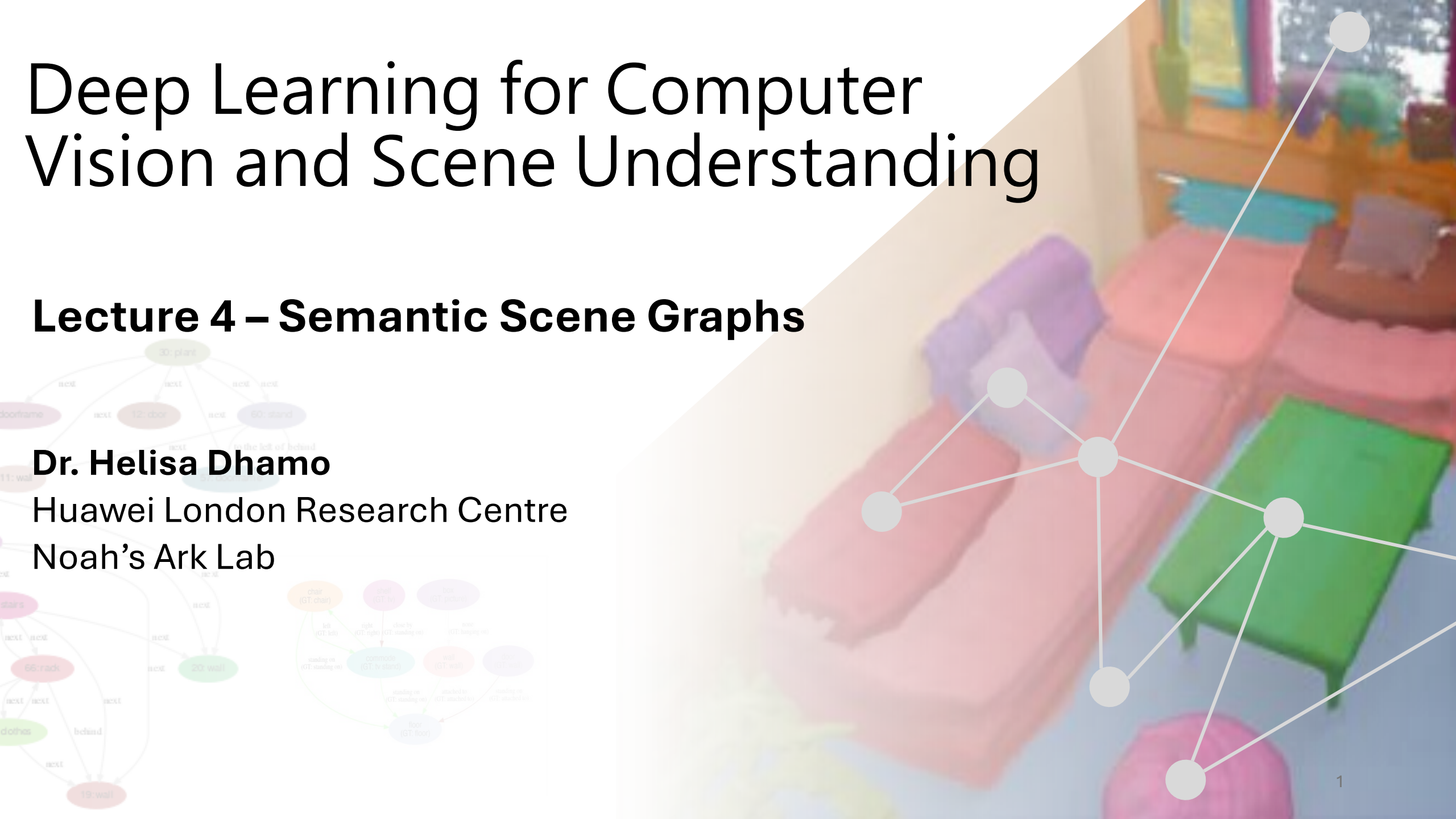
Deep Learning for Computer Vision and Scene Understanding

Lecture 4 – Semantic Scene Graphs

Dr. Helisa Dhamo
Huawei London Research Centre
Noah's Ark Lab

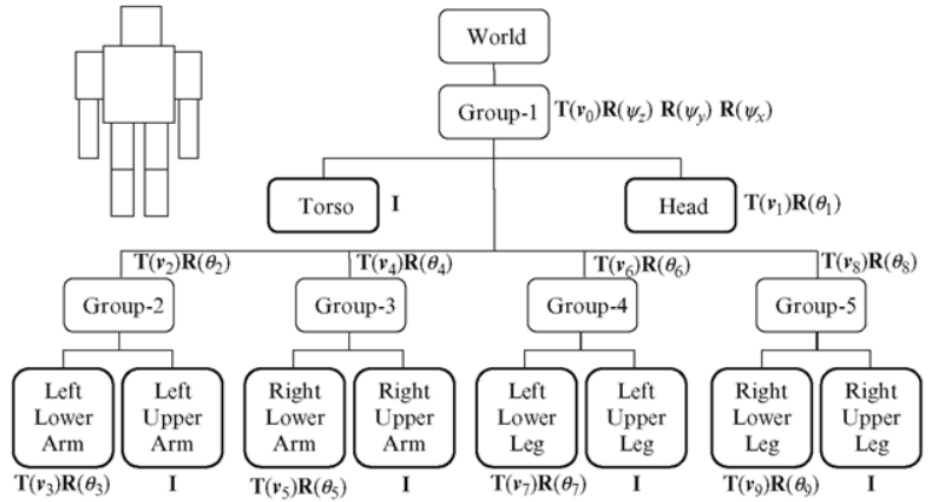


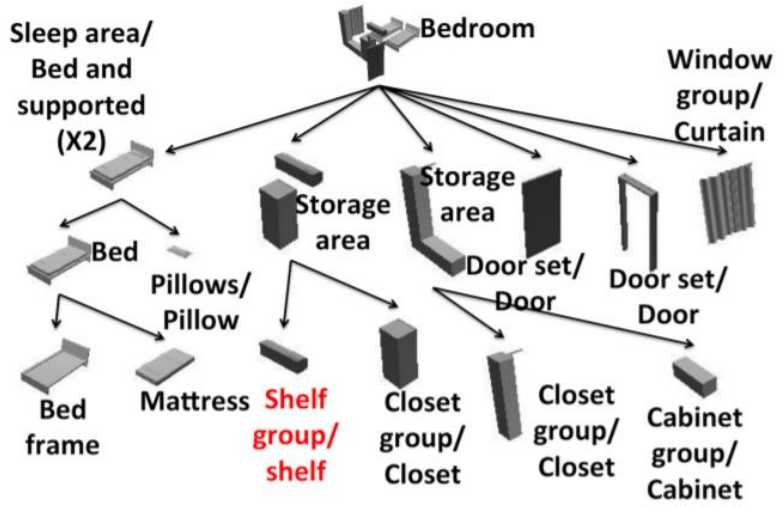
Invited lecture: Novel View Synthesis

Semantic Scene Graphs

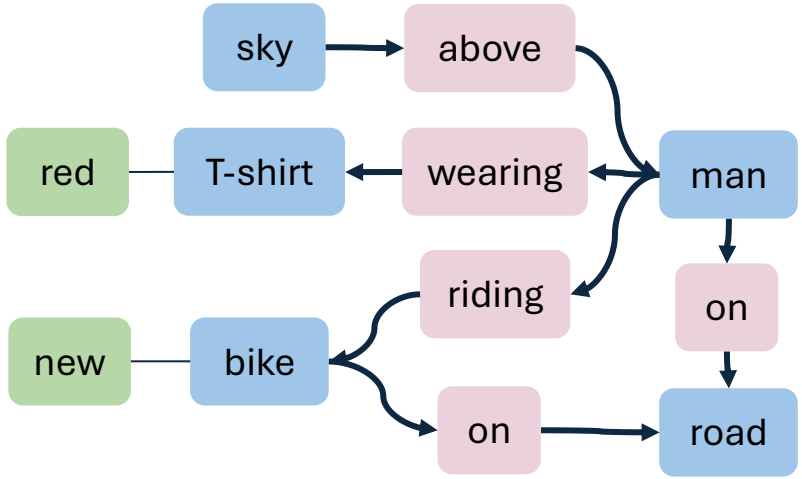
Lecture 4

Scene graph





Scene hierarchy [Liu TOG 2014]



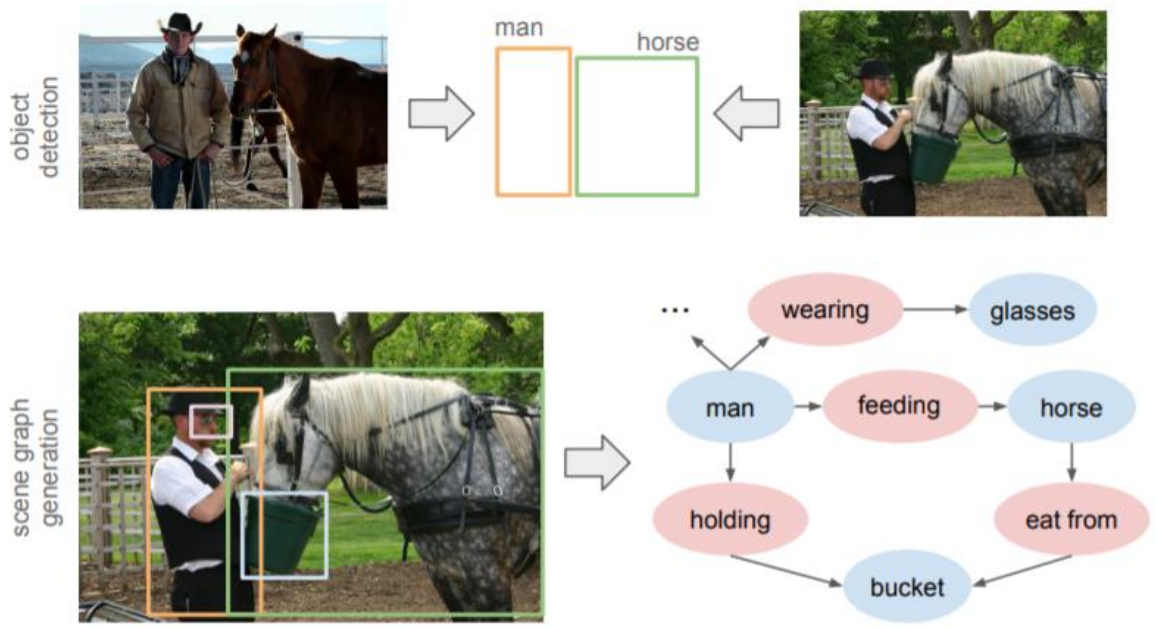
Semantic nodes and edges [Johnson CVPR 2015]

[1]Source: http://what-when-how.com/advanced-methods-in-computer-graphics/

(Recap) Scene Understanding Beyond Objects

Scene Graphs

- **Nodes:** objects in the scene
- **Edges:** relationships between objects (interaction, relative position)



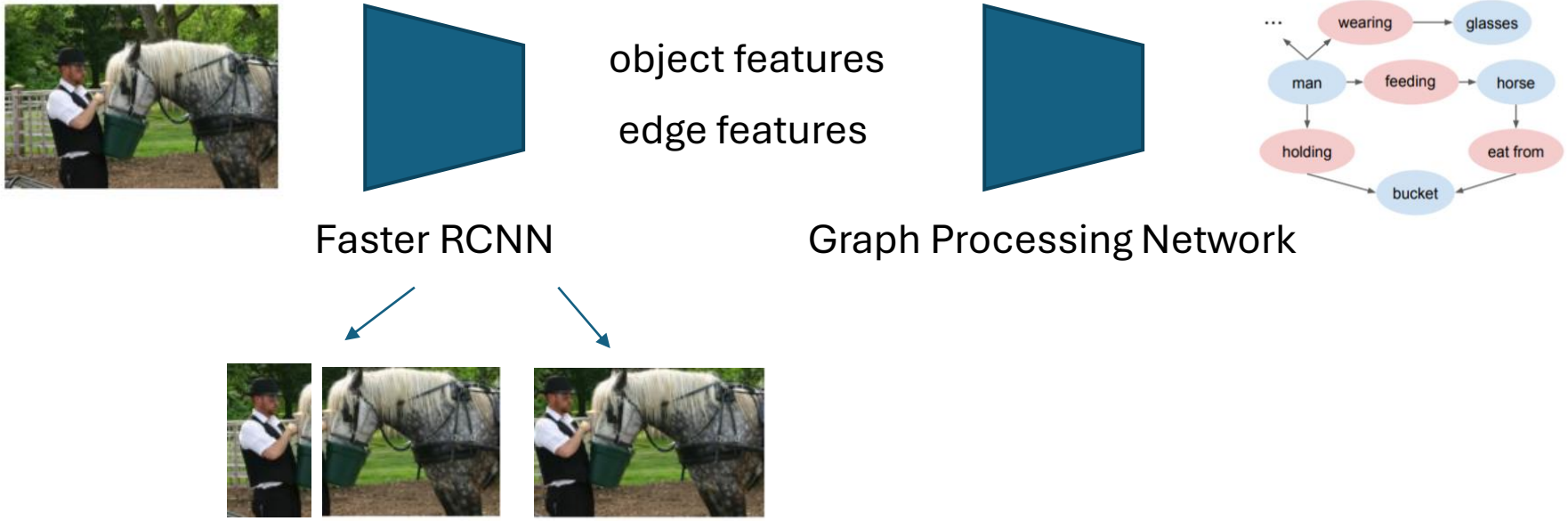
Xu et al., “Scene graph generation by iterative message passing.” CVPR’17

(Recap) Scene Understanding Beyond Objects

Scene Graphs

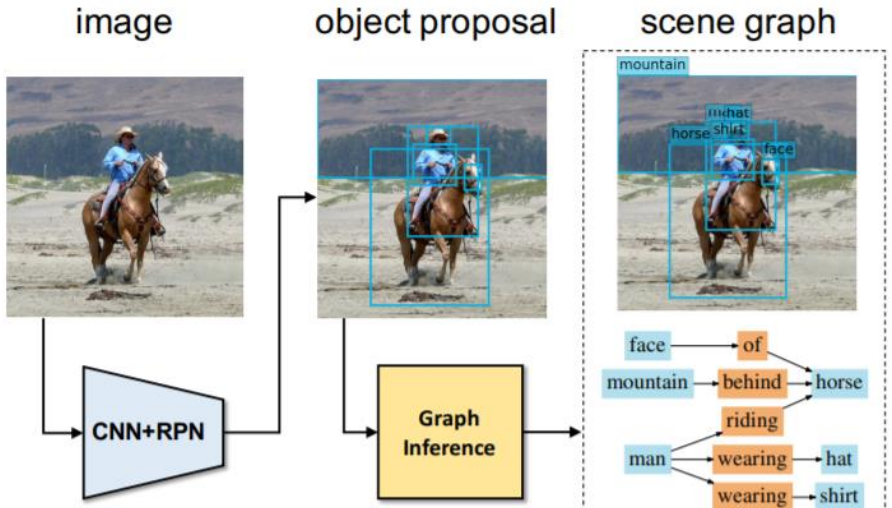
- **Nodes:** objects in the scene
- **Edges:** relationships between objects (interaction, relative position)

Scene graph generation networks are usually build on top of an object detector



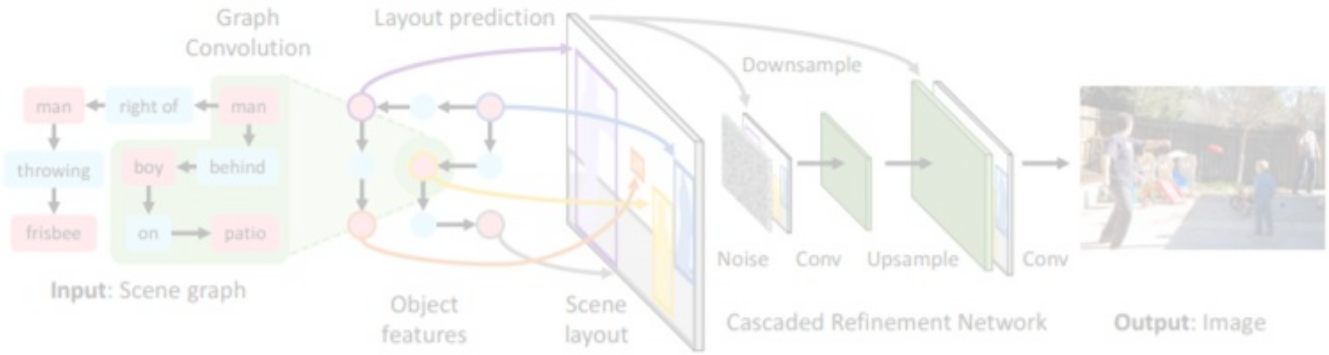
Semantic Scene Graphs and Images

From **image** to **scene graph**



[Xu CVPR 2017]

From **scene graph** to **image**



[Johnson CVPR 2018] Purely semantic nodes (object class)

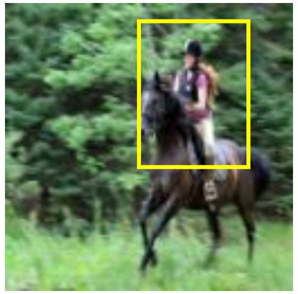
Scene Understanding

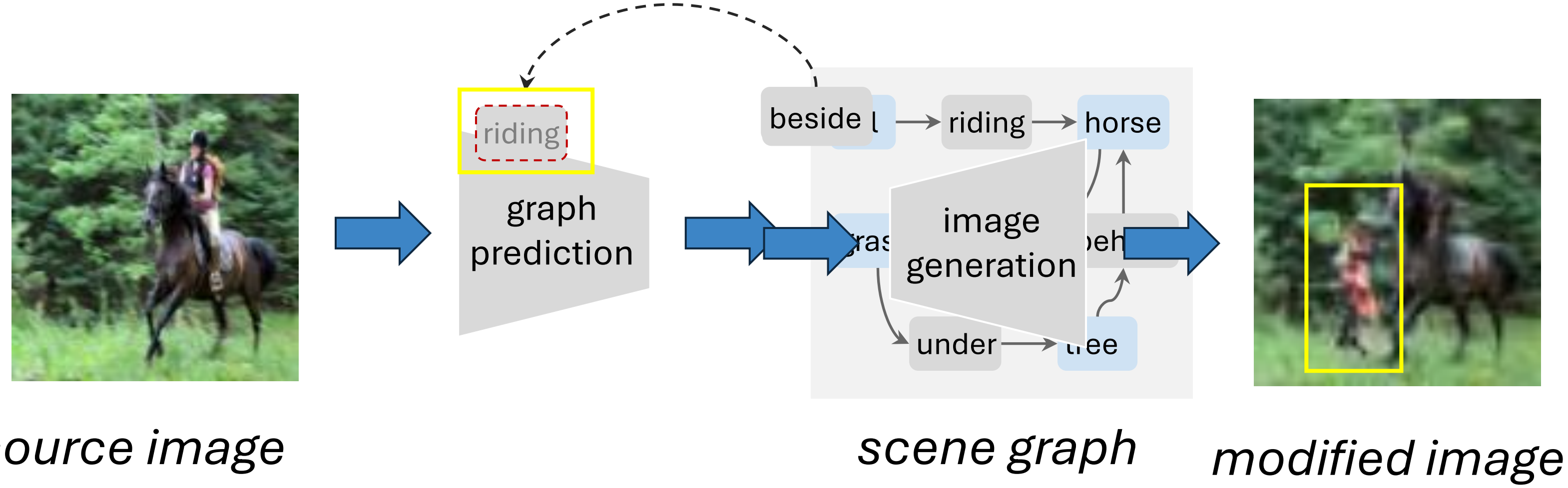
Scene Synthesis

J. Johnson, A. Gupta, and FF. Li. Image generation from scene graphs. **CVPR 2018**.

D. Xu, Y. Zhu, C.r Choi, FF. Li. Scene Graph Generation by Iterative Message Passing. **CVPR 2017**

Semantic Image Editing The combined problem…



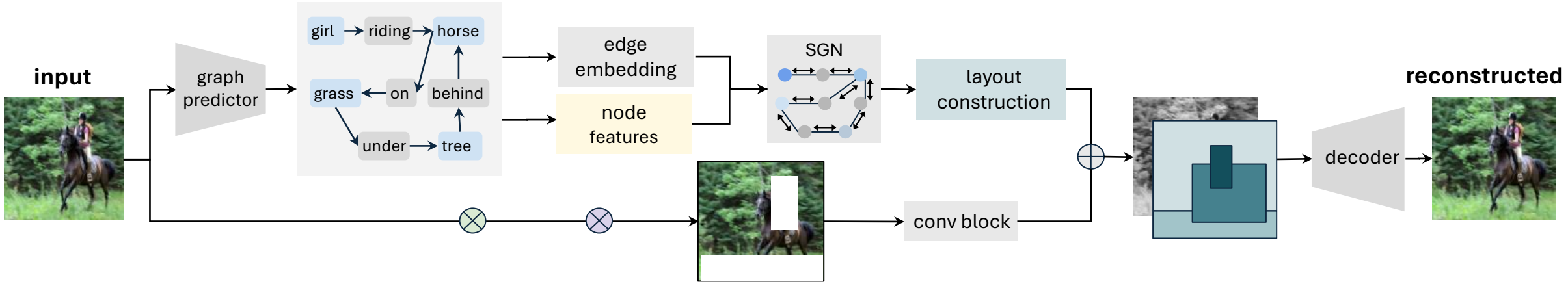


Challenge: No real image pairs with changes!

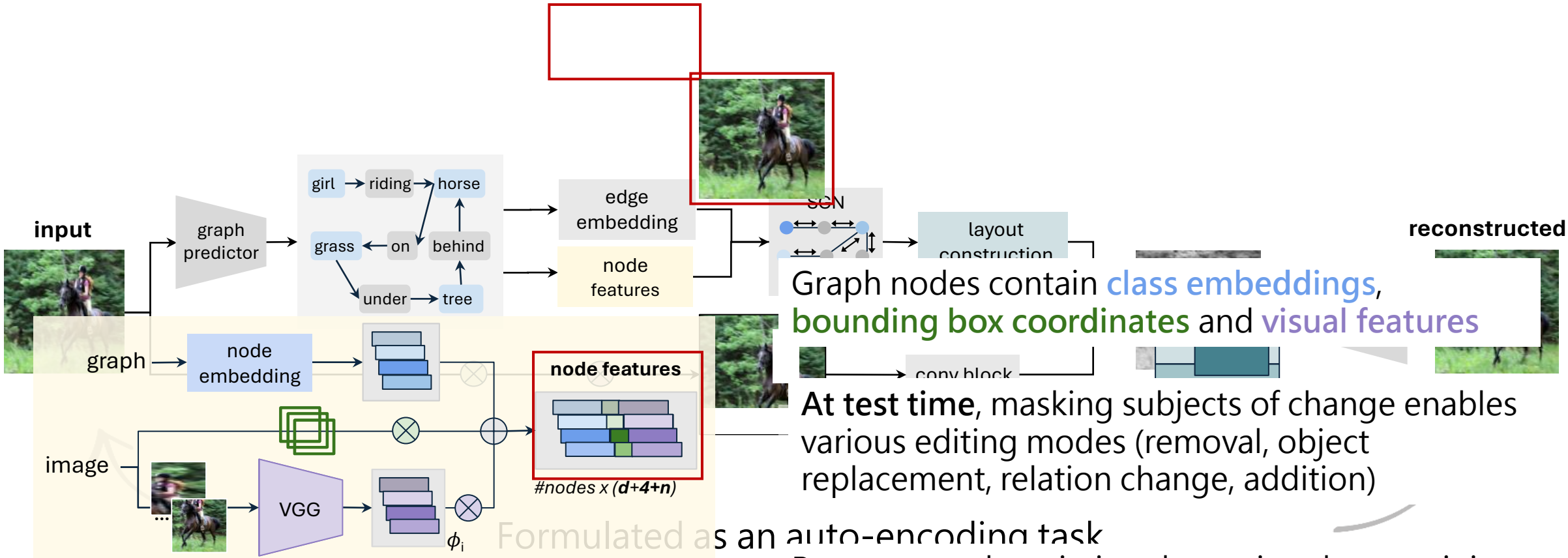
at test time

Semantic Image Editing Training

Our training strategy does not require pairs for the editing task



at training time



Graph nodes contain class embeddings, bounding box coordinates and visual features

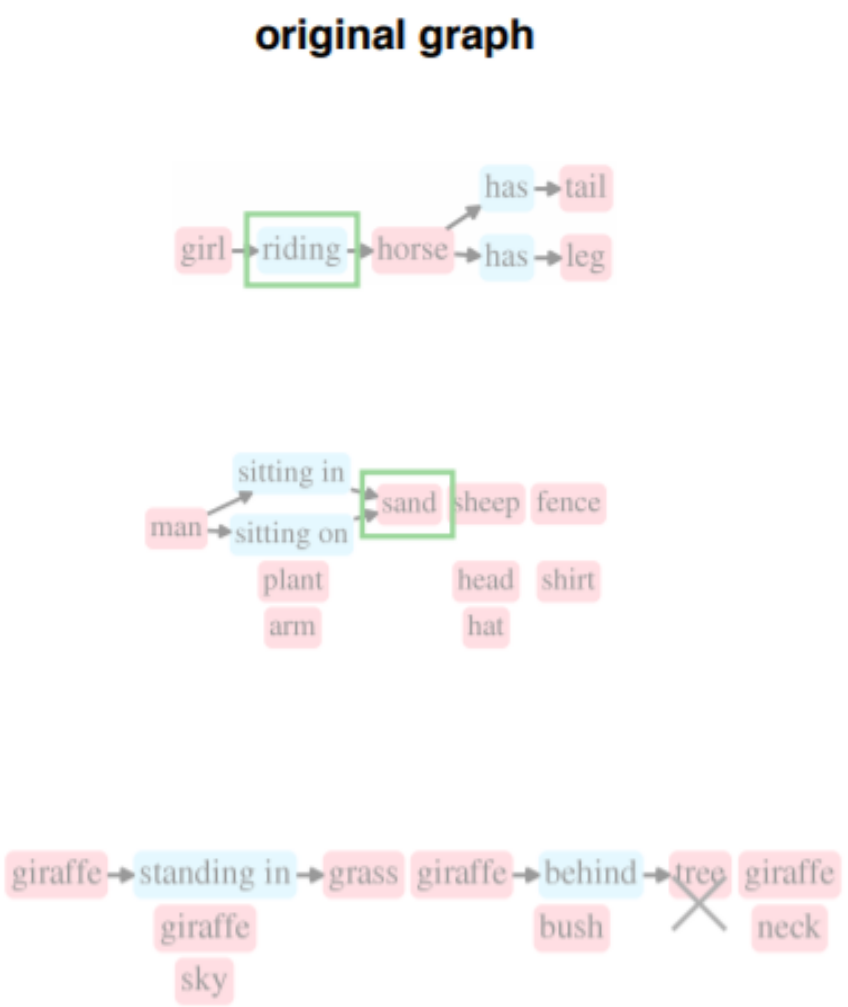
At test time, masking subjects of change enables various editing modes (removal, object replacement, relation change, addition)

Formulated as an auto-encoding task

Reconstruct the missing data using the remaining parts of the graph

at training time

Semantic Image Editing Results



relationship change

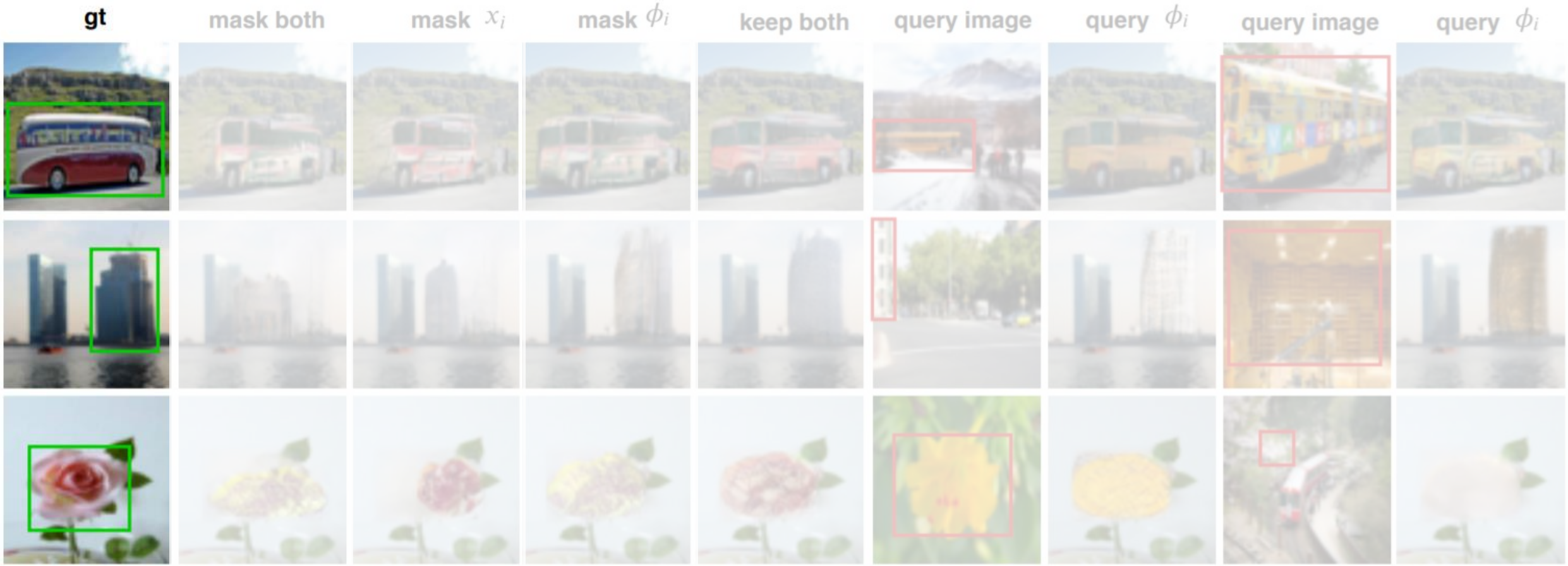
object replacement

object removal

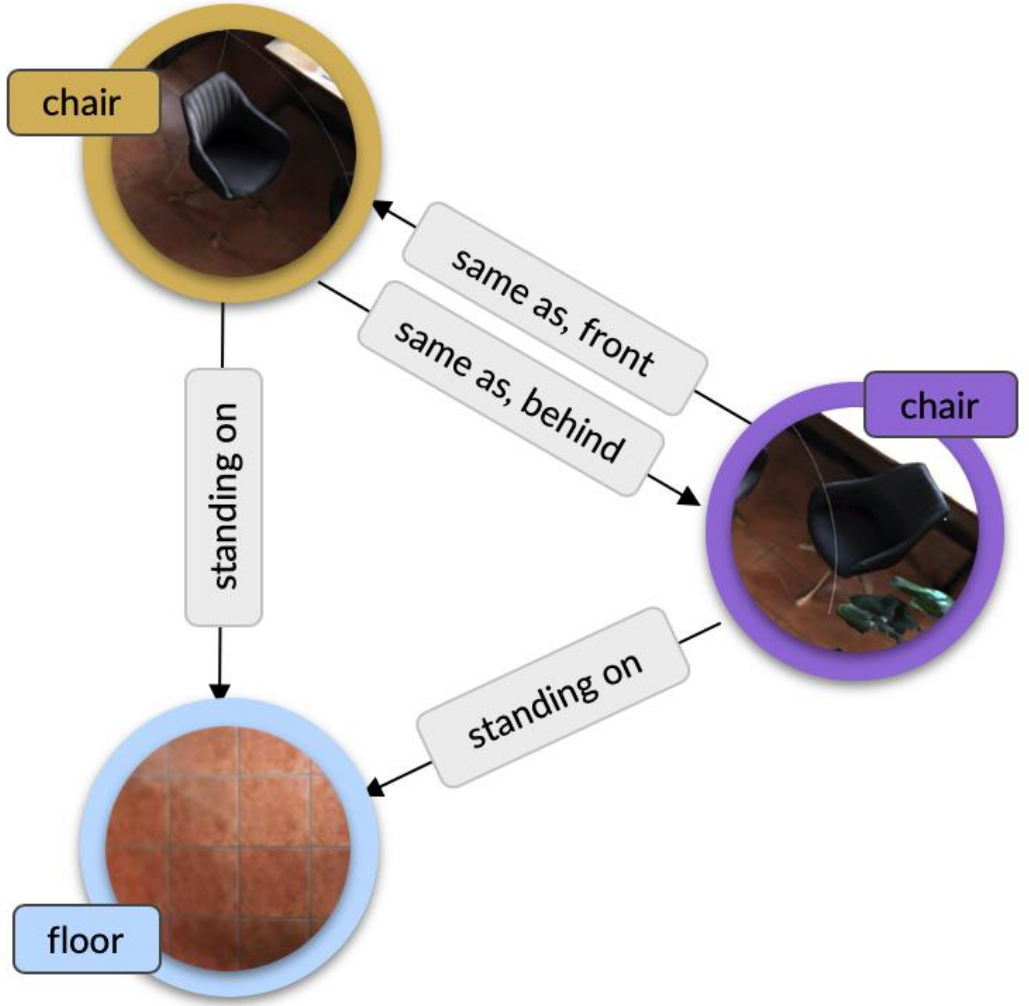
Krishna,..., FF Li. Visual genome: Connecting language and vision using crowdsourced dense image annotations. **IJCV 2017**

Results on **Visual Genome**

Semantic Image Editing Results



3D Semantic Scene Graphs



J. Wald*, **H. Dhamo***, N. Navab, F. Tombari. Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions. **CVPR 2020**

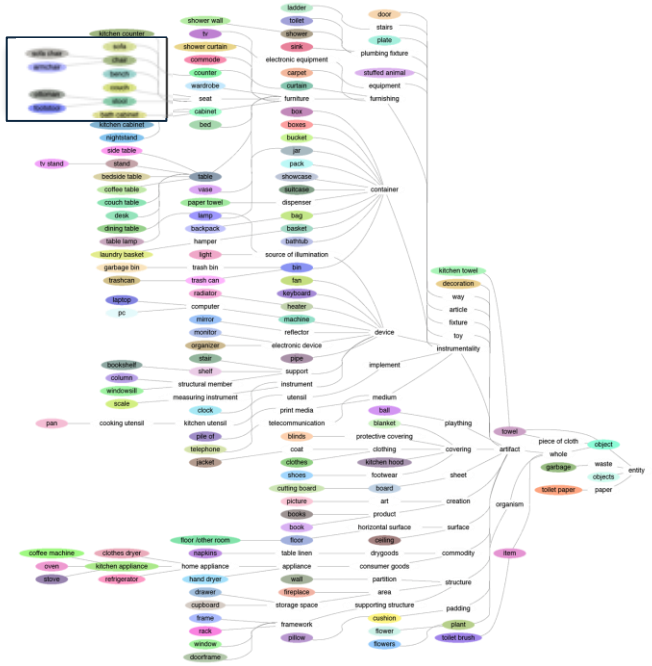

3DSSG
3D Semantic Scene Graphs

3DSSG Dataset

Based on 3RScan Dataset [Wald 2019]

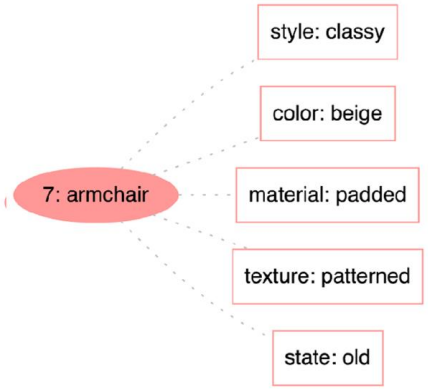
Nodes

Hierarchy of semantic class labels



Attributes

Static (color, material, shape) and **dynamic** attributes (tidy/messy, open/closed) and **affordances** (sitting, eating)



Relationships

Proximity (left/right, front/behind, close by)
Support (lying in, hanging, leaning against)
Comparative (smaller than, same as)

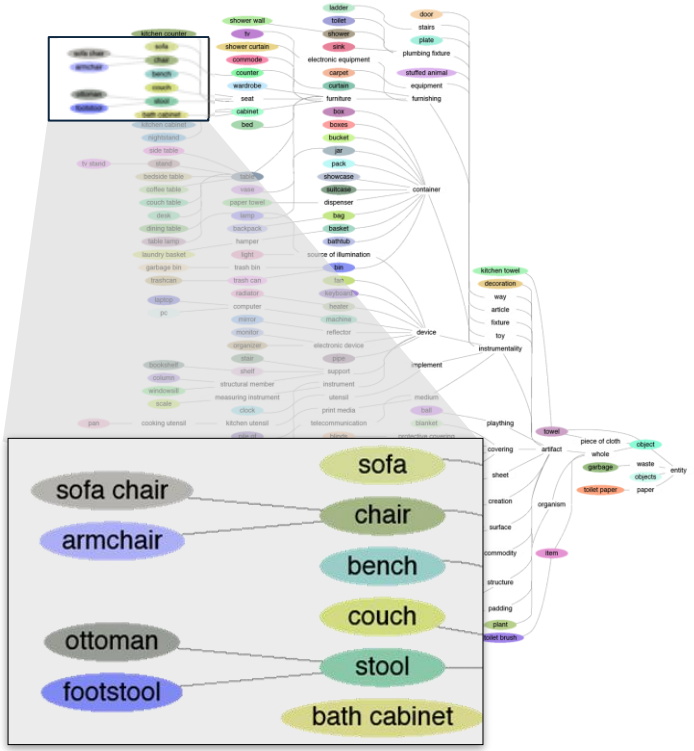
J. Wald*, **H. Dhamo***, N. Navab, F. Tombari. Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions. **CVPR 2020**
J. Wald, A. Avetisyan, N. Navab, F. Tombari*, M. Nießner*. RIO: 3D Object Instance Re-Localization in Changing Indoor Environments. **ICCV 2019**

3DSSG Dataset

Based on 3RScan [Wald 2019]

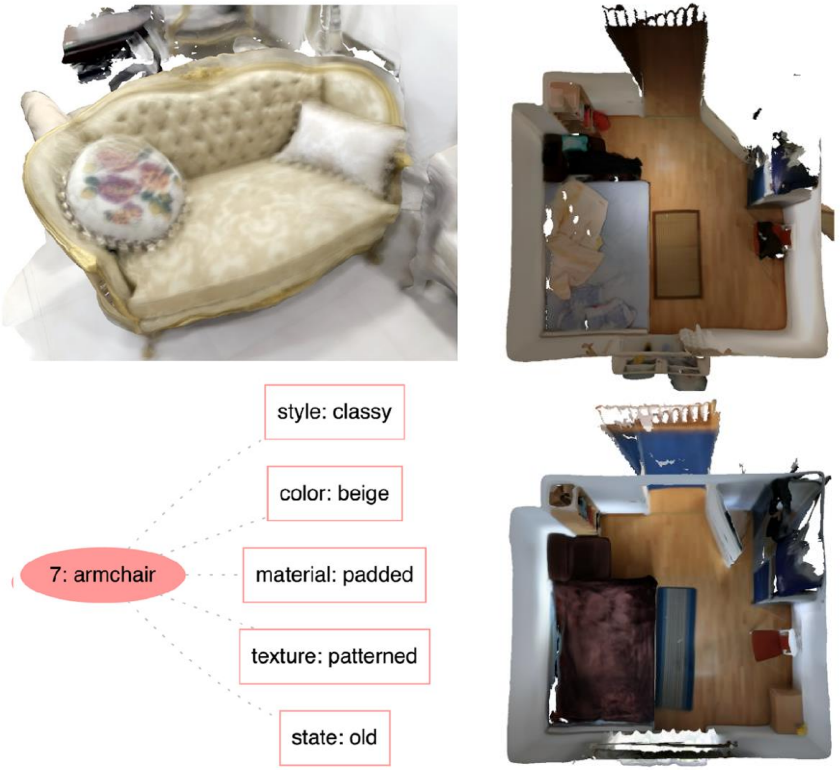
Nodes

Hierarchy of semantic class labels



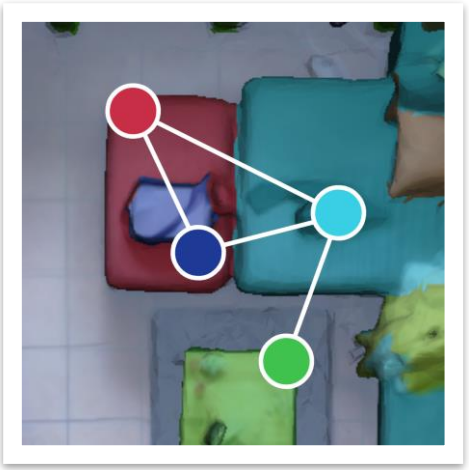
Attributes

Static (color, material, shape) and **dynamic** attributes (tidy/messy, open/closed) and **affordances** (sitting, eating)



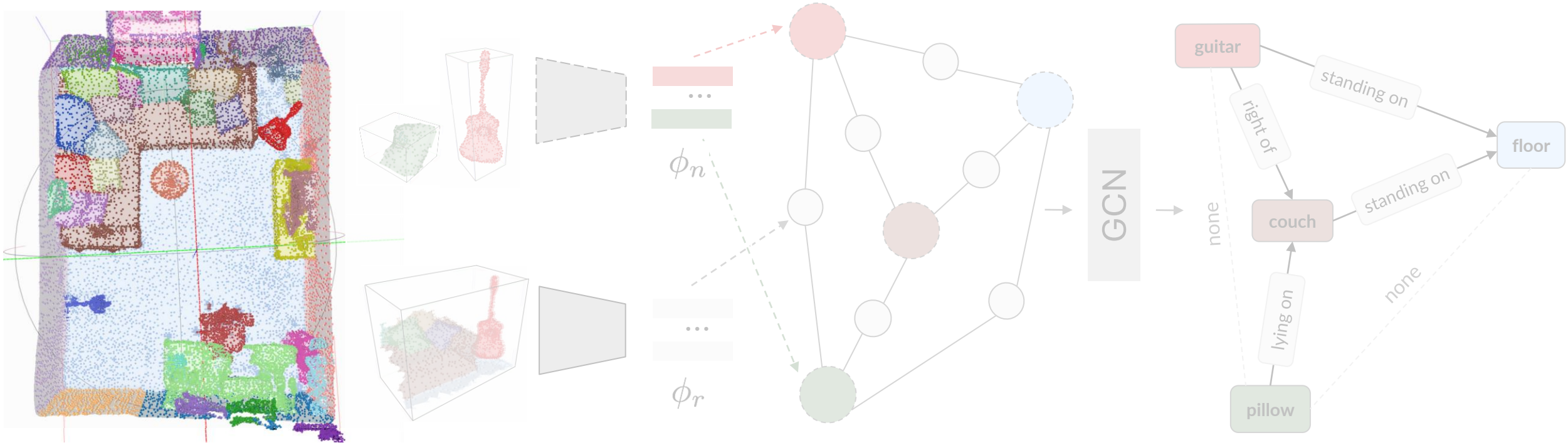
Relationships

Proximity (left/right, front/behind, close by)
Support (lying in, hanging, leaning against)
Comparative (smaller than, same as)



J. Wald*, **H. Dhamo***, N. Navab, F. Tombari. Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions. **CVPR 2020**
J. Wald, A. Avetisyan, N. Navab, F. Tombari*, M. Nießner*. RIO: 3D Object Instance Re-Localization in Changing Indoor Environments. **ICCV 2019**

Learning 3D Semantic Scene Graphs

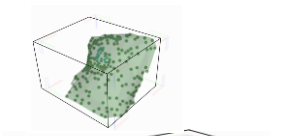


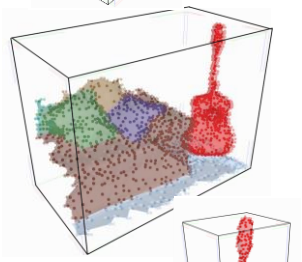
$$\mathcal{L}_{\text{total}} = \lambda_{obj}\mathcal{L}_{\text{obj}} + \mathcal{L}_{\text{pred}} \qquad \mathcal{L} = -\alpha_t(1-p_t)^{\gamma}\log p_t$$

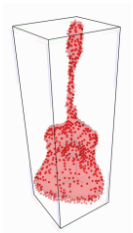
none or multiple predicate predictions per edge

J. Wald*, **H. Dhamo***, N. Navab, F. Tombari. Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions. **CVPR 2020**

Network Details - one GraphConvNet layer







...

...

MLP1

MLP1

propagate features through triplets

aggregate all features of the **same** node

MLP2

MLP2

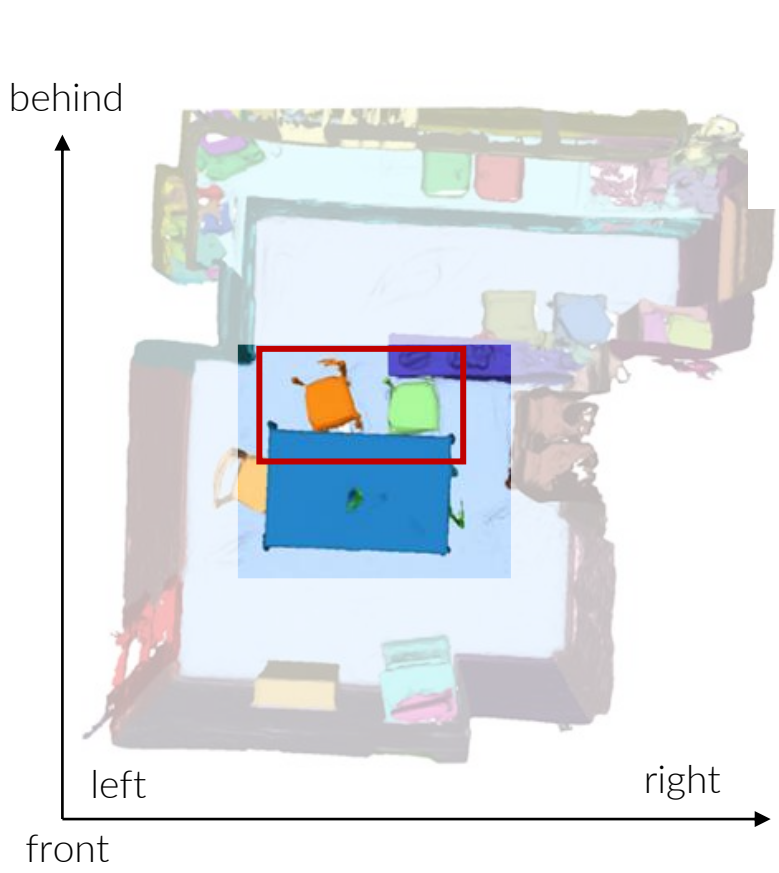
MLP2

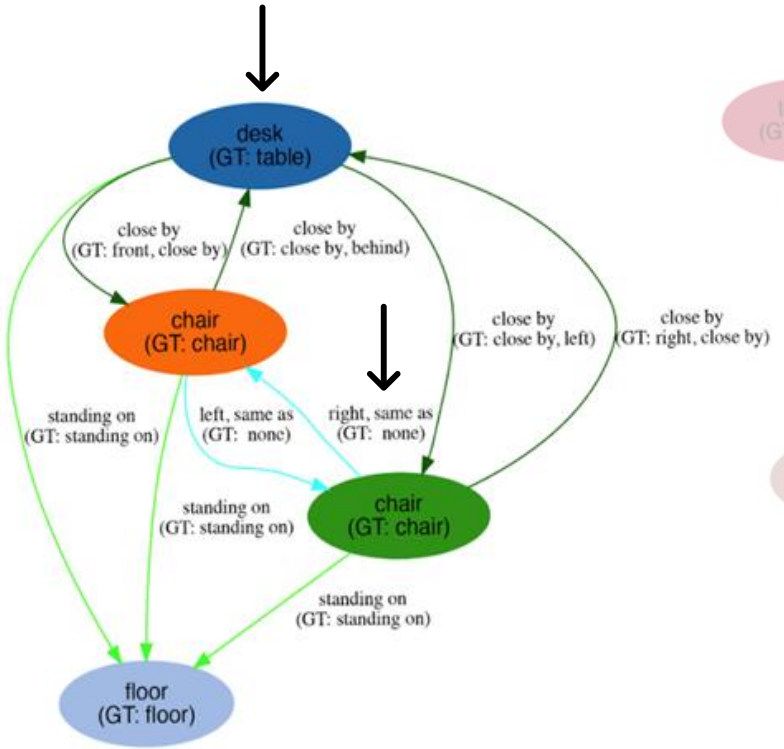
update each node

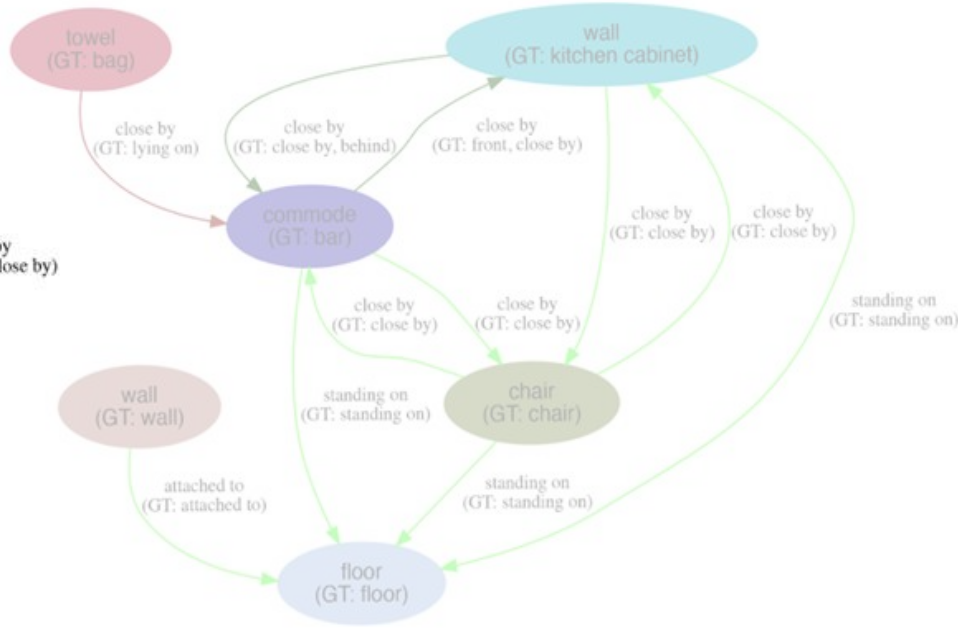
Use to populate graph - ready for next layer

Output node and edge features

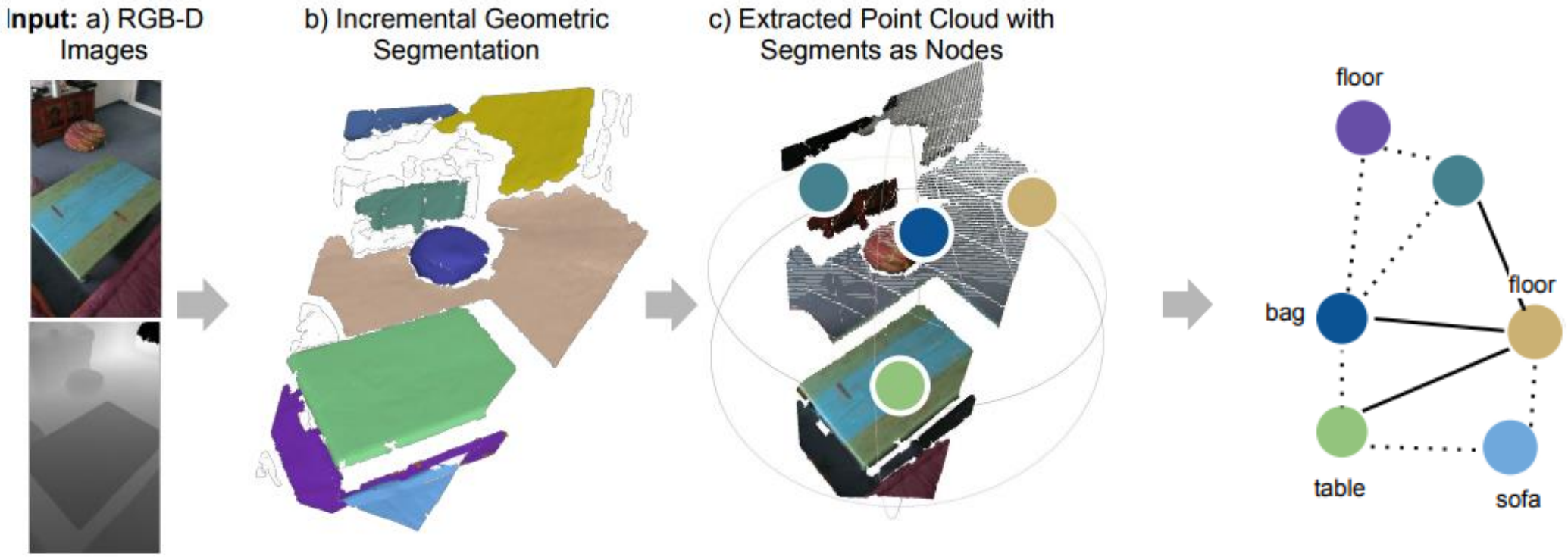
Learning 3D Semantic Scene Graphs Results







Follow-up research Towards real world requirements



Build scene graph incrementally as the scene is reconstructed

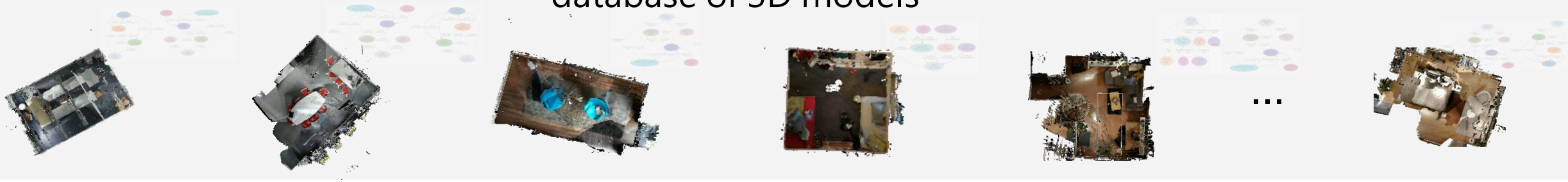
No need for class-agnostic instance segmentation

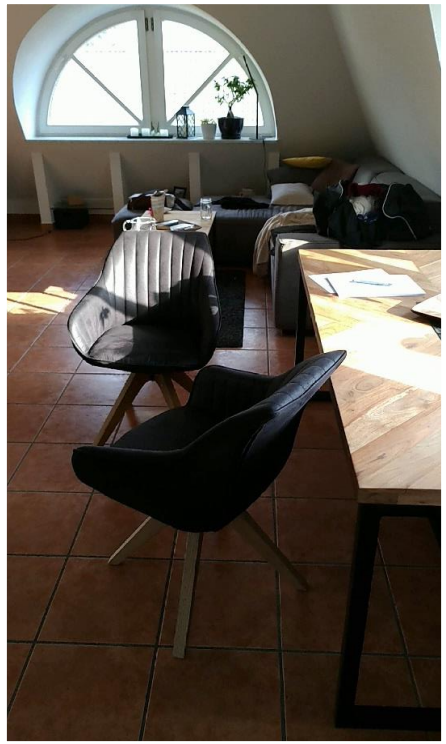
S. Wu, J. Wald, K. Tateno, N. Navab, F. Tombari. SceneGraphFusion: Incremental Scene Graph Prediction from RGB-D Sequences. **CVPR 2021**

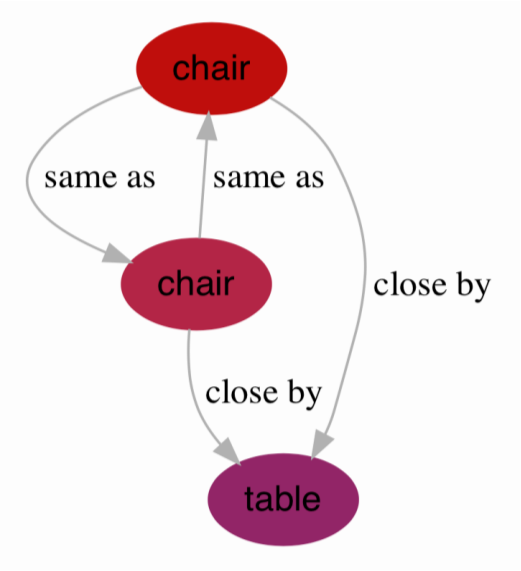
Applications

Domain agnostic scene retrieval

database of 3D models



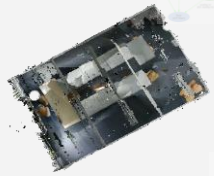




query photo and its scene graph

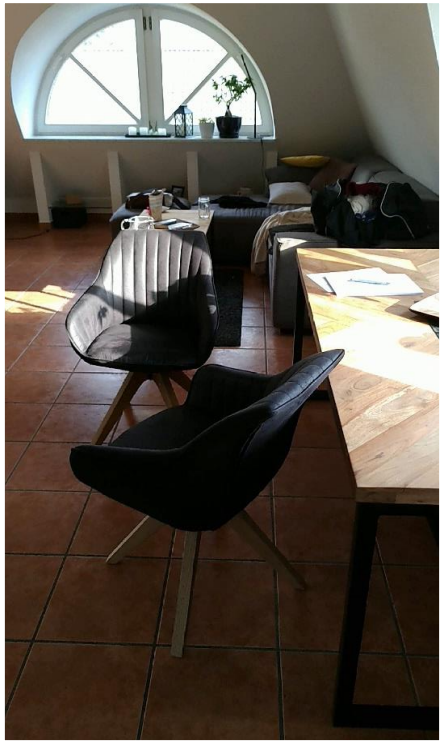
J. Wald*, **H. Dhamo***, N. Navab, F. Tombari. Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions. **CVPR 2020**

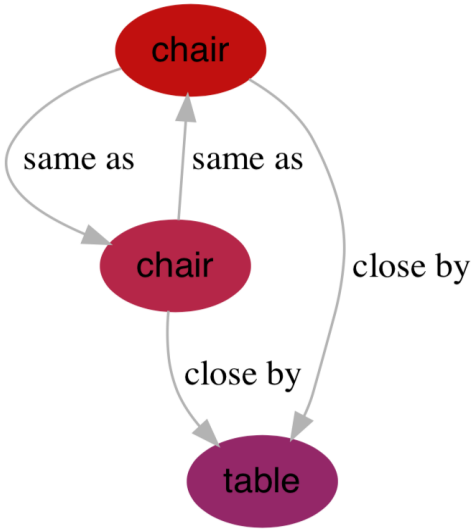
database of 3D models





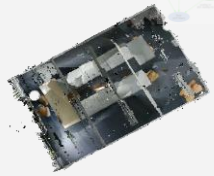

...



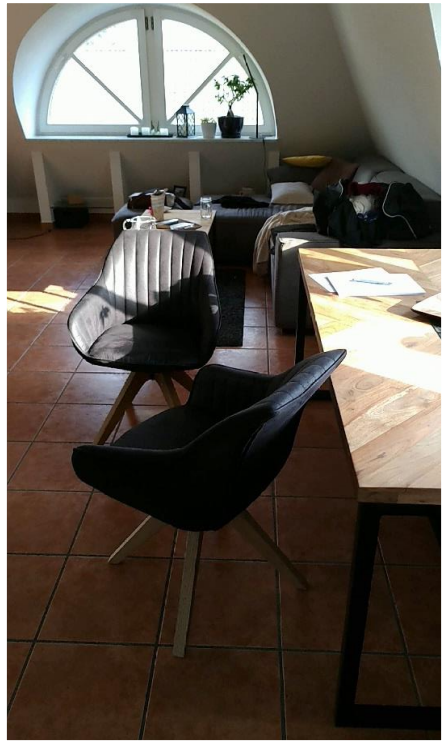
query photo and its scene graph

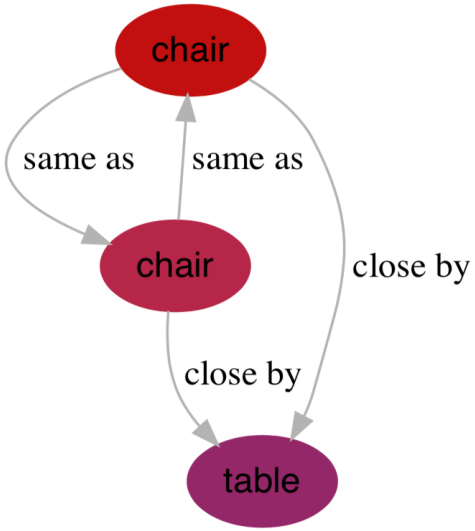
database of 3D models





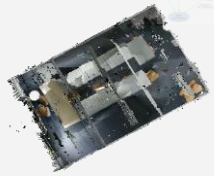

...



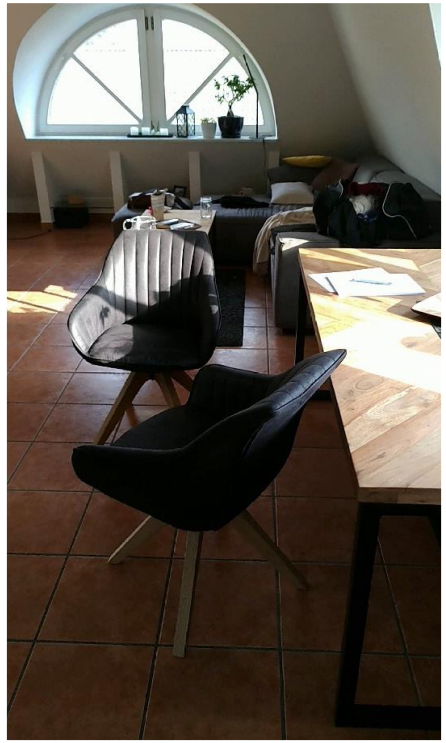
query photo and its scene graph

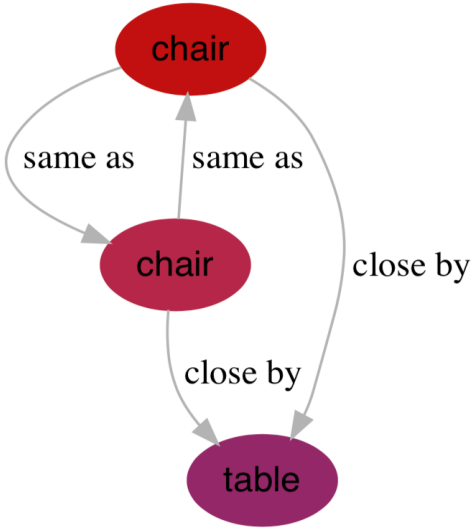
database of 3D models






...

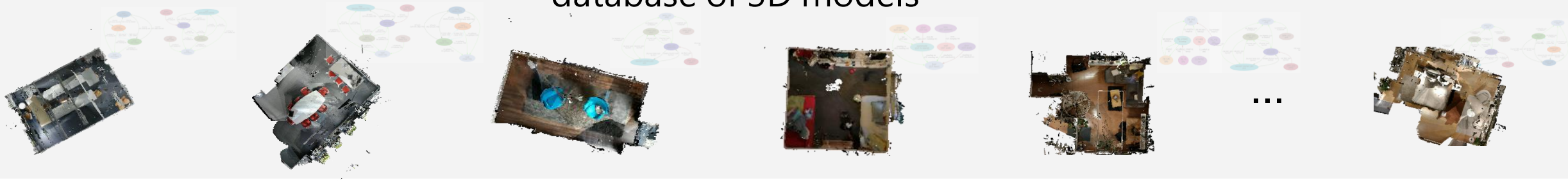


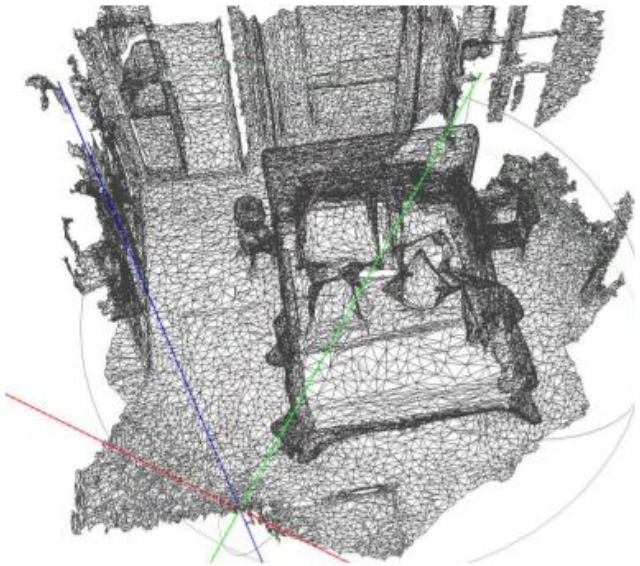
Query based on node set and triplet set

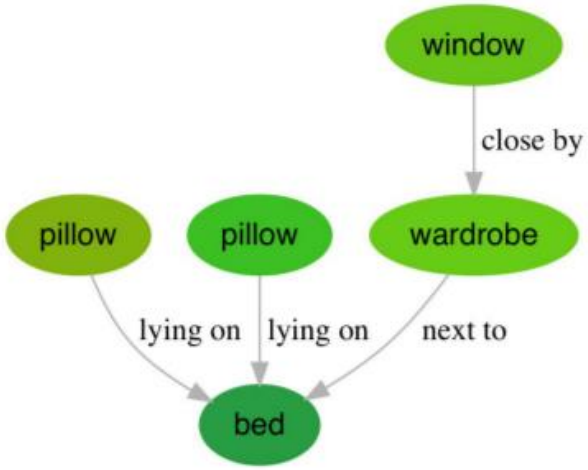
query photo and its scene graph

best match

database of 3D models





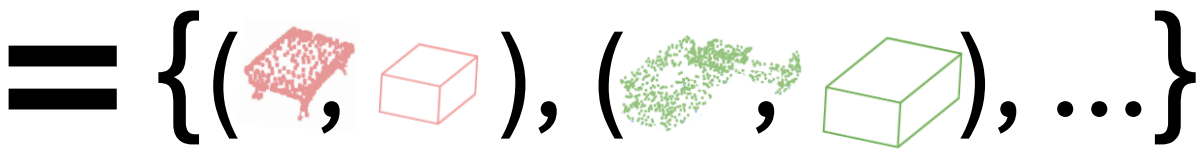


query 3D scene and its scene graph

best match

Applications

Scene generation and editing



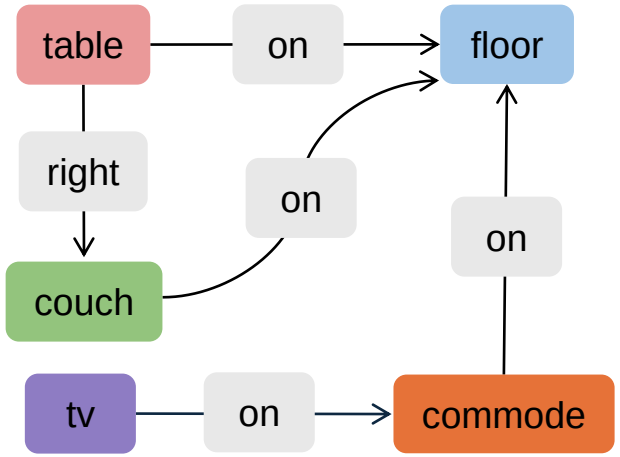
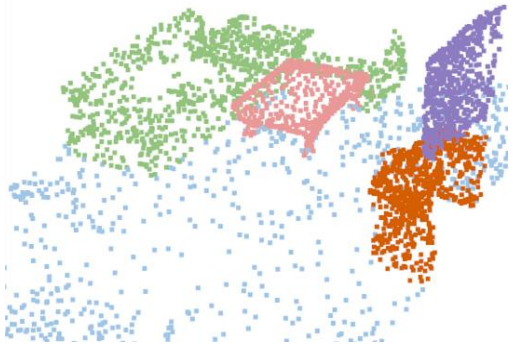
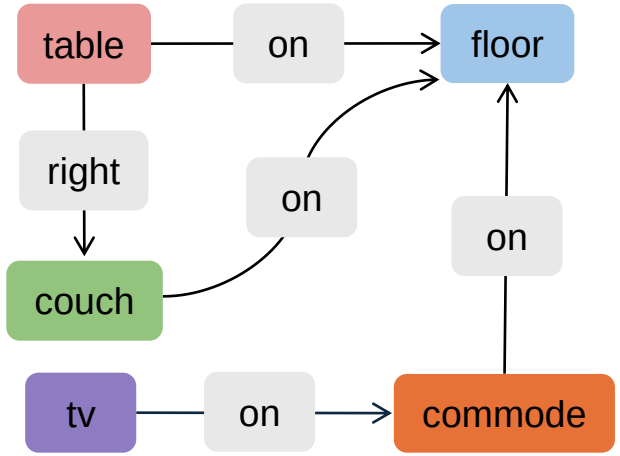
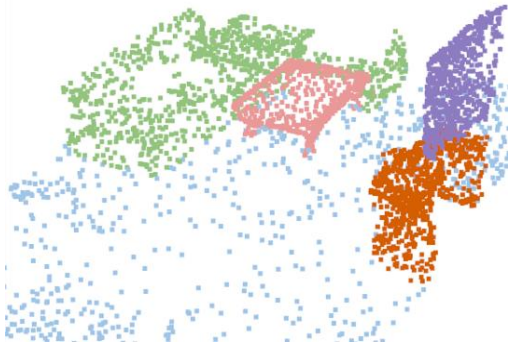


table
on
floor
right
on
on
couch
tv
on
commode



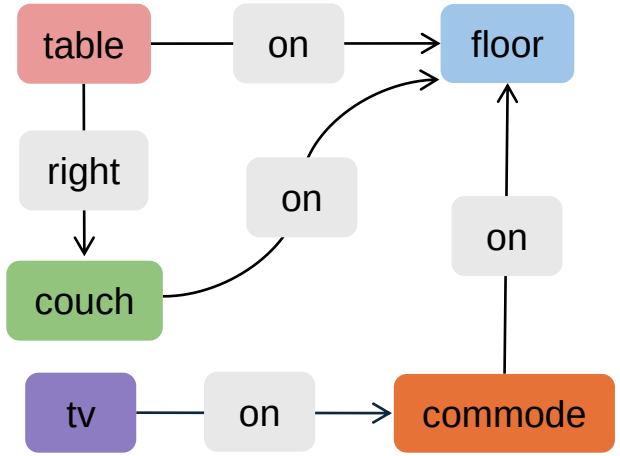


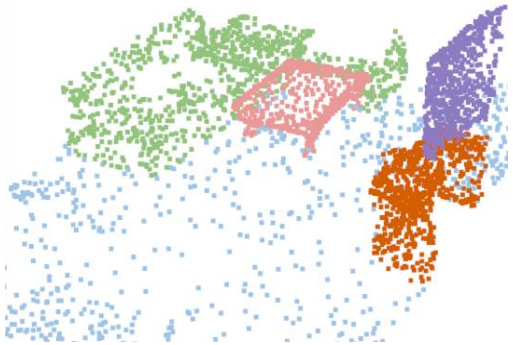


$= \{(\,,\,), (\,,\,), \ldots\}$

$\mathcal{E}_{\text{gen}}$

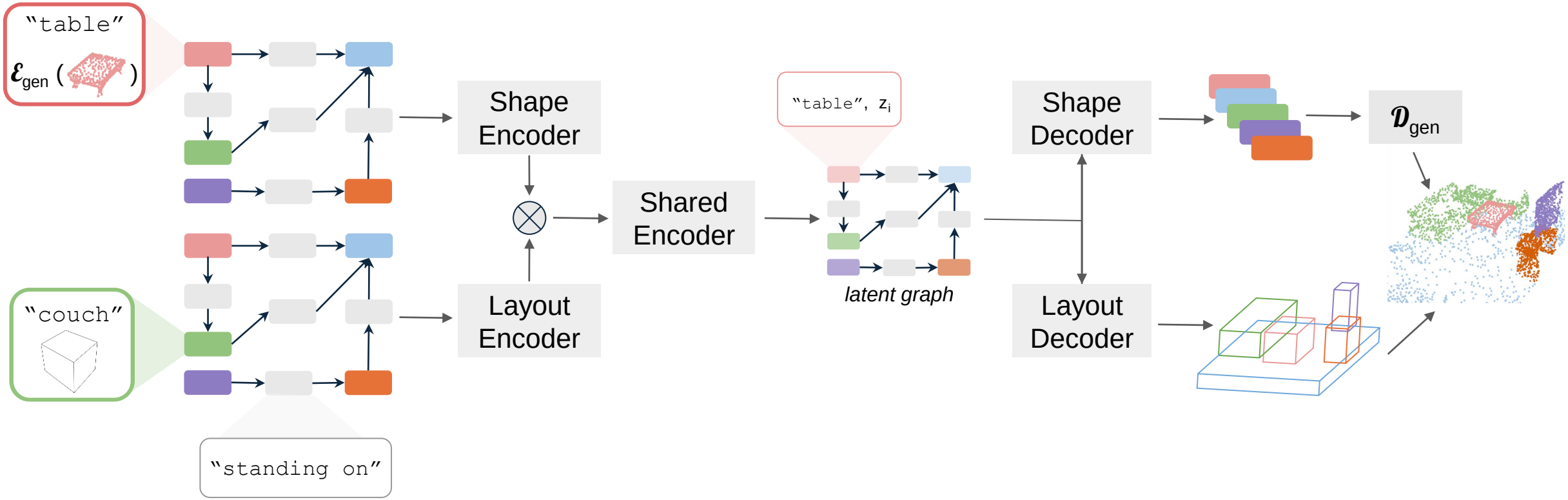
H. Dhamo*, F. Manhardt*, N. Navab, F. Tombari. Graph-to-3D: End-to-end Generation and Manipulation of 3D Scenes Using Scene Graphs. **ICCV 2021**

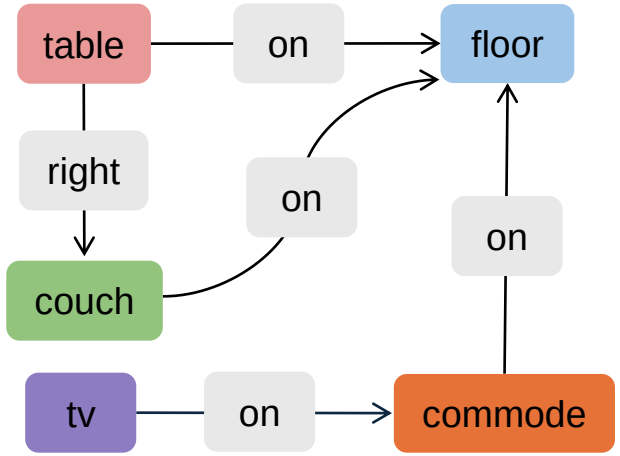


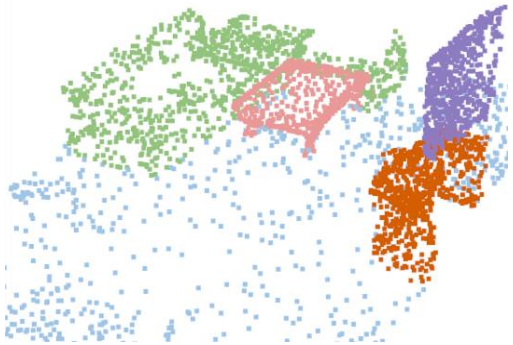


$= \{(\,,\,), (\,,\,), \ldots\}$

$\mathcal{E}_{\text{gen}}$

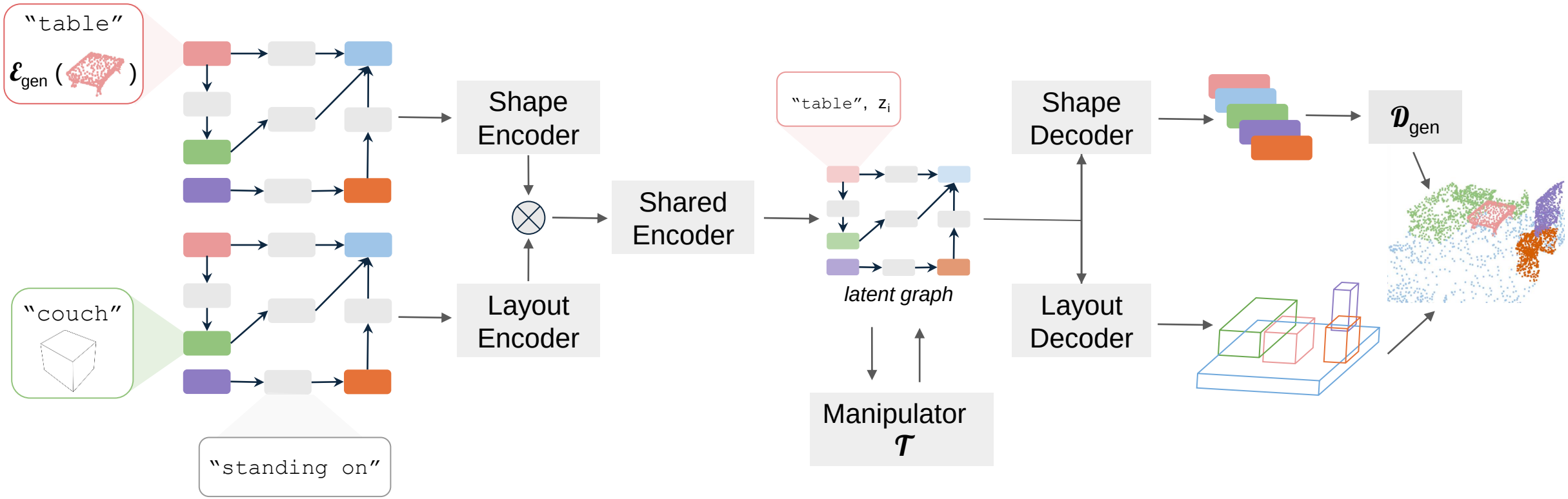




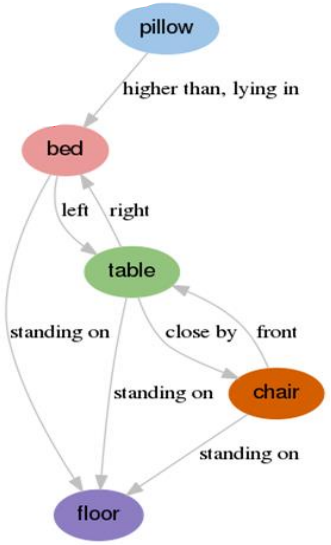


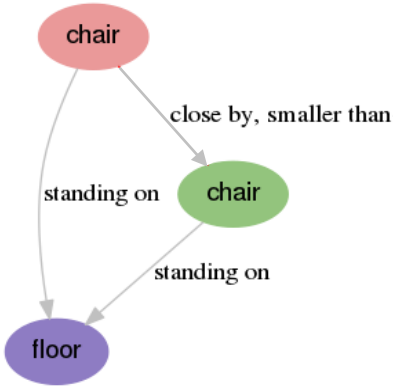
$= \{(\,,\,), (\,,\,), \ldots\}$

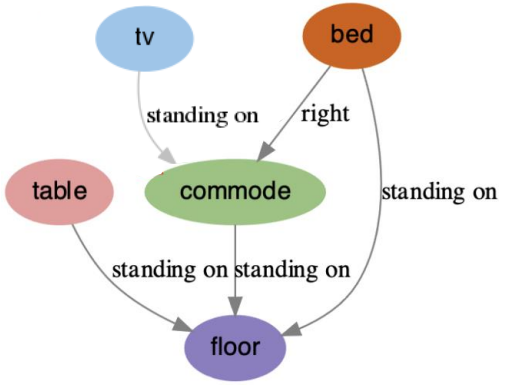
$\mathcal{E}_{\text{gen}}$

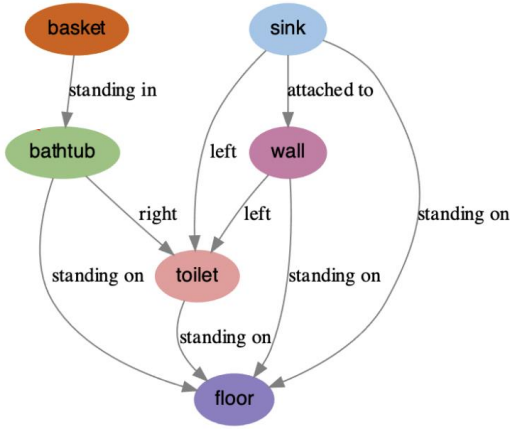


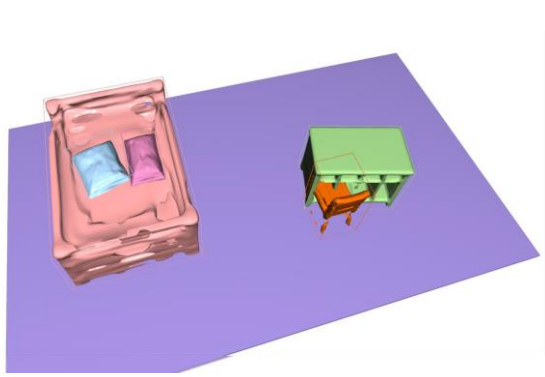
Graph-to-3D Results

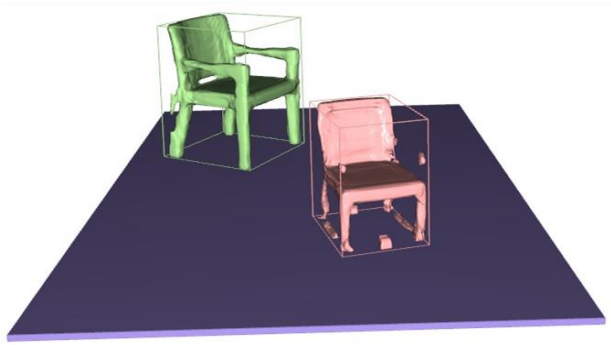




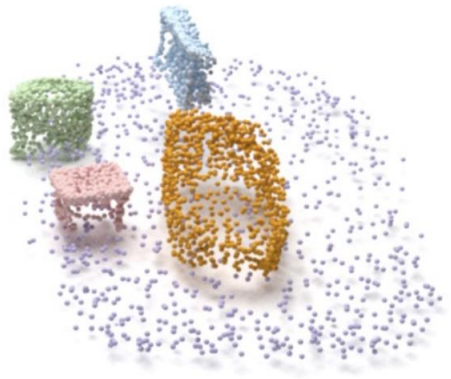


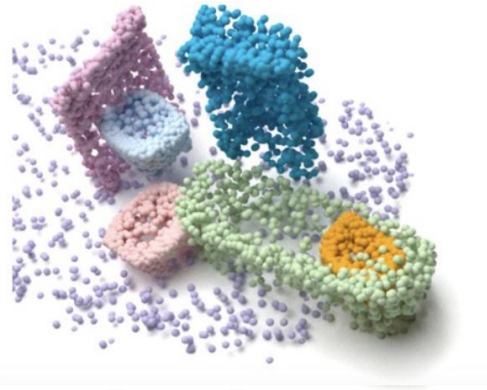






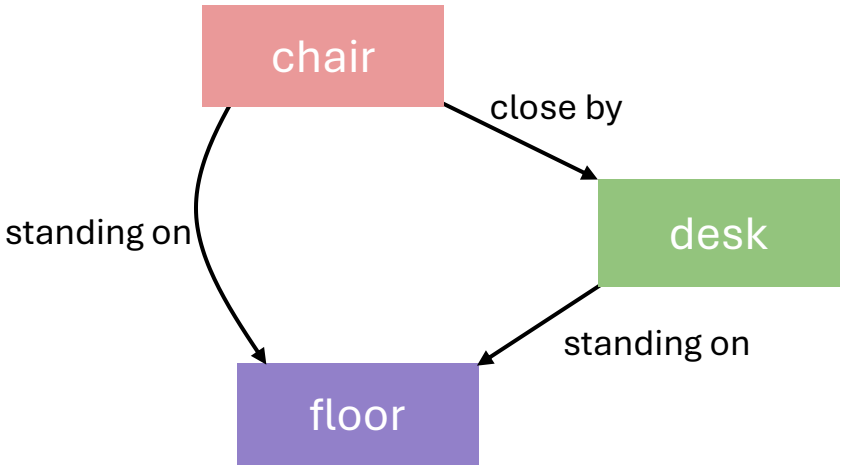
Meshes from DeepSDF encoding

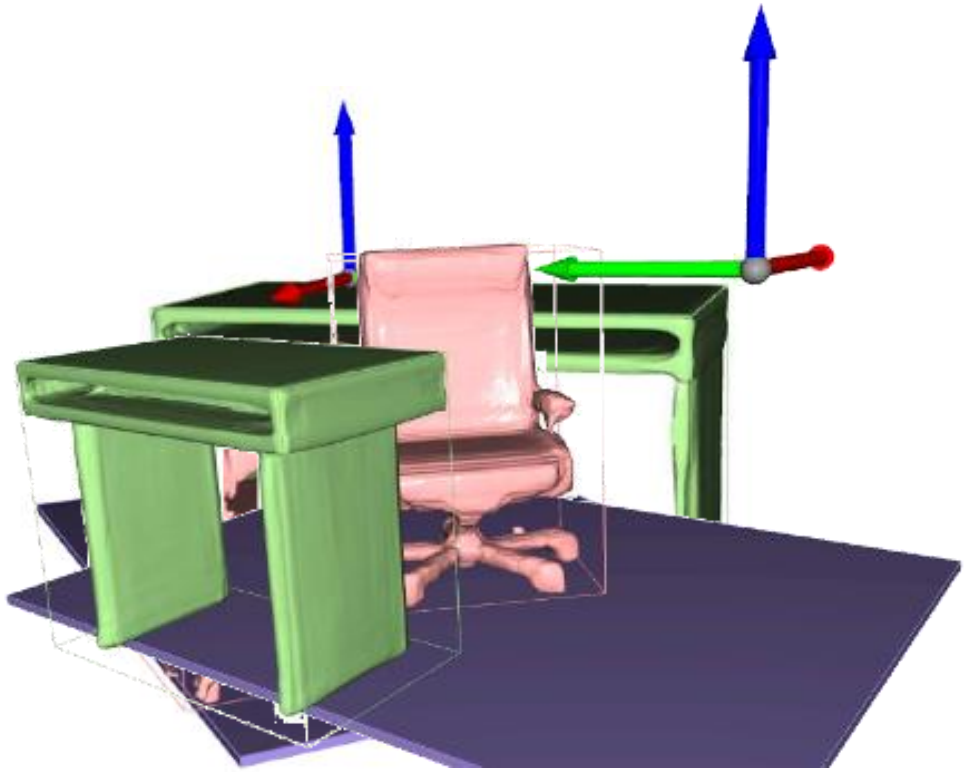




Point clouds from AtlasNet encoding

Graph-to-3D Context learning results

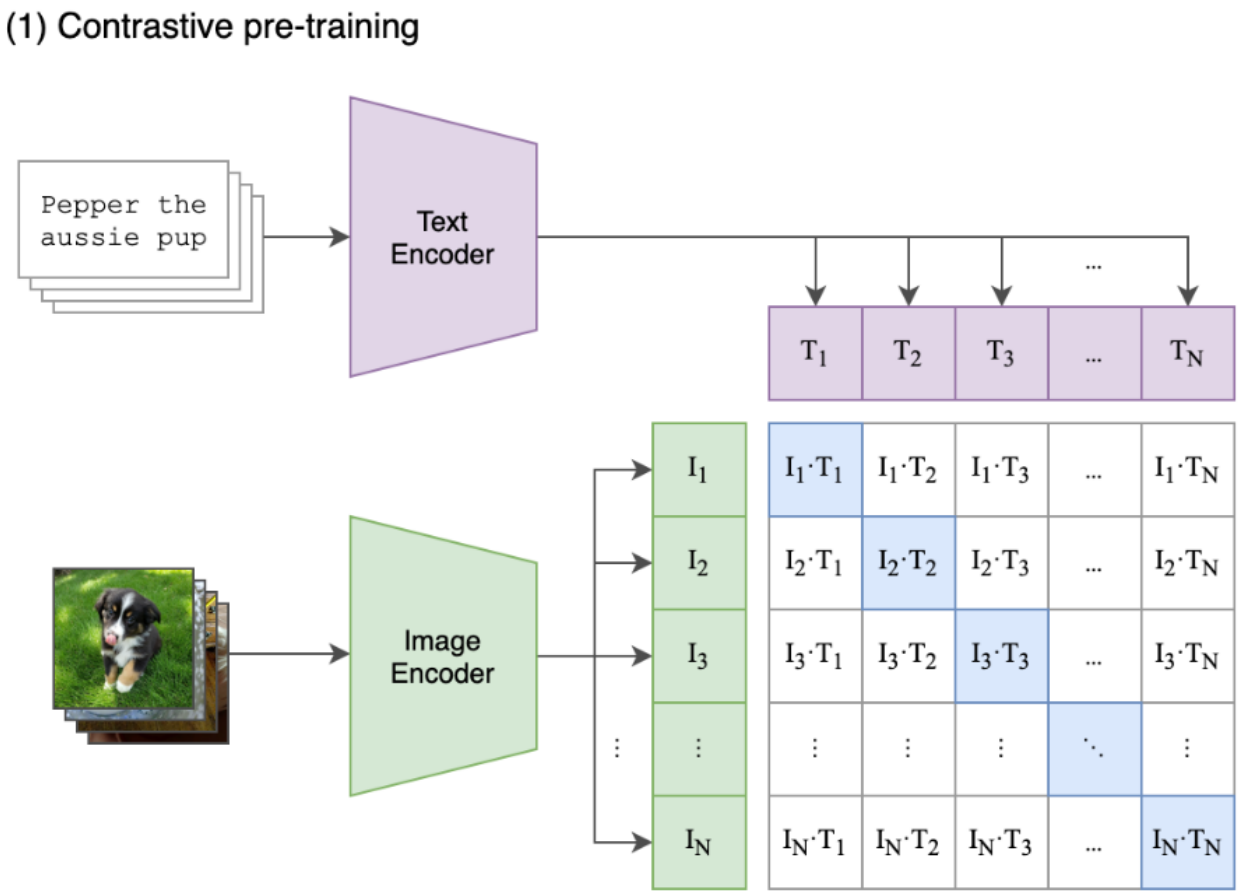


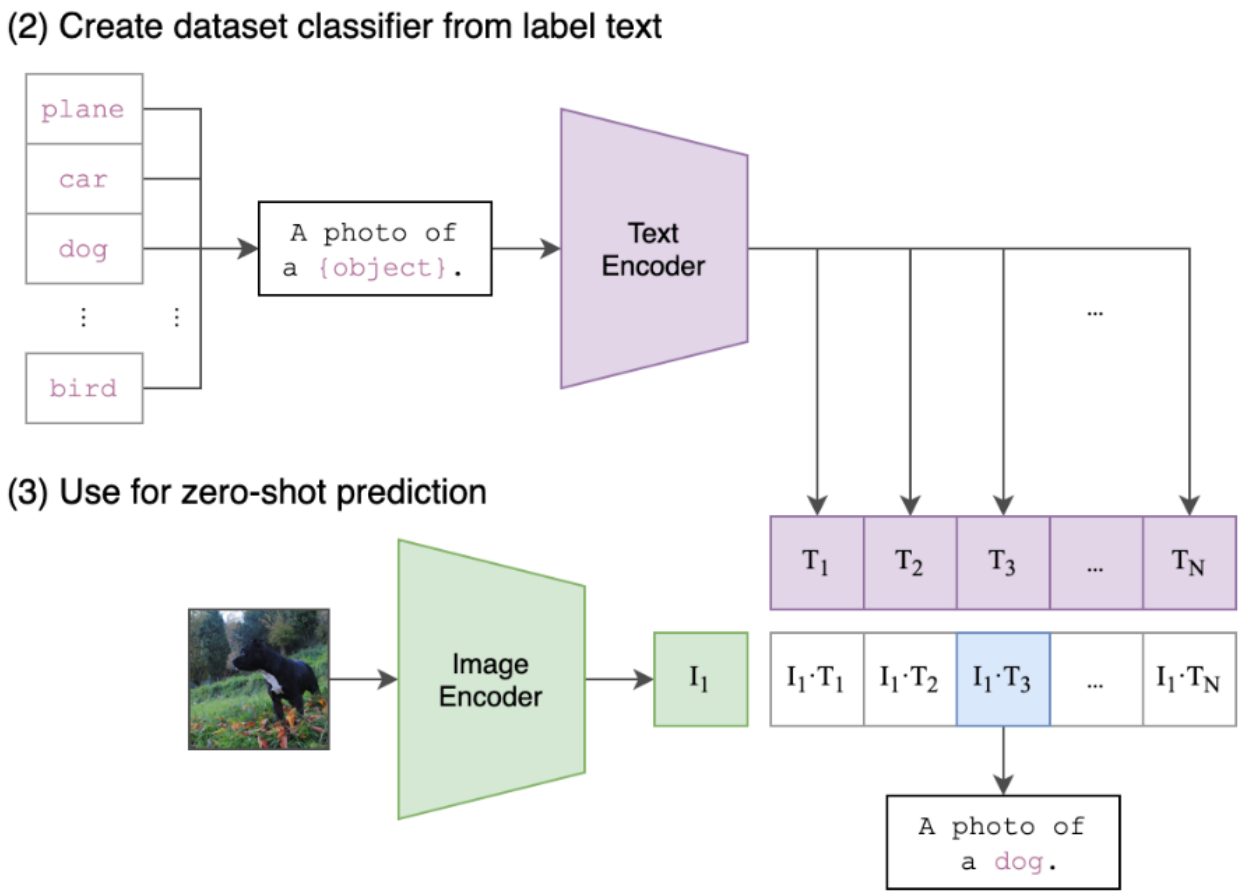


Open problem

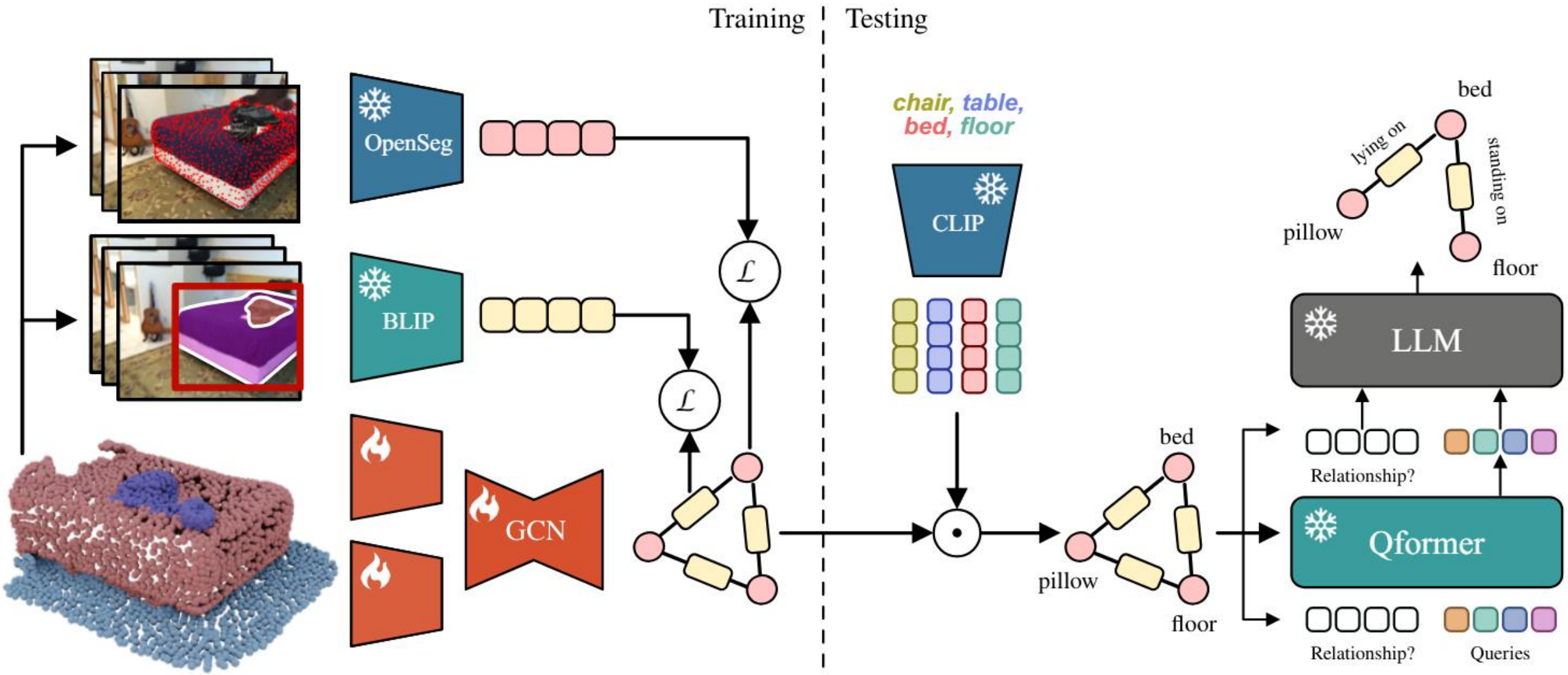
- **Fixed vocabulary**: Pre-defined set of semantic class categories for objects and relationships
- Why is that a problem?
- Solution **Open vocabulary** 3D Scene Graphs

CLIP: Contrastive Language-Image Pretraining



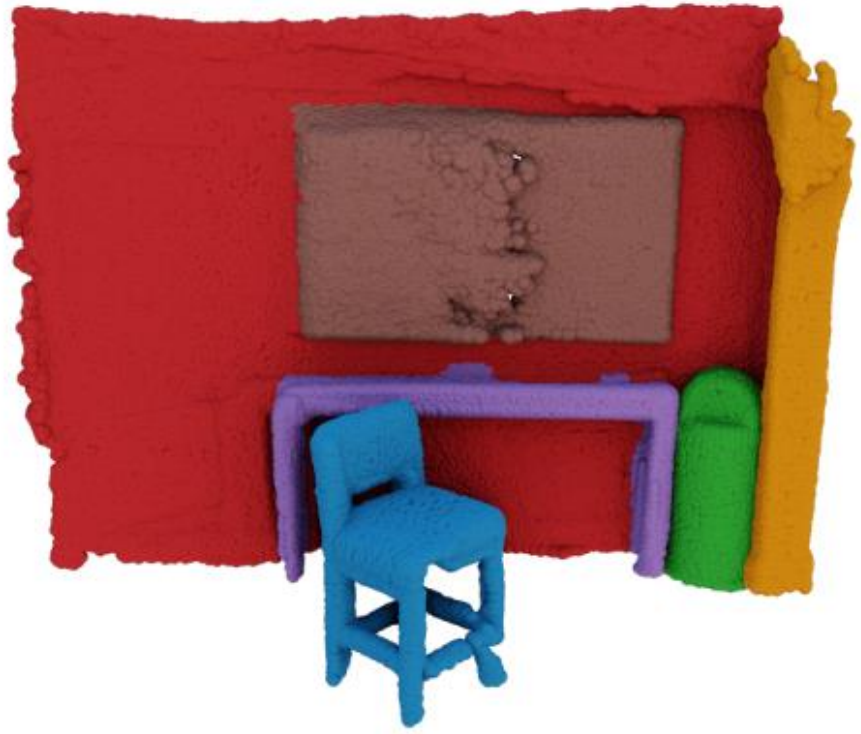


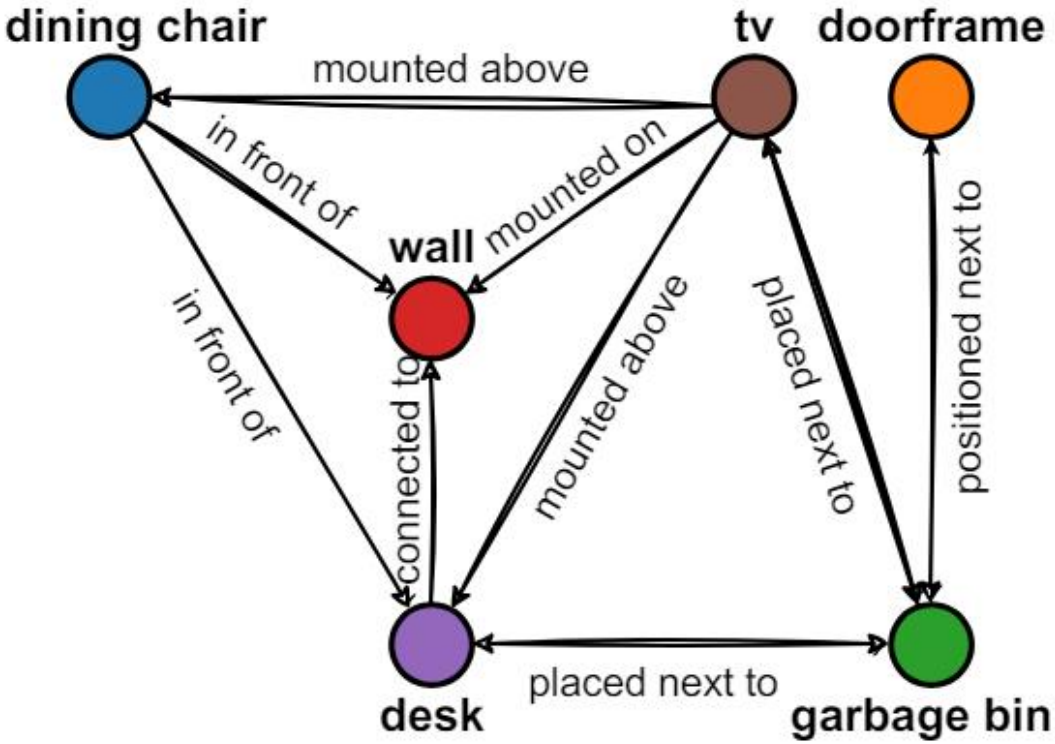
Open 3DSG



Koch el al. Open3DSG: Open-Vocabulary 3D Scene Graphs from Point Clouds with Queryable Objects and Open-Set Relationships. CVPR 2024

Open 3DSG





Koch el al. Open3DSG: Open-Vocabulary 3D Scene Graphs from Point Clouds with Queryable Objects and Open-Set Relationships. CVPR 2024