

Reverse Connection with Objectness Prior Networks for Object Detection III

Liangjie Cao

21 May 2018

1 Training and Testing

In this section, the authors firstly introduce the multi-task loss function for optimizing the networks. And then they explain how to optimize the network jointly and perform inference directly.

categories. Then they use a multi-task loss L to jointly train the networks end-to-end for objectness prior, classification and bounding-box regression:

$$L = \alpha \frac{1}{N_{obj}} L_{obj} + \beta \frac{1}{L_{loc}} L_{loc} + (1 - \alpha - \beta) \frac{1}{N_{cls|obj}} N_{cls|obj}$$

3 Joint Training and Testing

They optimize the models described above jointly with end-to-end training. At the beginning of training, the objectness prior maps are in chaos. However, along with the training progress, the objectness prior maps are more concentrated on areas covering objects. [2] Actually at training phase, each SGD mini-batch is constructed from N images chosen uniformly from the dataset. However, along with the training progress, the objectness prior maps are more concentrated on areas covering objects.

4 Combining Objectness Prior with Detection

In this section, we explain how RON combines objectness prior with object detection. We assist object detection with objectness prior both in training and testing phases. And at the back-propagation phase, the network firstly generates the objectness prior, then for detection, samples whose objectness scores are high than threshold are selected (Figure 2). Then Table 1 shows the result comparisons of the methods. I will continue learning in the following days. ¹

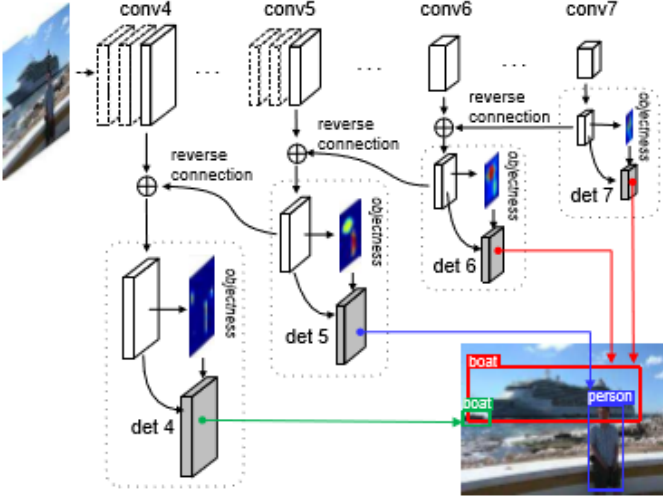


Figure 1: RON object detection overview

2 Loss Function

They have three output branches. The first outputs the objectness confidence score $p^{obj} = \{p^{obj0}, p^{obj1}\}$, computed by a Softmax over the $2 \times A$ outputs of objectness prior ($A = 10$ in this paper, as there are 10 types of default boxes). The second branch outputs the boundingbox regression loss, denoted by L_{loc} . It targets at minimizing the smoothed L1 loss [4] between the predicted location offsets $t = (t_x, t_y, t_w, t_h)$ and the target offsets $t1 = (t1_x, t1_y, t1_w, t1_h)$. The third branch outputs the classification loss $L_{cls|obj}$ for each box, over $K+1$

¹Actually they note that the latest SSD uses new training tricks (color distortion, random expansion and online hard example mining), which makes the results much better. We expect these tricks will also improve our results, which is beyond the focus of this paper.

Method	map	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Fast R-CNN [4]	70.0	77.0	78.1	69.3	59.4	38.3	81.6	78.6	86.7	42.8	78.8	68.9	84.7	82.0	76.6	69.9	31.8	70.1	74.8	80.4	70.4
Faster R-CNN [1]	73.2	76.5	79.0	70.9	65.5	52.1	83.1	84.7	86.4	52.0	81.9	65.7	84.8	84.6	77.5	76.7	38.8	73.6	73.9	83.0	72.6
SSD300 [3]	72.1	75.2	79.8	70.5	62.5	41.3	81.1	80.8	86.4	51.5	74.3	72.3	83.5	84.6	80.6	74.5	46.0	71.4	73.8	83.0	69.1
SSD500 [3]	75.1	79.8	79.5	74.5	63.4	51.9	84.9	85.6	87.2	56.6	80.1	70.0	85.4	84.9	80.9	78.2	49.0	78.4	72.4	84.6	75.5
RON320	74.2	75.7	79.4	74.8	66.1	53.2	83.7	83.6	85.8	55.8	79.5	69.5	84.5	81.7	83.1	76.1	49.2	73.8	75.2	80.3	72.5
RON384	75.4	78.0	82.4	76.7	67.1	56.9	85.3	84.3	86.1	55.5	80.6	71.4	84.7	84.8	82.4	76.2	47.9	75.3	74.1	83.8	74.5
RON320++	76.6	79.4	84.3	75.5	69.5	56.9	83.7	84.0	87.4	57.9	81.3	74.1	84.1	85.3	83.5	77.8	49.2	76.7	77.3	86.7	77.2
RON384++	77.6	86.0	82.5	76.9	69.1	59.2	86.2	85.5	87.2	59.9	81.4	73.3	85.9	86.8	82.2	79.6	52.4	78.2	76.0	86.2	78.0

Table 1: **Detection results on PASCAL VOC 2007 test set. The entries with the best APs for each object category are bold-faced.**

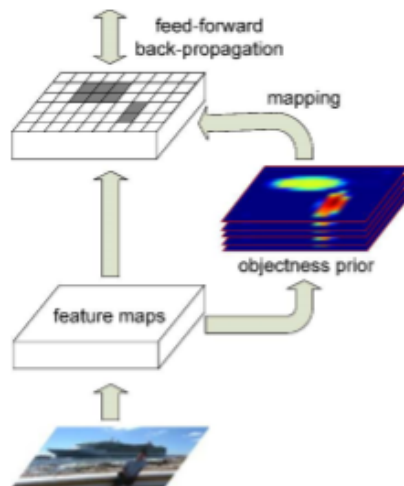


Figure 2: Mapping the objectness prior with object detection

References

- [1] S. Ren *etal.* Towards real-time object detection with region proposal networks. *In NIPS*, 2015.
- [2] Tao Kong *etal.* Ron: Reverse connection with objectness prior networks for object detection. *In CVPR*, 2017.
- [3] W. Liu *etal.* Single shot multibox detector. *In ECCV*, 2014.
- [4] R. Girshic. Fast r-cnn. *In ICCV*, 2015.