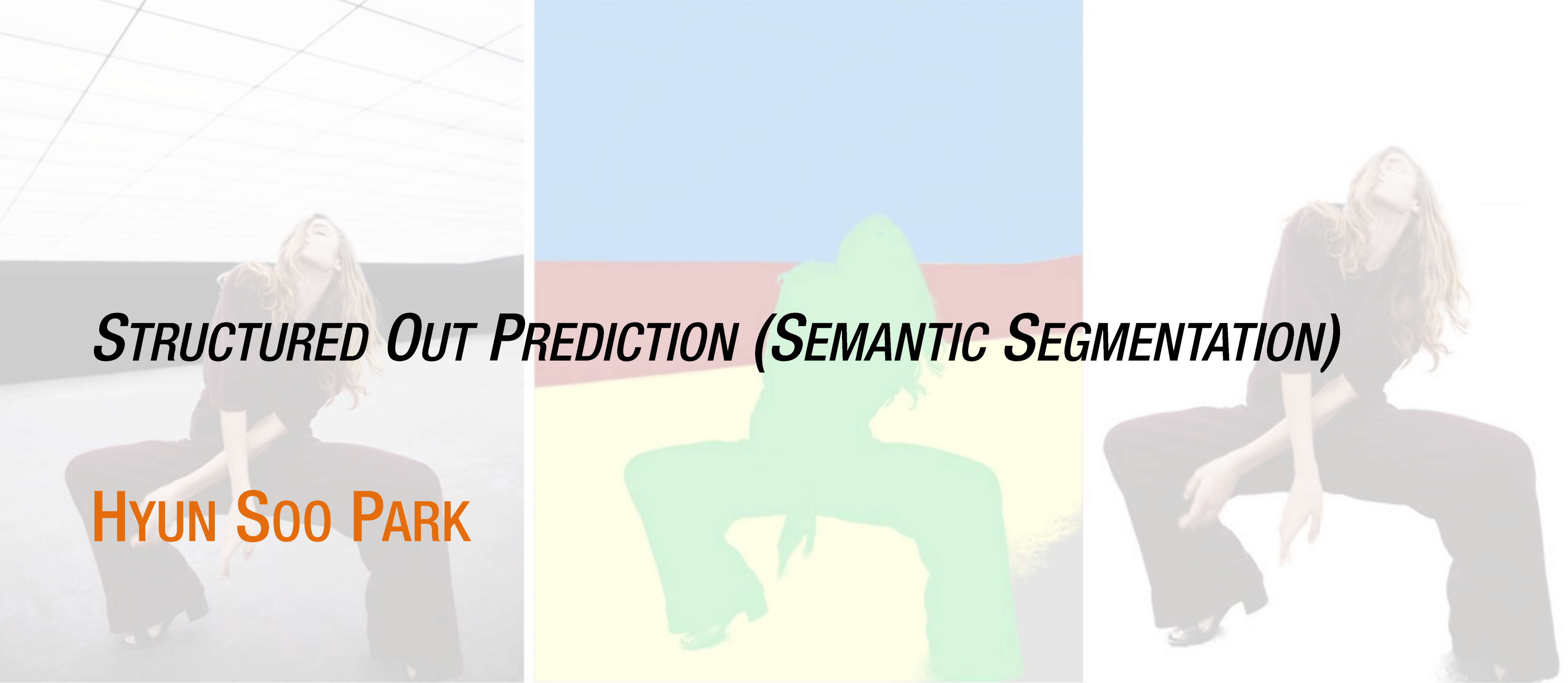


Structured Out Prediction (Semantic Segmentation)

Hyun Soo Park

CHALLENGES OF VISUAL RECOGNITION

- Appearance
    - DOF: texture, illumniation, material, shading, …
- Shape
    - DOF: object category, geometric pose, viewpoint, …

CHALLENGES OF VISUAL RECOGNITION

- Appearance
  - DOF: texture, illumniation, material, shading, …
- Shape
  - DOF: object category, geometric pose, viewpoint, …

Human

$$f(I) = l_{human}$$

Semantic segmentation

Semantic Segmentation: Pixel Classification

- 0: Background/Unknown
- 1: Person
- 2: Purse
- 3: Plants/Grass
- 4: Sidewalk
- 5: Building/Structures

Semantic Segmentation Formulation

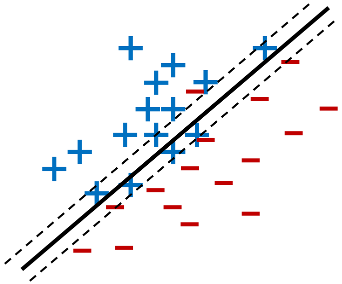
Unsupervised superpixel segmentation

x_i

Visual feature

- Color histogram
- BoW
- SIFT
- HOG

Classification



$L = \phi(x_i)$ e.g., $\phi(x_i) = w_{tree} \cdot x_i$

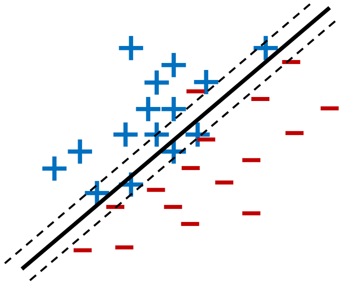
Semantic Segmentation Formulation

Unsupervised superpixel segmentation

Visual feature

- Color histogram
- BoW
- SIFT
- HOG

Classification

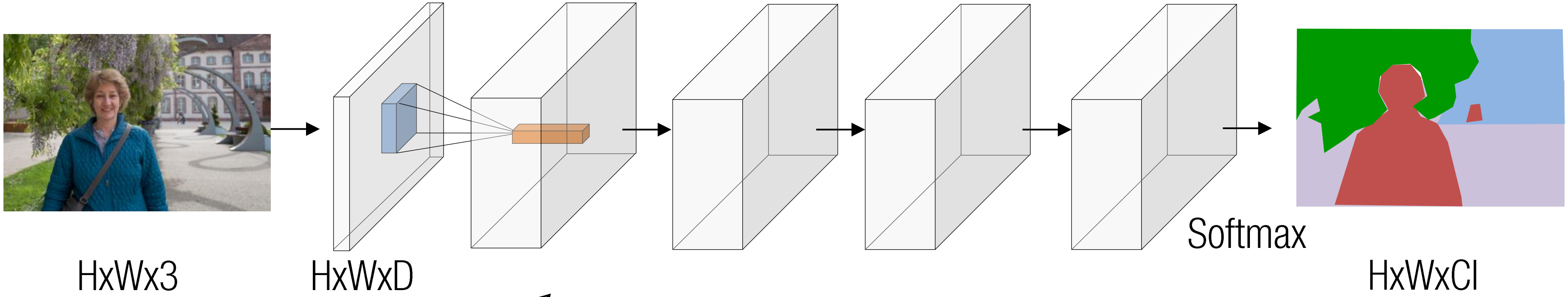


$$L = \phi(x_i) + \underbrace{\varphi(x_i, x_j)}_{\text{Context}}$$

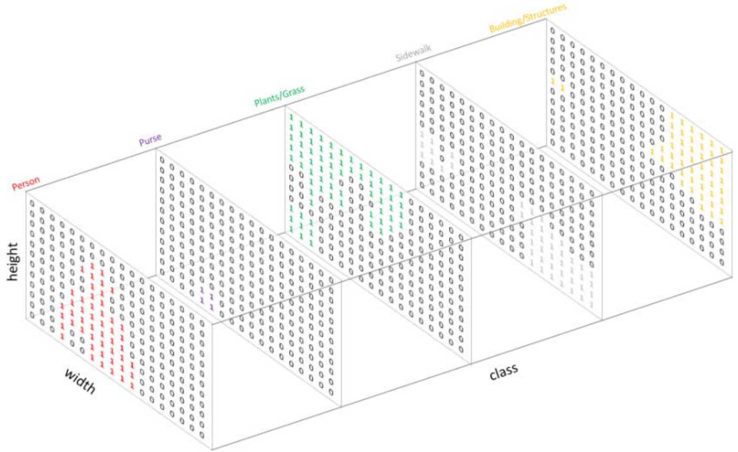
CRF: Conditional Random Field

aka. joint classifications (structured pred.)

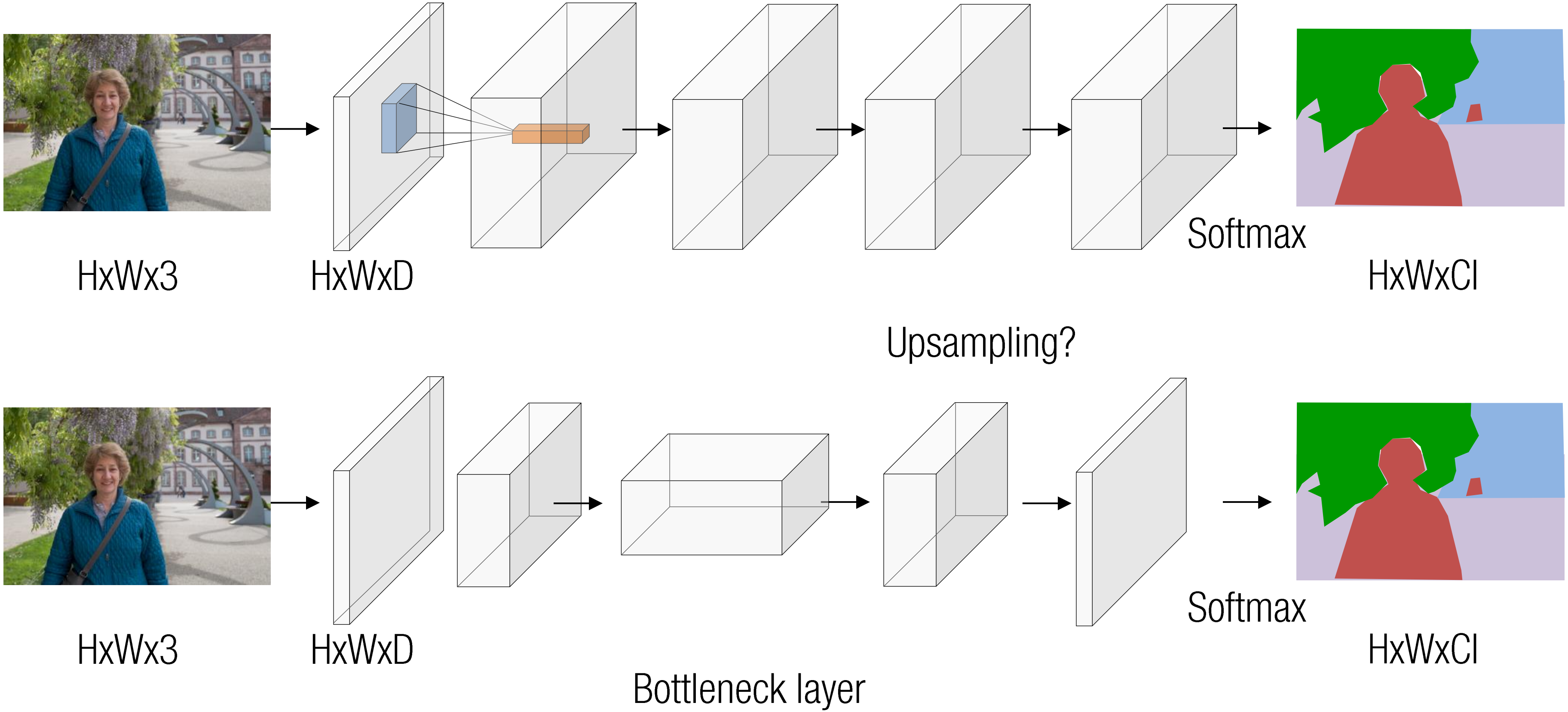
Holistic Prediction via Deep Learning



Computationally inefficient



Holistic Prediction via Deep Learning



Revisited: Spatial Pooling (Down-sampling)

3	2	3	1
0	**5**	3	**4**
1	2	2	2
7	3	1	**7**

4×4

5	**4**
7	**7**

2×2

Max-pooling (window size 2x2, stride 2)

SPATIAL UNPOOLING (UP-SAMPLING)

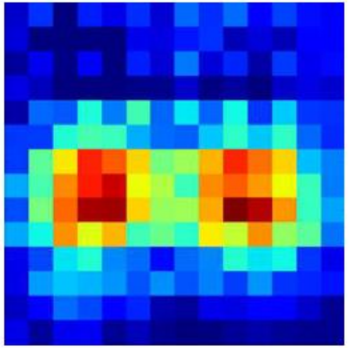
Max-**un**pooling (window size 2x2, stride 2)

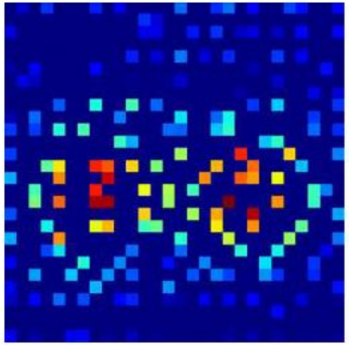
5	**4**
7	**7**

0	0	0	0
0	**5**	0	**4**
0	0	0	0
7	0	0	**7**

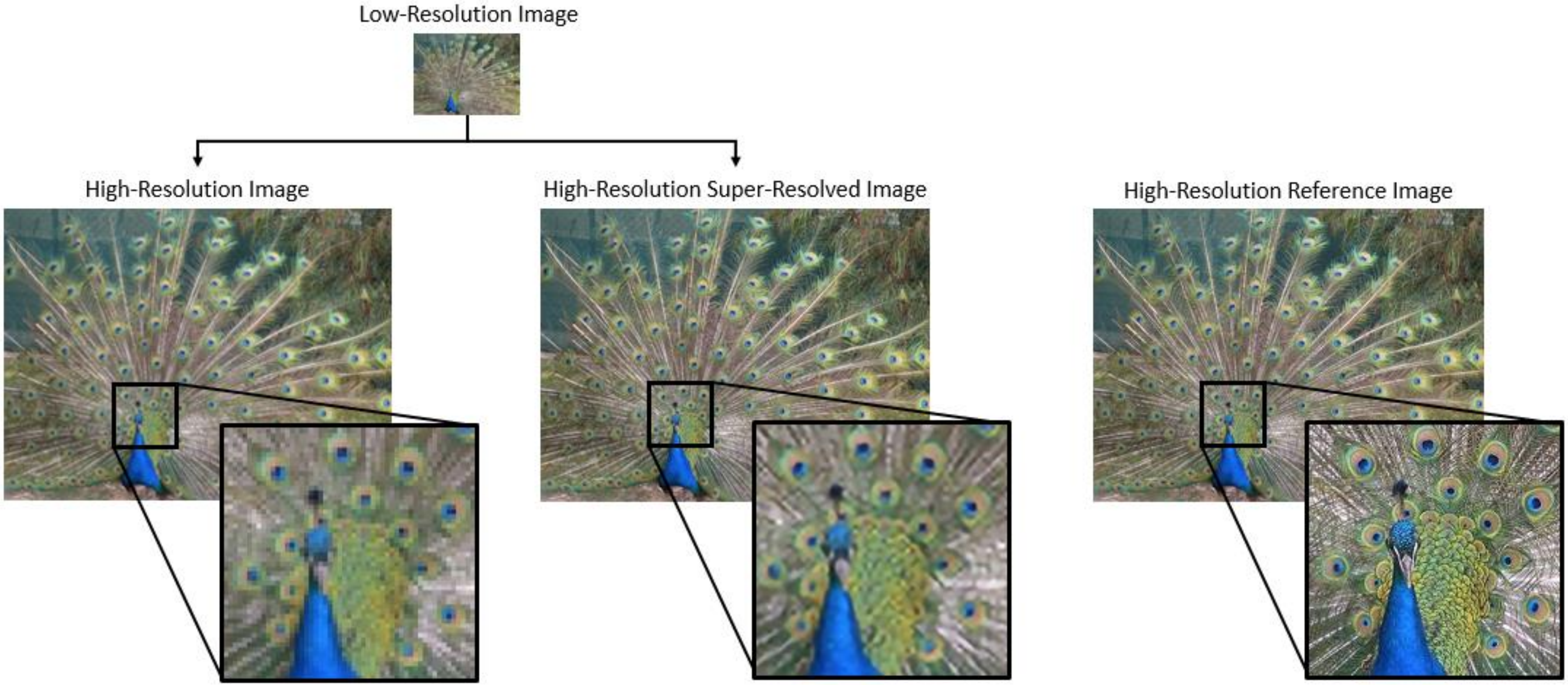
Learnable parameter?

Can we learn upsampling?





LEARNING UPSAMPLING



SPATIAL UNPOOLING (UP-SAMPLING)

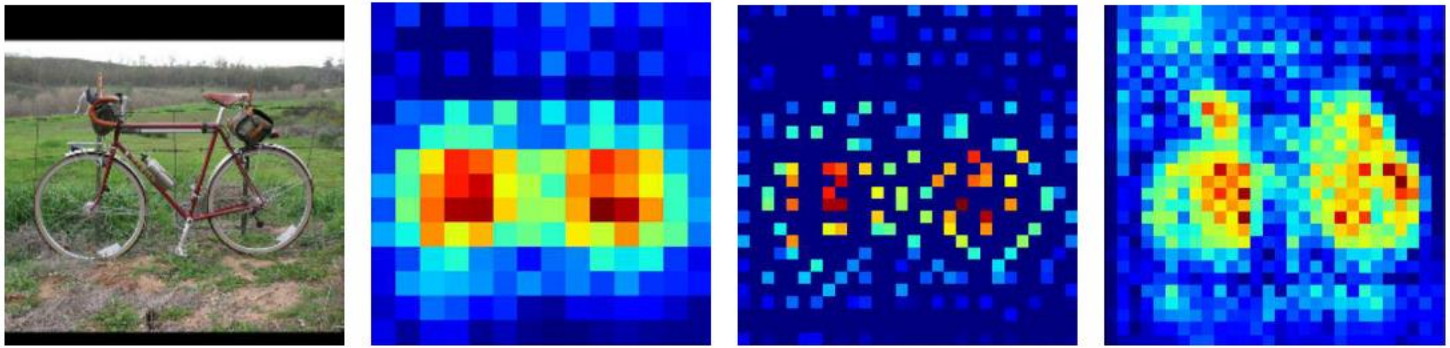
Max-**un**pooling (window size 2x2, stride 2)

5	**4**
7	**7**

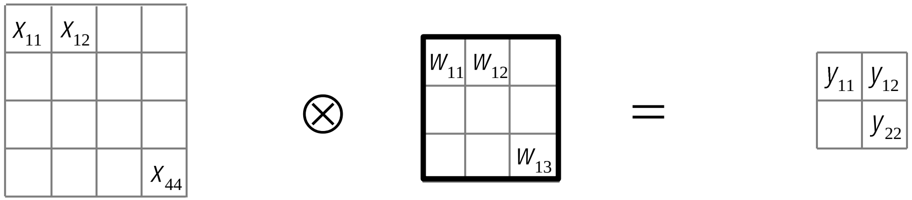
0	0	0	0
0	**5**	0	**4**
0	0	0	0
7	0	0	**7**

Learnable parameter?

Can we learn upsampling?



REVISITED: CONVOLUTION



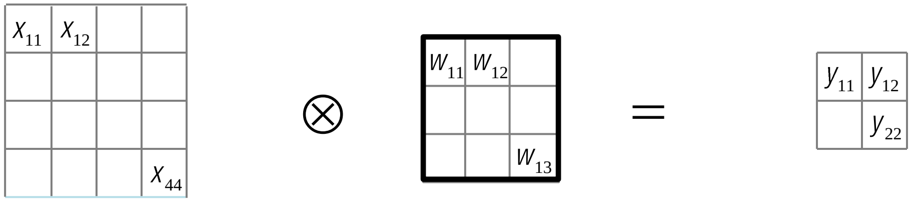
$$y_{11} = w_{11}x_{11} + w_{12}x_{12} + \cdots + w_{33}x_{33}$$

$$y_{12} = w_{11}x_{12} + w_{12}x_{13} + \cdots + w_{33}x_{34}$$

$$y_{21} = w_{11}x_{21} + w_{12}x_{21} + \cdots + w_{33}x_{43}$$

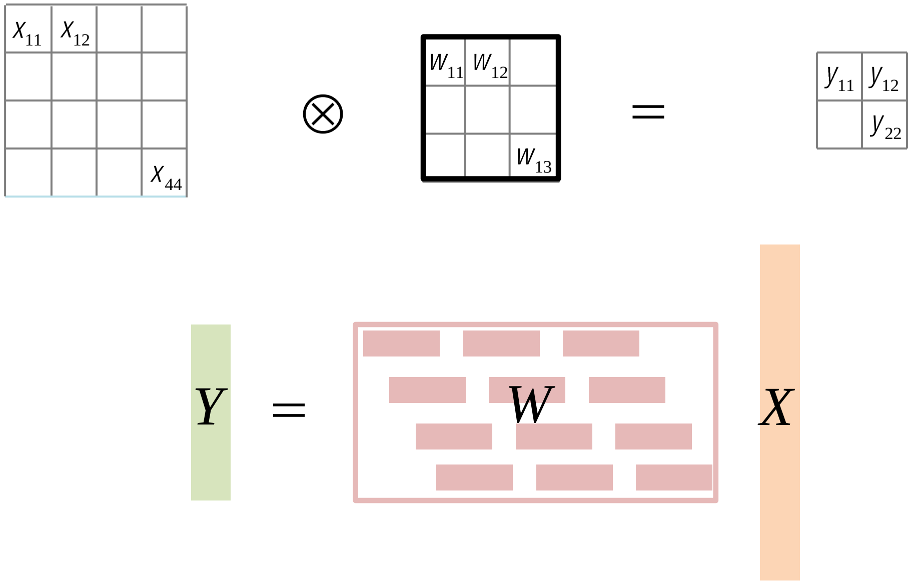
$$y_{22} = w_{11}x_{22} + w_{12}x_{23} + \cdots + w_{33}x_{44}$$

Revisited: Convolution



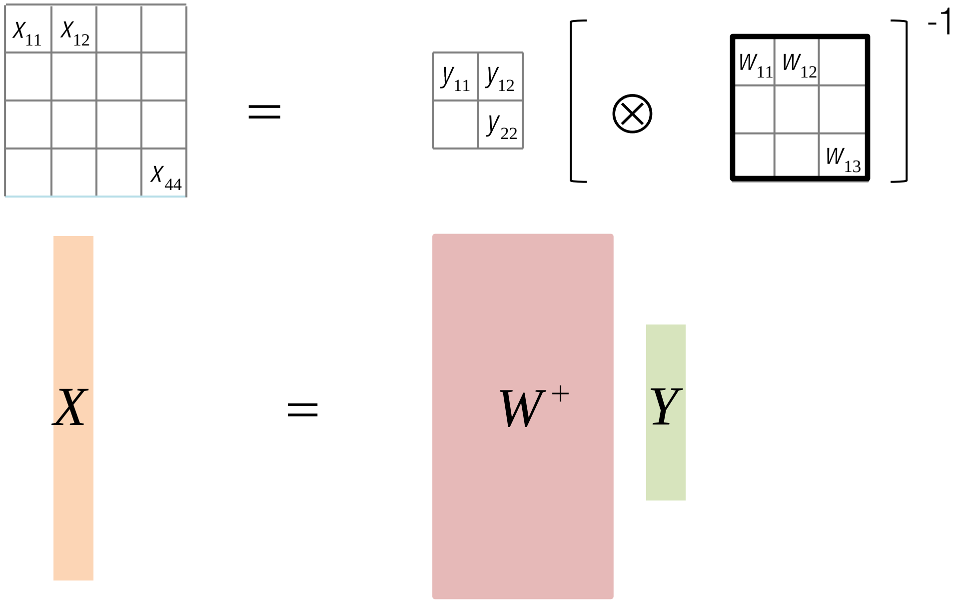
$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{44} \end{bmatrix}$$

REVISITED: CONVOLUTION



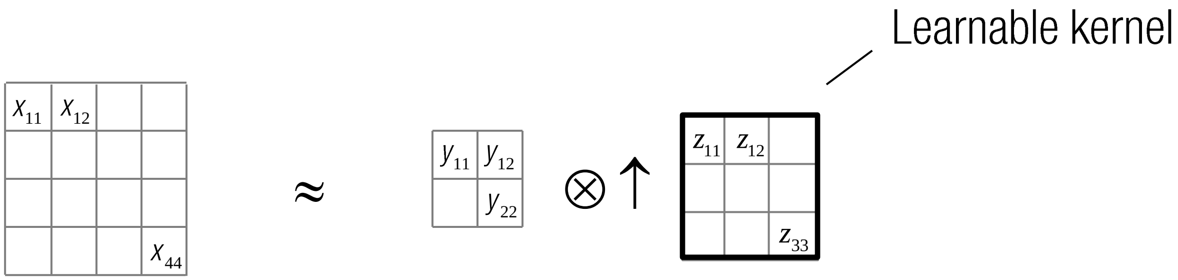
Non-zero entries of each row encode local connectivity.

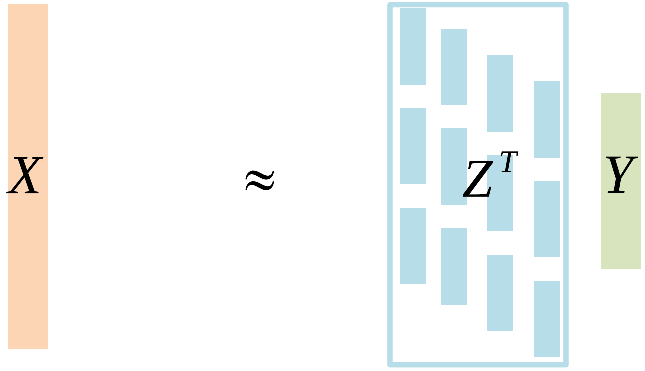
INVERSE OF CONVOLUTION



Inverse of W does not preserve local connectivity.

Upconvolution ~ Inverse of Convolution



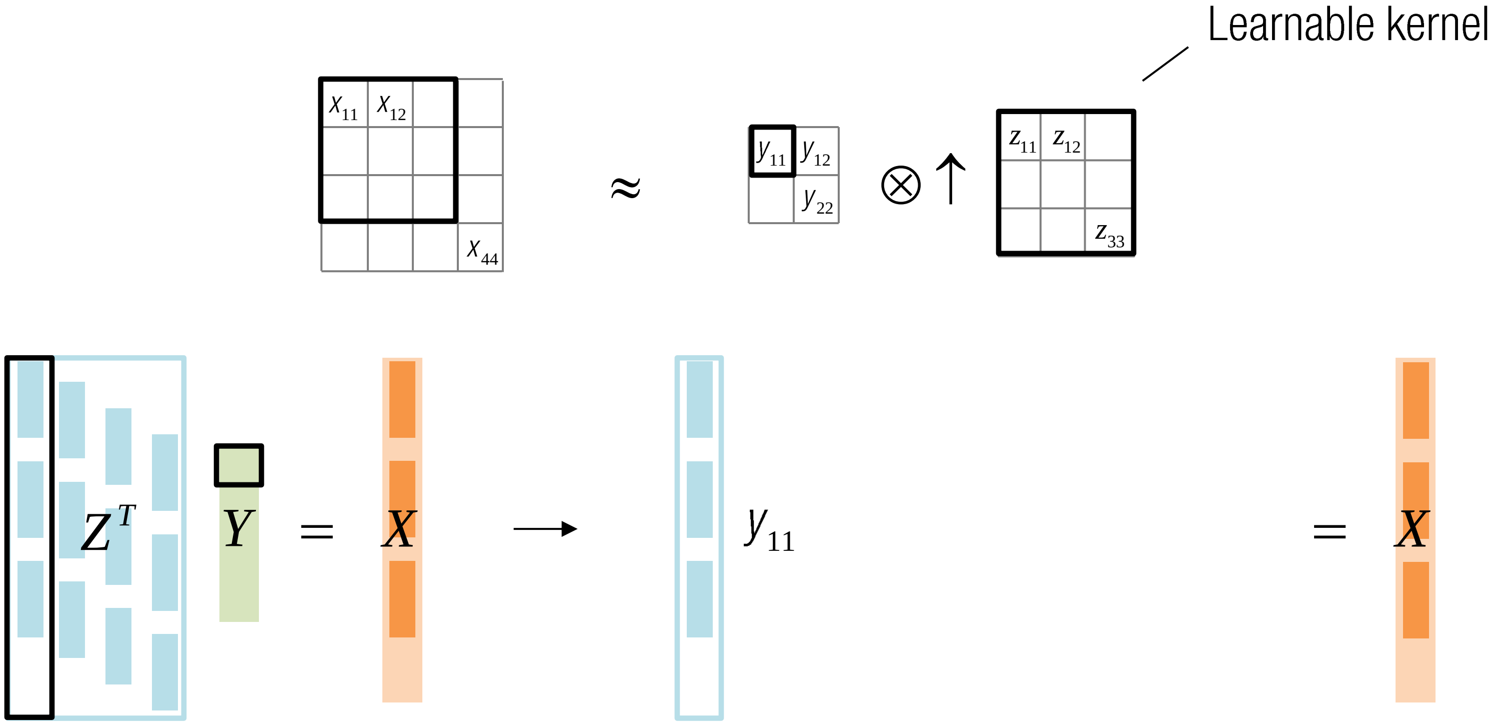


$$W = Z$$

if $\quad W^T W \approx I$

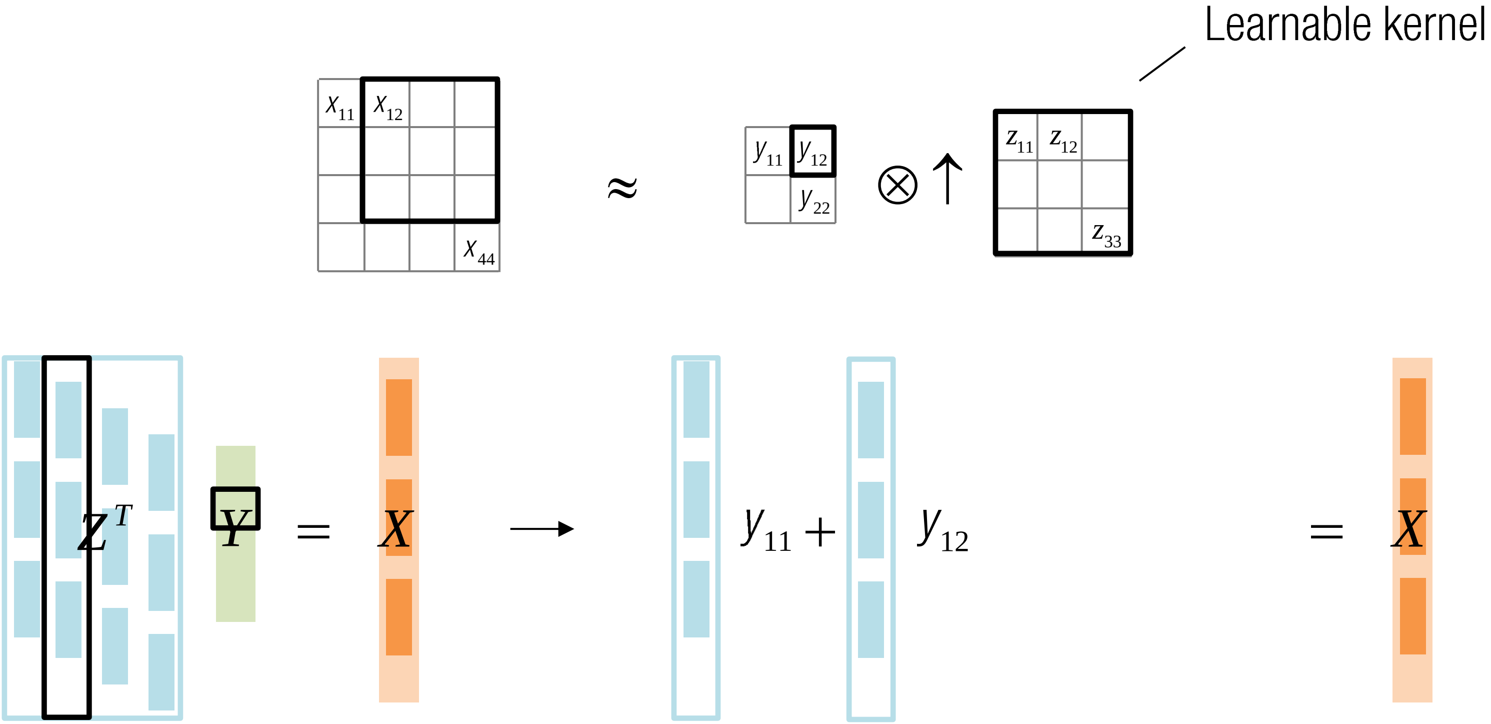
Non-zero entries of column encode local connectivity.

UPCONVOLUTION ~ INVERSE OF CONVOLUTION



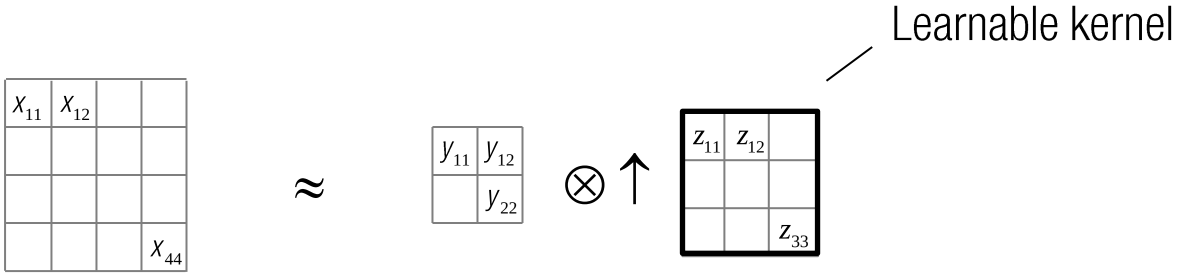
Non-zero entries of column encode local connectivity.

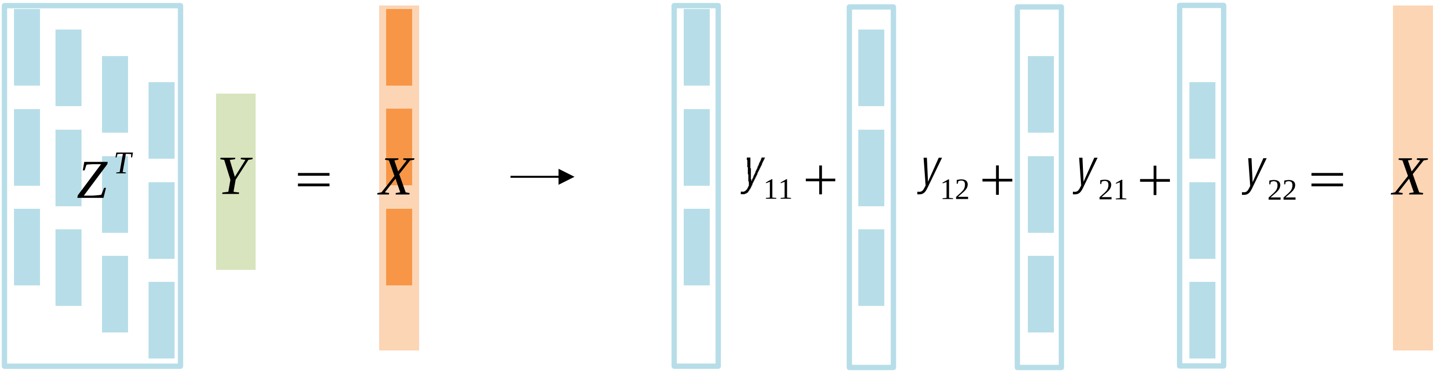
Transposed Convolution ~ Inverse of Convolution



Non-zero entries of column encode local connectivity.

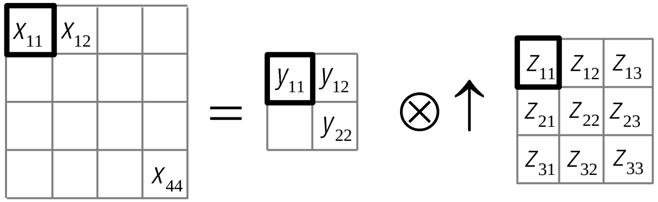
Transposed Convolution ~ Inverse of Convolution



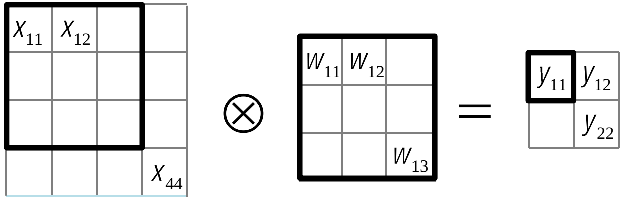


Non-zero entries of column encode local connectivity.

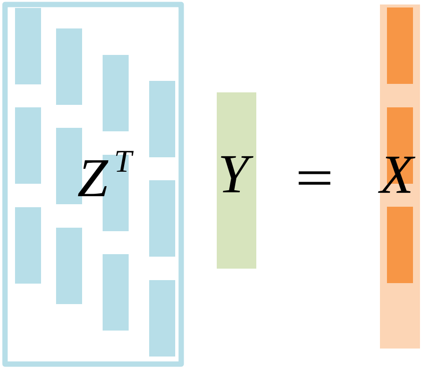
Transposed Convolution as Upconvolution



Upconvolution



Cf) convolution



$$x_{11} = z_{11} y_{11}$$

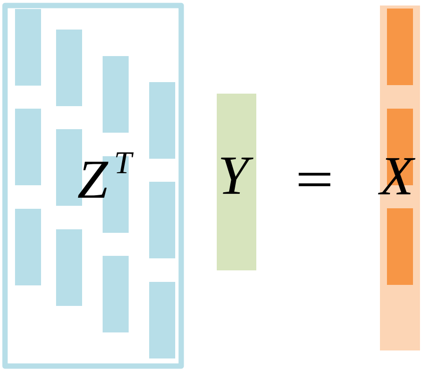
x_{11} is dependent only on y_{11}.

x_{11} contributes only to y_{11}.

Transposed Convolution as Upconvolution

$$\begin{bmatrix} x_{11} & \boxed{x_{12}} & & \\ & & & \\ & & & \\ & & & x_{44} \end{bmatrix} = \begin{bmatrix} \boxed{y_{11} \quad y_{12}} \\ \qquad\quad y_{22} \end{bmatrix} \otimes\uparrow \begin{bmatrix} \boxed{z_{11} \quad z_{12}} & z_{13} \\ z_{21} \quad z_{22} & z_{23} \\ z_{31} \quad z_{32} & z_{33} \end{bmatrix}$$

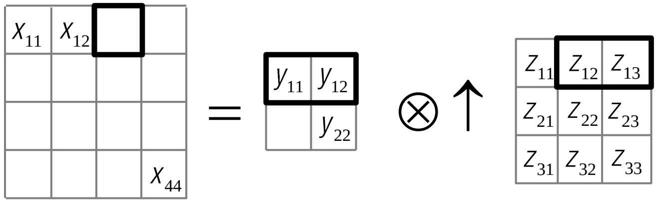
Upconvolution



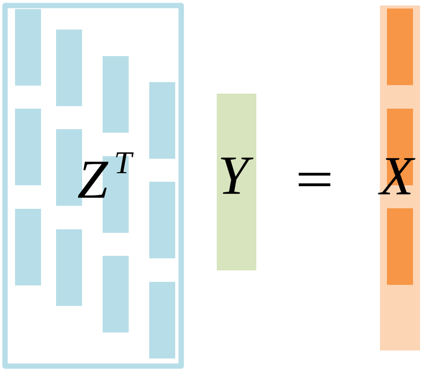
$$x_{11} = z_{11} y_{11}$$

$$x_{12} = z_{12} y_{11} + z_{11} y_{12}$$

Transposed Convolution as Upconvolution



Upconvolution

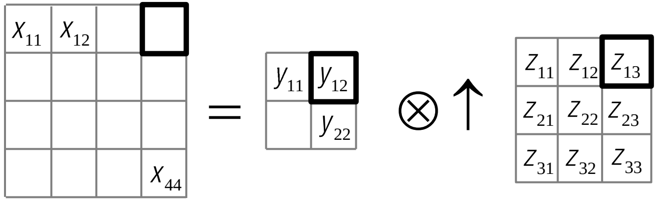


$$x_{11} = z_{11}y_{11}$$

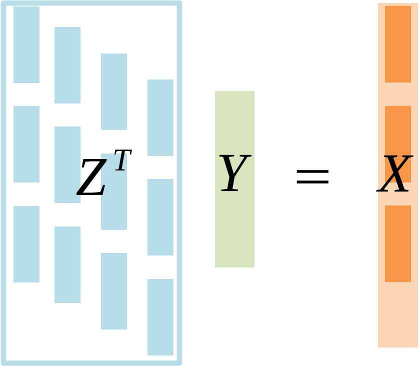
$$x_{12} = z_{12}y_{11} + z_{11}y_{12}$$

$$x_{13} = z_{13}y_{11} + z_{12}y_{12}$$

Transposed Convolution as Upconvolution



Upconvolution



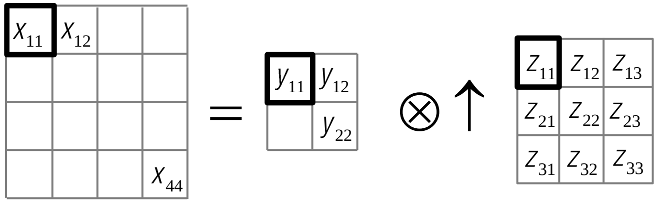
$$x_{11} = z_{11} y_{11}$$

$$x_{12} = z_{12} y_{11} + z_{11} y_{12}$$

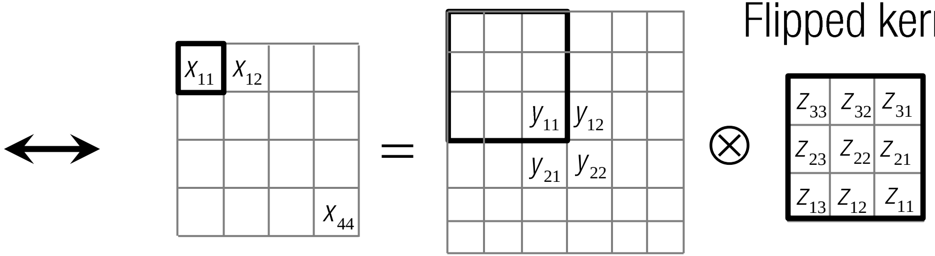
$$x_{13} = z_{13} y_{11} + z_{12} y_{12}$$

$$x_{14} = z_{13} y_{12}$$

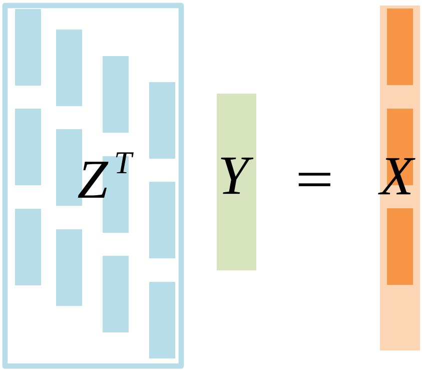
Transposed Convolution as Upconvolution



Upconvolution

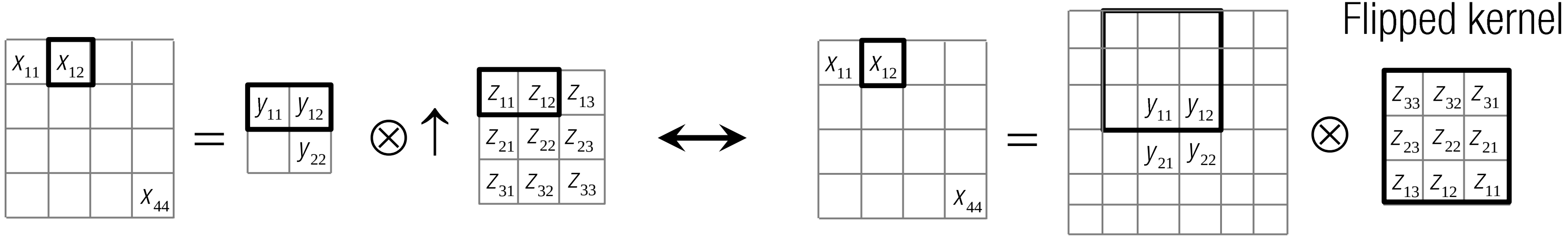


Convolution with zero padding



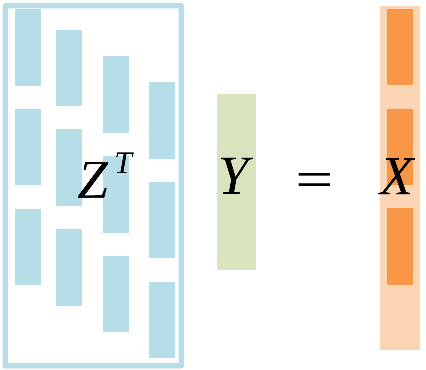
$$x_{11} = z_{11} y_{11}$$

TRANSPOSED CONVOLUTION AS UPCONVOLUTION



Upconvolution

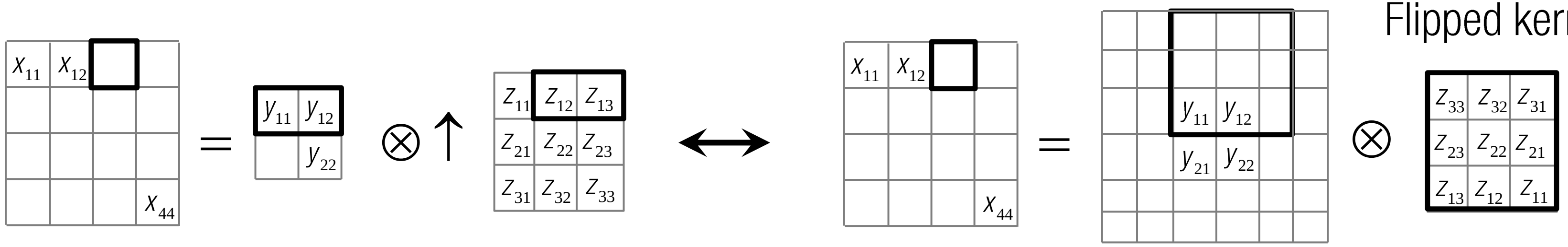
Convolution with zero padding



$$x_{11} = z_{11} y_{11}$$

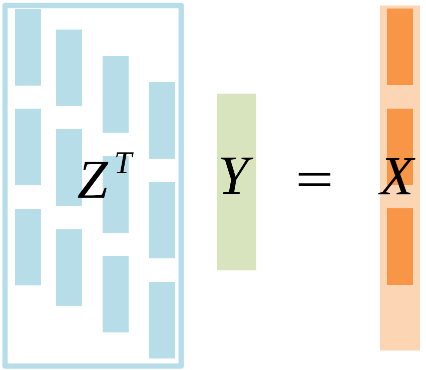
$$x_{12} = z_{12} y_{11} + z_{11} y_{12}$$

Transposed Convolution as Upconvolution



Upconvolution

Convolution with zero padding

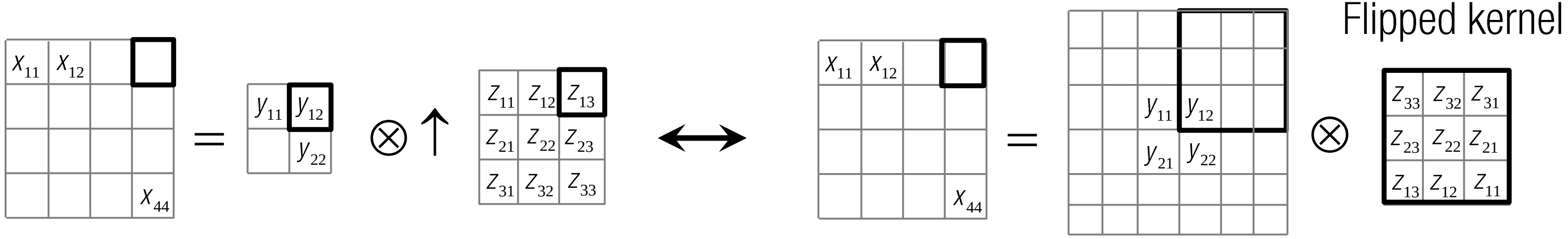


$$x_{11} = z_{11}y_{11}$$

$$x_{12} = z_{12}y_{11} + z_{11}y_{12}$$

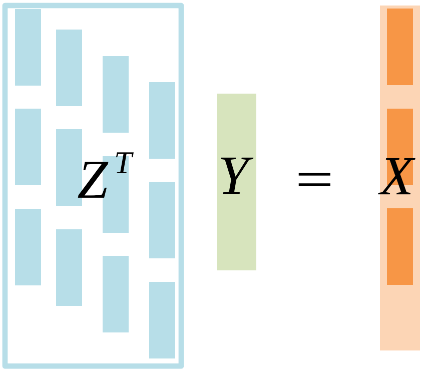
$$x_{13} = z_{13}y_{11} + z_{12}y_{12}$$

Transposed Convolution as Upconvolution



Upconvolution

Convolution with zero padding



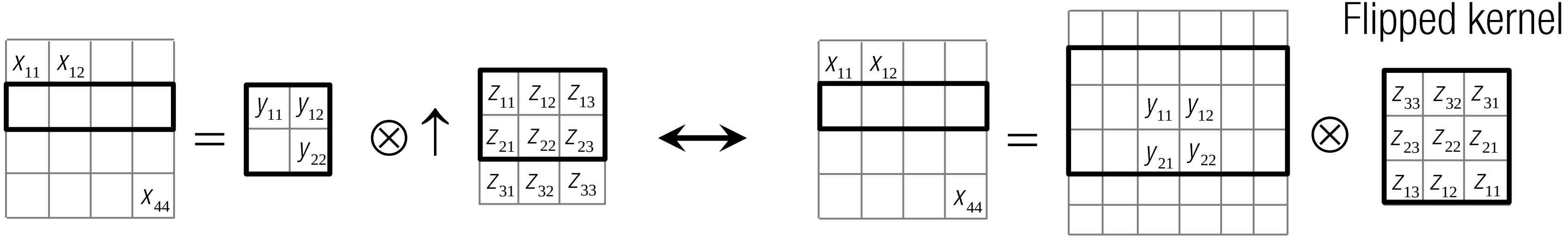
$$x_{11} = z_{11} y_{11}$$

$$x_{12} = z_{12} y_{11} + z_{11} y_{12}$$

$$x_{13} = z_{13} y_{11} + z_{12} y_{12}$$

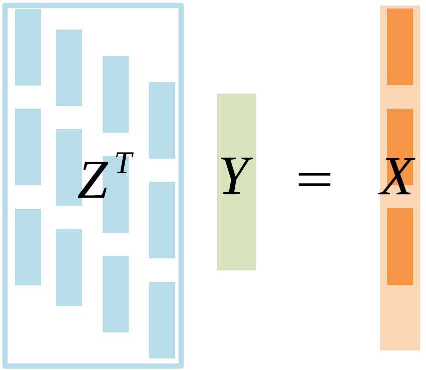
$$x_{14} = z_{13} y_{12}$$

Transposed Convolution as Upconvolution



Upconvolution

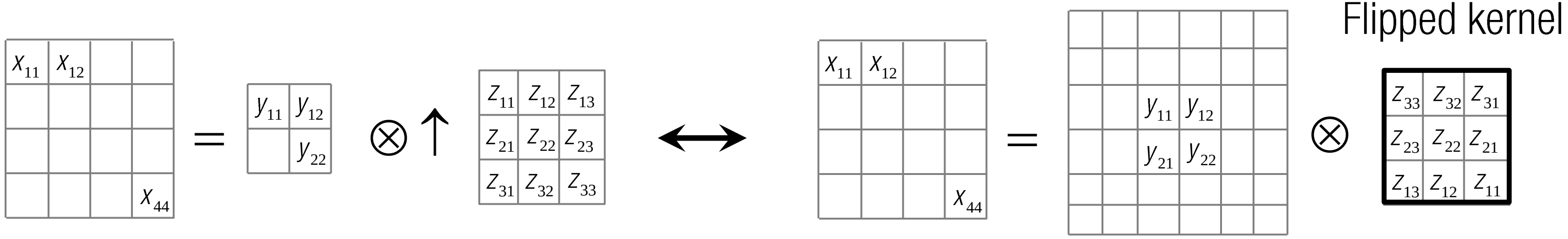
Convolution with zero padding



$$x_{11} = z_{11}y_{11}$$
$$x_{12} = z_{12}y_{11} + z_{11}y_{12}$$
$$x_{13} = z_{13}y_{11} + z_{12}y_{12}$$
$$x_{14} = z_{13}y_{12}$$

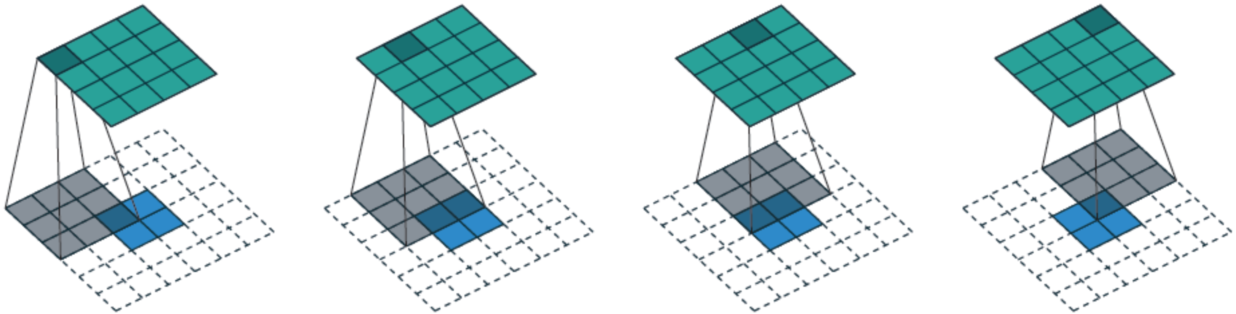
$$x_{21} = z_{21}y_{11} + z_{11}y_{21}$$
$$x_{22} = z_{22}y_{11} + z_{21}y_{12} + z_{12}y_{21} + z_{11}y_{22}$$
$$x_{23} = z_{23}y_{11} + z_{22}y_{12} + z_{13}y_{21} + z_{12}y_{22}$$
$$x_{24} = z_{23}y_{12} + z_{13}y_{22}$$

TRANSPOSED CONVOLUTION AS UPCONVOLUTION



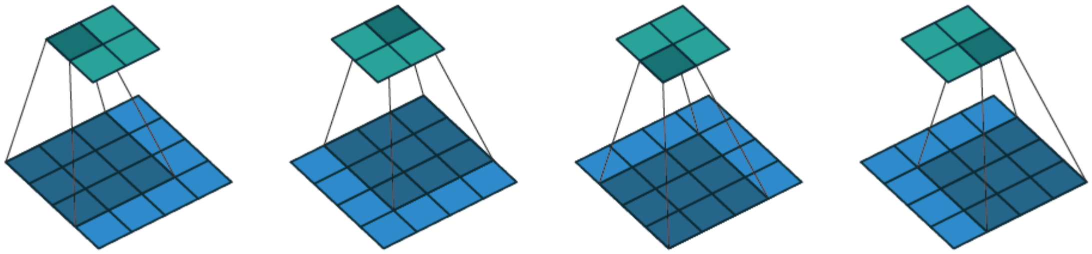
Upconvolution

Convolution with zero padding

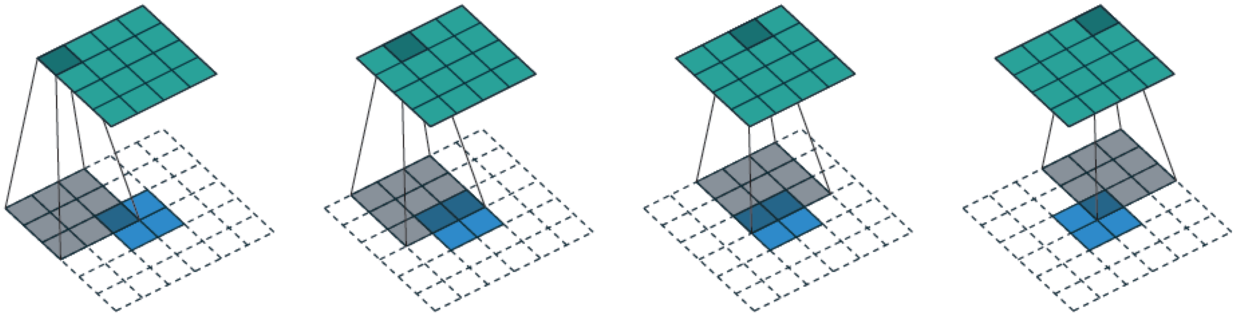


Upconvolution

DUALITY: CONVOLUTION VS. UPCONVOLUTION

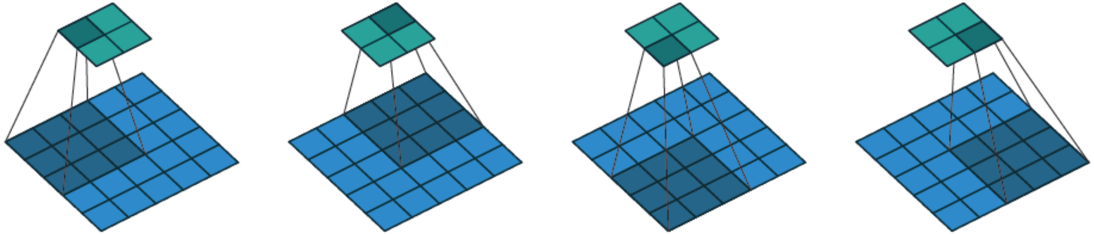


Convolution (3x3 kernel, no zero padding, 1 stride)

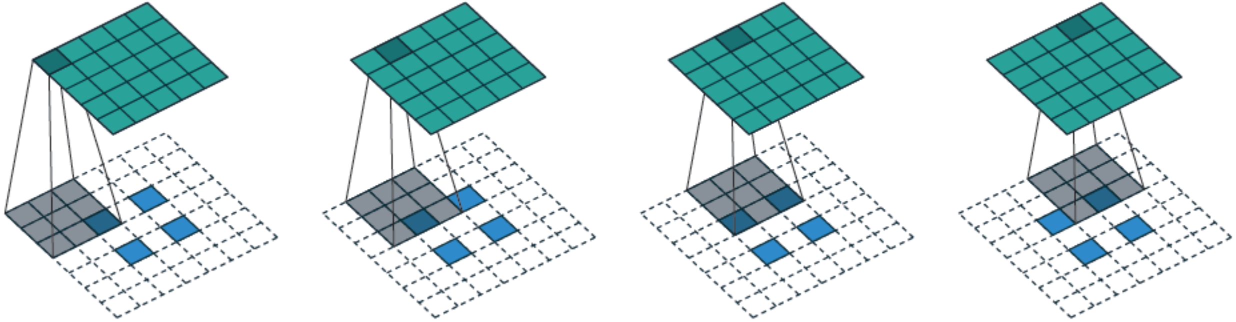


Upconvolution (3x3 kernel, 2x2 zero padding, 1 stride)

DUALITY: CONVOLUTION VS. UPCONVOLUTION

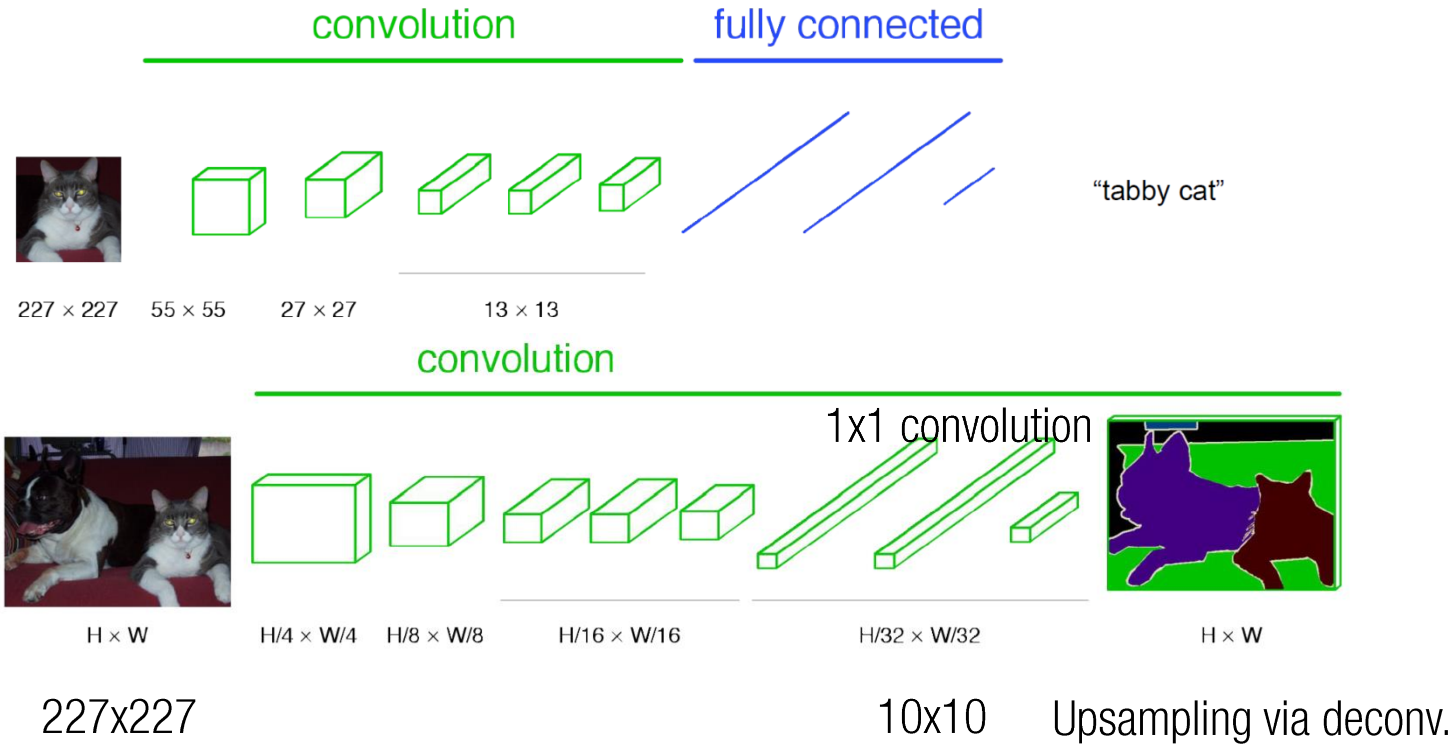


Convolution (3x3 kernel, no zero padding, 2 stride)



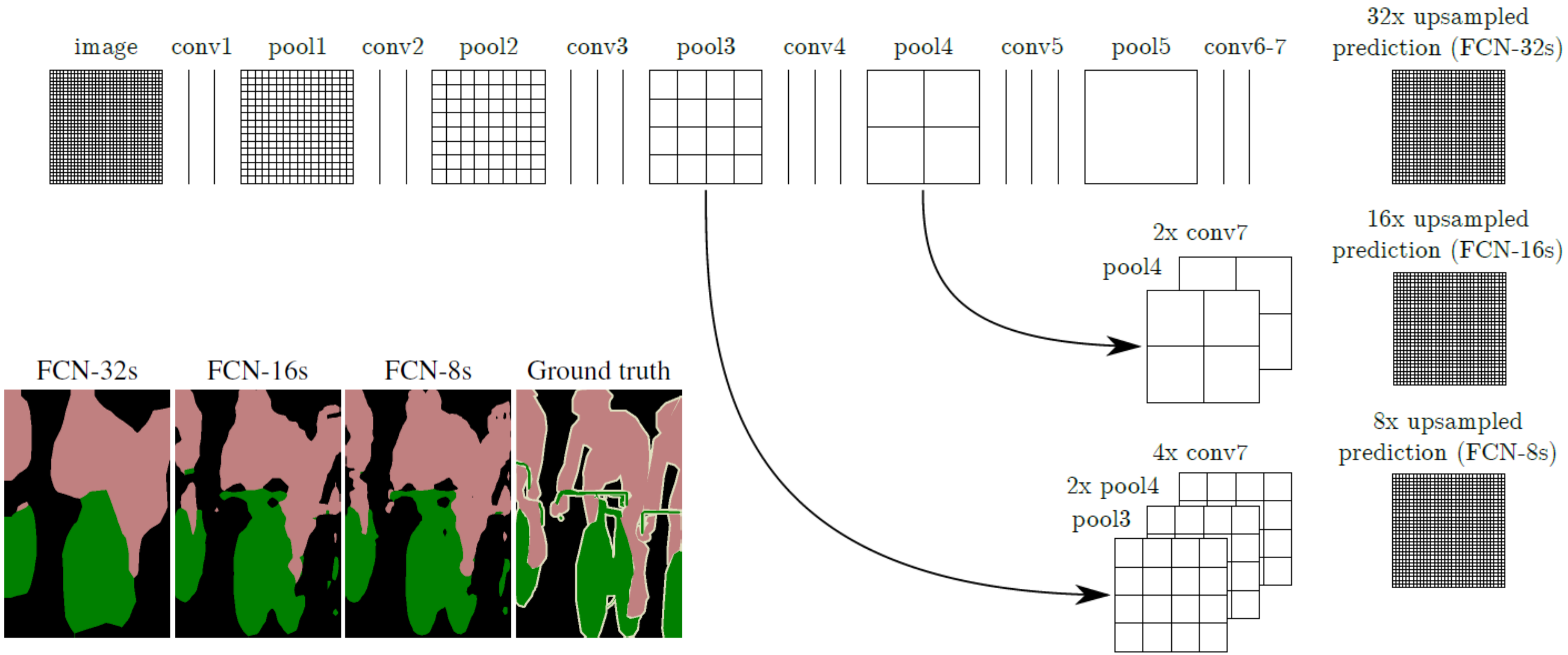
Upconvolution (3x3 kernel, 2x2 zero padding, 1 zero between input, 2 stride)

Fully Convolutional Network

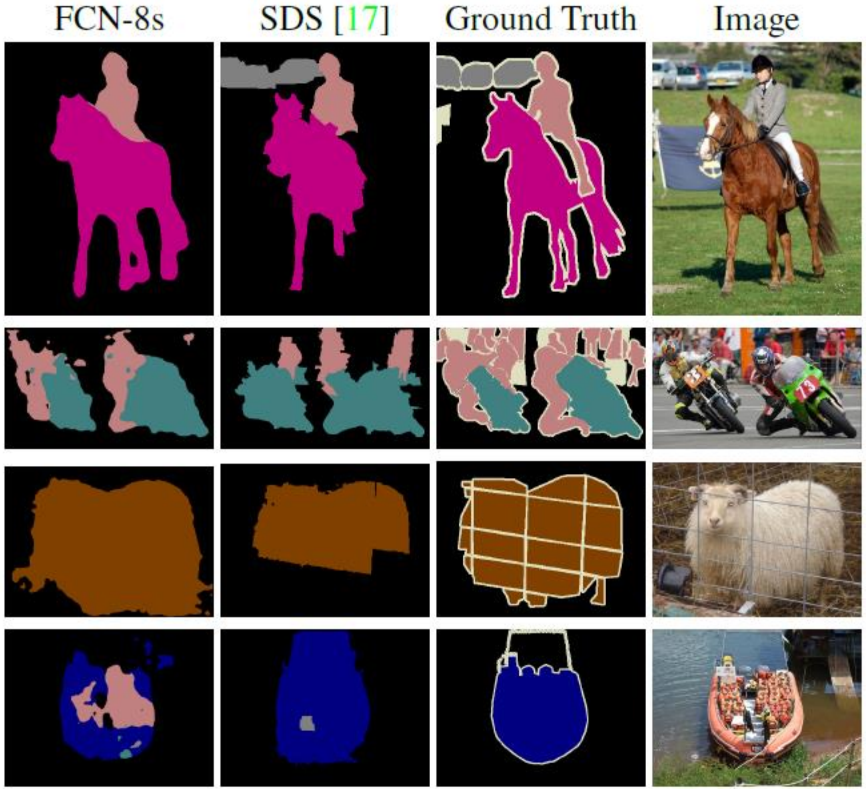


Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

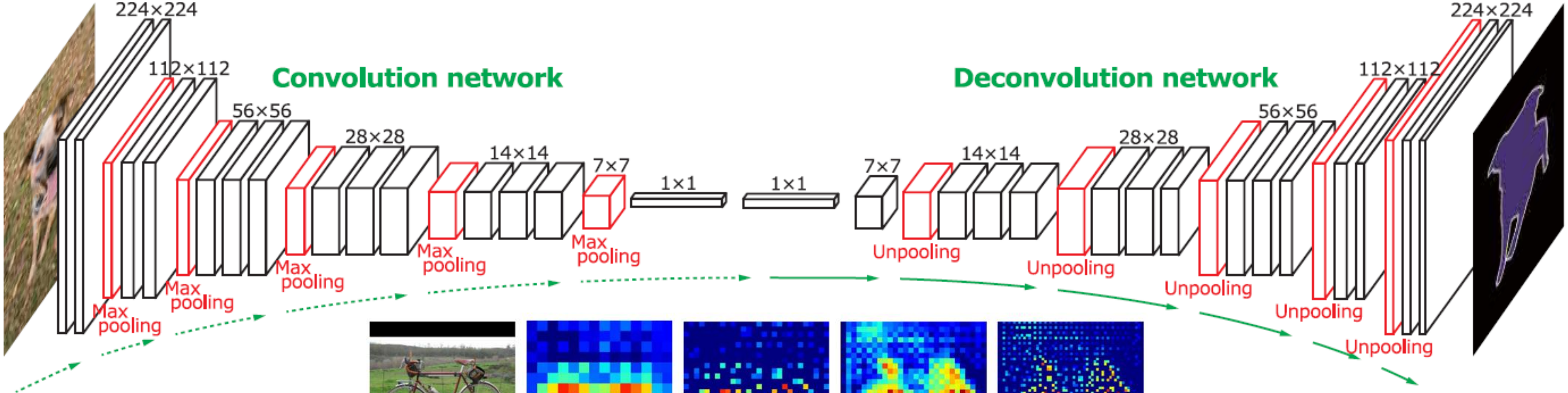
FULLY CONVOLUTIONAL NETWORK

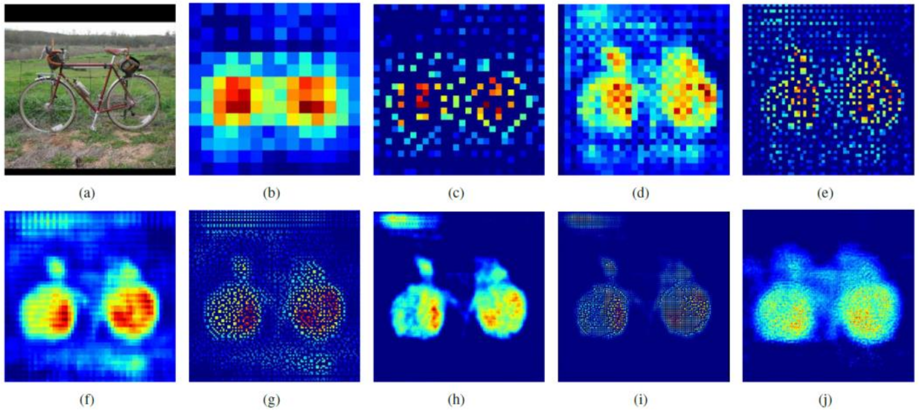


Long, Shelhamer, and Darrell, “Fully Convolutional Networks for Semantic Segmentation”, CVPR 2015

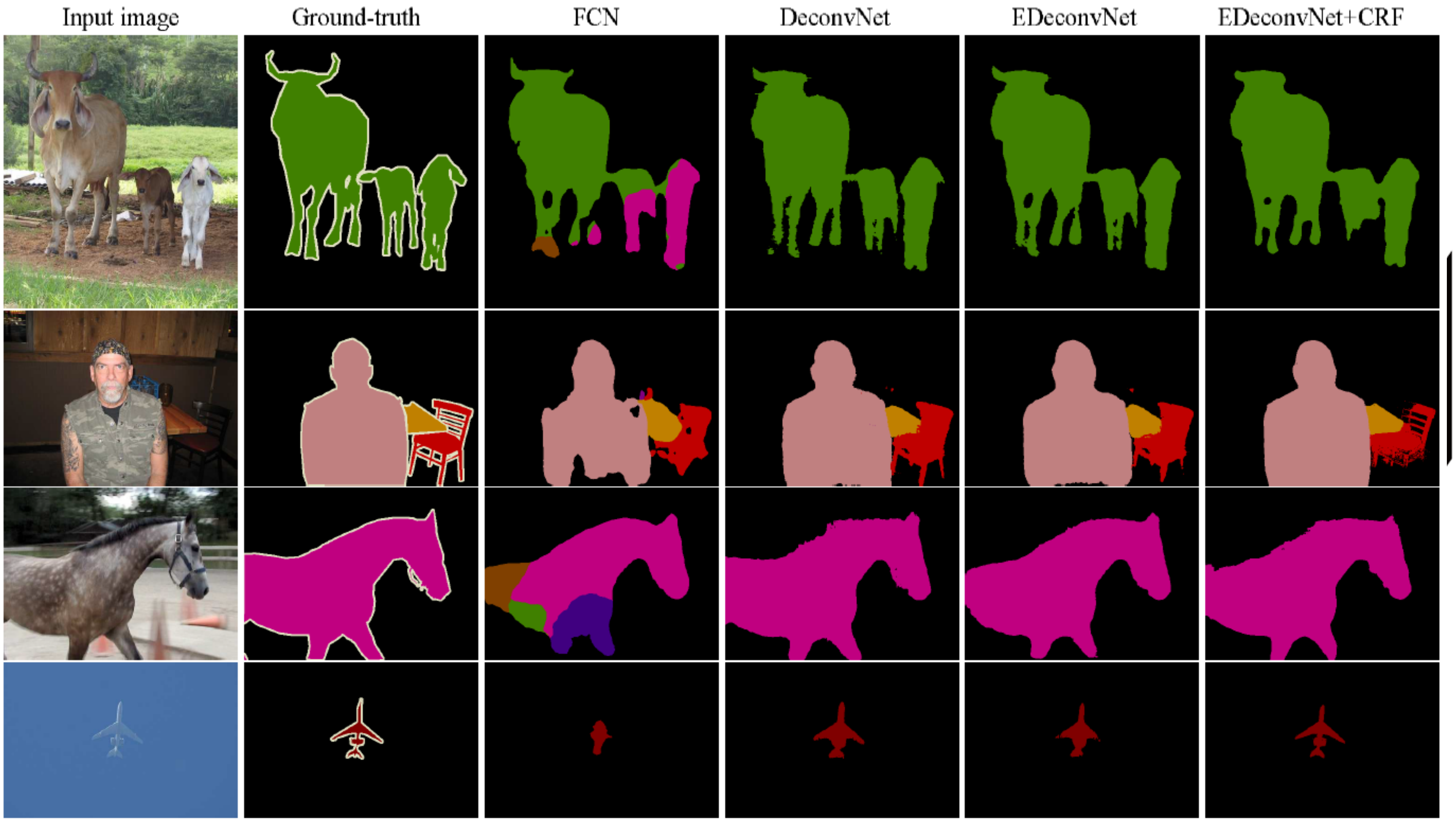

FCN-8s
SDS [17]
Ground Truth
Image

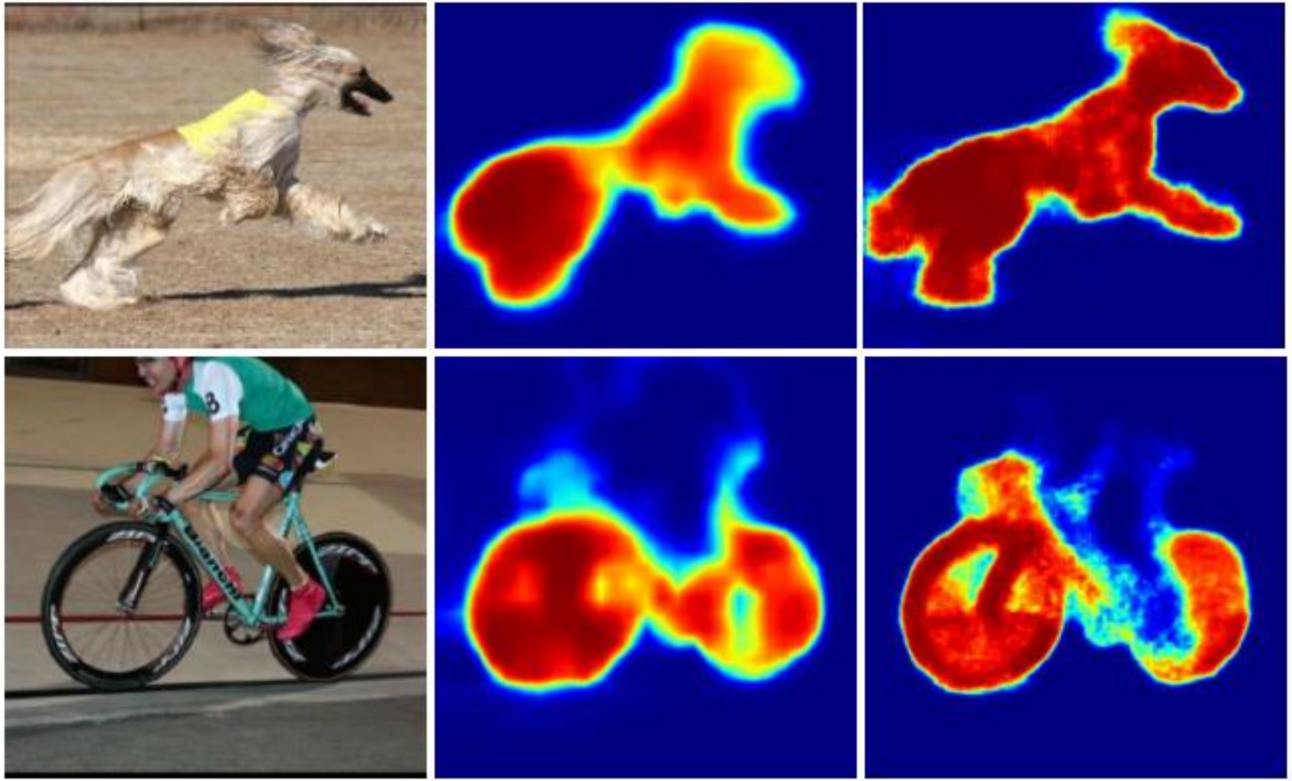
DEEPER UPCONVOLUTIONAL LAYERS





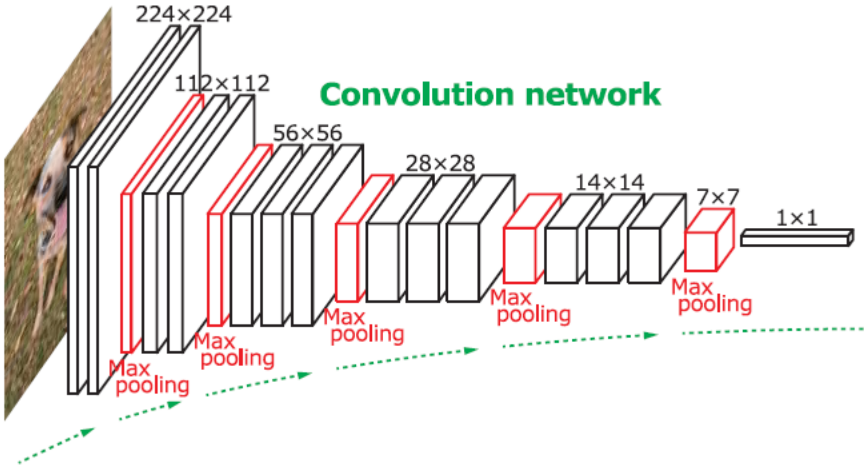
Noh, Hong, and Han, “Learning Deconvolutional Network for Semantic Segmentation”, ICCV 2015


Input image
Ground-truth
FCN
DeconvNet
EDeconvNet
EDeconvNet+CRF



(a) Input image (b) FCN-8s (c) Ours

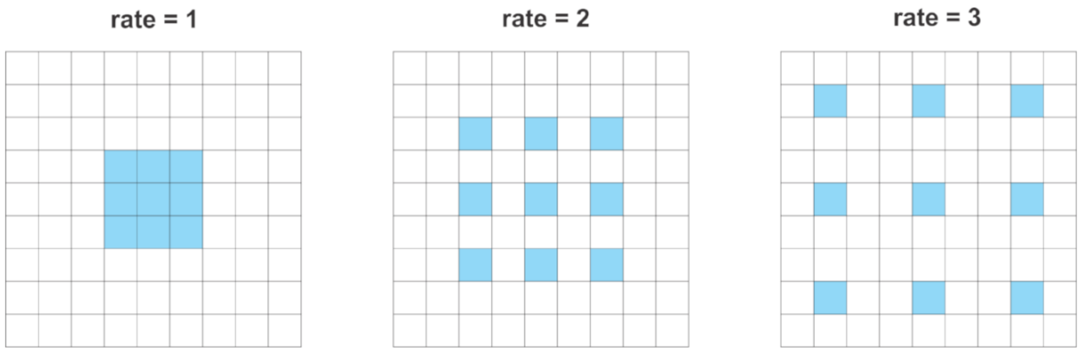
Receptive Field vs. Resolution

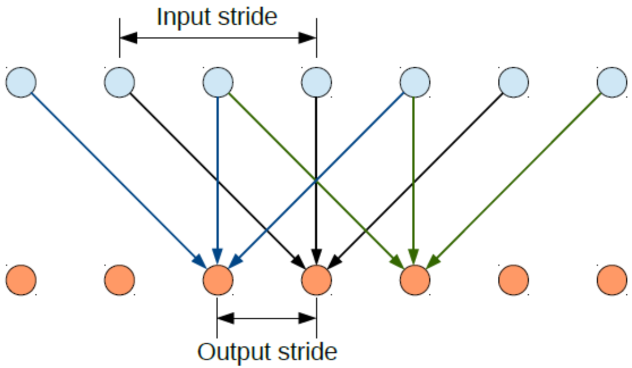


Desired properties:

- Larger receptive field (bigger spatial context) ⟶ reducing resolution
- Higher output resolution ⟶ reducing receptive field

DeepLab: Dilated Convolution

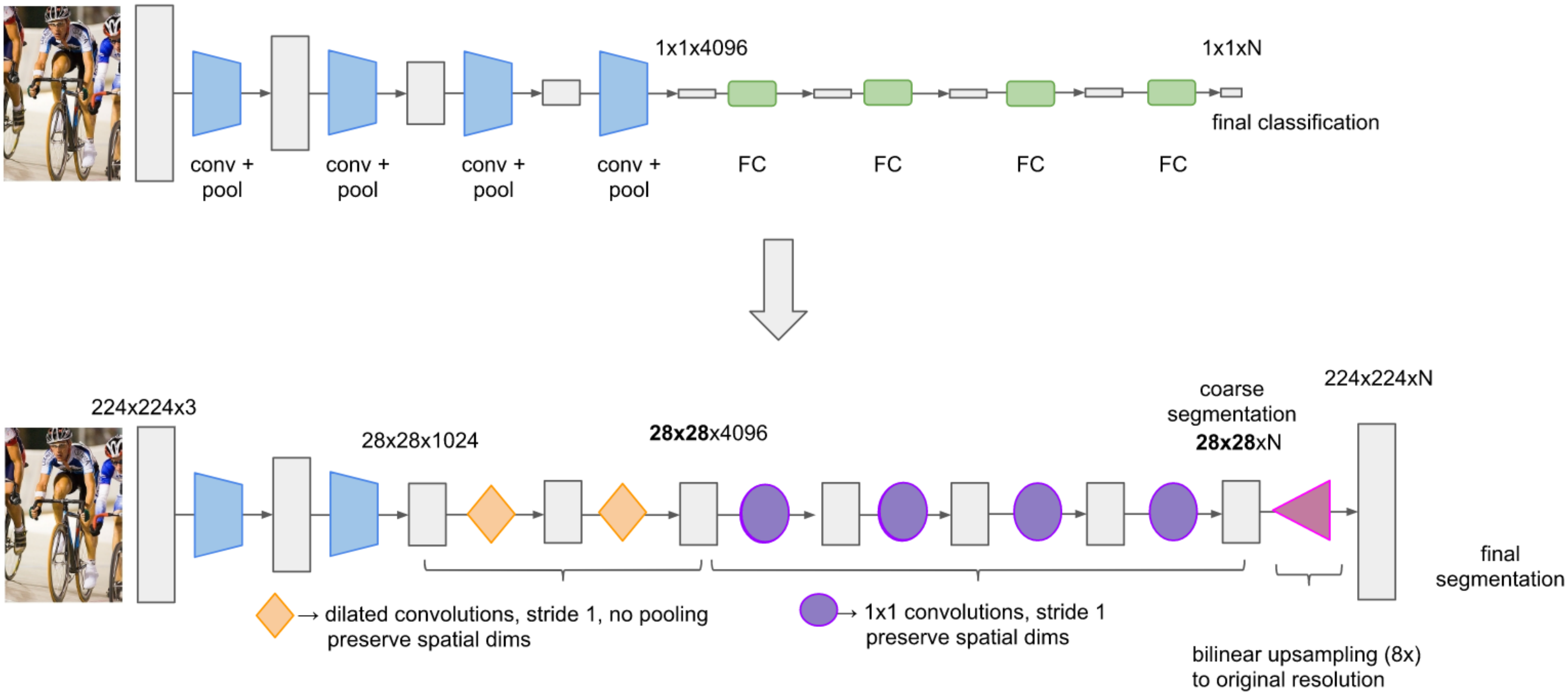




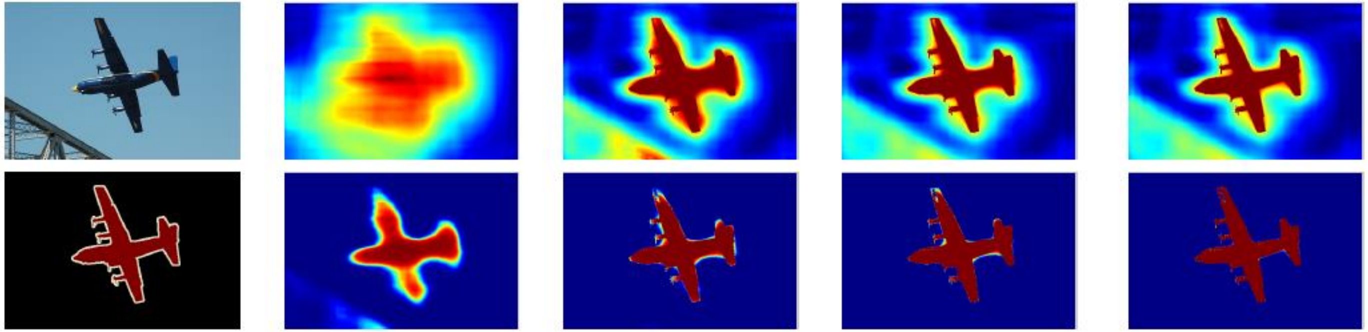
Enlarging receptive field without increasing reducing resolution

Chen, Papandreou, Kokkinos, Murphy, and Yuille, “Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs”, ICLR 2015

DeepLab: Dilated Convolution



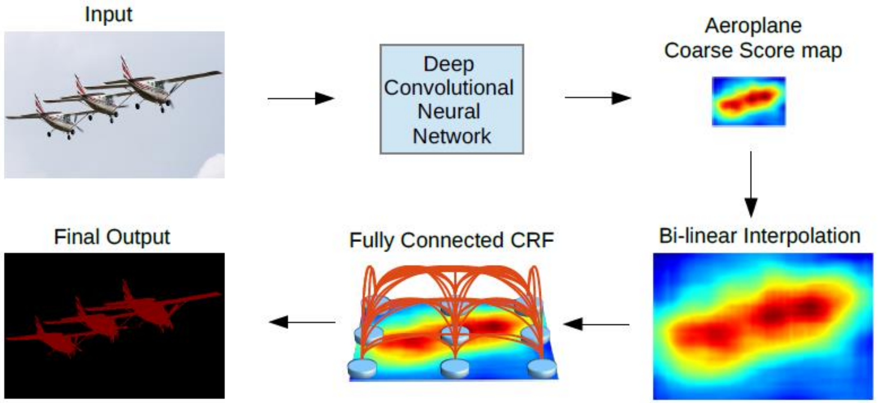
DeepLab: CRF Post-processing

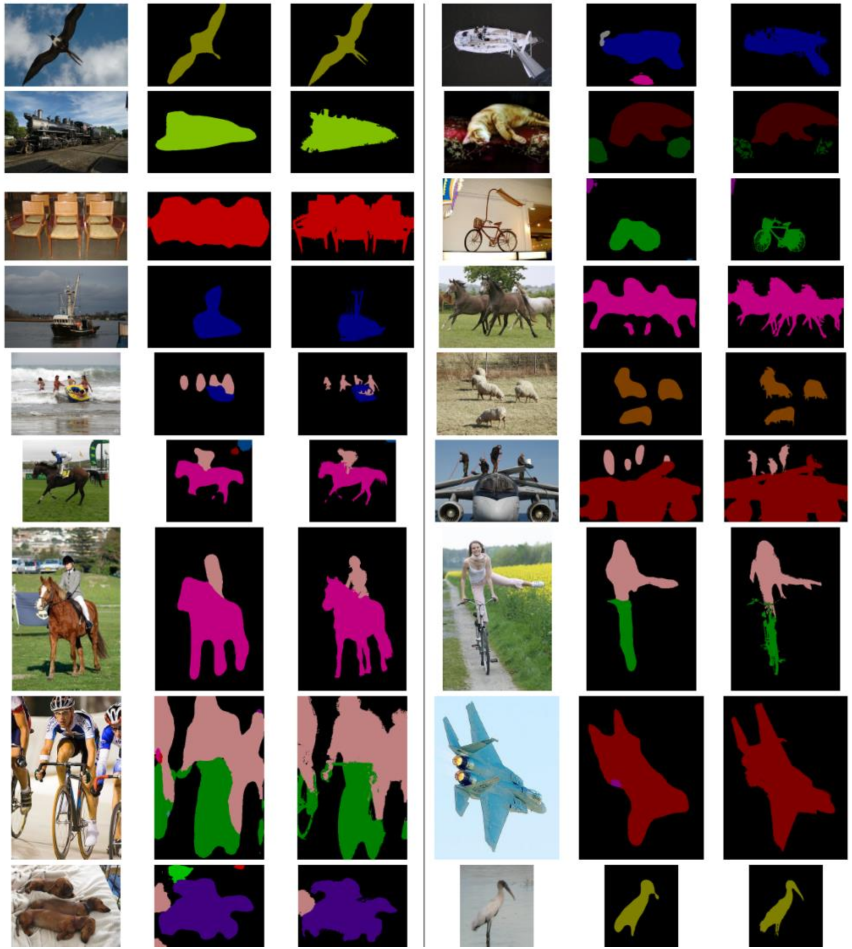


Image/G.T.	DCNN output	CRF Iteration 1	CRF Iteration 2	CRF Iteration 10

Chen, Papandreou, Kokkinos, Murphy, and Yuille, “Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs”, ICLR 2015

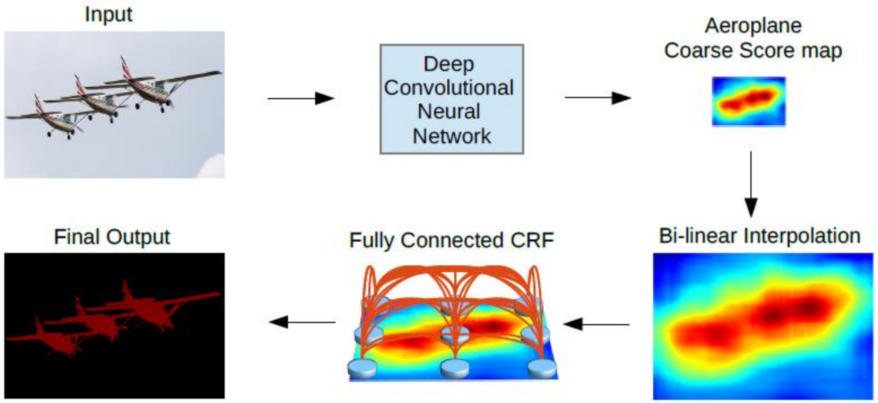
DEEPLAB

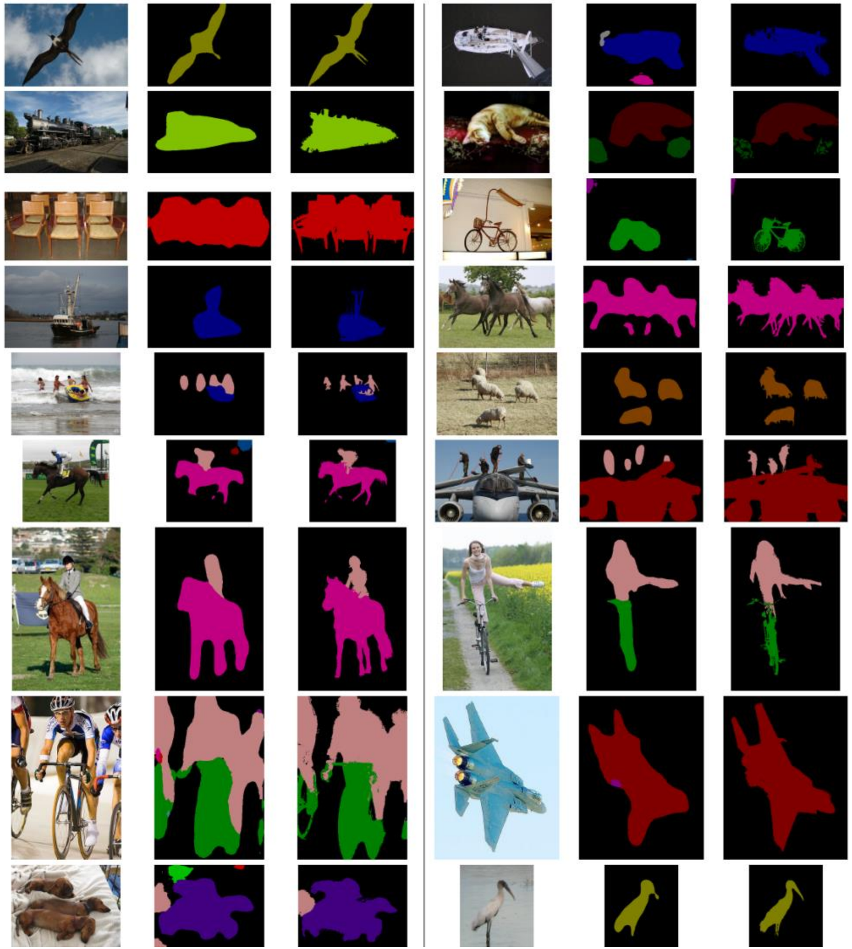




Chen, Papandreou, Kokkinos, Murphy, and Yuille, “Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs”, ICLR 2015

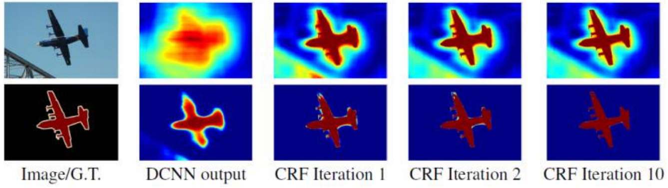
DEEPLAB

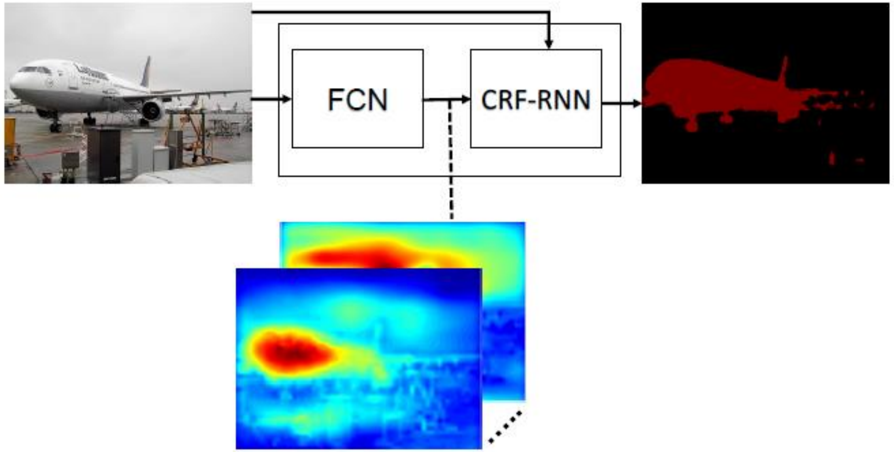




Chen, Papandreou, Kokkinos, Murphy, and Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs", ICLR 2015

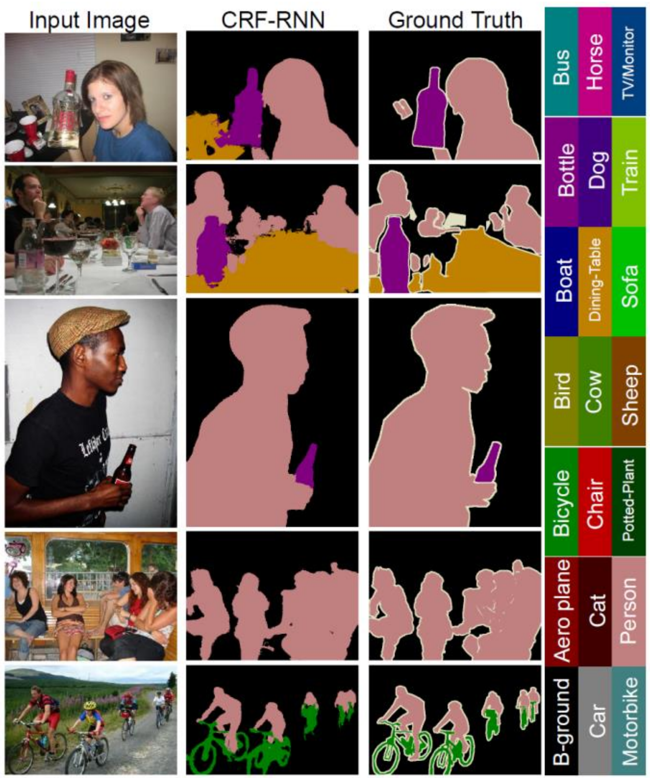
CRF AS RECURRENT NEURAL NET (END-TO-END)





Zheng et al., "Conditional Random Fields as Recurrent Neural Networks", ICCV 2015

CRF as Recurrent Neural Net (End-to-End)



Zheng et al., “Conditional Random Fields as Recurrent Neural Networks”, ICCV 2015