Bottom-up Instance Segmentation using Deep Higher-Order CRFs

Anurag Arnab, Philip H.S. Torr

UNIVERSITY OF OXFORD

1. Introduction

- Object Detection localises objects
  - but does not segment them.
- Semantic Segmentation labels individual pixels
  - but has no notion of different instances of the same class.
- *Instance Segmentation* recognises and localises objects at a pixel level. It's the intersection of Object Detection and Semantic Segmentation.

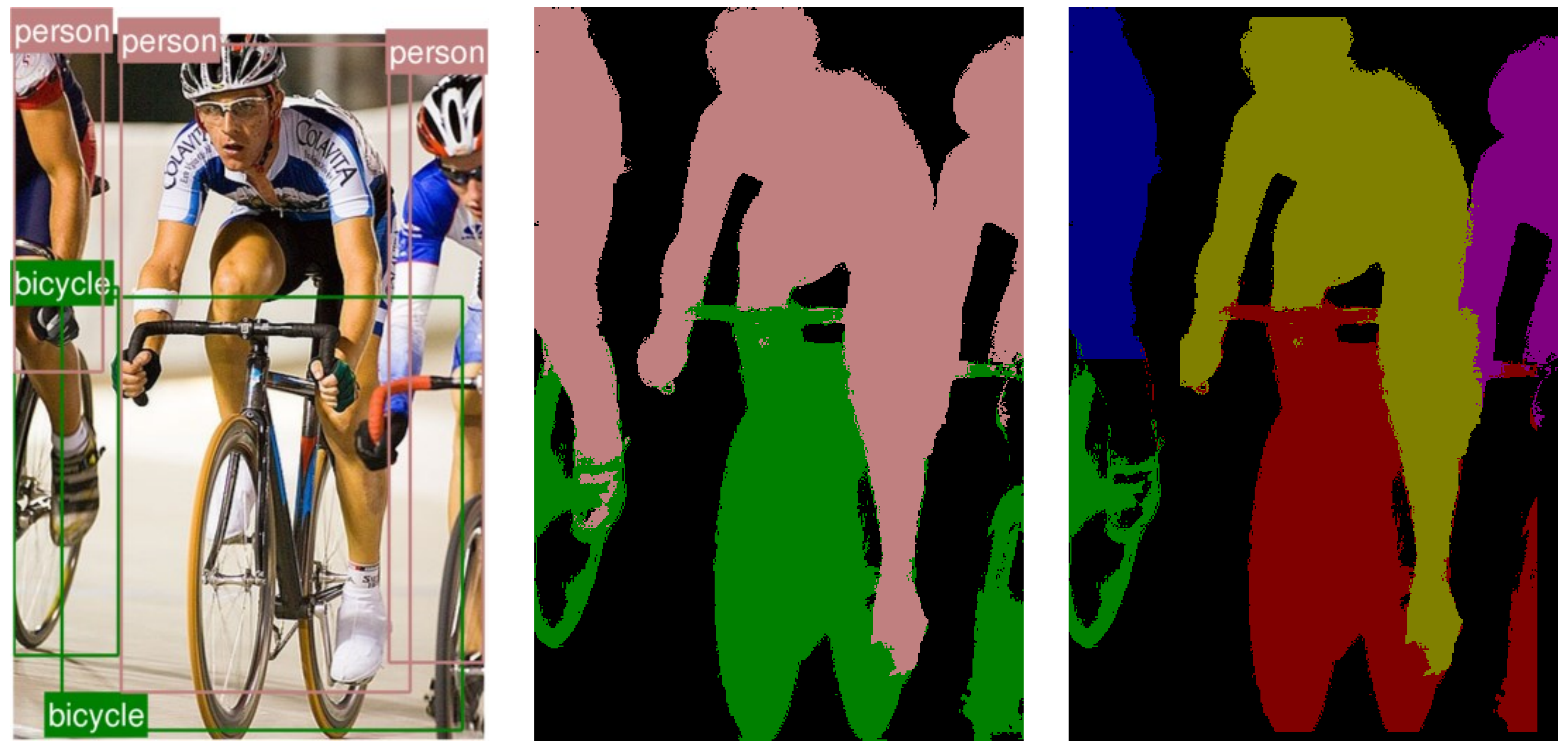

Figure 1: Instance Segmentation (right) is at the intersection of Object Detection (left) and Semantic Segmentation (middle).

2. Network Overview

We propose a simple end-to-end trainable network that leverages the great advances made in Semantic Segmentation [1] and Object Detection [2] to address the related problem of Instance Segmentation.

- Initially perform semantic segmentation of the image.
- From this category level segmentation, we reason about instances.
- We can identify instances using:
  - The outputs of an object detector [2].
  - Higher Order Detection Potentials [1] in the segmentation network, which are robust to false positive detections and recalibrate detection scores.

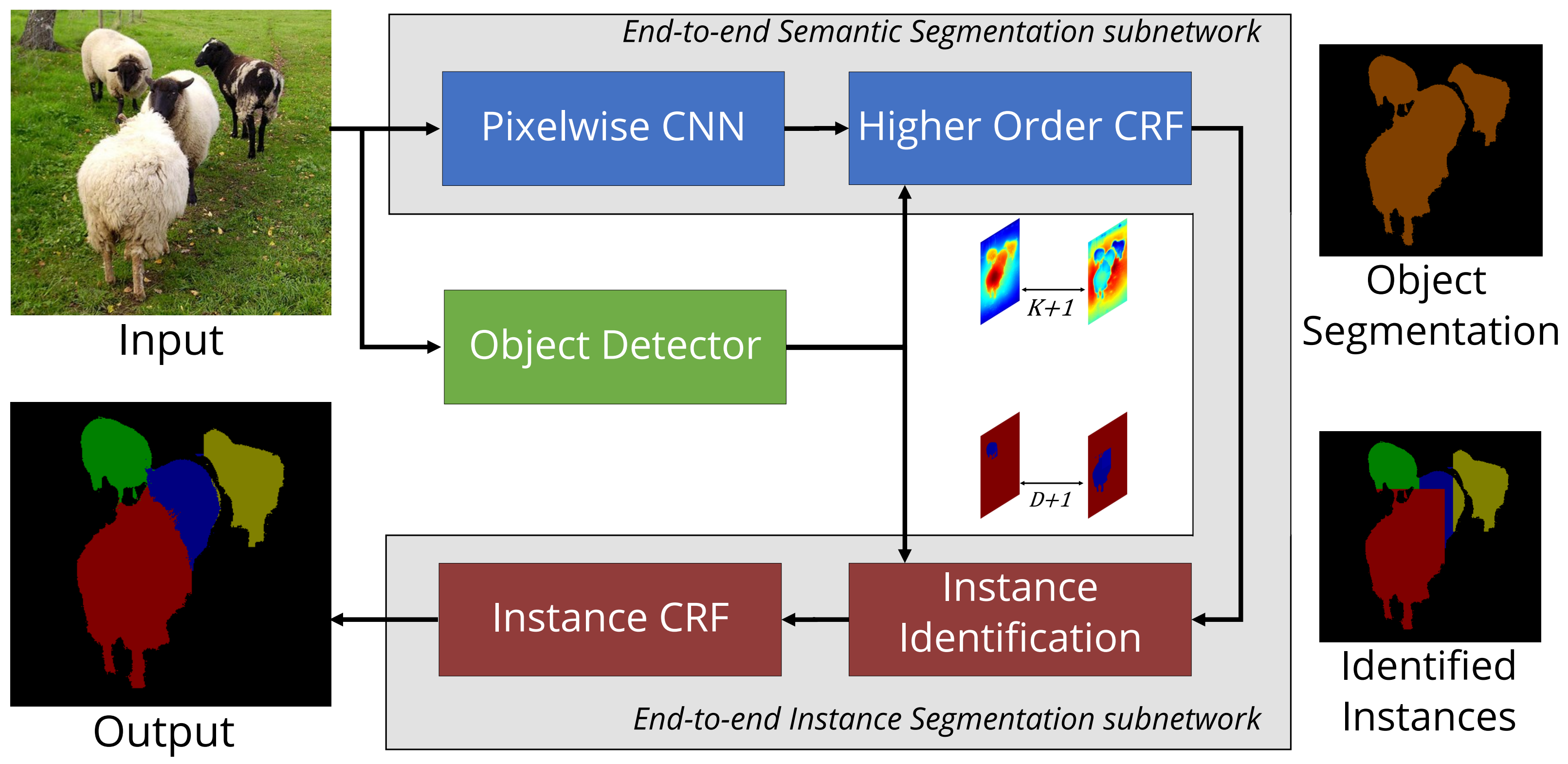

Figure 3: Overview of our proposed end-to-end method. Our system consists of an initial network for semantic segmentation, and then additional modules for instance segmentation. All modules are fully differentiable.

3. Instance Segmentation Network

From our category-level segmentation, each pixel is assigned to an object instance. Each of the D detections defines a possible instance, resulting in a problem of $D+1$ labels, including background.

If a pixel falls within the bounding box B of a detection, we probabilistically assign the pixel to that instance. The probability is proportional to the recalibrated detection score, Y (obtained from the Detection Potentials in the Higher Order CRF), and the semantic segmentation confidence for that detected class:

$$\Pr(v_i = k) = \begin{cases} \frac{1}{Z(\mathbf{Y},\mathbf{Q})} Q_i(l_k) \Pr(Y_k = 1) & \text{if } i \in B_k \\ 0 & \text{otherwise.} \end{cases}$$

Here, v_i is a multinomial random variable indicating the "identified instance" at pixel i, $Q_i(l)$ is the output of the initial category-level segmentation stage of our network and denotes the probability of pixel i taking the label l, and $Z(\mathbf{Y},\mathbf{Q})$ is the normalisation factor. This then acts as the unary potentials of a Dense CRF with pairwise terms encouraging appearance and spatial consistency [3].

4. Results

- As in Object Detection we calculate the mean Average Precision.
- However, we use the AP^r metric [4] where a prediction is considered correct if the predicted and ground truth *regions* have an Intersection over Union (IoU) above a certain threshold.
- In Object Detection, the IoU between *bounding boxes* is used.
- Perform particularly well at high thresholds which require precise segmentations.

Table 1: Results on the Pascal VOC 2012 Validation Set

Method	AP^r at 0.5	AP^r at 0.7	AP^r at 0.9	AP^r_{vol}
SDS [4]	43.8	21.3	0.9	-
Chen *et al.* [5]	46.3	27.0	2.6	-
PFN [6]	**58.7**	42.5	15.7	52.3
Ours	58.3	**45.4**	**20.1**	**53.1**

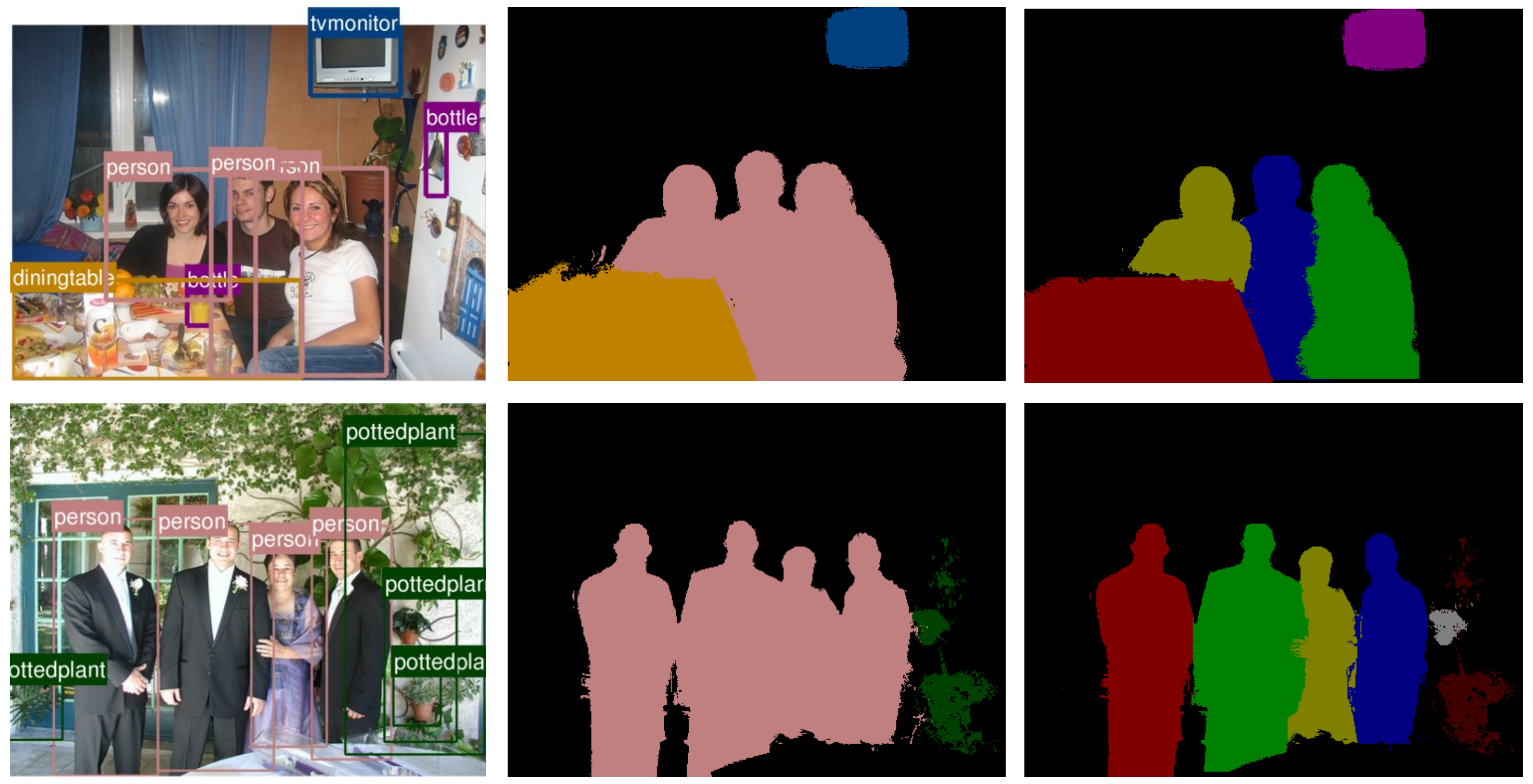

Figure 3: Left: Input image and object detections. Middle: Semantic Segmentation output. Right: Instance Segmentation output

5. Conclusion

We have presented a simple end-to-end method that effectively leverages state-of-the-art Semantic Segmentation and Object Detection networks to perform the increasingly relevant problem of Instance Segmentation. We have outperformed competing methods, particularly at high IoU thresholds which shows that utilising bottom-up segmentations enables more precise outputs.

[1] A Arnab *et al.* Higher Order Potentials in Deep Neural Networks. In *ECCV 2016*

[2] Ren *et al.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *NIPS* 2015

[3] P. Krahenbuhl and V Koltun. Efficient Inference in fully connected CRFs with Gaussian edge potentials. In *NIPS,* 2011

[4] Hariharan *et al.* Simultaneous Detection and Segmentation. In *ECCV,* 2014

[5] Y Chen *et al*. Multi-instance object segmentation with occlusion handling. In *CVPR*, 2015.

[6] X Liang *et al*. Proposal-free network for instance-level object segmentation. arXiv preprint arXiv:1509.02636, 2015.

www.robots.ox.ac.uk/~aarnab/instances | aarnab@robots.ox.ac.uk