3D Vision and Machine Perception

Prof. Kyungdon Joo

3D Vision & Robotics Lab.

AI Graduate School (AIGS) & Computer Science and Engineering (CSE)

This lecture is jointly developed with **Kyungdon Joo (UNIST)** and **Tae-Hyun Oh (POSTECH).**
Some materials, figures, and slides (used for this course) are from textbooks, published papers, and other open lectures

Logistics

- **Instructor**: Kyungdon Joo (kyungdon@unist.ac.kr)
  - Office: #801-10, E106 (제3공학관 106동 801-10호)
  - Office hours: Monday/Wednesday, 2:30-4:00 PM (send an email to avoid conflicts)

- **TA: Three TAs**
  - Dongjun Gu (구동준; djku1020@unist.ac.kr)
  - Junsu Kim (김준수; joonsu0109@unist.ac.kr)
  - Hyungyu Park (박현규; hyungyu@unist.ac.kr)

- **Course page**
  - Main platform: Blackboard
  - Lecture material, program assignments, announcements, etc.

Logistics (cont.)

- **Classroom:** 106-T201
  - Unless mentioned, this lecture will be an offline lecture
- **Class hours**:
  - Monday & Wednesday 1:00 PM~2:15 PM
- **No exam**:
  - The mid-term/final exam will be replaced with the PAs and projects

Logistics (cont.)

- **Grading**
  - Attendance (5%) + PAs (50%) + Final project with a presentation (45%)

- **PAs and Projects**
  - Submitted reports should be compiled in a PDF format from the CVPR **latex template**
    - https://github.com/cvpr-org/author-kit/releases/tag/CVPR2024-v2
    - Author Guidelines > Paper formatting (using overleaf)
    - If this does not hold, your report may not be graded at all

  - All the PAs and projects will be conducted in Python
    - Even if PAs are not related to deep learning,
      it is required to use PyTorch as a matrix computation library.

Logistics (cont.)

- Late policy
  - A late PA (or project) will be penalized by 30% for each day
  - ONE ticket for PA (NOT final project)– A ticket allows a one-day delay

- Policy
  - All the PAs will be done SOLO / A term project will be done by a TEAM
  - Please do not leave any code public on GitHub (or the like) at the end of the semester!
  - We reserve the right to run an automated code copying service on submitted code

Warning! One-strike Out policy
- Please comply with the academic honor code
- Be honest

Logistics (cont.)

- **Target**: Graduate students in AIGS and Dept. of CSE, but proficient in Python

- **Prerequisites**
  - **I assume that you know**
  - Linear algebra
  - Calculus
  - Probability & statistics or probability theory
  - Programming (Python)
  - Signal processing, digital signal processing
  - Machine learning & deep learning

Course overview (tentative)

* The contents are subject to change slightly.

- NO lecture on holidays

Week	Topic
1	Intro
2	Imaging Pipeline
3	Feature and Matching
4	Image Transform
5	Camera Models
6	Camera Models II
7	Epipolar Geometry
8	Project Proposal
9	Stereo Systems
10	Dense Correspondence
11	Structure from Motion
12	Multi-view Stereo
13	3D Scene Understanding
14	Guest Lecture
15	Final Project Presentation I
16	Final Project Presentation II

Textbook

- Multiple View Geometry in Computer Vision

  Hartley and Zisserman, Cambridge University Press

- Computer Vision: Algorithms and Applications, 2nd ed.

  R. Szeliski, Springer (online draft: https://szeliski.org/Book)

- Machine Learning: A Probabilistic Perspective

  Murphy, MIT Press

- Pattern Recognition and Machine Learning

  Bishop, Springer

- Or ask keywords (you are interested in) to Google God! (or chatGPT…?)

What is “Perception”?

Why is Perception Important?

- It's (input, output) data

Why is Perception Important?

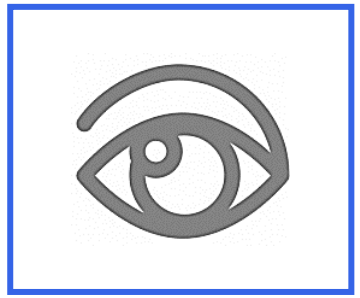
- It's (input, output) data
- As humans grow, we learn the world by interacting with it

Why is Perception Important?

- It's (input, output) data
- As humans grow, we learn the world by interacting with it
- Gather informative signal from multi-modal association like human

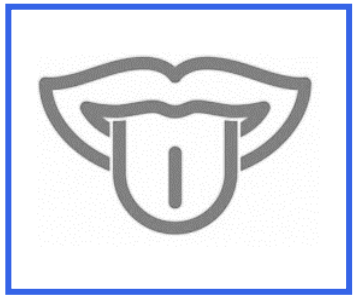
Why is Perception Important?

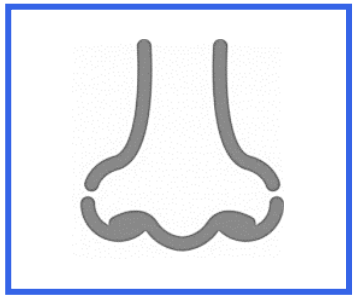
- It's (input, output) data
- As humans grow, we learn the world by interacting with it
- Gather informative signal from multi-modal association like human


Sight


Sound

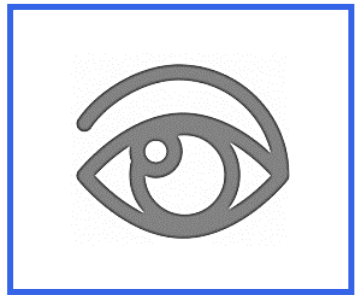

Touch


Taste


Smell

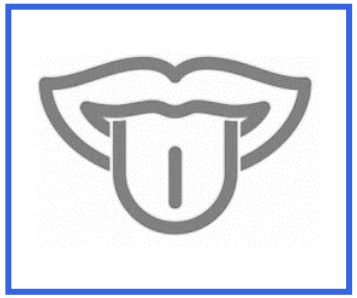
Why is Perception Important?

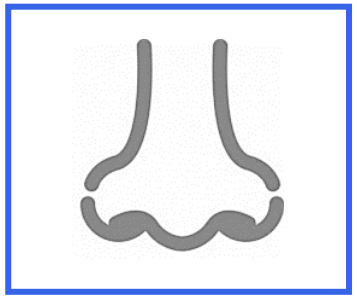
- It's (input, output) data
- As humans grow, we learn the world by interacting with it
- Gather informative signal from multi-modal association like human


Sight


Sound


Touch


Taste


Smell


Social

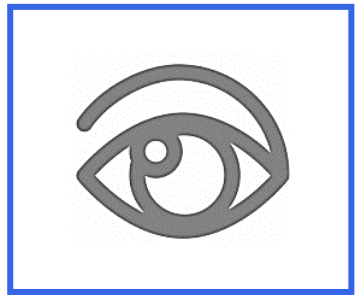

Face


Touch


Speech

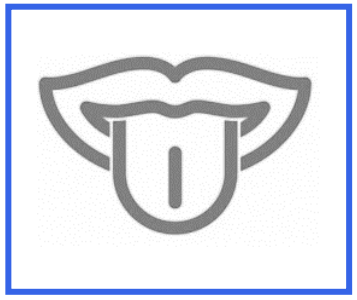
Why is Perception Important?

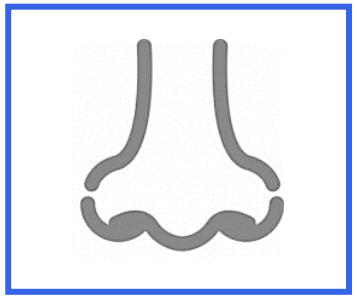
- It's (input, output) data
- As humans grow, we learn the world by interacting with it
- Gather informative signal from multi-modal association like human
- Developing machine perception is still open research area


Sight


Sound


Touch


Taste


Smell


Social


Face


Touch


Speech

Why is Visual Perception Important?

- Cambrian explosion
  - The most important evolutionary event in the history of life on Earth
  - Placed about 542 million years ago
  - A unique event when all major animal categories started appearing in the fossil record

Image from Wikipedia (Cambrian explosion). Source comes from Standford CS231

Why is Visual Perception Important?

- Cambrian explosion
  - The most important evolutionary event in the history of life on Earth
  - Placed about 542 million years ago
  - A unique event when all major animal categories started appearing in the fossil record

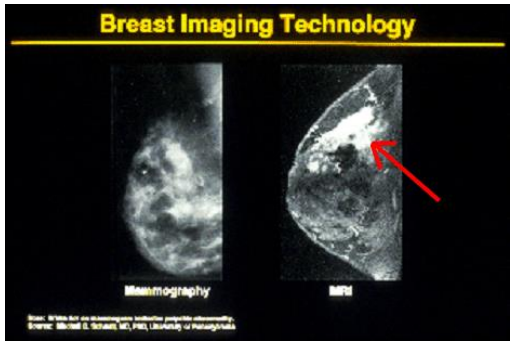
- One of possible cause (hypothesis): Evolution of eyesight [A. Parker]
  - Predator-prey relationships changed dramatically

Image from Wikipedia (Cambrian explosion). Source comes from Standford CS231

Why is Visual Perception (Computer Vision) Important?

- Visual information is the most dominant information source to humans
  - A picture is worth a thousand words
  - More than 50% of brain is devoted to processing visual info. -- William G. Allyn

Safety

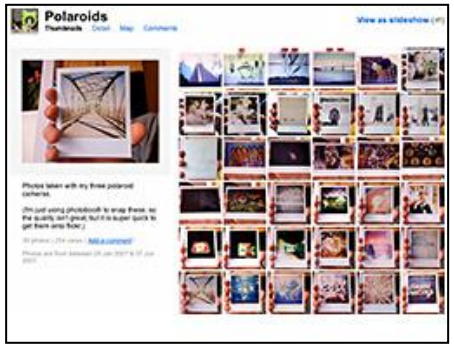


Health

Security

Comfort

Fun



Access

Image courtesy of James Hays

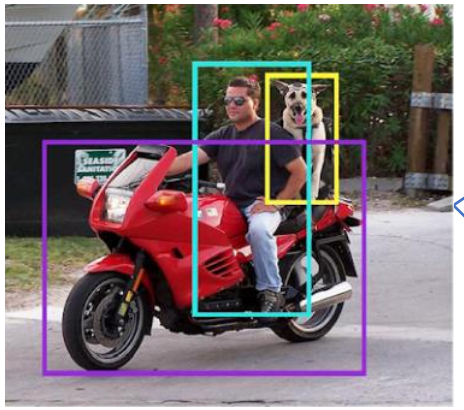
Visual Perception (Computer Vision)

- Visual perception & intelligence
  - Input: visual data (image or video)

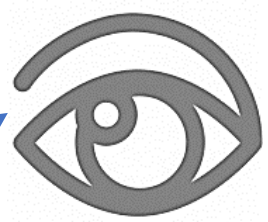
- Visual perceptions
  - Color perception
  - Motion perception
  - 3D perception
  - Semantic object-level perception
  - Social perception (Emotion perception)
  - Etc.

Computer Vision

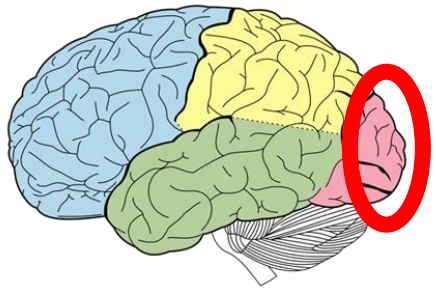
Visual world



Sensing device



Interpreting Device



Visual Cortex

Interpretation

The example image from [Kim et al., Dense Relational Captioning, CVPR 2019]
The other images are under CC-license

Computer Vision

Larry Roberts (Father of ARPANET),
1st Computer Vision thesis, 1963

MACHINE PERCEPTION OF THREE-DIMENSIONAL SOLIDS

by

LAWRENCE GILMAN ROBERTS

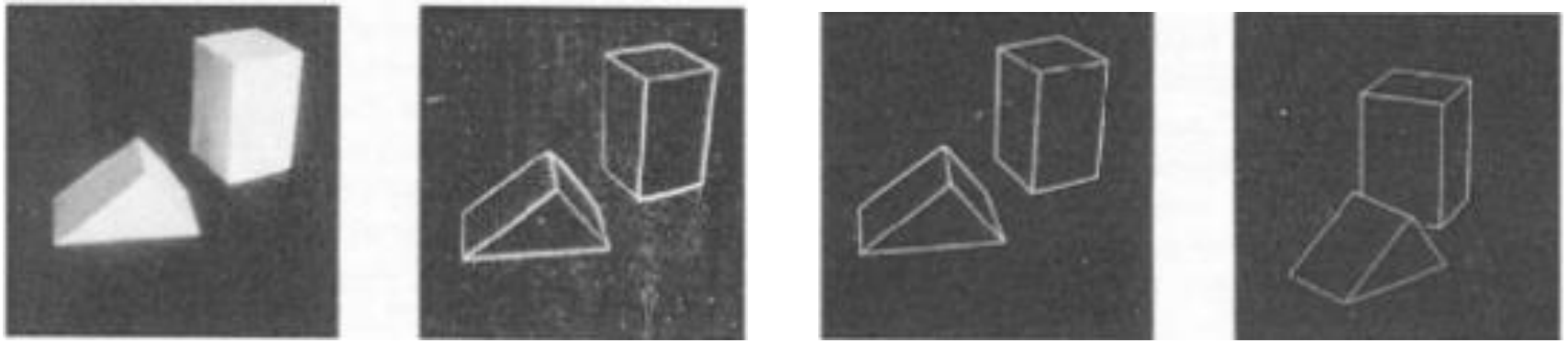
S.B., Massachusetts Institute of Technology
(1961)

M.S., Massachusetts Institute of Technology
(1961)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June, 1963



Computer Vision

- Difference with image processing
  - Image processing is techniques that processes low-level signals (intensity, color, frequency, etc.)

Computer Vision

- Difference with image processing
  - Image processing is techniques that processes low-level signals (intensity, color, frequency, etc.)

Computer Vision

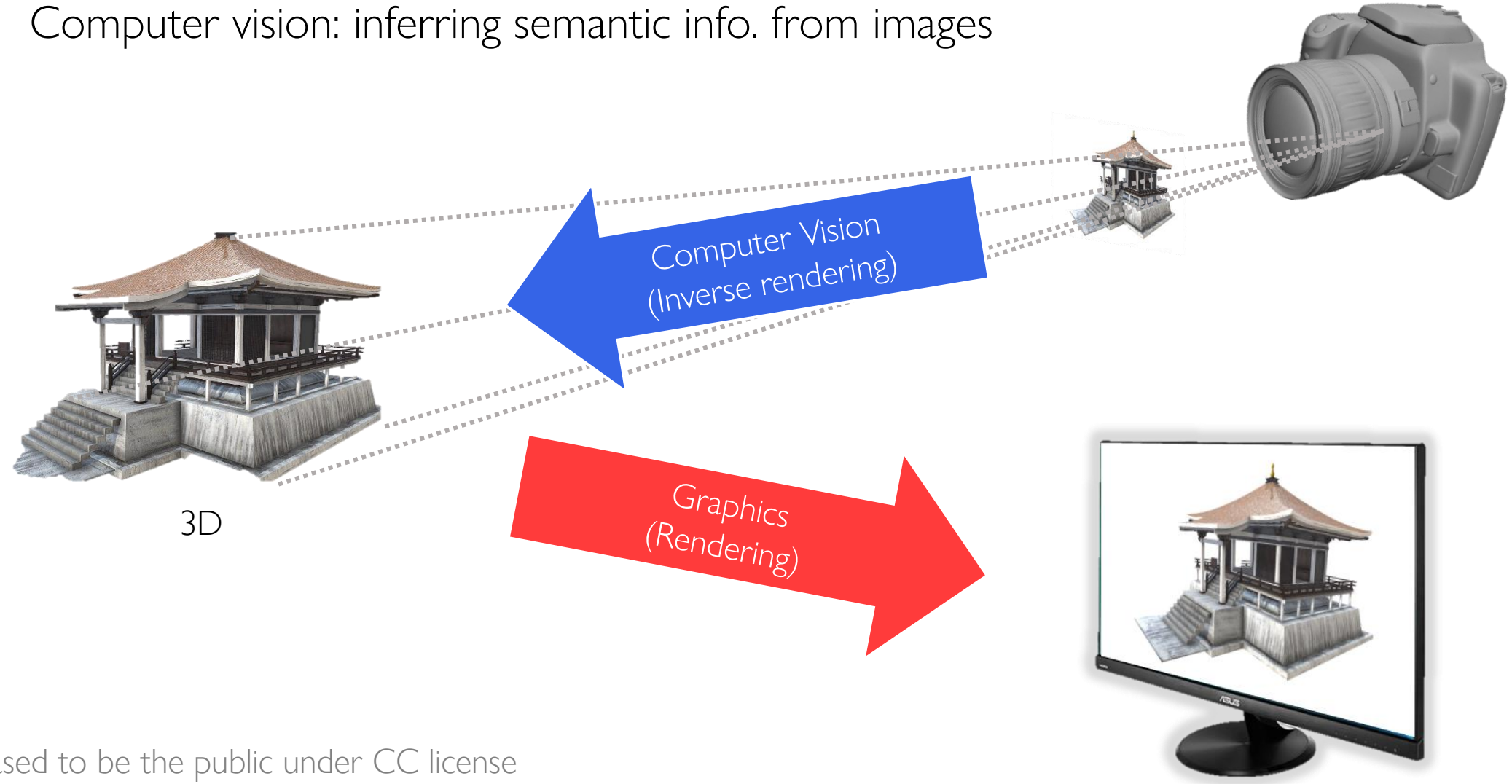
- Difference with graphics
  - From semantic info. (physical values), it renders images

3D

Graphics
(Rendering)

Computer Vision

- Difference with graphics
  - From semantic info. (physical values), it renders images
  - Computer vision: inferring semantic info. from images



Computer Vision

Input \ Output	Image	Semantic model (e.g., physical values, high-level knowledge, etc.)
Image	Image processing	Computer Vision
Semantic model	Computer Graphics	-

Recent progress in the field

- Color-texture perception – Neural Image Stylization (style transfer)

Photograph of Tübingen

Van Gogh's Starry Night

Output of Neural Image Stylization

Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

Image credit: Leon Gatys et al.

Recent progress in the field

- Motion perception – Subtle video motion magnification

Oh et al., Learning-based Video Motion Magnification, ECCV 2018

Recent progress in the field

- 3D perception

Given a single photo

Weng et al., Photo Wake-Up: 3D Character Animation from a Single Photo, CVPR 2019

Recent progress in the field

- 3D perception

Rematas et al., Soccer On Your Tabletop, CVPR 2018

Recent progress in the field

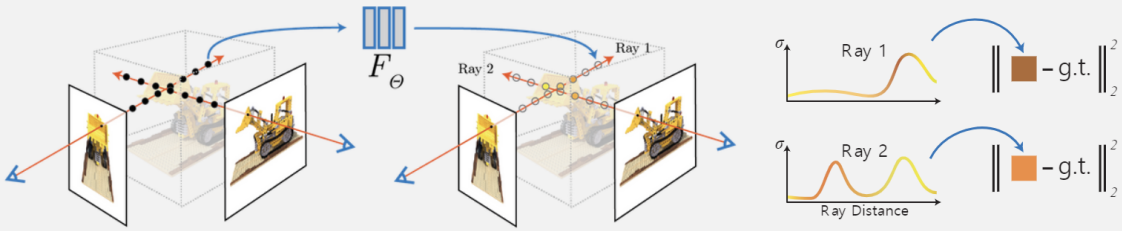
- 3D perception (by implicit model)

Abstract & Method

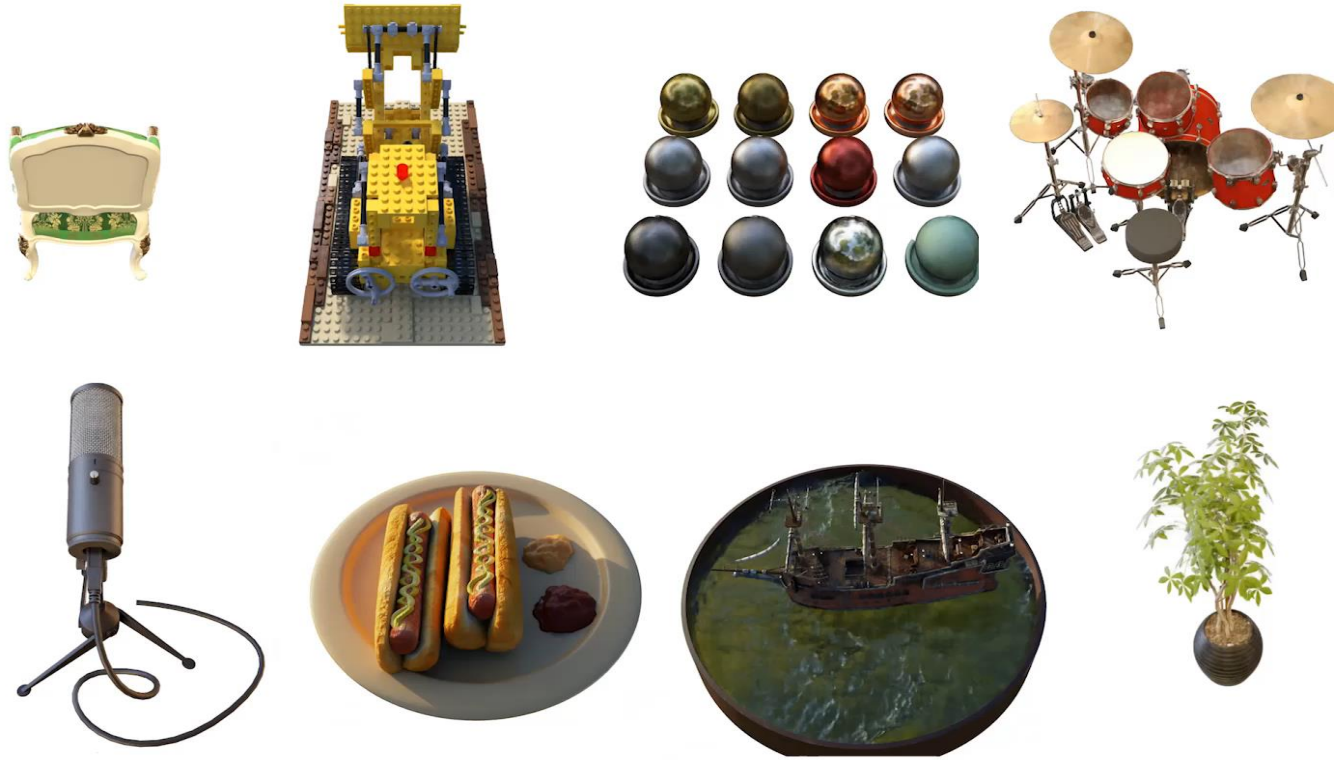
We present a method that achieves state-of-the-art results for synthesizing novel views of complex scenes by optimizing an underlying continuous volumetric scene function using a sparse set of input views.

$$(x,y,z,\theta,\phi) \rightarrow \boxed{\quad} \rightarrow (RGB\sigma)$$
$$F_\Theta$$

Our algorithm represents a scene using a fully-connected (non-convolutional) deep network, whose input is a single continuous 5D coordinate (spatial location (x, y, z) and viewing direction (θ, φ)) and whose output is the volume density and view-dependent emitted radiance at that spatial location.



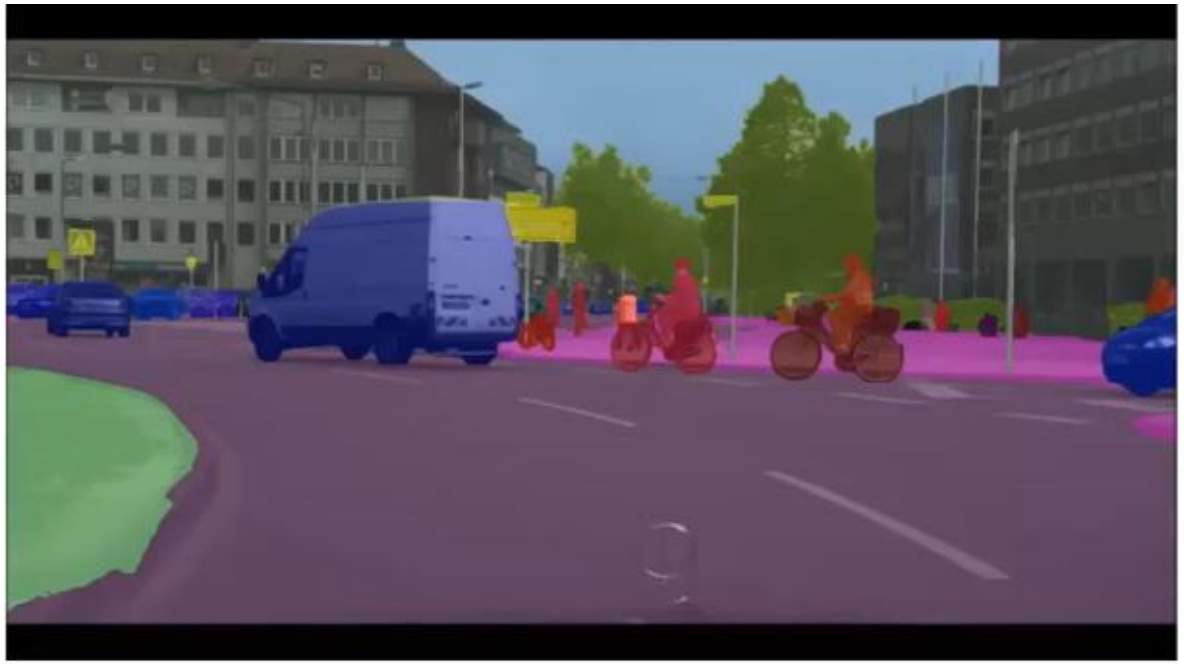
We synthesize views by querying 5D coordinates along camera rays and use classic volume rendering techniques to project the output colors and densities into an image. Because volume rendering is naturally differentiable, the only input required to optimize our representation is a set of images with known camera poses. We describe how to effectively optimize neural radiance fields to render photorealistic novel views of scenes with complicated geometry and appearance, and demonstrate results that outperform prior work on neural rendering and view synthesis.



Mildenhall et al., Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020

Recent progress in the field

- Semantic object-level perception



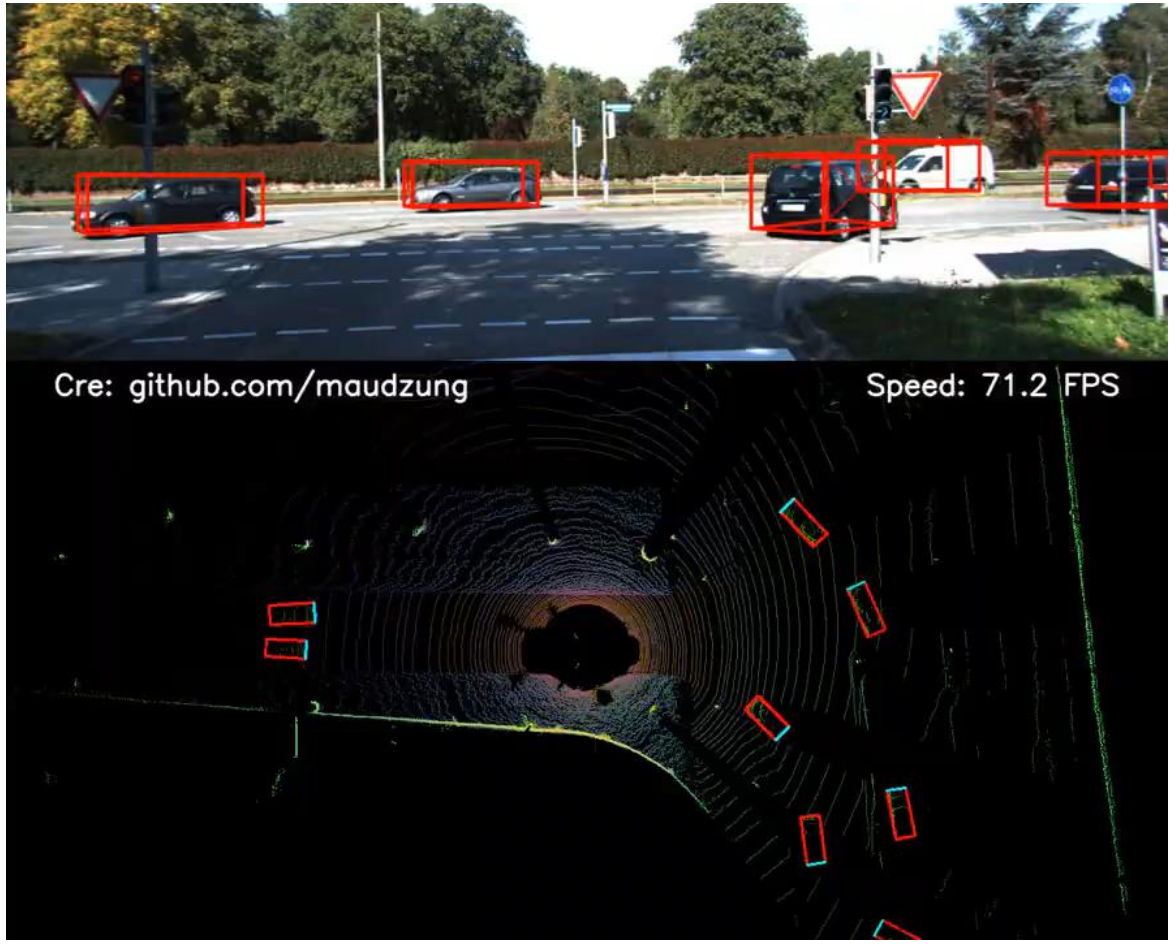
Base : image-level

VPSNet

Kim et al., Video Panoptic Segmentation, CVPR 2020

Recent progress in the field

- 3D & Semantic object-level perception



Li et al., RTM3D: Real-time Monocular 3D Detection from Object Keypoints for Autonomous Driving, Arxiv

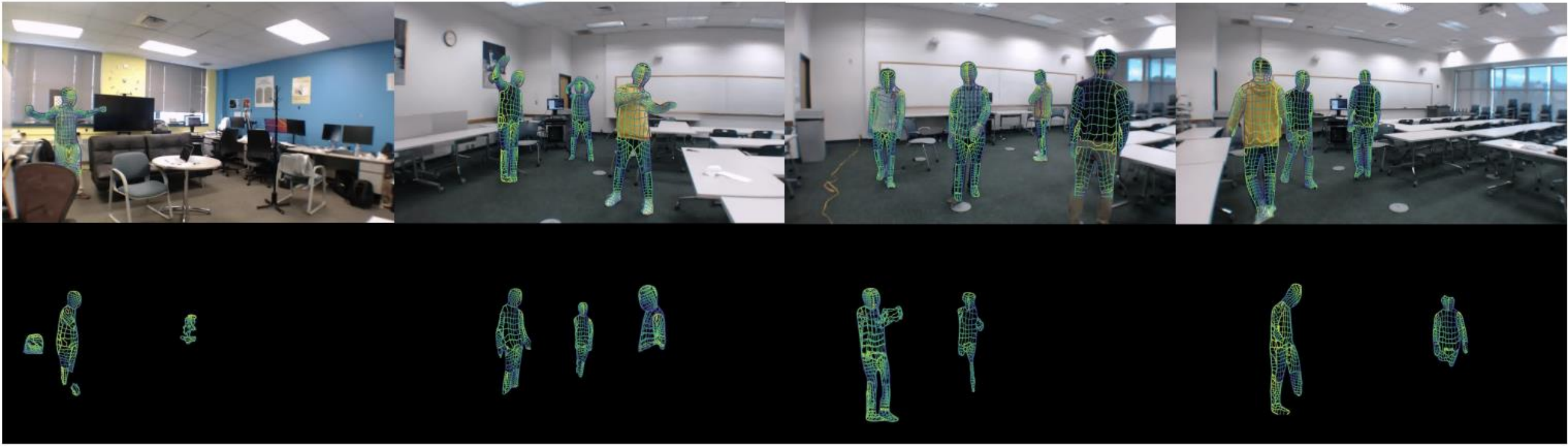
Recent progress in the field

- Multi-modal perception – WiFi signal to human body pose

Zhao et al., Through-Wall Human Pose Estimation Using Radio Signals, CVPR 2018

Recent progress in the field

- Multi-modal perception – WiFi signal to human body (dense) pose



Geng et al., DensePose From WiFi, arXiv 2022

Academic Research Fields

	Name (Abbr.)	Impact factor	Description
Journals	TPAMI	24.314	1st rank in all CS/EE/AI (Top 0.2%)
	IJCV	13.369	
	TIP	11.041	
	TOG (SIGGRAPH)	7.403	1st rank in Graphics (Top 2.4% in CS)
Conference	NeurIPS	In BK top CS conference list, recognized as IF 4.0	Acceptance rate ~20% (oral accept. rate ~2%)
	ICML		
	ICLR		
	AAAI		
	CVPR (1st rank in CS) ICCV ECCV		

Google Scholar

- My profile
- My library
- Metrics
- Alerts
- Settings

카테고리 ▾ 영어 ▾

	발행처	h5-색인	h5-중앙값
1.	Nature	444	667
2.	The New England Journal of Medicine	432	780
3.	Science	401	614
4.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	389	627
5.	The Lancet	354	635
6.	Advanced Materials	312	418
7.	Nature Communications	307	428
8.	Cell	300	505
9.	International Conference on Learning Representations	286	533
10.	Neural Information Processing Systems	278	436
11.	JAMA	267	425
12.	Chemical Reviews	265	444
13.	Proceedings of the National Academy of Sciences	256	364
14.	Angewandte Chemie	245	332
15.	Chemical Society Reviews	244	386
16.	Journal of the American Chemical Society	242	344
17.	IEEE/CVF International Conference on Computer Vision	239	415
18.	Nucleic Acids Research	238	550
19.	International Conference on Machine Learning	237	421
20.	Nature Medicine	235	389

Course overview

- How to build machine perception capability
  - like human, or
  - beyond human
    (super-human performance)

- Mostly computer vision
  + other modalities

- Classic + Modern deep learning in 3D