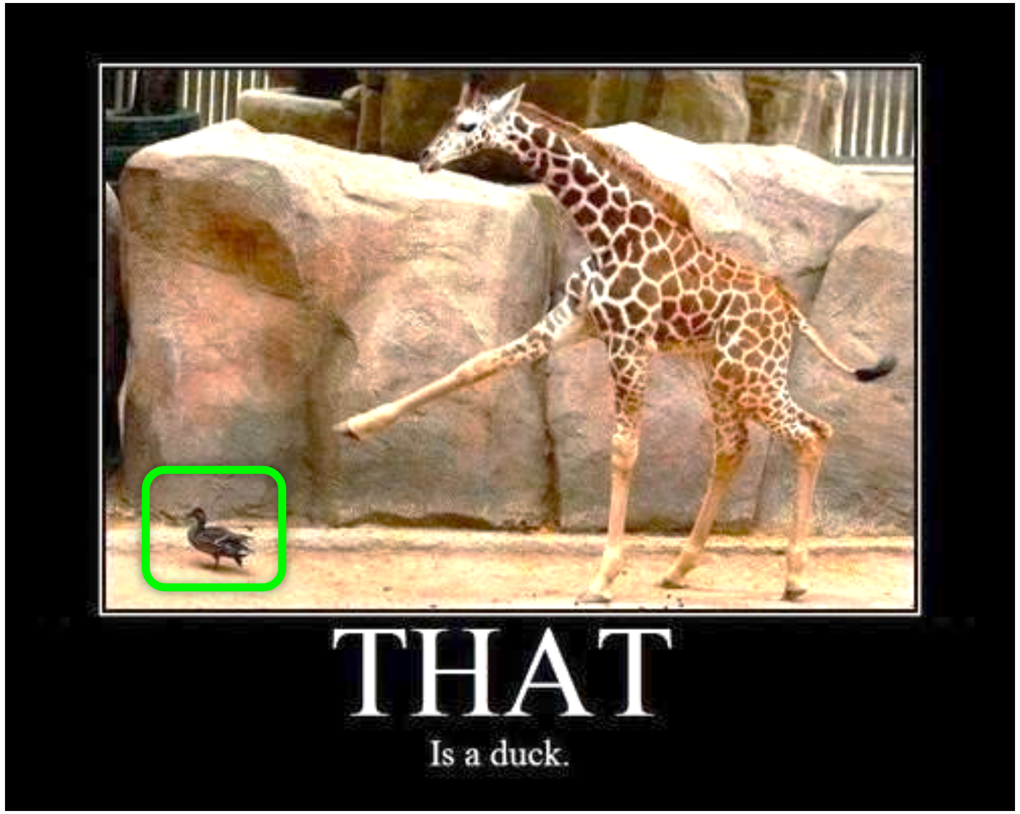
Introduction to Recognition

Presented by Abe Davis

Most Slides Originally from Noah Snavely



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 is “Recognition”?

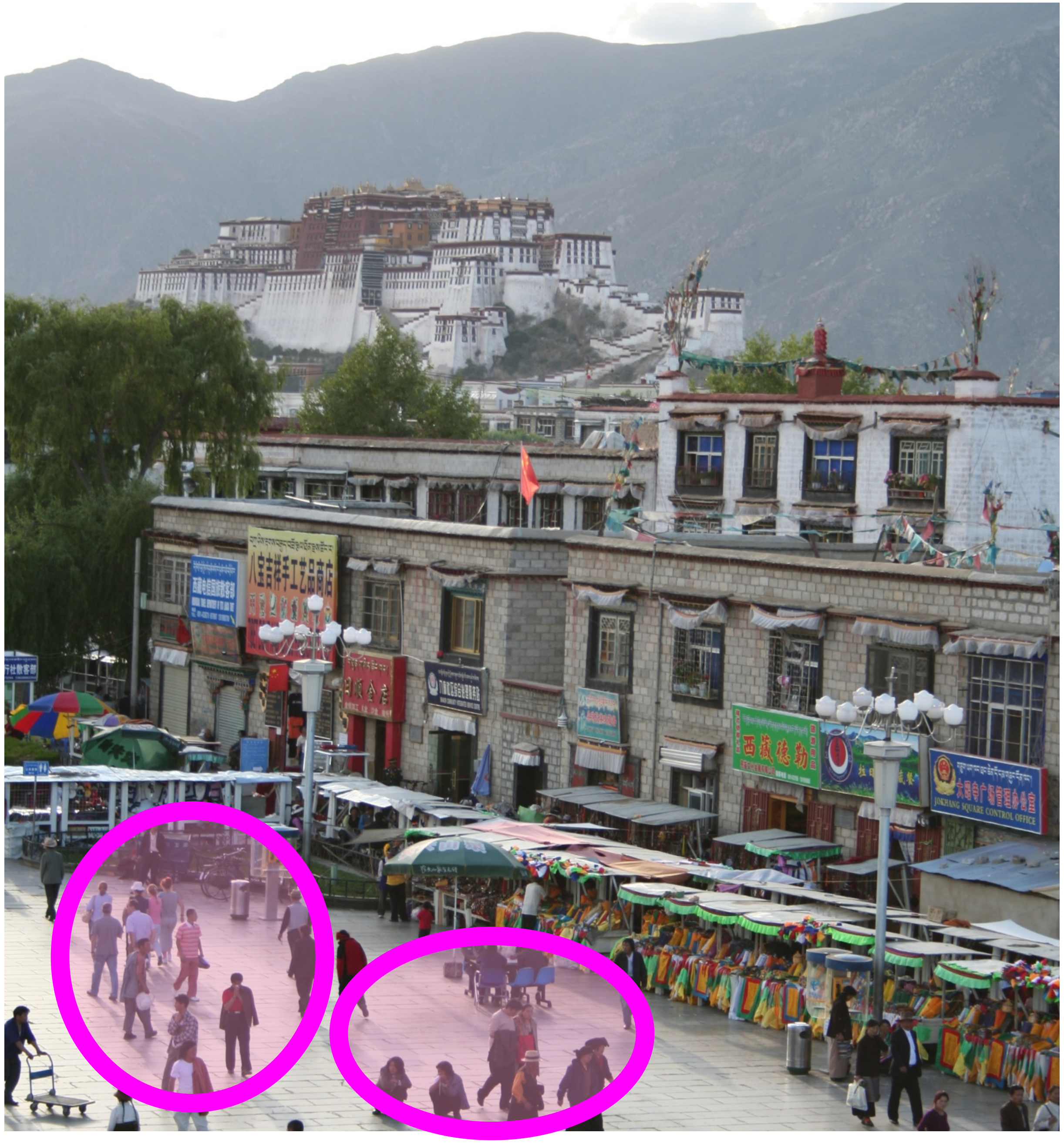
Next few slides adapted from Li, Fergus, & Torralba’s excellent short course on category and object recognition

What is “Recognition”?

- Verification: is that a lamp?

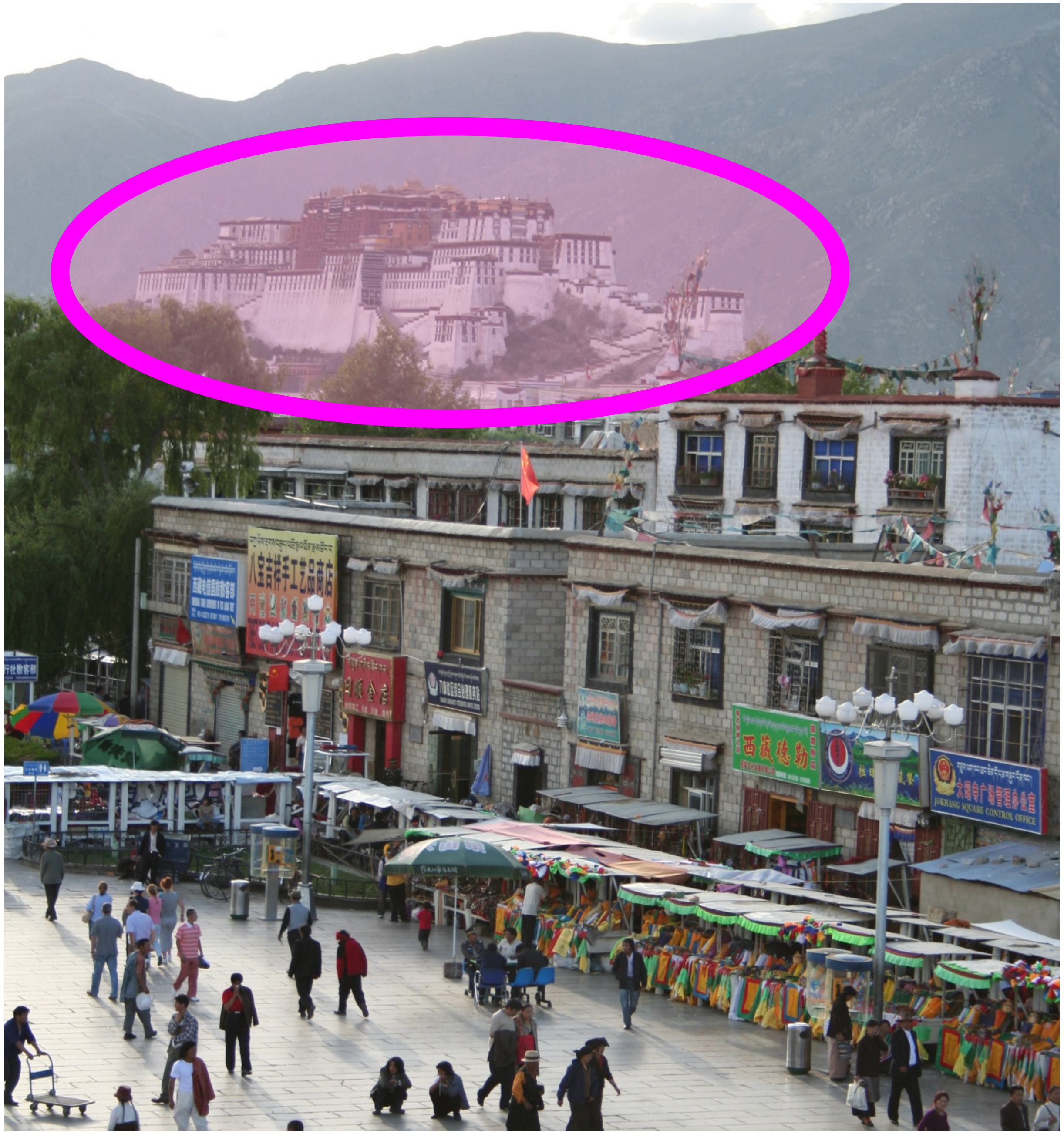
What is “Recognition”?

- 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?

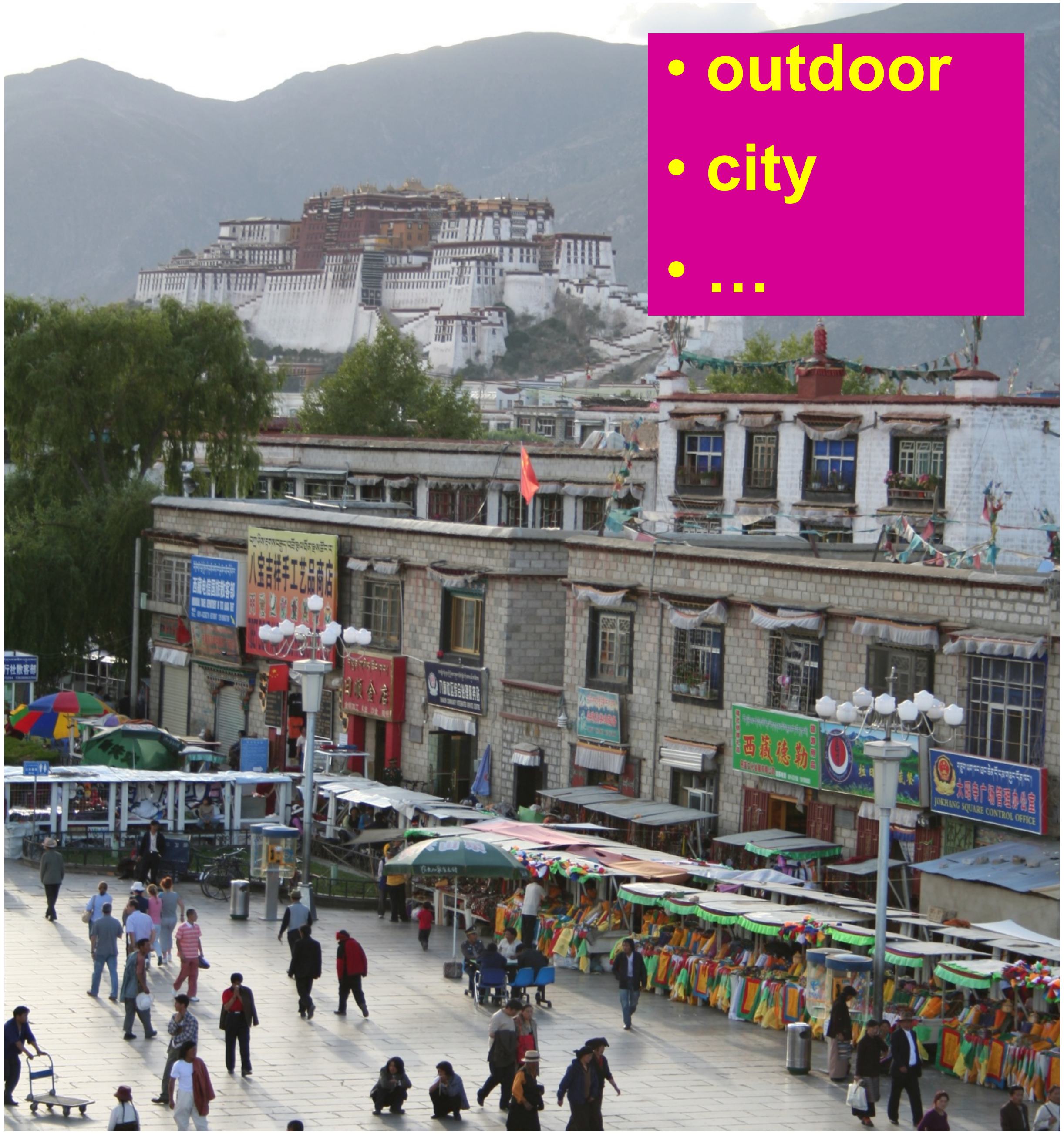


What is “Recognition”?

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

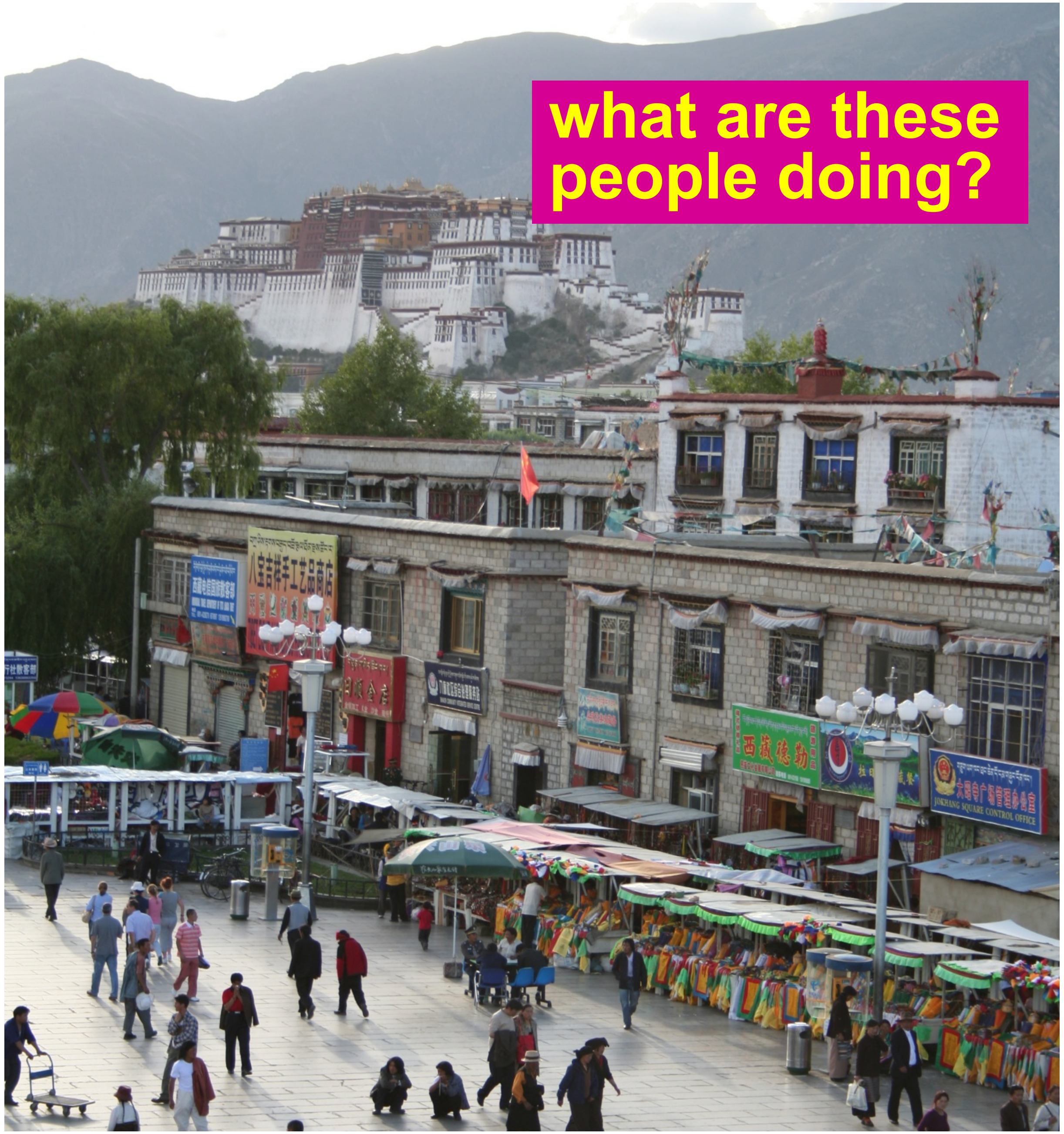
What is “Recognition”?

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



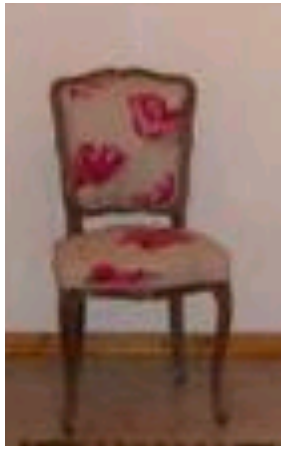
What is “Recognition”?

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

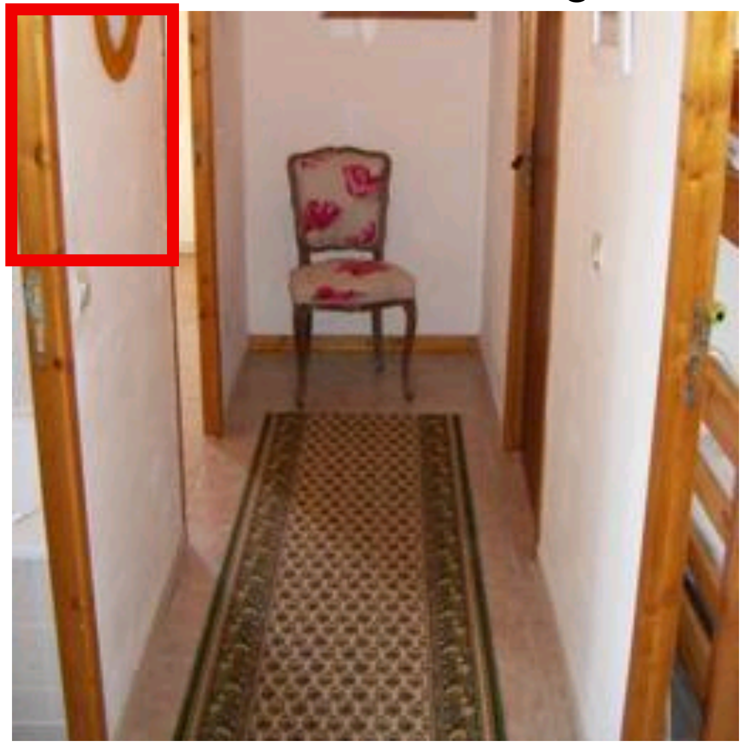


Object recognition: Is it really so hard?

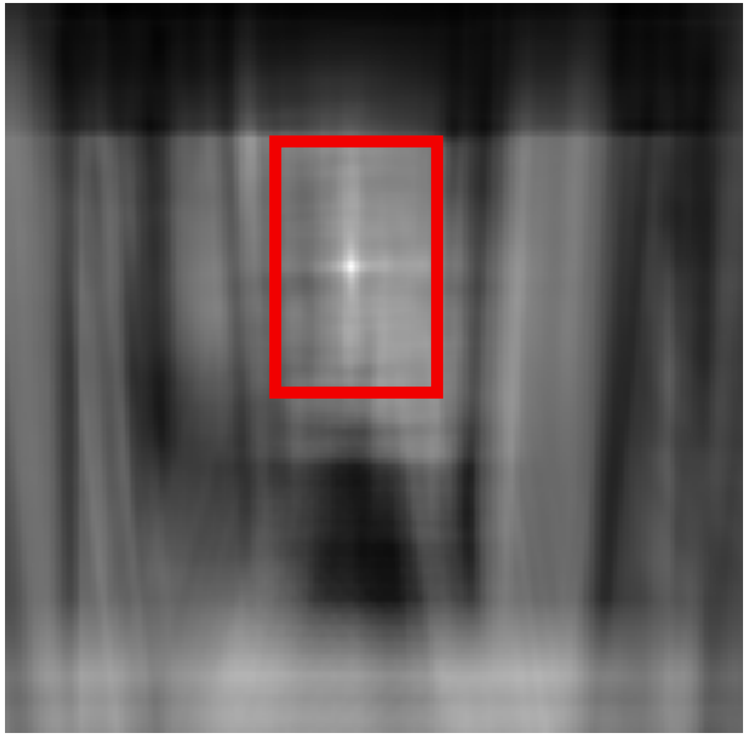
This is a chair



Find the chair in this image

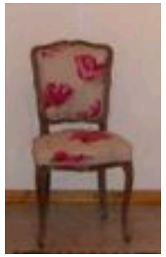


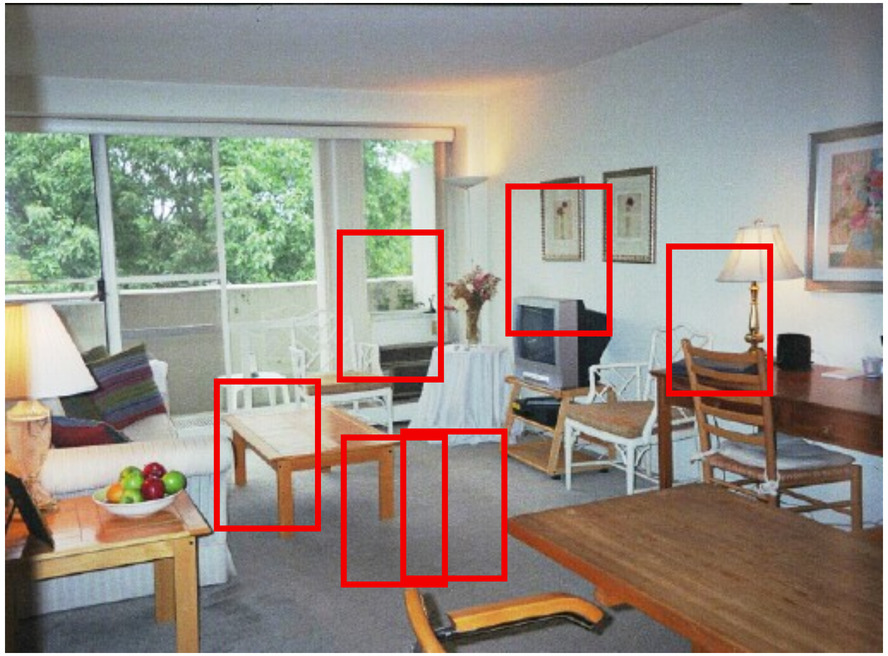
Output of normalized correlation

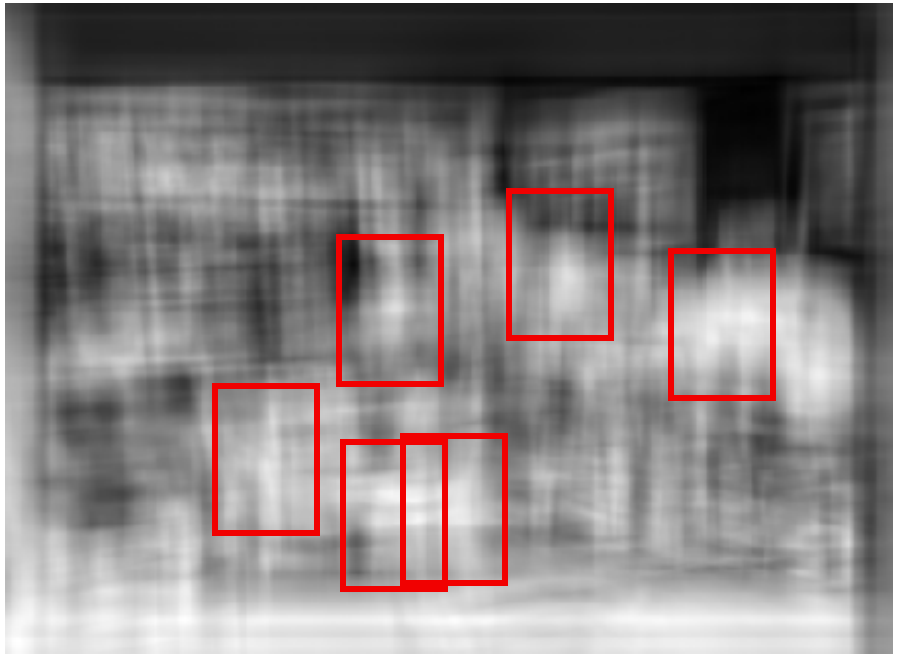


Object recognition: Is it really so hard?

Find the chair in this image





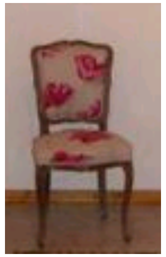


Pretty much garbage

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?

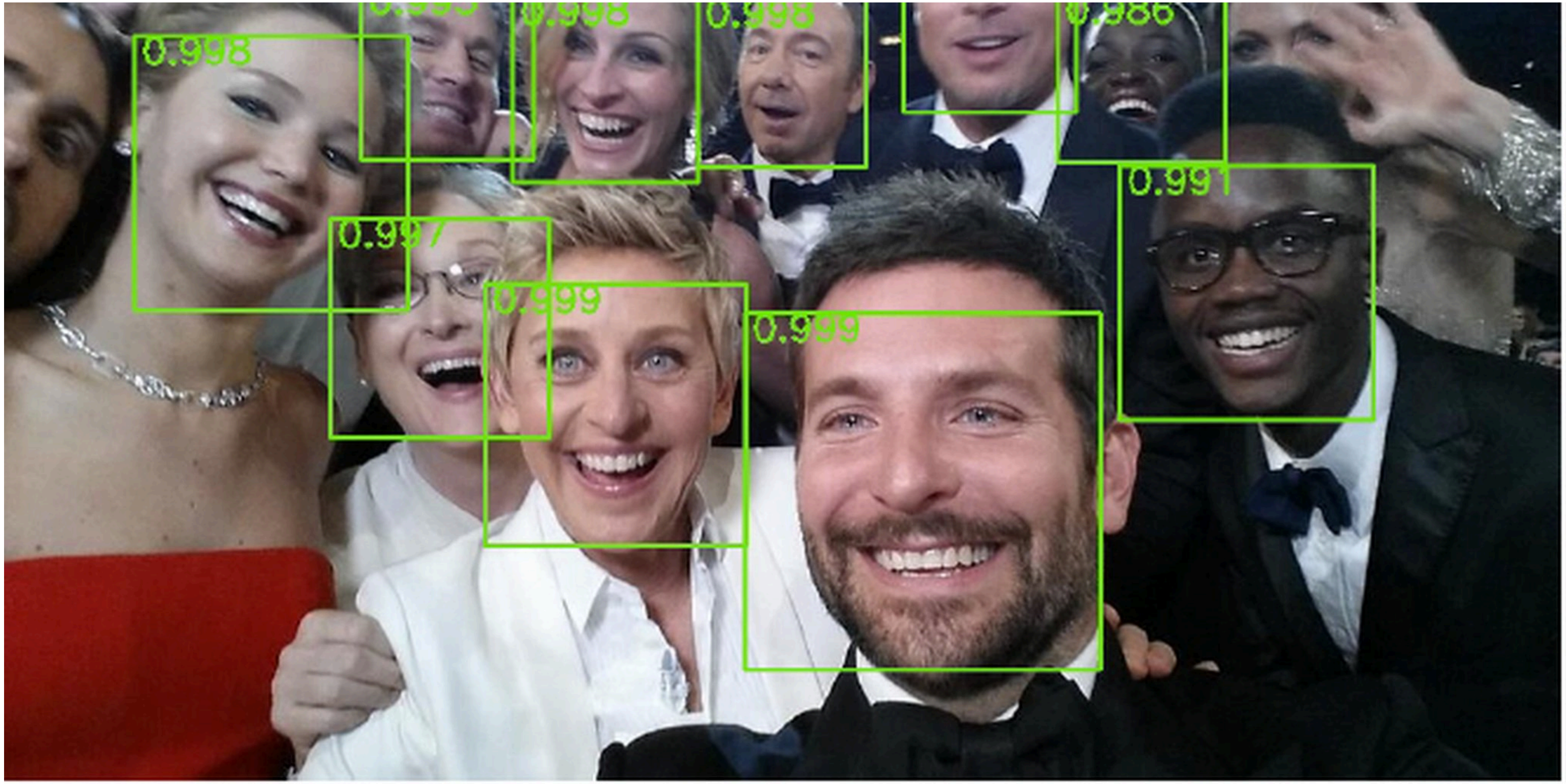
- Works well for object *instances* (or distinctive images such as logos)

- Not great for generic object *categories*

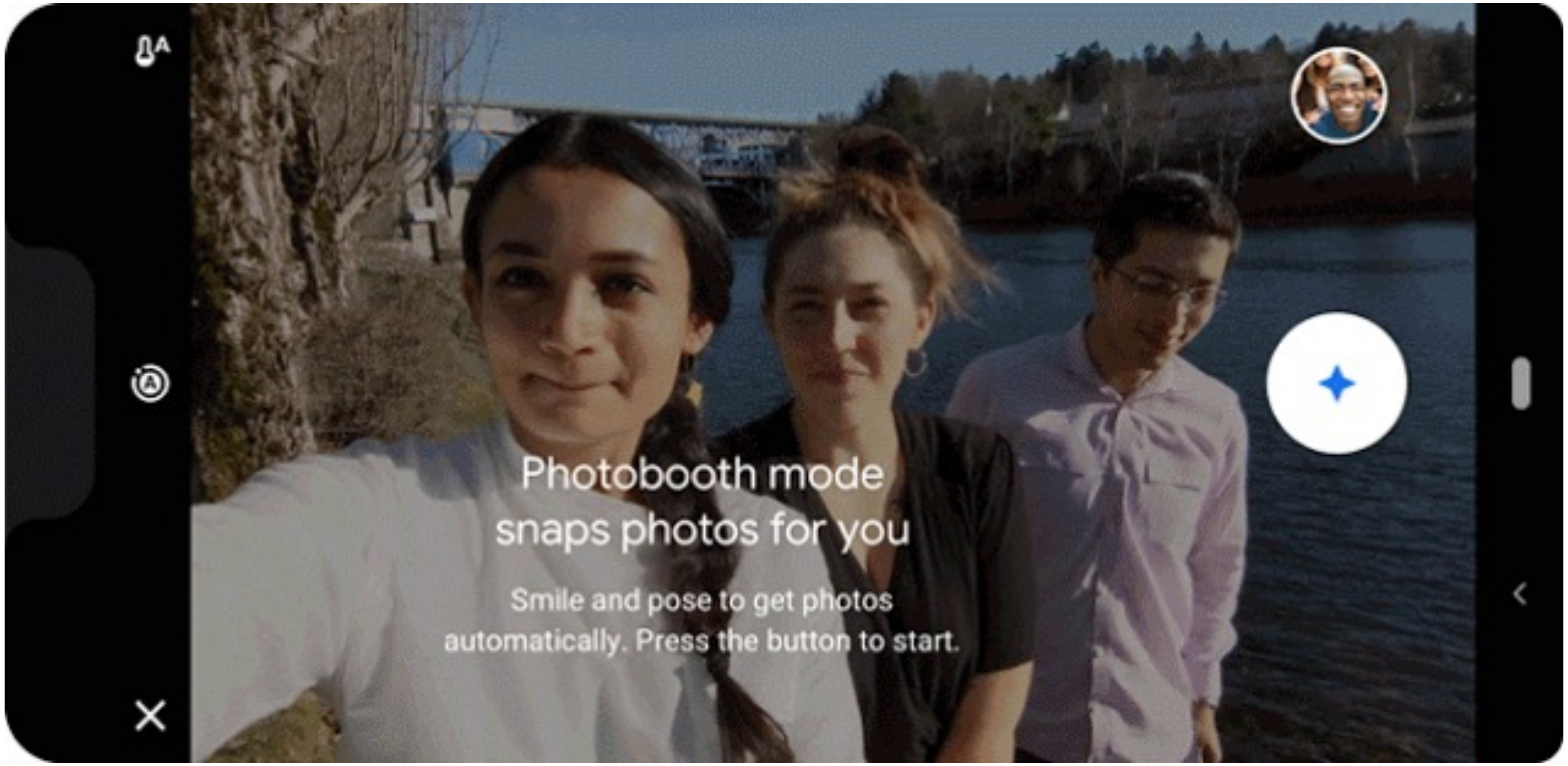
And it can get a lot harder

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

Applications: Photography



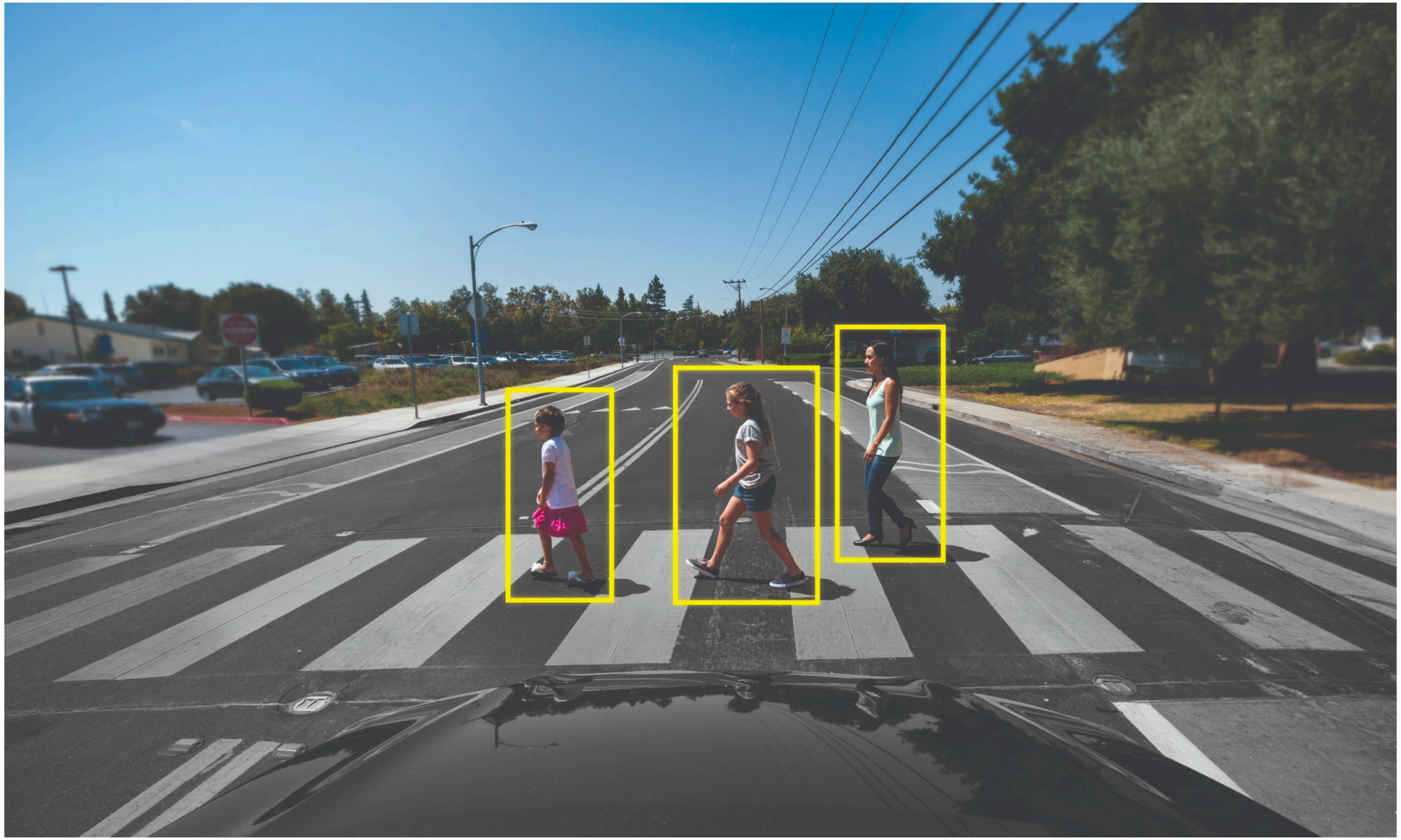
Applications: Shutter-free Photography



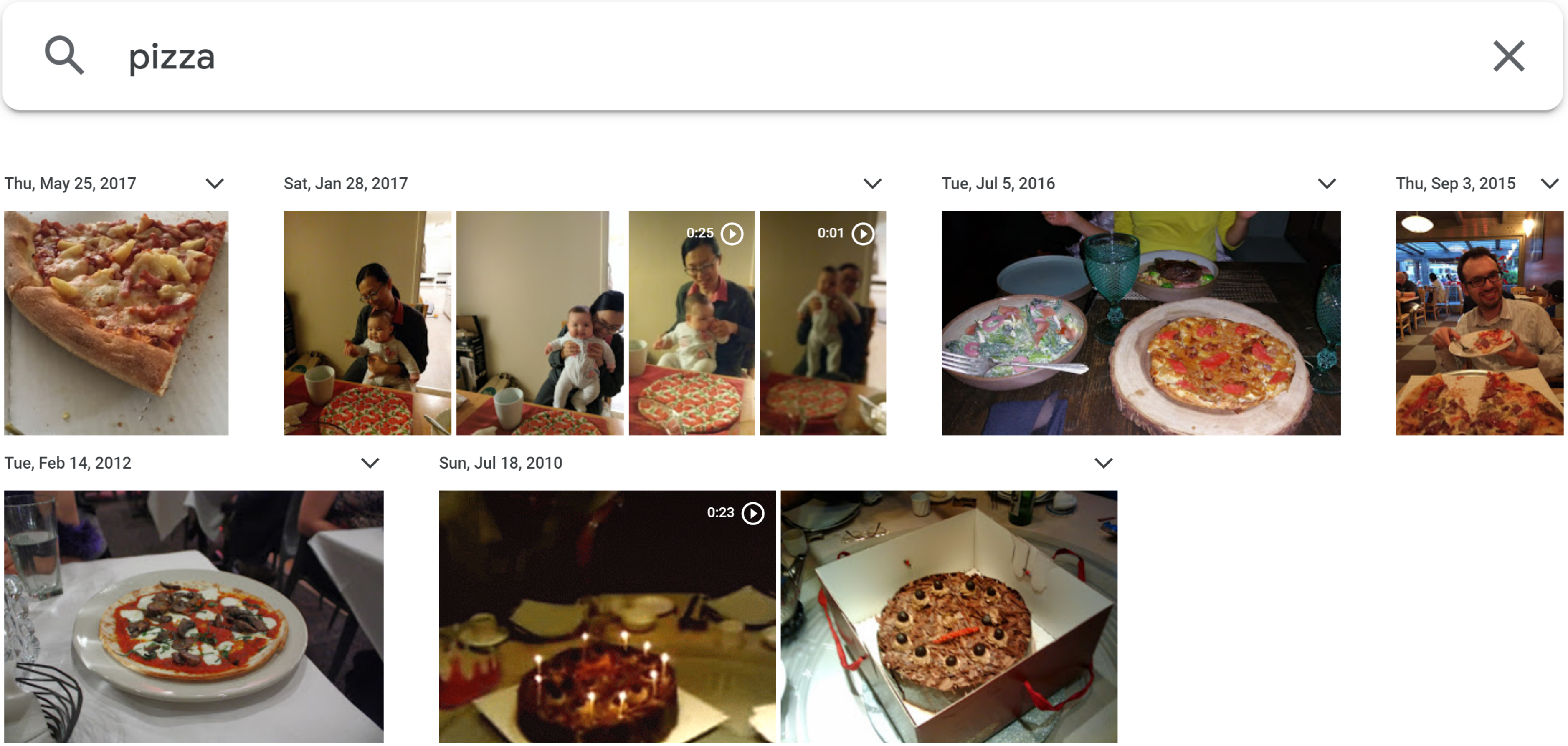
Take Your Best Selfie Automatically, with Photobooth on Pixel 3

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

Applications: Assisted / autonomous driving

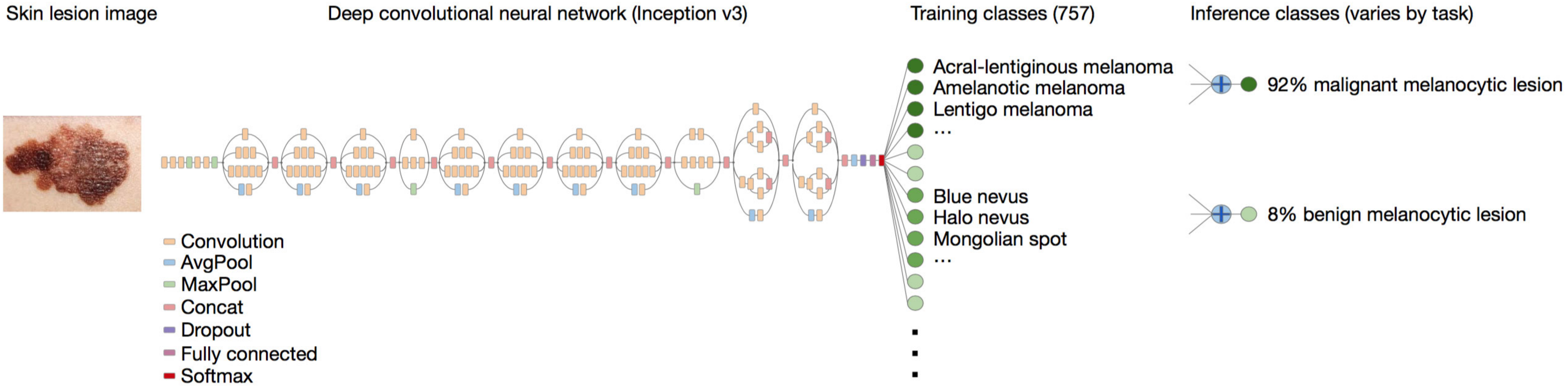


Applications: Photo organization



Source: Google Photos

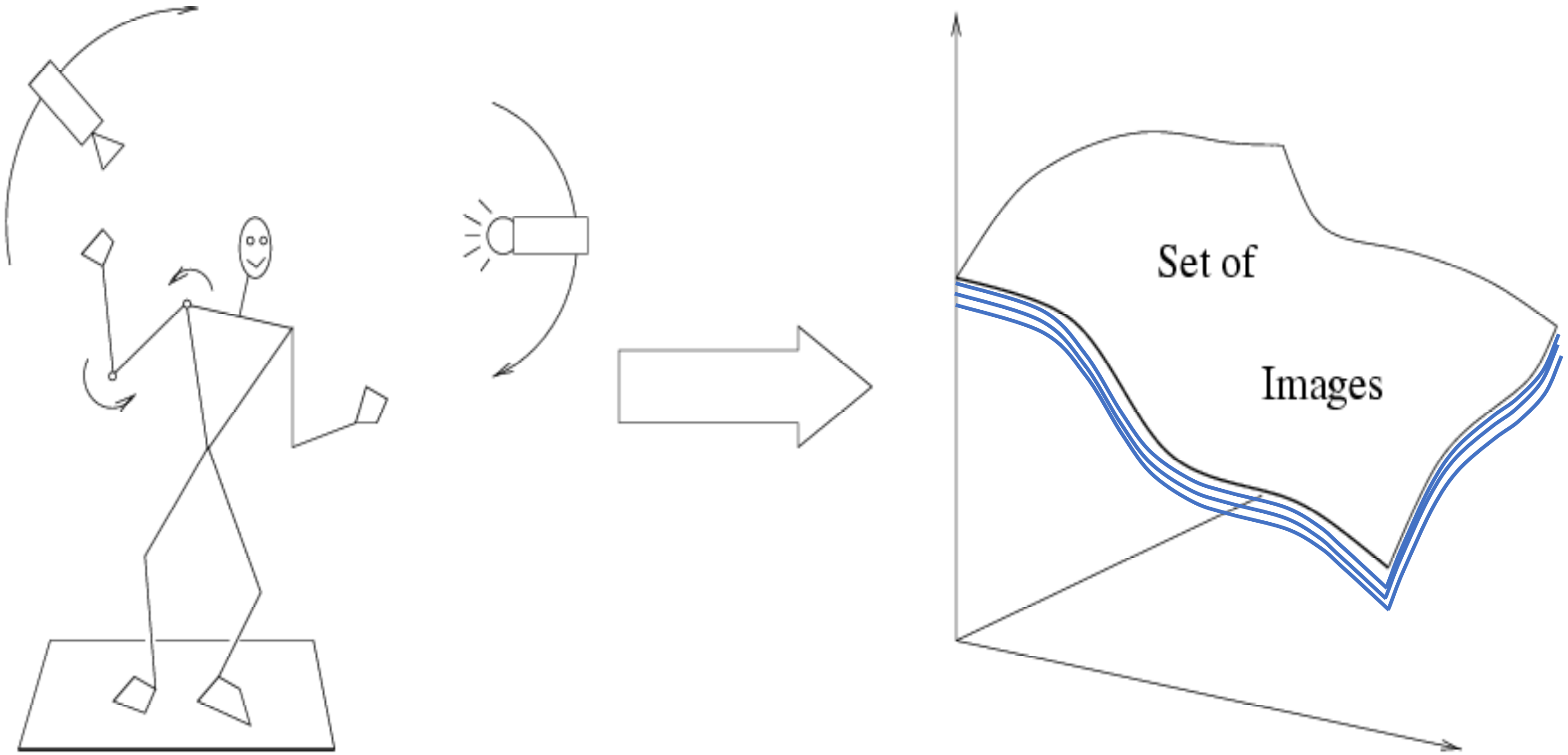
Applications: medical imaging



Dermatologist-level classification of skin cancer

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

Why is recognition hard?

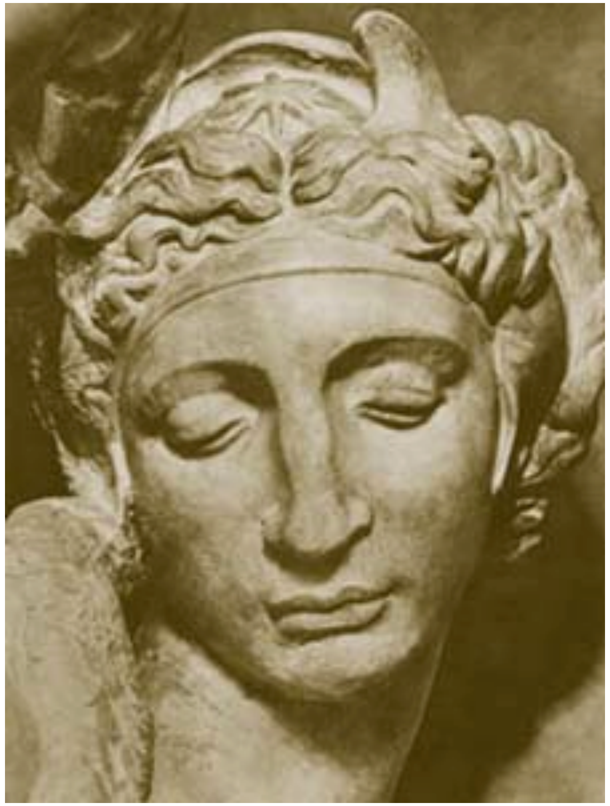


Variability: Camera position,
Illumination,
Shape,
etc…

Svetlana Lazebnik

Challenge: lots of potential classes

Challenge: variable viewpoint



Michelangelo 1475-1564

Challenge: variable illumination

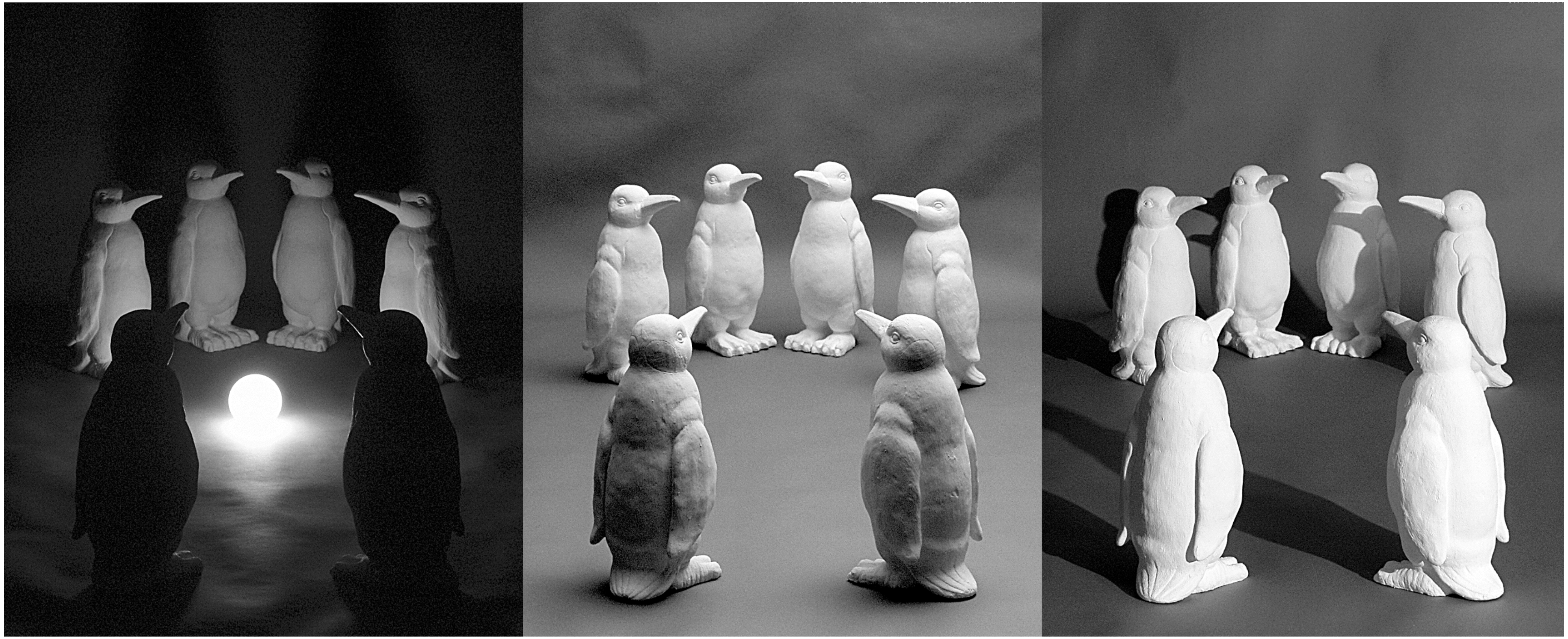
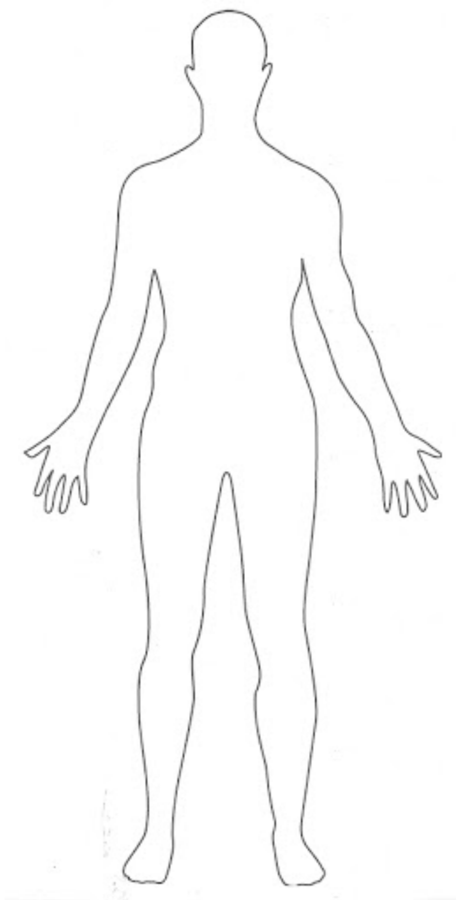


image credit: J. Koenderink

Challenge: scale

Challenge: deformation



Challenge: Occlusion

Magritte, 1957

Challenge: background clutter

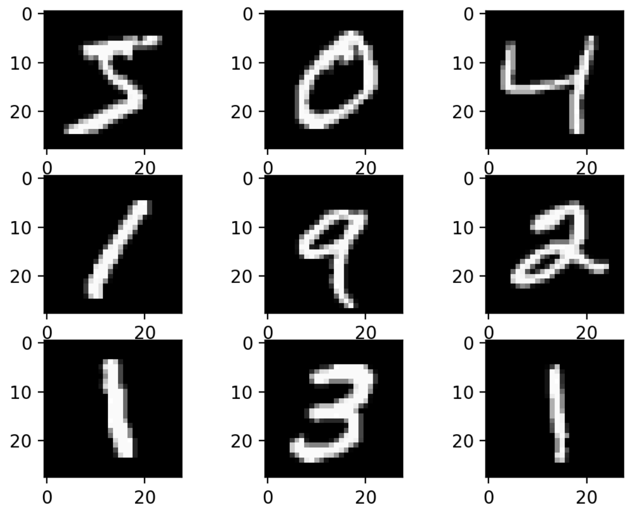
Kilmeny Niland. 1995

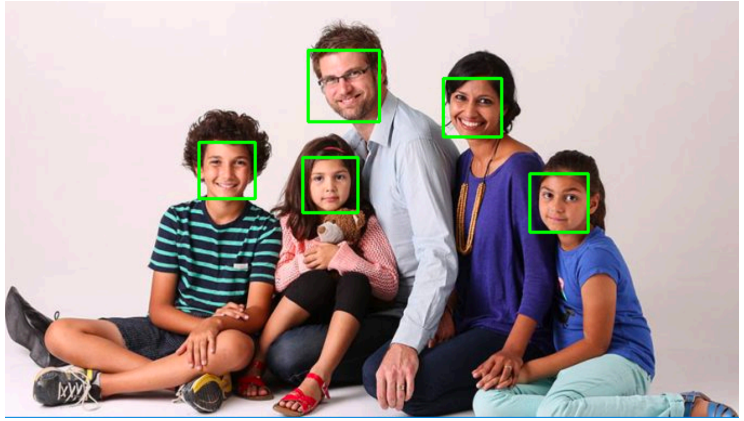
Challenge: intra-class variations

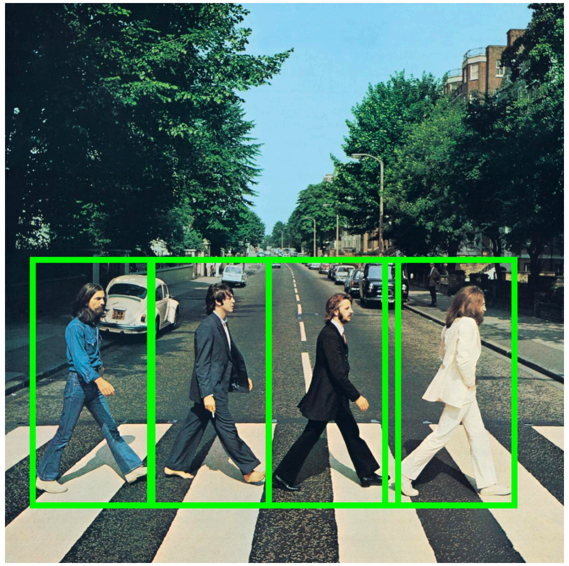
Svetlana Lazebnik

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)
  - 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.

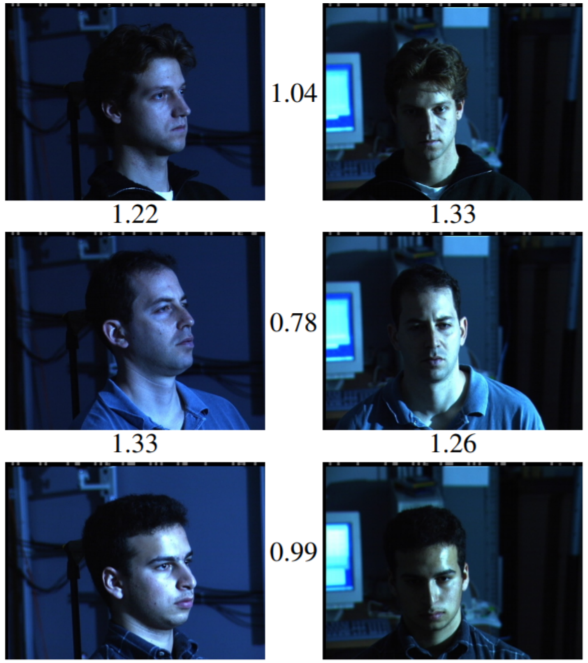


Figure 1. **Illumination and Pose invariance.** Pose and illumination have been a long standing problem in face recognition. This figure shows the output distances of FaceNet between pairs of faces of the same and a different person in different pose and illumination combinations. A distance of 0.0 means 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)

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.

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

Flickr Photos From 1 Day in 2011

http://www.kesselskramer.com/exhibitions/24-hrs-of-photos

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

Large Scale Visual Recognition Challenge (ILSVRC)

2010-2017

IMAGENET

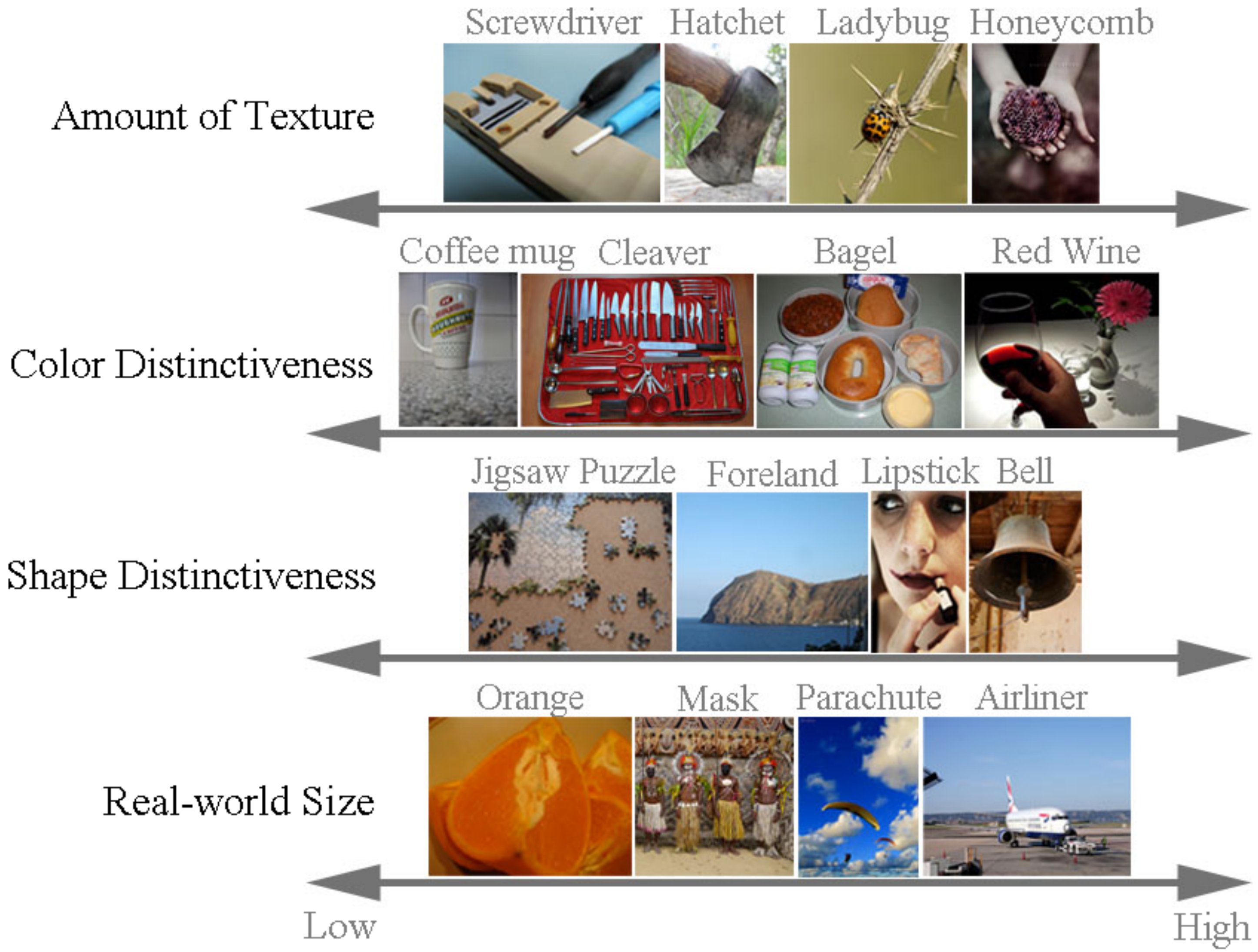
~~20 object classes~~ ~~22,591 images~~

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

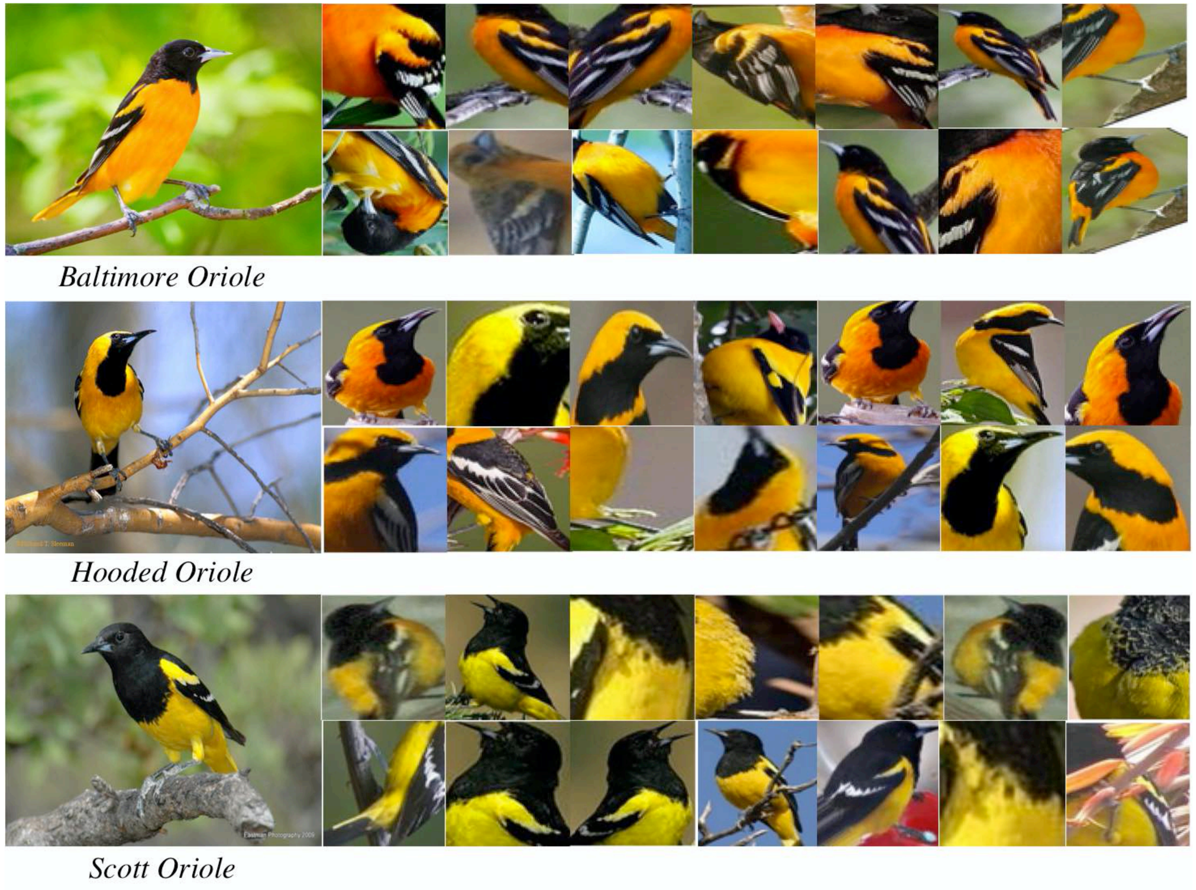


What's Still Hard?

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

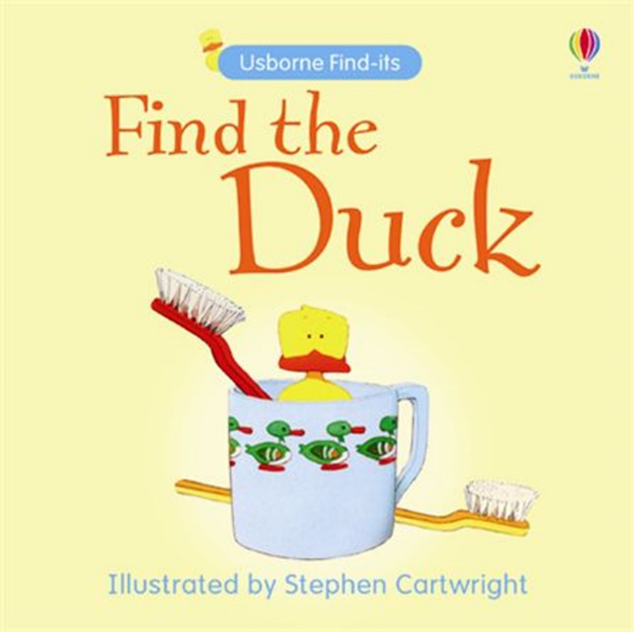
- Fine-grain classification
  - How do we distinguish between more subtle class differences?

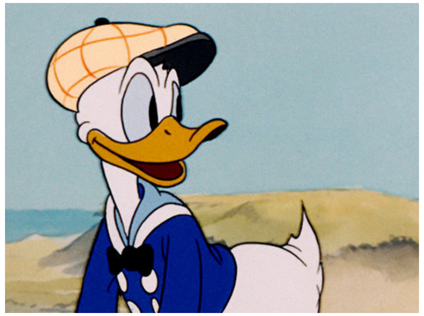
Animal->Bird->Oriole…

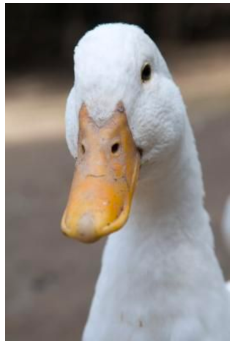


What’s Still Hard?

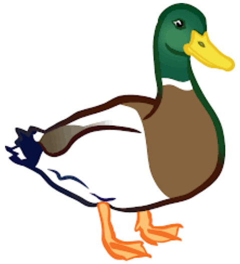
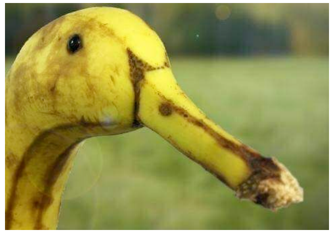
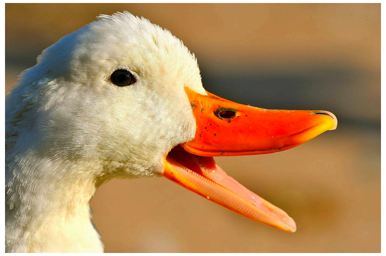
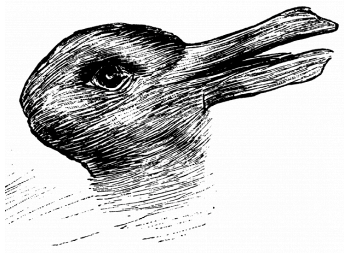
- Few shot learning
  - How do we generalize from only a small number of examples?









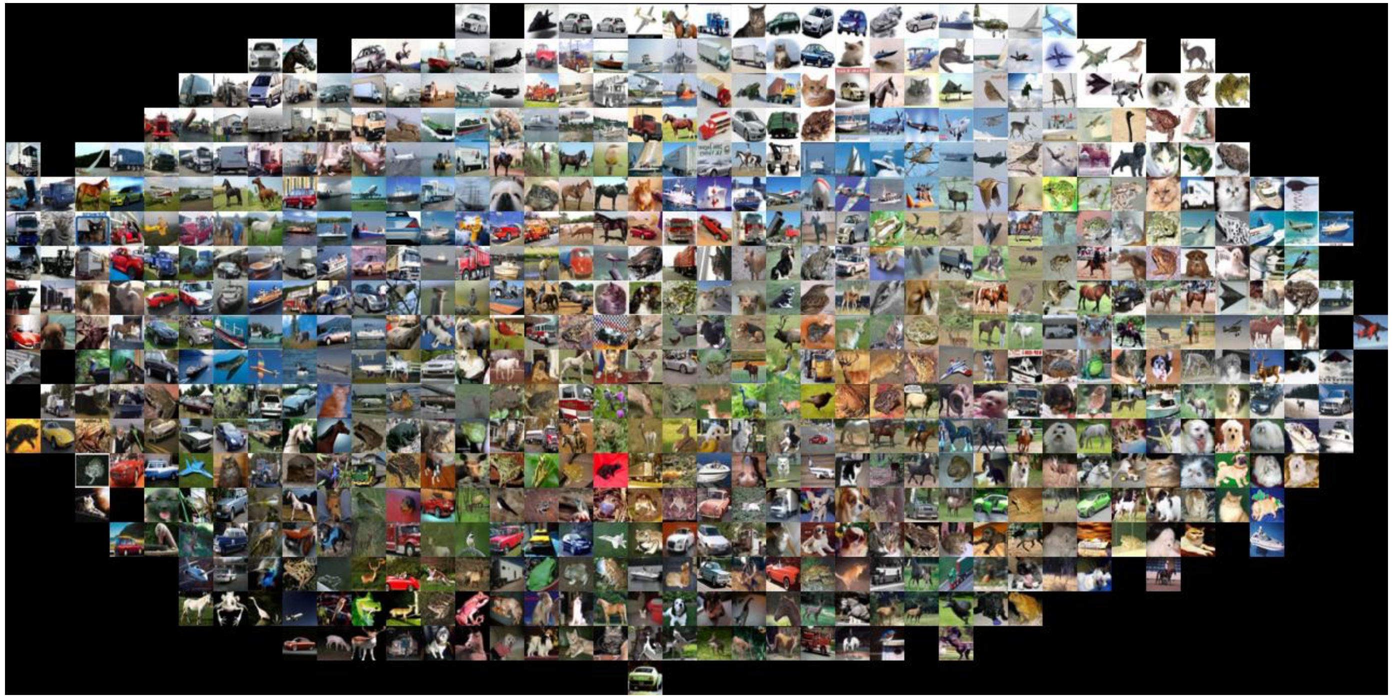





Questions?

CS5670: Computer Vision

Next Time…

Image Classification



Some Slides from Fei-Fei Li, Justin Johnson, Serena Yeung
http://vision.stanford.edu/teaching/cs231n/

Next Time

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

- Building up to CNNs for learning
  - Next four lectures on deep learning

Image Classification: A core task in Computer Vision

- Assume given set of discrete labels
  - e.g. {cat, dog, cow, apple, tomato, truck, … }

f() = “apple”

f() = “tomato”

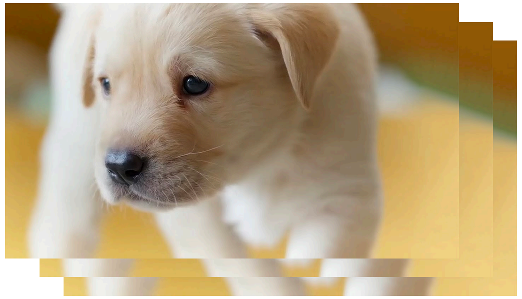
f() = “cow”

Dataset: ETH-80, by B. Leibe   Slide credit: L. Lazebnik

Classification

“Cat”

“Toaster”



“Dog”

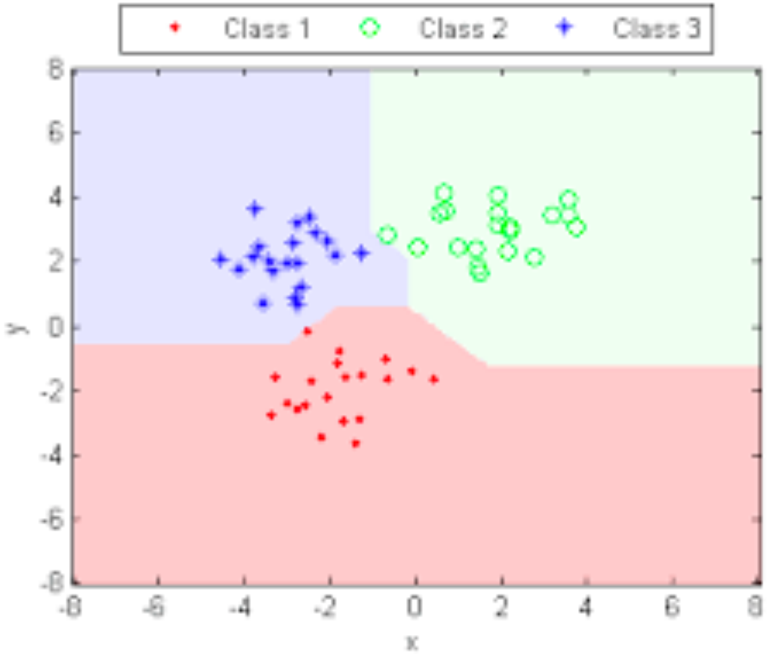
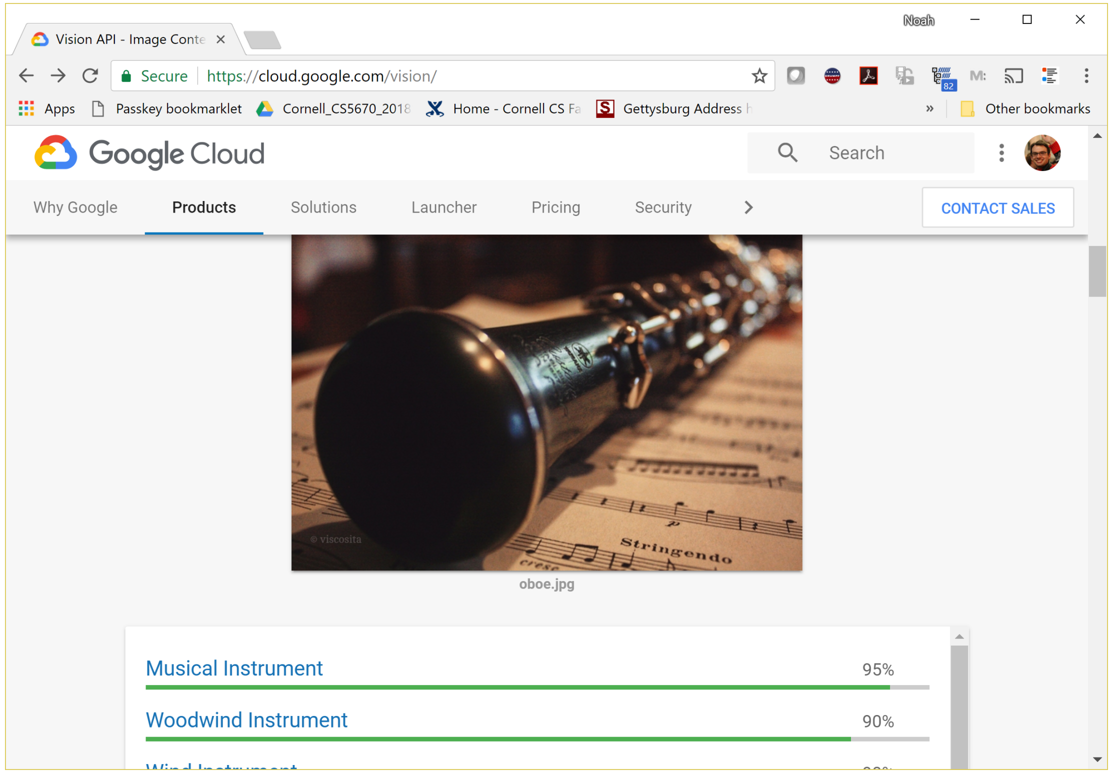


Image classification demo



https://cloud.google.com/vision/

See also:

https://aws.amazon.com/rekognition/

https://www.clarifai.com/

https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/

…

Questions?