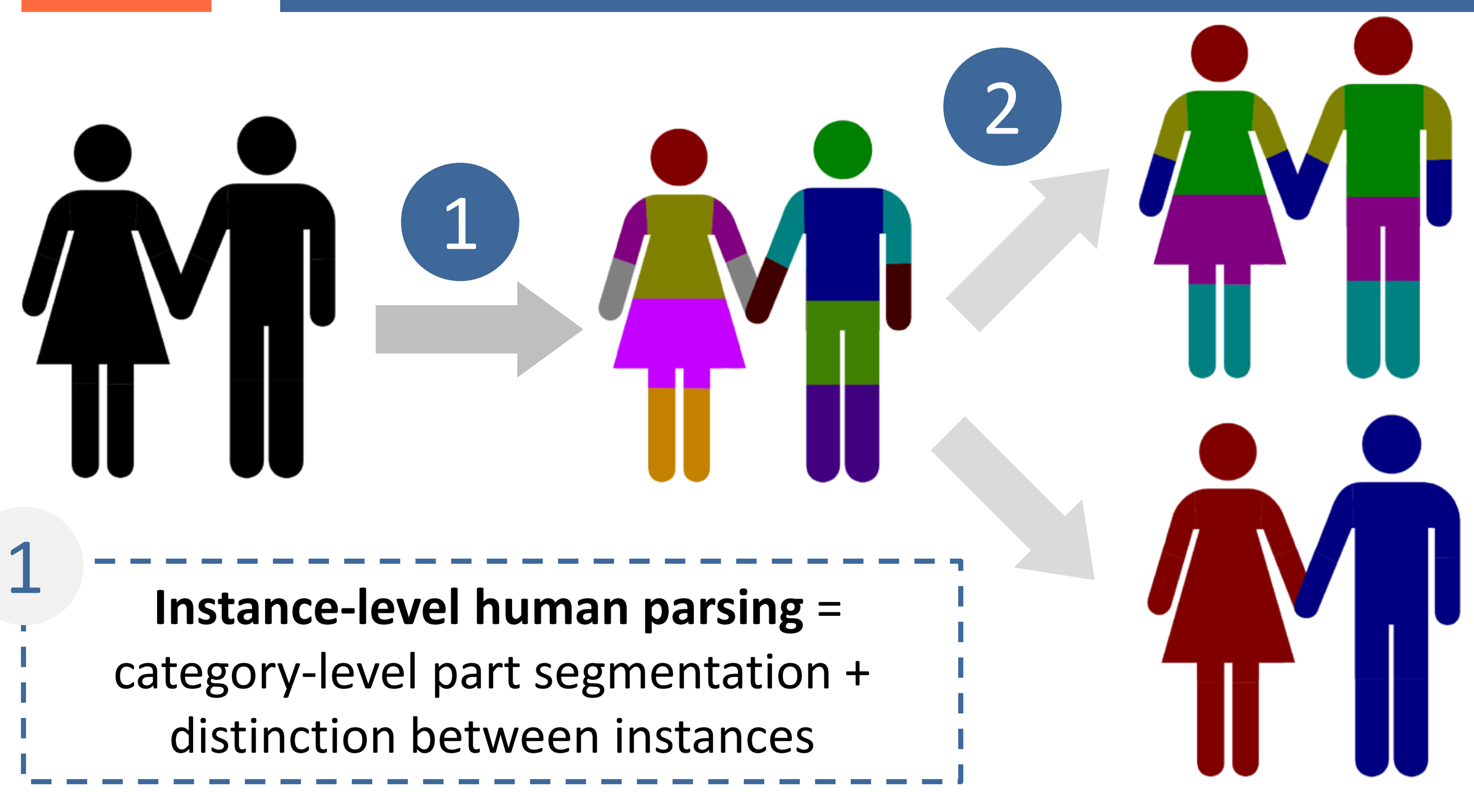
HOLISTIC, INSTANCE-LEVEL HUMAN PARSING

Qizhu Li*, Anurag Arnab*, Philip H.S. Torr

INTRODUCTION



Instance-level human parsing = category-level part segmentation + distinction between instances

A ***holistic*** solution to instance-level part segmentation is one which readily produces:

i. Instance-level human segmentation
ii. Semantic part segmentation

In contrast to existing instance segmentation methods [1], our approach segments humans at multiple granularities in a single forward pass through our network.

NETWORK OVERVIEW

Category-level Segmentation Module
- Semantically segments body parts

Human Detector
- Localises humans with bounding boxes and scores

Instance-level Segmentation Module
- Converts category-level unary potential to instance-level
- Encourages visual and spatial consistency

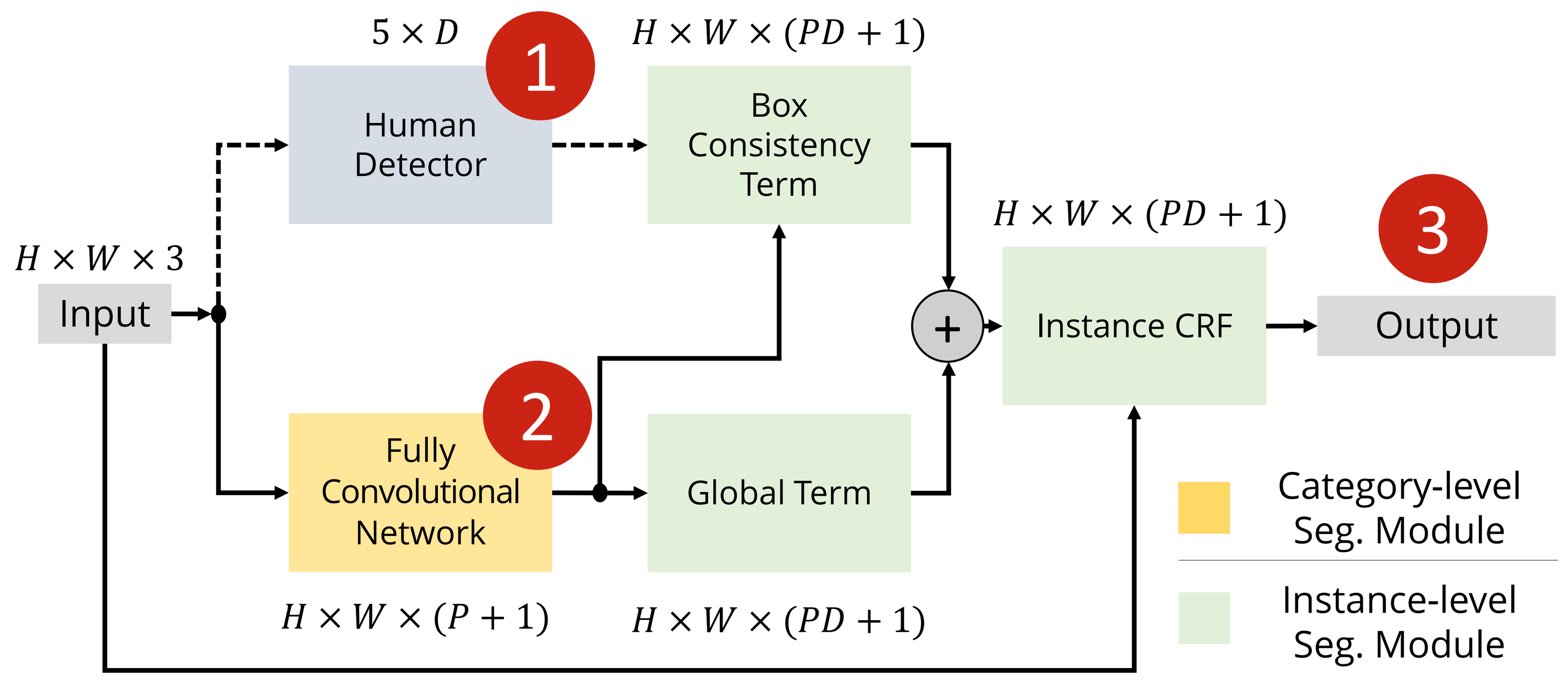


Figure 1. Overview of our proposed **end-to-end** network (given detections), containing a category-level segmentation module, a human detector, and an instance-level segmentation module. All modules are **fully differentiable**.

INSTANCE SEG. MODULE

The category-level seg. module assigns each pixel to one of the P body parts. Each of the D detections defines a possible human instance, resulting in a label space of $\{1,2,\ldots D\} \times \{1,2,\ldots P\} \cup \{(0,0)\}$, including background (0,0). A label of (i,j) denotes part j of human i.

- The **box consistency term** ψ_{Box} encourages pixels inside a human bounding box B_i to associate with the i-th human detection:

$$\psi_{Box}\left(V_k = (i,j)\right) = \begin{cases} s_i Q_k(j), & k \in B_i \\ 0, & \text{otherwise} \end{cases}$$

- The **global term** ψ_{Global} handles poor detection localisation by assuming equal likelihood for a pixel to belong to any of the detected humans:

$$\psi_{Global}\left(V_k = (i,j)\right) = Q_k(j)$$

- We formulate a **Dense CRF** [2] over these V variables:

$$E(\boldsymbol{V} = \boldsymbol{v}) = -\sum_{i}^{N} \ln\left(w_1 \psi_{Box}(v_i) + w_2 \psi_{Global}(v_i) + \varepsilon\right) + \sum_{i<j}^{N} \psi_{Pairwise}(v_i, v_j)$$

RESULTS

- We evaluate our **instance-level *part* segmentation** method on the Pascal Person-Parts (PPP) Dataset and **obtain state-of-the-art results** using Multi-task Network Cascades (MNC) [1] as a strong baseline.

The AP^r metric [3] is used to compare to other methods. A prediction is only considered correct if it has an intersection over union (IoU) with a ground truth instance above a certain threshold.

$$\mathrm{AP}^r_{\mathrm{vol}} = \sum_{i=1}^{9} \mathrm{AP}^r_{i/10}$$

Method	IoU threshold 0.5	0.6	0.7	AP^r_{vol}
MNC [1]	38.8	28.1	**19.3**	36.7
Ours, piecewise, box term only	38.7	28.9	17.5	36.7
Ours, piecewise	39.7	29.7	18.7	37.4
Ours, end-to-end	**40.6**	**30.4**	19.1	**38.4**

Table 1. Comparison of AP^r for instance-level part segmentation on PPP val. set

- Converting our output to **instance-level *human* segmentation** only involves mapping the predicted label (i,j) to i. We compare to other instance segmentation methods on the human category of the VOC12 val. set, and also **achieve state-of-the-art performance**.

Method	IoU threshold 0.5	0.6	0.7	0.8	0.9	AP^r_{vol}
SDS [3]	47.8	31.8	15.7	3.3	0.1	-
Chen *et al.* [4]	48.3	35.6	22.6	6.5	0.6	-
PFN [5]	48.4	38.0	26.5	16.5	5.9	41.3
Arnab *et al.* [6]	58.6	52.6	41.1	30.4	10.7	51.8
R2-IOS [7]	60.4	51.2	33.2	-	-	-
Arnab *et al.* [8]	65.6	58.0	46.7	33.0	14.6	57.4
Ours, piecewise	64.0	59.8	51.0	38.3	**20.1**	57.2
Ours, end-to-end	**70.2**	**63.1**	**54.1**	**41.0**	19.6	**61.0**

Table 2. Comparison of AP^r for instance-level human segmentation on VOC12 val. set

[1] J Dai, *et al*. Instance-aware semantic segmentation via multi-task network cascades. In *CVPR,* 2016.
[2] P Krahenbuhl and V Koltun. Efficient Inference in fully connected CRFs with Gaussian edge potentials. In *NIPS*, 2011
[3] B Hariharan *et al*. Simultaneous detection and segmentation. In *ECCV*, 2014.
[4] Y Chen, *et al*. Multi-instance object segmentation with occlusion handling. In *CVPR*, 2015.
[5] X Liang, *et al*. Proposal-free network for instance-level object segmentation. *arXiv preprint arXiv: 1509.02636*, 2015
[6] A Arnab, *et al*. Bottom-up instance segmentation with deep higher order crfs. In *BMVC*, 2016.
[7] X Liang, *et al*. Reversible recursive instance-level object segmentation. In *CVPR*, 2016.
[8] A Arnab, *et al*. Pixelwise instance segmentation with a dynamically instantiated network. In *CVPR*, 2017.

robots.ox.ac.uk/~aarnab/hihp
* Indicates equal contribution by the authors