南京大学
NANJING UNIVERSITY

计算机视觉表征与识别

Chapter 1： Introduction to computer vision

王利民

媒体计算课题组

http://mcg.nju.edu.cn/

Overview

- What is computer vision about?
- Computer vision is useful.
- Computer vision is difficult.
- History and progress of computer vision
- Course overview

Introduction





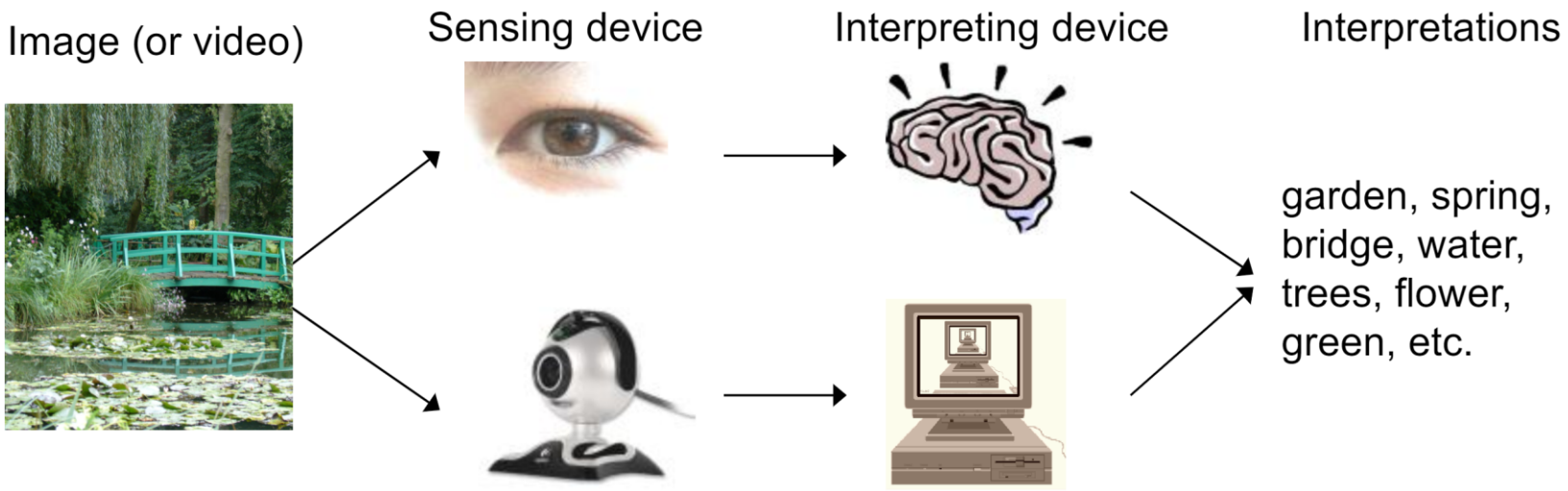
Camera is everywhere






What is computer vision

What is computer vision



The goal of computer vision

- To extract “meaning” from pixels

What we see

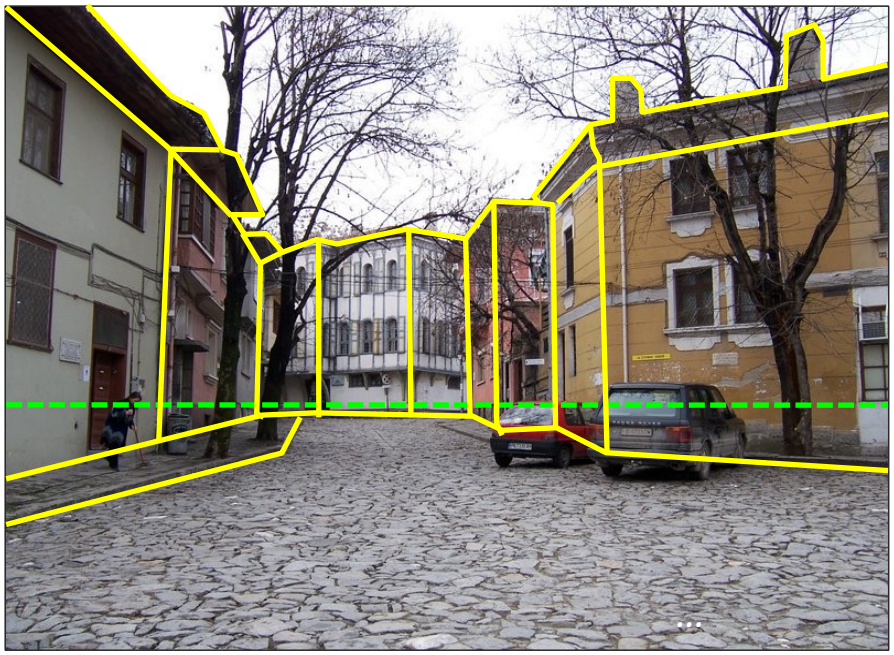
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: S. Narasimhan

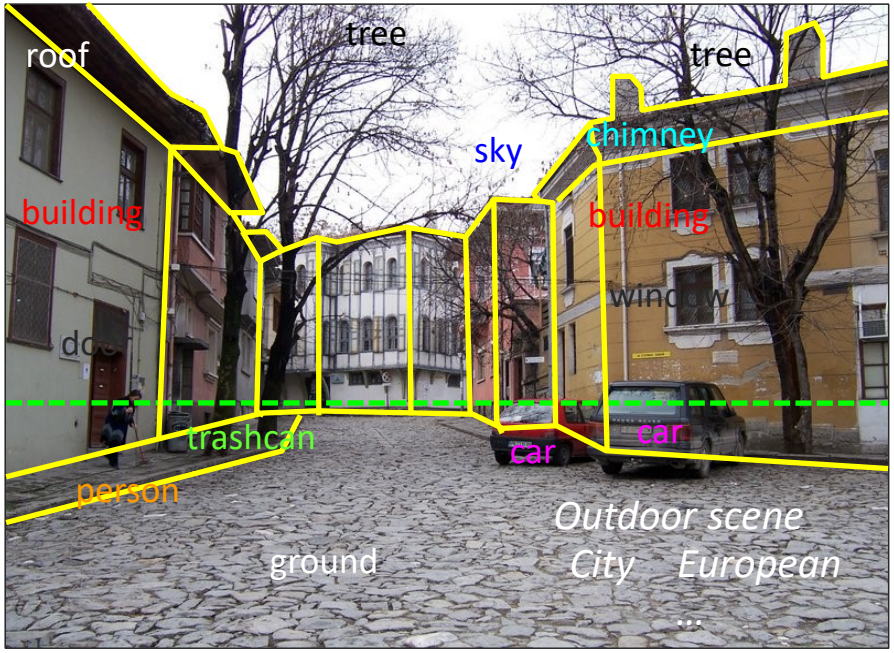
What kind of information

What kind of information



Geometric information

What kind of information



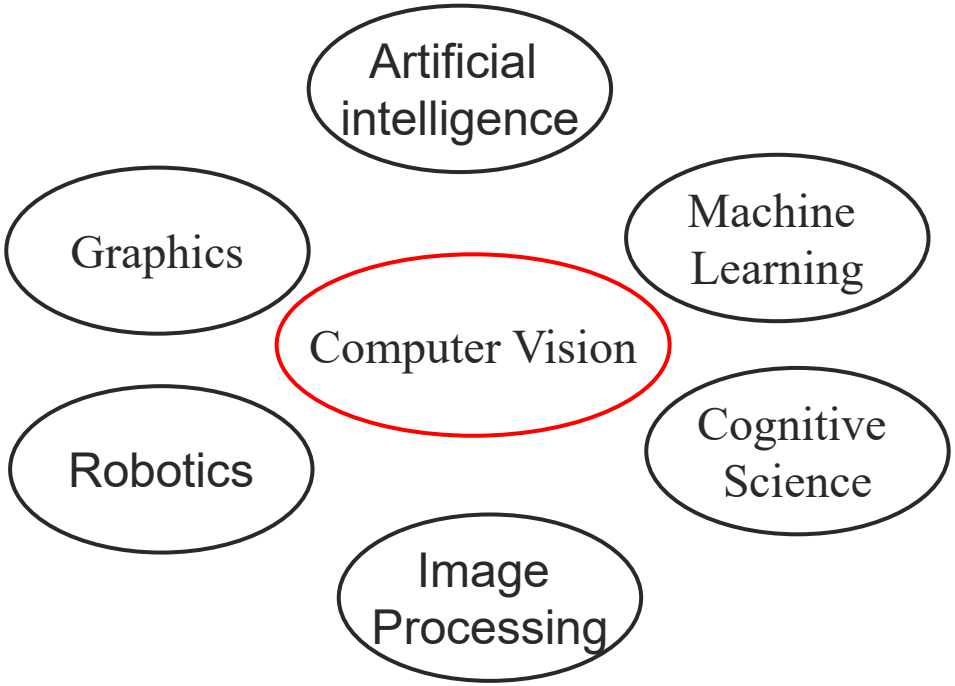
Geometric information

Semantic information

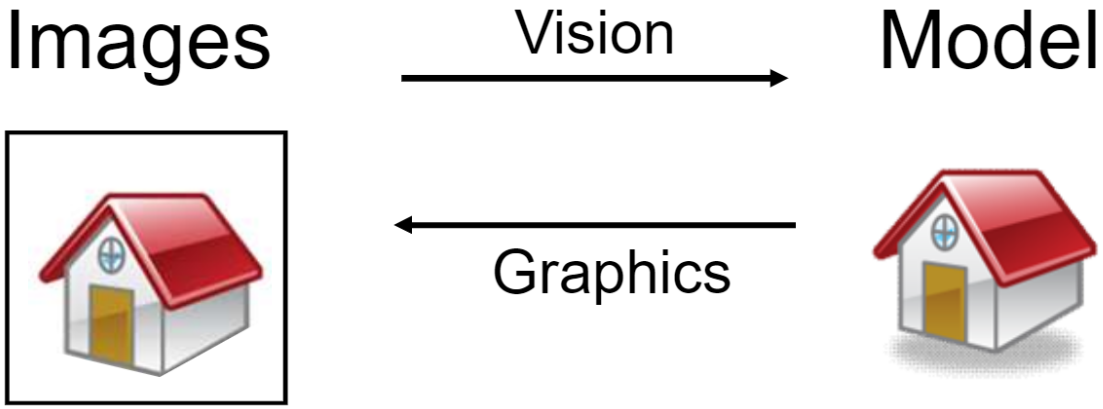
Computer vision

- Automatic understanding of images and video
  - 1. Computing properties of the 3D world from visual data (measurement)
  - 2. Algorithms and representations to allow a machine to recognize objects, scene, and people (perception and interpretation)

Related disciplines



Vision and graphics



Inverse problems: analysis and synthesis.

Why study computer vision

- Computer vision is useful
- Computer vision is difficult
- Computer vision is fast developing

Vision is useful

- For people vision is their most crucial sense, for good reason:
  - half our brain is devoted to it
  - developed many times during evolution
  - it can be implemented with high resolution
  - yields color, texture, depth, motion, shape

Faces and digital cameras

Camera waits for everyone to smile to take a photo [Canon]

Setting camera focus via face detection


Revision3

Pig face recognition

A system made by Yingzi Techology, a small Chinese company, scanning a barn to recognize pig faces.
Yingzi Technology

Video-based interfaces

Human joystick, NewsBreaker Live

Assistive technology systems
Camera Mouse, Boston College

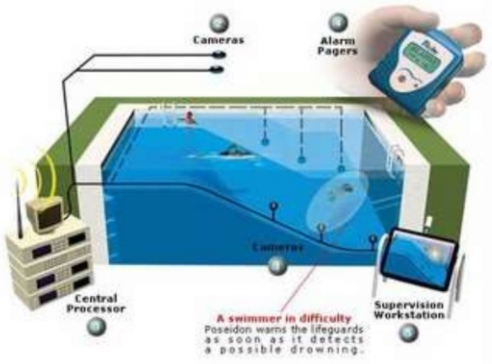
Microsoft Kinect


THE
SOCIAL
MEDIA SHOW


/SUS

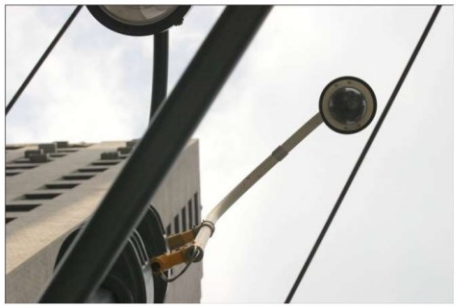
Safety & security

Navigation,
driver safety



Monitoring pool
(Poseidon)

Pedestrian detection
MERL, Viola et al.



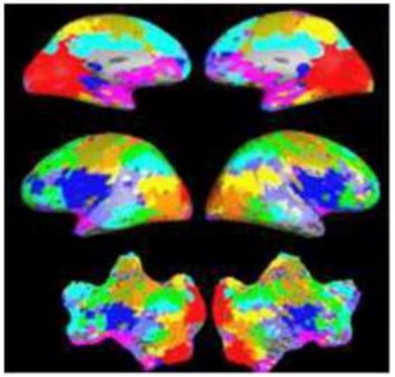
Surveillance

The system converts image data taken by 4 super-wide angle cameras, to display a virtual image of the vehicle from above.


LEARNING
NVIDIA

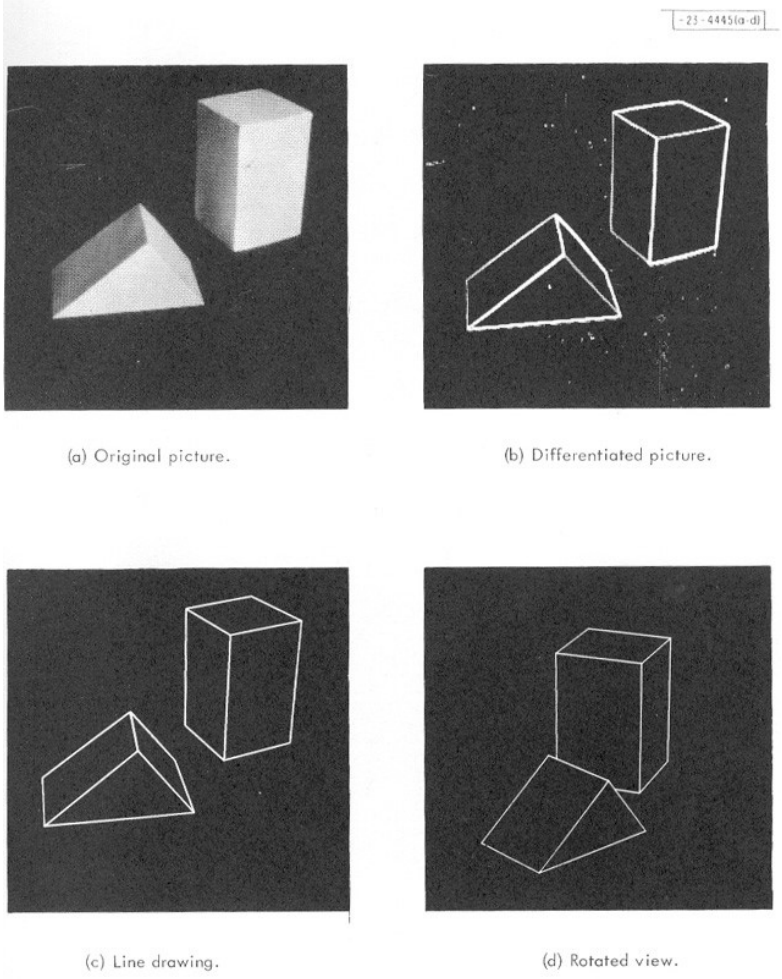
Vision for medical & neuroimages

Image guided surgery
MIT AI Vision Group



fMRI data
Golland et al.

Origins of computer vision



L. G. Roberts *Machine Perception of Three Dimensional Solids*
Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Origins of computer vision

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966

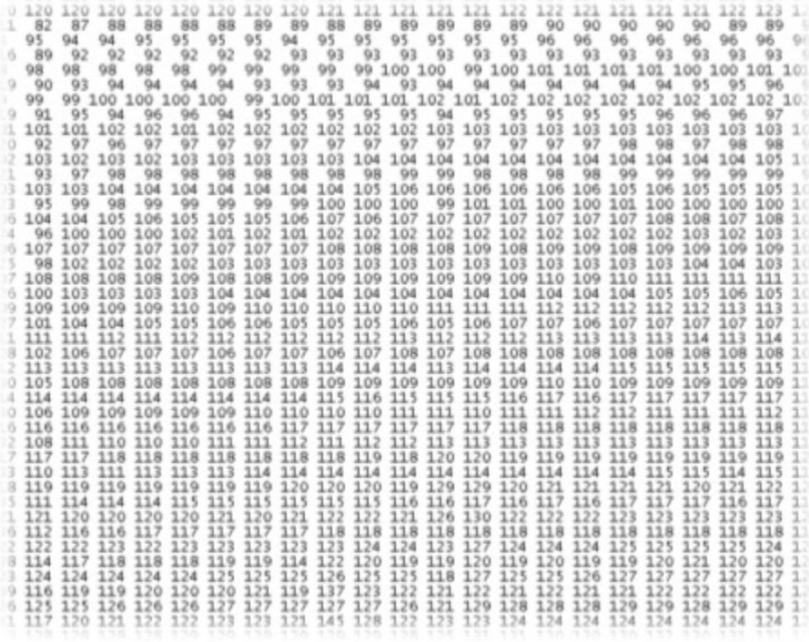
Vision Memo. No. 100.

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".

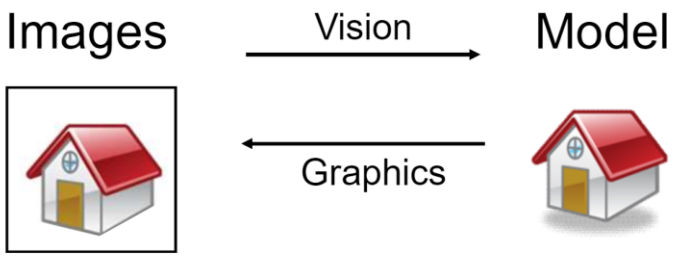
Computer vision is difficult



Gap between low level signal and high level meanings

Challenges: ill posed problem

- Ill-posed problem: real world much more complex than what we can measure in images.
- Impossible to literally “invert” image formation process.



Inverse problems: analysis and synthesis.

Challenges: large variations

Illumination

Object pose

Clutter

Occlusions

Intra-class appearance

Viewpoint

Challenge: intra-class variation

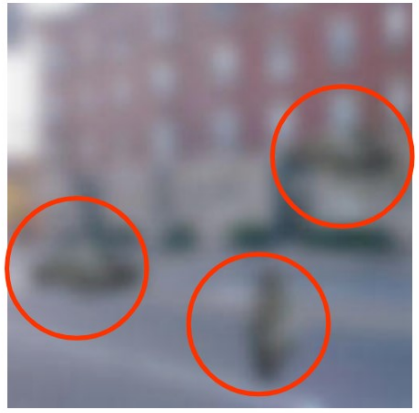





slide credit: Fei-Fei, Fergus & Torralba

Challenge: context

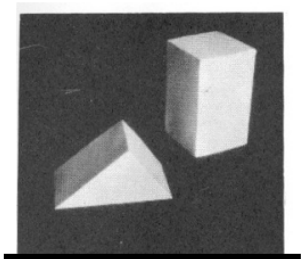
All encircled patterns are identical:



Challenges: complexity

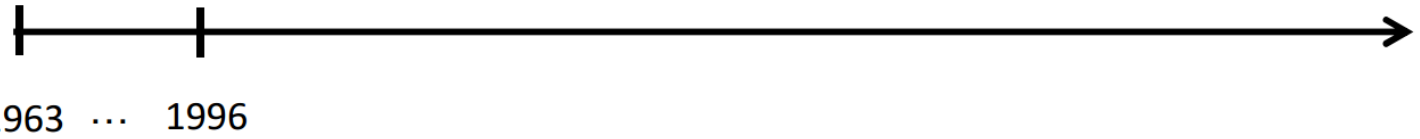
- High dimension of input data
  - Millions of pixels in an image
- Large number of categories
  - 30,000 recognizable object categories
- Large number of visual data
  - Billions of images online
  - 144K hours of new video on YouTube daily
- About half of the cerebral cortex in primates is devoted to processing visual information

Progress charted by dataset

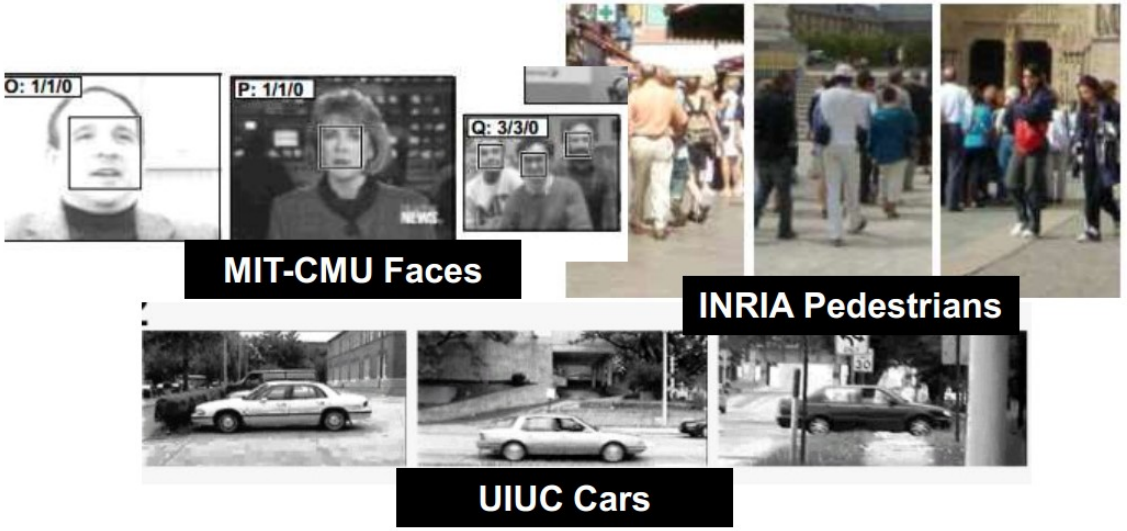


Roberts 1963

COIL



Progress charted by dataset

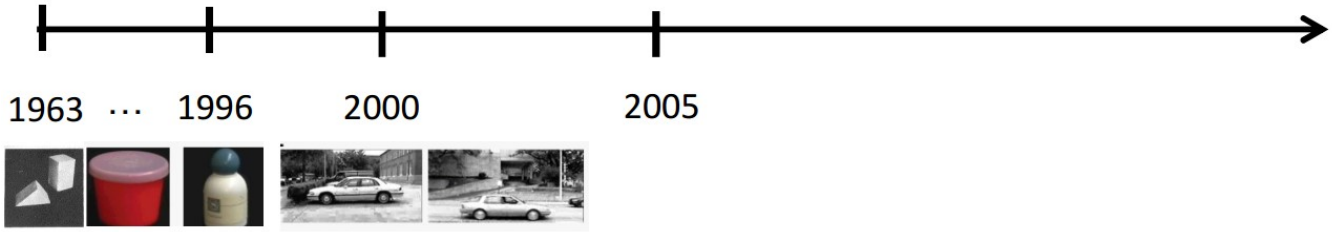


Progress charted by dataset

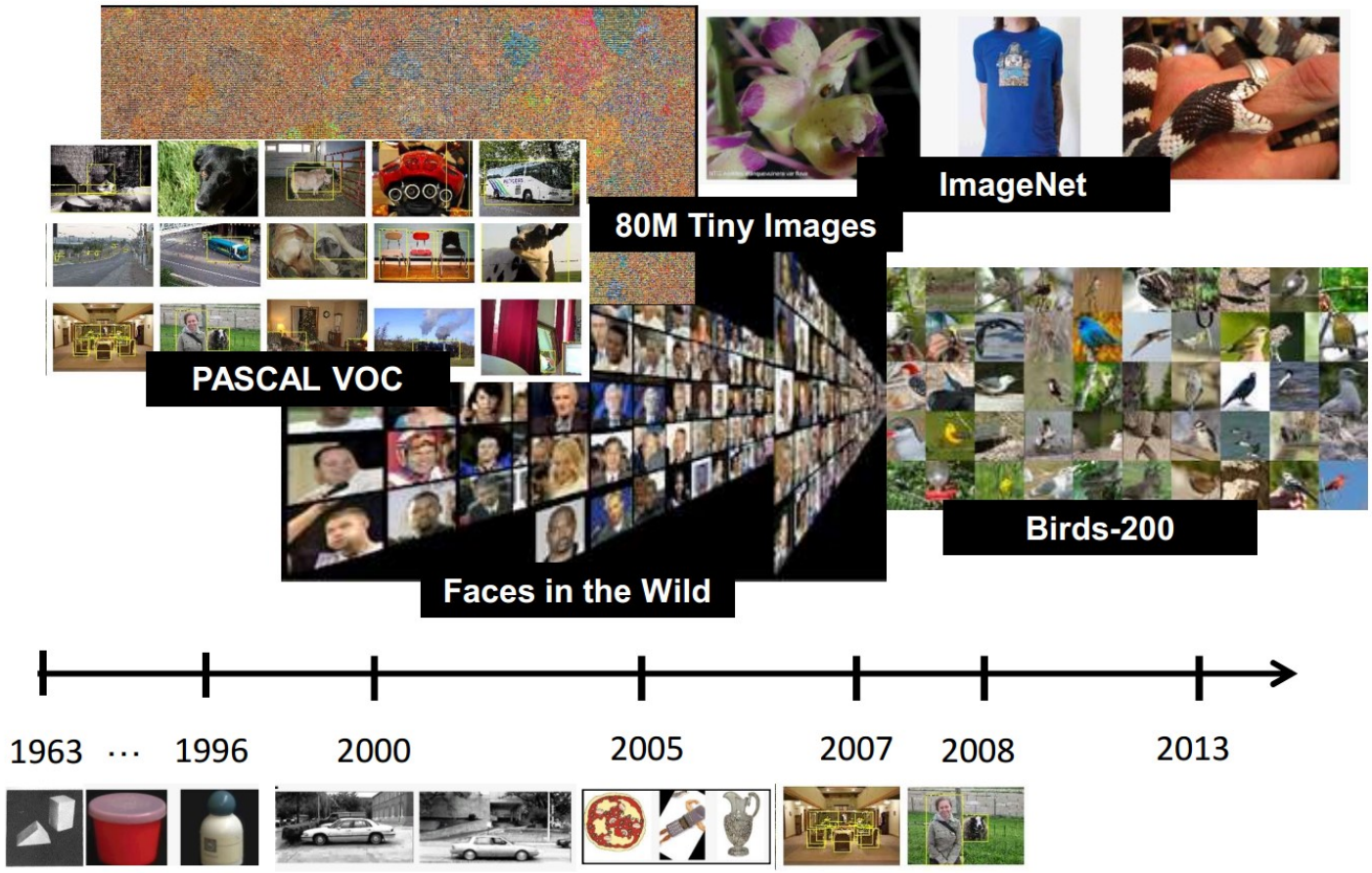
MSRC 21 Objects

Caltech-101

Caltech-256

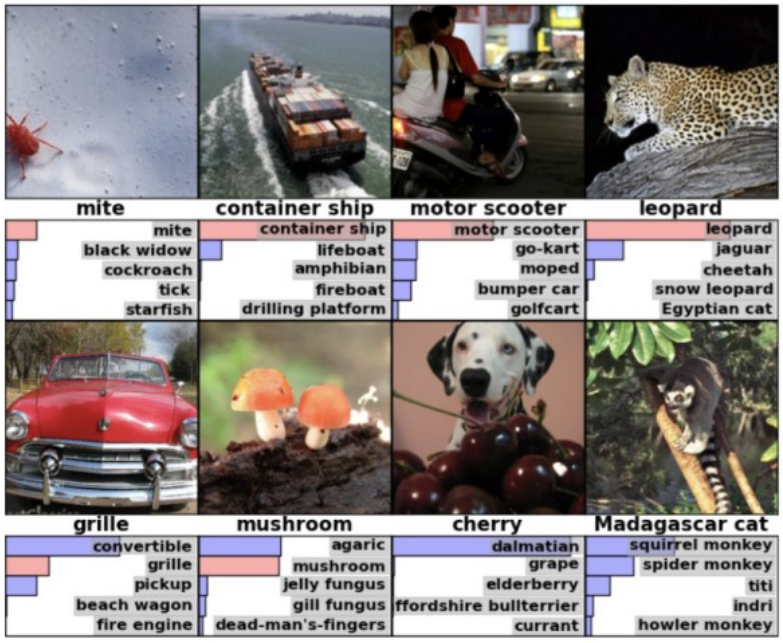


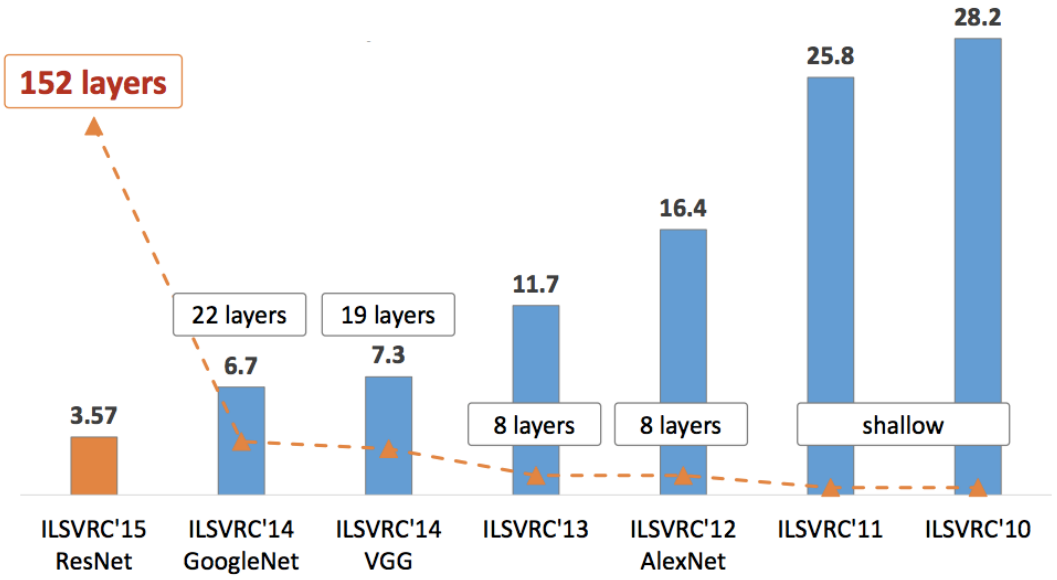
Progress charted by dataset



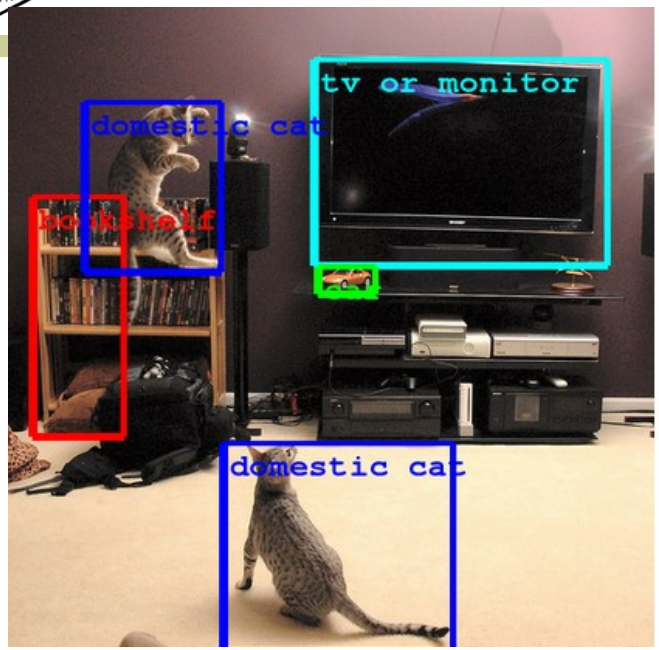
Recognition: General categories

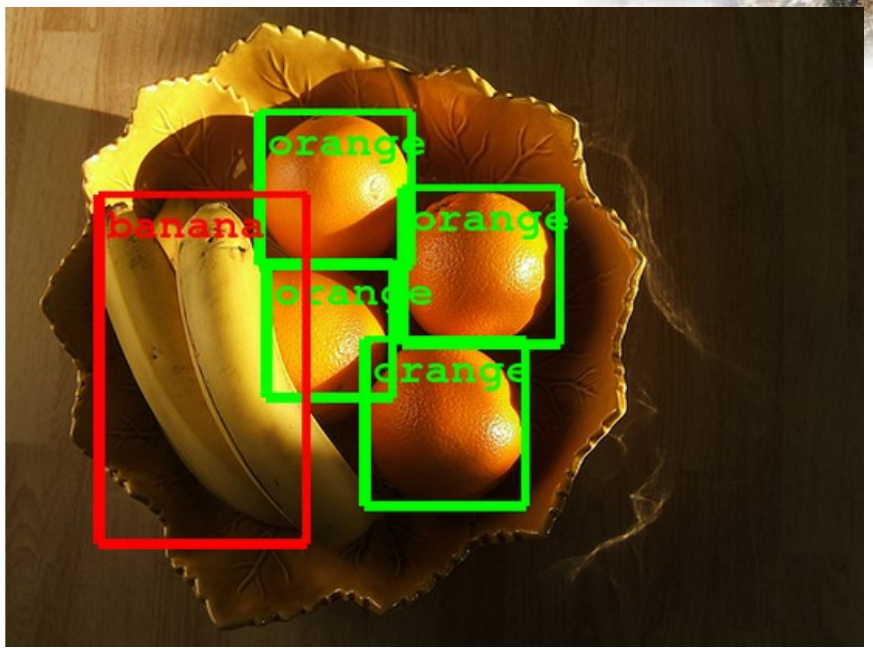
- ImageNet challenge



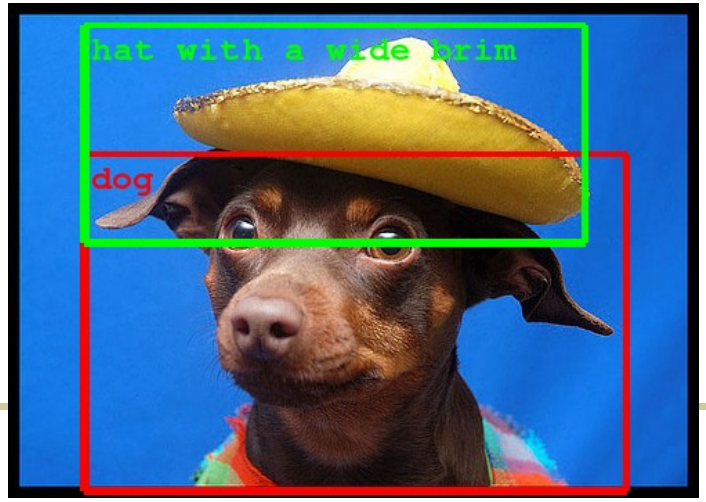


Recognition: General categories





- Computer Eyesight Gets a Lot More Accurate, NY Times Bits blog, August 18, 2014
- Building A Deeper Understanding of Images, Google Research Blog, September 5, 2014



Large scale recognition

clarifai ABOUT TECHNOLOGY API NEWS BLOG CAREERS CONTACT

Paste a url here...

USE THE URL CHOOSE A FILE INSTEAD

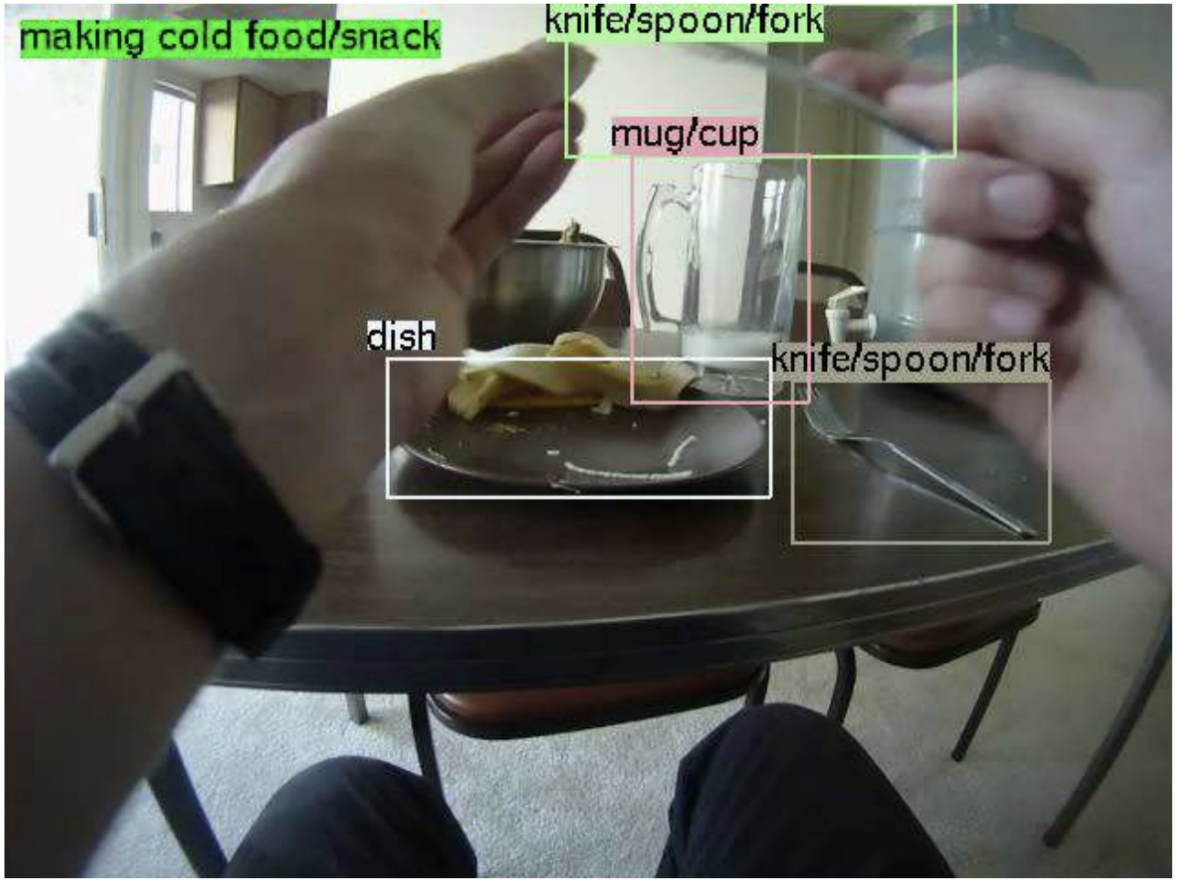
*By using the demo you agree to our terms of service

Predicted Tags

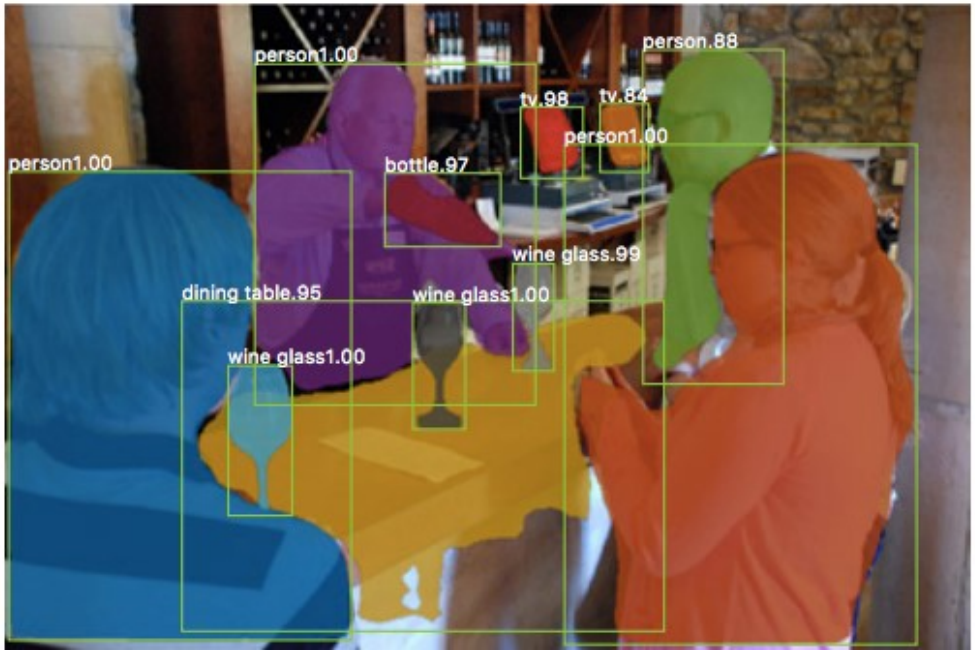
mammal livestock cattle pasture agriculture bovine farm nobody meadow grass

Similar Images

Recognition in first-person view



Object detection, instance segmentation



K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask R-CNN,
ICCV 2017 (Best Paper Award)

Image captioning

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 A person riding a motorcycle on a dirt road.	 Two dogs play in the grass.	 A skateboarder does a trick on a ramp.	 A dog is jumping to catch a frisbee.
 A group of young people playing a game of frisbee.	 Two hockey players are fighting over the puck.	 A little girl in a pink hat is blowing bubbles.	 A refrigerator filled with lots of food and drinks.
 A herd of elephants walking across a dry grass field.	 A close up of a cat laying on a couch.	 A red motorcycle parked on the side of the road.	 A yellow school bus parked in a parking lot.

Image generation

- Faces: 1024x1024 resolution, CelebA-HQ

T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, ICLR 2018

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv 2018

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv 2018

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

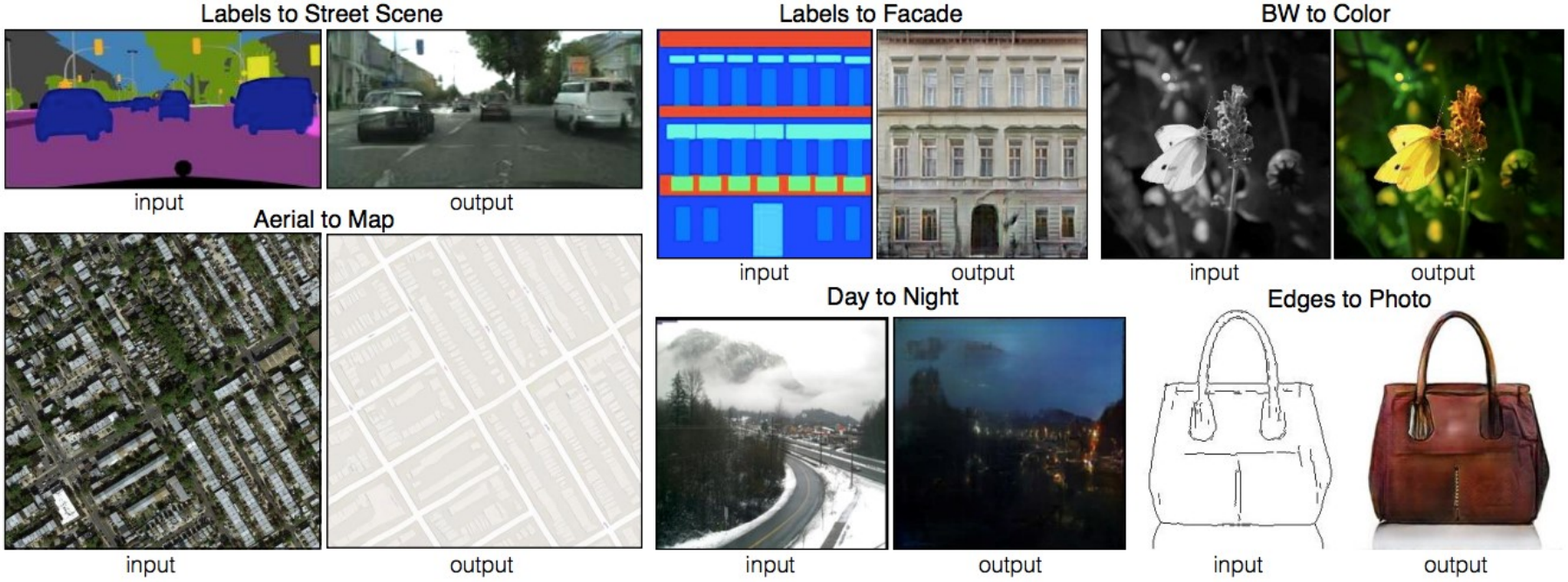
Easy classes

Difficult classes

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv 2018

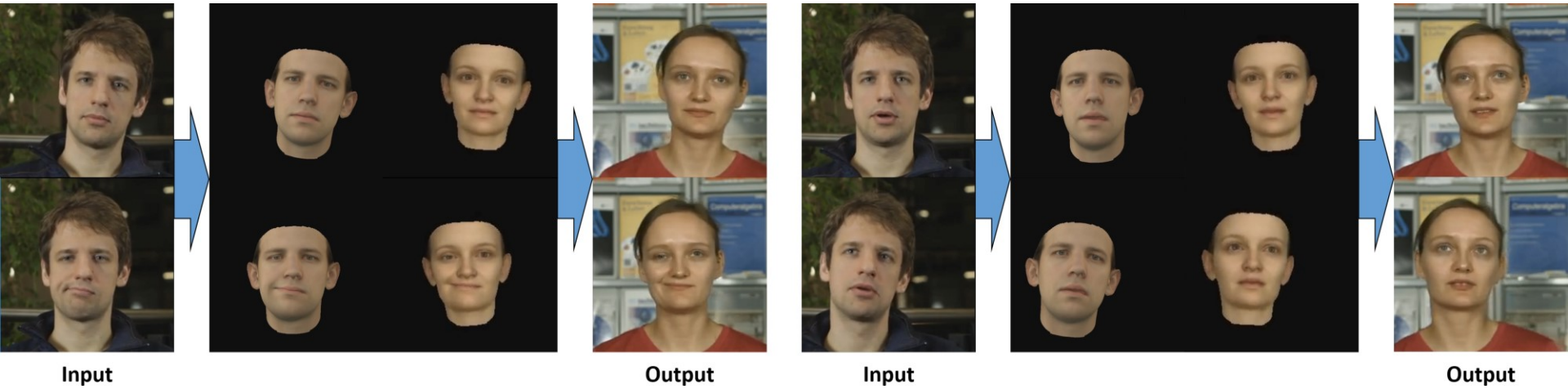
Image generation

- Image-to-image translation



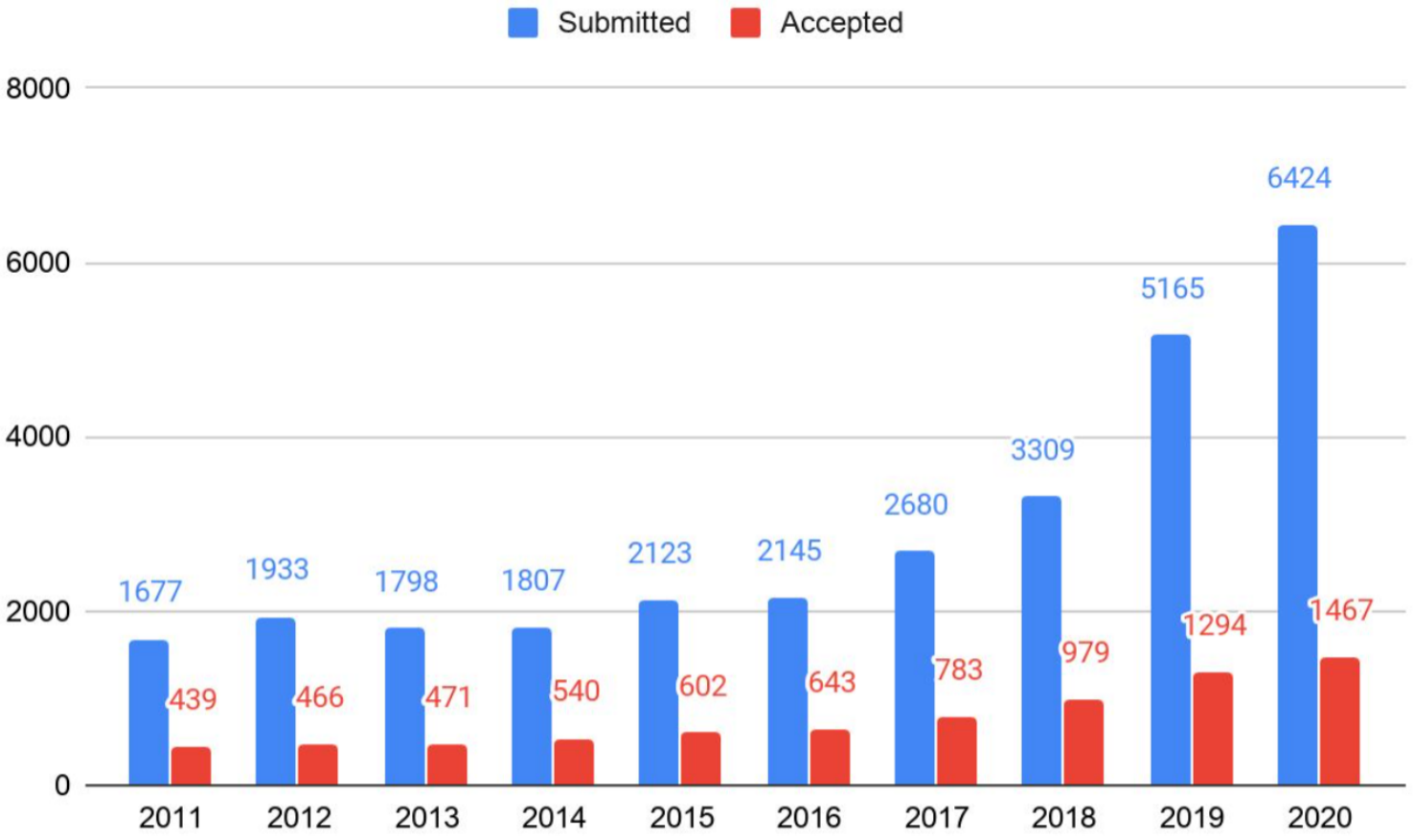
P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017

DeepFakes



"A quiet wager has taken hold among researchers who study artificial intelligence techniques and the societal impacts of such technologies. They're betting whether or not someone will create a so-called Deepfake video about a political candidate that receives more than 2 million views before getting debunked by the end of 2018" – IEEE Spectrum, 6/22/2018

CVPR:10 YEARS



Growth of the field

Diamond Sponsors

amazon Baidu 百度 Face++ 旷视 facebook Google Microsoft NVIDIA 商汤 sensetime

Platinum Sponsors

Alibaba Group 阿里巴巴集团 ANT FINANCIAL AURORA ByteDance 字节跳动 TOUTIAO AI LAB COGNEX DiDi HUAWEI IBM intel AI lyft

MOMENTA NAVER LINE oppo Qualcomm SAMSUNG 依图 YITU Tencent 腾讯 UBER XILINX

Gold Sponsors

360 Adobe AiDEE ARGO Ascent CRUISE CVTE Dream-Future 视源股份 Disney Research dji gumgum here HIKVISION HRI Honda Research Institute US

intellifusion 云天励飞 iRobot JD.COM MALONG 码隆 Mighty AI MITSUBISHI ELECTRIC RESEARCH LABORATORIES, INC NIO nullmax 纽劢科技 Snap Inc. WAYMO YANGTZE DELTA HIGH-TECH CITY 张江长三角科技城

Silver Sponsors

3M AutoX AXON cadence CEVA Cirrascale CLOUD SERVICES cloudfactory coupang DATATANG DeepMotion DroneDeploy figure eight HITACHI Inspire the Next Horizon Robotics 地平线

iMerit human-empowered computing ISEE JingChi 景驰科技 JUMIO Kitware KPST Lawrence Livermore National Laboratory magic leap mapbox MathWorks Matroid LABS NAVER LABS ObjectVideo LABS OCTI OUSTER

中国平安 平安科技 PING AN TECHNOLOGY Playment SAP scale Skydio SPORTLOGIQ SURFING TECH THE HOME DEPOT TOYOTA RESEARCH INSTITUTE VCC VISUAL COMPUTING CENTER Verisk 唯品会 vip.com wayfair Zillow GROUP ZOOX

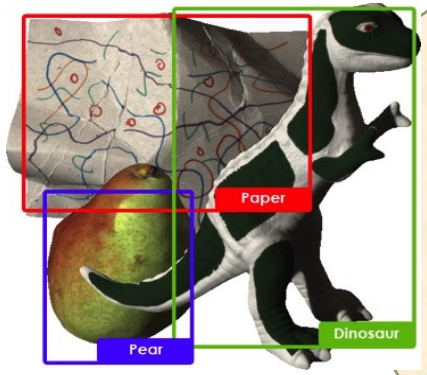
IEEE computer society

CVF

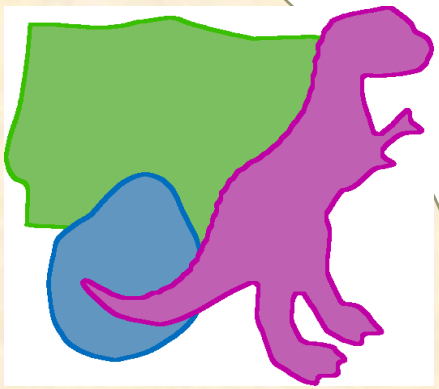
Marr's vision framework

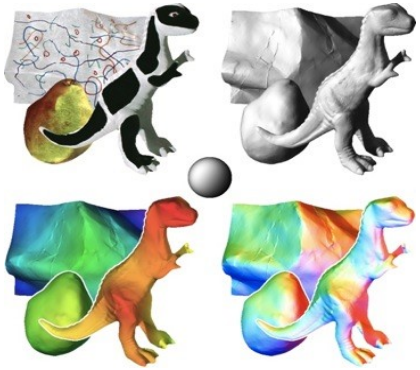
- computational level: what does the system do (e.g.: what problems does it solve or overcome) and similarly, why does it do these things
- algorithmic/representational level: how does the system do what it does, specifically, what representations does it use and what processes does it employ to build and manipulate the representations
- implementational/physical level: how is the system physically realised (in the case of biological vision, what neural structures and neuronal activities implement the visual system)

Malik's Perspective



Recognition

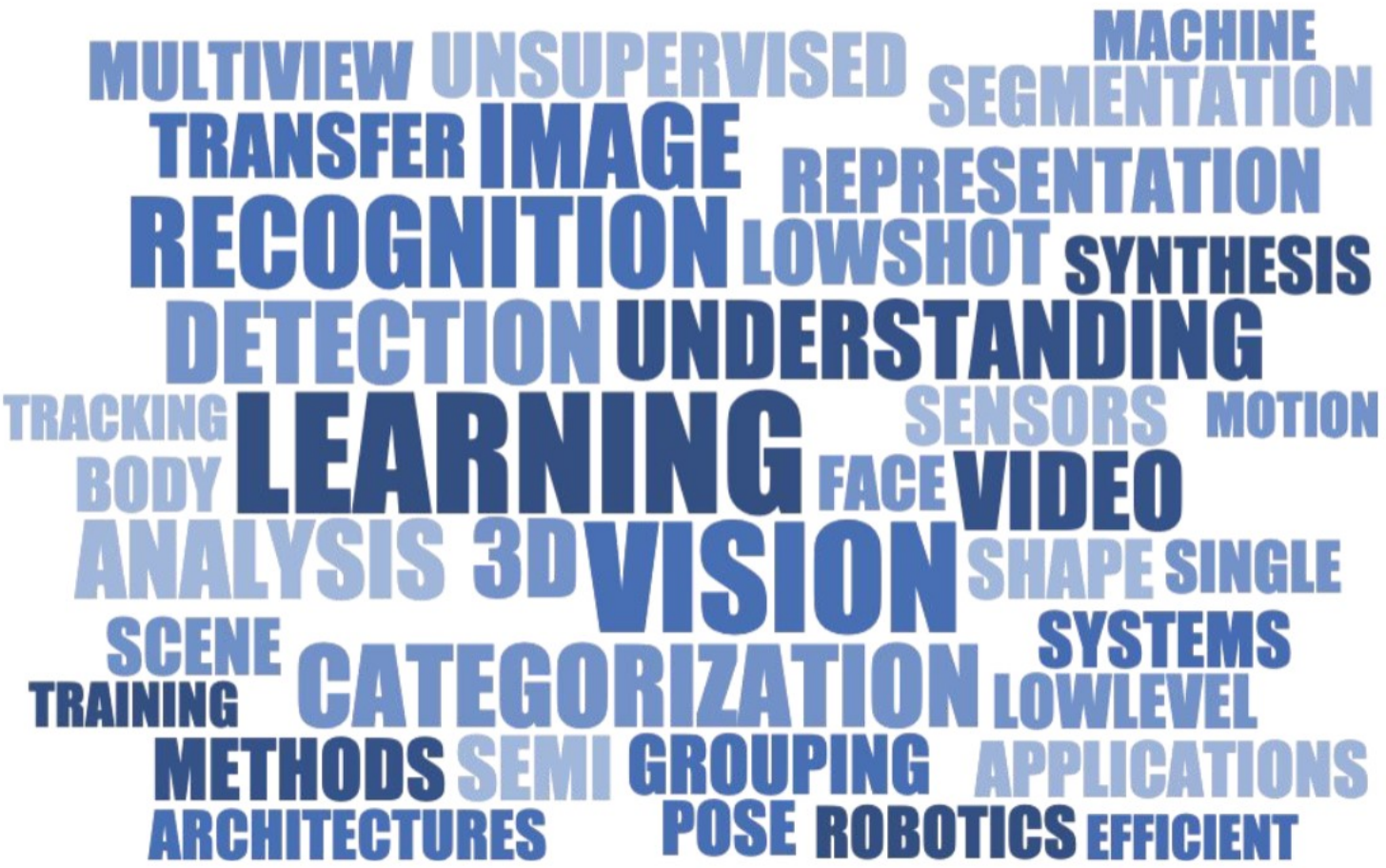




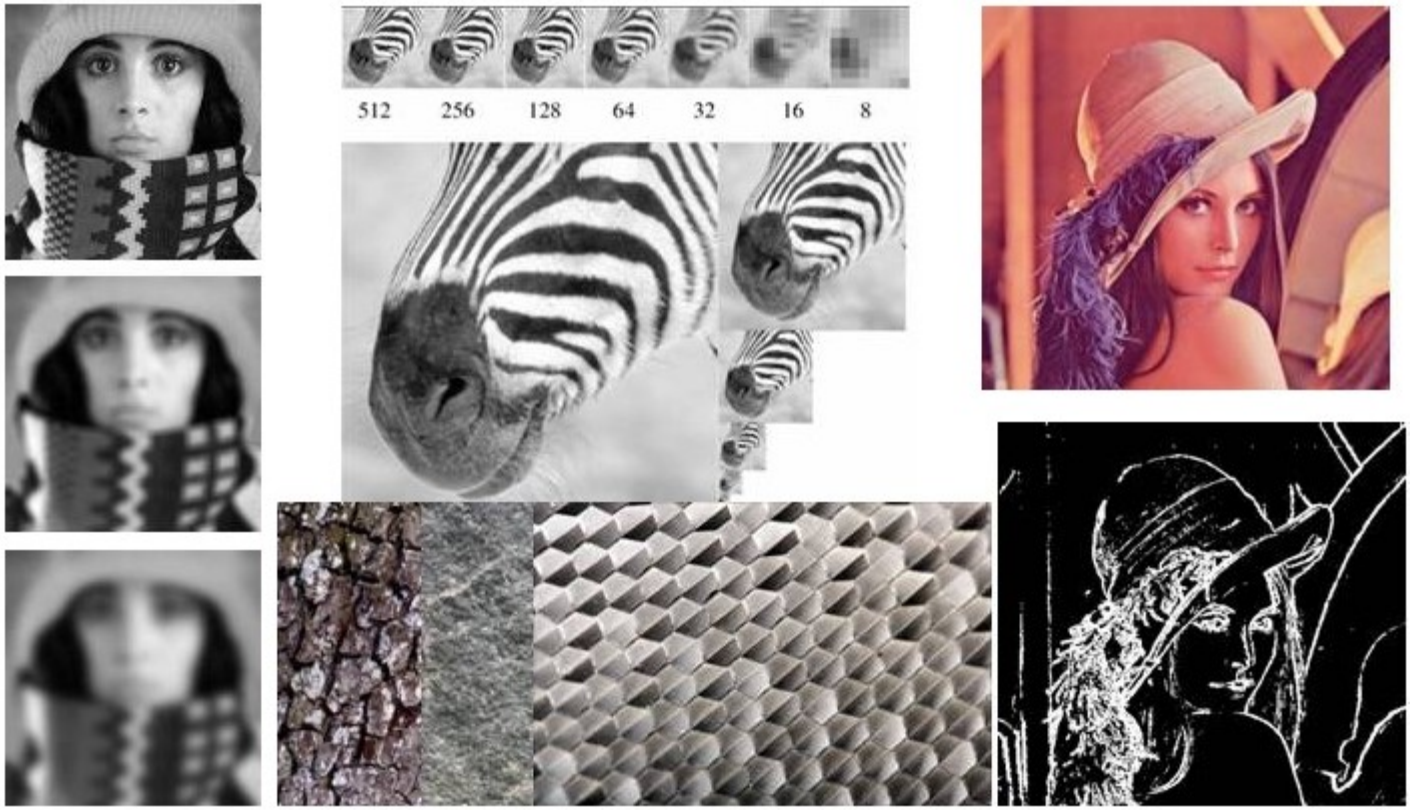
Reconstruction

Reorganization

Popular Areas in CVPR

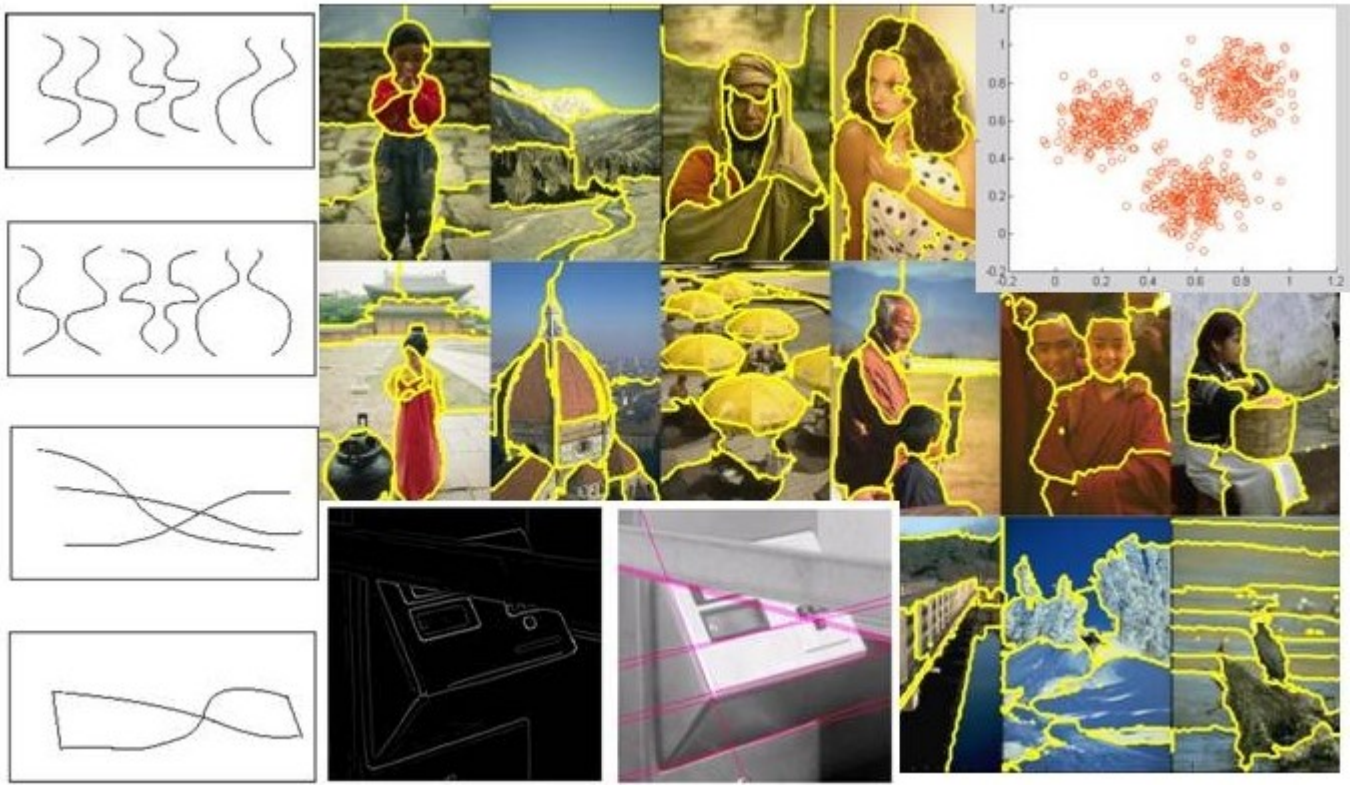


I. Features and filters



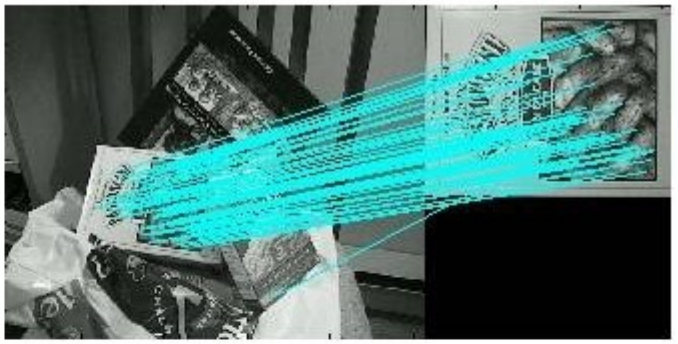
Transforming and describing images: textures, colors, edges

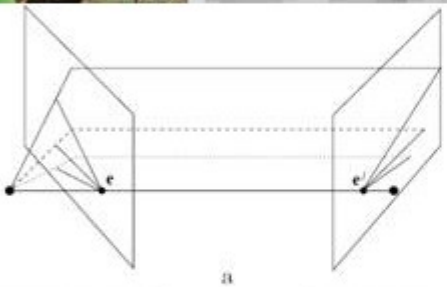
II. Grouping and fitting



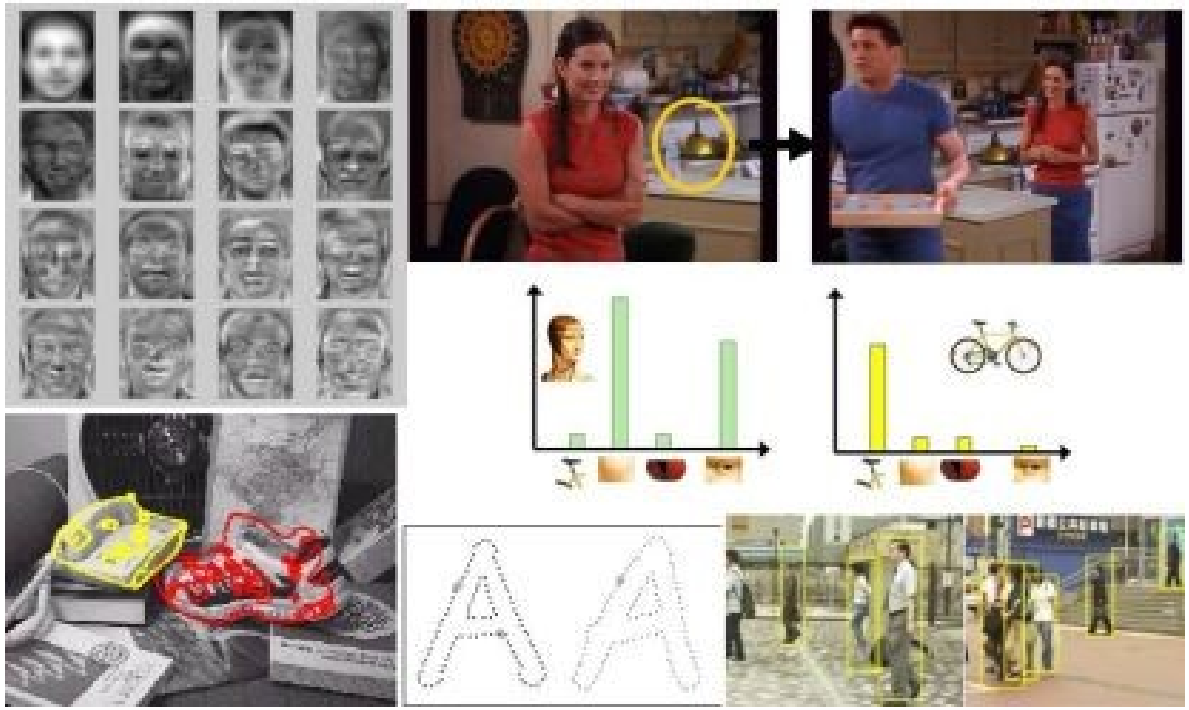
Clustering, segmentation, fitting: what parts belong together

III Matching and Alignment



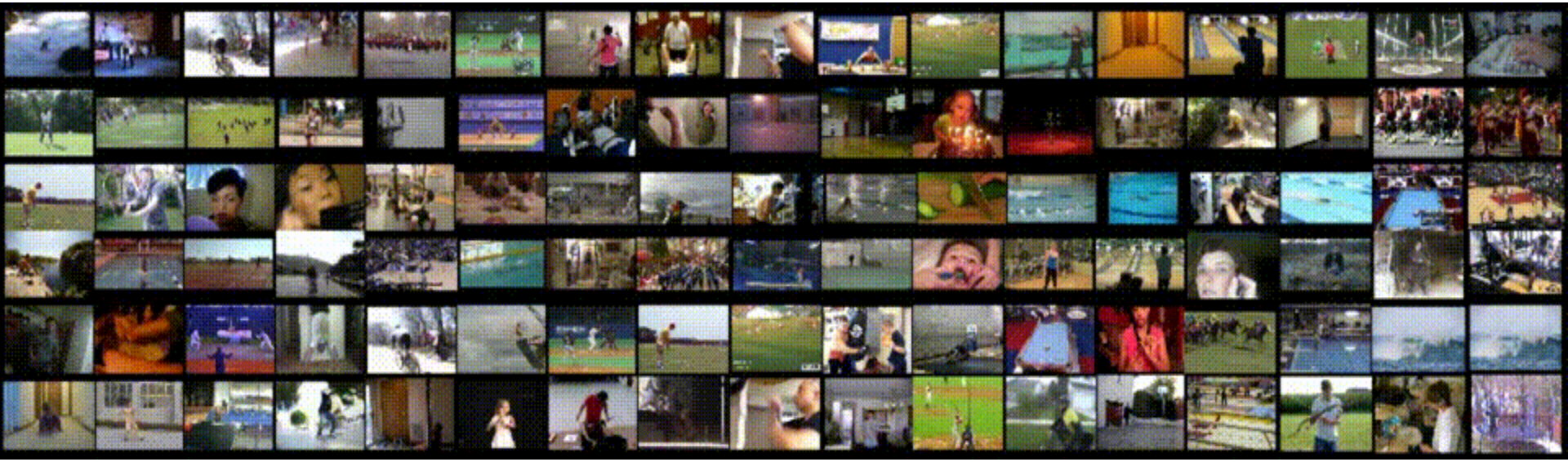


IV. Recognition



Recognizing categories and deep learning techniques

V. Video Understanding



Understanding videos: motion, action, event etc.

Course overview

- Chapter 1. Introduction
- Chapter 2. Images and Filter
- Chapter 3. Frequency Domain and Sampling
- Chapter 4. Template, Pyramid, and Filter Banks
- Chapter 5. Edges
- Chapter 6. Segmentation and Grouping
- Chapter 7 & 8. Interest Points
- Chapter 9. Fitting and Alignment
- Chapter 10. Alignment and Instance Recognition
- Chapter 11. Image Classification
- Chapter 12. Object Detection
- Chapter 13. Course summary

Summary

- What is computer vision
- Computer vision is useful
- Computer vision is difficult
- Computer vision is fast developing
- Course overview

Important note:

In general, computer vision does not work

(except in certain situations/conditions)

Hope you enjoy the course!