

YUTO SEMANTIC: A LARGE SCALE AERIAL LIDAR DATASET FOR SEMANTIC SEGMENTATION

S. Yoo^{1*}, C. Ko¹, G. Sohn^{1*}, H. Lee¹

¹ GeoICT Lab; Dept. of Earth and Space Science and Engineering, York University, Toronto, ON, M3J 1P3 Canada -
(jacobyoo, cko, gsohn, hyungju)@yorku.ca

KEY WORDS: semantic segmentation, aerial imagery, laser scanning, evaluation, test

ABSTRACT:

Creating virtual duplicates of the real world has garnered significant attention due to its applications in areas such as autonomous driving, urban planning, and urban mapping. One of the critical tasks in the computer vision community is semantic segmentation of outdoor collected point clouds. The development and research of robust semantic segmentation algorithms heavily rely on precise and comprehensive benchmark datasets. In this paper, we present the York University Teledyne Optech 3D Semantic Segmentation Dataset (YUTO Semantic), a multi-mission large-scale aerial LiDAR dataset specifically designed for 3D point cloud semantic segmentation. The dataset comprises approximately 738 million points, covering an area of 9.46 square kilometers, which results in a high point density of 100 points per square meter. Each point in the dataset is annotated with one of nine semantic classes. Additionally, we conducted performance tests of state-of-the-art algorithms to evaluate their effectiveness in semantic segmentation tasks. The YUTO Semantic dataset serves as a valuable resource for advancing research in 3D point cloud semantic segmentation and contributes to the development of more accurate and robust algorithms for real-world applications. The dataset is available at https://github.com/Yacovitch/YUTO_Semantic.

1. INTRODUCTION

The growing significance of the accuracy and robustness of datasets, particularly for the creation of large-scale virtual duplicates of outdoor environments, lead to the adoption of Airborne Laser Scanning (ALS) as a prominent remote sensing technique. ALS harnesses the power of laser technology to measure the three-dimensional structure of outdoor scenes from aerial vehicles, thereby providing invaluable insights into the spatial characteristics of our surroundings.

In recent years, remarkable advancements in LiDAR and drone technologies revolutionized the field, enabling the development of more precise and large virtual duplicates of the real world. Notable examples of these LiDAR datasets include Sensat-Urban (Hu et al., 2021), DALES (Varney et al., 2020), and DublinCity (Zolanvari et al., 2019), which emerged as leading benchmark datasets for testing algorithms. These benchmark datasets played a pivotal role in the research and development of 3D spatial data analysis and visualization.

One critical aspect of processing LiDAR datasets is semantic segmentation, a fundamental task that involves assigning a semantic class to each point in a point cloud. This task was traditionally executed manually. It was laborious and time-consuming, with the result of demanding extensive human effort and expertise. Consequently, there is a growing interest in exploring automated approaches instead. Particularly, utilizing deep learning-based 3D semantic segmentation algorithms such as PointNet (Qi et al., 2017a), RandLA (Hu et al., 2020), KPConv (Thomas et al., 2019), and EyeNet (Yoo et al., 2023), demonstrated promising results in simplifying point cloud labeling.

In this context, we present the York University Teledyne Optech 3D Semantic Segmentation Dataset (YUTO Semantic) as

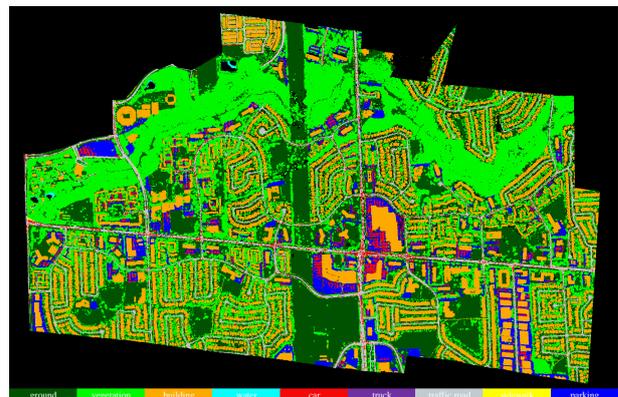


Figure 1. The YUTO dataset visualization

a comprehensive aerial LiDAR dataset specifically designed for semantic segmentation purposes. YUTO Semantic encompasses a vast area of 9.46 square kilometers, covering the sprawling York University Campus situated in Toronto, Canada. This extensive dataset was acquired using an ALS system deployed on multiple missions, ensuring comprehensive coverage and capturing diverse environmental characteristics.

YUTO Semantic comprises astounding 738 million points, resulting in an average point density of 100 points per square meter. Each point in the dataset has been meticulously labelled with one of nine semantic classes, including ground, traffic road, sidewalk, water, and various others. The careful and detailed annotation of these points enables researchers and practitioners to leverage YUTO Semantic for various applications ranging from urban planning and infrastructure development to autonomous navigation systems and environmental analysis.

Category	Datasets	Year	Data Type	Spatial Size/ Length	Classes	Points	Point Density	Type of Sensors
Outdoor Ground-Level	Oakland 3D (Munoz et al., 2009)	2009	Point Cloud	1.5 km	5	1.6M	–	MLS
	Paris-rue-Madame (Serna et al., 2014)	2014	Point Cloud	0.16 km	17	20M	–	MLS
	Semantic3D (Hackel et al., 2017)	2017	Point Cloud	–	8	4000M	–	TLS
	Paris-Lille-3D (Roynard et al., 2018)	2018	Point Cloud	1.94 km	9	143M	–	MLS
	SemanticKITTI (Behley et al., 2019)	2019	Point Cloud	39.2 km	25	4549M	–	MLS
	Toronto-3D (Tan et al., 2020)	2020	Point Cloud	1.0 km	8	78.3M	–	MLS
Outdoor Airborne	ISPRS (Rottensteiner et al., 2012)	2012	Point Cloud	0.1 km ²	9	1.2M	1.52ppm ²	ALS
	DublinCity (Zolanvari et al., 2019)	2019	Point Cloud	2 km ²	11	260M	105.25ppm ²	ALS
	DALES (Varney et al., 2020)	2020	Point Cloud	10 km ²	8	505M	42.83ppm ²	ALS
	LASDU (Yusheng Xu, 2020)	2020	Point Cloud	1.02 km ²	5	3M	3.45ppm ²	ALS
	Campus3D (Li et al., 2020)	2020	Mesh	1.58 km ²	14	937M	65.34ppm ² *	Photogrammetry
	SensatUrban (Hu et al., 2021)	2021	Point Cloud	6 km ²	13	2847M	261.54ppm ²	Photogrammetry
	SUM (Gao et al., 2021)	2021	Mesh	4 km ²	6	19M	458.13ppm ² *	Photogrammetry
	YUTO(Ours)	2023	Point Cloud	9.46 km ²	9	738M	100ppm ²	ALS

Table 1. Benchmark Dataset Comparison. * indicates that the point density is based on point cloud that is generated through the mesh.

By introducing YUTO Semantic, we aim to contribute to the growing body of research on semantic segmentation of LiDAR datasets, providing a valuable resource for benchmarking. YUTO Semantic highlights the remarkable progress in ALS technology, revolutionizing geospatial data analysis. By providing an extensive and meticulously annotated dataset, we anticipate an acceleration of research efforts, promoting innovation and advancements in 3D semantic segmentation and its various applications.

2. RELATED WORK

Creating reliable and comprehensive benchmark datasets for 3D point cloud semantic segmentation is a challenging task that requires significant human effort and careful planning. However, it provides valuable opportunities for the field of 3D semantic segmentation algorithms. Existing benchmark datasets for 3D point cloud semantic segmentation can be categorized into three main types.

The first category is **the Indoor 3D Point Cloud Dataset**, which focuses on understanding indoor scenes for semantic segmentation. These datasets are typically collected by the stationary survey LiDAR or depth sensors combined with RGB image data. Examples of datasets in this category include SUN RGB-D (Song et al., 2015), S3DIS (Armeni et al., 2016), SceneNN (Hua et al., 2016), and ScanNet (Dai et al., 2017).

The second category is **the Outdoor Ground-Level 3D Point Cloud Dataset**, which is primarily collected for the applications such as autonomous driving. These datasets involve the use of LiDAR and RGB sensors, while the sensors are typically mounted on a stationary platform or a moving vehicle. They often obtain the point cloud data through short scan-by-scan acquisitions, and the data would be paired with RGB images later. Examples of datasets in this category include OakLand3D (Munoz et al., 2009), KITTI (Geiger et al., 2013), Paris-rue-Madame (Serna et al., 2014), Semantic3D (Hackel et al., 2017), ParisLille-3D (Roynard et al., 2018), SemanticKitti (Behley et al., 2019), Toronto3D (Tan et al., 2020), and nuScenes (Caesar et al., 2020).

The third category is **the Outdoor Airborne 3D Point Cloud Dataset**, which aims to understand urban-level scenes for 3D semantic segmentation. These datasets involve the use of expensive airborne LiDAR systems mounted on drones or airplanes. Earlier datasets in this category ordinarily lacked RGB

information. However, recent advancements enabled the inclusion of RGB data. Additionally, one distinguishing characteristic of datasets in this category is their coverage of large areas. These datasets are often collected by conducting sweeping flights over extensive regions. Notable examples of datasets in this category include ISPRS (Rottensteiner et al., 2012), DublinCity (Zolanvari et al., 2019), DALES (Varney et al., 2020), LASDU (Yusheng Xu, 2020), SensatUrban (Hu et al., 2021), SAM (Gao et al., 2021), and Campus3D (Li et al., 2020). Comparisons between existing datasets and ours are shown in Table 1

3. THE YUTO SEMANTIC DATASET

3.1 Multi Mission Data Collection

The data collection through multiple missions conducted in different seasons is what differentiates our dataset from the others. Originally, the first dataset was collected on September 23, 2018, utilizing an ALS system, Teledyne Optech Galaxy Prime Airborne Lidar Terrain Mapper, mounted on an airplane. The initial flight covered an approximate area of 22 square kilometers. On top of initial data collection with the Galaxy Prime sensor, we carried out two more flights over the York University Campus in December 2019 and May 2021, covering an approximate area of 23 and 47 square kilometers, respectively. These flights encompassed both on-campus and off-campus regions.

During the flight missions, the ALS system maintained an average flight altitude of 1871 meters. To maximize coverage area and point density, multiple individual strips were acquired. The collected data was projected using the UTM zone 17N coordinate system, with the horizontal datum of NAD83. To ensure the accuracy of trajectory information, the PP-RTX base station was utilized. Finally, the trajectory information was extracted by the Applenix POSpac software. After boresight calibration, 80% of the data was found to maintain a height difference within 5cm, while 93% of the data was within 10cm from the boresight calibration.

3.2 Data Description

During the data collection phase, our main focus was specifically on the on-campus areas, which cover 9.46 square kilometers. The collected point cloud dataset comprised a total of 738 million points, resulting in a point density of 100 points per square meter. To facilitate data processing and analysis more efficiently, we divided the point cloud into 600-meter by 600-meter squares tiles, generating a total of 41 tiles. Among them,

we selected 32 tiles for training purposes and reserved 9 tiles for testing our algorithms. They were processed in *.ply* file format, which is the most universal LiDAR data file format. Each point of the dataset was associated with the following attributes:

- **x, y, z:** The position coordinates of each point recorded in UTM zone 17N using the NAD83 horizontal datum.
- **Intensity:** The normalized LiDAR intensity value of each point, ranging from 0 to 255.
- **Number of Return:** The number of times the laser pulse was reflected back.
- **GPS time:** The GPS time of each point, providing temporal information about the data acquisition.
- **Scan Angle:** The scan angle of each point, indicating the angle at which the laser beam hit the target.
- **Class:** The label assigned to each point, representing the semantic class or category of the object or surface the point belongs to.

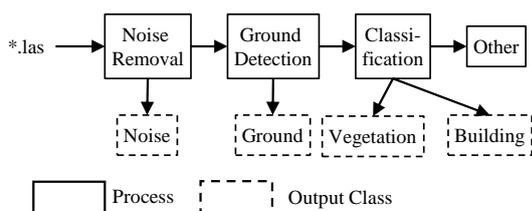


Figure 2. Schematic diagram for data annotation..

3.3 Data Annotation

In our study, we assigned labels to the point cloud dataset based on nine semantic classes. The assigned labels were as follows:

- **Ground:** This class includes unpaved surfaces, grass, and natural terrain.
- **Vegetation:** This class encompasses trees, bushes, and other forms of vegetation.
- **Building:** This class represents both commercial and residential buildings.
- **Water:** This class includes bodies of water such as lakes and rivers.
- **Car:** This class includes all types of vehicles except for commercial trucks.
- **Truck:** This class specifically represents commercial trucks.
- **Traffic Road:** This class corresponds to vehicle roads.
- **Sidewalk:** This class represents pedestrian walkways.
- **Parking:** This class represents parking lots.

For the initial data that was collected in 2018, the annotation process involved the use of two software tools: LAStools (rapidlasso GmbH, 2023) and TerraScan (Accurics, 2023). The schematic diagram of this process is depicted in Figure 2. Initially, noise removal was performed automatically by LAStools. Subsequently, the LAStools software was employed as an automatic labeling tool to annotate specific classes such as ground, vegetation, and building. The remaining classes were manually classified utilizing TerraScan. During this step, a manual cross-checking step was conducted as well to ensure the quality and consistency of the annotations.

For the data collected in 2019 and 2021, a proximity-based method was utilized for automatic labeling. The closest distance between the point cloud and a reference point cloud (from 2018, which was manually labelled) was determined, and the classification code from the reference point cloud was copied over to assign labels. It is important to note that the automatic transfer of labels was limited to specific classes, namely ground, vegetation, building, water, traffic road, sidewalk, and parking. The car and truck classes were excluded from this automatic transfer process because the characteristics and points associated with these objects could vary between different years. Therefore, manual labeling was deemed necessary to accurately classify car and truck points in each specific year.

To provide visual insights into the dataset, Figure 1 showcases a visualization of the point cloud data. Additionally, the distribution of labels for the 2018 data is displayed in Figure 4.

4. BENCHMARKS

4.1 3D Pointcloud Semantic Segmentation

Recently, 3D point cloud semantic segmentation gains significant attention due to its wide range of applications. This task involves assigning semantic labels to individual points in a point cloud, and it is typically approached by extracting features from points or local neighborhoods and employing machine learning techniques, such as deep learning, to predict the semantic labels.

Different methods are developed to address this task, including voxel-based approaches and point-based approaches. Voxel-based approaches, exemplified by VoxNet (Maturana and Scherer, 2015), involve dividing the point cloud space into smaller volumetric units known as voxels. However, voxel-based approaches suffer from limitations such as loss of fine-grained details due to voxelization, limited representation of objects due to fixed-size voxels, and high computational costs when processing the entire point cloud.

In contrast, point-based approaches such as PointNet (Qi et al., 2017a), PointNet++(Qi et al., 2017b), RandLA(Hu et al., 2020), KPConv (Thomas et al., 2019), and EyeNet (Yoo et al., 2023) shows promising performance in recent studies. The point-based approach directly operates on individual points without the need for voxelization or an explicit grid structure. Each point is treated as a separate entity, allowing more flexible and efficient processing.

Point-based approaches have proven to be highly effective in addressing the challenges associated with 3D point cloud semantic segmentation. These approaches excel in preserving fine-grained details, capturing local structures, and achieving

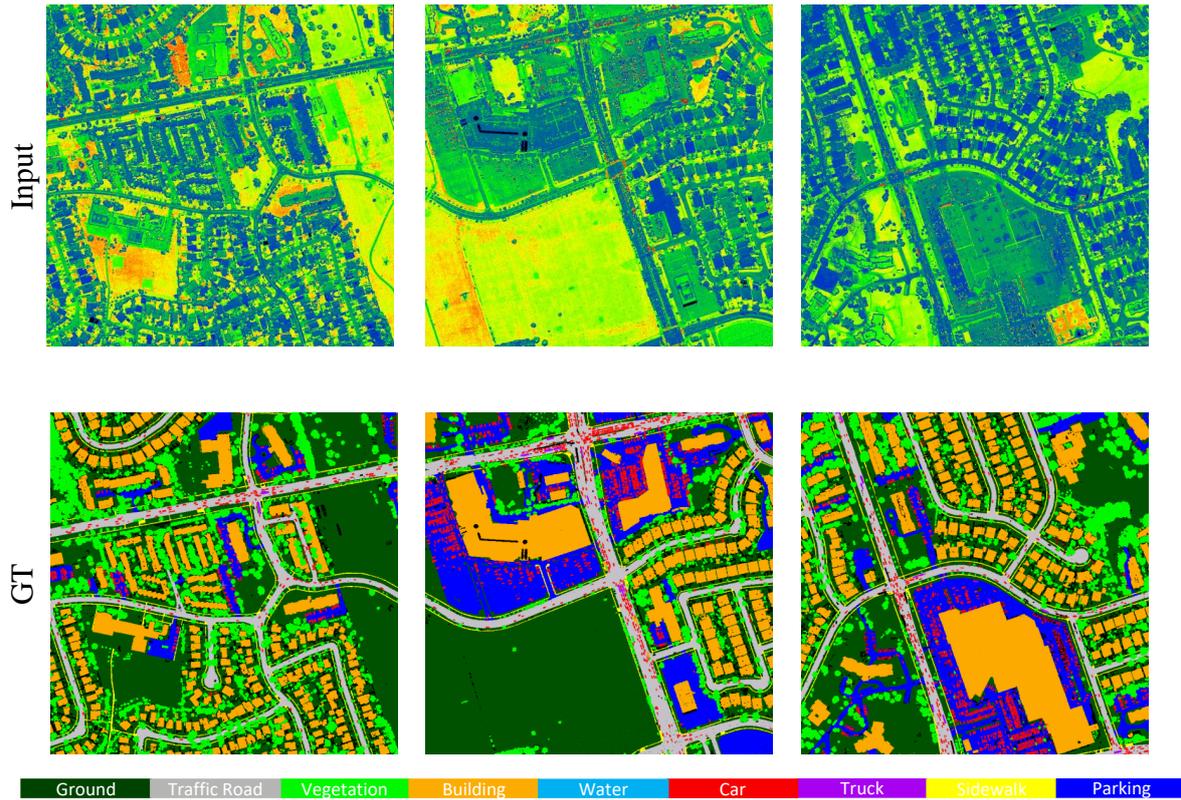


Figure 3. Visualization of the dataset: from top to bottom, the input point cloud with intensity and the point cloud with semantic labels, assigned with different colors.

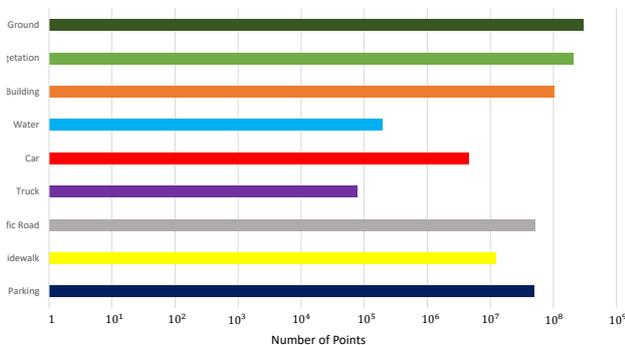


Figure 4. The distributions of labels

efficient computation. Their success in these areas highlights their potential for driving advancements in the field and enabling a wide range of applications that rely on accurate and robust semantic segmentation of 3D point clouds.

4.2 Baseline Networks

We conducted performance measurement experiments on three 3D semantic segmentation networks: RandLA, KPConv, and EyeNet.

RandLA (Hu et al., 2020) employs random point sampling as a simple and efficient method, avoiding the complexity of point selection techniques. However, random sampling poses a risk of discarding important features. To address this, RandLA introduces a local feature aggregation module that gradually expands the receptive field around each 3D point. This approach

effectively preserves geometric details and mitigates the potential loss of critical information caused by random sampling.

KPConv: (Thomas et al., 2019) directly processes point clouds without intermediate representations. It utilizes kernel points in Euclidean space to apply convolutional weights to nearby input points. This method offers flexibility and adaptability for handling point clouds. The continuous and learnable kernel point locations in KPConv enable deformable convolutions, allowing the network to adapt to local geometry and capture fine-grained details.

EyeNet: (Yoo et al., 2023) is a novel semantic segmentation network for point clouds inspired by human peripheral vision. It addresses the issue of functional coverage area size of inputs by introducing a multi-scale input and a parallel processing network with connection blocks. EyeNet overcomes the limitations of conventional networks and achieves state-of-the-art performance on the SensatUrban and Toronto3D datasets.

4.3 Evaluation Metrics

To evaluate the performance of these networks, we utilized overall accuracy (OA), per-class Intersection-over-Union (IoU_c), and mean Intersection-over-Union ($mIoU$) as the evaluation metrics. We first defined per-class IoU as:

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c} \quad (1)$$

where c , TP , FP , and FN are the class number, true positive, false positive, and false negative, respectively. Then, $mIoU$ is

calculated by finding the mean across all classes, which was defined as:

$$mIoU = \frac{\sum IoU_c}{C} \quad (2)$$

where C is the total number of classes. Lastly, OA was defined as:

$$mIoU = \frac{\sum TP_c}{Totalnumberofpoints} \quad (3)$$

The performance comparison results are presented in Table 2.

4.4 Experimental Configurations

All three networks—RandLA, KPConv, and EyeNet—were post-processed for the tests using a 0.2m grid sample. XYZ coordinates, intensity values, and the quantity of returns were used as input features throughout the training procedure.

The parameter settings for RandLA and EyeNet were adopted from the networks tested on the SensatUrban dataset. Similarly, the parameter settings for KPConv were taken from the network tested on the DALES dataset. It is important to note that while there was potential for improvement through hyperparameter optimizations, the tuning performed did not guarantee a completely fair comparison between the networks. Further fine-tuning and optimization could potentially yield better results.

The models were trained and tested using an Nvidia Quadro RTX 6000 GPU on the Determined AI server. This hardware setup provides the necessary computational power for efficient training and evaluation of the 3D semantic segmentation networks.

5. RESULTS AND DISCUSSION

5.1 Performances of Baseline Networks

The test results of the baseline networks are presented in Table 2. Additionally, visualization comparisons are provided in Figure 5.

RandLA (Hu et al., 2020) demonstrated satisfactory performance across all the classes. However, it showed a tendency to over-predict the sidewalk class, resulting in the lowest performance for the ground class among the tested networks. KPConv (Thomas et al., 2019), on the other hand, achieved the best performance in the ground, vegetation, and car classes. However, it struggled to predict the water and truck classes and rarely predicted the sidewalk class. EyeNet achieved the highest overall accuracy (OA) and mean Intersection over Union ($mIoU$) among the tested baseline networks. It also achieved the highest performance in the building, water, truck, traffic road, sidewalk, and parking classes.

5.2 Challenges

Accurately distinguishing terrain classes, including ground, water, traffic road, sidewalk, and parking, presents a significant challenge in the YUTO Semantic dataset. As evidenced in Table 2 and 5, the evaluated networks encountered difficulties

in accurately predicting these classes. This struggle can be attributed to two main factors: class imbalance and limited feature availability.

The class imbalance among the terrain classes poses a challenge for accurate segmentation. These classes have a relatively larger number of points compared to other classes, which can bias the predictions and make it difficult to achieve precise segmentation results.

Furthermore, the limited availability of features compounds the challenge of distinguishing these terrain classes. The YUTO Semantic dataset provides only intensity and the number of returns as features, without the inclusion of RGB information. The absence of RGB features restricts the networks' ability to leverage color cues and texture information, which are essential for effectively differentiating these classes. Consequently, accurately segmenting these terrain classes becomes inherently challenging.

Addressing these challenges requires further exploration and investigation. Strategies to mitigate the class imbalance and innovative approaches to leverage the available feature information effectively are crucial for enhancing the accurate segmentation of these terrain classes in the YUTO Semantic dataset. Future research efforts should focus on developing techniques that can compensate for the limited feature set and improve the distinction of these challenging classes.

6. CONCLUSION

In this study, we introduced YUTO Semantic, a multi-season large-scale aerial LiDAR dataset for semantic segmentation, obtained through multiple missions over the York University Campus using an ALS system. We evaluated three state-of-the-art 3D semantic segmentation networks: RandLA, KPConv, and EyeNet. Moving forward, we plan to expand the semantic classes and release labels for the remaining two missions, further enhancing the scope and utility of the YUTO Semantic dataset.

ACKNOWLEDGMENT

This research project has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC)'s Collaborative Research and Development Grant (CRD) - 3D Mobility Mapping Artificial Intelligence (3DMMAI) and Tele-dyne Geospatial Inc.

REFERENCES

- Accurics, 2023. Terrascan. <https://www.accurics.com/products/terrascan/>.
- Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3d semantic parsing of large-scale indoor spaces. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.
- Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J., 2019. Semantickitti: A dataset for semantic scene understanding of lidar sequences. *Proceedings of the IEEE/CVF international conference on computer vision*, 9297–9307.

Method	OA(%)	mIoU(%)	ground	vegetation	building	water	car	truck	traffic road	sidewalk	parking
RandLA (Hu et al., 2020)	84.19	58.37	80.61	94.44	95.39	3.34	74.59	13.87	78.10	23.43	61.56
KPConv (Thomas et al., 2019)	85.22	56.14	86.94	96.25	94.01	0.00	84.02	0.00	79.93	3.26	60.83
EyeNet (Yoo et al., 2023)	87.41	63.44	86.26	95.94	96.78	13.61	83.02	14.26	84.65	31.08	65.34

Table 2. YUTO Semantic Performance Comparison. Results of RandLA and KPConv are taken from internal experiments.

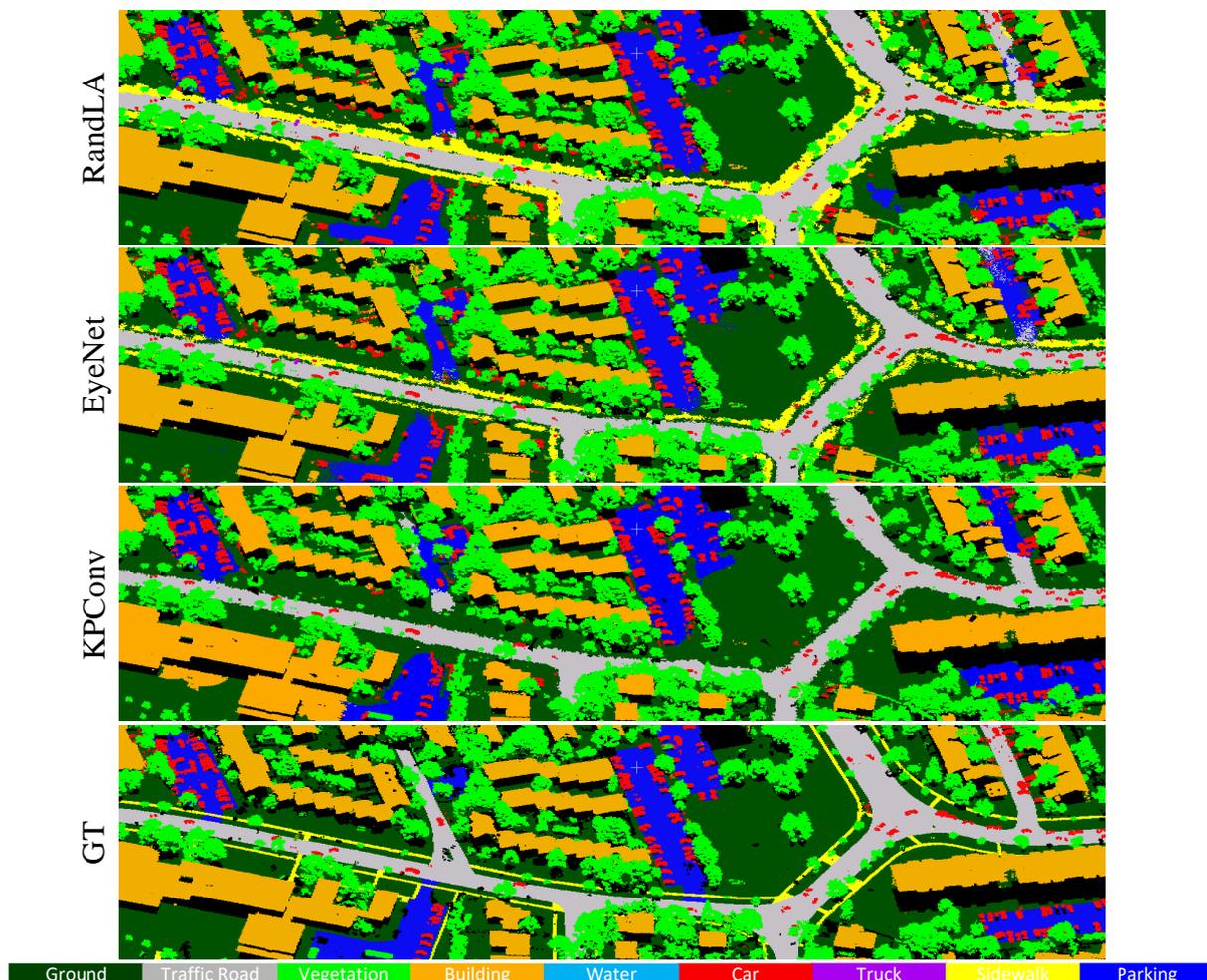


Figure 5. Visualization comparison of tested models: from top to bottom, results of RandLA, EyeNet, KPConv, and ground truth.

Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O., 2020. nuscenes: A multimodal dataset for autonomous driving. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5828–5839.

Gao, W., Nan, L., Boom, B., Ledoux, H., 2021. SUM: A Benchmark Dataset of Semantic Urban Meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179, 108–120. <https://www.sciencedirect.com/science/article/pii/S0924271621001854>.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11), 1231–1237.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., Pollefeys, M., 2017. Semantic3d. net: A new large-

scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*.

Hu, Q., Yang, B., Khalid, S., Xiao, W., Trigoni, N., Markham, A., 2021. Towards semantic segmentation of urban-scale 3d point clouds: A dataset, benchmarks and challenges. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4977–4987.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. Randla-net: Efficient semantic segmentation of large-scale point clouds. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11108–11117.

Hua, B.-S., Pham, Q.-H., Nguyen, D. T., Tran, M.-K., Yu, L.-F., Yeung, S.-K., 2016. Scenenn: A scene meshes dataset with annotations. *2016 fourth international conference on 3D vision (3DV)*, Ieee, 92–101.

Li, X., Li, C., Tong, Z., Lim, A., Yuan, J., Wu, Y., Tang, J., Huang, R., 2020. Campus3d: A photogrammetry point cloud

- benchmark for hierarchical understanding of outdoor scene. *Proceedings of the 28th ACM International Conference on Multimedia*, 238–246.
- Maturana, D., Scherer, S., 2015. Voxnet: A 3d convolutional neural network for real-time object recognition. *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 922–928.
- Munoz, D., Bagnell, J. A., Vandapel, N., Hebert, M., 2009. Contextual classification with functional max-margin markov networks. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 975–982.
- Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 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.
- rapidlasso GmbH, 2023. Lastools - efficient tools for lidar processing. <http://lastools.org>.
- Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3 (2012)*, Nr. 1, 1(1), 293–298.
- Roynard, X., Deschaud, J.-E., Goulette, F., 2018. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, 37(6), 545–557.
- Serna, A., Marcotegui, B., Goulette, F., Deschaud, J.-E., 2014. Paris-rue-madame database: a 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. *4th international conference on pattern recognition, applications and methods ICPRAM 2014*.
- Song, S., Lichtenberg, S. P., Xiao, J., 2015. Sun rgb-d: A rgb-d scene understanding benchmark suite. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 567–576.
- Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., Yang, K., Li, J., 2020. Toronto-3d: A large-scale mobile lidar dataset for semantic segmentation of urban roadways. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 202–203.
- Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE/CVF international conference on computer vision*, 6411–6420.
- Varney, N., Asari, V. K., Graehling, Q., 2020. Dales: A large-scale aerial lidar data set for semantic segmentation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 186–187.
- Yoo, S., Jeong, Y., Jameela, M., Sohn, G., 2023. Human vision based 3d point cloud semantic segmentation of large-scale outdoor scene.
- Yusheng Xu, Zhen Ye, R. H. X. L. K. L. X. L. L. H. X. T. U. S., 2020. LASDU: A Large-scale Aerial LiDAR Dataset for Semantic Labeling in Dense Urban Areas.
- Zolanvari, S., Ruano, S., Rana, A., Cummins, A., da Silva, R. E., Rahbar, M., Smolic, A., 2019. DublinCity: Annotated LiDAR point cloud and its applications. *arXiv preprint arXiv:1909.03613*.