COMP3065-Computer Vision

Topic1: Introduction to Computer Vision

Fiseha & Tianxiang

Team

- Conveners:
  - Dr. Fiseha B. Tesema:
    - Office: PMB 442,
    - Office hours: Wednesday 13:00-15:00,
    - Email: Fiseha-Berhanu.Tesema@nottingham.edu.cn
  - Dr. Tianxiang Cui:
    - Office: PMB 426,
    - Office hours: Thursday 14:00-16:00,
    - Email: tianxiang.cui@nottingham.edu.cn
- GTA:
  - Xiaoying Yang (xiaoying.yang@nottingham.edu.cn)
  - Nanjiang Du (nanjiang.du@nottingham.edu.cn)

Outline

- Module Overview
- Introduction

Module Overview

Timetabling

- Lectures:
  - Introduction of classical and latest computer vision techniques.
  - Lectures in-person in **DB-C06 (Tuesday 16:00-18:00)**
- Labs:
  - Practicing the techniques learned during lecture sessions.
  - Labs in person in **IAMET-406 (Wednesday 9:00-11:00)**
  - **Python** is the main programming language used in this module for labs and coursework. You can install it from an online repository. It is freely available. Please let us know if you have any problems.
  - You may also bring your own laptop, preferably one with GPU capability.
  - Lab-related materials will be made available in advance on Course Moodle page each week
  - The convener and GTA will support the lab session.
- Wechat group will be used for discussion regarding with learning materials or questions.

Learning Outcome

- Learn what **Computer Vision is about** and the **associated problems** that CV tries to solve
- Learn about **existing techniques** solving these problems and understand the algorithms behind them
- Knowledge **about existing current** research in CV
- To gain experience in **implementing CV** solutions to real world problems.
- Be able to write about computer vision approaches in a **technical form**

Prerequisites

- No prior experience with computer vision is assumed, although previous knowledge of **visual computing** or **signal processing** will be helpful. The following skills are necessary for this class:
  - Data structures: You'll be writing code that builds representations of images, features, and geometric constructions.
  - Programming: labs are to be completed in Python
  - Math: **Linear algebra**, **vector calculus**, and **probability**.
  - Machine learning basics,
    - linear and logistic regression
    - ANN

Topics Covered(tentative)

2024/25 Schedule of activities(Tentative)

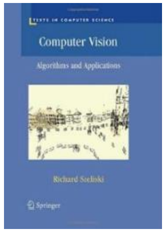
Week	Title	Date	Led by
1	Topic 1: Introduction to CV module		Tianxiang & Fiseha
2	Topic 2: Describing Image Regions and Patches		Tianxiang
3	Topic 3: Point Features: SIFT		
4	Topic 4: Image Stitching 1		
5	Topic 5: Image Stitching 2		
	Multiple Views and Motion		
6	Topic 6: Camera, Camera Calibration, Epipolar Geometry		Fiseha
7	Topic 7: Stereo		
8	Topic 8: Optical Flow		
	Recoginition and Generative Models		
9	Topic 9: Bag of features		
	Topic 10: Voila Jones methods		
10	Topic 11: Convolutional Neural Network (CNN) for CV		
11	Topic 12: Generative Adversarial Networks (GANs)		
12	Revision		Tainxiang & Fiseha

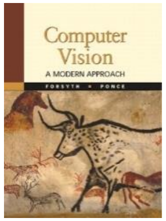
* Please check for Moodle announcements as these could still change

Materials

- The lecture materials are adapted from many professors' in the field including Tony Pridmore, Kristen Grauman, Fei-Fei Li, Svetlana Lazebnik, Prof. Andreas Geiger etc.
- Topics are tentative.
  - Please see Moodle page for regular update.

- Text book
  - Richard Szeliski, Computer Vision: Algorithms and Applications, 2nd ed., 2022 (free pdf version online, https://szeliski.org/Book/download.php .
  - David Forsyth, Jean Ponce, Computer Vision: A Modern Approach, Pearson, 2011.





COMP3065 - Assessment

- Coursework (**40%**)
  - Programming assignments/reports
  - Python applications with source code and documentation
  - Explanation & evaluation of results
  - Your work must be your own. We’ll look for cheating. Don’t talk at the level of code with other students.

- Exam (**60%**)
  - 2 hours, close book
  - Answer 3 or 4 questions with sub-questions

How to Get 70+

- **Studying…**
  - You are recommended to study the relevant notes before attending the lecture or lab.
  - Review as soon as possible to maximize retention.
- **Practice…**
  - **Do the lab exercise yourself** and repeat the practice for better learning.
  - If you get help on the labs, don't just blindly accept it, but try to understand what each part of the code is doing.
  - Do the Math in the lecture for better understanding.
- **Assignments…**
  - Start work on the assignment **when they are released**, and come up with a good plan to finish it.
  - Many times fixing problems in your program will take **longer** than you expect, so make sure you have plenty of time to complete.

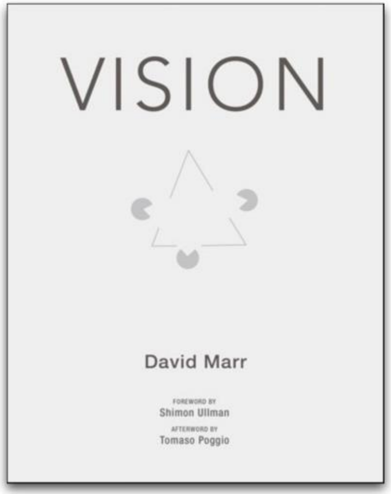
Introduction

Introduction

- Computer Vision(CV)
- CV vs Biological Vision
- CV vs. Image Processing
  - Application of Computer Vision
- Computer Vision Vs. Machine Learning
- Why is Visual Perception hard?
- History of Computer vision

Computer Vision

- What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking.



"CV is to discover from images what is present in the world, where things are, what actions are taking place, to predict and anticipate events in the world."

David Courtenay Marr (19 January 1945 – 17 November 1980) was a British **neuroscientist** and **physiologist**

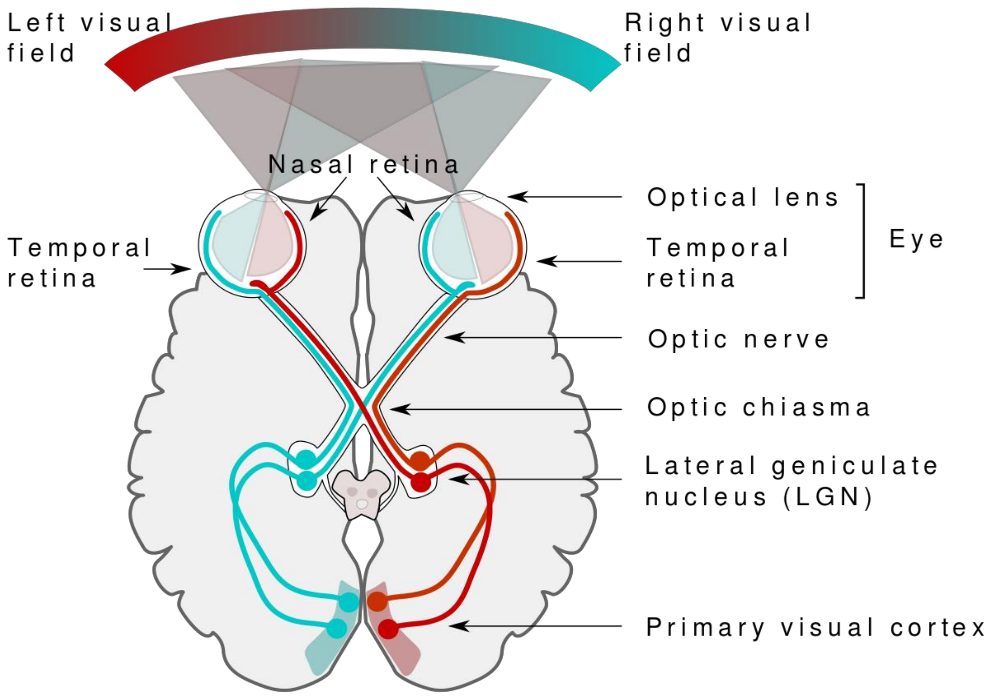
Computer Vision

- **Computer Vision** is concerned with the computational processes that allow representations of the viewed environment to be recovered from individual or sequences of images.

- “Computer vision is a field of computer science that focuses on enabling computers to identify and understand objects and people in images and videos.”

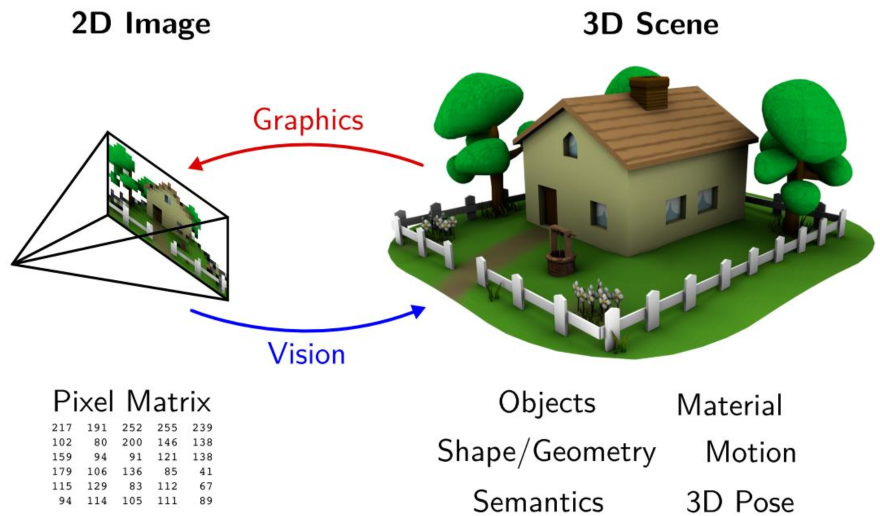
https://azure.microsoft.com/

Computer Vision Vs. Biological Vision



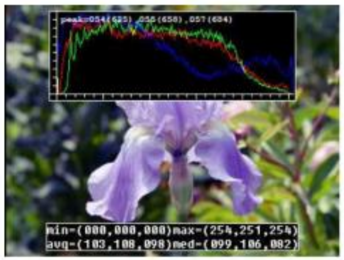
- Most of computer vision algorithm was inspired by **biological vision.**
- Over **50%** of the processing in the human brain is dedicated to **visual information**.

Computer Vision Vs. Computer Graphics



- Computer graphics is the forward problem:
  - Given scene geometry, reflectances, and lighting → synthesize an image
- Computer Vision is often thought of as inverse graphics:
  - Only raw pixel is accessible
  - Given single/multiple images recover the scene geometry, reflectances and illumination.

Computer Vision vs. Image Processing



Color adjustment

Edge Detection

Denoising

Image warping

(a) Subject 1 (b) Morph (c) Subject 2

U. Scherhag, Face Recognition Systems Under Morphing Attacks: A Survey, 2019

- **Image processing** concerned **with low level image manipulation** to improve their quality or prepare them for further analysis:
    - Deal with 2D images

Slide Credits: Rick Szeliski

Computer Vision vs. Image Processing

- Tasks:
  - Filtering: Smoothing, sharpening, noise reduction.
  - Enhancement: Contrast adjustment, brightness correction.
  - Transformation: Scaling, rotation, flipping, perspective corrections.
  - Segmentation: Dividing the image into regions based on pixel similarities (e.g., thresholding).
  - Feature Extraction: Detecting edges, corners, or blobs in the image.
- Applications: Image enhancement for visual quality (e.g., denoising, sharpening).
  - Preparing images for further analysis in areas like medical imaging, satellite imagery, and photography.

Computer Vision vs. Image Processing

- Historically **CV is deviated from image processing** primarily in the early stage because it is getting a more holistic scene representation, scene understating and recovering the 3D structure from 2D.
- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
    - Image Classification: categorizing images into user-defined classes for various applications.

Computer Vision vs. Image Processing

- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
  - Object Detection: Identifying and locating objects in an image (e.g., face, car,tiger).

Computer Vision vs. Image Processing

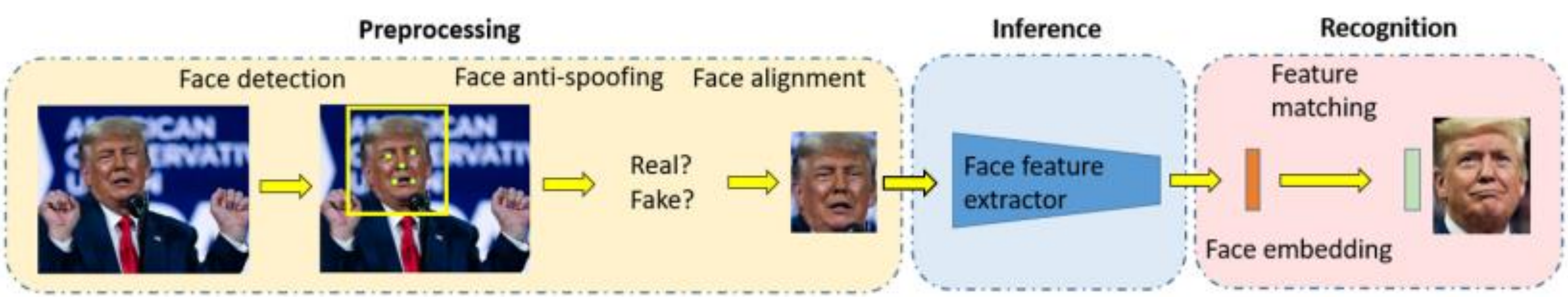
- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
  - Scene Understanding: Segmenting an image into semantically meaningful parts.

Semantic segmentation aims to identify each pixel within an image for a more detailed categorization.

Instance segmentation identifies each instance of the same object, making it more granular than semantic segmentation.

Computer Vision vs. Image Processing

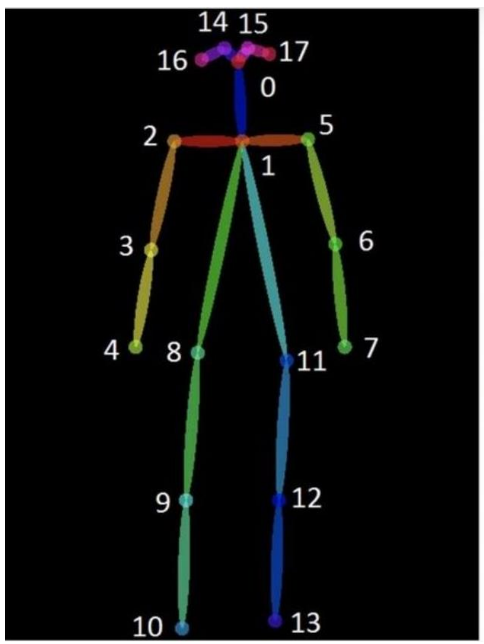
- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
  - Image recognition: is a broader concept that can involve **classifying** objects but also typically includes **localizing**, **identifying**, or **recognizing** objects in images with more details. It's about **identifying the objects or features** in an image.

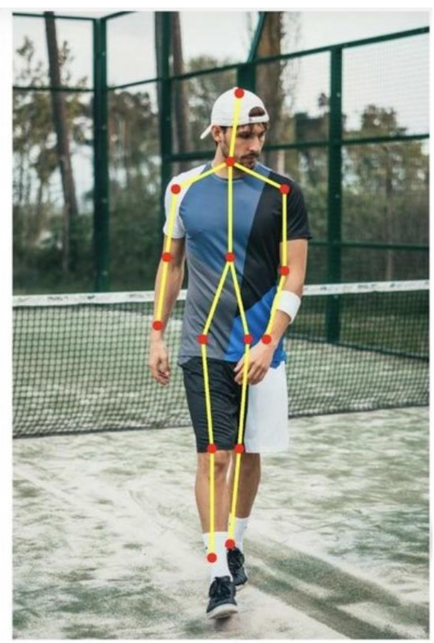


[A Survey of Face Recognition, Xin you etal, 2022]

Computer Vision vs. Image Processing

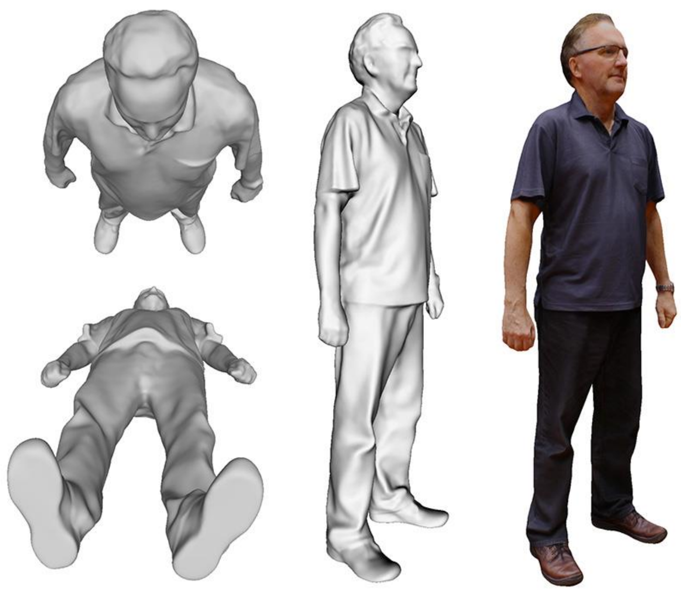
- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
  - Pose Estimation:
    - Identifies key semantic points on an object to track orientation.





Computer Vision vs. Image Processing

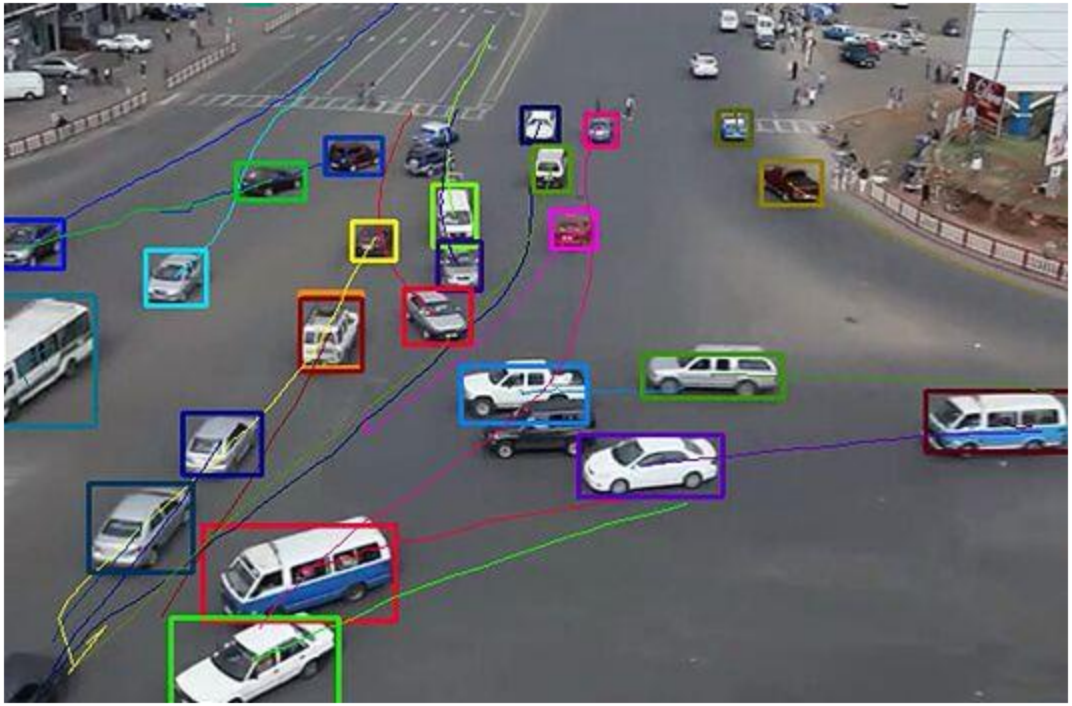
- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
    - 3D Reconstruction: Rebuilding 3D models from 2D images.



https://paulbourke.net/reconstruction/humanscan2/

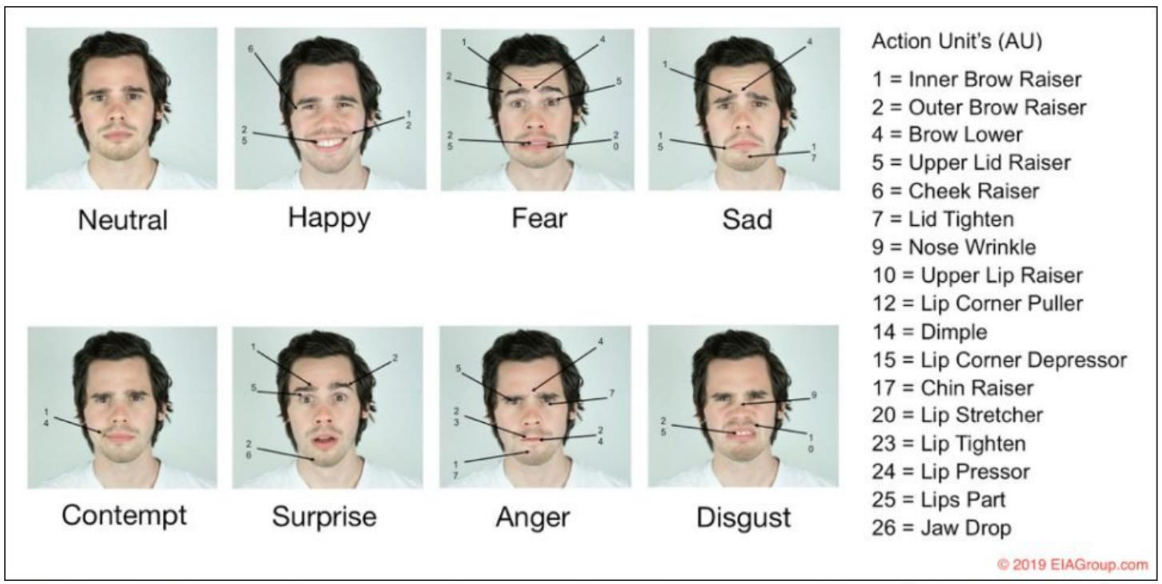
Computer Vision vs. Image Processing

- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
    - Tracking: Monitoring objects or people across video frames.



Computer Vision vs. Image Processing

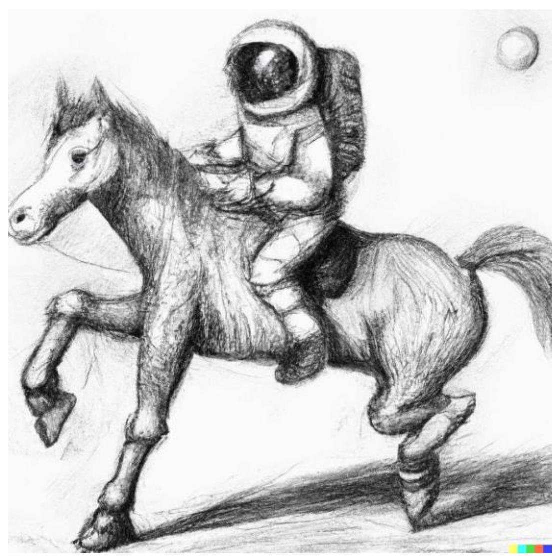
- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
  - Action/Emotion Recognition: Analyzing human activity or emotion in videos.



A slightly extended gamut of available facial expressions in the FACS system. Source: https://www.eiagroup.com/the-facial-action-coding-system/

Computer Vision vs. Image Processing

- CV Aims to interpret and understand visual data to gain meaningful insights or make decisions.
- Key Tasks:
  - Image Generation and Synthesis: AI algorithms generate novel images, artwork, designs, etc., based on training data.



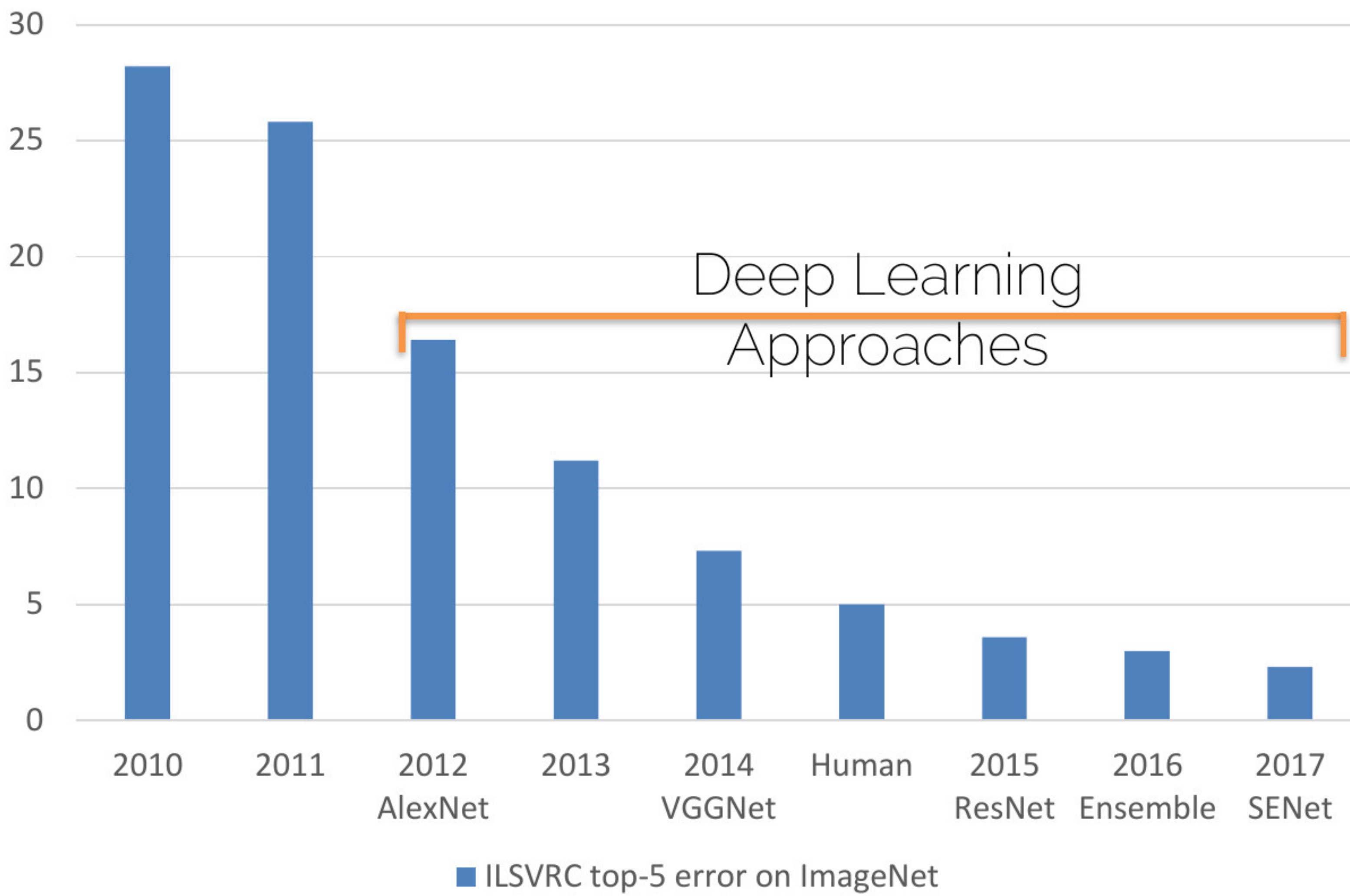
AI art generated with Dalle2 Open AI

Computer Vision Vs. Machine Learning

- A lot of developed machine learning algorithms are used everywhere in CV
- Made computer vision successful and commercially successful.
- CV also brought a several breakthrough for ML and in other field of ML such as language processing by accelerating the pace if the research.

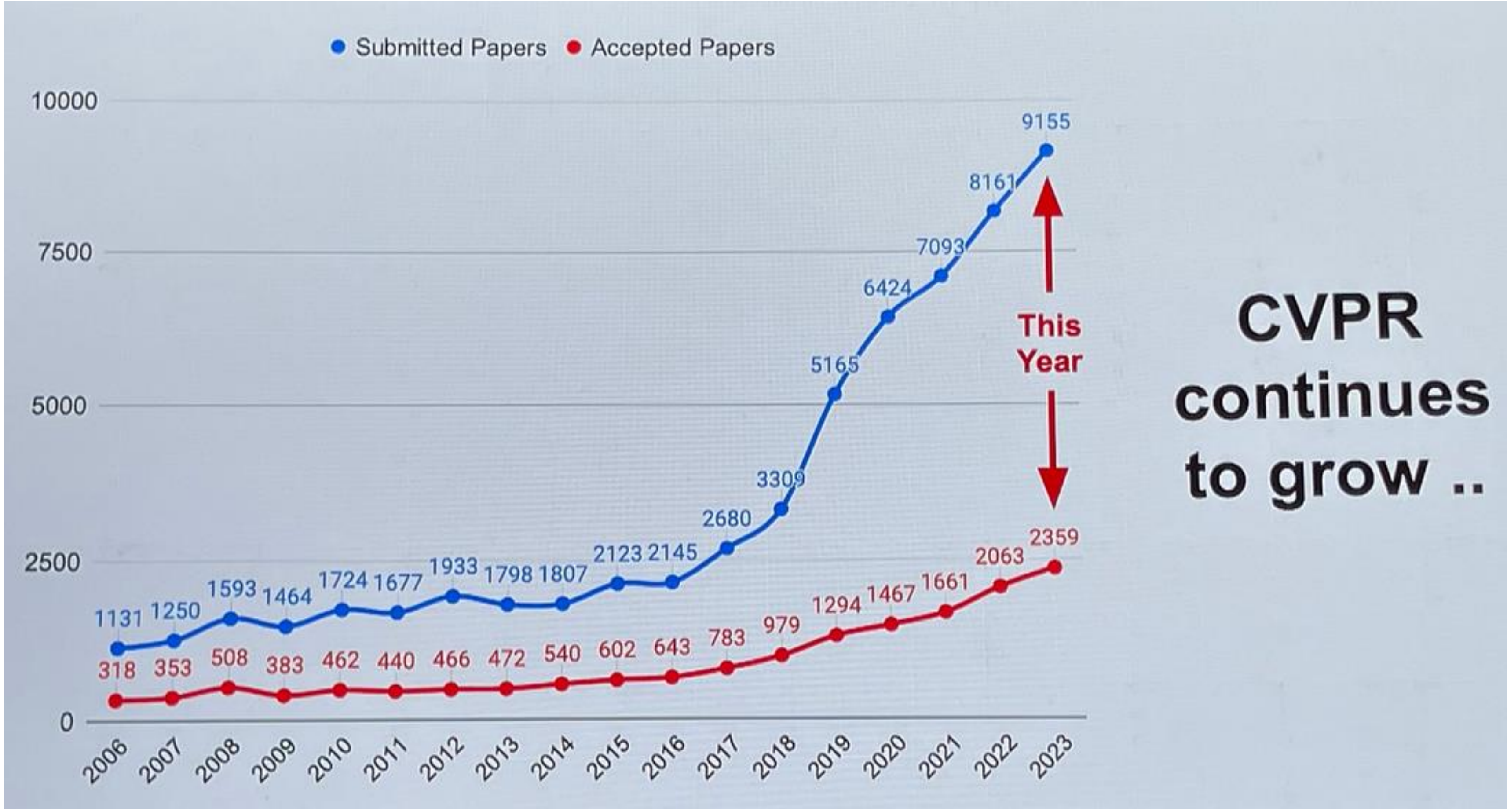
Deng, Dong, Socher, Li, Li and Li: ImageNet: A large-scale hierarchical image database. CVPR, 2009.

The Deep Learning Revolution



Slide Credit: Matthias Niessner

CVPR Submitted and Accepted Papers



[https://medium.com/@dobko_m/cvpr-2023-summary-ad271d383404]

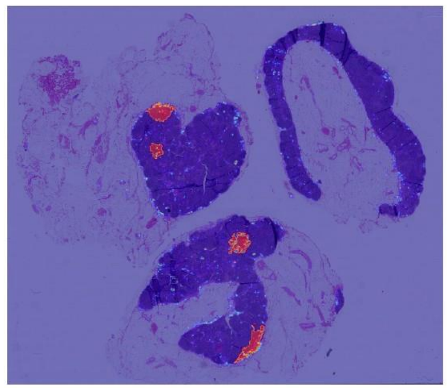
CVPR 2023 Sponsors

[https://twitter.com/CVPR/status/1671181305596485632/photo/4]

Computer Vision Application

Robotics

Medical applications



3D modeling

Driving

Mobile devices

Accessibility

Many more applications: https://viso.ai/applications/computer-vision-applications/

Computer Vision Application

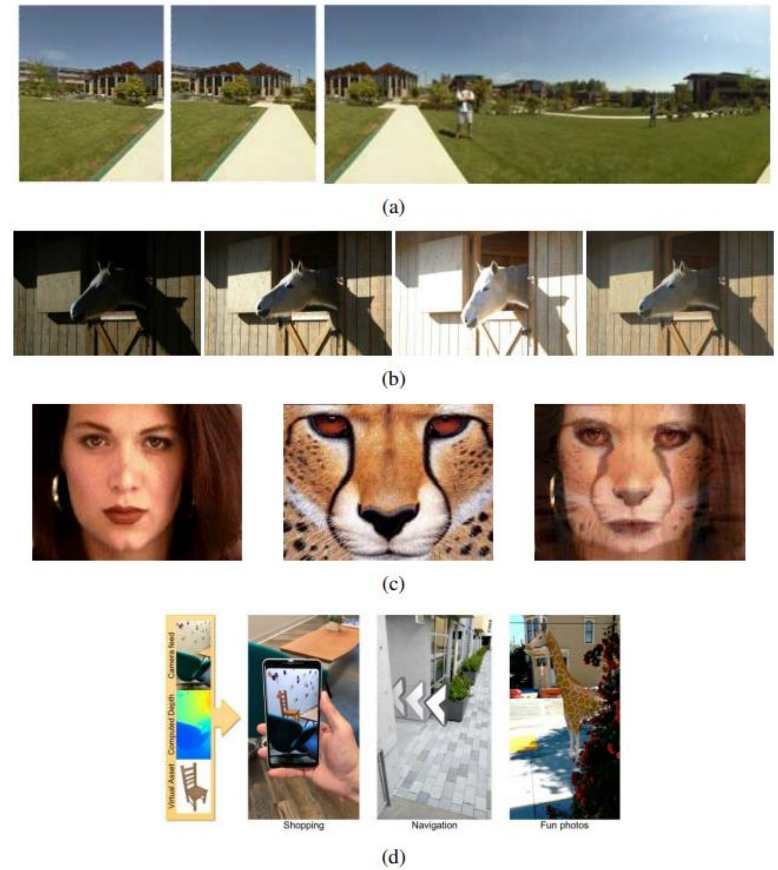


Figure 1.5 *Some consumer applications of computer vision: (a) image stitching: merging different views (Szeliski and Shum 1997) © 1997 ACM; (b) exposure bracketing: merging different exposures; (c) morphing: blending between two photographs (Gomes, Darsa* et al. *1999) © 1999 Morgan Kaufmann; (d) smartphone augmented reality showing real-time depth occlusion effects (Valentin, Kowdle* et al. *2018) © 2018 ACM.*

[Szeliski, Computer Vision: Alogrithsm and Applications, 2nd Ed.]

Computer Vision Application

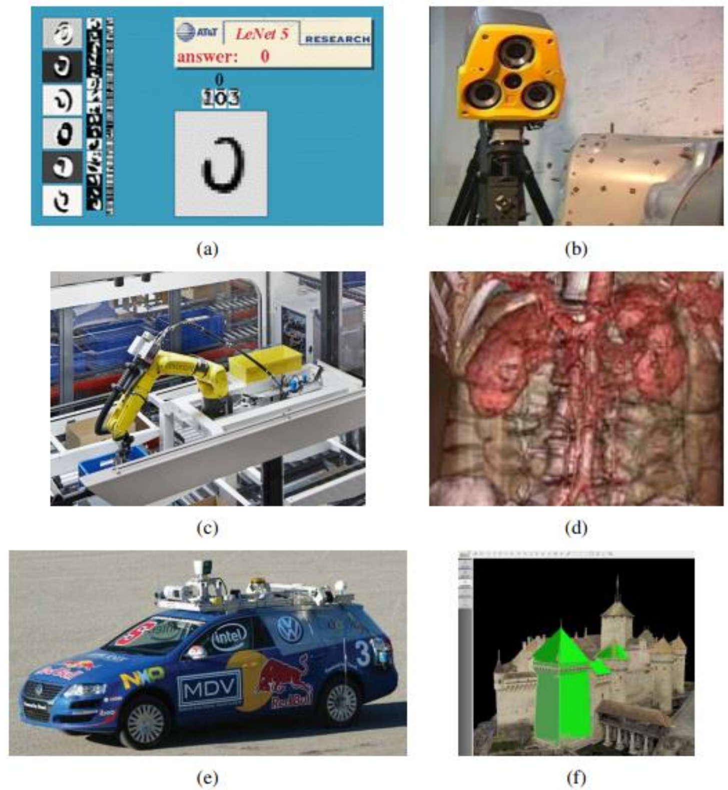


Figure 1.4 *Some industrial applications of computer vision: (a) optical character recognition (OCR), http://yann.lecun.com/exdb/lenet; (b) mechanical inspection, http://www.cognitens.com; (c) warehouse picking, https://covariant.ai; (d) medical imaging, http://www.clarontech.com; (e) self-driving cars, (Montemerlo, Becker* et al. *2008) © 2008 Wiley; (f) drone-based photogrammetry, https://www.pix4d.com/blog/mapping-chillon-castle-with-drone.*

[Szeliski, Computer Vision: Alogrithsm and Applications, 2nd Ed.]

Why is Visual Perception hard?

- In the beginning it was underestimated!!!!

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".

Papert: The Summer Vision Project. MIT AI Memos, 1966.

Why is visual perception hard?

What we see

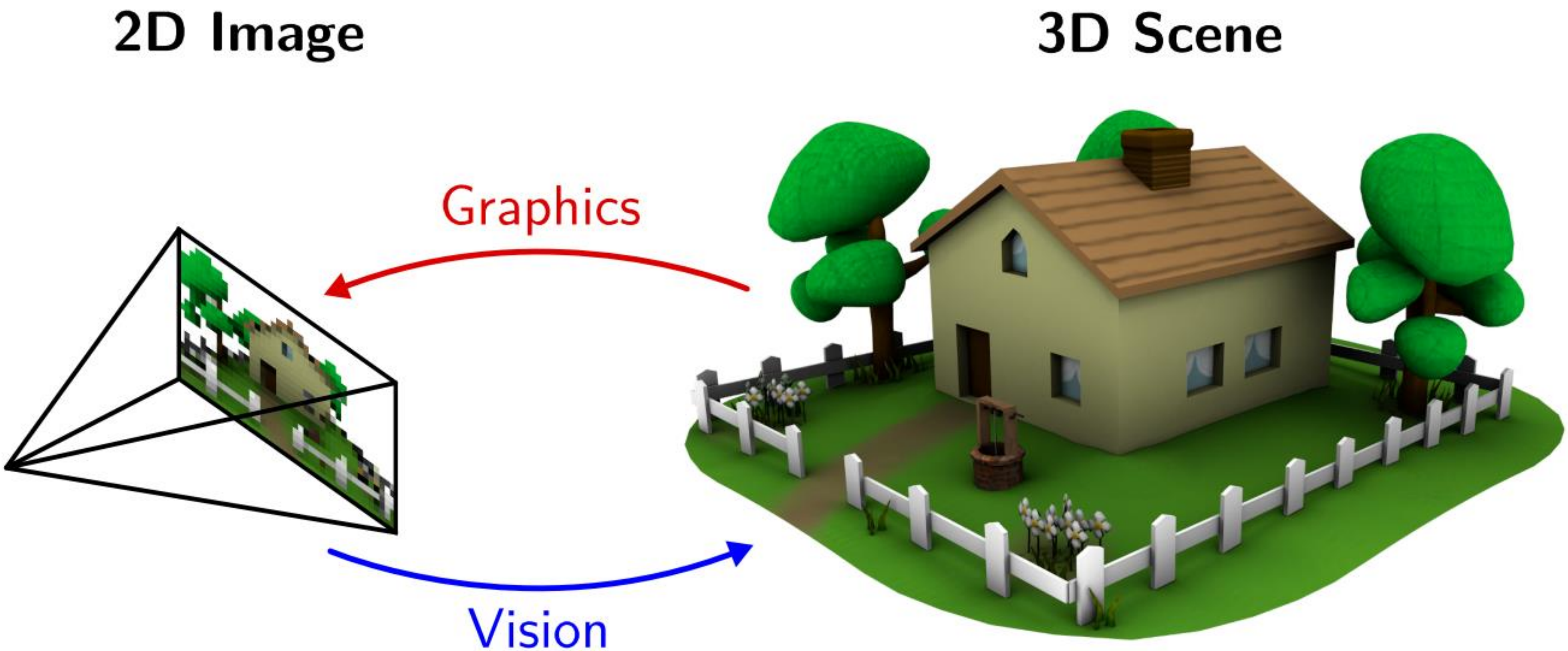
200	133	110	103	117	90	47	30	32	79	66	65
197	122	123	138	98	100	46	45	22	11	43	55
140	116	165	159	90	56	58	47	26	13	54	102
132	148	119	108	123	57	64	46	21	22	79	94
125	121	80	143	101	55	61	38	20	21	81	65
50	71	74	63	52	39	41	39	32	26	97	66
51	59	62	44	40	40	36	28	27	31	29	44
59	62	70	50	48	35	34	35	26	21	24	32
49	59	65	64	58	34	40	28	26	21	23	124
39	45	47	64	54	34	40	24	19	47	133	207
37	42	39	38	39	50	75	74	105	170	197	167
37	47	33	35	50	108	162	184	184	157	125	112
45	48	35	37	75	148	183	156	83	91	91	116
49	48	54	50	75	158	110	66	74	128	155	149
48	51	57	50	65	91	79	92	101	105	132	132
51	58	66	55	58	52	91	91	88	115	158	174
57	60	61	52	56	61	60	55	92	146	188	190
65	50	54	56	57	51	54	56	80	115	177	187
67	40	40	61	65	48	39	30	36	75	151	181
53	32	36	35	61	43	37	26	29	35	126	189
29	42	107	20	28	41	40	26	30	36	113	200
30	21	32	24	34	37	33	23	25	39	105	171
32	28	19	23	29	36	47	89	132	169	183	128
31	25	62	54	47	44	81	190	227	231	206	155
44	66	99	72	67	63	89	128	127	115	109	157
53	47	47	41	29	32	25	20	41	81	89	175
38	44	61	73	54	48	37	87	90	111	126	189
39	41	83	97	86	91	74	134	131	153	143	185
42	56	98	102	112	111	94	137	121	141	146	181
94	114	114	114	122	113	77	117	117	154	149	169
157	176	116	121	130	139	103	161	148	180	145	125
143	178	182	178	139	153	129	168	175	187	170	152
127	183	203	197	153	164	143	180	195	182	165	211
88	107	127	125	101	107	100	123	149	186	167	215

What the computer sees

Slide Credits: Li Fei-Fei

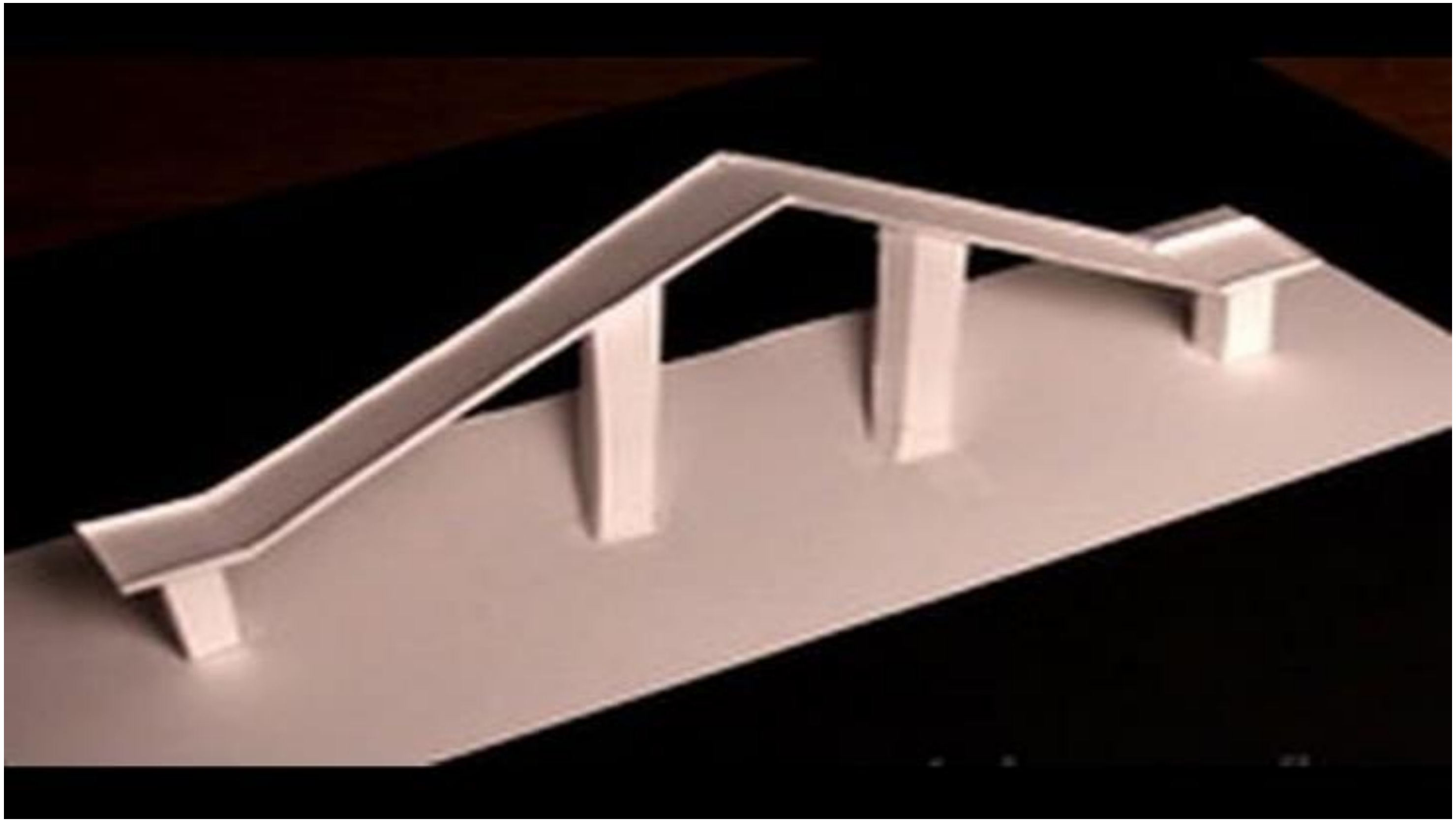
Pixel values represent the brightness and colour of the viewed objects, but give no indication of what object, incident, and so on.

Challenges: Image are 2D Projections of the 3D world



Ames Room Illusion

Perspective Illusion



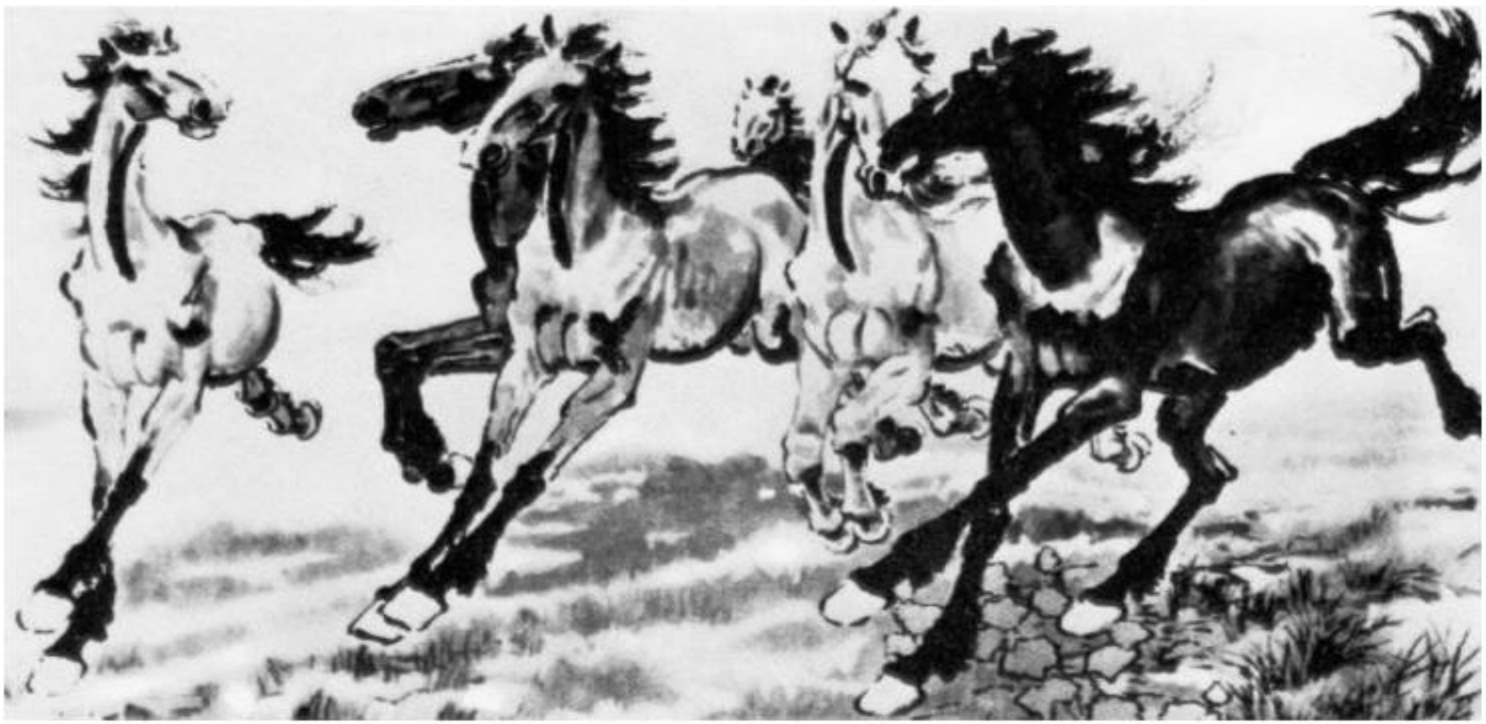
https://www.youtube.com/watch?v=vmkaVoLoFEU

Challenges: Viewpoint Variation

- Due to viewpoint change the pixel varies and it is challenging to recognize that these images are the same

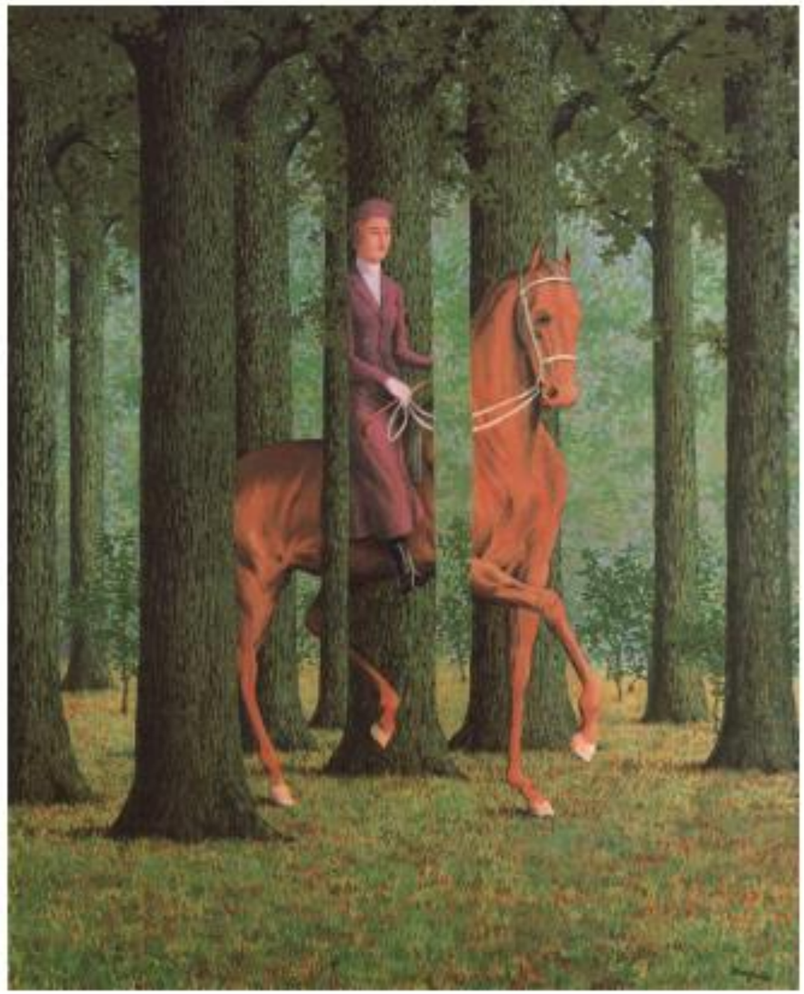
Michelangelo (1475-1564)

Challenges: Deformation



Xu Beihong (1943)

Challenges: Occlusion



Ren’e Magritte (1957)

Challenges: Illumination

Slide Credits: Antonio Torralba

Challenges: Illumination

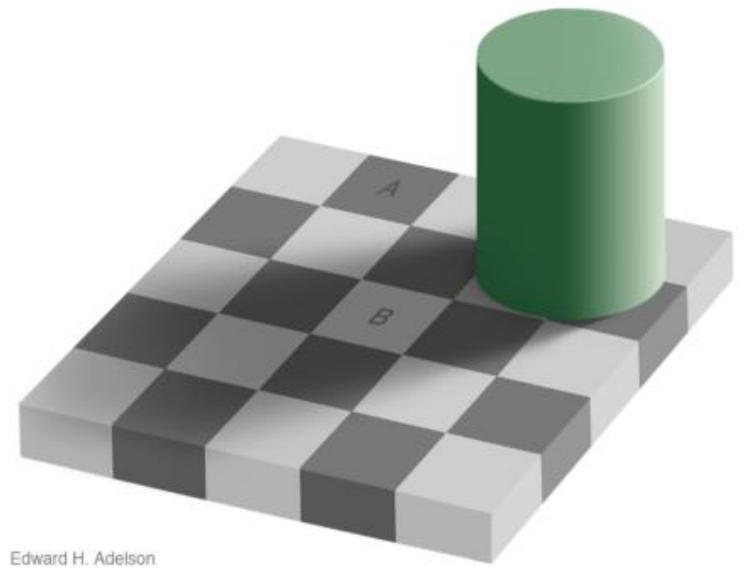
Slide Credits: Antonio Torralba

Challenges: Illumination

Slide Credits: Antonio Torralba

Challenges: Motion

Challenges: Perception

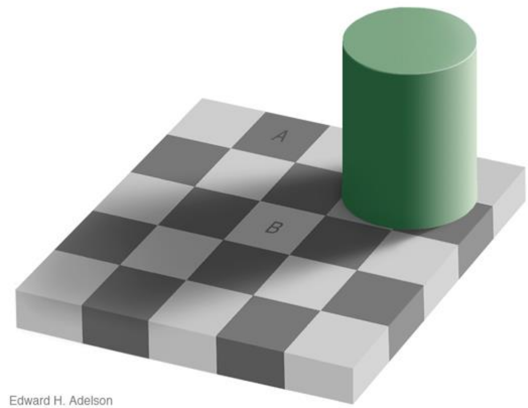


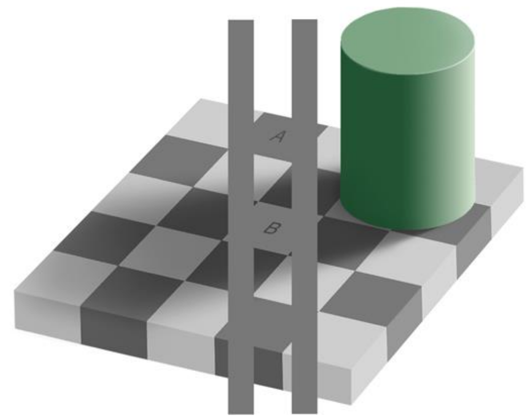
Checker Shadow Illusion

http://persci.mit.edu/gallery/checkershadow
https://www.youtube.com/watch?v=GALLMJxLvgA
https://www.youtube.com/watch?v=0i8ZKTyIaQc

Image Credits: Edward H. Adelson

Challenges: Perception



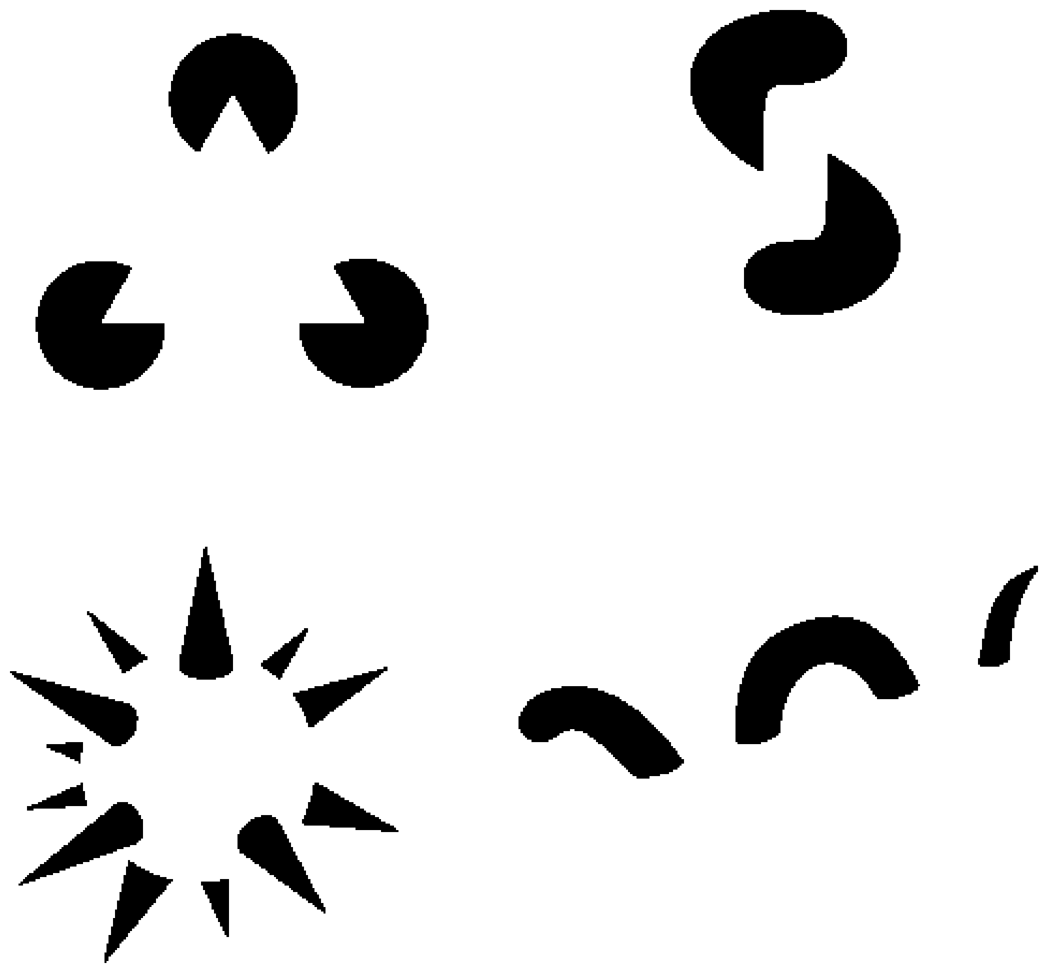


Proof

http://persci.mit.edu/gallery/checkershadow
https://www.youtube.com/watch?v=GALLMJxLvgA
https://www.youtube.com/watch?v=0i8ZKTyIaQc

Image Credits: Edward H. Adelson

Challenges: Perception



Gaetano Kanizsa (1976)

Challenges: Perception

Challenges: Local Ambiguities

- Images in the right have the same pixel arrangement with the small segment on the left.
- From the local component it is impossible to determine what an object is.
- It become clear when we interpret the scene as a whole.

Image Credits: Antonio Torralba

Challenges: Local Ambiguities

- Prior knowledge may lead to wrong perception.

Image Credits: Antonio Torralba

Challenges: Local Ambiguities

Image Credits: Antonio Torralba

Challenges: Intra Class Variation

http://www.homeworkshop.com/

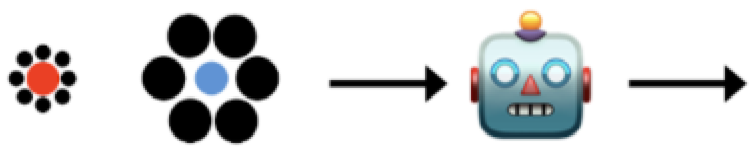
Image Credits: Antonio Torralba

Challenges: Number of Object Categories

Image Credits: Antonio Torralba

Current models still make very silly mistakes

Prompt: “Which is bigger, the blue circle or the red circle?”



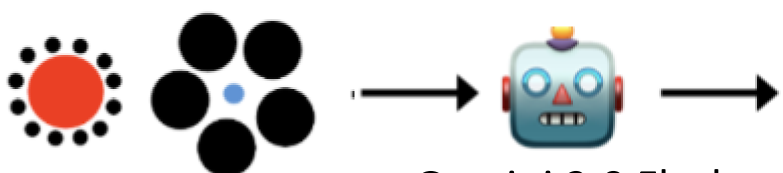
Gemini 2.0 Flash

This is a classic **optical illusion**!

The **red and blue circles are actually the same size.**

Our perception is tricked by the surrounding circles. The small black circles surrounding the red circle make it appear larger, while the larger black circles surrounding the blue circle make it appear smaller. This is due to how our brains interpret context and relative size.

You can verify this by measuring the circles with a ruler or other tool, or by simply covering up the surrounding circles – the red and blue circles will then appear identical in size.



Gemini 2.0 Flash

This is a classic optical illusion demonstrating how our perception of size can be distorted by context.

The red and blue circles are exactly the same size.

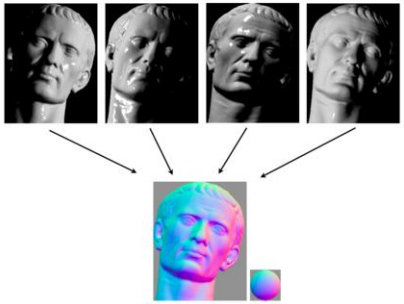
The reason the red circle appears larger is because it's surrounded by smaller circles. Our brains use the surrounding circles as a reference point, making the red circle seem bigger by comparison. Conversely, the blue circle is surrounded by larger circles, making it appear smaller.

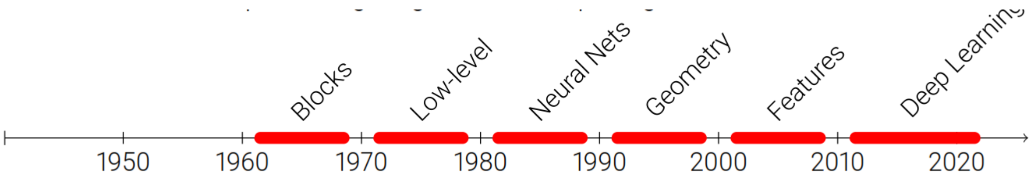
If you were to measure the circles or isolate them from their surroundings, you would see that they are indeed identical in size.

History of Computer vision

Waves of development:

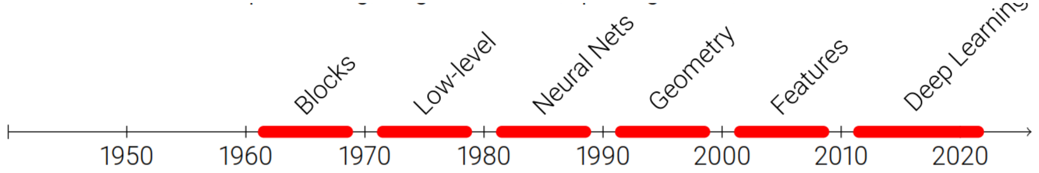
- **1970-1981: Low Level**
  - **Shape-from-shading**: A technique to recover 3D shapes from a single 2D image by analyzing the way light interacts with surfaces.
  - **Photometric Stereo**: A method for recovering 3D information from multiple 2D images taken from the same viewpoint but under different lighting conditions.
  - **Optical Flow Algorithm**: Used to detect motion in video by tracking the movement of objects between consecutive frames.





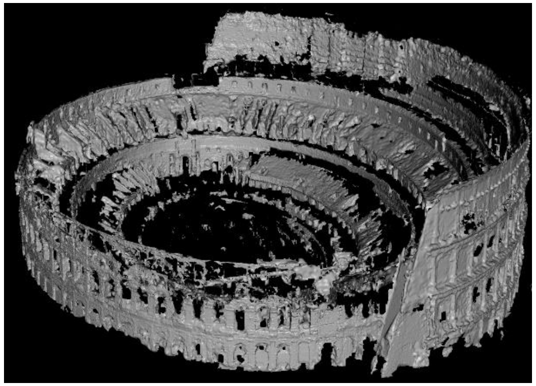
Waves of development:

- **1985-1988: Neural Networks**
  - **Backpropagation**: Introduced as a way to train multi-layer neural networks, enabling advances in machine learning for vision tasks.
  - **Self-driving cars**: The early stages of vision-based algorithms for autonomous vehicle development began during this time.

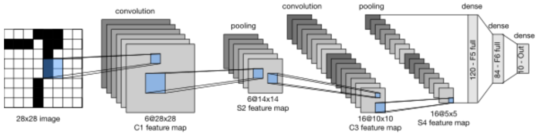


Waves of development:

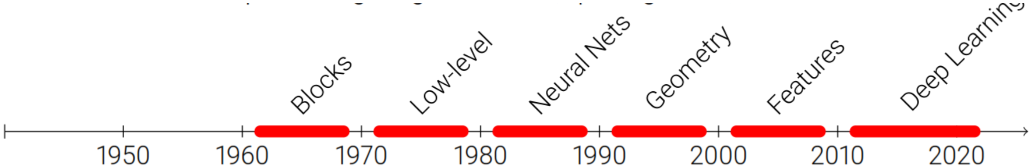
- **1990-2000: Dense Stereo and Multi-view Stereo**
  - **Dense stereo**: Improved algorithms to estimate depth across entire images, building richer 3D models from multiple 2D images.
  - **Multi-view Stereo**: Techniques to reconstruct 3D structures by taking advantage of images from multiple viewpoints.



  - **1998 - CNNs for digit recognition**: Convolutional Neural Networks (CNNs) were introduced for digit recognition but didn't gain much traction at the time.
  - **1999 - SIFT (Scale Invariant Feature Transform)**: A method for detecting and describing local features in images, which became foundational for many vision tasks.

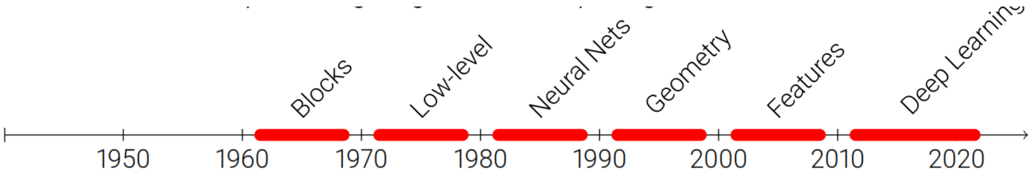


[LeNet-5]



Waves of development:

- **2000-2010: Features**
  - **Large-scale 3D reconstruction**: Using images from the internet (e.g., from photo-sharing websites) to build 3D models of environments.
- **2010-now: Deep Learning**
  - **Deep learning**: The rise of deep neural networks for computer vision, supported by large datasets, powerful GPUs, and rapid advancements in algorithms.
  - **Large datasets**: The availability of massive labeled datasets like ImageNet has greatly contributed to the improvement of deep learning models.
  - **Quick growth and commercialization**: Deep learning techniques have been applied to a wide range of applications, from object detection to facial recognition, leading to rapid growth in both research and commercial applications.
- **2020's: Widespread Autonomous Vehicles?**
  - Potentially, the widespread adoption of autonomous vehicles could transform industries and drive further research in computer vision to handle real-time, complex visual tasks in dynamic environments.



Readings

- Readings
  – Szeliski, Chapter 1 (Introduction)