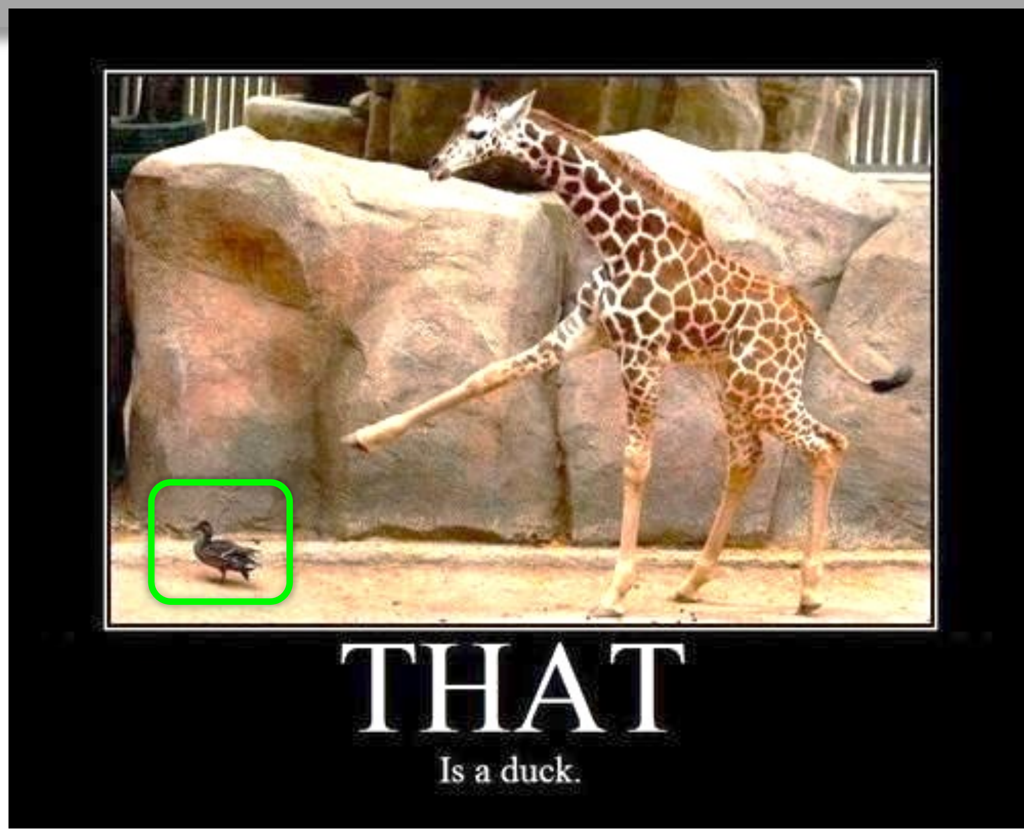
CS5670: Computer Vision

Introduction to Recognition



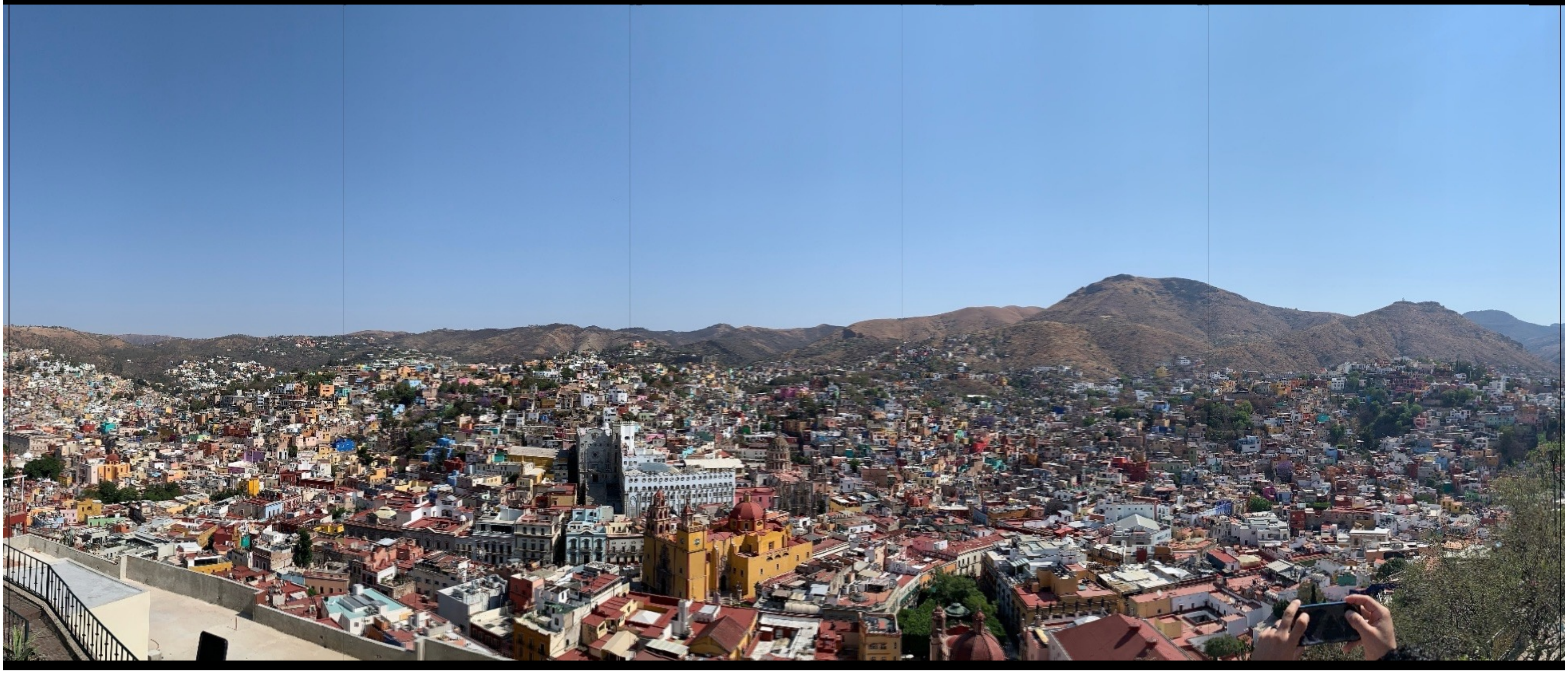
Announcements

- One more project to go – Project 5: Neural Radiance Fields
  - Tentative release date: Thursday, April 18
  - Tentative due date: Wednesday, May 1

- In-class Final Exam during the last lecture: Tuesday, May 7

Third Place

Giacomo Glotzer and Kirby Leo



Second Place

Akhil Raj and Justin Ryan Olson

First Place

Genki Miyasato and Philip Ian Tempelman

Where we go from here

- What we know: Geometry
  - What is the shape of the world?
  - How does that shape appear in images?
  - How can we infer that shape from one or more images?

- What's next: Recognition
  - What are we looking at?
  - **Representations of visual content**

- New representations for 3D geometry
- Generative models

What is “Recognition”?

Next few slides adapted from Li, Fergus, & Torralba’s excellent short course on category and object recognition

What is “Recognition”?

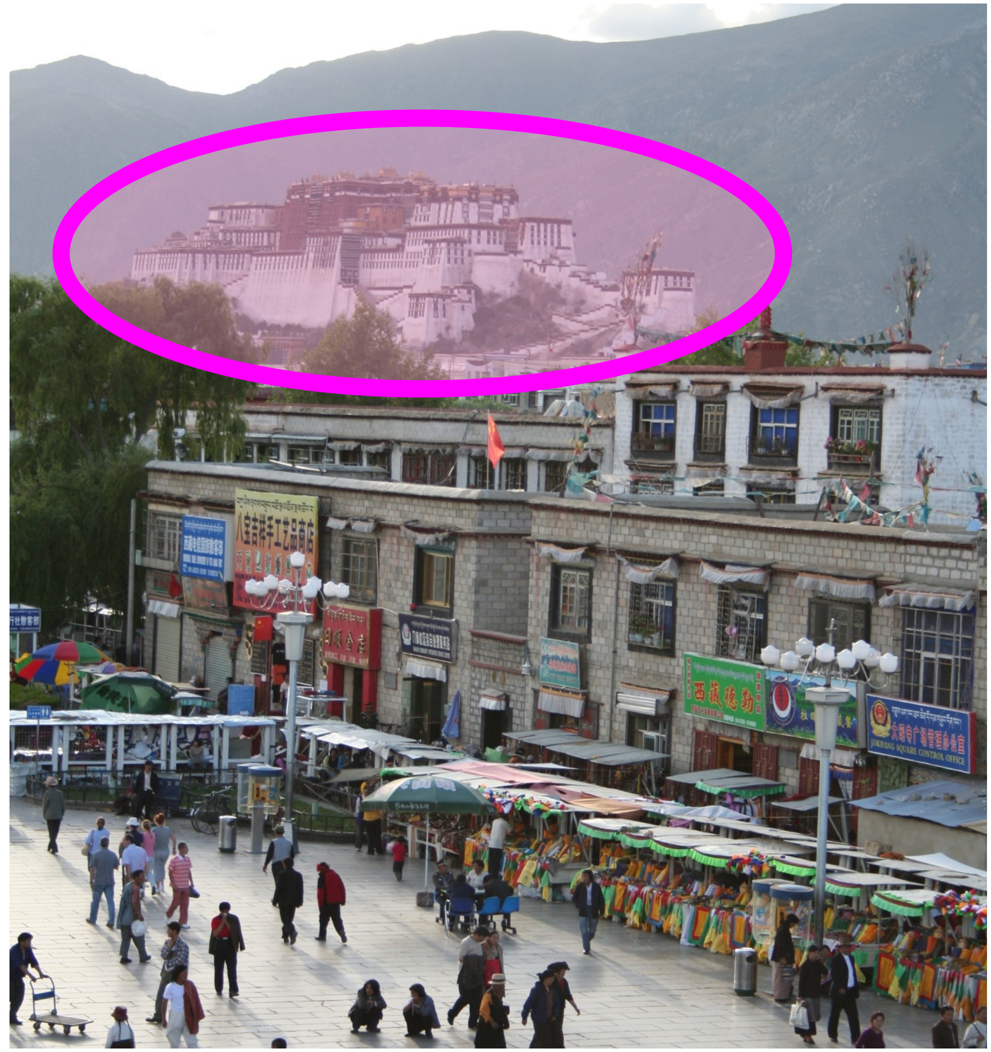
- Verification: is that a lamp?

What is “Recognition”?

- Verification: is that a lamp?
- Detection: where are the people?

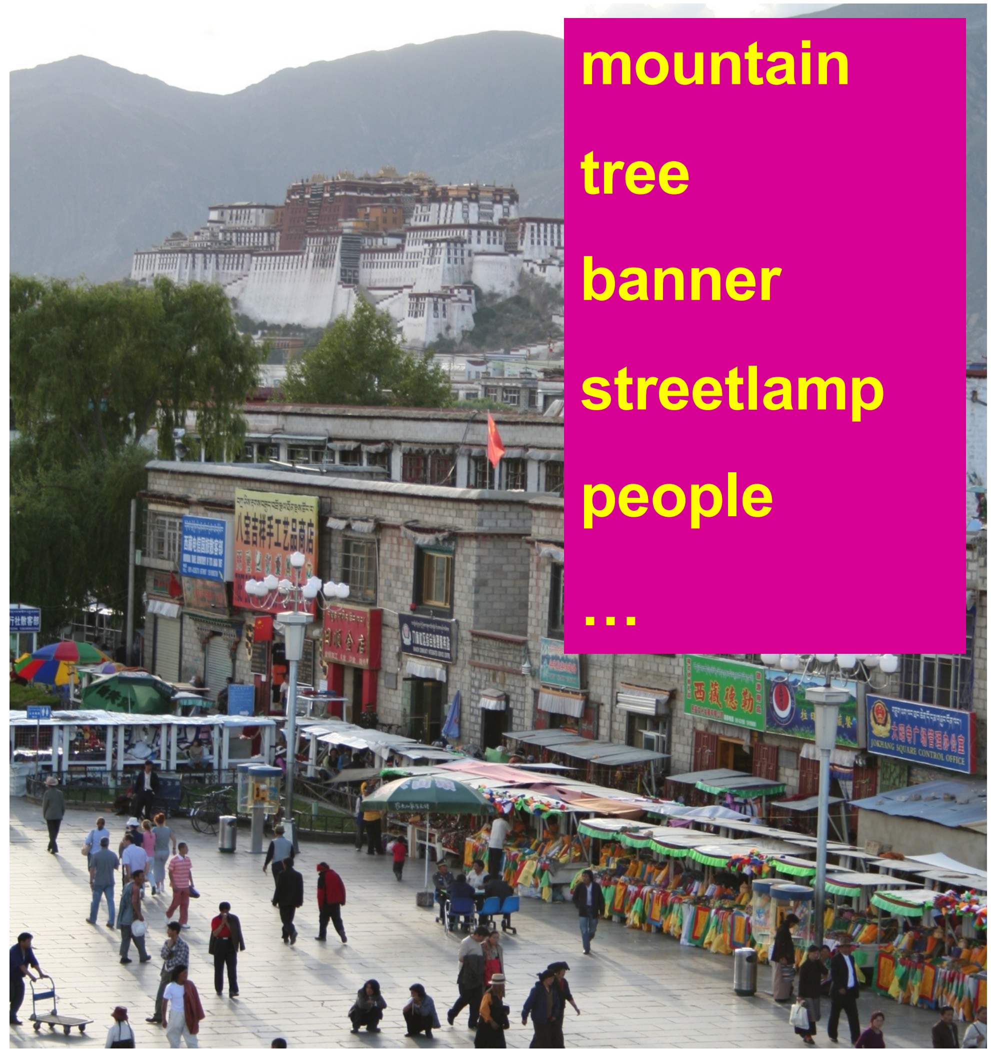
What is “Recognition”?

- Verification: is that a lamp?
- Detection: where are the people?
- Identification: is that Potala Palace?



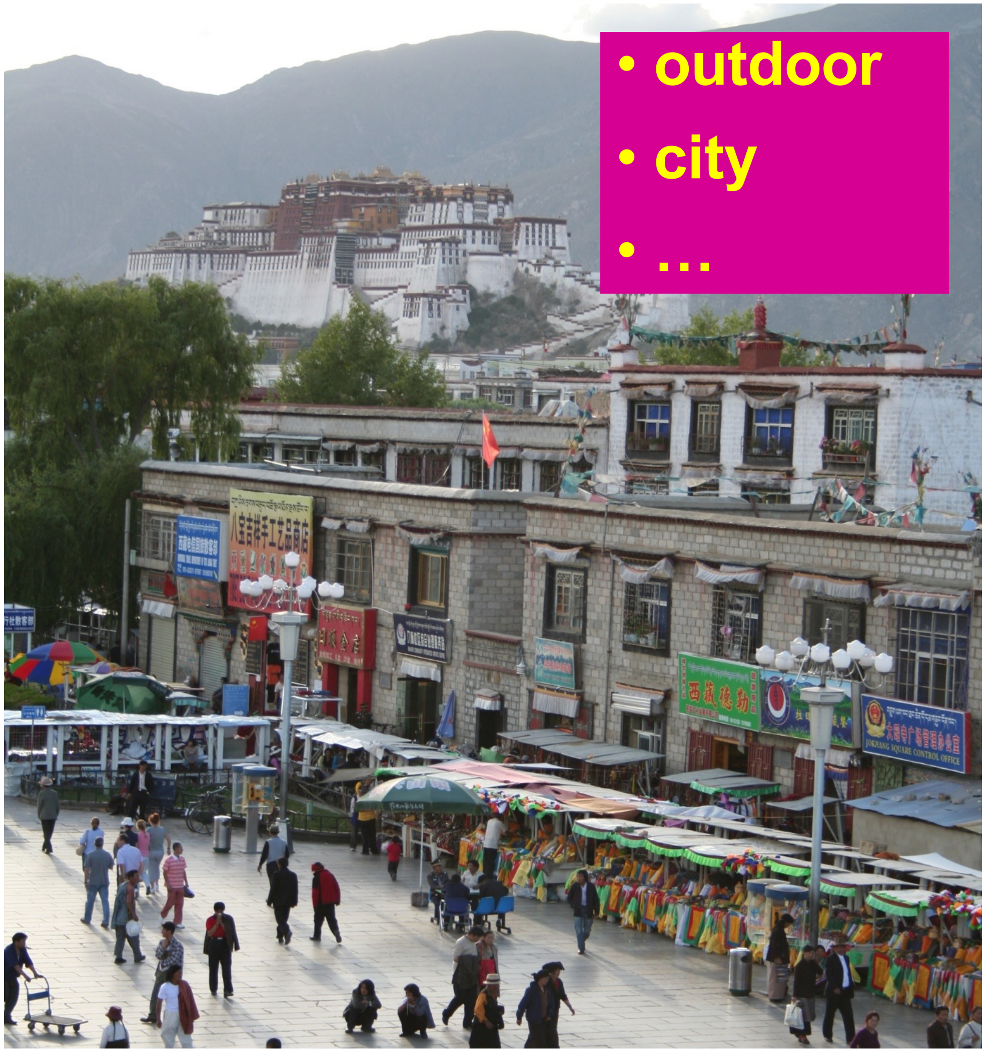
What is "Recognition"?

- Verification: is that a lamp?
- Detection: where are the people?
- Identification: is that Potala Palace?
- Classification: what objects are present?



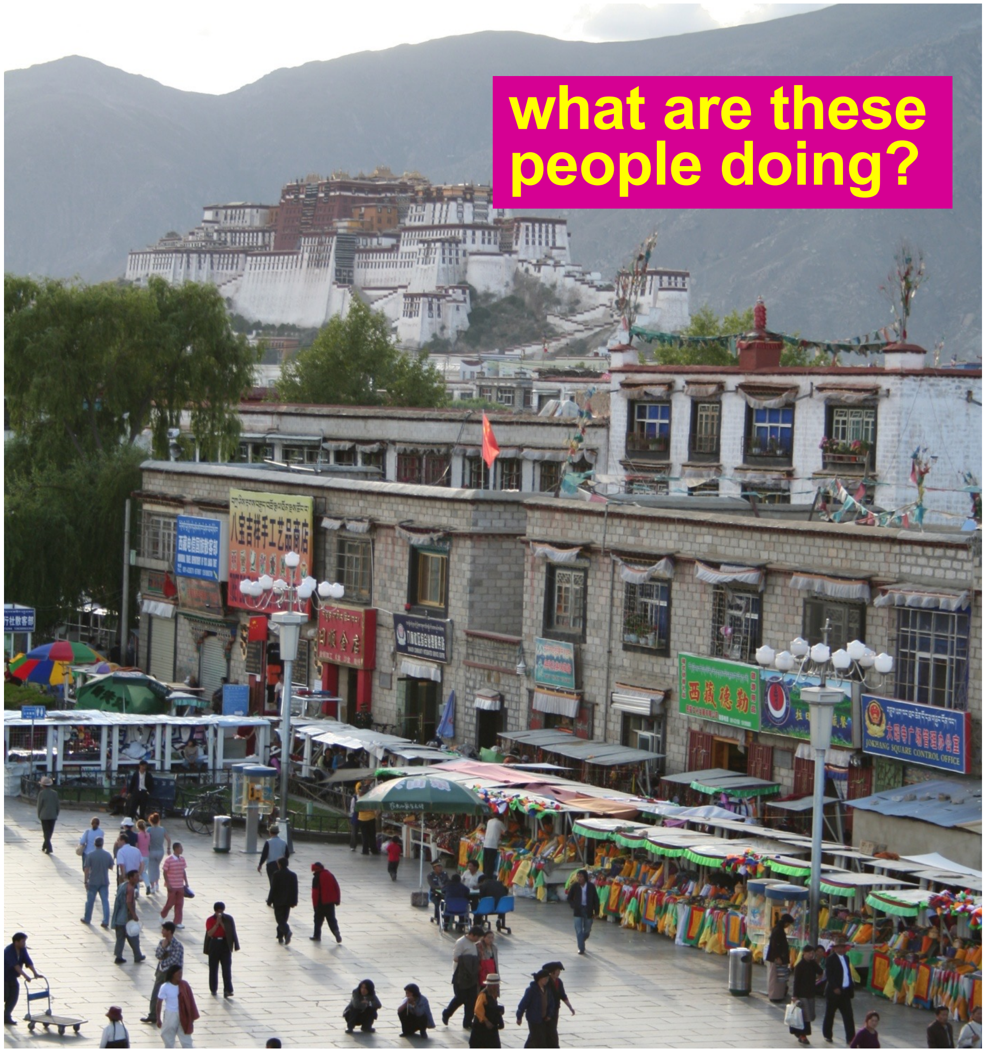
What is “Recognition”?

- Verification: is that a lamp?
- Detection: where are the people?
- Identification: is that Potala Palace?
- Classification: what objects are present?
- Scene and context categorization



What is “Recognition”?

- Verification: is that a lamp?
- Detection: where are the people?
- Identification: is that Potala Palace?
- Classification: what objects are present?
- Scene and context categorization
- Activity / Event Recognition

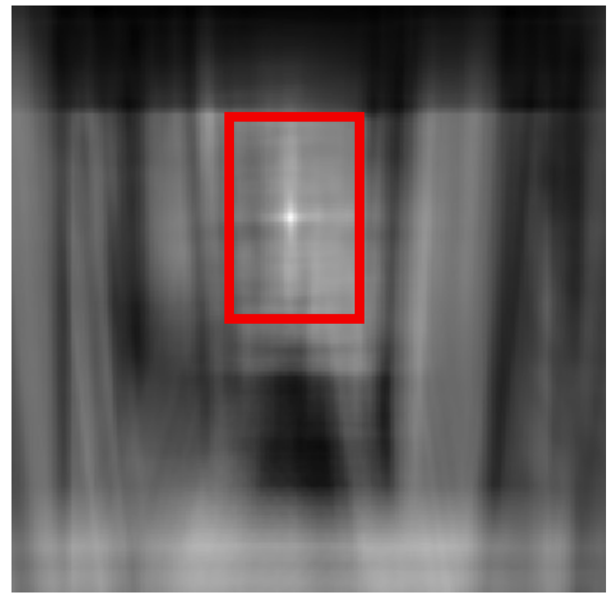


Object recognition: Is it really so hard?

This is a chair

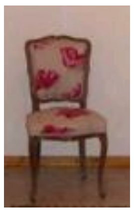
Find the chair in this image

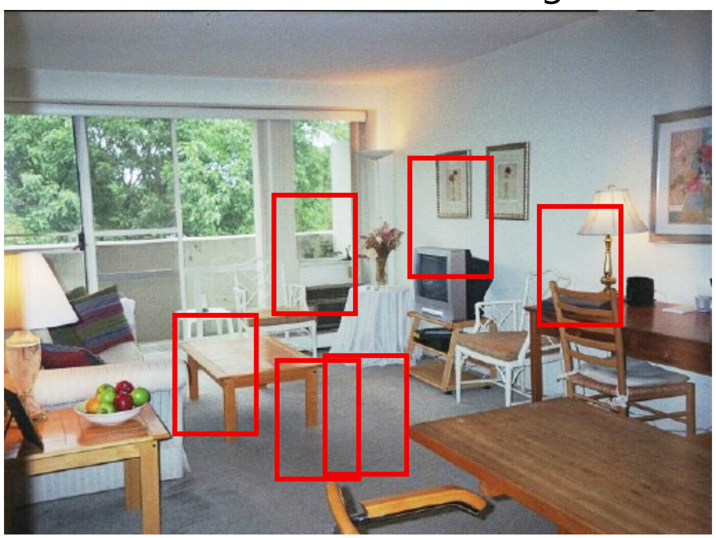
Output of normalized correlation

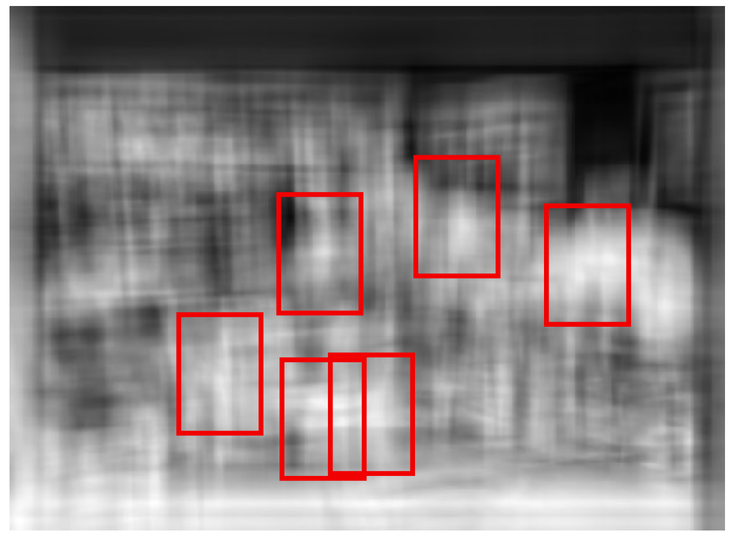


Object recognition: Is it really so hard?

Find the chair in this image



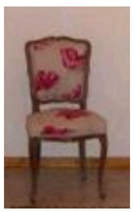




Pretty much garbage:
Simple template matching is not
going to do the trick

Object recognition: Is it really so hard?

Find the chair in this image



A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.

Why not use SIFT matching for everything?

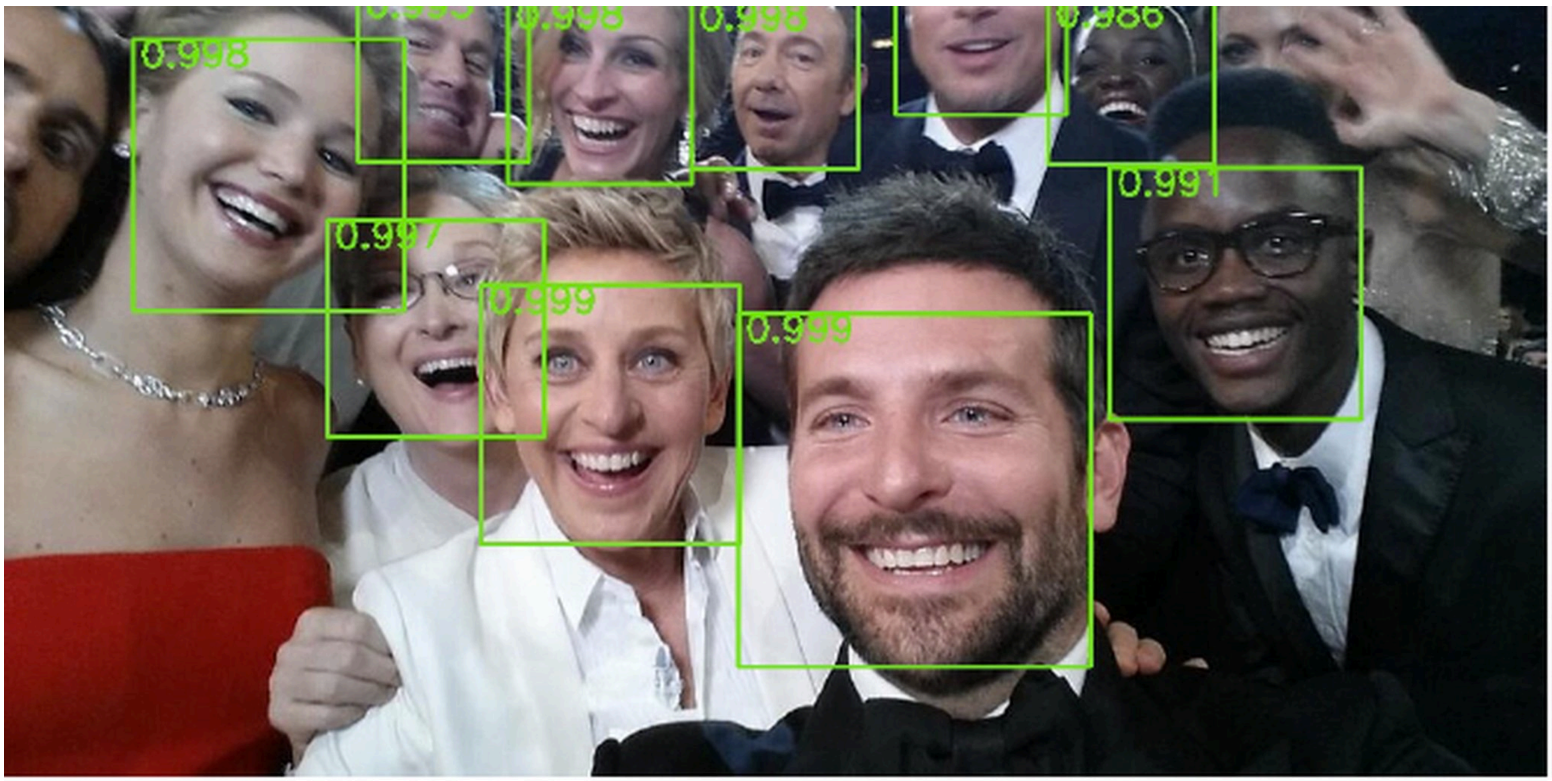
- Works well for object *instances* (or distinctive images such as logos)

- Not great for generic object *categories*

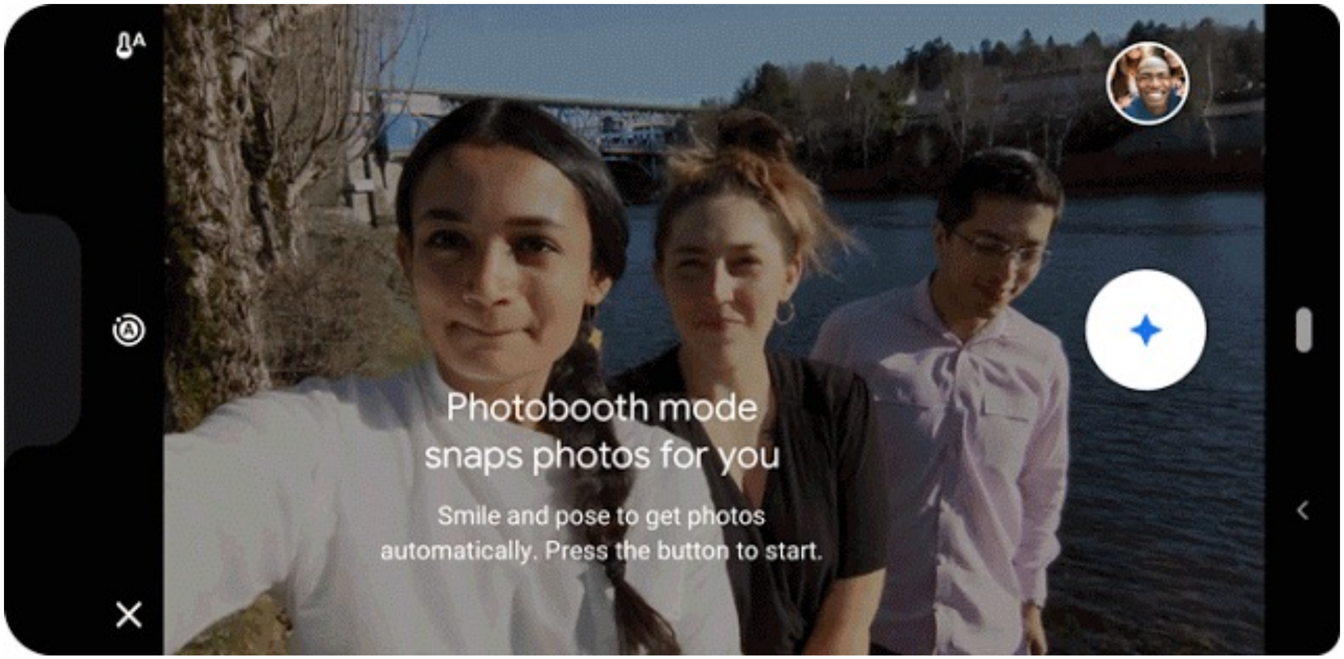
And it can get a lot harder

Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. J Vis, 3(6), 413-422

Applications: Photography



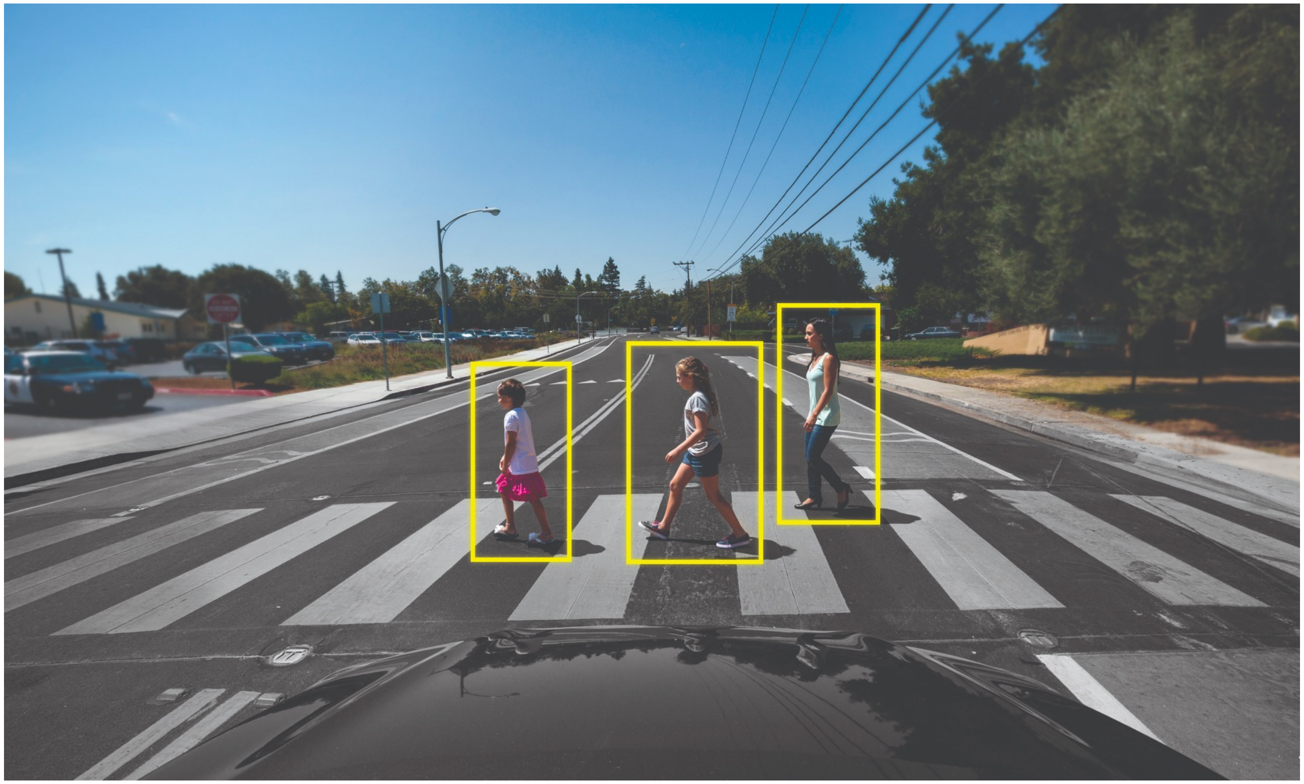
Applications: Shutter-free Photography



Take Your Best Selfie Automatically, with Photobooth on Pixel 3

https://ai.googleblog.com/2019/04/take-your-best-selfie-automatically.html
(Also features "kiss detection")

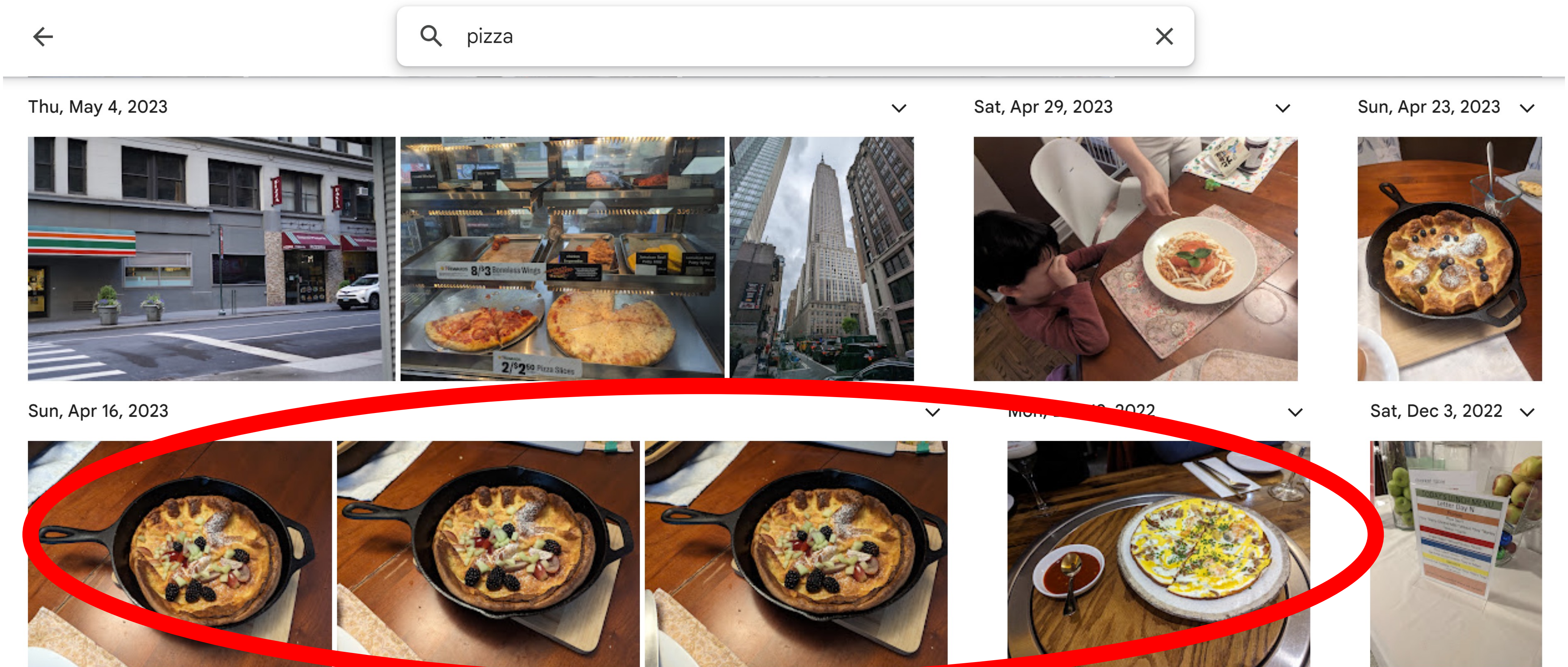
Applications: Assisted / autonomous driving



Applications: Robotics

https://arc.cs.princeton.edu/

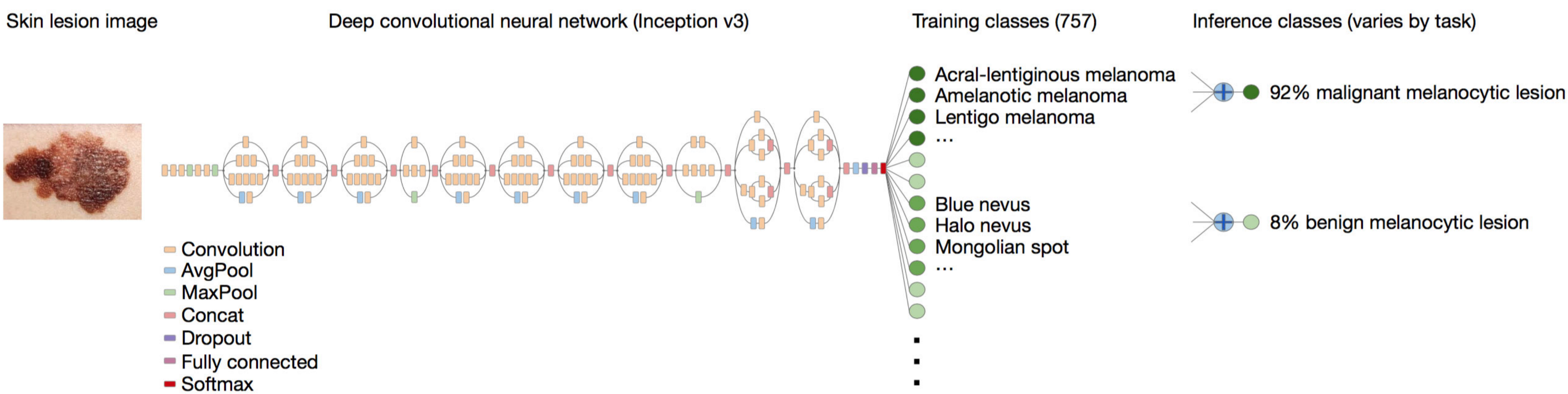
Applications: Photo organization



Source: Google Photos

Not Pizzas!

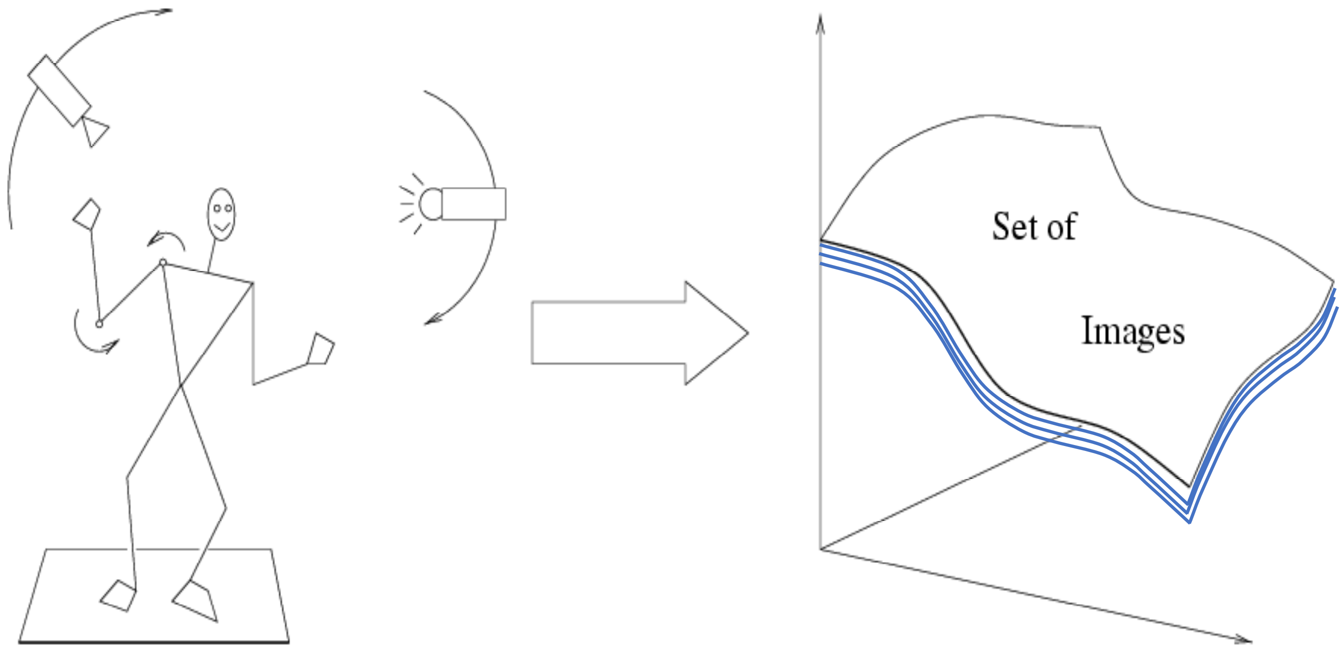
Applications: medical imaging



Dermatologist-level classification of skin cancer

https://cs.stanford.edu/people/esteva/nature/

Why is recognition hard?



Variability: Camera position,
Illumination,
Shape,
etc…

Svetlana Lazebnik

Challenge: lots of potential classes

Challenge: variable viewpoint

Michelangelo 1475-1564

Challenge: variable illumination

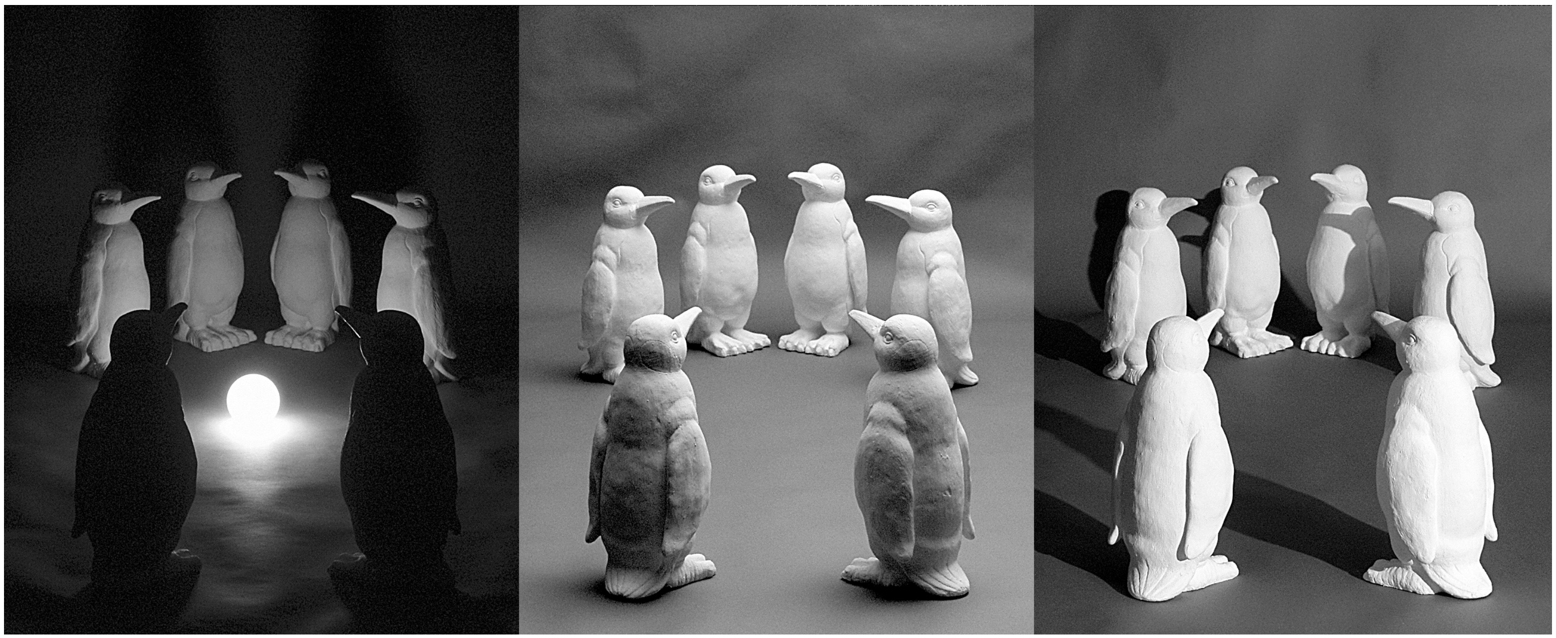
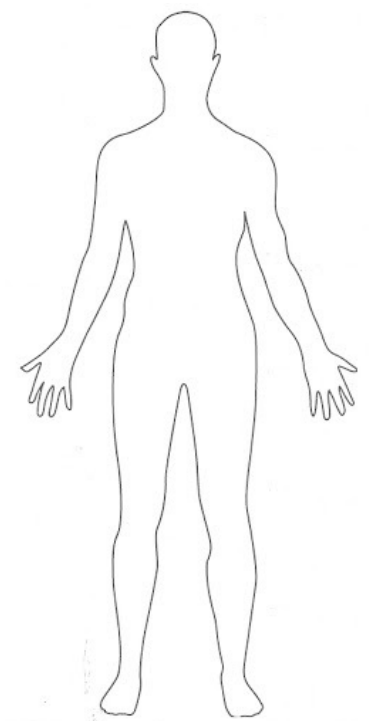


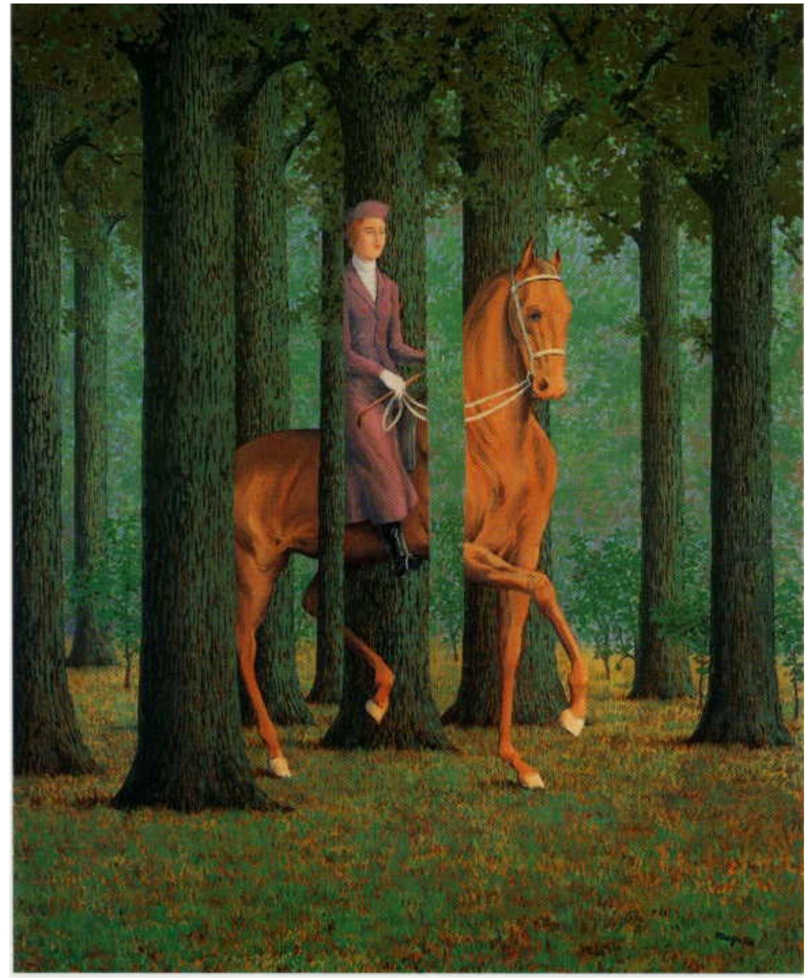
image credit: J. Koenderink

Challenge: scale

Challenge: deformation

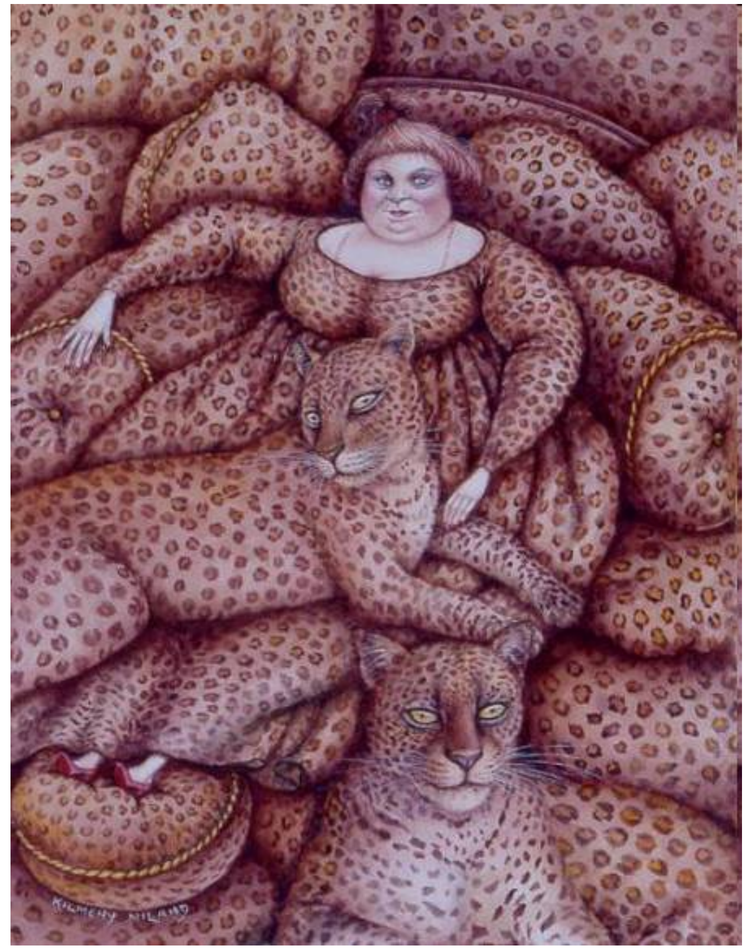


Challenge: Occlusion



Magritte, 1957

Challenge: background clutter



Kilmeny Niland.
1995

Challenge: intra-class variations

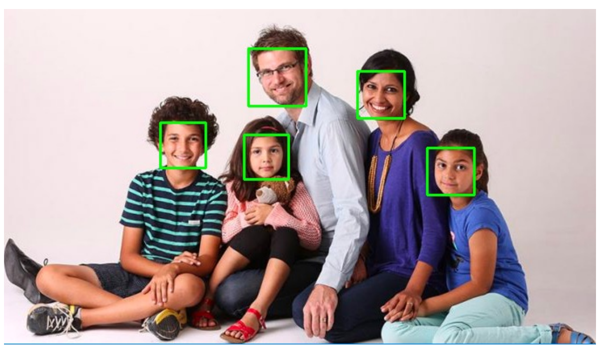


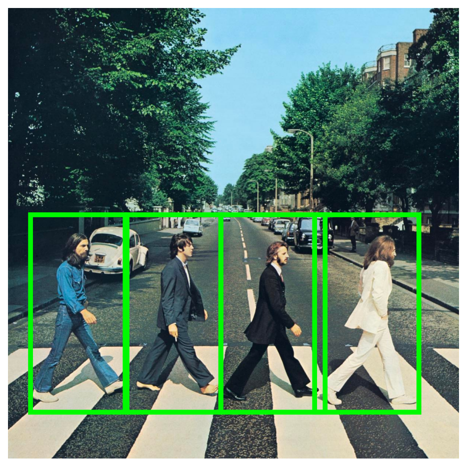


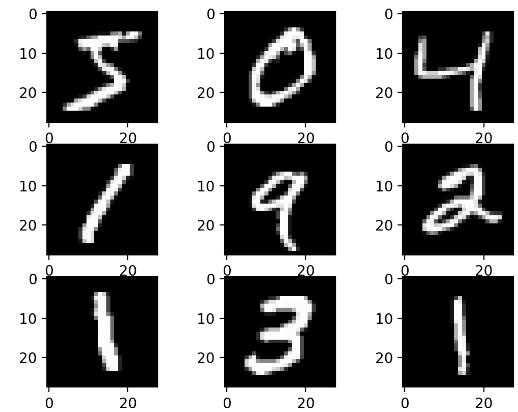

Svetlana Lazebnik

A brief history of image recognition

- What worked in 2011 (pre-deep-learning era in computer vision)
  - Optical character recognition
  - Face detection
  - Instance-level recognition (what logo is this?)
  - Pedestrian detection (sort of)
  - … that's about it







A brief history of image recognition

- What works now, post-2012 (deep learning era and beyond)
  - Robust object classification across thousands of object categories (rivalling human capabilities)

“Spotted salamander”

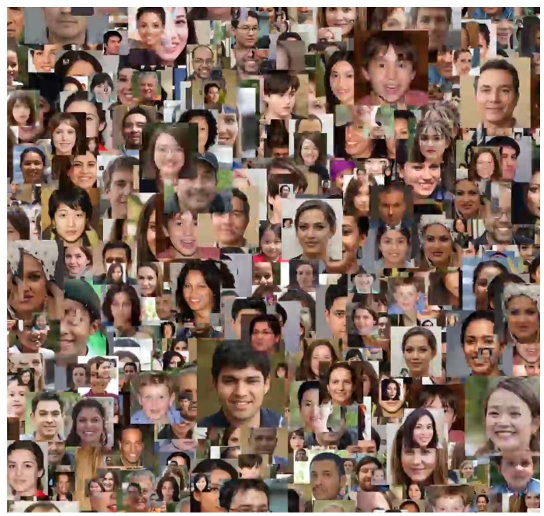
A brief history of image recognition

- What works now, post-2012 (deep learning era and beyond)
  - Face recognition at scale

The New York Times   Account

The Secretive Company That Might End Privacy as We Know It

A little-known start-up helps law enforcement match photos of unknown people to their online images — and “might lead to a dystopian future or something,” a backer says.



https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

FaceNet: A Unified Embedding for Face Recognition and Clustering

Florian Schroff
fschroff@google.com
Google Inc.

Dmitry Kalenichenko
dkalenichenko@google.com
Google Inc.

James Philbin
jphilbin@google.com
Google Inc.

FaceNet, CVPR 2015

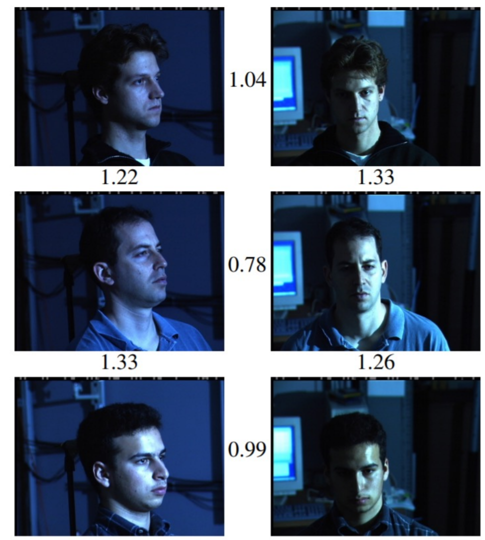


Figure 1. **Illumination and Pose invariance.** Pose and illumination have been a long standing problem in face recognition. This figure shows the output distances of FaceNet between pairs of faces of the same and a different person in different pose and illumination combinations. A distance of 0.0 means the faces are identical, 4.0 corresponds to the opposite spectrum, two different identities. You can see that a threshold of 1.1 would classify every pair correctly.

A brief history of image recognition

- What works now, post-2012 (deep learning era and beyond)
  - High-quality image/video synthesis

A Style-Based Generator Architecture for Generative Adversarial Networks
Tero Karras (NVIDIA), Samuli Laine (NVIDIA), Timo Aila (NVIDIA)
http://stylegan.xyz/paper

These people are not real – they were produced by our generator that allows control over different aspects of the image.

A brief history of image recognition

- What works now, post-2012 (deep learning era and beyond)
  - High-quality image/video synthesis

DALL-E 3

An illustration of an avocado sitting in a therapist's chair, saying 'I just feel so empty inside' with a pit-sized hole in its center. The therapist, a spoon, scribbles notes.

Several giant wooly mammoths approach treading through a snowy meadow, their long wooly fur lightly blows in the wind as they walk, snow covered trees and dramatic snow capped mountains in the distance…

Societal impacts

- Privacy invasion (e.g., face/person recognition, biometrics)
- Bias in AI methods (e.g., recognition systems that perform worse on certain demographics)
- Bias in training data (e.g., used to learn or perpetuate biased associations)
- Sources of training data (copyright issues, consent issues, etc.)
- Generative media (e.g., deepfakes, disinformation)
- …

What Matters in Recognition?

- Learning Techniques
  - E.g. choice of classifier or inference method
- Representation
  - Low level: SIFT, HoG, GIST, edges
  - Mid level: Bag of words, sliding window, deformable model
  - Deep learned features
  - Latent diffusion models
- Data
  - More is always better (as long as it is good data)
  - Annotation (labeling data) has historically been a key challenge
  - Now we are seeing powerful models trained from more noisy labels

What Matters in Recognition?

- Learning Techniques
  - E.g. choice of classifier or inference method
- Representation
  - Low level: SIFT, HoG, GIST, edges
  - Mid level: Bag of words, sliding window, deformable model
  - **Deep learned features**
  - **Latent diffusion models**
- Data
  - **More is always better (as long as it is good data)**
  - Annotation (labeling data) has historically been a key challenge
  - Now we are seeing powerful models trained from more noisy labels

24 Hrs in Photos

Flickr Photos From 1 Day in 2011

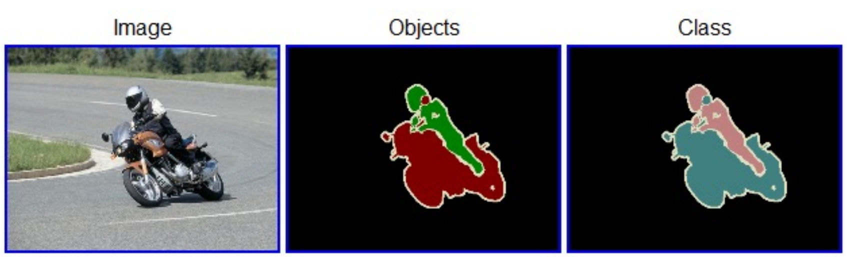
https://www.kesselskramer.com/project/24-hrs-in-photos/

Datasets

- PASCAL VOC [2005-2012]
  - *Not* Crowdsourced, bounding boxes, 20 categories
- CIFAR-10 [2009]
  - 60000 32x32 color images in 10 classes (6000 images per class)
- ImageNet [2010 – current]
  - Huge, Crowdsourced, Hierarchical, *Iconic* objects
- COCO (Common Objects in Context) [2014 – current]
  - Crowdsourced, large-scale objects
- LAION 5B [2022 – current]
  - 5.85 billion noisy image-text pairs

The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

- 20 object categories (aeroplane to TV/monitor)
- Three challenges:
  - Classification challenge (is there an X in this image?)
  - Detection challenge (draw a box around every X)
  - Segmentation challenge (which class is each pixel?)



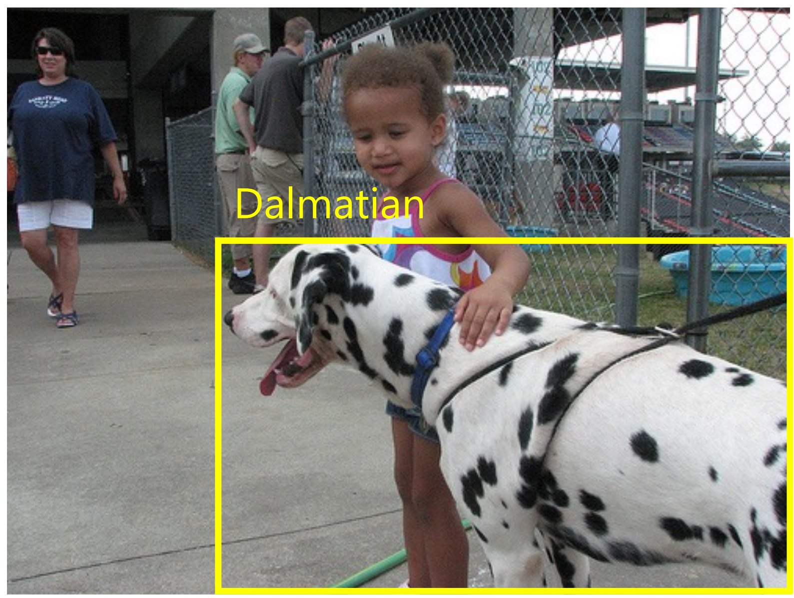
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

2010-2017

IMAGENET

~~20 object classes~~ ~~22,591 images~~

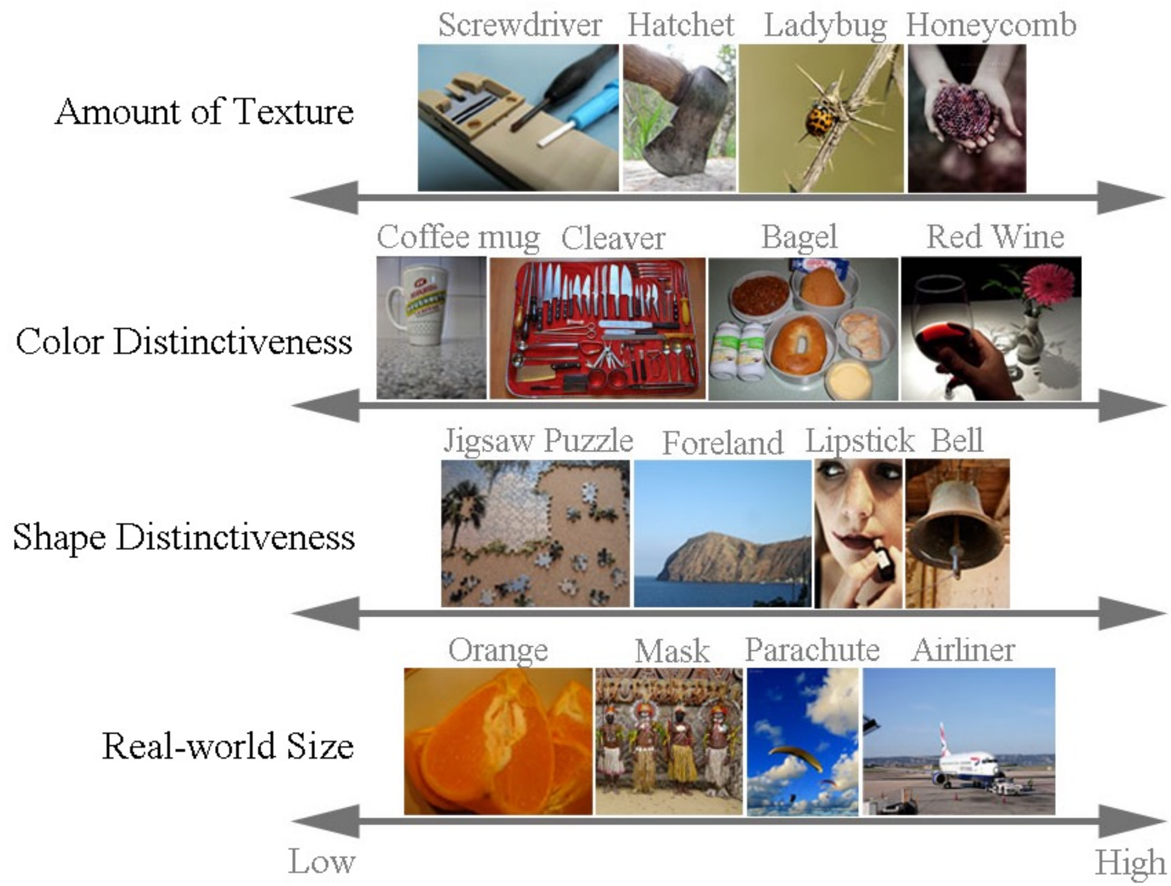
1000 object classes **1,431,167 images**



http://image-net.org/challenges/LSVRC/{2010,2011,2012}

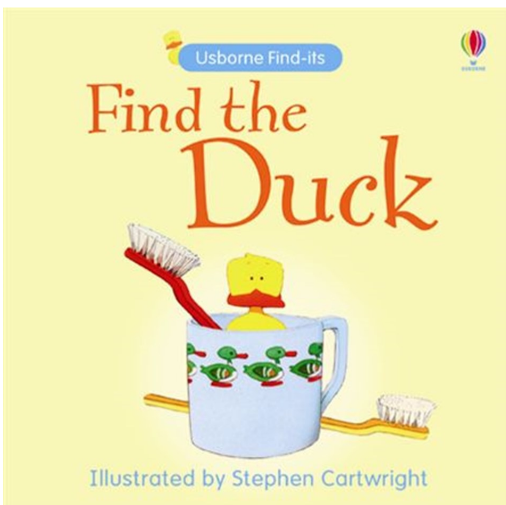
Variety of object classes in ILSVRC

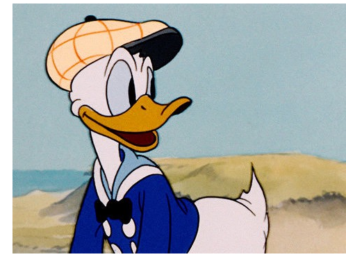
Variety of object classes in ILSVRC

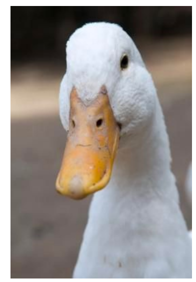


What’s Still Hard?

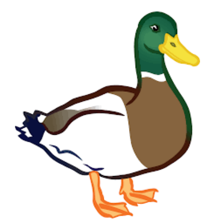
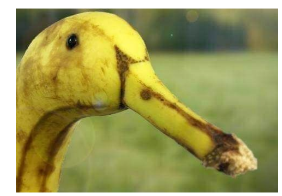
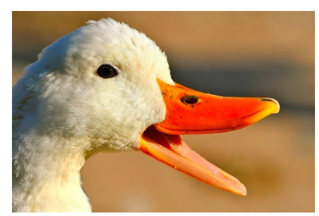
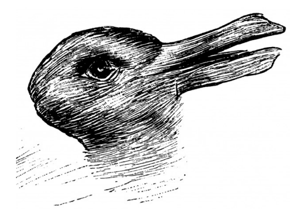
- Few shot learning
  - How do we generalize from only a small number of examples?









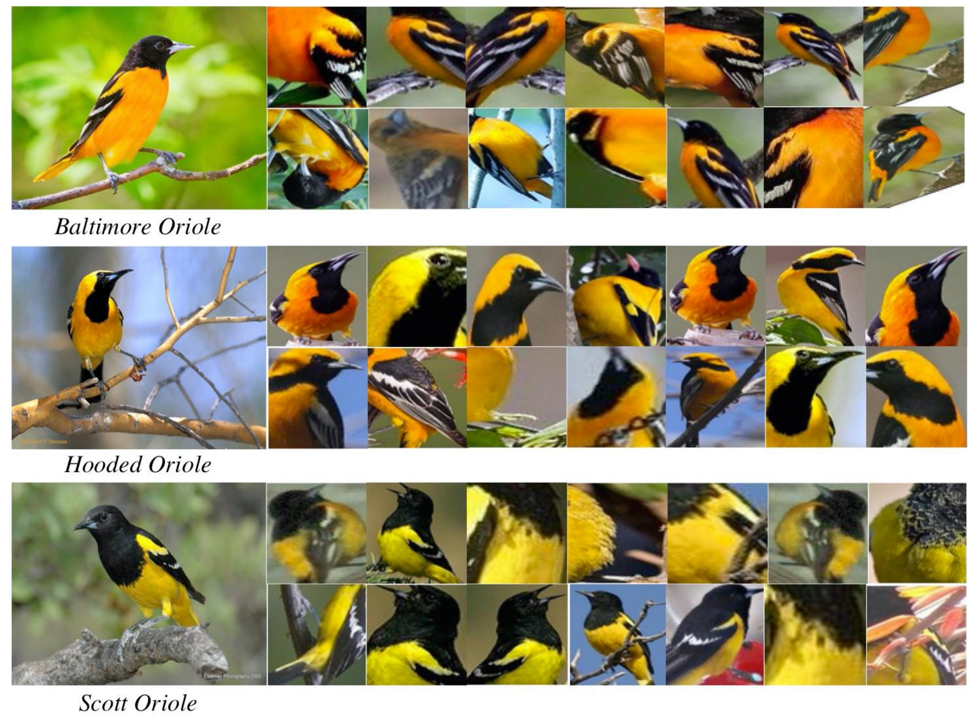





What’s Still Hard?

- Few shot learning
  - How do we generalize from only a small number of examples?

- Fine-grained classification
  - How do we distinguish between more subtle class differences?

Animal->Bird->Oriole…



Questions?

Next

- Image classification pipeline
- Training, validation, testing
- Nearest neighbor classification
- Linear classification

- Building up to Convolutional Neural Networks (CNNs) and beyond