

25 YEARS ANNIVERSARY
SOKT

ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Lesson 10: Image semantic segmentation

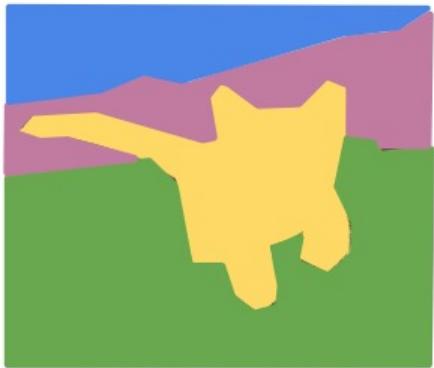
Outline

- Introduction to image segmentation problem
- Up-sampling layers
- Objective functions
- Some typical image segmentation networks

Introduction to image segmentation problem

Computer vision problems

Semantic Segmentation



GRASS, CAT,
TREE, SKY

No objects, just pixels

Classification + Localization



CAT

Single Object

Object Detection



DOG, DOG, CAT

Multiple Object

Instance Segmentation

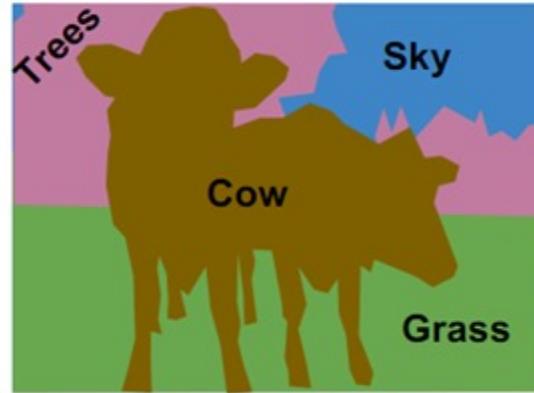
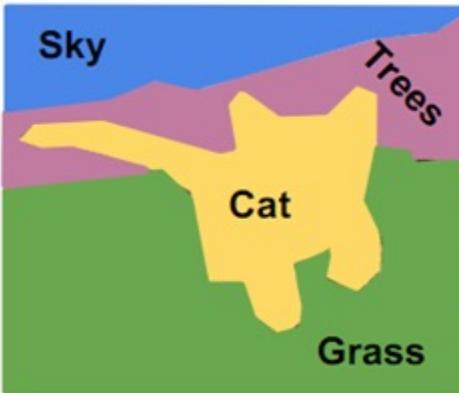


DOG, DOG, CAT

This image is CC0 public domain

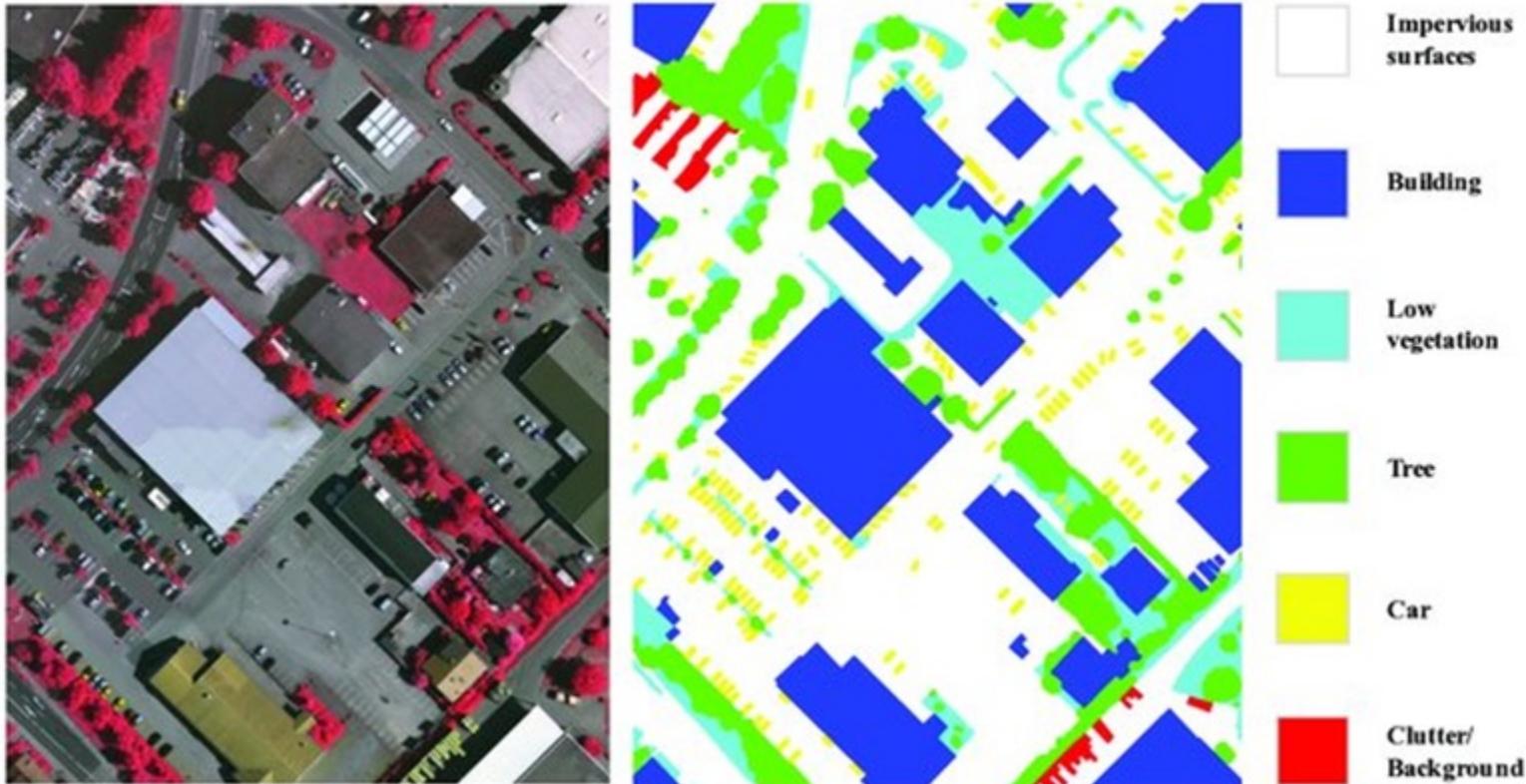
Image segmentation

- Classify each pixel in an image
- Does not distinguish objects of the same class in the image



Some image segmentation applications

- Satellite and aerial image segmentation



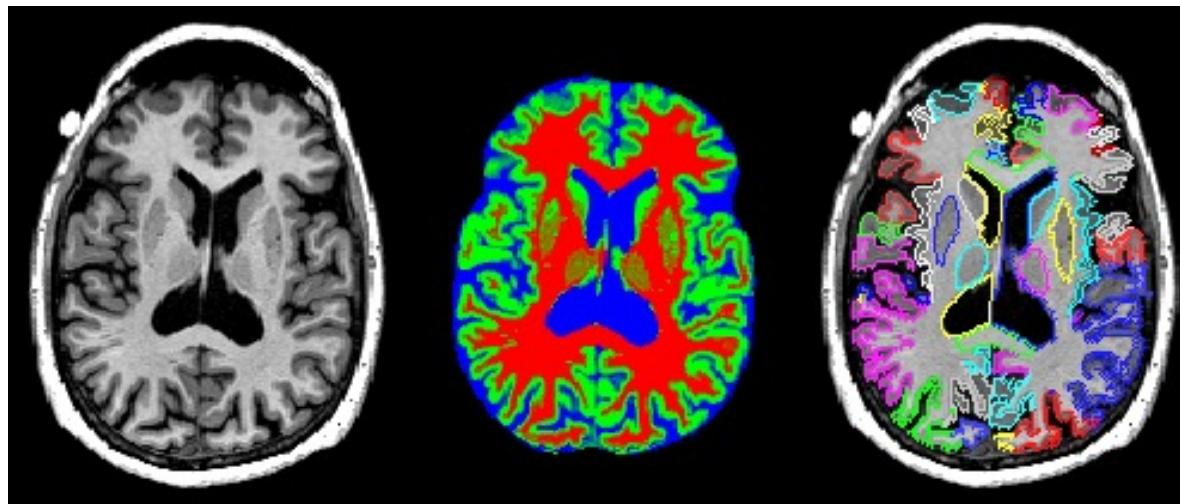
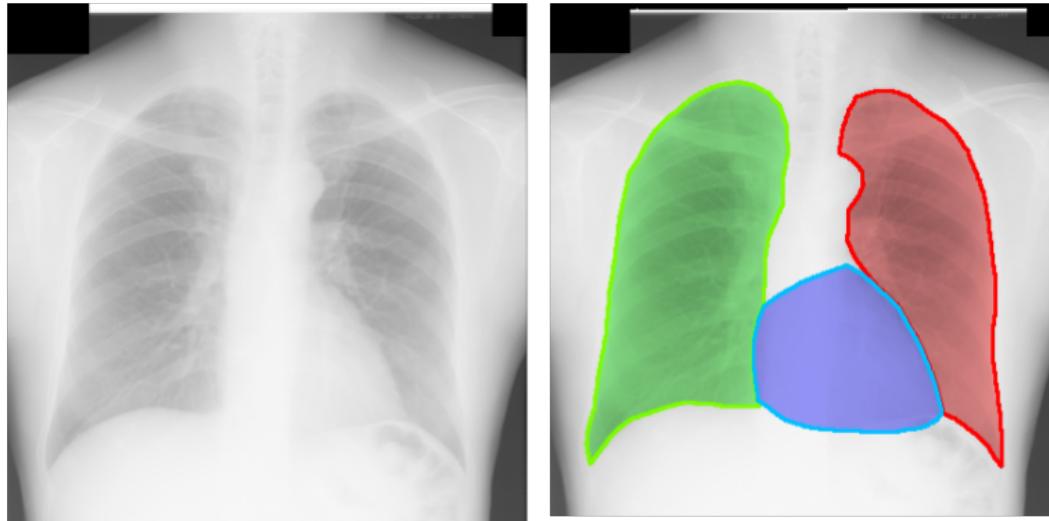
Some image segmentation applications

- Autonomous vehicle



Some image segmentation applications

- Healthcare



Some image segmentation applications

- OCR

T.2 David Goldblatt wrote in his world

Brenda Koobie and Cecilia Foreign Minister Motsoane Beavis will meet with President Hamid Karzai and other Afghan officials. The Afghan Foreign Ministry said: "It is hard to understand the country that I am delighted it's going to be tough, because they are very vocal, loud and can make a cause happen on some of the things."

Category	Count	Percentage (%)
100% true false	2	0.00
100% true false F10	2	0.00
100% true false F100	8	0.11
100% true false F1000	8	0.00
100% true false F10000	9	0.00
100% true false F100000	10	0.00
100% true false F1000000	10	0.00



Figure 8 a train on a train track with traffic lights

countries who repeatedly came from his seat and checked his bags. Developing nations such as China, Brazil and India, on the other hand, have welcomed the measure.

Figure 9 a black table with dishes, and many different food items



T.2.1 Jeff overheard like I heard me in

choose at full because he is not being a capital master change. Justina appeared happy with their lead while (follows), with an easy go in the bag, she seemed happy to start up shop. While there are some ways to business and we have had all these very carefully and listened to business I think this is really something and should not be the problem some people suggest, said Derry. Whistlers praised the many important accomplishments of Leonard's years in his part - including the April 2006 audit report , resulting in the reclassification of government documents

asked by the Clinton administration who are not that every single-day strongly supporting him. He said: "It is no surprise that a national newspaper has been around the same was centralized and centralized sections. But it also has 500 local and regional business chartlike break and reflected first, and then no running back from a broken oil

oil price was bad. He was on

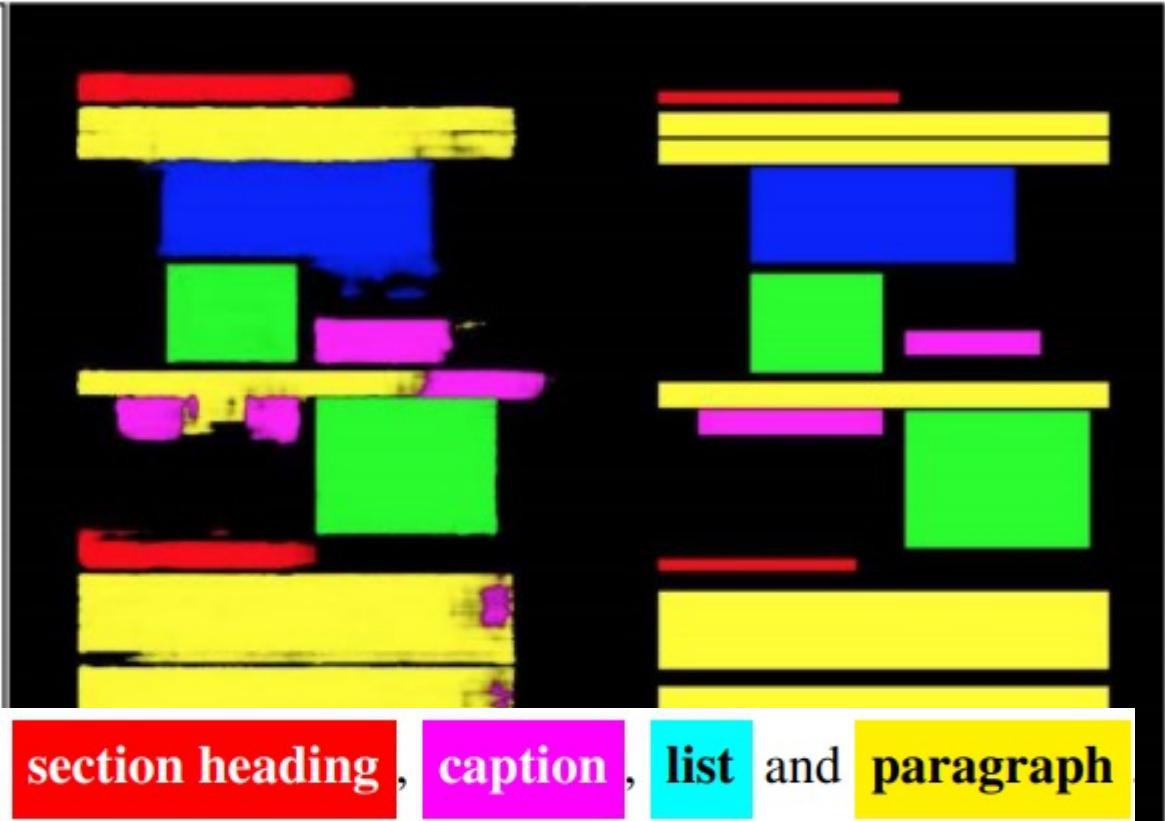
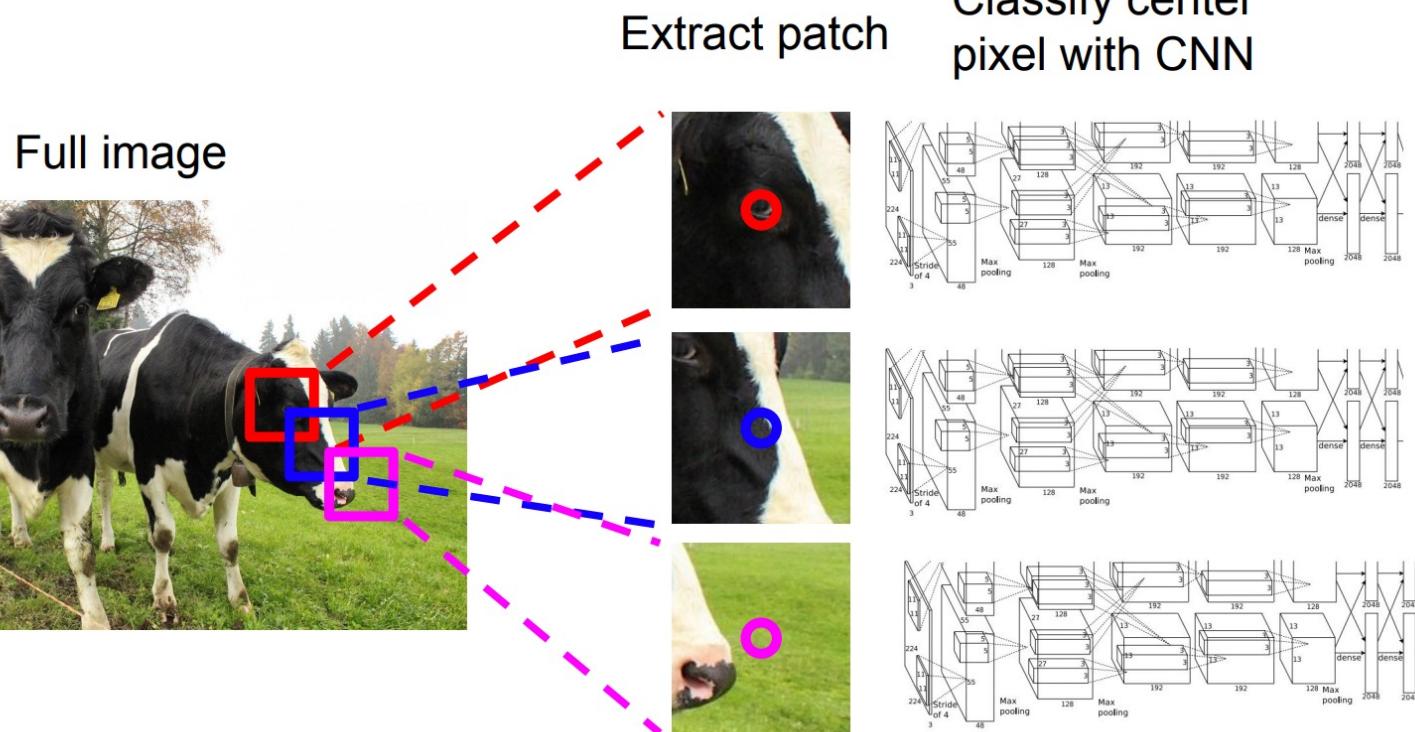
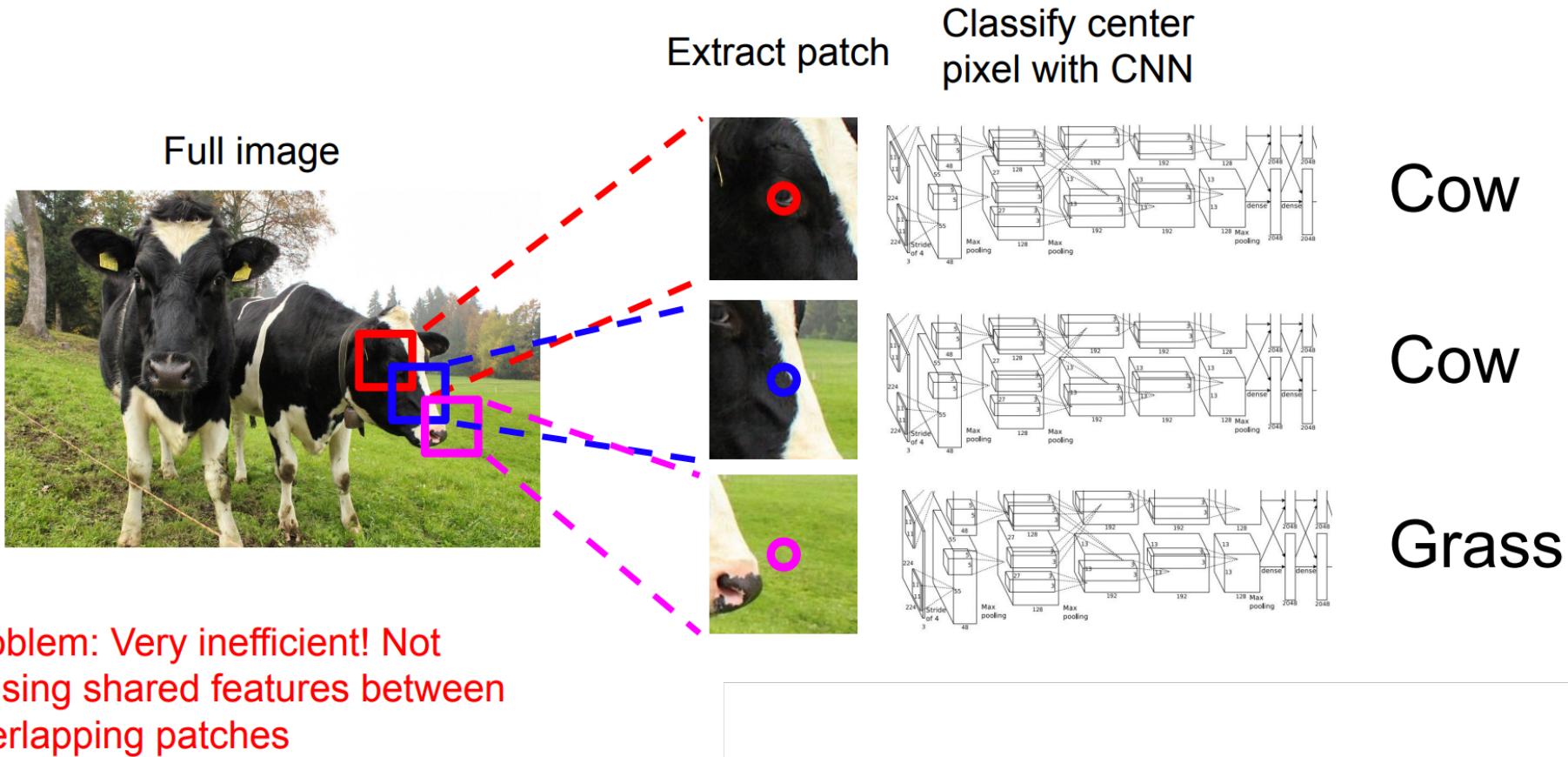


figure , **table** , **section heading** , **caption** , **list** and **paragraph**

Sliding window approach

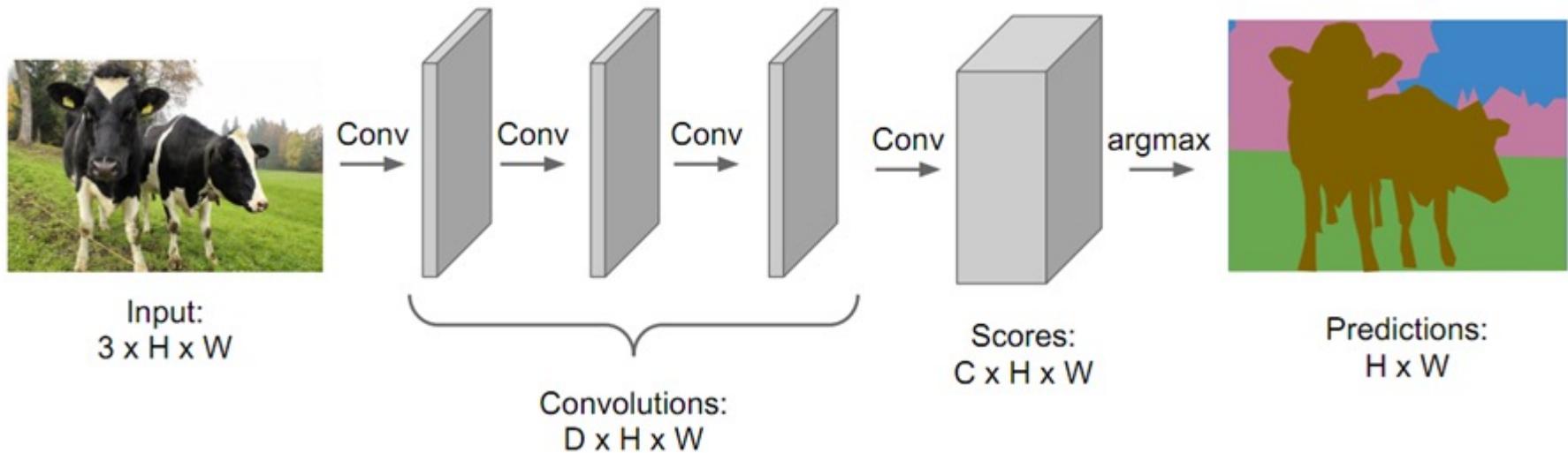


Sliding window approach [2]



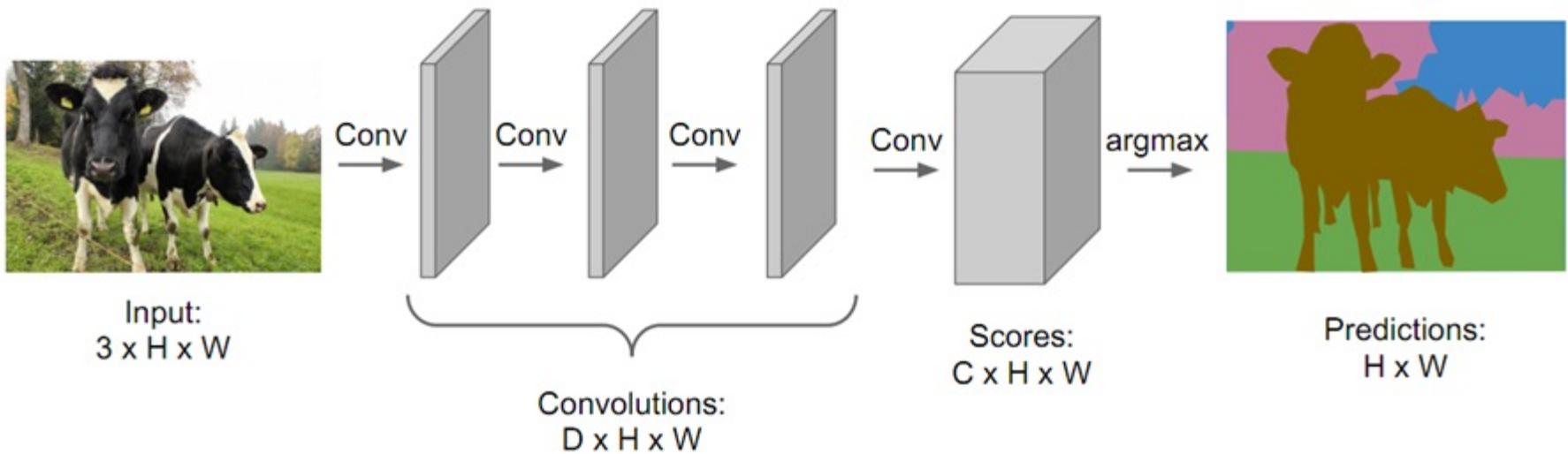
Semantic Segmentation Idea: Convolution

- An intuitive idea: encode the entire image with conv net and do semantic segmentation on top.



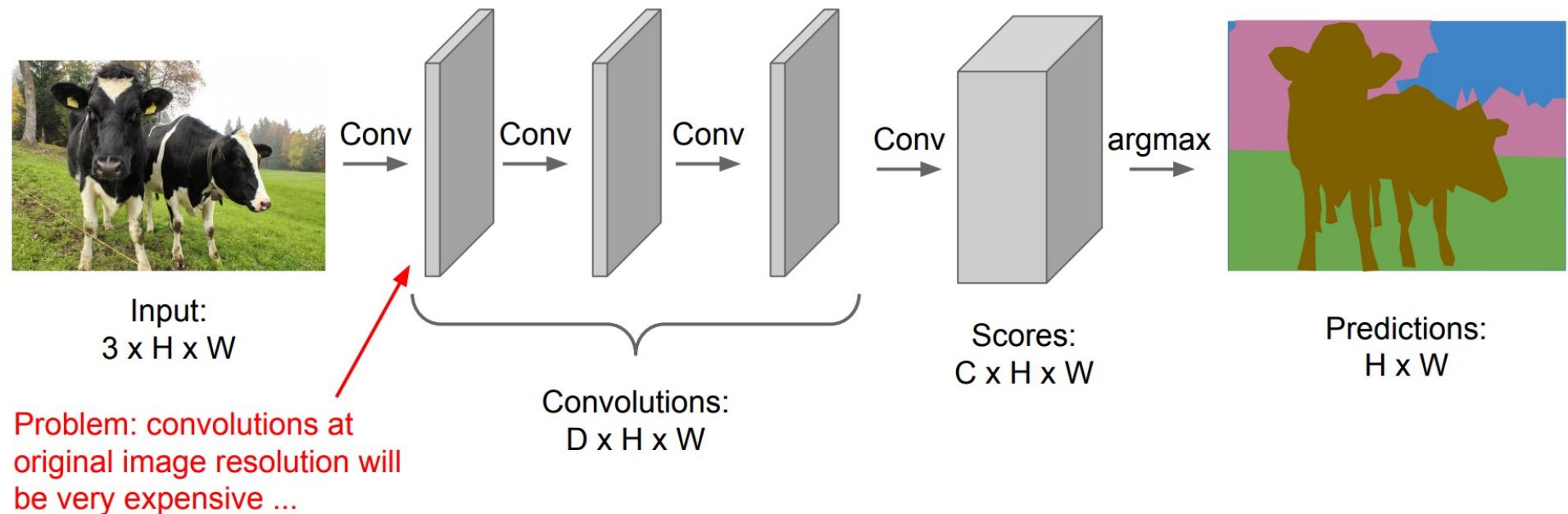
Semantic Segmentation Idea: Convolution

- An intuitive idea: encode the entire image with conv net and do semantic segmentation on top.
- Problem: classification architectures often reduce feature spatial sizes to go deeper, but semantic segmentation requires the output size to be the same as input size.



Semantic Segmentation Idea: Convolution

Design a network with only convolutional layers without downsampling operators to make predictions for pixels all at once!

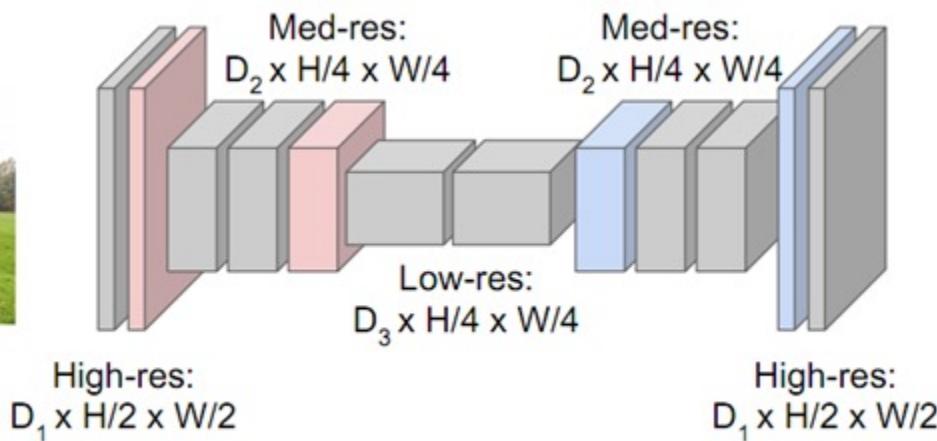


Semantic Segmentation Idea: Convolution

- Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!



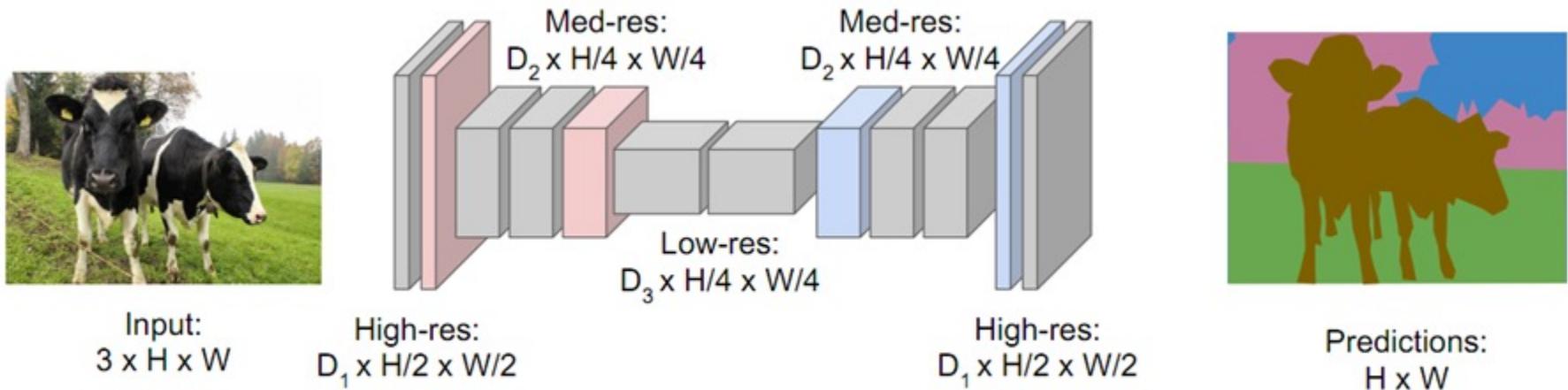
Input:
 $3 \times H \times W$



Predictions:
 $H \times W$

Semantic Segmentation Idea: Convolution

- Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!
- **Downsampling:** Pooling, strided convolution
- **Upsampling:** ???



Upsampling layers

Unpooling

- These layers have no parameters

Nearest Neighbor

1	2
3	4



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

Output: 4 x 4

Input: 2 x 2

“Bed of Nails”

1	2
3	4



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

Output: 4 x 4

Input: 2 x 2

Max 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

Max Unpooling

Use positions from pooling layer

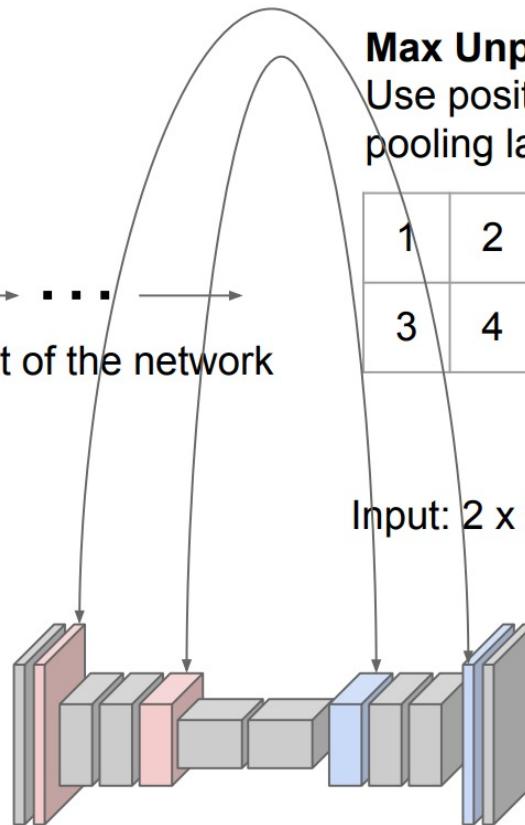
1	2
3	4

Input: 2 x 2

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

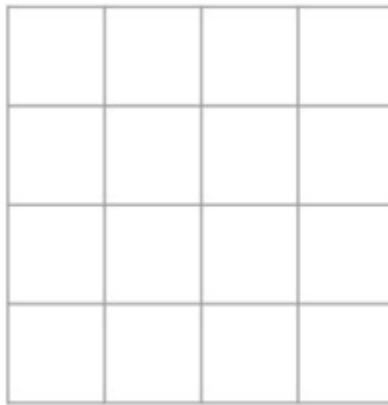
Output: 4 x 4

Corresponding pairs of
downsampling and
upsampling layers

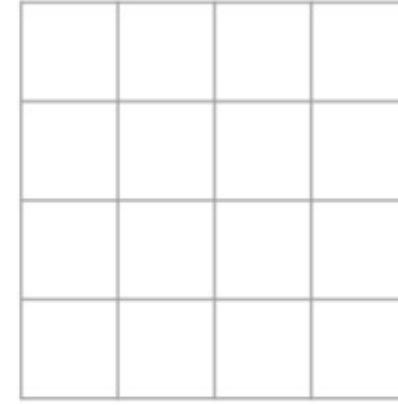


Learnable Upsampling: Transposed Convolution

- Recall: Normal 3×3 convolution, stride 1 pad 1

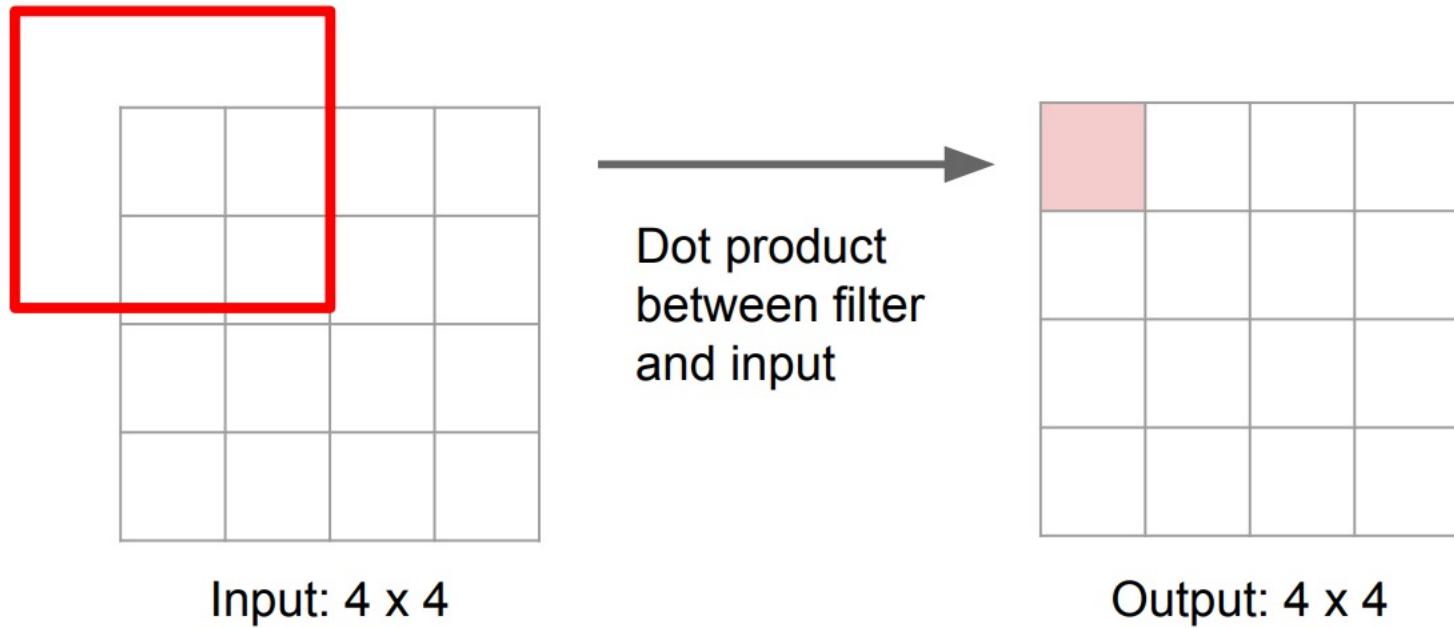


Input: 4×4

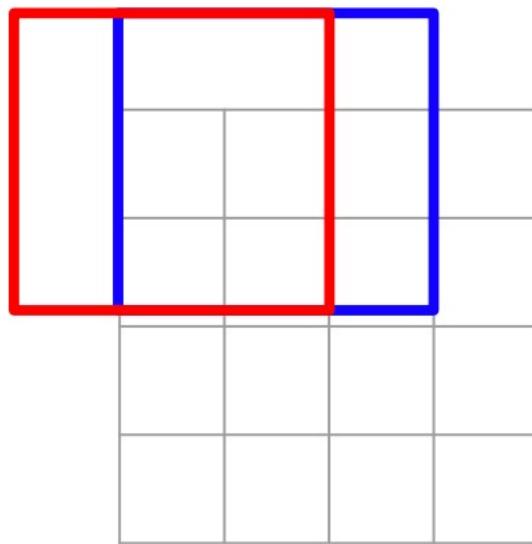


Output: 4×4

Recall: Normal 3 x 3 convolution, stride 1 pad 1

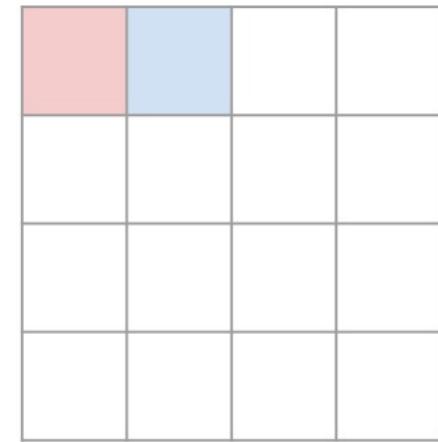


Recall: Normal 3×3 convolution, stride 1 pad 1



Input: 4×4

Dot product
between filter
and input

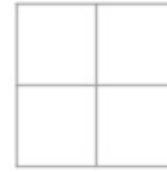


Output: 4×4

Recall: Normal 3 x 3 convolution, stride 2 pad 1

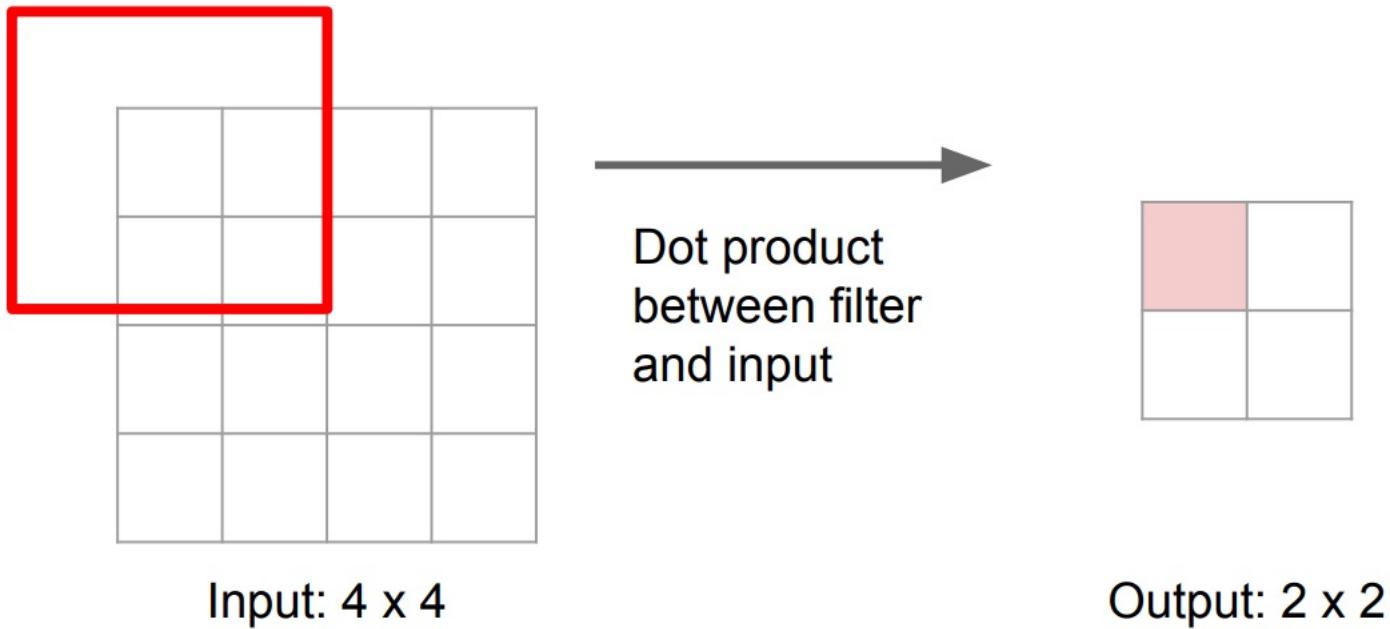


Input: 4 x 4

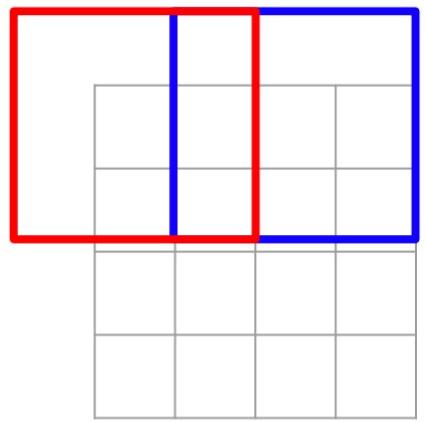


Output: 2 x 2

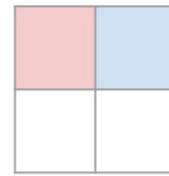
Recall: Normal 3×3 convolution, stride 2 pad 1



Recall: Normal 3 x 3 convolution, stride 2 pad 1



Dot product
between
filter
and input



Output: 2 x 2

Filter moves 2 pixels in
the input for every one
pixel in the output

Stride gives ratio between
movement in input and
output

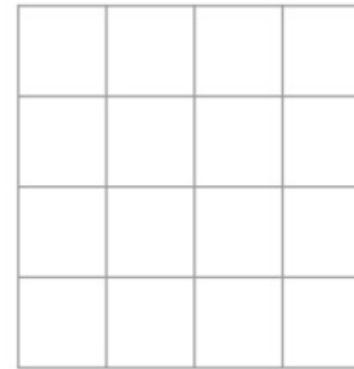
We can interpret strided
convolution as “learnable
downsampling”.

Learnable Upsampling: Transposed Convolution

- It is an upsampling that contains trainable parameters
- 3×3 transpose convolution, stride 2 pad 1



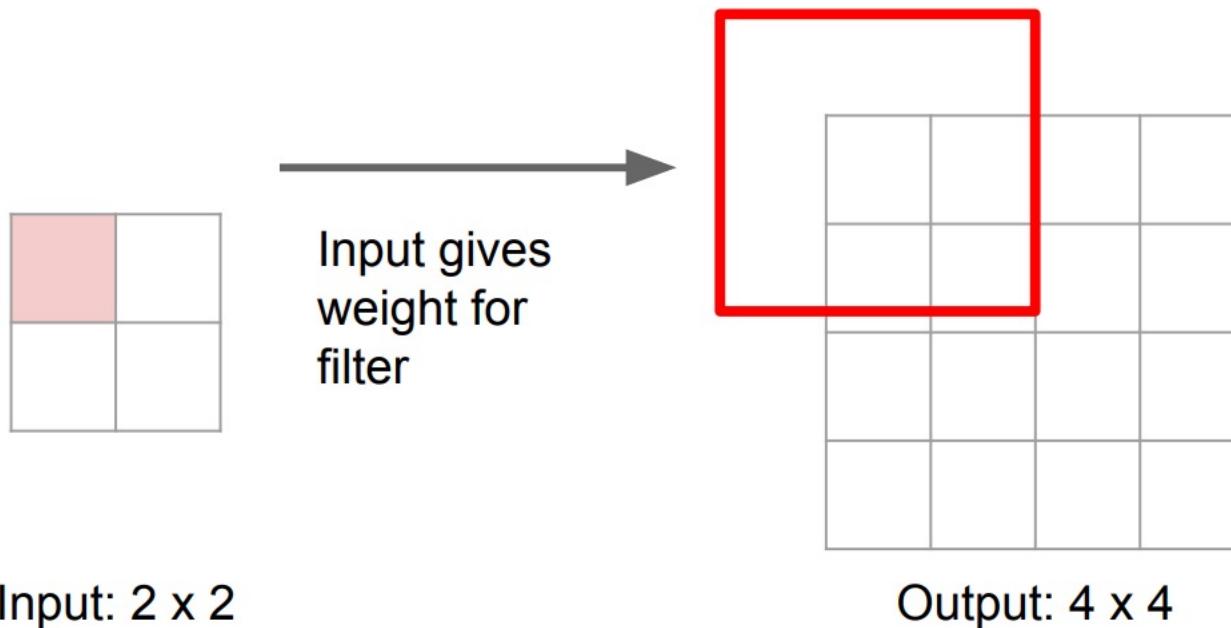
Input: 2×2



Output: 4×4

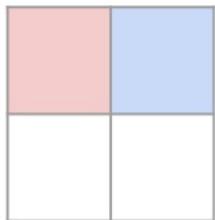
Learnable Upsampling: Transposed Convolution

3 x 3 transpose convolution, stride 2 pad 1

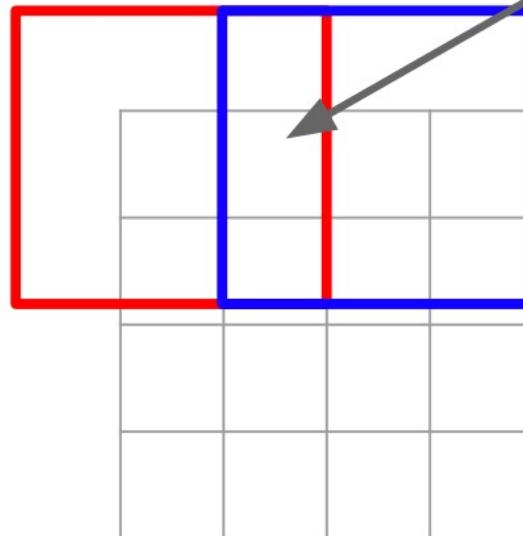


Learnable Upsampling: Transposed Convolution

3 x 3 transpose convolution, stride 2 pad 1



Input gives weight for filter



Output: 4 x 4

Sum where output overlaps

Filter moves 2 pixels in the output for every one pixel in the input

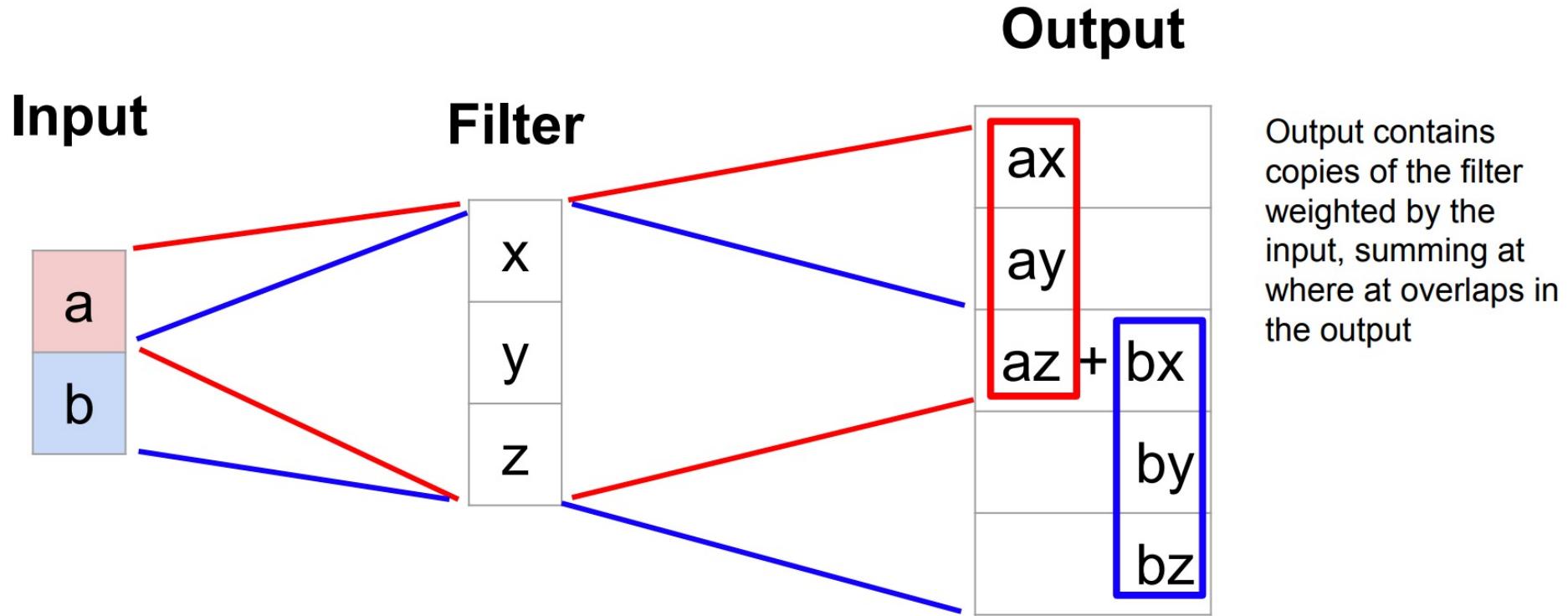
Stride gives ratio between movement in output and input

Input: 2 x 2

Learnable Upsampling: Transposed Convolution

- Other names?:
 - Deconvolution (should not, easy to misunderstand)
 - Upconvolution
 - Fractionally strided convolution
 - Backward strided convolution

Learnable Upsampling: 1D Example



Convolution as Matrix Multiplication (1D Example)

We can express convolution in terms of a matrix multiplication

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

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

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

Convolution transpose multiplies by the transpose of the same matrix:

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

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

Example: 1D transpose conv, kernel size=3, stride=2, padding=0

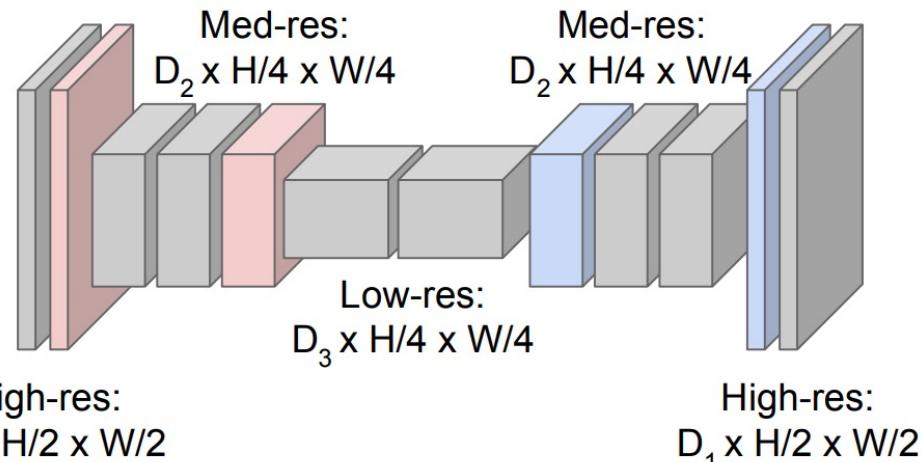
Recap: Semantic Segmentation Idea

Downsampling:
Pooling, strided convolution



Input:
 $3 \times H \times W$

High-res:
 $D_1 \times H/2 \times W/2$



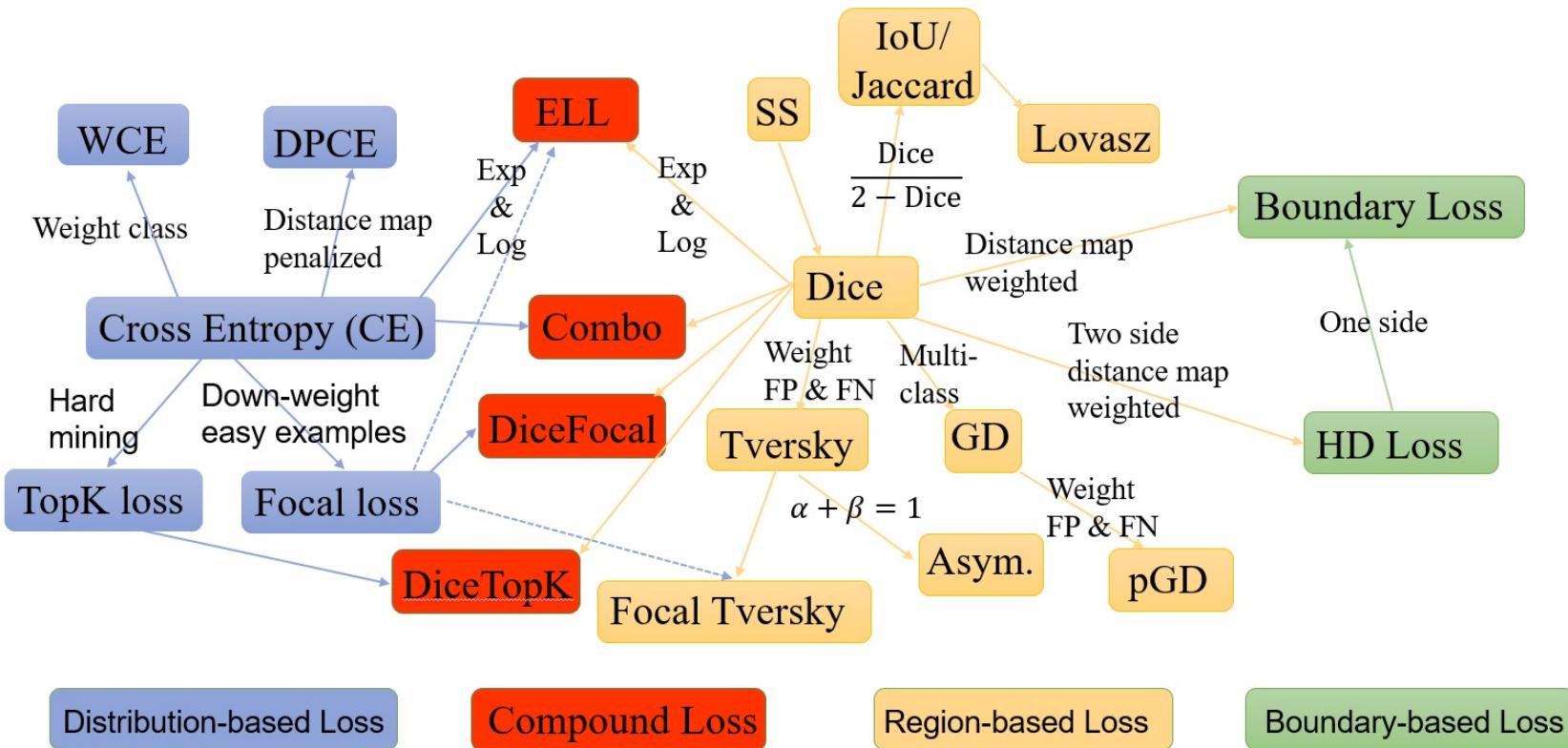
Upsampling:
Unpooling or strided transpose convolution

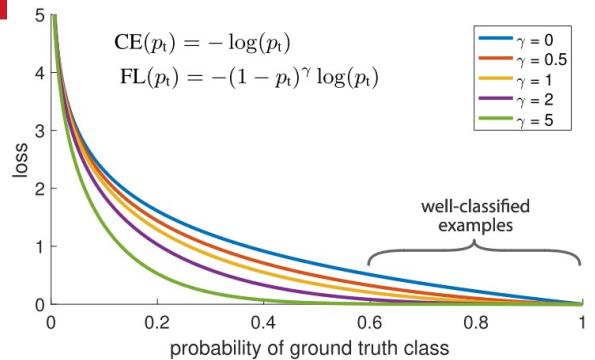
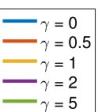


Predictions:
 $H \times W$

Objective functions for image segmentation

Objective functions





Distribution-based objective functions

- Cross Entropy (CE):

$$\text{CE}(p, \hat{p}) = -(p \log(\hat{p}) + (1 - p) \log(1 - \hat{p}))$$

- Weighted CE: Each class has a different weight

$$\text{WCE}(p, \hat{p}) = -(\beta p \log(\hat{p}) + (1 - p) \log(1 - \hat{p}))$$

- Focal loss: solve the problem of large imbalance between the background layer and the object layer of interest. The objective function value for the easy to classify objects is reduced so that the network focuses more on difficult objects.

$$\text{FL}(p, \hat{p}) = -(\alpha(1 - \hat{p})^\gamma p \log(\hat{p}) + (1 - \alpha)\hat{p}^\gamma(1 - p) \log(1 - \hat{p}))$$

Area-based objective functions

- Dice coefficient and IoU:

$$\text{DC} = \frac{2TP}{2TP + FP + FN} = \frac{2|X \cap Y|}{|X| + |Y|}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}$$

- Dice loss: $\text{DL}(p, \hat{p}) = 1 - \frac{2p\hat{p} + 1}{p + \hat{p} + 1}$
- Tversky loss:

$$\text{TI}(p, \hat{p}) = \frac{p\hat{p}}{p\hat{p} + \beta(1-p)\hat{p} + (1-\beta)p(1-\hat{p})}$$

Combined objective functions

- Dice loss + CE:

$$\text{CE}(p, \hat{p}) + \text{DL}(p, \hat{p})$$

- Dice loss + Focal loss

$$\text{CE}(p, \hat{p}) + \text{FL}(p, \hat{p})$$

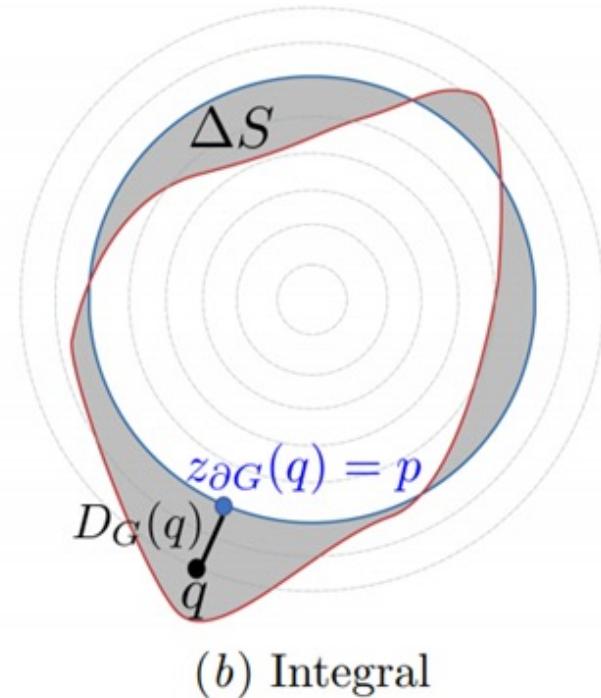
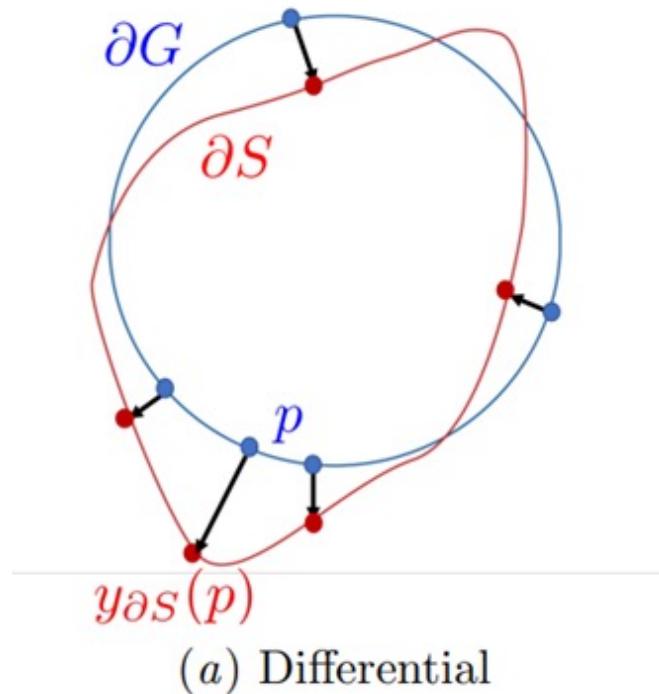
- ...

Boundary loss

distance metric on the space of contours, not regions

$$\text{Dist}(\partial G, \partial S) = \int_{\partial G} \|y_{\partial S}(p) - p\|^2 dp$$

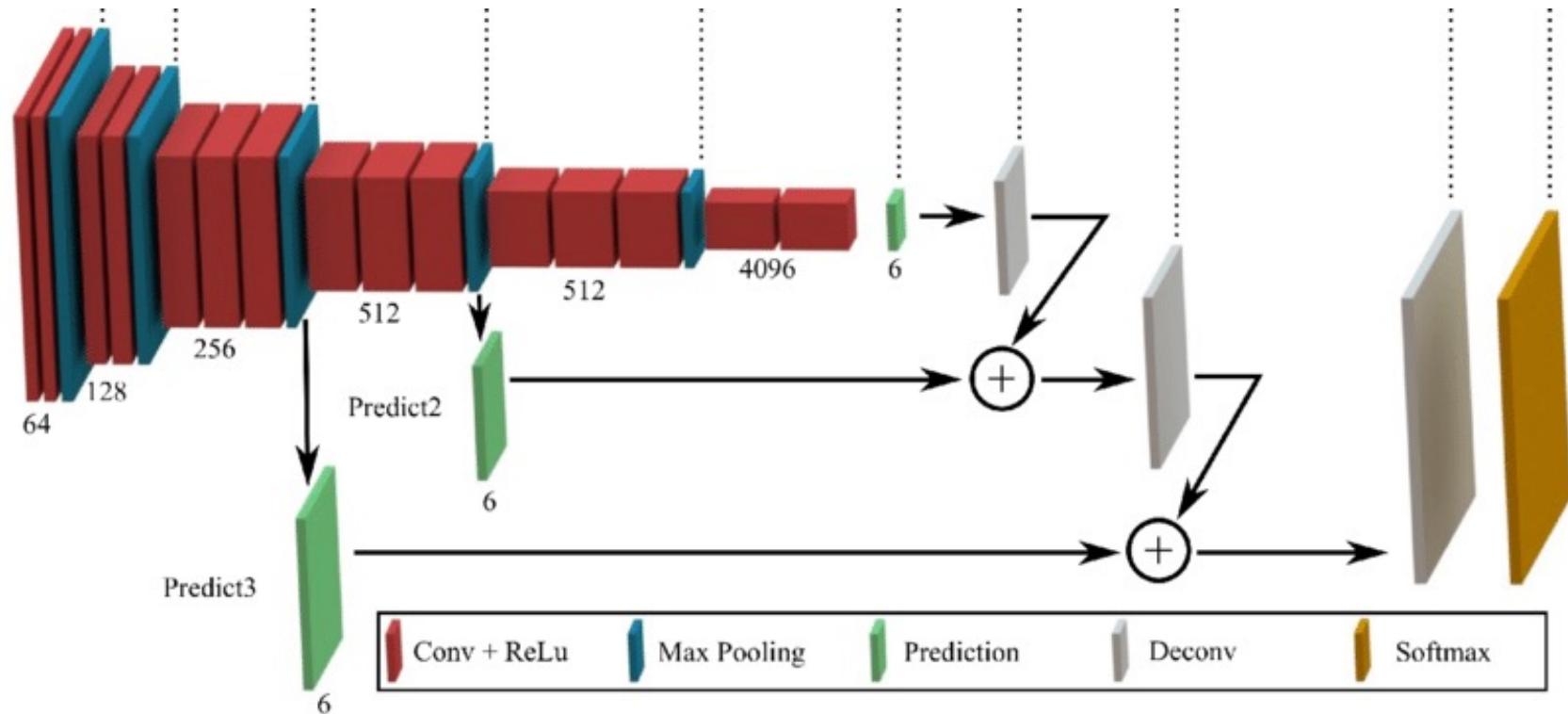
$$\text{Dist}(\partial G, \partial S) = 2 \int_{\Delta S} D_G(q) dq$$



$$\frac{1}{2} \text{Dist}(\partial G, \partial S) = \int_{\Omega} \phi_G(q) s(q) dq - \int_{\Omega} \phi_G(q) g(q) dq$$

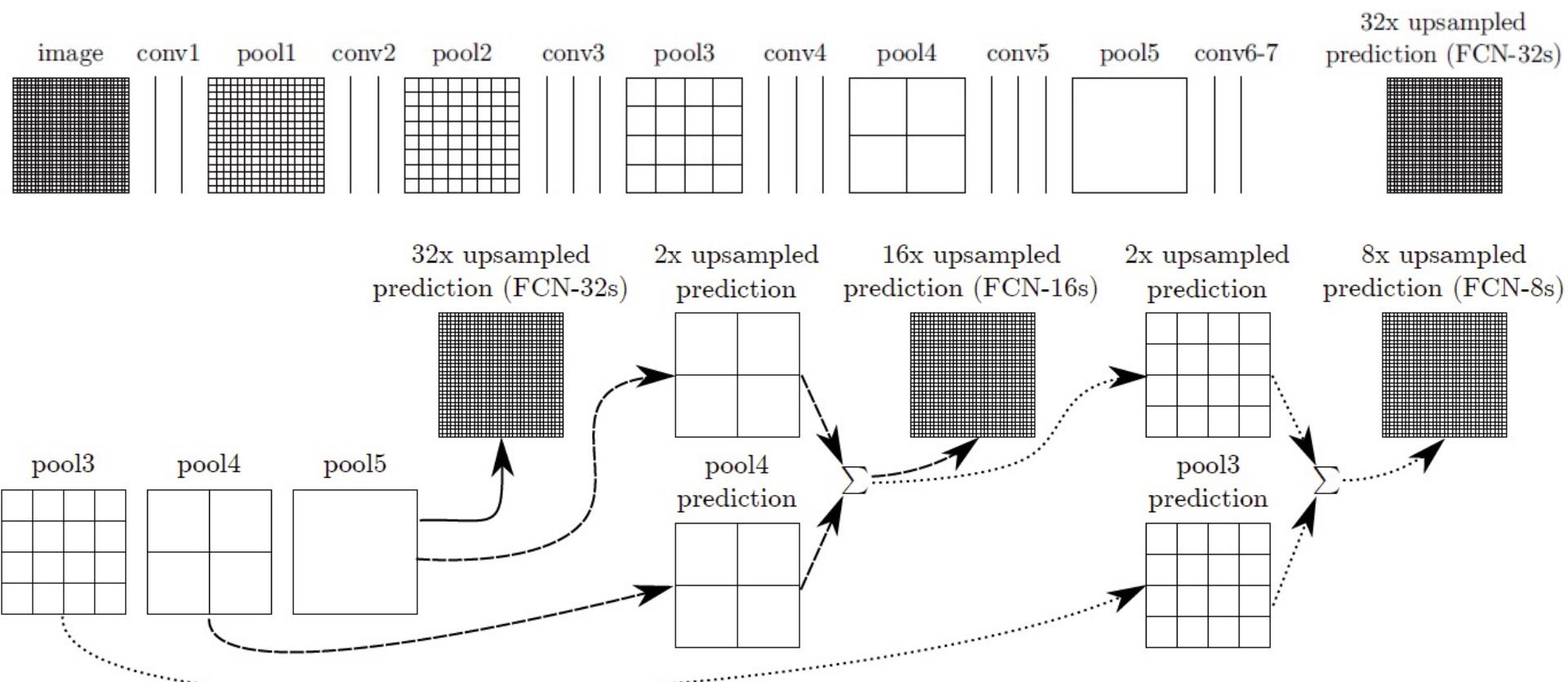
Some typical image segmentation networks

FCN with 2 shortcuts

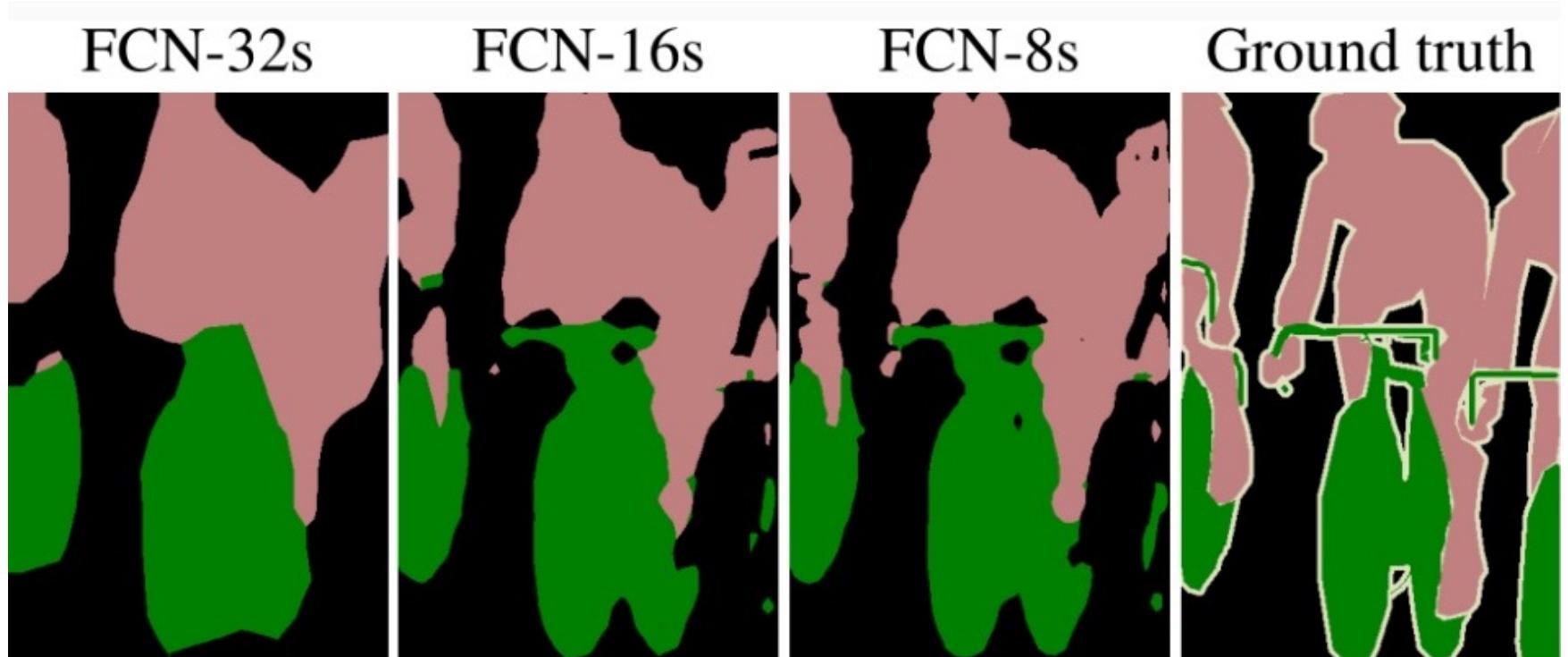


FCN with 2 shortcuts (2)

- Illustrate FCN results with different levels of resolution

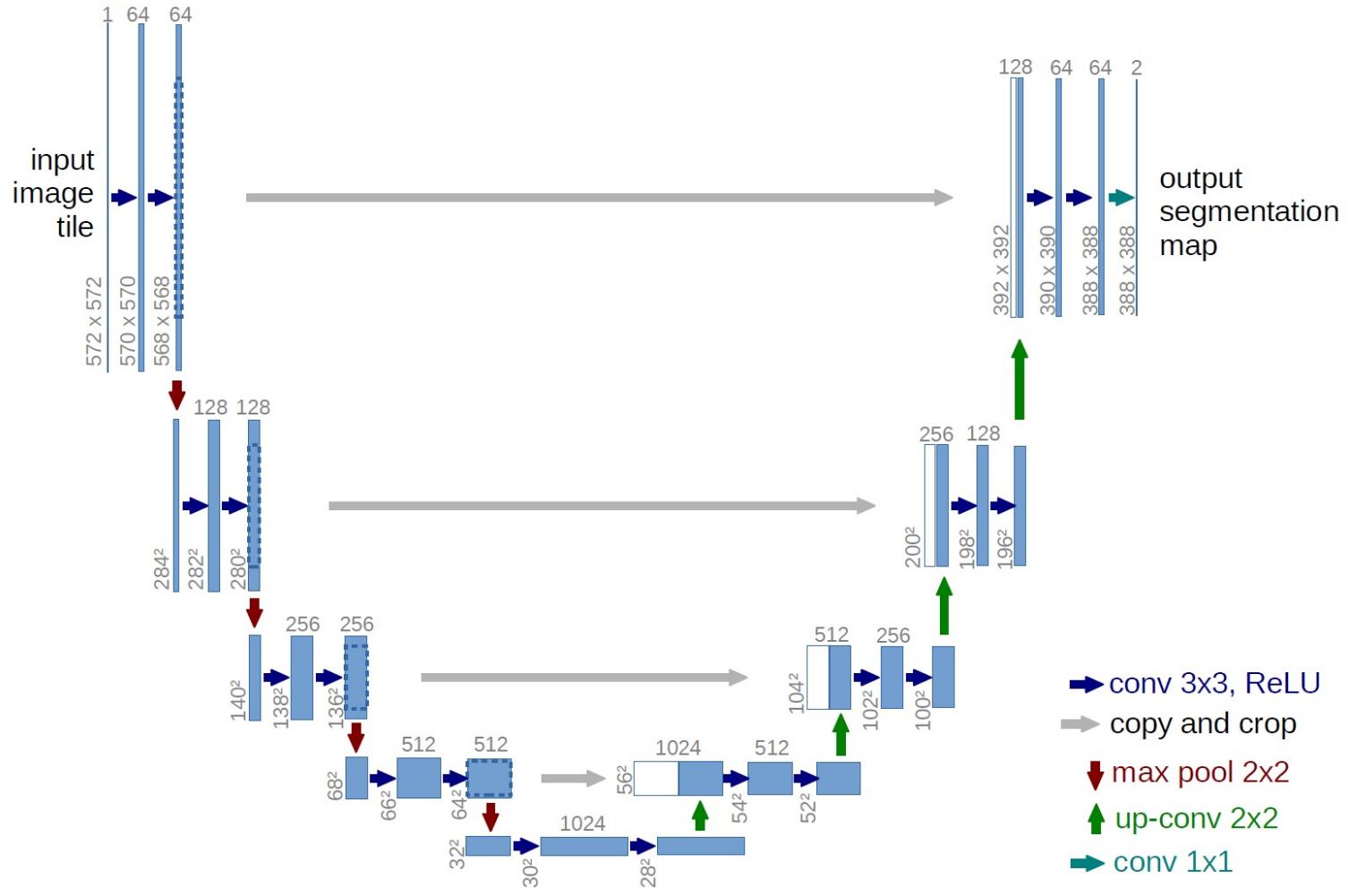


Results

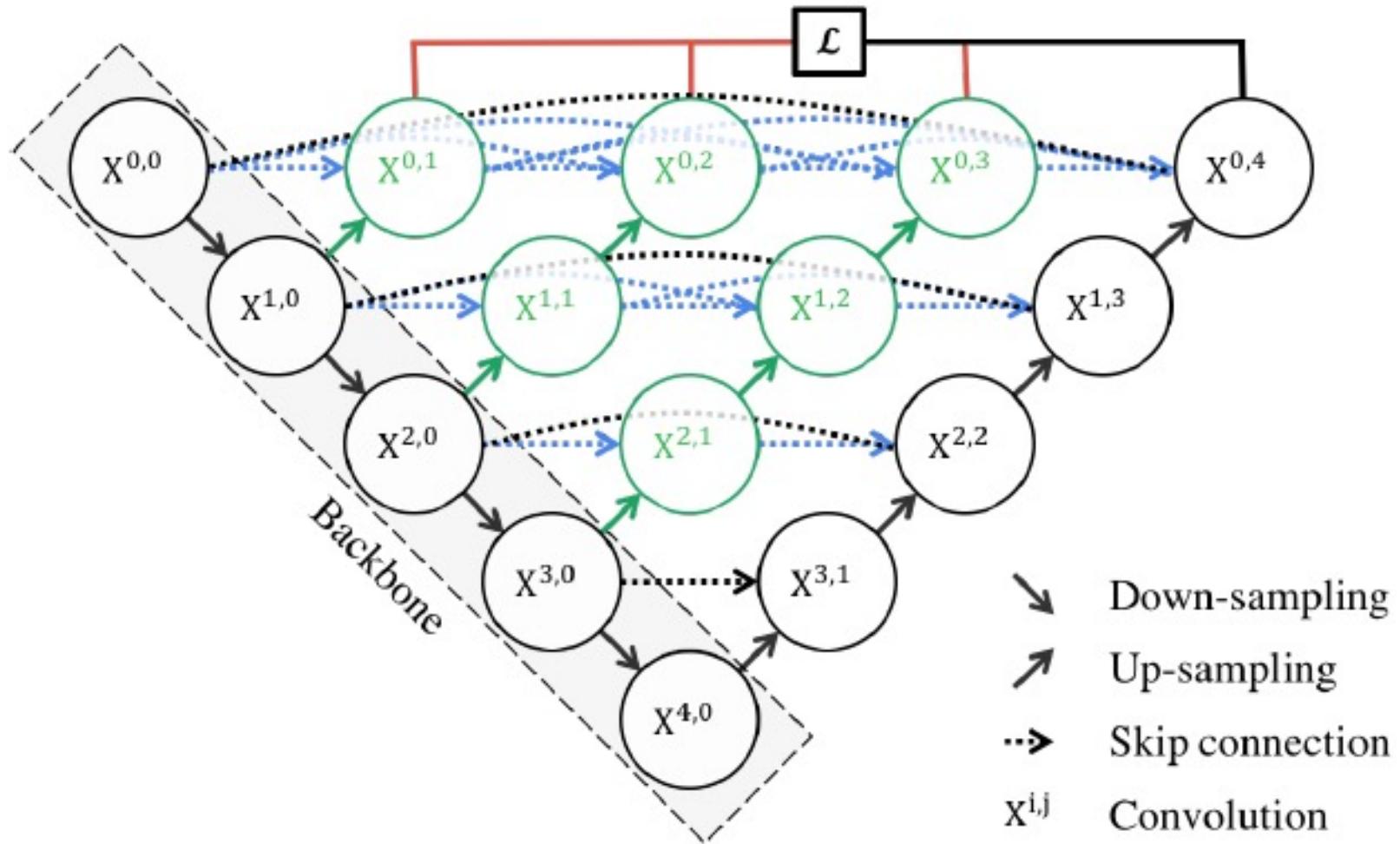


U-Net

- Widely used in medical images

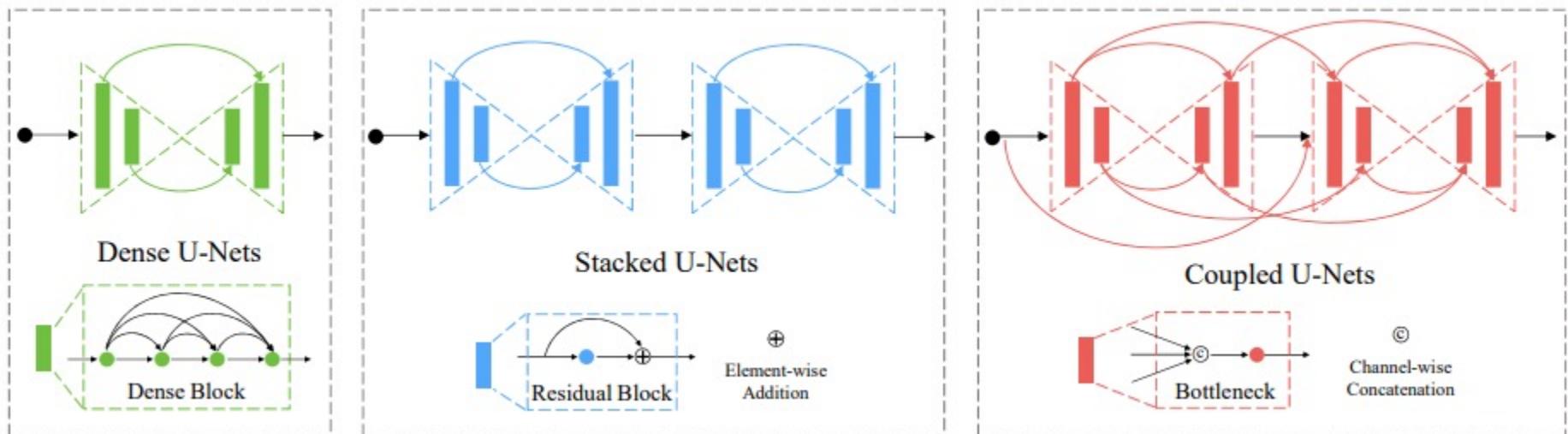


U-Net++



Stacked UNets and CUNets

- Stacked UNets: stack many UNets in series
- CUNets: also stack many UNets in series, but there are additional short connections between UNets



References

1. cs231n:

<http://cs231n.stanford.edu>

2. Losses for segmentation

<https://lars76.github.io/neural-networks/object-detection/losses-for-segmentation/>



25
YEARS ANNIVERSARY
SOICT

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you
for your
attentions!





HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY