IEEE International Conference on Image Processing
25-28 October 2020, United Arab Emirates
ICIP 2020
IEEE Signal Processing Society

Point Set Attention Network for Semantic Segmentation

AUTHOR: JIE JIANG, JING LIU, JUN FU, XINXIN ZHU, HANQING LU

REPORTER: JIE JIANG

NATIONAL LABORATORY OF PATTERN RECOGNITION, INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES

SCHOOL OF ARTIFICIAL INTELLIGENCE, UNIVERSITY OF CHINESE ACADEMY OF SCIENCES

Content

1. Background
2. Method
3. Experiments
4. Conclusion

Content

1. Background
2. Method
3. Experiments
4. Conclusion

Background: Semantic segmentation task

Image semantic segmentation aims to assign semantic labels to every pixel in an image

Background: Applications

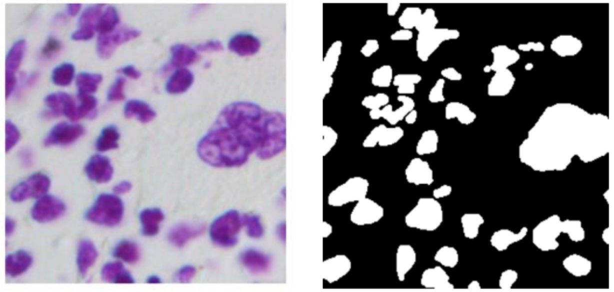
Instance segmentation

Human parsing

Saliency Detection

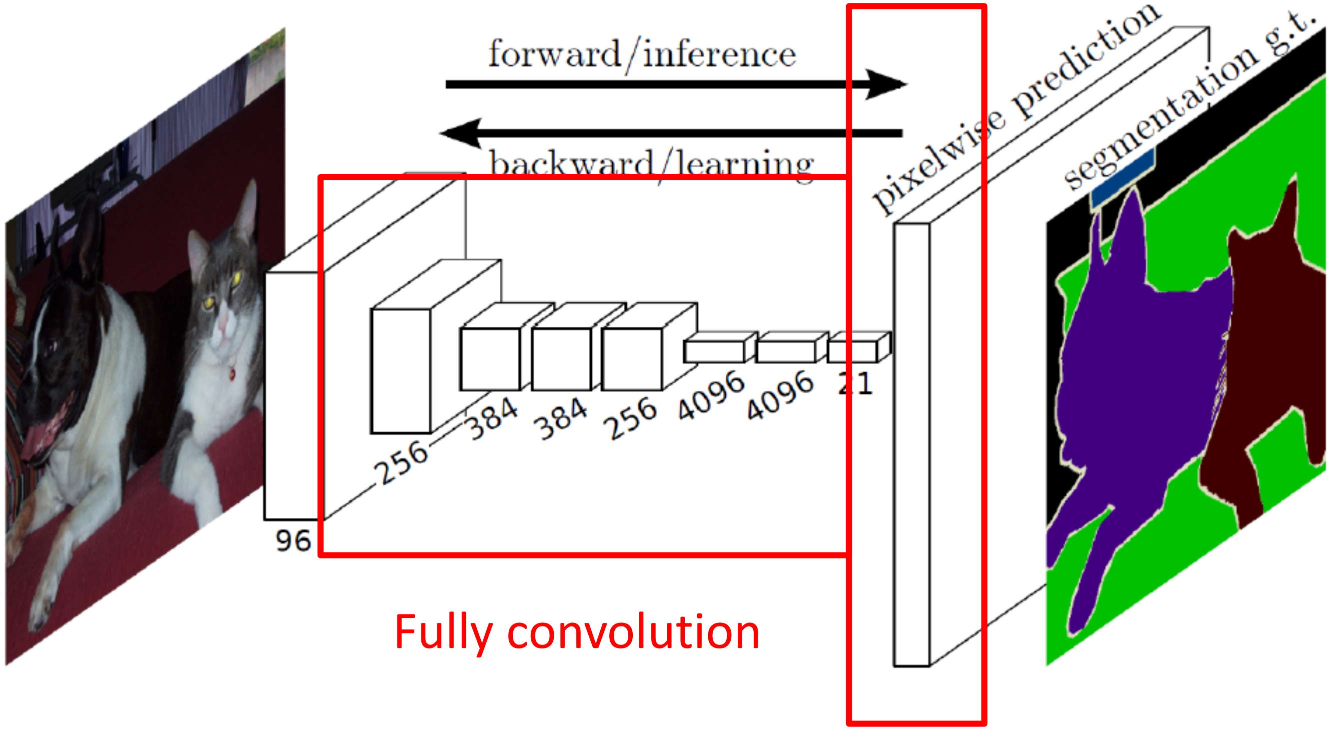
Automatic drive

Photograph



Medical diagnosis

Background: Fully convolutional network (FCN)



32X bilinear interpolation

Consecutive down-sampling operation

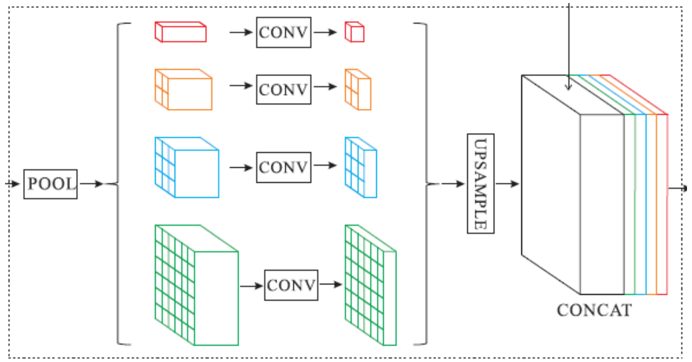
Missing small objects and details

Fully convolution

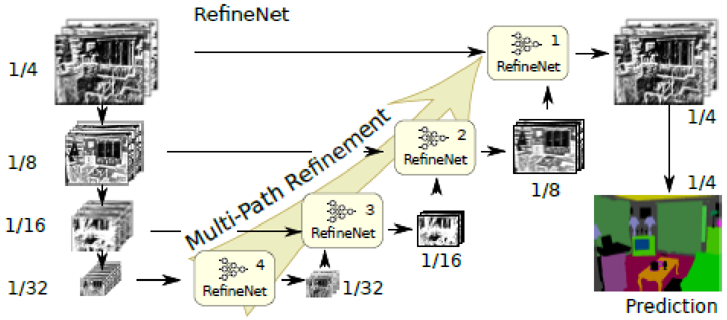
Inconsistent prediction in big objects

Background: Contextual information

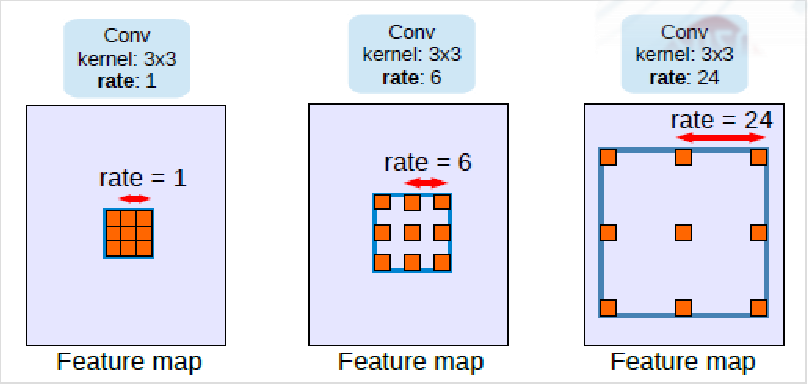
Multi-scale Context: aggregate multi-scale contextual information equally



Zhao, Hengshuang, et al. "Pyramid scene parsing network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

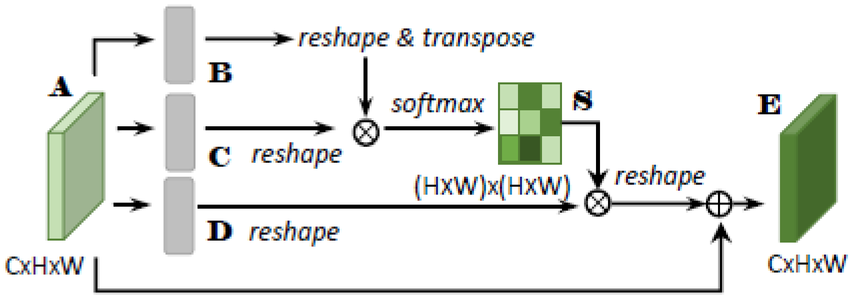


Lin, Guosheng, et al. "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation." CVPR, 2017.

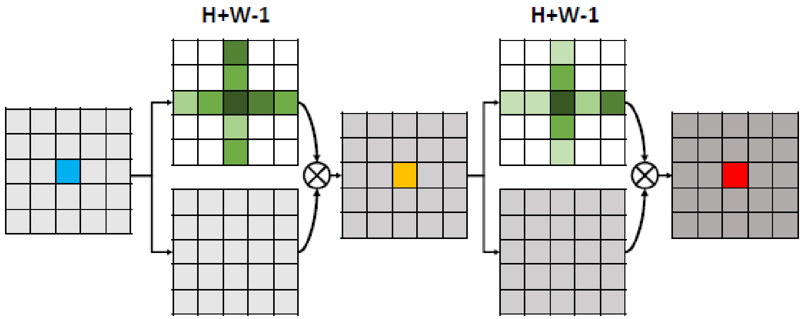


Chen, Liang-Chieh, et al. “Encoder-decoder with atrous separable convolution for semantic image segmentation.” ECCV. 2018.

Relation Context: self-attention mechanism captures the long-range dependencies between pixels



Fu, Jun, et al. "Dual attention network for scene segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

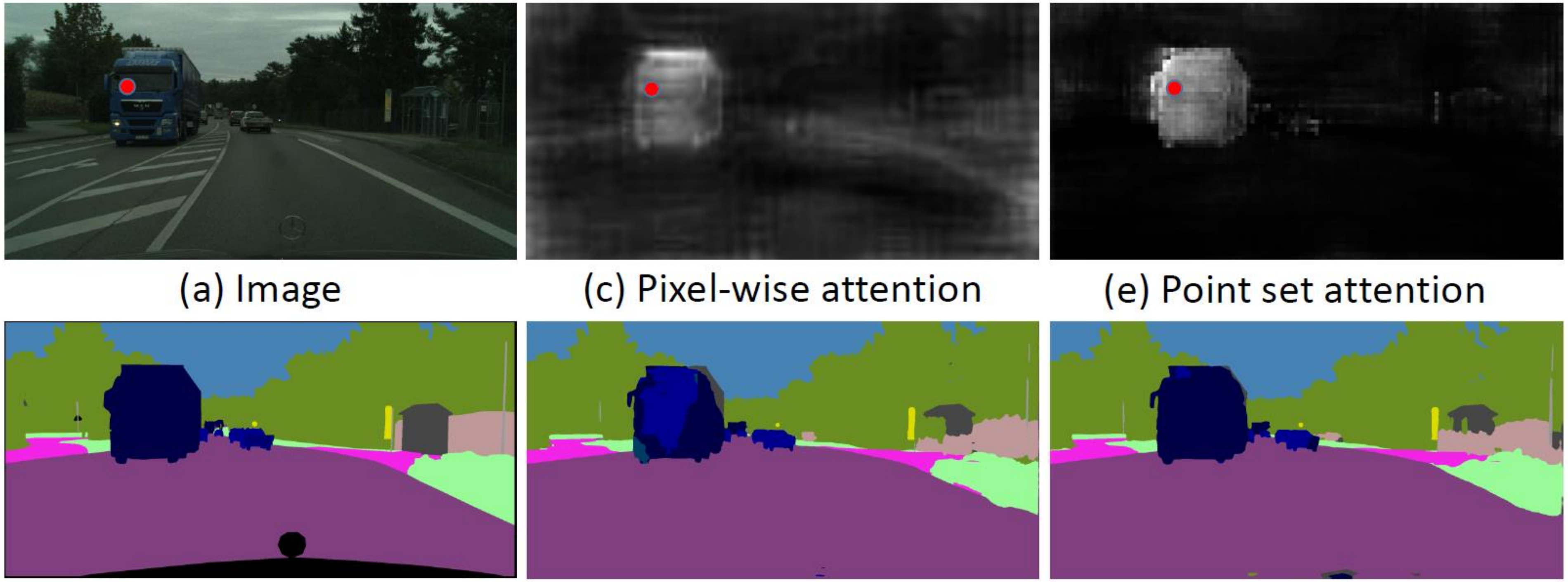


Huang, Zilong, et al. "Ccnet: Criss-cross attention for semantic segmentation." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

Content

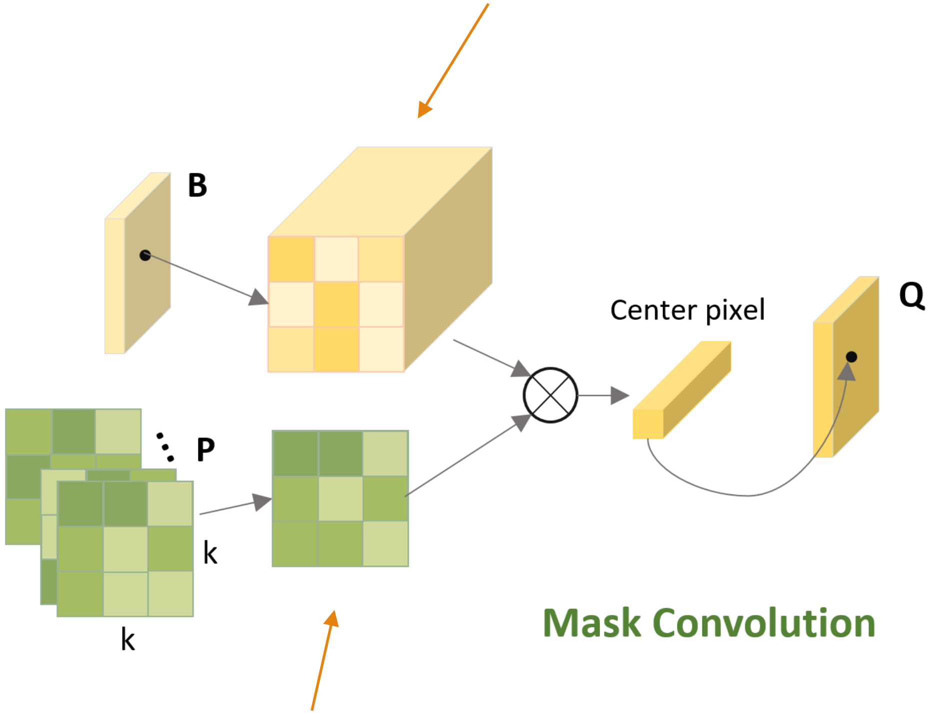
1. Background
2. Method
3. Experiments
4. Conclusion

Method: the defect of self-mechanism


(a) Image (c) Pixel-wise attention (e) Point set attention
(b) Groundtruth (d) Pixel-wise segmentation (f) Point set segmentation

Method: Point set attention network (PSANet)

Point set: current pixel and pixels in its neighborhood

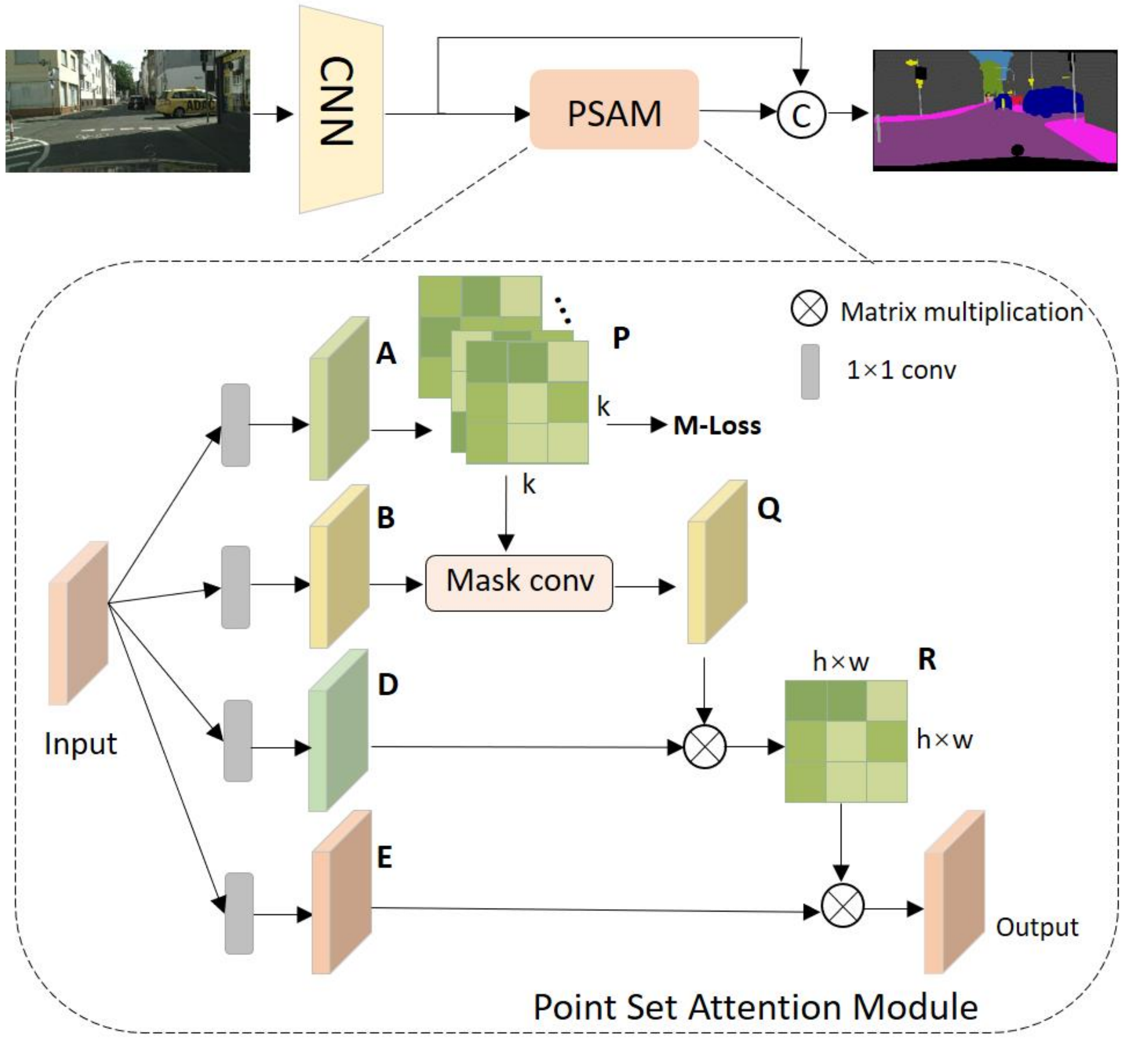


$$Q_{l,i,j} = \sum_{m=-s}^{s} \sum_{n=-s}^{s} P_{((m+s)\times k+(n+s)),i,j} \cdot B_{l,i+n,j+m} \quad (1)$$

where $s = \lfloor k/2 \rfloor$, $Q_{l,i,j} \in \mathbf{Q}$, $l \in [1, 2, ..., C']$, $i \in [1, 2, ..., H]$ and $j \in [1, 2, ..., W]$.

Mask: pixels with the same class as current pixel would response highly

Method: Point set attention network (PSANet)



1. Generate mask feature P
2. Apply a mask convolution on feature B with mask feature P, obtaining an updated feature Q
3. Model the relation between updated feature Q and context feature D with self-attention mechanism
4. Adopt M-Loss to regularize the training

Content

1. Background
2. Method
3. Experiments
4. Conclusion

Experiments: Ablation experiments

Table 1. Ablation experiments of point set attention on Cityscapes validation set. Os denotes the output feature size compared with input image. Point set(w/o mask) denotes the mask feature in PSAM is set to uniform distribution.

Method	os=16(mIoU%)	os=8(mIoU%)
baseline	69.06	70.73
+pixel-wise	72.77	75.10
+point set(w/o mask)	73.78	75.37
+point set	**74.18**	**76.54**

Table 2. Ablation experiments of point set size on Cityscapes validation set. k denotes the size of the point set.

k	mIoU%	GFLOPs	Memory(MB)
1	72.77	7.10	679.1
3	73.73	8.69	739.3
5	**74.18**	9.32	762.1
7	73.30	11.32	795.8

Experiments: Compare with state-of-the-art methods

Table 3. Comparisons with state-of-the-art approaches on Cityscapes test set.

Method	**BaseNet**	**mIoU%**
DeepLab-v2 [12]	Res-101	70.4
PSPNet [3]	Res-101	78.4
BiSeNet [15]	Res-101	78.9
OCNet[6]	Res-101	80.1
DenseASPP [13]	Dense-161	80.6
PSANet	Res-101	**81.5**

Table 4. Comparisons with state-of-the-art approaches on the testing set of PASCAL-Context validate set.

Method	**BaseNet**	**mIoU%**
PSPNet [3]	Res-101	47.8
EncNet [11]	Res-101	51.7
DANet [5]	Res-101	52.6
CFNet [17]	Res-101	54.0
ACNet [18]	Res-101	54.1
PSANet	Res-101	**55.1**

Content

1. Background
2. Method
3. Experiments
4. Conclusion

Conclusion

Main contribution:

(1) We propose a Point Set Attention Network (PSANet) to improve self-attention mechanism in noisy pixels and generate intra-class common features for semantic segmentation.

(2) We introduce context-aware mask feature to assist pixels to contribute intra-class mutual improvement.

(3) The proposed PSANet achieves state-of-the-art performance on Cityscapes and PASCAL Context datasets. In particular, PSANet obtains 81.5% mIoU on Cityscapes test set without using coarse data and 55.1% mIoU on PASCAL Context validate set.

Thank you