

→ Mask R-CNN

OBJECT INSTANCE SEGMENTATION AND HUMAN POSE ESTIMATION

Kaiming He
RESEARCH SCIENTIST

FACEBOOK AI RESEARCH (FAIR)

Georgia Gkioxari
POSTDOC

Piotr Dollár
RESEARCH SCIENTIST

Ross Girshick
RESEARCH SCIENTIST

Classic Computer Vision Problems



Source: PASCAL Dataset

Image classification

- ✓ boat
- ✓ person

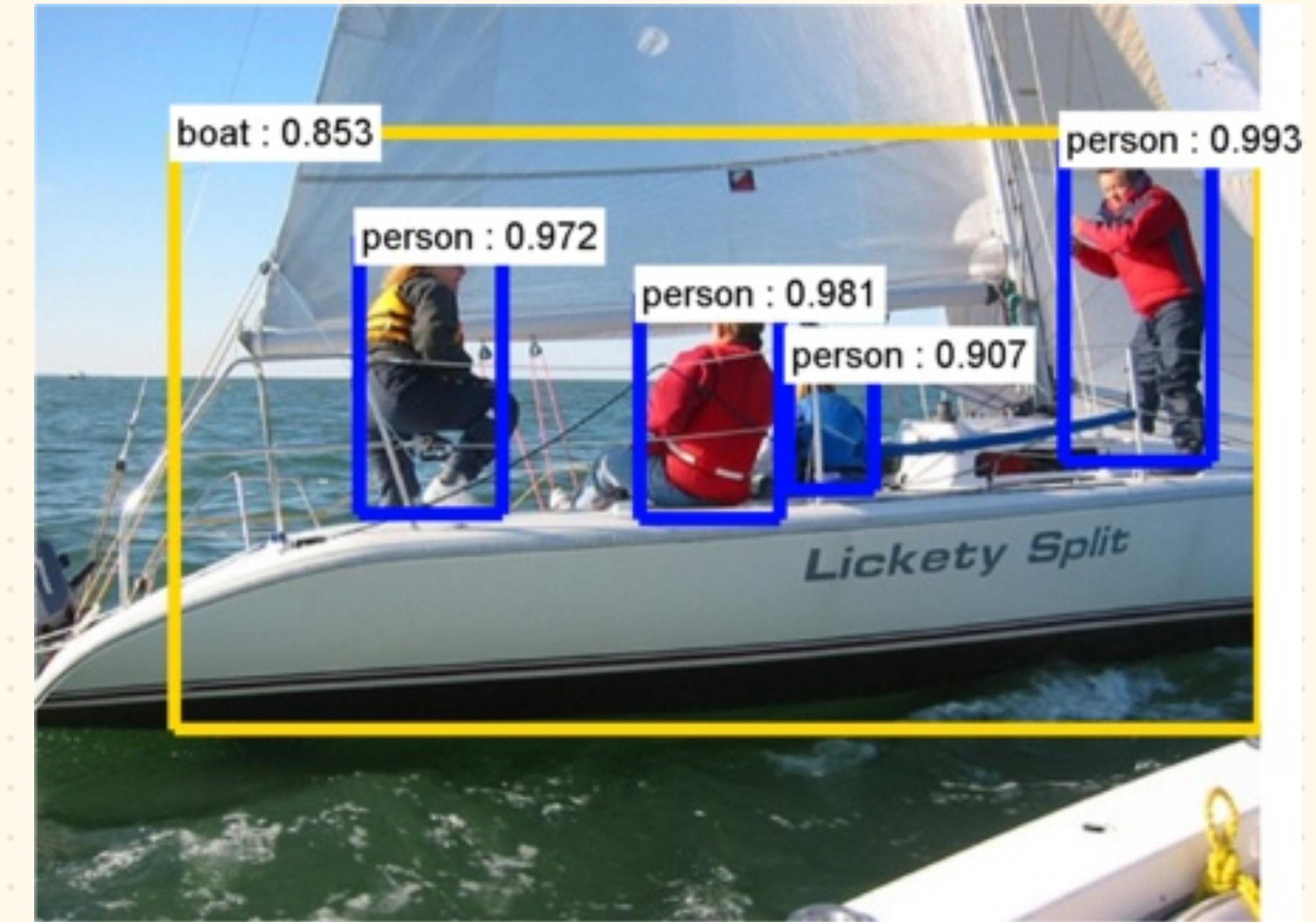
Classic Computer Vision Problems



Source: PASCAL Dataset

Image classification

- ✓ boat
- ✓ person

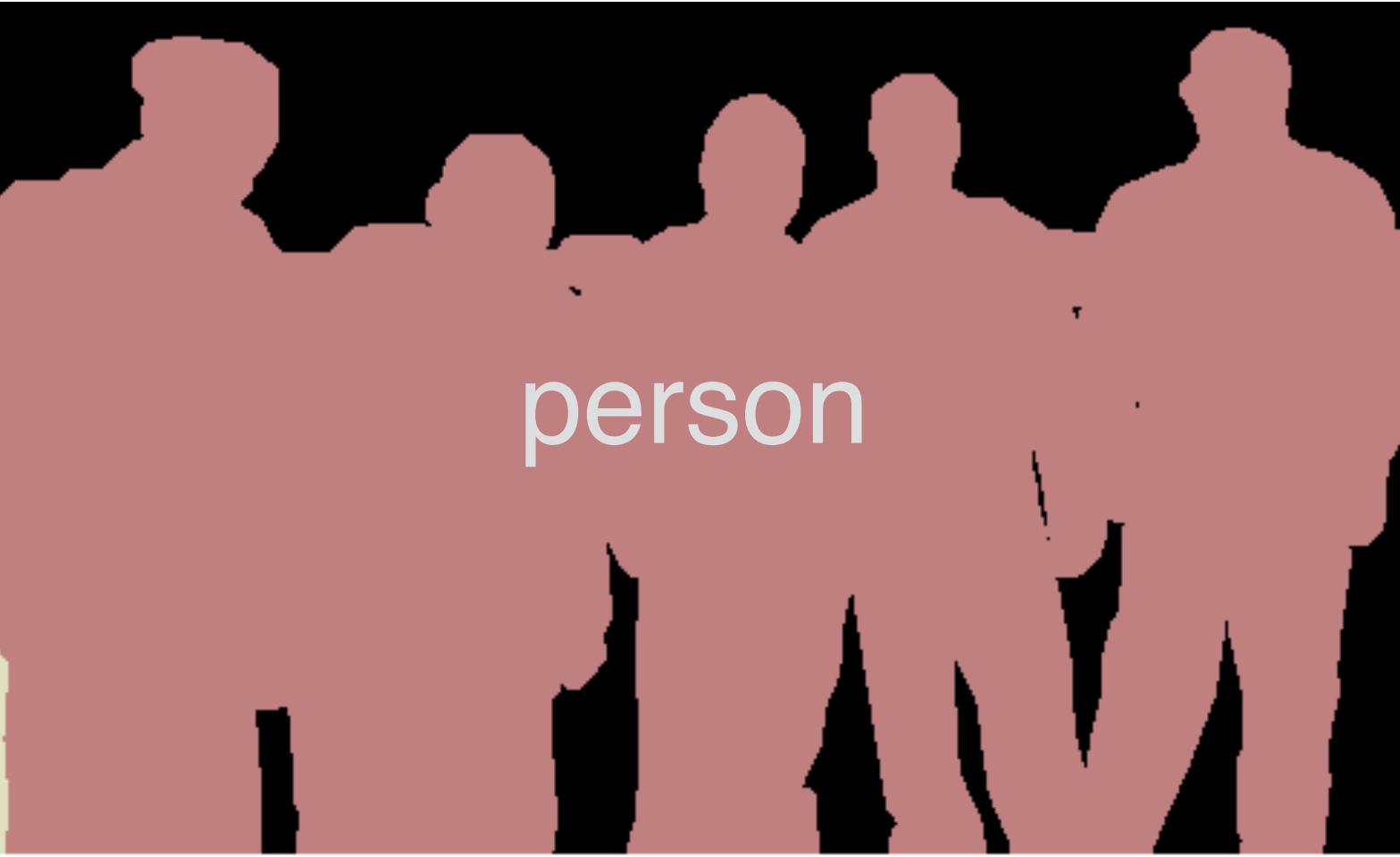


Object detection

Semantic Segmentation

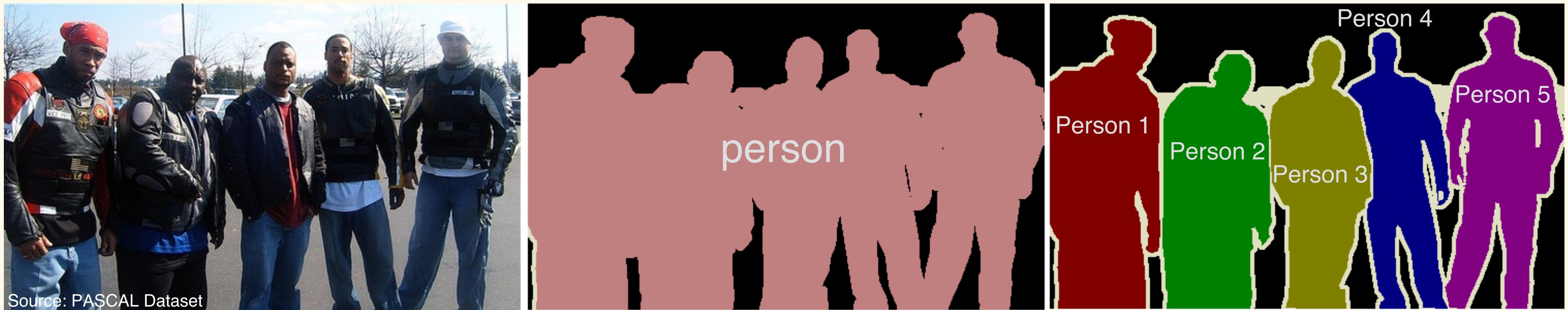


Source: PASCAL Dataset



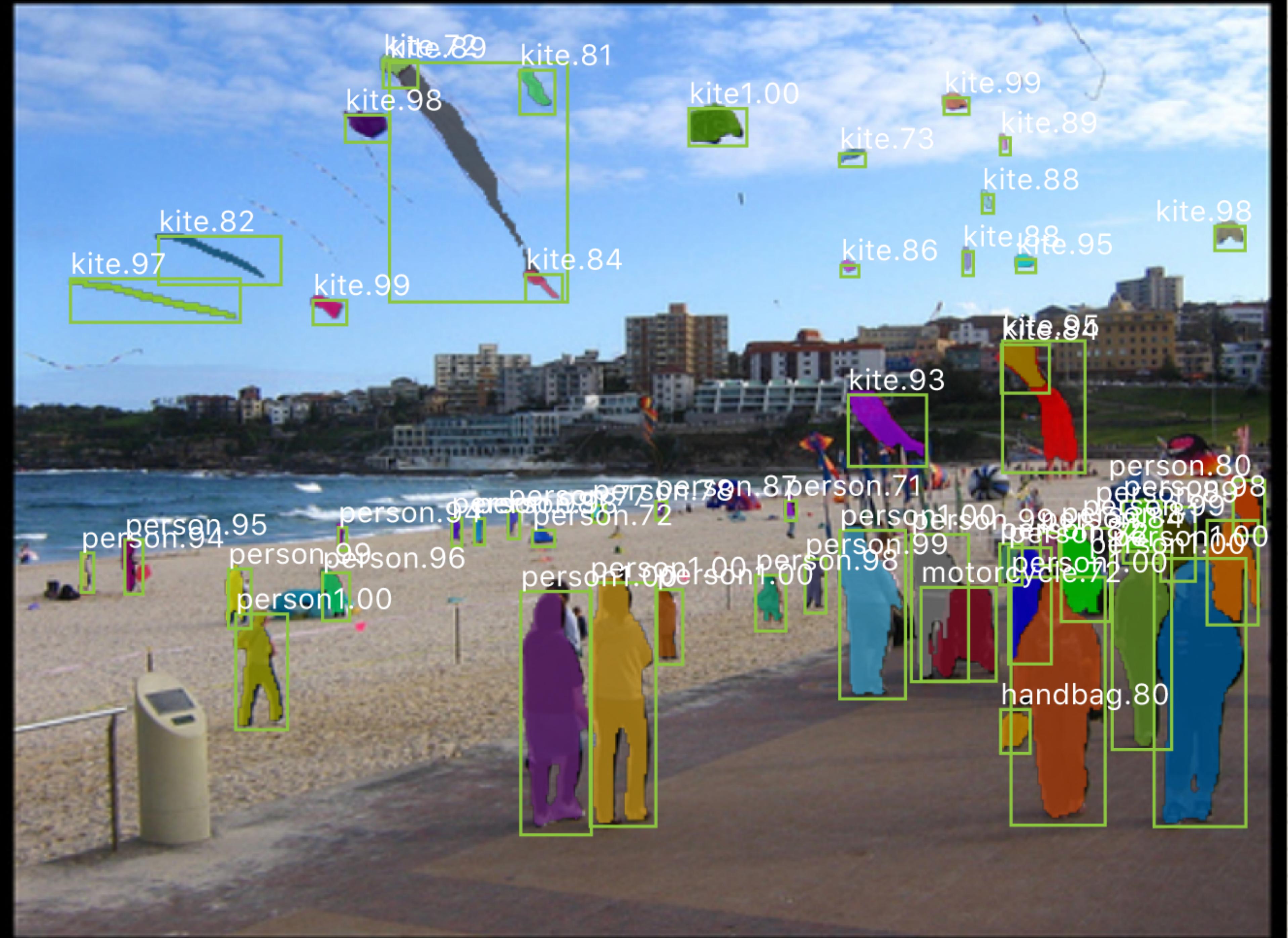
Semantic segmentation
(pixel-level classification)

The Instance Segmentation Task



Semantic segmentation
(pixel-level classification)

Instance segmentation
(pixel-level detection)





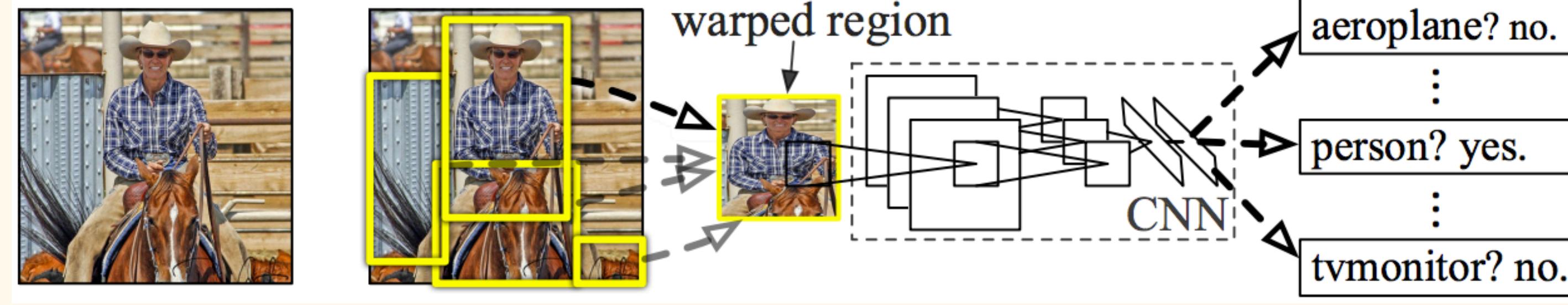
Mask R-CNN

TALK OUTLINE

- Mask R-CNN
 - Object instance segmentation
 - Human pose estimation
- Role of Caffe2 in our research
- Conclusions

Object Detection: R-CNN

REGION-BASED CONVOLUTION NEURAL NETWORK



Image

Region
proposals
(External
algorithm)

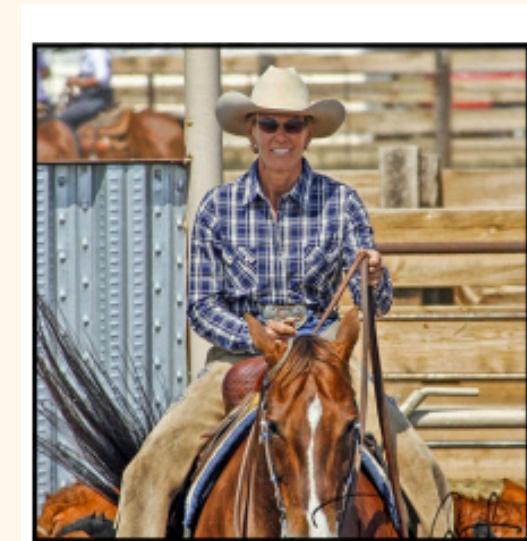
Per-region
classification by a CNN

SOURCE: GIRSHICK, DONAHUE, DARRELL, MALIK.

RICH FEATURE HIERARCHIES FOR ACCURATE OBJECT DETECTION AND SEMANTIC SEGMENTATION. CVPR 2014

Object Detection: R-CNN

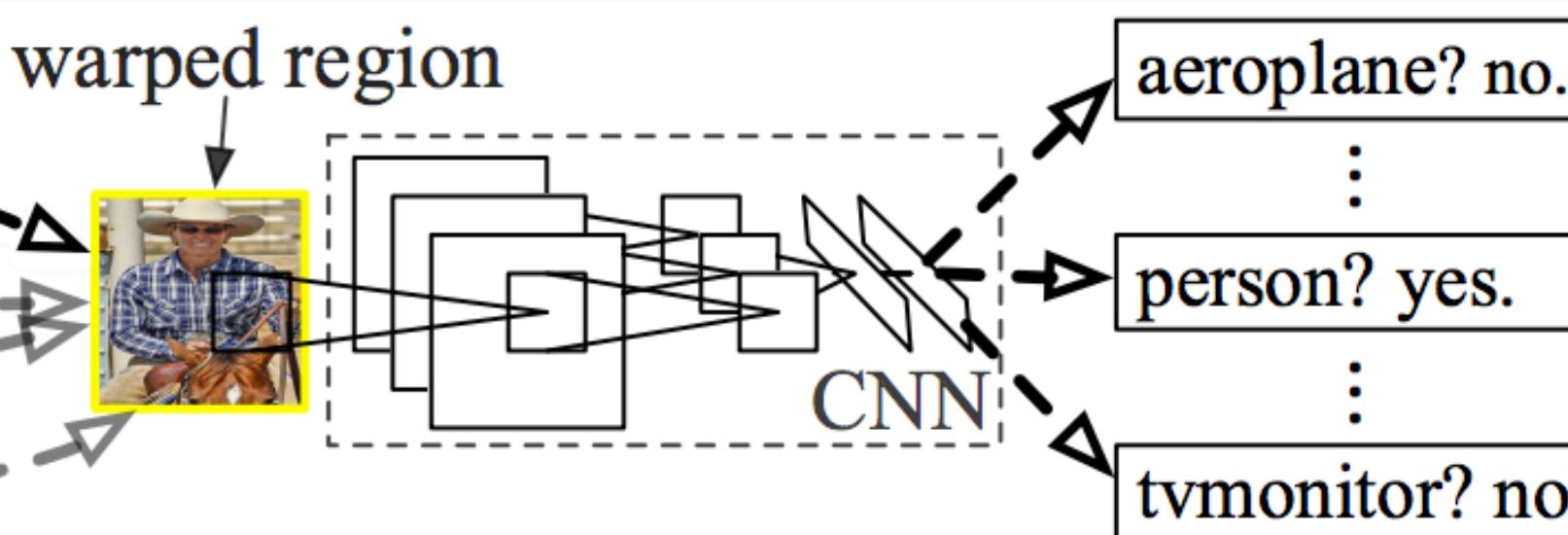
REGION-BASED CONVOLUTION NEURAL NETWORK



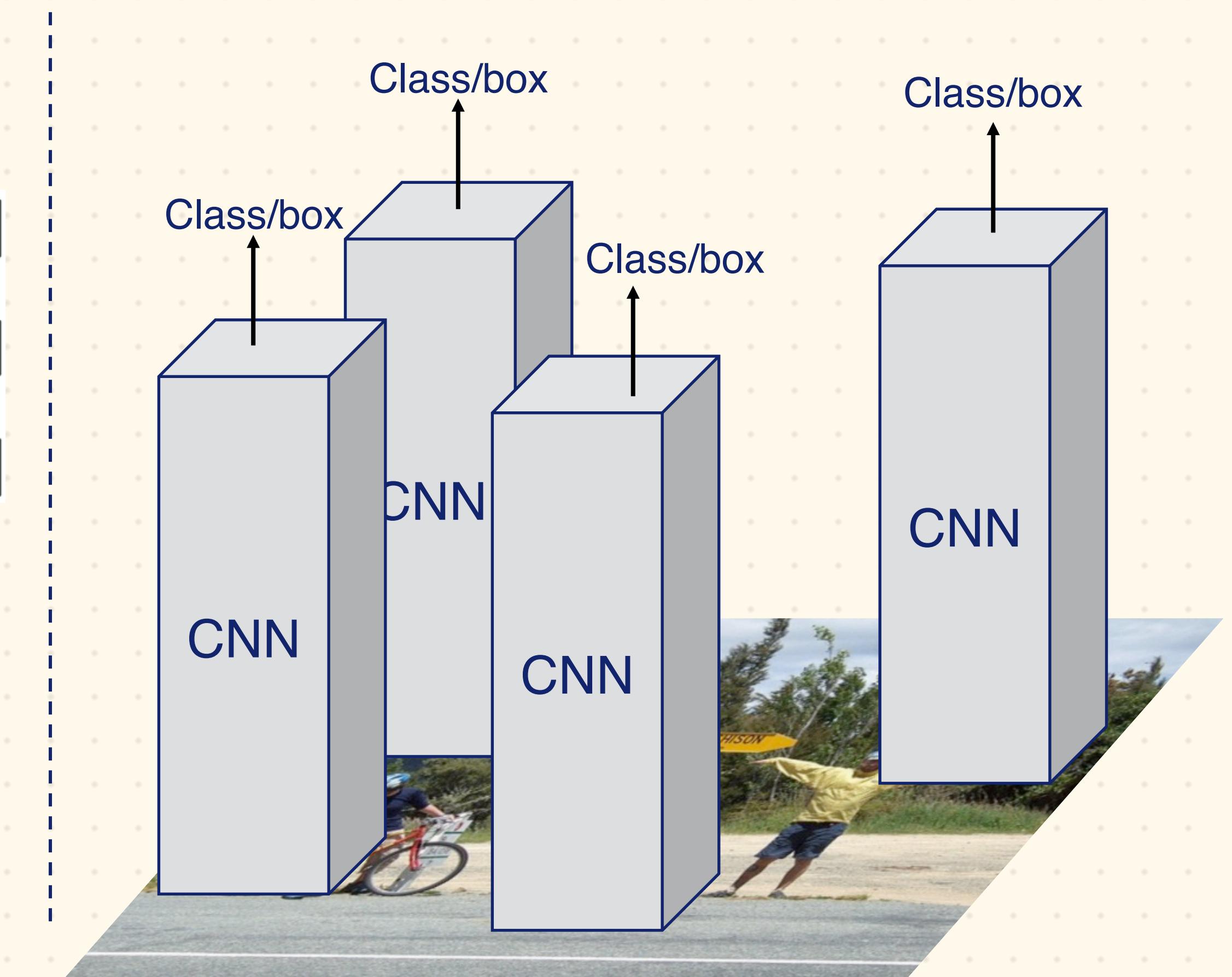
Image



Region
proposals
(External
algorithm)



Per-region
classification by a CNN



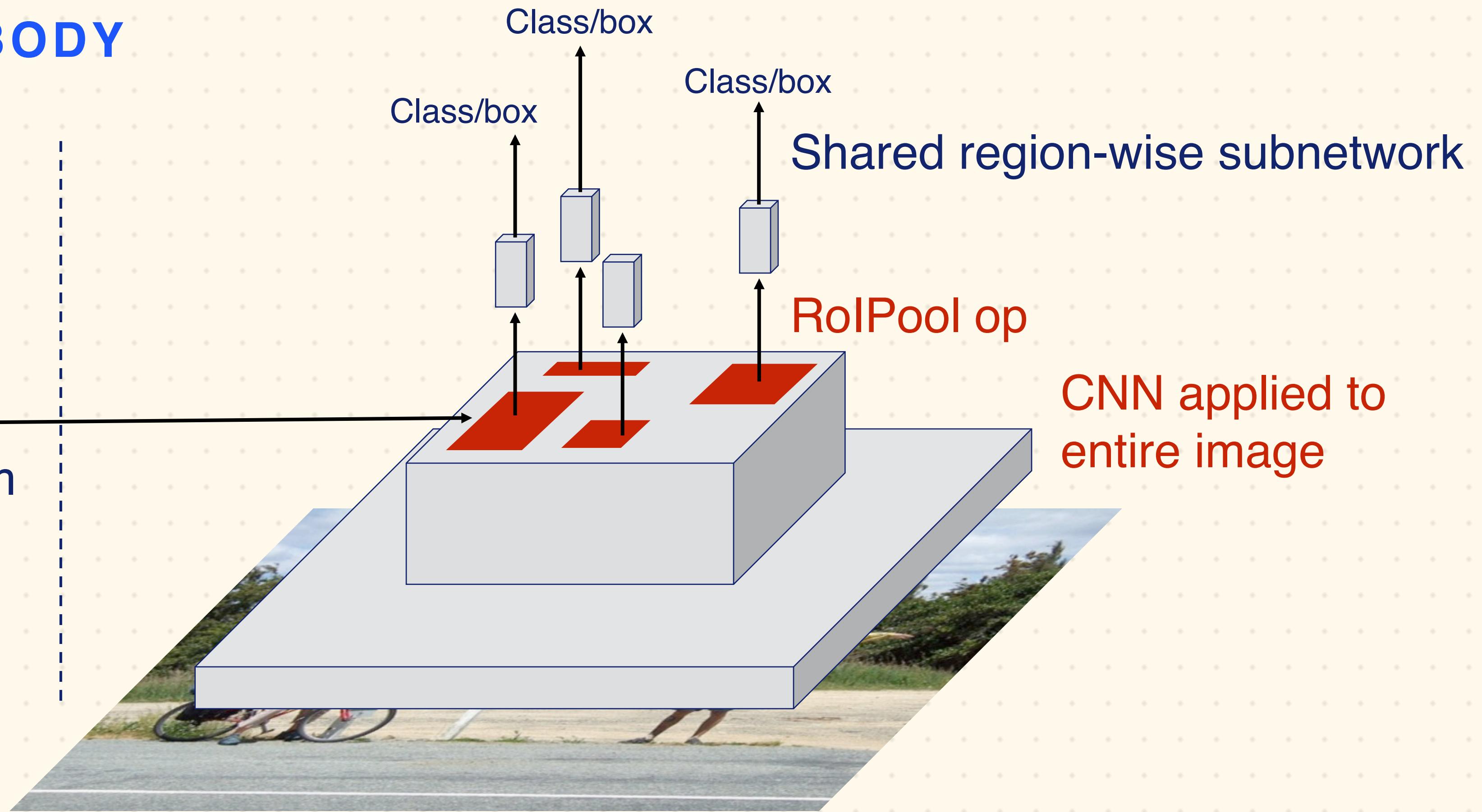
SOURCE: GIRSHICK, DONAHUE, DARRELL, MALIK.

RICH FEATURE HIERARCHIES FOR ACCURATE OBJECT DETECTION AND SEMANTIC SEGMENTATION. CVPR 2014

Fast R-CNN

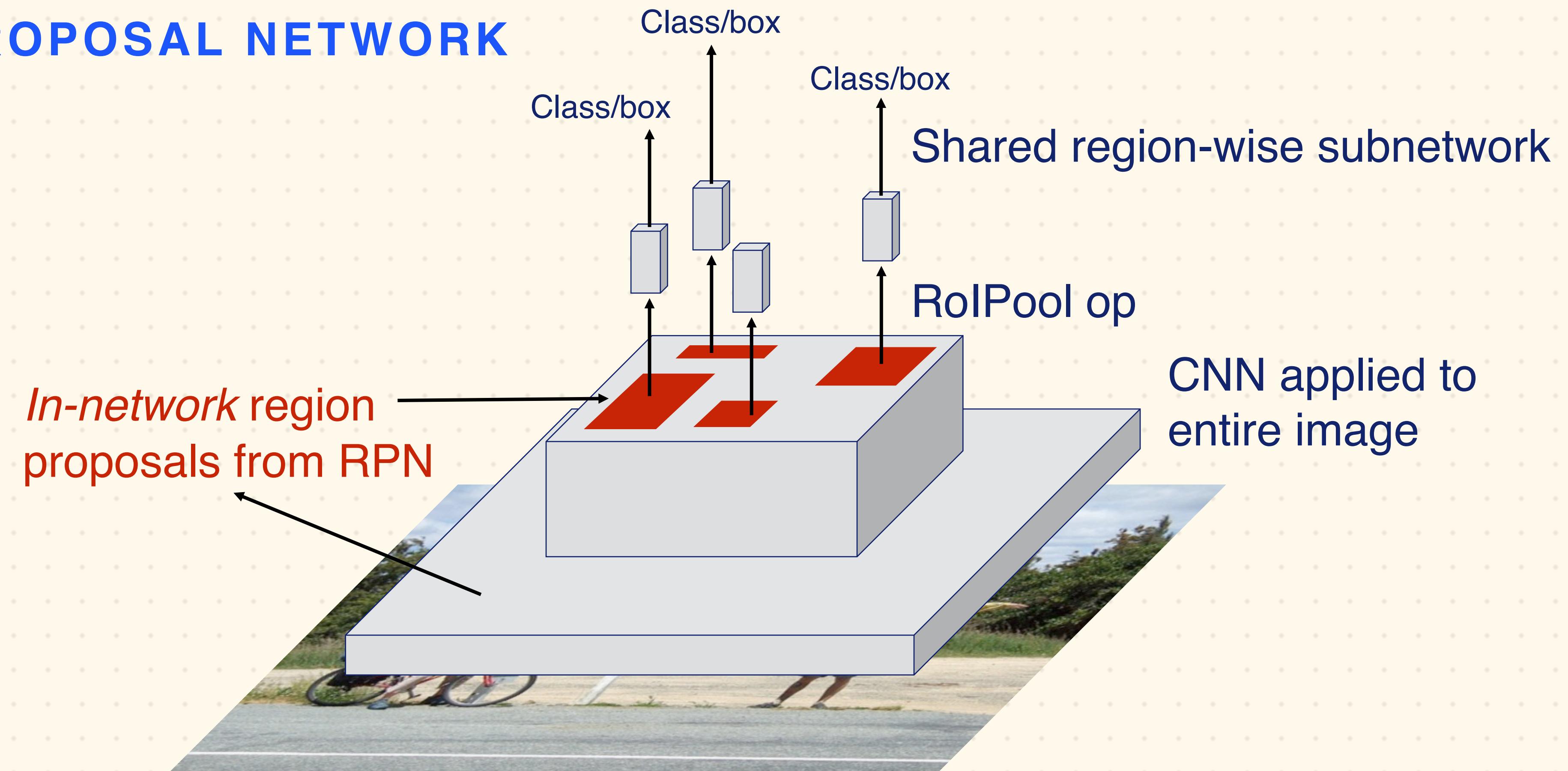
A SHARED CNN BODY

*External region
proposal algorithm
(same as R-CNN)*



Faster R-CNN

REGION PROPOSAL NETWORK



SOURCE: REN, HE, GIRSHICK, SUN.

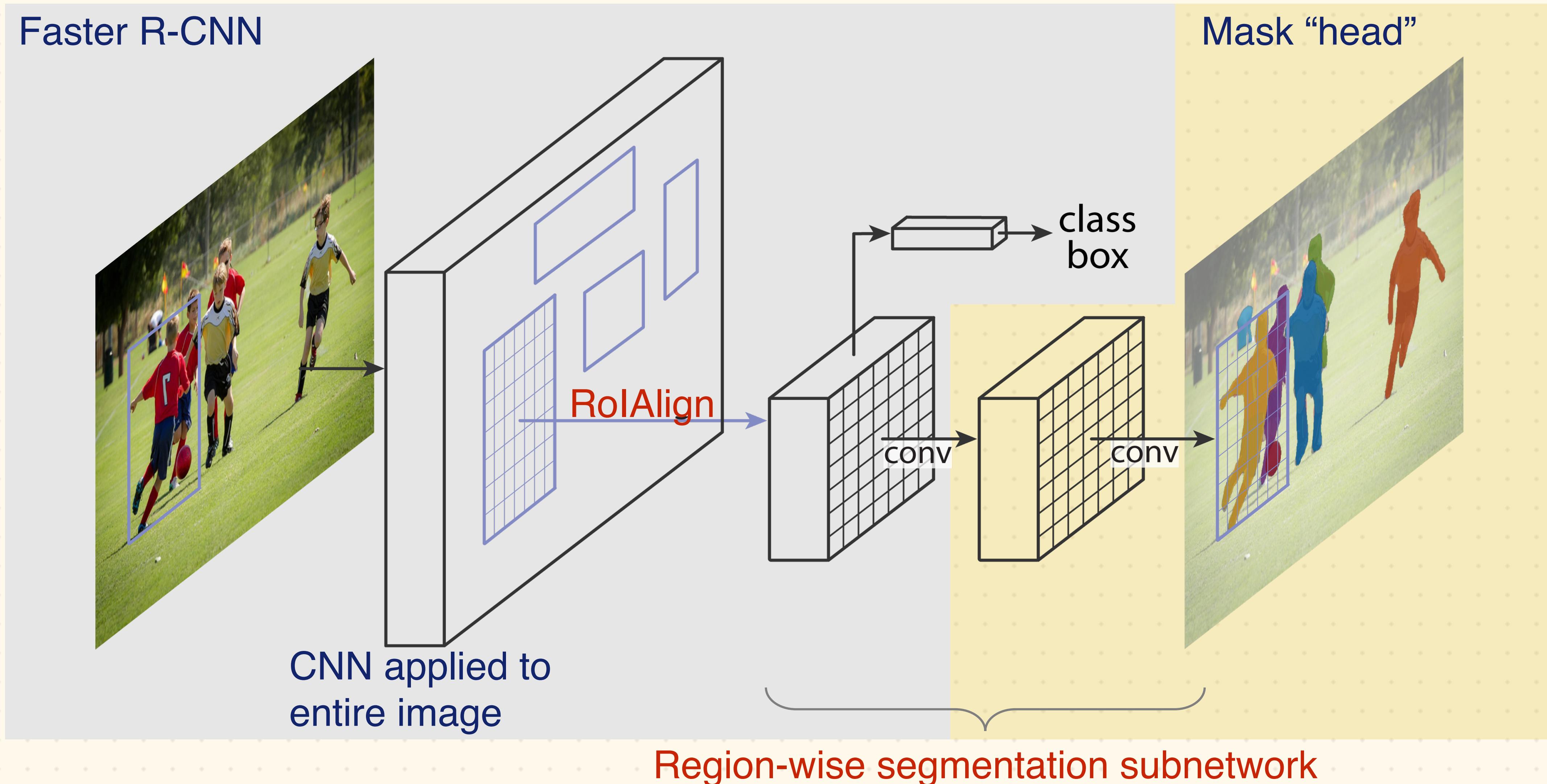
FASTER R-CNN: TOWARDS REAL-TIME OBJECT DETECTION WITH REGION PROPOSAL NETWORKS. NIPS 2015

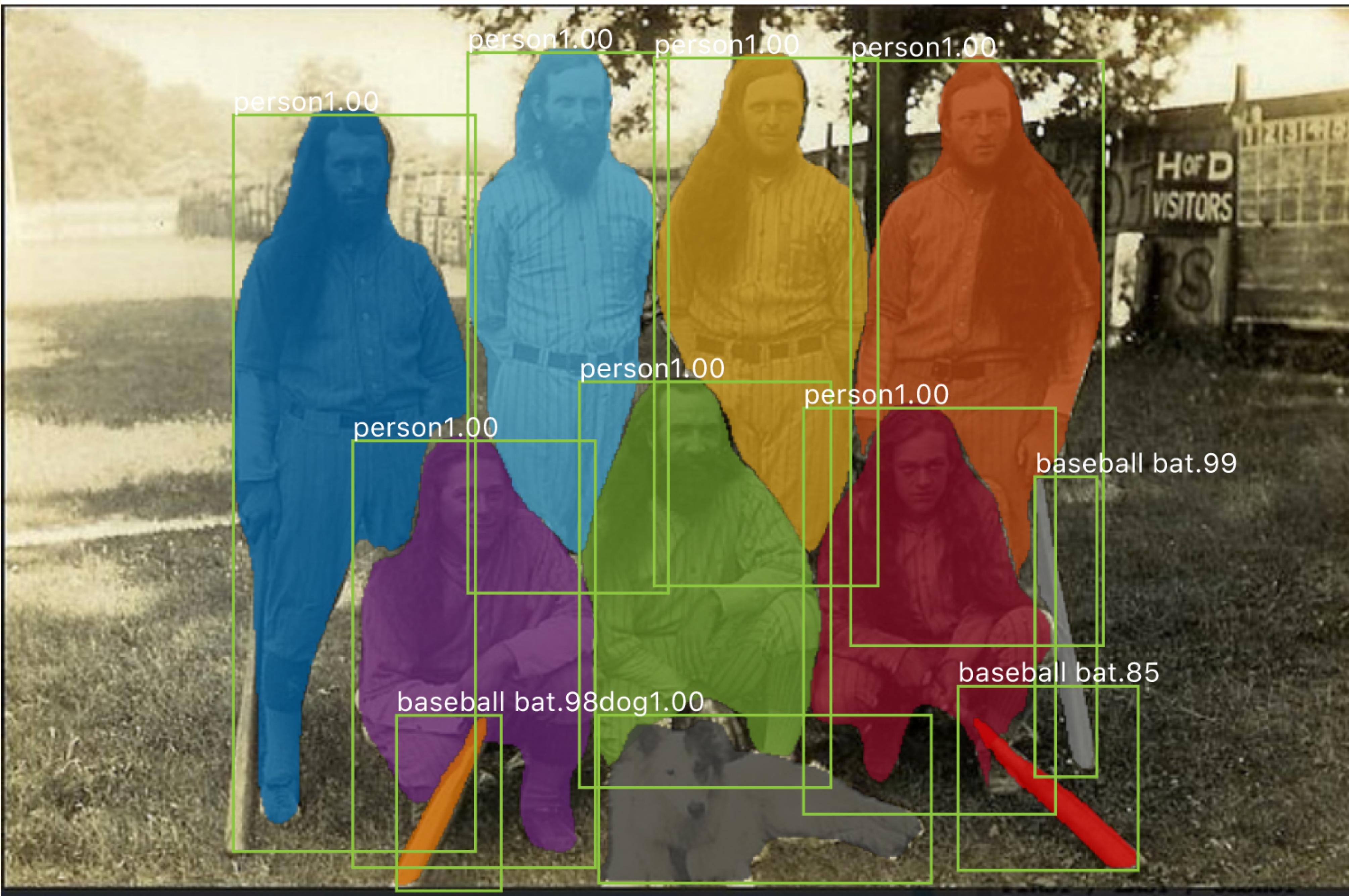
Mask R-CNN for Instance Segmentation

OVERVIEW

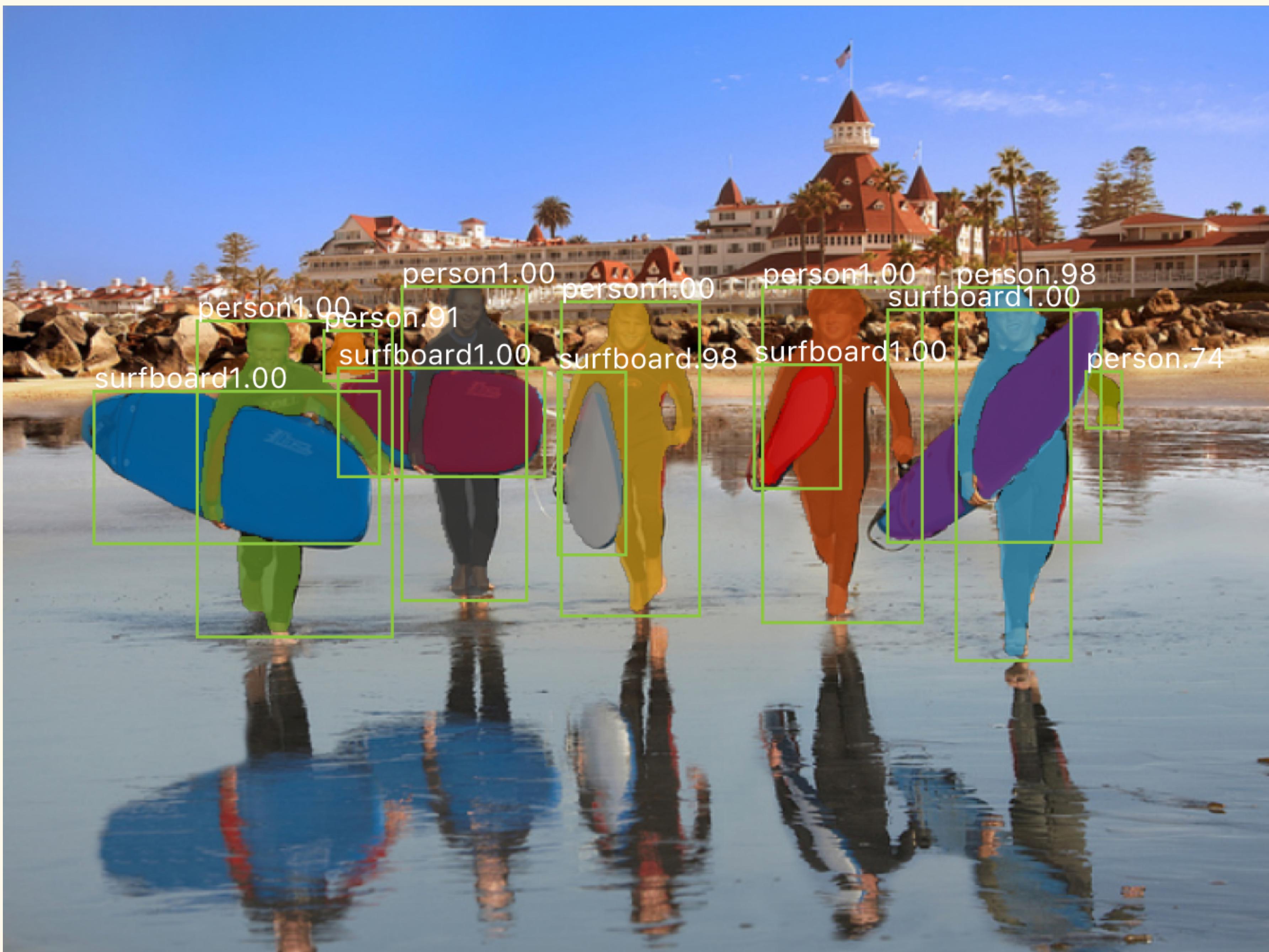
- An extension of Faster R-CNN
- Surprisingly simple
- Fast: 200 ms / im
- Accurate: state of the art on COCO

Mask R-CNN for Instance Segmentation

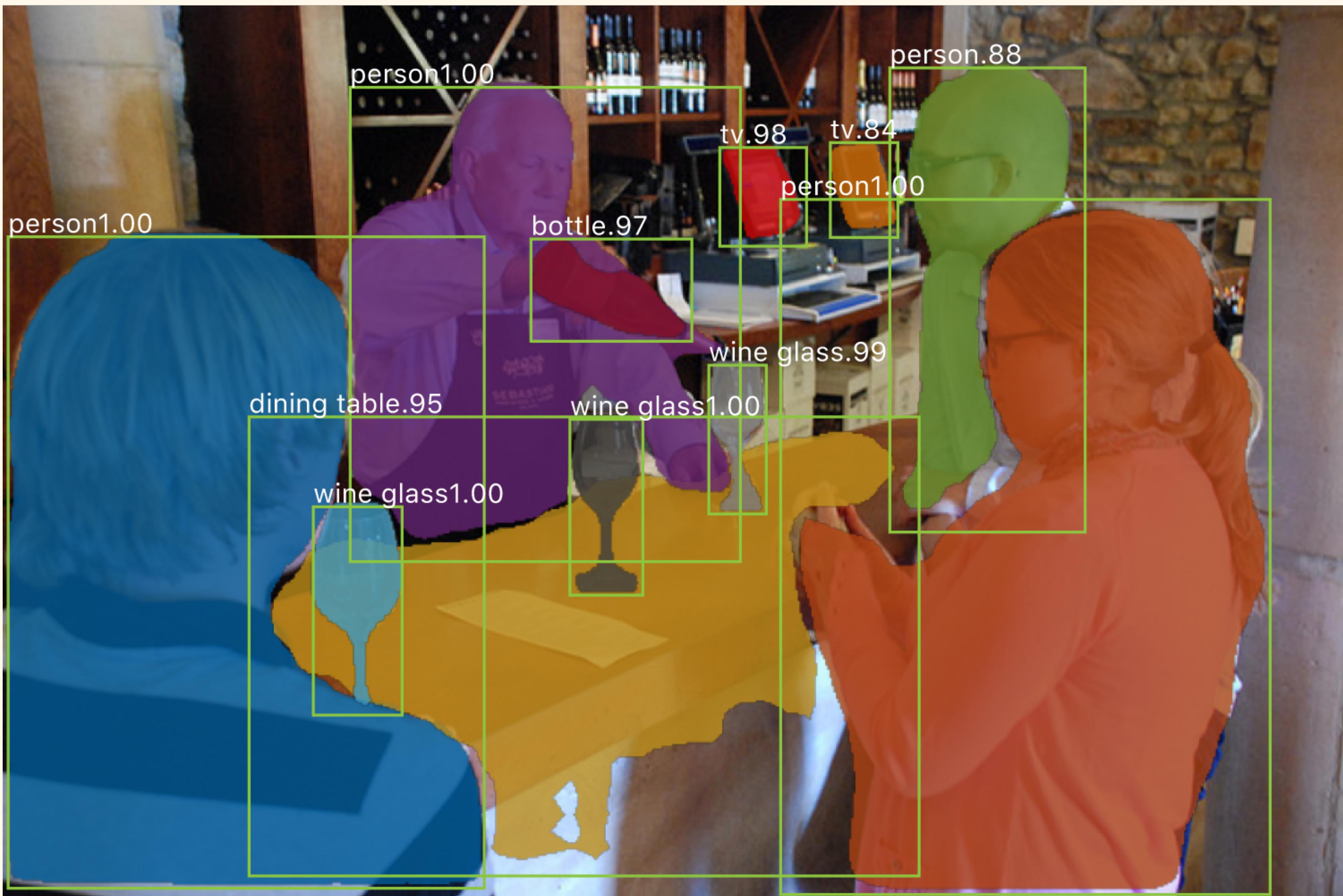




Mask R-CNN results on COCO



Mask R-CNN results on COCO



Mask R-CNN results on COCO

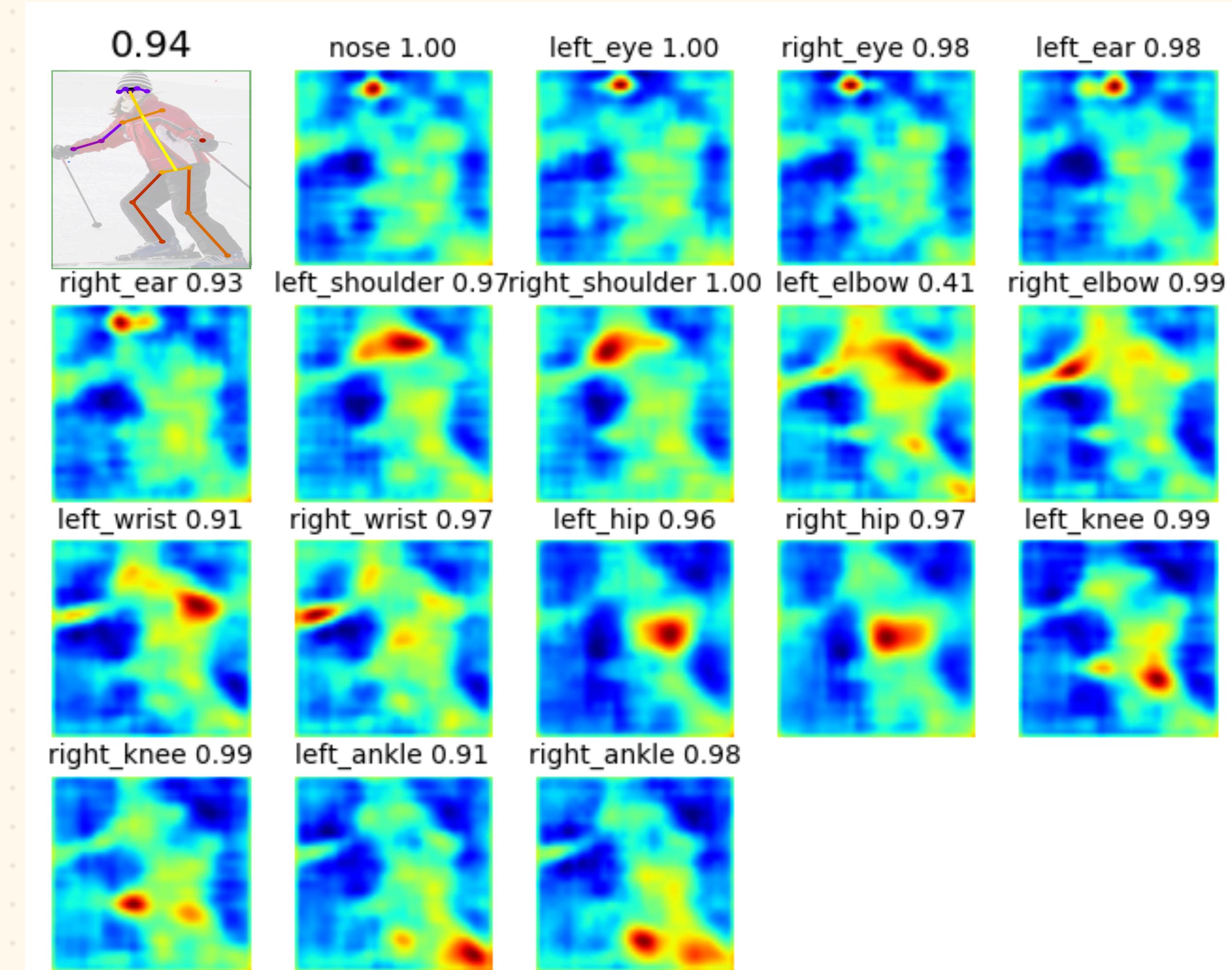
Quantitative Results

		backbone	mask AP
2015 COCO winner <small>[seconds per image]</small>	MNC	ResNet-101-C4	24.6
	FCIS w/ OHEM	ResNet-101-C5-dilated	29.2
2016 COCO winner	FCIS+++ w/ OHEM	ResNet-101-C5-dilated	33.6
	Mask R-CNN	ResNet-101-C4	33.1
Our 200ms version	Mask R-CNN	ResNet-101-FPN	35.7
	Mask R-CNN	ResNeXt-101-FPN	37.1

Mask R-CNN for Human Pose Estimation

OVERVIEW

- Keypoint = 1-hot mask
- Human pose = 17 keypoints
- Represent pose as 17 masks





Mask R-CNN results on COCO



Mask R-CNN results on COCO



Mask R-CNN results on COCO

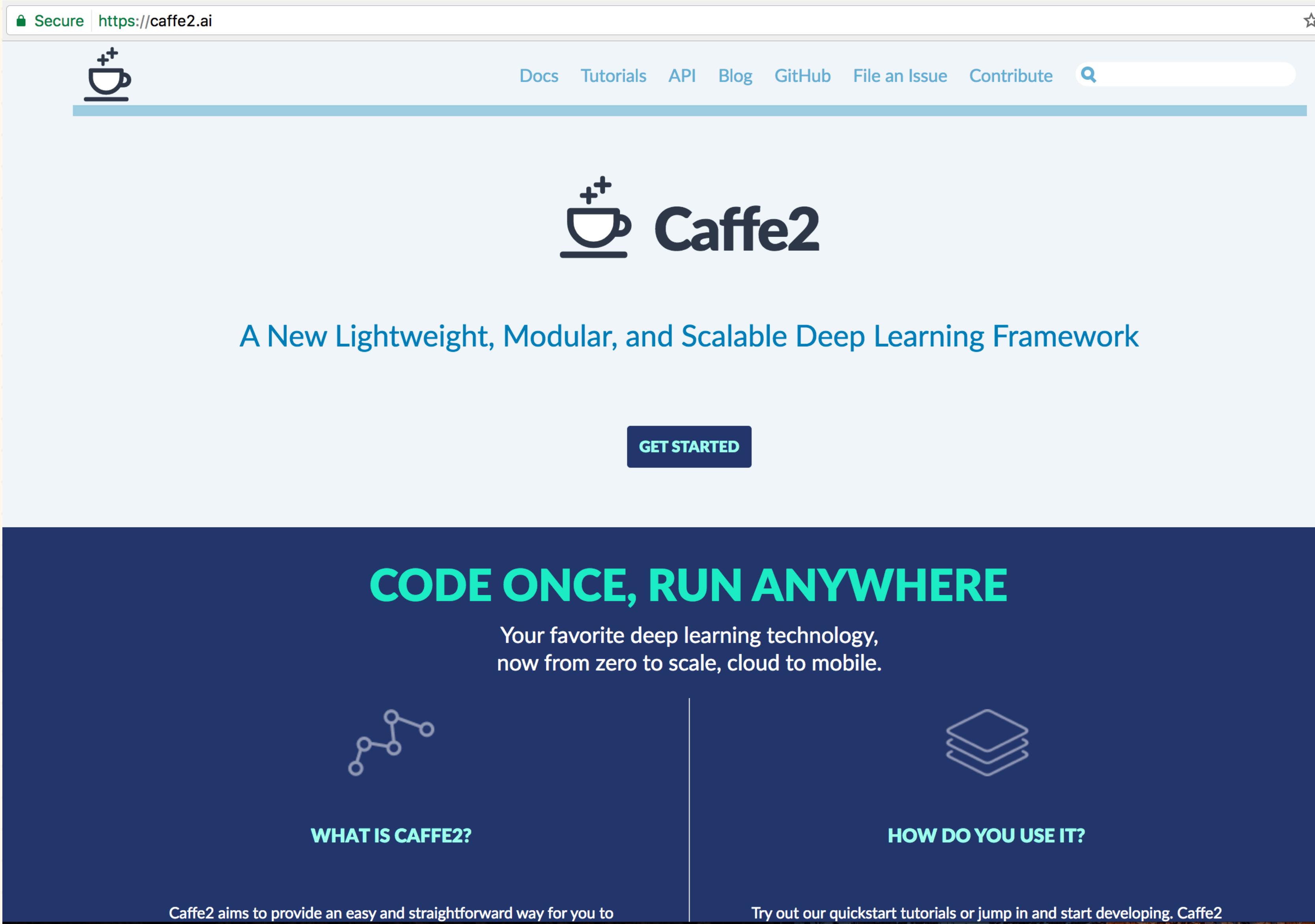


Mask R-CNN results on COCO

Quantitative Results

		keypoint AP
2016 COCO winner [seconds per image]	CMU-Pose+++	61.8
	G-RMI [w/ extra data]	62.4
	Mask R-CNN [keypoint-only]	62.7
Our 200ms version	Mask R-CNN [keypoint & mask]	63.1

Caffe2 Accelerated Research



The screenshot shows the official website for Caffe2. At the top, there's a header bar with a lock icon indicating it's secure (<https://caffe2.ai>), a star icon for bookmarking, and a navigation menu with links to Docs, Tutorials, API, Blog, GitHub, File an Issue, Contribute, and a search icon.

The main content area features the Caffe2 logo, which consists of a stylized coffee cup icon with two plus signs above it, followed by the text "Caffe2". Below the logo, a blue banner with white text reads "A New Lightweight, Modular, and Scalable Deep Learning Framework". A prominent "GET STARTED" button is located in the center of this banner.

Below the banner, there's a dark blue section with white text and icons. The text "CODE ONCE, RUN ANYWHERE" is displayed in large, bold, cyan letters. Below it, a smaller text block says "Your favorite deep learning technology, now from zero to scale, cloud to mobile." To the left of this text is a small network graph icon, and to the right is a stack of three rectangular blocks icon.

The dark blue section is divided into two columns. The left column contains the text "WHAT IS CAFFE2?" below a small network graph icon. The right column contains the text "HOW DO YOU USE IT?" below a stack of three rectangular blocks icon.

At the very bottom of the page, a thin white footer bar spans across the width of the screen, containing the text "Caffe2 aims to provide an easy and straightforward way for you to" on the left and "Try out our quickstart tutorials or jump in and start developing. Caffe2" on the right.

Caffe2 Object Detection Platform

RAPID IDEA ITERATION IS A KEY ENABLING FACTOR IN RESEARCH

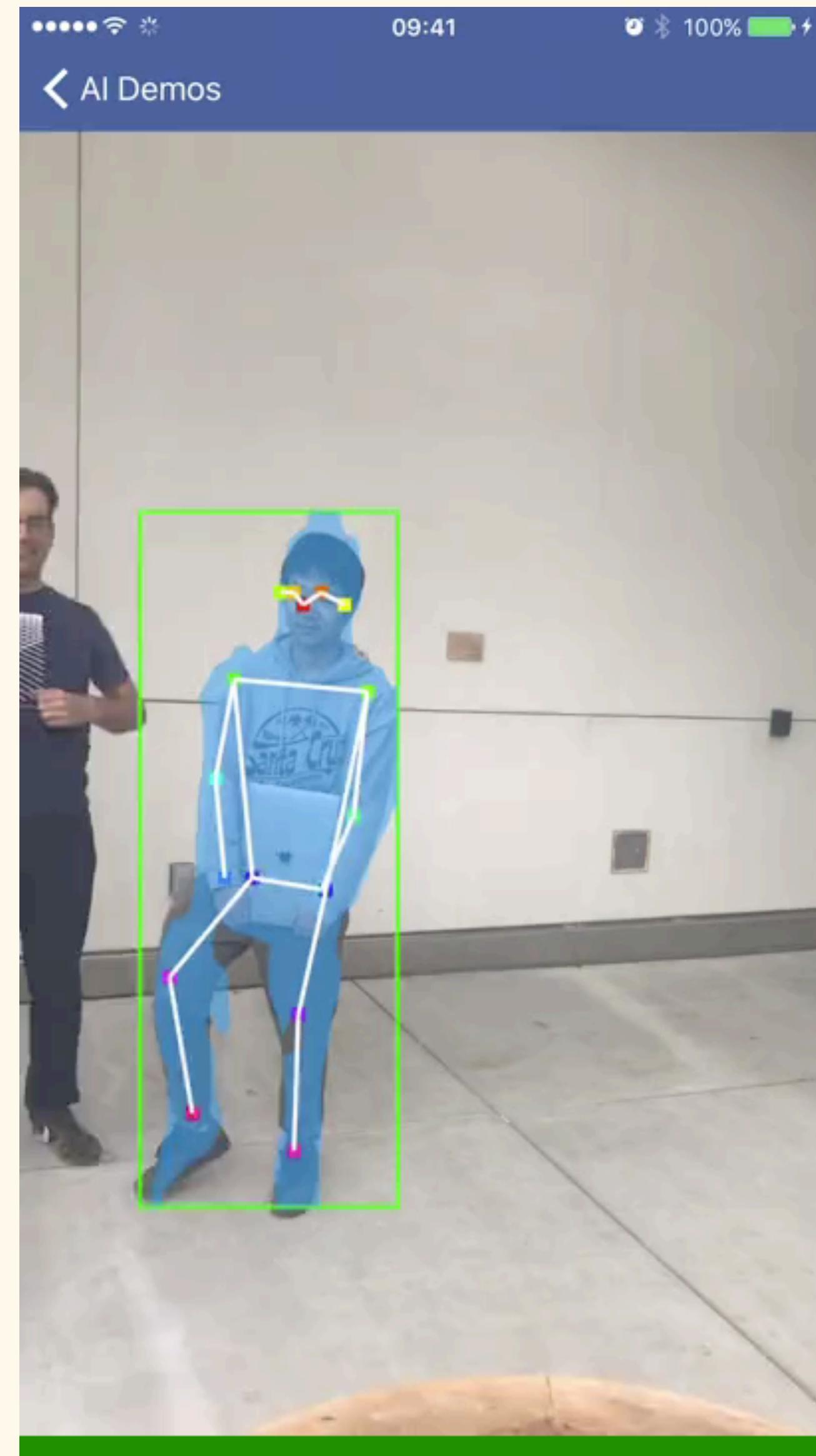
- Early alpha users starting in May 2016
- Ported py-faster-rcnn from Caffe to Caffe2
- Key design choices
 - Flexible framework for implementing object detection models
 - Parallelize data loading with forward/backward computation

Caffe2 Object Detection Platform

RAPID IDEA ITERATION IS A KEY ENABLING FACTOR IN RESEARCH

- Sync SGD with 8 GPUs [Tesla M40] in a BigSur server
- Rapid prototyping of Mask R-CNN models in 8-12 hours
- SOTA Mask R-CNN models train in 44 hours
- Previous systems: ~ 4 days training time [experience from MSRA]

From Research to Mobile with Caffe2



Conclusions

- Simple and effective
- Fast inference
- Box, mask, and pose all-in-one network *and* method
- Caffe2 enables extremely fast prototyping, critical to our success

20

F8

17