CS5670: Computer Vision

Introduction to Recognition

THAT
Is a duck.
Announcements

• One more project to go – Project 5: Neural Radiance Fields
  – Tentative release date: Thursday, April 20
  – Tentative due date: Wednesday, May 3

• In-class Final Exam during the last lecture: Tuesday, May 9
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

what are these people doing?
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
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

Applications: Photography
Applications: Shutter-free Photography

Take Your Best Selfie Automatically, with Photobooth on Pixel 3

(Also features “kiss detection”)

Photobooth mode
snaps photos for you
Smile and pose to get photos automatically. Press the button to start.
Applications: Assisted / autonomous driving
Applications: Photo organization

Source: Google Photos

Not Pizzas!
Applications: medical imaging

Skin lesion image

Deep convolutional neural network (Inception v3)

Training classes (757)
- Acral-lentiginous melanoma
- Amelanotic melanoma
- Lentigo melanoma
- ...

Inference classes (varies by task)
- 92% malignant melanocytic lesion
- 8% benign melanocytic lesion

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

How many object categories are there?

~10,000 to 30,000
Challenge: variable viewpoint

Michelangelo 1475-1564
Challenge: variable illumination

image credit: J. Koenderink
Challenge: scale

and small things from Apple.
(Actual size)
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

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.
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
  - 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 Crowdsourced, bounding boxes, 20 categories
- **ImageNet**
  - Huge, Crowdsourced, Hierarchical, *Iconic* objects
- **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
Data Sets

• PASCAL VOC
  • Not Crowdsourced, bounding boxes, 20 categories

• ImageNet
  • Huge, Crowdsourced, Hierarchical, Iconic objects

• 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
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

IMAGENET

20 object classes  22,591 images
1000 object classes  1,431,167 images

Dalmatian

Variety of object classes in ILSVRC

**PASCAL**
- birds
  - bird

**ILSVRC**
- birds
  - flamingo
  - cock
  - ruffed grouse
  - quail
  - partridge

- bottles
  - bottle
  - pill bottle
  - beer bottle
  - wine bottle
  - water bottle
  - pop bottle

- cars
  - car
  - race car
  - wagon
  - minivan
  - jeep
  - cab
Variety of object classes in ILSVRC

Amount of Texture
- Screwdriver
- Hatchet
- Ladybug
- Honeycomb

Color Distinctiveness
- Coffee mug
- Cleaver
- Bagel
- Red Wine

Shape Distinctiveness
- Jigsaw Puzzle
- Foreland
- Lipstick
- Bell

Real-world Size
- Orange
- Mask
- Parachute
- Airliner
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?
What’s Still Hard?

• Few shot learning
  • How do we generalize from only a small number of examples?
Questions?
Next Time

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

• Building up to CNNs for learning