Sharing Session

Mask R-CNN using Detectron2

We will learn together about some of computer vision problems, and will try Mask R-CNN architecture to solve the instance segmentation problems

Benedict Aryo

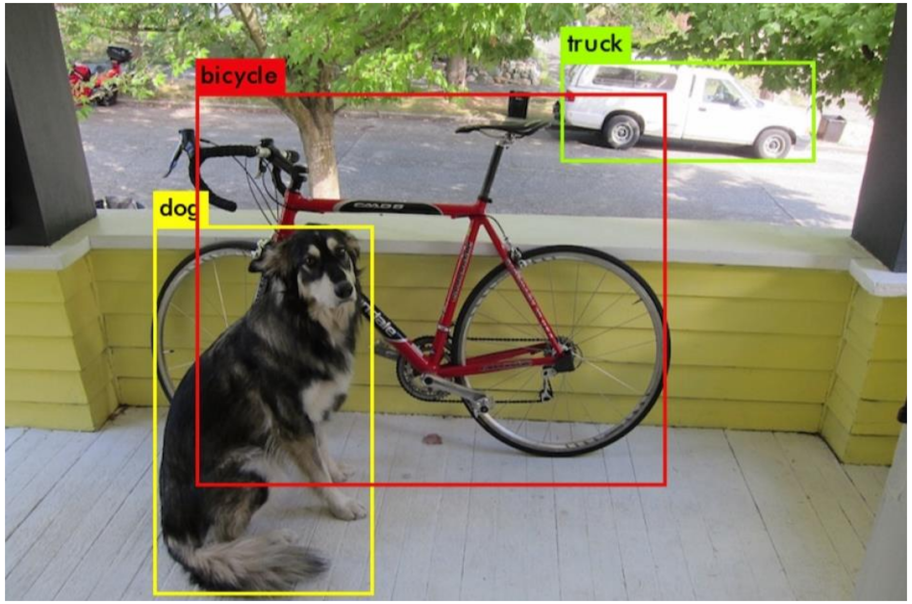
Mask R-CNN using Detectron2

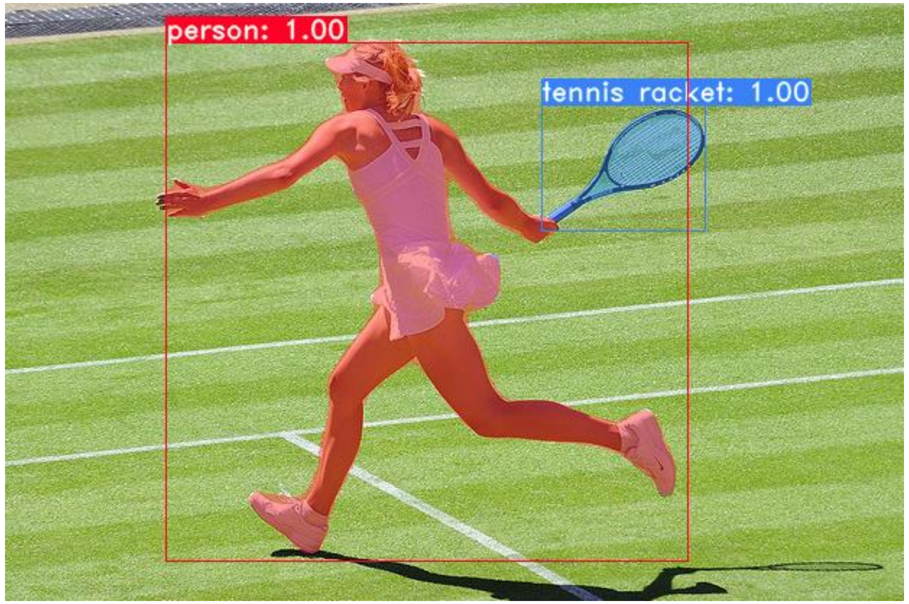
Agenda:

Computer Vision Problems
RCNN Family
Mask RCNN
Why Detectron2
Practice using Detectron2
Training using Detectron2 (optional)

What I was thinking about...

Faster R-CNN VS Mask R-CNN





I though that...

Mask R-CNN = Faster R-CNN + Steroids

(or some magic)

But, is it true ?

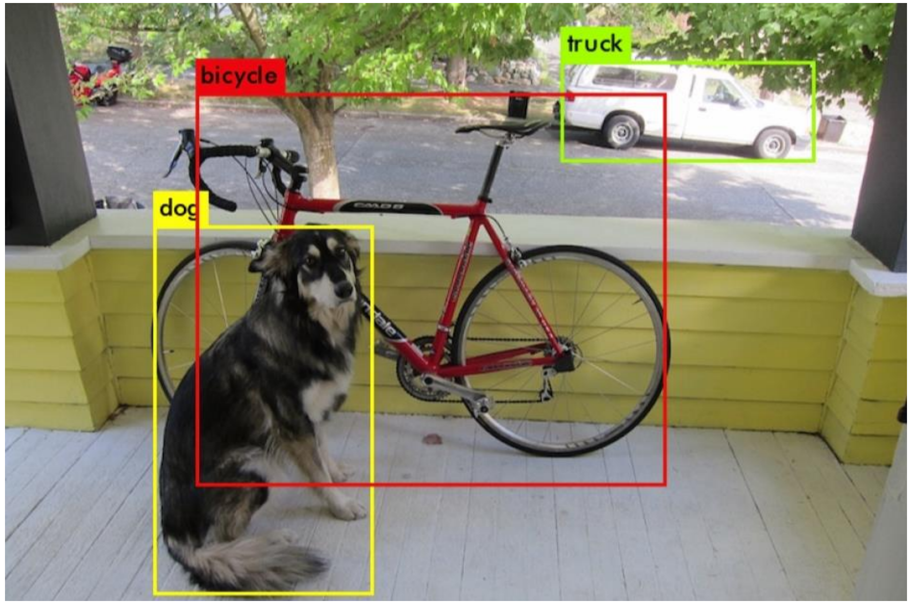
Now I think that...

Mask R-CNN & Faster R-CNN
serve **different purposes**

Wait... what?

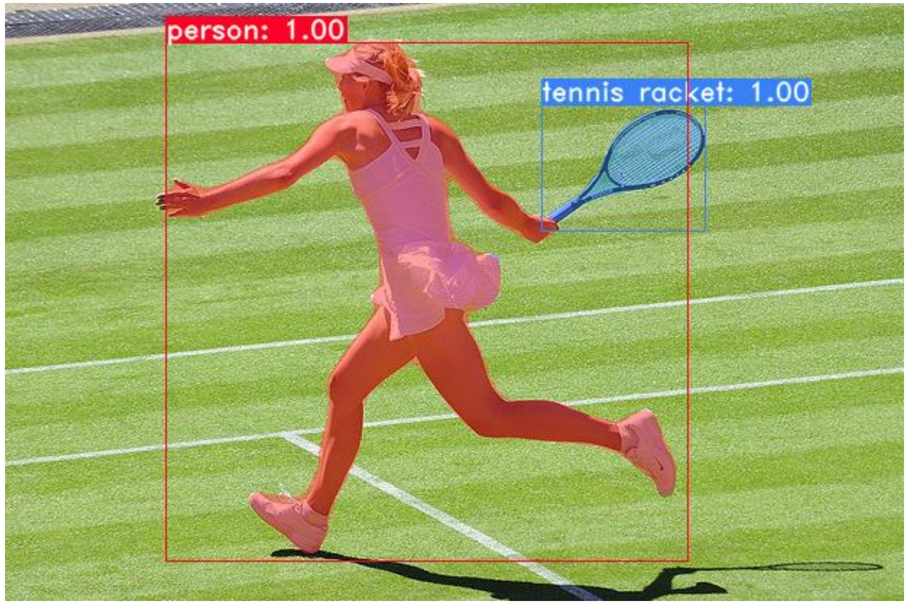
In fact...

This is
Not a Faster R-CNN



YOLOv3:An Incremental Improvement. Joseph Redmond, Ali Farhadi. 2018

This is
Not a Mask R-CNN

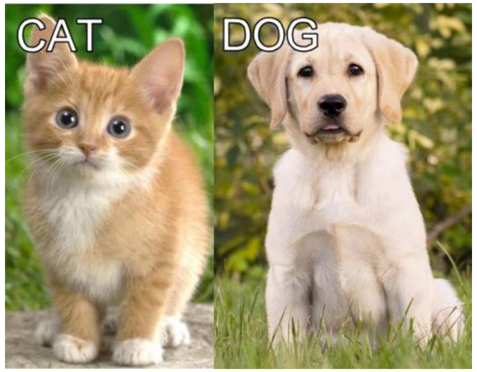


YOLACT: Real-time Instance Segmentation. Daniel Bolya, et. al. 2019

Understanding the **objectives** is essential to solving the problems

Problems in Computer Vision **Image Recognition**

Image Classification



Classification + Localization

Object Detection

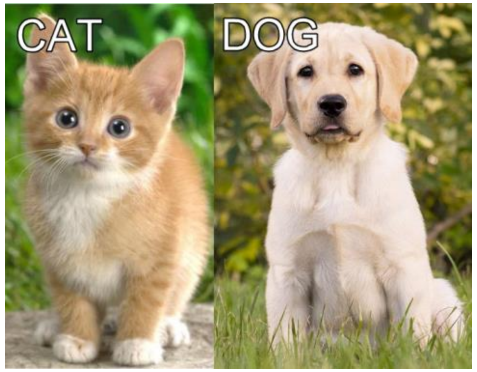
Instance Segmentation

Semantic segmentation

Panoptic segmentation

Problems in Computer Vision **Image Recognition**

Image Classification

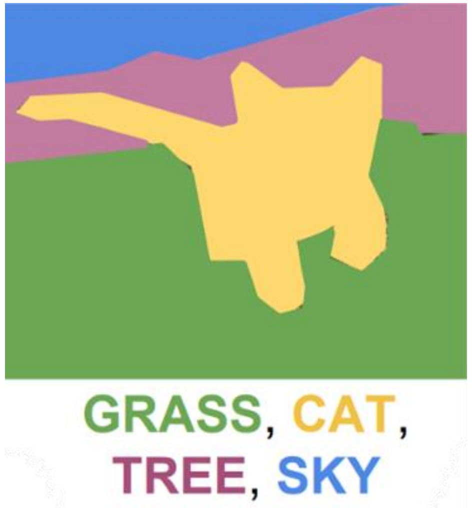


Classification + Localization

Object Detection

Instance Segmentation

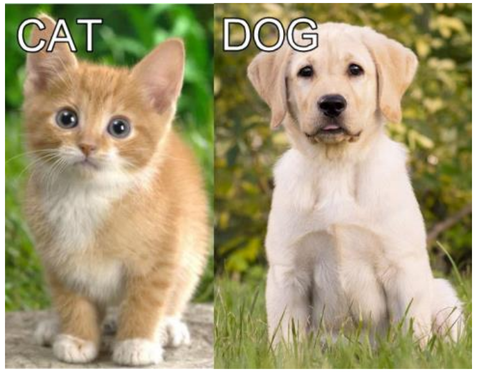
Semantic segmentation



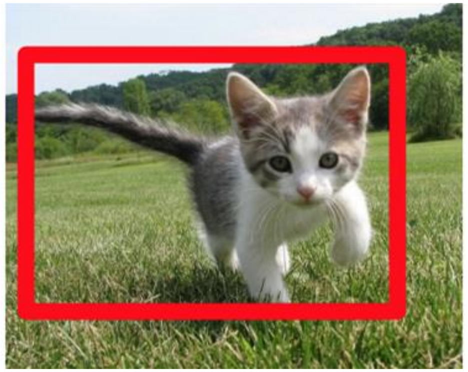
Panoptic segmentation

Problems in Computer Vision **Image Recognition**

Image Classification



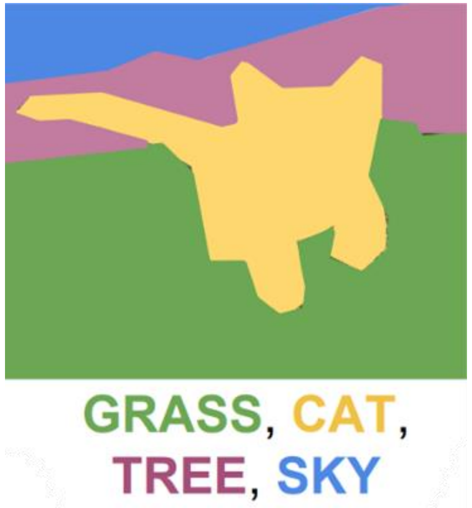
Classification + Localization



Object Detection

Instance Segmentation

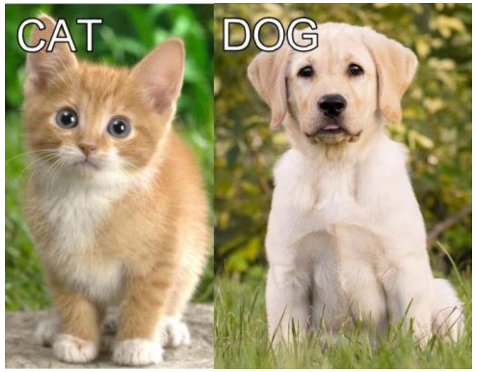
Semantic segmentation



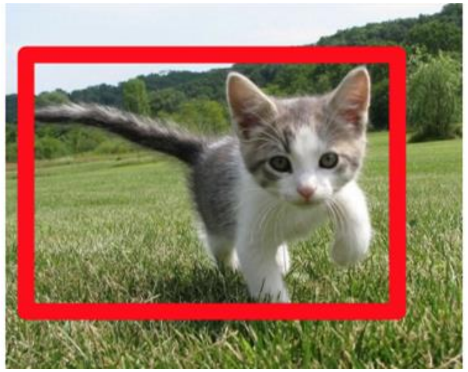
Panoptic segmentation

Problems in Computer Vision **Image Recognition**

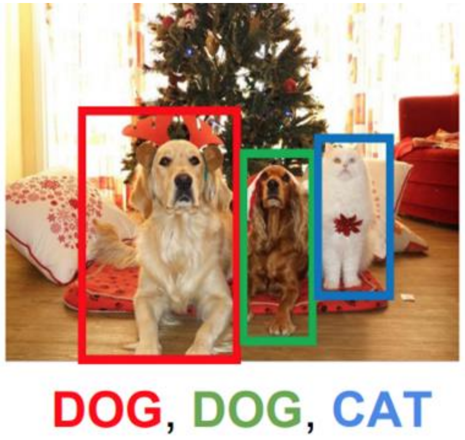
Image Classification



Classification + Localization

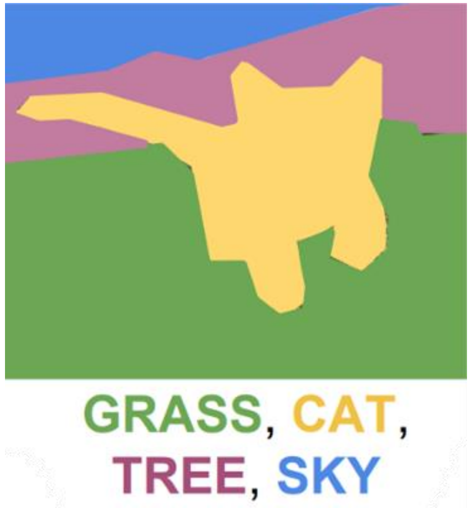


Object Detection



Instance Segmentation

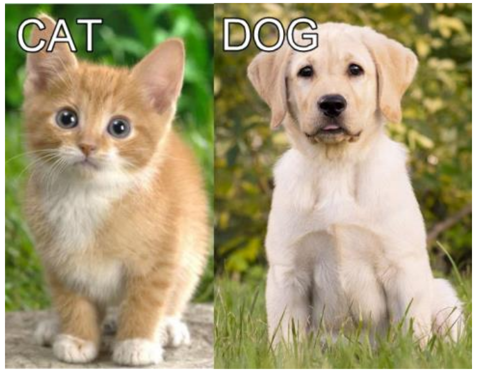
Semantic segmentation



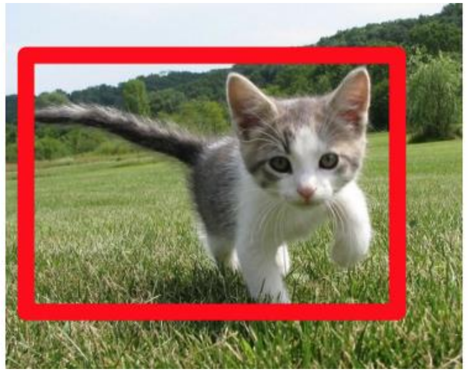
Panoptic segmentation

Problems in Computer Vision **Image Recognition**

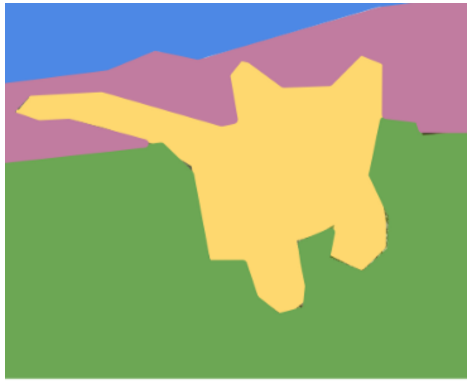
Image Classification



Classification + Localization

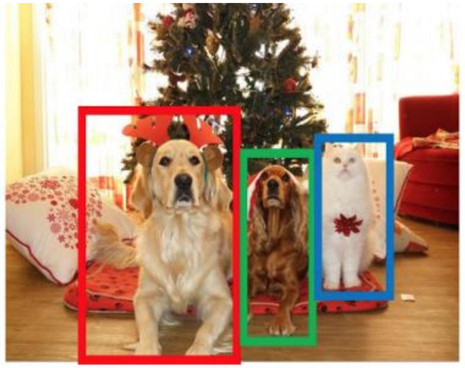


Semantic segmentation



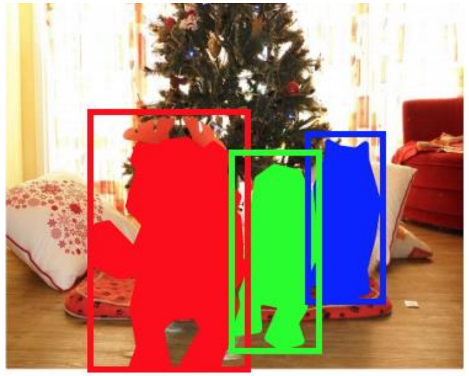
GRASS, CAT,
TREE, SKY

Object Detection



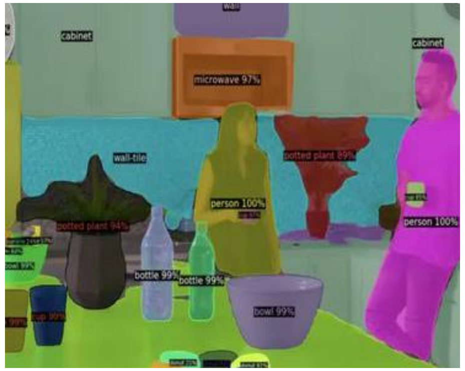
DOG, DOG, CAT

Instance Segmentation

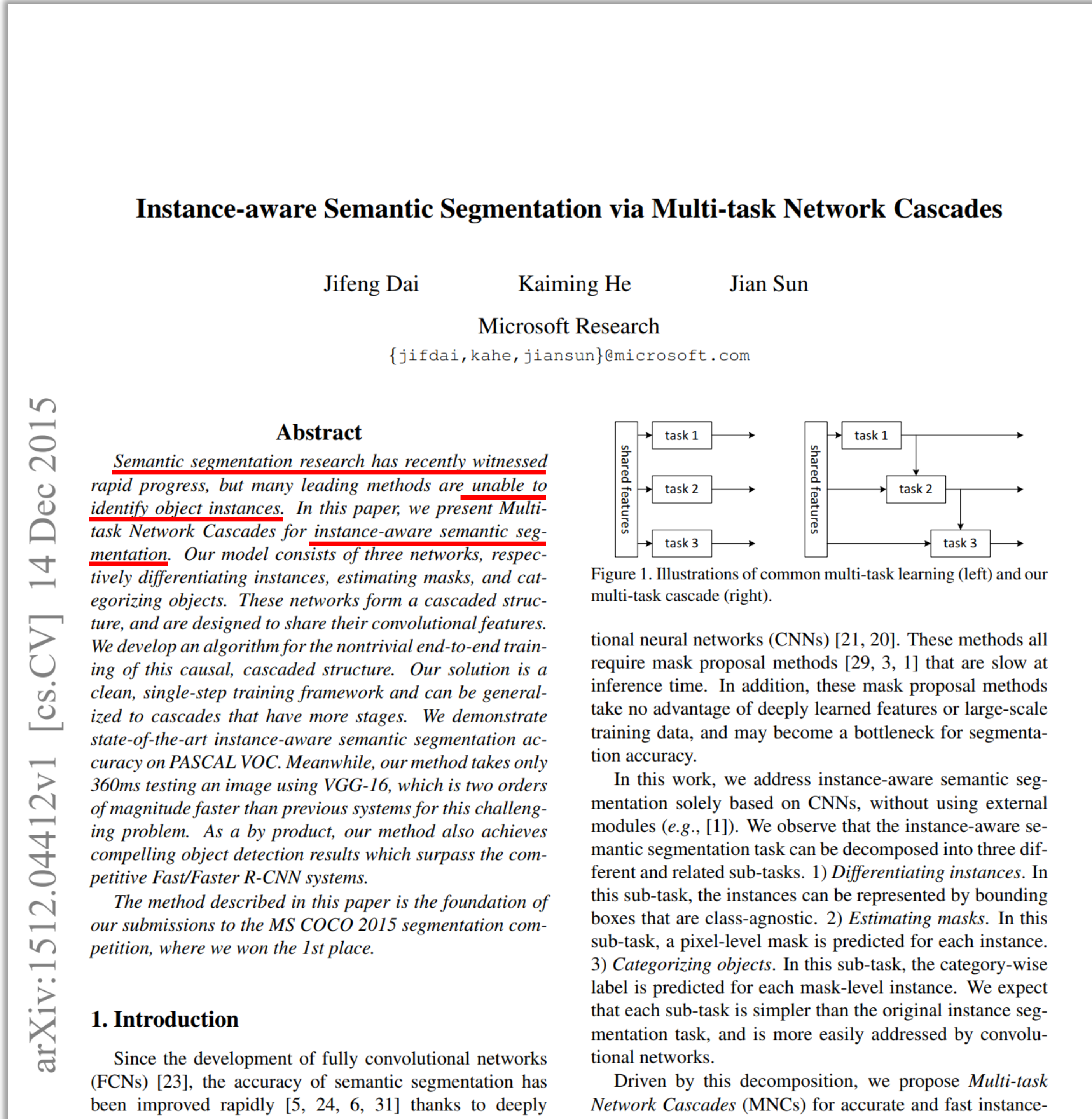


DOG, DOG, CAT

Panoptic segmentation



Instance-aware Semantic Segmentation



Instance-aware Semantic Segmentation via Multi-task Network Cascades

Jifeng Dai Kaiming He Jian Sun

Microsoft Research

{jifdai,kahe,jiansun}@microsoft.com

Abstract

Semantic segmentation research has recently witnessed rapid progress, but many leading methods are unable to identify object instances. In this paper, we present Multi-task Network Cascades for instance-aware semantic segmentation. Our model consists of three networks, respectively differentiating instances, estimating masks, and categorizing objects. These networks form a cascaded structure, and are designed to share their convolutional features. We develop an algorithm for the nontrivial end-to-end training of this causal, cascaded structure. Our solution is a clean, single-step training framework and can be generalized to cascades that have more stages. We demonstrate state-of-the-art instance-aware semantic segmentation accuracy on PASCAL VOC. Meanwhile, our method takes only 360ms testing an image using VGG-16, which is two orders of magnitude faster than previous systems for this challenging problem. As a by product, our method also achieves compelling object detection results which surpass the competitive Fast/Faster R-CNN systems.

The method described in this paper is the foundation of our submissions to the MS COCO 2015 segmentation competition, where we won the 1st place.

1. Introduction

Since the development of fully convolutional networks (FCNs) [23], the accuracy of semantic segmentation has been improved rapidly [5, 24, 6, 31] thanks to deeply

Figure 1. Illustrations of common multi-task learning (left) and our multi-task cascade (right).

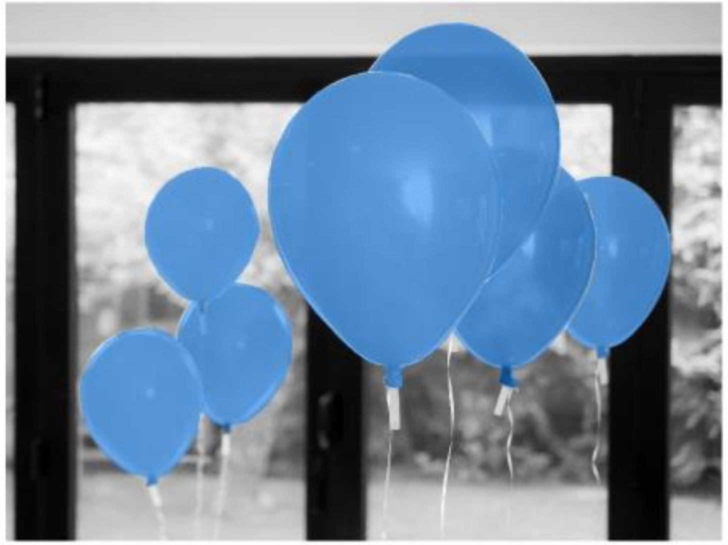
tional neural networks (CNNs) [21, 20]. These methods all require mask proposal methods [29, 3, 1] that are slow at inference time. In addition, these mask proposal methods take no advantage of deeply learned features or large-scale training data, and may become a bottleneck for segmentation accuracy.

In this work, we address instance-aware semantic segmentation solely based on CNNs, without using external modules (*e.g.*, [1]). We observe that the instance-aware semantic segmentation task can be decomposed into three different and related sub-tasks. 1) *Differentiating instances*. In this sub-task, the instances can be represented by bounding boxes that are class-agnostic. 2) *Estimating masks*. In this sub-task, a pixel-level mask is predicted for each instance. 3) *Categorizing objects*. In this sub-task, the category-wise label is predicted for each mask-level instance. We expect that each sub-task is simpler than the original instance segmentation task, and is more easily addressed by convolutional networks.

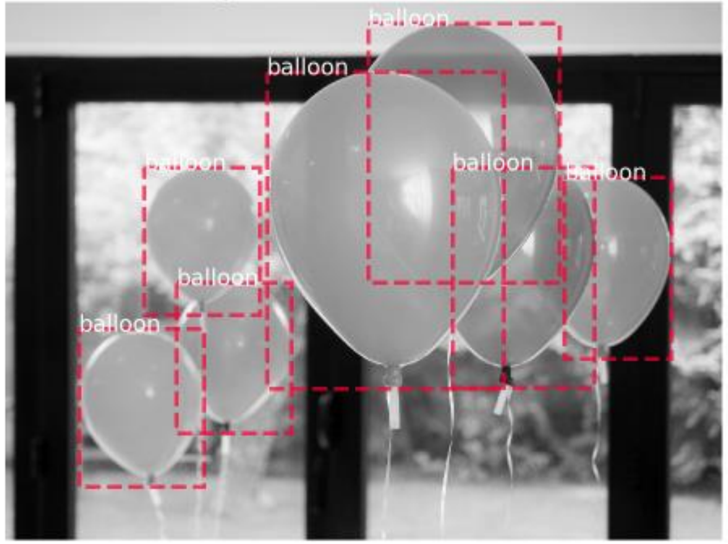
Driven by this decomposition, we propose *Multi-task Network Cascades* (MNCs) for accurate and fast instance-

The differences

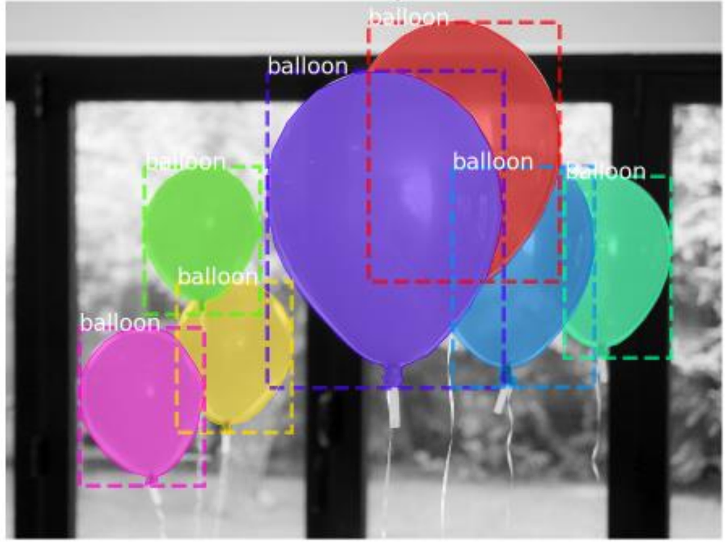
Semantic Segmentation



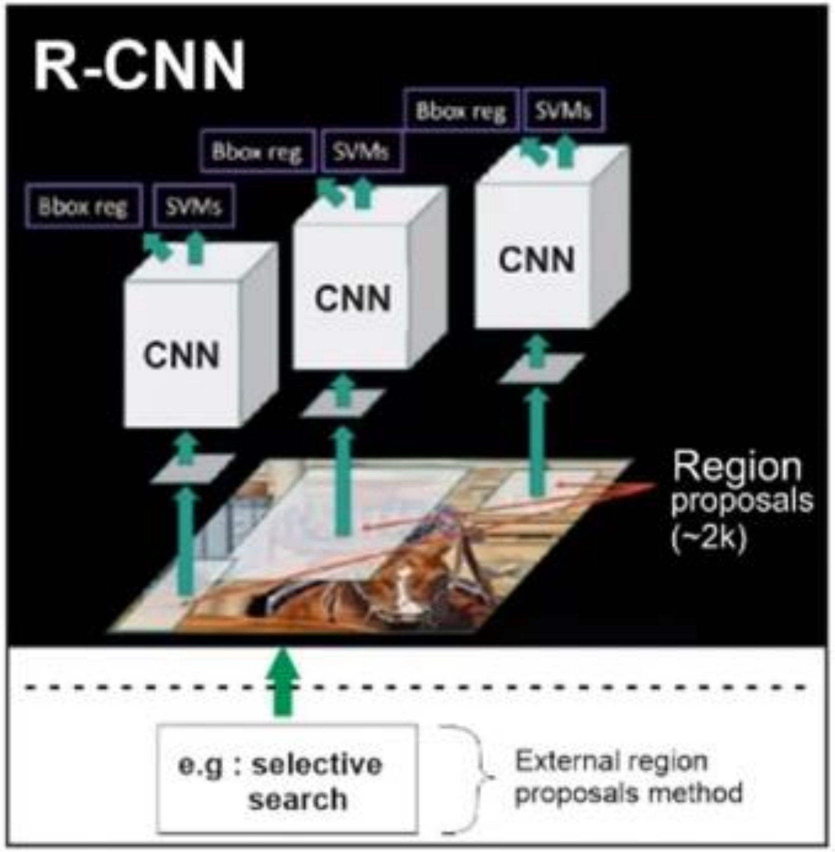
Object Detection

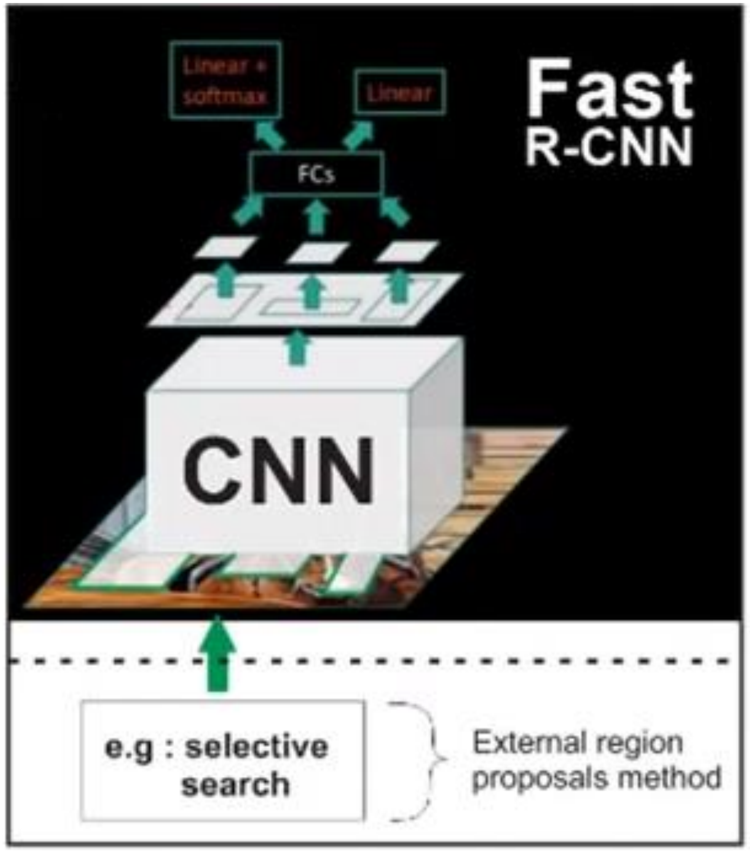


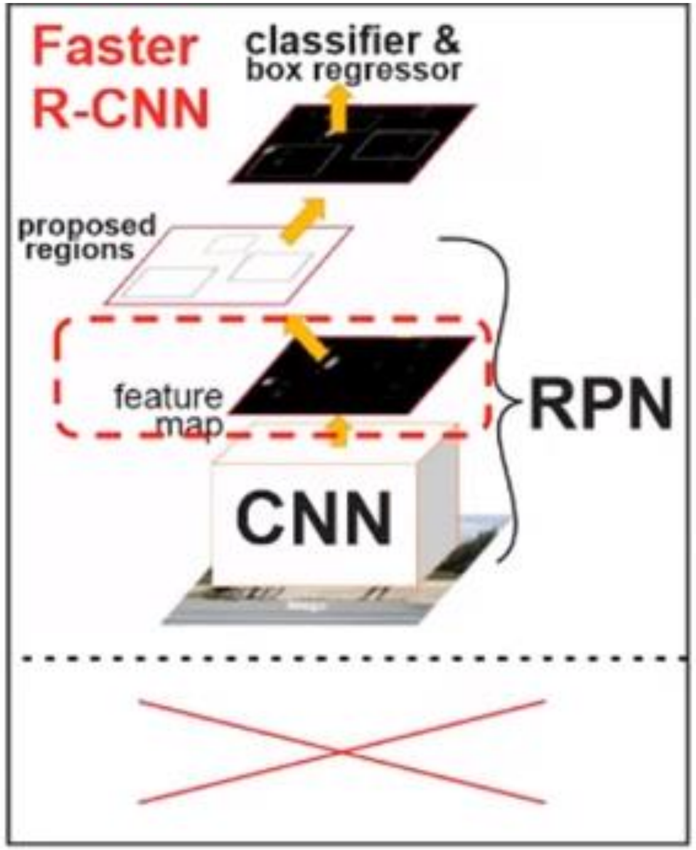
Instance Segmentation



RCNN Family for Object Detection

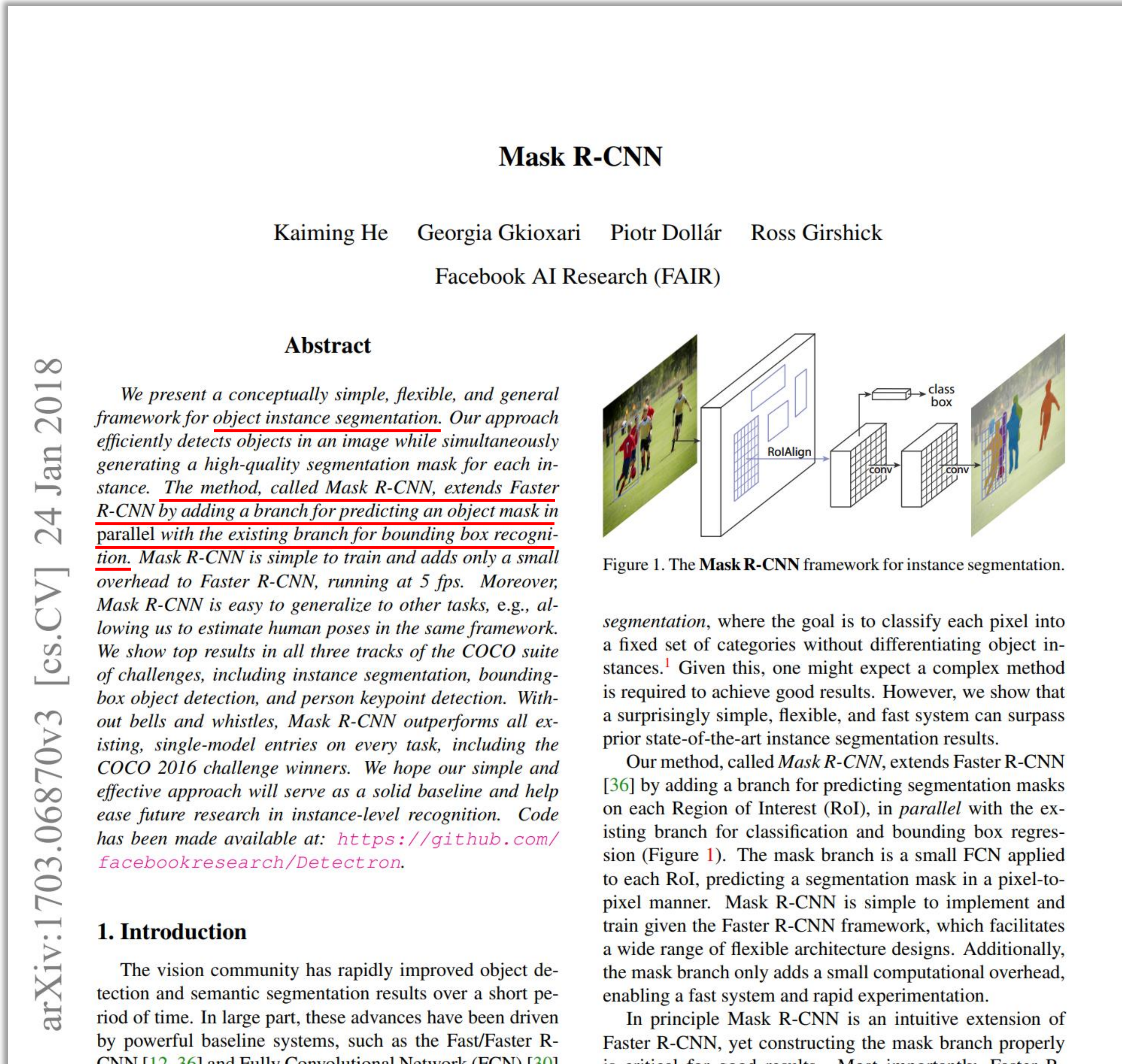






Mask R-CNN

Best Paper Award (**Marr Prize**) at the 16th International Conference on Computer vision (**ICCV**) *2017*



Mask R-CNN

Kaiming He Georgia Gkioxari Piotr Dollár Ross Girshick
Facebook AI Research (FAIR)

arXiv:1703.06870v3 [cs.CV] 24 Jan 2018

Abstract

We present a conceptually simple, flexible, and general framework for object instance segmentation. Our approach efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance. The method, called Mask R-CNN, extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. Mask R-CNN is simple to train and adds only a small overhead to Faster R-CNN, running at 5 fps. Moreover, Mask R-CNN is easy to generalize to other tasks, e.g., allowing us to estimate human poses in the same framework. We show top results in all three tracks of the COCO suite of challenges, including instance segmentation, bounding-box object detection, and person keypoint detection. Without bells and whistles, Mask R-CNN outperforms all existing, single-model entries on every task, including the COCO 2016 challenge winners. We hope our simple and effective approach will serve as a solid baseline and help ease future research in instance-level recognition. Code has been made available at: https://github.com/facebookresearch/Detectron.

1. Introduction

The vision community has rapidly improved object detection and semantic segmentation results over a short period of time. In large part, these advances have been driven by powerful baseline systems, such as the Fast/Faster R-CNN [12, 36] and Fully Convolutional Network (FCN) [30]

Figure 1. The **Mask R-CNN** framework for instance segmentation.

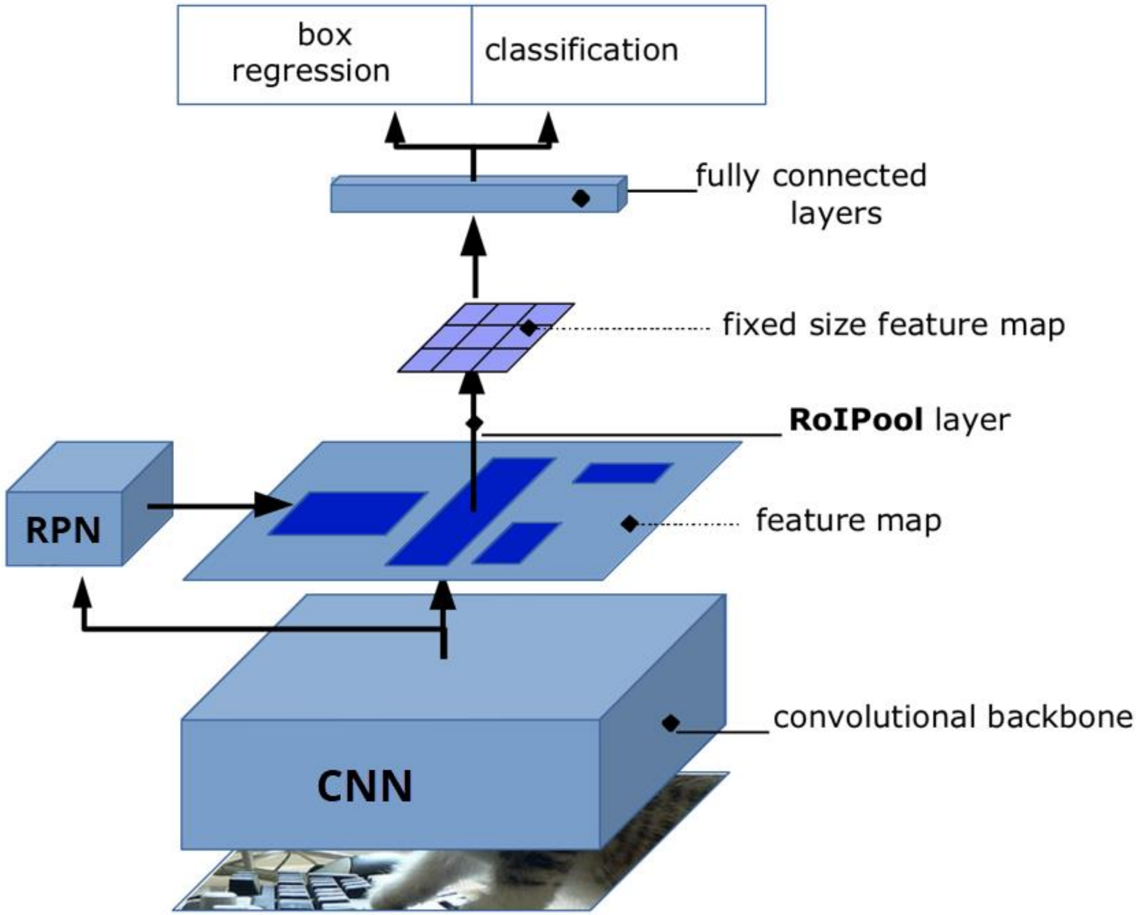
segmentation, where the goal is to classify each pixel into a fixed set of categories without differentiating object instances.[1] Given this, one might expect a complex method is required to achieve good results. However, we show that a surprisingly simple, flexible, and fast system can surpass prior state-of-the-art instance segmentation results.

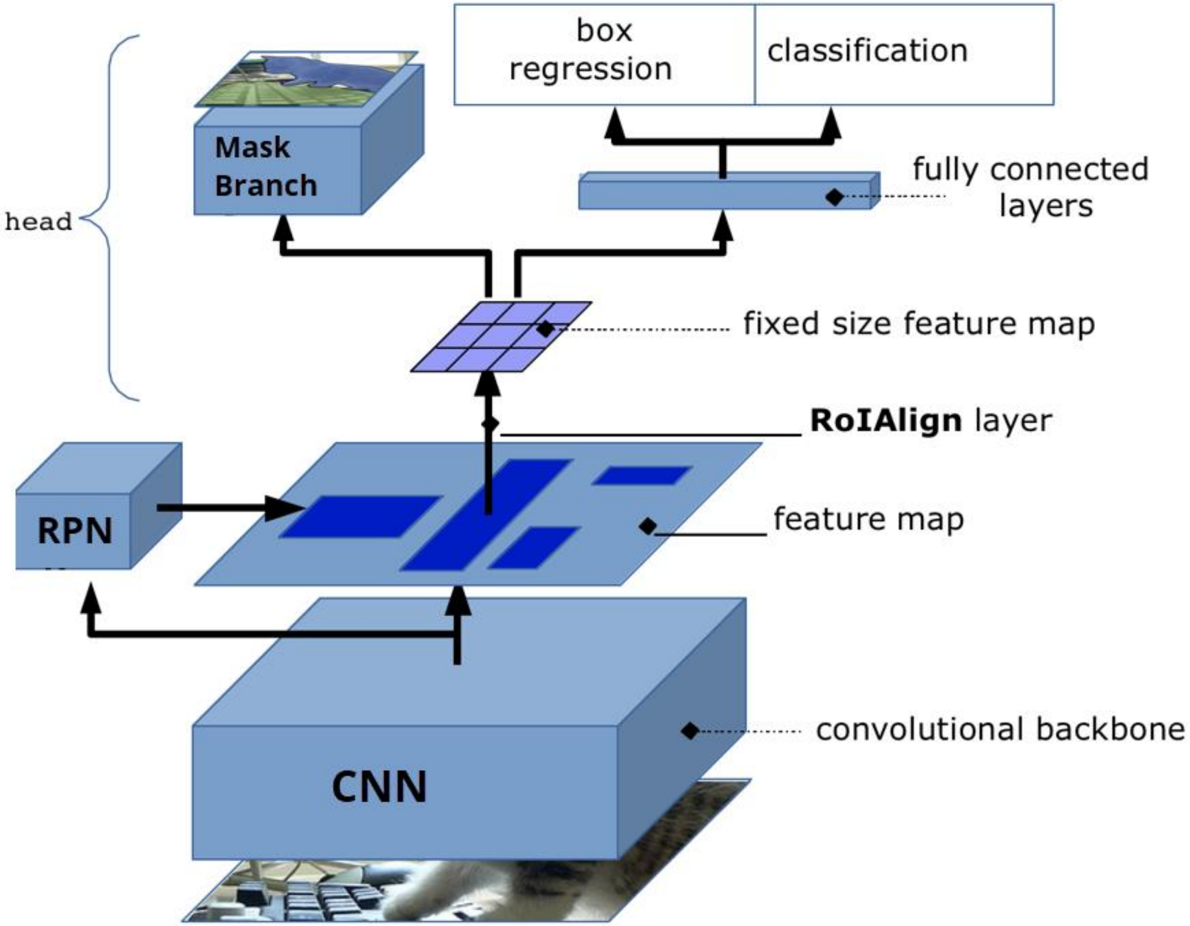
Our method, called *Mask R-CNN*, extends Faster R-CNN [36] by adding a branch for predicting segmentation masks on each Region of Interest (RoI), in *parallel* with the existing branch for classification and bounding box regression (Figure 1). The mask branch is a small FCN applied to each RoI, predicting a segmentation mask in a pixel-to-pixel manner. Mask R-CNN is simple to implement and train given the Faster R-CNN framework, which facilitates a wide range of flexible architecture designs. Additionally, the mask branch only adds a small computational overhead, enabling a fast system and rapid experimentation.

In principle Mask R-CNN is an intuitive extension of Faster R-CNN, yet constructing the mask branch properly is critical for good results. Most importantly, Faster R-

Model Architecture

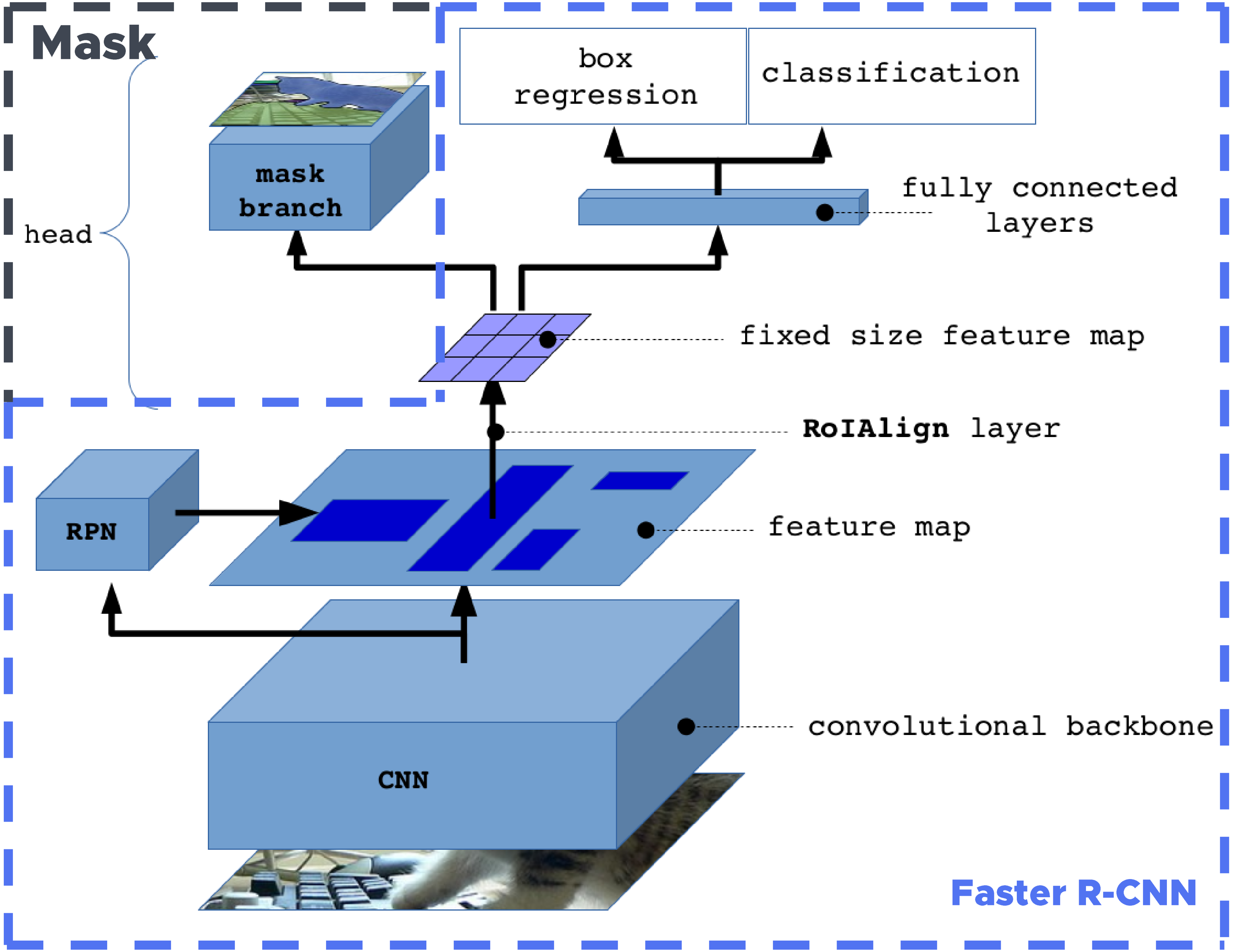
Faster R-CNN vs Mask R-CNN





Mask R-CNN

Architecture



Mask R-CNN Implementation

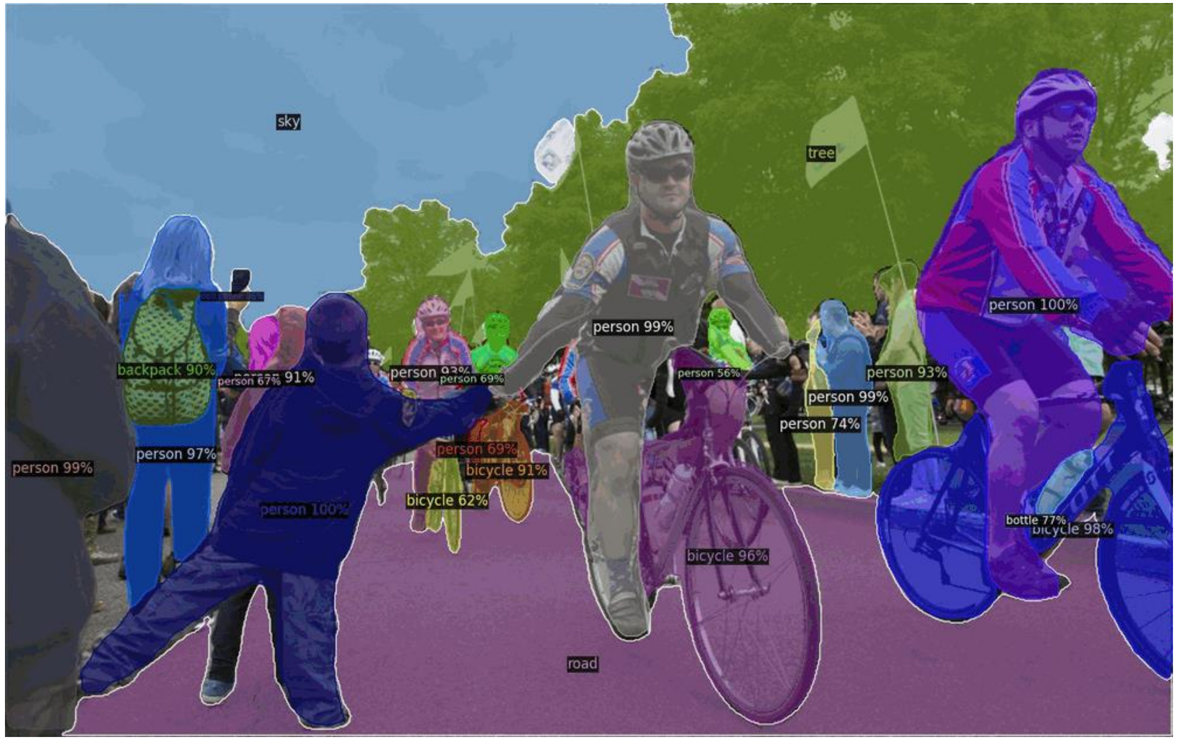
Facebookresearch/detectron

- Original Implementation from paper
- Using Caffe2 Framework
- Deprecated, move to detectron2 which is a ground-up rewrite of Detectron in PyTorch

Matterport/Mask_RCNN

- Most Popular implementation of Mask R-CNN
- Using Tensorflow v1 Framework
- No Longer maintained, many compatibility issue to new tensorflow version

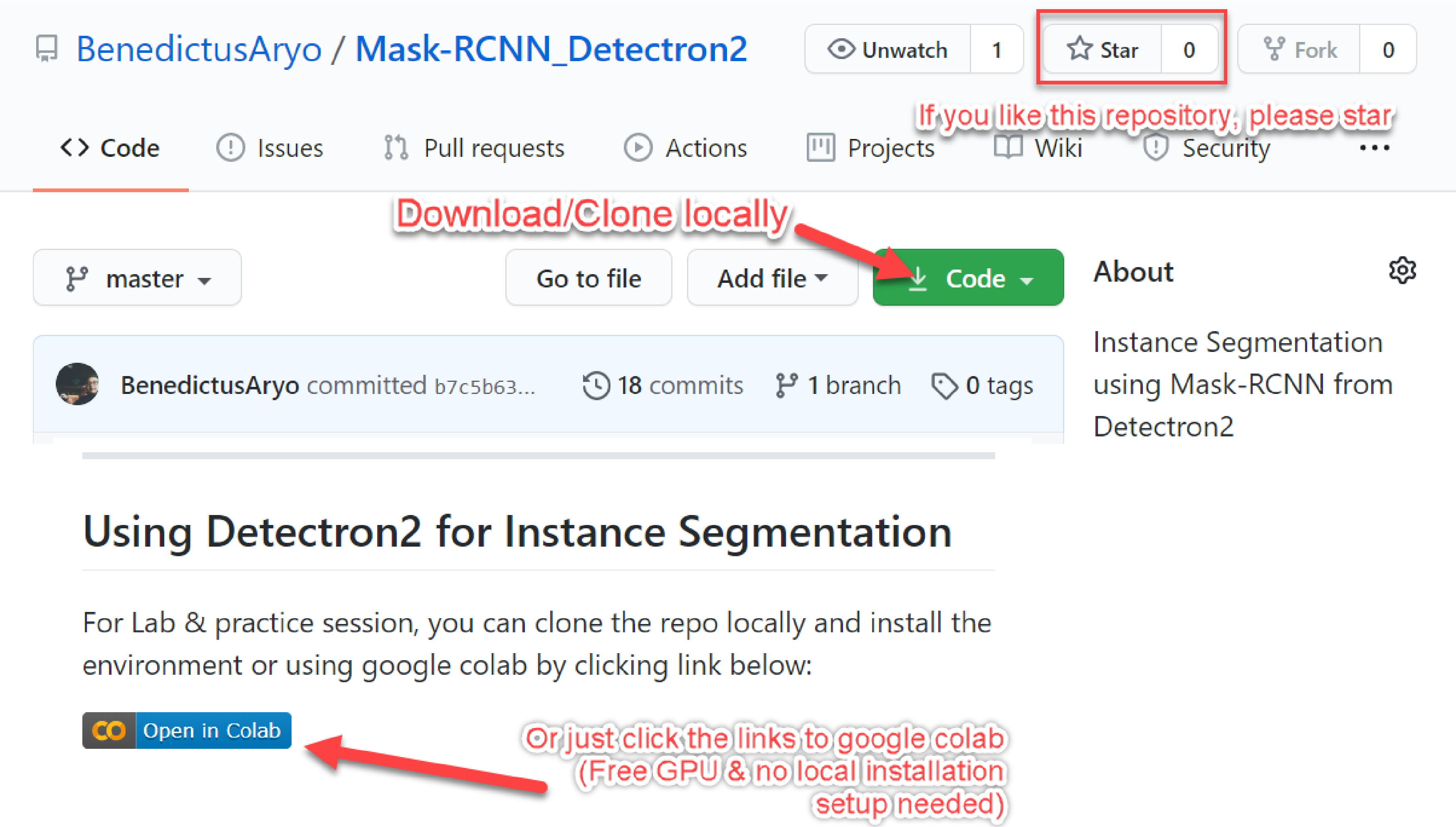
Mask R-CNN Implementation using Detectron2



Detectron2 is Facebook AI Research's next generation software system that implements state-of-the-art object detection algorithms. It is a ground-up rewrite of the previous version, Detectron, and it originates from maskrcnn-benchmark powered by the PyTorch deep learning framework.

Detectron2 Lab:

bit.ly/detectron2_app



About

Instance Segmentation using Mask-RCNN from Detectron2

Using Detectron2 for Instance Segmentation

For Lab & practice session, you can clone the repo locally and install the environment or using google colab by clicking link below:

If you like this repository, please star

Download/Clone locally

Or just click the links to google colab
(Free GPU & no local installation
setup needed)