The progress of the reading plan:

08 / 50

Paper Information

Paper Title :

Learning a Discriminative Feature Network for Semantic Segmentation

Conference :

CVPR 2018 **Spotlight**

Authors and Institutions

Authors

- Changqian Yu 1
- Jingbo Wang 2
- Chao Peng 3
- Changxin Gao 1
- Gang Yu 3
- Nong Sang 1

Institutions

- 1 Key Laboratory of Ministry of Education for Image Processing and Intelligent Control, School of Automation, Huazhong University of Science and Technology
- 2 Key Laboratory of Machine Perception, Peking University
- 3 Megvii Inc. (Face++)

Official Codes

Inofficial Code1 Inofficial Code2

Some articles to comprehend this paper

Official in Chinese l 旷视科技Face++提出用于语义分割的判别特征网络DFN 论文阅读 - Learning a Discriminative Feature Network for Semantic Segmentation (CVPR2018)

Network Structure

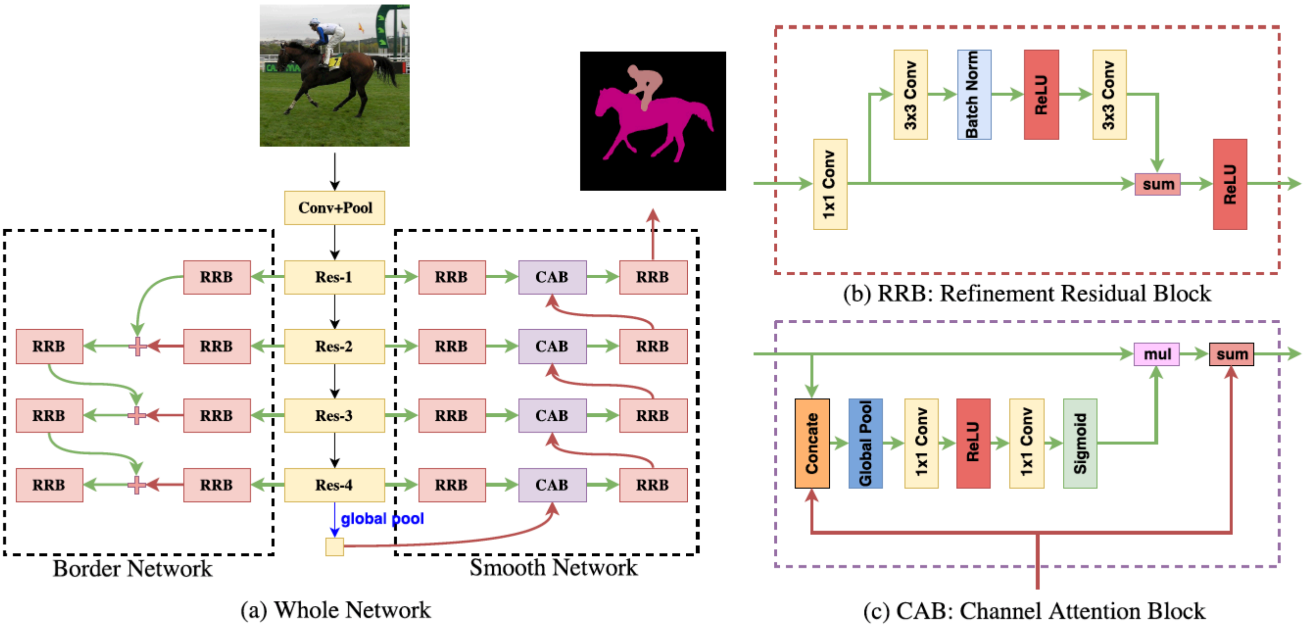


Figure 2. An overview of the Discriminative Feature Network. (a) Network Architecture. (b) Components of the Refinement Residual Block (RRB). (c) Components of the Channel Attention Block (CAB). The red and blue lines represent the upsample and downsample operators, respectively. The green line can not change the size of feature maps, just a path of information passing.

Note

1

To remedy the imbalance of different classes, focal loss [22] is used here to supervise the output of the Border Network.

$$FL(p_k) = -(1 - p_k)^{\gamma} \log p_k$$

p_k is the estimated probability for class k. **Reference** + [22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll´ar. Focal loss for dense object detection. In IEEE International Conference on Computer Vision, 2017.

2

We use a parameter lamada to balance the segmentation loss l_s and the boundary loss l_b:

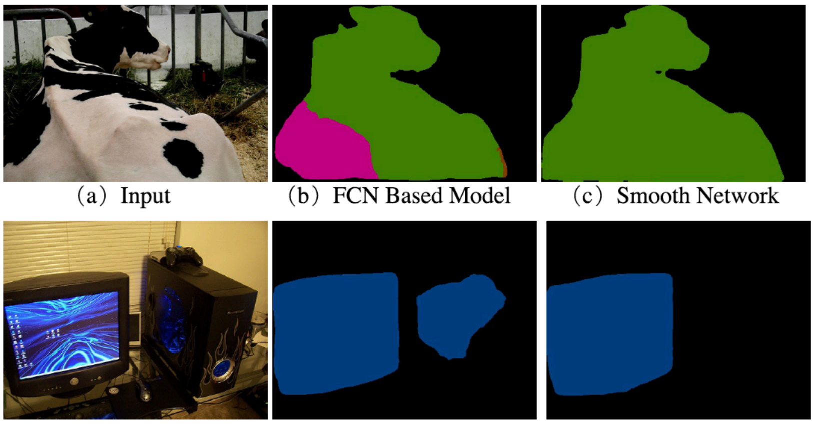
$L = l_s + lamada * l_b$

Five questions about this paper:

1. [Problem Definition / Motivation] What problem is this paper trying to

solve?

Intra-class inconsistency and inter-class indistinction is two important problems in semantic segmentation.



(a) Input (b) FCN Based Model (c) Smooth Network

(d) Input (e) FCN Based Model (f) Border Network

2. [Contribution / Method] What's new in this paper? / How does this paper solve the above problems?

1. intra-class inconsistency problem: a Smooth Network with Channel Attention Block and global average pooling to select the more discriminative features.
   - multi-scale features are used to enrich the information: U-net structure
   - global average pooling is used to capture the global information
   - CAB is designed to use the high level information to guide the low layer.

2. inter-class indistinction: To amplify the distinction of features, a Border Network is proposed to make the bilateral features of boundary distinguishable with deep semantic boundary supervision.

Using Canny algorithm on the semantic ground truth to obtain the supervisory signal, which is used to refine the low stage feature.

3. Details about the experiment

3.1 Which Datasets are used?

- PASCAL VOC 2012
- Cityscapes

3.2 How is the experiment set up?

- multi-scale data augmentation

Table 1. The performance of ResNet-101 with and without random scale augmentation.

Method	Random_Scale	Mean IOU(%)
Res-101		69.26
Res-101	√	72.86

Data augmentation is necessary!

3.3 What's the evaluation metric?

mIoU

3.4 Ablation study

Smooth Network

Detailed performance comparison of our proposed Smooth Network. RRB: refinement residual block. GP: global pooling branch. CAB: channel attention block. DS: deep supervision.

Index	Method	Mean IOU(%)
1	Res-101	72.86
2	Res-101+RRB	76.65
3	Res-101+RRB+GP	78.2
4	Res-101+RRB+GP+CAB	79.31
5	Res-101+RRB+DS	77.08
6	Res-101+RRB+GP+DS	78.51
7	Res-101+RRB+GP+CAB+DS	79.54
2-1	RRB	3.79
3-2	GP	1.55
4-3	CAB	1.11
5-2	DS	0.43
6-5	GP	1.43
7-6	CAB	1.03

We can find that the biggest improvementa are coming from RRB, GP and CAB.

Border Network

Table 3. Combining the Border Network and Smooth Network as Discriminative Feature Network. **SN**: Smooth Network. **BN:** Border Network. **MS_Flip:** Adding multi-scale inputs and left-right flipped inputs.

Method	Mean IOU(%)
Res-101+SN	79.54
Res-101+SN+BN	79.67
Res-101+SN+MS_Flip	79.90
Res-101+SN+BN+MS_Flip	80.01

The improvement of BN is not as big as RRB, GP and CAB.

Figure 4. Results of Smooth Network on PASCAL VOC 2012 dataset.

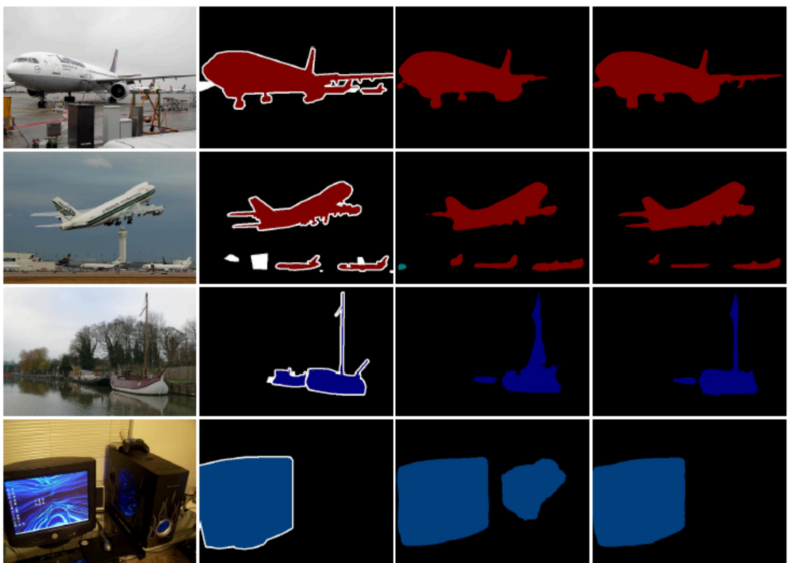


Figure 5. Results of Border Network on PASCAL VOC 2012 dataset. The boundary on prediction is refined by the Border Network.

3.5 What is the ranking of the experiment results?

Table 5. Performance on PASCAL VOC 2012 test set. Methods pre-trained on MS-COCO are marked with $^+$.

Method	Mean IOU(%)
FCN [27]	62.2
Zoom-out [29]	69.6
ParseNet [24]	69.8
Deeplab v2-CRF [5]	71.6
DPN [26]	74.1
Piecewise [20]	75.3
LRR-CRF [11]	75.9
PSPNet [40]	82.6
Ours	**82.7**
DLC$^+$ [18]	82.7
DUC$^+$ [34]	83.1
GCN$^+$ [30]	83.6
RefineNet$^+$ [19]	84.2
ResNet-38$^+$ [35]	84.9
PSPNet$^+$ [40]	85.4
Deeplab v3$^+$ [6]	85.7
Ours$^+$	**86.2**

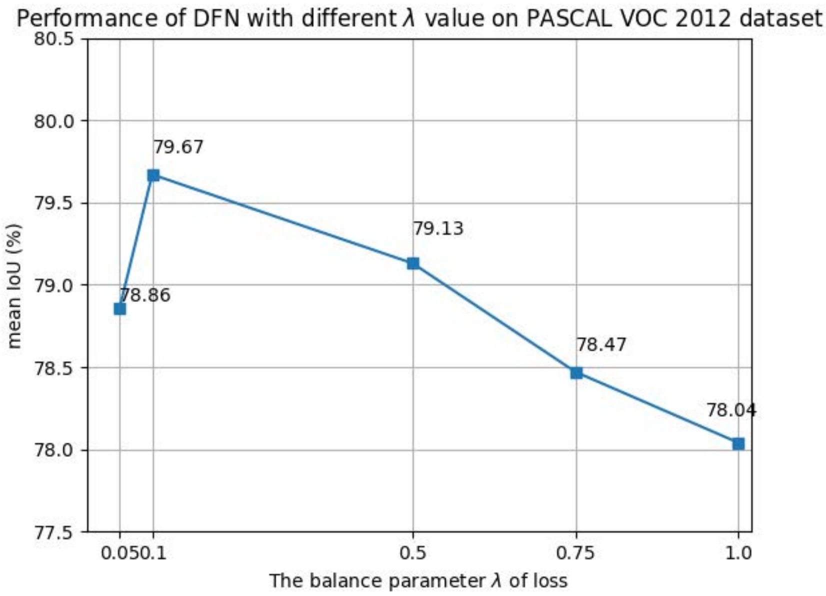
Table 6. Performance on Cityscapes test set. The "-" indicates that the method do not present this result in its paper.

Method	Mean IOU(%)	
	w/o coarse	w/ coarse
CRF-RNN [41]	62.5	-
FCN [27]	65.3	-
DPN [26]	66.8	59.1
LRR [11]	69.7	71.8
Deeplab v2-CRF [5]	70.4	-
Piecewise [20]	71.6	-
RefineNet [19]	73.6	-
SegModel [10]	78.5	79.2
DUC [34]	77.6	80.1
PSPNet [40]	78.4	80.2
Ours	**79.3**	**80.3**

4. Advantages (self-summary rather than the author's)

This paper was accepted as the spotlight paper in CVPR 2018. Here are some reasons I think important: 1. The ablation study is sufficient. 2. The idea of refinement the low stage feature (DS: deep supervision) is interesting. Though DS doesn't improve a lot performance, just 0.43% according to Table 2, this idea is easy to touch the reviewers and readers.

5. Disadvantages (self-summary rather than the author's)



The parameter lamada actually can be set to some value(like 0) at first, and gradually be learned to a proper position during the learning.