Pixelwise Instance Segmentation with a Dynamically Instantiated Network

Anurag Arnab, Philip H.S. Torr

UNIVERSITY OF OXFORD

1. Introduction

Instance Segmentation is at the intersection of Object Detection and Semantic Segmentation. It is the task of labelling each pixel in an image with its object class, and its instance identity.

2. Limitations of prior work

Most Instance Segmentation approaches are based on modifying object detectors to output segments instead of bounding boxes. However, these approaches have numerous limitations (Fig 1):

- Multiple object proposals are processed independently.
- One pixel can be assigned to multiple instances.
- Segmentation maps of the image are not naturally produced, rather, a ranked list of proposed instances which need to post-processed.

3. Advantages of our approach

- Precise labelling due to our initial Semantic Segmentation network
- Reasons about entire image holistically
  - Pixels are assigned unique instance labels, forcing network to learn to handle occlusions (unlike detector-based methods).
- Outputs a variable number of instances depending on the image.
- Trained end-to-end with a permutation-invariant loss.

Figure 1: The winner of the latest COCO challenge, FCIS, processes each proposal independently. As a result, it struggles with false detections, overlapping instances and cannot segment outside its initial box-based proposal. We have none of these limitations.

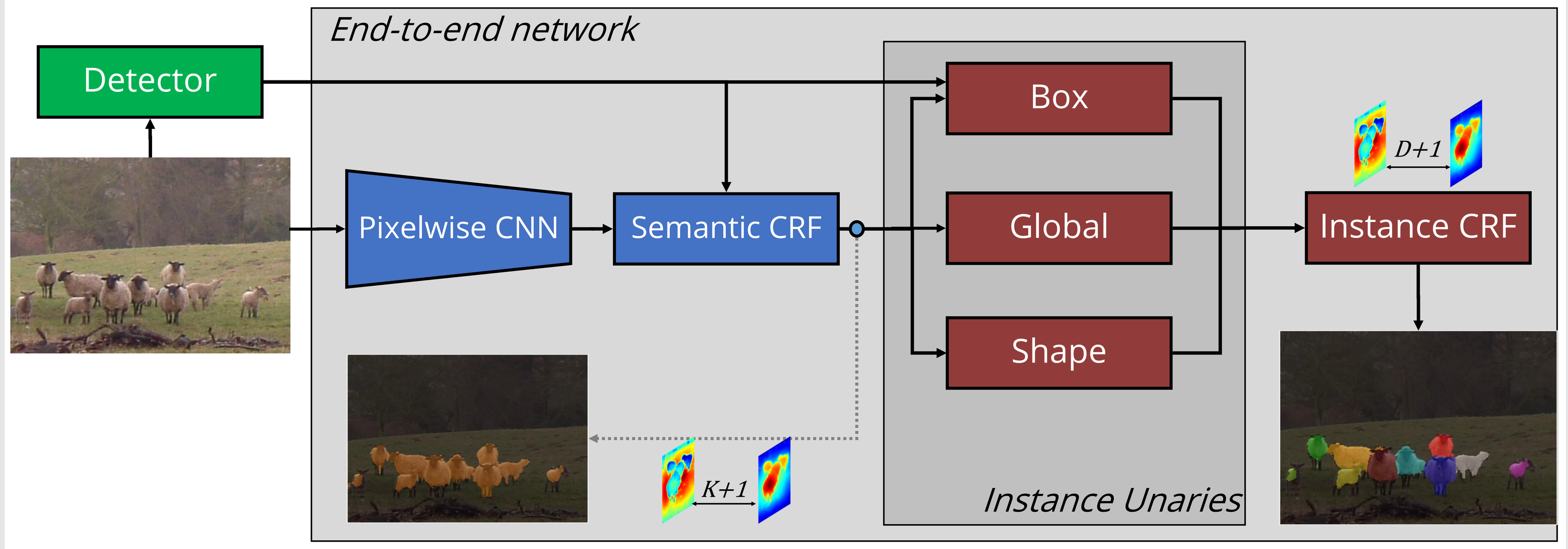


Figure 2: Overview of our proposed end-to-end method, given object detections. Weight-sharing in the Instance subnetwork allows for a dynamic number of instances per image.

4. End-to-end trained Network

Our network (Fig. 2) consists of an initial subnetwork for semantic segmentation (blue). The following instance subnetwork (red) has a CRF defined over a dynamic number of instances. It associates pixels to instances by using the cues of an object detector.

$$E(\mathbf{V} = \mathbf{v}) = \sum_i U(v_i) + \sum_{i<j} P(v_i, v_j).$$

$$U(v_i) = -\ln[w_1 \psi_{Box}(v_i) + w_2 \psi_{Global}(v_i) + w_3 \psi_{Shape}(v_i)].$$

4.1 Box Term

Encourages pixel to be an instance if it falls within its bounding box:

$$\psi_{Box}(V_i = k) = \begin{cases} Q_i(l_k) s_k & \text{if } i \in B_k \\ 0 & \text{otherwise} \end{cases}$$

Figure 3: Semantic and Instance Segmentation

4.2 Global Term

Allows us to deal with poorly localised bounding boxes

$$\psi_{Global}(V_i = k) = Q_i(l_k).$$

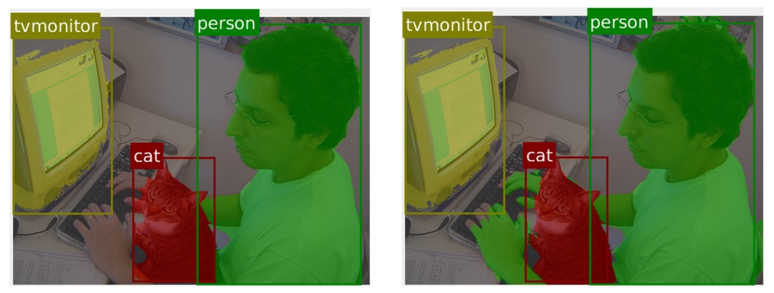


Figure 4: Without (left) and with (right) global term

4.3 Shape Term

Helps to reason about occluded objects that look the same. Shape templates learnt by network

$$t^* = \arg\max_{t \in \mathcal{T}} \frac{\sum \mathbf{Q}_{B_k}(l_k) \odot t}{\|\mathbf{Q}_{B_k}(l_k)\| \, \|t\|}$$

$$\psi(\mathbf{V}_{B_k} = k) = \mathbf{Q}_{B_k}(l_k) \odot t^*.$$

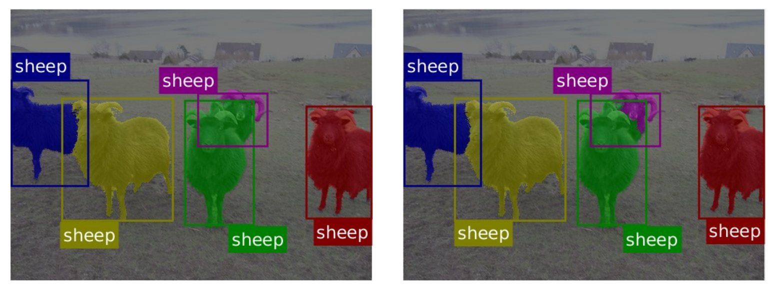
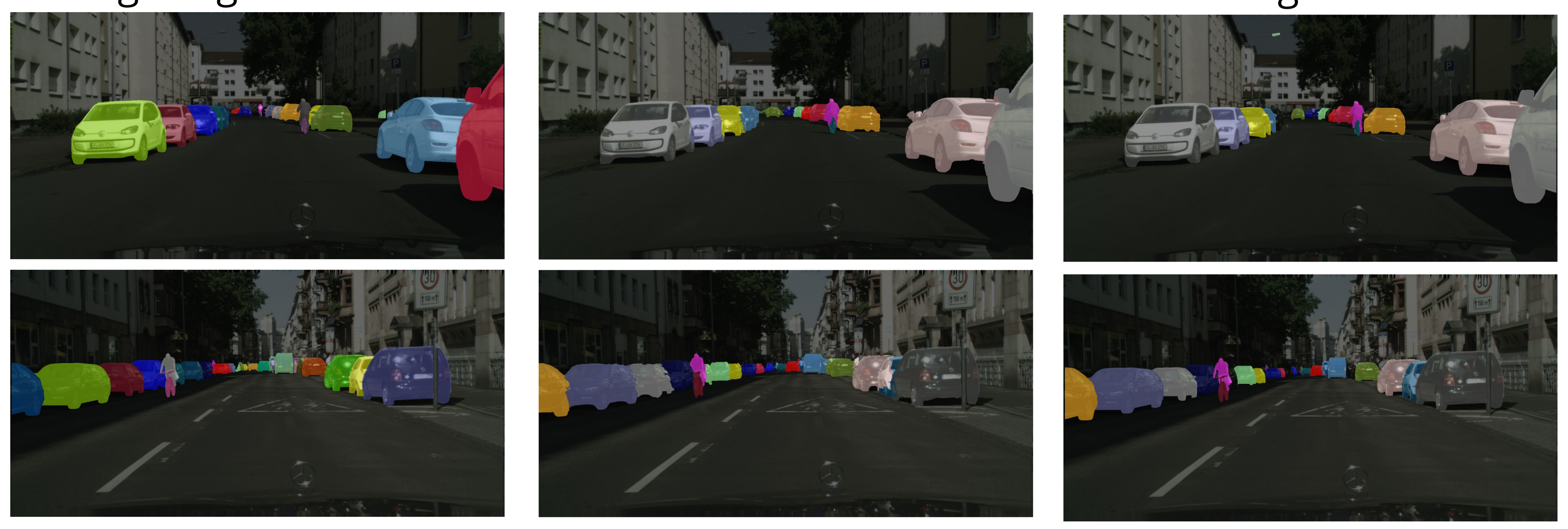


Figure 5: Without (left) and with (right) shape term

5 Loss Function

Match ground truth to prediction. Then use cross-entropy (or any other loss). Bipartite matching can be done efficiently with the Hungarian algorithm.



6. Results

Table 1: Results on Cityscapes Test Server

Method	AP^r	Method	AP^r
Ours			**20.0**
SAIS [2]	17.4	DWT [3]	15.6
InstanceCut [4]	13.0	Rec. Attend [5]	9.5

Table 2: Results on SBD Validation Set

Method	AP^r at 0.5	AP^r at 0.7	AP^r_{vol}	Matching IoU
SDS [6]	49.7	25.3	41.4	-
MPA 3-scale [7]	61.8	-	52.0	-
MNC [8]	**63.5**	41.5	-	39.0
Ours	62.0	**44.8**	**55.4**	**47.3**

Table 3: Effect of end-to-end training

Dataset	Piecewise		End-to-end	
	Semantic Seg. IoU	Instance AP^r_{vol}	Semantic Seg. IoU	Instance AP^r_{vol}
VOC	74.2	55.2	75.1	57.5
SBD	71.5	52.3	72.5	55.4

7. Conclusion

- Dynamic network, variable number of instances per image.
- Segmentation maps generated naturally; one pixel cannot belong to multiple instances.
- Training for instances improves semantic segmentation too.
- State-of-art results on Cityscapes, Pascal VOC and SBD.

[1] H Qi *et al.* Fully Convolutional Instance-Aware Semantic Segmentation. In *CVPR, 2017*
[2] Z Hayder *et al.* Boundary-aware Instance Segmentation. In *CVPR,* 2017
[3] M Bai and R Urtasun. Deep Watershed Transform for Instance Segmentation. In *CVPR,* 2017
[4] A Kirillov *et al.* Instancecut: From Edges to Instances with Multicut. In *CVPR,* 2017
[5] M Ren and R Zemel. End-to-End Instance Segmentation with Recurrent Attention. In *CVPR,* 2017
[6] B Hariharan *et al.* Simultaneous Detection and Segmentation. In *ECCV,* 2014
[7] S Liu *et al.* Multiscale Patch Aggregation for Simultaneous Detection and Segmentation. In *CVPR,* 2016.
[8] J Dai *et al.* Instance-aware Semantic Segmentation via Multi-task Network Cascades. In *CVPR,* 2016.

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