End-to-End Optimization of Scene Layout

3dsln.csail.mit.edu

Andrew Luo[1]

Zhoutong Zhang[2]

Jiajun Wu[3]

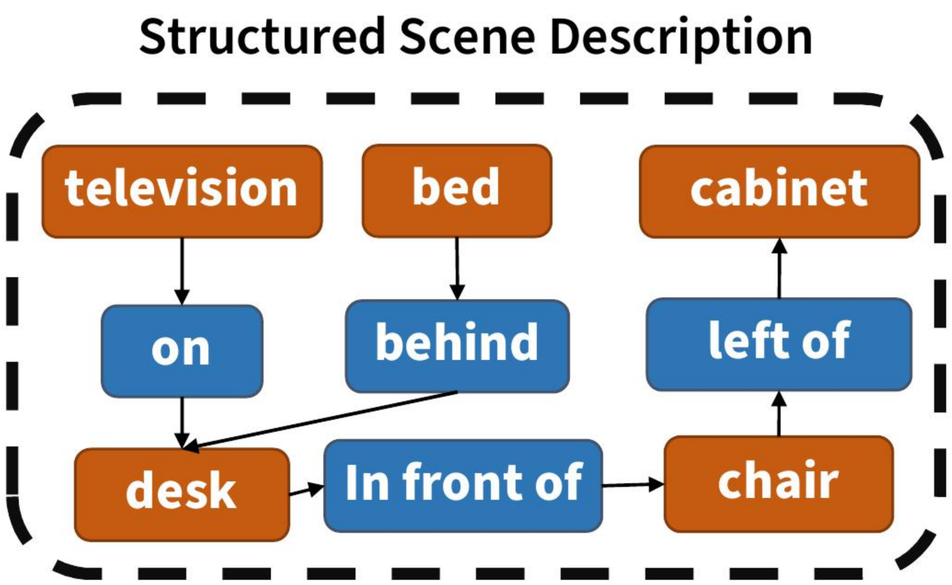
Joshua B. Tenenbaum[2]

1 Carnegie Mellon University

2 MIT

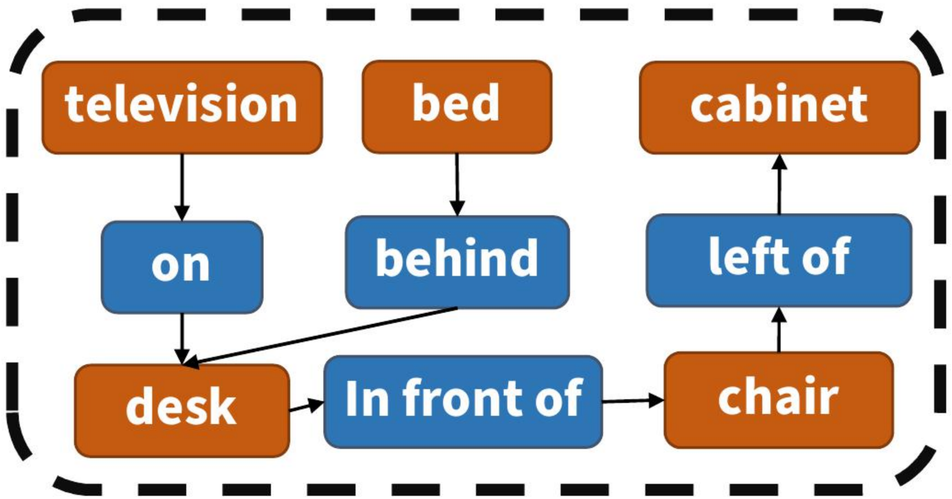
3 Stanford University

Motivation


Structured Scene Description
television
bed
cabinet
on
behind
left of
desk
In front of
chair

Motivation

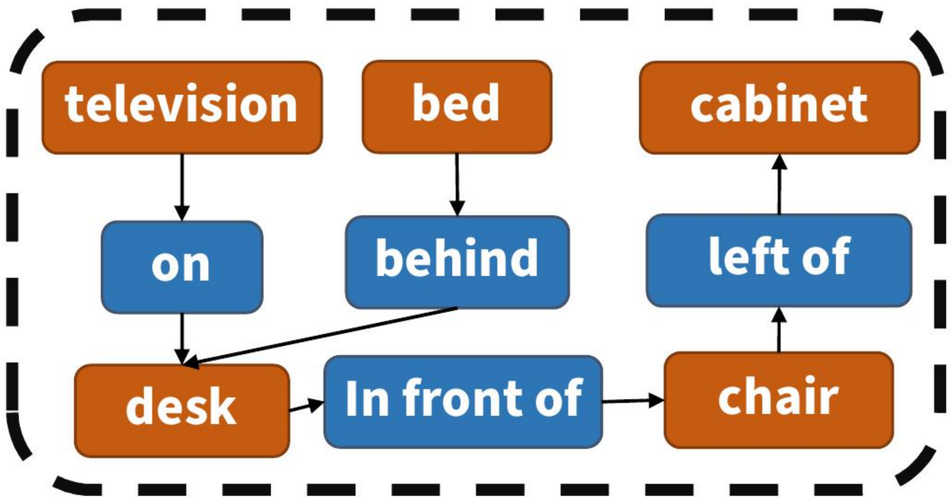
Structured Scene Description



How do we position the objects?

Motivation

Structured Scene Description



How do we position the objects?

Previous approaches

- Traditional scene graph based image generation[1]
  - Model objects relations in image space
  - No explicit concept of 3D objects (only pixels)

[1] Johnson et al., CVPR ’18 - *Image Generation from Scene Graphs*

Previous approaches

- Traditional scene graph based image generation[1]
  - Model objects relations in image space
  - No explicit concept of 3D objects (only pixels)

- Layout Generation[2]
  - No spatial conditioning
  - Auto-regressive (slow)

[1] Johnson et al., CVPR '18 - *Image Generation from Scene Graphs*
[2] Wang et al., SIGGRAPH '18 - *Deep Convolutional Priors for Indoor Scene Synthesis*

Previous approaches

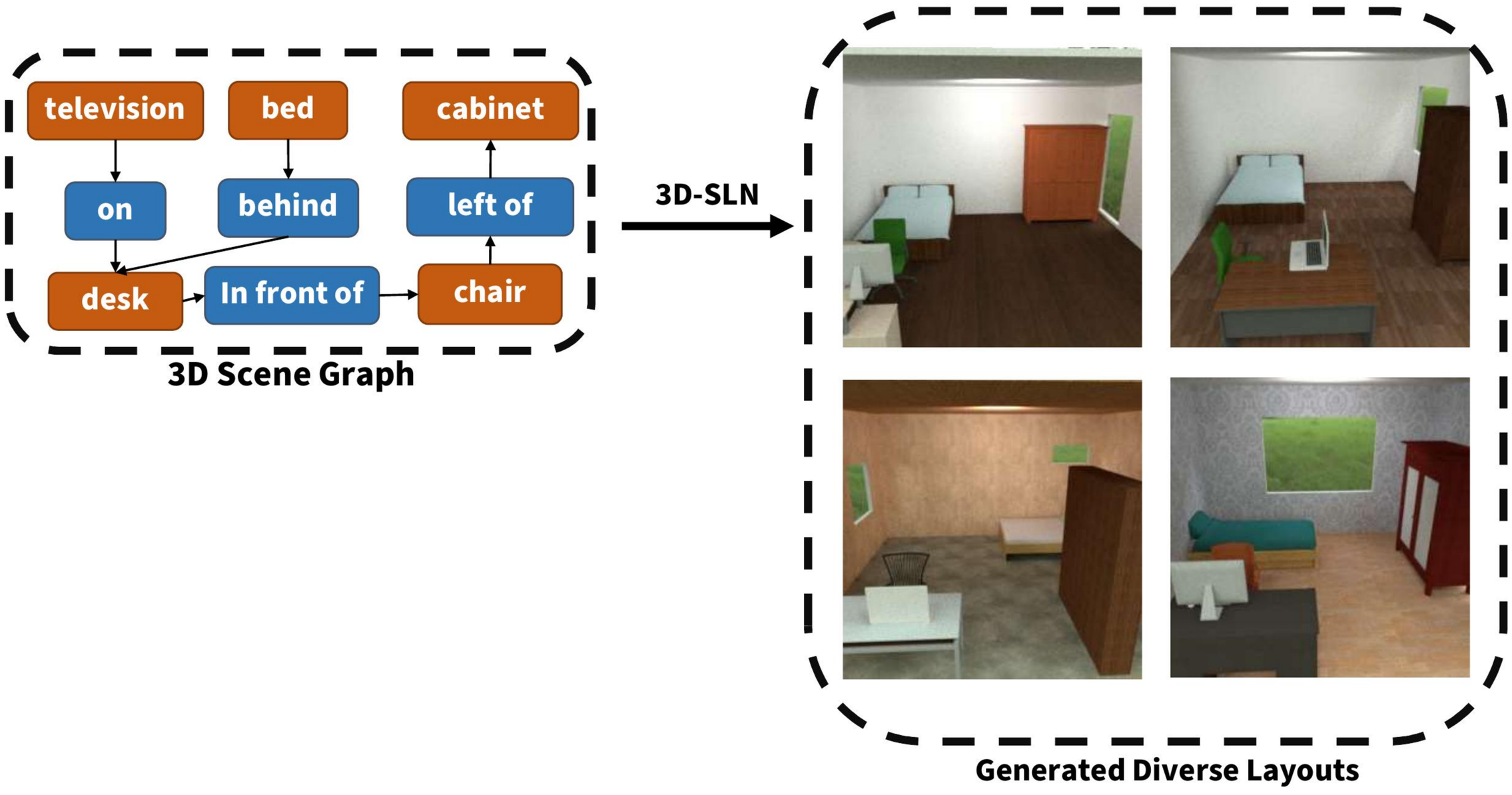
- Traditional scene graph based image generation[1]
  - Model objects relations in image space
  - No explicit concept of 3D objects (only pixels)

- Layout Generation[2]
  - No spatial conditioning
  - Auto-regressive (slow)

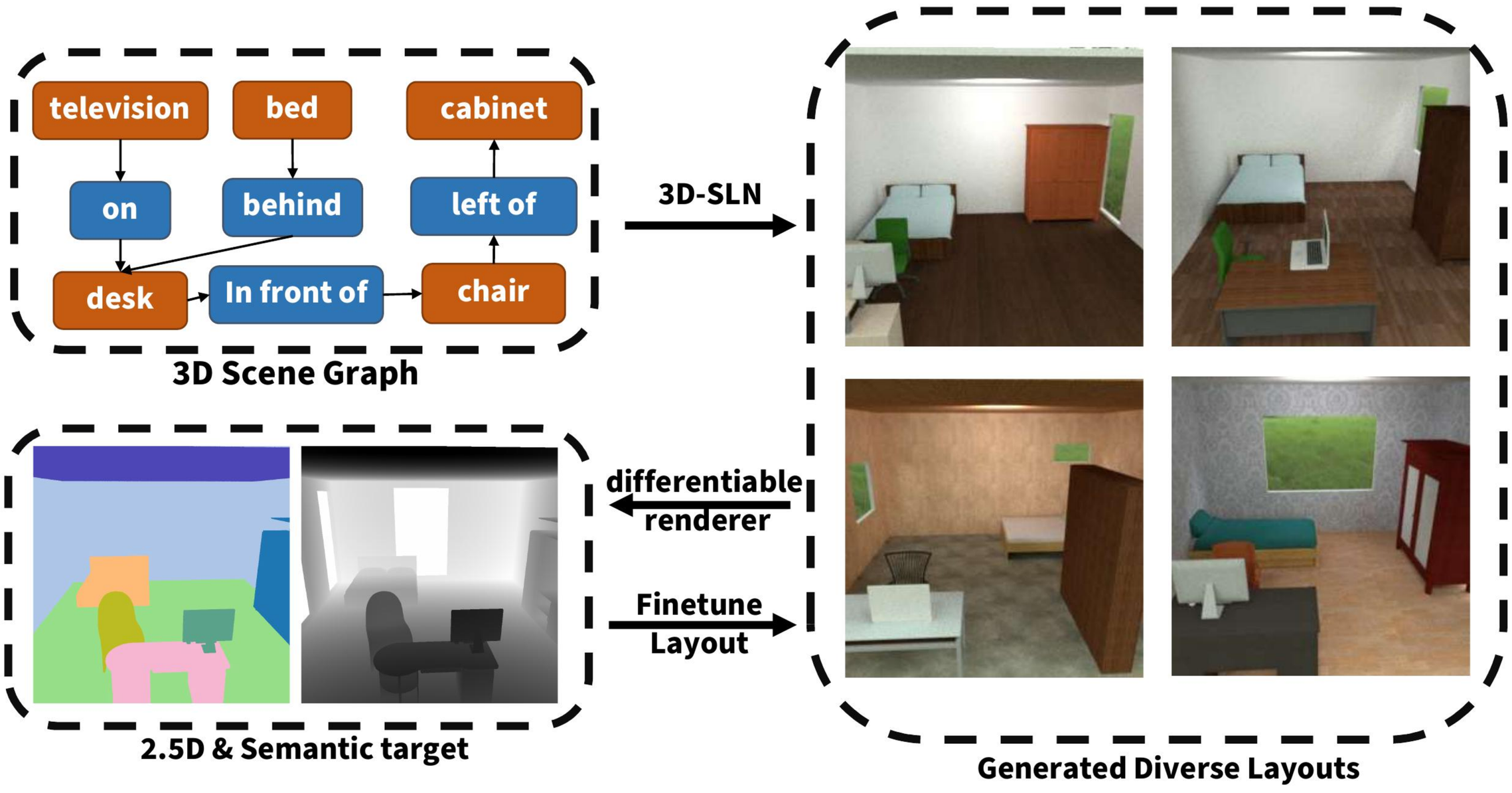
Issues to be addressed here
  - 3D relationships in scene space
  - Disentangled layout, shape, and image formation
  - Refinement of object locations based on 2.5D + semantic target

[1] Johnson et al., CVPR '18 - *Image Generation from Scene Graphs*
[2] Wang et al., SIGGRAPH '18 - *Deep Convolutional Priors for Indoor Scene Synthesis*

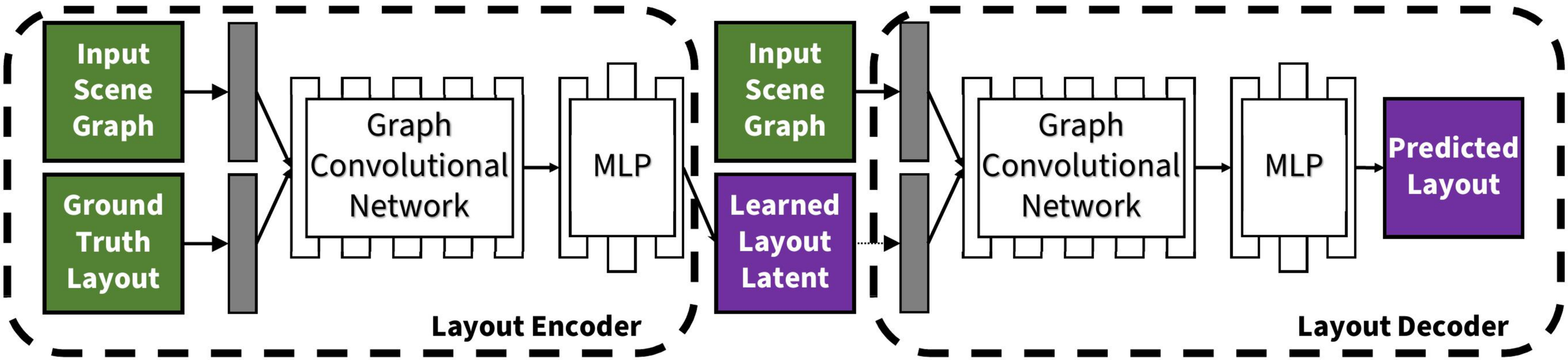
Scene Generation



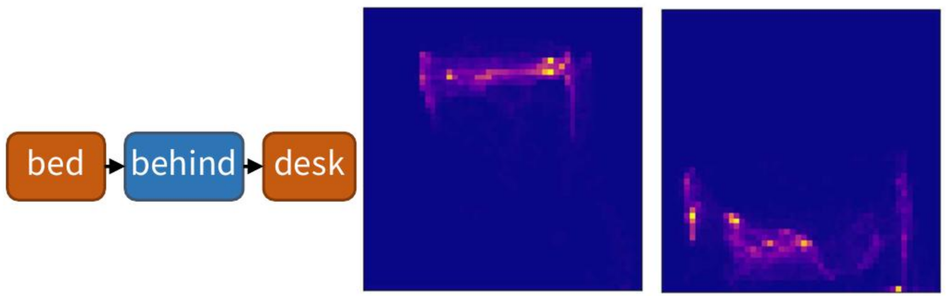
Scene Generation + Refinement

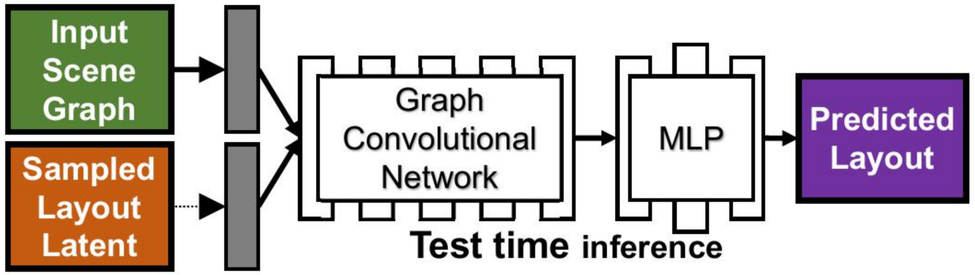


3D-SLN:
variational autoencoder graph CNN



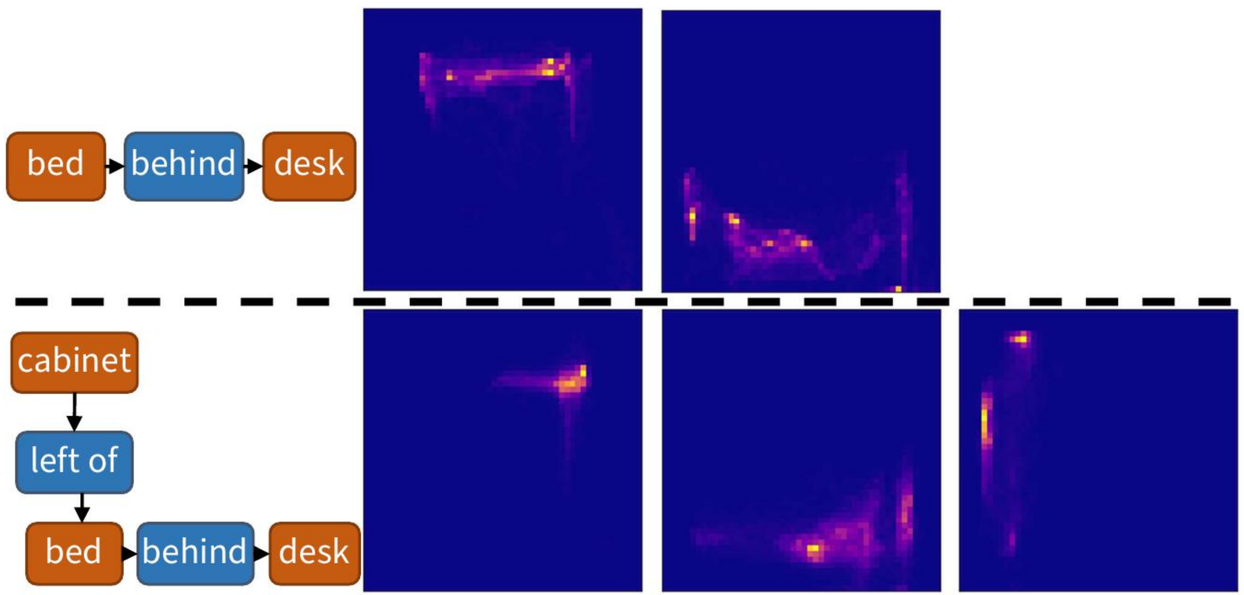
Diverse Layout Generation

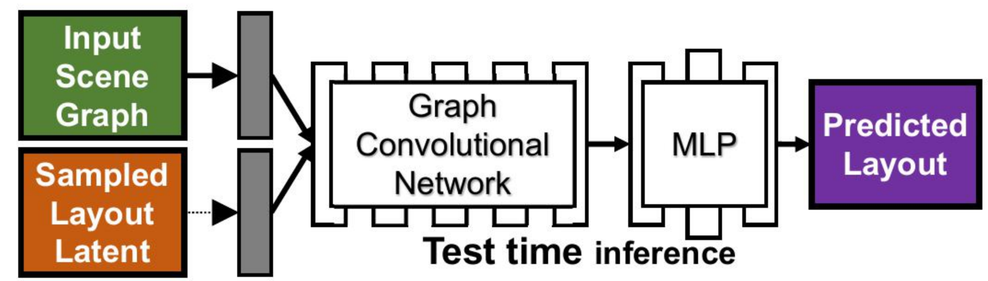




Bed **Desk** **Cabinet** **Chair** **Lamp**

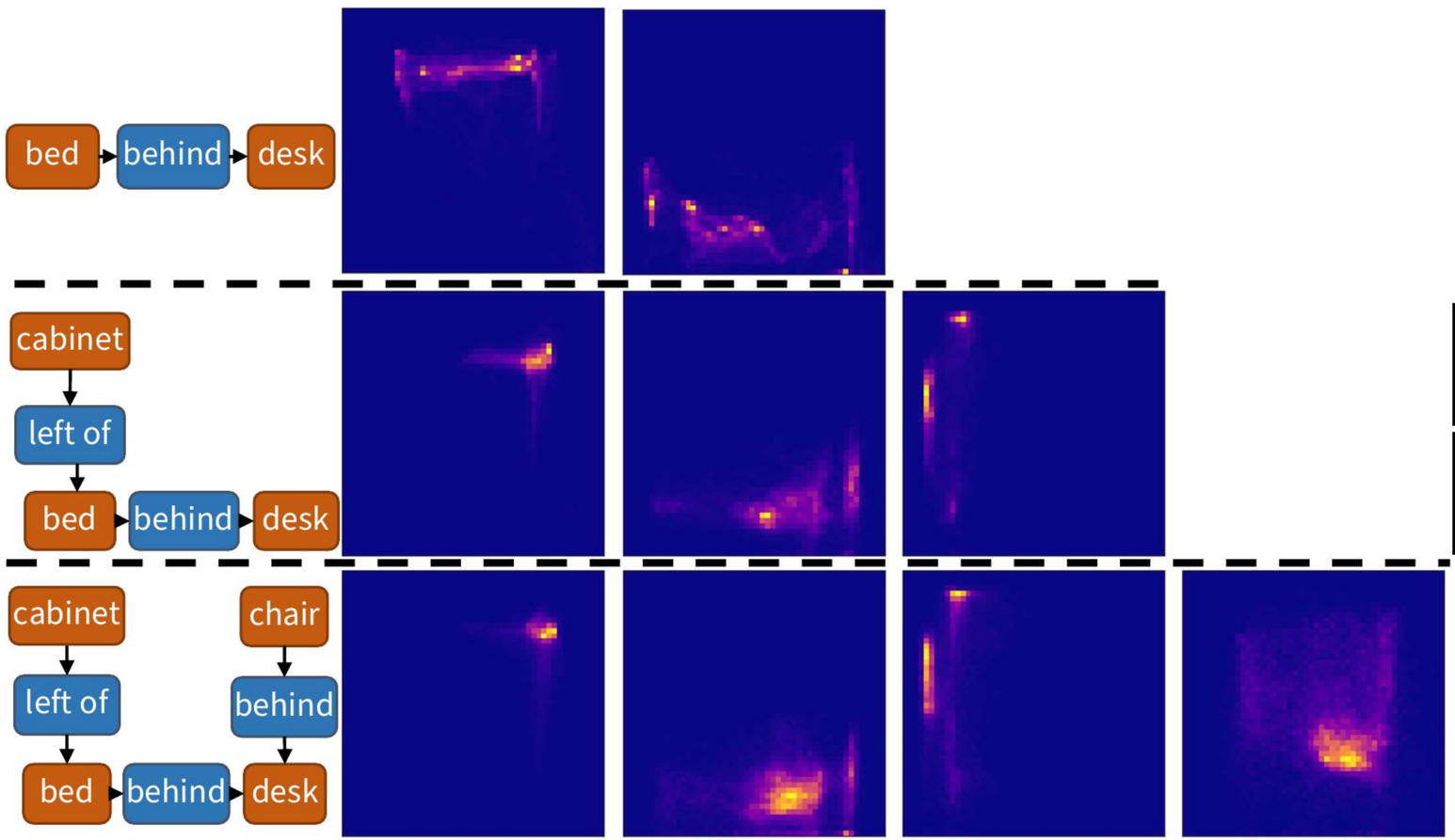
Diverse Layout Generation

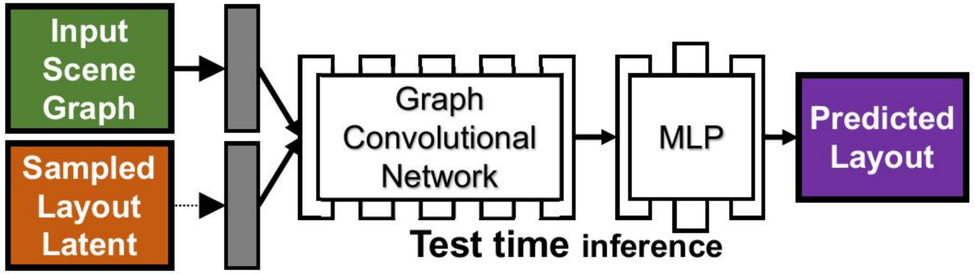




Bed **Desk** **Cabinet** **Chair** **Lamp**

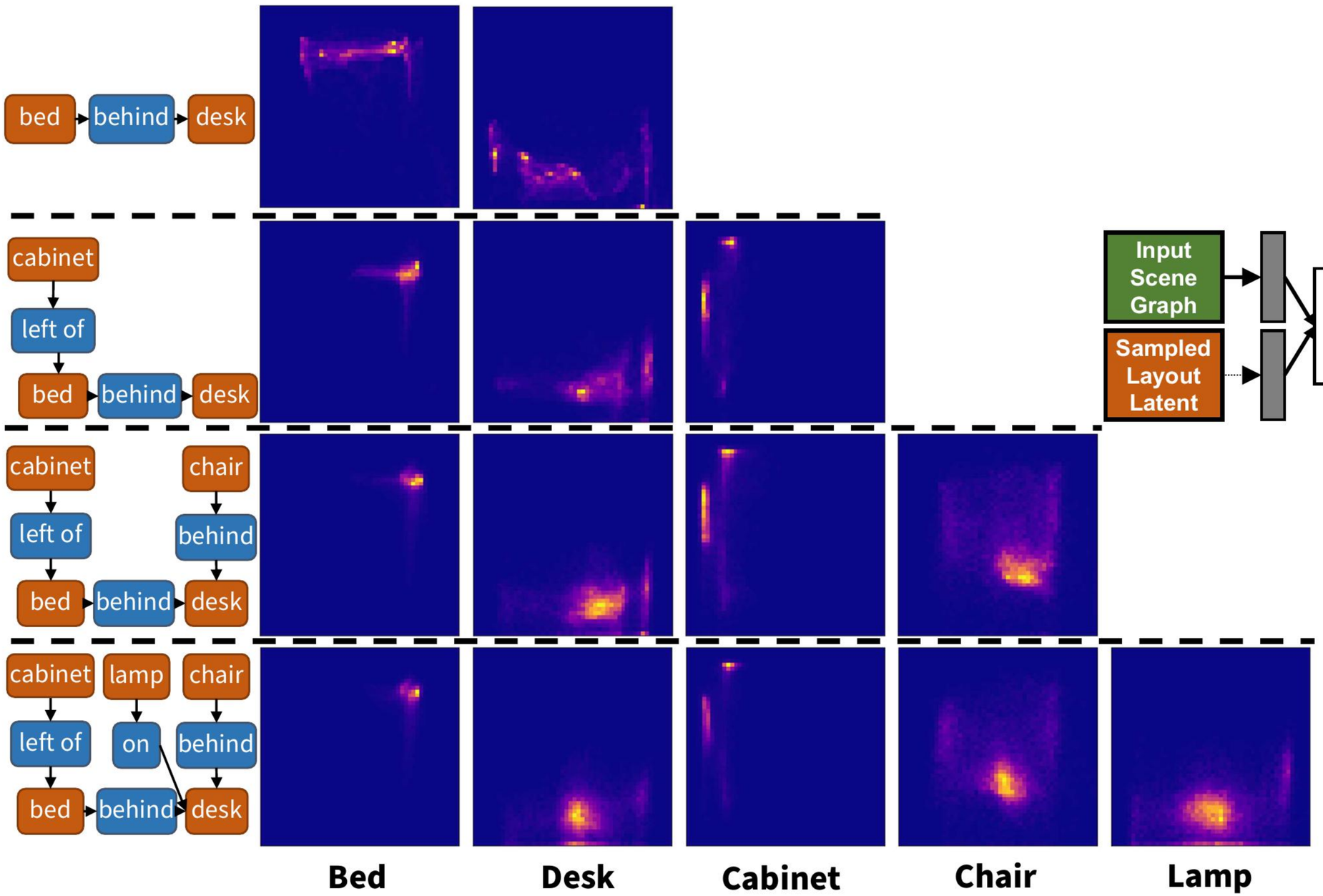
Diverse Layout Generation

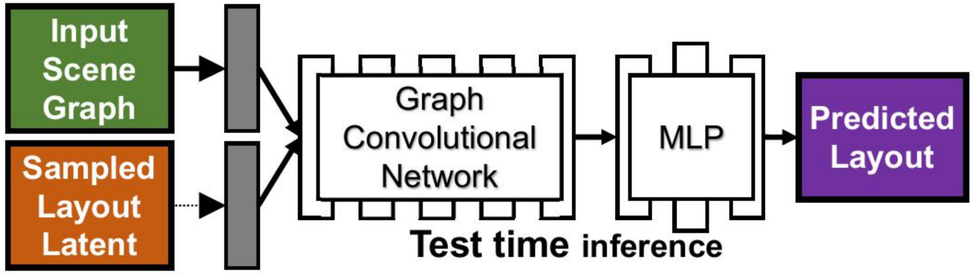




Bed **Desk** **Cabinet** **Chair** **Lamp**

Diverse Layout Generation





Quantitative Results

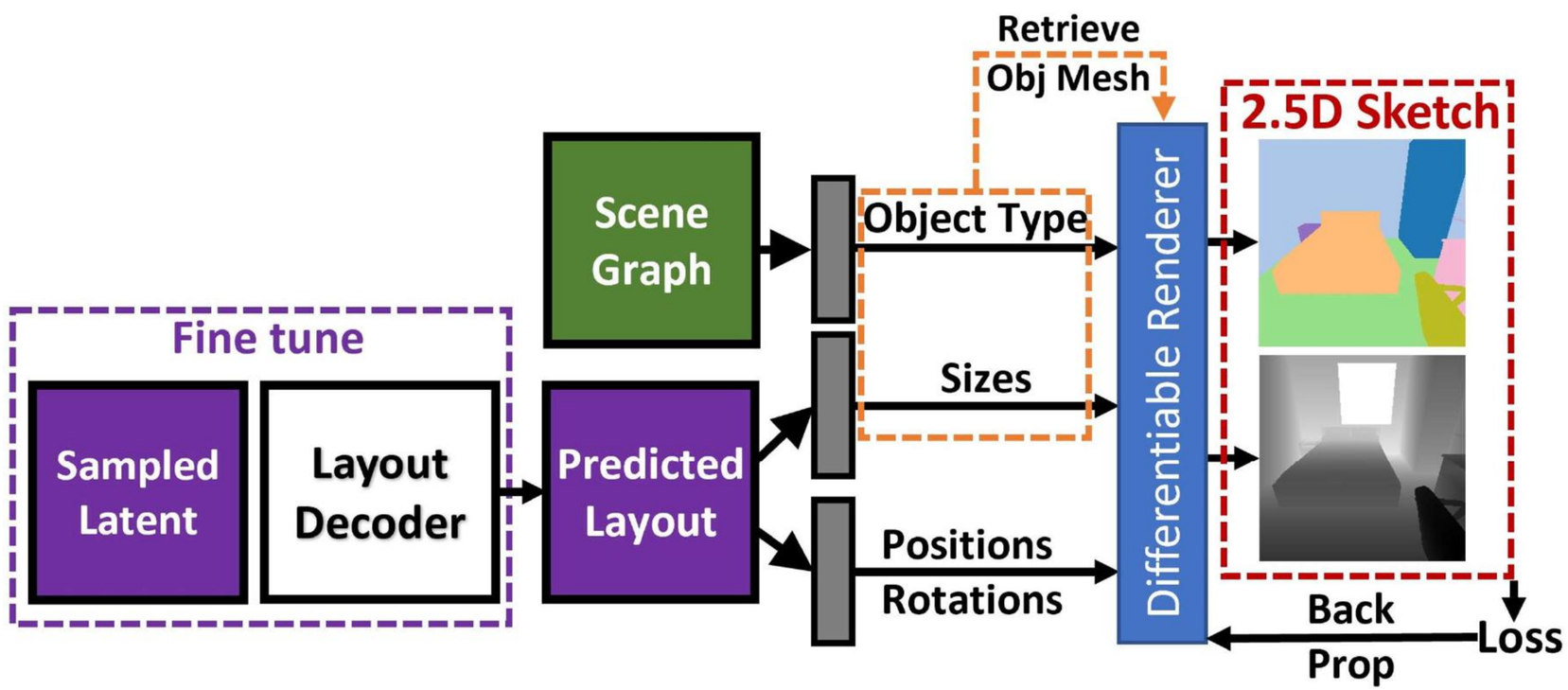
Model	Scene Graph Acc. (%)	L_1 box loss	STD (size)	STD (position)	STD (rotation)
Random Layout	57.1	0.317	0.000	0.244	6.48
Perturbed Layout	82.6	0.080	0.000	0.100	3.00
DeepSynth[1]	N/A	N/A	N/A	0.129	2.27
GCN	86.3	0.111	0.000	0.000	0.00
GCN + noise	86.9	**0.109**	0.001	0.002	0.18
3D-SLN (ours)	**94.3**	0.148	**0.026**	**0.078**	**4.77**

Note: rotation is measured in units of 15° increments

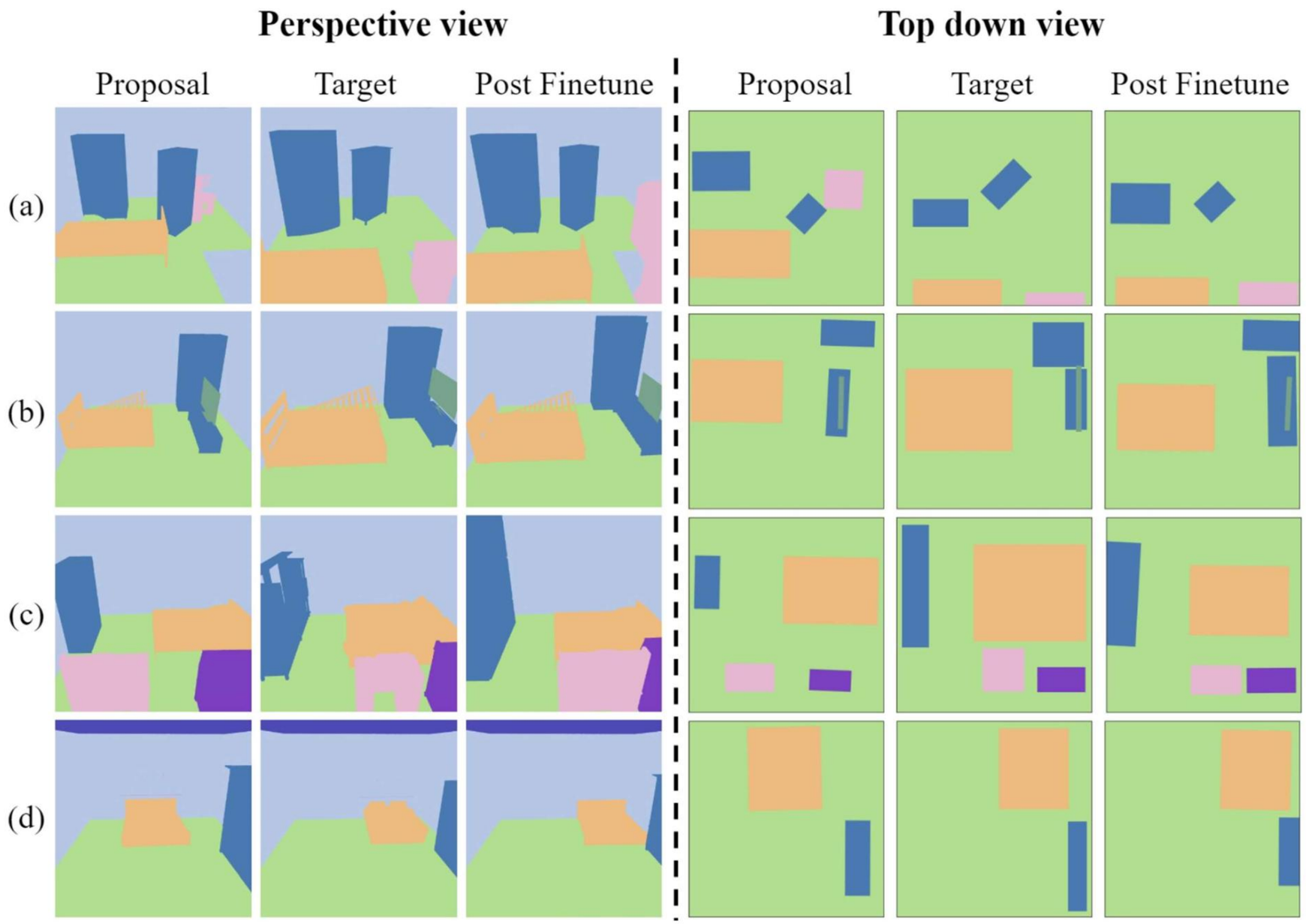
[1] Wang et al., SIGGRAPH '18 - *Deep Convolutional Priors for Indoor Scene Synthesis*

Refinement of Scene Layout

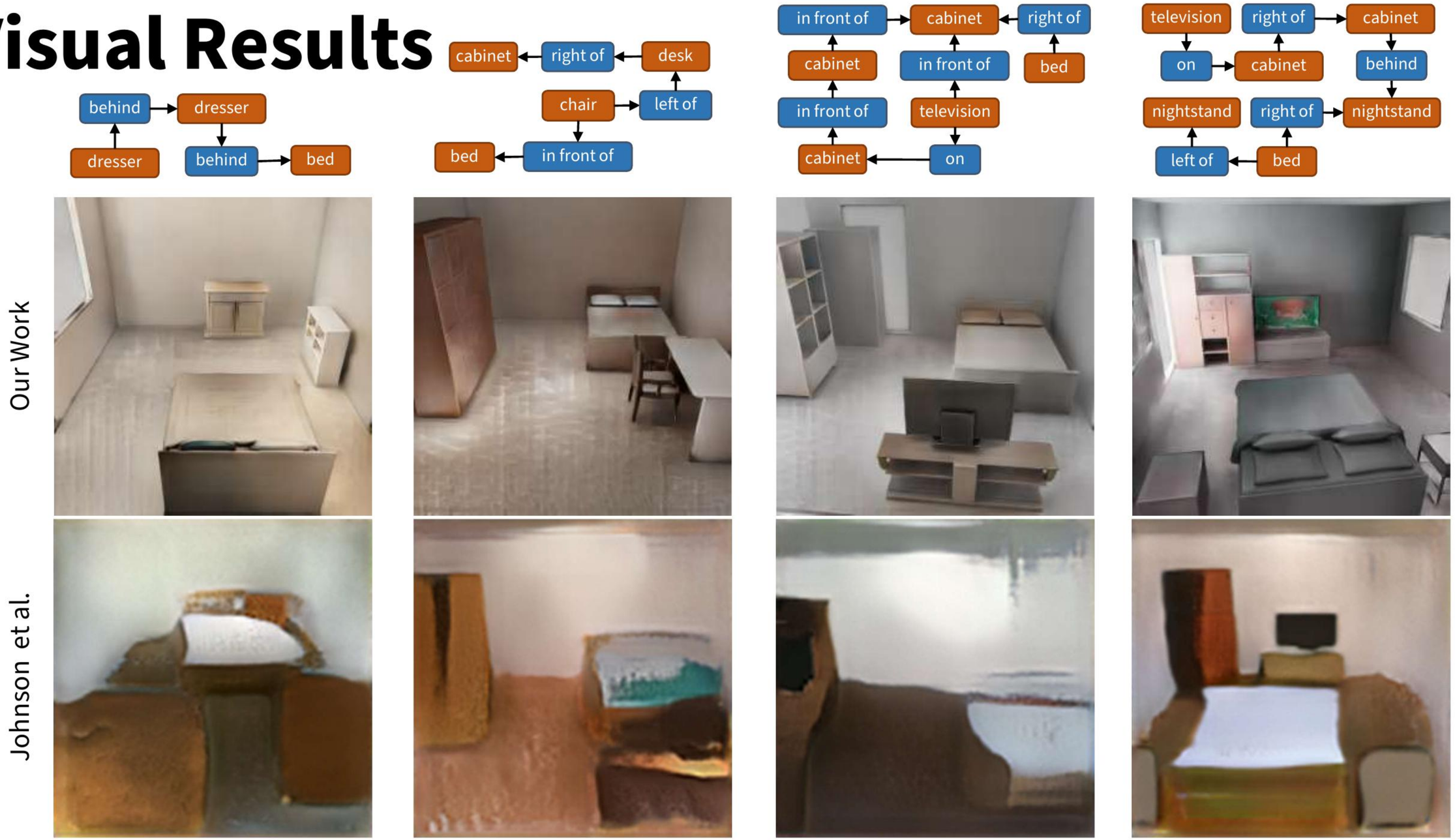
Using 2.5D + Semantic to fine-tune 3D scene layouts



Object Layout Refinement

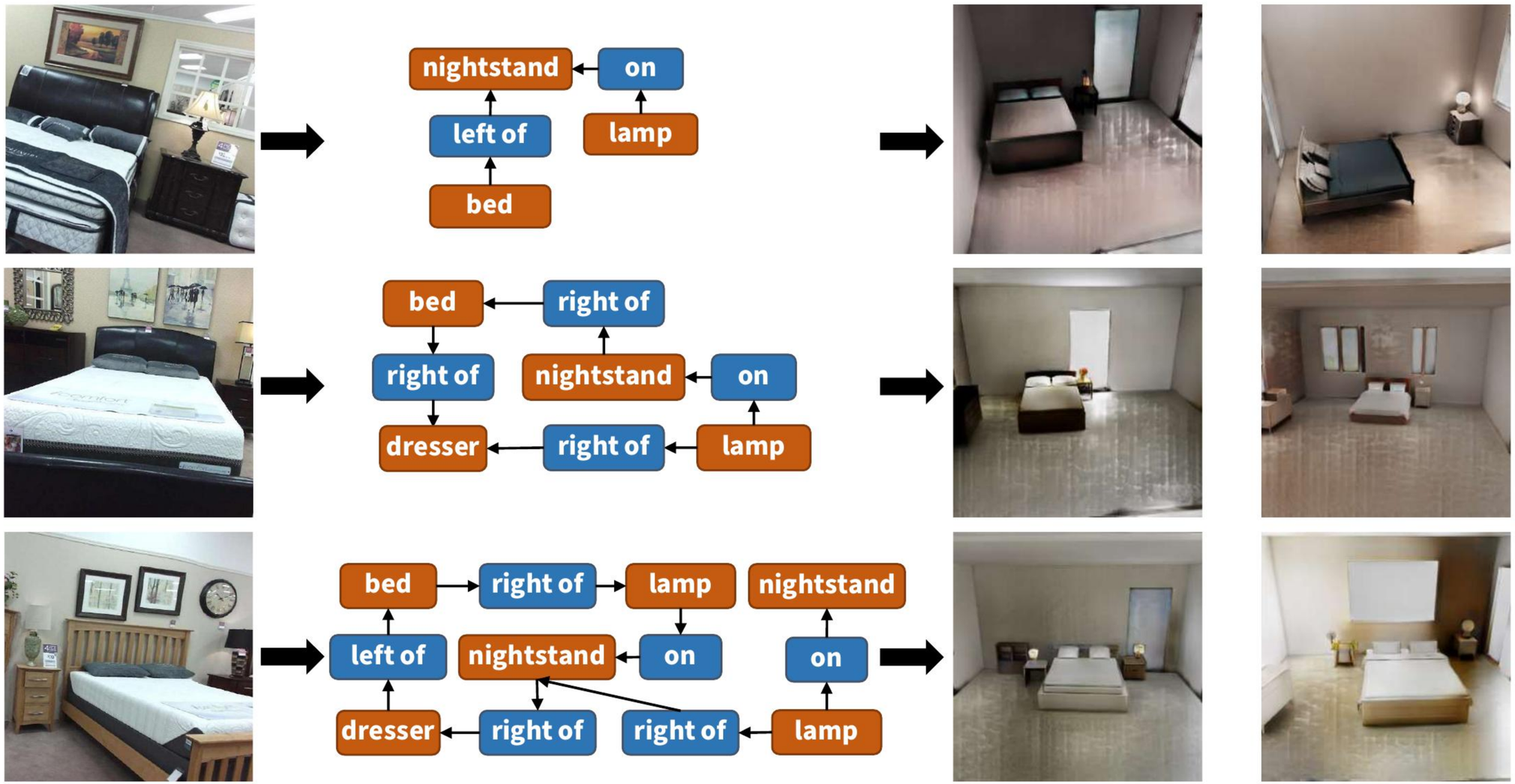


Visual Results



- Park et al., CVPR '19 - *Semantic Image Synthesis with Spatially-Adaptive Normalization*
- Johnson et al., CVPR '18 - *Image Generation from Scene Graphs*

New Layouts from Exemplar Images



3D Obj bbox parsed with: Huang et al., NeurIPS '18 - *Cooperative Holistic Scene Understanding*

Conclusion

What's cool:

- The 3D-SLN model generates **diverse and accurate** scene layouts from a scene graph
- The 3D scene layouts can be finetuned using 2.5D + semantic information
- Useful for layout synthesis and image synthesis.

- Code will be released soon!

http://3dsln.csail.mit.edu/