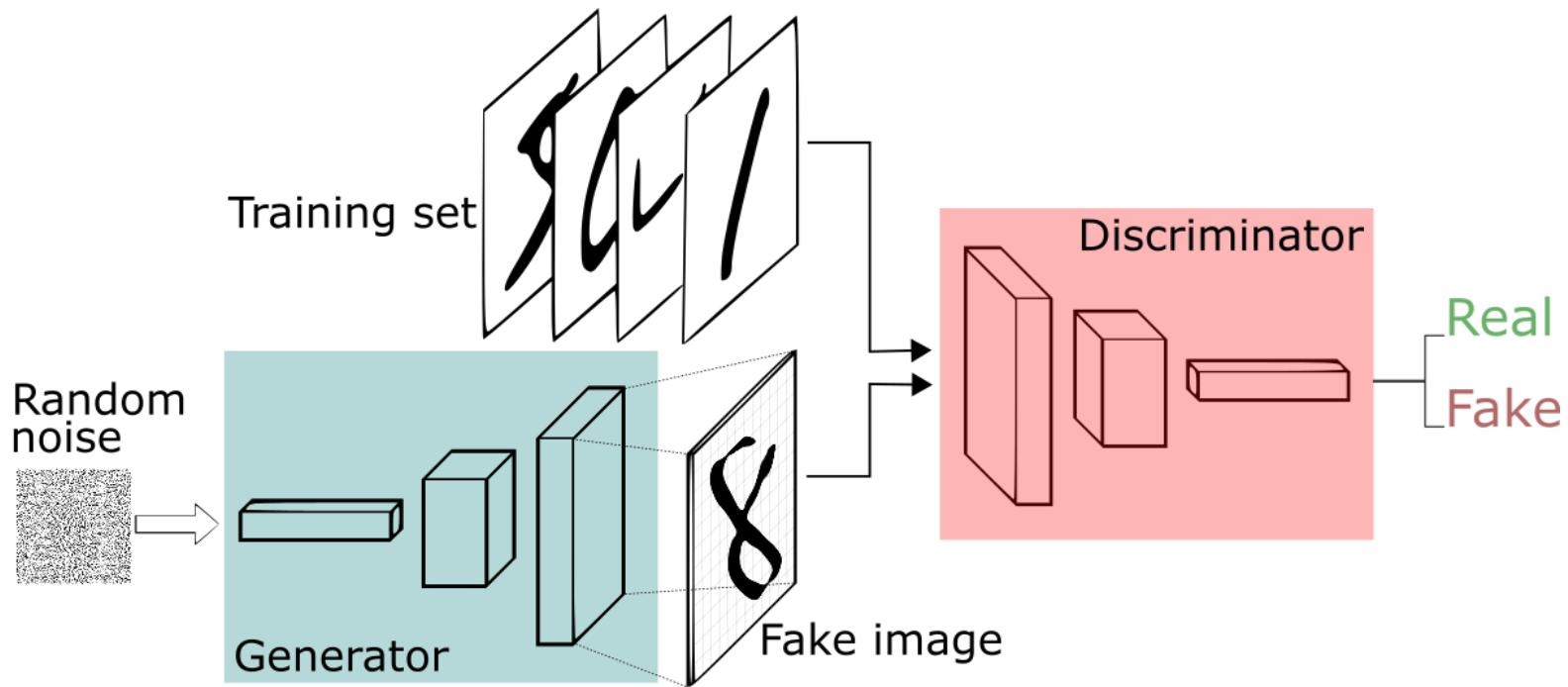

Segmentation and Object detection

— Tuan Nguyen - AI4E —

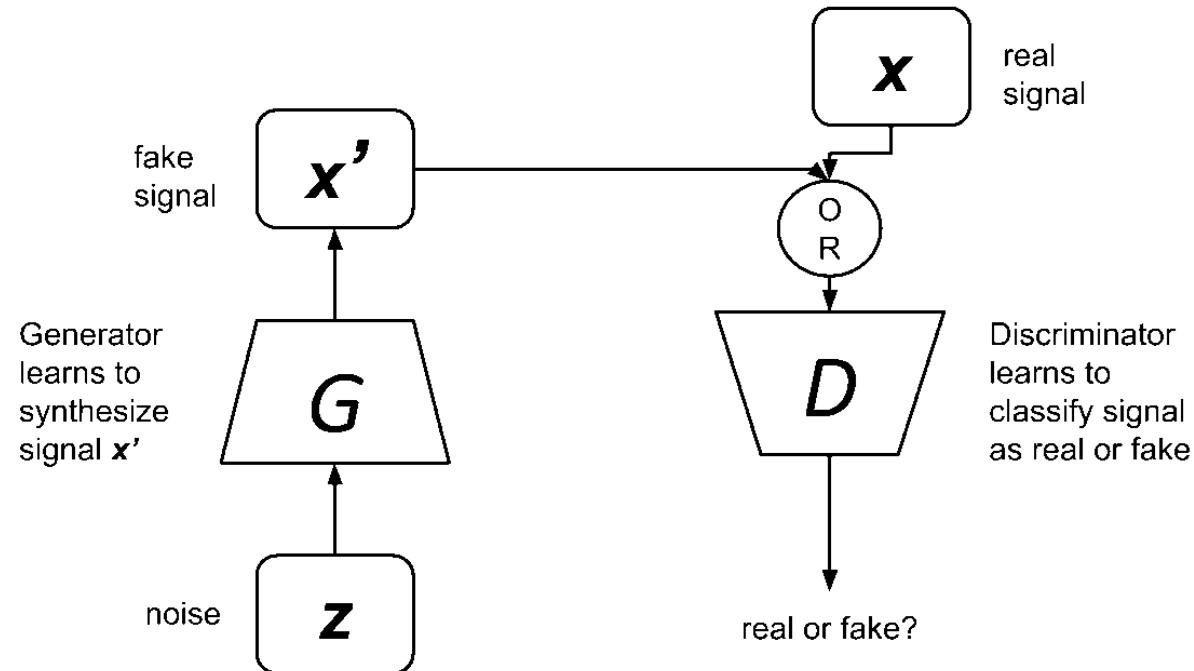
Outline

- GAN reviews
- Computer vision task
- Semantic vs Instance segmentation
- Fully convolutional segmentation
- U-Net
- Object Detection
- R-CNN, Fast R-CNN, Faster R-CNN

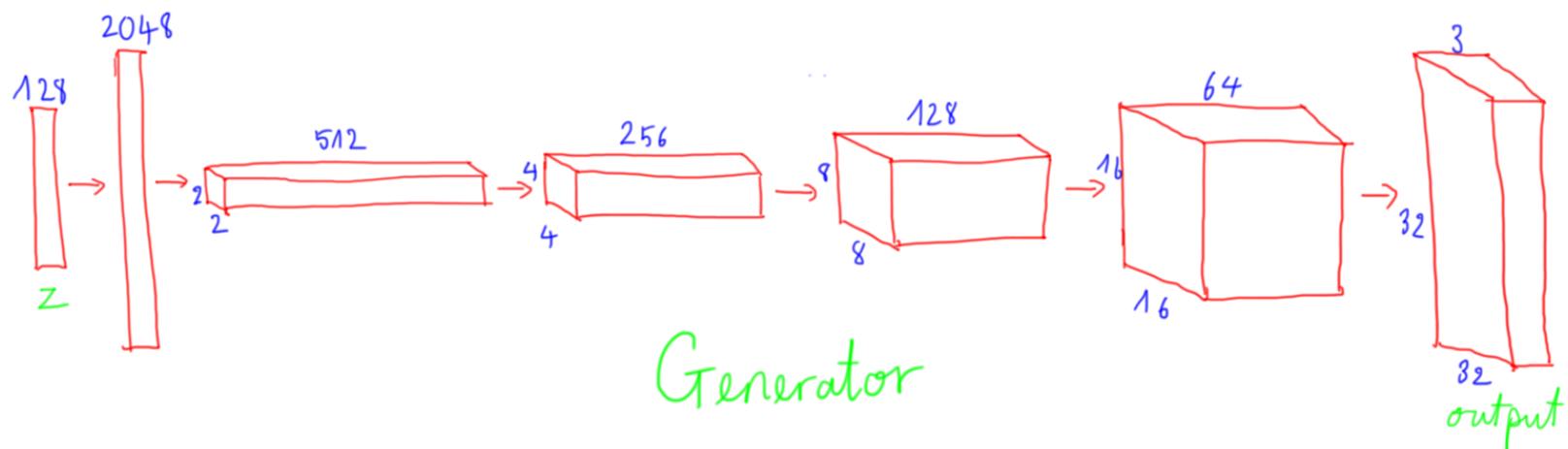
GAN model



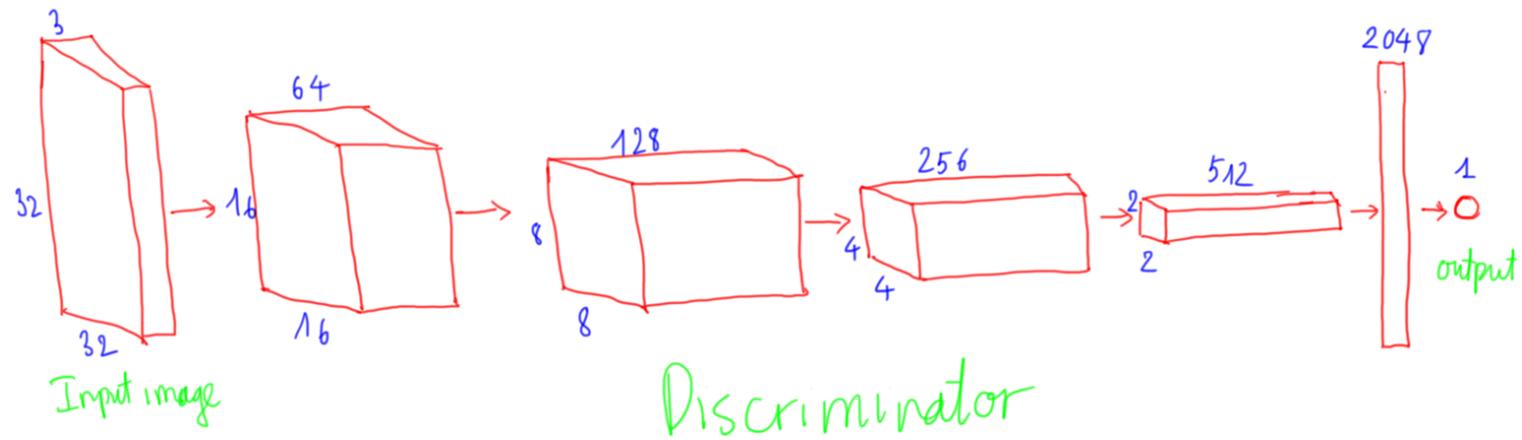
GAN model



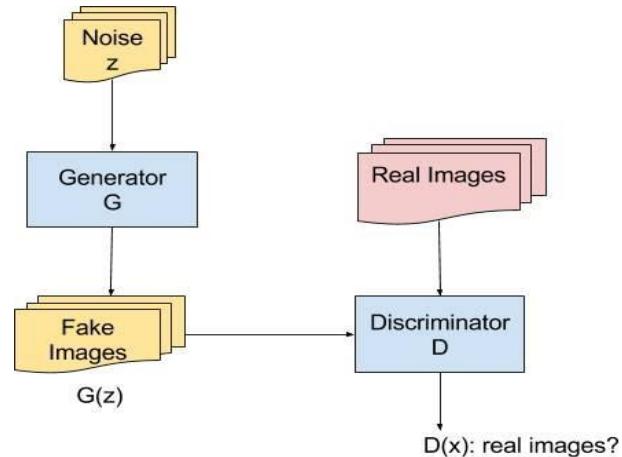
Generator



Discriminator



Loss function



$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))].$$

Training

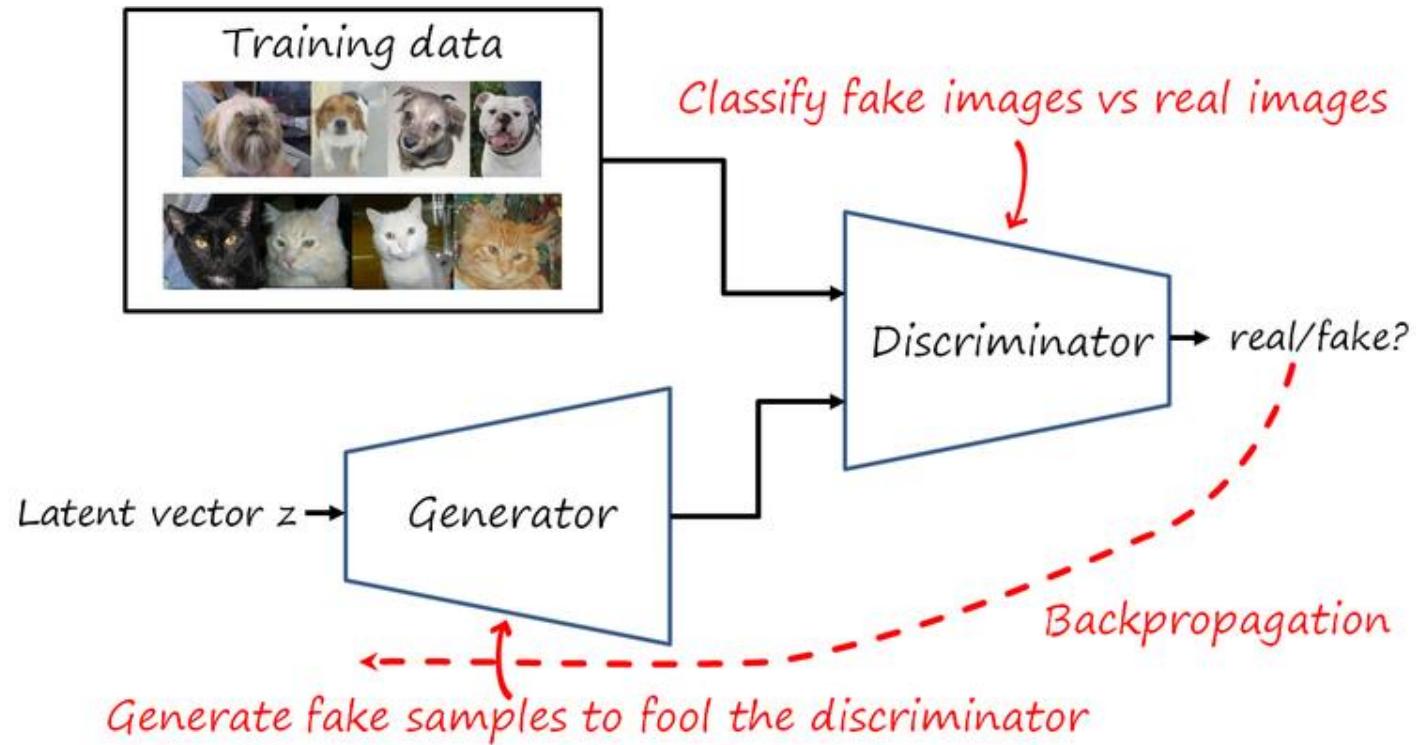
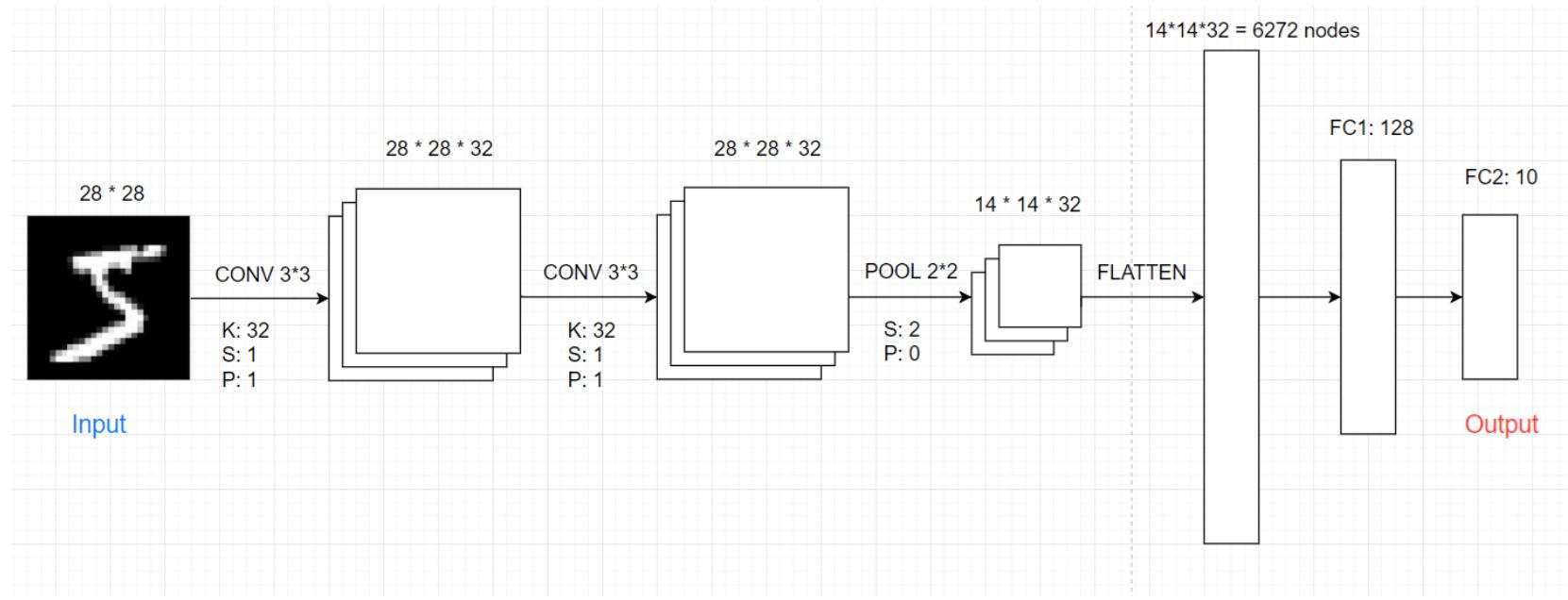
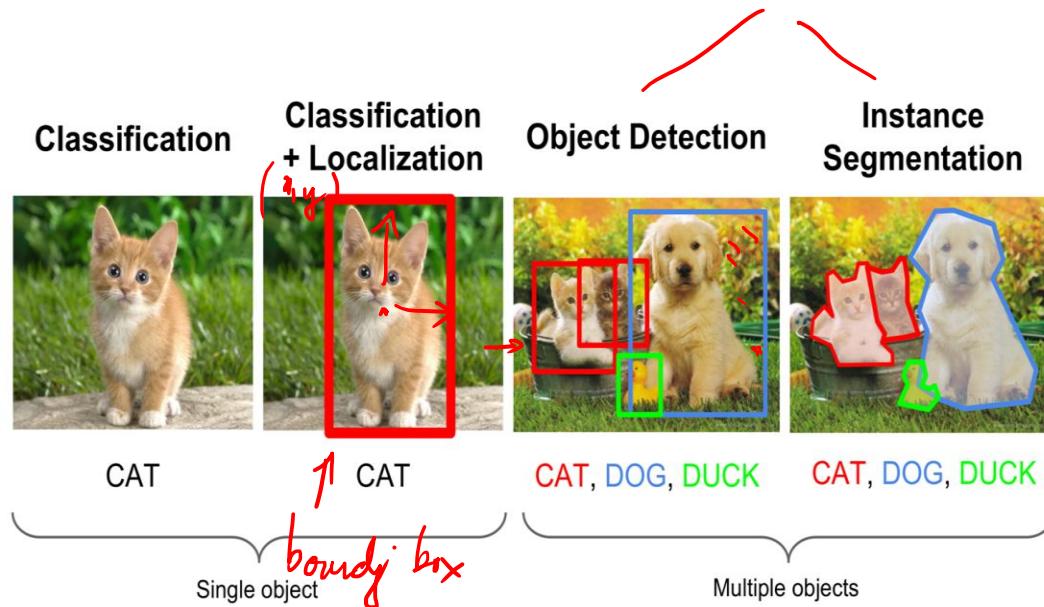


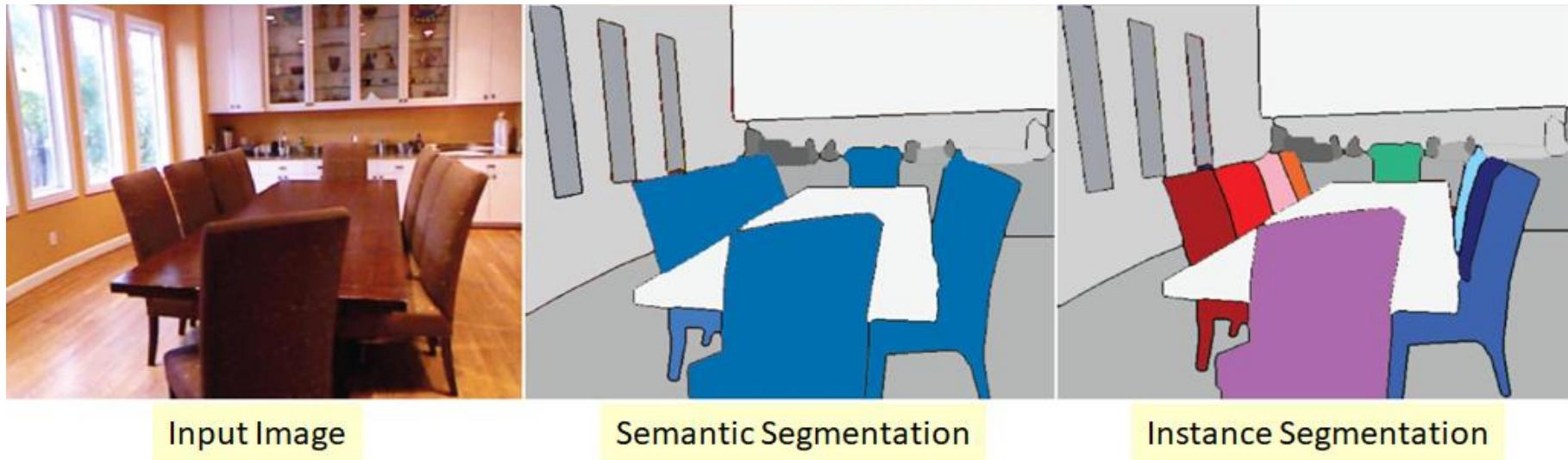
Image classification



Computer vision tasks



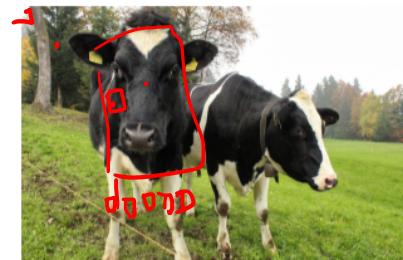
Semantic vs Instance segmentation



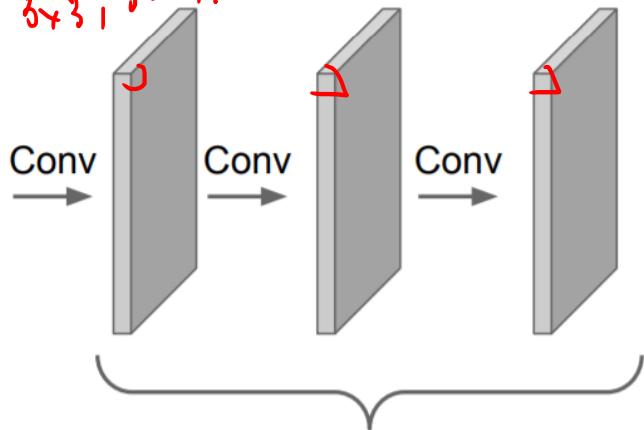
Fully convolutional

phân loại môt pixel → thuỷ lấp nòi truy & lấp

$3 \times 3, s=1, p=1$.

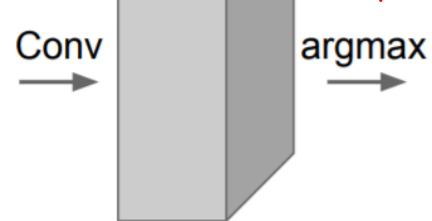


Input:
3 x H x W



Convolutions:
D x H x W

loss
Categorical-crossentropy
tùi pixel



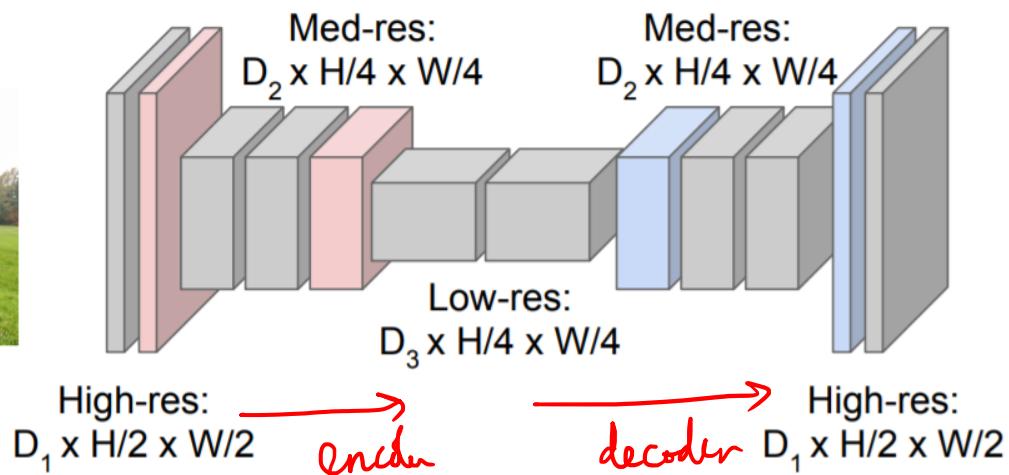
Scores:
C x H x W

Predictions:
H x W

Fully convolutional improvement



Input:
 $3 \times H \times W$

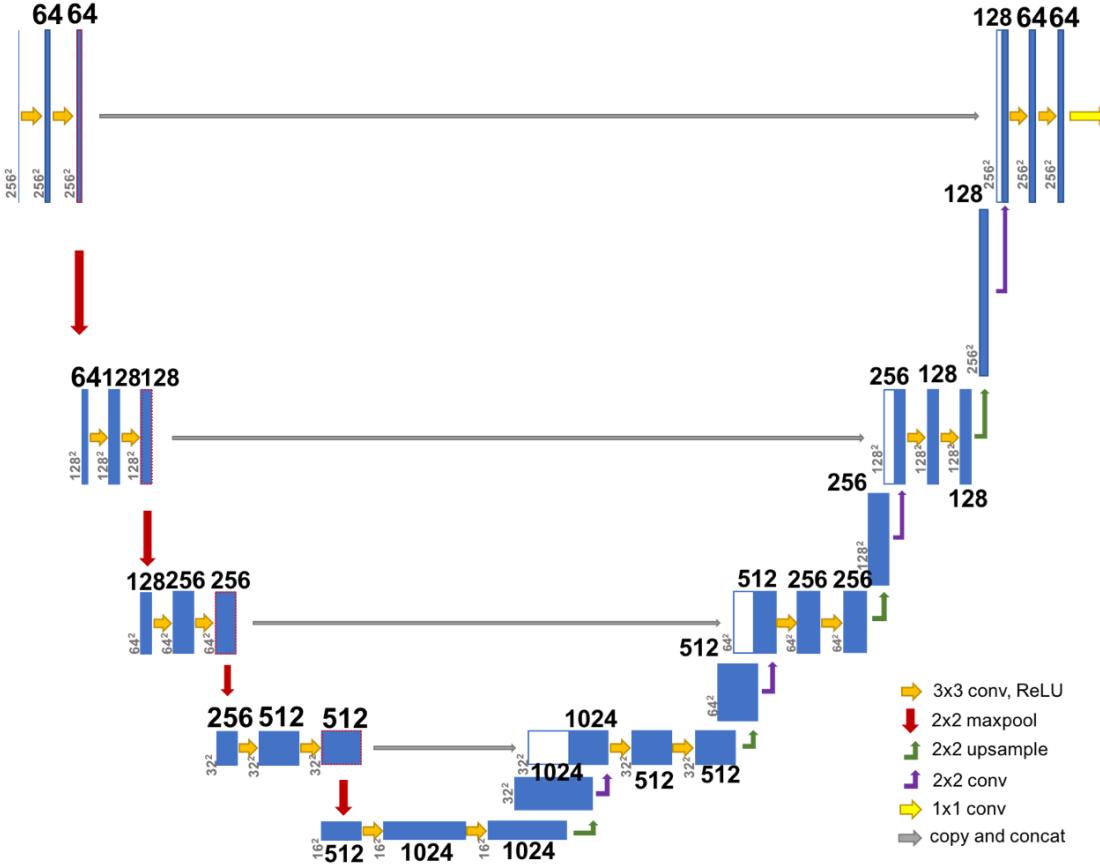


Predictions:
 $H \times W$

HAWC..

dersnet .

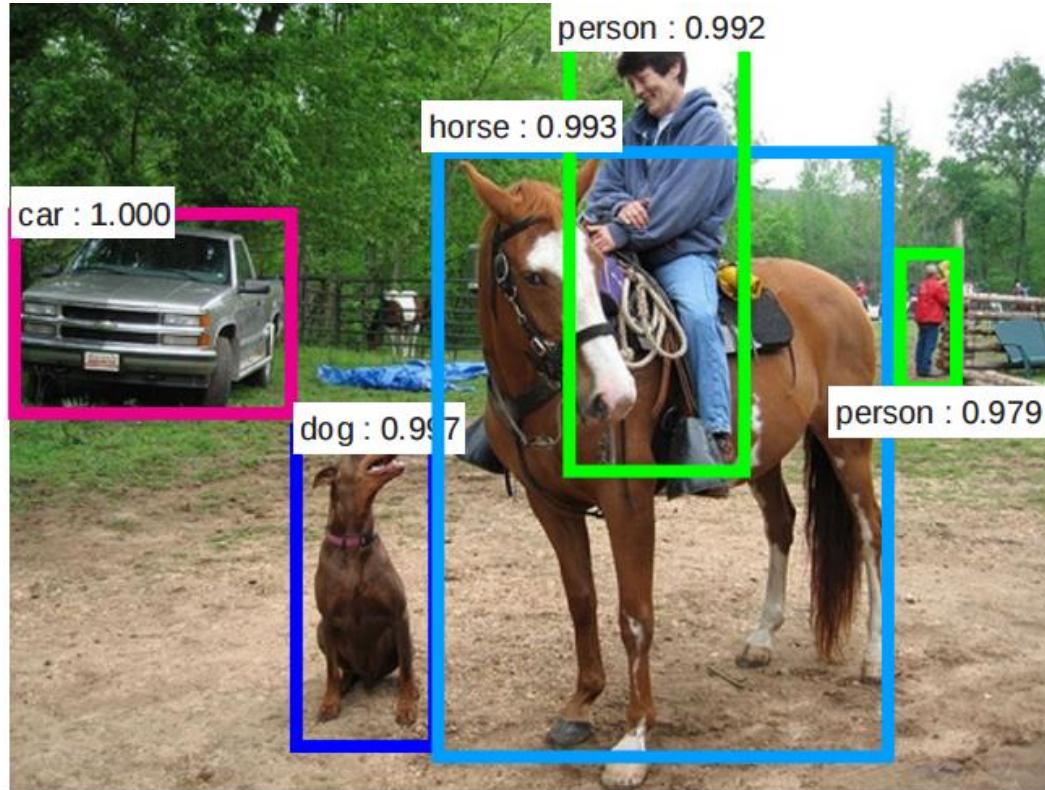
Unet



Code

<https://github.com/zhixuhao/unet/blob/master/model.py>

Object detection



R-CNN

- Bước 1: Dùng Selective Search algorithm để lấy ra khoảng 2000 bounding box trong input mà có khả năng chứa đối tượng.
- Bước 2: Với mỗi bounding box ta xác định xem nó là đối tượng nào (người, ô tô, xe đạp,...)

Graph based image segmentation



Input Image



Output Image

R-CNN

R-CNN: *Regions with CNN features*

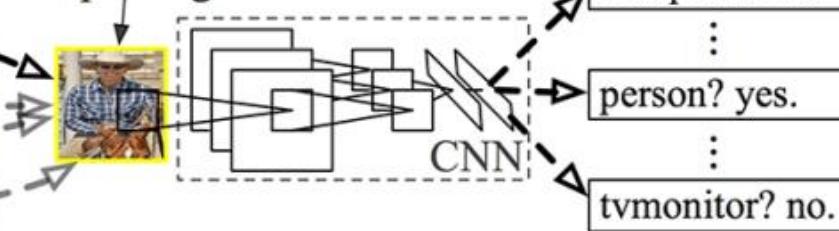


1. Input
image



2. Extract region
proposals (~2k)

warped region



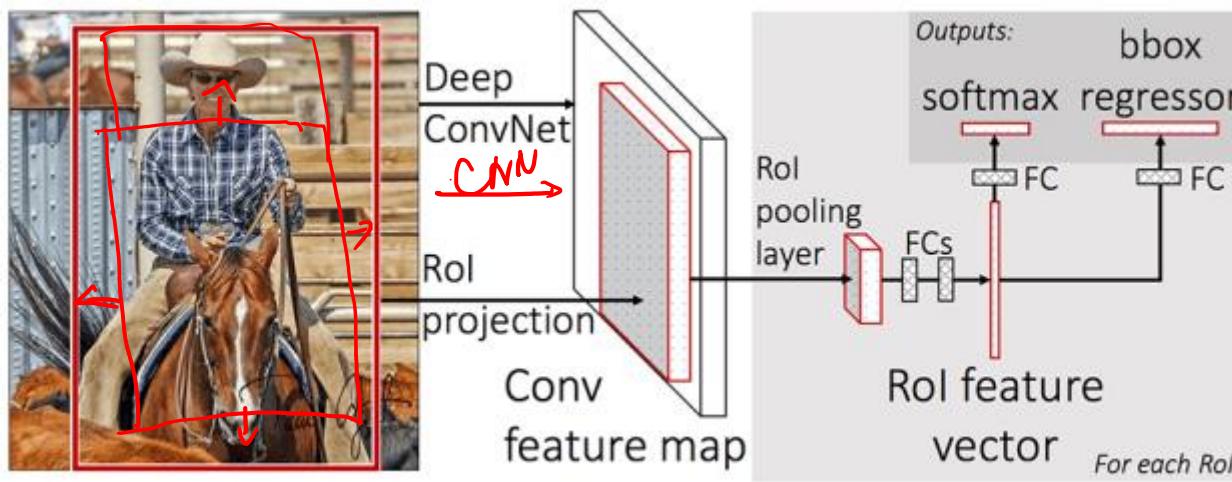
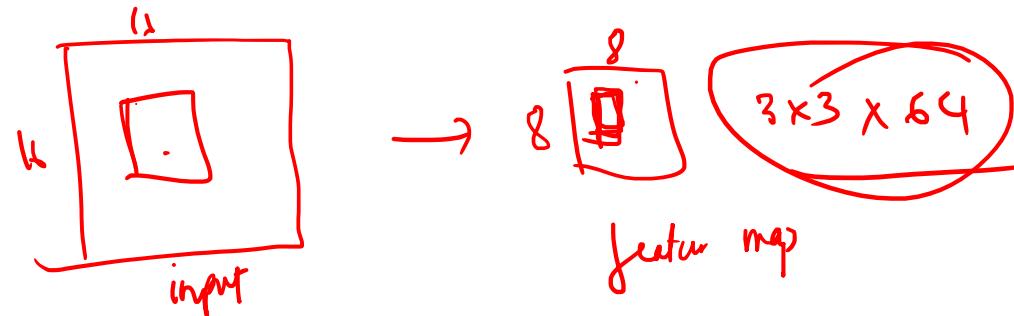
3. Compute
CNN features

4. Classify
regions

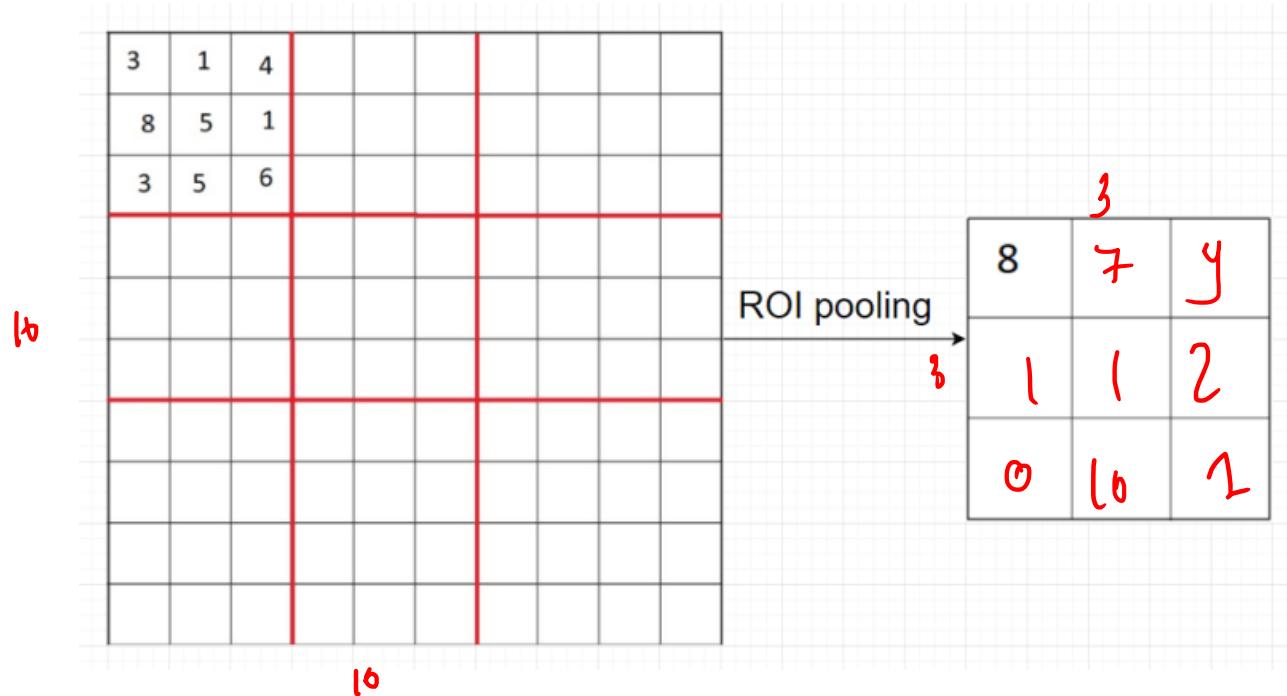
Problem?

- Vì với mỗi ảnh ta cần phân loại các class cho 2000 region proposal nên thời gian train rất lâu.
- Không thể áp dụng cho real-time thì mỗi ảnh trong test set mất tới 47s để xử lý.

Fast R-CNN

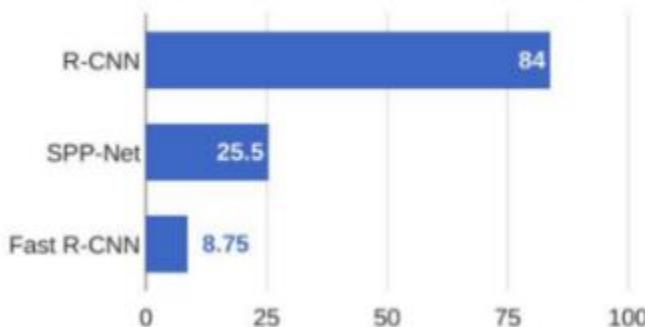


Region of Interest (ROI) pooling

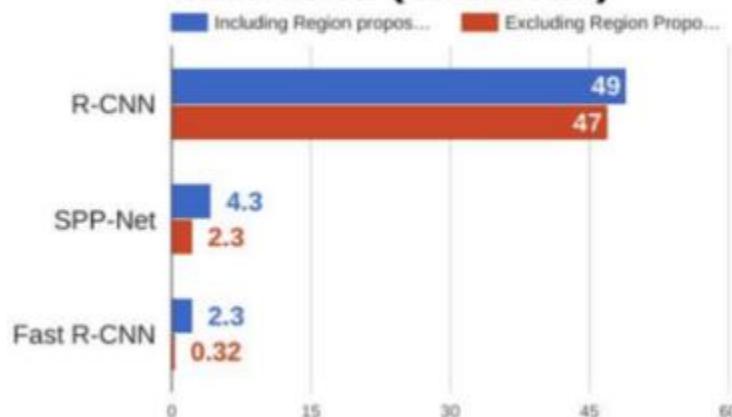


R-CNN vs Fast R-CNN

Training time (Hours)

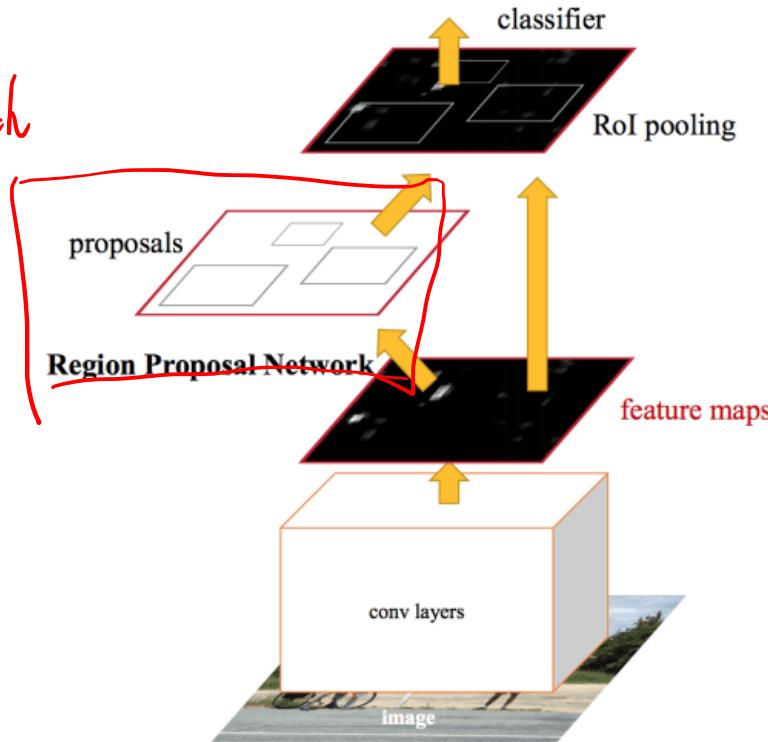


Test time (seconds)



Faster R-CNN

K^* obj selective search



Anchors

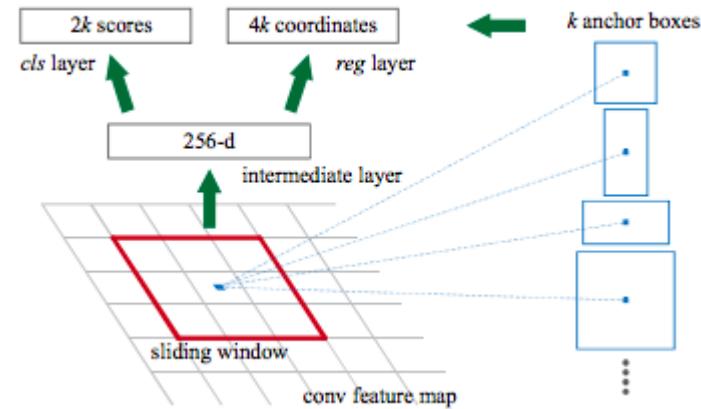


600 tam → 5400 box

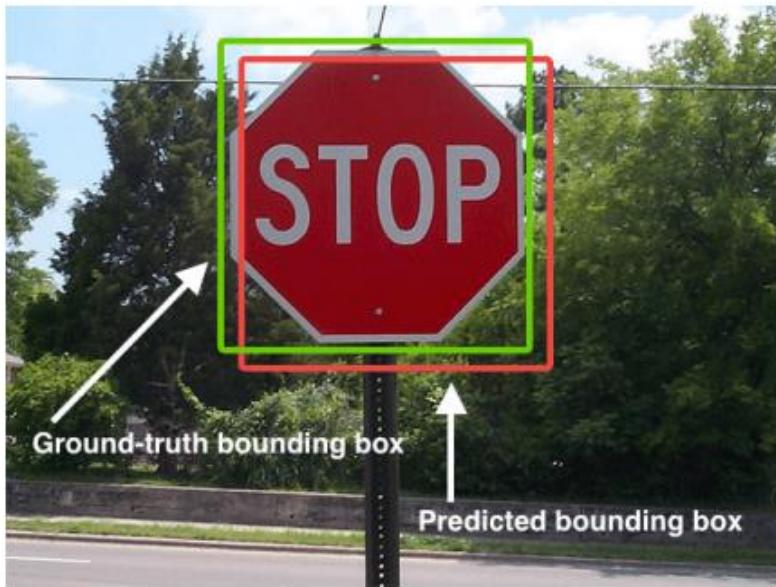
mõi tam
3 wj dt : 600, 200, 300
3 t'li : 1:1, 2:1, 3:2



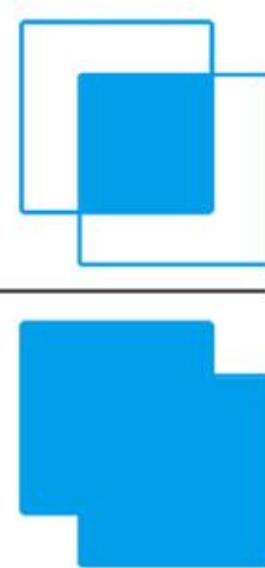
Region proposal network



Intersection over Union (IoU)

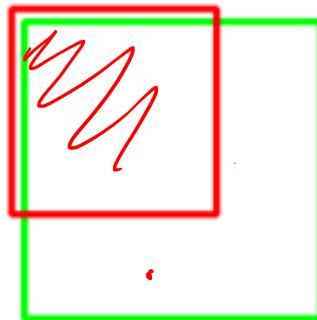


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



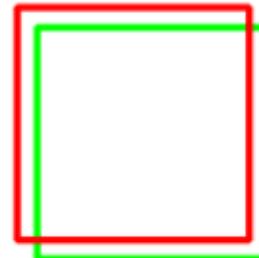
IoU

IoU: 0.4034



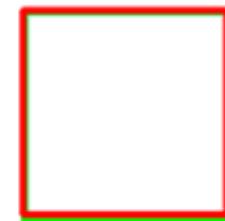
Poor

IoU: 0.7330



Good

IoU: 0.9264



Excellent

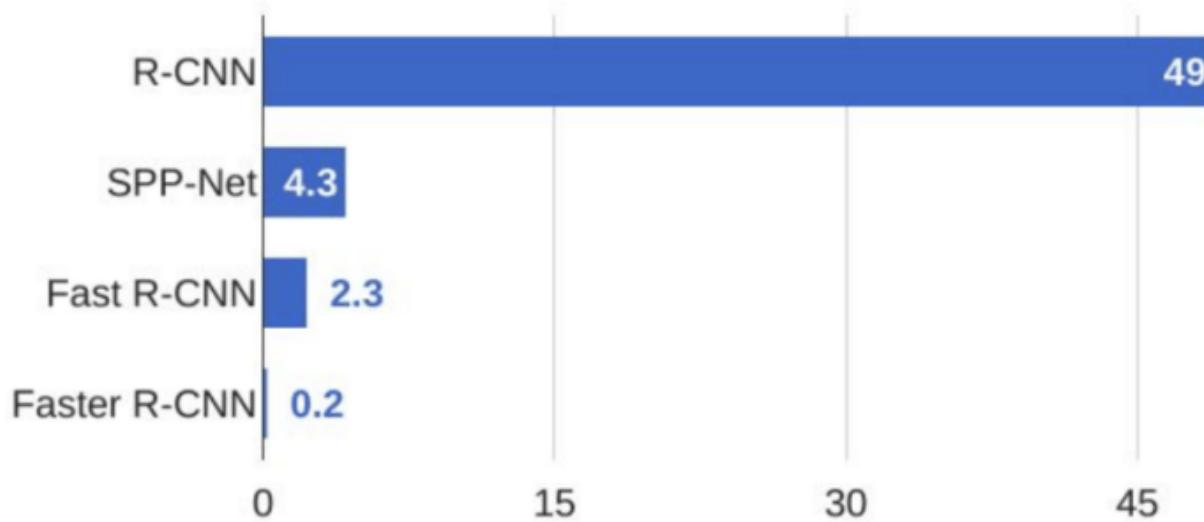
Non-maxima suppression

Output : 5400 anchors → 100 anchors

- Bước 1: Chọn ra anchor box (A) có xác xuất là foreground lớn nhất trong tập Input
- Bước 2: Thêm A vào tập Output.
- Bước 3: Loại bỏ A và các anchor box trong tập Input mà có hệ số IoU với A lớn hơn 0.5 ra khỏi tập Input.
- Bước 4: Kiểm tra nếu tập Input rỗng hoặc tập Output đủ 100 anchor thì dừng lại, nếu không quay lại bước 1.

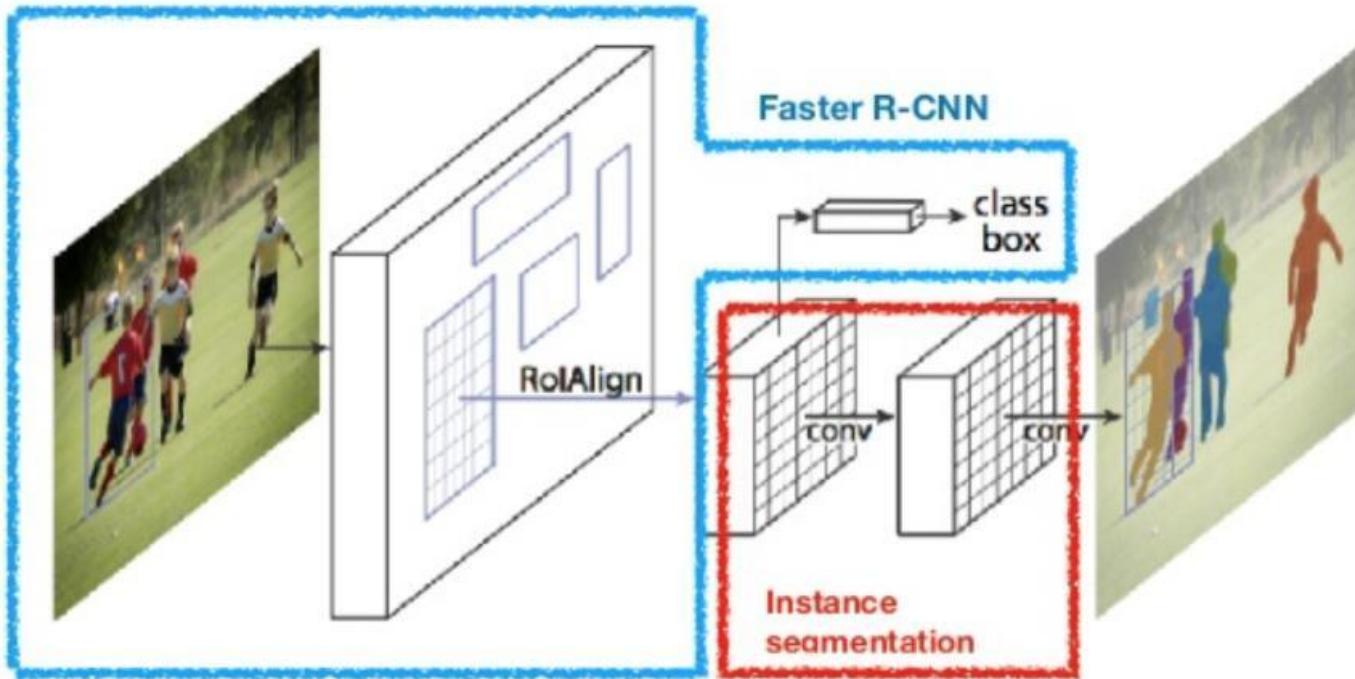
Comparison

R-CNN Test-Time Speed

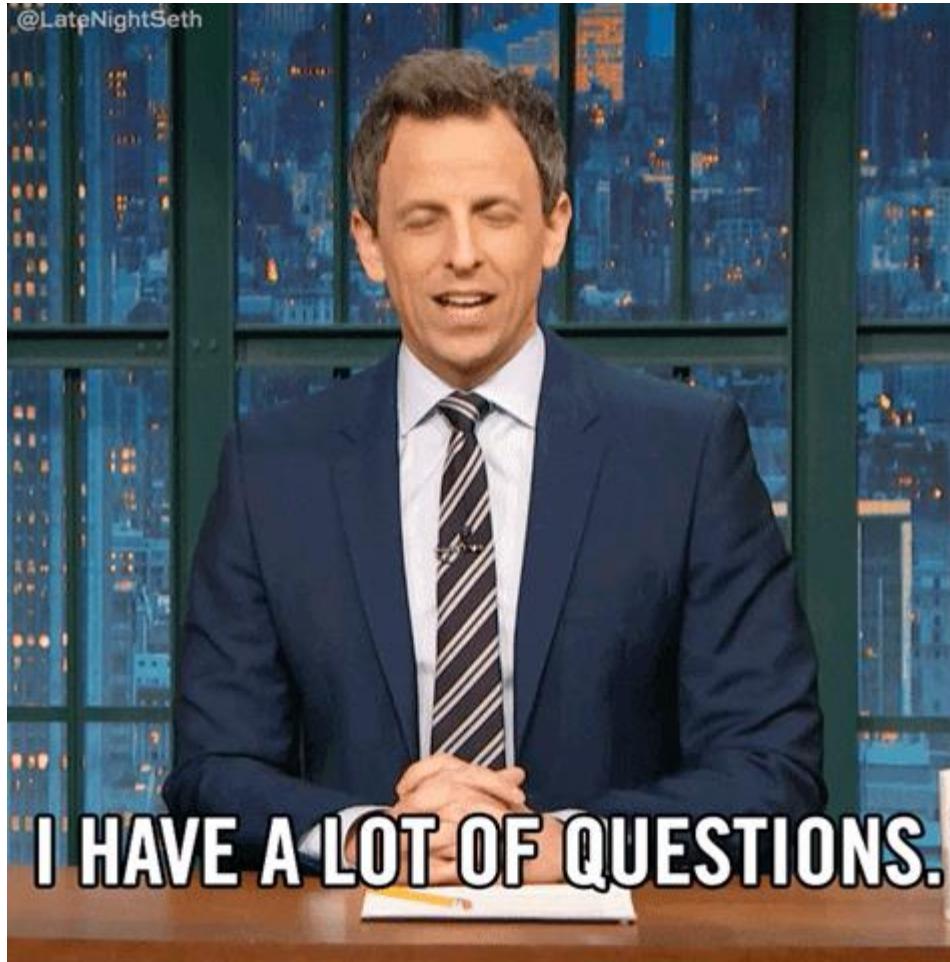


Comparison of test-time speed of object detection algorithms

Mask R-CNN



Q&A



@LateNightSeth

