

# LayoutGMN: Neural Graph Matching for Structural Layout Similarity

## \*\*Supplementary Material\*\*

Akshay Gadi Patil<sup>1</sup> Manyi Li<sup>1†</sup> Matthew Fisher<sup>2</sup> Manolis Savva<sup>1</sup> Hao Zhang<sup>1</sup>

<sup>1</sup>Simon Fraser University      <sup>2</sup>Adobe Research

### 1. Additional Results

We show more top-5 retrieved results for different queries of UI layouts and floorplans in Figure 1, 5, 6 and 7. In presenting these results, we randomly picked the queries from the set of N queries (N=50 for UIs and 100 for floorplans) which were used to get *Precision@k* scores for the baselines discussed in the main paper and also tabulated in Table 1 there. In Figure 1, we present results using the IoU metric, alongside retrieved results using the state-of-the-art GCN-CNN network [2] and LayoutGMN. In Figures 5, 6, 7, we show comparative results on only two learned methods, viz., GCN-CNN [2], and LayoutGMN. We would also like to point out that in the main paper (L 571), we promised to show results on document layouts, but were unaware of the submission policy for supplementary material (which prohibits presenting results on additional datasets). We, therefore, skip showing results on document layouts.

### 2. Attention Visualizations

LayoutGMN compares two layouts structurally via attention-based Graph Matching mechanism, in addition to message propagation within individual graphs. The former provides local structural correspondences, whereas the latter provides global structural prior for comparing two layouts. Specifically, if there exist  $m$  semantic elements in layout  $I_1$  and  $n$  semantic elements in layout  $I_2$ , the attention-weight matrix for matching elements in  $I_2$  w.r.t elements in  $I_1$  is of size  $n \times m$ , and vice-versa. These attention weights change from layer-to-layer depending on the the structural match. In Figure 2, we present two examples of floorplans with attention weights visualized in all 6 layers, with layer-0 being the layer where weights are initialized prior to training. For brevity, we just present floorplan attention visualization, and only show the largest attention weights, omitting all other (insignificant) connections.

### 3. Crowd-sourced Relevance Judgements

In the main paper, we mentioned that the *Precision@k* scores [3] were obtained using crowd-annotated responses on the relevance of the returned results for a given query. This crowd annotation was done on Amazon Mechanical Turk (AMT), for both, UI layouts [1] and floorplans [4]. The design of the questions plays a crucial role in validating the performance of a network employed for retrieval task. We, therefore, design our AMT response study on UI layouts in a similar manner as carried out in Manandhar et al. [2]. A snapshot of a question visible to turkers for tagging structurally similar results for a given query of UI layouts is shown in Figure 3. Such set of questions are shown for all the baseline methods enumerated in Section 5.1 in the main paper. For floorplans, we design our crowd-annotation study on AMT in a similar fashion. The set of instructions on which a user should base her relevance judgments for a given floorplan query are shown in Figure 4.

### 4. Fully connected vs Adjacency Graphs

All the quantitative results presented in the main paper are based on fully-connected graphs, for all the methods. We observed, both quantitatively and qualitatively (Fig 6, Table 1,2,3 in the main paper), that fully-connected graphs are a good input representation for learning structural similarity on layouts. We also experimented with adjacency graphs, on both, floorplans as well as UI layouts. As explained in the main paper, we observed that, for floorplans (where the graph node count is small), the quality of retrievals improved in the case of LayoutGMN, but degraded for GCN-CNN. A set of results for the same is shown in Figure 5. This is mainly because GCN-CNN obtains independent graph embeddings for each input graph and when the graphs are built only on adjacency connections, some amount of global structural prior is lost. On the other hand, GMNs obtain better contextual embeddings by now matching the sparsely connected adjacency graphs, as a result of narrower search space. However, for UIs (where the graph

<sup>†</sup> Corresponding Author: manyil@sfu.ca

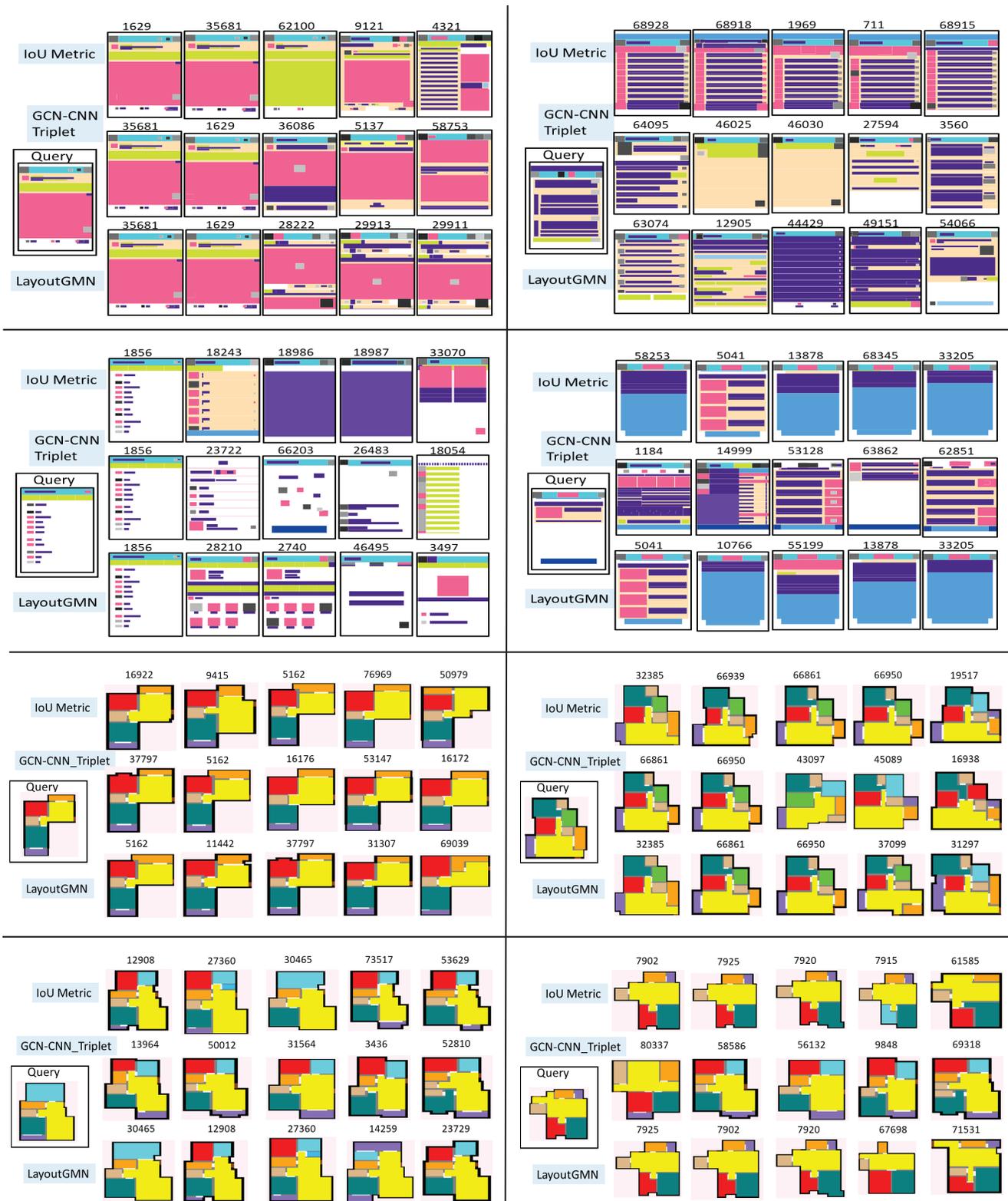


Figure 1. More Results: Top-5 retrieved results for an input query based on IoU metric, GCN-CNN\_Triplet [2] and LayoutGMN, on UI designs (first two rows), followed by floorplans. These set of queries were randomly chosen, and are a part of the larger set of  $N$  queries ( $N=50$  for UIs and 100 for floorplans) used to get  $Precision@k$  scores via crowd-sourced relevance judgements. Also see Fig 5, 6, 7.

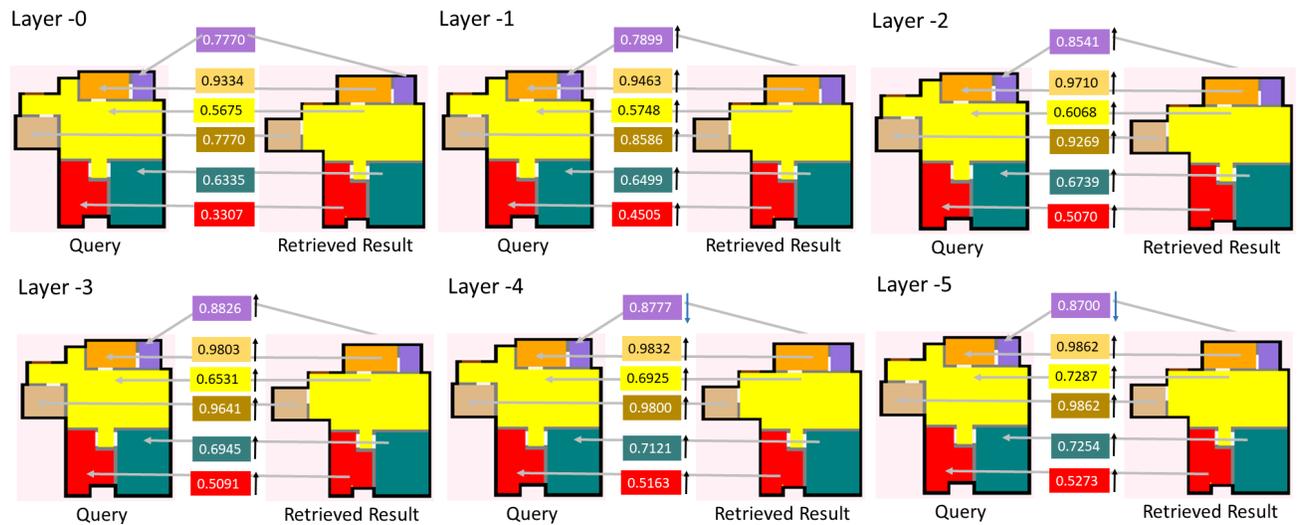
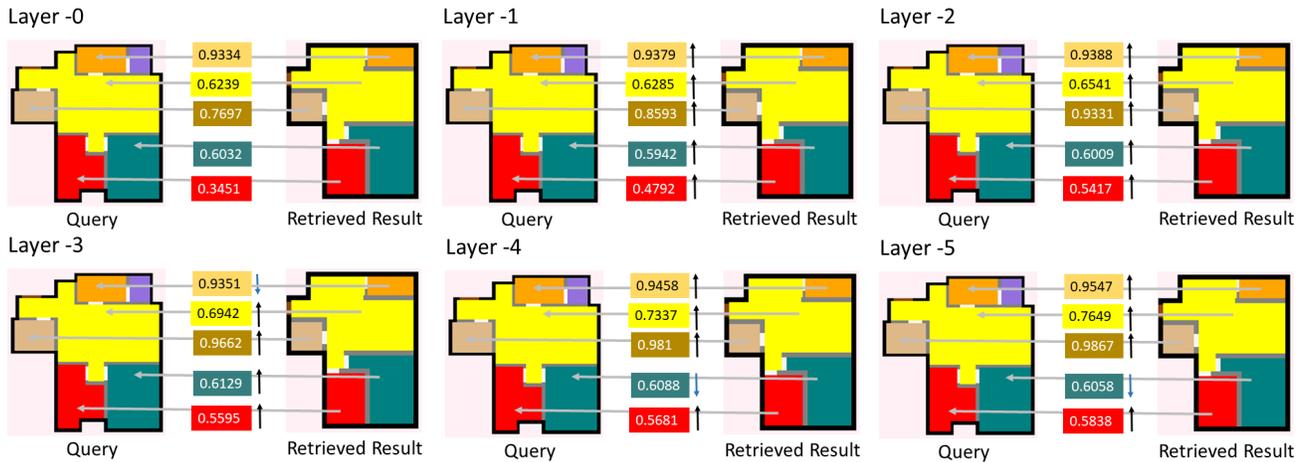


Figure 2. Given a query (on the left) and its retrieved result (on the right), we show attention weights in different layers of message propagation, leading to element correspondences, from which structural similarity is driven and partly established. Layer-0 to Layer-5 show *learned* attention weights in different layers of propagation. For brevity, we only show the largest weights.

node count is large), the elements are scattered all over the layout, and no one heuristic is able to capture adjacency relations perfectly. The quality of retrievals for both the networks degraded when using adjacency graphs on UI layouts. A set of such results is shown in Figure 6.

## 5. Retrieval Stability

In the main paper, we developed a new metric, called *Overlap@k* scores, to measure the stability of retrievals using different methods. This score measures the ability of the layout similarity metric to replicate the distance field implied by a query according to its top-ranked retrieval. Quantitative results for the same are shown in Table 2 in the main paper. In this manuscript, we present qualitative results for the same in Figure 7.

## References

- [1] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afegan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. Rico: A mobile app dataset for building data-driven design applications. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 845–854, 2017. 1
- [2] Dipu Manandhar, Dan Ruta, and John Collomosse. Learning structural similarity of user interface layouts using graph networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1, 2, 4, 6, 7, 8
- [3] Christopher D Manning, Hinrich Schütze, and Prabhakar Raghavan. Chapter 8: Evaluation in information retrieval in “Introduction to Information Retrieval”. pages 151–175. Cambridge university press, 2008. 1
- [4] Wenming Wu, Xiao-Ming Fu, Rui Tang, Yuhan Wang, Yu-

## Title: Tag structurally similar images for a given query.

### Description:

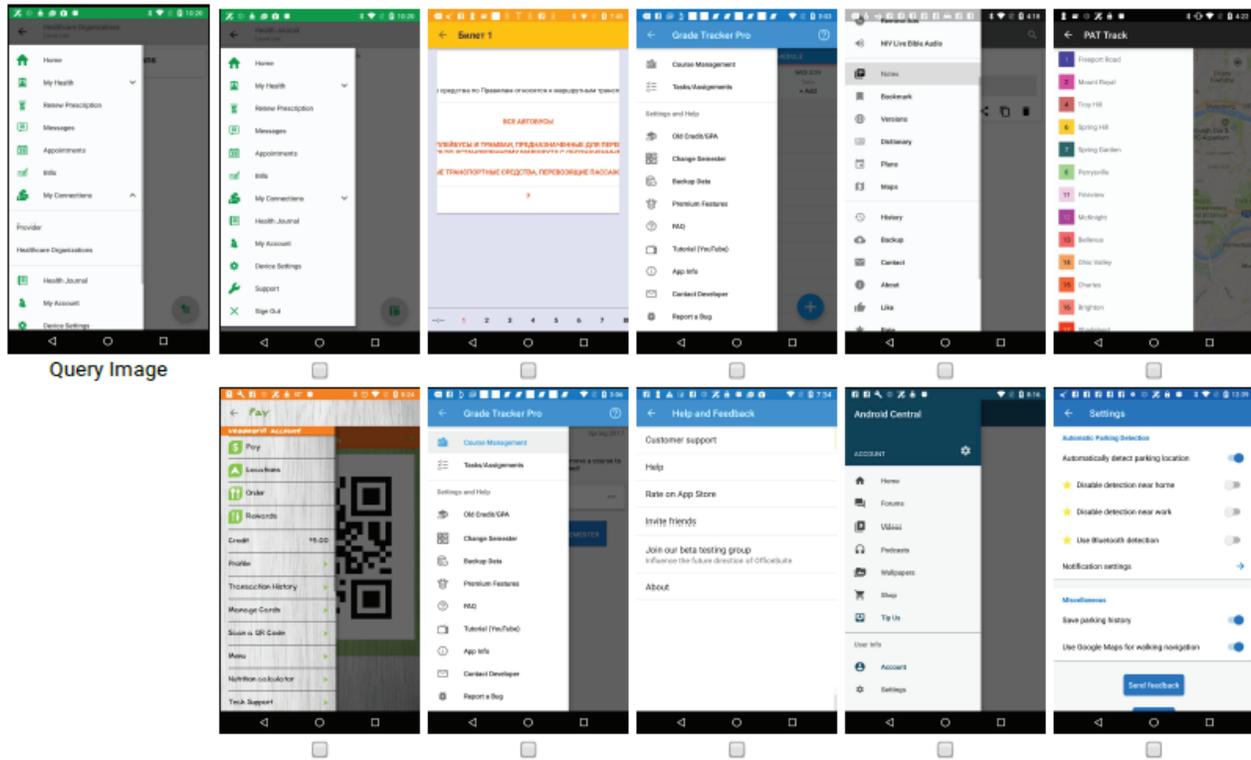
A set of mobile User-Interface snapshots are shown below.

Specifically, a query image is shown on the left, and 10 different images are shown on the right against it.

Taking the query image as the reference, select all the images that you think are structurally similar to the given query image.

### Note:

Ignore the color and content of the texts and visuals, and make a selection if the structure of the UI images shown on the right match that of the query image on the left.



Submit

Figure 3. Snapshot of a question visible to turkers on Amazon Mechanical Turk (AMT) to get relevance judgments of returned results for a given UI layout query. Our design of this study on UI layouts is similar to the one used in the state-of-the-art work on structural similarity by Manandhar et al. [2].

Hao Qi, and Ligang Liu. Data-driven interior plan generation for residential buildings. *ACM Transactions on Graphics (TOG)*, 38(6):1–12, 2019. 1

## Title: Tag structurally similar images for a given query.

### Description:

A building floorplan is represented as an image, where the colored boxes represent different kinds of rooms.

A query image is shown on the left, and 10 different images are shown on the right against it.

Taking the query image as the reference, select all the images that you think are *structurally* similar to the given query image.

### Note:

The box colors matter. They should NOT be ignored.

The interface displays a query image on the left, labeled "Query Image", which is a floorplan composed of several colored rectangular rooms: teal, red, yellow, light blue, brown, purple, and orange. To the right of the query image are two rows of five candidate images. Each candidate image is a floorplan with a similar layout to the query image but with some color variations. Below each candidate image is a small grey checkbox for selection. At the bottom left of the interface is an orange button with the text "Submit".

Figure 4. Example of a question presented to turkers on AMT to get relevance judgements on of the returned results for a given floorplan query.

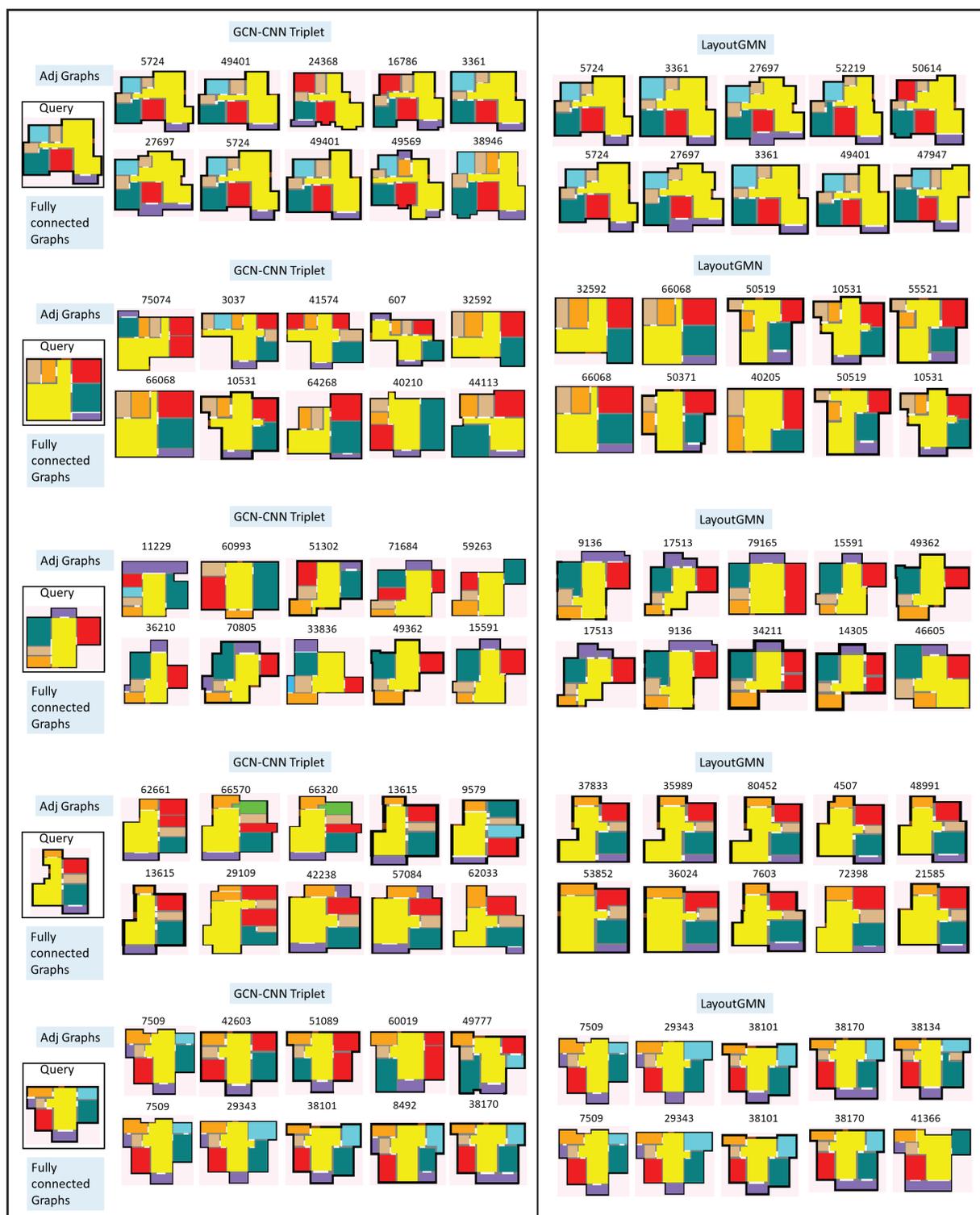


Figure 5. Additional retrieved results on floorplan queries, using adjacency graphs, and fully-connected graphs, using both, GCN-CNN [2] (left column), and LayoutGMN (right column). Note that all the quantitative results shown in the main paper are based on fully-connected graphs, following the design choice of [2].

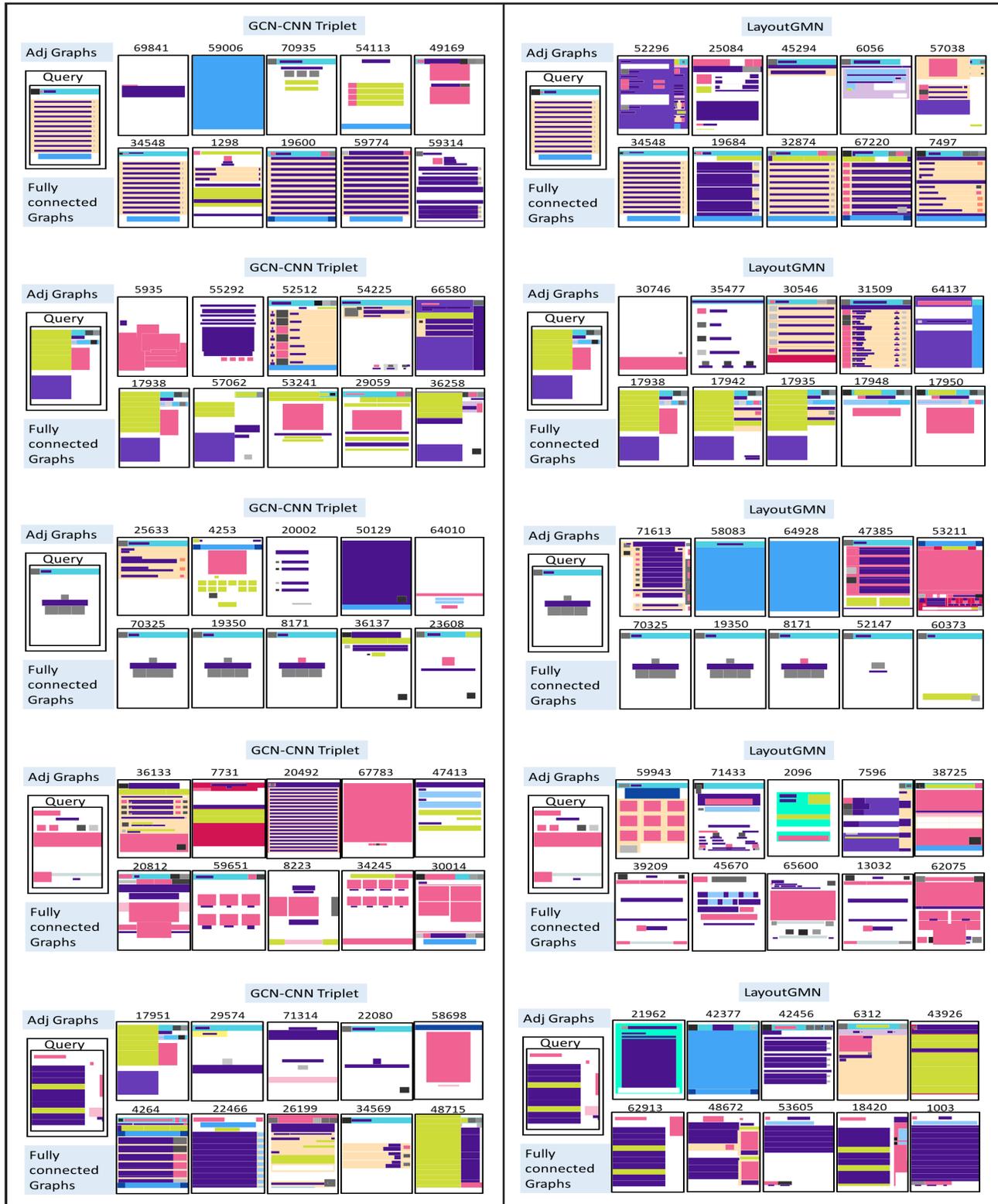


Figure 6. Additional retrieved results on UI layout queries, using adjacency graphs, and fully-connected graphs, using both, GCN-CNN [2] (left column), and LayoutGMN (right column). Note that all the quantitative results shown in the main paper are based on fully-connected graphs, following the design choice of [2].

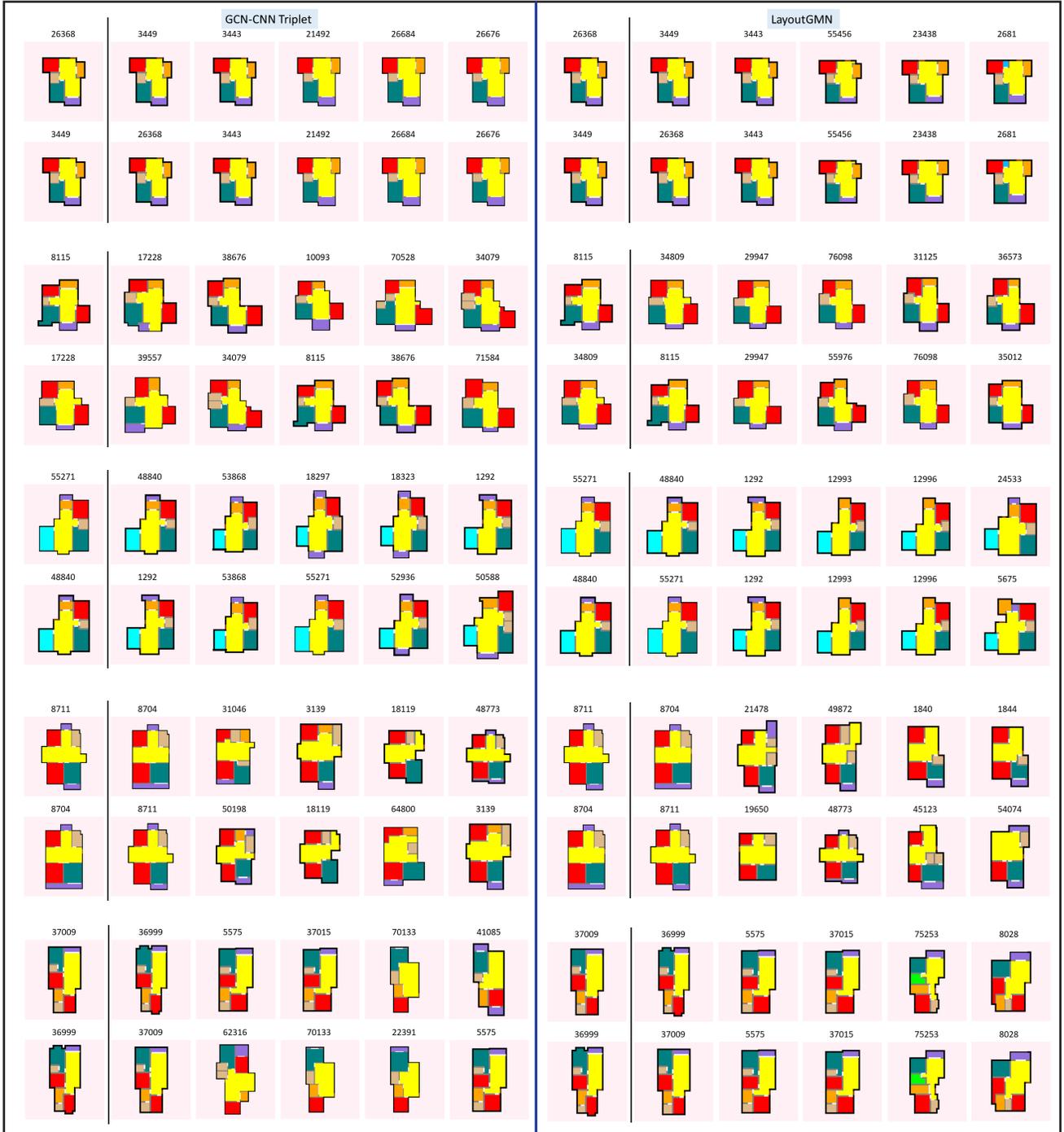


Figure 7. Retrieved results for a given query and its top ranked retrieval, using GCN-CNN [2] (left column) and LayoutGMN (right column). In every set of paired results (row-wise), the first row represents a query  $q$  and its top-5 retrievals. In the second row, the query is the top-1 result of query  $q$  in the first row, denoted by  $q^{top-1}$ . Its top-5 retrievals shown against it.