清華大學
Tsinghua University

商汤
sensetime

THU×SENSETIME - 80231202

Advanced Computer Vision

Friday, February 26, 2021

商汤
sensetime

清華大学
Tsinghua University

Content

Part 1	Course Introduction
Part 2	What's Computer Vision
Part 3	History of Computer Vision

Course Introduction

Overview

This course involves **computer vision**, **deep learning** and other fields of knowledge. It elaborates with the latest academic achievements and practical cases of industrial scenes and explain the classic and state-of-the-art methods in computer vision.

What we have

- Focus on Both Classics and Frontiers
- Combination of Academia and Industry
- Teaching from the shallower to the deeper
- GPU clusters for experiments

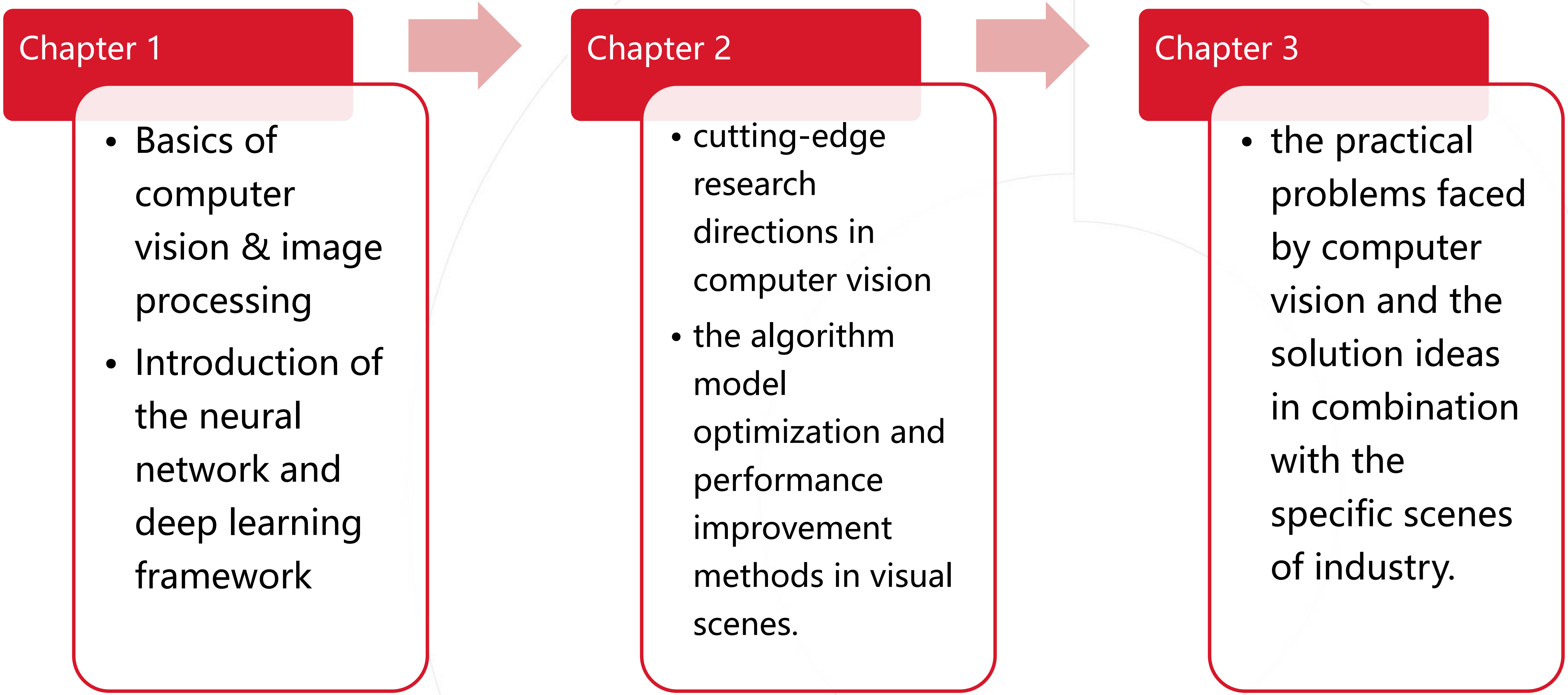
What you will learn

- Basic theories and advanced methods in Computer Vision
- Understand and explore practical problems in the industry
- Improve your research ability and innovative ability

What you need

- **Mathematics**
  - Calculus
  - Linear Algebra
  - Basic Probability and Statistics
- **Coding ability**
  - **Python** is recommended
- Machine Learning

Course Introduction



Chapter 1

- Basics of computer vision & image processing
- Introduction of the neural network and deep learning framework

Chapter 2

- cutting-edge research directions in computer vision
- the algorithm model optimization and performance improvement methods in visual scenes.

Chapter 3

- the practical problems faced by computer vision and the solution ideas in combination with the specific scenes of industry.

Syllabus

Chapter 1 - Computer Vision Overview and Deep Learing Basics

1.Computer Vision Basics

2.Image and Video Processing

3.Feature Detection

4.CNN & High-level Feature Extraction

5.Training Framework and Model Optimization

Chapter 2 - Advanced Computer Vision Tasks

6.Image Classification

7.Object Detection

8.Image Segmentation

9.Video Understanding and Sequence Analysis

10.Low-Level Computer Vision Task

11.Neural network Model Acceleration and Compilation

12.3D Vision

13.Representation Learning in Vision Tasks

Chapter 3 - Lectures on industry applications

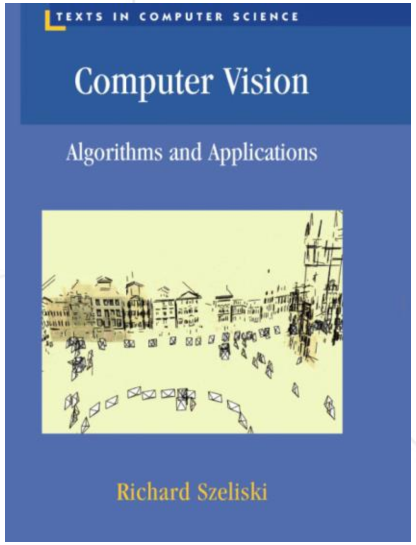
14.Smart City

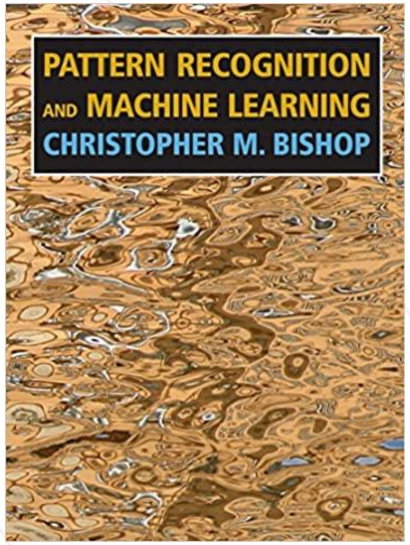
15.AutoPilot

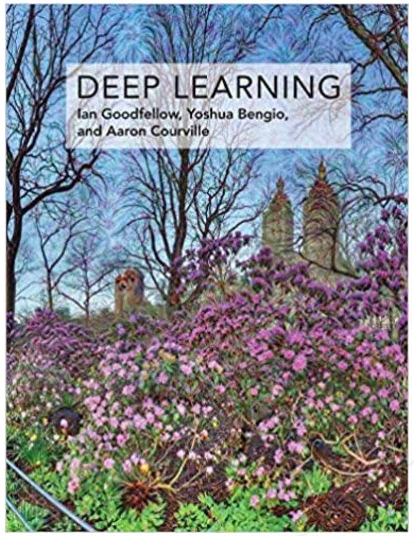
16.3D Vision and Augmented Reality

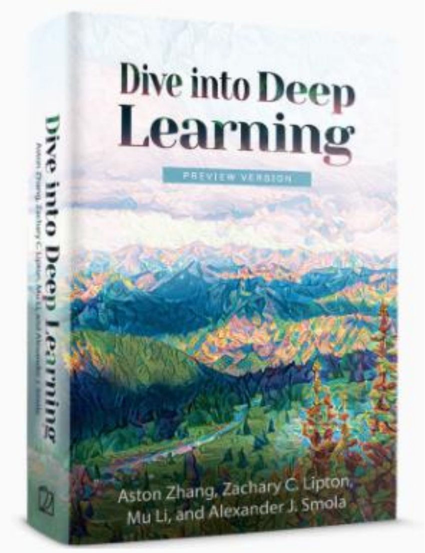
Course Introduction

- **Textbook**
  - ***Computer Vision Algorithms and Applications***
    - by Richard Szeliski
    - Preview version： [Link]
  - ***Pattern Recognition and Machine Learning***
    - by Christopher Bishop
    - Free online version： [Link]
  - ***Deep Learning***
    - by Goodfellow, Bengio, and Courville
    - Index： [Link]
  - ***Dive into deep learning***
    - An interactive deep learning book with code, math, and discussions, based on the NumPy interface
    - Free online version： [Link]








Dive into Deep Learning
PREVIEW VERSION
Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola

Course Introduction

- **Grading Policy**

Quizzes

(20%)

2 Quizzes in class, completed by one person

Assignments

(30%)

3 Assignments finish after class by one person

Final Project

(50%)

Choose one topic , submit 3-4 pages paper and make a oral presentation during the seminar.

Collaboration in groups of up to three people.

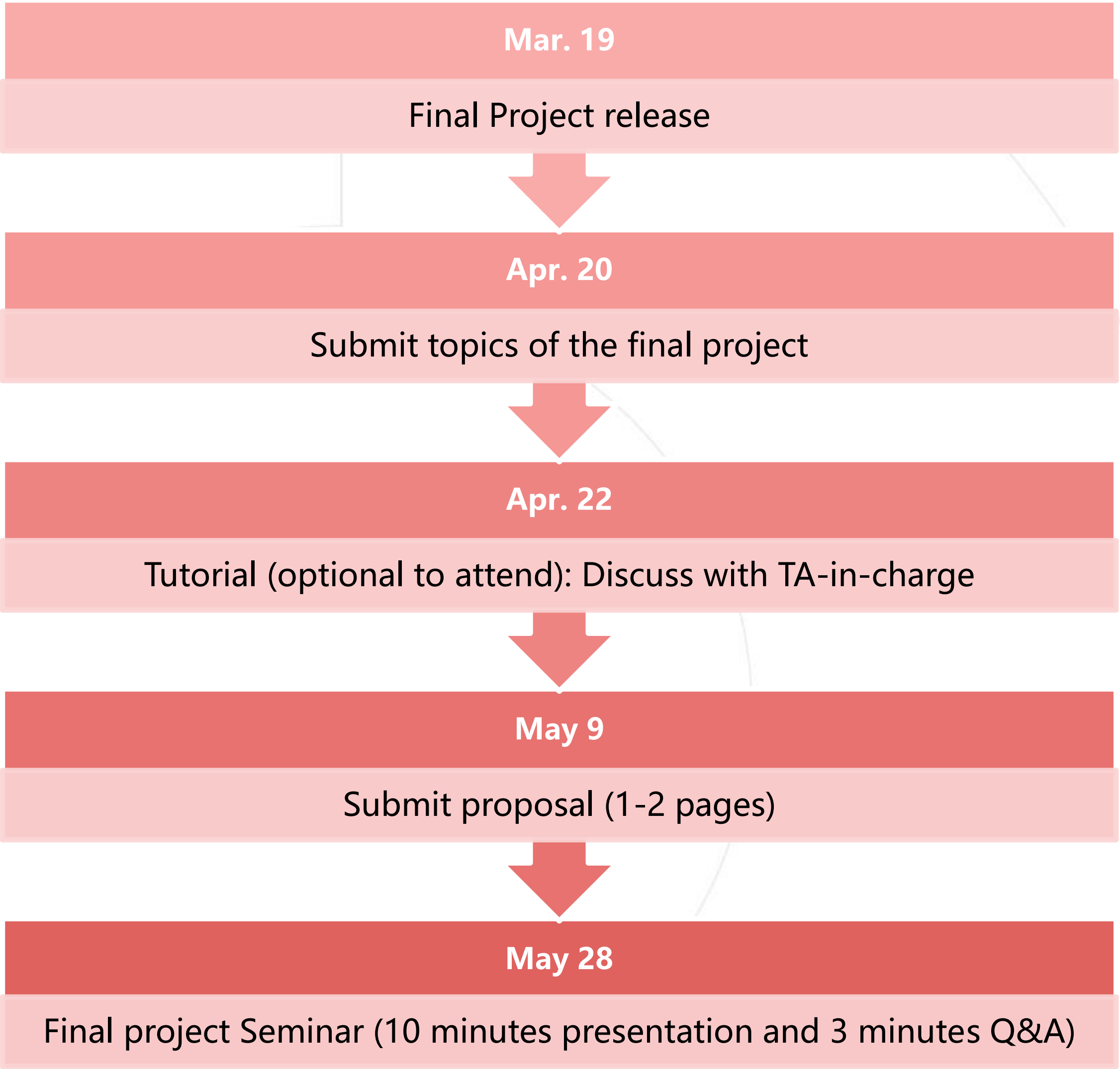
Course Introduction

- **Assignment**
  - All assignments should be finished by one person
  - You can finish assignment on your local machines or on clusters provided by SenseTime
  - More details will be update on Course Homepage

Assignment	Released Date	Due Date	Topic
Assignment 1	Mar. 12	Mar. 26	Deep learning training framework and model optimization implementation
Assignment 2	Apr. 2	Apr. 16	Advanced Computer Vision Tasks
Assignment 3	May. 7	May. 14	Lightweight Model Quantization and Model Pruning

Course Introduction

- **Final Project**
  - Choose one topic and finish the project
  - You should submit
    - One page proposal and discuss it with TAs (topic, idea, method, experiments)
    - A term paper of 4 pages (excluding figures) in maximum, double column, font size is equal or larger than 10
    - Code and sample data
    - Project presentation
  - Collaboration in groups of up to three people



Mar. 19

Final Project release

Apr. 20

Submit topics of the final project

Apr. 22

Tutorial (optional to attend): Discuss with TA-in-charge

May 9

Submit proposal (1-2 pages)

May 28

Final project Seminar (10 minutes presentation and 3 minutes Q&A)

Course Introduction

- **Instructors**

Dr. Li Yali	liyali13@mail.tsinghua.edu.cn
Dr. Dai Jifeng	daijifeng@sensetime.com
Dr. Liu Yu	liuyu@sensetime.com
Dr. Li Hongyang	lihongyang@sensetime.com

- **TAs**

Wang Han	i@hann.wang
Wang Cheng	wangcheng@senseauto.com
Song Guanglu	songguanglu@sensetime.com
Niu Yazhe	niuyazhe@sensetime.com

Course Introduction

- **Lecture Time & Venue**
  - **Friday**, 9:50am-11:25am
  - **4203**, No.4 Teaching Building
- **Optional Tutorials & QA Time**
  - **Thursday**, 19:00-20:00
  - Tencent Meeting Room：785 271 5223
- **Course Homepage**
  - https://thu-acv.github.io
- **Discussions**
  - WeChat Group
  - Tencent Meeting Room：785 271 5223

THU 高等计算机视觉 课程群

商汤泰坦小助手

中国

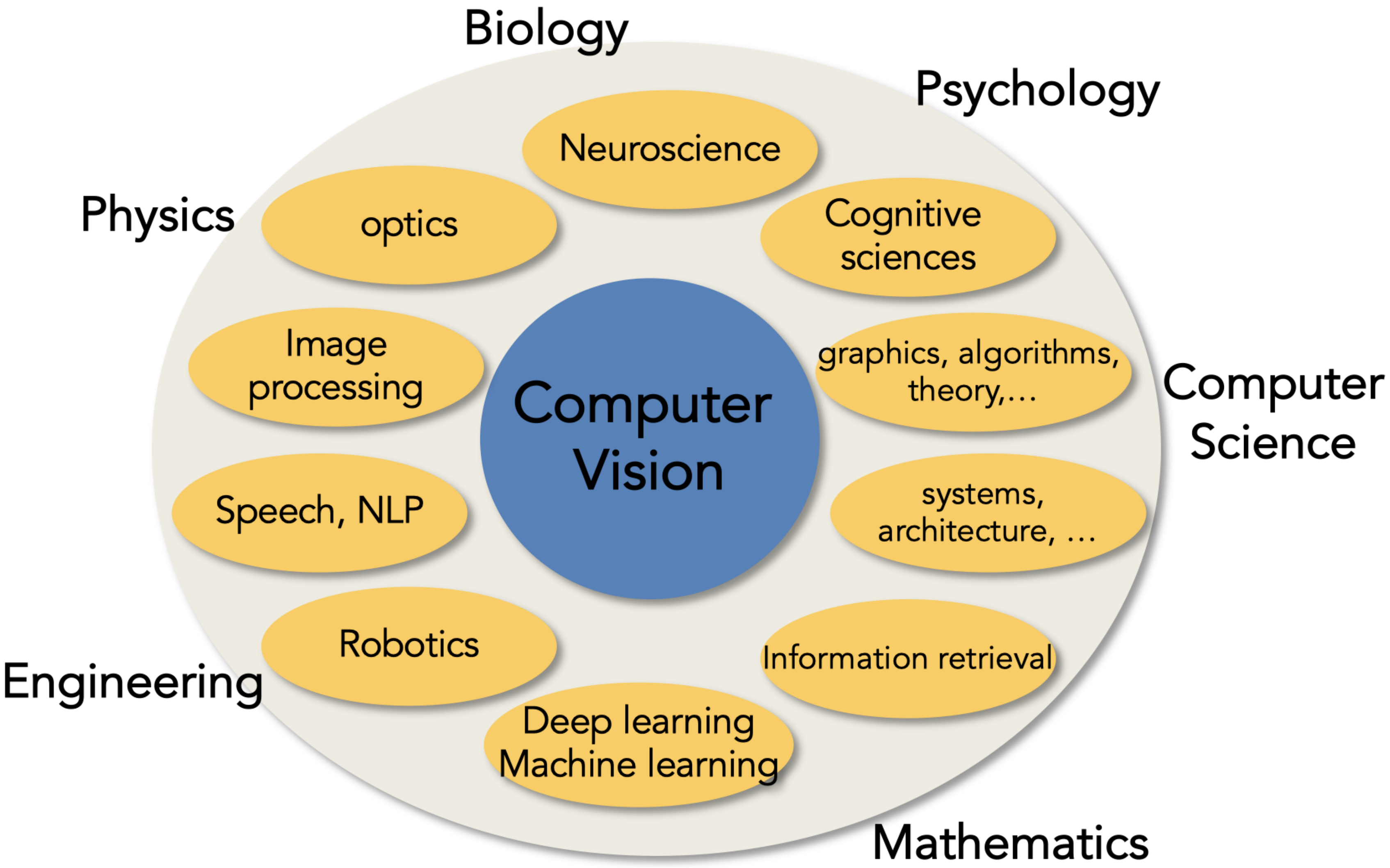
Content

Part 1 Course Introduction

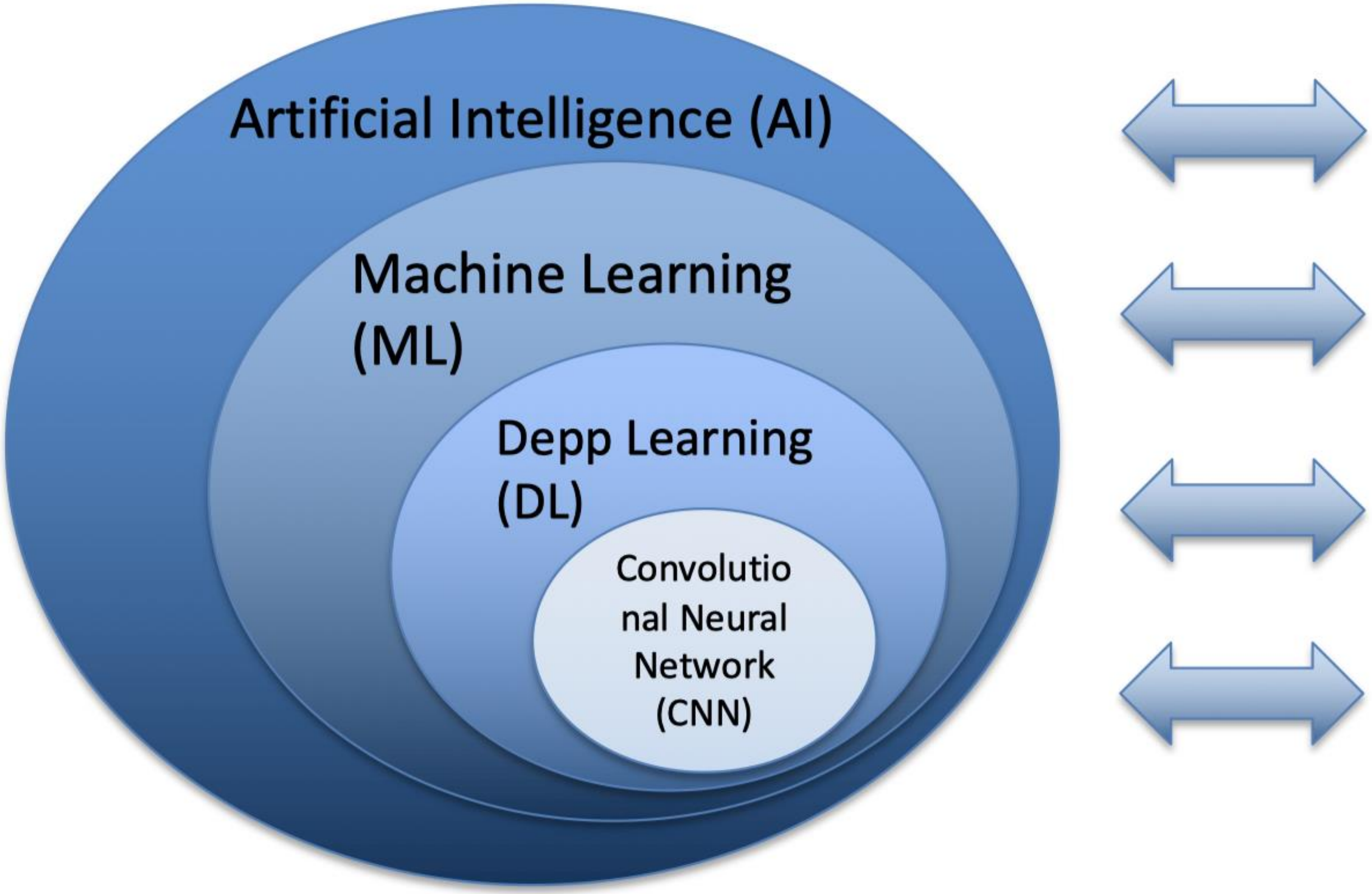
Part 2 What's Computer Vision

Part 3 History of Computer Vision

What's Computer Vision



What's Computer Vision



Artificial Intelligence (AI)

Machine Learning (ML)

Depp Learning (DL)

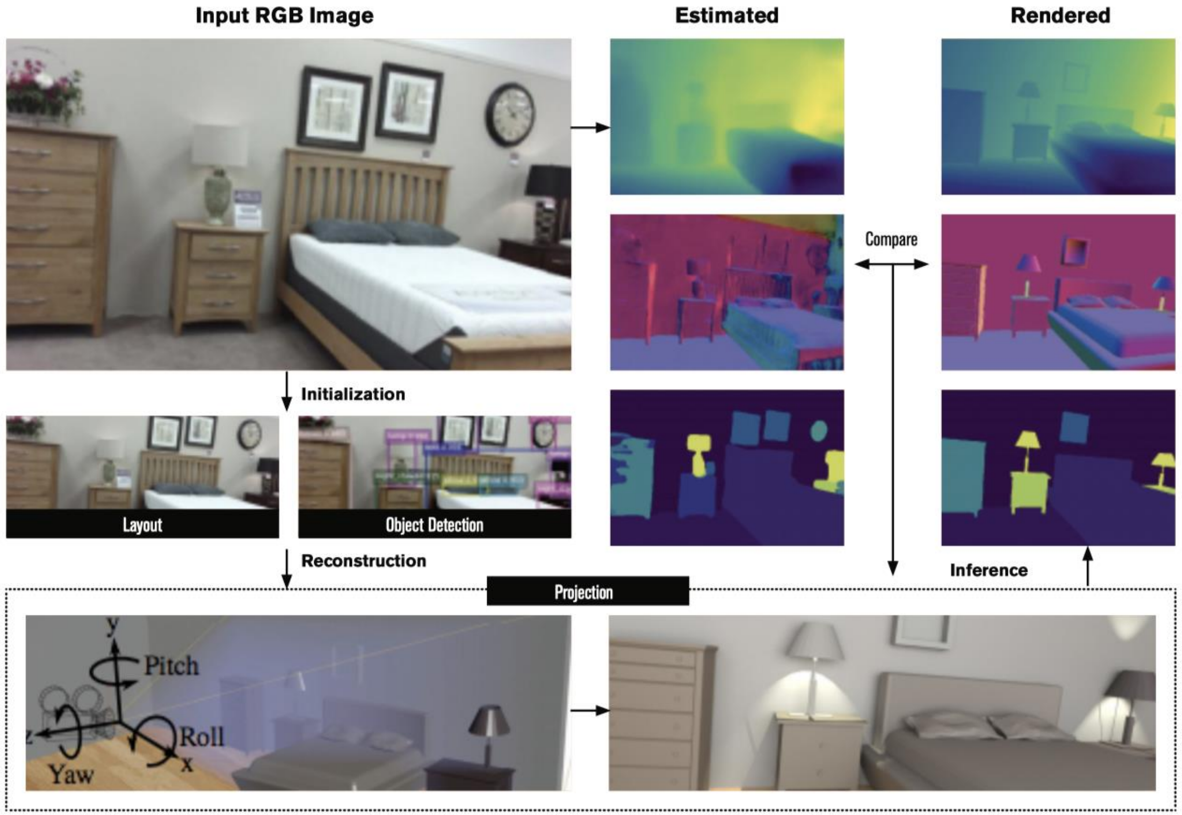
Convolutional Neural Network (CNN)

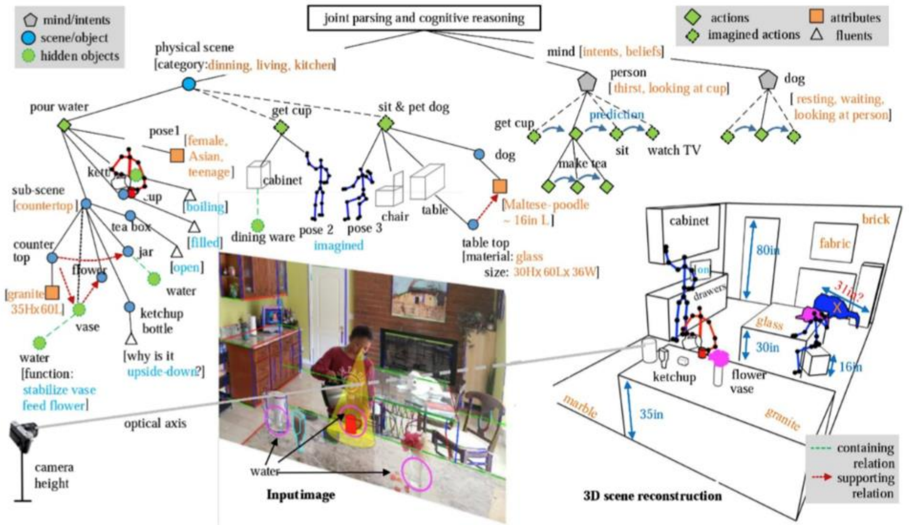
Computer Vision

- Object detection
- Object classification
- Scene understanding
- Semantic scene segmentation
- 3D reconstruction
- Object tracking
- Human pose estimation
- Activity recognition
- VQA
- ....

What's Computer Vision

Vision is the most important source of information for the human brain and is the “**entrance hall**” of AI.





商汤
sensetime

清華大學
Tsinghua University

Content

Part 1 Course Introduction

Part 2 What's Computer Vision

Part 3 History of Computer Vision

History of Computer Vision

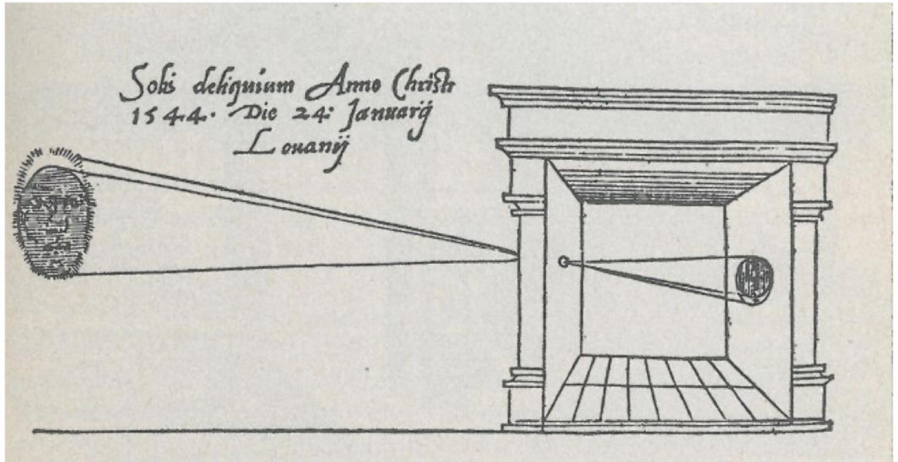
- **Biological Vision**

History of Computer Vision

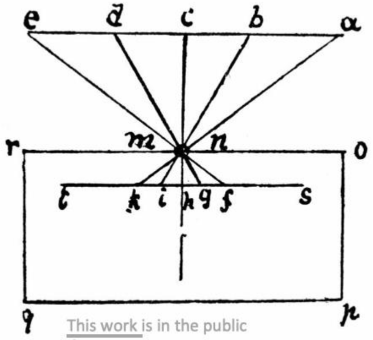
- **Ancient Human Vision**

Camera Obscura

Gemma Frisius, 1545



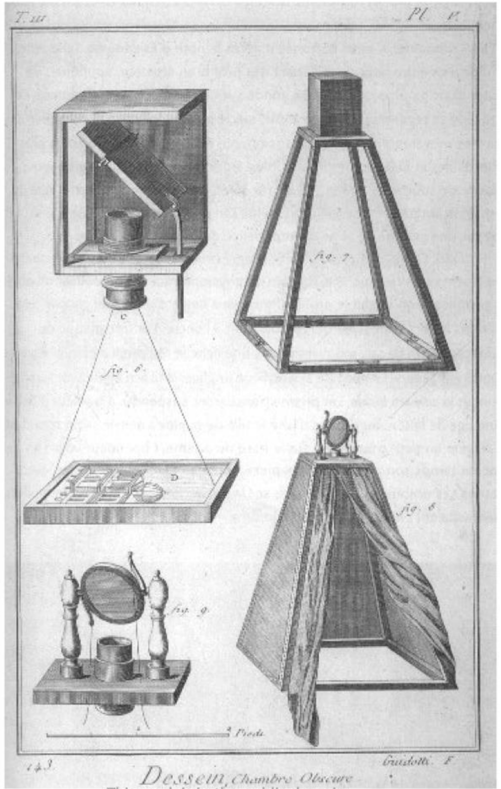
This work is in the public domain



This work is in the public domain

Leonardo da Vinci, 16th Century AD

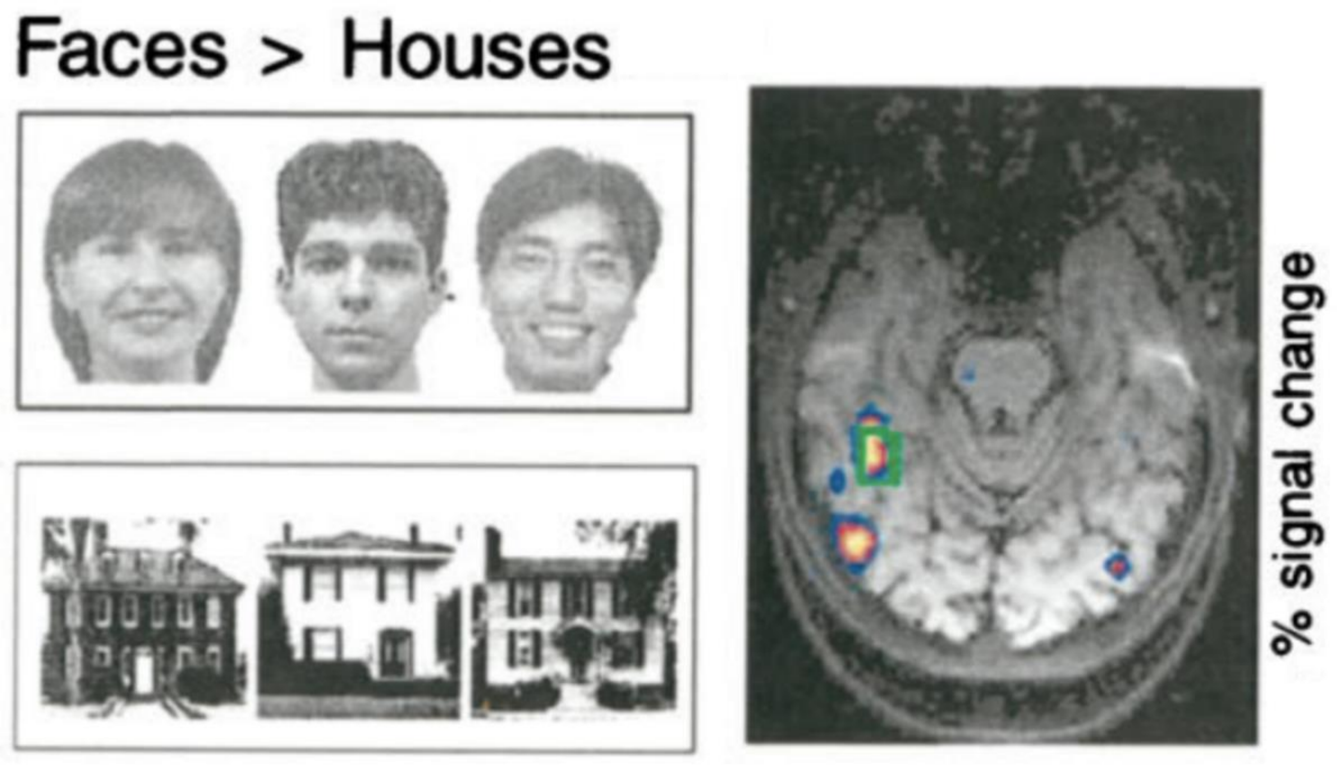
Encyclopedia, 18th Century



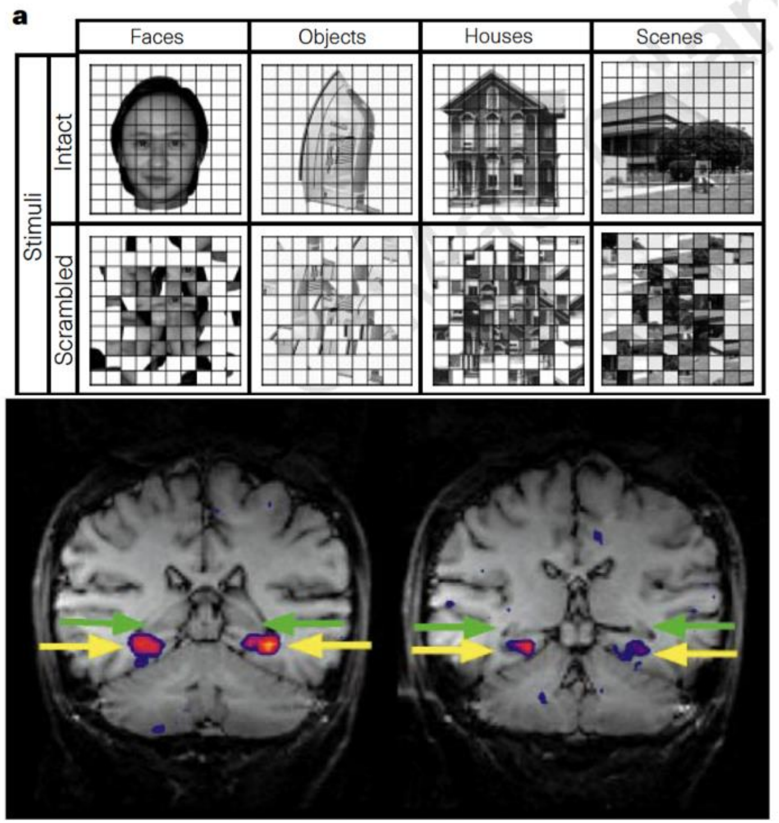
This work is in the public domain

History of Computer Vision

- **Neuroscience and Vision**



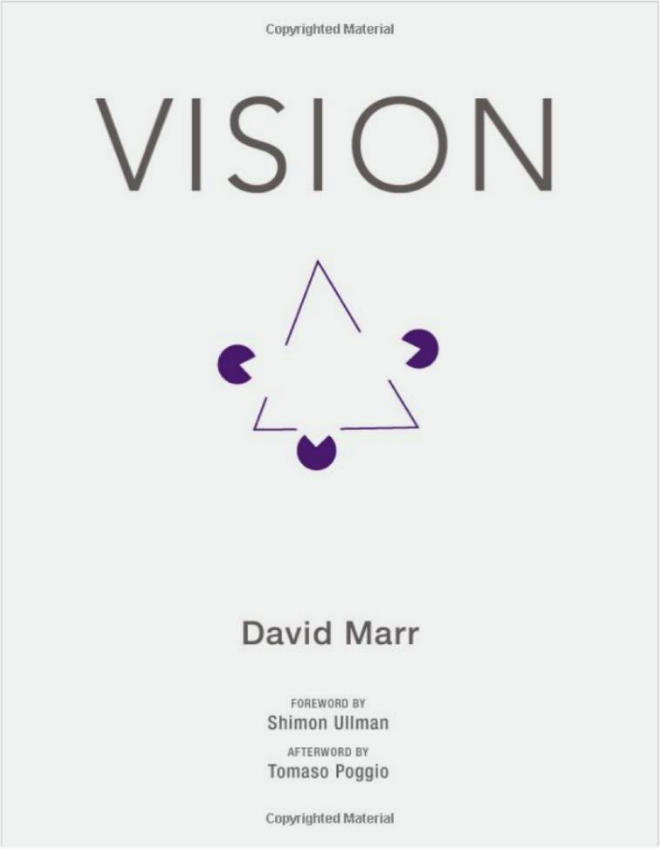
Kanwisher et al. J. Neuro. 1997



Epstein & Kanwisher, Nature, 1998

History of Computer Vision

- **Marr Computational Vision**



Copyrighted Material

VISION

David Marr

FOREWORD BY
Shimon Ullman

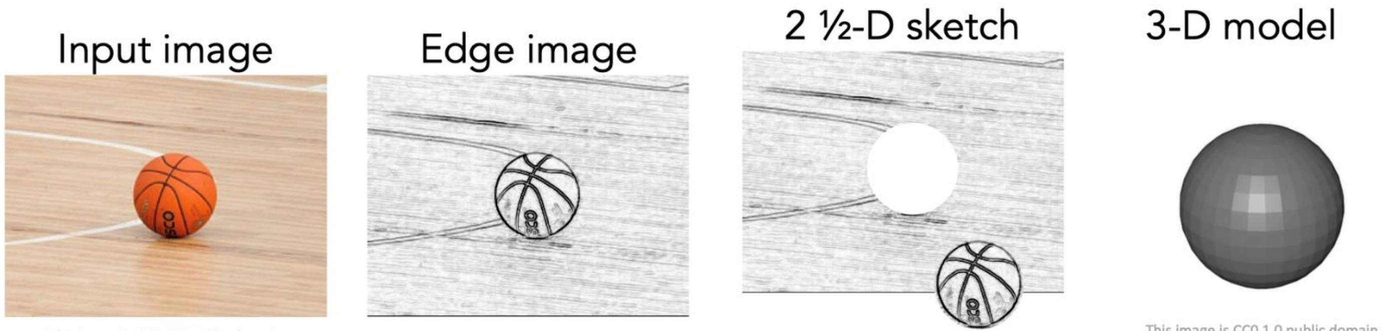
AFTERWORD BY
Tomaso Poggio

Copyrighted Material

3D Reconstruction
Not talent, but
computation

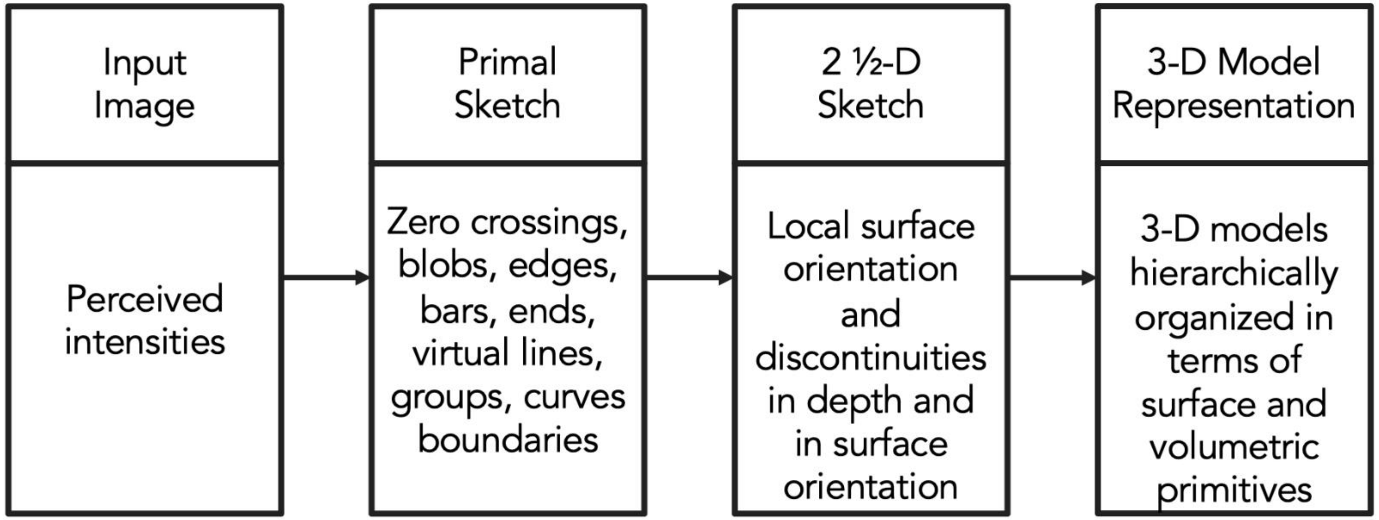
History of Computer Vision

- **Marr Computational Vision**



This image is CC0 1.0 public domain

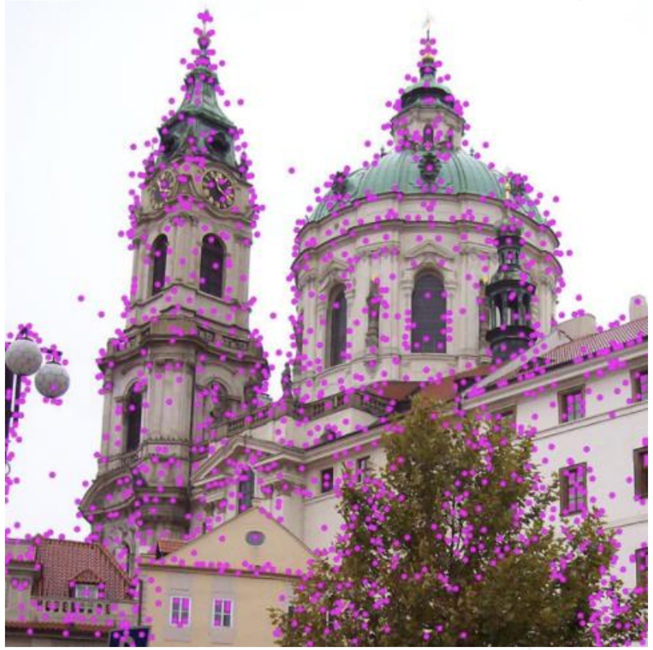
This image is CC0 1.0 public domain

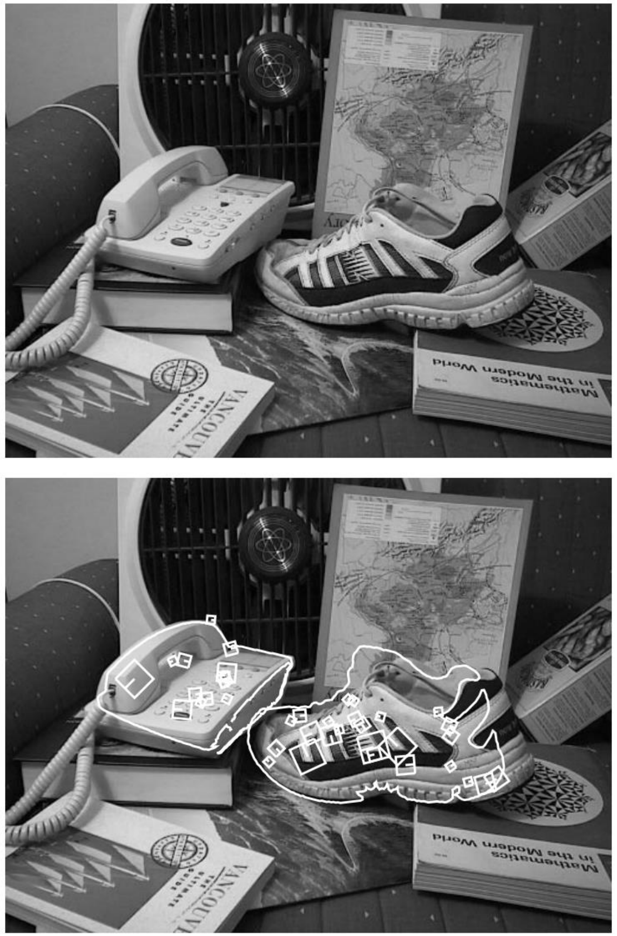

Input Image
Perceived intensities
Primal Sketch
Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries
2 ½-D Sketch
Local surface orientation and discontinuities in depth and in surface orientation
3-D Model Representation
3-D models hierarchically organized in terms of surface and volumetric primitives

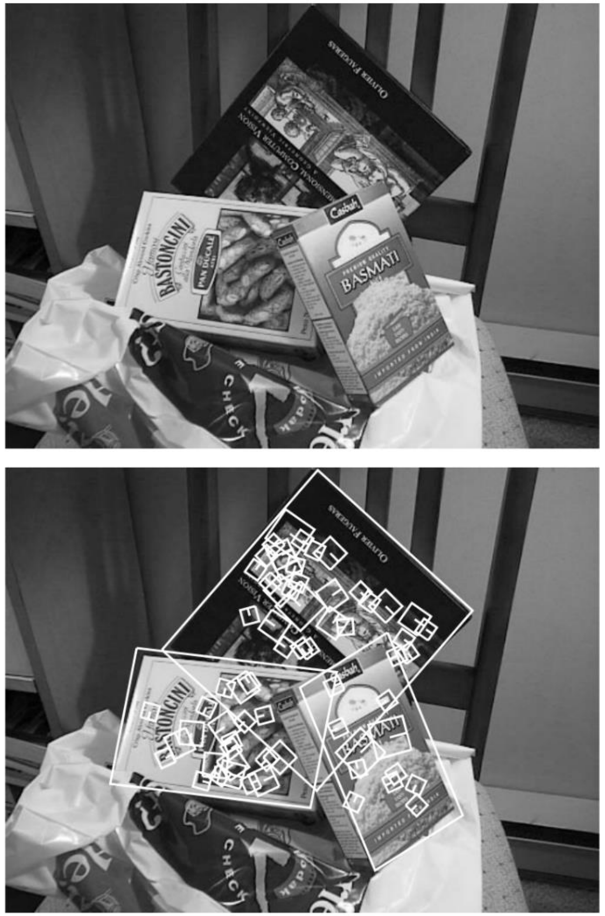
Stages of Visual Representation, David Marr, 1970s

History of Computer Vision

- **Feature Detection——SIFT**

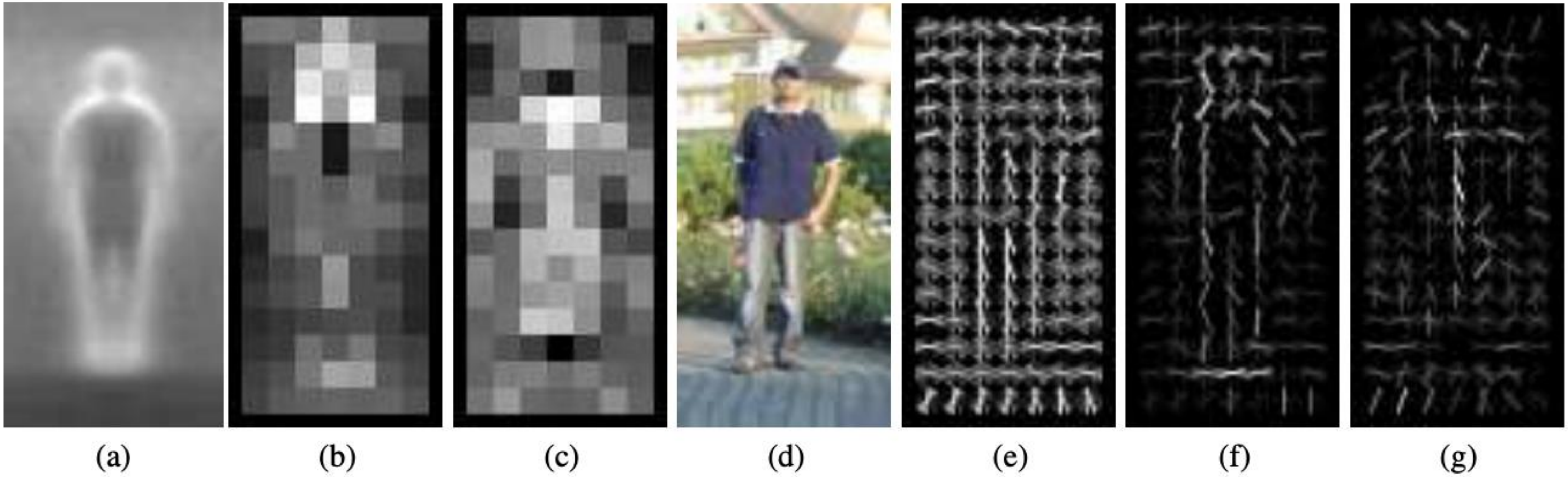






https://www.cs.ubc.ca/~lowe/papers/iccv99.pdf

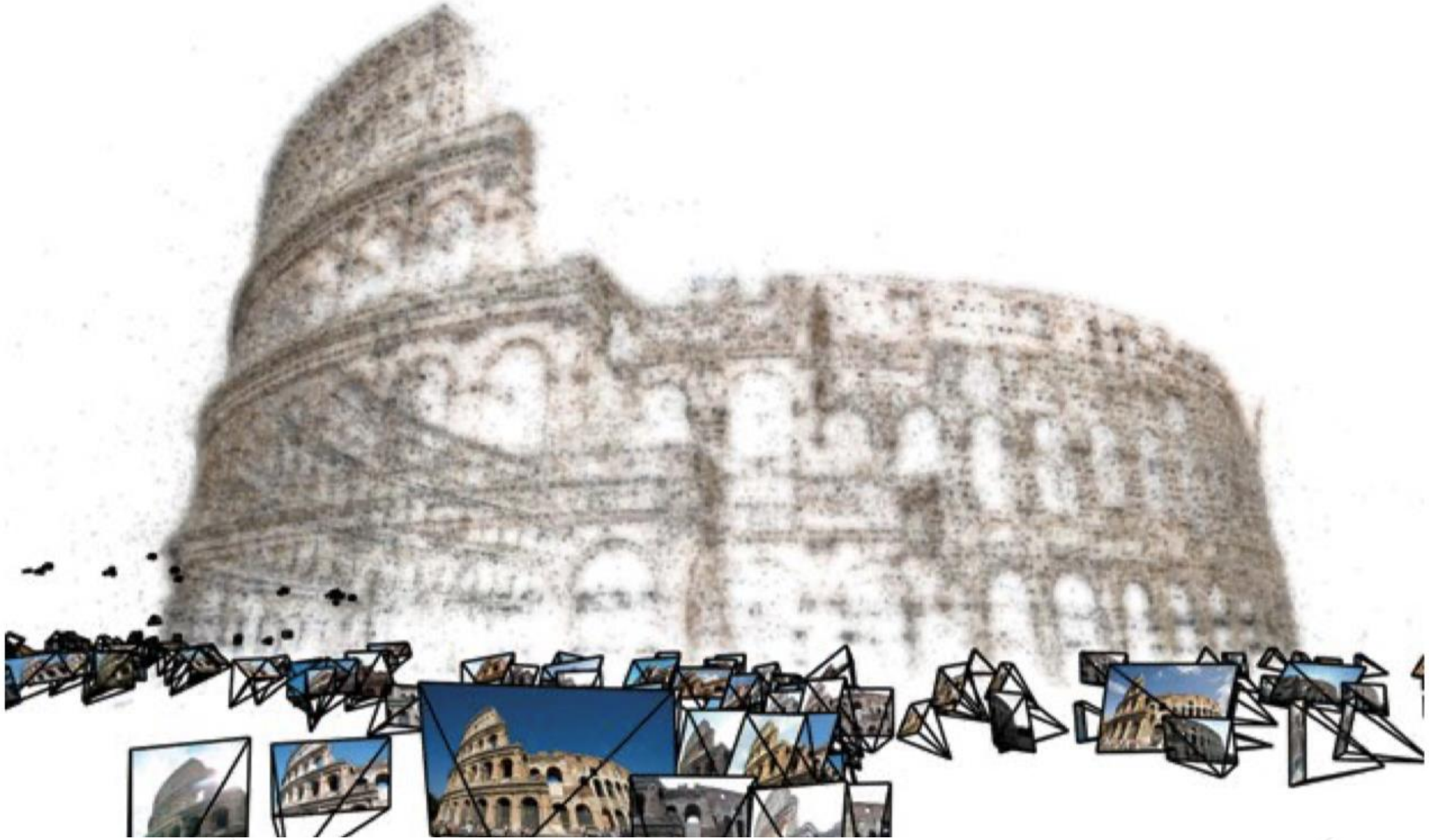
- **Feature Detection——HOG**



https://web.archive.org/web/20110408220331/
http://www.acemedia.org/aceMedia/files/document/wp7/2005/cvpr05-inria.pdf

History of Computer Vision

- **3D reconstruction**



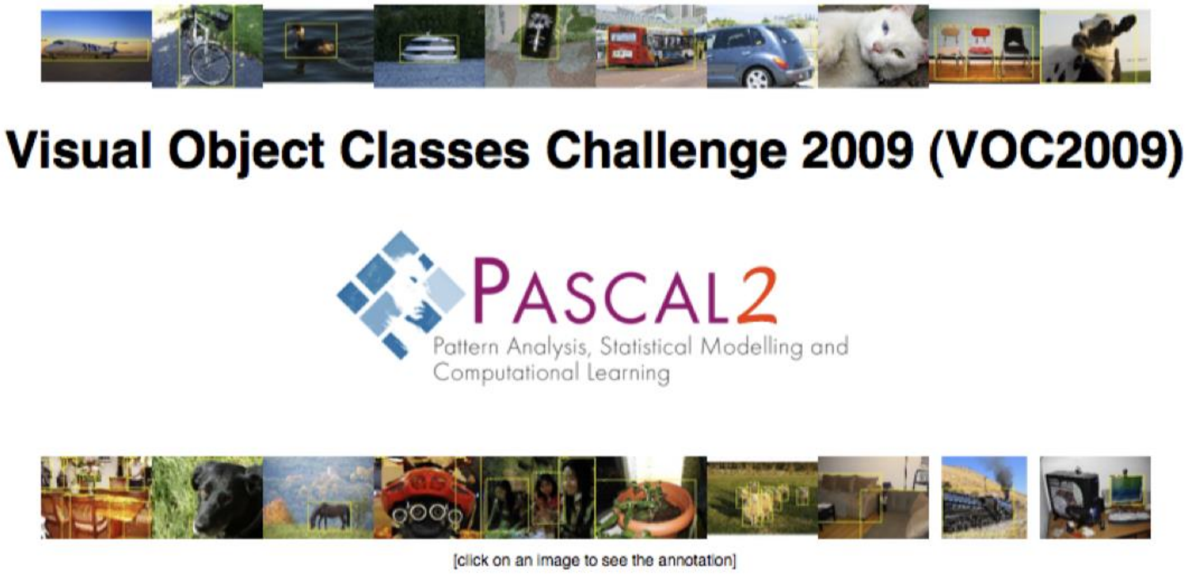
Agarwal et al.
ICCV, 2009

History of Computer Vision

- **Image Classification**

Caltech 101 images

Fei-Fei et al. 2004



Visual Object Classes Challenge 2009 (VOC2009)

PASCAL2

Pattern Analysis, Statistical Modelling and Computational Learning

[click on an image to see the annotation]

Everingham et al. 2006-2012

History of Computer Vision

- **IMAGENET Challenge**

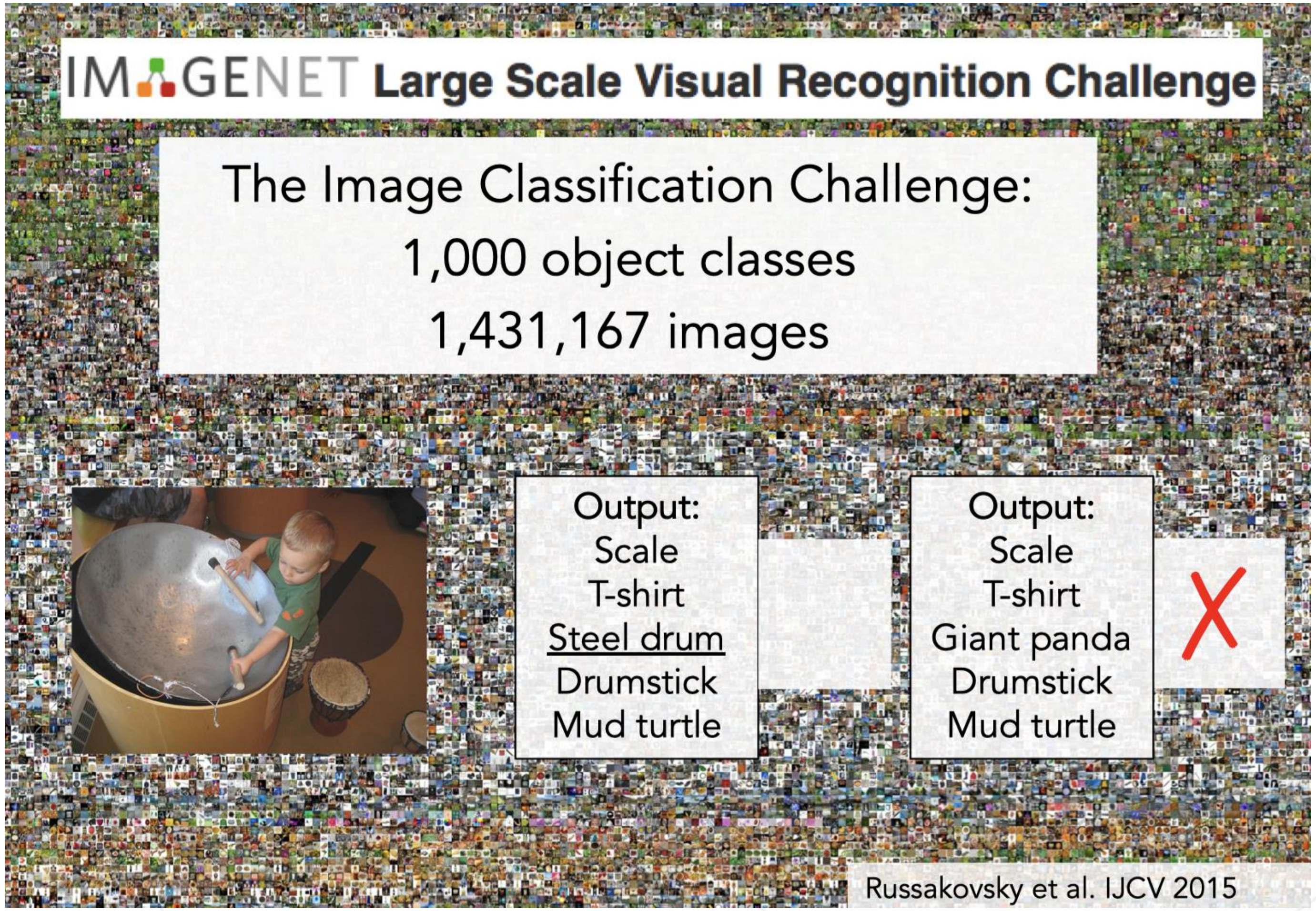
22,000 categories

15,000,000 images

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li & L. Fei-Fei. *CVPR*, 2009.

History of Computer Vision

- **IMAGENET Challenge**



IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images

Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

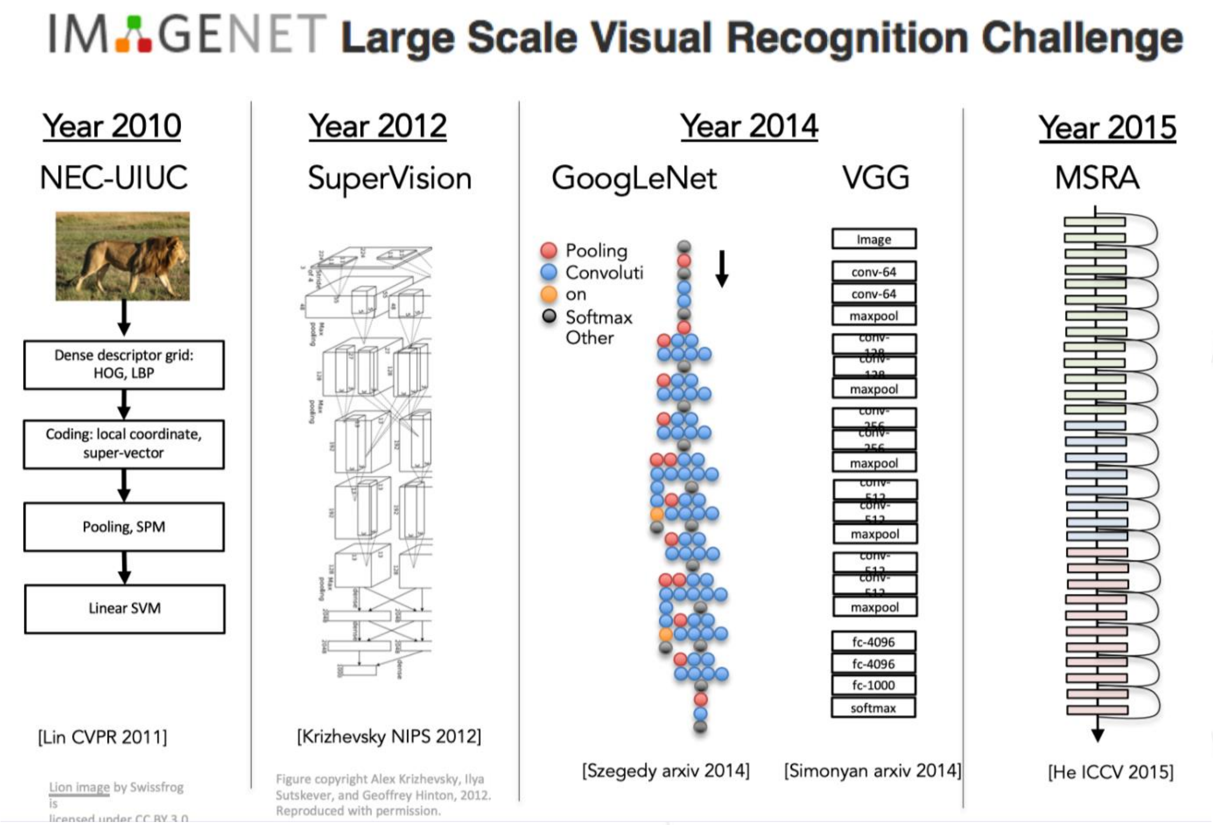
Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

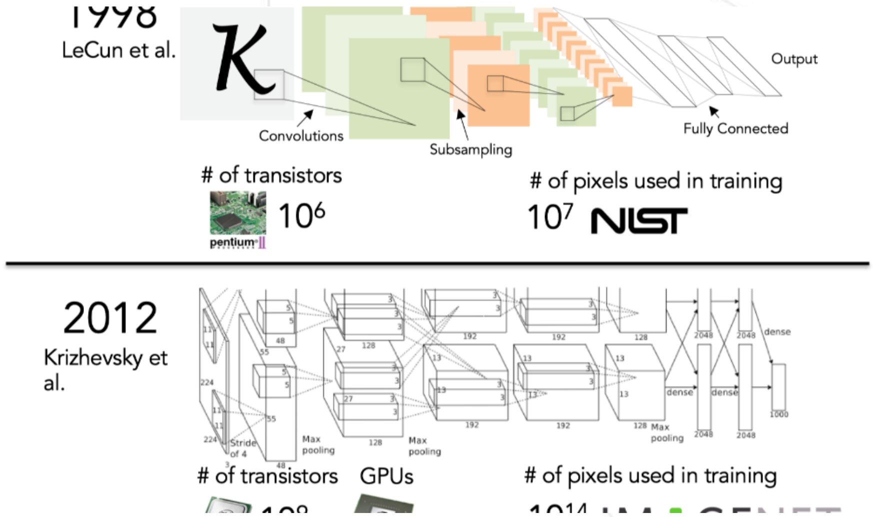
✗

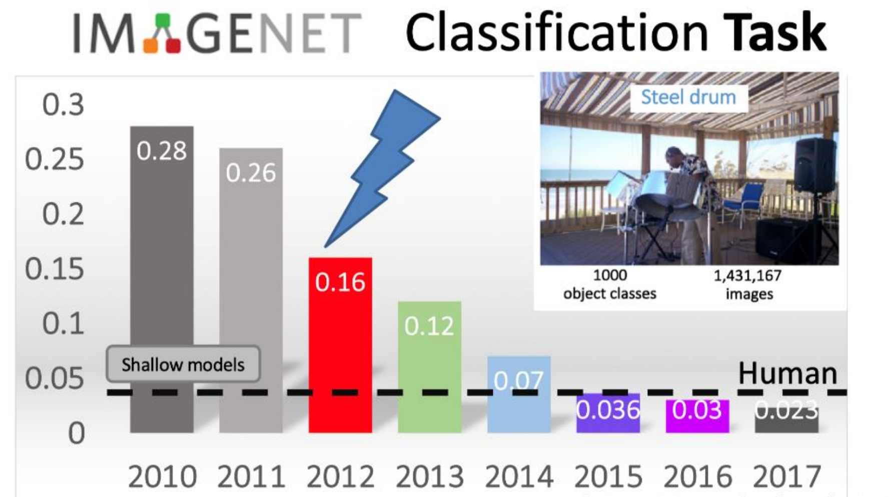
Russakovsky et al. IJCV 2015

History of Computer Vision

- **IMAGENET Challenge**





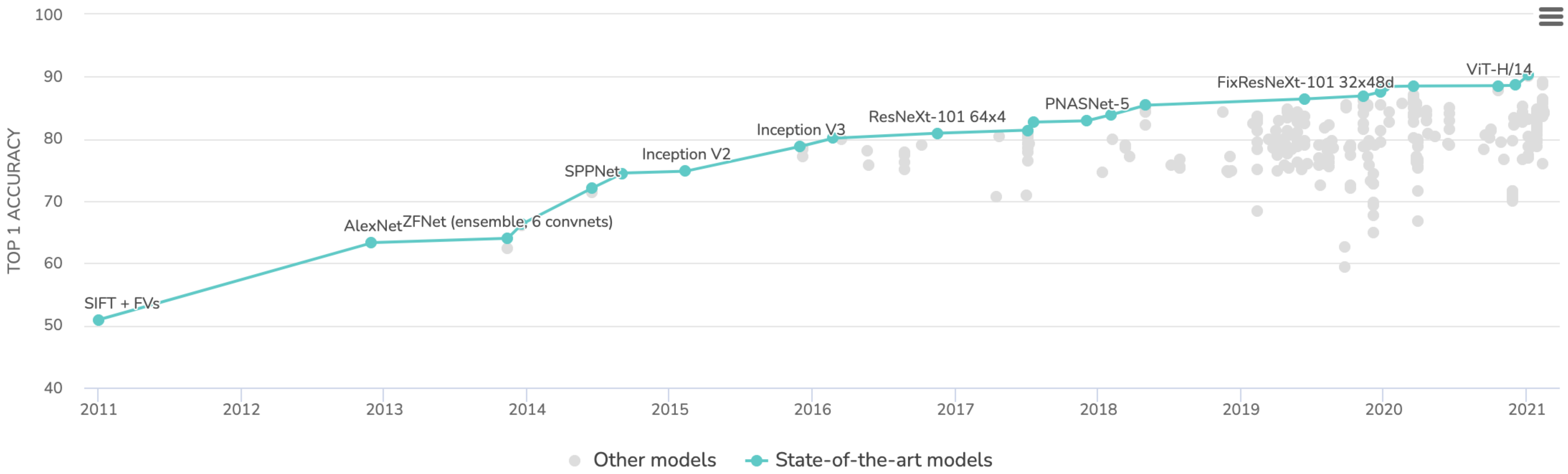


Deng et al. *CVPR*, 2009; Russakovsky et al. *IJCV*, 2

History of Computer Vision

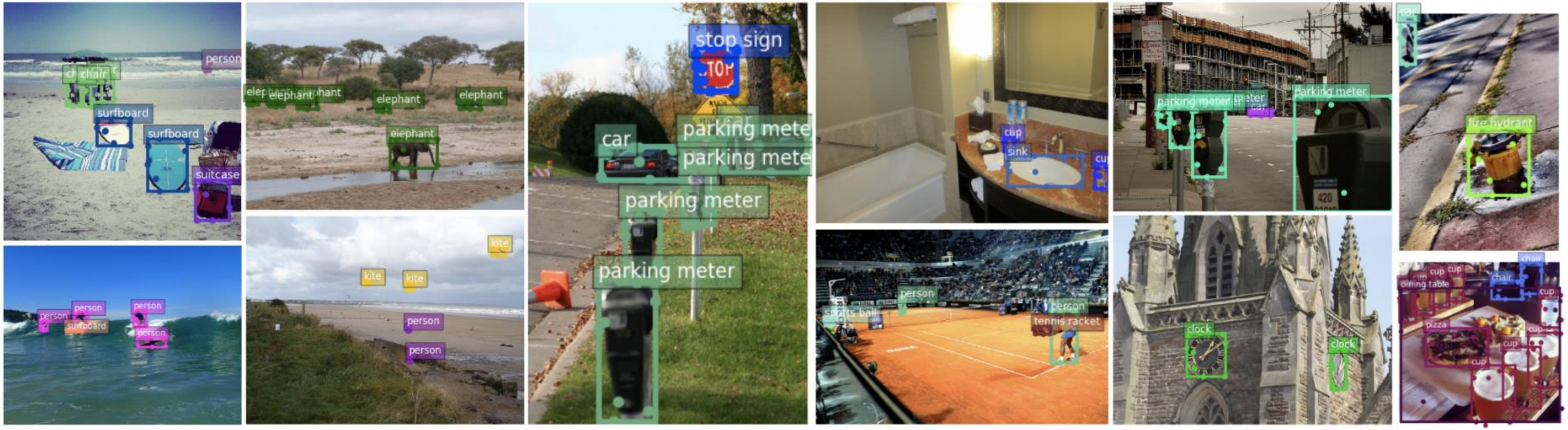
Image Classification on ImageNet

Leaderboard Dataset



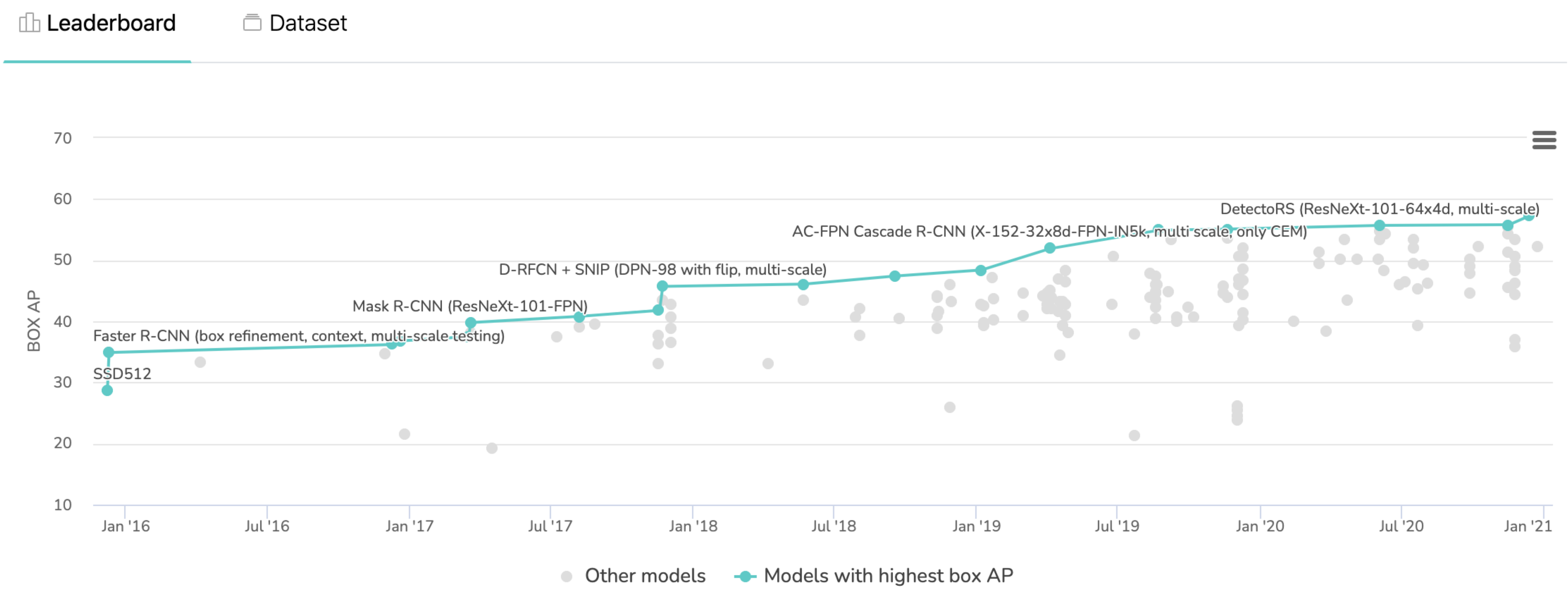
History of Computer Vision

- **Object Detection**



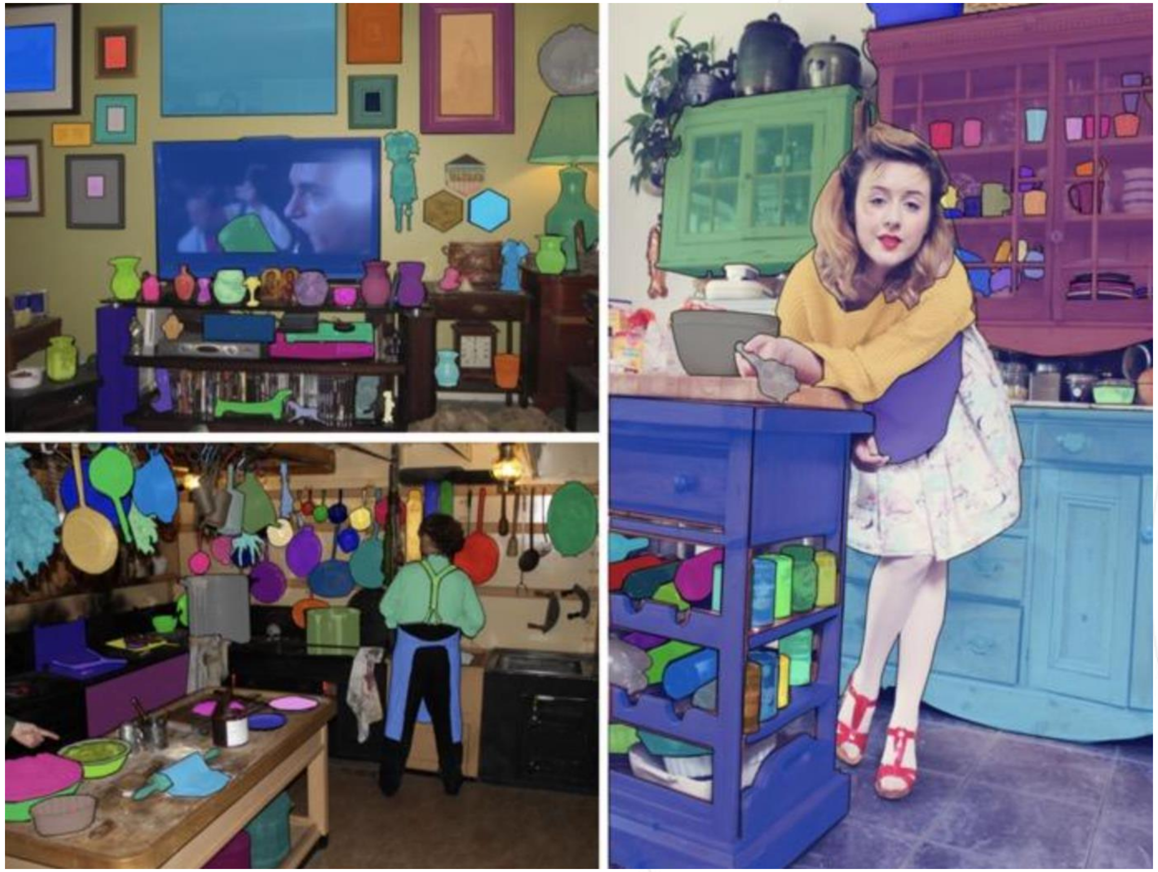
https://cocodataset.org/

Object Detection on COCO test-dev

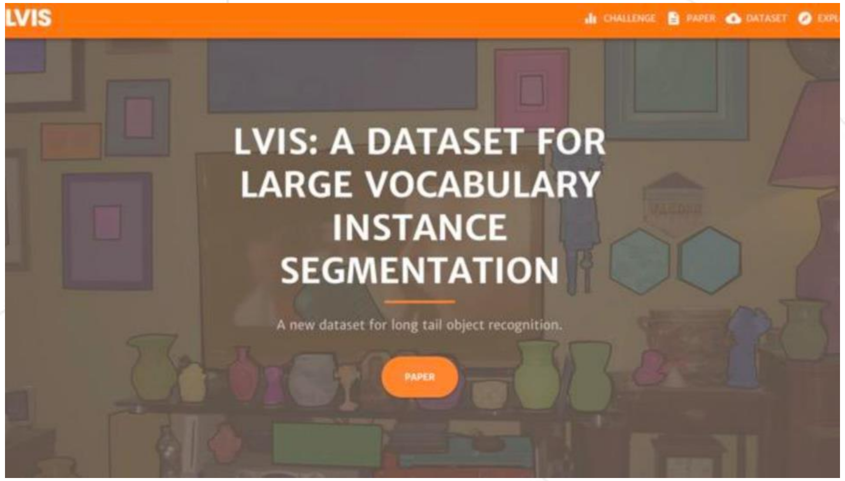


History of Computer Vision

- **Instance Segmentation**

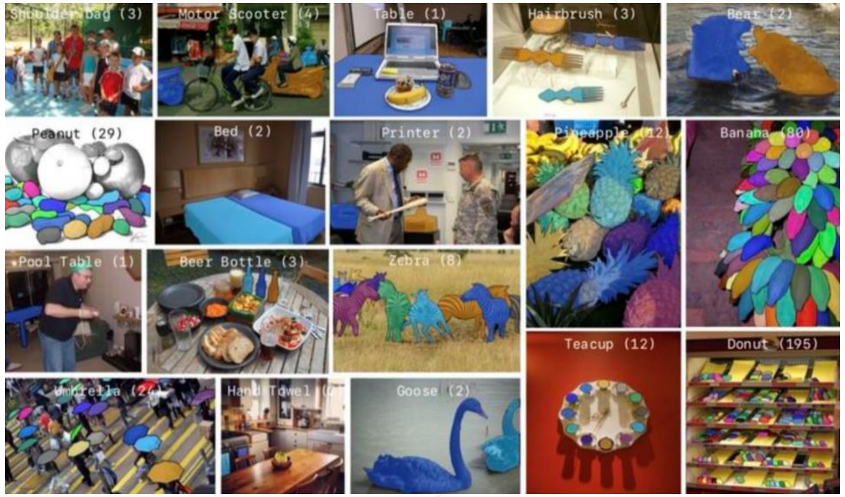


https://www.lvisdataset.org/explore



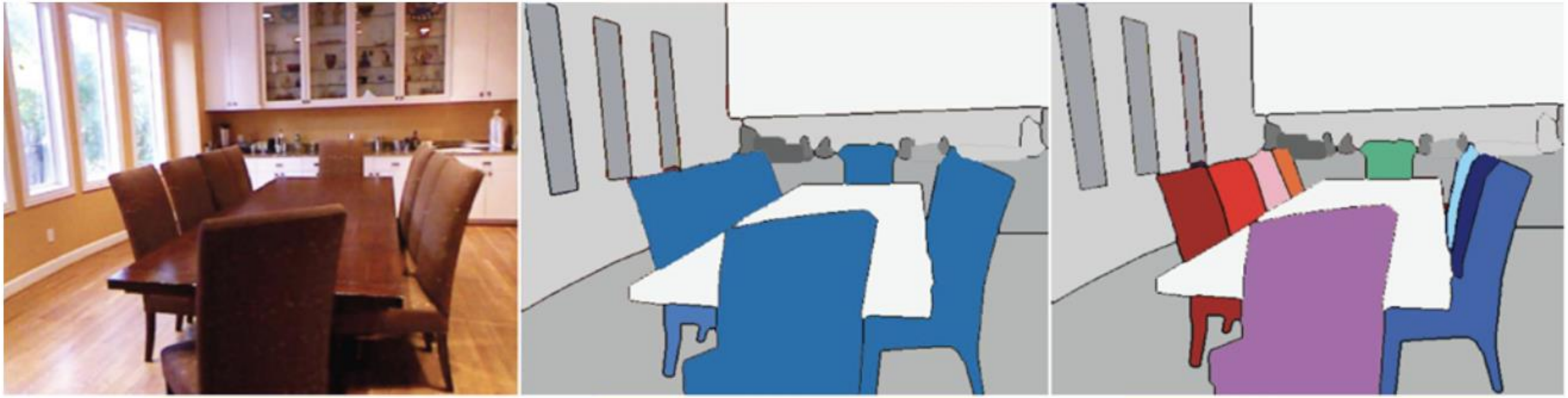
LVIS: A DATASET FOR LARGE VOCABULARY INSTANCE SEGMENTATION

A new dataset for long tail object recognition.



History of Computer Vision

- **Semantic Segmentation and Instance Segmentation**

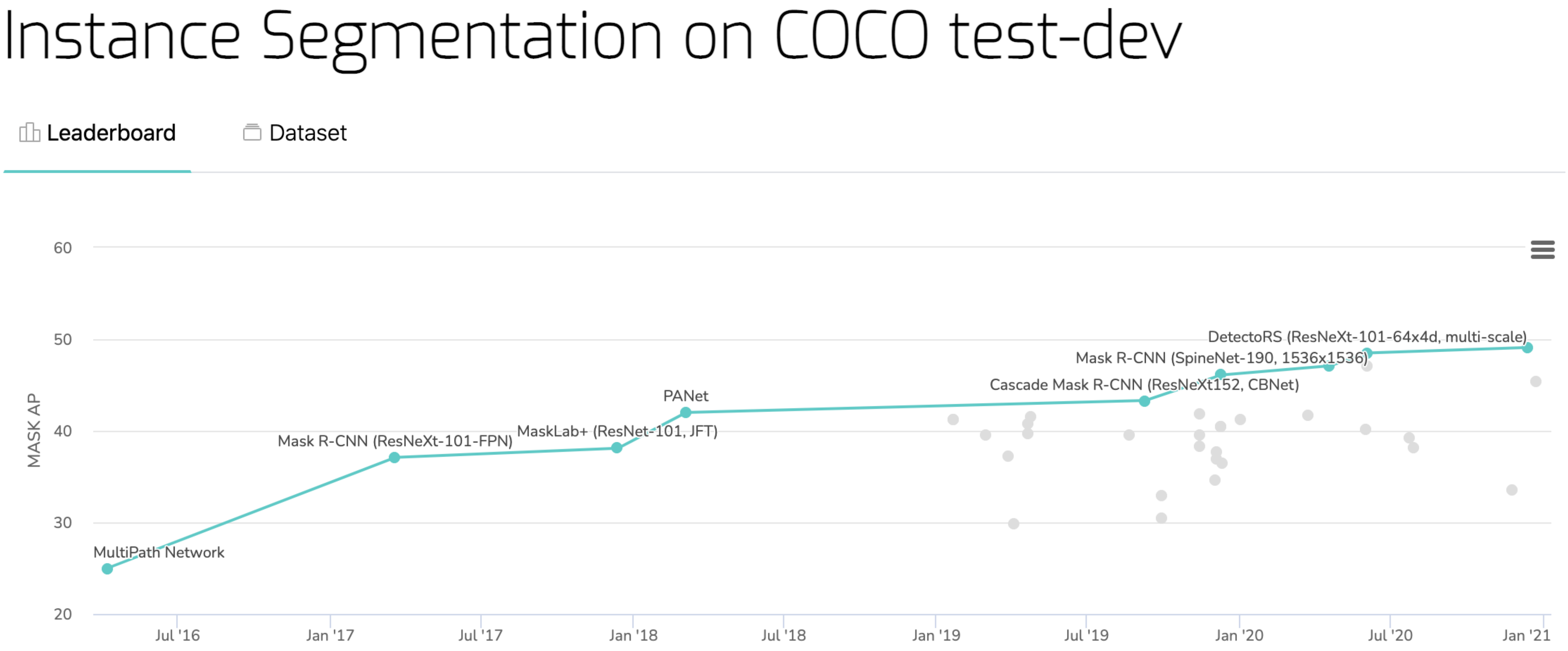


Input Image

Semantic Segmentation

Instance Segmentation

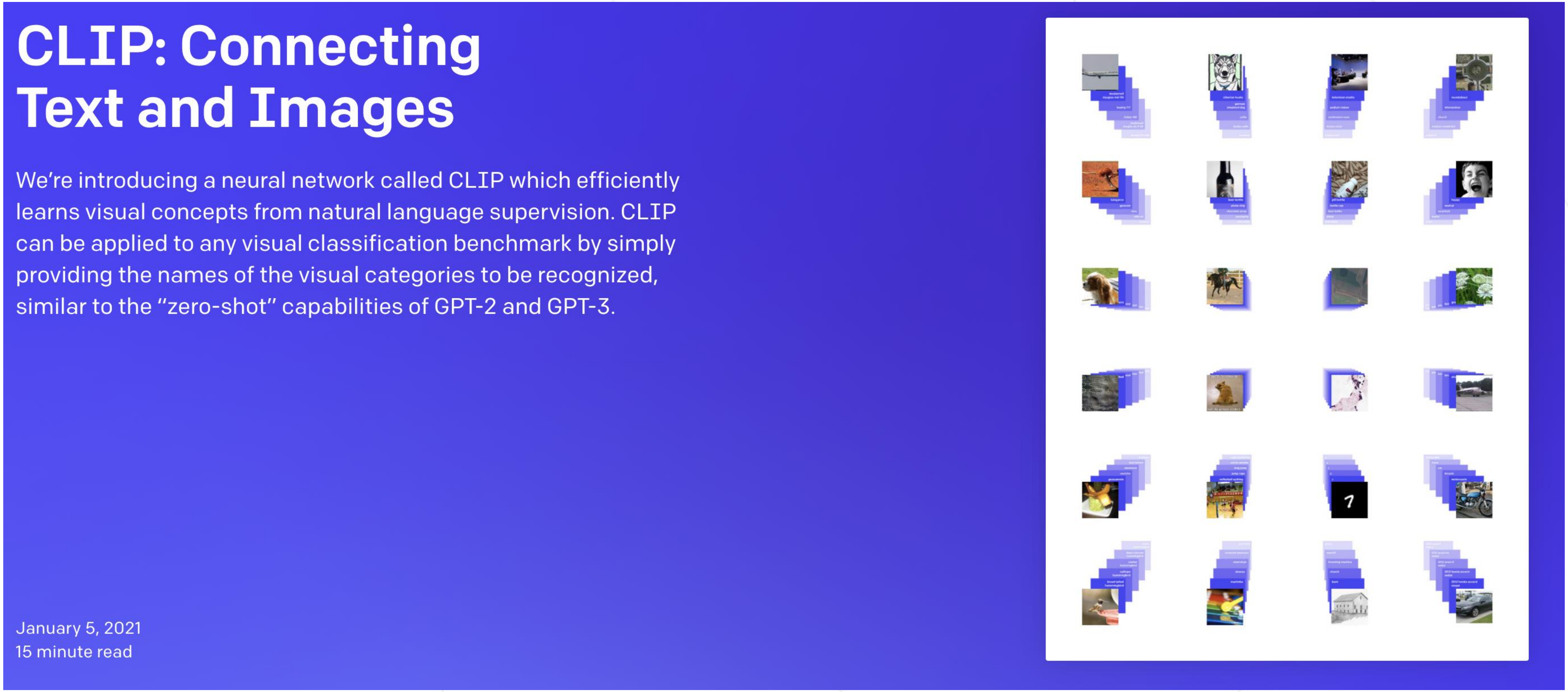
History of Computer Vision



Instance Segmentation on COCO test-dev

Leaderboard
Dataset
MASK AP
60
50
40
30
20
Jul '16
Jan '17
Jul '17
Jan '18
Jul '18
Jan '19
Jul '19
Jan '20
Jul '20
Jan '21
MultiPath Network
Mask R-CNN (ResNeXt-101-FPN)
MaskLab+ (ResNet-101, JFT)
PANet
Cascade Mask R-CNN (ResNeXt152, CBNet)
Mask R-CNN (SpineNet-190, 1536x1536)
DetectoRS (ResNeXt-101-64x4d, multi-scale)

History of Computer Vision



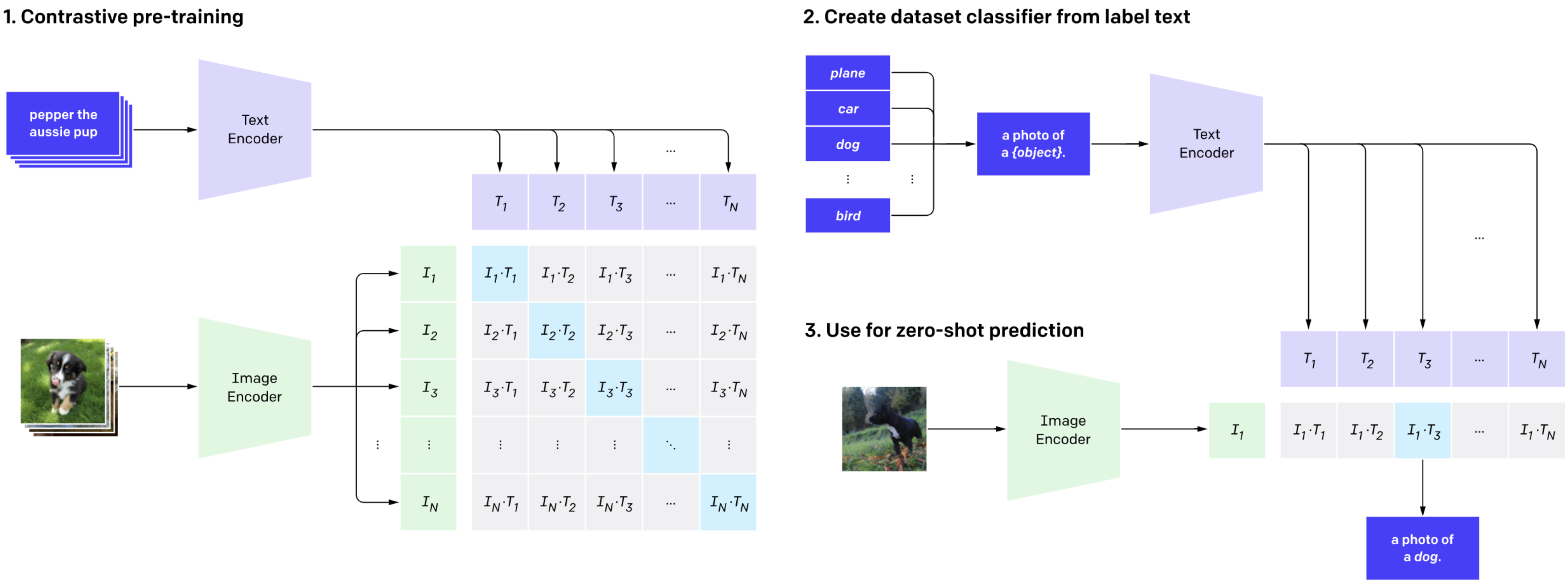
CLIP: Connecting Text and Images

We're introducing a neural network called CLIP which efficiently learns visual concepts from natural language supervision. CLIP can be applied to any visual classification benchmark by simply providing the names of the visual categories to be recognized, similar to the "zero-shot" capabilities of GPT-2 and GPT-3.

January 5, 2021
15 minute read

History of Computer Vision (Learning-based Vision)

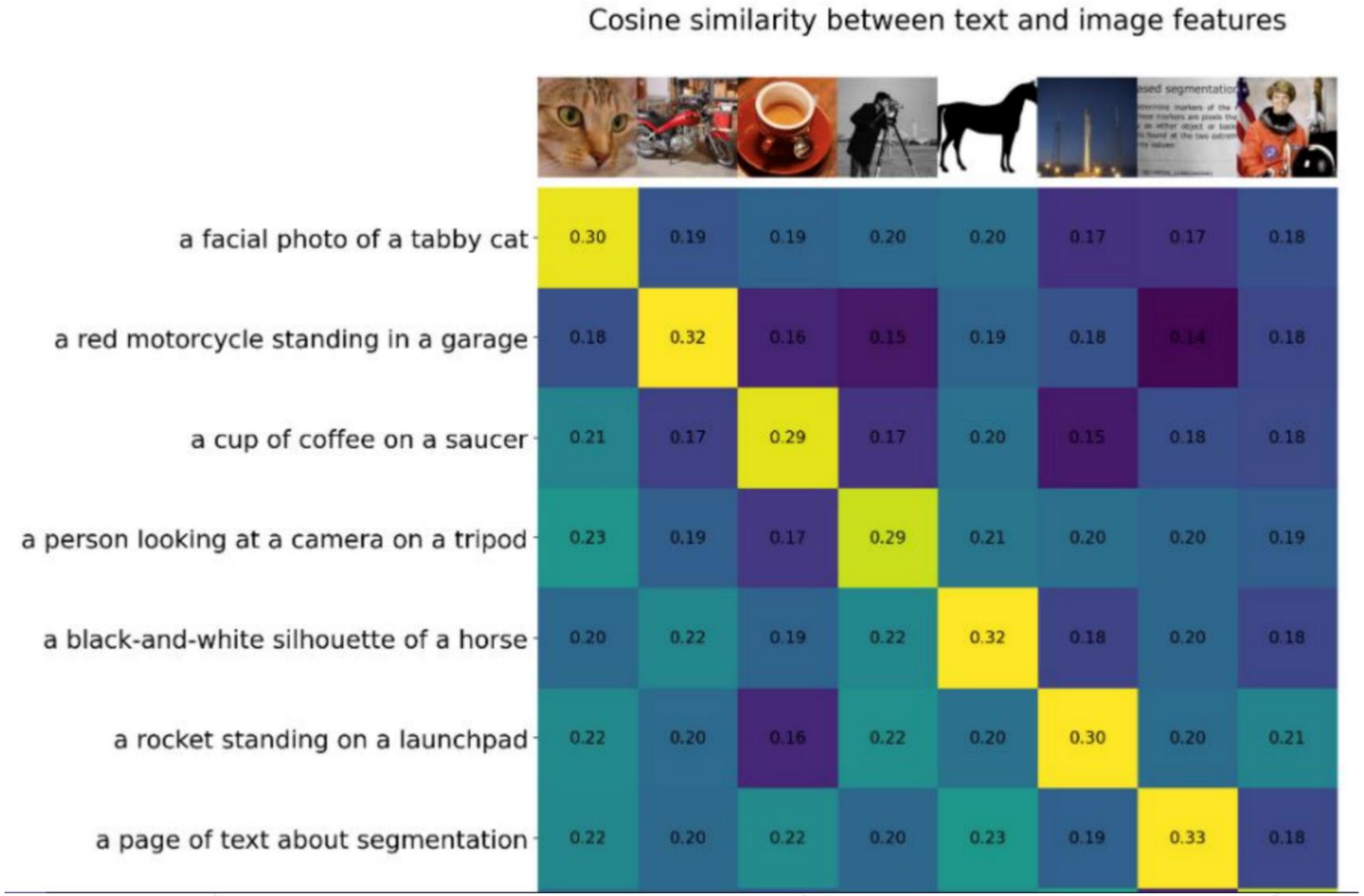
- **CLIP: Connecting Text and Images**


1. Contrastive pre-training
pepper the aussie pup
Text Encoder
Image Encoder
2. Create dataset classifier from label text
plane
car
dog
⋮
bird
a photo of a *{object}*.
Text Encoder
3. Use for zero-shot prediction
Image Encoder
I_1
T_1 T_2 T_3 … T_N
$I_1 \cdot T_1$ $I_1 \cdot T_2$ $I_1 \cdot T_3$ … $I_1 \cdot T_N$
a photo of a *dog*.

	T_1	T_2	T_3	…	T_N
I_1	$I_1 \cdot T_1$	$I_1 \cdot T_2$	$I_1 \cdot T_3$	…	$I_1 \cdot T_N$
I_2	$I_2 \cdot T_1$	$I_2 \cdot T_2$	$I_2 \cdot T_3$	…	$I_2 \cdot T_N$
I_3	$I_3 \cdot T_1$	$I_3 \cdot T_2$	$I_3 \cdot T_3$	…	$I_3 \cdot T_N$
⋮	⋮	⋮	⋮	⋱	⋮
I_N	$I_N \cdot T_1$	$I_N \cdot T_2$	$I_N \cdot T_3$	…	$I_N \cdot T_N$

History of Computer Vision

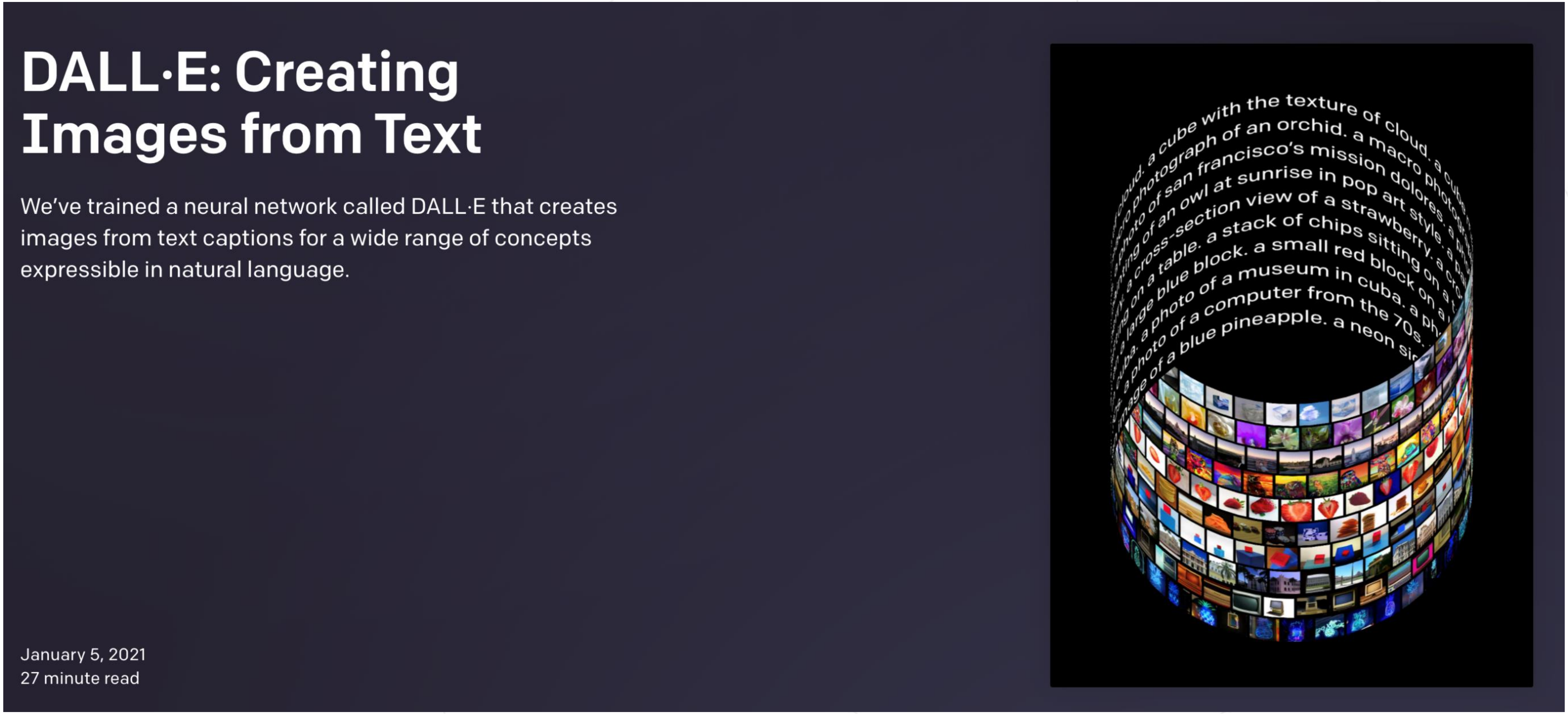
- **CLIP: Image-Text Match**



Cosine similarity between text and image features

a facial photo of a tabby cat	0.30	0.19	0.19	0.20	0.20	0.17	0.17	0.18
a red motorcycle standing in a garage	0.18	0.32	0.16	0.15	0.19	0.18	0.14	0.18
a cup of coffee on a saucer	0.21	0.17	0.29	0.17	0.20	0.15	0.18	0.18
a person looking at a camera on a tripod	0.23	0.19	0.17	0.29	0.21	0.20	0.20	0.19
a black-and-white silhouette of a horse	0.20	0.22	0.19	0.22	0.32	0.18	0.20	0.18
a rocket standing on a launchpad	0.22	0.20	0.16	0.22	0.20	0.30	0.20	0.21
a page of text about segmentation	0.22	0.20	0.22	0.20	0.23	0.19	0.33	0.18

History of Computer Vision



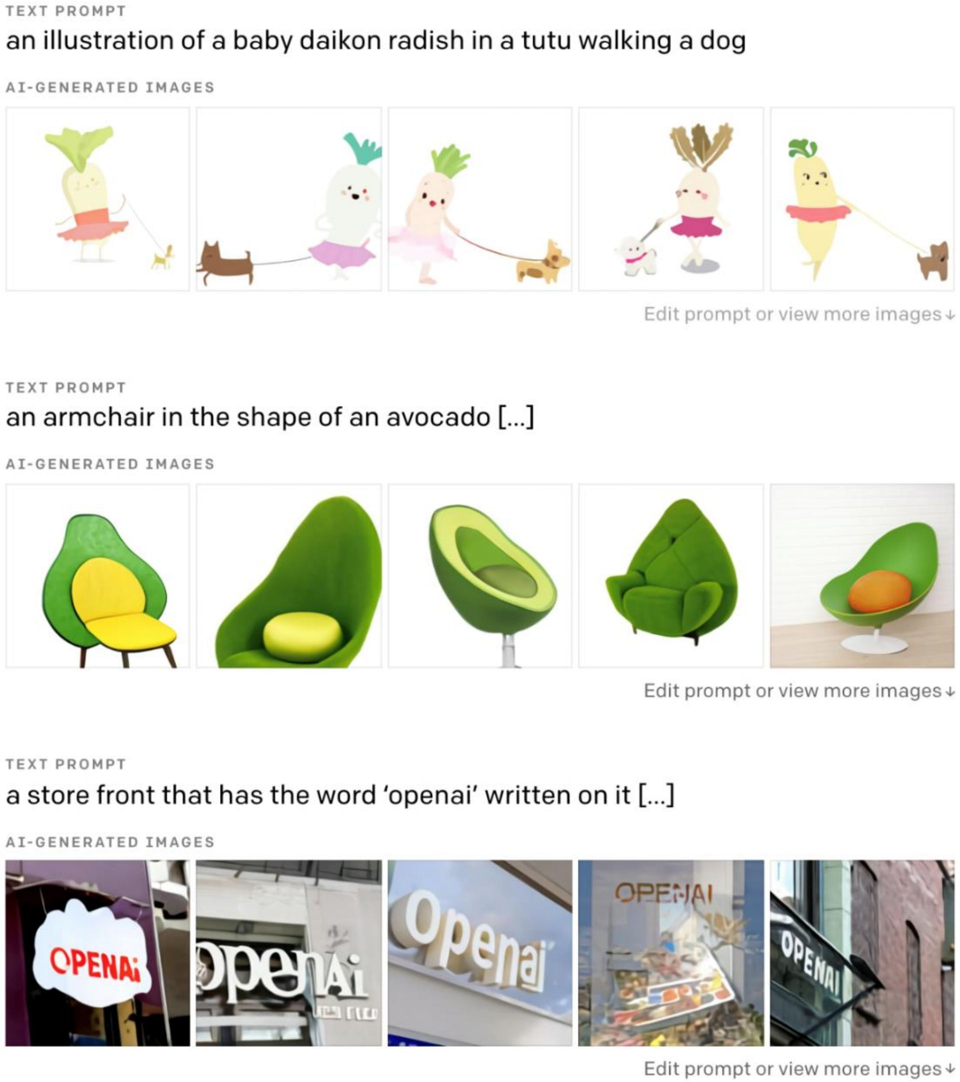
DALL·E: Creating Images from Text

We've trained a neural network called DALL·E that creates images from text captions for a wide range of concepts expressible in natural language.

January 5, 2021
27 minute read

a cube with the texture of cloud. a cube
photograph of an orchid. a macro photograph
of san francisco's mission dolores
of an owl at sunrise in pop art style
cross-section view of a strawberry
a table. a stack of chips sitting on a
blue block. a small red block on a
a photo of a museum in cuba. a ph
photo of a computer from the 70s
of a blue pineapple. a neon sig

History of Computer Vision



TEXT PROMPT

an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES

Edit prompt or view more images↓

TEXT PROMPT

an armchair in the shape of an avocado [...]

AI-GENERATED IMAGES

Edit prompt or view more images↓

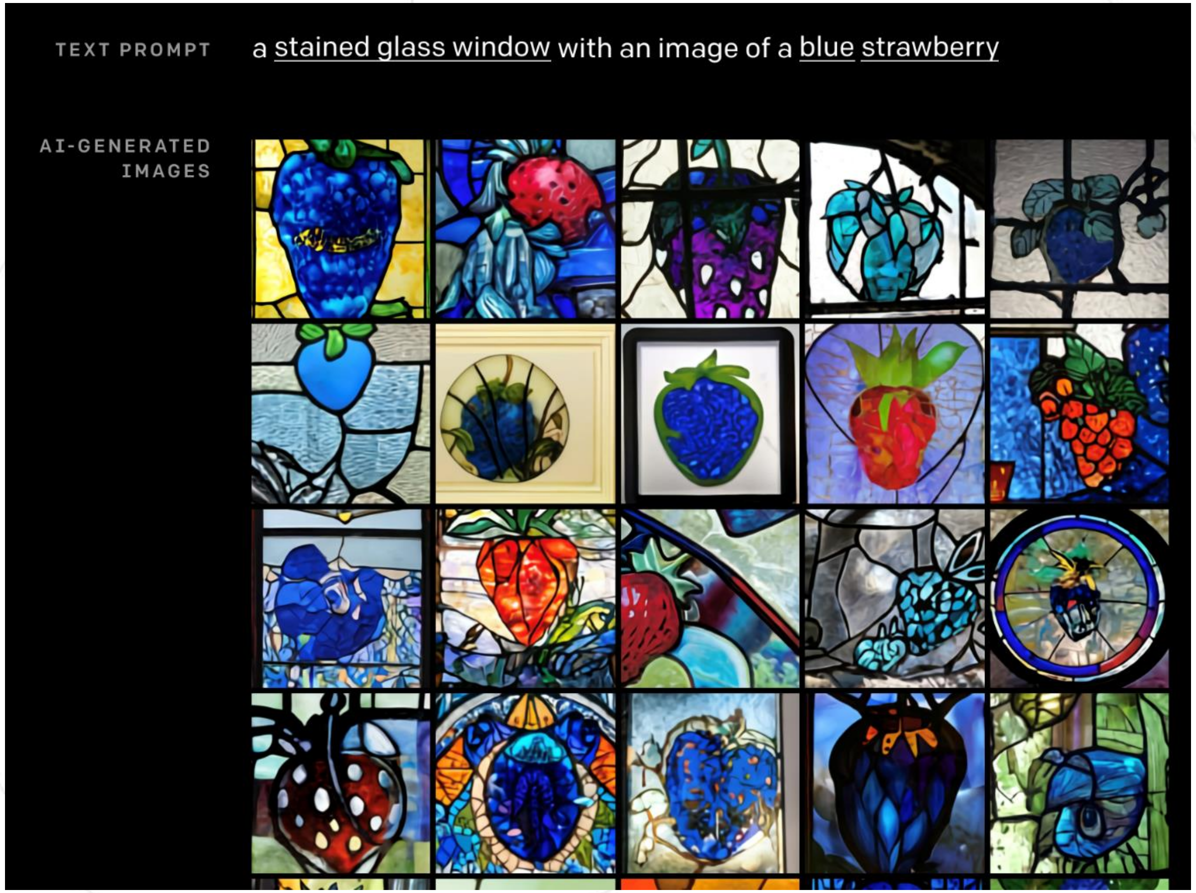
TEXT PROMPT

a store front that has the word 'openai' written on it [...]

AI-GENERATED IMAGES

Edit prompt or view more images↓

DALL-E
Creating Images from Text

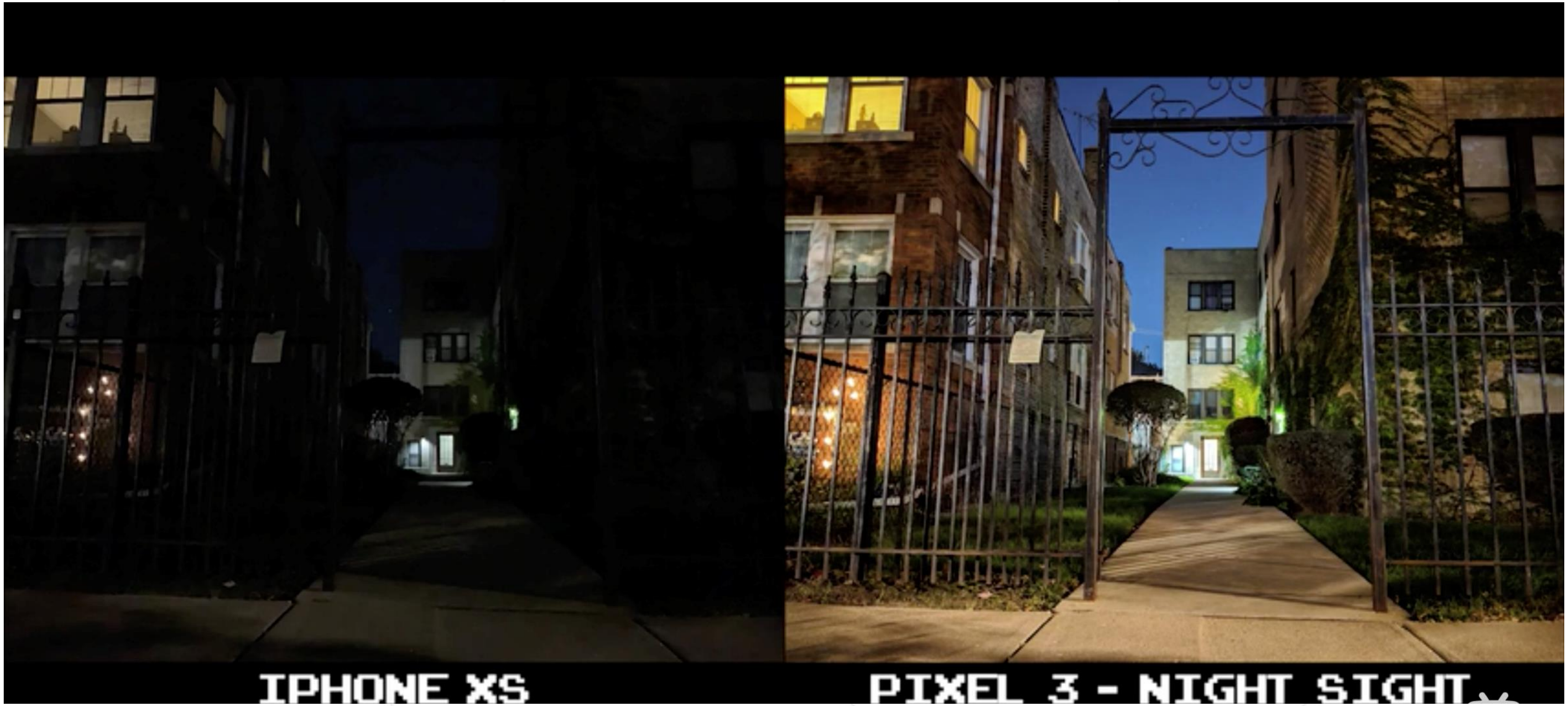


TEXT PROMPT a stained glass window with an image of a blue strawberry

AI-GENERATED IMAGES

History of Computer Vision

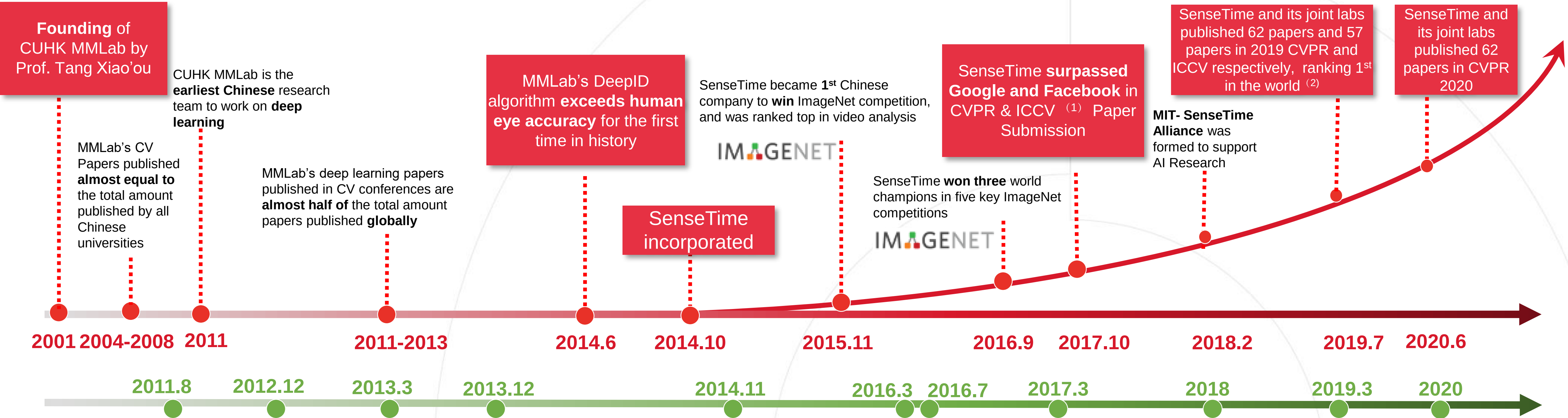
- **Low-level Vision**



SenseTime – Pioneer in Deep Learning and Computer Vision

Tsinghua University

sensetime



Founding of CUHK MMLab by Prof. Tang Xiao'ou

2001

MMLab's CV Papers published **almost equal to** the total amount published by all Chinese universities

2004-2008

CUHK MMLab is the **earliest Chinese** research team to work on **deep learning**

2011

MMLab's deep learning papers published in CV conferences are **almost half of** the total amount papers published **globally**

2011-2013

MMLab's DeepID algorithm **exceeds human eye accuracy** for the first time in history

2014.6

SenseTime incorporated

2014.10

SenseTime became 1st Chinese company to **win** ImageNet competition, and was ranked top in video analysis

IMAGENET

2015.11

SenseTime **won three** world champions in five key ImageNet competitions

IMAGENET

2016.9

SenseTime **surpassed Google and Facebook** in CVPR & ICCV (1) Paper Submission

2017.10

MIT- SenseTime Alliance was formed to support AI Research

2018.2

SenseTime and its joint labs published 62 papers and 57 papers in 2019 CVPR and ICCV respectively, ranking 1st in the world (2)

2019.7

SenseTime and its joint labs published 62 papers in CVPR 2020

2020.6

2011.8

Microsoft

Microsoft significantly improved the accuracy of Deep Learning-driven voice recognition

2012.12

IMAGENET

Hinton, the originator of Deep Learning, won the ImageNet visual recognition competition

2013.3

Google

Google employed Prof. Hinton for US$50 million

2013.12

facebook

Facebook established an AI Lab in New York and employed Yann LeCun

2014.11

Google DeepMind

Google acquired DeepMind, a Deep Learning company, for US$660 million

2016.3

AlphaGo

DeepMind's AlphaGo, the Go AI, beats Lee Sedol in Go

cruise GM

GM acquired Cruise Automation, a startup in autonomous vehicle technology for US$1 billion

2016.7

SoftBank ARM

Softbank launched a US$100 billion investment fund that focuses on AI and completed US$32 billion acquisition of ARM

2017.3

intel MOBILEYE

Intel acquired Mobileye, a leader in computer vision for autonomous driving technology

2018

TESLA WAYMO

Tesla launched self-driving vehicle. Waymo launched self-driving taxi service

NVIDIA

NVIDIA launches its own GPU cloud

2019.3

Hinton, LeCun & Bengio received the Turing Award, the "Nobel Prize" of Computing

2020

Global tech companies are exploring the use of AI to combat the outbreak of COVID-19.

(1) CVPR、ICCV、ECCV are the top 3 computer vision conferences worldwide with highest impact factor They accept the best work on computer vision and deep learning

(2) Based on statistics released by different companies and organizations to date

How to Generate the Best AI

Fundamental research & technological capabilities determine rate of innovation

Expertise

Large amount of high quality data fuels the algorithm iteration

Data

Super fast computing power ensures speed of training

Computing Power

Vertical partnerships ensure technology and data feedback for adaptive improvement

Positive Feedback Loop

商汤 sensetime

SenseTime Excels at All of These Core Capabilities

SenseTime – World Leading AI Innovation Platform

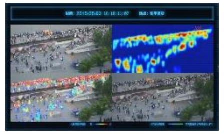
Smart City

Smart Surveillance

Smart City Management System

Smart Traffic Management

Fire Detection


Smart Crowd Management

Abnormal Behavior Detection

Garbage Detection

Illegal Parking Detection

Illegal Occupation Detection

Abnormal Objects Detection on Road

Business Intelligence

Retail Analytics Solutions

Intelligent Hotel Check in System

Smart Airport Solution

Smart Metro Solution

Smart Office Management System

Smart Tourism Area Management

Smart Entertainment Solution

Smart Campus Solution

Smart Amusement Park Solution

Real Estate Sales Management

Mobile Solution

Face Unlock

Photo Processing

Image Super Resolution

3D Face Beautification

AR Platform

AR Live Streaming

AR Game

AR Classroom

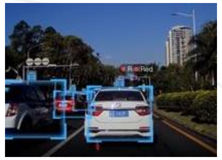
AR Effect

Autonomous Driving

Guide Line Prediction

Human Face Prediction

Lane Detection


Front Vehicle Detection

Intelligence Cabin Sensing

Face Unlock

Gaze Tracking

Gesture Tracking

Drowsiness detection

AI Education Package

AI Textbook

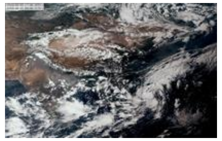
AI Experiment Platform

AI RobotCar

AI Lab

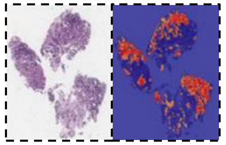
Remote Sensing

Road Network Extraction


Cloud and Snow Detection

AI-Enabled Diagnosis, Treatment and Rehabilitation

Lung AI Application


Pathology Application

IN THE MOOD FOR LOVE

WONG KAR-WAI'S

IN THE MOOD FOR LOVE