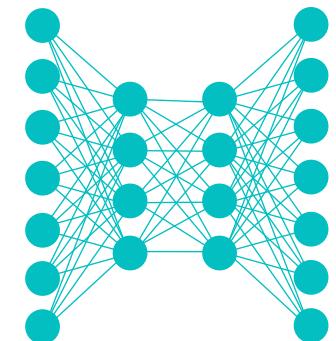


Lecture Notes for **Neural Networks** **and Machine Learning**



Fully Convolutional Learning
Instance Segmentation

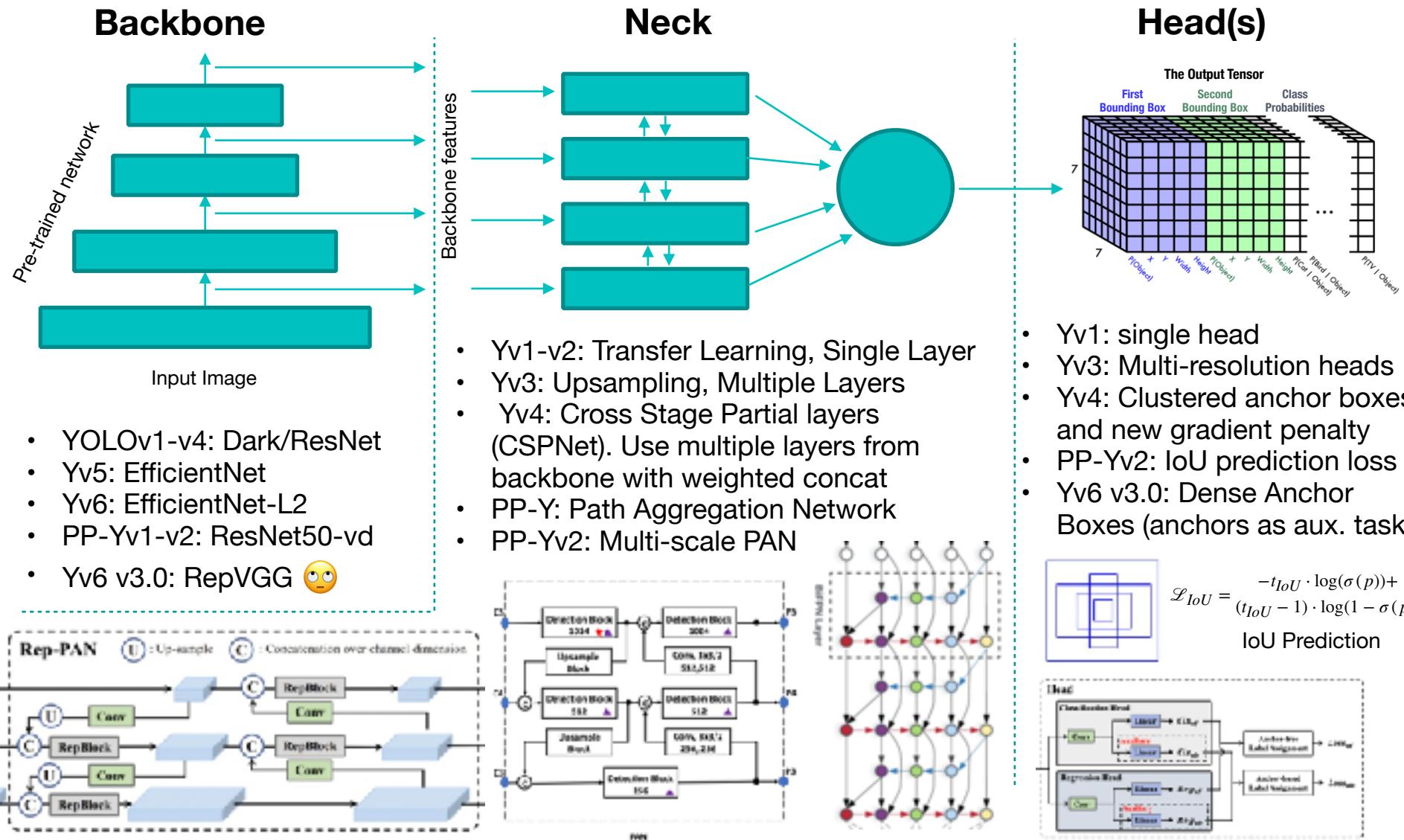


Logistics and Agenda

- Logistics
 - Labs due soon
- Agenda
 - Fully Convolutional Learning
 - ◆ Semantic Segmentation (last last time)
 - ◆ Object Detection (mostly last time, finish today)
 - ◆ Instance Segmentation (this time)



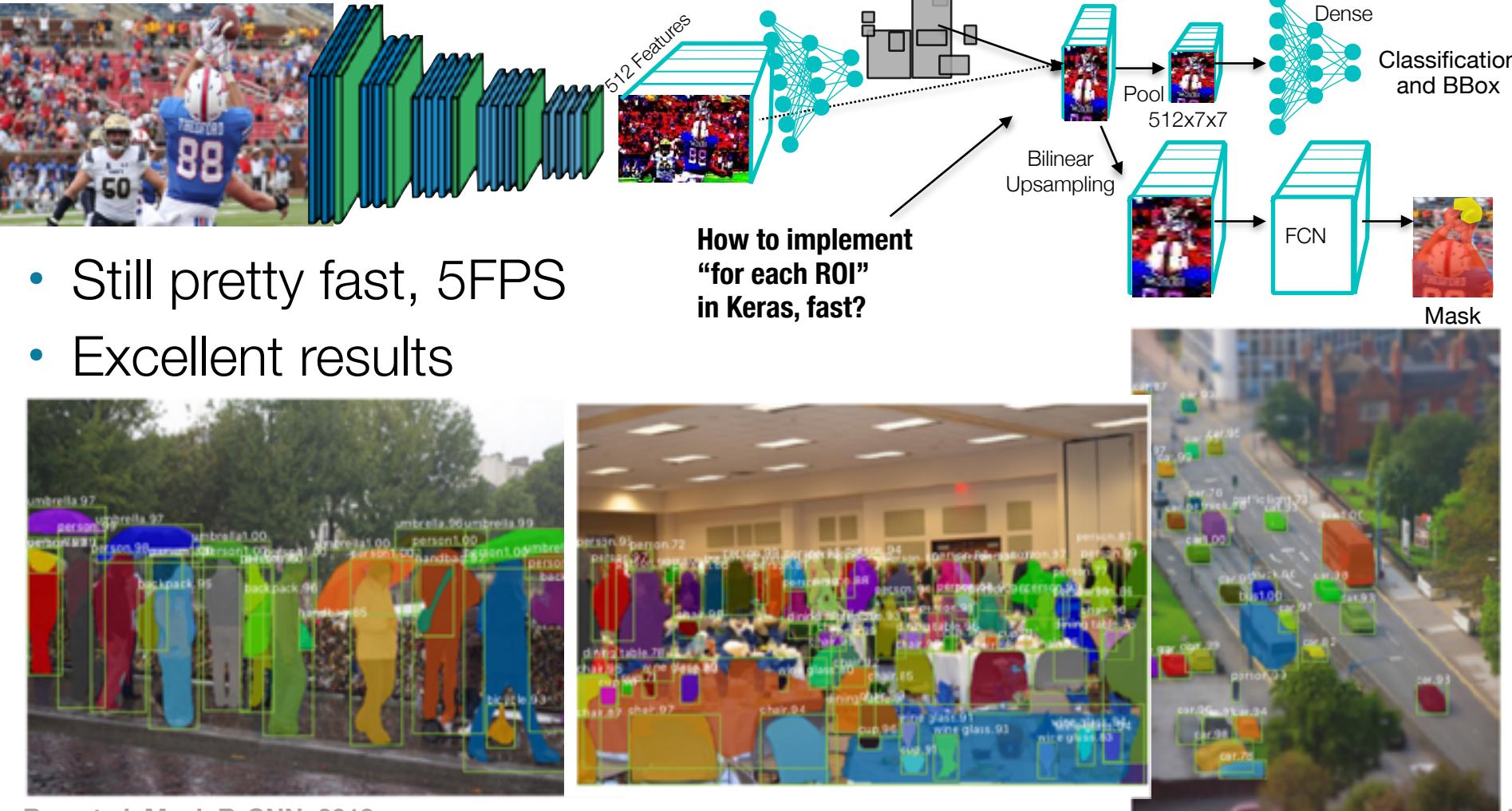
Last Time: YOLO



Instance Segmentation



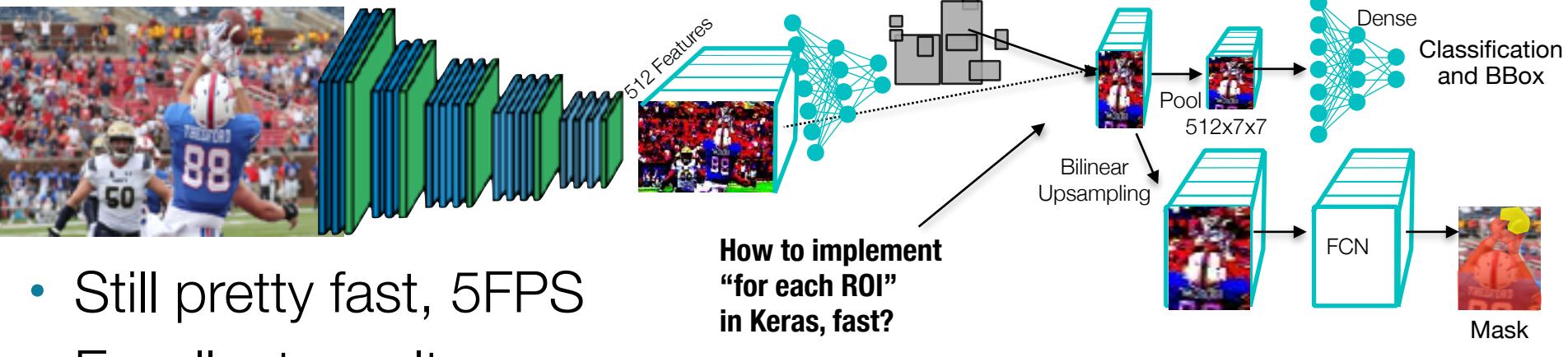
2018: Mask R-CNN



Ren et al. Mask R-CNN, 2018



2018: Mask R-CNN



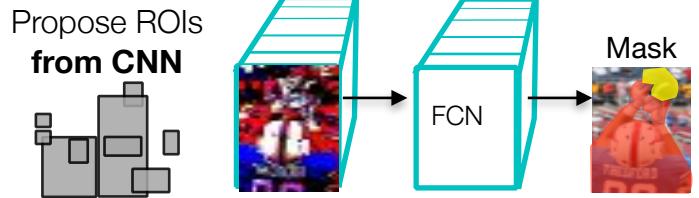
- Still pretty fast, 5FPS
- Excellent results

An Excellent, well documented Implementation here:
[https://github.com/matterport/Mask RCNN](https://github.com/matterport/Mask_RCNN)

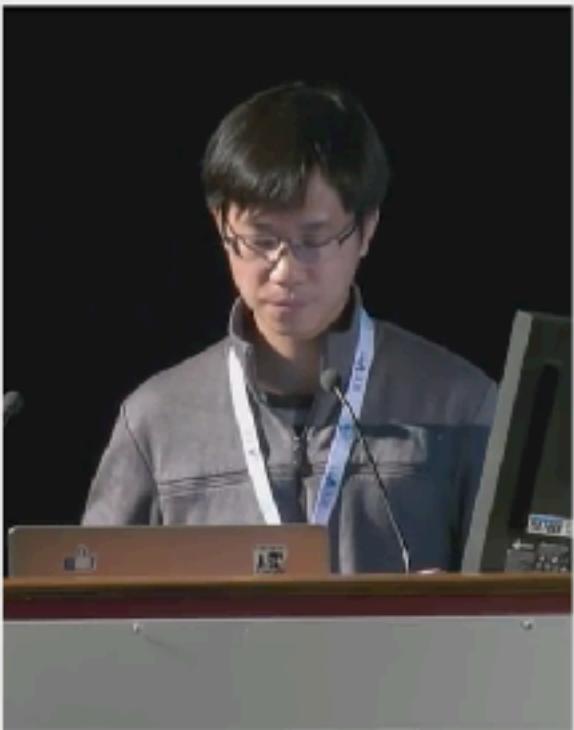
```
# Use shape of first image. Images in a batch must have the same size.  
image_shape = parse_image_meta_graph(image_meta)['image_shape'][0]  
# Equation 1 in the Feature Pyramid Networks paper. Account for  
# the fact that our coordinates are normalized here.  
# e.g. a 224x224 ROI (in pixels) maps to P4  
image_area = tf.cast(image_shape[0] * image_shape[1], tf.float32)  
roi_level = log2_graph(tf.sqrt(h * w) / (224.0 / tf.sqrt(image_area)))  
roi_level = tf.minimum(5, tf.maximum(
```



2018: Mask R-CNN



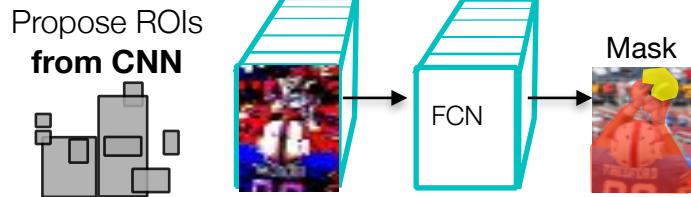
Can also provide **key point detection** from same FCN features (not real time, post processed)



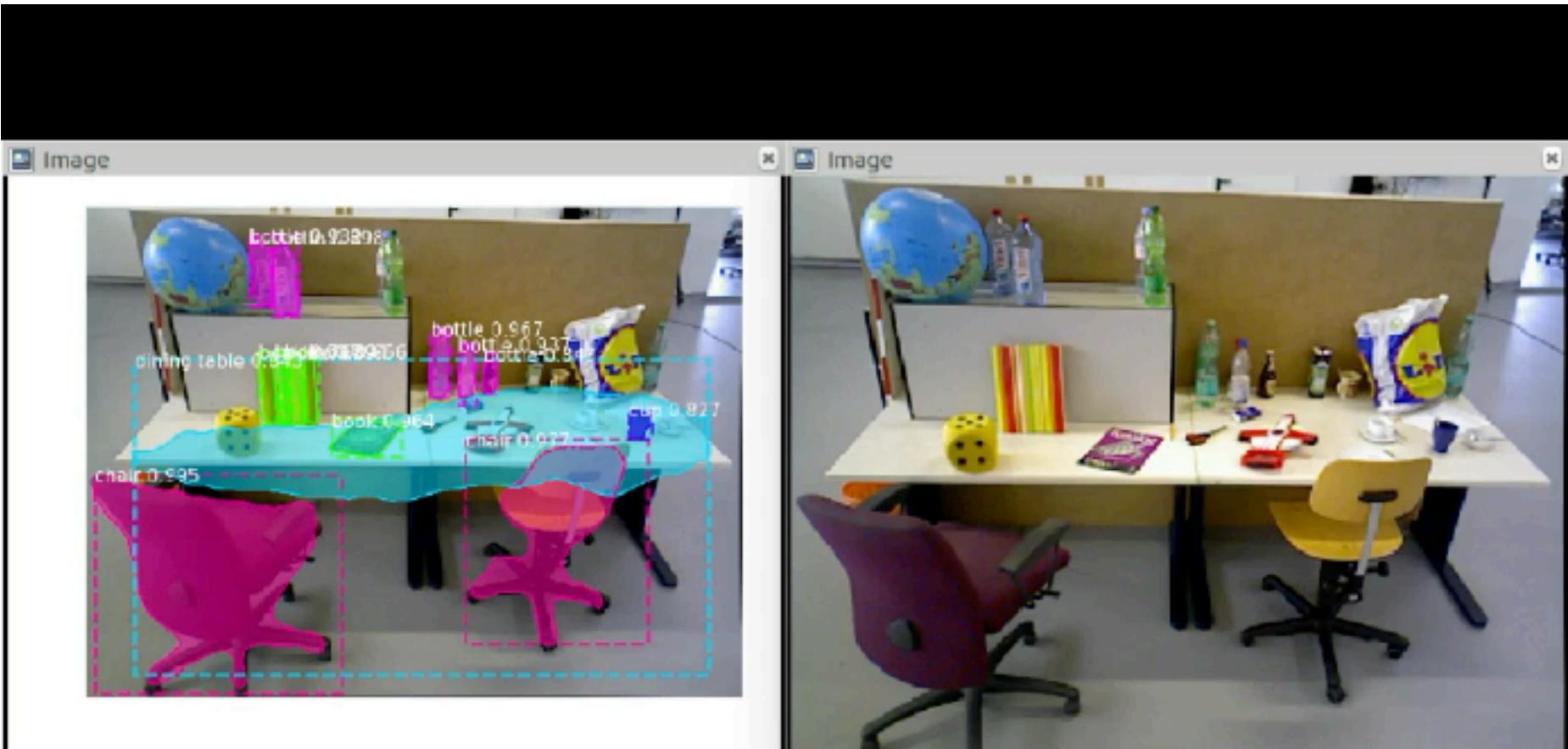
ICCV17



2018: Mask R-CNN



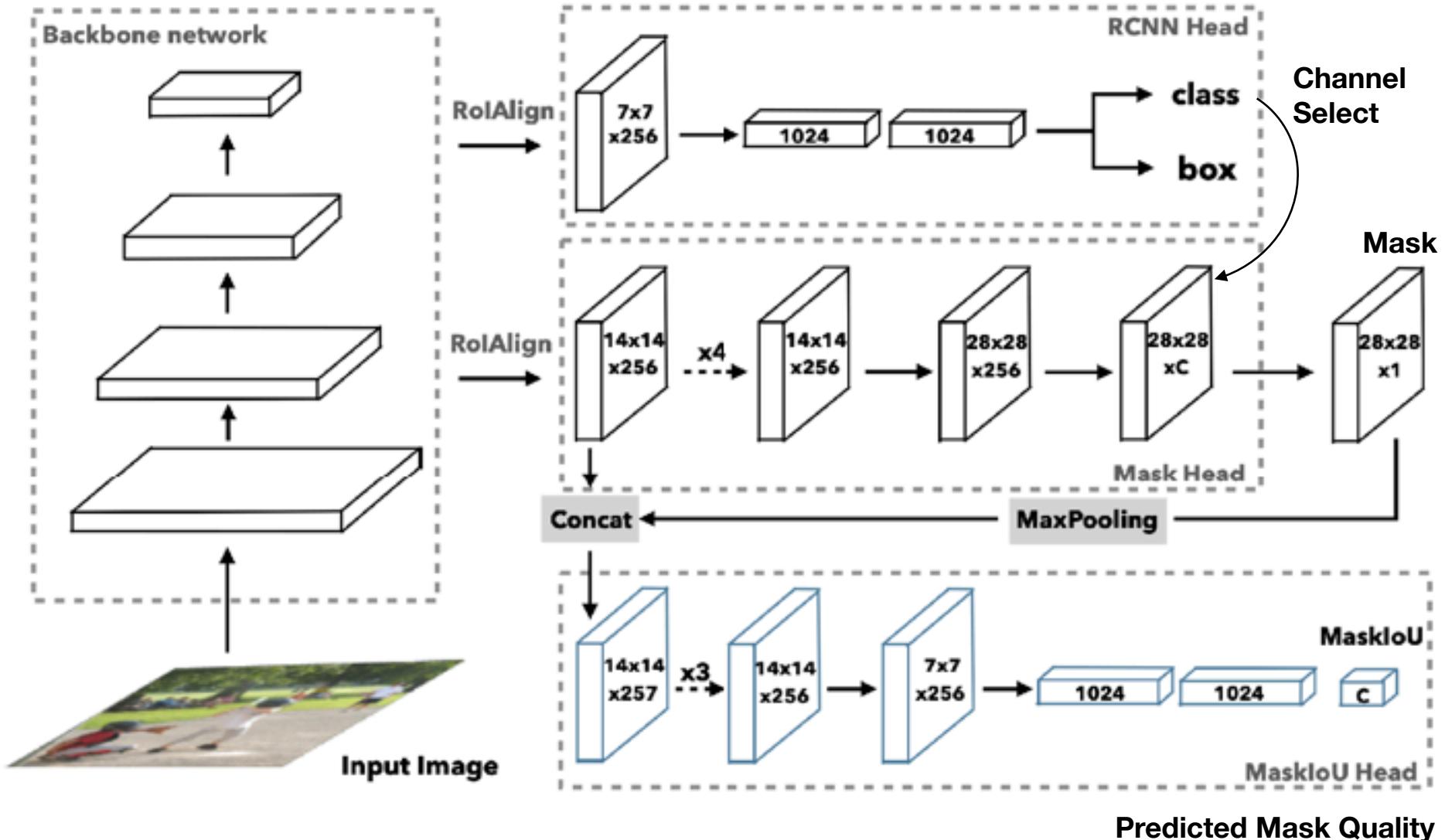
- Real time, Mask R-CNN



<https://www.youtube.com/watch?v=nEug0-pD0Ms>



March 2019: Mask Scoring RCNN, MS-RCNN



December 2019: YOLACT++

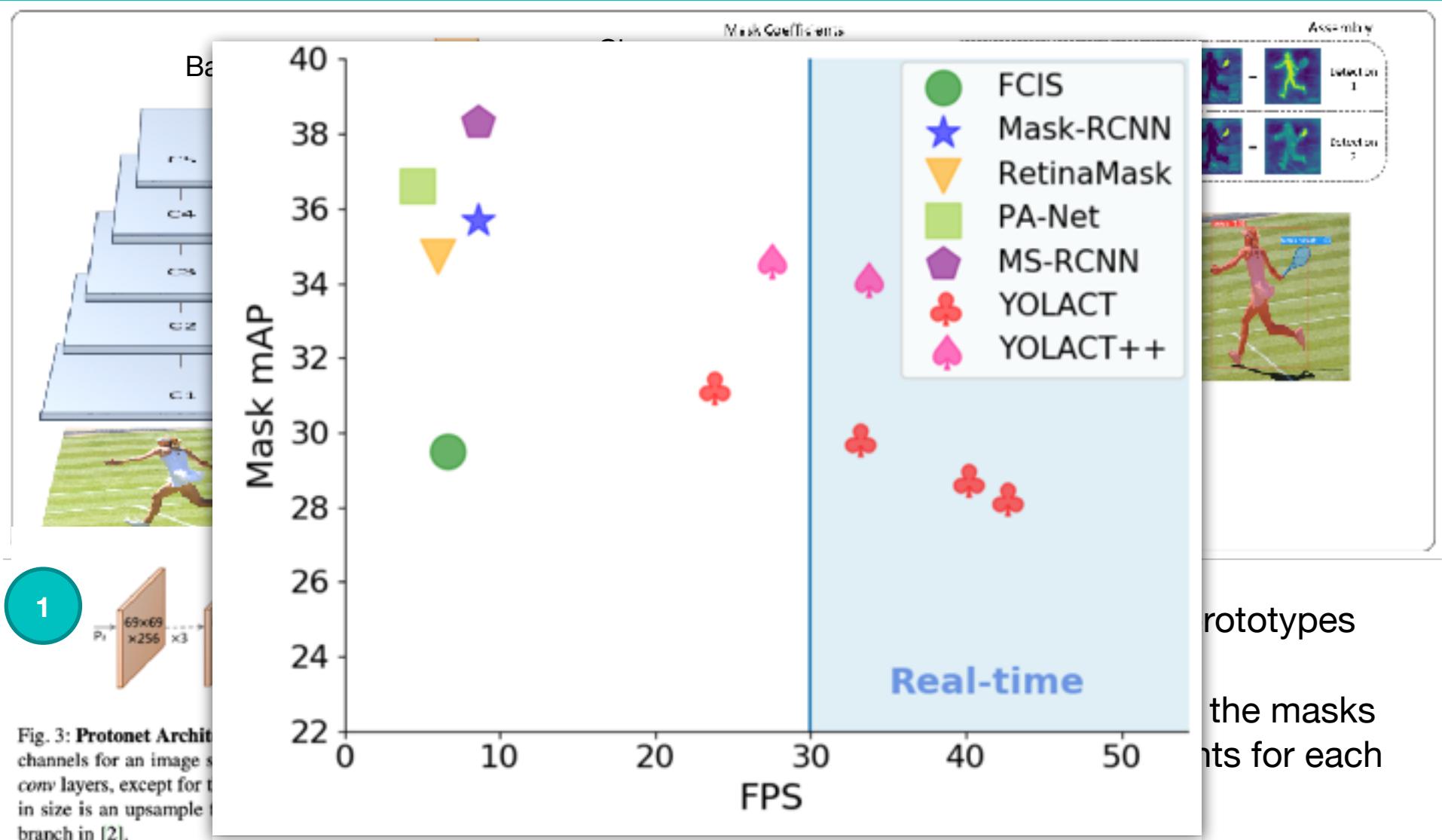


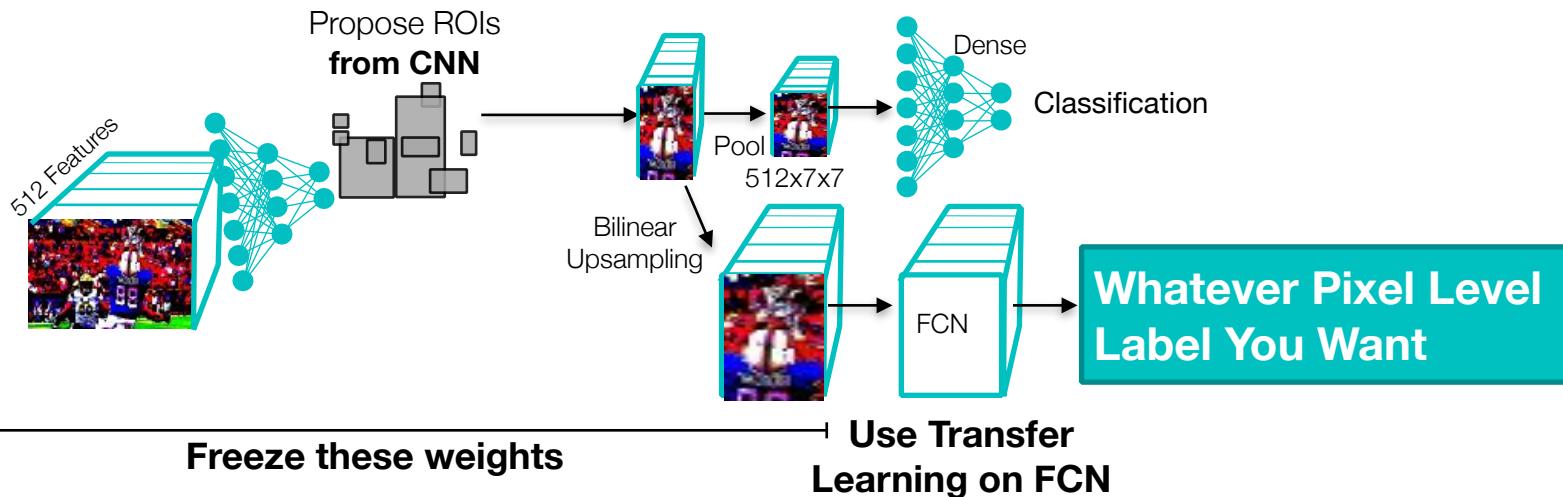
Fig. 3: Protonet Architecture. The diagram shows a backbone with layers c1, c2, c3, c4, c5, and c6. A 69x69 input is processed through a 256-channel layer P1. The architecture then branches into two parallel paths: one for detection and one for segmentation. The detection path uses a series of conv layers, while the segmentation path uses a series of deconv layers. The final output consists of detected objects and their corresponding masks.

YOLACT++ Example Video (Real time)



Expanding Masking

- **Key insight:** features that can be used for getting mask of object, are good at doing other things:
 - Like human pose estimation
 - ...Depth processing and more
- Just connect FCN to image features and learn any label



Operationalizing Masks: Ripeness Detection

Fruit Sorting: YOLO at 130FPS, ripe versus not ripe



Expanding Masks

3D Building Reconstruction (mask becomes 3D point cloud)



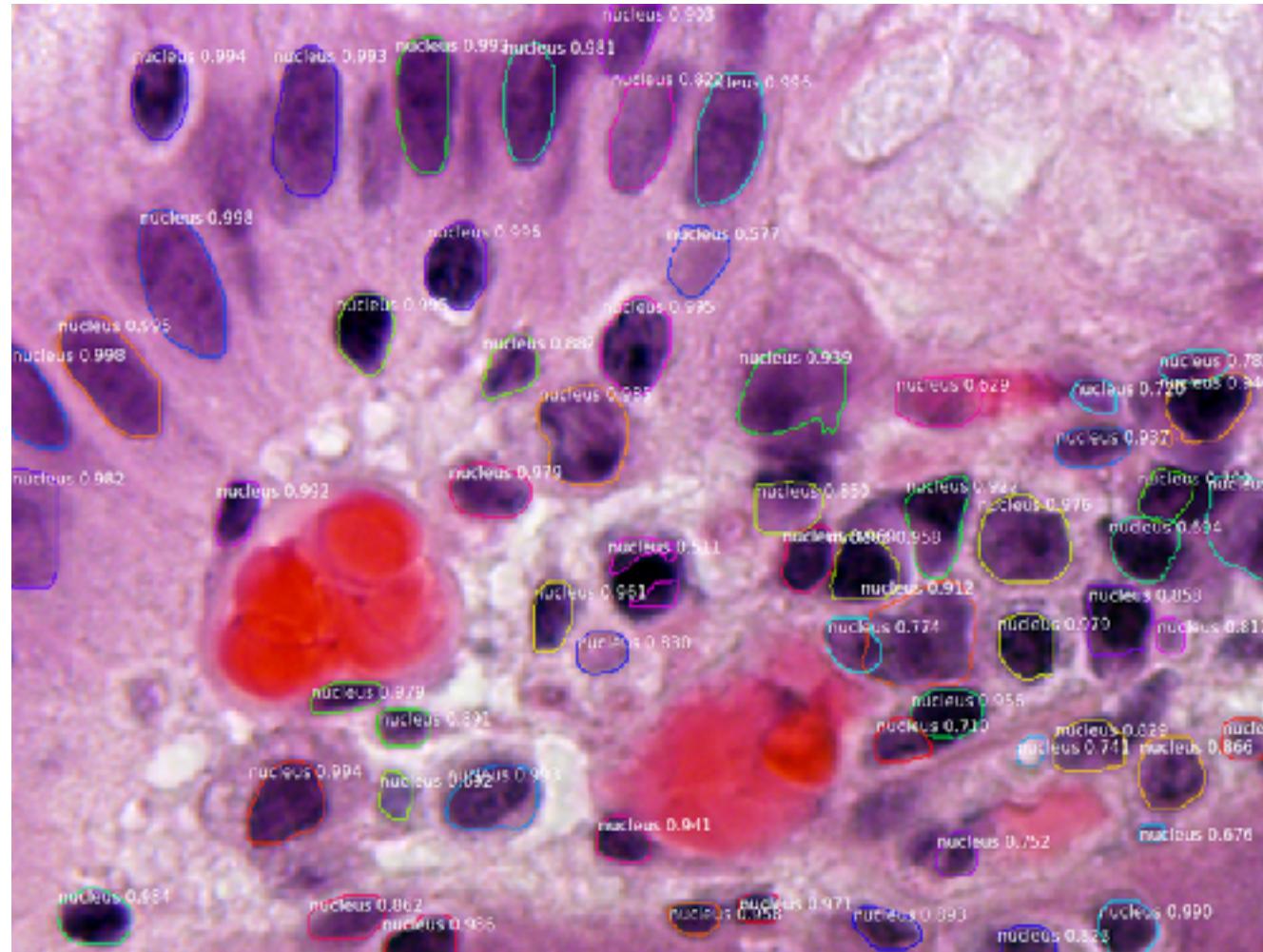
https://github.com/matterport/Mask_RCNN

14



Retraining Masks

Segmenting Nuclei

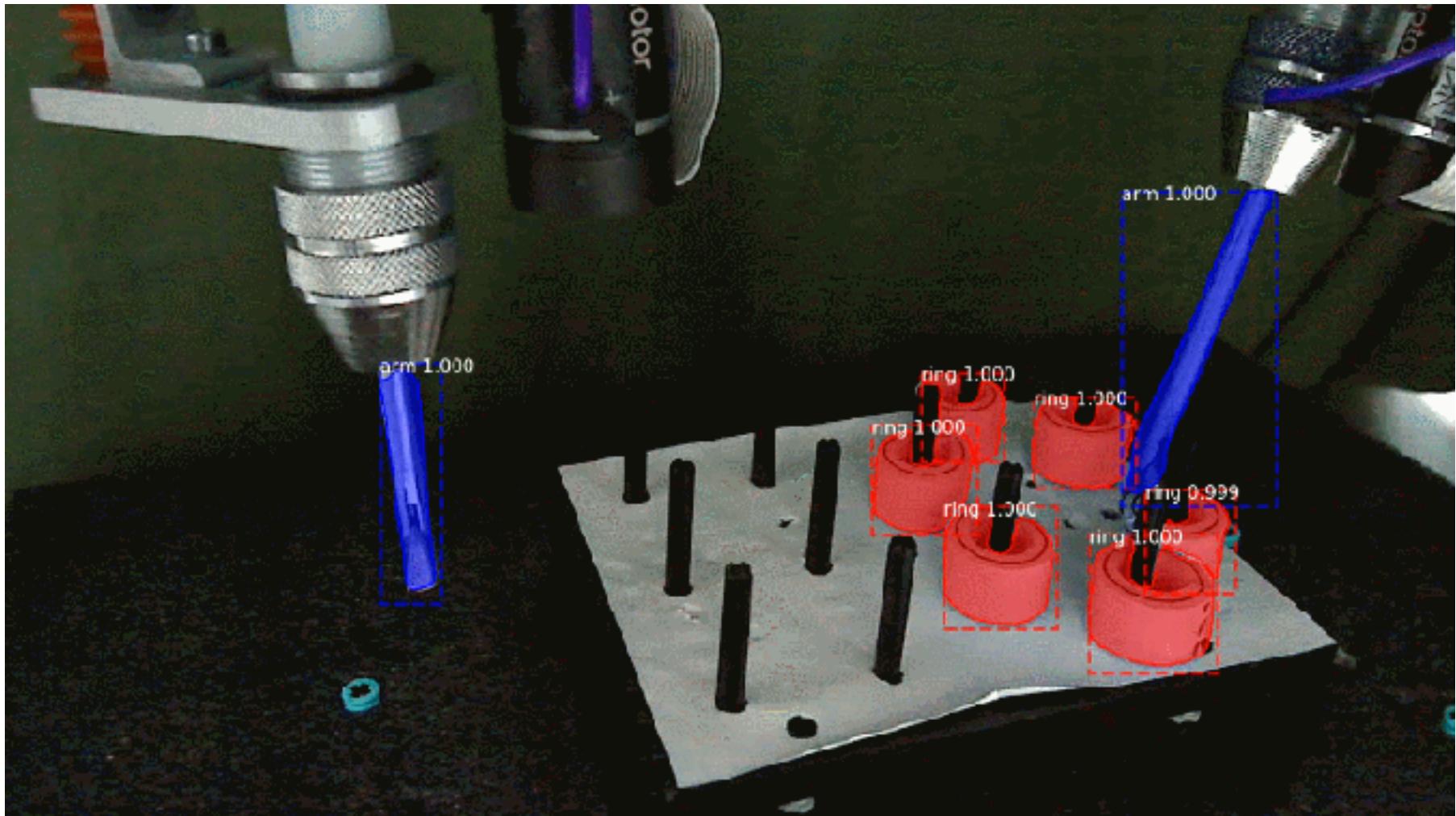


https://github.com/matterport/Mask_RCNN

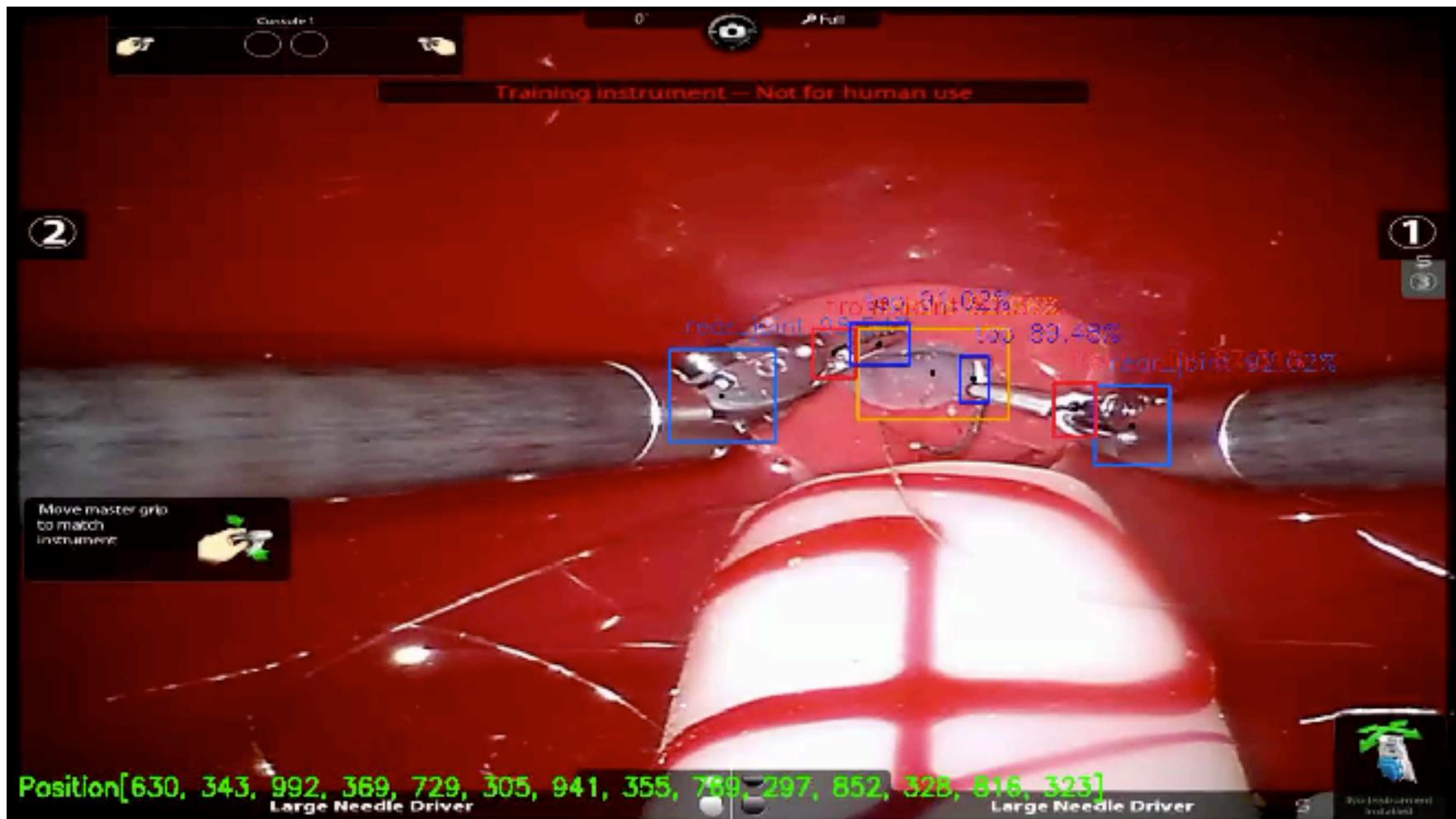


Repurposing for Robotics

Robotic Movement (like surgery)



Robotic Surgery Assessment



X. Qu, M. El-Saied, J. Gahan, R. Steinberg, and E.C. Larson (2019). "Machine Learning using a Multi-task Convolutional Neural Networks Can Accurately Provide Robotic Skills Assessment." 2019 World Congress of Endourology.

Y. Wang, J. Dai, T. Morgan, M. Elsaid, A. Garbens, X. Qu, R. Steinberg, J. Gahan, and E.C. Larson (2021). "Evaluating Robotic-Assisted Surgery Training Videos with Multi-task Convolutional Neural Networks." Journal of Robotic Surgery (JORS), 2021. Doi: 10.1007/s11701-021-01316-2

Learning Depth and 3D Shapes

Mesh R-CNN, Facebook AI January 2020



<https://ai.facebook.com/blog/pushing-state-of-the-art-in-3d-content-understanding/>

In summary

- **Semantic segmentation** through FCN is active research area
 - DeepLabV3+ or GSCNN are excellent choices, but have performance issues
 - EVA: state of the art vanilla ViT for many tasks (1B params)
- **Object detection** is excellent, ready for use in industry (Apple's ObjectDetector uses YOLO variant)
 - Already deployed in a many of Apps
 - At 60+ FPS, supports tracking applications and AR
 - Can backoff to CPU only at about 10 FPS (on phone)
- **Instance Segmentation** is ready for deployment in a number of areas, and is now better than realtime with good performance
 - Mask-RCNN, YOLACT,
 - EVA: state of the art vanilla ViT for many tasks (1B params)



Lecture Notes for **Neural Networks** **and Machine Learning**

FCN: Instance Segmentation



Next Time:
Image Style Transfer
Reading: Chollet 8.1– 8.3

