You Only Look Once

SKT Fellowship

LEE JINKYU

Department of Civil, Environmental and Architectural Engineering, Korea University

February 24, 2023

SKT AI Fellowship

Introduction

About YOLO

1. Before YOLO

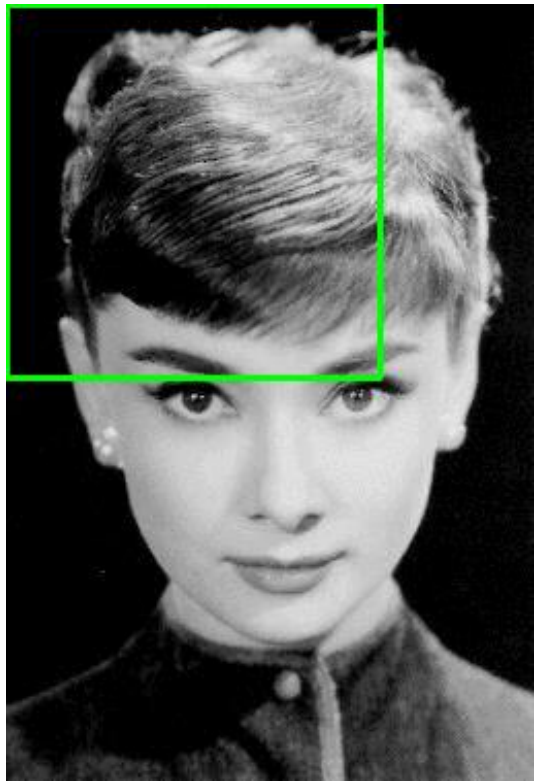
2. YOLO – You Only Look Once

3. Structure of YOLO

4. Training Methods

Before YOLO

Object Detection before YOLO



[Classification 기반 detection]

- Classification model을 detection에 맞게 변형 = Sliding window approach를 통해 각 approach마다 classify를 진행

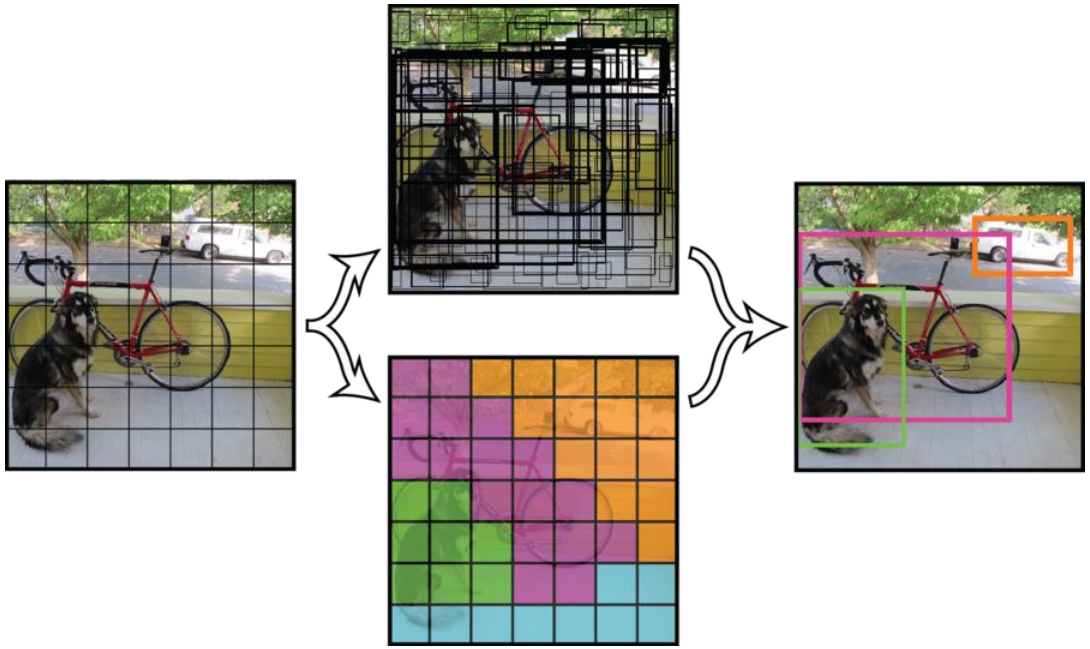
- R-CNN(YOLO 직전의 model)의 경우도 bounding box를 확률에 따라 제안하고 classification을 수행 (region proposal / classification / box regression 이라는 3단계 파이프 라인으로 진행, 복잡 + inference time이 길다)

→ **Classification이 아닌 Regression으로 model을 정의한다**

→ **Sliding window가 아닌 직접 bounding box의 좌표와 각 클래스의 확률을 구한다**

YOLO – You Only Look Once

About YOLO



YOLO : End-to-End의 통합된 구조로 이미지를 CNN에 한번 넣어 추론함으로써 다수의 bounding box와 class를 추론 (Sliding window와는 완전히 다른 개념)

→ 다른 모델에 비해 real time 동작이 가능 + 그에 비해 높은 Map
→ Sliding window가 아닌 전체 이미지를 파악해 이미지 전체적인 맥락 파악에 유리, 각 class에 대해 더욱 잘 예측
→ Object의 일반적인 특성을 잘 학습 (natural로 학습된 yolo를 artwork에 동작 시켜도 타 모델에 비해 좋은 성능)

But,

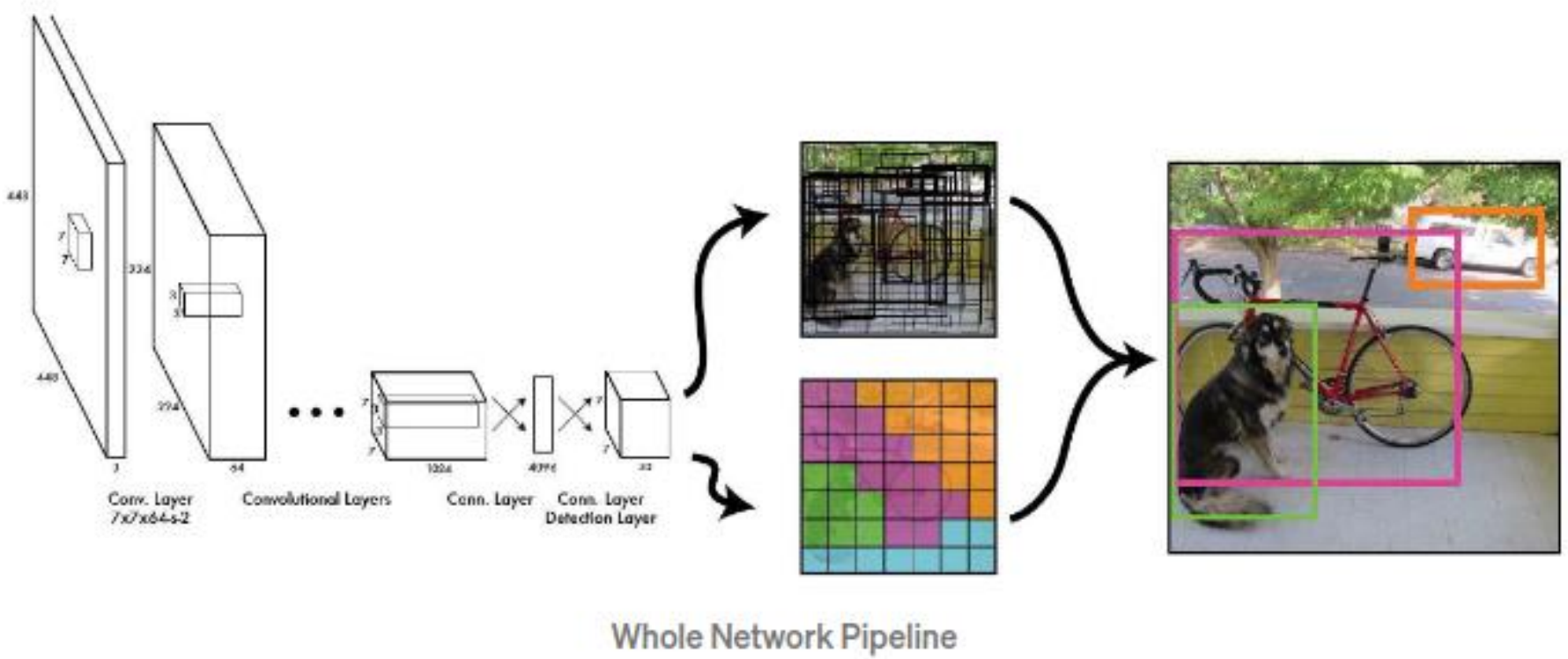
→ SOTA에 못 미치는 성능
→ 작은 물체에 대한 detect는 잘 못함

Structure of YOLO – Unified detection

End to End real time model

YOLO : Unified Detection으로 하나의 CNN을 통해 Object detection을 위한 feature extract / bounding box regression / class prediction을 모두 수행

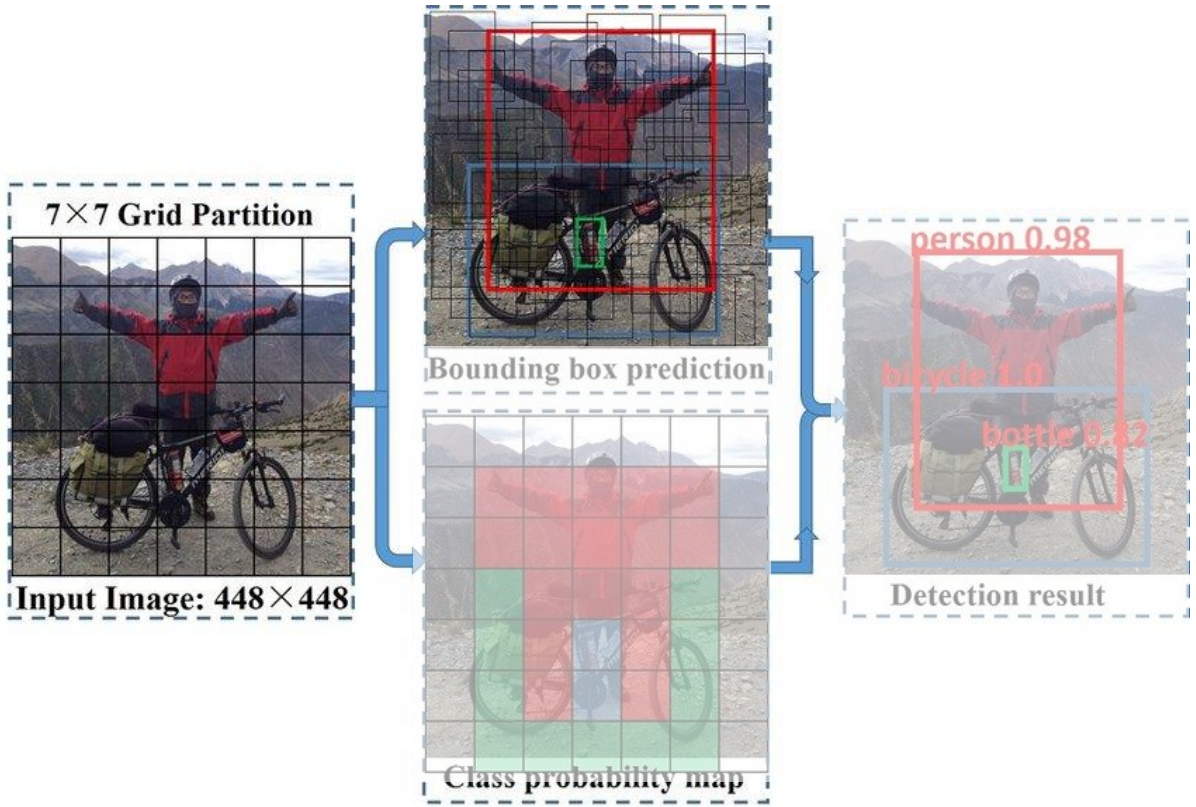
→ Bounding box regression, multi class regression을 동시에 진행
→ 높은 Map, end to end model, real time inference



Whole Network Pipeline

Structure of YOLO - Grid

S x S grid image



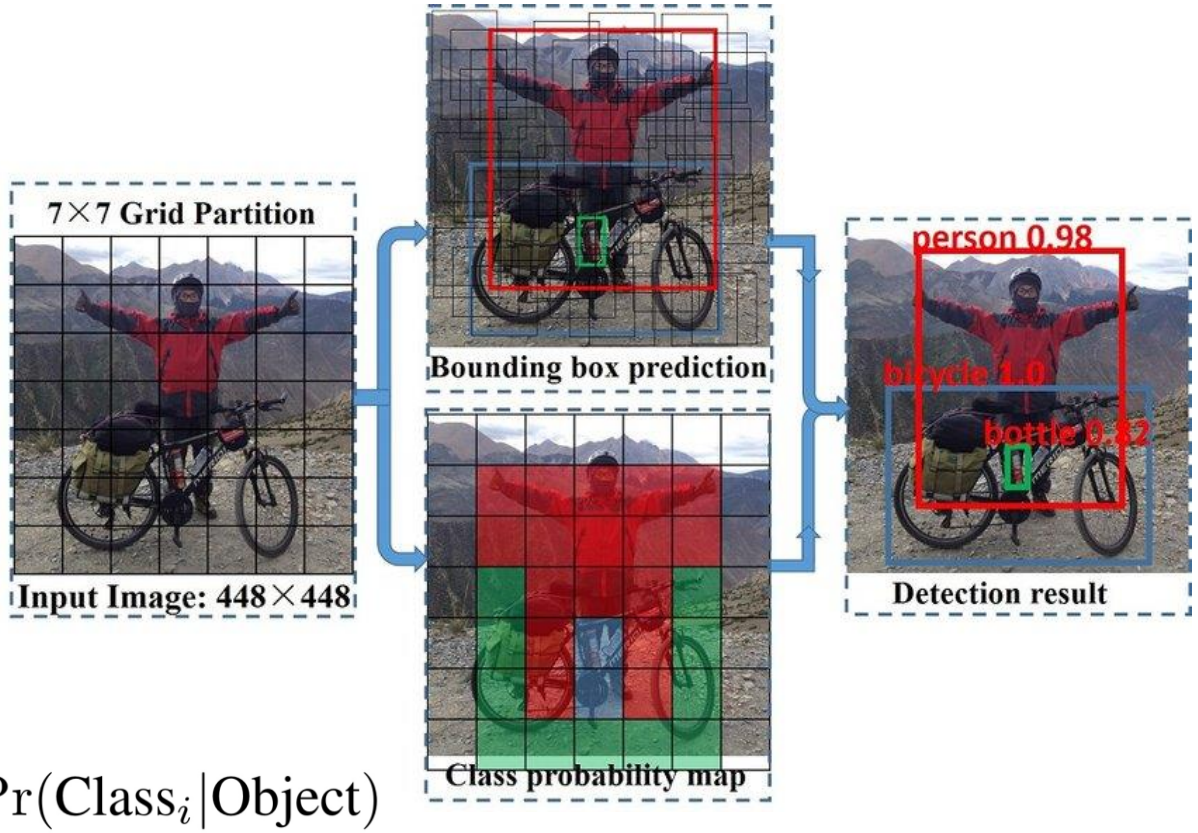
이미지를 s x s grid로 나눈다.
이때, 어떤 객체의 중점이 그리드에 존재할 때, 그 그리드가 해당 객체를 인식하고 bounding box를 그려야 한다.
(각 그리고 별로 B개의 box 생성 + 각 box마다 object 유무에 대한 confidence score 생성)

bounding box : (x ,y ,w ,h ,confidence)로 구성

Ex) S = 7, B = 2, C = 20 → 7 x 7 x 30

Structure of YOLO – Probability C per grid

Conditional Probability C



$$\Pr(\text{Class}_i|\text{Object})$$

그리드 당 하나의 확률 값 C를 예측한다. 이때 C는 B와 상관없이 오직 하나의 값만을 도출한다

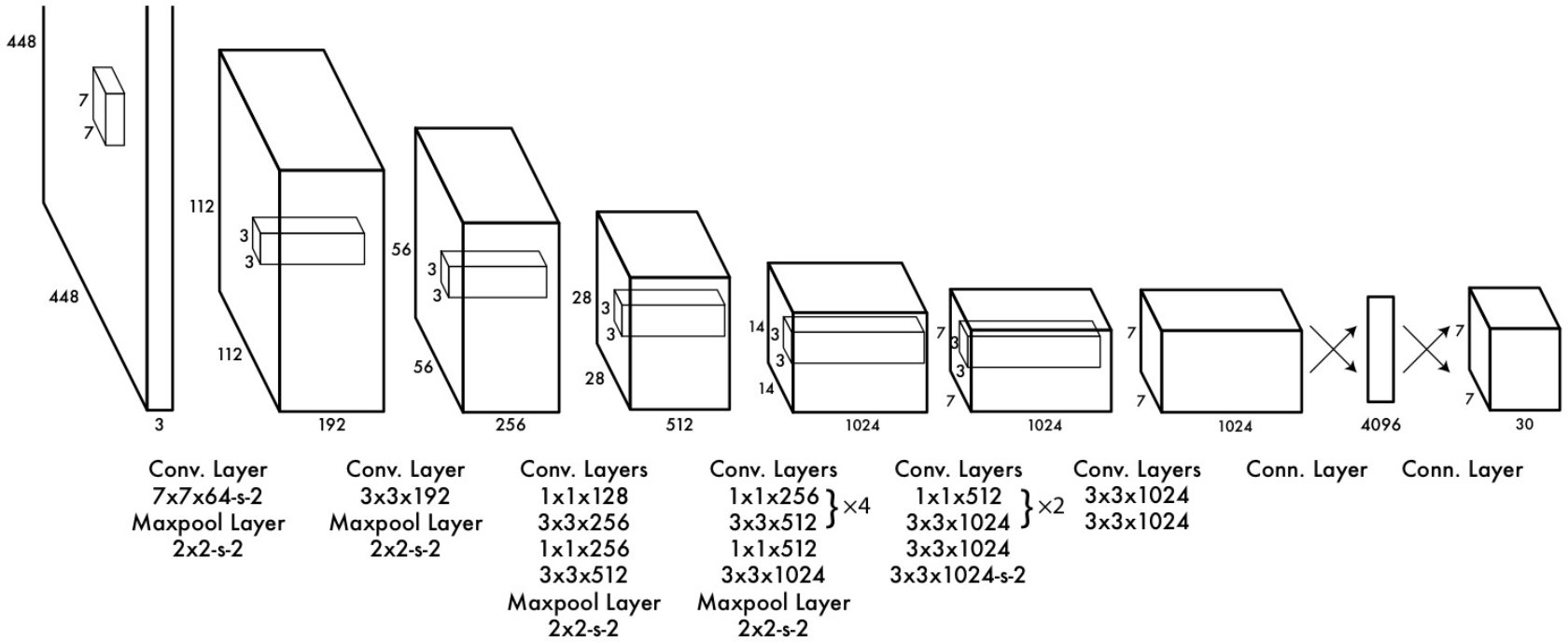
→ 도출된 C를 기반으로 bounding box의 confidence score를 생성

Structure of YOLO – CNN structure

Feature extraction & box regression

Conv. Layer : 이미지의 특징을 추출

Fully Connected layer : 클래스의 확률, bounding box의 좌표, 크기를 추출



Training Methods

To make YOLO effective

- Pretraining network : ImageNet으로 Conv layers를 pretrain
- Normalized bounding box : class의 확률, bounding box의 좌표는 모두 0 ~ 1 사이로 고정한다
- Nonlinearity : 활성화 함수로 leaky relu를 사용
- 학습 시 loss function의 가중치를 목적에 맞게 할당 → (small box / large box, localization constraint / classification constraint, co ord / no obj)

$$
\begin{aligned}
&\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
&+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2
\end{aligned}
$$