ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

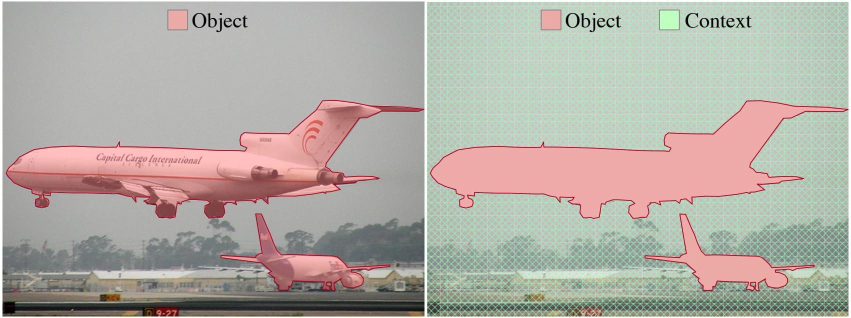
WWW.ECCV2020.EU

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], Bei Yu[1]

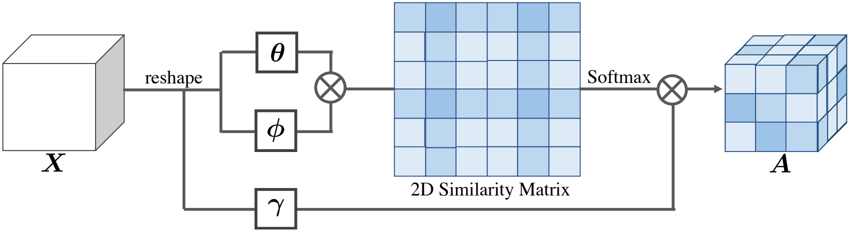
[1]CSE Department, Chinese University of Hong Kong
[2]Shanghai Jiao Tong University
[3]Shenzhen Institutes of Advanced Technology
[4]SmartMore

Introduction



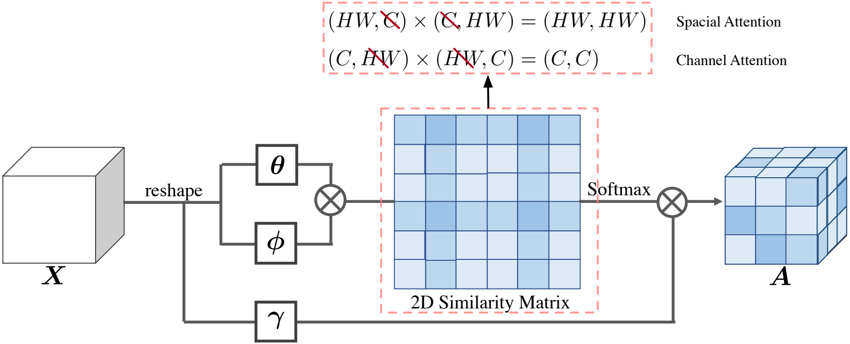
Context information plays an indispensable role in the success of semantic segmentation.

Introduction



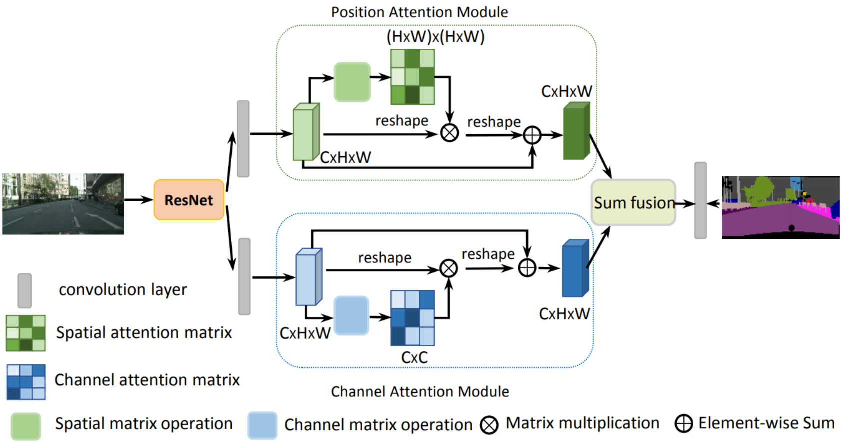
Non-local attention based methods become the main stream of semantic segmentation.

Introduction



Spatial or channel attention? A dilemma in Non-local self-attention based approaches.

Introduction



Architecture of DANet [1], which contains 2 stream of non-local attentions.

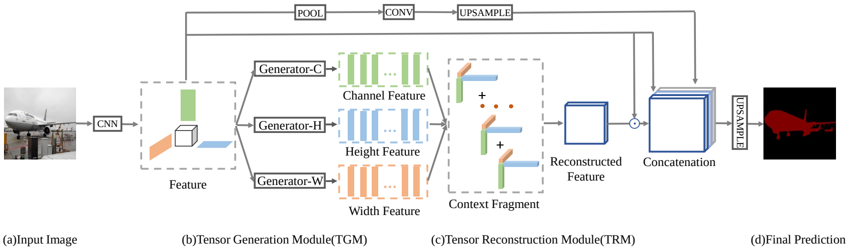
Introduction

Can we obtain spatial and channel attention **simultaneously**?

- Better context representation.
- Smaller computational cost.

Our Proposed RecoNet

Tensor Reconstruction Network (RecoNet).



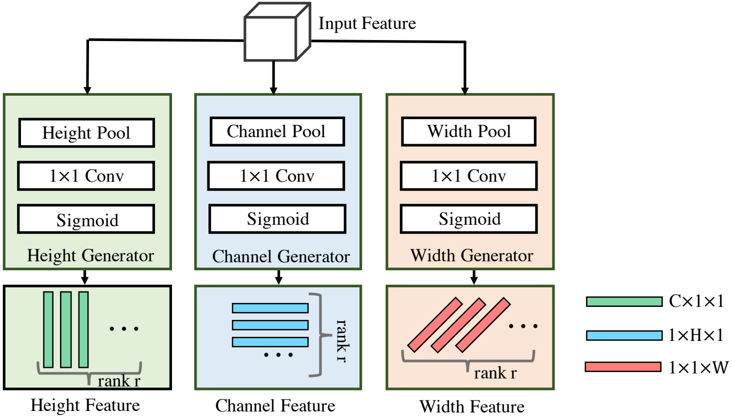
The pipeline of our framework. Two major components are involved, Tensor Generation Module (TGM) and Tensor Reconstruction Module (TRM). TGM peroforms the low-rank tensor generation while TRM achieves the high-rank tensor reconstruction via CP construction theory.

Our Proposed RecoNet

Tensor canonical-polyadic decomposition (CP decomposition).
Assuming we have $3r$ vectors in C/H/W directions $\boldsymbol{v}_{ci} \in \mathbb{R}^{C\times 1\times 1}$, $\boldsymbol{v}_{hi} \in \mathbb{R}^{1\times H\times 1}$ and $\boldsymbol{v}_{wi} \in \mathbb{R}^{1\times 1\times W}$, where $i \in r$ and r is the tensor rank. These vectors are the CP decomposed fragments of $\boldsymbol{A} \in \mathbb{R}^{C\times H\times W}$, then tensor CP rank-r reconstruction is defined as:

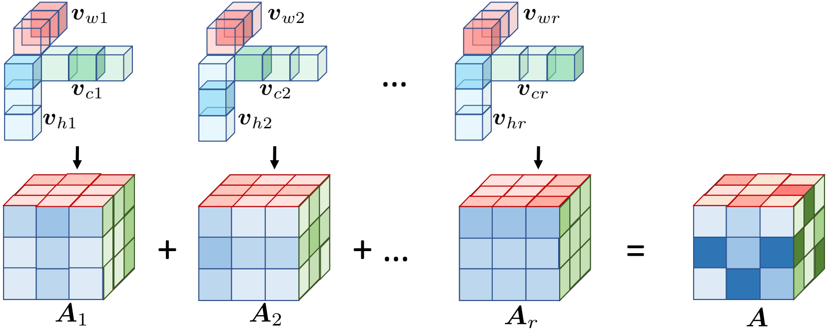
$$\boldsymbol{A} = \sum_{i=1}^{r} \lambda_i \boldsymbol{v}_{ci} \otimes \boldsymbol{v}_{hi} \otimes \boldsymbol{v}_{wi}, \tag{1}$$

Tensor Generation Module



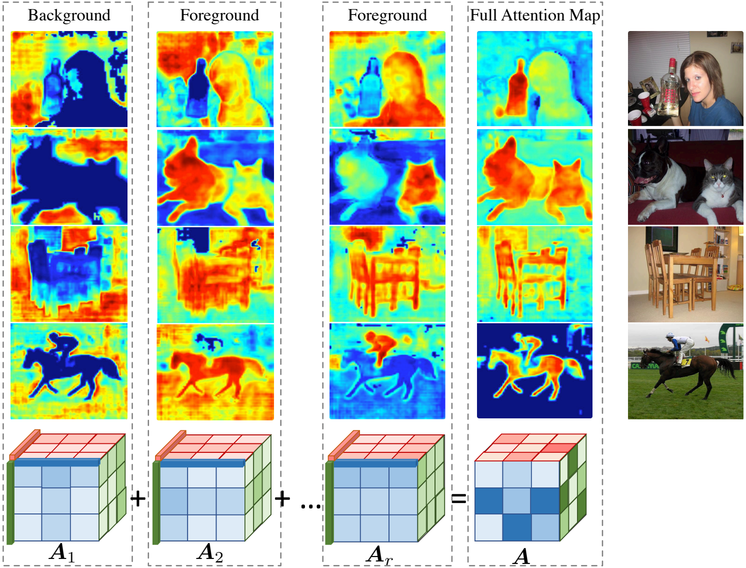
Tensor Generation Module. Channel Pool, Height Pool and Width Pool are all global average pooling.

Tensor Reconstruction Module



Tensor Reconstruction Module (TRM). The pipeline of TRM consists of two main steps, sub-attention map generation and global context reconstruction. The processing from top to bottom (see $\downarrow$) indicates the sub-attention map generation from three dimensions (channel / height / width). The processing from left to right (see $\boldsymbol{A}_1 + \boldsymbol{A}_2 + \cdots + \boldsymbol{A}_r = \boldsymbol{A}$) denotes the global context reconstruction from low-rank to high-rank.

Visualization



Results on PASCAL-VOC12 w/o COCO-pretrained model

	FCN [2]	PSPNet [3]	EncNet [4]	APCNet [5]	CFNet [6]	DMNet [7]	RecoNet
aero	76.8	91.8	94.1	95.8	95.7	**96.1**	93.7
bike	34.2	71.9	69.2	75.8	71.9	**77.3**	66.3
bird	68.9	94.7	**96.3**	84.5	95.0	94.1	95.6
boat	49.4	71.2	**76.7**	76.0	76.3	72.8	72.8
bottle	60.3	75.8	86.2	80.6	82.8	78.1	**87.4**
bus	75.3	95.2	96.3	96.9	94.8	**97.1**	94.5
car	74.7	89.9	90.7	90.0	90.0	**92.7**	92.6
cat	77.6	95.9	94.2	96.0	95.9	96.4	**96.5**
chair	21.4	39.3	38.8	42.0	37.1	39.8	**48.4**
cow	62.5	90.7	90.7	93.7	92.6	91.4	**94.5**
table	46.8	71.7	73.3	75.4	73.0	75.5	**76.6**
dog	71.8	90.5	90.0	91.6	93.4	92.7	**94.4**
horse	63.9	94.5	92.5	95.0	94.6	95.8	**95.9**
mbike	76.5	88.8	88.8	90.5	89.6	91.0	**93.8**
person	73.9	89.6	87.9	89.3	88.4	90.3	**90.4**
plant	45.2	72.8	68.7	75.8	74.9	76.6	**78.1**
sheep	72.4	89.6	92.6	92.8	**95.2**	94.1	93.6
sofa	37.4	**64**	59.0	61.9	63.2	62.1	63.4
train	70.9	85.1	86.4	88.9	**89.7**	85.5	88.6
tv	55.1	76.3	73.4	79.6	78.2	77.6	**83.1**
mIoU	62.2	82.6	82.9	84.2	84.2	84.4	**85.6**

Computational Cost

Table: Computational cost and GPU occupation of TGM+TRM. FLOPs (FLoating point Operations). We use tensor rank $r = 64$ for evaluation

Method	Channel	FLOPs	GPU Memory
Non-Local [8]	512	19.33G	88.00MB
APCNet [5]	512	8.98G	193.10MB
RCCA [9]	512	5.37G	41.33MB
A^2Net [10]	512	4.30G	25.00MB
AFNB [11]	512	2.62G	25.93MB
LatentGNN [12]	512	2.58G	44.69MB
EMAUnit [13]	512	2.42G	24.12MB
TGM+TRM	512	**0.0215G**	**8.31MB**

Contact

Thanks for watching!
Please feel free to contact with me.
E-mail: 1155137828@link.cuhk.edu.hk
WeChat: ChenWanLi11410579

[1] J. Fu, J. Liu, H. Tian, Z. Fang, and H. Lu, “Dual attention network for scene segmentation,” *arXiv preprint arXiv:1809.02983*, 2018.

[2] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. CVPR*, 2015, pp. 3431–3440.

[3] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proc. CVPR*, 2017, pp. 2881–2890.

[4] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, “Context encoding for semantic segmentation,” in *Proc. CVPR*, 2018, pp. 7151–7160.

[5] J. He, Z. Deng, L. Zhou, Y. Wang, and Y. Qiao, “Adaptive pyramid context network for semantic segmentation,” in *Proc. CVPR*, 2019, pp. 7519–7528.

[6] H. Zhang, H. Zhang, C. Wang, and J. Xie, “Co-occurrent features in semantic segmentation,” in *Proc. CVPR*, 2019, pp. 548–557.

[7] J. He, Z. Deng, and Y. Qiao, “Dynamic multi-scale filters for semantic segmentation,” in *Proc. ICCV*, 2019, pp. 3562–3572.

[8] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in *Proc. CVPR*, 2018, pp. 7794–7803.

[9] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. ICCV*, 2019, pp. 603–612.

[10] Y. Chen, Y. Kalantidis, J. Li, S. Yan, and J. Feng, "Aˆ 2-Nets: Double attention networks," in *Proc. NIPS*, 2018, pp. 352–361.

[11] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *Proc. ICCV*, 2019, pp. 593–602.

[12] S. Zhang, X. He, and S. Yan, "LatentGNN: Learning efficient non-local relations for visual recognition," in *Proc. ICML*, 2019, pp. 7374–7383.

[13] X. Li, Z. Zhong, J. Wu, Y. Yang, Z. Lin, and H. Liu, "Expectation-maximization attention networks for semantic segmentation," in *Proc. ICCV*, 2019, pp. 9167–9176.

ECCV'20 ONLINE 23-28 AUGUST 2020