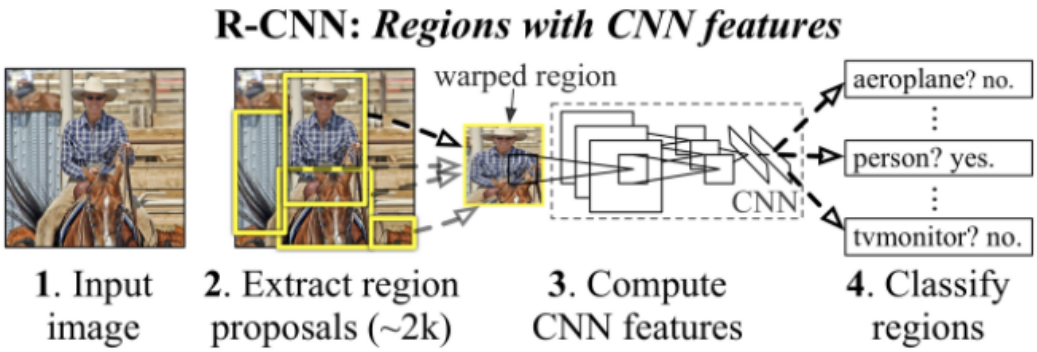
Detection

classification, localization 문제

이미지 내 관심 있는 객체의 위치(Region Of Interest, ROI)를 Bounding Box를 그려 localization

다수의 Bounding Box를 다양한 object로 classification

- **R-CNN**



  Multi-stage Detector

  object 위치 찾는 Region Proposal & object 분류하는 Region Classfication

  - **Region Proposal**

    object 위치 찾는 단계, Selective Search로 약 2천개 후보 영역 추출

    추출된 후보영역의 사이즈, 해당 사이즈의 이미지로 warping

  - **CNN**

    약 2천개의 warped image, 각각 CNN에 넣어 feature 추출

    fixed-length feature vector(4096-dimensional feature vector) 추출

  - **SVM(Support Vector Machine) & Bounding Box Regression**

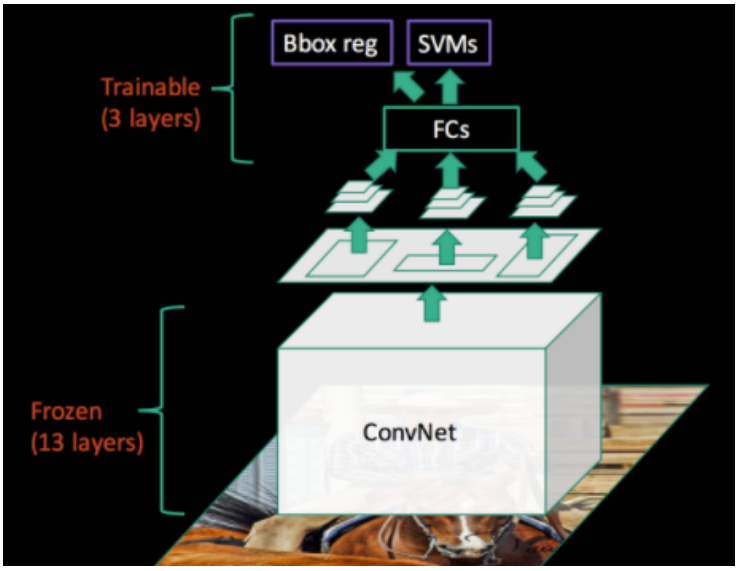
    추출된 fixed-length feature vector 활용해 object 분류

    - CNN fine-tunning을 위한 학습데이터가 시기상 많지 않음, Softmax 적용 => 오히려 성능 낮아짐 => SVM 사용

    BBox Regressor로 정확하지 않은 bounding box 위치 조정

  - 성능 뛰어남
  - But, 시간 오래 걸림, 복잡, End-to-end learning(종단간(처음부터 끝까지, 입력에서 출력까지 파이프라인 네트워크 없이 한 번에 처리) 기계학습) 불가능

- **SPPNet (Spatial Pyramid Pooling Network)**



- CNN 문제점
  - 2천번의 연산
  - 입력 사이즈 227×227 고정 => warped image 넣음 => image distortion 발생

R-CNN과 같지만, bounding box에 해당하는 텐서만 떼와서 학습

(CNN을 2천번이 아닌 1번 돌림)

- **Multi-stage Detector**

  구조 R-CNN와 다름

  But, 3가지 pipeline(CNN, SVM, BBox Regressor)으로 이루어진 Multi-stage detector은 같음

- **Reduce CNN Operation**

  입력 이미지에 대해 CNN 연산을 먼저 적용, feature map에 기반한 region propoal 과정 거침

  => **CNN 연산 1번**

- **Spatial Pyramid Pooling**

  Max pooling 연산

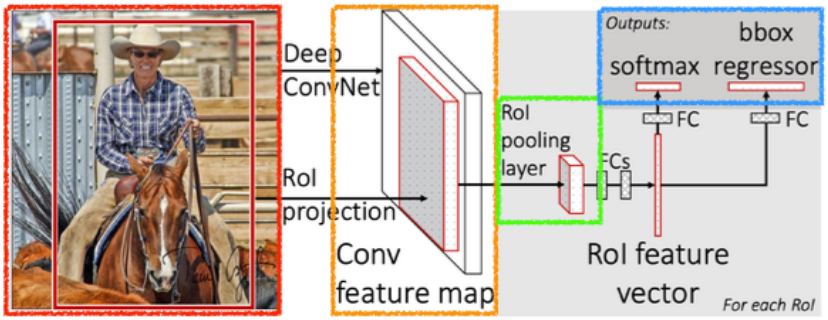
  ROI feature size 에 따라 kernel size 와 stride 만 설정 시, 다양한 resolution을 설정할 수 있음

  => warping 작업이 필요가 없음, 다양한 resolution 가짐

  => Spatial Pyramid Pooling

- R-CNN보다 소요시간 더 빠름
- But, 복잡, end-to-end learning 불가능

- **Fast R-CNN**



SPPNet과 거의 비슷하지만, 뒷단에 neural net (ROI feature vector 단)이 쓰임

- **ROI Pooling Layer**

  7×7 single level spatial bin을 사용하여, over-fitting 문제 피함

- **Softmax Classifier**

  SVM classifier가 아닌 Softmax classifier 이용

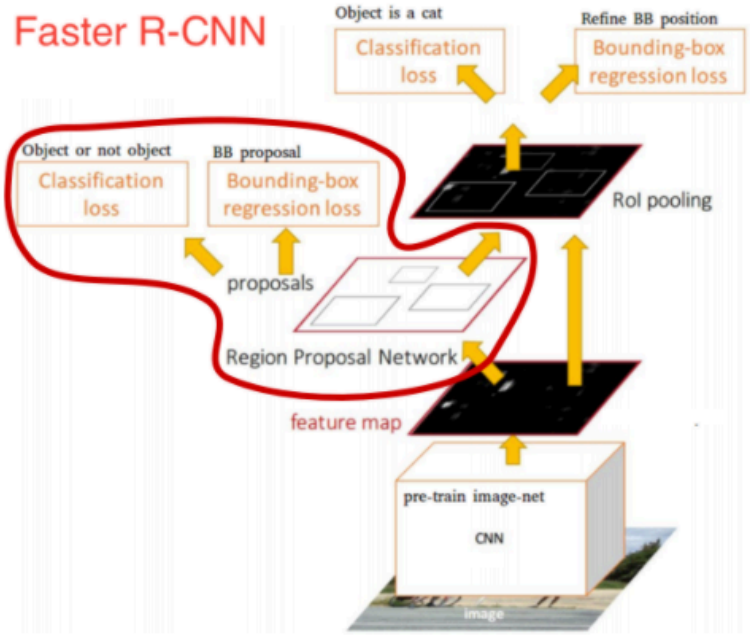
- **Multi-task loss**

  하나의 loss fucntion을 학습하여 End-to-End learning 가능

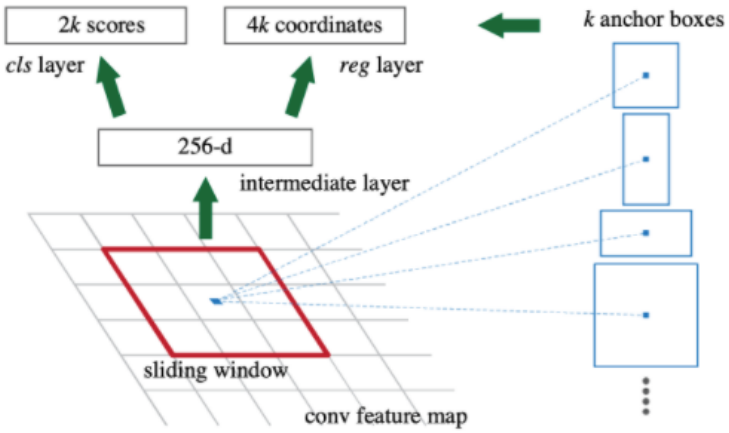
- **Truncated SVD**

  SVD(Single Value Decomposition)로 compression하여 파라미터 수 줄여, 소요시간 줄임

- **Faster R-CNN**



- region proposal, 즉 bounding box를 뽑는 것도 학습으로 해결
- R-CNN + RPN
- **RPN(Region proposal network)**



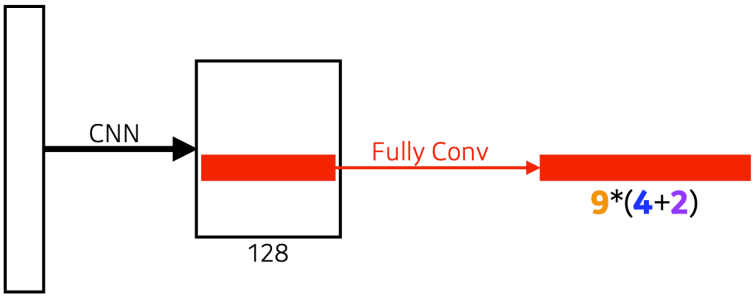
ZFNet 기준의 그림이어서 256-d 으로 표현되었지만 VGG-16의 경우 512-d의 feature map 출력

- **anchor box**

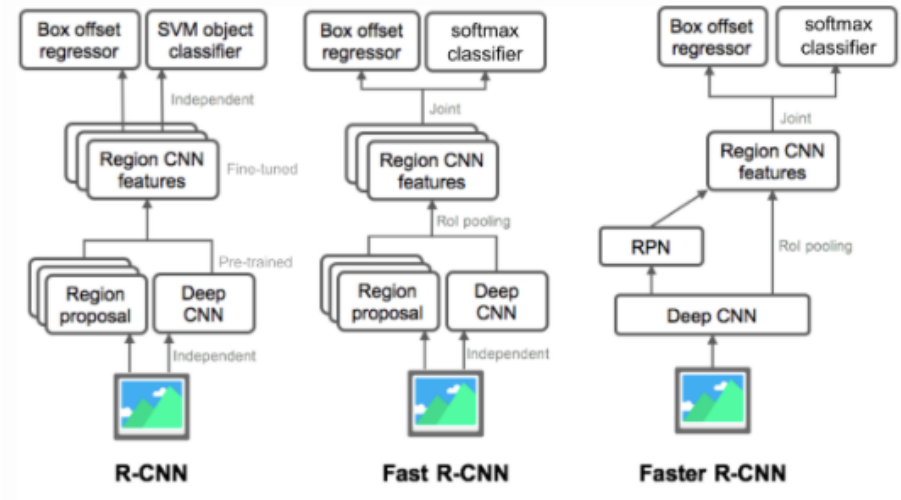
  미리 정해 둔 bounding box의 크기 (템플릿)

  detection box with predefined sizes

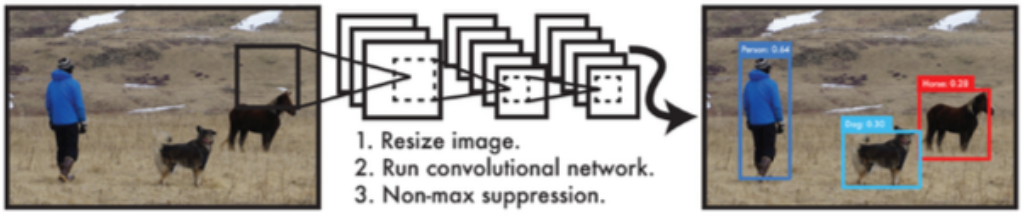
- **fully conv**

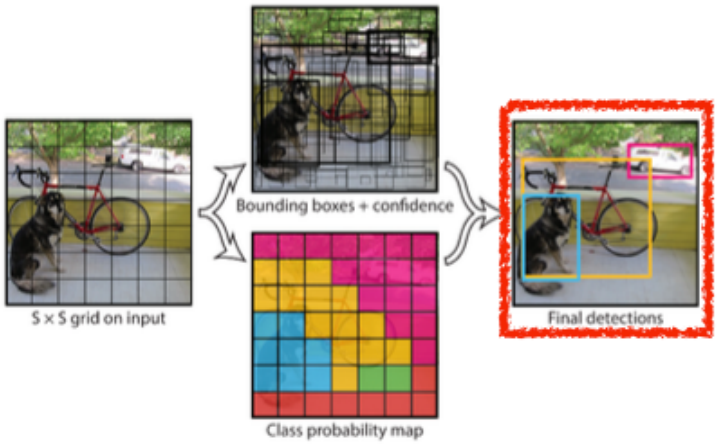


(3 region sizes * 3 ratios) * (4 bounding box params + 2 box classification(yes/no)) = 54



- **YOLO** (v1)





- You Only Look Once: 이미지 한 장에서 바로 output 출력
- **Faster R-CNN보다 훨씬 빠름**

  localization(바운딩 박스 찾는 것), classfication(클래스 찾는 것) 동시에 진행
- no explicit bounding box sampling
- output tensor

  (그리드의 셀 개수) * (바운딩박스 offset 5개 + 클래스 개수)