6.S980

MACHINE LEARNING FOR INVERSE GRAPHICS

6.S980

Prof. Vincent Sitzmann

About Me

I want to build algorithms that can learn to perceive the world the way you can!

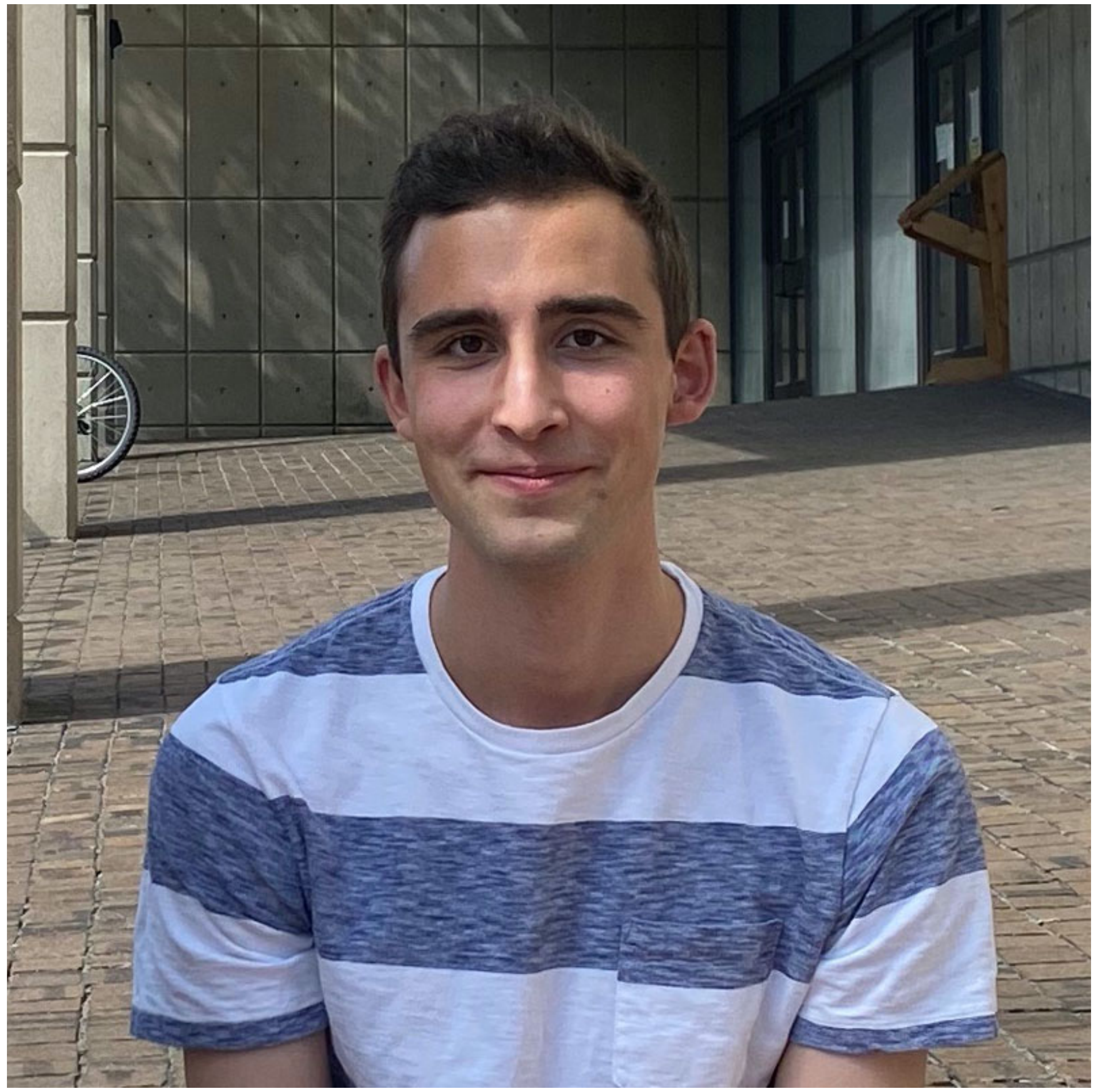
My research group @ MIT:

SCENE
REPRESENTATION
GROUP

Ulm -> TU Munich -> Stanford -> MIT

Vincent Sitzmann

TA: David Charatan



PhD Year: 2nd

Advisors: Vincent Sitzmann

Administrativa

Enrollment

Enrollment

- I have scheduling conflicts.
  - *We'll try to get all materials (videos/slides) online (but do attend when you can!)*

Enrollment

- I have scheduling conflicts.
  - *We'll try to get all materials (videos/slides) online (but do attend when you can!)*

- I am not yet sure if this course is for me…
  - *Hopefully today can help you decide.*

Prerequisites

- **Proficiency in Python**: All class assignments are Google Colabs.
- **Proficiency in Numpy:** You need to be able to manipulate multi-dim arrays.
- **Deep Learning:** Training NNs, CNNs in PyTorch / TF / Jax, Stochastic Gradient Descent, Loss Functions, Optimization, Backpropagation
- **Linear Algebra**: matrix-vector products, matrix-matrix products, norms, SVD

Related Courses & Credits

- CMU 16-889: Learning for 3D Vision
  Prof. Shubham Tulsiani
- CMU 16-385: Computer Vision
  Prof. Kris Kitani
- MIT 6.819/6.869: Advances in Computer Vision,
  Profs. Bill Freeman, Phillip Isola, Antonio Torralba
- University of Amsterdam: Deep Learning II, Geometric Deep Learning
  Prof. Erik Bekkers
- Stanford CS348I: Computer Graphics in the Era of AI
  Profs. C. Karen Liu and Jiajun Wu
- University of Tübingen: Computer Vision
  Prof. Andreas Geiger
- **Ben Mildenhall:** Volume Rendering Slides

Related Courses & Credits

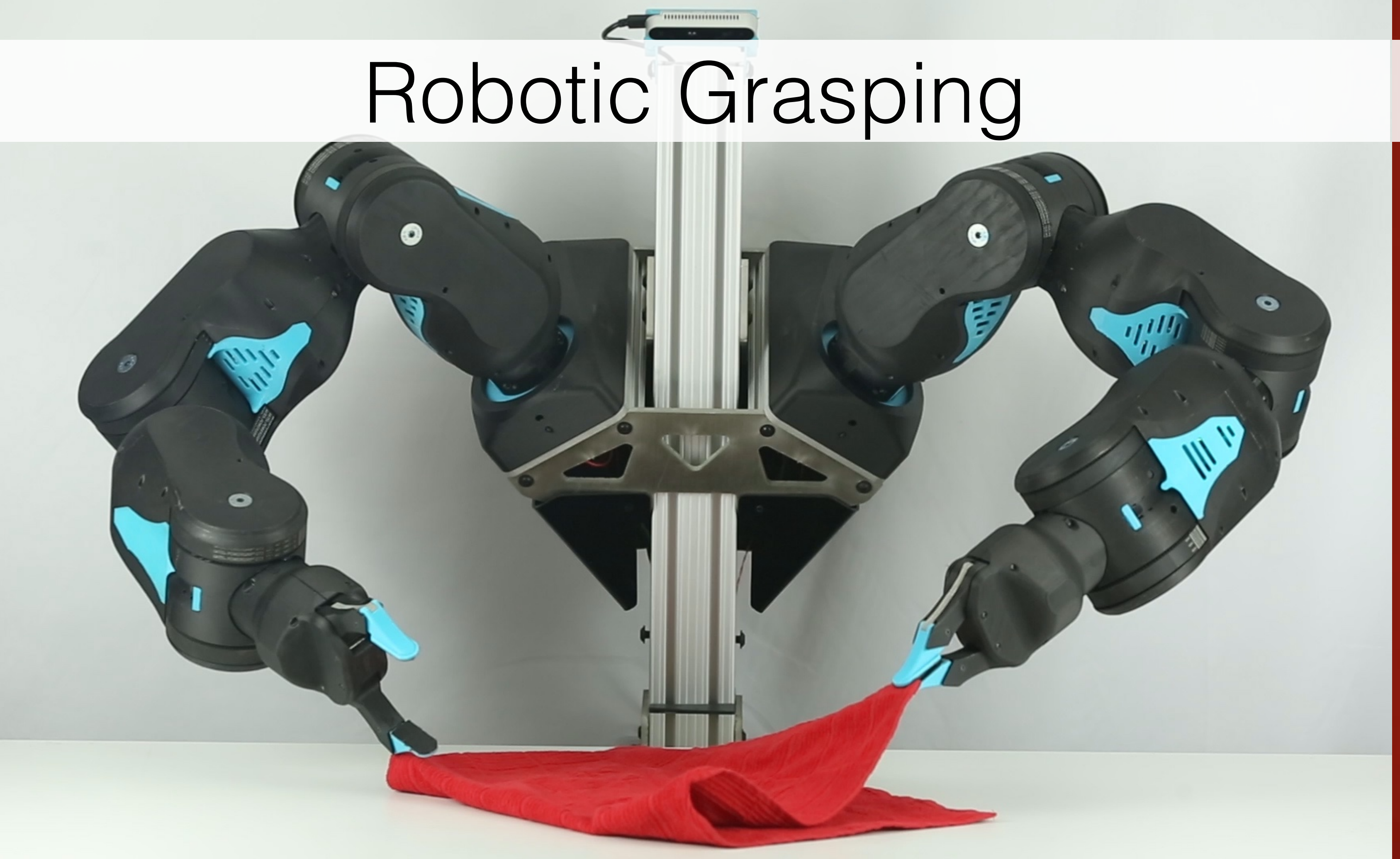
- CMU 16-889: Learning for 3D Vision
  Prof. S
- CMU
  Prof.
- MIT 6
  Profs.
- Unive
  Prof.
- Stanf
  Profs.
- University of Tübingen: Computer Vision
  Prof. Andreas Geiger
- **Ben Mildenhall:** Volume Rendering Slides

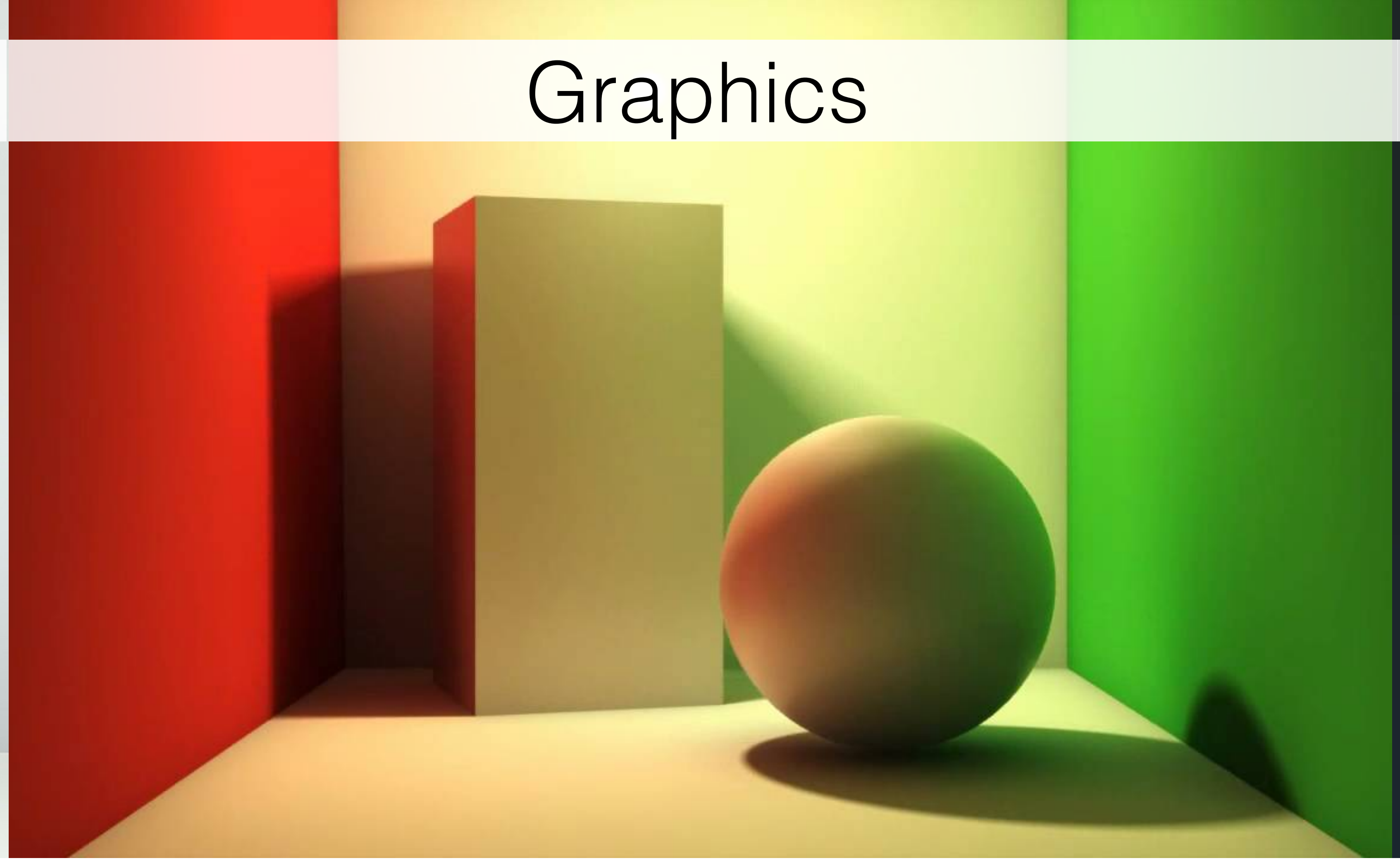
All of these folks let me use / adapt / build on top of their courses.
Look out for their mentions in the slide credits!

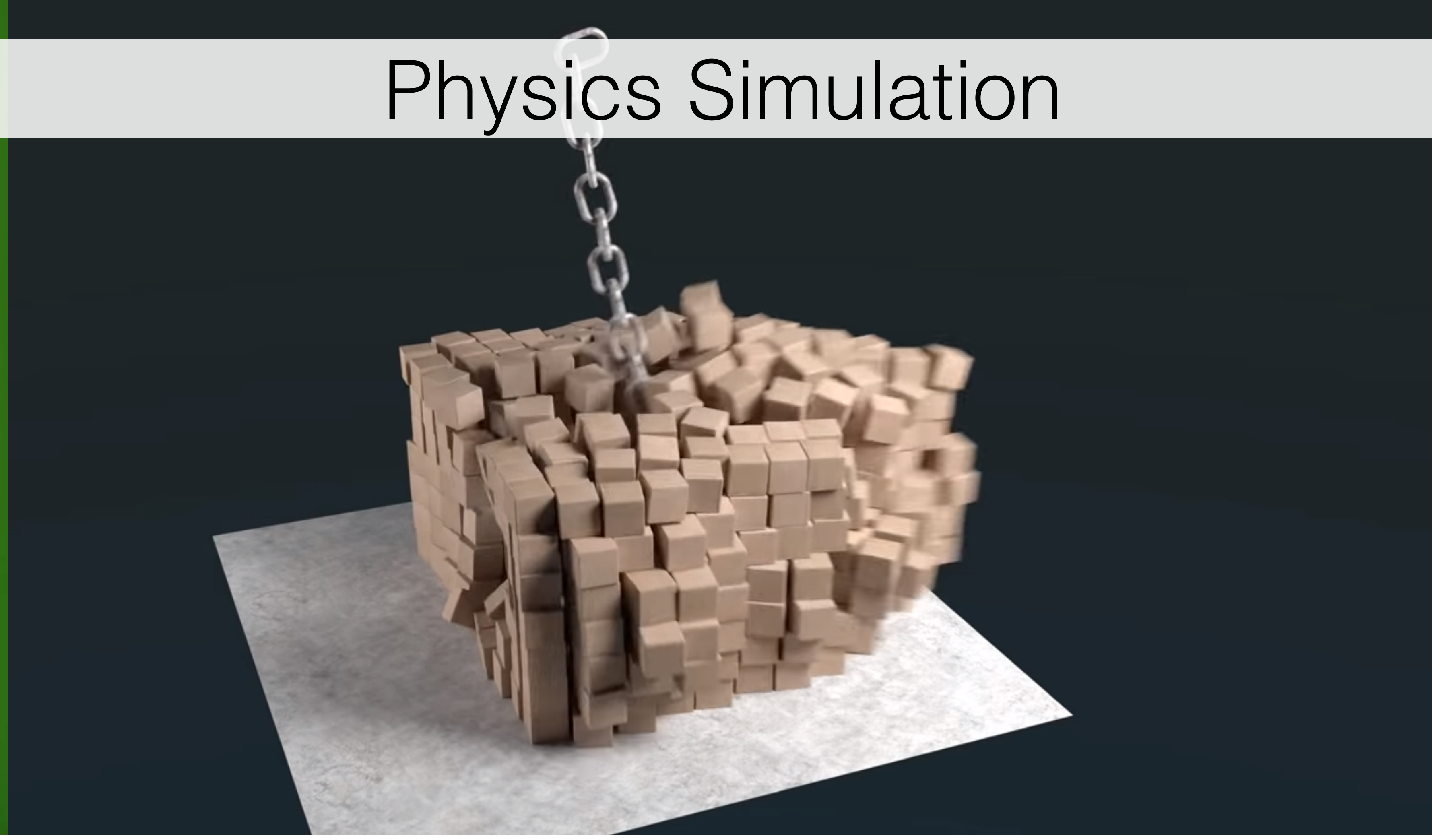
INVERSE GRAPHICS

LECTURE 0:

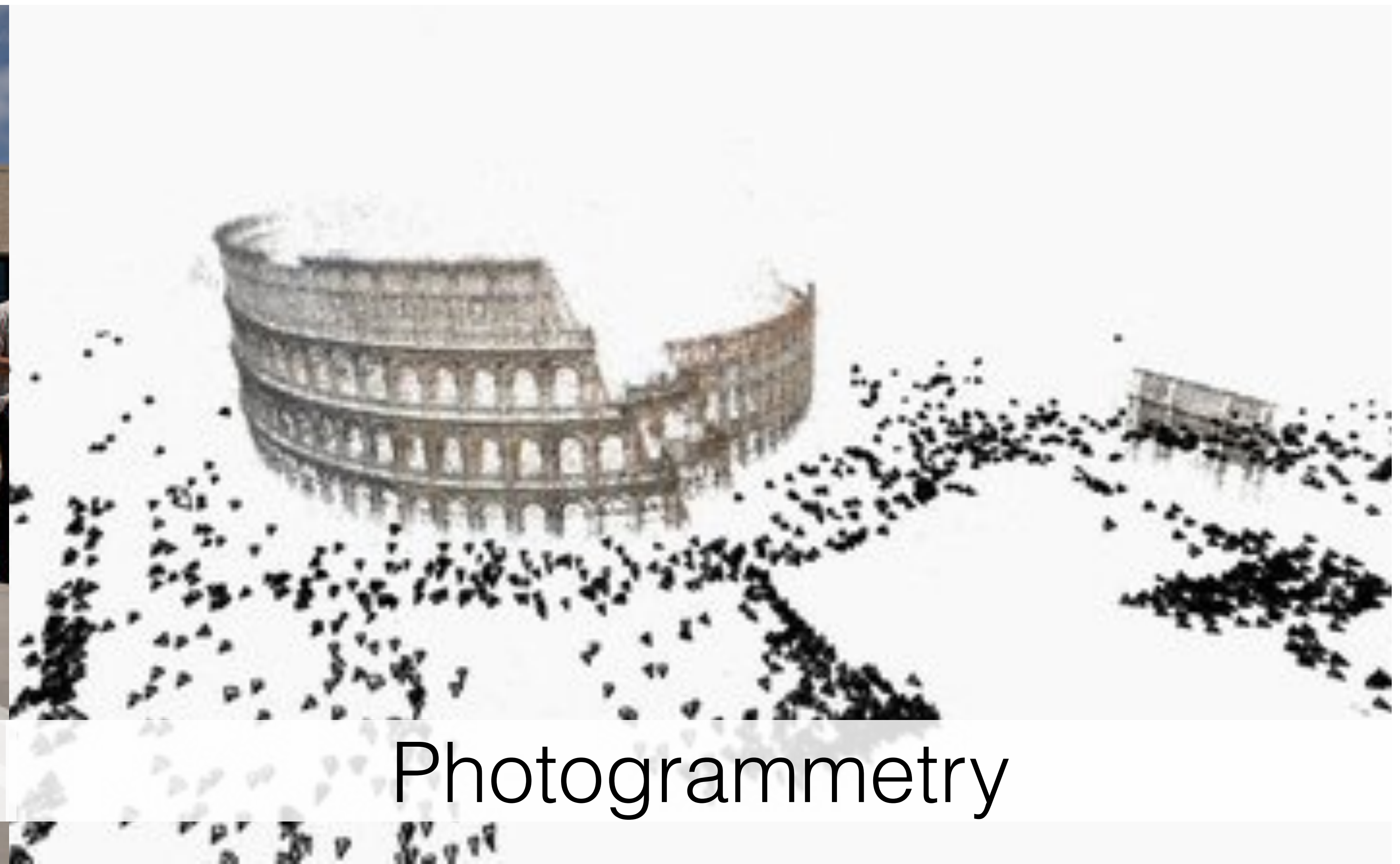
Introduction

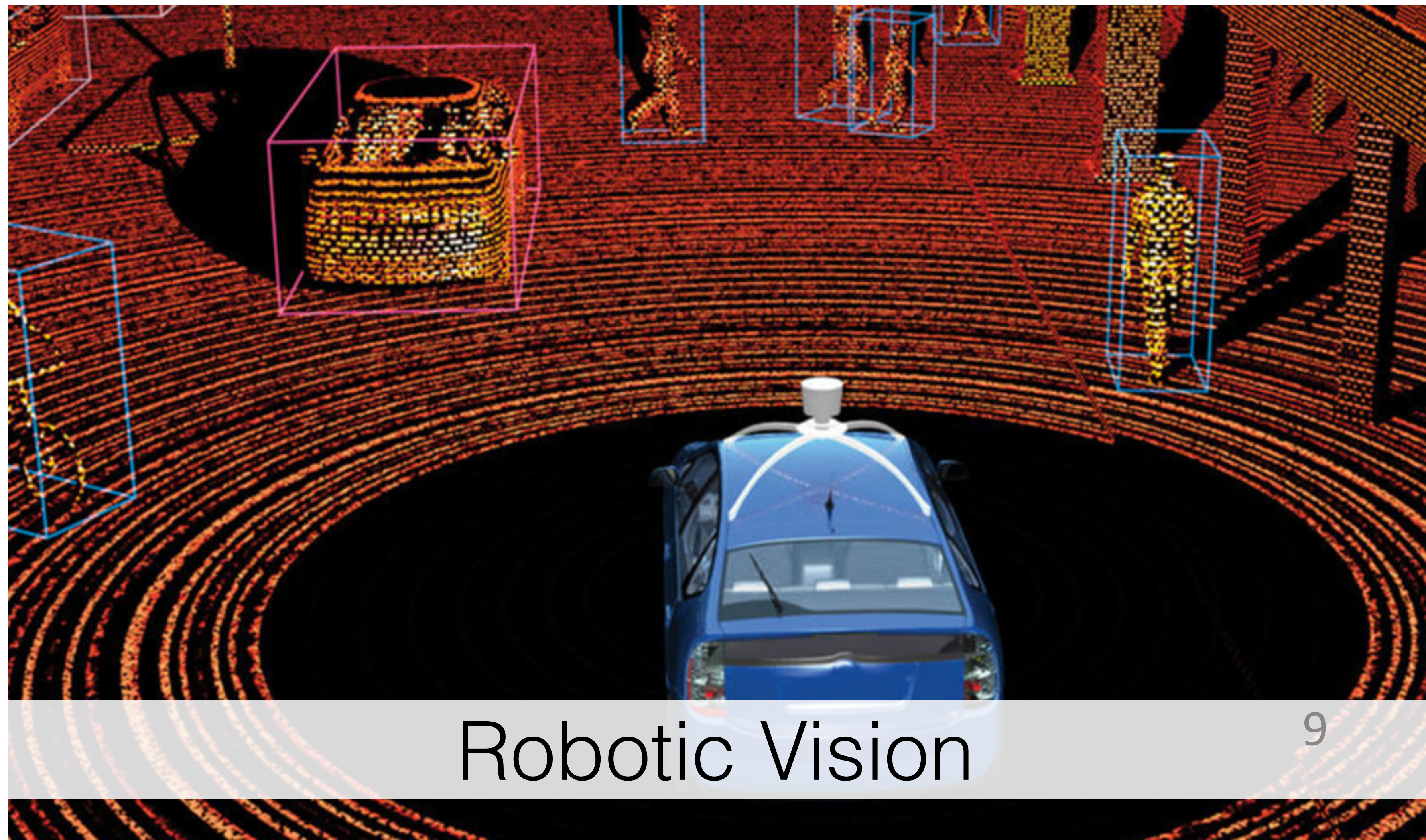


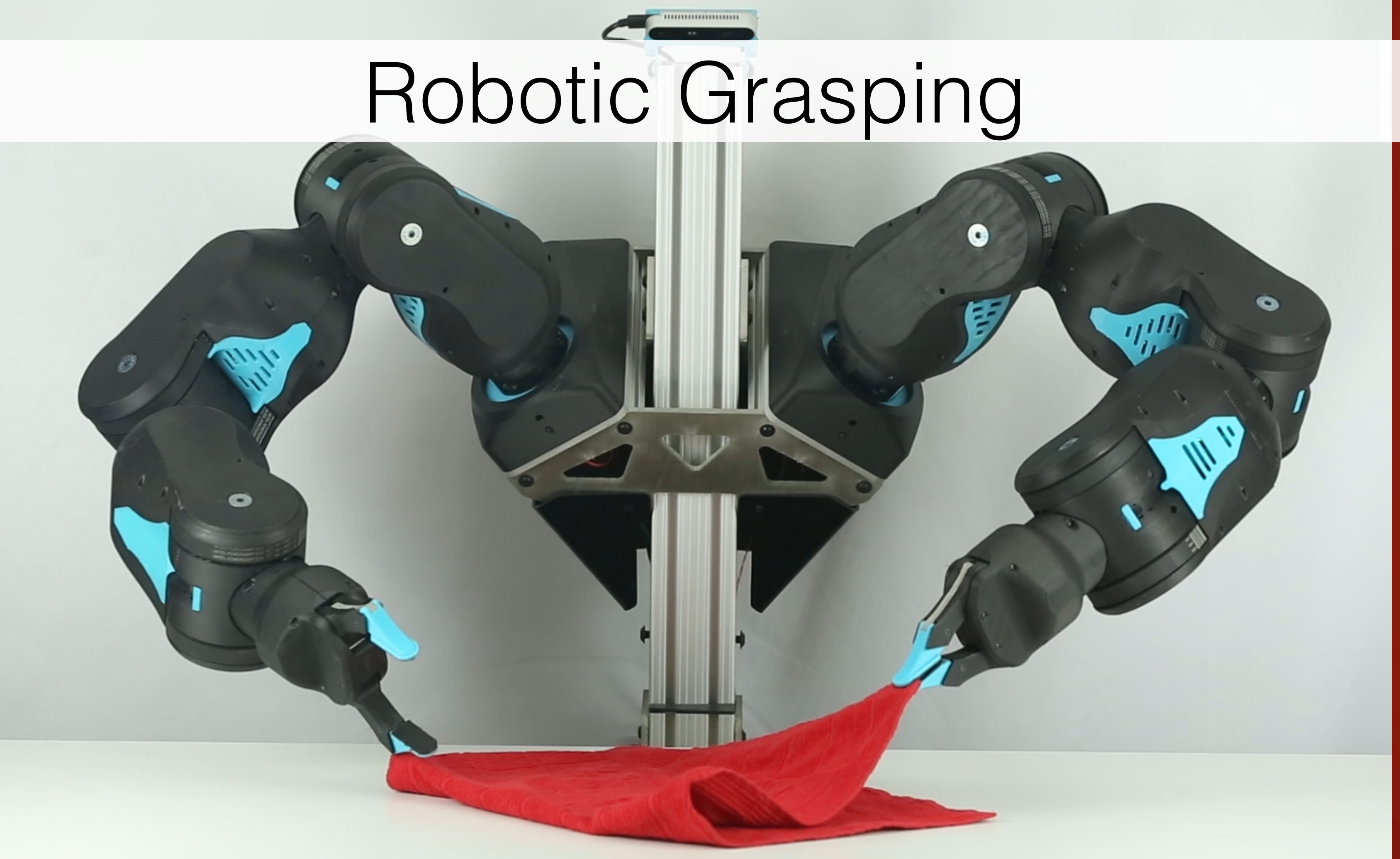


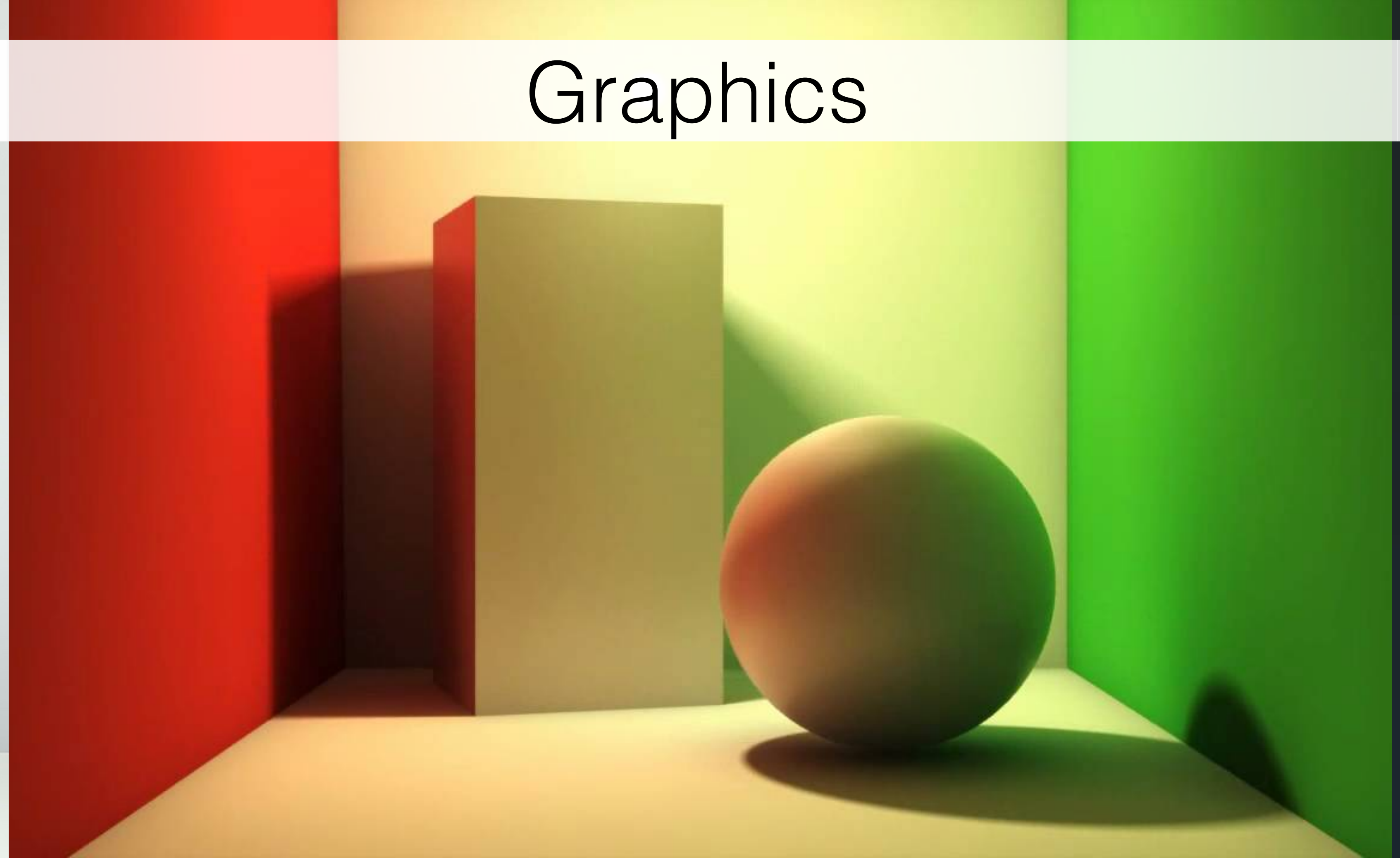


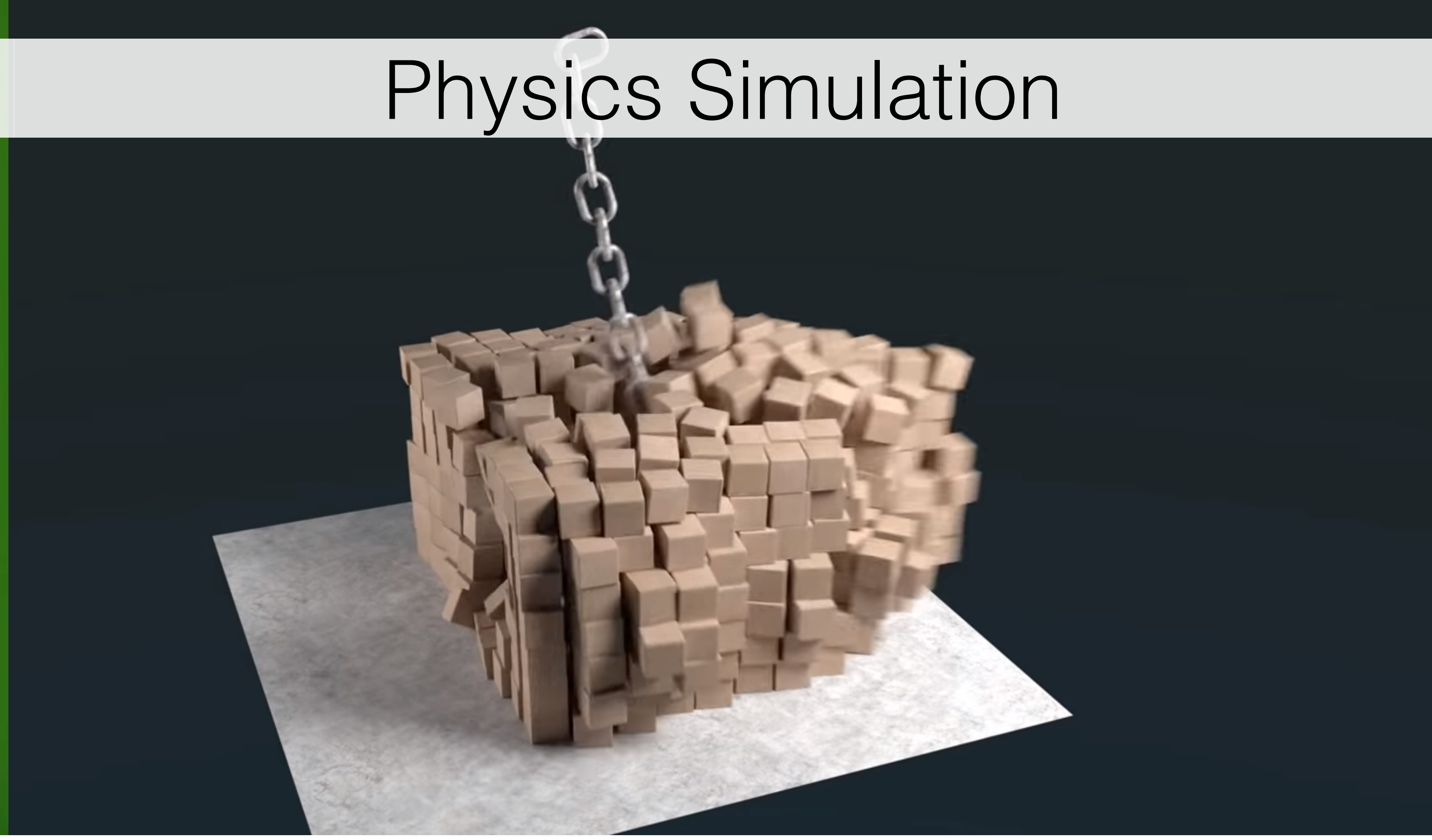
Scene Representation






Robotic Grasping

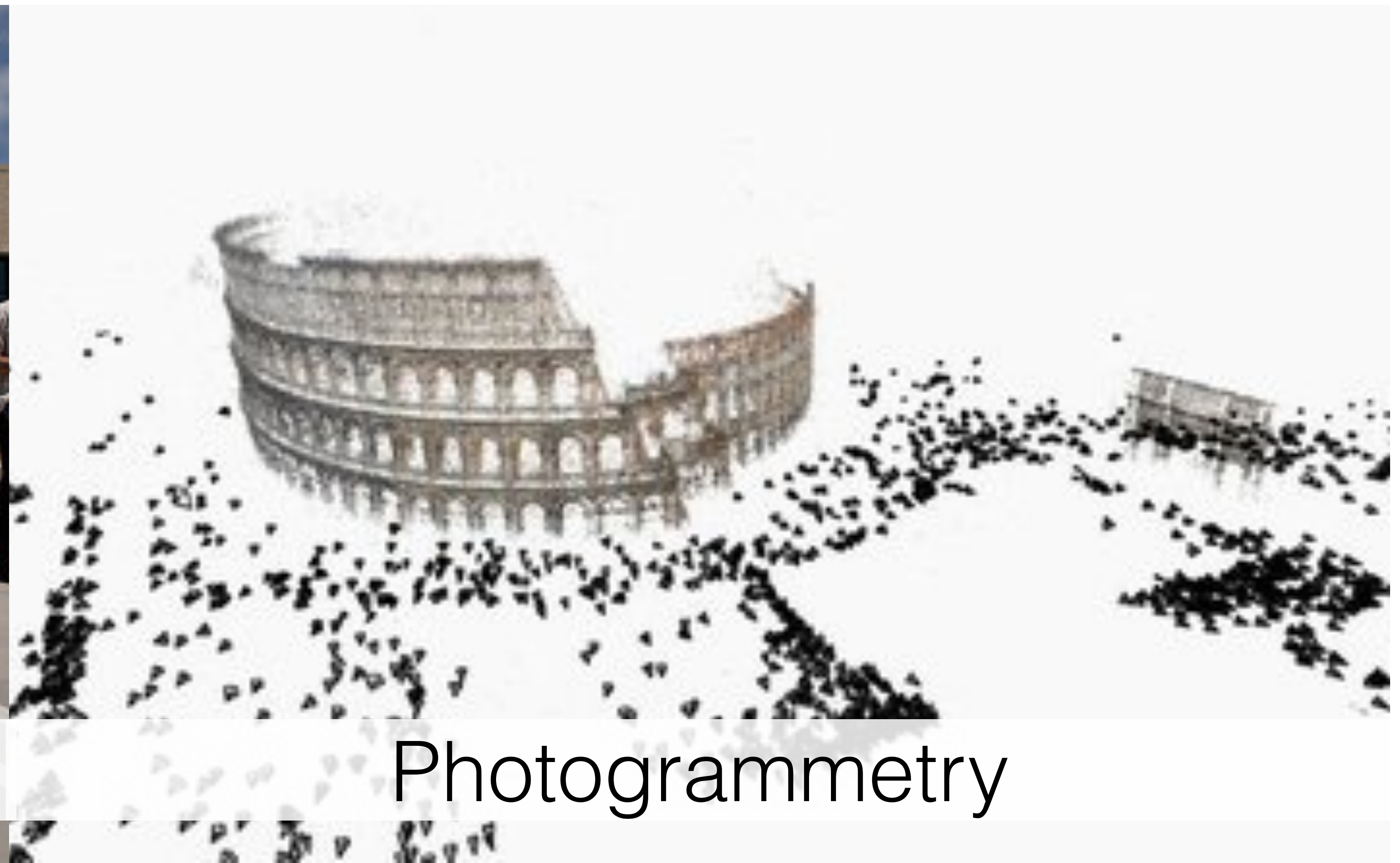

Graphics

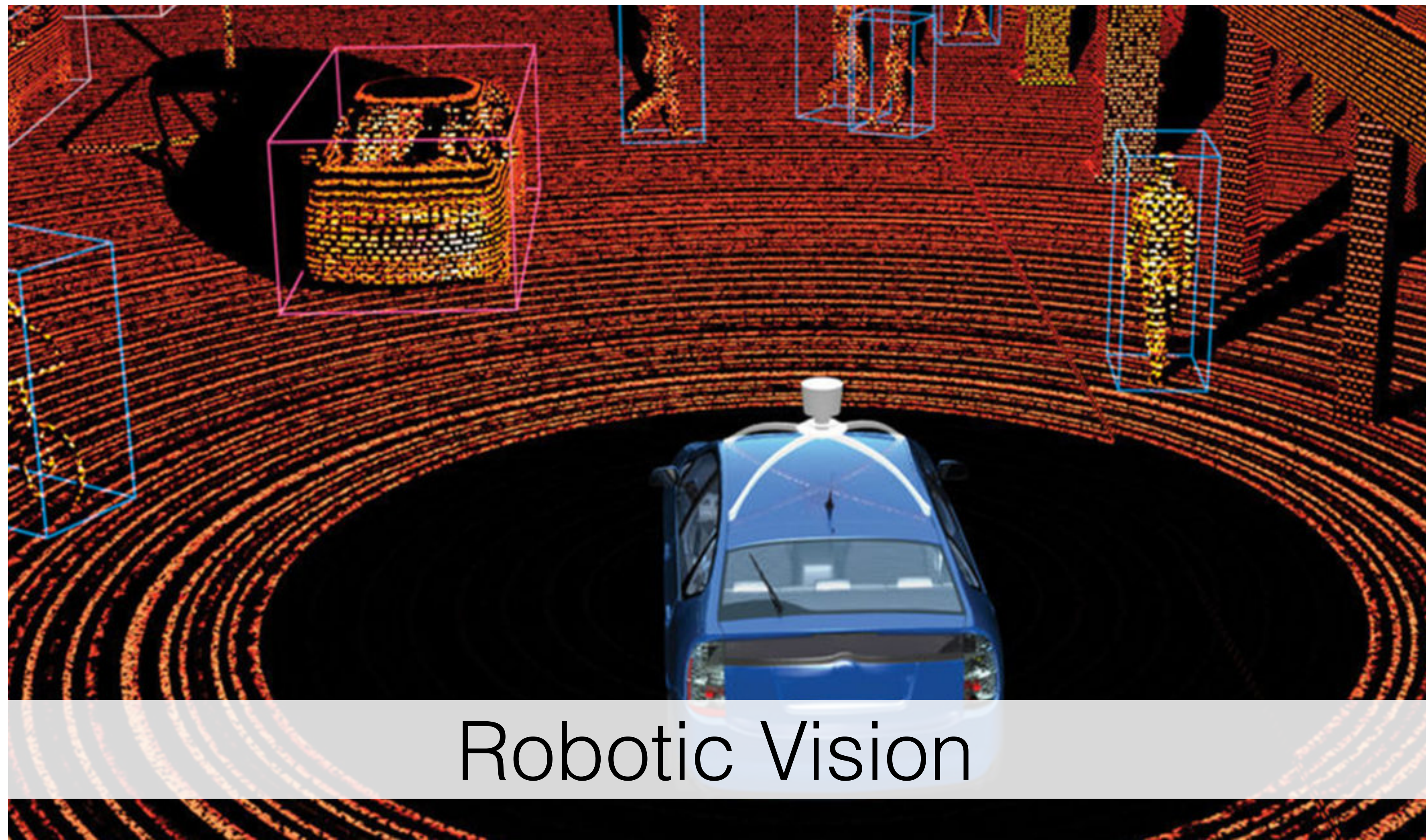

Physics Simulation

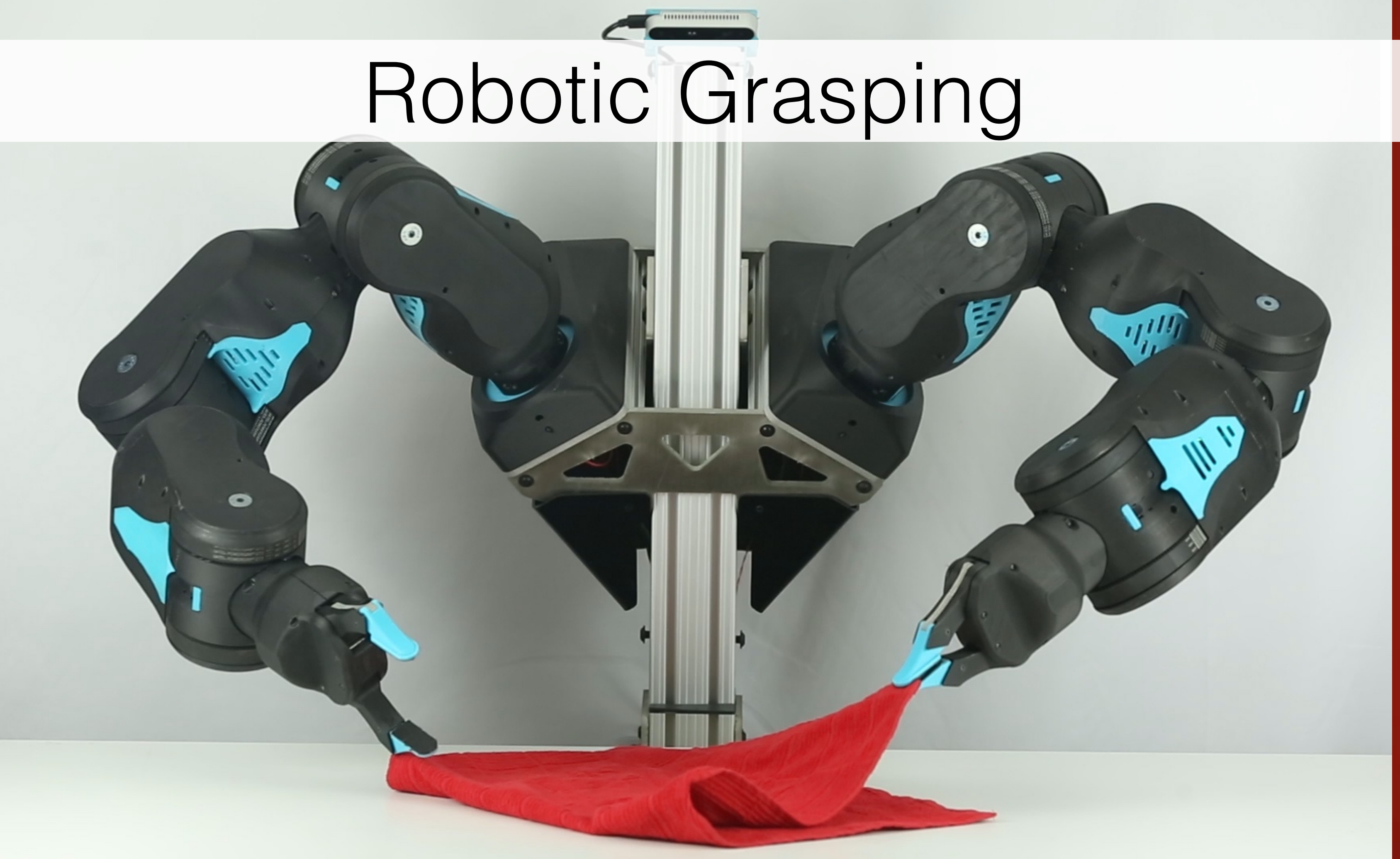
Handcrafted Scene Representation

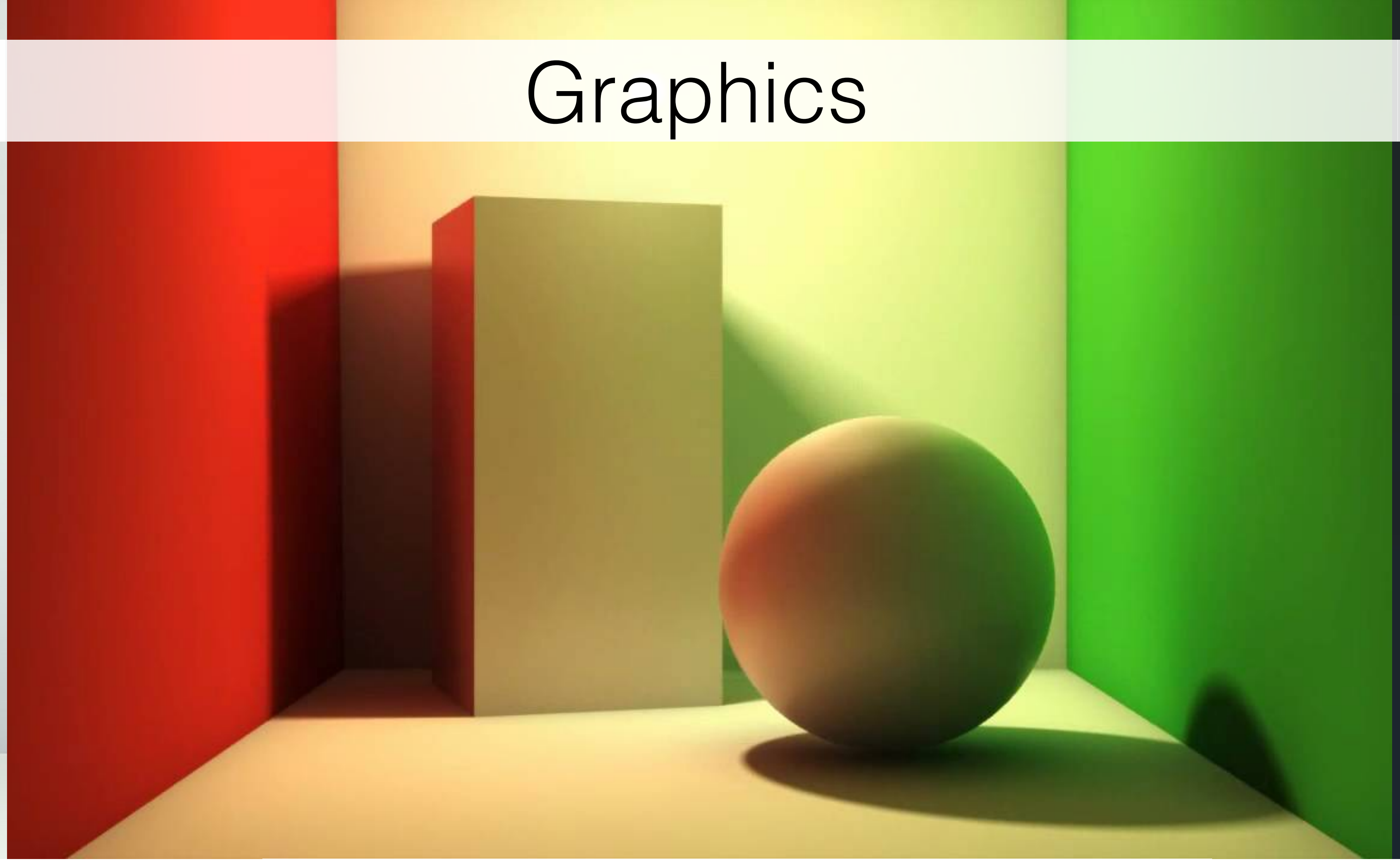
Features are hand-crafted for each application.

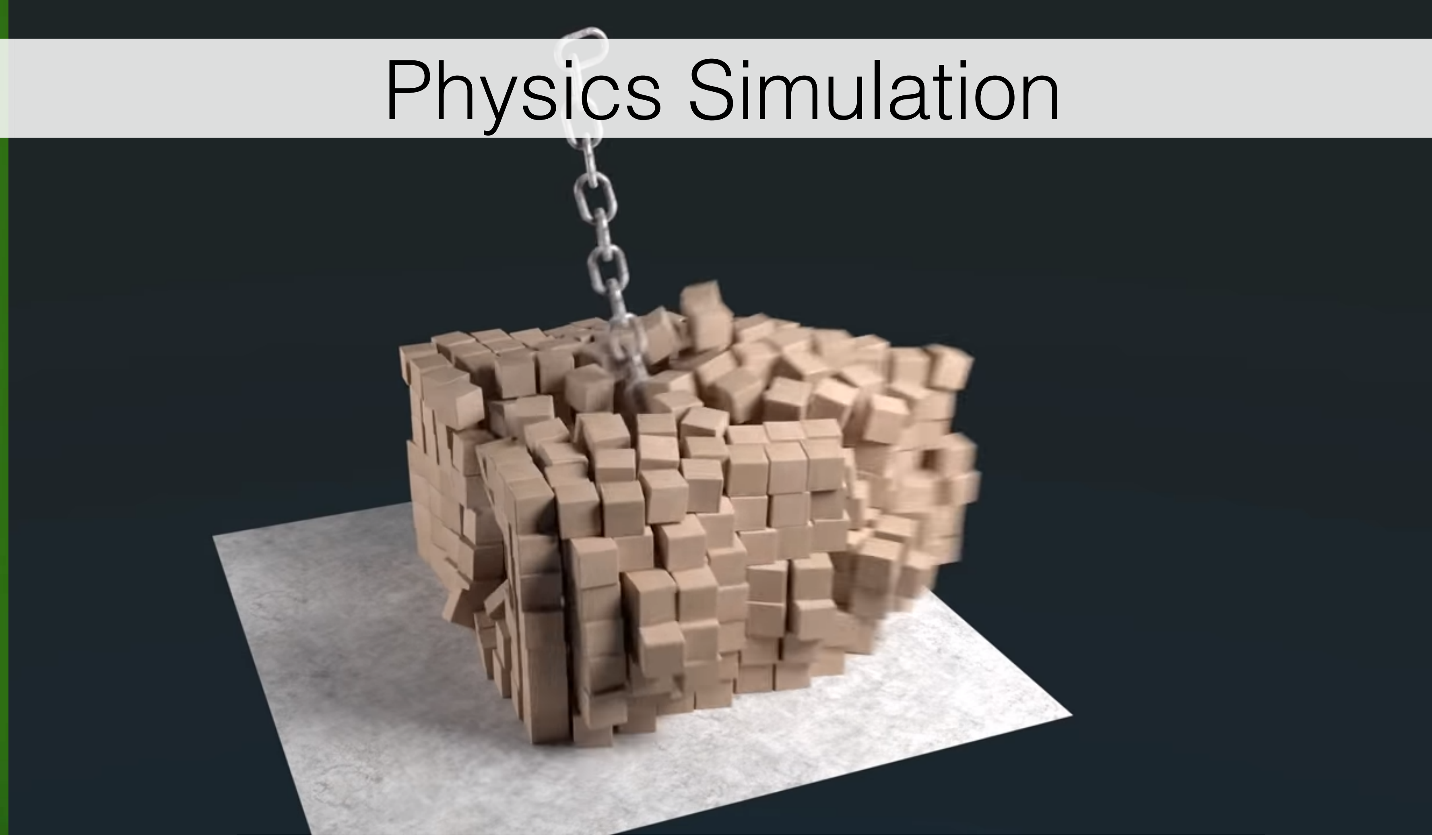
Autonomous Navigation & Planning


Photogrammetry


Robotic Vision







geometry, affordance, …

albedo, specular color, …

mass, friction coefficients, …

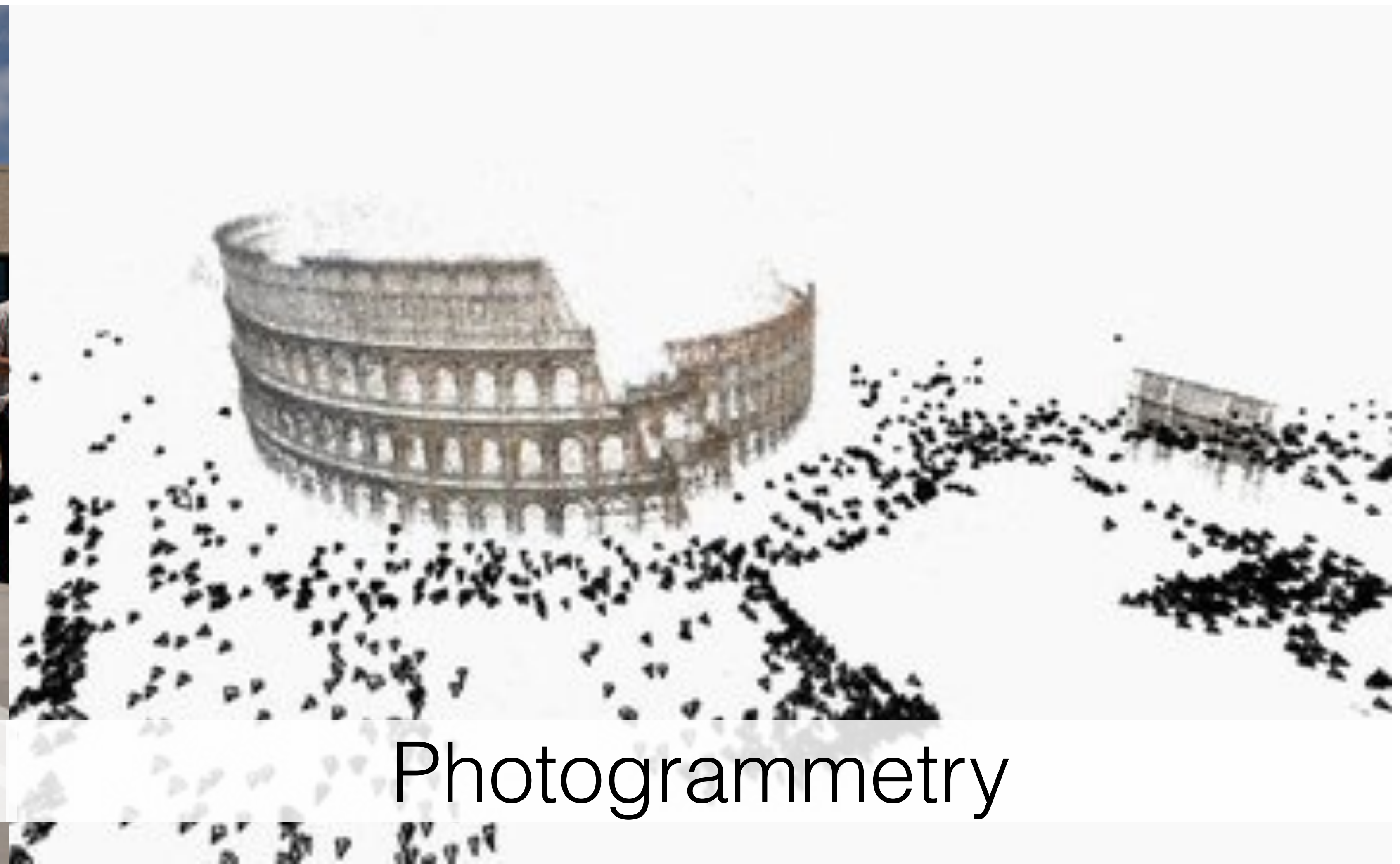
Handcrafted Scene Representation

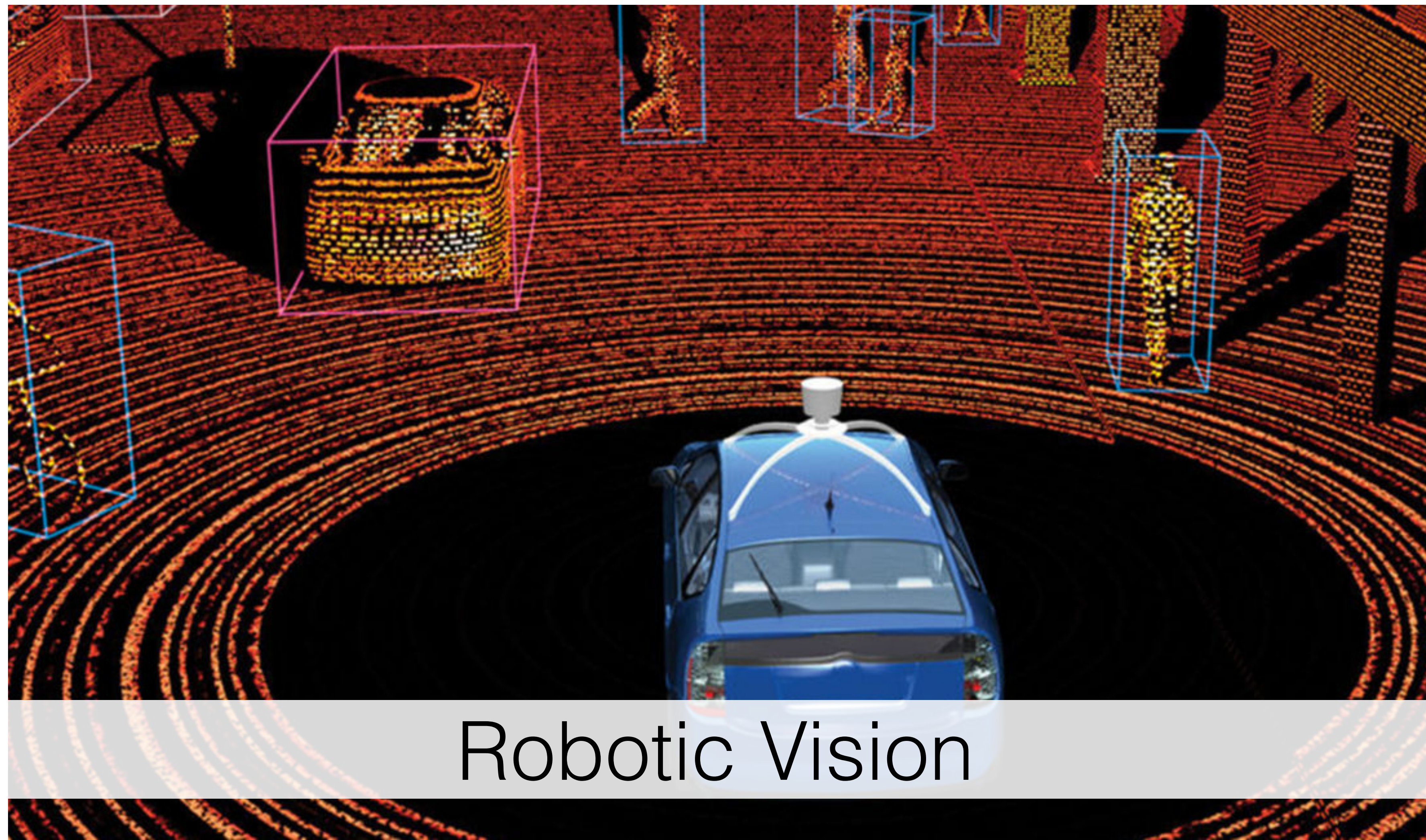
Features are hand-crafted for each application.

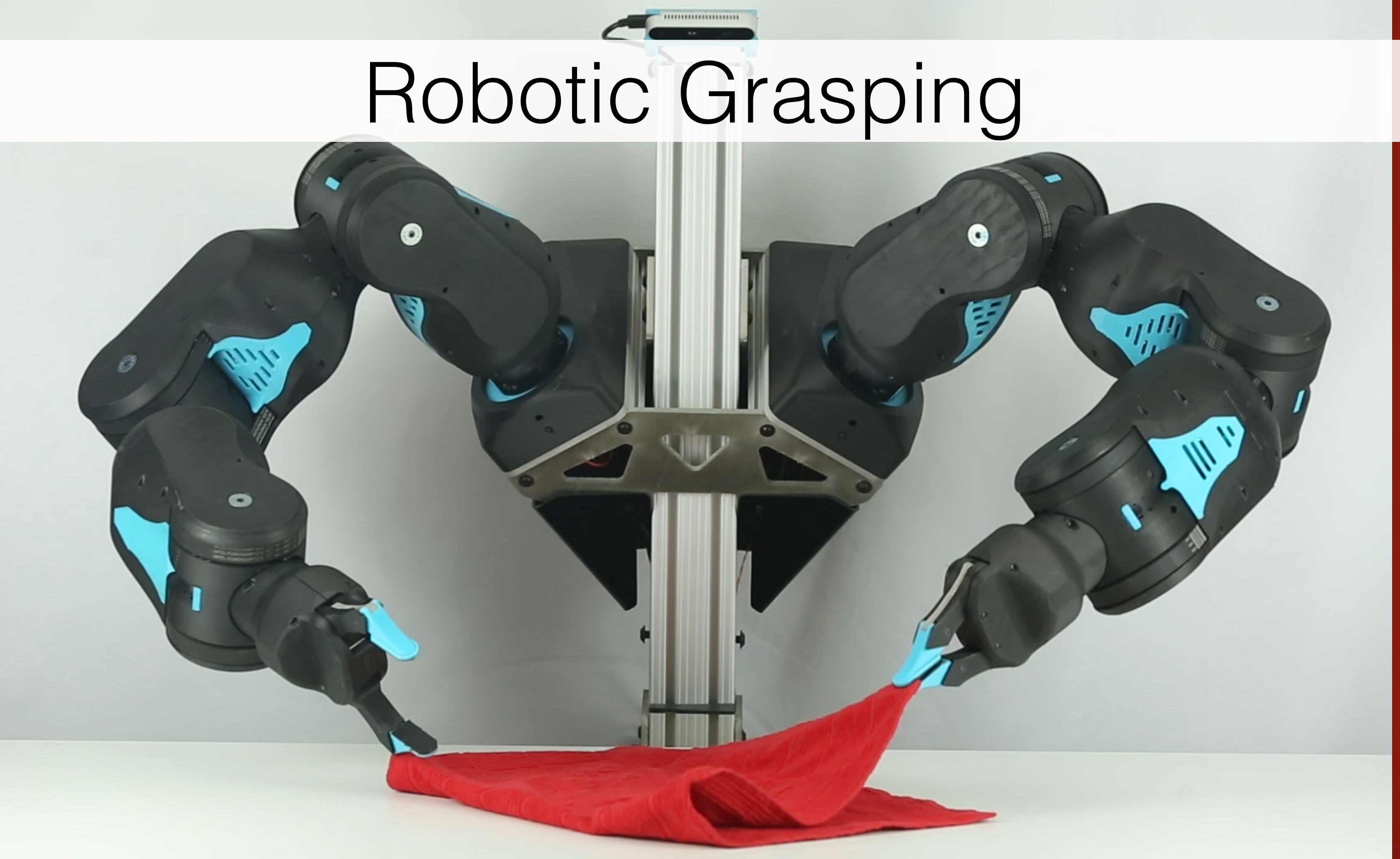
accessibility, obstacle, …

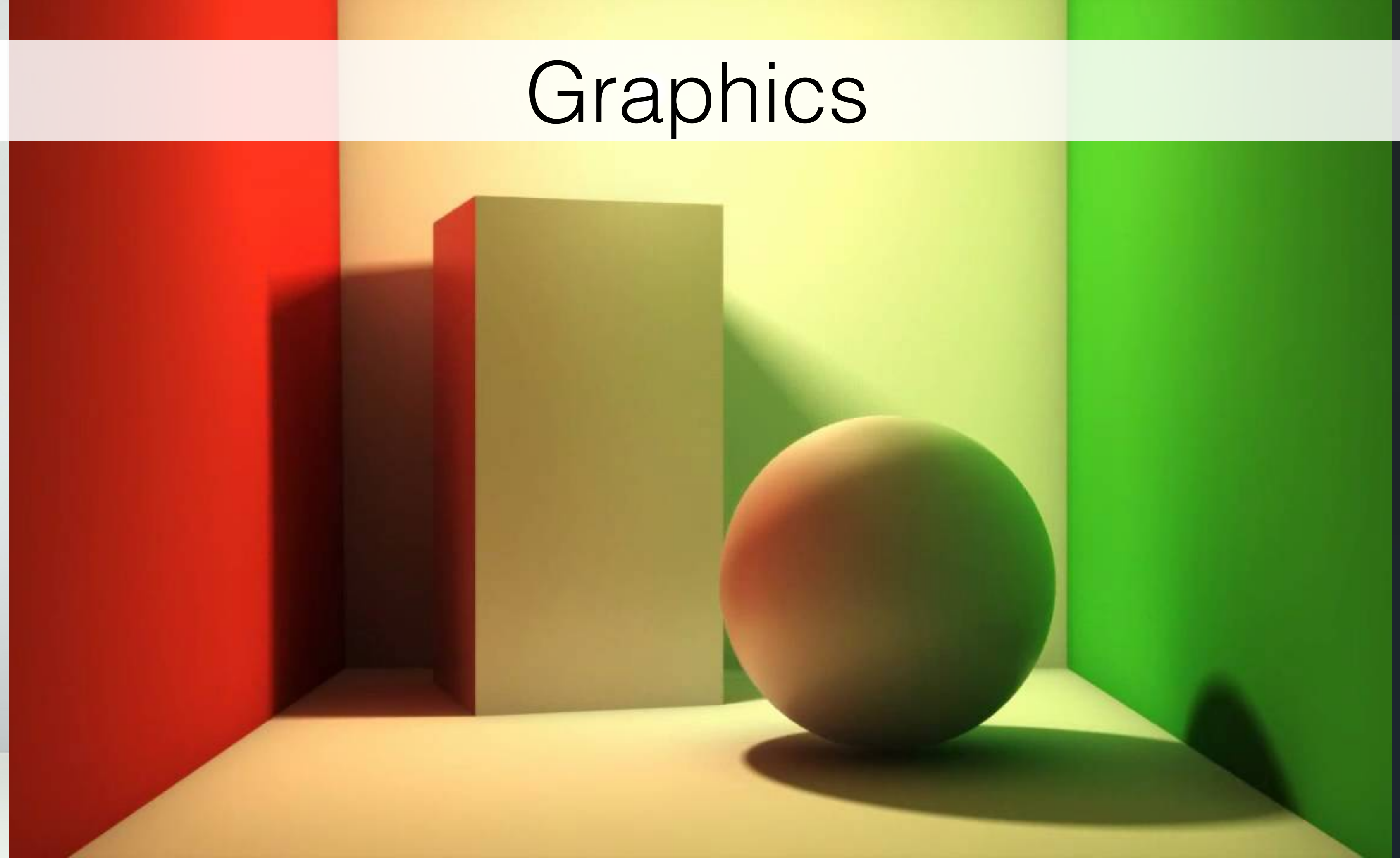
geometry, color, …

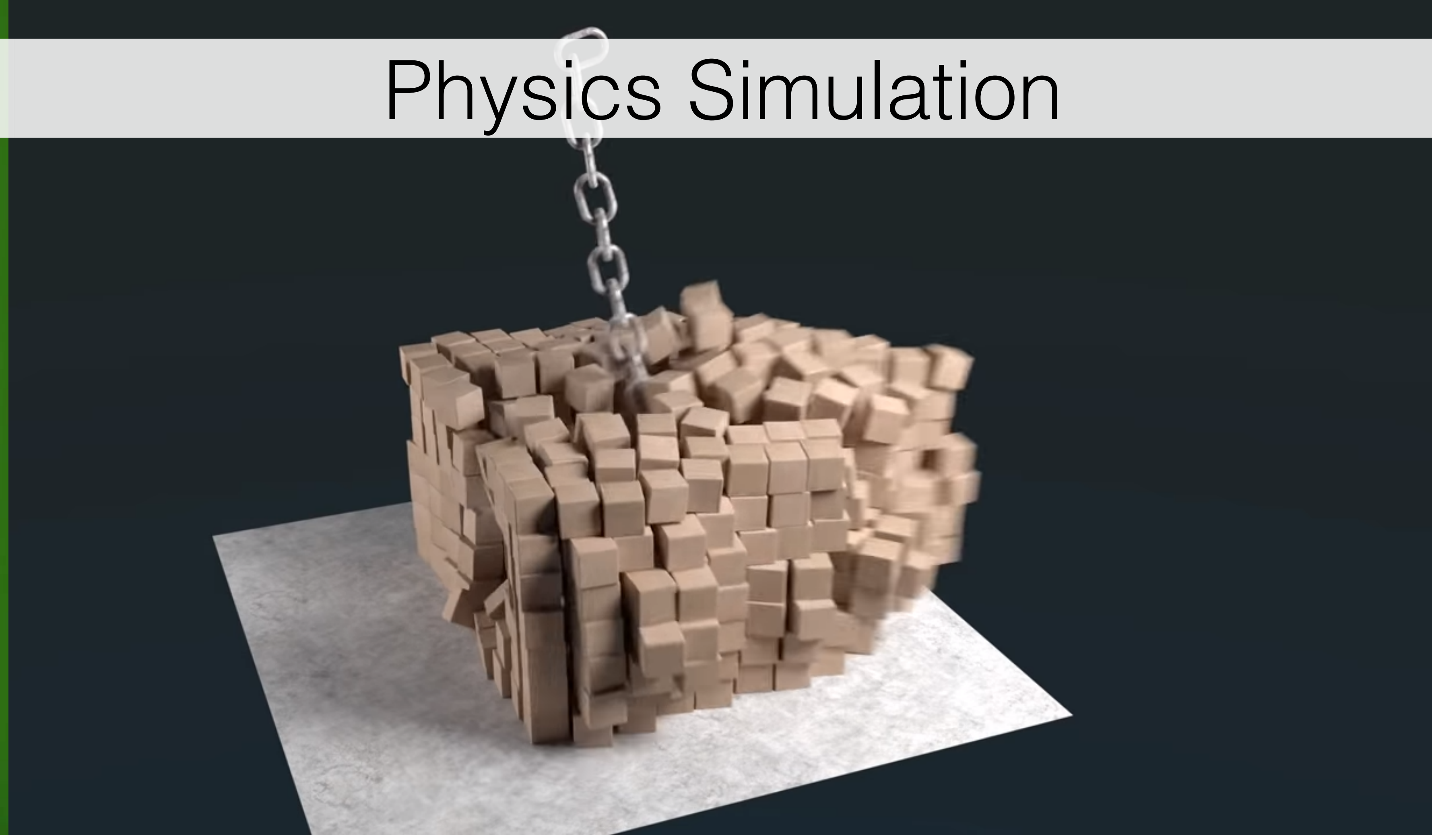
semantic class, geometry, …






Robotic Grasping

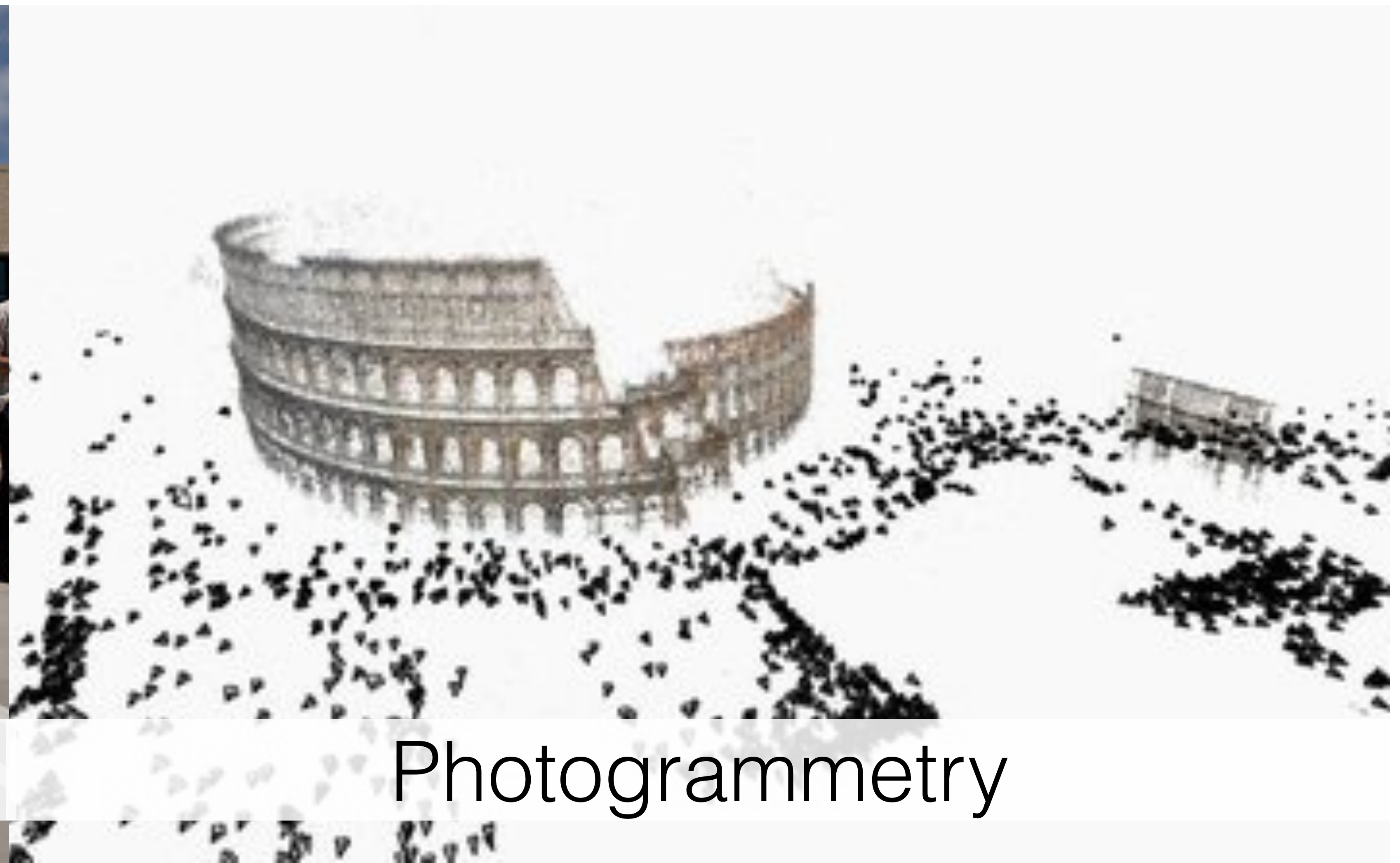

Graphics

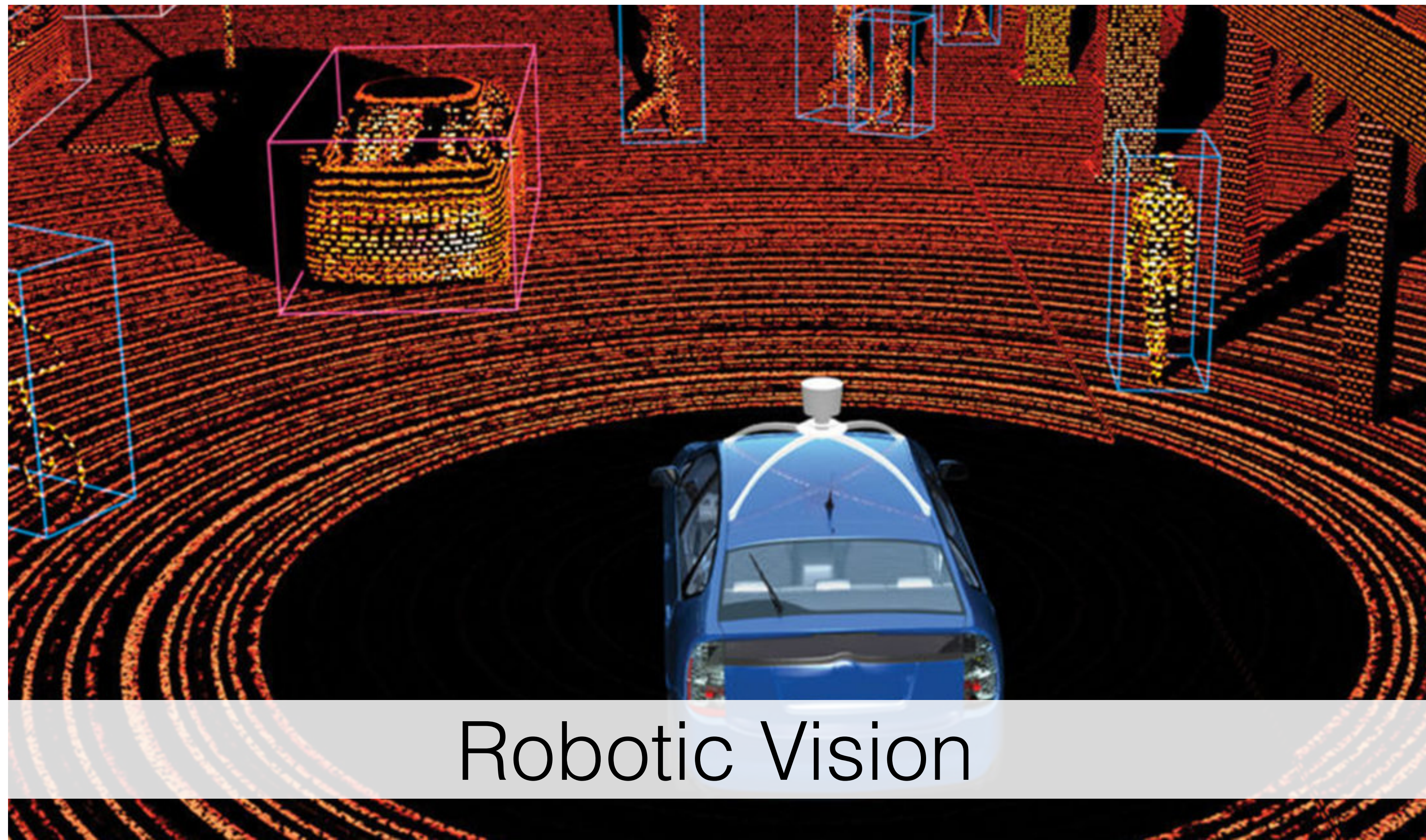

Physics Simulation

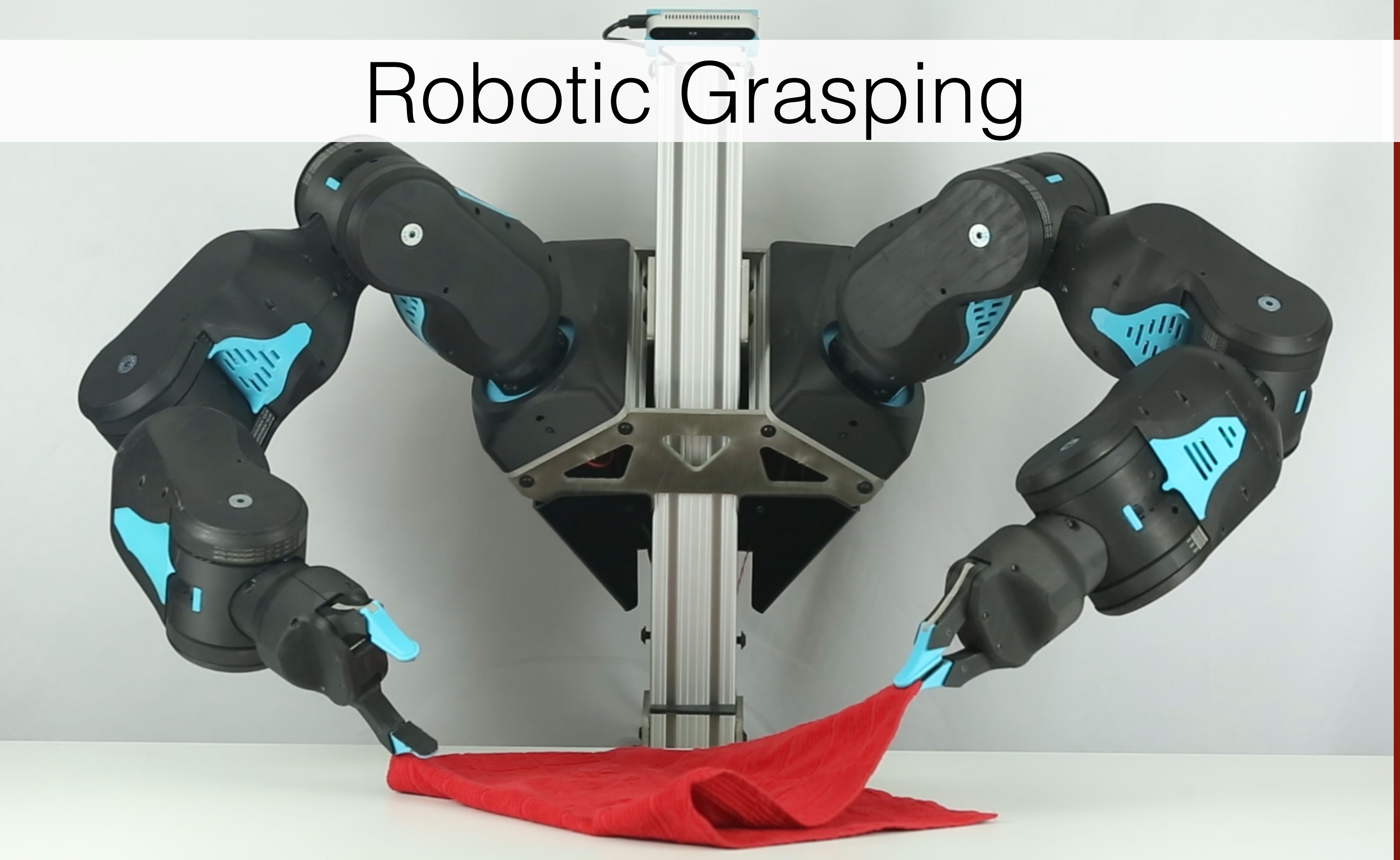
Handcrafted Scene Representation

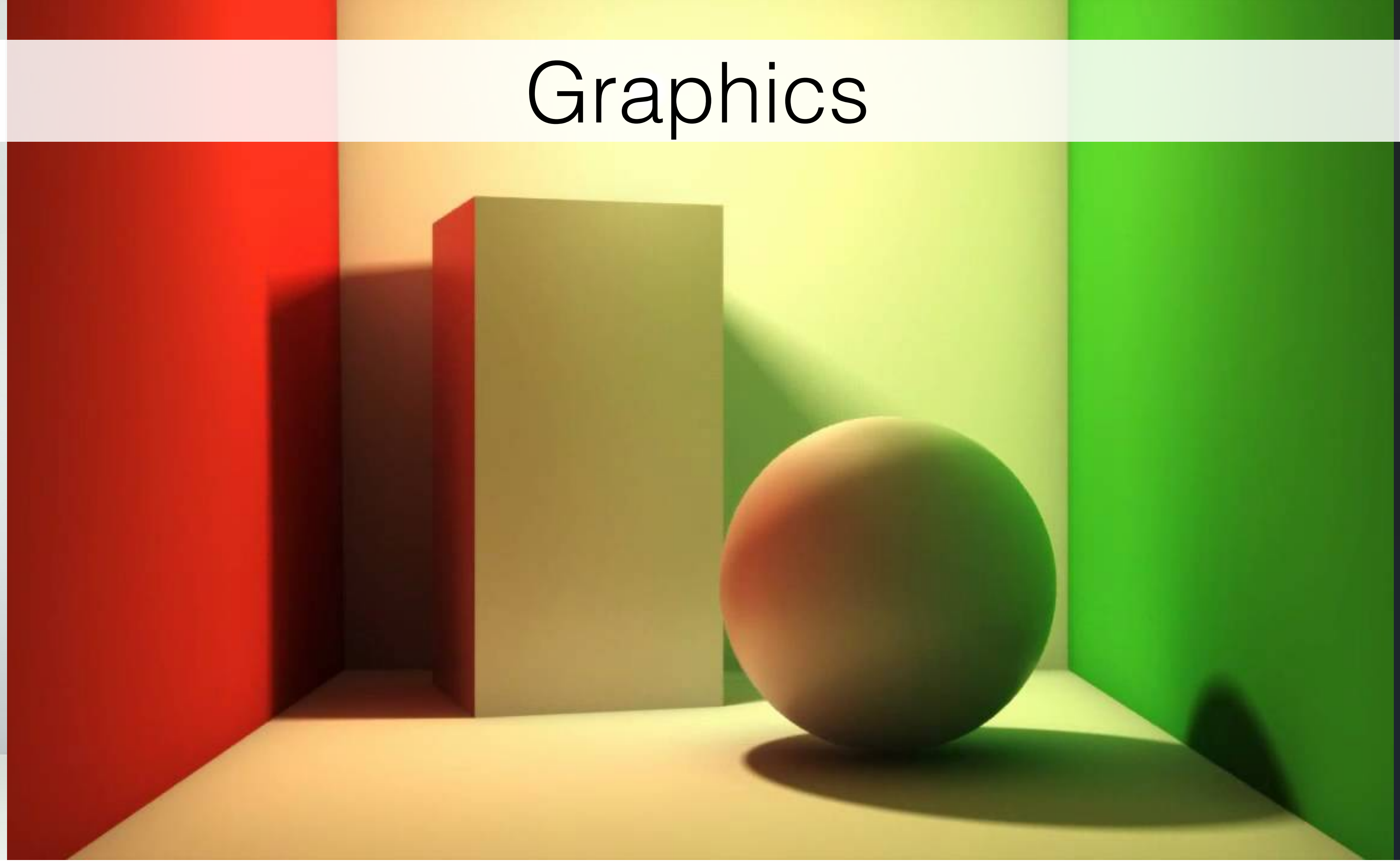
Not differentiable. Hard to infer. Don’t represent all properties of scenes.

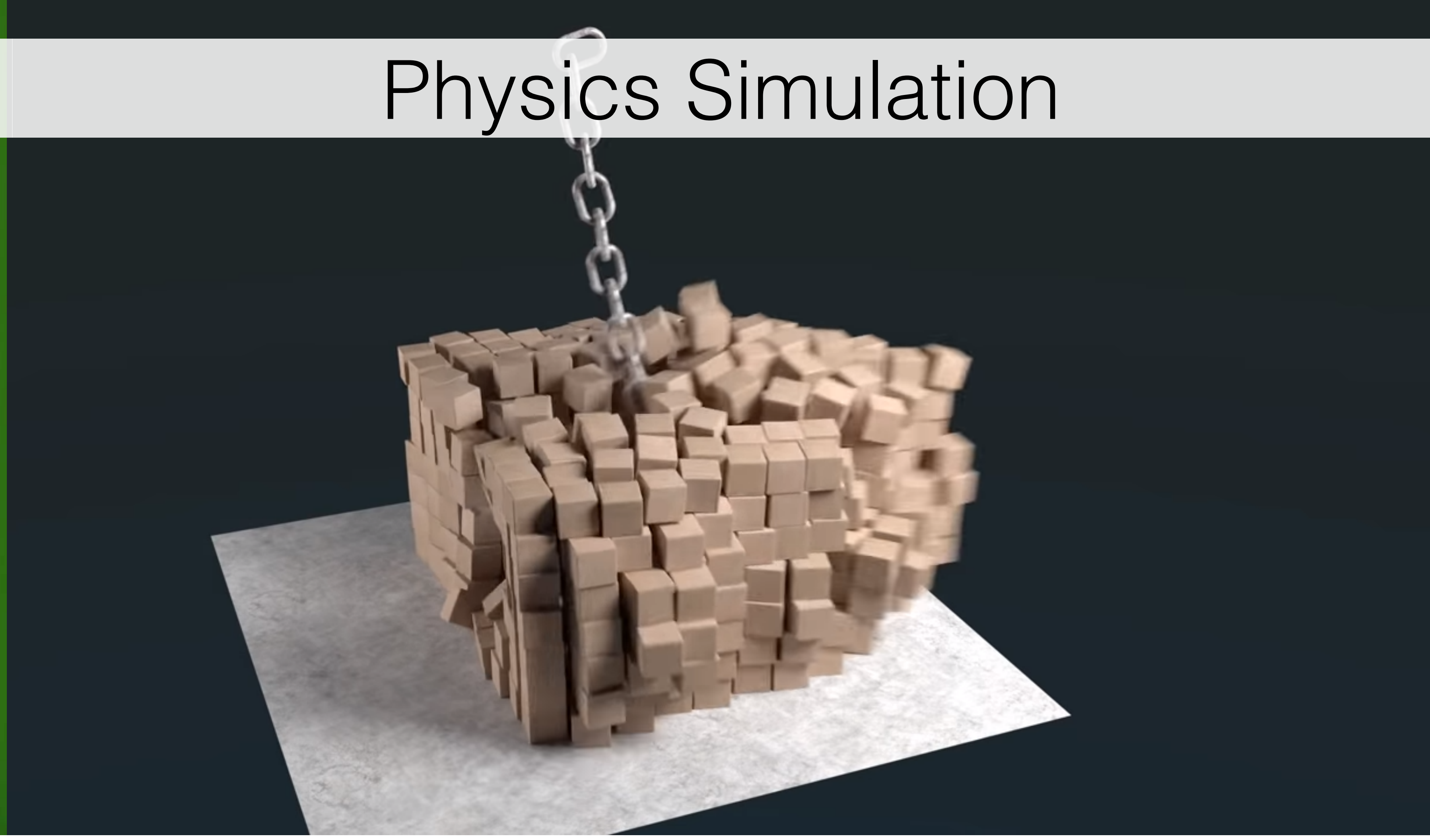
Autonomous Navigation & Planning


Photogrammetry


Robotic Vision

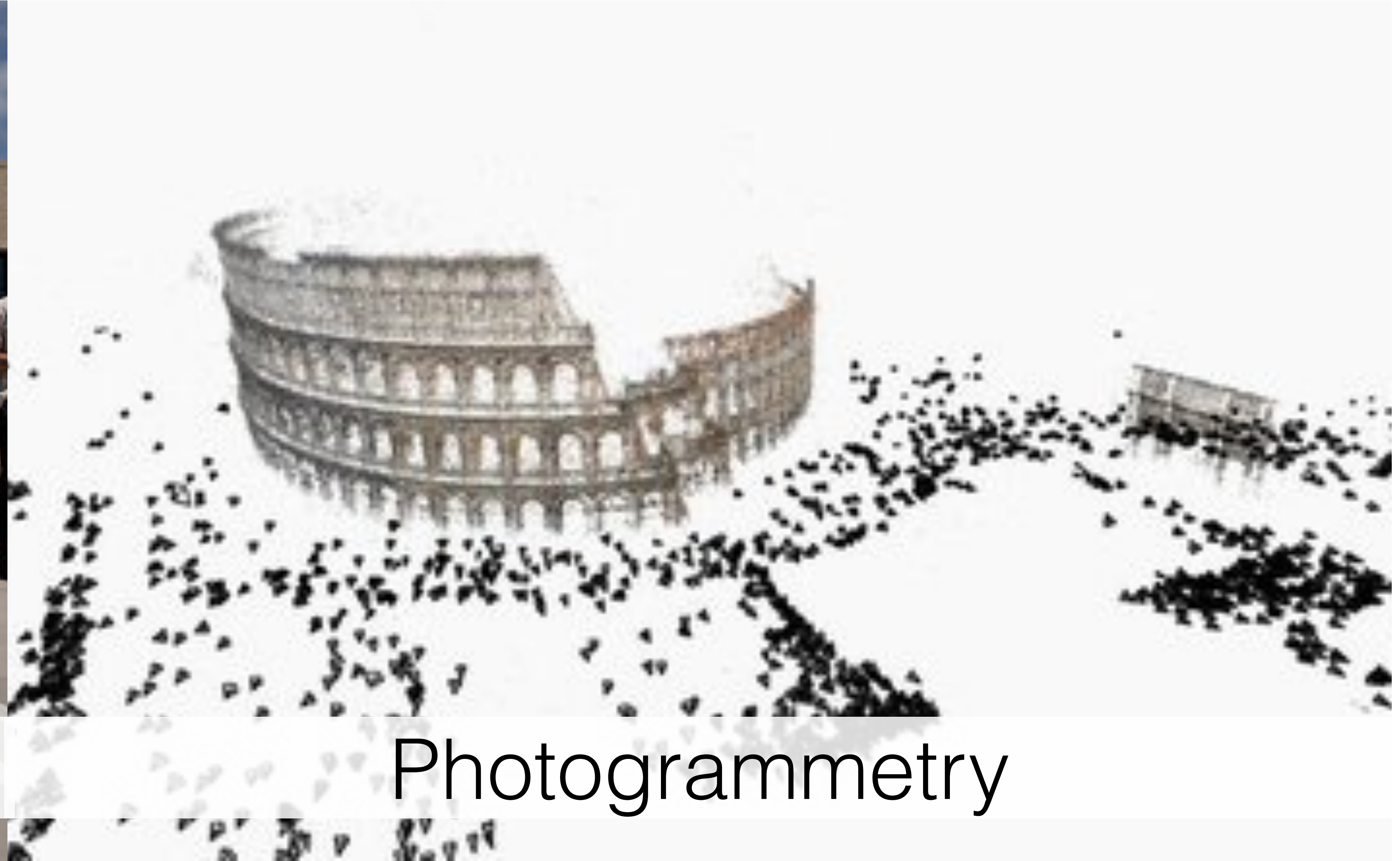


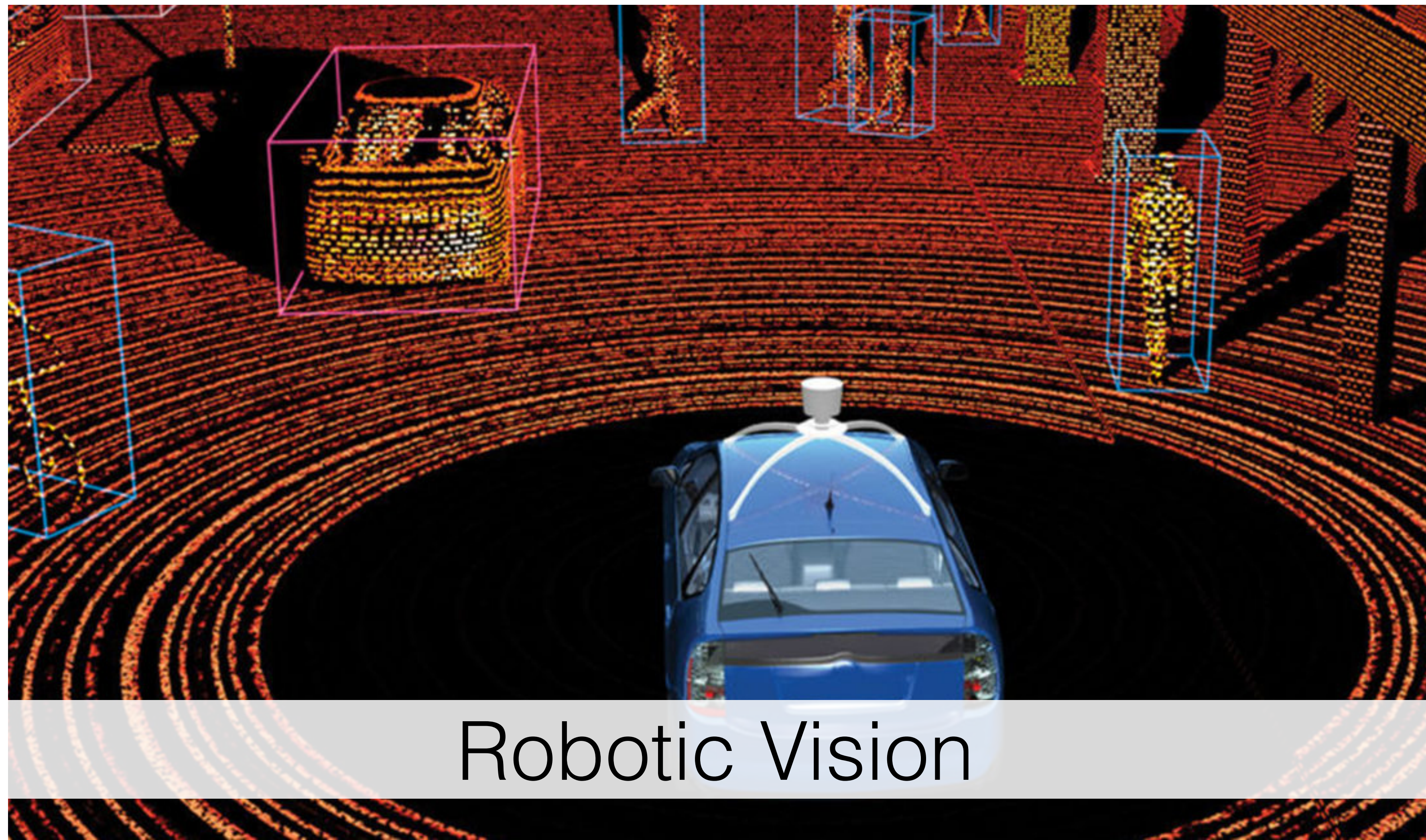




Goal: Neural Scene Representation

Learned feature representation of 3D scenes.

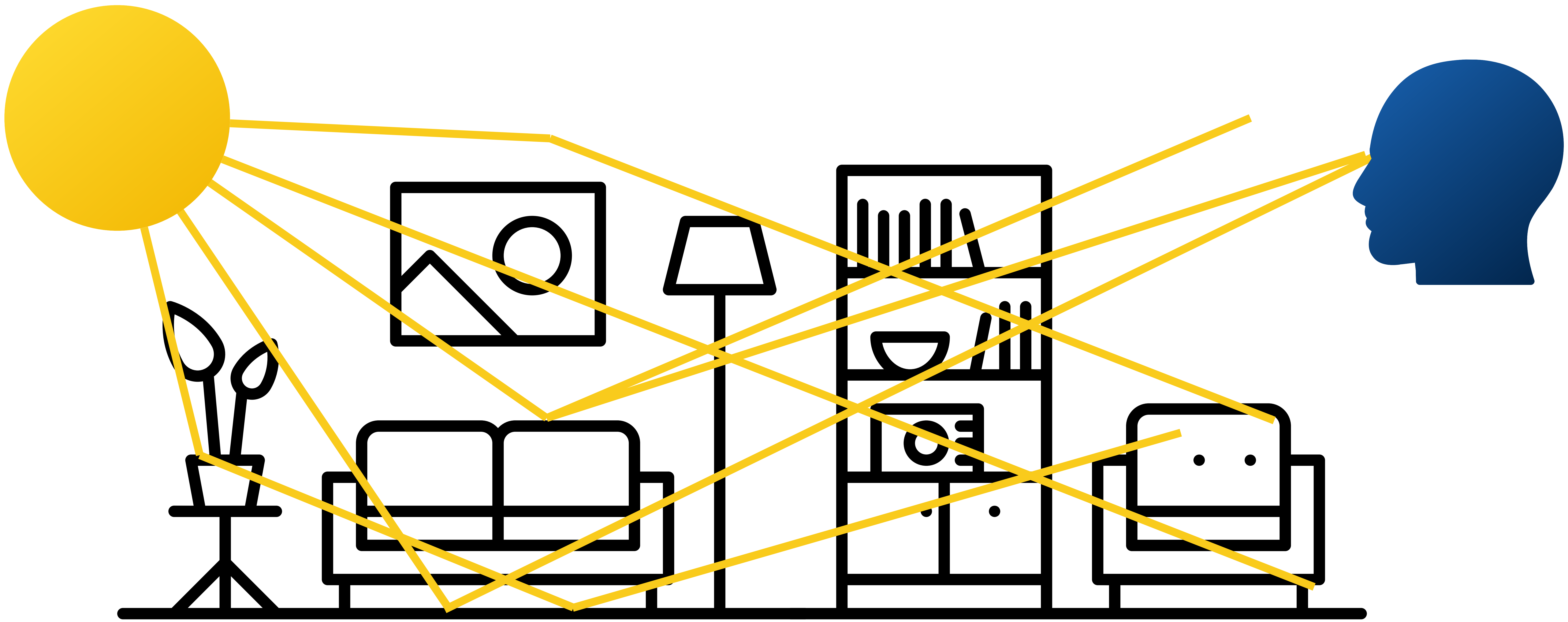




What is a scene?

materials, light sources, 3D shape, color, weight, density, friction coefficients, etc

How do we observe scenes?



An eye (or a camera) observes a subset of all the light rays in a scene.


3D Scene
Image Formation
Graphics


3D Scene
Image Formation
Graphics
Neural Scene Representation


3D Scene
Image Formation
Inference
Neural Scene Representation
Graphics

3D Scene
Image Formation
Inference
Neural Scene Representation
Graphics
“inverse graphics”…?


3D Scene
Image Formation
Inference
Neural Scene Representation
Downstream Tasks
Graphics
“inverse graphics”...?

This course

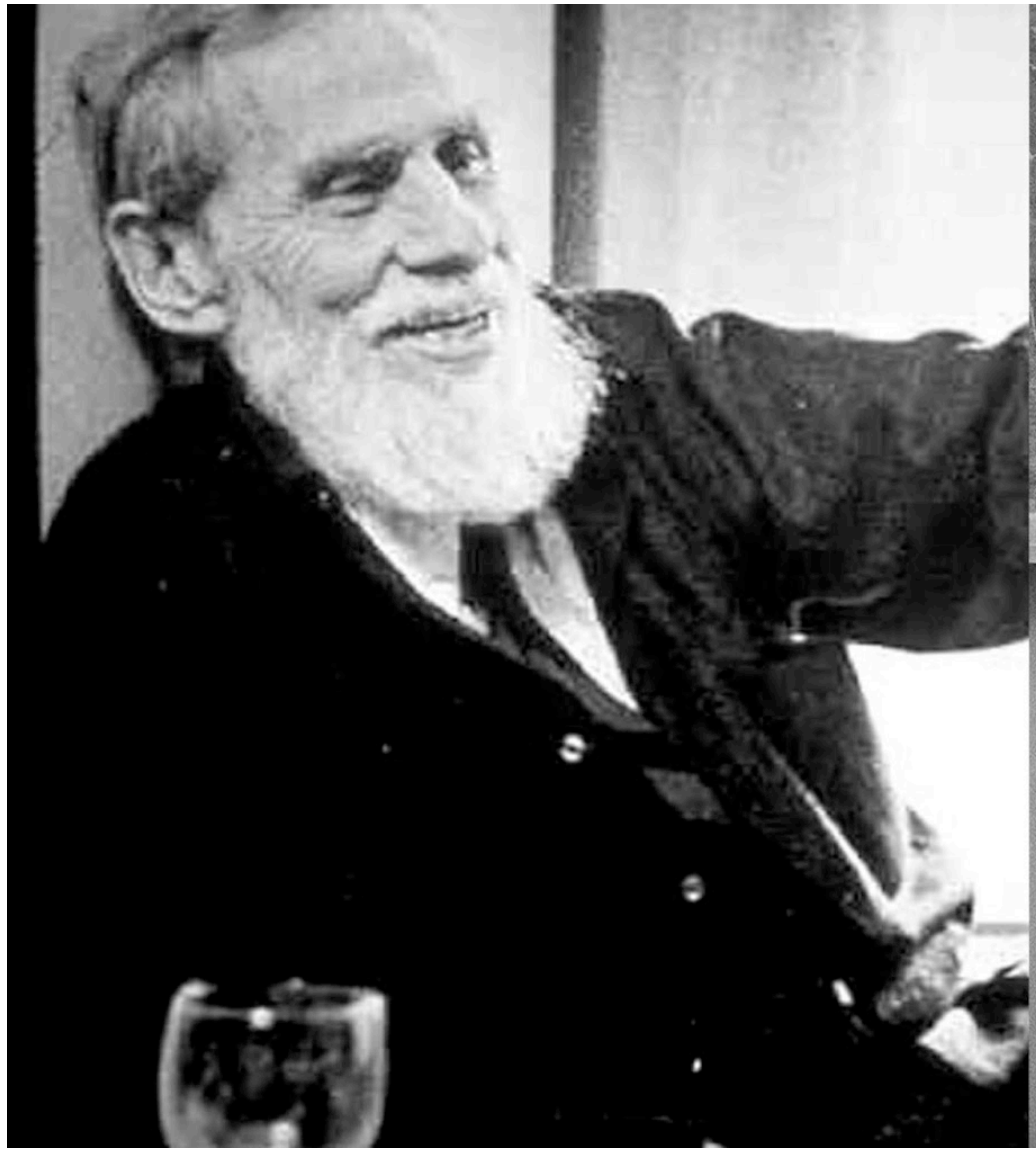
Inference

Neural Scene Representation

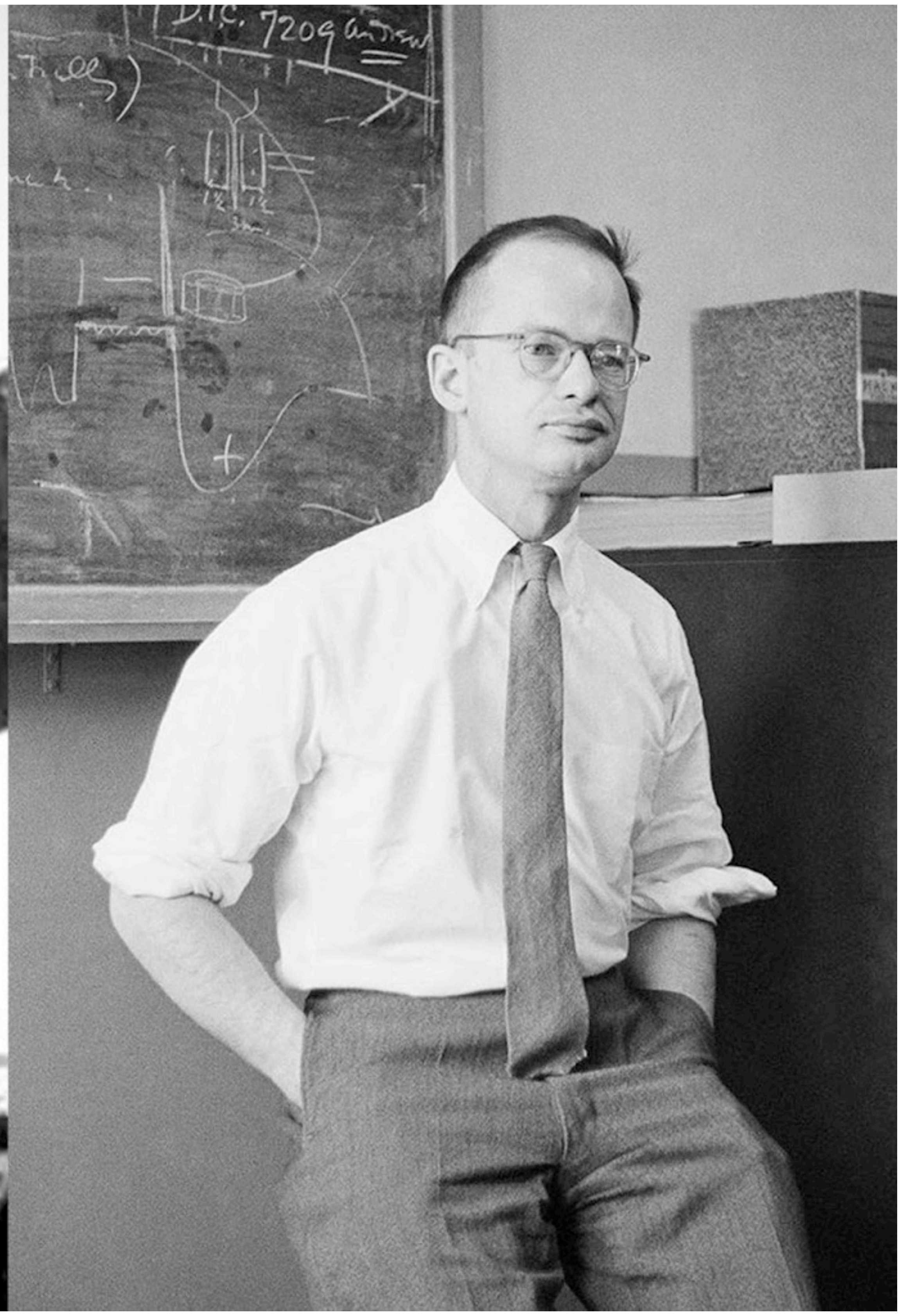
Downstream Tasks

Brief history of AI for Computer Vision

Perceptron (1943)

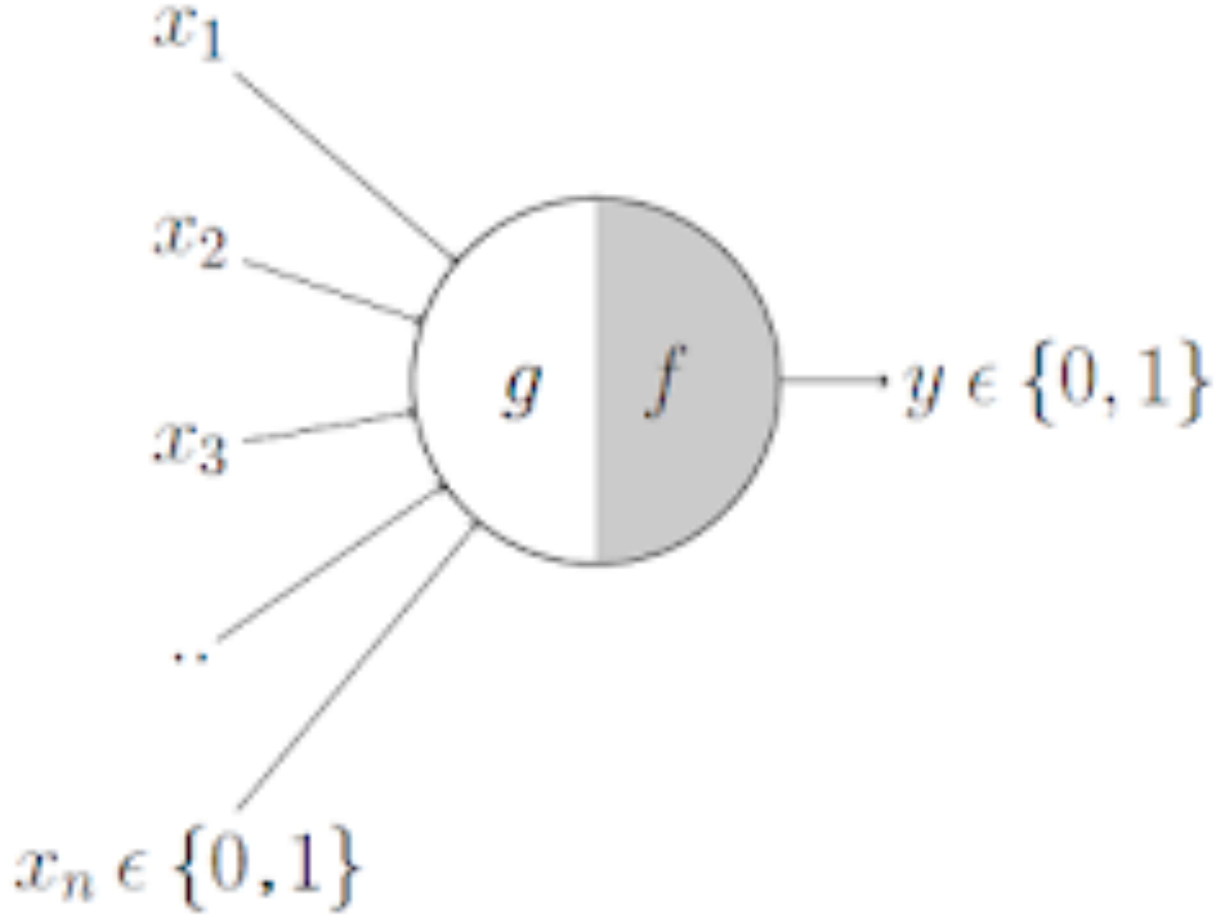


Warren S. McCulloch



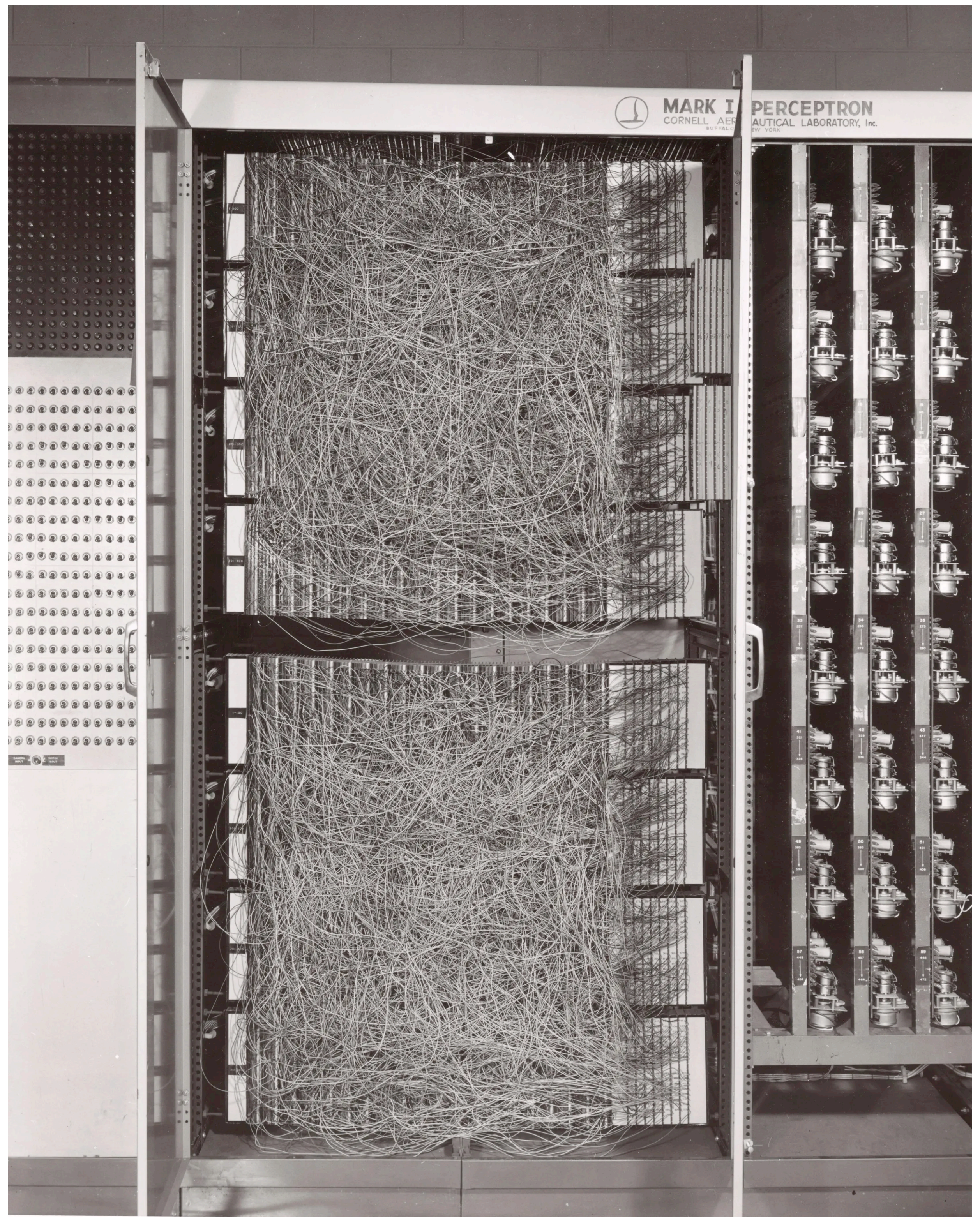
Walter Pitts

First mathematical model of a neuron
(no implementation yet)



Perceptron, Mark I (1958)

Frank Rosenblatt
Often referred as the father of Deep Learning



Perceptron, Mark I
"The first device to think as the human brain"

1960s: The "Logic" approach to AI and Computer Vision: A Summer Project

Marvin Minsky, Seymour Papert, Gerald Sussman
MIT, Project Mac

"connect a camera to a computer and do something with it"
- Minsky to Sussman (1966)

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".

Counter Arguments Against Deep Learning

Marvin Minsky and Seymour Papert
Perceptron (Book), 1969, lays out problems with neural networks

Often mis-cited: Neural Networks cannot solve the XOR problem!?!

Only single-layer Perceptrons can't…

The Foundations of Modern Deep Learning

BEYOND REGRESSION:
NEW TOOLS FOR PREDICTION AND ANALYSIS
IN THE BEHAVIORAL SCIENCES

A thesis presented
by
Paul John Werbos
to
The Committee on Applied Mathematics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
statistics

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].

There have been many attempts to design self-organizing neural networks. The aim is to find a powerful synaptic modification rule that will allow an arbitrarily connected neural network to develop an internal structure that is appropriate for a particular task domain. The task is specified by giving the desired state vector of the output units for each state vector of the input units. If the input units are directly connected to the output units it is relatively easy to find learning rules that iteratively adjust the relative strengths of the connections so as to progressively reduce the difference between the actual and desired output vectors[2]. Learning becomes more interesting but

Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network

LONG SHORT-TERM MEMORY

Neural Computation 9(8):1735–1780, 1997

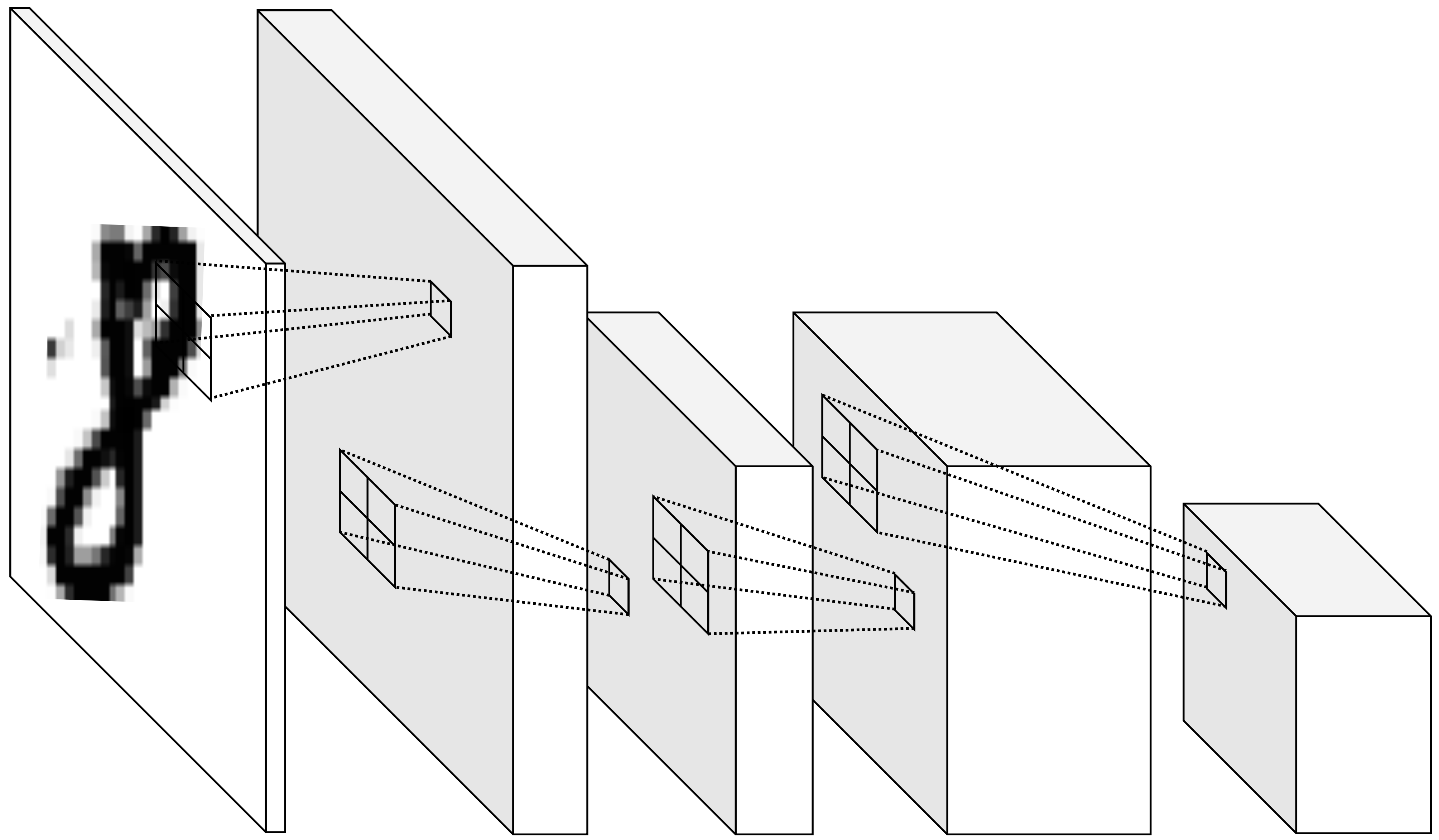
Sepp Hochreiter
Fakultät für Informatik
Technische Universität München
80290 München, Germany
hochreit@informatik.tu-muenchen.de
http://www7.informatik.tu-muenchen.de/~hochreit

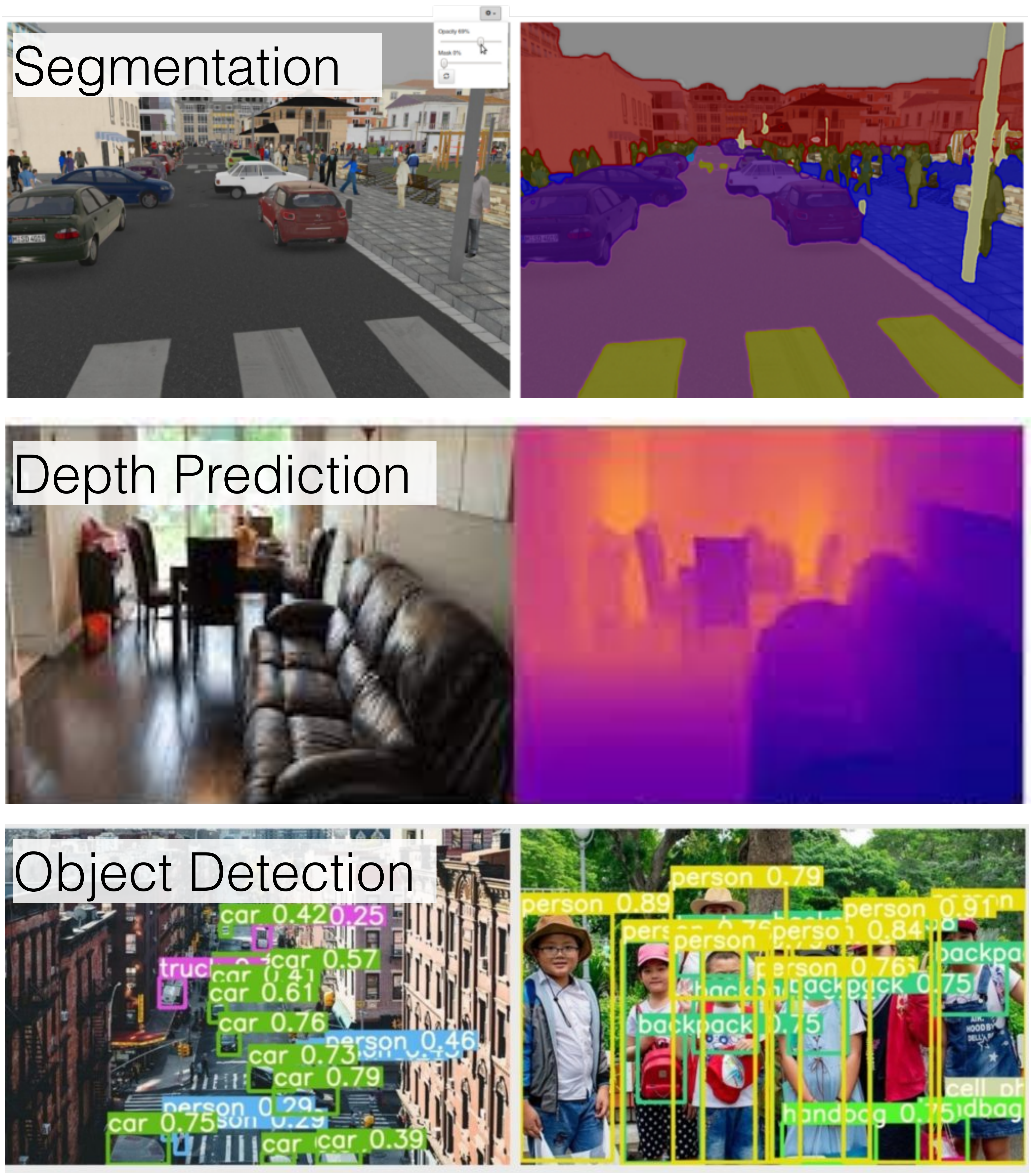
Jürgen Schmidhuber
IDSIA
Corso Elvezia 36
6900 Lugano, Switzerland
juergen@idsia.ch
http://www.idsia.ch/~juergen

Paul John Werbos proposed to train Perceptrons via Backprop in his thesis.

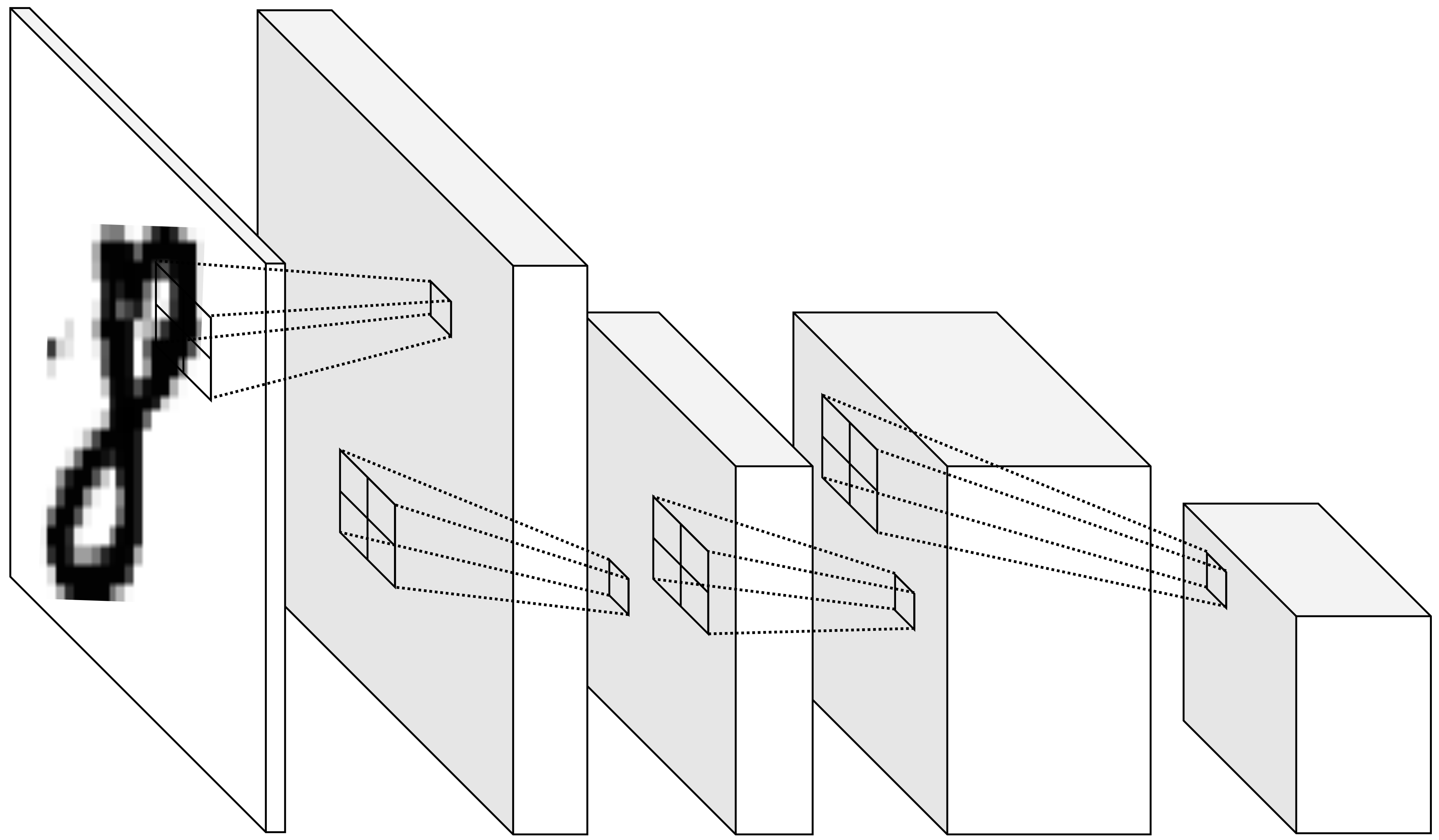
The broader community became aware of the promise of this approach via progress made in publications by Hinton, LeCun, Hochreiter, Schmidhuber, Bengio and others in the late 1990s and 2000s.

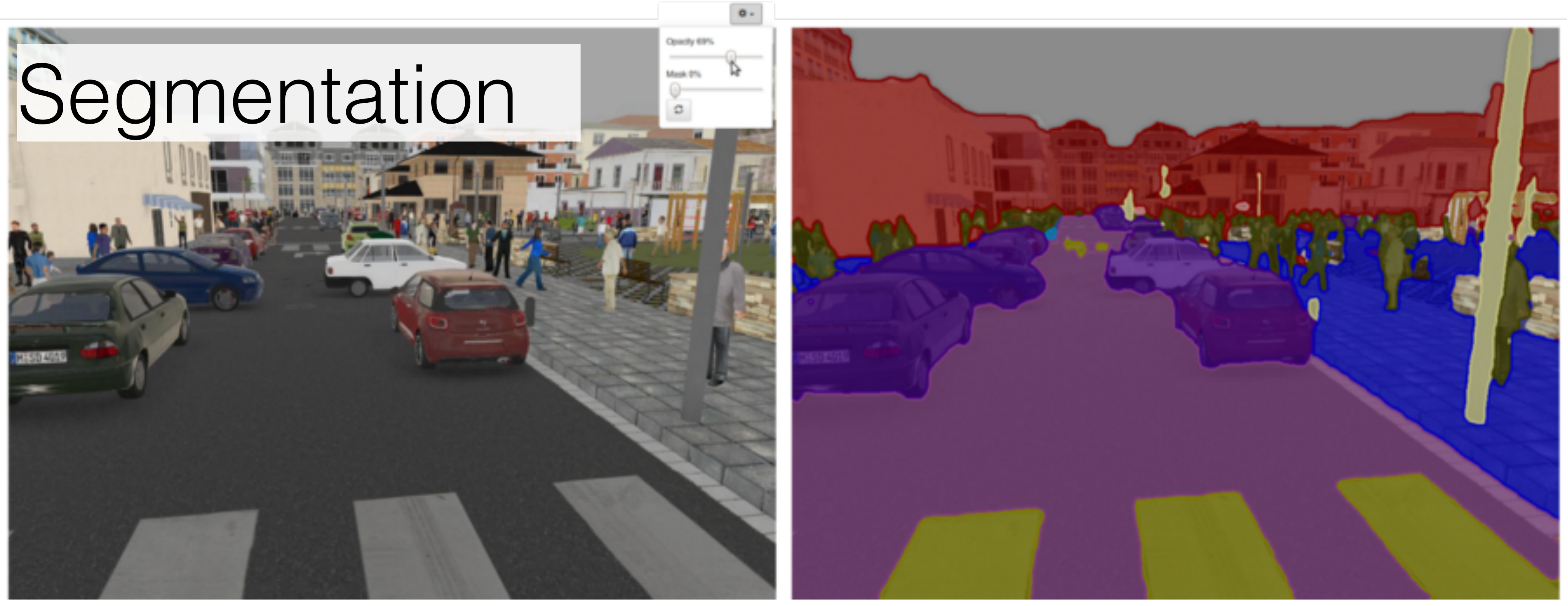
2012: Breakthrough in Supervised Learning





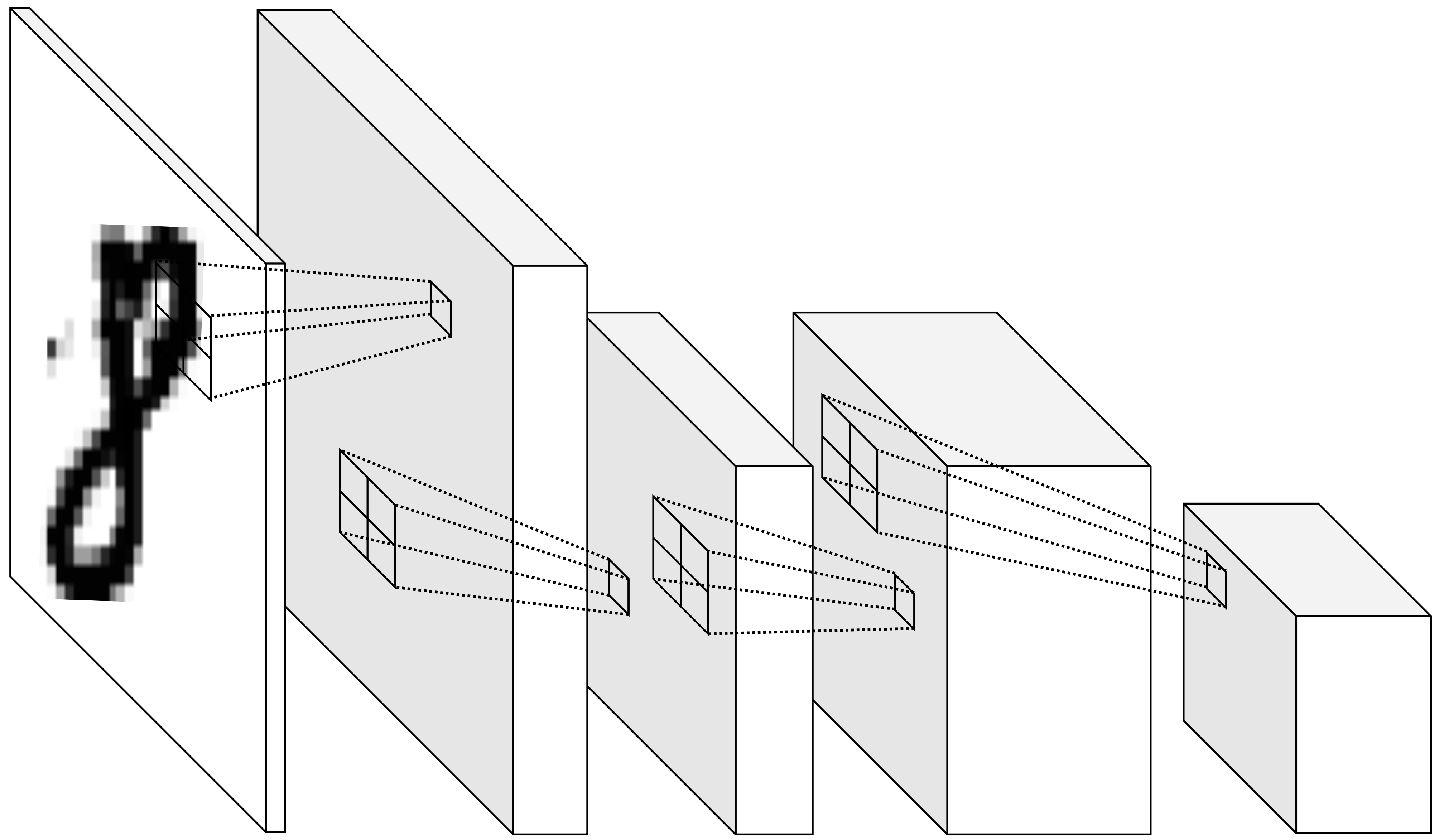
2012: Breakthrough in Supervised Learning

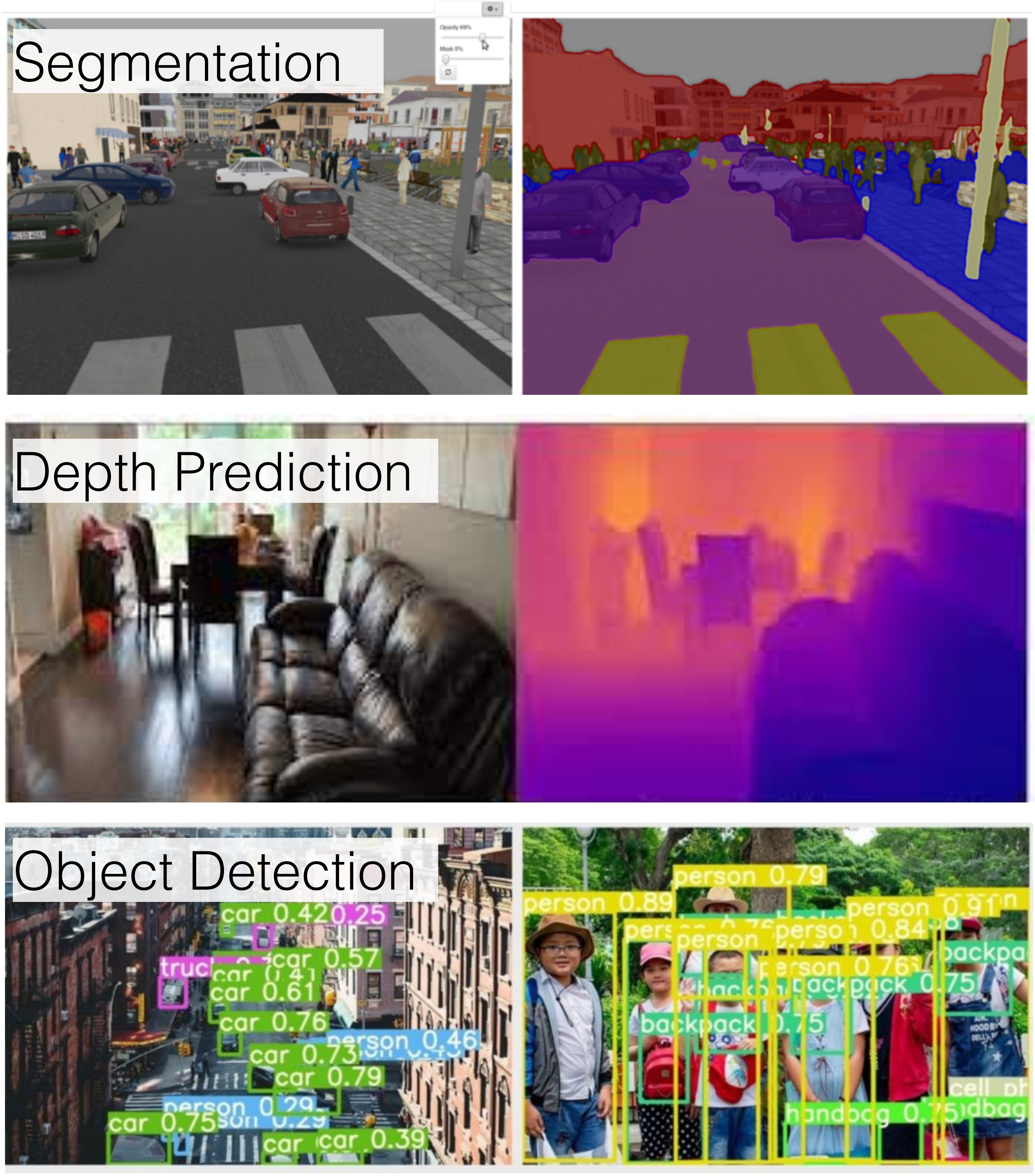




Labeling is expensive :/

2012: Breakthrough in Supervised Learning

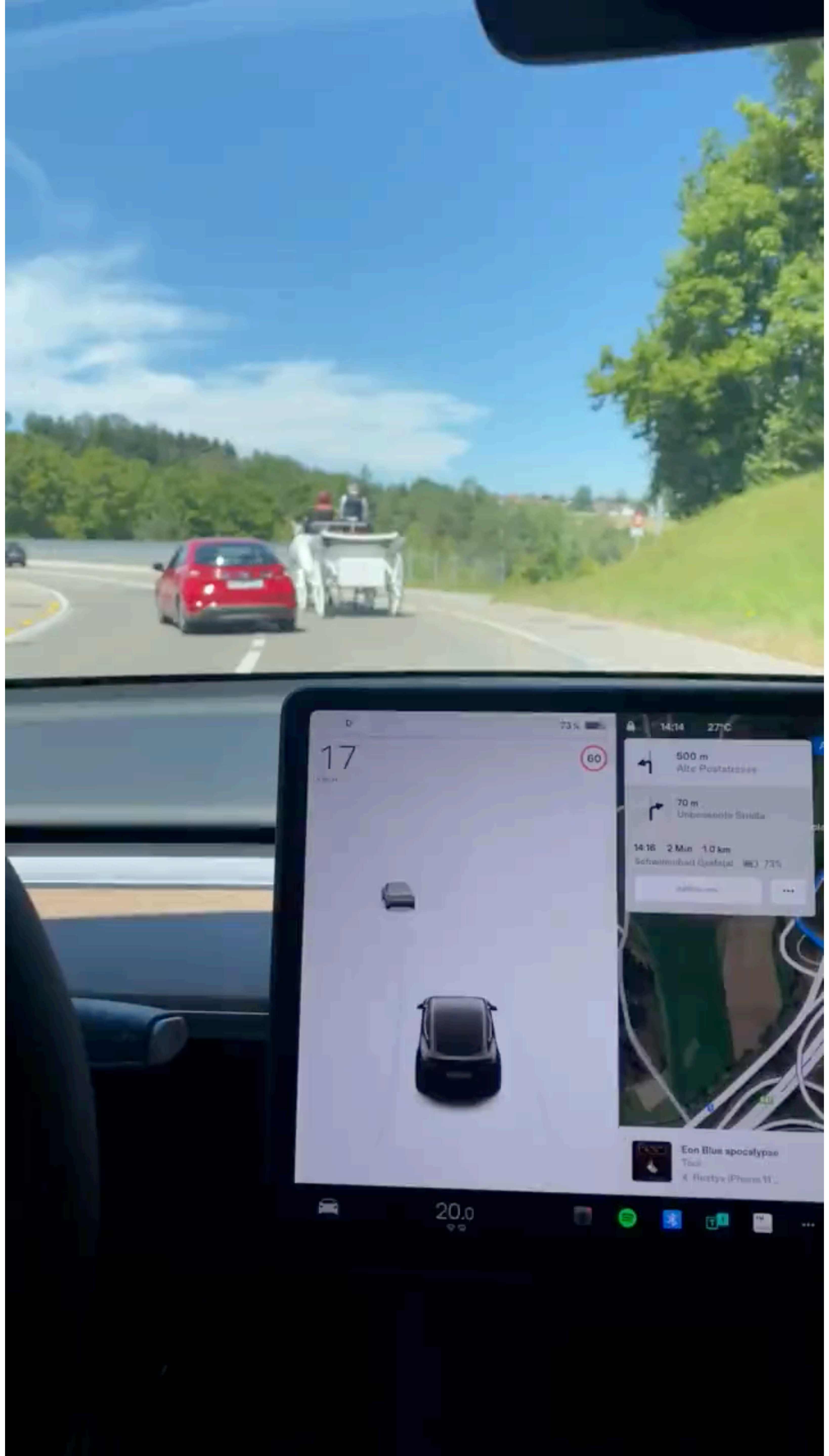




Labeling is expensive :/

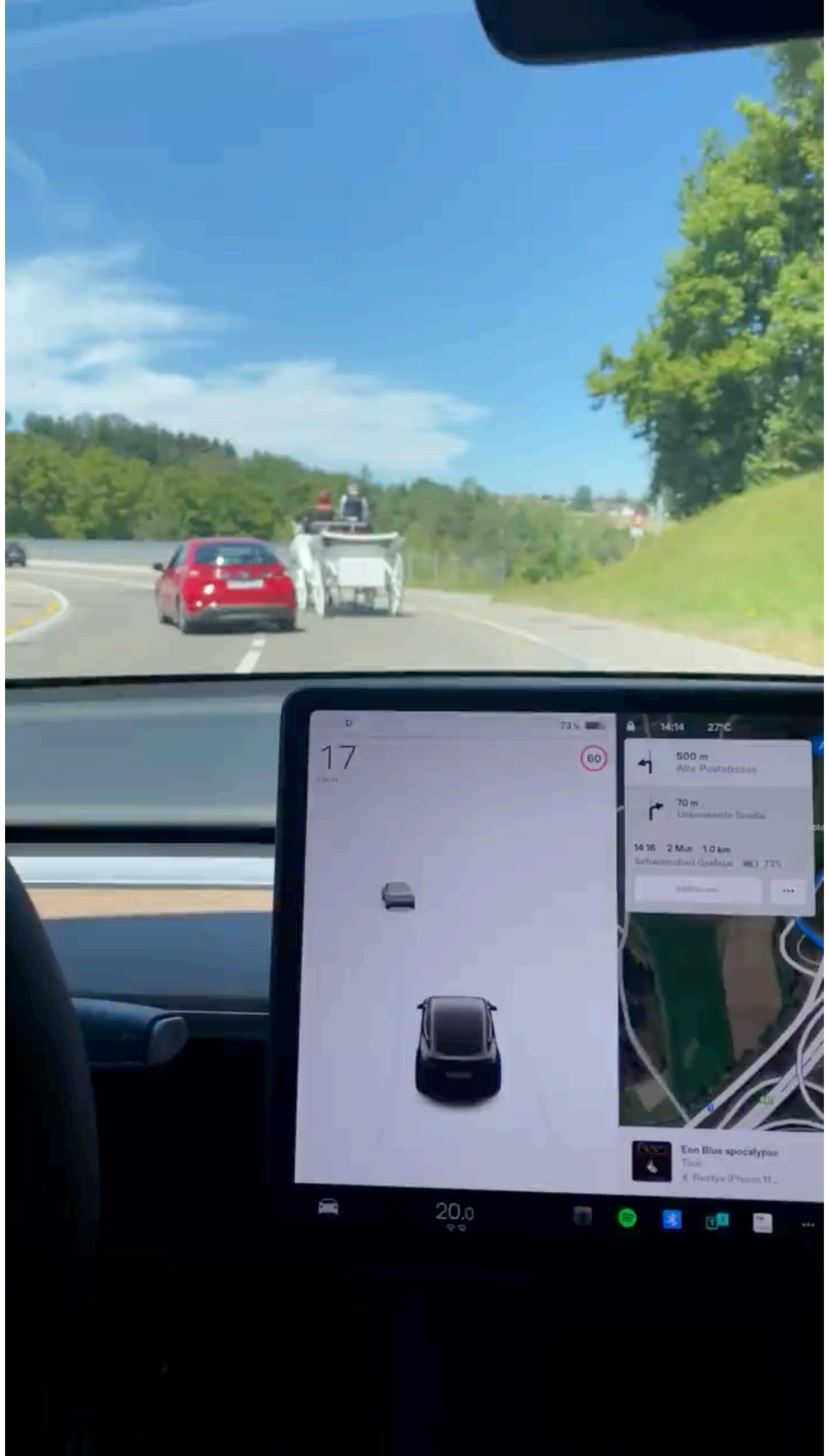
Supervised is “Closed-Set”: Can only learn what we can label!

Closed-Set Predictions



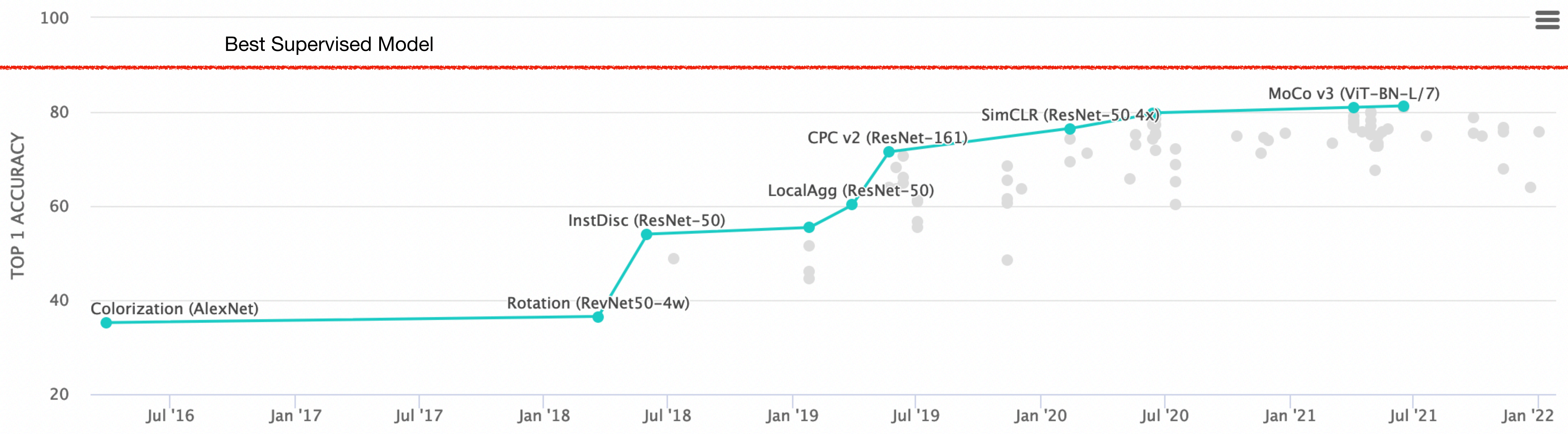
https://www.youtube.com/watch?v=tpOg87AQvbo&t=1s

Closed-Set Predictions



https://www.youtube.com/watch?v=tpOg87AQvbo&t=1s

Self-supervised Learning getting closer to supervised learning!



Recent progress in Self-supervised Learning

Emerging Properties in Self-Supervised Vision Transformers

Mathilde Caron[1,2] Hugo Touvron[1,3] Ishan Misra[1] Hervé Jegou[1]
Julien Mairal[2] Piotr Bojanowski[1] Armand Joulin[1]

[1] Facebook AI Research [2] Inria* [3] Sorbonne University

[cs.CV] 24 May 2021

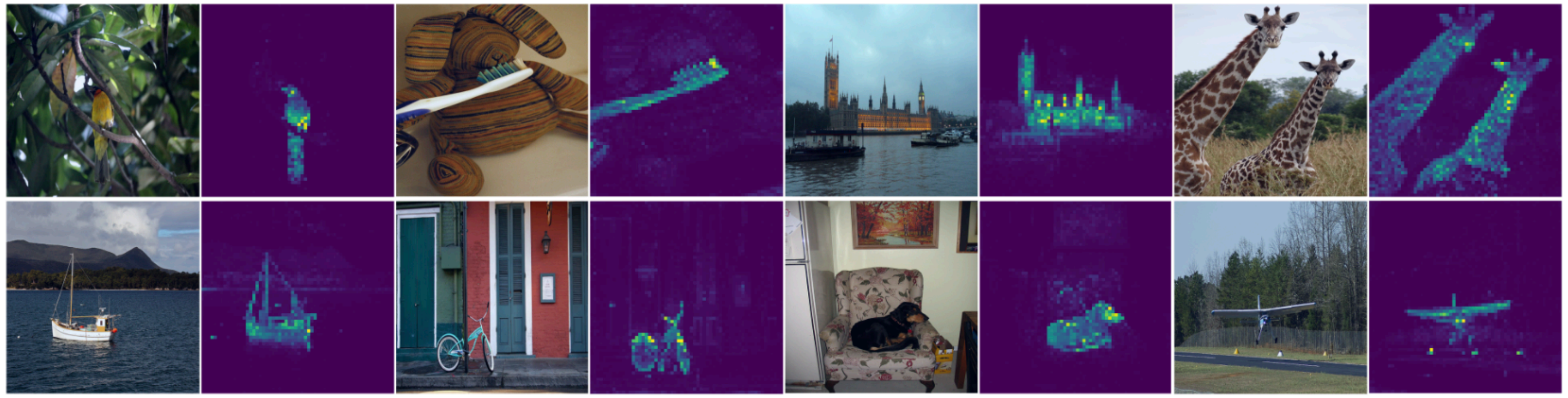


Figure 1: **Self-attention from a Vision Transformer with** 8×8 **patches trained with no supervision.** We look at the self-attention of the `[CLS]` token on the heads of the last layer. This token is not attached to any label nor supervision. These maps show that the model automatically learns class-specific features leading to unsupervised object segmentations.

Recent progress in Self-supervised Learning

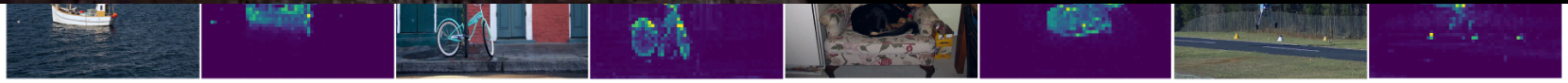
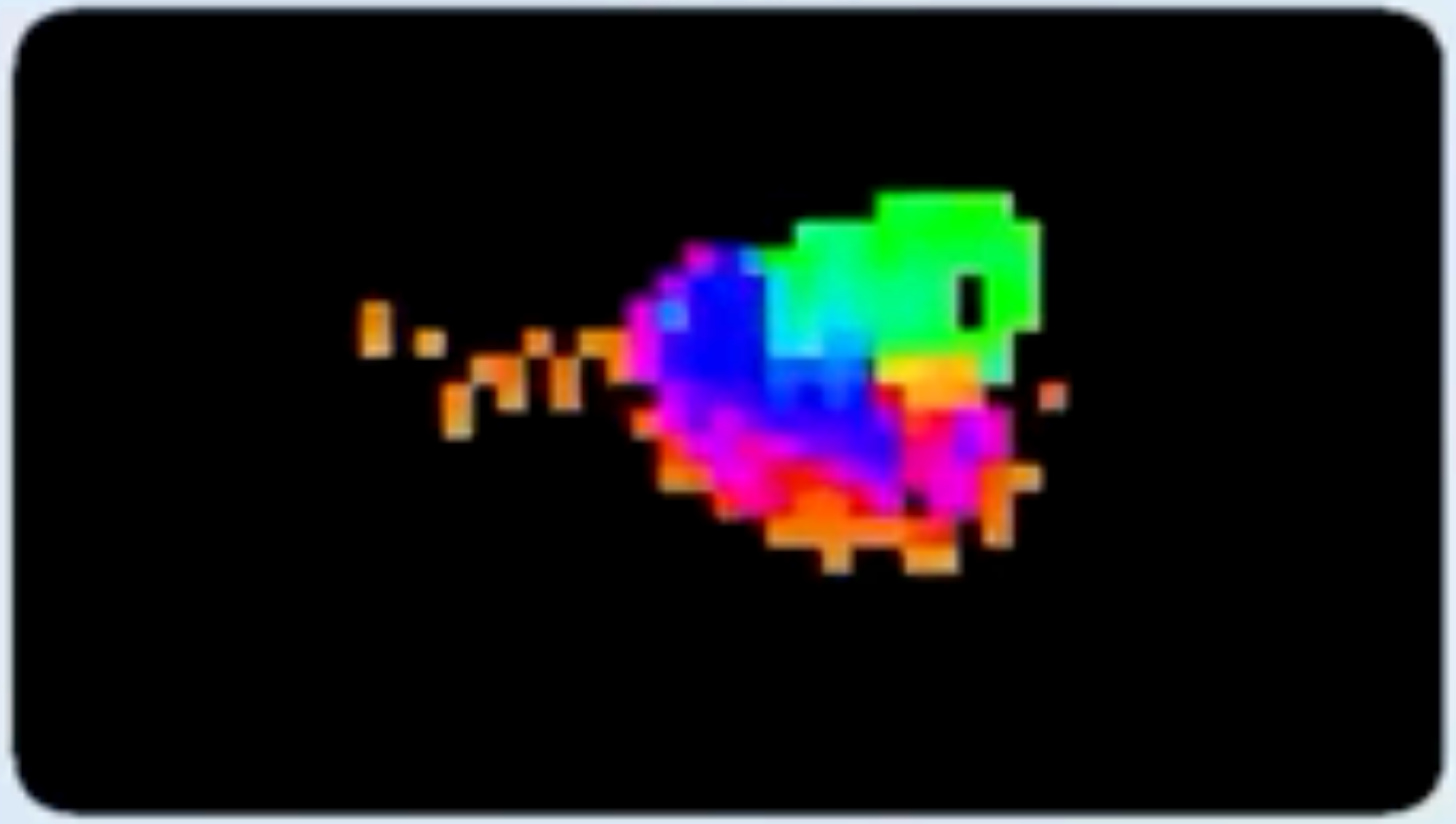


Figure 1: **Self-attention from a Vision Transformer with 8×8 patches trained with no supervision.** We look at the self-attention of the [CLS] token on the heads of the last layer. This token is not attached to any label nor supervision. These maps show that the model automatically learns class-specific features leading to unsupervised object segmentations.

Recent progress in Self-supervised Learning

DINO

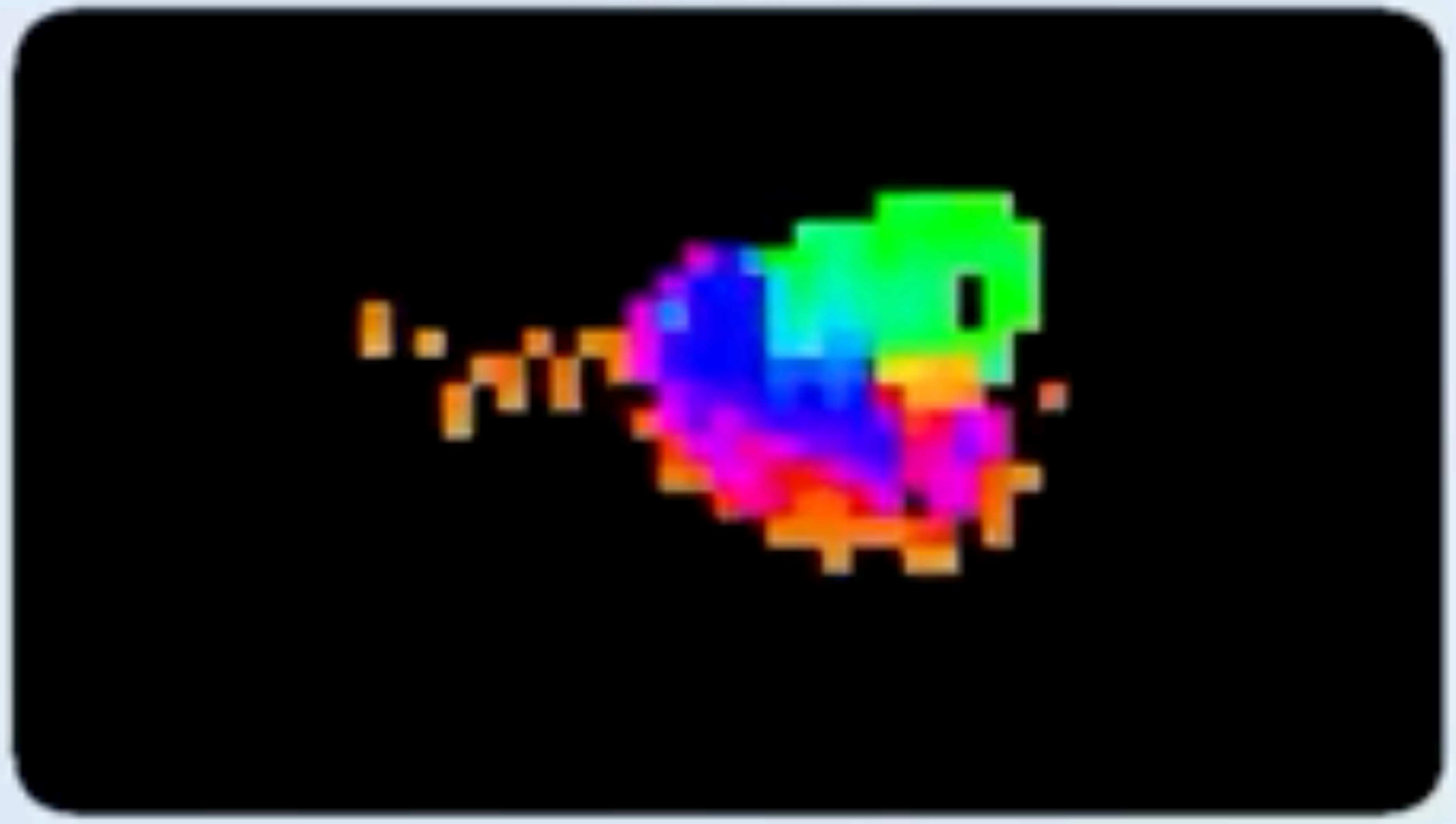


DINOv2

DINOv2, Oquab et al., arXiv 2023

Recent progress in Self-supervised Learning

DINO



DINOv2

DINOv2, Oquab et al., arXiv 2023

Recent progress in Self-supervised Learning

DINOv2, Oquab et al., arXiv 2023

Generative Modeling

https://thealgorithmicbridge.substack.com/p/stable-diffusion-is-the-most-important

Epic-Kitchens, Damen et al. 2018

Epic-Kitchens, Damen et al. 2018

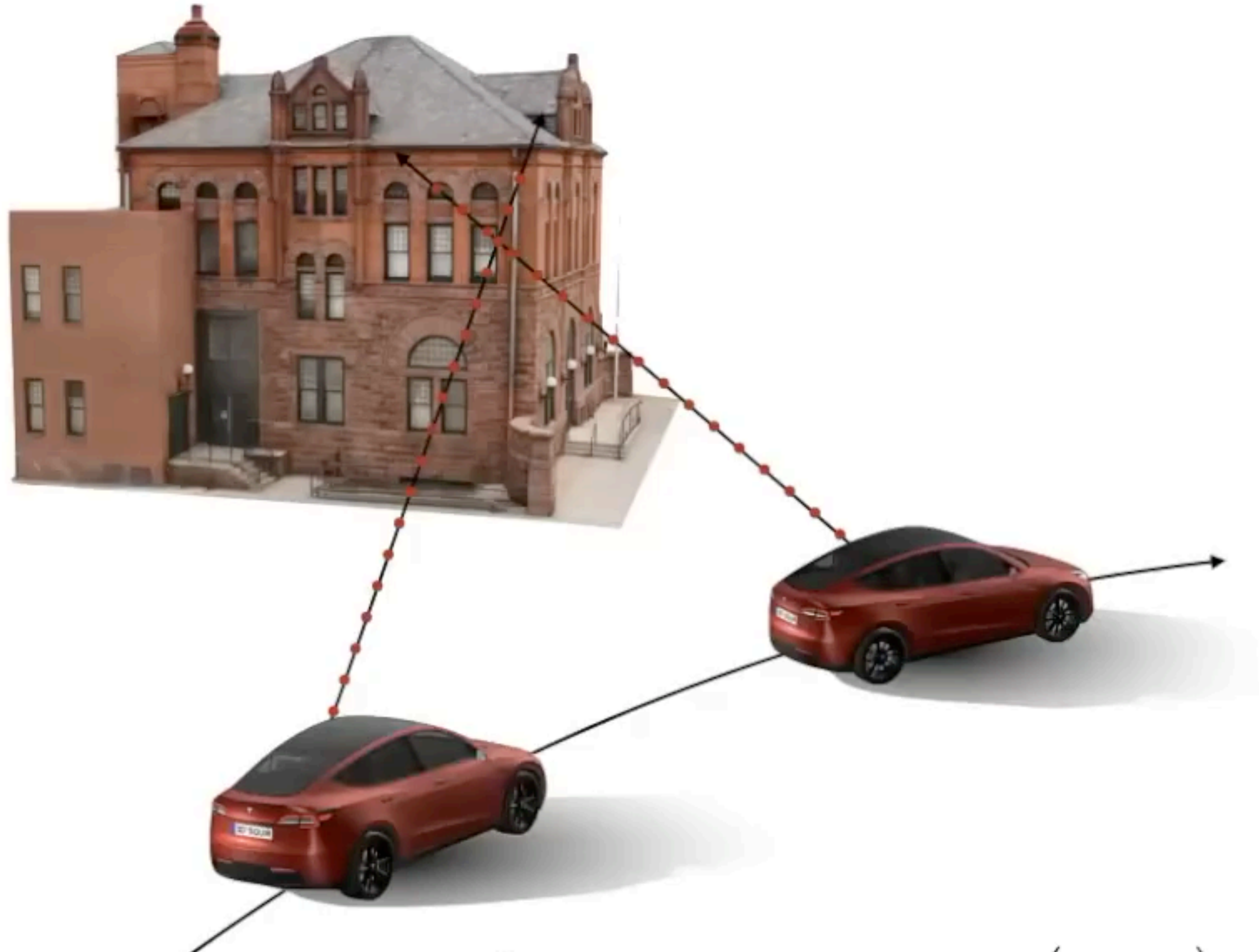
Epic-Kitchens, Damen et al. 2018

The role of 3D in ML and Vice-Versa

- Any model that interacts with our world has to make predictions about 3D.
- Need 3D structure to be robust to 3D transformations. Augmentation can help, but does not guarantee generalization.
- Necessary across perception, robotics, graphics, augmented reality, virtual reality, self-driving, simulation…

Moving away from closed-set predictions towards 3D!

From CVPR Workshop on Autonomous Driving Keynote, Ashok Elluswamy, Tesla

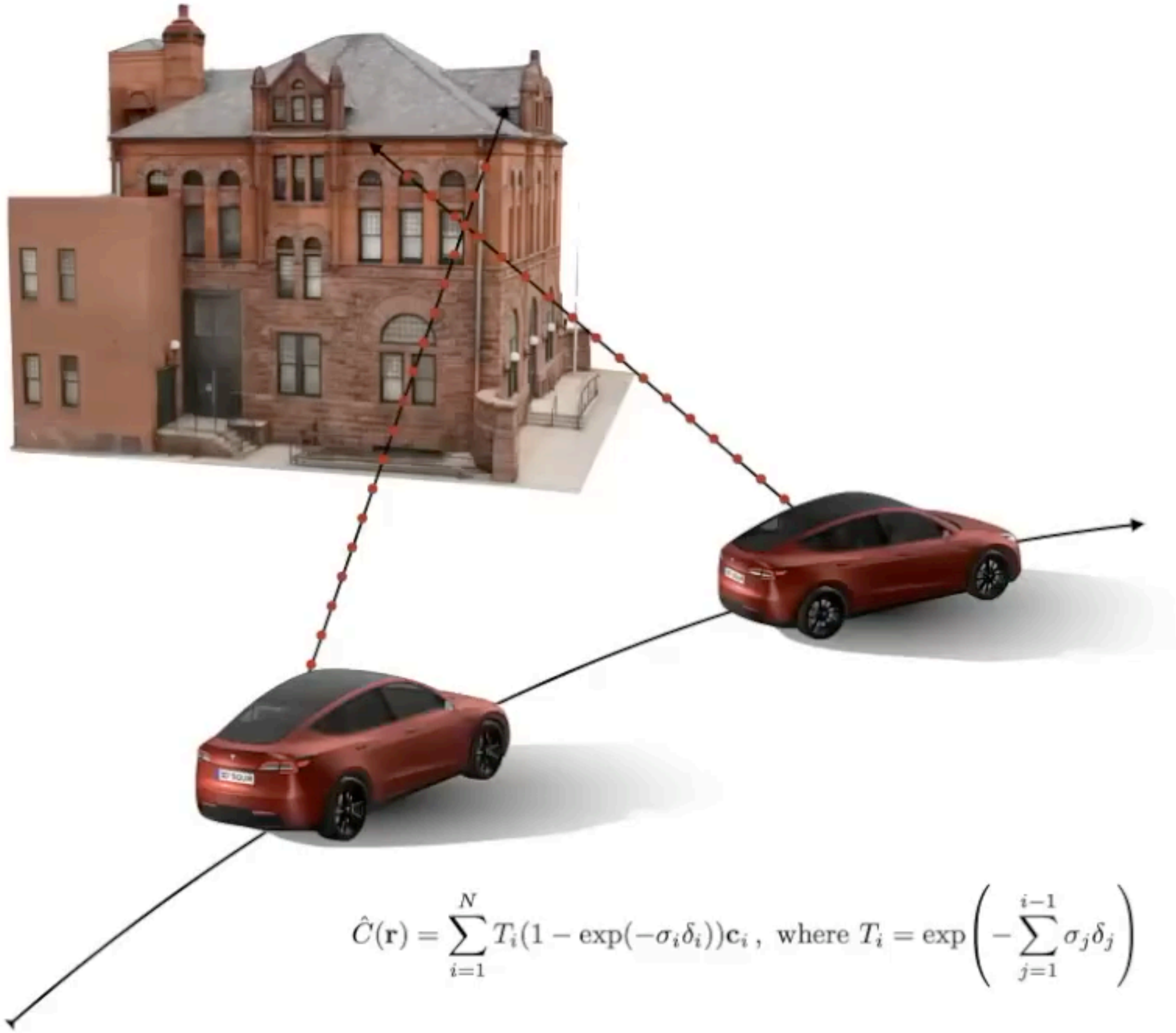


$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i \delta_i))\mathbf{c}_i\,, \text{ where } T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$$

[1] NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis
[2] Plenoxels: Radiance Fields without Neural Networks

Moving away from closed-set predictions towards 3D!

From CVPR Workshop on Autonomous Driving Keynote, Ashok Elluswamy, Tesla



$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i \delta_i))\mathbf{c}_i\,, \text{ where } T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$$

[1] NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis
[2] Plenoxels: Radiance Fields without Neural Networks

3D Computer Vision for AR / VR

No Scene understanding yet - that's our job!

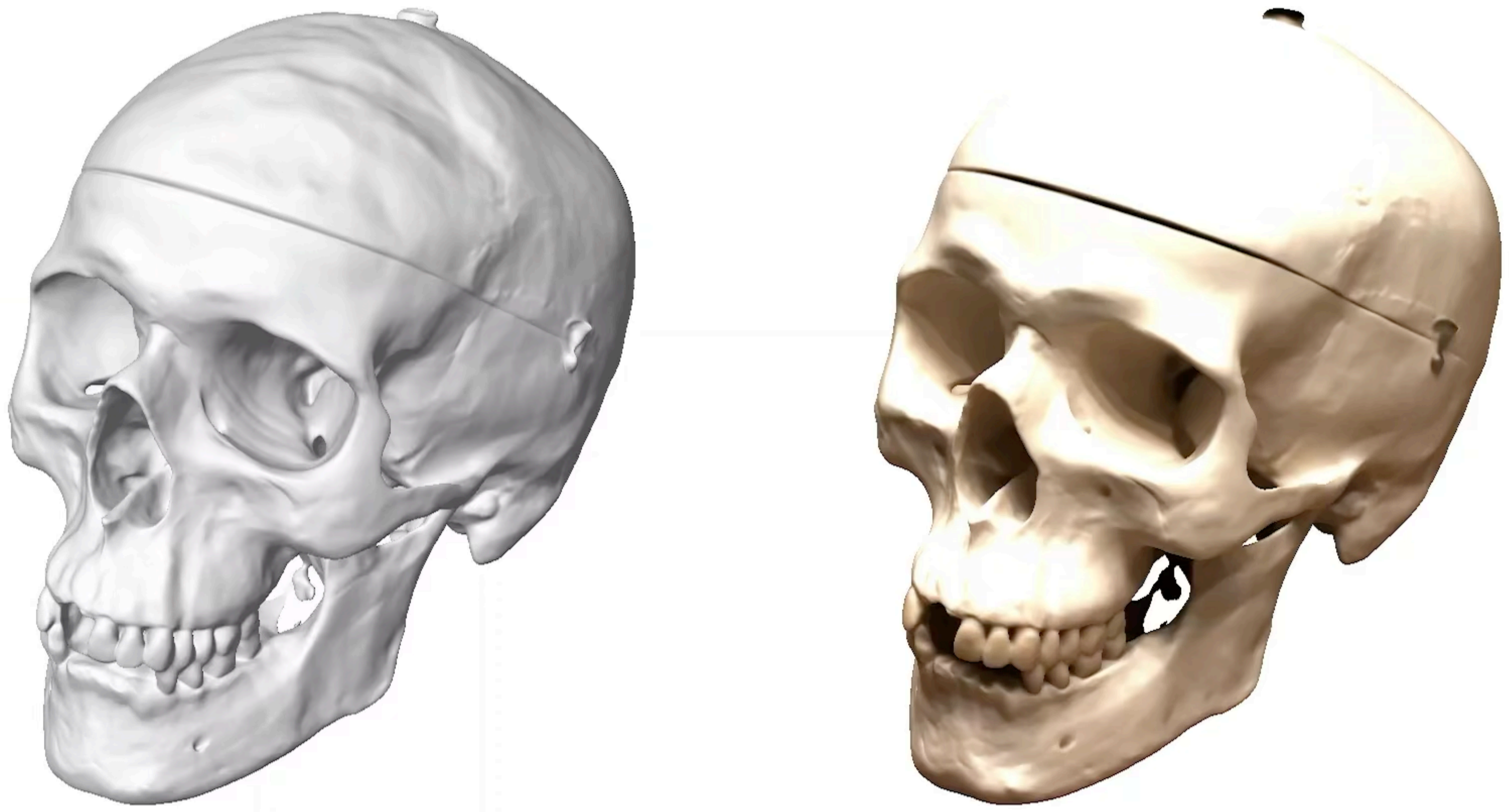
https://ai.facebook.com/blog/powered-by-ai-oculus-insight/

3D Computer Vision for AR / VR

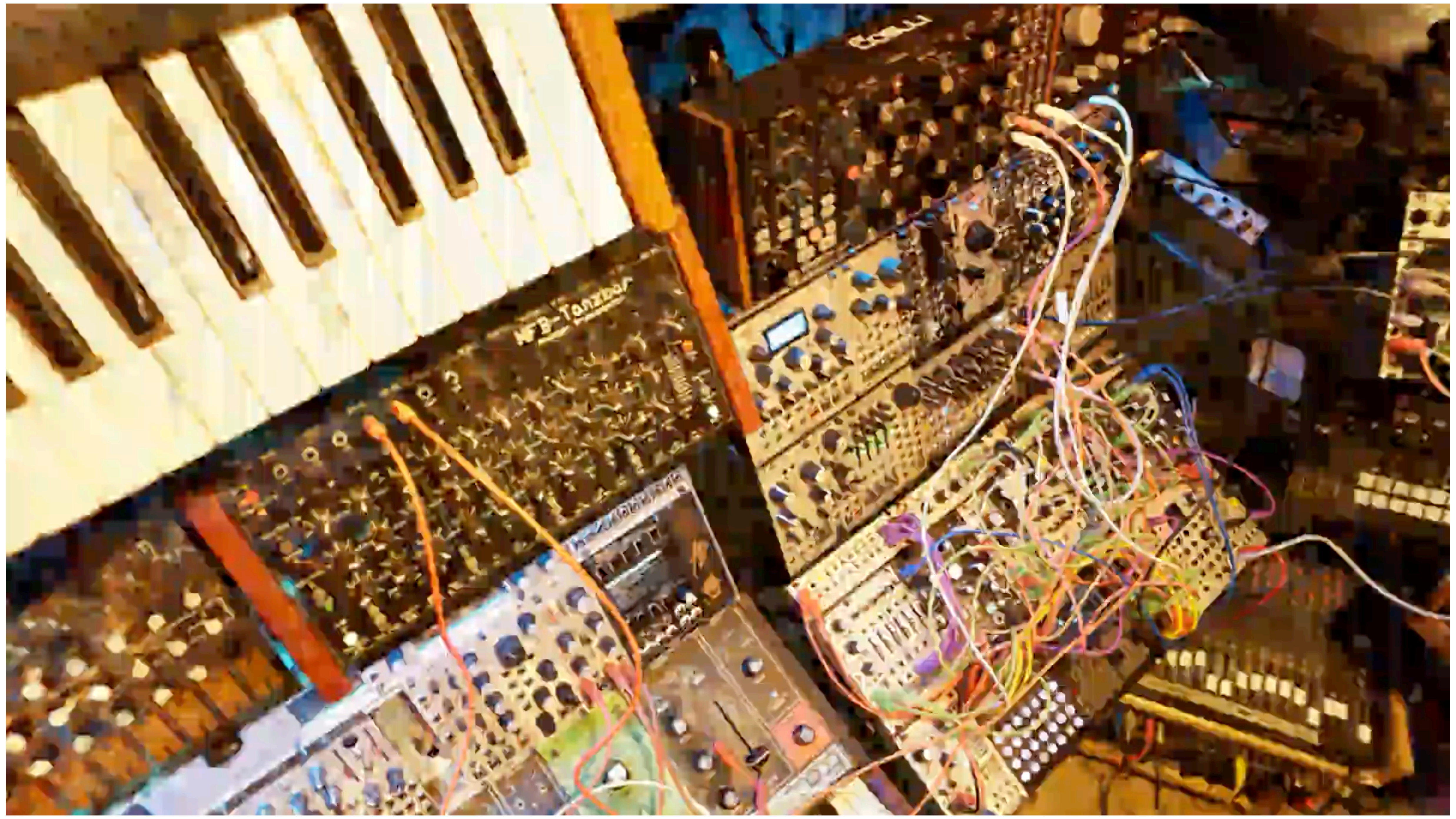
No Scene understanding yet - that's our job!

https://ai.facebook.com/blog/powered-by-ai-oculus-insight/

Graphics

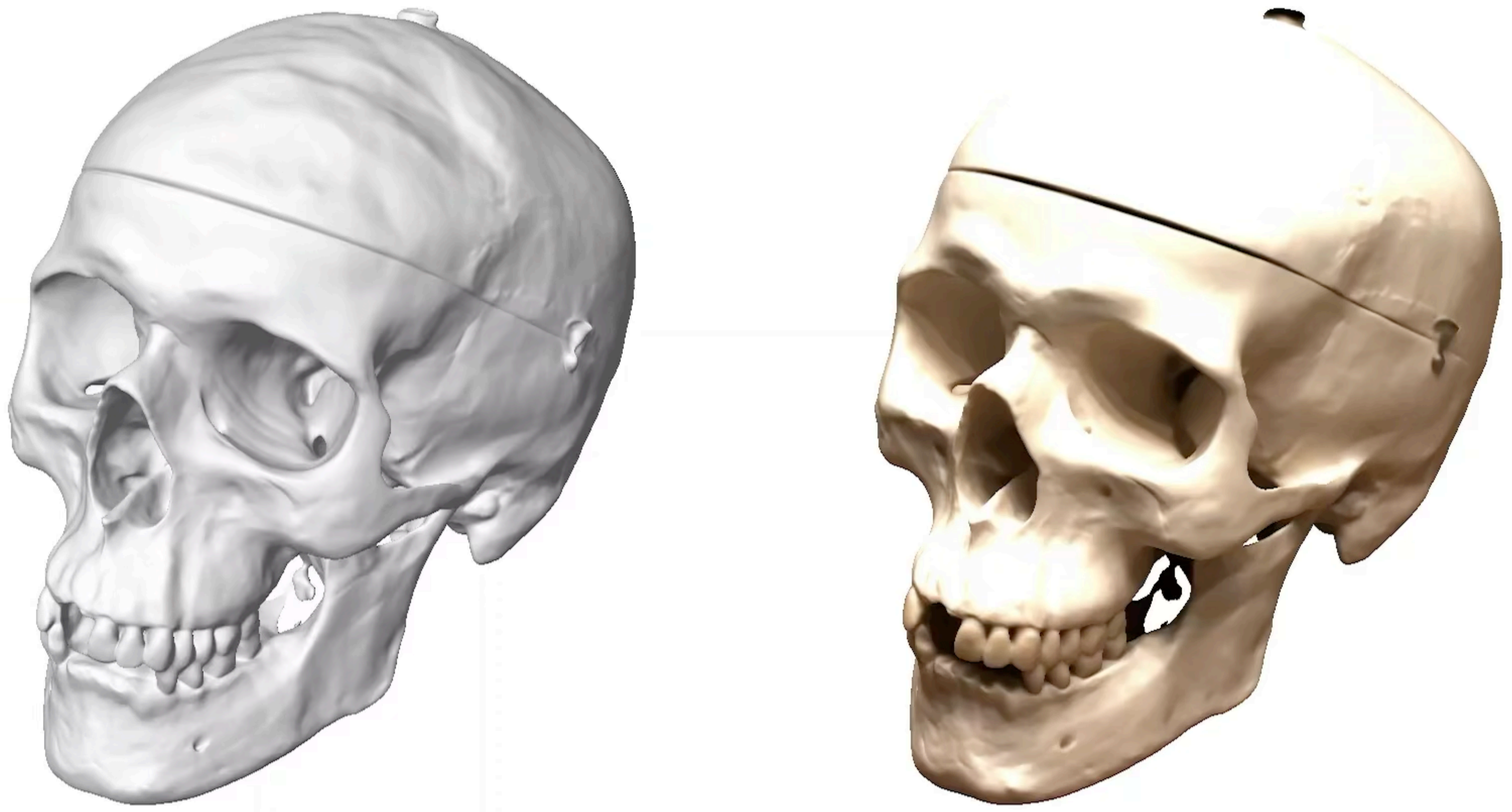


Implicit Differentiable Renderer, Yariv et al. 2020

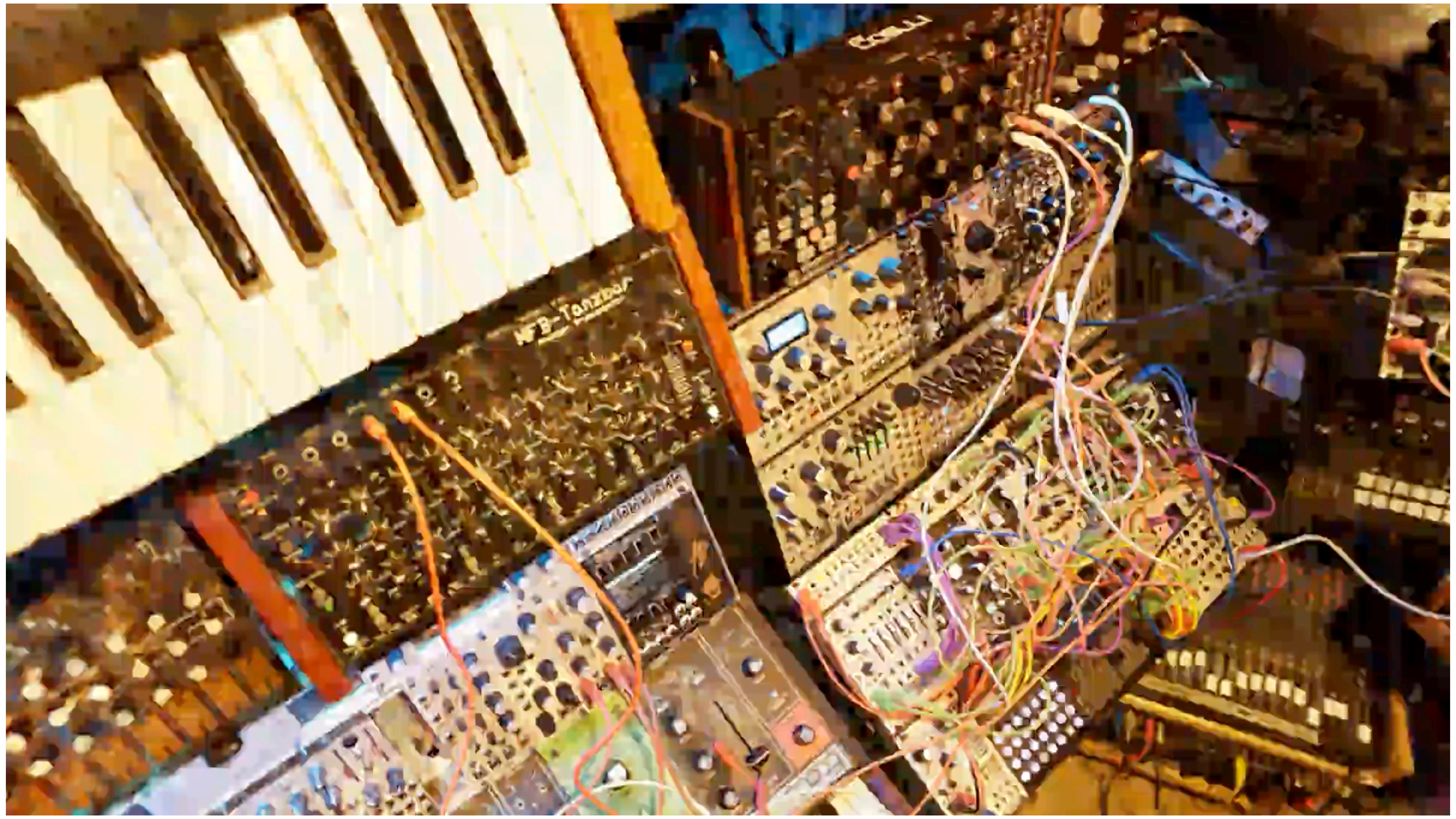


InstantNGP, Müller et al. 2022

Graphics



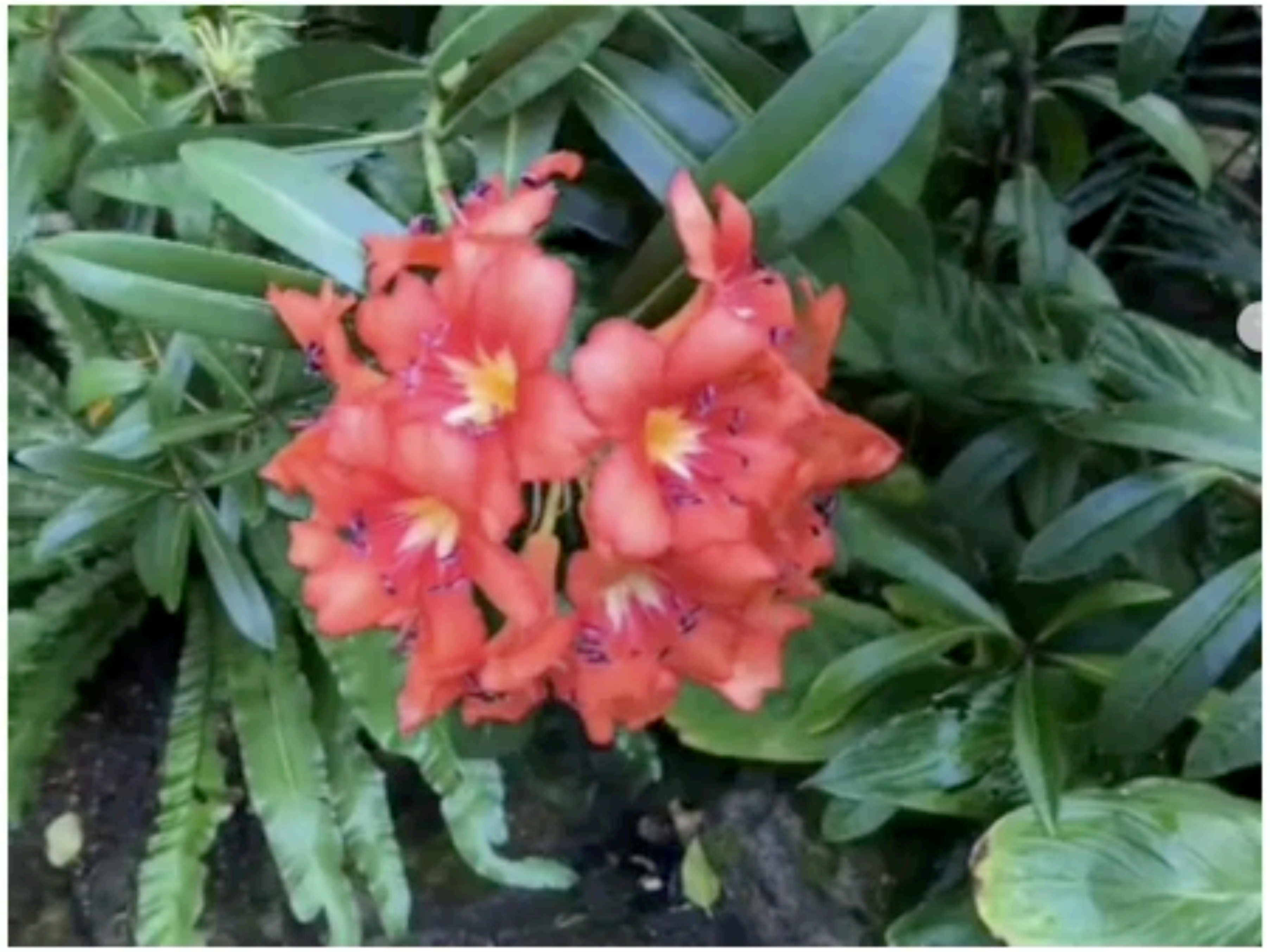
Implicit Differentiable Renderer, Yariv et al. 2020



InstantNGP, Müller et al. 2022

Graphics

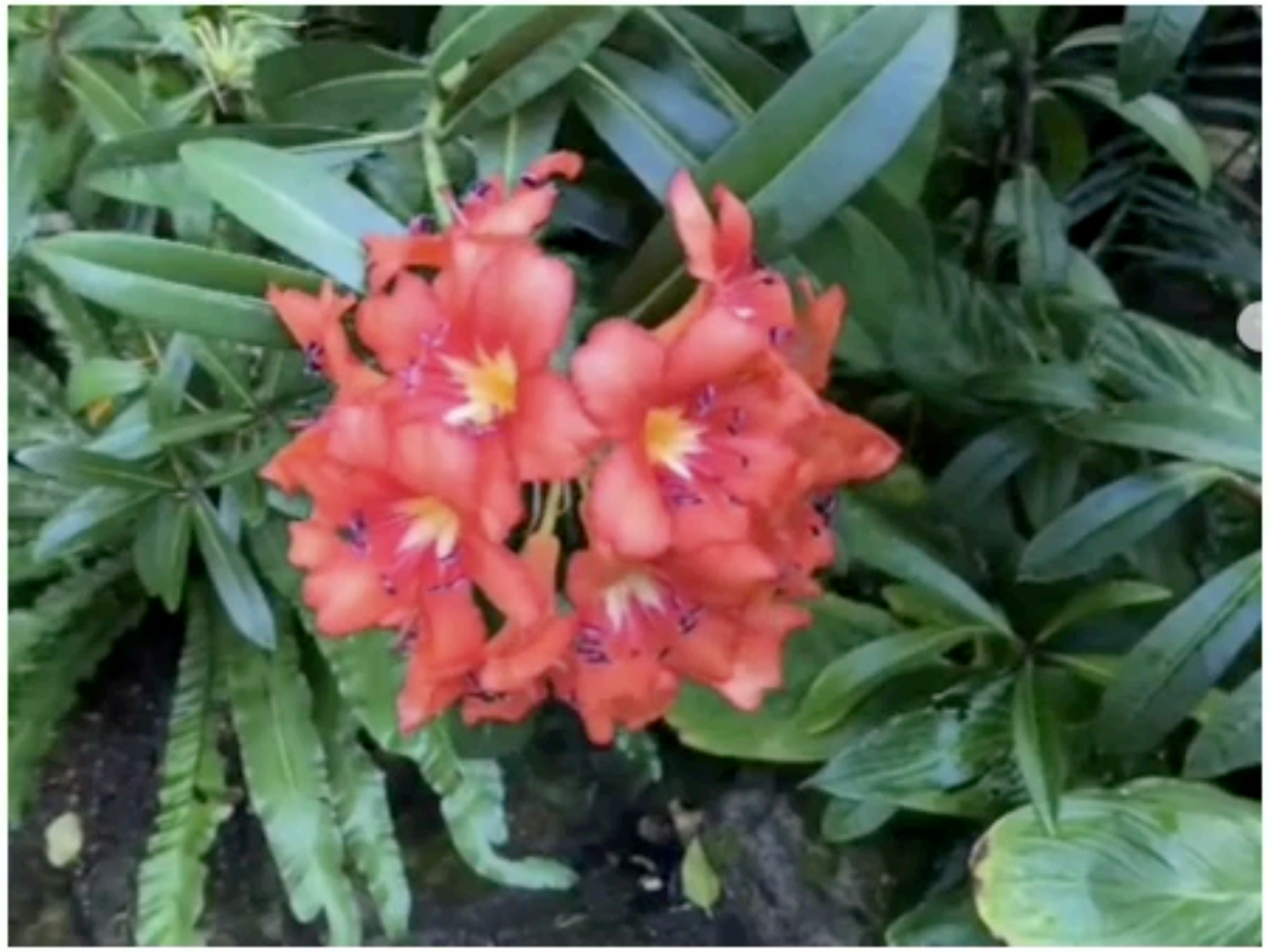
EG3D, Chan et al. 2022



Decomposing NeRF for Editing via Feature Field Distillation, Kobayashi et al. 2022

Graphics

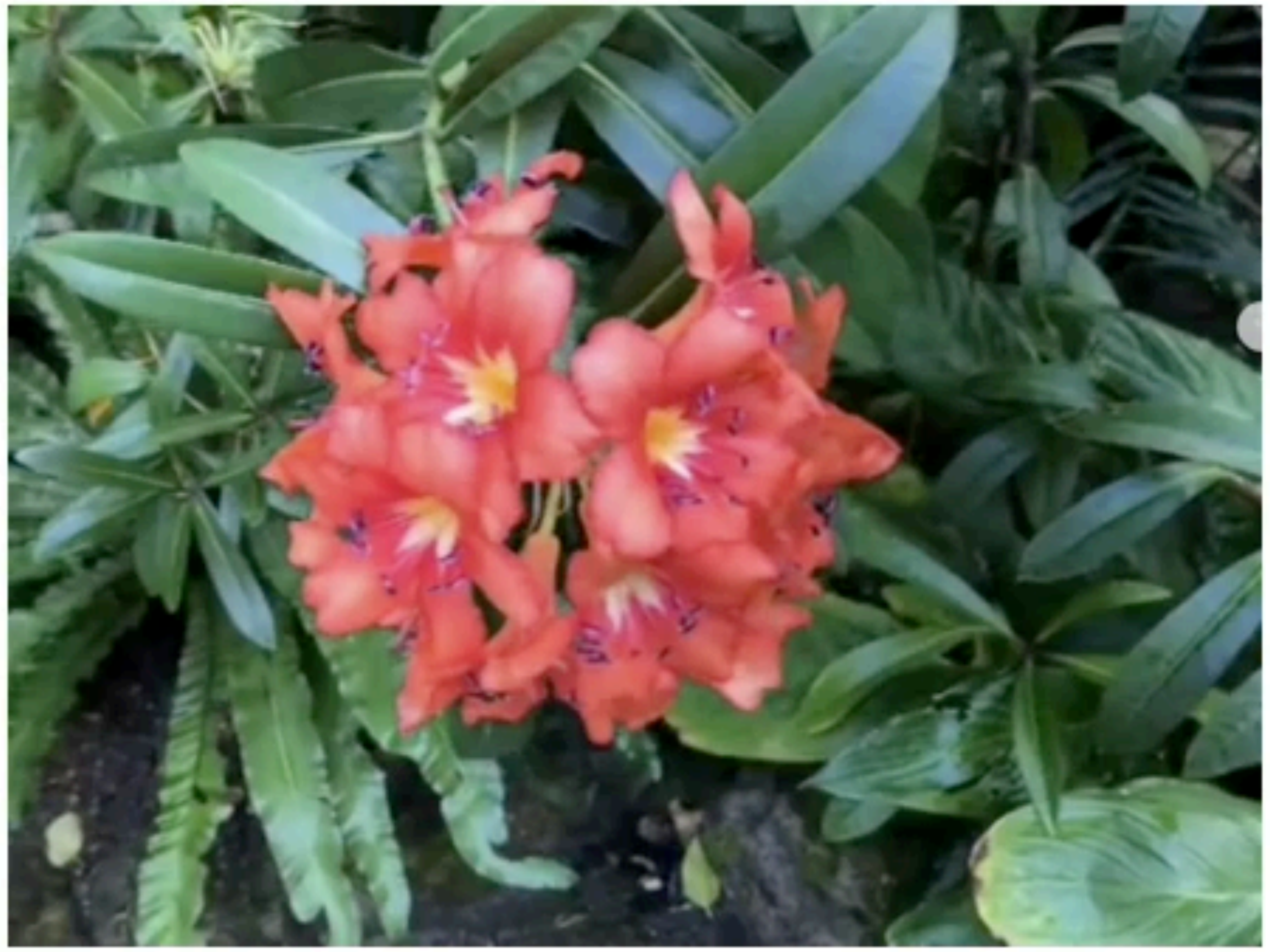
EG3D, Chan et al. 2022



Decomposing NeRF for Editing via Feature Field Distillation, Kobayashi et al. 2022

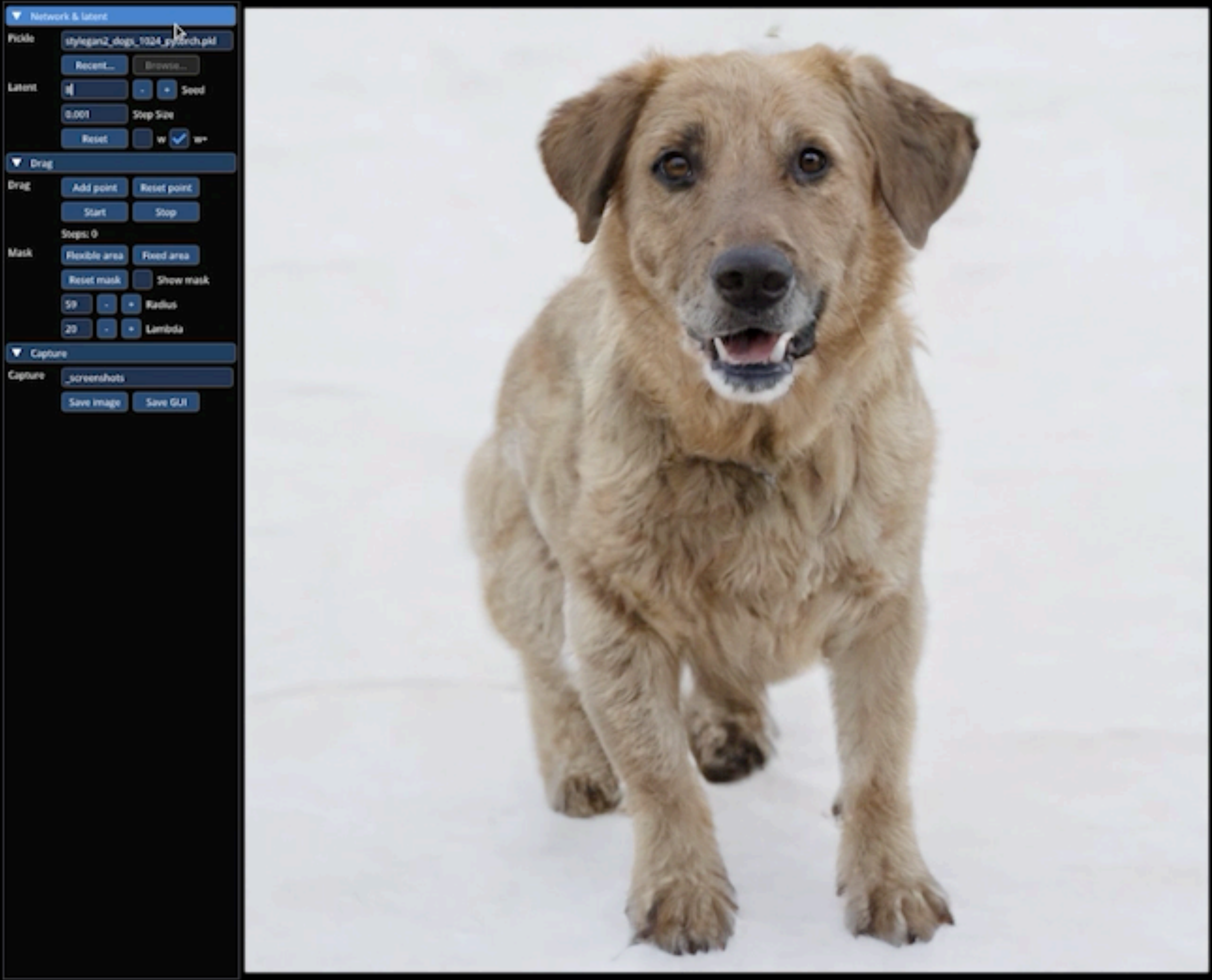
Graphics

EG3D, Chan et al. 2022



Decomposing NeRF for Editing via Feature Field Distillation, Kobayashi et al. 2022

Drag Your GAN, Pan et al., SIGGRAPH 2023

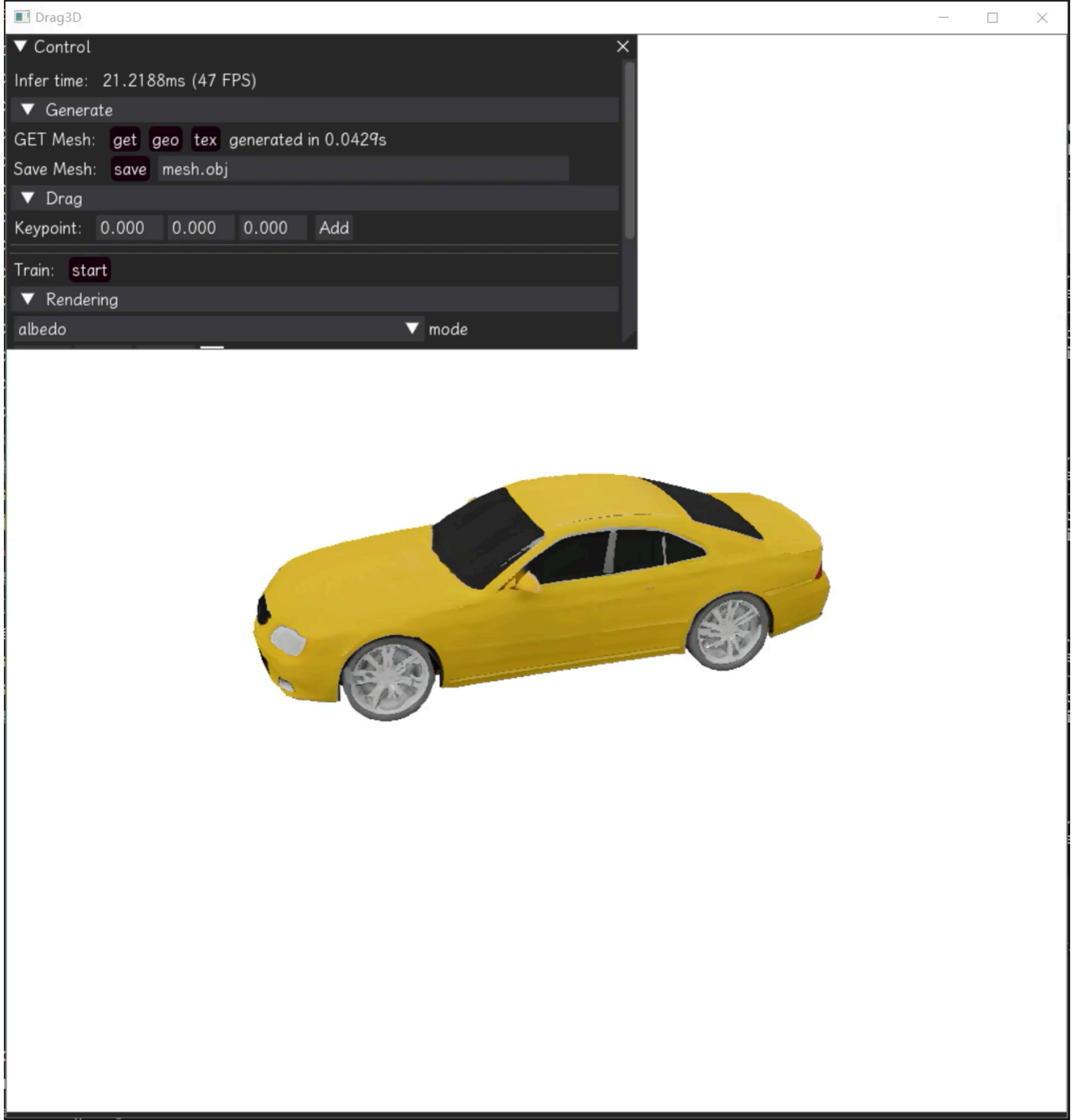


Drag Your GAN, Pan et al., SIGGRAPH 2023

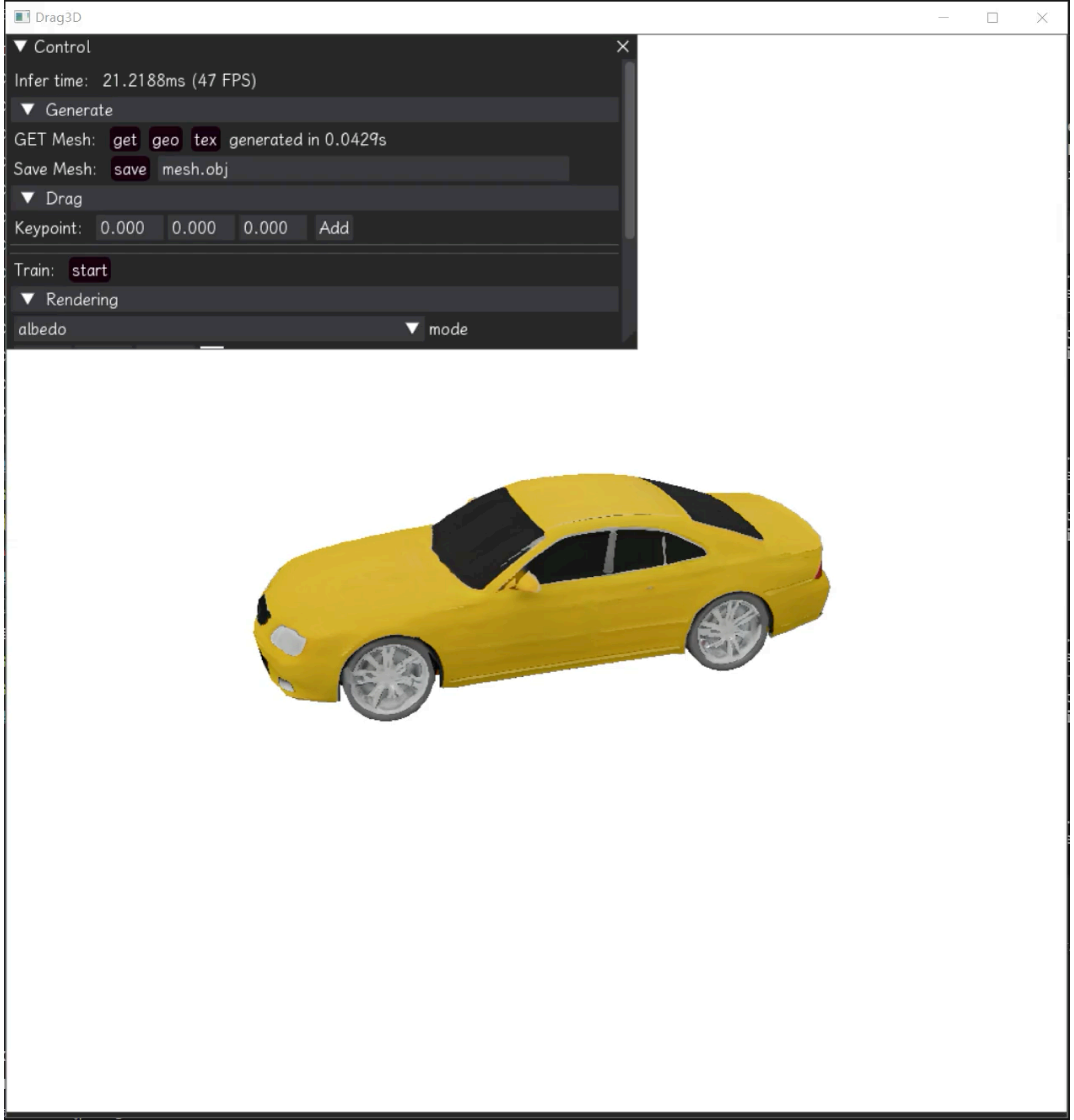
https://www.youtube.com/watch?app=desktop&v=bahZegyn1aU

https://www.youtube.com/watch?app=desktop&v=bahZegyn1aU

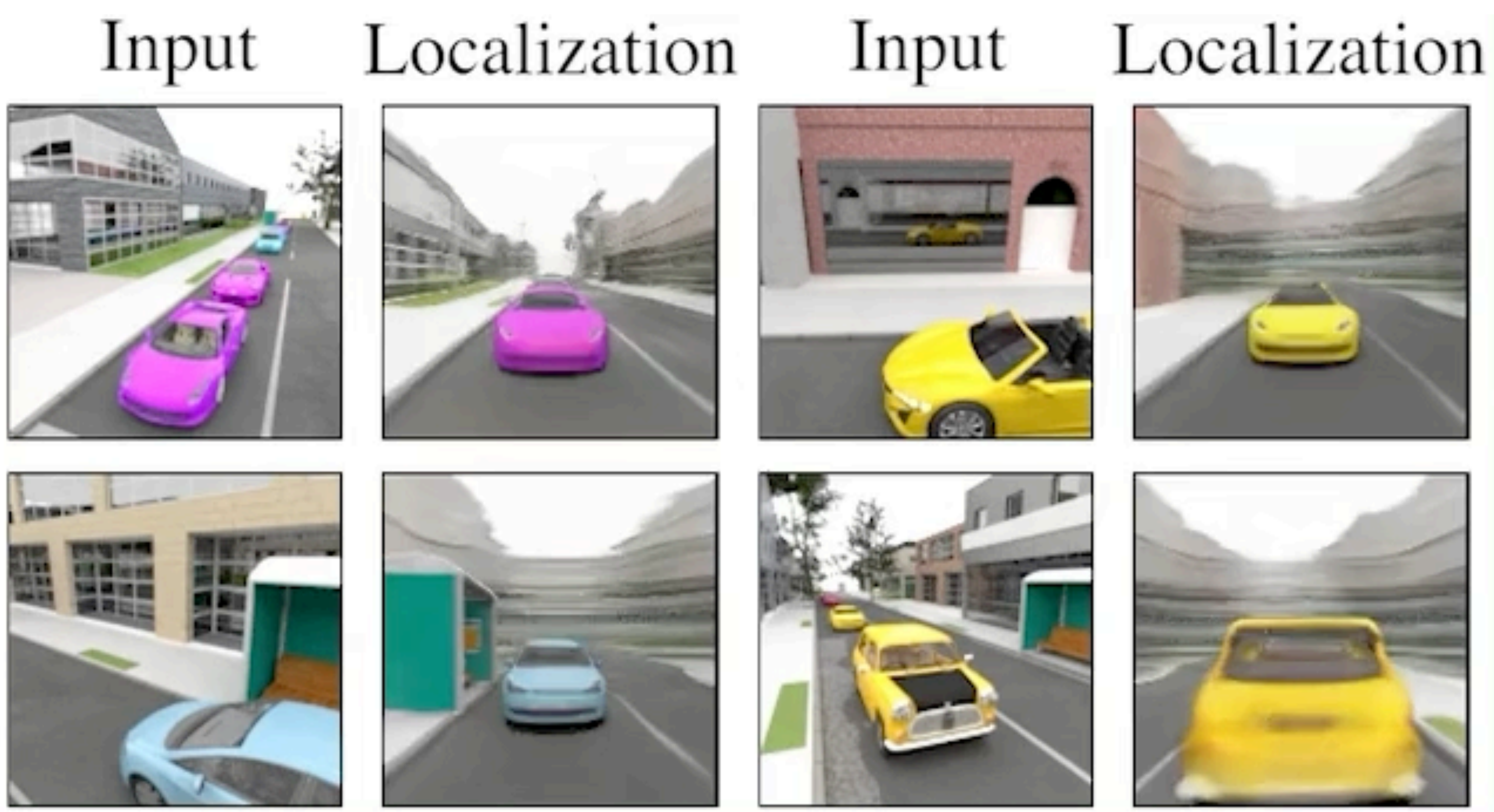
https://github.com/ashawkey/Drag3D



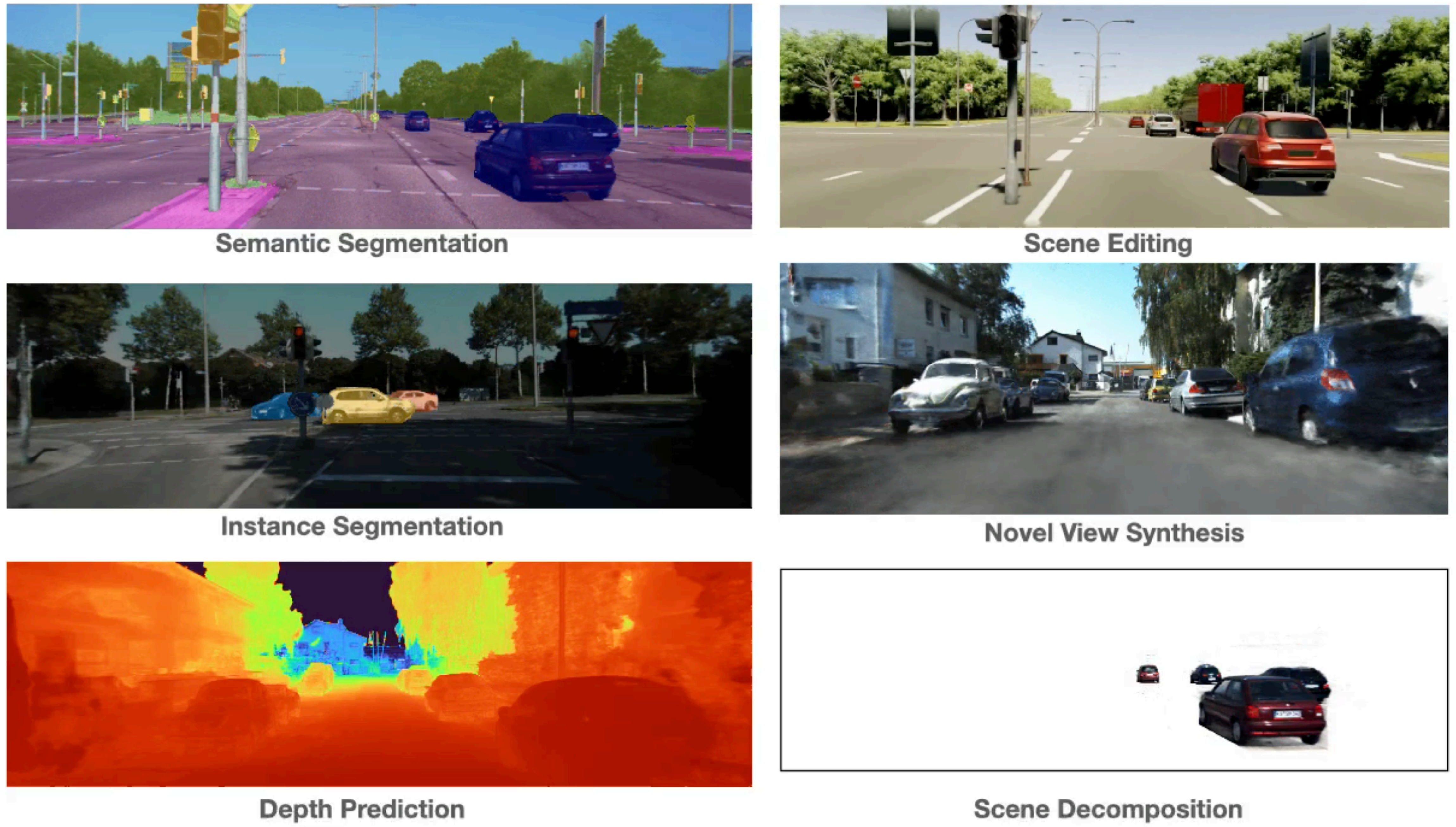
https://github.com/ashawkey/Drag3D



Vision

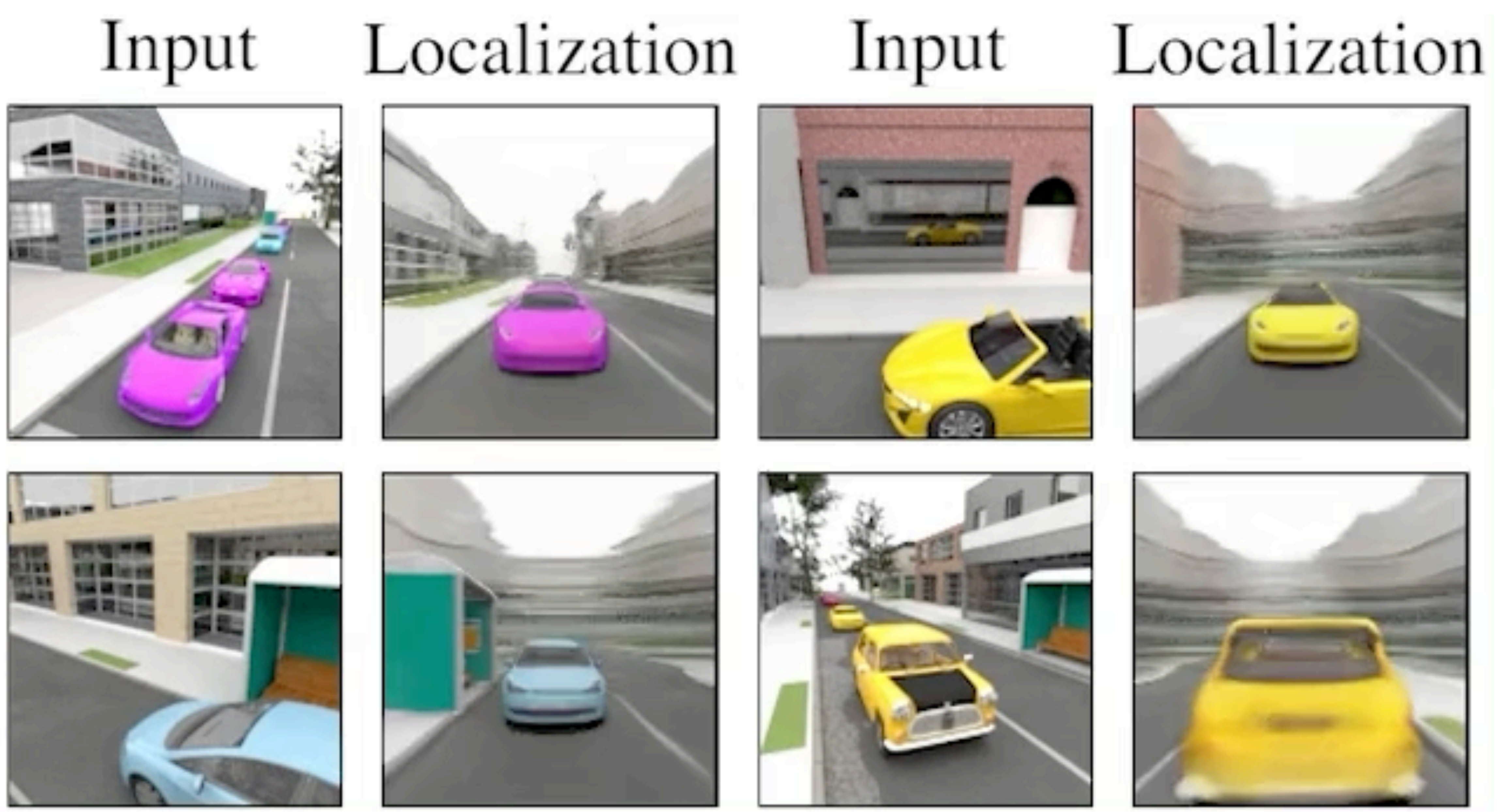


See3D, Sharma et al. 2022

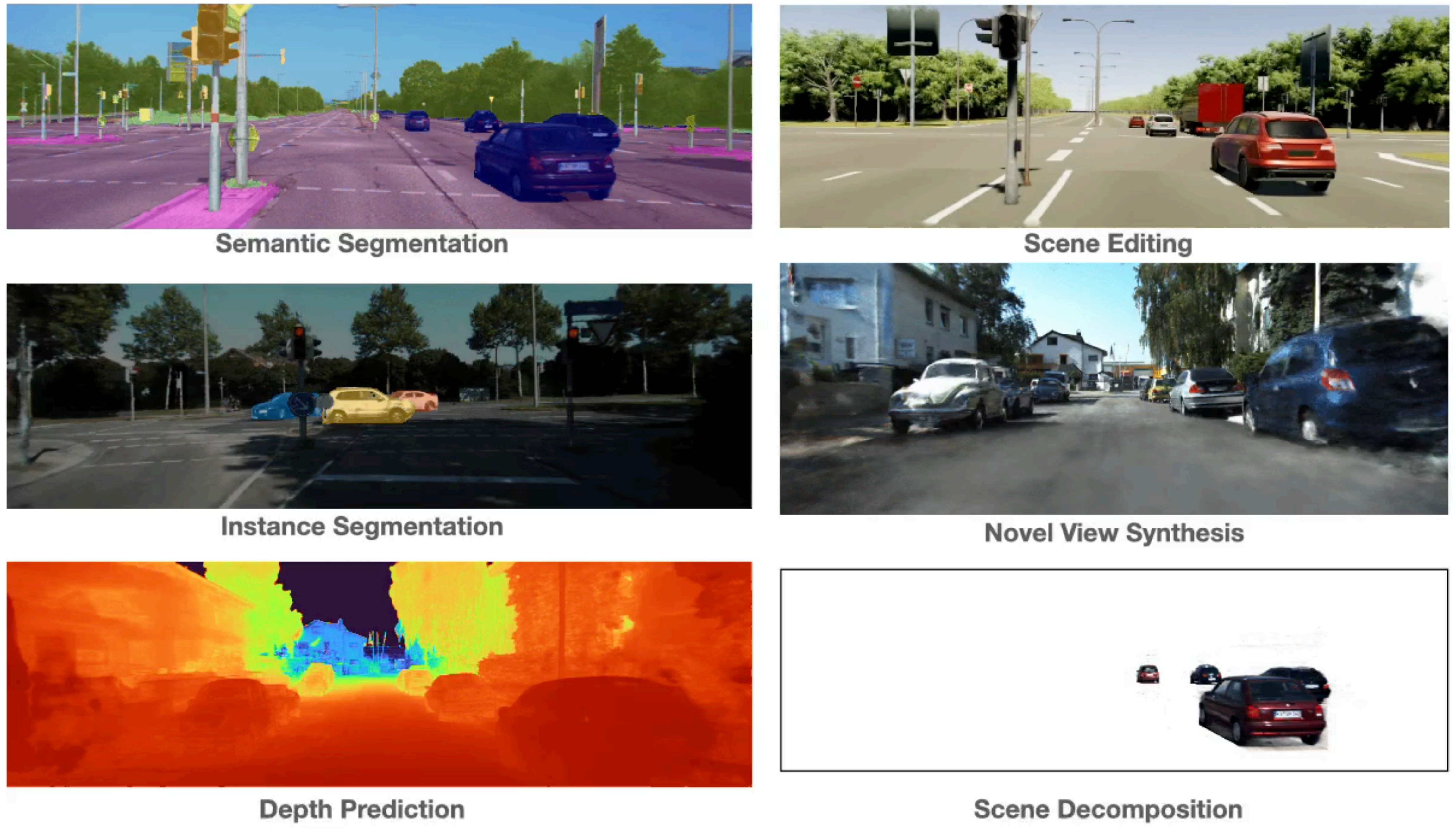


Panoptic Neural Fields, Kundu et al. 2022

Vision



See3D, Sharma et al. 2022



Panoptic Neural Fields, Kundu et al. 2022

3D Reconstruction from a Single Image

?

Diffusion with Forward Models: Solving Stochastic Inverse Problems Without Direct Supervision, Tewari et al. 2023

3D Reconstruction from a Single Image

?

Diffusion with Forward Models: Solving Stochastic Inverse Problems Without Direct Supervision, Tewari et al. 2023

3D Reconstruction from a Single Image

?

Diffusion with Forward Models: Solving Stochastic Inverse Problems Without Direct Supervision, Tewari et al. 2023

3D Reconstruction from a Single Image

?

Diffusion with Forward Models: Solving Stochastic Inverse Problems Without Direct Supervision, Tewari et al. 2023

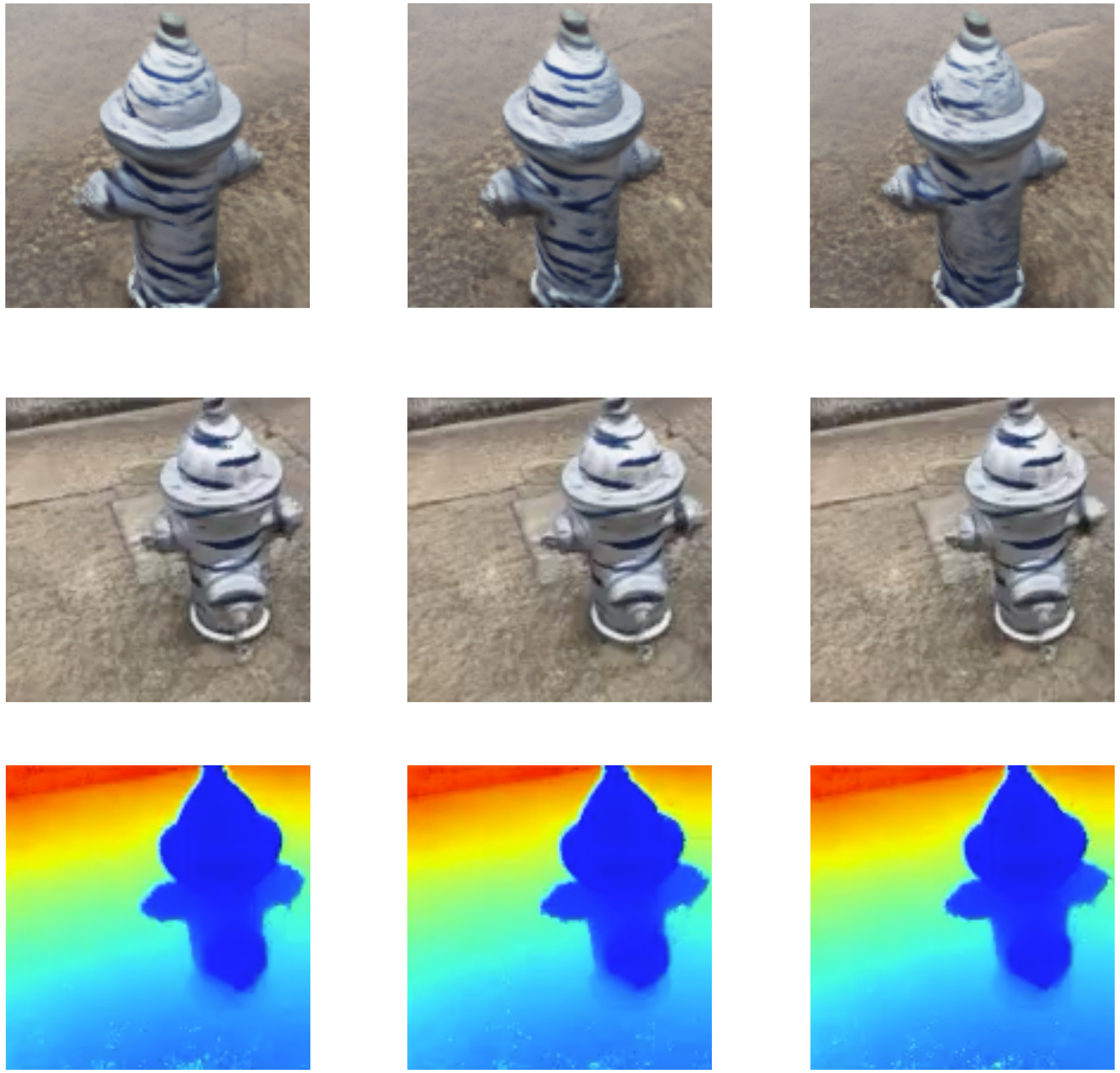
3D Reconstruction from a Single Image

?

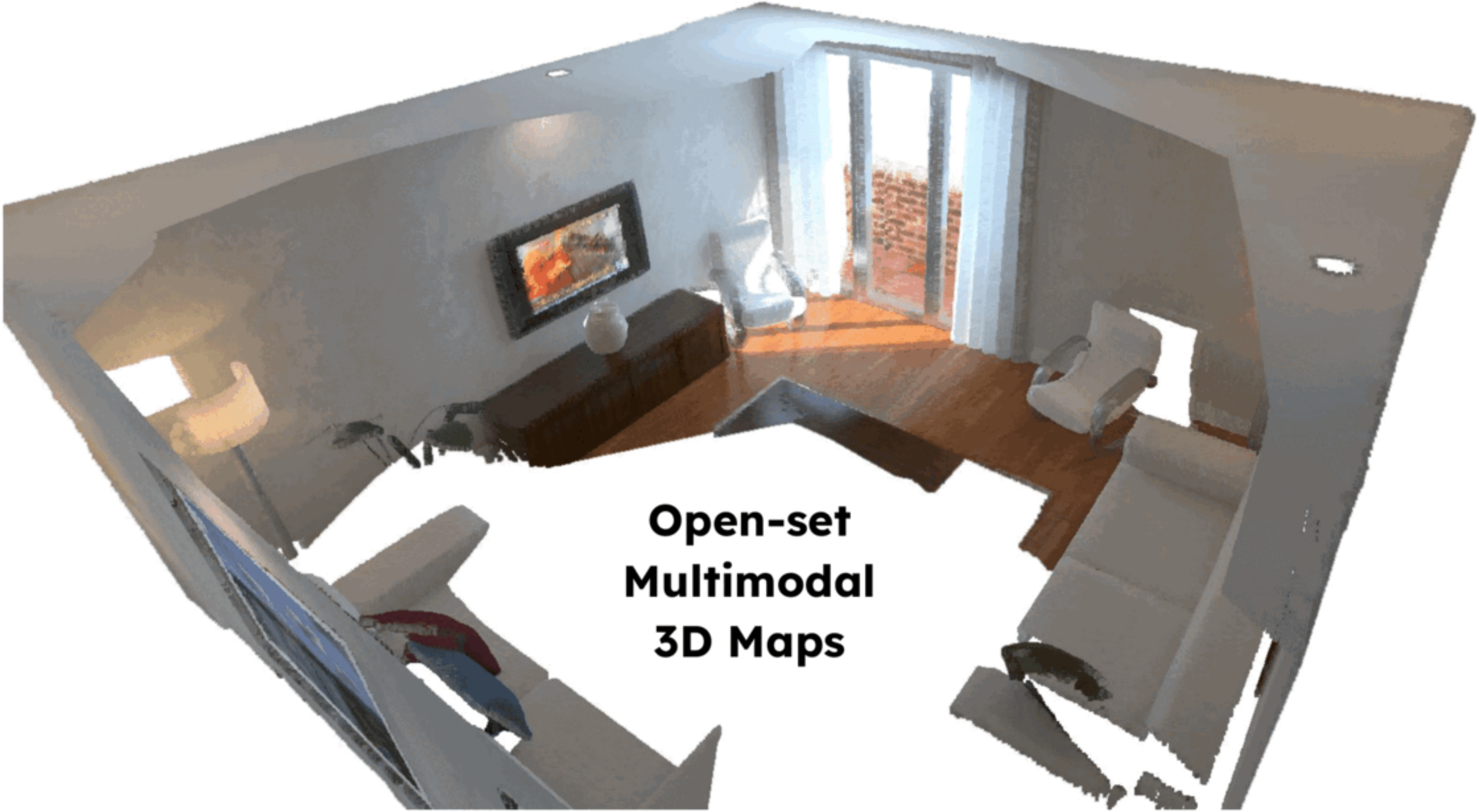
Diffusion with Forward Models: Solving Stochastic Inverse Problems Without Direct Supervision, Tewari et al. 2023

3D Reconstruction from a Single Image

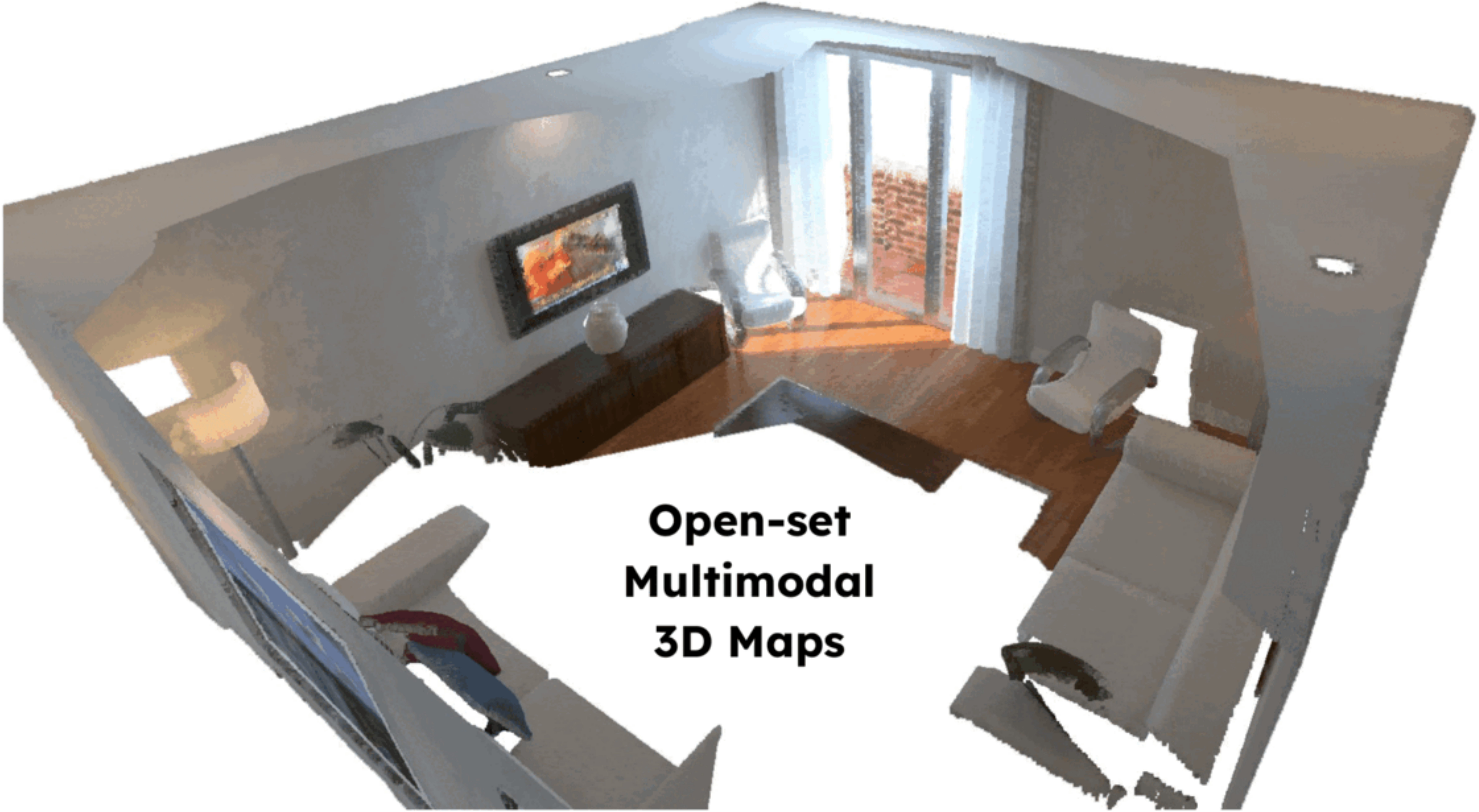
?



Diffusion with Forward Models: Solving Stochastic Inverse Problems Without Direct Supervision, Tewari et al. 2023



ConceptFusion, Jatavallabhula et al., RSS 2023



ConceptFusion, Jatavallabhula et al., RSS 2023

RoDynRF: Robust Dynamic Radiance Fields, Liu et al. 2023, CVPR

RoDynRF: Robust Dynamic Radiance Fields, Liu et al. 2023, CVPR

TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement, Doersch et al., 2023

TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement, Doersch et al., 2023

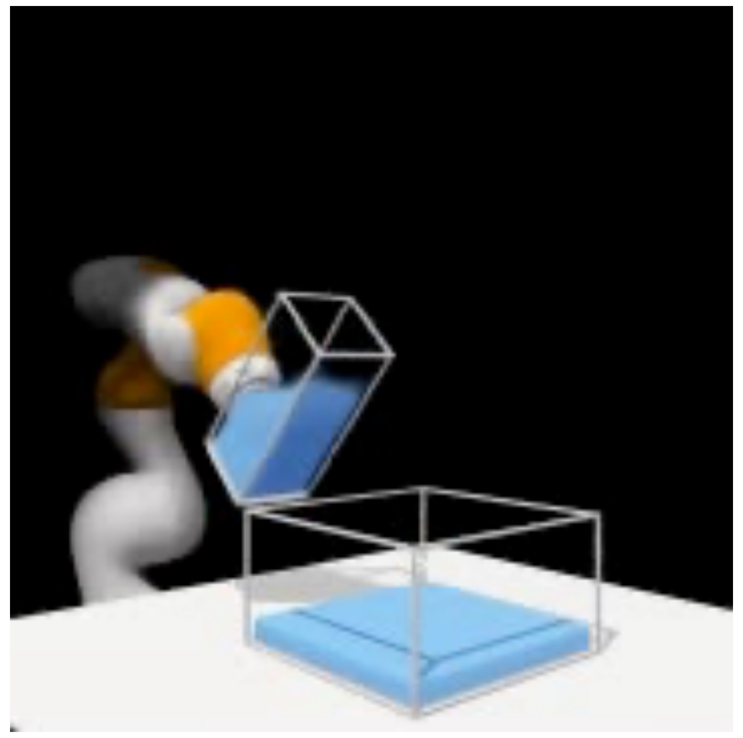
Input RGBD video

Input RGBD video

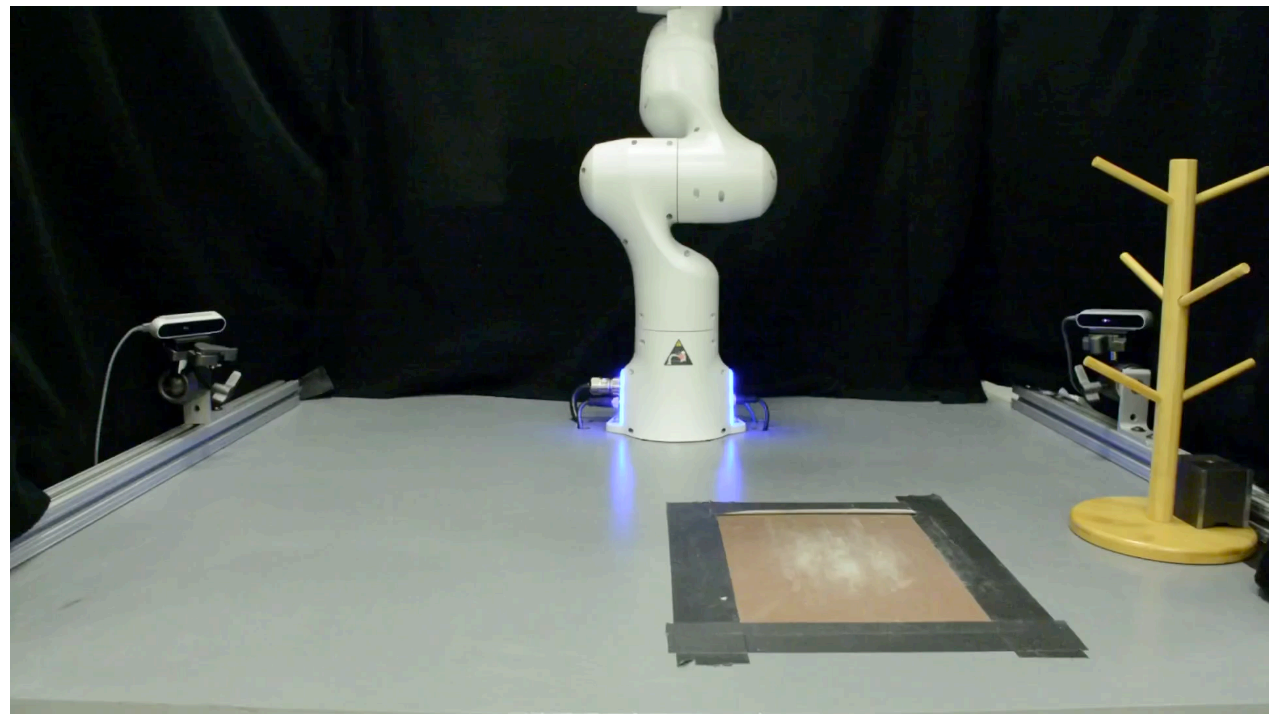
Total-Recon, Song et al., ICCV 2023

Robotics

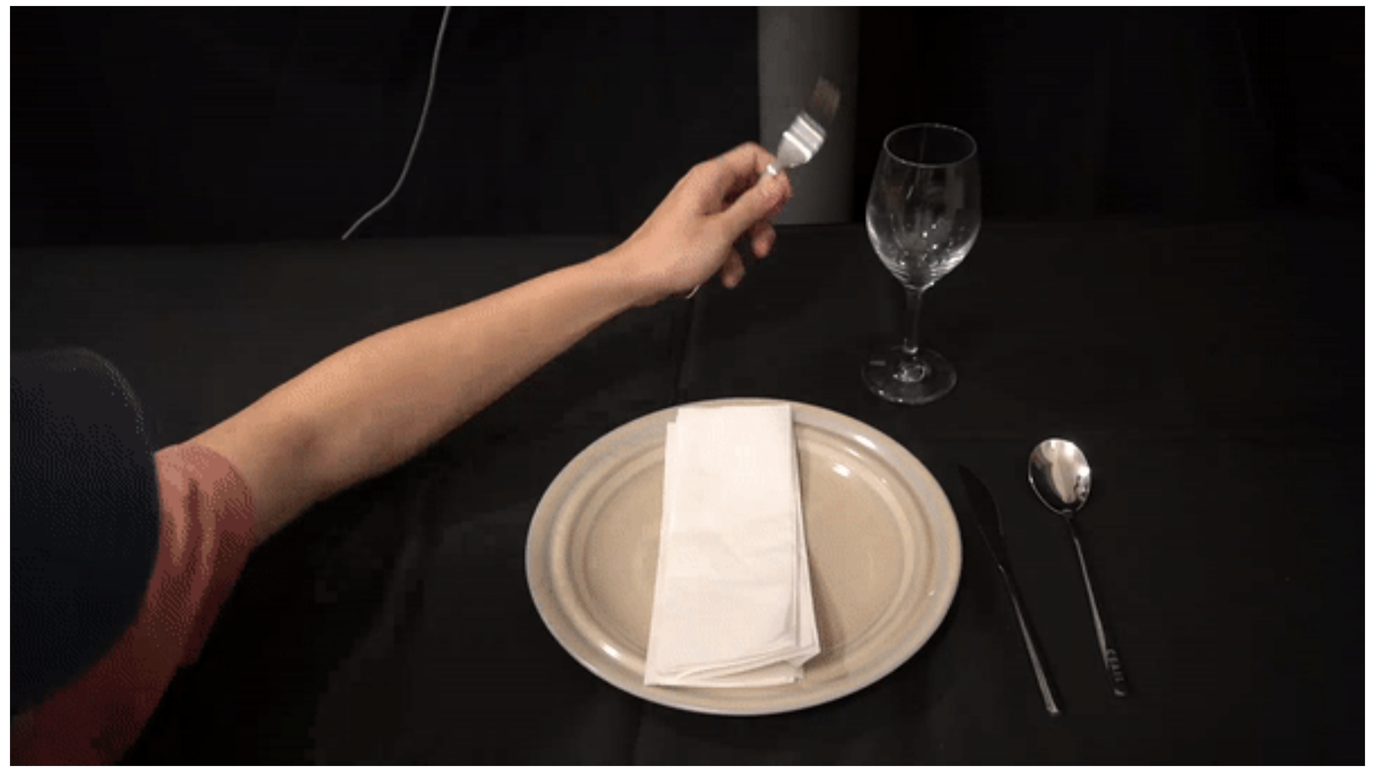
3d neural scene representations for visuomotor control
Li & Li et al. CORL 2022



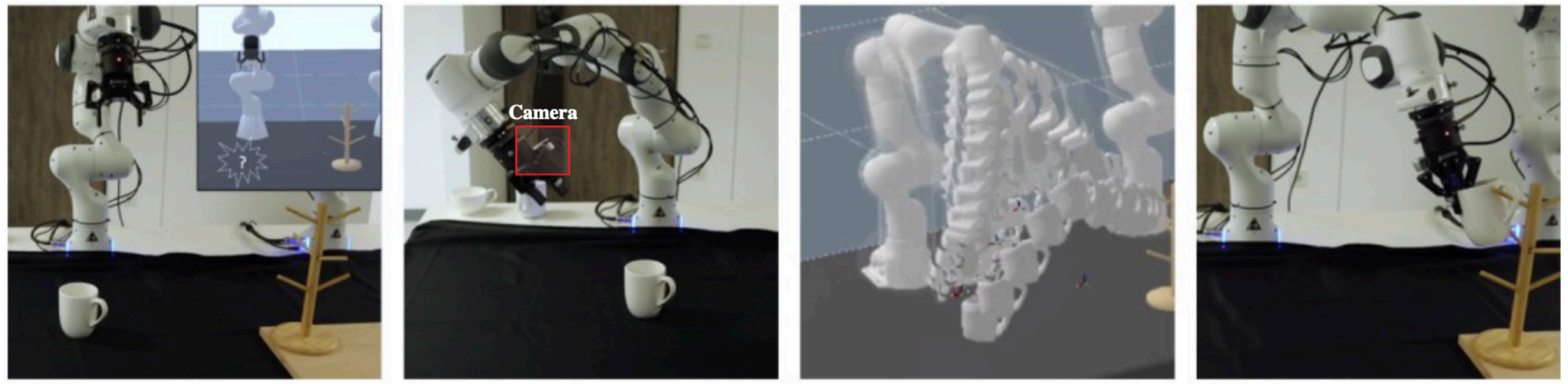
Neural descriptor fields: Se (3)-equivariant object representations for manipulation
Simeonov & Du et al., ICRA 2022



NeRF-Supervision: Learning Dense Object Descriptors from Neural Radiance Fields
Lin et al. ICRA 2022



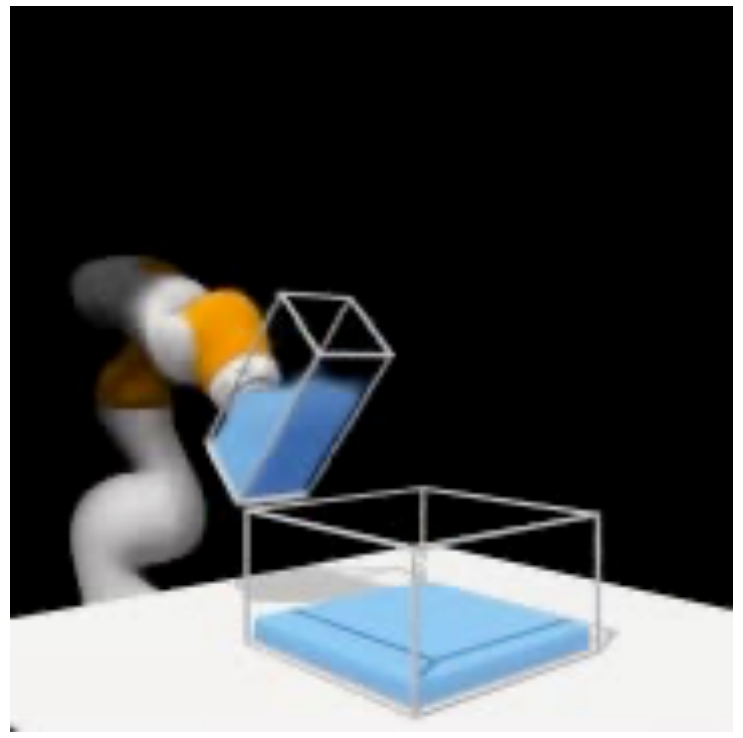
Deep Visual Constraints: Neural Implicit Models for Manipulation Planning from Visual Input
Ha et al. arXiv 2021



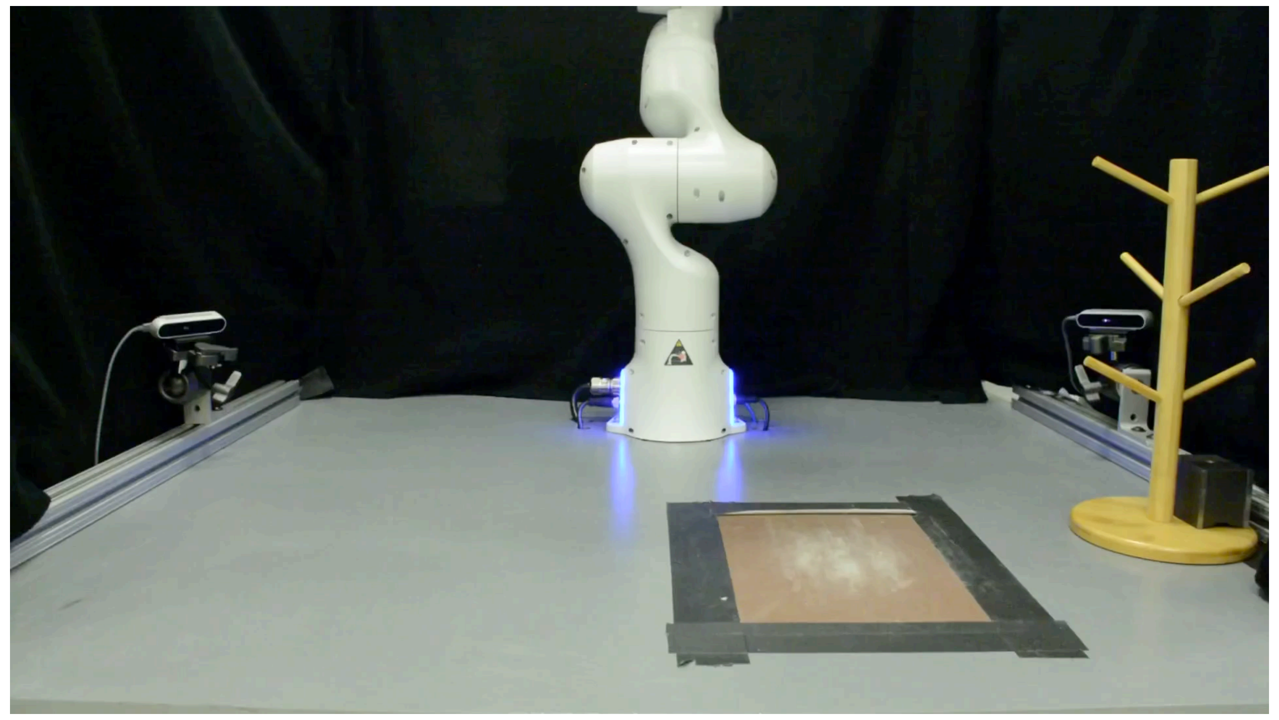
(a) No object model (b) See (c) Plan (d) Act

Robotics

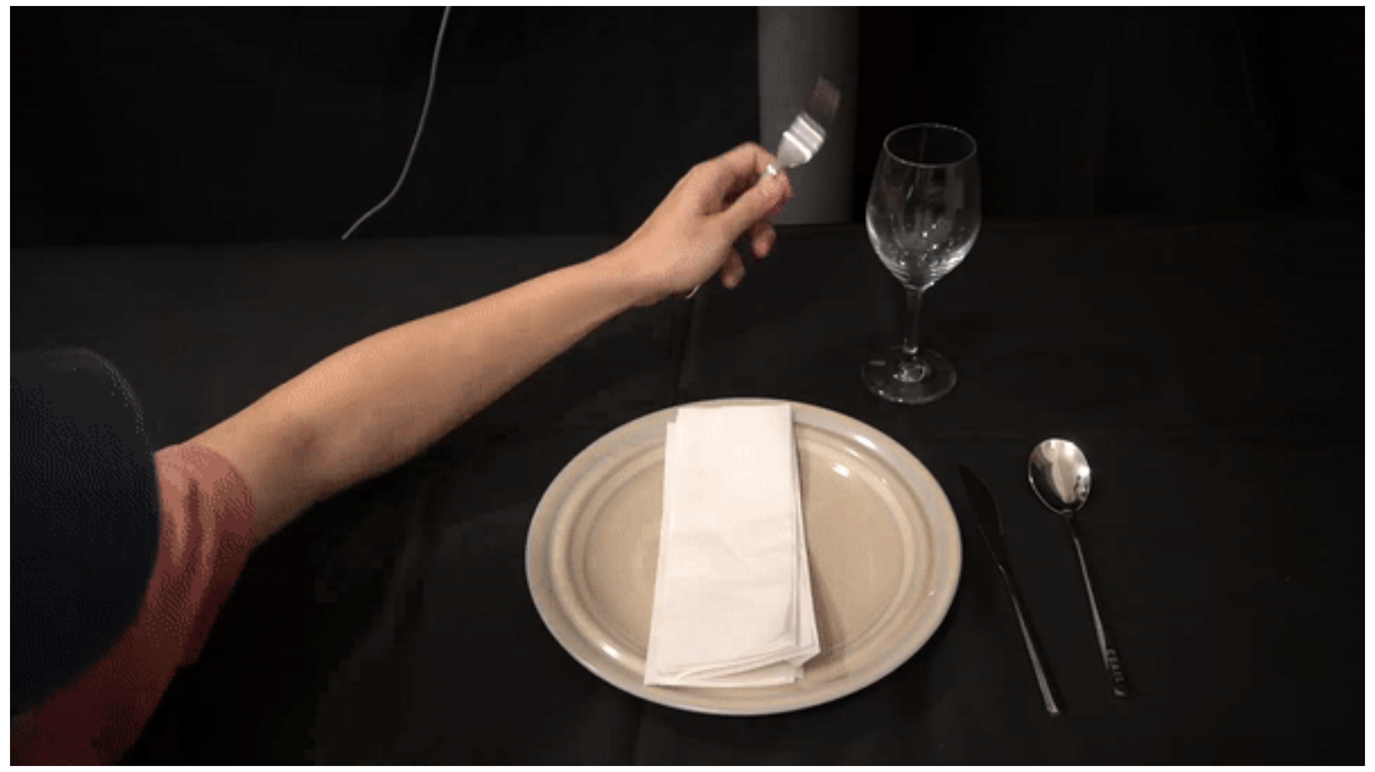
3d neural scene representations for visuomotor control
Li & Li et al. CORL 2022



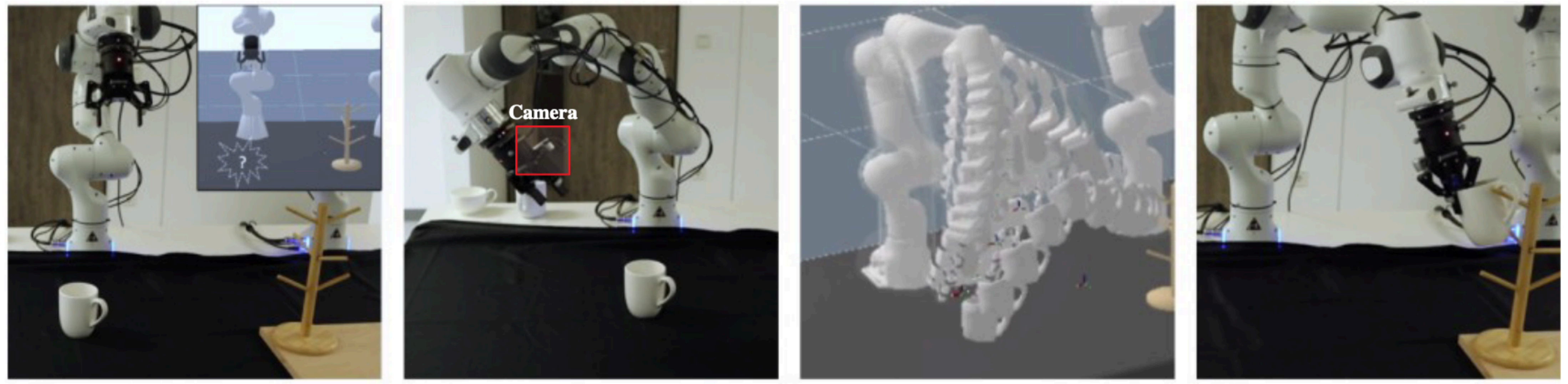
Neural descriptor fields: Se (3)-equivariant object representations for manipulation
Simeonov & Du et al., ICRA 2022



NeRF-Supervision: Learning Dense Object Descriptors from Neural Radiance Fields
Lin et al. ICRA 2022

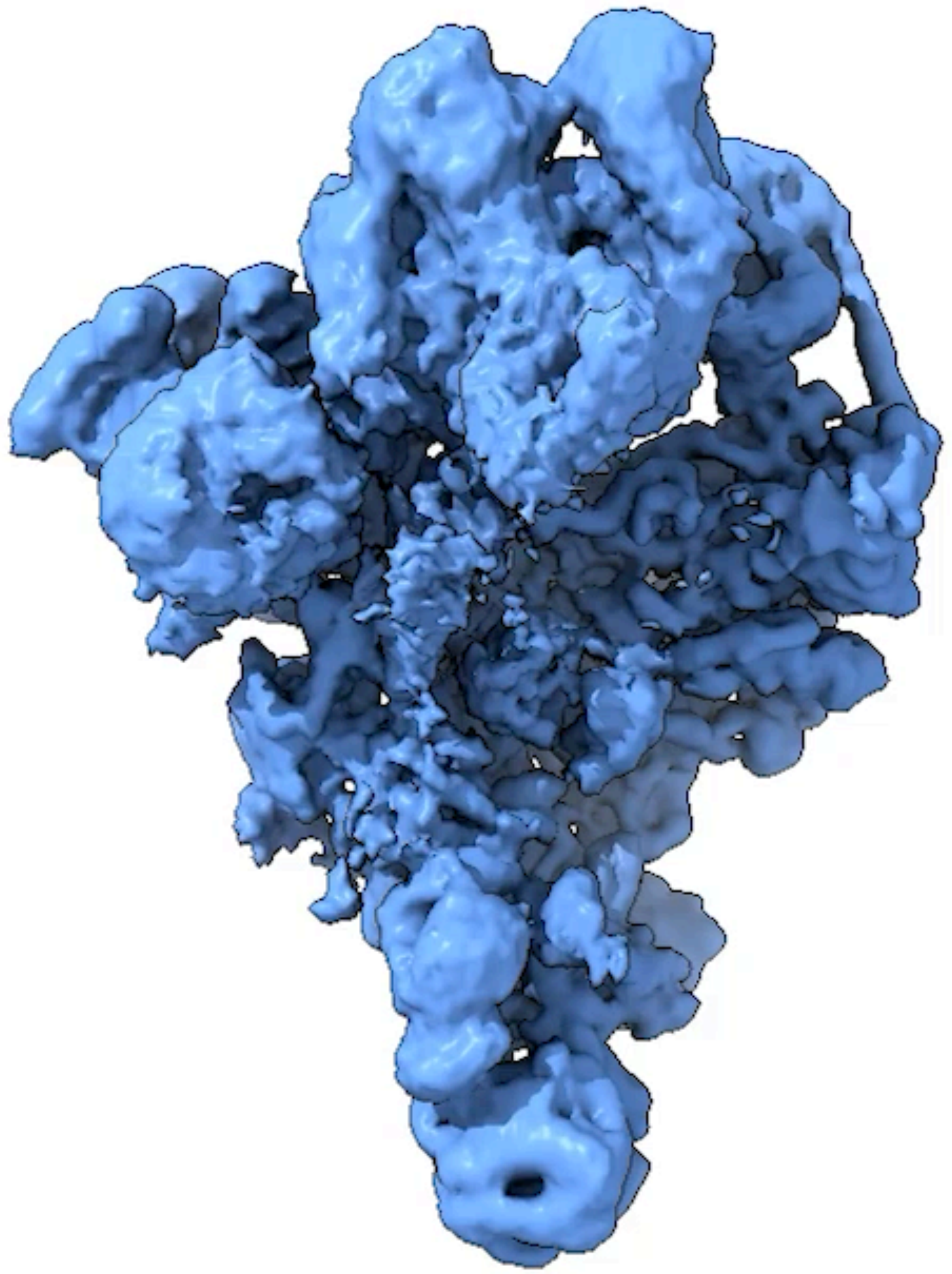


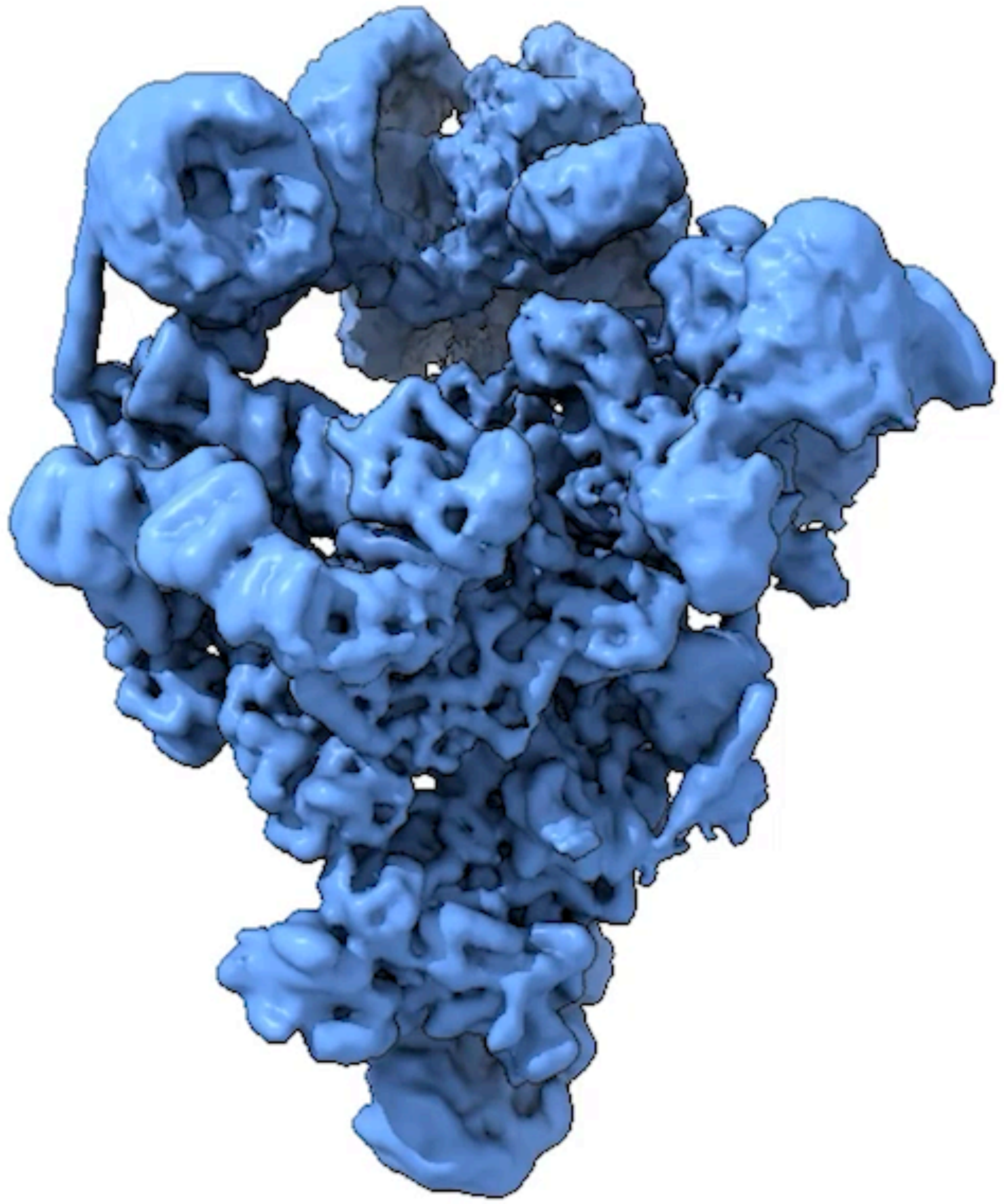
Deep Visual Constraints: Neural Implicit Models for Manipulation Planning from Visual Input
Ha et al. arXiv 2021

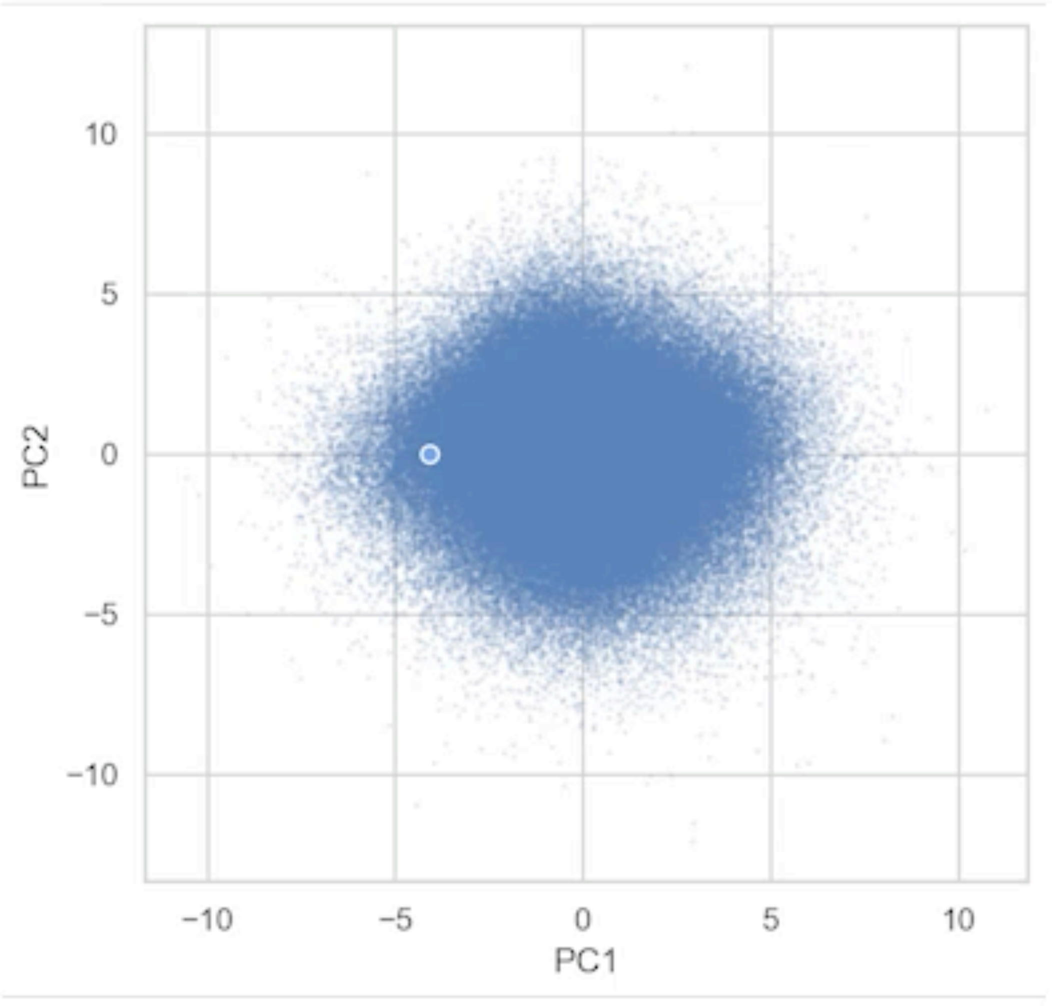


(a) No object model (b) See (c) Plan (d) Act

Scientific Discovery

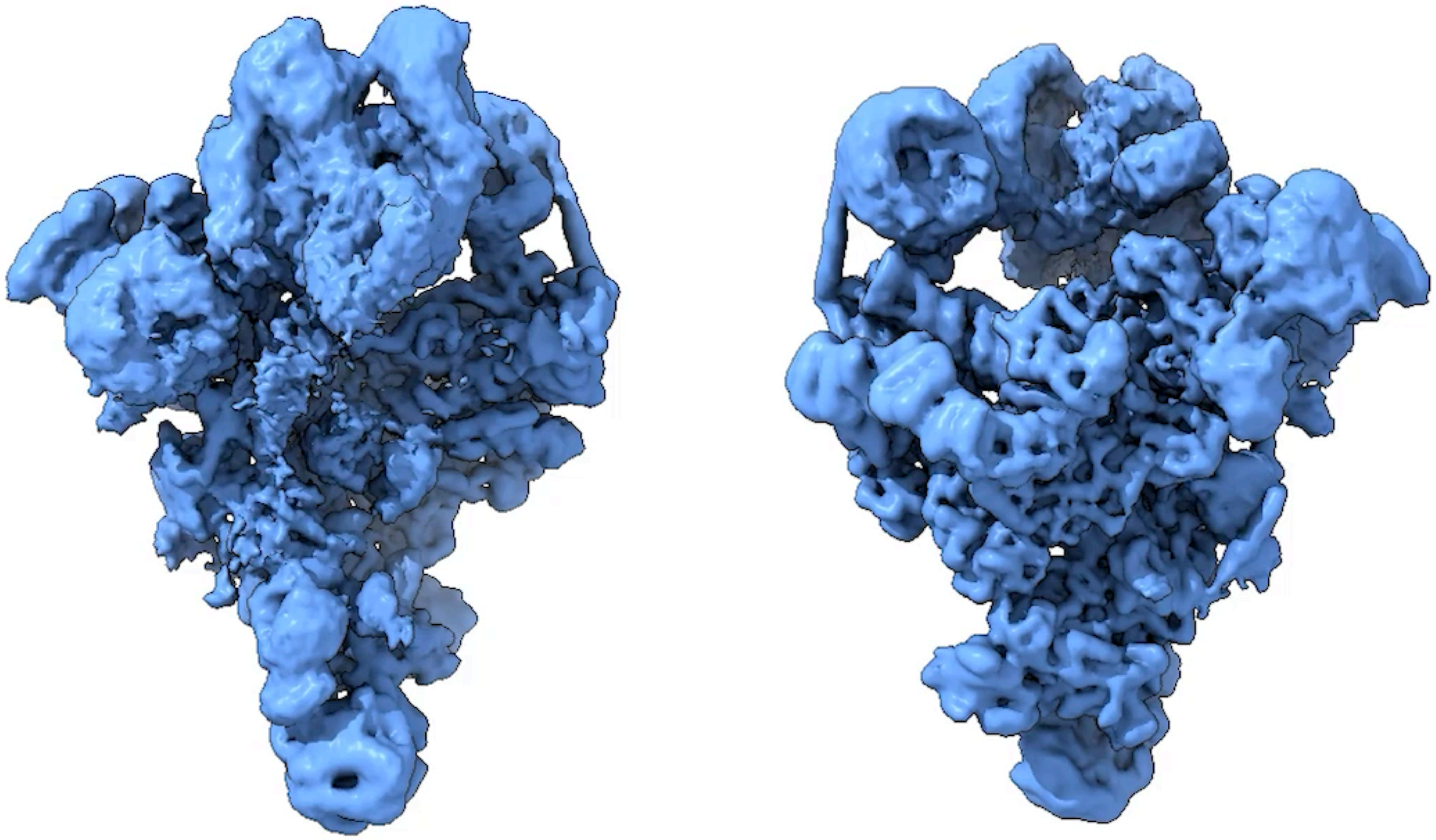


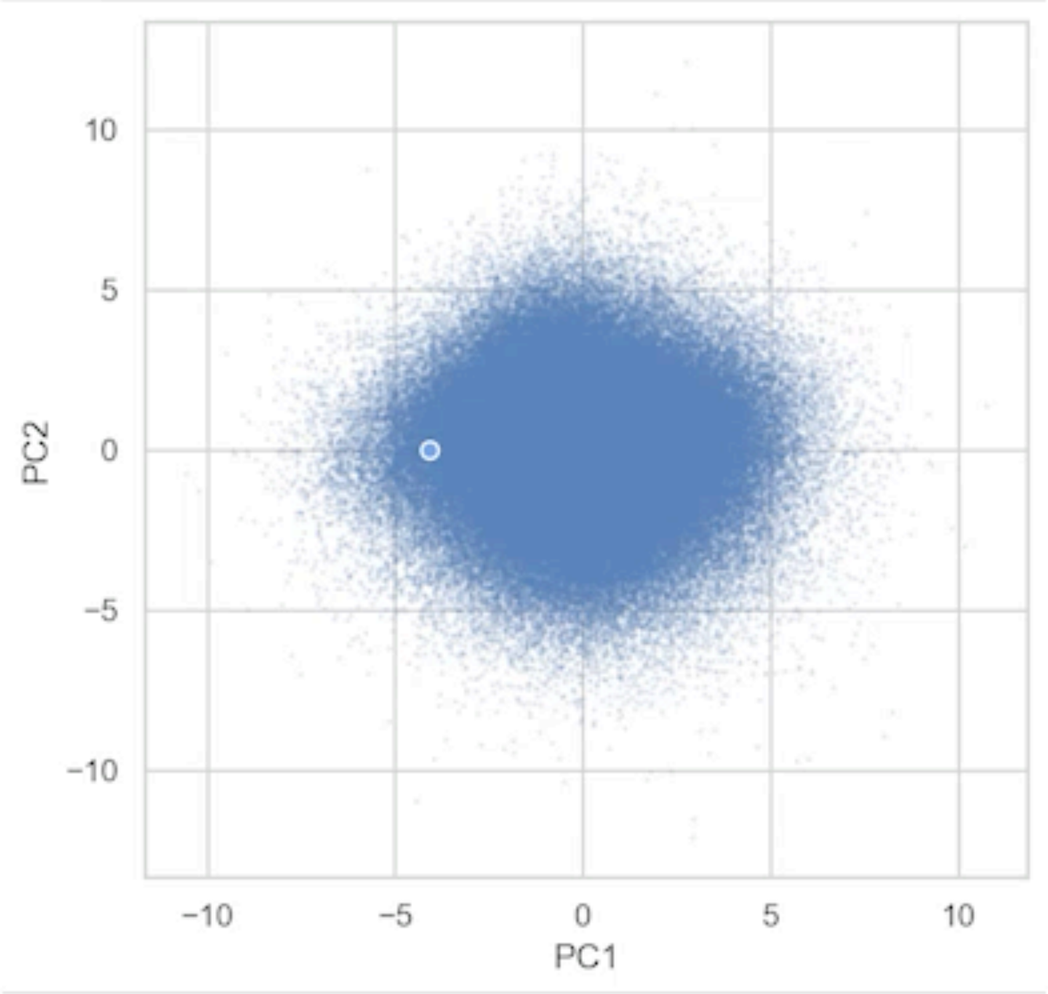




CryoDRGN, Zhong et al. 2021

Scientific Discovery





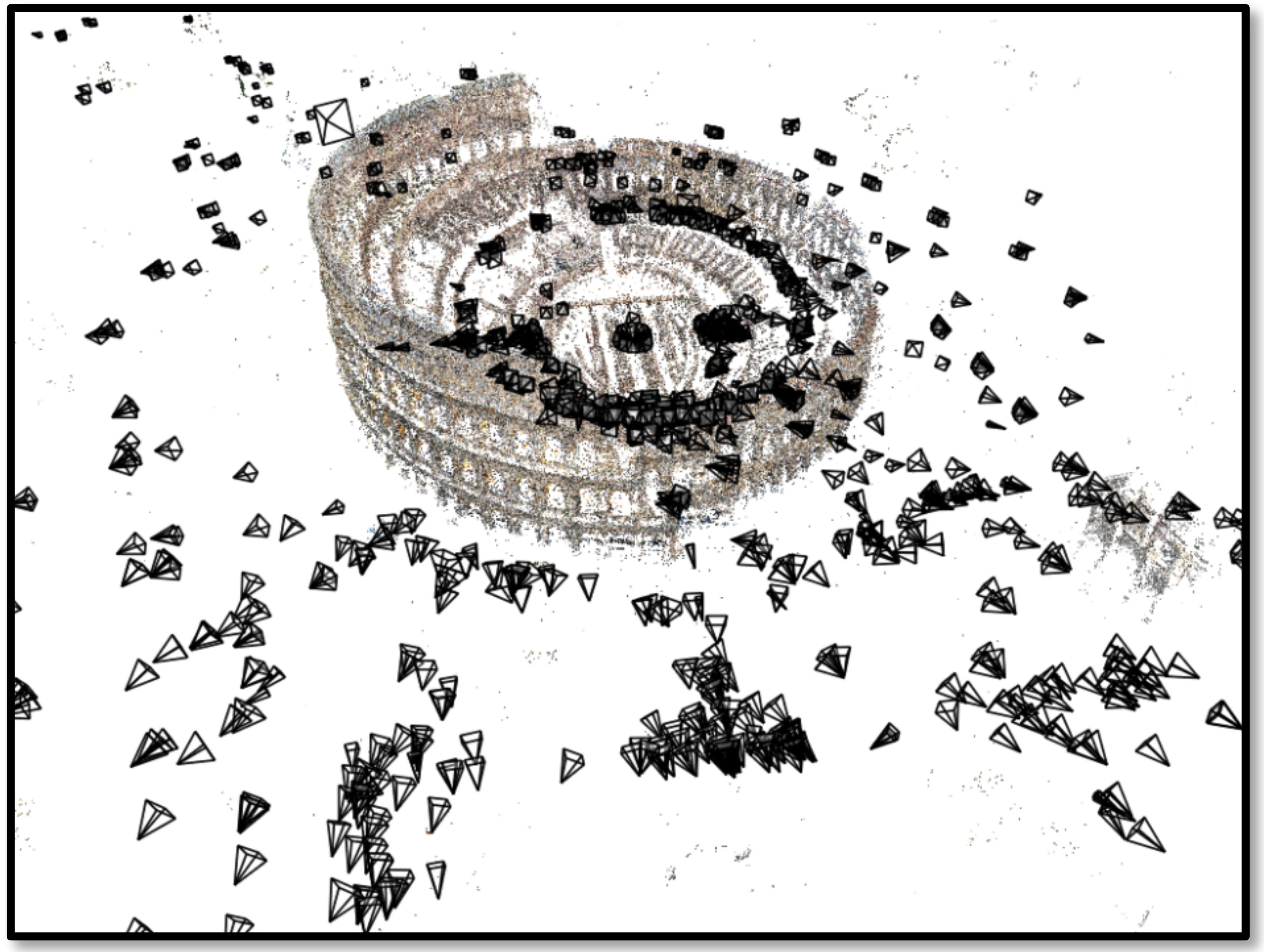
CryoDRGN, Zhong et al. 2021

Tentative Schedule

Tentative Schedule

- **Module 1: Image Formation and Multi-View Geometry (Lectures 2 & 3, Assignment 1)**
  - Rigid-body-transforms, projective geometry, camera transforms, camera models.

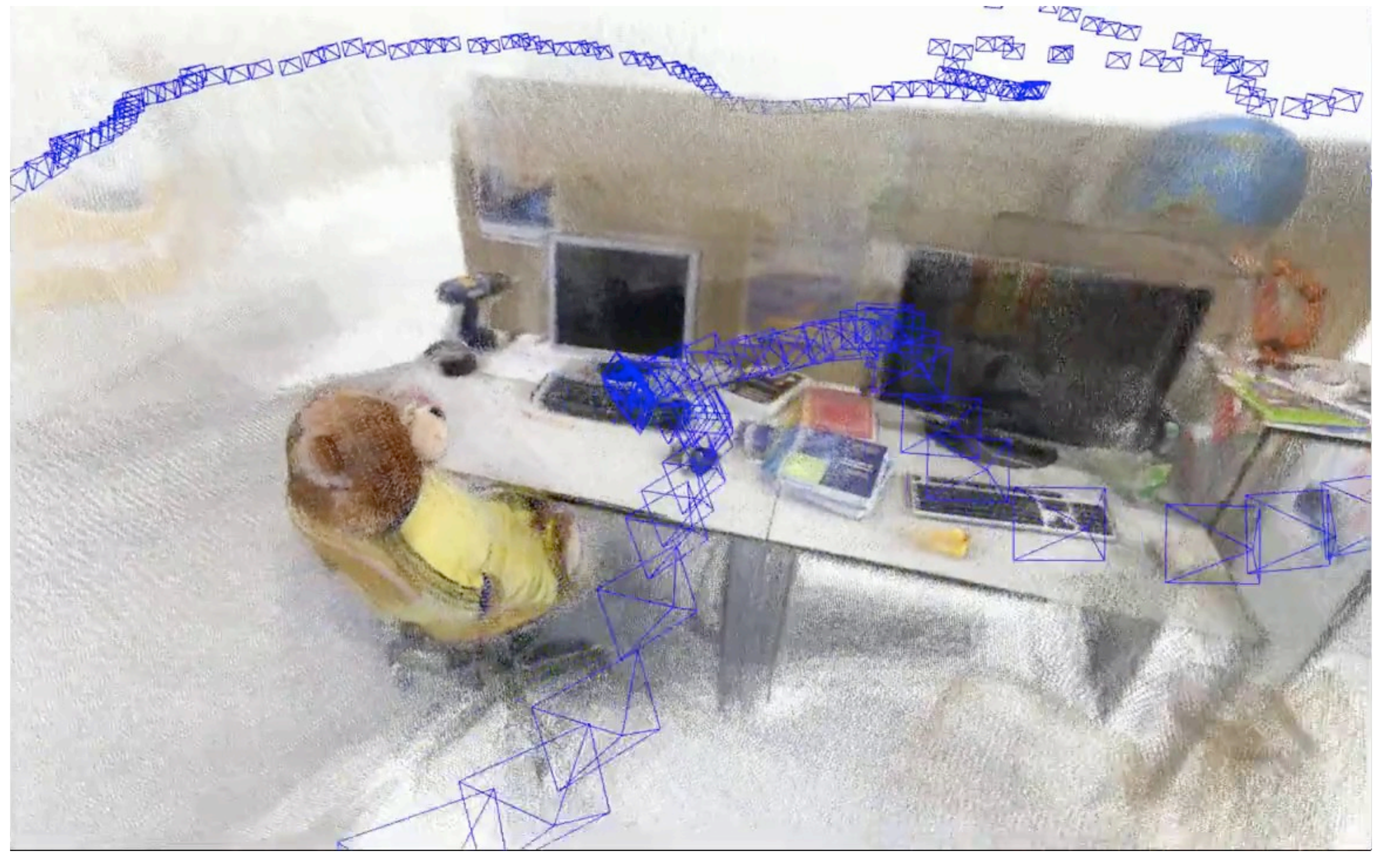
Module 1: Image Formation and Multi-View Geometry



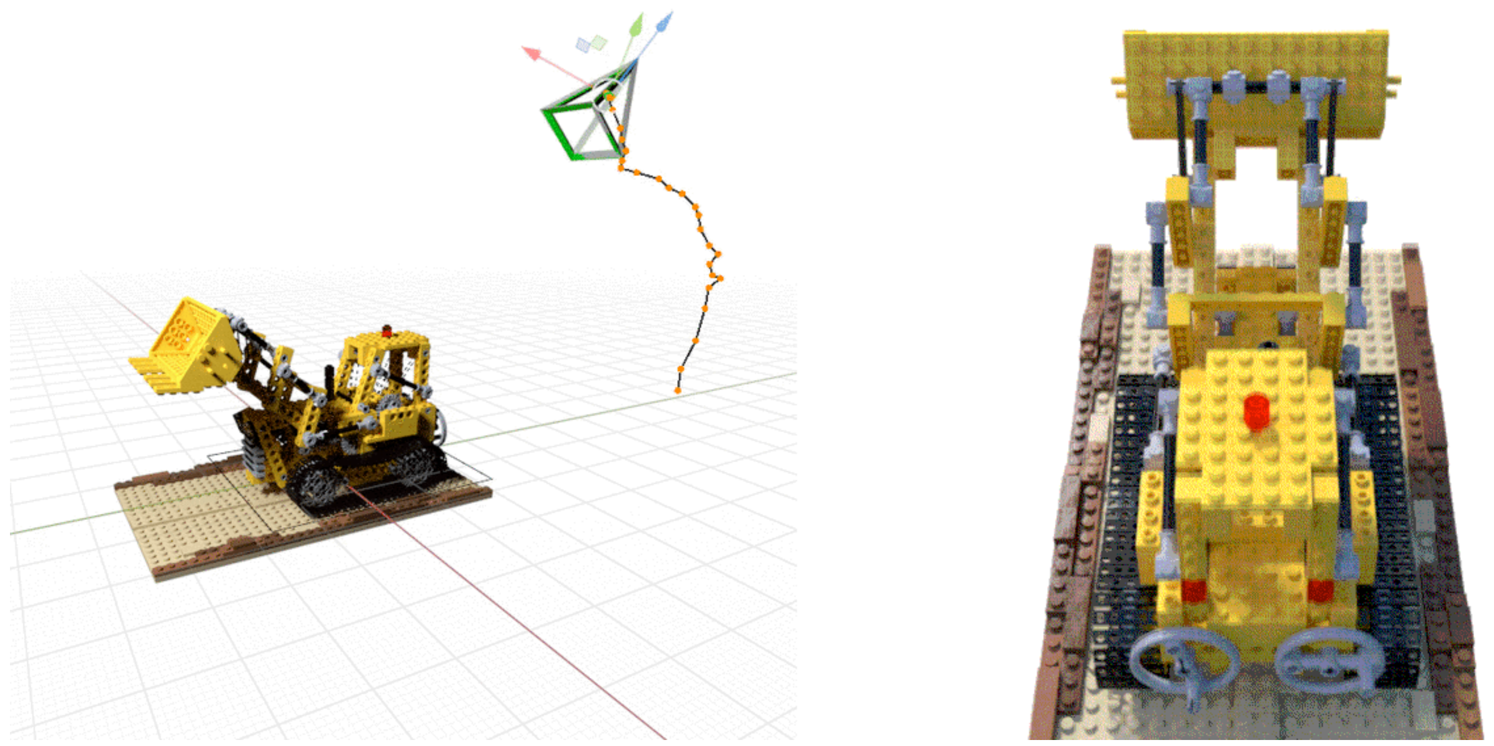
Why?	We want to understand 3D world only from 2D observations (images). For that, we need to have a mathematical understanding of how they are connected.
What you'll learn.	Mathematical model of cameras. Reconstruct camera poses, approximate geometry, and camera parameters from 2D images of a scene.

Image credits: *Noah Snavely*

Existing Approaches - Challenges



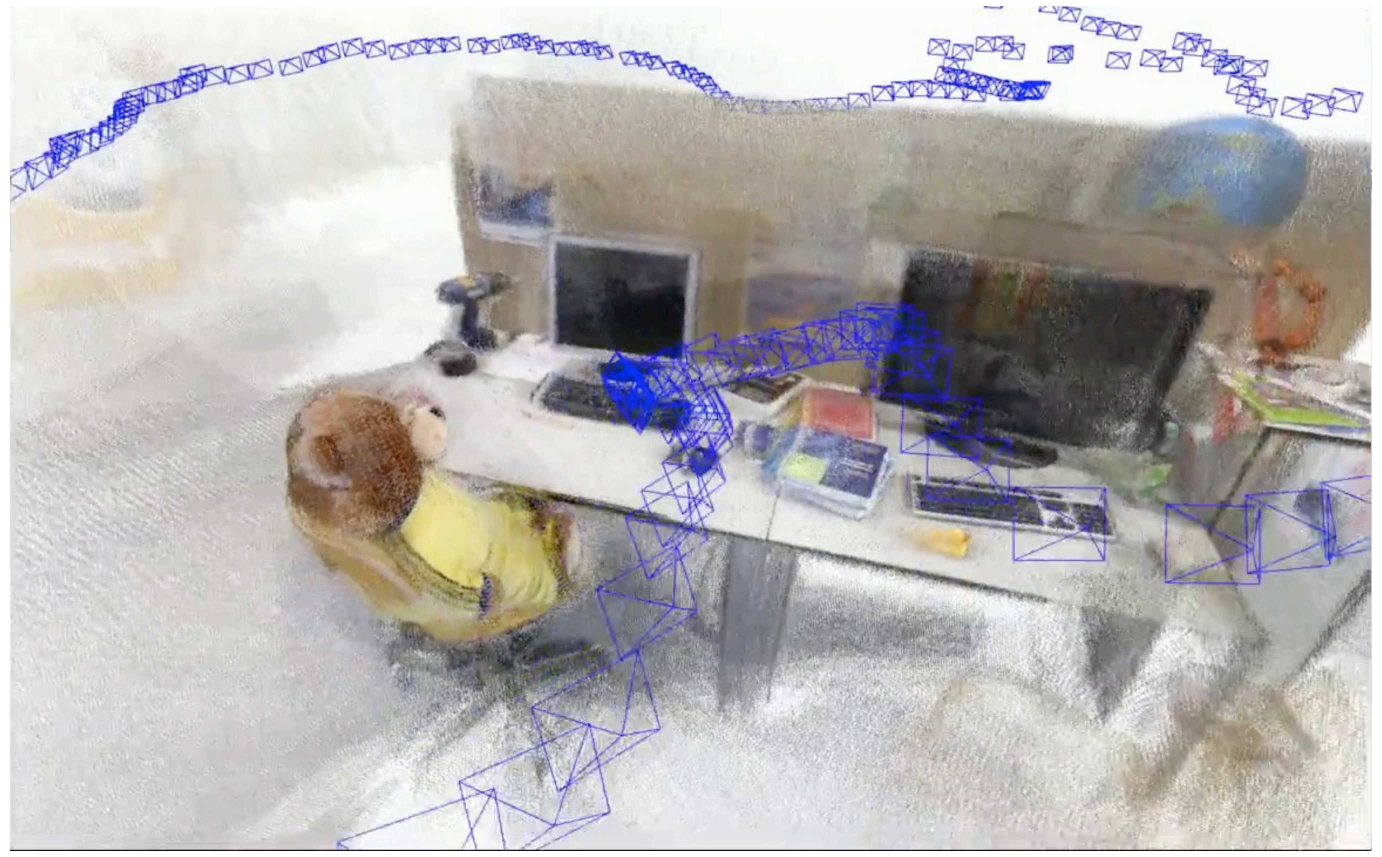
SFM & SLAM



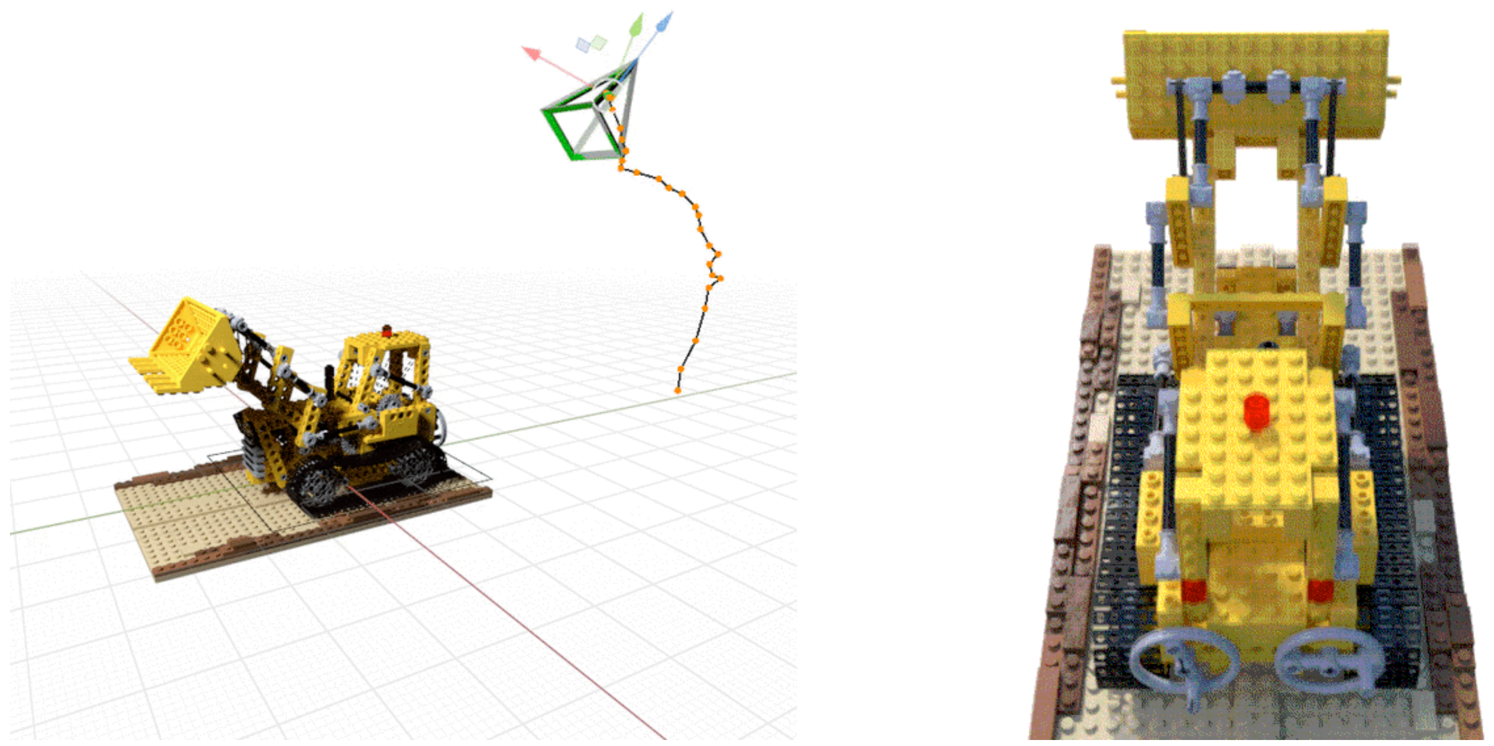
Joint NeRF & pose optim.

Joint depth & pose pred.

Existing Approaches - Challenges



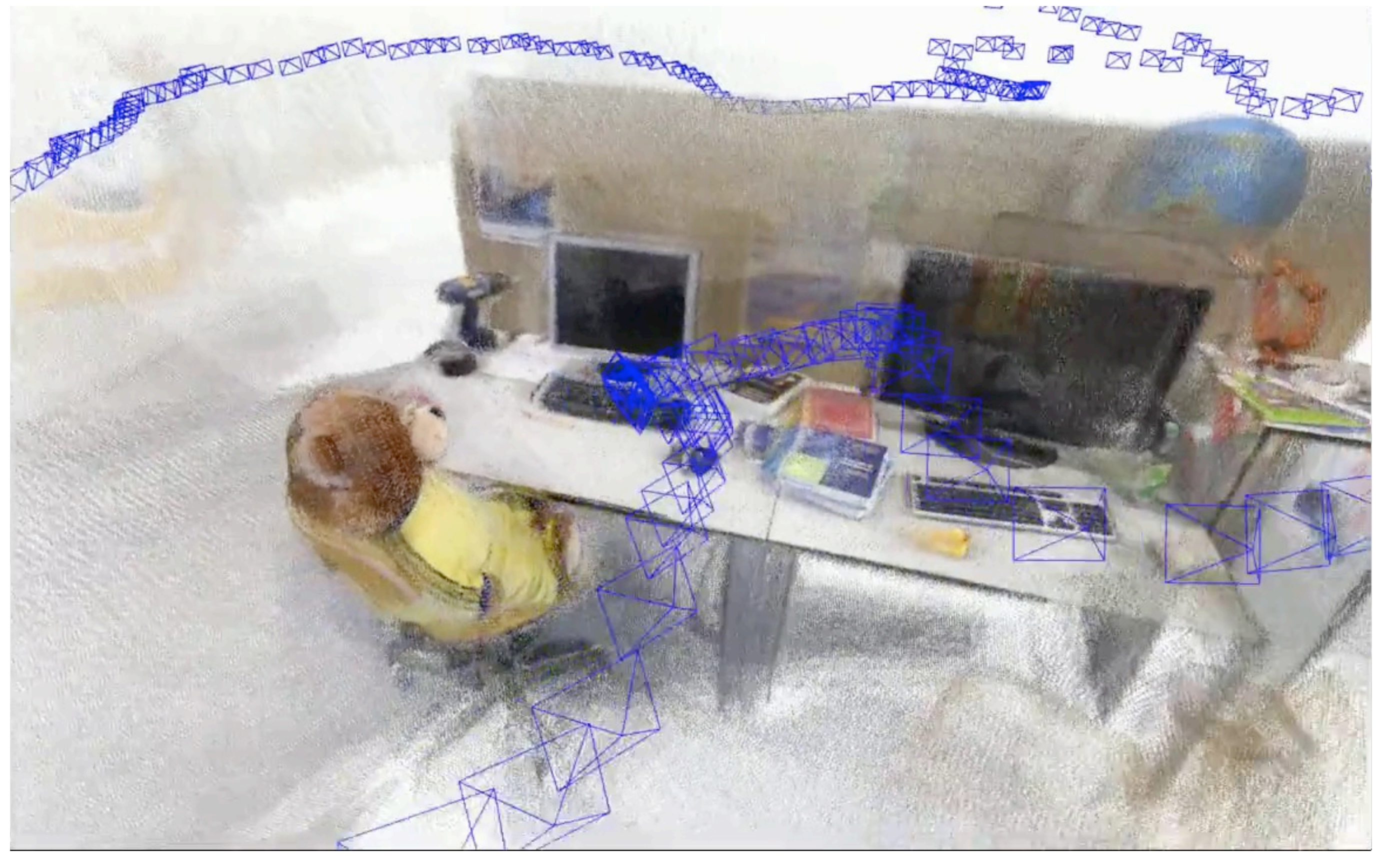
SFM & SLAM

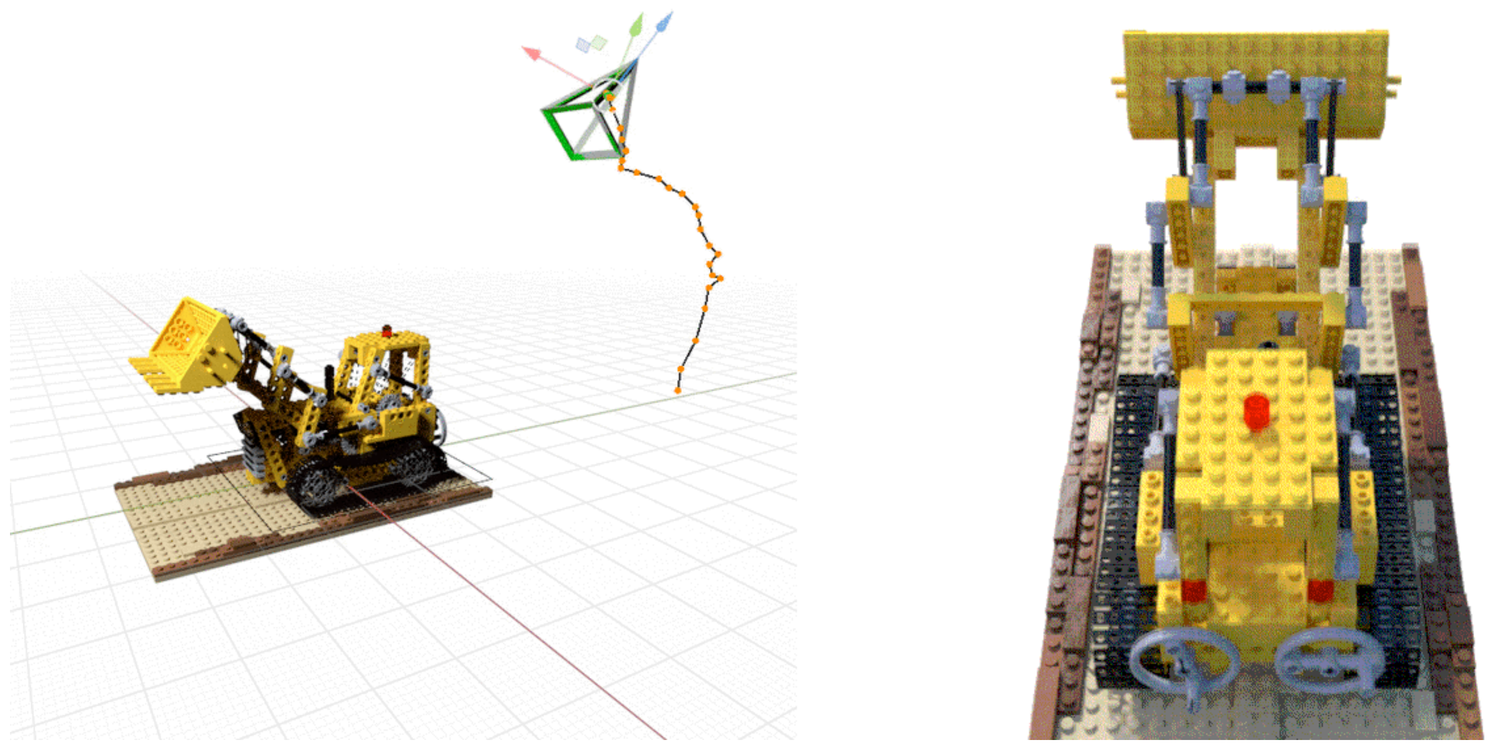


Joint NeRF & pose optim.

Joint depth & pose pred.

Existing Approaches - Challenges





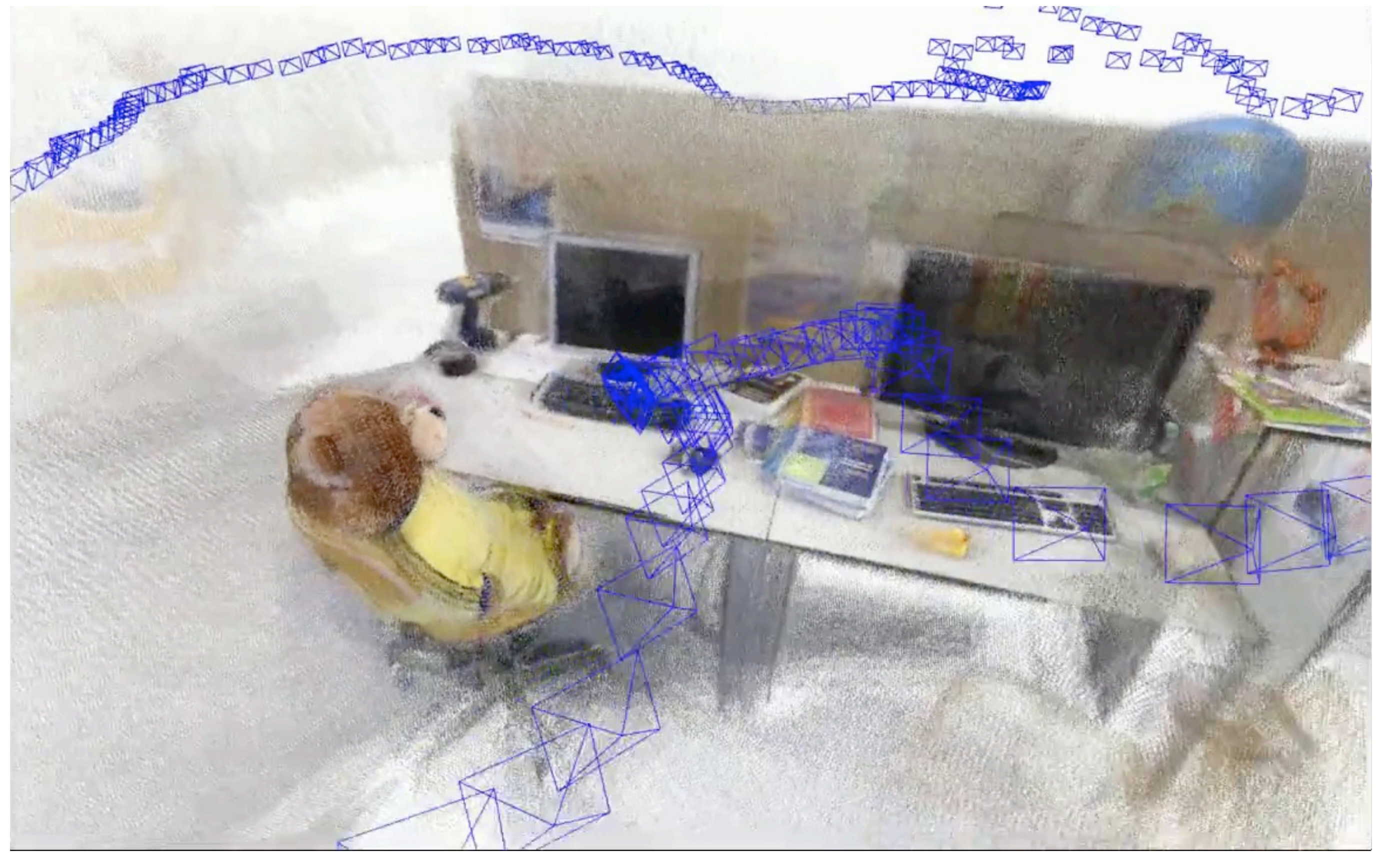
SFM & SLAM

Joint NeRF & pose optim.

Joint depth & pose pred.

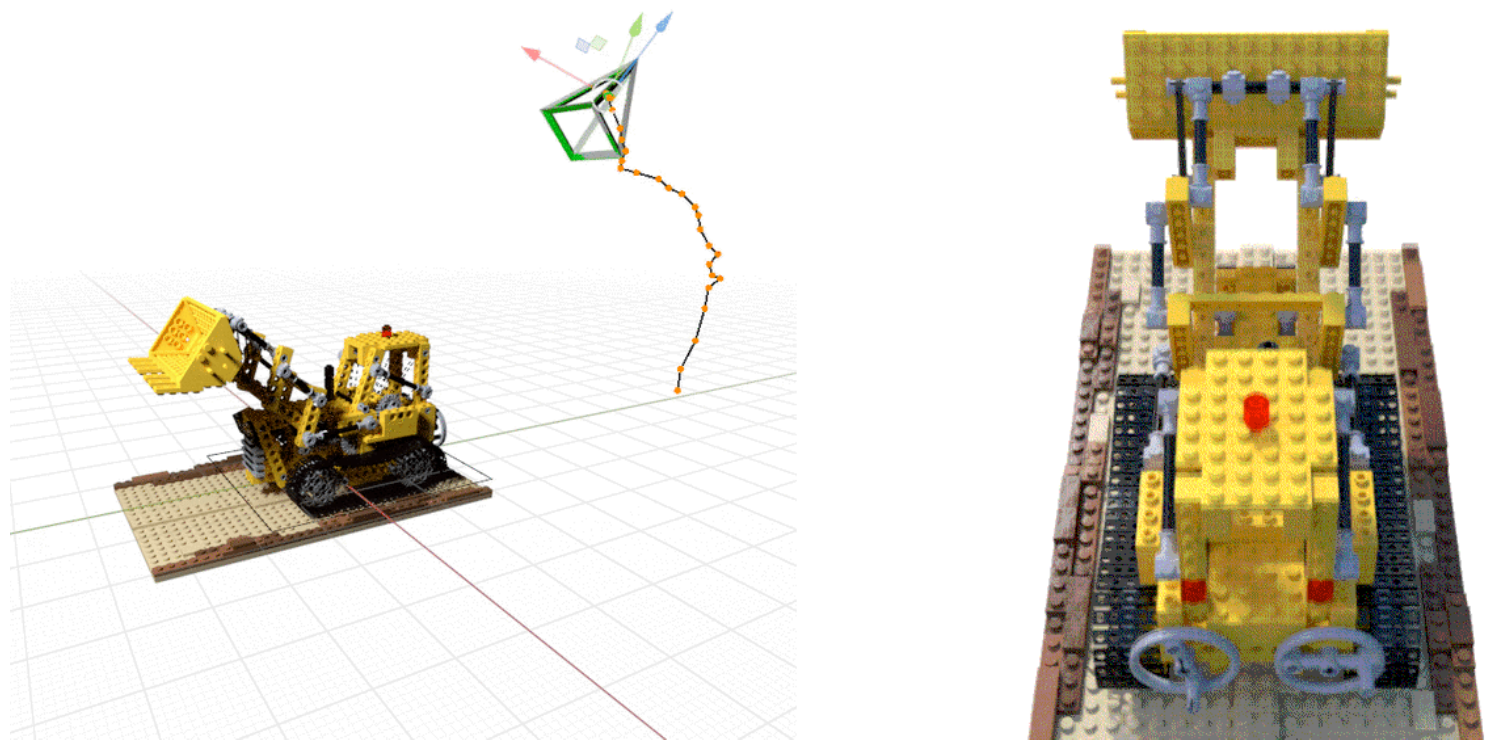
- Per-scene optimization
- No learned 3D scene priors
- SLAM: Prone to failure for particular trajectories

Existing Approaches - Challenges



SFM & SLAM

- Per-scene optimization
- No learned 3D scene priors
- SLAM: Prone to failure for particular trajectories

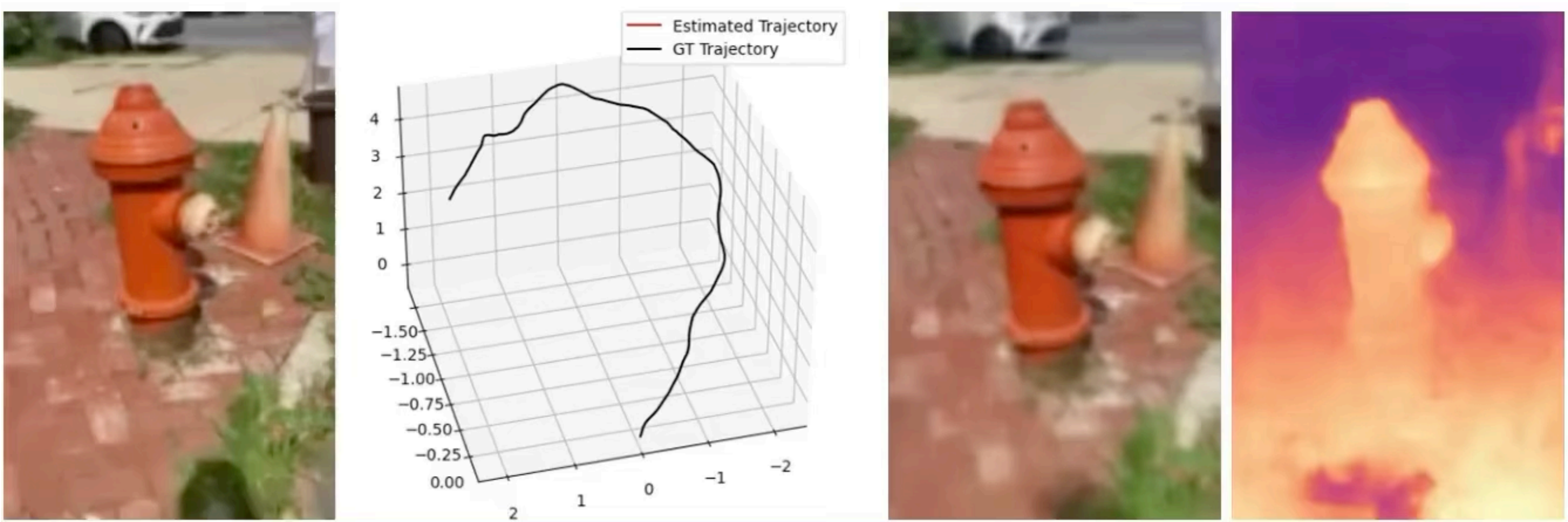


Joint NeRF & pose optim.

Joint depth & pose pred.

- Online!
- Limited to simple trajectories
- Depth only, no 3D scene

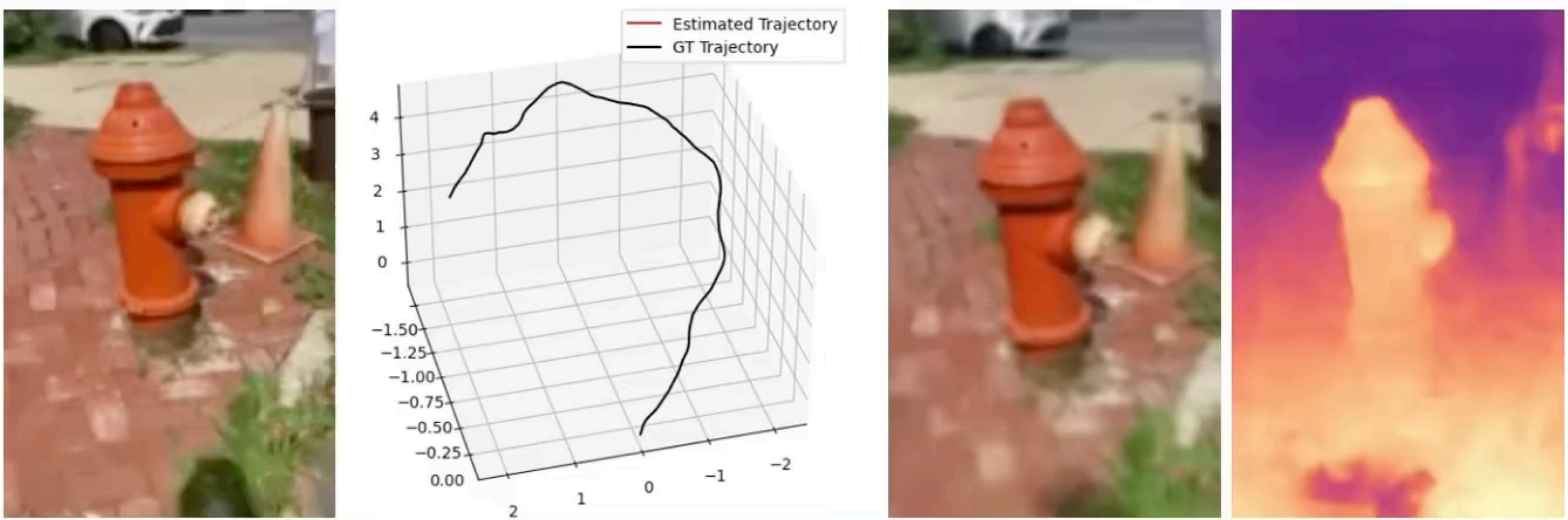
FlowCam:
Generalizable Neural Scene Representations without Camera Poses



Renderings from Generalizable Neural Scene Representation.
Trained end-to-end, without camera poses, only on raw video!

FlowCam, Smith et al., arXiv 2023

FlowCam:
Generalizable Neural Scene Representations without Camera Poses



Renderings from Generalizable Neural Scene Representation.
Trained end-to-end, without camera poses, only on raw video!

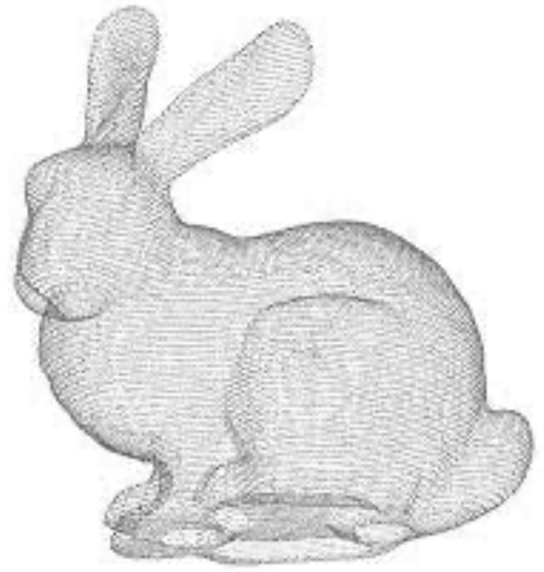
FlowCam, Smith et al., arXiv 2023

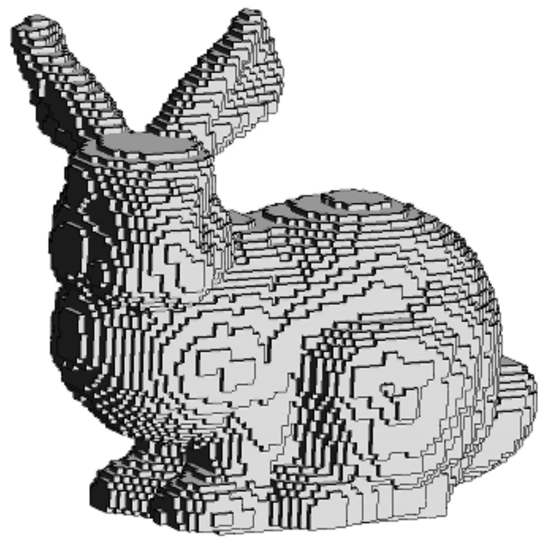
Tentative Schedule

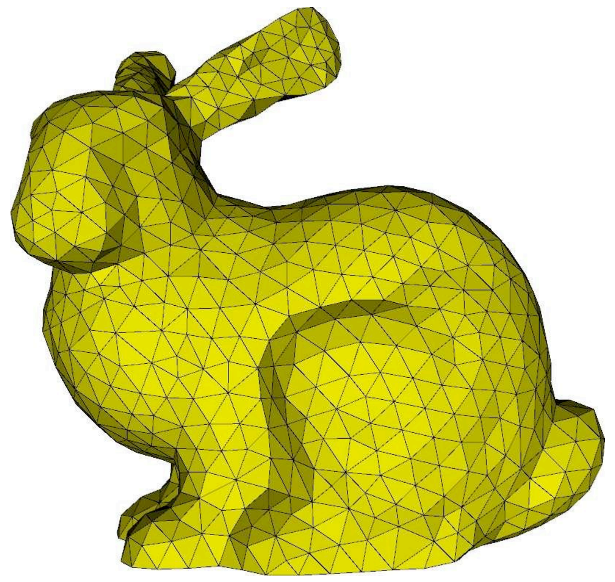
- **Module 1: Image Formation and Multi-View Geometry (Lectures 2 & 3, Assignment 1)**
  - Rigid-body-transforms, projective geometry, camera transforms, camera models.

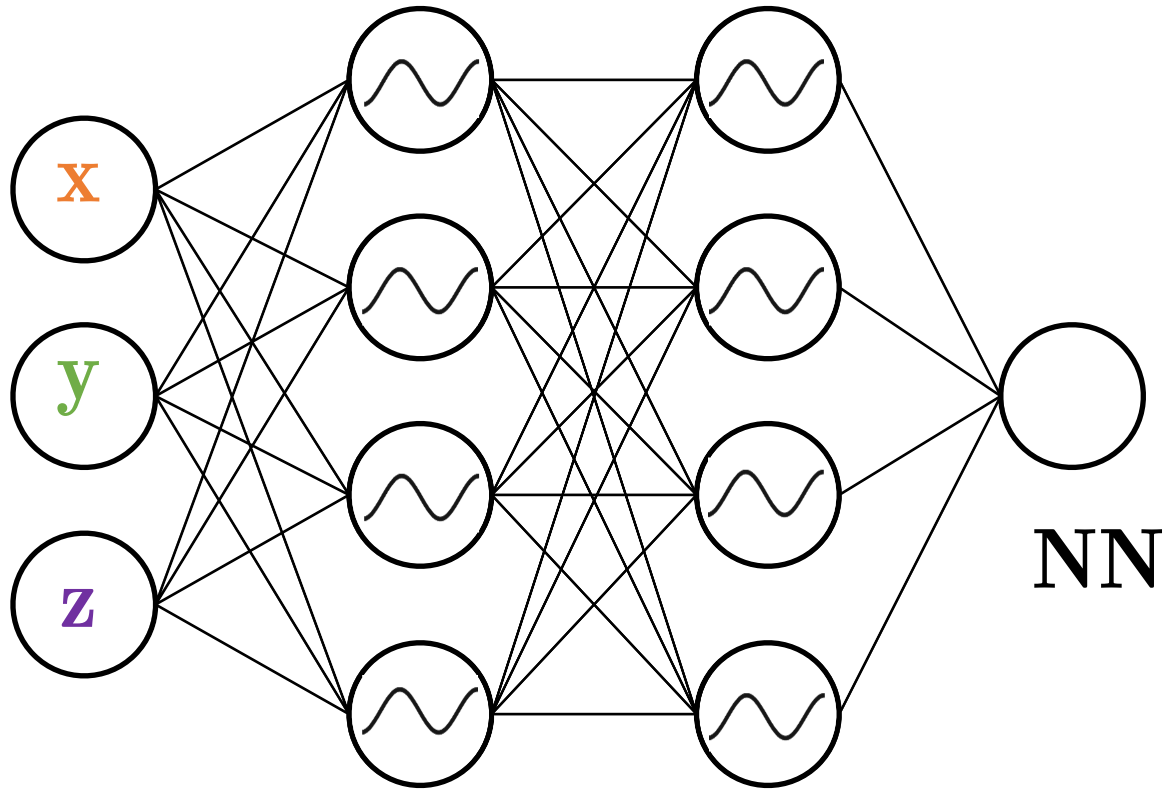
- **Module 2: 3D Representations, Rendering, and Differentiable Rendering (Lectures 4-7, Assignment 2)**
  - Surface vs. volume representations, continuous vs. discrete representations
  - Conventional rendering in computer graphics
  - Differentiable rendering & neural rendering

3D Representations

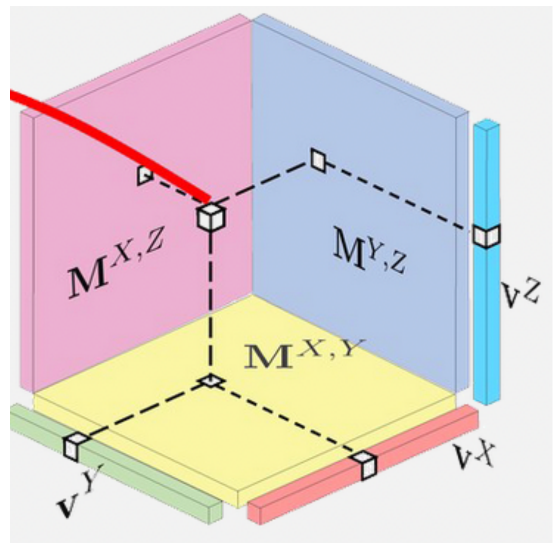




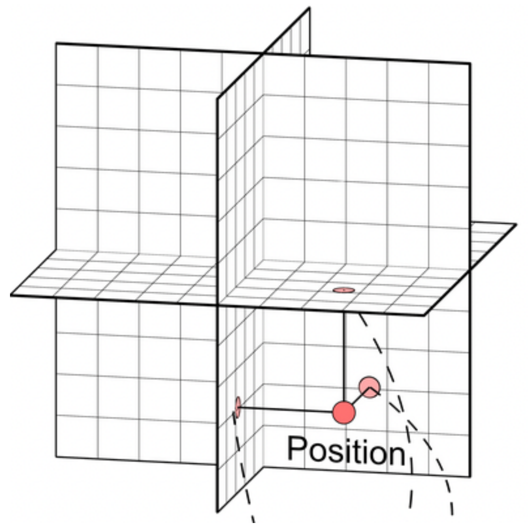




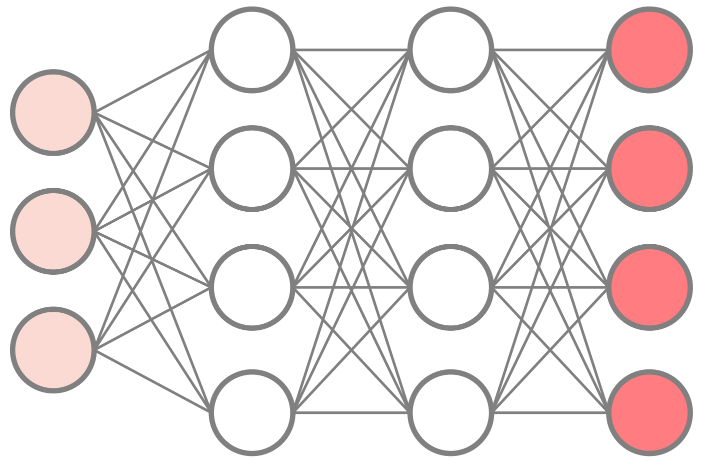
$\mathrm{NN}(x, y, z)$



TensoRF, Chen & Xu et al. 2022

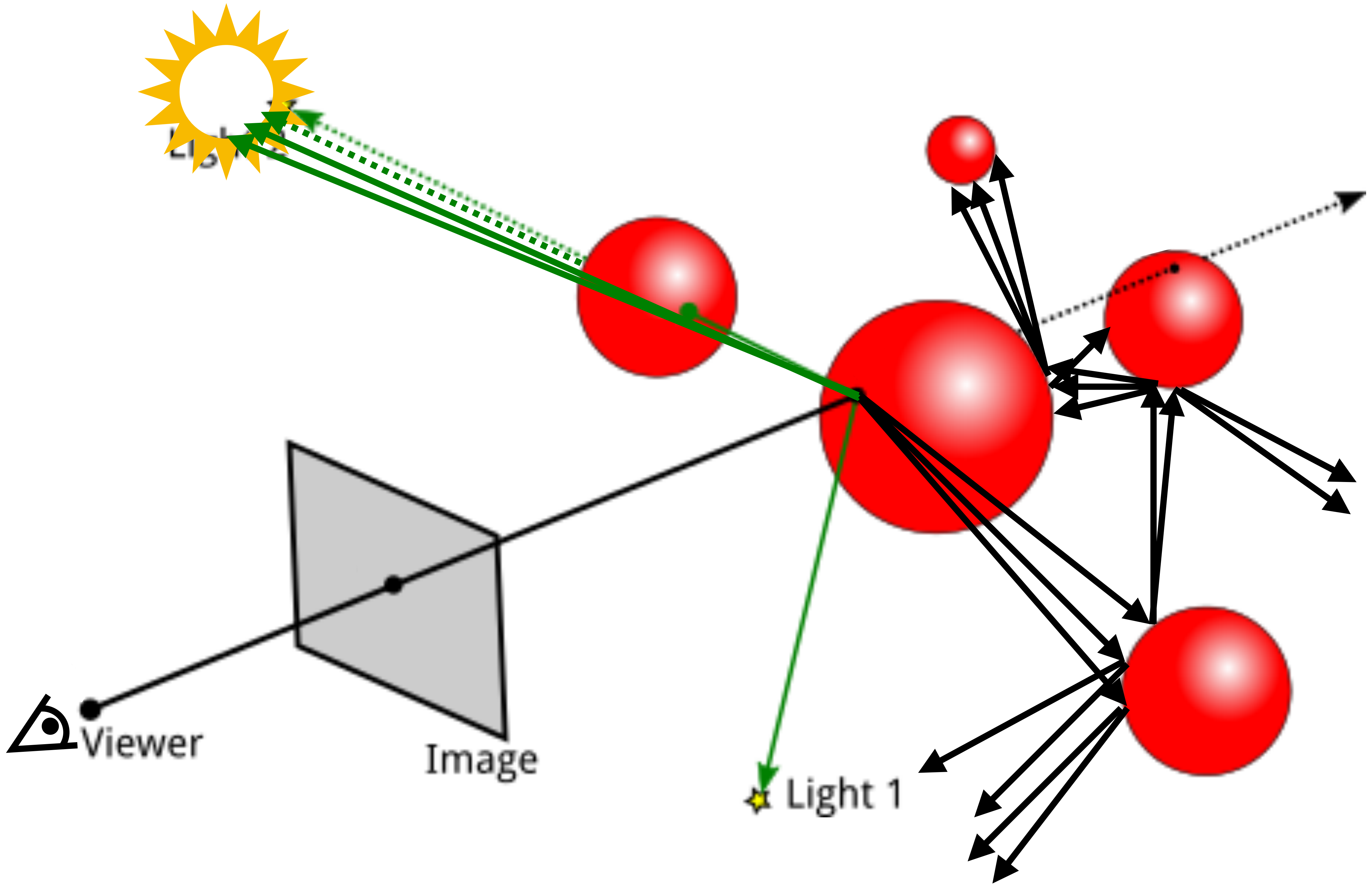


EG3D, Chan et al. 2022



Why?	At the end of the day, we want to make predictions about 3D scenes. For that, we need to know how we can represent 3D scenes computationally.
What you'll learn.	Surface-based representations, volumetric representations, discrete representations, continuous representations.

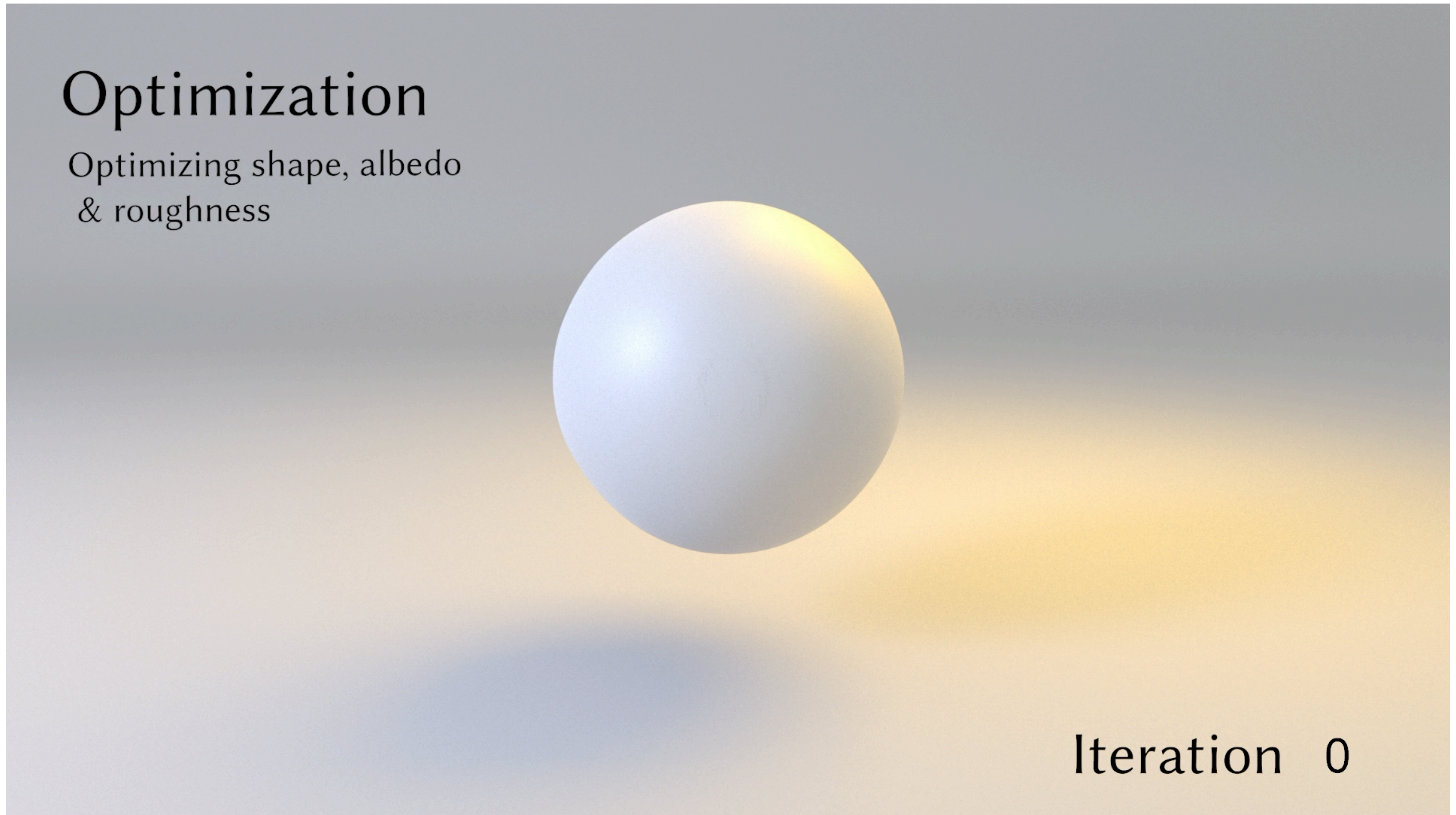
Models of Light Transport



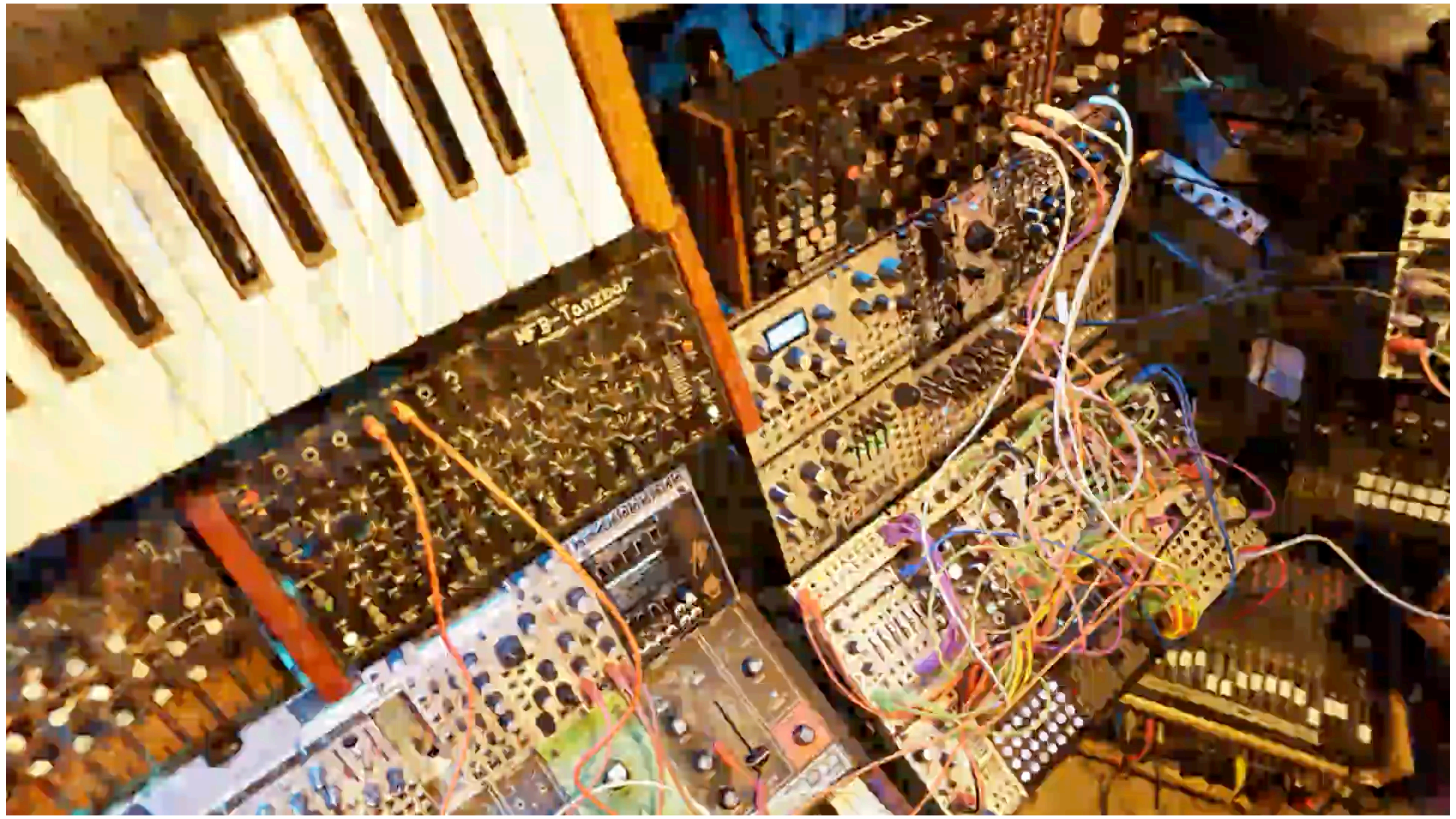
From "Computer Graphics in the Age of AI", C. Karen Liu & Jiajun Wu

Why?	In order to reason about the 3D world from images, we need to understand how 3D properties such as materials, lighting, etc. relate to the measurements observed by a camera.
What you'll learn.	The rendering equation, simulating light transport via multi-bounce ray-tracing, bidirectional radiance distribution functions, rasterization, rendering.

Differentiable Rendering



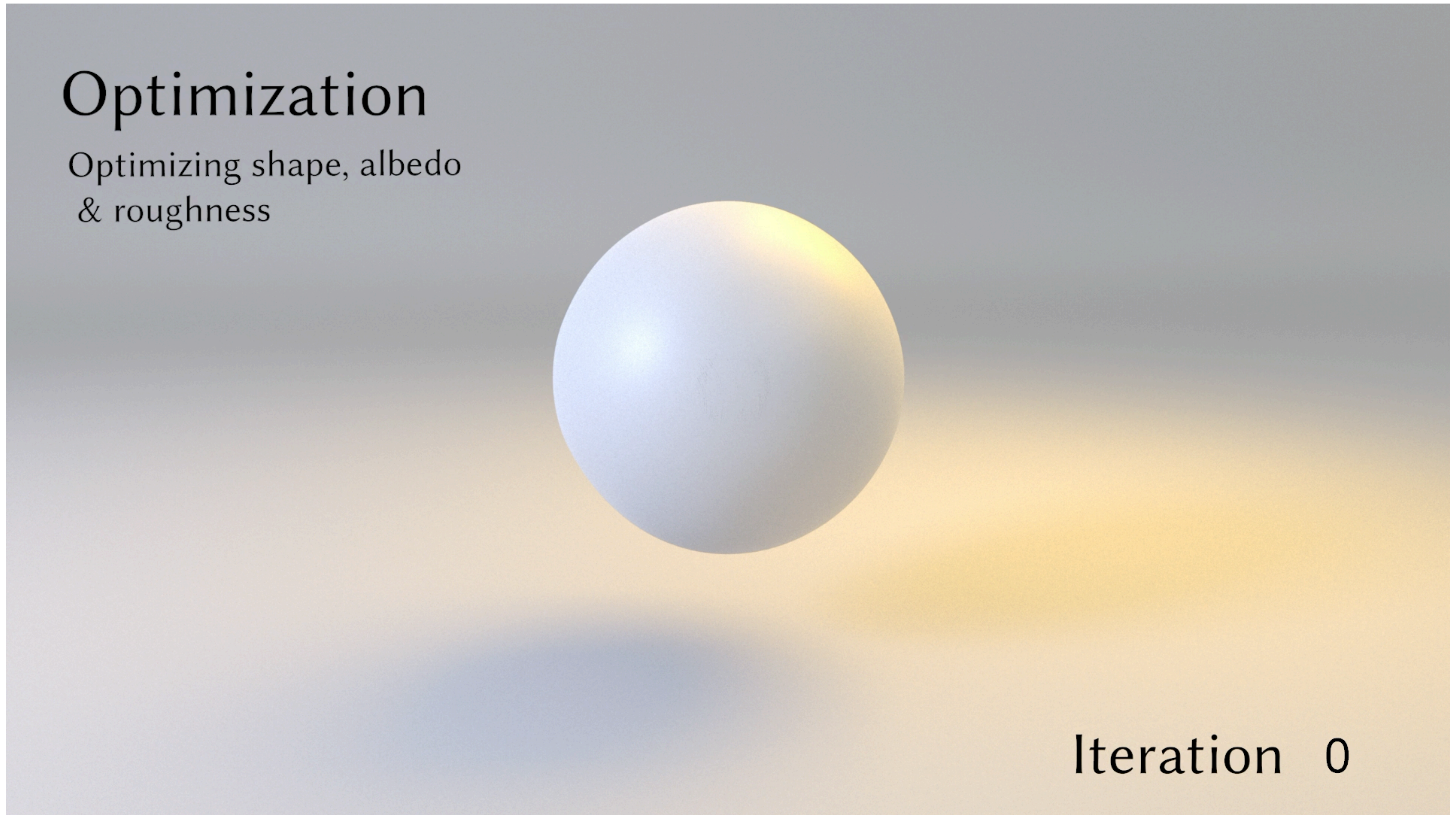
Differentiable Signed Distance Function Rendering,
Vicini et al. 2022



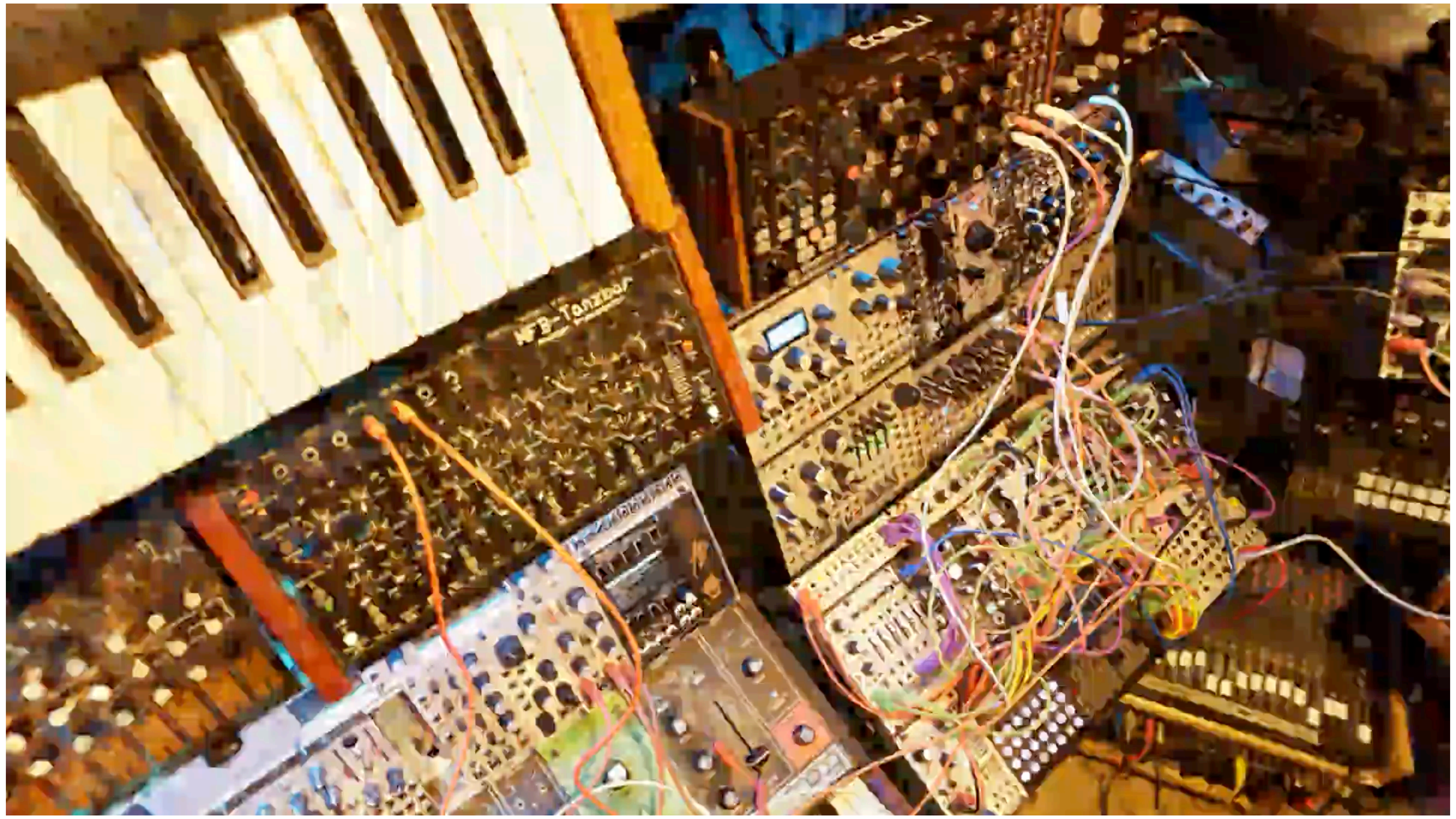
InstantNGP, Müller et al. 2022

Why?	Part of our problem relates to **inverting** the rendering process: Given 2D images, we want to reconstruct 3D scenes. Differentiable rendering is one way of exactly inverting the rendering process.
What you'll learn.	The structure of differentiable renderers, pros and cons and assumptions of different algorithms.

Differentiable Rendering



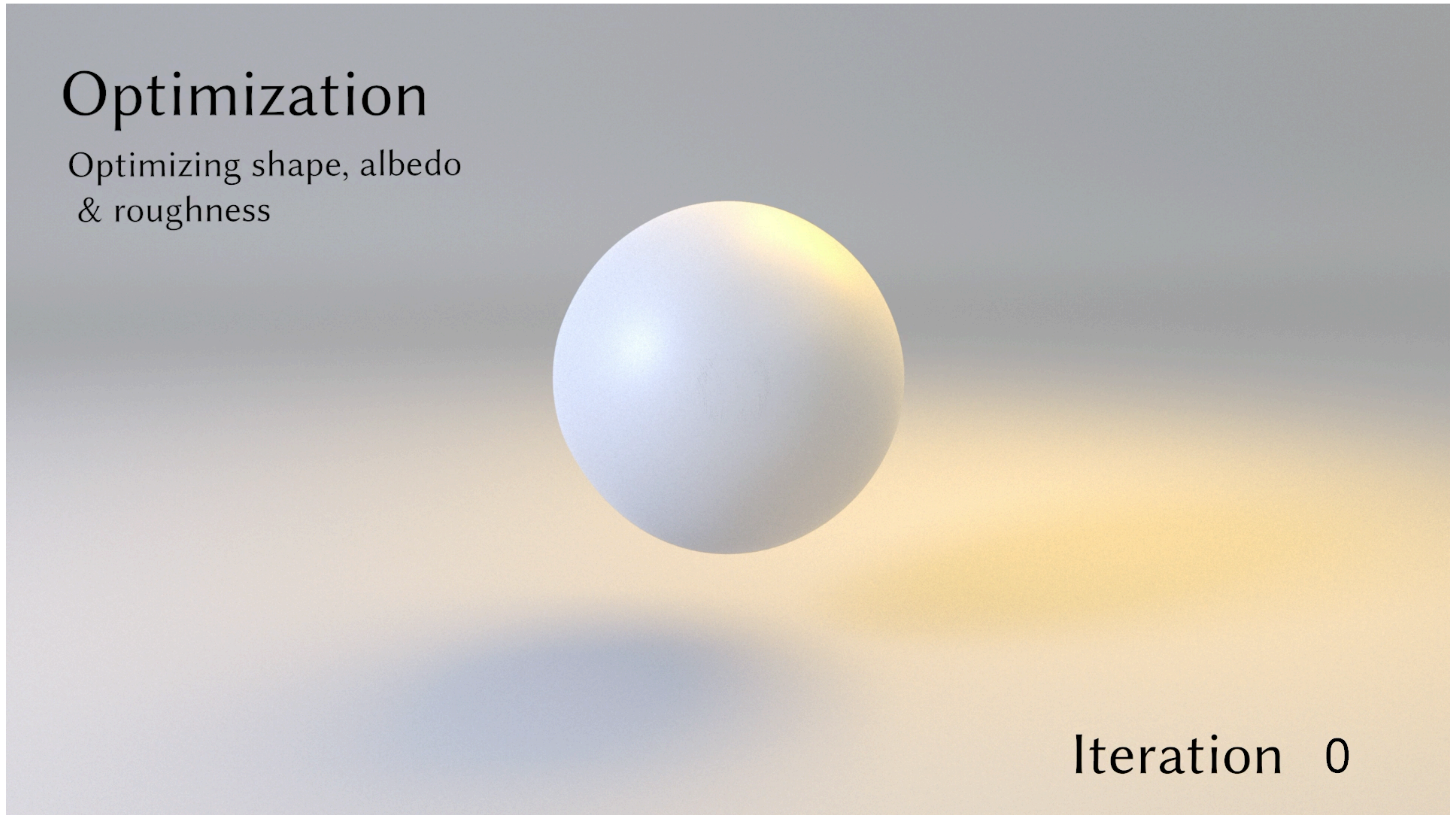
Differentiable Signed Distance Function Rendering,
Vicini et al. 2022



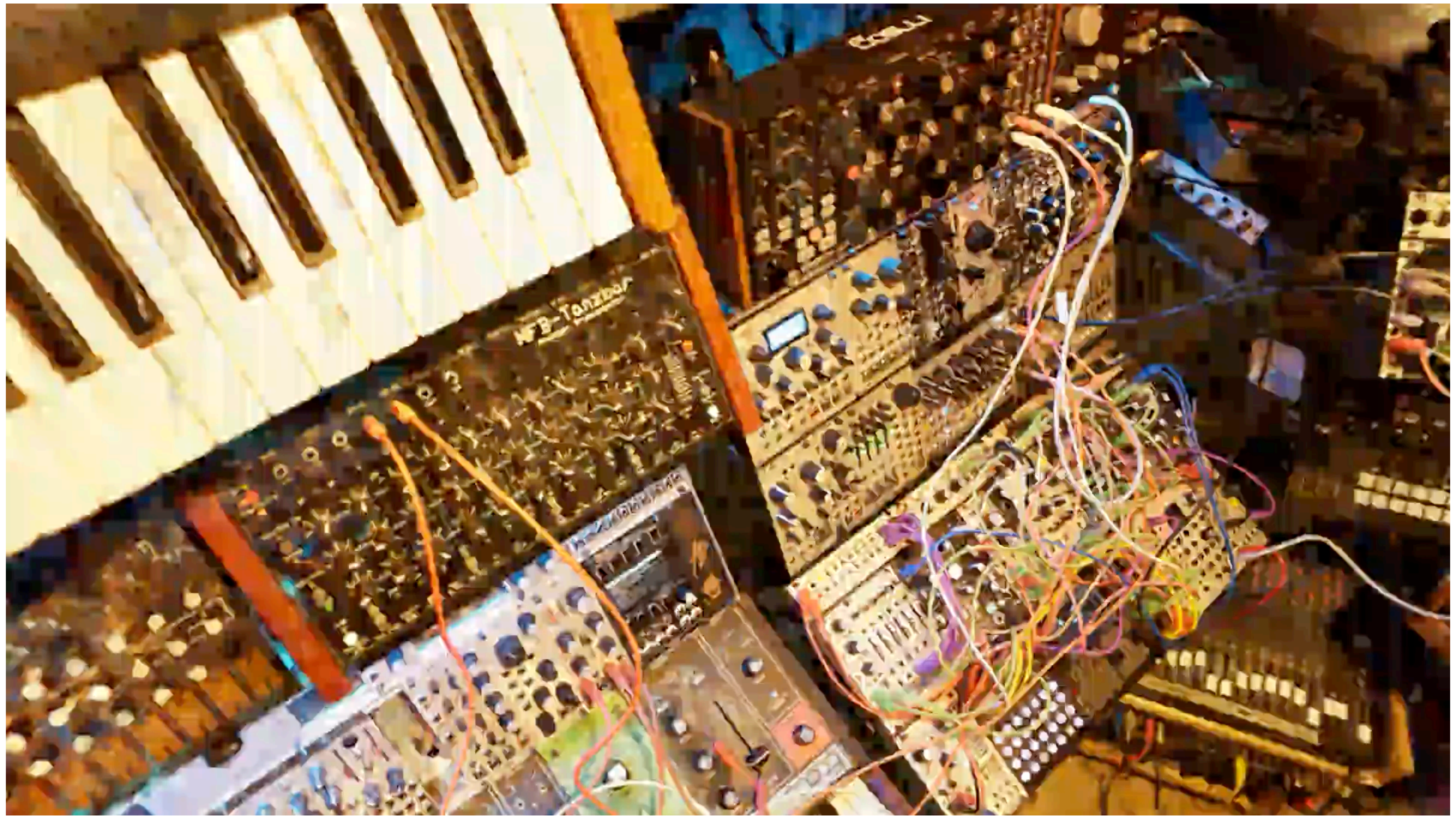
InstantNGP, Müller et al. 2022

Why?	Part of our problem relates to **inverting** the rendering process: Given 2D images, we want to reconstruct 3D scenes. Differentiable rendering is one way of exactly inverting the rendering process.
What you'll learn.	The structure of differentiable renderers, pros and cons and assumptions of different algorithms.

Differentiable Rendering



Differentiable Signed Distance Function Rendering,
Vicini et al. 2022



InstantNGP, Müller et al. 2022

Why?	Part of our problem relates to **inverting** the rendering process: Given 2D images, we want to reconstruct 3D scenes. Differentiable rendering is one way of exactly inverting the rendering process.
What you'll learn.	The structure of differentiable renderers, pros and cons and assumptions of different algorithms.

Tentative Schedule

- **Module 1: Image Formation and Multi-View Geometry (Lectures 2 & 3, Assignment 1)**
  - Rigid-body-transforms, projective geometry, camera transforms, camera models.

- **Module 2: 3D Representations, Rendering, and Differentiable Rendering (Lectures 4-7, Assignment 2)**
  - Surface vs. volume representations, continuous vs. discrete representations
  - Conventional rendering in computer graphics
  - Differentiable rendering & neural rendering

- ***Module 3: Deep learning for 3D reconstruction & processing, representation learning (Lectures 8-12, Assignment 3)***
  - Deep learning for 3D reconstruction
  - Deep learning on 3D data
  - Geometric deep learning

Prior-based Reconstruction

Seeing 3D Objects in a Single Image via Self-Supervised Static-Dynamic Disentanglement, Sharma et al. 2022

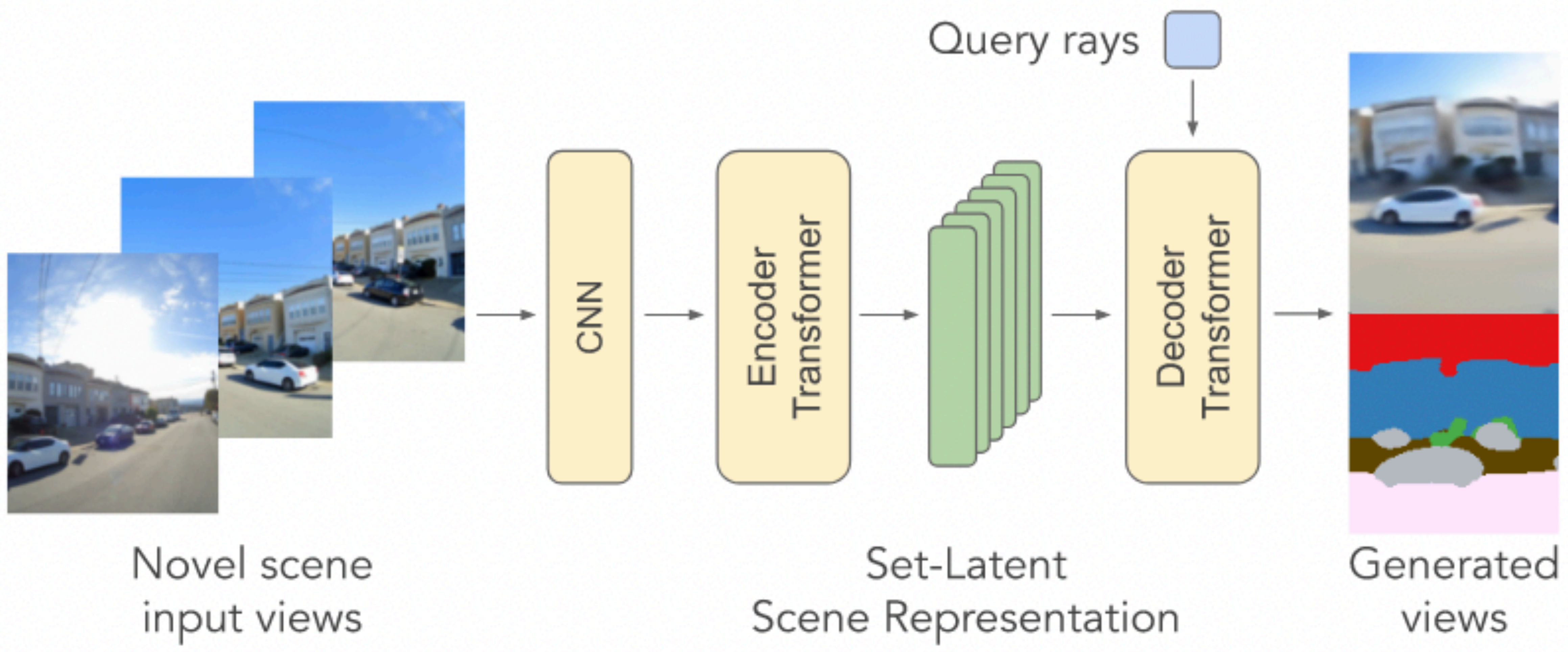
Why?	We humans can reconstruct 3D from **incomplete** observations, by using knowledge that we have learned about the world. Deep learning is the best way we know to date to learn such priors from data.
What you'll learn.	How to express priors over 3D scenes using deep learning, different ways of doing inference (encoding, auto-decoding)

Prior-based Reconstruction

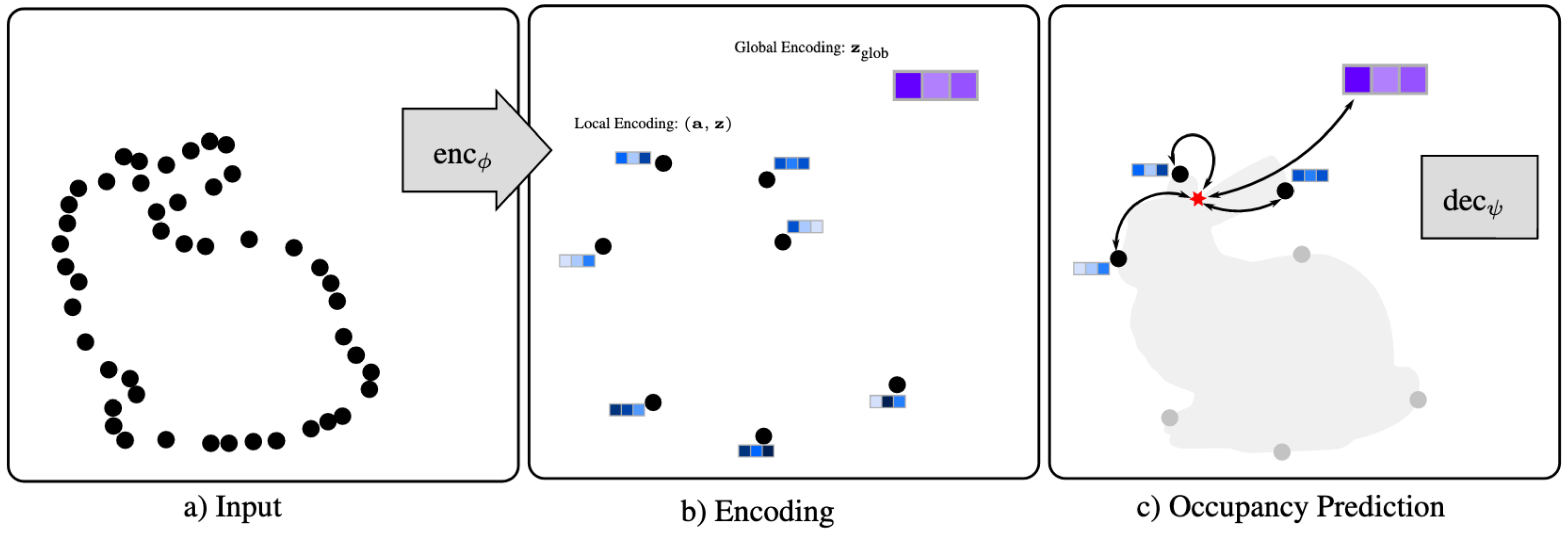
Seeing 3D Objects in a Single Image via Self-Supervised Static-Dynamic Disentanglement, Sharma et al. 2022

Why?	We humans can reconstruct 3D from **incomplete** observations, by using knowledge that we have learned about the world. Deep learning is the best way we know to date to learn such priors from data.
What you'll learn.	How to express priors over 3D scenes using deep learning, different ways of doing inference (encoding, auto-decoding)

Advanced Inference Topics



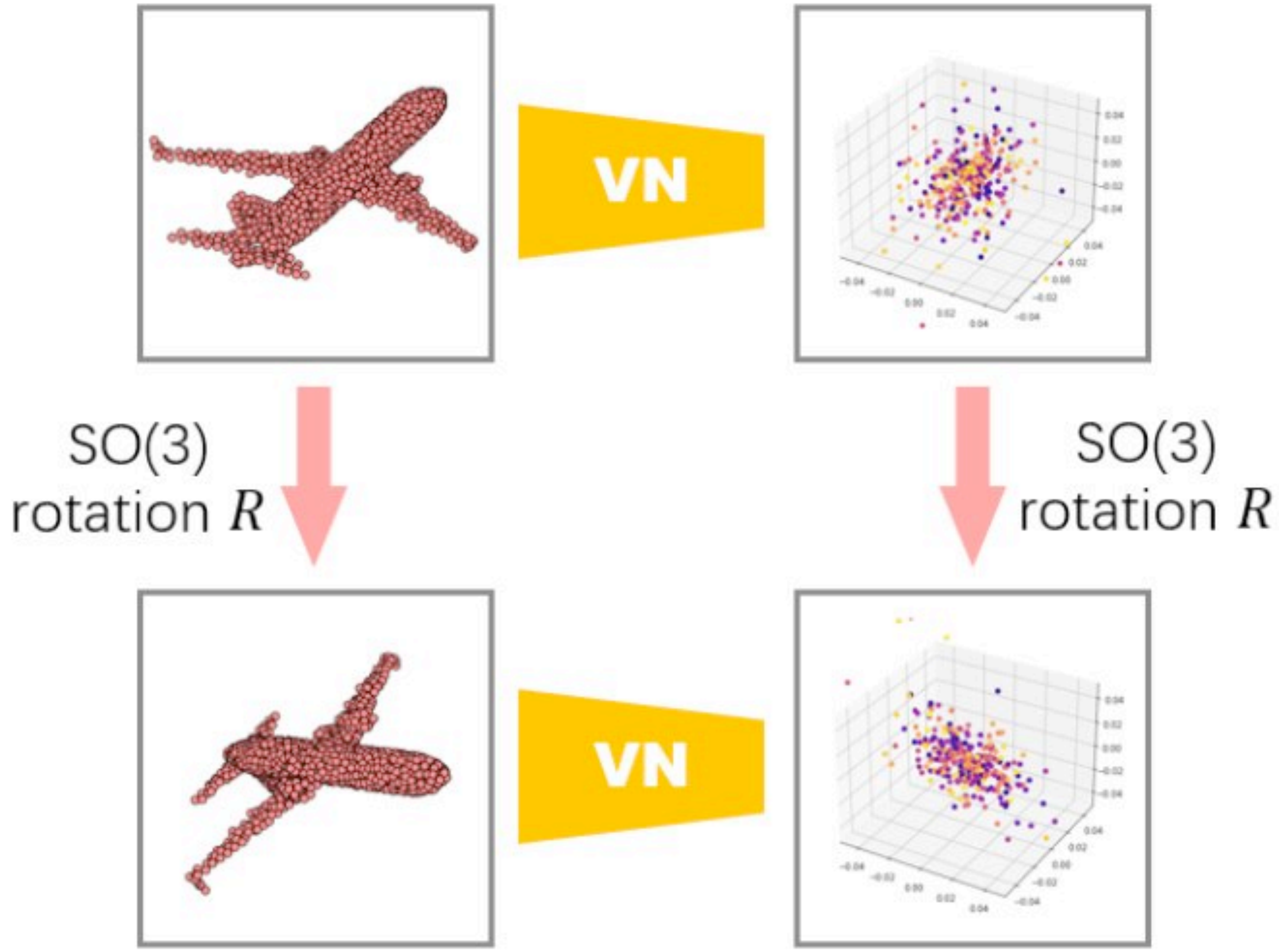
Scene Representation Transformer, Sajjadi et al. 2022



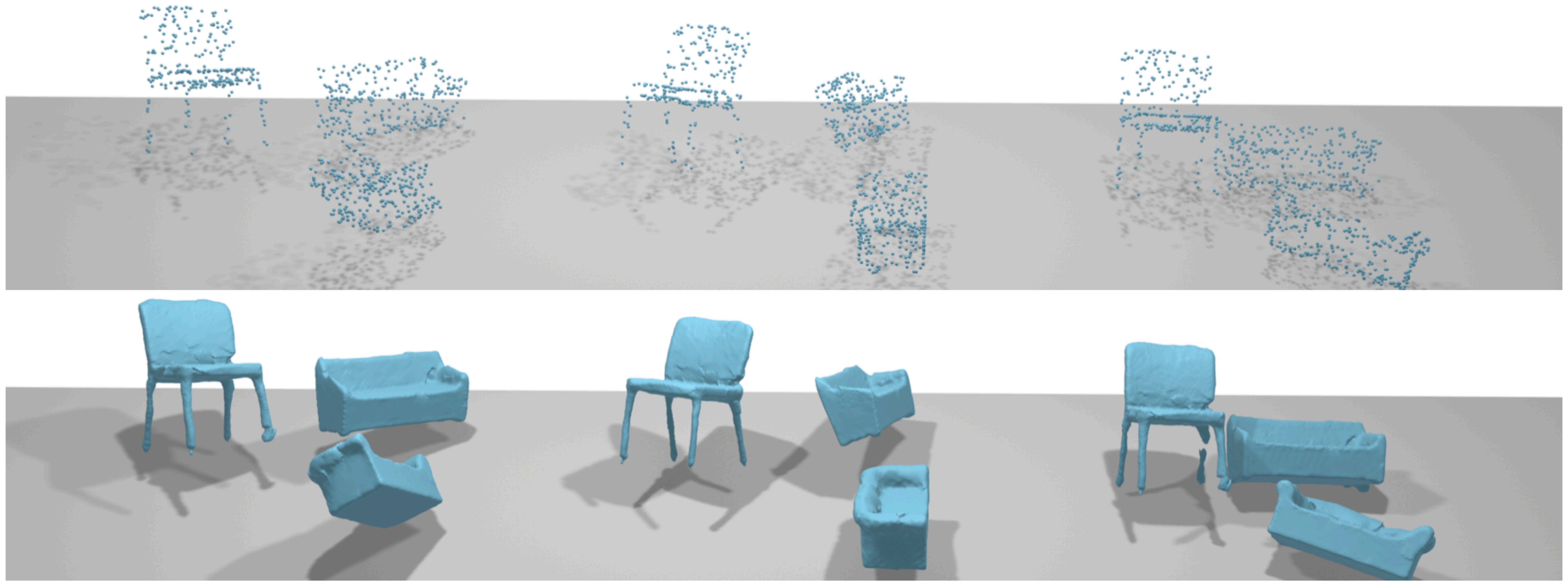
AIR-Nets, Giebenhain et al. 2021

Why?	Deep learning affords us alternative ways of formulating 3D reconstruction that don't involve exact forward models (such as multi-bounce ray-marching). This is potentially more tractable & biologically plausible.
What you'll learn.	Transformer-based inference, attention-based conditioning, gradient-based meta-learning, light field neural scene representations.

Geometric Deep Learning



Vector Neurons, Deng et al. 2021



SE(3)-Equivariant Attention Networks for Shape Reconstruction in Function Space, Chatzipantazis & Pertigkiozoglou et al.

Why?	**To guarantee generalization to certain transformations (such as 3D translations and rotations), we need to build special neural network architectures.**
What you'll learn.	**Basics of group theory, neural network architectures that respect symmetries (shift, rotation, scale equivariance)**

Tentative Schedule

- **Module 1: Image Formation and Multi-View Geometry (Lectures 2 & 3, Assignment 1)**
  - Rigid-body-transforms, projective geometry, camera transforms, camera models.

- **Module 2: 3D Representations, Rendering, and Differentiable Rendering (Lectures 4-7, Assignment 2)**
  - Surface vs. volume representations, continuous vs. discrete representations
  - Conventional rendering in computer graphics
  - Differentiable rendering & neural rendering

- ***Module 3: Deep learning for 3D reconstruction & processing, representation learning (Lectures 8-12, Assignment 3)***
  - Deep learning for 3D reconstruction
  - Deep learning on 3D data
  - Geometric deep learning

- ***Module 4: Motion & Objectness (Lectures 13 - 14)***
  - Deep learning for 3D reconstruction
  - Geometric deep learning

Dynamic Neural Scene Representations & Physics

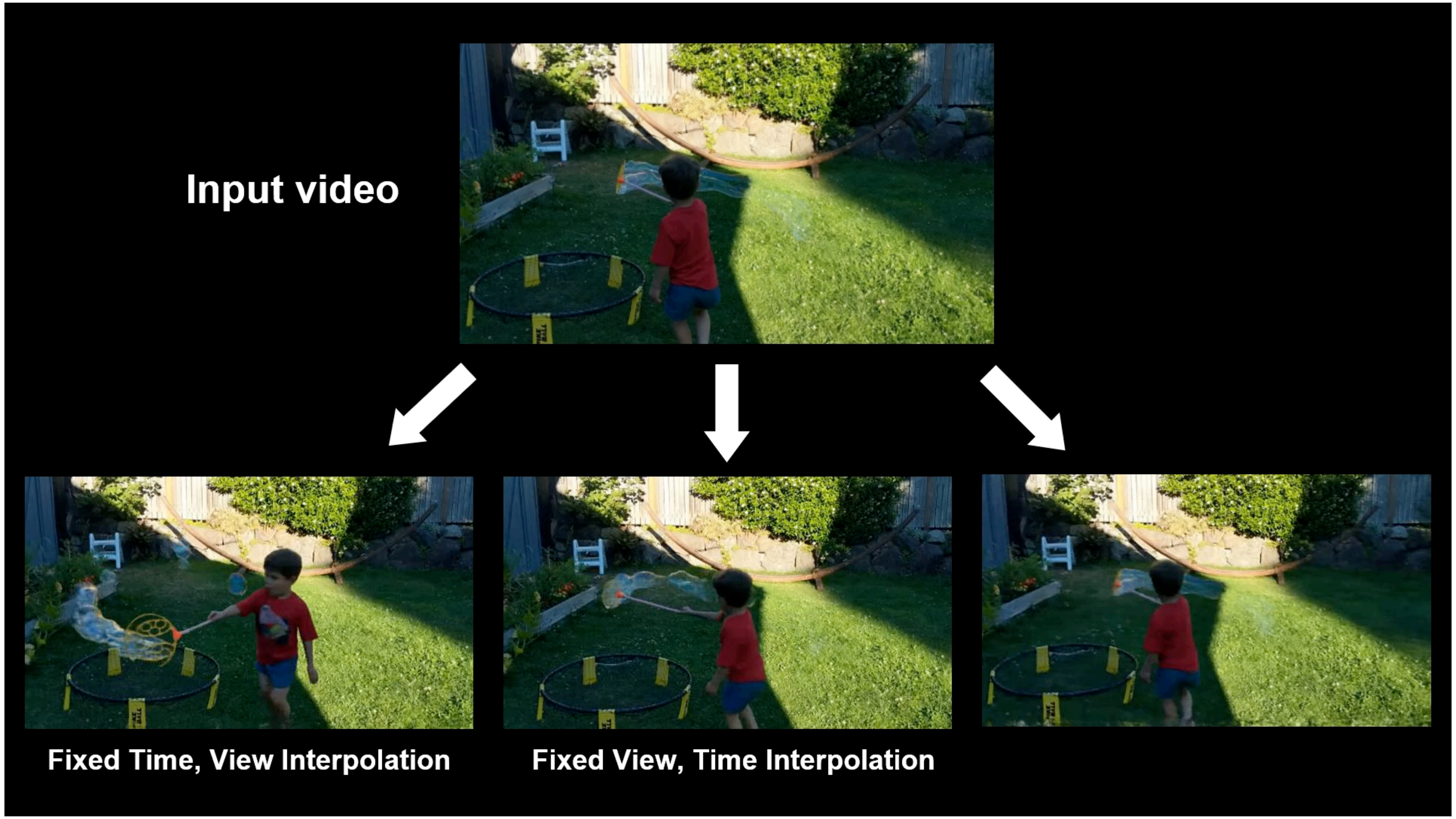


Image Courtesy: Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes. *Li et. al.*

Why?	Our world is not static, but dynamic. We can reason about dynamic processes (rigid and non-rigid), such as tracking a car over time or realizing when something is deforming. We need mathematical tools for this.
What you'll learn.	How to reconstruct dynamic scenes. Shortcomings and open research directions.

Dynamic Neural Scene Representations & Physics

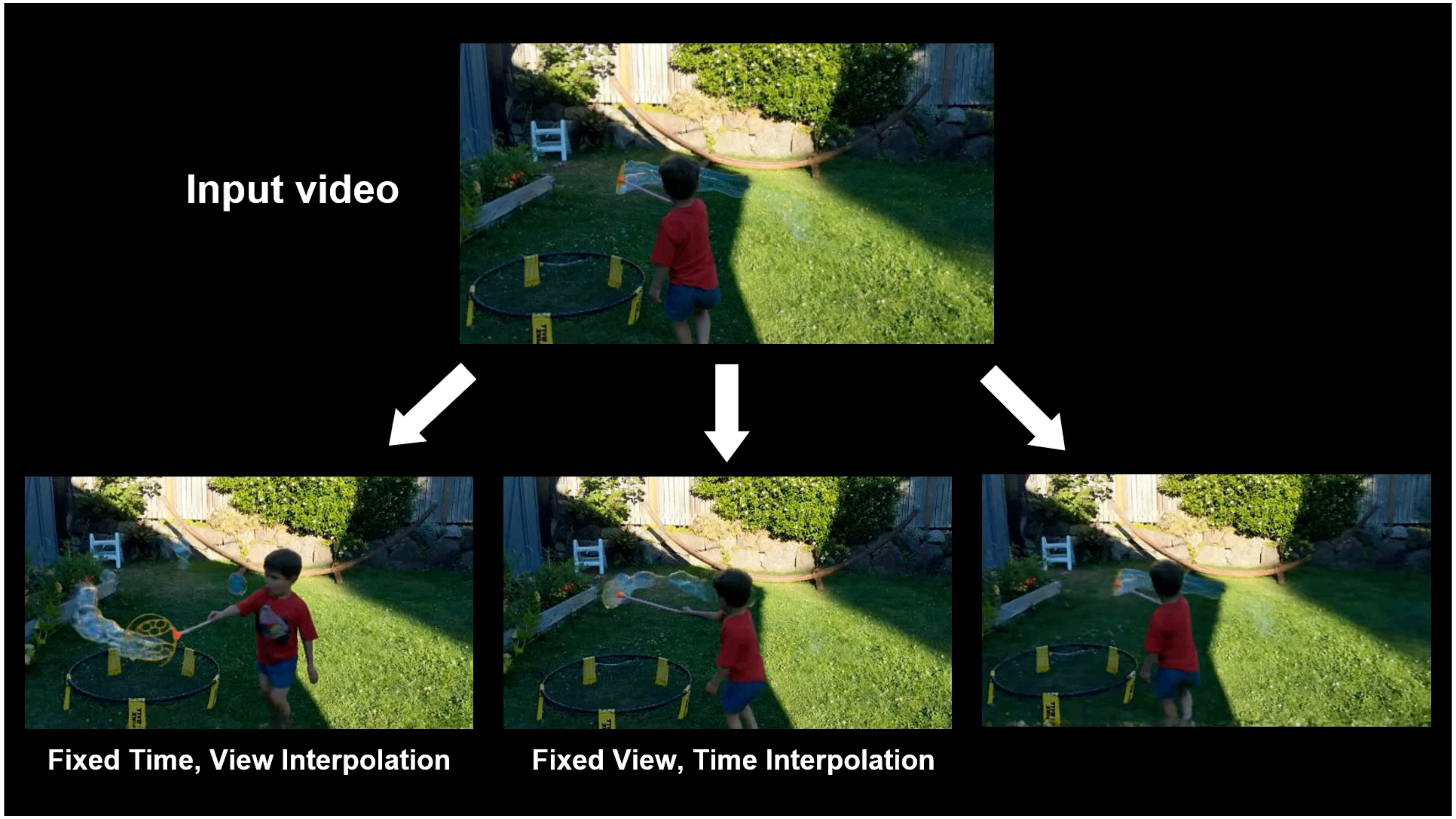


Image Courtesy: Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes. *Li et. al.*

Why?	Our world is not static, but dynamic. We can reason about dynamic processes (rigid and non-rigid), such as tracking a car over time or realizing when something is deforming. We need mathematical tools for this.
What you'll learn.	How to reconstruct dynamic scenes. Shortcomings and open research directions.


RoDynRF: Robust Dynamic Radiance Fields, Liu et al. 2023, CVPR

RoDynRF: Robust Dynamic Radiance Fields, Liu et al. 2023, CVPR

TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement, Doersch et al., 2023

TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement, Doersch et al., 2023

Input RGBD video

Total-Recon, Song et al., ICCV 2023

Input RGBD video

Tentative Schedule

- **Module 1: Image Formation and Multi-View Geometry (Lectures 2 & 3, Assignment 1)**
  - Rigid-body-transforms, projective geometry, camera transforms, camera models.

- **Module 2: 3D Representations, Rendering, and Differentiable Rendering (Lectures 4-7, Assignment 2)**
  - Surface vs. volume representations, continuous vs. discrete representations
  - Conventional rendering in computer graphics
  - Differentiable rendering & neural rendering

- ***Module 3: Deep learning for 3D reconstruction & processing, representation learning (Lectures 8-12, Assignment 3)***
  - Deep learning for 3D reconstruction
  - Deep learning on 3D data
  - Geometric deep learning

- ***Module 4: Motion & Objectness (Lectures 13 - 14)***
  - Deep learning for 3D reconstruction
  - Geometric deep learning

- ***Module 5: Applications & Guest Lectures (Lectures 15-17)***
  - Robotics
  - Scientific Discovery
  - …

What this class will *not* cover

- Advanced Computer Graphics:
  - Mesh processing, computational geometry
  - Multi-bounce rendering, ray-tracing, path-tracing
  - Material rendering
- Basics of Deep Learning: Backpropagation, differentiable programming, Stochastic Gradient Descent
- Supervised deep learning

Evaluation

Evaluation

- **Assignments**: *70%*
  - 4 assignments, 17.5% each
  - Individual submissions (feel free to discuss, but not share code)
  - 5 late days total (25% penalty for each extra late day)

Evaluation

- **Assignments**: *70%*
  - *4* assignments, 17.5% each
  - Individual submissions (feel free to discuss, but not share code)
  - 5 late days total (25% penalty for each extra late day)

- **Final Projects**: *30%*
  - Teams of **1-3**
  - Proposal (5%), Mid-term Presentation (5%), Final Report (10%), Final Presentation / Poster Session (10%)
  - No late days for report!

Honor Code

Write all code by yourself!

But, I am stuck on this problem?

- Talk to your fellow classmates about your approach.
- Reach out on Piazza
- Try solving a toy version of the problem where you know the solution.
- Write visualization code for inputs, intermediates, and outputs.

Discussing ideas (not code) is fair game.

I am running late? Use late days. If you ran out of late days, reach out.

What can get me into trouble?

- Copying code from anywhere including the internet, classmate's solution, whiteboard, etc.
- Using AI assisted code completion.

Communication

Communication

- **Lectures**
  - Tuesdays and Thursdays between 2:30pm and 4:00pm @ 32-124 (Stata ground floor).
  - Slides will be posted on the course website shortly before each lecture
  - All lectures will be recorded and uploaded to Canvas after the lecture.
  - Please ask questions - don’t let me move on if something is unclear!

Communication

- **Lectures**
  - Tuesdays and Thursdays between 2:30pm and 4:00pm @ 32-124 (Stata ground floor).
  - Slides will be posted on the course website shortly before each lecture
  - All lectures will be recorded and uploaded to Canvas after the lecture.
  - Please ask questions - don't let me move on if something is unclear!

- **Canvas**
  - *https://canvas.mit.edu/courses/16653*

Communication

- **Lectures**
  - Tuesdays and Thursdays between 2:30pm and 4:00pm @ 32-124 (Stata ground floor).
  - Slides will be posted on the course website shortly before each lecture
  - All lectures will be recorded and uploaded to Canvas after the lecture.
  - Please ask questions - don't let me move on if something is unclear!

- **Canvas**
  - *https://canvas.mit.edu/courses/16653*

- **Piazza**
  - *https://piazza.com/class/lm6rtmrt6c6lo*
  - Please post any/all questions - also helps other students in this class.
  - Create private posts to instructors instead of emailing (really helps keep track!)

Communication

- **Lectures**
  - Tuesdays and Thursdays between 2:30pm and 4:00pm @ 32-124 (Stata ground floor).
  - Slides will be posted on the course website shortly before each lecture
  - All lectures will be recorded and uploaded to Canvas after the lecture.
  - Please ask questions - don't let me move on if something is unclear!

- **Canvas**
  - *https://canvas.mit.edu/courses/16653*

- **Piazza**
  - *https://piazza.com/class/lm6rtmrt6c6lo*
  - Please post any/all questions - also helps other students in this class.
  - Create private posts to instructors instead of emailing (really helps keep track!)

- **Office Hours**
  - Great opportunities to clear doubts, discuss project ideas, get debugging help etc.
  - We are happy to chat - please come say hi!
  - David's office hours: TBD
  - Vincent's office hours: TBD

Communication

- **Lectures**
  - Tuesdays and Thursdays between 2:30pm and 4:00pm @ 32-124 (Stata ground floor).
  - Slides will be posted on the course website shortly before each lecture
  - All lectures will be recorded and uploaded to Canvas after the lecture.
  - Please ask questions - don't let me move on if something is unclear!

- **Canvas**
  - *https://canvas.mit.edu/courses/16653*

- **Piazza**
  - *https://piazza.com/class/lm6rtmrt6c6lo*
  - Please post any/all questions - also helps other students in this class.
  - Create private posts to instructors instead of emailing (really helps keep track!)

- **Office Hours**
  - Great opportunities to clear doubts, discuss project ideas, get debugging help etc.
  - We are happy to chat - please come say hi!
  - David's office hours: TBD
  - Vincent's office hours: TBD

- **Website**
  - *https://www.scenerepresentations.org/courses/inverse-graphics/*
  - We will link all course materials here (in addition to canvas)

Disclaimer - New Course (and Instructor)

From CMU 16-889, Shubham Tulsiani.

Disclaimer - New Course (and Instructor)

There will likely be bugs in assignments and lectures - apologies in advance!

From CMU 16-889, Shubham Tulsiani.

Disclaimer - New Course (and Instructor)

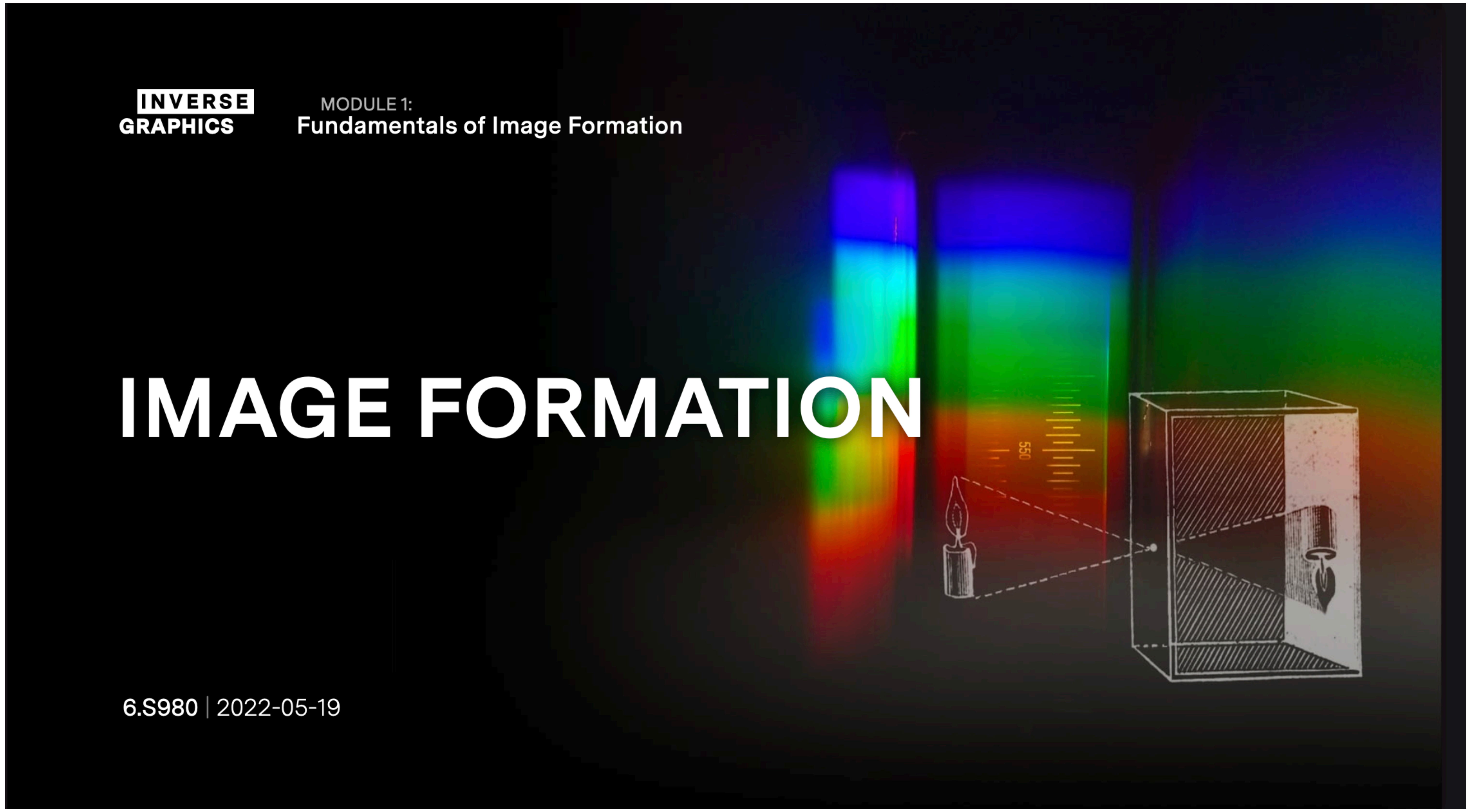
There will likely be bugs in assignments and lectures - apologies in advance!

Anonymous feedback after a few weeks to help improve

From CMU 16-889, Shubham Tulsiani.

Questions?

Next up!


INVERSE
GRAPHICS
MODULE 1:
Fundamentals of Image Formation
IMAGE FORMATION
6.S980 | 2022-05-19