컴퓨터비전특론 9th week summary

[Lec10: Detection and Segmentation] 이어서

**Classification + Localization**

: Single object

Pretrained된 CNN 모델에 대해 전이학습을 거친 뒤 사물을 구분하는 classification을 위한 FC layer와 bounding box의 좌표를 구하는 regression을 위한 FC layer로 나눠져 Softmax loss와 L2 loss를 더한 multitask loss를 계산한다.

**Object Detection**

: Multiple objects: 여러 bounding box & 여러 물체를 구분

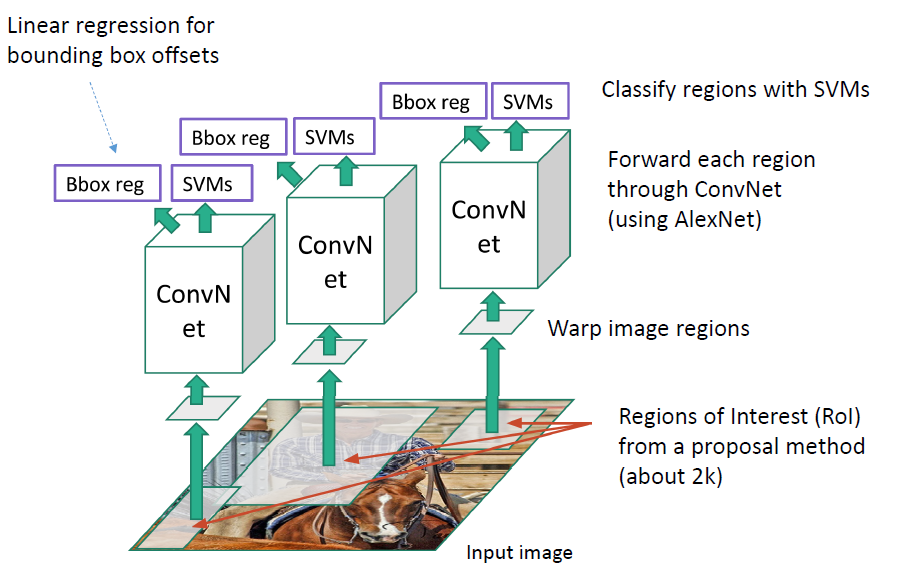
물체가 몇 개 있는지 알 수 없는 문제

🡪 여러 크기의 Window를 이미지 위에서 계속 슬라이딩 하면서 CNN을 적용

그러나 이 방법은 고려해야할 location과 scale이 너무 많으며 계산이 매우 복잡하다.

Region Proposal: Selective search를 통해 object를 포함할 가능성이 높은 image region을 찾음.

**R-CNN**



- Feature extraction

Positive samples: Region proposal과 overlap되는 부분이 0.5이상인 sample

Negative samples: Positive가 안인 region proposal의 나머지 영역

- Object category classifier: Alexnet의 final layer output에 SVM classifier를 적용

- Bounding box regression: AlexNet의 pool5 layer output을 이용하여 regression network

R-CNN Training

step 1) ImageNet으로부터 분류모델 학습(AlexNet)

step 2) Pretrained model의 마지막 FC layer를 Re-initialize하여 분류하고자 하는 이미지의 클래스 수에 맞게 fine tuning -> classification (positive / negative region 학습)

step 3) pool5 feature로부터 extract features

step 4) Region feature 분류를 위해 binary SVM 학습(학습하고자 하는 class vs 나머지 class)

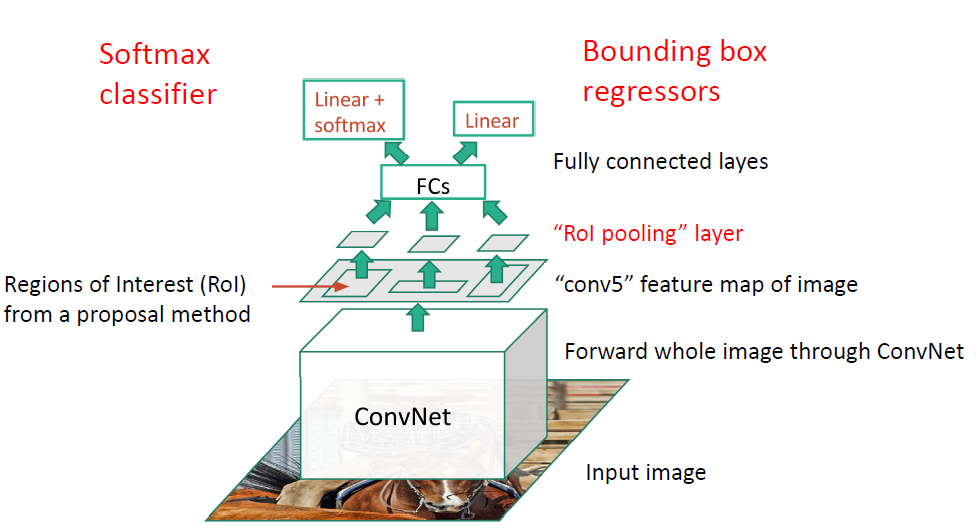
step 5) Bounding box regression: offset (dx, dy, dw, dh)

Problems

- object detection 문제에만 적용가능

- 학습 및 detection 속도가 ㄴ림

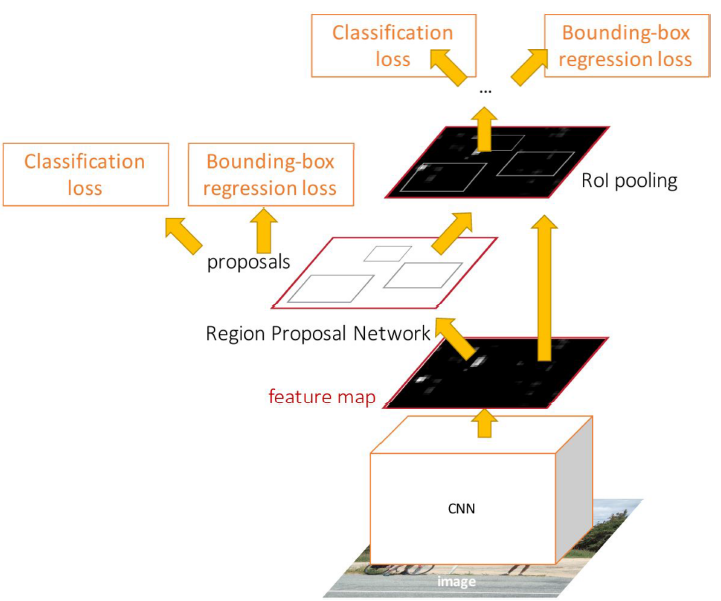
**Fast R-CNN**



- R-CNN과 달리 input 이미지에 바로 CNN을 적용한 후 region proposal 부분을 빼내며 end-to-end training으로 이루어진다.

- Rol Pooling layer: 임의의 사이즈인 projected region proposalㅣ을 7x7 grid로 resize하기 위해 pooling을 적용한다.

- R-CNN과 비교했을 때 계산시간이 훨씬 적다. 그러나 region proposal에 대한 계산시간이 region proposal을 제외한 계산 시간에 비해 7배 정도 커서 계산의 대부분을 차지한다는 문제점이 있다.

**Faster R-CNN**

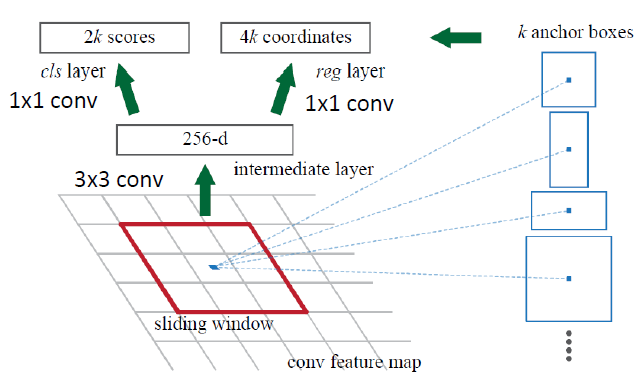
Feature를 이용하여 proposal을 예측하기 위해 Region Proposal Network(RPN)를 추가

🡪 RPN을 통한 classification/regression과 final classification/regression 과정이 진행됨: 4개의 loss를 통한 학습

- Region Proposal Network

R-CNN과 Fast R-CNN에서 selective search를 통해 region proposal을 찾던 과정을 딥러닝 네트워크로 해결

: 물체를 담고 있을 가능성이 있는 k의 anchor boxes를 만들어 놓고 모든 boxes에 대해 시도해본다.



**YOLO**

Big convolutional network를 이용하여 proposal 없이 detection을 진행하는 모델.

이미지를 7x7 grid로 나누고 각 grid cell을 중심으로 한 set of base boxes를 형성한다.

-> Regress final box with 5 numbers: (x, y, h, w, confidence)

이 때 confidence는 object인지 아닌지 판단하는 지표이다.

-> output: 7 x 7 x (5\*B+C), B: base 박스 개수, C: 클래스 개수