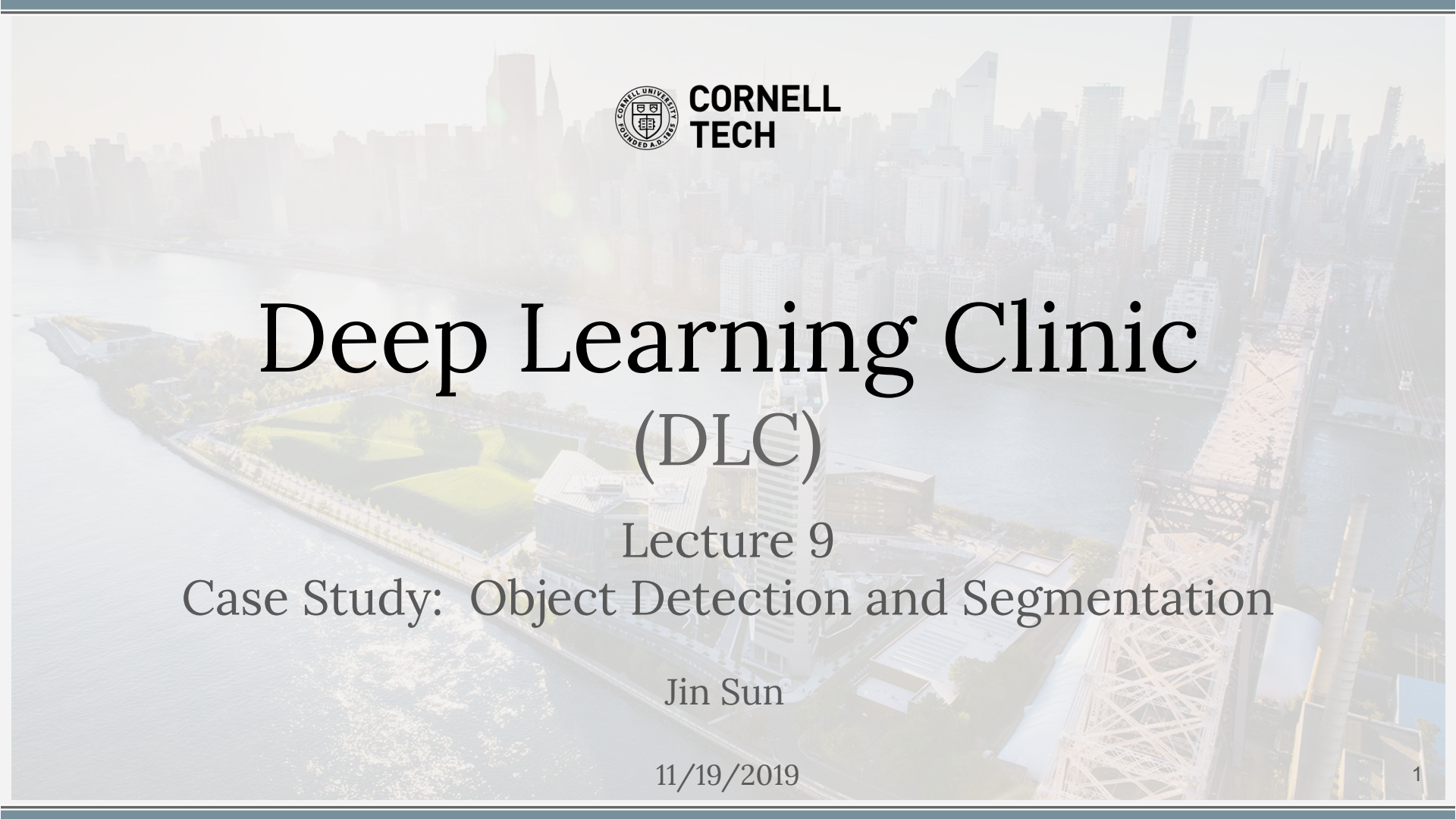


CORNELL TECH

Deep Learning Clinic (DLC)

Lecture 9
Case Study: Object Detection and Segmentation

Jin Sun

Today - Object Detection and Segmentation

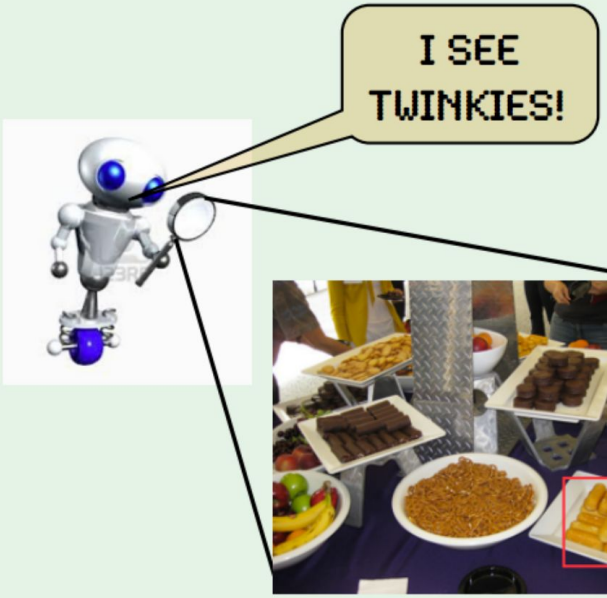
- **Overview**
- Object Detection
- Segmentation
- Detection + Segmentation: Instance Segmentation
- What's Next

It's All About Objects

Snack time in the lab

Slides from Ross Girshick

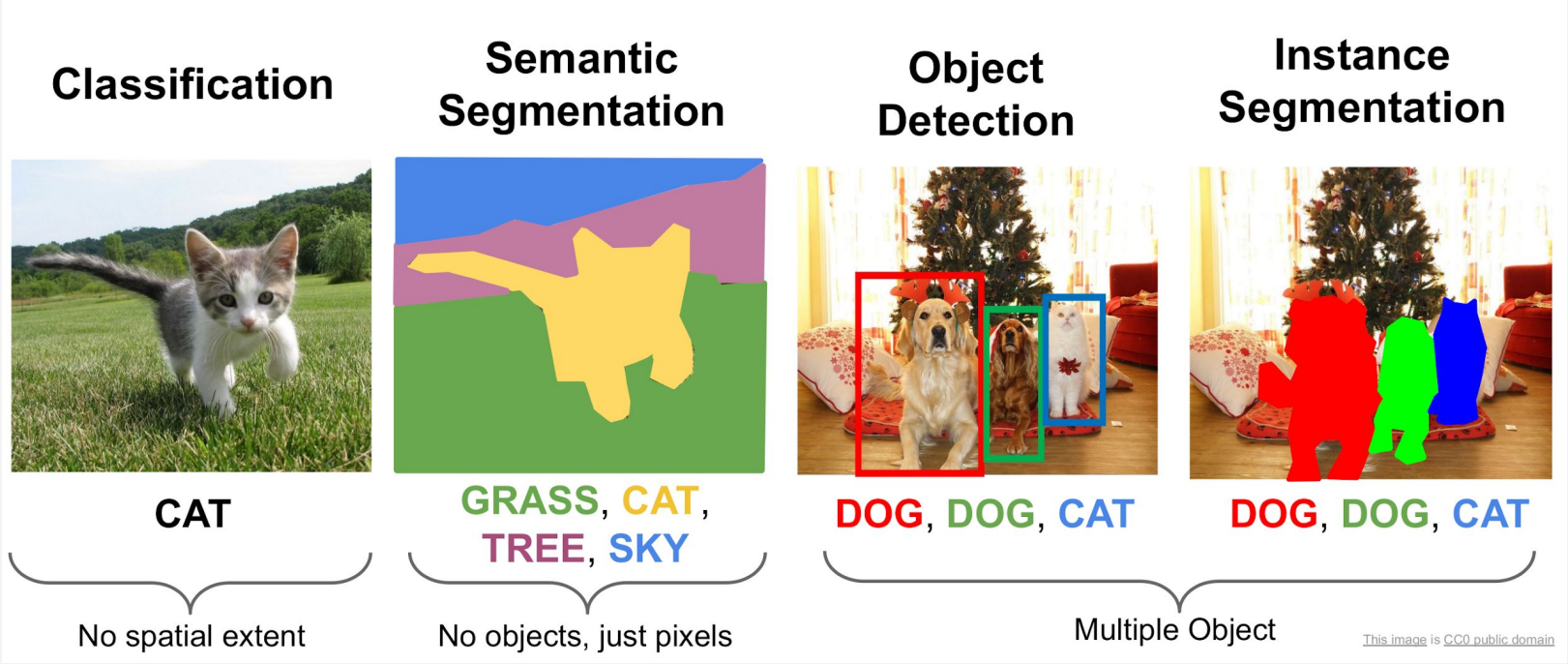
What objects are where?



robot: “I see a table with twinkies, pretzels, fruit, and some mysterious chocolate things...”

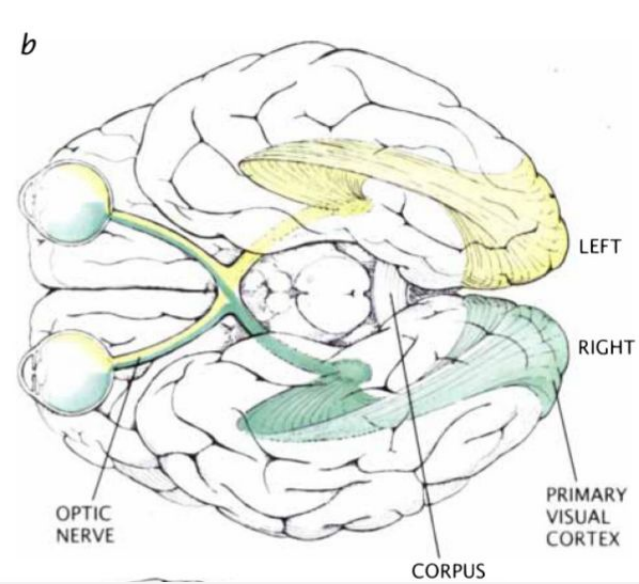
Slides from Ross Girshick

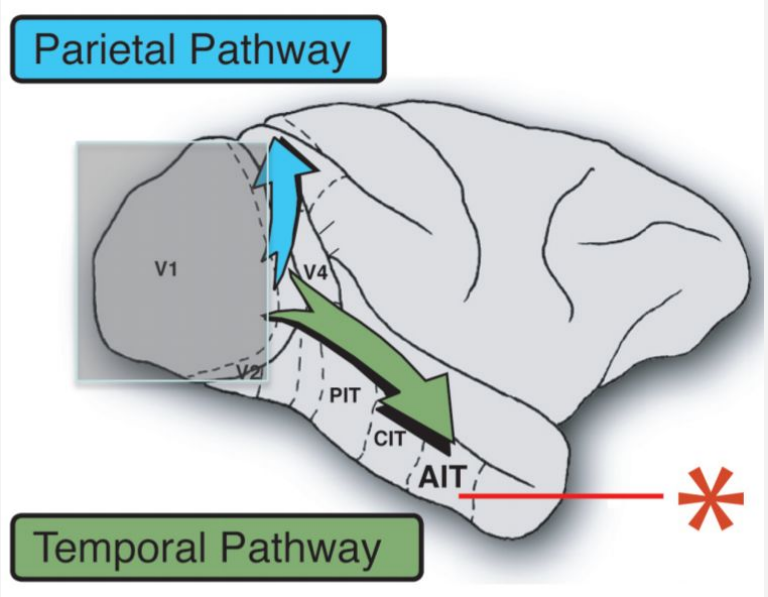
Computer Vision Tasks



Slides from Stanford cs231n

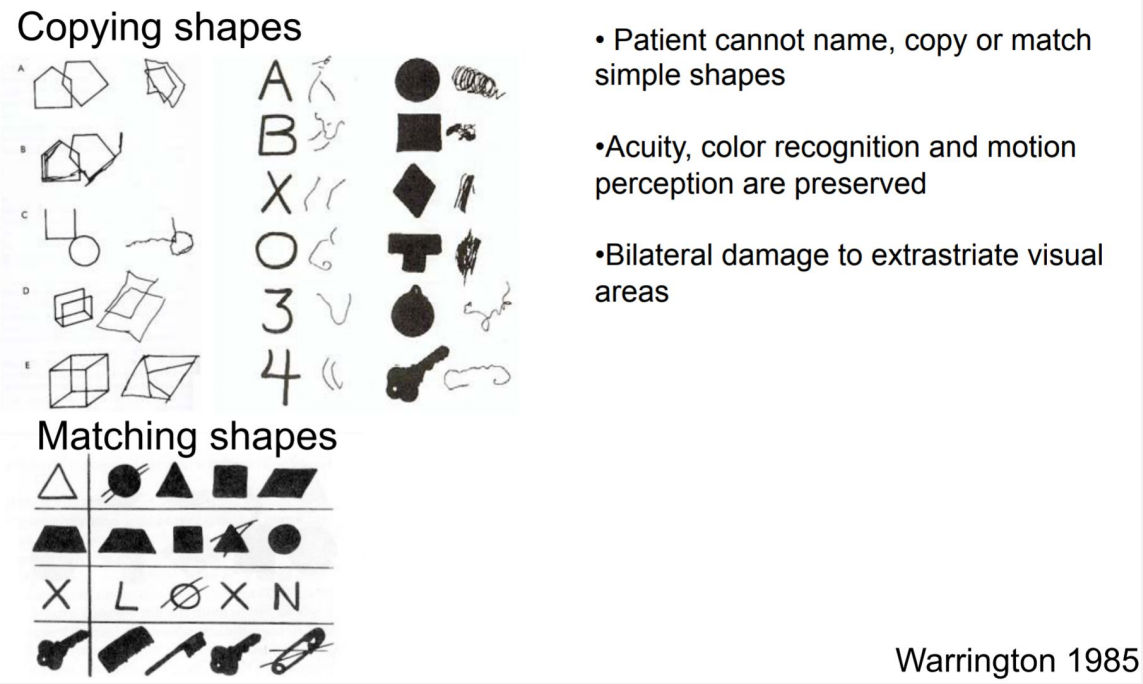
Object Recognition in Brains





http://klab.tch.harvard.edu/academia/classes/Neuro230/2016/slides/hms230_2016_Lecture3.pdf

Lesion Study on Object Recognition



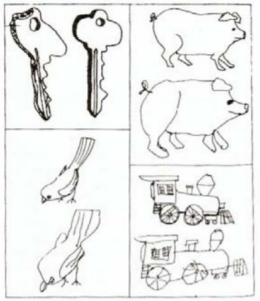
- Patient cannot name, copy or match simple shapes
- Acuity, color recognition and motion perception are preserved
- Bilateral damage to extrastriate visual areas

Warrington 1985

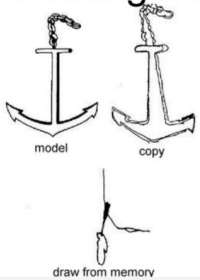
http://klab.tch.harvard.edu/academia/classes/Neuro230/2016/slides/hms230_2016_Lecture3.pdf

Lesion Study on Object Recognition

Copying from templates



Drawing from memory



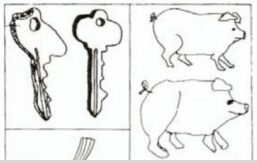
- Subject can copy complex drawings, match complex shapes and use the objects correctly
- Subject cannot identify (name) those shapes
- Subject cannot draw from memory
- Acuity, color recognition and motion perception are preserved
- Bilateral lesion of the anterior inferior temporal lobe

Warrington 1985

http://klab.tch.harvard.edu/academia/classes/Neuro230/2016/slides/hms230_2016_Lecture3.pdf

Lesion Study on Object Recognition

Copying from templates



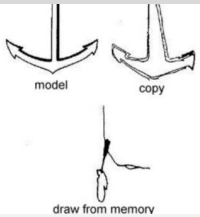
• Subject can copy complex drawings, match complex shapes and use the objects correctly

• Subject cannot identify (name) those

Object Recognition seems to be effortless for human

But actually it is quite complicated!

temporal lobe



Warrington 1985

http://klab.tch.harvard.edu/academia/classes/Neuro230/2016/slides/hms230_2016_Lecture3.pdf

Why This Is A Hard Problem

Viewpoint

Slides from Rob Fergus

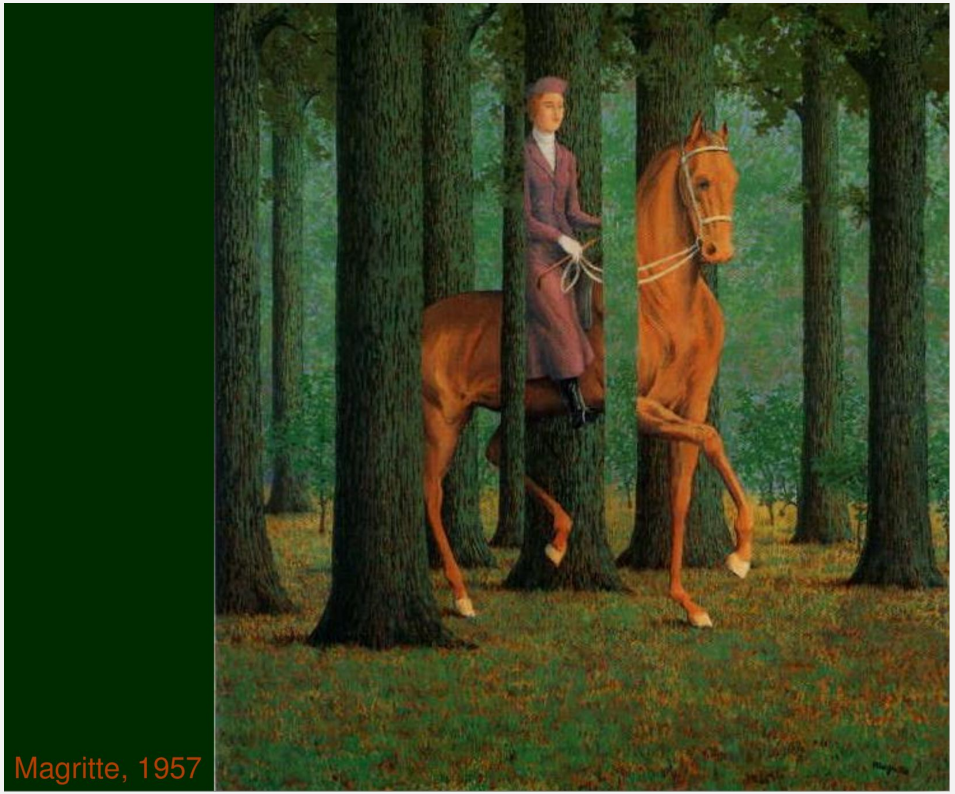
Why This Is A Hard Problem

Illumination

Slides from Rob Fergus

Why This Is A Hard Problem

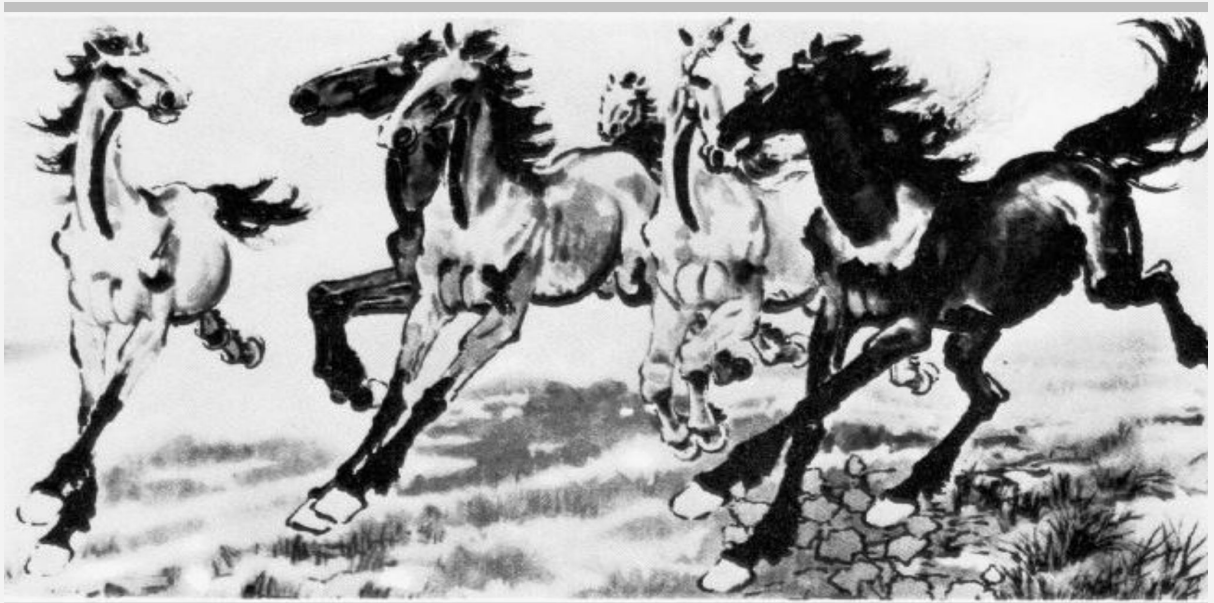
Occlusion



Slides from Rob Fergus

Why This Is A Hard Problem

Deformation



Xu, Beihong 1943

Slides from Rob Fergus

Why This Is A Hard Problem

Background Clutter

Slides from Rob Fergus

Why This Is A Hard Problem

In-class variations

Slides from Rob Fergus

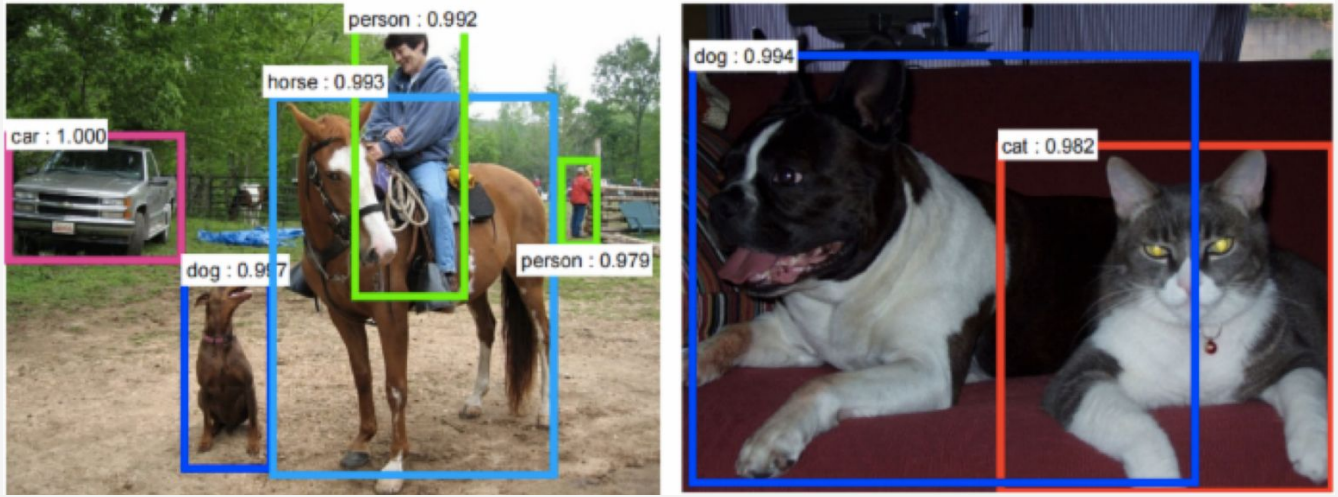
Applications of Object Recognition

- Face Detection
- Face Verification
- Assistive Driving
- Cashier-free Shopping
- Drones
- And much more!

Today - Object Detection and Segmentation

- Overview
- **Object Detection**
- Segmentation
- Detection + Segmentation: Instance Segmentation
- What's Next

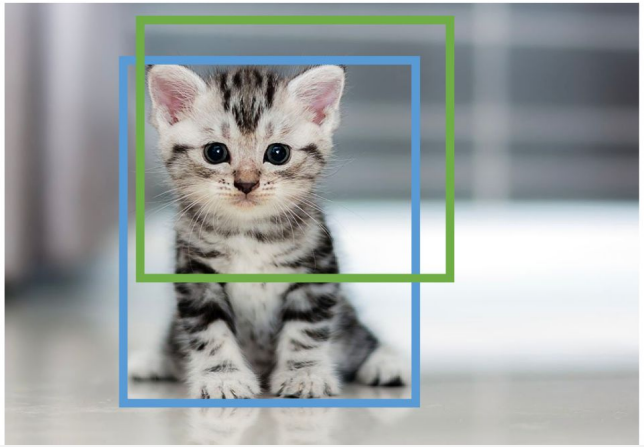
Task Formulation



Slides from Abhinav Shrivastava

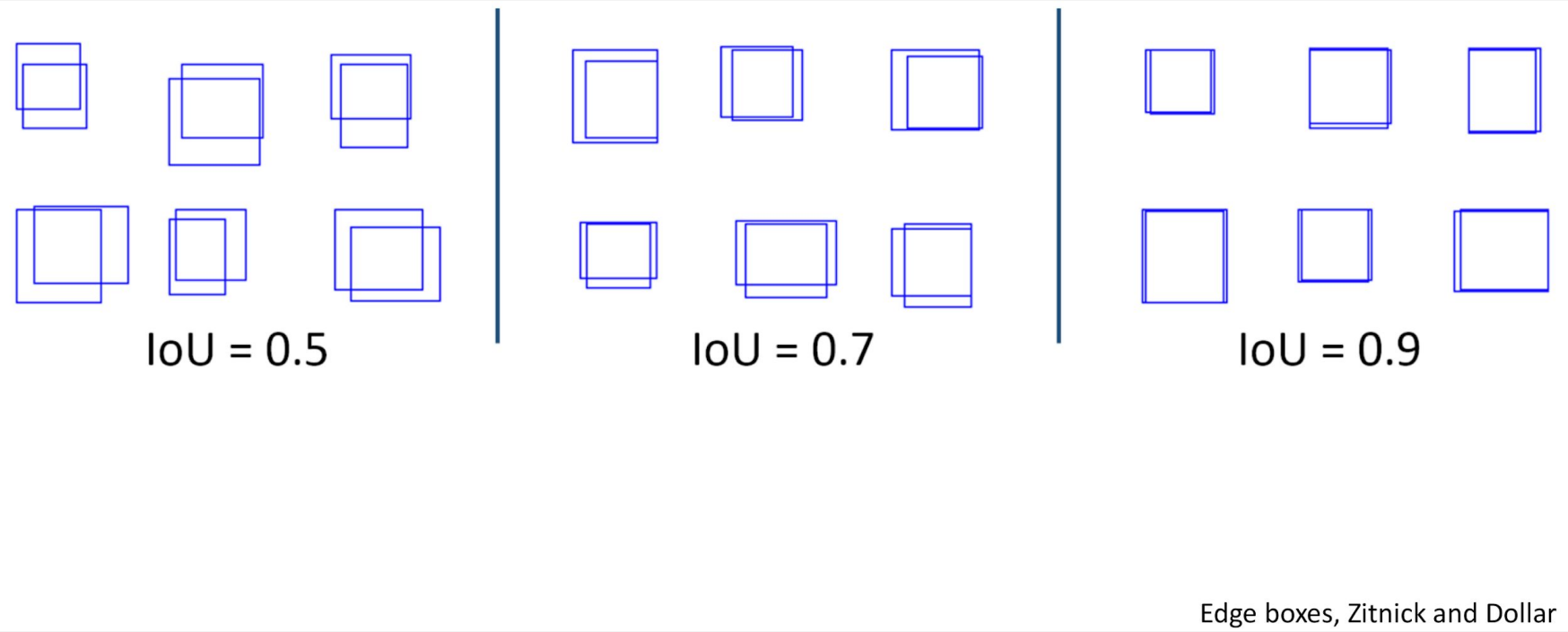
Evaluation Metric

$$\text{Overlap} = \frac{\square \cap \square}{\square \cup \square} > \text{threshold} \quad \checkmark$$



Slides from Abhinav Shrivastava

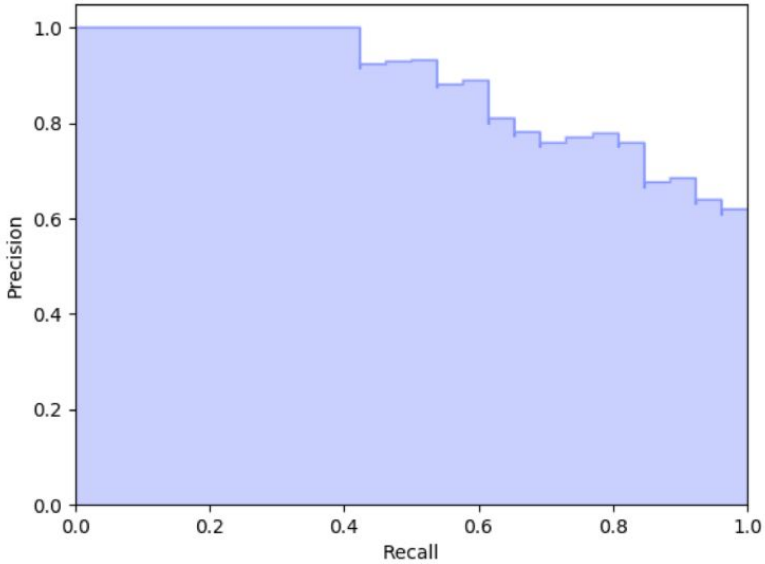
Evaluation Metric



Edge boxes, Zitnick and Dollar

Slides from Abhinav Shrivastava

Evaluation Metric



Precision:
true positive detections / total detections
Recall:
true positive detections / total positive test instances

Average Precision: Area under curve

Slides from Abhinav Shrivastava

Simple Detector: Sliding Window with Correlation

Example 1: Find Waldo!

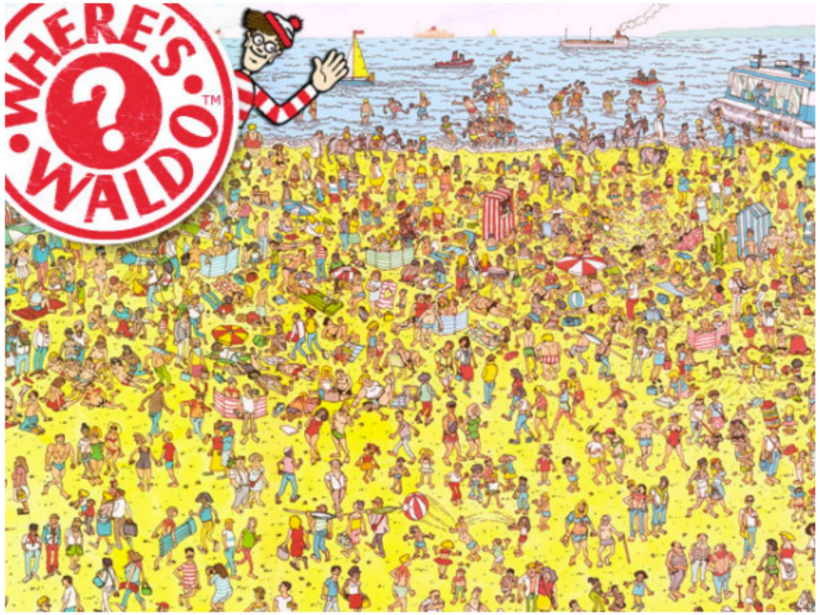
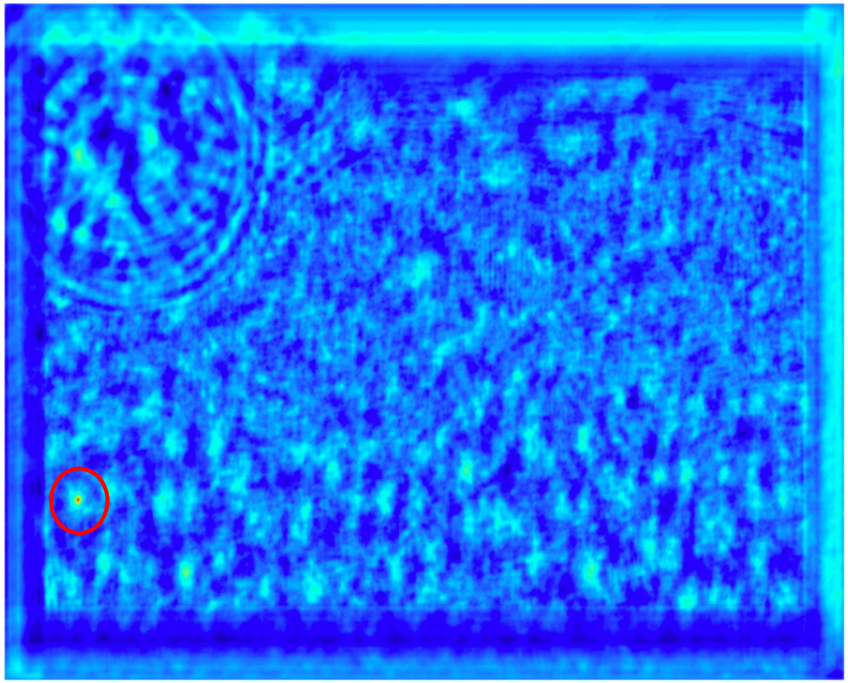


image *I*

slide credit: Chris McIntosh

Simple Detector: Sliding Window with Correlation

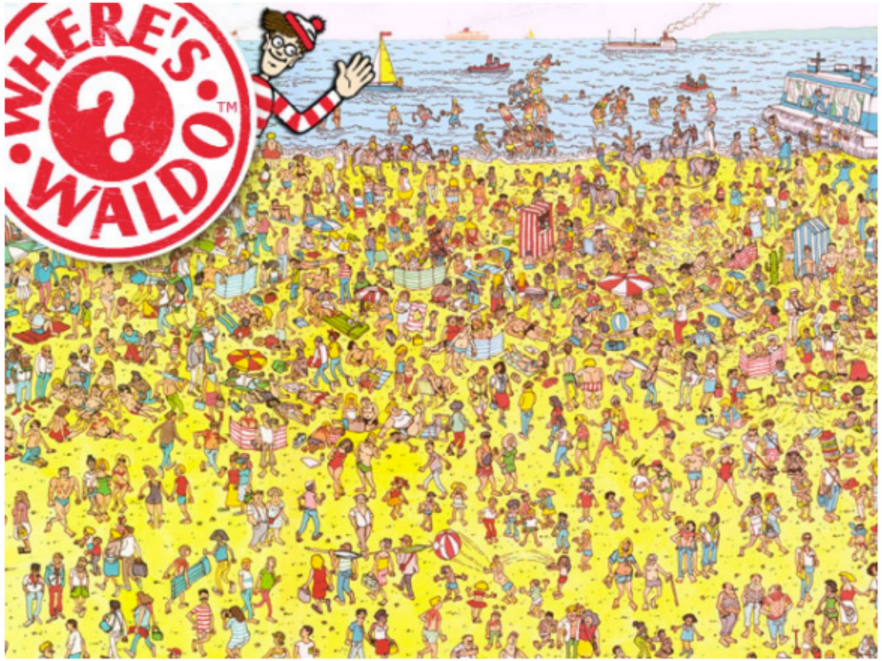
2. Result of normalized cross-correlation

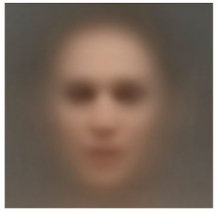


slide credit: Chris McIntosh

Simple Detector: Sliding Window with Correlation

Example 2: Find all persons?





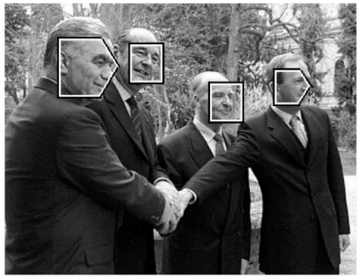
A template for all instances?

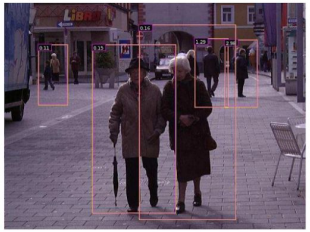
We need features!

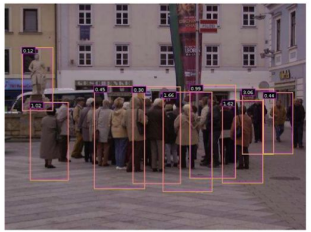
slide credit: Chris McIntosh

Pre-Deep Learning Era

- Wide variety of poses
- Variable appearance
- Complex backgrounds
- Unconstrained illumination
- Occlusions, different scales
- …







Histogram of Oriented Gradients (HOG)

Scan image(s) at all scales and locations

↓

Extract features over windows

↓

Run linear SVM classifier on all locations

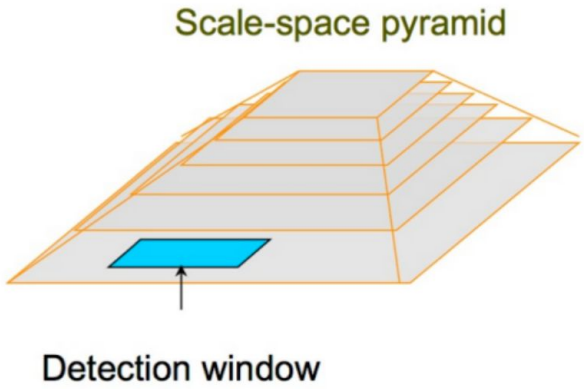
↓

Fuse multiple detections in 3-D position & scale space

↓

Object detections with bounding boxes

locations



scales

Slide credit: Sanja Fidler

Histogram of Oriented Gradients (HOG)

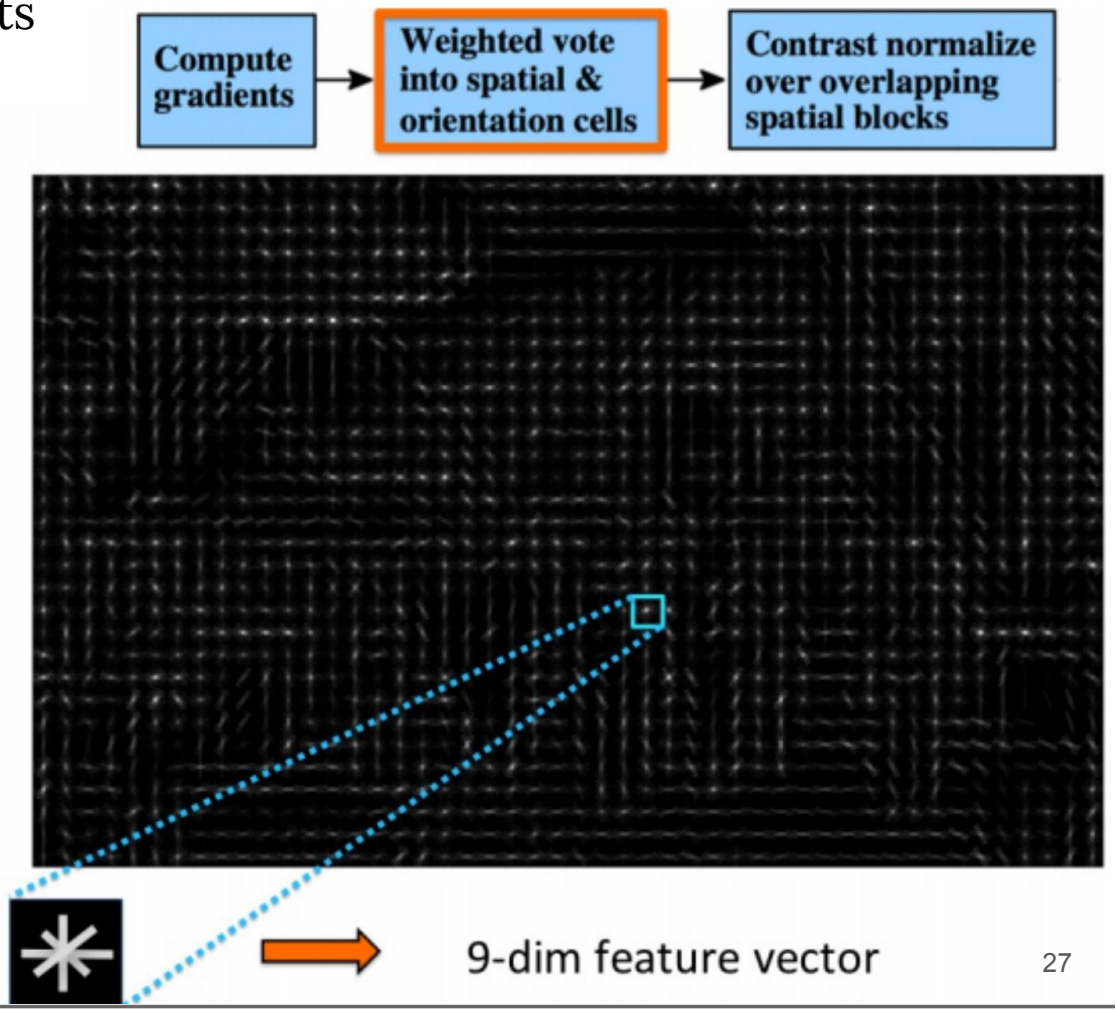
Scan image(s) at all scales and locations

Extract features over windows

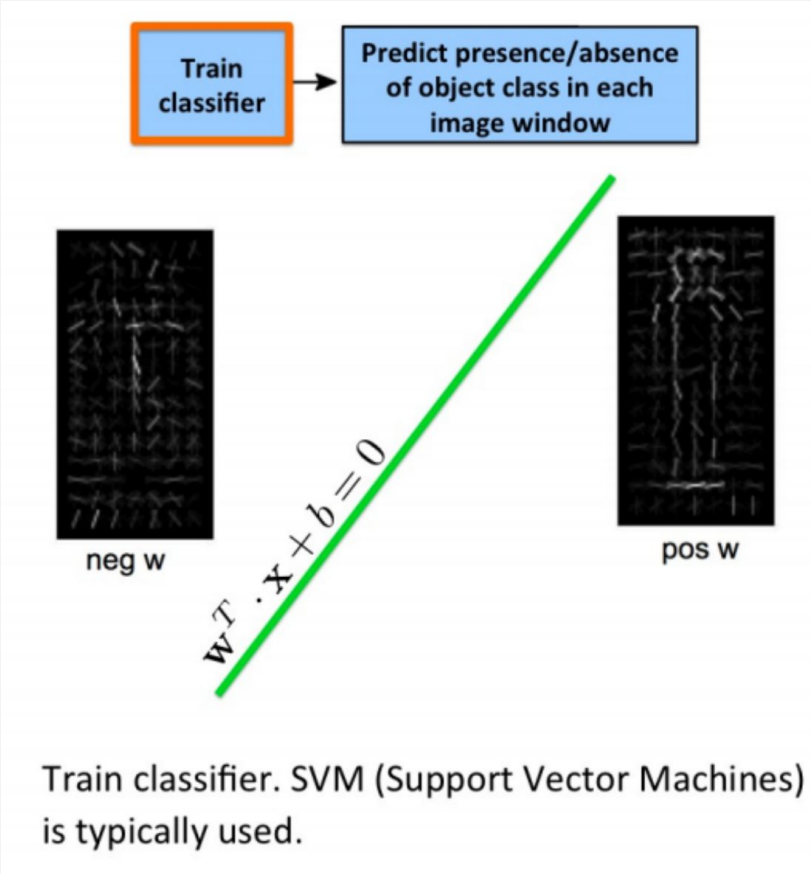
Run linear SVM classifier on all locations

Fuse multiple detections in 3-D position & scale space

Object detections with bounding boxes



Slide credit: Sanja Fidler

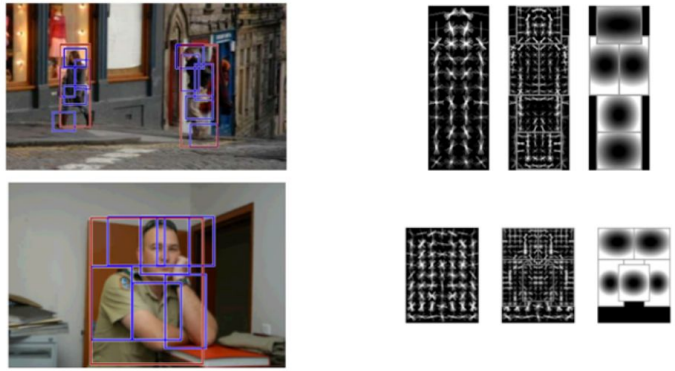


Train classifier. SVM (Support Vector Machines) is typically used.

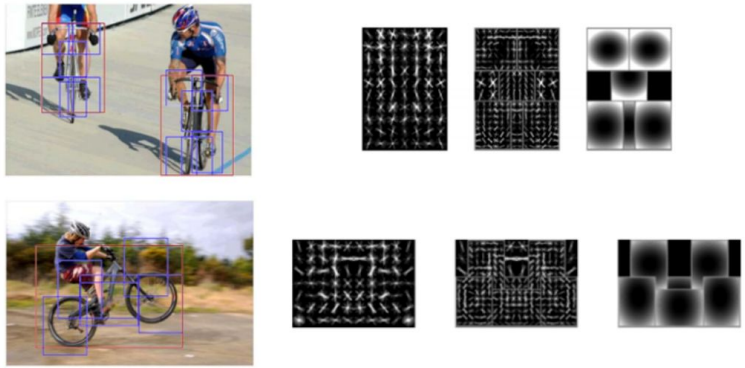
Slides from Sanja Fidler

Deformable Parts Model

Mixture Model Example - Person

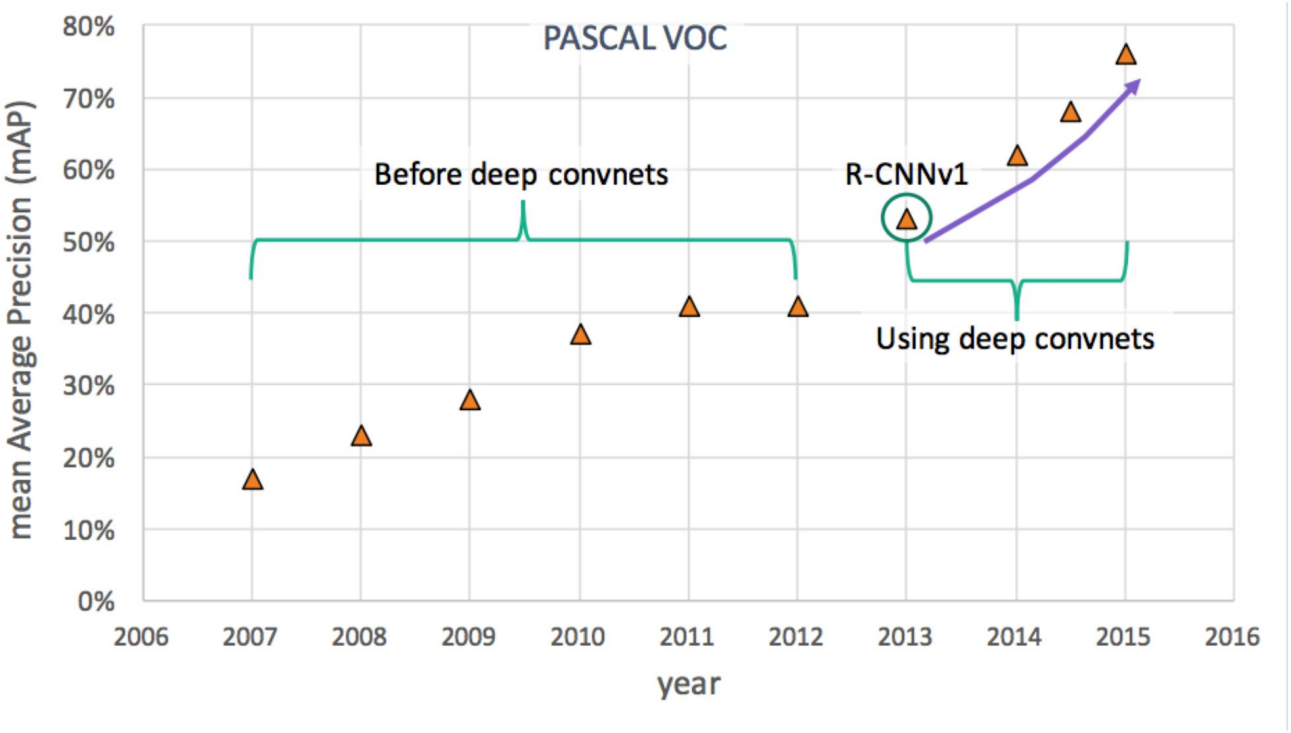


Mixture Model Example - Bicycle



Slide credit: Mubarak Shah

A Significant Boost from Deep Learning

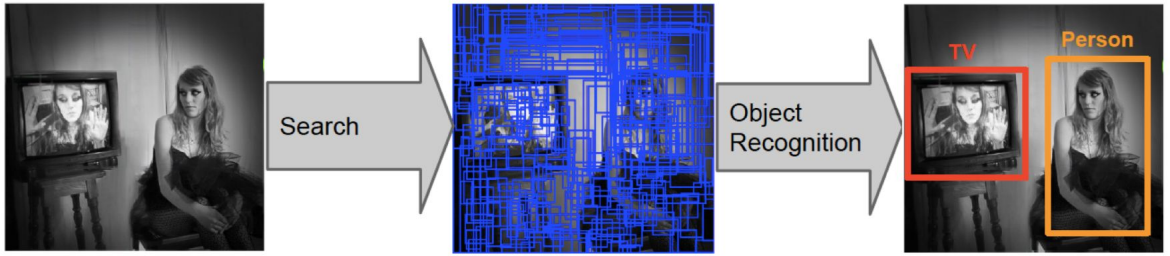


Slide credit: Renjie Liao

Two Types of Detection Paradigm

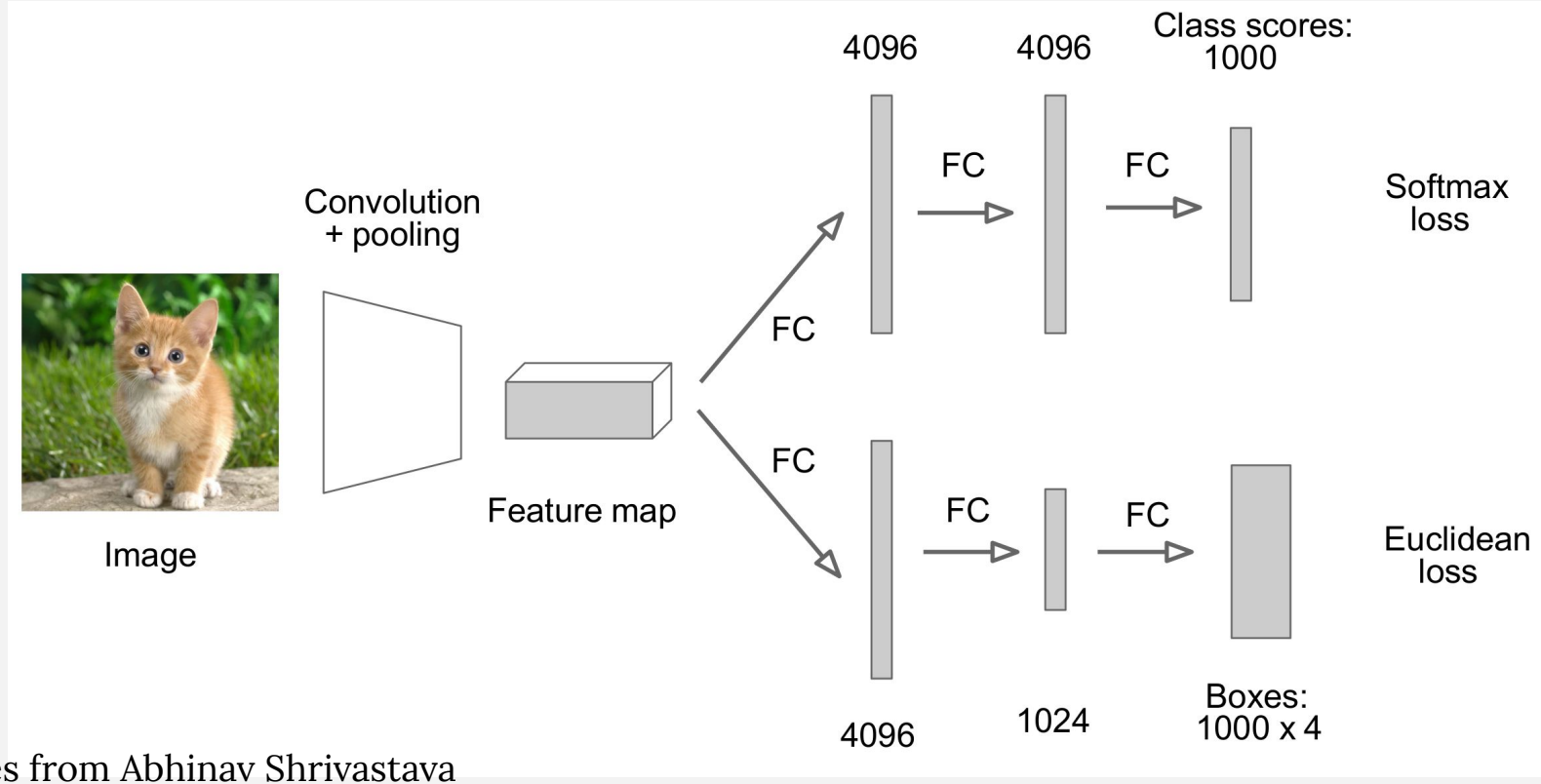
Single-stage Detectors (I-Stage)

Two-stage Detectors (II-Stage):



Slides from Abhinav Shrivastava

Single Stage - Overfeat



Slides from Abhinav Shrivastava

Single Stage - YOLO

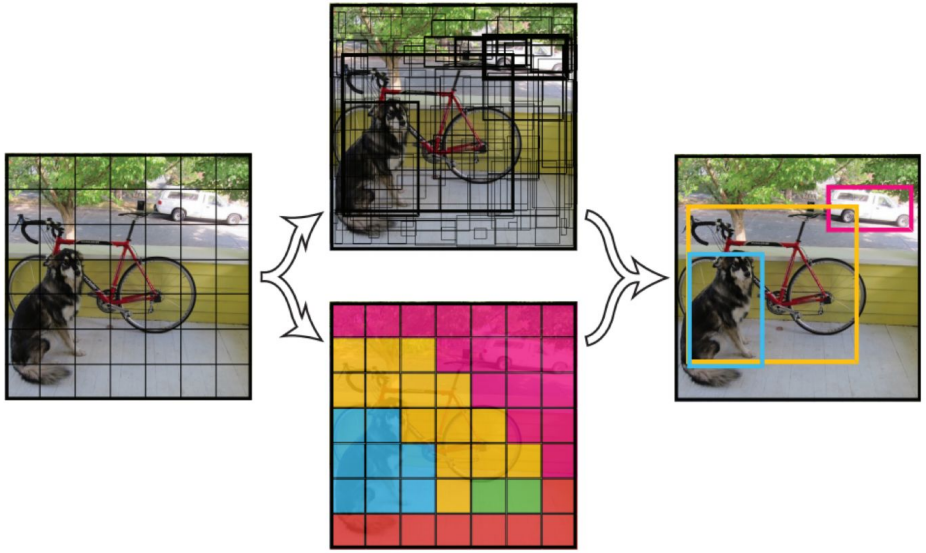
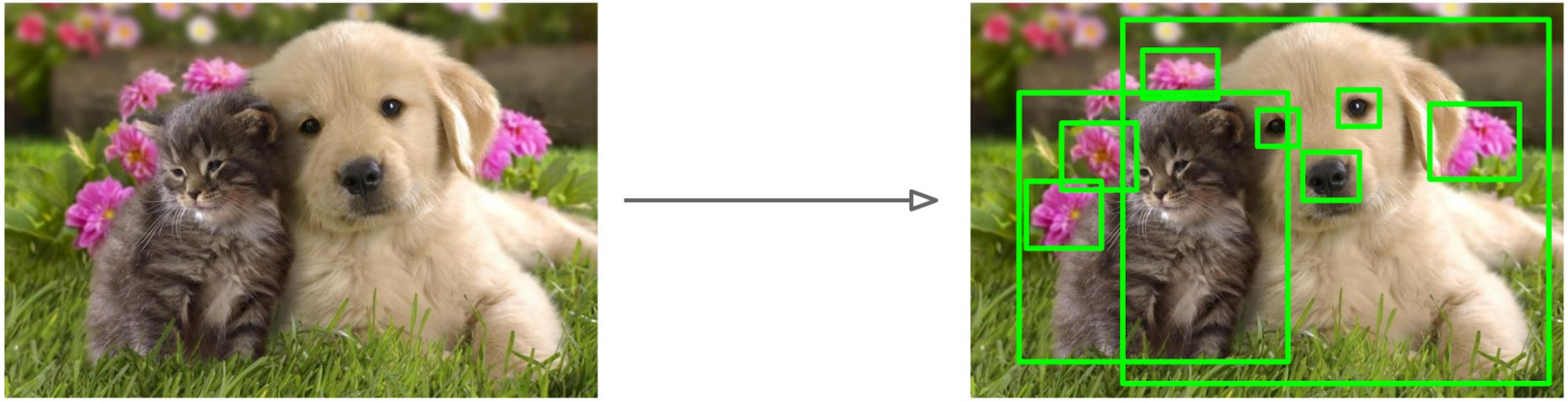


Figure 2: The Model. Our system models detection as a regression problem. It divides the image into an even grid and simultaneously predicts bounding boxes, confidence in those boxes, and class probabilities. These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

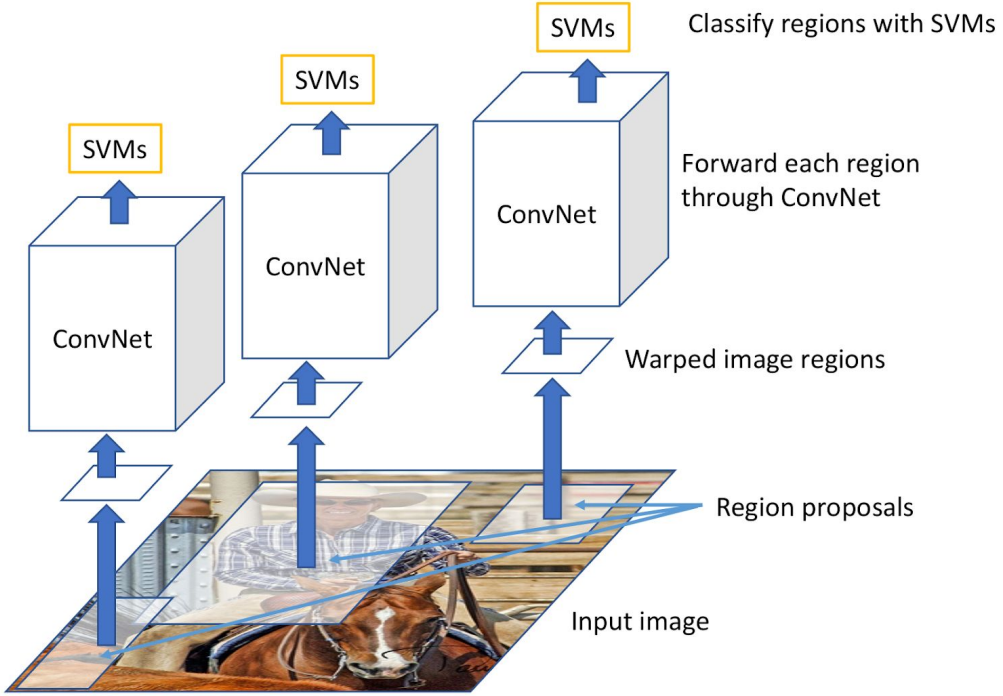
Region Proposals in Two Stage Detectors

- Find “blobby” image regions that are likely to contain objects
- “Class-agnostic” object detector
- Look for “blob-like” regions



Slide credit: CS231n: Justin Johnson, Andrej Karpathy, Fei-Fei Li

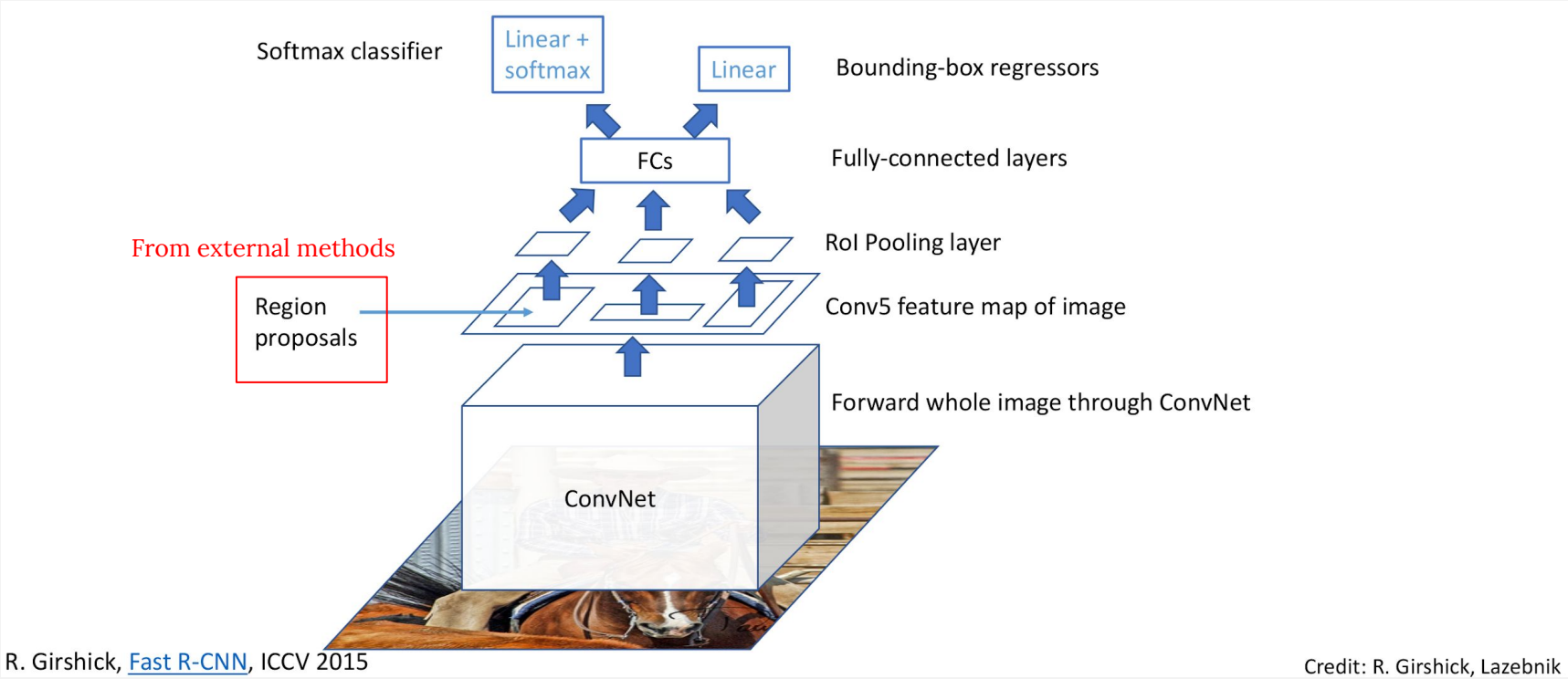
RCNN: Region proposals + CNN features



Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, CVPR 2014.

Credit: R. Girshick, Lazebnik

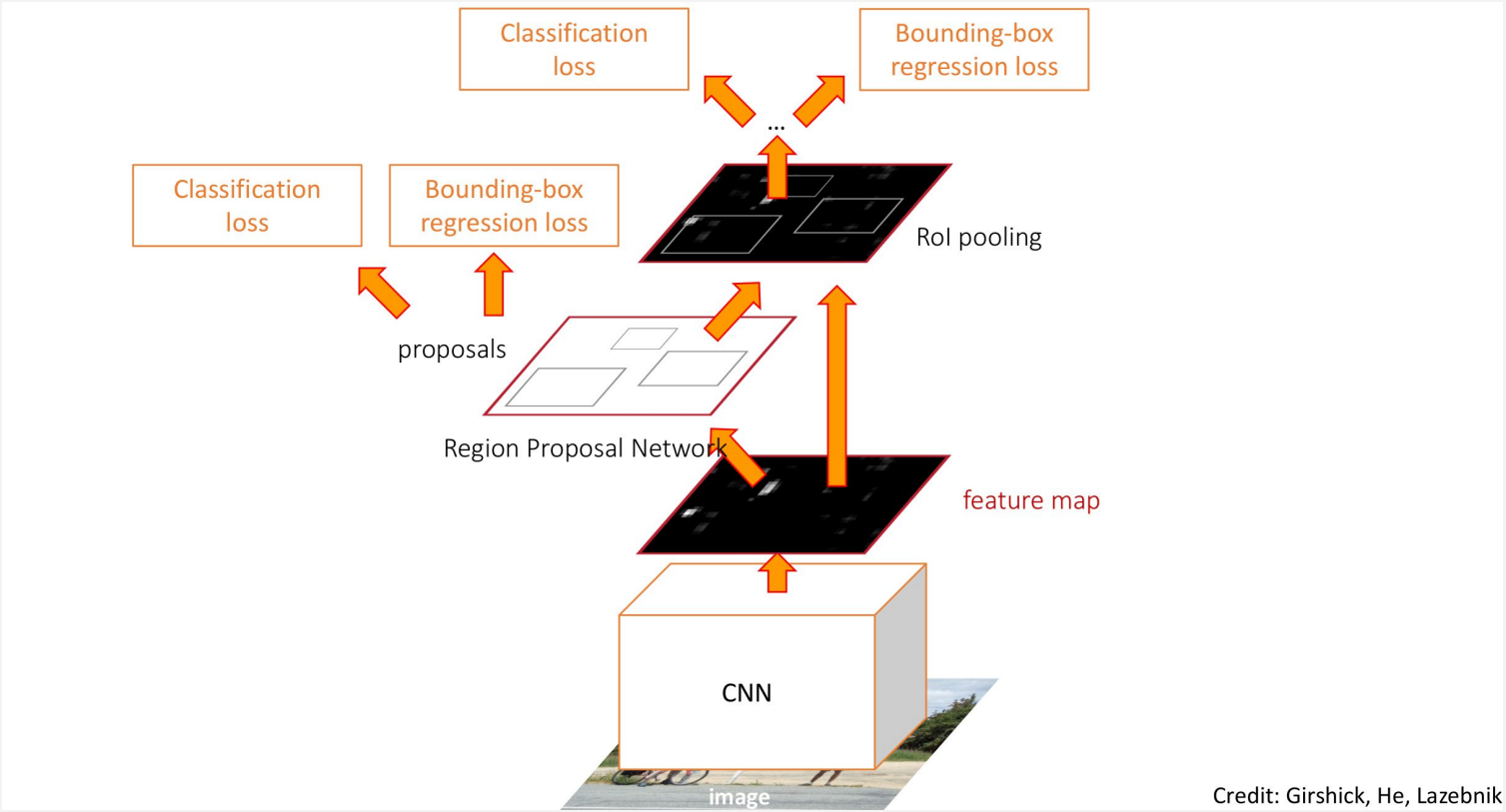
Fast RCNN



R. Girshick, Fast R-CNN, ICCV 2015

Credit: R. Girshick, Lazebnik

Faster RCNN



Credit: Girshick, He, Lazebnik

All the Details

Better Training

- Online hard example mining

Network Variants

- Feature Pyramid Network, RetinaNet

Ensemble

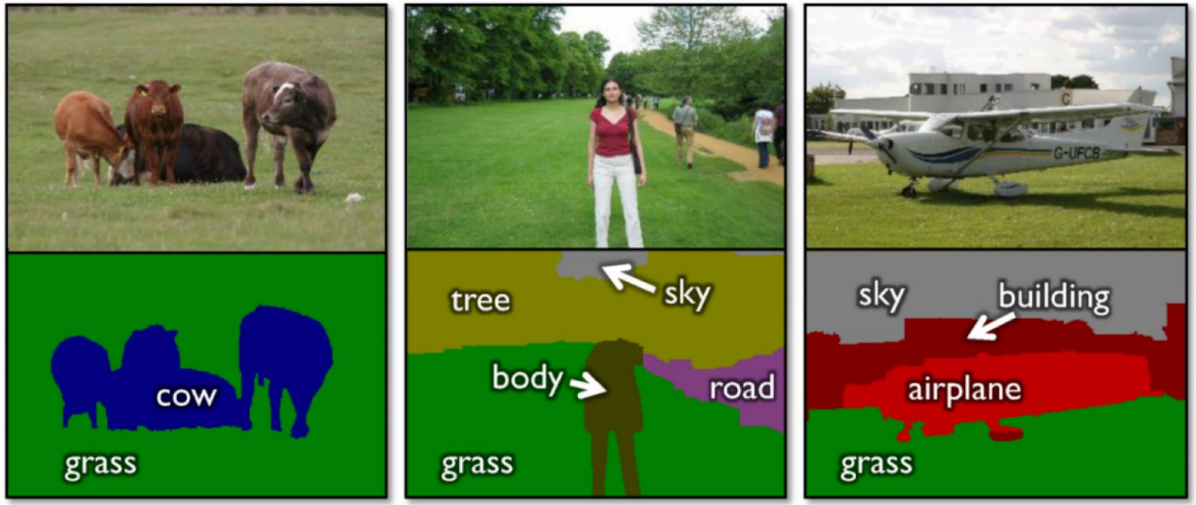
NMS

- Remove duplicate detections

Today - Object Detection and Segmentation

- Overview
- Object Detection
- **Segmentation**
- Detection + Segmentation: Instance Segmentation
- What's Next

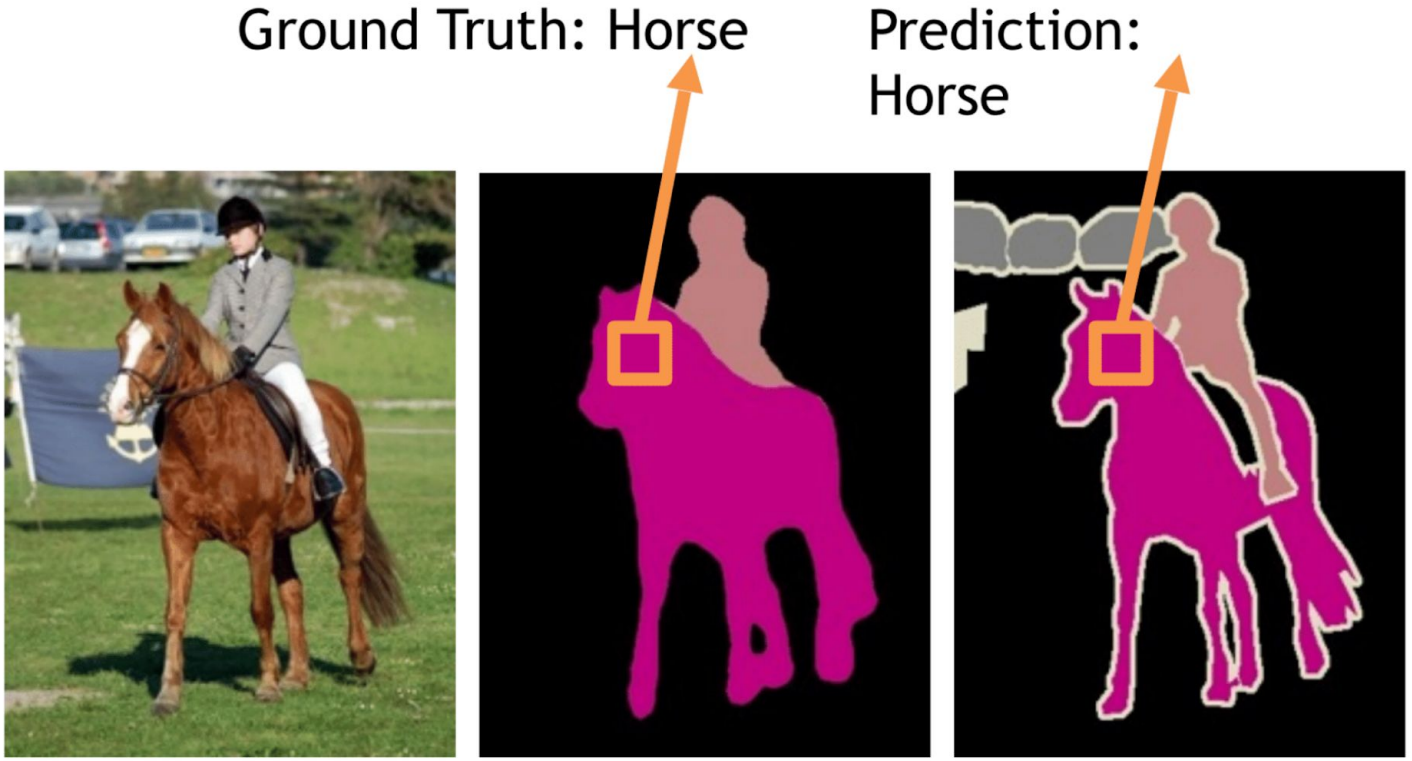
Task Formulation



object classes	building	grass	tree	cow	sheep	sky	airplane	water	face	car
bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat

Credit: Gupta

Evaluation Metric



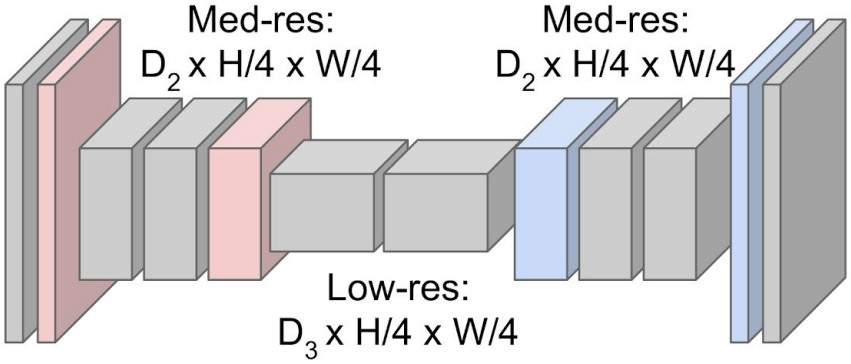
Pixel accuracy, Class accuracy, IOU

Slide credit: Abhinav Gupta

Network Structure for Dense Predictions

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Input:
3 x H x W



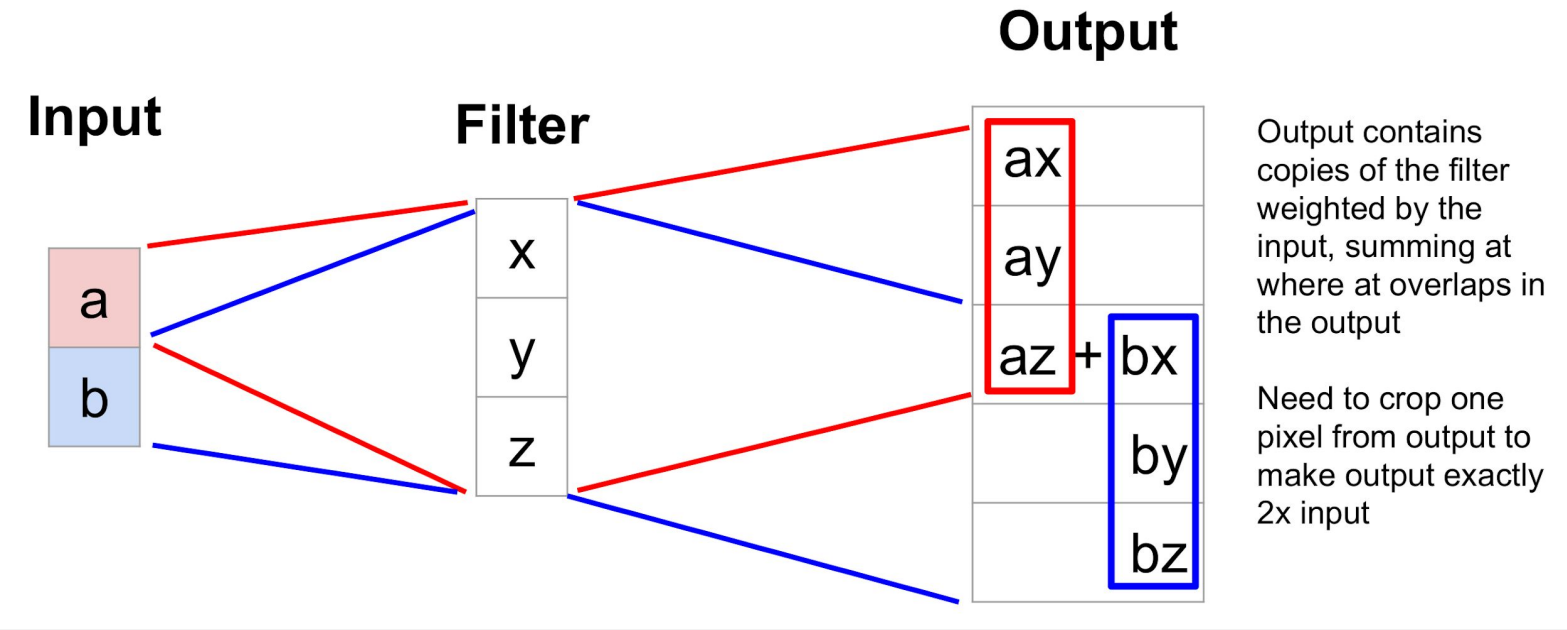
High-res:
D1 x H/2 x W/2

High-res:
D1 x H/2 x W/2

Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

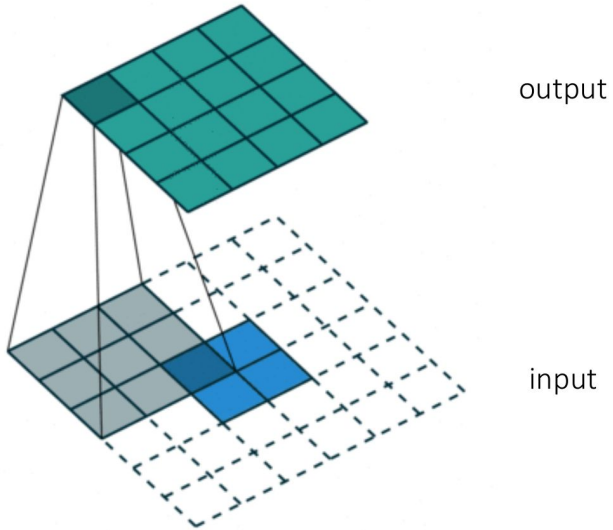
Upconvolution



Slides from Stanford cs231n

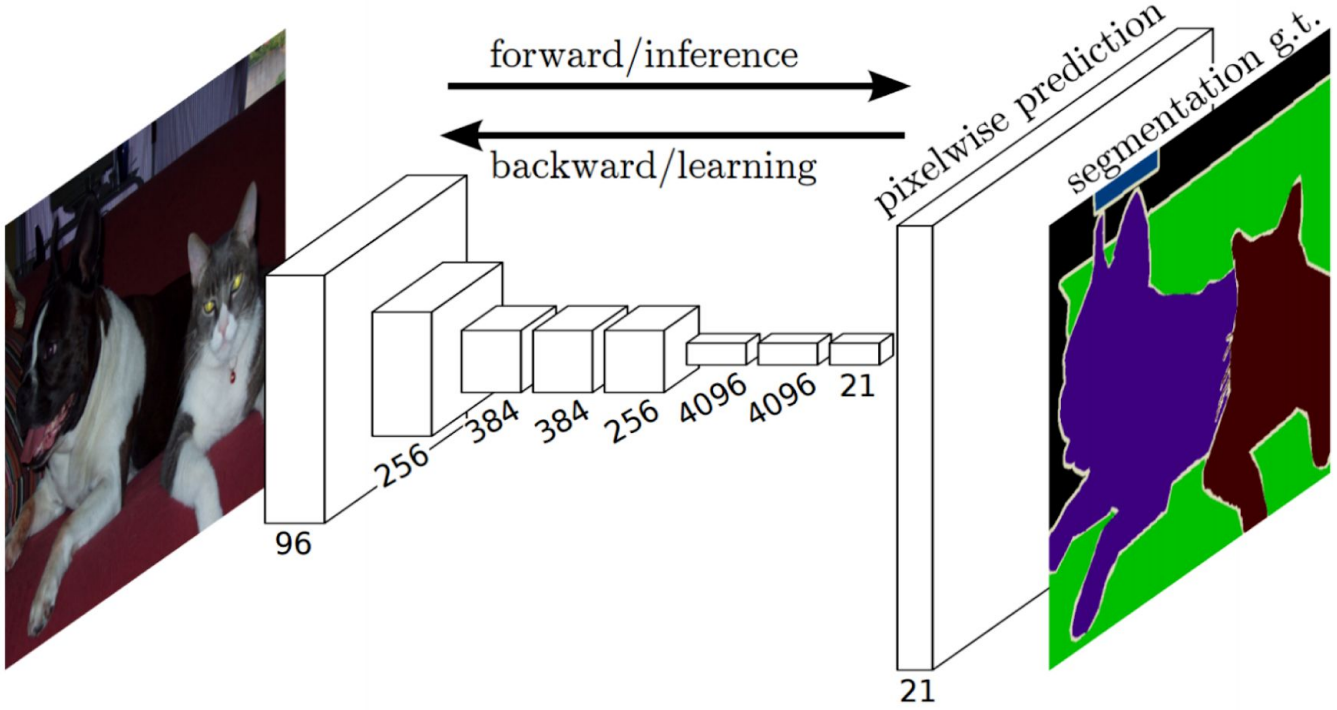
Upconvolution

Transposed convolution



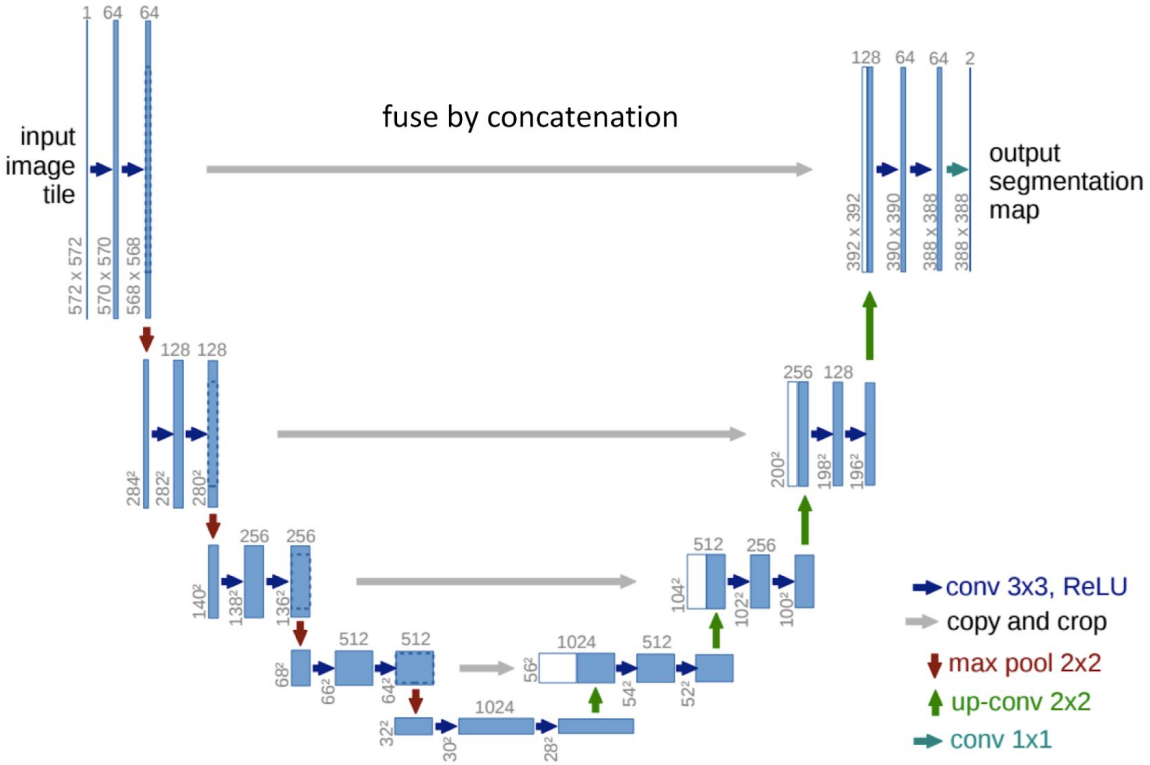
Credit: Lazebnik; V. Dumoulin and F. Visin, A guide to convolution arithmetic for deep learning, arXiv 2018

Fully Convolutional Networks



Fully convolutional Networksfor Semantic Segmentation, CVPR 2015

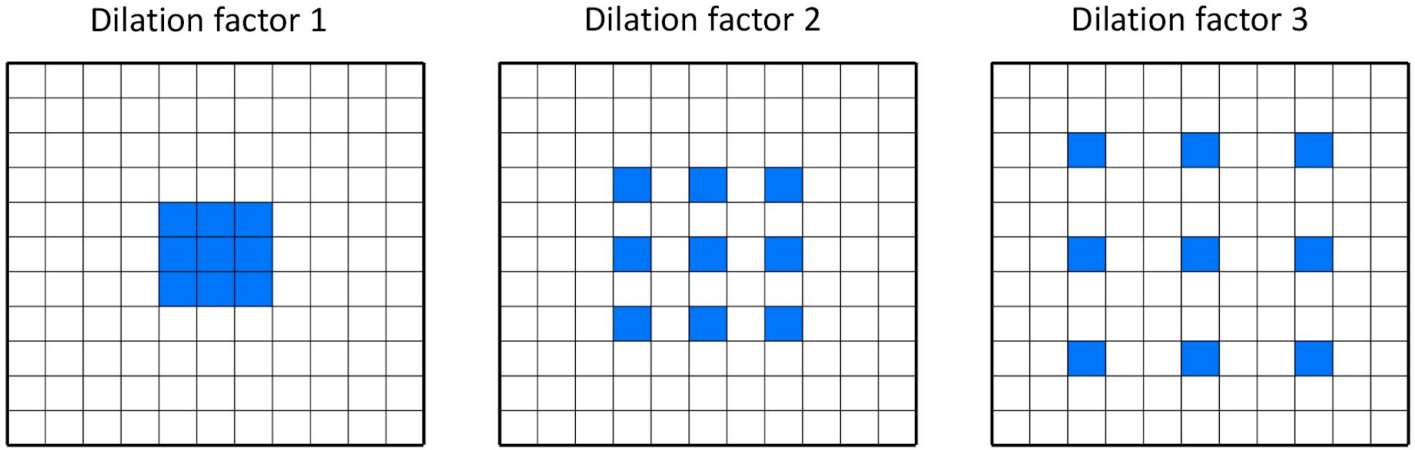
UNet



Credit: Lazebnik, Badrinarayanan et al., SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

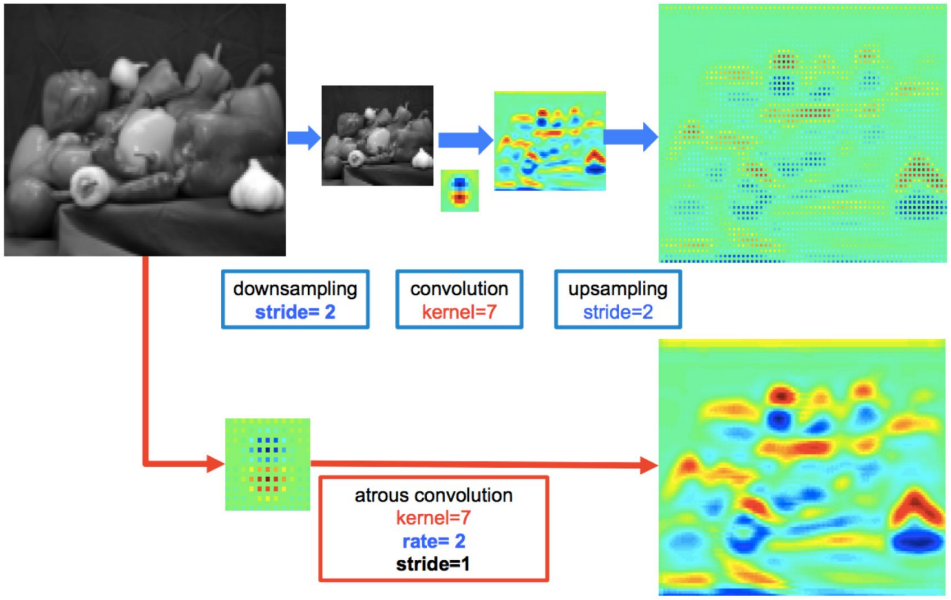
Dilated Convolutions

Instead of reducing spatial resolution of feature maps, use a large sparse filter



Credit: Lazebnik, image source

Deeplab

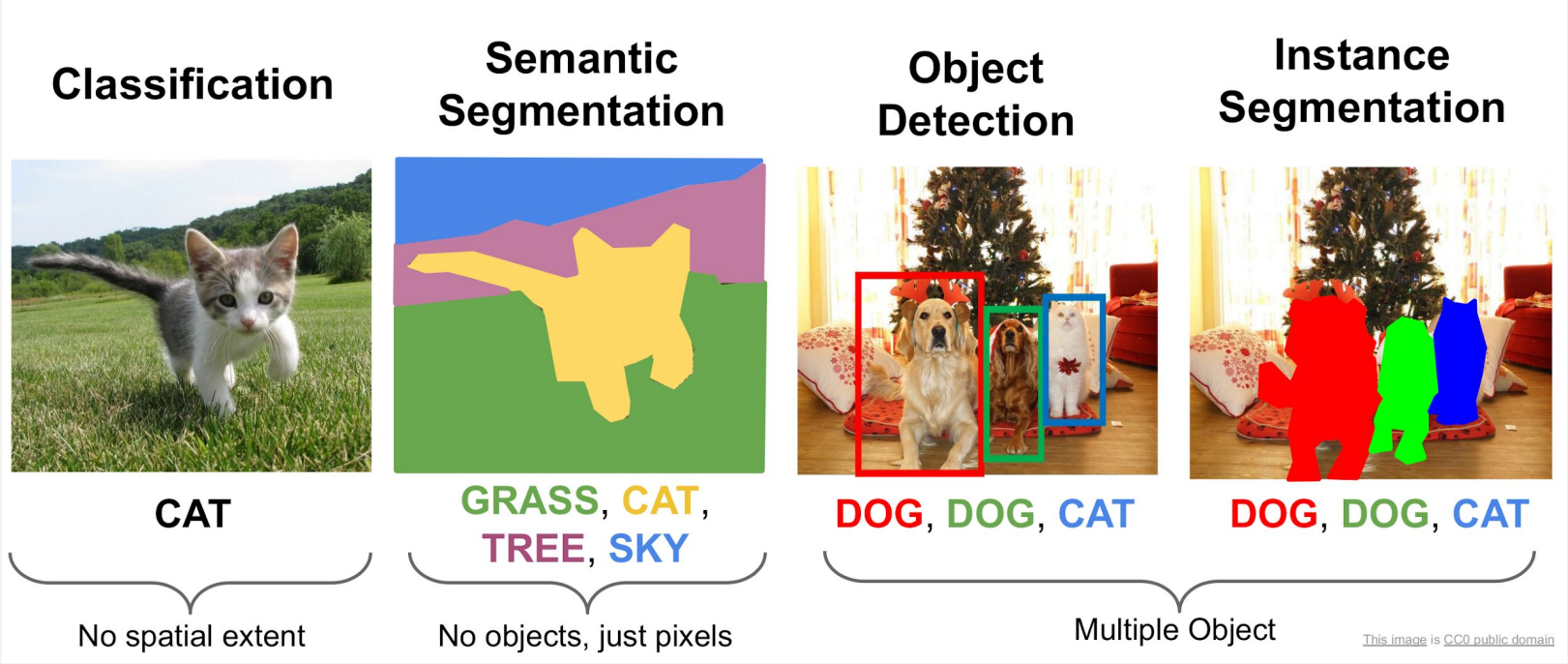


DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs

Today - Object Detection and Segmentation

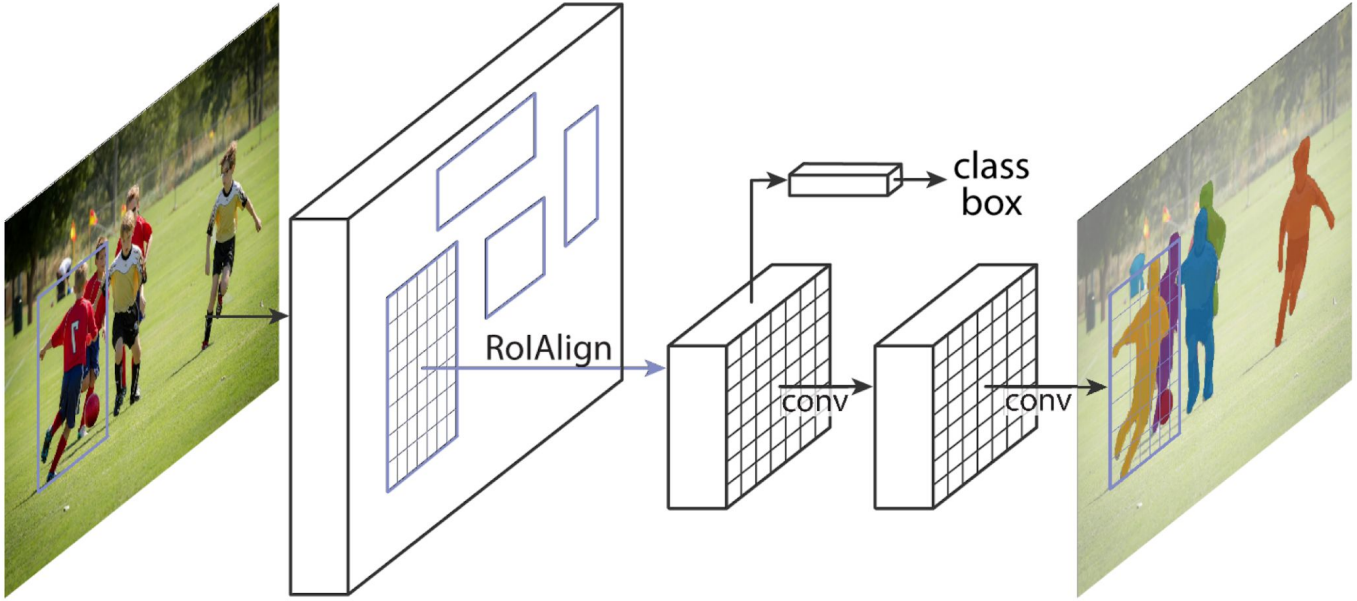
- Overview
- Object Detection
- Segmentation
- **Detection + Segmentation: Instance Segmentation**
- What's Next

Task Formulation



Slides from Stanford cs231n

Mask RCNN



Mask branch: separately predict segmentation for each possible class

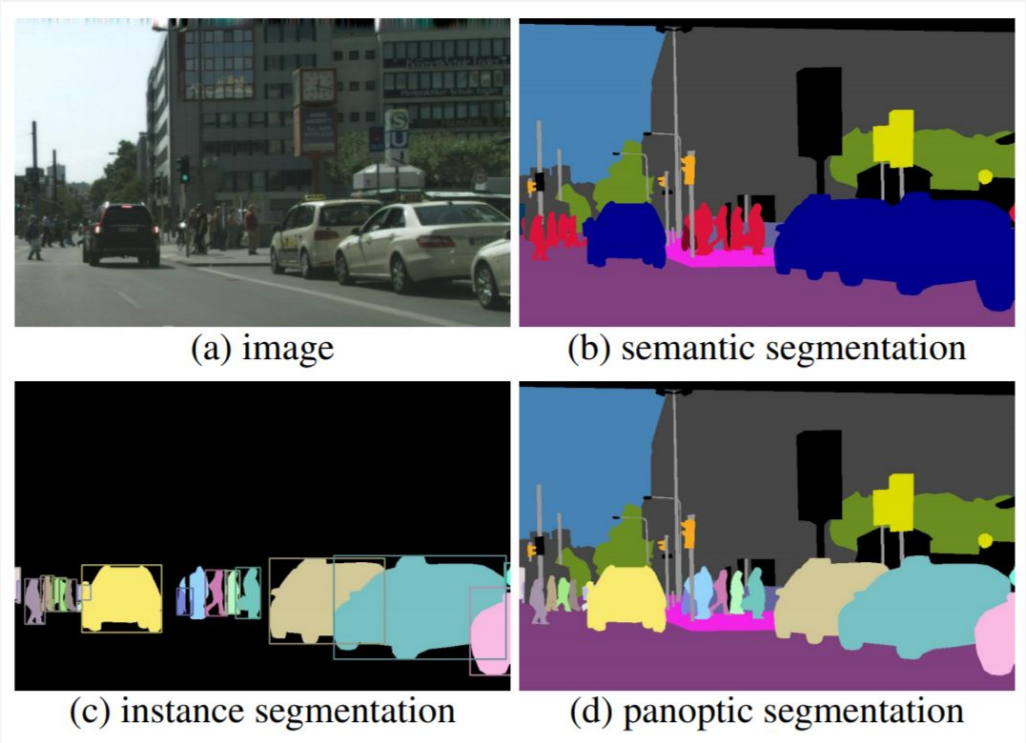
Slide credit: Kaiming He

Mask RCNN Results

Today - Object Detection and Segmentation

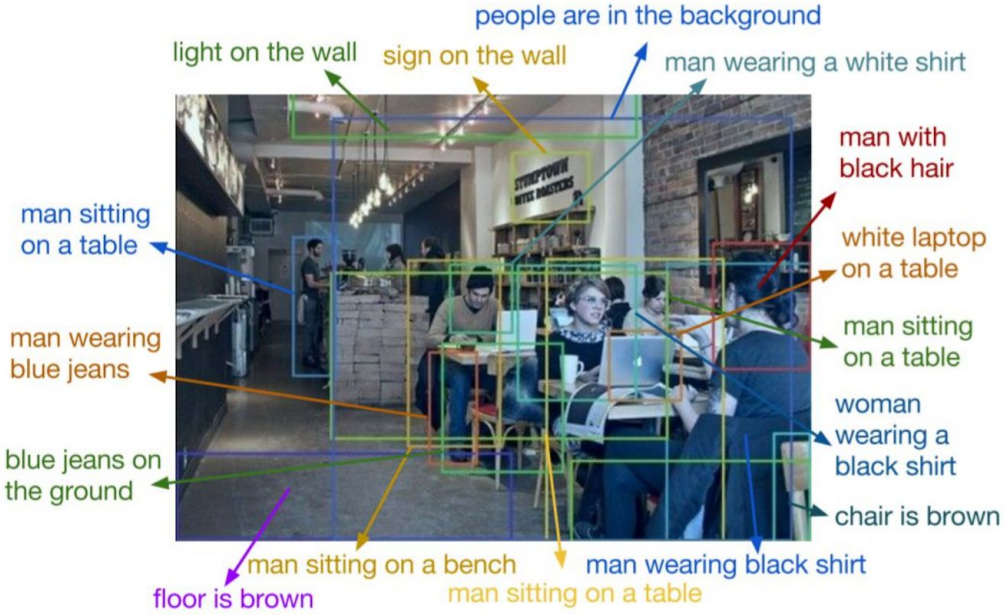
- Overview
- Object Detection
- Segmentation
- Detection + Segmentation: Instance Segmentation
- **What's Next**

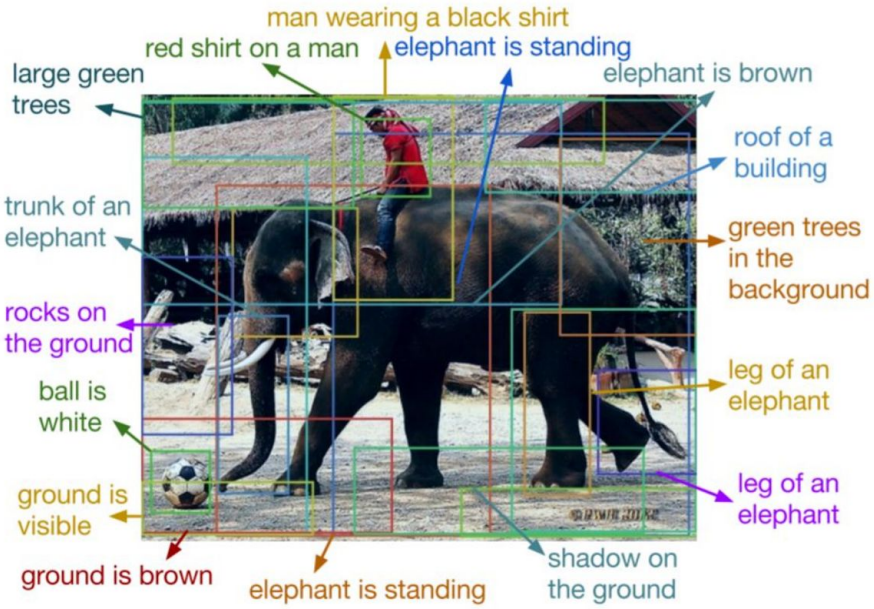
Panoptic Segmentation



Kirillov, Alexander, et al. "Panoptic segmentation." *CVPR*. 2019.

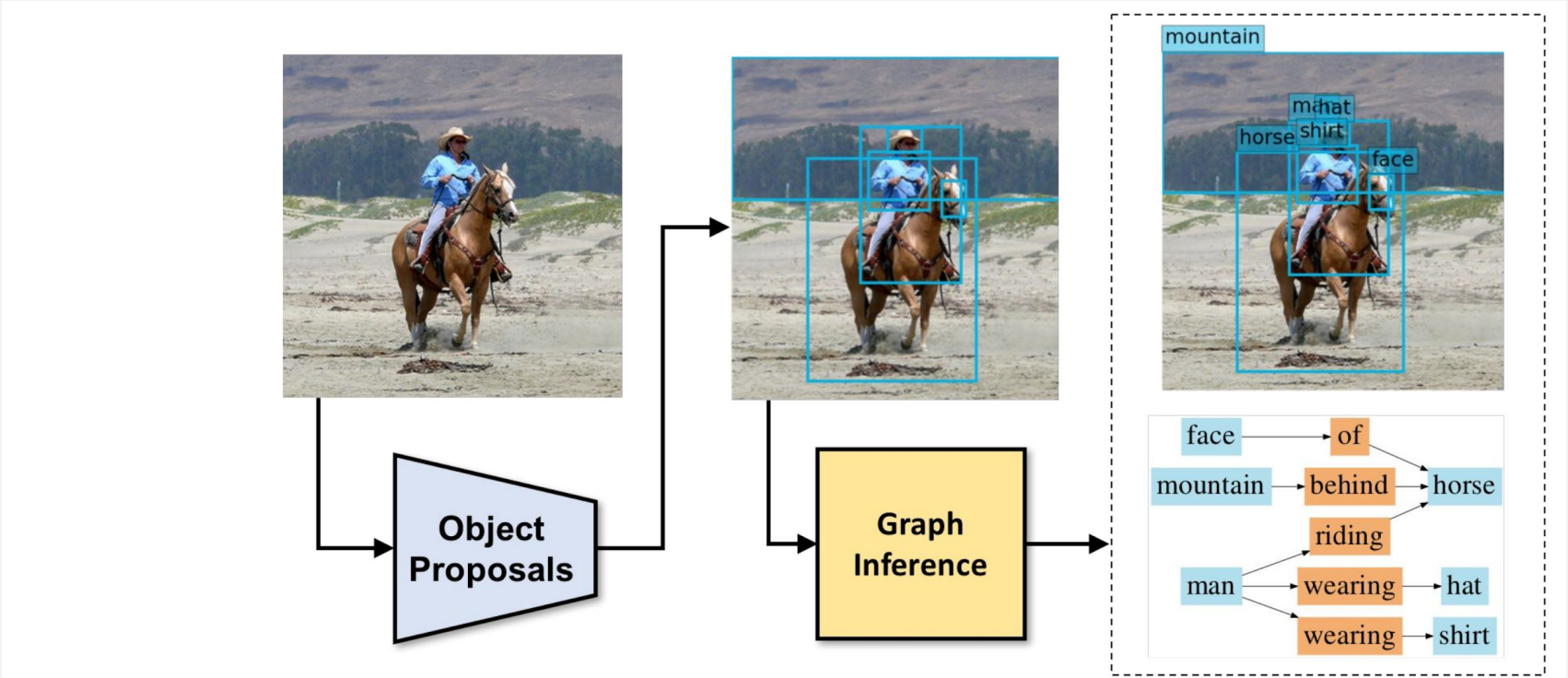
Captioning





Johnson, Karpathy, and Fei-Fei, "DenseCap: Fully Convolutional Localization Networks for Dense Captioning", CVPR 2016

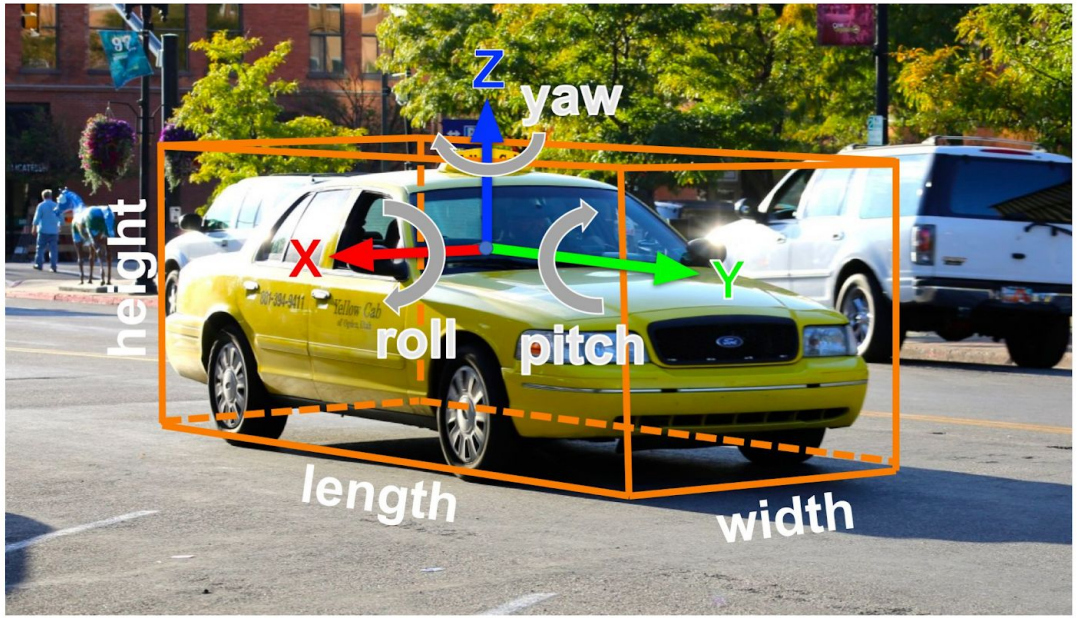
Scene Graph



Xu, Zhu, Choy, and Fei-Fei, "Scene Graph Generation by Iterative Message Passing", CVPR 2017

Slides from Stanford cs231n

3D Detection



2D Object Detection:
2D bounding box
(x, y, w, h)

3D Object Detection:
3D oriented bounding box
(x, y, z, w, h, l, r, p, y)

Simplified bbox: no roll & pitch

Much harder problem than 2D object detection!

Summary - Object Detection and Segmentation

- Overview
- Object Detection
- Segmentation
- Detection + Segmentation: Instance Segmentation
- What's Next