Attention to Scale: Scale-aware Semantic Segmentation – supplementary material –

Liang-Chieh Chen
lcchen@cs.ucla.edu

Yi Yang, Jiang Wang, Wei Xu
{yangyi05, wangjiang03, wei.xu}@baidu.com

Alan L. Yuille
yuille@stat.ucla.edu
alan.yuille@jhu.edu

Abstract

The supplementary material contains: (1) more qualitative results on some videos from MPII Human Pose dataset. (2) few more experimental result on subset of MS-COCO 2014 dataset, (3) more qualitative results on PASCAL-Person-Part, PASCAL VOC 2012, and subset of MS-COCO 2014 datasets.

1. Test on unseen dataset

We apply our trained model (*i.e.*, has been trained with PASCAL-Part dataset) to some videos from MPII Human Pose dataset [1]. The model is not fine-tuned on the dataset and the result is run frame-by-frame. In the videos, for each frame we show in the first row: (1) input image frames, (2) part segmentation results and (3) results overlapped with images. In the second row, we show (4) attention for scale = 1, (5) attention for scale = 0.75, and (6) attention for scale = 0.5. As shown in the video, even for images not seen during training (*i.e.*, cross domain), our model is able to produce reasonably and visually good part segmentation results and it also infers meaningful attention for different scales. The video is available at `http://liangchiehchen.com/video/mpii.avi`.

2. Results on subset of MS-COCO

Due to the limited space in the main paper, we report in this supplementary material the qualitative results of our proposed methods on a subset of MS-COCO 2014 dataset [4] in Fig. 1.

3. More qualitative results

We show more qualitative results on PASCAL-Person-Part [2] in Fig. 2, on PASCAL VOC 2012 [3] in Fig. 3, and on subset of MS-COCO 2014 [4] in Fig. 4, respectively.

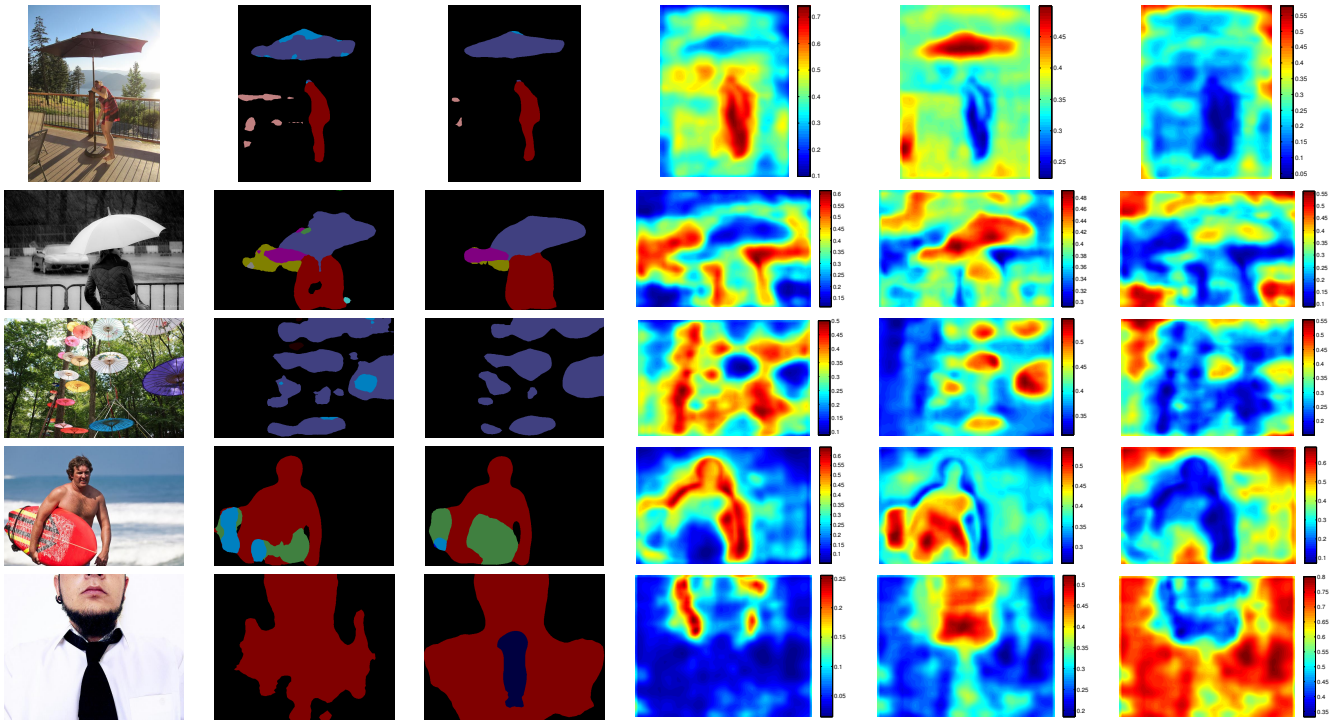
References

[1] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *CVPR*, 2014. 1

[2] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille. Detect what you can: Detecting and representing objects using holistic models and body parts. In *CVPR*, 2014. 1

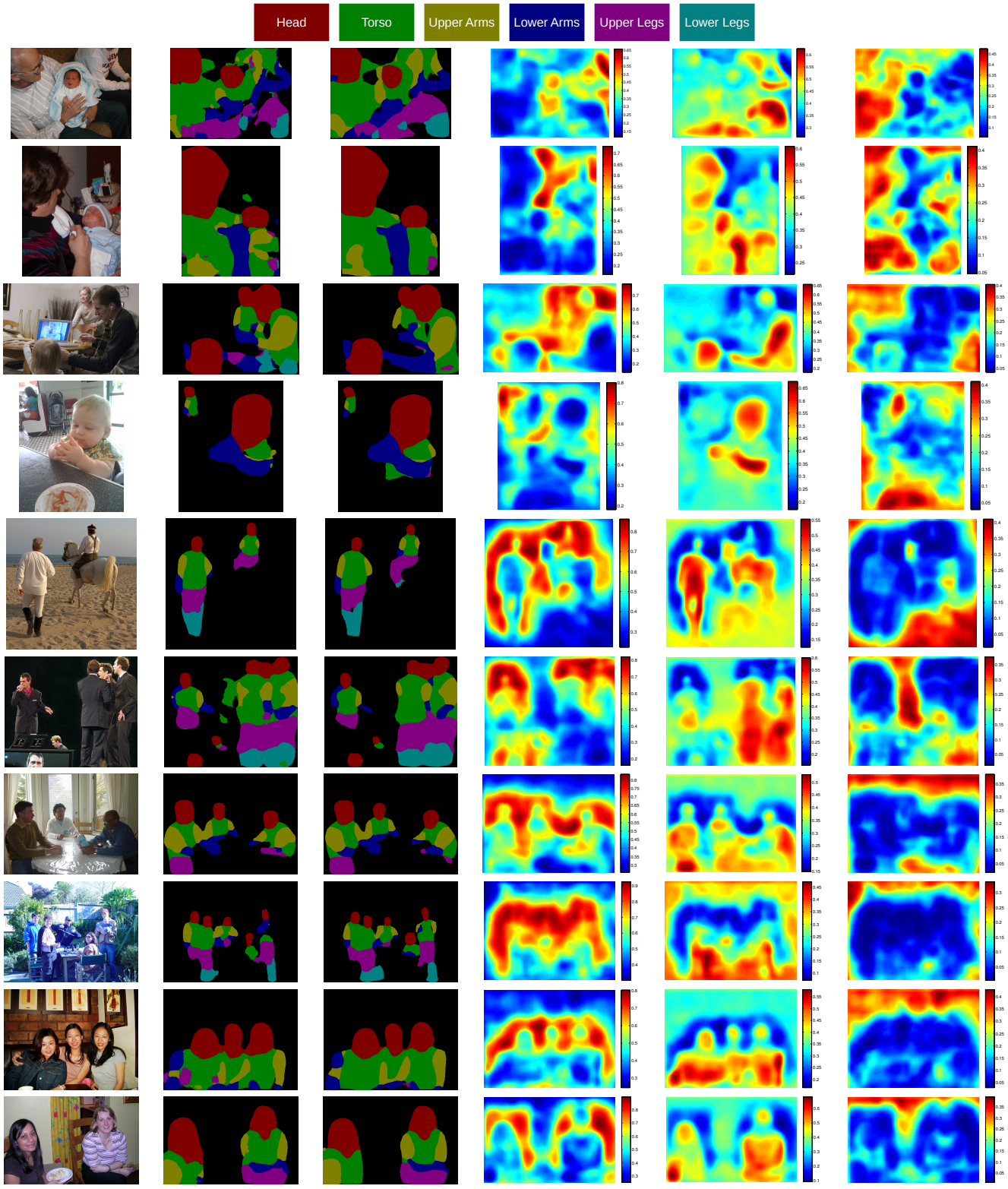
[3] M. Everingham, S. A. Eslami, L. V. Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *IJCV*, 111(1):98–136, 2014. 1

[4] 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



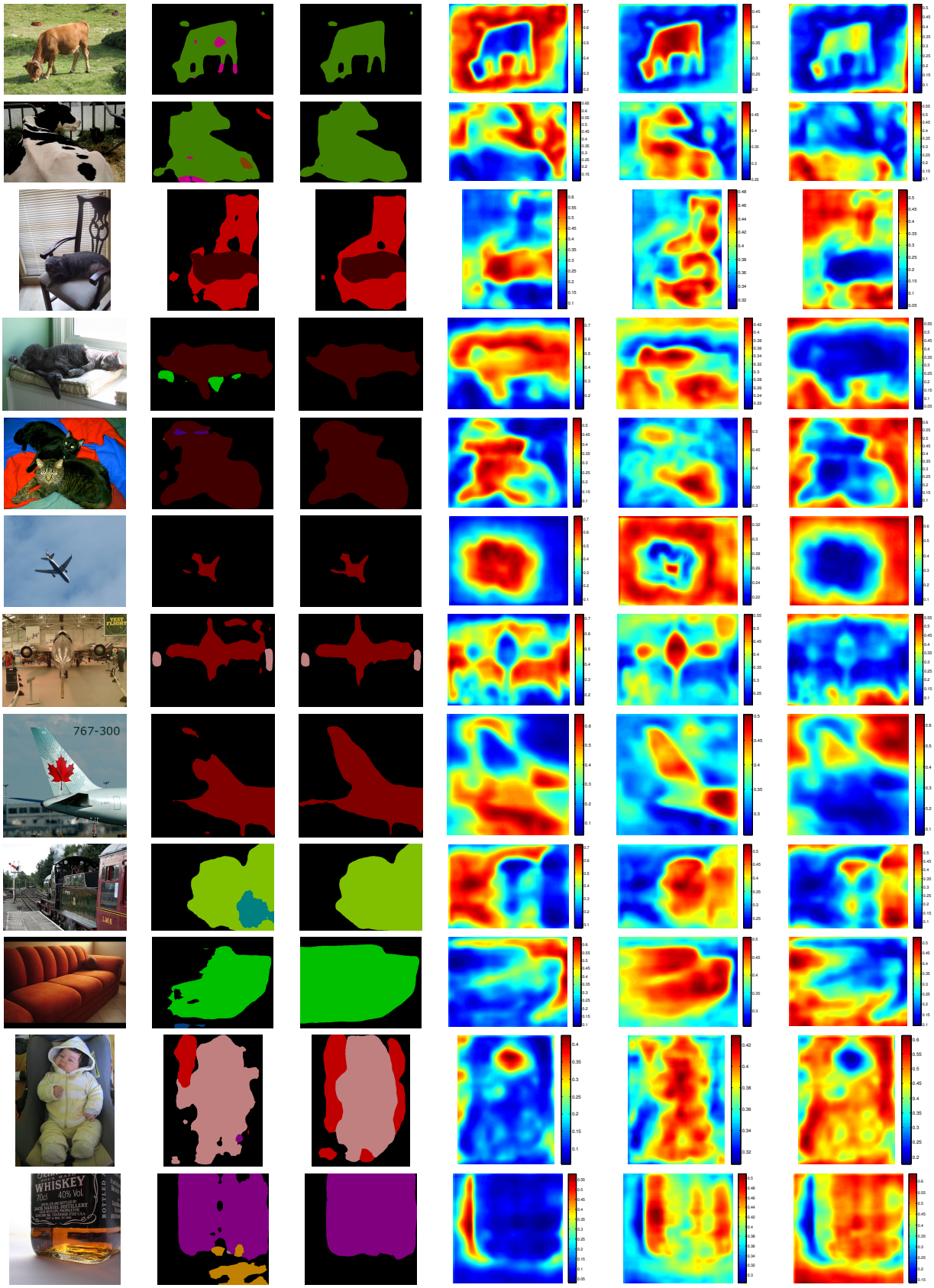
(a) Image (b) Baseline (c) Our model (d) Scale-1 Attention (e) Scale-0.75 Attention (f) Scale-0.5 Attention

Figure 1. Results on subset of MS-COCO 2014 *validation* set. DeepLab-LargeFOV with one scale input is used as baseline. Our model employs three scale inputs, attention model and extra supervision. Scale-1 attention captures small-scale person (dark red label) and umbrella (violet label). Scale-0.75 attention concentrates on middle-scale umbrella and head, while scale-0.5 attention catches large-scale person torso.



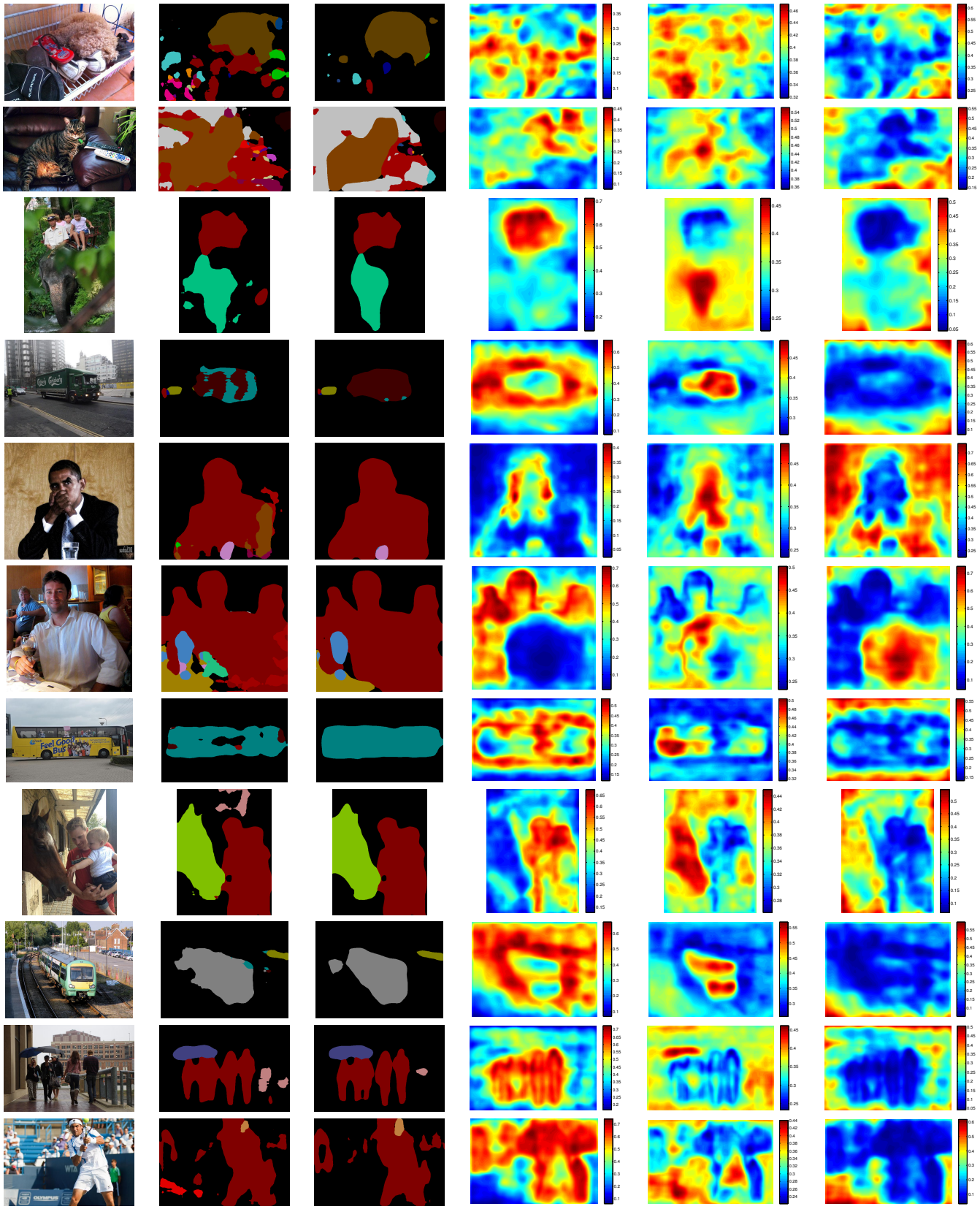
(a) Image (b) Baseline (c) Our model (d) Scale-1 Attention (e) Scale-0.75 Attention (f) Scale-0.5 Attention

Figure 2. Qualitative segmentation results on PASCAL-Person-Part validation set.



(a) Image (b) Baseline (c) Our model (d) Scale-1 Attention (e) Scale-0.75 Attention (f) Scale-0.5 Attention

Figure 3. Qualitative segmentation results on PASCAL VOC 2012 validation set.



(a) Image (b) Baseline (c) Our model (d) Scale-1 Attention (e) Scale-0.75 Attention (f) Scale-0.5 Attention

Figure 4. Qualitative segmentation results on subset of MS-COCO 2014 validation set.