

# Room Classification on Floor Plan Graphs using Graph Neural Networks

Abhishek Paudel  
George Mason University  
Fairfax, Virginia, USA

Roshan Dhakal  
George Mason University  
Fairfax, Virginia, USA

Sakshat Bhattarai  
George Mason University  
Fairfax, Virginia, USA

## ABSTRACT

We present our approach to improve room classification task on floor plan maps of buildings by representing floor plans as undirected graphs and leveraging graph neural networks to predict the room categories. Rooms in the floor plans are represented as nodes in the graph with edges representing their adjacency in the map. We experiment with House-GAN dataset that consists of floor plan maps in vector format and train multilayer perceptron and graph neural networks. Our results show that graph neural networks, specifically GraphSAGE and Topology Adaptive GCN were able to achieve accuracy of 80% and 81% respectively outperforming baseline multilayer perceptron by more than 15% margin.

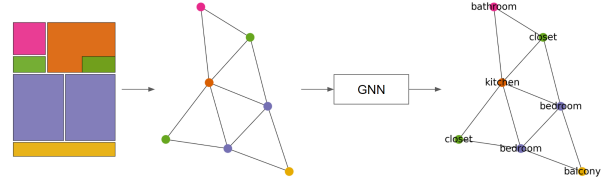
## KEYWORDS

floor plan map, graph convolutional network, graph neural network, node classification

## 1 INTRODUCTION

A floor plan is a map representing the layout of a floor in a building. Floor plans generally illustrate the locations of walls, doors windows and other structural properties in the floor. Although it has its canonical usage in architectural design, various types of maps resembling a floor plan is widely used in the context of indoor navigation by mobile robots. **Robotic mapping techniques like SLAM are commonly leveraged to construct a map of floor** of a building which the robot can then use for localization and navigation. Such generated maps contain no explicit information about room usage. However, information about room usage can be useful for robot navigation. In this paper, we present our approach to improve room classification task on floor plan maps of buildings by **representing floor plans as undirected graphs and leveraging graph neural networks to predict the room categories**. The key motivation towards this approach lies in the intuition that category of a room has some predictable relationship with the neighboring rooms. This relationship can be captured using graph neural networks that take into account the influence of neighboring nodes in the graph through message passing mechanism between nodes.

**We generate room connectivity graph from floor plans in House-GAN dataset by representing each room as nodes in the graph with edges between them representing their adjacency.** We experiment with the dataset using four types of graph neural networks: graph convolutional networks (GCN)[12], graph attention network (GAT)[21], GraphSAGE[10] and topology adaptive graph convolutional networks (TAGCN)[9], along with a baseline multi-layer perceptron (MLP). Our experiments demonstrate that GraphSAGE and TAGCN significantly outperform all other models in terms of classification accuracy. Our experiments with deepening the layers



**Figure 1: Floor plan map can be represented as an undirected graph with rooms as nodes, and edges representing their adjacency in the map. A graph neural network then takes this graph as input to make prediction of room categories as a node classification task.**

in the network also show that GraphSAGE and TAGCN do not underperform upon deepening the layers in the network in contrast to what one would generally expect due to oversmoothing[13].

## 2 RELATED WORK

Detecting and classifying elements of floor plans have been studied as a part of architectural floor plan analysis for many years using various rule-based heuristics approaches and image processing approaches[1][7][8][16][3]. **Later, machine learning approaches that use fully connected neural networks and convolutional neural networks were also used for tasks like vectorization, segmentation and detection** [14][5][6][22]. In [3] and [7], elements in floor plans are represented as a graph structure for analysis and recognition of floor plan. Most of these aforementioned tasks have focused on semantic understanding of floor plan in the context of architectural design.

With the introduction of graph neural networks (GNNs) in the recent years, a new frontier of learning over graph structures has been opened. In [11], the authors use relational graph convolutional network (RGCN) in addition to traditional machine learning to infer the usage of rooms in public buildings using floor plan represented as graph. However, their results show that RGCN does not outperform traditional machine learning approaches like random forest.

In the field of robotics, a robotic agent uses a map of the environment for localization and navigation. These maps are broadly classified into two categories: metric maps and topological maps[20]. Metric maps like occupancy grid maps[17] represent the environment in 2D grid distinguishing obstacles like walls and objects from free space where robot can navigate. Topological maps represent the environment as graph where nodes represent notable places and edges represent navigability between them generally without

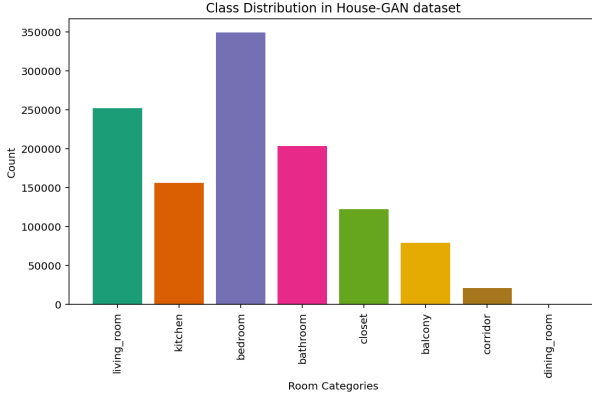


Figure 2: Class distribution in House-GAN dataset

metric information. In this context, making inferences about unknown spaces is an active field of research in robotic navigation. In [15], a method is proposed to infer semantic information from occupancy grid map represented as a scene graph. In [2], the authors propose to predict the topology and the categories of rooms given a partial map. The map of the environment is represented as a graph with rooms as nodes and their connectivity as edges. Our work aims to use graph neural networks to demonstrate that inference on such graph representations can be improved with the recent advancements in learning over graph structures.

### 3 DATASET

We experiment with **House-GAN dataset** [18] that consists of 143,184 floor plans extracted in vector format from real floor plan images in Lifull Home’s dataset [19]. There are 8 categories of rooms<sup>1</sup>.

The dataset consists of the following.

- Room bounding boxes
- Room categories
- Wall coordinates
- Wall to room mapping
- Presence of door in walls
- Name of the original RGB image from Lifull Home’s dataset

### 4 APPROACH

Our approach consists of three stages. First, we generate undirected graphs from the vectorized floor plan maps in the House-GAN dataset. Second, we extract relevant node features from the geometric and structural properties of rooms that are available in the floor plans. Finally, we train multiple graph neural networks on the floor plan graphs in inductive settings as node classification task to predict the categories of the rooms.

<sup>1</sup>The original size of the dataset and number of classes are slightly different. This is because we perform some data cleaning and also ignore irrelevant samples like those consisting of only one room in floor plan which would result in empty edges when represented as a graph.

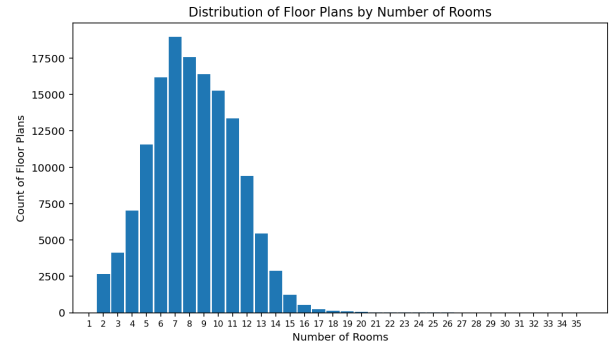


Figure 3: Distribution of floor plans by number of rooms

#### 4.1 Graph Generation

We generate the undirected graphs from vectorized floor plan maps with nodes in the graph representing the rooms, and edges representing the adjacency between two rooms. We consider two rooms to be adjacent if the distance between them is smaller than a certain threshold. This is done on the normalized bounding boxes of rooms with the threshold value set to 0.03 meaning that two rooms are adjacent if they are within 3% of the length of floor plan map.

#### 4.2 Node Feature Extraction

We extract the following node features corresponding to each room in the floor plan.

- Area of the room
- Length of the room
- Width of the room
- Number of doors in the room
- Whether a room is a parent room
- Whether a room is a child room

The first four features: area, length, width, and number of doors in the list above are self-explanatory. The last two features in the list above are useful for nested rooms i.e. room within a room. This is because some rooms in the floor plan are inside or intersect with another room and could be useful as a distinguishing feature, for example the closet could be inside a bedroom, the dining room could be inside the kitchen, and so on. If the intersection of a room with another room is greater than 70% of its area, we assume that room to be a child room and the another room that it intersects with to be a parent room. These are both binary features with two possible values: 0 or 1.

#### 4.3 Models

We experiment with multilayer perceptron (MLP) as a baseline model, and graph neural network (GNN) models. The multilayer perceptron model only takes node features as input without any graph structures that define connectivity between nodes, while the graph neural network models take both node features and the graph structures represented as a list of edges between nodes as input. For graph neural network models, we experiment with the following four variants.

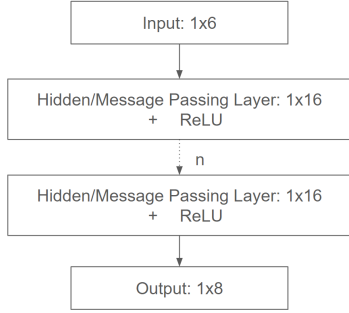


Figure 4: General Model Architecture

- **Graph Convolutional Network**

Proposed by Kipf et. al.[12], Graph Convolutional Network (GCN) is one of the first models that proposed a general approach for convolution on graphs in transductive settings. Intuitively, this method aggregates features from neighboring nodes after a linear transformation to obtain new features for the node.

- **Graph Attention Network**

Proposed by Veličković et. al.[21], Graph Attention Network (GAT) improves graph convolutions by leveraging masked self-attentional layers to address the shortcomings of prior methods and thus enabling the model to learn different weights for different nodes in a neighborhood.

- **GraphSAGE**

Proposed by Hamilton et. al.[10], GraphSAGE extends previous approaches for graph convolution to inductive settings by learning to generate node embeddings by sampling and aggregating features from the node’s local neighborhood.

- **Topology Adaptive Graph Convolutional Network**

Proposed by Du et. al.[9], Topology Adaptive Graph Convolutional Network (TAGCN) extends previous approaches by learning a fixed-size filter to perform convolution on graphs that resembles convolutions on images. It does not approximate convolution, unlike previous methods.

We implement a general model architecture as shown in Figure 4 for MLP and all GNN models for a fair comparison between them. The input layer takes in a vector of size 6 corresponding to 6 node features. We experiment with varying number of hidden or message passing layers each with a size of 16 and followed by ReLU activation function. The output layer has a size of 8 corresponding to 8 categories of rooms.

## 5 EXPERIMENTAL SETUP

We split the House-GAN dataset into two parts for training and testing. We take 120,000 floor plans for training and the remaining 23,184 floor plans for testing. Preliminary small-scale experiments were run on a small fraction of the dataset for hyperparameter tuning before training on the full dataset. For large-scale training on full dataset, we use the following setup and hyperparameters for training.

Table 1: Classification Accuracy

Model	Train Accuracy	Test Accuracy
MLP	0.6585	0.6574
GCN	0.5469	0.5485
GAT	0.5269	0.5250
GraphSAGE	0.8068	0.8026
TAGCN	0.8140	0.8107

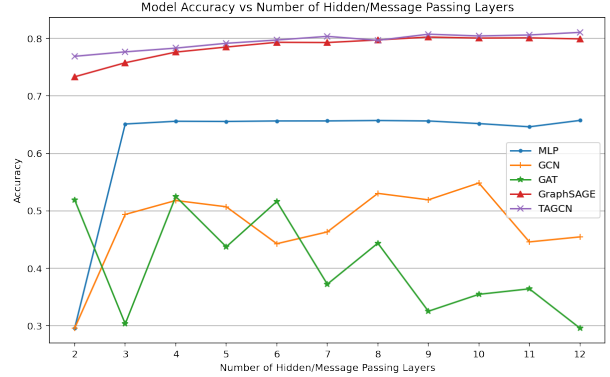


Figure 5: Visualization of test accuracy for each model with respect to the number of hidden/message passing layers

- GPU: NVIDIA Tesla K80 with 8GB memory
- Number of training epochs: 100
- Loss function: Cross-entropy loss
- Gradient descent optimizer: Adam
- Batch size: 128
- Learning rate: 0.004
- Learning rate scheduling: Exponential with step size of 10 and gamma of 0.8

We use PyTorch and PyTorch Geometric for our experiments. Large-scale experiments were run on ARGO, a research computing cluster provided by the Office of Research Computing at George Mason University. The cumulative training time to run all experiments with all models was approximately 75 hours.

## 6 RESULTS

The best results obtained in terms of classification accuracy for each model are shown in Table 1. The best performance is achieved by TAGCN with 81.07% test accuracy followed by GraphSAGE with test accuracy of 80.26%. We observe that GCN and GAT had very low test accuracy of 54.85% and 52.50% respectively, and performed even worse than MLP which had the test accuracy of 65.74%. It should also be noted that although we have not used any explicit regularization techniques in our models, the results demonstrate that overfitting is not an issue since all the models seem to have generalized well to the test set.

To study the effects of oversmoothing which is common in graph neural networks, we train our models with a varying number of hidden/message passing layers ranging from 2 to 12 layers. The results

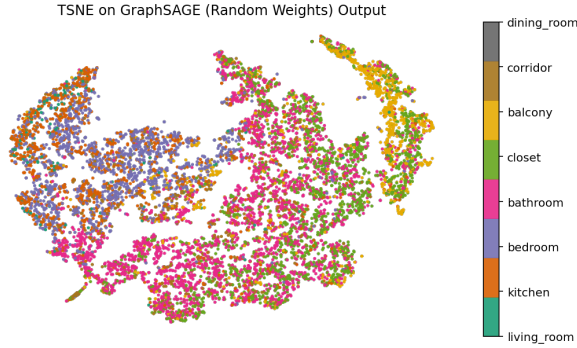


Figure 6: t-SNE on node embeddings obtained from GraphSAGE model initialized with random weights. Partial clusters for some classes are distinguishable.

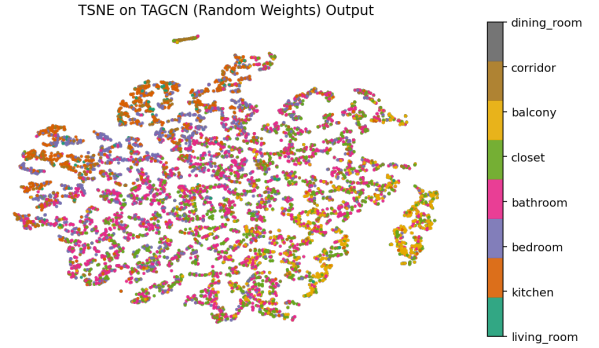


Figure 7: t-SNE on node embeddings obtained from TAGCN model initialized with random weights. Partial clusters for some classes are distinguishable.

are shown in Figure 5. We observe that MLP has no improvement in accuracy after 3 hidden layers. GraphSAGE and TAGCN have considerable improvements early on with only slight improvements as the number of layers increase. However, we do not see GCN and GAT perform better with more message passing layers. In fact, the performance of GAT generally keeps decreasing upon increasing the number of message passing layers which suggests that GAT might be more susceptible to oversmoothing. Surprisingly, both GraphSAGE and TAGCN have no decrease in performance upon increasing the number of message passing layers. This suggests that GraphSAGE and TAGCN are more robust to oversmoothing than other models.

To understand why GCN and GAT were not able achieve a good performance unlike TAGCN and GraphSAGE both of which had superior performance, we study the characteristics of these models in terms of their abilities to capture relational inductive biases that are inherent to the graph structures due to the edges in the graph affecting how node features interact during message passing in the graph neural network[4]. To study this, we instantiate the GNN models with random weights and obtain the node embeddings for 10,000 samples of floor plan graphs from test split by propagating them through the untrained models. Then, we perform non-linear dimensionality reduction on the obtained node embeddings using t-Distributed Stochastic Neighbor Embedding (t-SNE). We observe that GraphSAGE and TAGCN were able to partially capture relational inductive biases even before training the network as demonstrated in Figure 6 and 7. In both of the figures, we can see partial clusters for some of the classes in the dataset. For example, bedroom, bathroom and balcony have partially distinguishable clusters. We did not observe any meaningful clusters with GCN and GAT suggesting that these models were not quite capable of capturing relational inductive biases in the dataset. Trivially, no meaningful clusters were observed with MLP or with original features as one would generally expect strengthening our hypothesis that GraphSAGE and TAGCN are better at capturing relational inductive biases in the dataset resulting in better performance after training.

## 7 CONCLUSION AND FUTURE WORK

Overall, our experiments demonstrate that graph neural networks can be leveraged to improve performance on room classification using floor plans represented as undirected graphs. Specifically, we observe that TAGCN and GraphSAGE show superior performance compared to other models like GCN, GAT and MLP we experimented with for this task. It was a surprising finding that GCN and GAT, which leveraged graph structures could not even perform on par with MLP which did not leverage graph structures. Additionally, our experiments to study relational inductive bias on untrained graph neural networks suggest that GCN and GAT are likely not very capable for capturing relations between the nodes in the graph resulting in subpar performance. However, further experiments would be needed to confirm this hypothesis. We also observe decreased performance of GAT as number of message passing layers were increased while GCN only has decreased performance when number of layers is 11 or 12. These results are likely due to oversmoothing. However, it was intriguing that GraphSAGE and TAGCN had either improved or similar performance as more layers were stacked. This suggests that GraphSAGE and TAGCN are more robust to oversmoothing than the other models. And again, more experiments would be needed to conclusively assert this statement.

In terms of improving the performance over our current approach, it would be interesting to explore whether adding more node features could be useful. **One could also eliminate the need for manual feature extraction by using CNN. One possible direction for this could be to use the image of the floor plan with specific room segmented as node property and leveraging CNN to extract features from these images to be used as node features.** This would however be computationally more expensive than our current approach.

## ACKNOWLEDGMENTS

All experiments were run on ARGO, a research computing cluster provided by the Office of Research Computing at George Mason University. We would like to thank the Office of Research Computing for providing the computing infrastructure to run these experiments.

## REFERENCES

- [1] Sheraz Ahmed, Marcus Liwicki, Markus Weber, and Andreas Dengel. 2012. Automatic room detection and room labeling from architectural floor plans. In *2012 10th IAPR international workshop on document analysis systems*. IEEE, 339–343.
- [2] Alper Aydemir, Patric Jensfelt, and John Folkesson. 2012. What can we learn from 38,000 rooms? Reasoning about unexplored space in indoor environments. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 4675–4682.
- [3] Alessio Barducci and Simone Marinai. 2012. Object recognition in floor plans by graphs of white connected components. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 298–301.
- [4] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. 2018. Relational inductive biases, deep learning, and graph networks. arXiv:1806.01261 [cs.LG]
- [5] Lluís-Pere de las Heras, Sheraz Ahmed, Marcus Liwicki, Ernest Valveny, and Gemma Sánchez. 2014. Statistical segmentation and structural recognition for floor plan interpretation. *International Journal on Document Analysis and Recognition (IJ DAR)* 17, 3 (2014), 221–237.
- [6] Samuel Dodge, Jiu Xu, and Björn Stenger. 2017. Parsing floor plan images. In *2017 Fifteenth IAPR international conference on machine vision applications (MVA)*. IEEE, 358–361.
- [7] Bernardino Dominguez, ÁL García, and Francisco R Feito. 2012. Semiautomatic detection of floor topology from CAD architectural drawings. *Computer-Aided Design* 44, 5 (2012), 367–378.
- [8] Philippe Dosch, Karl Tombre, Christian Ah-Soon, and Gérald Masini. 2000. A complete system for the analysis of architectural drawings. *International Journal on Document Analysis and Recognition* 3, 2 (2000), 102–116.
- [9] Jian Du, Shanghang Zhang, Guanhang Wu, Jose M. F. Moura, and Soumya Kar. 2017. Topology Adaptive Graph Convolutional Networks. arXiv:1710.10370 [cs.LG]
- [10] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. arXiv:1706.02216 [cs.SI]
- [11] Xuke Hu, Hongchao Fan, Alexey Noskov, Zhiyong Wang, Alexander Zipf, Fuqiang Gu, and Jianga Shang. 2021. Room semantics inference using random forest and relational graph convolutional networks: A case study of research building. *Transactions in GIS* 25, 1 (2021), 71–111. <https://doi.org/10.1111/tgis.12664>
- [12] Thomas N. Kipf and Max Welling. 2016. Semi-Supervised Classification with Graph Convolutional Networks. arXiv:1609.02907 [cs.LG]
- [13] Guohao Li, Matthias Müller, Ali Thabet, and Bernard Ghanem. 2019. DeepGCNs: Can GCNs Go as Deep as CNNs? arXiv:1904.03751 [cs.CV]
- [14] Chen Liu, Jiajun Wu, Pushmeet Kohli, and Yasutaka Furukawa. 2017. Raster-to-vector: Revisiting floorplan transformation. In *Proceedings of the IEEE International Conference on Computer Vision*. 2195–2203.
- [15] Ziyuan Liu and Georg von Wichert. 2014. Extracting semantic indoor maps from occupancy grids. *Robotics and Autonomous Systems* 62, 5 (2014), 663–674.
- [16] Sébastien Macé, Hervé Locteau, Ernest Valveny, and Salvatore Tabbone. 2010. A system to detect rooms in architectural floor plan images. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. 167–174.
- [17] Hans Moravec and Alberto Elfes. 1985. High resolution maps from wide angle sonar. In *Proceedings. 1985 IEEE international conference on robotics and automation*, Vol. 2. IEEE, 116–121.
- [18] Nelson Nauata, Kai-Hung Chang, Chin-Yi Cheng, Greg Mori, and Yasutaka Furukawa. 2020. House-GAN: Relational Generative Adversarial Networks for Graph-constrained House Layout Generation. arXiv:2003.06988 [cs.CV]
- [19] Informatics Research Data Repository. 2020. LIFULL HOME’S Dataset. <https://www.nii.ac.jp/dsc/idr/en/lifull/>
- [20] Sebastian Thrun. 1998. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence* 99, 1 (1998), 21–71.
- [21] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. Graph Attention Networks. arXiv:1710.10903 [stat.ML]
- [22] Zhiliang Zeng, Xianzhi Li, Ying Kin Yu, and Chi-Wing Fu. 2019. Deep floor plan recognition using a multi-task network with room-boundary-guided attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9096–9104.