

Mining Cross-Image Semantics for Weakly Supervised Semantic Segmentation

Supplementary Material

Guolei Sun¹, ✉Wenguan Wang¹, Jifeng Dai^{2,3}, and Luc Van Gool¹

¹ ETH Zurich ² SenseTime Research

³ Qing Yuan Research Institute, Shanghai Jiao Tong University

https://github.com/GuoleiSun/MCIS_wsss

Overview

- **Sec. 1. Segmentation Results:** comparisons among qualitative segmentation results produced by different methods
- **Sec. 2. Implementation Details of Segmentation Network:** hyper parameters used for training segmentation network

1 Segmentation Results

More visual comparisons of segmentation results produced by different methods are shown in Fig. 1. We can find that our method significantly outperforms two top-performing methods, *i.e.*, PSA [1] and OAA+ [3].

2 Implementation Details of the Segmentation Network

After producing the localization maps, the pseudo ground-truth masks are generated to train the fully segmentation network (DeepLab-LargeFOV [2]). Following [3], we use a mini-batch of 10 images and terminate the training at 15,000 iterations under most standard setting. For other two settings where extra single-label images or web images are used, we terminate the training at 20,000, and 25,000 iterations respectively since more images are used. All other hyper-parameters are the same as [2]. Specifically, the initial learning rate is 0.001 and "poly" learning rate policy is adopted, *i.e.*, the learning rate is multiplied by $(1 - \frac{iter}{max.iter})^{power}$, where the *power* is set to be 0.9. We also use the momentum of 0.9 and weight decay of 0.0005. For more details, please refer to [2].

✉ Corresponding author: *Wenguan Wang* (wenguanwang.ai@gmail.com).

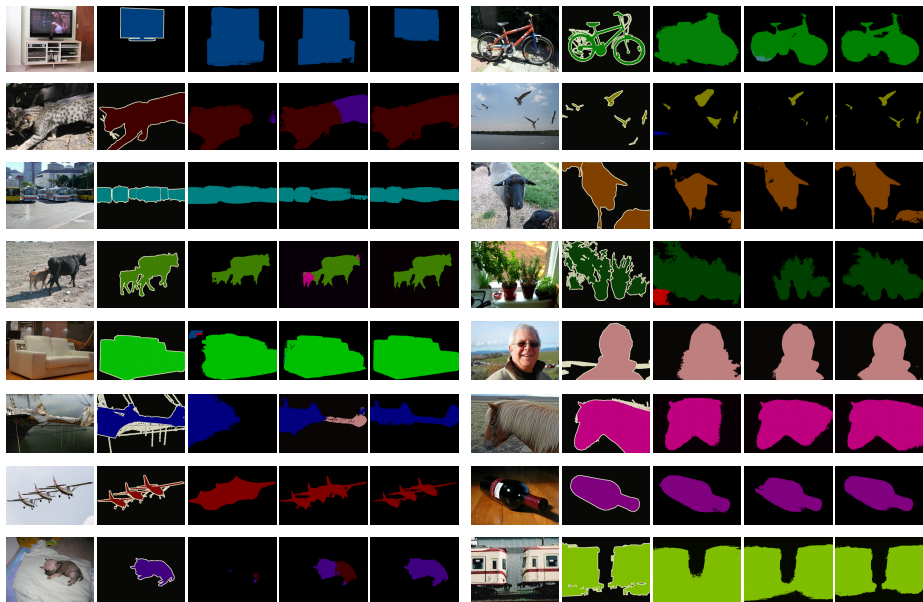


Fig. 1: Visual comparisons of segmentation results on PASCAL VOC12 val set. From *left to right*: input image, ground truth, results for PSA [1], OAA+ [3], and our method.

References

1. Ahn, J., Kwak, S.: Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: CVPR (2018) 1, 2
2. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. TPAMI 40(4), 834–848 (2017) 1
3. Jiang, P.T., Hou, Q., Cao, Y., Cheng, M.M., Wei, Y., Xiong, H.K.: Integral object mining via online attention accumulation. In: ICCV (2019) 1, 2