International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7645
***Article Received:** 04 June 2023 **Revised:** 29 July 2023 **Accepted:** 14 August 2023*

Detection of 3D Object in Point Cloud: Cloud Semantic Segmentation in Lane Marking

Hemlata Arya[1], Parul Saxena[2], Jaimala Jha[3]
[1]Assistant Professor, Department of Computer Science and Engineering
Madhav Institute of Technology and Science,
Gwalior, India
hemlataarya21@mitsgwalior.in
[2]Assistant Professor, Department of Computer Science and Engineering
Madhav Institute of Technology and Science,
Gwalior, India
gaurparul2007@mitsgwalior.in
[3]Assistant Professor, Department of Computer Science and Engineering
Madhav Institute of Technology and Science,
Gwalior, India
jaimala.jha@mitsgwalior.in

Abstract— Managing a city efficiently and effectively is more important than ever as growing population and economic strain put a strain on infrastructure like transportation and public services like keeping urban green areas clean and maintained. For effective administration, knowledge of the urban setting is essential. Both portable and stationary laser scanners generate 3D point clouds that accurately depict the environment. These data points may be used to infer the state of the roads, buildings, trees, and other important elements involved in this decision-making process. Perhaps they would support "smart" or "smarter" cities in general. Unfortunately, the point clouds do not immediately supply this sort of data. It must be eliminated. This extraction is done either by human specialists or by sophisticated computer programmes that can identify objects. Because the point clouds might represent such large locations, relying on specialists to identify the things may be an unproductive use of time (streets or even whole cities). Automatic or nearly automatic discovery and recognition of essential objects is now possible with the help of object identification software. In this research, In this paper, we describe a unique approach to semantic segmentation of point clouds, based on the usage of contextual point representations to take use of both local and global features within the point cloud. We improve the accuracy of the point's representation by performing a single innovative gated fusion on the point and its neighbours, which incorporates the knowledge from both sets of data and enhances the representation of the point. Following this, we offer a new graph point net module that further develops the improved representation by composing and updating each point's representation inside the local point cloud structure using the graph attention block in real time. Finally, we make advantage of the global structure of the point cloud by using spatial- and channel-wise attention techniques to construct the ensuing semantic label for each point.

Keywords- point clouds, 3D data, Voxel-based Approach, Deep learning, Inference time.

I. INTRODUCTION

Researchers are starting to pay greater attention to the point clouds generated by 3D scanners, in particular for challenges requiring point cloud interpretation, such as 3D item classification [13, 14], 3D object identification [21, 27], and Segmenting 3D data semantically [25, 13, 14, 23, 10]. The task of giving labels of the same class to each point in a 3D environment is known as 3D semantic segmentation and is both difficult and common. The first difficulty is that 3D scanners only collect a small number of points at a time, making it hard to train a single, robust deep model for semantic segmentation. Second, the arguments aren't always presented in the proper sequence and are often disorganised. It's very hard to put into words and show the link between the places.

As developments in 3D acquisition technology continue, devices like 3D scanners, LiDARs, and RGB-D cameras (such the Kinect, RealSense, and Apple depth cameras) are becoming more widely accessible and cheap [1]. The 3D data gathered by these sensors may provide insight into the object's geometry, shape, and size [2, 3]. When 3D data is combined with 2D photos, robots may get a greater awareness of their immediate surroundings. Autonomous cars, robotics, remote sensing, and medicine are just a few of the many areas that employ 3D data [4]. There are several methods in which three-dimensional data may be represented. depth pictures, point clouds, meshes, and volumetric grids are just a few of the most prevalent types. Point clouds, a popular file format, save the original, non-discretized 3D geometric data. This representation is used by many applications, such as robots and autonomous vehicles, that need to understand their surroundings. Current cutting-edge research in fields as diverse as computer vision, voice recognition, and natural language processing often relies on deep learning methods. Due to their high dimensionality, which is compounded by their intrinsic lack of organisation and the small amount of accessible datasets, deep learning on 3D point clouds still has a long way to go [5]. As a result, the major goal of this research is to investigate deep learning techniques currently being used to the processing of 3D point clouds.

Fig 1. 3D representations of Point Cloud

To accomplish the necessary semantic segmentation, several current approaches first convert point clouds into conventional 3D voxel grids or collections of pictures [25, 5, 22]. Such a transformation method might make advantage of the structural data conveyed by the interspot links. When dealing with 3D volumetric data, in particular, the memory and processing power needed by such systems is enormous. For effective and efficient point management, many modern deep learning architectures on point clouds have been suggested, such as PointNet [13] and PointNet++ [14]. PointNet accurately learns a spatial encoding for each point before combining their attributes into a global representation.

HD Map processing and sensor-based autonomous driving rely heavily on target object recognition from point cloud data.

Common types of things on the road include the road itself, lane markers, pavement areas, support structures (such as railing and curb), signs, light poles, and other unrelated elements like trees, people, and buildings.

This research aimed to automate the process of lane detection using point cloud information, which is crucial for the decision-making processes of autonomous cars.

II. LITERATURE REVIEW

Recently, deep models have shown their ability to learn features on computer vision tasks utilising traditional data structures. However, there are still many obstacles to overcome because of the constraints of the data representation technique when working with a 3D point cloud, which contains asymmetrical data structures. Based on their respective 3D data representation methodologies, the currently available strategies may be roughly classified as either 3D voxel-based [5, 25, 22, 7, 9], multiview-based [18, 12], or set-based [13, 14] methods.

Fig 2. Point cloud 3D appearance

3D Voxel-based Approach: Point clouds may be converted into standard 3D voxel grids utilising 3D voxel-based techniques, enabling 3D CNN to be used directly in the same way as an image or video. When dealing with spatial data, Wu et al. [25] suggest using a 3D ShapeNets network based on complete voxels. Some information is lost during the discretization process because of the bounds of any representation. While a higher resolution voxel requires more memory and computing power, higher resolution voxels are not always possible. There has been a recent suggestion to stop processing voxels that are completely empty in order to save money on computer power. There were contributions from Oct-Net [16], Kd-Net [7], and O-CNN [22].

Multiview-based Approach:The target point cloud must be photographed from a variety of angles, and the multiview-based methods must provide several images for each viewpoint. In the next step, regular 2D CNN processes [18] may be applied to each individual image. Recent work has employed the multiview image CNN [12] to successfully segment 3D forms. The multiview-based solutions aid in minimising the operational memory and computational expenses. However, there is some data loss throughout the process of creating pictures from the 3D point cloud. However, there is still no clear solution for the difficult open issue of deciding how many views to collect and how to arrange them to adequately display the 3D structure.

Set-based Approach: To directly learn the representation, Point Net [13] is the first set-based technique. fromunorganized point clouds. Using a hierarchical learning method, Point Net++ [14] expands Point Net to gather data about nearby structures. When trying to extract local context data, PointCNN [10] suggests using the canonical arrangement of points.

Deep learning on regular domain: To learn the representation from disorganised point clouds, PointNet [13] is the first set-based technique. PointNet++ [14] builds on PointNet using a hierarchical learning approach to collect information about adjacent buildings. The canonical arrangement of points is recommended by PointCNN [10] when attempting to extract local context data.

DOI: https://doi.org/10.17762/ijritcc.v11i10s.7645
***Article Received:** 04 June 2023 **Revised:** 29 July 2023 **Accepted:** 14 August 2023*

Deep learning on irregular domains:A number of state-of-the-art systems that directly analyse point clouds using complex networks to extract attributes have been inspired by the pioneering work on PointNet [12, 13]. Techniques like this may be broken down into four broad classes: nearby feature pooling [13, 15, 18, 21, 22], graph message passing [16, 23, 24], kernel-based convolution [14, 25-28], and transformer-based learning [29, 30]. Straightforward point clouds may be employed with these techniques. Point clouds are limited in their capacity for inductive learning because of their incoherence over space.

III. METHODOLOGY

Semantic segmentation of 3D point clouds works to identify each point with a single semantic class.

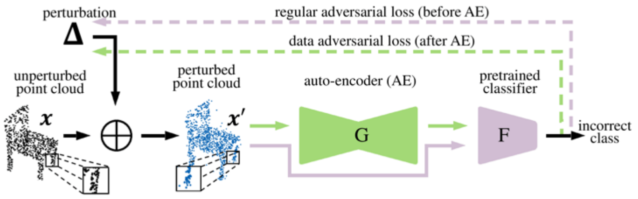

Fig 3. Semantic cloud point cloud segmentation model proposed

Point Enrichment. Accurate class predictions require considering not just the information of the target point but also the information of any surrounding or contextual points inside the complicated point cloud structure.

Feature Representation. We use the typical encoder-decoder architecture with lateral connections to learn the feature representation for each point using the enhanced point representation.

Prediction. We leverage channel-wise and spatial-wise attentions based on the acquired semantic representations to make better use of the point cloud's overall structure. This makes it possible to infer the semantic label for each node in the graph.

As a whole, the lane's characteristics are summed up as follows for the purpose of detection:

1. If you look at the point cloud data for a road, you'll see that the lane is contained inside the road surface, which is a thin, flat region.
2. Some areas along the road, called lane points, are more reflective than others.
3. A lane's profile is segmented in a straight line.
4. Each lane runs perpendicular to the others.

Based on these observations, we developed the following methods for detecting lanes in point cloud data:

1. Find the road surface once all the bumps have been taken off.
2. points of possible lane intensity over which they are deemed too weak to be used.
3. Try to guess the lane's orientation.
4. Using cluster analysis, calculate the line equations for each lane.

Preprocessing

Starting with the following, we preprocess the point cloud data:

1. Change points’ coordinate from (latitude, longitude) to (x, y)
2. Down sample
3. Filter noise

The raw point cloud data for each row is stored as follows:

1. latitude (degree)
2. longitude (degree)
3. altitude (meter)
4. point reflecion intensity (0~100)

For simpler presentation and processing, we converted the points' (latitude, longitude) values to a local Cartesian coordinate (x, y) based on the average position of all cloud points.

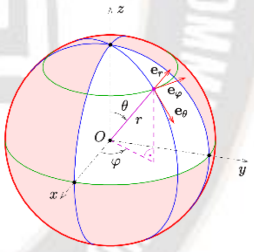

Fig 4. Coordinates on a Sphere Angle of Approach Location in Polar Space

Global coord is {O}

Local coord is {e_r}

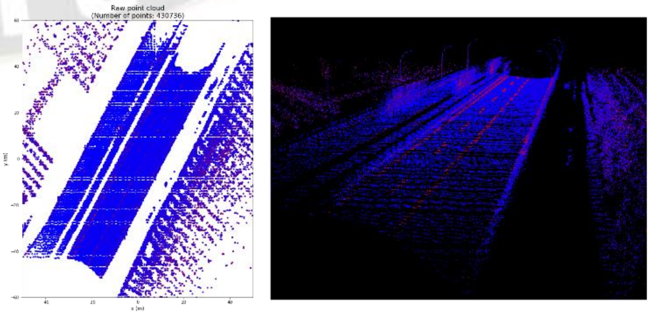

Fig 5. Left: Point cloud projected on x-y plane. Right: 3D point cloud. Points with higher intensity (i.e. more reflective) are more red.

After resampling the coordinates to capture the highest point intensities, we obtained a point cloud with a lower resolution.

Steps:

For each voxel of size 0.1m in the world:

> Find all points {Pi} inside the voxel. Replace them with a single point Q.
>
> Q is placed at the center of the voxel.
>
> Intensity of Q is set as the max of intensities of {Pi}.

Reasons of down sampling:

1. Trim the amount of points to speed up the calculation.
2. Densifying the point cloud will make it easier to determine algorithm parameters.
3. It would be easier to distinguish lanes if the lane markers were larger.

The data points representing the road surface are clustered together, whereas the noise data points are spread out. So, we employed a method called "radius outlier removal" to get rid of the noisy areas.

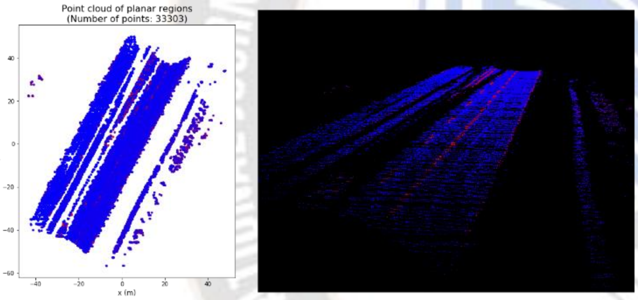

Fig 6. Planar regions Point Cloud

The previous limit of 430k cloud points has been lowered to 84k. The noisy spots have also been removed.

IV. RESULT ANALYSIS

Threshold on point intensity

A point's reflection intensity may be anything between 0 and 1, inclusive. After completing an experiment, we determined that the lane points were more intense than 0.5. We thus created a criteria of 0.5 to choose possible lane places. The result may be seen in the table below:

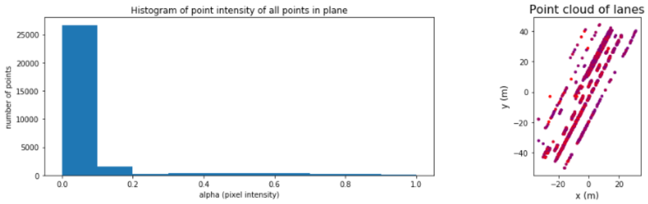

Fig 7. Point intensity of all points in plane

Here, in red, are the eight lanes that were counted. We eliminated the non-planar segments so that we could concentrate on the most probable spots for the road's surface and lanes.

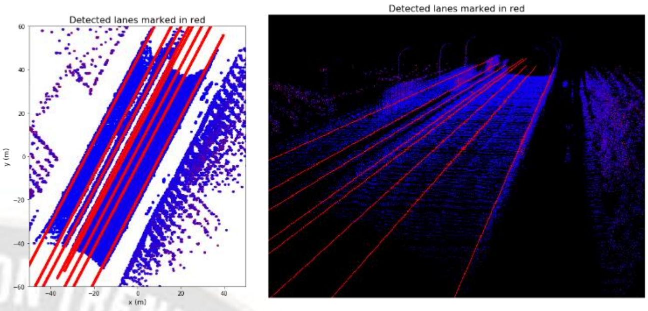

Fig 8. Red markers on the Detected lanes

The graph below shows the model complexity comparison

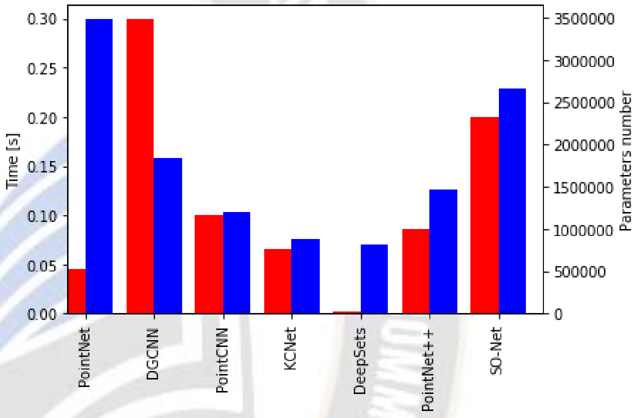

Fig 9. Comparison of the model complexity

The inference time of a machine learning model is the time it takes to make predictions on new, untested data after it has been trained.

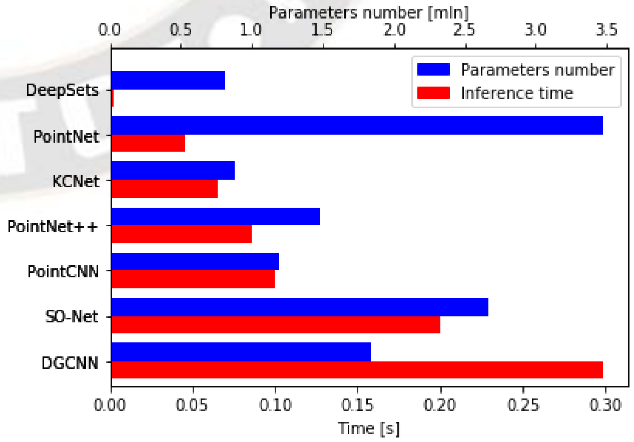

Fig 10. Inference time of the models

In the ensemble learning method known as bagging without replacement, several models are trained on different random subsets of the training data. The goal is to improve the generalisation performance of the ensemble and reduce its variance by training each model on a separate subset of the data.

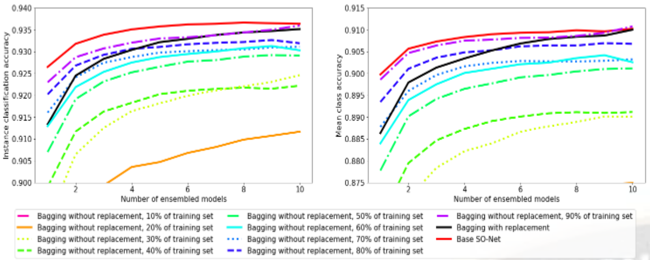

Fig 11. Class accuracy Vs Nubmer of Ensembled models

V. CONCLUSION

The point cloud data supplied in this research allowed us to successfully locate all eight lanes. As a foundation for our method, we have found that

1. The road's surface stands out as a distinct, somewhat flat region in the point cloud.
2. There are lanes in the road, and they reflect more light than the rest of the pavement. Also, the lanes are perfectly straight and parallel to one another.

We developed the following lane detecting algorithm based on these results.

1. First, find the road surface, and then remove the irregular parts.
2. Limit of reflected light intensity at which lane inflections are useless.
3. Find out which way the lane goes. Cluster analysis might help you tell the lanes apart.
4. As a further step, we make an approximation of the line equation for each lane.

The point cloud data presented shows that this framework is effective in lane detection, validating the validity and precision of the proposed method.

The proposed method was tested on a single point cloud and may not be applicable to other real-world scenarios. A bigger dataset is needed to develop a more stable and general method for lane identification.

The lanes may either be straight or curved. A quadratic function, as opposed to a linear model, might be used to fit the curvature of each lane, with feature reduction then being performed on the associated manifold in future studies.

REFERENCES

[1] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3D semantic parsing of large-scale indoor spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1534–1543, 2016.

[2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5828–5839, 2017.

[3] Francis Engelmann, Theodora Kontogianni, Alexander Hermans, and Bastian Leibe. Exploring spatial context for 3D semantic segmentation of point clouds. In Proceedings of the IEEE International Conference on Computer Vision, pages 716–724, 2017.

[4] Jun Fu, Jing Liu, Haijie Tian, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3146–3154, 2019.

[5] Benjamin Graham, Martin Engelcke, and Laurens van der Maaten. 3D semantic segmentation with submanifold sparse convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3577–3586, 2018.

[6] Qiangui Huang, Weiyue Wang, and Ulrich Neumann. Recurrent slice networks for 3D segmengtation of point clouds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2626–2635, 2018.

[7] Roman Klokov and Victor Lempitsky. Escape from cells: Deep Kd-networks for the recognition of 3D point cloud models. In Proceedings of the IEEE International Conference on Computer Vision, pages 863–872, 2017.

[8] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4558–4567, 2018.

[9] Truc Le and Ye Duan. Pointgrid: A deep network for 3D shape understandings. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 9204–9214, 2018.

[10] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. PointCNN: Convolution on X-transformed points. In Advances in Neural Information Processing Systems, pages 820–830, 2018.

[11] Yongcheng Liu, Bin Fan, Shiming Xiang, and Chunhong Pan. Relation-shape convolutional neural network for point cloud analysis. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 8895–8904, 2019.

[12] Guan Pang and Ulrich Neumann. 3D point cloud object detection with multi-view convolutional neural network. In 2016 23rd International Conference on Pattern Recognition (ICPR), pages 585–590. IEEE, 2016.

[13] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 652–660, 2017.

[14] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a

metric space. In Advances in Neural Information Processing Systems, pages 5099–5108, 2017.

[15] Xiaojuan Qi, Renjie Liao, Jiaya Jia, Sanja Fidler, and Raquel Urtasun. 3D graph neural networks for RGBD semantic segmentation. In International Conference on Computer Vision (ICCV), pages 5199–5208, 2017.

[16] Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger. Octnet: Learning deep 3D representations at high resolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3577–3586, 2017.

[17] Ji Hou, Angela Dai, and Matthias Nießner. 3d-sis: 3d semantic instance segmentation of rgb-d scans. In CVPR, 2019.

[18] Paul VC Hough. Machine analysis of bubble chamber pictures. In Conf. Proc., 1959.

[19] Vadetay Saraswathi Bai, T. Sudha. (2023). A Systematic Literature Review on Cloud Forensics in Cloud Environment. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 565–578. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2727

[20] Allison Janoch, Sergey Karayev, Yangqing Jia, Jonathan T Barron, Mario Fritz, Kate Saenko, and Trevor Darrell. A category-level 3d object dataset: Putting the kinect to work. In Consumer depth cameras for computer vision. 2013.

[21] Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In CVPR, 2017.

[22] Jan J Koenderink and Andrea J Van Doorn. Affine structure from motion. JOSA A, 8(2):377–385, 1991.

[23] Jason Ku, Melissa Mozifian, Jungwook Lee, Ali Harakeh, and Steven L Waslander. Joint 3d proposal generation and object detection from view aggregation. In IROS. IEEE, 2018.

[24] Jean Lahoud and Bernard Ghanem. 2d-driven 3d object detection in rgb-d images. In CVPR, pages 4622–4630, 2017.

[25] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In CVPR, 2019.

[26] Chen Wang, Danfei Xu, Yuke Zhu, Roberto Mart´ın-Mart´ın, Cewu Lu, Li Fei-Fei, and Silvio Savarese. Densefusion: 6d object pose estimation by iterative dense fusion. In CVPR, pages 3343–3352, 2019.

[27] Weiyao Wang, Du Tran, and Matt Feiszli. What makes training multi-modal networks hard? arXiv preprint arXiv:1905.12681, 2019.

[28] Carlos Silva, David Cohen, Takashi Yamamoto, Maria Petrova, Ana Costa. Ethical Considerations in Machine Learning Applications for Education. Kuwait Journal of Machine Learning, 2(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/192

[29] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. Sun3d: A database of big spaces reconstructed using sfm and object labels. In ICCV. IEEE, 2013.

[30] Danfei Xu, Dragomir Anguelov, and Ashesh Jain. Pointfusion: Deep sensor fusion for 3d bounding box estimation. In CVPr, pages 244–253, 2018.

[31] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-CNN: Octree-based convolutional neural networks for 3D shape analysis. ACM Transactions on Graphics (TOG), 36(4):72, 2017.

[32] Weiyue Wang, Ronald Yu, Qiangui Huang, and Ulrich Neumann. SGPN: Similarity group proposal network for 3D point cloud instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2569–2578, 2018.

[33] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics (TOG), 2019.

[34] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1912–1920, 2015.

[35] Saining Xie, Sainan Liu, Zeyu Chen, and Zhuowen Tu. Attentional shapecontextnet for point cloud recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4606–4615, 2018.

[36] Mark White, Thomas Wood, Maria Hernandez, María González , María Fernández. Enhancing Learning Analytics with Machine Learning Techniques. Kuwait Journal of Machine Learning, 2(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/184

[37] Yin Zhou and Oncel Tuzel. VoxelNet: End-to-end learning for point cloud based 3D object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4490–4499, 2018.