DeepPool: 3D Part Segmentation of Point Clouds

Gautam Kumar*‡, Preetham P *‡, Siddharth Srivastava†, Swati Bhugra†, Brejesh Lall†
‡Department of Mathematics and Computing, Indian Institute of Technology, Delhi, India
†Department of Electrical Engineering, Indian Institute of Technology, Delhi, India
{mt6140555, mt6140560, eez127506, eez138301, brejesh}@iitd.ac.in

Abstract

We utilize the recently proposed PointNet architecture for directly processing the 3D point clouds. The PointNet architecture is augmented with pyramid pooling leveraging skip connections to encode local information. The network is trained end-to-end on the provided training data to achieve the final part segmentation results.

1. Method Details

The overall architecture of the proposed method is shown in Figure 1. The initial processing of the point cloud is performed directly on the 3D point cloud (without any voxelization etc.) using the recently proposed PointNet [1] architecture. Sunequently, PointNet is augmented with a Pyramid Pooling technique motivated from PSPNet [2] which has been recently found to provide excellent results on pixel level labelling in 2D images.

In the PointNet architecture, we use the features from intermediate layers and augment them with a deep spatial pooling module and train the resulting network end-to-end. Additionally, we perform several modifications to the Spatial Pooling Architecture of PSPNet to suit 3D data. The methodology along with specific contributions is described below:

- Like its 2D counterpart, we form four levels of pyramid pooling. However, instead of directly convolving them and upsampling them using bilinear interpolation, we form a network of convolution and deconvolution (Conv-Deconv in Figure 1) to get upsampled feature maps. Also, we add another layer of mlp in PointNet architecture prior to pooling the features. The intuition being that in later stage we concat these features to exploit characteristic of local regions to strengthen the feature at various levels of spatial pyramid pooling.

- The output of encoder for each level is concatenated to form a global descriptor (1xK1). Subsequently, we form point features by concatenating this global feature by forming skip connections from PointNet layers.

- Now instead of performing convolution with the obtained feature map, we stack convolution and dropout layers, whose output is concatenated with the final layer of PointNet. This is finally convolved to obtain the final probabilistic vector of point labels.

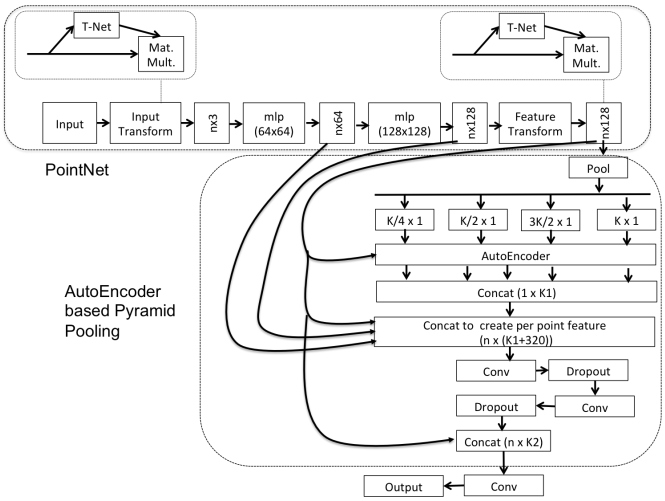

Figure 1. Overall Architecture

References

[1] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1

[2] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1

*Equal Contribution