A PROJECTION-BASED POINT CLOUD SEMANTIC SEGMENTATION APPROACH USING DEEP LEARNING AND 3D GEOMETRIC FEATURES

Muhammed Enes Atik[1] and Zaide Duran[1]
[1]Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Turkey
Email: atikm@itu.edu.tr, duranza@itu.edu.tr

KEY WORDS: Point Cloud, Semantic Segmentation, Range Image, Geometric Feature, Deep Learning.

ABSTRACT: Modeling, understanding, and interpreting of the environment become an important task for autonomous systems in vehicles. The ability to sense the environment accurately and robustly in real time is essential for autonomous driving. Mobile point clouds are data obtained using laser scanners mounted on a moving vehicle. An accurate perception of the environment and precise location are essential for autonomous cars to reliably navigate and operate safely in complicated dynamic contexts. For these purposes, semantic segmentation of point clouds is an essential requirement. This study presents a projection-based point cloud semantic segmentation approach that combines 3D data structure and 2D segmentation techniques. Range images are created by projecting the irregular structure of the point cloud to the 2D plane. Each pixel in the range image is defined by vectors containing 3D geometric features. Experiments were carried out on Pandaset, a mobile lidar scanning point cloud. The PandaSet contains 4800 unorganized lidar point cloud scans of the various city scenes captured using the Pandar 64 sensor. The data set provides semantic segmentation labels for 42 different classes including car, road, and pedestrian. U-Net was used as the segmentation algorithm. As a result of the study, 91.89% overall accuracy and 60.82% mIoU were obtained with the U-Net architecture.

1. INTRODUCTION

Robust and real-time sensing of the environment has become an essential task for autonomous driving systems (Atik and Duran, 2022). Point clouds are widely used to reconstruct the shape and surface of objects (Akyol and Duran, 2014). LiDARs are an essential sensor type for capturing direct point clouds (Biasutti *et al.*,2019). Many current studies in photogrammetry, remote sensing, computer vision, and robotics focus on the extraction of information from three-dimensional (3D) data (Atik *et al.*, 2022). Artificial intelligence approaches such as machine learning and deep learning are used in many applications in photogrammetry and remote sensing (Atik and Ipbuker, 2020). An accurate perception of the environment and precise location are essential for autonomous cars to navigate and operate safely in complicated dynamic contexts reliably.

Current studies in autonomous driving focus on point cloud segmentation with deep learning (Atik and Duran, 2022; Nagy and Benedek, 2019). Point-wise multi-layer perceptrons (MLPs) approaches directly process the point cloud and learn the properties using shared MLPs. (Qi et al., 2017; Qi et al., 2017b; Hu et al., 2020; Jiang et al., 2018). point convolution methods recognize weights based on features learned by convolutions, considering the relationships of the points (Li *et al.*, 2018; Xu *et al.*, 2018, Zhou *et al.*, 2021). Besides, there are graph-based methods (Wang *et al.*, 2018; Landrieu and Simonovsky, 2018) that process point cloud directly and approaches using recurrent neural network (RNN) (Atik and Duran, 2021; Huang *et al.*, 2018, Ye *et al.*, 2018). There are also approaches that apply certain transformations to the input data instead of using the point cloud directly. Voxel-based methods process the point cloud by transforming it into structures of regular shapes (Maturana and Scherer, 2015). Projection-based methods define point cloud semantic segmentation as an image processing problem by reducing point cloud from 3D space to 2D plane (Atik and Duran, 2022, Biasutti *et al.*,2019). In the literature, there are also studies on point cloud classification by using 3D geometric features with machine learning methods (Weinmann *et al.*, 2015; Duran *et al.*, 2021; Atik *et al.*, 2021).

This study proposes a deep learning-based approach for point cloud semantic segmentation, in which geometric features and U-Net architecture are used together. Experiments were carried out on Pandaset, a mobile LiDAR point cloud. Segmentation was applied to the created image with U-net architecture.

2. MATERIAL AND METHODS

2.1 PandaSet Dataset

The PandaSet (Heisai and Scale, 2020) contains 4800 unorganized lidar point cloud scans of the various city scenes captured using the Pandar 64 sensor. The data set provides semantic segmentation labels for 42 classes, including car, road, and pedestrian. This example uses a subset of PandaSet containing 2560 preprocessed organized point clouds. Each point cloud is specified as a 64-by-1856 matrix. The method reorganizes point clouds into a 3-dimensional structure. Point clouds are defined within structures with dimensions Nxmx3. Thus, the point cloud has an image-like structure.The corresponding ground truth contains the semantic segmentation labels for 13 classes.

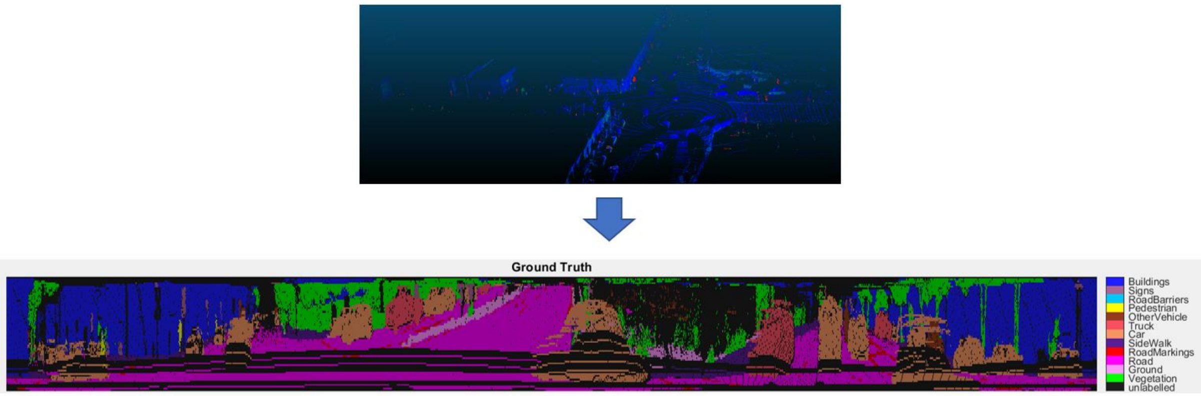

Figure 1. Point cloud is projected on 2D plane.

2.2 U-Net

U-Net (Ronneberger *et al.*, 2015) architecture is a commonly used CNN network in image segmentation. It has an encoder-decoder structure. That is, firstly, the input data is gradually reduced for feature extraction and then brought back to its original size. U-Net architecture, which has 23 convolutional layers, includes multichannel feature maps and has a contracting and expansive paths. The U-Net model generally consists of a convolution layer with a window size of 3 × 3 followed by a series of subsampling layers with a window size of 2 × 2 (Atik and Ipbuker, 2021). It is widely used for many computer vision segmentation tasks because it works with very few training data and still yields precise segmentation results. The general structure of U-Net is shown in the figure.

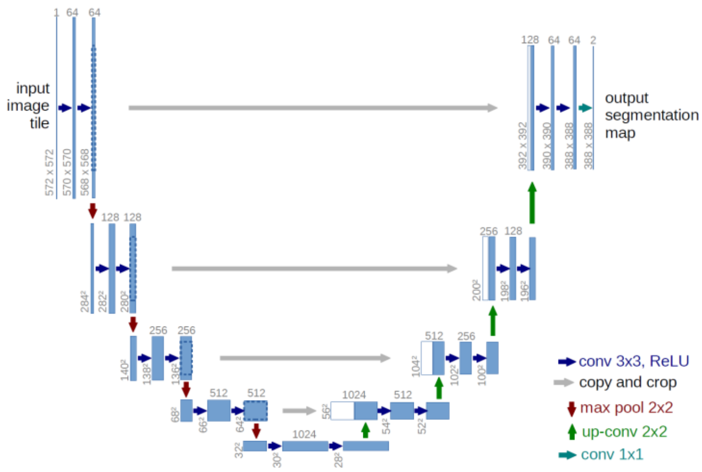

Figure 2. General structure of U-Net architecture (Ronneberger *et al.*, 2015).

2.3 3D Geometric Features

The local geometry surrounding the point is described by geometric features, which are frequently used in LiDAR processing. A sphere or other geometric shape with that point as the center can be used to calculate the neighboring points surrounding a given location (Weinmann *et al.*, 2015). This neighborhood area is called the support area. In this study, the support area was determined with a sphere. In addition to the geometric features calculated from the neighborhood of a point, 3 additional features have been added. These features can be increased or decreased depending on the point cloud structure. The 3D geometric features used in this study are presented in Table 1.

Table 1. 3D geometric features were used in this study.

Feature	Explanation
Omnivariance	$\sqrt[3]{\lambda_1 \lambda_2 \lambda_3}$
Eigenentropy	$\sum_{i=1}^{3} \lambda_i \ln \lambda_i$
Anisotropy	$(\lambda_1 - \lambda_3)/\lambda_1$
Planarity	$(\lambda_2 - \lambda_3)/\lambda_1$
Linearity	$(\lambda_1 - \lambda_2)/\lambda_1$
Sphericity	λ_3/λ_1
Verticality	$1 - \lvert\langle[0\ 0\ 1], \lambda_3\rangle\rvert$
Height value	Z_i
Normal change rate	
Intensity	Other features
Range	

2.4 Experiment

First, the irregular point cloud structure is projected onto a regular 2D plane. However, this is not a point cloud conversion to an image. Each dot corresponds to a pixel. Thus, there is no information loss in the point cloud. Calculated geometric features are also considered as bands of an image. Segmentation was applied to the created image with U-net architecture. The transformation is based on the SqueezeNet algorithm (Iondola et al., 2016) in literature. SqueezeNet, a point cloud segmentation architecture, uses the organized point cloud as input. There are studies in the literature to develop this approach. Transformation is based on spherical projection. Depending on the parameters of the laser scan, the point is projected onto the 2D plane. The calculated features are combined into the NxMxd structure. where d is the number of features. Once the point cloud is designed this way, it is possible to implement deep learning networks operating on a 2D image. The point cloud can be used directly in the U-Net architecture input data. For the experiments, i7-11800H, 2.30 GHz processor, GTX 3070 graphics card, and 32 GB RAM hardware is used.

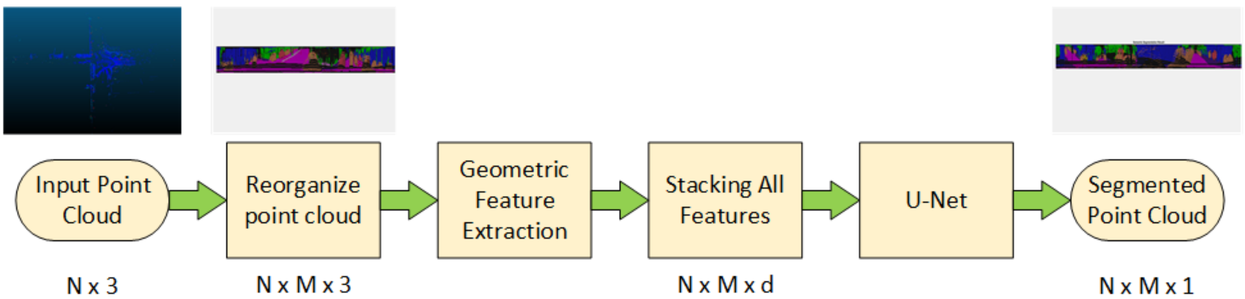

Figure 3. Workflow of study.

3. RESULTS AND DISCUSSION

Overall accuracy, mean accuracy, mean IoU, weighted IoU, and F1 score, were used as evaluation metrics. While the overall accuracy was 91.89%, the F1 score was 76.13%. Considering the class metrics, the best results were obtained in the buildings and road classes, regardless of the unlabeled class. The F1 score for building class and road class is 99.10% and 99.42%, respectively. The lowest metrics were obtained in the road barrier class. The reason for this is that the road barrier class is low in the data set and it is mixed with geometrically different structures. The F1 score of the Other vehicles class is 50.27. Vehicles that do not fall into the car and truck class are mixed with cars and trucks as a result of classification. Traffic signs class could not be determined geometrically. General dataset metrics are presented in Table 2 and class-based results are presented in Table 3. Classified range image example is presented in Figure 4.

Table 2. Evaluation results based on different support radius for eigen-based features.

Overall Accuracy	**Mean Accuracy**	**Mean IoU**	**Weighted IoU**	**F1 Score**
91.89 %	67.92 %	60.82 %	85.36 %	76.13 %

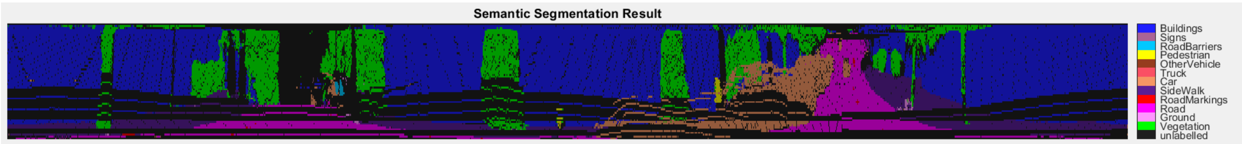

Figure 4. Predicted point cloud sample.

Table 3. Evaluation results based on different training samples for each class. The values given as %.

Number of points	**Precision**	**Recall**	**F1 Score**
Unlabeled	95.98	92.29	99.99
Vegetation	81.88	70.83	96.86
Ground	78.22	66.90	53.42
Road	94.09	85.77	99.42
Road markings	35.12	31.60	76.36
Sidewalk	83.38	70.46	96.63
Car	94.39	85.56	96.83
Truck	56.74	50.95	57.36
Other vehicles	73.45	66.78	50.27
Pedestrian	46.40	36.41	56.00
Road barriers	25.03	23.90	18.38
Signs	22.91	20.79	48.09
Buildings	95.32	88.44	99.10

The approach has some advantages and disadvantages. First, it is important to use the point cloud without losing information. Since 2D techniques have more comprehensive literature, they allow easy development of the method. It can be used with point clouds obtained from different sensors. On the other hand, the approach has some limitations. First, a preprocessing step is required. Reorganization of the point cloud should be done. However, while it is possible for this preprocessing step to be embedded in the architecture, it will greatly slow down the training and testing phases. In addition, since the method depends on the geometric features, the parameters determined in the calculation of the geometric features affect the accuracy. In particular, the neighborhood radius of the point cloud should be determined again for each point cloud.

The proposed approach is also compared with the current algorithms in the literature. It has been demonstrated that the proposed approach is superior to the other approaches for the Pandaset dataset. The highest evaluation metrics were obtained in the method proposed in our study.

Table 4. Comparative results with current studies in literature. The values given as %.

Model	**Overall accuracy**	**Mean IoU**	**Weighted IoU**	**F1 Score**
SqueezeSegv2 (Wu *et al.*, 2019)	89.72	54.53	81.81	74.54
SalsaNext (Cortinhal *et al.*, 2020)	-	57.80	-	-
PointSeg (Wang *et al.*, 2018)	80.07	46.24	71.37	71.92
Our study	**91.89**	**60.82**	**85.36**	**76.13**

3. CONCLUSIONS

In this study, a projection-based point cloud semantic segmentation approach using traditional image processing algorithms is presented. In order to increase the original value of the study, a new approach can be suggested instead of U-Net in future studies. In addition to geometrical properties, radiometric properties can also be used. Deep learning approaches have great potential for semantic segmentation of complex and large point clouds. The use of powerful algorithms is an important need for safe autonomous driving.

REFERENCES

Akyol, O., Duran, Z. 2014. Low-cost laser scanning system design. Journal of Russian Laser Research, 35(3), pp. 244-251.

Atik, M. E., Duran, Z. 2021. Classification of Aerial Photogrammetric Point Cloud Using Recurrent Neural Networks. Fresenius Environ. Bull, 30, pp. 4270-4275.

Atik, M. E., Duran, Z. 2022. An Efficient Ensemble Deep Learning Approach for Semantic Point Cloud Segmentation Based on 3D Geometric Features and Range Images. Sensors, 22(16), 6210.

Atik, M. E., Duran, Z., Seker, D. Z. 2021. Machine learning-based supervised classification of point clouds using multiscale geometric features. ISPRS International Journal of Geo-Information, 10(3), 187.

Atik, S. O., Atik, M. E., Ipbuker, C. 2022. Comparative research on different backbone architectures of DeepLabV3+ for building segmentation. Journal of Applied Remote Sensing, 16(2), 024510.

Atik, S. O., Ipbuker, C. 2020. Instance Segmentation Of Crowd Detection In The Camera Images. In Proceeding of Asian Conference on Remote Sensing.

Atik, S. O., Ipbuker, C. 2021. Integrating convolutional neural network and multiresolution segmentation for land cover and land use mapping using satellite imagery. Applied Sciences, 11(12), 5551.

Biasutti, P., Lepetit, V., Aujol, J. F., Brédif, M., Bugeau, A. 2019. Lu-net: An efficient network for 3d lidar point cloud semantic segmentation based on end-to-end-learned 3d features and u-net. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 0-0).

Cortinhal, T., Tzelepis, G., Erdal Aksoy, E. 2020. SalsaNext: Fast, uncertainty-aware semantic segmentation of LiDAR point clouds. In International Symposium on Visual Computing. pp. 207-222. Springer, Cham.

Duran, Z., Ozcan, K., Atik, M. E. 2021. Classification of Photogrammetric and Airborne LiDAR Point Clouds Using Machine Learning Algorithms. Drones, 5(4), 104.

Hesai and Scale. 2020. Pandaset. https://scale.com/open-av-datasets/pandaset.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Markham, A. (2020). Randla-net: Efficient semantic segmentation of large-scale point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11108-11117.

Huang, Q., Wang, W., Neumann, U. 2018. Recurrent slice networks for 3d segmentation of point clouds. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2626-2635.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., Keutzer, K. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360.

Jiang, M., Wu, Y., Zhao, T., Zhao, Z., Lu, C. 2018. Pointsift: A sift-like network module for 3d point cloud semantic segmentation. arXiv preprint arXiv:1807.00652.

Landrieu, L., Simonovsky, M. 2018. Large-scale point cloud semantic segmentation with superpoint graphs. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4558-4567.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B. 2018. Pointcnn: Convolution on x-transformed points. Advances in neural information processing systems, 31.

Maturana, D., Scherer, S. 2015. Voxnet: A 3d convolutional neural network for real-time object recognition. In 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). pp. 922-928. IEEE.

Nagy, B., & Benedek, C. (2019). 3D CNN-based semantic labeling approach for mobile laser scanning data. IEEE

Sensors Journal, 19(21), 10034-10045.

Qi, C. R., Su, H., Mo, K., Guibas, L. J. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652-660.

Qi, C. R., Yi, L., Su, H., Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30.

Ronneberger, O., Fischer, P., Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention. pp. 234-241. Springer, Cham.

Wang, Y., Shi, T., Yun, P., Tai, L., Liu, M. 2018. Pointseg: Real-time semantic segmentation based on 3d lidar point cloud. arXiv preprint arXiv:1807.06288.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., Solomon, J. M. 2019. Dynamic graph cnn for learning on point clouds. Acm Transactions On Graphics (tog), 38(5), pp. 1-12.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C. 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry and Remote Sensing, 105, 286-304.

Wu, B., Zhou, X., Zhao, S., Yue, X., Keutzer, K. 2019. 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). pp. 4376-4382. IEEE.

Xu, Y., Fan, T., Xu, M., Zeng, L., Qiao, Y. 2018. Spidercnn: Deep learning on point sets with parameterized convolutional filters. In Proceedings of the European Conference on Computer Vision (ECCV). pp. 87-102.

Ye, X., Li, J., Huang, H., Du, L., Zhang, X. 2018. 3d recurrent neural networks with context fusion for point cloud semantic segmentation. In Proceedings of the European conference on computer vision (ECCV). pp. 403-417.

Zhou, H., Feng, Y., Fang, M., Wei, M., Qin, J., Lu, T. 2021. Adaptive graph convolution for point cloud analysis. In Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4965-4974.