

섹션 3. YOLO Overview

1. YOLO 모델 리뷰

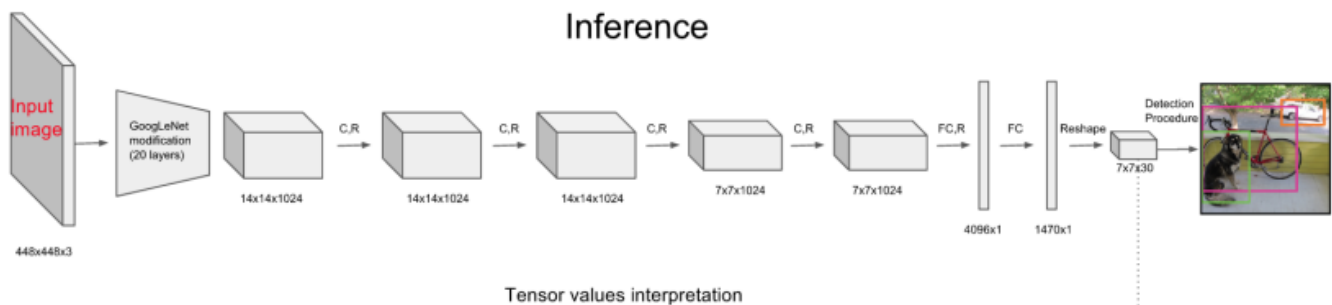
YOLO(You Only Look Once)

- YOLO 모델의 기본 컨셉 : 이미지를 $S \times S$ Grid Cell로 나누고 Grid Cell별로 B개의 Bounding Box를 예측
- 즉, 원본 이미지를 넣으면 고정 사이즈로 resizing 한 후, resizing한 이미지를 CNN모델에 넣은 후 도출되는 아웃풋이 $S \times S$ Grid Cell의 결과로 나온다.
- YOLO 모델의 최종 Output : $S \times S \times (5 * B + C)$ 이다. 여기서 S = grid cell , B =bounding box 개수, C =class 개수, 5 =bounding box 하나에 mapping 되는 정보($x, y, w, h, \text{confidence}$)
 - x, y : grid cell 내의 x, y 의 위치
 - w, h : 전체 이미지 대비의 bounding box의 width, high
 - confidence : 이미지 내에 object가 있을 것이라고 확신하는 정도



- YOLO version.1이 제안 될 당시에 정확도는 좀 떨어지지만 속도(초당 처리할 수 있는 이미지 양)면에서 아주 뛰어났다.

YOLO 동작 과정 (YOLO version.1)



24depth의 CNN과정을 거친 후, 마지막에 나온 convolution feature map에 대해서 fully connected layer로 4096 노드로 넘긴 후, reshape하여 최종 output 생성

2. Non-Maximum Suppression (NMS)

Non -Maximum Suppression

: YOLO의 경우 Object를 총 98개를 도출해준다. 여러 개의 Object가 감지 되었을 때, 가장 대표성을 띄는 하나의 Bounding Box만 남기는 것을 NMS라고 한다.

Non -Maximum Suppression 동작 과정

1. confidence ≤ 0.6 이하의 bounding box 제거
2. class별로 confidence가 가장 높은 bounding box가 앞으로 오도록 내림차순 정렬
3. 가장 confidence가 높은 bounding box와 나머지 bounding box를 비교해서 2개의 $\text{IoU} \geq 0.5$ 이면 confidence가 작은 bounding box를 제거
4. 제거되지 않은 bounding box중에서 confidence가 가장 높은 bounding box와 나머지 bounding box간에 3번 과정을 반복
5. 3~4 과정을 전체 bounding box에 대해서 진행
6. 2~5과정을 전체 class에 대해서 진행