Reaction Report

Traditional approach of deep learning and neural network usually operates on structured data while most of the data collected via 3D sensors is irregular and unstructured. To learn from such data is totally different task for classification, segmentation and semantic segmentation. There were two major categories of solutions proposed over the years. One focuses on designing models which process 3D point cloud data as irregular as it is and design a model which can learn from unstructured data while other solution focused on preprocessing the data and convert it into some structured format and design a learning model to wrap it accordingly. First category of solution has been proposed by various researchers such as PointNet, PointNet++, SuperPoint Graph and Deep Parametric Continuous CNNs. These methodologies have processed irregular set of 3D point clouds and tried to learn from these points using different neighborhood consideration strategies. PointNet++ used hierarchical approach to add local context on multiscale and multi resolution level, Superpoint graphs formed superpoints from 3D point clouds and used edge featured to include the context for learning while Deep Parametric continuous CNNs used KD Tree for local context and eventually used PointNet like structure to aggregate global features to local features of each 3D point cloud. All these have limitation such as in PointNet and PointNet++ flexibility of spatial connectivity over points is not facilitated. Pooling from hierarchical approaches results in surface information loss. Superpoints utilized the edge features which are calculated and mostly context included in Superpoints and Deep parametric continuous CNNs is geometric instead of semantic. Superpoint used geometric shapes which are not consistent and have various spectral bases which is different for every shape and leads to inefficient generalization. Other category of solutions tried to convert the data in 2D images or 3D voxel grids which has different limitations including artifacts and loss of significant information. One such approach which tried to address the limitation of different systems and proposed a new data structure to map 3D point clouds.

Methodology proposed in SPLATNet: Sparse Lattice Networks for Point Cloud Processing by Hang Su and team has discussed the usage of lattice based network for point cloud classification and segmentation. They proposed the usage of bilateral convolution layers (BCL) to map the irregular 3D point clouds on lattice structure with flexible specification which goes under convolution. It is very simple and fairly smooth framework which translates the 3D point cloud to spare lattice structure, perform convolution and interpolate them back to original input data. One of the key features of BCL is to convolve only where data is present by using hash table which computationally efficient and helps dealing with sparsity of 3D point cloud. It also helps in calculating features which consider the locality and hierarchical characteristics of 3D point cloud. It also fuse 2D images with 3D points cloud for feature mapping on same lattice structure and provide capability of end to end learning from both source to help the segmentation and classification. It also let 2D images to be mapped onto 3D space which is very helpful once 2D images are convoluted and output 2D features can be mapped into 3D space using first two operation of BCL splat and slice. These signals from 2D images and 3D point clouds are later convoluted and used to produce point wise segmentation and 2D image which were transformed can be converted back to 2D segmented images which have taken the advantage of 3D space and fusion of 3D points. It is a flexible framework which doesn’t regulate the input and output space features for lattice space and point features. It allows the longer connectivity range in deeper layers of BCL to increase the receptive field in between 3D point cloud. Softmax layers produce labels for every point and aggregates the information from every layer of the lattice scale neighborhood to include the context. Lattice scale has major effect on the receptive field and it helps in manipulating the efficiency of network. Strategy used to increase receptive field is computationally efficient and better than alternative is to make lattice coarser in deeper layers. This framework calculates the spatial and hierarchically aware features which helps in processing irregular 3D point cloud and outperformed most the state of art methodology on majorly available dataset. Challenges faced during fusion was mislabeling of certain elements for part segmentation in available dataset that complicated the classification and segmentation and confused the network. Aggregating output from all lattice scales and network layers can give the liberation of adding context and local and hierarchical information for classification and point-wise segmentation.

Proposed methodology has some limitation and room for improvements as well. One of the major issues that has been dealt in 3D point cloud processing is sparsity through hashtable which is definitely extra computational cost to refer and maintain for dense areas. Since these 3D point clouds are randomly distributes over lattice it is addressed using convolution over each input point adding computational cost and complexity to the system. Since deeper layer of BCL has long range of connectivity that does increase the receptive field but also increase the computational cost and takes longer time to train such network. Idea situation would be every point connected to all others what will be tradeoff for range of connectivity and computational efficiency.