraheem Alhashim, Matt Olson, and Hao Zhang. Mean curvature skeletons. In *Computer Graphics Forum*, volume 31, pages 1735–1744. Wiley Online Library, 2012. 1, 2, 3

[8] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 38(5):1–12, 2019. 1

[9] Shihao Wu, Hui Huang, Minglun Gong, Matthias Zwicker, and Daniel Cohen-Or. Deep points consolidation. *ACM Transactions on Graphics (TOG)*, 34(6):1–13, 2015. 1, 3

[10] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, 2015. 1, 2