

A MultiPath Network for Object Detection

Sergey Zagoruyko, Adam Lerer, Tsung-Yi Lin, Pedro O. Pinheiro, Sam Gross, Soumith Chintala, Piotr Dollár

Facebook AI Research (FAIR)

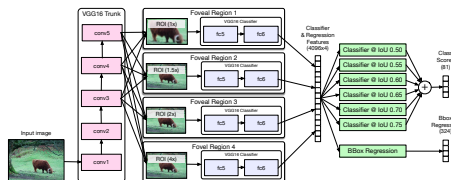


Figure 1: Proposed MultiPath architecture. COCO contains objects at multiple scales, in context and among clutter, and under frequent occlusion. Moreover, the COCO evaluation metric rewards high quality localization. To address this, we propose the MultiPath network pictured above, which contains three key modifications: skip connections, foveal regions, and an integral loss. Together these modifications allow information to flow along multiple paths through the network, enabling the classifier to operate at multiple scales, utilize context effectively, and perform more precise object localization. Our MultiPath network, coupled with DeepMask object proposals [4, 5], achieves major gains on COCO detection.

The recent COCO dataset presents several new challenges for object detection. In particular, it contains objects at a broad range of scales, less prototypical images, and requires more precise localization. To address these challenges, we test three modifications to the standard Fast R-CNN object detector: (1) skip connections that give the detector access to features at multiple network layers, (2) a foveal structure to exploit object context at multiple object resolutions, and (3) an integral loss function and corresponding network adjustment that improve localization.

The result of these modifications is that information can flow along multiple paths in our network, including through features from multiple network layers and from multiple object views. We refer to our modified classifier as a ‘MultiPath’ network. We couple our MultiPath network with DeepMask object proposals, which are well suited for localization and small objects, and adapt our pipeline to predict segmentation masks in addition to bounding boxes. The combined system improves results over the baseline Fast R-CNN detector with Sel-Search by 66% overall and by 4× on small objects.



Figure 2: Selected detection results on COCO. Only high-scoring detections are shown. While there are missed objects and false positives, many of the detections are quite good.

	AP	AP ⁵⁰	AP ⁷⁵	AP ^S	AP ^M	AP ^L
ResNet [3]	27.9	51.2	27.6	8.6	30.2	45.3
MultiPath	25.0	45.4	24.5	7.2	28.8	39.0
ResNet [3]	37.1	58.8	39.8	17.3	41.5	52.5
MultiPath	33.2	51.9	36.3	13.6	37.2	47.8
ION [1]	30.7	52.9	31.7	11.8	32.8	44.8
Fast R-CNN* [2]	19.3	39.3	19.9	3.5	18.8	34.6
Faster R-CNN* [6]	21.9	42.7	—	—	—	—

Table 1: **Top:** COCO test-standard segmentation results. **Bottom:** COCO test-standard bounding box results (top methods only). Leaderboard snapshot from 01/01/2016. *Note: Fast R-CNN and Faster R-CNN results are on test-dev as reported in [6], but results between splits tend to be quite similar.

Our system placed second in both the COCO 2015 detection and segmentation challenges, without using ResNets. Source code is available.

- [1] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural nets. In *CVPR*, 2016.
- [2] R. Girshick. Fast R-CNN. In *ICCV*, 2015.
- [3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [4] P. O. Pinheiro, R. Collobert, and P. Dollár. Learning to segment object candidates. In *NIPS*, 2015.
- [5] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár. Learning to refine object segments. In *ECCV*, 2016.
- [6] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.