AGILE: GIScience Series, 3, 2, 2022. https://doi.org/10.5194/agile-giss-3-2-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Understanding the Imperfection of 3D point Cloud and Semantic Segmentation algorithms for 3D Models of Indoor Environment

Guoray Cai [1] and Yimu Pan [1]

[1]College of Information Sciences and Technology, Penn State University, University Park, PA 16802, USA

Correspondence: Guoray Cai (gxc26@psu.edu)

Abstract. Point clouds data provides new potentials for automated construction of more geometrically accurate and semantically rich 3D models for indoor environments. Recent advances in deep learning methods on point cloud semantic segmentation demonstrated impressive accuracy in labeling points of 3D surfaces with object classes. However, it remains challenging to reconstruct the shape of semantic objects from semantically-labeled 3D points, due to imperfection of such data and the under-determination of object construction algorithms. We have little empirical knowledge about how data imperfections affect the reconstruction of 3D indoor room objects. This paper contributes to understanding the nature of such imperfection of 3D point cloud data and semantic segmentation algorithms by analyzing the reconstructability of indoor room objects from semantically-labeled point cloud. 181 rooms from Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS) were used in our experiment. After generating semantic labels on point-clouds using PointNet++ segmentic segmentation algorithm, we use human coders to judge the reconstructability of indoor objects, following a qualitative coding scheme. Human exploration of object shape imperfection was assisted by a visual analytic tool in making their judgement. We found that high point-level accuracy achieved through semantic segmentation of point cloud data does not guarantee high object-level accuracy. The extent of this problem varies widely among different spatial settings and configurations. We discuss the significance of these findings on the choice of 3D reconstruction methods.

Keywords. 3D Models; indoor environment; 3D reconstruction; point clouds processing

1 Introduction

Indoor spaces support majority of human activities (Worboys, 2011). Studies show that the average person spends around 90% of their time indoors (Klepeis et al., 2001). With rapid urbanization, large indoor spaces (such as high-rise business complex, public buildings, airports, and train stations) are increasingly used to provide services and infrastructures (Kang and Li, 2017). These indoor environments are increasingly complex and difficult to navigate, manage, and use. Therefore, accurate three-dimensional (3D) models for such indoor spaces are of profound importance for a wide range of applications, such as construction, indoor navigation and real estate management (Zlatanova et al., 2013; Lehtola et al., 2017).

The construction of 3D models of indoor environments still poses mounting challenges due to the complicated layout of the indoor structure, and the complex interactions between objects, clutter, and occlusions (Naseer et al., 2019; Zlatanova et al., 2013). When indoor space is large and complex, models typically incorporate some levels of subdivisions to reflect the hierarchical structure of the physical, functional, and social space (Richter et al., 2011). For example, a shopping mall has a number of stores, warehouses, control rooms, cinemas, sports centers, subway stations, etc., each of which has unique requirements and functions. Indoor 3D model specifications, such as IndoorGML (Kang and Li, 2017), typically represent indoor space subdivisions that are composed of rooms (or cells) connected by corridors. Rooms are themselves complex structures, commonly surrounded by architectural components (such as walls, ceilings and floors) and could be populated by doors, windows, and furniture (e.g., chairs, tables, lamps, computers, cabinets) (Lorenz et al., 2006; Becker et al., 2009). Basic elements in a room have strong mutual relationships with known priors (e.g., a chair stands on the floor, a monitor rests on the table).

3D models for indoor environment could be derived from many sources, such as photogremmetry, survey, BIM, 2D imagery, and modern 3D sensors and scanners (Zlatanova and Isikdag, 2017). In particular, 3D Point cloud data derived from 3D sensors provides rich geometric, shape and scale information, and can be used to estimate the surface geometry and material composition of the reflecting surface. Point cloud represents objects in space by collating

a large number of single point measurements. Each point represents a single laser scan measurement (in {X, Y, Z} geometric coordinates) on an sampled surface of a spatial object. With these data sources, 3D models can be derived through a series of processes, commonly known as *3D reconstruction* (Berger et al., 2017; Chen and Clarke, 2020). The ultimate goal of 3D reconstruction is to characterize object's location, geometries, semantics, and their spatial arrangement and relationships. The fundamental tasks are the discovery of structural elements, such as rooms, walls, doors, and indoor objects, and their combination in a consistent structured 3D shape and semantics (Pintore et al., 2020).

The major challenge of solving the reconstruction problem using 3D point cloud data is that inferring the geometric shape of an objects from sampled 3D points (point cloud representation) is ill-posed. Pintore et al (Pintore et al., 2020) showed that an infinite number of 3D surfaces may fit a 3D point cloud scan of a surface due to the under-sampled or partially missing data. For this reason, existing 3D object construction algorithms all rely on some form of pre-defined knowledge *priors* to restrict the candidate matching and to make reconstruction tractable (Berger et al., 2017). The knowledge priors may come from architectural principles (such as *rooms are bounded by walls, floor and ceiling*) or functional principles (such as *chairs are typically next to tables*). This approach was only partially successful due to the complexity and variability of interior environments. The effectiveness of these methods vary significantly in dealing with imperfection of point cloud data, such as noisy data, missing data, non-uniform sampling, and outliers Kang et al. (2020). This results in several dozens of 3D object reconstruction algorithms, each was optimized on very specific expected indoor structures and objects, or to combat specific types of imperfections in the point cloud representation. Users of these algorithms face difficulties in choosing the 'right' algorithms when facing new situations. Therefore, it is important that we develop good understanding of the nature of the imperfections of the data. Further more, we have little empirical knowledge about how data imperfections affect the reconstruction of 3D room objects.

This paper contributes to the above knowledge gap by offering insights to the nature of imperfection of point cloud data from the perspective of 3D reconstruction. We explored the imperfections of a semantically-labeled point cloud datasets derived from semantic segmentation of Stanford 3D Indoor Space Dataset (S3DIS), and assessed the reconstructability of indoor objects using human coders who were assisted by a visual analytic tool in making their judgement. We found that high point-level accuracy achieved through semantic segmentation of point cloud data does not guarantee high object-level accuracy. The extent of this problem varies widely among different spatial settings and configurations and is also sensitive to the hyperparameters of semantic segmentation algorithms. We discuss the significance of these findings on the choice of 3D reconstruction methods.

2 Related work

The automated reconstruction of 3D models from 3D point cloud data has been one of the central topics in computer graphics and computer vision for decades (Pintore et al., 2020). A subfield of this domain concerns with automatic reconstruction of indoor environments, which derives a 3D representation of an interior scene from 3D scan data. Indoor objects in the context of this study include not only architectural (fixed) elements (walls, floors, ceilings) but also movable objects of room interiors (furniture, windows, doors, boards, etc). Indoor 3D reconstruction is the process by which a 3D indoor objects are inferred, or 'reconstructed', from a collection of discrete points that sample the shape (Berger et al., 2017). Indoor objects (such as furniture) are considered as integral part of the indoor environment. The occurrence and arrangement of indoor objects in room interior offer important clues to understand the purpose and functions of indoor environments (Zhang et al.).

Most indoor object reconstruction methods have focused on finding the geometric shapes of architectural elements that bound the interior of rooms (Ochmann et al., 2016, 2019; Shi et al., 2019; Macher et al., 2017; Kang et al., 2020). They typically start with detection of primitive geometric features (such as planes, lines, and corners) and then compose them into cuboids, using adjacency relationships and Manhattan-World (MW) prior. Room interior objects are harder to reconstruct due to the need for detailed surface representations, complex shapes, and occlusion. To overcome this problem, room objects reconstruction typically leverage semantic information to construct geometries. Figure 1 shows the generic workflow of 3D reconstruction process that derive shape and semantic representation of room objects from point cloud representation. It shows that the input to the object reconstruction algorithm is the *semantically-labeled point cloud* (SLPC) representation of room interiors, which is generated from the process of semantic segmentation on point clouds. This will allow the reconstruction of object shapes to exploit data-driven priors in the form of a collection of known shapes (e.g. a shape library of different furniture objects) (Li et al., 2015; Nan et al., 2012). Using SLPC as input, an object reconstruction algorithm will try to grow surfaces using clusters of points with similar object labels and use these partially constructed surfaces to match shapes of the same object class. The performance of object reconstruction algorithm is determined by three factors:

(1) *The quality of surface reconstruction algorithms*. Substantial progress has been made in 3D surface reconstruction methods. Berger et al. (2017) reviewed thirty-two point cloud modelling methods where they identified the knowledge priors of each algorithm explicitly. They

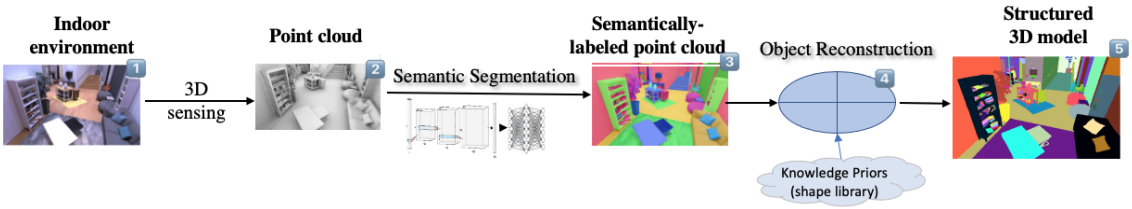

Figure 1. Workflow of room objects 3D reconstruction from point cloud

showed that the effectiveness of these methods vary significantly in dealing with imperfection of point cloud data, such as noisy data, missing data, non-uniform sampling, and outliers. So far, there is no method that is generally good fit to all situations, and each of the methods was designed to deal with specific type of objects and imperfections.

(2) *The quality of shape library*. Significant efforts have been invested in building shape libraries for various applications, together with methods of approximate shape matching and retrieval (Xie et al., 2017). Such shape matching can be mediated by the use of domain-specific ontology of objects shape, parts and relationships (Poux et al., 2018).

(3) *The quality of input data*. As shown in Figure 1, the input data to the object reconstruction is the *semantically-labeled point cloud* (SLPC) representation of room interiors ((box 3). This representation is typically generated by supervised machine learning methods that are designed to infer the object class on each 3D point (also known as *semantic segmentation*)(Xie et al., 2020). The process of matching point cloud representation with shape priors is vulnerable to various imperfections of point cloud data. Berger et al Berger et al. (2017) reviewed five commonly known point cloud artefacts (non-uniform sampling density, missing data, noise, outliers, and misalignment) and discussed how they impact the fidelity of reconstructed surfaces. However, we know very little on the imperfection of *semantically-labeled point cloud* (SLPC) representation in relation to 3D indoor object reconstruction.

Our work contributes to filling the knowledge gap identified above (in(3)) by exploring the imperfections of SLPC representation that is produced by automated semantic segmentation methods. We use the following research questions to guide our exploration:

Question 1 *What are the error and confusions in the semantic labels generated by automated semantic segmentation methods*? Existing work on semantic segmentation of indoor scenes only reported average point-level accuracy for the whole dataset, and have never explored where and why errors occurred. In this paper, we explored the accuracy and confusion measures of semantic labels for commonly encountered room object classes (furniture, doors, windows, boards, etc) and offer some insights on how accuracy varies among different object classes (see Section 3).

Question 2 *Can 3D object shapes be reconstructed from semantically labeled point cloud representation?* We answer this question by comparing human coding of object geometry quality with the semantic label accuracy (see Section 4). We demonstrated that high point-level accuracy of point labels does not guarantee shape reconstructability in object-level geometry.

3 Assessing Imperfections of Semantic Segmentation

The work of this section is to address *research question* 1. Given a point cloud representation of an indoor scene, the goal of semantic segmentation is to separate a point cloud into several subsets according to their semantic object categories (tables, chairs, doors.etc). With the availability of multiple semantically annotated 3D point cloud datasets for indoor environment ((Armeni et al., 2016; Dai et al., 2017)), supervised machine learning models based on deep learning architecture have achieved superior performance over other methods Guo et al. (2020). The key advantage of deep learning methods is that it does not require human-guided design of discrimination features in segmentation tasks. Based on recent survey of automated semantic segmentation techniques (Xie et al., 2020; Guo et al., 2020; Liang and Fu, 2019; Li et al., 2018), PointNet++ Qi et al. (2017) and its close variants are the latest and best performing network structures for 3D semantic segmentation on point cloud. Since PointNet++ represents the state-of-the-art semantic segmentation algorithms for indoor objects, we will use PointNet++ as the proxies of the best semantic labeling methods on 3D points.

Our subsequent experiments use PointNet++ Qi et al. (2017) as the semantic segmentation algorithm to observe the imperfections caused by the algorithmic artefacts. To better understand the algorithmic artefacts associated with PointNet++, we will briefly review the principles of PointNet++ and its associated algorithmic artefacts, followed by an experimental studies for understanding the imperfection.

3.1 PointNet++: Semantic Segmentation on Point Cloud

PointNet++ Qi et al. (2017) is a hierarchical neural network which process a set of 3D points sampled in a metric space. Figure 2 illustrates the architecture of PointNet++

framework. It consists of a "Hierarchical Point Set Feature Learning" phase and a "segmentation" phase. During hierarchical feature learning, multiple levels of abstraction form a hierarchy to pool features towards more compact and abstract feature representation. During segmentation, a hierarchical point feature propagation coupled with distance based interpolation to derive semantic label for all the points.

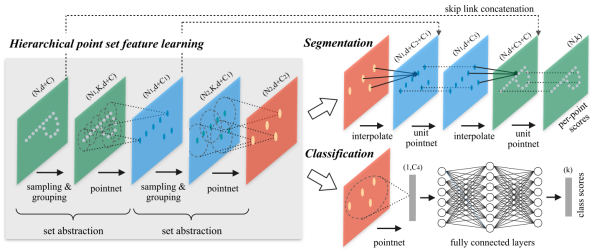

Figure 2. PointNet++ framework Qi et al. (2017)

The imperfections of object class labels generated by PointNet++ model can be understood from the way that PointNet++ constructs the feature space in the learning model. This can be summarized into the following key ideas:

- *It partitions the set of points into overlapping local regions by the distance metric of the underlying space*. It uses the *point sphere model* around a set of *centroids* to define local regions.
- *It extracts local features capturing fine geometric structures from small neighborhoods*. This is done by a mini-PointNet - a multi-layer perceptron (MLP) network.
- *Such local features are further grouped into larger units and processed to produce higher level features*. PointNet++ uses MSG (Multi-Scale Grouping) to extract the local features at each radius and combines them together and use MRG (multi-resolution grouping) to combine the global features of different layers. This process is repeated until the features of the whole point set is obtained.

PointNet++ has multiple algorithmic artefacts that could contribute to the imperfection of object class labels on points.

As PointNet++ method of semantic segmentation is heavily data-driven, some parameters were made adjustable for the detection and recognition of objects of different shapes and sizes. Any applications of PointNet++ method on indoor object semantic segmentation will face a number of decisions to make related to the choices of parameters that can be tuned by the modeler to avoid the worst performance and achieve the best possible outcome.

By analyzing the source code and algorithms of PointNet++, we have identified a number of parameter settings that users may adjust to maximize the performance of semantic segmentation on a specific dataset. In the pre-processing step, the raw data was chunked (to cubes defined by 1.5m-1.5m square of floor areas) and down-sampled (using random sample 8192 points) which satisfied the data quality without losing information. Then, then 1024 local regions are defined on each cube, where the *centroid* of those local regions are determined by furthest point sampling (FPS) and a *radius*. These parameters are given in Table 1.

Table 1. Parameters and default settings in PointNet++.

Para	Explanation	Default
S_c	The size of the floor square used to chunk the raw data	1.5m x 1.5m
N_s	The number of points to down-sample raw data	8192
R	Radius of the spheres that define local regions around centroids	0.1
K	Number of of points sampled from each local region	1024

The success in extracting local features depends on proper setting of the above parametersZhang et al. (2020). Due to the entanglement of feature scale and non-uniformity of input point set, the choice of these above parameters could lead to failure in searching local features. This is a source of imperfection in semantic segmentation. To understand the effect of such model artefacts on the imperfection of the semantic segmentation outcome, we conducted the following experiment.

3.2 Experiment 1: Comparing Imperfections Across Object Classes

We conducted an experiment to understand the extent to which PointNet++ semantic segmentation algorithm contributes to the imperfection of object class labels for indoor room objects. To do so, we replicated PointNet++ semantic segmentation algorithm exactly as was done in the original paper Qi et al. (2017). The purpose is to understand imperfection in terms of accuracy and confusion rates.

System design. We followed the original design of PointNet++ semantic scene labeling [1], which uses 4 set of abstractions. The four layers have 1024, 256, 64, 16 nodes, respectively. All parameters settings in the source code were unchanged.

Dataset description and preprocessing. To verify the ability of the adopted approach to achieve more semantic and acceptable results for indoor 3D modeling, we used a publicly available dataset, *Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS)* Armeni et al. (2016) to train and test a PointNet++ semantic segmentation model. The dataset contains 3D scans from Matterport scanners in 6

[1] https://github.com/charlesq34/pointnet2

areas including 272 rooms (see Figure 3). Each point in the scan has been annotated with one of the semantic labels from 13 categories (chair, table, floor, wall, window, sofa, door, etc.). The dataset has (X, Y, Z) coordinates and (r, g, b) color information, but we only use (X, Y, Z) for semantic segmentation.

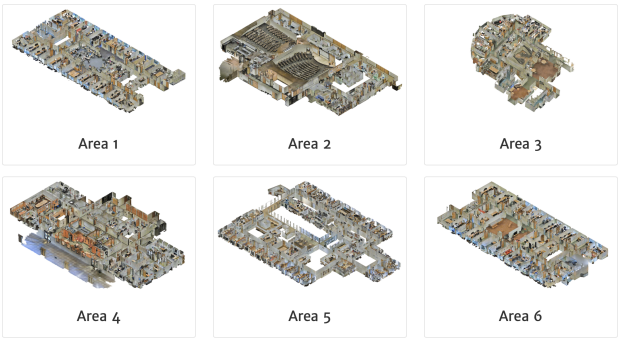

Figure 3. Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS)

Among the 272 rooms, we filtered out hallway, WC, and storage rooms, so that we can focus on "offices" and "conference" rooms. That brings the total rooms used down to 181. Among these rooms, we randomly sampled 151 rooms(83%) as training set and remaining 30 rooms are reserved for testing.

To prepare the training data, we cut each room in the training set into a number of cubes where each cube corresponds to 1.5m by 1.5m squares on the floor space. In each cube, we randomly selects 8,192 points to participate the model training. Then, combine all the rooms we get the training data of shape (n, 8192, 3) and label of shape (n, 8192,1) where n is the number of 1.5m by 1.5m cubes. Testing data is prepared the same way.

During each epoch of training, we randomly select n/2 cubes from the prepared training data. We trained the model with 200 epoch and batch size of 8. This will guarantee all data cubes will be covered during 200 epoch of training. The model is tested every 5 epoch, the testing is done on all prepared testing data with batch size of 8. Finally, we save the best performing model.

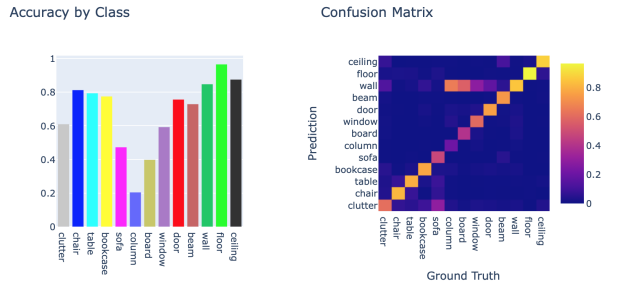

Figure 4. Point-level accuracy and error rates on semantic labeling

Results and Findings.

The experiment resulted in 83.2% in overall class label accuracy. Figure 4 breaks down the output by 13 object classes and shows the per-class accuracy as a bar chart on the left, and a confusion matrix (on the right) that communicates the degree of errors and confusions among classes. The following observations can be made:

Observation 1. *Relatively high level of accuracy was achieved on 'wall,' 'floor,' and 'ceiling' (90%+).* This is due to their simple geometric features and dominant amount of training samples in the data.

Observation 2. *Columns, boards, and windows have relatively low accuracy, and they are often confused as walls* (as reflected in the confusion matrix of Figure 4.

Observation 3. *Tables and chairs achieved around 80% accuracy on average, which is quite impressive*. There is some degree of confusion between tables and chairs, because they tend to be cluttered together.

The above results were collected using the default value of radius = 0.1 meter. We hypothesize that the choice of search radius of local regions has significant impact on the imperfection of object class label of points, as measured by accuracy and confusion.

3.3 Experiment 2: Understanding the Impact of Model Parameters on the Imperfection of Point Semantic Labels

Among those model parameters identified in Table 1, setting the proper *radius* value for local neighborhood balls is the most sensitive action, due to the entanglement of feature scale and non-uniformity of input point set Qi et al. (2017). In PointNet++, radius determines the receptive field of the multi-layer perceptron (MLP) used to extract local features. When it is set to a smaller value, the perceptive field is reduced, but it is going to see more details in the local, since each local region sphere will be sampled on the same number of points used in each training epoch. Conversely, a larger radius value will be equivalent to the change of receptive field when you zoom out on a camera lens. Larger radius allows detection of structures that span larger areas. Subsequent feature learning layers in the training model depend heavily on what the MLP can detect within the radius from a centroid point. Based on this observation, we hypothesize that *the choice of radius value has significant impact on the performance of PointNet++ semantic segmentation*. Next, we will conduct an experiment to observe how the performance of PointNet++ semantic segmentation model varies with the choice of *radius*.

Experimentation design

We will conduct the same experiment as described in Section 3.2, except that we will repeat that experiment 13 times to collect performance data on different radius values r = [0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.3, 0.35, 0.4]. To keep the PointNet++ model structure intact while changing radius, we sample the same

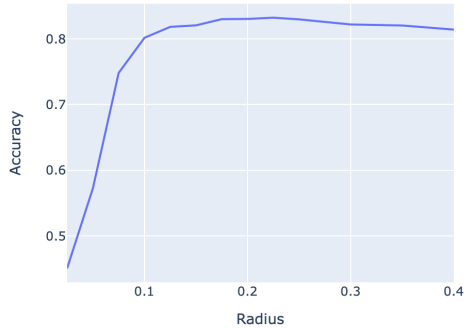

Figure 5. PointNet++ semantic segmentation performance as a function of radius parameter

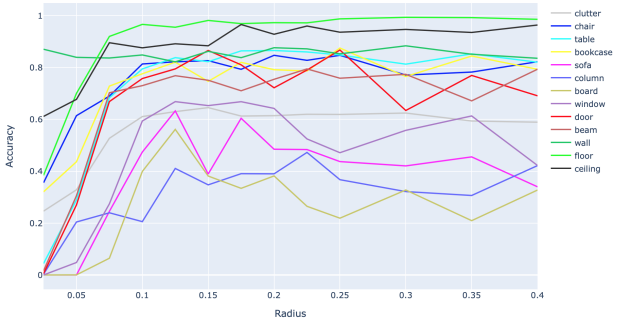

Figure 6. Radius-modulated accuracy curves by object classes

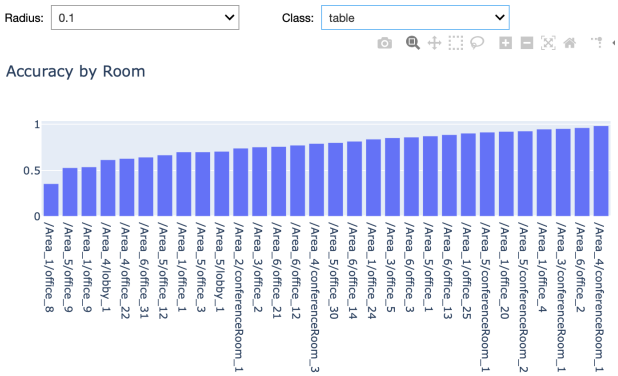

Figure 7. The effect of room configuration on errors and confusion

number of points from each local region during model training. The test dataset was prepared once and the same sampled data points were used for all r values. The purpose is to collect semantic label data on the same set of points across all r values, so that performance data are easier to be correlated for better comparison.

Experimental Results and Findings

We compiled the semantic label accuracy of all the testing points and computed average accuracy for each radius value. The result is shown as a performance curve in Figure 5. The low performance at smaller radius suggests that narrow field-of-view degrades capacity to learn local features. The performance of PointNet++ model for ScanNet data improve rapidly as radius gets larger, until it plateaued around r=0.1. This proves that the default setting of r=0.1 in PointNet++ was a reasonable choice for ScanNet data. Although overall accuracy is a good indicator of performance, accuracy on interior objects (such as furniture) is more important than on architectural elements such as walls, ceilings, and floors. To further investigate this, we created (Accuracy, Radius) curves for each of the 13 object classes in ScanNet (see Figure 6). A few observations can be made here:

Observation 4. *The change of radius has vastly different effect on different object class*. For walls, ceilings and floors, it seems that the choice of radius value does not matter as long as it is over 0.1m. In contrasts, there seem to be some "sweet spots" for doors and sofas that certain radius values produce the best possible outcome.

Observation 5. *selection of the radius value to optimize performance is very challenging.*. This was echoed by Qi et al. (2017), but we now have more details to appreciate the challenge. We are intrigued by the complex performance response to radius change as shown in Figure 6, and we do not have good enough theory to explain the phenomena. For example, r=0.15 seems to be an optimal choice for detecting doors, but it is a poor choice for detecting sofas.

Observation 6. *The change of radius has complex interactions with the configuration of rooms to create more performance variations across rooms*. Figure 7 shows the variation of accuracy across rooms for a given radius setting (r=0.1) and an object class (tables). The rooms are sorted by the accuracy measure on tables. An interesting observation here is that tables in conference rooms seem to be more accurately labeled than those tables in offices. We explored the room configuration (point cloud view) to see the unique and common aspects of those rooms. We found that Conference rooms tend to have tables that are in large size, centered in the room, have common oval shape or L shape. In contract, tables in offices tend to be smaller, less common in shape, and scattered along walls.

3.4 Enabling Interactive Exploration of Data Imperfection

Given the insights we derived from the above two experiments, we have come to a conclusion that it is important for users of indoor semantic segmentation algorithms (such as PointNet++). leveraging automated semantic segmentation on point cloud data for 3D indoor object reconstruction is far more complicated than a plug-and-play process. The modelers must have deep insight into how the model works and how the choice of parameters (such as radius) would impact the quality of point-level semantics. Modelers need a way to experiment, evaluate, and tune model parameters. Such activity can be supported by an interactive visual environment where users can explore

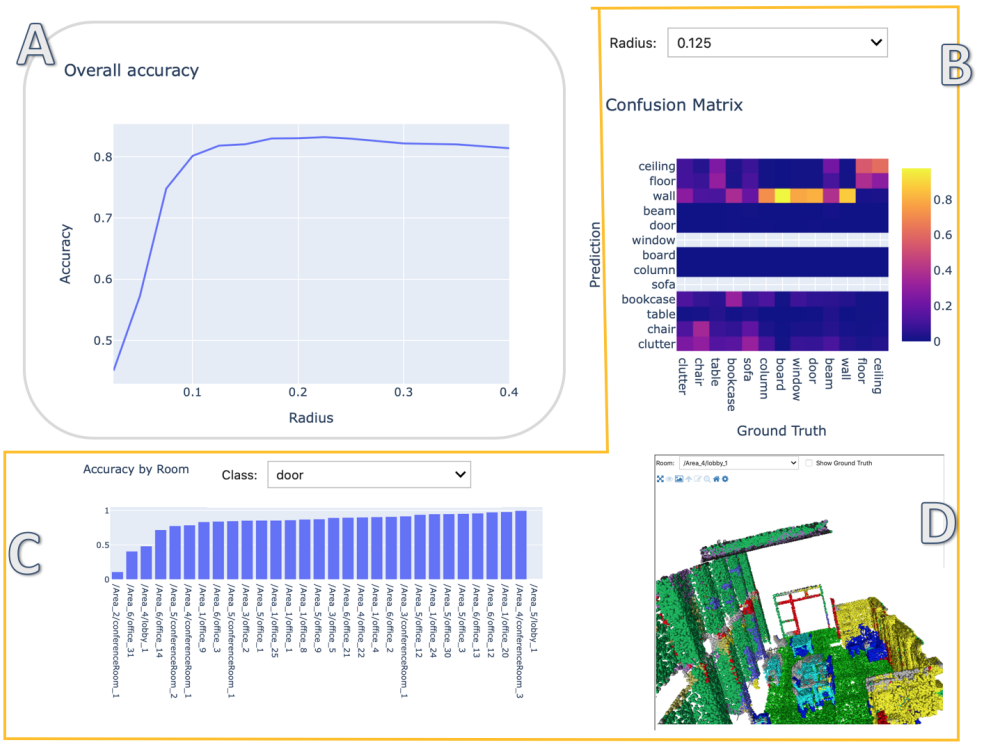

Figure 8. Interactive Dashboard for Exploratory Semantic Segmentation

the nature of semantic imperfections by themselves. Figure 8) shows our vision of such a system. The design of this tool followed the coordinated multiple view approach (Roberts, 2007). It has four component views A, B, C, and D. View A serves as the initial overview that provides high-level summary of overall accuracy at different radius settings. View B shows confusion matrix under the current radius value (indicated on the top of View B). View C is the breakdown of accuracy measures by rooms, and users can further filter this by object type ("class"). View D allows full interactive inspection of point cloud representation of a selected room, with the ability to zoom and rotate a 3D object in six degree of freedom.

The four views are linked to support state sharing, mutual filtering, overview+details, brushing and highlighting actions. Users may desire to investigate more details of performance for any radius value by clicking on a segment of view A. That will cause View B and View C to refresh with the accuracy and errors for the new radius. When users attempt to explain the patterns observed in View C, the can choose to inspect the raw point cloud data or segmented data (View D) by clicking on a bar in View C. That will cause View D to refresh with the new room data.

This tool was designed with the following goals in mind:

- *Provide an overview of segmentation accuracy*
- *Provide an overview of segmentation errors and confusion*
- *For a given radius setting, explore details in variation of accuracy and confusions by object classes*
- *Allow users to drill down to a subset of classes for ease of comparison and prioritization.*
- *Allow users to drill down to a subset of rooms taking a closer inspection of room configuration and geometric structures.*

4 Assessing the Reconstructability of indoor objects from SLPC

This section will conduct an experiment to answer Question 2, which is to understand whether and to what extent indoor room objects can be reconstructed from semantically-labeled point cloud representation of room scenes. One challenge of this task is that we do not have automated methods that we can trust to perform shape reconstruction (Laga et al., 2019), as it involves geometric quality judgment based on human knowledge about each class of objects. To get around this difficulties, we employed human intelligence in judging the degree of reconstructability on object instances.

In our experiment, two human coders were hired to generate a rating on the quality of each object in a subset of rooms, following a given codebook shown in Figure 9. This coding scheme was compiled by inspecting common types of imperfections and ranking them in the order

of their damage to reconstructability. Given a distribution of semantic labels in a room objects, a coder uses his/her world knowledge to judge if certain object shape can be inferred with some degree of confidence. This judgment is translated to a reconstructability value between 0 to 10, following the coding scheme.

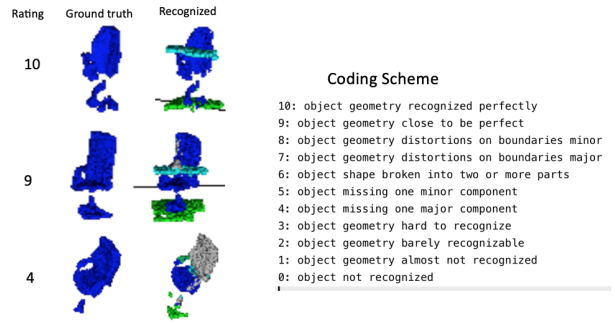

Figure 9. Object-level quality scales and examples

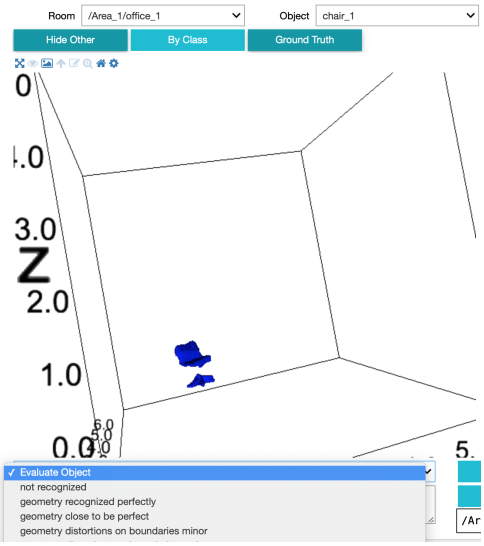

Figure 10. Interactive interface for annotating object quality

Judging the reconstructability following the coding scheme requires careful inspection of the geometric patterns of semantic labeled points. For example, the top surface and legs of a table are the main functional parts of table, while other decorative parts may not be as important. It is important that the point cloud semantic labels allow table top geometry to be clearly delineated. To easy the coding work, we developed an interactive tool that streamlined the process of selecting an object, comparing recognized object with its ground truth, and adding a rating. Figure 10 shows the interface. The tool allows a human coder to choose a room and an object in that room to be the working object at anytime. The system will retrieve the ground truth version and the predicted version of the object and visualize them in a volume view which can be freely rotated, zoomed, and tilted for inspecting the geometric quality from different angles. By turning on/off the ground truth, the coder will easily see the differences and judge what are missing.

Understanding that qualitative coding scheme is subject to variations of human interpretation, certain degree of coding errors and inconsistencies is expected. We assigned two coders to work independently on objects in those rooms in Area 1, 2, 3, and 4 of Stanford 3D Indoor Spaces (S3DIS) dataset (Figure 3). Inter-coder reliability based on Hallgren (2012) is 0.76.

We used the average value from the two coders as the indicator of the recognized object quality, and visualized it together with the point-level accuracy of each object as a scatterplot (see Figure 11). Each dot is an object, and its position on the scatterplot corresponds to the point class label accuracy on the horizontal axis and the object quality measure on the vertical axis. Dots are color-coded by their object class. We made the scatterplot interactive to enable exploration of the overall relationships and outliers among the set of objects. Cursor-over on any dot will prompt more details of the object to be shown as tool-tip. Clicking on a dot will invoke a view of the object in 3D volume to see both the ground truth and the recognized versions.

A number of patterns can be observed from the result in Figure 11:

Observation 7. *High point-level accuracy of class labels on an object does not necessarily mean high level of shape reconstructability*. Take "chairs" as example (shown as blue dots in the figure). Those chairs that have point-level-accuracy between 80%-90% may end up with an object quality rating from very good to very bad. When errors are distributed evenly over different parts of an object, they may not create much distortions on the shape and boundaries of an object and human eyes have little problem recovering the object shape. In contrast, if error points are clustered and cause significant parts of the object to be missing, the object quality will be significantly lowered.

Observation 8. *The overall reconstructability of major furniture objects are better than expected from point-level accuracy*. There are generally more tables, chairs, and sofas (in blue, green and pink colors) that are located on the top region of the scatterplot in Figure 11. This is a good news for indoor modeling research.

5 Conclusions and future work

In this paper, we developed a few experiments to explore the imperfection of semantically-labeled point cloud representation of room scenes for the purpose of 3D reconstruction. The insight we gained from these experiments can inform modelers in choosing indoor object reconstruction algorithms to achieve the best outcome. It will also inform future development of novel 3D reconstruction algorithms and workflows. We understand that our findings are far from conclusive, and the results of our experiments suggest much more questions needs to be answered on the

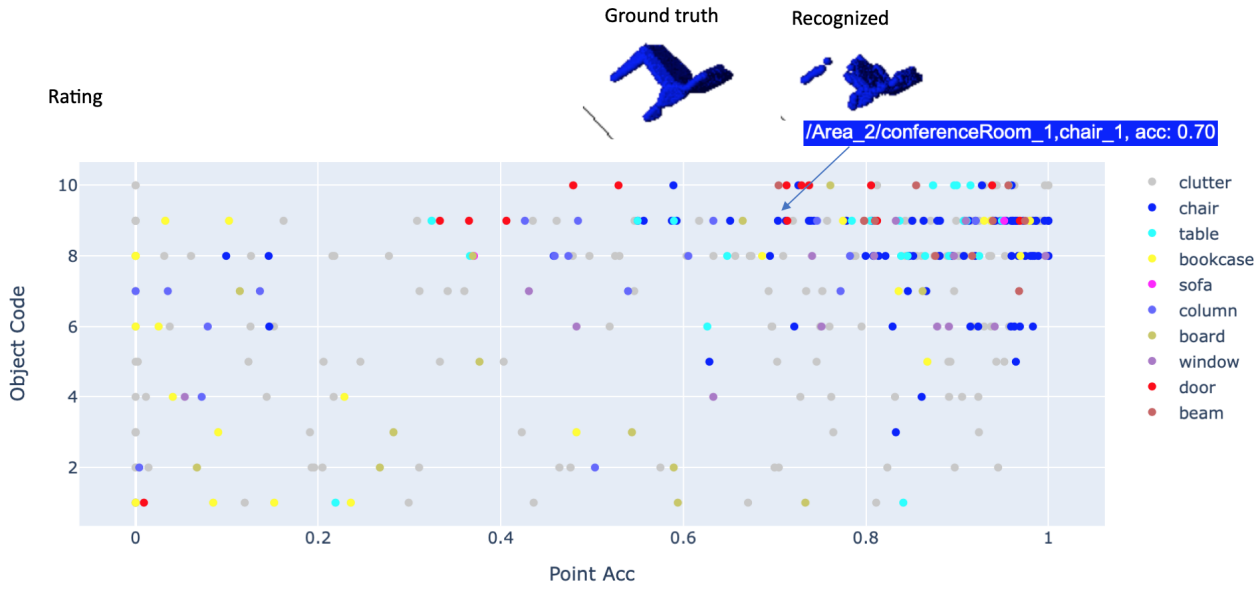

Figure 11. Relationship between point-level accuracy and object-level quality

entangled relationships among the complexity of indoor environment, the alternative configuration of deep learning models and data quality. Nevertheless, our findings contributes to an initial theory of deep learning model behavior on semantic segmentation and object reconstruction. The interactive tools we developed for this work can also be shared with other scholars who need tools to seamlessly integrate the process of model tuning and performance evaluation.

6 Data and software availability

The 3D indoor point cloud data used in this paper can be requested from S3DIS project website (http://buildingparser.stanford.edu/dataset.html). PointNet++ semantic segmentation algorithm is available from thr Github https://github.com/charlesq34/pointnet2. The Interactive data exploration and coding tools were written in Python and can be accessed as JupyterNotebook projects from Github at https://github.com/gxc26/PointClouds).

References

Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., and Savarese, S.: 3D semantic parsing of large-scale indoor spaces, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 1534–1543, https://doi.org/10.1109/CVPR.2016.170, 2016.

Becker, T., Nagel, C., and Kolbe, T. H.: Supporting contexts for indoor navigation using a multilayered space model, in: Proceedings - IEEE International Conference on Mobile Data Management, pp. 680–685, IEEE, 2009.

Berger, M., Tagliasacchi, A., Seversky, L. M., Alliez, P., Guennebaud, G., Levine, J. A., Sharf, A., and Silva, C. T.: A Survey of Surface Reconstruction from Point Clouds, Computer Graphics Forum, 36, 301–329, 2017.

Chen, J. and Clarke, K. C.: Indoor cartography, Cartography and Geographic Information Science, 47, 95–109, 2020.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., and Nießner, M.: ScanNet: Richly-annotated 3D reconstructions of indoor scenes, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pp. 2432–2443, 2017.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M.: Deep Learning for 3D Point Clouds: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 42, 1–24, http://arxiv.org/abs/1912.12033, 2020.

Hallgren, K. A.: Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial, Tutorials in Quantitative Methods for Psychology, 8, 23–34, https://doi.org/10.20982/tqmp.08.1.p023, 2012.

Kang, H.-K. and Li, K.-J.: A standard indoor spatial data model—OGC IndoorGML and implementation approaches, ISPRS International Journal of Geo-Information, 6, 116, 2017.

Kang, Z., Yang, J., Yang, Z., and Cheng, S.: A review of techniques for 3D reconstruction of indoor environments, ISPRS International Journal of Geo-Information, 9, 2020.

Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., Behar, J. V., Hern, S. C., and Engelmann, W. H.: The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants, Journal of Exposure Analysis and Environmental Epidemiology, 11, 231–252, 2001.

Laga, H., Guo, Y., Tabia, H., Fisher, R. B., and Bennamoun, M.: 3D Shape Analysis: Fundamentals, Theory, and Applications, John Wiley & Sons, 2019.

Lehtola, V. V., Kaartinen, H., Nüchter, A., Kaijaluoto, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaja, M. T., Virtanen, J. P., Kurkela, M., El Issaoui, A., Zhu, L., Jaakkola, A., and Hyyppä, J.: Comparison of the selected state-of-the-art 3D indoor scanning and point cloud generation methods, Remote Sensing, 9, 1–26, 2017.

Li, Y., Dai, A., Guibas, L., and Nießner, M.: Database-Assisted Object Retrieval for Real-Time 3D Reconstruction BT - Computer Graphics Forum, Computer Graphics Forum, 34, 435–446, 2015.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B.: Pointcnn: Convolution on x-transformed points, in: Advances in neural information processing systems, pp. 820–830, 2018.

Liang, X. and Fu, Z.: MHNet: Multiscale Hierarchical Network for 3D Point Cloud Semantic Segmentation, IEEE Access, 7, 173 999–174 012, 2019.

Lorenz, B., Ohlbach, H. J., and Stoffel, E. P.: A hybrid spatial model for representing indoor environments, in: International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2006), vol. 4295 LNCS, pp. 102–112, 2006.

Macher, H., Landes, T., and Grussenmeyer, P.: From point clouds to building information models: 3D semi-automatic reconstruction of indoors of existing buildings, Applied Sciences (Switzerland), 7, 1–30, 2017.

Nan, L., Xie, K., and Sharf, A.: A search-classify approach for cluttered indoor scene understanding, ACM Transactions on Graphics, 31, 1–10, 2012.

Naseer, M., Khan, S., and Porikli, F.: Indoor Scene Understanding in 2.5/3D for Autonomous Agents: A Survey, IEEE Access, 7, 1859–1887, 2019.

Ochmann, S., Vock, R., Wessel, R., and Klein, R.: Automatic reconstruction of parametric building models from indoor point clouds, Computers and Graphics, 54, 94–103, http://dx.doi.org/10.1016/j.cag.2015.07.008, 2016.

Ochmann, S., Vock, R., and Klein, R.: Automatic reconstruction of fully volumetric 3D building models from oriented point clouds, ISPRS Journal of Photogrammetry and Remote Sensing, 151, 251–262, https://doi.org/10.1016/j.isprsjprs.2019.03.017, 2019.

Pintore, G., Mura, C., Ganovelli, F., Fuentes-Perez, L., Pajarola, R., and Gobbetti, E.: State-of-the-art in Automatic 3D Reconstruction of Structured Indoor Environments, Computer Graphics Forum, 39, 667–699, 2020.

Poux, F., Neuville, R., Nys, G. A., and Billen, R.: 3D point cloud semantic modelling: Integrated framework for indoor spaces and furniture, Remote Sensing, 10, 1–26, 2018.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space, in: Advances in neural information processing systems, pp. 5099–5108, 2017.

Richter, K. F., Winter, S., and Santosa, S.: Hierarchical representations of indoor spaces, Environment and Planning B: Planning and Design, 38, 1052–1070, 2011.

Roberts, J. C.: State of the art: Coordinated & multiple views in exploratory visualization, in: CMV 2007: Proceedings - Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization, pp. 61–71, 2007.

Shi, W., Ahmed, W., Li, N., Fan, W., Xiang, H., and Wang, M.: Semantic geometric modelling of unstructured indoor point cloud, ISPRS international journal of geo-information, 8, 9, 2019.

Worboys, M.: Modeling indoor space, in: Proceedings of the 3rd ACM SIGSPATIAL international workshop on indoor spatial awareness, edited by Kulik, L., Guting, R. H., and Lu, H., ACM, 2011.

Xie, J., Dai, G., Zhu, F., Wong, E. K., and Fang, Y.: DeepShape: Deep-Learned Shape Descriptor for 3D Shape Retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1335–1345, 2017.

Xie, Y., Tian, J., and Zhu, X. X.: Linking Points With Labels in 3D - A review of point cloud semantic segmentation, IEEE Geoscience and Remote Sensing Magazine, 8, 38 – 59, 2020.

Zhang, Q., Cheng, J., Wang, S., Xu, C., and Gao, X.: Point-selection and multi-level-point-feature fusion-based 3D point cloud classification, Electronics Letters, 56, 290–293, https://doi.org/10.1049/el.2019.2856, 2020.

Zhang, Y., Song, S., Tan, P., and Xiao, J.: .

Zlatanova, S. and Isikdag, U.: 3D Indoor Models and Their Applications, in: Encyclopedia of GIS, edited by Shekhar, S., Xiong, H., and Zhou, X., Springer, 2017.

Zlatanova, S., Sithole, G., Nakagawa, M., and Zhu, Q.: Problems in indoor mapping and modelling, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 40, 63–68, 2013.