


Object Detection

- Classification과 bounding box를 동시에 추정하는 문제
- 특정 객체들을 bounding box 형태로 위치를 측정하고 해당 박스 안의 객체의 Category까지 측정하는 기술

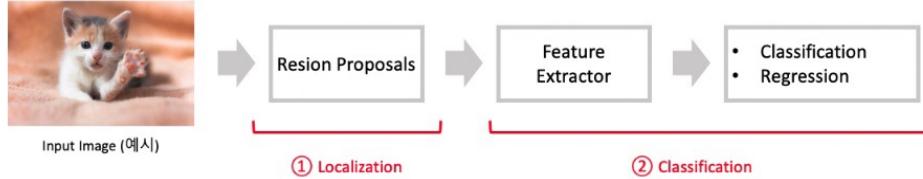
1-Stage 방식

- Localization 문제와 Classification 문제를 한 번에 해결



Input Image (예시)

- 2-Stage 방식보다 속도는 빠르지만 정확도는 낮다는 특징을 가짐 (대표적으로 YOLO 계열)



① Localization

② Classification

- 물체의 위치를 찾는 문제와 이미지가 무엇인지 분류하는 문제를 순차적으로 해결

Image가 주어졌을 때, 물체가 있을 법한 위치들을 찾은 후 각각의 위치에 대해 Feature를 추출하여 이를 토대로 Class를 분류한다.

(위치 정보를 정확히 조절하는 Regression 단계도 Classification과 함께 가짐.)

- 대표적으로 R-CNN 계열의 알고리즘이 있다. (R-CNN, Fast R-CNN, Faster R-CNN 등)

Detection 용어정리

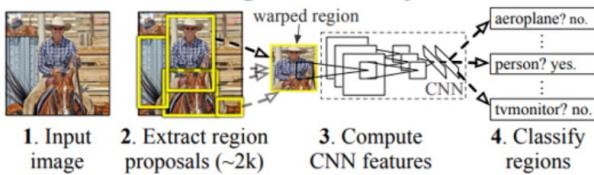
- Classification : 이미지가 주어졌을 때, 그 물체가 무엇인지 맞추는 문제. 예전에 사용되는 CNN 아키텍처들이 object detection이나 Segmentation과 Backborn으로 사용된다.
- Classification + Localization : Object의 위치를 찾고, 무엇인지까지 맞추는 문제
- Object Detection : 이미지에 한 개 이상의 object가 있을 때, 각각의 object에 대해 위치와 무엇인지까지 맞추는 문제
- Instance Segmentation : 이미지에 한 개 이상의 object가 있을 때, 각각의 object에 광활 단위의 위치와 무엇인지까지 맞추는 문제

★ Object Detection은 Box 형태의 위치를 찾고, Segmentation은 광활 단위의 위치를 찾는다는 차이가 있음.
- Bounding Box : 하나의 객체 전체를 포함하는 가장 작은 직사각형
- IoU (Intersection Over Union) : 실제값 (Ground Truth)과 모델이 예측한 값이 얼마나 겹쳐지는지를 나타내는 치표.
$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$
- Confidence Score : Object Detection 알고리즘에서 찾은 Bounding Box 안에 물체가 있을 확률 (물체가 없으면 0, 있으면 1)
- NMS (Non-Maximum Suppression) : 동일한 물체를 가리키는 여러 박스의 중복을 제거하기 위해 각 물체별 가장 좋은 Box 한 개만 남기고 나머지는 다 차움
- AP (Average Precision) : Confidence threshold에 따른 Precision과 Recall의 변화를 선 그래프로 그렸을 때, 선의 아래 면적
ROC-AUC와 비슷하다. MAP(mean AP)는 각 Class마다 AP를 계산하고 이를 평균화한 값.

체지 팀지 알고리즘 간단 설명

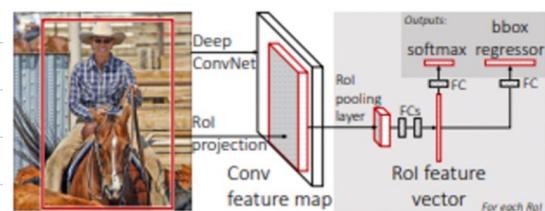
• R-CNN

R-CNN: Regions with CNN features



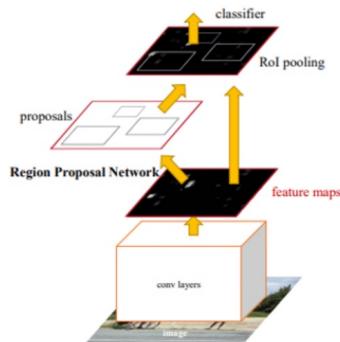
- Selective Search를 이용해 이미지에 대한 후보영역(Region Proposal)을 생성. 생성된 각 후보 영역을 고정된 크기로 Wrapping하여 CNN의 input으로 사용. CNN에서 과연 Feature map으로 SVM을 통해 분류, Regression을 통해 Bounding-Box를 조정함. 침체로 크기를 맞추기 위한 Wrapping으로 이미지의 변형이나 손실이 일어나고 후보 영역만큼 CNN을 돌아야하기 때문에 큰 저장공간을 요구하고 느리다는 단점을 지님.

• Fast R-CNN

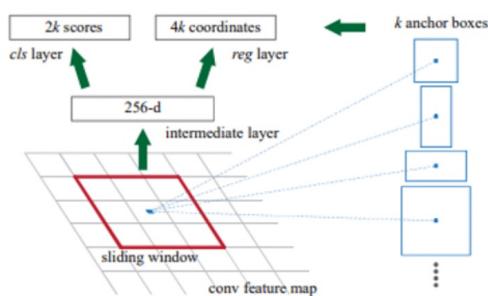


- 각 후보영역에 CNN을 적용하는 R-CNN과 달리 이미지 전체에 CNN을 적용하여 생성된 Feature map에서 후보 영역을 생성. 생성된 후보영역은 ROI Pooling을 통해 고정 사이즈의 Feature Vector로 추출함. Feature Vector에 FC layer를 거쳐 Soft max를 통해 분류, Regression을 통해 Bounding-Box를 조정

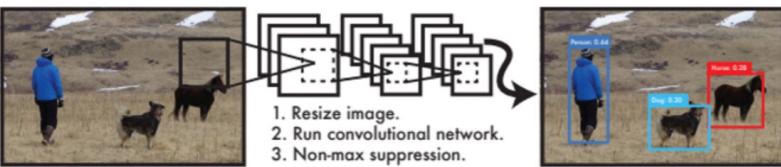
• Faster R-CNN



- Selective Search 부분을 딥러닝으로 바꾼 Region Proposal Network (RPN)을 사용. RPN은 Feature map에서 CNN 연관의 Sliding-Window과 같은 치열마다 Anchor-box로 후보영역 예측. Anchor-box란 이미지 자체에는 여러 개의 빙울과 크기의 Bounding-box. RPN에서 얻은 후보영역을 IoU 순으로 정렬하여 NMS 알고리즘을 통해 최종 후보 영역 선택. 선택된 후보영역의 크기를 맞추기 위해 ROI Pooling을 거치고 이후 Fast R-CNN과 동일하게 진행.



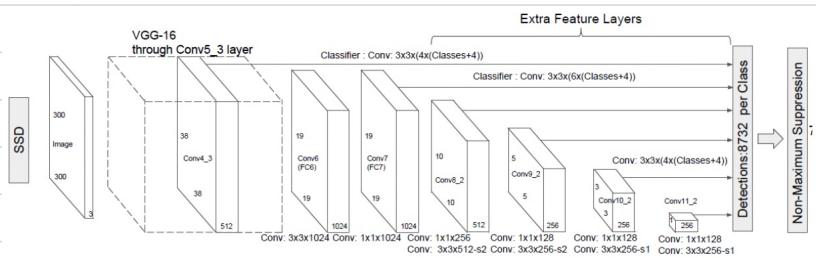
YOLO



· Bounding-box와 Class Probability를 하나의 풀체로 간주하여 개체의 종류와 위치를 한 번에 예측함.

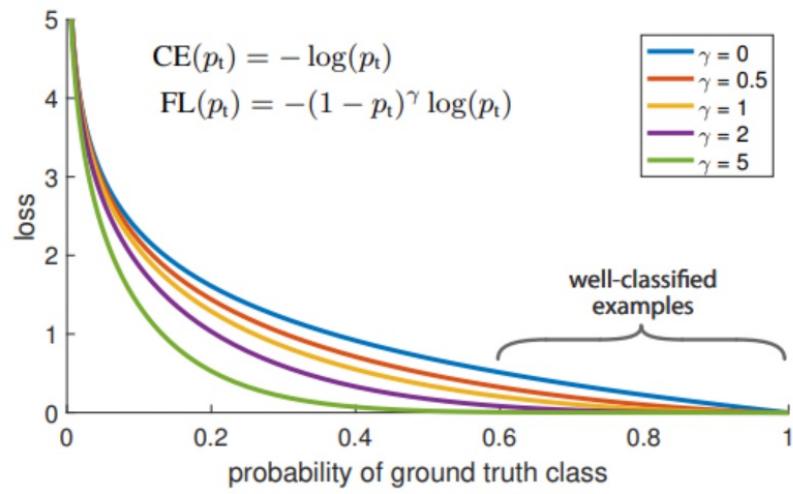
이미지를 일정 크기의 그리드로 나눠 각 그리드에 따른 Bounding-box를 예측함. Bounding-box의 Confidence Score와 그리드 속의 Class Score의 값으로 학습함. 간단한 처리 과정으로 속도가 매우 빠르지만 작은 개체에 대해서는 성능이 낮음.

SSD



· 각 Convolution Layer 이후에 나오는 Feature Map마다 Bounding-box의 Class 점수와 Offset (위치좌표)를 구하고, NMS 알고리즘을 통해 최종 Bounding-box를 결정. 이는 각 Feature map마다 스케일이 다르기 때문에 작은 물체와 큰 물체를 모두 탐지할 수 있다는 장점을 가짐

Retina Net



· 모델 학습 시 계산하는 손실함수에 변화를 주어 기존 One-Stage Detector들이 가지 못한 성능을 개선함.

One-Stage Detector는 맵마다 상관관계까지의 후보군 차리를 통해 확률을 전파함. 그 중 실제 개체인 것은 일반적으로 10배 이하이며, 나머지 후보군이 background 클래스로 간주됨.

상대적으로 분류하기 쉬운 background 후보군에 따른 loss 값을 줄여줌으로써 분류하기 어려운 실제 개체들의 loss 비중을 높이고, 그에 따라 실제 개체들에 대한 학습에 집중하게 함.

Retina Net은 속도가 빠르면서 Two-Stage Detector와 유사한 성능을 보임

One-Stage Detection과 Two-Stage Detection의 특징

• One-Stage Detection

- YOLO : 이미지를 일정 크기의 그리드로 나눠 각 그리드에 대한 Bounding-Box를 예측
간접적 차례방법으로 속도가 매우 빠르지만 작은 객체에 대해서는 상대적으로 정확도가 낮음
- SSD : NMS 알고리즘을 통해 최종 Bounding-Box를 결정
Feature map마다 스케일이 다르기 때문에 작은 물체의 큰 물체를 모두 탐지할 수 있다는 단점.
- Retina Net : 손실 함수에 변화를 주어 기존 One-Stage detector들이 차단 낮은 성능 개선
속도가 빠르면서 Two-Stage Detector와 유사한 성능

• Two-Stage Detection

- R-CNN : 이미지에서 객체 후보군 생성
Wrapping으로 이미지의 ~~부록~~이나 손질이 발생하며 큰 계층 공간이 필요하고 느리다.
- Fast R-CNN : Feature map에서 후보영역을 생성
- Faster R-CNN : RPN을 사용하여 Feature map에서 후보영역을 생성