Reaction Report

3D point cloud segmentation and classification suffer performance issues due to nature of data. Traditional CNN architecture uses rigid grid structures and voxelization for input but that results in loss of information and introduction of quantization artifacts. Methods used before either construct the feature by hand which is very trivial for every specific task, due to sparsity existing deep learning model could not process large scale data, 2D images conversion of 3D point cloud has better performance but its challenging to extend it for scene semantic, point classification or shape classification. These challenges have been addressed by Charles R. Qi, Hao Su, Kaichun, Mo Leonidas and J. Guibas in their paper PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation which proposed a methodology for using simple points instead of 3D voxel grid of collection of 2D images obtained from 3D point cloud for segmentation and classification.

Methodology proposed was very simple and targeted three major areas classification, part segmentation and semantic segmentation. One of the key feature of the system was to process each point individually and similarly and represented by x, y and z along some computed local and global features. It inputs the point cloud and transform those point cloud using T-Net which makes it flexible towards the alignment invariance for classification and whole network outputs the scores for k object classes. Also learnt global features are input as point features for local descriptors which helps in classifying the scene and part segmentation using local and global features of input 3D point cloud. Interaction among points can be captured through neighboring points to understand the scene semantic. Three major component of the proposed system are i) max pooling layer as symmetric function which aggregates information from all points ii) structure for local and global features combination and iii) two joint networks which map and align the input points and point features. Main achievement was to build a network that learnt to summarize a shape by a sparse set of key points. Results of experiment shows that proposed methodology has outperformed due to its invariance characteristics of outliers and transformations also it has processed points separately. It has also have better performance on space and time complexity index as compare to existing state of art systems discussed in the paper.

In my opinion this paper has opened the door of deep learning in 3D point cloud segmentation and classification. It has proposed the independent and identical processing of points which helped in processing data without loss or additive Gaussian noise due to processing. One of the major limitation was lack of contextual information usage in the methodology which can help the classification and increase performance. Idea of combination of local and global feature does tries to accommodate semantic of scene but contextual information fed or aggregated as input could have increase the performance effectively.