

A 3D BUILDING INDOOR-OUTDOOR BENCHMARK FOR SEMANTIC SEGMENTATION

Yuwei Cao*, Marco Scaioni

Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, via Ponzio 31, 20133, Milan, Italy - (yuwei.cao, marco.scaioni)@polimi.it

KEY WORDS: 3D Building Benchmark, Point Cloud, Mesh, Indoor Outdoor Dataset, Random Forest

ABSTRACT:

Both machine learning (ML) and deep learning (DL) algorithms require high-quality training samples as well as precise and thorough annotations in order to work effectively. The 3D building indoor-outdoor dataset (BIO dataset), which is a highly accurate, high level of detail, and high coverage dataset for 3D building point cloud and mesh semantic segmentation, is established as a canonical benchmark dataset. It contains 100 building models, in which building structural elements are annotated into 11 semantic categories. Each building in this dataset has an average of 75,587 triangular faces, and the total area of the dataset is 481,769 square meters. Furthermore, semantic segmentation of the dataset was carried out using the Random Forest ML algorithm to verify the dataset's accessibility. A weighted F1 score of 96.64% was obtained with 10% of the segments of each building randomly chosen as training data. For applications involving building geometry data, the BIO dataset can support a broad class of recently developed ML and DL methodologies.

1. INTRODUCTION

Recent developments in artificial intelligence (AI) have demonstrated great promise for enabling a variety of applications that need an accurate and thorough understanding of complex environments, including indoor navigation (Isikdag et al., 2013), autonomous driving, energy efficiency (O'Donnell et al., 2019), disaster response (Nikoohemat et al., 2020), cultural heritage building digitalization (Cao et al., 2022), and sustainable urban planning (Schrotter and Hürzeler, 2020). High-quality datasets with precise and thorough annotations are crucial for the training and testing of AI models for these applications.

Both machine learning (ML) and deep learning (DL) algorithms require millions of training samples to work properly (Géron, 2022). Unlike 2D images that comfortably exist on the Internet, collecting real-world 3D scene datasets usually involves traversing the environment in real life and scanning with 3D sensors. Therefore, the number of building scenes that can be scanned might be limited, and current 3D indoor-only or outdoor-only labeled building datasets are often limited in their coverage, diversity, and accuracy, hindering the development of new AI applications that require a detailed understanding of complex indoor-outdoor environments.

There has been substantial growth in the number of 3D models available online over the last decade, with repositories like the Trimble 3D Warehouse providing millions of 3D polygonal models covering thousands of object and scene categories. In this paper, we present a new building indoor-outdoor dataset (BIO dataset) consisting of 100 labeled building models in both mesh and point cloud formats. The motivation behind this work is to provide a comprehensive and accurate dataset that can enable new AI applications that require a detailed understanding of complex indoor-outdoor environments. To generate our indoor-outdoor labeled building dataset, we first collected 3D building models from online repositories such as 3D Warehouse. These

models were then pre-processed (e.g., data format conversion, mesh repair, etc.) using an automatic pipeline.

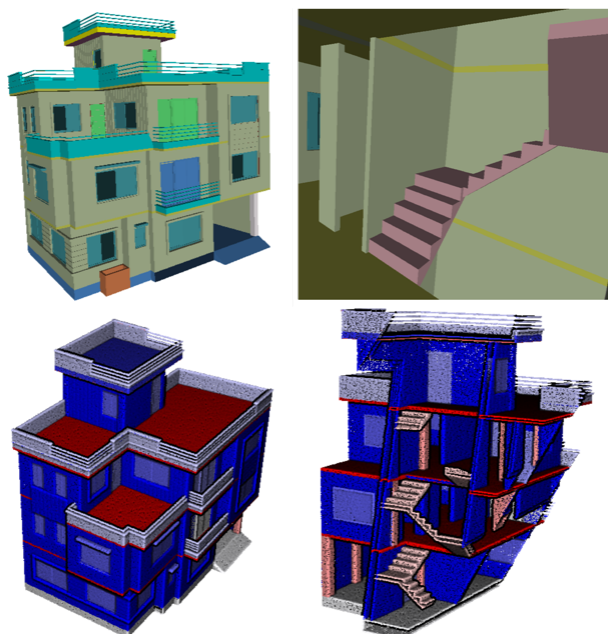


Figure 1. An example of BIO Dataset. Top row is labelled mesh of a residential building, left is the outside, while right is the indoor part. Bottom row is labelled point cloud of the same building, left the exterior part, and right is a slice of indoor part.

After the models were pre-processed, they were manually labeled and checked for accuracy. The labeling process involved identifying and labeling polygons of the building models as indoor or outdoor structural elements in 11 categories, including wall, roof, column, door, ceiling, windows, balcony, floor, stairs,

* Corresponding author

slab, and beam. The labeling was done by using an existing annotation platform (Gao et al., 2022) that was designed for urban dataset labeling to ensure accuracy and consistency across the dataset.

To facilitate the use of the dataset in AI applications, we also generate point cloud samples from the labeled meshes using a uniform sampling method. This allowed us to represent the complex geometry of the buildings in a more efficient and manageable format. As shown in Figure 1, the resulting dataset contains both mesh and point cloud versions of each building model, along with their corresponding indoor and outdoor labels. The use of automated mesh repair and point cloud sampling, combined with manual labeling and checking, ensures a high level of detail, accuracy, and consistency in the dataset.

We also propose a pipeline with the state-of-the-art ML method, Random Forest (Breiman, 2001), to evaluate the availability of the established dataset. When we trained classifiers with the extracted geometric features (Weinmann et al., 2017) based on the covariance matrix and annotations of a portion of each building, our averaged results reached 96.64% in terms of the F1 score, which demonstrates the feasibility and effectiveness of the proposed pipeline. In addition, the results also suggest that the RF algorithm can be combined with new online models to quickly generate larger and more diverse datasets, enabling the dataset to be scaled up in the future.

In summary, our new indoor-outdoor labeled building dataset and pipeline can enable new indoor-outdoor AI applications that require accurate and detailed understanding of complex environments. By providing a large-scale, richly annotated dataset, we can also promote a broad class of recently resurgent machine learning and neural network methods for applications dealing with geometric data.

2. LITERATURE REVIEW

Buildings in urban scenes, indoor scene datasets, and building exterior datasets are a few different types of 3D building datasets that are frequently used for AI applications. The following list includes several widely used datasets:

- The Stanford Large-Scale 3D Indoor Spaces (S3DIS) dataset (Armeni et al., 2016) contains labeled point cloud data for 271 rooms in six indoor spaces. The S3DIS dataset offers a sizable, high-quality dataset for tasks involving object recognition and understanding indoor scenes. However, it excludes outdoor scenes and only depicts indoor environments.
- There are 4 billion points in the Semantic3D (Hackel et al., 2017) dataset that have semantic labels for outdoor urban scenes. It consists of a variety of building types, including residential, commercial, and industrial structures. When compared to indoor datasets, it offers useful information about the outdoor environment but is labeled in LoD1 with insufficient detail.
- ScanNet (Dai et al., 2017): The ScanNet dataset consists of 1,500 labeled 3D scans of indoor environments and textured meshes, as well as pointwise semantic labels and 3D object instance labels. The ScanNet is frequently used for indoor scene understanding tasks like semantic segmentation and object recognition because it has high-quality

annotations and comprehensive coverage of actual indoor objects.

- DublinCity is the first labeled dataset of high-density aerial laser scanning (ALS) point clouds at the city scale (SM Iman Zolanvari et al., 2019). This dataset, a benchmark in the field of computer vision, consists of over 260 million manually annotated point clouds. Using hierarchical levels of detail, objects are classified into 13 classes, ranging from the coarse level of buildings, vegetation, and ground to the fine level of windows, doors, and trees.
- SensatUrban (Hu et al., 2018): SensatUrban is an extensively used urban-scale photogrammetric dataset acquired by unmanned-aerial vehicle (UAV) cameras. This dataset includes three cities in the UK and covers an area of 7.6 km². The availability of this dataset can facilitate the development of new applications and services that will be useful to researchers and urban planners alike.
- BuildingNet (Selvaraju et al., 2021): BuildingNet is a sizable collection of 2K building models and 513K annotated mesh primitives in 3D building models with consistently labeled exteriors. The dataset includes data regarding a variety of building types, such as homes, churches, town halls, and castles. BuildingNet dataset makes an important contribution to building understanding.

For researchers working on building-related AI applications, these datasets collectively provide a useful resource. These datasets have a few coverage limitations. To be more precise, many existing 3D building datasets only include a small number of buildings or only include interior or exterior scenes, which may not accurately represent the entire range of building types and environments.

3. DATASET

In this section, we outline the process used to create the annotation pipeline for defining, collecting, processing, labeling, and evaluating the BIO dataset. The primary objective that inspired the development of our framework was to enable, with the assistance of the defined pipeline, the quick generation of semantically labeled meshes and point clouds of indoor and outdoor building scenes (see Figure 2). As a result, the framework must be easy to use, the data pre- and post-processing techniques must be reliable and automatic, and the semantic annotation process must be straightforward and quick. The following subsections provide descriptions of the details.

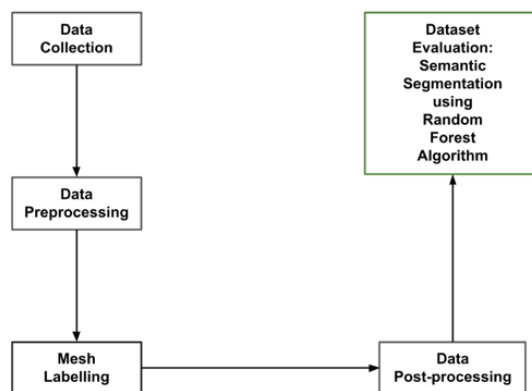


Figure 2. Annotating Pipeline.

3.1 Building Types

As depicted in Figure 3, the constructed BIO dataset consists of 100 indoor-outdoor building models, with each model representing one of the four most typical building types according to the building's intended use:

1. Residential buildings are those used by residents as living space, such as single-family houses (e.g., bungalows and cottages), multi-family houses, and public official residences.
2. Commercial buildings are those occupied by businesses, including office buildings, shopping centers, and some special purpose buildings like theaters.
3. Industrial buildings are used primarily for the manufacturing, storage, and distribution of goods. Buildings such as manufacturing plants, warehouses, and storage facilities.
4. Institutional buildings are mainly built for public use, such as medical spaces, educational facilities, libraries, religious premises, and government places (e.g., city halls).

When collecting building models, the scales of buildings and orientation information are checked and corrected in the SketchUp Software.

3.2 Annotations

Although semantic annotation can be applied to all different kinds of architectural elements, at this point we specifically focus on the enrichment of structural elements. A common semantic information model for the representation of 3D urban objects is defined by the CityGML Conceptual Model Standard and can be used by various applications. Furthermore, IFC (ISO 16739-1:2018) is a standardized, digital description of the built environment, including buildings and civil infrastructure. It is an open, global standard that is intended to be vendor-neutral, or agnostic, and usable across a wide range of hardware devices, software platforms, and interfaces for many different use cases, enabling faster and more effective utilization. The semantic annotations of our BIO dataset are identified in accordance with CityGML 3.0 (Kutzner et al., 2020) and IFC standards (ISO, 2018) to emphasize the reusability of information within lifecycle thinking. In addition, the classes included in the ArCH dataset (Matrone et al., 2020) and the indoor S3DIS dataset (Armeni et al., 2016) were taken into account to identify the semantic annotations in our study. As a result, 11 classes — *wall*, *roof*, *window*, *door*, *balcony*, *floor*, *stairs*, *column*, *ceiling*, *beam*, and *slab* — have been selected.

3.3 Dataset Preprocessing

Prior to semantic enrichment, a series of automatic data preprocessing steps were performed. The PyMeshLab (Muntoni and Cignoni, 2023) and Trimesh (Dawson-Haggerty, 2022) libraries are employed in this step. Specifically, these steps include data format conversion, material and texture information transfer and geometric error repair.

3.4 Dataset Annotating

In 3D point clouds, pointwise annotation will take an enormous amount of time and effort, so we use the UrbanMeshAnnotator (Gao et al., 2021) as our annotation tool to directly label buildings

on the mesh. This tool, however, was made specifically for labeling massive urban mesh scenes. As a result, we customized this tool for our dataset. For instance, the original system requires the input of a manifold mesh, which is then segmented using a region-growing algorithm (Lafarge and Mallet, 2012), and the segments are finally labeled. However, 1) the majority of the building models downloaded from online repositories do not comply with the manifold geometry requirement; 2) the effectiveness of the current manifold repair algorithm is not ensured; and 3) the texture information is not preserved.

As a result, we changed the tool's semantic annotations to conform to our specifications. As a result, our annotation system does not require a manifold mesh or over-segmented segments. In our study, we directly labeled the faces of the input 3D model without the segmentation step, taking advantage of the fact that each face in a mesh belongs to only one category and the faces of building models are very easy to label. Figure 4 depicts a sample of annotated meshes. As you can see, each category is represented by a different color.

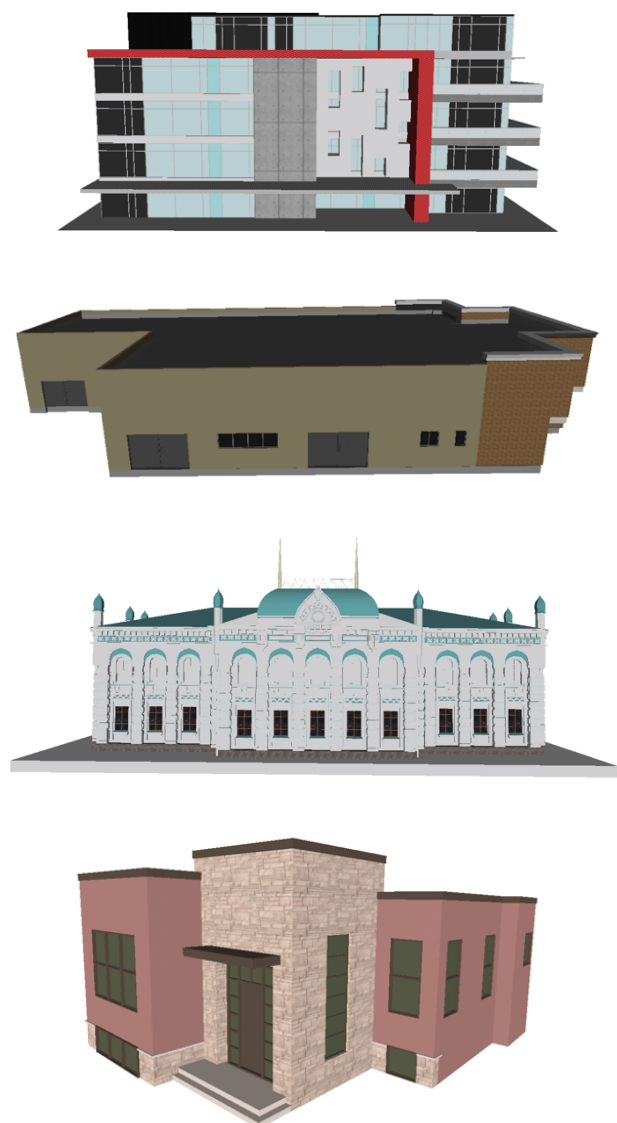


Figure 3. Four different building types. From top to bottom: commercial building, industrial building, institutional building, and residential building.

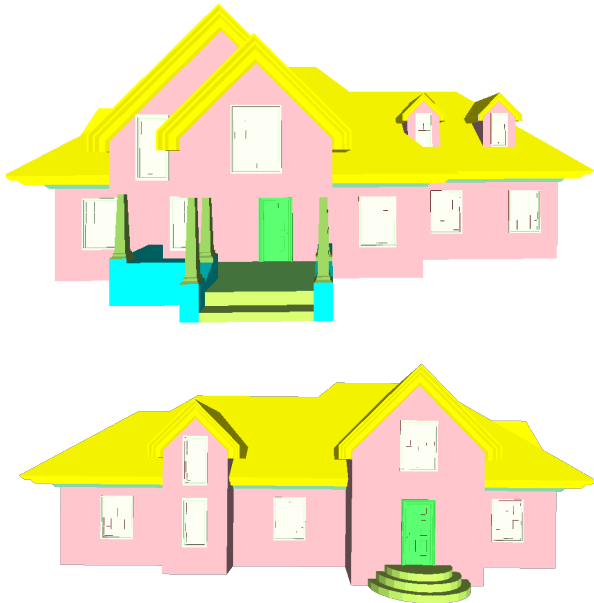


Figure 4. MeshLab Software's visualization of a mesh with annotations. Top: Building's front view; bottom: Building's back view.

3.5 Dataset Post-processing

3.5.1 Point cloud sampling: As seen in Figure 5, Point clouds are densely and uniformly sampled on the labeled meshes. We employed a method of sampling point clouds in accordance with the size of each mesh face in a mesh to produce a uniform point cloud on each building, yielding 3,500,000 points per building. The point densities between various buildings vary depending on the scale of the buildings, as seen in Figure 6. The semantic labels and color information on each mesh face were converted into points within the corresponding face during the sampling process, in addition to maintaining the geometric information. Following sampling, the small faces without labels are classified as the unclassified category.

3.5.2 Data Augmentation: It is important to provide the data augmentation methods used in future research rather than the augmented data itself, so that others could use their own augmentation techniques if they desired. In the context of the building dataset, some examples of data augmentation techniques that could be used are provided along with the dataset:

1. Rotation: randomly rotating the building models in different directions to simulate changes in viewpoint.
2. Translation: randomly translating the building models in different directions to simulate variations in location.
3. Scaling: To simulate size variations, the building models were scaled arbitrarily.
4. Flip: randomly flip the building models in either a horizontal or vertical direction to simulate changes in orientation.
5. Adding noise: randomly introducing Gaussian noise to the building models to simulate variations in real-world conditions.

By using these types of data augmentation techniques, researchers can create a more diverse and comprehensive dataset

for training ML and DL algorithms. However, it is important to note that not all data augmentation techniques are applicable or suitable for all types of ML/DL algorithms or building-related tasks.

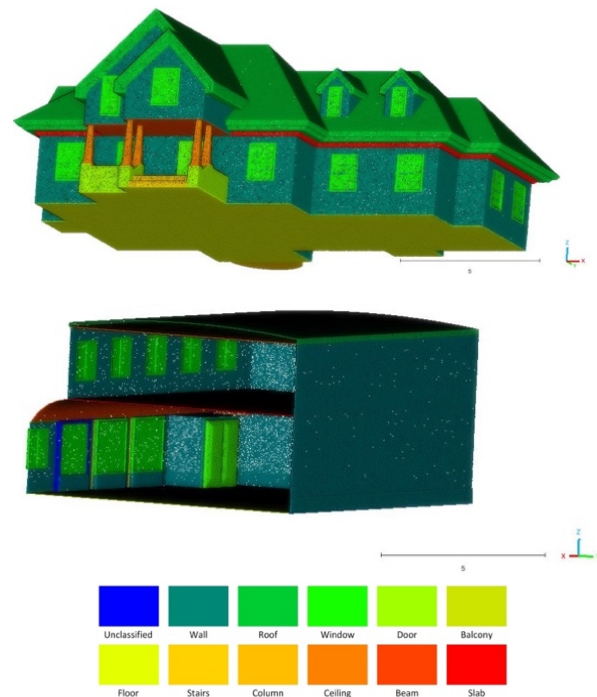


Figure 5. Sampled point clouds on labelled meshes. Top: outdoor of a building, bottom: indoor part of a building.

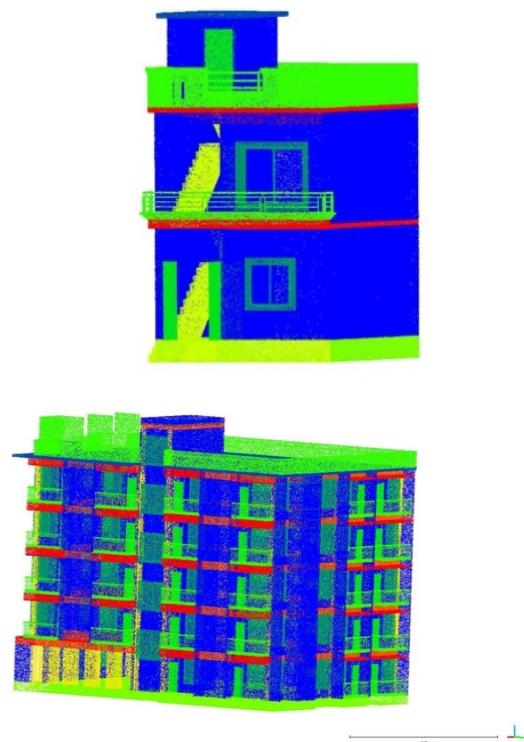


Figure 6. Different point densities in different buildings. Two buildings are displayed with the same point size in the CloudCompare Software.

4. BENCHMARK DATASET

By counting the area of the labeled meshes, the total area of this dataset is 481,769 m², including the indoor part. An average of 75,587 triangular faces in each building. The distribution of semantic categories in this dataset is depicted in Figure 7.

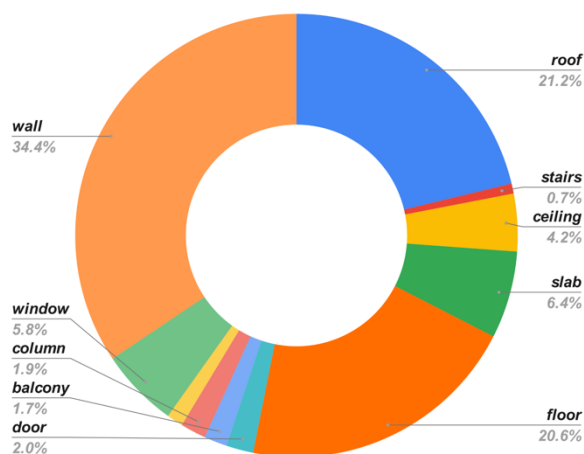


Figure 7. Category distribution.

To establish the benchmark for the BIO dataset and check the availability of our benchmark dataset, we used 1% and 10% portions of each building as training data and tested the outcomes with the remaining portions. It should be noted that we randomly chose 1% and 10% of the training data rather than manually choosing a few segments for each building. To be more precise, we divided each building into 1m×1m blocks and then randomly sampled 8,192 points from each block. After generating the blocks, we randomly selected 1% or 10% of the blocks (see Figure 8), and then trained on the selected blocks. The semantic segmentation was carried out on the remaining test data for each building after the ML model had been trained on the training data for each building. The final dataset can be classified by the RF classifier without the need for a sizable amount of manually annotated data, but it needs significant geometrical features as input that can draw attention to the discontinuities between elements.

Following earlier research (Weinmann et al., 2017), we first chose a set of features that are relevant to the problem in order to determine the most effective set of features. These features place a strong emphasis on the point cloud's structure within the predetermined radius of the points. We then rank the significance of each feature in predicting the target variable using the random forest feature importance method. Finally, 16 features are used in our experiment, including x, y, z, r, g, b, normalized color, verticality_0.1m, verticality_0.2m, anisotropy_0.2m, surface_variation_0.2m, omnivariance_0.2m, verticality_0.4m, linearity_0.4m, and planarity_0.4m (see Figure 9). The search radii used when calculating geometric covariance features are indicated by the numbers that come after the name of the geometric features.

To choose the RF algorithm's hyperparameters, such as the number of trees and the maximum depth of each tree. We use the weighted F1 score to compare the performance of models and choose the model with the best performance and the most suitable hyperparameters using the grid search technique.

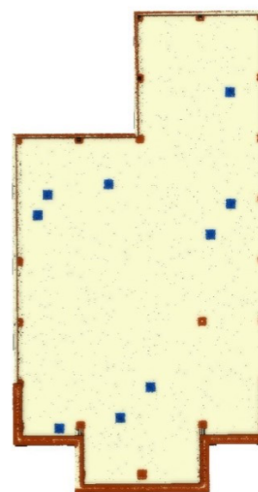


Figure 8. Training data in a building: the blue blocks denote the blocks randomly selected as training data (for better visualization, we just visualize 1% blocks here).

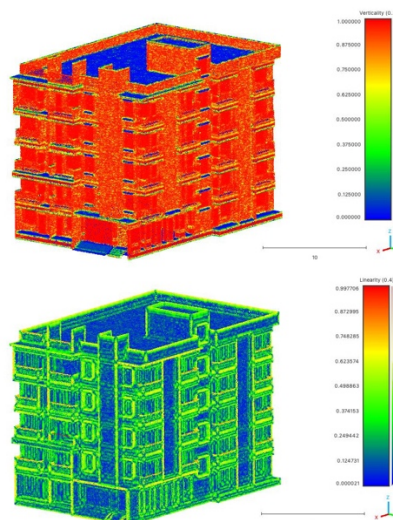


Figure 9. Examples of geometric features. Top: verticality_0.2m, bottom: linearity_0.4m.

We evaluated their overall accuracy and weighted F1 score using the test data. Table 1 displays the outcomes of semantic segmentation using different numbers of generated building blocks. According to the findings, the semantic segmentation achieves an average weighted F1 score of 86.02% when only 1% of the blocks are used as training data. Additionally, with 10% of the blocks serving as training data, we were able to produce results that were encouraging, with an average weighted F1 score of 96.64%.

Data Percent	Weighted_F1	Overall Accuracy
1%	86.02%	87.84%
10%	96.64%	96.87%

Table 1. Semantic segmentation results.

Figure 10 demonstrates the prediction errors using training data with 1% and 10% portions of building. A three-building school complex serves as the chosen demonstration site. As you can see, when only using 1% of the blocks as training data, the RF model has some difficulty in correctly predicting the regions containing windows and doors. With 10% of the blocks are used as training data in each building, the RF model can perform accurate classification on complex building models.

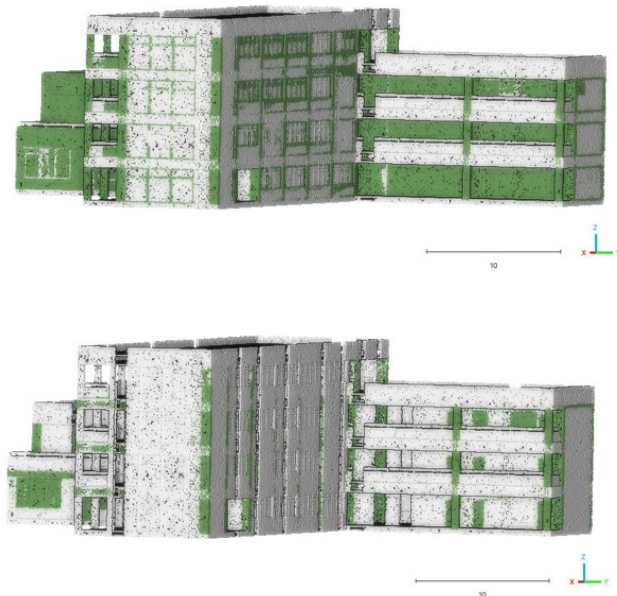


Figure 10. Prediction errors of a complex of school buildings (indicated in green), top: training with 1% blocks of this building, bottom: training with 10% blocks of this building.

5. CONCLUSION

We have introduced a new indoor-outdoor labeled building dataset in this paper, which includes a variety of building types and was created using a semi-automatic framework that includes data collection, pre-processing, manual labeling, and automatic post-processing. Using the machine learning (ML) algorithm Random Forest (RF), we then assessed the applicability of this dataset for the semantic segmentation task. The outcomes of our tests show that the RF algorithm attained a high level of accuracy. To sum up, our research makes a significant contribution to the field of creating datasets for AI applications.

More ML and deep learning algorithms will be tested on the BIO dataset in the future. Additionally, the possibility of using this dataset to improve the performance of real-world datasets will be investigated. In order to define and develop the BIO dataset as a crucial dataset with lasting impact, we would like to involve the larger research community.

ACKNOWLEDGEMENTS

Financial support from the program of the China Scholarships Council (Grant No. 201906860014) is acknowledged. We thank the anonymous reviewers for their constructive comments, which helped improve this paper. We thank Dr. Gao et al. (2022) for the UrbanMeshAnnotator software.

REFERENCES

- Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3D Semantic Parsing of Large-Scale Indoor Spaces, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 1534–1543. <https://doi.org/10.1109/CVPR.2016.170>
- Breiman, L., 2001. Random forests. *Machine learning*, 45, pp.5–32. <https://doi.org/10.1023/A:1010933404324>.
- Cao, Y., Teruggi, S., Fassi, F., Scaioni, M., 2022. A Comprehensive Understanding of Machine Learning and Deep Learning Methods for 3D Architectural Cultural Heritage Point Cloud Semantic Segmentation, in: Borgogno-Mondino, E., Zamperlin, P. (Eds.), *Geomatics for Green and Digital Transition, Communications in Computer and Information Science*. Springer International Publishing, Cham, pp. 329–341. https://doi.org/10.1007/978-3-031-17439-1_24.
- Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., 2008. MeshLab: an Open-Source Mesh Processing Tool. *Eurographics Italian Conference*. <https://doi.org/10.2312/LOCALCHAPTEREVENTS/ITALCHAP/ITALIANCHAPCONF2008/129-136>.
- Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Niessner, M., 2017. ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, pp. 2432–2443. <https://doi.org/10.1109/CVPR.2017.261>
- Dawson-Haggerty, M., 2022. Trimesh Software. Available at: <https://trimsh.org/trimesh.html> (Accessed: Jul. 2023)
- Gao, W., Nan, L., Boom, B. and Ledoux, H., 2021. SUM: A benchmark dataset of semantic urban meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179, pp.108-120. doi.org/10.1016/j.isprsjprs.2021.07.008.
- Géron, A., 2022. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. SEMANTIC3D.NET: A NEW LARGE-SCALE POINT CLOUD CLASSIFICATION BENCHMARK. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* IV-1/W1, 91–98. <https://doi.org/10.5194/isprs-annals-IV-1-W1-91-2017>.
- Isikdag, U., Zlatanova, S. and Underwood, J., 2013. A BIM-Oriented Model for supporting indoor navigation requirements. *Computers, Environment and Urban Systems*, 41, pp.112-123. [10.1016/j.compenvurbsys.2013.05.001](https://doi.org/10.1016/j.compenvurbsys.2013.05.001).
- ISO, 2018. 16739-1: 2018: Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries—Part 1: Data Schema. International Organisation for Standardisation: Geneva, Switzerland.

Kutzner, T., Chaturvedi, K., Kolbe, T.H., 2020. CityGML 3.0: New Functions Open Up New Applications. *PFG* 88, 43–61. <https://doi.org/10.1007/s41064-020-00095-z>.

Matrone, F., Lingua, A.M., Pierdicca, R., Malinverni, E.S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T., 2020. A BENCHMARK FOR LARGE-SCALE HERITAGE POINT CLOUD SEMANTIC SEGMENTATION. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLIII-B2-2020, 1419–1426. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1419-2020>.

Muntoni, A., Cignoni, P., 2021. cnr-isti-vclab/PyMeshLab: PyMeshLab. Zenodo. <https://doi.org/10.5281/zenodo.4438750>.

Nikoohemat, S., Diakit , A.A., Zlatanova, S. and Vosselman, G., 2020. Indoor 3D reconstruction from point clouds for optimal routing in complex buildings to support disaster management. *Automation in construction*, 113, p.103109. [10.1016/j.autcon.2020.103109](https://doi.org/10.1016/j.autcon.2020.103109).

O'Donnell, J., Truong-Hong, L., Boyle, N., Corry, E., Cao, J. and Laefer, D.F., 2019. LiDAR point-cloud mapping of building facades for building energy performance simulation. *Automation in Construction*, 107, p.102905. [10.1016/j.autcon.2019.102905](https://doi.org/10.1016/j.autcon.2019.102905)

Schrotter, G. and Hürzeler, C., 2020. The digital twin of the city of Zurich for urban planning. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1), pp.99-112. [10.1007/s41064-020-00092-2](https://doi.org/10.1007/s41064-020-00092-2).

Selvaraju, P., Nabail, M., Loizou, M., Maslioukova, M., Averkiou, M., Andreou, A., Chaudhuri, S., Kalogerakis, E., 2021. BuildingNet: Learning to Label 3D Buildings, in: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, pp. 10377–10387. <https://doi.org/10.1109/ICCV48922.2021.01023>.

SM Iman Zolanvari, Susana Ruano, Rana, A., Cummins, A., da Silva, R.E., Morteza Rahbar, Aljosa Smolic, 2019. DublinCity: Annotated LiDAR Point Cloud and its Applications, in: *Proceedings of the British Machine Vision Conference (BMVC)*. Presented at the 30th British Machine Vision Conference, BMVA Press, Cardiff, UK. <https://doi.org/10.5244/C.33.127>.

Weinmann, M., Jutzi, B., Mallet, C., and Weinmann, M.: GEOMETRIC FEATURES AND THEIR RELEVANCE FOR 3D POINT CLOUD CLASSIFICATION, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-1/W1, 157–164. <https://doi.org/10.5194/isprs-annals-IV-1-W1-157-2017>, 2017.

3D Warehouse. Available at: <https://3dwarehouse.sketchup.com> (Accessed: Jul. 2023).