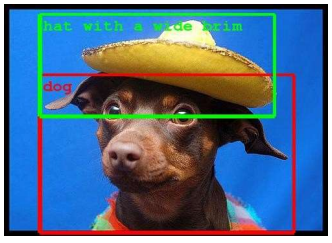
PITT | SWANSON ENGINEERING
ELECTRICAL & COMPUTER

ECE 1390/2390 Image Processing and Computer Vision – Fall 2021

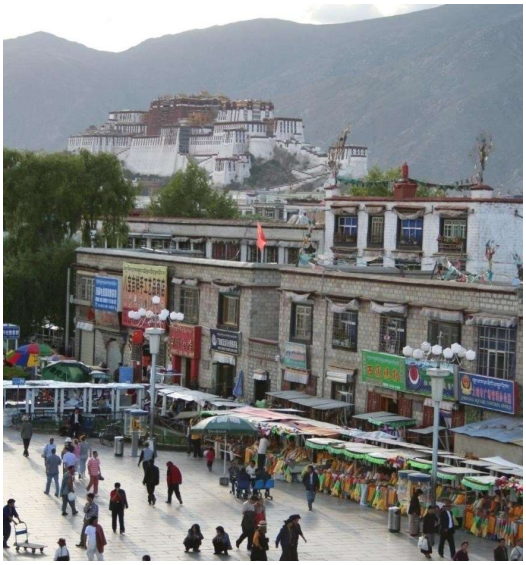
Introduction to recognition

Ahmed Dallal

Introduction to recognition

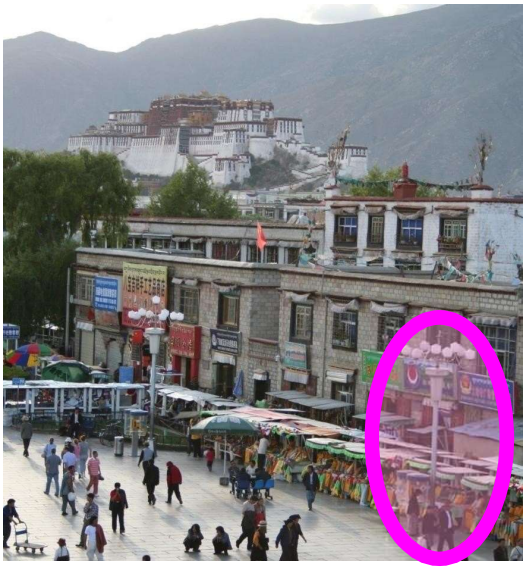


What does recognition involve?

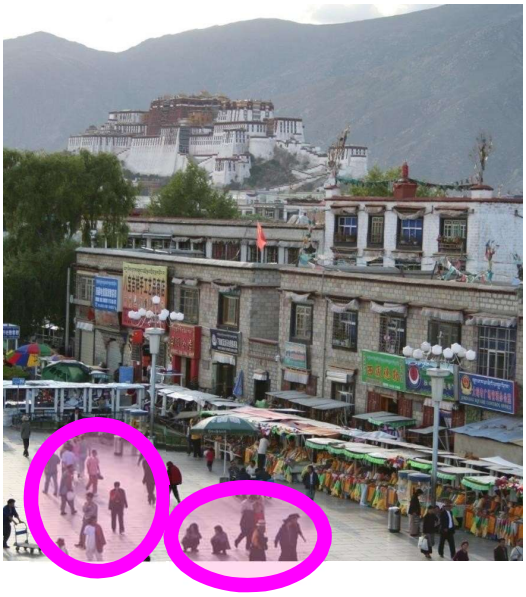


Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

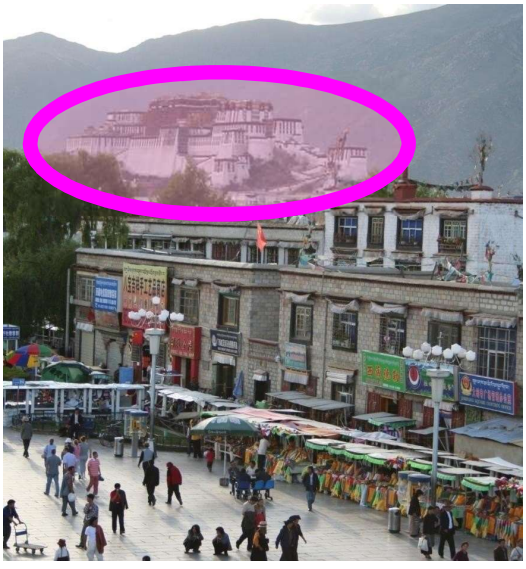
Verification: is that a lamp?



Detection: are there people?

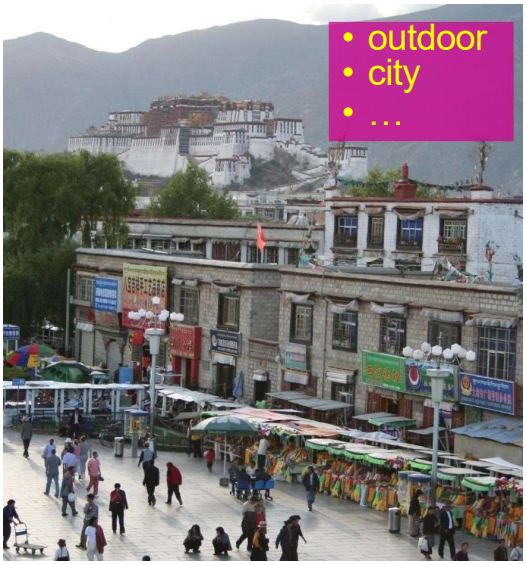


Identification: is that Potala Palace?



Object categorization

Scene and context categorization



Instance-level recognition problem

Generic categorization problem





Object Categorization

Task: *Given a (small) number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label.*

K. Grauman, B. Leibe

Object Categorization

Which categories are the best for visual identification?

Which should we use? Dog - most specific that is still very well known

Visual Object Categories

Basic Level Categories in human categorization
[Rosch 76, Lakoff 87]

- The highest level at which category members have similar perceived shape i.e. dog
- The highest level at which a single mental image reflects the entire category

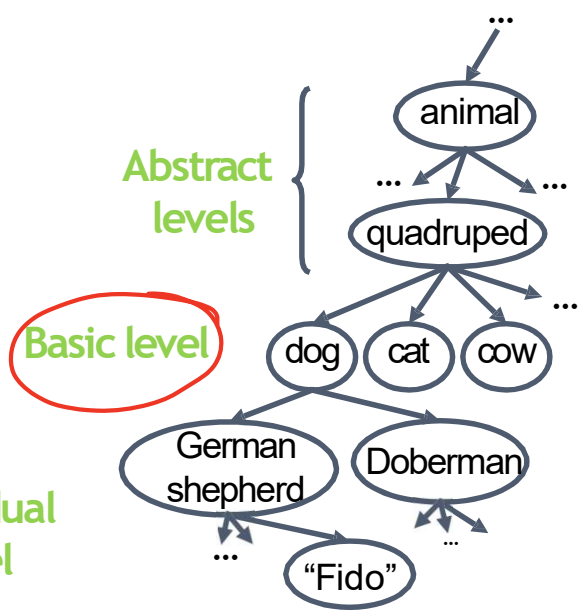
Visual Object Categories

Basic Level Categories in human categorization
[Rosch 76, Lakoff 87]

- The level at which human subjects are usually fastest at identifying category members
- The first level named and understood by children
- The highest level at which a person uses similar motor actions for interaction with category members

Visual Object Categories

- Basic-level categories in humans seem to be defined predominantly visually.
- There is evidence that humans (usually) start with basic-level categorization *before* doing identification.



Object Categorization

Which categories are the best for visual identification?

Other Types of Categories

Functional Categories

e.g. chairs = *"something you can sit on"*





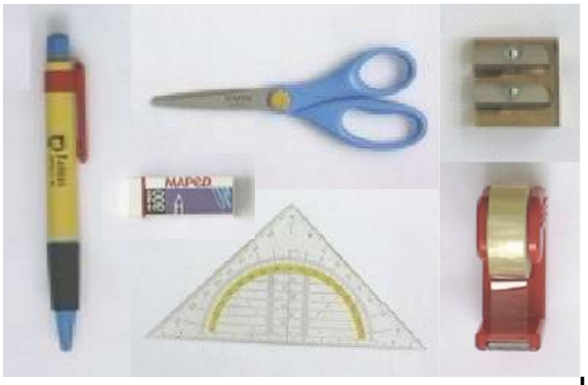



K. Grauman, B. Leibe

Other Types of Categories

Ad-hoc categories

e.g. *"something you can find in an office environment"*

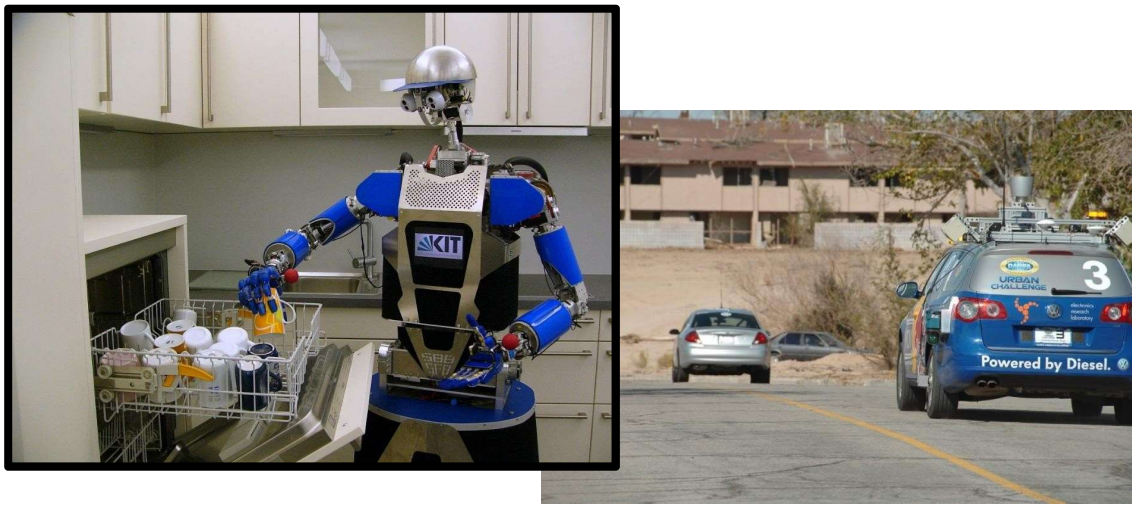


K. Grauman, B. Leibe

Words: 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
- Because it is a very human way of thinking about things…

Autonomous agents able to detect objects

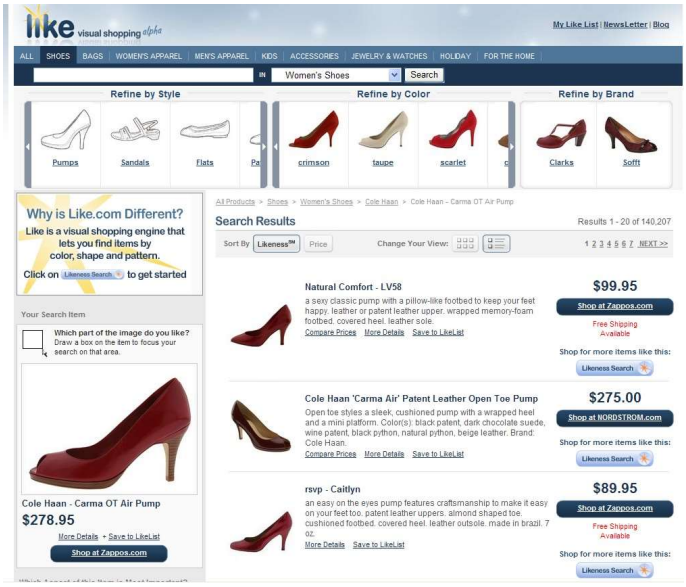


Labeling people

Posing visual queries

Belhumeur et al.

Finding visually similar objects



So why is this hard?

Challenges: Robustness

Illumination

Pixels have similar color

Object pose

Different poses

Clutter

Challenges: Robustness

Occlusions

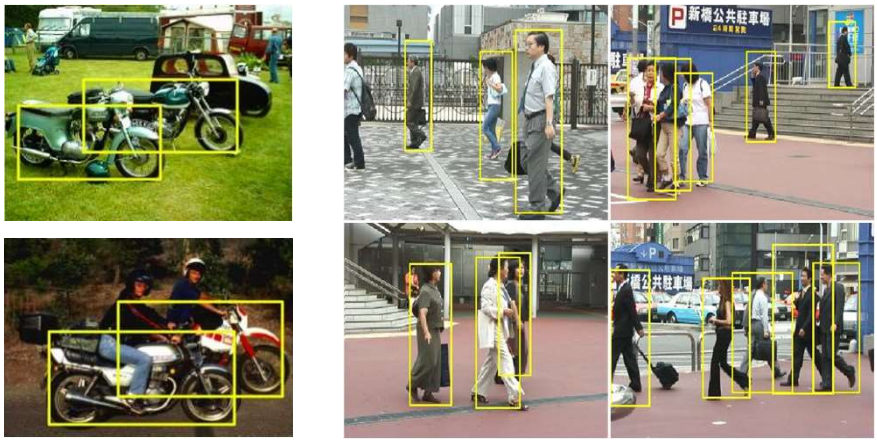
Intra-class appearance

Different breeds, for example

Viewpoint

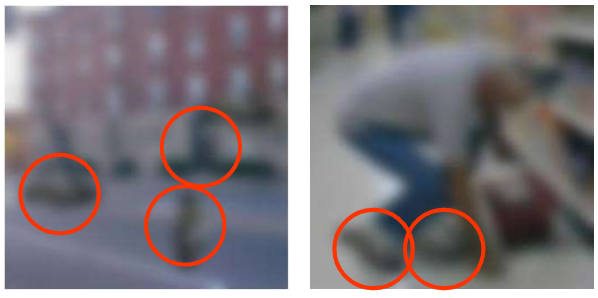
Challenges: Robustness

Realistic scenes are crowded, cluttered, have overlapping objects.



Challenges: Importance of context

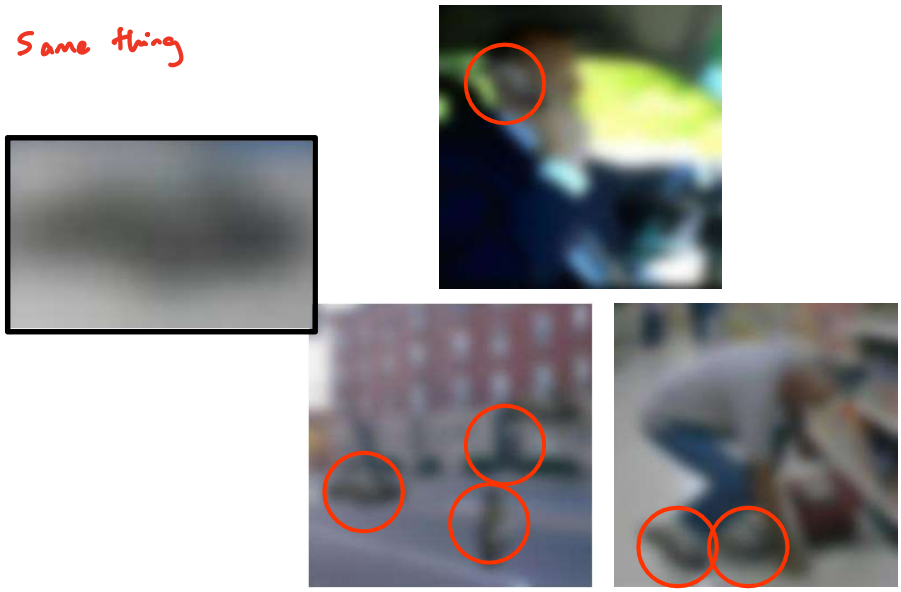
Can't tell which image this belongs to - need a window



Fei-Fei, Fergus & Torralba

Challenges: Importance of context

Same thing



Fei-Fei,
Fergus & Torralba

Challenges: complexity

NNs - need millions of images

- Thousands to millions of pixels in an image Fast and robust
- 3,000-30,000 human recognizable object categories Focus on specific problem
- 30+ degrees of freedom in the pose of articulated objects (humans)
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991] Biologically complex → computationally complex

Kristen Grauman

So what works?

What worked most reliably

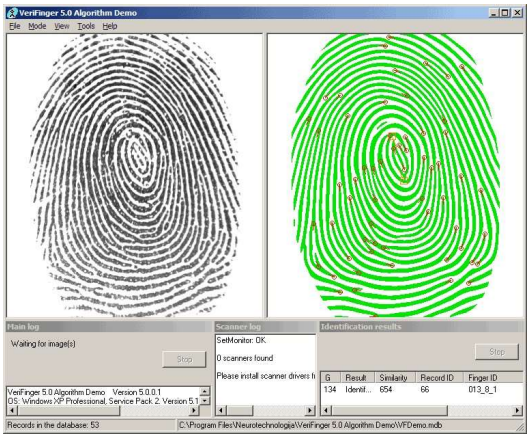
EZ Pass

- Reading license plates (real easy), zip codes, checks

Lana Lazebnik

What worked most reliably

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

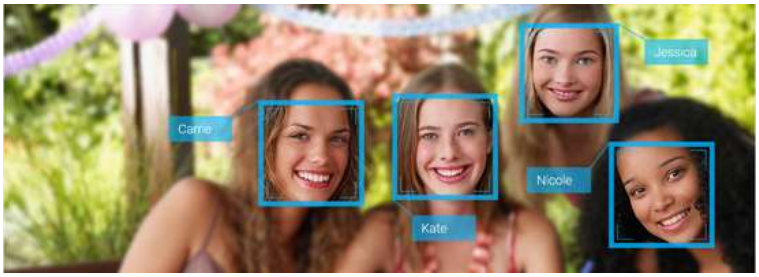


Lana Lazebnik

What worked most reliably

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

(Today recognition)



What worked most reliably

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

(Today recognition)

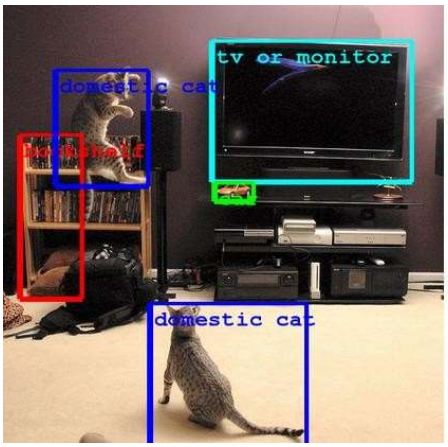
SIFT →

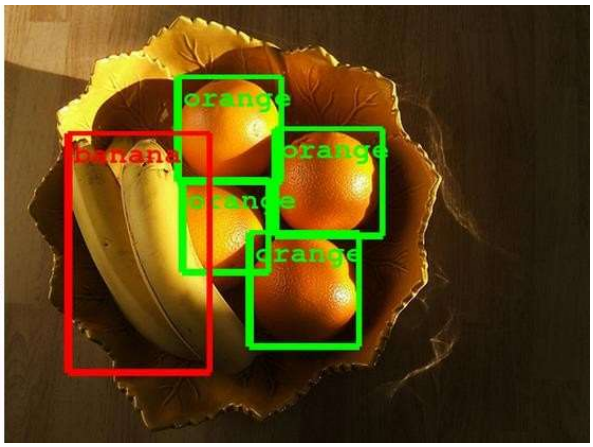
- Recognition of flat textured objects

(CD covers, book covers, etc.)

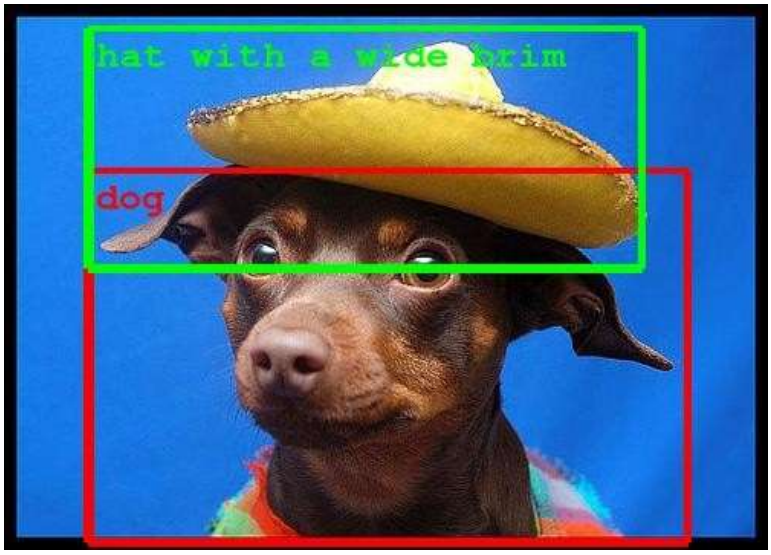
Lana Lazebnik

GoogleNet2014 (NN for objects without context)





GoogleNet—no context needed?



Going forward

- Much of strong label "recognition" is really machine learning applied to patterns of pixel intensities. (image)
- We'll focus on some principles of generative vs discriminative methods and image representations that they use.
- Then we'll spend some time on *activity recognition* from video (time permits).