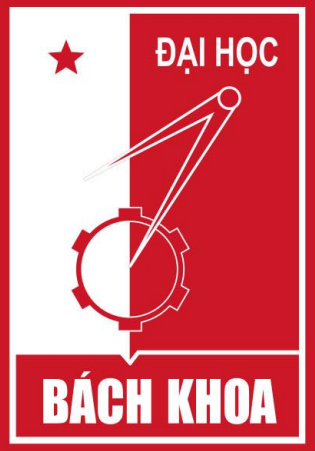


ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

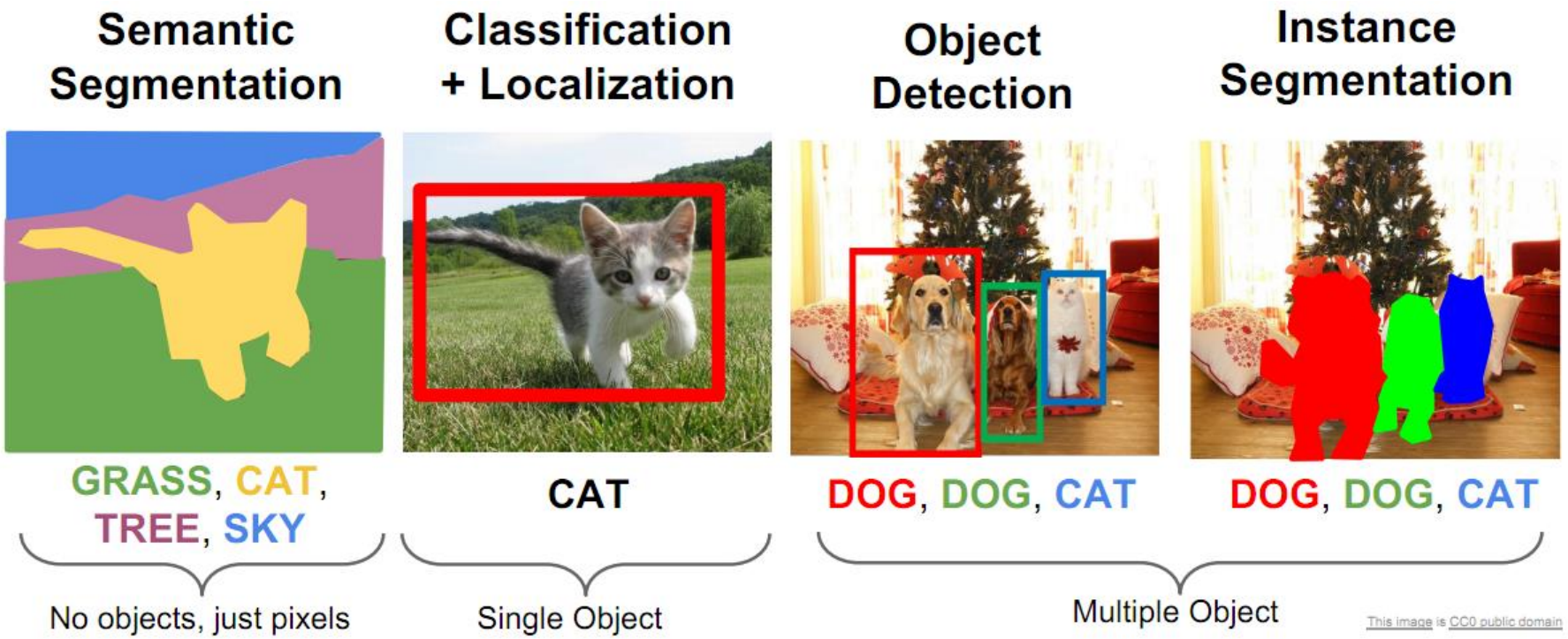
Chương 6:
Một số ứng dụng học sâu trong thị giác máy (Phần 2)

Nội dung

- Giới thiệu bài toán phân đoạn ảnh
- Lớp tăng độ phân giải upsampling
- Hàm mục tiêu
- Một số mạng phân đoạn ảnh tiêu biểu

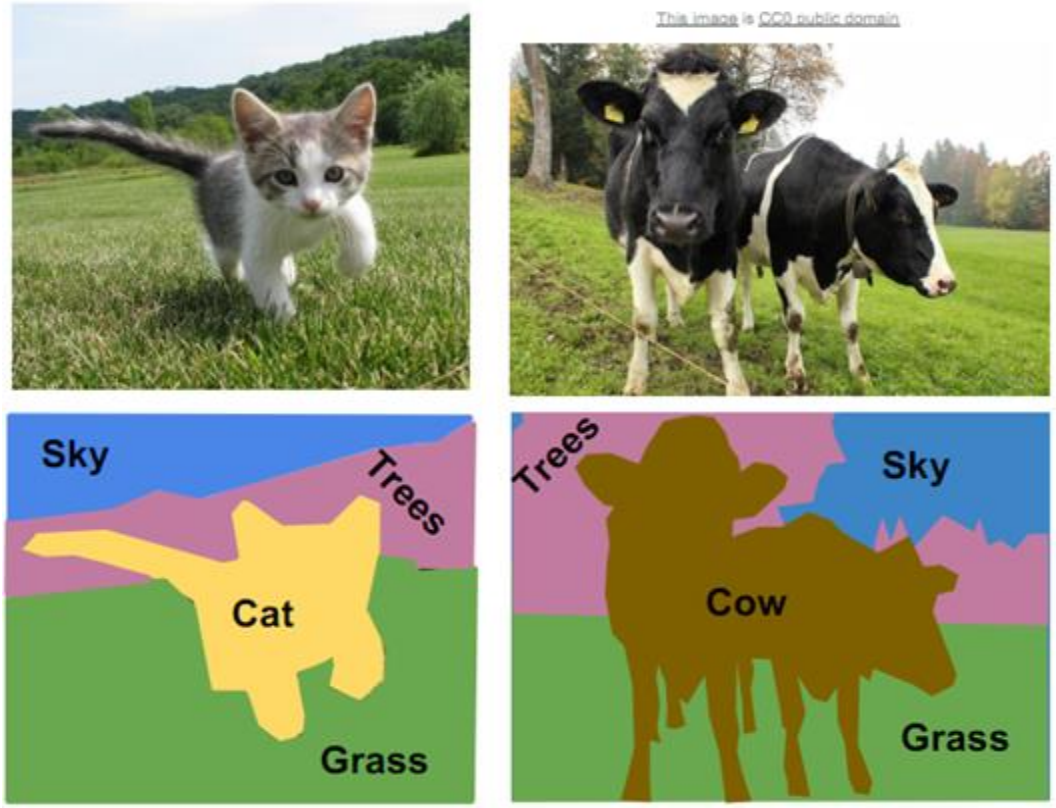
Giới thiệu bài toán phân đoạn ảnh

Các bài toán thị giác máy



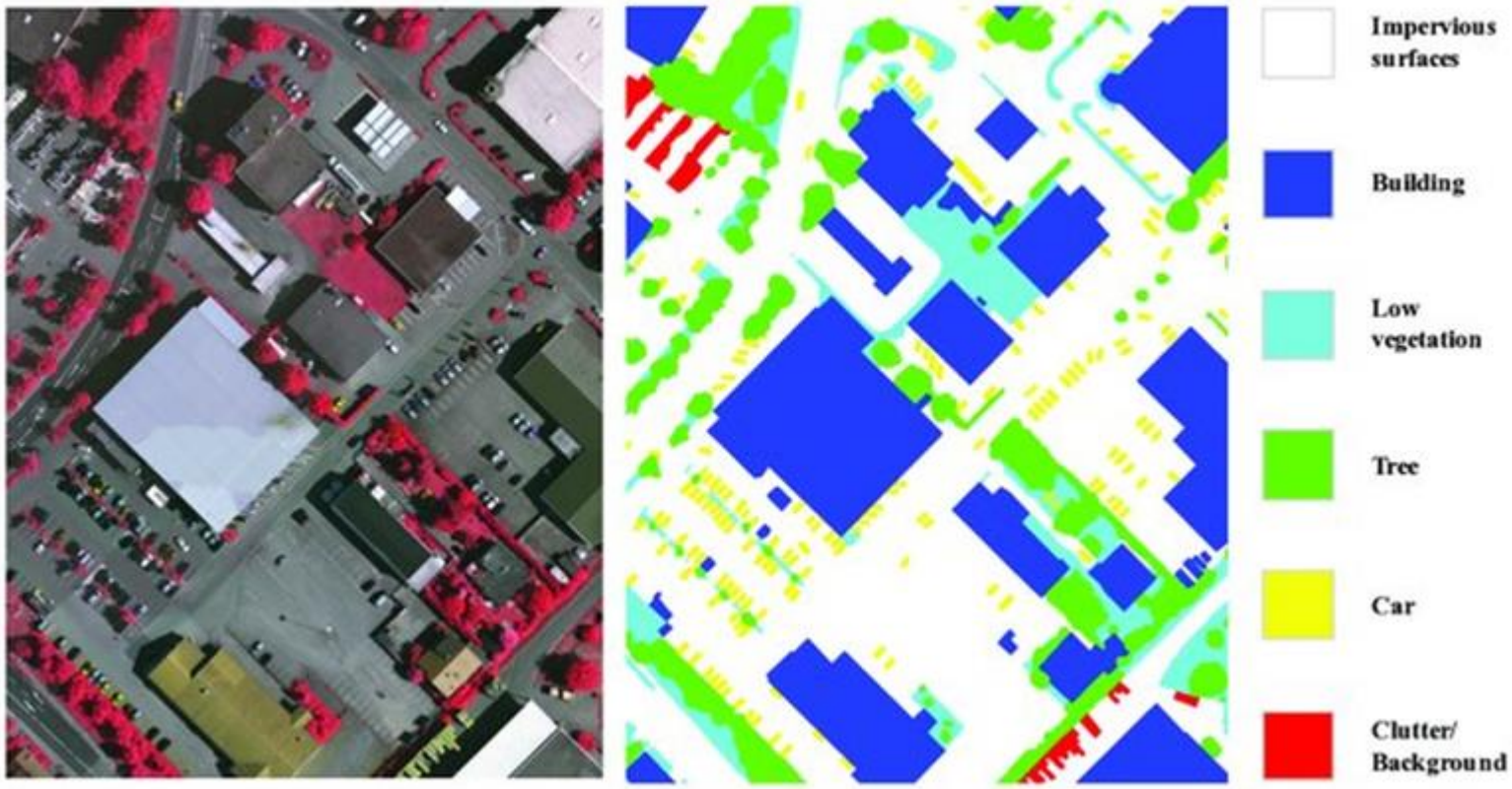
Phân vùng

- Phân lớp từng điểm ảnh trong ảnh
- Không phân biệt các đối tượng cùng lớp trong ảnh



Một số ứng dụng phân đoạn ảnh

- Phân đoạn ảnh vệ tinh và hàng không

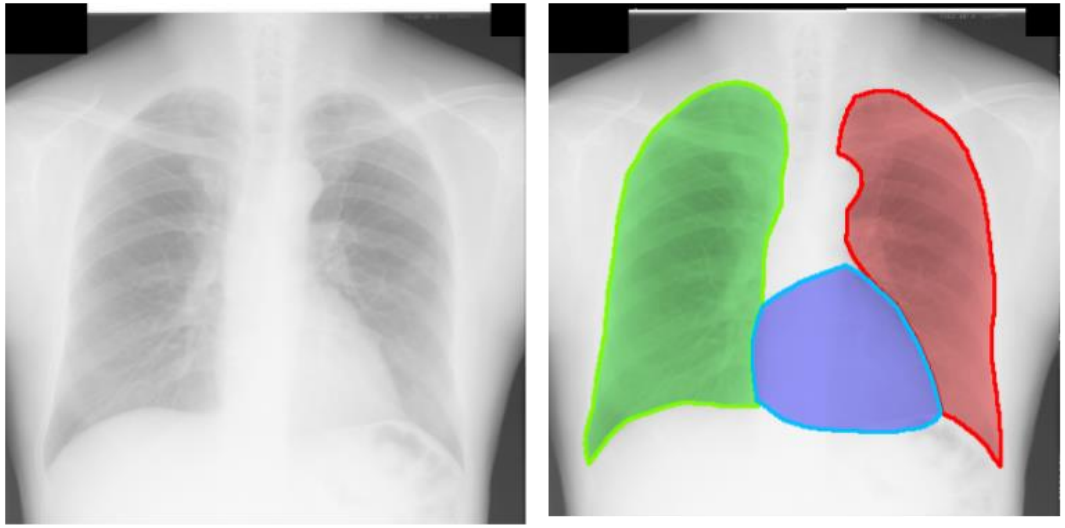


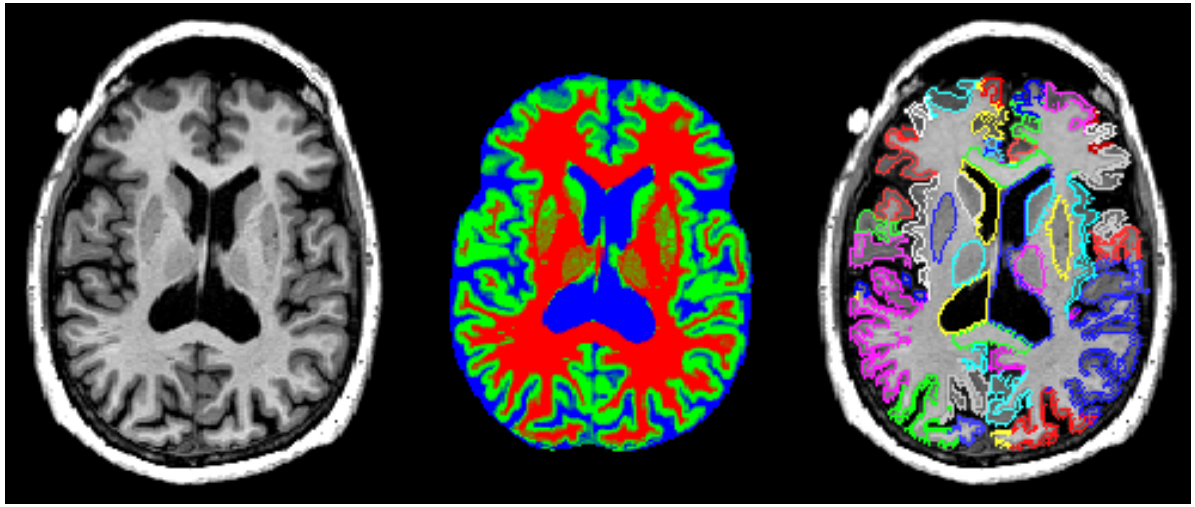
Một số ứng dụng phân đoạn ảnh

- Xe tự hành

Một số ứng dụng phân đoạn ảnh

- Y tế

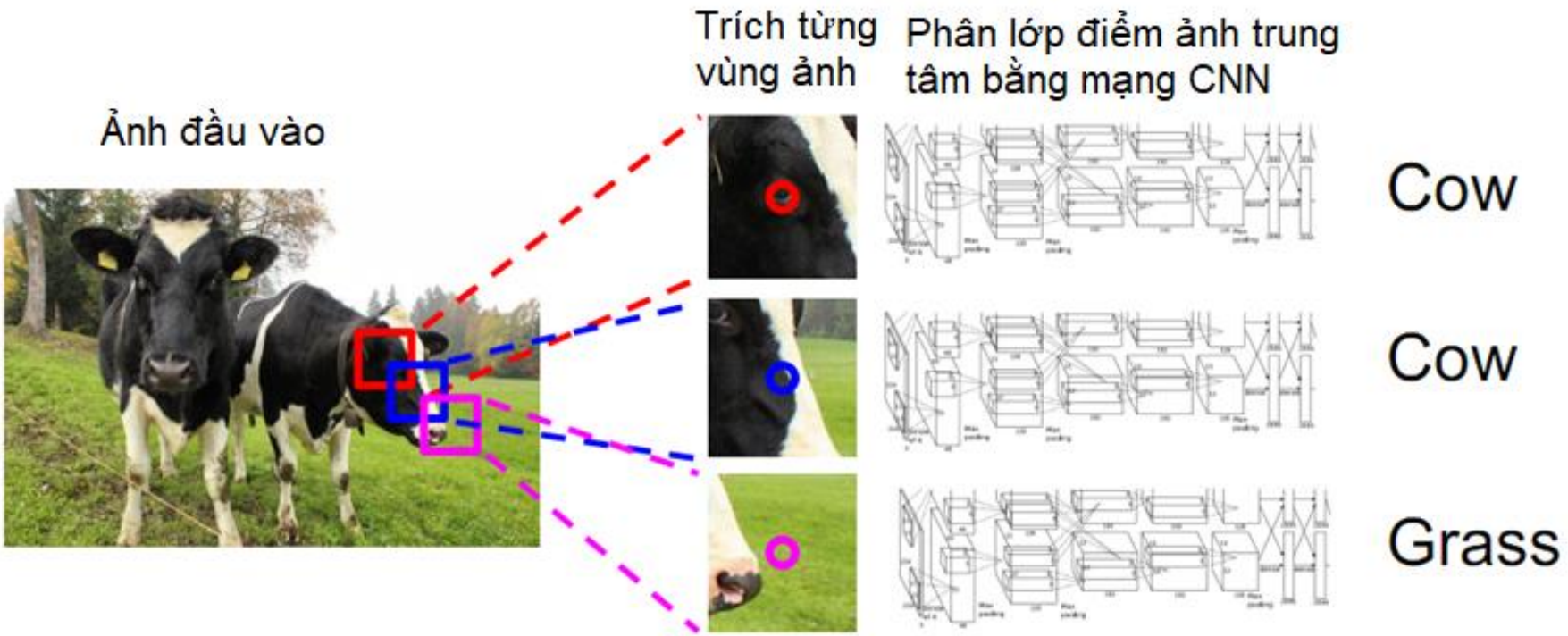




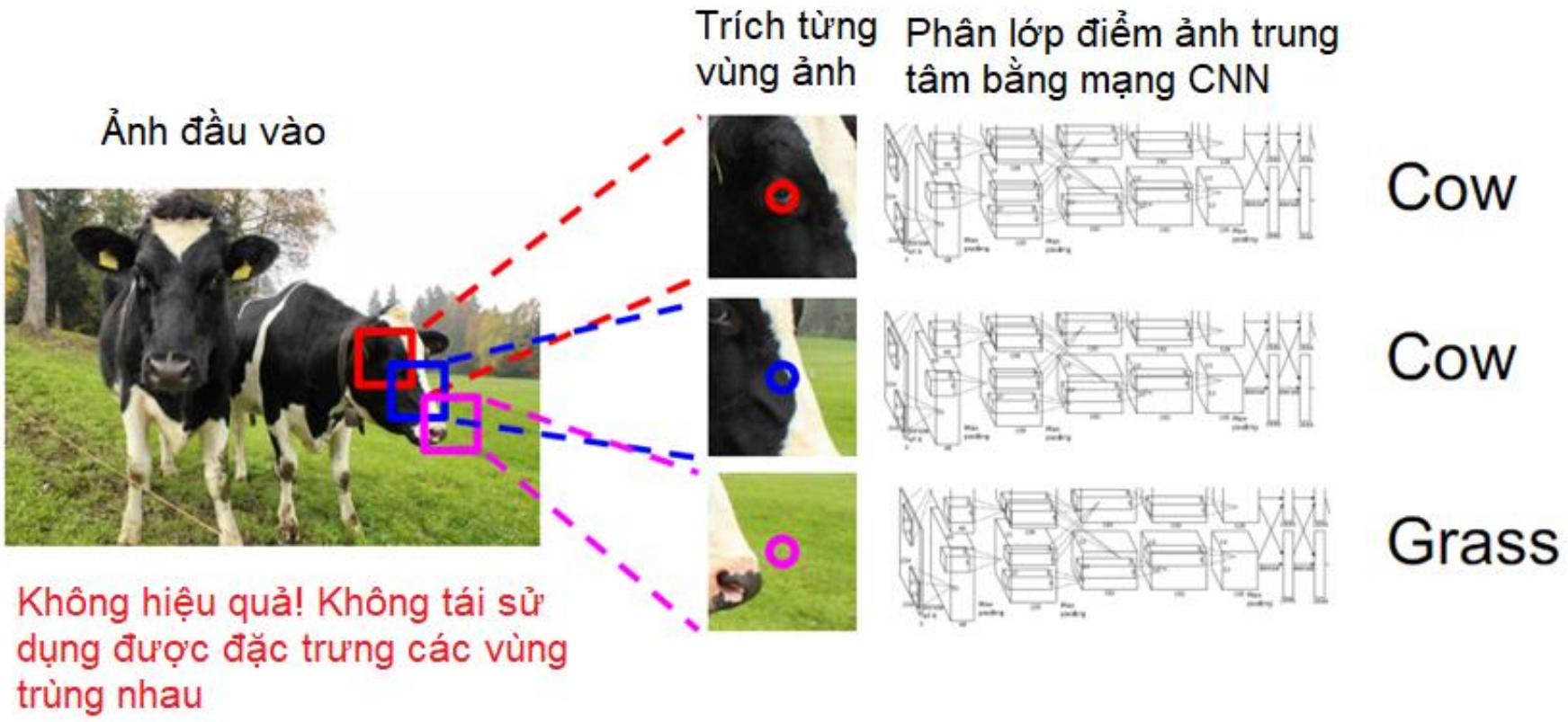
Một số ứng dụng phân đoạn ảnh

- OCR

Trượt cửa sổ

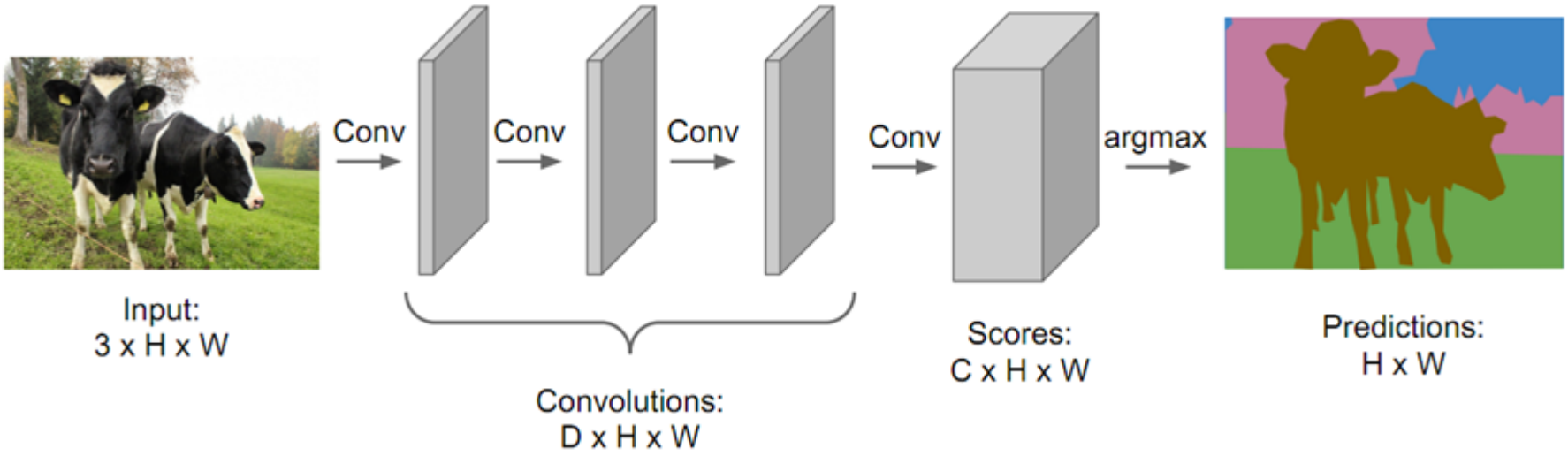

Ảnh đầu vào
Trích từng vùng ảnh
Phân lớp điểm ảnh trung tâm bằng mạng CNN
Cow
Cow
Grass

Trượt cửa sổ


Trích từng vùng ảnh
Phân lớp điểm ảnh trung tâm bằng mạng CNN
Ảnh đầu vào
Cow
Cow
Grass
Không hiệu quả! Không tái sử dụng được đặc trưng các vùng trùng nhau

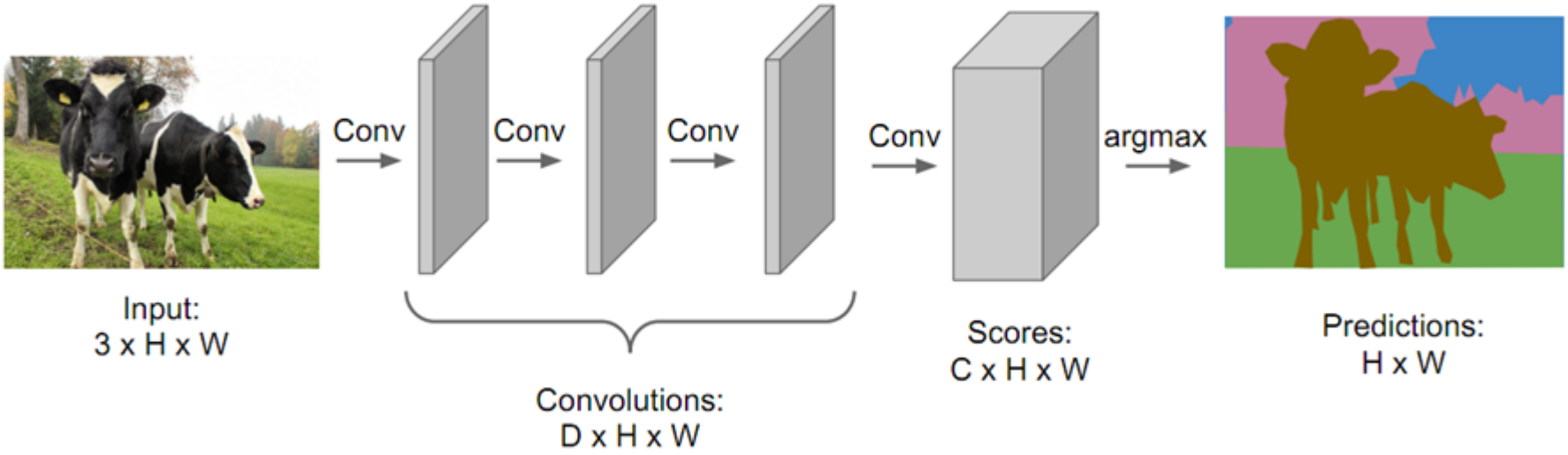
Tích chập hoàn toàn (Fully Convolutional)

- Thiết kế mạng CNN gồm nhiều lớp tích chập để phân lớp đồng thời tất cả các điểm ảnh.



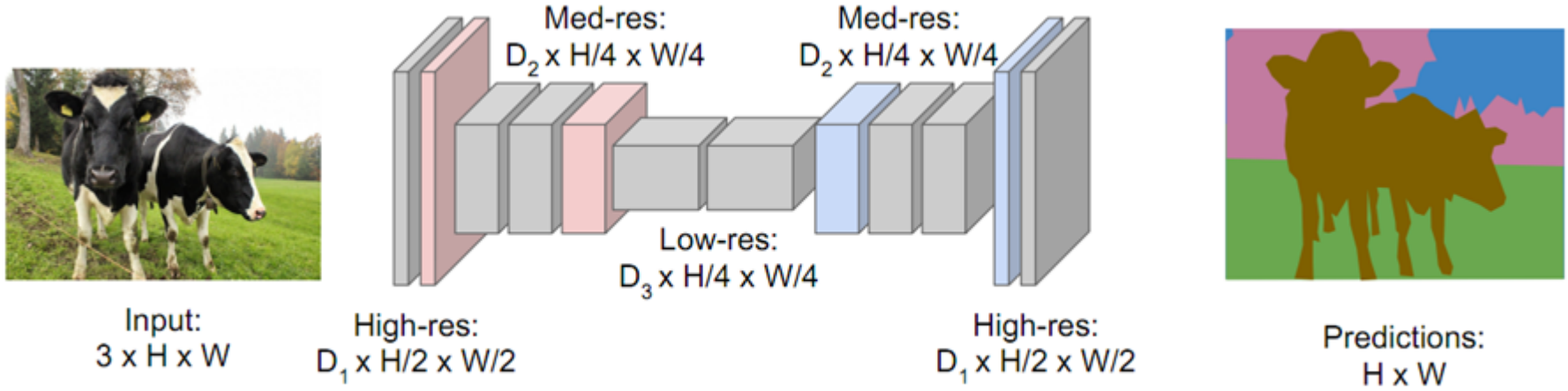
Tích chập hoàn toàn (Fully Convolutional)

- Thiết kế mạng CNN gồm nhiều lớp tích chập để phân lớp đồng thời tất cả các điểm ảnh.
- Vấn đề: Tích chập với các lớp đầu vào có độ phân giải cao đòi hỏi nhiều chi phí tính toán



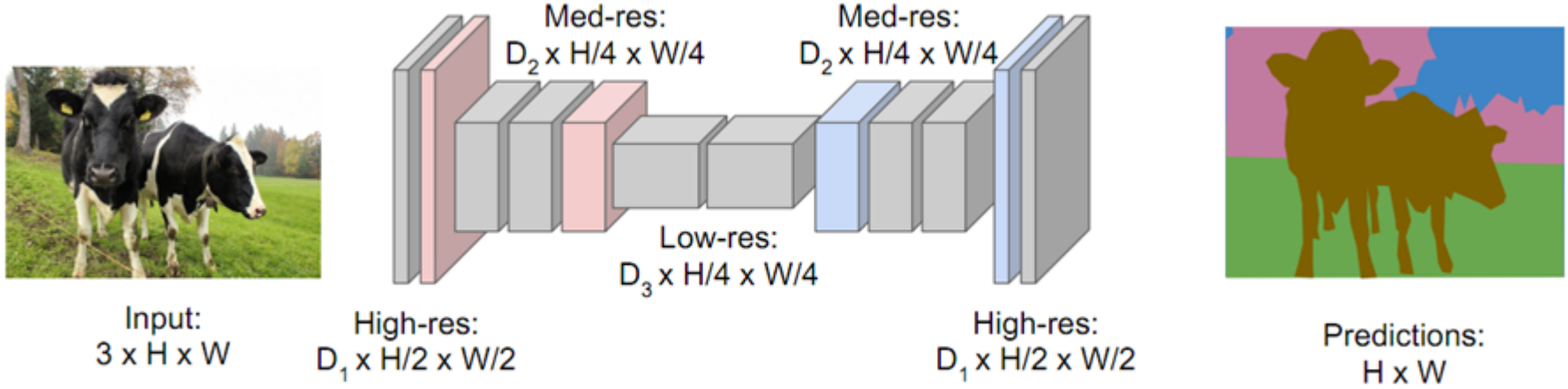
Tích chập hoàn toàn (Fully Convolutional)

- Thiết kế mạng CNN với các lớp giảm độ phân giải (downsampling) và tăng độ phân giải (upsampling)



Tích chập hoàn toàn (Fully Convolutional)

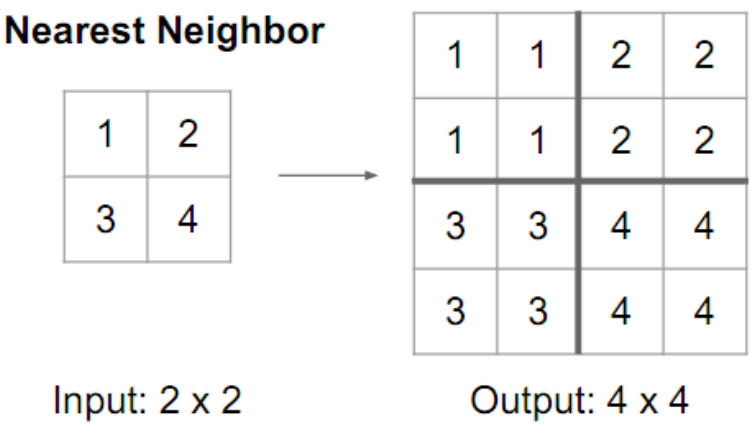
- Thiết kế mạng CNN với các lớp giảm độ phân giải (downsampling) và tăng độ phân giải (upsampling)
- Giảm độ phân giải: max pooling hay strided conv
- Tăng độ phân giải?

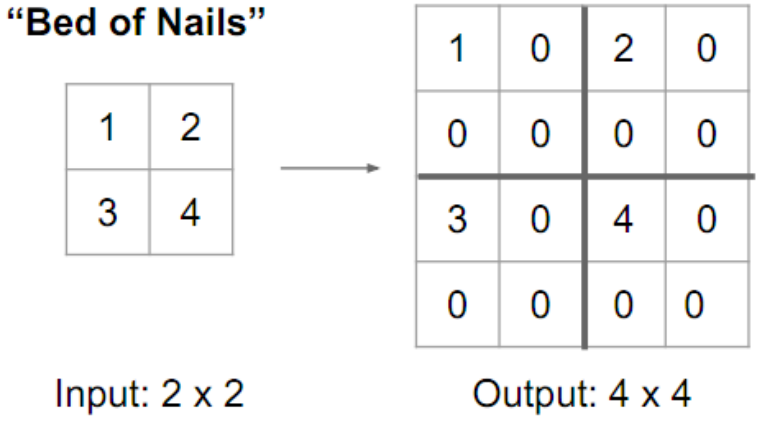


Lớp tăng độ phân giải upsampling

Lớp Unpooling

- Các lớp này không có tham số





Lớp Max Unpooling

Max Pooling

Ghi nhớ vị trí phần tử lớn nhất

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

Sử dụng vị trí đã ghi nhớ khi pooling

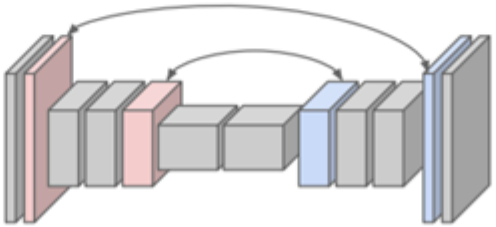
1	2
3	4

Input: 2 x 2

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

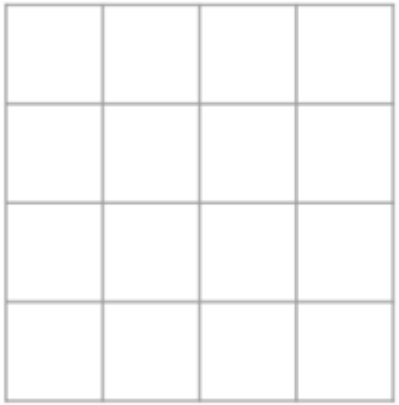
Output: 4 x 4

Các cặp max pooling và max unpooling được dùng đối xứng nhau trong mạng

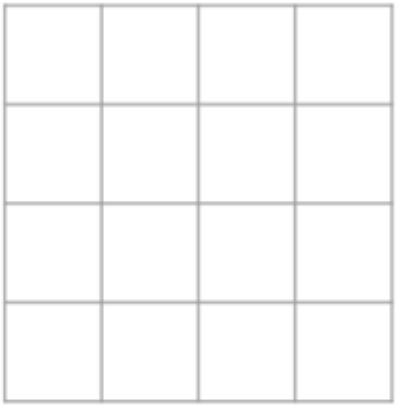


Tích chập chuyển vị (Transposed convolution)

- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



Input: 4 x 4

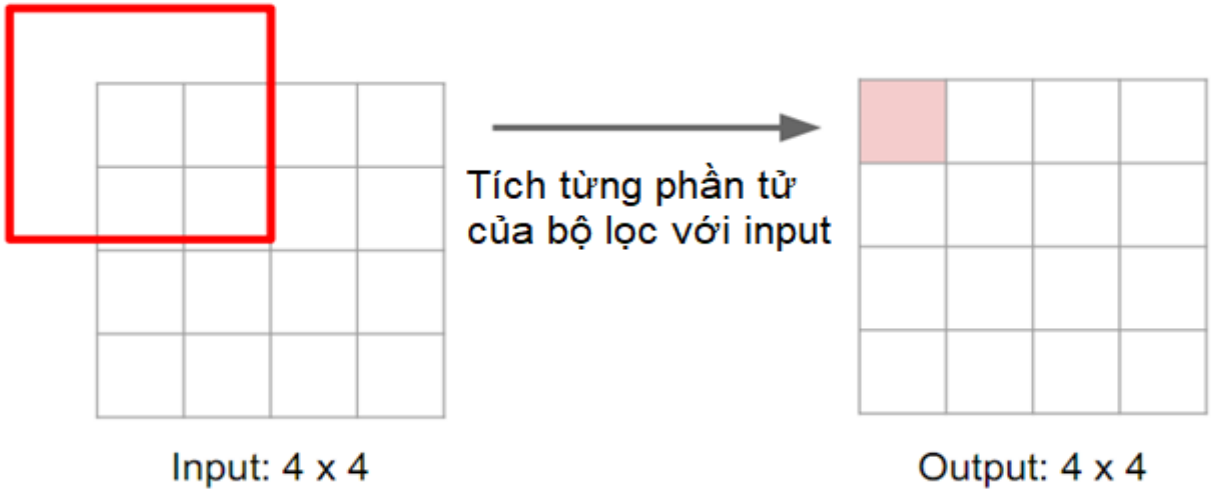


Output: 4 x 4

- Xem lại ví dụ tích chập conv 3x3, bước nhảy stride 1 và thêm viền padding 1

Tích chập chuyển vị

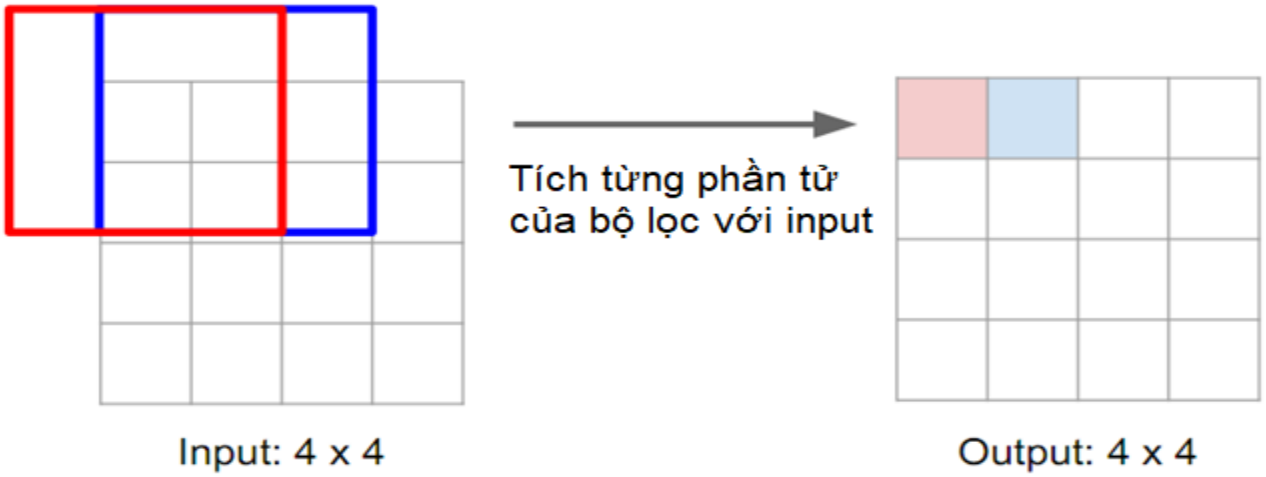
- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



- Xem lại ví dụ tích chập conv 3x3, bước nhảy stride 1 và thêm viền padding 1

Tích chập chuyển vị

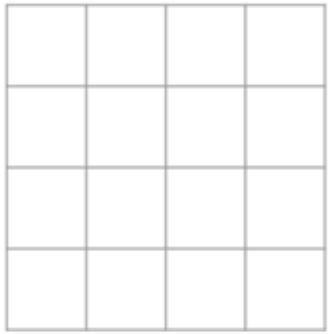
- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



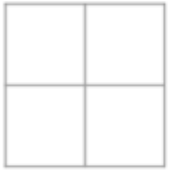
- Xem lại ví dụ tích chập conv 3x3, bước nhảy stride 1 và thêm viền padding 1

Tích chập chuyển vị

- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



Input: 4 x 4

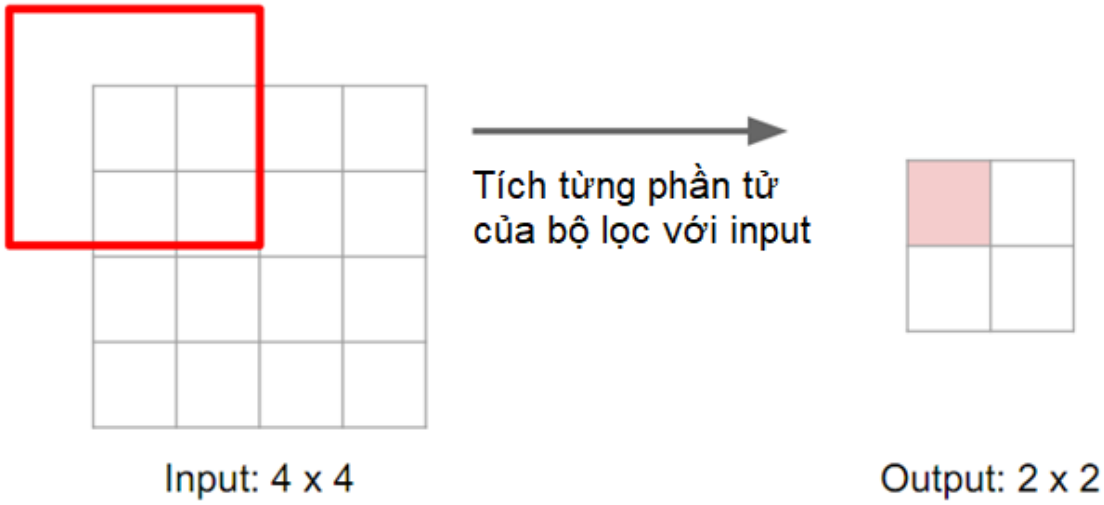


Output: 2 x 2

- Xem lại ví dụ tích chập conv 3x3, bước nhảy stride 2 và thêm viền padding 1

Tích chập chuyển vị

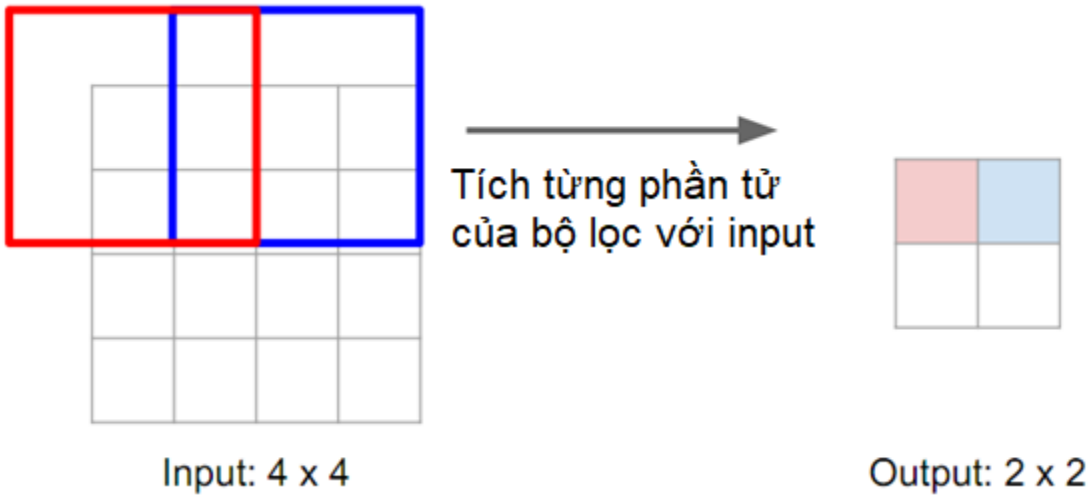
- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



- Xem lại ví dụ tích chập conv 3x3, bước nhảy stride 2 và thêm viền padding 1

Tích chập chuyển vị

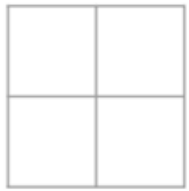
- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



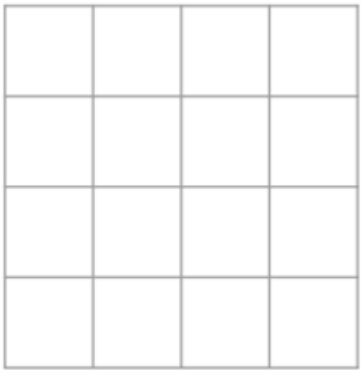
- Xem lại ví dụ tích chập conv 3x3, bước nhảy stride 2 và thêm viền padding 1

Tích chập chuyển vị

- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



Input: 2 x 2

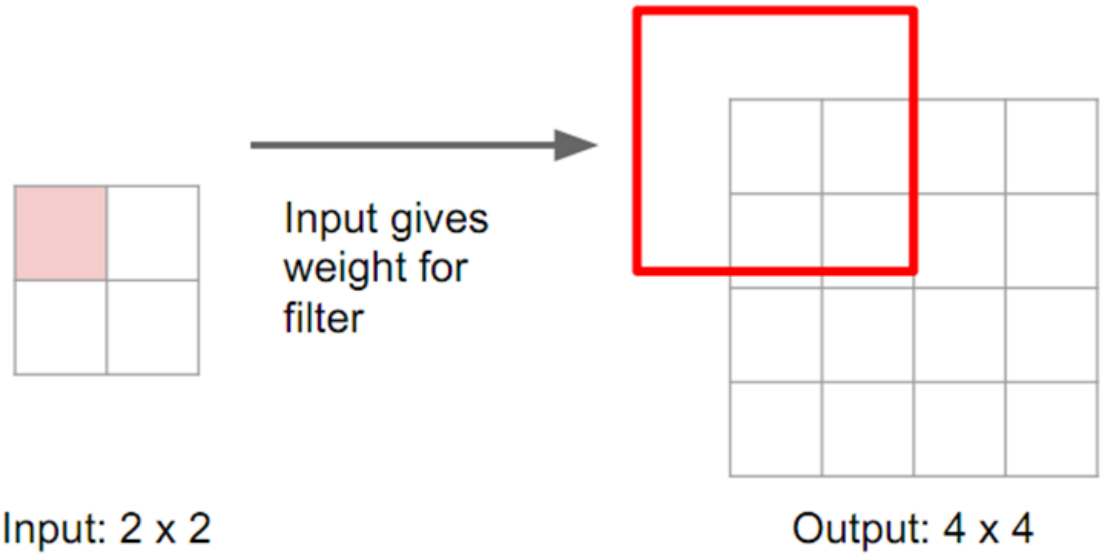


Output: 4 x 4

- Tích chập chuyển vị conv 3x3, bước nhảy stride 2 và thêm viền padding 1

Tích chập chuyển vị

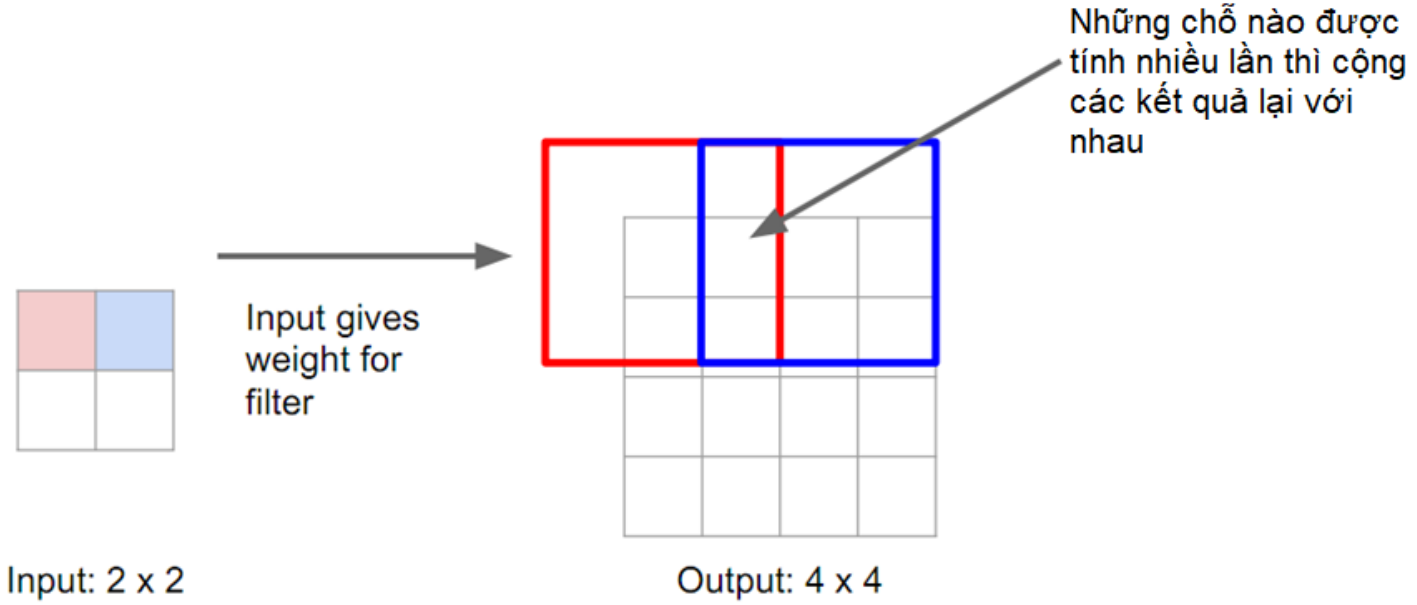
- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



- Tích chập chuyển vị conv 3x3, bước nhảy stride 2 và thêm viền padding 1

Tích chập chuyển vị

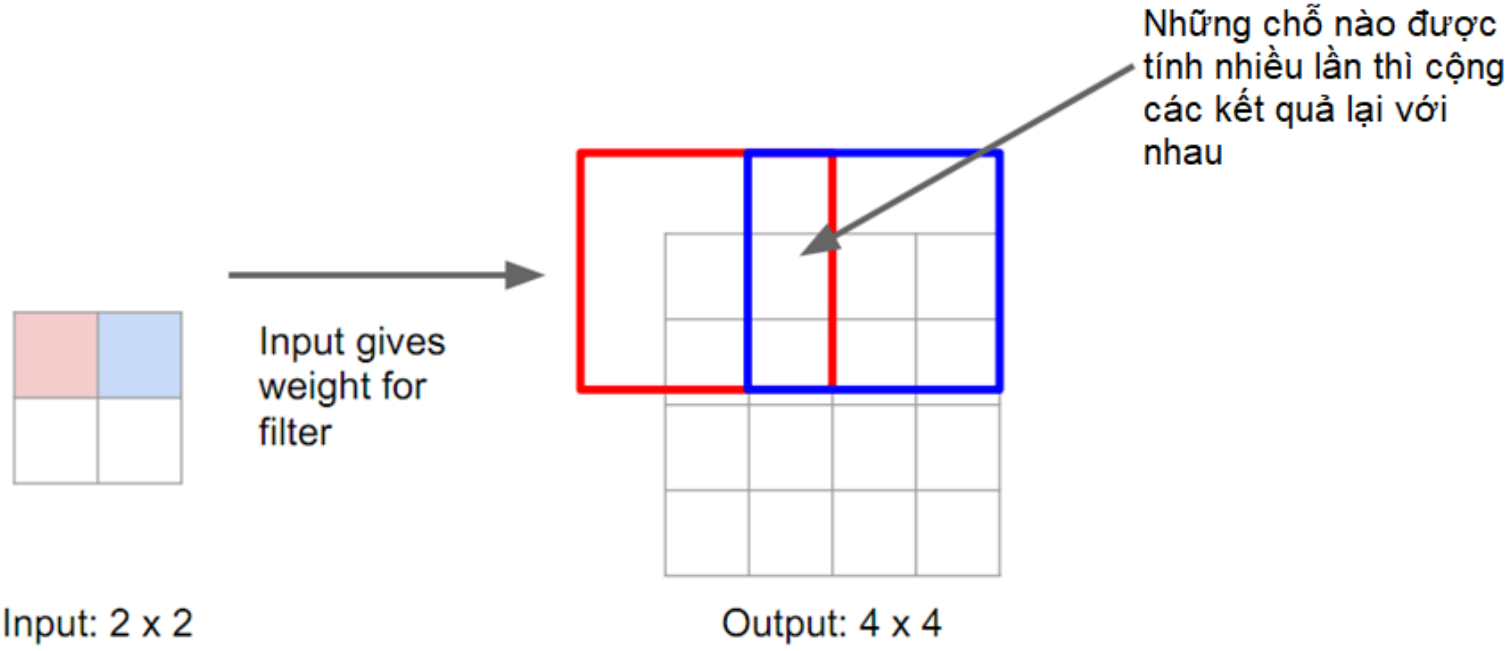
- Là phép tăng độ phân giải (upsampling) có chứa các tham số có thể huấn luyện được



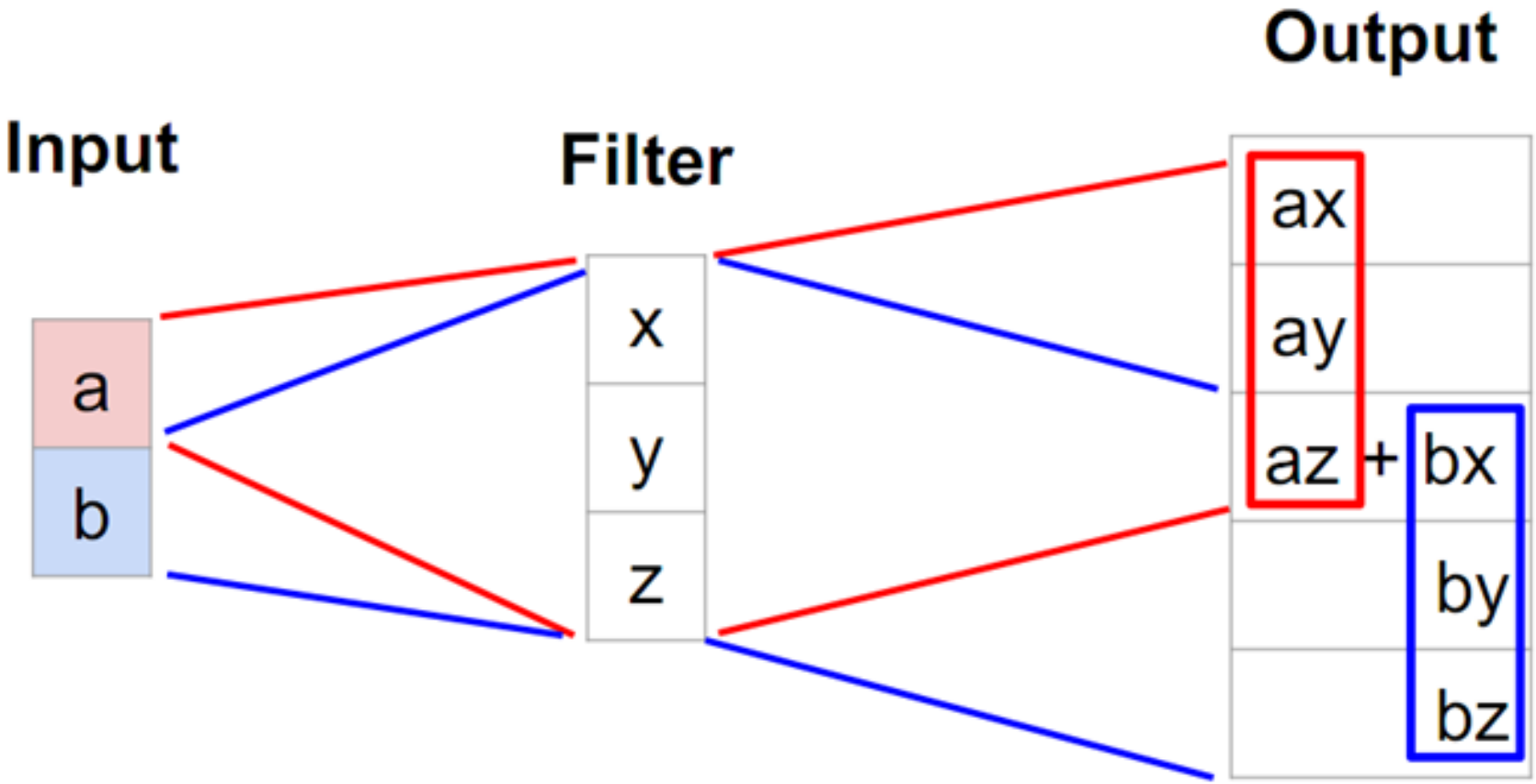
- Tích chập chuyển vị conv 3x3, bước nhảy stride 2 và thêm viền padding 1

Tích chập chuyển vị

- Tên gọi khác:
  - Deconvolution (không nên, dễ gây hiểu nhầm)
  - Upconvolution
  - Fractionally strided convolution
  - Backward strided convolution

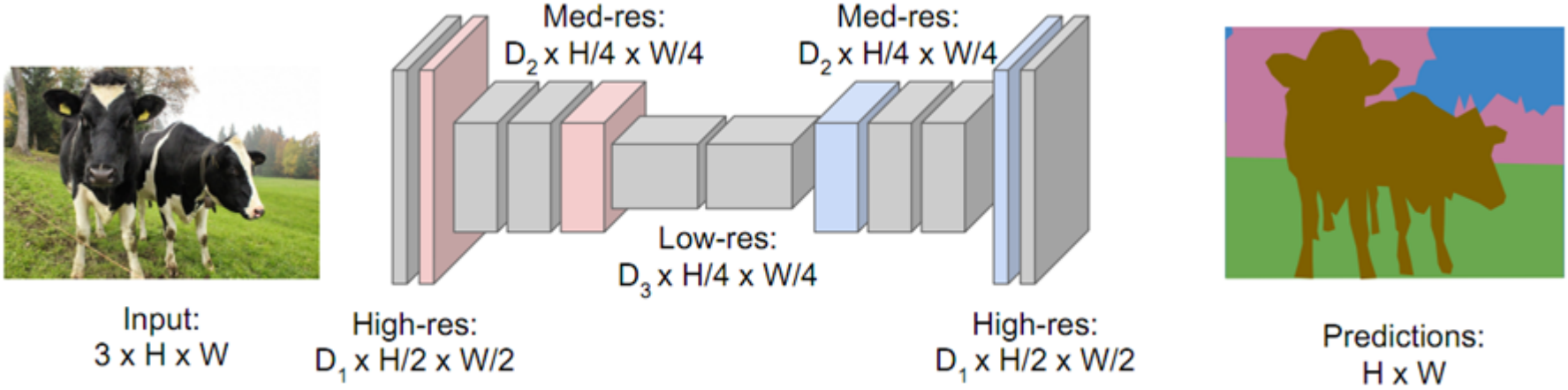


Ví dụ tích chập chuyển vị trong 1D



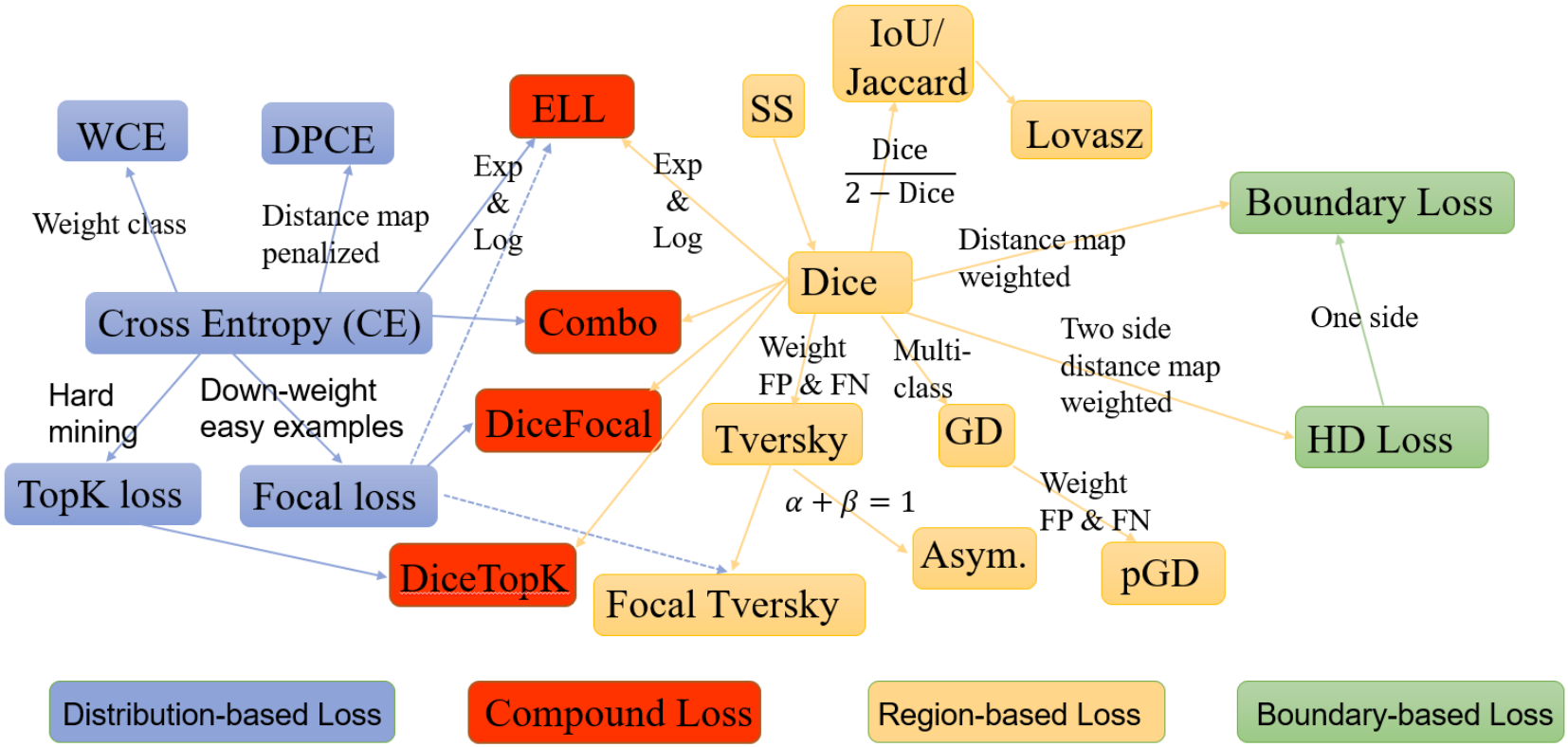
Tích chập hoàn toàn (Fully Convolutional)

- Thiết kế mạng CNN với các lớp giảm độ phân giải (downsampling) và tăng độ phân giải (upsampling)
- Giảm độ phân giải: max pooling hay strided conv
- Tăng độ phân giải: unpooling hoặc transpose conv

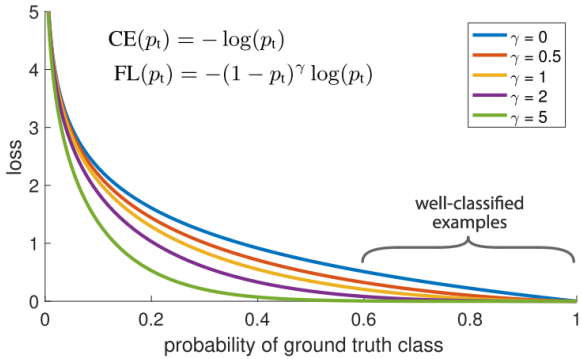


Hàm mục tiêu
cho bài toán phân đoạn ảnh

Hàm mục tiêu


WCE
DPCE
Weight class
Distance map penalized
Cross Entropy (CE)
Hard mining
Down-weight easy examples
TopK loss
Focal loss
ELL
Exp & Log
Combo
DiceFocal
DiceTopK
Exp & Log
SS
IoU/ Jaccard
Lovasz
Dice
$\frac{\text{Dice}}{2-\text{Dice}}$
Distance map weighted
Boundary Loss
One side
HD Loss
Two side distance map weighted
Weight FP & FN
Multi-class
Tversky
GD
$\alpha+\beta=1$
Asym.
Weight FP & FN
pGD
Focal Tversky
Distribution-based Loss
Compound Loss
Region-based Loss
Boundary-based Loss

Hàm mục tiêu dựa trên phân phối



- Cross Entropy (CE):

$$\mathrm{CE}\left(p, \hat{p}\right) = -\left(p\log(\hat{p}) + (1-p)\log(1-\hat{p})\right)$$

- Weighted CE: mỗi lớp có trọng số khác nhau

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

- Focal loss: giải quyết vấn đề mất cân bằng lớn giữa lớp nền và lớp đối tượng quan tâm. Giá trị hàm mục tiêu đối với những mẫu dễ phân loại được giảm xuống thấp để mạng tập trung hơn vào mẫu khó.

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

Hàm mục tiêu dựa trên vùng

- Dice coefficient và IoU:

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

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

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

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

Hàm mục tiêu kết hợp

- Dice loss + CE:

$$\mathrm{CE}\left(p, \hat{p}\right) + \mathrm{DL}\left(p, \hat{p}\right)$$

- Dice loss + Focal loss

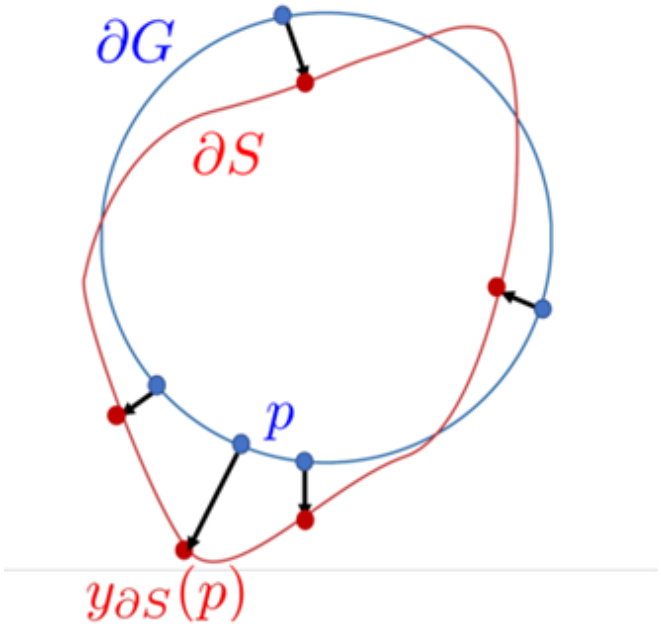
$$\mathrm{CE}\left(p, \hat{p}\right) + \mathrm{FL}\left(p, \hat{p}\right)$$

- …

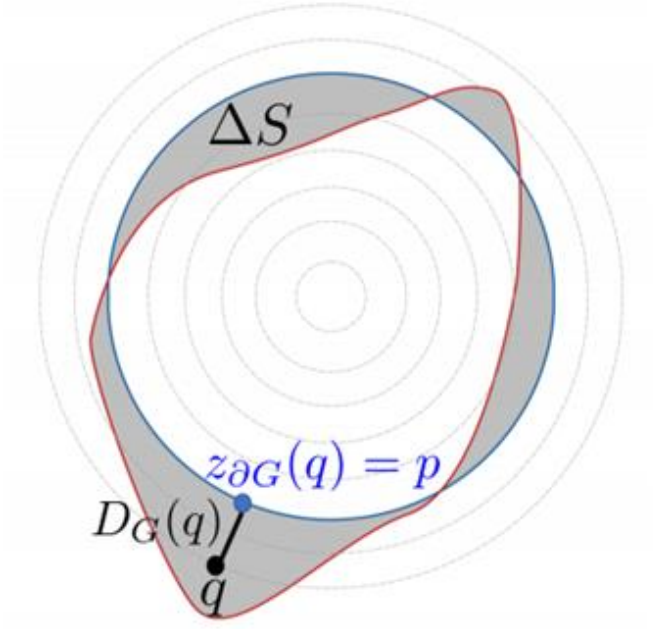
Hàm mục tiêu boundary loss

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

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



(a) Differential

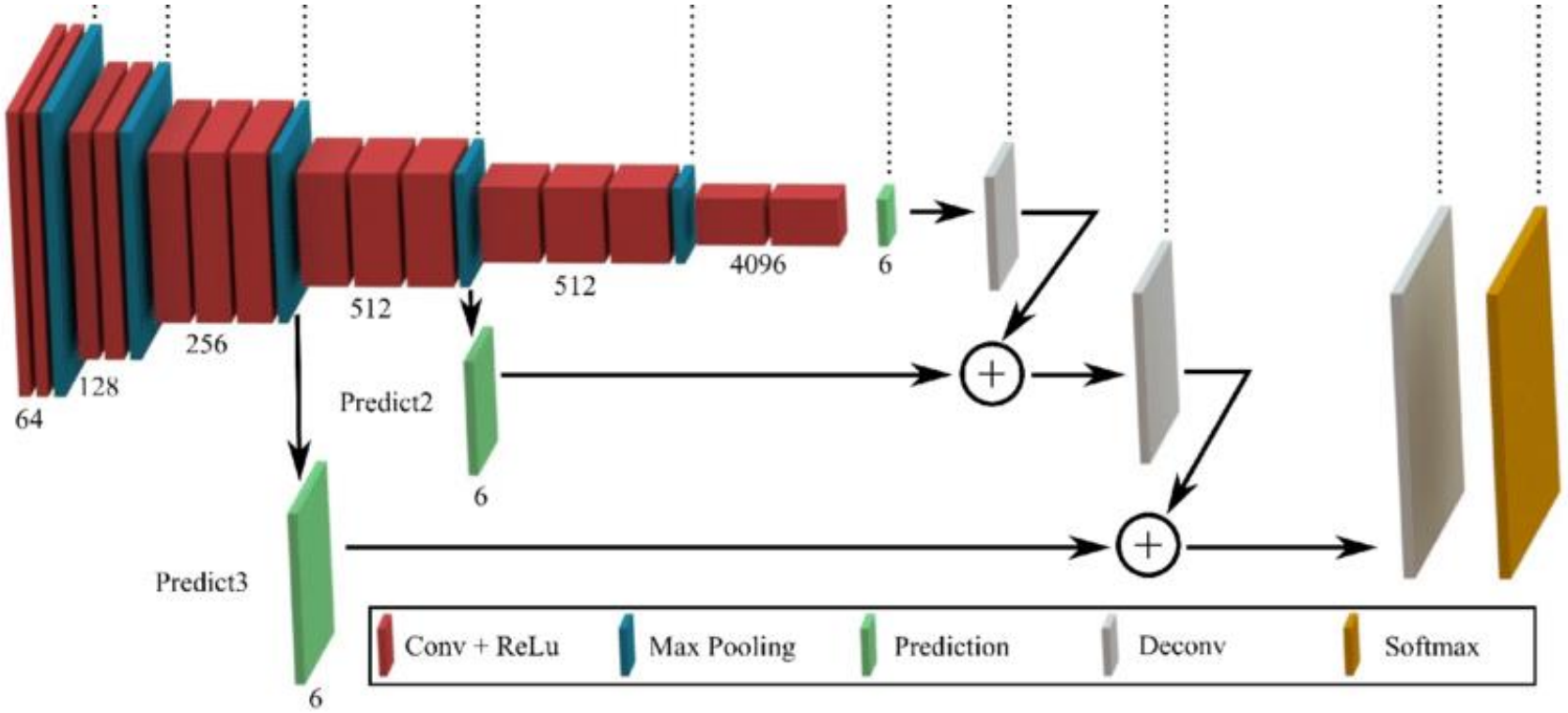


(b) Integral

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

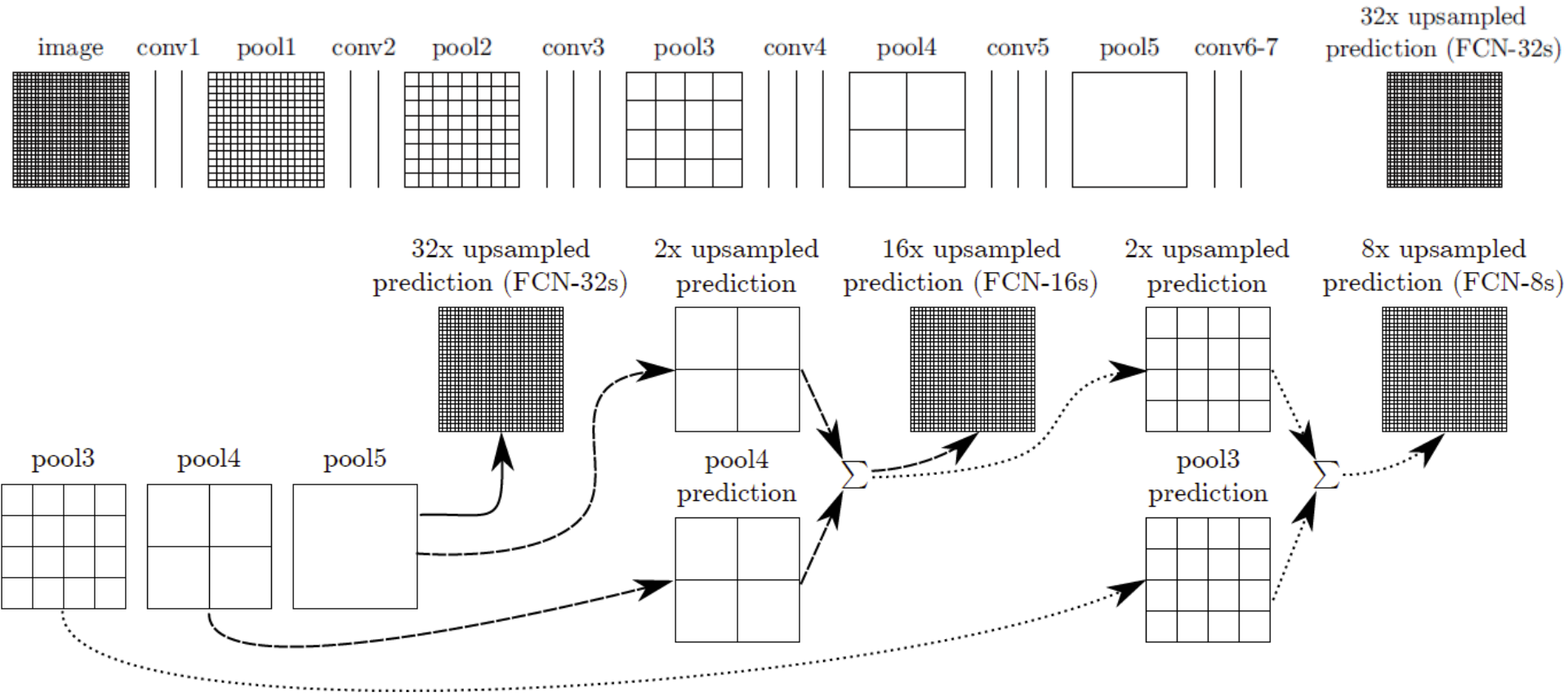
Một số mạng phân đoạn ảnh tiêu biểu

FCN với 2 kết nối tắt

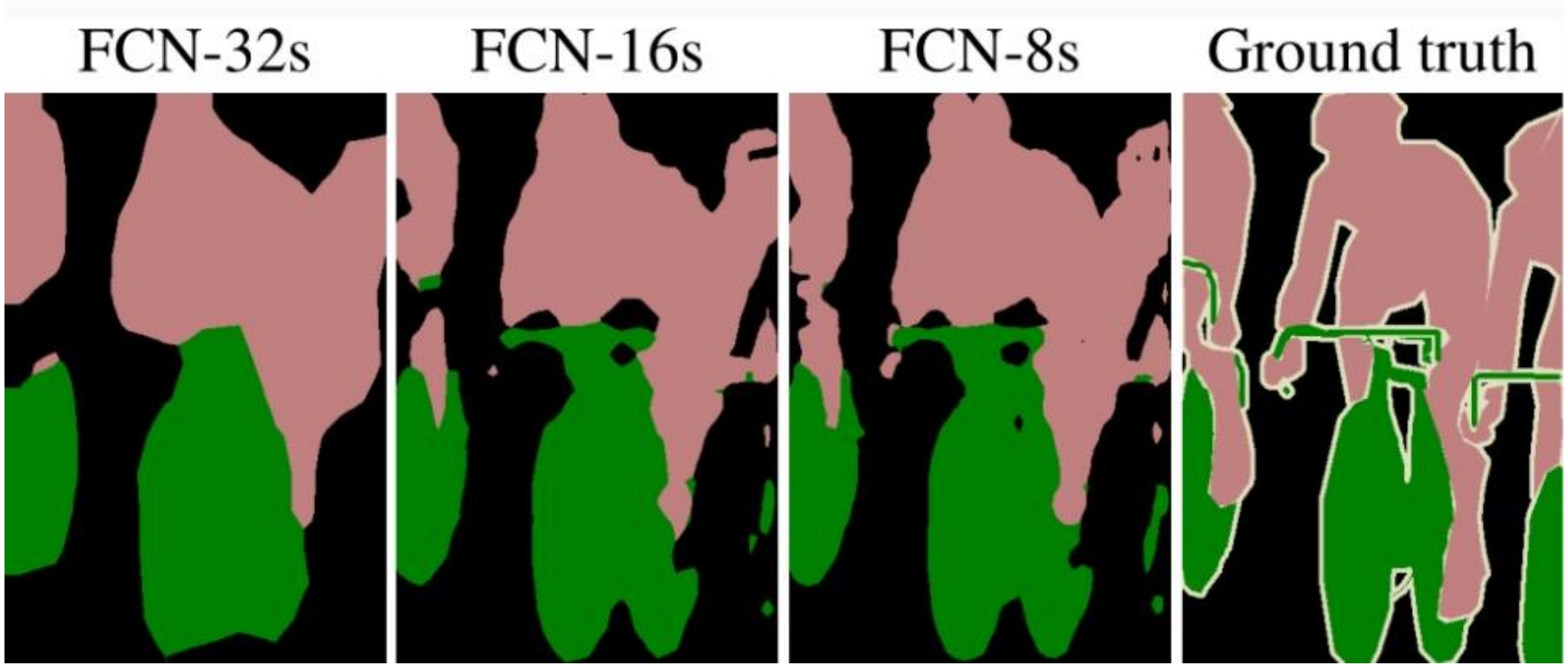


FCN với 2 kết nối tắt

- Minh họa kết quả FCN với các mức độ phân giải khác nhau

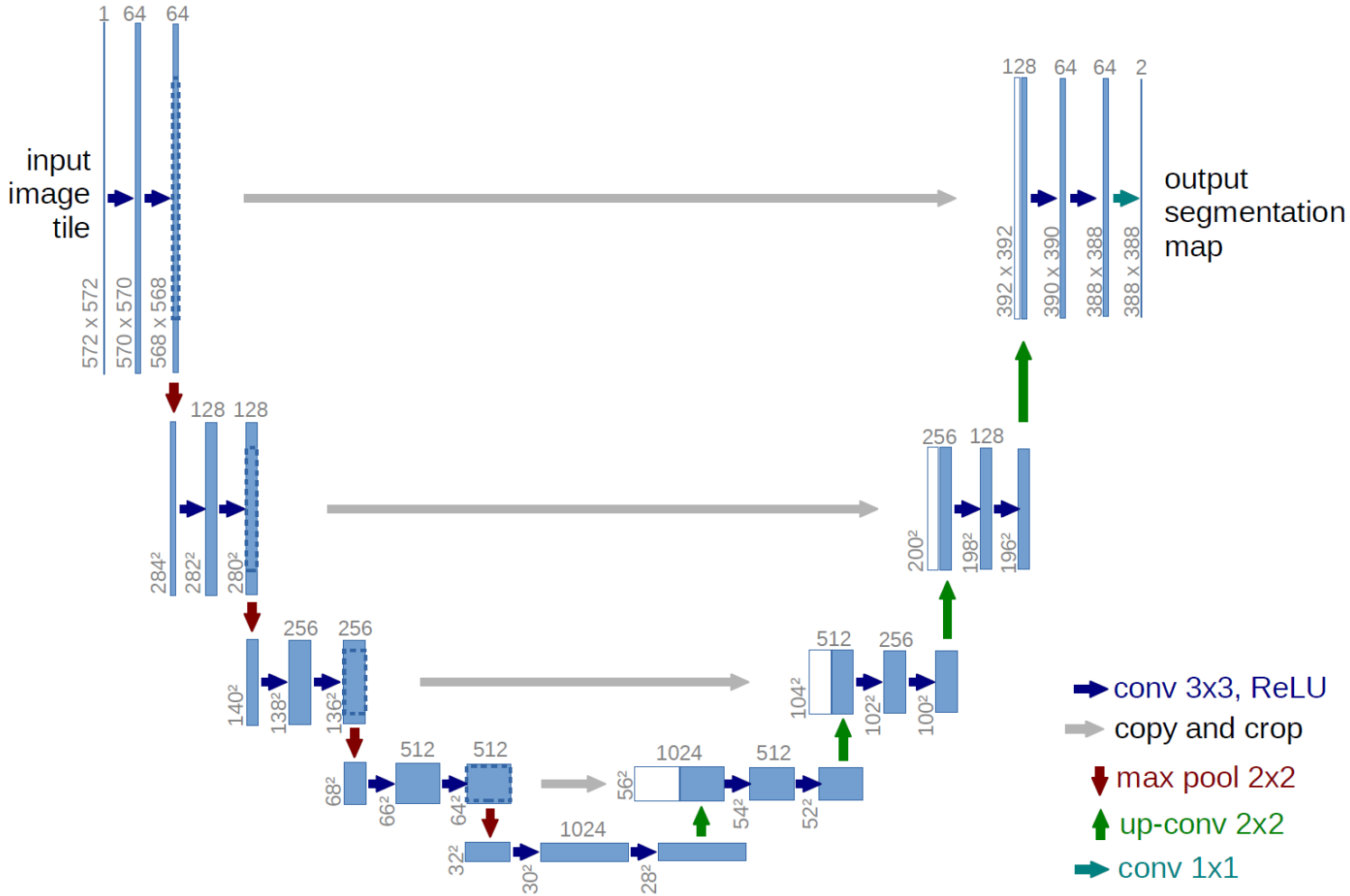


So sánh kết quả

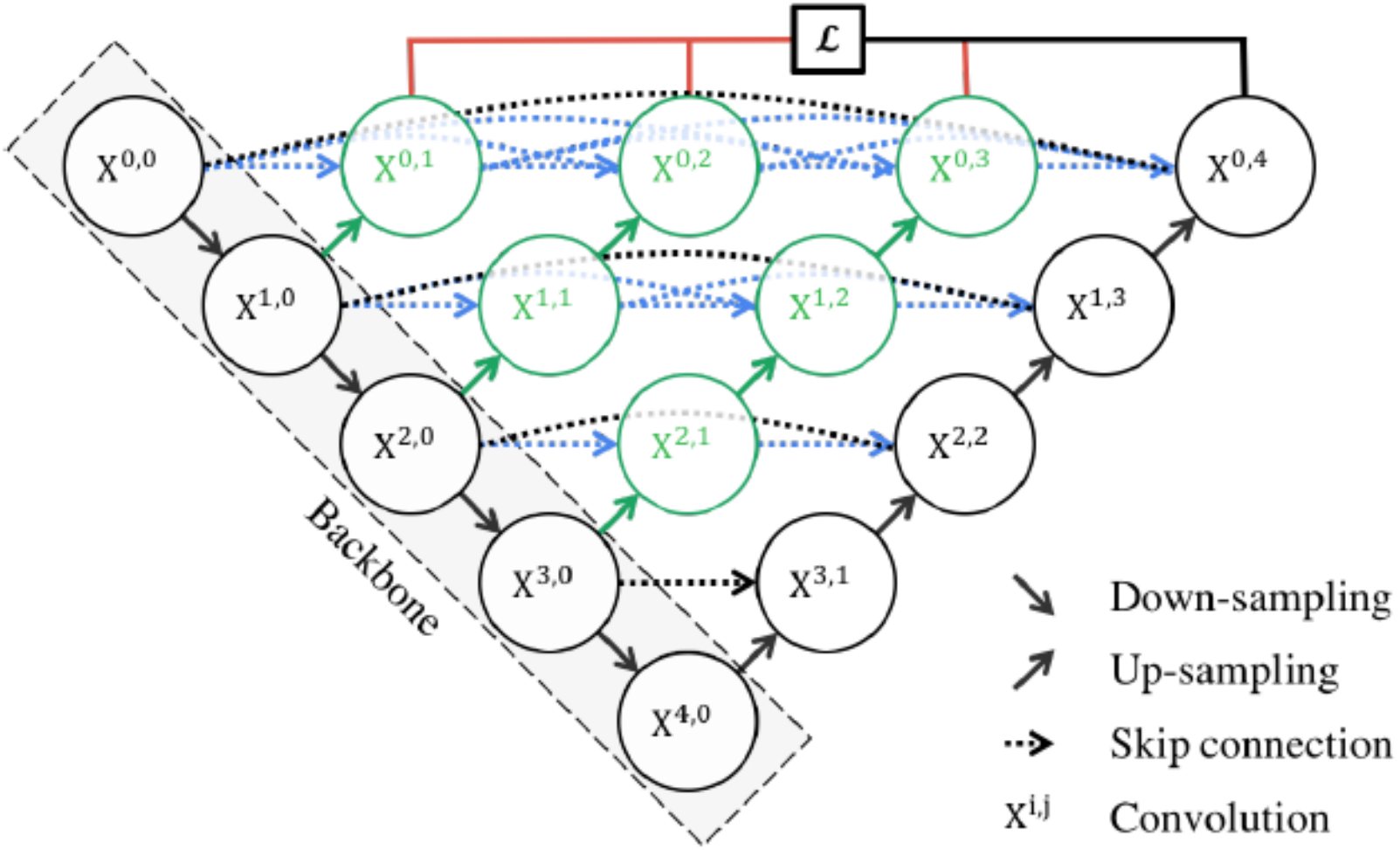


U-Net

- Được sử dụng rộng rãi trong y tế

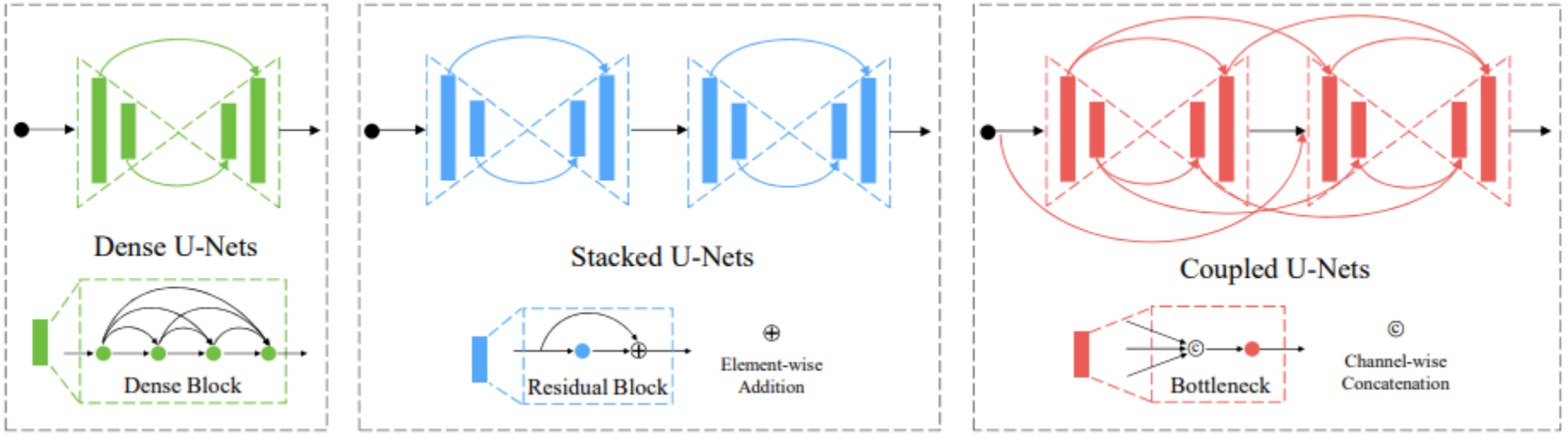


U-Net++


$\mathcal{L}$
$X^{0,0}$
$X^{0,1}$
$X^{0,2}$
$X^{0,3}$
$X^{0,4}$
$X^{1,0}$
$X^{1,1}$
$X^{1,2}$
$X^{1,3}$
$X^{2,0}$
$X^{2,1}$
$X^{2,2}$
$X^{3,0}$
$X^{3,1}$
$X^{4,0}$
Backbone
Down-sampling
Up-sampling
Skip connection
$X^{i,j}$
Convolution

Stacked UNets và CUNets

- Stacked UNets: ghép nhiều UNet nối tiếp nhau
- CUNets: cũng ghép nhiều UNet nối tiếp nhau nhưng có thêm các kết nối tắt giữa các UNet với nhau



Tài liệu tham khảo

1. Khóa cs231n của Stanford:

http://cs231n.stanford.edu

2. Hàm mục tiêu cho bài toán phân đoạn ảnh:

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

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

Thank you for your attentions!