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

Phân đoạn ảnh

1

Nội dung buổi học

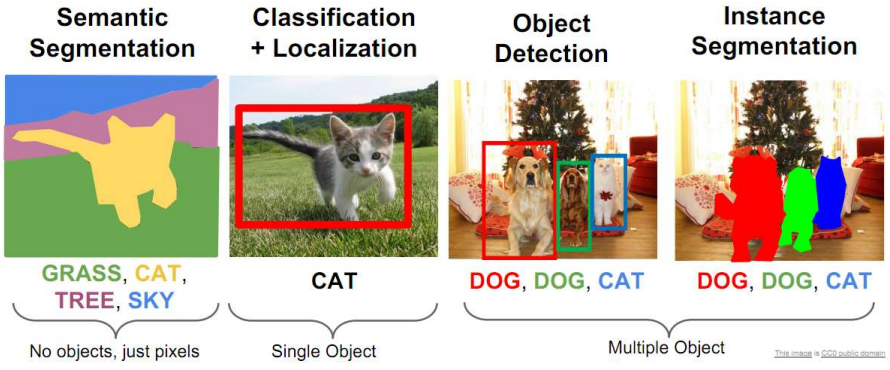
- Giới thiệu về bài toán phân đoạn ảnh
- Phân đoạn ảnh dựa trên ngưỡng
- Phân đoạn ảnh dựa vào cạnh
- Phân đoạn ảnh dựa vào miền
- Phân đoạn ảnh phân cụm
- Kỹ thuật học sâu

2

Giới thiệu bài toán phân vùng ảnh

3

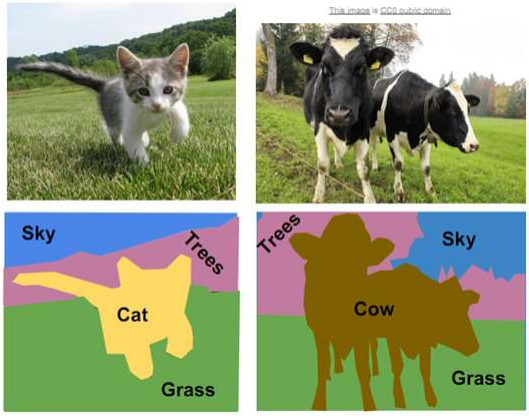
Các bài toán thị giác máy



4

Phân đoạn ảnh

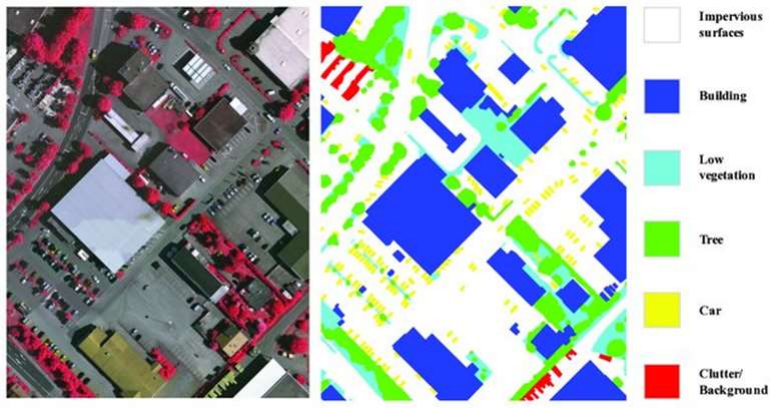
- Phân lớp từng điểm ảnh trong ảnh
- Không phân biệt các đối tượng cùng lớp trong ảnh



5

Một số ứng dụng phân đoạn ảnh

- Phân đoạn ảnh vệ tinh và hàng không



6

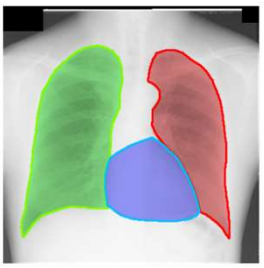
Một số ứng dụng phân đoạn ảnh

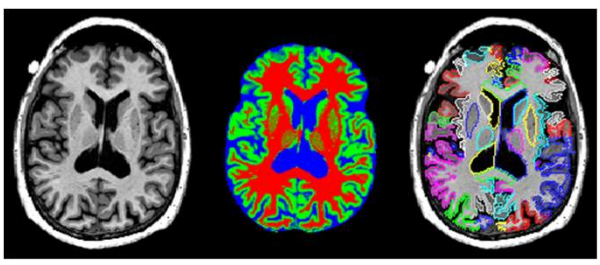
- Xe tự hành

7

Một số ứng dụng phân đoạn ảnh

- Y tế





8

Một số ứng dụng phân đoạn ảnh

- OCR

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

9

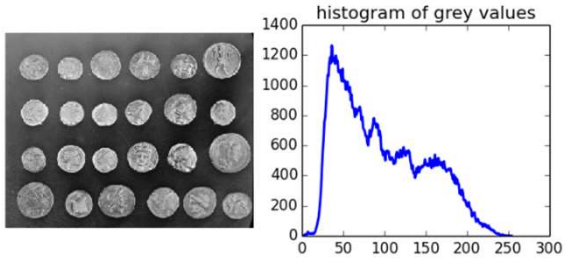
Phân loại các giải pháp phân đoạn ảnh

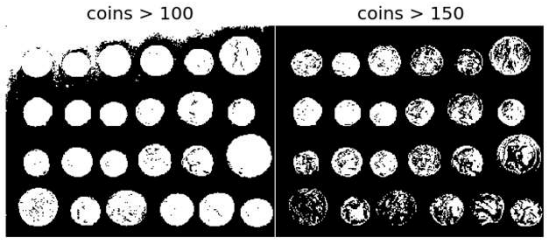
- Phân đoạn dựa vào ngưỡng (thresholding)
- Phân đoạn dựa trên phát hiện biên/ cạnh (edge-based)
- Phân đoạn dựa trên vùng (region-based)
- Phân đoạn dựa trên phân cụm (clustering)
- Phân đoạn dựa trên kỹ thuật học sâu

10

Giải thuật phân đoạn dựa trên ngưỡng

- Phân đoạn ảnh dựa trên ngưỡng đơn → ảnh nhị phân



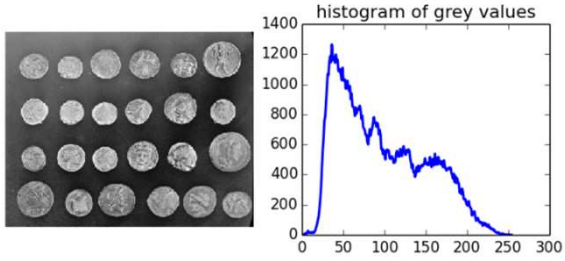


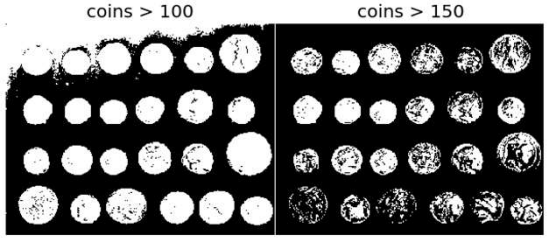
11

11

Giải thuật phân đoạn dựa trên ngưỡng

- Phân đoạn ảnh dựa trên ngưỡng đơn → ảnh nhị phân



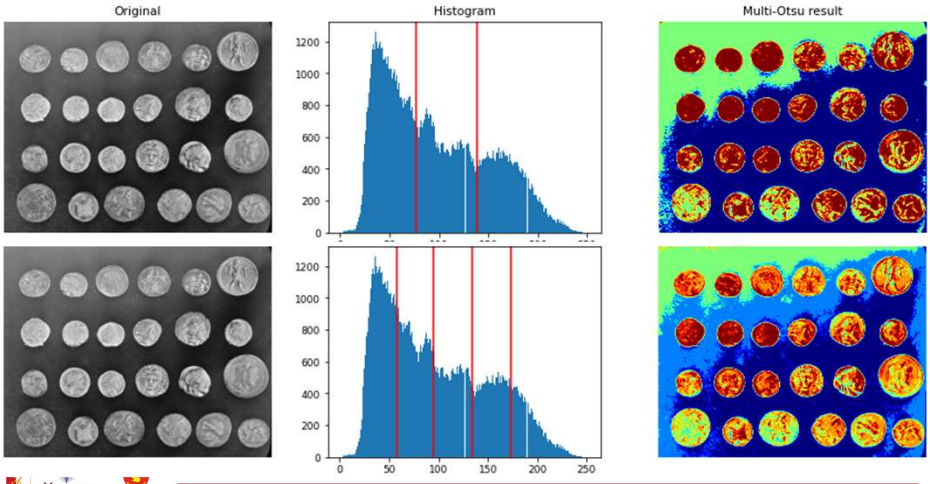


12

12

Giải thuật phân đoạn dựa trên ngưỡng (2)

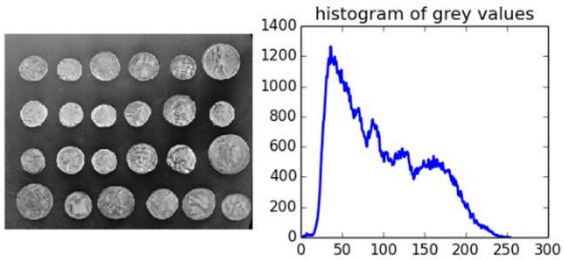
- Phân đoạn ảnh dựa trên đa ngưỡng
  - Vd: Multi-otsu thresholding

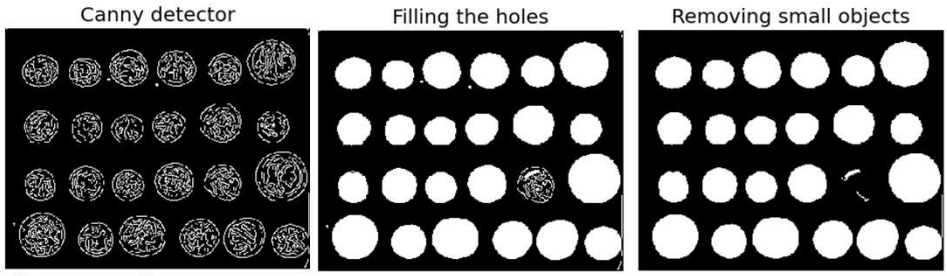


13

13

Giải thuật phân đoạn dựa vào biên





14

14

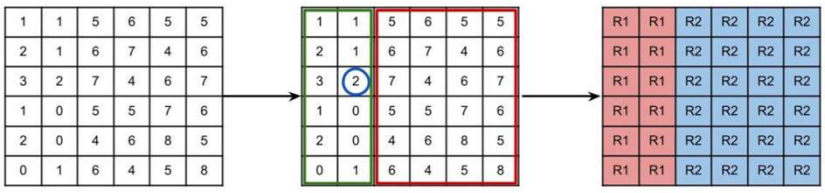
Giải thuật phân đoạn dựa trên vùng

- Lan tỏa/ Phát triển vùng (Region Growing)
- Tách và hợp (Splitting and Merging)

15

Lan tỏa vùng (Region Growing)

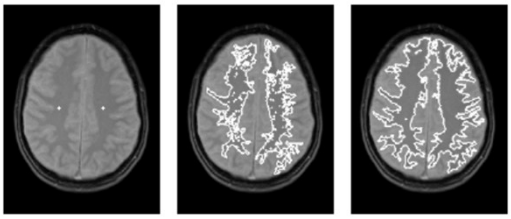
- B1: Lựa chọn một điểm seed (theo tiêu chí người dùng)
- B2: kiểm tra các điểm lân cận để tìm ra các điểm tương đồng (sai lệch độ sáng nhỏ hơn ngưỡng T), thêm điểm tương đồng vào vùng
- B3: Lặp lại bước 2 với tất cả các điểm trong vùng



16

Lan tỏa vùng (Region Growing)

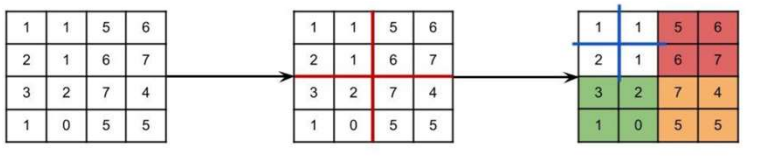
- Ưu điểm
  - Đơn giản, chỉ cần các điểm seed point
  - Dễ dàng tách những vùng đồng nhất
  - Kết quả tốt với ảnh có biên rõ ràng
- Nhược điểm
  - Chi phí tính toán lớn
  - Tính toán cục bộ, không phải toàn cục
  - Nhạy với nhiễu



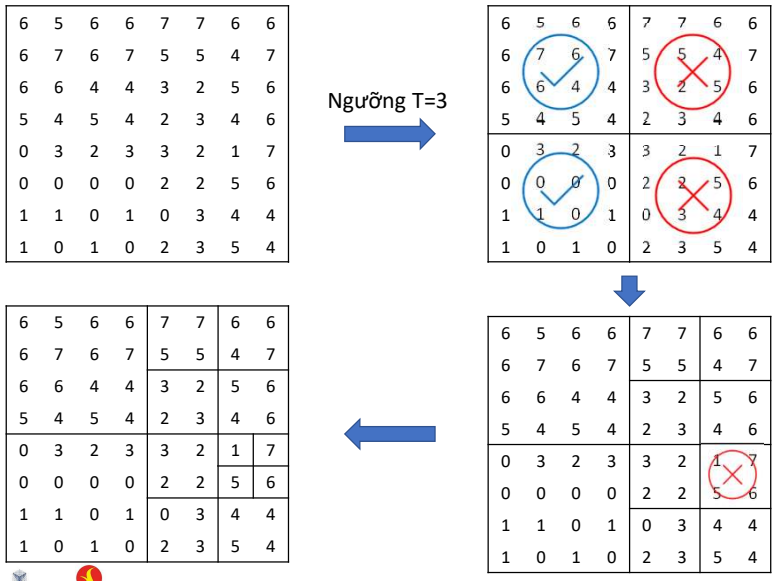
Ví dụ lan tỏa vùng từ 2 seed points

Tách và hợp (Splitting and Merging)

- B1: Chia
  - Chia ảnh thành 4 phần
  - Kiểm tra tính đơn nhất trong mỗi phần (maxP – minP <= T)
  - Với mỗi vùng không đơn nhất, tiếp tục chia nhỏ
- B2: Hợp
  - Từ những vùng đã có, hợp lại thành nhiều vùng lớn hơn dựa vào ngưỡng T



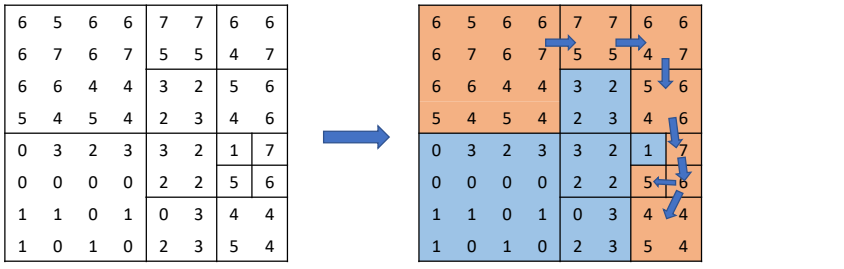
Quá trình tách (ví dụ)



19

Quá trình hợp (ví dụ)

Hợp 2 vùng lân cận nếu
Max(1) – min(2) <= T
và max(2) – min(1) <= T



Ảnh gốc | Ảnh sau khi tách | Ảnh sau hợp

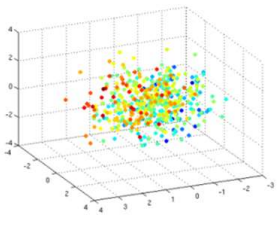
20

Các giải thuật phân cụm

Segmentation as Clustering

- Điểm ảnh được biểu diễn trong không gian nhiều chiều

- Màu sắc: 3d
- Màu sắc + vị trí: 5d



Giải thuật K-means

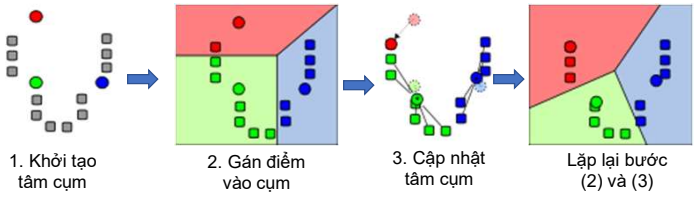
- B1: Ngẫu nhiên khởi tạo k điểm trung tâm c_1, c_2, … c_k
- B2: Với các điểm trung tâm hiện tại của từng phân cụm, xác định các điểm còn lại thuộc phân cụm nào
  - Với mỗi điểm p, tìm điểm trung tâm gần nhất và thêm p vào cụm tương ứng
- B3: Với từng phân cụm, tính toán lại điểm trung tâm
  - Điểm trung tâm là trung bình các điểm trong phân cụm
- B4: Nếu c_i thay đổi, lặp tại bước 2

Đặc điểm:

- Luôn hội tụ về một phương án nào đó
- Có thể là "cực tiểu cục bộ"
  - Không đảm bảo luôn tìm thấy hàm mục tiêu cực tiểu toàn cục

$$\sum_{i=1}^{k} \sum_{x \in C_i} \|x - c_i\|^2$$

Minh họa giải thuật K-means

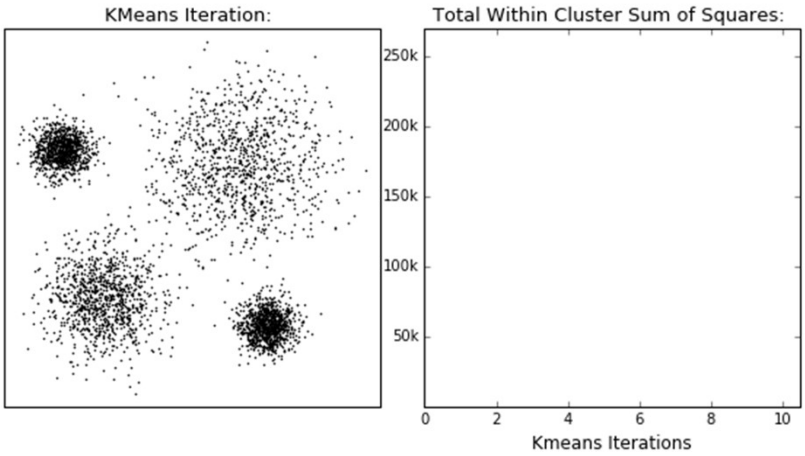


- Demo giải thuật K-means:

https://www.naftaliharris.com/blog/visualizing-k-means-clustering

http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html

Minh họa giải thuật K-means (2)



25

25

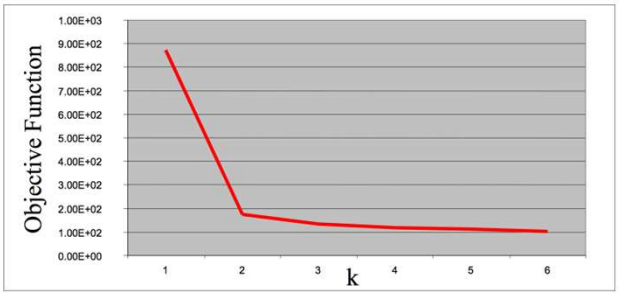
Ví dụ áp dụng K-mean lên ảnh

26

26

Lựa chọn số cụm như thế nào?

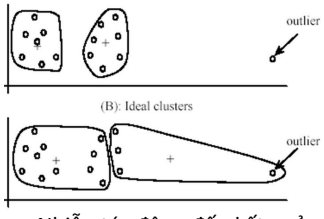
- Thử các giá trị K khác nhau và quan sát kết quả trên tập validation.



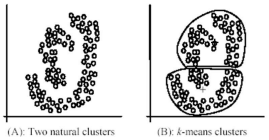
- Có sự thay đổi đột ngột hàm mục tiêu tại k=2, nhiều khả năng số nhóm là 2.

K-Means: Ưu và nhược điểm

- Ưu điểm
  - Đơn giản, nhanh, dễ cài đặt
- Nhược điểm
  - Cần chọn K
  - Nhạy cảm với ngoại biên (outliers)
  - Hội tụ về lời giải địa phương
  - Làm việc không tốt nếu phân bố dữ liệu các cụm không giống nhau
  - *Có thể chậm: O(KNd)
- Sử dụng
  - Phân cụm không giám sát
  - Ít dùng cho phân đoạn ảnh hoặc là bước trung gian cho phân đoạn

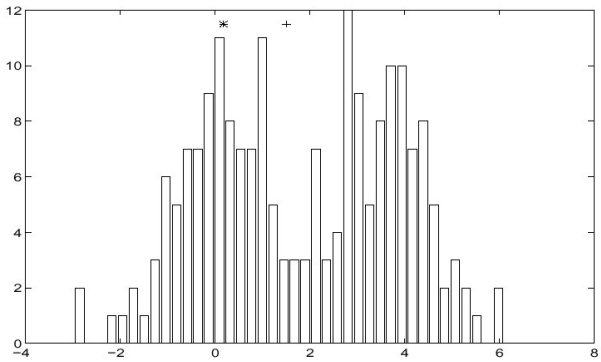


Nhiễu tác động đến kết quả



Phân bố dữ liệu khác nhau

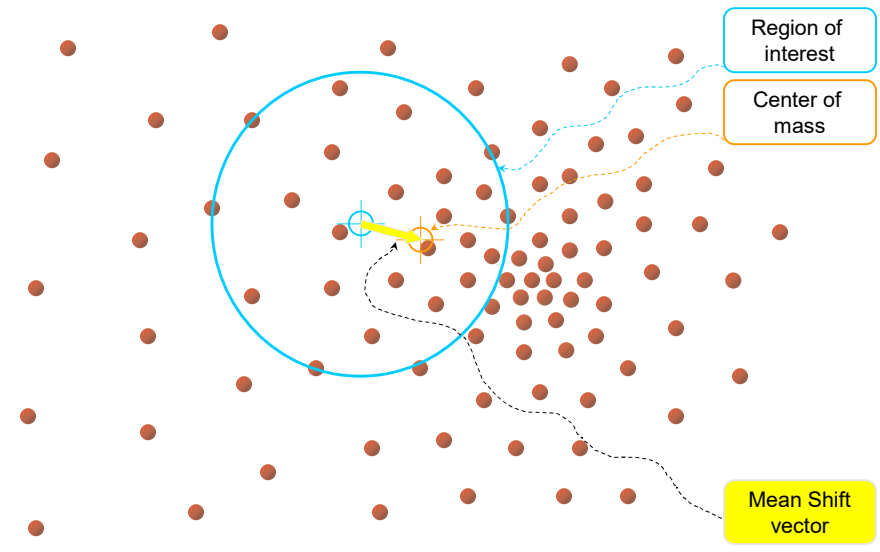
Thuật toán Mean-Shift



- Tìm kiếm đỉnh (mode) bằng thủ tục lặp
  - Khởi tạo các điểm ngẫu nhiên, và cửa sổ W
  - Tính trọng tâm ("mean") của cửa sổ W: $\sum_{x \in W} xH(x)$
  - Tịnh tiến cửa sổ tìm kiếm tới trọng tâm
  - Lặp lại bước 2 cho tới khi hội tụ

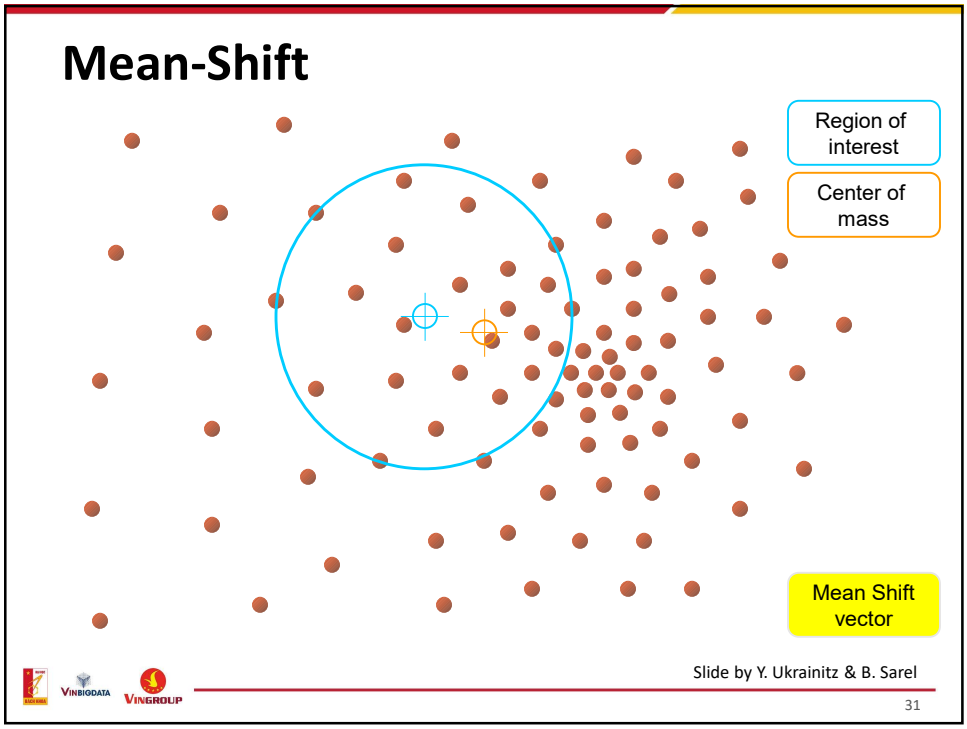
29

Mean-Shift

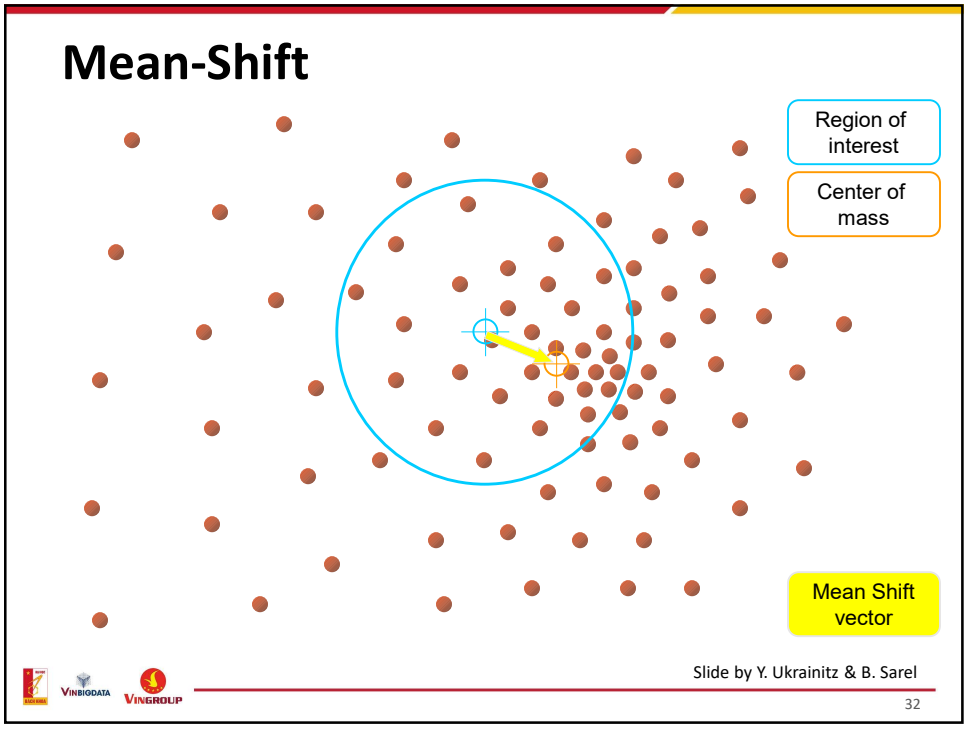


Slide by Y. Ukrainitz & B. Sarel

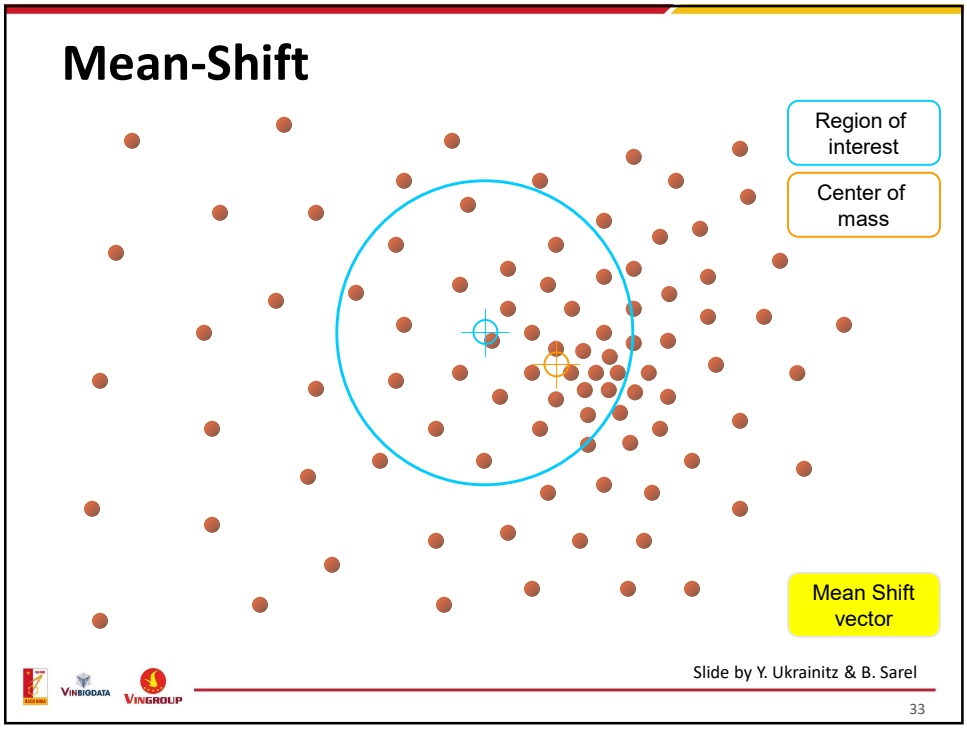
30


Mean-Shift
Region of interest
Center of mass
Mean Shift vector
Slide by Y. Ukrainitz & B. Sarel
31

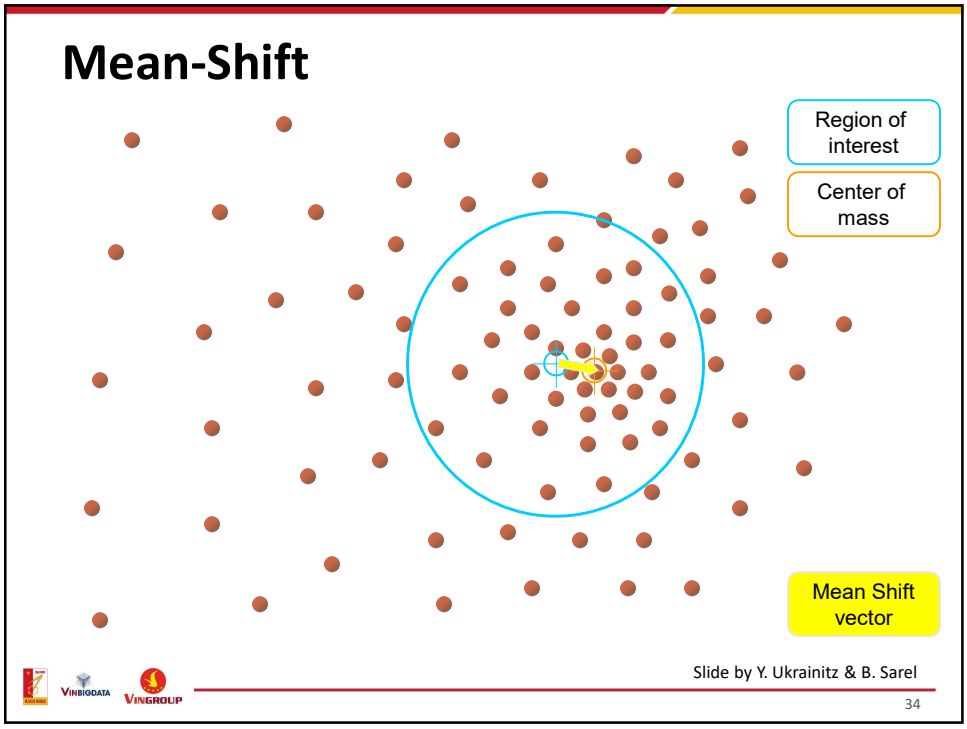
31


Mean-Shift
Region of interest
Center of mass
Mean Shift vector
Slide by Y. Ukrainitz & B. Sarel
32

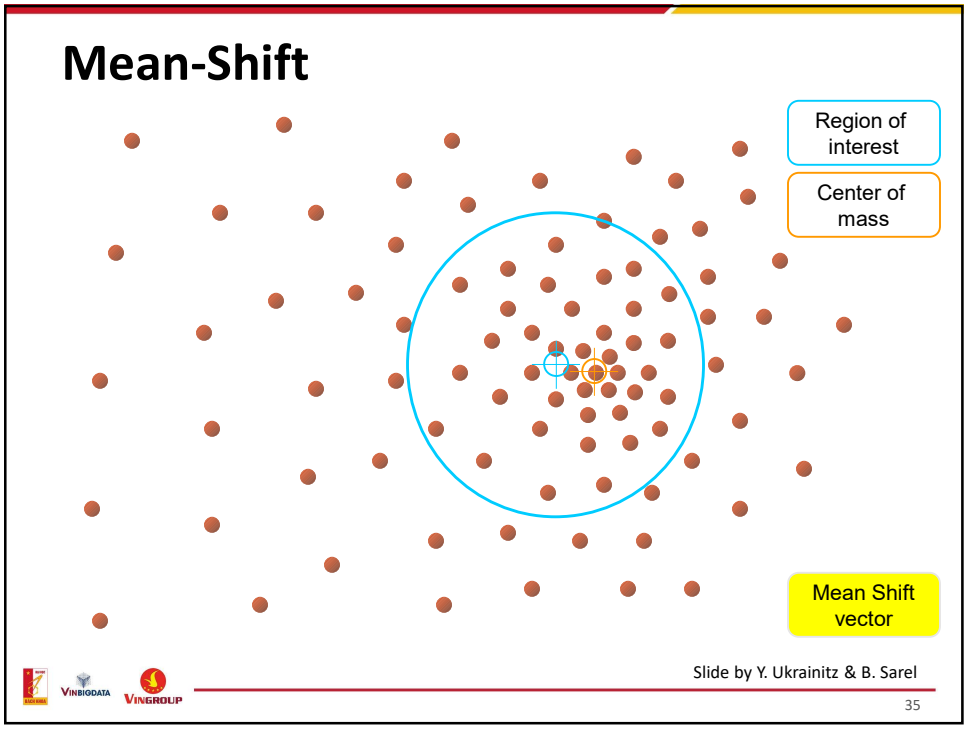
32


Mean-Shift
Region of interest
Center of mass
Mean Shift vector
Slide by Y. Ukrainitz & B. Sarel
33

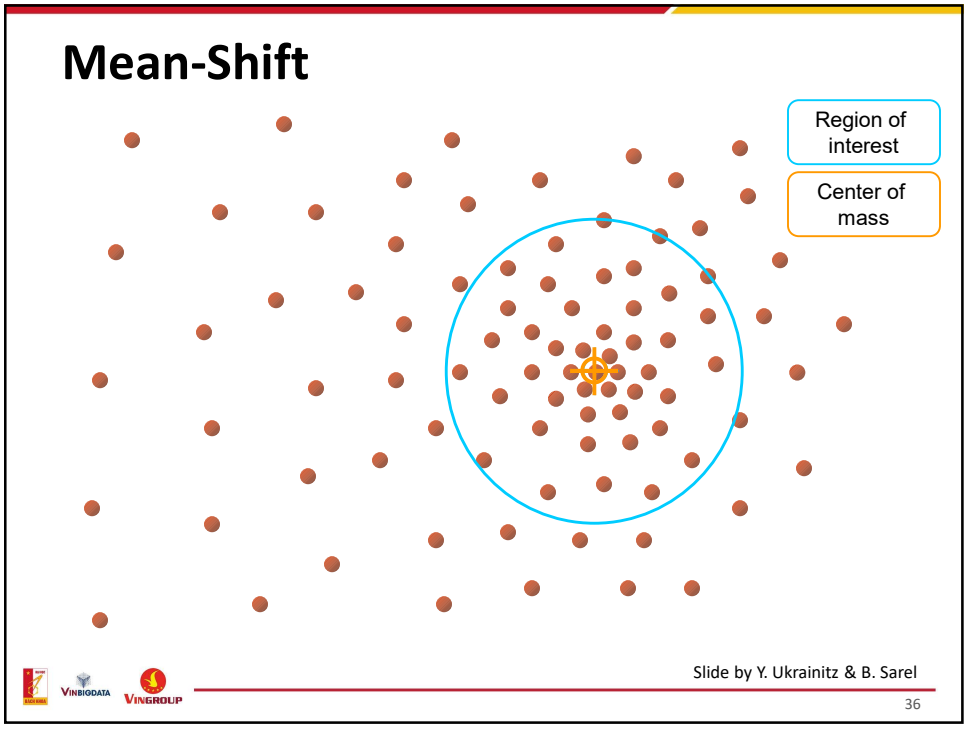
33


Mean-Shift
Region of interest
Center of mass
Mean Shift vector
Slide by Y. Ukrainitz & B. Sarel
34

34

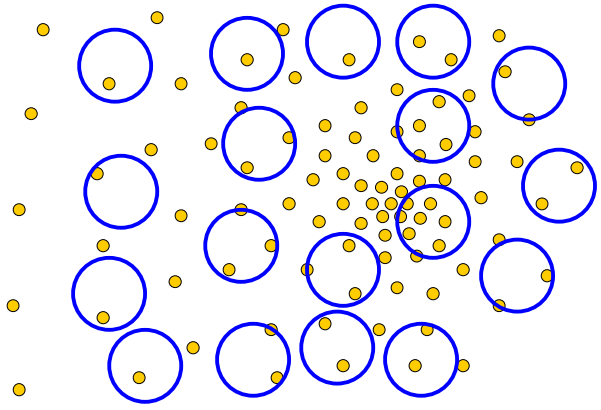

Mean-Shift
Region of interest
Center of mass
Mean Shift vector
Slide by Y. Ukrainitz & B. Sarel
35

35


Mean-Shift
Region of interest
Center of mass
Slide by Y. Ukrainitz & B. Sarel
36

36

Real Modality Analysis

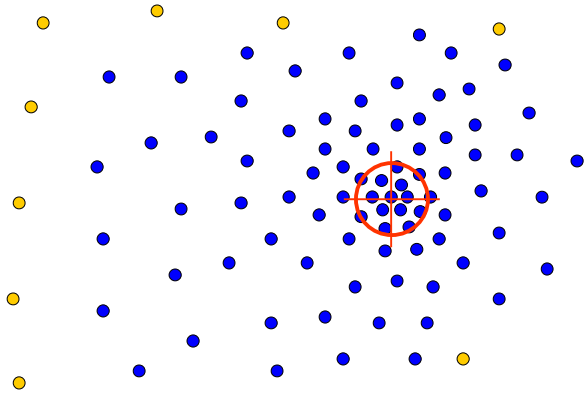


Tessellate the space with windows

Chạy song song các quá trình

37

Real Modality Analysis

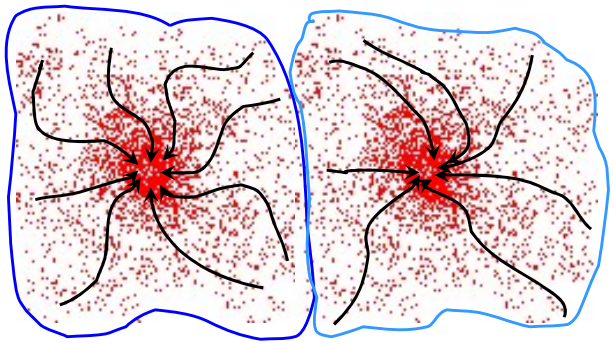


Các điểm màu xanh data là các điểm được các cửa sổ tìm kiếm quét qua trong quá trình hội tụ tới đỉnh (mode).

38

Phân cụm dùng Mean-Shift

- Cluster: tất cả dữ liệu trong vùng "attraction basin" của một đỉnh mode
- Attraction basin: là vùng là tất cả các cửa sổ tìm kiếm trong đó đều hội tụ về cùng một đỉnh.

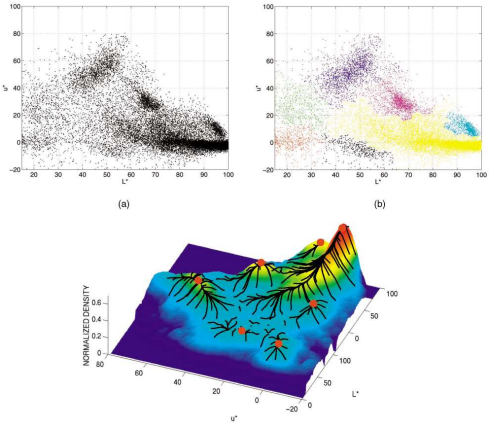


39

39

Phân cụm/phân đoạn dùng Mean-Shift

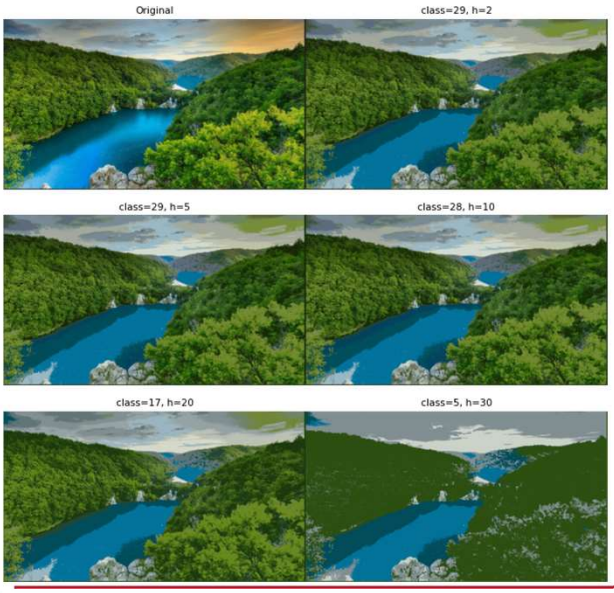
- Trích xuất đặc trưng (color, gradients, texture, etc)
- Khởi tạo các cửa sổ tìm kiếm tại các điểm ảnh khác nhau
- Thực hiện mean shift đối với mỗi cửa sổ cho tới khi hội tụ
- Nhập lại (merge) tất cả các cửa sổ hội tụ về gần cùng một đỉnh



40

40

Ví dụ áp dụng Mean-shift lên ảnh



41

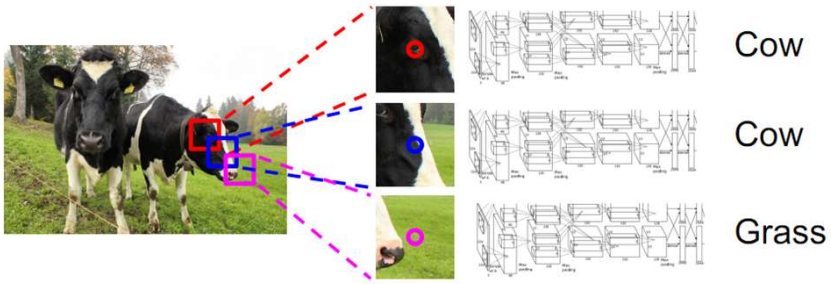
Tổng kết về Mean-Shift

- Ưu điểm
  - Một công cụ tổng quát cho nhiều bài toán
  - Không cần biết phân bố của các cụm dữ liệu (spherical, elliptical, etc.)
  - Chỉ có một tham số duy nhất (kích thước cửa sổ h)
  - Có thể tìm thấy nhiều cụm
  - Không bị ảnh hưởng bởi dữ liệu ngoại lai outliers
- Nhược điểm
  - Kết quả phụ thuộc kích thước cửa sổ
  - Lựa chọn kích thước cửa sổ là không tầm thường
  - Tính toán chậm (~2s/ảnh)
  - Không có khả năng mở rộng tốt khi số chiều không gian đặc trưng tăng

42

Kỹ thuật học sâu

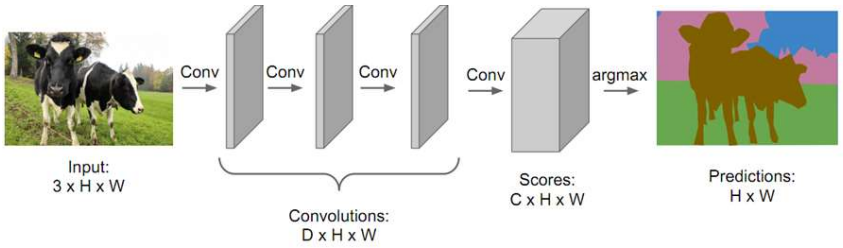
Phương pháp cửa sổ trượt



- Phù hợp cho bài toán nhận dạng đối tượng
- Không hiệu quả trong phân đoạn ảnh

Tích chập hoàn toàn (FCN)

- Vấn đề: Tích chập ở độ phân giải cao tốn chi phí



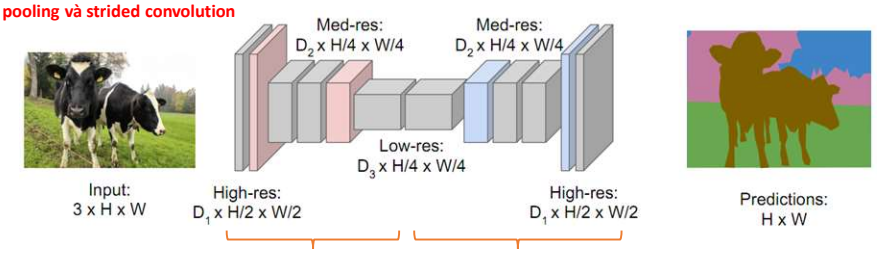
45

Tích chập hoàn toàn (FCN)

- Thiết kế mạng có nhánh downsampling và upsampling

Downsampling:
pooling và strided convolution

Upsampling:

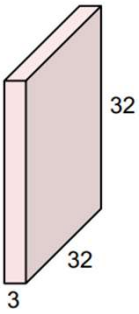


46

Lớp tích chập

Tùy thuộc số band ảnh

Ảnh 32x32x3 (RGB)



Mặt nạ 5x5x3

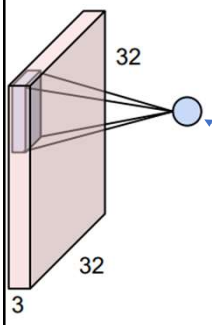
Nhân chập: Thực hiện nhân chập giữa ảnh đầu vào và mặt nạ (nhân) lọc

47

Lớp tích chập

Quá trình nhân chập

Ảnh 32x32x3
Mặt nạ 5x5x3 (w)



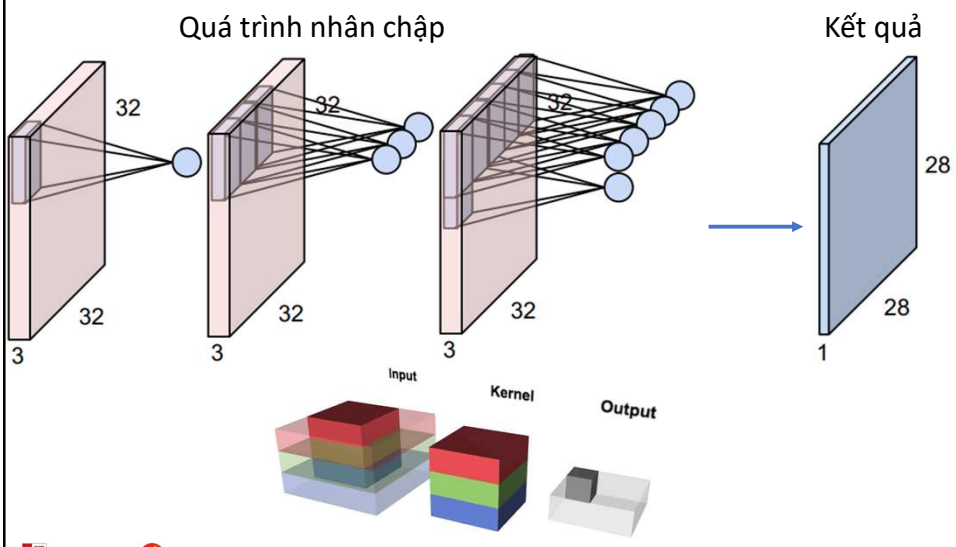
Một giá trị đơn (5*5*3+1 chiều)

$$w^T x + b$$

w là bộ lọc, x là ảnh, b là bias

48

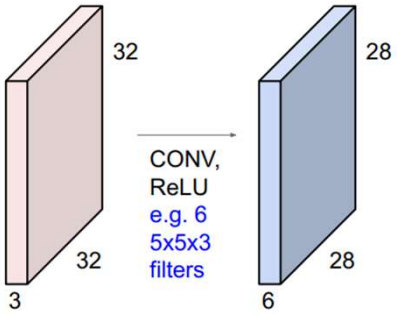
Lớp tích chập



49

Lớp tích chập

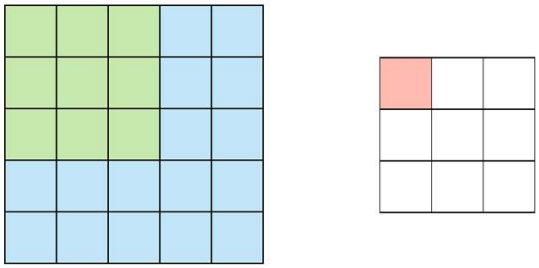
- Kết hợp với các hàm kích hoạt (Activation function)



50

Padding và Stride

- Padding
  - Thêm viền bên ngoài để tăng kích thước ảnh đầu ra
  - Thường dùng để giữ kích thước ảnh đầu ra giống đầu vào
- Stride
  - Bước nhảy, bình thường là 1, có thể lớn hơn 1



Stride 1 và pad 0

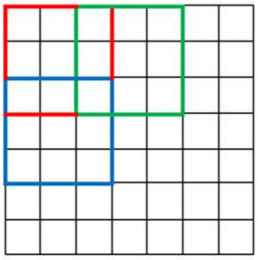
51

51

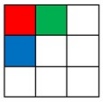
Padding và Stride (2)

Stride = (2, 2)

Ảnh đầu vào 7x7



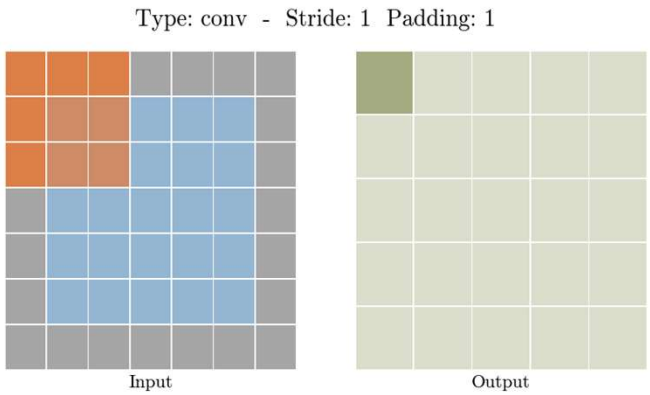
Ảnh đầu ra 3x3



52

52

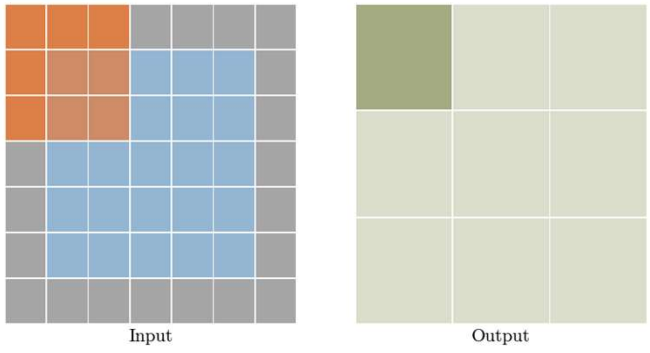
Padding và Stride (2)



Padding: zero padding và nearest neighbour

53

Padding và Stride (4)



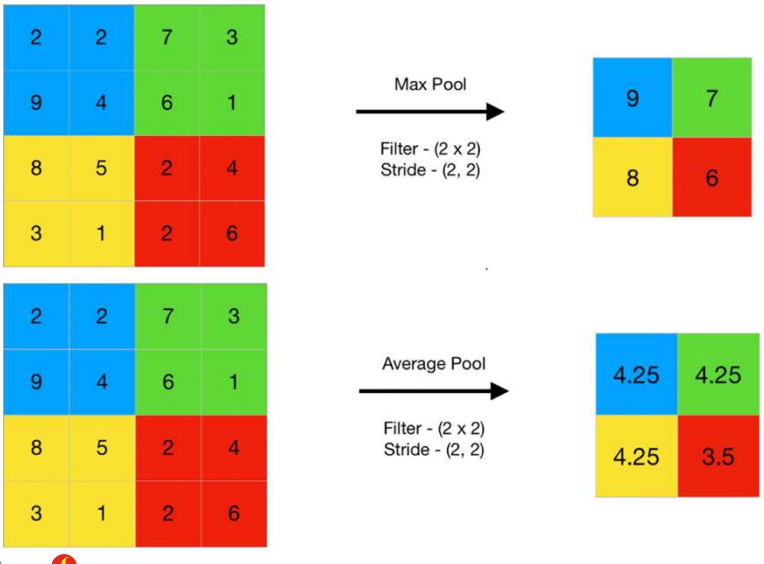
55

Pooling layer

- Là kỹ thuật để giảm kích thước ảnh
- Các kiểu pooling
  - Max (cực đại): giá trị lớn nhất trong cửa sổ
  - Average (trung bình): giá trị trung bình trong cửa sổ
  - Global (toàn cục): mỗi kênh chuyển về một giá trị duy nhất (cực đại hoặc trung bình)

56

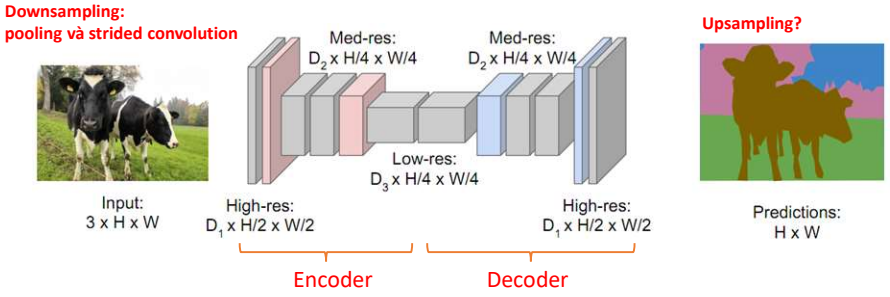
Pooling layer (ví dụ)



57

Tích chập hoàn toàn (FCN)

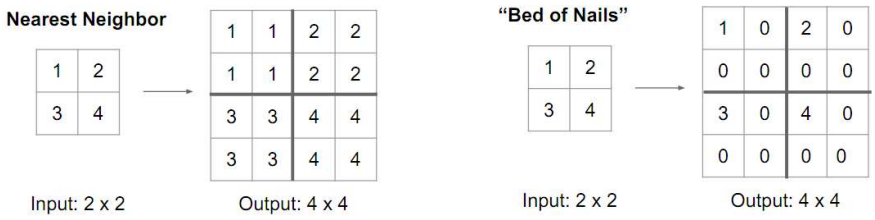
- Thiết kế mạng có nhánh downsampling và upsampling



58

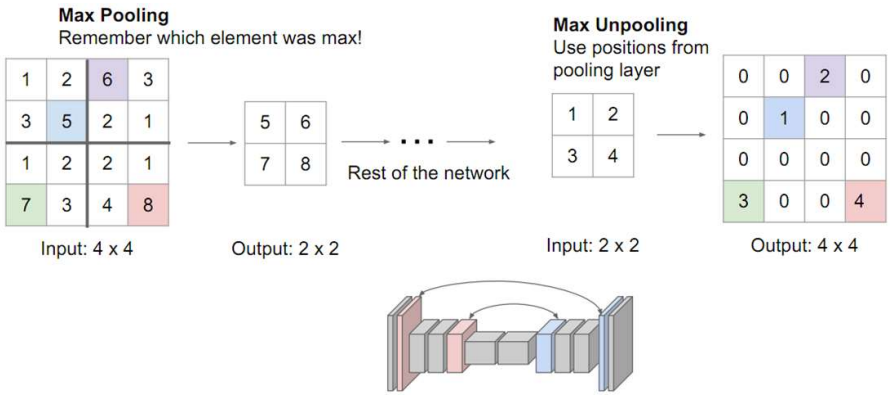
Unpooling

- Thực hiện ngược lại pooling



59

"Max Unpooling"



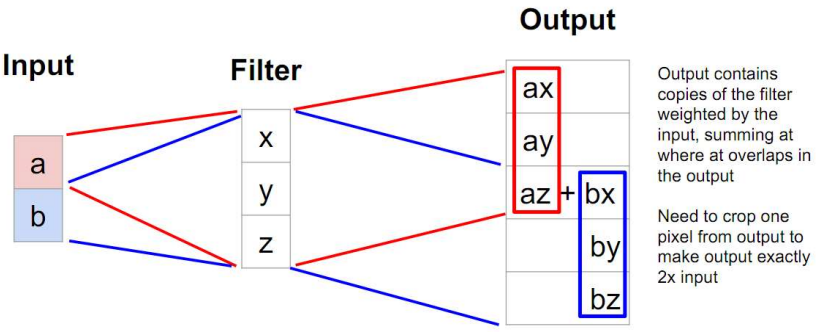
60

Transpose convolution layer

- Thực hiện ngược lại phép nhân chập để nâng độ phân giải.
- Một số tên gọi khác
  - Deconvolution (không nên dùng)
  - Upconvolution
  - Backward strided convolution
  - Fractionally strided convolution

61

Transpose convolution: Ví dụ 1D

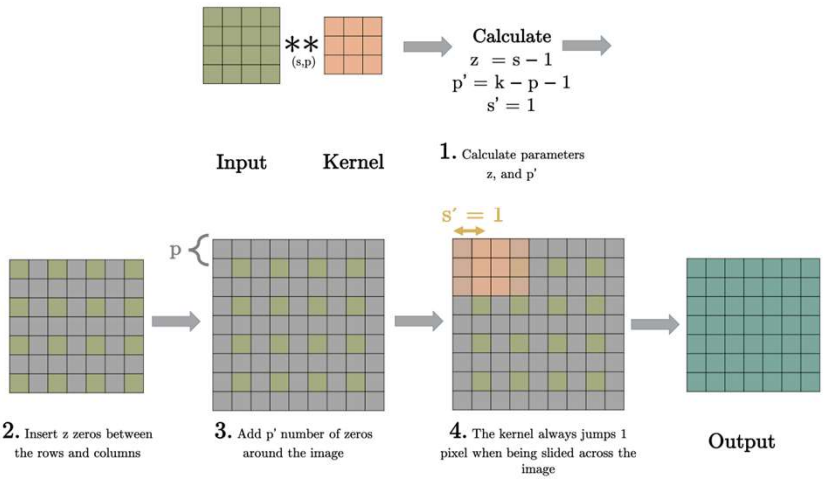


Output contains copies of the filter weighted by the input, summing at where at overlaps in the output

Need to crop one pixel from output to make output exactly 2x input

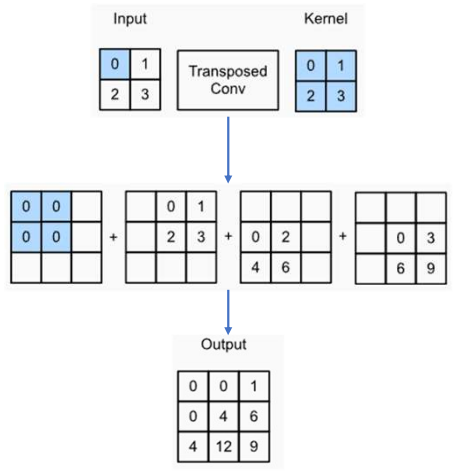
62

Tranpose convolution: 2D



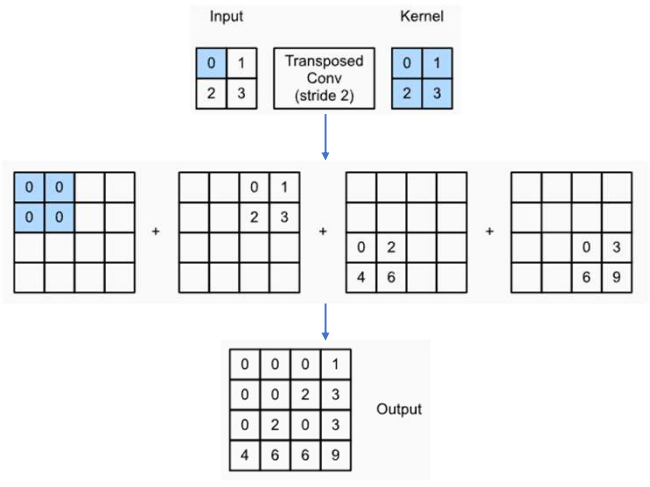
63

Tranpose convolution: ví dụ 2D



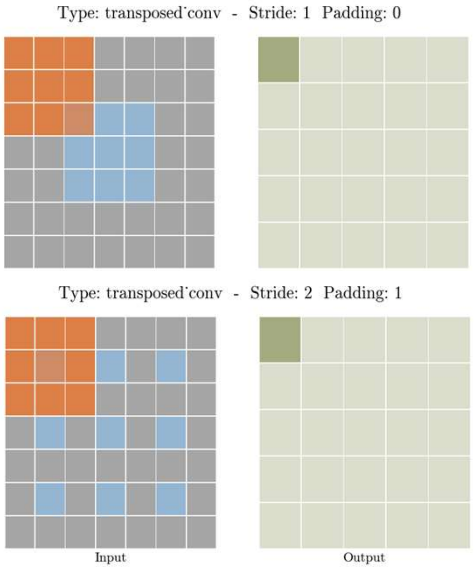
64

Tranpose convolution: ví dụ 2D



65

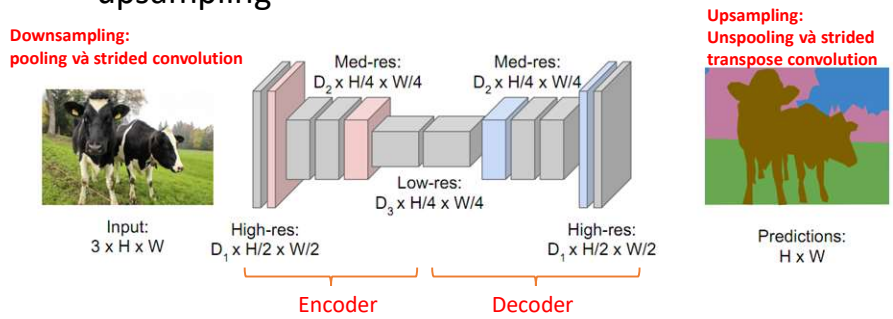
Tranpose convolution: ví dụ



66

Tích chập hoàn toàn (FCN)

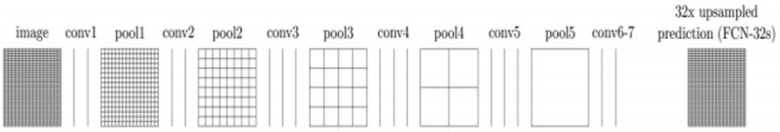
- Thiết kế mạng có nhánh downsampling và upsampling

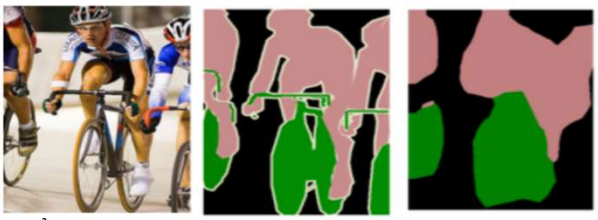


67

Skip connection (kết nối tắt)

- Quá trình downsampling/ mã hóa làm giảm độ phân giải đầu vào → quá trình upsampling/ giải mã khó phân đoạn chi tiết ảnh



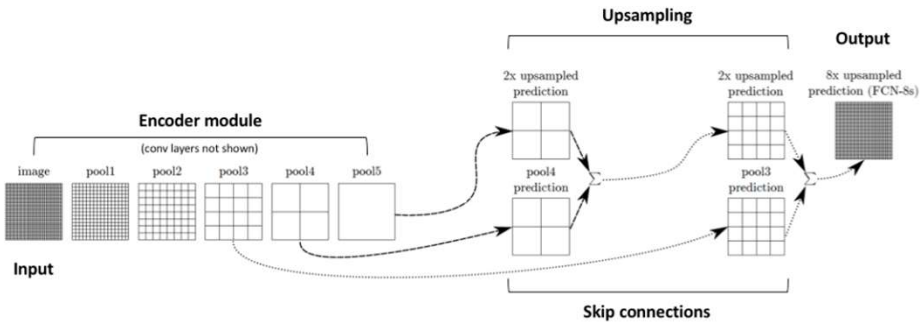


Ảnh đầu vào | Ground truth | Kết quả phân đoạn

68

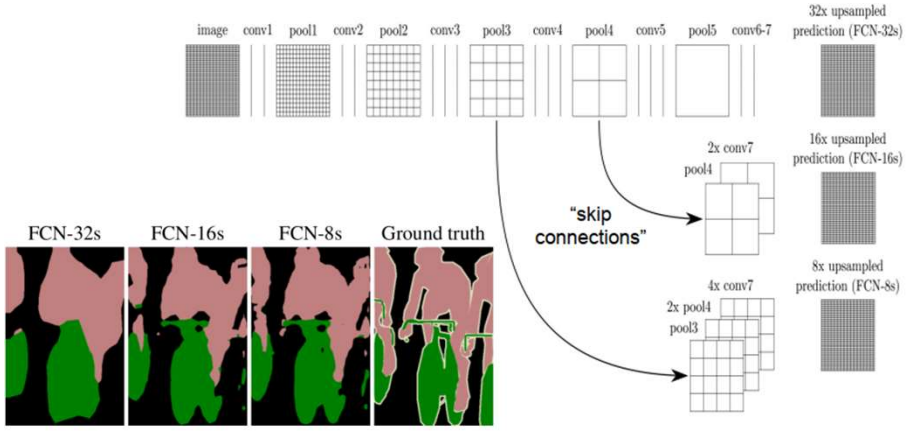
Skip connection (kết nối tắt)

- Thêm các kết nối tắt các dữ liệu độ phân giải cao hơn (đã bị loại bỏ) vào kết quả trong quá trình upsampling → tăng độ chi tiết phân loại



69

Kết quả giữa có và không có kết nối tắt



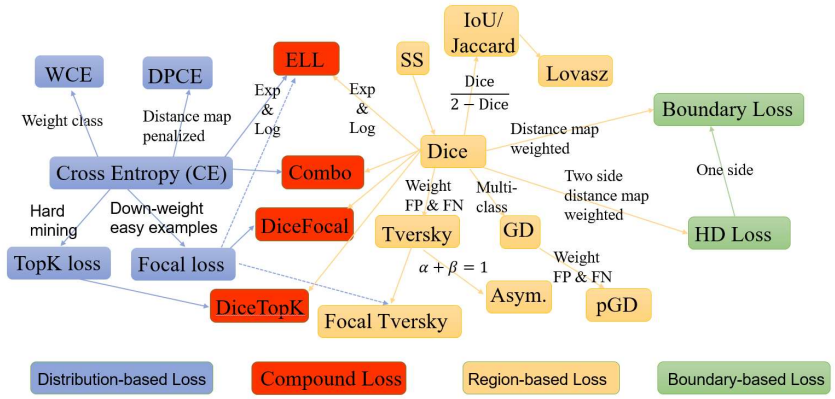
70

Lưu ý khi dùng kết nối tắt

- Không gian giữ liệu phải giống nhau
- Hai phép toán phổ biến
  - Cộng (addition)
  - Nối (concatenation)
- Hai loại kết nối tắt
  - Kết nối tắt ngắn (short skip connection) : những lớp conv liên tiếp không có sự thay đổi kích thước dữ liệu (vd: ResNet)
  - Kết nối tắt dài (long skip connection): dùng cho những mạng đối xứng có sự thay đổi kích thước dữ liệu, cần áp dụng cho từng lớp mã hóa/ giải mã tương ứng cùng kích thước (FCN, Unet)

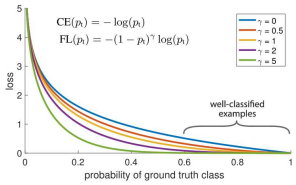
71

Hàm mục tiêu



72

Hàm mục tiêu dựa trên phân phối



- Cross Entropy (CE):

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

- Weighted CE: mỗi lớp có trọng số khác nhau

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

- 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.

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

73

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:

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

- Dice loss + Focal loss

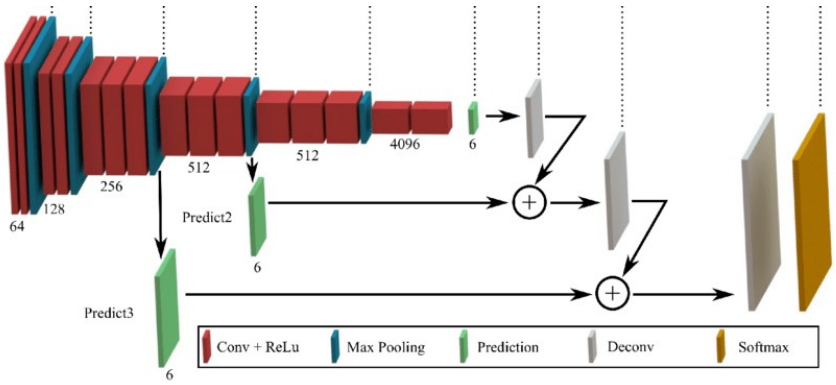
$$\mathrm{CE}(p, \hat{p}) + \mathrm{FL}(p, \hat{p})$$

- ...

Một số mạng phân đoạn ảnh tiêu biểu

78

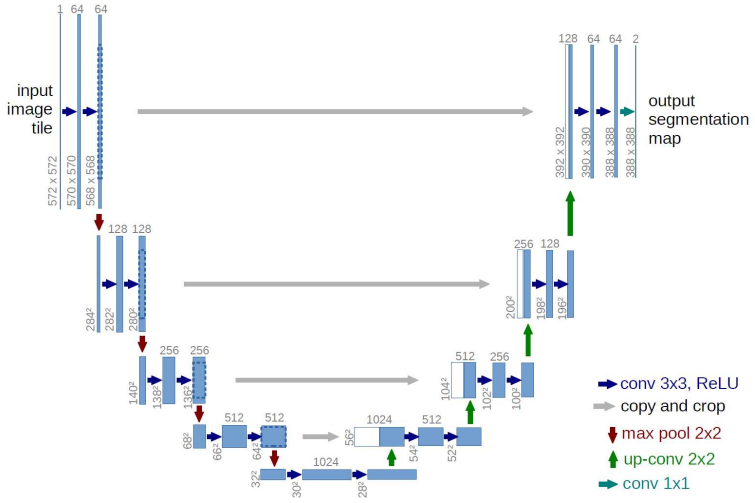
FCN với 2 kết nối tắt



79

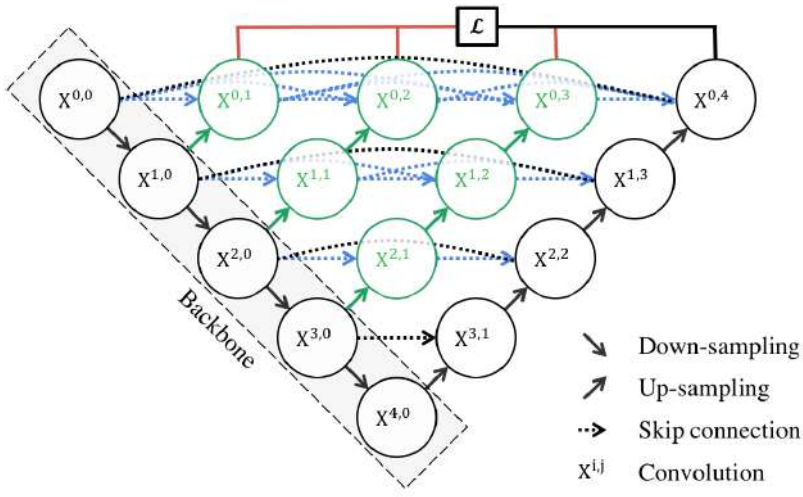
U-Net

- Được sử dụng rộng rãi trong y tế



81

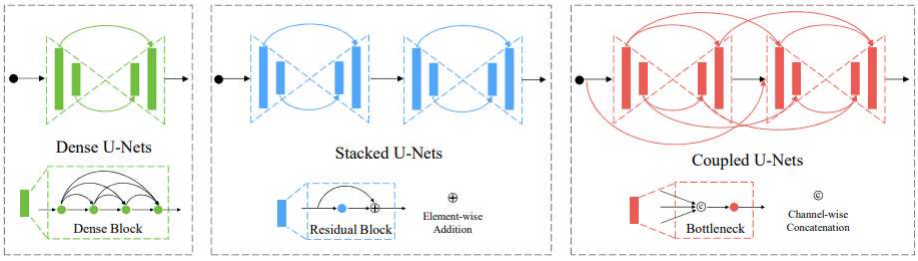
U-Net++



82

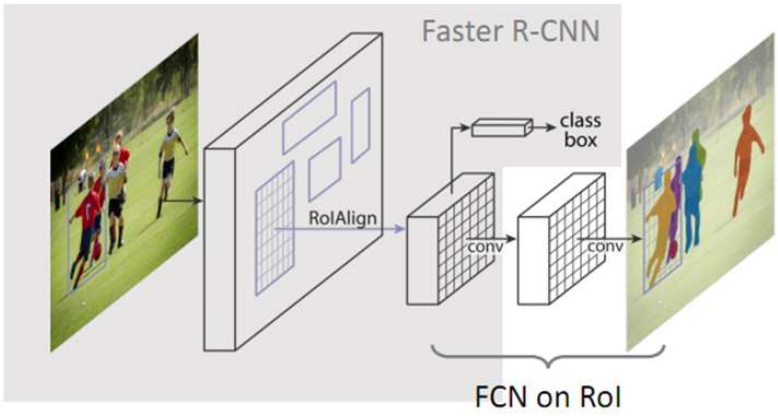
Stacked UNets và CUNets

- 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



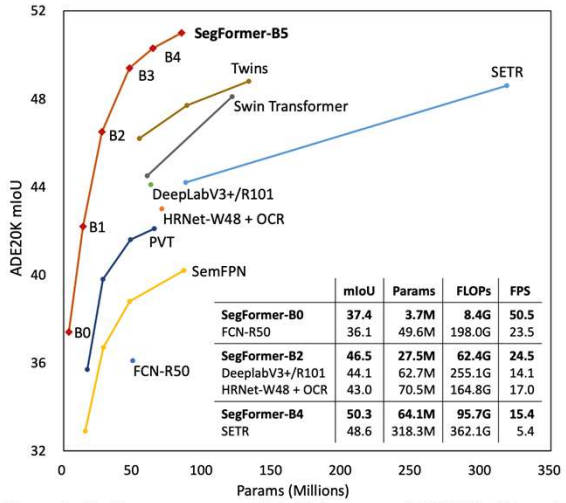
83

Mask R-CNN



84

Transformer-based models



	mIoU	Params	FLOPs	FPS
SegFormer-B0	**37.4**	**3.7M**	**8.4G**	**50.5**
FCN-R50	36.1	49.6M	198.0G	23.5
SegFormer-B2	**46.5**	**27.5M**	**62.4G**	**24.5**
DeeplabV3+/R101	44.1	62.7M	255.1G	14.1
HRNet-W48 + OCR	43.0	70.5M	164.8G	17.0
SegFormer-B4	**50.3**	**64.1M**	**95.7G**	**15.4**
SETR	48.6	318.3M	362.1G	5.4

Figure 1: **Performance *vs.* model efficiency on ADE20K.** All results are reported with single model and single-scale inference. SegFormer achieves a new state-of-the-art 51.0% mIoU while being significantly more efficient than previous methods.

Tài liệu tham khảo

- Kristen Grauman (CS 376: Computer Vision, Spring 2018, The University of Texas at Austin)
- Stanford CS231n Course: http://cs231n.stanford.edu/


25
SOICT
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY
Thank you for your attentions!
soict.hust.edu.vn/
fb.com/groups/soict
B1

88