

# Data-driven Interior Plan Generation for Residential Buildings

*Supplementary material*

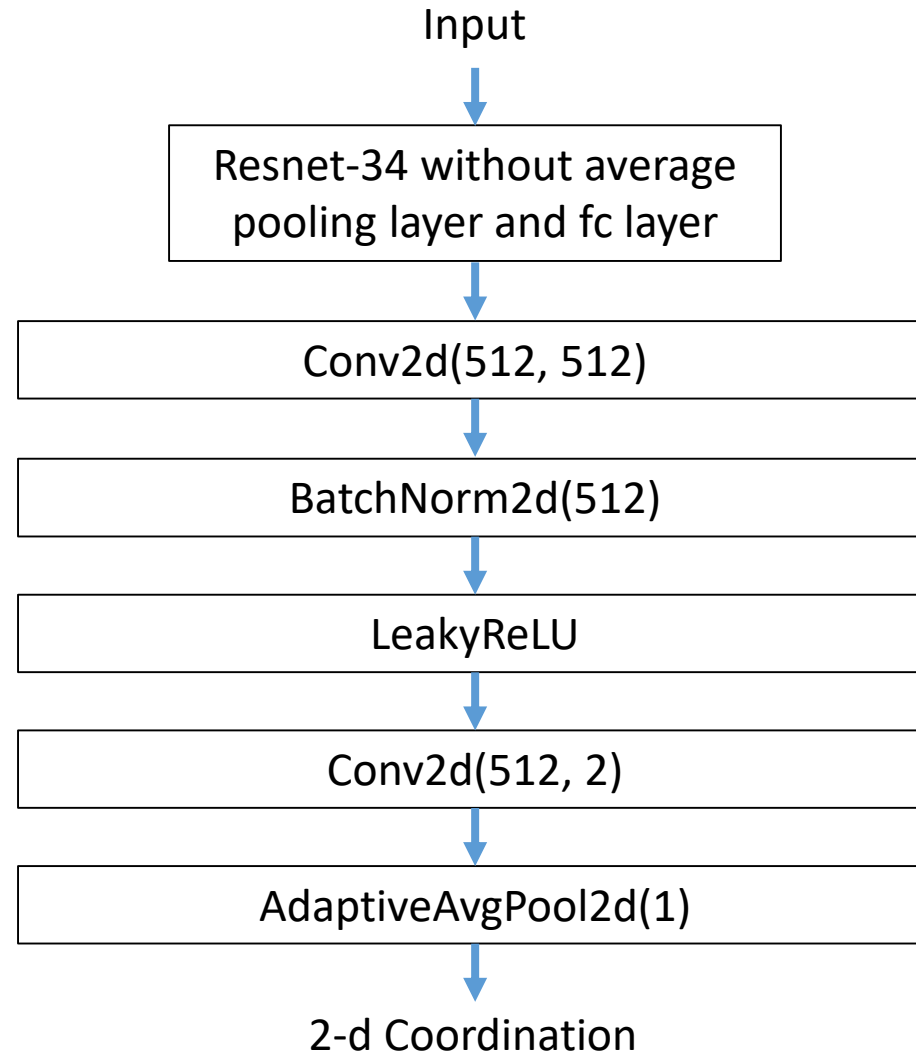
# Outline

- Networks and implementation details
- Floor plans generated by our method
- Human-created floor plans for Figs. 11, 12, 13 in the paper
- Questionnaire for comparison to human-created floor plans

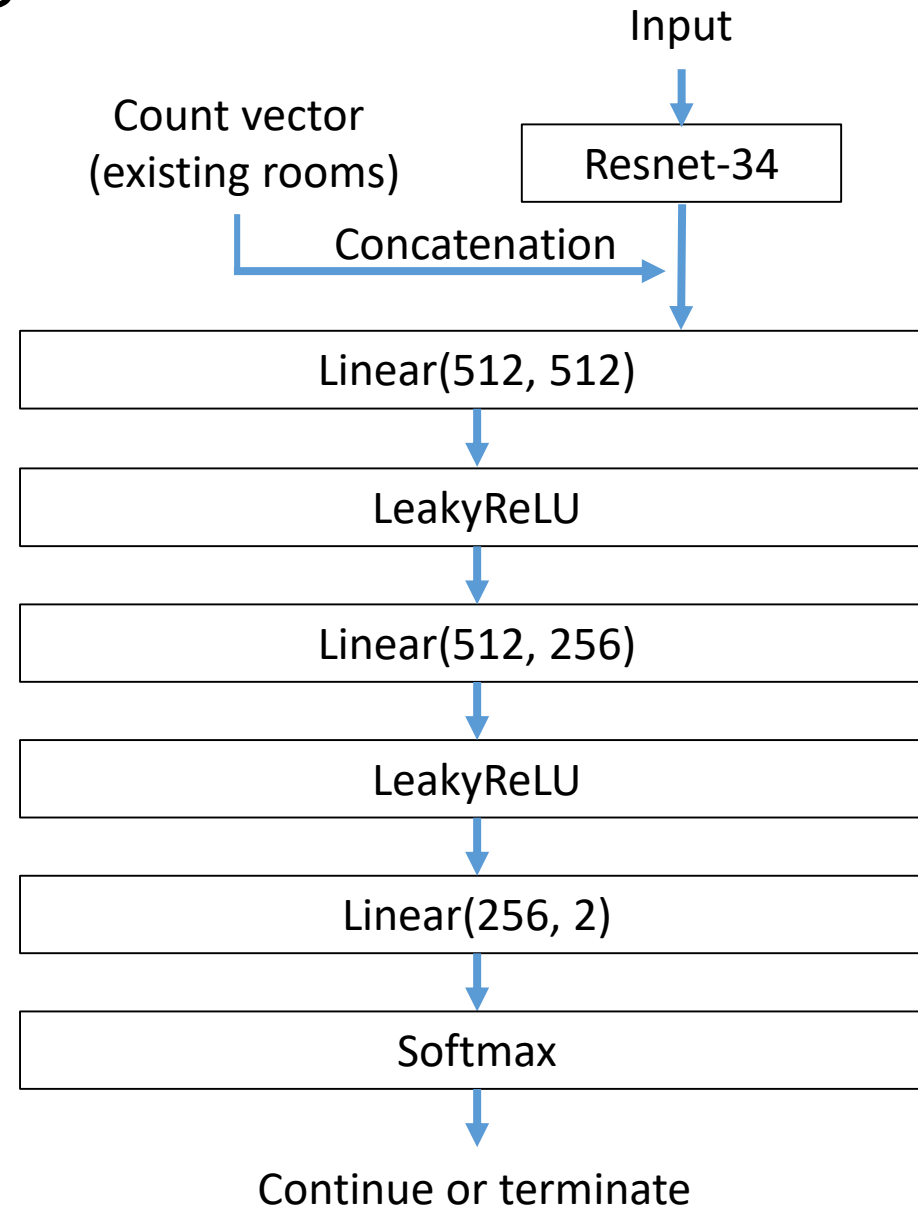
# Outline

- Networks and implementation details
- Floor plans generated by our method
- Human-created floor plans for Figs. 11, 12, 13 in the paper
- Questionnaire for comparison to human-created floor plans

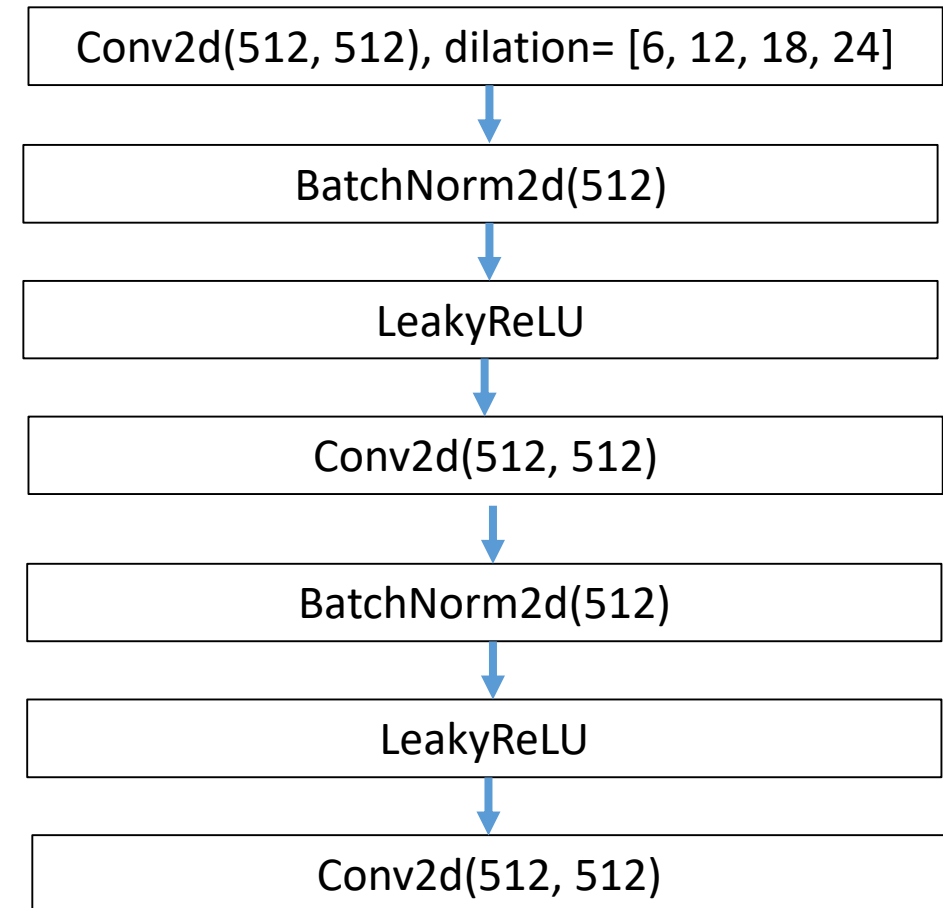
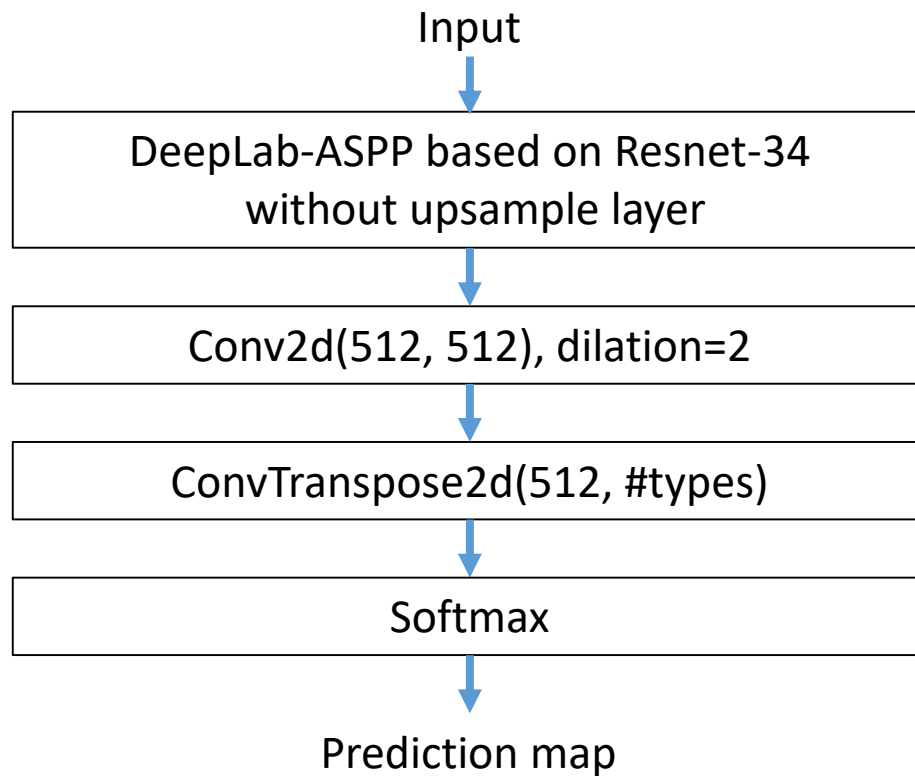
# Our living room regression network



# Our continuing network



# Network architecture of our room locating network and wall locating network



The main architecture is shown on the left and the modified ASPP block of DeepLab-ASPP is shown on the right.

# Hyper parameters

Hyper parameters	Value
Batch size	16
Epoch (regression)	300
Epoch (continuing)	300
Epoch (room locating)	100
Epoch (wall locating)	100
Optimizer	Adam
Learning rate*	1e-4
Weight decay	1e-4

\*For the regression network and continuing network, we adopt a “step” learning rate policy: the learning rate is multiplied by 0.5 every 30 epoch. For the room locating network and wall locating network, we adopt a “poly” learning rate policy: the learning rate is multiplied by  $(1 - \frac{current\_epoch}{max\_epoch})^{power}$  with power = 1.5.

# Loss function

- The regression network

$$L_{regression} = \frac{1}{N} \sum_i^N (Smooth_{L1}(w_i - w_i^*) + Smooth_{L1}(h_i - h_i^*))$$

Where N is the number of the training data in one min batch.  $(w_i, h_i)$  and  $(w_i^*, h_i^*)$  are the predicted and target locations of the living room, respectively. Here

$$Smooth_{L1}(x - x^*) = \begin{cases} 0.5(x - x^*)^2, & \text{if } |x - x^*| < 1 \\ |x - x^*| - 0.5, & \text{otherwise} \end{cases}$$



# Loss function

- The continuing network

$$L_{continuing} = \frac{1}{N} \sum_i^N \left\{ -\log \left( \frac{e^{x_i^*}}{e^{x_i} + e^{x_i^*}} \right) \right\}$$

Where N is the number of the training data in one min batch.  $x_i^*$  and  $x_i$  are predicated score for continuing and not continuing, respectively.

- The locating network

$$L_{locating} = \frac{1}{N} \sum_i^N \sum \omega_t \left\{ -\log \left( \frac{e^{x[t]}}{\sum_j e^{x[j]}} \right) \right\}$$

Where N is the number of the training data in one min batch. The second  $\Sigma$  means the sum on all pixels.  $\omega_t$  is the weight for each category.  $t$  is the index of the target category for each pixel.  $x$  is the predicted score vector computed on each pixel.

# Implementation of ISSNet [Wang et al. 2018]

We adopt ISSNet as the baseline room locating model, which predicates room locations one by one. In each iterative step, it first analyzes the input scene to determine whether to add another room by the continuing network and then uses the locating network to decide both which category of room to add and also where to add it.

The continuing network has no difference with the one used in our proposed method. The locating network computes  $p_{cat}(c|S, x, y)$ , the probability that  $c$  is the category which should be added to scene  $S$  at location  $(x, y)$ .

$$f_{cat}(c|S, x, y) = \text{MLP}([\text{couts}(S), \text{CNN}[S, M_{attn}(x, y)]])$$

$$p_{cat}(c|S, x, y) = \frac{\exp(f_{cat}(c|S, x, y))}{\sum_{c'} \exp(f_{cat}(c'|S, x, y))}$$

Here MLP is a multilayer feed-forward neural network,  $\text{couts}(S)$  is a vector of counts of each category of room already present in the scene.  $M_{attn}(x, y)$  is an image containing a small 18x18 mask centered about  $(x, y)$ . CNN extracts high-level features from  $S$  and  $M_{attn}(x, y)$ .

In the sampling, the joint distribution can be constructed as

$$p_{cat}(c, x, y|S) = p_{cat}(c|S, x, y) \cdot \frac{1}{NN}$$

Then we suppress noise in the above distribution to obtain the final distribution  $p^*$  using the two-step tempering scheme mentioned in [Wang et al. 2018]. At synthesis time, we use a 16x16 grid to sample a location  $(x, y)$  and a room category  $c$  and then choose the location which maximizes  $p^*(x, y|c)$  within the previously-sampled grid cell.

# Implementation of MIQP [Wu et al. 2018]

We adopt MIQP as the baseline wall locating method. Given a set of room locations predicted by ISSNet or our method, the room locations serve as the position constraints for MIQP to generate walls. A polygonal room can be represented as the union of a set of room rectangles with the same label. A room rectangle is described by a tuple  $(x_i, y_i, w_i, d_i, l_i)$  where  $(x_i, y_i)$  denotes the position of the bottom-left corner of the rectangle,  $(w_i, d_i)$  denotes the width and depth, and  $l_i$  denotes its label. Given the boundary as input, the goal of the floor plan design is to arrange a set of room rectangles inside of the domain according to some constraints.

We first introduce two types of basic constraints. The first one is the inside constraint which ensures that all rooms are inside of the layout domain. The second one is the non-overlap constraint. We require that there is no overlap between any pairs of rooms. Then the room locations are formulated into the position constraints: for each room location  $(x^*, y^*)$ , the position constraint requires the room to cover the position,

$$\begin{cases} x_i \leq x^* \leq x_i + w_i \\ y_i \leq y^* \leq y_i + d_i \end{cases}$$

The objective function is defined as the combination of the coverage term and the size error,

$$E = \lambda_{cover} E_{cover} + \lambda_{size} E_{size},$$
$$E_{cover} = Area - \sum_i w_i \cdot d_i, E_{size} = \sum_i (w_i - w_i^*)^2 + (d_i - d_i^*)^2$$

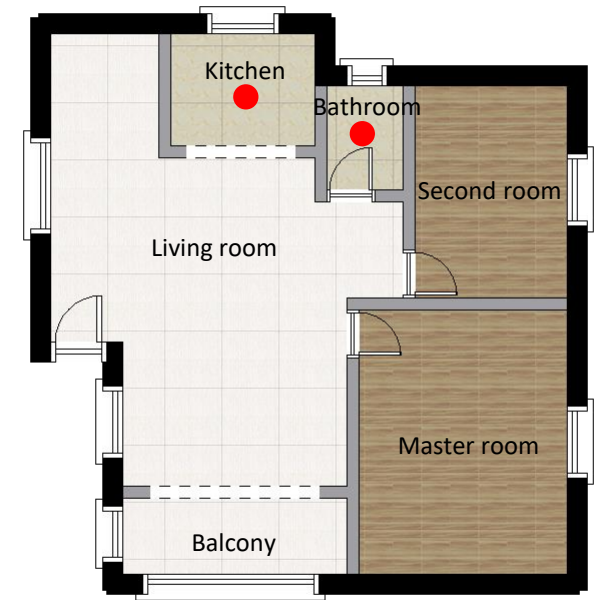
Where  $\lambda_{cover}$  and  $\lambda_{size}$  are weights, Area is the total area of the floor plan and  $(w_i^*, d_i^*)$  is the target size of room  $i$ .

We then propose a hierarchical framework to improve the results. We first generate an initial floor plan using the above algorithm. There may exist some region that is not covered by any rooms, and we select the sub-region for further improvement. We initialize the optimization by separating rooms in the sub-domain. The constraints for rectangular rooms in the sub-domain are updated. Then we apply the same method to generate the new floor plan in the sub-domain. We repeat this procedure to get the final result.

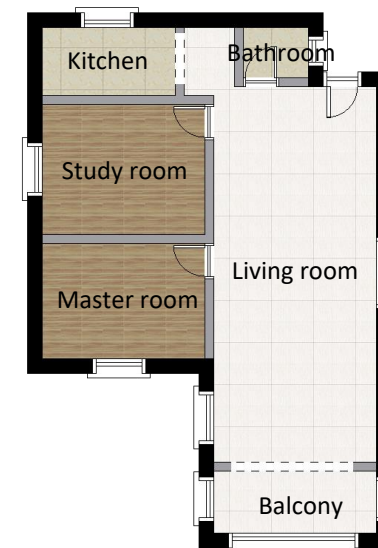
# Outline

- Networks and implementation details
- Floor plans generated by our method
- Human-created floor plans for Figs. 11, 12, 13 in the paper
- Questionnaire for comparison to human-created floor plans

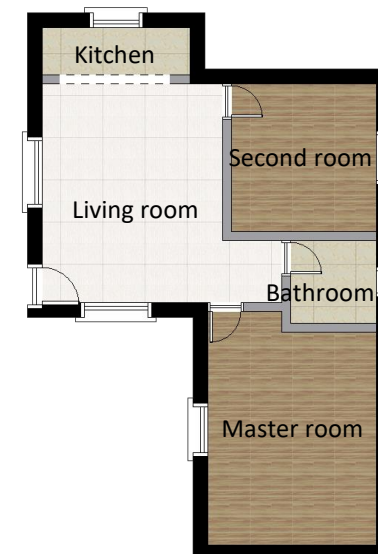
# Room constraints



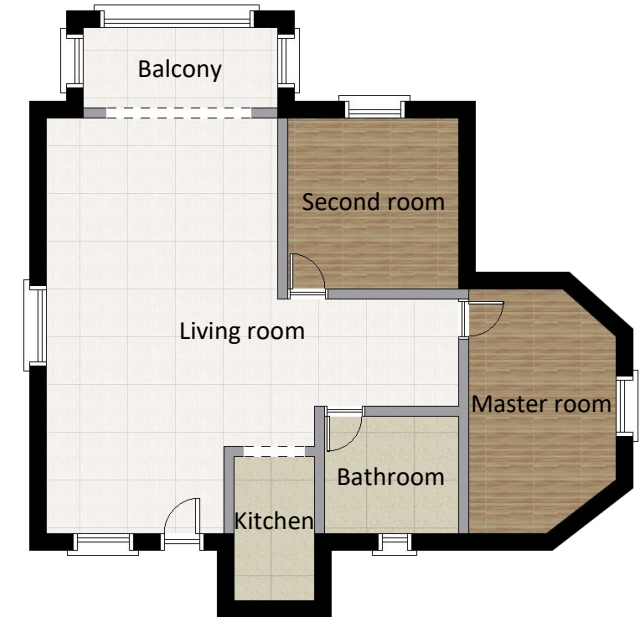
## The nearest neighbors



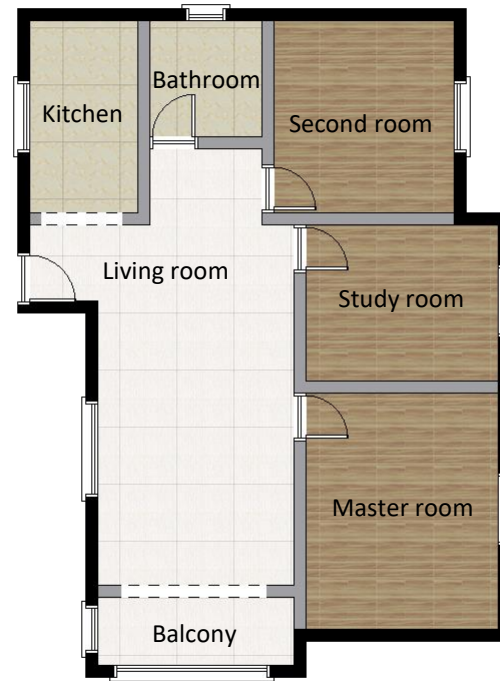
## Our synthesized results



# Non axis-aligned walls

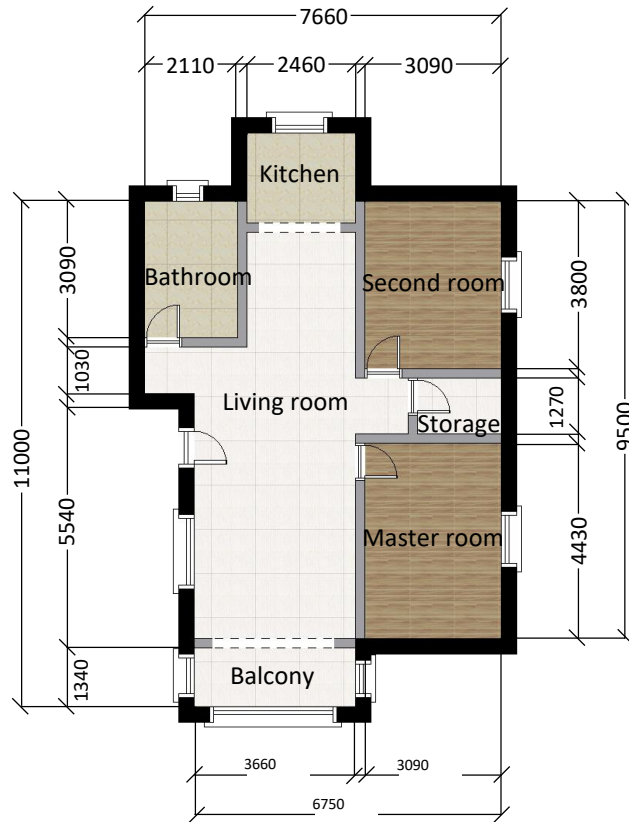


# Multiple floor plans

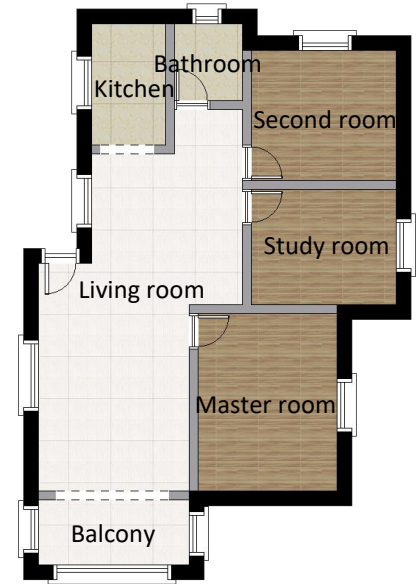


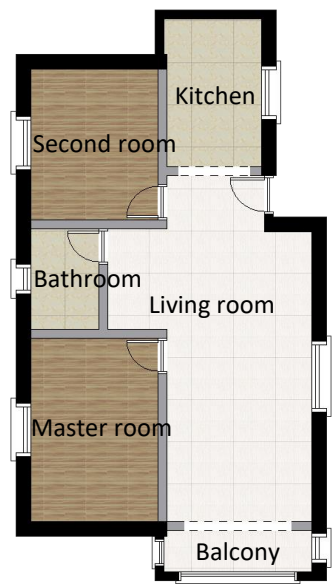


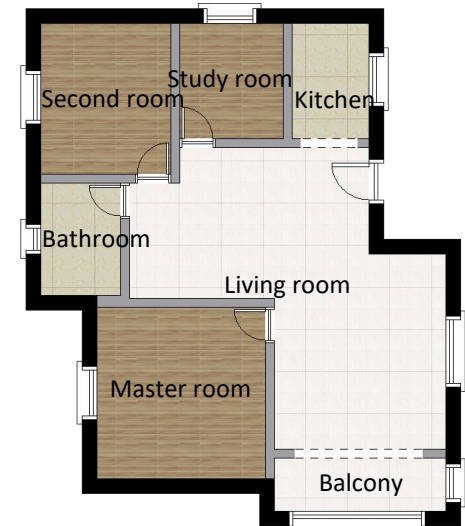
# Room dimensions



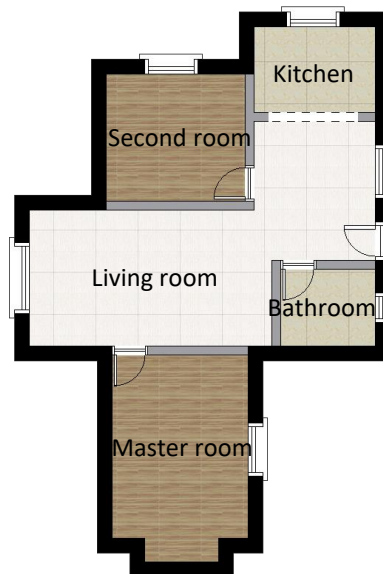
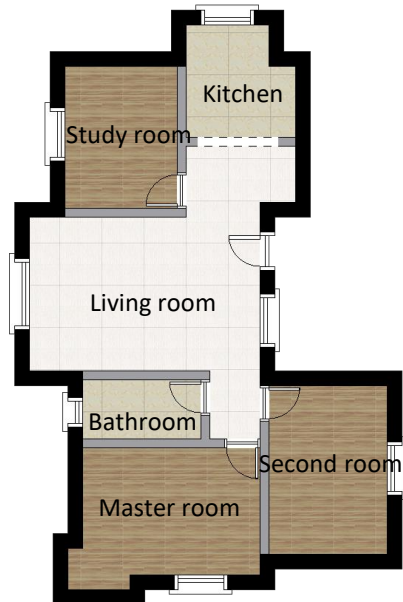
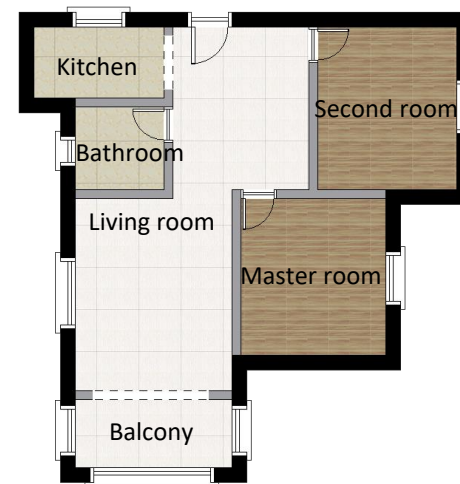
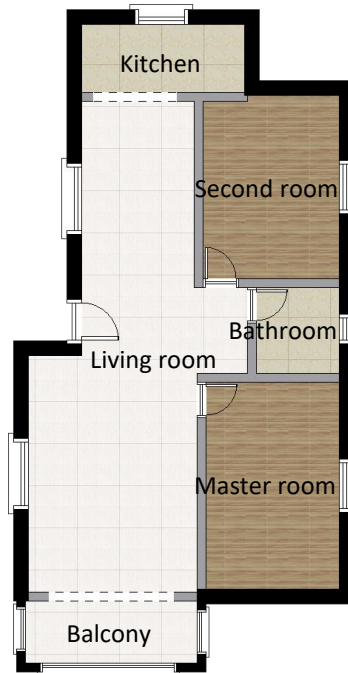
# More results







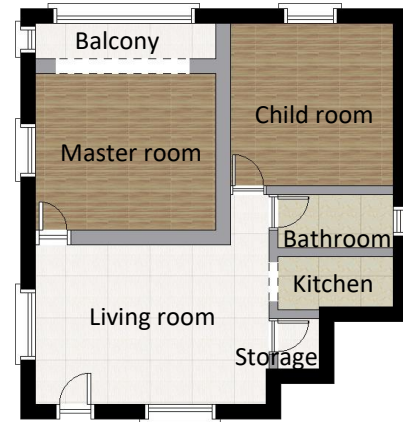
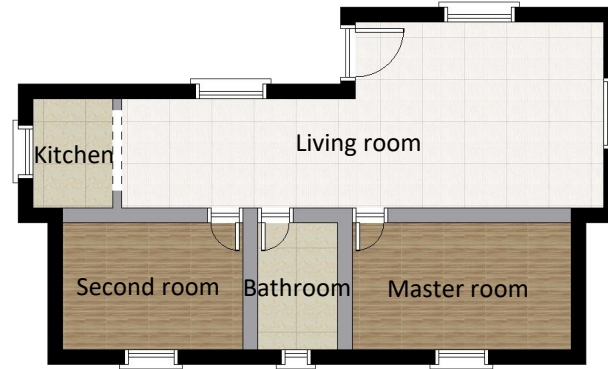




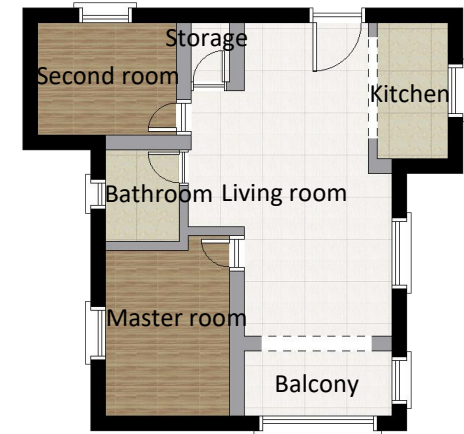
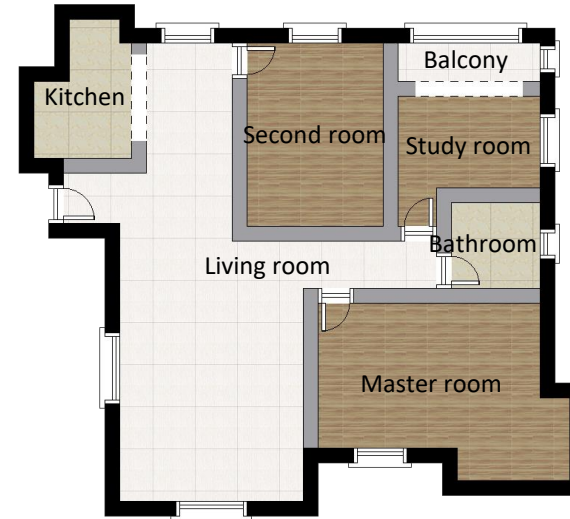
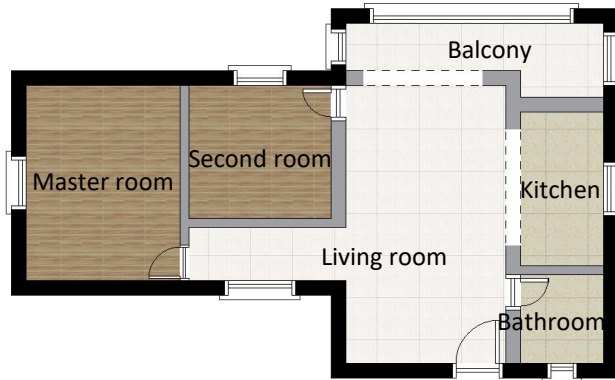
# Outline

- Networks and implementation details
- Floor plans generated by our method
- Human-created floor plans for Figs. 11, 12, 13 in the paper
- Questionnaire for comparison to human-created floor plans

# Human-created floor plans for the comparisons to ISSNet+MIQP (Fig.11 in the paper)



# Human-created floor plans for the comparisons to Stage1+MIQP (Fig.12 in the paper)





# Human-created floor plans for the comparisons to ISSNet+Stage2 (Fig.13 in the paper)



# Outline

- Networks and implementation details
- Floor plans generated by our method
- Human-created floor plans for Figs. 11, 12, 13 in the paper
- Questionnaire for comparison to human-created floor plans

Which one do you think is more plausible in the following pair of floor plans with same boundary?







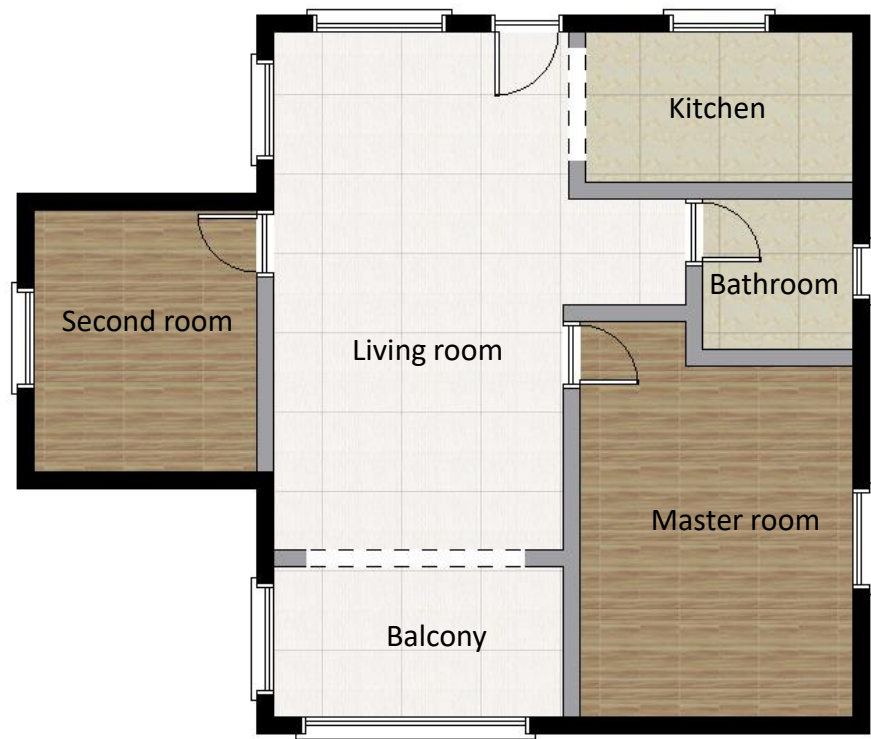


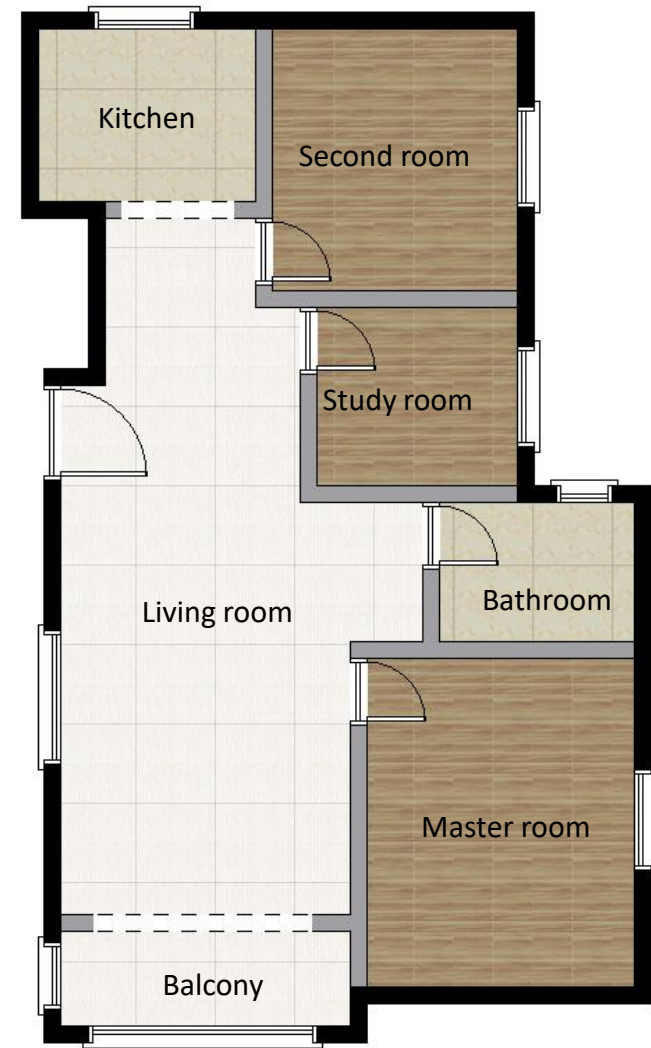




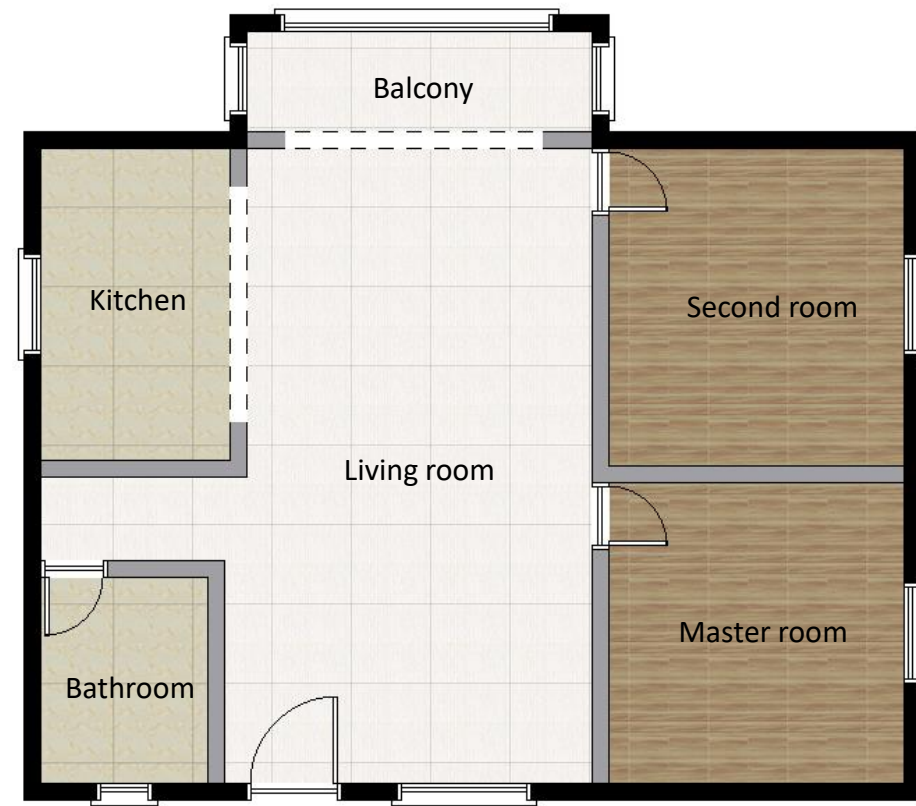
















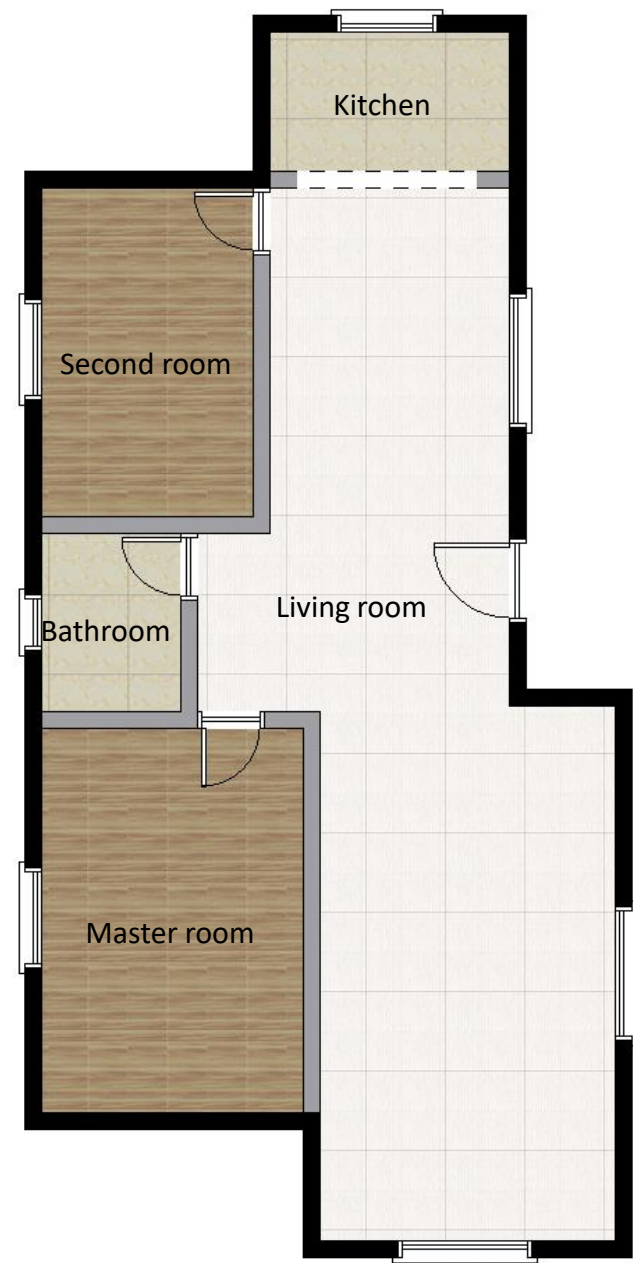
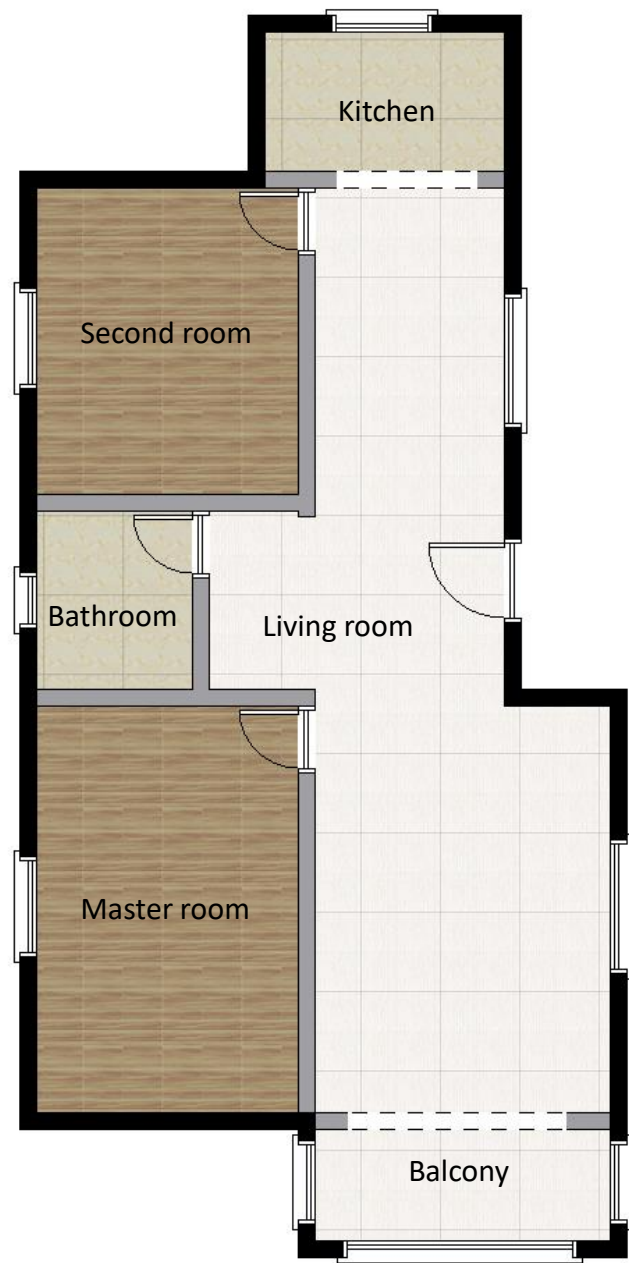


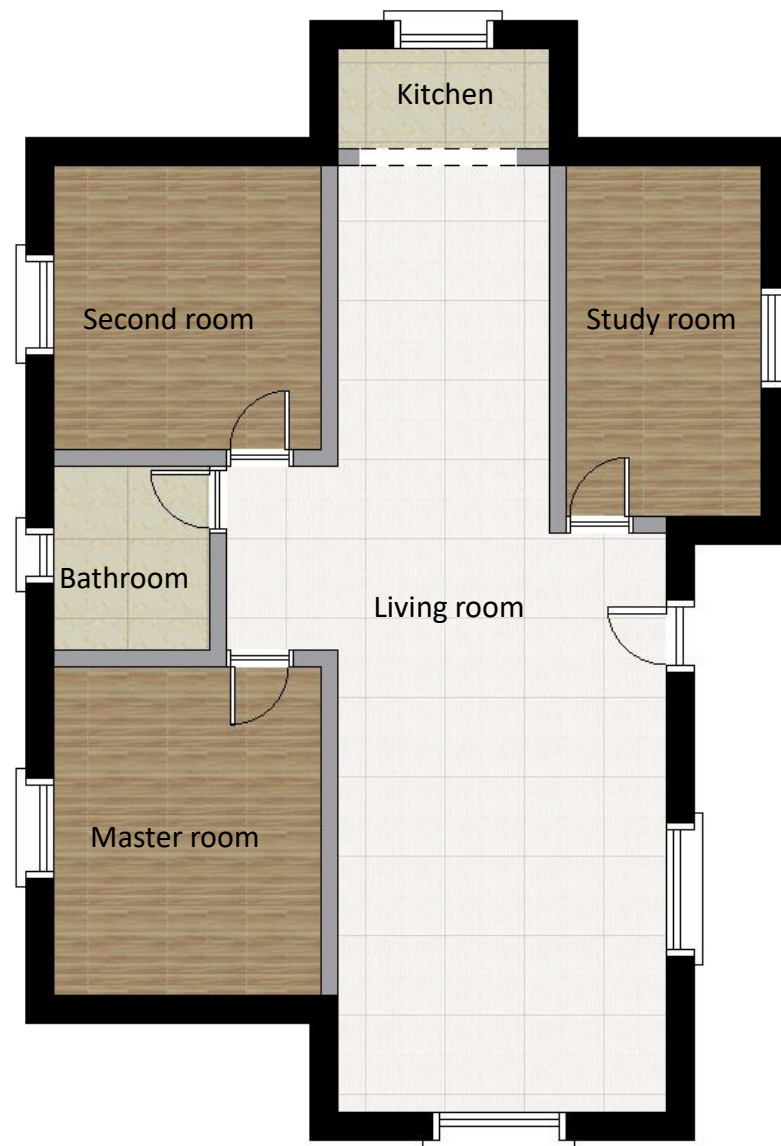








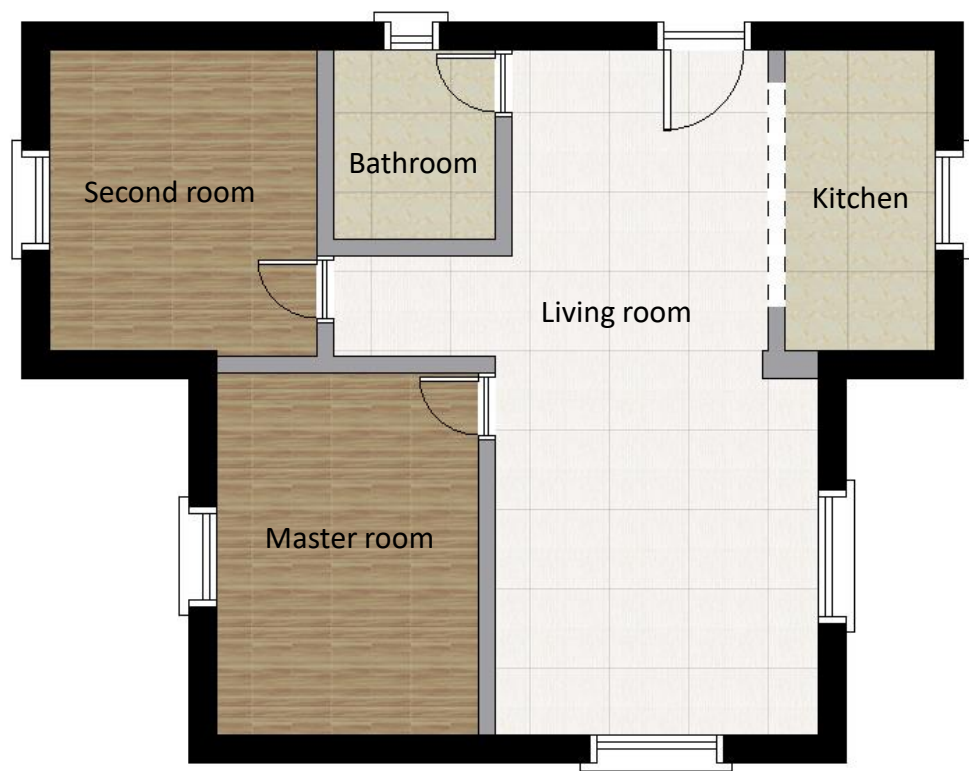














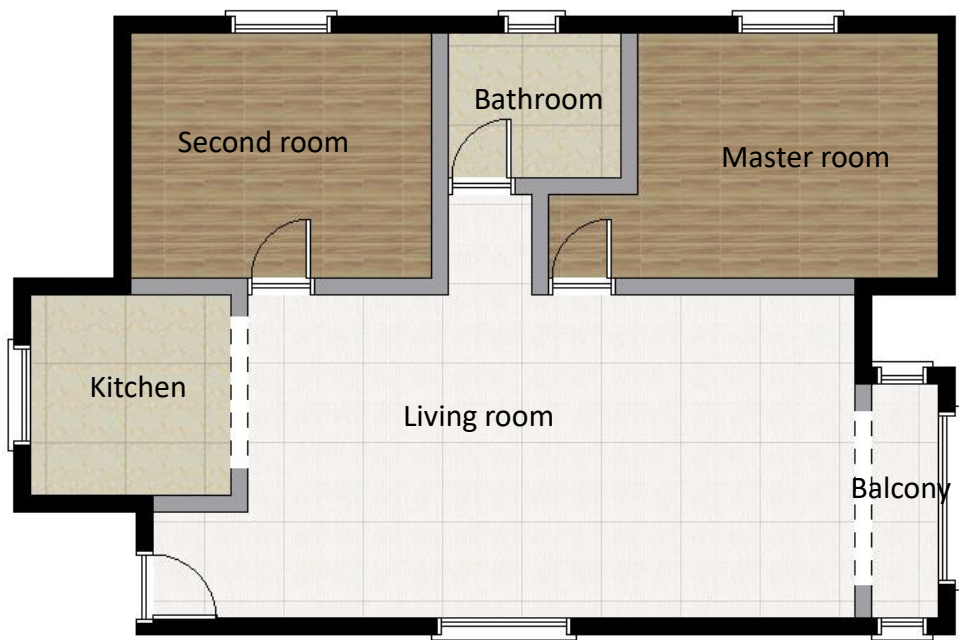


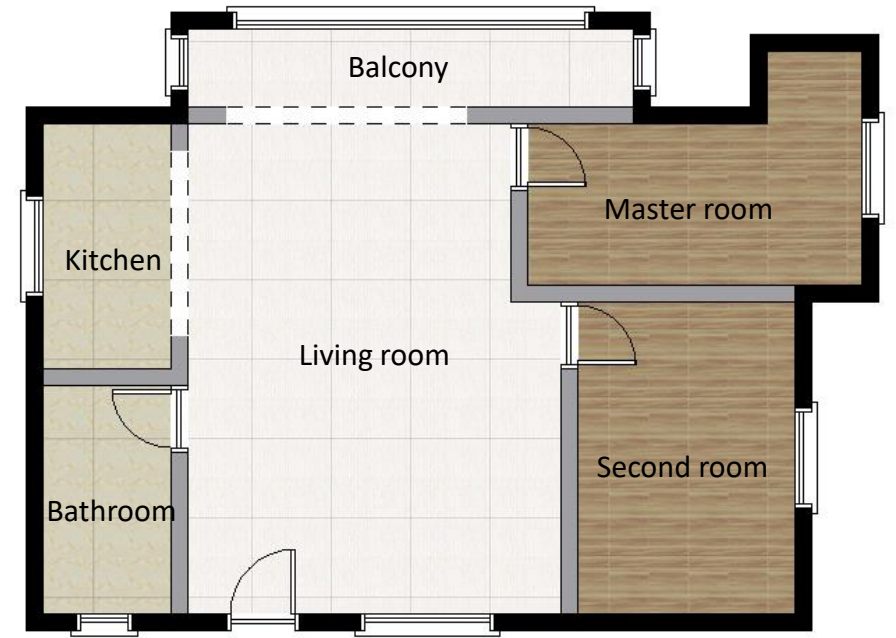










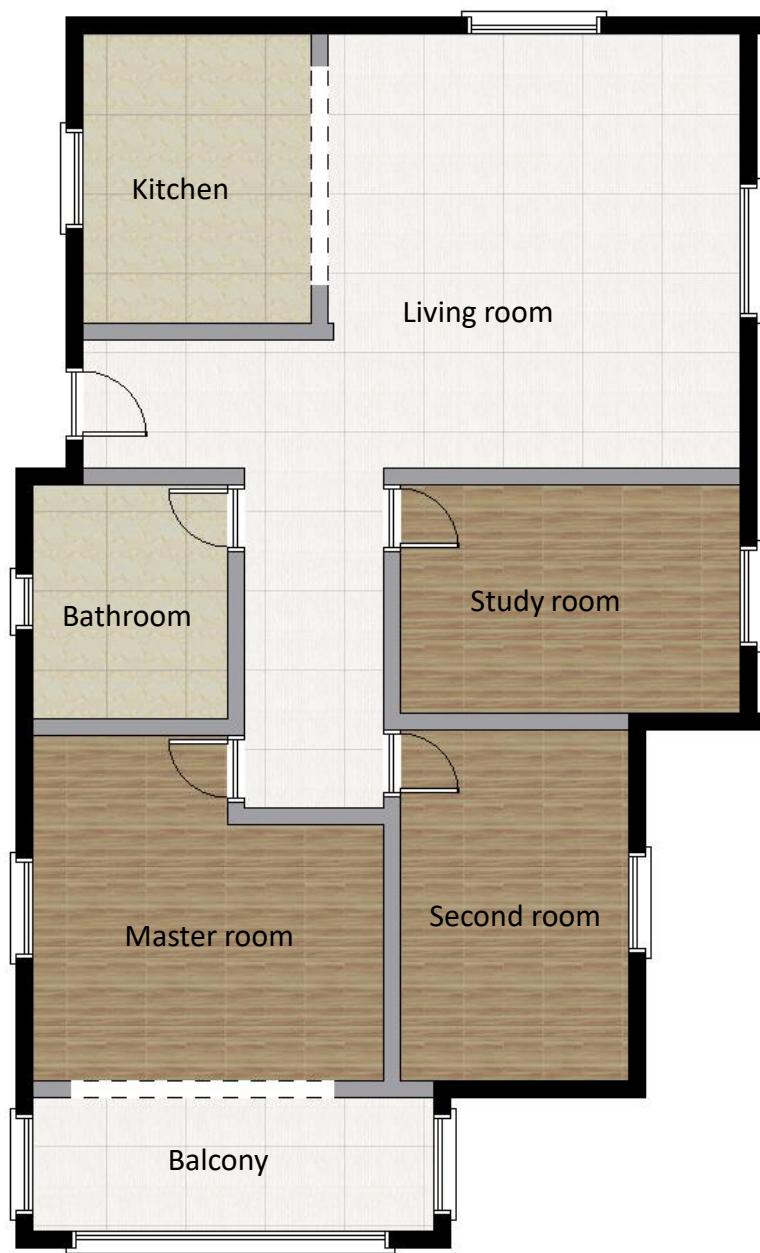












Thank you