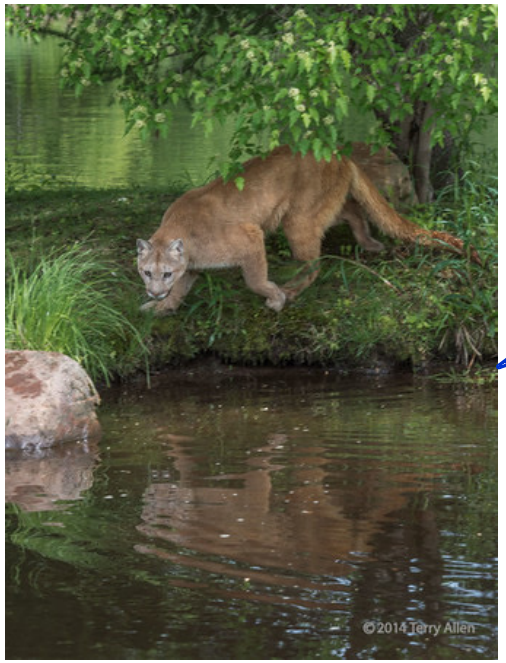
Vision

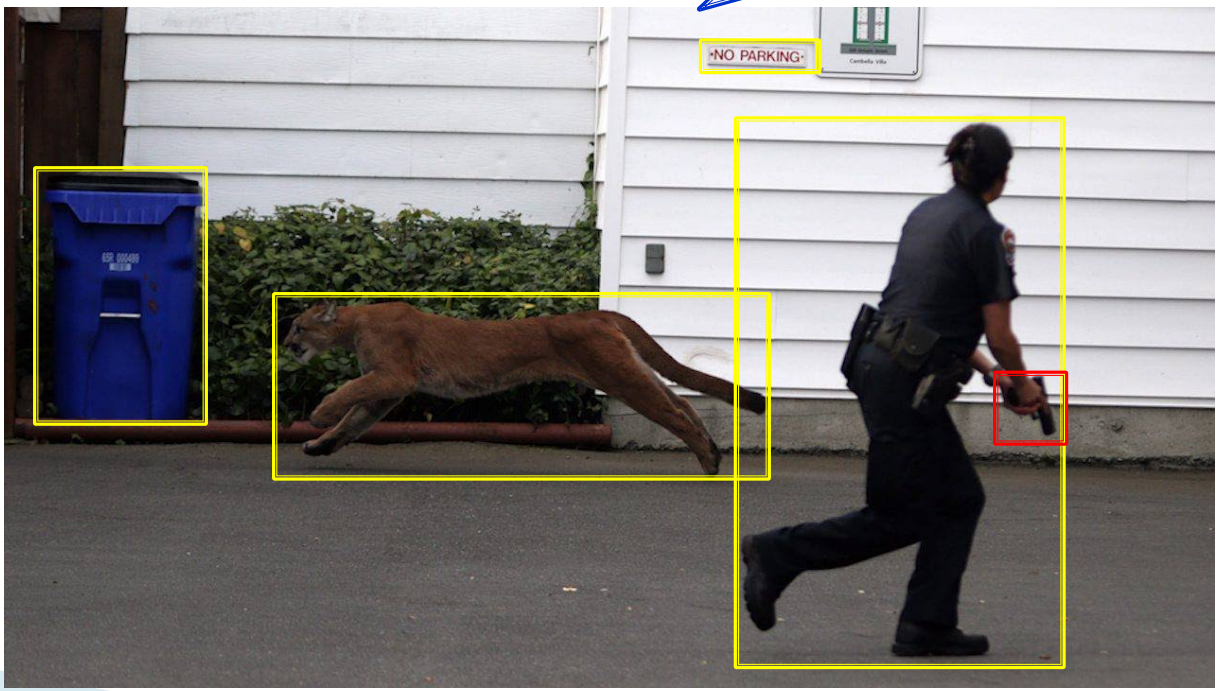
Larry Holder
School of EECS
Washington State University

Goals



- I see a cougar.
- I see two cougars?
- I see water and grass.
- Pretty tree.
- Distance to cougar: 10m.
- Should I leave?

- Animal, cat, cougar; moving left.
- Person, police, gun; moving right.
- Trash bin.
- Danger!
- Can't park here.

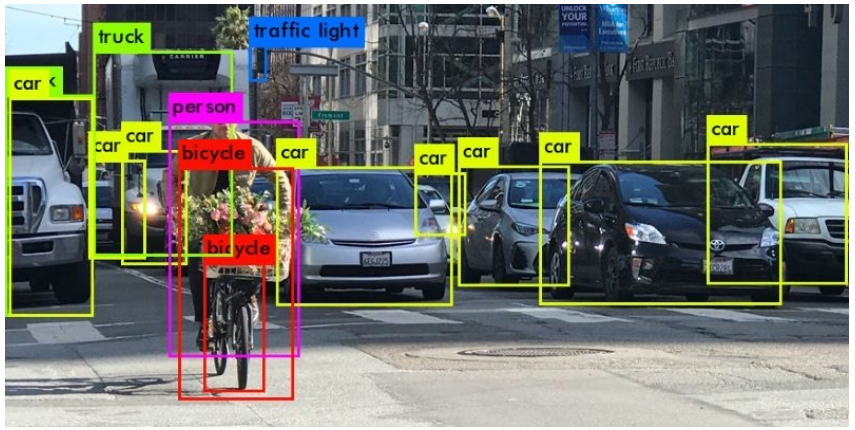


Scene Understanding

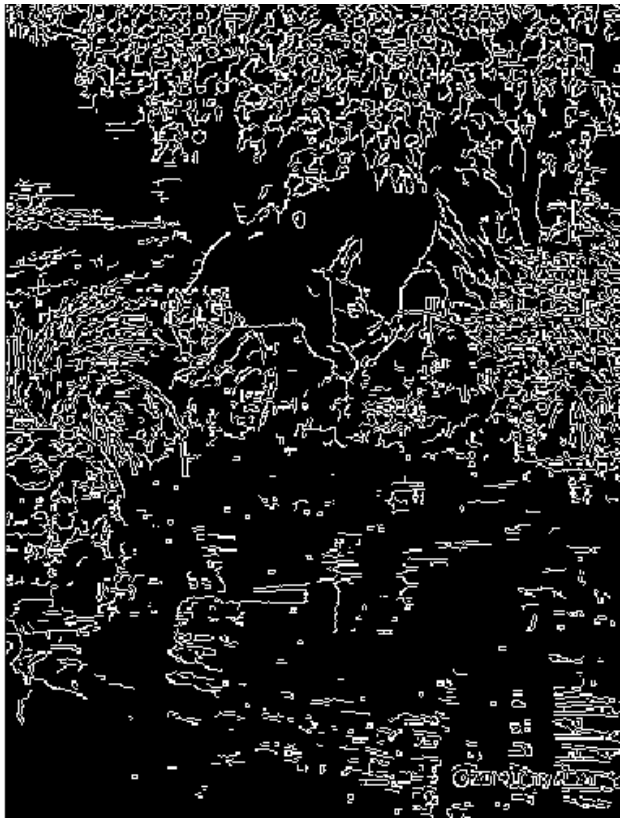
cloud.google.com/vision

Techniques

- Edge detection
- Shape detection
- Motion: optical flow, tracking
- Object detection
- Image classification
- Scene understanding



Edge Detection



Canny Edge Detector

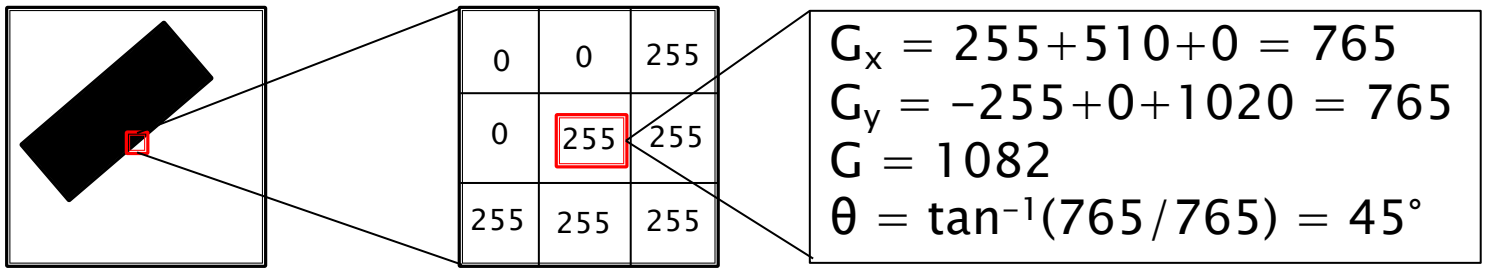
- Given image intensity I(x,y) (e.g., grayscale)
- Step 1: Smooth image
  - Gaussian $N_\sigma(x,y)$
  - Replace intensity I(x,y) with $I * N_\sigma$ (i.e., convolve)

Blur (5x5)

Blur (103x103)

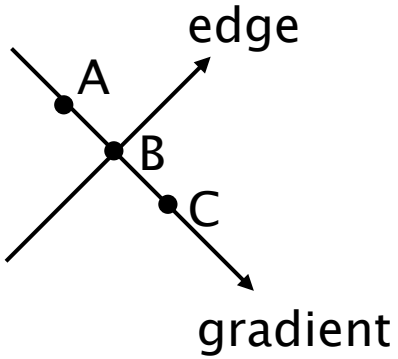
Canny Edge Detector

- Step 2: Find intensity gradient G
  - Convolve image with Sobel operator
  - $G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \mathrm{I} \qquad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * \mathrm{I}$
  - $G = \sqrt{G_x^2 + G_y^2} \qquad \theta = \tan^{-1}\left(\frac{G_y}{G_x}\right)$



Canny Edge Detector

- Step 3: Thin edges



If B < max(A,B,C)
Then B = 0

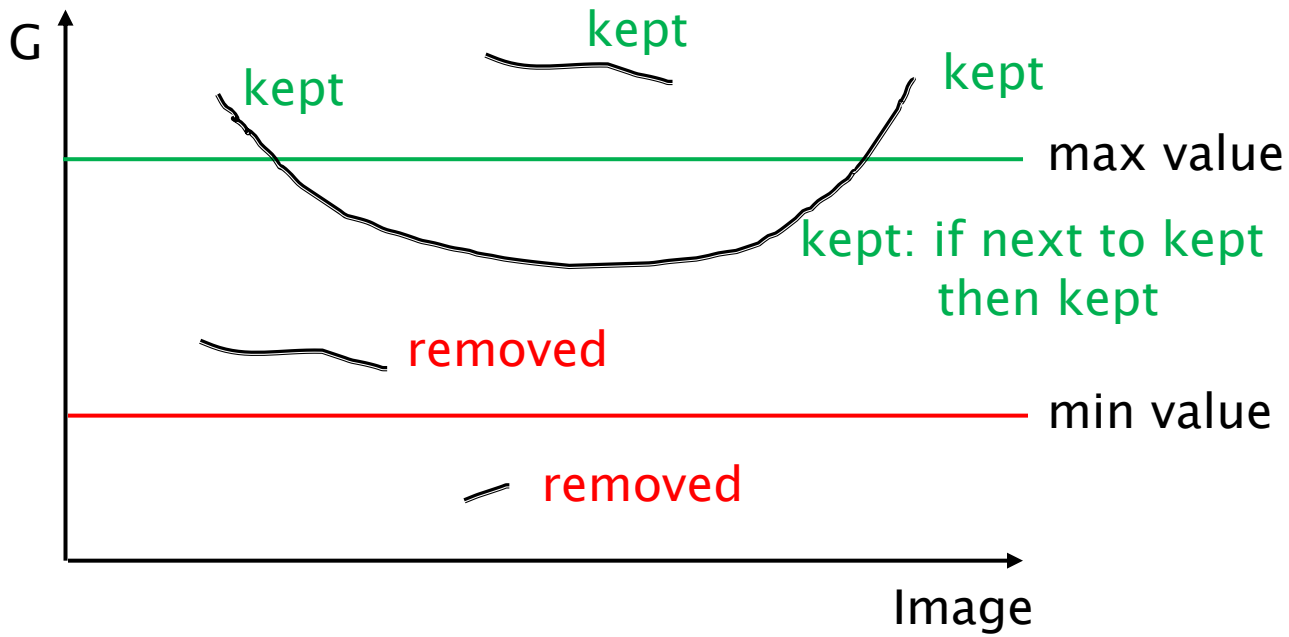
4	3	255
5	255	255
255	255	255

→

0	0	255
0	255	255
255	255	255

Canny Edge Detector

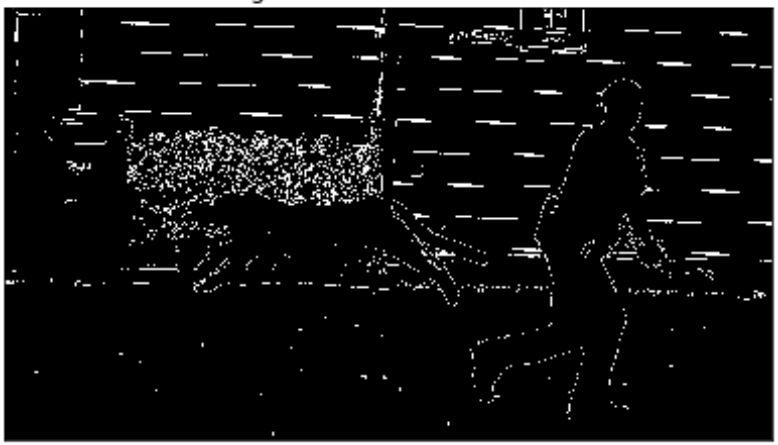
- Step 4: Apply double threshold with hysteresis



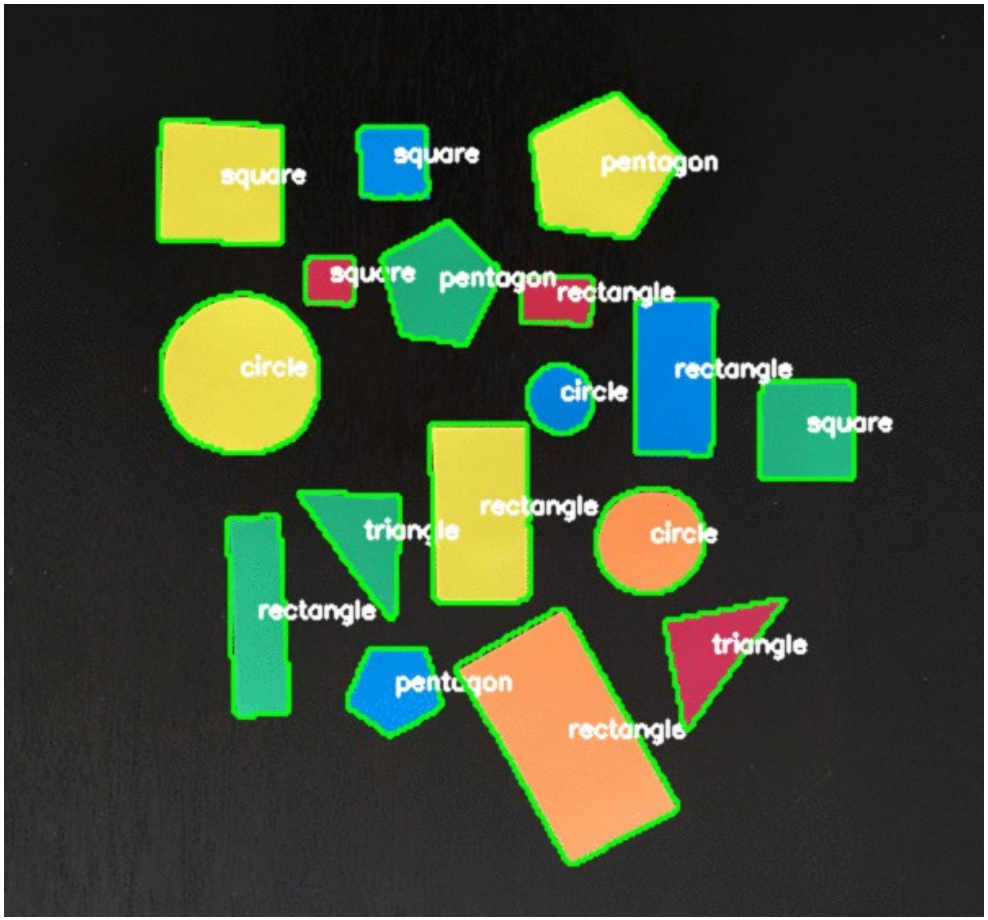
Canny Edge Detection

Original Image

Edges (min=100,max=200)

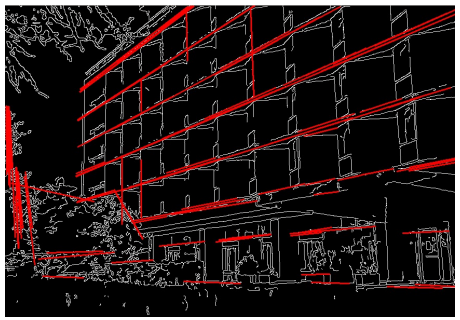


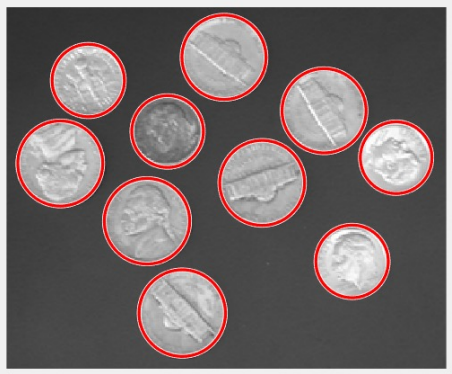
Shape Detection



Hough Transform

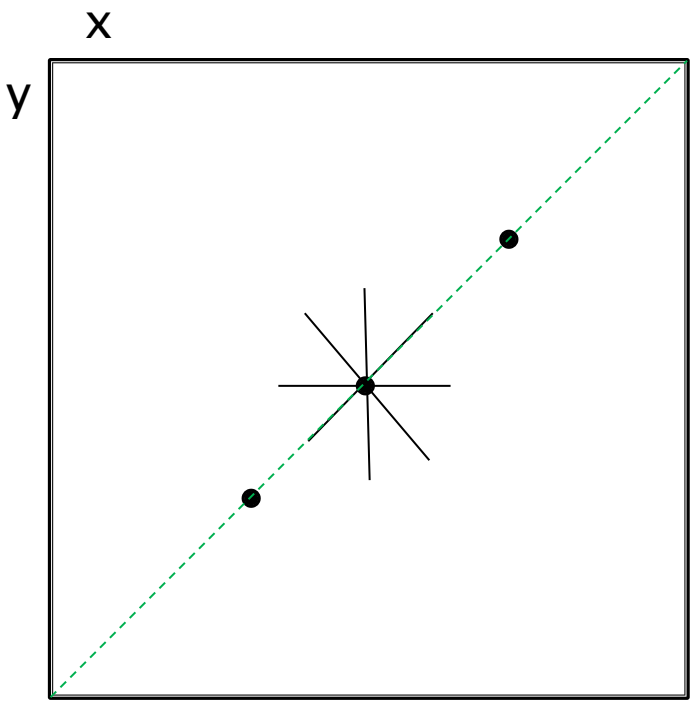
- Detect simple parametric shapes
  - Lines, circles, etc.
- Noise tolerant
- Approach
  - For each edge
    - Increment models consistent with edge
  - Choose models with most votes



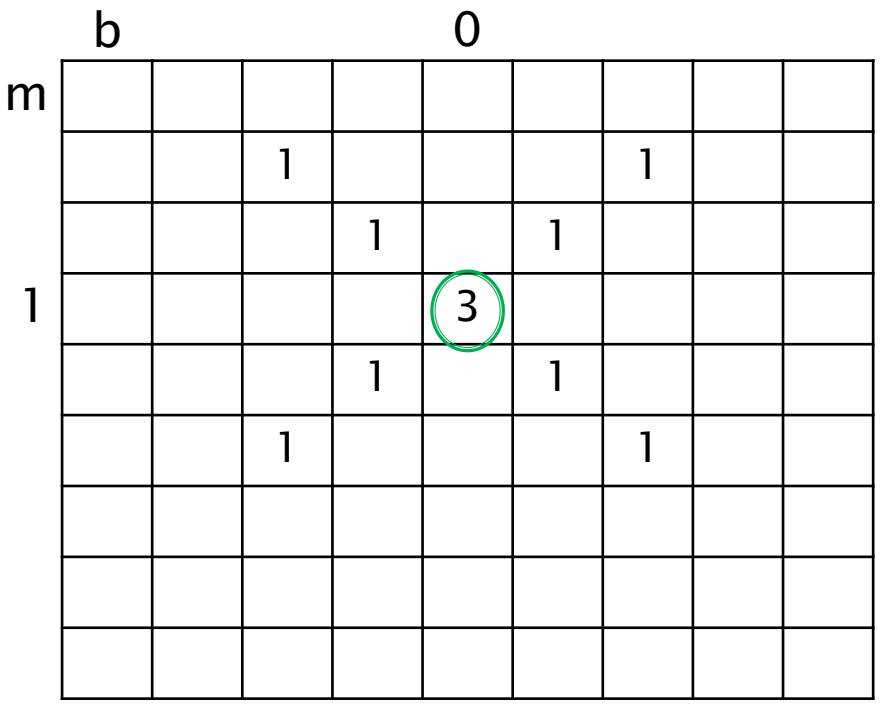


Hough Line Detection

Model: $y = mx + b$

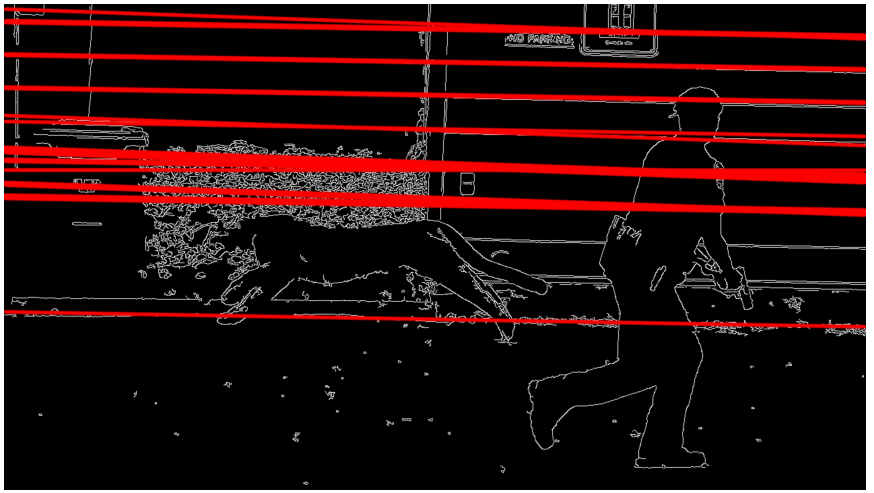


Image



Parameter Space

Hough Line Detection



Probabilistic Hough Transform:
Line segments

Motion: Optical Flow

Optical Flow

- Assumptions
  - Brightness constancy
  - Small motion

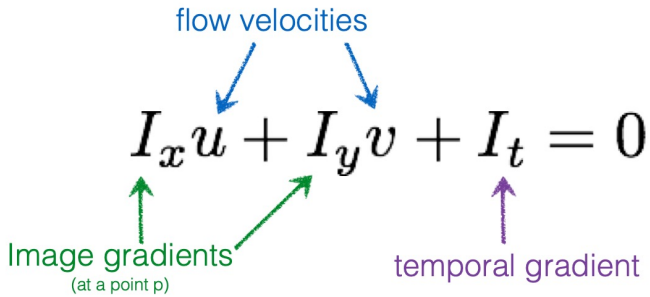
$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$

Brightness Constancy Equation

$$I_x u + I_y v + I_t = 0$$

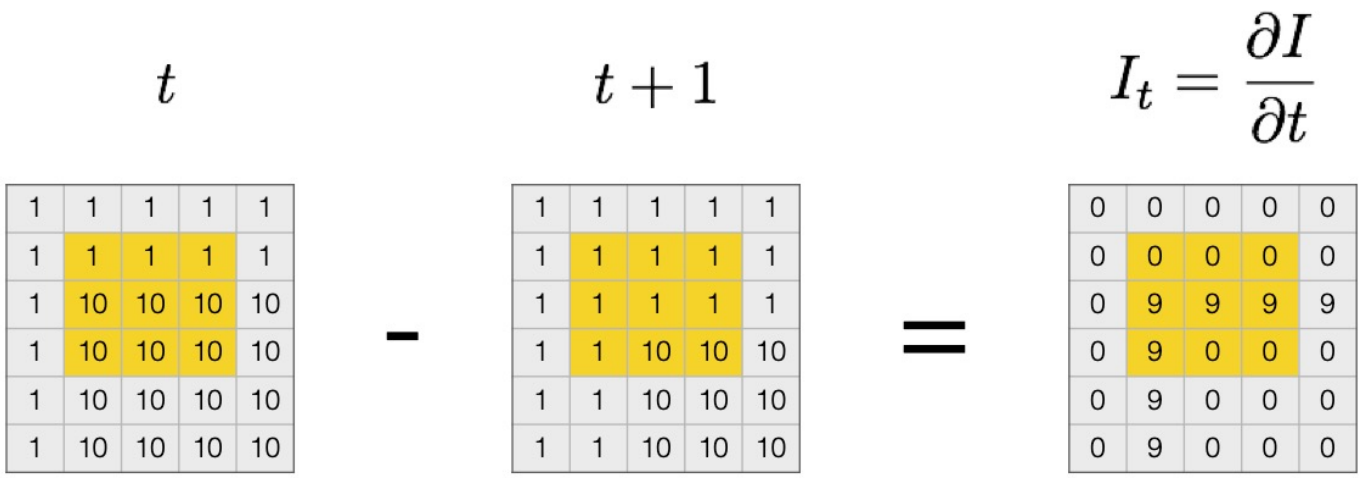
(x-flow) (y-flow)

shorthand notation

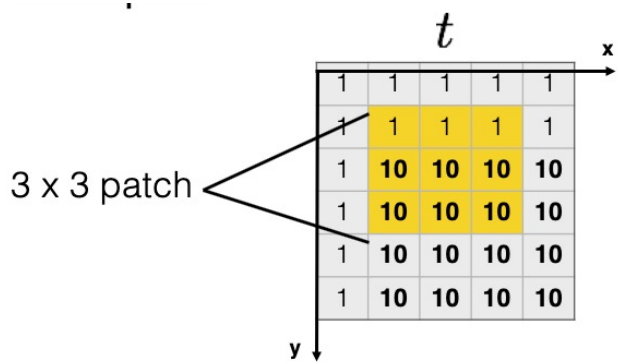


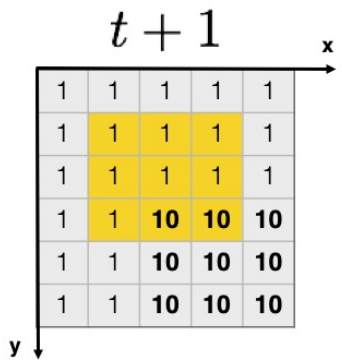
Optical Flow

- Frame differencing

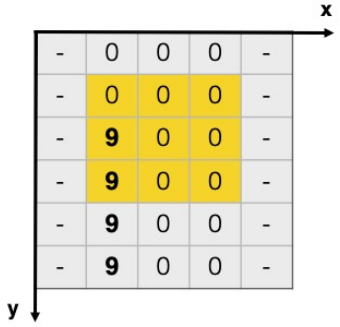


Optical Flow: Frame Differencing



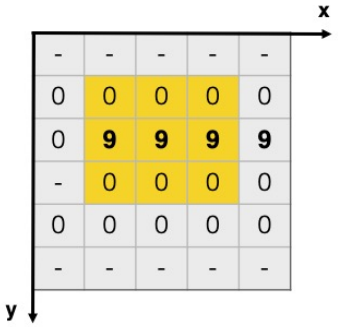


$$I_x = \frac{\partial I}{\partial x}$$



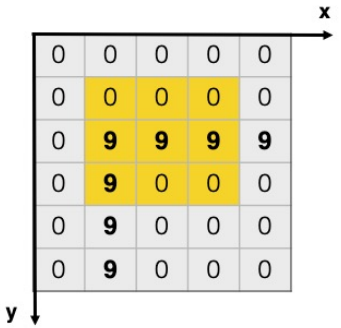
-1 0 1

$$I_y = \frac{\partial I}{\partial y}$$

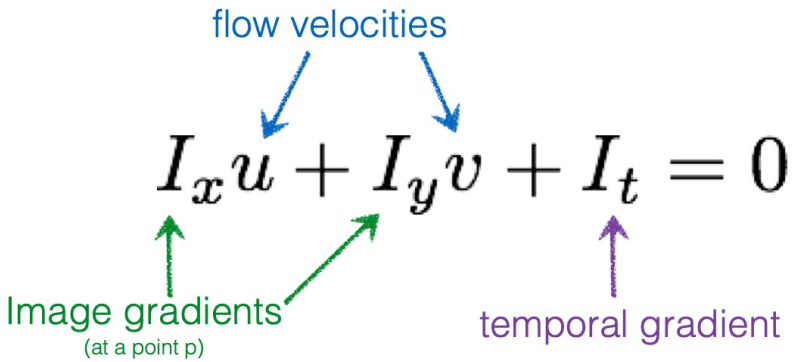


-1
0
1

$$I_t = \frac{\partial I}{\partial t}$$



Optical Flow

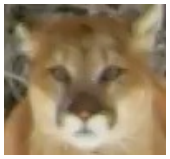


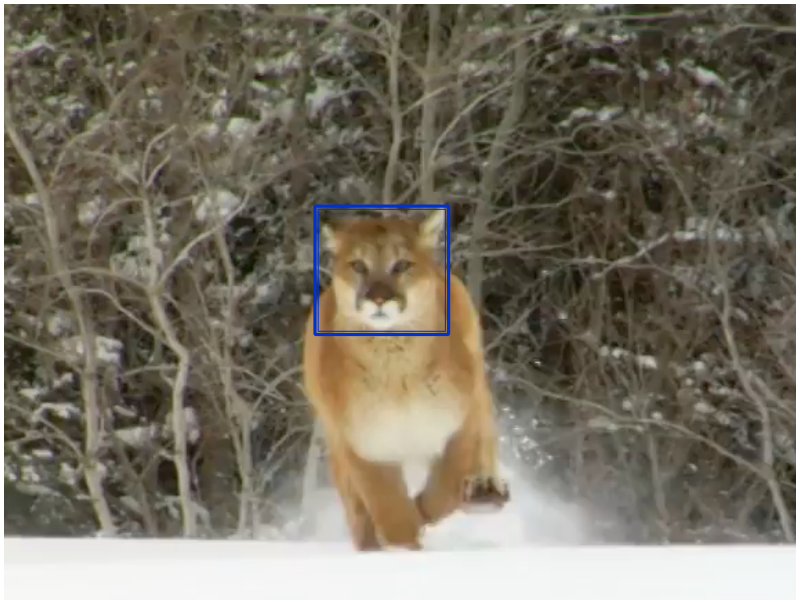
Have I_x, I_y, I_t. Solve for u and v.

Optical Flow

Motion: Tracking

- Given an image feature to track
  - E.g., bounding box
- Find it as the image changes





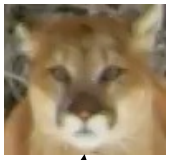
Tracking

- As an image alignment problem...

$$\min_{\boldsymbol{p}} \sum_{\boldsymbol{x}} \left[I(\mathbf{W}(\boldsymbol{x};\boldsymbol{p})) - T(\boldsymbol{x}) \right]^2$$

warped image — template image

I(x)



T(x)

W(x;p)

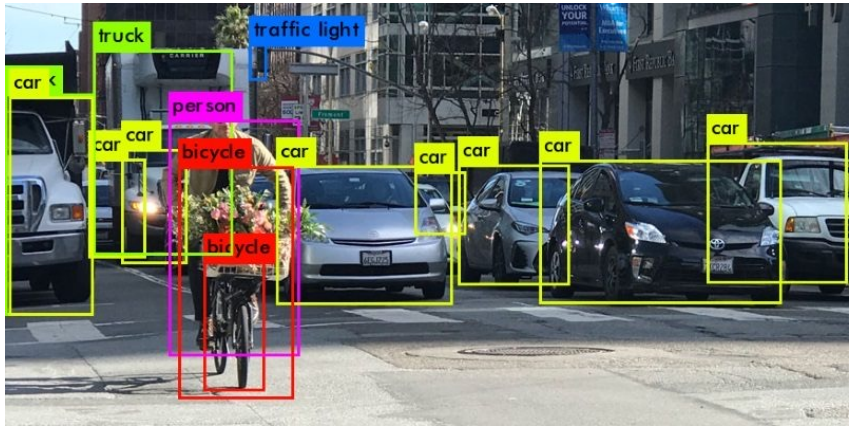
p involves translation, rotation, scaling

Tracking: Demo

What if we don't have a template...?

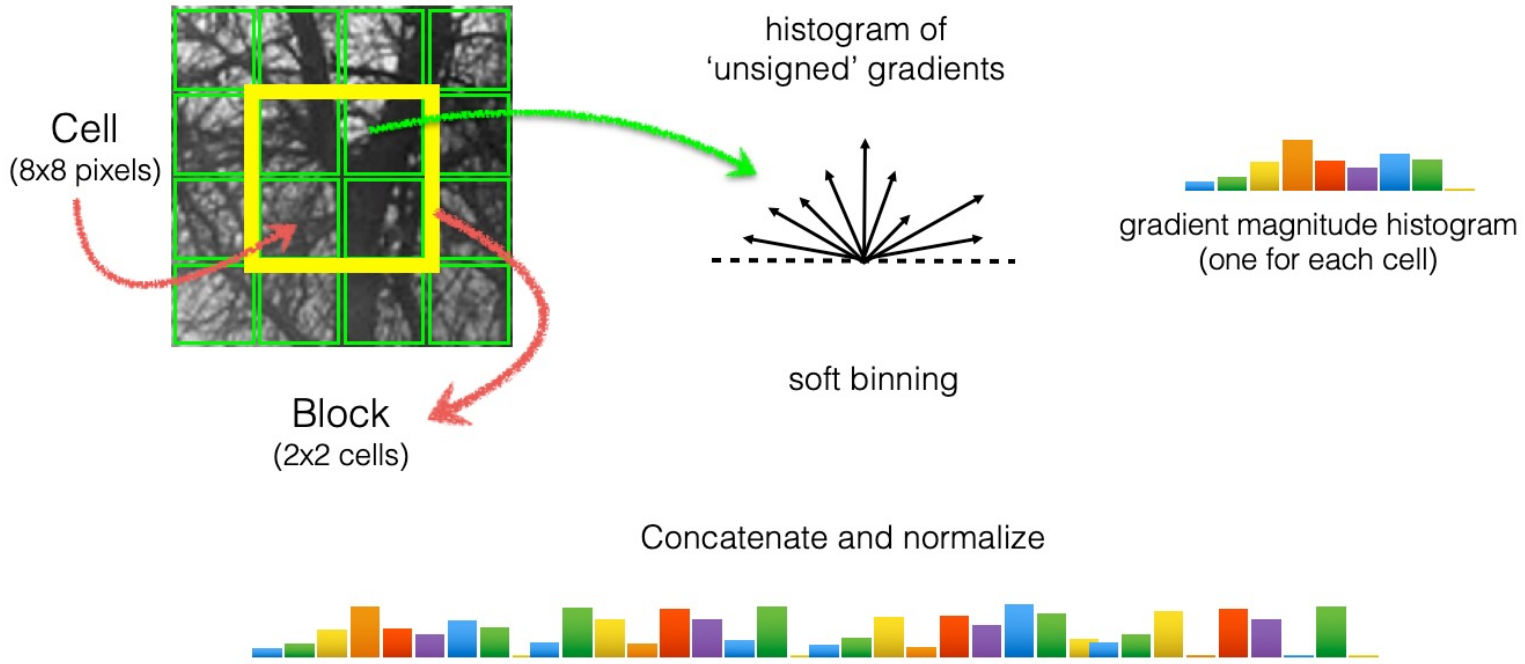
Object Detection

- Approach #1: Feature-based
  - Define various image features
  - Model object in terms of these features
  - Look for feature-level matches in image
- Approach #2: Network-based
  - Train a deep neural network on lots of images with known objects in known locations
  - Use network to locate objects in an image



Object Detection: Feature-based

- Features
  - Histograms of Oriented Gradients (HOG)



Object Detection: Feature-based

- Features
  - Scale Invariant Feature Transform (SIFT)

1. Find extreme points

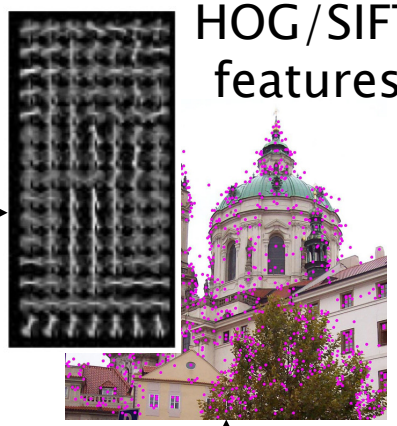
2. Discard low-contrast points

3. Filter points on edges

"keypoints"

Object Detection: Feature-based

Training:



HOG/SIFT features

Machine Learning

airplane ✓

bicylcle

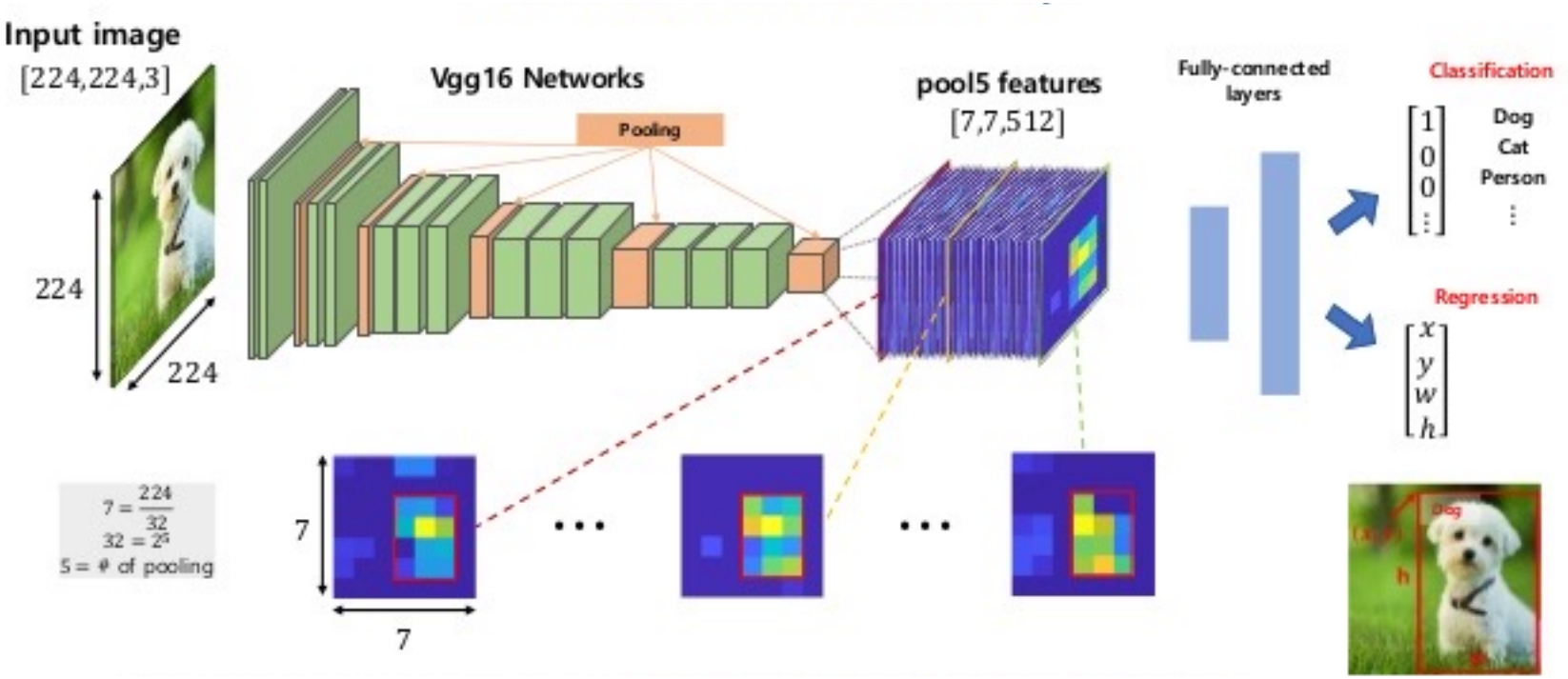
.
.
.

truck ✓

Testing:

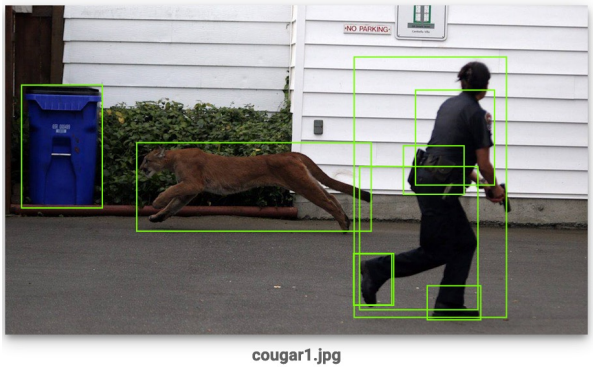
Object Detection: Network-based

Figures out features automatically



Need a lot of training data...

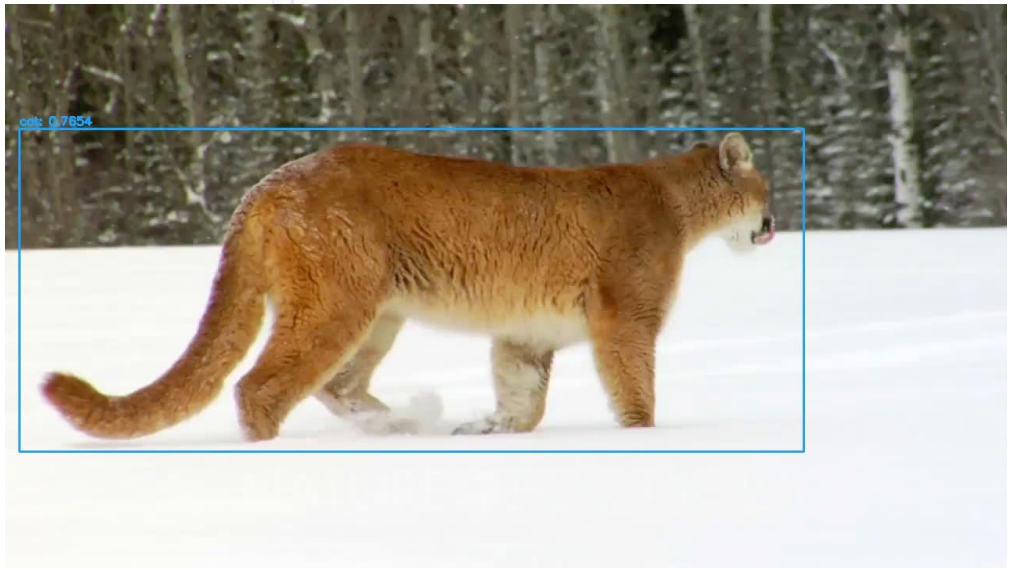
Object Detection: Demo



Person	90%
Pants	90%
Footwear	84%
Shoe	77%
Animal	77%
Luggage & bags	60%
Footwear	56%
Top	54%

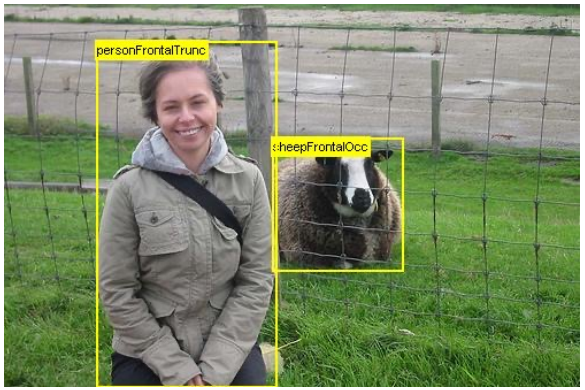
Google Vision

YOLO (You Only Look Once)
Real-time Object Recognition



Object Detection Data

- PASCAL Visual Object Classes (VOC)
  - host.robots.ox.ac.uk/pascal/VOC
  - 12K images
  - 20 classes



- ImageNet
  - image-net.org
  - 14M images
  - 22K classes

- airplane
- helmet
- sunglasses
- person

Image Classification

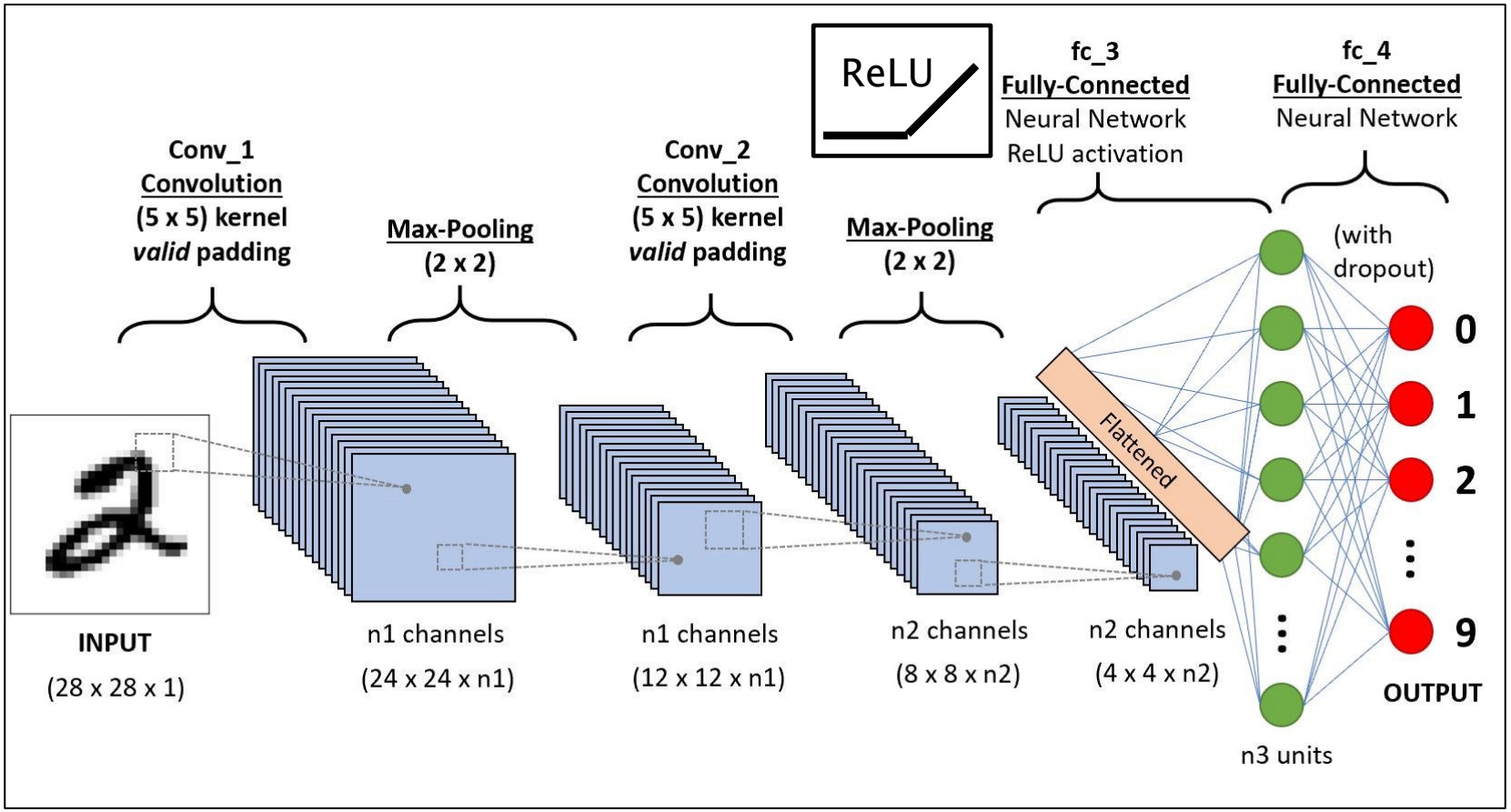
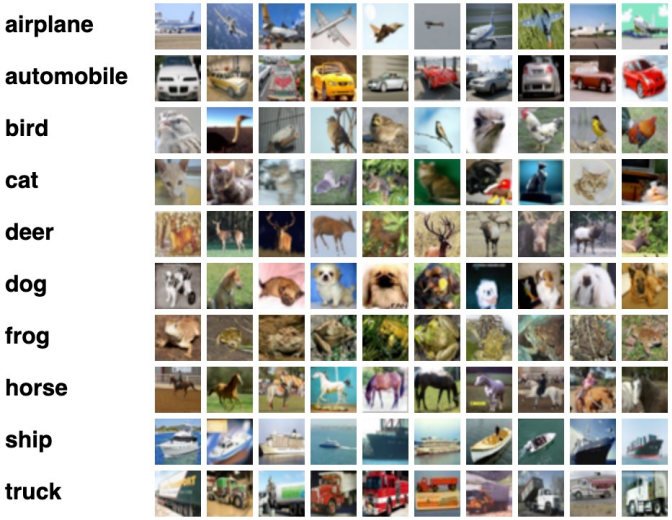


Image Classification Data

- CIFAR-10
  - 60,000 images
  - 10 classes



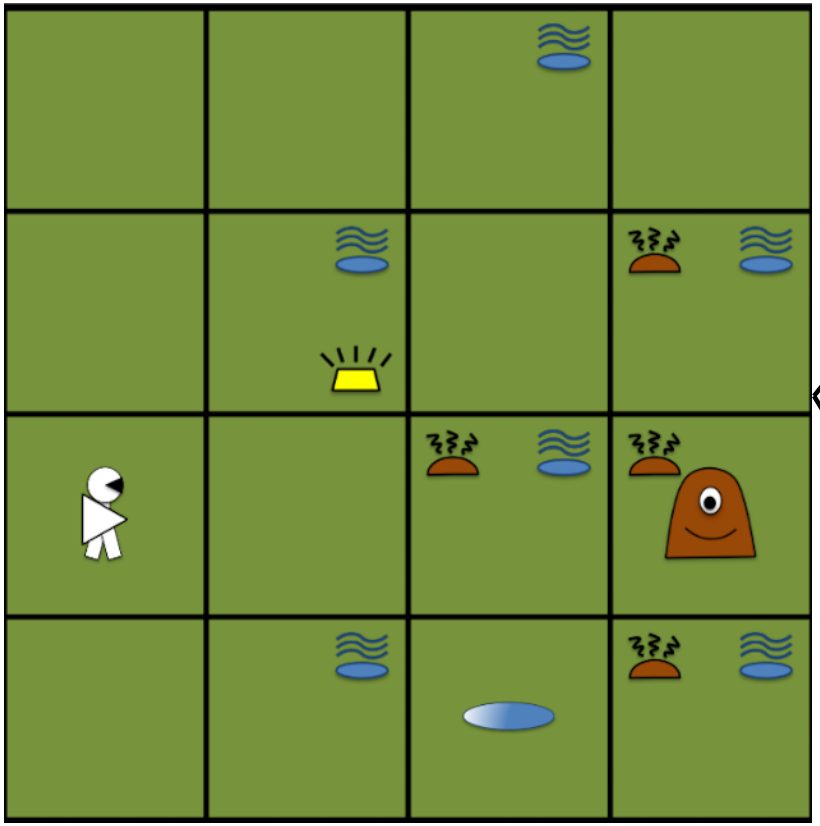
- CIFAR-100
  - 60,000 images
  - 100 classes

Superclass	Classes
aquatic mammals	beaver, dolphin, otter, seal, whale
fish	aquarium fish, flatfish, ray, shark, trout
flowers	orchids, poppies, roses, sunflowers, tulips
food containers	bottles, bowls, cans, cups, plates
fruit and vegetables	apples, mushrooms, oranges, pears, sweet peppers
household electrical devices	clock, computer keyboard, lamp, telephone, television
household furniture	bed, chair, couch, table, wardrobe
insects	bee, beetle, butterfly, caterpillar, cockroach
large carnivores	bear, leopard, lion, tiger, wolf
large man-made outdoor things	bridge, castle, house, road, skyscraper
large natural outdoor scenes	cloud, forest, mountain, plain, sea
large omnivores and herbivores	camel, cattle, chimpanzee, elephant, kangaroo
medium-sized mammals	fox, porcupine, possum, raccoon, skunk
non-insect invertebrates	crab, lobster, snail, spider, worm
people	baby, boy, girl, man, woman
reptiles	crocodile, dinosaur, lizard, snake, turtle
small mammals	hamster, mouse, rabbit, shrew, squirrel
trees	maple, oak, palm, pine, willow
vehicles 1	bicycle, bus, motorcycle, pickup truck, train
vehicles 2	lawn-mower, rocket, streetcar, tank, tractor

Yes, I know mushrooms aren't really fruit or vegetables and bears aren't really carnivores.

www.cs.toronto.edu/~kriz/cifar.html

Scene Understanding

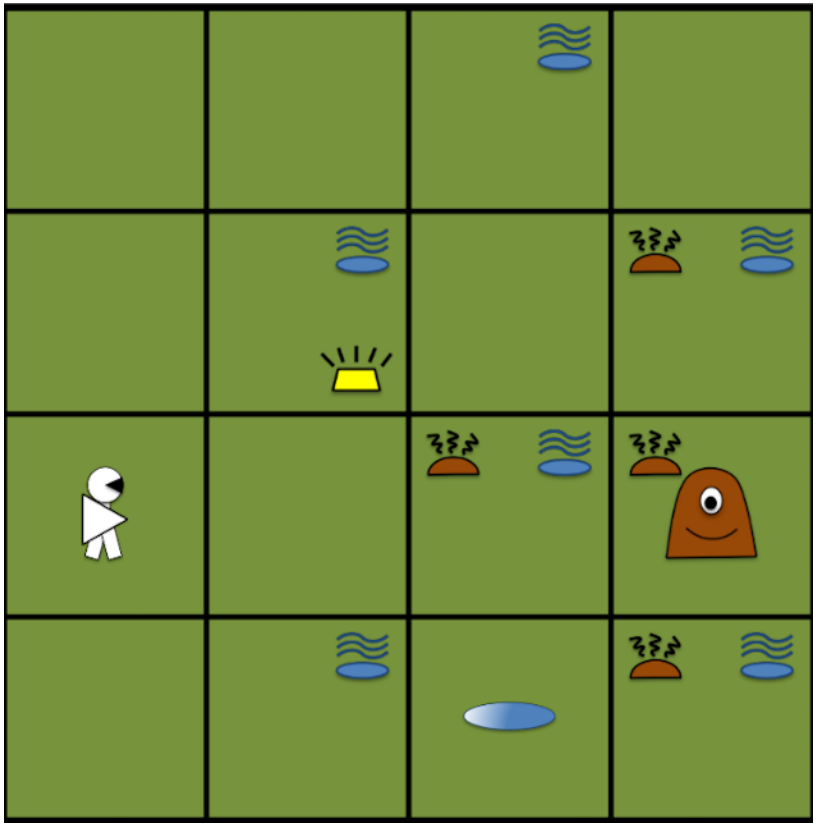


?

I see an astronaut walking on grass. There is a bright gold dog bowl nearby. There are several puddles of water, some with wavy lines above them. There are brown bumps on the grass that are smelly. There is a happy alien with one eye.

location(agent,1,2)
orientation(agent,right)
location(wumpus,4,2)
location(gold,2,3)
location(pit,3,1): 1.0
location(pit,3,3): 0.8
dimensions(4,4)
bestAction(goforward)

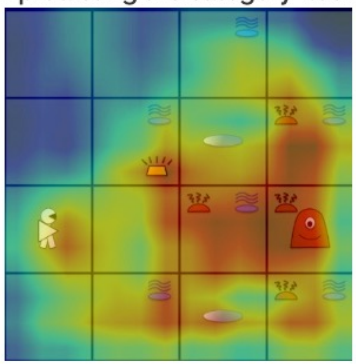
Scene Understanding



MIT Places
Uses 16-layer ConvNet

Predictions:

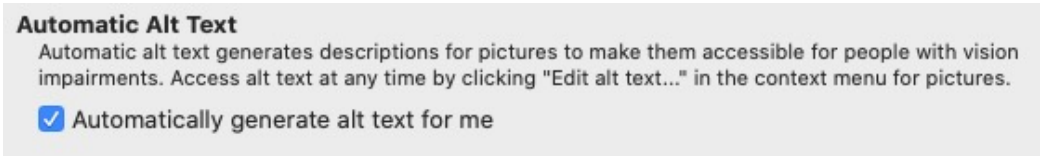
- **Type of environment:** indoor
- **Scene categories:** locker_room (0.503), elevator/door (0.299)
- **Scene attributes:** no horizon, enclosed area, man-made, metal, indoor lighting, glossy, vertical components, competing, sports
- **Informative region for predicting the category "locker_room" is:**



http://places2.csail.mit.edu/demo.html

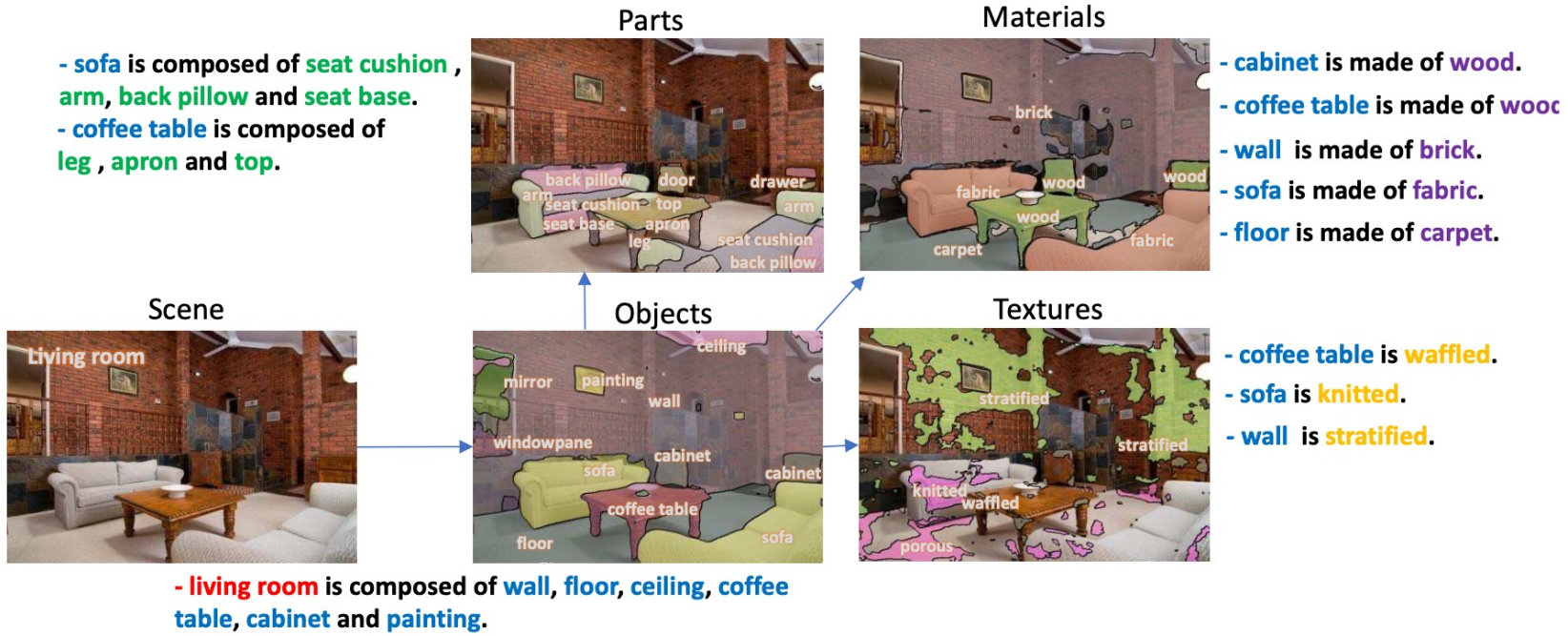
Scene Understanding by Microsoft

- Drag image here
- Edit Alt Text...



"A person walking a dog on a leash in front of a building."

Scene Understanding



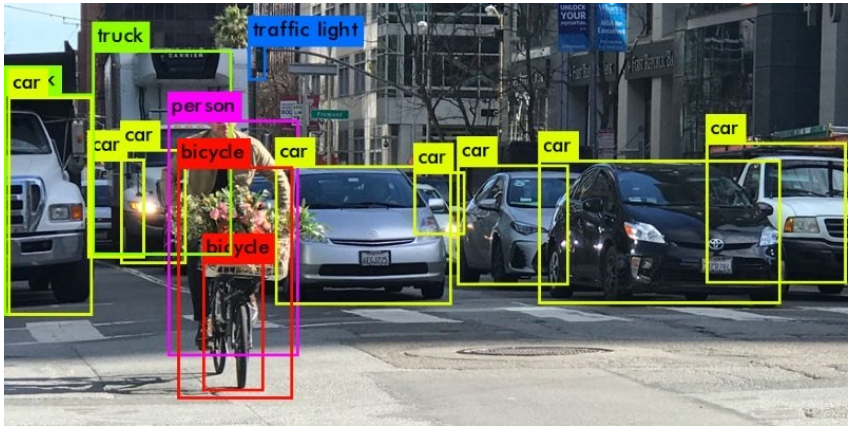
[Xiao et al., ECCV 2018]

Vision Tools

- OpenCV (opencv.org)
- scikit-image (scikit-image.org)
- scikit-video (scikit-video.org/)

Summary: Vision

- Techniques depend on goals of vision
  - Edge detection
  - Shape detection
  - Motion: optical flow, tracking
  - Object detection
  - Image classification



- Future
  - Scene understanding
  - Video summarization
  - Fake images and video