SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

Computer Vision Introduction

A bit about myself

- Currently: EMEA Lead for AI Inference at INTEL. Assistant professor at IE (previously mathematics and statistics)
- Previously: Data and AI architect at Repsol, Advanced analytics manager at Avanade, research fellow in Massachusetts Institute of Technology, research scientist at the Spanish National Center for Cardiovascular research.
- Studies: PhD studies in the Edinburgh joint research institute (led by university of Edinburgh and heriot watt university) with focus on computer vision applications, electronics engineer from ENSEIRB (Bordeaux, France)

ie
SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

Computer vision: objectives

- Explore the different techniques and frameworks that we can use to analyze and manipulate images
- Build machine learning models that enable artificial intelligence software to recognize and generate visual patterns
- Help you take your first steps into the exciting area of computer vision.
- This will be a blended course: examples and real scenarios will motivate and accompany all elements of theory, and will be consolidated using coding exercises.

Computer Vision: Timeline

Session	Objective
Session 1	INTRODUCTION TO THE COURSE AND ITS ORGANIZATION
Session 2	BASICS OF IMAGE PROCESSING Image transformations: (crop, resize, flip),Image basics with Python, Image normalization: histogram equalization, Image filtering: blurring, edge detection, Thresholding and segmentation
Session 3	IMAGE CLASSIFICATION Definition of image classification, Classical approaches for image classification, and limitations
Session 4	DEEP LEARNING FOR COMPUTER VISION Introduction to deep learning, Convolutional Neural Networks (CNN Vision Transformers (ViT)
Session 5-6	TRANSFER LEARNING AND OBJECT DETECTION Introduction to transfer learning. Introduction to object detection, YOLO versions
Session 7	IMAGE SEGMENTATION Introduction to semantic and instance segmentation
Session 8	HUMAN POSE ESTIMATION Human 2D pose estimation, localizing anatomical key points or parts of individuals using top down and bottom up approaches with deep learning techniques
Session 9	OBJECT TRACKING Detecting and tracking moving objects in video data. Computer vision approaches and algorithms for tracking objects through video, and methods for performance evaluation of tracking algorithms. We review traditiona approaches and deep learning algorithms
Session 10-11	Advanced Generative Models: Architectures and Applications of Variational Autoencoders, Diffusion Models, and Other Novel Approaches
Session 12	Multimodal Learning: Principles, Techniques, and Applications
Session 13	ADVERSARIAL ATTACKS AND HOW TO PREVENT THEM IN IMAGES. Adversarial examples are inputs to machi learning models that an attacker intentionally designed to cause the model to make mistakes, how to they work and ho to prevent them.
Session 14	COMPUTER VISION AT THE EDGE How computer vision is a catalyst giving rise to revolutionary leaps in Internet of Things (IoT) innovations and applications.

Computer Vision: Assignments

Hands on projects using python

Session	Objective
Session 1	INTRODUCTION TO THE COURSE AND ITS ORGANIZATION
Session 2	BASICS OF IMAGE PROCESSING
Session 3	IMAGE CLASSIFICATION
Session 4	DEEP LEARNING FOR COMPUTER VISION
Session 5-6 Individual assignment 40%	TRANSFER LEARNING AND OBJECT DETECTION
Session 7	IMAGE SEGMENTATION Introduction to semantic and instance segmentation
Session 8	HUMAN POSE ESTIMATION
Session 9	OBJECT TRACKING
Session 10-11 Individual Assignment due date	GENERATIVE MODELS
Group assignment 40%	
Session 12	MULTIMODAL LEARNING
Session 13	ADVERSARIAL ATTACKS AND HOW TO PREVENT THEM IN IMAGES
Session 14 Group assignment due date	COMPUTER VISION AT THE EDGE
Session 15	GROUP PROJECT PRESENTATION

SCHOOL OF HUMAN SCIENCES & TECHNOLOGY

To make the most of this course

- Throughout this course, you will acquire the basic knowledge and required skills to process images and interact with the different python libraries that allow for building computer vision applications.
- You are required to **practice** solving the exercises mentioned in class to increase your **confidence** in dealing with the topic.
- Throughout the course, we will make **intensive** use of **Python**, and some **Python Computer vision libraries**.

ie
SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

Introduction to Computer Vision

What is Computer vision?

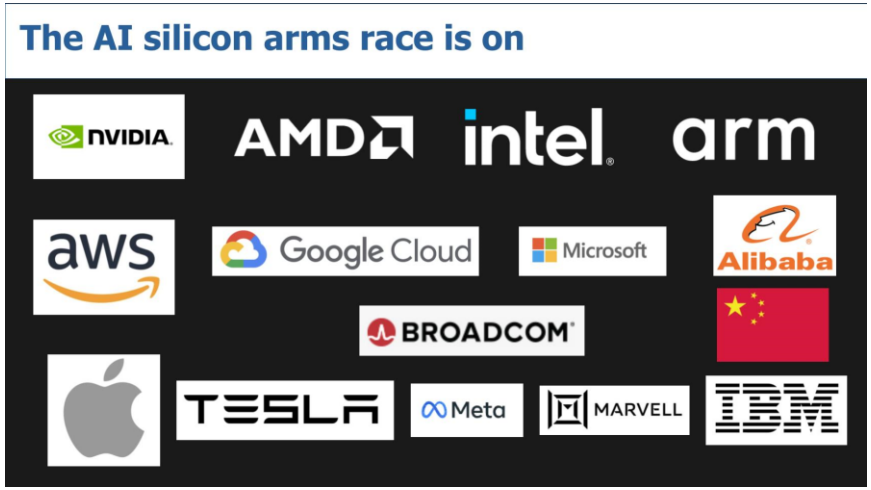
- Make computers understand images and videos.
- Infer useful information, interpret and make decisions
- It automates tasks that the human visual system can do.
- One of the most exciting fields of AI and ML

- Where is this?
- What are they doing?
- Who are these people?
- What are their emotions?
- What is their age/gender?

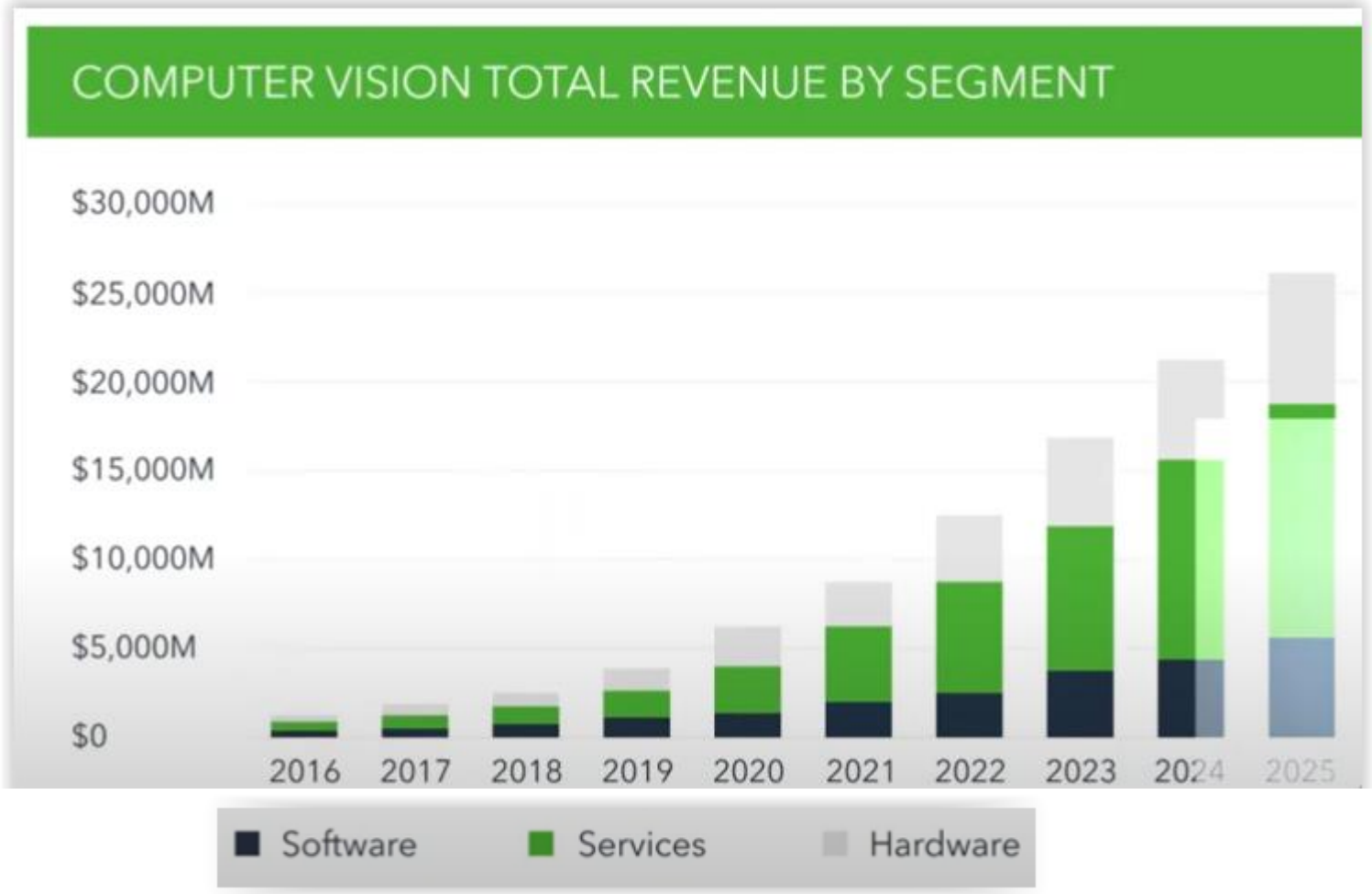
Computer Vision: Motivation

- Emergence of deep learning
- Advancement in Hardware
- Availability of large-scale data:
  - imageNet
  - OpenImages
  - YFCC100M
  - Youtube-8M
  - Kinetics

  …

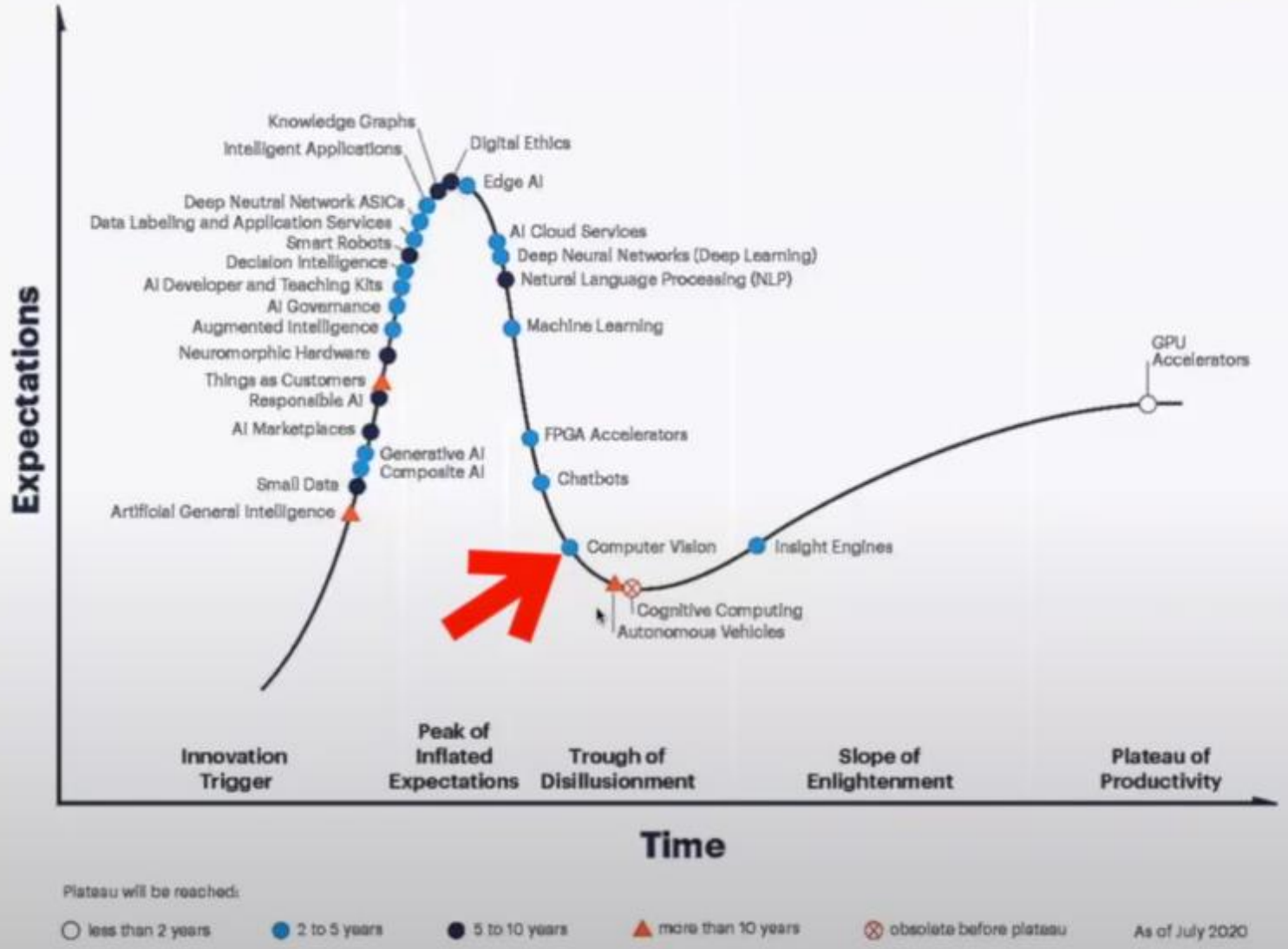


Market of Computer Vision



- Source: tractica.com

Hype Cycle for Artificial Intelligence, 2020



gartner.com/SmarterWithGartner

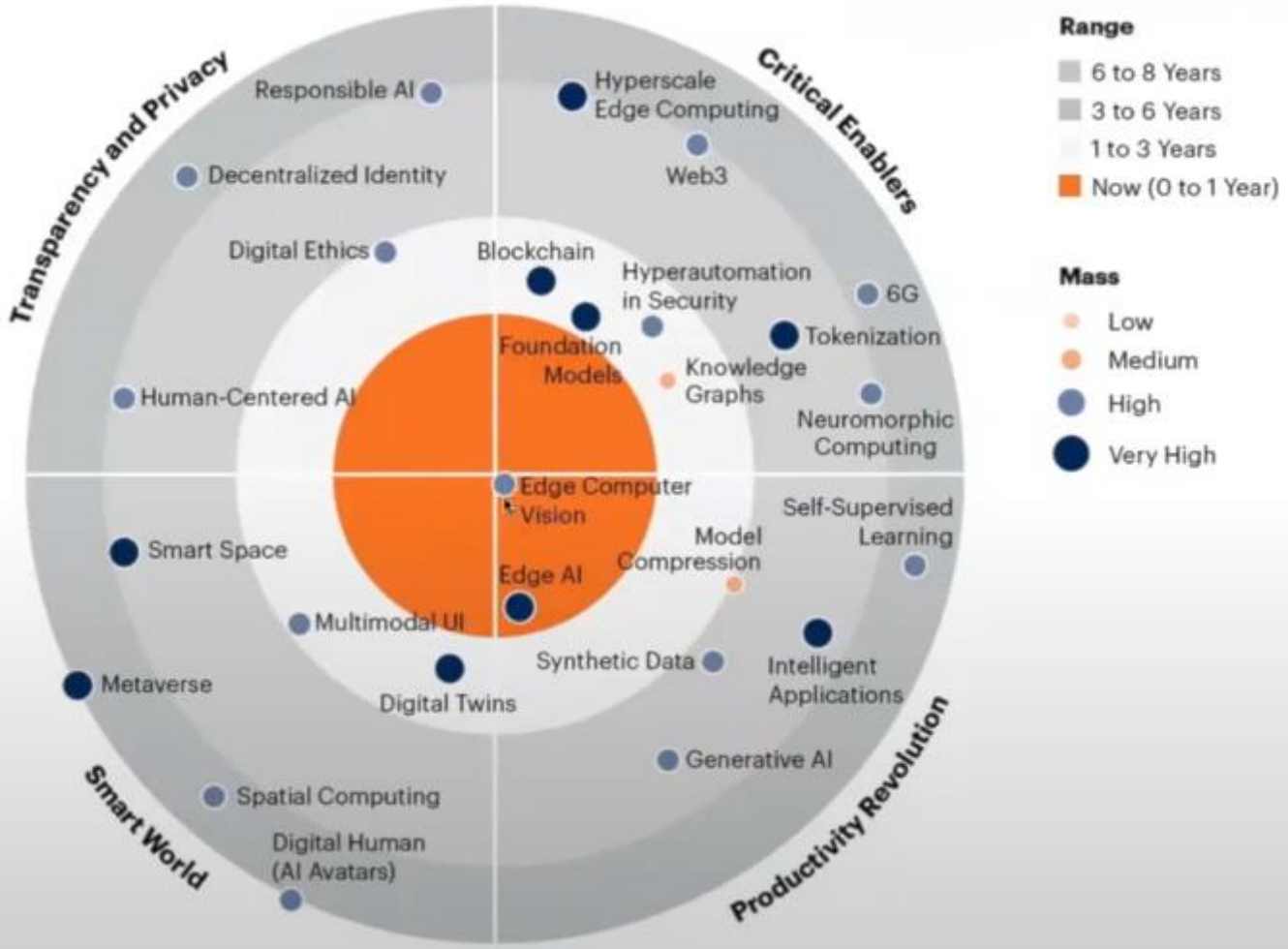
Gartner.

Hype Cycle for Artificial Intelligence, 2022

gartner.com

Gartner

2023 Gartner Emerging Technologies and Trends Impact Radar



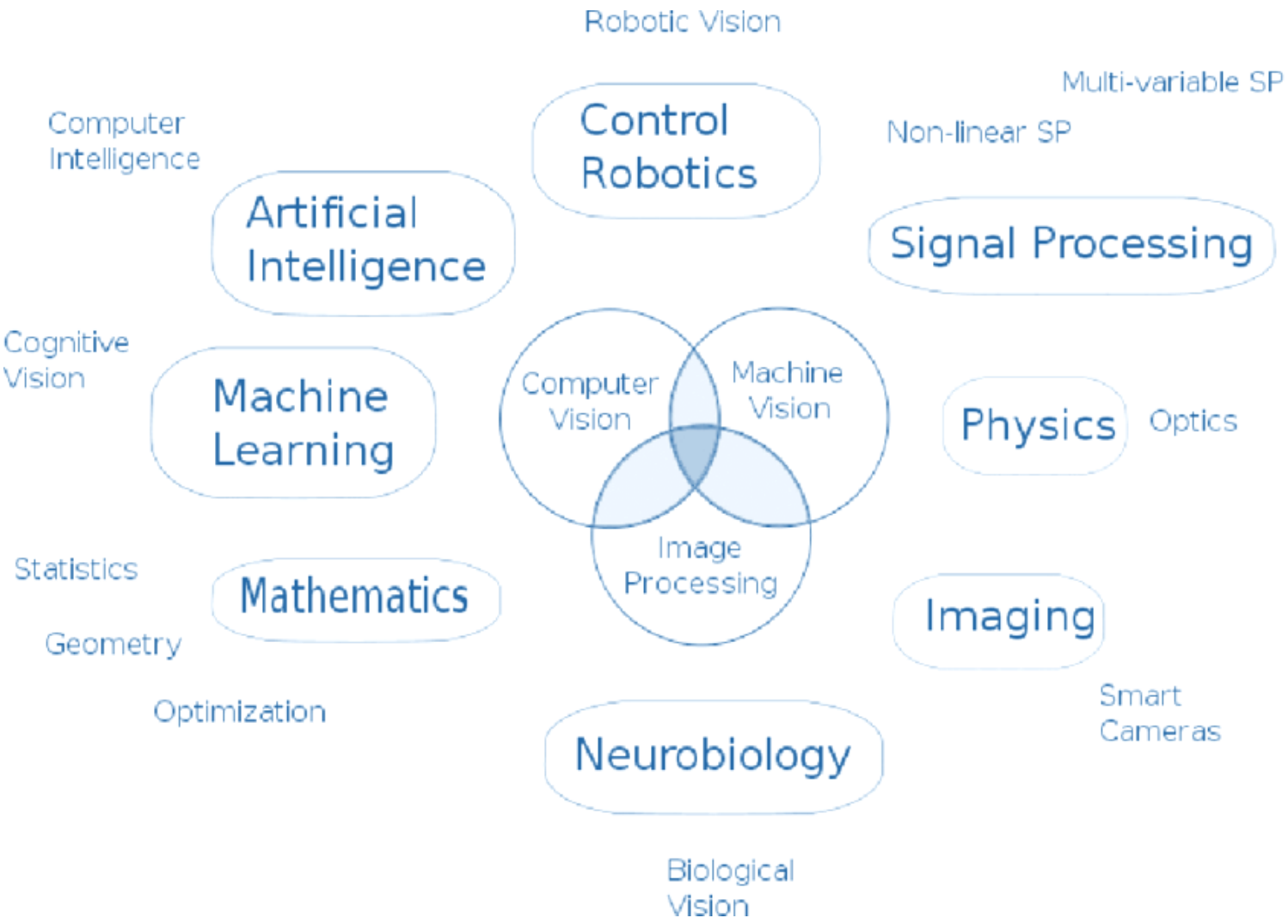
gartner.com

Note: Range measures number of years it will take the technology/trend to cross over from early adopter to early majority adoption. Mass indicates how substantial the impact of the technology or trend will be on existing products and markets.

Source: Gartner

Gartner

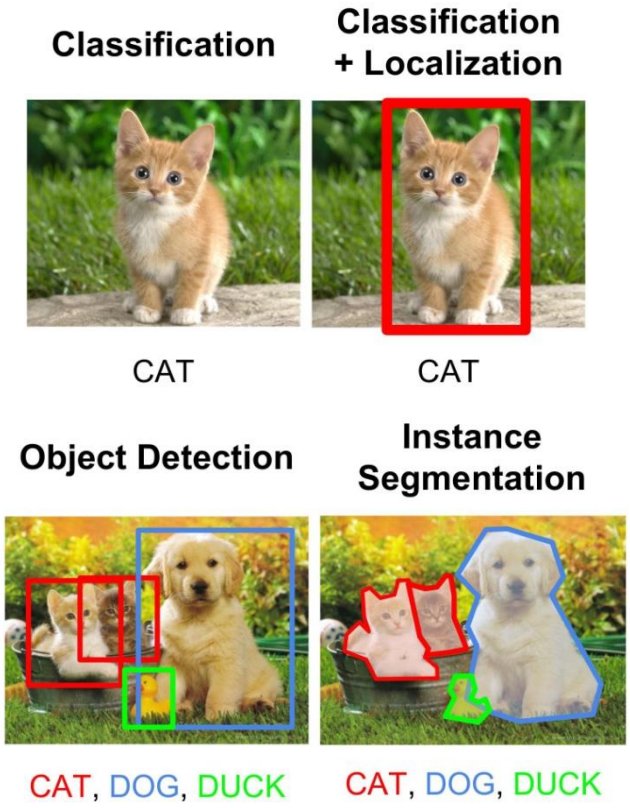
Computer Vision: Intersection of various fields



Computer Vision: Typical tasks

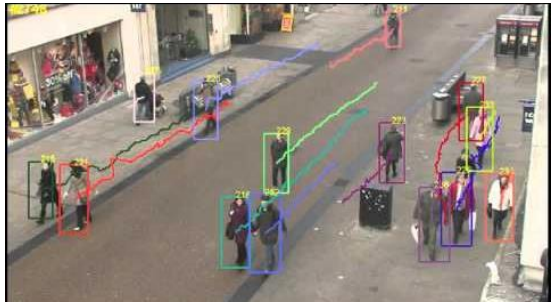
Focus of our course

- Image Classification
- Object Detection
- Object Segmentation



- Pose estimation

- Object tracking

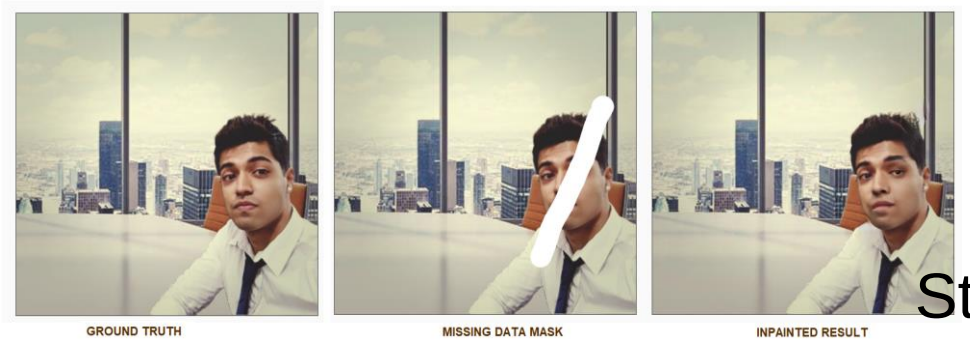


Computer Vision: Typical tasks

Focus of our course

- Generative AI.

inpainting



Coloring photos

Style transfer

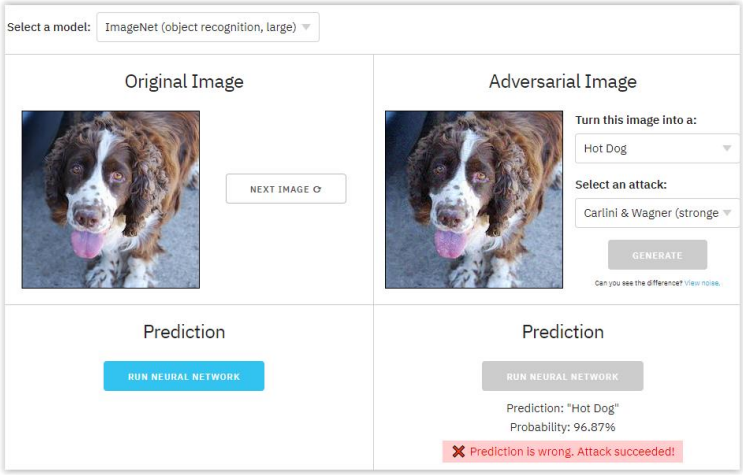
Computer Vision: Typical tasks

Focus of our course

- Text to image

- Adversarial attacks

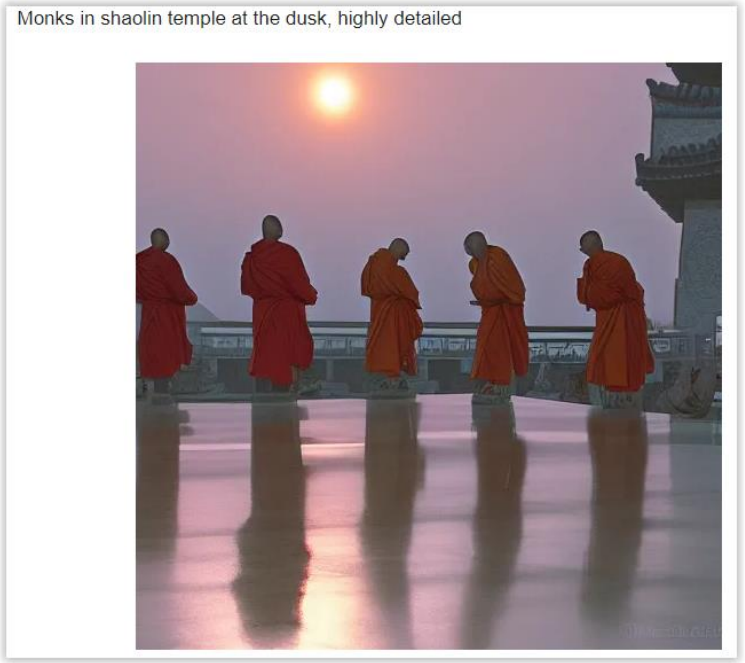
https://kennysong.github.io/adversarial.js



Computer Vision: Typical tasks

Focus of our course

- Text to image

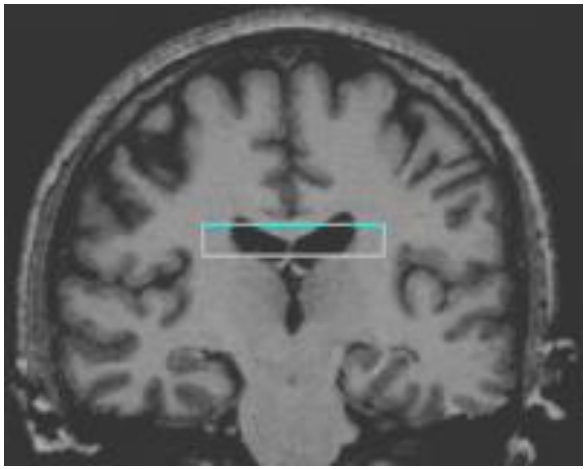


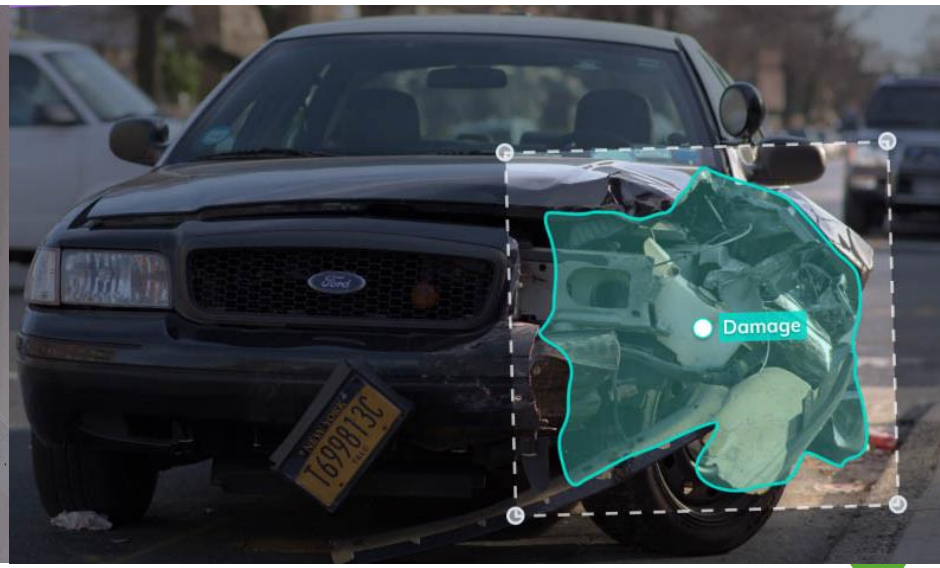
- Text to video

Video Generation-Sora

Computer Vision: Typical applications by Vertical

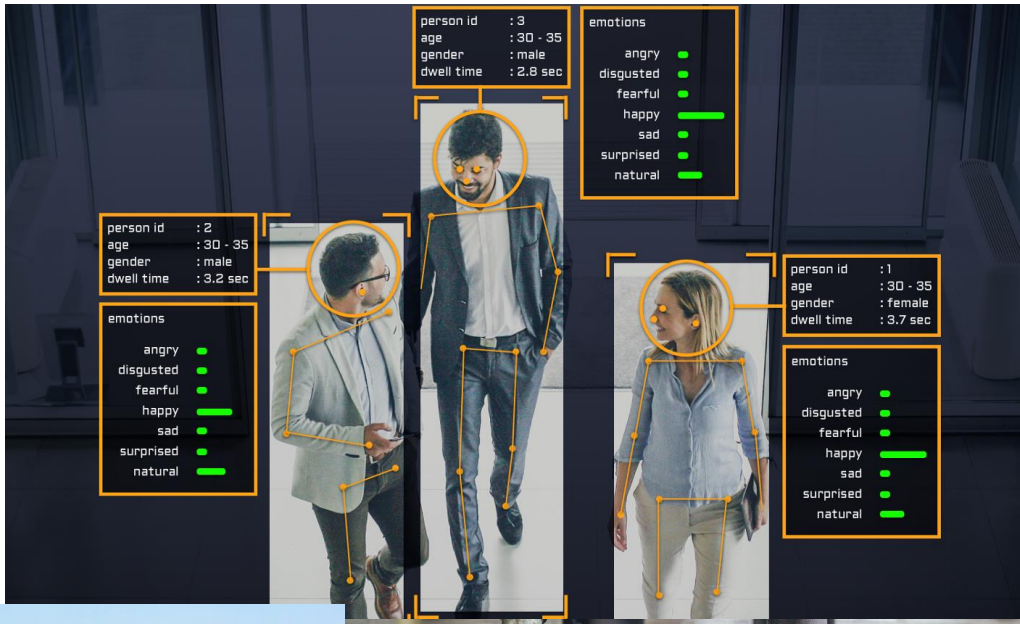
- Healthcare: Diagnosis
- Retail: Product layout, inventory management, theft prevention
- Insurance: Car Damage estimation





Computer Vision: Typical applications by Vertical

- Construction: Construction progress monitoring
- Hospitality/Service: Customer queue monitoring, customer satisfaction
- Energy: Predictive

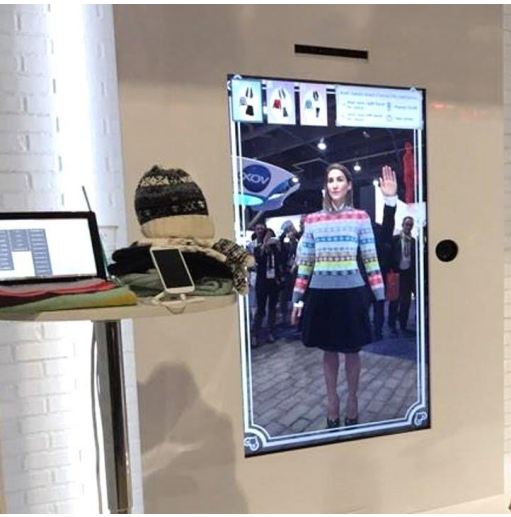


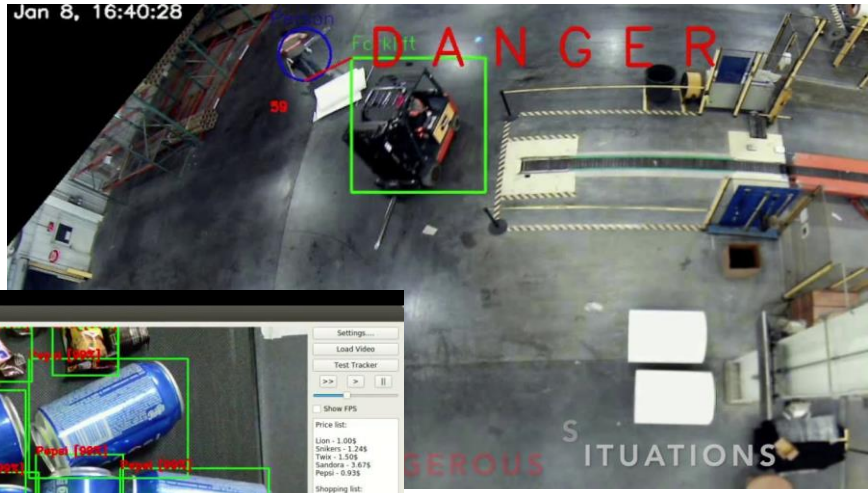
Computer Vision: Typical applications by Vertical

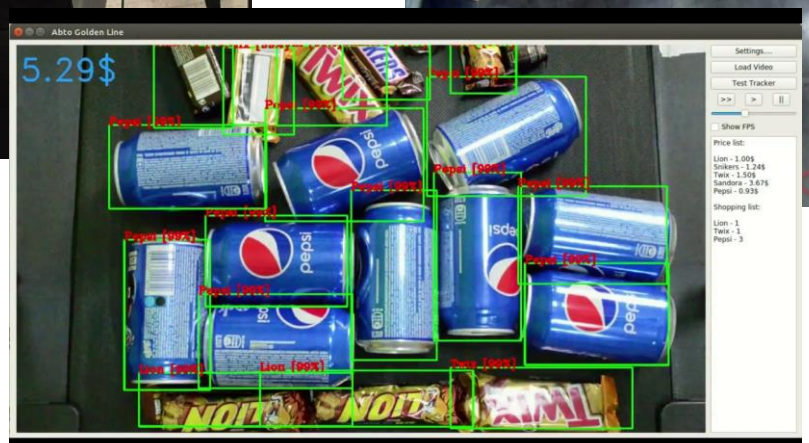
- Entertainment: Filters, augmented reality
- Manufacturing: Quality control , product sorting
- Agriculture: Quality of yield
- Surveillance/military: identity, biometrics, access control
- Transport: assistance to driver, navigation, autonomous vehicles

Computer Vision: Typical applications by Vertical

- Safety: Reduce unsafe behavior in work sites
- Retail: Virtual fitting rooms, cashier-less stores



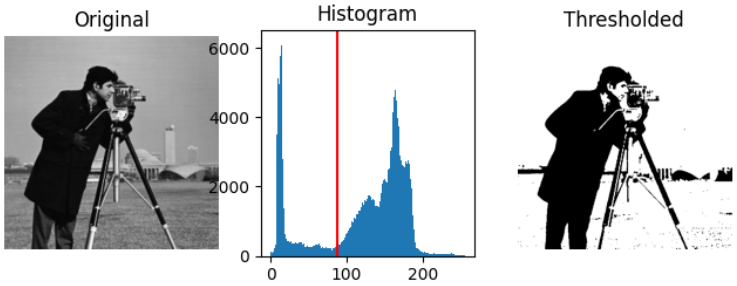




The old era of Computer Vision

Hand crafted features

- Thresholding



- Edge detection (Canny, Sobel)

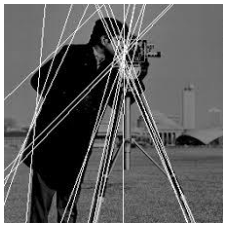
- Sliding windows filters

Figure 4: result of Gaussian low pass filter

Figure 5: Gaussian high pass filter

- Corner detection (Harris, Shi & Tomasi)

- Ridge detection (Hough transform)



- Featre descriptions
  - Histogram of oriented gradients (HOG)
  - Scale invariant feature transform (SIFT)

The old era of Computer Vision

- Computer vision typical workflow
- 1. Image acquisition (camera, sensors)
- 2. Pre-processing (sampling, noise reduction, hist equalization)
- 3. Feature extraction (lines, edges, regions, points)
- 4. Detection and segmentation
- 5. Post-processing (verification, estimation, recognition)
- 6. Decision making: Ability of a machine to step back and interpret the big picture of those pixels

Limitation

- Computationally expensive (hundreds of features in image, millions in database) and complex due to errors and mismatches

The new era of Computer Vision

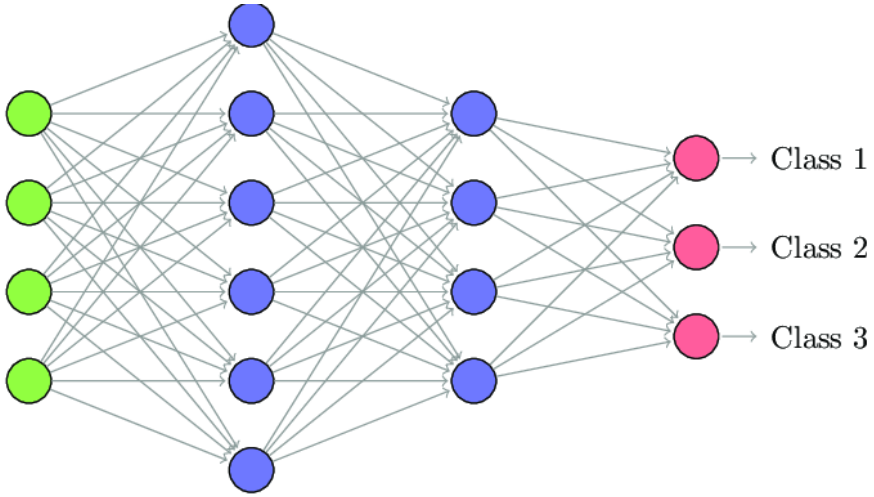
- Artificial neural networks
  - Fundamentals - artificial neuron

- Elementary building block
- Inspired by biological neurons
- Mathematical function y=f(wx+b)
- Learnable weights

The new era of Computer Vision

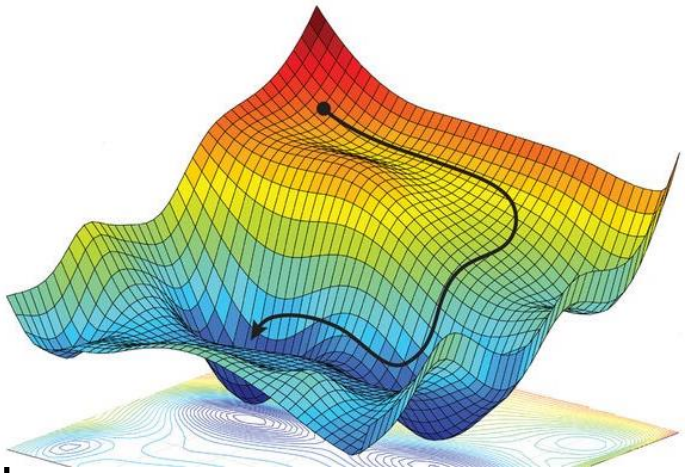
- Artificial neural networks
  - Fundamentals - artificial neural networks

- Collection of neurons organized in layers
- Universal approximators



The new era of Computer Vision

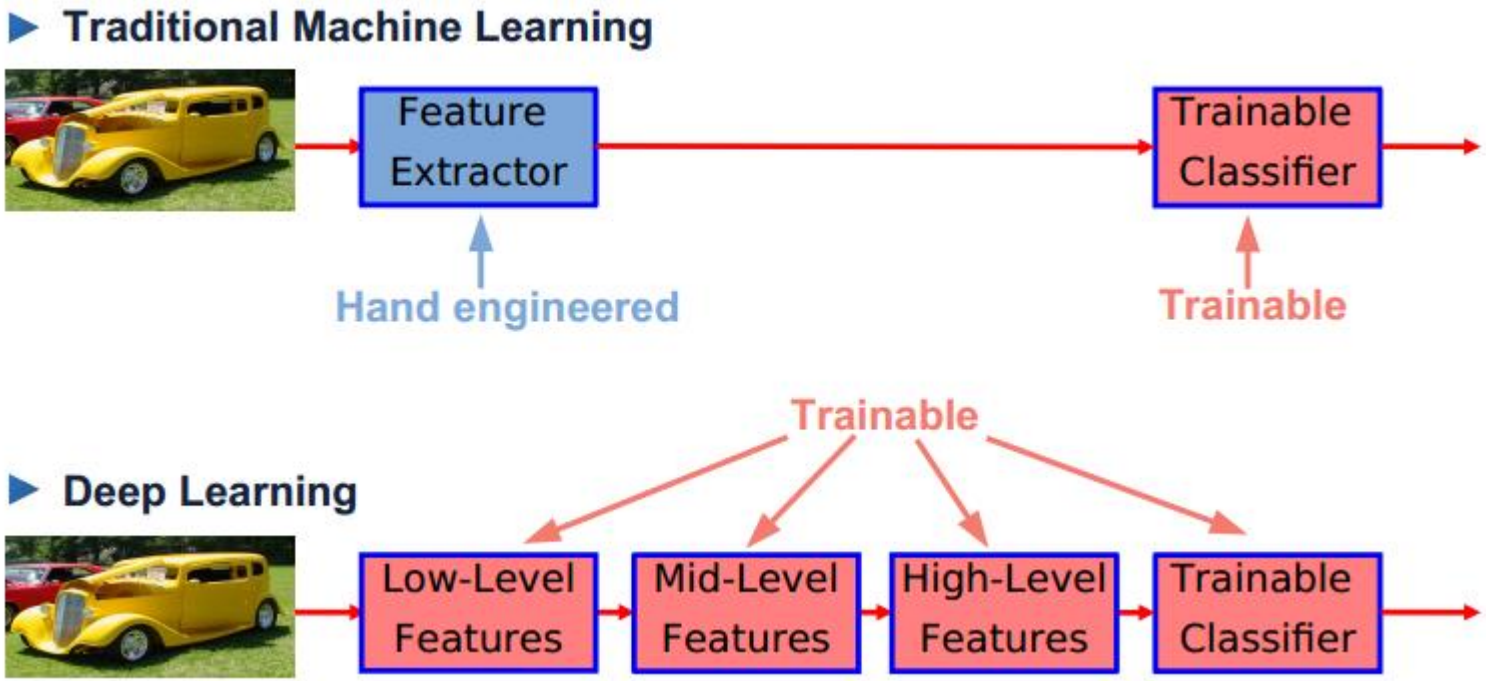
- Artificial neural networks
  - Fundamentals - training
- Basically, an optimization problem
- Find minimum of a loss function by an iterative process (training)
- Designing the loss function is sometimes tricky



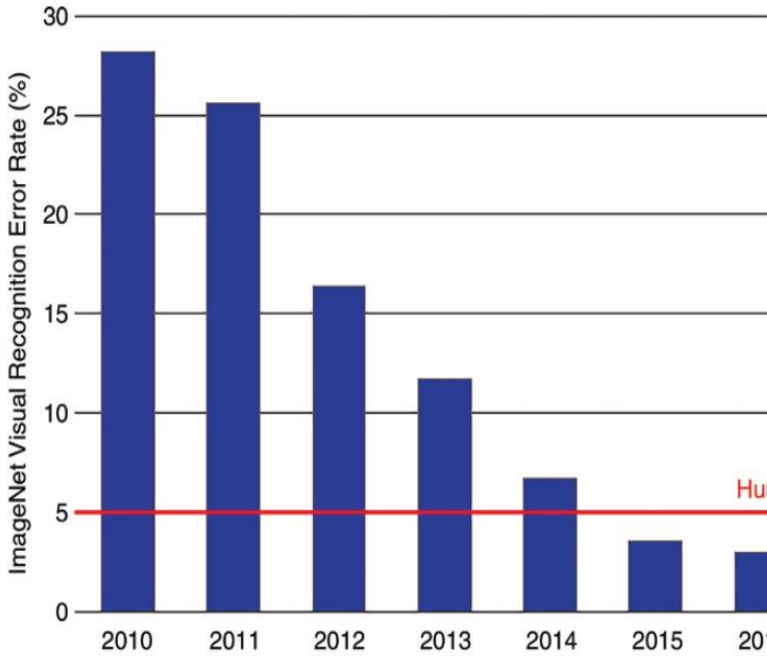
Simple optimizer algorithm:

1. Forward pass with a batch of data
2. Calculate error between actual and wanted output
3. Nudge weights in proportion to error into the correct direction (same data would result in smaller error)
4. Repeat until convergence

Then and Now

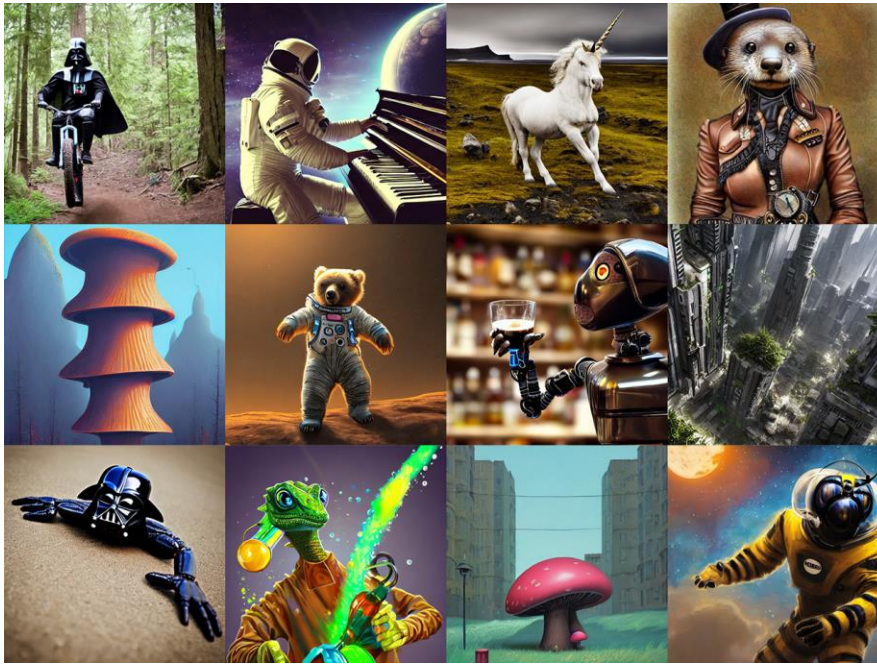

Traditional Machine Learning
Feature Extractor
Trainable Classifier
Hand engineered
Trainable
Deep Learning
Trainable
Low-Level Features
Mid-Level Features
High-Level Features
Trainable Classifier

Today

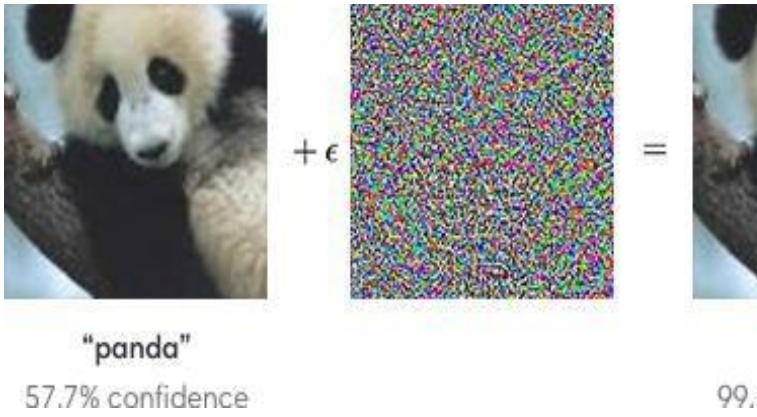


Computer Vision Algorithms

- Better than humans for classifying image and video
- Better than a person to recognize a person
- Generate almost real footage, and very artistic images

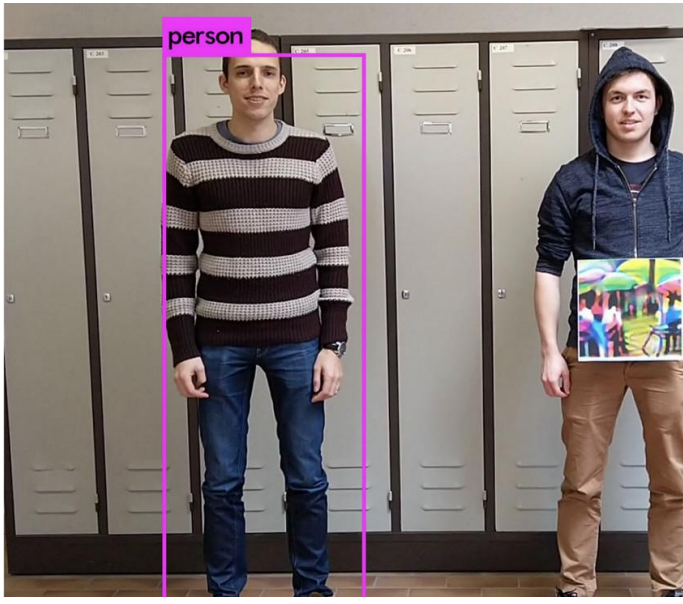


Today



BUT

- "Confusing" or "fooling" the neural network called making a **physical adversarial attack** or a real-world adversarial attack is possible
- These attacks, initially based on intricately altered pixel values, confuse the network (based on its training data) into labeling the object differently or as "unknown" or simply ignoring it.




How AI will understand the real world
TED

ie
SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

Chapter Summary

- Presented course objective
- Presented course timeline
- Got an introduction to computer vision and its applications

Computer Vision: Milestones

- 1959 – The first digital image scanner was invented by transforming images into grids of numbers.
- 1963 – Larry Roberts, the father of CV, described the process of deriving 3D info about solid objects from 2D photographs.
- 1966 – Marvin Minksy instructed a graduate student to connect a camera to a computer and have it described what it sees.
- 1980 – Kunihiko Fukushima built the 'neocognitron', the precursor of modern Convolutional Neural Networks.
- 1991-93 – Multiplex recording devices were introduced, together with cover video surveillance for ATM machines.
- 2001 – Two researchers at MIT introduced the first face detection framework (Viola-Jones) that works in real-time.
- 2009 – Google started testing robot cars on roads.
- 2010 – Google released Goggles now google lens, an image recognition app for searches based on pictures taken by mobile devices.
- 2010 – To help tag photos, Facebook began using facial recognition.
- 2011 – Facial recognition was used to help confirm the identity of Osama bin Laden after he is killed in a US raid.

Computer Vision: Milestones

- 2012 – Google Brain's neural network recognized pictures of cats using a deep learning algorithm.
- 2015 – Google launched open-source Machine learning-system TensorFlow.
- 2016 – Google DeepMind's AlphaGo algorithm beat the world Go champion.
- 2017 – Waymo sued Uber for allegedly stealing trade secrets.
- 2017 – Apple released the iPhone X in 2017, advertising face recognition as one of its primary new features.
- 2018 – Alibaba's AI model scored better than humans in a Stanford University reading and comprehension test.
- 2018 – Amazon sold its real time face recognition system Rekognition to police departments.
- 2019 – The Indian government announced a facial recognition plan allowing police officers to search images through mobile app.
- 2019 – The US added four of China's leading AI start-ups to a trade blacklist.(sensetime…)
- 2019 – The UK High Court ruled that the use of automatic facial recognition technology to search for people in crowds is lawful.
- 2020 – Intel launched the Intel Xe graphics card pushing into the GPU market.
- 2025 – By this time, regulation will significantly diverge between China and US/Europe.
- 2030 – At least 60% of countries globally will be using AI surveillance technology