Lecture 26: Robot Learning Overview and Deep Learning Frontiers

Danfei Xu

Administrative

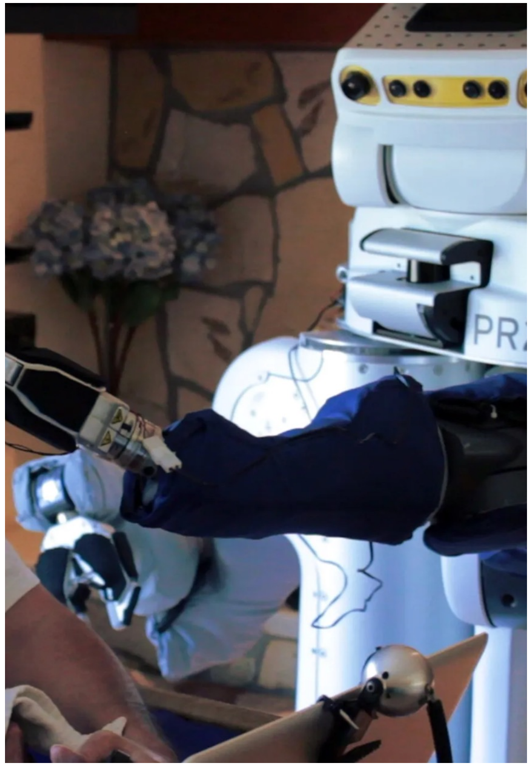
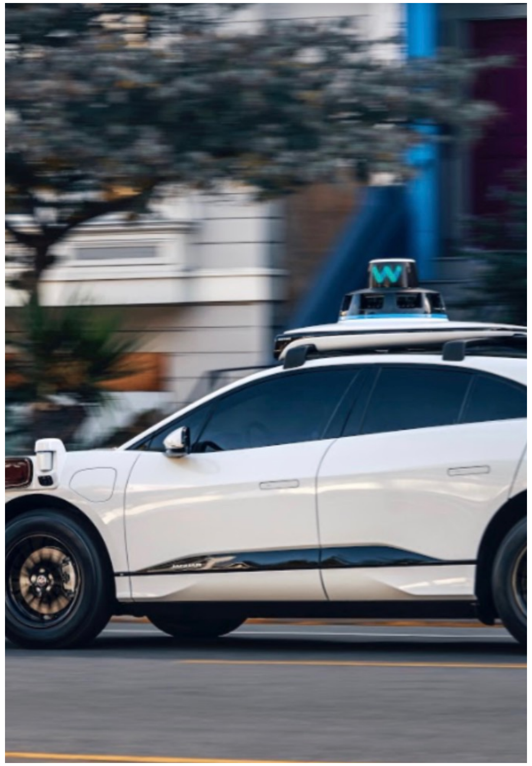
Remember to fill CIOS evaluation!

Poster session Dec 5th 5pm-6:30pm

- Bring your poster. We will provide easels.
- You will be given an easel number the day of the event.
- The TAs will start by grading half of the posters in the first 45 min, and the other half in the second 45 min.
- You will know which batch you are in at the event.
- Check out other posters if your batch is not being graded.
- We will have pizza and dessert available
- We will announce a **best project award** at the end of the poster session.
- The event is open to the GT community. Expect many attendees, so bring your best work. And tell your friends to come too!

Past & present: robots in factories & semi-structured environments

Future: robots everywhere!

How we program these robots today …

Image source

Manual programming is not enough!

diverse tasks

messy environments

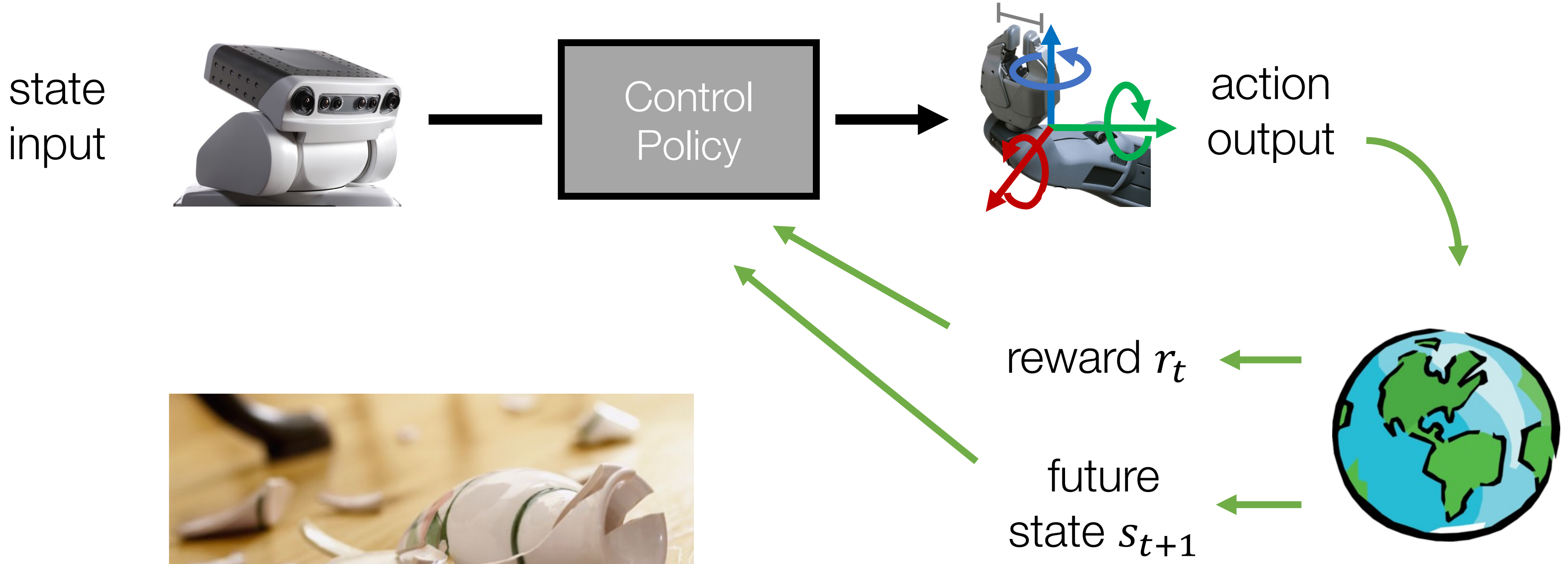
The Moravec's paradox

Moravec's paradox is the observation … contrary to traditional assumptions, reasoning requires very little computation, but **sensorimotor and perception skills require enormous computational resources**. (Wikipedia)

Marvin Minsky: "In general, we're least aware of what our minds do best" … "we're more aware of simple processes that don't work well than of complex ones that work flawlessly".

Can we teach robots through data / examples?

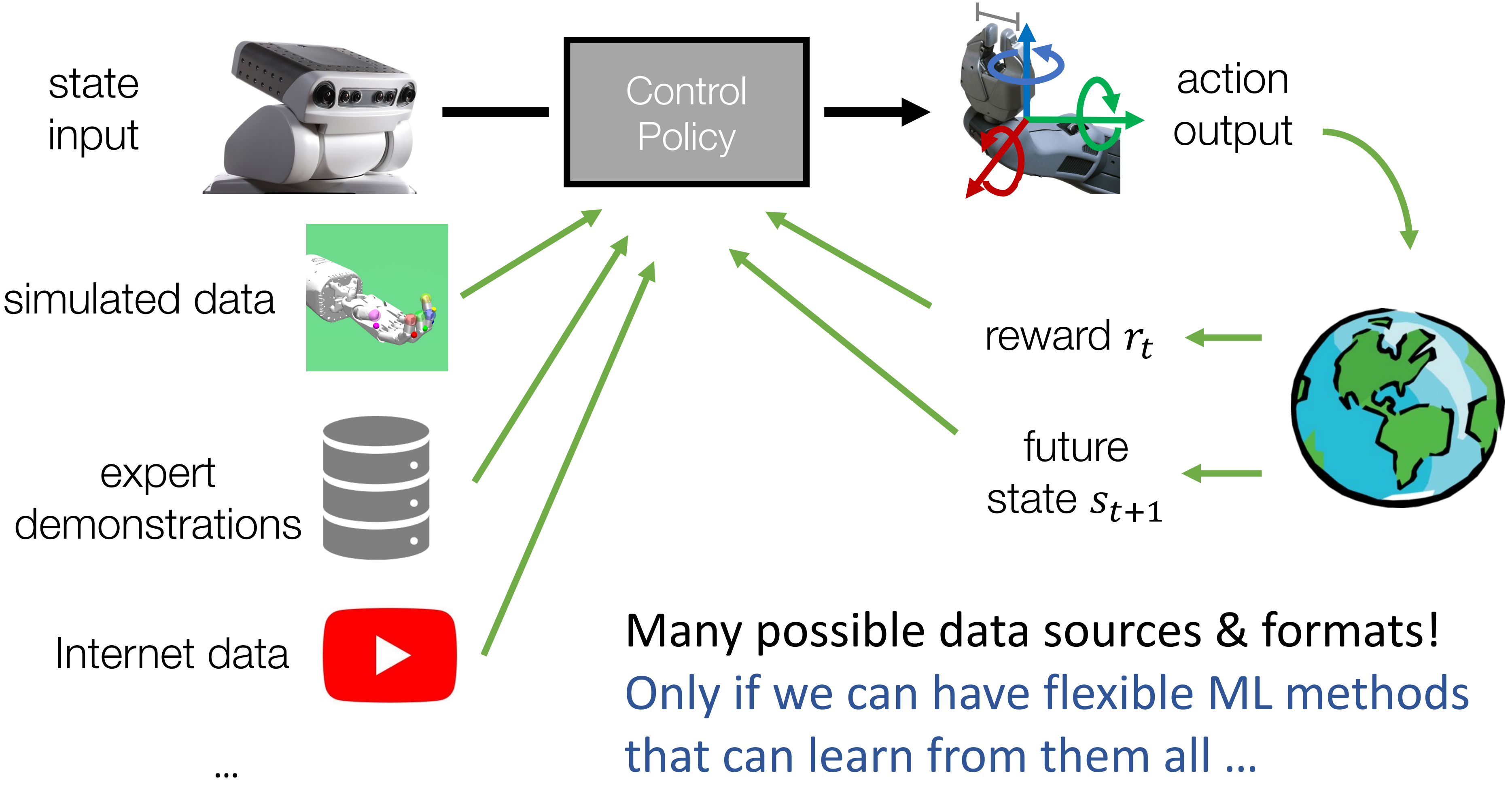
Can we teach robots through data / examples?



Very useful, but expensive to acquire in the physical world

Data source: experience from trial-and-error

Can we teach robots through data / examples?

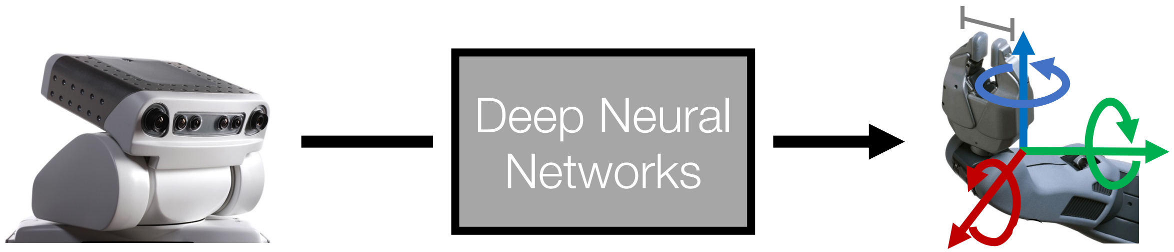


Many possible data sources & formats!

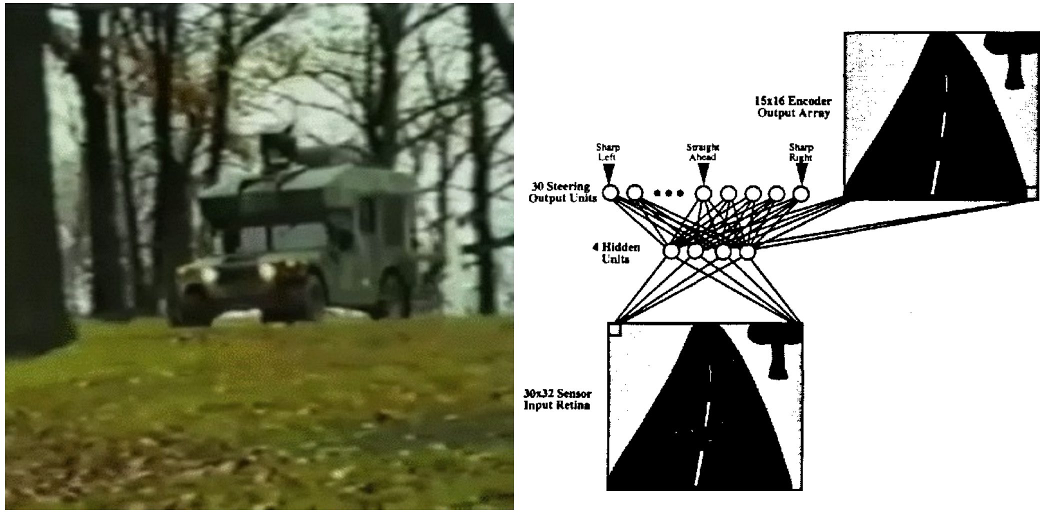
Only if we can have flexible ML methods that can learn from them all …

Deep Learning for Robotics

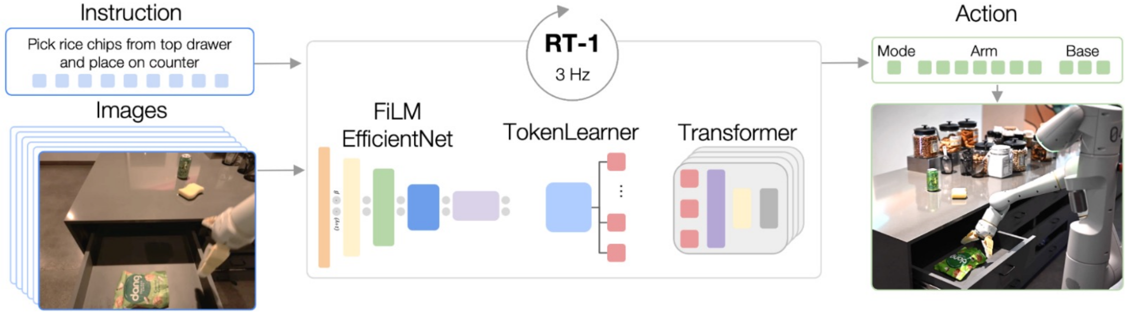
state
input



action
output



The ALVINN project at CMU
(Pomerleau 1988)

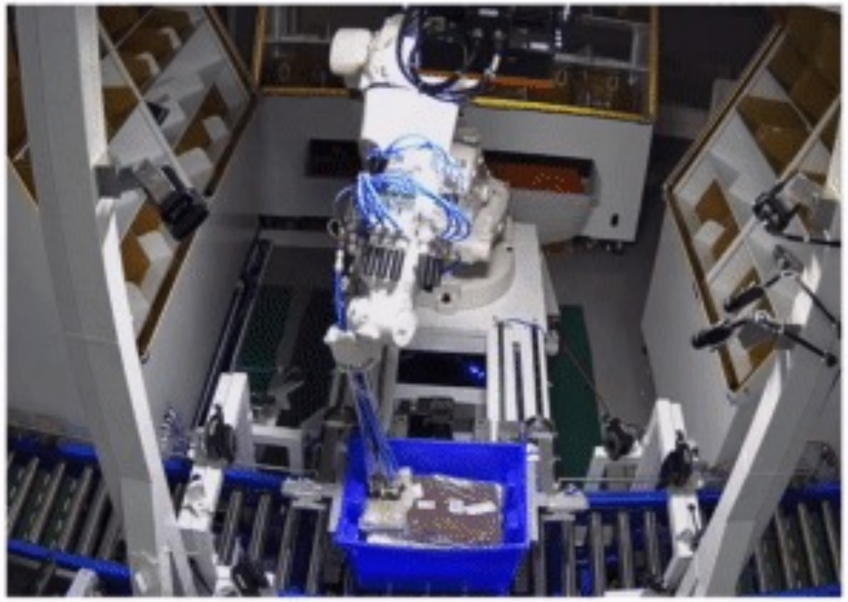


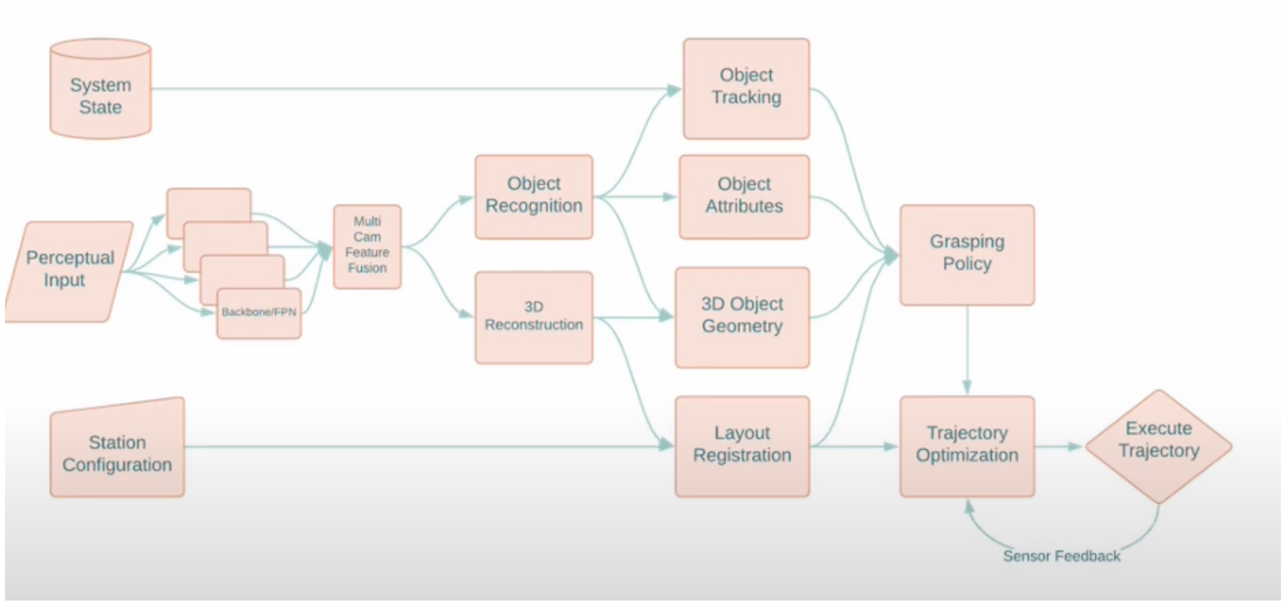
"Robot Transformer (RT1)" from
Google Robotics (2023)

Deep Learning is **NOT** all you need!

Deep Learning is NOT all your need

Robots today have some deep learning components, but nothing is fully "end-to-end".

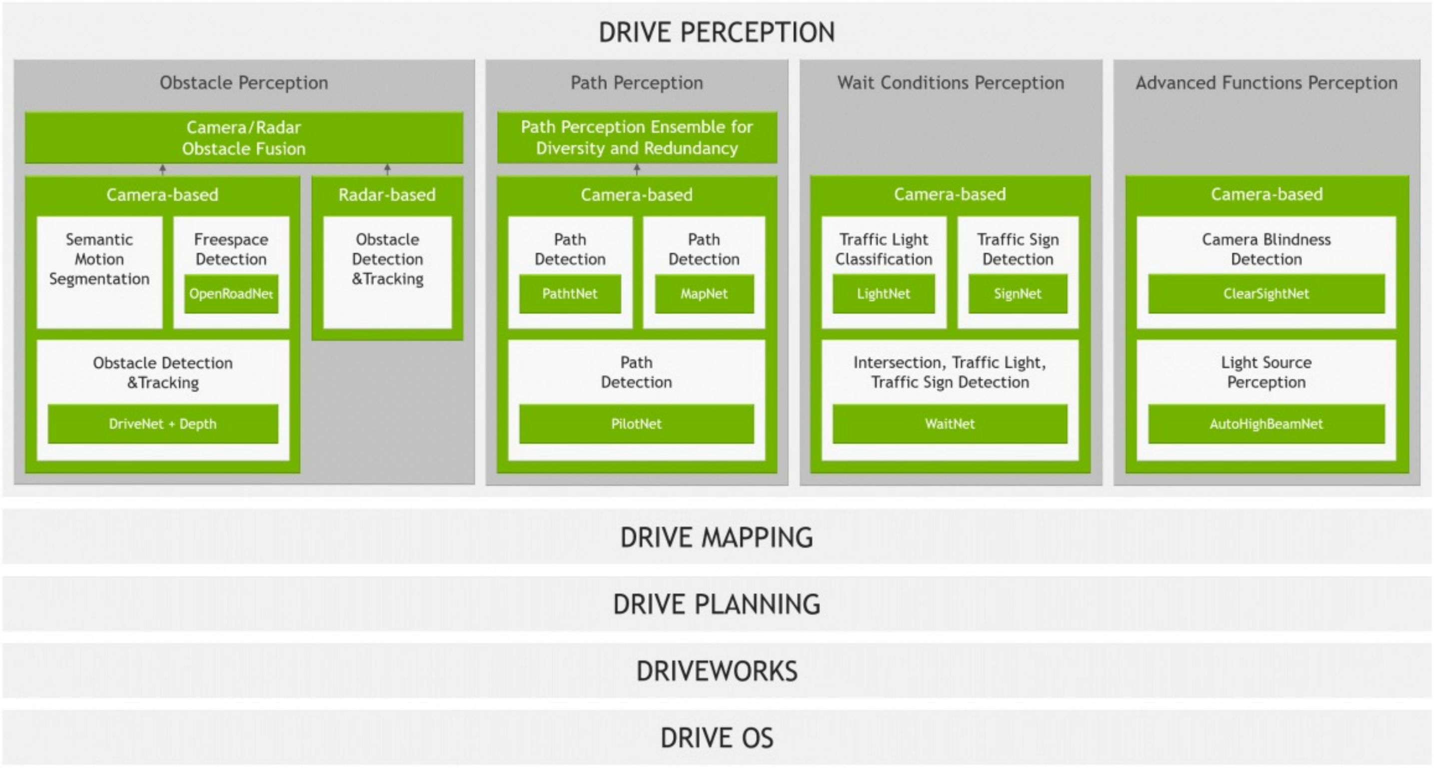




The "control policy" of a learning robot for e-commerce fulfillment.
Covariant AI (video source)

Deep Learning is NOT all your need

Robots today have some deep learning components, but nothing is fully "end-to-end”.



The perception pipeline of an autonomous driving stack NVIDIA (image source)

Robot Learning

Robot learning is a research field at the intersection of machine learning and robotics. It studies techniques allowing a robot to acquire novel skills or adapt to its environment through learning algorithms. (Wikipedia)

More concise version:

Principles, algorithms, and systems that allow robots to improve by learning from data.

Robot Learning research today (2023): *what* and *how* to learn.

Robot Learning: ML don’t need to (and shouldn’t) be applied to everything!

The reason that we want to use machine learning is to deal with variation, noise, and things that are hard to model.



Unlike computer vision and natural language understanding, robotics often deal with physics, which we know well. So we don’t need to learn everything!

Both a challenge and an opportunity for robot learning: how to best combine what we know and what we need to learn.

State of Robot Learning Research

Mastery: be able to solve tasks that are hard / infeasible to solve by manual programming.

Scaling: apply a method / framework to a broad range of tasks by scaling up data sources.

Generalization: solve new tasks in new environments and scenarios; show emerging behaviors that are not in the training data.

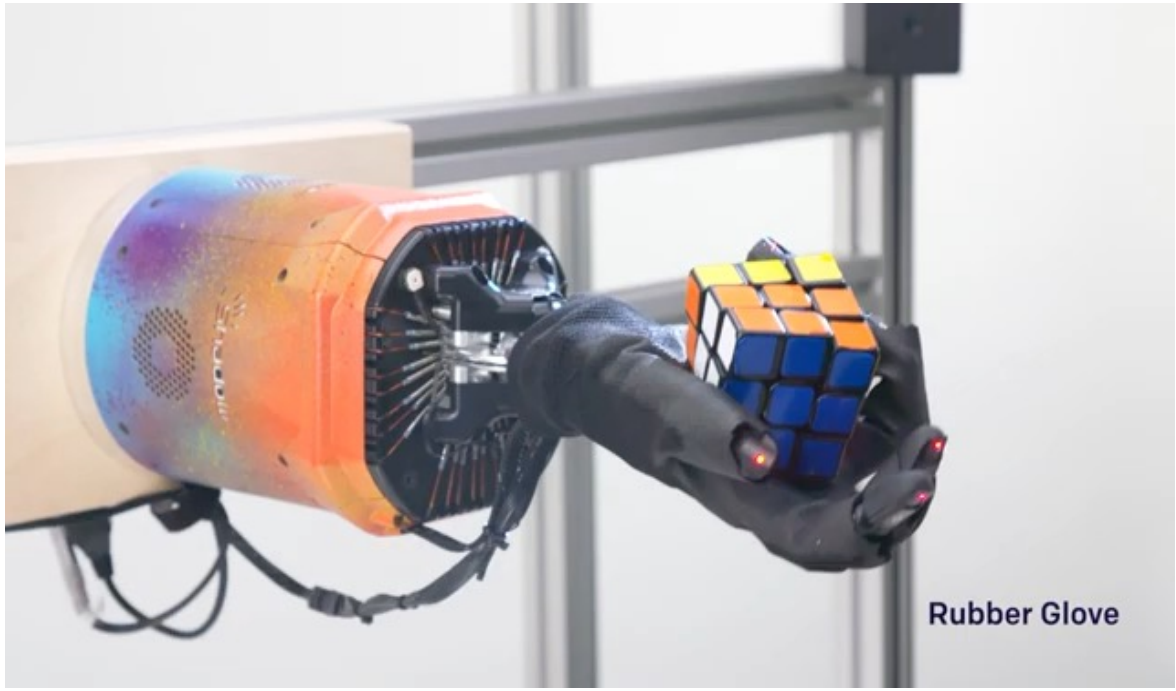
State of Robot Learning Research

Mastery: be able to solve tasks that are hard / infeasible to solve by manual programming (successes in some domains).

Scaling: apply a method / framework to a broad range of tasks by scaling up data sources (ongoing progress).

Generalization: solve new tasks in new environments and scenarios; show emerging behaviors that are not in the training data (holy grail, no real progress yet).

Examples of mastering hard tasks



Source: OpenAI

Source: ETH Zurich

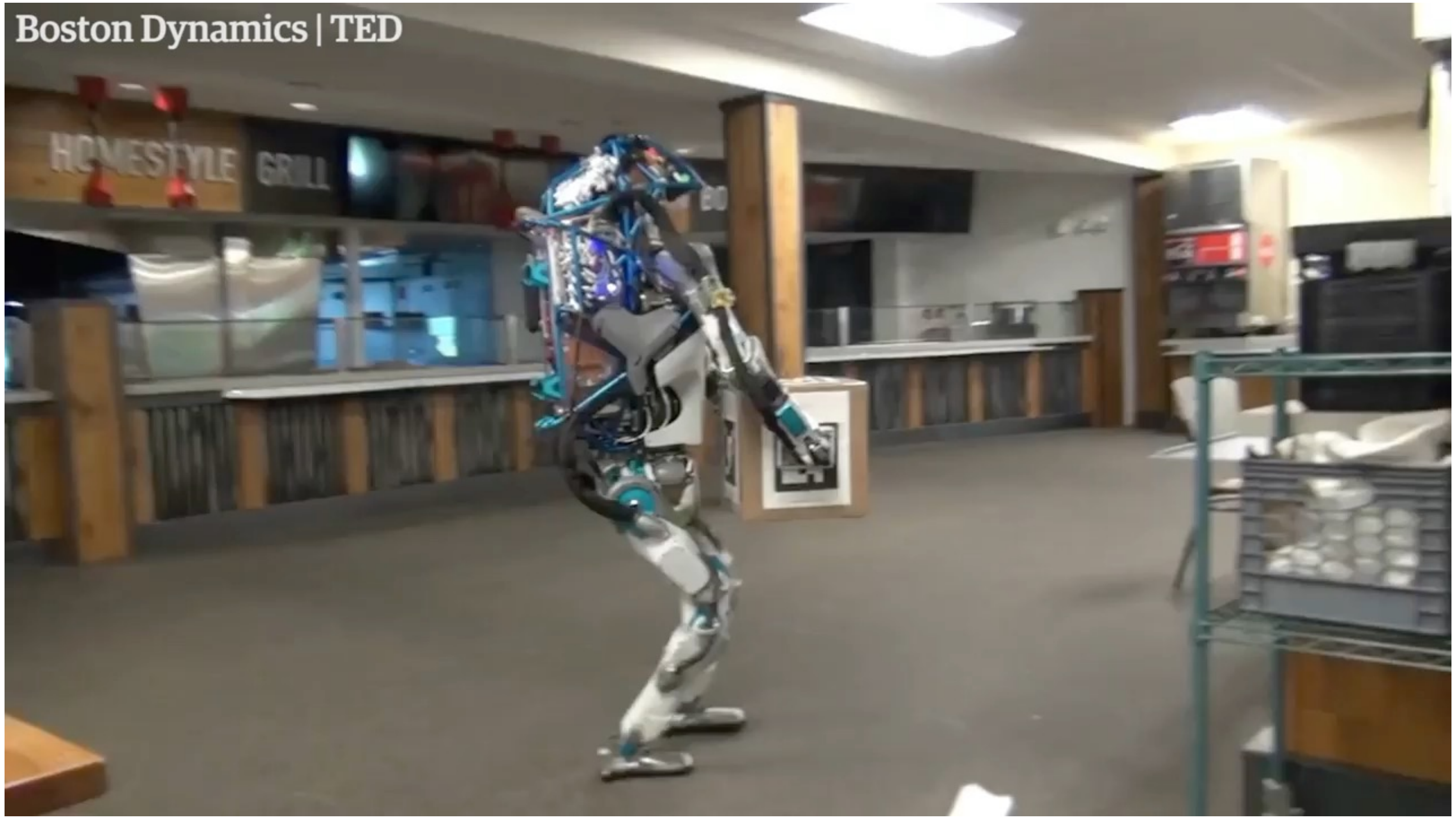
Sim-to-real Reinforcement Learning

Examples of scaling up data sources

RT1: Imitation learning from 130k demonstrations
collected over the course of 17 months

https://robotics-transformer.github.io/assets/rt1.pdf

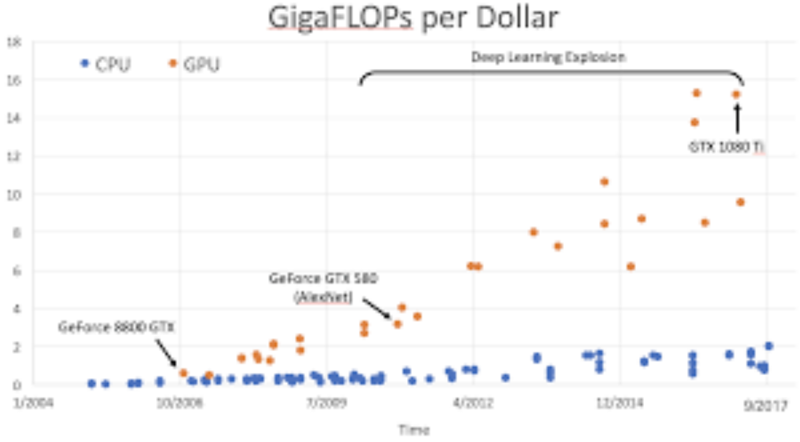
No where near generalizable decision making!



https://www.ted.com/talks/marc_raibert_meet_spot_the_robot_dog_that_can_run_hop_and_open_doors?language=en

It’s a great time to work on robot learning!

general-purpose
learning algorithms



more compute

general-purpose
robot hardware

Deep Learning for Robotics (CS 8803-DLM): an overview

2D/3D Perception and Grasping	Act without Models: Reinforcement Learning and Imitation Learning	Model-based Decision Making: Learning for Planning and Control
Learning to grasp: DexNet family	Model-free RL: TRPO, SAC, DDPG	Model-based RL
Learning to grasp: visual affordances and action-as-perception	Offline Reinforcement Learning	Learning Planning Representations
VLM for Manipulation	Imitation Learning: Behavior Cloning, Learning from human data	Learning Control Representations
Tactile Sensing	Imitation Learning: Inverse RL, Generative Adversarial Imitation,	Task and Motion Planning
Multimodal Representation Learning	Sim-to-real transfer	Learning for Task and Motion Planning
	Curiosity and Exploration	Language Model for Robotics
	Human-in-the-loop Robot Learning	

Frontiers of Deep Learning

Topics we didn't get time to cover:

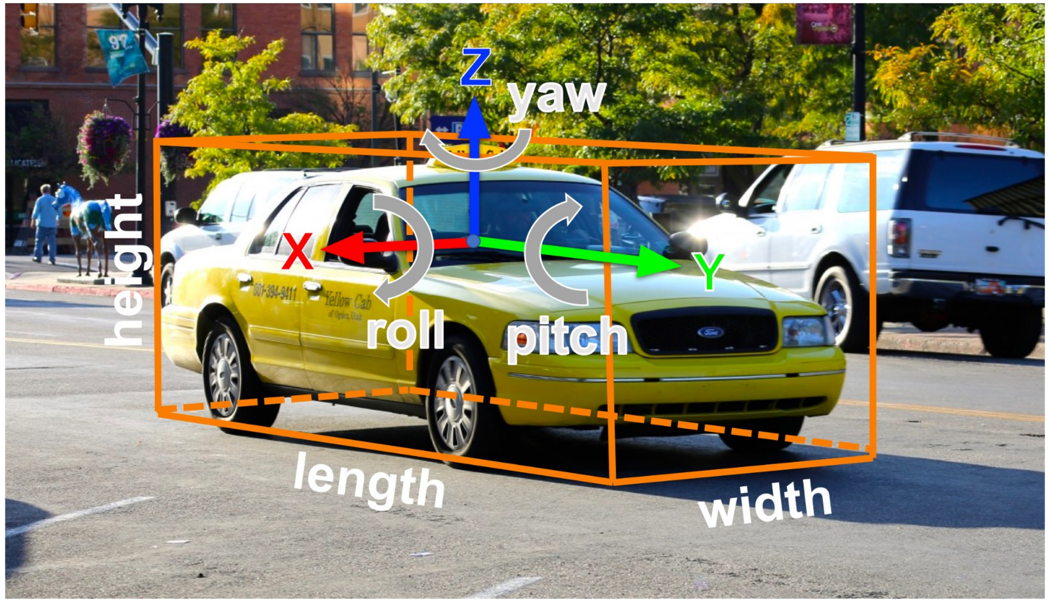
- Vision Transformers
- Graph Neural Nets
- Metric learning
- AutoML
- 3D perception & reconstruction
- Memory modeling
- Few-shot / meta learning
- Neural Radiance Field (NeRF) / implicit representations
- Adversarial learning and robustness
- Continual / lifelong learning
- Visual reasoning
- Neural Theorem Proving
- Neural Program Induction / Synthesis
- MLSys
- Many topics in NLP …

Frontiers of Deep Learning

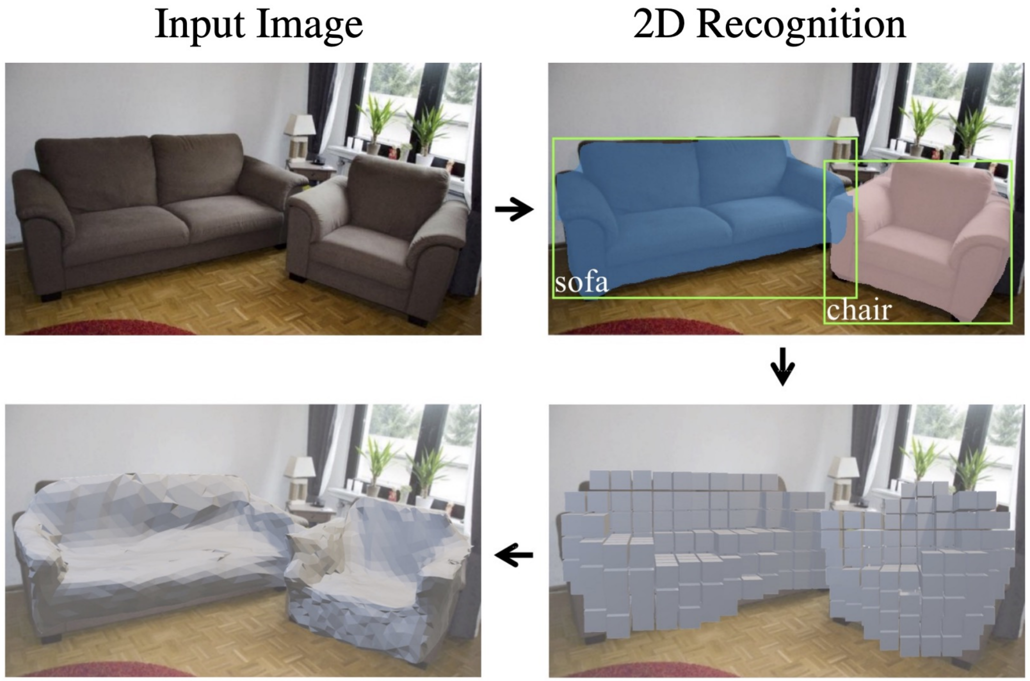
Topics we didn't get time to cover:

- Vision Transformers
- Graph Neural Nets
- Metric learning
- AutoML
- **3D perception & reconstruction**
- Memory modeling
- Few-shot / meta learning
- **Neural Radiance Field (NeRF) / implicit representations**
- Adversarial learning and robustness
- Continual / lifelong learning
- Visual reasoning
- Neural Theorem Proving
- Neural Program Induction / Synthesis
- MLSys
- Many topics in NLP …

3D Perception



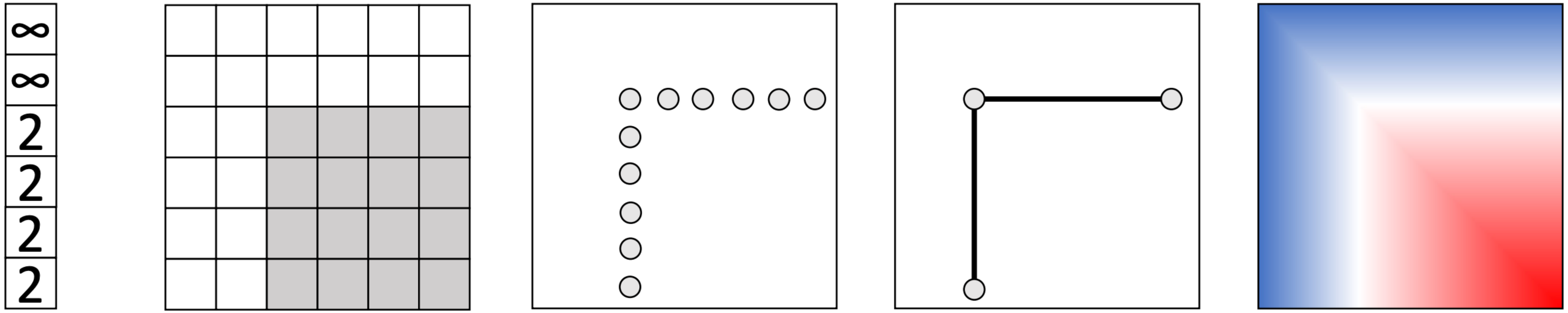
3D Object Detection / Pose Estimation



3D Object Reconstruction

3D Perception

Many possible ways to represent the 3D world …

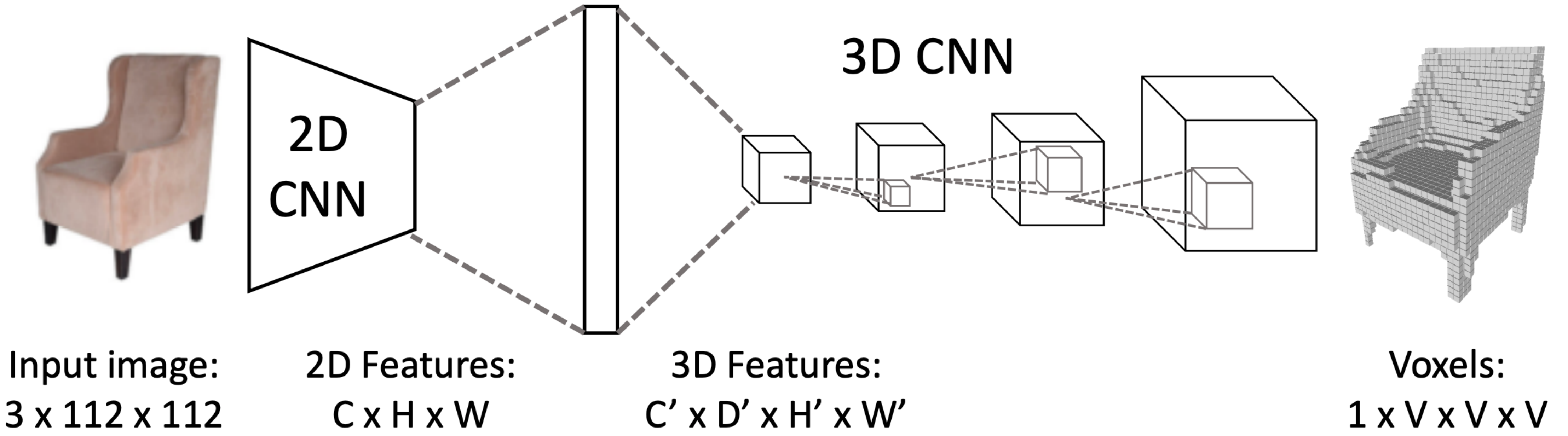


Depth Map | Voxel Grid | Pointcloud | Mesh | Implicit Surface

Each representation requires different neural network architectures!

Figure credit: Justin Johnson

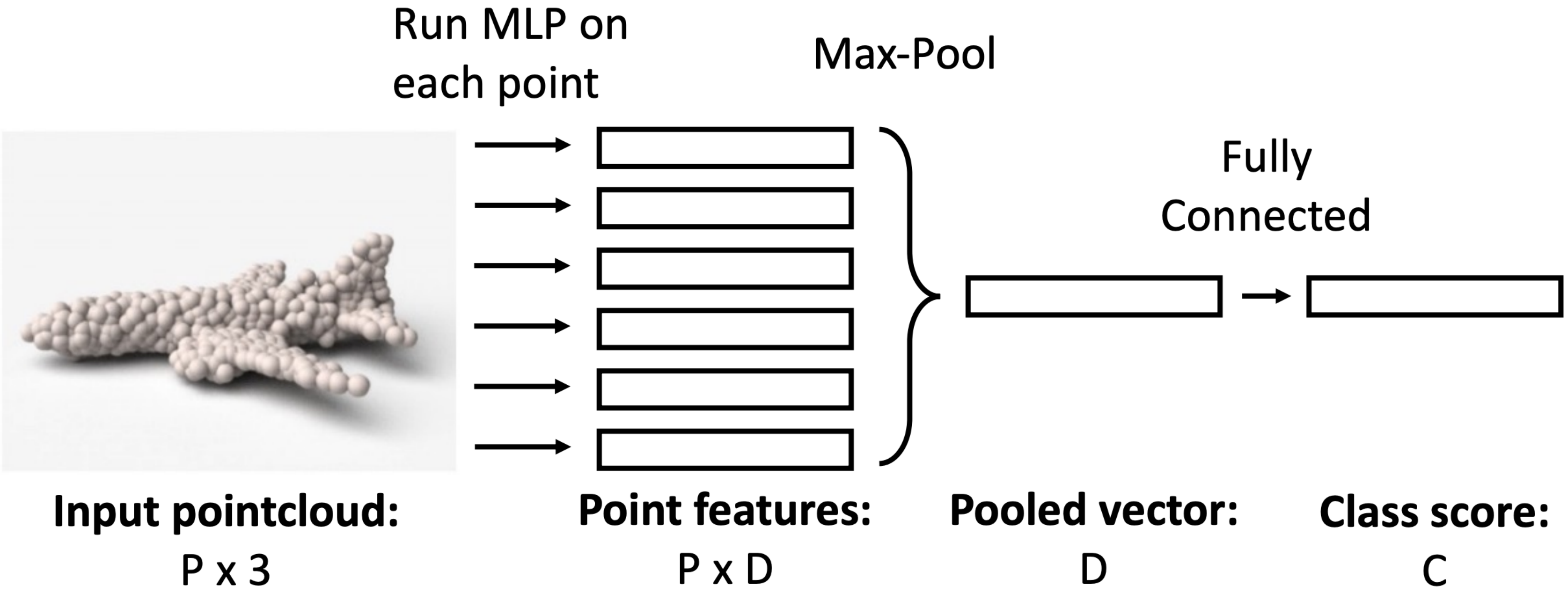
3D Perception



3D Convolution for Voxel-based 3D Reconstruction

Choy et al., 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction, ECCV 2016

3D Perception

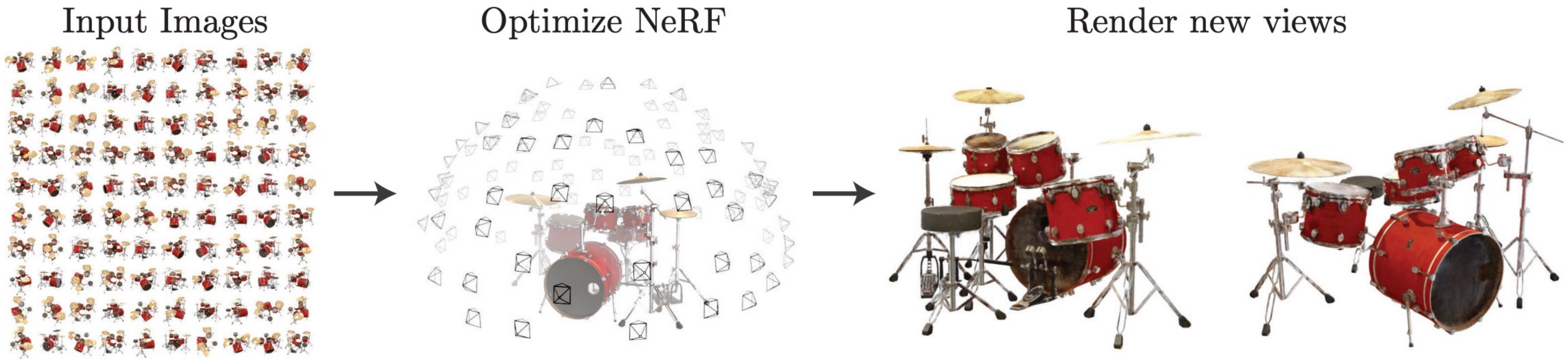


(Simplified) PointNet architecture for 3D point cloud classification

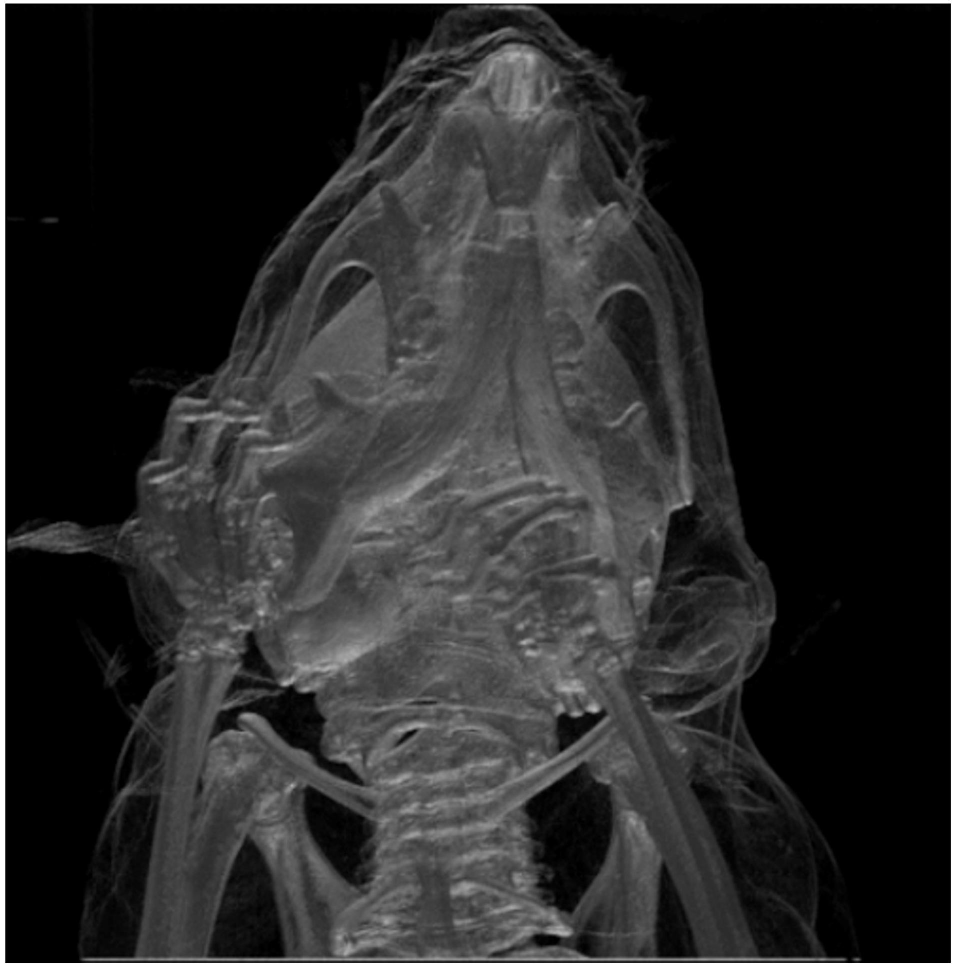
Choy et al., 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction, ECCV 2016

Neural Radiance Field

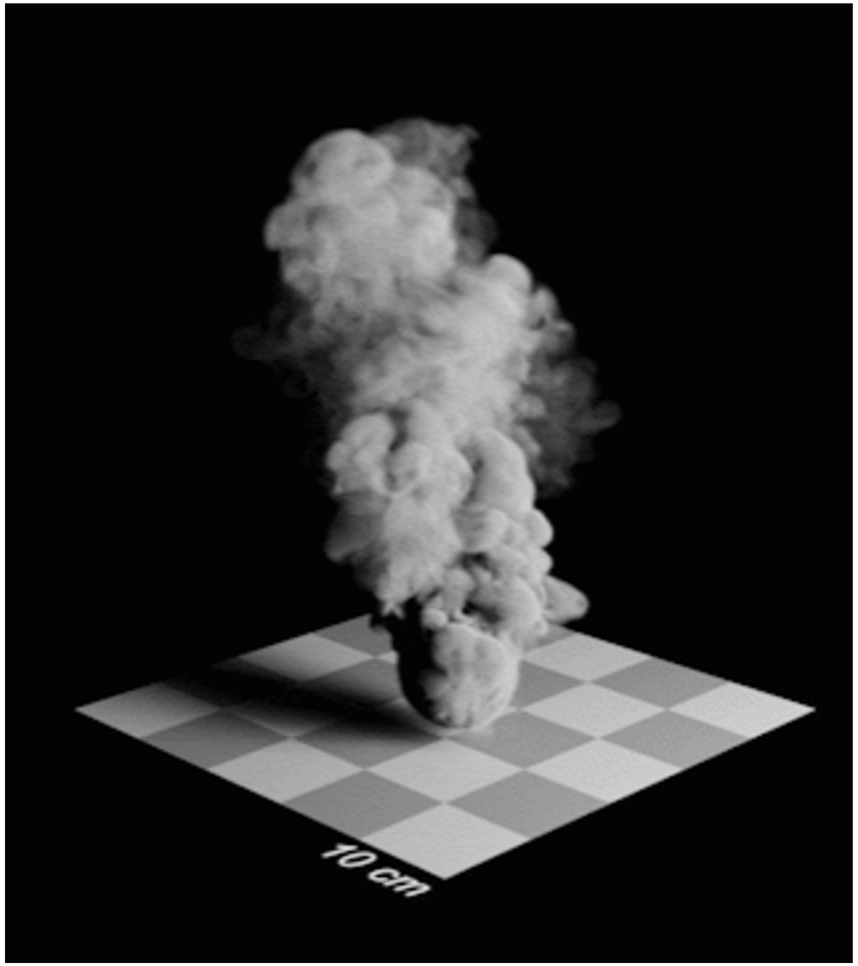
Neural Radiance Field: View Synthesis



Volume Rendering

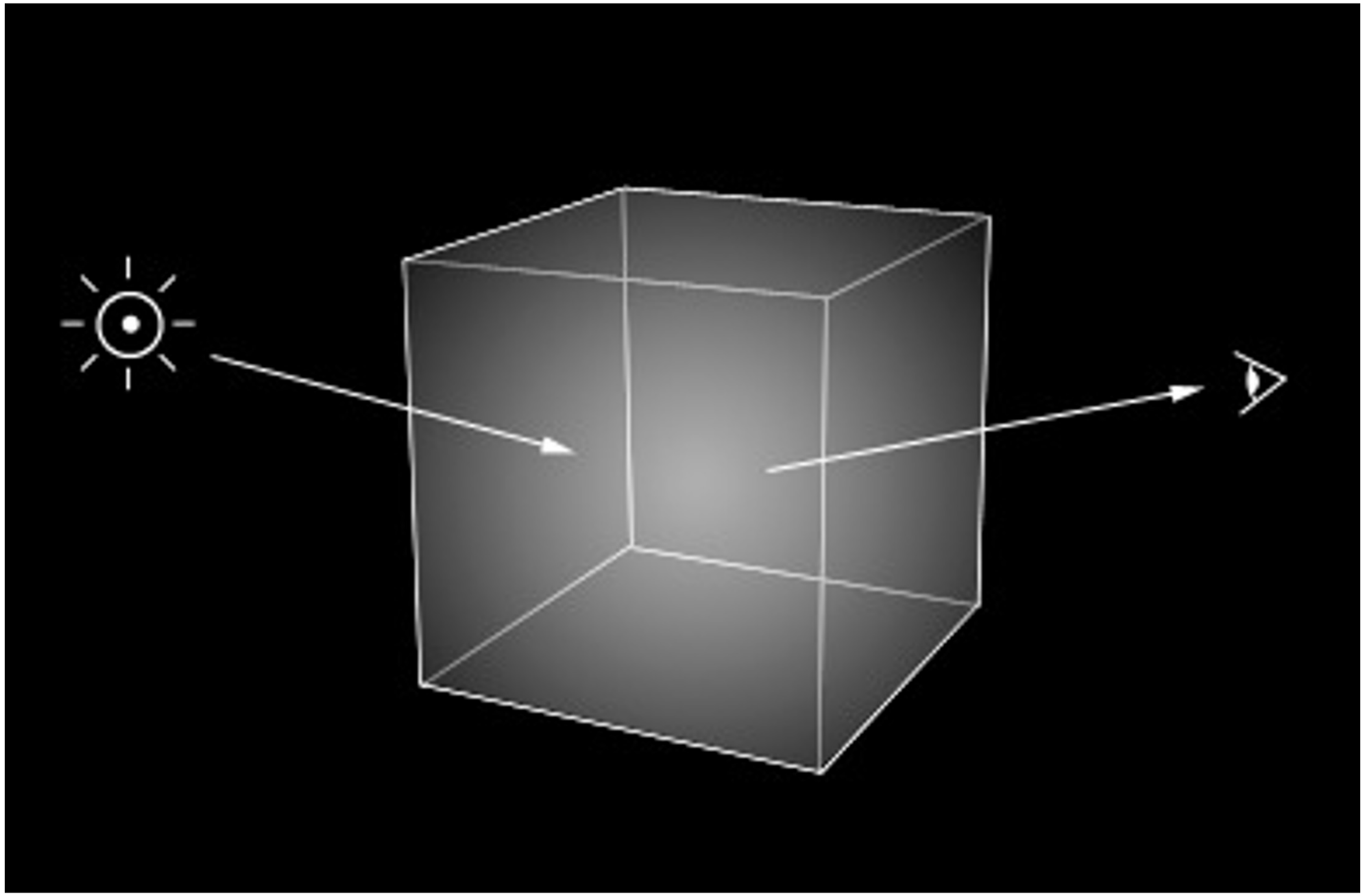


https://en.wikipedia.org/wiki/Volume_rendering

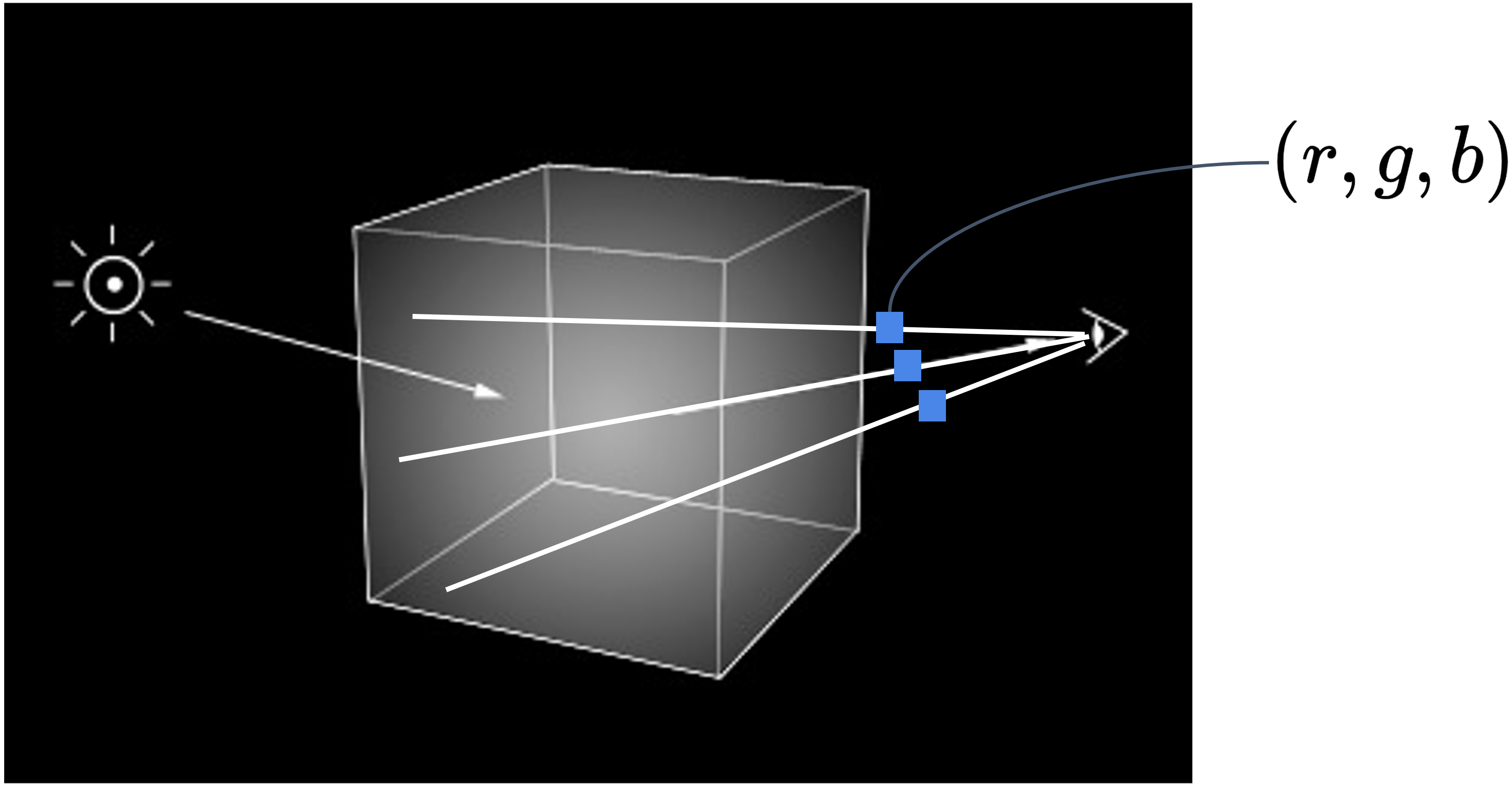


https://coronarenderer.freshdesk.com/support/solutions/articles/12000045276-how-to-use-the-corona-volume-grid-

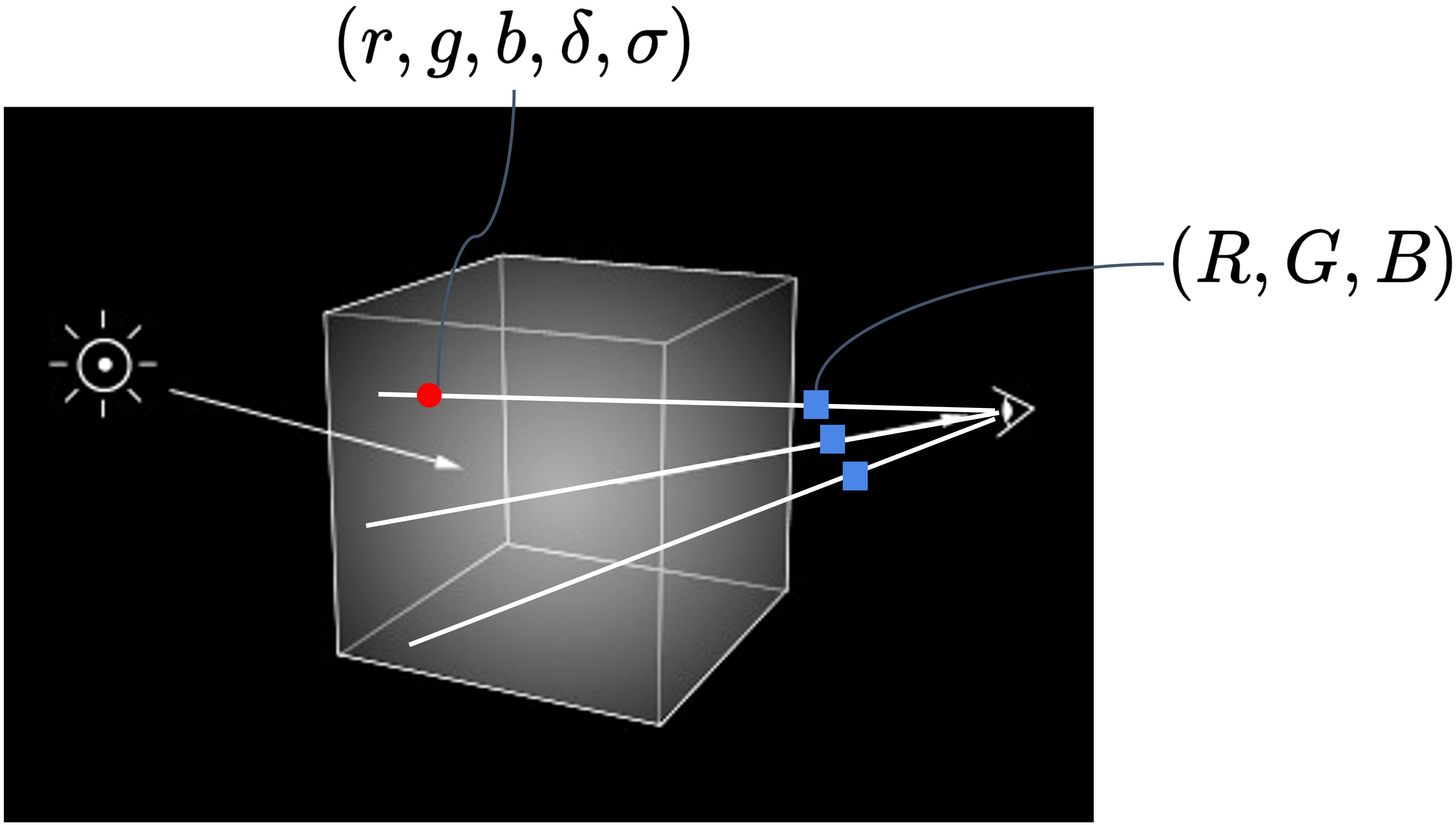
Volume Rendering: Ray Marching



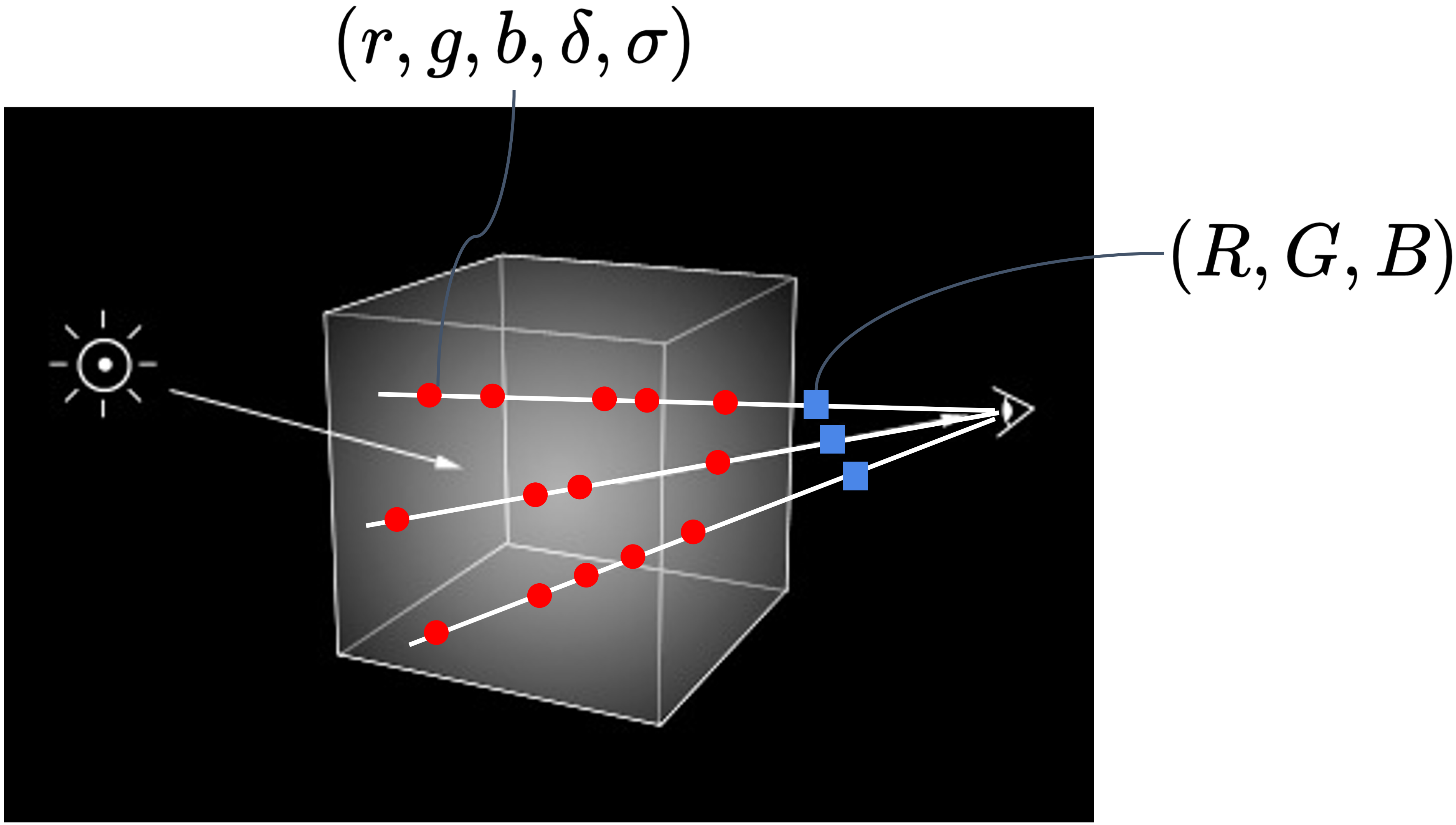
Volume Rendering: Ray Marching



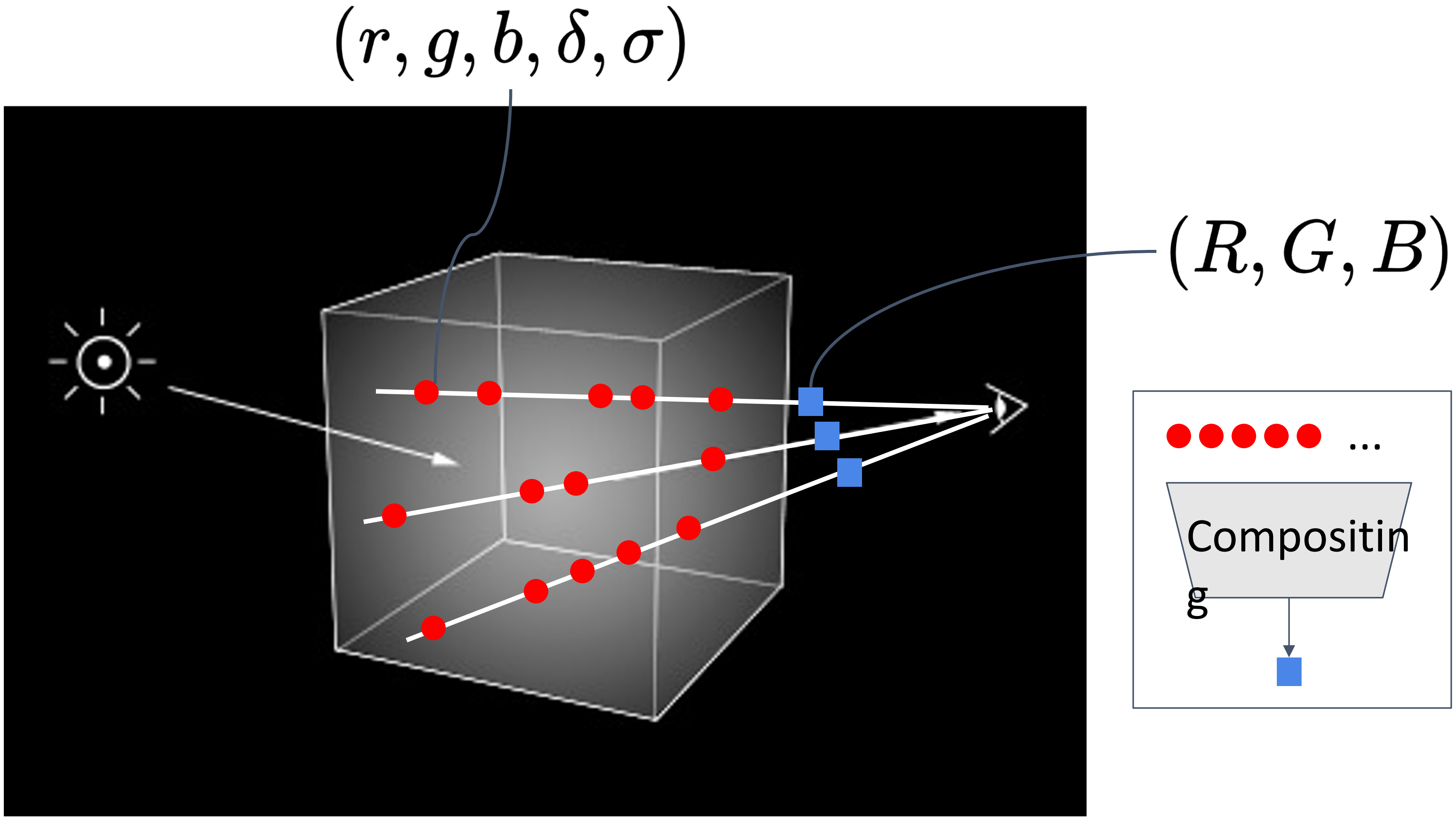
Volume Rendering: Ray Marching



Volume Rendering: Ray Marching



Volume Rendering: Ray Marching

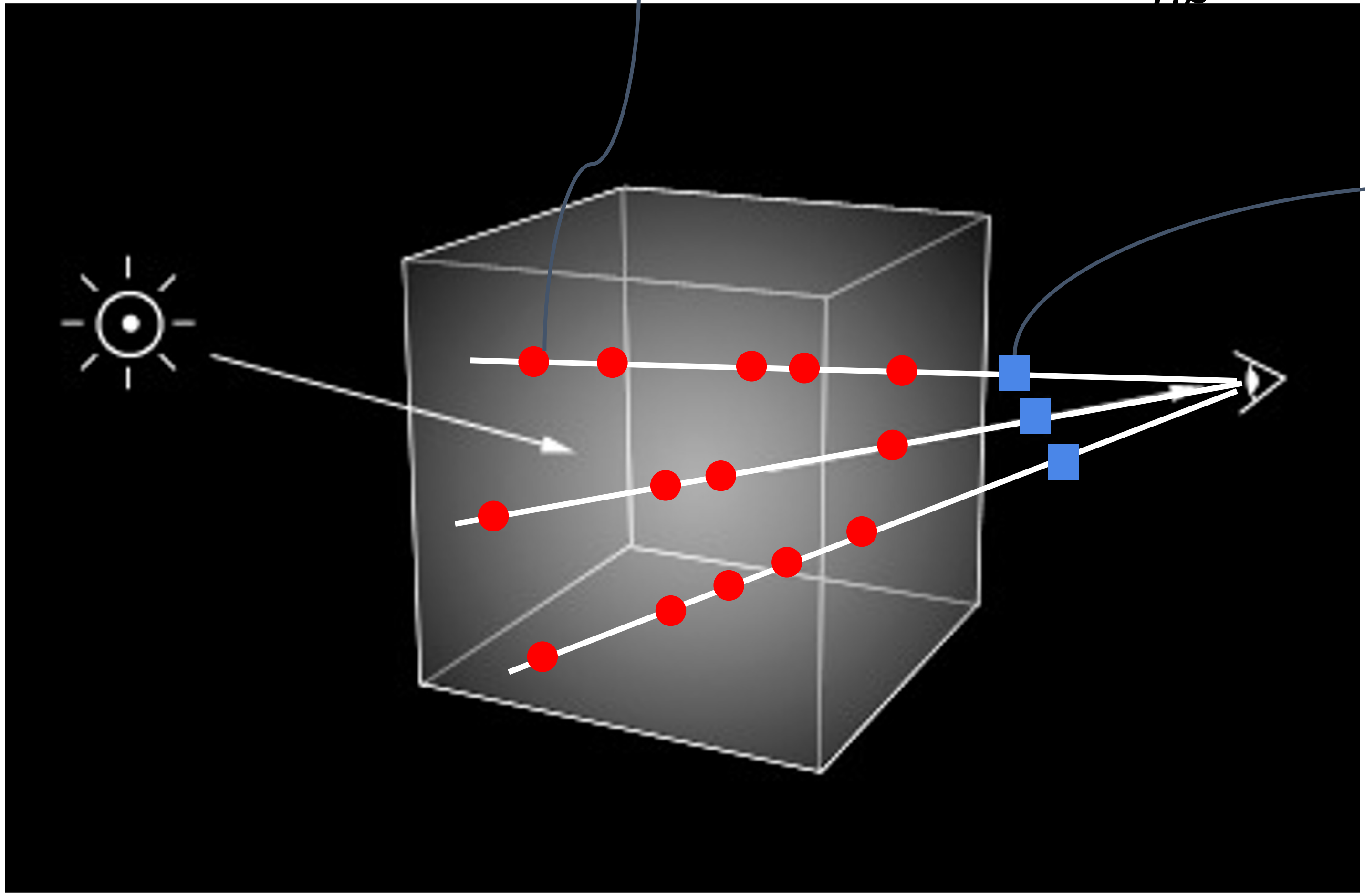


Volume Rendering: Ray Marching

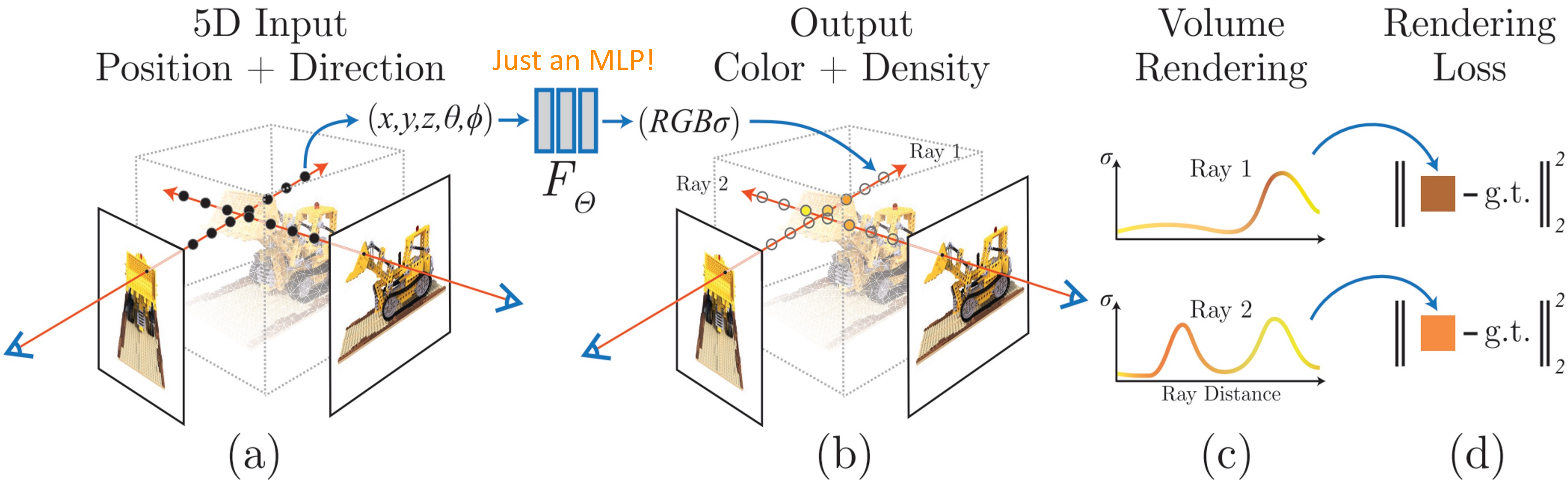
$$(x, y, z, \theta, \phi) \xrightarrow{F_\Theta} (r, g, b, \delta, \sigma)$$

Compositi
ng

$$(R, G, B)$$



Neural Radiance Field



Very slow to train & render!

Requires many tricks to render high-quality images

One model per scene

Instant NeRF

Müller et al., 2022

Frontiers of Deep Learning

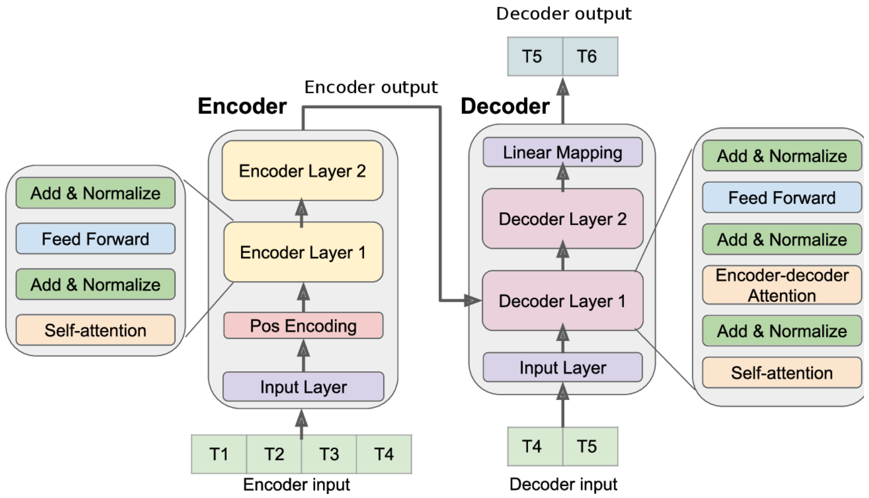
Topics we didn't get time to cover:

- Vision Transformers
- **Graph Neural Nets**
- Metric learning
- AutoML
- **3D perception & reconstruction**
- Memory modeling
- Few-shot / meta learning
- **Neural Radiance Field (NeRF) / implicit representations**
- Adversarial learning and robustness
- Continual / lifelong learning
- Visual reasoning
- Neural Theorem Proving
- Neural Program Induction / Synthesis
- MLSys
- Many topics in NLP …

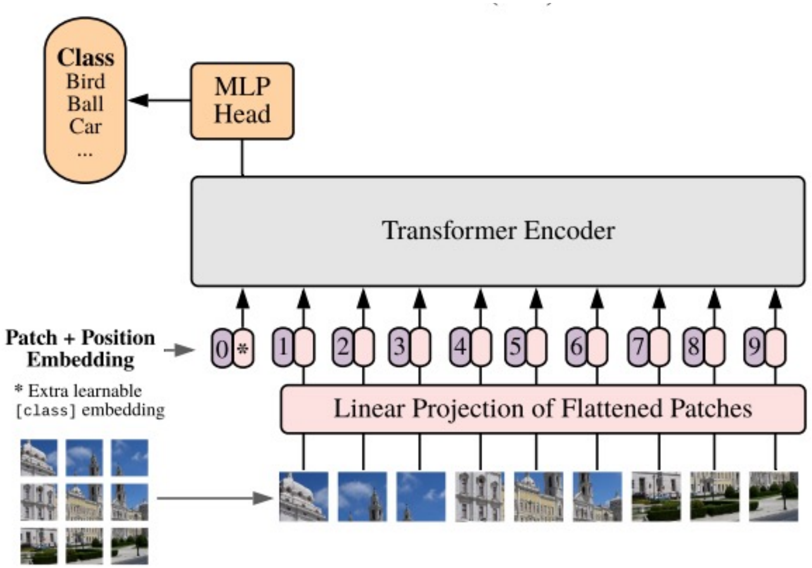
Homogenization of Deep Learning

Homogenization is the **consolidation** of methodologies for building machine learning systems across a wide range of applications.

Example: The Transformer Models (Vaswani *et al.*, 2017)



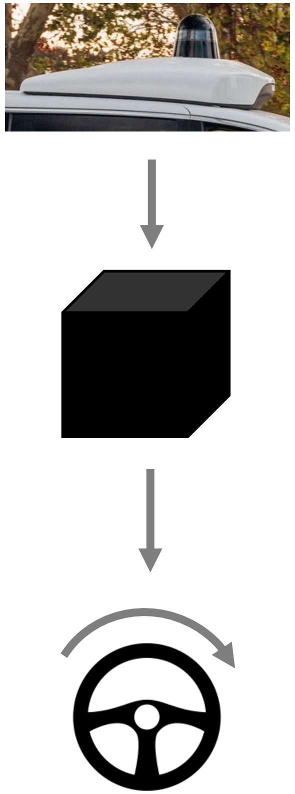
Transformer Models originally designed for NLP



Almost identical model (Visual Transformers) can be applied to Computer Vision tasks

Lack of interpretability

Why did the robot do that?



What have we learned this semester?

Deep Learning Fundamentals

Linear classification & kNNs
Loss functions
Optimization
Optimizers
Backpropagation
Computation Graph
Multi-layer Perceptrons

Neural Network Components and Architectures

Hardware & software
Convolutions
Convolution Neural Networks
Pooling
Activation functions
Batch normalization
Transfer learning
Data augmentation
Architecture design
RNN/LSTMs
Attention & Transformers

Applications & Learning Algorithms

Object Detection
Semantic & instance Segmentation
Reinforcement Learning
Large-language Models
Variational Autoencoders
Diffusion Models
Generative Adversarial Nets
Self-supervised Learning
Vision-Language Models
VLM for Robotics
Graph Neural Networks

Thank you!


Danfei Xu


Head TA: Mihir Bafna


Krishanu Agarwal


Manav Agrawal


Anshul Ahluwalia


Aditya Akula


Matthew Bronars


Will Held


Vikranth Keerthipati


Renzhi Wu


Wei Zhou