Learning Functional Properties of Rooms in Indoor Space from Point Cloud Data: A Deep Learning Approach

Guoray Cai
Penn State University
University Park, PA, USA
gxc26@psu.edu

Yimu Pan
Penn State University
University Park, PA, USA
ymp5078@psu.edu

ABSTRACT
This paper presents a method to derive functional labels of rooms from the spatial configuration of room objects detected from 3D point clouds representation. The method was inspired by the intuition that spatial configuration of room objects has intimate link with the intended functional purposes. To explore the possibility of inferring the room usage information from its spatial configuration, we designed and trained a deep learning model to learn the important features of spatial configuration of room scenes and examined the predictive power of the model in inferring room usage. We present an experiment on using the model to predict room function category on Stanford 3D (S3DIS) dataset, and achieved reasonable performance. Analysis of accuracy and confusion rates allows us to draw a number insight on the separability of rooms among top level categories (such as offices, conference rooms, lounge, hallways, and storage rooms). Our findings suggested that our method is promising, with an accuracy of 91.8% on predicting room function categories. Future work should further validate and refine our method using data with more balanced training samples on the range of room types as they become available.

CCS CONCEPTS
• Applied computing → Cartography; • Computing methodologies → Image processing; Neural networks; Object recognition.

KEYWORDS
3D Models, indoor environment, room semantics, point clouds processing, deep learning

1 INTRODUCTION
People spend most of their time in indoor spaces [10]. With rapid urbanization, large indoor structures (such as high-rise business complex, public buildings, airports, and train stations) are increasingly used to provide services and infrastructures [9]. Accurate and semantically rich three-dimensional (3D) models for such indoor spaces are of profound importance to the effective use and management of space[17]. However, majority of the existing buildings do not have 3D models. Making advances in constructing 3D models of indoor space requires solving two subproblems: (1) recognizing indoor objects (structural elements and furniture), and (2) representing semantic properties of indoor space, such as their occupancy, usage, and safety. Location and geometric properties of indoor space can be derived from many sources, such as photogrammetry, 2D imagery, and 3D sensors and scanners. However, adding semantic information to 3D models of indoor environment remains largely a manual process, where human make inference on semantic information from world knowledge and tagging objects manually. Such process is time consuming and error-prone [12]. Automated methods to derive semantics of indoor space become a new research front [6].

Our work contributes to automated method in understanding the functional aspects of indoor environment at room level. First, human conceptualization of indoor space is foremost functional and social structures [16]. 3D indoor models are expected to capture semantic aspects of indoor space, in addition to the geometric, topological properties [4]. Second, semantics of indoor space should capture meanings in multiple levels of details (or granularity), from objects, to rooms, and to buildings. Current specification and implementation of 3D indoor models, such as IndoorGML [9], has been focused on the description of objects in rooms, such as doors, corridors, stairs, chairs, tables, and windows, but room-level semantics has been missing. Several recent work have pointed out the need to enrich 3D indoor models with semantic information at the room level [3, 5, 9, 14]. In particular, the usage of rooms (or function of a room) is required in applications such as safety management and code compliance [3]. Unfortunately, room usage information is not directly observable through sensed data, which presents difficulties for automation.

This paper addresses the above challenge by developing a room function classifier that infers the type of room usage based on spatial configuration of room objects in a room scene. This is based on the insight that people can infer the intended function (office, conference rooms, etc) of a room by a quick peek, without having to construct detailed knowledge about room object geometries and their topology. Human are fast and relatively effortless when interpreting the spatial layout in a room (i.e. the way that room...
With the goal of inferring room usage directly from point cloud (SLPC) representation of objects in a room and generates room (2) a slow, serial process of perceiving objects and their relationship with neighbors. These two stages activate two different neural cortex areas, the parahippocampal place area (PPA) and the retrosplenial cortex (RSC). Our room function classifier simulates the first stage of human visual processing of indoor scenes. It operates on point cloud representation of room scenes and reasons about the intended use of the rooms based on the spatial configurations of room objects. We presents an experiment to train a room function classifier using the Stanford 3D Indoor Scene Dataset (S3DIS) (http://buildingparser.stanford.edu/) and evaluate its performance in predicting intended room function. The performance of our classifier is promising, with 91.8% accuracy achieved.

The contributions of this paper are:

- We show evidence that the intended functions of a room can be (partially) inferred from the spatial configuration of semantic objects in the room with good confidence.
- We show that automated generation of room usage labels is possible.
- We forcefully argued that point cloud representation of indoor environment is a powerful source of information for enriching indoor models with semantic knowledge.

2 OVERVIEW OF OUR APPROACH

With the goal of inferring room usage directly from point cloud representation of a room in mind, we will train a room function classifier which operates on the semantically-labeled point cloud (SLPC) representation of objects in a room and generates room function labels using pre-defined categories. Our classifier uses a deep learning architecture as shown in Figure 1, which is a variant of PointNet++ [13]. It consists of a multi-layer perceptron (MLP) for local feature detection and three fully-connected layers for feature classification. The MLP takes \( n \) points as input \(^1\) and applies feature detection and transformations through (MSG1, MSG2, and MSG3).

- MSG1 takes in 32768 points and output 2048 point features. It searches features on four radius (0.4m, 0.6m, 0.8m, 1m), and will sample 16 points from each radius. Points sampled from each radius are processed by a MLP(16, 16, 32), and features detected will be synthesized via max-pooling.
- MSG2 takes in 2048 points and output 512 point features. It searches features on three radius (1.2m, 1.4m, 1.6m), and will sample 16 points from each radius. Points sampled from each radius are processed by a MLP(32, 32, 64), and features detected will be synthesized via max-pooling.
- MSG3 takes in 512 points and output 2048 point features. It searches features on three radius (1.8m, 2.0m, 2.2m), and will sample 16 points from each radius. Points sampled from each radius are processed by a MLP(64, 64, 128), and features detected will be synthesized via max-pooling.

The feature classifier (FC(512,256,128, k)) has three fully connected layers, and it takes the learned global features and outputs classification scores for \( k \) classes. A dropout rate 0.5 is applied after each layer except for the output layer (layer \( n \)).

To train the Room Function Classifier, we assume that there exists data on a set of rooms that have been annotated with room function types and object labels. We assume that there are \( M \) number of functional categories for rooms. These functional categories reflect the different intended usage of rooms. For the purpose of this study, we consider common room functions in a typical office building, such as conference rooms, lounge, offices, or storage, hallways, and WC (see Figure 2). The second condition for training the

\[\text{Figure 1: Architecture of the room function classifier}\]

\[\text{Figure 2: Six categories of room usage in office buildings}\]

Room Function classifier is that all 3D points in a room scene are semantically labeled by the object class it detected. Input to the classifier is semantically-labeled point cloud representation of room scenes, which is a four-dimensional vector \((x, y, z, \text{cls})\) (3D coordinates plus a semantic tag ‘cls’ indicating the object class it belongs to). An example of Semantically-labeled point cloud representation of a room is shown in Figure 3. We further assume that there are a limited number of object classes \(\{O_1, O_2, \ldots, O_n\}\). For indoor environment, these object classes typically include structural elements (walls, ceiling, and floors), structural surface elements (doors, windows, columns, picture frames), and furniture objects (tables, chairs, sofas, etc). Semantic labels on 3D points can be generated by either manually (by human annotators) or by an automatic processes. Automated semantic segmentation of point clouds into semantic objects is a research problem of itself, and the accuracy of even the best algorithms remains low [5, 7]. For this study, we will primarily use human manual processing of point cloud data to generated highly reliable SLPC representations of objects in rooms. This will allow us to focus on training of the room function classifier with known quality of training data.

3 EXPERIMENTAL DESIGN

We presents an experiment to explore the effectiveness of our method in inferring room usage type from semantically-labelled

\(^1\) \(n\) is set to 32768 for the experiment on S3DIS dataset
Learning Functional Properties of Rooms in Indoor Space from Point Cloud Data: A Deep Learning Approach

SIGSPATIAL '21, November 2–5, 2021, Beijing, China

Figure 3: An example of room scene represented in Semantically-labeled point cloud

point cloud representation of room scenes. We choose to use a publicly available dataset, Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS) [2] because it is already in the form of semantically-labelled point clouds. The dataset contains 3D scans from Matterport scanners in 6 areas including 272 rooms (see Figure 4).

Figure 4: Six areas in (S3DIS) Dataset

Rooms in S3DIS dataset have been manually labeled as one of the twelve (12) room types: office, conference room, auditorium, lobby, lounge, copy room, pantry, open space, storage, hallway, and WC. For the purpose of training our room function classifier, we inherited the room categories in S3DIS, but made the following changes:

- We eliminated “auditorium,” “copy room,” “open space,” and “pantry,” due to their extreme small samples (less then 5 samples).
- We also combined “lobby” and “lounge” rooms to form a new category “Lounge1.” Lobby and lounge rooms are very similar in their functions and layout.
- We employed a research assistant to go through all the rooms and verified that the room category label are consistent with

the spatial layout of this room. In case of obvious inconsistency, we believed it was an error in the original S3DIS data, and set it to the correct category. We found 5 conference rooms that were mislabeled as offices, and manually fixed them.

After this cleaning process, we end up with 224 rooms properly labeled by six categories: conference rooms, lounge, storage, hallway, offices, and WC (see figure 2). The distribution of room types is shown in Figure 5. The rooms are highly imbalanced and dominated by offices.

For each room, 3D points in the scene have been annotated with one of the semantic labels from 13 categories (chair, table, sofa, bookcase, board, floor, wall, ceiling, window, door, beam, column, and clutter). We should also note that some of the room objects were labeled as “clutter” which means they were not one of the first twelve categories. For example, there are three ceiling lights (the three bars in grey color) that were labeled as “clutter.” This is the limitation of the data.

Model training

We first divide the 224 rooms into a training set (RL) and a test set (RT). The training set RL contains 2/3 of rooms from each category in schema1, and the test set RT contains the remaining 1/3 rooms. Second, we train a model using the training set RL following the principles described in Section 2. Finally, we use the model to recognize room usage category of test set.

Measurement of performance

Accuracy: measures the number of examples classified correctly, divided by the total number of test rooms. Misclassification: measures the error rate, i.e., the number of rooms classified wrong, divided by the total number of test rooms. Empirical evidence proves that using accuracy and error rate to evaluate a classifier can be misleading. In particularly, confusions between two classes are likely to be strongly biased to favor the majority class and might produce misleading conclusions [15].

4 RESULTS

Our classifier achieved an accuracy of 91.8%. Only 8.2% of the testing rooms were classified wrong. Figure 6 shows the confusion matrix among room function categories. This result suggests the following findings:

Finding 1: (Office, conference) rooms are clearly distinct from (hallway, storage, WC) rooms in spatial layout. The confusion matrix shows that the two subsets are rarely confused by our classifier.

Finding 2: Some of the conference rooms share similar spatial layout with some of the offices. There are 33% of the conference rooms that were misclassified as offices, but only 3% of offices were misclassified as conference room. The effect of training sample imbalance is certainly at work in this case.

Finding 3: WC and hallways are clearly separable by our classifier, while storage rooms can be confused with WC or hallways,
or offices. In Figure 6, no confusion reported between 'WC' and 'hallway'. Prediction on 'storage' is 40% accurate, with confusion with hallway or WC. After checking the original point cloud data, we noticed that (WC, hallway, storage) rooms are mostly empty in furniture objects other than 'clutter'. If these 'clutter' objects were properly labeled to their true semantics (such as 'toilet'), we believe that our classifier will potentially perform better. This will be explored in the future.

Figure 6: Performance of Room Function Classifier
Accuracy: 91.8%; Misclassification: 8.2%

5 DISCUSSION AND CONCLUSION

Findings from this work offered good evidence for us to claim that spatial layout of room furniture and structural objects can be used to infer the functional properties of rooms in indoor space. Compared to other methods of inferring room semantics in the literature [2, 3, 6, 8], our method operates on the scene level representation of rooms (rather than object level representations). The closest work [2, 3, 6, 8], our method operates on the scene level representation of rooms (rather than object level representations). The closest work [2, 3, 6, 8], our method operates on the scene level representation of rooms (rather than object level representations). The closest work [2, 3, 6, 8], our method operates on the scene level representation of rooms (rather than object level representations). The closest work [2, 3, 6, 8], our method operates on the scene level representation of rooms (rather than object level representations).

Despite the promises of our method, the experimental work will affect amount of available information in the point cloud representation of room scenes.

Future work will refine our deep learning architecture and parameter settings to improve the chance of discovering signature features of room spatial layout relevant to their functions. Much more extensive testing of our method is needed using a wider variety of indoor environment (residential, commercial, office buildings, public buildings, etc), but the availability of semantically labeled large indoor point cloud datasets will continue to be the bottleneck in the foreseeable future.

REFERENCES