# Models for 3D semantic segmentation

Lokesh Veeramacheneni

September 20, 2021

## 1 Summary of models

| Model | Description |
|---|---|
| PointNet [12] | • Process point cloud directly, no projection involved.<br>• MultiLayerPercepton (MLP) is used to extract features which are permutation invariant<br>• Then max pool is used to extract global features.<br>• A transformation network generates transformation matrix to spatially align input PC and features to achieve transformation invariance.<br>• Cannot account the relation between points and local neighbourhood information, so in large scenes critical information is lost |
| PointNet++ [12] | • Extension to PointNet by adding sampling layer and grouping layer<br>• First selects some points from input points as centroids of their local areas using Farthest Point Sampling (FPS) algorithm.<br>• The applies local region grouping module to construct local regions.<br>• Recursively PointNet is applied to extract local features<br>• Independently process the points so the relation such as distance and direction between the points are not taken into consideration |

| | |
|---|---|
| SPGraph [6] | • 3D point cloud is represented as multiple superpoint graphs (SPG) based on linearity, planarity and scattering.<br>• Each SPG represent a simple primitive shape and these SPGs are downsampled and embedded into fixed size vector using PointNet<br>• Then graph convolutions are employed to label each superpoint based in neighbours<br>• High computation cost and moreover same performance as PointNet++ over SemanticKITTI dataset |
| SPLATNet [9] | • Direct application over points and no preprocessing such as voxel conversion or image projection is necessary<br>• Projects input points onto high dimensional lattice and then employ Bilateral Convolution Layers (BCL) for feature aggregation<br>• The BCL operations (splat, conv and slice) are not learned and moreover the repeated application of BCL degrade point cloud features as BCL act as gaussain filters as propsed in [8] |
| SqueeseSeg V1 [10], and V2 [11] | • All the three variations of the Squeeseseg are of projection based models<br>• Projects the point cloud onto the sphere of dense, grid based representation (3D cartesian coordinates to 2D spherical coordinates<br>• Use an 2D encoder-decoder like SqueeseNet to perform semantic segmentation and the apply Conditional Random Field (CRF) to convert 2D semantic maps to 3D point clouds<br>• SqueeseSeg V2 improves on V1 by applying Context Aggregation Module (CAM) in between and also employs batch normalization and focal loss to improve the training accuracys |

| | |
|---|---|
| SalsaNet [1] | • This is also a projection based method, projecting as Bird Eye View (BEV)<br>• A new SalsaNet is proposed to do the 2D semantic segmentation of the BEV projected point cloud and also a process of auto data labelling is proposed to annotate the unannotated point cloud data<br>• This method also tests and compare the performance among BEV and Spherical Front View (SFV) |
| RangeNet [7] | • RangeNet is a novel method which uses spherical projection for 3D to 2D projections<br>• The architectures for the 2D semantic segmentation used are DarkNet-21, DarkNet-53 leading to various models such as RangeNet21, RangeNet53 and RangeNet53++ where ++ means post processing enabled<br>• The labelled 3D point clouds will suffer from loss of annotations from shadows, to cover this post processing method such as gpu enabled KNN is employed to annotate the shadowed points<br>• RangeNet suffers from higher number of parameters leading to higher trianing times |
| LatticeNet [8] | • LatticeNet is a novel method to apply raw 3D point cloud as input<br>• This network uses hybrid approach by employing PointNet for low level features and 3D convolutions for extracting global context and architecture resembles 3d U-Net architecture<br>• 3D data is projected onto a permutohedral lattice by using a projection from hyperplane<br>• LatticeNet suffers from memory issues because it has to store the lattices for computations and copy from CPU to GPU |

| | |
|---|---|
| RandLA-Net [4] | • It is efficient and light weight architecture which feeds on 3D raw point clouds<br>• The method is similar to PointNet but the sampling algorithm is random sampling instead of farthest point sampling method<br>• Random sampling might not yield good feature points, so local feature aggregation module to acheive the more receptive field<br>• The key modules in RandLA-Net are local feature aggregation module, attentive pooling and dilated residual blocks, the network resembles the encoder-decoder architecture<br>• This network is one of the few 3D semantic segmentation models with low number of parameters gaining higher performance. |
| SalsaNext [3] | • Its an upgrade version of the SalsaNet, with an addition of new context module and dilated convlution stack replacing the encoder and decoder module comes with a addition of pixel shuffle layer<br>• A combination of weighted cross entropy and lovasz softmax loss is used for optimization<br>• This is byfar one of the two studies involving the uncertainty estimation in 3D models and also acheive highest performance with less number of parameters |
| 3D MiniNet [2] | • This network combines information from 3D point cloud and 2D projection data.<br>• The network structure is divided into two modules, Projection learning module and MiniNet backbone<br>• A post processing module gpu enabled KNN same as in RangeNet is applied to label the shadowed points in output |
| KPRNet [5] | • This is also a projection based method similar to RangeNet<br>• To obtain accuracte 3D annotations, novel KPConv is proposed and utilized as post processing module<br>• It utilizes ResNext-101 module as encoder and decoder similar to Panoptic Deeplab network followed by KPConv, and classifier |

| | |
|---|---|
| PolarNet [13] | • This network takes polar bird eye view projection with learnt polar grid as an input<br>• IT utilizes polar BEV encoder network for grid learning and then simplified PointNet with only MLP's are applied on this BEV grid to learn the semantic labels<br>• The downside to this network is that the BEV grid generation is quite complex |

Table 1: Short summary of the 3D semantic segmentation models evaluated on SemanticKITTI dataset. Models highlighted in Cyan are the models that are selected to apply for OOD detection

# 2 Model Vs Parameters

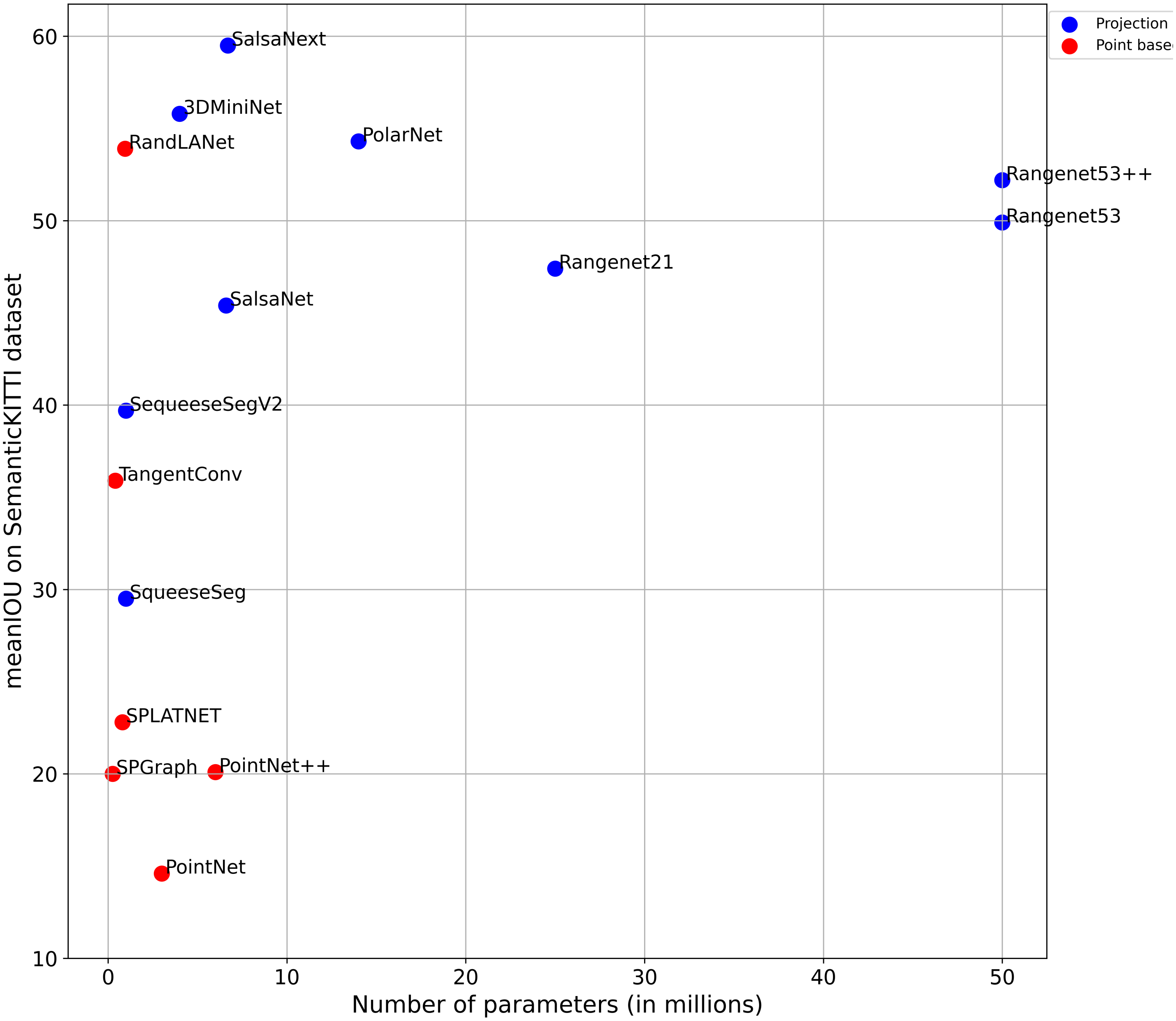

Figure 1: Model performance vs number of parameters on semantic KITTI dataset

# References

[1] Eren Erdal Aksoy, Saimir Baci, and Selcuk Cavdar. Salsanet: Fast road and vehicle segmentation in lidar point clouds for autonomous driving. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 926–932, 2020.

[2] Iñigo Alonso, Luis Riazuelo, Luis Montesano, and Ana C. Murillo. 3d-mininet: Learning a 2d representation from point clouds for fast and efficient 3d lidar semantic segmentation. *IEEE Robotics and Automation Letters*, 5(4):5432–5439, 2020.

[3] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In George Bebis, Zhaozheng Yin, Edward Kim, Jan Bender, Kartic Subr, Bum Chul Kwon, Jian Zhao, Denis Kalkofen, and George Baciu, editors, *Advances in Visual Computing*, pages 207–222, Cham, 2020. Springer International Publishing.

[4] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[5] Deyvid Kochanov, Fatemeh Karimi Nejadasl, and Olaf Booij. Kprnet: Improving projection-based lidar semantic segmentation. *arXiv preprint arXiv:2007.12668*, 2020.

[6] Chakri Lowphansirikul, Kvoung-Sook Kim, Poliyapram Vinayaraj, and Suppawong Tuarob. 3d semantic segmentation of large-scale point-clouds in urban areas using deep learning. In *2019 11th International Conference on Knowledge and Smart Technology (KST)*, pages 238–243, 2019.

[7] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet ++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4213–4220, 2019.

[8] Radu Alexandru Rosu, Peer Schütt, Jan Quenzel, and Sven Behnke. Latticenet: Fast point cloud segmentation using permutohedral lattices. *arXiv preprint arXiv:1912.05905*, 2019.

[9] Hang Su, Varun Jampani, Deqing Sun, Subhransu Maji, Evangelos Kalogerakis, Ming-Hsuan Yang, and Jan Kautz. Splatnet: Sparse lattice networks for point cloud processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[10] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1887–1893, 2018.

[11] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4376–4382, 2019.

[12] Jiaying Zhang, Xiaoli Zhao, Zheng Chen, and Zhejun Lu. A review of deep learning-based semantic segmentation for point cloud. *IEEE Access*, 7:179118–179133, 2019.

[13] Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zerong Xi, Boqing Gong, and Hassan Foroosh. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.