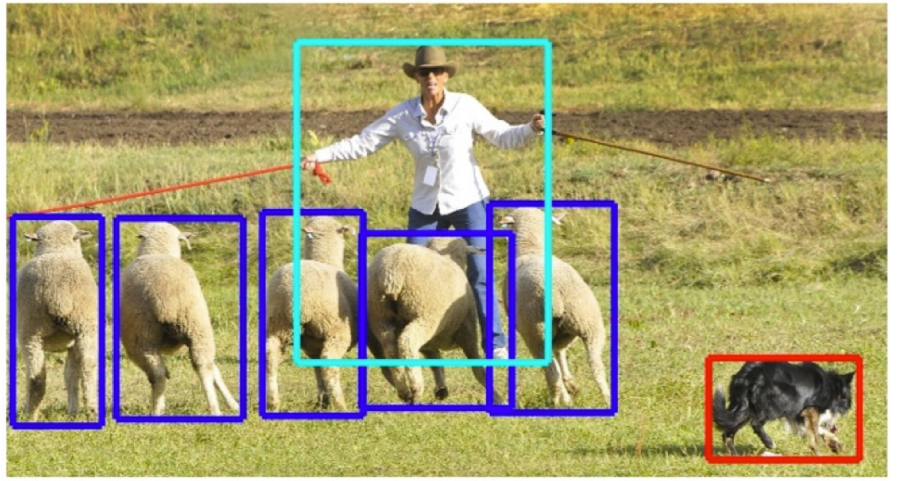
Semantic Segmentation: From FCN to DeepLab v3+ (I)

Zhixin Piao

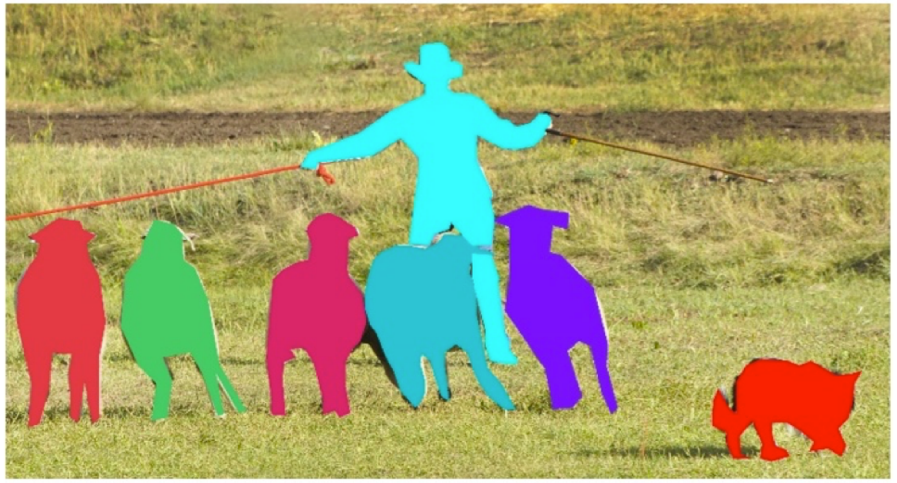
Semantic Segmentation

(a) Image classification



(b) Object localization

(c) Semantic segmentation



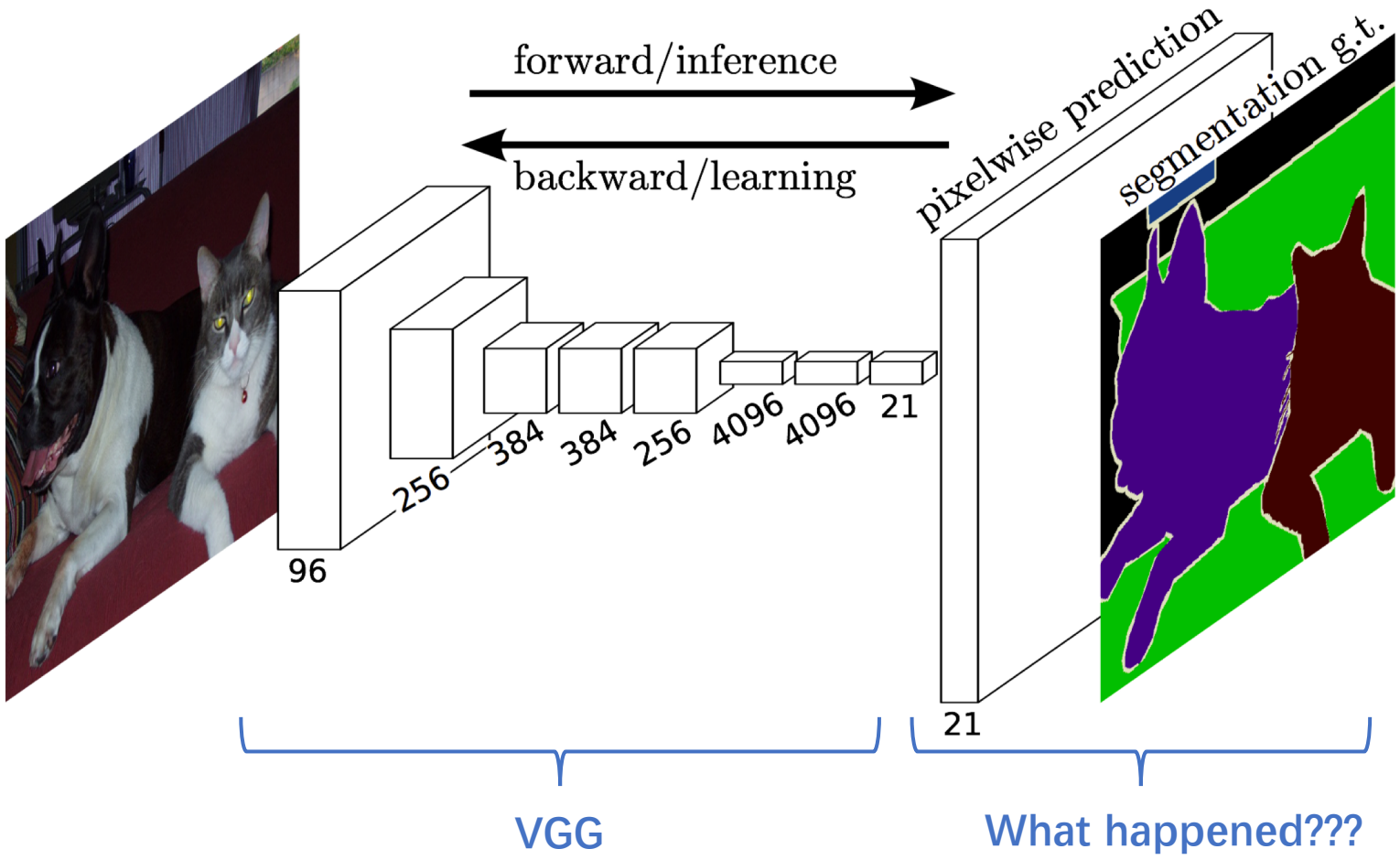
(d) **Instance Segmentation**

Outline

- FCN
- SegNet
- U-Net
- DeepLab v1/v2
- PSPNet
- Large-Kernel-Matters
- DeepLab v3/v3+

Fully Convolutional Networks for Semantic Segmentation

Jonathan Long, Evan Shelhamer, Trevor Darrell



In-Net Upsampling: Transposed convolution

- **I heard that it is called deconvolution...**

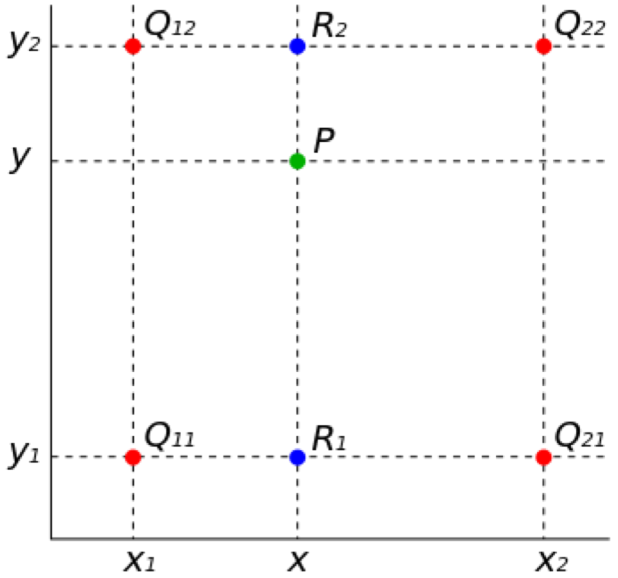
  It has many names: **Transposed convolution**, **Deconvolution**(**not recommand**), **Upconvolution, Fractionally stride convolution, Backward convolution...**

  But deconvolution is not accurate!

- **Why called In-Net Upsampling?**

  Because the parameters of these kind of unsamling are **learnable**.

  We also have **fixed upsampling**, e.g. bilinear upsampling.



bilinear upsampling

Why called Transposed convolution?

We can express convolution in terms of a matrix multiplication

$$\vec{x} * \vec{a} = X\vec{a}$$

$$\begin{bmatrix} x & y & x & 0 & 0 & 0 \\ 0 & x & y & x & 0 & 0 \\ 0 & 0 & x & y & x & 0 \\ 0 & 0 & 0 & x & y & x \end{bmatrix} \begin{bmatrix} 0 \\ a \\ b \\ c \\ d \\ 0 \end{bmatrix} = \begin{bmatrix} ay + bz \\ ax + by + cz \\ bx + cy + dz \\ cx + dy \end{bmatrix}$$

Example: 1D conv, kernel size=3, stride=1, padding=1

Convolution transpose multiplies by the transpose of the same matrix:

$$\vec{x} *^T \vec{a} = X^T \vec{a}$$

$$\begin{bmatrix} x & 0 & 0 & 0 \\ y & x & 0 & 0 \\ z & y & x & 0 \\ 0 & z & y & x \\ 0 & 0 & z & y \\ 0 & 0 & 0 & z \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} ax \\ ay + bx \\ az + by + cx \\ bz + cy + dx \\ cz + dy \\ dz \end{bmatrix}$$

When stride=1, convolution transpose is just a regular convolution (with different padding rules)

Why called Transposed convolution?

We can express convolution in terms of a matrix multiplication

$$\vec{x} * \vec{a} = X\vec{a}$$

$$\begin{bmatrix} x & y & z & 0 & 0 & 0 \\ 0 & 0 & x & y & z & 0 \end{bmatrix} \begin{bmatrix} 0 \\ a \\ b \\ c \\ d \\ 0 \end{bmatrix} = \begin{bmatrix} ay + bz \\ bx + cy + dz \end{bmatrix}$$

Example: 1D conv, kernel size=3, stride=2, padding=1

Convolution transpose multiplies by the transpose of the same matrix:

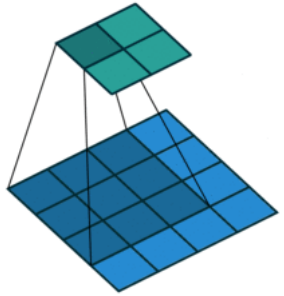
$$\vec{x} *^T \vec{a} = X^T \vec{a}$$

$$\begin{bmatrix} x & 0 \\ y & 0 \\ z & x \\ 0 & y \\ 0 & z \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} ax \\ ay \\ az + bx \\ by \\ bz \\ 0 \end{bmatrix}$$

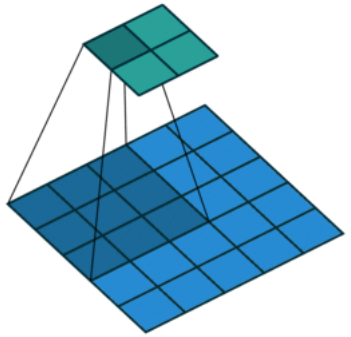
When stride>1, convolution transpose is no longer a normal convolution!

Why called Transposed convolution?

Convolution:

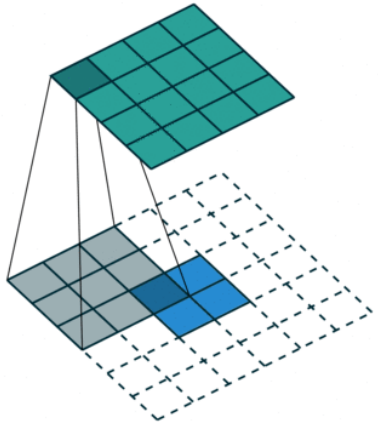


No padding, no strides

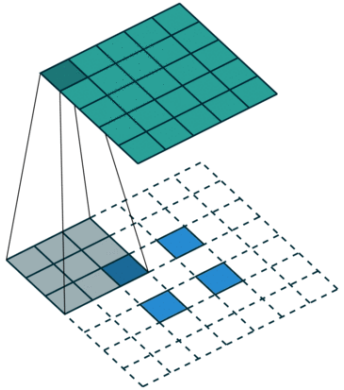


No padding, strides

Transposed Convolution:



No padding, no strides



No padding, strides

Why not called Deconvolution?

Convolution

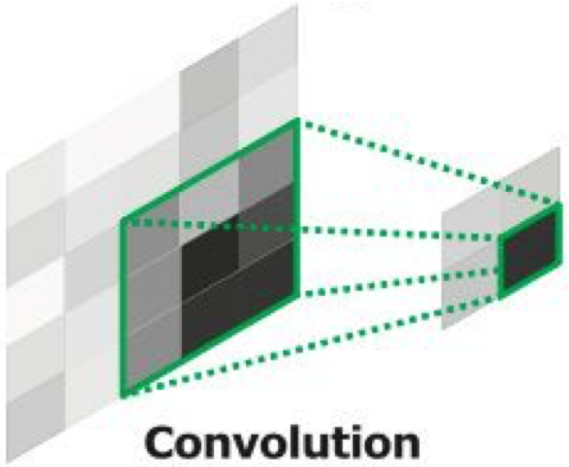
$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} * \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 37 & 47 \\ 67 & 77 \end{pmatrix}$$

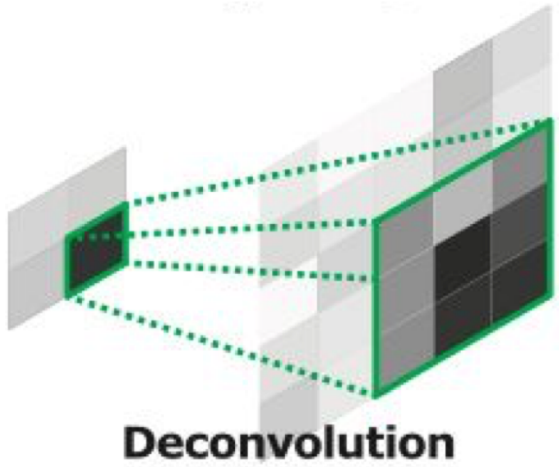
“Deconvolution”

$$\begin{pmatrix} 37 & 47 \\ 67 & 77 \end{pmatrix} *^{T} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 148 & 299 & 141 \\ 342 & 640 & 278 \\ 134 & 221 & 77 \end{pmatrix}$$

It should recover convolution result to input matrix If it is deconvolution!

Why called backward convolution?



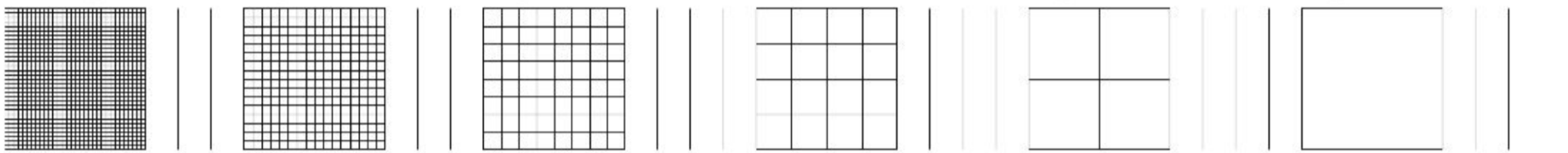


Very similar with gradient back propagation...

- **Tansposed convlution can be implemented very simply:**

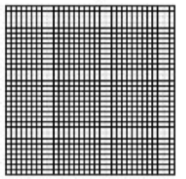
  just **exchange** **forward** function and **backward** function

Skip–Layer



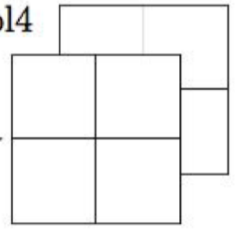
transposed conv

32x upsampled prediction (FCN-32s)

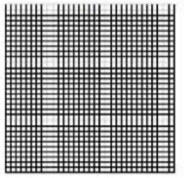


2x conv7

pool4



16x upsampled prediction (FCN-16s)

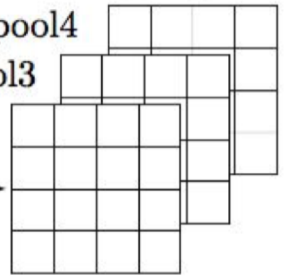


need 1x1 conv to make each layer have same channel size(1)

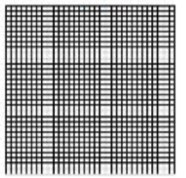
4x conv7

2x pool4

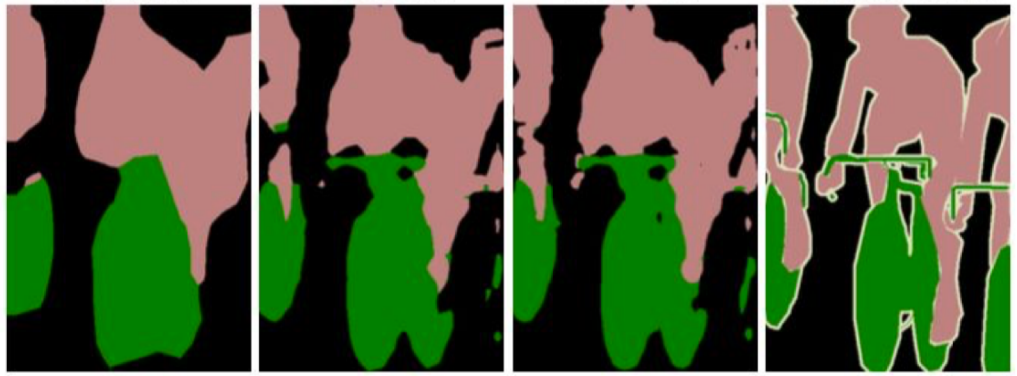
pool3



8x upsampled prediction (FCN-8s)



FCN-32s FCN-16s FCN-8s Ground truth



FCN Summary

- First work using CNN to solve the semantic segmentation
- Introducing skip-net framework
- Large Improvement! (60 vs 30)

Score:

FCN-8s	62.2	12-Nov-14

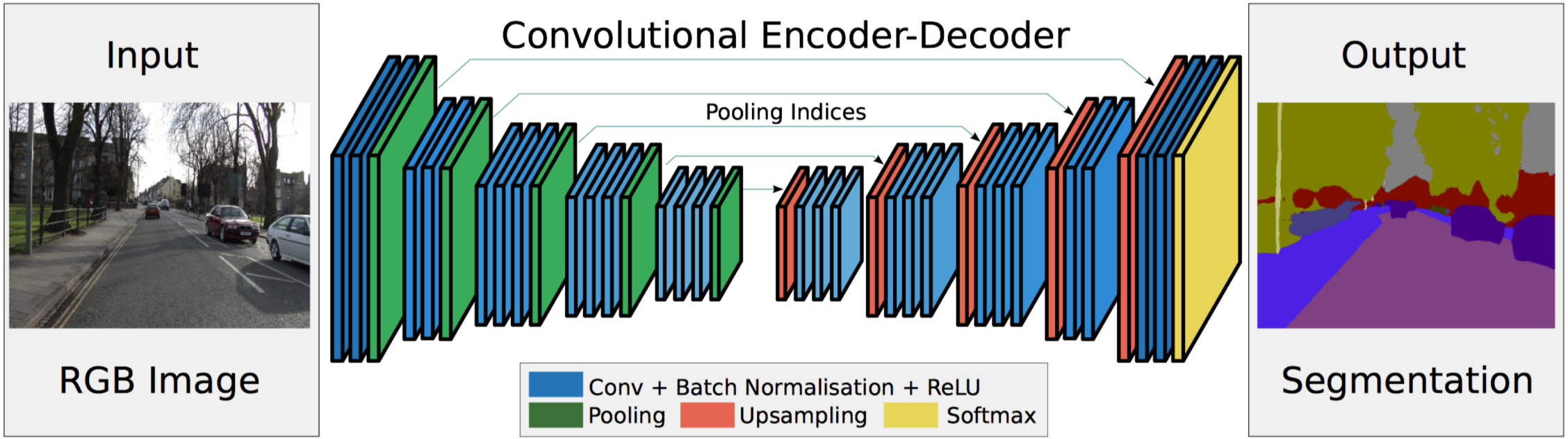
Outline

- FCN
- SegNet
- U-Net
- DeepLab v1/v2
- PSPNet
- Large-Kernel-Matters
- DeepLab v3/v3+

SegNet (TPAMI 2017) (2 Nov 2015)

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla



Unpooling

Max Pooling
Remember which element was max!

1	2	6	3
3	5	2	1
1	2	2	1
7	3	4	8

Input: 4 x 4

5	6
7	8

Output: 2 x 2

...

Rest of the network

Max Unpooling
Use positions from pooling layer

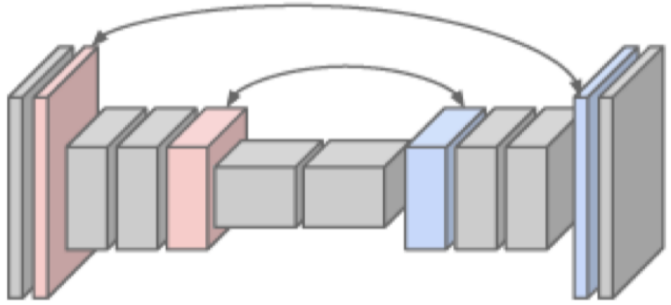
1	2
3	4

Input: 2 x 2

0	0	2	0
0	1	0	0
0	0	0	0
3	0	0	4

Output: 4 x 4

Corresponding pairs of downsampling and upsampling layers



SegNet Summary

- Unpooling
- Encoder-Decoder Model

Score:

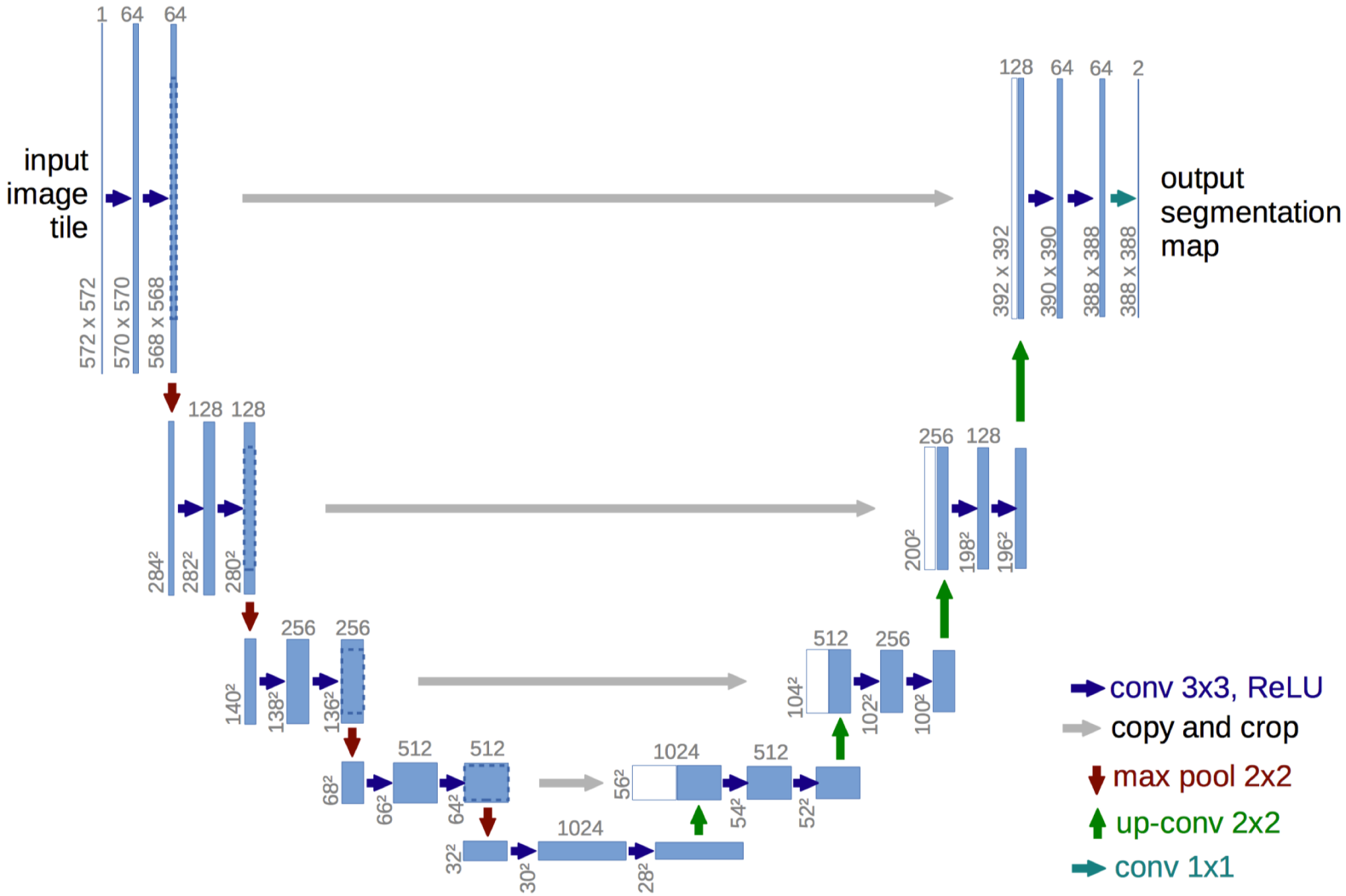
FCN-8s	**62.2**	**12-Nov-14**
SegNet	59.9	10-Nov-2015

Outline

- FCN
- SegNet
- U-Net
- DeepLab v1/v2
- PSPNet
- Large-Kernel-Matters
- DeepLab v3/v3+

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, Thomas Brox



U-Net Summary

- Concat feature map of encoder and decoder

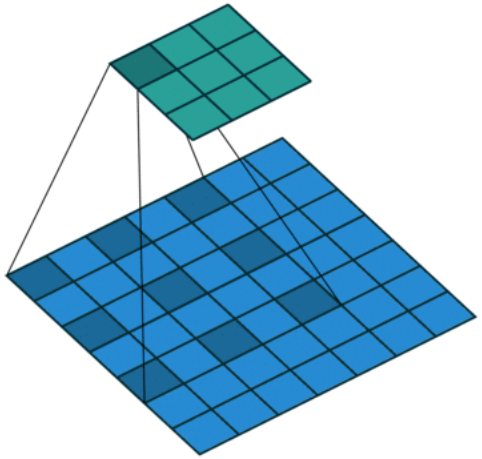
Outline

- FCN
- SegNet
- U-Net
- DeepLab v1/v2
- PSPNet
- Large-Kernel-Matters
- DeepLab v3/v3+

DeepLab v1 (ICLR 2015)(22 Dec 2014)

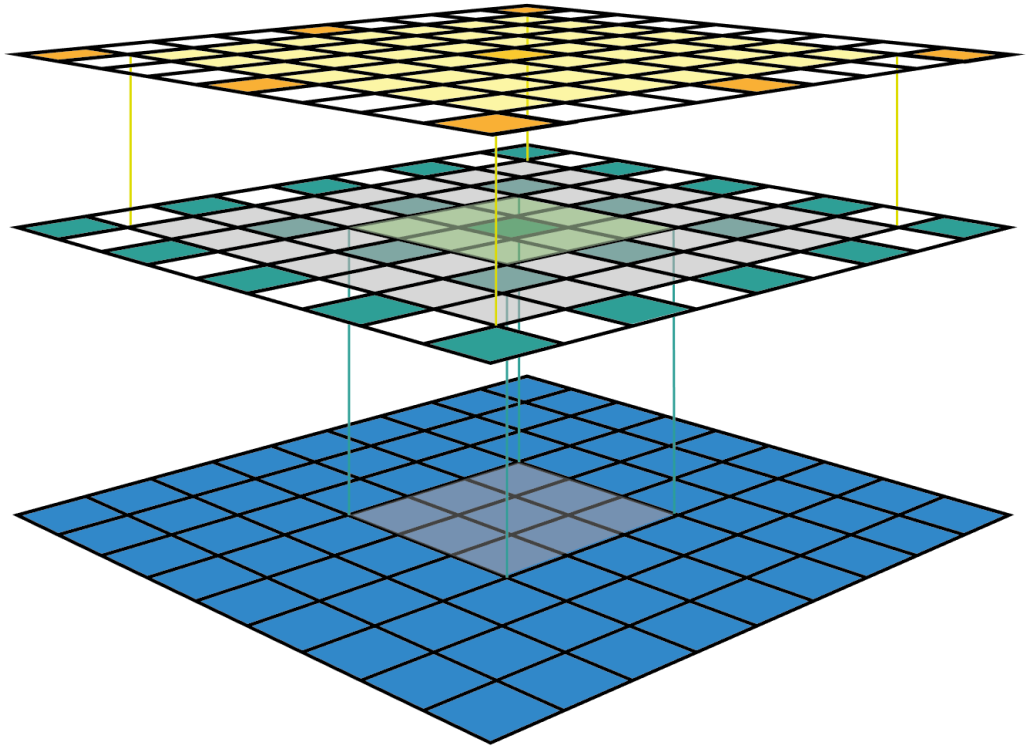
Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs

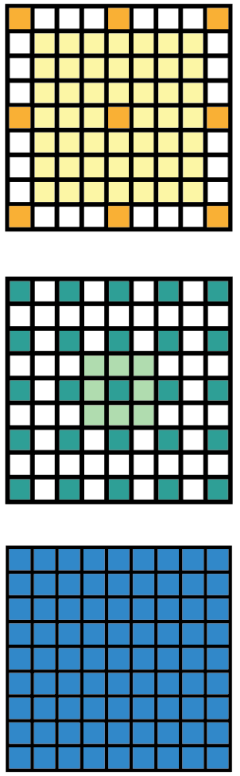
Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille



Dilated / Atrous Convolution

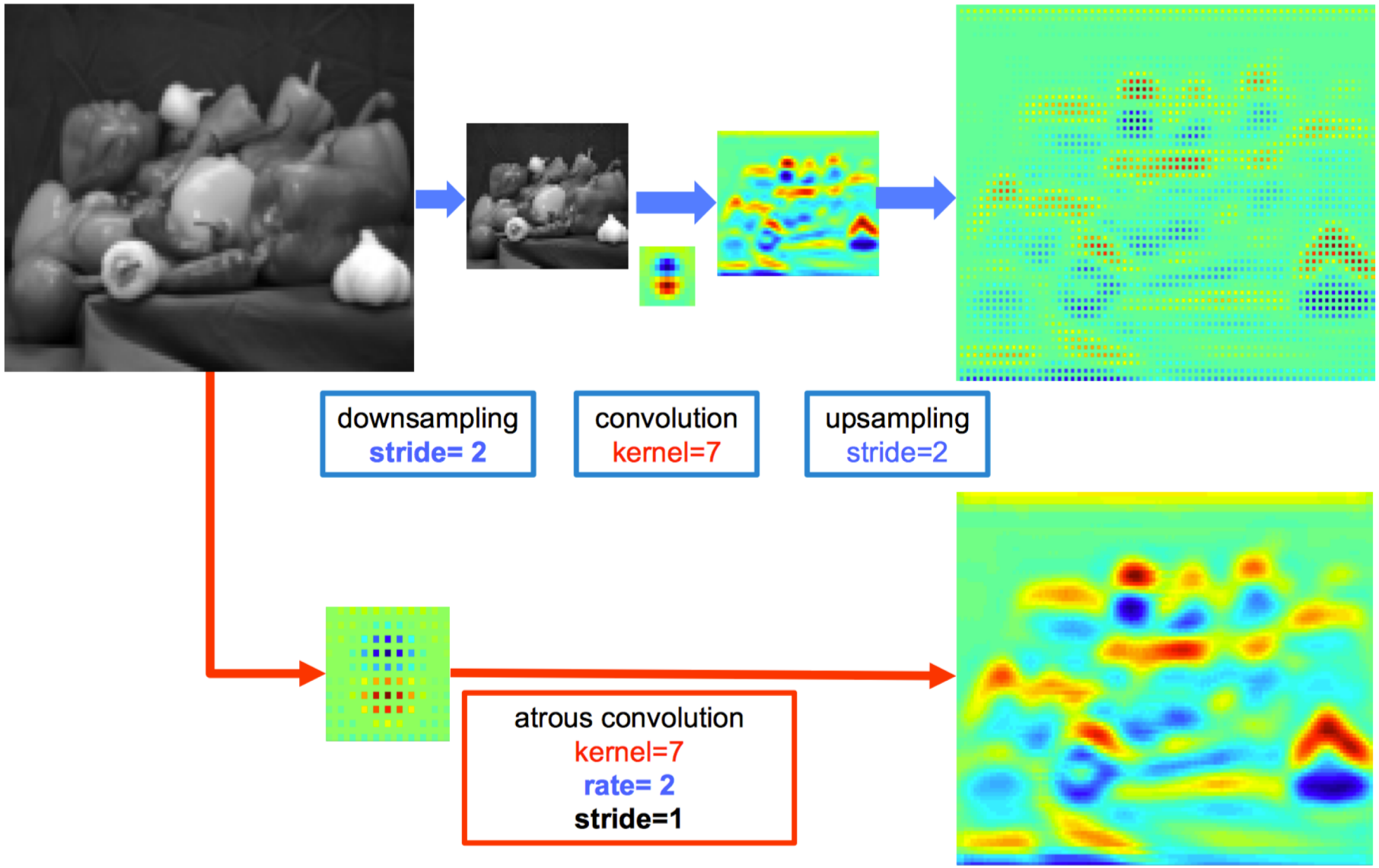
Dilation=2, Kernel_size=3





Field Of View

Dilated Convolution



DeepLab v1 Summary

- Dilated Convolution
- Use CRF to refine

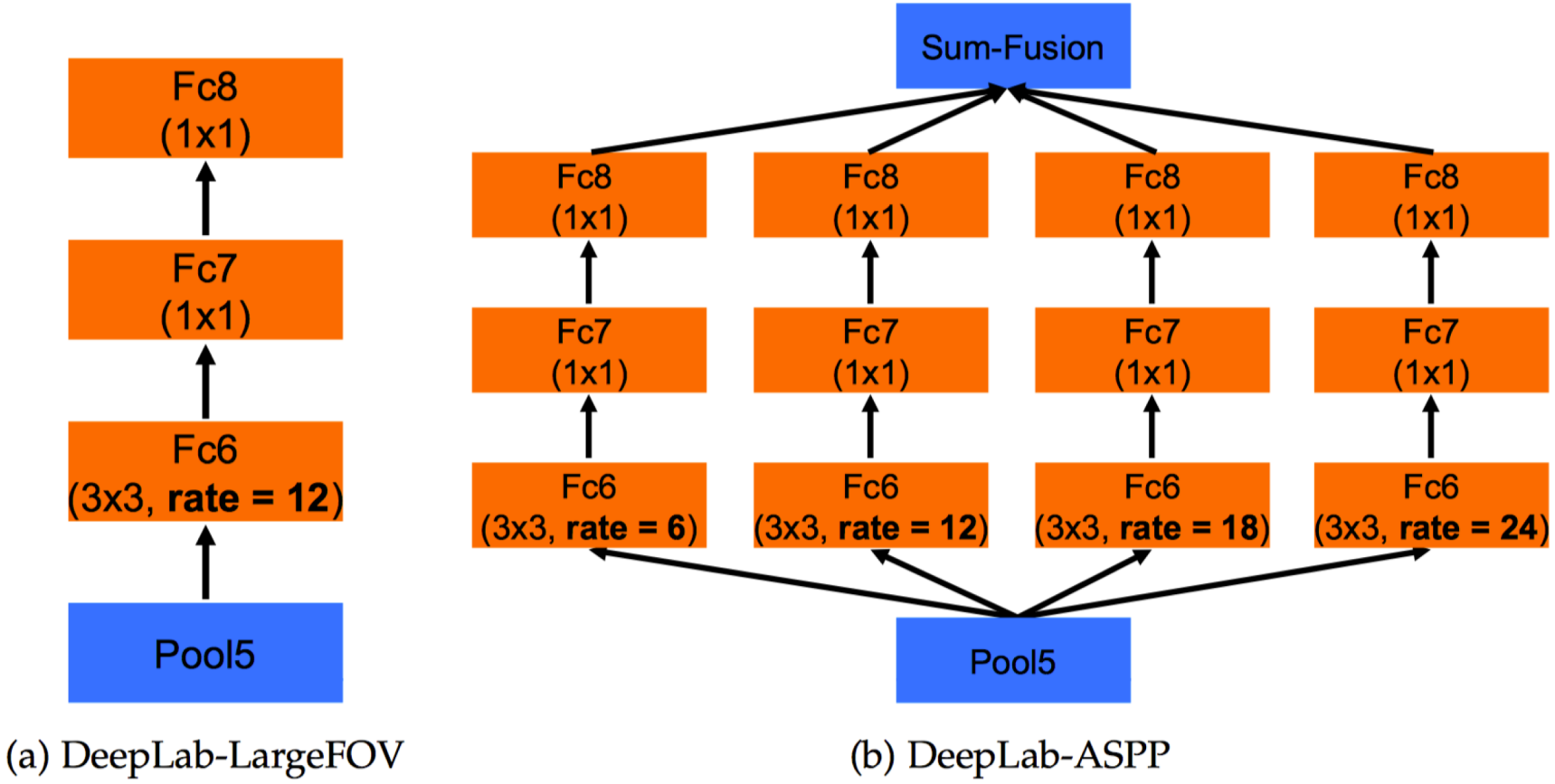
Score:

DeepLab-CRF-COCO-LargeFOV	72.7	18-Mar-2015
FCN-8s	62.2	12-Nov-2014
SegNet	59.9	10-Nov-2015

DeepLab v2 (TPAMI 2017) (2 Jun 2016)

DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille



(a) DeepLab-LargeFOV (b) DeepLab-ASPP

atrous spatial pyramid pooling

DeepLab v2 Summary

- ASPP

Score:

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

Score > 80

- Need more boundary details
- Need more attention about small object