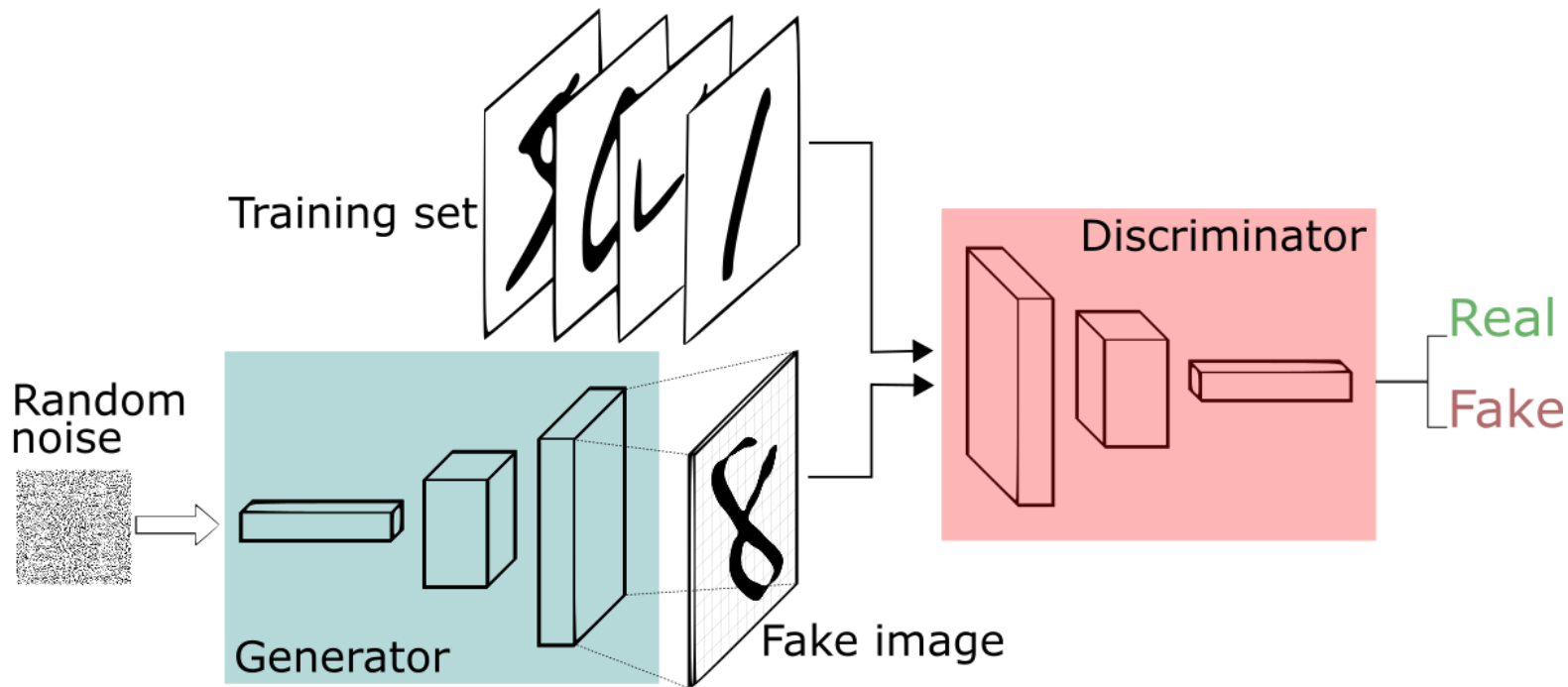

Segmentation and Object detection

— Tuan Nguyen - AI4E —

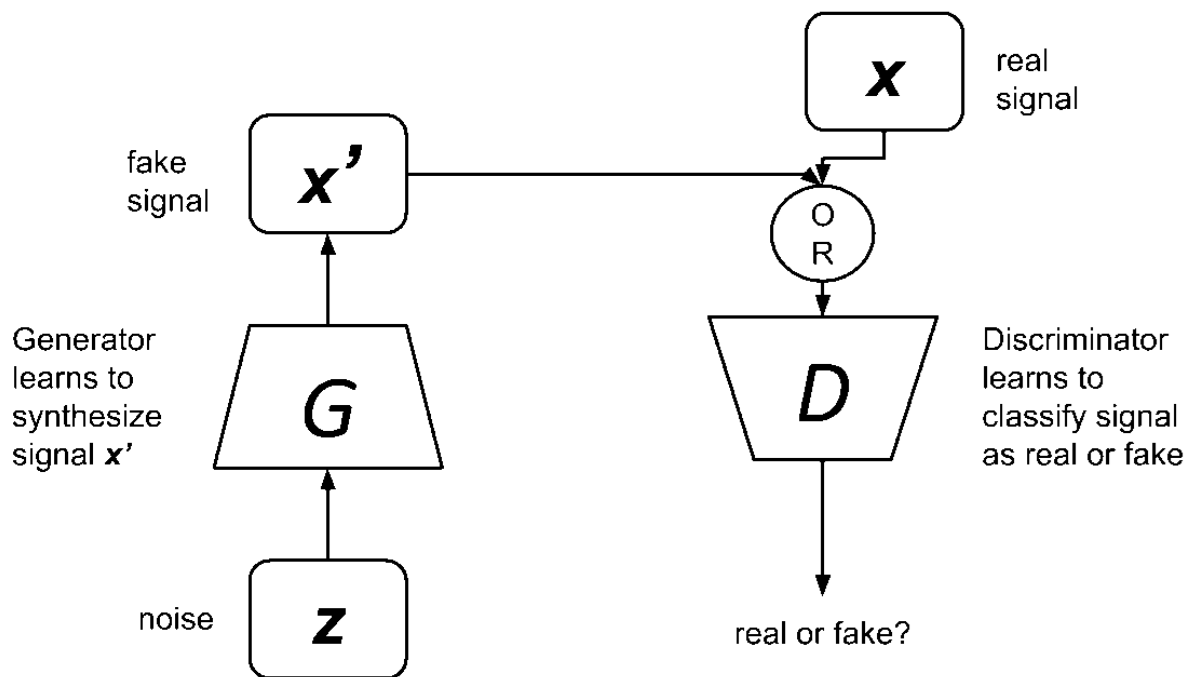
Outline

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

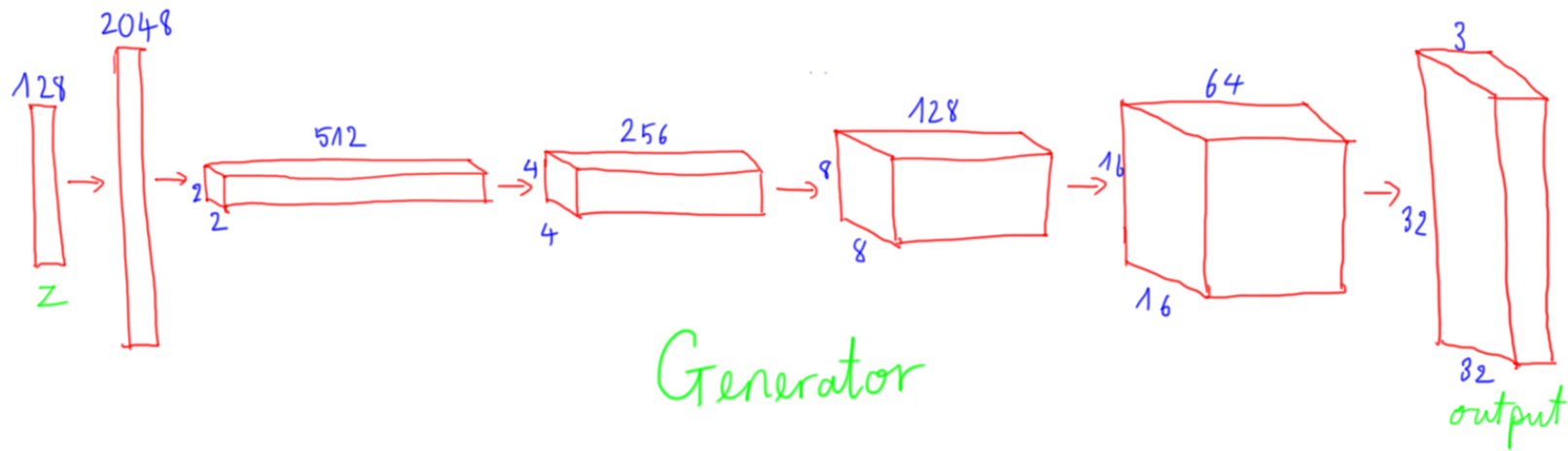
GAN model



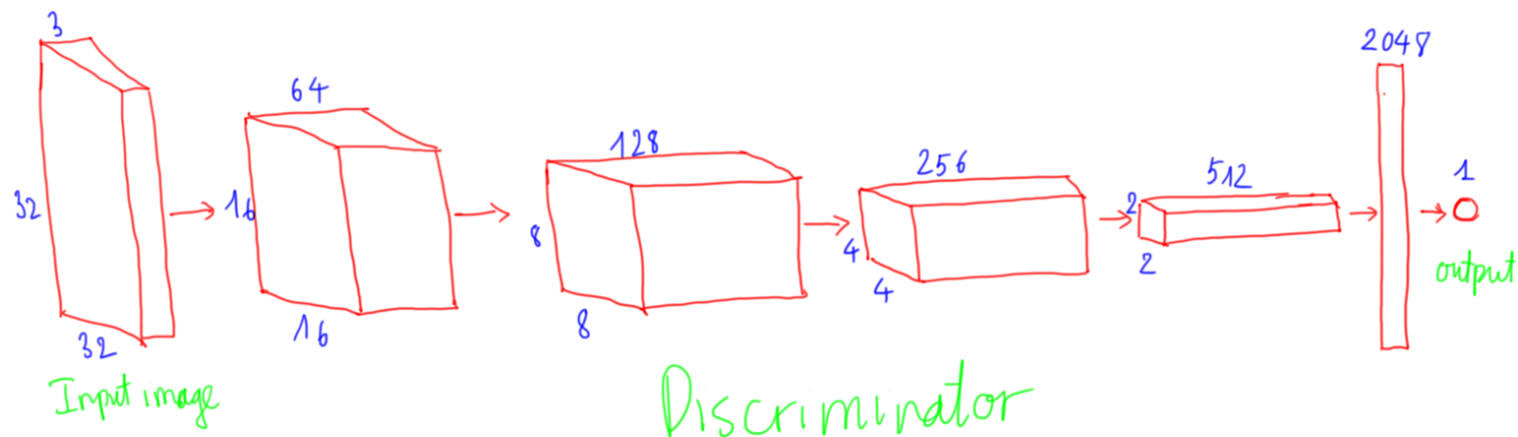
GAN model



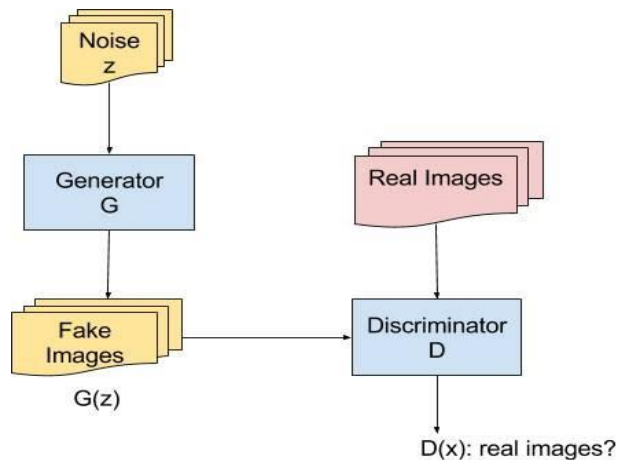
Generator



Discriminator



Loss function



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

Training

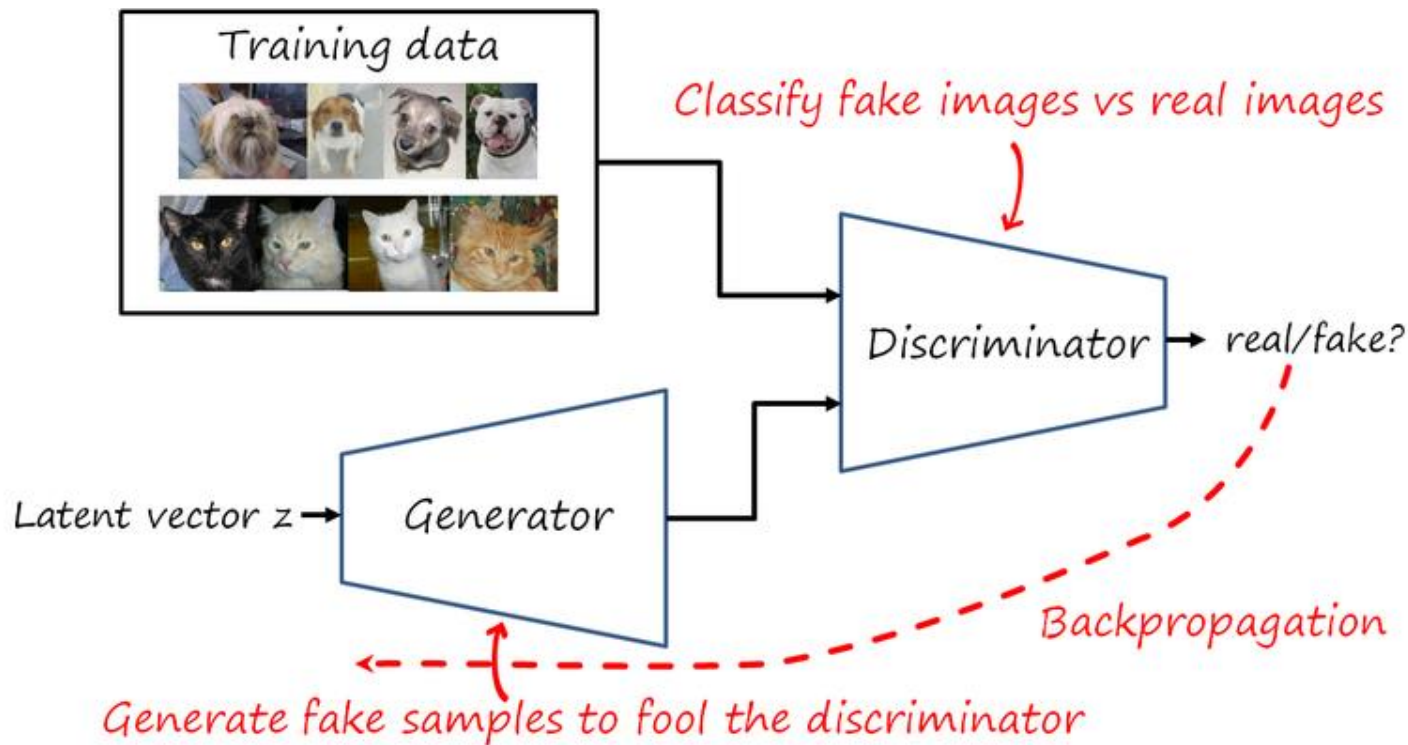
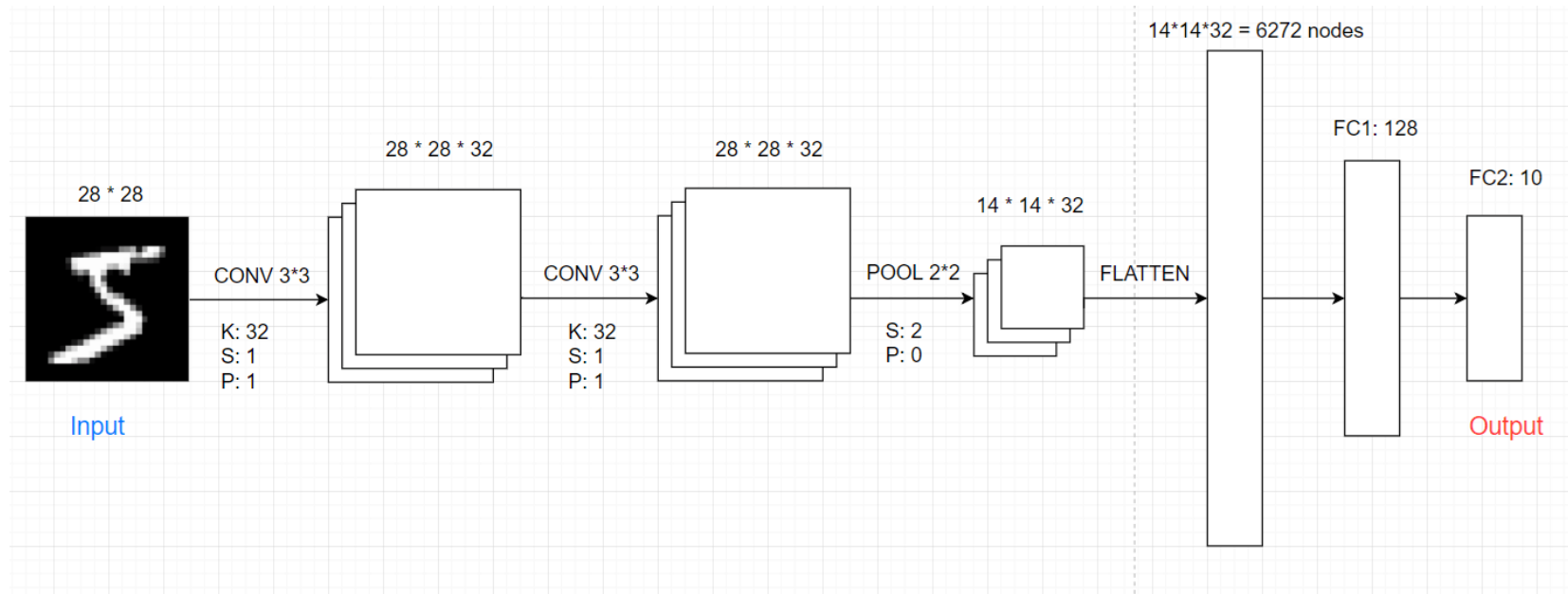
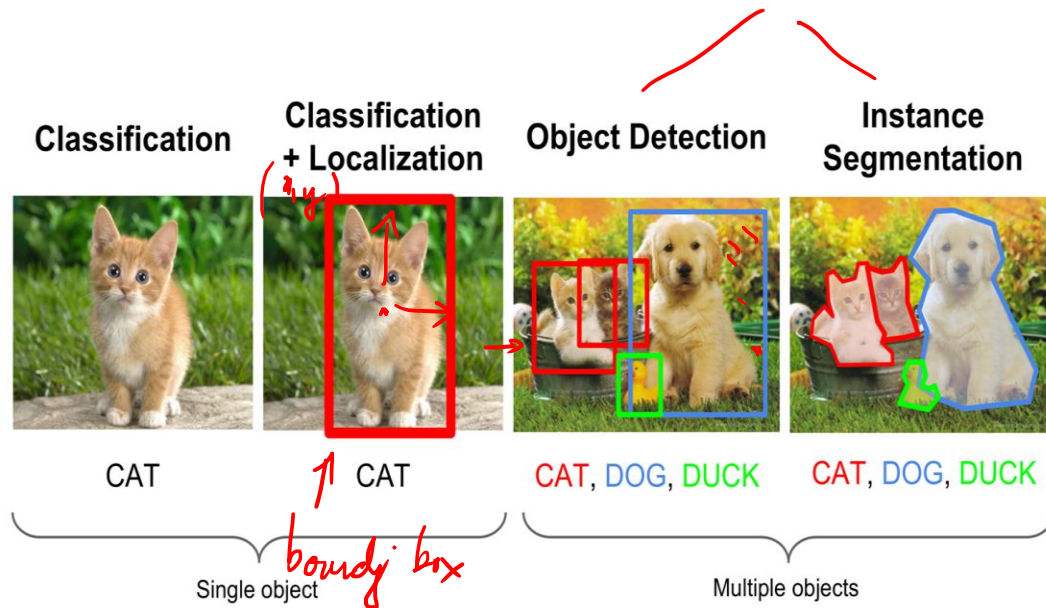


Image classification



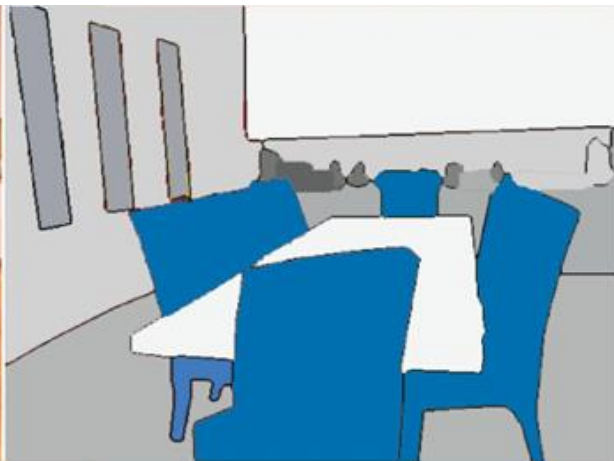
Computer vision tasks



Semantic vs Instance segmentation



Input Image



Semantic Segmentation



Instance Segmentation

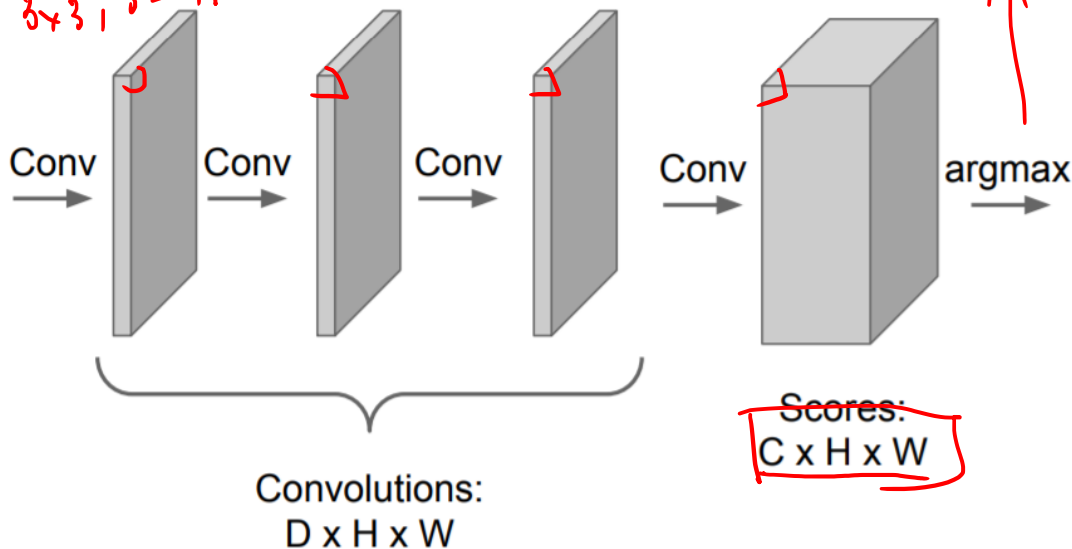
Fully convolutional

phân loại mỗi pixel \rightarrow thuộc lớp nào trong 4 lớp

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



Input:
 $3 \times H \times W$

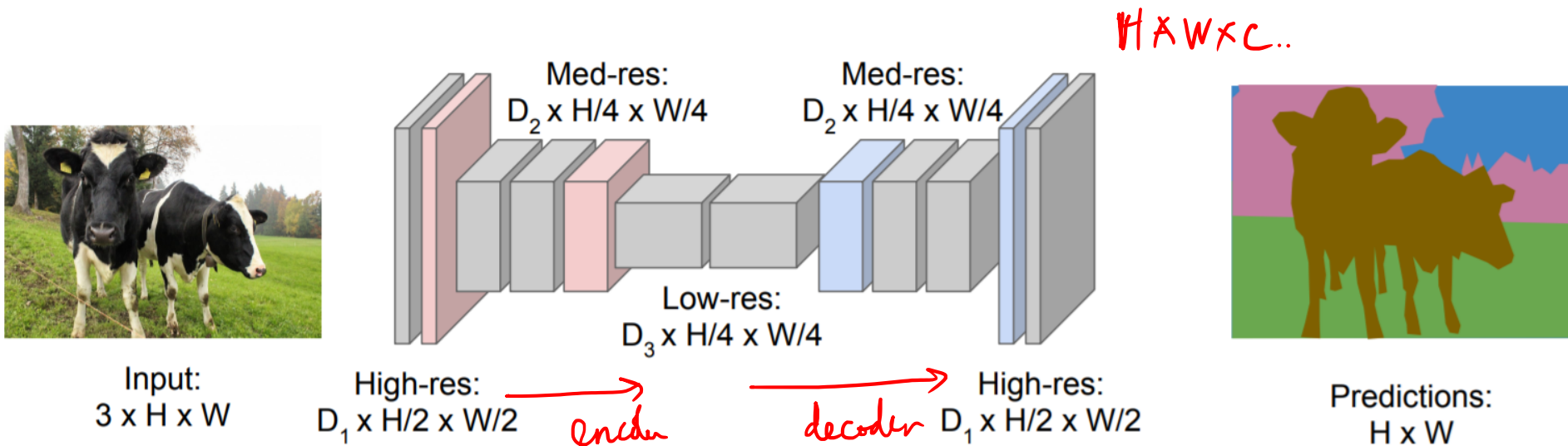


loss: Categorical-Crossentropy
tính pixel



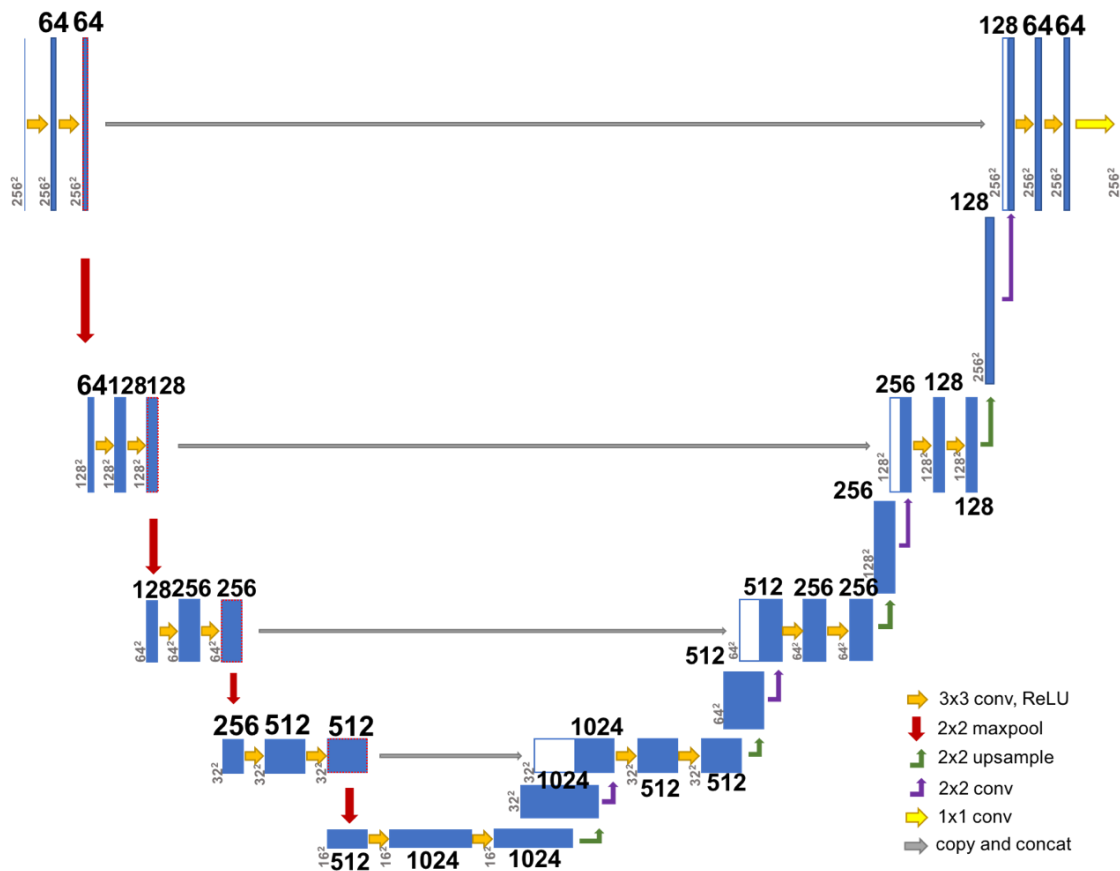
Predictions:
 $H \times W$

Fully convolutional improvement



Unet

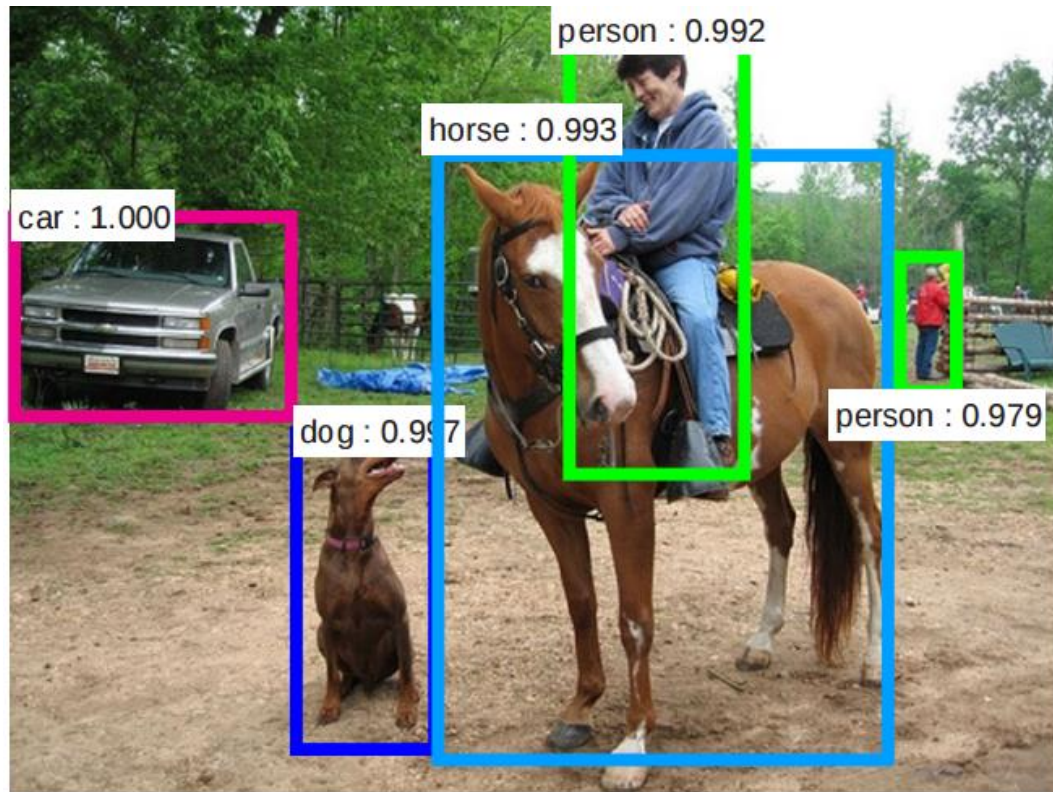
dersnet.



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



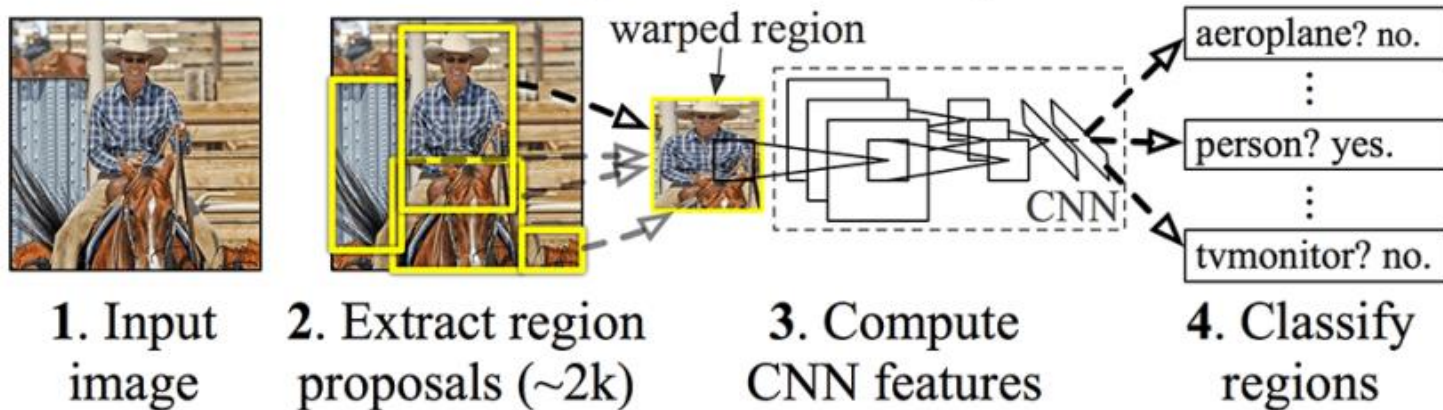
Input Image



Output Image

R-CNN

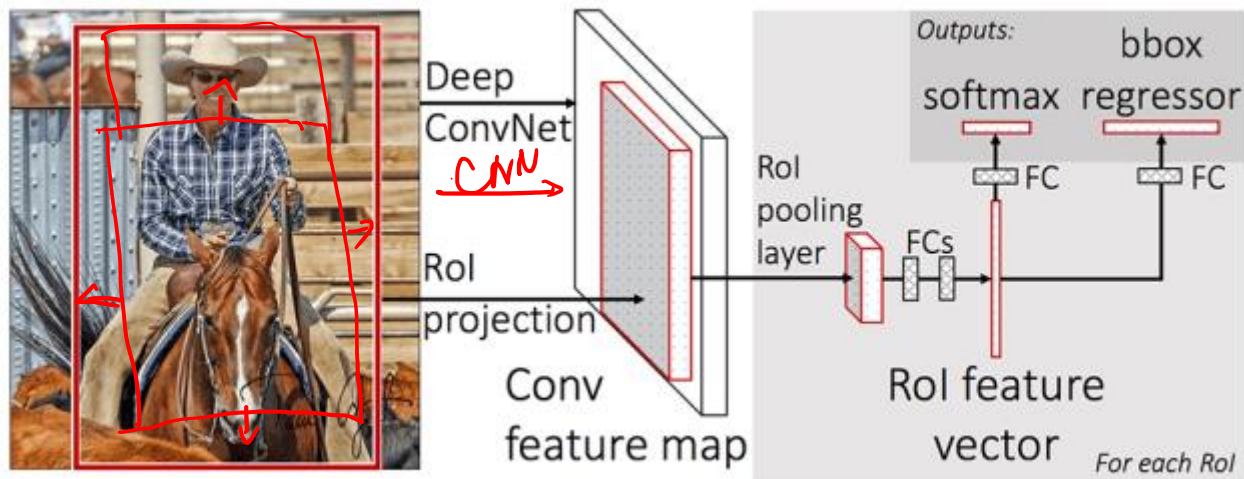
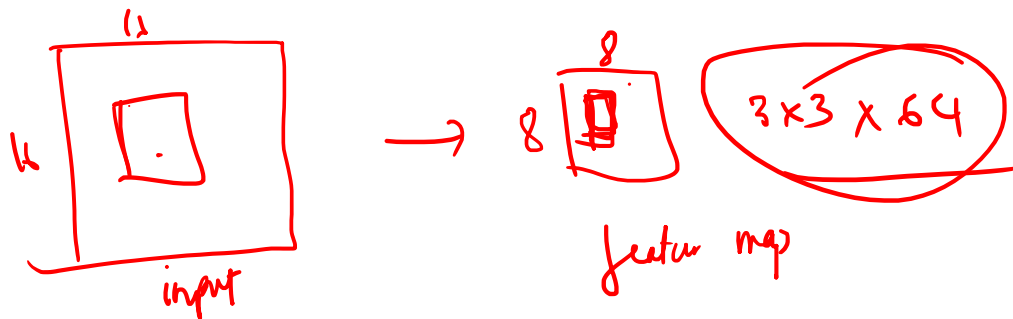
R-CNN: *Regions with CNN features*



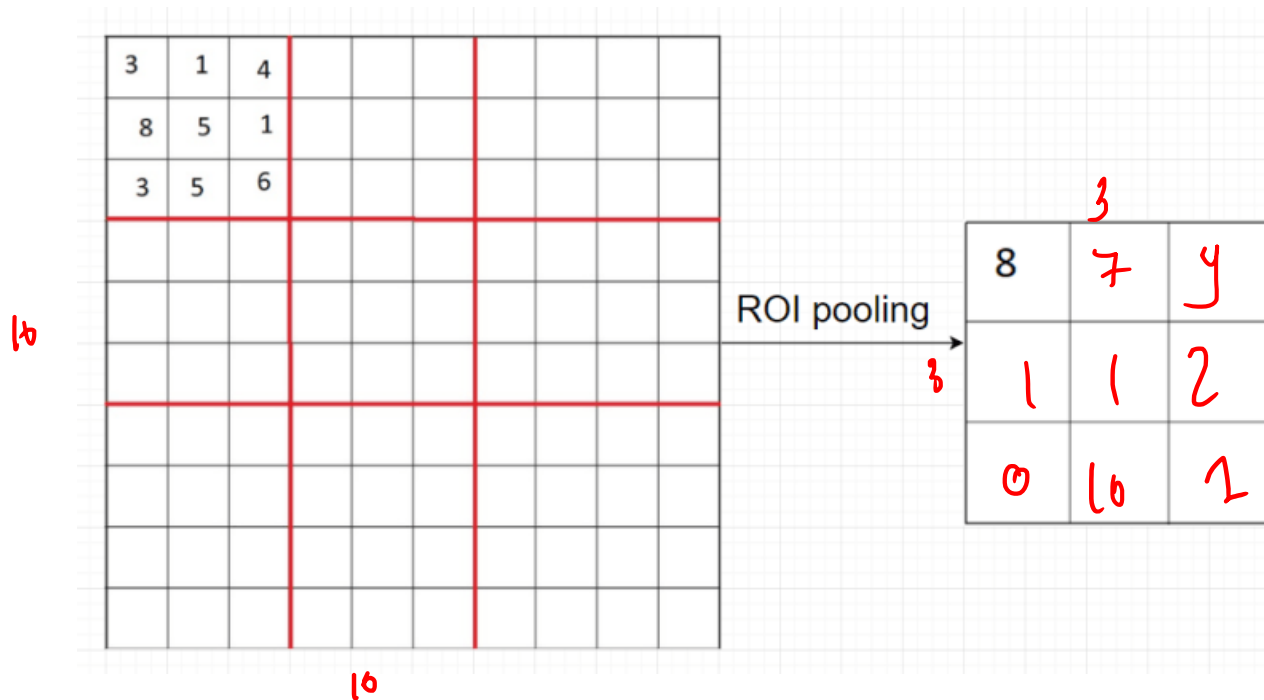
Problem?

- 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.

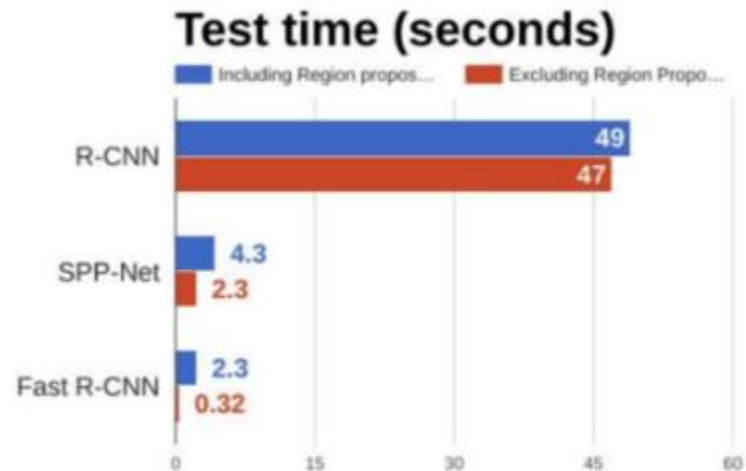
Fast R-CNN



Region of Interest (ROI) pooling

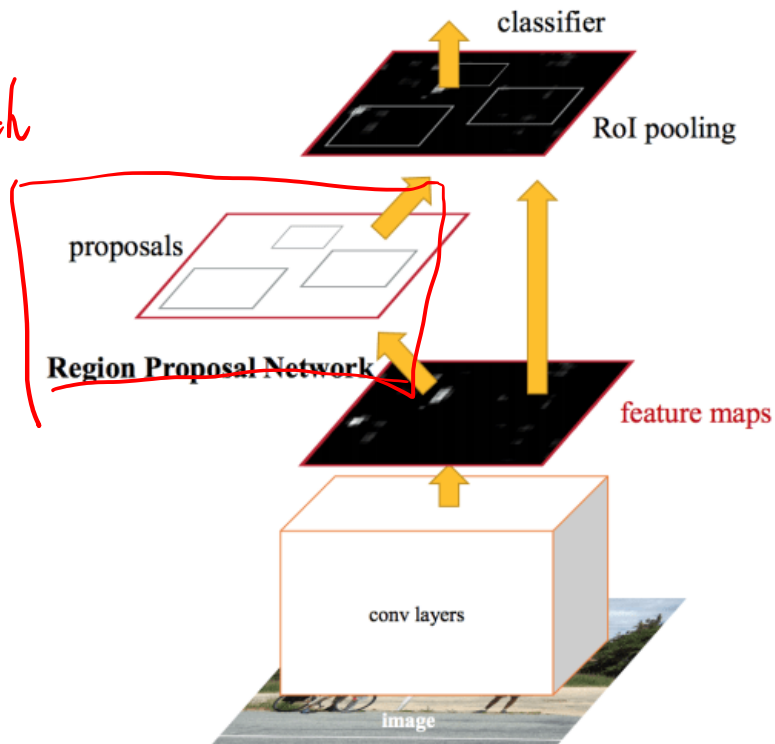


R-CNN vs Fast R-CNN



Faster R-CNN

k^{th} deg selective search



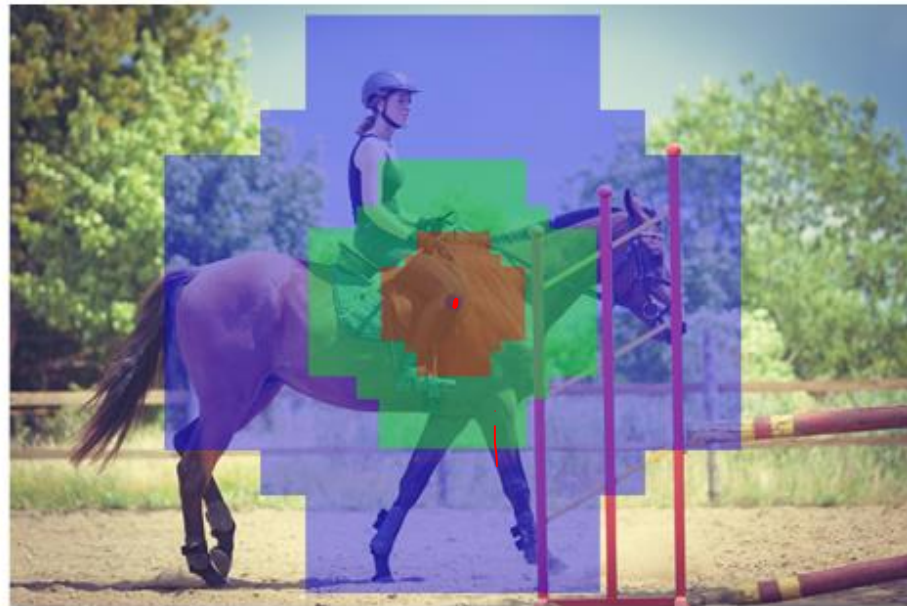
Anchors

mỗi train

3 tỷ dt : 100, 200, 300

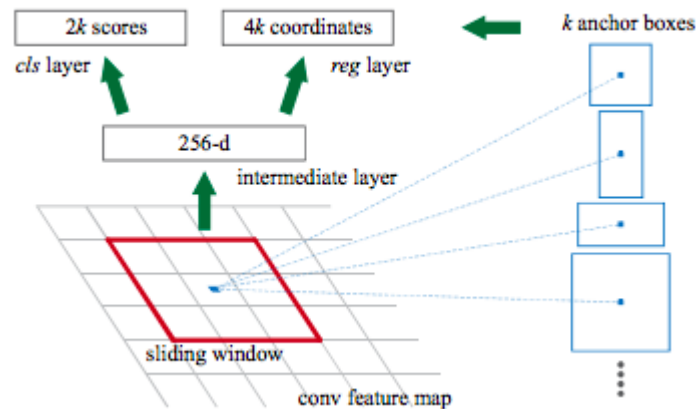
3 tỉ lệ : 1:1, 2:1, 1:2

30

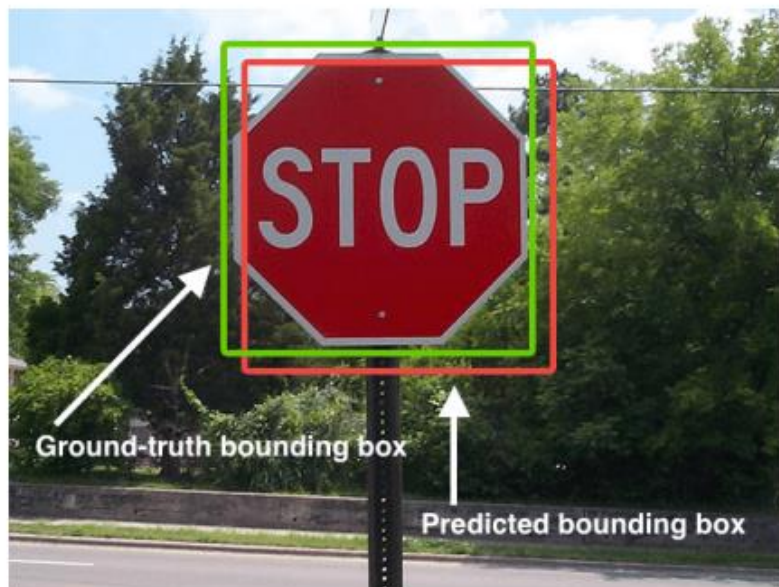


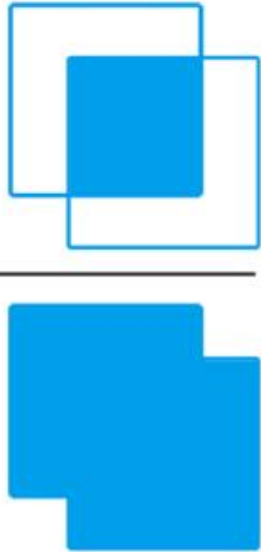
600 train → 5400 box

Region proposal network



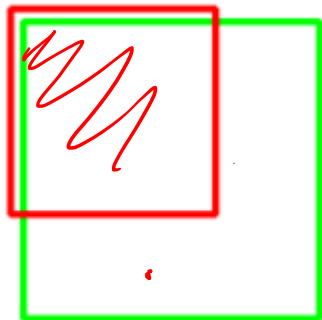
Intersection over Union (IoU)



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


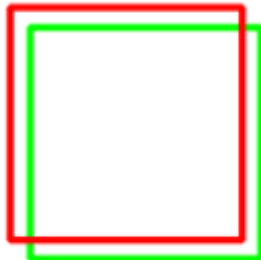
IoU

IoU: 0.4034



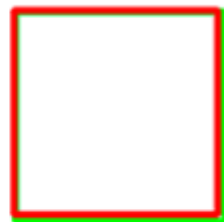
Poor

IoU: 0.7330



Good

IoU: 0.9264



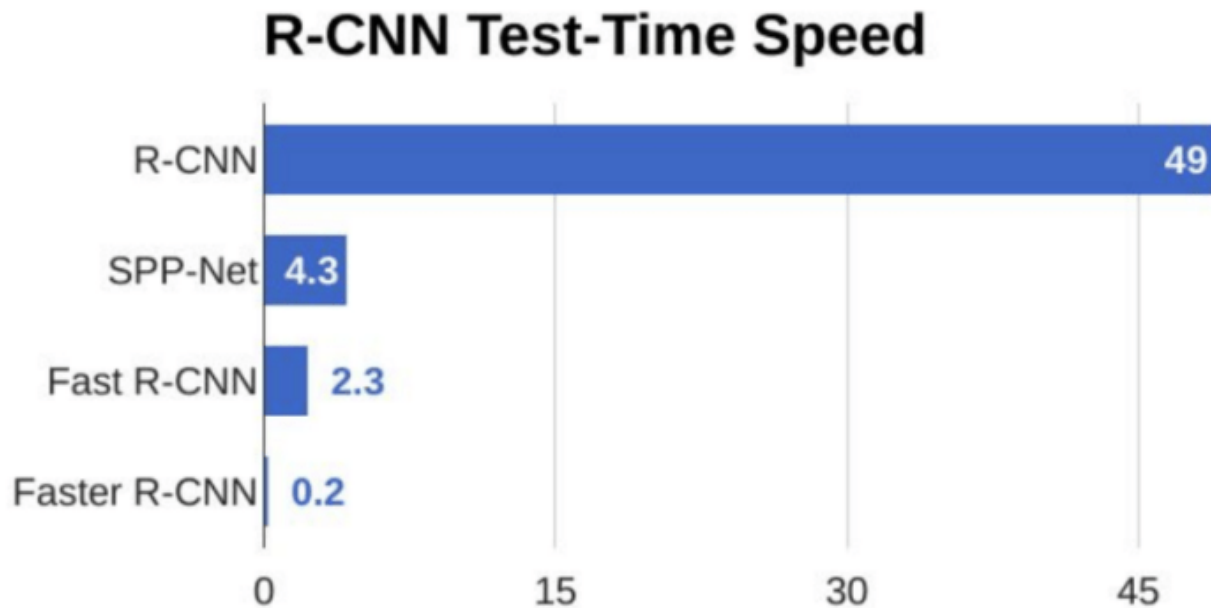
Excellent

output : 5400 anchors \rightarrow 100 anchors

Non-maxima suppression

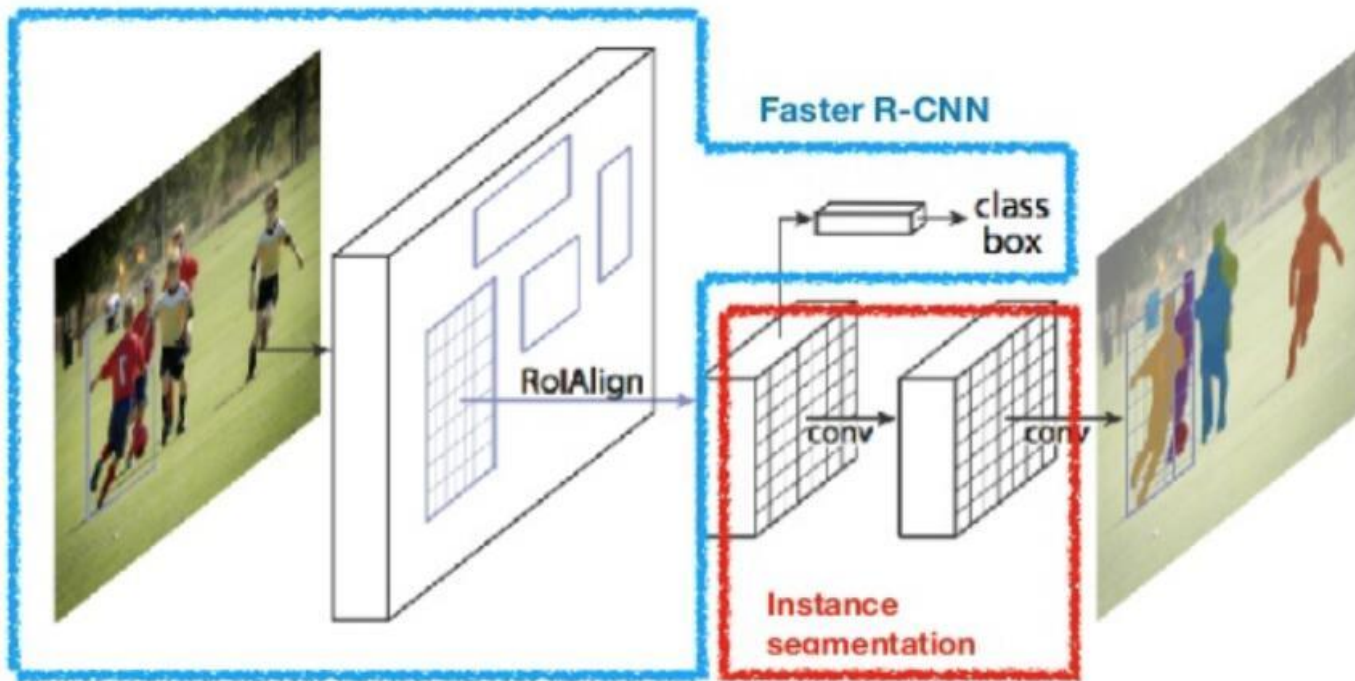
- Bước 1: Chọn ra anchor box (A) có xác suất là foreground lớn nhất trong tập Input
- Bước 2: Thêm A vào tập Output.
- Bước 3: Loại bỏ A và các anchor box trong tập Input mà có hệ số IoU với A lớn hơn 0.5 ra khỏi tập Input.
- Bước 4: Kiểm tra nếu tập Input rỗng hoặc tập Output đủ 100 anchor thì dừng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

Comparison



Comparison of test-time speed of object detection algorithms

Mask R-CNN



Q&A

