ECS 269: Visual Recognition
Fall 2019

Yong Jae Lee
Department of Computer Science

Plan for today

- Topic overview
- Course requirements and administrivia
- Introductions
- Syllabus tour

Computer Vision

- Let computers see!
- Automatic understanding of visual data (i.e., images and video)
  - Computing properties of the 3D world from visual data *(measurement)*
  - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities *(perception and interpretation)*

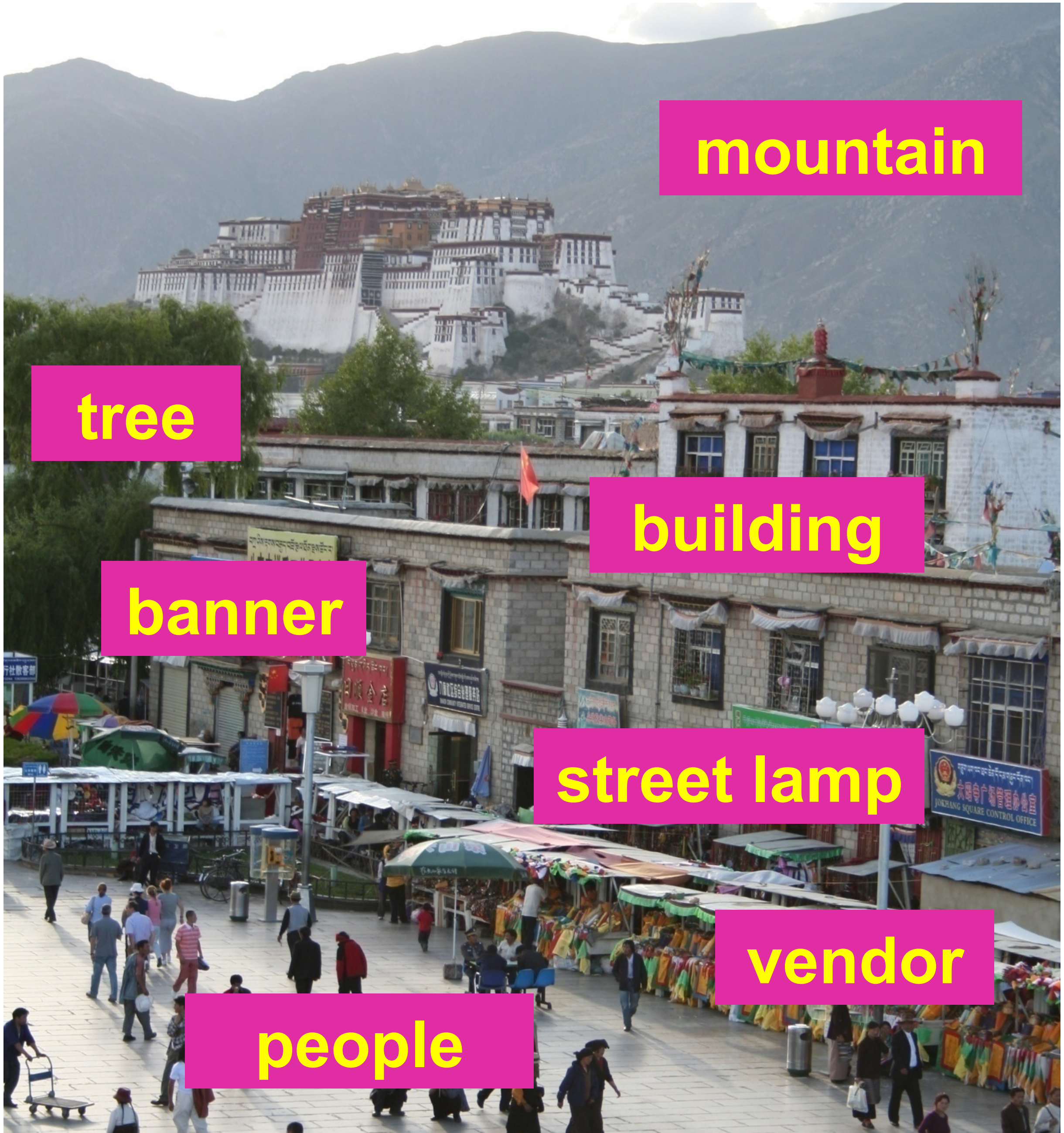
What does recognition involve?

Fei-Fei Li

Detection: are there people?

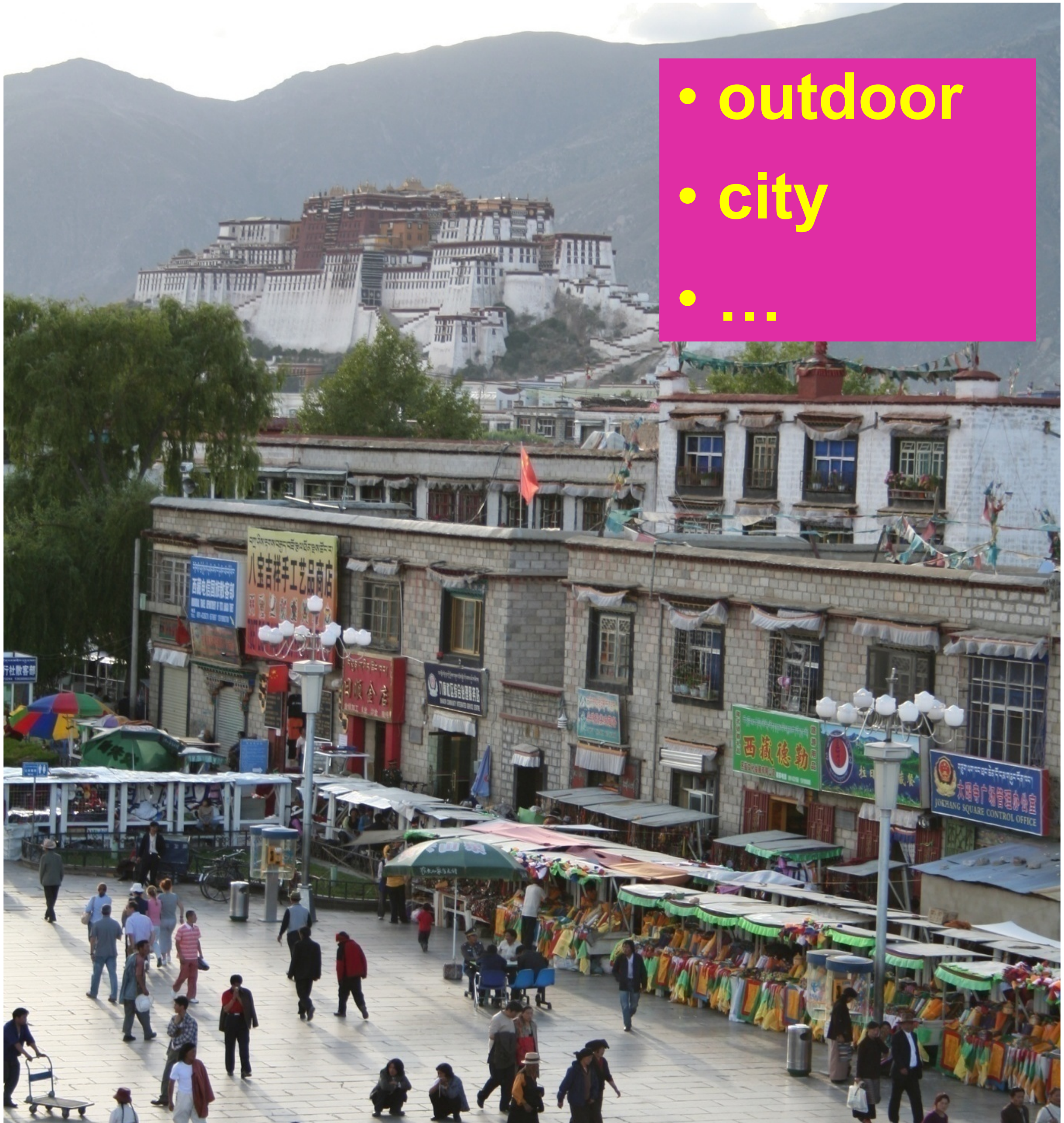
Activity: What are they doing?

Object categorization



Instance recognition

Scene categorization

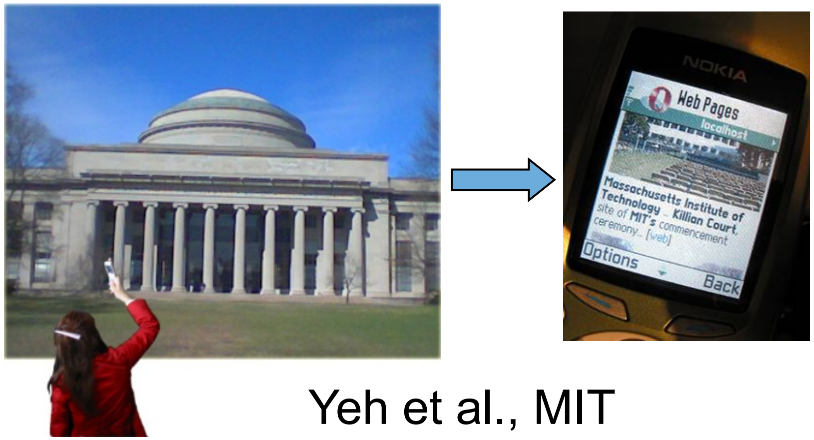


Attribute recognition

Why recognition?

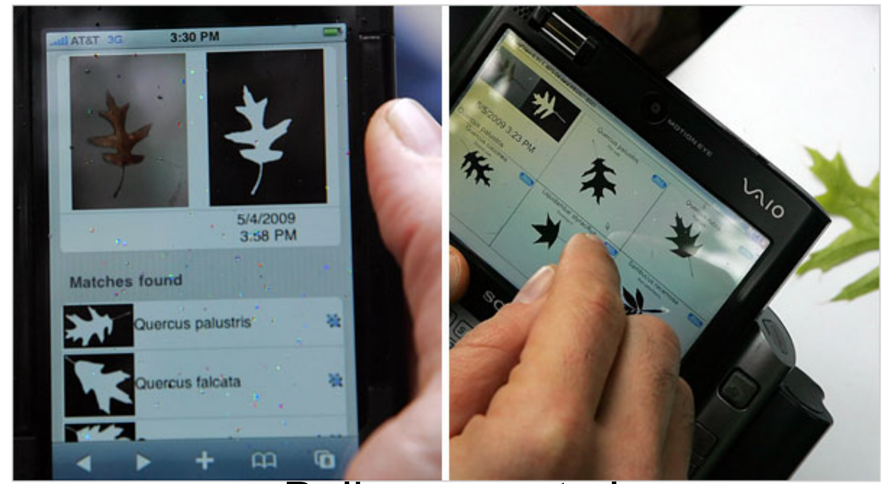
- Recognition a fundamental part of perception
  - e.g., robots, autonomous agents
- Organize and give access to visual content
  - Connect to information
  - Detect trends and themes

Posing visual queries



Yeh et al., MIT

Digital Field Guides Eliminate the Guesswork

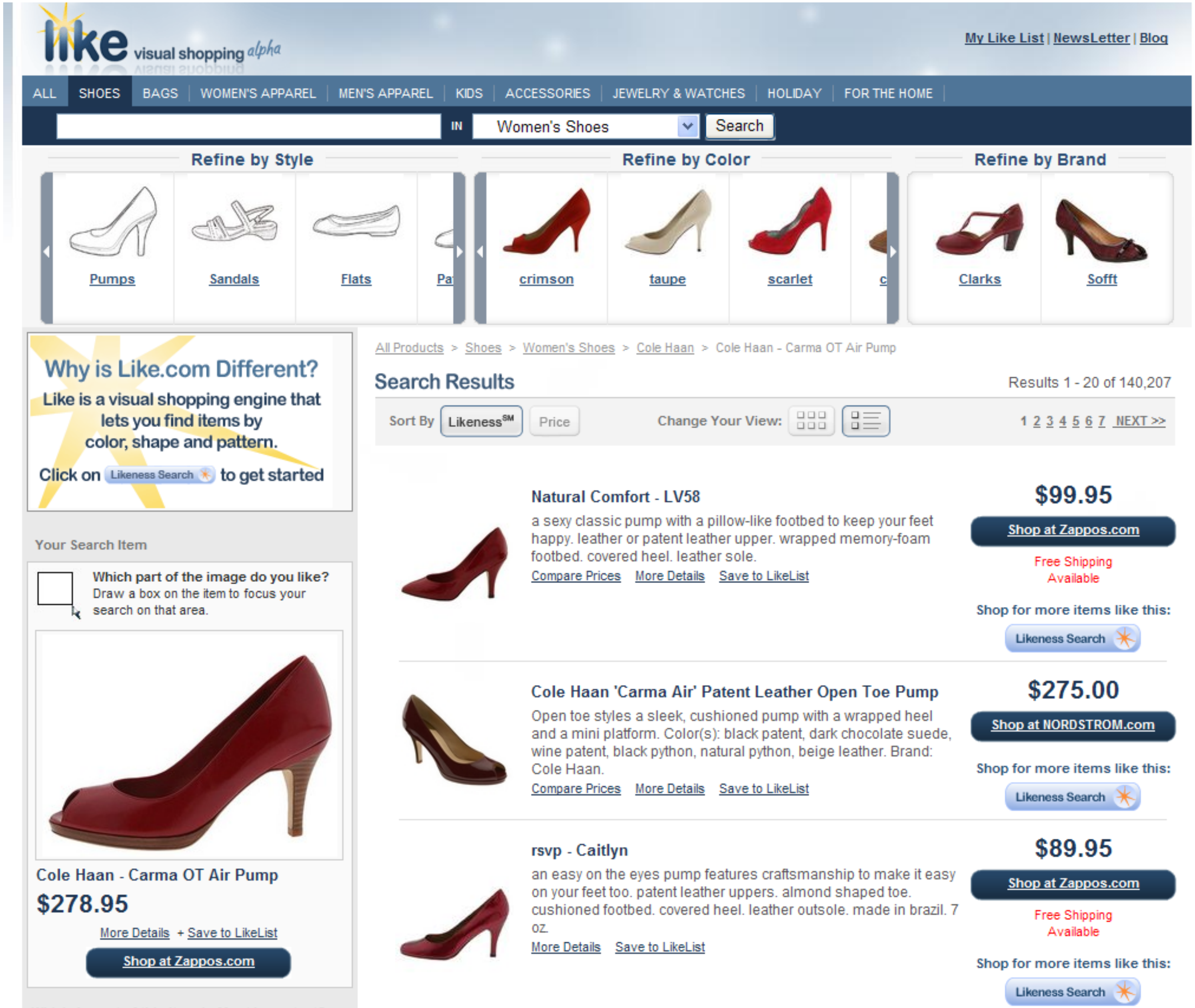


Belhumeur et al.

Kooaba, Bay & Quack et al.

Kristen Grauman

Finding visually similar objects



Interactive systems

Shotton et al.

Autonomous agents

Mars rover

Google self-driving car

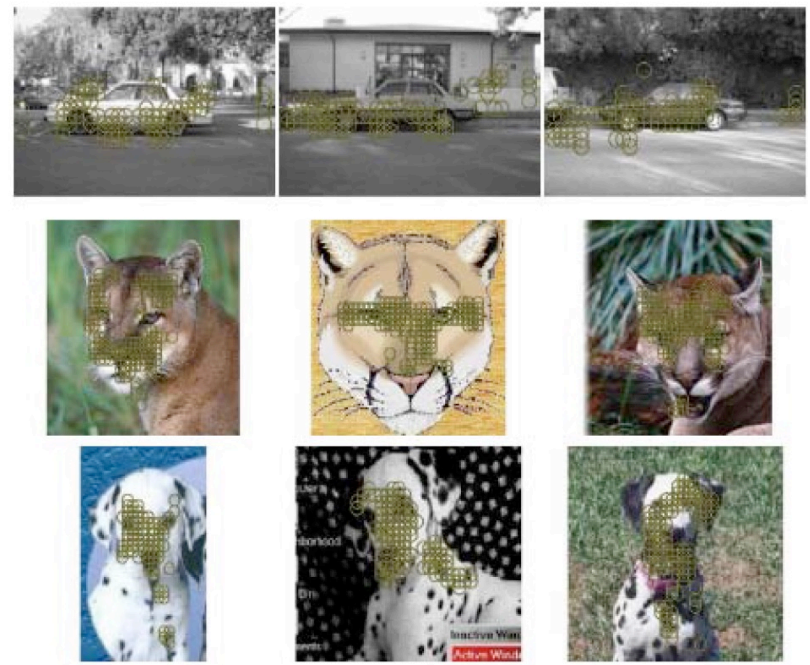
Discovering visual patterns

Wang et al.

Actions

Doersch et al.

Characteristic elements



Lee & Grauman

Categories

What are the challenges?

What we see

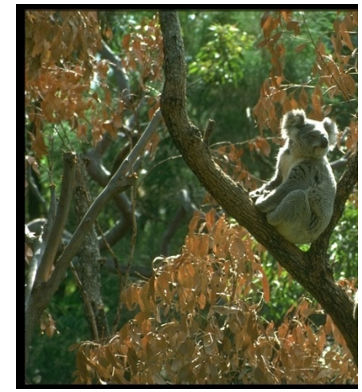
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Challenges: robustness

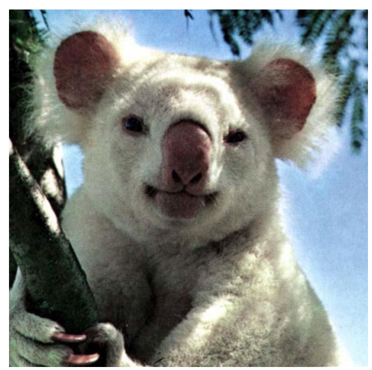
Illumination

Object pose

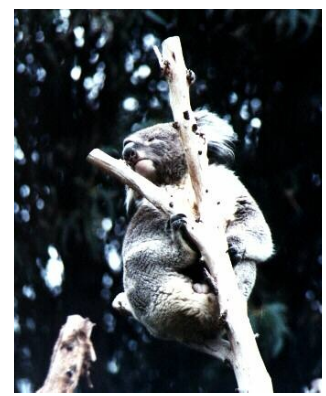


Clutter

Occlusions



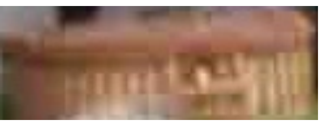
Intra-class appearance



Viewpoint

Kristen Grauman

Challenges: context and human experience



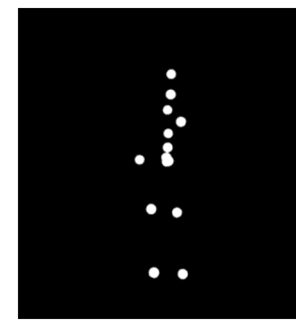
Context cues

Kristen Grauman

Challenges: context and human experience

Context cues

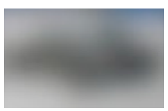
Function



Dynamics

Kristen Grauman

Challenges: context and human experience



Fei Fei Li, Rob Fergus, Antonio Torralba

Challenges: context and human experience

Fei Fei Li, Rob Fergus, Antonio Torralba

Challenges: context and human experience

Fei Fei Li, Rob Fergus, Antonio Torralba

Challenges: scale, efficiency

- Half of the cerebral cortex in primates is devoted to processing visual information

Challenges: scale, efficiency

10 billion images

250 billion images

1 billion images served daily

300 hours uploaded per minute

10 billion images

From :

Almost 90% of web traffic is visual!

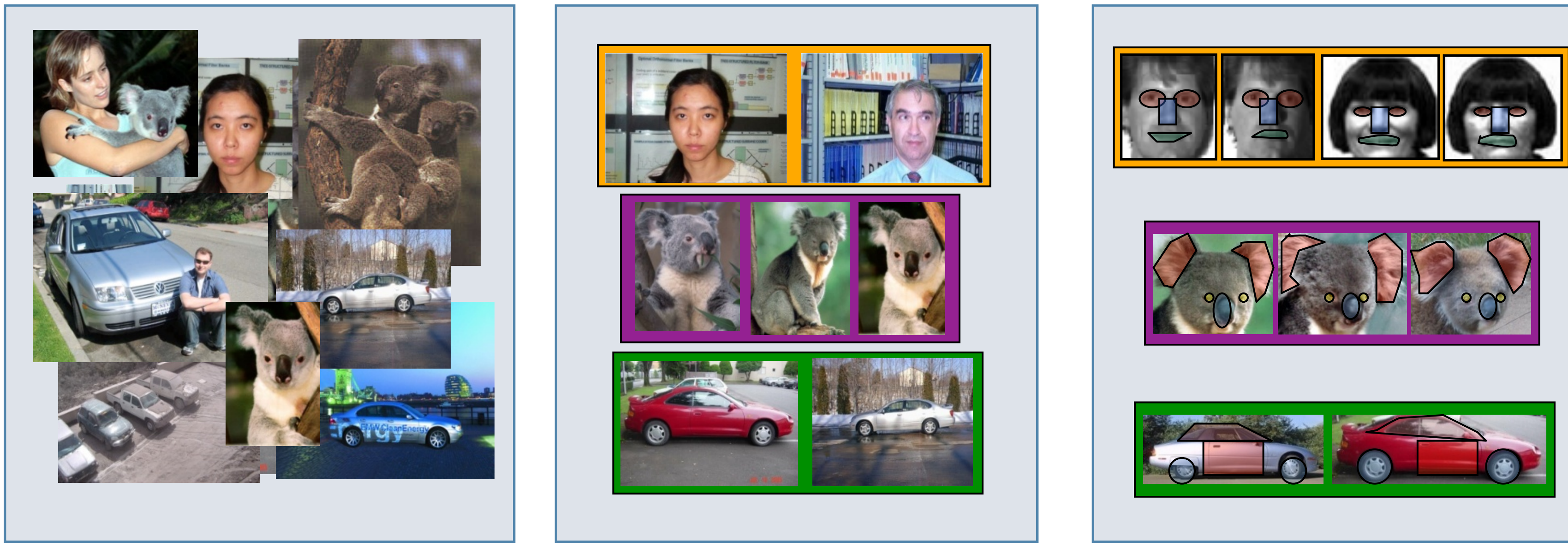
Challenges: scale, efficiency

Biederman 1987


~10,000 to 30,000

Challenges: learning with minimal supervision

Less — **More**



Unlabeled, multiple objects

Classes labeled, some clutter

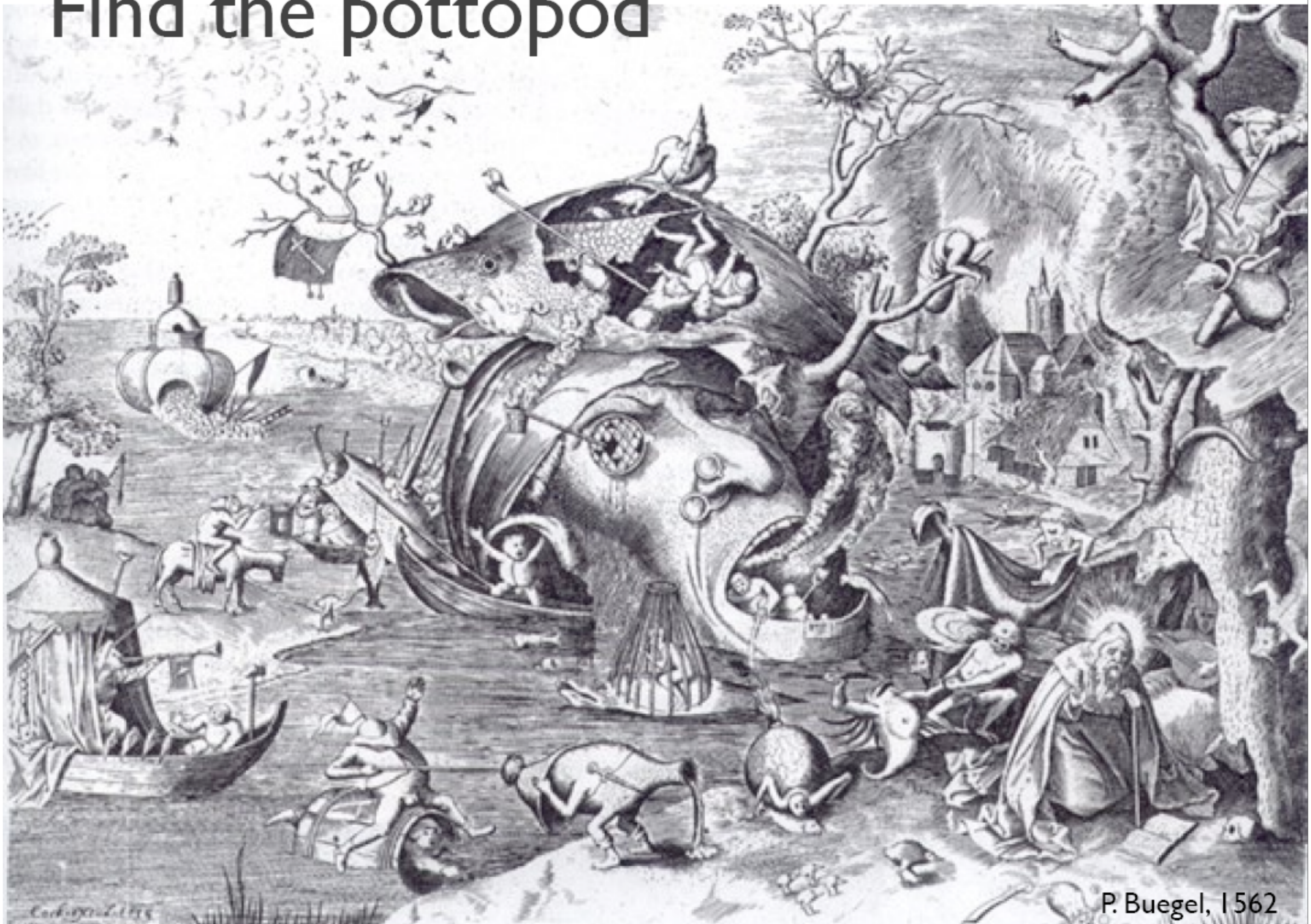
Cropped to object, parts and classes labeled

Kristen Grauman

This is a pottopod

Slide from Pietro Perona, 2004 Object Recognition workshop

Find the pottopod



Slide from Pietro Perona, 2004 Object Recognition workshop

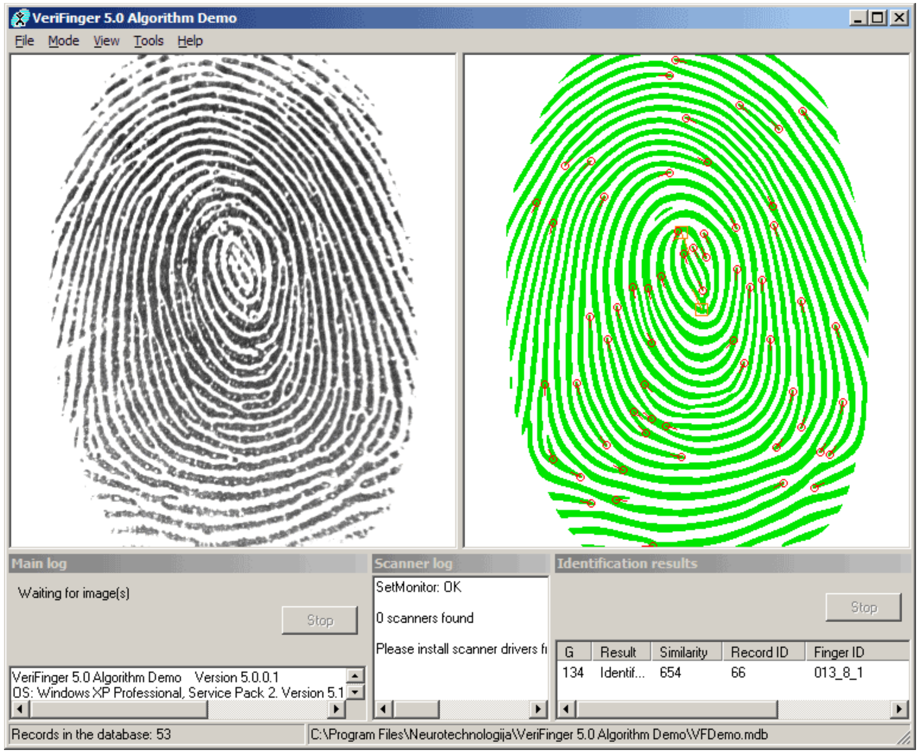
What works today

- Reading license plates, zip codes, checks

Source: Lana Lazebnik

What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition



Source: Lana Lazebnik

What works today

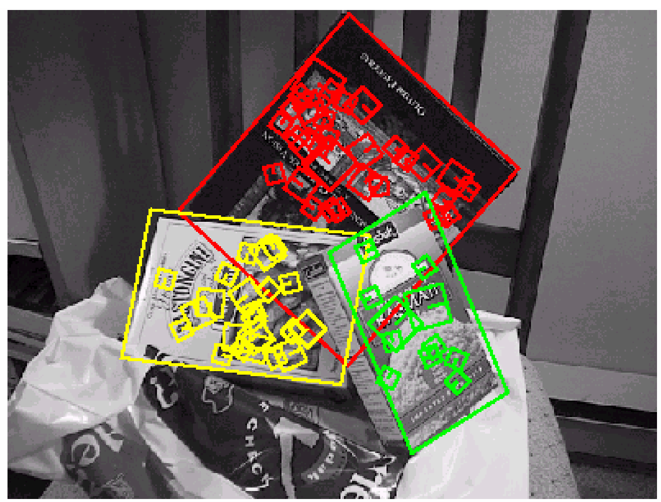
- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

[Face priority AE] When a bright part of the face is too bright

Source: Lana Lazebnik

What works today

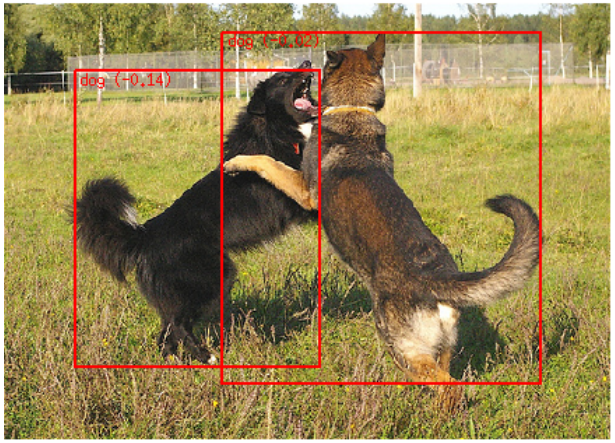
- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)

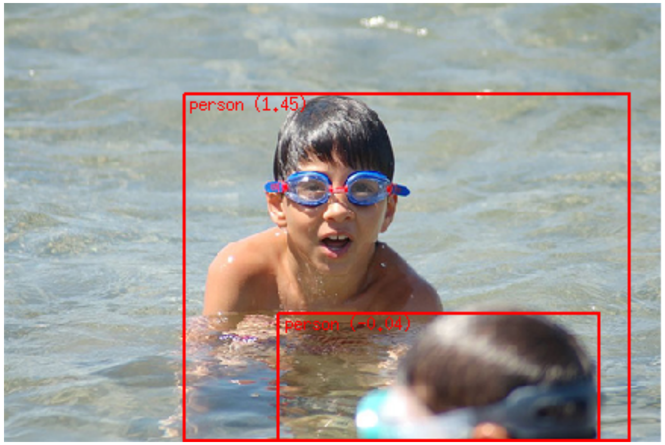


Source: Lana Lazebnik

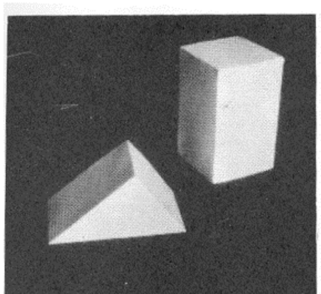
What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)
- Recognition of generic categories(*)!





Progress charted by datasets



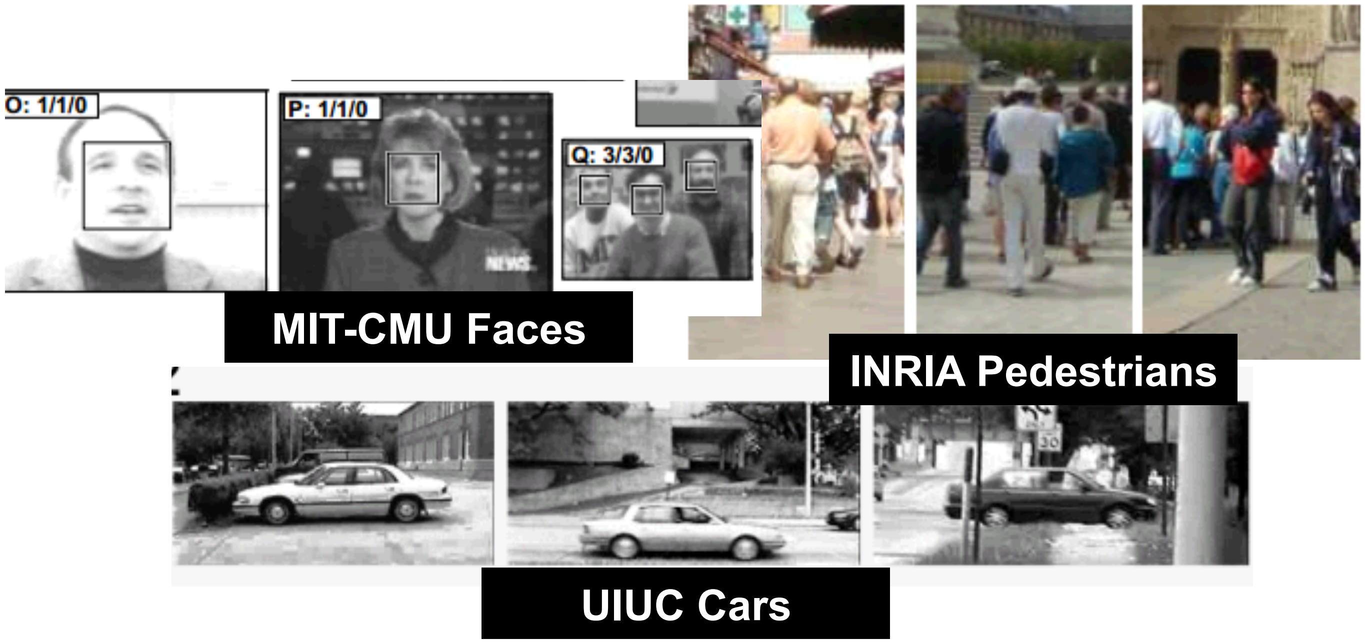
Roberts 1963

COIL

1963 ··· 1996

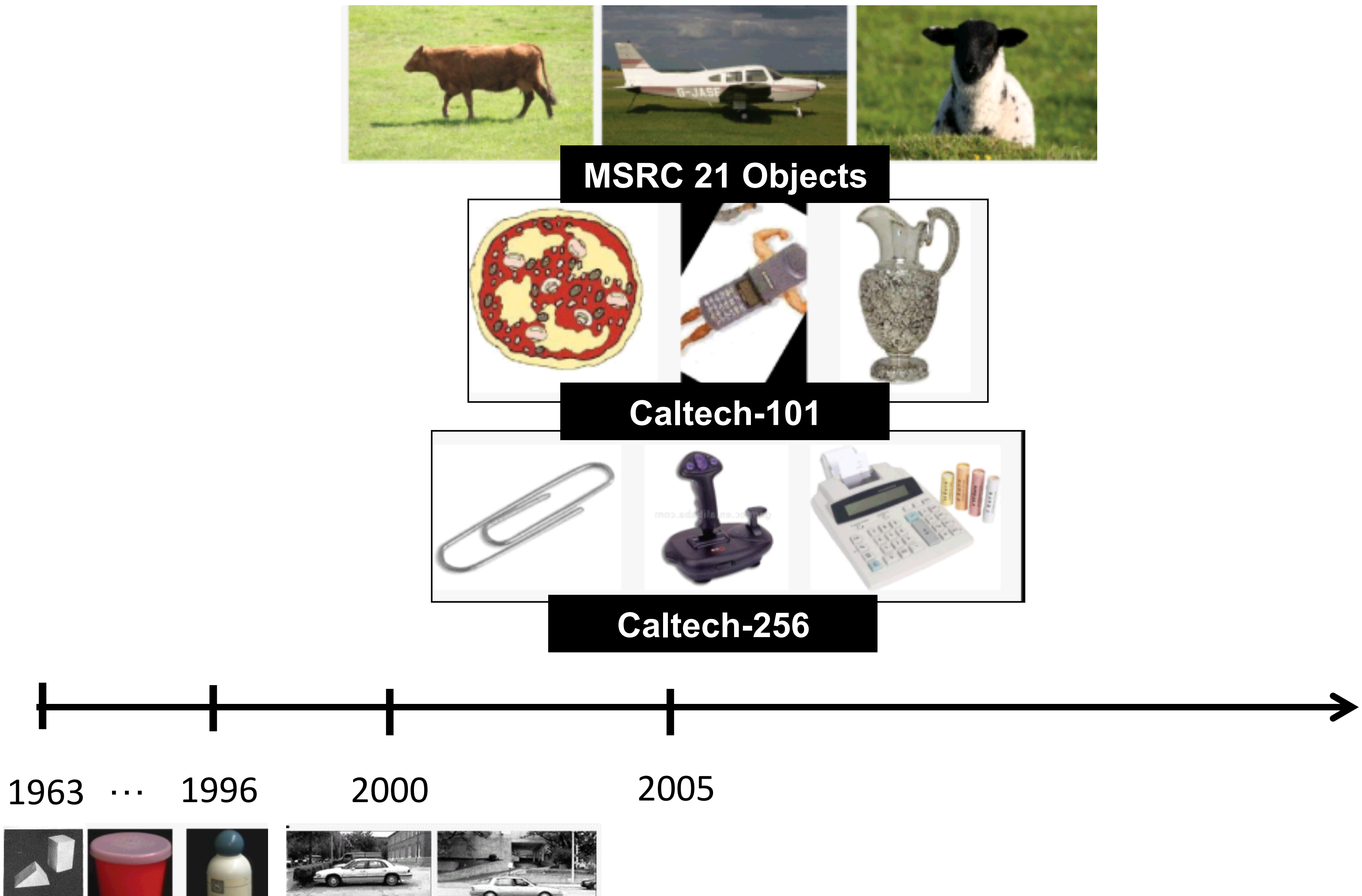
Slide credit: Kristen Grauman

Progress charted by datasets



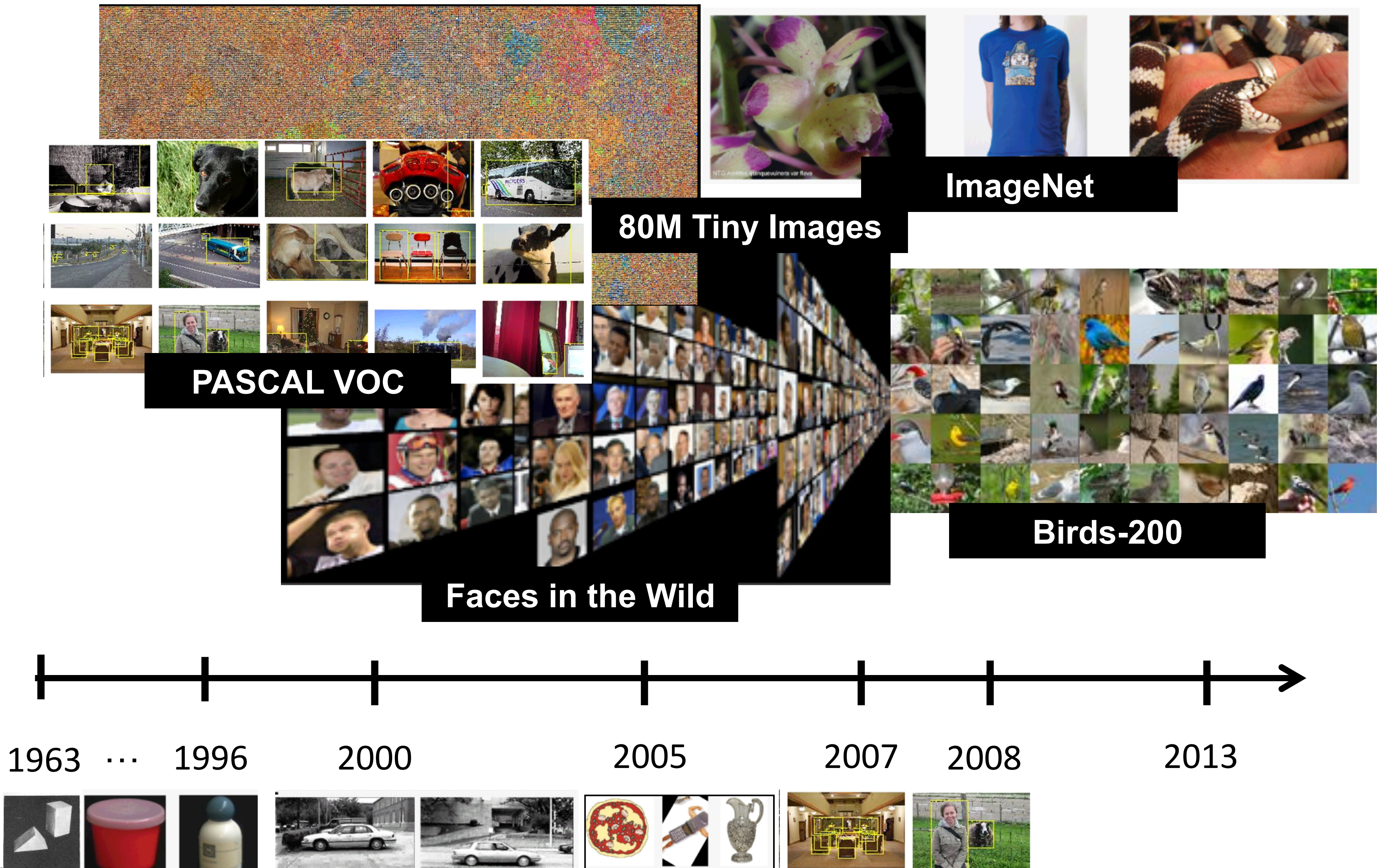
Slide credit: Kristen Grauman

Progress charted by datasets



Slide credit: Kristen Grauman

Progress charted by datasets



Slide credit: Kristen Grauman

Evolution of methods

- Hand-crafted models
- 3D geometry
- Hypothesize and align

- Hand-crafted features
- Learned models
- Data-driven

- "End-to-end" learning of features and models*

→

* Labeled data, fast compute (GPUs).

Course Information

- Website: https://sites.google.com/a/ucdavis.edu/ecs-269-visual-recognition-fall-2019/

- Canvas:

  https://canvas.ucdavis.edu/courses/361432

- 4 Units
- Office hours: by appointment (email)
- Email: start subject with "[ECS269]"

TA

- Krishna Kumar Singh
- ecs269fall2019TA@gmail.com
- PhD student in CS

- Yuheng Li
- ecs269fall2019TA@gmail.com
- MS student in CS

This course

- Survey state-of-the-art research in computer vision
- High-level vision and learning problems with focus on recent deep learning techniques

Goals

- Understand, analyze, and discuss state-of-the-art techniques
- Identify interesting open questions and future directions

Prerequisites

- Course in computer vision (ECS 174) or machine learning (ECS 171)

Requirements

- Class Participation **[15%]**
- Paper Reviews **[25%]**
- Paper Presentations (1 time) **[30%]**
- Final Project **[30%]**

Class participation

- Actively participate in discussions
- **Read assigned paper** before each class
- Email TA after class if you contributed in any way to the discussions that day (1 sentence summary)

Paper reviews

- **Detailed review of paper** before each class
- *At most* one page in length
  - A summary of the paper (2-3 sentences)
  - Main contributions
  - Strengths and weaknesses
  - Experiments convincing?
  - Extensions?
  - Additional comments, including unclear points, open research questions, and applications

- Upload to Canvas by **11:59 pm the day before each class**

Paper presentations

- 1 time during the quarter (in groups of 2)
- ~20 minutes long, well-organized, and polished:
  - Clear statement of the problem
  - Motivate why the problem is interesting, important, and/or difficult
  - Describe key contributions and technical ideas
  - Experimental setup and results
  - Strengths and weaknesses
  - Interesting open research questions, possible extensions, and applications
- Slides should be clear and **mostly visual** (figures, animations, videos)
- Look at background papers as necessary

Paper presentations

- Search for relevant material, e.g., from the authors' webpage, project pages, etc.
- **Each slide that is not your own must be clearly cited**
- Meet with TA **one day before your presentation** with prepared slides; your responsibility to email TA to schedule a time
- Upload slides to Canvas on day of presentation (after class is fine)
- Skip paper review the day you are presenting

Final project

- One of:
  - Design and evaluation of a novel approach
  - An extension of an approach studied in class
  - In-depth analysis of an existing technique
  - In-depth comparison of two related existing techniques

- Work in groups of 4 or 5
- Talk to me/TA if you need ideas

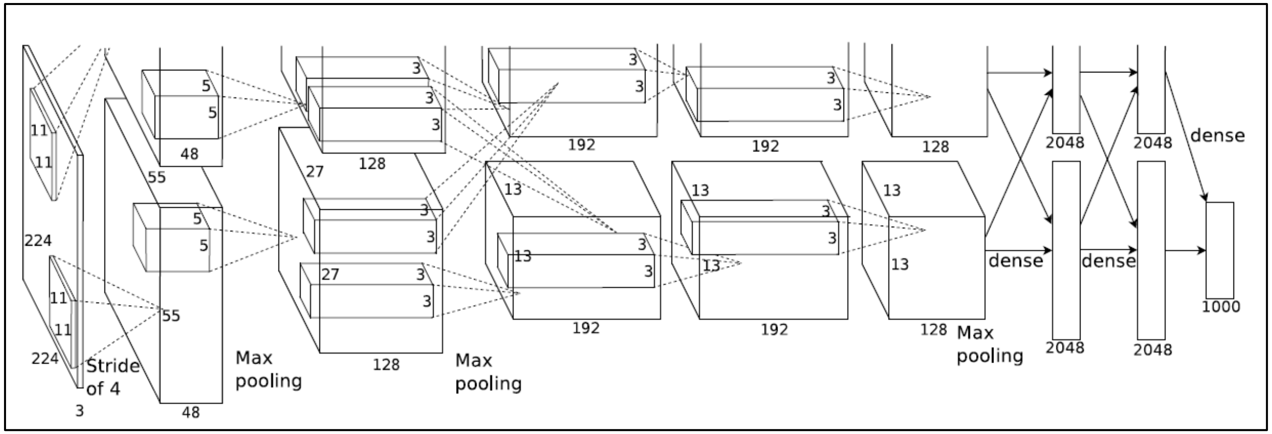
- Google Cloud Credits (will be explained in a future class)

Final project

- Final Project Proposal (5%) due **10/23**
  - 1 page
- Final Project Presentation (10%) due **12/2, 12/4, and 12/6**
  - 10 minutes
- Final Project Paper (15%) due **12/9**
  - 6-8 pages
- Upload pdf/slides to Canvas

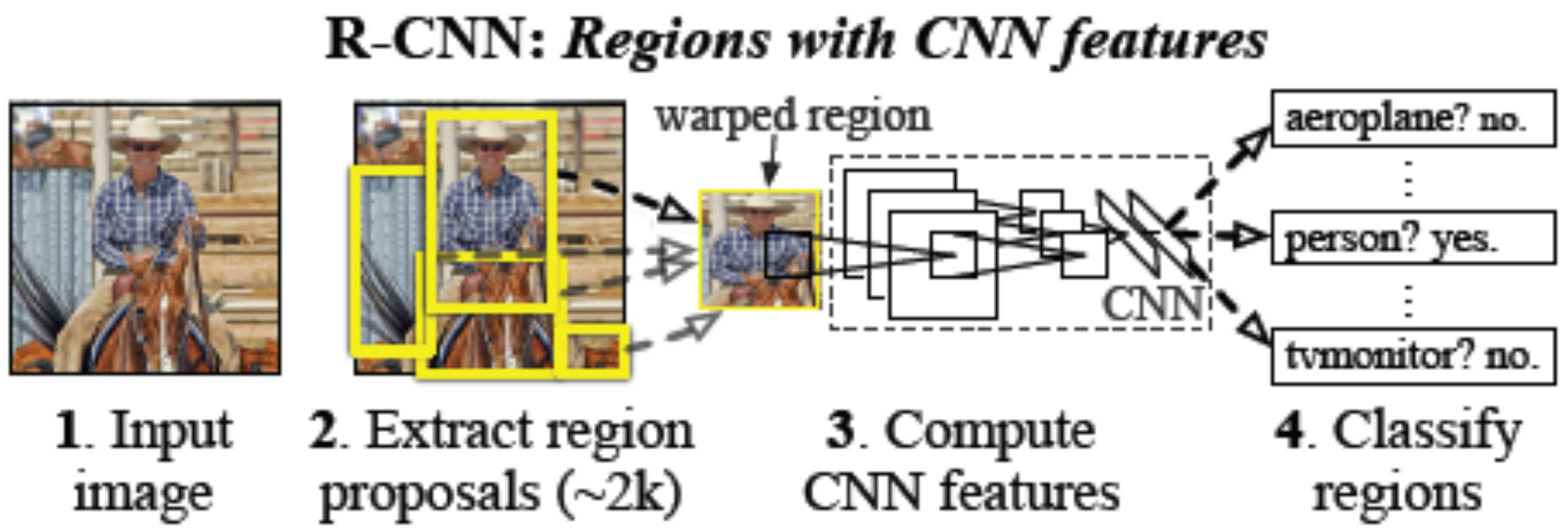
Introductions

Image classification



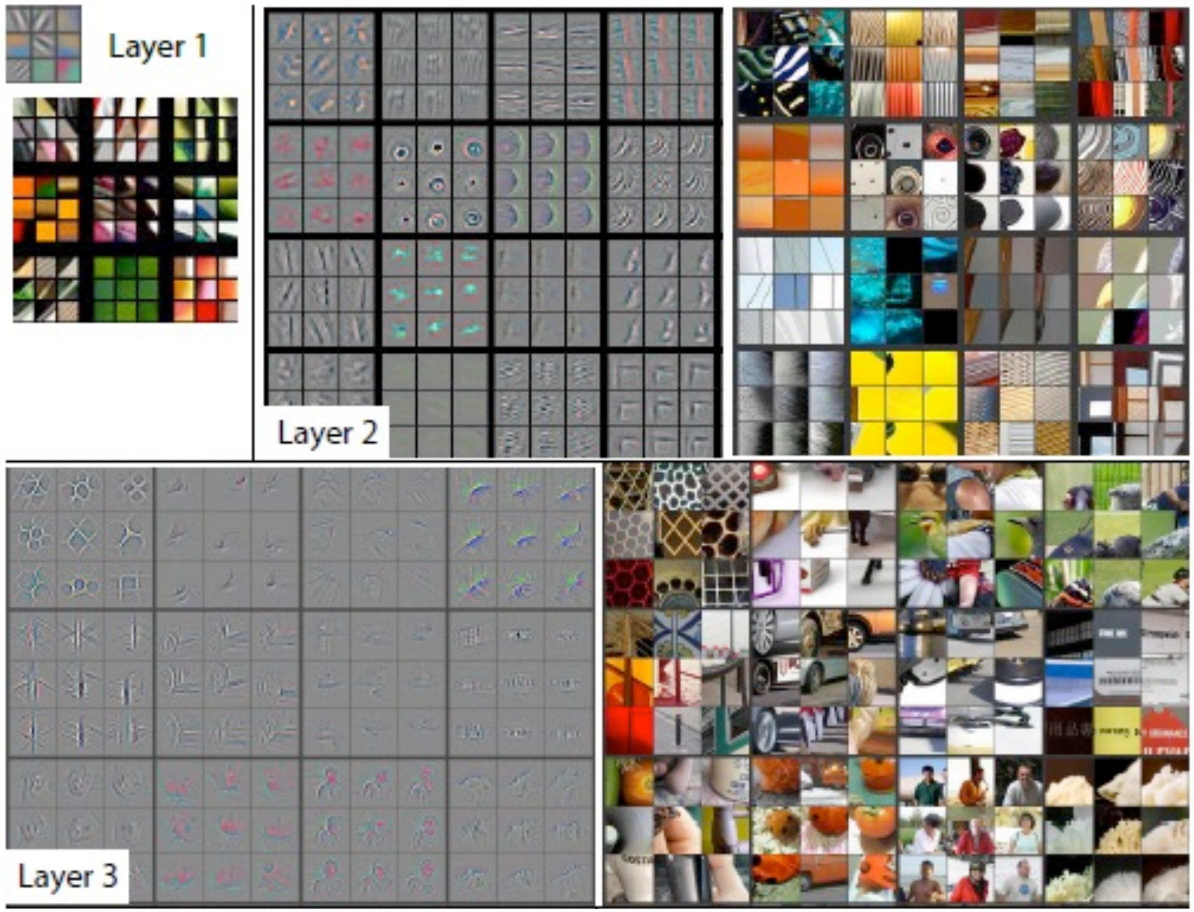
- Convolutional neural networks AlexNet, ResNet

Object detection



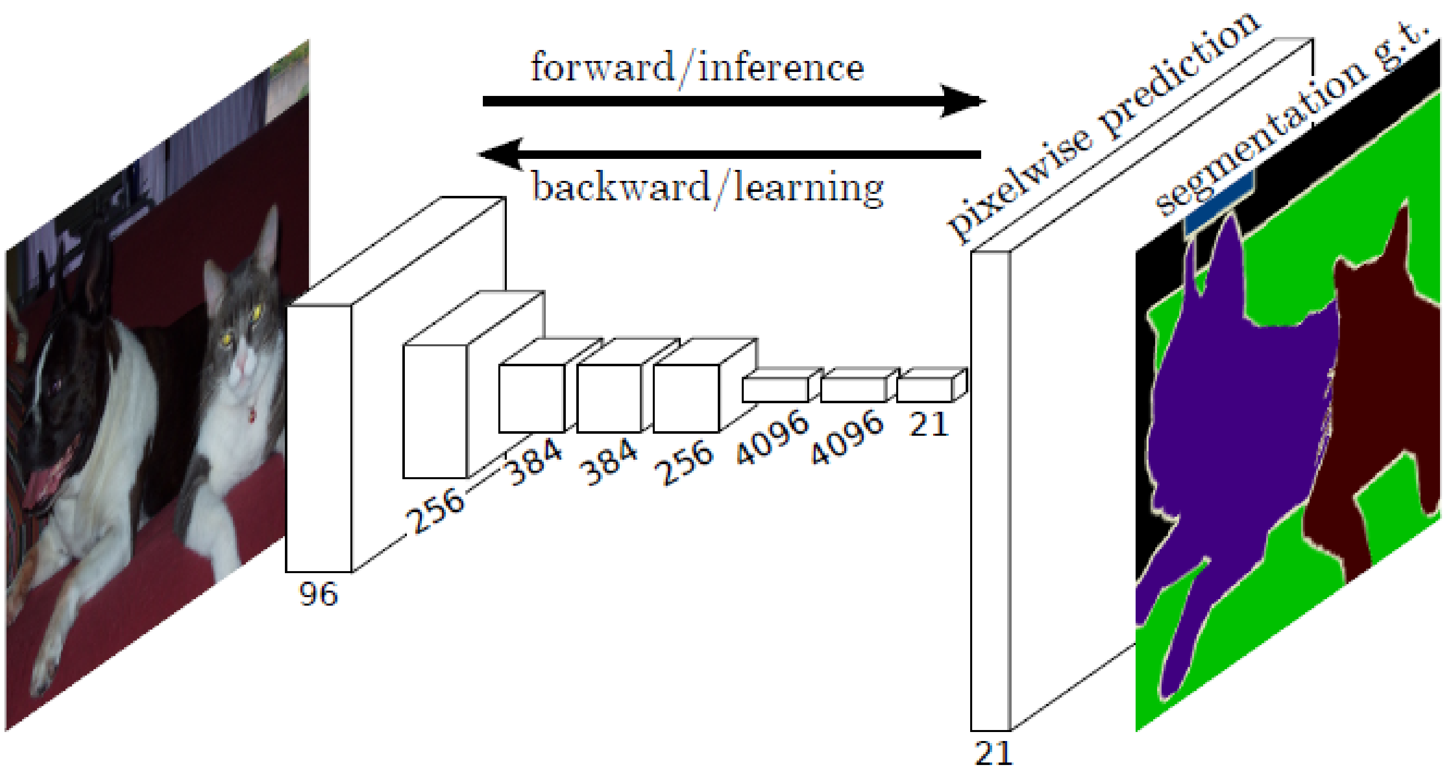
- Two-stage and single-stage frameworks

CNN Visualization and Analysis

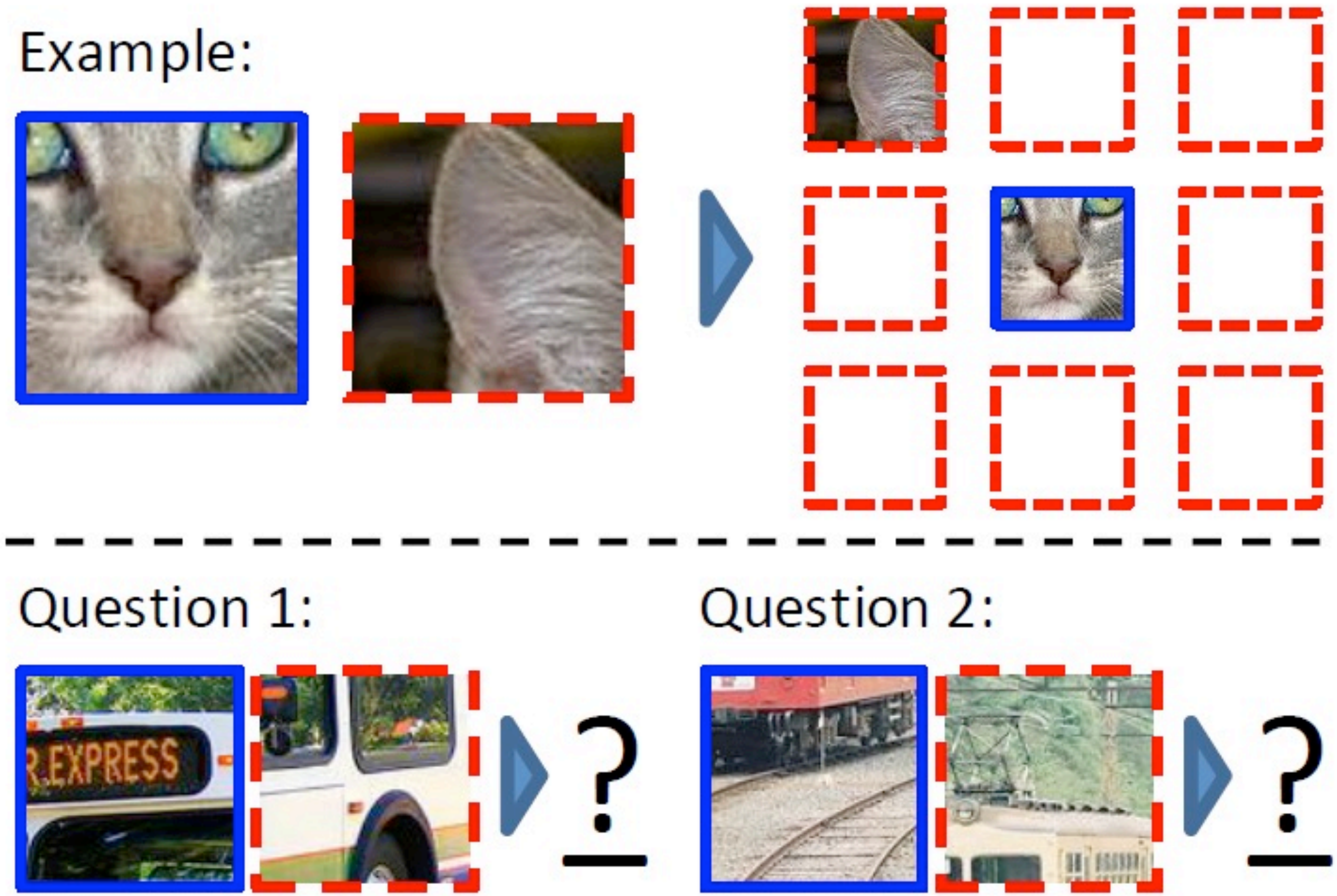


- What is encoded by a CNN?

Semantic Segmentation



Self-supervision



Neural Network Art

Coming up

- Carefully read class webpage
- Sign-up for papers you want to present (spreadsheet will be emailed shortly)
- Next class: CNN basics