SCIENCE CHINA
Information Sciences

• LETTER •

Indoor Layout Programming via Virtual Navigation Detectors

Qiang Fu[1], Hongbo Fu[2], Zhigang Deng[3] & Xueming Li[1]*

[1]*School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing* 100876*, China;*
[2]*School of Creative Media, City University of Hong Kong, Hong Kong* 999077*, China;*
[3]*Department of Computer Science, University of Houston, Houston, TX* 77204*, USA*

Citation Fu Q, Fu H B, Deng Z G, Li X M. Indoor Layout Programming via Virtual Navigation Detectors. Sci China Inf Sci, for review

Dear editor,

As one of the fundamental tasks in indoor scene modeling, indoor layout programming has been extensively studied in the past decades. Some existing approaches adopt a heuristic exploration that generates the arrangement of indoor objects from an initial state (e.g., a messy indoor scene or manual arrangement by users), following the pre-specified constraints such as object-object or object-human relations of indoor scenes. However, these works always focused on the *relative* object arrangement, while ignoring the impact of the room environment. That is mainly because man-made guidelines for layout programming can only cover limited factors such as object alignment, while actual situations could involve more sophisticated factors.

Many design details and priors are embedded in the layouts of indoor scenes created by professional interior designers; thus, some methods attempted to leverage layout examples to generate plausible indoor layouts. However, even though they can achieve *global* solutions that consider room configurations such as the room size, the positions of windows and doors, the variations of their solutions are fundamentally limited by the shapes of the indoor scene examples. In other words, given a room with irregular shape (e.g., non-rectangle rooms or rooms with curved walls), it would be difficult for these methods to use their defined room features which always focus on rectangle rooms to describe the input room, and also not easy to find proper layout examples. We observe that in reality, two rooms with different configurations might have similar paths between two certain locations, and the observations on the paths are also similar. Therefore, an intriguing, widely open research question arises: *Can we leverage limited indoor scene examples to tackle the layout programming of indoor scenes with irregular room configurations?*

In this work, we propose to leverage a virtual navigation detector to explore plausible paths of the given room for certain indoor objects to the door, in order to determine their positions/directions. Specifically, we choose the indoor scene example with similar paths to the given room with respect to certain related objects as the reference, which is a drastic turn from existing methods that choose an example solely based on the similarity of room configurations. In this way, the task of layout programming can be re-formulated as searching for certain paths of the indoor objects guided by indoor scene examples (Figure 1).

Virtual Navigation Detector. The virtual navigation detector contains both the navigation and classification models. The first one is a reinforcement learning model aiming at generating candidate paths in a given room, while the second one is a deep classifier based on transfer learning via the ImageNet-pretrained ResNet-50 [1]. We sample the 2D bounding box of the floor plan for both the given room and the dataset scenes (from the SUNCG dataset [2] with furniture removed), and set cameras with four different directions (i.e., up, down, left, right) to capture its observation images at each cell.

We adopt the method of [3] as the network architecture of our navigation model, and use the same reward mechanism as [3] to train the reinforcement learning model with four actions including move forward, move back, turn left, or turn right. The trained navigation model is able to generate plausible paths that suit for the configuration of the given room. The classification model is trained via transfer learning, which fine-turns the ImageNet-pretrained ResNet-50 with the observation images of the dataset indoor scenes. In this manner, given a series of observation images captured along a candidate path of the given room, the classifier can find a proper class, i.e., a certain indoor scene example, as the reference to guide the layout programming.

Layout Programming. For a given room, let $\mathbf{P}_i = \{p^i_0, p^i_1, \cdots, p^i_k\}$ be the i-th candidate path that has a total of k states along the path. The ResNet feature of the associated observation image for the state p is denoted as $f(p)$. Let $\tilde{p}^j$ be the state where a certain category of object was ever placed in the reference room from the datasets, suggested by the trained classification model (e.g., of the j-th class), we define the following distance metric for path

*Corresponding author (email: lixm@bupt.edu.cn)

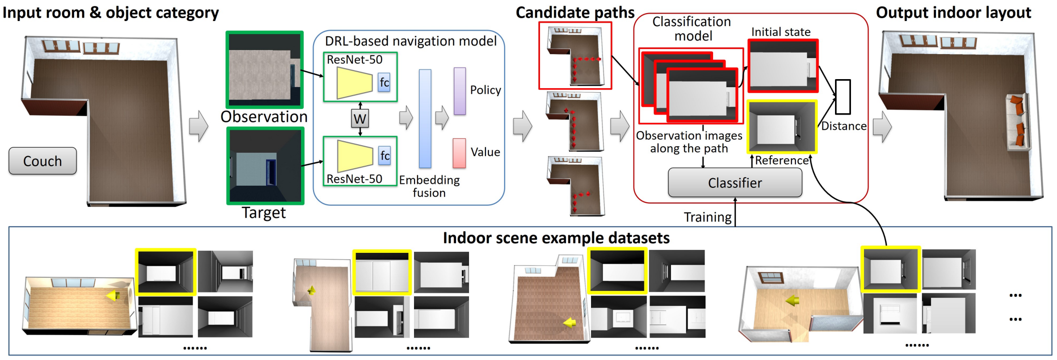

Figure 1 The pipeline of our method. Aiming at placing the given objects into the input room, our method has a navigation model trained by the input room with the observation at the current state and the view towards a door as the target, and a classification model trained by lightweight indoor scene datasets with respect to different object categories. The navigation model is to generate candidate paths, and the classification model is to evaluate the paths and choose the proper one, to determine the position/direction for the placement of the given object.

selection:

$$d(P_i) = ||f(p_0^i) - f(\tilde{p}^j)||_2^2 + \delta(p_0^i). \tag{1}$$

In the above equation, j is determined by $\arg\max_j N(\mathbf{P}_i, j)$, where $N(\mathbf{P}_i, j)$ is the number of states (in the set of $\mathbf{P}_i$), whose observation images are classified to the j-th class by the trained classification model. That is, we choose the class that most observation images are classified into for a certain path. $\delta(p_0^i)$ is used to detect whether object collisions will occur in the position where the object would be placed, namely, if no collision exists in the initial state of the path, $\delta(p_0^i) = 0$; and $\delta(p_0^i) = \inf$ otherwise. Note that, if multiple objects are placed in the same room, the order of placements will impact the programmed layout.

From the candidate paths, we choose the one with the minimum distance $d(P_i)$, and use the initial state of the selected path as the position to place the associated indoor object. We try all four directions at this position to determine the object direction which would have the minimum feature distance. Our method allows the user to browse the suggested paths in turns to choose the preferred one. We adopt a flexible sampling step size to generate the grid of the room, to limit the number of cells on the grid and ensure the time complexity of the navigation model training stage to be close to constant. To improve computing efficiency, we only consider *core furniture* in the room and use the positions/directions of the core furniture to arrange the associated objects. For example, in a given living room, we train the classification model for couch (the core furniture) but ignore the associated objects such as coffee table and TV set. These associated objects can be arranged by pre-specifying the possible ranges of the relative positions/directions to their associated core furniture.

Experiments. We show several indoor scenes whose layouts are programmed by the virtual navigation detector in Appendix A. On average, the time cost for training the navigation model for a given room takes less than 10 minutes to process 1 million frames of observation images on an off-the-shelf PC. Once the virtual navigation detector is trained, the layout programming takes less than 10 seconds. The experiments are conducted on a PC with Intel Core i7-8700K 3.70GHz CPU with 32GB RAM and NVIDIA GeForce RTX 2080 Ti GPU. To validate the effectiveness of our method for layout programming, especially for non-rectangle rooms, we conducted a user study to compare the indoor layouts created by our method, and those by the state-of-the-art data-driven indoor synthesis methods [4,5]. The results (see Appendix B) show that our method performs better on non-rectangle rooms compared with [4,5], benefited from the navigation model which makes our results adaptive to the non-rectangle room inputs.

Conclusion. We proposed a novel approach to programming indoor scene layouts, based on the local observation features captured by virtual navigation detectors. It extends the applicability of limited indoor scene examples as the references to guide the layout creation of rooms with complex configurations. The effectiveness of the proposed method was demonstrated through several synthesized indoor scenes and comparisons.

Acknowledgements This work was partially supported by grants from the NSFC (No. 61902032), NSF (IIS-1524782), City University of Hong Kong (No. 7004915), the Fundamental Research Funds for the Central Universities, the Open Project Program of the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (VRLAB2018C11, VRLAB2019B01) and Shenzhen Research Institute, City University of Hong Kong.

Supporting information Appendix A-B. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778.

2 Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. 2017. Semantic Scene Completion from a Single Depth Image. IEEE Conference on Computer Vision and Pattern Recognition (2017), 190–198.

3 Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J Lim, Abhinav Gupta, Li Fei-Fei, and Ali Farhadi. 2017. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In 2017 IEEE international conference on robotics and automation (ICRA). IEEE, 3357–3364.

4 Qiang Fu, Xiaowu Chen, Xiaotian Wang, Sijia Wen, Bin Zhou, and Hongbo Fu. 2017. Adaptive synthesis of indoor scenes via activity-associated object relation graphs. ACM Transactions on Graphics (TOG) 36, 6 (2017), 201.

5 Kai Wang, Manolis Savva, Angel X Chang, and Daniel Ritchie. 2018. Deep convolutional priors for indoor scene synthesis. ACM Transactions on Graphics (TOG) 37, 4 (2018), 70.