The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W3-2023
2nd GEOBENCH Workshop on Evaluation and BENCHmarking of Sensors, Systems and
GEOspatial Data in Photogrammetry and Remote Sensing, 23–24 October 2023, Krakow, Poland

A NEW DATASET AND METHODOLOGY FOR URBAN-SCALE 3D POINT CLOUD CLASSIFICATION

O. C. Bayrak [1], F. Remondino [2], M. Uzar [1]

[1] Dept. of Geomatics Engineering, Faculty of Civil Engineering, Yildiz Technical University, Istanbul, Turkey
Email: (onurcb, auzar)@yildiz.edu.tr

[2] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy – Email: remondino@fbk.eu

Commission II

KEY WORDS: point cloud, dataset, classification, semantic segmentation, deep learning, benchmarking

ABSTRACT:
Urban landscapes are characterized by a multitude of diverse objects, each bearing unique significance in urban management and development. With the rapid evolution and deployment of Unmanned Aerial Vehicle (UAV) technologies, the 3D surveying of urban areas through high resolution point clouds and orthoimages has become more feasible. This technological leap enhances our capacity to comprehensively capture and analyze urban spaces. This contribution introduces a new urban dataset, called YTU3D, which covers an area of approximately 2km^2 and encompasses 45 distinct classes. Notably, YTU3D exceeds the class diversity of existing datasets, thereby enhancing its suitability for detailed urban analysis tasks. The paper presents also the application of three popular deep learning methods in the context of 3D semantic segmentation, along with a multi-level multi-resolution (MLMR) integration. Significantly, our work marks the first application of deep learning with MLMR in the literature and shows that a MLMR approach can improve the classification accuracy. The YTU3D dataset and research findings are publicly available at https://github.com/3DOM-FBK/YTU3D.

1. INTRODUCTION

Fully semantically enriched 3D point clouds play a significant role in urban planning, documentation, and management. Semantic segmentation of large-scale point clouds is the process to assign each point a semantic label, such as road, tree, roof, sidewalk, etc. Machine and deep learning algorithms offer crucial methodologies for understanding 3D point clouds at various scales (Matrone et al., 2020a; Xie et al., 2020; Grilli et al., 2021). However, the semantic enrichment of large-scale point clouds remains an open research challenge. Available urban datasets generally lack class diversity, impeding the representation and understanding of detailed urban areas. Consequently, imbalanced classes are often present in the available datasets/benchmarks, and most learning methods struggle to accurately predict such imbalanced classes. Numerous techniques for 3D semantic segmentation, with deep learning standing out, have gained prominence within the realms of computer vision and photogrammetry (Guo et al., 2020). However, annotating large urban scenarios is time-consuming and prone to errors, resulting in training data riddled with issues that negatively affect learning methods. Deep learning methods heavily rely on the quality and quantity of provided training data.

Despite the growing availability of different point cloud benchmarks/datasets (Niemeyer et al., 2014; Matrone et al., 2020b; Kölle et al., 2021; Hu et al., 2021; Li et al., 2023), 3D semantic segmentation still faces major challenges such as low accuracy, dealing with imbalanced classes, misclassifications, generalization, upscaling, etc.

To address these challenges, authors have proposed various solutions, such as neuro-symbolic logic rules (Grilli et al., 2023), data augmentation (Achlioptas et al., 2018; Chen et al., 2020), class weighting (Lin et al., 2017; Griffiths and Boehm, 2019), oversampling/undersampling techniques (Lin and Nguyen, 2020; Ren and Xia, 2023), or a multi-level multi-resolution approach (Teruggi et al., 2020).

1.1 Paper motivations and aims

Urban areas are complex environments featuring a multitude of objects, changing scales, irregular point distribution, etc. A proper 3D semantic mapping of urban areas holds significant importance for city management and monitoring operations. However, existing methods for 3D classification fall short on the abovementioned challenges also due to the lack of proper datasets for methods development and evaluation. There is a need for (i) large-scale point cloud datasets that encompass a wide class diversity, including imbalanced classes, (ii) a reduction in misclassifications through a better understanding of complex urban areas, allowing for detailed 3D scene representation and (iii) very accurate annotation data.

Therefore, the objectives of this paper are:

- to introduce a new urban dataset, called YTU3D, with rich semantic annotations;
- to evaluate various deep learning methods to assess issues related to generalization and performance;
- to propose a novel methodology for classifying large datasets while minimizing misclassifications using a hierarchical classification approach.

The contribution of this study lies in the creation of a unique 3D photogrammetric point cloud dataset, which contains the most extensive range of semantic classes specific to urban areas found in the literature. Additionally, for the first time, this study pioneers the application of a multi-level multi-resolution (MLMR) classification approach coupled to deep learning in the context of urban scenes. We comparatively assess three popular 3D semantic segmentation methods (Point Transformer (Zhao et al., 2018), KPConv (Thomas et al., 2019), and RandLA-Net (Hu et al., 2020)) and their integration with MLMR approach, to provide a benchmark and investigate the effect of hierarchical 3D classification for large urban-scale scenarios.

1.2 Related works

Urban-level airborne 3D point cloud datasets can be categorized based on the inclusion of color information, number of classes or used sensor (Table 1). The ISPRS - Vaihingen (Niemeyer et al., 2014), DublinCity (Zolanvari et al., 2019), DALES (Varney et al., 2020), and LASDU (Ye et al., 2020) datasets, despite their large scale, deny the use of color-related features. In contrast, the Campus3D (Li et al., 2020), Swiss3DCities (Can et al., 2021),

Name	Reference	Classes	Points (mil)	Spatial Size (m^2)	RGB	Sensor
ISPRS	Niemeyer et al. (2014)	9	1.2	1.6×10^5	No	ALS / LiDAR
DublinCity	Zolanvari et al. (2019)	13	260	2×10^6	No	ALS / LiDAR
IEEE-GRSS	Bosch et al. (2019)	5	102	34×10^6	No	ALS / LiDAR
DALES	Varney et al. (2020)	8 (9)	505	10×10^6	No	ALS / LiDAR
LASDU	Ye et al. (2020)	5	3.12	1.02×10^6	No	ALS / LiDAR
Campus3D	Li et al. (2020)	24	937	1.58×10^6	Yes	UAV Photo
Swiss3DCities	Can et al. (2021)	5	226	2.7×10^6	Yes	UAV Photo
Hessigheim 3D	Kölle et al. (2021)	11	73	8×10^4	Yes	UAV LiDAR
OpenGF	Qin et al., (2021)	2	500	47×10^6	No	ALS
SensatUrban	Hu et al. (2022)	13 (31)	2847	7.64×10^6	Yes	UAV Photo
STPLS3D	Chen et al. (2022)	6 (18)	-	6×10^6	Yes	Synthetic + UAV Photo
HRHD-HK	Li et al. (2023)	7	273	9×10^6	Yes	UAV Photo
YTU3D (our)	**-**	**45**	**1700**	**2×10^6**	**Yes**	**UAV Photo**

Table 1. A summary of some representative aerial datasets for urban-scale 3D point cloud classification. The number of classes is the one used for evaluation while in brackets we report the annotated classes. ALS: Airborne Laser Scanning, Photo: Photogrammetry.

and Hessigheim3D (Kölle et al., 2021) datasets, which do contain color information, cover smaller areas and have few class labels. STPLS3D (Chen et al., 2022) and HRHD-HK (Li et al., 2023) were acquired on large urban areas but they exhibit a small number of classes. Furthermore, since the STPLS3D dataset is predominantly synthetic, it contains a lack of noise and discontinuities, limiting its usability (and generalization) to learn complex real-life scenarios. The HRHD-HK dataset, despite encompassing a vast area, does not exhibit a significant contrast compared to existing datasets due to its limited number of classes. The proposed YTU3D dataset is similar to Zachar et al. (2023), being a large-scale photogrammetric acquisition with more than 30 classes.

2. MLMR CLASSIFICATION

The multi-level multi-resolution (MLMR) methodology follows a hierarchical approach to classify 3D data based on diverse geometric resolutions, thereby enhancing the learning process and optimizing classification outcomes (Teruggi et al., 2020; Russo et al., 2021; Mazzacca et al., 2022). By adopting this approach also for urban-scale point cloud:

- it becomes possible to alleviate classification errors that may arise when dealing with visually similar objects belonging to distinct categories, due to the consideration of their geometric relations;
- the use of sub-sampled points facilitates the extraction of primary objects while reducing processing time, allowing a more detailed classification at each level;
- complex structures and large scale variations could be better tackled;
- model's generalization capability can be boosted by allowing for the identification and extraction of small-scale objects within the primary classes, particularly in scenarios characterized by an imbalanced distribution of data;
- class standardization could be supported, having the primarily levels alike to various users and the final/sub-levels specific for other users.

We propose the use of MLMR, in combination with a deep learning network, for the classification of large-scale high-resolution point clouds over urban scenarios. For the MLMR methodology (Figure 2), the full-resolution point cloud (Level-5) is subsampled at 20 cm (Level-1), 10 cm (Level-2), 6 cm (Level-3), and 3 cm (Level-4). At Level-1, models are trained for the initial target classes: ground and non-ground. Classification results from this process are then interpolated back to a higher resolution, i.e., Level-2, using a nearest-neighbour algorithm. A new classification stage is then applied to identify new classes in Level-2, including impervious surface, low vegetation, and soil/gravel classes for ground, as well as building, high vegetation, and non-building classes for non-ground. These operations are subsequently repeated until the original resolution is reached (Level-5) and all subclasses are identified.

At Level-3, the focus shifted to the identification of classes such as stairs, tennis courts/football pitches, sidewalks, streets/roads, and parking lots for impervious surfaces, and soil/gravel, stone, and disorganized regions for soil/gravel classes. Meanwhile, classes like roof and façade are identified for buildings, shrubs and trees for high vegetation, and urban furniture, people, pets, and vehicle classes for non-building categories, all within Level-3. Level-4 continued with the identification of subclasses, including façade (Window, pipeline, other accessories, ventilation, and façade surface), roof (chimney, window, solar panel, ventilation, pipeline, other accessories, and roof surface), urban furniture (wall, fence, playground, traffic signage, garbage box, lamp, street separator, tent, and pole), and vehicle classes (truck, van, motorbike/bicycle, bus, car, work machine), respectively. Finally, for the original resolution, dome, tile, and industrial roof surface types are classified.

3. YTU3D DATASET

The proposed UAV-photogrammetry derived dataset covers an area of approximately 2 km^2 over the Davutpasa Campus of Yildiz Technical University (YTU), Turkey (Figure 1a). A DJI Matrice 300 RTK was used for the flights, acquiring nadiral images with 75-75 along/across-track overlap. After the aerial triangulation, a dense point cloud of the area was produced in Pix4D software with an average point density of ca 1000 points/m2 (Figure 1b). To represent the complex structures and increase 3D scene understanding, we annotated 45 classes (Figure 1c and Table 3) in CloudCompare (Girardeau-Montaut, 2016) following a MLMR idea, as illustrated in Figure 2. From the 43 regions of the dataset, 20 are used for training, 12 for validation and 8 for testing. More details and results are available at https://github.com/3DOM-FBK/YTU3D.

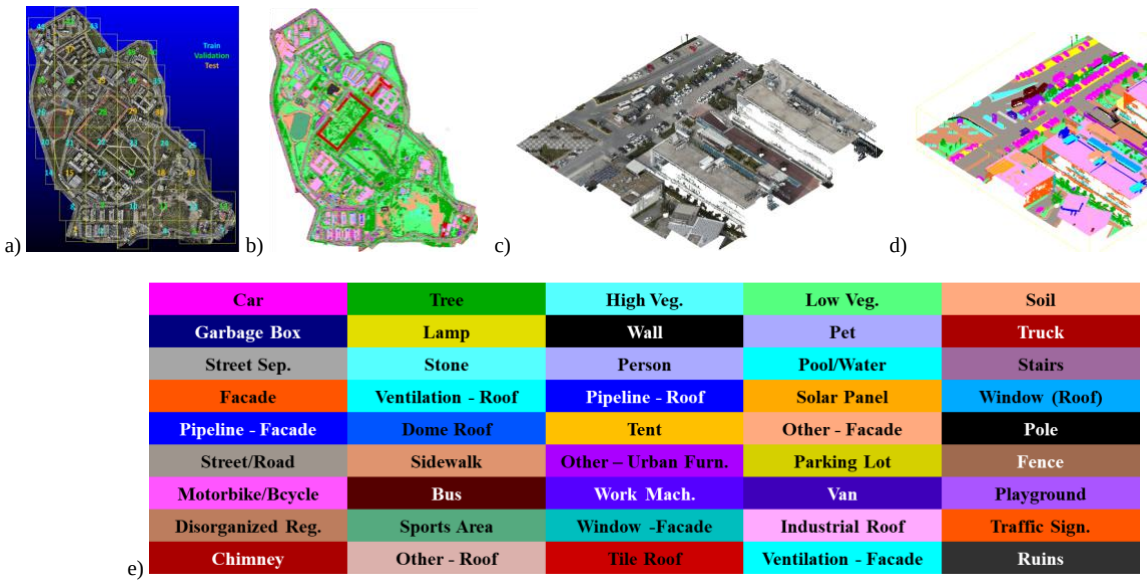

Figure 1: The proposed YTU3D dataset divided in 43 regions, each on of ca 210 x 210 sqm (a); orthoviews of the ground truth (GT) annotations on the entire are (b); sample region (c) and corresponding annotationed point cloud (d); 45 classes (e).

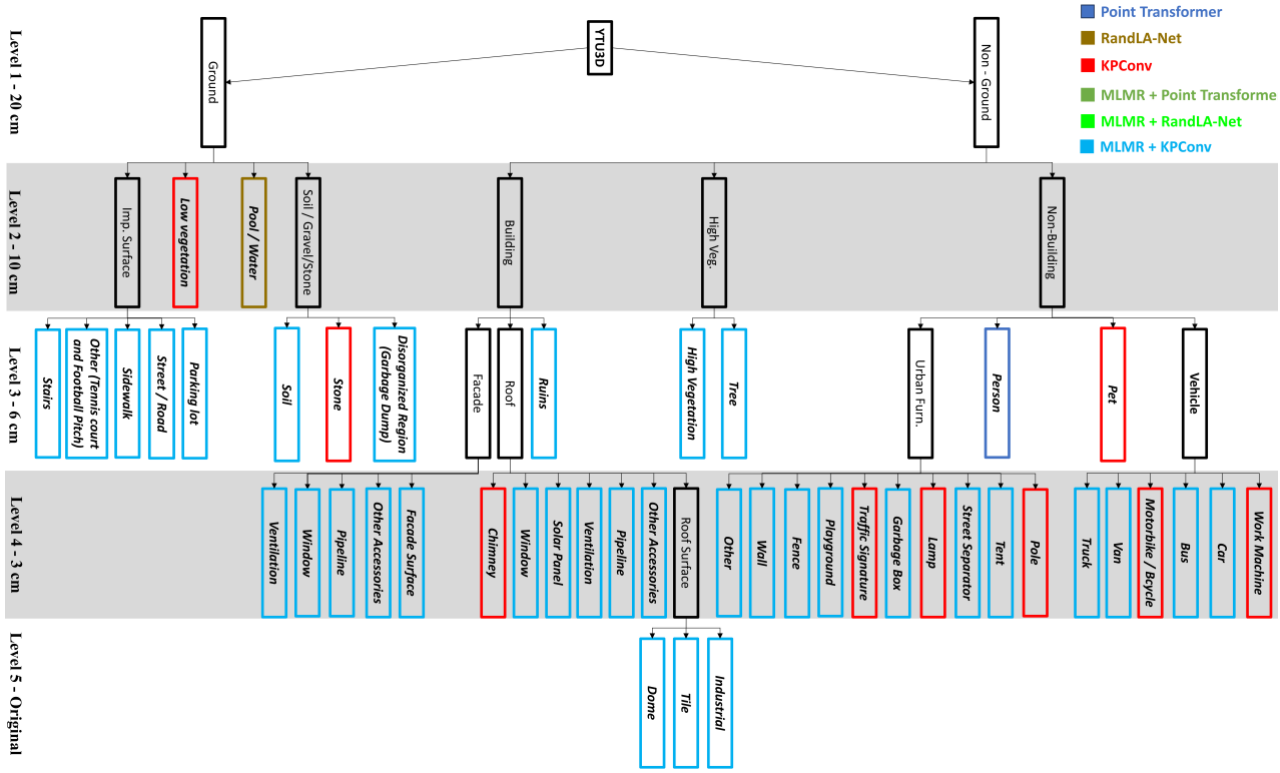

Figure 2: The MLMR hierarchical structure applied to the YTU3D dataset, with the different resolution levels (5) and classes (45). Labelled classes are shown in italic bold and classes which will be further sub-classified are shown with a black box. Classes are coloured according to the deep learning method that gives the most accurate results (Section 3).

4. EXPERIMENTS AND RESULTS

Using the YTU3D dataset, we evaluated the performance of some state-of-the-art algorithms and provide a comprehensive analysis of the results. In particular, we utilized Point Transformer (Zhao et al., 2018), a Kernel point convolution (KPConv) (Thomas et al., 2019), and RandLA-Net (Hu et al., 2020), which are commonly used in 3D semantic segmentation benchmarks. All tests are run on a GPU Nvidia RTX 4090, 64 GB RAM, processor 13th Gen Intel(R) Core(TM) CPU i9-13900K @ 5.20 GHz. Our findings indicate that the MLMR approach played a critical role in influencing the performance of these methods (Table 2 and Table 3). MLMR + KPConv demonstrated the highest mIoU score, showcasing the effectiveness of combining MLMR with KPConv to enhance segmentation accuracy. KPConv performed impressively in isolation, highlighting its intrinsic strength in spatial data analysis. We observed that when employing MLMR + KPConv, the accuracy achieved for specific classes such as Stone, Chimney, Traffic Signature, Lamp, Pole, Motorbike/Bicycle, Pet, and Work Machine was slightly lower compared to using KPConv in isolation (Figure 4).

Method	***mIoU***
Point Transformer	0.43
RandLA-Net	0.49
KPConv	**0.58**
MLMR + Point Transformer	0.49
MLMR + RandLA-Net	0.36
MLMR + KPConv	**0.62**

Table 2. Quantitative results on test set of the YTU3D dataset.

The comparative analysis suggests that the addition of MLMR may not always yield improvements for all classes, and its impact can vary depending on the specific class characteristics and dataset properties. While the MLMR + KPConv combination may excel in certain aspects of the segmentation task, it may encounter challenges in accurately identifying these particular classes.

RandLA-Net, when combined with MLMR, displayed a slightly lower mIoU score, indicating a decrease in accuracy compared to its standalone performance. This result suggests that the incorporation of MLMR may not be as beneficial for RandLA-Net in this context. Our analysis reveals that RandLA-Net has exhibited superior accuracy in predicting the class "Pool/Water" when compared to other methods.

Point Transformer exhibited a similar mIoU score to its MLMR-enhanced counterpart, MLMR + Point Transformer, implying that the Point Transformer method experienced only marginal improvements when combined with MLMR. Our observations indicate that Point Transformer has demonstrated superior predictive performance specifically for the "Person" class when compared to other methods. Interestingly, when integrated with MLMR (Multi-Level Multi-Resolution), Point Transformer did not exhibit significant improvements in predicting this particular class. These findings contribute significantly to our understanding of the effectiveness of various methods and their compatibility with the MLMR approach in the context of spatial data analysis. It is worth noting that these observations have important implications for researchers and practitioners seeking to optimize segmentation performance in urban environments.

Sample results performed by MLMR + KPConv on the test set of YTU3D are shown in Figure 3 and 4. Figure 5 reports further qualitative results of the various tested methods on YTU3D.

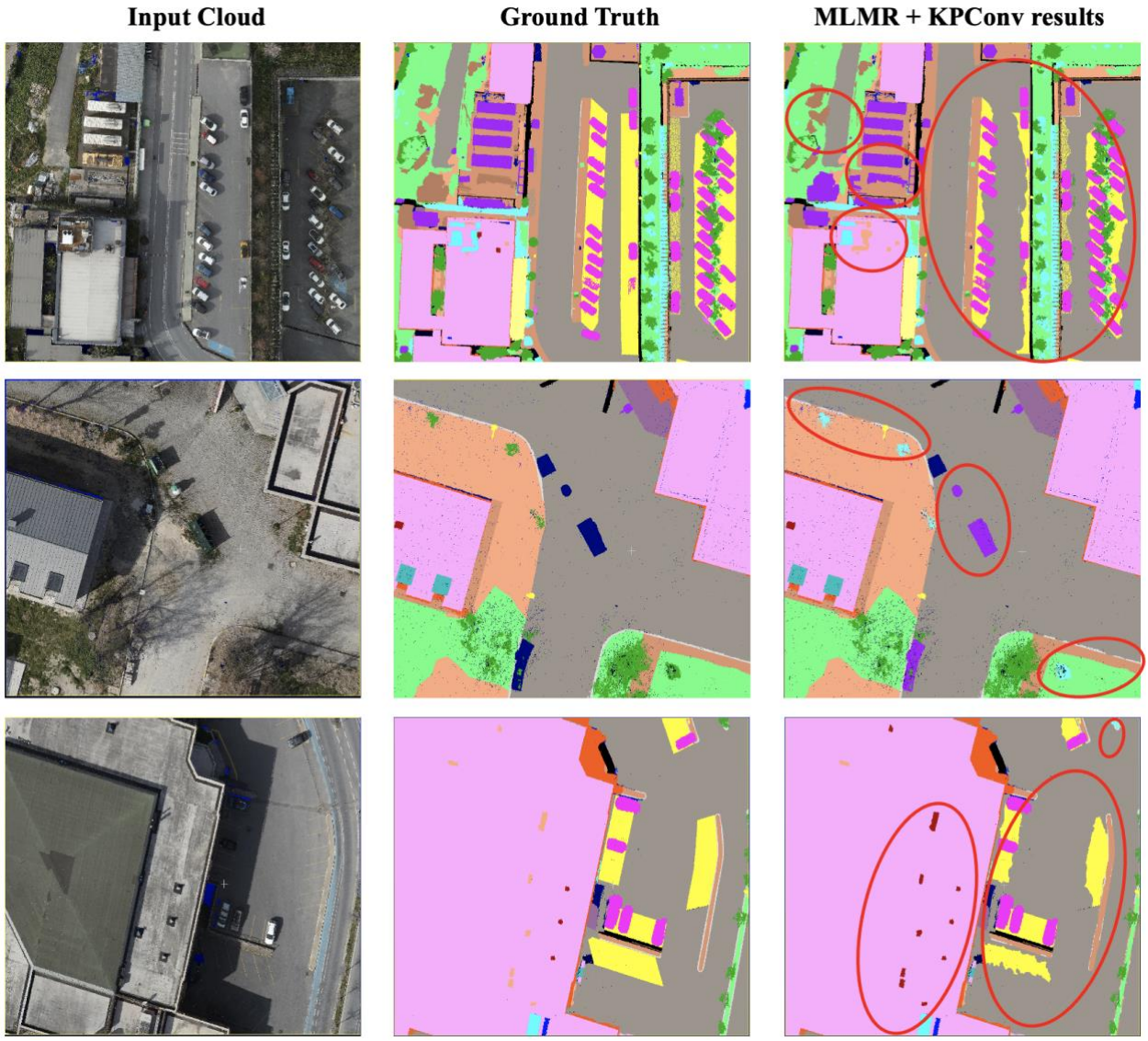

Figure 3: Qualitative results of MLMR + KPConv. Red boxes demonstrate inaccurate predictions with respect GT annotation. Classes colours as in Figure 1.

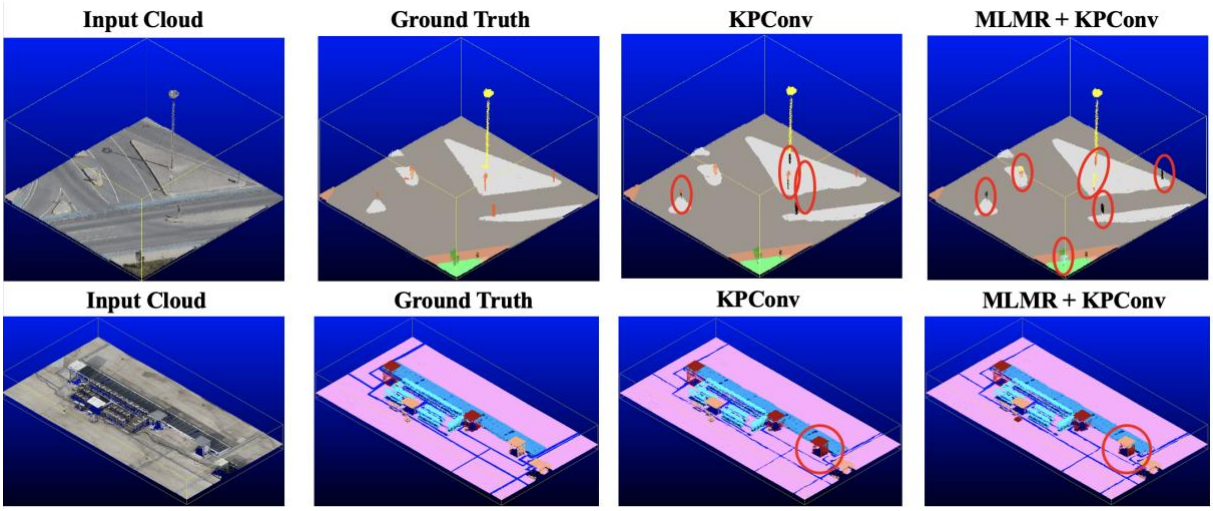

Figure 4: Qualitative results of MLMR + KPConv on vertical objects (top) and small under-represented objects on building roofs (bottom). Red boxes show inaccurate predictions with respect to GT annotation. Classes colours as in Figure 1.

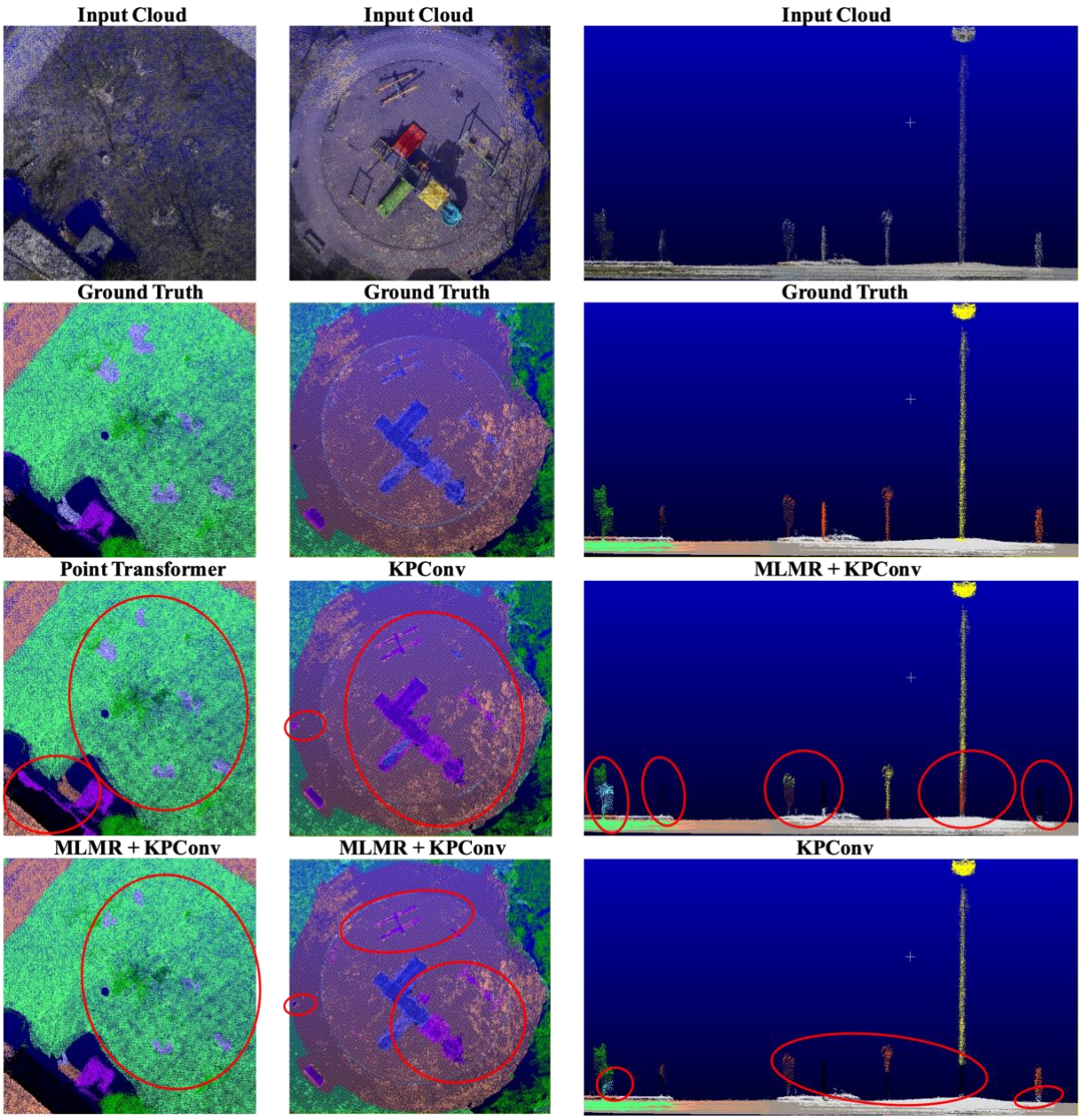

Figure 5: Qualitative results of the different methods: red boxes show inaccurate predictions with respect to GT.

https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-1-2023

5. CONCLUSIONS

In this study, we introduced (i) a new benchmark dataset for 3D segmentation of urban areas and (ii) evaluate the performance of three popular deep learning methods coupled to a hierarchical MLMR approach (Table 3). The shared data and research findings are publicly available at https://github.com/3DOM-FBK/YTU3D. While the dataset's spatial coverage may not be as extensive as other current datasets, YTU3D features the richest variety of semantic classes found in the literature to date. The primary distinction from existing datasets lies in the expanded sub-classes within vehicle, building, and urban object categories. The detailed characterization of these objects, which define urban areas across various regions, enhances the practical utility of 3D classification models.

The presented results indicate that the MLMR approach significantly improved the overall performance of classification models. However, we observed that this approach did not yield performance improvements for imbalanced classes, specifically 2-dimensional objects such as poles, lamps, and chimneys, as well as moving object classes like pets and people. Furthermore, our findings suggest that the MLMR approach may not be suitable for all methods, as it resulted in performance degradation for RandLA-Net. Nevertheless, we believe that a hierarchical approach could bring many advantages to large-scale scenarios and could play a fundamental role toward standards in classes.

Future studies may benefit from exploring a broader range of subsampling resolutions and examining the performance of different methods under varying conditions.

ACKNOWLEDGEMENTS

This work was supported by The Scientific and Technological Research Council of Türkiye (TUBITAK) [2214-A].

REFERENCES

Achlioptas, P., Diamanti, O., Mitliagkas, I., Guibas, L., 2018. Learning representations and generative models for 3d point clouds. Proc. *Int. Conference on Machine Learning*, pp. 40-49.

Bosch, M., Foster, K., Christie, G., Wang, S., Hager, G.D., Brown, M., 2019. Semantic Stereo for Incidental Satellite Images. *arXiv*:1811.08739.

Can, G., Mantegazza, D., Abbate, G., Chappuis, S. and Giusti, A., 2021. Semantic segmentation on Swiss3DCities: A benchmark study on aerial photogrammetric 3D pointcloud dataset. *Pattern Recognition Letters*, 150, pp.108-114.

Chen, Y., Hu, V.T., Gavves, E., Mensink, T., Mettes, P., Yang, P., Snoek, C.G., 2020. Pointmixup: Augmentation for point clouds. Proc. *ECCV*, pp. 330-345.

Chen, M., Hu, Q., Yu, Z., Thomas, H., Feng, A., Hou, Y., McCullough, K., Ren, F. and Soibelman, L., 2022. Stpls3d: A large-scale synthetic and real aerial photogrammetry 3d point cloud dataset. *arXiv*:2203.09065.

Girardeau-Montaut, D., 2016. *CloudCompare*. EDF R&D Telecom ParisTech, 11.

Griffiths, D., Boehm, J., 2019. Weighted point cloud augmentation for neural network training data class-imbalance. *arXiv*:1904.04094.

Grilli, E., Poux, F., Remondino, F., 2021. Unsupervised object-based clustering in support of supervised point-based 3D point cloud classification. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*., 43, 471-478.

Grilli, E., Daniele, A., Bassier, M., Remondino, F., Serafini, L., 2023. Knowledge Enhanced Neural Networks for Point Cloud Semantic Segmentation. *Remote Sensing*. 2023; 15(10):2590.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N. and Markham, A., 2020. Randla-net: Efficient semantic segmentation of large-scale point clouds. Proc. *CVPR*, pp. 11108-11117.

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. Proc. *CVPR*, pp. 4977–4987.

Hu, Q., Yang, B., Khalid, S., Xiao, W., Trigoni, N. and Markham, A., 2022. Sensaturban: Learning semantics from urban-scale photogrammetric point clouds. *Int. J. Computer Vision*, 130(2), pp.316-343.

Kölle, M., Laupheimer, D., Schmohl, S., Haala, N., Rottensteiner, F., Wegner, J.D. and 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, p.100001.

Lin, H.I., Nguyen, M.C., 2020. Boosting minority class prediction on imbalanced point cloud data. *Appl. Sci*., 10, 973.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. Proc. *ICCV*, pp. 2980-2988.

Li, X., Li, C., Tong, Z., Lim, A., Yuan, J., Wu, Y., Tang, J. and Huang, R., 2020, October. Campus3d: A photogrammetry point cloud benchmark for hierarchical understanding of outdoor scene. Proc. *28th ACM Int. Conference on Multimedia*, pp. 238-246.

Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., Remondino, F., 2020a. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS Int. J. Geo-Inf*., 9, 535.

Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T., 2020b. A benchmark for large-scale heritage point cloud semantic segmentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*., 43, pp. 1419-1426.

Mazzacca, G., Grilli, E., Cirigliano, G.P., Remondino, F., Campana, S., 2022. Seeing among foliage with lidar and machine learning: towards a transferable archaeological pipeline. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVI-2/W1-2022, 365–372.

Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp.152-165.

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. Proc. *CVPR*, 1082-1091.

Ren, P., Xia, Q., 2023. Classification method for imbalanced LiDAR point cloud based on stack autoencoder. *Electron. Res. Arch*., 31, 3453-3470.

Russo, M., Grilli, E., Remondinio, F., Teruggi, S., Fassi, F., 2021. Machine learning for cultural heritage classification. Proc. *REAACH-ID* Symposium.

Teruggi, S., Grilli, E., Russo, M., Fassi, F., Remondino, F., 2020. A hierarchical machine learning approach for multi-level and multi-resolution 3D point cloud classification. *Remote Sensing*, 12(16), 2598.

Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F. and Guibas, L.J., 2019. Kpconv: Flexible and deformable convolution for point clouds. Proc. *CVPR*, pp. 6411-6420.

Varney, N., Asari, V.K. and Graehling, Q., 2020. DALES: A large-scale aerial LiDAR data set for semantic segmentation. Proc. *CVPR*, pp. 186-187.

Xie, Y., Tian, J., Zhu, X.X., 2020. Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geosci. Remote Sens. Magazine*, 8, 38-59.

Ye, Z., Xu, Y., Huang, R., Tong, X., Li, X., Liu, X., Luan, K., Hoegner, L. and Stilla, U., 2020. Lasdu: A large-scale aerial lidar dataset for semantic labeling in dense urban areas. *ISPRS Int. J. Geo-Inf*., 9(7), p.450.

Zachar, P. Bakuła, K., Ostrowski, W., 2023. CENAGIS-ALS benchmark - New proposal for dense ALS benchmark based on the review of datasets and benchmarks for 3D point cloud segmentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*., Vol. XLVIII-1/W3-2023, in press.

Zolanvari, S.M., Ruano, S., Rana, A., Cummins, A., da Silva, R.E., Rahbar, M. and Smolic, A., 2019. DublinCity: Annotated LiDAR point cloud and its applications. *arXiv*:1909.03613.

Class	Train (%)	Validation (%)	Test (%)	Point Transformer	RandLA-Net	KPConv	MLMR + PT	MLMR + RandLA-Net	MLMR + KPConv
Stairs	0.100	0.122	0.132	0.41	0.43	0.56	0.48	0.28	**0.62**
Other (Sports Area)	0.775	0.327	3.100	0.54	0.56	0.64	0.63	0.41	**0.72**
Sidewalk	7.868	9.470	7.186	0.46	0.51	0.62	0.56	0.41	**0.66**
Street/Road	18.706	12.268	18.588	0.8	0.84	0.91	0.85	0.79	**0.93**
Parking Lot	0.753	1.048	1.598	0.36	0.4	0.68	0.49	0.25	**0.71**
Low Vegetation	36.975	33.827	25.891	0.83	0.85	**0.9**	0.85	0.84	0.89
Pool/Water	0.105	0.017	0.050	0.18	**0.21**	0.13	0.11	0.03	0.09
Soil	4.850	6.307	4.460	0.55	0.6	0.72	0.6	0.45	**0.74**
Stone	0.029	0.050	0.019	0.26	0.3	**0.46**	0.34	0.15	0.41
Disorganized Region	0.301	0.252	0.283	0.19	0.37	0.53	0.35	0.22	**0.56**
Ruins	0.477	0.274	1.996	0.17	0.23	0.21	0.27	0.08	**0.37**
High Vegetation	0.387	0.773	0.849	0.71	0.74	0.86	0.77	0.69	**0.89**
Tree	12.423	15.391	12.505	0.79	0.83	0.91	0.82	0.81	**0.92**
People	0.007	0.074	0.006	**0.18**	0.08	0.07	0	0	0.02
Pet	0.001	0.004	0.002	0.08	0.12	**0.18**	0	0	0.07
Ventilation - Facade	0.004	0.017	0.006	0.5	0.4	0.52	0.57	0.39	**0.67**
Window - Facade	0.464	0.385	0.614	0.59	0.65	0.81	0.62	0.66	**0.84**
Pipeline - Facade	0.003	0.012	0.041	0.46	0.53	0.61	0.56	0.38	**0.62**
Other - Facade	0.010	0.007	0.049	0.45	0.49	0.46	0.55	0.34	**0.6**
Facade	1.632	2.000	2.623	0.77	0.85	0.82	0.8	0.75	**0.92**
Chimney - Roof	0.045	0.073	0.041	0.17	0.29	**0.34**	0.16	0.11	0.27
Window - Roof	0.085	0.416	0.625	0.44	0.46	0.6	0.49	0.31	**0.61**
Solar Panel - Roof	0.094	0.018	0.008	0.34	0.41	0.51	0.33	0.23	**0.57**
Ventilation - Roof	0.080	0.106	0.050	0.28	0.44	0.55	0.4	0.44	**0.63**
Pipeline - Roof	0.208	0.361	0.073	0.4	0.41	0.56	0.44	0.24	**0.58**
Other - Roof	0.217	0.331	0.118	0.54	0.51	0.52	0.55	0.38	**0.59**
Dome - Roof	0.042	0.102	0.043	0.28	0.4	0.54	0.45	0.25	**0.68**
Tile - Roof	1.300	2.567	1.608	0.8	0.83	0.89	0.85	0.78	**0.91**
Industrial - Roof	7.691	9.431	12.555	0.81	0.82	0.9	0.88	0.67	**0.92**
Other - Urban Furn.	0.658	0.652	0.557	0.44	0.55	0.52	0.53	0.32	**0.57**
Wall	1.002	0.610	1.176	0.33	0.52	0.66	0.47	0.37	**0.73**
Fence	0.160	0.093	0.135	0.37	0.49	0.66	0.46	0.34	**0.74**
Playground	0.005	0.001	0.020	0.31	0.2	0.44	0.35	0.05	**0.54**
Traffic Signature	0.004	0.001	0.001	0.15	0.22	**0.34**	0.13	0.07	0.28
Garbage Box	0.030	0.018	0.041	0.29	0.33	0.51	0.39	0.17	**0.61**
Lamp	0.061	0.054	0.066	0.2	0.27	**0.33**	0.15	0	0.26
Street Separator	0.930	0.611	0.829	0.29	0.43	0.59	0.41	0.28	**0.63**
Tent	0.137	0.391	0.095	0.19	0.31	0.36	0.28	0.16	**0.66**
Pole	0.002	0.002	0.002	0.14	0.2	**0.29**	0.12	0	0.23
Truck	0.081	0.041	0.059	0.47	0.62	0.64	0.58	0.51	**0.73**
Van	0.169	0.078	0.122	0.57	0.6	0.66	0.64	0.6	**0.74**
Motorbike/Bicycle	0.003	0.005	0.008	0.6	0.59	**0.69**	0.63	0.44	0.65
Bus	0.063	0.065	0.199	0.71	0.81	0.86	0.79	0.72	**0.9**
Car	1.031	1.342	1.531	0.77	0.83	0.86	0.82	0.77	**0.91**
Work Machine	0.032	0.007	0.042	0.2	0.41	**0.62**	0.35	0.28	0.55
mIoU				0.43	0.49	**0.58**	0.49	0.36	**0.62**

Table 3. Class distributions and per class results for the different methods