Tensor Low-Rank Reconstruction for Semantic Segmentation

Wanli Chen[1], Xinge Zhu[1], Ruoqi Sun[2], Junjun He[2,3], Ruiyu Li[4], Xiaoyong Shen[4], and Bei Yu[1]

[1] The Chinese University of Hong Kong
{wlchen,byu}@cse.cuhk.edu.hk,zx018@ie.cuhk.edu.hk
[2] Shanghai Jiao Tong University
ruoqisun7@sjtu.edu.cn
[3] ShenZhen Key Lab of Computer Vision and Pattern Recognition, SIAT-SenseTime Joint Lab, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
hejunjun@sjtu.edu.cn
[4] SmartMore
{ryli,xiaoyong}@smartmore.com

1 Appendix

1.1 More Experimental Results

We conduct experiments on Cityscapes dataset [1], which is a famous scene segementation dataset that includes 19 semantic classes. It provides 2975/500/1525 images for training, validation and testing. Since the training setting of Cityscapes is very distinct to the implementation details that presented in main paper, we put the results in supplementary materials.

The input images are cropped into 512×1024 before input. The batch size we use is 8. Initially, the learning rate $lr = 0.01$. SGD optimizer with momentum $= 0.9$ and weight decay $= 0.0005$ is applied for training. The evaluation metrics and data augmentation strategies we use are the same as main paper. For the

Table 1: Results on Cityscapes *test* set

Method	Backbone	mIoU
PSANet [6]	ResNet-101	80.1
CFNet [5]	ResNet-101	79.6
AsymmetricNL [7]	ResNet-101	81.3
CCNet [3]	ResNet-101	81.4
DANet [2]	ResNet-101	81.5
ACFNet [4]	ResNet-101	81.8
RecoNet	ResNet-101	**82.3**

evaluation on $val/test$ set, we train 40K/100K iterations on $train/train + val$

Table 2: Ablation study on different components. The experiments are implemented using Cityscapes validation set

Method	TGM+TRM	GPM	Aux-loss	MS/Flip	mIoU %
ResNet-50				√	73.1
ResNet-50	√			√	78.9
ResNet-50	√	√		√	79.4
ResNet-50	√	√	√	√	79.8
ResNet-101	√	√	√		80.5
ResNet-101	√	√	√	√	81.6

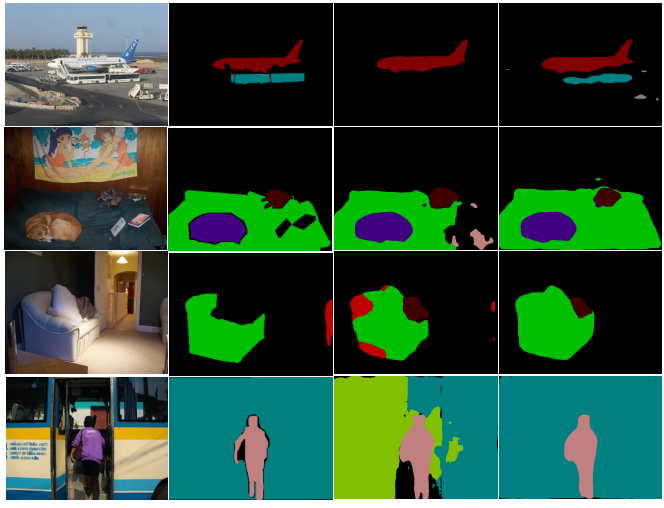
set respectively. The testing results are shown on Table 1, which collects current state-of-the-art attention based methods. RecoNet get better performance than these approaches. The online hard example mining (OHEM) strategy is not used in our implementation since it is time consuming. The result is avaliable on the website.[5]

In order to validate the consistency of RecoNet, we conduct additional ablation experiments on Cityscapes dataset. The tensor rank is set to $r = 64$ for ablation. In Table 2, it can be found that TGM+TRM contributes 5.8 % mIoU improvement (73.1% to 78.9%), which dominates the other modules. The experimental results show that RecoNet is consistent on different datasets.

1.2 More Visualization

Fig. 1 shows some results of RecoNet-101 on PACAL-VOC12 *validation* dataset. The figure shows that RecoNet has a better qualitative result, especially in the boundary, which also demonstrates its effectiveness of context modeling.

[5] https://www.cityscapes-dataset.com/anonymous-results/?id=7c7bfabc1026a9fd07b348bfd311c56a57ba0369969f3bd9fd9f036ce49a2934



(a) Image (b) GT (c) FCN (d) RecoNet

Fig. 1: Qualitative results on PASCAL-VOC12 *validation* dataset.

References

1. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3213–3223 (2016) 1
2. Fu, J., Liu, J., Tian, H., Fang, Z., Lu, H.: Dual attention network for scene segmentation. arXiv preprint arXiv:1809.02983 (2018) 1
3. Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W.: CCNet: Criss-cross attention for semantic segmentation. In: Proc. ICCV. pp. 603–612 (2019) 1
4. Zhang, F., Chen, Y., Li, Z., Hong, Z., Liu, J., Ma, F., Han, J., Ding, E.: Acfnet: Attentional class feature network for semantic segmentation. In: The IEEE International Conference on Computer Vision (ICCV) (October 2019) 1
5. Zhang, H., Zhang, H., Wang, C., Xie, J.: Co-occurrent features in semantic segmentation. In: Proc. CVPR. pp. 548–557 (2019) 1
6. Zhao, H., Zhang, Y., Liu, S., Shi, J., Change Loy, C., Lin, D., Jia, J.: PSANet: Point-wise spatial attention network for scene parsing. In: Proc. ECCV. pp. 267–283 (2018) 1
7. Zhu, Z., Xu, M., Bai, S., Huang, T., Bai, X.: Asymmetric non-local neural networks for semantic segmentation. In: Proc. ICCV. pp. 593–602 (2019) 1