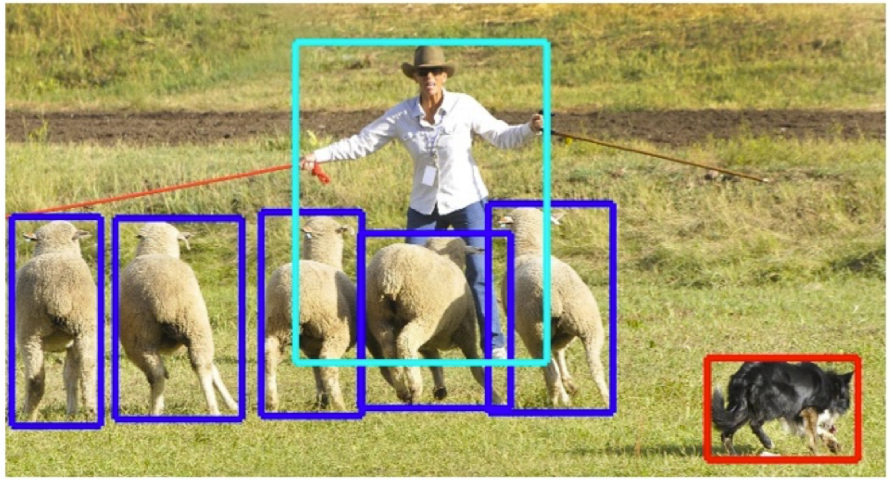
Large Kernel Matters：Improve Semantic Segmentation by Global Convolutional Network

Zhixin Piao

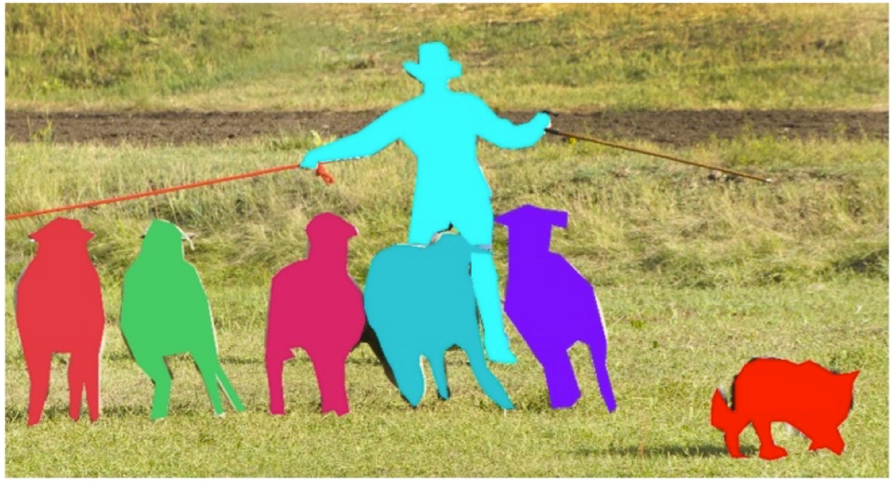
Semantic Segmentation

(a) Image classification



(b) **Image Detection**

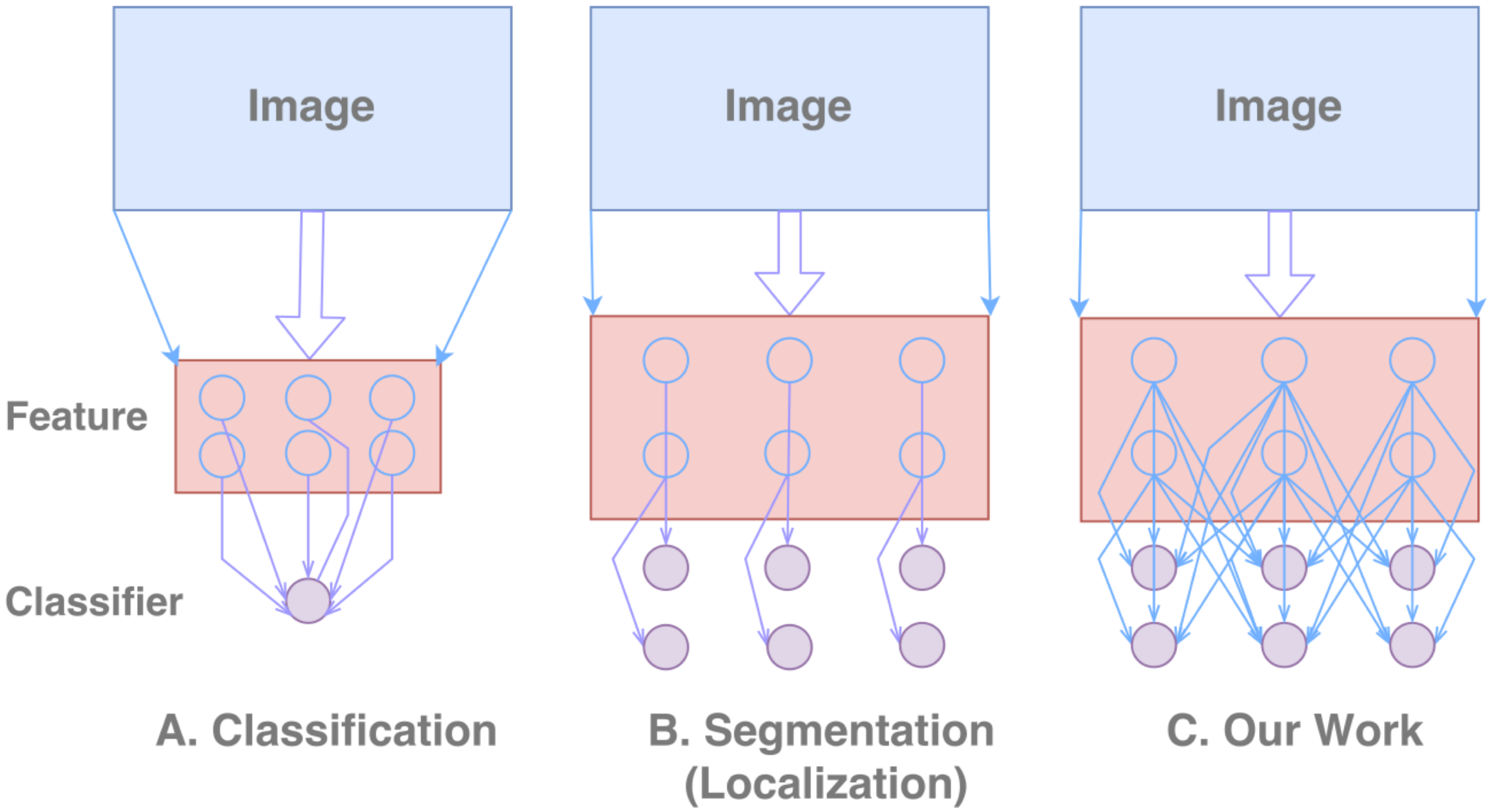
(c) Semantic segmentation



(d) **Instance Segmentation**

Motivation

- For the classification task, the models are required to be invariant to various transformations like translation and rotation.
- For the localization task, models should be transformation-sensitive, i.e., precisely locate every pixel for each semantic category.

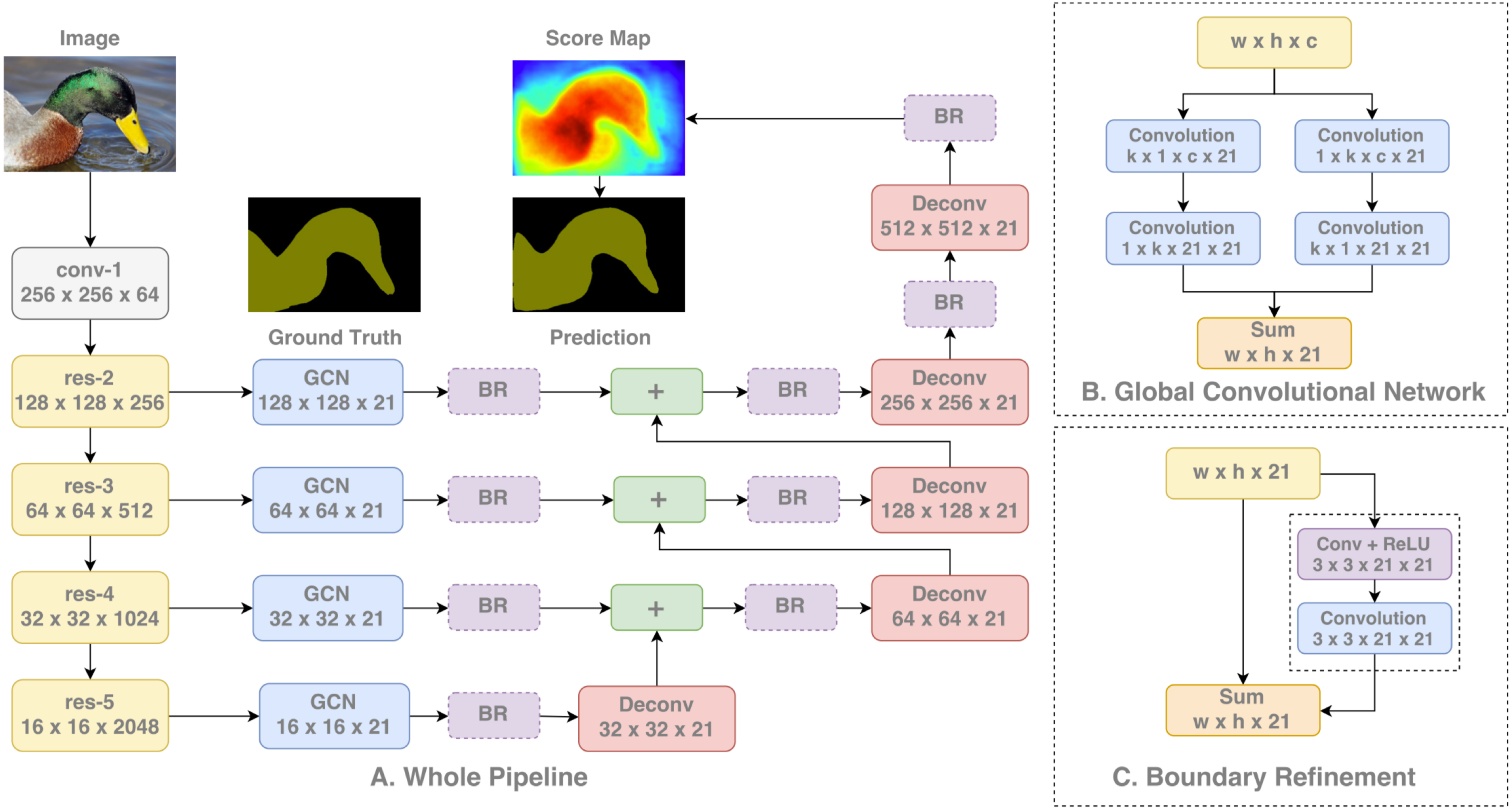


The conventional semantic segmentation algorithms mainly target for the localization issue,
But this might decrease the classification performance.

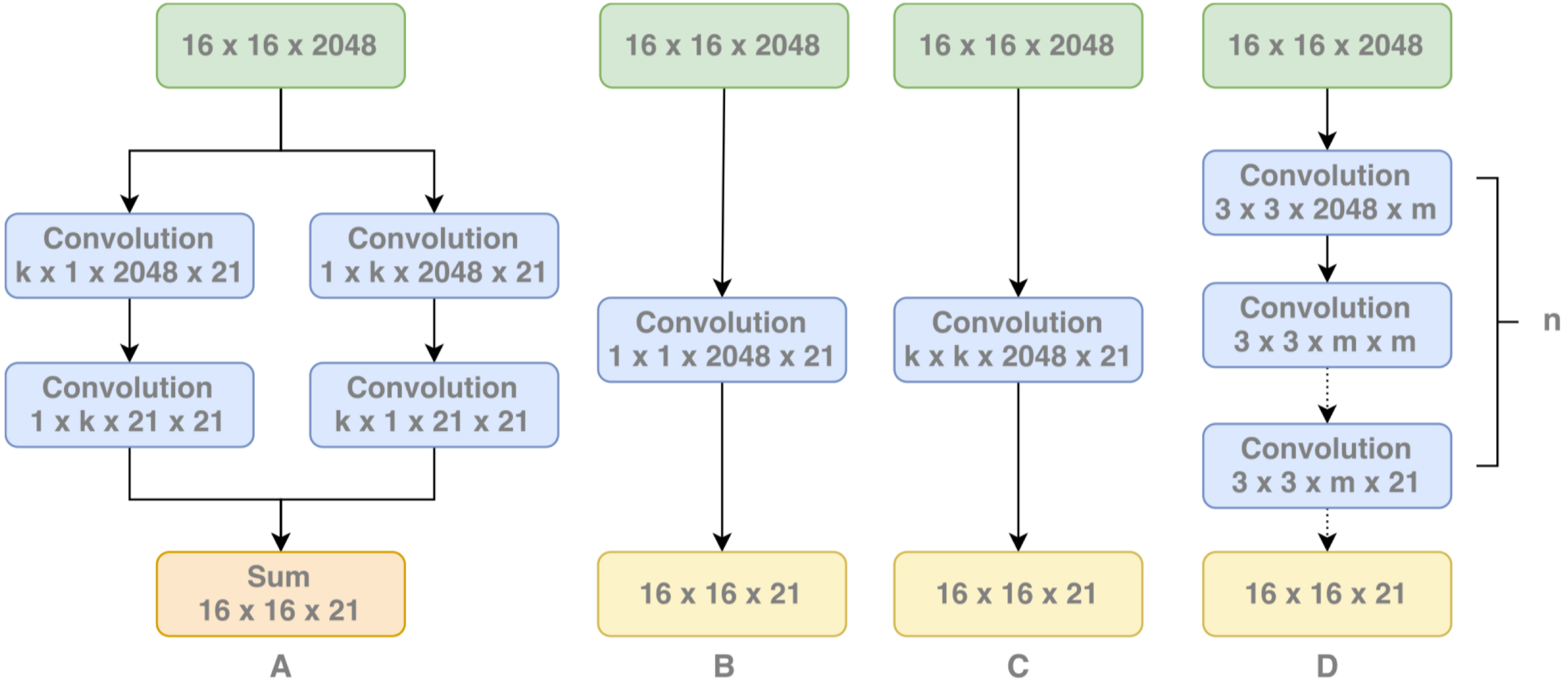
Large–Kernel–Matters (CVPR 2017) (8 Mar 2017)

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

Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, Jian Sun

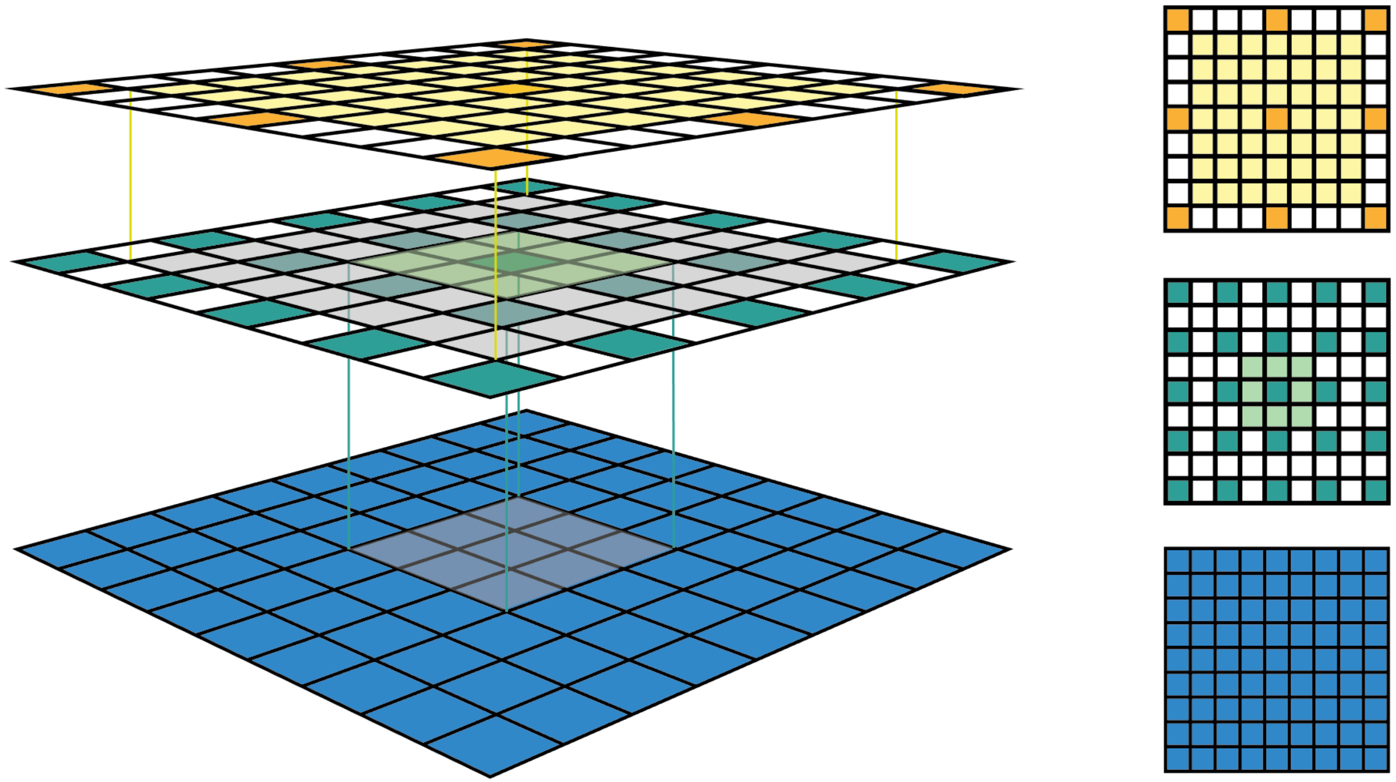


Global Convolutional Network (GCN)



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

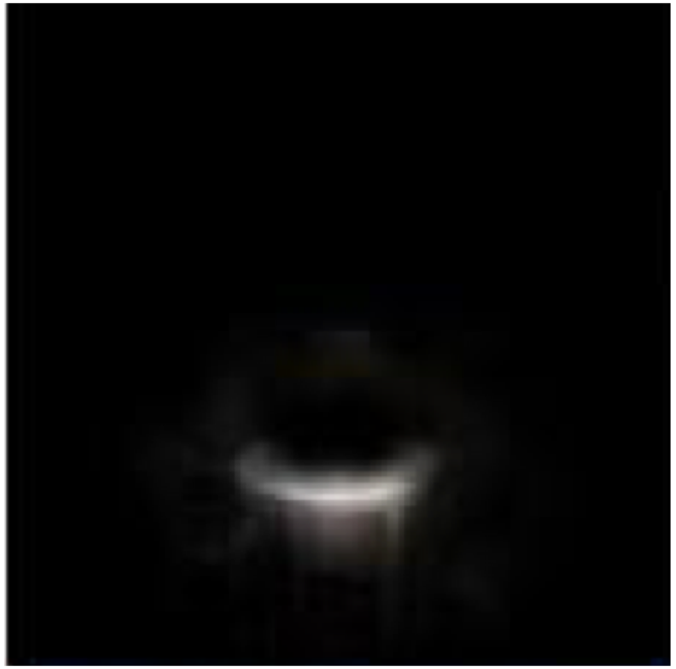
Receptive Field (RF)



Receptive Field of stacked 3x3 kernel

Valid Receptive Field (VRF)

- sliding-window stimuli contains a small randomized patch (example indicated by red arrow) at different spatial locations



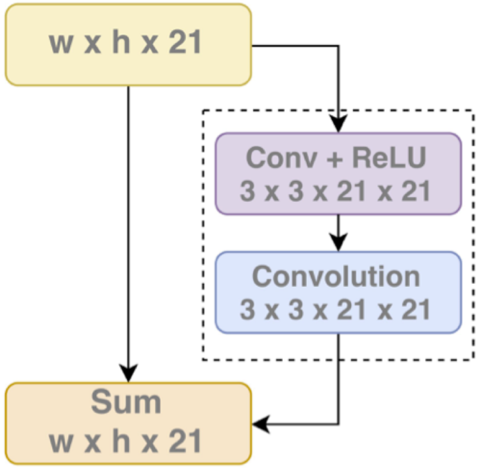
- comparing the activation response of the sliding-window stimuli with the activation response of the original image

Valid Receptive Field (VRF)

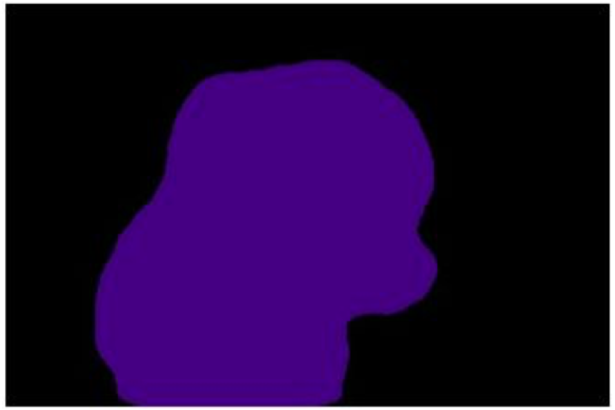
A B C

- 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, Global Convolution Network significantly enlarges the VRF (C).

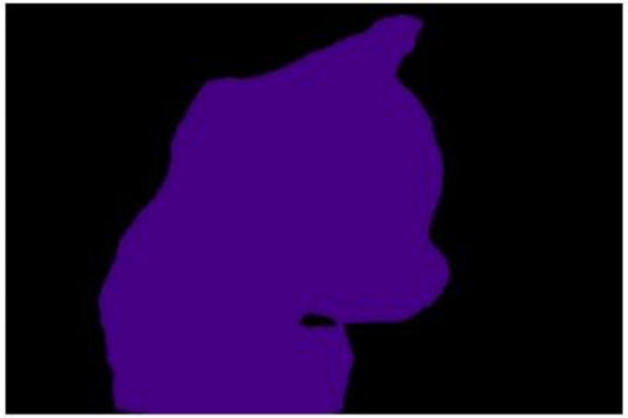
Boundary Refinement (BR)



Boundary Refinement (BR)



GCN



GCN + BR

Large-Kernel-Matters: Kernel Size

- **Large Kernel Size**

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.

Large–Kernel–Matters: Kernel Size

- **Compare with kxk conv**

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.

- **Compare score with stack**

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 ($k > 7$).

- **Compare Params with stack(k=7 3x3)**

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.

- **Compare without GCN**

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.

Large–Kernel–Matters: Result

Phase	Baseline	GCN	GCN + BR
Stage-1(%)	69.6	74.1	75.0
Stage-2(%)	72.4	77.6	78.6
Stage-3(%)	74.0	78.7	80.3
Stage-3-MS(%)			80.4
Stage-3-MS-CRF(%)			**81.0**

Table 7. Experimental results on PASCAL VOC 2012 validation set. The results are evaluated by standard mean IoU.

(1) In Stage-1, we mix up all the images from COCO, SBD and standard PASCAL VOC 2012, resulting in 109,892 images for training.

(2) During the Stage-2, we use the SBD and standard PASCAL VOC 2012 images, the same as Section 4.1.

(3) For Stage-3, we only use the standard PASCAL VOC 2012 dataset. The input image is padded to 640×640 in Stage-1 and 512×512 for Stage-2 and Stage-3.

Large–Kernel–Matters Summary

- Large Kernel
- Partially Solve the Receptive Field Problem
- Two key components (GCN and BRN)

Score:

PSPNet	85.4(82.6 in paper)	06-Dec-2016
Large-Kernel-Matters	83.6	16-Mar-2017
DeepLab-v2-CRF(ResNet-101)	79.7	06-Jun-2016
DeepLab-CRF-COCO-LargeFOV	72.7	18-Mar-2015
FCN-8s	62.2	12-Nov-2014
SegNet	59.9	10-Nov-2015