Computational Visual Media
DOI 10.1007/s41095-xxx-xxxx-x Vol. x, No. x, month year, xx–xx

Article Type (Research/Review)

Erroneous Pixel Prediction for Semantic Image Segmentation: Supplemental Material

Lixue Gong[1], Yiqun Zhang[1], Yunke Zhang[1],Yin Yang[2] and Weiwei Xu[1](✉)

1 Introduction

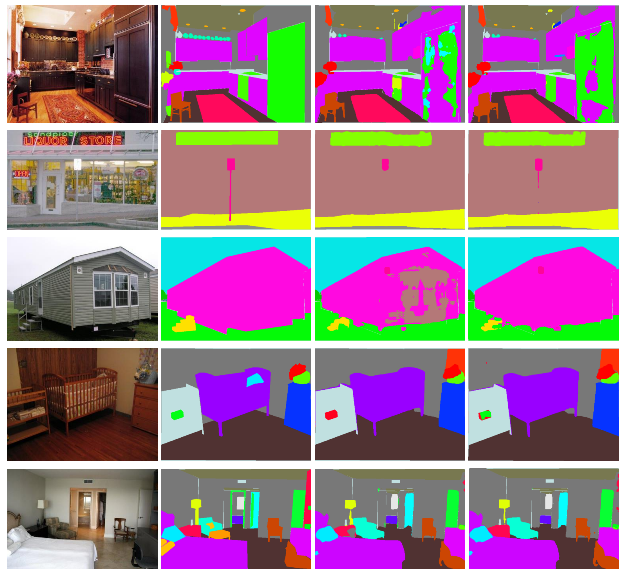
This supplemental material introduces the network architecture of each branch in our model in detail. And more visual results on ADE20K [7] and Cityscapes [4] of each branch are provided.

2 Network Architecture

The architecture of the *error-prediction branch* and the *detail branch* are shown in Tab. 1 and Tab. 2 respectively. The Global Attention Upsampling module (GAU) [5] and the Xception block [3] are used as building blocks in both branches.

3 More Visual Results

ADE20K validation set: In Fig. 1, we show more segmentation results of our model training on ADE20K training set with DeepLabv3+ as the *semantic branch*.
Cityscapes testing set: We illustrate additional visual results on Cityscapes testing dataset in Fig. 4. This model training on Cityscapes *trainval_fine* dataset takes Xception71-DPC [1] as the *semantic branch*.
Error-prediction: The estimated error probability maps $\mathbf{E}_{\mathrm{ep}}$ are illustrated in Fig. 2. Our method can detect most of ground truth erroneous pixels.
Layer Cascade, Hard-mining and Bagging: We have compared the quantitative results among Layer Cascade [6], Hard-mining and Bagging in the original paper. The visual results are show in Fig. 3. All these three strategies can improve the segmentation results while the improvement of our method is the most significant.



(a) Image (b) Ground-truth (c) DeepLabv3+ (d) Ours

Fig. 1 Visual improvements on ADE20K validation set.

1 the State Key Laboratory of CAD & CG at Zhejiang University, 310058, China. E-mail: L. Gong, gonglx@zju.edu.cn; W. Xu, xww@cad.zju.edu.cn(✉).

2 School of Computing Clemson University, South Carolina, 29634, U.S..

Manuscript received: 2014-12-31; accepted: 2015-01-30.

References

[1] L.-C. Chen, M. Collins, Y. Zhu, G. Papandreou, B. Zoph, F. Schroff, H. Adam, and J. Shlens. Searching for efficient multi-scale architectures for dense image prediction. In *NIPS*, pages 8713–8724, 2018.

[2] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, pages 801–818, 2018.

[3] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *CVPR*, pages 1251–1258, 2017.

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, pages 3213–3223, 2016.

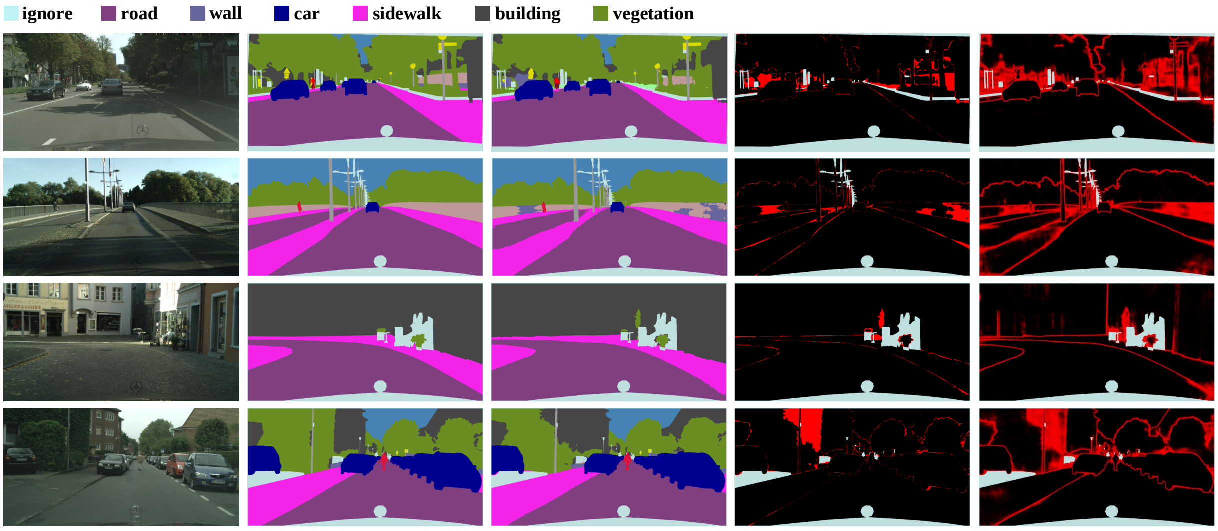
[5] H. Li, P. Xiong, J. An, and L. Wang. Pyramid attention

Tab. 1 Architecture of Error-prediction Branch

Block name	Layer type	Stride	Ch I/O	Input
sbp_feature	3x3conv+GN+ReLU	2	class_num/32	initial prediction $\mathbf{P}_{sb}$
rgb_feature	3x3conv+GN+ReLU	2	3/32	RGB Image
ep_input	concat	-	32+32/64	sbp_feature rgb_feature
conv1	3x3conv+GN+ReLU	1	64/64	ep_input
conv2	3x3conv+GN+ReLU	1	64/64	conv1
max_pool	2x2max_pool	2	64/64	conv2
gau1	GAU Block [5]	1	64,1280/128	max_pool aspp_concat_out
unit1	Xception Block [3]	1	128/128 128/128 128/128	GAU1
unit2	Xception Block+ReLU	1	128/256 256/256 256/256	unit1
error_probs	3x3Conv+sigmoid	1	256/1	unit2

Tab. 2 Architecture of Detail Branch

Block name	Layer type	Stride	Ch I/O	Input
gau3	GAU Block	1	728, 1280/512	Xception_entryflow_block5_unit1 aspp_concat_out
gau3_conv0	3x3seperable_conv2d	1	512/512	gau3
gau3_conv1	3x3seperable_conv2d	1	512/512	gau3_conv0
gau3_relu	Add+ReLU	-	512, 512/512	gau3_conv1 gau3
gau2	GAU Block	1	256, 512/256	Xception_entryflow_block3_unit1 gau3_relu
gau2_conv0	3x3seperable_conv2d	1	256/256	gau2
gau2_conv1	3x3seperable_conv2d	1	256/256	gau2_conv0
gau2_relu	Add+ReLU	-	256, 256/256	gau2_conv1 gau2
gau1	GAU Block	1	128, 256/128	Xception_entryflow_block1_unit1 gau2_relu
gau1$_conv0$	3x3seperable_conv2d	1	128/128	gau1
gau1$_conv1$	3x3seperable_conv2d	1	128/128	gau1_conv0
gau1_relu	Add+ReLU	-	128, 128/128	gau1_conv1 gau1
db_logits	3x3conv2d	1	19/19	gau1_relu



(a) Image (b) Ground-truth (c) Semantic branch prediction (d) Ground-truth error (e)Error prediction

Fig. 2 (a) Input images. (b) The ground-truth semantic label map provided by Cityscapes dataset. (c) The semantic segmentation result output by the semantic branch, DeepLabv3+ in this case. (d) The ground-truth error map. (e) The error probablity map generated by our error-prediction branch. the error probability of value 1 is colored in red, and 0 in black. White pixels indicate the unlabeled pixels in the dataset.

network for semantic segmentation. *arXiv preprint arXiv:1805.10180*, 2018.

[6] X. Li, Z. Liu, P. Luo, C. Change Loy, and X. Tang. Not all pixels are equal: Difficulty-aware semantic segmentation via deep layer cascade. In *CVPR*, pages 3193–3202, 2017.

[7] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba. Scene parsing through ade20k dataset. In *CVPR*, pages 633–641, 2017.

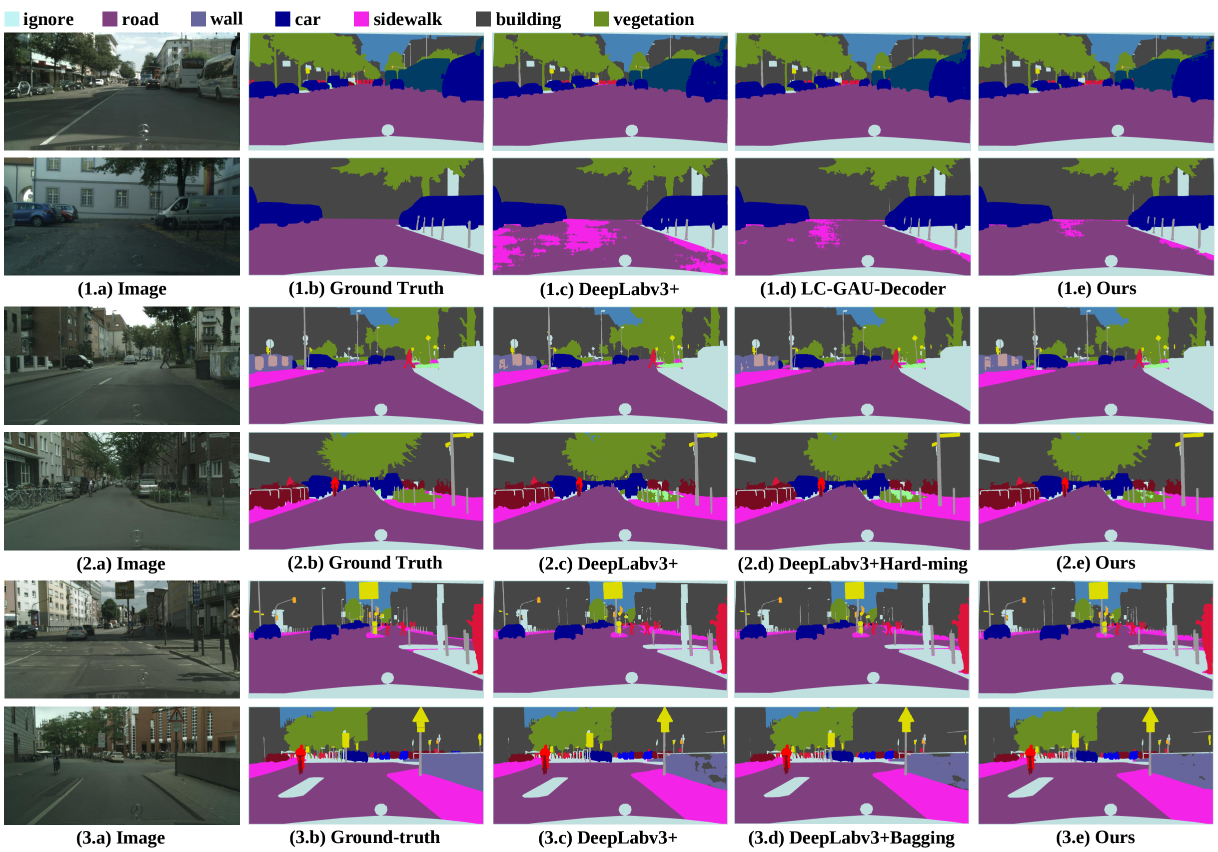


Fig. 3 Visual comparison. (1.a)-(1.e): Comparison with 'LC [6]-GAU-Decoder'; (2.a)-(2.e): Comparison with 'DeepLabv3+ [2]+Hard-mine'; (3.a)-(3.e): Comparison with 'DeepLabv3+ [2]+Bagging'.

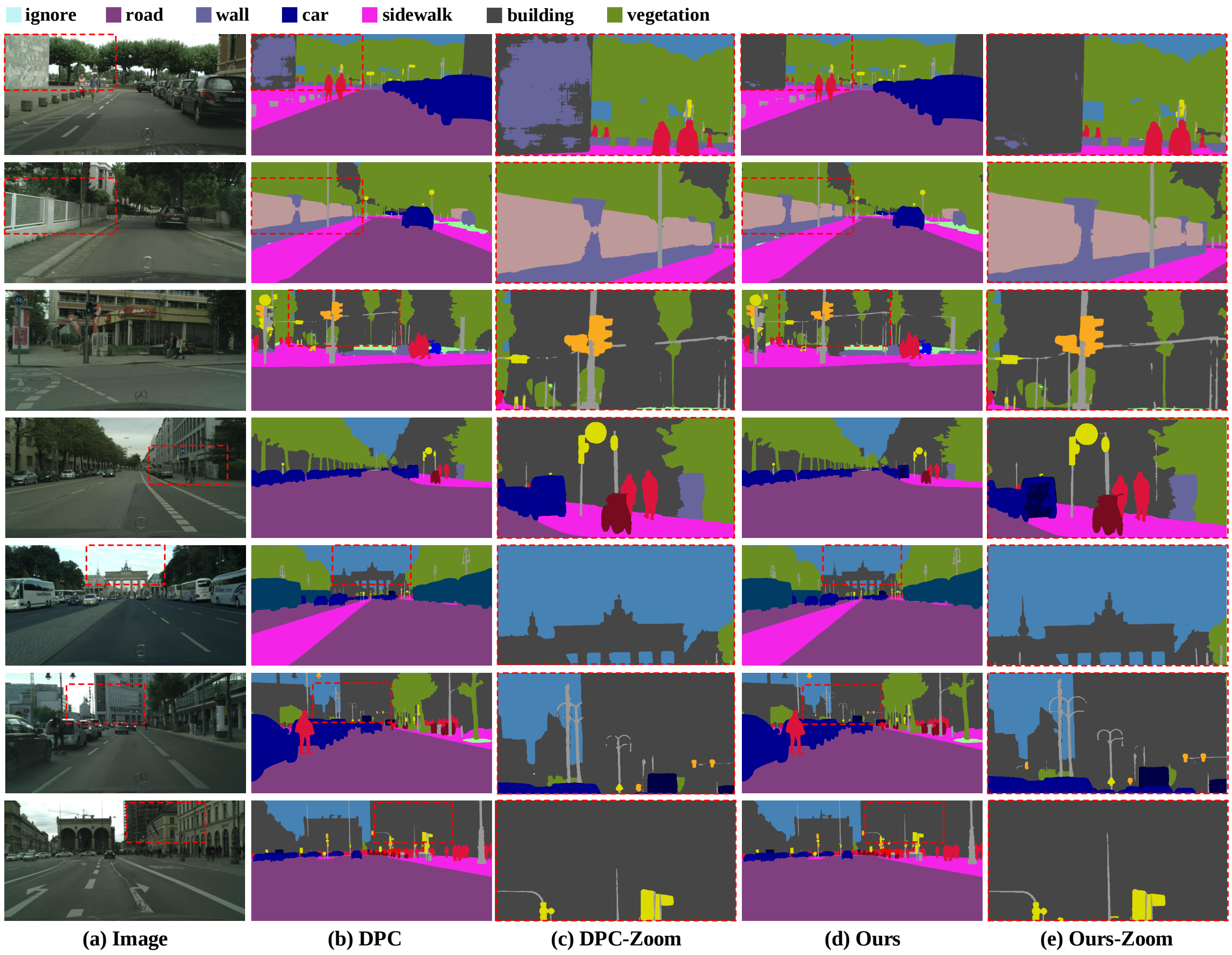


Fig. 4 Visual results on Cityscapes testing set. Comparing the details in segmentation result, our method has better performance than Xception71-DPC [1] which we take as the *semantic branch*. The dashed rectangles highlight the regions where our method can effectively correct the errors in the front end model results.