The progress of the reading plan:

12 / 50

Paper Information

Paper Title :

Large Kernel Matters——Improve Semantic Segmentation by Global Convolutional Network

Conference :

CVPR 2017

Authors and Institutions

Authors

- Chao Peng 1
- Xiangyu Zhang 2
- Gang Yu 2
- Guiming Luo 1
- Jian Sun 2

Institutions

- 1 School of Software, Tsinghua University
- 2 Megvii Inc. (Face++)

Inofficial Codes

https://github.com/ycszen/pytorch-segmentation

Some articles to comprehend this paper

Network Structure

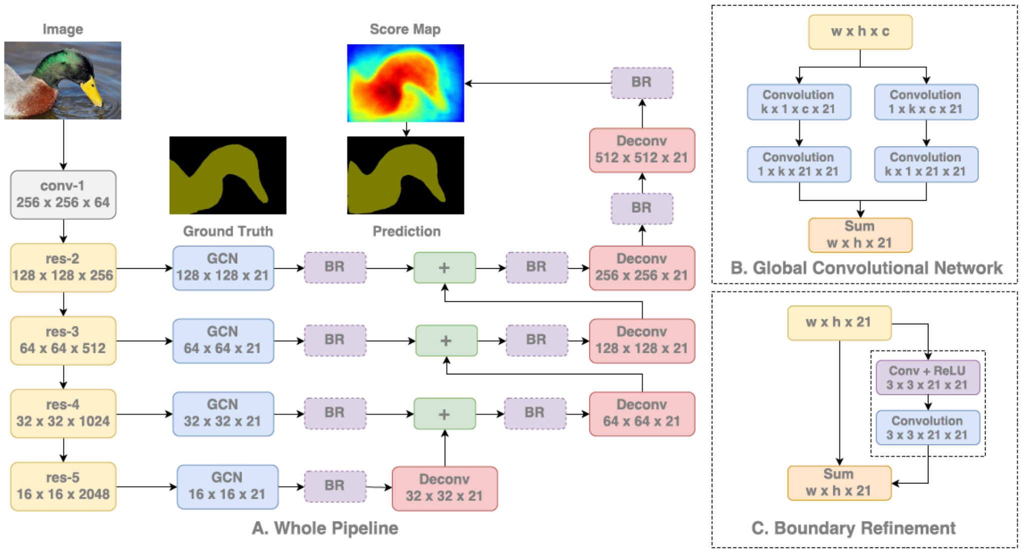


Figure 2. An overview of the whole pipeline in (A). The details of Global Convolutional Network (GCN) and Boundary Refinement (BR) block are illustrated in (B) and (C), respectively.

Note

To address the contradictory aspects including classification and localization, GCN is proposed.

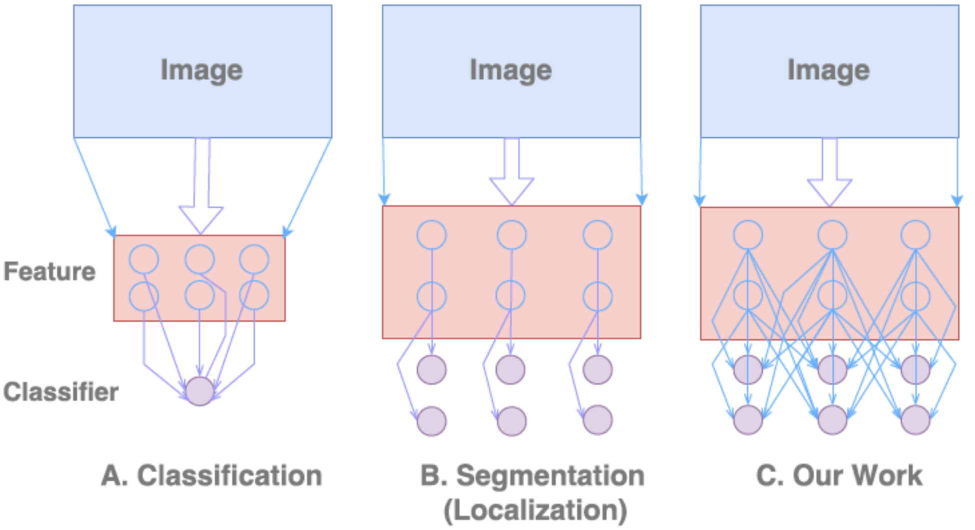


Figure 1. A: Classification network; B: Conventional segmentation network, mainly designed for localization; C: Our Global Convolutional Network.

Instead of directly using larger kernel or global convolution, our GCN module employs a combination of 1 x k + k x 1 and k x 1 + 1 x k convolutions, which enables densely connections within a large k x k region in the feature map.

And computation cost is lower.

As a comparison, our Global Convolution Network significantly enlarges the VRF

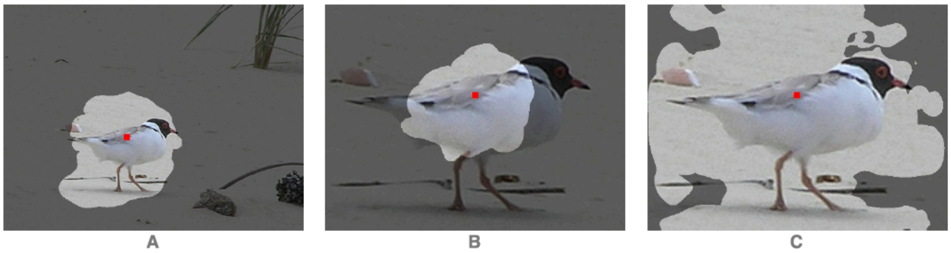


Figure 3. Visualization of *valid receptive field* (VRF) introduced by [38]. Regions on images show the VRF for the score map located at the center of the bird. For traditional segmentation model, even though the receptive field is as large as the input image, however, the VRF just covers the bird (A) and fails to hold the entire object if the input resized to a larger scale (B). As a comparison, our Global Convolution Network significantly enlarges the VRF (C).

Key Words

Five questions about this paper:

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

Classification and localization tasks are naturally contradictory.

For the classification task, the models are required to be invariant to various transformations like translation and rotation. But for the localization task, models should be transformation-sensitive to locate each pixel.

The conventional semantic segmentation algorithms mainly target for the localization issue, which might decrease the classification performance.

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

Global Convolutional Network (GCN) is proposed to deal with the above two challenges simultaneously.

- **from the localization view:** fully convolutional to retain the localization performance and no fully-connected or global pooling layers should be used as these layers will discard the localization information;
- **from the classification view:** large kernel size should be adopted in the network architecture to enable densely connections between feature maps and per-pixel classifiers, which enhances the capability to handle different transformations.

We introduce boundary refinement block to model the boundary alignment as a residual structure.

3. Details about the experiment

3.1 Which Datasets are used?

- PASCAL VOC 2012 (SBD)
- Cityscapes

3.2 How is the experiment set up?

3.3 What’s the evaluation metric?

3.4 Ablation Study

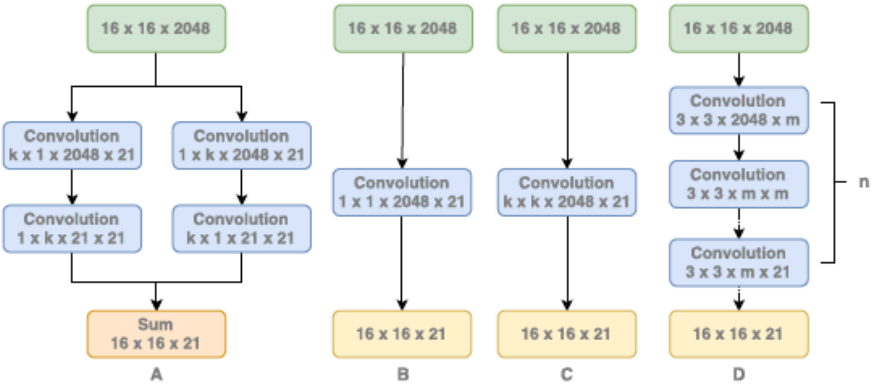
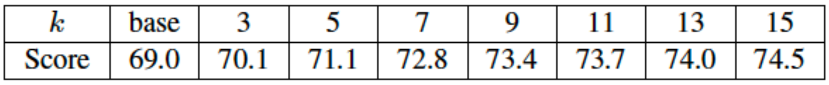

16 x 16 x 2048
Convolution k x 1 x 2048 x 21
Convolution 1 x k x 2048 x 21
Convolution 1 x k x 21 x 21
Convolution k x 1 x 21 x 21
Sum 16 x 16 x 21
A
16 x 16 x 2048
Convolution 1 x 1 x 2048 x 21
16 x 16 x 21
B
16 x 16 x 2048
Convolution k x k x 2048 x 21
16 x 16 x 21
C
16 x 16 x 2048
Convolution 3 x 3 x 2048 x m
Convolution 3 x 3 x m x m
Convolution 3 x 3 x m x 21
n
16 x 16 x 21
D

Figure 4. (A) Global Convolutional Network. (B) 1 × 1 convolution baseline. (C) $k \times k$ convolution. (D) stack of 3 × 3 convolutions.

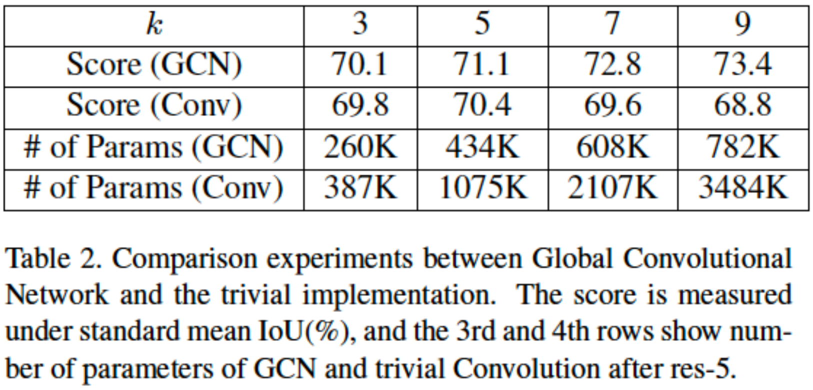
Global Convolutional Network — Large Kernel Matters



k	base	3	5	7	9	11	13	15
Score	69.0	70.1	71.1	72.8	73.4	73.7	74.0	74.5

Table 1. Experimental results on different k settings of Global Convolutional Network. The score is evaluated by standard mean IoU(%) on PASCAL VOC 2012 validation set.

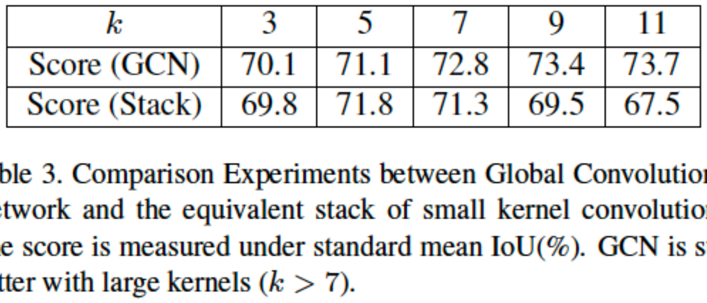
(1) Are more parameters helpful?



k	3	5	7	9
Score (GCN)	70.1	71.1	72.8	73.4
Score (Conv)	69.8	70.4	69.6	68.8
# of Params (GCN)	260K	434K	608K	782K
# of Params (Conv)	387K	1075K	2107K	3484K

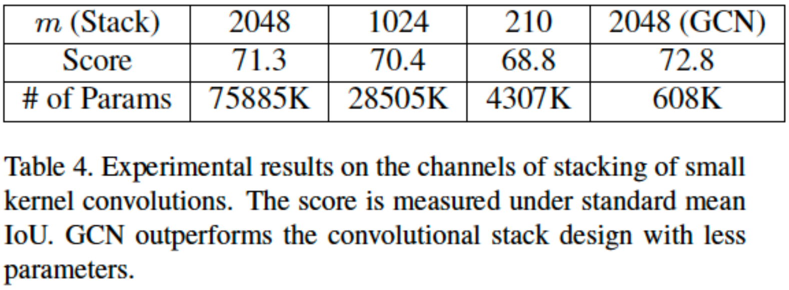
Table 2. Comparison experiments between Global Convolutional Network and the trivial implementation. The score is measured under standard mean IoU(%), and the 3rd and 4th rows show number of parameters of GCN and trivial Convolution after res-5.

(2) GCN vs. Stack of small convolutions



k	3	5	7	9	11
Score (GCN)	70.1	71.1	72.8	73.4	73.7
Score (Stack)	69.8	71.8	71.3	69.5	67.5

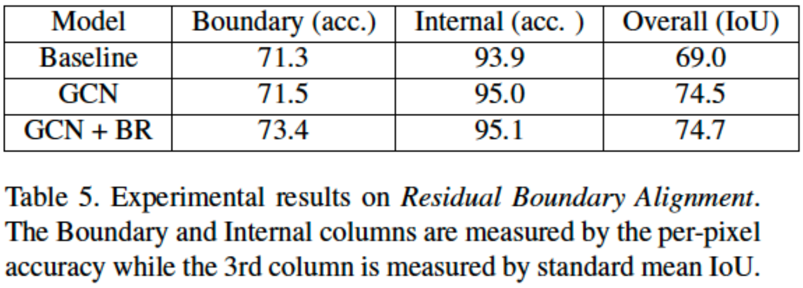
Table 3. Comparison Experiments between Global Convolutional Network and the equivalent stack of small kernel convolutions. The score is measured under standard mean IoU(%). GCN is still better with large kernels ($k > 7$).



m (Stack)	2048	1024	210	2048 (GCN)
Score	71.3	70.4	68.8	72.8
# of Params	75885K	28505K	4307K	608K

Table 4. Experimental results on the channels of stacking of small kernel convolutions. The score is measured under standard mean IoU. GCN outperforms the convolutional stack design with less parameters.

(3) How GCN contributes to the segmentation results?



Model	Boundary (acc.)	Internal (acc.)	Overall (IoU)
Baseline	71.3	93.9	69.0
GCN	71.5	95.0	74.5
GCN + BR	73.4	95.1	74.7

Table 5. Experimental results on *Residual Boundary Alignment*. The Boundary and Internal columns are measured by the per-pixel accuracy while the 3rd column is measured by standard mean IoU.

Global Convolutional Network for Pretrained Model

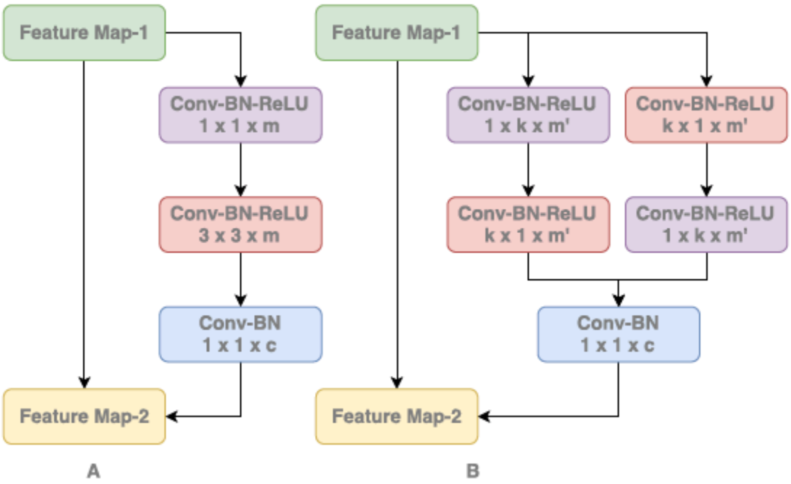
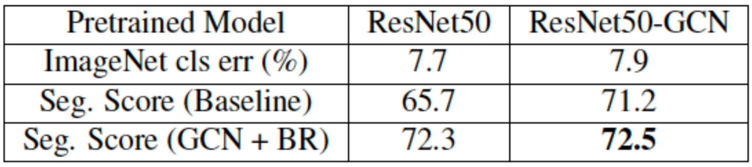


Figure 5. A: the bottleneck module in original ResNet. B: our *Global Convolutional Network* in ResNet-GCN.

Change the module in original ResNet to GCN. We can safely conclude that GCN mainly helps to improve segmentation performance, no matter in pretrained model or segmentation-specific structures.



Pretrained Model	ResNet50	ResNet50-GCN
ImageNet cls err (%)	7.7	7.9
Seg. Score (Baseline)	65.7	71.2
Seg. Score (GCN + BR)	72.3	**72.5**

Table 6. Experimental results on ResNet50 and ResNet50-GCN. Top-5 error of 224×224 center-crop on 256×256 image is used in ImageNet classification error. The segmentation score is measured under standard mean IoU.

3.5 What is the ranking of the experiment results?

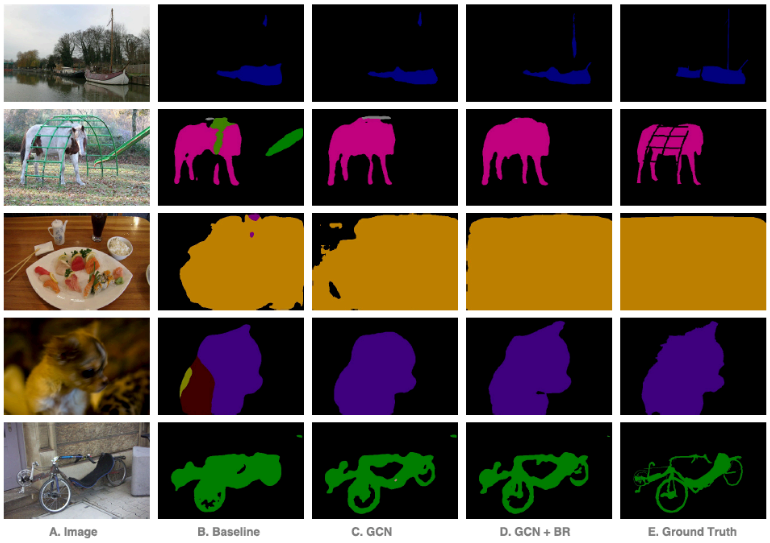
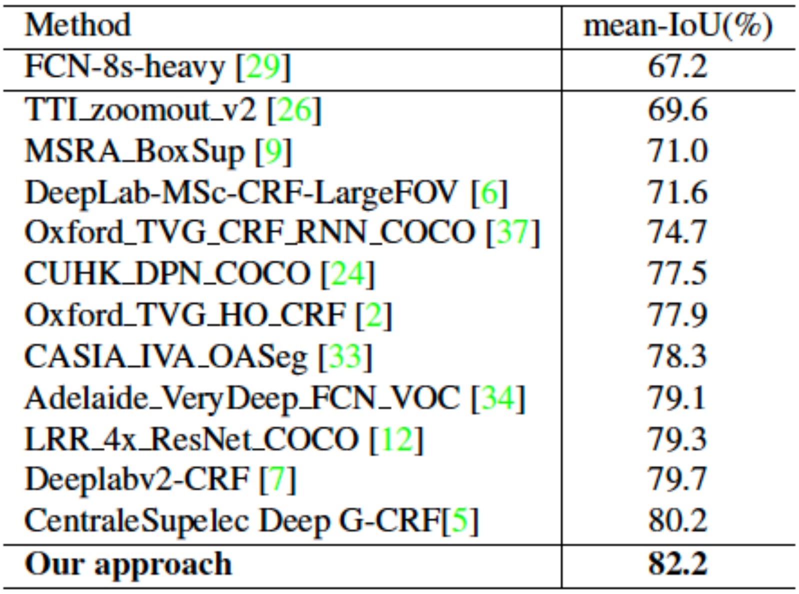
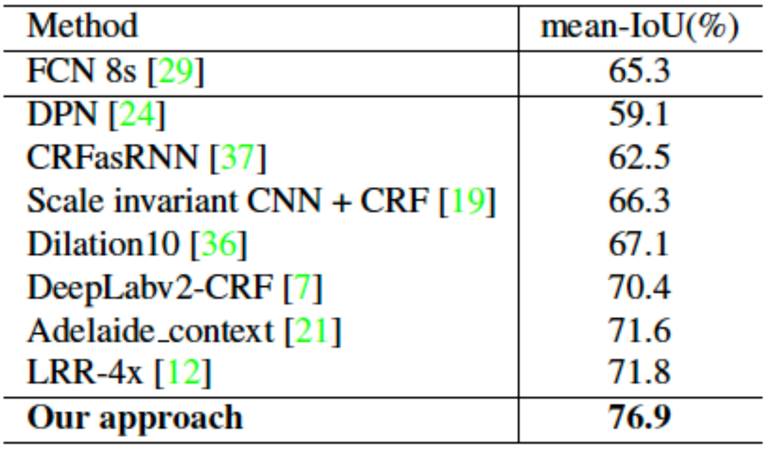


Figure 6. Examples of semantic segmentation results on PASCAL VOC 2012. For every row we list input image (A), 1 × 1 convolution baseline (B), Global Convolutional Network (GCN) (C), Global Convolutional Network plus Boundary Refinement (GCN + BR) (D), and Ground truth (E).



Method	mean-IoU(%)
FCN-8s-heavy [29]	67.2
TTI_zoomout_v2 [26]	69.6
MSRA_BoxSup [9]	71.0
DeepLab-MSc-CRF-LargeFOV [6]	71.6
Oxford_TVG_CRF_RNN_COCO [37]	74.7
CUHK_DPN_COCO [24]	77.5
Oxford_TVG_HO_CRF [2]	77.9
CASIA_IVA_OASeg [33]	78.3
Adelaide_VeryDeep_FCN_VOC [34]	79.1
LRR_4x_ResNet_COCO [12]	79.3
Deeplabv2-CRF [7]	79.7
CentraleSupelec Deep G-CRF[5]	80.2
Our approach	**82.2**

Table 8. Experimental results on PASCAL VOC 2012 test set.



Method	mean-IoU(%)
FCN 8s [29]	65.3
DPN [24]	59.1
CRFasRNN [37]	62.5
Scale invariant CNN + CRF [19]	66.3
Dilation10 [36]	67.1
DeepLabv2-CRF [7]	70.4
Adelaide_context [21]	71.6
LRR-4x [12]	71.8
Our approach	**76.9**

Table 10. Experimental results on Cityscapes test set.

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

1. The design of larger kernel and boundary refinement is interesting and useful.

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

1. I don’t agree with saying classification and localization are contradictory. Because localization is reached by classification originally, in the field of detection and segmentation. It’s helpful for telling story here, but not that big difference between them.