

CENAGIS-ALS BENCHMARK - NEW PROPOSAL FOR DENSE ALS BENCHMARK BASED ON THE REVIEW OF DATASETS AND BENCHMARKS FOR 3D POINT CLOUD SEGMENTATION

P. Zachar, K. Bakula, W. Ostrowski

Warsaw University of Technology, Faculty of Geodesy and Cartography, Department of Photogrammetry, Remote Sensing and Spatial Information Systems, Warsaw, Poland (paulina.zachar.dokt, krzysztof.bakula, wojciech.ostrowski@pw.edu.pl)

ABSTRACT

Benchmarking is an essential tool for scientific and technological progress. This article reviews the benchmarks for 3D point cloud segmentation and classification. Based on the analysis of the articles and the knowledge gathered, it can be concluded that there has been an increase in the number of benchmarks, allowing to compare research results against specific performance metrics independently. However, benchmarks vary regarding the number of classes, spatial size, nomenclature, and class division. In this article, we introduce a new annotated 3D dataset - CENAGIS-ALS Benchmark. We propose a benchmark of highly dense lidar point clouds acquired by Leica CityMapper-2 for the Centre of Warsaw, Poland. The area covers 2 km², and the data has a density of 275 pts/m². The dataset consists of a number of classes that are distinguishable for this type of data. In addition to the basic classes, more specialized classes, important from the perspective of urban space, are also distinguished. Moreover, the division of classes consists of three levels of detail from coarse (e.g., a building) to refined elements (e.g., roofs, chimneys, and other rooftop objects). This benchmark can contribute to geospatial societies, considering the large spatial size of the study area with unified data quality and the higher number of classes with the hierarchical division compared to other benchmarking data.

KEYWORDS: point clouds, ALS, benchmark, segmentation, datasets

1. INTRODUCTION

The analysis of three-dimensional urban space is constantly popular as an area of interest for many research communities. This is due to the continuous development of cities, and new challenges in urban space management follow. According to the United Nations, by 2050 approx. 68% of people will live in cities (Ritchie & Roser, 2018). Thus, the need to acquire new spatial data and further interpret it to extract information and knowledge is apparent. With the development of light detection and ranging (lidar) technology in recent years, it is possible to collect dense point cloud data (over 100 pts/m²) for large areas (e.g., at the scale of an entire city). The increased interest in 3D data, as well as advances in the development of algorithms to work on such datasets, has created a demand for access to well-annotated ground-truth point clouds (Zolanvari et al., 2019). Automatic identification of 3D shapes and objects requires a three-dimensional, densely labeled point cloud that includes various urban elements (e.g., different types of roofs, building facades, light poles, trees, and vehicles). Despite developing various methods and approaches for object segmentation and detection, acquiring precise datasets and generating labels is still tedious and costly, as it amounts to manual work. Currently, semantic and instance segmentation of point clouds is increasingly being done using deep learning techniques since they offer relatively efficient and high-accuracy means to process massive datasets (Wu et al., 2015). Access to high-quality training datasets allows models to be trained and makes it possible to compare the performance of different networks, so the amount of benchmark datasets grows.

National mapping agencies in many European countries, i.e. Head Office of Geodesy and Cartography (GUGK) in Poland, acquire Airborne Laser Scanning (ALS) lidar data sets for entire countries. ALS data has a classification assigned, but the

differentiation of these classes is limited to an ASPRS standard, where several basic classes are distinguished, including ground, buildings, and low, medium, and high vegetation. The limitation of classes is also related to the density of this data. Although in cities, the density is at least a dozen points per square metres, this is sometimes not enough for some essential objects to be mapped. A similar pattern is noticeable in other European countries. For the purposes of the present study, in the Table 1. We compiled ALS data specification for several European countries.

The above review and report (Kakoulaki et al., 2021) provide information about the density, spatial coverage, and classes of non-commercial lidar data in Europe. Despite many applications where data with such characteristics are sufficient, it is noticeable that an increasing number of datasets are being acquired for cities with significantly higher densities. Previously, such data as for Dublin (Zolanvari et al., 2019) (the average density of 348.43 points/m²), for example, could be acquired scientifically for research purposes. It is now possible to acquire such data in production for entire cities.

The development of technology allows the acquisition of very dense data in which objects and elements are very well mapped. Thus, it is possible to detect classes such as walls, fences, roofs, equipment on roofs, stairs, building entrances, etc. Concerning attributes, in addition to class labels, there are sometimes data such as intensity and RGB. The hierarchical division into classes is very rare but can be helpful in some applications. The challenges mentioned earlier are being met by the latest lidar technologies, such as CityMapper-2, for which the typical point density is 60-70 points/m² on the ground for a single strip. Therefore, the benchmark presented in the paper shows a high density for the selected study area and is about 275 pts/m². Thanks to this density, it is possible to map more elements and further distinguishing them.

Table 1 ALS data specifications for selected European countries.

| Country | Area of entire country | Density | Number of classes | Compliant with the ASPRS classes |
|----------------|------------------------|---|-------------------|----------------------------------|
| Poland | yes | from 4 pts / m ² to at least 12 pts / m ² (in the case of cities) | 9 | yes |
| Croatia | yes | 4 pts/m ² in non-urban areas and 8 pts/m ² in urban areas | 10 | yes |
| Austria | yes | at least 8 pts/m ² | 13 | yes |
| Denmark | yes | at least 8 pts/m ² | 9 | yes |
| Sweden | yes | 1-2 pts/m ² | 6 | yes |
| Slovenia | yes | 5 pts/m ² (10 pts/m ² *) | 8 | no |
| Slovakia | yes | at least 5 pts/m ² | 12 | yes |
| Spain | yes | 0.5-4 pts/m ² | 10 | yes |
| Portugal | yes | 5 pts/m ² | 6 | yes |
| Greece | yes* | planned: 4 pts/m ² for the entire area and 10 pts/m ² for major urban areas | ND | yes* |
| Czech Republic | yes | 1 point/m ² | ND | ND |
| Italy | no | 0.5 - 5 pts/m ² | ND | ND |
| Lithuania | yes | 6,5 pts / m ² | 9 | yes |
| Finland | yes | 0.5 - 5 pts/m ² | 10 | yes |

*planned; ND – No Data.

To propose desirable classes for urban ALS datasets with such high density, a literature review was performed. With a focus on recent scientific articles and post-conference materials on benchmarks for point cloud segmentation, the adopted methodology of the study was to search by keywords related to the topic. Finally, twenty-six scientific articles were selected, for which the table included information such as authors, title, year of publication, data type (MLS, ALS, point clouds from DIM, etc.), number of classes, class names, and dataset size created. This list was used for further in-depth analysis. The most important information was collected and summarised in Table 2.

Most of the datasets for the 3D point cloud segmentation task involve MLS data (11), and of the 26, only 8 use ALS data. Other data types include TLS (1), UAV point clouds (5), and multispectral lidar (1). In general, the classes that stand out are sparse; often, it is a few or a dozen classes. There is no standardisation in class naming and division.

Table 2 Comparison of benchmarking datasets dealing with the segmentation of 3D point clouds.

| Name | Number of classes | Technology | Coverage [km ²] |
|-----------------------|-------------------|--------------|-----------------------------|
| SensatUrban | 13 | UAV Photogr. | 7,6 |
| Hessigheim 3D (H3D) | 11 | UAV lidar | 0,19 |
| Swiss3DCities | 5 | UAV Photogr. | 2,7 |
| OpenGF | 2 | ALS | 47,7 |
| DALES | 8 | ALS | 10 |
| LASDU | 5 | ALS | 1,02 |
| Campus3D | 24 | UAV Photogr. | 1,58 |
| DublinCity | 13 | ALS | 2 |
| Vaihingen (ISPRS) | 9 | ALS | 0,1 |
| STPLS3D-Real | 6 | UAV Photogr. | 1,27 |
| SemanticKITTI | 25 | MLS | 39,2 |
| Paris-Lille3D | 9 | MLS | 1,94 |
| Toronto3D | 8 | MLS | 1 |
| Semantic3D | 8 | TLS | 0 |
| A2D2 | 38 | MLS | 0 |
| Waymo dataset | 4 | MLS | 0 |
| Paris-CARLA-3D | 23 | MLS | 0,55 |
| SemanticPOSS | 14 | MLS | 0 |
| CSPC-dataset | 6 | MLS | 0 |
| TerraMobilita/iQmulus | 8 | MLS | 10 |
| Paris-rue-Madame | 17 | MLS | 0,16 |
| Oakland | 5 | MLS | 1,5 |
| YTU | 45 | UAV Photogr. | 2,2 |

The databases obtained by Mobile Laser Scanning (MLS) do not include information about the roofs of buildings or other elements that are not achieved from the ground, making the urban scene incomplete. Some of them contains large number of classes, for instance SemanticKITTI (Behley et al., 2019), YTU Bayrak et al. (2023), A2D2 (Geyer et al., 2020), Paris-CARLA-3D (Deschaud et al., 2021), SemanticPOSS (Pan et al., 2020), Paris-rue-Madame (Serna et al., 2014), Paris-Lille3D (Roynard et al., 2018). Other examples of outdoor terrestrial benchmarks are popular in the context of autonomous driving, such as Toronto3D (Tan et al., 2020), Waymo dataset (Sun et al., 2020), CSPC-dataset (Tong et al., 2020), Oakland (Munoz et al., 2009), TerraMobilita/iQmulus (Vallet et al., 2015). The Semantic3D dataset (Hackel et al., 2017) was generated from Terrestrial Laser Scanning (TLS). However, it only covers a small portion of a city with a limited number of elements (only eight classes).

Meanwhile, the second category contains aerial and UAV data. The H3D dataset (Kölle et al., 2021) consists of a high-density lidar point cloud of approximately 800 points/m² and includes eleven classes; the spatial size is relatively small (0.19 km²). This benchmark is the first ultra-high resolution 3D dataset acquired from a lidar system and cameras integrated into the same UAV platform. The DALES (Varney et al., 2020) dataset is slightly

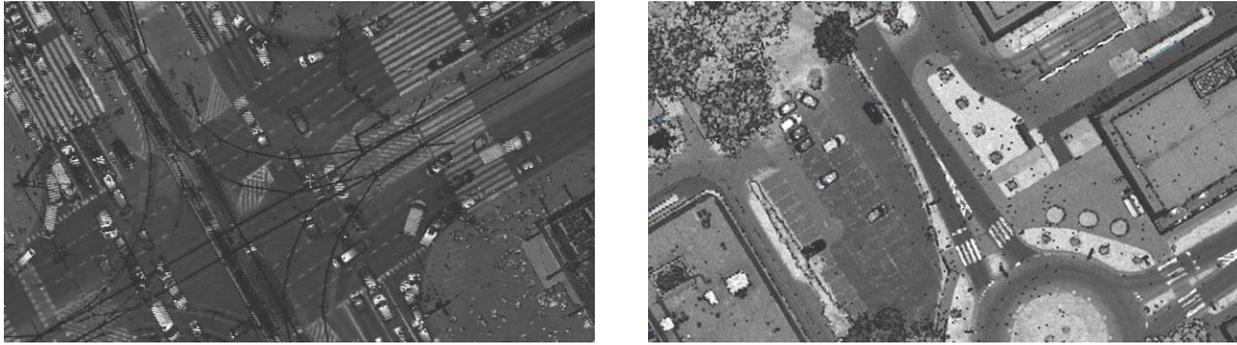


Figure 3 Warsaw ALS CityMapper-2: intensity rasters present this sensor's point cloud details.

Table 3 Division of classes in CENAGIS-ALS Benchmark.

| Level 1 Class | | Level 2 Class | | Level 3 Class | | | |
|---------------|---------------------------|---------------|-----------------------|---------------|---------------------|--|--|
| Code | Name | Code | Name | Code | Name | | |
| 6 | Building | 20 | Façade | | | | |
| | | 21 | Roof | | | | |
| | | 22 | Chimney | | | | |
| | | 23 | Other roof structures | | | | |
| | | 24 | Stairs | | | | |
| | | 25 | Balcony | | | | |
| 60 | Vegetation | 3 | Low vegetation | | | | |
| | | 4 | Shrub | | | | |
| | | 5 | Tree | | | | |
| 2 | Ground | 27 | Sidewalk | | | | |
| | | 28 | Bikeroad | | | | |
| | | 30 | Grass | | | | |
| | | 60 | Street | 33 | Speed bumper | | |
| | | | | 34 | Dashed/Solid line | | |
| | | | | 35 | Zebra crossing | | |
| | | | | 36 | Road | | |
| | | | | 37 | Parking | | |
| | | 10 | Railway | 13 | Wire | | |
| | | | | 14 | Train pole | | |
| | | | | 16 | Rail-track | | |
| 31 | Other impervious surfaces | | | | | | |
| 32 | Other-ground | | | | | | |
| 9 | Water | | | | | | |
| 19 | Vehicle | | | | | | |
| 17 | Bridge | | | | | | |
| 61 | Other | 24 | Stairs | | | | |
| | | 26 | Underground entrance | | | | |
| | | 62 | Pole-like | 38 | Light Pole | | |
| | | | | 39 | Power Pole | | |
| | | | | 40 | Traffic Signal | | |
| | | | | 41 | Other Pole-like | | |
| | | 63 | Urban furniture | 42 | Billboards | | |
| | | | | 43 | Announcement pole | | |
| | | | | 44 | City bike station | | |
| | | | | 45 | Fountains | | |
| | | | | 46 | Benches | | |
| | | | | 47 | Playgrounds | | |
| | | | | 48 | Monuments | | |
| | | | | 49 | Postboxes | | |
| | | 64 | Fences/hedges | 50 | Sidebars | | |
| | | | | 51 | Fence (gate) | | |
| | | | | 52 | Wall-like | | |
| | | | | 53 | Other fences/hedges | | |
| | | 54 | Other structures | 55 | Bus-stop | | |
| 56 | Parkingmeter | | | | | | |
| 57 | Vertical surfaces | | | | | | |
| 58 | Shelter | | | | | | |
| 59 | Kiosk | | | | | | |
| | | | | | | | |
| 0 | Never classified | | | | | | |
| 1 | Unassigned | | | | | | |
| 7 | Noise | | | | | | |

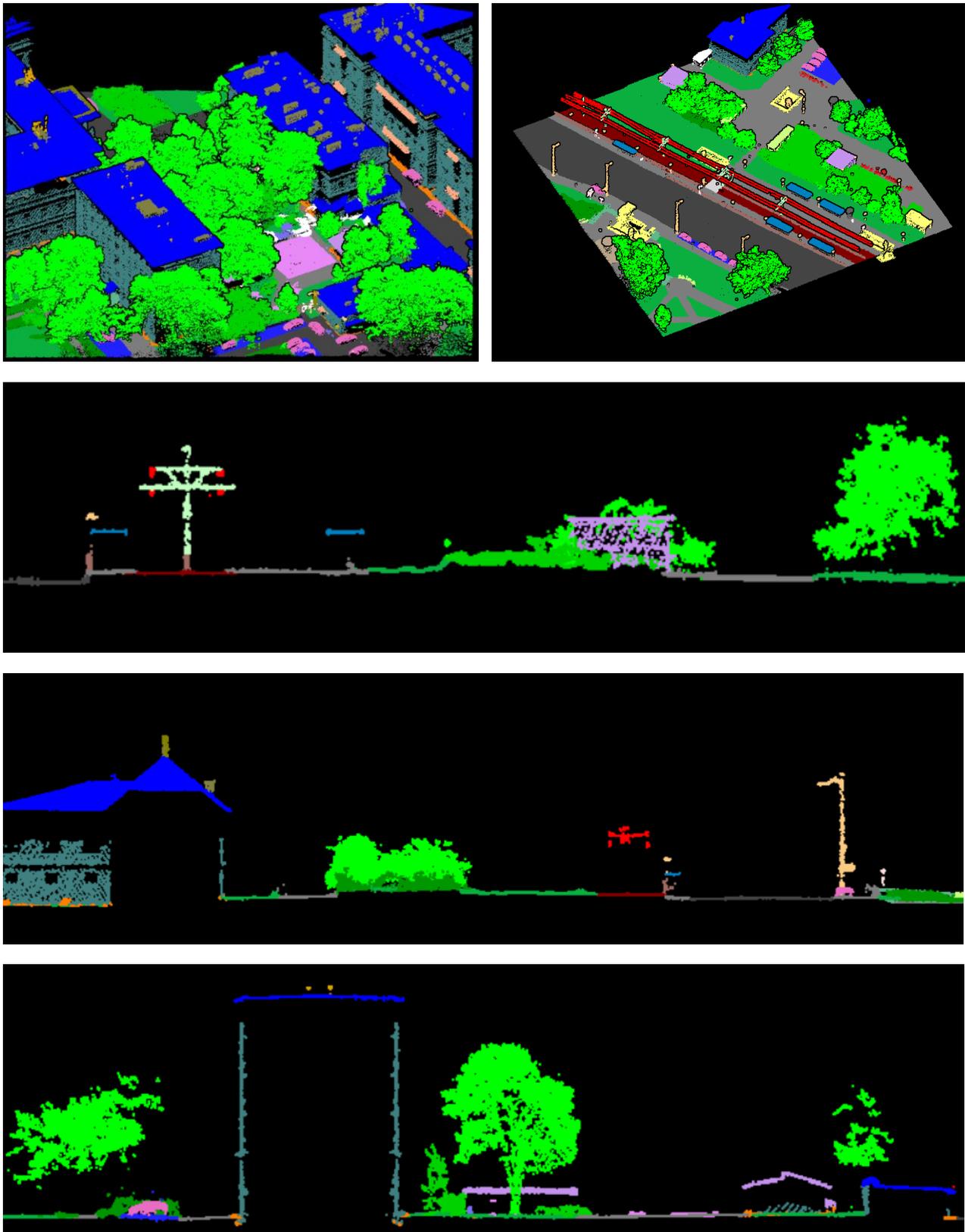


Figure 4 Some sample images from the classified point cloud of CENAGIS-ALS Benchmark showing the selected classes: light green - trees, green – shrub, dark green – low vegetation, dark blue – roofs, khaki – chimney, bright orange – other roof structures, yellow – street lamp, red – wire, burgundy - rail-track, pink – vehicles, celadon – train pole, turquoise – walls, lemon – underground entrance, salmon – balcony, blue - bus-stop, bright grey – sidewalk, dark grey – road, purple – shelter, light brown - fence.

Users who use our benchmark can use the accuracy rating. They can upload their classification results to the server, and they will get a confusion matrix that tells them about the accuracy against ground-truth data. In order to obtain performance metrics, the confusion matrix is calculated, showing the percentage of correctly and incorrectly classified points for the study area. The diagonal of the matrix shows the percentage of points classified correctly, in addition to which the user receives a feedback message about the accuracy, which is the average of all classes.

3. CONCLUSIONS

The proposed CENAGIS-ALS Benchmark offers a comprehensive point cloud data collection for Warsaw Downtown. It consists of a wide range of urban elements and objects, which makes it a valuable resource for various applications related to urban planning and analysis. With its high point density, accurate classification levels, and partitioning for efficient utilisation, this dataset serves as valuable, well-annotated ground-truth data for benchmarking and validating algorithms and models. The availability of such data supports various applications, but mainly, such data is used to classify 3D objects using state-of-the-art models, particularly for model learning or transfer learning.

The data will be provided using the CENAGIS infrastructure (the "Center for Scientific Geospatial Analyses and Satellite Computations") implemented at the Faculty of Geodesy and Cartography, Warsaw University of Technology.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the OPEGIEKA Sp.z o.o. company (former CENAGIS project partner) acquisition and providing the Leica CityMapper-2 dataset for the Centre of Warsaw, Poland.

REFERENCES

- Bayrak, O., Remondino, F., Uzar, M., 2023. A new dataset and methodology for urban-scale 3d point cloud classification. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Vol. XLVIII-1/W3-2023, in press
- 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. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9297-9307).
- Chen, M., Hu, Q., Hugues, T., Feng, A., Hou, Y., McCullough, K., & Soibelman, L. (2022). STPLS3D: A Large-Scale Synthetic and Real Aerial Photogrammetry 3D Point Cloud Dataset. *arXiv preprint*.
- Can, G., Mantegazza, D., Abbate, G., Chappuis, S., & Giusti, A. (2021). Semantic segmentation on Swiss3DCities: A benchmark study on aerial photogrammetric 3D pointcloud dataset. *Pattern Recognition Letters*, 150, 108-114.
- Deschaut, J. E., Duque, D., Richa, J. P., Velasco-Forero, S., Marcotegui, B., & Goulette, F. (2021). Paris-CARLA-3D: A real and synthetic outdoor point cloud dataset for challenging tasks in 3D mapping. *Remote Sensing*, 13(22), 4713.
- Geyer, J., Kassahun, Y., Mahmudi, M., Ricou, X., Durgesh, R., Chung, A. S., ... & Schuberth, P. (2020). A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*.
- 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*.
- Hebel, M., Arens, M., & Stilla, U. (2013). Change detection in urban areas by object-based analysis and on-the-fly comparison of multi-view ALS data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86, 52-64.
- 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. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4977-4987).
- Kakoulaki, G., Martinez, A., & Florio, P. (2021). Non-Commercial Light Detection and Ranging (LiDAR) Data in Europe. *Publications Office of the European Union: Luxemburg*.
- Kölle, M., Laupheimer, D., Schmohl, S., Haala, N., Rottensteiner, F., Wegner, J. D., & Ledoux, H. (2021). The Hessigheim 3D (H3D) benchmark on semantic segmentation of high-resolution 3D point clouds and textured meshes from UAV LiDAR and Multi-View-Stereo. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 1, 100001.
- Li, X., Li, C., Tong, Z., Lim, A., Yuan, J., Wu, Y., ... & Huang, R. (2020, October). Campus3d: A photogrammetry point cloud benchmark for hierarchical understanding of outdoor scene. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 238-246).
- Munoz, D., Bagnell, J. A., Vandapel, N., & Hebert, M. (2009). Contextual classification with functional max-margin markov networks. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*.
- Niemeyer, J., Rottensteiner, F., & Soergel, U. (2014). Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing*, 87,.
- Pan, Y., Gao, B., Mei, J., Geng, S., Li, C., & Zhao, H. (2020). SemanticPOSS: A point cloud dataset with large quantity of dynamic instances. In *2020 IEEE Intelligent Vehicles Symposium (IV)* (pp. 687-693).
- Qin, N., Tan, W., Ma, L., Zhang, D., & Li, J. (2021). Opengf: An ultra-large-scale ground filtering dataset built upon open als point clouds around the world. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1082-1091).
- Ritchie H. Roser M. (2018). Urbanization, <https://ourworldindata.org/urbanization> (20.09.2023)
- 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).
- 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. In 4th international conference on pattern recognition, applications and methods ICPRAM.
- Sun, P., Kretschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., ... & Angelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2446-2454).
- Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., ... & Li, J. (2020). Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 202-203).
- Tong, G., Li, Y., Chen, D., Sun, Q., Cao, W., & Xiang, G. (2020). CSPC-dataset: New LiDAR point cloud dataset and benchmark for large-scale scene semantic segmentation. *IEEE access*, 8, 87695-87718.
- Vallet, B., Brédif, M., Serna, A., Marcotegui, B., & Paparoditis, N. (2015). TerraMobilita/iQmulus urban point cloud analysis benchmark. *Computers & Graphics*, 49, 126-133.
- Varney, N., Asari, V., & Graehling, Q. (2020). DALES: A large-scale aerial LiDAR data set for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern rec.*
- Vosselman, G., Coenen, M., & Rottensteiner, F. (2017). Contextual segment-based classification of airborne laser scanner data. *ISPRS journal of photogrammetry and remote sensing*, 128, 354-371.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1912-1920).
- Xu, Y., Du, B., Zhang, L., Cerra, D., Pato, M., Carmona, E., ... & Le Saux, B. (2019). Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6), 1709-1724.
- Ye, Z., Xu, Y., Huang, R., Tong, X., Li, X., Liu, X., ... & Stilla, U. (2020). Lasdu: A large-scale aerial lidar dataset for semantic labeling in dense urban areas. *ISPRS International Journal of Geo-Information*.
- Zolanvari, S. M., 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*.