PointSpherical: Deep Shape Context for Point Cloud Learning in Spherical Coordinates

Hua Lin, Bin Fan, Yongcheng Liu, Yirong Yang,
Zheng Pan, Jianbo Shi, Chunhong Pan, Huiwen Xie
hua.lin@nlpr.ia.ac.cn, bin.fan@ieee.org

CASIA UISEE 驭势科技
中国科学院大学 University of Chinese Academy of Sciences

Introduction

The two critical distinguishing concepts in a convolutional neural network (CNN) are 1) **receptive field**: convolution has an explicit encoding neighborhood in xy plane, and 2) **parts**: the max-pooling has an explicit delineation of the parts. For 3D point cloud recognition and segmentation tasks, it has been challenging to find the right representation to operationalize these two concepts of receptive field and parts. We directly address the concepts of the receptive field and parts in 3D point cloud. Starting from N 3D point clouds, we sample M points using the furthest point method. We center a spherical coordinate on each 3D sampling point, divide the space into radial (r), polar angular (ϕ), and azimuthal angular (θ) bins. We apply 1x1 CNN convolution on each point and apply max pooling over points that fall within each bin to extract its ShapeContext feature. We apply 3D CNN operations, convolution, and max-pooling, on the spherical bins. Just as in 2D CNN, after a fixed number of stages, we condense information by merging nearby bins in (r, ϕ, θ) space after max-pooling. After three stages of the convolution, max-pooling, and subsampling in our current experiment setting, we reduce the information into one central bin.
Our idea can be thought of as 'picture-in-picture': we took a fovea image-like snapshot of the 3D neighborhood around each point. Each local Shape Context could be thought of as a tiny fovea picture inside the whole 3D shape volume: it has detailed shape information nearby and a blurred picture of faraway points. As such, we regained the concept of the receptive field: a regular-shaped neighbor for spatial convolution.

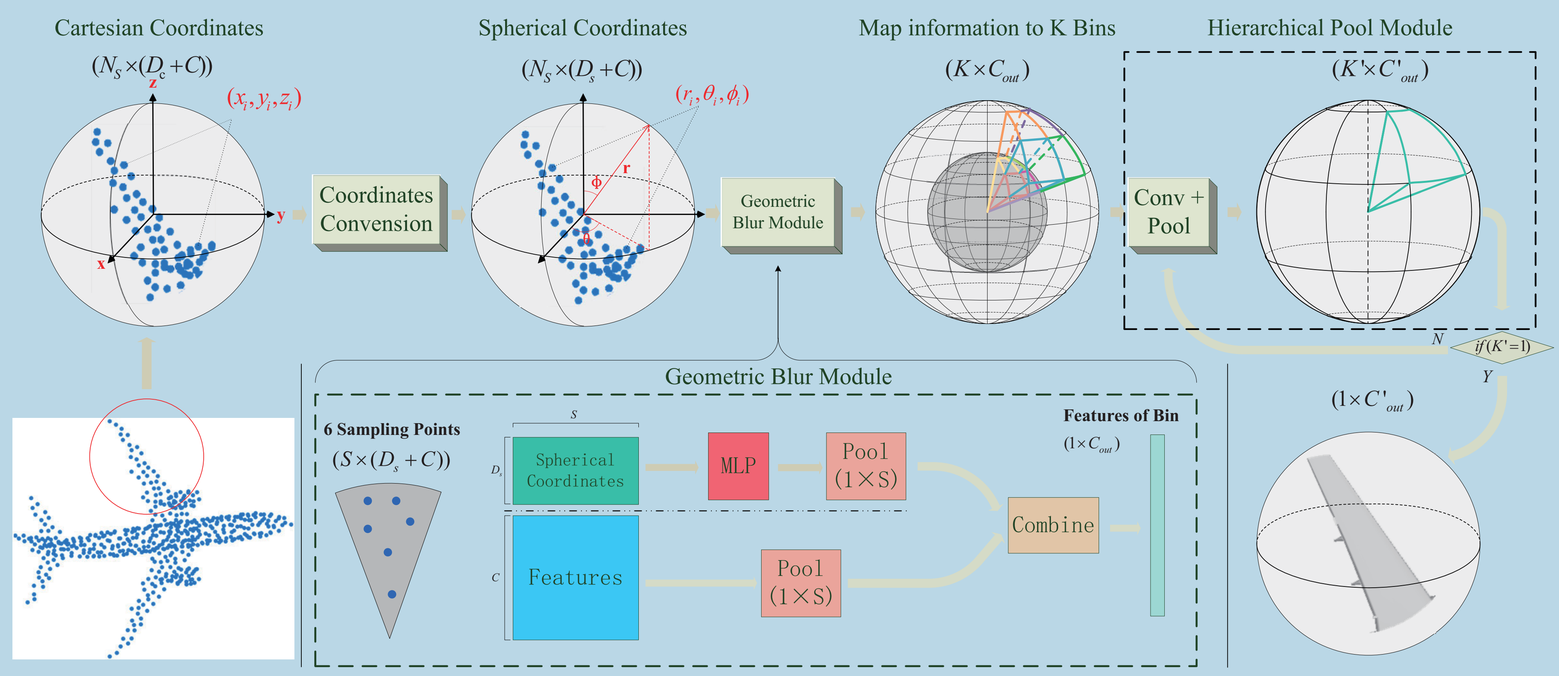
Conclusion

In this paper, a novel method is proposed for 3D point cloud modeling, Spherical Hierarchical. Inspired by Shape Context, we design a receptive field on each 3D point by placing a spherical coordinate on it. We sample points using the furthest point method and creating overlapping balls of points. For each ball, space is divided into radial, polar angular, and azimuthal angular bins on which we form a Spherical Hierarchy. 1x1 CNN convolution on points is applied to start the initial feature extraction. Repeated 3D CNN and max-pooling over the Spherical bins propagate contextual information until all the information is condensed in the center bin. Experiments show that the proposed method outperforms its counterparts by a large margin.

Acknowledgements

This research was supported by the Major Project for New Generation of AI under Grant No.2018AAA0100400, the National Key Research and Development Program under Grant No.2016YFB0501100, the National Natural Science Foundation of China under Grants 61976208, 61773377, and the Young Elite Scientists Sponsorship Program by CAST (2018QNRC001).

Proposed Method



Our SHPool contains three key modules: Shape Context Module (SCM, for short), Geometric Blur Module(GBM, for short), and Hierarchical Pool Module(HPM, for short).

- We follow 3D Shape Context (3DSC) to divide every spherical neighborhood $\{P_i \subset R^3\}$ into $K(K = n_r \times n_\phi \times n_\theta$, n_r is the partition num arcoss r and n_ϕ, n_θ can be analogied) sector bins $\{b_{i1}, b_{i2}, ..., b_{ik}\}$ but without log-polar.
- Geometric blur applies gaussian average pooling and blurring changes with the radial distance from the center.

$$F_{b_{ik}} = MLP^1(\mathcal{A}(g(MLP^2(D_{x_{ij}})), g(F_{x_{ij}}))), \forall x_{ij} \in \mathcal{N}(b_{ik})$$

- Aiming at integrating information of bins $K \times C_{bin}$, we propose the Hierarchy Pool Module, composed of iterative convolution and spherical max-pool.

Experiments & Visualization Results

Table 1. 3D shape classification and 3D semantic segmentation on ModelNet40, ScanNet and S3DIS Area5 benchmarks.

Method	ModelNet				ScanNet v1	S3DIS Area5
	Input	Point	mAcc(%)	Acc(%)	accuracy(%)	mIoU(%)
VoxNet	voxel	-	83.0	85.9	-	-
MVCNN	image	-	-	90.1	-	-
PointNet	xyz	1k	86.2	89.2	73.9	-
ShapeContextNet	xyz	1k	87.6	90.0	-	-
Kd-Net(depth=10)	xyz	1k	-	90.6	-	-
PointNet++	xyz	1k	-	90.7	84.5	-
Ψ-CNN	xyz	1k	88.7	92.0	-	-
PointCNN	xyz	1k	88.1	92.2	85.1	57.26
DGCNN	xyz	1k	90.2	92.2	-	58.27
PointWeb	xyz	1k	89.4	92.3	85.9	60.28
A-CNN	xyz	1k	90.3	92.6	85.4	-
RS-CNN	xyz	1k	-	92.9	-	-
ShellNet	xyz	1k	-	93.1	85.2	-
PointSpherical	xyz	1k	**90.7**	**93.2**	**86.54**	**62.68**

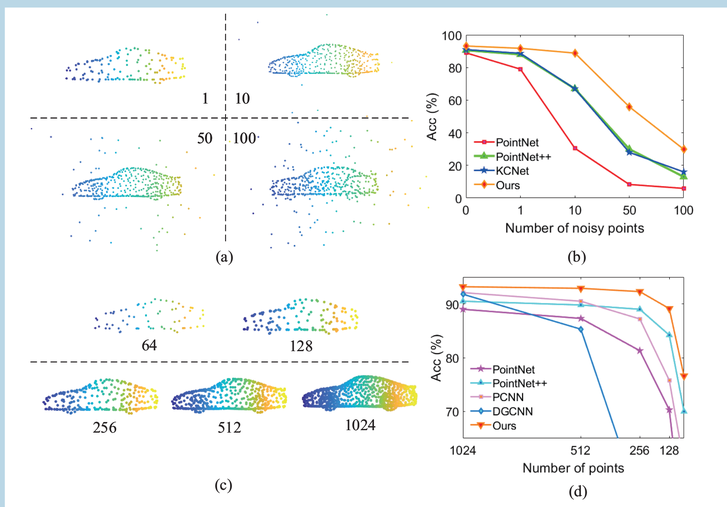


Figure 1. Robustness of Noise and Sampling Density. The upper is the visualization of noise and the upper for sample density.

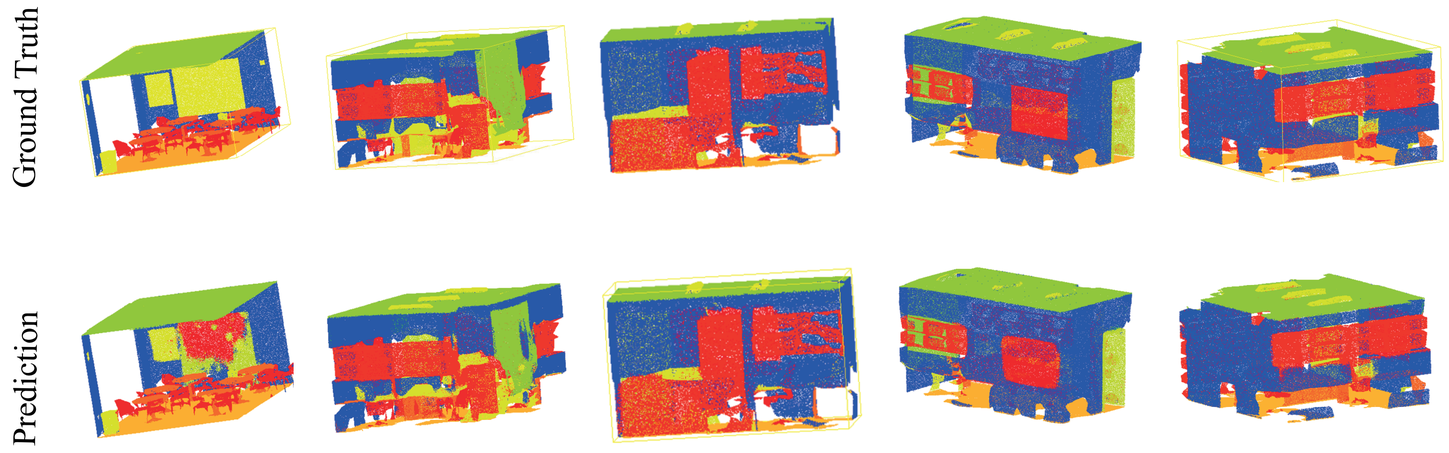


Figure 2. Semantic segmentation for indoor scenes in the S3DIS dataset.

References

[1] Belongie, Serge and Malik, Jitendra and Puzicha, Jan: *Shape matching and object recognition using shape contexts*, IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 4, pp. 509522, 2002.
[2] Berg, Alexander C and Malik, Jitendra: *Geometric blur for template matching*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2001.
[3] Körtgen, Marcel and Park, Gil-Joo and Novotni, Marcin and Klein, Reinhard: *3d shape matching with 3d shape contexts*, in The 7th central European seminar on computer graphics. Budmerice, 2003, pp. 517.