

# Toward large-scale interactive environments

Angel X. Chang

# NLP, Visual Computing & Robotics @ SFU

SFU overlooking Metro Vancouver



NLP



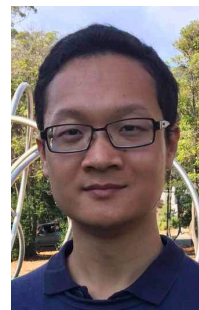
Anoop Sarkar



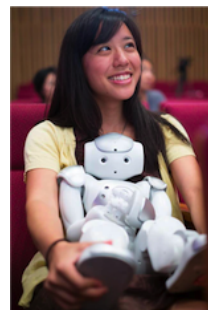
Angel Chang



Richard Vaughn



Mo Chen



Angelica Lim

Robotics

Visual Computing



Greg Mori



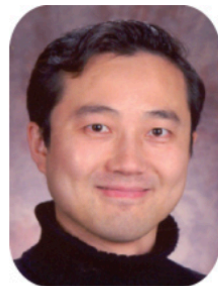
Ping Tan



Manolis Savva



KangKang Yin



Richard Zhang

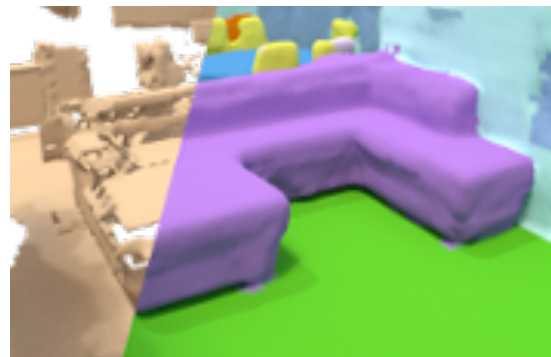


Yasutaka Furukawa

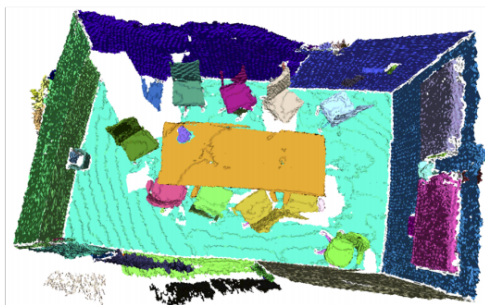
# Progress in 3D scene understanding



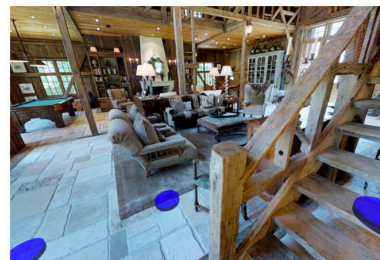
4D Spatio-Temporal ConvNets  
[Choy et al. 2019]



ScanComplete  
[Dai et al. 2018]



MASC  
[Liu and Furukawa 2019]

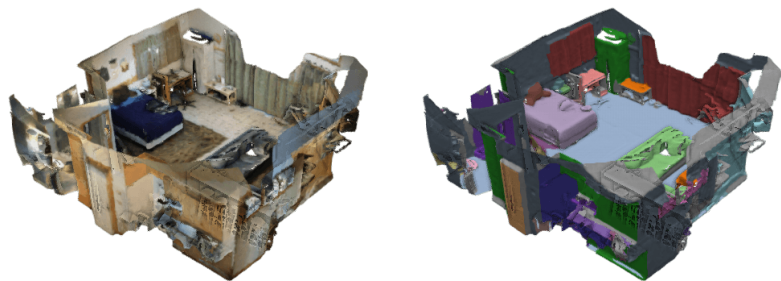


Instruction: Head upstairs and walk past the piano through an archway directly in front. Turn right when the hallway ends at pictures and table. Wait by the moose antlers hanging on the wall.

Vision-Language Navigation  
[Anderson et al. 2018]

# Environments for 3D scene understanding

ScanNet [Dai et al. 2017]



InteriorNet [Li et al. 2018]



# Learning through interaction



# Learning through interaction



# 3D environments for interaction

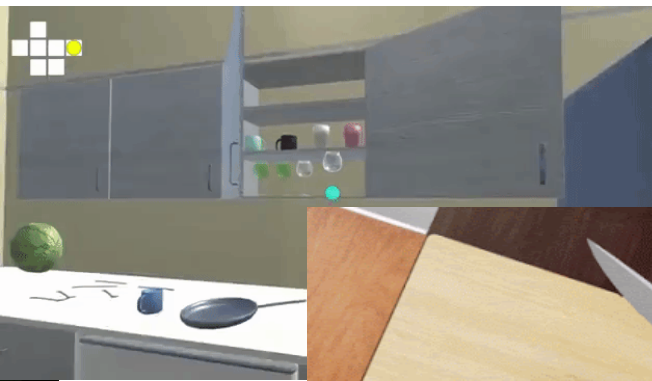
## AI2-THOR

[Kolve et al. 2017]



## Cornell CHALET

[Yan et al. 2018]



## VirtualHome

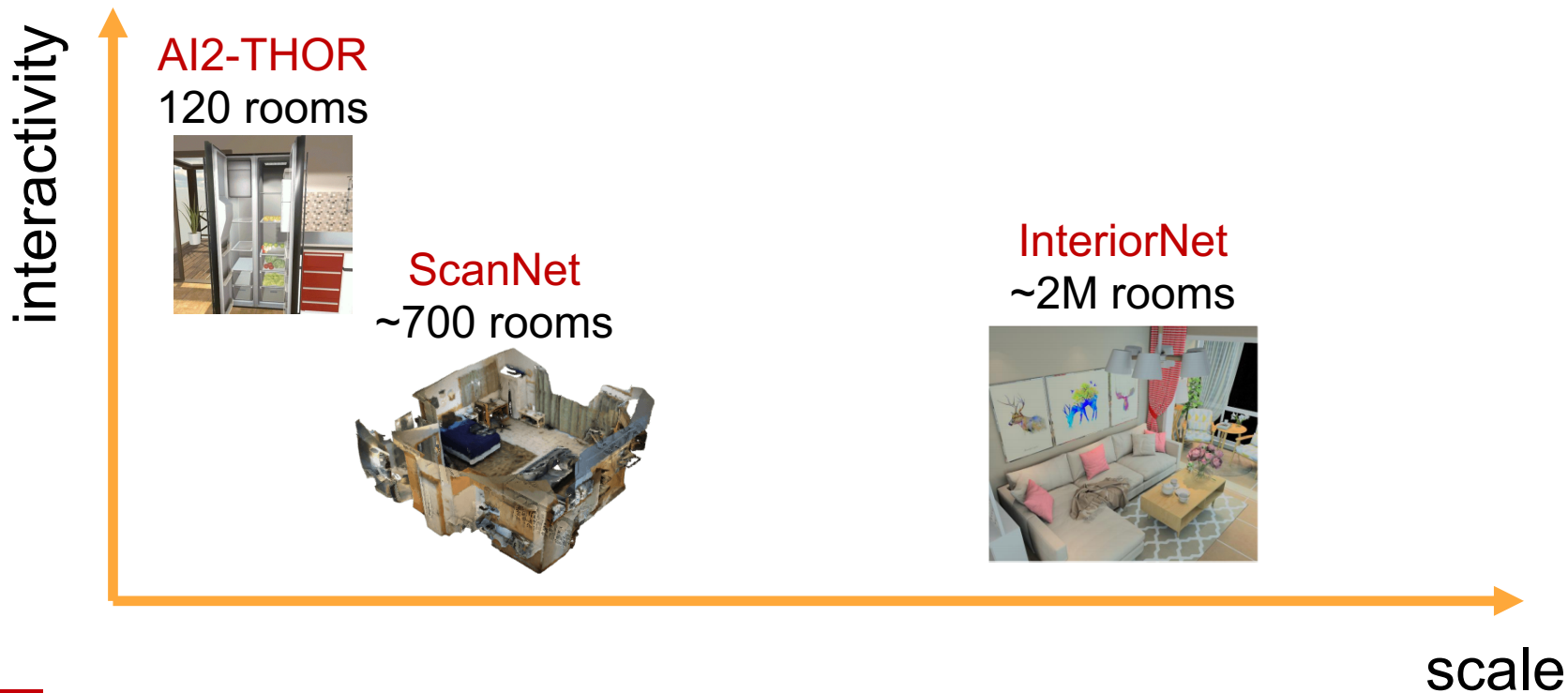
[Puig et al. 2018]



## VRKitchen

[Gao et al. 2019]

# Scale is limited compared to static datasets

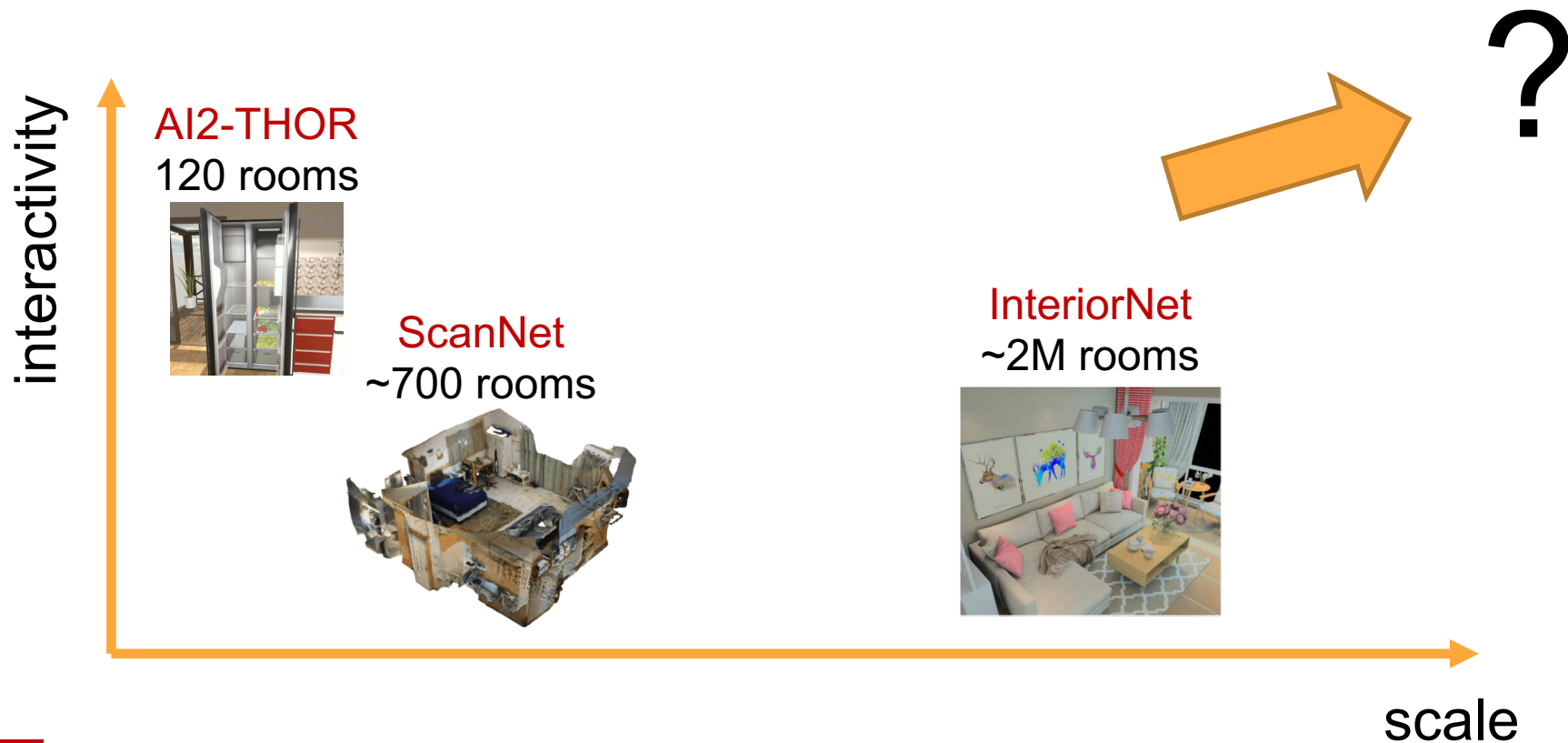




# Opportunities

- Understanding how an object can move → object detection, object tracking
- Understanding how parts move → human-object interaction understanding
- Understanding how sets of objects can co-occur → layout prediction, reconstruction, semantic segmentation

# How do we achieve large scale interactive environments?



# How do we achieve large scale interactive environments?



Environments are composed of objects

# Objects can be rearranged



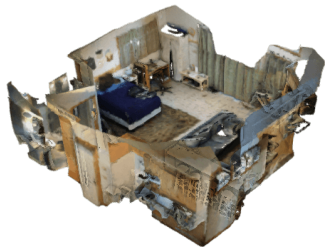
Similarly for objects



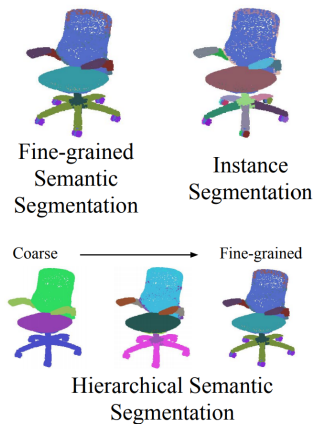
Objects are composed of parts

# Bootstrapping

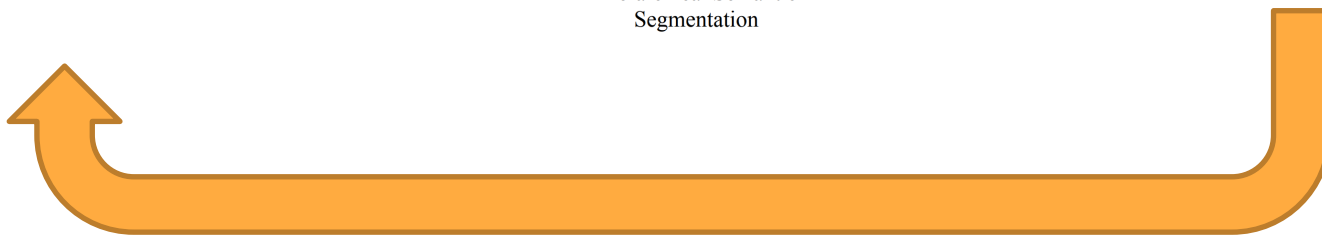
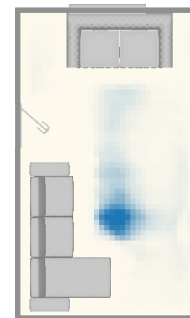
Initial annotated data



Learn to segment and parse



Learn generative models



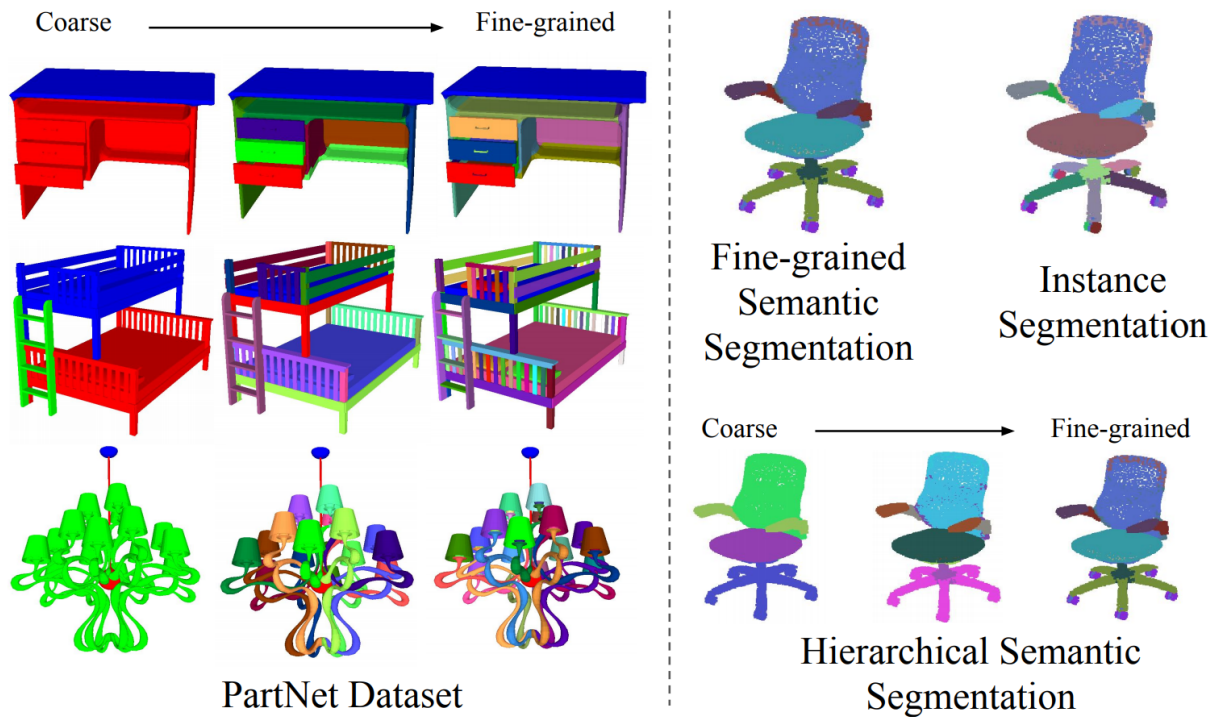
# What do we need to have?

- Structure of environments
  - Model environments as objects, objects as parts
  - Learn to **parse** objects from environments, parts from objects
  - Learn to **generate** objects from parts, environments from objects
- Interaction
  - What objects/parts can move? How do they move?
    - Seed set that is annotated, learn and propagate to more (fixup) and then iterate
  - What physical properties does the objects/parts have? Mass, friction, materials.

# Understanding object parts



# PartNet: A Large-scale Benchmark for Fine-grained and Hierarchical Part-level 3D Object Understanding



[Mo et al. CVPR 2019]

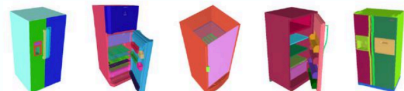
# PartNet

- 26,671 ShapeNet objects with 573,585 parts
- Hierarchical part segmentation

Mug



Refrigerator



Scissors



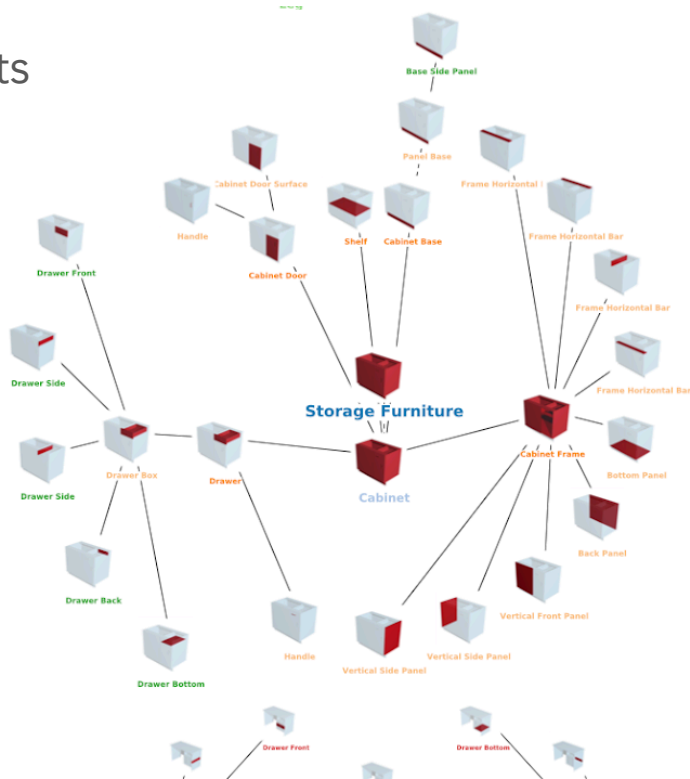
Table



Trash Can



Vase



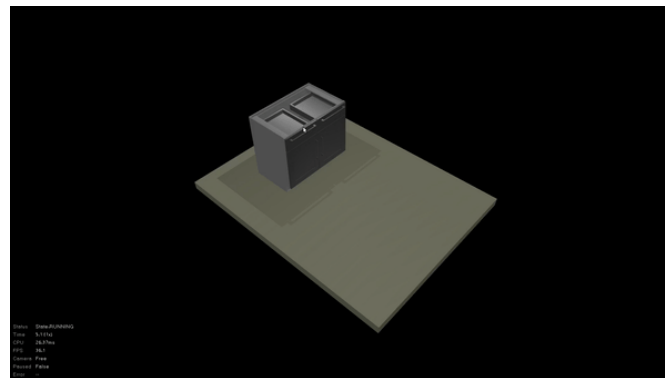
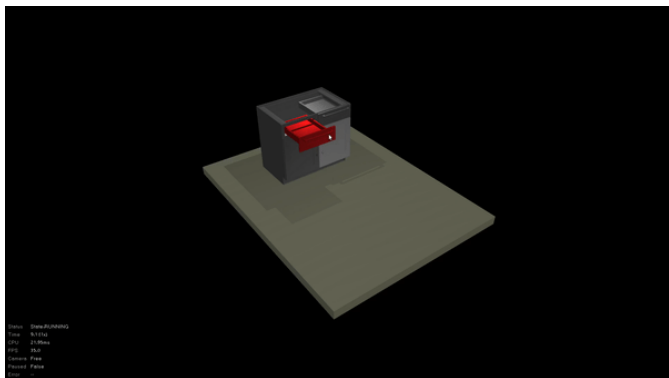
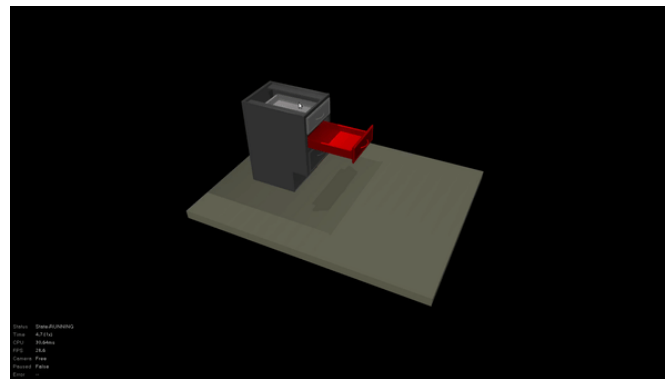
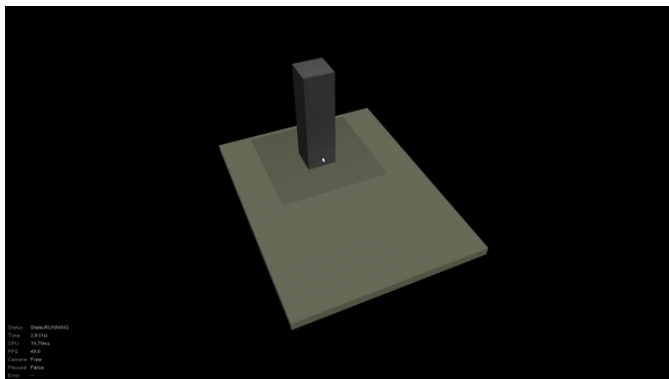
# Articulated PartNet

- Type of motion: rotation vs translation
- Axis and origin of motion
- Range of motion



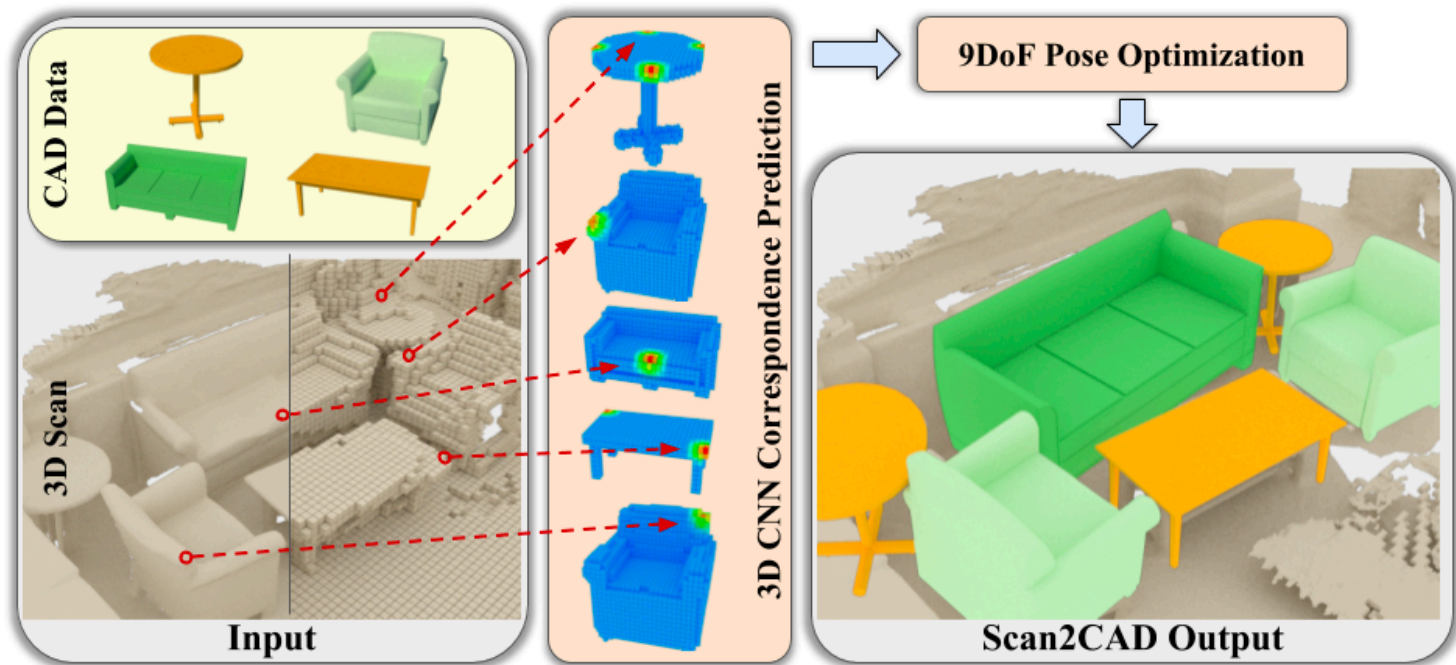
~300 annotated objects, 1060 joints  
(ongoing)

# Articulated PartNet



# From objects to environments

# Scan2CAD: Learning CAD Model Alignment in RGB-D Scans



*[Avetisyan et al. CVPR 2019]*

# Scan2CAD: Learning CAD Model Alignment in RGB-D Scans



1506 scans of ~700 rooms  
14225 scan objects annotated with CADs

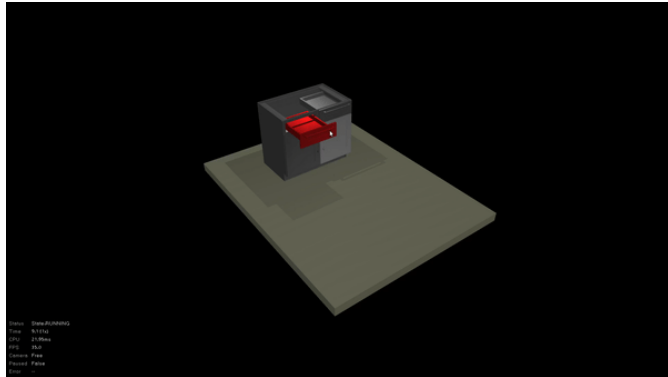
# Interactive environments

Scan2CAD



+

→ Interactive environments



articulated PartNet

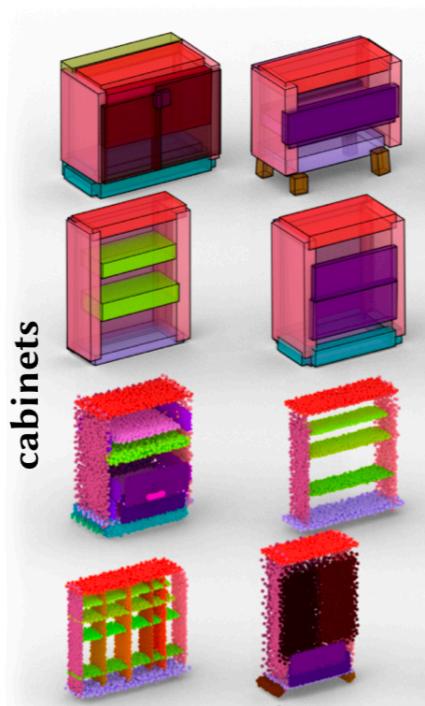


# Interactive environments



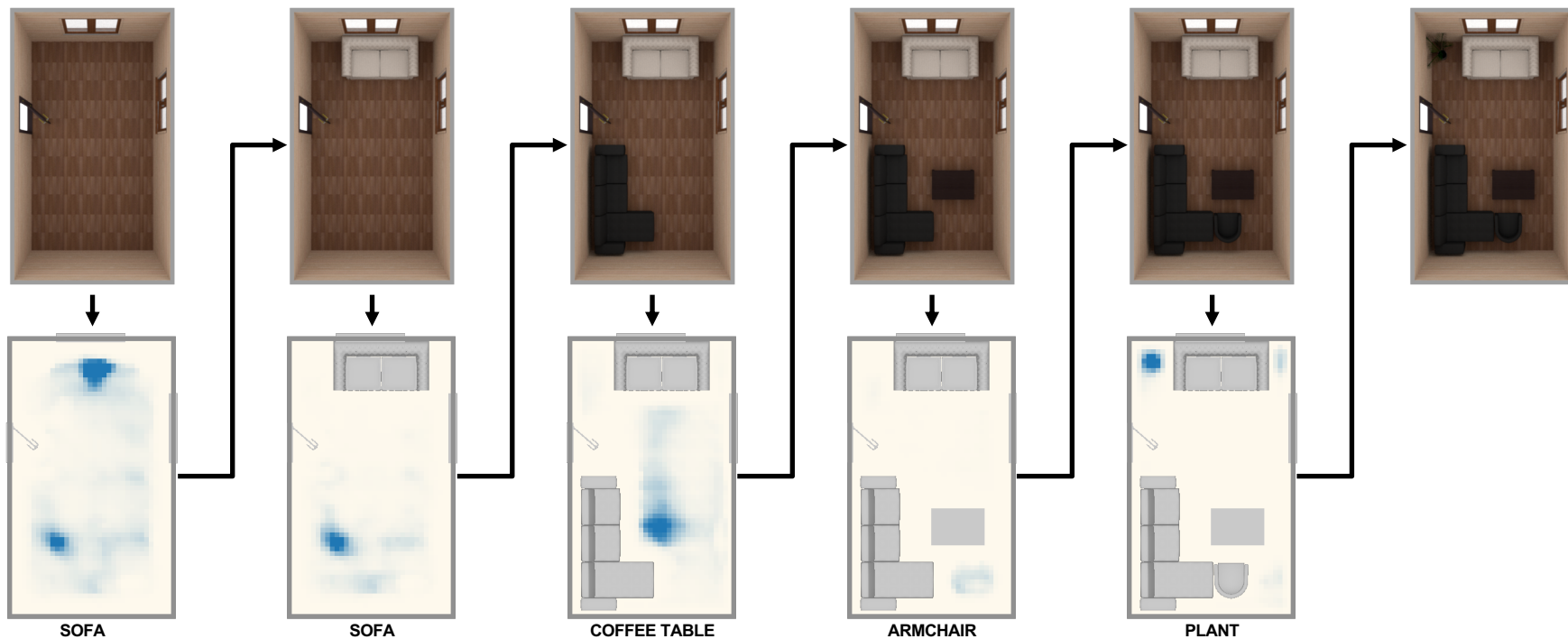
# Generating new environments and shapes

# Novel shapes from parts



[Mo et al, 2019]

# Arranging objects for new room layouts



Deep Convolutional Priors for Scene Synthesis [Wang et al, 2018]



# What is still missing?

- From rooms to houses
- Inferring object mobility in an environment
- Simulating object interaction with physics
- Inferring physical properties: mass distribution, materials etc.

## Ongoing work: interactive objects in AI Habitat



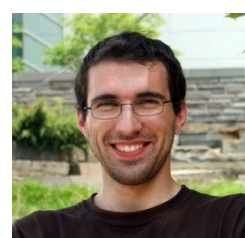
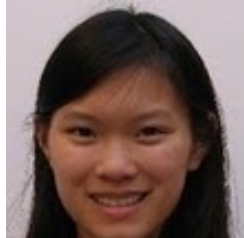
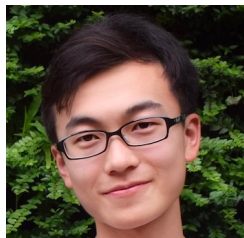
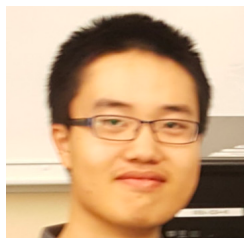
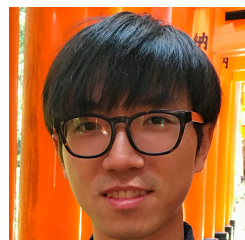
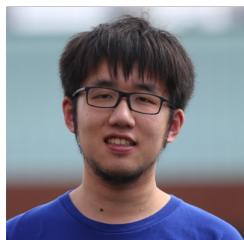
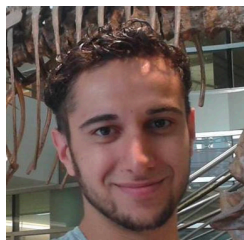
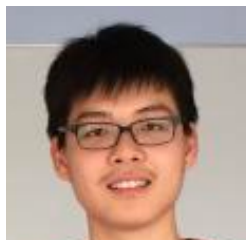
- Fully open sourced, high performance, generic 3D scene dataset support

# Object interaction with physics in Habitat





# Collaborators



Thank you! Questions?

