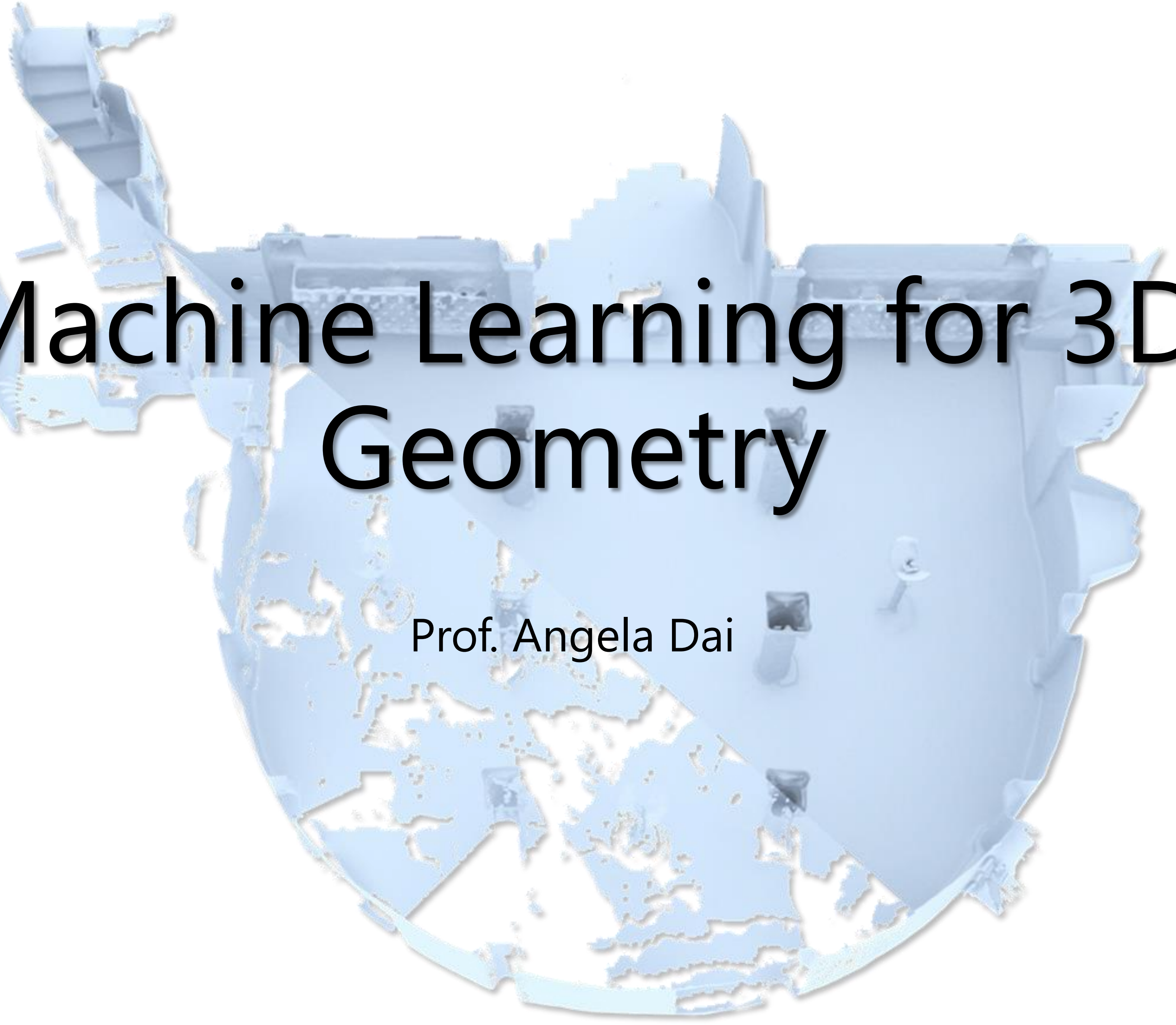


Machine Learning for 3D Geometry

Prof. Angela Dai

Welcome

- Why this course
- Course Overview
- Logistics

Team

Lecturers

Prof. Angela Dai

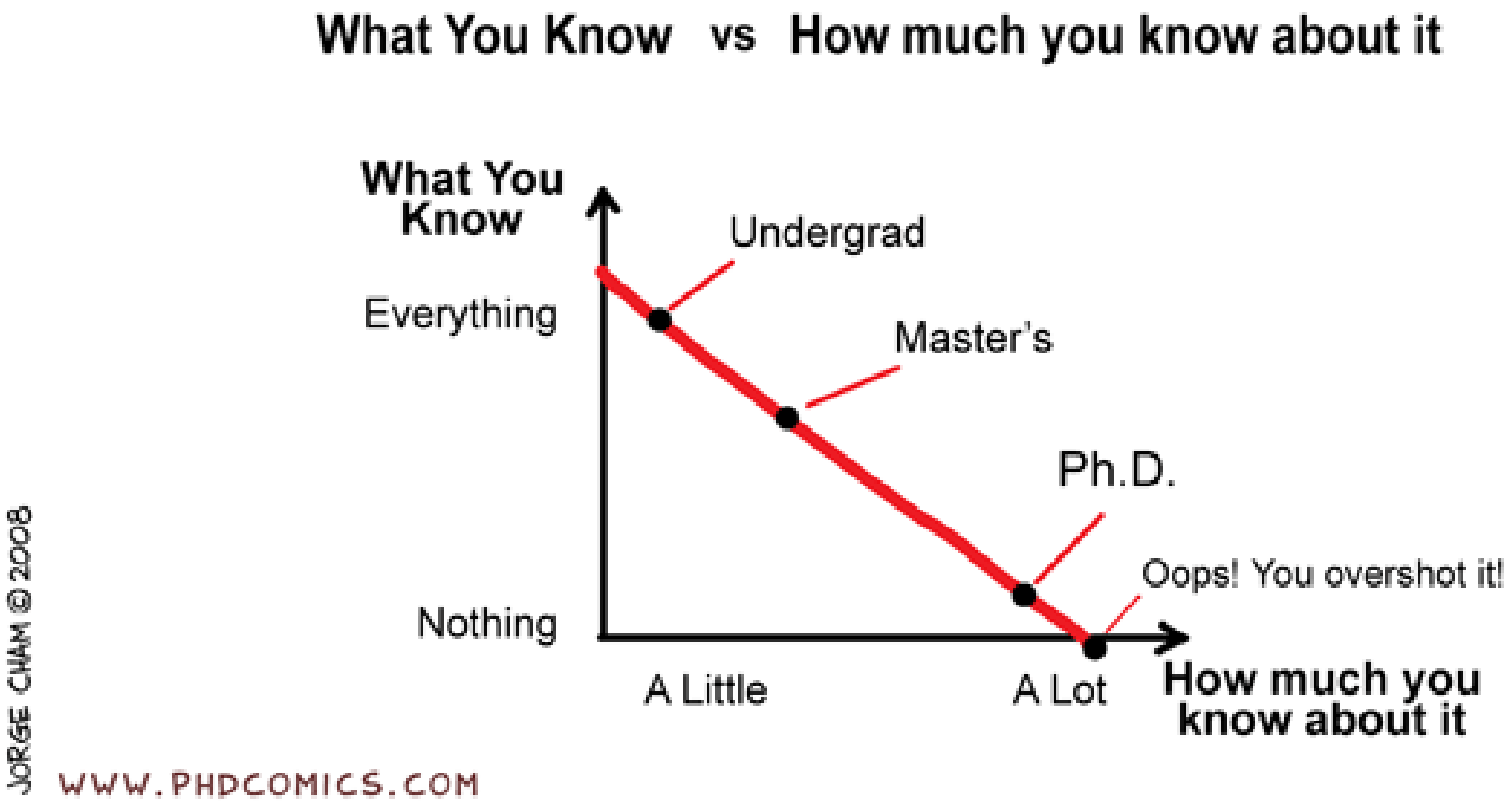
Teaching Assistants

Christian Diller

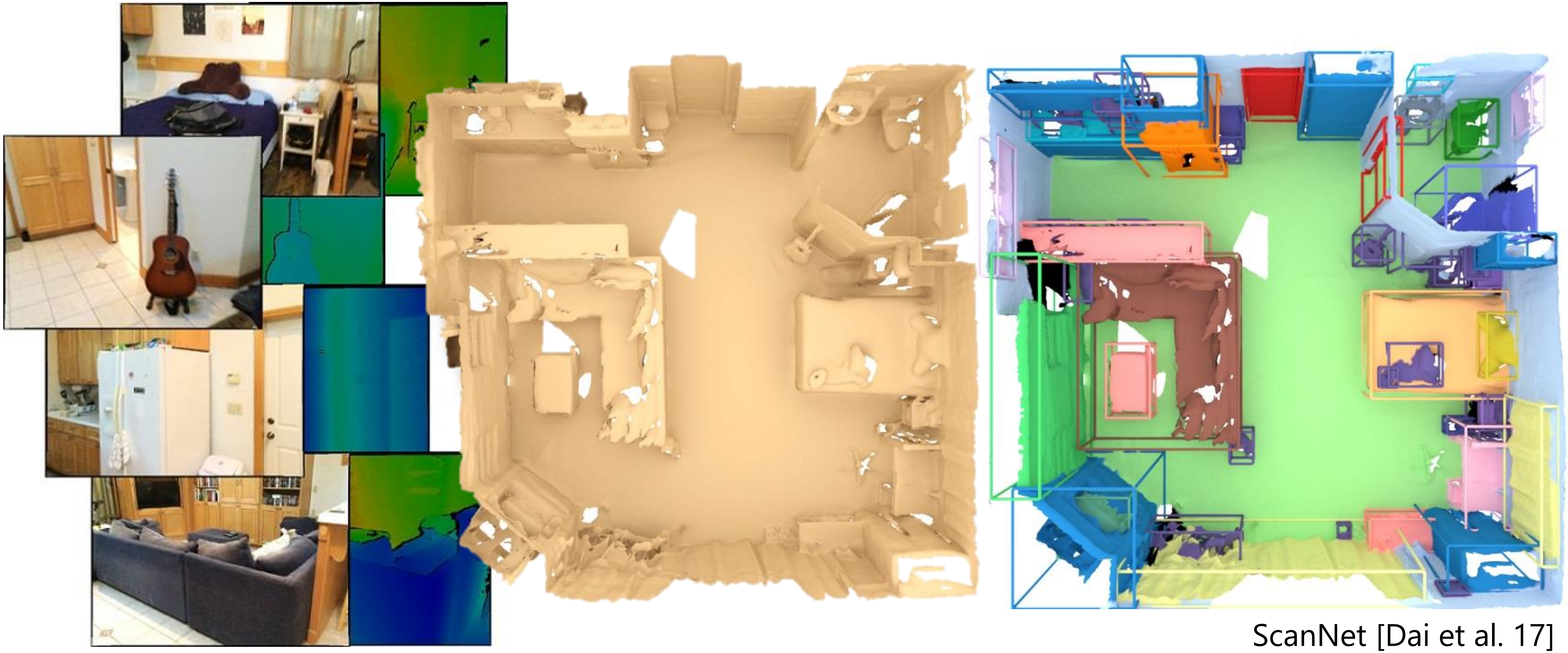
Can Guemeli

Jiapeng Tang

Research!


What You Know vs How much you know about it
What You Know
Everything
Nothing
Undergrad
Master's
Ph.D.
Oops! You overshot it!
A Little
A Lot
How much you know about it
JORGE CHAM © 2008
WWW.PHDCOMICS.COM

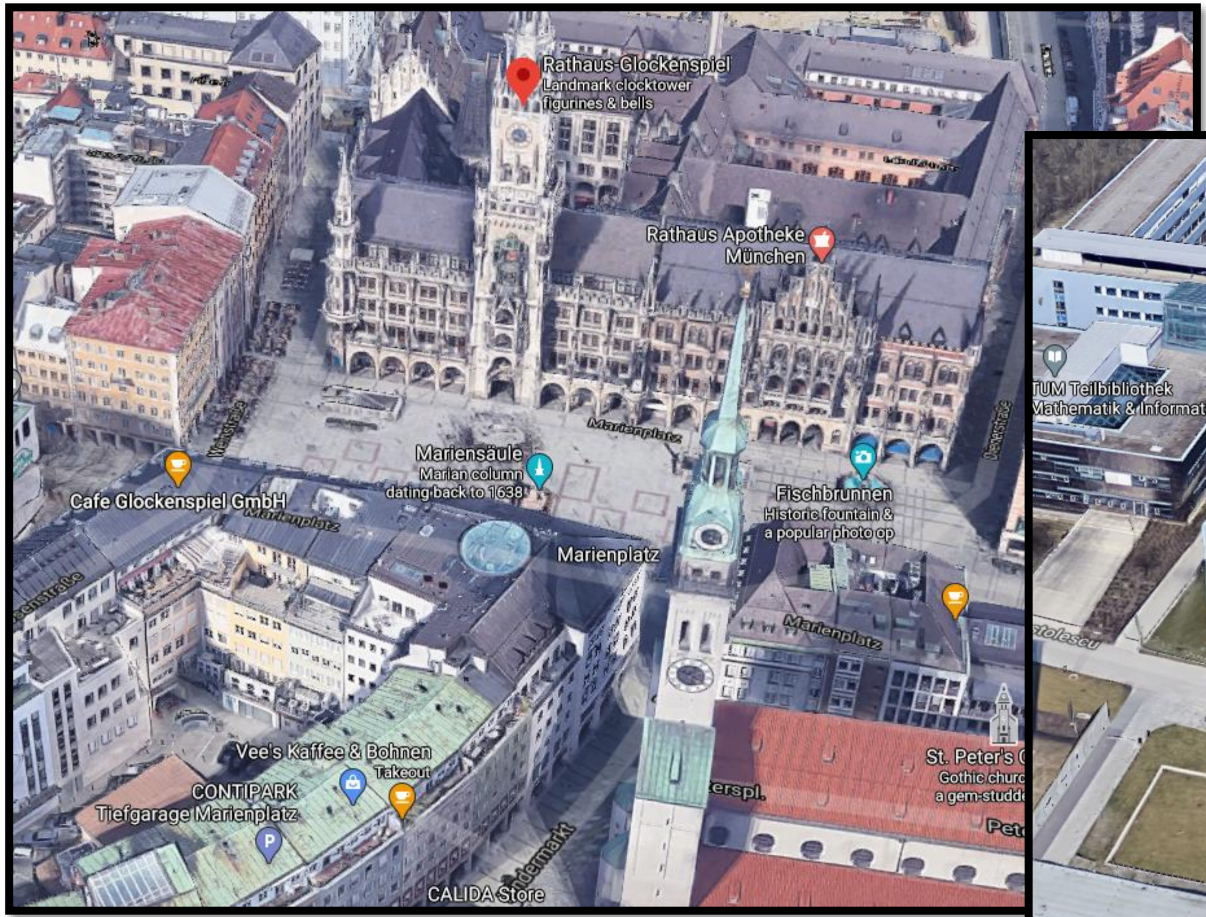
Machine Perception of Real-World Environments



ScanNet [Dai et al. 17]

Why 3D Perception?

We live in a 3D world



Google Maps

We *perceive* and *interact* with a 3D world

ASIMO, Honda

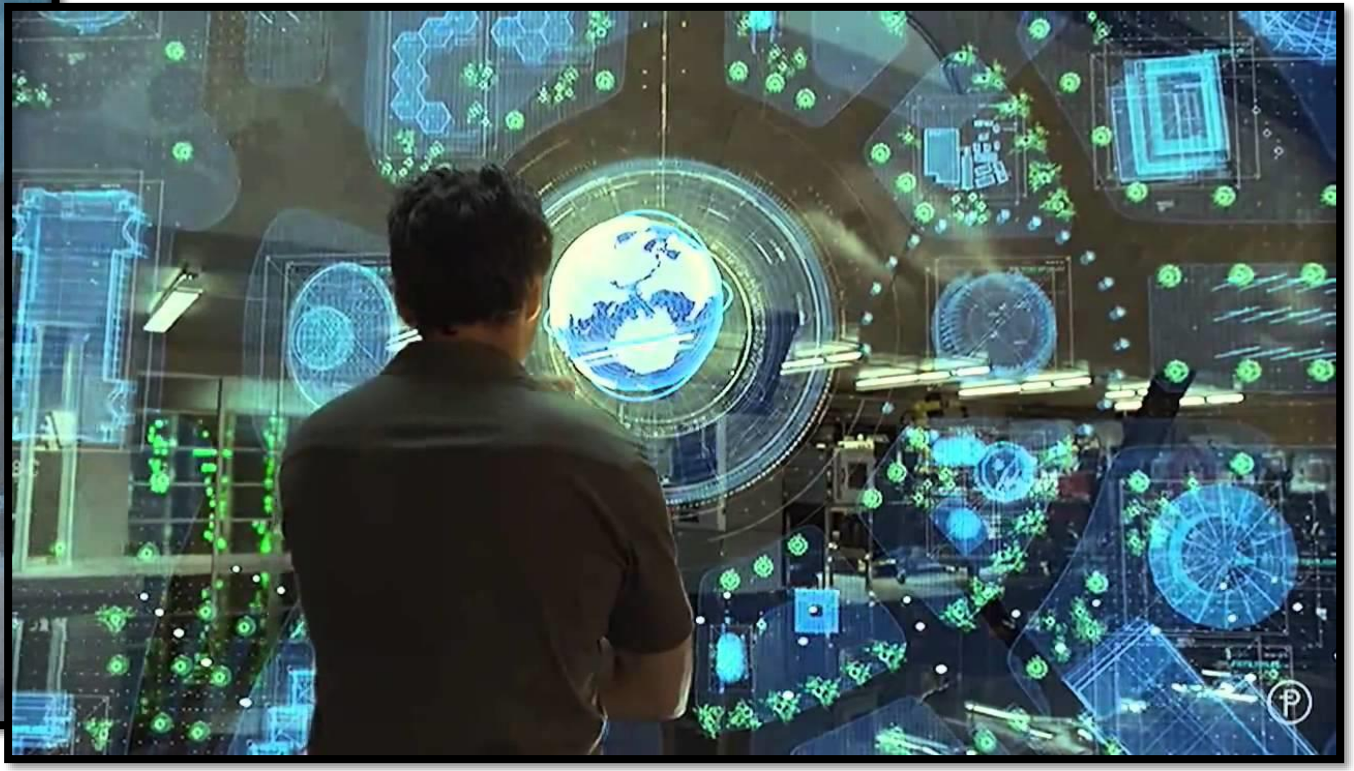
Star Trek TNG (Phantasms)

Insight to how we learn

iStock/RobHainer

Augmented Reality

Minority Report



Iron Man

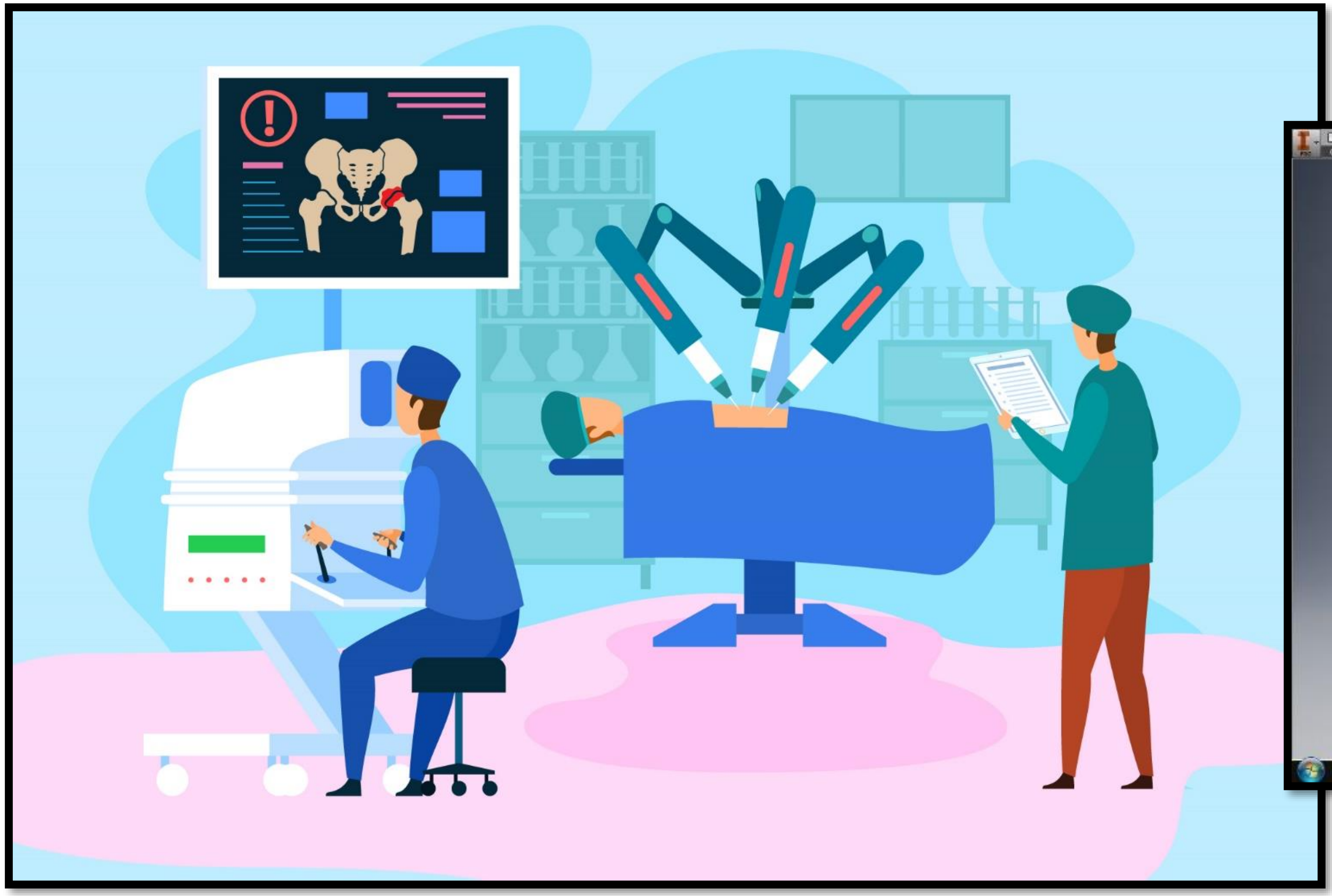
Virtual Reality

Star Trek

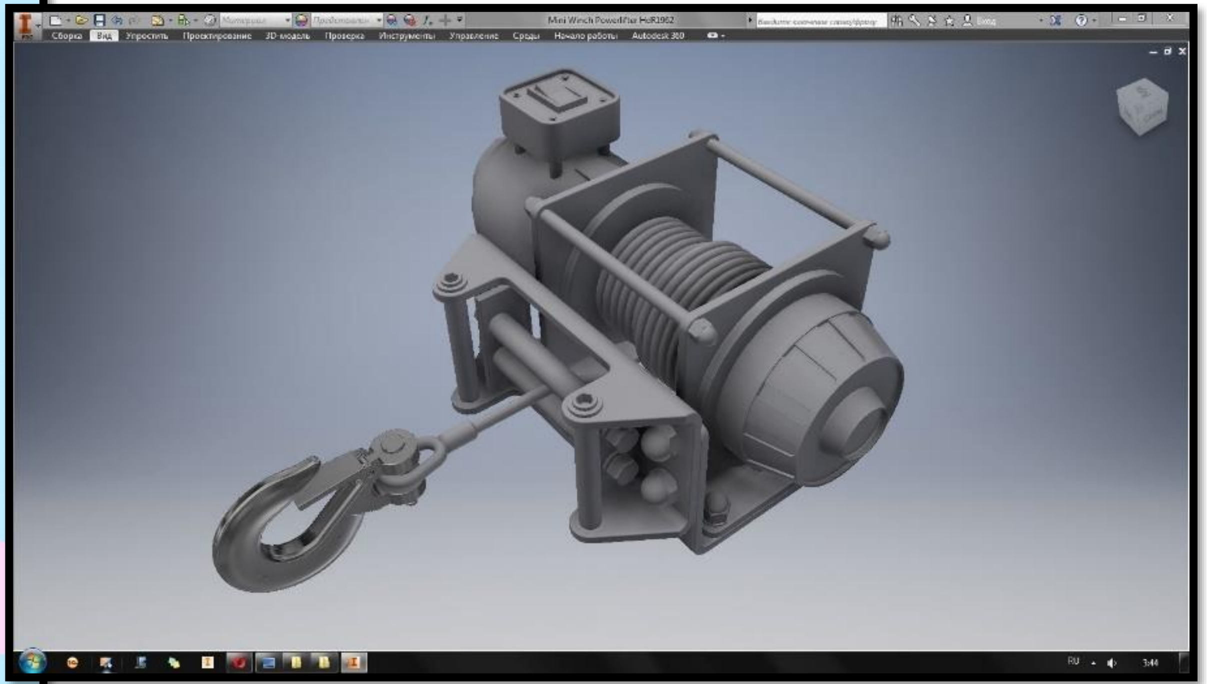
http://getsupernatural.com

Medical Analysis

Manufacturing



opmed.doximity.com

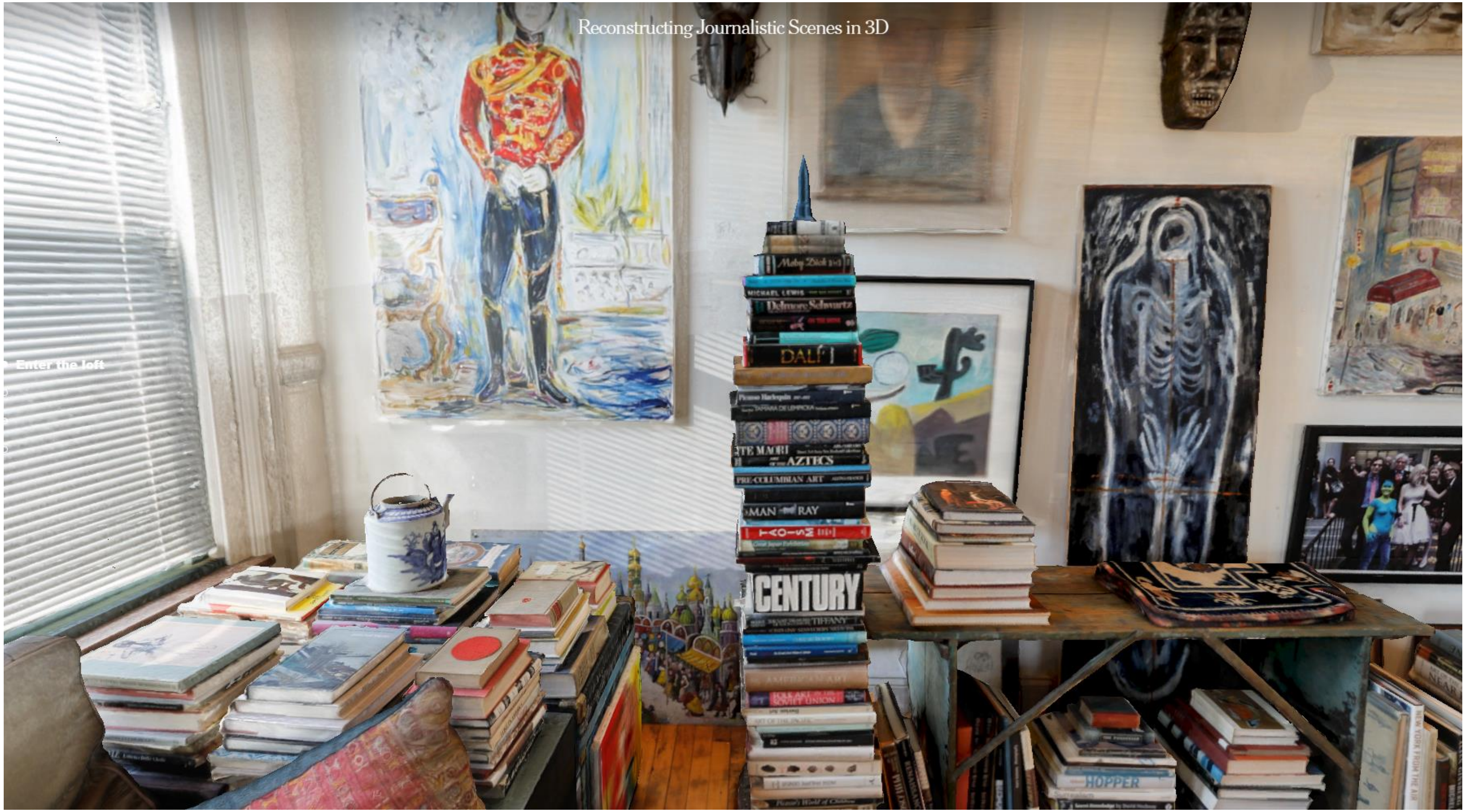


cgtrader.com

Virtual Try-On

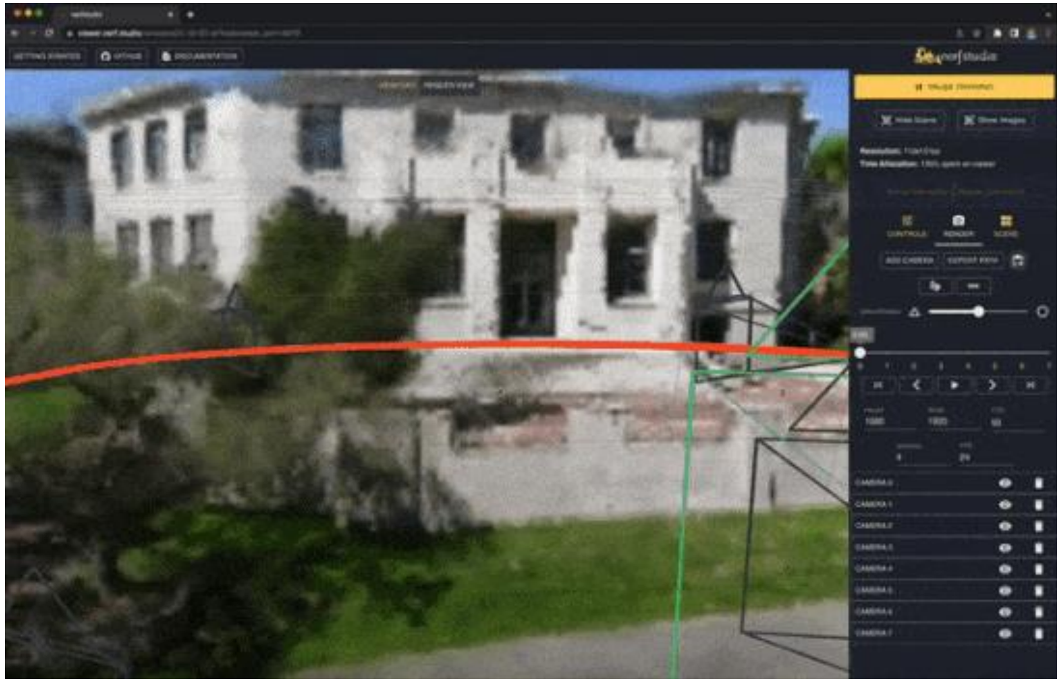
https://www.quytech.com/blog/how-virtual-mirror-technology-will-change-the-way-you-shop

Virtual Exploration



https://rd.nytimes.com/projects/reconstructing-journalistic-scenes-in-3d

Virtual Exploration



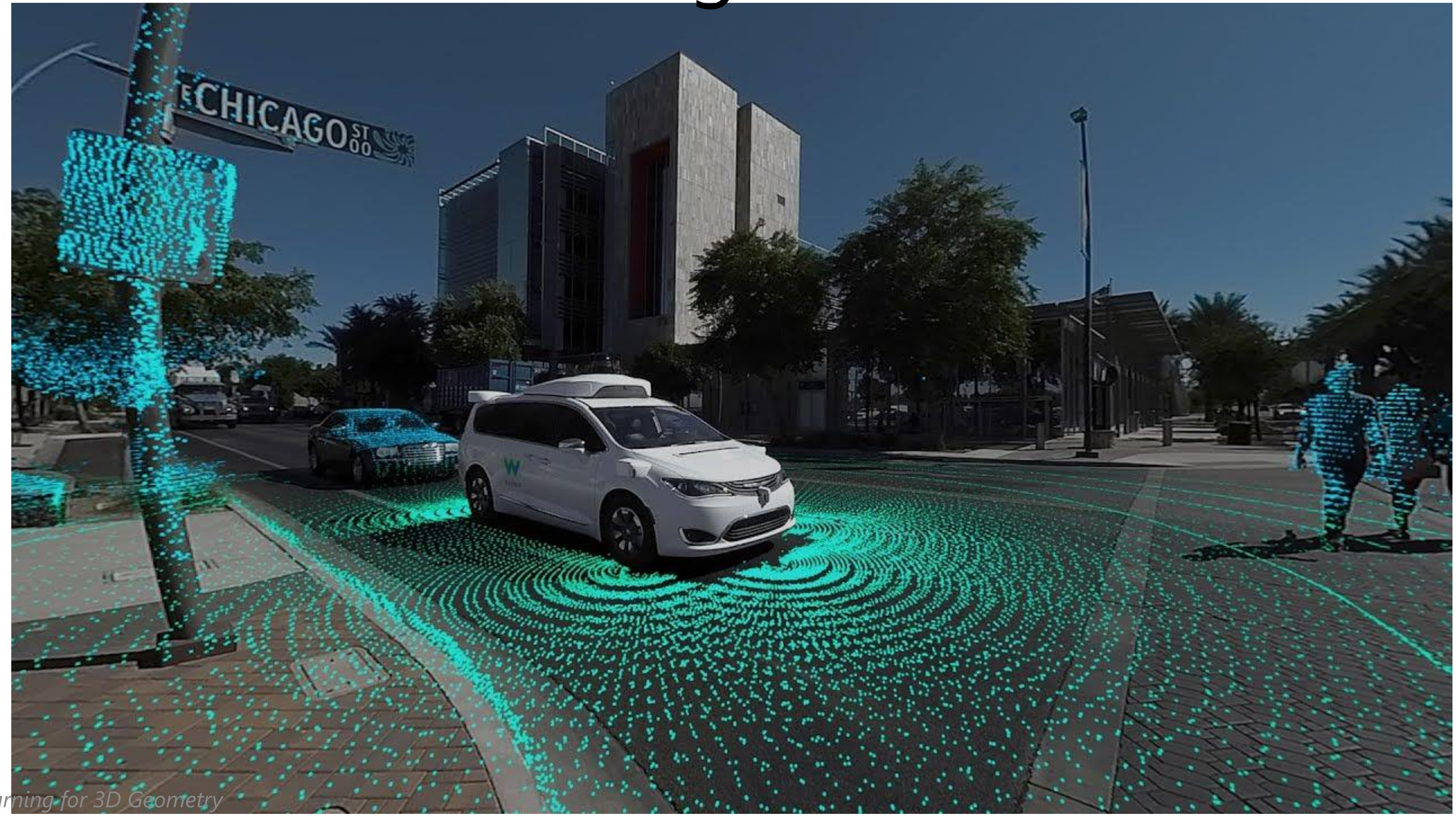
https://docs.nerf.studio/en/latest/index.html

Content Creation

IKEA

Assassin’s Creed

Autonomous Driving

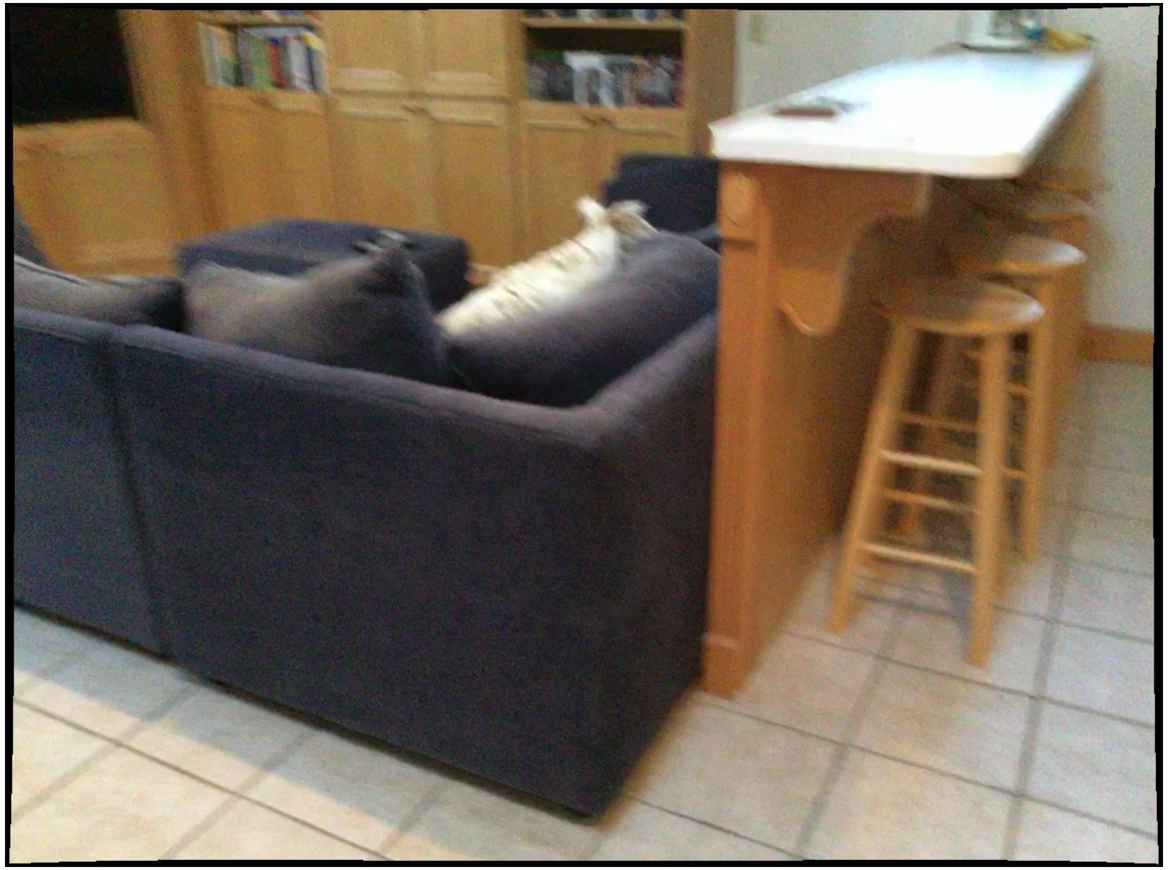


Waymo

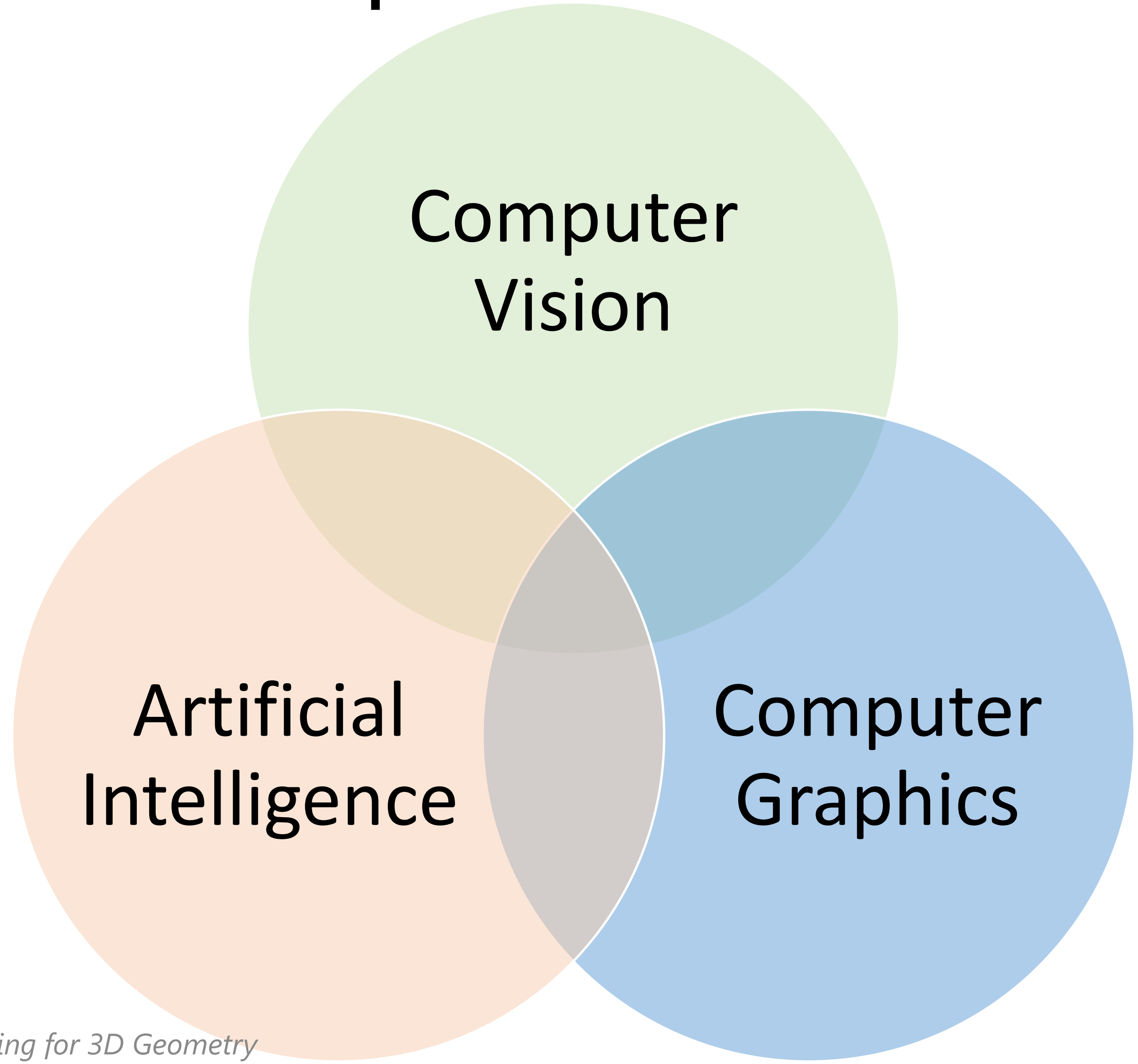
Robot Movement and Interactions

Boston Dynamics

2D views of a 3D world



3D Perception

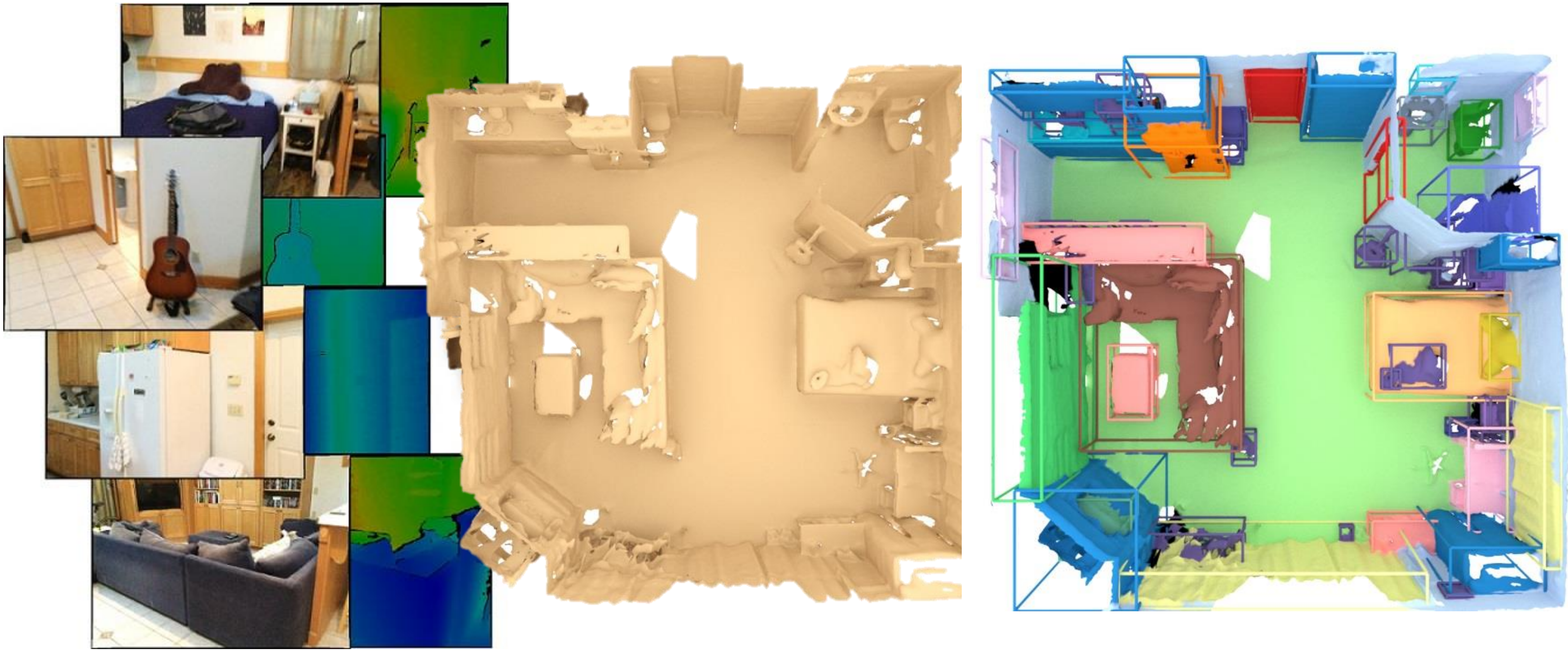


- Shape Classification
- Part Segmentation
- Semantic Scene Segmentation
- Object Reconstruction
- Scene Reconstruction
- Object Tracking
- Pose Estimation
- Affordance Prediction
- …

Goals

- Understanding the 3D World
- Perceiving and reconstructing 3D geometry of shapes and scenes
- Recognizing objects and their functionality
- Creating new, immersive virtual environments
- Helping robots perceive the world like humans

3D AI Lab



Capturing real-world environments

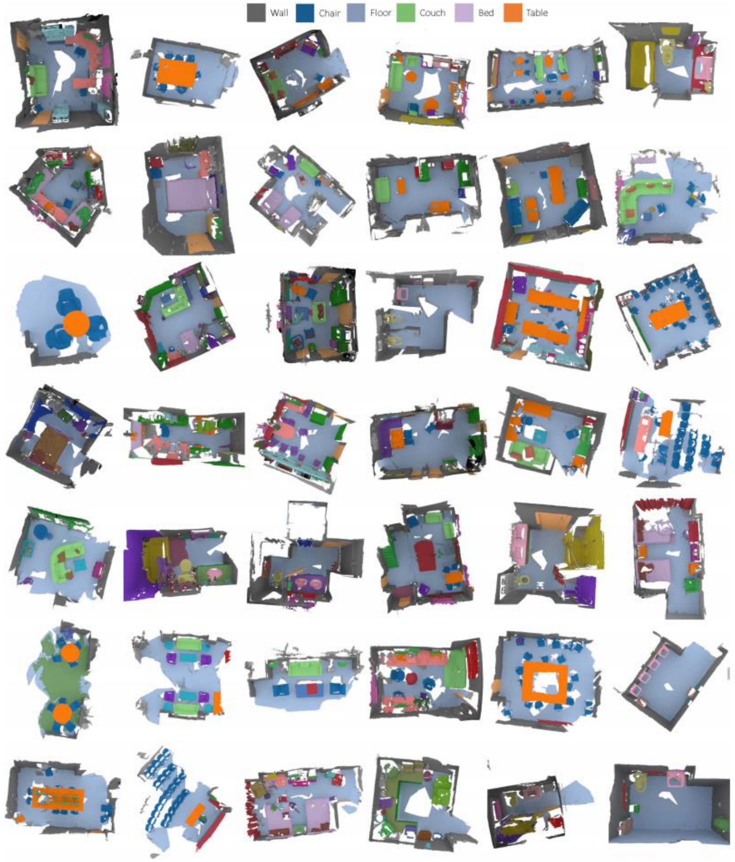
[Dai et al '17] BundleFusion

Capturing real-world environments

[Dai et al '17] BundleFusion

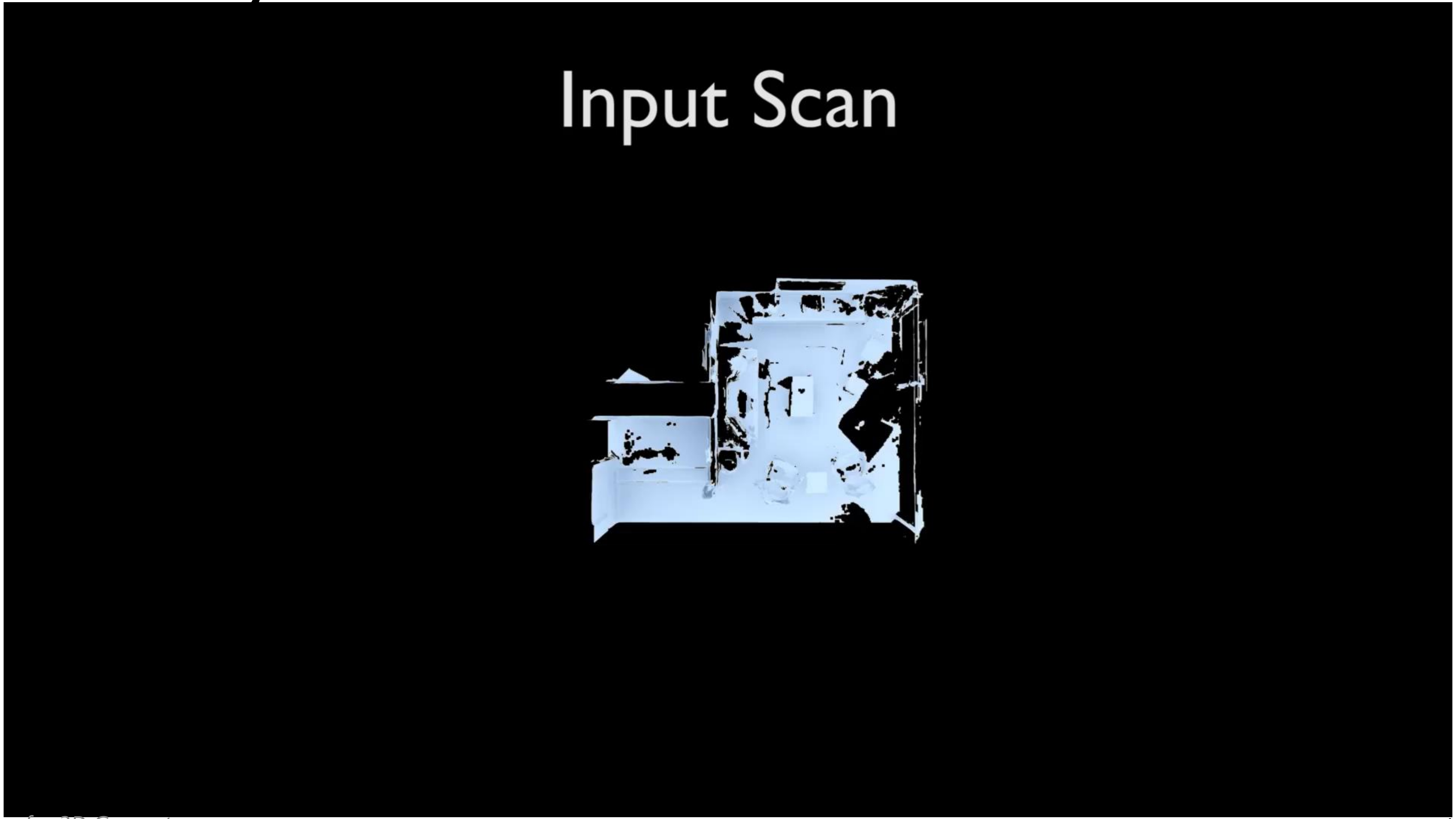
Learning from real-world environments

- ScanNet Dataset



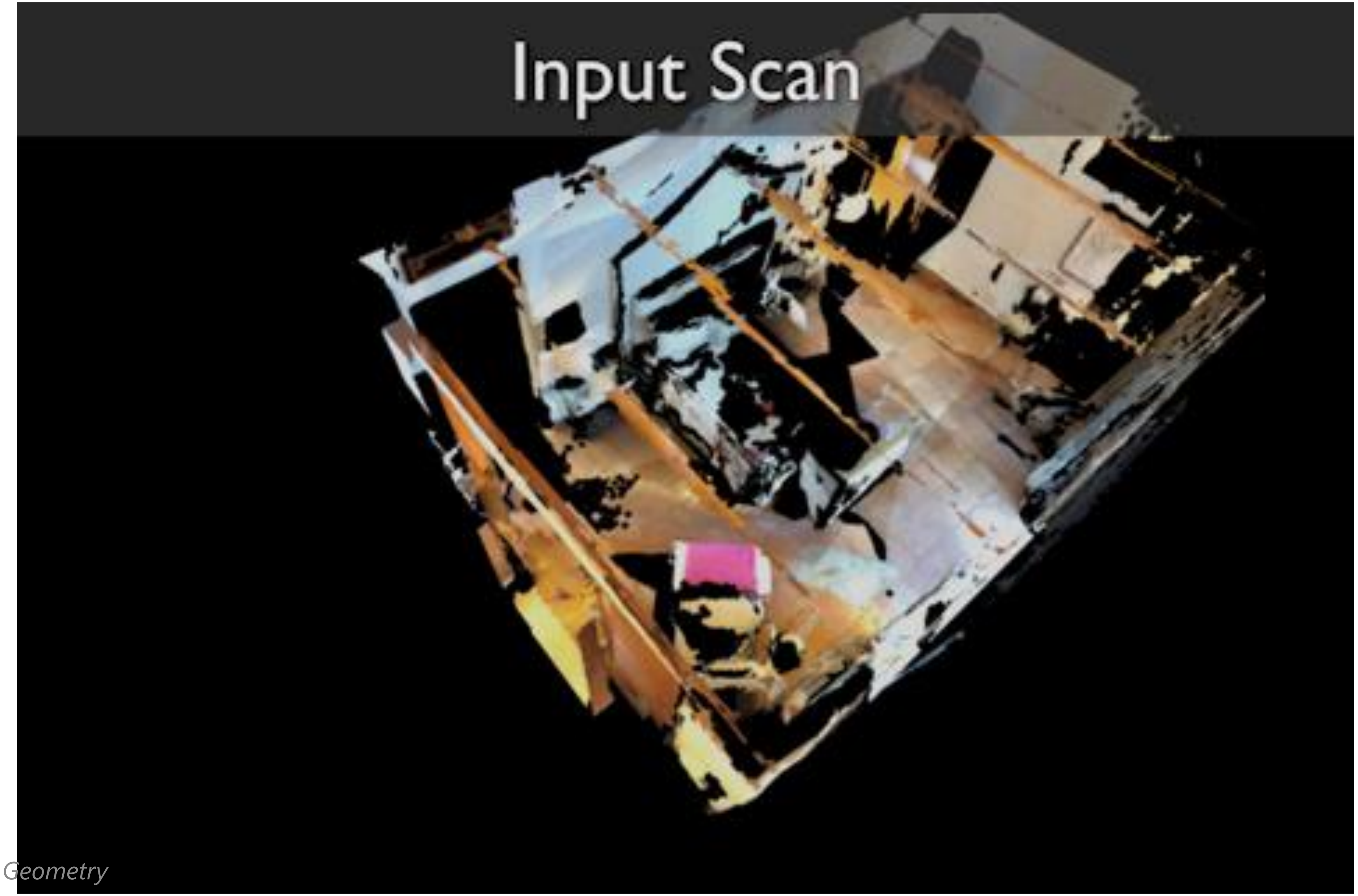
[Dai et al ‘17] ScanNet

Learning to reconstruct real-world scenes



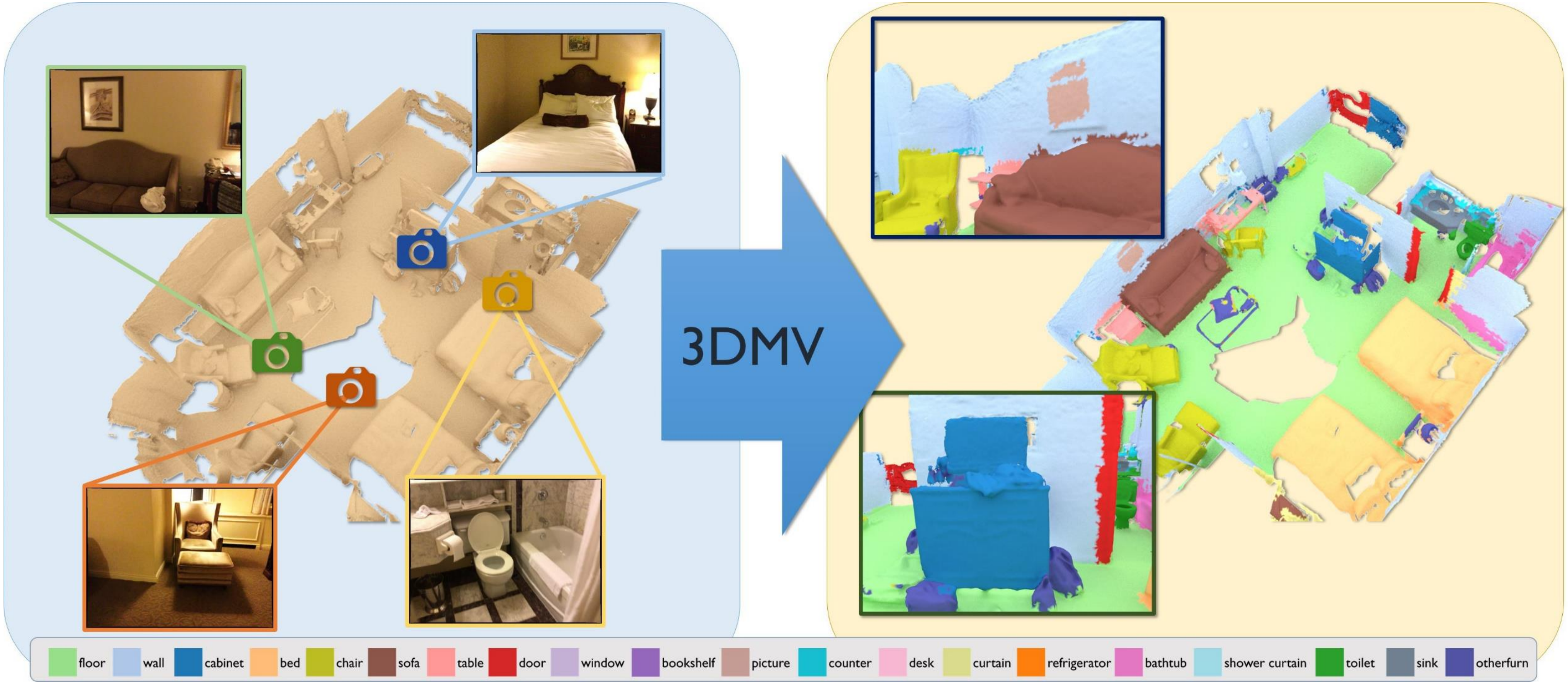
[Dai et al '20] SGNN

Learning to reconstruct real-world scenes



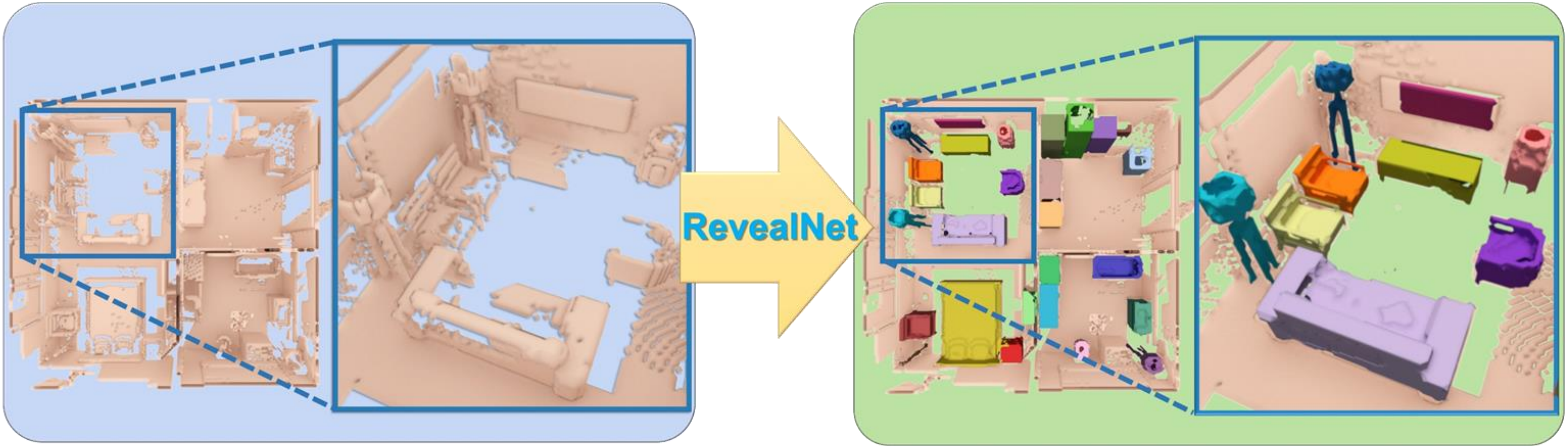
[Dai et al '20] SPSG

Learning to understand real-world scenes



[Dai et al '18] 3DMV

Learning to understand real-world scenes

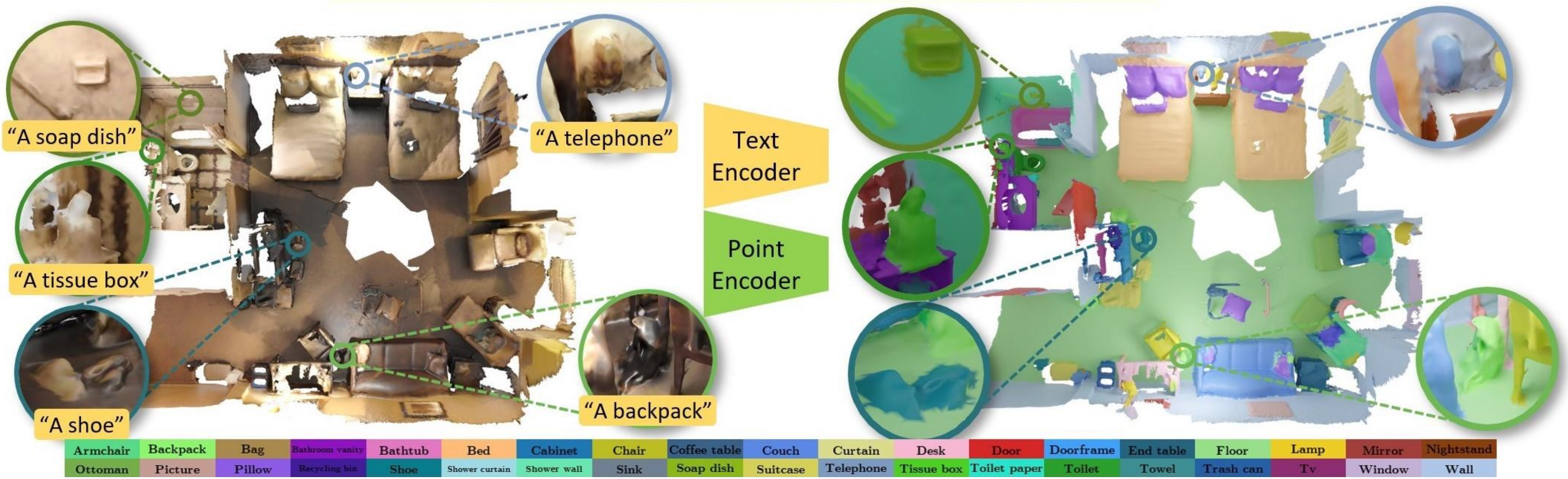


[Hou et al '20] RevealNet

Learning to understand real-world scenes

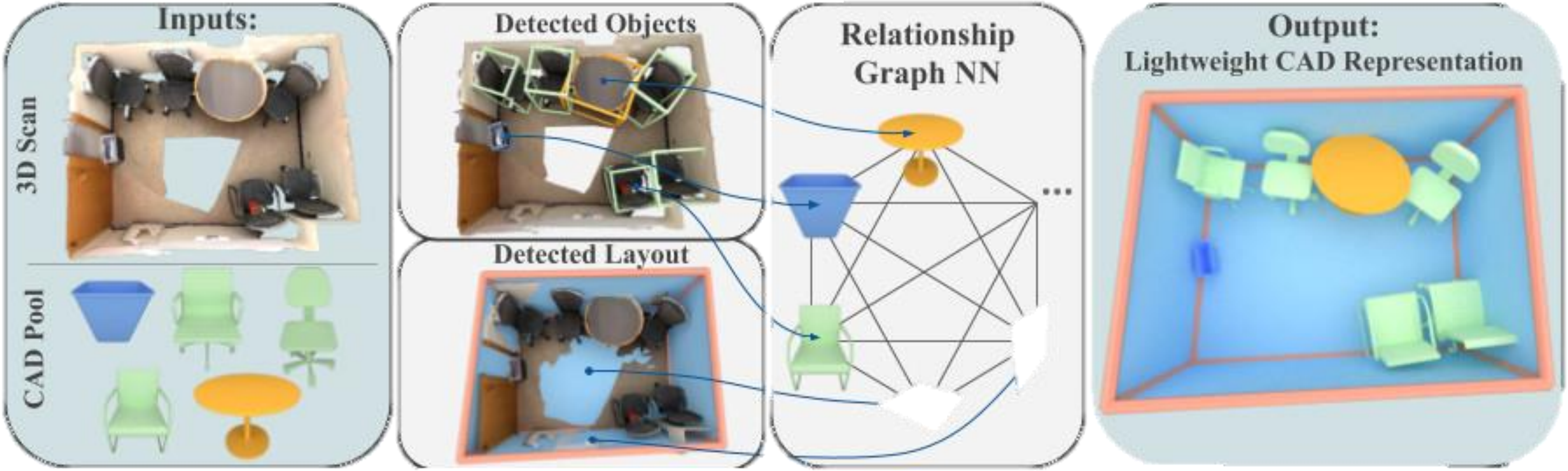
[Bokhovkin et al '23] Neural Part Priors

Learning to understand real-world scenes



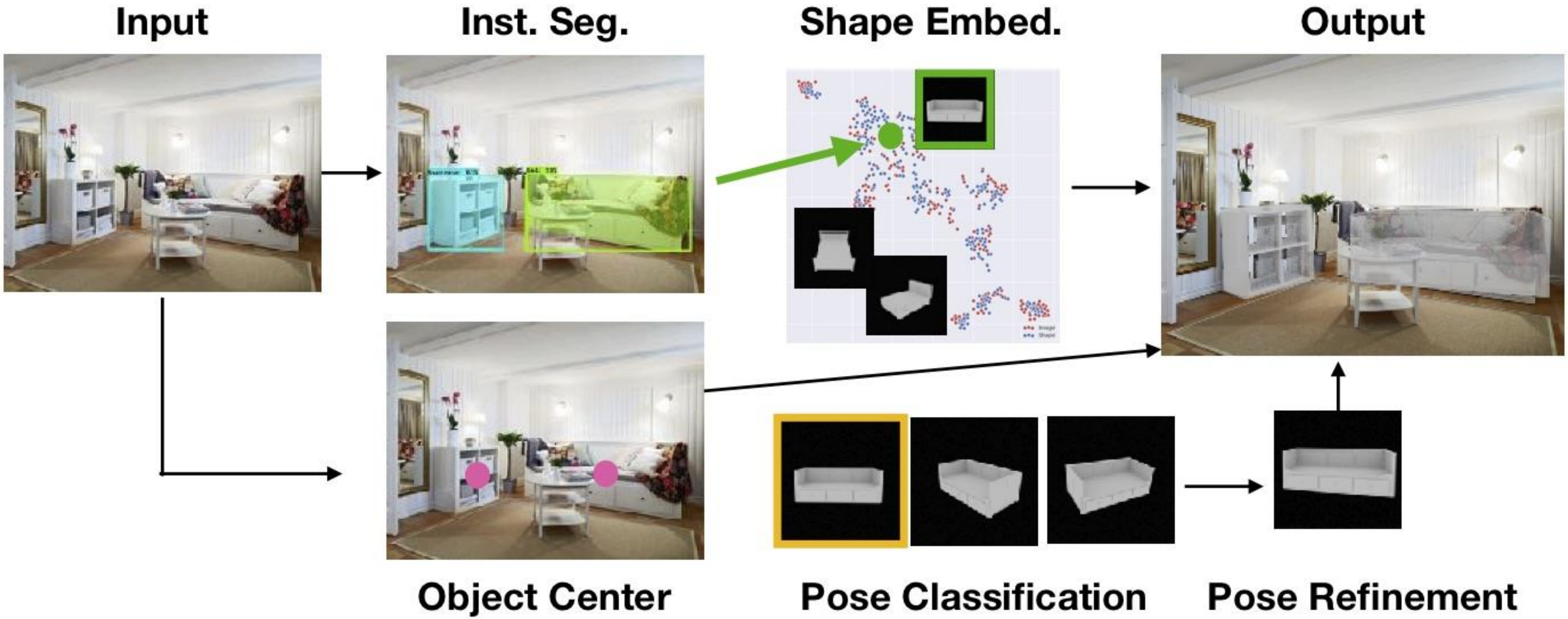
[Rozenberszki et al '22] ScanNet200

Learning to understand real-world scenes



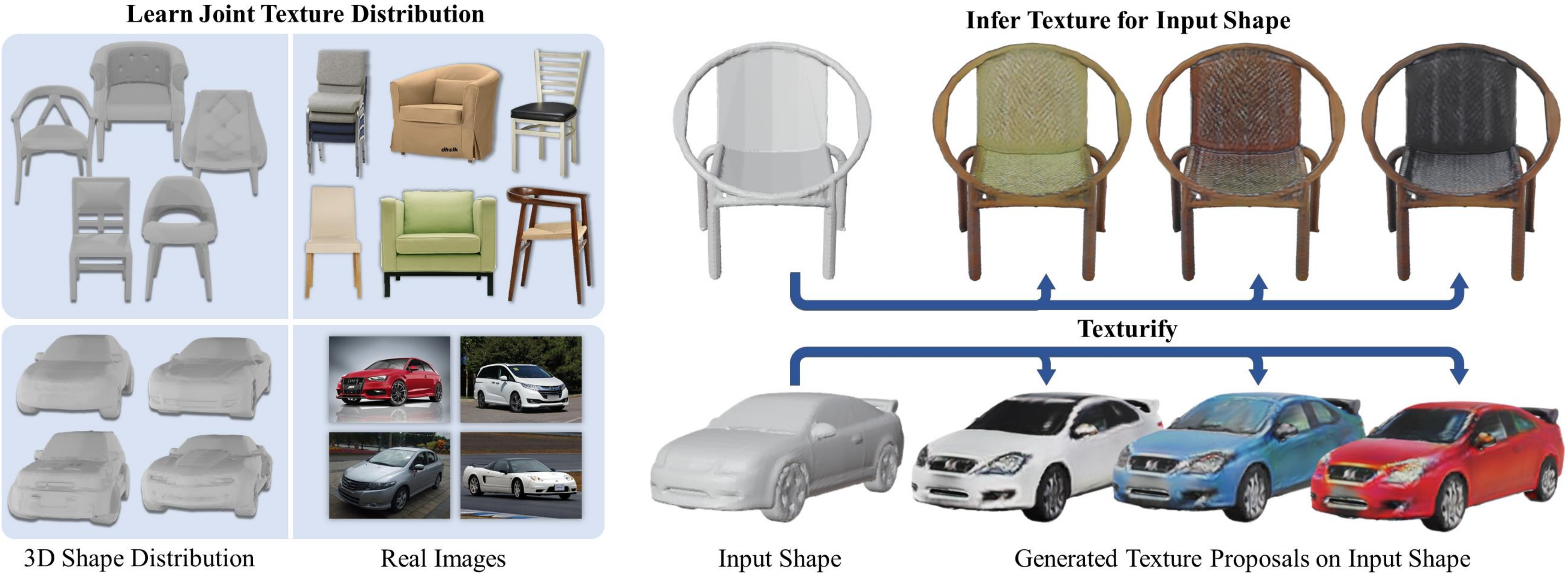
[Avetisyan et al '20] SceneCAD

Learning to understand real-world scenes



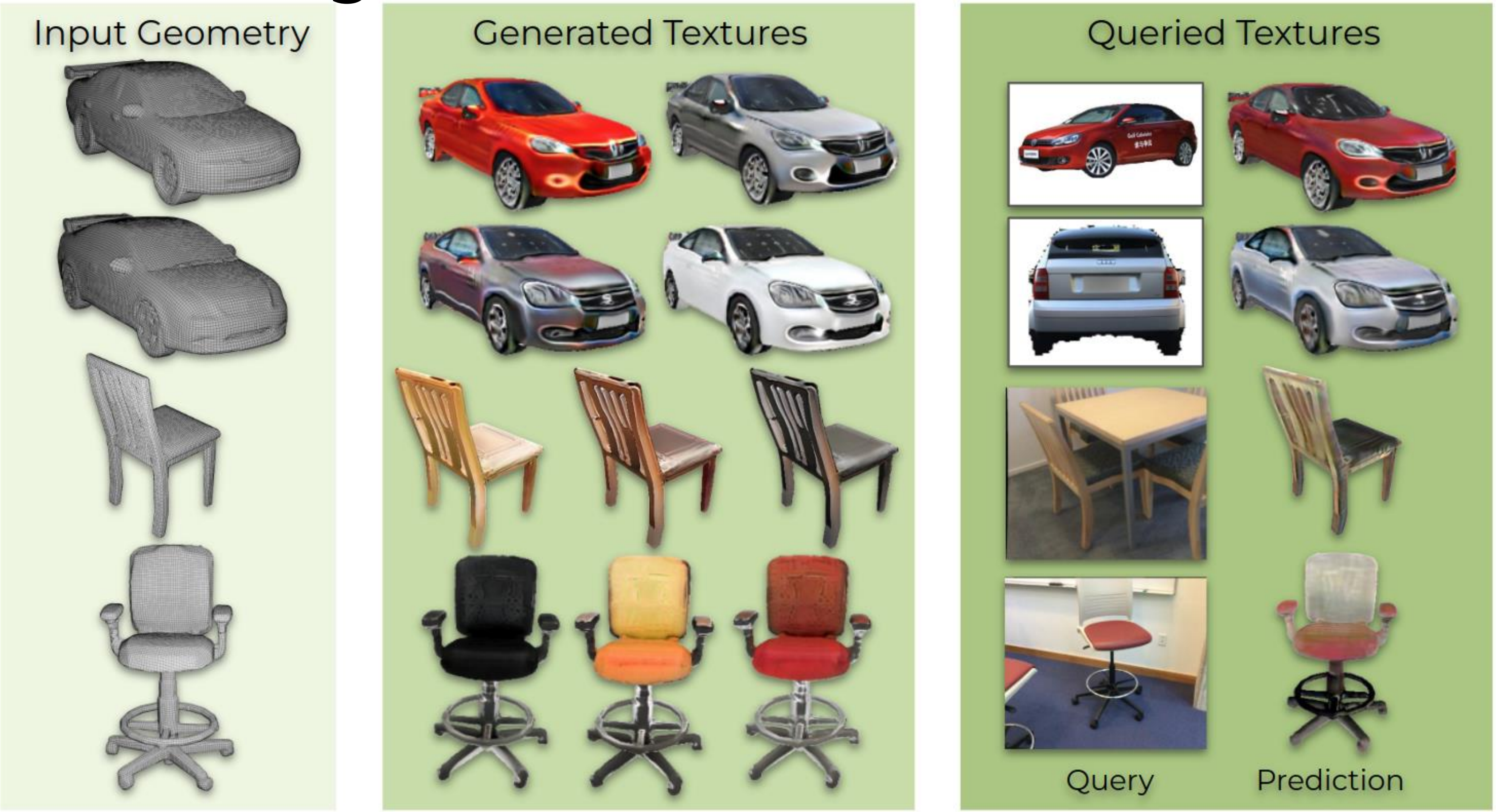
[Kuo et al ’20] Mask2CAD

Generating 3D Content



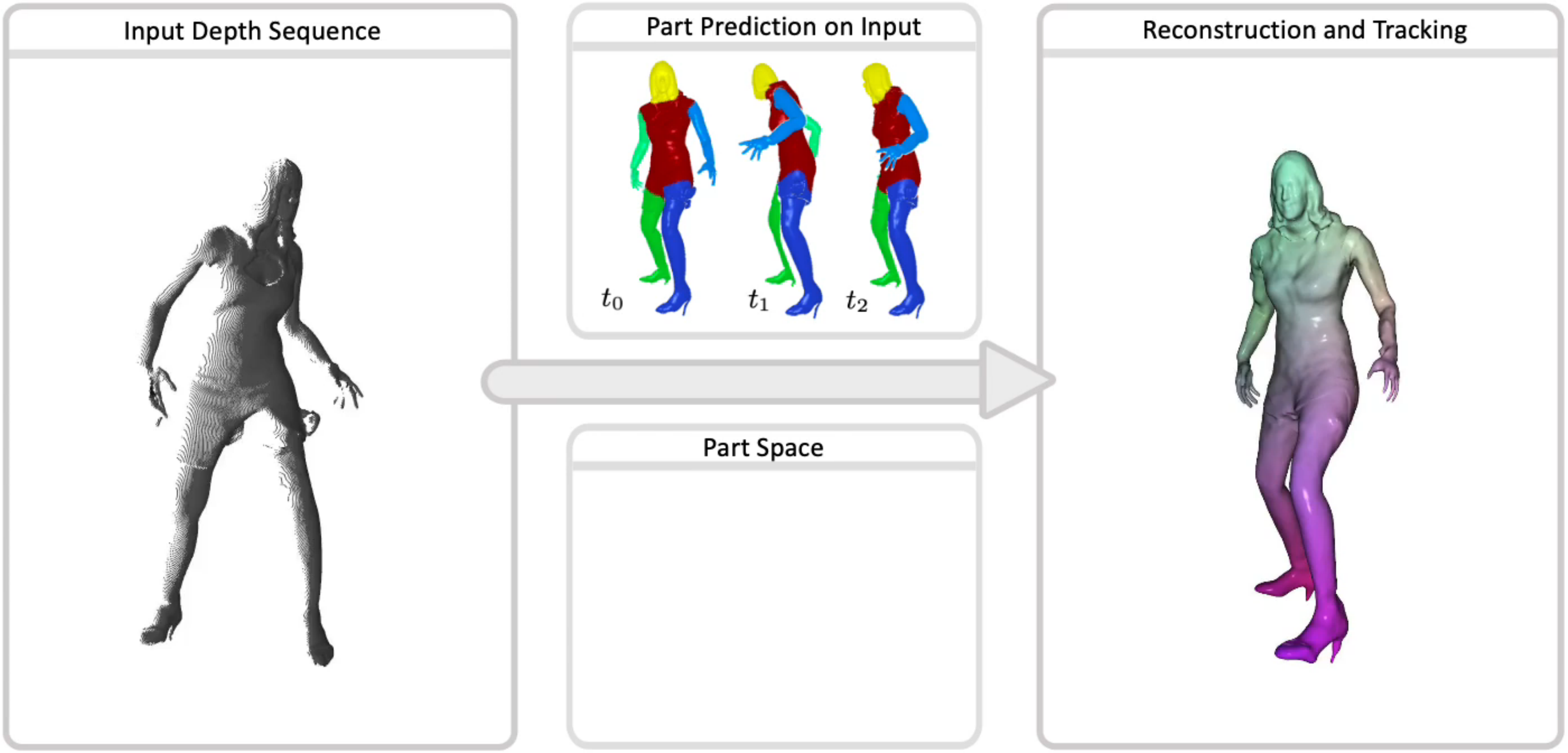
[Siddiqui et al ’22] Texturify

Generating 3D Content



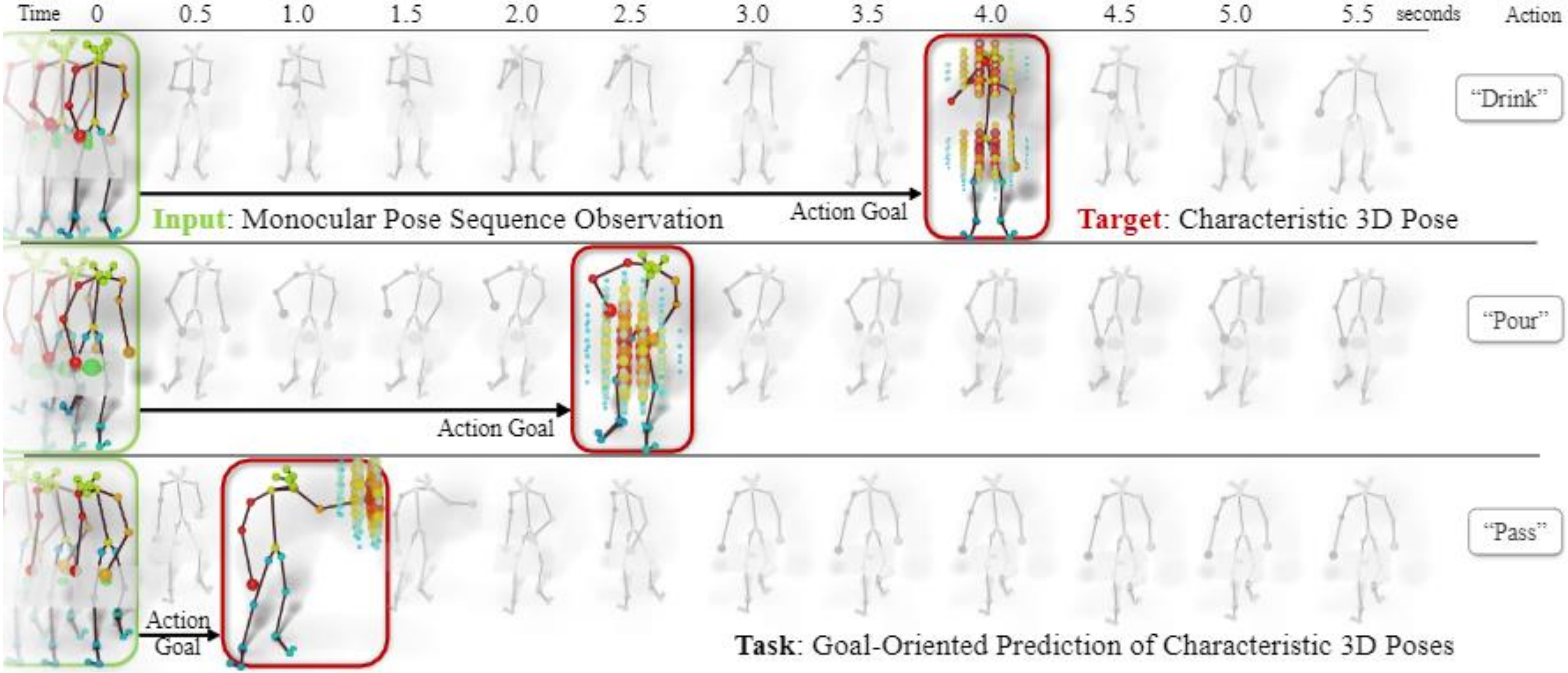
[Bokhovkin et al '23] Mesh2Tex

Understanding dynamics



[Palfox et al '22] SPAMs

Understanding people and functionality



[Diller et al '22] Forecasting Characteristic 3D Poses of Human Actions

Early machine perception

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

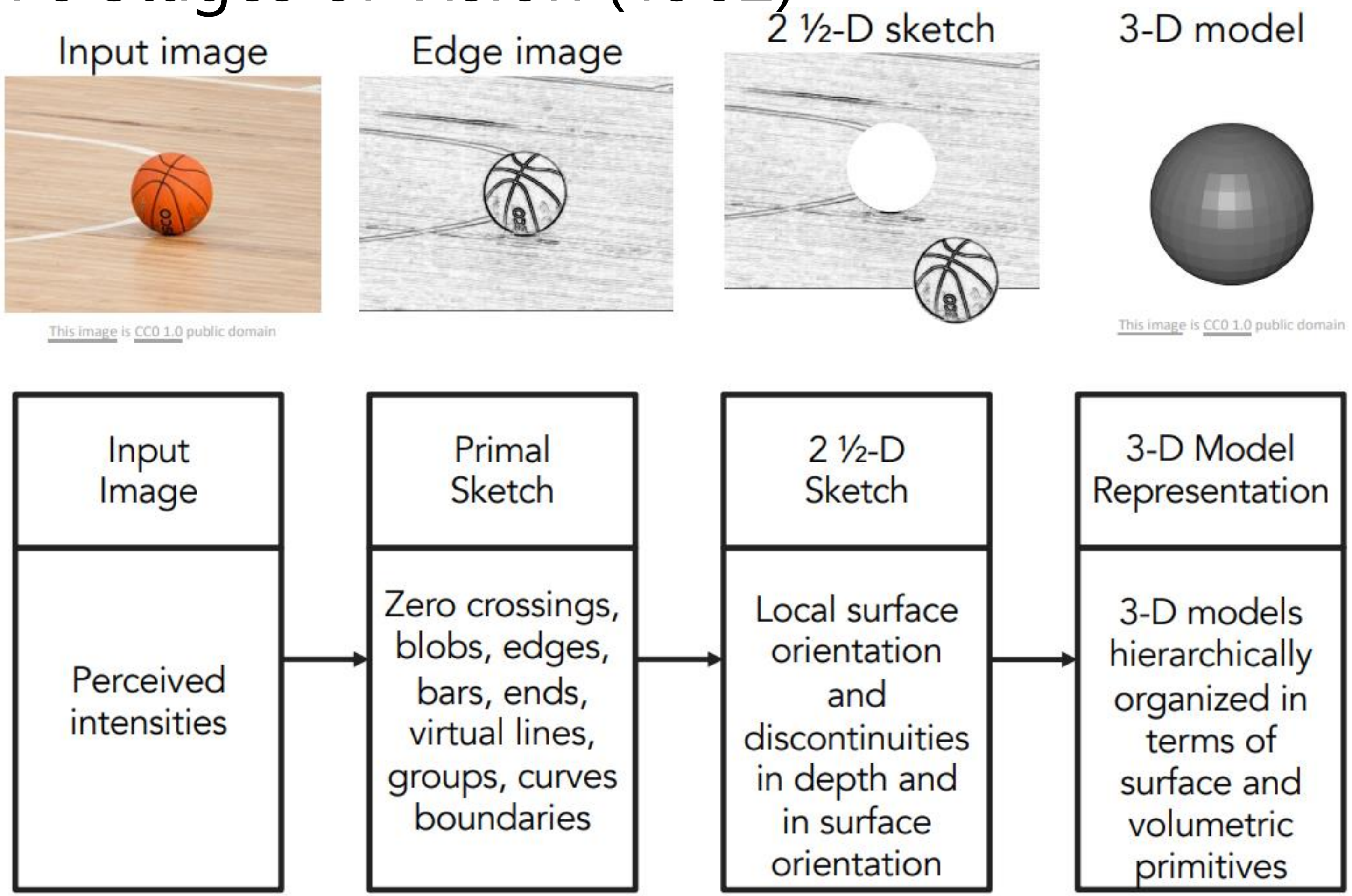
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

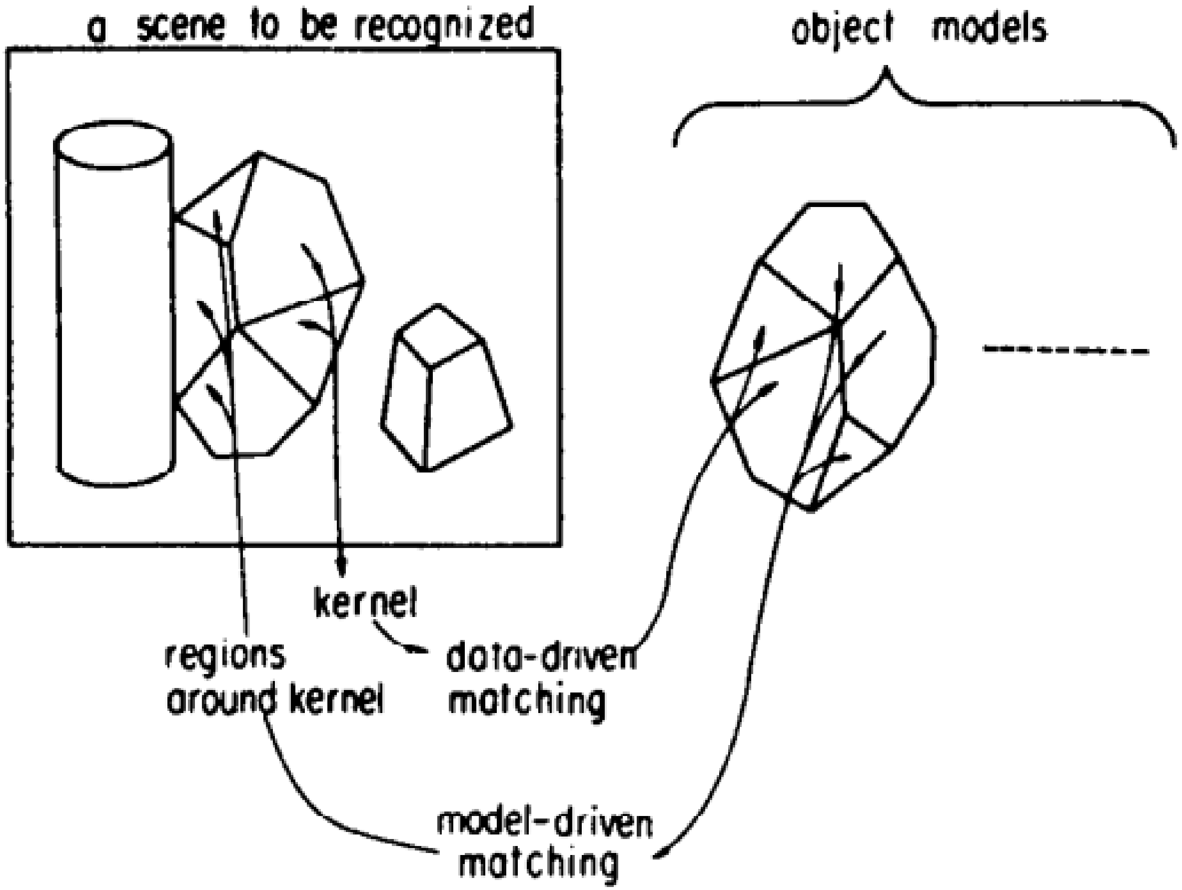
Early machine perception

- David Marr's Stages of Vision (1982)



images from Stanford cs231n

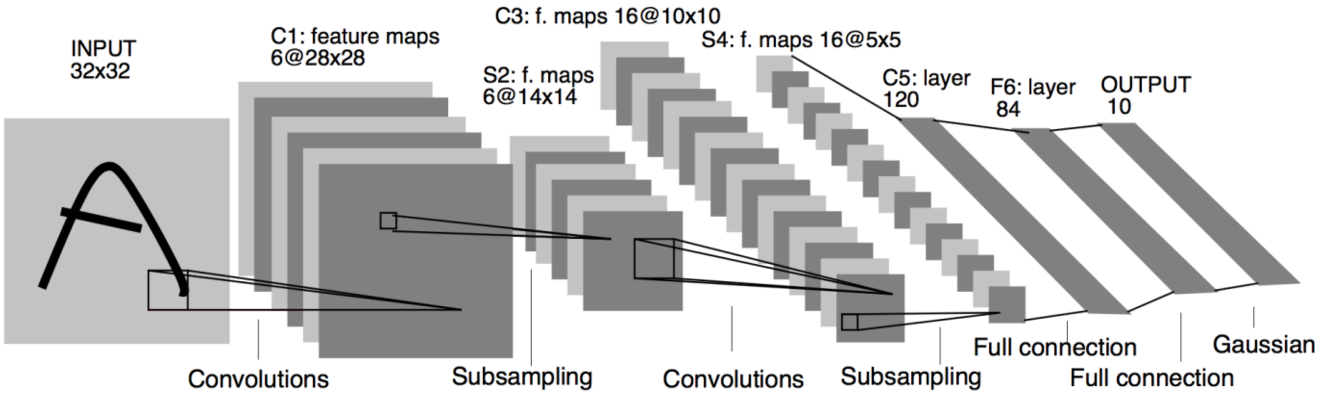
Early machine perception



[Chin et al. ’86] Model-based Recognition in Robot Vison

(fast forward)

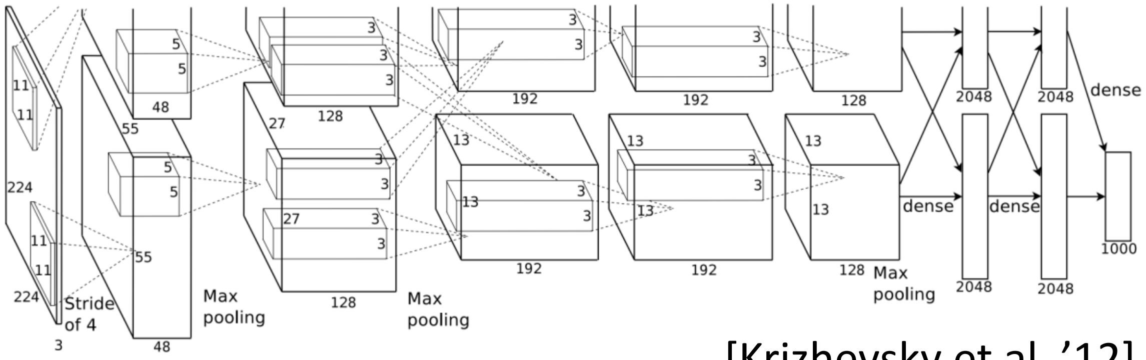
Today:



[Lecun et al. ’98]

NVIDIA

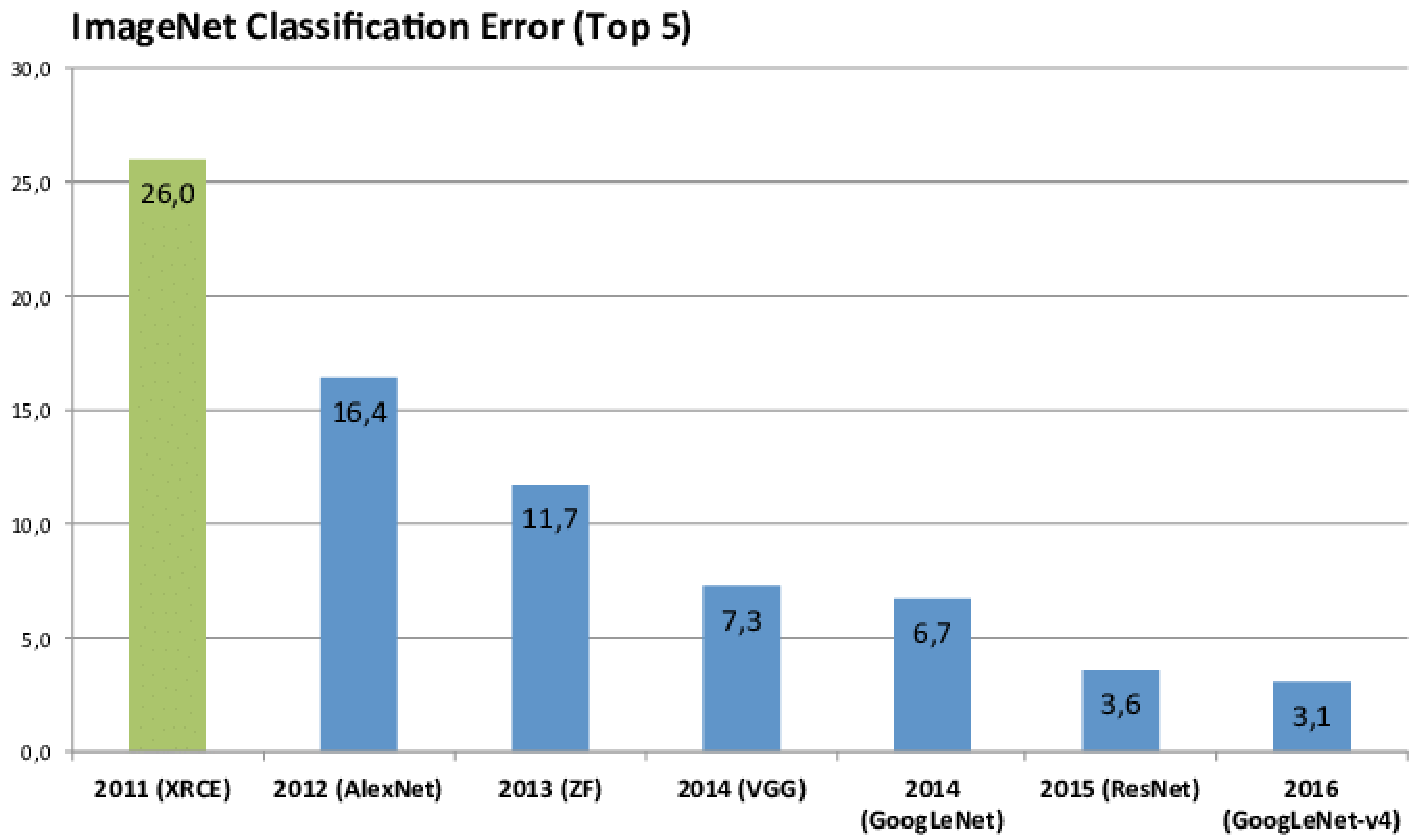
[Deng et al. ’09; Russakovsky et al. ‘14]



[Krizhevsky et al. ’12]

Many successes in 2D

- Image Classification



Many successes in 2D

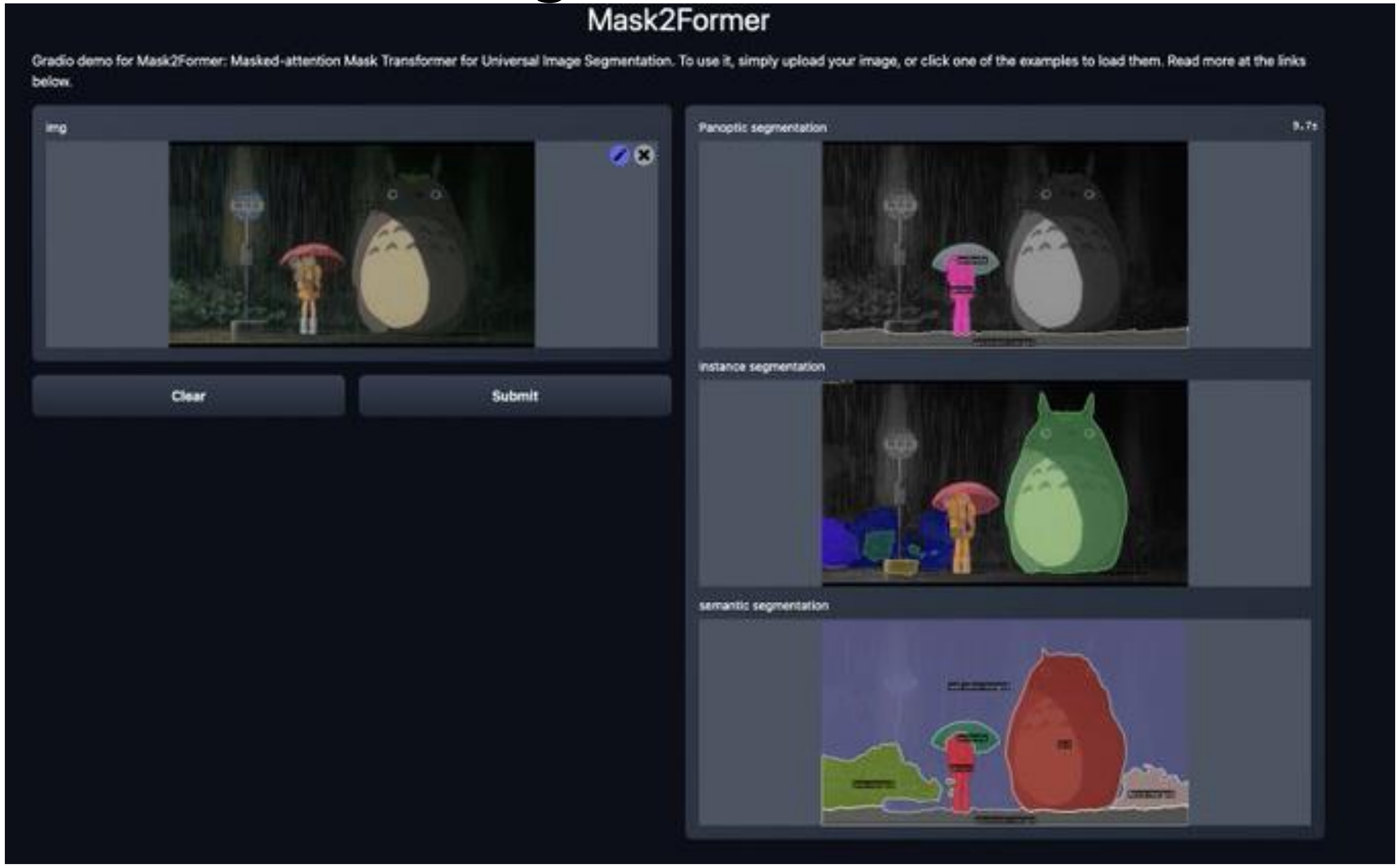
- Object Recognition

github.com/matterport/Mask_RCNN

[He et al. '17]
Mask R-CNN

Many successes in 2D

- 2D Panoptic Understanding

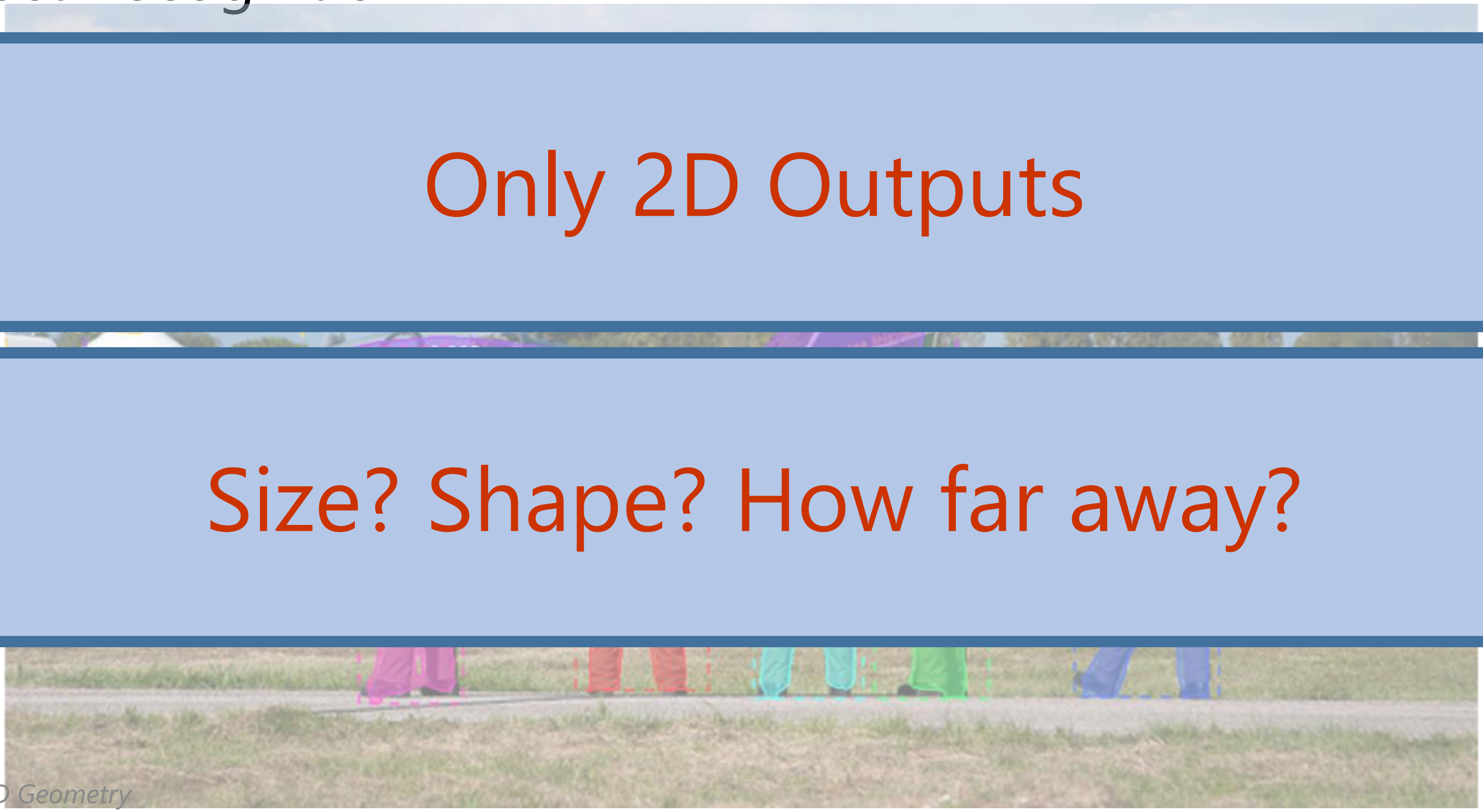


https://github.com/facebookresearch/Mask2Former

[Cheng et al. '22]
Mask2Former

Many successes in 2D

- Object Recognition

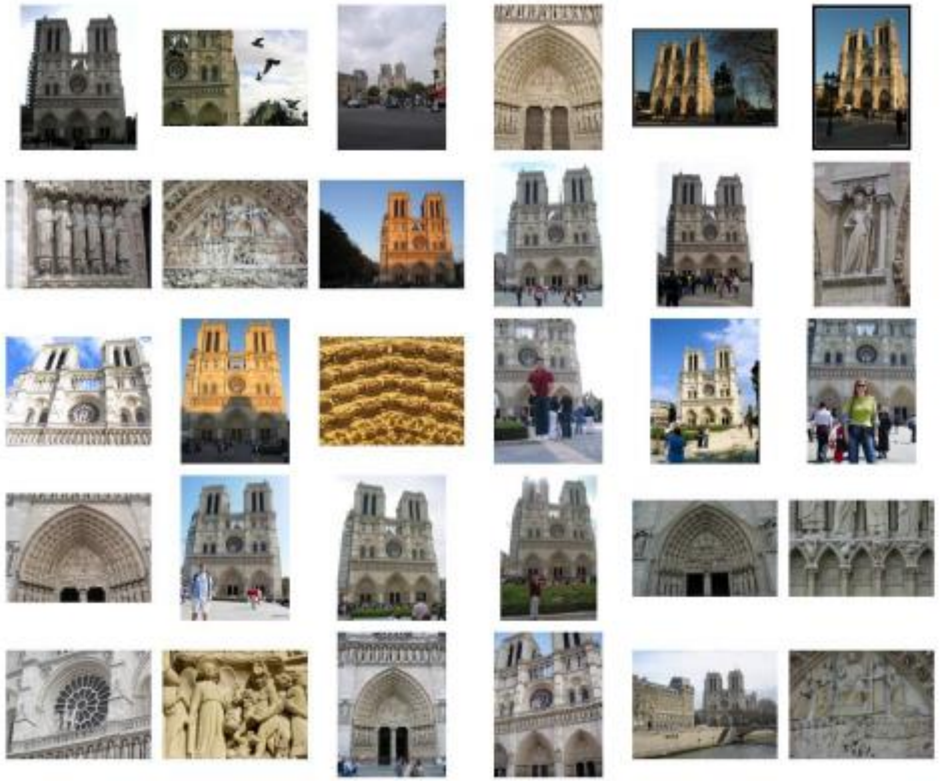


Only 2D Outputs

Size? Shape? How far away?

Understanding 3D

- Multi-View 3D Reconstruction



[Snavely et al. '06] Phototourism

Understanding 3D

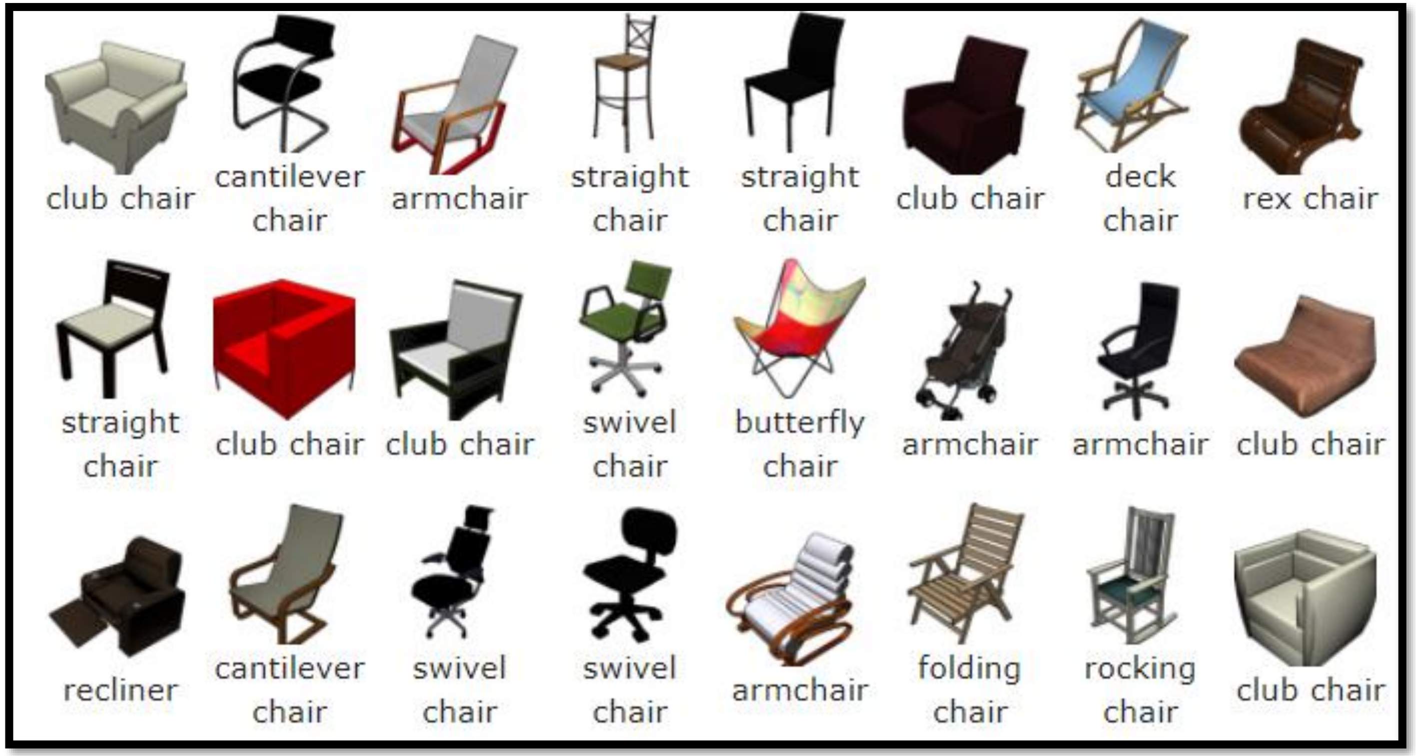
- RGB-D Reconstruction

[Newcombe et al '11; Izadi et al '11] KinectFusion

[Dai et al. '17] BundleFusion

Understanding 3D

- 3D Shape and Scene Datasets

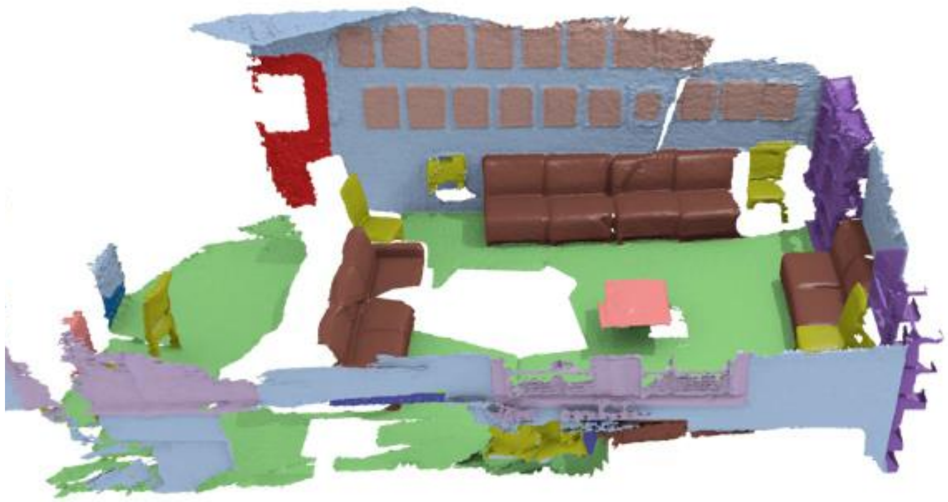


[Chang et al. '15] ShapeNet

[Dai et al. '17] ScanNet

Understanding 3D

- 3D Semantic Understanding

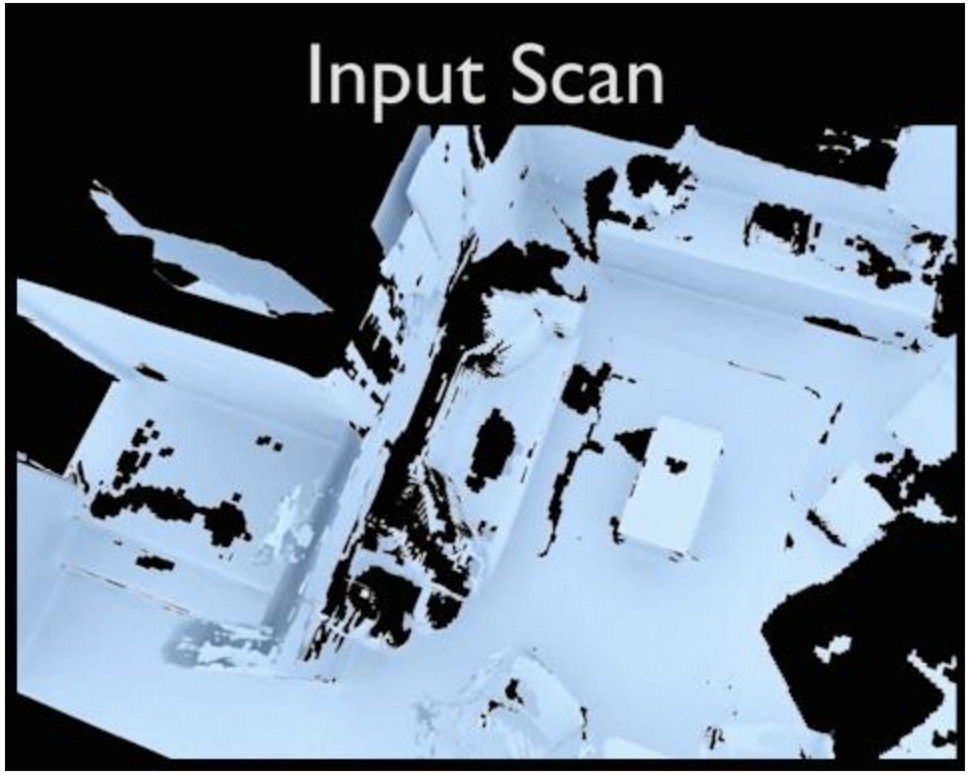


[Kundu et al ’20] Virtual MVFusion

Cool Challenges

- How to efficiently generate 3D? What representations?

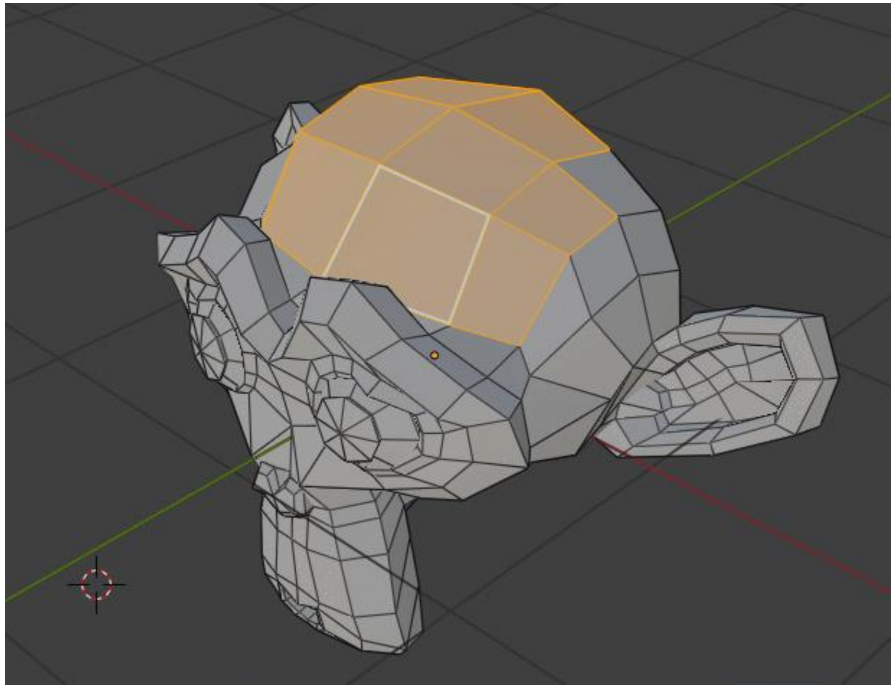
[Mescheder et al. ’19] Occupancy Networks

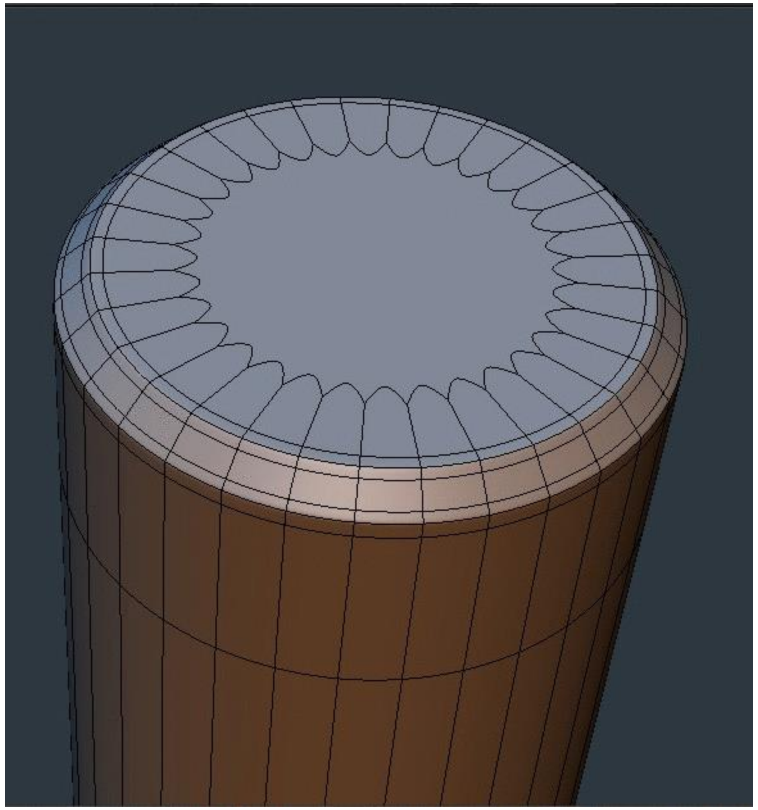


[Dai et al. ’20] SG-NN

Cool Challenges

- What about how people model 3D?

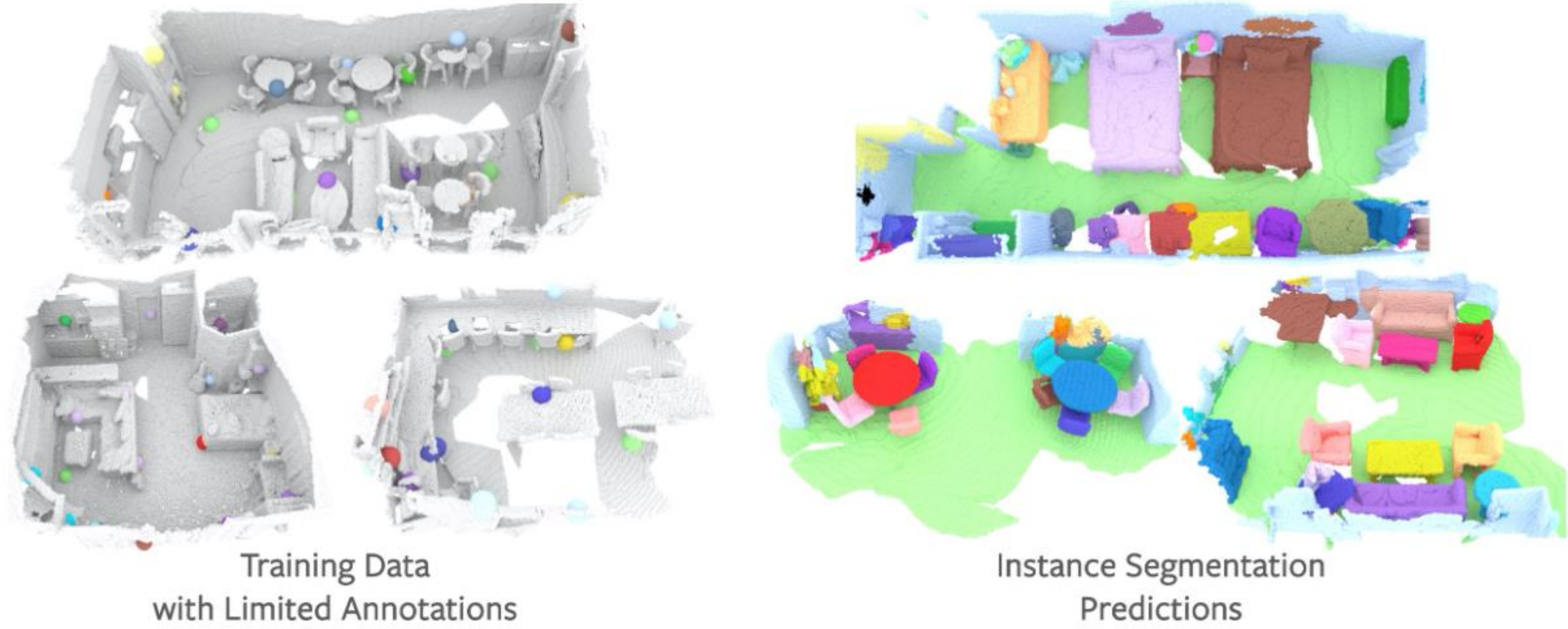




Blender (https://www.blender.org/)

Cool Challenges

- How to train with limited supervision?



[Hou et al '21] Exploring Data Efficient 3D Scene Understanding

Cool Challenges

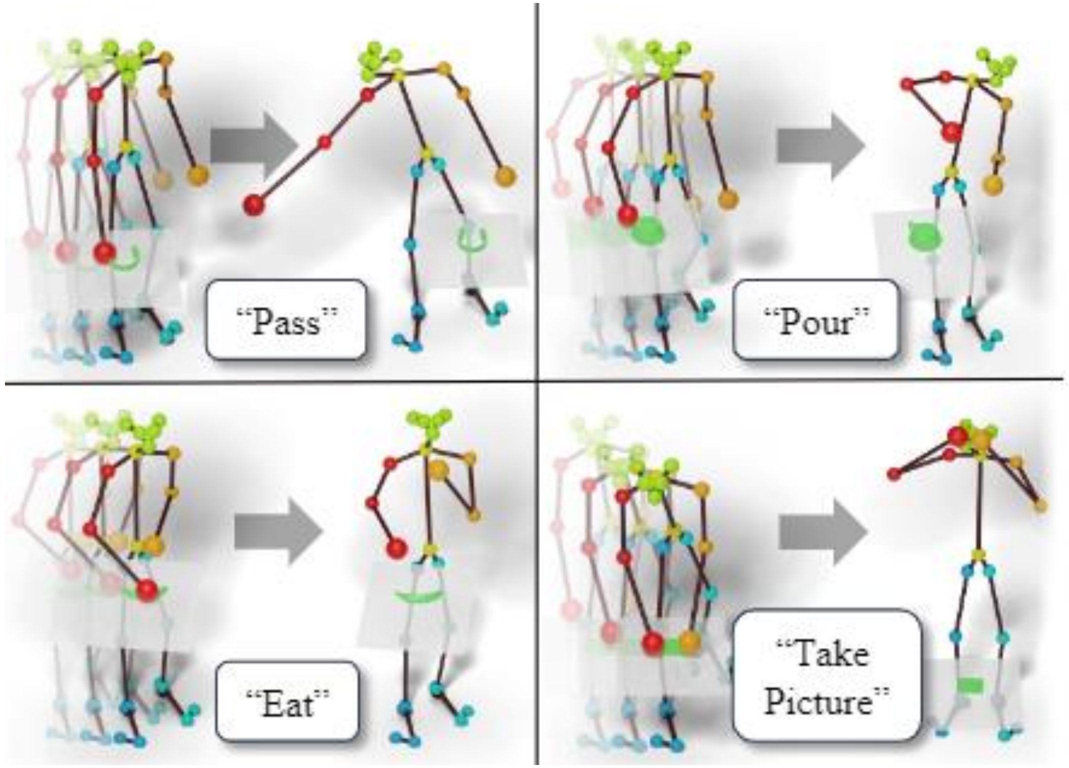
- How to train with *no* supervision?

[Rozenberszki et al '23] UnScene3D

Cool Challenges

- How to predict human-like behavior and functions?

[Savva et al '16] PiGraphs



[Diller et al '22] Characteristic 3D Poses

Lecture Topics

- Basics of 3D Representations
- Shapes: Alignment, Descriptors, Similarity
- Basics of Machine Learning
- Shapes Segmentation and Labeling
- Reconstructing and Generating Shapes

Lecture Topics

- Learning on 3D Representations: implicit representations, radiance fields, triplanes, coordinate networks, etc.
- Semantic Scene Understanding
- Reconstructing and Generating Scenes
- Differentiable Rendering for 3D Understanding
- Functional Analysis of Scenes

Organization

- Lectures and Tutorials
  - Knowledge of linear algebra (e.g., affine transforms, rigid transforms, rank, eigenvectors)
  - Introductory knowledge of deep learning
  - Python

Organization

- Lectures and Tutorials
  - Online videos
- Tutorials
  - Help on exercises and projects
- Office Hours
  - Contact TAs for zoom appointments
  - Use Moodle!

Tutorials

- Exercise: Basics of 3D
- Exercise: Shape Classification
- Exercise: Shape Generation

- Groups of two allowed
- Submit & pass all exercises for 0.3 bonus

Final Project

- Start: mid-semester
- 5+ weeks
- Groups of 3-4
- Proposal (1-2 pages)
- Project presentation and report

Final Exam

- Written Exam
- Questions covering general lecture content
- Questions about final project

Logistics

TAs:
- Christian Diller christian.diller@tum.de
- Can Guemeli can.guemeli@tum.de
- Jiapeng Tang jiapeng.tang@tum.de

Lectures:
- Online Videos + Office Hours

Tutorials:
- Uploaded with lectures
- Focus on exercises and projects

Office Hours:
- Email TAs for zoom appointments

Exercises:
- In total, 3 exercises with 1-2 weeks each for completion
- All exercises completed and passed gives 0.3 bonus on final grade
- Exact deadlines are the day before the lecture at 23:55

Projects:
- 1 final project with 5+ weeks for completion
- Project features in final exam
- Mid-semester: Introduction to projects
- Final exam date: deadline for projects