CRLNet: Cascaded Resolution Learning Network for Unstructured Semantic Segmentation in Natural Scenes

Wei Li[b,c], Shishun Tian[a,b,c], Guoguang Hua[b,c], Muxin Liao[a,c], Yuhang Zhang[b,c], Wenbin Zou[a,b,c,*]

[a]Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, China, 518060, Guangdong, shenzhen
[b]Shenzhen Key Laboratory of Advanced Machine Learning and Applications, Shenzhen University, China, 518060, Guangdong, shenzhen
[c]College of Electronics and Information Engineering, Shenzhen University, China, 518060, Guangdong, shenzhen

Keywords: Unstructured semantic segmentation, deep neural networks, partition-fuse channel attention module, features learning module

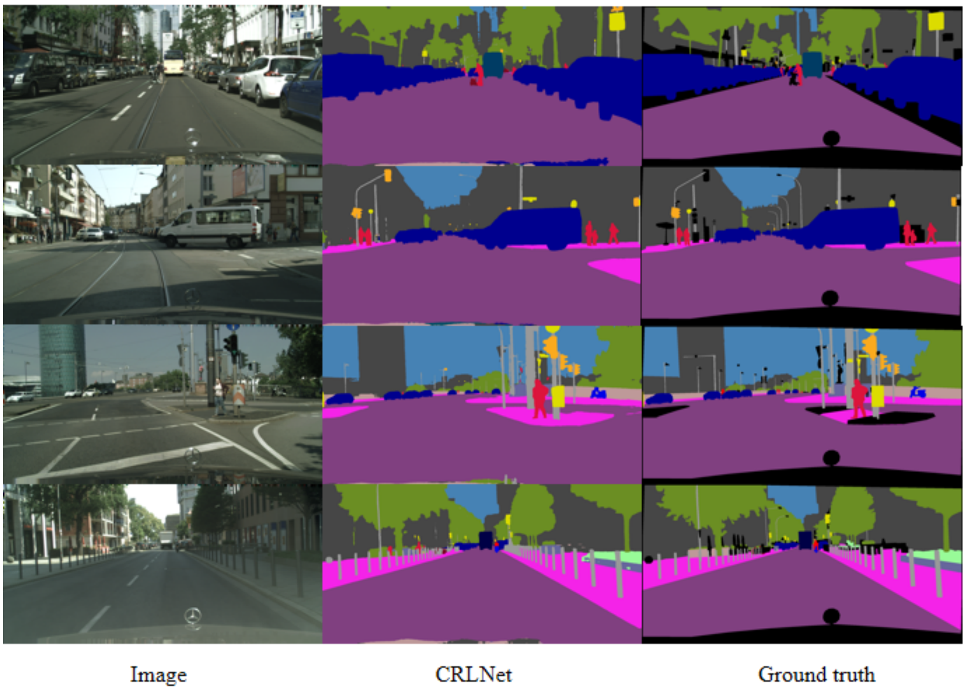

Figure 1: Qualitative examples of CRLNet on Citysacpes validation set. Please zoom in to view more details.

Moreover, we also validate the generalizability of CRLNet on another widely used outdoor street scene dataset, Cityscapes. We first demonstrate the segmentation performance of CRLNet compared to some recent state-of-the-art multi-scale multi-feature semantic segmentation algorithms in Table 2. From Table 2, it can be seen that our proposed

[★]This document is the results of the research project funded by the National Science Foundation, Guangdong Provincial Natural Science Foundation, Shenzhen Natural Science Foundation.

[*]Corresponding author at: College of Electronics and Information Engineering, Shenzhen University, Shenzhen, 518060, China. e-mail: wzou@szu.edu.cn (Wenbin Zou)

[1]College of Electronics and Information Engineering, Shenzhen University, Shenzhen, 518060, China.e-mail: 2250432002@szu.edu.cn

Table 1: Results on Cityscapes test set for segmentation models. The best-performing data is highlighted in bold, while the second-best data is underlined.

Method	Roa	Sid	Bui	Wal	Fen	Pol	TLi	TSi	Veg	Ter	Sky	Per	Rid	Car	Tru	Bus	Tra	Mot	Bic	mIoU↑	Iou-C↑
CGNet[1]	95.5	78.7	88.1	40	43	54.1	59.8	63.9	89.6	67.6	92.9	74.9	54.9	90.2	44.1	59.5	25.2	47.3	60.2	64.8	85.7
EDANet[2]	97.8	80.6	89.5	42	46	52.3	59.8	65	91.4	68.7	93.6	75.7	54.3	92.4	40.9	58.7	56	50.2	64	67.3	85.8
ERFNet[3]	97.2	80	89.5	41.6	45.3	56.4	60.5	64.6	91.4	68.7	94.2	76.1	56.4	92.4	45.7	60.6	27	48.7	61.8	66.3	85.2
ICNet[4]	97.1	79.2	89.7	43.2	48.9	61.5	60.4	63.4	91.5	68.3	93.5	74.6	56.1	92.6	51.3	72.7	51.3	53.6	70.5	69.5	86.4
LEDNet[5]	**98.1**	79.5	91.6	47.7	49.9	**62.8**	61.3	72.8	92.6	61.2	94.9	76.2	53.7	90.9	64.4	64	52.7	44.4	71.6	70.6	87.1
ESNet[6]	**98.1**	80.4	**92.4**	48.3	49.2	61.5	62.5	72.3	92.5	61.5	94.4	76.6	53.2	94.4	62.5	74.3	52.4	45.4	71.4	70.7	87.4
AGLNet[7]	97.8	80.1	91.0	51.3	50.6	58.3	63.0	68.5	92.3	**71.3**	94.2	80.1	59.6	93.8	48.4	68.1	42.1	52.4	67.8	70.1	87.0
NDNet[8]	97.4	80.8	91.0	49.1	51.9	55.7	64.1	72.1	91.3	61.0	93.5	76.9	58.4	92.9	70.7	79.1	71.5	60.4	73.2	64	73.2
DSANet[9]	96.8	78.5	91.2	50.5	50.8	59.4	64.0	71.7	92.6	70.0	94.5	81.8	61.9	92.9	56.1	75.6	50.6	50.9	66.8	71.4	87.4
FDDWNet[10]	98.0	82.4	91.1	52.5	51.2	59.9	64.4	68.9	92.5	70.3	94.4	80.8	59.8	**94.0**	56.5	68.9	48.6	55.7	67.7	71.5	-
FPANet[11]	97.1	78.1	90.0	45.0	50.0	58.4	65.9	72.0	92.3	69.4	94.4	81.9	62.7	93.9	50.2	71.7	60.8	57.1	69.6	72.0	87.7
JPANet[12]	**98.1**	82.7	90.6	47.1	48.2	53.1	60.2	66.1	91.6	69.2	94.5	78.6	60.5	**94.0**	64.2	75.9	66.5	63.7	65.9	71.6	86.7
CRLNet	98.0	**82.8**	92.2	**53.8**	**58.7**	62.3	**72.4**	**75.5**	**92.9**	69.7	**95.0**	**83.9**	**67.9**	93.8	**64.9**	**79.5**	**72.2**	**64.9**	**72.9**	**76.5**	**89.4**

[1] Roa:road; Sid:sidewalk; Bui:building; Wa:wall; Fen:fence; Pol:pole; TLi:traffice light; Tsi:traffice sign; Veg:vegetation; Ter:terrain; Sky:sky; Per:person; Rid:rider; Car:car; Tru:truck; Bus:bus; Tra:train; Mot:motorcycle; Bic:bicycle; Iou-C: Iou Categories.

Table 2: Results on Cityscapes dataset for different SOTA methods. The best-performing data is highlighted in bold.

Methods	Resolution	mIoU %(test/val) ↑
CSRNet [13]	768×768	76.0/77.3
FRNet [14]	512×1024	70.4/-
GDN [15]	512× 1024	75.6/-
Relaxnet [16]	512×1024	74.8/-
LEANet [17]	512×1024	71.9/-
LMFFNet [18]	512×1024	75.1/74.9
FBSNet [9]	512×1024	70.9/-
DABNet [19]	1024× 2048	70.1/69.6
CGNet [1]	1024× 2048	64.8/63.5
LSPNet [20]	1024×2048	74.9/76.5
CRLNet	1024× 2048	**76.5/82.09**

CRLNet achieves a mIoU of 76.5% on the cityscapes test set at a resolution of 1024x2048, which is 0.9% mIoU higher than GDN and 1.6% mIou higher than LSPNet, both recently published methods. Additionally, in Table 1, we present the classification results for individual objects on the Cityscapes test set. From Table 1, it can be seen that our network CRLNet exhibits good segmentation performance on regular objects such as walls, fences, traffic lights, and traffic signs. It also performs well on non-structured objects such as vegetation and trucks. This indicates the strong generalizability of CRLNet to various outdoor scenes. Besides, some results on the Cityscapes validation set are shown in Fig. 1, which intuitively show that our proposed CRLNet network can segment the urban street well.

1. CRediT authorship contribution statement

Wei Li: Methodology, Writing–original draft. **Shishun Tian**: Supervision, Methodology, Writing–review & editing. **Guoguang Hua**: Methodology & editing. **Muxin Liao**: Methodology & editing. **Yuhang Zhang**: Methodology & editing. **Wenbin Zou**: Supervision, Methodology, Writing–review & editing.

2. Data availability

Data will be made available on request.

3. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

4. Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under grants 62171294, 62101344, in part by the Natural Science Foundation of Guangdong Province, China under grants 2020A1515010959, in part by Natural Science Foundation of Shenzhen under grants JCYJ20200109105832261, JCYJ20190808122409660, in part by the Tencent "Phinoceros Birds" Scientific Research Foundation for Young Teachers of Shenzhen University and in part by the Interdisciplinary Innovation Team of Shenzhen University.

References

[1] T. Wu, S. Tang, R. Zhang, J. Cao, Y. Zhang, Cgnet: A light-weight context guided network for semantic segmentation, IEEE Transactions on Image Processing 30 (2020) 1169–1179.

[2] S.-Y. Lo, H.-M. Hang, S.-W. Chan, J.-J. Lin, Efficient dense modules of asymmetric convolution for real-time semantic segmentation, in: Proceedings of the ACM Multimedia Asia, 2019, pp. 1–6.

[3] E. Romera, J. M. Alvarez, L. M. Bergasa, R. Arroyo, Erfnet: Efficient residual factorized convnet for real-time semantic segmentation, IEEE Transactions on Intelligent Transportation Systems 19 (2017) 263–272.

[4] H. Zhao, X. Qi, X. Shen, J. Shi, J. Jia, Icnet for real-time semantic segmentation on high-resolution images, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 405–420.

[5] Y. Wang, Q. Zhou, J. Liu, J. Xiong, G. Gao, X. Wu, L. J. Latecki, Lednet: A lightweight encoder-decoder network for real-time semantic segmentation, in: 2019 IEEE international conference on image processing (ICIP), IEEE, 2019, pp. 1860–1864.

[6] W. Yu, Q. Zhou, J. Xiong, X. Wu, X. Jin, Esnet: An efficient symmetric network for real-time semantic segmentation, in: Pattern Recognition and Computer Vision: Second Chinese Conference, PRCV 2019, Xi'an, China, November 8–11, 2019, Proceedings, Part II 2, Springer, 2019, pp. 41–52.

[7] Q. Zhou, Y. Wang, Y. Fan, X. Wu, S. Zhang, B. Kang, L. J. Latecki, Aglnet: Towards real-time semantic segmentation of self-driving images via attention-guided lightweight network, applied soft computing 96 (2020) 106682.

[8] Z. Yang, H. Yu, M. Feng, W. Sun, X. Lin, M. Sun, Z.-H. Mao, A. Mian, Small object augmentation of urban scenes for real-time semantic segmentation, IEEE Transactions on Image Processing 29 (2020) 5175–5190.

[9] G. Gao, G. Xu, J. Li, Y. Yu, H. Lu, J. Yang, Fbsnet: A fast bilateral symmetrical network for real-time semantic segmentation, IEEE Transactions on Multimedia (2022).

[10] J. Liu, Q. Zhou, Y. Qiang, B. Kang, X. Wu, B. Zheng, Fddwnet: a lightweight convolutional neural network for real-time semantic segmentation, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 2373–2377.

[11] H. Yin, W. Xie, J. Zhang, Y. Zhang, W. Zhu, J. Gao, Y. Shao, Y. Li, Dual context network for real-time semantic segmentation, Machine Vision and Applications 34 (2023) 22.

[12] X. Hu, L. Jing, U. Sehar, Joint pyramid attention network for real-time semantic segmentation of urban scenes, Applied Intelligence 52 (2022) 580–594.

[13] J. Xiong, L.-M. Po, W.-Y. Yu, C. Zhou, P. Xian, W. Ou, Csrnet: Cascaded selective resolution network for real-time semantic segmentation, Expert Systems with Applications 211 (2023) 118537.

[14] M. Lu, Z. Chen, Q. J. Wu, N. Wang, X. Rong, X. Yan, Frnet: Factorized and regular blocks network for semantic segmentation in road scene, IEEE Transactions on Intelligent Transportation Systems 23 (2020) 3522–3530.

[15] D. Luo, H. Kang, J. Long, J. Zhang, X. Liu, T. Quan, Gdn: Guided down-sampling network for real-time semantic segmentation, Neurocomputing 520 (2023) 205–215.

[16] J. Liu, X. Xu, Y. Shi, C. Deng, M. Shi, Relaxnet: Residual efficient learning and attention expected fusion network for real-time semantic segmentation, Neurocomputing 474 (2022) 115–127.

[17] X.-L. Zhang, B.-C. Du, Z.-C. Luo, K. Ma, Lightweight and efficient asymmetric network design for real-time semantic segmentation, Applied Intelligence 52 (2022) 564–579.

[18] M. Shi, J. Shen, Q. Yi, J. Weng, Z. Huang, A. Luo, Y. Zhou, Lmffnet: a well-balanced lightweight network for fast and accurate semantic segmentation, IEEE Transactions on Neural Networks and Learning Systems (2022).

[19] G. Li, I. Yun, J. Kim, J. Kim, Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation, arXiv preprint arXiv:1907.11357 (2019).

[20] Y. Zhang, T. Yao, Z. Qiu, T. Mei, Lightweight and progressively-scalable networks for semantic segmentation, International Journal of Computer Vision (2023) 1–19.