Advanced Topics in Computer Vision

Prof. Jeong Joon Park

EECS 542

Welcome!

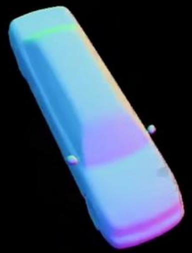
Introductory Concepts of Computer Vision

And Administrative Information

About the Instructor
- Jeong Joon Park

- 2023 – now: Assistant Professor at UMichigan, Computer Science and Engineering
- Before:
  - Postdoc at Stanford
  - PhD at University of Washington
  - B.S. at Caltech
- Research:
  - Computer vision & Computer graphics
  - 3D/4D Reconstruction and generative models

Estimating Scene Physical Properties

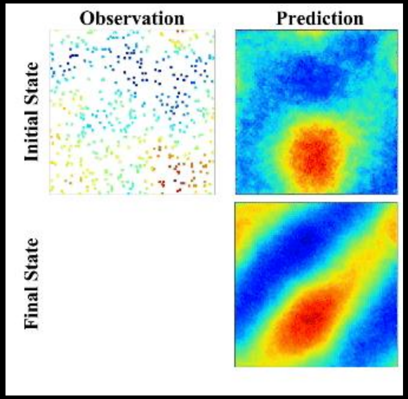


Neural 3D Representation

3D Generative Model

4D Generative Model

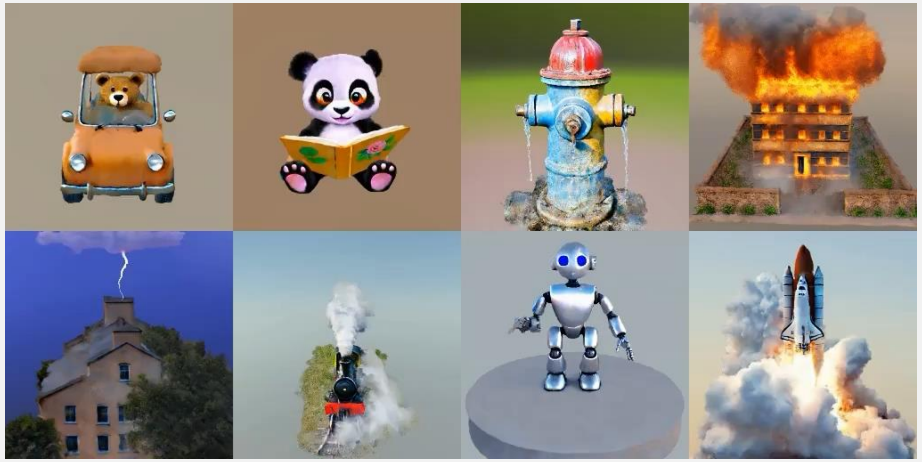
Novel-View Synthesis

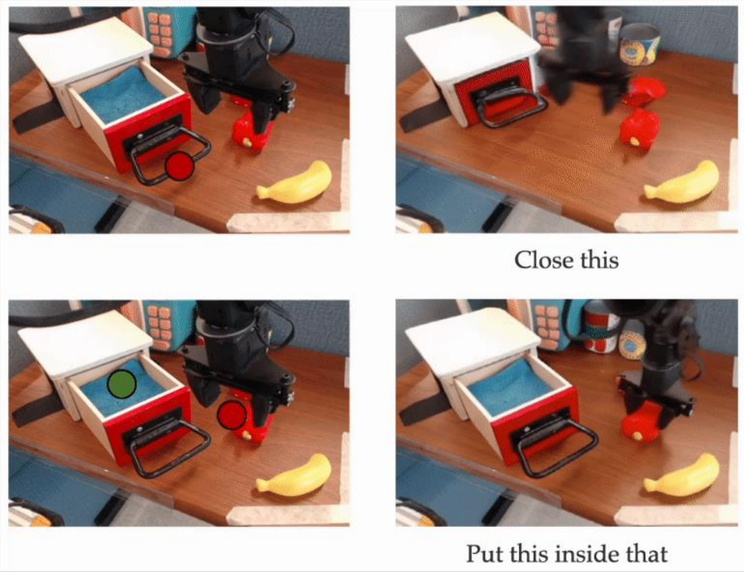


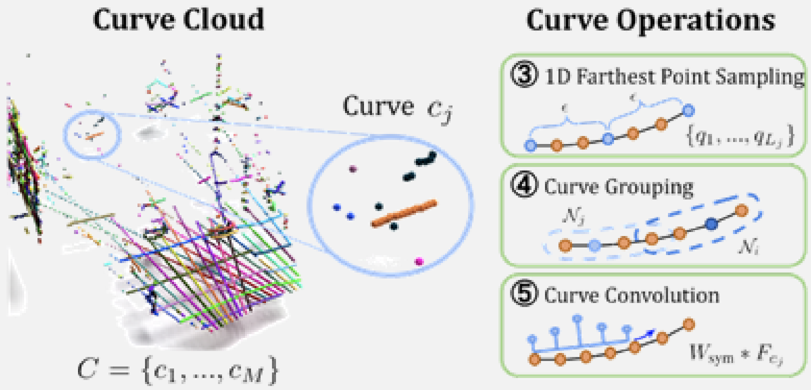
Generative Simulation

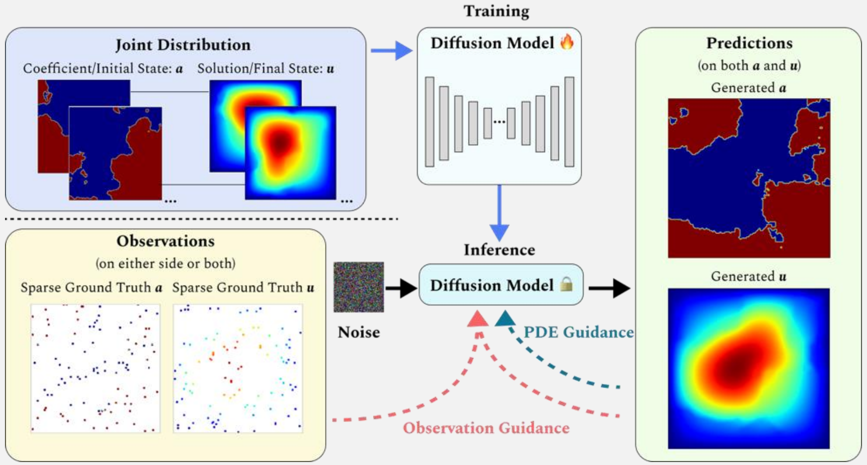
Jeong Joon Park – 3D Vision & Generative Models

Check my website for details: https://jjparkcv.github.io/









Goals of Computer Vision

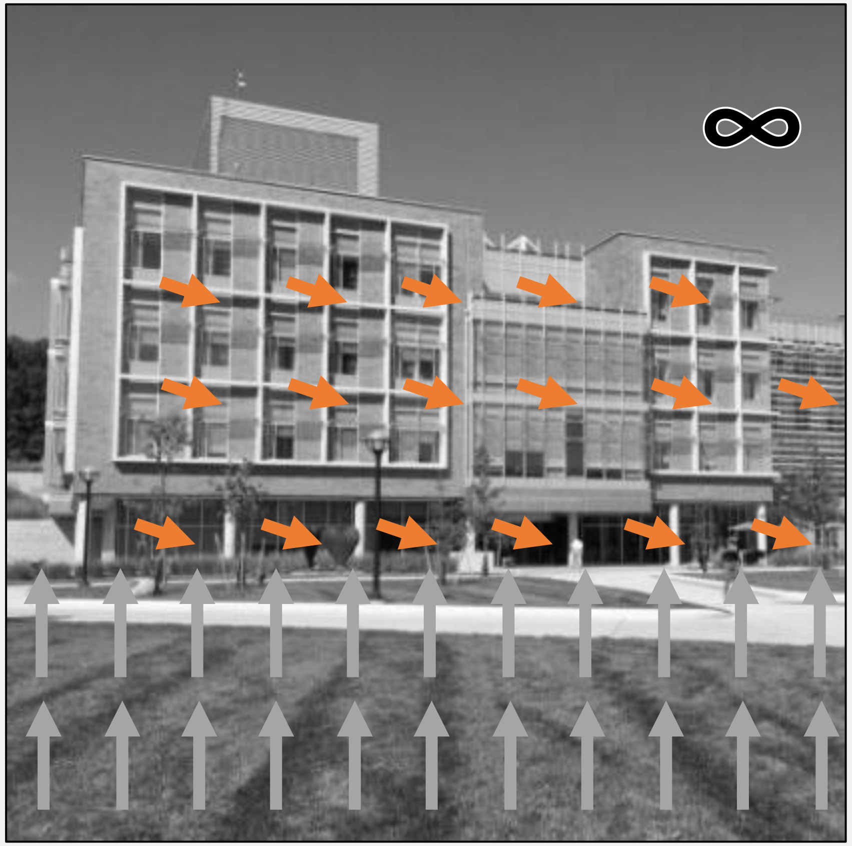
Get a computer to understand

Goal: Naming

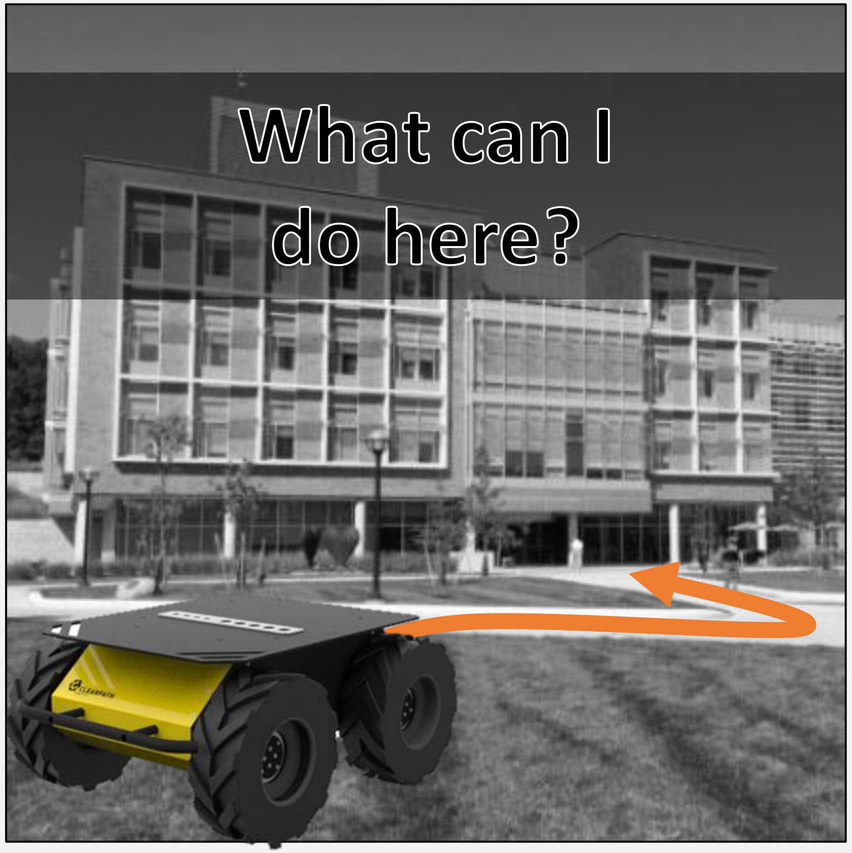
Goal: Naming

The picture shows a building with many windows and grass in front of it. There is a person walking on the right…

Goal: 3D



Goal: Actions



Seems Obvious, Right?

- **Key concept to keep in mind throughout the course**: you see with both your eyes **and** your brain.
- Which were developed over the course of evolution and your lifetime.

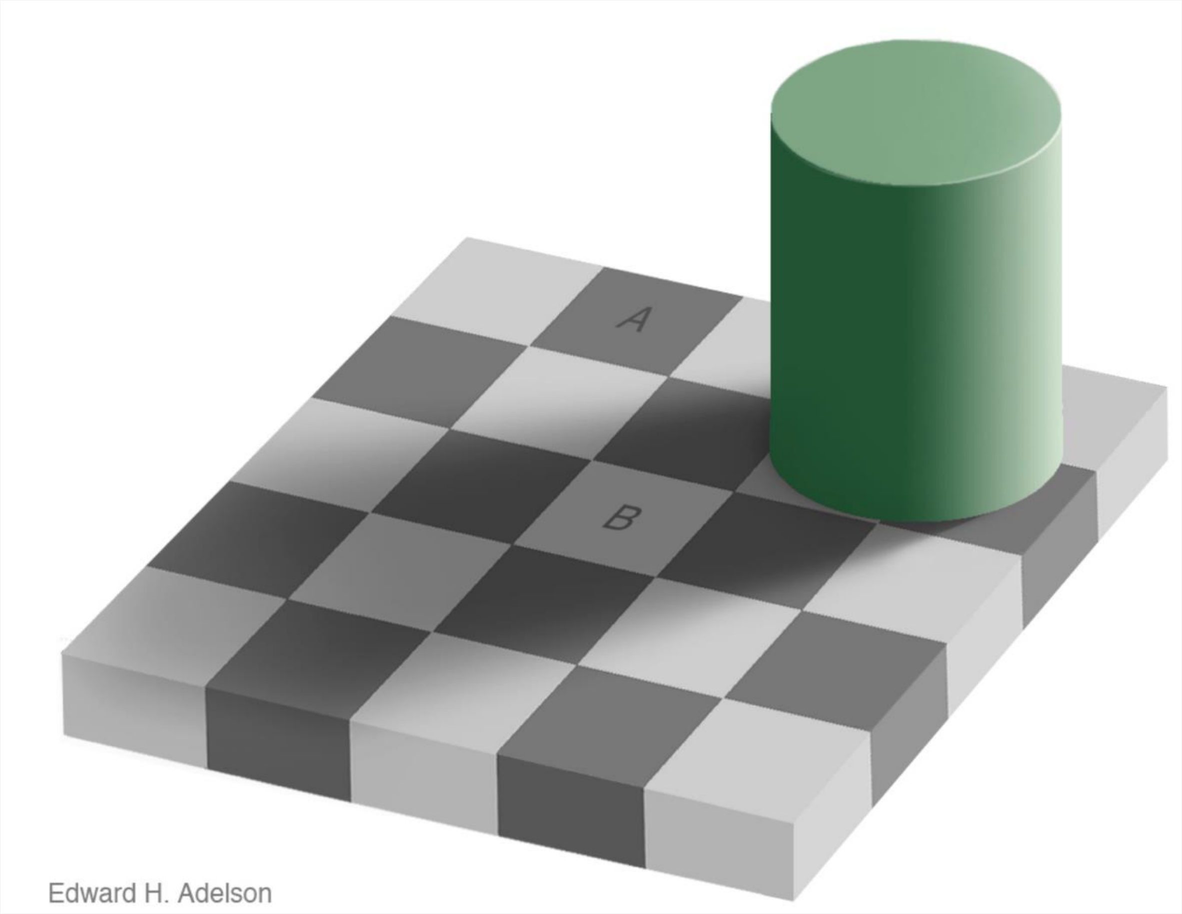
Why is it Hard?

097	097	097	097	097	097	097	097	096	097	097	096	096	096
100	100	100	100	100	100	101	101	102	101	100	100	100	099
105	105	105	105	105	105	105	103	102	102	101	103	104	105
109	109	109	109	109	110	107	118	145	132	120	112	106	103
113	113	113	112	112	113	110	129	160	160	164	162	157	151
118	117	118	123	119	118	112	125	142	134	135	139	139	175
123	121	125	162	166	157	149	153	160	151	150	146	137	168
127	127	125	168	147	117	139	135	126	147	147	149	156	160
133	130	150	179	145	132	160	134	150	150	111	145	126	121
138	134	179	185	141	090	166	117	120	153	111	153	114	126
144	151	188	178	159	154	172	147	159	170	147	185	105	122
152	157	184	183	142	127	141	133	137	141	131	147	144	147
130	147	185	180	139	131	154	121	140	147	107	147	120	128
035	102	194	175	149	140	179	128	146	168	096	163	101	125

We need to invent eyes and brains to interpret list of numbers!

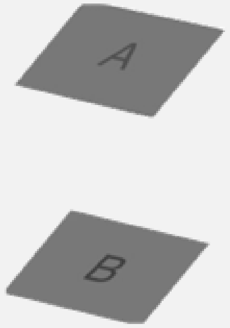
Finding patterns, concepts, and meanings of the world from these numbers!

To See: Perception vs. Measurement

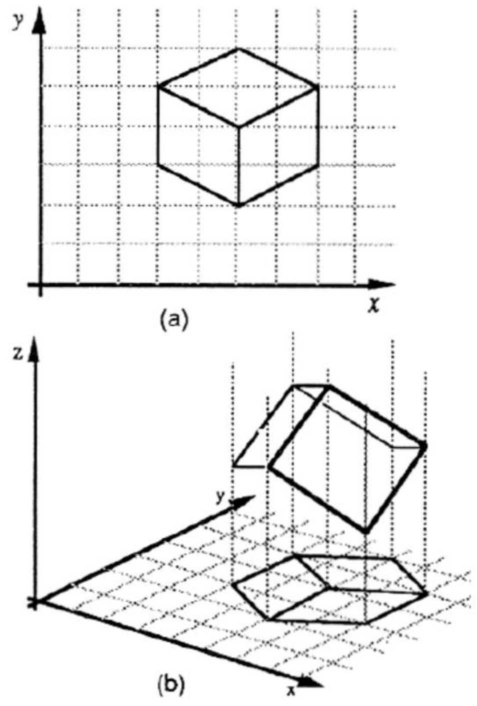


To See: Perception vs. Measurement

- We don’t just read the numbers!



Fundamental Ambiguities (2D)



[Sinha & Adelson, 1993]

Fundamental Ambiguities

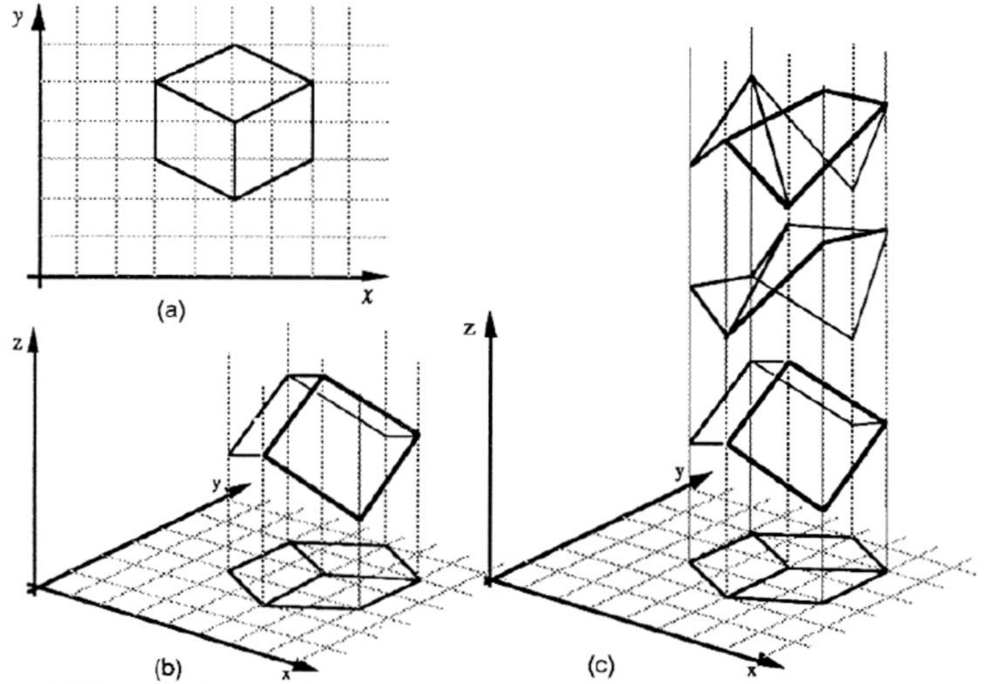


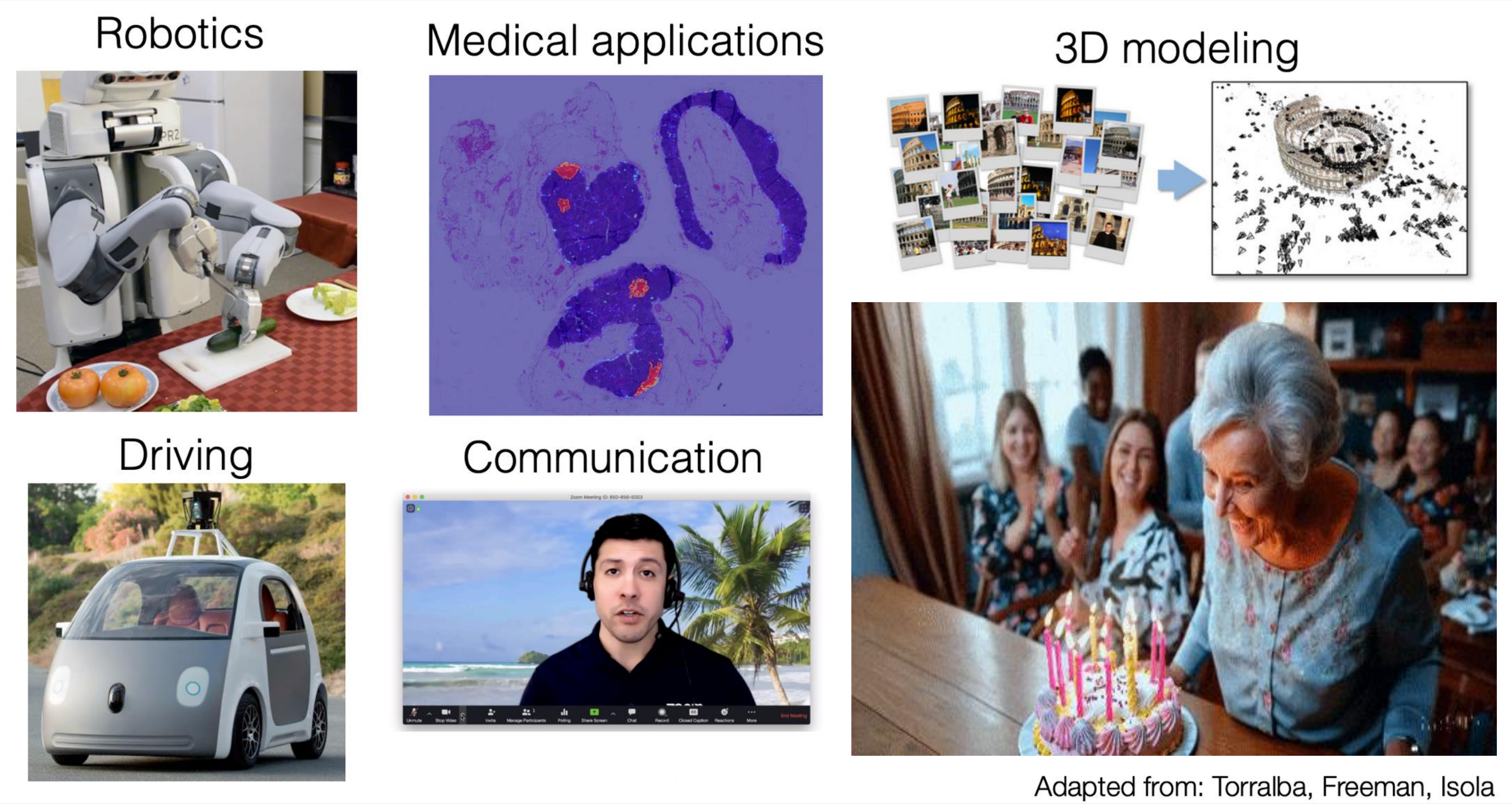
Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent with infinitely many 3D structures.

[Sinha & Adelson, 1993]

Despite This, We’ve Made Progress

- Few of these problems are **solved** *(and there are lots of dangers to pretending things are solved when they aren’t)*
- But we do have systems with performance ranging from non-embarrassing to super-human (with the right caveats)

Exciting Times for Computer Vision

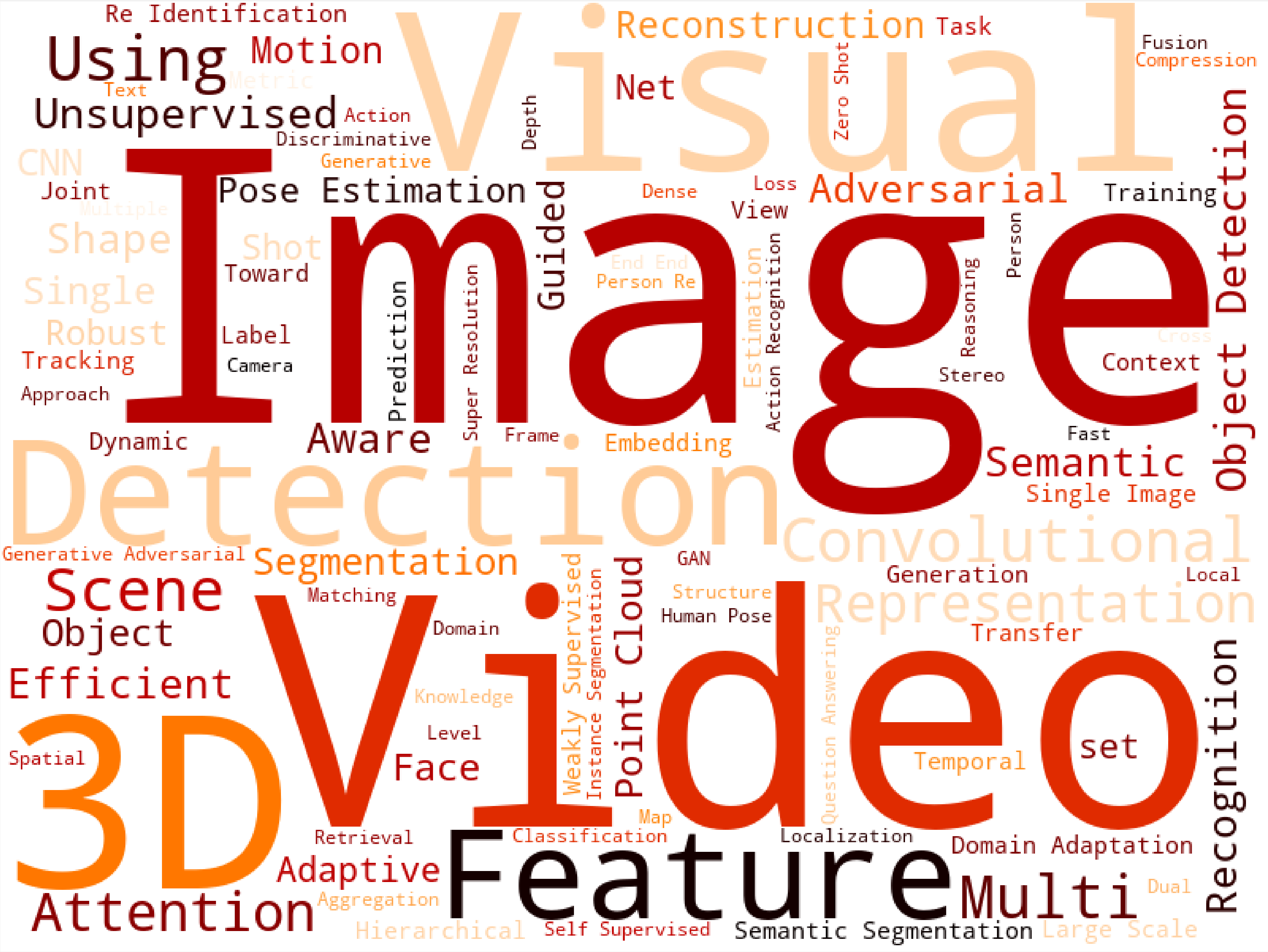


Content
Generation

What is Computer Vision?

Computer vision is a subfield of artificial intelligence (AI) that equips machines with the ability to process, analyze and interpret visual inputs such as images and videos. It uses machine learning to help computers and other systems derive meaningful information from visual data.

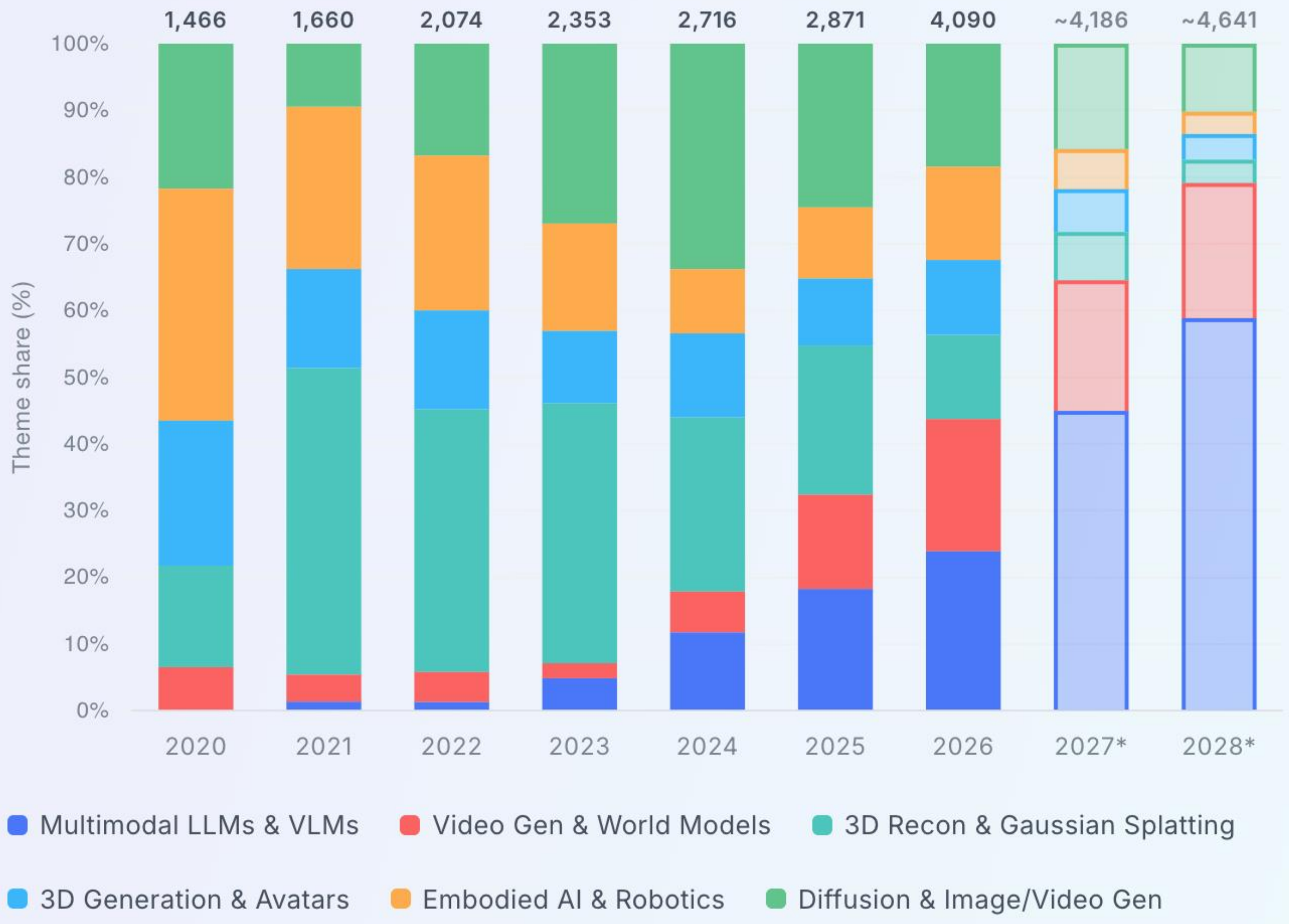
- Any ways of processing (typically) visual inputs
- Any ways of extracting meaningful information from continuous physical sensing data



CVPR 2019 Keywords

Research Trend Tracker

Theme share by year (2020–2028) · forecast by TimesFM

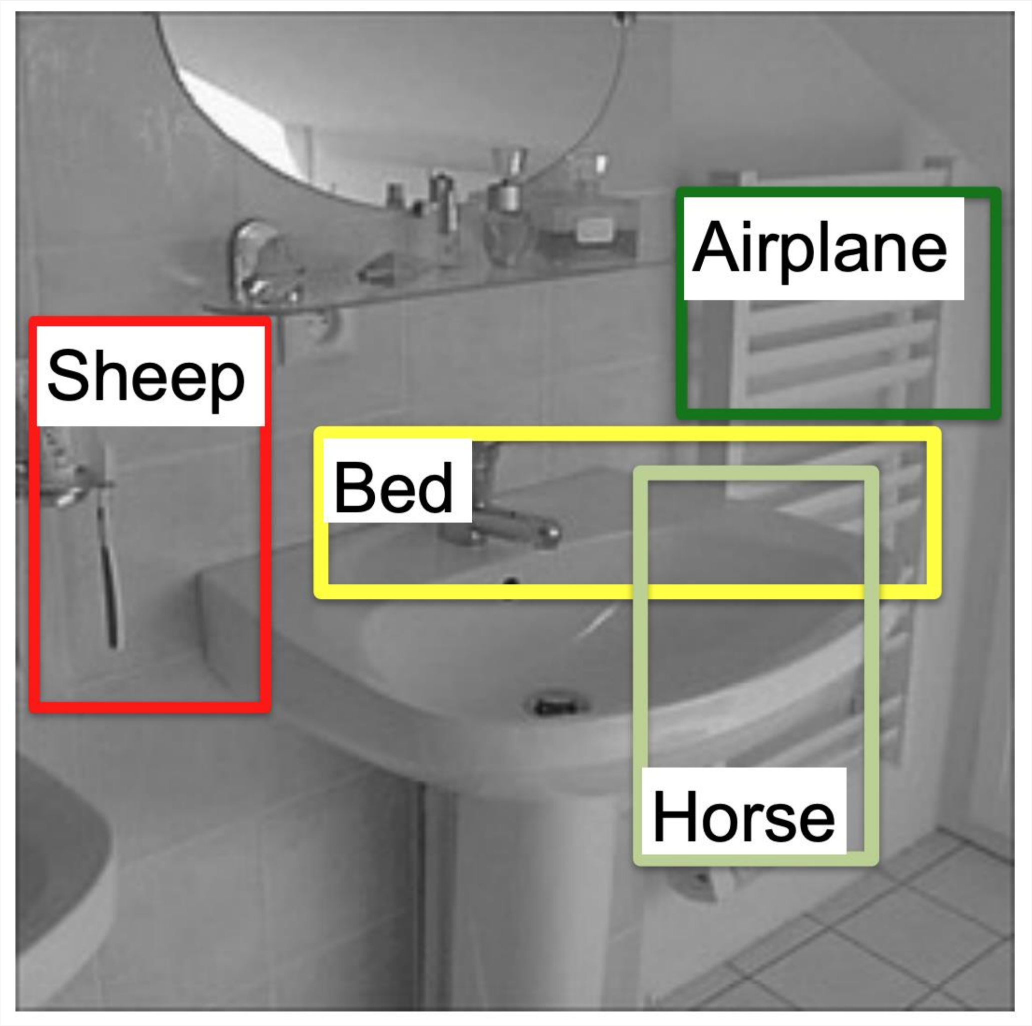


Dashed borders = TimesFM forecast · Numbers above bars = total accepted papers

Recognition & Perception

- Typically means: predicting semantic labels from image/videos
- **Semantic** means relating to the meaning of words, signs, or symbols in language.
- Object detection, segmentation, tracking, language + vision, etc

Object Detection Circa 2010



Object Recognition Now

Kirillov et al. Segment Anything, 2023

Visi

- Perc Vision Language Models (VLMs)

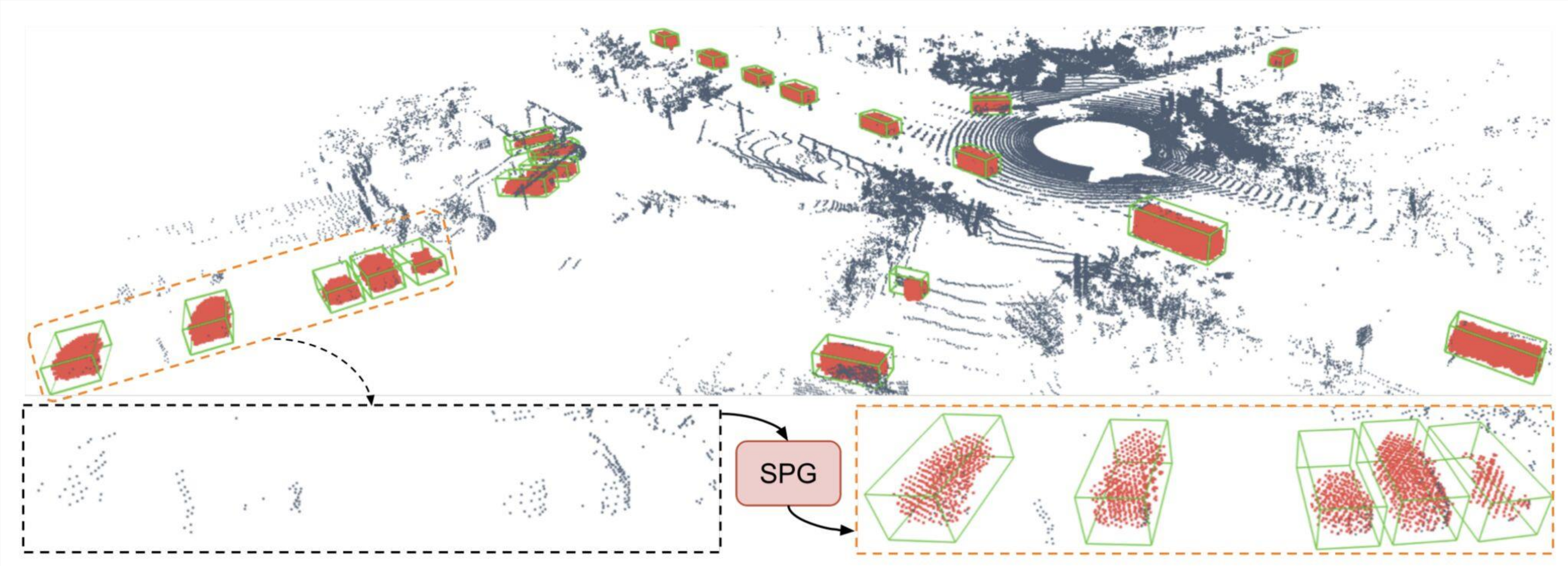
ChatGPT Prompt
“Give me the bounding box coordinates for vehicles on the non-congested side”

3D Reconstruction

Barron, et al., 2023

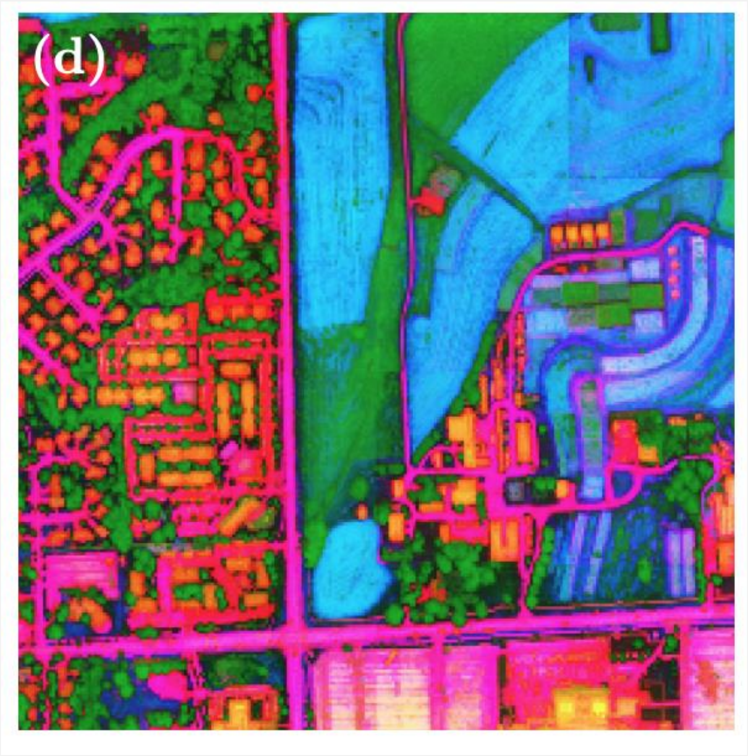
3D Perception

- Perception on 3D visual data, e.g., point clouds
- Very important for Robotics and Self-driving



Representation Learning

- Automatically learning features using deep learning
- Self-supervised learning: Vision training without training labels

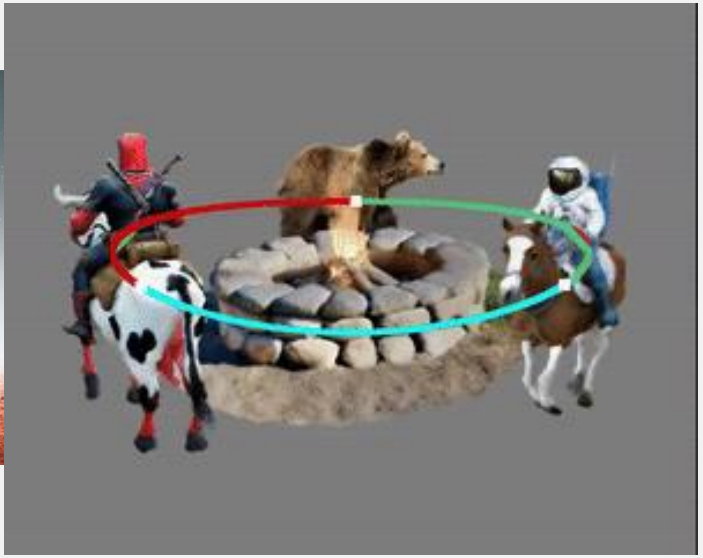


Generative Models

- Automatically generate realistic imagery using AI
- Long and rich history
  - VAEs, GANs, Diffusion Models, Flow-based Models, etc
- 3D/4D Generation is blooming

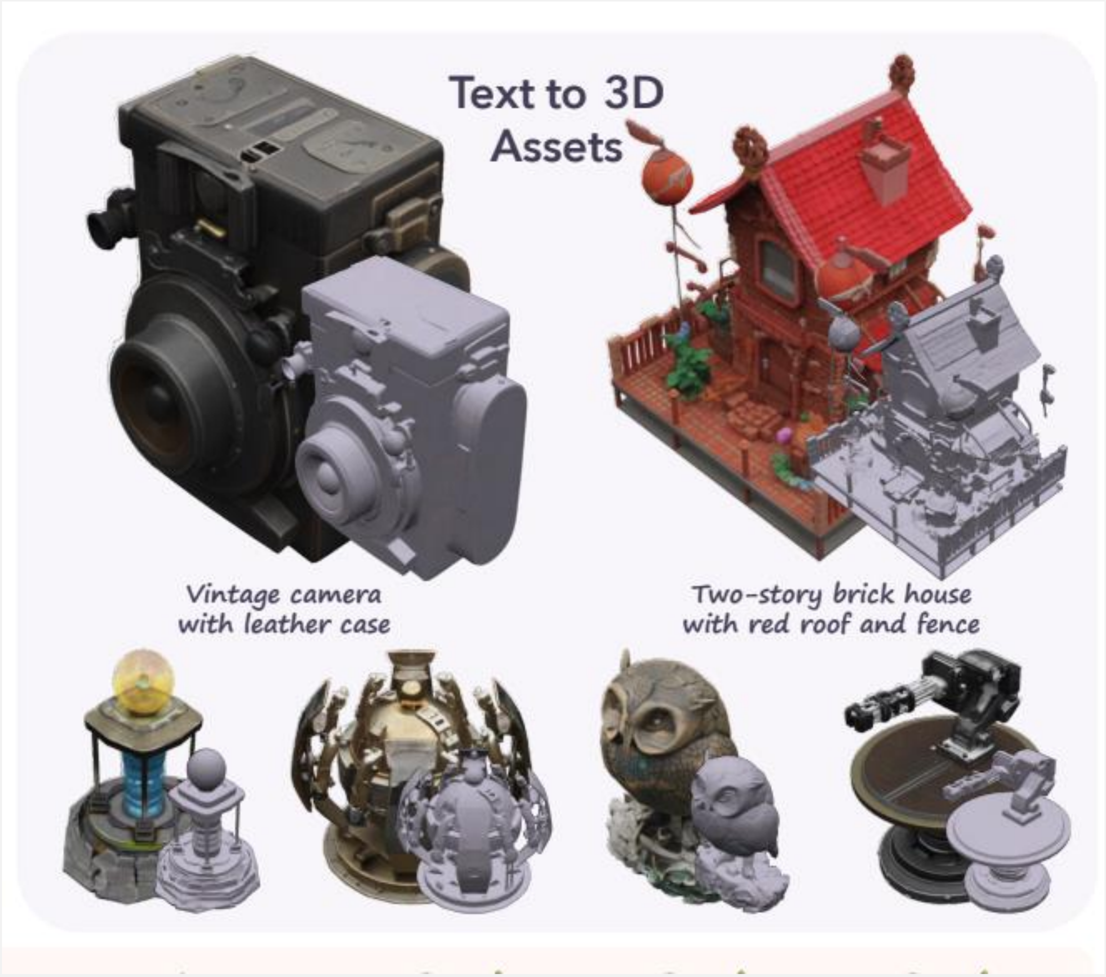
Goodfellow et al. 2024

Stable Diffusion 3. 2024



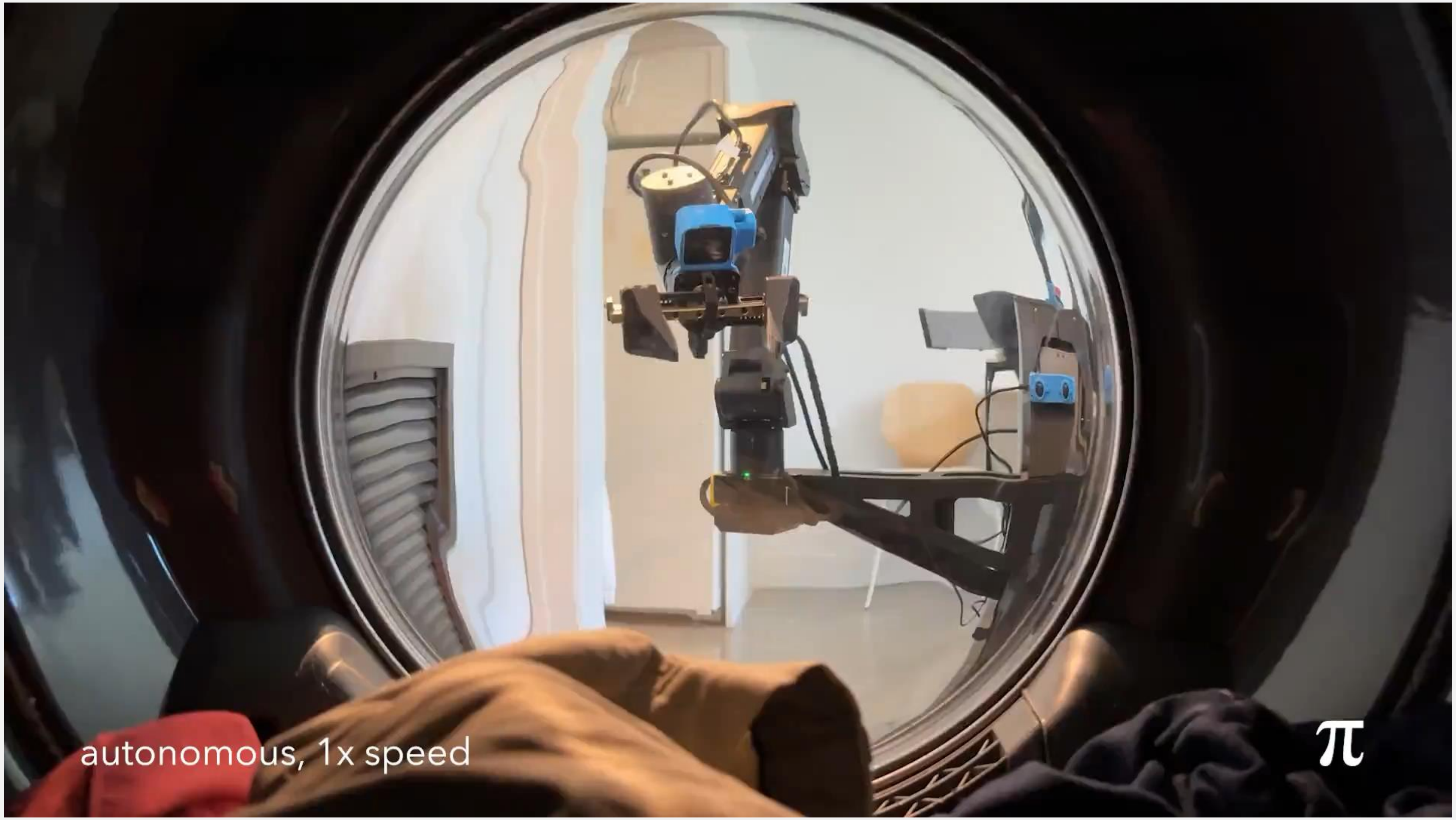
Bahmani et al. 2024

3D Generative Models



Trellis

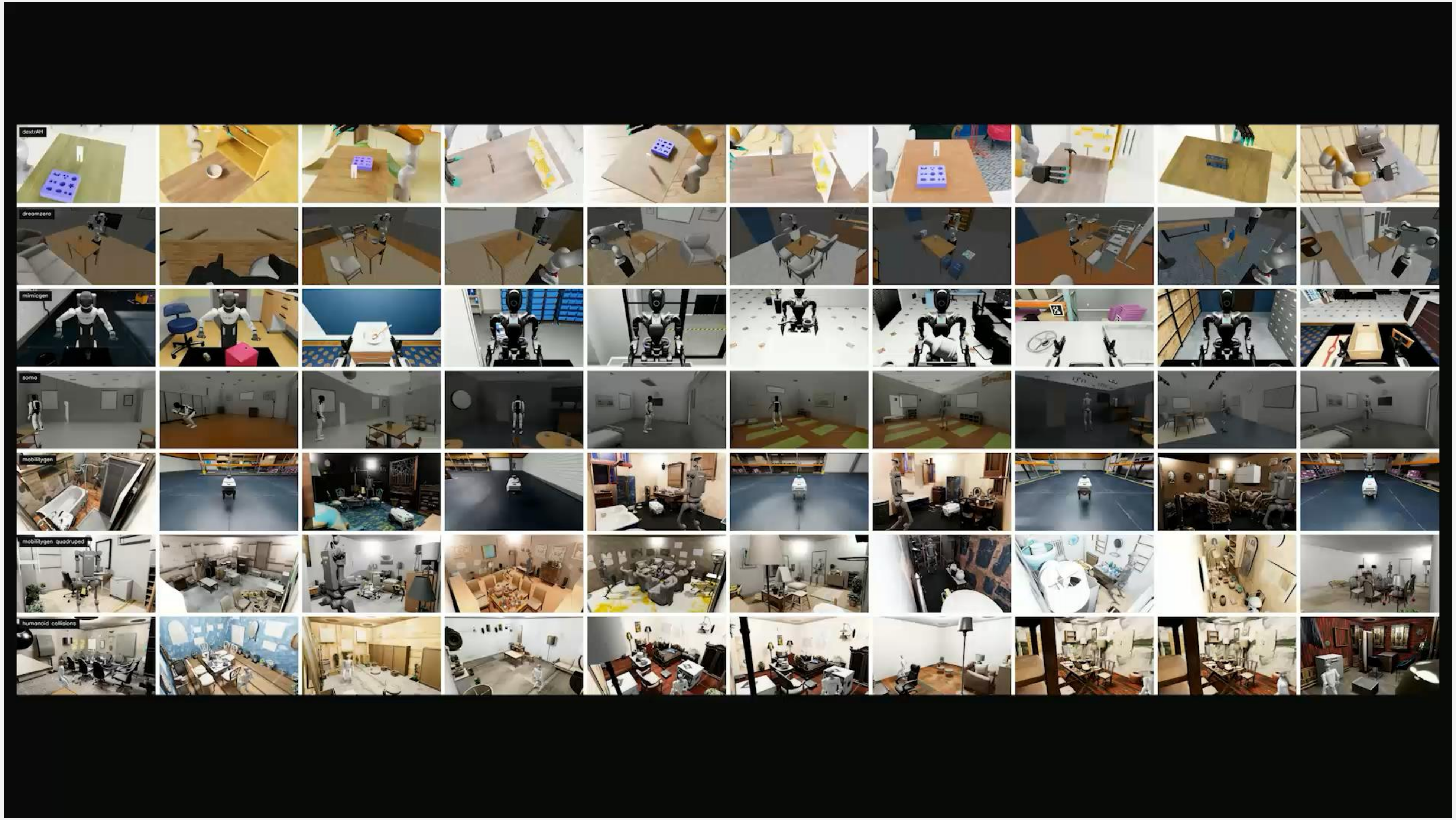
Robot Vision



pi0

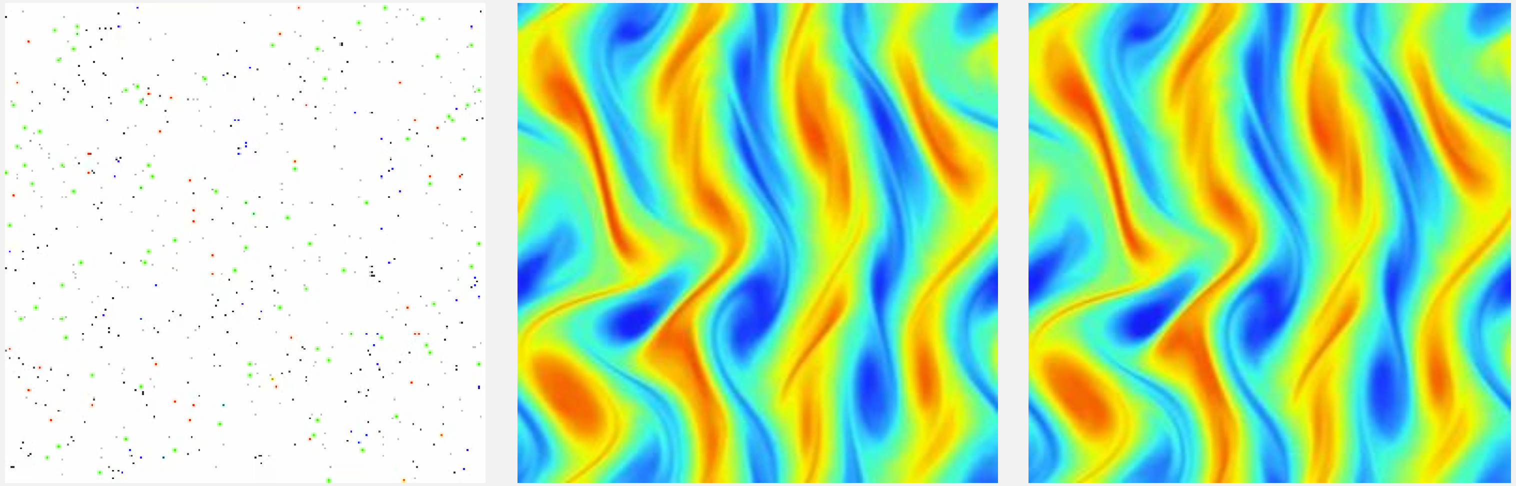
World Models

- Predict the environment's future state



Physics-Aware Machine Learning

- Instead of natural images (pixel + time or 3D points), model physical quantities.



Input Video | Our reconstruction (VideoPDE) | Ground truth

This course is NOT about

- Deep learning course (PyTorch)
- Instead:
  - Familiarize with general literature of computer vision
  - Both classic papers and recent breakthroughs
  - Prepare you for cutting-edge industry applications and academic research
  - Apply to relevant areas

EECS 542 · ADVANCED TOPICS IN COMPUTER VISION

Course Format & Logistics

Fall 2026 · Mon & Wed, 10:30 AM–12:00 PM · 2505 GGBL

Instructor: Jeong Joon Park | GSIs: Yunhao Luo, Hyelin Nam

OVERVIEW

What we'll cover

1

Lecture Format

80 minutes, two papers per class

2

Paper Reviews

CVPR-style, due before class

3

Paper Presentations

Teams of 2–3, 20–25 min

4

Discussion

10 min in small groups

5

Final Project

50% of your grade

6

Grading & Policies

Piazza, email, deadlines

General Information

- Course Website:
  - https://um-eecs542-fa2026.github.io/
  - All course information and links provided here
  - Syllabus, slides, reading materials, etc.
- Course email:
  park-eecs542@umich.edu
  - Send all emails to here (don't use individual emails!)
- GradeScope:
  - https://www.gradescope.com/courses/1356343
  - Piazza:
  - https://piazza.com/class/msyw1hcgw4l29y
  - Has all the information you need on website

Course Information

- Teaching Assistants
  - Hyelin Nam
    - Expertise in computer vision & video diffusion models
  - Yunhao Luo
    - Expertise in computer vision & robot learning

Each lecture is ~80 minutes

80

minutes per lecture

2 papers · every class

The goal is not only to understand the technical content of each paper.

It is also to practice:

1. **Critically reading research papers**
2. **Presenting technical work to peers**
3. **Discussing and debating open questions**

A typical lecture, from Lecture 3 on

Paper 1	**Presentation** 20–25 min · last 3–5 min is the critique	**Discussion** 10 min · small groups
Paper 2	**Presentation** 20–25 min · last 3–5 min is the critique	**Discussion** 10 min · small groups

Before class: all students read both assigned papers — every week, whether or not you are presenting.
Remaining <10 minutes: project presentations | questions to the instructor and the TAs.

Project check-in presentations

Oct 21 & Oct 26

Two dedicated project presentation days. Most teams present during these two class sessions.

~1 team per lecture

Remaining teams present after class, roughly one team per lecture across the 2–3 weeks before and after the dedicated days.

Format

7 min presentation + **3 min Q&A and feedback**

A tentative schedule is on the project sign-up sheet.

Submit a review for every assigned paper

1

Before class

Reviews are due before the lecture in which the paper is discussed.

2

On Gradescope

All reviews are submitted through Gradescope.

3

Yes, even your own paper

Submit a review for the paper you are presenting too.

Reviews + active discussion are worth **23%** of your final grade.

CVPR-style review format

Written sections

- Summary
- Strengths
- Weaknesses — major and minor

Write the review you would want to receive: specific, evidence-based, and fair to what the paper actually claims.

Ratings

Rate 1–5 (Poor, Fair, Good, Very good, Excellent) with justification:

Soundness

Presentation

Contribution

No strict character or page limit — write as much as the paper warrants, but you can't be too short

For presenters

2–3

students per presentation team

20–25

minutes of presentation time

3–5

discussion questions for the class

Signing up

Choose a paper and a presentation slot on the presentation sign-up sheet (linked from the course website). Sign up early — slots are first come, first served.

What to cover

1. **Motivation & problem setting**

   What task does the paper address, and why was prior work insufficient?

2. **Technical approach**

   The proposed method, model or architecture, loss functions, training procedure, and key algorithmic details.

3. **Experiments**

   Datasets, baselines, evaluation metrics, main results, and ablations.

Give enough background for the class to follow the method — but keep the focus on the paper's own technical contributions.

The critique — last 3–5 minutes

Your own take on the paper, framed as food for thoughts.

Close your presentation with 3–5 discussion questions. These questions drive the 10-minute discussion that follows.

Good questions

Open-ended and debatable — they invite disagreement.

Weak questions

Factual recall — answered by rereading one line of the paper.

Slide submission policy

Submit your slides through the course Google Form both before and after your presentation.

Pre-presentation draft

Monday talks	**due Friday EOD**
Wednesday talks	**due Sunday EOD**

- Need not be polished — but must show the structure and flow.
- You can keep editing after this submission.
- Teaching staff may send feedback if needed.

Final version

Due by the end of your presentation day.

5%

of the course grade is the pre- and post-presentation slide submission alone.

The 10-minute discussion

1

Presenters pose 3–5 questions

at the end of their talk

→

2

Break into groups of 2–6

and discuss the questions

→

3

Submit a summary to Gradescope

one submission per group

Everyone participates — discussion counts toward the 23% review and participation component.

The largest component of the course

50%

of your final grade

3–5

students per team

GPU

Great Lakes computing credits provided

Forming a team

- Use Piazza to pitch project ideas and find teammates.
- Register your team on the project sign-up sheet and list every member.
- Hands-on research in computer vision or a closely related area.

Milestones and deadlines

Oct 2

Written proposal

Problem & motivation, related work, direction. Written only — no oral presentation.

Oct 21 & 26

Progress presentation

7 min talk + 3 min Q&A on method, progress, setup and preliminary results.

Dec 7

Poster session

Final poster presentation of the completed project.

Dec 9

Report due

CVPR-style final report, including a per-member contribution statement.

Any dates not yet fixed will be announced in class as they approach.

Progress talk

Progress presentation

7 min talk + 3 min Q&A

- Problem, motivation and related work
- Method
- Progress so far
- Experimental setup, baselines, metrics
- Preliminary results — bring as much as you can
- Remaining work and next steps

final report

Final report

CVPR-style paper format

- Abstract
- Introduction
- Related Work
- Method
- Experiments
- Conclusion and Limitations

final report

Collaboration

- At the end of the report, include a brief **contribution statement** indicating the specific component(s) of the project led by each team member.
- Every team member should be responsible for at least one specific and meaningful part of the project.

Grading breakdown

- **50%** — **Final Project**
- **25%** — **Class Presentation**
  5% slide submission · 20% in-class talk
- **23%** — **Paper Reviews + Active Discussion**
- **2%** — **Course Feedback**

Piazza

Piazza is our forum for conceptual and technical discussion among classmates. It is moderated by the teaching staff.

1. **There is no anonymity**

   Your name is attached to every post. Be respectful to your classmates.

2. **Post in the right folder**

   Every paper we read has its own dedicated discussion thread.

3. **Search before you post**

   Check for an existing answer — clogging Piazza makes it less usable for everybody.

4. **Prefer Piazza over email**

   Public questions and answers benefit the whole class.

Email

park-eecs542@umich.edu

Use Piazza first. If you must email, use this course address.

Emails to individual teaching staff members will be ignored.

This keeps requests in one queue so nothing gets lost.

BEFORE YOU LEAVE

Three things to do this week

1. **Join Piazza & Gradescope**
   Both are linked from the course website.

2. **Sign up for a paper slot**
   Teams of 2–3 on the presentation sign-up sheet.

3. **Start forming a project team**
   3–5 students. Proposal is due Oct 2.

Key dates

Oct 2	Proposal
Oct 21 & 26	Progress talks
Dec 7	Poster session
Dec 9	Report due

Questions? Ask on Piazza.

In-Class Discussion

- Q1: What are you most excited about the applications of computer vision?
- Q2: Which machine learning aspects of computer vision are you most excited about?

Go to https://um-eecs542-fa2026.github.io/

Log-in to Gradescope

Fill out the discussion questions

Ask the TAs if needed