YOLO
(You Only Look Once)

– Grid Cell · Bounding Box · IoU –

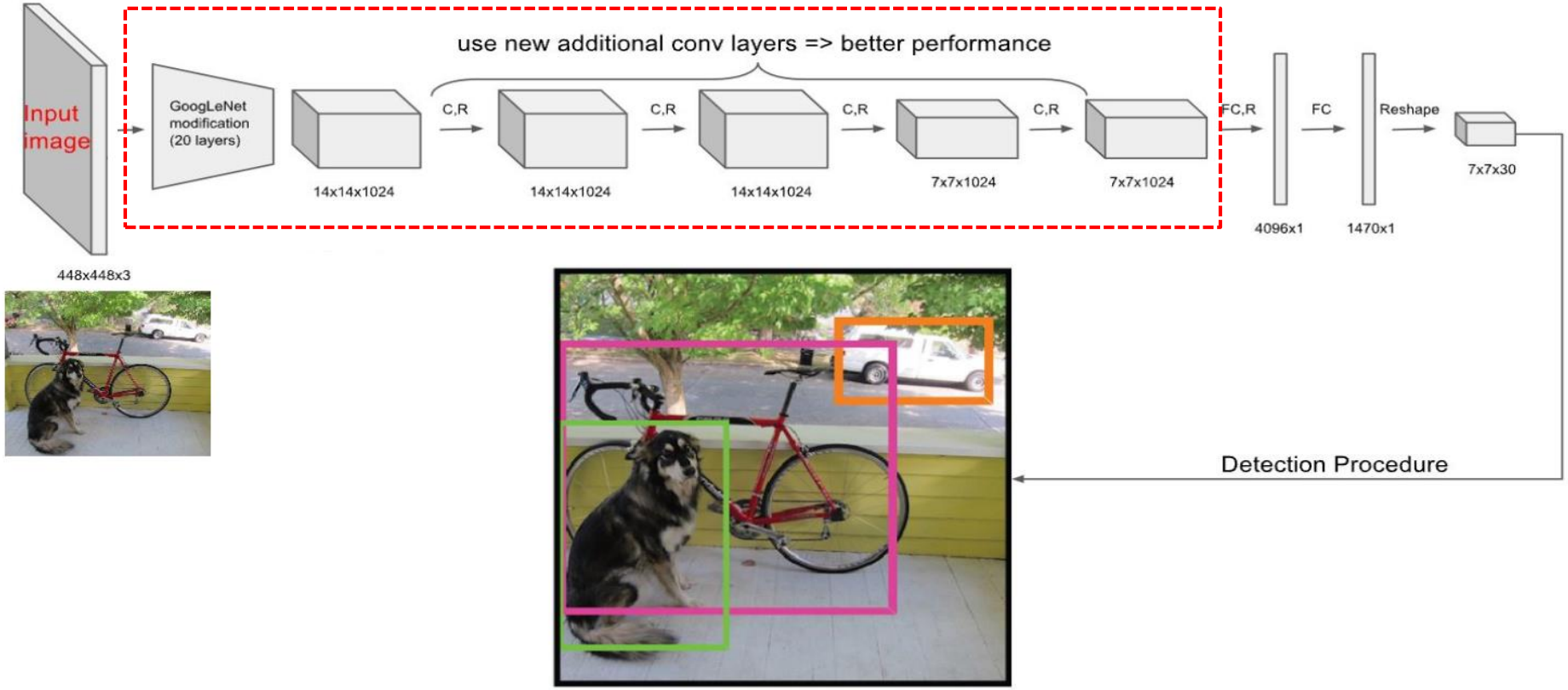
박성호 (neowizard2018@gmail.com)

YOLO (You Only Look Once)

- YOLO는 You Only Look Once 약어로서, 2015년 Joseph Redmon이 워싱턴 대학교에서 여러 친구들과 함께 YOLO 아키텍처를 논문과 함께 발표함

  - ✓ 당시만 해도 two-shot-detection 방식인 Faster R-CNN (Region with CNN)가 가장 좋은 성능을 내지만 실시간 성이 굉장히 부족함 (7 FPS 최대)

  - ✓ 이때 one-shot-detection 방식으로 동작하는 YOLO 가 등장하여 평균 45 FPS을 보여주었고 빠른 버전의 경우 최대 155 FPS을 기록하며 사람들을 놀라게 함.

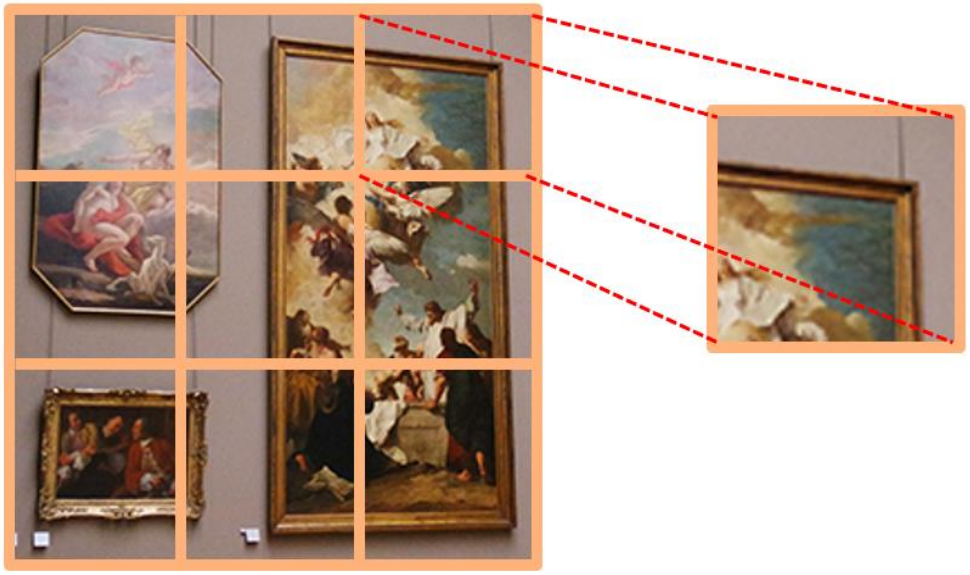


- YOLO 아키텍처는 feature extractor 로서 Google LeNet + 5 개의 new layer 를 사용하며, Flatten layer 를 거쳐서 최종 출력층에서는 7x7x30 텐서를 만들어내는 아키텍처이다. 즉 YOLO 아키텍처는 기본적으로 yolov5s.pt 같은 pre-trained model 로서 제공되며, 사용자 데이터 이용해서 Transfer Learning 에서 fine-tuning 하는 과정임

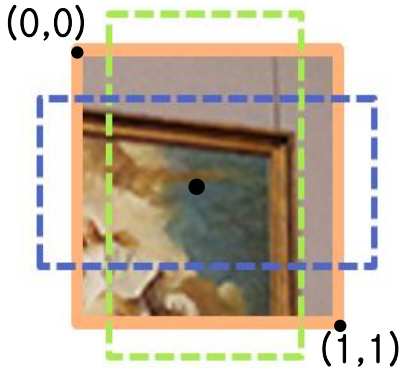
- YOLO는 입력 이미지에 대해서, Bounding Box와 classification 작업을 동시에 수행

✓ YOLO 아키텍처는, 먼저 입력 이미지를 S x S grid로 나눔.

(다음 이미지는 3 x 3 grid로 나누었으나, 실제 YOLO 논문에서는 7 x 7 grid로 분할함)



✓ YOLO 아키텍처를 학습할 경우, 각각의 그리드 셀은 B 개의 bounding box 를 가지고 있으며, 각각의 그리드 셀(grid cell)은 B개의 bounding box와 그 bounding box에 대한 confidence score를 가짐 . 즉 각각의 bounding box는 (x, y, w, h, confidence score) 같은 5개의 값으로 구성됨



(x, y) 는 bounding box의 중심점

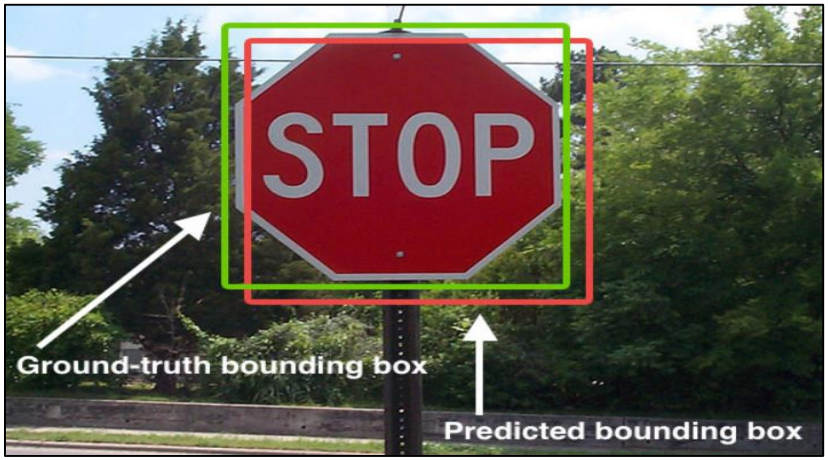
(w, h) 는 해당 셀에서 bounding box가 차지하는 width, height

$$\text{confidence score} = \Pr(\text{Object}) * \text{IoU}_{\text{pred}}^{\text{truth}}$$

Confidence: 해당 그리드 셀에 물체가 존재할 확률을 나타냄

[참고] Ground-truth, IoU

✓ 딥러닝에서 Ground-truth는 학습 데이터의 실제 값을 표현할 때 사용되는 개념. 즉 학습데이터에서 주어지는 정답(label)으로 생각해도 무방함.



✓ 딥러닝 모델이 얼마나 예측을 잘 했는지를 알기 위해 IoU(Intersection over Union) 사용

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$