ESTATE: A Large Dataset of Under-Represented Urban Objects for 3D Point Cloud Classification

Onur Can Bayrak[1,2], Zhenyu Ma[2], Elisa Mariarosaria Farella[2], Fabio Remondino[2], Melis 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: <obayrak><zma><elifarella><remondino>@fbk.eu

Commission II - WG2

KEY WORDS: point cloud, deep learning, dataset, object classification, under-represented urban object

ABSTRACT: Cityscapes contain a variety of objects, each with a particular role in urban administration and development. With the rapid growth and implementation of 3D imaging technology, urban areas are increasingly surveyed with high-resolution point clouds. This technical advancement extensively improves our ability to capture and analyse urban environments and their small objects. Deep learning algorithms for point cloud data have shown considerable capacity in 3D object classification but still face problems with generally under-represented objects (such as light poles or chimneys). This paper introduces the ESTATE dataset (https://github.com/3DOM-FBK/ESTATE), which combines available datasets of various sensors, densities, regions, and object types. It includes 13 classes featuring intensity and/or colour attributes. Tests using ESTATE demonstrate that the dataset improves the classification performance of deep learning techniques and could be a game-changer to advance in the 3D classification of urban objects.

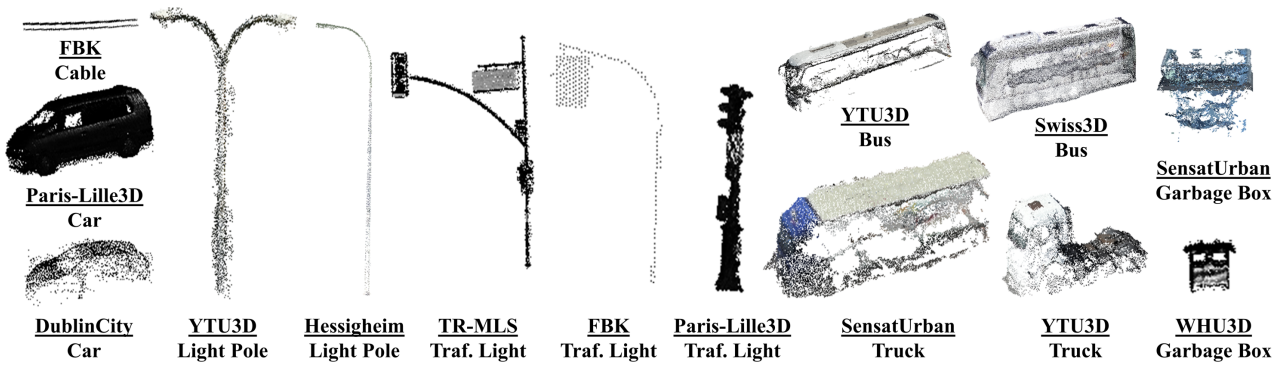

Figure 1: Examples of some objects included in the ESTATE dataset realized to improve the identification and classification of normally under-represented objects in urban point clouds.

1. INTRODUCTION

Urban point clouds have recently played an important role in 3D scene interpretation (Xie et al., 2020; Grilli et al., 2021). The growing use of reality-based 3D techniques is sparking research efforts to develop solutions for point cloud analyses useful for building modelling (Özdemir and Remondino, 2018), urban management (Zolanvari et al., 2019), street furniture extraction (Bai et al., 2021), digital twin generation (Ismail et al., 2023).
Operative approaches for point cloud categorization rely on hand-crafted feature extraction rules and a variety of machine learning-based classifiers (Zhang et al., 2023). With the advancements in deep learning-based techniques, the use of deep neural networks has gained traction (Hu et al., 2020; Hu et al., 2021; Mao et al., 2022; Ren et al., 2023), including the combination of logic rules (Grilli et al., 2023).
Although there have been positive findings, the classification of 3D point clouds still encounters numerous difficulties when applied in real-world scenarios. While current methods perform well on single datasets, they generally struggle to generalize when faced with cross-dataset circumstances, where the training and test data are obtained from distinct distributions (Wang et al., 2021). Figure 1 shows how point clouds from distinct datasets can vary in terms of density, colour, noise and shape. This variation is particularly visible in small and generally under-represented objects, such as cables, traffic lights and garbage boxes. With the advancement of 3D digitization techniques and the increase in point cloud density, machine and deep learning methods suffer even more in detecting small urban objects.
The presence of under-represented classes limits the performance of state-of-the-art neural networks. Therefore, to allow an effective utilization of deep learning-based algorithms in real-world contexts, especially for supporting the needs of municipalities and mapping agencies, a set of discriminated 3D urban elements must be incorporated. The use of reality-based 3D surveying data with respect to synthetic ones (Wu et al., 2015; Chang et al., 2015; Deitke et al., 2022) can enhance dataset generalization and expand its applications.

1.1 Paper motivation and aims

The motivations behind the presented research activities include:

- presence of under-represented urban objects in available datasets/benchmarks which can hardly be segmented with state-of-the-art neural networks;
- generalization issues across datasets using current deep learning approaches due to the heterogeneity of sensors, data, and locations.

The goal of the study is thus to create an urban point cloud dataset fulfilling the following criteria:

- data should come from several locations to represent diverse

Name	Reference	Classes	Tot. Numb. of Objects	Target	Scene Type
Sydney Urban Objects	De Deuge et al. (2013)	14	588	Outdoor scenes	Real World
ModelNet10	Wu et al. (2015)	10	4596	Indoor objects	Synthetic
ModelNet40	Wu et al. (2015)	40	12311	Indoor objects	Synthetic
ShapeNet	Chang et al. (2015)	55	51190	Indoor objects	Synthetic
ScanNet	Dai et al. (2017)	17	12283	Indoor scenes	Real World
ScanObjectNN	Uy et al. (2019)	15	2902	Indoor objects	Real World
Objaverse	Deitke et al. (2022)	21K +	10 million +	Indoor objects	Synthetic
ModelNet40-C	Sun et al. (2022)	15	185000	Indoor objects	Synthetic
ESTATE (Our)	-	13	6528	Outdoor scenes	Real World

Table 1. A summary of some representative datasets for object classification in point clouds.

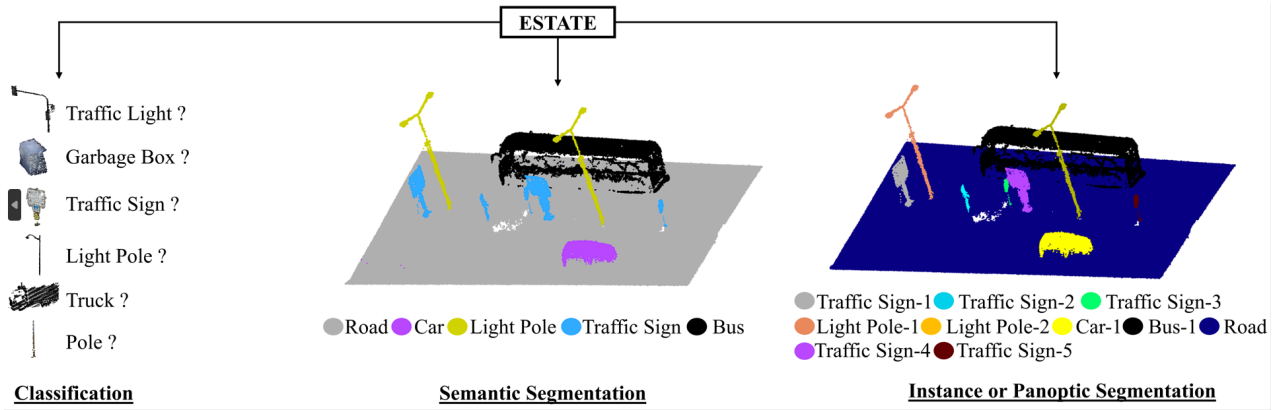

Figure 2: Potential uses of the proposed ESTATE dataset.

landscapes, environments and construction styles;
- data should include various under-represented urban objects, such as poles, vehicles and infrastructural units;
- data should have been acquired using different sensors, such as Mobile Laser Scanner (MLS), Airborne Laser Scanner (ALS), Unmanned Aerial Vehicle (UAV-Photo) or Airborne Photogrammetry, and include sensor-specific features (colour and intensity);
- data must feature various resolutions (density);
- data should support object classification.

Therefore, the aim of the paper is to present ESTATE, a new dataset to improve the identification and classification of normally under-represented objects in urban point clouds, including generalization capabilities (Figure 2). ESTATE contains 13 objects (classes). It is produced by combining in-house and available heterogeneous datasets and could also be used for semantic segmentation purposes. The reported analyses demonstrate that ESTATE improves the classification performances of deep learning techniques.

2. STATE OF THE ART

3D point cloud datasets for object classification purposes can be broadly categorised based on the location (indoor vs outdoor) and scene type (synthetic vs real-world) (Table 1).

Among the widely recognized datasets, Objaverse (Deitke et al., 2022), ModelNet10 (Wu et al., 2015), ModelNet40 (Wu et al., 2015), ModelNet40-C (Sun et al., 2022) and ShapeNet (Chang et al., 2015) consist of synthetic and indoor object samples. Despite their sizes, the synthetic contents limit their applicability in real-world scenarios with environmental variability and unpredictability.

On the other hand, datasets like ScanNet (Dai et al., 2017) and ScanObjectNN (Uy et al., 2019) provide real-world data captured from indoor environments. Although these datasets introduce more realistic scenarios compared to their synthetic counterparts, they still have disadvantages in representing generally under-represented urban objects, such as traffic lights or street furniture, which are crucial for applications like autonomous driving and urban planning.

The Sydney Urban Objects dataset (De Deuge et al., 2013) addresses some of these gaps by comprising reality-based point clouds of outdoor objects. However, the insufficient sample size of this dataset hampers its ability to generalize across the broad range of objects and complex urban conditions.

In summary, while significant progress has been made in the development of 3D datasets for object classification, the field should continue to evolve with an increasing focus on enhancing the diversity, realism and practical applicability of the real-world dataset by including a sufficient variety and number of under-represented urban objects.

3. THE ESTATE DATASET

To overcome the above-mentioned gaps, as well as generalization limitations of neural networks with existing datasets, we provide ESTATE (Figures 2 and 3), which includes various urban objects normally under-represented in the publicly available datasets. ESTATE contains 13 different classes of 3D points (with colour and/or intensity information) extracted and merged from 11 MLS/ALS/UAV-Photogrammetry datasets, which were created for 3D segmentation purposes:

- WHU-Urban3D (Han et al., 2024): a large-coverage ALS and MLS annotated datasets (three subsets) containing urban scenes and roads from different cities (one of the subsets contains 37 annotated classes);
- DublinCity (Zolanvari et al., 2019): a benchmark dataset including 13 manually annotated object classes from a LiDAR point cloud depicting the city of Dublin;

Dataset properties and classes	Without Feature	Without RGB				With RGB						Number of Objects in ESTATE
		ALS		MLS		UAV-Photogrammetry				ALS	MLS	
	WHU3D	DublinCity	FBK	Paris-Lille3D	TR-MLS	Sensat Urban	Swiss3D City	STPLS3D	YTU3D	Hessigheim	Toronto3D	
Approx. point density (pts/m²)	600	348	140	2000	700	400	1000	100	1000	800	1000	
Light Pole	337	258	70	52	48	8	5	116	346	32	64	1336
Traffic Light	79	3	2	15	17	-	-	16	6	-	26	164
Pole	135	71	27	24	28	-	13	67	39	18	32	454
Electr. Pole	7	-	83	2	-	-	-	43	-	-	41	176
Traffic Sign	231	5	74	124	114	9	-	36	20	14	43	670
Pylon	-	8	125	-	-	-	-	-	-	-	-	133
Cable	-	81	43	-	-	-	-	-	-	-	183	307
Garbage Box	87	-	-	162	13	369	17	-	120	66	-	834
Car	85	80	-	274	7	130	-	-	801	28	78	1483
Truck	10	-	5	-	2	20	14	6	64	-	-	121
Bus	7	30	-	-	-	3	2	2	38	-	-	82
Chimney	-	54	65	-	-	-	232	-	234	40	-	625
Ventilation	-	-	-	-	-	38	-	-	105	-	-	143
Total	978	590	494	653	229	577	283	286	1773	198	467	**6528**

Table 2. Selected datasets and extracted objects (classes) featuring the proposed ESTATE dataset. ALS and MLS data include also intensity values.

- Hessigheim (Kölle et al., 2021): a high-density UAV laser scanning point cloud dataset acquired over the village of Hessigheim (Germany), segmented into 11 classes;
- Paris-Lille3D (Roynard et al., 2018): about 2 km of MLS point cloud, acquired in Lille and Paris with 50 classes;
- SensatUrban (Hu et al., 2022): a photogrammetric urban-scale UAV dataset consisting of 13 object classes;
- STPLS3D (Chen et al., 2022): a synthetic aerial photogrammetric point cloud annotated with 19 classes;
- Swiss3DCities (Can et al., 2021): a LiDAR point cloud acquired over three Swiss cities with 5 semantic classes;
- Toronto3D (Tan et al., 2020): an MLS point cloud which features 8 semantic classes;
- YTU3D (Bayrak et al., 2023): a UAV photogrammetric point cloud acquired over the Davutpasa Campus of Yildiz Technical University (YTU) in Turkey, semantically labelled in 45 classes;
- two in-house datasets from Turkey (TR-MLS) and Italy (FBK).

The ESTATE dataset encompasses and refines semantically segmented 3D data from each of the presented datasets, focusing on 13 specific classes. It was noted that most of the datasets, due to semi-automatic or manual labelling procedures, contain many labelling errors; therefore, only higher quality and manually refined data were included in ESTATE. The dataset characteristics are summarized in Table 2.

4. EXPERIMENTS

The ESTATE dataset is generated to facilitate the accurate classification of generally under-represented objects. In the experiments, the data were split into the train (70%) and test (%30) with three different input configurations:

- XYZ
- XYZ + RGB
- XYZ + intensity.

Among the available recent deep learning methods (Wag et al., 2019; Wu et al., 2019; Guo et al., 2021; Lu et al., 2022), we evaluated the performance of KPConv (Thomas et al., 2019), which is commonly used in 3D semantic segmentation, object classification and SLAM segmentation benchmarks. KPConv utilizes radius neighbourhoods as input and applies weights spatially determined by a small set of kernel points. KP-CNN is a convolutional network with 5 layers for classification. Every layer consists of two convolutional blocks, with the exception of the first layer, where the first block is not stride. The convolutional blocks are structured similarly to bottleneck ResNet blocks (He et al., 2016), utilizing a KPConv instead of the traditional image convolution, along with batch normalization and leaky ReLu activation. After the final layer, the features undergo aggregation through global average pooling and are then processed by the fully connected and softmax layers. Only deformable kernels are utilized in the last 5 KPConv blocks. These kernels have proven to be highly effective in learning local shifts that can accurately adapt to the point cloud geometry and local structures. An optimizer is used to minimize the cross-entropy loss by implementing a Momentum gradient Descent approach. The batch size is set to 16, while the momentum is set to 0.98. The initial learning rate is set to 10^{-3}. The learning rate was set to decrease exponentially, with a chosen exponential decay that guarantees a division by 10 every 100 epochs during training of 300 epochs. A dropout probability of 0.5 is employed in the fully connected layers at the end. The initial subsampling grid size was set to 1 cm.

The purpose of ESTATE data is to make deep learning models invariant to various density, sensor, and object types belonging to the same class. Thus, in order to determine whether the ESTATE dataset improves (i) the classification performance and (ii) generalization capability of deep learning methods, two different training and testing approaches were applied:

- Approach-1 (Single-Train Single-Test - STST): randomly selected 70% of the objects in each dataset as training and 30% as test and perform training and testing for each dataset;
- Approach-2 (All-Train All-Test - ATAT): all train and test sets of STST are merged to examine the classification performance in general; the model is then trained and tested on a dataset basis (All-Train Single-Test (ATST) and compared to STST.

It is expected to obtain more accurate results for ATST since it includes training samples from all datasets.

Class	XYZ	Train Samples	Test Samples	XYZ+Intensity	Train Samples	Test Samples	XYZ+RGB	Train Samples	Test Samples
Light Pole	0.92	930	406	**0.96**	360	159	0.93	397	174
Traffic Light	**0.76**	112	52	0.67	42	21	0.67	33	15
Pole	0.80	312	142	**0.82**	136	64	0.81	116	53
Elect. Pole	0.76	121	55	**0.80**	87	39	0.74	58	26
Traffic Sign	0.88	464	206	**0.90**	258	116	0.77	84	38
Pylon	0.95	92	41	**0.98**	92	41	-	-	-
Cable	0.80	214	93	0.89	214	93	**0.96**	128	55
Garbage Box	0.89	581	253	**0.94**	168	73	0.87	399	173
Car	**0.99**	1034	449	0.98	324	143	**0.99**	724	313
Truck	**0.82**	82	39	0.00	4	3	**0.82**	71	33
Bus	0.88	55	27	**1.00**	21	9	0.75	30	15
Chimney	0.86	435	190	0.88	110	49	**0.89**	353	153
Ventilation	0.68	99	44	-	-	-	**0.72**	99	44
Total		4531	1997		1816	810		2492	1092

Table 3. ATAT results with KPConv on the 13 objects of ESTATE.

For thorough performance analysis, these procedures are repeated for Intensity and RGB features to evaluate their importance (therefore: XYZ, XYZ+Intensity and XYZ+RGB). For the XYZ, XYZ+Intensity and XYZ+RGB input configurations, STST comprises 11, 6 and 6 training and testing operations, respectively. For the XYZ, XYZ+Intensity, and XYZ+RGB input configurations, ATST incorporates 1 training process and 11, 6 and 6 test operations, respectively. Since ATAT comprises a single training and testing operation, a cumulative sum of 29 training operations and 49 test processes were executed within our analyses.

Table 3 shows ATAT results obtained for XYZ, XYZ+Intensity, and XYZ+RGB input configurations and the number of samples used for training and testing. For Traffic Light, the best scores are obtained with XYZ input (F1-score of 0.76), while scores of 0.99 and 0.82 are obtained for Car and Truck classes, respectively, for both XYZ and XYZ+RGB configurations. With XYZ+Intensity inputs, the most successful scores were obtained for Light Pole (0.96), Pole (0.82), Electrical Pole (0.80), Traffic Sign (0.90), Pylon (0.98), Garbage Box (0.94) and Bus (1.00) classes. For XYZ+RGB, the most successful results were obtained for Cable (0.96), Chimney (0.89) and Ventilation (0.72) classes. For the Traffic Light class, the addition of Intensity and RGB features decreased the classification accuracy. It was observed that the RGB attribute decreased the classification accuracy of objects with various colour ranges in different datasets such as Traffic Light, Electrical Pole, Traffic Sign, Garbage Box and Bus. This finding is similar to Sun et al. (2020), where the classification performance decreases with the addition of colour information. However, the best results were obtained with the addition of Intensity in Light Pole, Pole, Electrical Pole, Traffic Sign, Pylon, Garbage Box and Bus classes. According to these results, it is seen that the employed network uses the point cloud geometry predominantly, while RGB attributes (that can difference among datasets) decrease the generalization ability (which instead also increase with the Intensity attribute).

Table 4 reports classification results using only XYZ information. The model trained on the ESTATE dataset using XYZ features improved the classification results. The F1-scores obtained for STST and ATST, respectively, have relatively small improvements of 0.95 to 0.96 for Light Pole in DublinCity, 0.85 to 0.89 for Traffic Sign in FBK, 0.63 to 0.64 for TR-MLS Pole. But, at the same time, the ESTATE dataset allows meaningful improvements for Bus in DublinCity (from 0.2 to 1.00), Traffic Light in FBK (from 0.00 to 0.4), Light Pole in SensatUrban (from 0.50 to 0.80), Electrical Pole in Paris-Lille3D (from 0.67 to 1.00), Garbage Box in Swiss3DCities (from 0.44 to 0.73), Traffic Light (from 0.00 to 0.86) and Electrical Pole (from 0.24 to 0.69) in Toronto3D. However, no improvement was achieved in YTU3D and Hessigheim datasets. This is probably due to the high similarity of objects in the same class in those datasets.

Table 5 reports the obtained results with XYZ+Intensity features for STST and ATST. It is seen that ESTATE dataset improves the classification results. For example, relatively small improvements such as 0.92 to 0.97 for Light Pole in DublinCity, 0.95 to 0.96 and 0.85 to 0.90 for Pylon and Traffic Sign in FBK, 0.94 to 1.00 for Light Pole in Paris-Lille3D, 0.80 to 0.88 for Chimney in Hessigheim, 0.76 to 0.89 and 0.97 to 0.99 for Light Pole and Cable Toronto3D are observed. On the other hand, relatively higher improvements for Pylon in DublinCity (from 0.00 to 0.50), Traffic Light in FBK (from 0.00 to 0.50), Electrical Pole in Paris-Lille3D (from 0.67 to 1.00), Pole (from 0.29 to 0.80) and Traffic Sign (from 0.57 to 1.00) in Hessigheim, Traffic Light (from 0.40 to 0.75), Electrical Pole (from 0.50 to 0.85) and Traffic Sign (from 0.55 to 0.88) in Toronto3D are visible.

The results of STST and ATST for XYZ+RGB features (Table 6) show that ESTATE improves the classification results. Relatively small improvements for Car in SensatUrban (from 0.97 to 0.99), Pole in Swiss3DCities (from 0.89 to 1.00), Light Pole in STPLS3D (from 0.79 to 0.80), Light Pole (from 0.96 to 0.97), Truck (from 0.81 to 0.86) in YTU3D, Pole (from 0.77 to 0.83) in Hessigheim and Cable (from 0.97 to 0.98) in Toronto3D are attained. On the other hand, we have achieved relatively significant improvements for the class Truck in SensatUrban (from 0.67 to 0.80) and Swiss3DCities (from 0.67 to 0.83), Traffic Light in STPLS3D (from 0.67 to 0.89) and YTU3D (from 0.00 to 0.67), Traffic Light (from 0.22 to 0.62) and Electrical Pole (from 0.41 to 0.83) in Toronto3D.

Input: XYZ F1-Scores	WHU3D		DublinCity		FBK		Paris-Lille3D		TR-MLS		SensatUrban		Swiss3DCity		STPLS3D		YTU3D		Hessigheim		Toronto3D	
	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST
Light Pole	0.96	0.94	0.95	**0.96**	0.79	**0.85**	0.94	**1.0**	0.83	0.74	0.50	**0.80**	0.80	0.80	0.88	0.86	0.99	0.97	0.95	0.95	0.64	**0.74**
	337		258		70		52		48		8		5		116		346		32		64	
Traffic Light	0.65	**0.68**	0.00	0.00	0.00	**0.4**	0.80	0.77	0.44	**0.80**	-	-	-	-	0.75	**0.80**	1.00	0.80	-	-	0.00	**0.86**
	79		3		2		15		17		-		-		16		6		-		26	
Pole	0.86	0.85	0.83	0.74	0.76	0.73	1.00	0.93	0.63	**0.64**	-	-	0.75	**1.00**	0.88	0.80	0.92	0.85	1.00	0.80	0.74	**0.87**
	135		71		27		24		28		-		13		67		39		18		32	
Electrical Pole	0.00	0.00	-	-	0.86	0.85	0.67	**1.00**	-	-	-	-	-	-	0.75	0.85	-	-	-	-	0.24	**0.69**
	7		-		83		2		-		-		-		43		-		-		41	
Traffic Sign	0.86	0.80	0.00	0.00	0.85	**0.89**	0.96	**0.97**	0.85	**0.94**	0.40	**1.00**	-	-	0.82	0.82	0.91	**0.92**	1.00	0.89	0.76	**0.87**
	231		5		74		124		114		9		-		36		20		14		43	
Cable	-	-	0.98	**1.00**	0.96	0.96	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.99	0.99
	-		81		43		-		-		-		-		-		-		-		183	
Pylon	-	-	0.00	0.00	0.95	0.92	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-		8		125		-		-		-		-		-		-		-		-	
Garbage Box	0.96	0.88	-	-	-	-	1.00	0.95	0.86	**0.86**	0.93	0.93	0.44	**0.73**	-	-	0.75	0.70	0.92	0.86	-	-
	87		-		-		162		13		369		17		-		120		66		-	
Car	1.00	0.96	0.98	**1.00**	-	-	1.00	0.99	0.67	0.67	0.97	0.92	-	-	-	-	0.99	0.99	0.95	0.95	1.00	0.98
	85		80		-		274		7		130		-		-		801		28		78	
Truck	0.86	0.86	-	-	1.00	1.00	-	-	0.00	0.00	0.92	0.91	0.67	**1.00**	0.80	0.67	0.90	0.80	-	-	-	-
	10		-		5		-		2		20		14		6		64		-		-	
Bus	0.80	**1.00**	0.2	**1.00**	-	-	-	-	-	-	1.00	0.67	0.00	**1.00**	0.00	**0.67**	0.82	0.81	-	-	-	-
	7		30		-		-		-		3		2		2		38		-		-	
Chimney	-	-	1.00	0.83	0.98	0.83	-	-	-	-	-	-	0.96	**0.97**	-	-	0.91	0.81	0.87	0.81	-	-
	-		54		65		-		-		-		232		-		234		40		-	
Ventilation	-	-	-	-	-	-	-	-	-	-	0.53	0.50	-	-	-	-	0.78	0.72	-	-	-	-
	-		-		-		-		-		38		-		-		105		-		-	

Table 4. Classification results (F1-score) and number of instances per object with XYZ input for STST and ATST.

5. CONCLUSIONS

The paper (i) introduced a new dataset for the 3D classification of urban objects and (ii) evaluated its benefits on a deep learning method with various input configurations. The shared data and research findings are publicly available at https://github.com/3DOM-FBK/ESTATE.

The detailed collection of the 13 objects, which include some urban objects normally under-represented in commonly available datasets, enhances the practical utility of 3D object classification models. The reported experimental results indicate that the ESTATE dataset improved the overall performance of classification models. Intensity feature obtained more successful results than RGB colour inputs. This shows that additional features are able to improve the classification performance, but the inclusion of various colour features couldn't provide the expected improvement. In order to use the models to be trained on the ESTATE dataset in real-life scenarios, only the coordinate values of the objects can be considered since the results show that there are relatively low differences between the results of XYZ and XYZ+Intensity input configurations.

Furthermore, the ESTATE dataset has the potential to be used for object classification as well as semantic segmentation, instance or panoptic segmentation (Figure 2), where objects in complex urban areas can be extracted using traditional pre-processing methods or unsupervised learning, graphs, etc. Future studies may evaluate the performances of other neural networks and focus on improving and integrating supervised and unsupervised learning techniques into a complementary process.

input: XYZ+I F1-Scores	**DublinCity**		**FBK**		**Paris-Lille3D**		**TR-MLS**		**Hessigheim**		**Toronto3D**	
	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST
Light Pole	0.92	**0.97**	0.79	**0.95**	0.94	**1.00**	0.85	**0.90**	1.00	1.00	0.76	**0.89**
Traffic Light	0.00	**1.00**	0.00	**0.50**	0.80	0.80	0.29	**0.80**	-	-	0.40	**0.75**
Pole	0.76	**0.83**	0.76	**0.90**	1.00	0.93	0.70	0.70	0.67	**0.92**	0.87	0.82
Electrical Pole	-	-	0.86	0.86	0.67	**1.00**	-	-	-	-	0.50	**0.85**
Traffic Sign	0.00	0.00	0.85	**0.90**	0.96	0.95	0.90	**0.92**	0.57	**1.00**	0.55	**0.88**
Cable	0.94	**0.98**	0.96	0.96	-	-	-	-	-	-	0.97	**0.99**
Pylon	0.00	**0.50**	0.95	**0.96**	-	-	-	-	-	-	-	-
Garbage Box	-	-	-	-	1.00	0.99	0.40	**0.75**	0.92	0.92	-	-
Car	0.91	**1.00**	-	-	1.00	1.00	0.29	**0.80**	0.90	**1.00**	1.00	0.98
Truck	-	-	1.00	0.00	-	-	0.00	0.00	-	-	-	-
Bus	0.63	**1.00**	-	-	-	-	-	-	-	-	-	-
Chimney	0.65	**0.94**	0.98	0.86	-	-	-	-	0.80	**0.88**	-	-
Ventilation	-	-	-	-	-	-	-	-	-	-	-	-

Table 5. Classification results with XYZ+Intensity input for STST and ATST (number of instances per object as in Table 4).

Input: XYZ+RGB F1-Scores	**SensatUrban**		**Swiss3DCity**		**STPLS3D**		**YTU3D**		**Hessigheim**		**Toronto3D**	
	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST	STST	ATST
Light Pole	0.86	0.86	1.00	0.80	0.79	**0.80**	0.96	**0.97**	1.00	0.91	0.62	**0.86**
Traffic Light	-	-	-	-	0.67	**0.89**	0.00	**0.67**	-	-	0.22	**0.62**
Pole	-	-	0.89	**1.00**	0.80	0.78	0.82	0.76	0.77	**0.83**	0.58	**0.80**
Electrical Pole	-	-	-	-	0.75	0.70	-	-	-	-	0.41	**0.83**
Traffic Sign	0.75	0.75	-	-	0.74	0.70	0.91	0.91	0.89	0.89	0.53	**0.76**
Cable	-	-	-	-	-	-	-	-	-	-	0.97	**0.98**
Garbage Box	0.94	0.94	0.91	0.29	-	-	0.85	0.68	1.00	0.97	-	-
Car	0.97	**0.99**	-	-	-	-	0.99	0.99	1.00	0.94	1.00	1.00
Truck	0.67	**0.80**	0.67	**0.83**	0.80	0.50	0.81	**0.86**	-	-	-	-
Bus	1.00	0.50	0.00	0.00	0.00	0.00	1.00	0.92	-	-	-	-
Chimney	-	-	0.97	0.97	-	-	0.86	0.83	0.92	0.88	-	-
Ventilation	0.63	0.61	-	-	-	-	0.76	**0.83**	-	-	-	-

Table 6. Classification results with XYZ+RGB input for STST and ATST (number of instances per object as in Table 4).

ACKNOWLEDGEMENTS

Onur Can Bayrak is supported within the scope of 2214-A Programme with 1059B142201684 project code by The Scientific and Technological Research Council of Türkiye (TUBITAK).

The work is partly funded by the EU project USAGE - Urban Data Space for Green Deal (https://www.usage-project.eu/) which has received funding from the European Union’s Horizon Europe Framework Programme for Research and Innovation under the Grant Agreement no 101059950 - call HORIZONCL6-2021-GOVERNANCE-01-17 (IA).

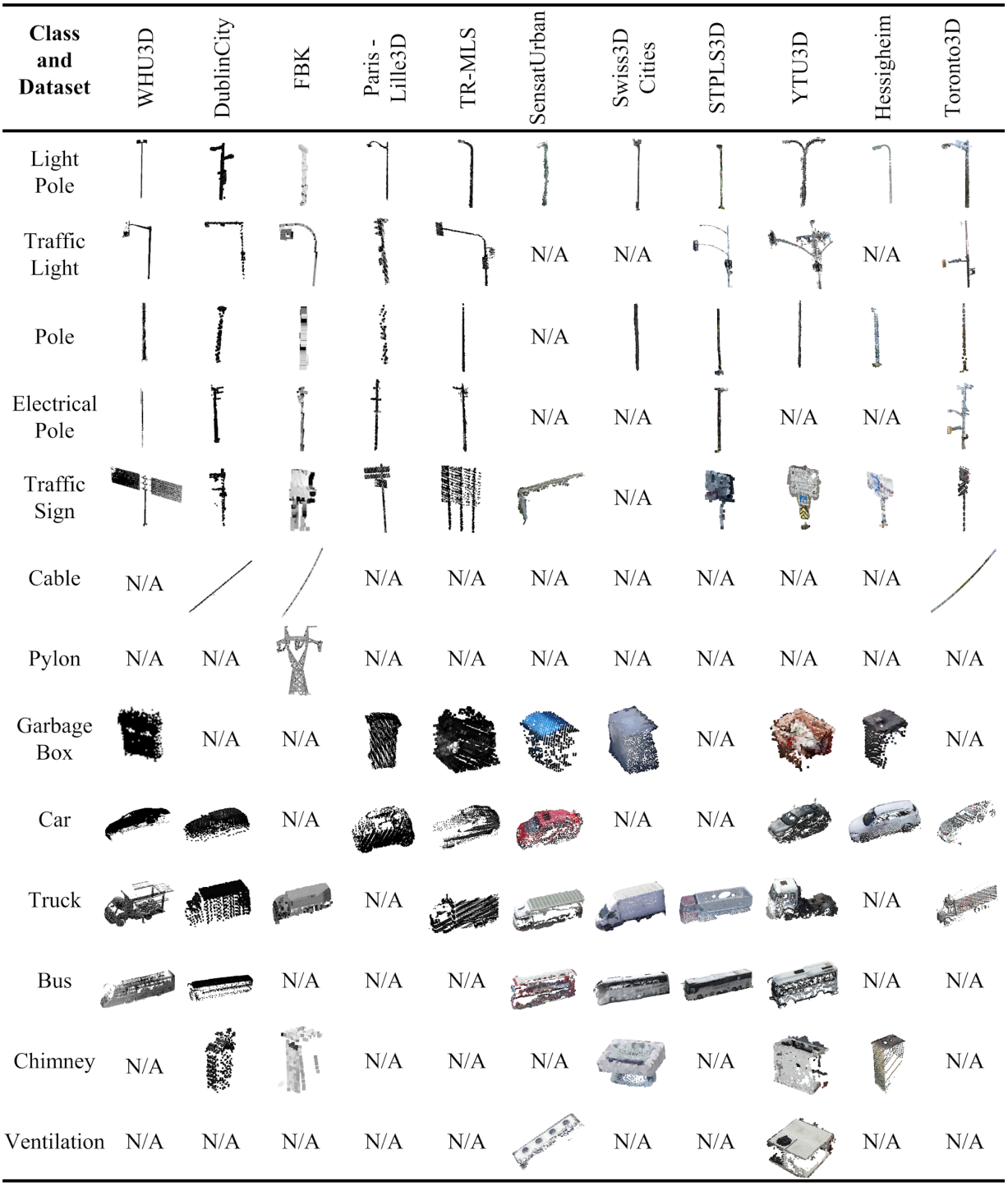

Figure 3: Examples of instances (rows) available in the proposed ESTATE dataset and collected from various available datasets (columns).

REFERENCES

Bai, Q., Lindenbergh, R.C., Vijverberg, J. and Guelen, J.A.P., 2021. Road type classification of MLS point clouds using deep learning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*., *43*, pp.115-122.

Bayrak, O.C., Remondino, F. and Uzar, M., 2023. A New Dataset and Methodology for Urban-Scale 3D Point Cloud Classification. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*.,48, pp.1-8.

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

Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H. and Xiao, J., 2015. Shapenet: An information-rich 3d model repository. *arXiv:1512.03012*.

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*.

Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T. and Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. *Proc. CVPR*, pp. 5828-5839.

De Deuge, M., Quadros, A., Hung, C. and Douillard, B., 2013, December. Unsupervised feature learning for classification of outdoor 3D scans. Proc. *Australasian conference on robotics and automation*, Vol. 2, No. 1.

Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., VanderBilt, E., Schmidt, L., Ehsani, K., Kembhavi, A. and Farhadi, A., 2023. Objaverse: A universe of annotated 3d objects. *Proc. CVPR,* pp. 13142-13153.

Grilli, E., Daniele, A., Bassier, M., Remondino, F., Serafini, L., 2023. Knowledge Enhanced Neural Networks for Point Cloud Semantic Segmentation. *Remote Sensing*, 5(10):2590.

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.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L. and Bennamoun, M., 2021. Deep learning for 3d point clouds: A survey. *IEEE TPAMI*, 43, pp. 4338-4364.

Han, X., Liu, C., Zhou, Y., Tan, K., Dong, Z. and Yang, B., 2024. WHU-Urban3D: An urban scene LiDAR point cloud dataset for semantic instance segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 209, pp.500-513.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

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

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., 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.

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.

Ismail, M. H., Shaker, A., and Li, S., 2023. Developing complete urban digital twins in busy environments: a framework for facilitating 3d model generation from multi-source point cloud data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W2-2023

Li, Q., Zhuang, Y. and Huai, J., 2023. Multi-sensor fusion for robust localization with moving object segmentation in complex dynamic 3D scenes. *International Journal of Applied Earth Observation and Geoinformation*, 124, p.103507.

Lu, D., Xie, Q., Wei, M., Member, S., Gao, K., Member, S., Xu, L. and Li, J., 2022. Transformers in 3D point clouds: A survey. *arXiv:2205.07417*.

Mao, Y., Chen, K., Diao, W., Sun, X., Lu, X., Fu, K. and Weinmann, M., 2022. Beyond single receptive field: A receptive field fusion-and-stratification network for airborne laser scanning point cloud classification. ISPRS Journal of Photogrammetry and Remote Sensing, 188, pp.45-61.

Özdemir, E. and Remondino, F., 2018. Segmentation of 3D photogrammetric point cloud for 3D building modeling. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLII-4/W10, 135–142

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

Roynard, X., Deschaud, J.E. and 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), pp.545-557.

Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., Yang, K. and Li, J., 2020. Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways. *Proc. CVPR*, pp. 202-203.

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. ICCV*, pp. 6411-6420.

Uy, M.A., Pham, Q.H., Hua, B.S., Nguyen, T. and Yeung, S.K., 2019. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. *Proc. ICCV*, pp. 1588-1597.

Wang, F., Li, W. and Xu, D., 2021. Cross-dataset point cloud recognition using deep-shallow domain adaptation network. *IEEE Transactions on Image Processing,* 30, pp.7364-7377.

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):1–12.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X. and Xiao, J., 2015. 3d Shapenets: A deep representation for volumetric shapes. *Proc. CVPR*, pp. 1912-1920.

Wu, W., Qi, Z., and Fuxin, L., 2019. Pointconv: Deep convolutional networks on 3d point clouds. *Proc. CVPR*, pp. 9621-9630

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.

Zhang, H., Wang, C., Tian, S., Lu, B., Zhang, L., Ning, X. and Bai, X., 2023. Deep learning-based 3D point cloud classification: A systematic survey and outlook. *Displays*, p.102456.

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 preprint arXiv:1909.03613.