

Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform – Supplementary Material –

Liang-Chieh Chen
lcchen@cs.ucla.edu

Jonathan T. Barron, George Papandreou, Kevin Murphy
{barron, gpapan, kpmurphy}@google.com

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

Abstract

The supplementary material contains: (1) Detailed quantitative results for the proposed methods, showing per-class semantic segmentation IOU on the PASCAL VOC 2012 test set. (2) Qualitative edge detection and semantic segmentation results on additional images.

1. Detailed quantitative image segmentation results

We provide per-class semantic segmentation IOU on the PASCAL VOC 2012 test set. We compare with the DeepLab-LargeFOV and DeepLab-CRF-LargeFOV baselines. In Tab. 1 we show performance of models that have only been pretrained on the Imagenet 2012 image classification task [5], while in Tab. 2 we show performance of models that have also been pretrained on the MS-COCO 2014 semantic segmentation task [3].

Method	mean	bkg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
DeepLab-LargeFOV [1]	65.1	90.7	74.7	34.0	74.3	57.1	62.0	82.6	75.5	79.1	26.2	65.7	55.8	73.0	68.0	78.6	76.2	50.6	73.9	45.5	66.6	57.1
DeepLab-CRF-LargeFOV [1]	70.3	92.6	83.5	36.6	82.5	62.3	66.5	85.4	78.5	83.7	30.4	72.9	60.4	78.5	75.5	82.1	79.7	58.2	82.0	48.8	73.7	63.3
DT-SE	67.8	91.7	78.8	33.5	78.7	60.6	64.5	84.5	77.4	81.3	29.0	69.1	59.4	76.1	70.8	80.6	77.9	53.4	77.9	46.0	70.1	62.5
DT-EdgeNet	69.0	92.1	79.8	34.8	79.6	61.3	67.0	85.0	78.5	83.2	30.2	70.3	58.9	77.9	72.3	82.3	79.5	55.0	79.8	47.9	70.8	62.5
DT-EdgeNet + DenseCRF	71.2	92.8	83.6	35.8	82.4	63.1	68.9	86.2	79.6	84.7	31.8	74.2	61.1	79.6	76.6	83.2	80.9	58.3	82.6	49.1	74.8	65.1

Table 1. Segmentation IOU on the PASCAL VOC 2012 test set, using the trainval set for training. Model only pretrained on the Imagenet image classification task.

Method	mean	bkg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
DeepLab-COCO-LargeFOV [4]	68.9	92.1	81.6	35.4	81.4	60.1	65.9	84.3	79.3	81.8	28.4	71.2	59.0	75.3	72.6	81.5	80.1	53.5	78.8	50.8	72.7	60.3
DeepLab-CRF-COCO-LargeFOV [4]	72.7	93.4	89.1	38.3	88.1	63.3	69.7	87.1	83.1	85.0	29.3	76.5	56.5	79.8	77.9	85.8	82.4	57.4	84.3	54.9	80.5	64.1
DT-SE	70.7	92.6	83.8	35.0	85.5	61.9	67.6	85.4	80.3	84.4	30.2	73.6	60.4	77.8	74.8	82.3	81.0	54.9	81.2	52.3	75.5	64.1
DT-EdgeNet	71.7	93.0	85.6	36.0	86.4	63.0	69.3	86.0	81.2	85.9	30.7	75.1	60.8	79.3	76.1	83.2	82.0	56.2	82.8	53.3	75.9	64.4
DT-EdgeNet + DenseCRF	73.6	93.5	88.3	37.0	89.8	63.6	70.3	87.3	82.0	87.6	31.1	79.0	61.9	81.6	80.4	84.5	83.3	58.4	86.1	55.9	78.2	65.4
DeepLab-CRF-Attention-DT	76.3	94.3	93.2	41.7	88.0	61.7	74.9	92.9	84.5	90.4	33.0	82.8	63.2	84.5	85.0	87.2	85.7	60.5	87.7	57.8	84.3	68.2

Table 2. Segmentation IOU on the PASCAL VOC 2012 test set, using the trainval set for training. Model pretrained on both the Imagenet image classification task and the MS-COCO semantic segmentation task.

2. Qualitative edge detection and image segmentation results

We show additional edge detection and semantic segmentation results on PASCAL VOC 2012 val set in Figs. 1 and 2. We compare results obtained with the proposed domain transform when using our learned EdgeNet edges vs. the SE edges of [2].



(a) Image (b) Baseline (c) SE (d) DT-SE (e) EdgeNet (f) DT-EdgeNet

Figure 1. Visualizing results on VOC 2012 val set. For each row, we show (a) Image, (b) Baseline (DeepLab-LargeFOV) segmentation result, (c) edges produced by Structured Edges, (d) segmentation result with Structured Edges, (e) edges generated by EdgeNet, and (f) segmentation result with EdgeNet. Similar to Fig. (9) of main paper.



(a) Image

(b) Baseline

(c) SE

(d) DT-SE

(e) EdgeNet

(f) DT-EdgeNet

Figure 2. Visualizing results on VOC 2012 val set. Continued from Fig. 1.

References

- [1] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *ICLR*, 2015.
- [2] P. Dollár and C. L. Zitnick. Structured forests for fast edge detection. In *ICCV*, 2013.
- [3] T.-Y. Lin et al. Microsoft COCO: Common objects in context. In *ECCV*, 2014.
- [4] G. Papandreou, L.-C. Chen, K. Murphy, and A. L. Yuille. Weakly- and semi-supervised learning of a dcnn for semantic image segmentation. In *ICCV*, 2015.
- [5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 2015.