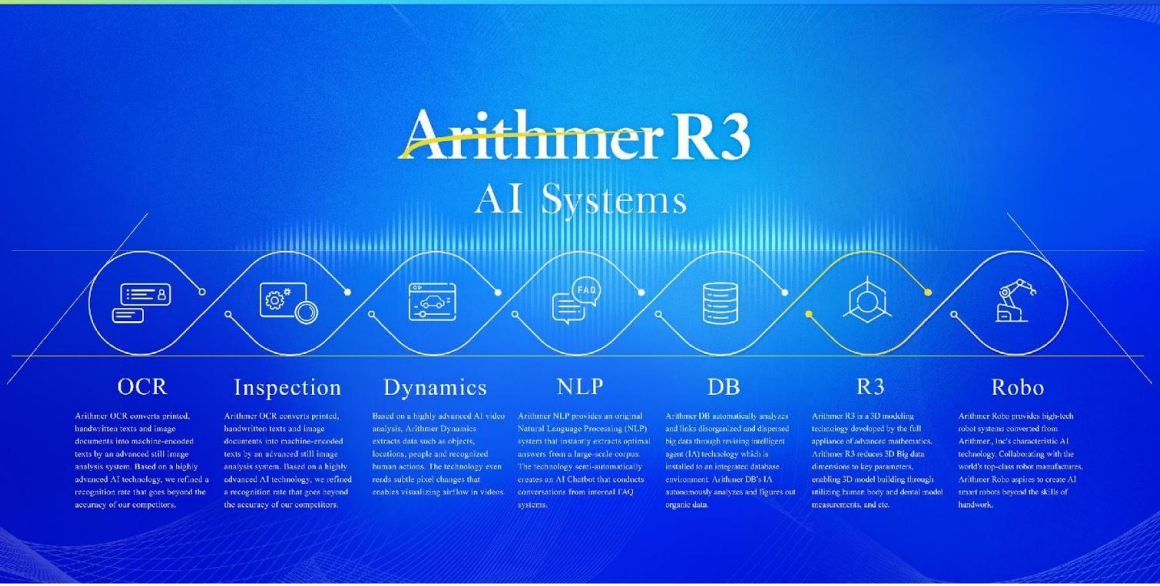


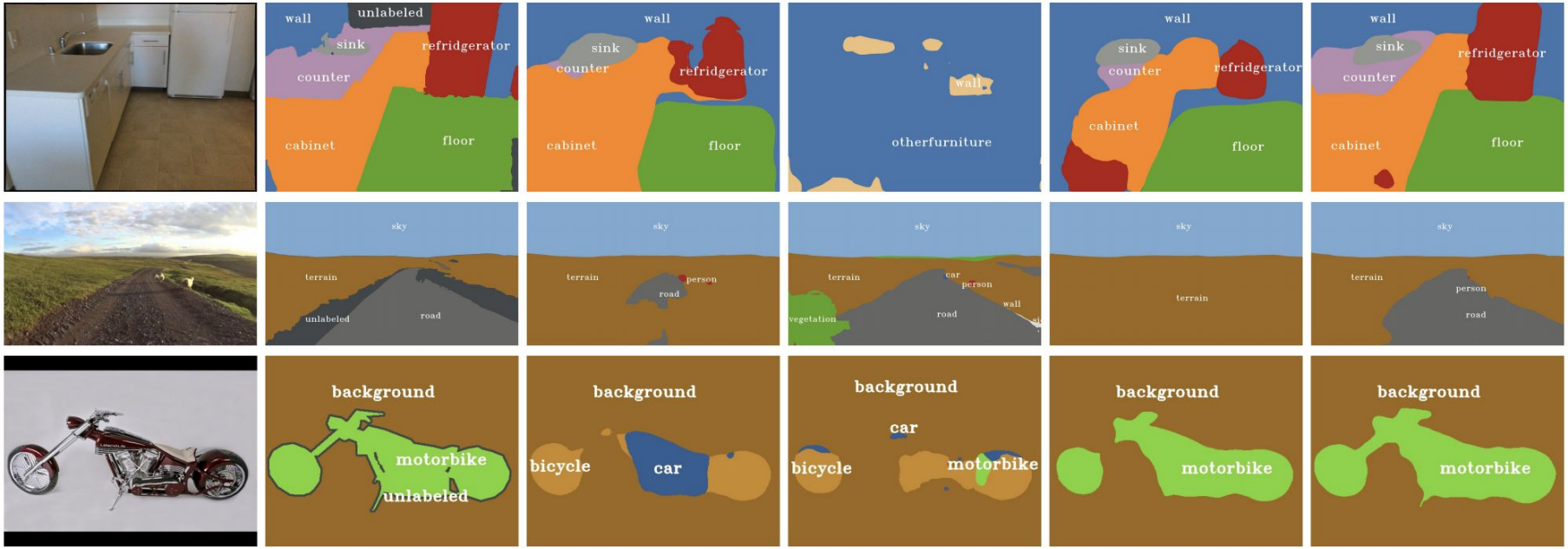
Instance Segmentation

How-to for real-time predictions

Arithmer R3 Div. Enrico Rinaldi

Enrico Rinaldi (Ph.D.)

- **Education**
  - Bachelor of Science in Physics from the University of Milan
  - Master of Science in Theoretical Physics from the University of Milan
  - Ph.D. in Theoretical Particle Physics from the University of Edinburgh
    - Computational Physics and (Big) Data Analysis
    - Particle Physics with Composite Higgs, Dark Matter, and Extra Dimensions
    - Supported by Scottish Universities Physics Alliance (SUPA) fellowship and Japanese Society for the Promotion of Science (JSPS) fellowship at the Kobayashi-Maskawa Institute of Nagoya University
- **Associate Researcher** at the Lawrence Livermore National Laboratory
  - High-Performance Computing simulations of particle physics, nuclear physics and string theory
  - Markov Chain Monte Carlo sampling and Optimization problems
- **Special Postdoctoral Fellow** at the RIKEN BNL Research Center
  - Junior PI of project on numerical simulations for composite dark matter theories
  - GPU computing for HPC simulations of nuclear physics properties
  - Leveraging machine learning techniques to accelerate physics discoveries
- **R&D Researcher and Engineer** at Arithmer Inc.
  - Research on Geometric Learning, Graph Learning, 3D perception, Computer Vision
  - Applications to robotics, automated measurements systems, optimization problems



Input image | Ground truth | ADE20K model | Mapillary model | COCO model | MSeg model

From MSeg (paper)

Instance Segmentation

- ➔ *“Instance segmentation is the task of detecting and delineating each distinct object of interest appearing in an image”* -- source
- ➔ Sub-task of:
  - ◆ “Object Detection”
  - ◆ “Semantic Segmentation”
- ➔ Improvements in baselines (R-CNN, FCN) for the “parent” tasks do not automatically apply to the “daughter” task
- ➔ Typically combines:
  - ◆ detection of boxes for all objects
  - ◆ segmentation of pixels

Methodology

- **Based on the Mask R-CNN model:**
  - Approach is “detect” and THEN “segment”: two-steps
  - A Region-based CNN (Faster R-CNN) outputs class labels and bounding-box offset for each candidate
    - Start with a Region Proposal Network (RPN)
    - Extract features from RoI and predict class and bbox
  - Additionally adds a branch to output the pixel mask of the object
    - Uses Fully Convolutional Networks (FCN) sharing weights and maintaining spatial correspondence
    - Needs alignment between pixels and feature maps (RoIAlign)

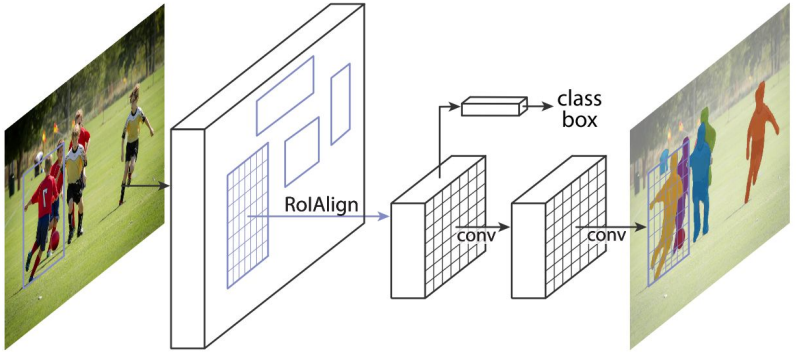
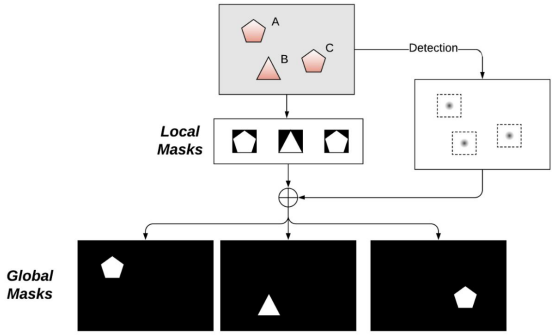


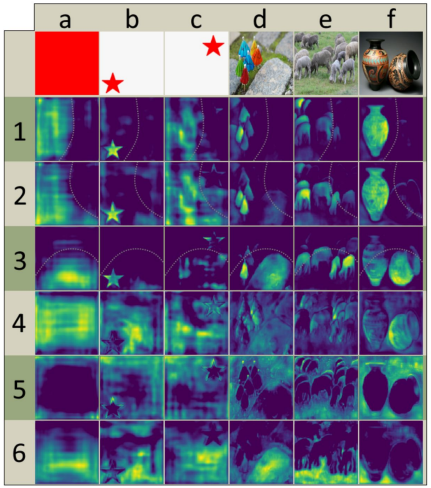
Figure 1. The **Mask R-CNN** framework for instance segmentation.

Methodology

- **Based on the YOLACT model:**
  - Approach is single-step, like anchor-free object detection (e.g. CenterNet) (but still has anchors)
  - Uses a “global mask” instead of separate masks for instances: no loss of quality due to reduced resolution
  - Performs 2 parallel tasks:
    - Generate prototype “global” masks (entire image)
    - Predict linear combination of coefficients for each instance (hence the name: **Y**ou **O**nly **L**ook **A**t **C**oefficien**T**s)
  - Instance masks are constructed by combining prototypes with the learned coefficients in an assembly step (crop to bbox)
  - Computation cost is constant with #instance



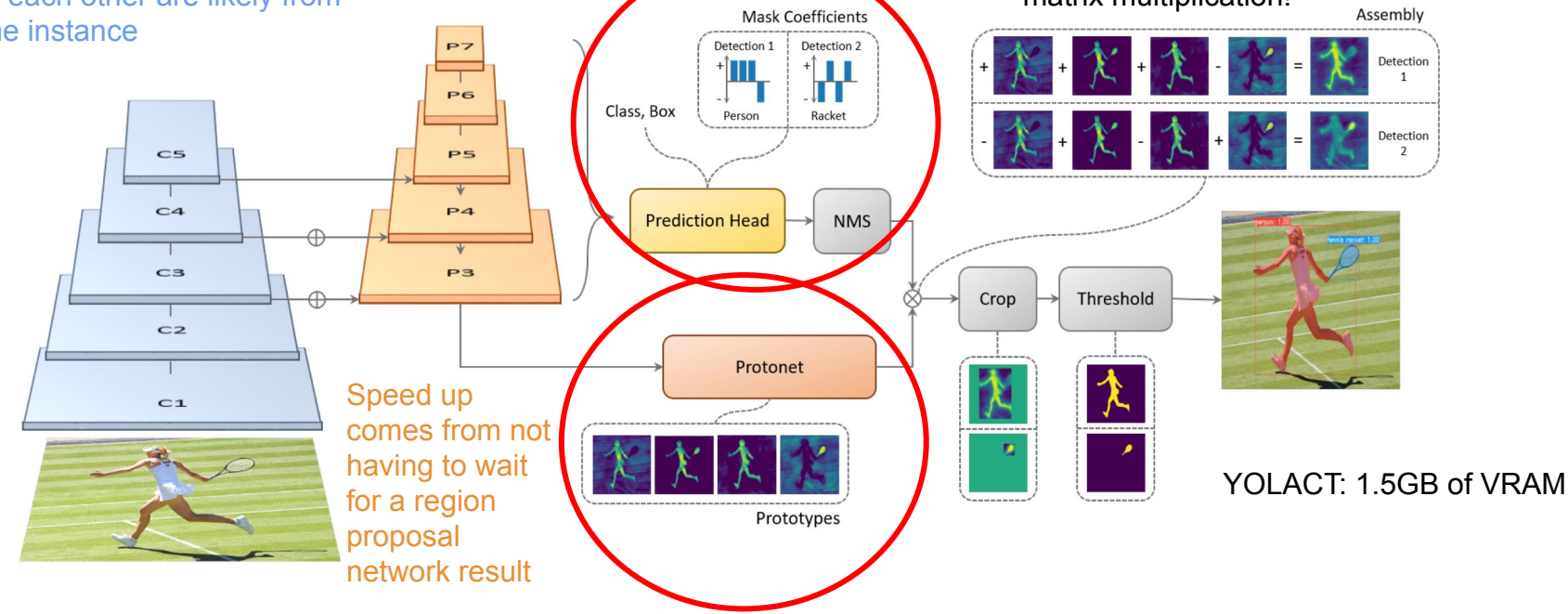
Mask representation: Local Masks and Global Masks



Use fully convolutional network: preserves spatial coherence, pixels close to each other are likely from the same instance

Parallel steps:

Overhead coming from the Assembly step: it is *just* matrix multiplication!



Speed up comes from not having to wait for a region proposal network result

YOLACT: 1.5GB of VRAM

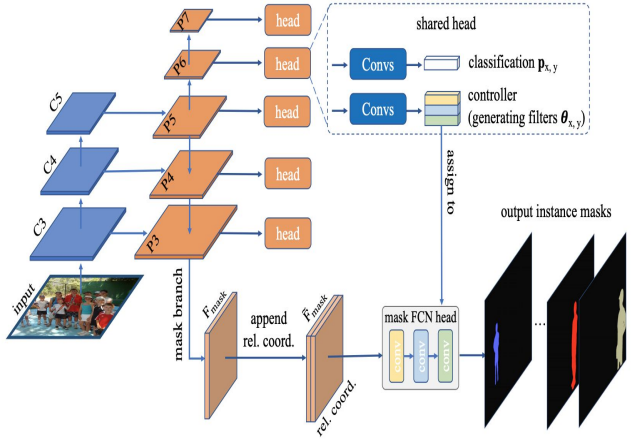
YOLACT architecture. FCN backbone, ProtoNet, Assembly step.

Related work

YOLACT (and YOLACT++) are *similar* to:

1. BlendMask (CVPR20) uses attention maps instead of coefficients
2. CenterMask (CVPR20) based on anchor-free Obj. Detection
3. CondInst removes dependency on bbox in assembly step
4. SOLO and SOLOv2 entirely bbox free: predicts instance category directly pixel by pixel

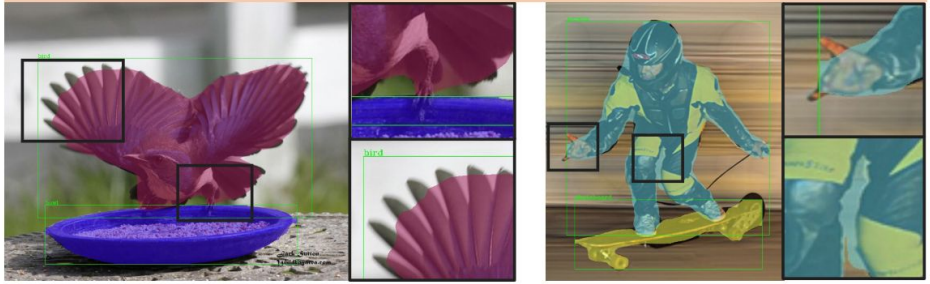
CondInst



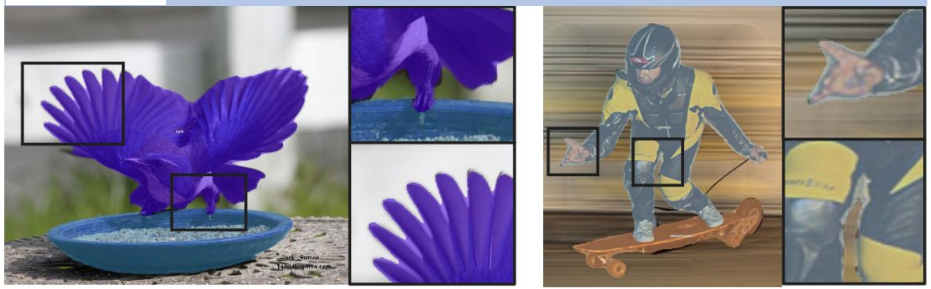
Exceptional resources on the open-source instance-segmentation toolbox from Adelaide University (on top of detectron2): AdelaiDet

Mask accuracy details

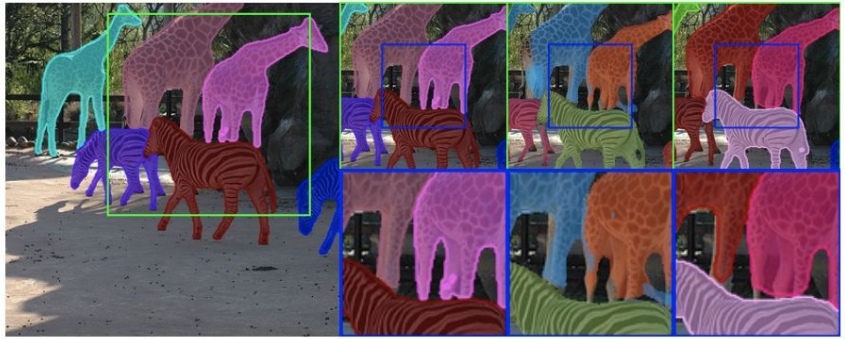
Mask R-CNN:



SOLOv2



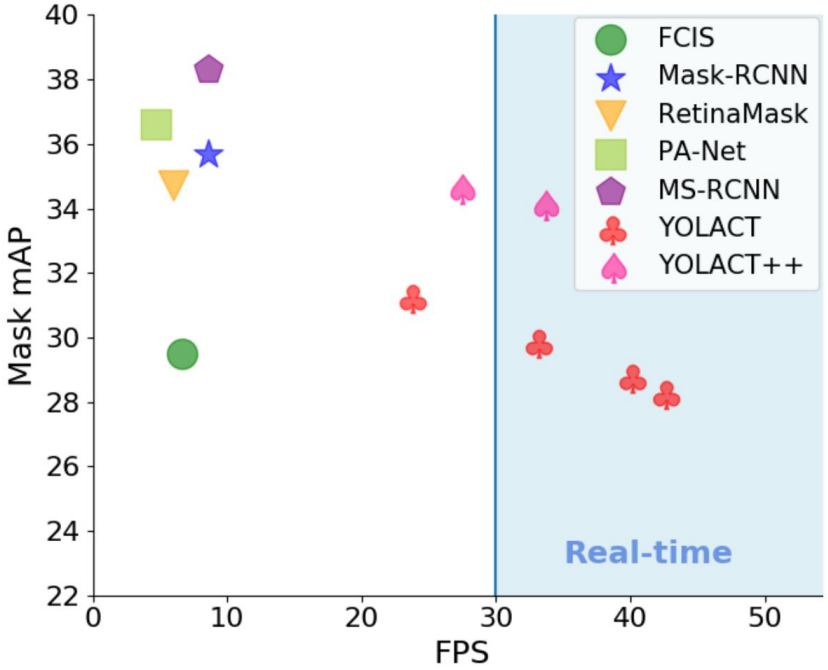
CondInst CondInst YOLACT M-RCNN



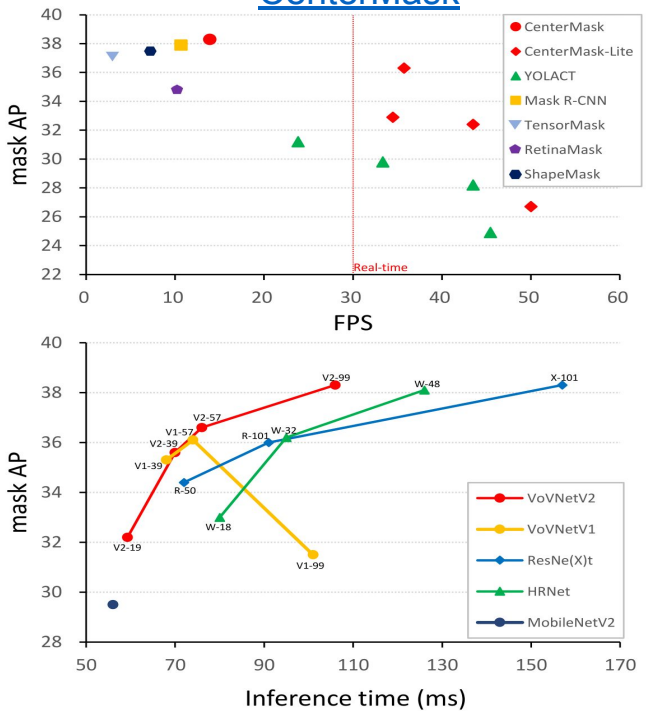
CondInst CondInst YOLACT M-RCNN

Real-time instance segmentation

YOLACT++



CenterMask



More Real-Time instance segmentation