Please fill in the following information after you read the paper.
Name: Darshit Miteshkumar Desai; UID: 118551722; Dir ID: darshit

[Paper tile] Pointnet: Deep learning on point sets for 3d classification and segmentation
[Summary] Describe the key ideas, experiments, and their significance.
Pointnet is one of the foundational research papers in the domain of point cloud processing which kicked off the field. In the paper the authors describe their motivation behind pursuing this area of research at the time, which boils down to the need of converting point clouds to quantized voxel grid representations to apply architectures like 3D CNN for further processing. This has its own disadvantages one of which is the voluminous representation of data and the another one is the reduction in detail and change in variances which are cause by using quantized voxels. For a generic point cloud processor, it has to be invariant to permutations of the Nx3 point cloud list and also to the rigid body motions of the point cloud. Towards this end the authors explore various related works to justify that why their architecture is state of the art. One area they explore is the point cloud features which at the time were categorized into intrinsic and extrinsic features. They also discuss various Deep learning architectures which work on 3D data and Unordered sets of data. The network architecture is fairly simple, the property of point sets are defined by the authors which form the requirements of the model. One of which was the unordered property of point clouds as they can be shuffled in the data structure and still represent the same object i.e., the network should be capable of identifying the object even after these variances in order. Another one is interaction among points, in 3D space certain groups of points may represent a meaningful sub part of that object. The network should be able to capture that. The third is invariance to rigid body transformations of the point cloud as a whole. To account for all the properties mentioned above, the authors mention individual parts of the network architecture which takes care of it, for the unordered input requirement the authors explore various solutions like presorting the input array but that is tricky since in a 3 dimensional space/ higher dimensional space it is difficult to do it, they also explore RNNs for achieving the objective but RNNs have been disproven to be invariant of orders in the work “OrderMatters” where they have shown order cannot be neglected for such architectures. They finally decide to use a symmetric function, which can be empirically denoted by a multi-layer perceptron. For the rigid body transformation invariance, the authors achieve this by using T-net whose specific purpose is to predict point cloud affine transformation matrix, the same is done for the feature layers too. For the last requirement the authors come up with a combo of using global features which are calculated by max-pooling and a set of local features from the intermediate MLP layers and stack them back to back to capture local and global context for the segmentation network. They also provide theoretical analysis of using these methods as a validation for their empirical approach. The experiments mainly focus on 3 applications, object recognition,

scene segmentation, object part segmentation. For classification they use ModelNet40 and achieve state of the art performance on overall accuracy metric. For part segmentation they use Shapenet part dataset and they achieve state of the art mean IOU and achieve reasonable segmentation results. For scene segmentation the network is tested on Stanford 3d dataset and 3d scans from matterport.

In conclusion the network is the first work in the area of point cloud processing which consumes raw point clouds and is able to do various computer vision tasks like recognition and segmentation.

[Strengths] Consider the aspects of key ideas, experimental or theoretical validation.

The strengths of the paper are mentioned below:

1. First deep learning architecture to process unordered point sets and address rigid body variances.
2. The same network can be used for feature extraction of point clouds which can be further used for other complicated downstream tasks
3. The authors in the main paper and supplementary paper provide sufficient architecture analysis and theoretical analysis to prove the stability and efficiency of their network.
4. The network can also be used for other applications like Model retrieval and shape correspondences between two different point clouds

[Weaknesses] Consider the aspects of key ideas, experimental or theoretical validation, writing quality, and data contribution (if relevant). Explain clearly why these are weak aspects of the paper

Being one of the seminal papers there were not many weaknesses mentioned in the paper but here are some of them that I found while reading other follow up works-

1. From this paper it can be fairly assumed that there might be a lot of areas of improvement and future work required, the authors failed to mention that in their paper which can be attributed to the writing quality issue.
2. In pointnet++ the authors highlight a major weakness of pointnet which is the inability of the network to capture local structures as pointnet only learns specific point wise features which are then aggregated later.
3. Another example of such a work which accounts for local relations between points is the point transformer which inherently encodes such features due to the property of self attention module and achieves better results.

[Reflection] Share your thoughts about the paper. What did you learn? How can you further improve the work?

From the paper it can be said that efficiently breaking down the problem and addressing each of it with simple solutions can be better then coming up with a complicated architecture, that is exactly what PointNet authors do, they use first principles to design a network which handles each of the problems and properties of the pointnet representation.

Other following works already identify areas of improvement on pointnet one such was capturing and encoding local relations in 3d space using pointnet++ or point transformers.