PSANet: Point-wise Spatial Attention Network for Scene Parsing – Supplementary Material

Hengshuang Zhao[1⋆], Yi Zhang[2⋆], Shu Liu[1], Jianping Shi[3],
Chen Change Loy[4], Dahua Lin[2], and Jiaya Jia[1,5]

[1]The Chinese University of Hong Kong
[2]CUHK-Sensetime Joint Lab, The Chinese University of Hong Kong
[3]SenseTime Research [4]Nanyang Technological University [5]Tencent Youtu Lab
{hszhao,sliu,leojia}@cse.cuhk.edu.hk, {zy217,dhlin}@ie.cuhk.edu.hk,
shijianping@sensetime.com, ccloy@ntu.edu.sg

More Numerical Results In Table 1, we give a comparison with current top ranked approaches on ADE20K test set. Our single PSANet101 gets final score as 55.46%, surpasses PSPNet269, the winner in challenge 2016 with a deeper backbone, and matches CASIA_IVA_JD, the winner entry in 2017 with multiple models ensembling. In Table 2, we compare PSANet with several state-of-the-art methods on VOC 2012 validation set. Table 3 and Table 4 list the detailed per-class results.

More Visual Results Fig. 1 shows the visual improvements on validation set of VOC 2012. Fig. 2 and Fig. 3 contain visual comparisons between different methods. Fig. 4 includes several visual predictions on Cityscapes dataset.

Table 1. Methods comparison on ADE20K *test* set. Score is the average of mean IoU and pixel accuracy.

Method	score
CASIA_IVA	54.33
360+MCG-ICT-CAS_SP	54.68
Adelaide/WiderNet[32]	56.41
SenseCUSceneParsing/PSPNet269[6]	55.38
WinterIsComing	55.44
CASIA_IVA_JD	55.47
PSANet101	55.46

Table 2. Methods comparison with models trained on *train_aug* set and evaluated on *val* set of VOC 2012.

Method	mIoU(%)
DeepLabv2[5]	77.69
PSPNet[6]	78.83
DeepLabv3[37]	79.77
PSANet	79.77

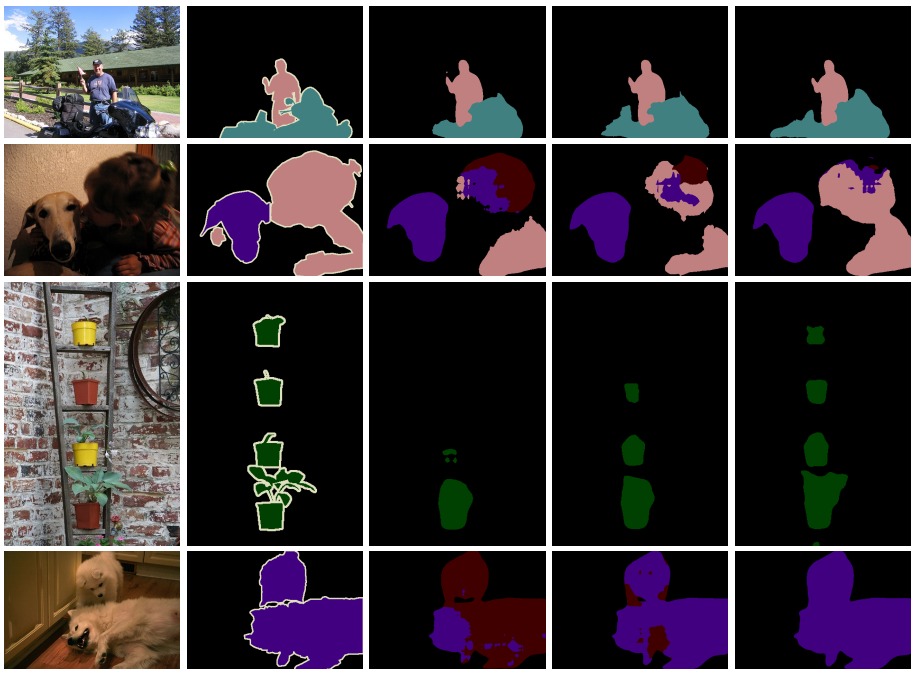
⋆ indicates equal contribution.

Table 3. Detailed per-class results on VOC 2012 *test* set.

Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mIoU
LRR [4]	92.4	45.1	94.6	65.2	75.8	95.1	89.1	92.3	39.0	85.7	70.4	88.6	89.4	88.6	86.6	65.8	86.2	57.4	85.7	77.3	79.3
DeepLabv2 [1]	92.6	60.4	91.6	63.4	76.3	95.0	88.4	92.6	32.7	88.5	67.6	89.6	92.1	87.0	87.4	63.3	88.3	60.0	86.8	74.5	79.7
SegModel [10]	93.6	60.2	93.6	69.1	76.4	96.3	88.2	95.5	37.9	90.8	73.3	91.1	94.3	88.6	88.6	64.8	90.1	63.7	87.3	78.2	81.8
LC [5]	85.5	66.7	94.5	67.2	84.0	96.1	89.8	93.5	47.2	90.4	71.5	88.9	91.7	89.2	89.1	70.4	89.4	70.7	84.2	79.6	82.7
DUC_HDC [11]	92.1	64.6	94.7	71.0	81.0	94.6	89.7	94.9	45.6	93.7	74.4	92.0	95.1	90.0	88.7	69.1	90.4	62.7	86.4	78.2	83.1
LKM [8]	95.3	68.7	94.1	72.6	82.4	96.0	89.3	93.0	47.8	89.6	70.8	89.2	93.3	90.1	91.2	72.0	89.8	67.8	88.9	76.9	83.6
RefineNet [6]	95.0	73.2	93.5	78.1	84.8	95.6	89.8	94.1	43.7	92.0	77.2	90.8	93.4	88.6	88.1	70.1	92.9	64.3	87.7	78.8	84.2
ResNet-38 [12]	96.2	75.2	95.4	74.4	81.7	93.7	89.9	92.5	48.2	92.0	79.9	90.1	95.5	91.8	91.2	73.0	90.5	65.4	88.7	80.6	84.9
PSPNet [13]	95.8	72.7	95.0	78.9	84.4	94.7	92.0	95.7	43.1	91.0	80.3	91.3	96.3	92.3	90.1	71.5	94.4	66.9	88.8	82.0	85.4
DeepLabv3 [2]	96.4	76.6	92.7	77.8	87.6	96.7	90.2	95.4	47.5	93.4	76.3	91.4	97.2	91.0	92.1	71.3	90.9	68.9	90.8	79.3	85.7
PSANet	95.8	76.1	94.1	78.2	84.6	96.5	92.2	94.6	44.7	92.4	78.8	90.2	97.1	93.2	91.5	70.5	94.8	66.0	88.7	83.5	85.7

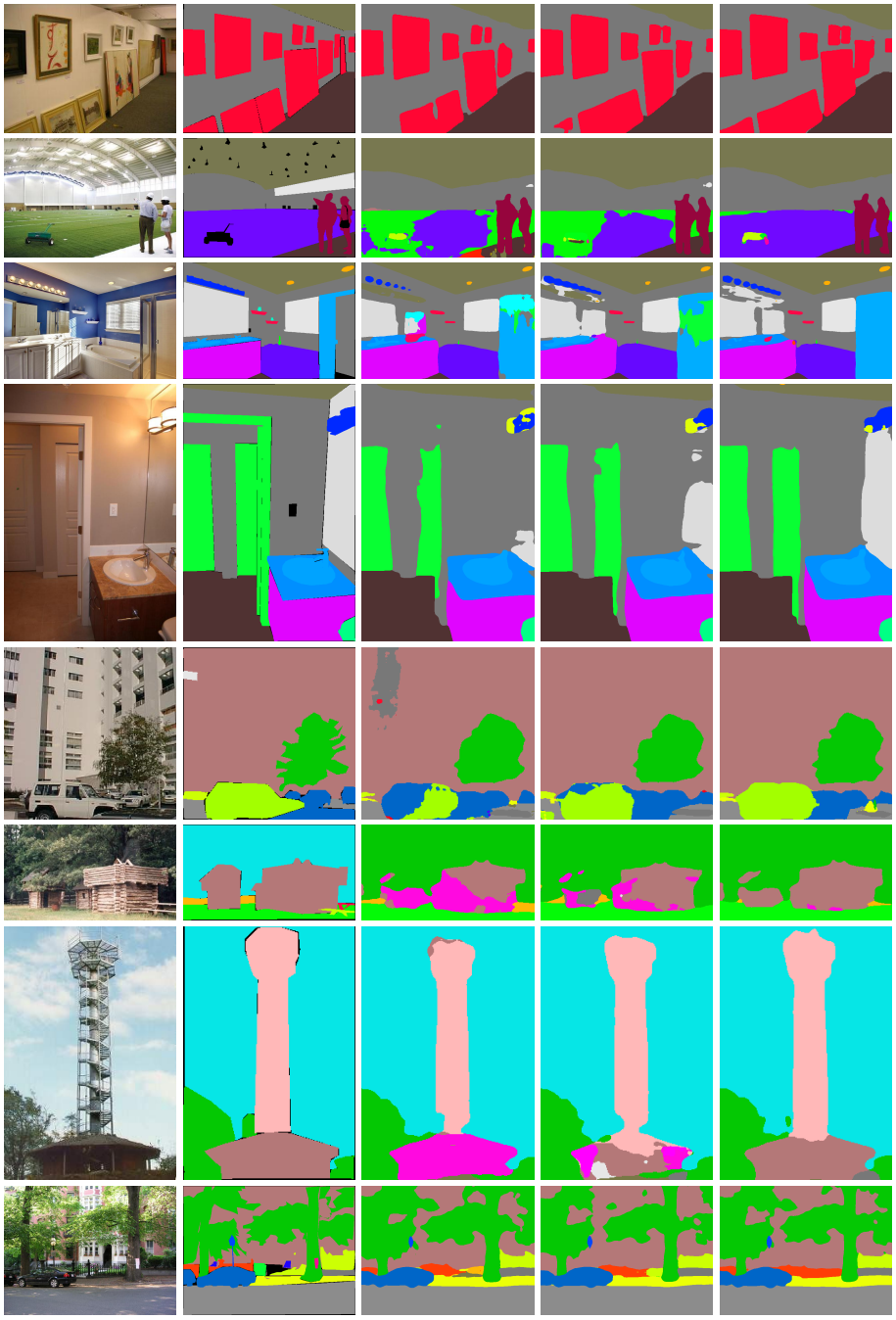
Table 4. Detailed per-class results on Cityscapes *test* set. Methods are trained using both *fine* and *coarse* data.

Method	road	swalk	build.	wall	fence	pole	tlight	sign	veg.	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
LRR-4x [4]	97.9	81.5	91.4	50.5	52.7	59.4	66.8	72.7	92.5	70.1	95.0	81.3	60.1	94.3	51.2	67.7	54.6	55.6	69.6	71.8
SegModel [9]	98.6	86.2	93.0	53.7	60.4	64.2	73.5	78.5	93.4	72.2	95.5	85.3	68.6	95.8	77.9	87.0	78.0	68.0	75.1	79.2
DUC_HDC [11]	98.5	85.9	93.2	57.7	61.1	67.2	73.7	78.0	93.4	72.3	95.4	85.9	70.5	95.9	76.1	90.6	83.7	67.4	75.7	80.1
Netwarp [3]	98.6	86.7	93.4	60.6	62.6	68.6	75.9	80.0	93.5	72.0	95.3	86.5	72.1	95.9	72.9	89.9	77.4	70.5	76.4	80.5
ResNet-38 [12]	98.7	86.9	93.3	60.4	62.9	67.6	75.0	78.7	93.7	73.7	95.5	86.8	71.1	96.1	75.2	87.6	81.9	69.8	76.7	80.6
PSPNet [13]	98.7	86.9	93.5	58.4	63.7	67.7	76.1	80.5	93.6	72.2	95.3	86.8	71.9	96.2	77.7	91.5	83.6	70.8	77.5	81.2
DeepLabv3 [2]	98.6	86.2	93.5	55.2	63.2	70.0	77.1	81.3	93.8	72.3	95.9	87.6	73.4	96.3	75.1	90.4	85.1	72.1	78.3	81.3
PSANet	98.7	87.0	93.5	58.9	62.5	67.8	76.0	80.0	93.7	72.6	95.4	87.0	73.0	96.2	79.3	91.2	84.9	71.1	77.9	81.4



(a) Image (b) Ground Truth (c) Baseline (d) PSA-COL (e) PSA-COL-DIS

Fig. 1. Visual improvements on *val* set of VOC 2012.



(a) Image (b) Ground Truth (c) DeepLab (d) PSPNet (e) PSANet

Fig. 2. Visual comparison on ADE20K dataset. (a) Image. (b) Ground Truth. (c) DeepLab [1]. (d) PSPNet [13]. (e) PSANet.

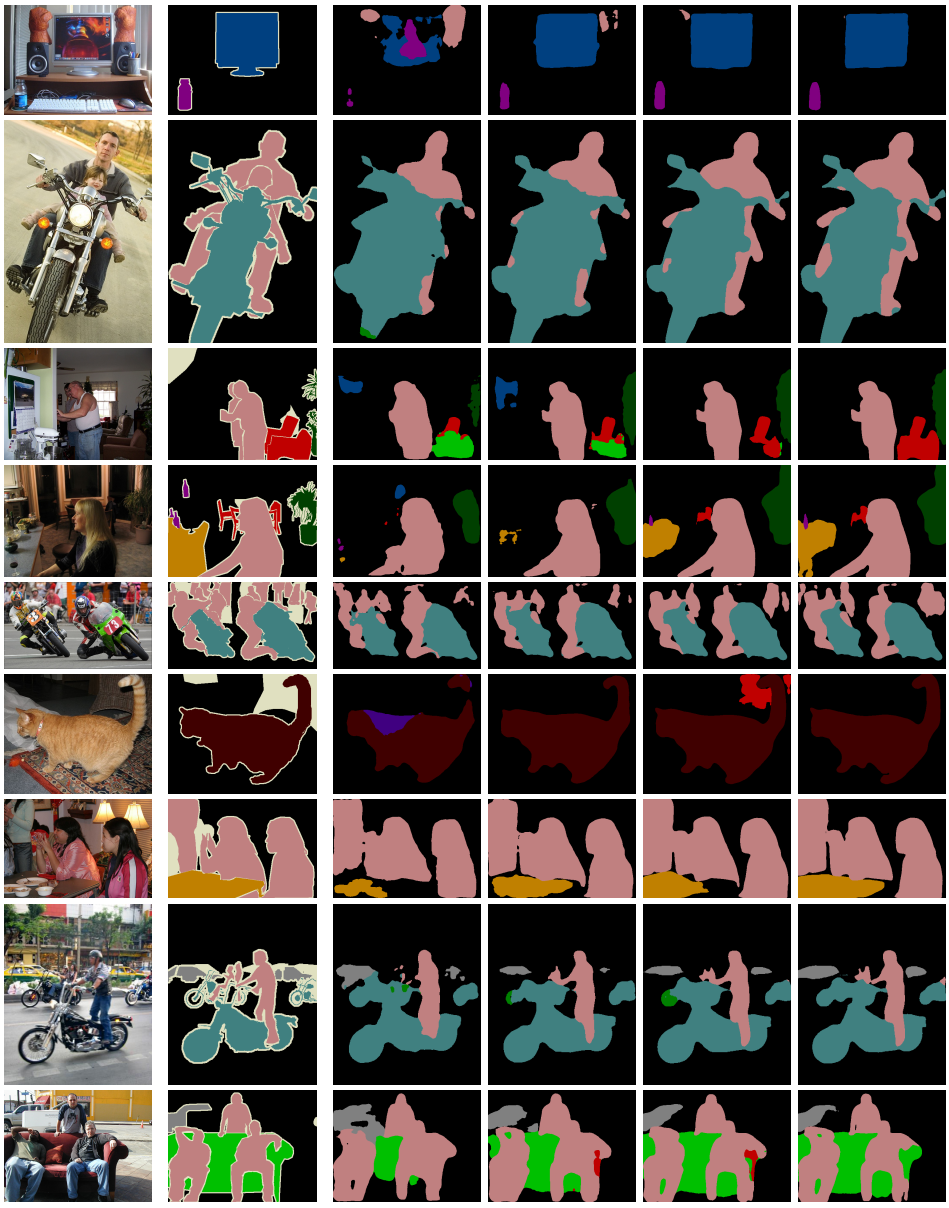
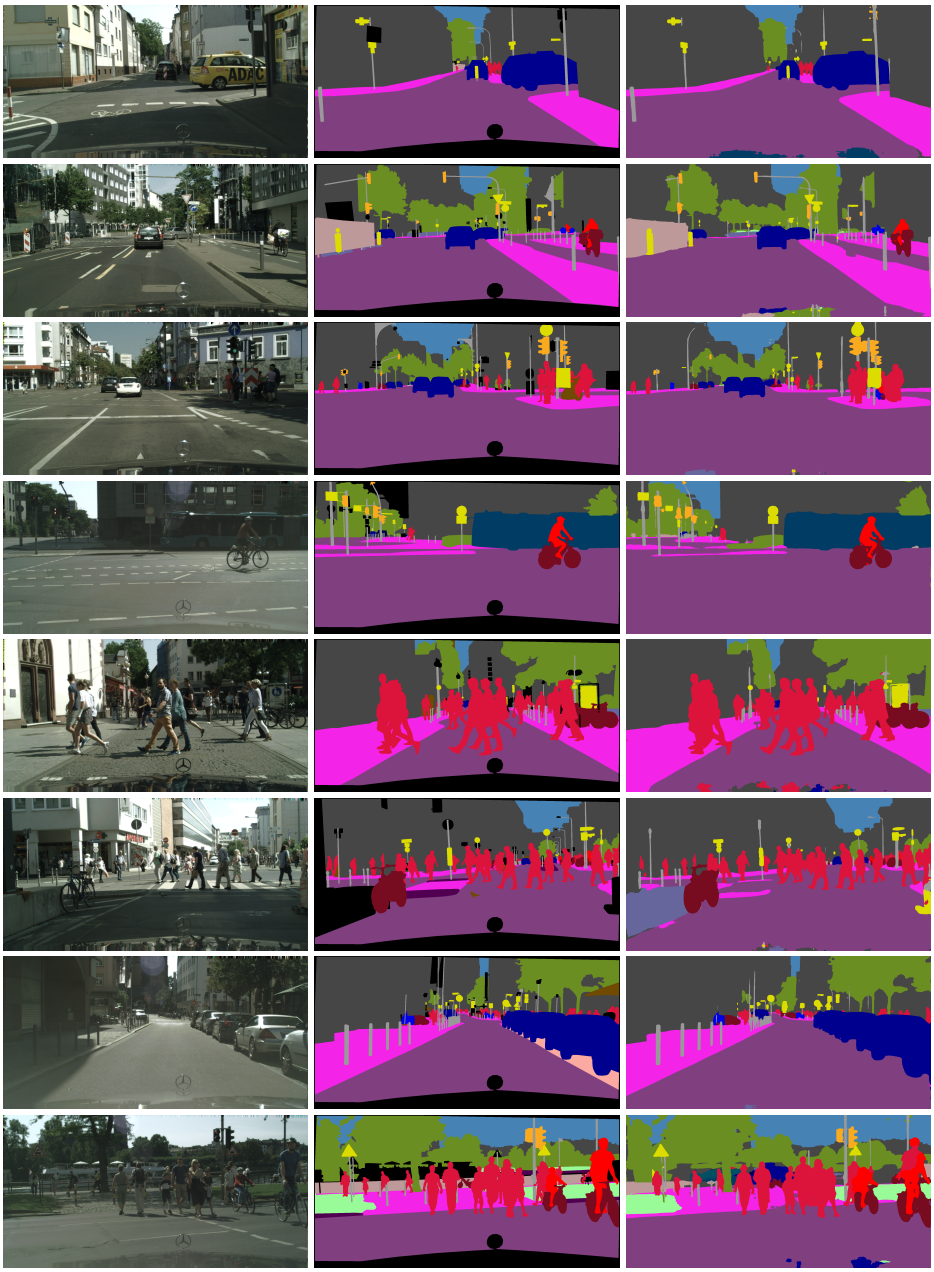


Fig. 3. Visual comparison on PSACAL VOC 2012 dataset. (a) Image. (b) Ground Truth. (c) FCN [7]. (d) DeepLab [1]. (e) PSPNet [13]. (f) PSANet.



(a) Image (b) Ground Truth (c) PSANet

Fig. 4. Visual prediction on Cityscapes dataset. (a) Image. (b) Ground Truth. (c) PSANet.

References

1. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. TPAMI (2018)
2. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587 (2017)
3. Gadde, R., Jampani, V., Gehler, P.V.: Semantic video cnns through representation warping. In: ICCV (2017)
4. Ghiasi, G., Fowlkes, C.C.: Laplacian pyramid reconstruction and refinement for semantic segmentation. In: ECCV (2016)
5. Li, X., Liu, Z., Luo, P., Loy, C.C., Tang, X.: Not all pixels are equal: difficulty-aware semantic segmentation via deep layer cascade. In: CVPR (2017)
6. Lin, G., Milan, A., Shen, C., Reid, I.D.: Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In: CVPR (2017)
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR (2015)
8. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters–improve semantic segmentation by global convolutional network. In: CVPR (2017)
9. Shen, F., Gan, R., Yan, S., Zeng, G.: Semantic segmentation via structured patch prediction, context crf and guidance crf. In: CVPR (2017)
10. Shen, F., Zeng, G.: Fast semantic image segmentation with high order context and guided filtering. arXiv:1605.04068 (2016)
11. Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G.W.: Understanding convolution for semantic segmentation. In: WACV (2018)
12. Wu, Z., Shen, C., van den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. arXiv:1611.10080 (2016)
13. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR (2017)