

I. Một số định nghĩa

1. Region Proposal

Khái niệm: Là các bounding box, là tập vị trí hình chữ nhật có thể chứa vật thể.

2. Thuật toán Selective search algorithm

- Input là ảnh màu.
- Output là khoảng 2000 region proposal (bounding box) mà có khả năng chứa đối tượng.

3. ROI pooling (Region of Interest Pooling)

- Là một dạng pooling layer. Với bất kỳ kích thước của tensor input, ROI pooling luôn cho ra output là một kích thước cố định
- Input đầu vào là $m \times n$, output là $h \times k$:
 - o Ta chia chiều rộng thành h phần, (h-1) có kích thước m/h , phần cuối có kích thước $m/h + m \% h$
 - o Tương tự như chiều dài
 - o Với mỗi ô ta thực hiện chọn giá trị max => đưa ra kích thước như mong muốn

4. Anchor.

- Là một box. Được xác định bằng 4 tham số x_center , y_center , width, height.
- Là các prediction box được định nghĩa trước khi huấn luyện.
-

5. RPN (Region Proposal Network)

- Input là feature map và output là các region proposal, và xác suất chứa vật thể tương ứng.
- Là một mạng fully convolution network nên không cần một kích thước đầu vào cố định.
- RPN sử dụng một Conv với 3×3 filters, 1 padding, 512 kernels. Đầu ra được chia làm 2 nhánh 1 binary object classification, 1 cho bounding box regression (với classification để dự đoán xem box đó có phải đối tượng hay không)
- RPN thực hiện các bước:
 - o Dự đoán xem anchor đây là chứa object hay không chứa đối tượng.
 - o Dự đoán 4 offset value cho x_center , y_center , width, height cho các anchor.

6. NMS (Non- maximum suppression)

- Do có rất nhiều anchor bị chồng lên nhau nên NMS được dùng để loại bỏ các anchor chồng lên nhau.
- Giả sử có 900 anchor box (tập Input) nhưng chỉ muốn giữ lại 100 anchor cho output làm region proposal ta sẽ làm:
 - o Chọn ra anchor box(A) có chứa foreground(chứa object) lớn nhất trong tập input.
 - o Thêm A vào tập output
 - o Loại bỏ A và các anchor box trong tập Input nếu có hệ số IOU với A > 0.5 ra khỏi tập Input.
 - o Kiểm tra nếu Input rỗng hoặc Output đủ 100 anchor thì dừng lại, nếu không qua lại bước 1.

7. IoU (Intersection over Union)

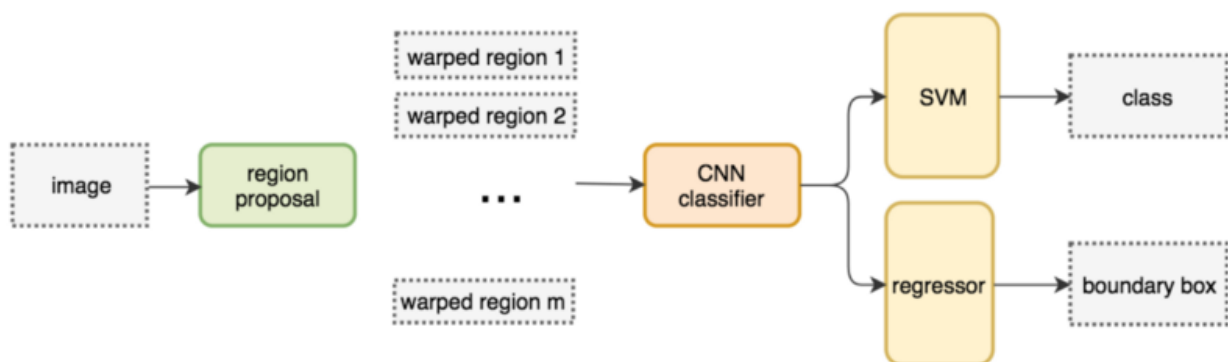
- Dùng cho bài toán object detection, để đánh giá xem bounding box dự đoán đối tượng khớp với ground truth thật của đối tượng.
- Có giá trị từ $[0, 1]$. Càng gần 1 thì bounding box gần ground truth.

II. RCNN (Region with CNN feature)

1. Ý tưởng thuật toán :

- Dùng Selective Search Algorithm để lấy ra khoảng 2000 bounding box trong input mà có khả năng chứa đối tượng.
- Với mỗi bounding box ta xác định xem nó có phải đối tượng không

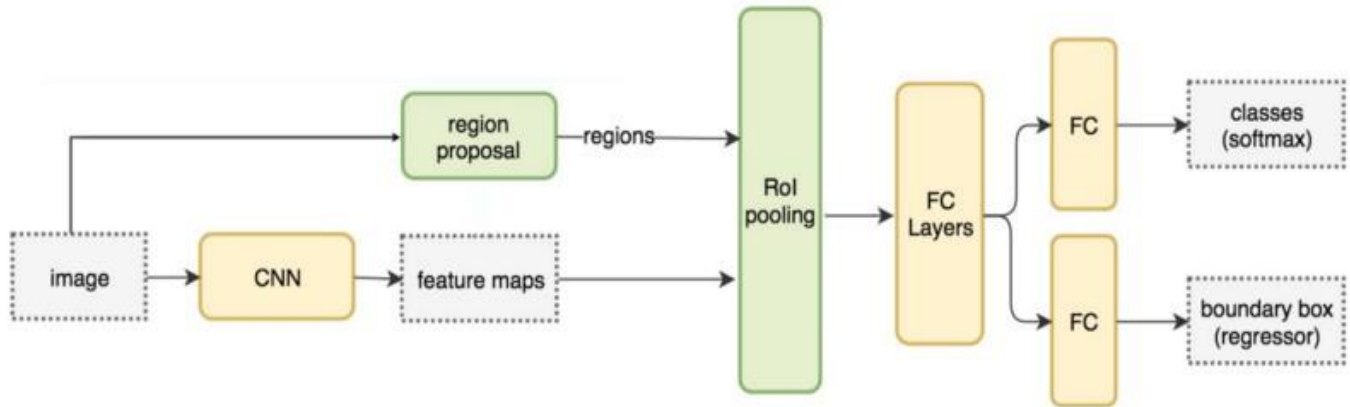
2. Các bước trong RCNN



III. Fast RCNN

- Tương tự như RCNN vẫn dùng thuật toán selective search để lấy ra các region proposal. Tuy nhiên Fast RCNN cho cả bức ảnh vào ConvNet (Conv+pooling) để tạo ra convolution feature map. Sau đó các vùng region proposal được lấy ra tương

ứng từ Conv feature map. Sau đó tiếp là Flatten và thêm 2 Fully Connected layer (FCs) để dự đoán lớp của Region proposal và giá trị xác suất của bounding box.



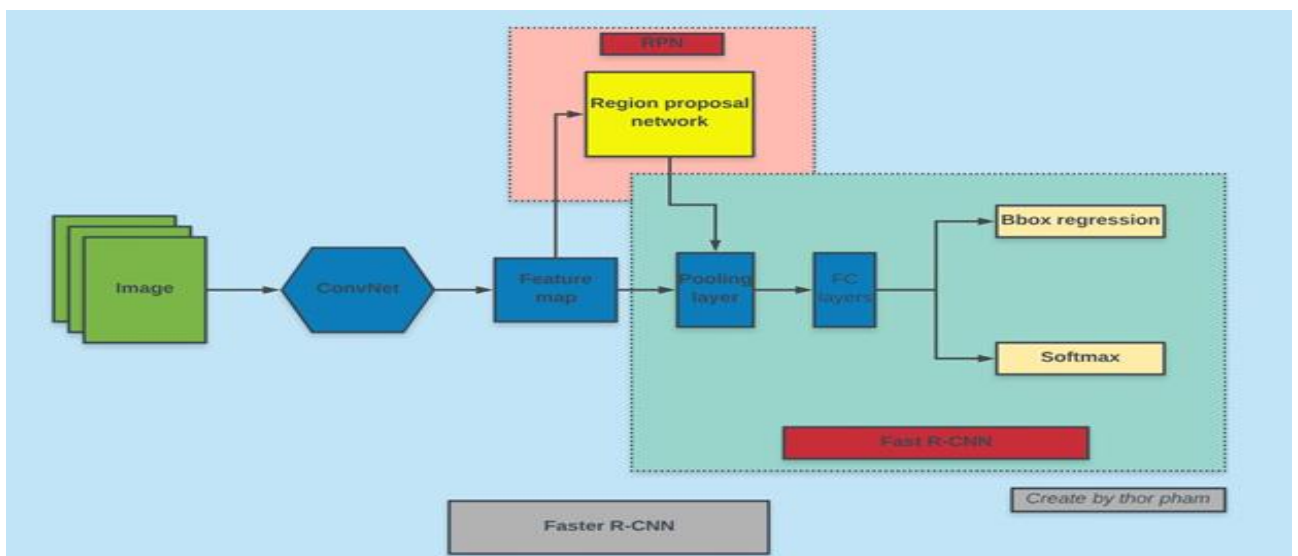
- Các bước thực hiện thuật toán:

Dùng pretrained model (VGG,...) để lấy feature map.

1. Sử dụng selective search để lấy region proposal .
2. Feature sẽ được đưa qua ROI pooling để fix size.
3. Cuối cùng là Fully Connect layer để classification và regression box.

IV. Faster RCNN

- Không sử dụng thuật toán selective search để lấy ra các region mà sử dụng một mạng CNN mới là Region Proposal Network (RPN) để tìm ra các region proposal. Phần còn lại tương tự như Fast RCNN.



- Ý tưởng thuật toán:
 - Trước tiên nó sẽ sử dụng ConvNet để trích xuất các đặc trưng từ hình ảnh đầu vào.
 - Sau đó sẽ được chuyển qua một Region Proposal Network (RPN) và sẽ trả về các bounding box tại các vùng có thể có đối tượng với các kích thước khác nhau.
 - Sau đó thêm vào lớp ROI pooling layer với mục đích gộp tất cả các bounding box trên cùng 1 đối tượng với các kích thước khác về cùng 1 kích cỡ.
 - Và cuối cùng được chuyển tới một fully connected layer để phân loại và đầu ra là 1 bounding box cho từng đối tượng.

V. Mask RCNN

- Là phần mở rộng của Faster RCNN.
- Khi đầu vào là một hình ảnh ngoài việc trả về label và bounding box của từng object thì sẽ có thêm các object mask.

