CS-GY 9223G Project Proposal: 3D Point Cloud Classification using Deep Learning

Jiaying Li
New York University
Tandon School of Engineering
jl10919@nyu.edu

Zili Xie
New York University
Tandon School of Engineering
zx979@nyu.edu

1. PROBLEM STATEMENT

Point clouds are important datasets that represent objects or space, which are of great significance in 3D modeling, mapping and reconstruction. Each point represents the X, Y, and Z geometric coordinates of a single point on an underlying sampled surface and sometimes plus extra feature channels such as color. The classification of 3D objects has always been a hot research topic in 3D computer vision. Given hundreds to thousands of three-dimensional coordinate points, how to automatically find the category of such a geometric body among many categories is a complicated problem. This problem is mainly different from ordinary image classification problems in the following aspects.

- First, our input has changed from a four-dimensional image matrix of size $n*H*W*c$ to a three-dimensional vector set of length n, where n is the number of input points. Also, the relationship between 3d points are much subtler than the relationship between 2d pixels.
- For geometric objects, their corresponding categories should remain unchanged under certain geometric transformations, such as rotation, translation, scaling and folding in space. Therefore, we cannot naively judge the category of objects merely through some single perspective.
- Finally, the order of the points of the geometry does not affect the category of the geometry, which means that a 3D object composed of n points, its points can be arranged in at most $n!$ ways, and given any of these $n!$ permutations our classifier should give the same classification result.

2. PRELIMINARY LITERATURE SURVEY

Many classification methods for three-dimensional objects have been proposed by researchers in recent years. In fact, because a geometric body can have many different manifestations, the classification method will be very different according to the different forms of representations. An intuitive method is multi-view input learning. Multiple 2D rendered images or projections on a two-dimensional plane can be obtained from the 3D model. Each 2D image is trained through its own CNN network and then aggregated for pooling. Then set CNN for feature extraction, and finally output the classification of items.[3]

In other studies, 3D objects are discretized into uniform-sized voxel grids, and CNN is directly applied to the three-dimensional objects.[4, 1] Two-dimensional CNN and three-dimensional CNN are used together to learn the properties of geometry, and the outputs of multiple CNNs are fused before generating prediction results.

Finally, it is the learning network pointNet that directly uses the point cloud as input without special processing on the input geometry.[2]

3. GOAL

The expected output of this project would be a well-trained deep neural network model that has high performance on 3D object classification. The model will be verified both on indoor and outdoor scenes.

References

[1] M. Nießner A. Dai M. Yan C. R. Qi, H. Su and L. Guibas. Volumetric and multi-view cnns for object classification on 3d data., 2016. In Proc. Computer Vision and Pattern Recognition(CVPR), IEEE. 1

[2] Charles R. Qi* Hao Su* Kaichun Mo Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation., 2017. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1

[3] H. Su M. Aono B. Chen D. Cohen-Or W. Deng H. Su S. Bai X. Bai et al. M. Savva, F. Yu. Shrec'16 track large-scale 3d shape retrieval from shapenet core55, 2016. 1

[4] D. Maturana and S. Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition, 2015. In IEEE/RSJ International Conference on Intelligent Robots and Systems. 1