Mask R-CNN

https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf

Team members:

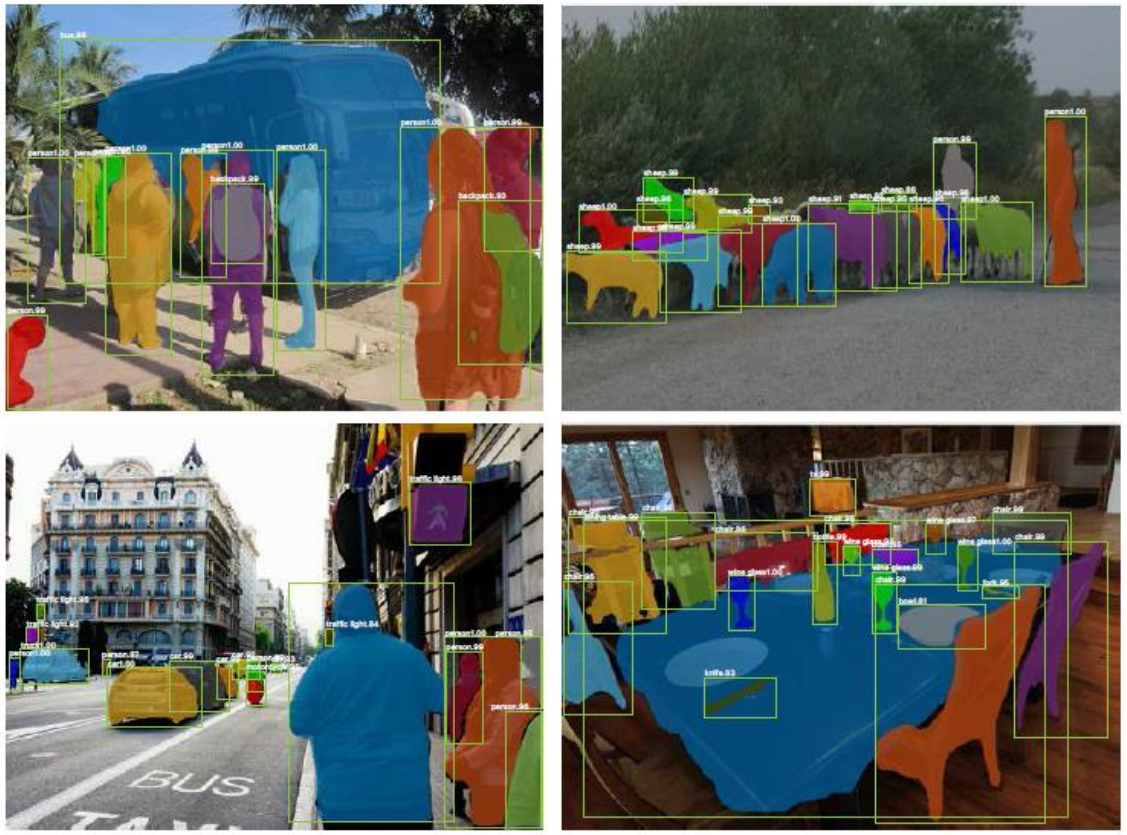
Hoàng Văn Quyết

Đỗ Thị Thùy Linh

Dương Quốc Khang

What is motivation of this work?

- **Topic:** *"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."*
- **Problem:** *"****Instance segmentation is challenging because it requires the correct detection of all objects in an image while also precisely segmenting each instance.*** *It therefore combines elements from the classical computer vision tasks of object detection, where the goal is to classify individual objects and localize each using a bounding box, and semantic segmentation, where the goal is to classify each pixel into a fixed set of categories without differentiating object instances.* ***Given this, one might expect a complex method*** **is required to achieve good results."**



What is the proposed solution?

- “Our method, called Mask R-CNN, **extends Faster R-CNN 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. **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**.”

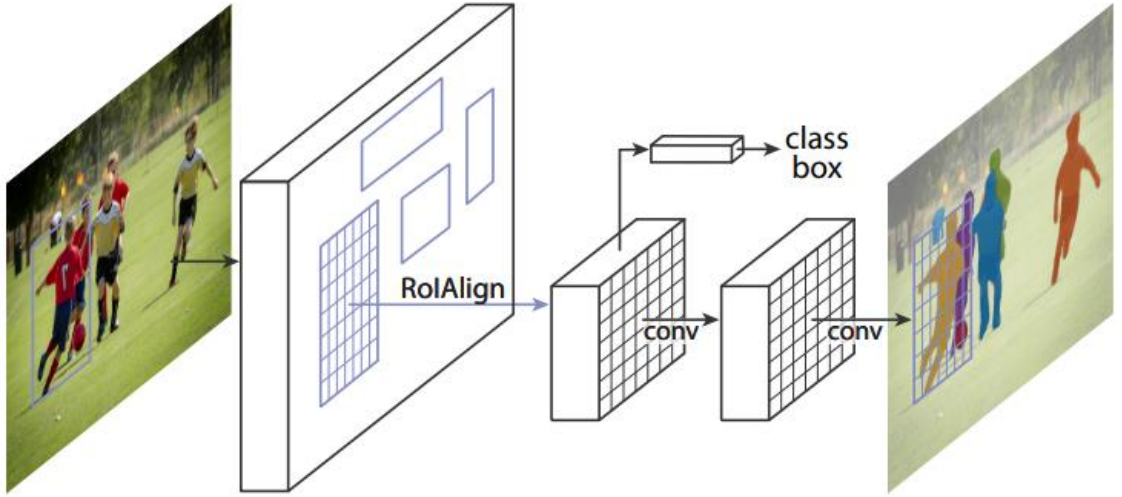


Figure 1. The **Mask R-CNN** framework for instance segmentation.

What is the work's evaluation of the proposed solution?

- *"Without bells and whistles,* ***Mask R-CNN surpasses all previous state-of-the-art single-model results on the COCO instance segmentation task, including the heavily engineered entries from the 2016 competition winner.*** *As a by-product, our method also excels on the COCO object detection task. In ablation experiments, we evaluate multiple basic instantiations, which allows us to demonstrate its robustness and analyze the effects of core factors."*

	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
MNC [7]	ResNet-101-C4	24.6	44.3	24.8	4.7	25.9	43.6
FCIS [21] +OHEM	ResNet-101-C5-dilated	29.2	49.5	-	7.1	31.3	50.0
FCIS+++ [21] +OHEM	ResNet-101-C5-dilated	33.6	54.5	-	-	-	-
Mask R-CNN	ResNet-101-C4	33.1	54.9	34.8	12.1	35.6	51.1
Mask R-CNN	ResNet-101-FPN	35.7	58.0	37.8	15.5	38.1	52.4
Mask R-CNN	ResNeXt-101-FPN	**37.1**	**60.0**	**39.4**	**16.9**	**39.9**	**53.5**

Table 1. **Instance segmentation** *mask* AP on COCO `test-dev`. MNC [7] and FCIS [21] are the winners of the COCO 2015 and 2016 segmentation challenges, respectively. Without bells and whistles, Mask R-CNN outperforms the more complex FCIS+++, which includes multi-scale train/test, horizontal flip test, and OHEM [30]. All entries are *single-model* results.