PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science
https://doi.org/10.1007/s41064-025-00374-7

ORIGINAL ARTICLE

Combining 3D Urban Objects from All Around the World to Improve Object Classification and Semantic Segmentation

Onur Can Bayrak[1] · Zhenyu Ma[2,3] · Elisa Mariarosaria Farella[2] · Fabio Remondino[2] · Melis Uzar[1]

Received: 9 June 2025 / Accepted: 15 December 2025

Abstract
Given the growing number of applications in urban planning and large-scale digital twins, the development of effective solutions for urban point cloud classification is of extreme interest for the R&D community and commercial sector. State-of-the-art neural networks commonly lack adequate cross-dataset generalisation ability, mainly due to varying sensors and data collection platforms, object shape differences, as well as the presence of under-represented objects and imbalanced classes, especially in case of dense and high-resolution reality-based 3D data. This work demonstrates how the recently released ESTATE dataset (A large dataset of under-represented urban objects—https://github.com/3DOM-FBK/ESTATE), full of thousands of under-represented urban objects, such as traffic lights, electrical poles, pylons, and ventilation units, spread over 13 classes, can improve the performance of state-of-the-art point cloud classification algorithms. Experiments with different neural networks and several testing configurations with sensor-specific inputs (coordinate, intensity, and colour) show the effectiveness of this dataset in enhancing the classification capabilities and increasing cross-dataset generalisation. Moreover, reported results show not only the adaptation of object classification networks to the semantic segmentation pipeline, but also an improvement of semantic segmentation performance by increasing the distribution of under-represented classes with the ESTATE dataset.

Keywords Point cloud · 3D deep learning · Dataset · 3D object classification · Under-represented urban object

1 Introduction

With the recent progress of technologies, sensors and methods, more and more accurate and dense urban 3D data are nowadays available and used in several fields, such as autonomous driving, robotics or geospatial studies (Liang et al. 2019; Guo et al. 2020; Xue et al. 2020; Bloembergen and Eijgenstein 2021). Urban 3D point clouds, in particular, have been increasingly exploited for many applications, like smart city development (Iliopoulou and Feloni 2022), building modelling (Özdemir and Remondino 2018), urban management (Zolanvari et al. 2019), street furniture extraction (Bai et al. 2021), and digital twin generation (Ismail et al. 2023). When enriched with colorimetric information (Sarker et al. 2024, p. 3) D data offer a comprehensive understanding of urban scenes and environments (Guo et al. 2020; Xie et al. 2020; Grilli et al. 2021). Among the various point cloud processing tasks, semantic enrichment and classification is an area of considerable interest within the research community due to the increasing requests for semantically enriched 3D assets. Operative methods for point cloud classification utilize manually designed feature extraction criteria and a range of operative machine learning classifiers (Zhang et al. 2023). But recently, the utilization of deep neural networks has become more prevalent due to the progress made by deep learning methods (Hu et al. 2020, Hu et al. 2021; Mao et al. 2022; Ren and Xia 2023), which also includes the integration of logic rules (Grilli et al. 2023).

In order to assess the capability of different classification algorithms, several benchmark datasets have been released in the last few years (Roynard et al. 2018; Zolanvari et al. 2019; Hu et al. 2021; Kölle et al. 2021; Chen et al. 2022). A benchmark dataset is a collection of data used by scien-

✉ Onur Can Bayrak
onurcb@yildiz.edu.tr

[1] Department of Geomatic Engineering, Yildiz Technical University, 34220 Istanbul, Türkiye

[2] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy

[3] School of Data Science, Qingdao University of Science and Technology, Qingdao, China

Published online: 14 January 2026

tists to compare the effectiveness of sensors or processing algorithms. Benchmarks serve as a reference for evaluating their performance against a reliable and accurate ground truth (Bakuła et al. 2019).

Nevertheless, the classification of 3D point clouds faces significant challenges when applied to large real-world scenarios and datasets. Indeed, real-world urban scenarios suffer extreme class imbalance for critical categories such as various pole-like and urban objects that only occupy a small proportion of the total number of points. This issue is particularly noticeable in small and generally under-represented objects, such as cables, light poles, traffic signs, traffic lights and garbage boxes. Moreover, it is necessary to clarify the impact of different object shapes, sensors, colour and intensity information, as well as data acquisition protocols for comprehensive usage in real-life scenarios. The higher density of point clouds derived from more and more performant 3D reconstruction algorithms and advanced sensors compromises the ability of machine and deep learning methods in recognizing small urban objects and to work with imbalanced classes. Finally, and potentially most significantly, none of the existing deep neural network methods have demonstrated generalization capabilities among benchmarks. Although current approaches demonstrate strong performance on individual datasets, they often encounter difficulties in generalizing when the training and test data are sourced from different locations (Wang et al. 2021) or sensors/platforms (Han et al. 2024).

Therefore, in order to enable effective utilization of deep learning-based algorithms in real-world contexts (especially for classifying under-represented objects typically present in urban areas), a set of discriminated 3D urban elements must be incorporated into the learning process. To overcome the current limitations, it is necessary to enhance the diversity, realism and practical applicability of real-world datasets by including a sufficient variety and number of under-represented urban objects in classification tasks.

This paper aims to resolve the above-mentioned issues, extending the ESTATE work presented in Bayrak et al. (2024). The extensions and contribution of this paper are multifold:

- To further explain in detail the ESTATE dataset realized to improve the 3D classification of under-represented urban objects (Sect. 3);
- To evaluate the benefits of including ESTATE in the learning process of state-of-the-art neural networks for 3D urban object classification (Sects. 4.2 and 4.3);
- To analyze (i) the impact of combined training data, (ii) the effects of input features and (iii) the influence of object shapes on classification performance (Sects. 4.4 and 4.5);
- To present, for the first time, an innovative adaptation of trained object classification methods to semantic segmentation (Sect. 4.6);
- To improve semantic segmentation accuracy, the inclusion of objects from the ESTATE dataset into the training set leads to significant improvements in semantic segmentation accuracy (Sect. 4.7).

The paper is organised as follows: Sect. 2 reports an overview of the state-of-the-art benchmark datasets, under-represented objects in urban areas, and deep learning-based algorithms for 3D object classification. Section 3 describes the collection and creation of the ESTATE dataset, while Sect. 4 presents experiments and classification results using the ESTATE data. Section 5 and 6 discuss achievements and conclude the work.

2 Related Works

2.1 Datasets/Benchmarks for Object Classification

The research community has released various datasets and benchmarks to support object classification in 3D point clouds (Table 1):

- Sydney Urban Objects (De Deuge et al. 2013): it comprises reality-based point clouds of outdoor objects. However, the low point density of this dataset hampers its ability to generalize across a broad range of objects and complex urban conditions. The Sydney Urban Ob-

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

Dataset	Classes	Objects	Scene	Type
Sydney Urban Objects	14	588	Outdoor	Real
ModelNet10	10	4596	Indoor	Synthetic
ModelNet40	40	51,190	Indoor	Synthetic
ShapeNet	55	51,190	Indoor	Synthetic
ScanNet	17	12,283	Indoor	Real
ScanObjectNN	15	2902	Indoor	Real
Objaverse	21K+	10+mil	Indoor	Synthetic
ModelNet40-C	40	185,000	Indoor	Synthetic
ESTATE	13	6528	Outdoor	Real

jects dataset was created by the Australian Centre and contains a variety of common urban road objects, including 631 scanned objects in the categories of vehicles, pedestrians, signs, trees, etc. These objects are captured using Velodyne LiDAR sensors in Sydney CBD, providing detailed point clouds that are essential for classification tasks. This dataset is specifically designed to enhance 3D object recognition systems with real-world complexities, such as variable point densities and occlusions, commonly encountered in urban environments. The dataset supports and facilitates the advancement of outdoor 3D object recognition technologies in academic and practical fields.

- ModelNet40 (Wu et al. 2015): it is a large-scale synthetic CAD-based dataset developed by the Vision and Robotics Laboratory at Princeton University. ModelNet40 contains 12,311 CAD models of 40 object categories, including different objects such as airplanes, cars, plants, and lamps. The dataset is the most widely used benchmark for 3D object classification algorithms due to its well-structured data and clear shapes. For training and testing machine learning algorithms, ModelNet40 provides 9843 models for training and 2468 for testing. Each model is pre-processed to align with the origin and scaled to fit within a unit sphere, ensuring uniformity in data handling. This dataset is crucial for advancing research in 3D computer vision, offering a standardized set of objects for developing and benchmarking new technologies.
- ModelNet10 (Wu et al. 2015): it is a subset of ModelNet40, containing only 10 classes, and it is divided into 3991 training and 908 testing shapes.
- ModelNet40-C (Sun et al. 2022): it contains 185,000 different point clouds and was created based on ModelNet40 validation set. This dataset is mainly used to benchmark damage robustness for 3D point cloud recognition, with 15 damage types and 5 severity levels for incompleteness or non-uniformity issues, such as noise or density.
- ShapeNet (Chang et al. 2015): it is developed collaboratively by Stanford University, Princeton University and the Toyota Technological Institute of Chicago. The dataset is an extensive repository of 3D CAD models organized into a comprehensive taxonomy based on WordNet. This repository features over 300 million models, with 220,000 models meticulously categorized into 3135 distinct classes. ShapeNet provides a wealth of semantic annotations for each model, such as physical sizes, keywords, rigid alignments, bilateral symmetry planes, and other semantic details. This level of detail supports more sophisticated data-driven geometric analyses and provides a robust benchmark for evaluating and comparing algorithms.
- ScanNet (Dai et al. 2017): it is a richly annotated RGB-D dataset containing over 2.5 million images from 1513 scans of 707 unique indoor environments with about 90% surface coverage. This dataset is categorized into 20 classes of annotated 3D voxelized objects. The dataset is detailed with annotations like camera poses, surface reconstructions, textured meshes, and aligned CAD models.
- ScanObjectNN (Uy et al. 2019): it is a real-world dataset developed by researchers from the Hong Kong University of Science and Technology along with other global institutions. The dataset encompasses 2902 objects categorized into 15 classes representing common indoor environments. It includes objects often occluded or embedded within clutter, reflecting typical conditions faced in robotics and autonomous vehicle applications. It features various perturbations, such as noise and incomplete data, to test the robustness of classification algorithms.
- Objaverse (Deitke et al. 2023): the dataset provides an extensive collection of over 800,000 3D models from a diverse community of over 100,000 artists via Sketchfab. The dataset covers a wide range of object categories, including animals, humans, vehicles, and architectural structures (interiors and exteriors). They are paired with detailed metadata, including descriptive captions and tags, making them useful for a variety of applications in 3D vision and machine learning in simulated environments.

As summarized in Table 1, among these widely recognized datasets, Objaverse, ModelNet40, ModelNet40-C, ModelNet10 and ShapeNet consist of synthetic and various object samples. The corresponding points are uniformly sampled from the CAD mesh surface and then further preprocessed by translations and scaling. Despite their large scale, their synthetic nature limits the applicability to real-world scenarios. On the other hand, datasets like ScanNet and ScanObjectNN provide real-world data captured from only indoor environments. Although these datasets introduce more realistic scenarios compared to their synthetic counterparts, they still have disadvantages in representing outdoor urban objects, such as traffic lights or street furniture, which are crucial for applications like autonomous driving and urban planning. Thus, they are not suitable for the identification of small and generally under-represented urban objects.

In summary, while significant progress has been made in the development of 3D datasets for object classification, the available data still lack in terms of realism and practical applicability to real-world urban scenarios.

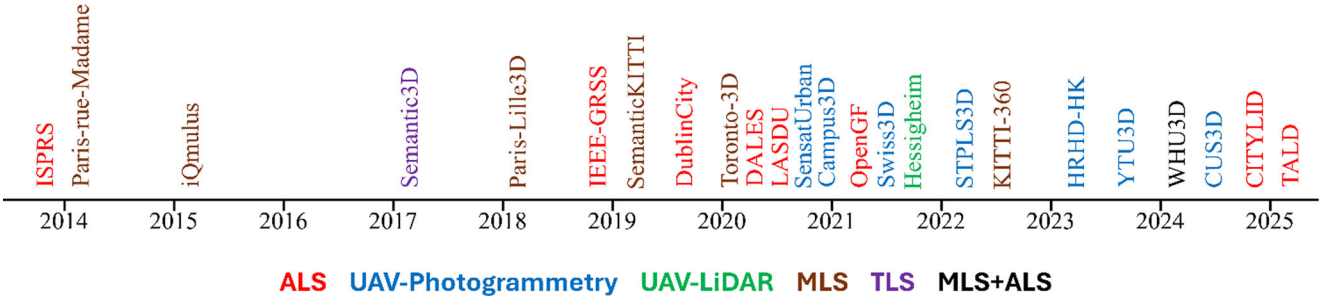

Fig. 1 Timeline of the most referenced urban-level semantic segmentation datasets/benchmarks

2.2 Datasets/Benchmarks for Point Cloud Semantic Segmentation

The research community has released various datasets and benchmarks to support point cloud semantic segmentation in urban environments (Fig. 1 and Table 2):

- ISPRS (Niemeyer et al. 2014): it is a pioneer ALS dataset over Vaihingen. The dataset covers 9 classes, including common classes such as buildings, ground and vegetation, and sub-classes such as roofs and facades.
- Paris-rue-Madame (Serna et al. 2014): is a high-resolution dataset acquired by MLS of a 160 m long street in Paris, with 17 annotated classes.
- iQmulus (Vallet et al. 2015): it is an MLS point cloud dataset covering approximately 10 km of streets within a square km in Paris. It provides manual annotations of 101 classes under a three-level hierarchical manner. This dataset features four main classes, providing a general division of urban objects and facilitating object-level analysis.
- Semantic3D (Hackel et al. 2017): a TLS point cloud dataset obtained from various urban landscapes, consists of over 4 billion points, covering outdoor scenes across Central Europe. It includes eight semantic classes.
- Paris-Lille3D (Roynard et al. 2018): a point cloud dataset acquired with an MLS prototype produced at the Center for Robotics of Mines ParisTech. Data are captured in two areas (Lille and Paris) and feature relatively high density but with anisotropic patterns due to the multi-beam LiDAR sensor. The dataset is annotated with 50 classes.
- IEEE-GRSS (Bosch et al. 2019): it is a large-scale ALS dataset covering four USA cities. However, only 5 classes are available, and it suffers from low point density.
- SemanticKITTI (Behley et al. 2019): the dataset is a MLS point clouds including sensor odometry information from roads around Karlsruhe and developed for autonomous vehicle and transportation purposes. It offers rich class annotations across 28 categories, including specific instances such as drivers of various vehicles and diverse urban objects. However, it has a low point density and various classes (e.g., bicycles, motorcycles, other object) have very few points whereas other classes (e.g., road, sidewalk and building) overwhelm the dataset.
- DublinCity (Zolanvari et al. 2019): it is a LiDAR point cloud dataset over Dublin. A hierarchical approach was followed for the manual labelling process and the annotation of 13 classes. Labelled data are structured following three hierarchical levels, moving from a coarse labelling that includes four classes to a finer division of urban elements into refined classes.
- Toronto3D (Tan et al. 2020): it is a large-scale MLS annotated dataset depicting Toronto for semantic segmentation tasks with a focus on urban roadways. Some of the 8 labelled classes are quite unusual, such as road marking and utility lines.
- DALES (Varney et al. 2020): it is another ALS dataset that covers approximately 10 km2 area of the City of Surrey in British Columbia and features 8 classes. Despite its large-scale coverage, it has relatively low point density, similar to SemanticKITTI and LASDU.
- LASDU (Ye et al. 2020): a large-scale point cloud ALS dataset over the Heihe River Valley. The dataset covers approximately 1 km^2 and consists of 3.12 million points with a low point density of 3–4 points/m^2. It includes five labelled classes.
- SensatUrban (Hu et al. 2021): it is an urban-scale UAV photogrammetric dataset covering three UK cities. The dataset features 13 annotated categories, including some common classes (e.g. vegetation, ground and buildings), as well as unusual groups like rail, bridge and water.
- Swiss3DCities (Can et al. 2021): it is a dataset covering three Swiss cities with different characteristics, obtained from images acquired by a multirotor and high-resolution oblique and nadir camera. The dataset was manually segmented into 5 categories and chimneys were primarily extracted and included in the ESTATE dataset.
- Campus3D (Li et al. 2020a): a photogrammetric point cloud dataset acquired via UAV imagery over the National University of Singapore (NUS) campus, covering an area of 1.58 km^2, and is designed for hierarchical out-

door scene understanding. It features hierarchical multi-level annotations with 24 semantic classes.

- OpenGF (Qin et al. 2021): an ultra-large-scale ALS dataset specifically designed for ground and non-ground separation, covering an area of 47 km² across four countries (Netherlands, New Zealand, the US, and Canada). Although it has ultra-large-scale coverage, the dataset was categorized under two classes, ground and non-ground, and average point density is relatively low to differentiate urban objects.
- Hessigheim3D (Kölle et al. 2021): the annotated dataset (11 classes) is a LiDAR point cloud acquired with a UAV platform over the village of Hessigheim (Germany). The dataset stands out for its very high point densities and unusual classes like garbage boxes and chimneys.
- STPLS3D (Chen et al. 2022): it is a synthetic aerial photogrammetric point cloud dataset created by simulating the real pattern of a UAV flight on different synthetic urban and rural areas. The semantic annotation of 6 classes is generated in a fully automated way while rendering the 2D images.
- KITTI-360 (Liao et al. 2022): a large-scale autonomous driving dataset collected by MLS which covers 73.7 km of suburban roads with 1 billion labeled 3D points and 37 annotated classes. However the point clouds are quite sparse, with poor resolution for small or distant objects and they feature different vertical scan patterns.
- HRHD-HK (Li et al. 2023): a photogrammetric point cloud dataset focusing on high-rise, high-density (HRHD) urban scenes in Hong Kong. The dataset covers 9375 km² and contains 273 million colorized 3D points. It includes seven semantic classes.
- YTU3D (Bayrak et al. 2023): it is a UAV photogrammetric point cloud covering the Davutpasa Campus of Yildiz Technical University (Turkey). Data are annotated into 45 classes following a hierarchical multi-level and multi-resolution (MLMR) approach (Teruggi et al. 2020).
- WHU-Urban 3D (Han et al. 2024): the dataset includes both ALS and MLS point clouds (along with street-level panorama images) of two urban areas in China. The MLS dataset has 30 annotated classes, while the ALS dataset has 8. The MLS dataset covers more than 10 km of urban roads divided into 38 scenes. The ALS contains a large annotated urban area divided into 80 blocks of 200 × 200 m. The predominant categories are represented by light poles, followed by traffic signs.
- CUS3D (Gao et al. 2024): it is a UAV photogrammetry dataset which consists of point clouds, aerial images and meshes. The semantic annotation was performed for 10 classes, including unusual categories such as farmland and playground.
- CITYLID (Verma et al. 2025): a large-scale ALS dataset specifically designed for street-level urban analysis in Berlin. The dataset covers the entire city with 15 billion points and features eight urban classification classes. Unlike other datasets, CITYLID includes shadow classes, derived from solar radiation analysis and supports urban shading studies.
- TALD (Vijaywargiya and Ramiya 2025): a benchmark ALS dataset covering 9 km² in Thiruvananthapuram, Kerala, featuring a complex tropical urban landscape with high land-cover diversity of four semantic classes, with an average point density of 12 points/m².

2.3 Under-represented Objects in Urban Areas

Within the point cloud classification framework, an imbalanced dataset features classes with very different numbers of samples, some of them under-represented. The accurate classification of under-represented samples is one of the main problems for machine learning techniques (Rezvani and Wang 2023). Real-world point clouds suffer from class imbalance due to the typical class imbalance observed in the environment. For example, an urban environment consists mostly of buildings, roads, sidewalks, and trees, whereas other objects, such as street furniture and poles, are under-represented (Griffiths and Boehm 2019). To address this challenge, various methods have been proposed in the literature, including data augmentation (Achlioptas et al. 2018; Chen et al. 2020), class weighting (Lin et al. 2017; Griffiths and Boehm 2019; Sander 2020), graph-based (Ma et al. 2024), oversampling/undersampling techniques (Lin and Nguyen 2020; Ren and Xia 2023), and decoupling optimization (Li et al. 2024; Zhang et al. 2024). Even though many approaches have been developed (Li et al. 2020b; Ji et al. 2023; Ye et al. 2025), the correct classification of under-represented classes and the handling of imbalanced classes remain an open research problem (Grilli et al. 2023).

In this study, as a contribution to the challenge of imbalanced classes, we propose two solutions by (I) introducing a novel approach that integrates both semantic segmentation and object classification, and demonstrate its applicability on unseen data, and (II) enhancing classification performance by registering additional objects into the point cloud used for training. The latter could be regarded as some kind of 3D data augmentation are presented in some works (Xiao et al. 2022; Zhu et al. 2024).

2.4 Deep Learning Algorithms for 3D Point Cloud Classification and Semantic Segmentation

The representative approaches for point cloud classification can be categorized into four groups, namely multi-view based, voxel-based, point cloud-based, and polymorphic fusion-based approaches (Zhang et al. 2023) (Fig. 2).

Multi-view-based methods such as MVCNN (Su et al. 2015), MHBN (Yu et al. 2018), PointView-GCN (Mohammadi et al. 2021), PointOfView (Ren et al. 2024), SelectiveMV (Alzahrani et al. 2024), and MSCV (Kim et al. 2025) utilize deep learning techniques with 2D images as input, each referred to as a specific view (Guo et al. 2020). Multi-view-based classification methods comprise three steps:(i) point clouds projection into multiple views, (ii) feature extraction by deep learning, and (iii) classification of 3D point clouds by fusing the extracted features. While leveraging the computational efficiency and relative ease of training offered by 2D convolutional neural networks (CNNs), multi-view-based methods inherently suffer from the loss of intrinsic 3D geometric information and are highly sensitive to the selection of view-points, which significantly impacts classification performance.

Voxel-based approaches such as VoxNet (Maturana and Scherer 2015), Super-Voxel (Lin et al. 2018), VBEC (Kang et al. 2018), VV-Net (Meng et al. 2019), MVPNet (Li et al. 2023), iBALR3D (Zhang et al. 2024), HyperG-PS (Bie et al. 2025) and MSVC (Štroner et al. 2025) involve transforming a 3D point cloud model into voxels that approximates the shape of an object. Each voxel block contains a group of associated points and 3D CNNs are utilized to classify the voxels. While the voxel-based models address the issue of unorder and lack of structure in point cloud data, the sparse and incomplete nature of the data still hampers the efficiency of classification tasks, preventing the full utilization of the information contained in the point cloud. Although the quantization process, similar to that in multi-view based approaches, causes a loss of fine-grained detail, voxel-based methods facilitate efficient spatial operations due to their structured inputs and regularized volumetric grids.

Unlike multi-view and voxel-based approaches, point-based methods such as PointNet (Qi et al. 2017a), PointNet++ (Qi et al. 2017b), MinkowskiNet (Choy et al. 2019), KPConv (Thomas et al. 2019), Point Transformer (Zhao et al. 2021), Point-BERT (Yu et al. 2022), RFFS-Net (Mao et al. 2022), Point Transformer-v2 (Wu et al. 2022), PointContrast (Wu et al. 2023), PointGPT (Chen et al. 2023), Octformer (Wang 2023), MOS-module (Li et al. 2023), Point Transformer-v3 (Wu et al. 2024), PointGT (Zhang et al. 2024), Oneformer3D (Kolodiazhnyi et al. 2024), MDSCNet (Xia et al. 2025), SAPFormer (Xiao et al. 2025), and CyDConv (Mao et al. 2025) prioritize the direct processing of point clouds through deep learning techniques. The feature aggregation operator plays a crucial role in point cloud processing as it facilitates the transfer of information among individual points. By directly processing raw 3D points, point-based methods preserve fine-grained geometric details; however, they are inherently sensitive to noise and face challenges in capturing local features effectively.

Polymorphic fusion-based methods such as PointGrid (Le and Duan 2018), PointCLIP (Zhang et al. 2022), CrossPoint (Afham et al. 2022) combine voxel-based, point-based and multi-view-based ideas and utilize components of these approaches together. Polymorphic fusion-based approaches leverage the complementary strengths of multiple representations, offering enhanced generalization and increased robustness to variations in shape and scale; however, they often entail higher computational overhead and may exhibit limited performance in highly complex regions.

All the above-mentioned methods offer different performances based on the used network/algorithm. Determining an outstanding method is quite challenging due to various factors such as dataset size, number/type of classes, object type and sensor features. These variables should be considered when assessing the performance of different methods and the quality of their outcomes.

3 The ESTATE Dataset

ESTATE (Bayrak et al. 2024) stems from the need to (i) overcome the limitations of available datasets for semantic object classification (Table 1) and (ii) improve the generalization capabilities of neural networks by providing an extensive set of annotated and under-represented urban objects. The dataset encompasses 13 classes of elements, including different kinds of poles, vehicles, and roof items. Collected objects (more than 6000) come from available public datasets released for semantic segmentation tasks, which entail real-world 3D data (except a synthetically generated dataset) acquired with several sensors (LiDAR or photogrammetric-based) and from different perspectives (aerial, UAV-based or terrestrial). The main characteristics of ESTATE are:

1. The inclusion of many objects around the world: ESTATE contains data from China, Canada, France, Germany, Ireland, Italy, Switzerland, Turkey, UK, and USA, along with 6528 real-world urban objects belonging to 13 classes: light pole, traffic light, pole, electrical pole, traffic sign, pylon, cable, garbage box, car, truck, bus, chimney and ventilation (Fig. 3).
2. The richness of sensor sources and geometric resolutions: ESTATE includes data collected with Mobile and Aerial Laser Scanning (MLS/ALS) and UAV-Photogrammetry techniques, featuring different shapes, uneven density, incompleteness and different attributes (RGB and intensity).

Among the outdoor datasets presented in Sect. 2.2 and Table 2, ESTATE includes objects from WHU3D, DublinCity, Paris-Lille3D, TR-MLS, SensatUrban, Swiss3D, STPLS3D, Hessigheim3D and Toronto3D. More-

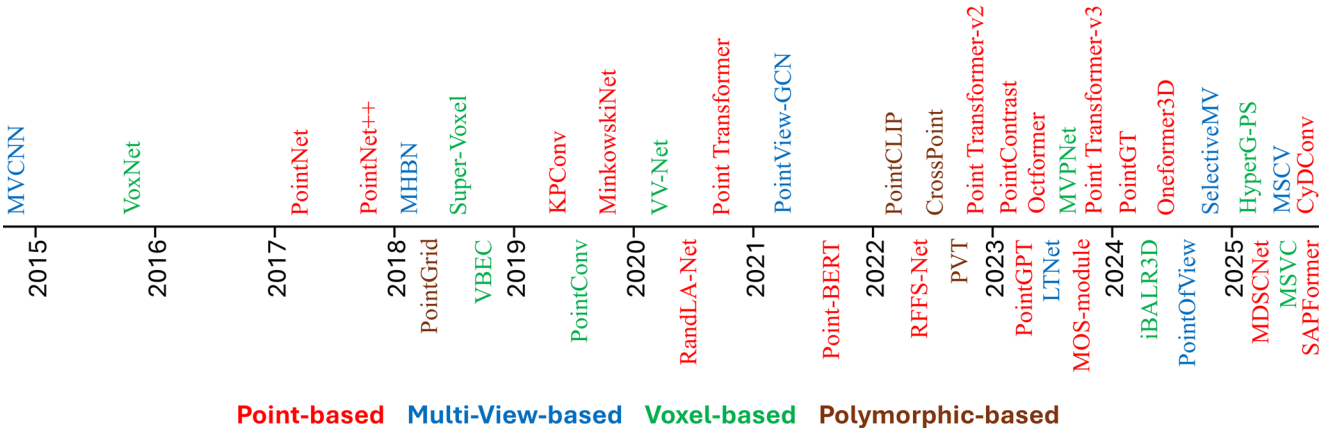

Fig. 2 Timeline with key 3D classification methods

over, data from two in-house datasets are also used: one from Turkey (TR-MLS) and the other from Italy (FBK). Both datasets are LiDAR-based, with the first one acquired through MLS and the second through ALS. They were exploited mainly to derive traffic signs and electricity pylons, respectively.

Although datasets such as SemanticKITTI and KITTI-360 are available, they were not considered since they are tailored for autonomous driving or robotic navigation and thus do not provide sufficiently high spatial resolution of urban objects similarly to other datasets generated from aerial surveys. Moreover, MLS point clouds generally features quite incomplete objects and their inclusion in ESTATE would not support the improvement of identification and classification tasks.

3.1 Data Preparation and Class Statistics

Firstly, for the already pre-classified objects available in the different source datasets mentioned in Table 2, a vis-

Table 2 Specifications of some representative datasets for urban-level semantic segmentation in point clouds.

Name	Classes	Points (mil)	Spatial Size (m2)	RGB	Sensor
ISPRS (Niemeyer et al. 2014)	9	1.2	1.6 × 106	No	ALS
Paris-rue-Madame (Serna et al. 2014)	17	20	0.16 * 103	No	MLS
iQmulus (Vallet et al. 2015)	8 (22)	300	10 × 103	No	MLS
Semantic3D (Hackel et al. 2017)	8	4009	–	No	TLS
Paris-Lille3D (Roynard et al. 2018)	9 (50)	143	1.94 × 103	No	MLS
IEEE-GRSS (Bosch et al. 2019)	5	102	34 × 106	No	ALS
SemanticKITTI (Behley et al. 2019)	22 (28)	4549	39.2 × 103	No	MLS
DublinCity (Zolanvari et al. 2019)	13	260	2 × 106	No	ALS
Toronto3D (Tan et al. 2020)	8	78.3	1 × 103	Yes	MLS
DALES (Varney et al. 2020)	8	505.3	10 × 106	No	ALS
LASDU (Ye et al. 2020)	5	3.12	1.02 × 106	No	ALS
SensatUrban (Hu et al. 2021)	13	2847.1	7.64 × 106	Yes	UAV-Photo
Swiss3DCities (Can et al. 2021)	5	226	2.7 × 106	Yes	UAV-Photo
Campus3D (Li et al. 2020a)	24	937.1	1.58 × 106	Yes	UAV-Photo
OpenGF (Qin et al. 2021)	2	500	47 × 106	No	ALS
Hessigheim3D (Kölle et al. 2021)	11	125.7	8 × 104	Yes	UAV-LiDAR
STPLS3D (Chen et al. 2022)	6	–	6 × 106	Yes	UAV-Photo
KITTI-360 (Liao, Xie et al. 2022)	37	1000	Ca 70 km	No	MLS
HRHD-HK (Li et al. 2023)	7	273	9 × 106	Yes	UAV-Photo
YTU3D (Bayrak et al. 2023)	45	1700	2 × 106	Yes	UAV-Photo
WHU3D (Han et al. 2024)	37	393	6.5 × 103	No	MLS + ALS
CUS3D (Gao et al. 2024)	10	152.3	2.85 × 106	Yes	UAV-Photo
CITYLID (Verma et al. 2025)	9	15000	1060 × 106	No	ALS
TALD (Vijaywargiya and Ramiya 2025)	4	121	9 × 106	No	ALS

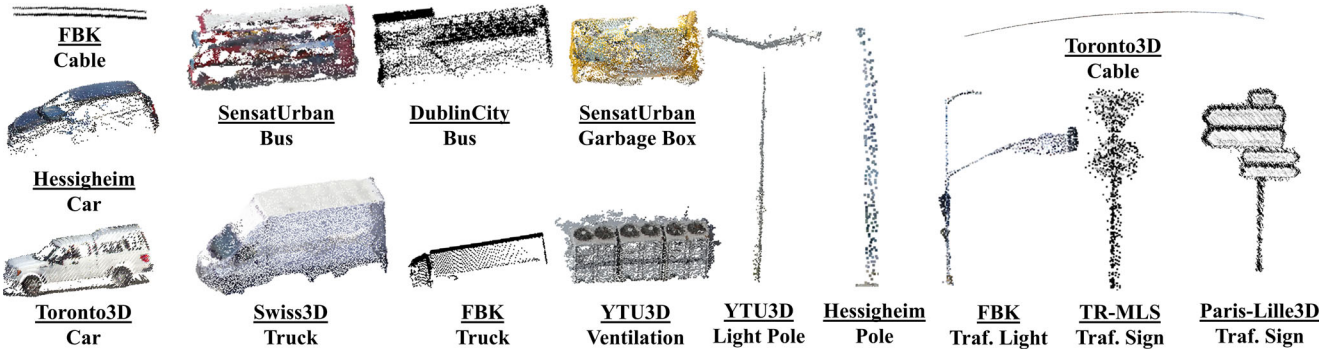

Fig. 3 Examples of some objects included in the ESTATE dataset realized to improve the identification and classification of normally under-represented objects in urban point clouds

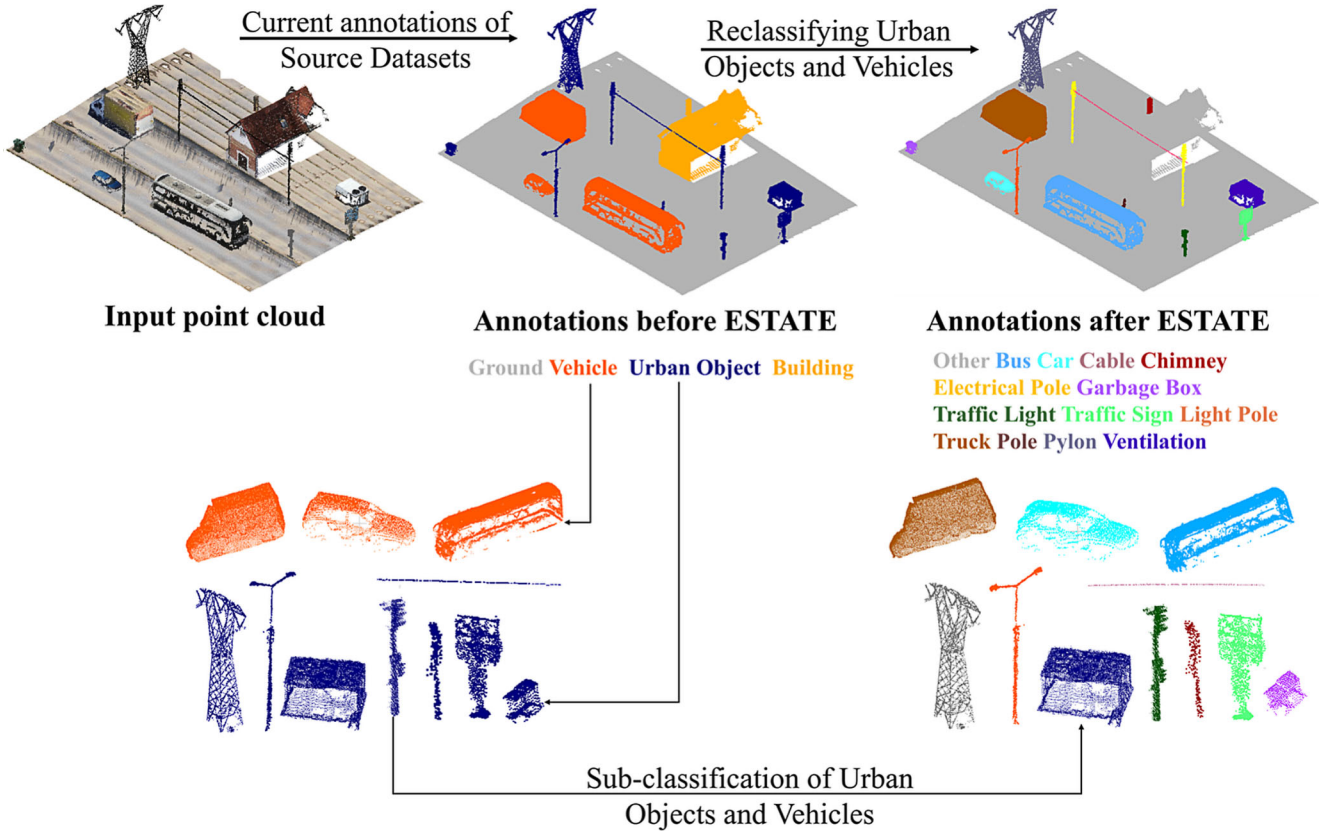

Fig. 4 Creation process of the ESTATE annotated 3D data: source point clouds. Class annotation of source datasets, and detailed final annotations in ESTATE. Please note that the class "building" is not considered as it is not an under-represented object

ual check and manual extraction are executed (Fig. 4) for every single object to guarantee that only distinguishable objects are retained (without removing noise to support the real-case learning of the networks). This operation was especially crucial in those datasets where semi-automatic labelling procedures—more prone to errors—were used to generate annotated data.

Details on the number of objects collected from each dataset and for each class are presented in Table 3. Certain classes contain a greater number of instances, such as cars, light poles, garbage boxes, traffic signs and chimneys. Conversely, some classes encompass fewer instances, including buses, trucks, 'pylons, ventilations, and traffic lights, which are considered minority classes among source datasets. Figure 5 summarizes class statistics based on the dataset features (geometry, intensity and RGB) and the contribution of each dataset to the composition of ESTATE classes. Then, using CloudCompare functionalities such as Segment tool and Cloud Layers plugins, the classes "Urban Object", "Vehicle" and "Other" are split and re-labelled

Table 3 The ESTATE dataset composition: reference datasets and extracted objects per class (visuals in Fig. 3).

Dataset Properties and Classes	XYZ	XYZ + Intensity				XYZ + RGB						Number of Objects in ESTATE
		ALS		MLS		UAV-Photogrammetry				ALS	MLS	
	WHU3D	DublinCity	FBK	Paris-Lille3D	TR-MLS	SensatUrban	Swiss3D	STPLS3D	YTU3D	Hessigheim3D	Toronto3D	
Approx. Point Density (points/m^2)	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*

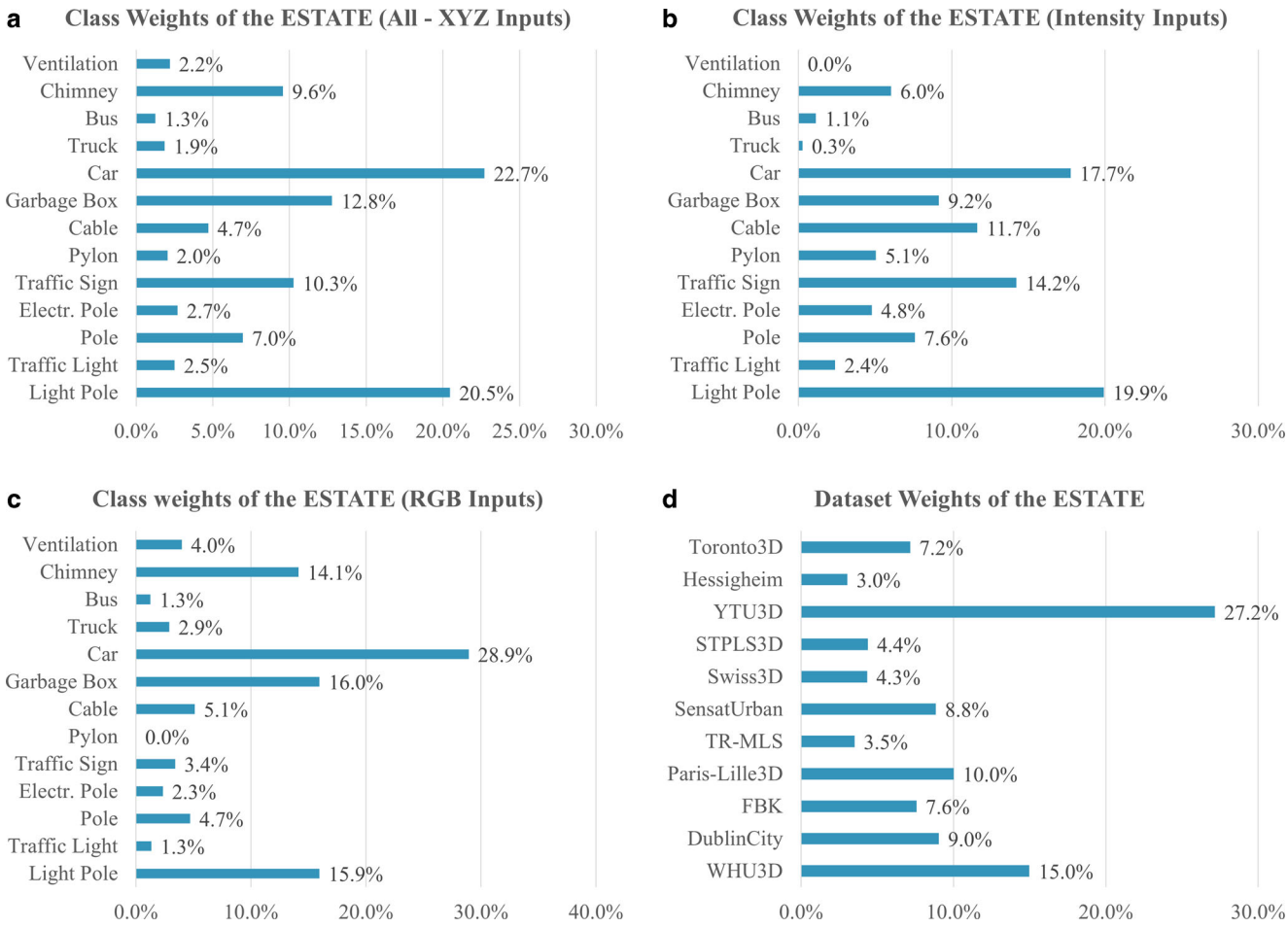

Fig. 5 Class weight statistics for various input configurations: All-XYZ inputs (**a**), XYZ + Intensity inputs (**b**), XYZ + RGB inputs (**c**) and contribution of each dataset to the composition of ESTATE classes (**d**)

into the (sub)classes they belong to. Indeed, the source datasets "Urban Object" and "Vehicle" classes usually comprise poles, lights, traffic lights, cars, busses, etc., whereas "Other" classes usually consist of wires, garbage boxes, etc. Therefore, these objects in the source datasets are extracted to enrich the relevant (sub)classes. For all these preparation steps, CloudCompare and its semi-automatic segmentation plugin Label Connected Component are used to isolate elements and again refine the results manually.

The following criteria were taken into consideration while collecting samples:

- The resolution of the dataset: according to the clarity and distinguishability of the objects;
- The size of the dataset and the presence of target classes: collecting samples of target classes by visual inspection of all regions in the data sets;
- The selection of different object types belonging to the same class: different object types and objects with various noise/deficiencies were selected to avoid merging uniform data.

The 13 classes included in the ESTATE are:

- Light Pole: Traditional, contemporary, architectural, double arm, floodlight, bollard lights, high-mast lights, etc.
- Traffic Light: Vertical and horizontal configuration, single and dual mast arm lights, etc.
- Pole: Security camera, bollard, flagpoles, etc.
- Electrical Pole: Wooden, concrete, guyed, distribution poles, etc.
- Traffic Sign: Round, square, hexagonal, octagonal fluted, telespar, cantilever sign poles, etc.
- Pylon: Lattice towers, H-frame, delta towers, etc.
- Cable: Wires connecting electrical poles and pylons
- Garbage Box: Round, rectangular, elliptical, recycling, residential bins, etc.
- Car: Sedan, hatchback, SUV, crossover, coupe, station wagon, etc.
- Truck: Box, flatbed, dump, tanker, refrigerated, tow truck, garbage, logging, livestock truck, van, etc.
- Bus: Single decker, double decker, articulated, minibuses, etc.

- Chimney: Tile, spiral, polygonal, multiple flue chimneys, etc.
- Ventilation: Air conditioning units outside or on top of buildings.

4 Experiments

In the following sections, the training and testing ratio is reported (Sect. 4.1), while the employed 3D object classification neural networks are explained in Sect. 4.2. The varying input configurations (XYZ, XYZ + Intensity, XYZ + RGB), training/testing strategies and results are described in Sect. 4.3 with further analyses and challenges in Sects. 4.4 and 4.5. Finally we describe how ESTATE can be exploit to improve deep learning procedures for semantic segmentation of under-represented objects (Sects. 4.6 and 4.7).

4.1 Train/test Split

In benchmarks such as Modelnet40, ModelNet40-C, ScanNet, or ScanObjectNN, the training/testing percentage is 80–20% respectively, while in our tests with the ESTATE dataset, we set it as 70–30% to examine the model performance under challenging conditions and with fewer training data. In order to investigate the effect of sensor-based features on classification accuracy, training and testing procedures were repeated considering point clouds featuring XYZ, XYZ + Intensity and XYZ + RGB input configurations and a training/testing ratio as reported in Table 4. In addition, in order to examine the effect of ESTATE data on the generalization capability of the tested neural networks, the training and testing procedures were repeated with 3 different strategies as follows:

- Single-Train Single-Test (STST): the training and testing processes are repeated on the same dataset using the training and test samples from each dataset (e.g., training on DublinCity and testing on DublinCity). This implies 11, 6, and 6 training and testing operations on all three input configurations.
- All-Train Single-Test (ATST): the models trained by using the training set of all datasets are then run on the test set of each dataset (e.g., training with all available sets and testing on DublinCity, FBK, Toronto3D, etc.). This means 1 training process and 11, 6, and 6 testing operations, respectively.

Table 4 Training and testing samples (70%–30%) for the ESTATE dataset.

Input/Samples	XYZ	XYZ + Intensity	XYZ + RGB
Training	4531	1816	2504
Testing	1997	810	1080

- All-Train All-Test (ATAT): training and testing sets of all datasets are used.

Since ATAT comprises a single training and testing operation, a total of 87 training operations and 147 testing processes are executed within our analyses.

4.2 Employed 3D Object Classification Neural Networks

Due to their outstanding achievements in several benchmarks such as DALES, WOD-C and ScanNet, and to ensure reproducibility while guiding researchers in selecting the optimal model, we benchmarked three representative network architectures:

- KPConv (Thomas et al. 2019): it is a point-based method which employs radius neighbourhoods as input and assigns weights based on the spatial arrangement of a limited number of kernel points. KP-CNN is a classification convolutional network consisting of convolutional blocks. The convolutional blocks are organized in a similar manner to bottleneck ResNet blocks (He et al. 2016), using a KPConv instead of the conventional image convolution. The following parameters were set: number of kernel points = 15, initial subsampling distance = 0.01 m, and a convolution radius = 2.5 m. These hyperparameters were selected to balance fine-grained local feature capture and a broader contextual understanding of the objects.
- MinkowskiNet (Choy et al. 2019): it is a voxel-based method which handles high-dimensional, spatially sparse data. The sparse tensors are only evaluated at occupied voxels, which is important for high-dimensional spaces. This approach decreases memory usage and computational time. MinkowskiNet was implemented with the following hyperparameters: the number of channels was set to [32, 64, 128, 256, 256, 128, 96, 96] across the respective layers, and the depth of these layers was defined as [2, 3, 4, 6, 2, 2, 2, 2]. Additionally, a patch size of 32 and a dilation rate of 4 were employed. These settings were chosen to optimize the balance between feature representation and computational efficiency, enabling effective extraction of both local and contextual information.
- OctFormer (Wang 2023): it is a voxel-based method which uses sorted shuffled keys of octrees to partition point clouds into voxels to speed up the computation and extract features. Although this operation produces a reduced receptive field, Octformer applies dilated attention to exponentially increase the receptive field for capturing global details. The number of channels was set to [96, 192, 384, 384], and the number of blocks for these layers

Table 5 STST and ATST F1-Scores obtained from KPConv, Octformer, and Minkowski methods for XYZ, XYZ + Intensity and XYZ + RGB input configurations. Cells marked with "–" indicate that the datasets do not possess the specific attributes required for those configurations and no prediction was performed.

Network	Input	Config	DublinCity	FBK	Paris-Lille3D	TR-MLS	SensatUrban	Swiss3D	STPLS3D	YTU3D	Hessigheim3D	Toronto3D	WHU3D	All
KPConv	XYZ	STST	0.549	0.794	0.910	0.611	0.750	0.603	0.697	0.897	0.948	0.624	0.772	0.85
		ATST	0.614	0.826	0.944	0.664	0.819	0.917	0.781	0.837	0.877	0.857	0.774	
	XYZ+ Intensity	STST	0.534	0.794	0.91	0.49	–	–	–	–	0.81	0.721	–	0.82
		ATST	0.802	0.766	0.953	0.696	–	–	–	–	0.953	0.88	–	
	XYZ+RGB	STST	–	–	–	–	0.831	0.74	0.65	0.796	0.93	0.619	–	0.826
		ATST	–	–	–	–	0.779	0.648	0.624	0.842	0.903	0.836	–	
Octformer	XYZ	STST	0.617	0.828	0.787	0.659	0.814	0.983	0.670	0.889	0.953	0.828	0.792	0.87
		ATST	0.658	0.783	0.783	0.811	0.776	0.908	0.807	0.879	0.930	0.837	0.754	
	XYZ+ Intensity	STST	0.636	0.819	0.787	0.574	–	–	–	–	0.918	0.809	–	0.80
		ATST	0.949	0.86	0.986	0.926	–	–	–	–	0.915	0.87	–	
	XYZ+RGB	STST	–	–	–	–	0.906	0.908	0.89	0.837	0.925	0.809	–	0.83
		ATST	–	–	–	–	0.766	0.753	0.797	0.793	0.927	0.849	–	
Minkowski	XYZ	STST	0.621	0.684	0.794	0.399	0.666	0.818	0.409	0.726	0.728	0.672	0.622	*0.41*
		ATST	0.444	0.320	0.640	0.609	0.480	0.403	0.303	0.427	0.743	0.589	0.474	
	XYZ+ Intensity	STST	0.669	0.77	0.743	0.434	–	–	–	–	0.895	0.724	–	0.77
		ATST	0.829	0.642	0.927	0.77	–	–	–	–	0.945	0.943	–	
	XYZ+RGB	STST	–	–	–	–	0.686	0.445	0.591	0.729	0.752	0.813	–	0.663
		ATST	–	–	–	–	0.761	0.687	0.31	0.715	0.852	0.547	–	

was configured as [2, 2, 18, 2]. Additionally, a patch size of 32 and a dilation rate of 4 were employed.

We augmented the training data at run-time with random anisotropic scaling in the range [0.8, 1.2], noise [0.001] and color (jitter) [0.01]. Train/test processes are performed on an NVIDIA GEFORCE RTX 4090 GPU. For training, the batch size is set to 16, Cross Entropy Loss function and Stochastic Gradient Optimizer are used with an initial learning rate of 10^{-3}, and a momentum of 0.98. The learning rate is set to decrease exponentially, with a chosen exponential decay that guarantees a division by 10 every 100 epochs during training of 200 epochs for each network. F1-Scores calculated from Precision and Recall values are used to assess the performance of the 3D object classification results.

4.3 STST Vs ATST Strategies and Results

Table 5 presents the results obtained using the three different networks reported in Sect. 4.2 (KPConv, Octformer, and Minkowski) and by varying input configurations (XYZ, XYZ + Intensity, XYZ + RGB) and training/testing strategies. These strategies include Single Train Single Test (STST), All Train Single Test (ATST) and All Train All Test (ATAT), with the combined elements forming the ESTATE dataset. The performance analysis across different networks and configurations reveals significant insights into the efficacy of various input combinations and training strategies. The ATST strategy generally improves performance across all networks by leveraging a broader training dataset, thus enhancing the model's generalization capabilities. The ATAT strategy, reported in the final ALL column of Table 5, shows the overall performance across all datasets, reflecting the network's ability to generalize when trained on the combined ESTATE dataset.

The KPConv network, using XYZ input under the STST configuration, shows a wide range of performance across different datasets. For instance, in Paris-Lille3D, the network achieves a high accuracy of 0.910, while in TR-MLS, the accuracy is considerably lower at 0.611. This variability suggests that KPConv's effectiveness with XYZ input is highly dependent on the specific characteristics of the datasets. In the ATST configuration, the overall performance improves. For instance, the performance in datasets such as DublinCity (from 0.549 to 0.614), FBK (from 0.794 to 0.826), Paris-Lille3D (from 0.910 to 0.944), TR-MLS (from 0.611 to 0.664), SensatUrban (from 0.750 to 0.819), Swiss3DCities (from 0.603 to 0.917), STPLS3D (from 0.697 to 0.781), Toronto3D (from 0.624 to 0.857), and WHU3D (from 0.772 to 0.774) indicates that training on the combined ESTATE dataset helps in generalizing better across individual datasets. The overall score for KPConv with XYZ input is 0.85, indicating strong performance but with noticeable variability across datasets. Octformer displays better performance than KPConv, with XYZ input under the STST configuration, especially in datasets like DublinCity (0.617), FBK (0.828), TR-MLS (0.659), SensatUrban (0.814), Swiss3DCities (0.983), Hessigheim3D (0.953), Toronto3D (0.828), and WHU3D (0.792). The high scores in these datasets demonstrate Octformer's capability to effectively utilize geometric information alone. The STST configuration yields an overall score of 0.87, reflecting the network's strong ability to generalize across different datasets even without additional attributes. In the ATST configuration, improvements for DublinCity (from 0.617 to 0.658), TR-MLS (from 0.659 to 0.811), STPLS3D (from 0.670 to 0.807), and Toronto3D (0.828 to 0.837) datasets were obtained. The Minkowski network shows moderate performance with XYZ input, with the lowest scores in several datasets, such as TR-MLS (0.399) and STPLS3D (0.409). However, the scores of DublinCity (0.621), Swiss3DCities (0.818), and Toronto3D (0.672) outperform KPConv results. This indicates that Minkowski may require additional attributes or more comprehensive training data to perform optimally. The overall score for Minkowski with XYZ input is 0.41, highlighting its limitations when relying solely on geometric information.

Adding intensity data slightly improves KPConv's performance in the ATST configuration. The network achieves notable scores such as DublinCity (from 0.534 to 0.802), Paris-Lille3D (from 0.910 to 0.953), TR-MLS (from 0.490 to 0.696), Hessigheim3D (from 0.810 to 0.953), and Toronto3D (from 0.721 to 0.88). This enhancement underscores the value of intensity data in improving the network's generalization capabilities. However, the overall score decreased from 0.85 to 0.82 when using XYZ + Intensity input, demonstrating that the inclusion of intensity attributes leads to lower performance across diverse datasets. Octformer benefits greatly from the addition of intensity data in STST and ATST configuration. Under the STST configuration, it achieves poorer results than XYZ inputs except for DublinCity (0.636). In the ATST configuration, the network performance was improved for DublinCity (from 0.636 to 0.949), FBK (from 0.819 to 0.86), Paris-Lille3D (from 0.787 to 0.986), TR-MLS (from 0.574 to 0.926), and Toronto3D (from 0.809 to 0.87). Despite that, the overall performance decreased from 0.87 to 0.80 after the inclusion of the intensity feature. This indicates that intensity data is beneficial for dataset-basis classification by enhancing its ability to differentiate between different classes more effectively. The ATST configuration further boosts performance, showing how aggregated training data can enhance model generalization. The Minkowski network shows significant improvement with the inclusion of intensity data, particularly in the ATST configuration. For example, performance in DublinCity increases from 0.669

Table 6 Per-class outcomes of highest mean F1-Scores from Table 5.

Class	XYZ (Octformer)				XYZ + Intensity (KPConv)				XYZ + RGB (Octformer)			
	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy
Bus	1.00	0.85	*0.92*	*0.94*	0.75	1.00	0.86	0.88	0.75	0.8	0.77	0.79
Cable	0.97	0.97	0.97	0.98	0.98	0.98	*0.98*	*0.99*	0.96	0.95	0.95	0.96
Car	0.99	0.99	*0.99*	*0.99*	0.99	0.98	0.99	0.99	0.93	0.99	0.96	0.97
Chimney	0.91	0.89	*0.90*	*0.91*	0.89	0.86	0.88	0.89	0.9	0.83	0.86	0.88
Light Pole	0.90	0.97	0.93	0.95	0.96	0.96	*0.96*	*0.97*	0.89	0.92	0.9	0.92
Pole	0.84	0.80	0.82	0.83	0.77	0.88	0.82	0.85	0.9	0.87	*0.88*	*0.91*
Pylon	0.92	0.88	*0.90*	*0.91*	0.97	0.83	0.89	0.94	–	–	–	–
Traffic Light	0.66	0.71	0.69	0.72	0.72	0.62	0.67	0.71	0.69	0.73	*0.71*	*0.74*
Traffic Sign	0.90	0.87	0.89	0.92	0.87	0.83	*0.90*	*0.92*	0.82	0.84	0.83	0.84
Electr. Pole	0.80	0.64	0.71	0.76	0.89	0.82	*0.80*	*0.84*	0.82	0.69	0.75	0.79
Truck	0.92	0.87	*0.89*	*0.92*	0.00	0.00	0.00	0.00	0.81	0.67	0.73	0.8
Garbage Box	0.90	0.92	0.91	0.92	0.98	0.89	*0.94*	*0.95*	0.85	0.9	0.88	0.9
Ventilation	0.76	0.73	0.74	0.76	–	–	–	–	0.83	0.77	*0.8*	*0.81*
Mean	0.88	0.85	0.87	0.89	0.81	0.80	0.81	0.83	0.85	0.83	0.84	0.86

to 0.829, in Paris-Lille3D from 0.743 to 0.942, from 0.434 to 0.77 in TR-MLS, from 0.895 to 0.945 in Hessigheim3D, and from 0.724 to 0.943 in Toronto3D. These improvements lead to the overall performance being increased from 0.41 to 0.77 after the inclusion of the intensity feature. This demonstrates that Minkowski benefits significantly from aggregated training data and intensity features.

The performance of KPConv with XYZ + RGB input is less consistent than XYZ inputs for the STST configuration. This inconsistency shows potential challenges for sensor-specific applications in training with this combination. Under the STST configuration, KPConv scores were improved for just SensatUrban (from 0.750 to 0.831), and Swiss3Dcities (from (0.603 to 0.740), whereas decreased for STPLS3D (from 0.697 to 0.65), YTU3D (0.897 to 0.796), Hessigheim3D (from 0.948 to 0.930), and Toronto3D (from 0.624 to 0.619). Similar situations were obtained for the ATST configuration of XYZ + RGB input, such as SensatUrban (from 0.819 to 0.779), Swiss3Dcities (from 0.917 to 0.648), STPLS3D (from 0.781 to 0.624), and Toronto3D (from 0.857 to 0.836). Octformer performs well with XYZ + RGB input in the STST configuration, in particular for SensatUrban (from 0.814 to 0.906), and STPLS3D (from 0.670 to 0.89). Under the ATST configuration, classification performance was decreased in SensatUrban (from 0.906 to 0.766), Swiss3Dcities (from 0.908 to 0.753), STPLS3D (0.89 to 0.797), and YTU3D (0.837 to 0.793)), respectively. In addition, the overall performance decreased from 0.87 to 0.83 after the inclusion of XYZ + RGB inputs. The results for Minkowski with XYZ + RGB input are sparse, indicating either the absence of RGB data in many datasets or challenges in utilizing this configuration.

It should be noted that the number of training samples for ATST with XYZ input is higher than XYZ + RGB and XYZ + Intensity inputs; therefore, less training data may decrease the classification performance. The ATST strategy significantly improves performance across all networks. By training on the combined ESTATE dataset, the models gain a more comprehensive understanding of the variations and nuances present in different datasets, leading to enhanced generalization and accuracy.

4.4 Class-wise Performance of Different Inputs

Table 6 shows the class-wise performance of the highest average scores obtained from XYZ, XYZ + Intensity and XYZ + RGB input configurations. For the XYZ configuration, Octformer produced the most successful results with the highest scores found in Bus (0.92), Car (0.99), Chimney (0.90), Pylon (0.90) and Truck (0.89) classes compared to Intensity and RGB configurations. For the XYZ + Intensity configuration, the best results are produced for Cable (0.98),

Light Pole (0.96), Traffic Sign (0.90), Electrical Pole (0.80) and Garbage Box (0.94), while only Pole (0.99), Traffic Light (0.71) and Ventilation (0.80) scores are obtained for the XYZ + RGB configuration. It is observed that XYZ input achieves the best results in the detection of volumetric objects such as Bus, Car, Chimney, Pylon and Truck. It is observed that pole-like objects such as Cable, Light Pole, Traffic Sign, Electrical Pole are more successful in XYZ + Intensity configuration, but Pole and Traffic Light objects in the same group gave the best results in XYZ + RGB configurations. On the other hand, it is observed that XYZ + Intensity and XYZ + RGB inputs obtained the best results for Garbage Box and Ventilation, respectively.

According to these results, we can safely say that deep learning methods produce more accurate results with XYZ information in the classification of volumetric objects; on the other hand, since Garbage Box is homogeneous and different in colour from other objects (Fig. 6, c.5) and black and white colours are dominant in Ventilation (Fig. 6, c.3), more successful results are obtained in XYZ + RGB configuration. This information can be used to better distinguish Ventilation, which is geometrically similar to Garbage Box and Chimney classes. Although they are Pole-like objects, Pole and Traffic Light classes achieved the best classification scores in XYZ + RGB configuration. As can be seen in Fig. 6, c.10 and Fig. 6, c.11, it is concluded that colour information is an important discriminating factor since Pole class consists of homogeneous objects and Traffic Light class consists of heterogeneous objects. The fact that the Cable class is classified with high accuracy because it is thinner than the Pole-like objects and curved according to the situation, and the relatively lower score (0.95) in the XYZ + RGB input compared to XYZ (0.97) and XYZ + Intensity (0.98) confirms that colour information is not advantageous in all cases. On the other hand, even though the dominant colours of objects can be distinguished with RGB, the diversity of colours can cause disagreement, which is in line with the findings in (Sun et al. 2022). Since the Bus, Car, Chimney and Truck classes are both volumetric and less similar to other Pole-like objects, it is observed that XYZ information is effective in the detection of these classes. When Fig. 6, c.1 and Fig. 6, c.8 are analysed, it is observed that although these objects do not have homogeneous density, they are classified correctly due to their shape and size differences, but according to Fig. 6, a.2 and Fig. 6, c.4, Bus and Car objects are classified as Garbage Box because different types of noise change the object boundary information. Another finding is that the Pylon object (Fig. 6, c.10), which has a different pattern in terms of size and object type, has achieved the most successful results with the use of XYZ information. This result shows that although it is a sparse Pylon, as can be seen in Fig. 6, b.10, it can be classified correctly because the object dimensions and boundary are generally the same. As can be seen in Fig. 6, a.7, Fig. 6, b.5, Fig. 6, b.8, and Fig. 6, c.7, it is observed that there are misclassifications between Pole and other Pole-like object classes, and the reason for this is that the relatively small differences in Traffic Sign and Electrical Pole objects with short arms may not be detected by deep learning methods. Similarly, in Fig. 6, a.6, It was observed that misclassification occurred when the horizontal arms at the center of the Electrical Pole object were mistakenly predicted as a Light Pole. Due to the change in object boundary information caused by the noise, the Light Pole object in Fig. 6, c.8 was classified as Traffic Light, while the Chimney object in Fig. 6, b.2 was classified as Traffic Sign. In the cases where relatively less noise does not change the object shape, as in Fig. 6, a.11, Fig. 6, b.11 and Fig. 6, c.11, it is concluded that the classification process for Traffic Light is successful. This finding is consistent with the results in Fig. 6, a.9, Fig. 6, b.6, Fig. 6, b.7, Fig. 6, b.9, Fig. 6, b.6, Fig. 6, c.9 and Fig. 6, c.10 and proves that coordinate information plays a key role in 3D object classification.

4.5 Challenging Conditions On Classification Performance

One of the most significant strengths of the ESTATE dataset lies in its realistic representation of point cloud characteristics encountered in real-world urban environments, such as sensor-induced noise, occlusions, and incompleteness, stemming from both data acquisition conditions and the nature of the objects themselves. A key objective of this work is to assess and encourage robustness in deep learning models, enabling accurate object classification even under noisy and incomplete input conditions.

However, these models do not always yield reliable results across all scenarios. The main motivation here is to foster the development of more resilient algorithms within the research community. To contribute toward this goal, we present in Fig. 7 selected failure cases from the KPConv, MinkowskiNet, and OctFormer architectures, demonstrating examples of misclassified urban objects during our experimental evaluations.

As illustrated in Fig. 7.1 through Fig. 7.4, Fig. 7.6, and Fig. 7.10, objects such as buses and cars—despite their geometric similarity to the target classes—were misclassified due to missing side-view data, which caused a loss of discriminative geometric features critical for model inference.

In addition to data incompleteness, noise emerges as a major factor influencing classification accuracy. Figure 7.7, Fig. 7.8, and Fig. 7.9 reveal that in the presence of noise, especially for pole-like structures, even subtle geometric deformations can lead to incorrect predictions. For instance, in Figs. 7.9, although an object belongs to the Pole class and is visually consistent with a straight

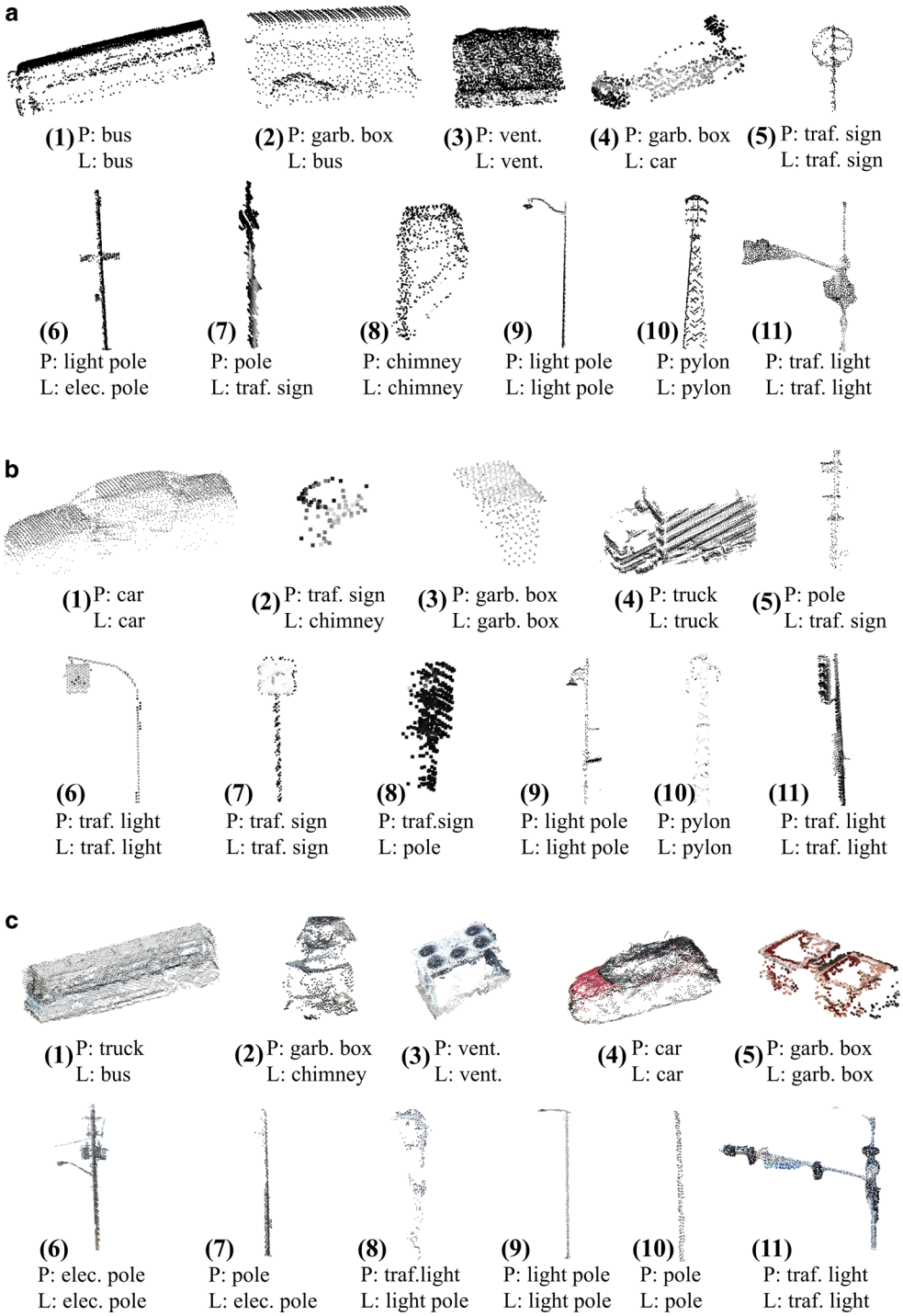

Fig. 6 Sample predictions for XYZ (**a**), XYZ + Intensity (**b**) and XYZ + RGB (**c**) (input configurations. *P* Predicted, L Label.)

cylindrical pole, localized noise along the shaft resulted in erroneous labeling by the network.

It is important to emphasize that geometric corruption is not the sole factor impacting performance. Class imbalance, particularly the under-representation of rare object types in training data, also contributes to decreased classification accuracy. For example, as shown in Fig. 7.11, an object belonging to the Traffic Light class may be mistaken for a Light Pole due to its sharp vertical form, despite differing in the structural details of the signal lamp section. Similarly, Fig. 7.12 shows a Traffic Sign with relatively low noise but an inclined shape, which leads the model to confuse it with a light pole, further highlighting the model's tendency to rely on shape over fine-grained semantics.

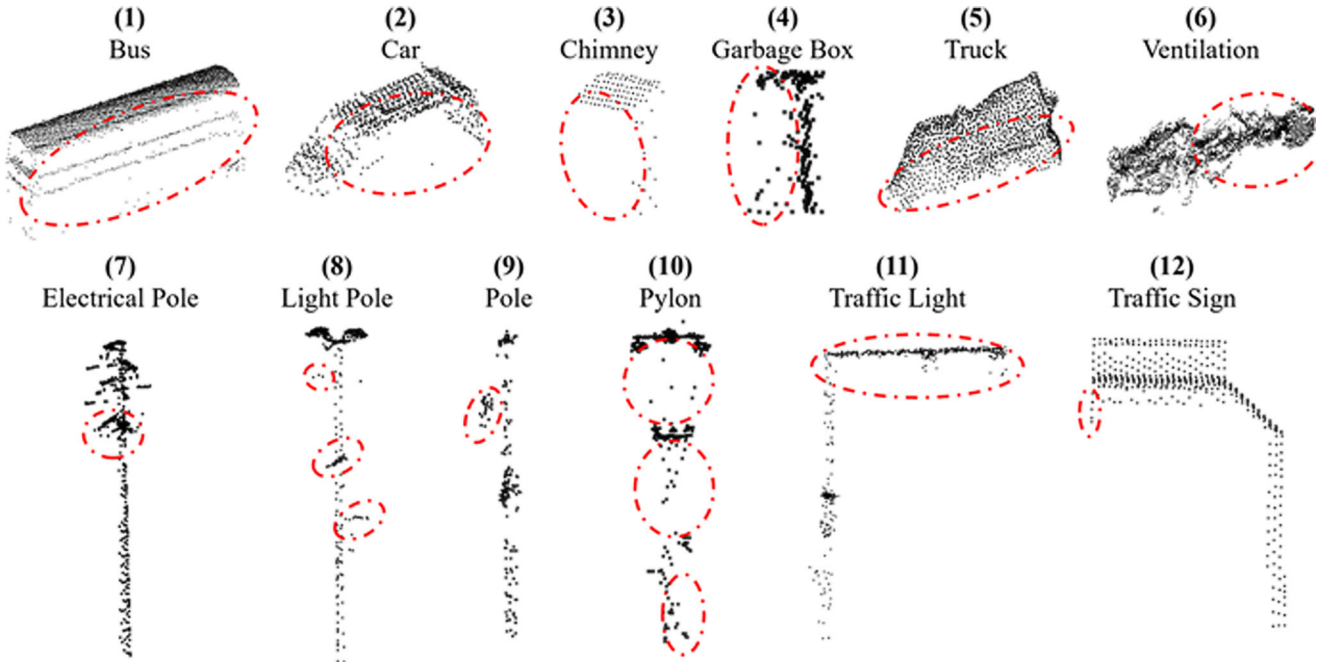

Fig. 7 Objects misclassified by KPConv, Minkowski and Octformer due to noisy or incomplete 3D objects

4.6 How to Use ESTATE for Semantic Segmentation of Under-represented Objects

In order to use the ESTATE dataset for semantic segmentation applications, a three-stage pipeline—tested on the YTU3D dataset—is proposed. This pipeline (Fig. 8) involves:

A. semantic segmentation of the Ground, Building, High Vegetation, Urban Object, and Vertical Surface classes (Fig. 9, a),
B. grouping points classified as Urban Objects (Fig. 9, b) and extracting them by using the Label Connected Component (LCC) plugin in CloudCompare (Fig. 9, c)
C. classifying the grouped points belonging to the Urban Object class using re-trained Octformer weights from the ESTATE dataset (Fig. 9, d).

It should be noted that for the process adapted to the YTU3D dataset, objects belonging to YTU3D were removed from the ESTATE. The primary objective of this application is to explore the applicability of the ESTATE dataset in scenarios where fine-grained annotations for Urban Object subclasses are unavailable. Specifically, the proposed framework aims to infer subclass labels, such as Pole, Light Pole, Traffic Sign (representing slender vertical structures), and Car, Truck, Garbage Box (representing volumetric objects), for instances broadly categorized as Urban Objects in a separate target domain. This is achieved through a class-discriminative deep learning pipeline that enables

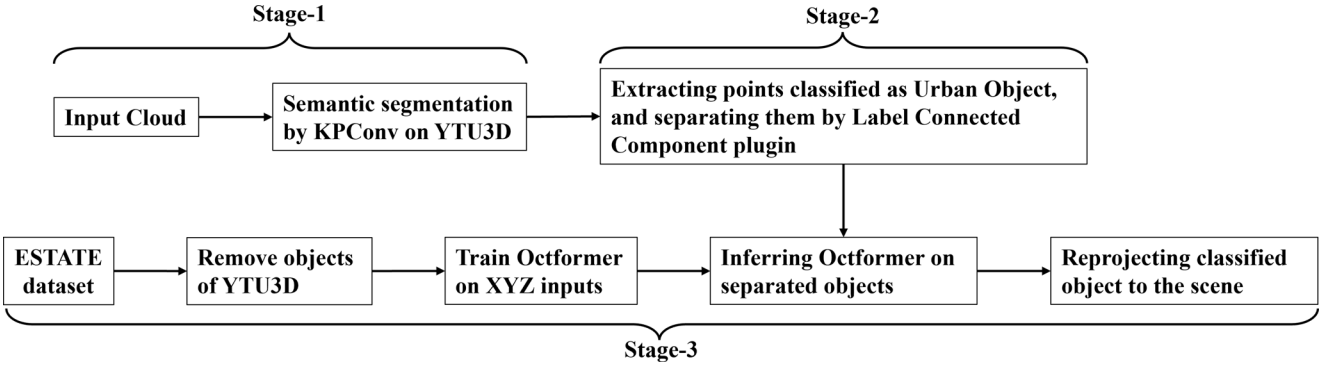

Fig. 8 The steps of proposed pipeline for adapting and using ESTATE for semantic segmentation. Visual of each step are shown in Fig. 9

knowledge transfer and subclass classification beyond the scope of the ESTATE annotations.

Step-1: Semantic Segmentation on the YTU3D Dataset: the 45 classes in the YTU3D dataset were merged into five classes as follows: Ground, Building, High Vegetation, Urban Object, and Vertical Surface (Fig. 10 and 11) shows class distributions after merging operations.

KPConv architecture was configured with 30 kernel points and an input radius of 19.0 m, while the initial subsampling distance was set to 0.2 m and the convolution radius to 2.5 m. For training, in the same way as object classification process, the batch size is set to 3, Cross Entropy Loss function and Stochastic Gradient Optimizer are used with an initial learning rate of 10^{-3}, and a momentum of 0.98. The learning rate is set to decrease exponentially, with a chosen exponential decay that guarantees a division by 10 every 100 epochs during training of 300 epochs.

In order to enable a comparative evaluation of the proposed approach with conventional methods, feature extraction was carried out following the procedures described in studies of optimal neighborhood-based feature extraction (Weinmann et al. 2015). Let $p = (X, Y, Z) \in R^3$ represents a three-dimensional point in a point cloud $P = \{p_o, \ldots, p_n, \}$ with n points. In order to obtain suitable neighborhoods for each point, 3D tensor $S \in R^{3x3}$ can be noted as $S = \frac{1}{h+1} \sum_{i=0}^{h} \left(P_i - \overline{P}\right)\left(P_i - \overline{P}\right)^T$, representing a 3D

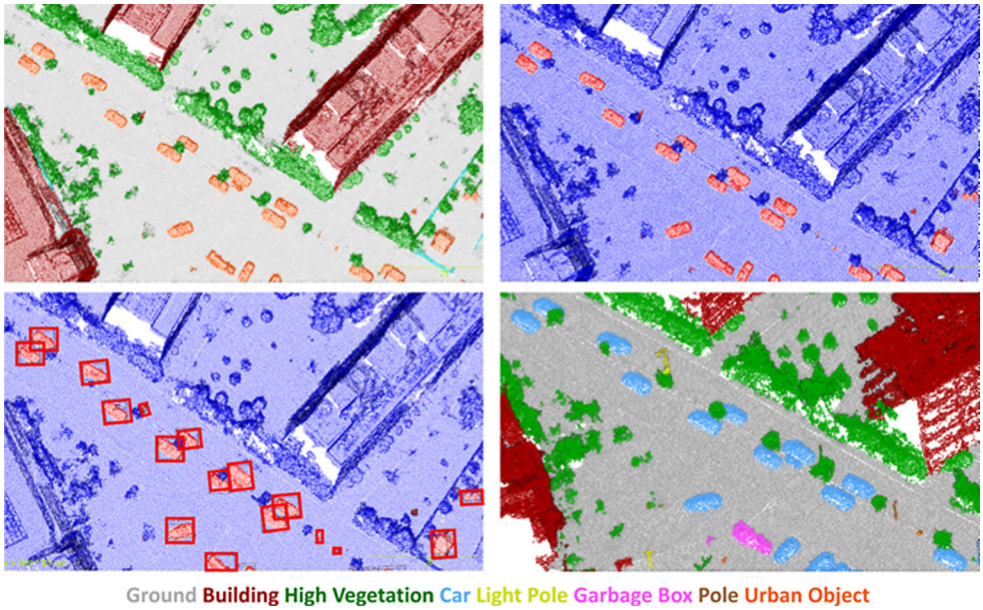

Fig. 9 Point cloud classified with KPConv—note how the class Urban Object includes multiple elements (**a**); Selection of all points classified as Urban Objects (**b**); Separation through Label Connected Components of points assigned to the Urban Objects class (**c**); object classification with OctFormer and reprojection onto—note how the class Urban Objects now is split in multiple elements (**d**)

Ground	Building	Vegetation	Urban Object	Vertical Surface
Stairs	Façade – Ventilation	High Vegetation	Playground	Wall
Tennis Court	Façade – Window	Tree	Traffic Signature	Fence
Sidewalk	Façade – Pipeline		Other	
Street/Road	Façade – Other		Garbage Box	
Parking Lot	Façade Surface		Light Pole	
Low Vegetation	Chimney		Pole	
Pool/Water	Roof – Window		Person	
Soil	Roof – Solar Panel		Pet	
Stone	Roof – Ventilation		Truck	
Disorganized Region	Roof – Pipeline		Van	
Street Separator	Roof – Other		Motorbike/Bcycle	
	Roof – Dome		Bus	
	Roof – Tile		Car	
	Roof – Industrial		Work Machine	
	Ruins			
	Tent			

Fig. 10 Source and merged classes of the YTU3D dataset

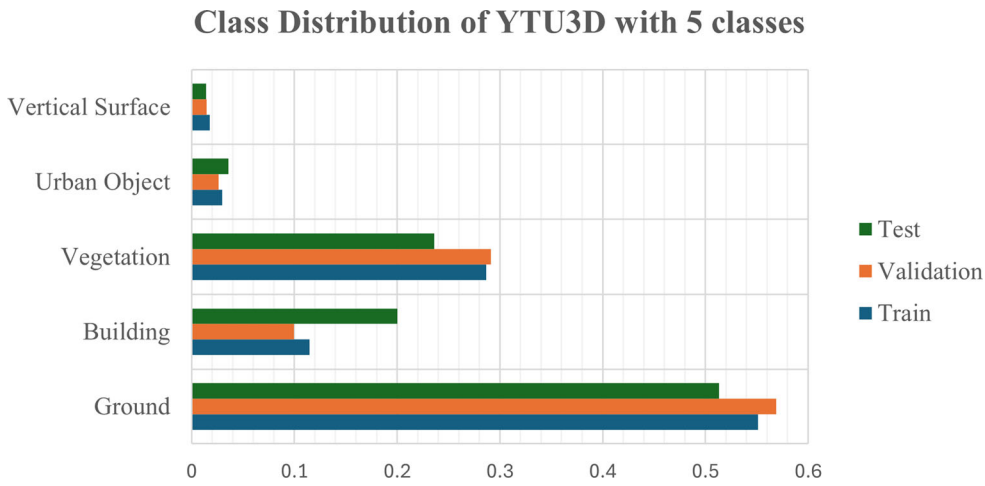

Fig. 11 Class distribution after merging YTU3D into 5 classes

covariance matrix for a given 3D point $P = P_0$, with the inclusion of its h closest neighbors P_i with $i = 1, \ldots, h$. The geometric center $\overline{P}$ is therefore defined as $\overline{P} = \frac{1}{h+1} \sum_{i=0}^{h} P_i$. Since the 3D structure tensor corresponds to an orthogonal system of eigenvectors, the general case of a structure tensor with rank 3 can be denoted by eigen values λ_1, λ_2, andλ_3 with $\lambda_1, \lambda_2, \lambda_3 \in R$, and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ represents the extent of a 3D ellipsoid along principal axes. In this way, the local 3D shape can be characterized by eigen values of the 3D structure tensor. To define the local 3D shape characteristics, as performed in (Blomley et al. 2016; Atik et al. 2021; Duran et al. 2021; Sevgen and Abdikan 2023; Yilmaz et al. 2024): planarity, scattering, omnivariance, anisotropy, eigenentropy, sum of eigen values, change of curvature, density, verticality, maximum height difference, and eigen values were calculated.

Subsequently, Random Forest (RF) (Liang et al. 2025; Breiman 2001) and XGBoost (Ozendi et al. 2023) classifiers were employed for the classification task, due to their outstanding performance in (Weinmann et al. 2015; Atik et al. 2021; Yilmaz et al. 2024).

Metric and visual results obtained are presented in Table 7 and Fig. 12, respectively. The Ground class achieves the highest performance, with an IoU of 89.77% and an F1-Score of 95.12%. Building and Vegetation classes follow closely, showing IoU values of 81.84% and 84.65%, respectively, and F1-Scores near to 90% and 92%. In contrast, Urban Object and Vertical Surface classes exhibit lower performance, likely due to more complex geometry and greater intra-class variability, with IoU values of 61.07% and 41.28% and F1-Scores of 75.85% and 58.65%, respectively. Overall, the mean IoU and F1-Score across all classes are 71.72% and 82.26%.

The RF model performs considerably lower classification performance. The Ground class gets an IoU of 80.70% and an F1-Score of 89.32%, followed by Vegetation and Building classes with IoU values of 68.23% and 46.11%, and F1-Scores of 81.11% and 63.12%, respectively. Performance drops significantly for Urban Object and Vertical Surface, with IoU values of 11.64% and 14.00% and F1-Scores of 20.85% and 24.56%, respectively. The overall mean IoU is 44.14%, and the mean F1-Score is 55.79%. Similarly, the XGBoost model achieves moderate results. The Ground class attains an IoU of 75.62% and an F1-Score of 86.12%. The Building and Vegetation classes score IoUs of 52.18% and 61.41%, and F1-Scores of 68.57% and 76.10%, respectively. Performance for Urban Object and Vertical Surface remains low, with IoU values of 17.75% and 12.66%, and F1-Scores of 30.15% and 22.47%, respectively. The average IoU and F1-Score across all classes are 43.92% and 56.68%.

Table 7 Semantic segmentation results of KPConv, Random Forest and XGBoost on the YTU3D data.

Class	KPConv		Random Forest		XGBoost	
	IoU (%)	F1-Score (%)	IoU (%)	F1-Score (%)	IoU (%)	F1-Score (%)
Ground	89.77	95.12	80.70	89.32	75.62	86.12
Building	81.84	89.97	46.11	63.12	52.18	68.57
Vegetation	84.65	91.73	68.23	81.11	61.41	76.10
Urban Object	61.07	75.85	11.64	20.85	17.75	30.15
Vertical Surface	41.28	58.65	14.00	24.56	12.66	22.47
Mean	71.72	82.26	44.14	55.79	43.92	56.68

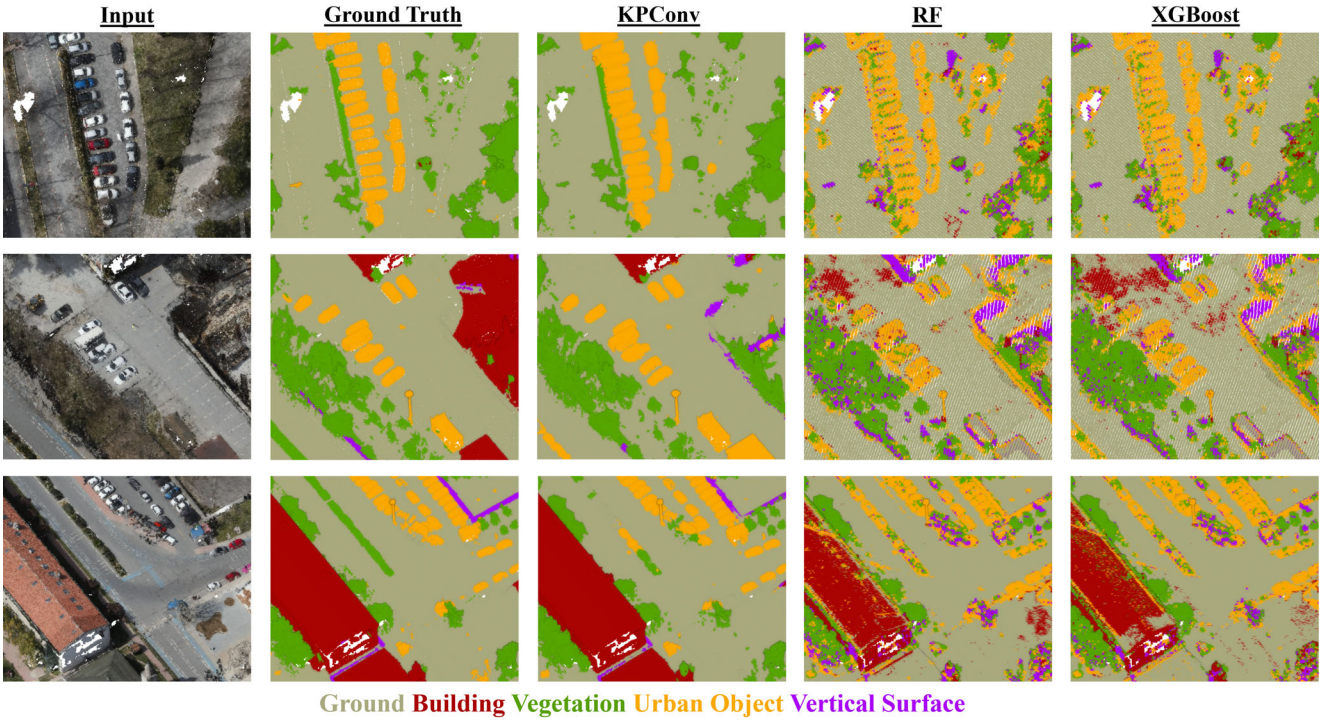

Fig. 12 Visualized segmentation results on YTU3D with KPConv, RF and XGBoost

Analysis of the visual results (Fig. 12) reveals that RF and XGBoost exhibit comparatively effective performance in semantic segmentation but yielding noisy outcomes. Generally, while the segmented objects approximate the ground truth, the outcomes derived from machine learning techniques are excluded from the pipeline due to the classification of points as Vegetation and Vertical Surface within the Urban Object class, which will induce anomalies in object separation during Step-2 of the proposed pipeline.

Nonetheless, a comparison of the KPConv results with the ground truth reveals that inaccuracies predominantly occur between the Ground/Vertical Surface and Vertical Surface/Vegetation classes, indicating that the Urban Object class emphasised in the proposed methodology is comparatively less impacted.

Step-2 and Step-3—Separating the points classified as Urban Object with LCC plugin and classifying separated point clouds with Octformer: since the YTU3D dataset does not include Pylon and Cable classes, and Chimney and Ventilation classes were part of the Building class in Step-1, these objects were excluded from the ESTATE dataset for retraining. Consequently, the training process was conducted solely for the Bus, Car, Electrical Pole, Garbage Box, Light Pole, Pole, Traffic Light, Traffic Sign, and Truck objects. The optimal classification scores were achieved with OctFormer utilising the XYZ configuration, whereby objects distinct from those in YTU3D were employed for training, and the objects delineated by LCC were allocated as the test set. LCC was performed with an octree level of 12 and a minimum of 50 points per component. This procedure was applied to the test regions of the YTU3D dataset, resulting in the extraction of 976 objects. The extracted objects were then subjected to an accuracy assessment in the ESTATE dataset using the corresponding objects from the same regions.

Table 8 displays metric results for different Urban Object classes, highlighting varying levels of classification performance. Notably, the Car class achieves the highest F1-Score (0.932), suggesting that it is more distinctly recognized by the model due to its prevalence in YTU3D. In contrast, Bus (0.51) and Truck (0.541) show lower scores.

Table 8 Object classification results of Octformer with XYZ configuration on YTU3D.

Class	F1-Score	Accuracy
Bus	0.51	0.596
Car	0.932	0.948
Garbage Box	0.706	0.765
Light Pole	0.852	0.87
Pole	0.62	0.698
Traffic Light	0.793	0.822
Traffic Sign	0.761	0.79
Truck	0.541	0.611
Mean	0.714	0.763

Mid-range performances, such as Pole (0.62) and Garbage Box (0.706), indicate moderate model effectiveness, while Light Pole (0.852), Traffic Light (0.793), and Traffic Sign (0.761) demonstrate fairly high classification scores.

Fig. 13 shows the results of the proposed approach. When analyzed under various scenarios, the first case shows that despite KPConv producing noisy outputs, the proposed method successfully classified the Car objects. In contrast, in the second scenario, parking poles located on the road surface were misclassified due to KPConv failing to assign them to the Ground class. Nevertheless, in this case, traffic signs embedded within vegetation were accurately detected by the proposed method.

In the third scenario, which includes regions containing buses, trucks, and garbage boxes, results indicate that KPConv misclassified Urban Object points, especially those overlapping with vegetation, as Ground or Vegetation, leading to the complete misclassification of garbage boxes. Another noted limitation is the presence of points within the Urban Object class that are incorrectly labeled as Vegetation. However, the proposed method mitigated this by accurately distinguishing the object as a whole and correctly assigning it to the Bus class. As for the Truck class, the example illustrates that due to KPConv assigning different parts of the object to Vertical Surface, Urban Object, and Vegetation classes, only the front and rear segments were labeled as Truck, while the cargo section remained incorrectly labeled as Vegetation and Vertical Surface.

Thanks to the proposed workflow, utilizing the ESTATE dataset, Urban Objects located in a completely different region, featuring object structures not included in the model's training, were classified. Although this three-step process relies on semantic segmentation performance and the proximity of target objects, it still holds the potential to offer preliminary insights into how many urban objects of each class are present in a study area.

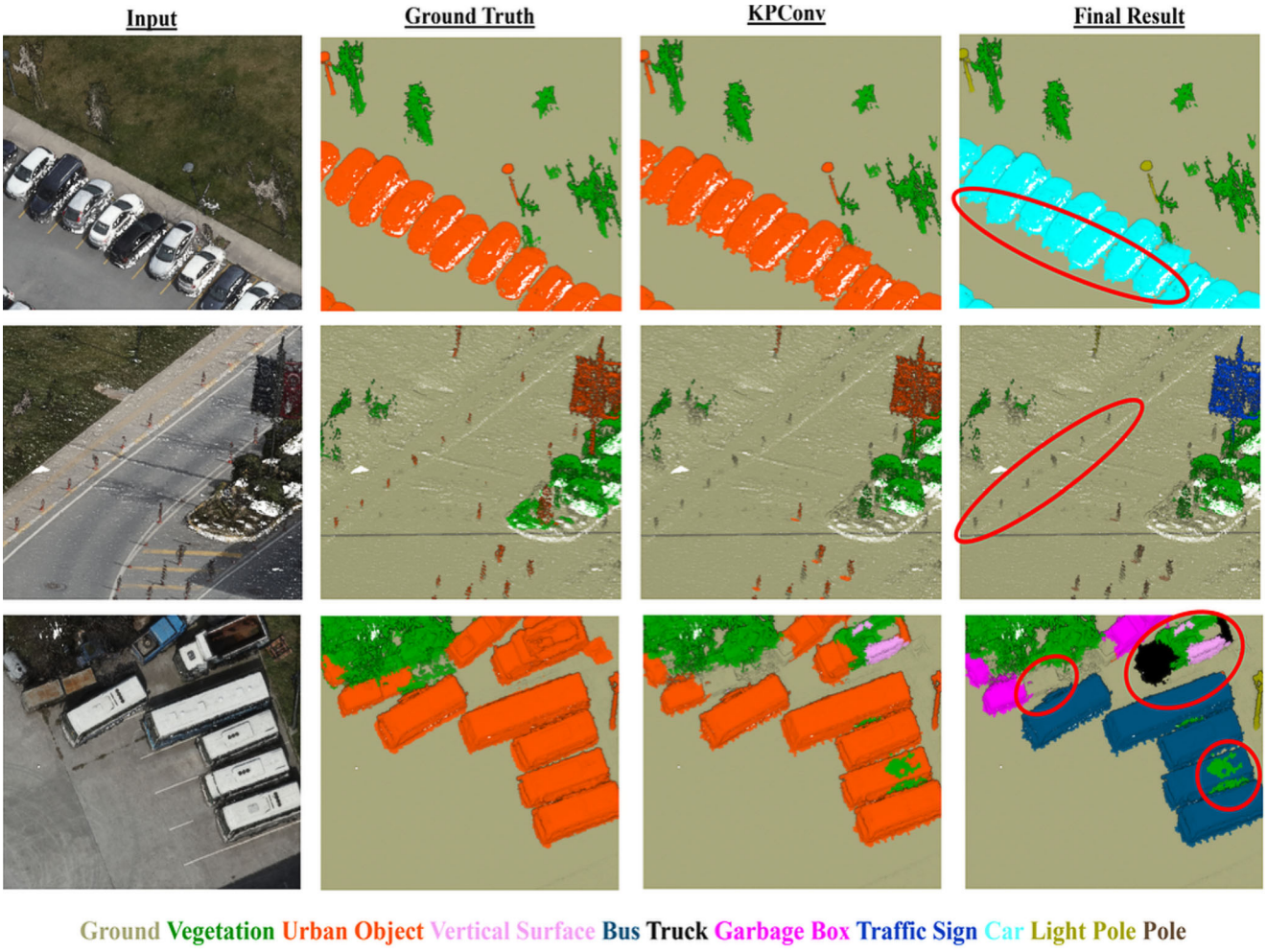

Fig. 13 Segmentation results of the proposed approach able to further segment the scene although inheriting previous errors of the neural network

4.7 How ESTATE Can Improve Semantic Segmentation Performances of Under-represented Urban Objects

This section proofs how ESTATE can be used to improve the semantic segmentation performance of specific classes. The Hessigheim3D dataset (Kölle et al. 2021) and its 'Urban Object' and ‚Vehicle' classes are considered. The primary objective is to investigate the impact of augmenting the training set with diverse object instances available in ESTATE, thereby addressing class imbalance and improving classification performance. Specifically, the approach involves increasing the representation of under-represented classes in the primary dataset through targeted incorporation of ESTATE data.

Semantic segmentation experiments were conducted using KPConv (i) on the original Hessigheim3D dataset and (ii) on an augmented version that includes 40 additional objects from the ESTATE dataset. This augmentation included 20 vehicle instances—positioned on top of the Impervious Surface class—and 15 pole-like objects along with 5 garbage boxes integrated into the Urban Furniture class, situated over Impervious Surface and Low Vegetation areas (Fig. 14). The placement and alignment of these objects within the scenes were performed using the Registration functionality in CloudCompare. The distribution of the training, validation and test classes within the Hessigheim3D dataset, along with the updated distribution after incorporating objects from the ESTATE dataset, is shown in Table 9. As a result of the data augmentation/incorporating process, the distribution of the Vehicle class increased from 0.43% to 1.04%, while the distribution of the Urban Furniture class rose from 1.95% to 2.20%.

By considering Hessigheim3D dataset's density, KPConv architecture was configured with 10 kernel points and an input radius of 5.0 m, while the initial subsampling distance was set to 0.1 m and the convolution radius to 2.5 m. Parameters other than those explicitly mentioned were configured to be consistent with the model and training setup described in Sect. 4.6.

Table 10 Semantic segmentation results (IoU and F1-Scores) of KPConv on Hessigheim3D dataset. Metrics improvements for the two considered classes are notable.

Class	w/o the ESTATE		w/the ESTATE	
	IoU (%)	F1-Score (%)	IoU (%)	F1-Score (%)
Low Vegetation	65.07	78.84	67.98	80.94
Impervious Surface	61.84	78.28	61.21	77.94
Vehicle	32.87	49.47	58.99	74.21
Urban Furniture	31.52	48.08	38.67	55.58
Roof	85.48	92.17	82.96	90.69
Façade	63.97	78.03	65.43	79.11
Shrub	47.31	64.23	48.39	65.22
Tree	91.48	95.55	91.45	95.54
Soil/Gravel	0.04	0.07	23.18	37.64
Vertical Surface	56.88	72.51	56.75	72.41
Chimney	0.090	16.55	0.00	0.00
Mean	48.87	61.25	54.09	66.29

The quantitative results are presented in Table 10, while the qualitative visualizations are provided in Fig. 15. The augmentation of the Vehicle and Urban Furniture classes resulted in an improvement in classification performance for these classes by 26% and 7%, respectively, as shown in Table 9. Additionally, positive impacts were observed across most other classes, with the exception of Impervious Surface, Roof, Chimney, and Tree. However, performance gains in the other classes were relatively modest compared to those in Vehicle and Urban Furniture, particularly for the Impervious Surface class, with improvements generally ranging between 2% and 3%.

As it can be observed in Fig. 15, in the classification results without the incorporation of the ESTATE dataset (Fig. 15.w/o ESTATE), objects intended to belong to the Vehicle class were simultaneously misclassified as both Ve-

Table 9 Class frequency/distribution (%) of the Hessigheim3D data.

Class	Train	Train with the ESTATE	Validation	Test
Low Vegetation	35.96	35.64	25.85	34.32
Impervious Surface	17.53	17.37	22.21	24.10
Vehicle	0.43	1.04	1.27	0.63
Urban Furniture	1.95	2.20	3.15	2.09
Roof	10.56	10.47	21.10	17.03
Façade	2.02	2.00	3.82	2.86
Shrub	1.81	1.80	2.36	1.58
Tree	13.60	13.48	15.34	9.79
Soil/Gravel	14.45	14.32	4.10	6.83
Vertical Surface	1.64	1.63	0.70	0.69
Chimney	0.043	0.042	0.11	0.08

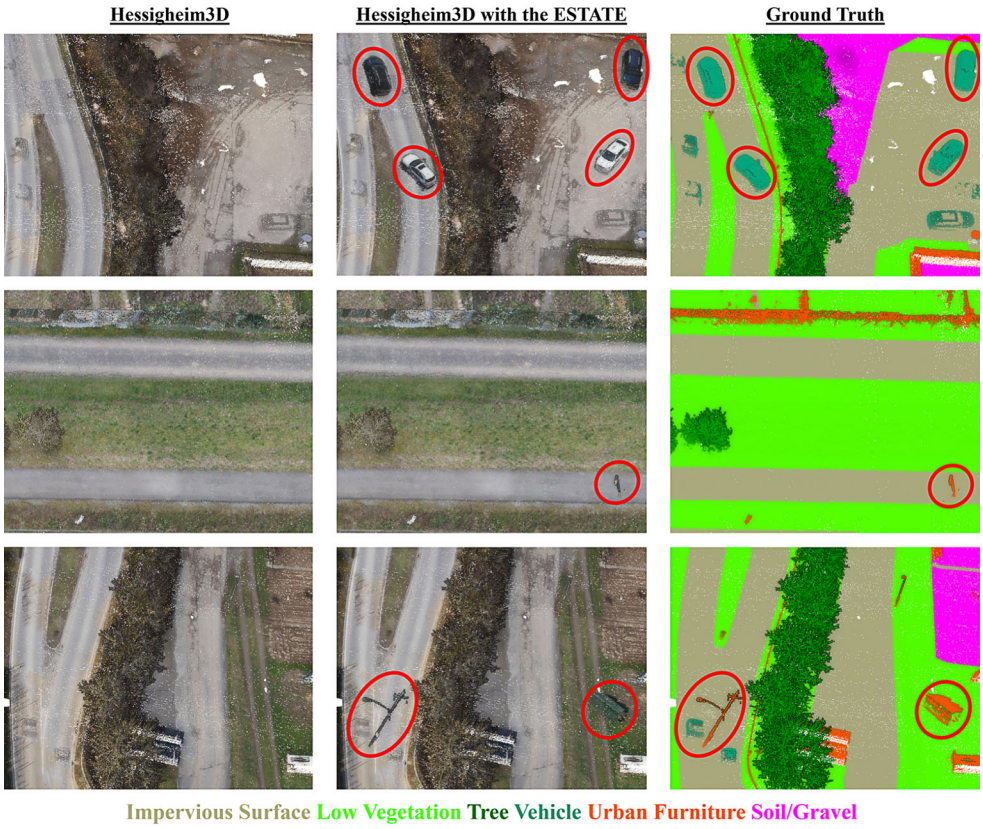

Fig. 14 Original Hessigheim3D data, inclusion of ESTATE objects (vehicles and urban furniture, highlighted with red circles) and annotations

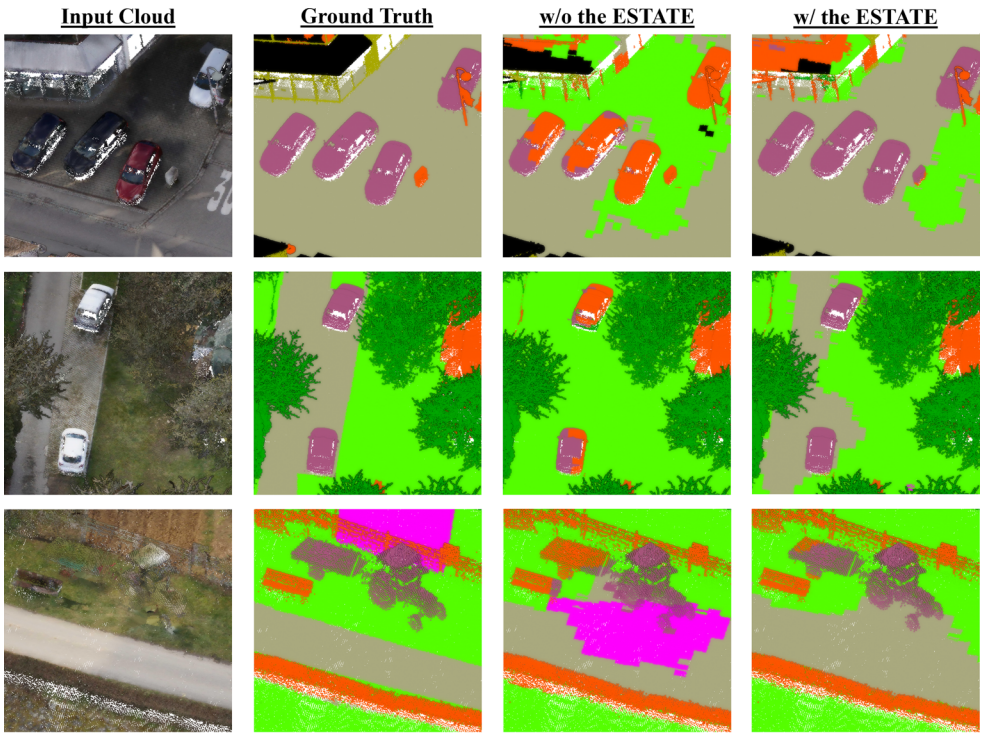

Fig. 15 Results after the inclusion of ESTATE objects for semantic segmentation: Input cloud; Ground truth of test data; KPConv results without the inclusion of ESTATE objects; KPConv results with the inclusion of ESTATE objects

hicle and Urban Furniture. However, following the inclusion of the ESTATE dataset (Fig. 15.w/ESTATE), noticeable improvements in classification performance were achieved. This improvement was observed not only for the Vehicle class but also for the Urban Furniture class, as evidenced in the second and third examples.

5 Discussion

The presented analyses underscore the value of combining multiple datasets and comprehensive training strategies to optimize the performance of neural networks in diverse and complex environments. Incorporating additional features, such as intensity and RGB, proves to be beneficial for improving classification accuracy on a dataset basis. These features provide supplementary information that helps the networks to differentiate between various classes more effectively since they increase the network's ability to capture fine-grained details that are not discernible from geometric information alone. The reported analyses showed that the OctFormer achieved the most successful outcomes through various experiments and input configurations. The principle of partitioning point clouds into local windows and applying dilated attention to exponentially increase the receptive field is providing an effective capturing of global details. These features enable OctFormer to efficiently process both local and global details, resulting in superior performance. The relatively smaller differences in results obtained from KPConv using XYZ coordinate, RGB and Intensity inputs indicate that KPConv is less affected by input features compared to OctFormer and Minkowski. From literature, it is known that KPConv ensures a consistent receptive field by incorporating a deformable operator to learn local shifts and focus on geometric patterns, making it less dependent on additional features such as colour and intensity. On the other hand, Minkowski struggles with learning training data acquired from diverse sources, requiring the use of different features to address this deficiency and enhance classification performance. This issue may stem from the sparse convolution operators causing inconsistencies during the learning phase of datasets with varying densities and noise levels. Even when objects exhibit non-uniform point cloud densities, accurate classification can still be achieved by leveraging distinctive geometric features such as shape and size. This suggests that the model can effectively generalize when clear structural cues are present. The use of XYZ coordinate information significantly improves classification accuracy, especially for sparsely sampled objects. Even when the point cloud density is low, the spatial structure provided by coordinate data allows the model to preserve semantic integrity. Object classes with similar global geometries, such as poles, traffic signs, and electrical poles, often result in mutual misclassifications. The model struggles to capture subtle differences, particularly when distinguishing features (e.g., short arms or attachments) are small or ambiguous. This highlights a limitation in detecting fine-grained structural variations that are critical for differentiating between certain classes.

Misclassifications occur when various types of noise distort the object boundaries, leading to incorrect predictions. This is particularly evident in cases where visually similar objects are assigned to the wrong class due to minor boundary perturbations. Such errors suggest a need for more robust boundary-preserving representations or denoising strategies during preprocessing.

The object classification outcomes, observed during the adaptation of the ESTATE dataset for semantic segmentation, demonstrate that models trained on this dataset exhibit strong generalization capabilities, effectively handling classification tasks regardless of variations in (i) sensor modality, (ii) spatial resolution and (iii) object geometry.

In the reported experiments, it has been observed that missing points and noise in the objects, despite the preservation of their overall geometric structure, can still lead to misclassification. Implementation of denoising methods such as (Ozendi et al. 2023; de Silva Edirimuni et al. 2024; Liang et al. 2025) or point cloud completion approaches such as (Zhang et al. 2021; Fei et al. 2025) could be useful to avoid classification errors based on geometrical issues.

Furthermore, we have shown that misclassification is not solely attributable to data incompleteness or noise (Fig. 7). A contributing factor is the under-representation of certain object types in the training dataset, which causes these instances to be mistakenly classified as more frequent or geometrically similar classes.

The reported learning-based classification approaches (Sect. 4.6; Table 8) were trained using conventional handcrafted features. The relatively limited representational capacity of these features, compared to the complex, high-level feature representations extracted by deep learning models, may account for the observed performance gap.

An additional point of concern is that deep learning methods operate directly on 3D point clouds using convolutional filters, enabling them to capture semantic relationships between neighboring objects. In contrast, traditional handcrafted features lack contextual or spatial neighborhood class information, which imposes a fundamental limitation on the performance of machine learning classifiers in this domain.

Finally, the incorporation of objects from the ESTATE dataset into the training set led to an improvement in classification performance. Although this integration was performed manually through registration, augmenting the training set with additional instances of under-represented classes such as 'Vehicle' and 'Urban Object' can effectively

enhance semantic segmentation performance in scenarios where class imbalance is present.

6 Conclusions

The paper extensively presented the ESTATE dataset and how it can support the achievement of better scores with deep learning methods in 3D classification tasks for urban scenes featuring various under-represented classes. Results showed how the meticulous collection of the 13 ESTATE classes from publicly available point cloud datasets enhances the performances of 3D object classification models. As ESTATE contains diverse point cloud characteristics (e.g. density, intensity, etc.) of urban objects, it can be adapted to various sensor-specific applications. It is foreseen that it can be useful in improving classification processes with deep learning methods, especially when objects feature relatively low geometric differences. In order to train models on the ESTATE dataset, only the coordinate values of the objects can be considered since the results show that XYZ input configuration offers the highest classification score. Nevertheless, the importance of different input configurations for various urban objects is emphasized and the feasibility of which input configurations can be applied according to the purpose is examined.

Future studies may (i) evaluate the performances of other neural networks, (i) assess the classification of objects against noise and other data-acquisition problems, (iii) extend the ESTATE dataset including further generally under-represented urban objects or (iv) use ESTATE for instance or panoptic segmentation, where objects in complex urban areas can be extracted using traditional pre-processing methods or unsupervised learning. Using highly effective semantic segmentation methods may improve classification results by better-determining object boundaries. In fact, combining the same classes from different datasets have enhanced the networks generalization capability, similar to the benefit provided by the variety of the ESTATE dataset in improving the network learning performance.

The employed data and research findings are publicly available at https://github.com/3DOM-FBK/ESTATE.

Funding The R&D activities of Onur Can Bayrak are supported by the The Scientific and Technological Research Council of (TUBITAK) 2214-A Programme with 1059B142201684 project code. The work is also 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).

Author Contribution Conceptualization, F.R. and O.C.B.; methodology, O.C.B. and Z.M; software, O.C.B. and Z.M.; validation, O.C.B., Z.M. and E.M.F. ; investigation, O.C.B., E.M.F., Z.M., F.R.; resources, F.R., M.U.; data curation, O.C.B., E.M.F., Z.M., F.R.; writing—original draft preparation, O.C.B., Z.M., E.M.F.; writing—review and editing, F.R., M.U.; visualization, O.C.B., E.M.F., Z.M.; supervision, F.R.. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Availability of data and material The shared data and research findings are publicly available at https://github.com/3DOM-FBK/ESTATE.

Declaration

Conflict of interest O.C. Bayrak, Z. Ma, E.M. Farella, F. Remondino and M. Uzar declare that they have no competing interests.

References

Achlioptas P, Diamanti O, Mitliagkas I, Guibas L (2018) Learning representations and generative models for 3d point clouds. International conference on machine learning, PMLR.

Afham M, Dissanayake I, Dissanayake D, Dharmasiri A, Thilakarathna K, Rodrigo R (2022) Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Alzahrani M, Usman M, Anwar S, Helmy T (2024) Selective multi-view deep model for 3d object classification. Proceedings of the ieee/cvf conference on computer vision and pattern recognition.

Atik ME, Duran Z, Seker DZ (2021) Machine learning-based supervised classification of point clouds using multiscale geometric features. ISPRS Int J Geo Inf 10(3):187

Bai Q, Lindenbergh R, Vijverberg J, Guelen J (2021) Road type classification of MLS point clouds using deep learning. Int Arch Photogramm Remote Sens Spatial Inf Sci 43:115–122

Bakuła K, Mills J, Remondino F (2019) A review of benchmarking in photogrammetry and remote sensing. Int Arch Photogramm Remote Sens Spatial Inf Sci 42:1–8

Bayrak OC, Remondino F, Uzar M (2023) A new dataset and methodology for urban-scale 3D point cloud classification. Int Arch Photogramm Remote Sens Spatial Inf Sci 48:1–8

Bayrak OC, Ma Z, Farella EM, Remondino F, Uzar M (2024) Estate: a large dataset of under-represented urban objects for 3D point cloud classification. Int Arch Photogramm Remote Sens Spatial Inf Sci 48:25–32

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.

Bie L, Xiao G, Li Y, Gao Y (2025) HyperG-PS: Voxel correlation modeling via hypergraph for LiDAR panoptic segmentation. Fundam Res. https://doi.org/10.1016/j.fmre.2024.03.033

Bloembergen D, Eijgenstein C (2021) Automatic labelling of urban point clouds using data fusion. arXiv preprint arXiv:2108.13757

Blomley R, Jutzi B, Weinmann M (2016) Classification of airborne laser scanning data using geometric multi-scale features and different neighbourhood types. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 3:169–176

Bosch M, Foster K, Christie G, Wang S, Hager GD, Brown M (2019) Semantic stereo for incidental satellite images. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE

Breiman L (2001) Random Forests. Mach Learn 45:5–32

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 Recognit Lett 150:108–114

Chang AX, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, Savarese S, Savva M, Song S, Su H (2015) Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012

Chen G, Wang M, Yang Y, Yu K, Yuan L, Yue Y (2023) Pointgpt: Auto-regressively generative pre-training from point clouds. Adv Neural Inf Process Syst 36:29667–29679

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

Chen Y, Hu VT, Gavves E, Mensink T, Mettes P, Yang P, Snoek CG (2020) Pointmixup: augmentation for point clouds. Computer Vision-ECCV 2020: 16th European Conference, Glasgow, August 23–28, 2020. Proceedings, Part III 16. Springer

Choy C, Gwak J, Savarese S (2019) 4d spatio-temporal convnets: Minkowski convolutional neural networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Dai A, Chang AX, 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.

De Deuge M, Quadros A, Hung C, Douillard B (2013) Unsupervised feature learning for classification of outdoor 3d scans. Australasian conference on robitics and automation. University of New South Wales, Kensington

Deitke M, Schwenk D, Salvador J, Weihs L, Michel O, VanderBilt E, Schmidt L, Ehsani K, Kembhavi A, Farhadi A (2023) Objaverse: A universe of annotated 3d objects. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Duran Z, Ozcan K, Atik ME (2021) Classification of photogrammetric and airborne lidar point clouds using machine learning algorithms. Drones 5(4):104

Fei B, Li Y, Yang W, Chen W-M, Li Z (2025) Multi-modality consistency for point cloud completion via differentiable rendering. IEEE Trans Artif Intell. https://doi.org/10.1109/tai.2025.3527922

Gao L, Liu Y, Chen X, Liu Y, Yan S, Zhang M (2024) CUS3D: A new comprehensive urban-scale semantic-segmentation 3D benchmark dataset. Remote Sens 16(6):1079

Griffiths D, Boehm J (2019) Weighted point cloud augmentation for neural network training data class-imbalance. arXiv preprint 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 Spatial 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 Sens 15(10):2590

Guo Y, Wang H, Hu Q, Liu H, Liu L, Bennamoun M (2020) Deep learning for 3d point clouds: a survey. IEEE Trans Pattern Anal Mach Intell 43(12):4338–4364

Hackel T, Savinov N, Ladicky L, Wegner JD, Schindler K, Pollefeys M (2017) Semantic3d. net: A new large-scale point cloud classification benchmark. arXiv preprint arXiv:1704.03847

Han X, Liu C, Zhou Y, Tan K, Dong Z, Yang B (2024) WHU-Urban3D: an urban scene LiDAR point cloud dataset for semantic instance segmentation. ISPRS J Photogramm Remote Sens 209:500–513

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

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.

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.

Iliopoulou P, Feloni E (2022) Spatial modelling and geovisualization of house prices in the greater Athens region, Greece. Geographies 2(1):111–131

Ismail M, Shaker A, 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 48:7–14

Ji A, Zhang L, Fan H, Xue X, Dou Y (2023) Dual attention-based deep learning network for multi-class object semantic segmentation of tunnel point clouds. Autom Constr 156:105131

Kang Z, Yang J, Zhong R, Wu Y, Shi Z, Lindenbergh R (2018) Voxel-based extraction and classification of 3-D pole-like objects from mobile LiDAR point cloud data. IEEE J Sel Top Appl Earth Observations Remote Sensing 11(11):4287–4298

Kim Y, Cho B, Ryoo S, Lee S (2025) Multi-view structural convolution network for domain-invariant point cloud recognition of autonomous vehicles. arXiv preprint arXiv:2501.16289

Kölle M, Laupheimer D, Schmohl S, Haala N, Rottensteiner F, Wegner JD, 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 J Photogramm Remote Sens 1:100001

Kolodiazhnyi M, Vorontsova A, Konushin A, Rukhovich D (2024) Oneformer3d: One transformer for unified point cloud segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Le T, Duan Y (2018) Pointgrid: A deep network for 3d shape understanding. Proceedings of the IEEE conference on computer vision and pattern recognition.

Li H, Guan H, Ma L, Lei X, Yu Y, Wang H, Delavar MR, Li J (2023) MVPNet: A multi-scale voxel-point adaptive fusion network for point cloud semantic segmentation in urban scenes. Int J Appl Earth Obs Geoinf 122:103391

Li M, Wu Y, Yeh A, Xue F (2023) HRHD-HK: A benchmark dataset of high-rise and high-density urban scenes for 3D semantic segmentation of photogrammetric point clouds. 2023 IEEE international conference on image processing challenges and workshops (ICIPCW).

Li M, Lin S, Wang Z, Shen Y, Zhang B, Ma L (2024) Class-imbalanced semi-supervised learning for large-scale point cloud semantic segmentation via decoupling optimization. Pattern Recognit 156:110701

Li Q, Zhuang Y, Huai J (2023) Multi-sensor fusion for robust localization with moving object segmentation in complex dynamic 3D scenes. Int J Appl Earth Obs Geoinformation 124:103507

Li X, Li C, Tong Z, Lim A, Yuan J, Wu Y, Tang J, Huang R (2020a) Campus3d: A photogrammetry point cloud benchmark for hierarchical understanding of outdoor scene. Proceedings of the 28th ACM International Conference on Multimedia.

Li X, Wang L, Wang M, Wen C, Fang Y (2020b) DANCE-NET: Density-aware convolution networks with context encoding for airborne LiDAR point cloud classification. ISPRS J Photogramm Remote Sens 166:128–139

Liang G, Cui X, Yuan D, Zhang L, Yang R (2025) An improved point cloud filtering algorithm applies in complex urban environments. Remote Sens 17(8):1452

Liang Z, Guo Y, Feng Y, Chen W, Qiao L, Zhou L, Zhang J, Liu H (2019) Stereo matching using multi-level cost volume and multi-scale feature constancy. IEEE Trans Pattern Anal Mach Intell 43(1):300–315

Liao Y, Xie J, Geiger A (2022) Kitti-360: a novel dataset and benchmarks for urban scene understanding in 2d and 3d. IEEE Trans Pattern Anal Mach Intell 45(3):3292–3310

Lin H-I, Nguyen MC (2020) Boosting minority class prediction on imbalanced point cloud data. Appl Sci 10(3):973

Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision.

Lin Y, Wang C, Zhai D, Li W, Li J (2018) Toward better boundary preserved supervoxel segmentation for 3D point clouds. ISPRS J Photogramm Remote Sens 143:39–47

Ma Z, Bayrak OC, Remondino F (2024) Automatic point cloud classification of under-represented pole-like objects based on hierarchical directed graph. IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE

Mao Y, Chen K, Diao W, Sun X, Lu X, Fu K, Weinmann M (2022) Beyond single receptive field: a receptive field fusion-and-stratification network for airborne laser scanning point cloud classification. ISPRS J Photogramm Remote Sens 188:45–61

Mao Y-Q, Bi H, Li X, Chen K, Wang Z, Sun X, Fu K (2025) Twin deformable point convolutions for airborne laser scanning point cloud classification. ISPRS J Photogramm Remote Sens 221:78–91

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

Meng H-Y, Gao L, Lai Y-K, Manocha D (2019) Vv-net: Voxel vae net with group convolutions for point cloud segmentation. Proceedings of the IEEE/CVF international conference on computer vision.

Mohammadi SS, Wang Y, Del Bue A (2021) Pointview-gcn: 3d shape classification with multi-view point clouds. 2021 IEEE International Conference on Image Processing (ICIP). IEEE

Niemeyer J, Rottensteiner F, Soergel U (2014) Contextual classification of lidar data and building object detection in urban areas. ISPRS J Photogramm Remote Sens 87:152–165

Özdemir E, Remondino F (2018) Segmentation of 3D photogrammetric point cloud for 3D building modeling. Int Arch Photogramm Remote Sens Spatial Inf Sci 42:135–142

Ozendi M, Akca D, Topan H (2023) A point cloud filtering method based on anisotropic error model. Photogramm Rec 38(184): 460–497

Qi CR, Su H, Mo K, Guibas LJ (2017a) Pointnet: deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition.

Qi CR, Yi L, Su H, Guibas LJ (2017b) Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Adv Neural Inf Process Syst 30:

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. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Ren P, Xia Q (2023) Classification method for imbalanced LiDAR point cloud based on stack autoencoder. Electron Res Arch 31(6)

Ren H, Wang J, Yang M, Velipasalar S (2024) Pointofview: a multimodal network for few-shot 3d point cloud classification fusing point and multi-view image features. Proceedings of the ieee/cvf conference on computer vision and pattern recognition.

Rezvani S, Wang X (2023) A broad review on class imbalance learning techniques. Appl Soft Comput 143:110415

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. Int J Rob Res 37(6): 545–557

Sander R (2020) Sparse data fusion and class imbalance correction techniques for efficient multi-class point cloud semantic segmentation. https://doi.org/10.13140/RG.2.2.12077.03042

Sarker S, Sarker P, Stone G, Gorman R, Tavakkoli A, Bebis G, Sattarvand J (2024) A comprehensive overview of deep learning techniques for 3D point cloud classification and semantic segmentation. Machine Vis Apps 35(4):67

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.

Sevgen E, Abdikan S (2023) Classification of large-scale mobile laser scanning data in urban area with LightGBM. Remote Sens 15(15):3787

de Silva Edirimuni D, Lu X, Li G, Wei L, Robles-Kelly A, Li H (2024) Straightpcf: straight point cloud filtering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Štroner M, Boušek M, Kučera J, Váchová H, Urban R (2025) Multi-size Voxel cube (MSVC) algorithm—A novel method for terrain filtering from dense point clouds using a deep neural network. Remote Sens 17(4):615

Su H, Maji S, Kalogerakis E, Learned-Miller E (2015) Multi-view convolutional neural networks for 3d shape recognition. Proceedings of the IEEE international conference on computer vision.

Sun J, Zhang Q, Kailkhura B, Yu Z, Xiao C, Mao ZM (2022) Benchmarking robustness of 3d point cloud recognition against common corruptions. arXiv preprint arXiv:2201.12296

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.

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 Sens 12(16):2598

Thomas H, Qi CR, Deschaud J-E, Marcotegui B, Goulette F, Guibas LJ (2019) Kpconv: Flexible and deformable convolution for point clouds. Proceedings of the IEEE/CVF international conference on computer vision.

Uy MA, Pham Q-H, Hua B-S, Nguyen T, Yeung S-K (2019) Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. Proceedings of the IEEE/CVF international conference on computer vision.

Vallet B, Brédif M, Serna A, Marcotegui B, Paparoditis N (2015) TerraMobilita/iQmulus urban point cloud analysis benchmark. Comput Graph 49:126–133

Varney N, Asari VK, 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.

Verma D, Mumm O, Carlow VM (2025) CITYLID: A large-scale categorized aerial lidar dataset for street-level research. Environ Plan B Urban Anal City Sci. https://doi.org/10.1177/23998083241312273

Vijaywargiya J, Ramiya AM (2025) Trivandrum aerial LiDAR dataset (TALD): a benchmark for complex urban point cloud. IEEE Access. https://doi.org/10.1109/access.2025.3546628

Wang P-S (2023) Octformer: Octree-based transformers for 3d point clouds. ACM Trans Graph 42(4):1–11

Wang F, Li W, Xu D (2021) Cross-dataset point cloud recognition using deep-shallow domain adaptation network. IEEE Trans Image Process 30:7364–7377

Weinmann M, Jutzi B, Hinz S, Mallet C (2015) Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS J Photogramm Remote Sens 105:286–304

Wu X, Lao Y, Jiang L, Liu X, Zhao H (2022) Point transformer v2: Grouped vector attention and partition-based pooling. Adv Neural Inf Process Syst 35:33330–33342

Wu X, Wen X, Liu X, Zhao H (2023) Masked scene contrast: a scalable framework for unsupervised 3d representation learning. Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition.

Wu X, Jiang L, Wang P-S, Liu Z, Liu X, Qiao Y, Ouyang W, He T, Zhao H (2024) Point transformer v3: Simpler faster stronger. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J (2015) 3d shapenets: a deep representation for volumetric shapes. Proceedings of the IEEE conference on computer vision and pattern recognition.

Xia P, Tian S, Yu L, Fan X, Zhu Z, Dong H, Qu N, Liu T, Yuan X (2025) Mdcsnet: multi-scale dynamic spatial information fusion with criticality sampling for point cloud classification. J Supercomput 81(2):387

Xiao A, Huang J, Guan D, Cui K, Lu S, Shao L (2022) PolarMix: a general data augmentation technique for LiDAR point clouds. Proc. NIPS, pp 11035–11048

Xiao G, Ge S, Zhong Y, Xiao Z, Song J, Lu J (2025) SAPFormer: Shape-aware propagation Transformer for point clouds. Pattern Recognit 164:111578

Xie Y, Tian J, Zhu XX (2020) Linking points with labels in 3D: a review of point cloud semantic segmentation. IEEE Geosci Remote Sens Mag 8(4):38–59

Xue F, Lu W, Chen Z, Webster CJ (2020) From LiDAR point cloud towards digital twin city: clustering city objects based on gestalt principles. ISPRS J Photogramm Remote Sens 167:418–431

Ye L, Xiao W, Weng Q (2025) Supervoxel-based instance segmentation of pole-like facilities from mobile laser scanning data using pyramid cascaded fisher vector modeling. IEEE Trans Geosci Remote Sens

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

Yilmaz Y, Bayrak OC, Soycan A (2024) Evaluation of machine learning algorithms for classification of infrastructure elements in complex structures. KSCE J Civ Eng 28(8):3489–3505

Yu T, Meng J, Yuan J (2018) Multi-view harmonized bilinear network for 3d object recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.

Yu X, Tang L, Rao Y, Huang T, Zhou J, Lu J (2022) Point-bert: Pre-training 3d point cloud transformers with masked point modeling. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Zhang H, Wang C, Tian S, Lu B, Zhang L, Ning X, Bai X (2023) Deep learning-based 3D point cloud classification: a systematic survey and outlook. Displays 79:102456

Zhang H, Wang C, Yu L, Tian S, Ning X, Rodrigues J (2024) Pointgt: a method for point-cloud classification and segmentation based on local geometric transformation. IEEE Trans Multimedia. https://doi.org/10.1109/tmm.2024.3374580

Zhang K, Cai R, Wu X, Zhao J, Qin P (2024) iBALR3D: imBalanced-aware long-range 3D semantic segmentation. Computer Sciences & Mathematics Forum, MDPI.

Zhang R, Guo Z, Zhang W, Li K, Miao X, Cui B, Qiao Y, Gao P, Li H (2022) Pointclip: point cloud understanding by clip. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Zhang S, Wang B, Chen Y, Zhang S, Zhang W (2024) Point and voxel cross perception with lightweight cosformer for large-scale point cloud semantic segmentation. Int J Appl Earth Obs Geoinf 131:103951

Zhang X, Feng Y, Li S, Zou C, Wan H, Zhao X, Guo Y, Gao Y (2021) View-guided point cloud completion. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Zhao H, Jiang L, Jia J, Torr PH, Koltun V (2021) Point transformer. Proceedings of the IEEE/CVF international conference on computer vision.

Zhu Q, Fan L, Weng N (2024) Advancements in point cloud data augmentation for deep learning: a survey. Pattern Recognit 153:110532

Zolanvari S, Ruano S, Rana A, Cummins A, Da Silva RE, Rahbar M, Smolic A (2019) DublinCity: Annotated LiDAR point cloud and its applications. arXiv preprint arXiv:1909.03613