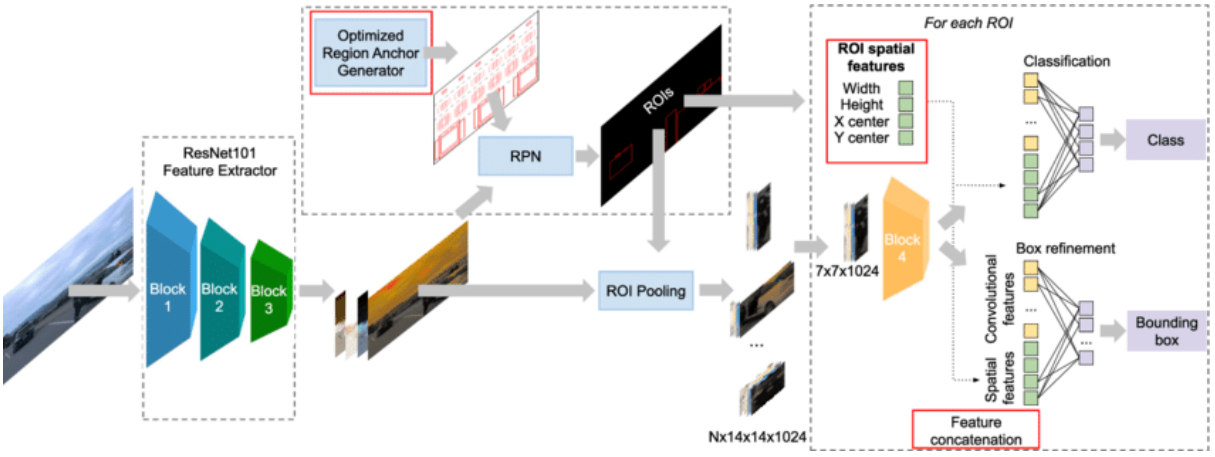
3. Faster R-CNN



Abstract

- 객체탐지 신경망은 객체의 위치를 가정하기 위해 region proposal 알고리즘에 의존함
- SPPnet, Fast R-CNN과 같은 발전은 region proposal 계산 때문에 일어나는 병목 현상에 노출되는 것을 줄여줌
- 이 연구에서는 탐지 네트워크와 Region Proposal Network(RPN)이 전체 이미지의 합성곱특성을 공유하며 region proposal에서의 비용을 줄여줌
- RPN은 고품질의 region proposal 생성을 위해 end - to - end 의 방식으로 훈련됨

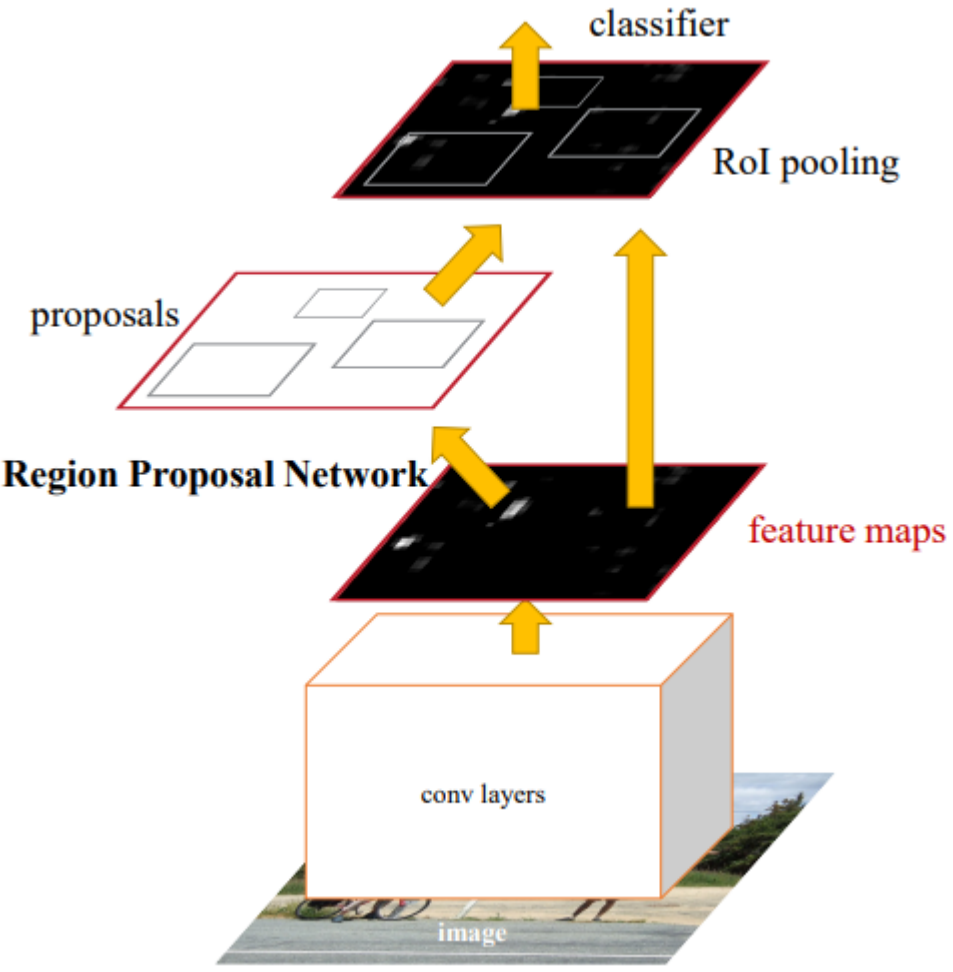
Introduce

- 객체탐지 기술은 region proposal과 region-based CNN 기술의 성공으로 발전
- region-based CNN은 계산비용이 높았지만, proposals 간의 합성곱을 공유함으로써 값이 감소함
- Fast R-CNN은 region proposals에 사용되는 시간을 무시하면, 매우 깊은 신경망을 사용해도 거의 실시간에 가까움
- region proposal은 일반적으로 저렴한 feature와 경제적인 추론기법에 의존
    - region proposal에 관한 연구는 CPU에서, fast region-based CNN은 GPU에서 구현됐기 때문에 동등한 비교 대상이 아님

Related Work

- Object Proposals

    - Object Proposal은 super-pixels의 그룹화, sliding windows등의 기술에 기반함
    - Object Proposal은 Detector와는 독립적인 외부모듈로 채택됨
- Deep Networks for Object Detection
    - R-CNN은 resion proposal을 객체의 카테고리 또는 배경으로 분류하기 위해 CNN을 end - to - end로 훈련시킴
    - R-CNN은 Object Bounds를 예측하는 것이 아닌 주로 분류기로써 작



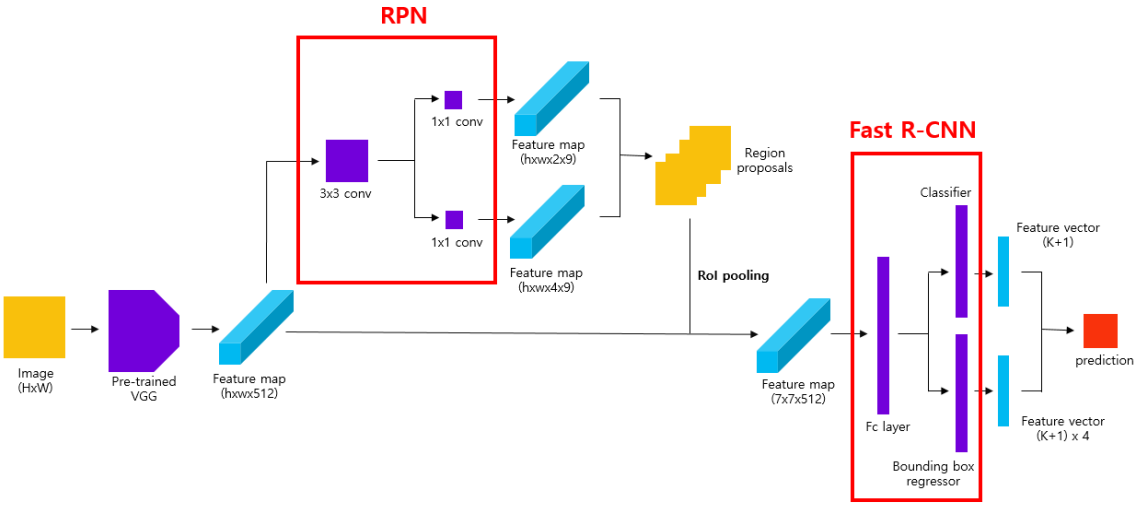
- Faster-Rcnn 흐름도
    - Conv Layers
        - 입력 이미지가 Convolutional Layers에 전달 → 여기서 이미지의 공간적 계층 구조를 캡처하기 위해 여러 개의 필터가 적용

        - 이 과정에서 이미지로부터 Feature Maps(특징 맵)이 생성 → 이 특징 맵은 이미지 내의 중요한 시각적 특징을 추출
    - Region Proposal Network(RPN)
        - 생성된 Feature Maps가 Region Proposal Network에 전달
        - RPN은 이미지에서 객체가 있을 가능성이 높은 영역들을 예측(즉, Proposals라 불리는 영역 후보)
        - 각 proposal은 해당 영역에 객체가 있을 가능성에 대한 점수와 위치를 나타냄 → 다양한 크기와 비율의 앵커 박스를 사용하여 객체의 위치와 크기를 예측
    - ROI Pooling
        - RPN에서 생성된 Proposals는 ROI Pooling층으로 전달
        - ROI Pooling은 제안된 각각의 영역을 동일한 크기의 출력으로 변환 → 다양한 크기의 영역들을 일관된 크기로 정규화하여, 이후의 단계에서 쉽게 다룰 수 있도록 함
        - 이 단계는 proposals를 고정된 크기의 Feature Maps로 변환하는 과정이라고 할 수 있음

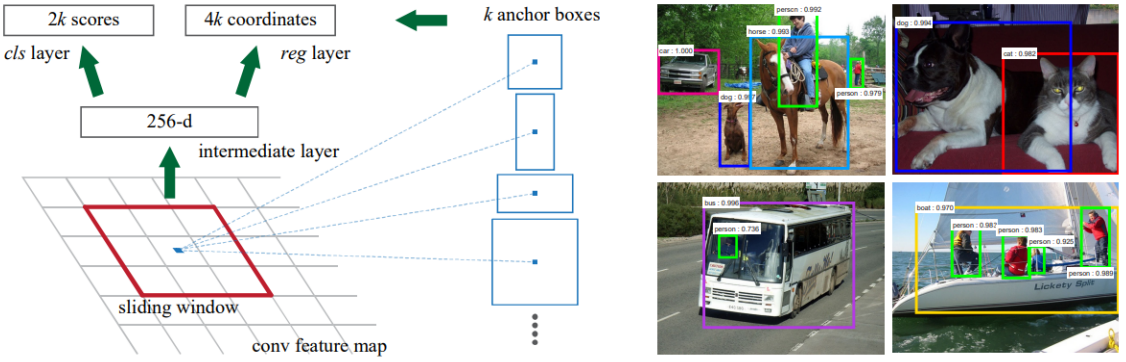
Faster R-CNN

- 2개의 모듈로 구성돼있음
    - region proposal을 진행하는 FCN
        - FCN
            - 전통적인 신경망 구조에서 완전연결층을 모두 제거하고, 오직 컨볼루션 레이어만으로 구성된 신경망을 의미 → 특히 이미지에서 픽셀 단위의 예측이 필요할 때 유용
            - FCN의 주요 특징
                - 완전연결층(Fully Connected Layers) 없음 : 일반적인 CNN에서는 최종적으로 이미지의 크기를 줄이고, 그 결과를 1차원 벡터로 변환하여 완전연결층에서 처리
                  → FCN에서는 이 과정이 없고 대신 모든 레이어가 컨볼루션으로 구성
                - 공간 정보 유지 : FCN은 입력 이미지의 공간 정보를 유지하며, 최종 출력도 입력과 동일한 공간적 해상도를 가짐
                  → FCN이 이미지의 모든 픽셀에 대해 예측을 수행할 수 있게 해줌

        - 가변 입력 크기 : FCN은 고정된 크기의 입력만 처리하는 것이 아니라, 다양한 크기의 입력 이미지를 처리할 수 있음
    - FCN을 이용하는 Fast R-CNN 탐지
        - RPN에서 FCN 사용
            - Feature Map 생성 : 입력 이미지는 먼저기존의 CNN에서 여러 층의 컨볼루션을 거치며 Feature Map으로 변환 → 이미지의 고수준 특징을 포함
            - Anchor Box 적용 : RPN은 각 위치에 대해 다양한 크기와 비율의 Anchor Boxes를 적용 → anchor boxes는 잠재적인 객체의 후보 영역을 나타냄
            - FCN을 통한 예측
                - 객체성 점수(Objectness Score) : 해당 anchor box가 객체를 포함하고 있을 확률을 예측
                - 바운딩 박스 조정(Boundig Box Regression) : anchor box의 위치를 조정하여 실제 객체의 위치에 더욱 가깝게 만듬
            - Region Proposal **생성** : FCN을 거친 후, RPN은 각 위치에서 객체가 있을 가능성이 높은 영역(Region Proposal)을 제안 → 이 제안된 영역들은 이후의 객체 탐지 단계에서 더 정확하게 분류되고, 바운딩 박스가 조정
        - 장점 및 중요성
            - 효율성 : FCN은 모든 픽셀에 대한 예측을 동시에 수행하므로 매우 효율적 → 이는 객체 탐지 모델이 더 빠르게 학습되고 예측할 수 있게 함
            - 일관된 Feature 사용 : RPN과 객체 탐지 네트워크는 동일한 Feature Map을 사용하여, 정보의 일관성을 유지하고 성능을 향상시킴
            - 가변 크기 처리 : FCN은 다양한 크기의 입력 이미지에 대해 유연하게 작동할 수 있어, 여러 상황에서 활용될 수 있음
- Region Proposal Networks

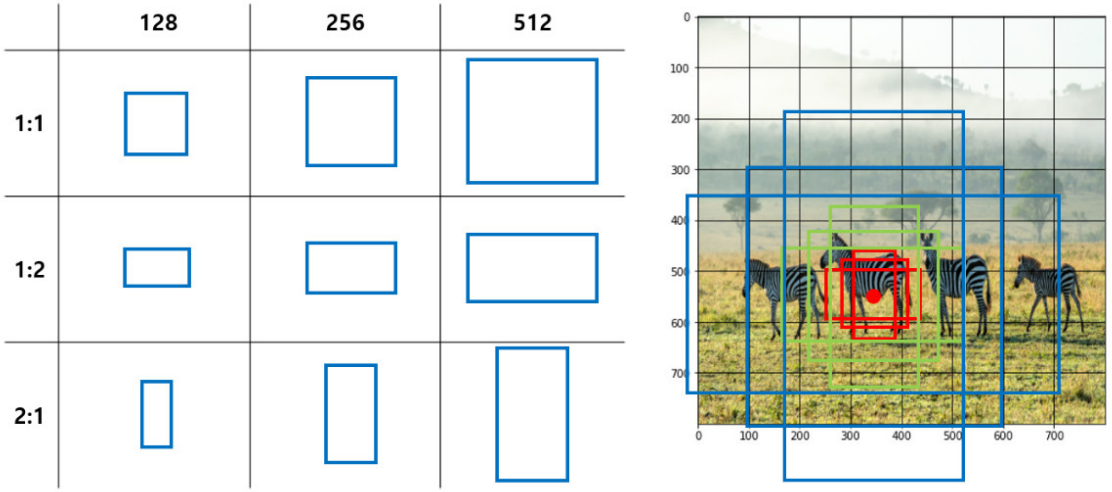


- RPN은 아무 크기의 이미지를 입력하면 각 객체에 대한 점수와 함께 사각형의 객체 proposal을 출력함 → 이 과정을 FCN을 이용
- region proposal을 생성하기위해 마지막 공유 합성곱층의 출력 위로 작은 신경망을 슬라이딩
- 이 작은 신경망은 입력된 합성곱 층의 n x n 크기의 창을 입력으로 받아들임
- 각각의 창은 저차원의 특성으로 맵핑(ZF모델은 256차원, VGG모델은 512차원) 이 특성들은 2개의 비슷한 전역합성곱 계층으로 입력
  - box-regression 계층(reg)
  - box-classification 계층(cls)



- 작은 신경망은 슬라이딩 윈도우 방식으로 작동하기 때문에 완전연결계층은 모든 위치에서 공유됨
- 이 구조는 자연스럽게 n x n 합성곱층에 이어 두 개의 형제 1 x 1 합성곱층(각각 reg와 cls)으로 구현

- Anchors



- 각 슬라이딩 윈도우는 동시에 여러 region proposal을 하며 각위치에서의 최대 proposal은 k로 표시
- eg는 k개의 상자의 좌표를 표시할 4k의 출력을, cls는 각 proposal에 대해 객체인지 아닌지의 확률을 추정하는 2k 점수를 출력
- k개의 proposal은 k 참조 박스를 기준으로 매개변수화 되며 이를 우리는 앵커(anchor)라고 부름
- 앵커는 슬라이딩 윈도우의 중앙에 위치하며 크기와 가로세로 비율과 연관되어 있음
- 3개의 크기와 3개의 비율을 디폴트로 사용하였고 각 슬라이딩 위치에서 k=9로 지정
  - W x H 크기의 합성곱맵에서 WHk개의 앵커가 나옴

Translation-Invariant Anchors

- 중요한 특성은 앵커와 앵커에 관련된 proposal을 계산하는 함수의 측면에서 모두 '이동 불변'이라는 것
- 이미지에서 객체가 이동할 경우 proposal은 이동되어야 하며 이는 같은 방법으로 각 위치에서 proposal을 예측할 수 있어야 함
  - **이동 불변성**이란 이미지 내 객체의 위치가 변경되더라도, 모델이 제안하는 **Region Proposal**이 동일하게 변경될 수 있는 특성을 의미
    → 객체가 이미지 내에서 이동을하면, 그에 맞게 proposal(제안된 영역)도 적절히 이동해야 함
    → 이동 불변성 덕분에 모델은 이미지의 어느 위치에서나 일관된 방식으로 proposal을 예측할 수 있고, 모델이 다양한 위치에서 객체를 인식하는 데 있어 매우 중요한 특성

Multi-Scale Anchors as Regression References

- 앵커 피라미드
  - 다양한 크기와 비율의 앵커박스를 참조하여 경계박스를 분류
  - 단일 비율의 이미지와 특성맵에만 의존하던 단일 크기의 필터(슬라이딩 윈도우)를 사용
  - 다중 크기 기반의 앵커 덕분에 Fast R-CNN에 의해 수행되는 것처럼 단일 이미지에서 계산된 합성곱 특성을 쉽게사용할 수 있음
  - 앵커는 추가적인 비용 없이 특성을 공유하기 위한 중요 요소
- Loss Function
  - RPN을 훈련하기 위해 각 앵커에 이진 클래스 레이블을 할당
    - positive 레이블에 2 종류의 앵커를 할당
      - 실측 박스와 가장 높은 IoU를 기록한 앵커
      - 어떤 박스든 0.7이상의 IoU가 나오는 앵커
    - 하나의 실측 박스는 여러 앵커에 positive 레이블을 지정할 수 있음
    - 일반적으로 두 번째 조건은 positive 샘플을 결정하기에 충분하지만, 일부 드문 경우에서 두 번째 조건은 양성 샘플을 찾을 수 없다는 이유로 여전히 첫 번째 조건을 채택
    - 모든 실측박스에 대해 IoU비율이 0.3보다 낮으면 non-positive 앵커에 negative 레이블을 할당
    - Positive 또는 Negative를 모두 갖지 않는 앵커는 훈련객체에 포함되지 않음
- Training RPNs
  - RPN은 역전파와 확률적 경사하강법을 이용하여 end - to - end 방식으로 훈련가능
  - 모든 앵커의 손실함수에 대해 최적화할 수 있지만, 이것은 negative가 많을때 이로 편향될 것
  - 미니 배치의 손실함수를 계산하기 위해 이미지에서 256개의 앵커를 무작위로 샘플링
    - 여기서 샘플링된 양의 앵커와 음의 앵커 비율은 최대 1:1
    - 이미지에 양성 샘플이 128개 미만인 경우 미니 배치를 음의 샘플로 패딩