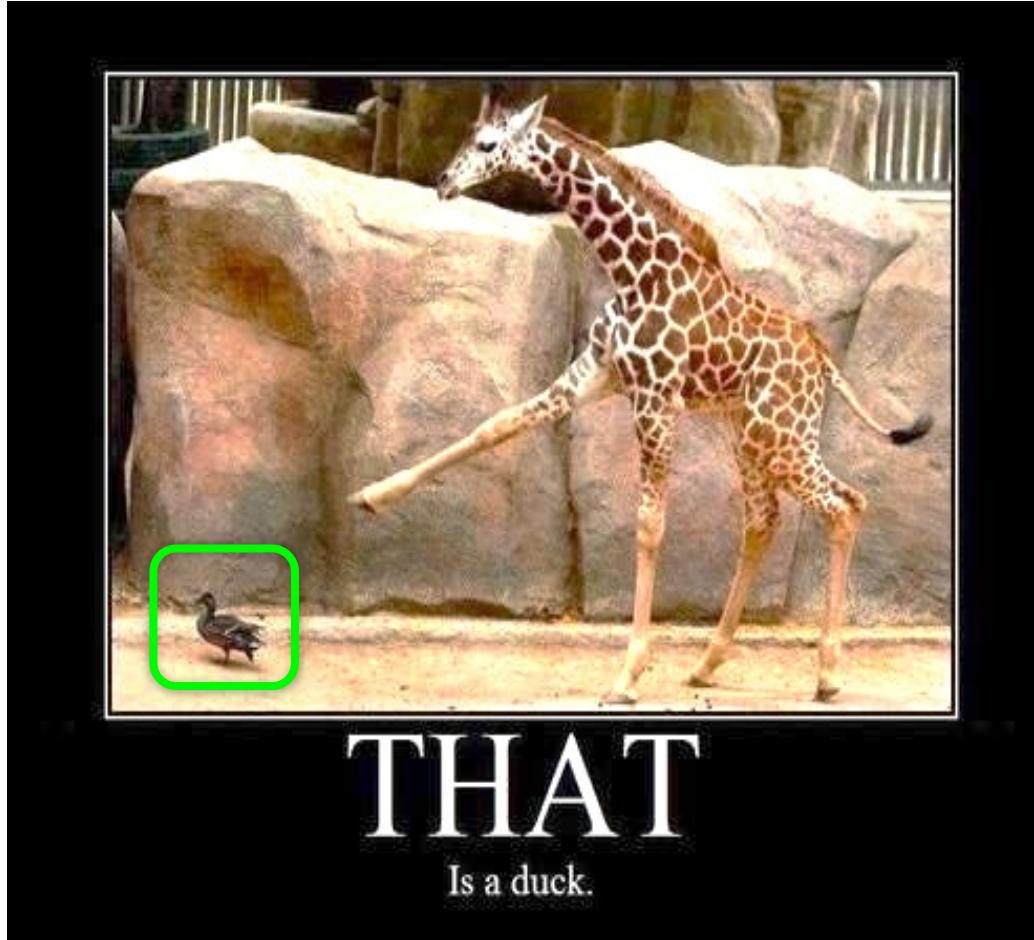


CS7GV1: Computer Vision

Introduction to Recognition



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?



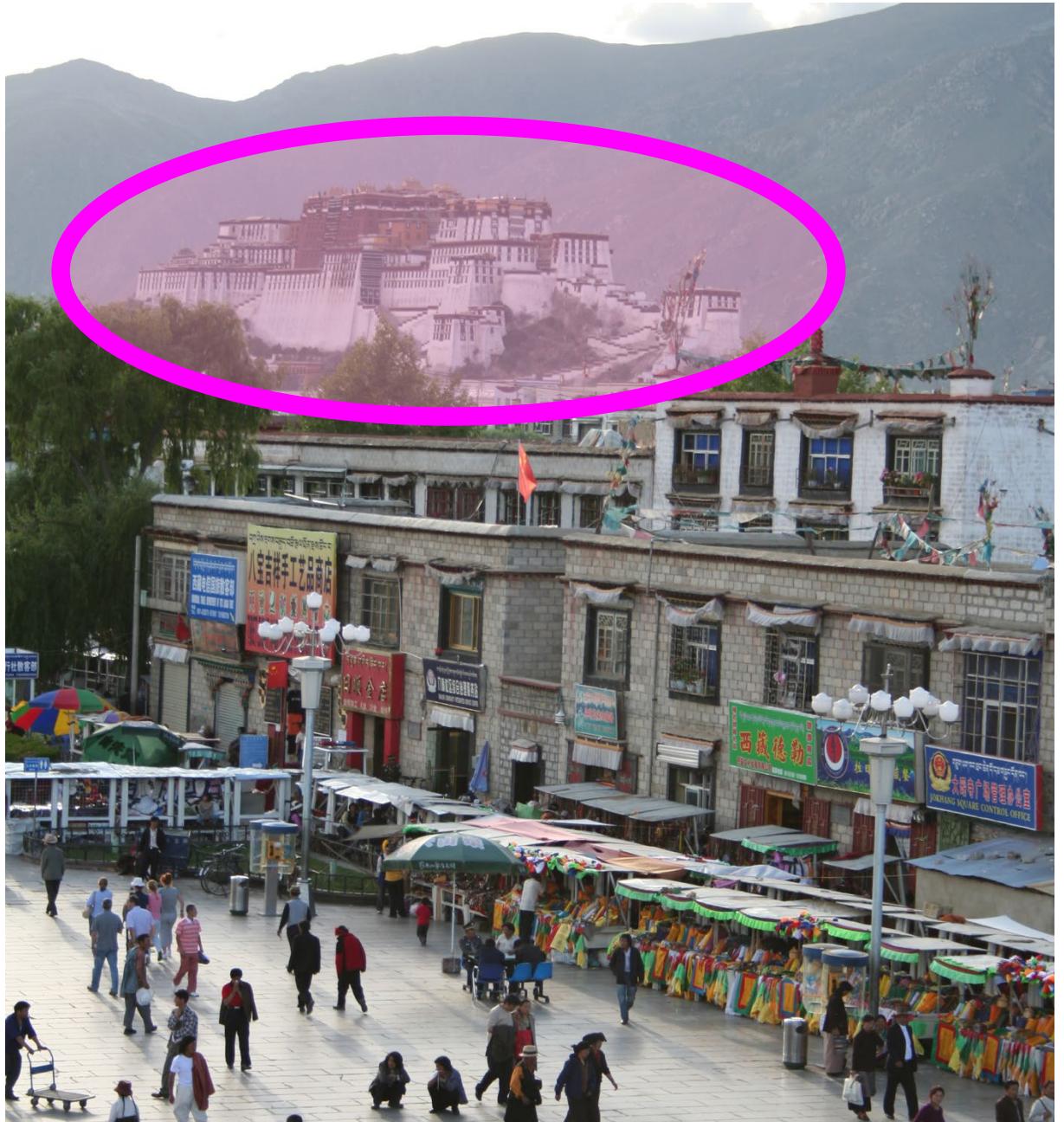
What is “Recognition”?

- 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



mountain

tree

building

banner

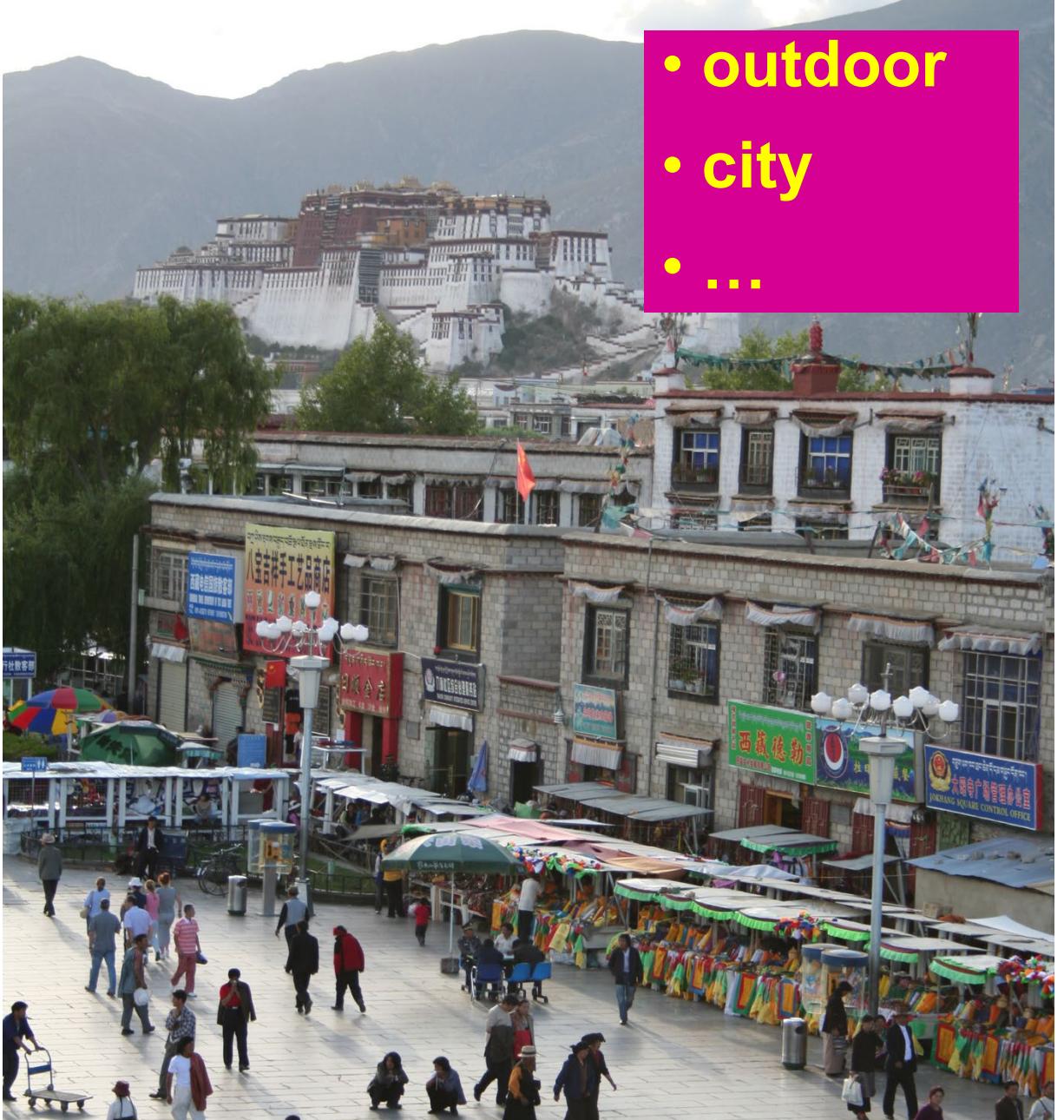
streetlamp

vendor

people

What is “Recognition”?

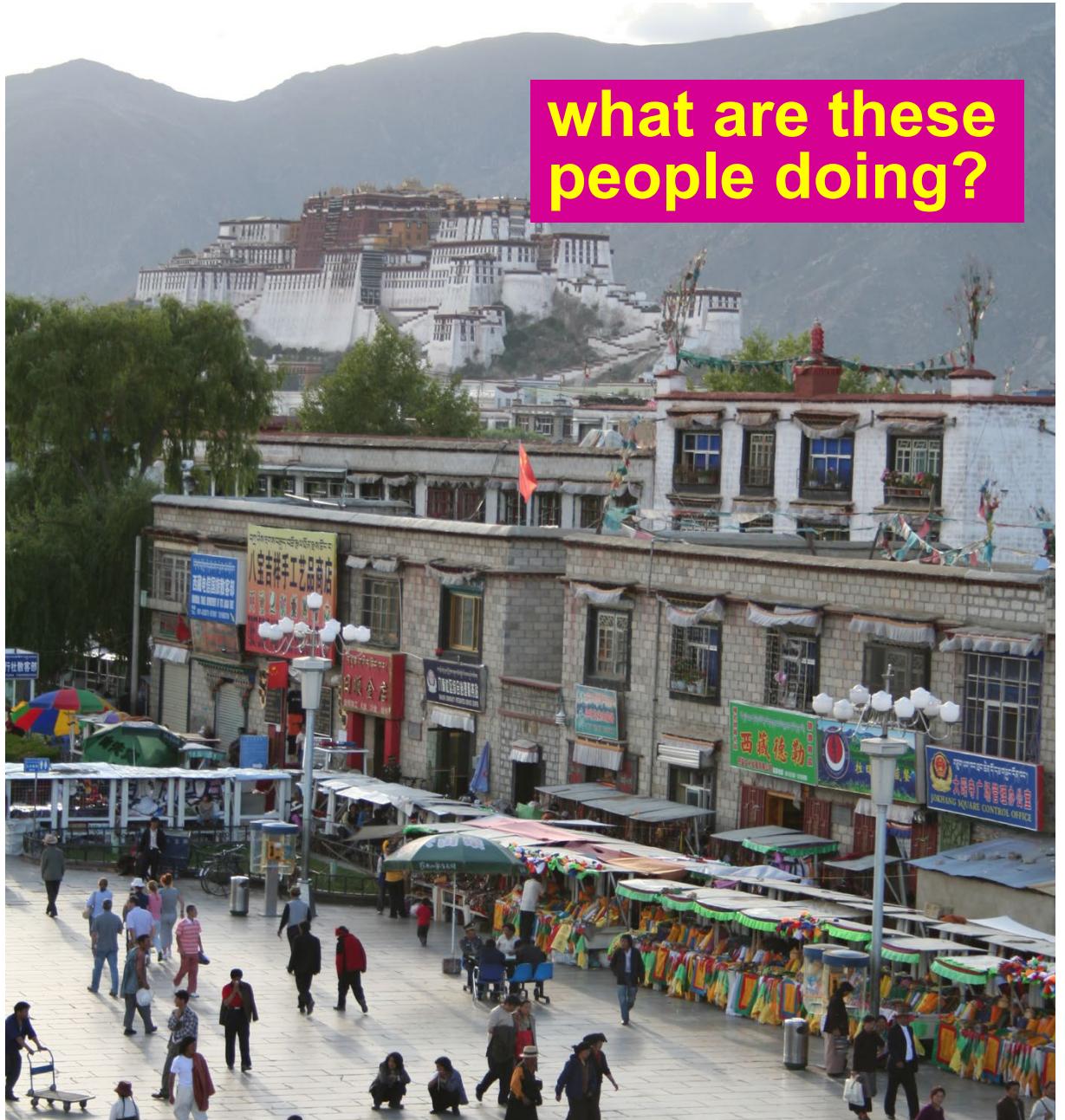
- Verification: is that a lamp?
- Detection: where are the people?
- Identification: is that Potala Palace?
- Object categorization
- Scene and context categorization



- outdoor
- city
- ...

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?

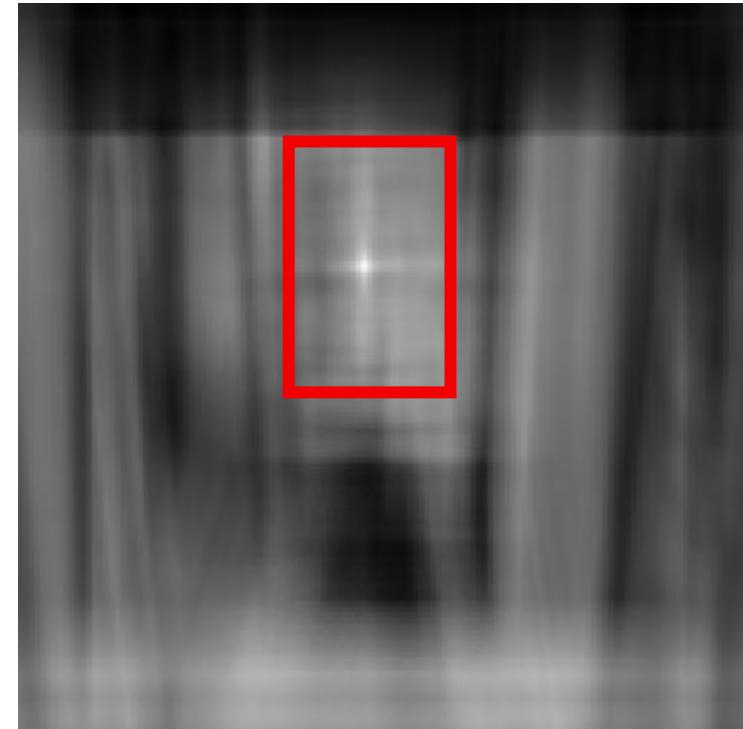
This is a chair



Find the chair in this image

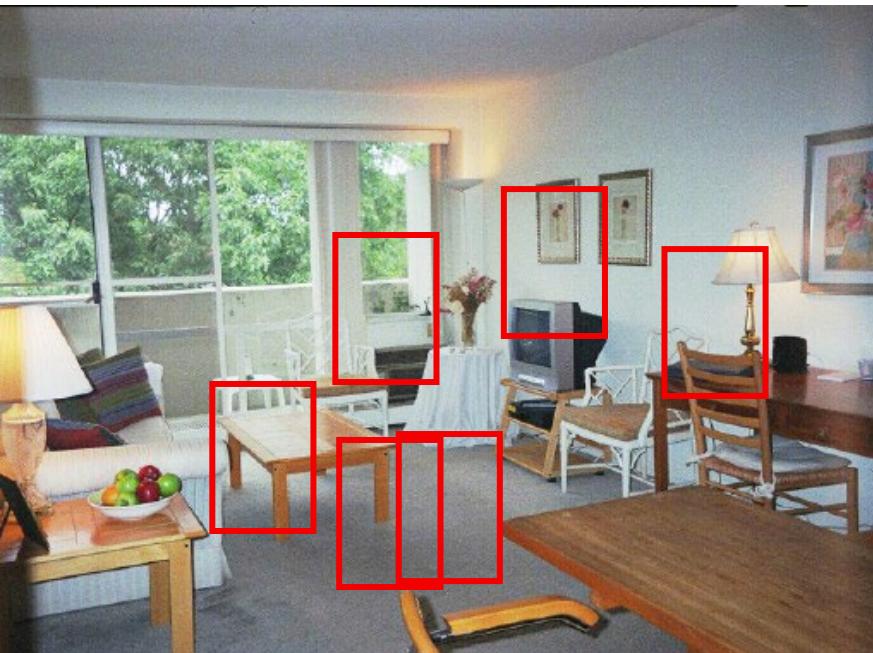


Output of normalized correlation



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*



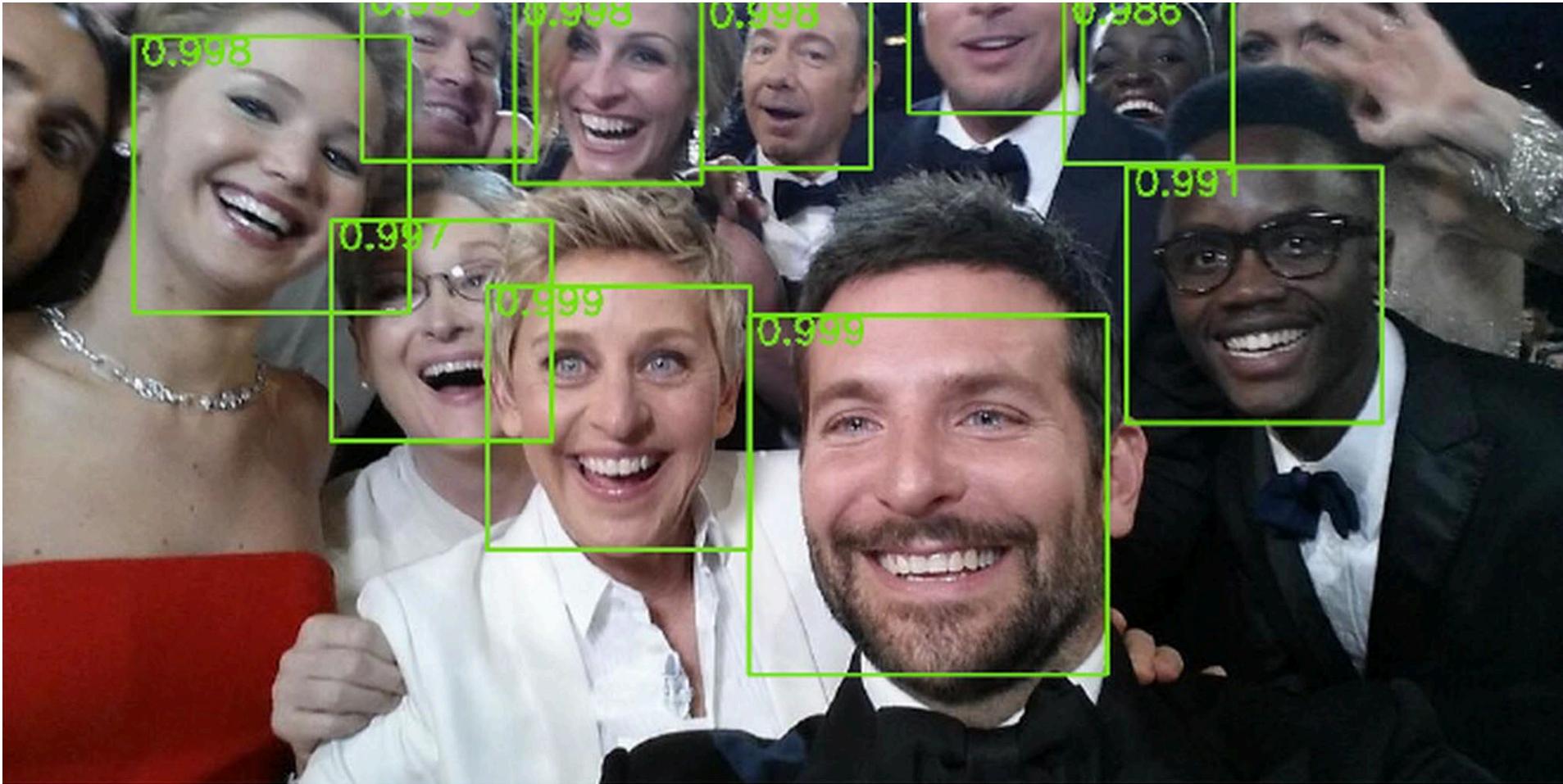
*scale-invariant feature transform (SIFT)

And it can get a lot harder

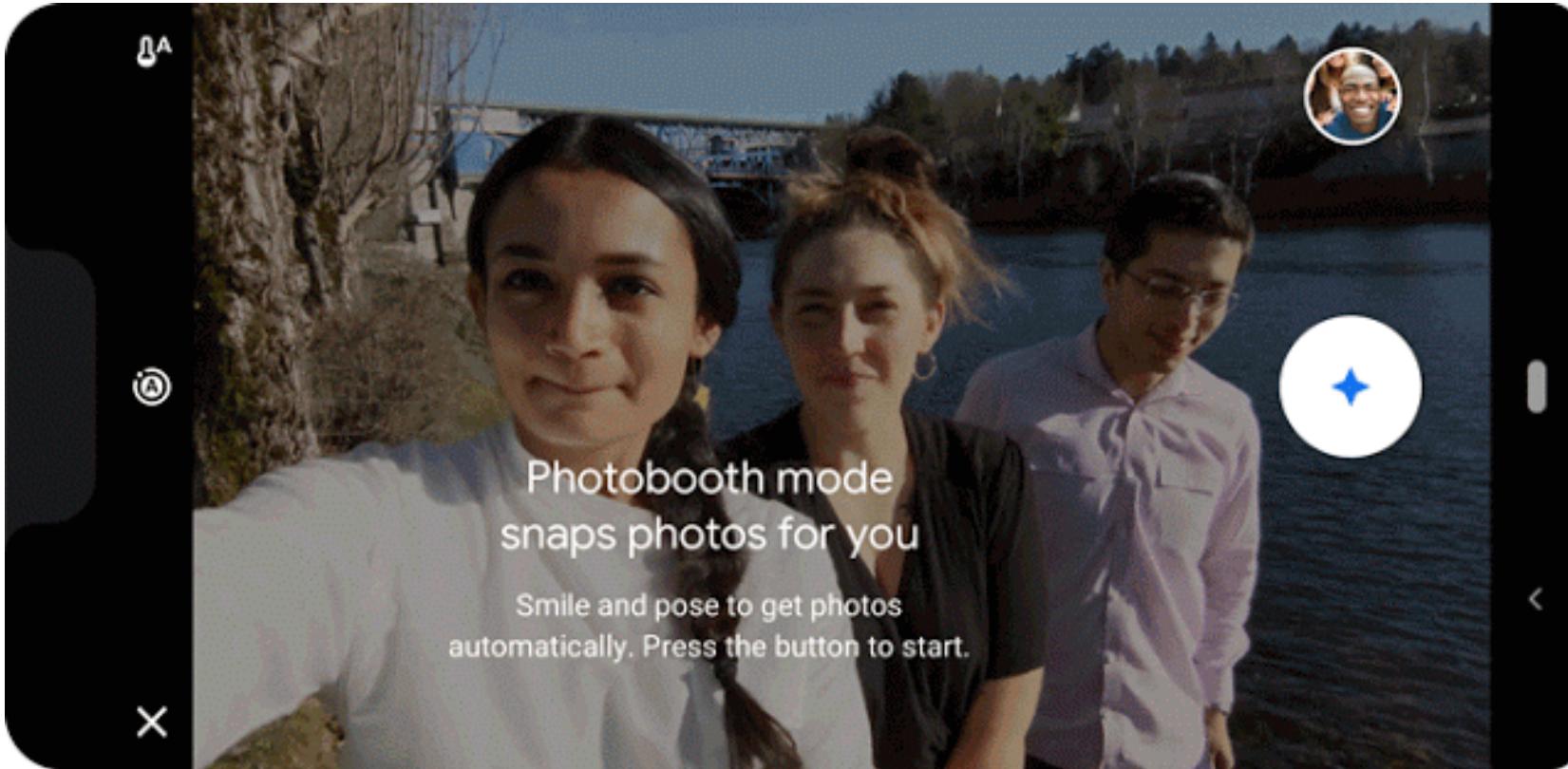


Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. *J Vis*, 3(6), 413-422

Applications: Photography



Applications: Shutter-free Photography

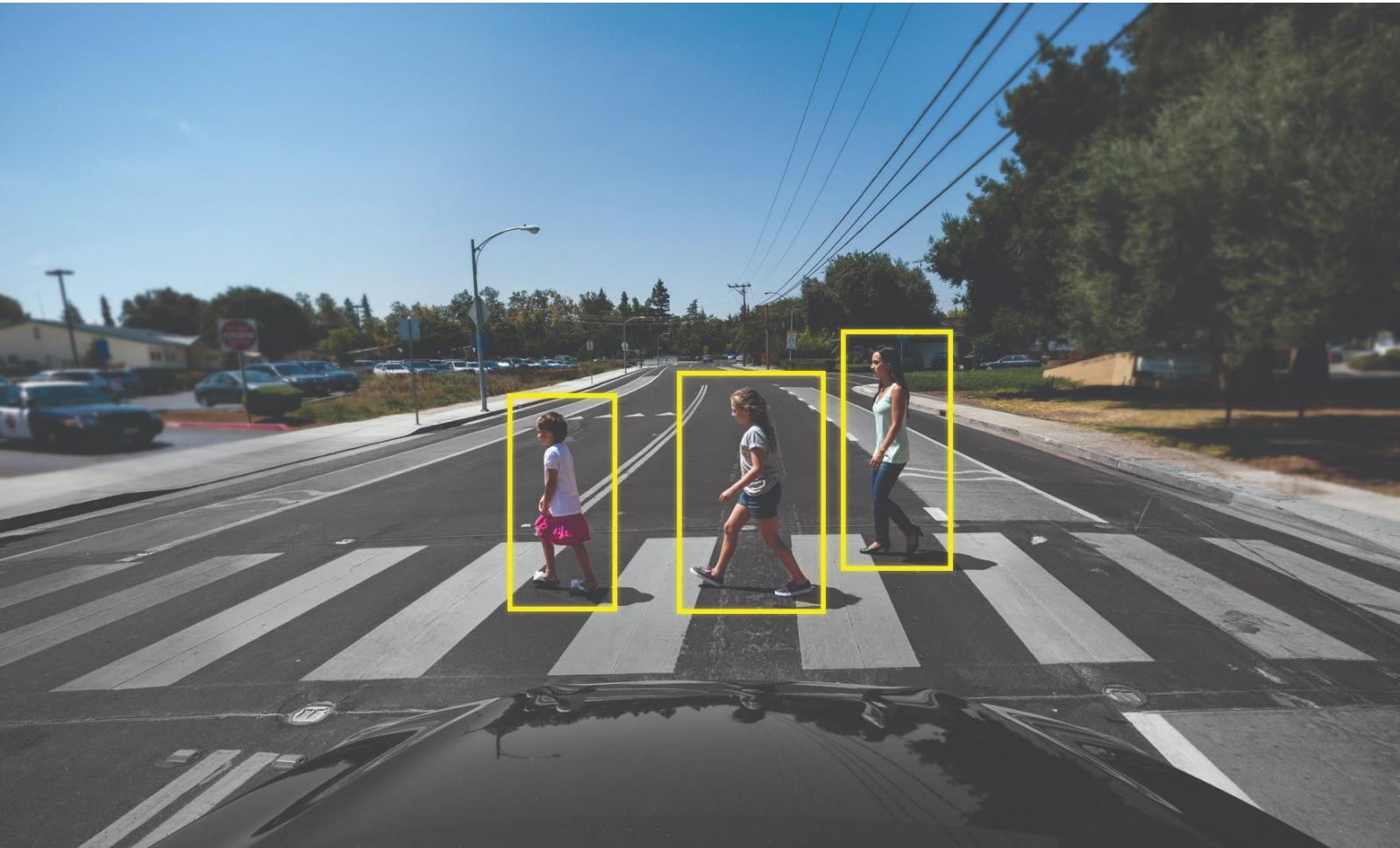


Take Your Best Selfie Automatically, with Photobooth on Pixel 3

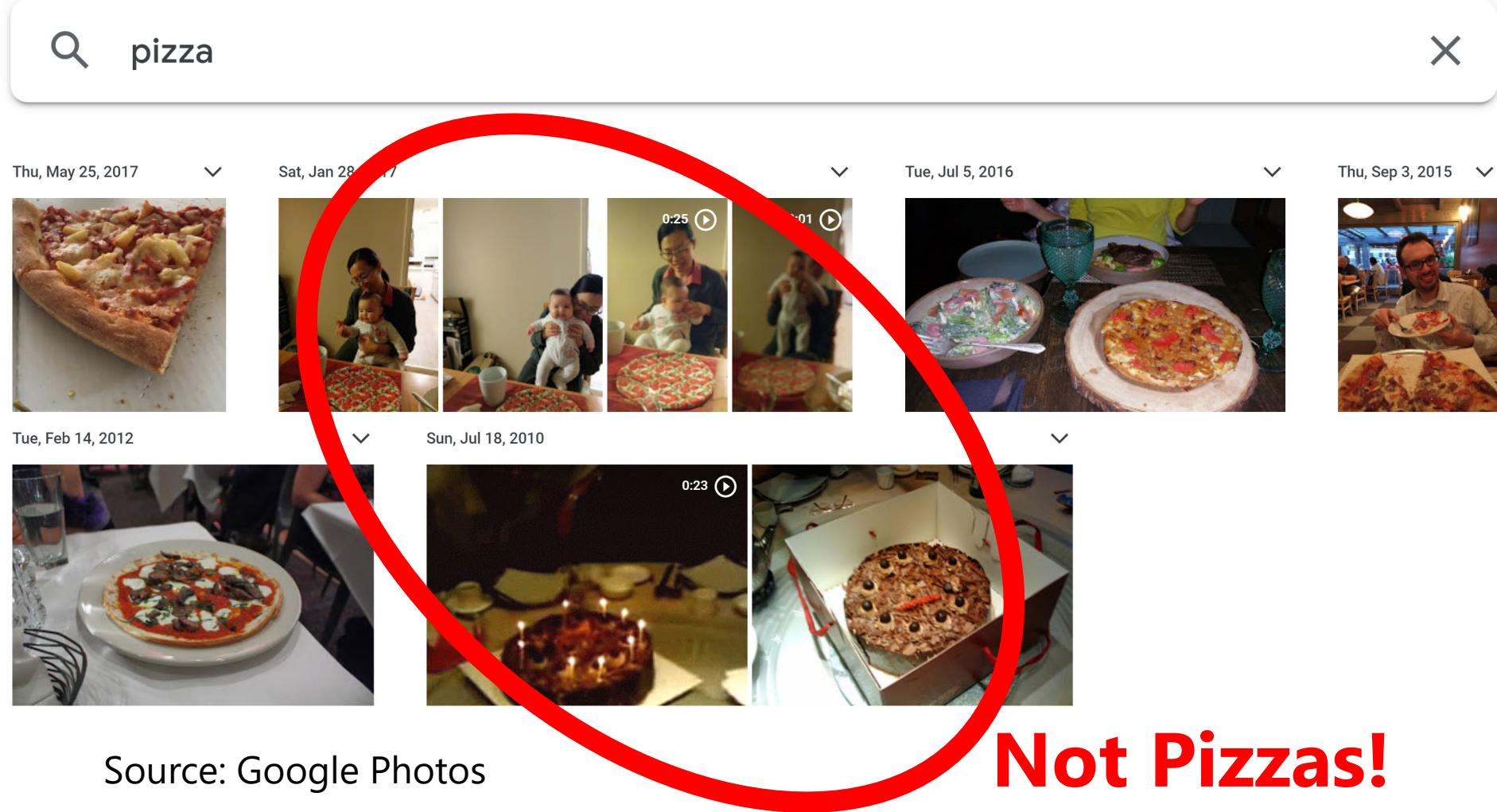
<https://ai.googleblog.com/2019/04/take-your-best-selfie-automatically.html>

(Also features "kiss detection")

Applications: Assisted / autonomous driving



Applications: Photo organization

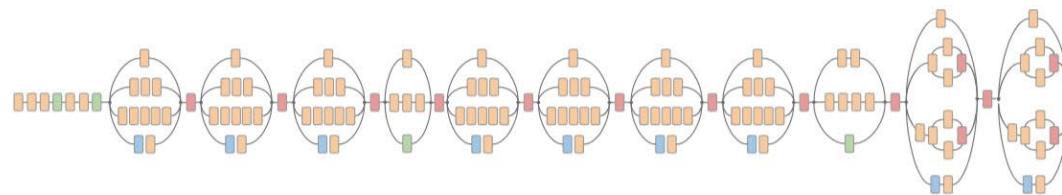


Applications: medical imaging

Skin lesion image



Deep convolutional neural network (Inception v3)

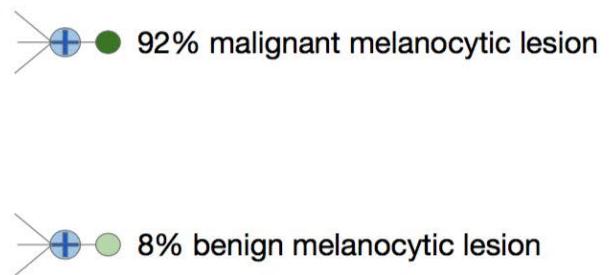


- Convolution
- AvgPool
- MaxPool
- Concat
- Dropout
- Fully connected
- Softmax

Training classes (757)

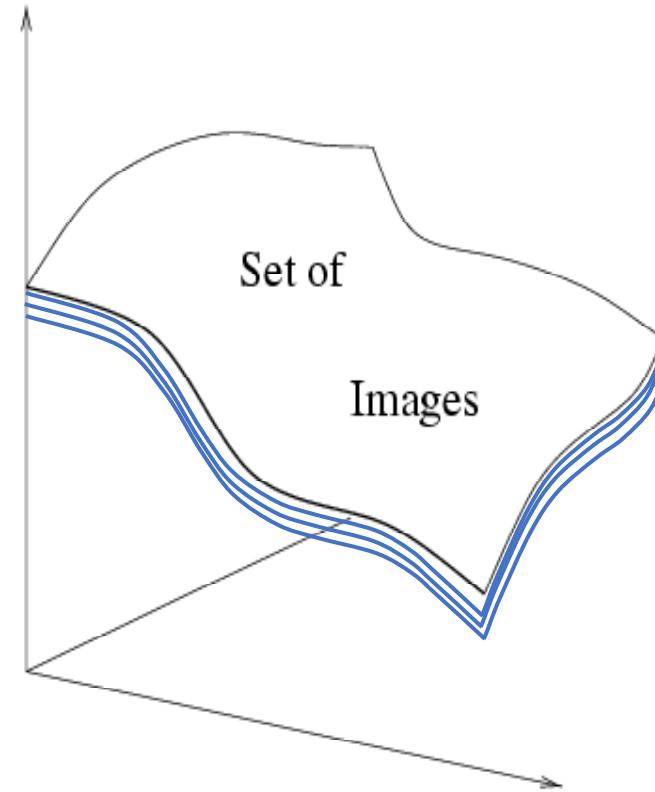
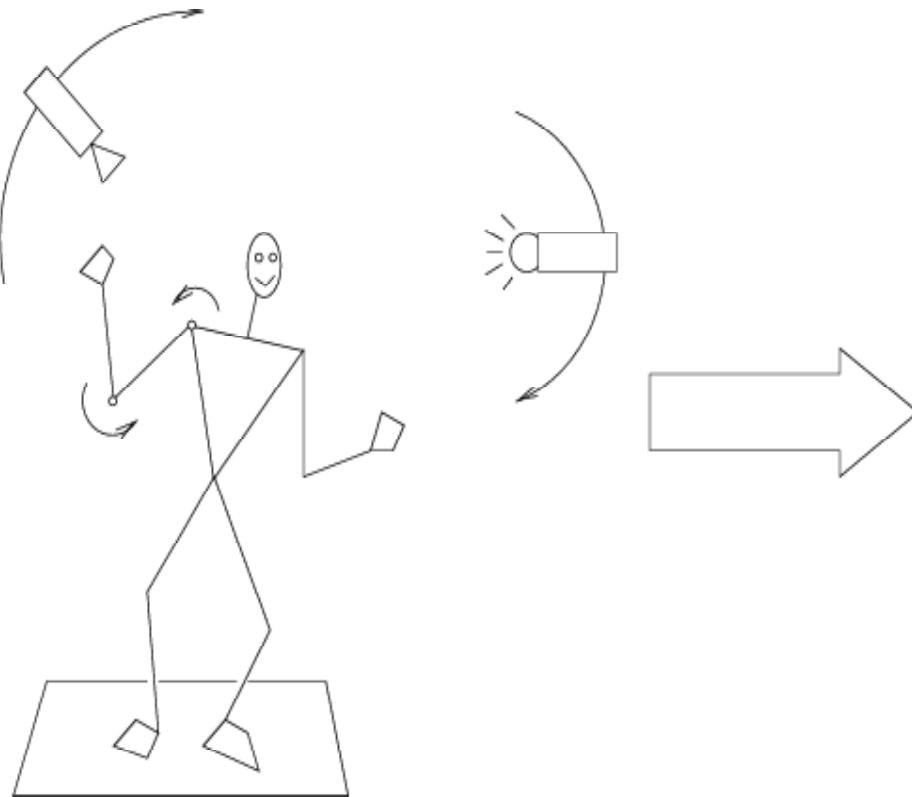
- Acral-lentiginous melanoma
- Amelanotic melanoma
- Lentigo melanoma
- ...
- Blue nevus
- Halo nevus
- Mongolian spot
- ...
- ...
- ...

Inference classes (varies by task)



Dermatologist-level classification of skin cancer
<https://cs.stanford.edu/people/esteva/nature/>

Why is recognition hard?

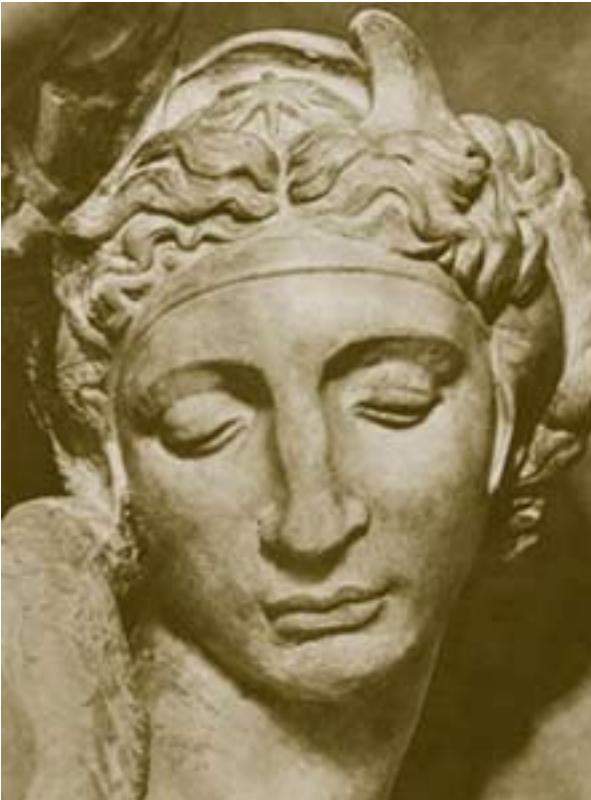


Variability: Camera position,
Illumination,
Shape,
etc...

Challenge: lots of potential classes



Challenge: variable viewpoint



Michelangelo 1475-1564

Challenge: variable illumination

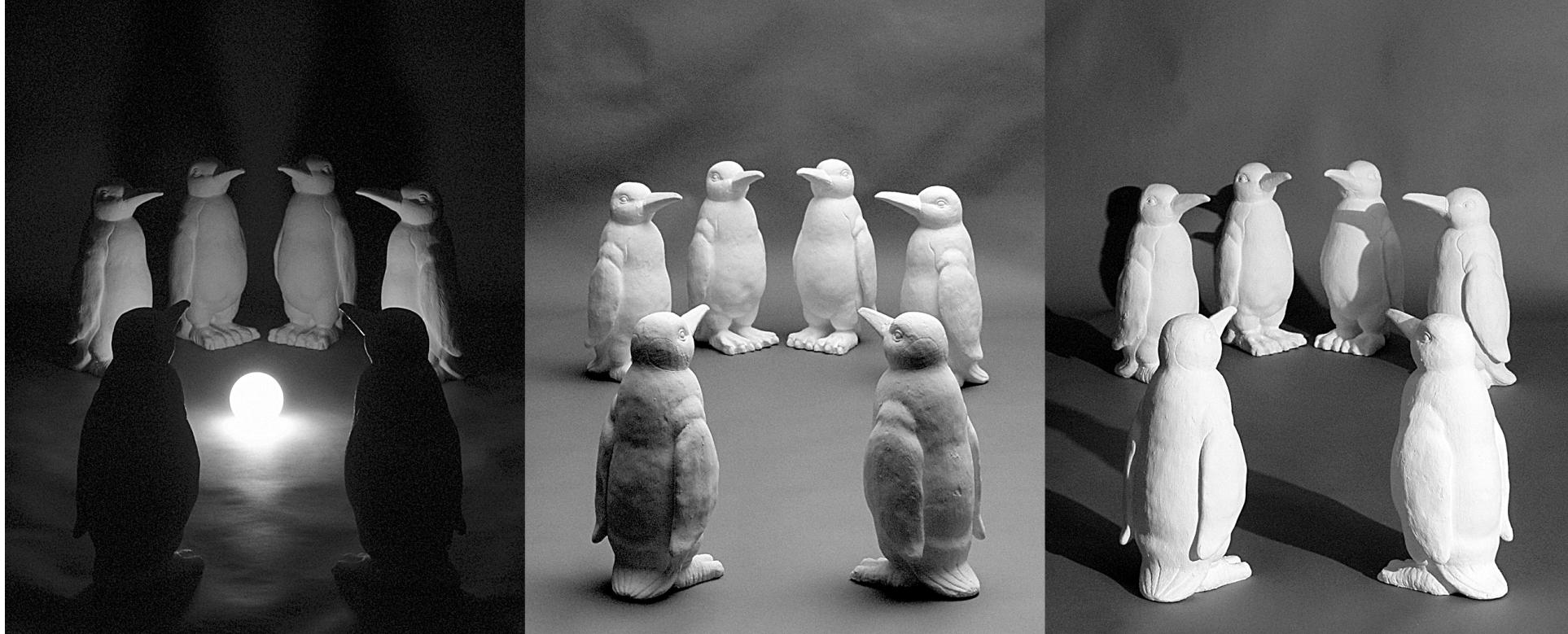


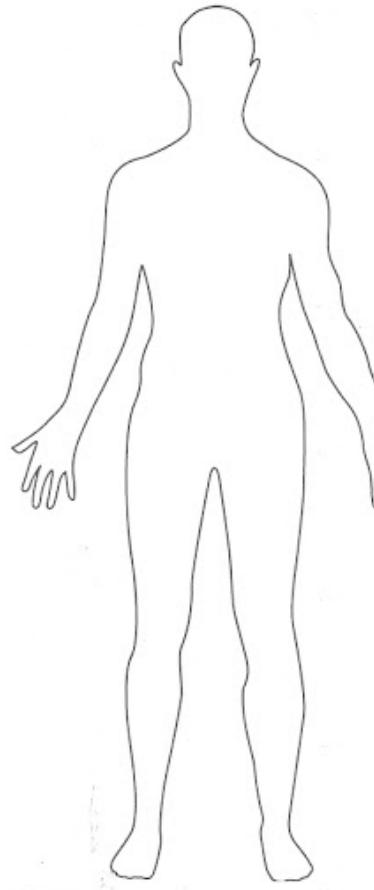
image credit: J. Koenderink

Challenge: scale

and small things
from Apple.
(Actual size)



Challenge: deformation



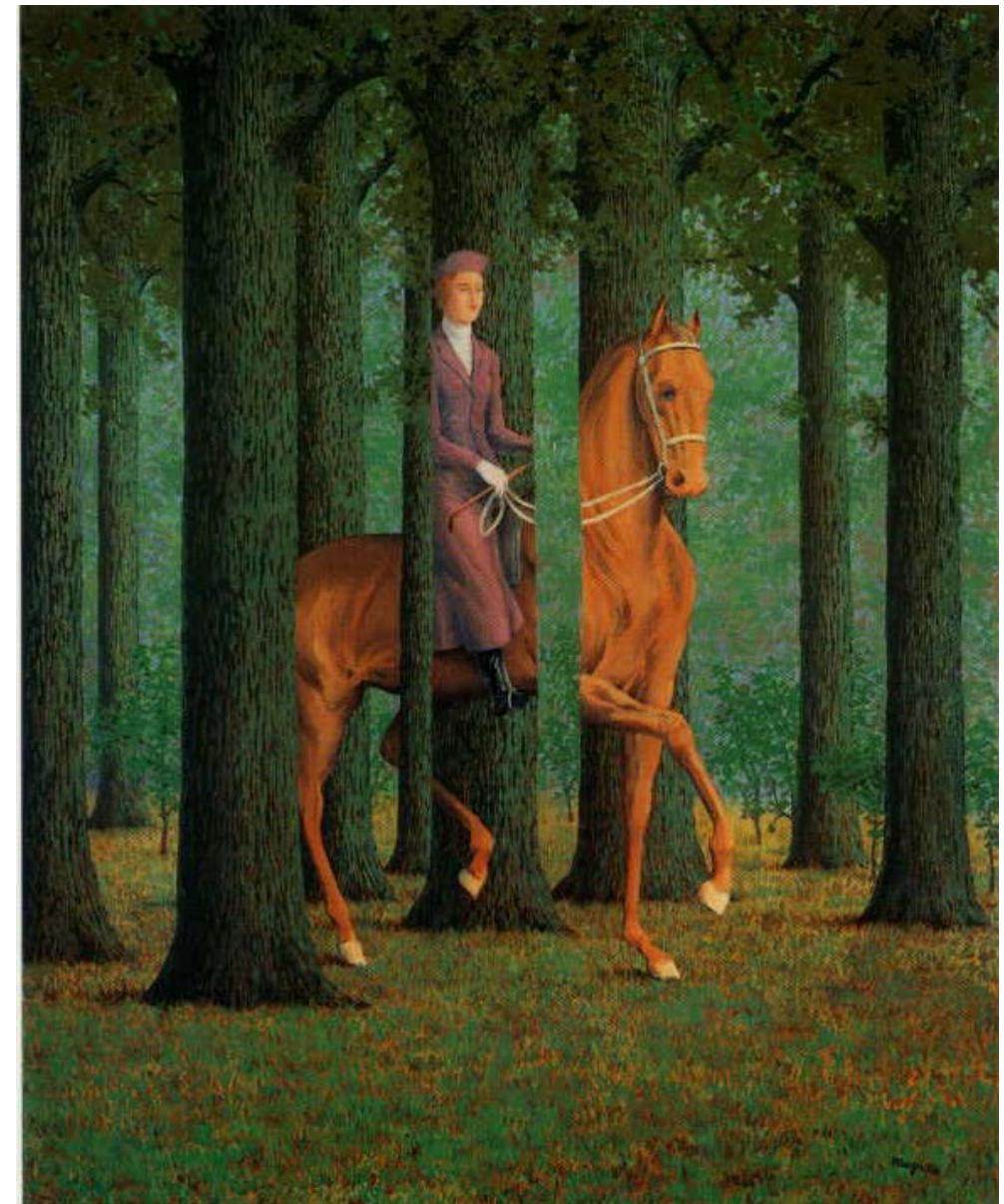
Punch Layout
© Corbis/Getty Images

Arabian
Double Front

Roundoff,
Back Handspring

Roundoff,
One-and-a-half Stepout

Challenge: Occlusion



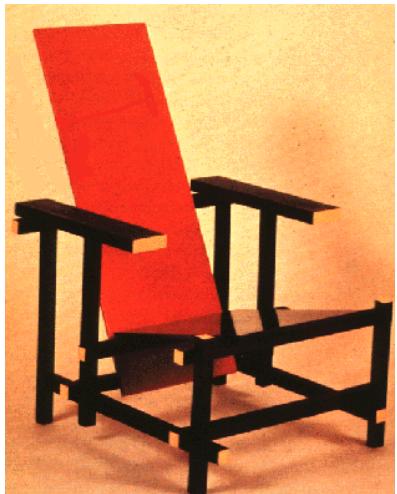
Magritte, 1957

Challenge: background clutter



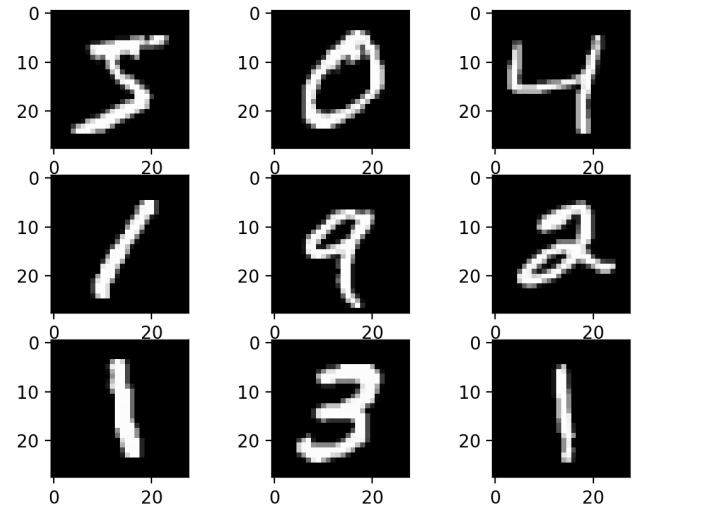
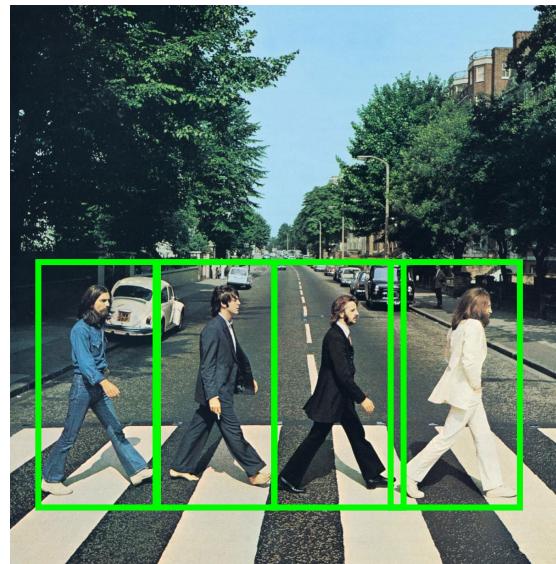
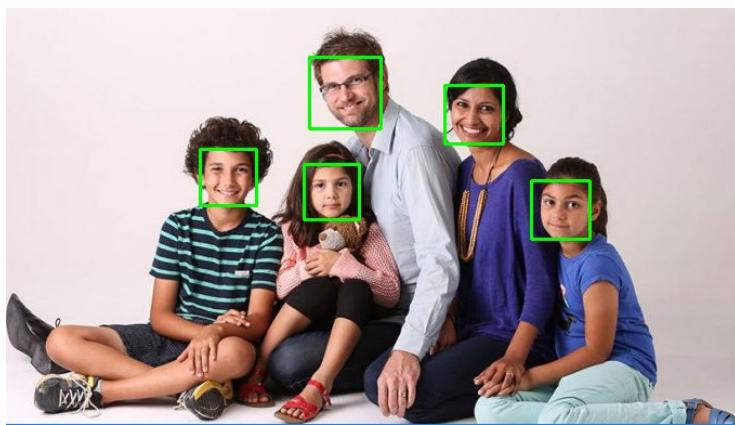
Kilmeny Niland.
1995

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

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

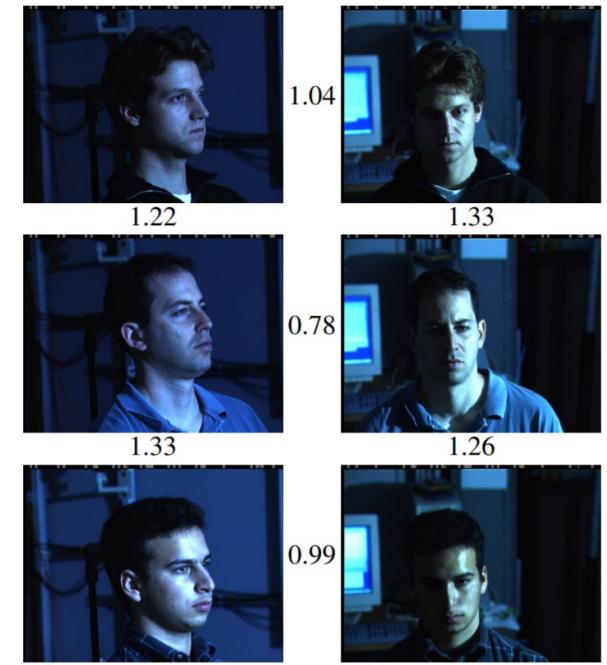


Figure 1. **Illumination and Pose invariance.** Pose and illumination have been a long standing problem in face recognition. This figure shows the output distances of FaceNet between pairs of faces of the same and a different person in different pose and illumination combinations. A distance of 0.0 means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



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

Data Sets

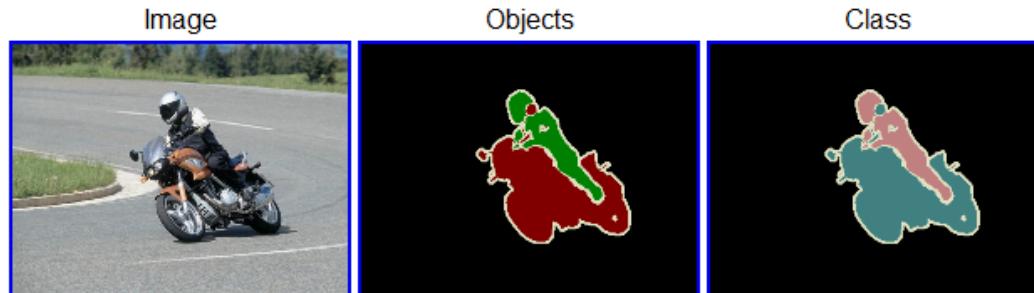
- PASCAL VOC
 - Not Crowd sourced, bounding boxes, 20 categories
- ImageNet
 - Huge, Crowd sourced, Hierarchical, *Iconic* objects
- SUN Scene Database, Places
 - Not Crowd sourced, 397 (or 720) scene categories
- LabelMe (Overlaps with SUN)
 - Sort of Crowd sourced, Segmentations, Open ended
- SUN Attribute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

Data Sets

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

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?)



Large Scale Visual Recognition Challenge (ILSVRC)

2010-2017

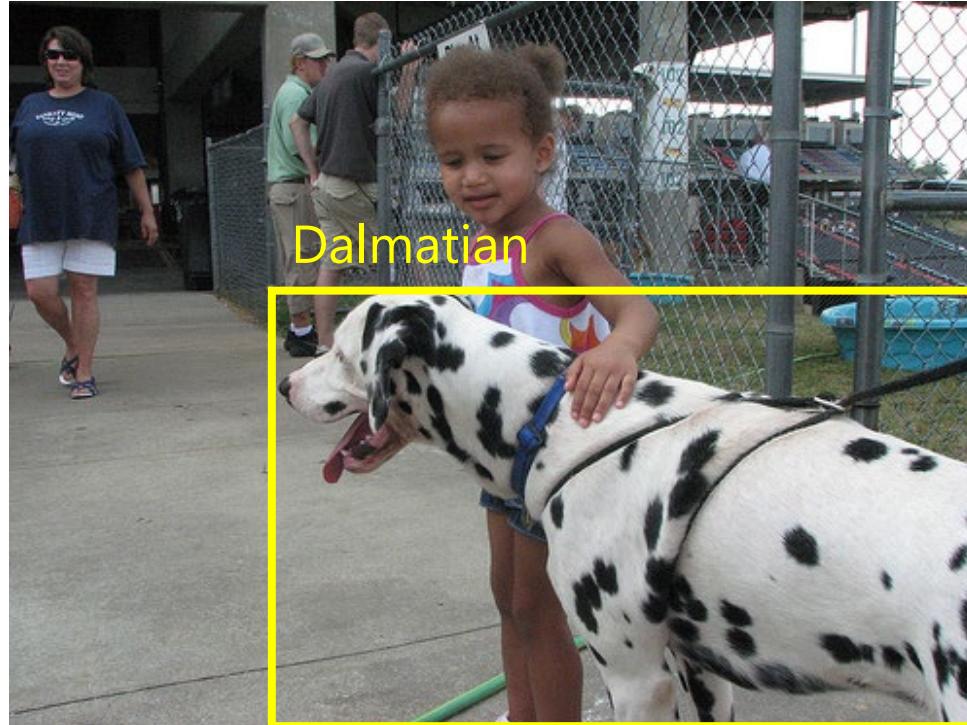


20 object classes

1000 object classes

22,591 images

1,431,167 images



Dalmatian

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

Variety of object classes in ILSVRC

PASCAL

birds



bird

bottles



bottle

cars



car

ILSVRC



flamingo



cock



ruffed grouse



quail



partridge

...



pill bottle



beer bottle



wine bottle



water bottle



pop bottle

...



race car



wagon



minivan



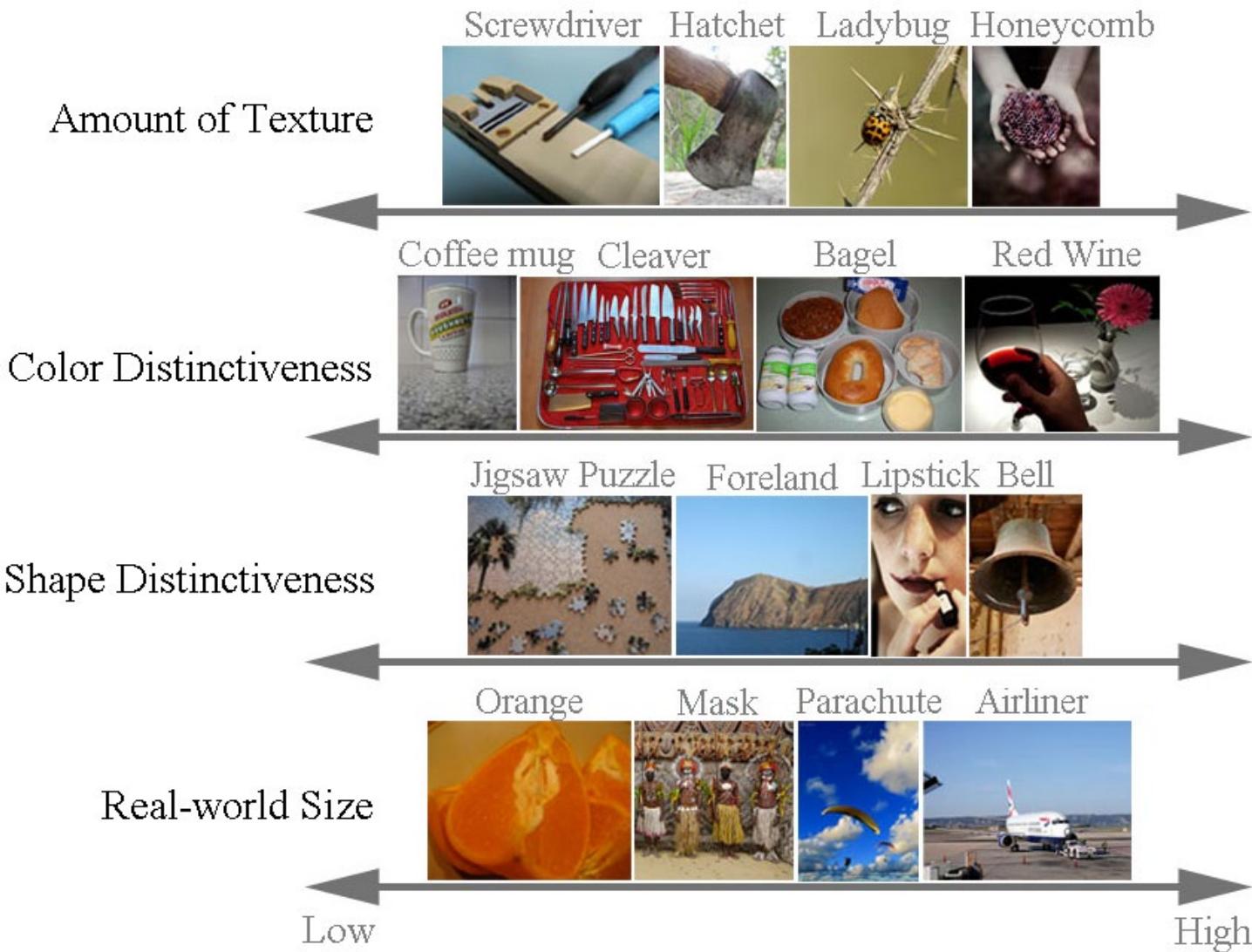
jeep



cab

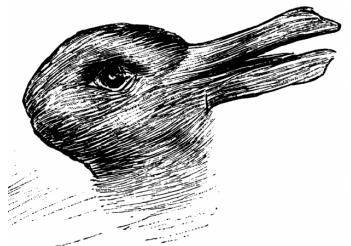
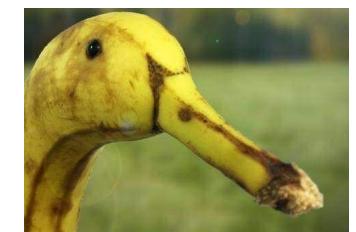
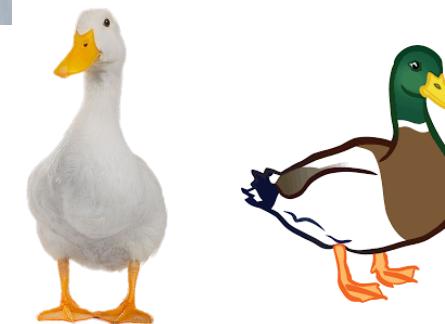
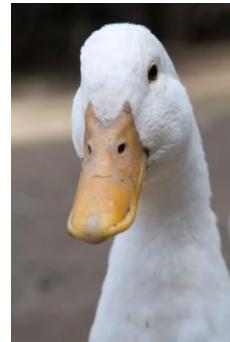
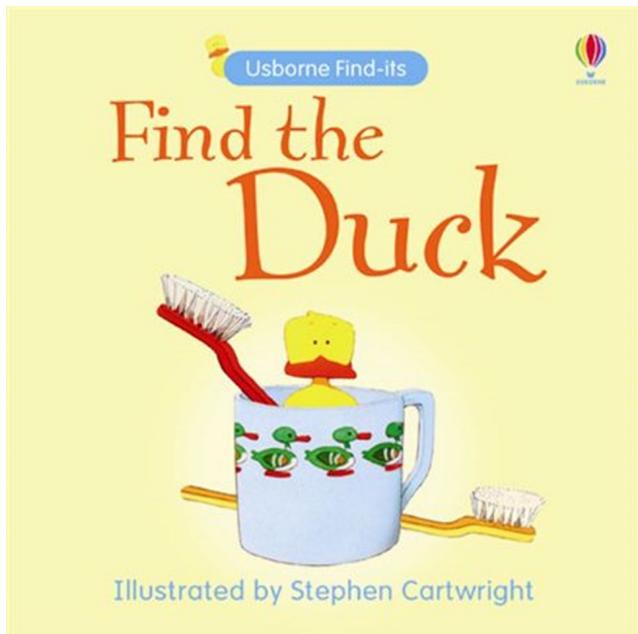
...

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

Animal->Bird->Oriole...



Baltimore Oriole



Hooded Oriole



Scott Oriole

Next

- Image classification
 - Training, validation, testing
 - Nearest neighbor classification
 - Linear classification
-
- Building up to CNNs for learning