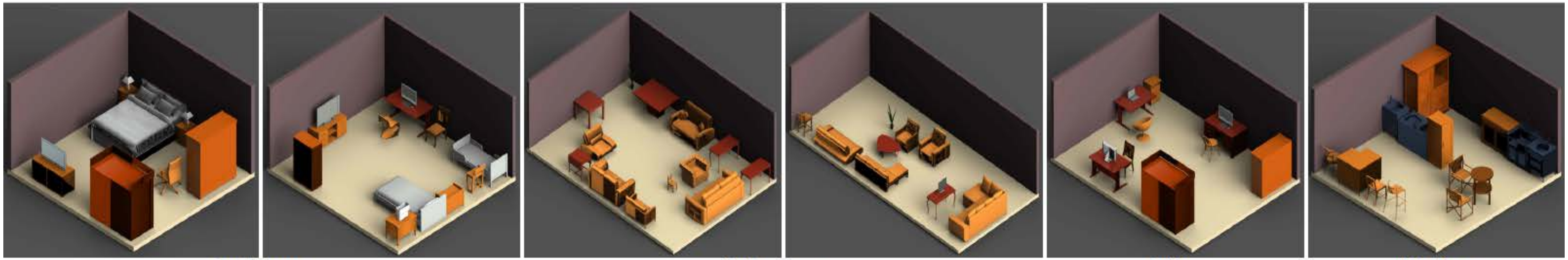
# GRAINS: Generative Recursive Autoencoders for INdoor Scenes

Manyi Li [1,2], Akshay Gadi Patil [2], Kai Xu [3,4], Siddhartha Chaudhuri [5,6], Owais Khan [6],
Ariel Shamir [7], Changhe Tu [1], Baoquan Chen [8], Daniel Cohen-Or [9], Hao (Richard) Zhang [2]

[1] Shandong University [2] Simon Fraser University [3] National University of Defense Technology [4] AICFVE Beijing Film Academy
[5] Adobe Research [6] IIT Bombay [7] The Interdisciplinary Center [8] Peking University [9] Tel-Aviv University



Bedrooms Living rooms Office Kitchen

# Outline

- Problem & Related work
- Method
  - Scene representation
  - Network
- Ablation study
- Results & Application

# Scene generation problem

- Generate plausible room layouts automatically, to replace or reduce human work.


Planner5d

# Related works: data-driven

- Graphical model methods

[Fisher et al. SIGA 2012], [Kermani et al. SGP 2016], [Qi et al. CVPR 2018]

☺ Respect object-object relations

☹ Too many parameters and rules to tune manually

☹ Time-consuming


Object appearance + Object positioning

Learning

Generation

# Related works: data-driven

- Graphical model methods
- Deep neural networks

[Wang et al. SIGGRAPH 2018], [Ritchie et al. CVPR 2019]

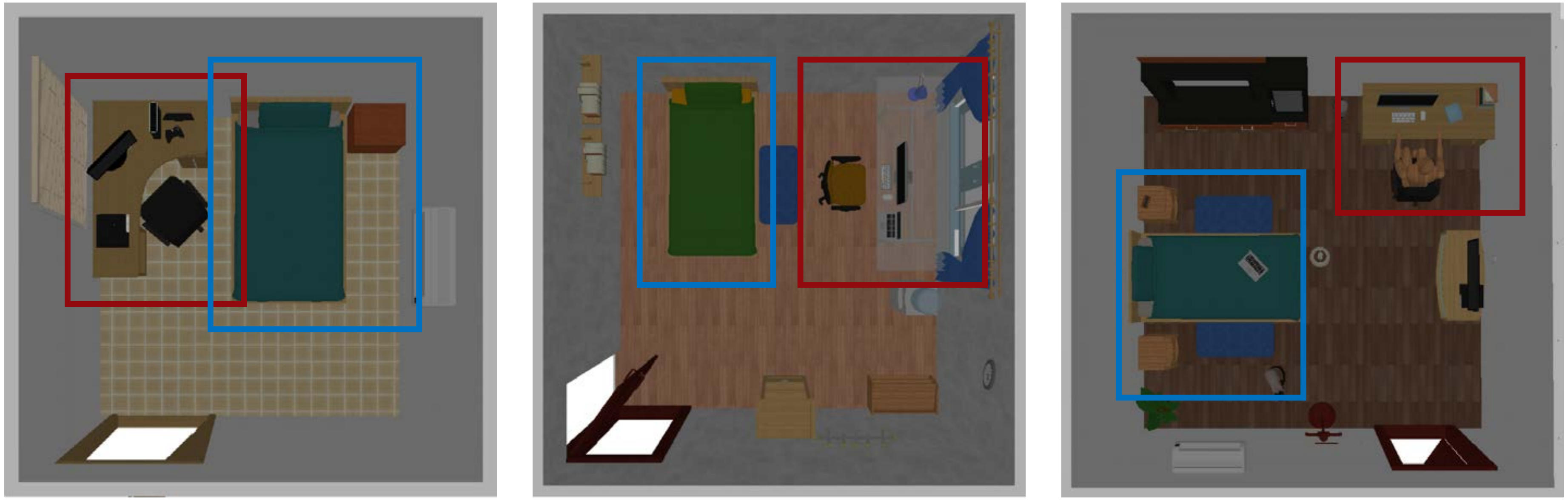
☺ Easy-to-use model and better plausibility

☹ No object-object relations

Indoor scene    Multi-channel image    CNN training

# Our method

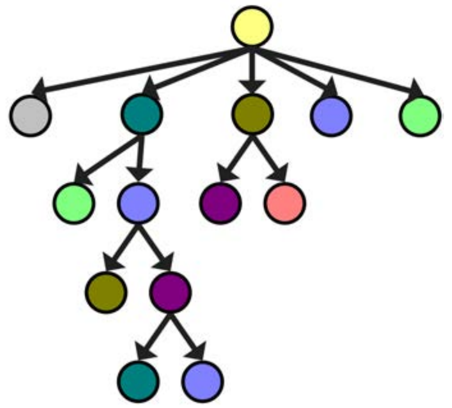
- Indoor scene structures are inherently hierarchical.



Indoor scenes share some common patterns in the sub-scenes.

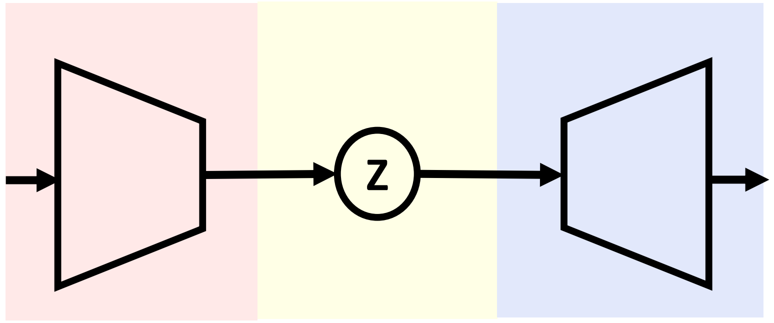
# Our method

- Indoor scene structures are inherently hierarchical.



Hierarchical scene representation

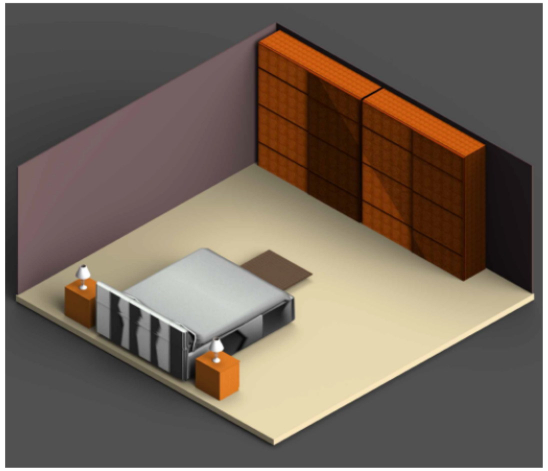
+

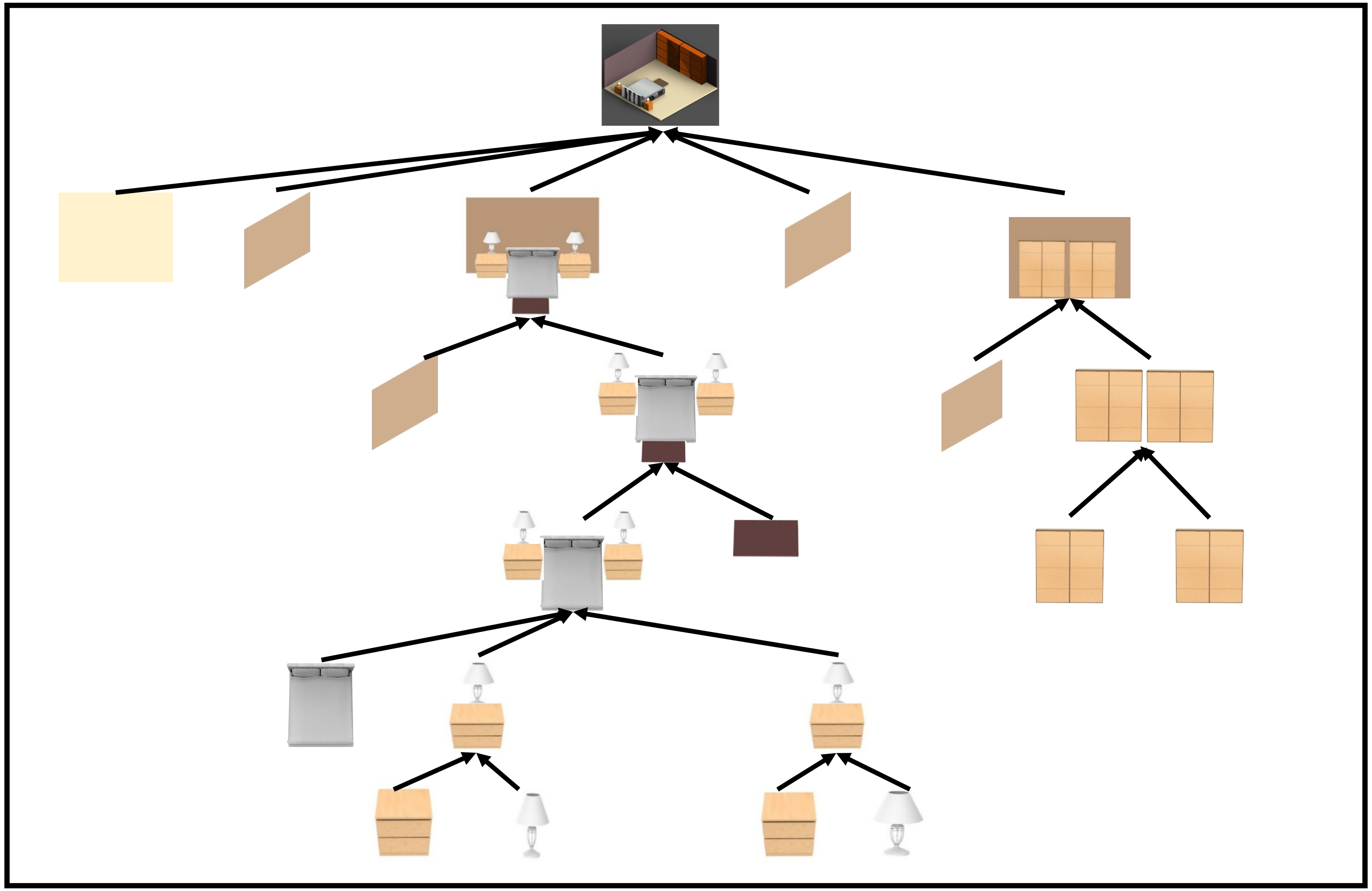



Recursive neural network - Variational Auto-encoder

# Scene Representation

Step1: Deciding the merge order



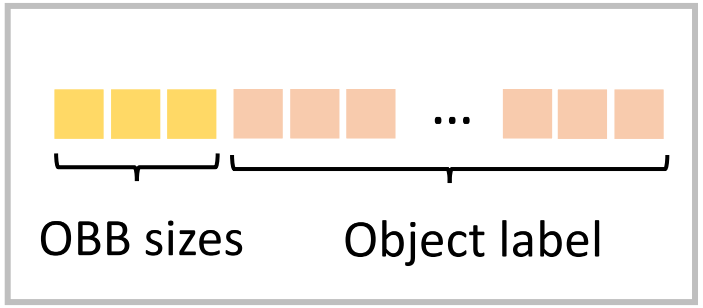


# Scene Representation

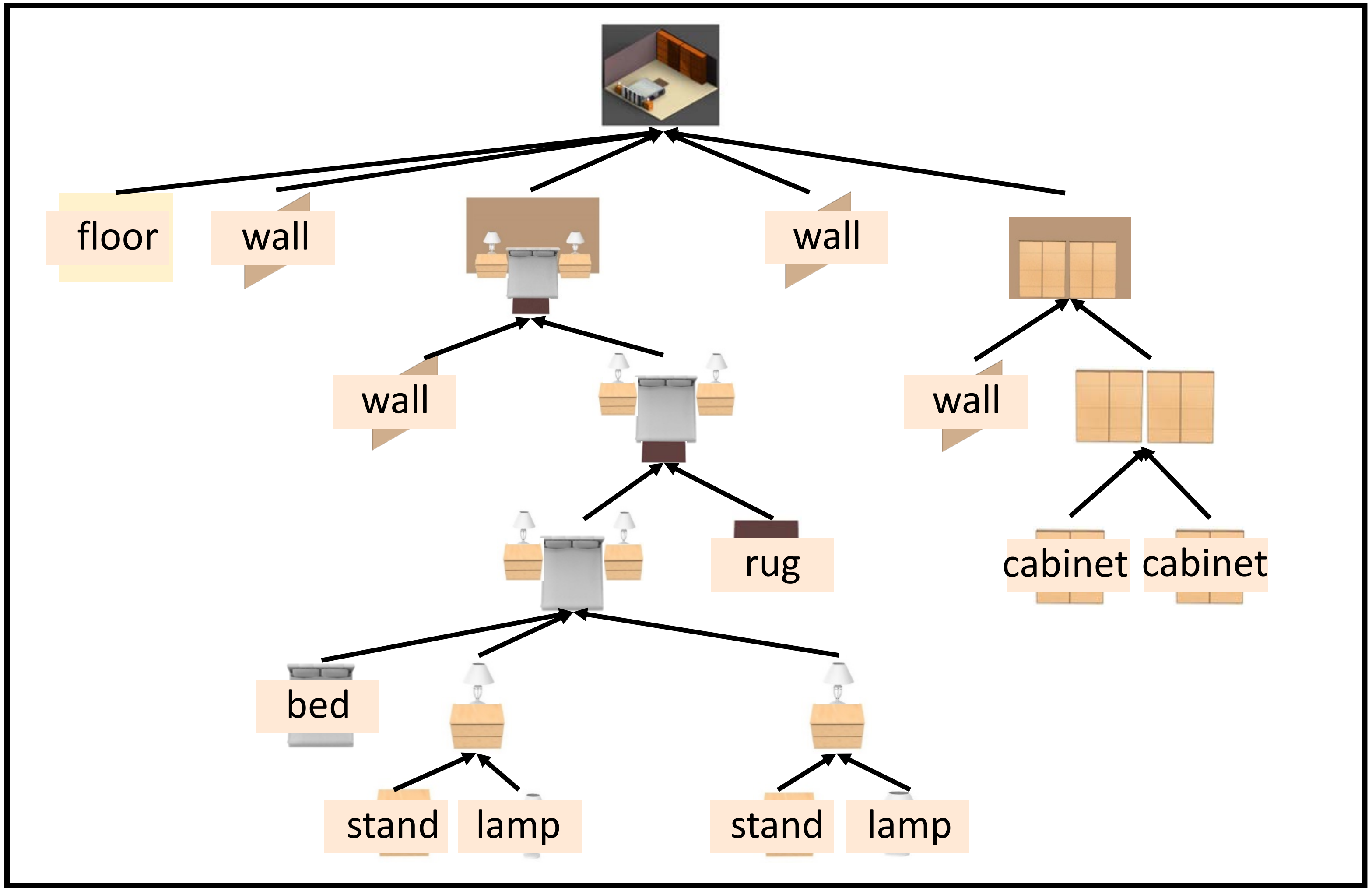
Step1: Deciding the merge order

Step2: Construct the nodes of the hierarchy

- Leaf nodes: objects




Leaf vector

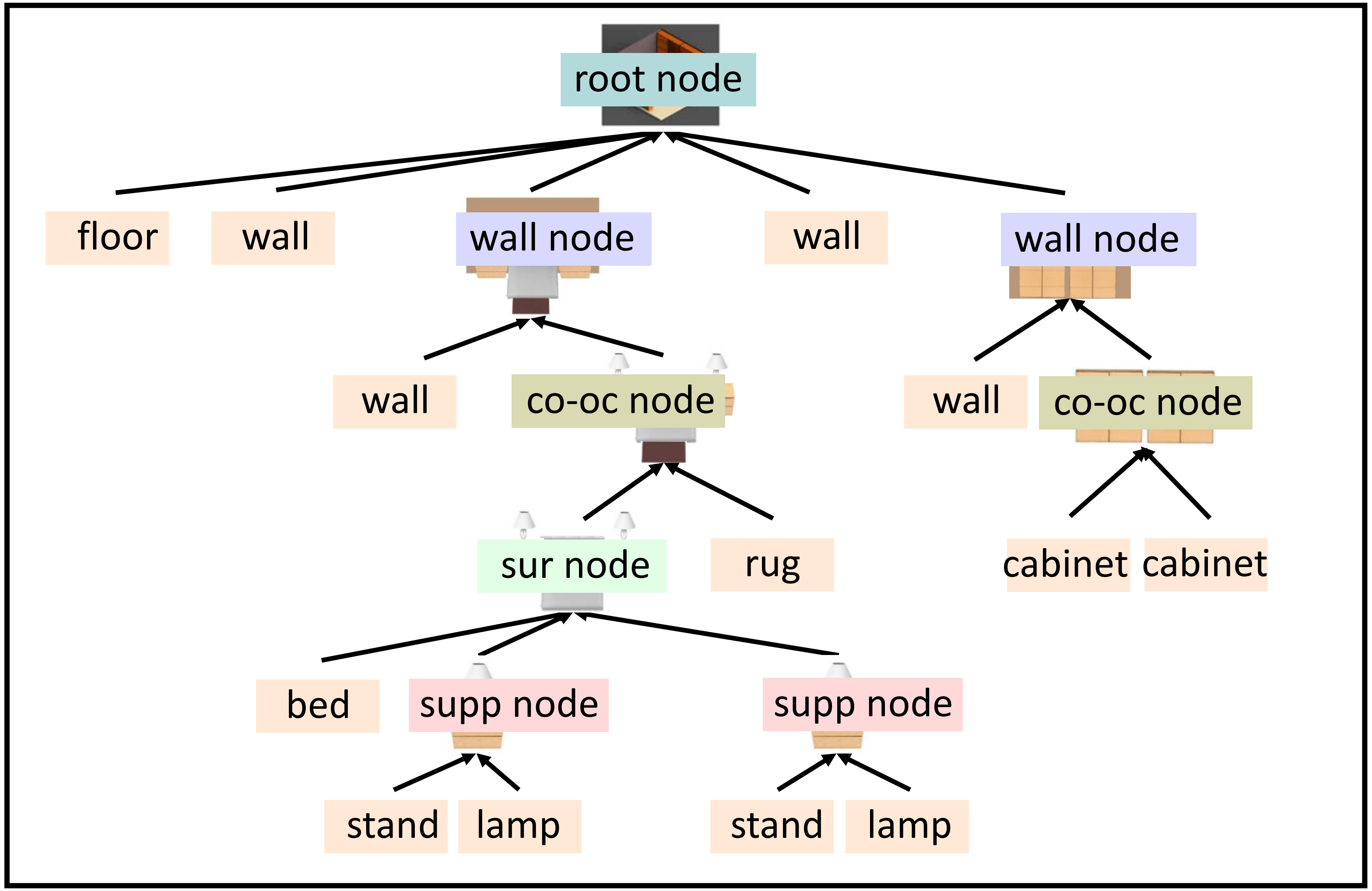


# Scene Representation

Step1: Deciding the merge order

Step2: Construct the nodes of the hierarchy

- Leaf nodes: objects
- Internal nodes: groups
  - Support node
  - Surround node
  - Co-occur node
  - Wall node
  - Root node



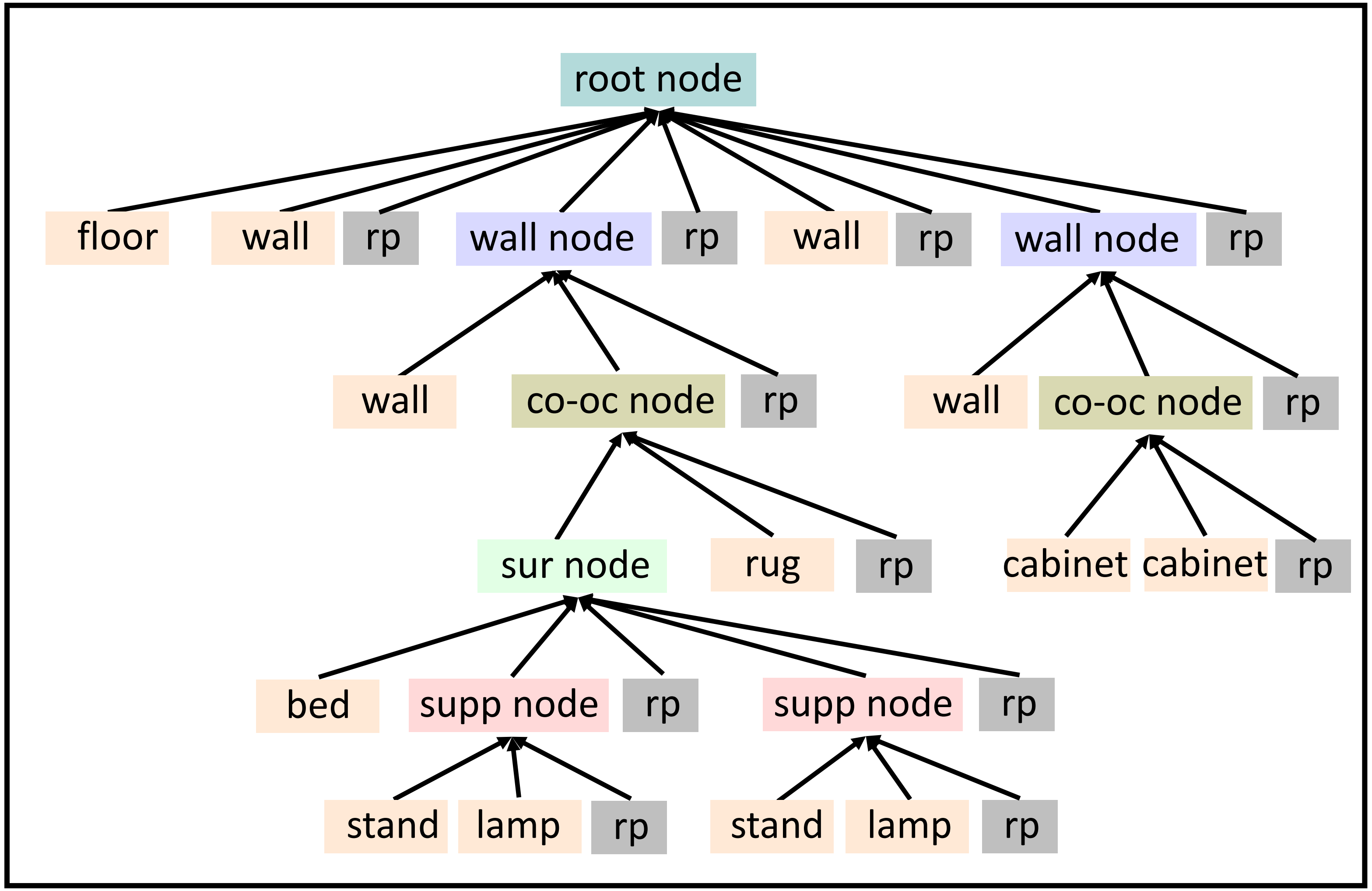
# Scene Representation

Step1: Deciding the merge order

Step2: Construct the nodes of the hierarchy

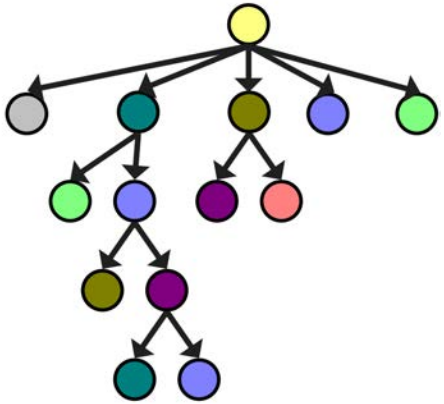
- Leaf nodes: objects
- Internal nodes: groups

Step3: Compute relative positions between sibling nodes



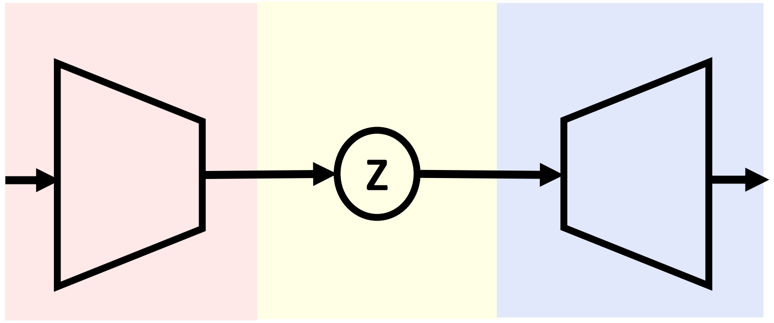
# Our method

- Indoor scene structures are inherently hierarchical.



Hierarchical scene representation

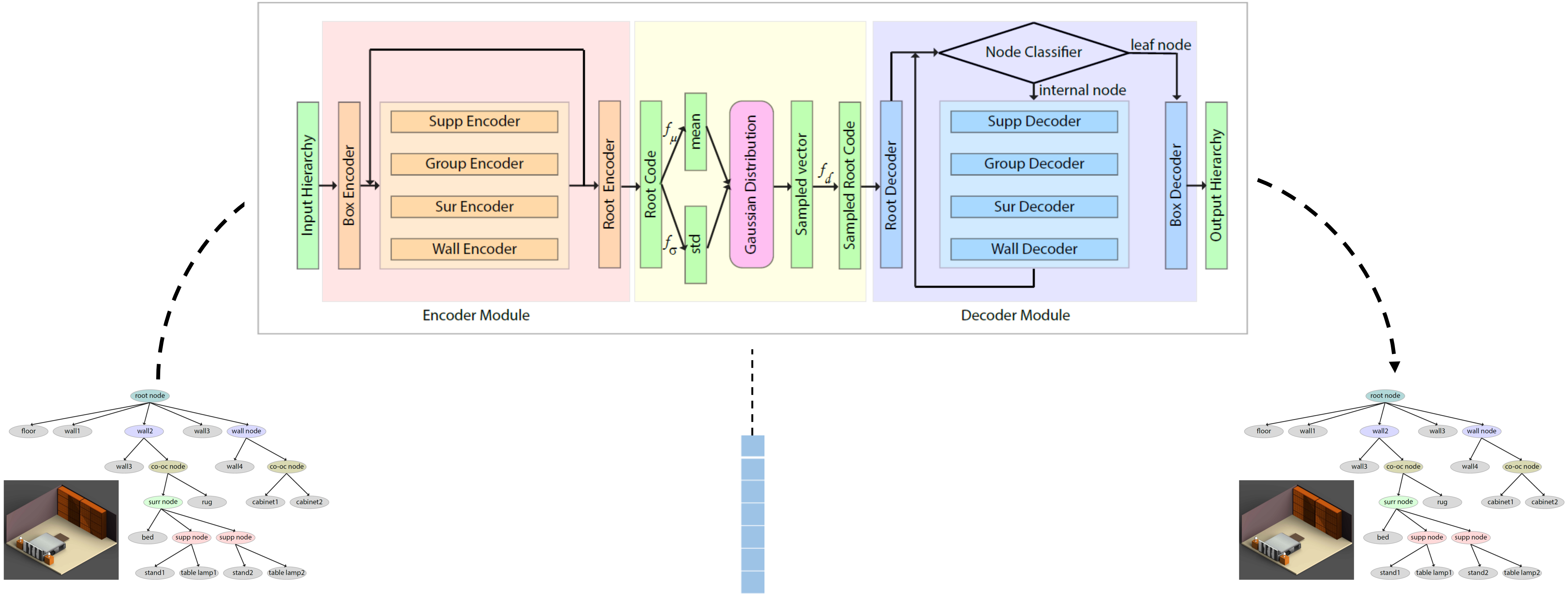
+



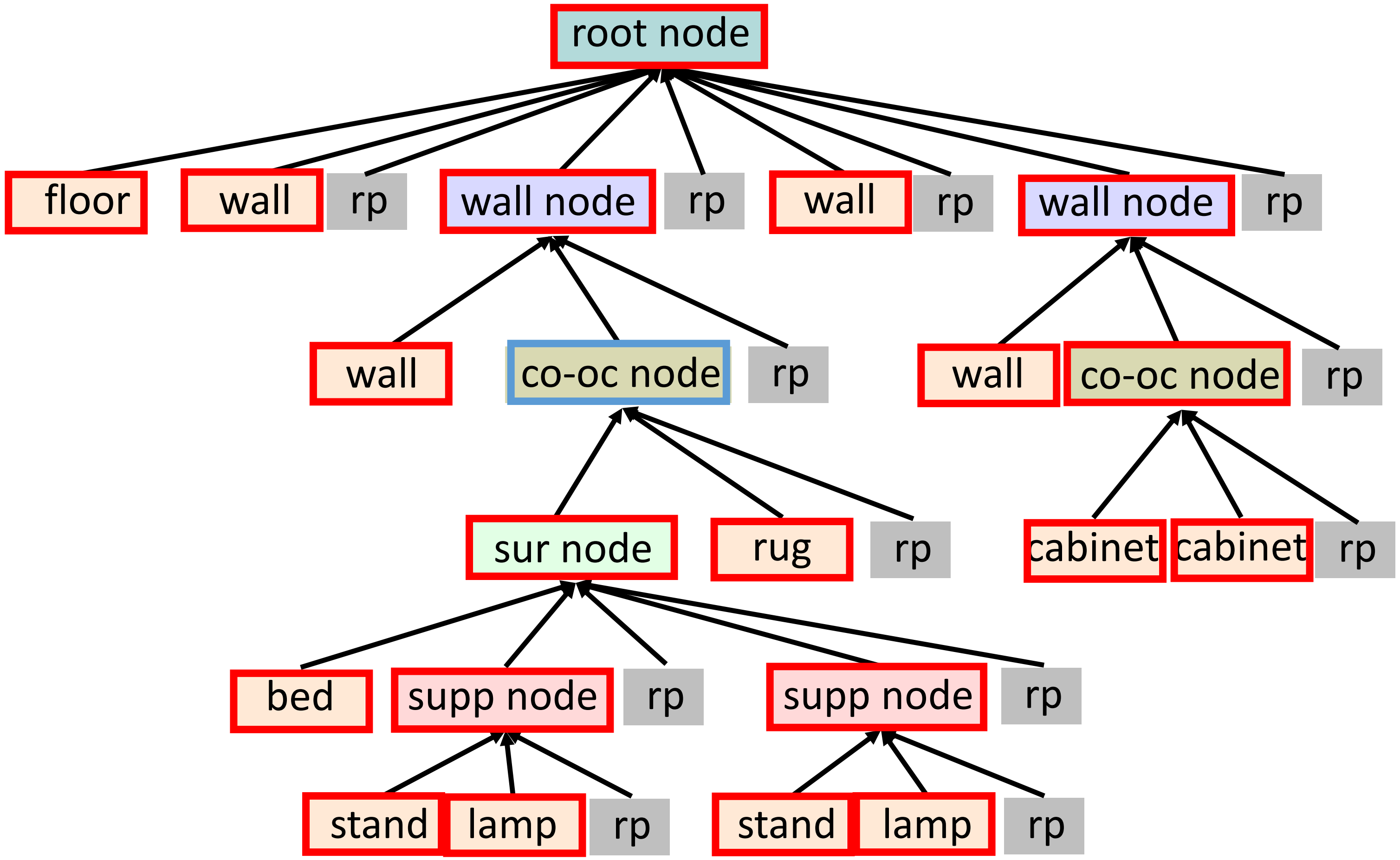

Recursive neural network - Variational Auto-encoder

# Network

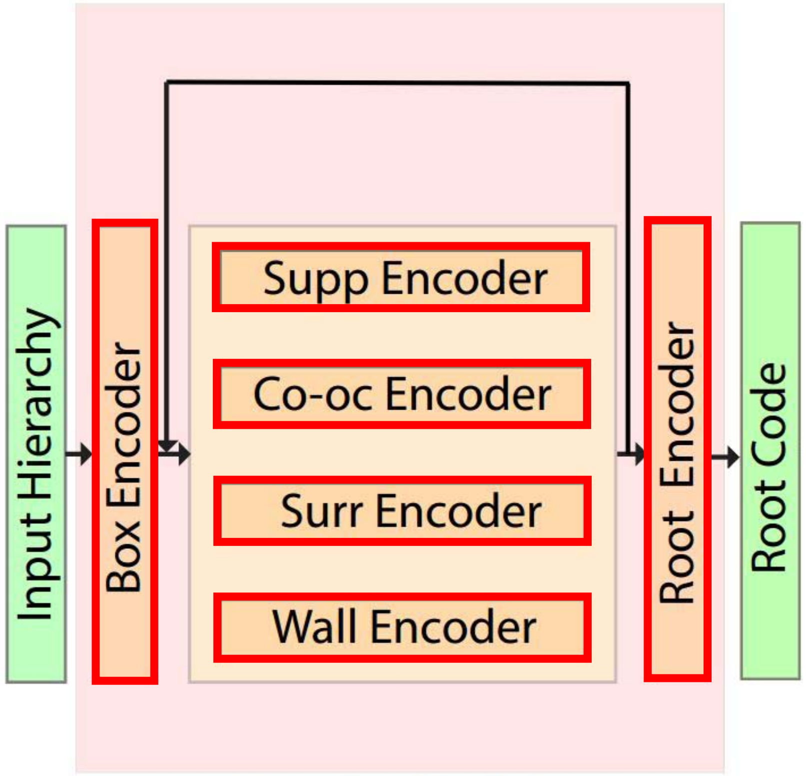
- Network: RvNN-VAE (Recursive Neural Network – Variational Auto-Encoder)



# Encoding Process

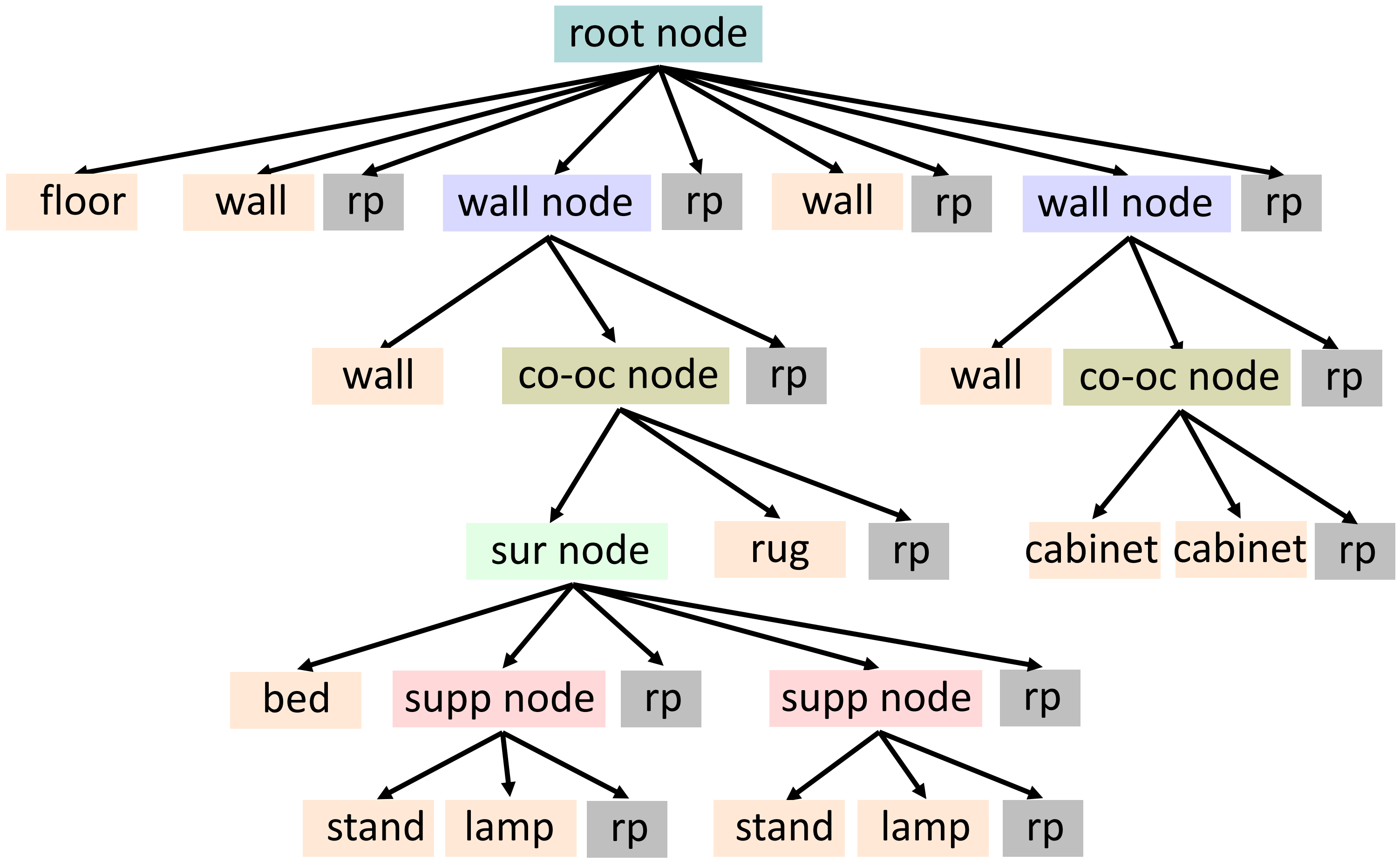



Input hierarchy

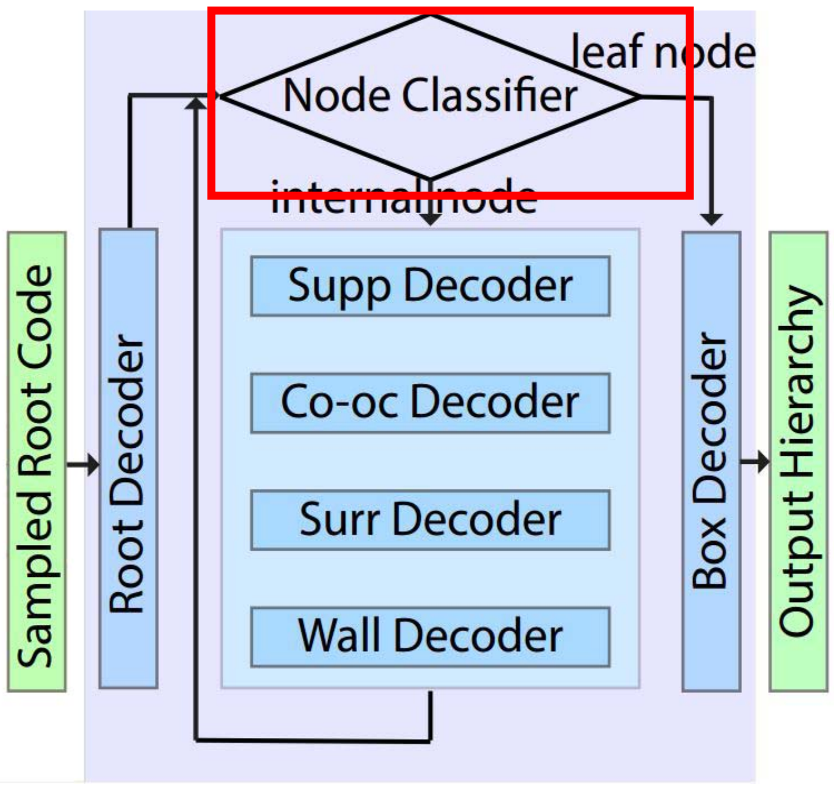



Encoder module

# Decoding process



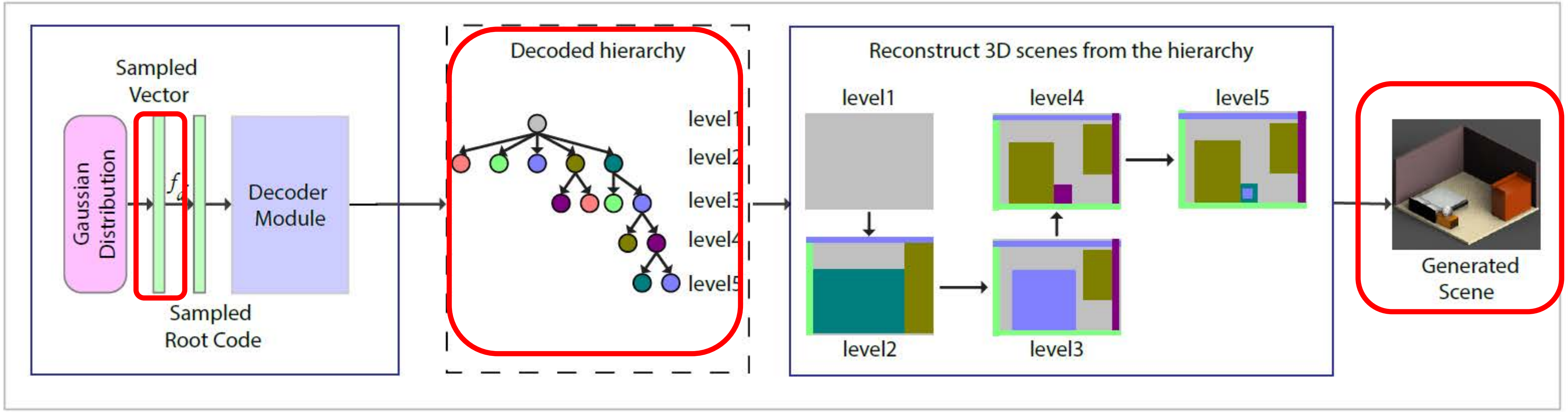

Output hierarchy




Decoder module

# Generation Pipeline

- The network learns to map a random vector to a plausible indoor scene.

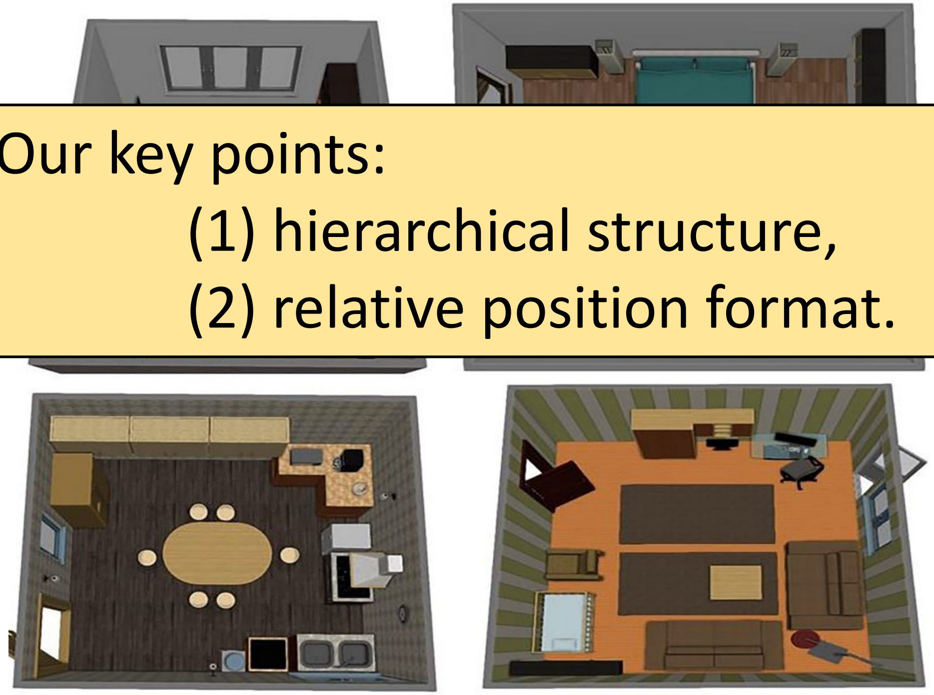



Generation pipeline

# Scene representation matters!

Indoor scenes are complex and diverse

Appropriate scene representation is the key to learning



Our key points:

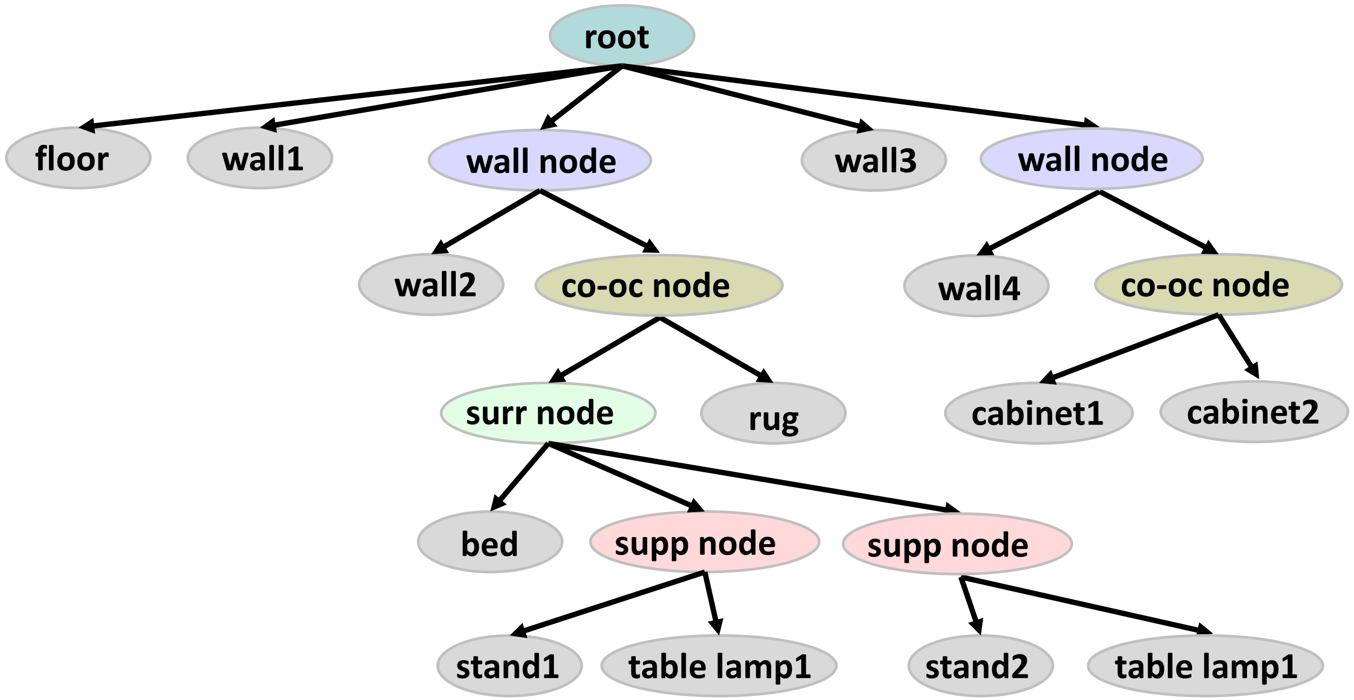
(1) hierarchical structure,

(2) relative position format.

SUNCG dataset [Song et al. 2017]

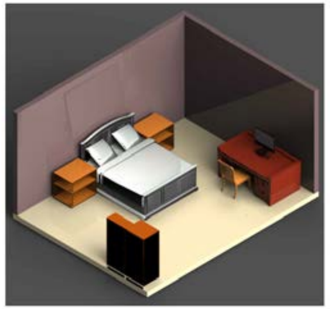
# Key point 1: Hierarchical structure

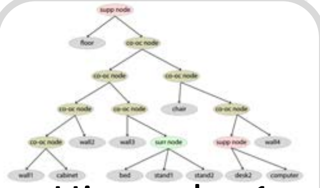
- We specifically define wall node and root node in our hierarchies.
- Reason: walls should serve the role of “grounding” the placement of objects in a room.



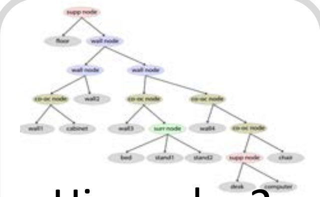

Our hierarchical structure

# Ablation Studies: Hierarchical structure

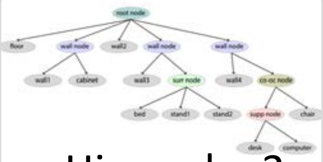




Hierarchy 1

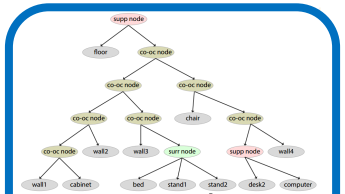


Hierarchy 2

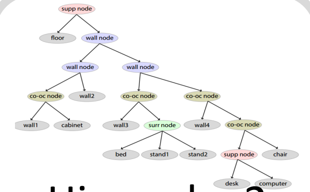


Hierarchy 3

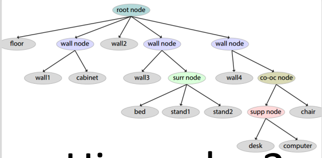
# Ablation Studies: Hierarchical structure



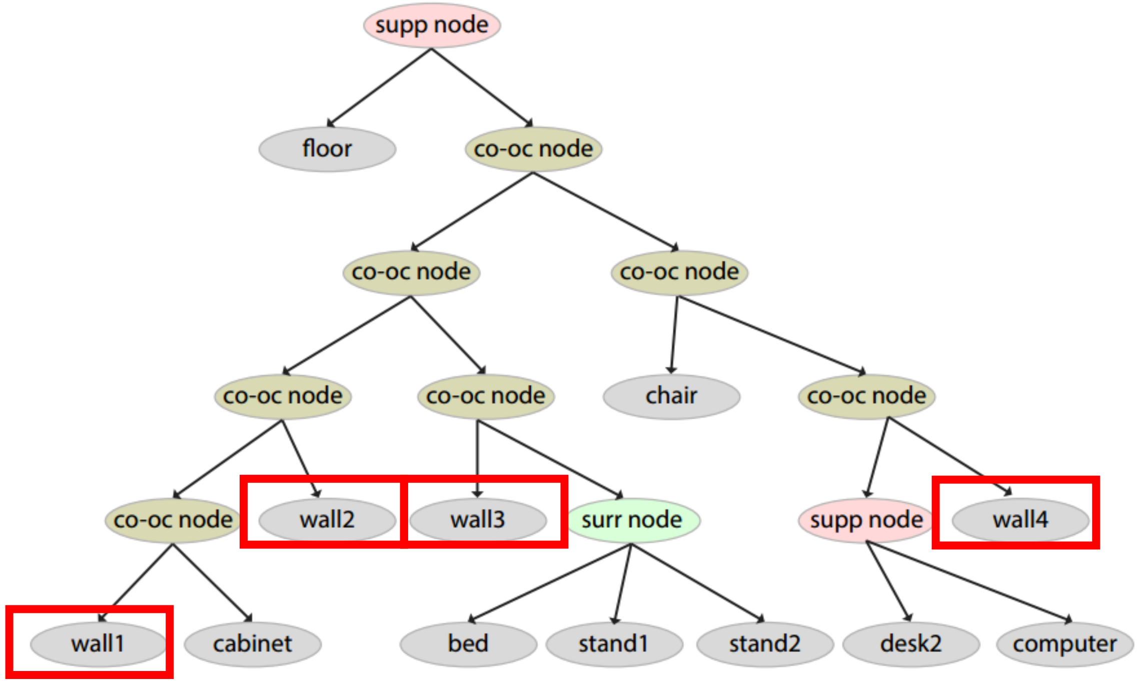
Hierarchy 1



Hierarchy 2



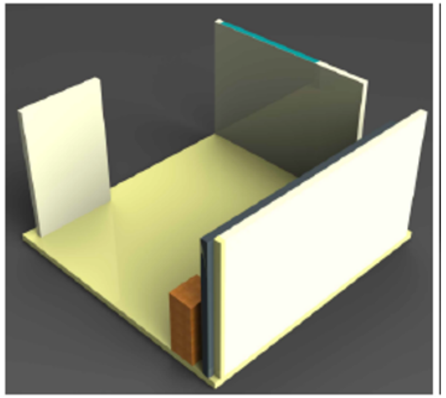
Hierarchy 3

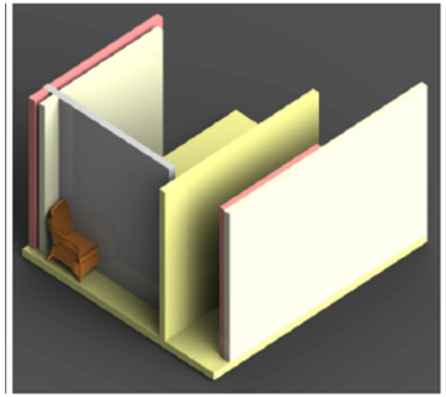



Hierarchy 1:

☹ **No** “wall node”s

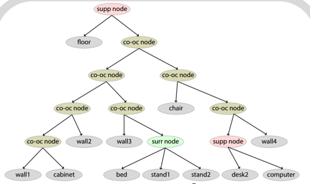
☹ **No** “root node”



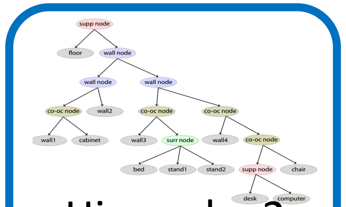


Generated scenes

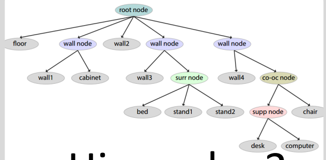
# Ablation Studies: Hierarchical structure



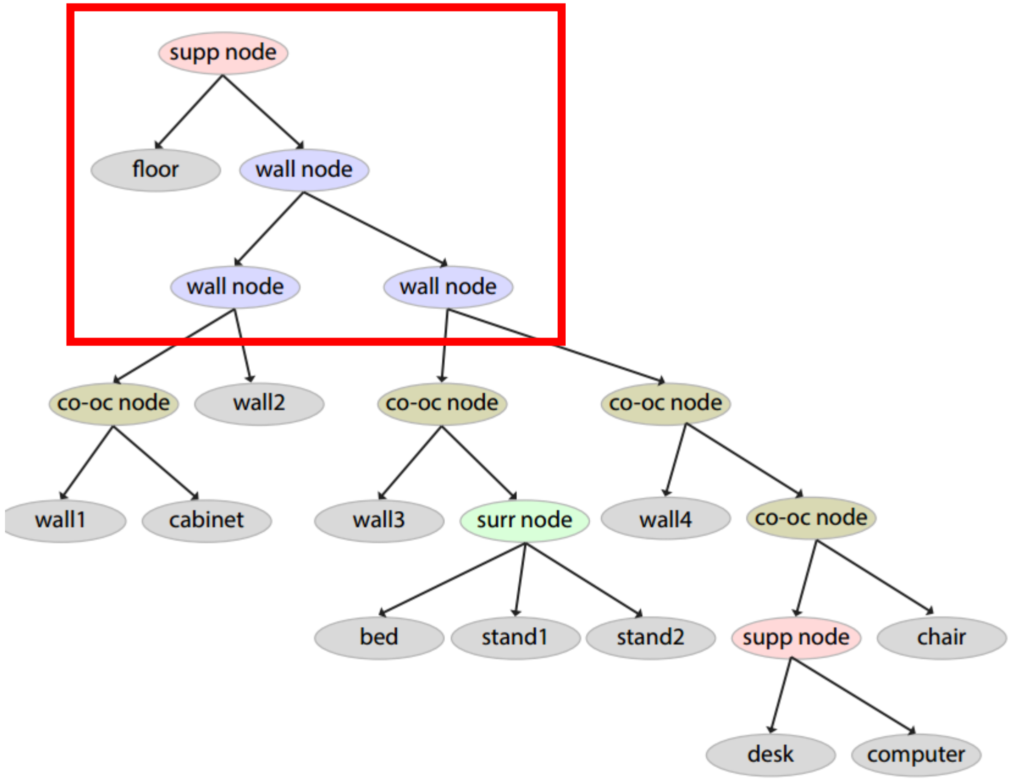
Hierarchy 1



Hierarchy 2



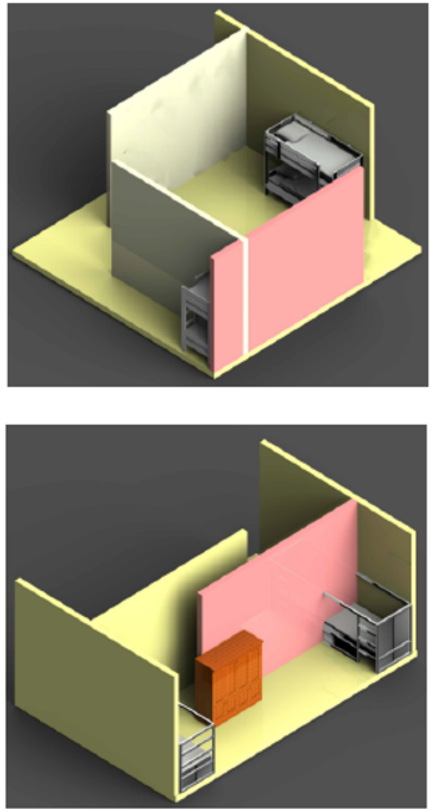
Hierarchy 3




Hierarchy 2:

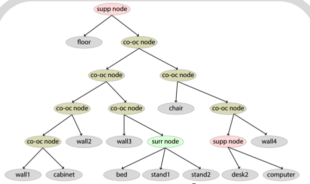
☺ “wall nodes”

☹ No “root nodes”

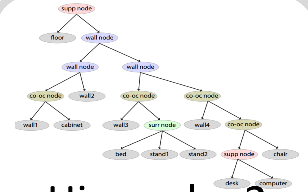


Generated scenes

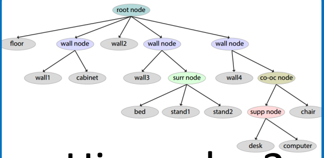
# Ablation Studies: Hierarchical structure



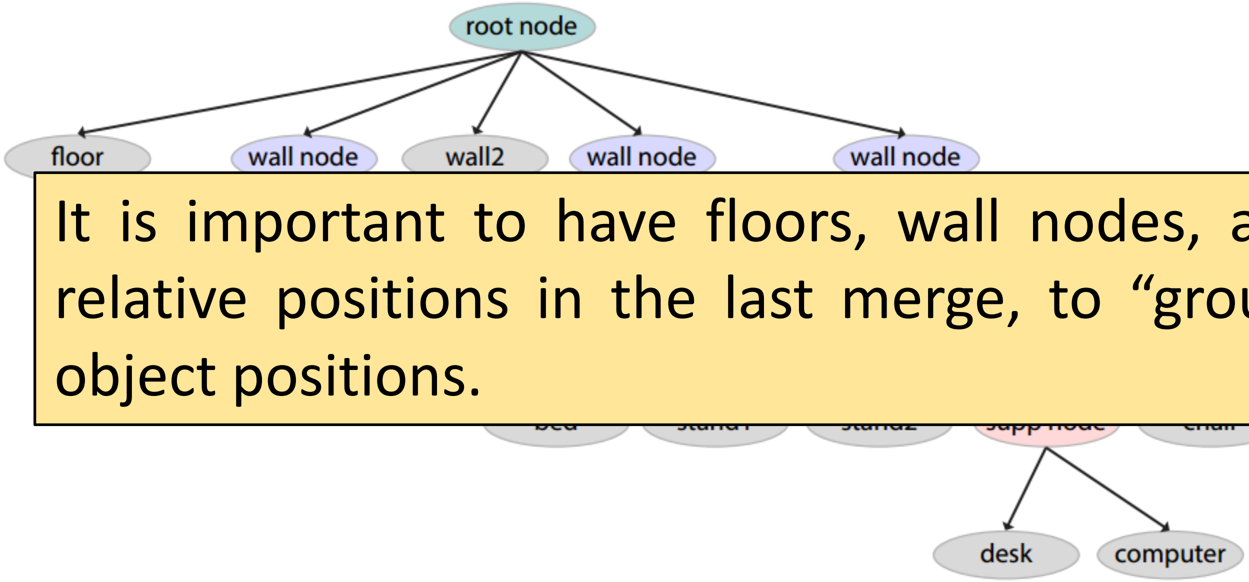
Hierarchy 1



Hierarchy 2



Hierarchy 3




It is important to have floors, wall nodes, and their relative positions in the last merge, to "ground" the object positions.

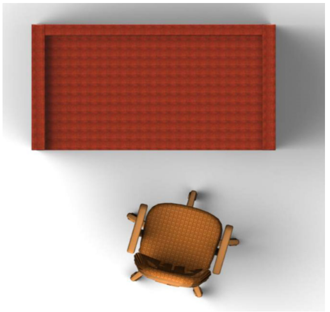
Hierarchy 3 (Ours):

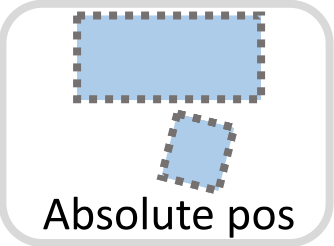
☺ "wall nodes"

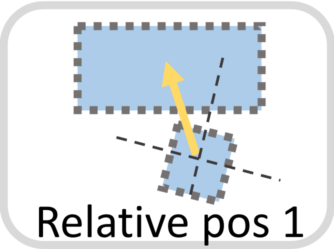
☺ "root nodes"

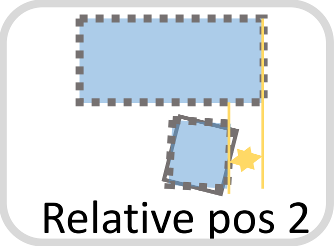
Generated scenes

# Key point 2: Relative Position format

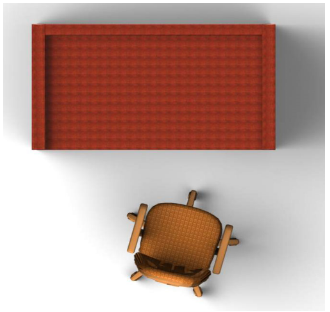


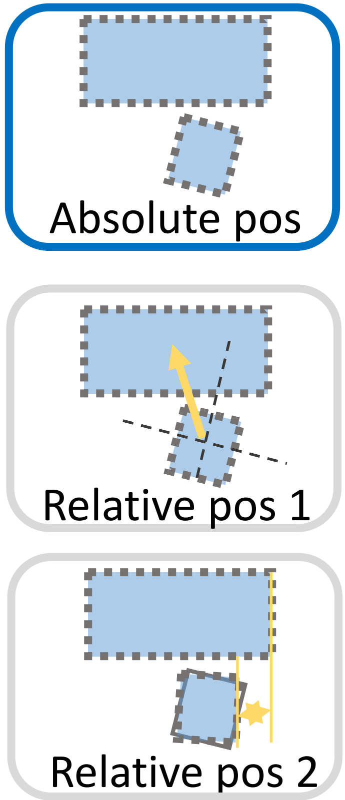


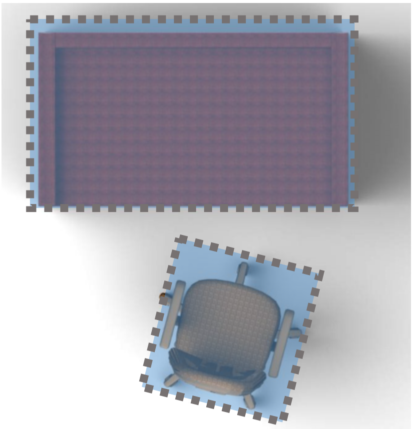



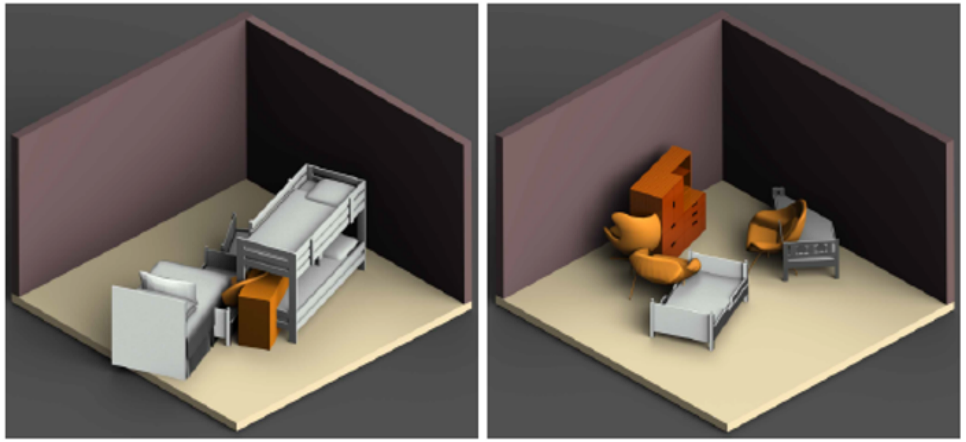
# Ablation Studies : Relative position format





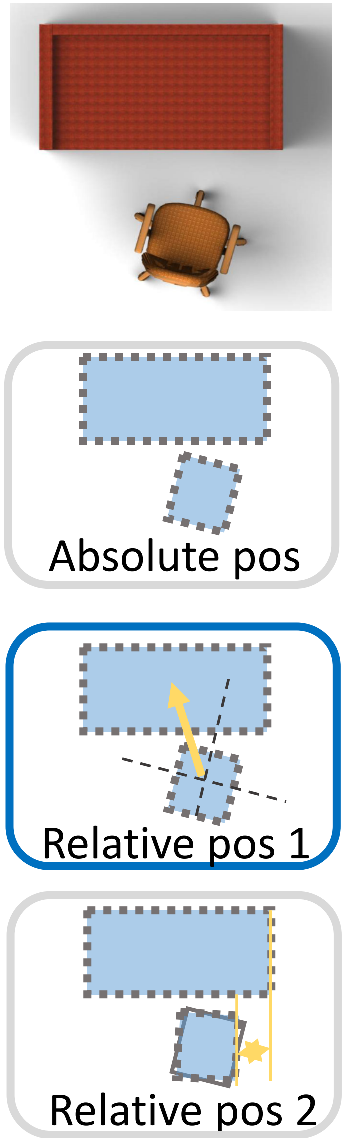



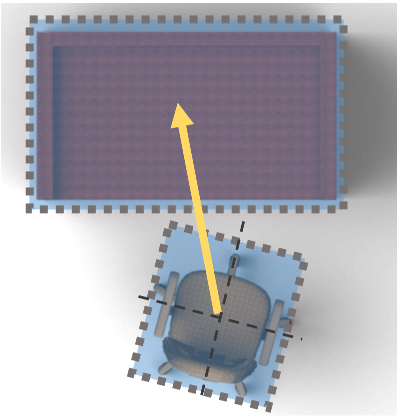
Object’s absolute position
in the leaf nodes



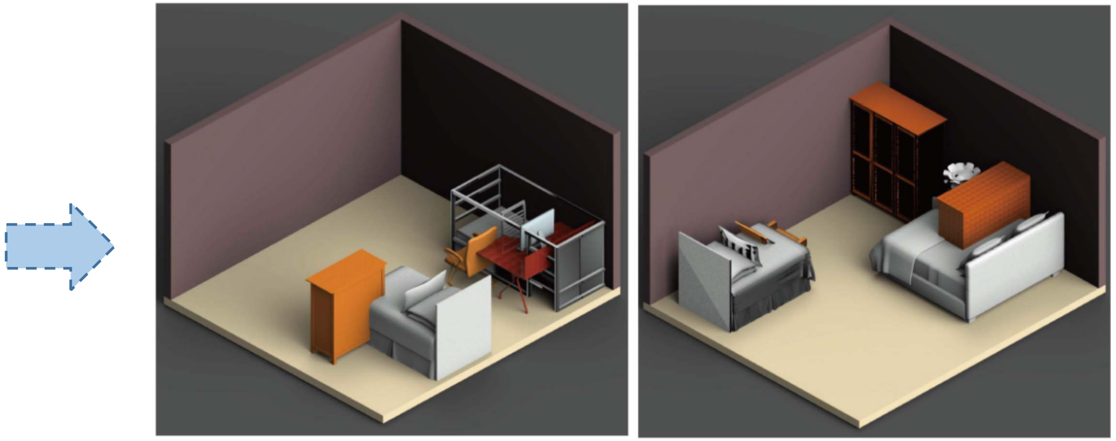
Generated scenes

# Ablation Studies : Relative position format



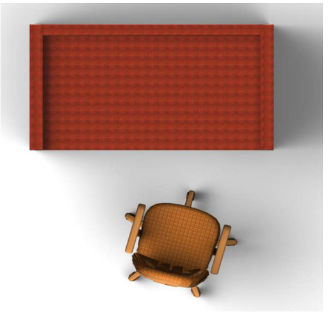



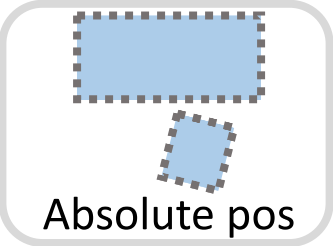
Relative position between
the object centers

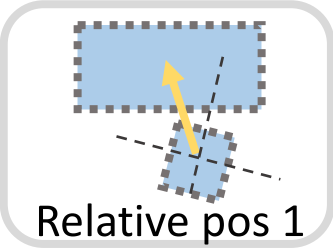


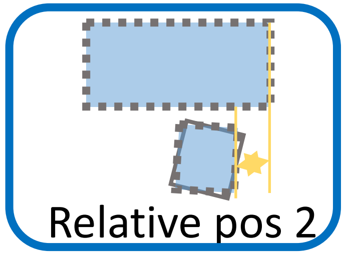
Generated scenes

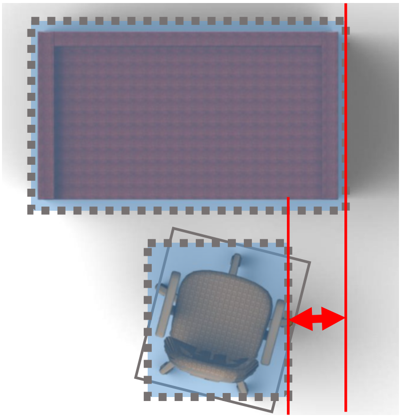
# Ablation Studies : Relative position format



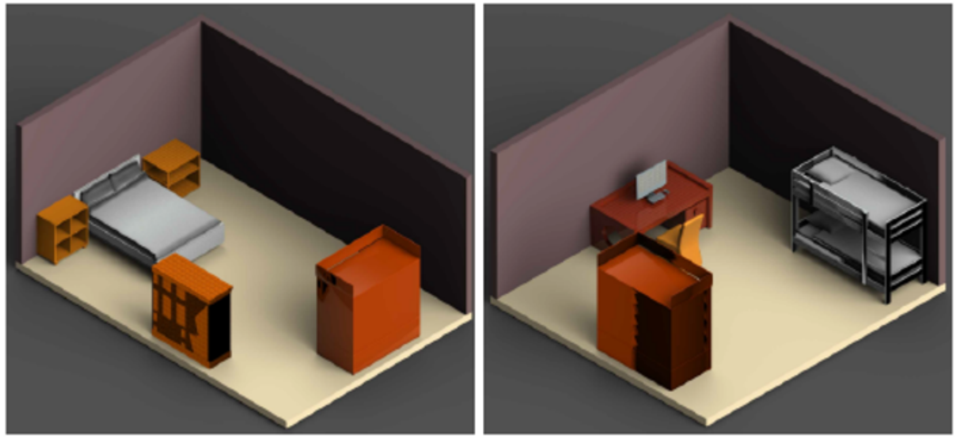


Relative position with offsets between closest edges (ours)

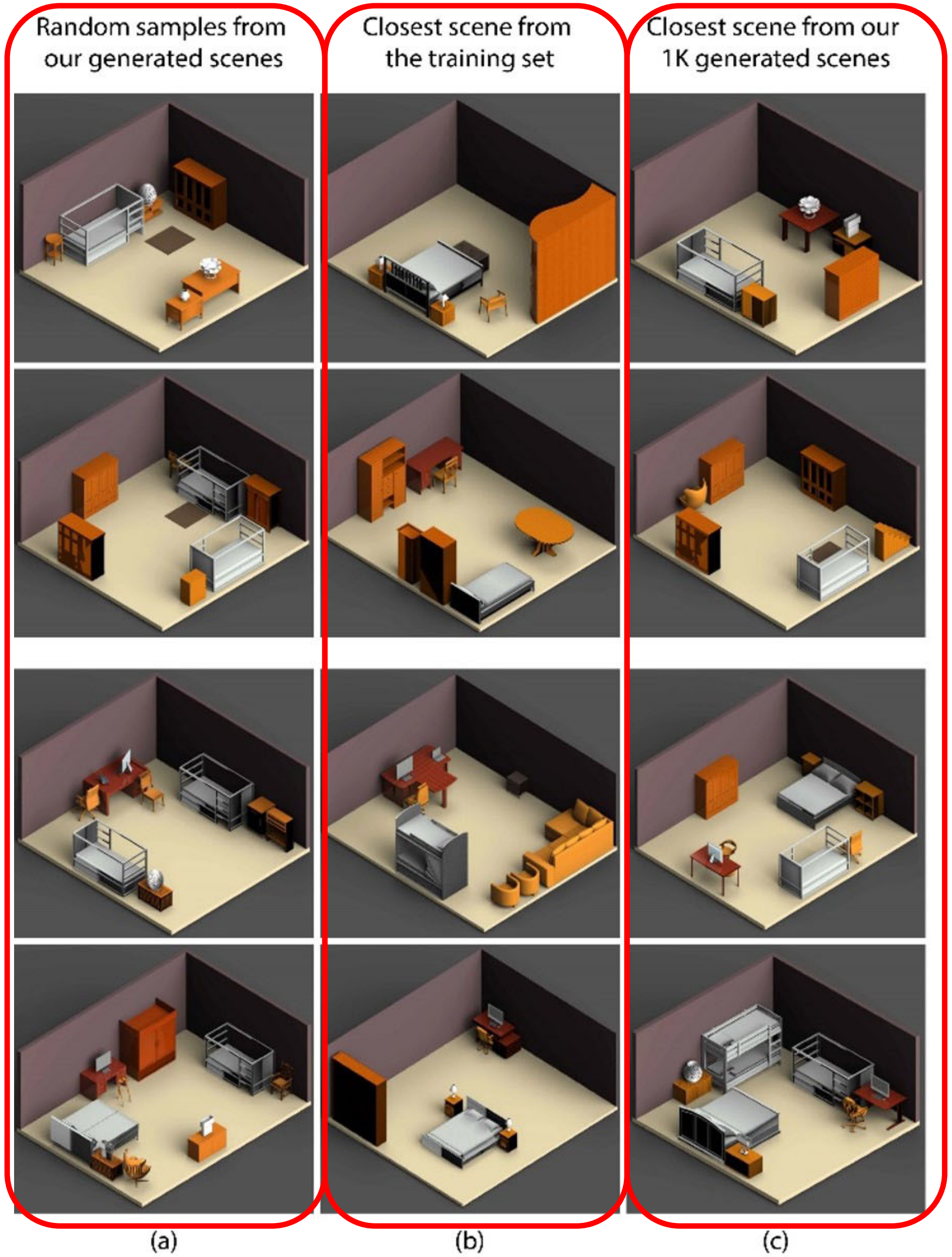


Generated scenes

# Results

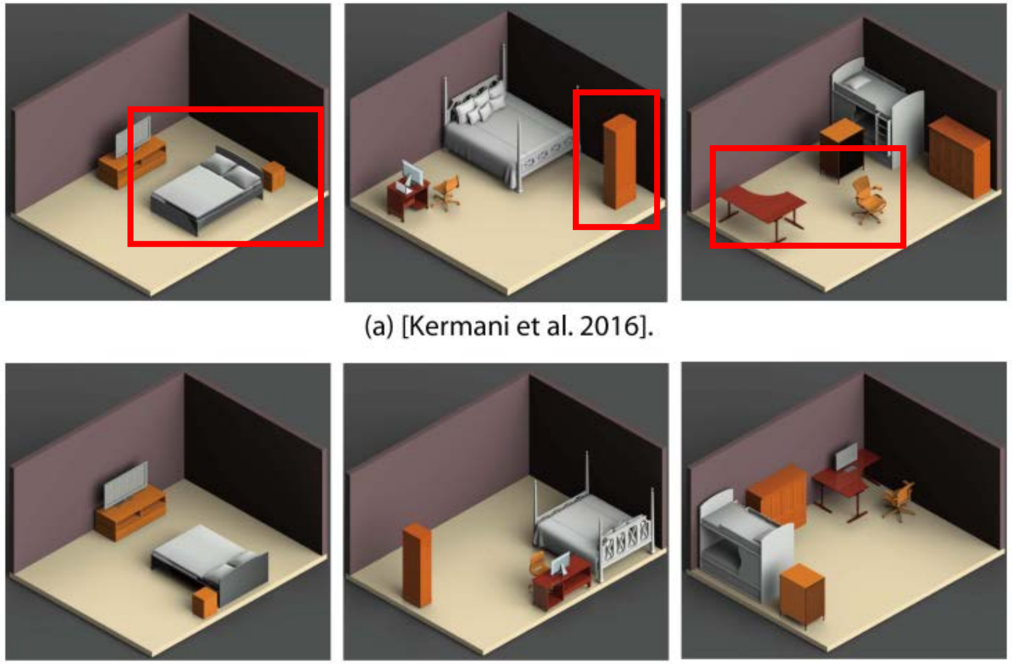
Generated scenes:

- Plausible
- Novel
- Diverse



# Comparison against a graphical model method

- For comparison, we select scenes with the same object shapes.



(a) [Kermani et al. 2016].

(b) Our results.

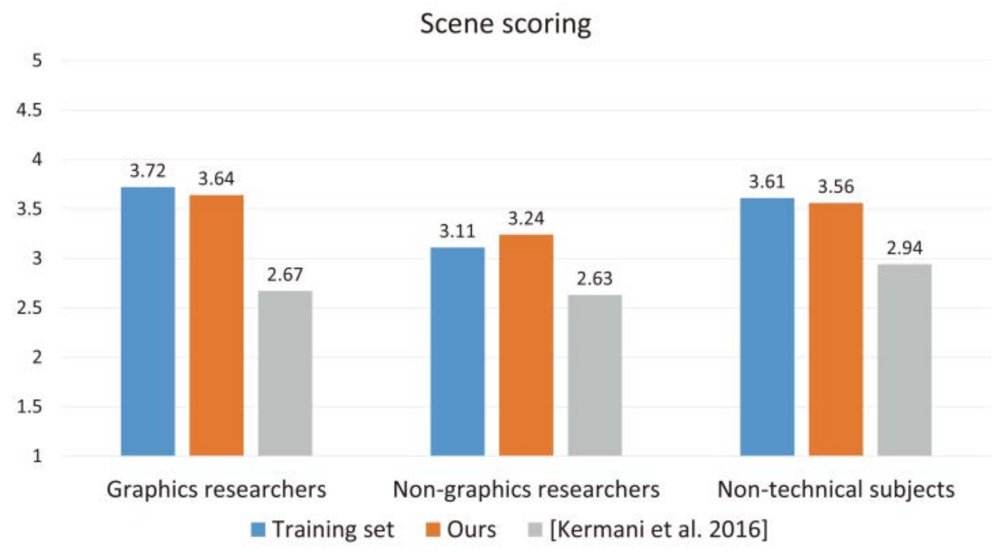
- 3-12min / scene.
- No guarantee on the exact alignment.
- More unreasonable object pairs.

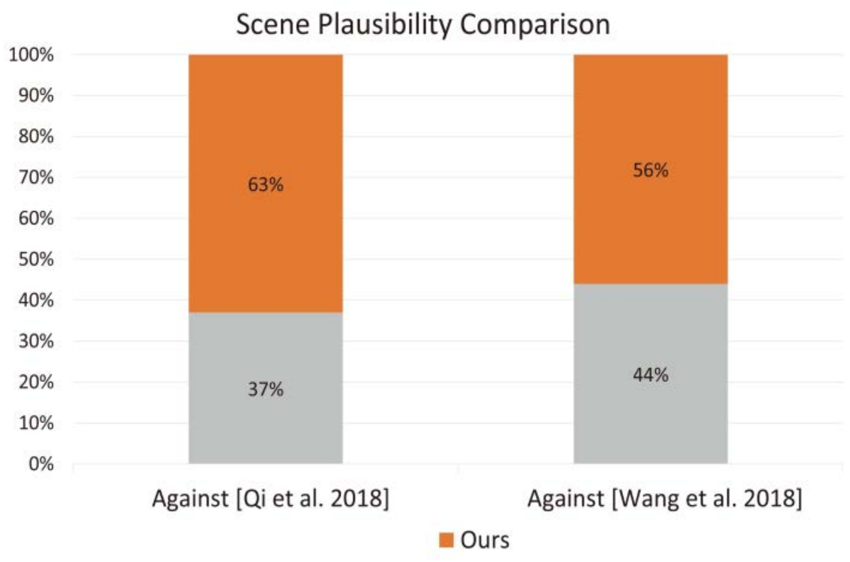
- 0.1027sec / scene.
- Relative positions with attachment and alignment information.

# Comparison: Perceptual studies

We ask users to score or select the scenes based on their plausibility.

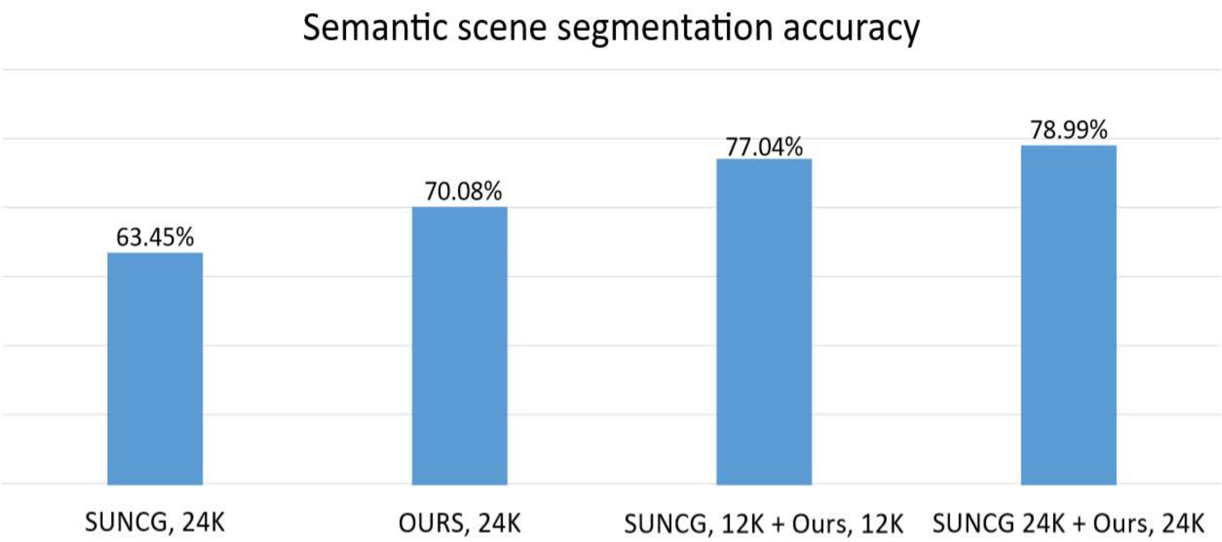
Comparisons are done against (1) the training set, (2) [Kermani et al. 2016]
(3) [Wang et al. 2018], (4) [Qi et al. 2018]



# Applications

- Data augmentation method for deep learning tasks

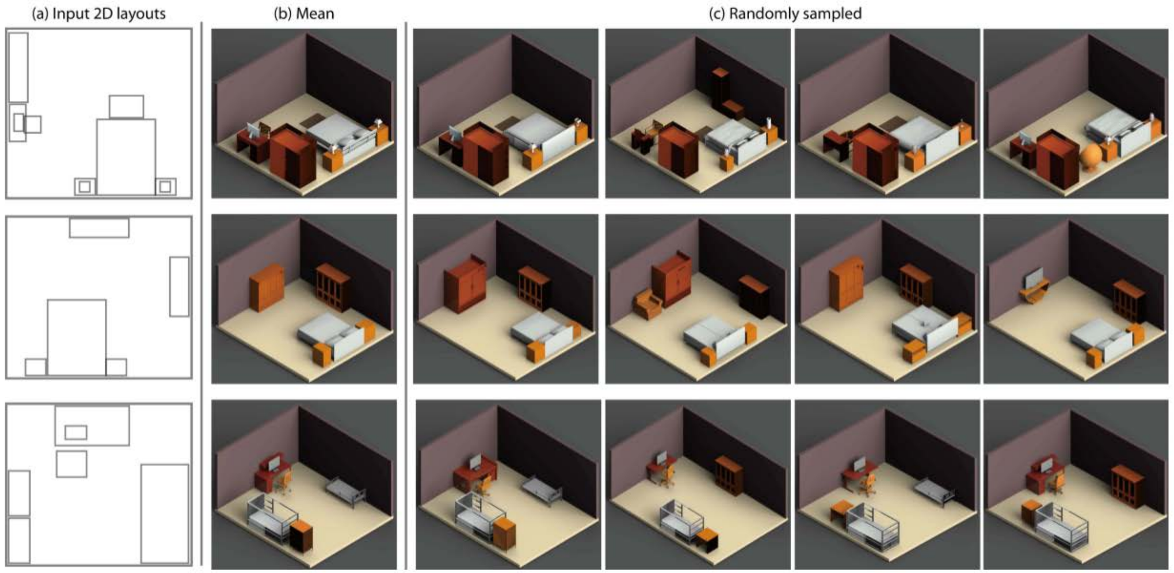



Semantic scene segmentation task:

- Network: PointNet [Qi et al. 2017]
- Dataset: Indoor scenes as point clouds
- Results: More *relevant* training data, better learning performance and generalization.

# Applications

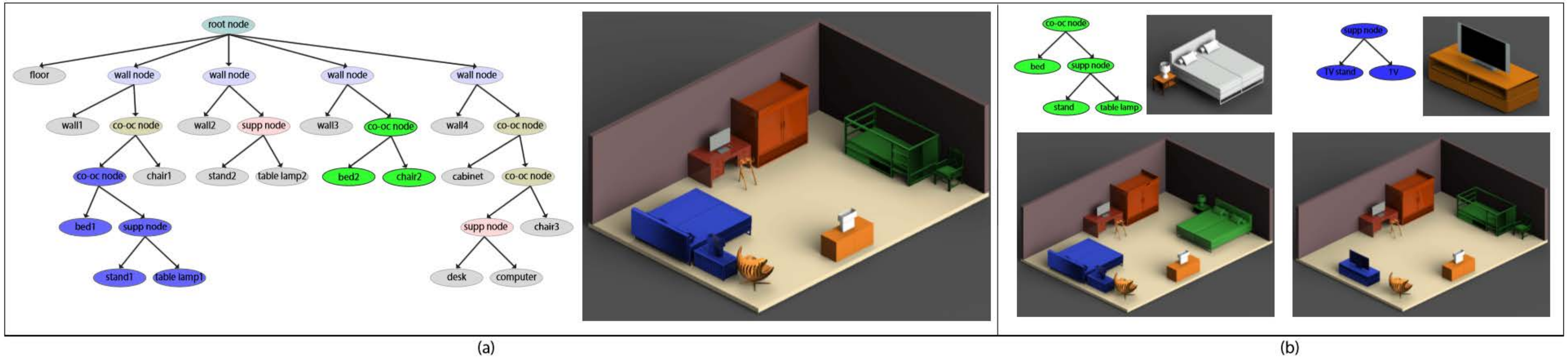
- Data augmentation method for deep learning tasks
- 2D layout guided 3D scene modeling




- Goal: 2D box layout to 3D indoor scene
- Network: Pre-trained RvNN-VAE on 3D scenes
- Result: Transform between multi-modal data which share the same hierarchical structures.

# Applications

- Data augmentation method for deep learning tasks
- 2D layout guided 3D scene modeling
- Hierarchy-guided scene editing




Hierarchical indoor scene structure helps designers to edit a scene at the sub-scene level.

# Conclusion

- We present a generative neural network which enables us to generate plausible 3D indoor scenes in large quantities and varieties, easily and highly efficiently.
- We study the influence of different scene representations on the learning ability of generative RvNNs.
- We show the applications of our generated scenes with the corresponding hierarchies.

# Thank you!