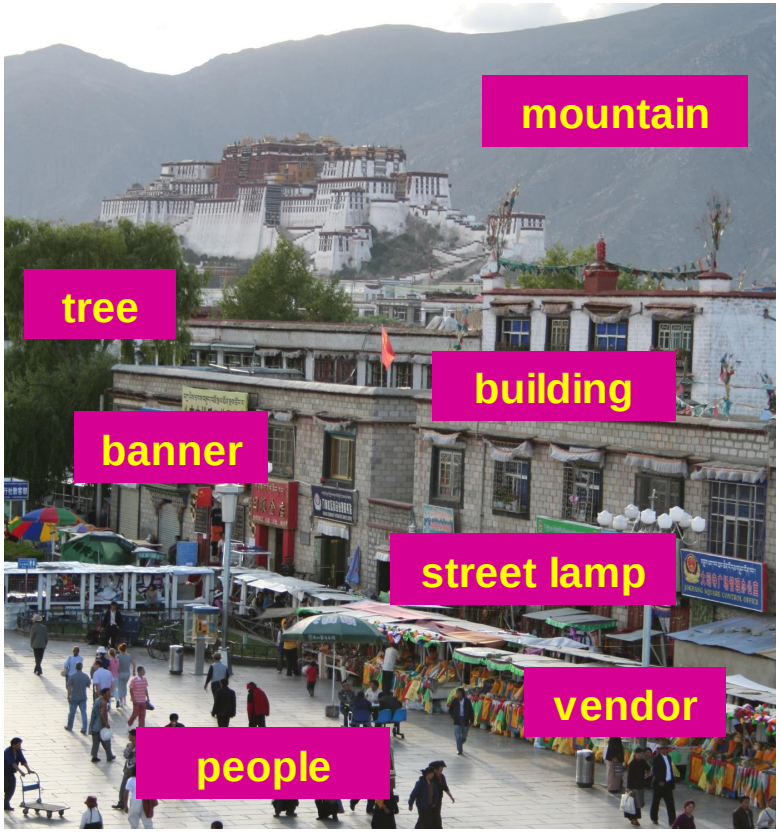
CS5670: Intro to Computer Vision

Noah Snavely

Introduction to Recognition



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?

What do we mean by “object recognition”?

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

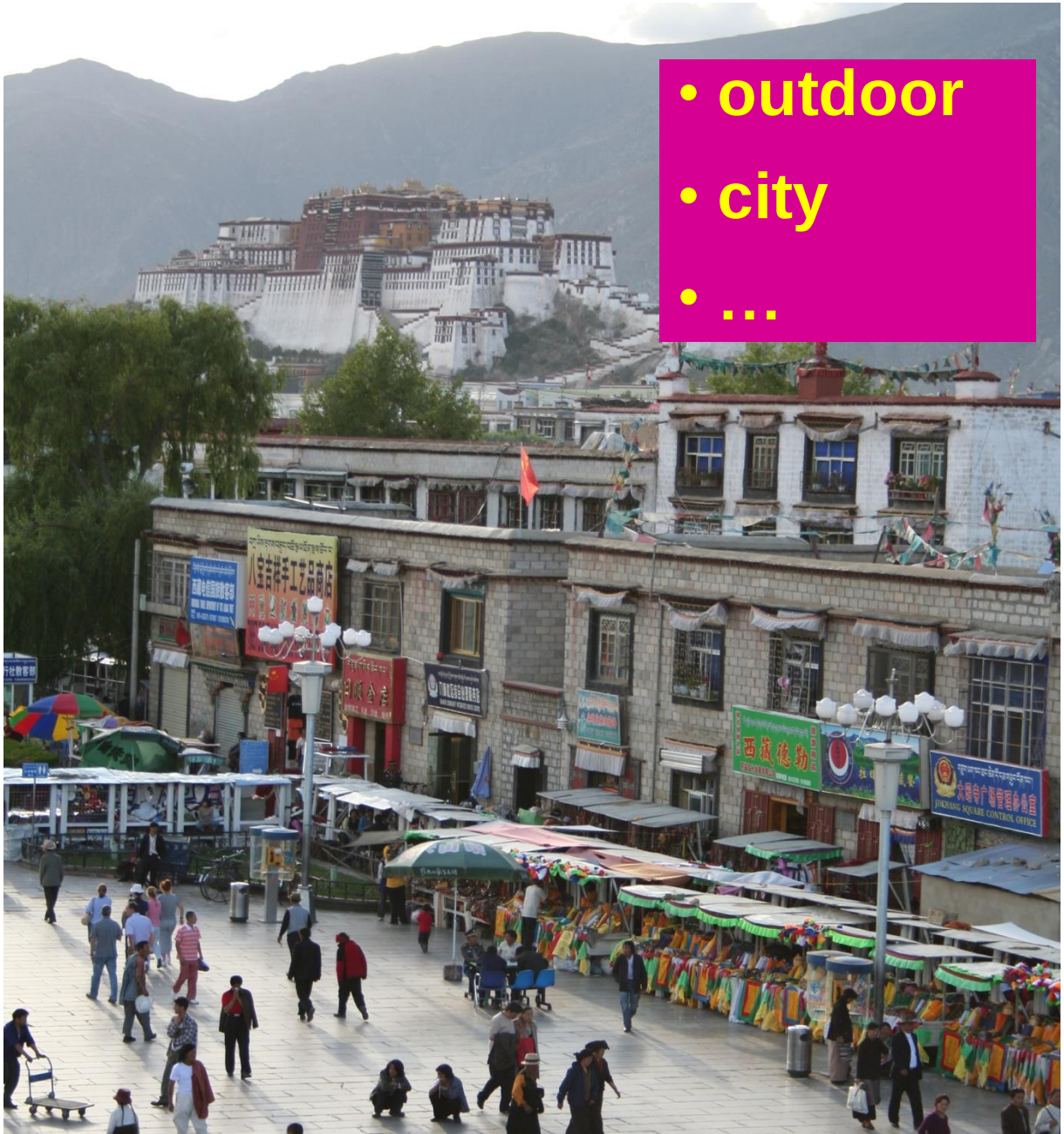
Verification: is that a lamp?

Detection: where are the people?

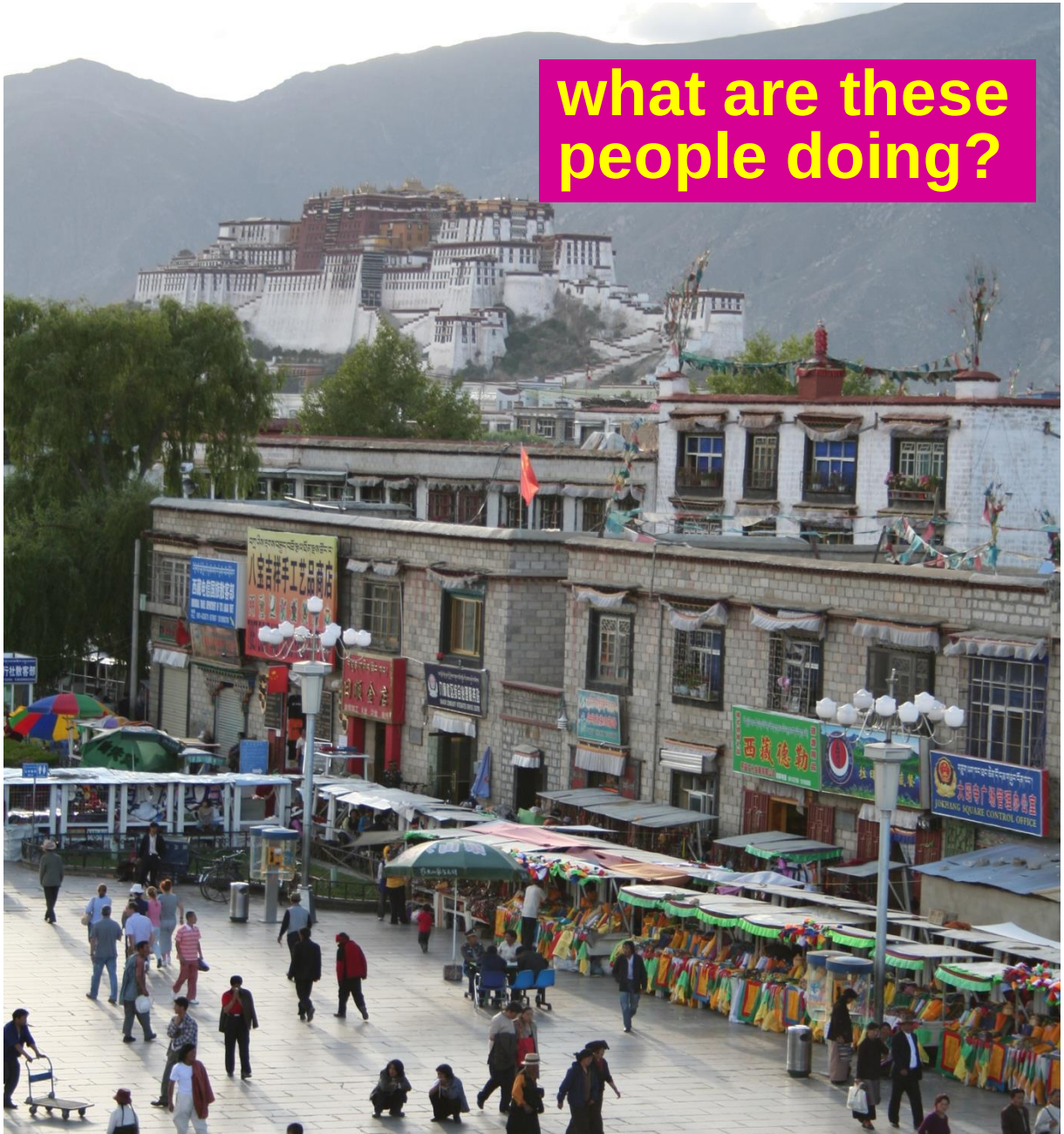
Identification: is that Potala Palace?

Object categorization

Scene and context categorization

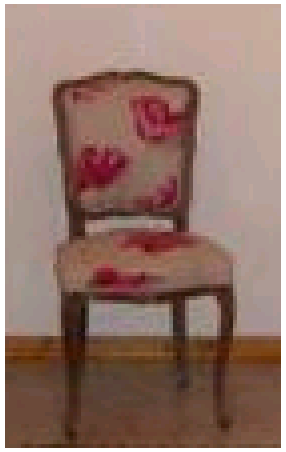


Activity / Event Recognition



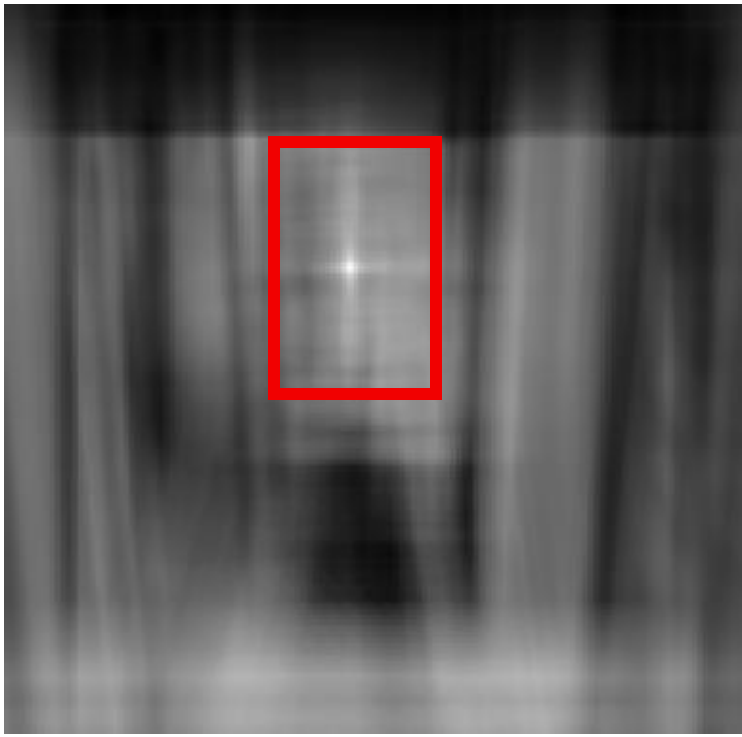
Object recognition
Is it really so hard?

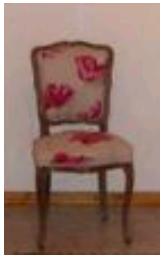
This is a chair



Find the chair in this image

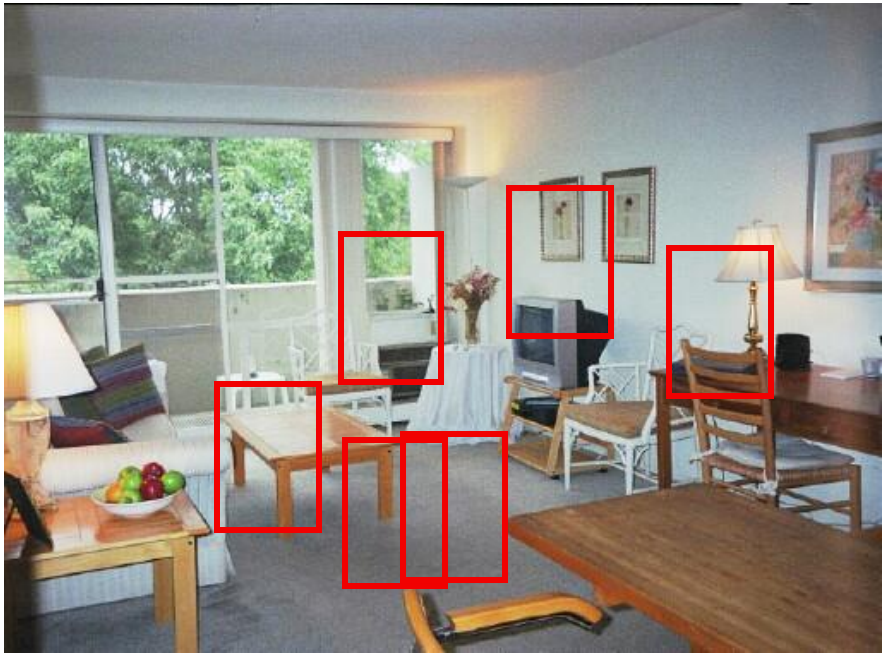
Output of normalized correlation

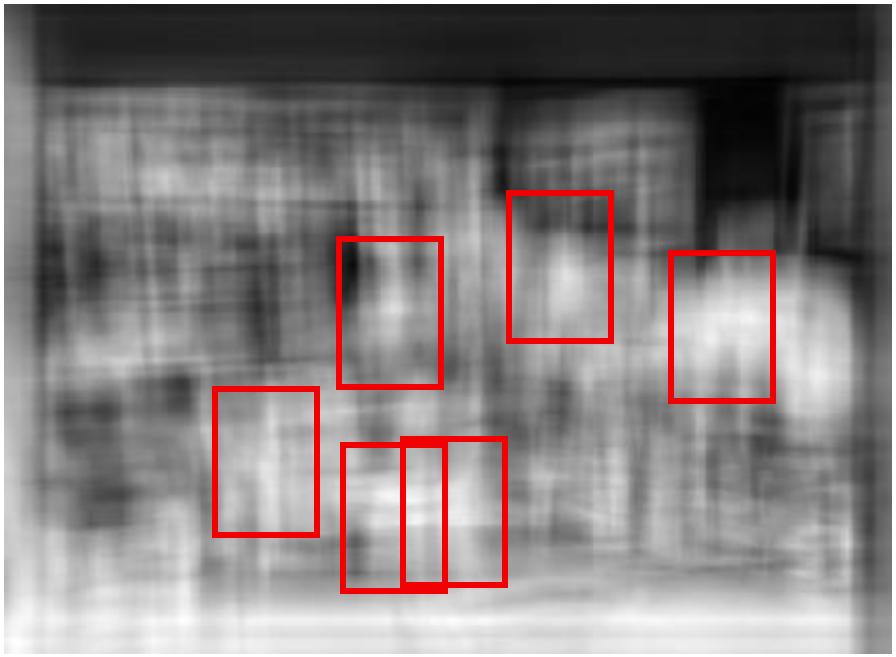




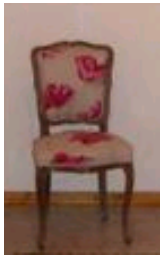
Object recognition
Is it really so hard?

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)

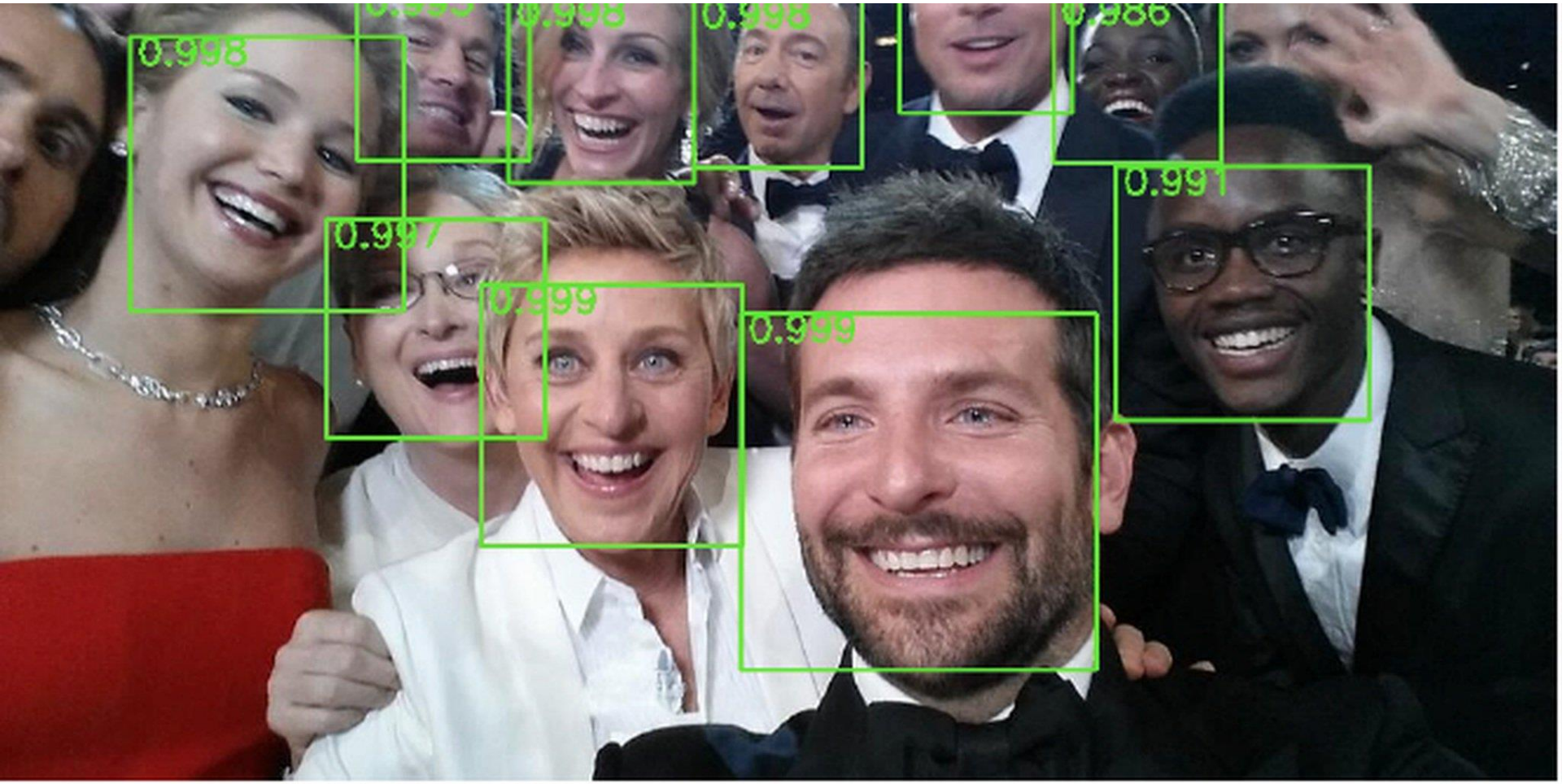
- Not great for generic object *categories*



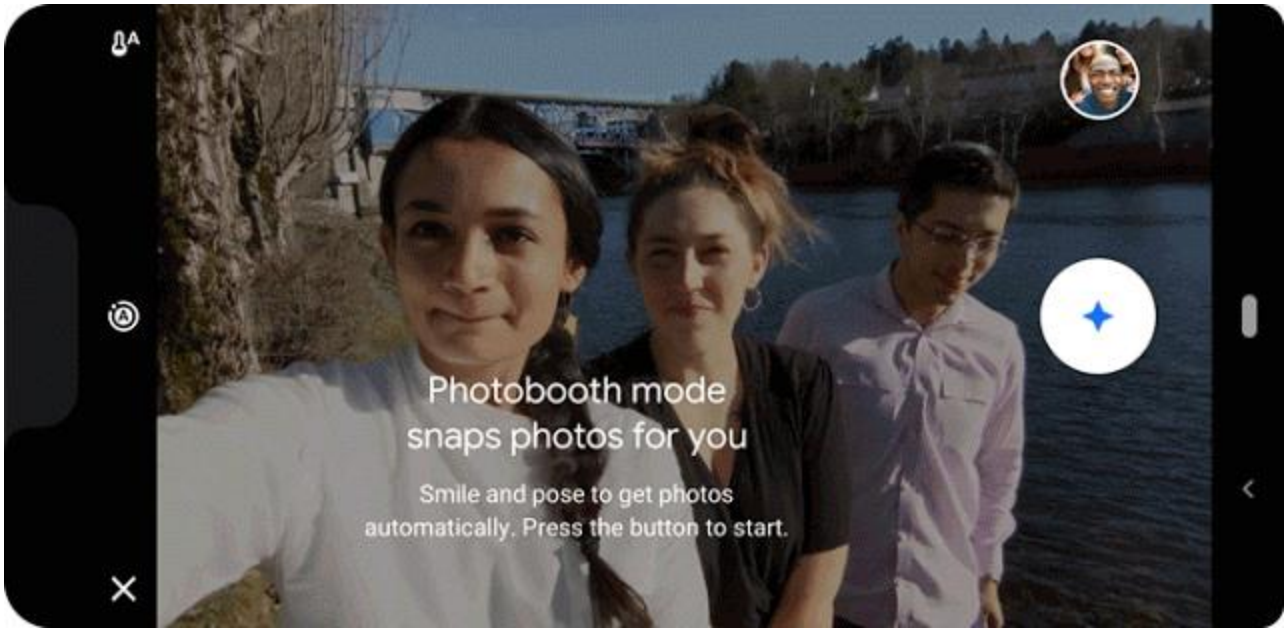
And it can get a lot harder

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

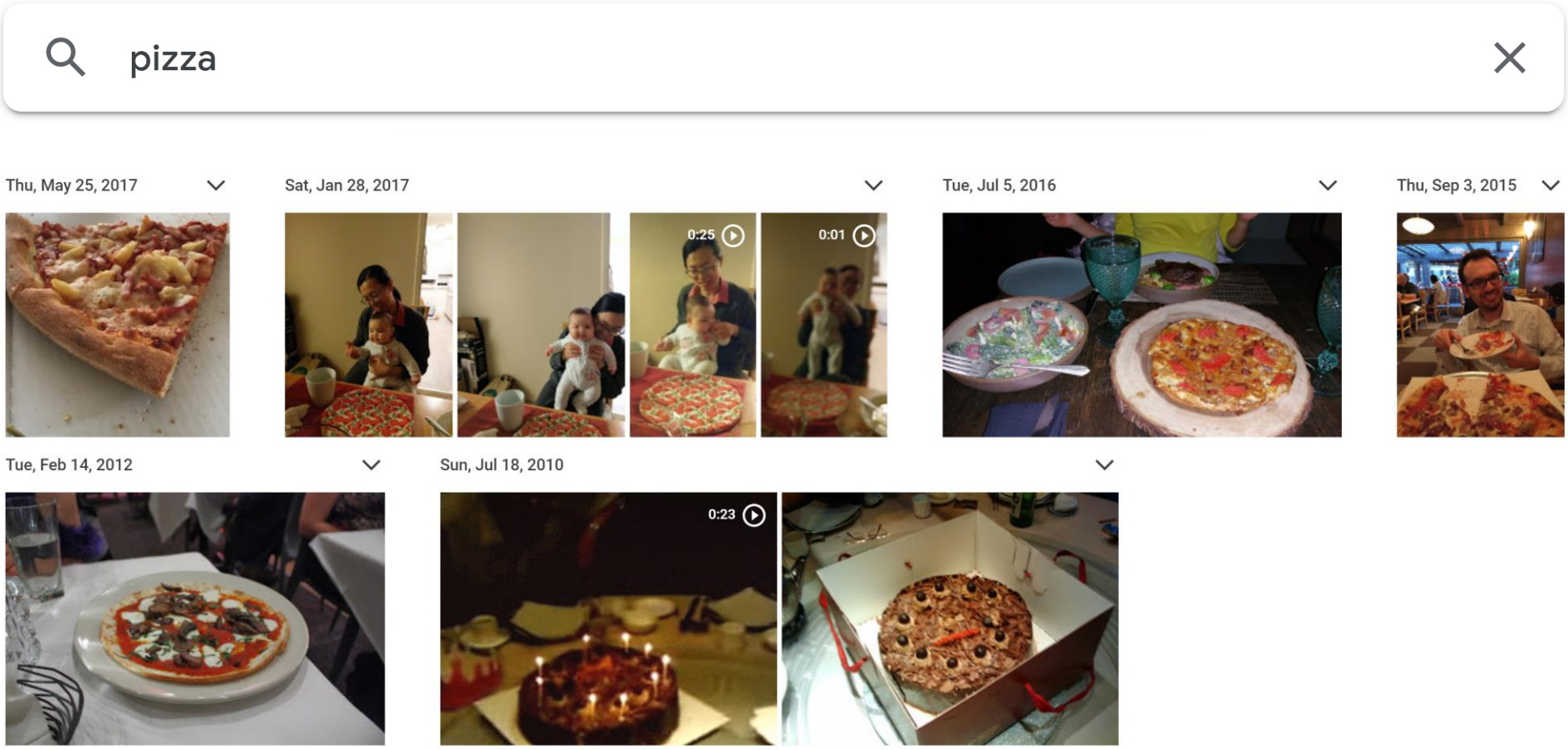
https://ai.googleblog.com/2019/04/take-your-best-selfie-automatically.html

(Also features “kiss detection”)

Applications:
Assisted / autonomous driving

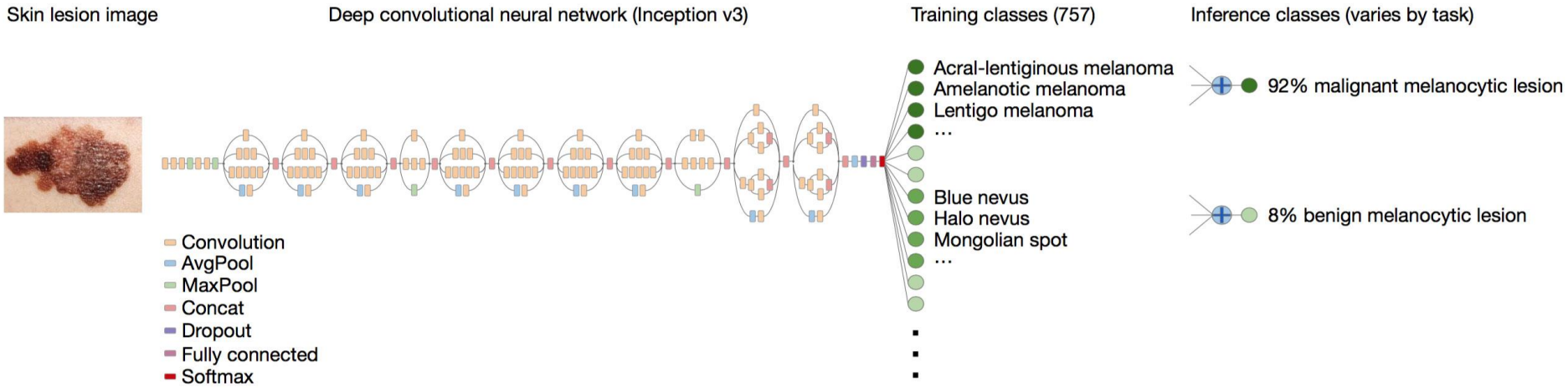
https://www.extremetech.com/extreme/226071-nvidia-goes-all-in-on-self-driving-cars-including-a-robotic-car-racing-league

Applications: Photo organization



Source: Google Photos

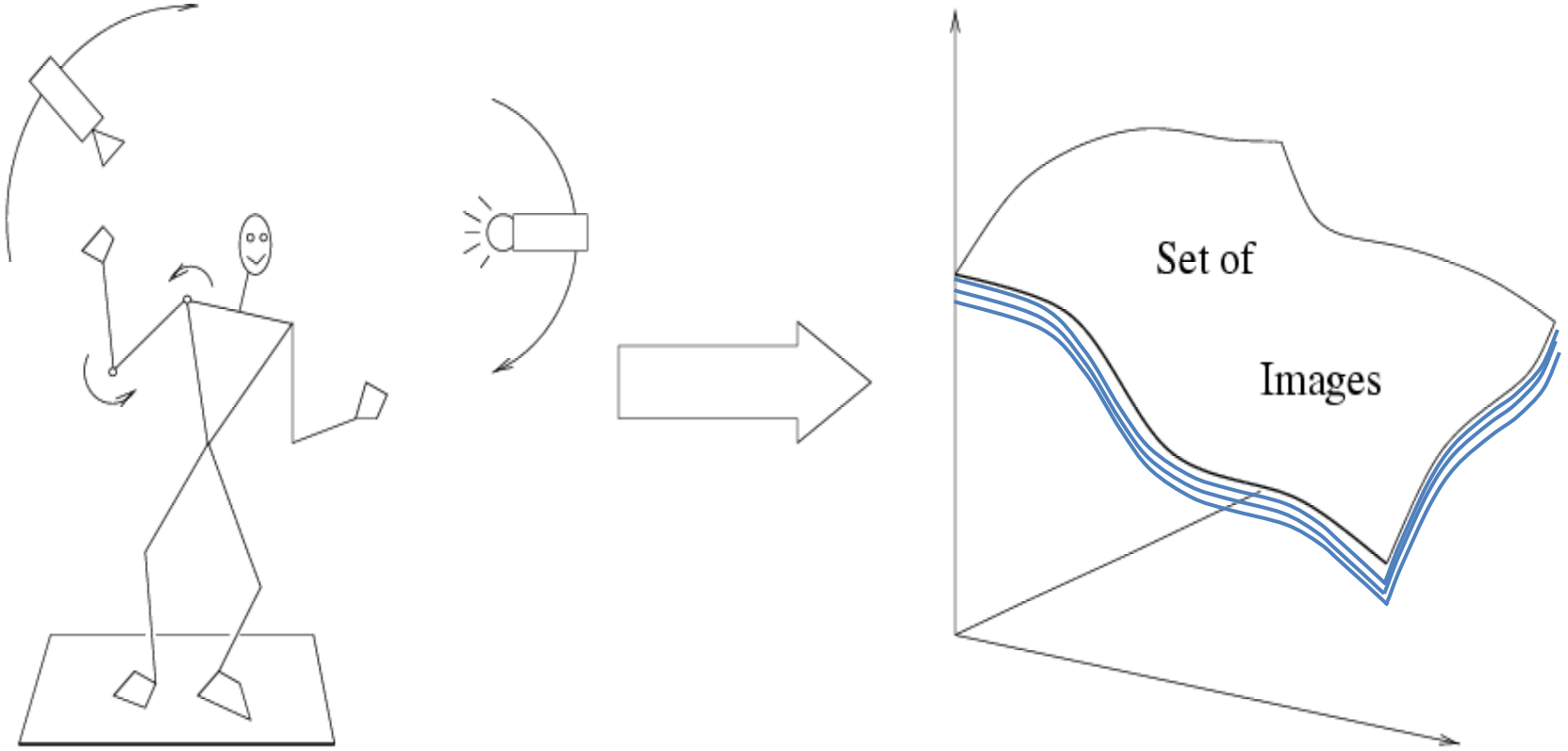
Applications: medical imaging



Dermatologist-level classification of skin cancer

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

Why is this hard?



Variability: Camera position
Illumination
Shape parameters

How many object categories are there?

Challenge: variable viewpoint

Michelangelo 1475-1564

Challenge: variable illumination

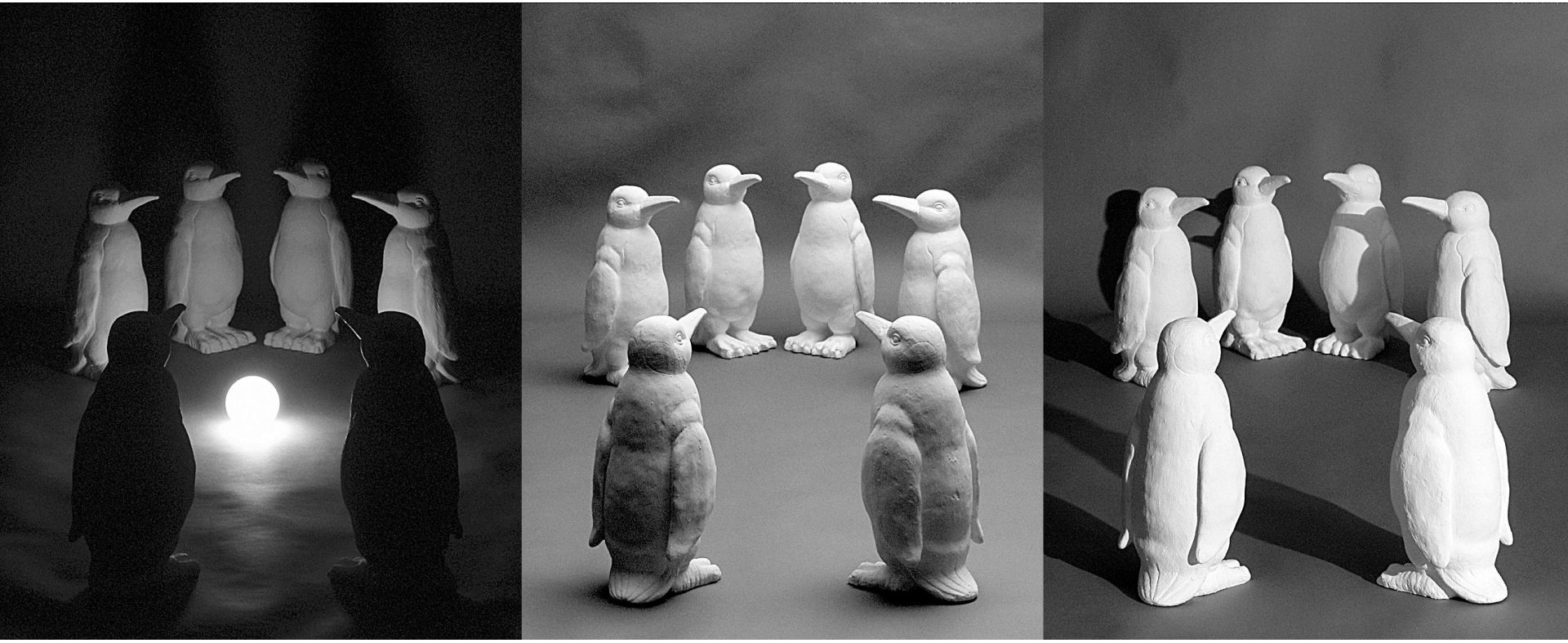
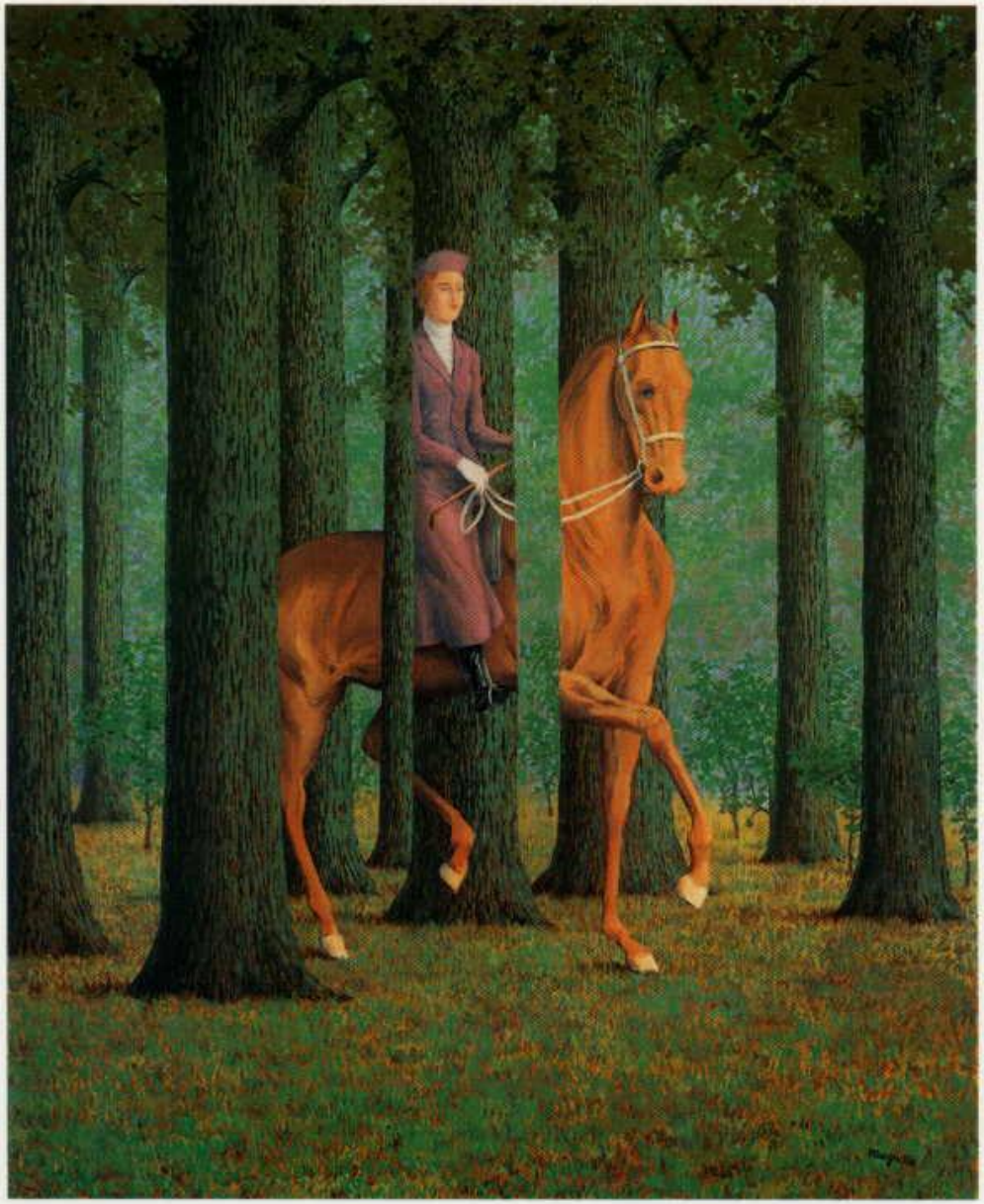


image credit: J. Koenderink

Challenge: scale

Challenge: deformation

Challenge: Occlusion

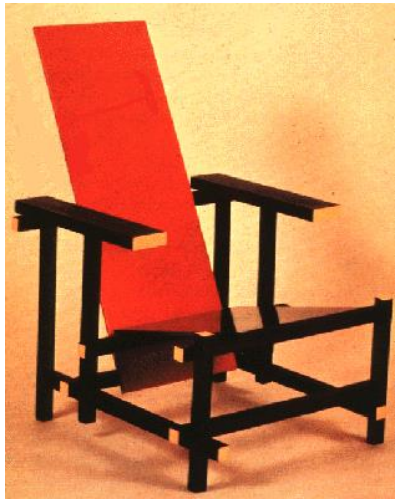


Magritte, 1957

Challenge: background clutter

Kilmeny Niland. 1995

Challenge: intra-class variations



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)
  - Robust object classification across thousands of object categories (outperforming humans)

“Spotted salamander”

A brief history of image recognition

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

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.

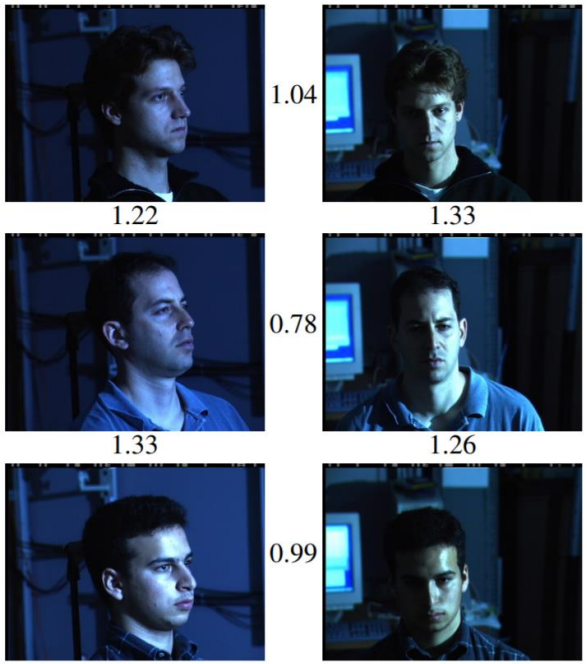


Figure 1. **Illumination and Pose invariance.** Pose and illumination have been a long standing problem in face recognition. This figure shows the output distances of FaceNet between pairs of faces of the same and a different person in different pose and illumination combinations. A distance of 0.0 means 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)
  - High-quality face synthesis (but not yet for completely general scenes)

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

A Style-Based Generator Architecture for Generative Adversarial Networks

Tero Karras (NVIDIA), Samuli Laine (NVIDIA), Timo Aila (NVIDIA)

http://stylegan.xyz/paper

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
  - High level: Contextual dependence
  - Deep learned features
- Data
  - More is always better (as long as it is good data)
  - Annotation is the hard part

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
  - High level: Contextual dependence
  - **Deep learned features**
- **Data**
  - More is always better (as long as it is good data)
  - Annotation is the hard part

24 Hrs in Photos

http://www.kesselskramer.com/exhibitions/24-hrs-of-photos installation by Erik Kessels

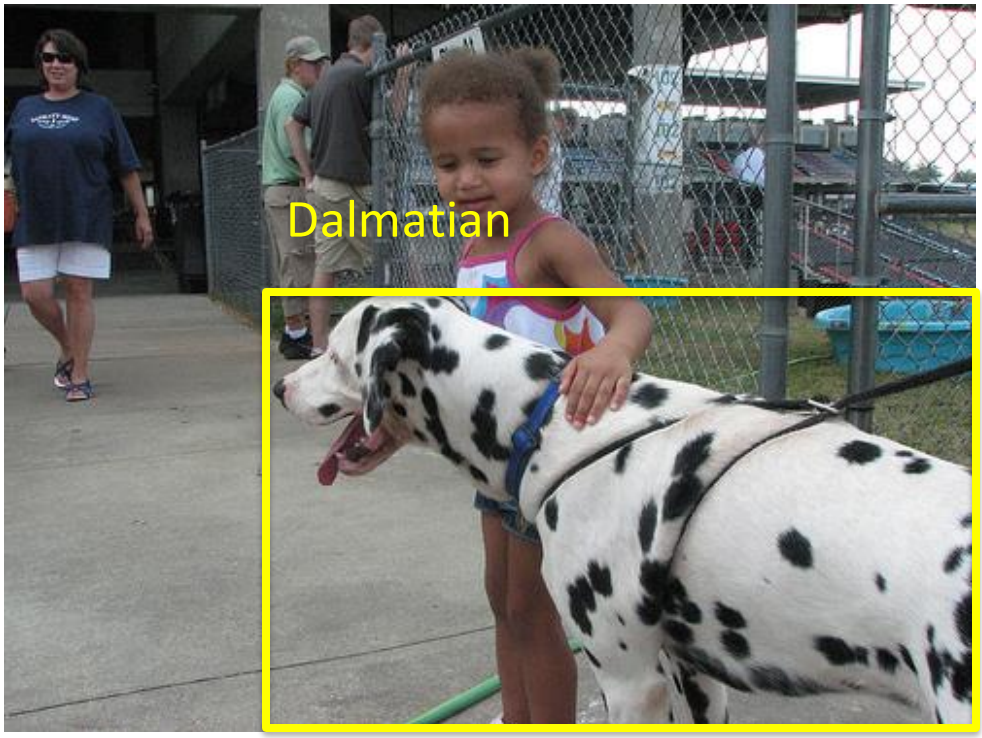
Data Sets

- ImageNet
  - Huge, Crowdsourced, Hierarchical, *Iconic* objects
- PASCAL VOC
  - *Not* Crowdsourced, bounding boxes, 20 categories
- SUN Scene Database, Places
  - *Not* Crowdsourced, 397 (or 720) scene categories
- LabelMe (Overlaps with SUN)
  - Sort of Crowdsourced, Segmentations, Open ended
- SUN *Attribute* database (Overlaps with SUN)
  - Crowdsourced, 102 attributes for every scene
- OpenSurfaces
  - Crowdsourced, materials
- Microsoft COCO
  - Crowdsourced, large-scale objects

IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) 2010-2012

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

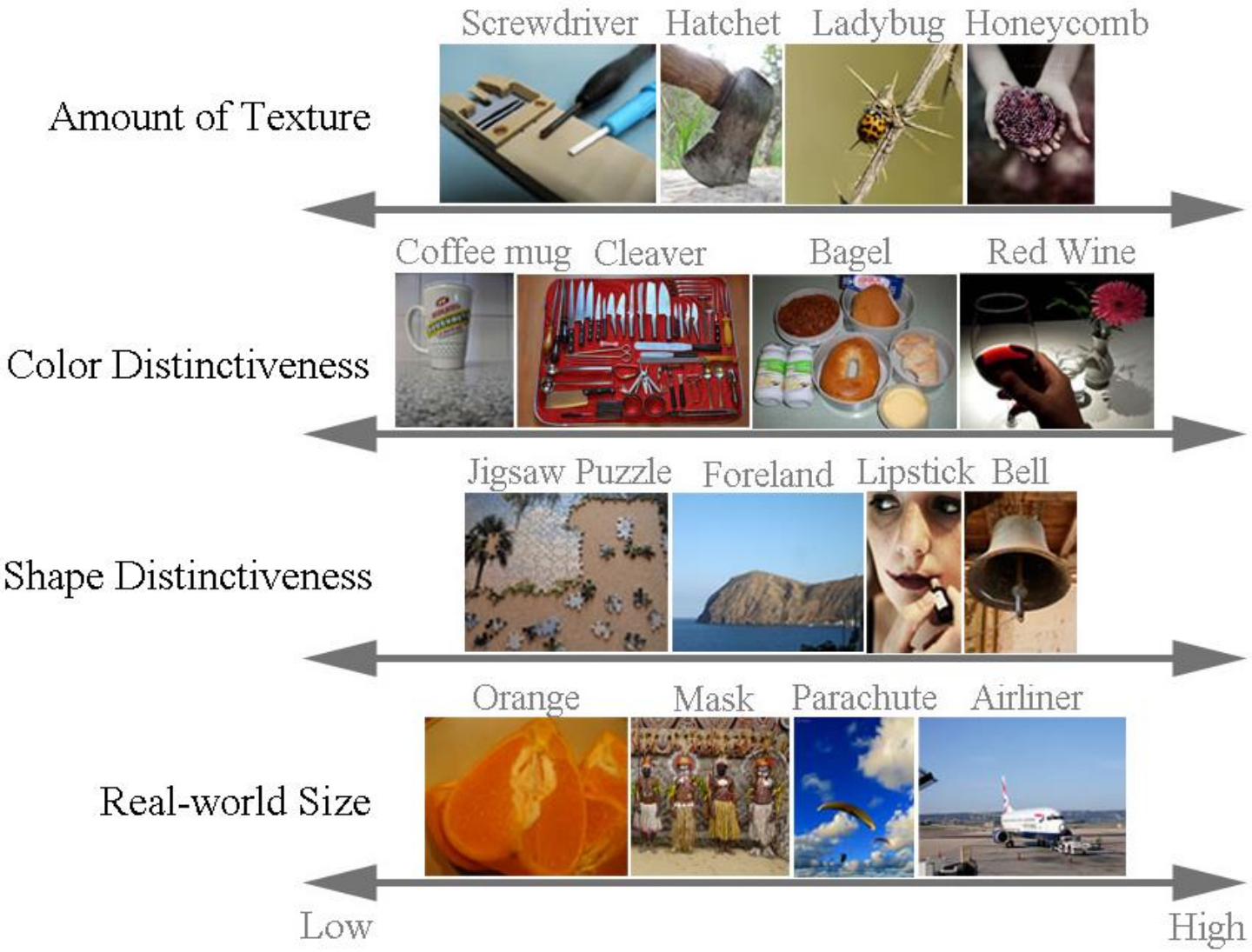
1000 object classes **1,431,167 images**



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

Variety of object classes in ILSVRC

Variety of object classes in ILSVRC



Questions?