Supplementary Material: FastMask: Segment Multi-scale Object Candidates in One Shot

Hexiang Hu* †
USC
hexiangh@usc.edu

Shiyi Lan* †
Fudan University
sylan14@fudan.edu.cn

Yuning Jiang
Megvii Inc.
jyn@megvii.com

Zhimin Cao
Megvii Inc.
czm@megvii.com

Fei Sha
USC
feisha@usc.edu

S1. Two-Stream FastMask

As mentioned in Section 5.1 of the main paper, we give all the details about our two-stream network here (See Figure S1). We craft the body network into two branches through manipulating the middle structure with different strides (*e.g.* 2 and 3 in our implementation). As a consequence, such manipulation produces differently scaled features as input to the neck module. It augments the body network to produce feature of diverse sizes, not necessarily limited to a multiple of two.

In our practice, we branch a 2×2 pooling on the feature downscaled by 8 to generate feature downscaled by factors of 16 and 24, and input these feature to the shared top convolutions. Then we apply our neck and head modules on these two streams to produce object segments in different scales. This technique adds more scales of feature, helps FastMask to be more robust to scale difference, but introduce limited extra computation. Note that we do not add any new parameters for learning through this branching technique.

S2. Visualizations

As promised in the main paper, we select more sample images from MS COCO[1], and visualize the results in Figures S2. As a comparison with existing methods, we have also provided the DeepMask$*$ [2] and SharpMask [2] results on MS COCO.

References

[1] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 1

[2] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár. Learning to refine object segments. In *ECCV*, 2016. 1, 2

*Equal contribution.

†Work was done during their internships at Megvii Inc.

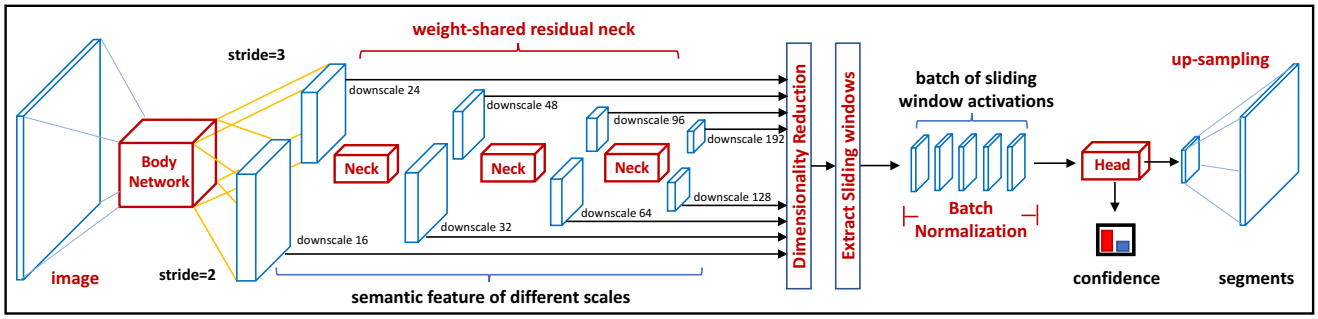

Figure S1. An overview of the proposed two-stream FastMask architecture.

Figure S2. Visualization of the object candidate segmentation results on sample MS COCO images. We compare our FastMask with DeepMask∗ [2] and SharpMask [2]. We also show the origin images and the ground-truth annotations for reference.