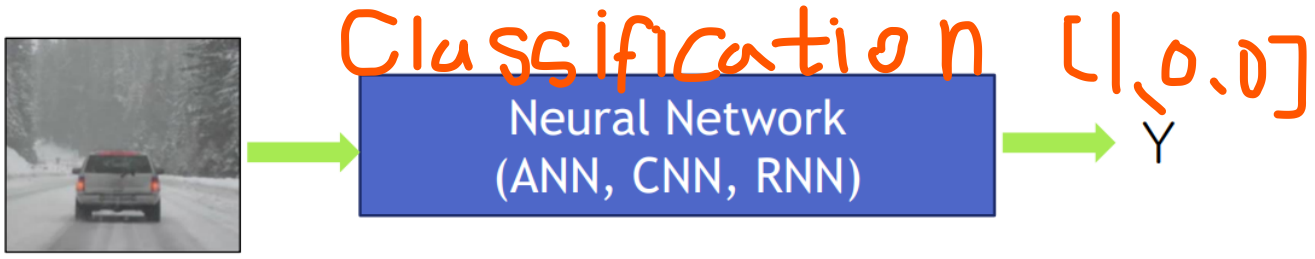
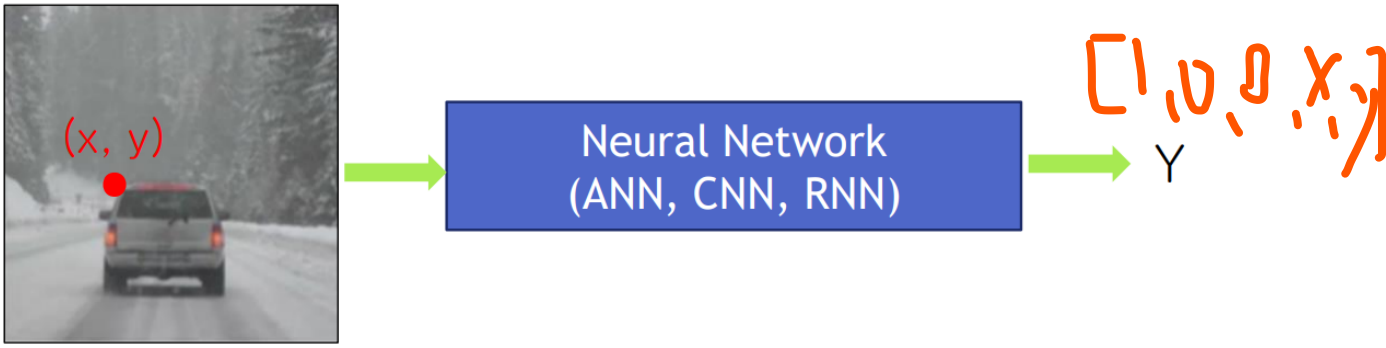
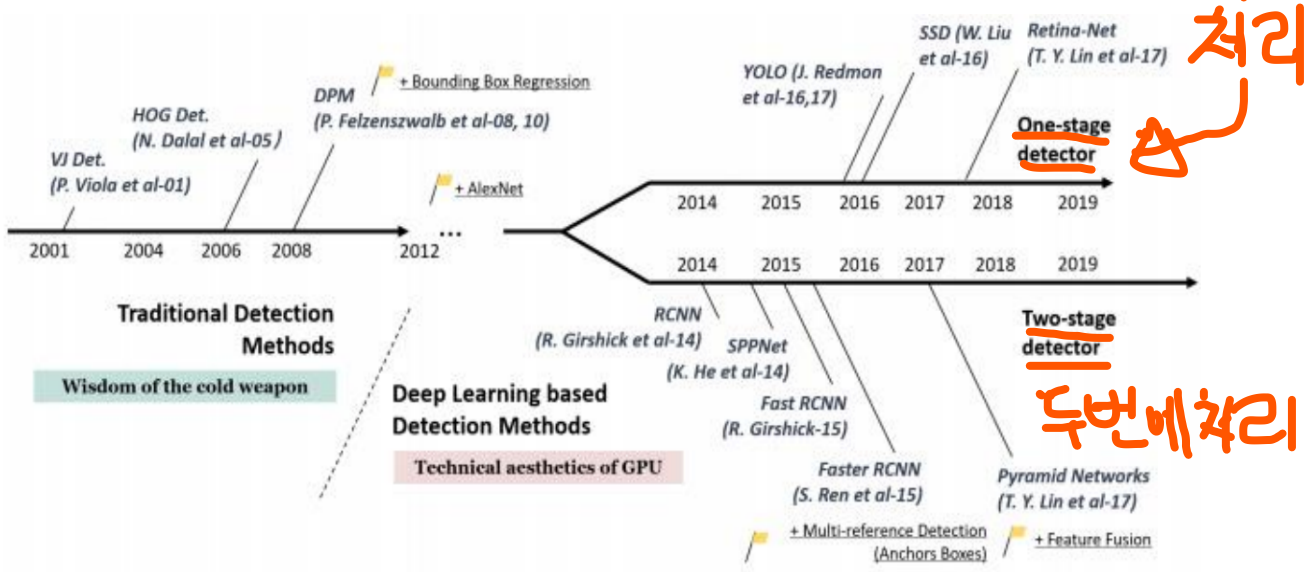
YOLO 교안

1. Object Detection



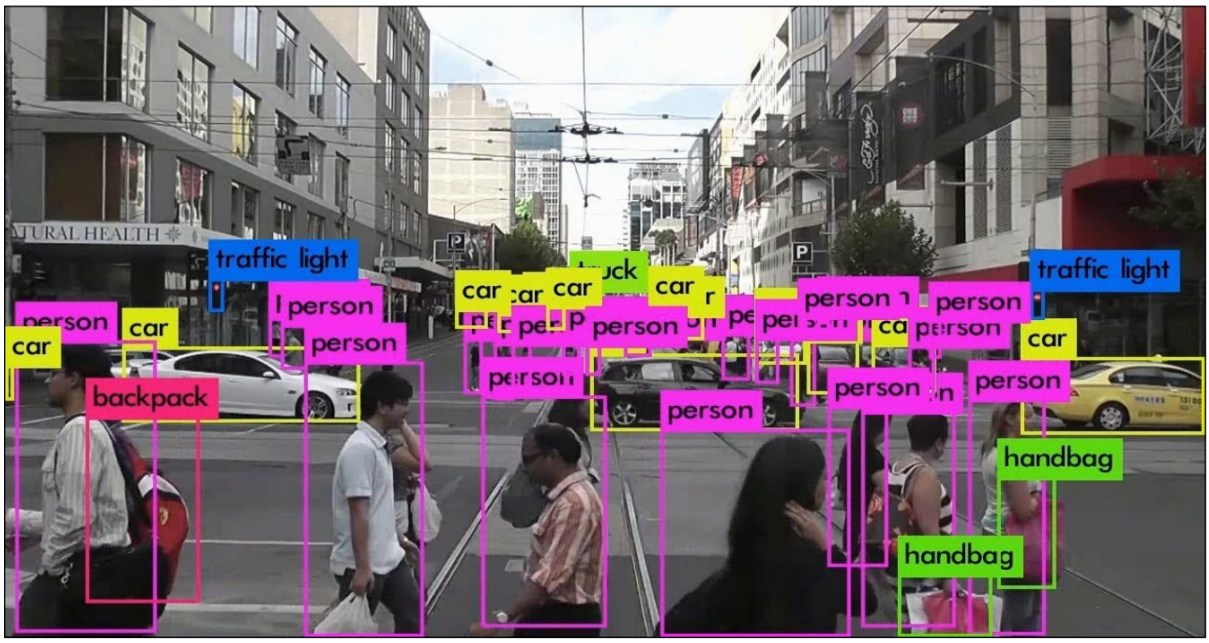


한번에 처리

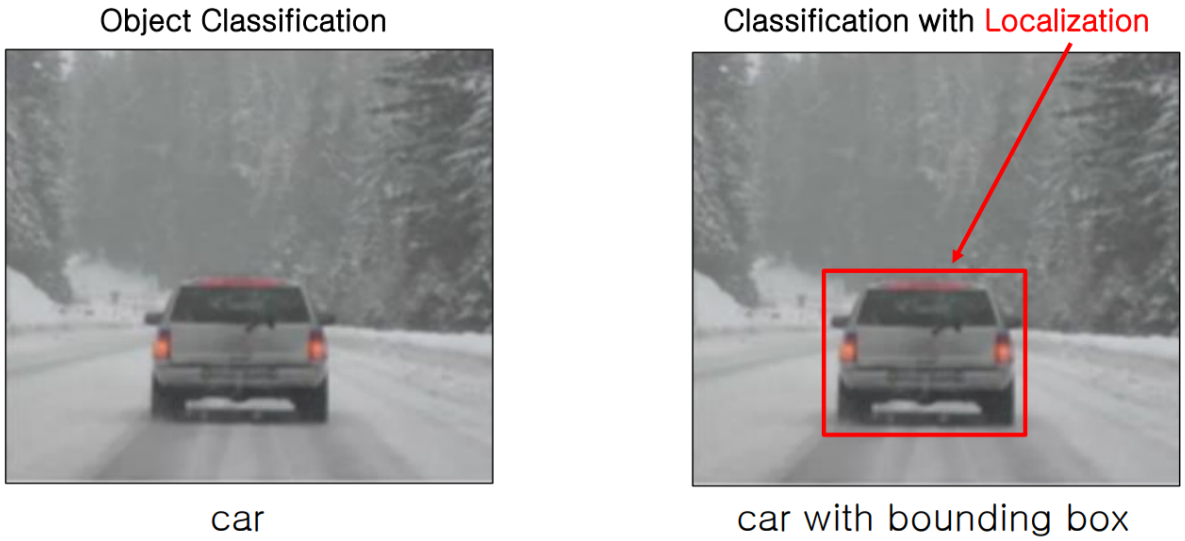


두번에처리

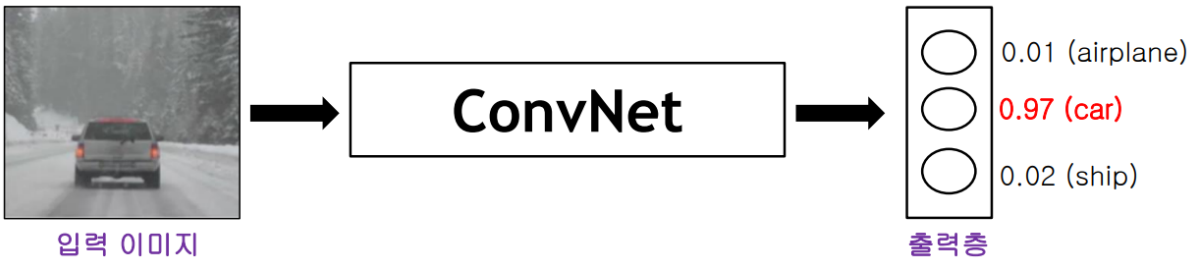
Object Detection = Object Classification + Object Localization
= Object Classification with Localization



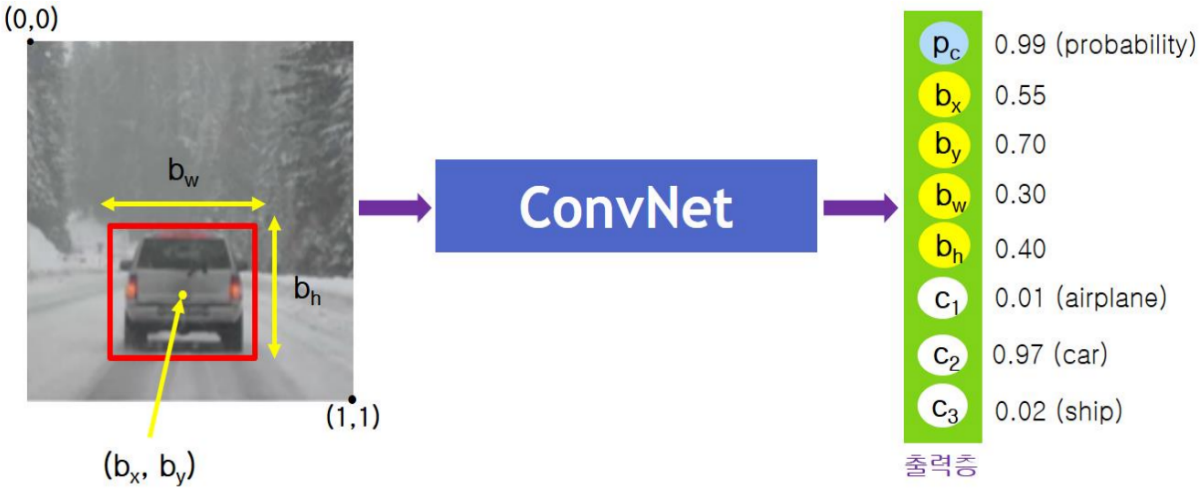
Object Classification with Localization - concept

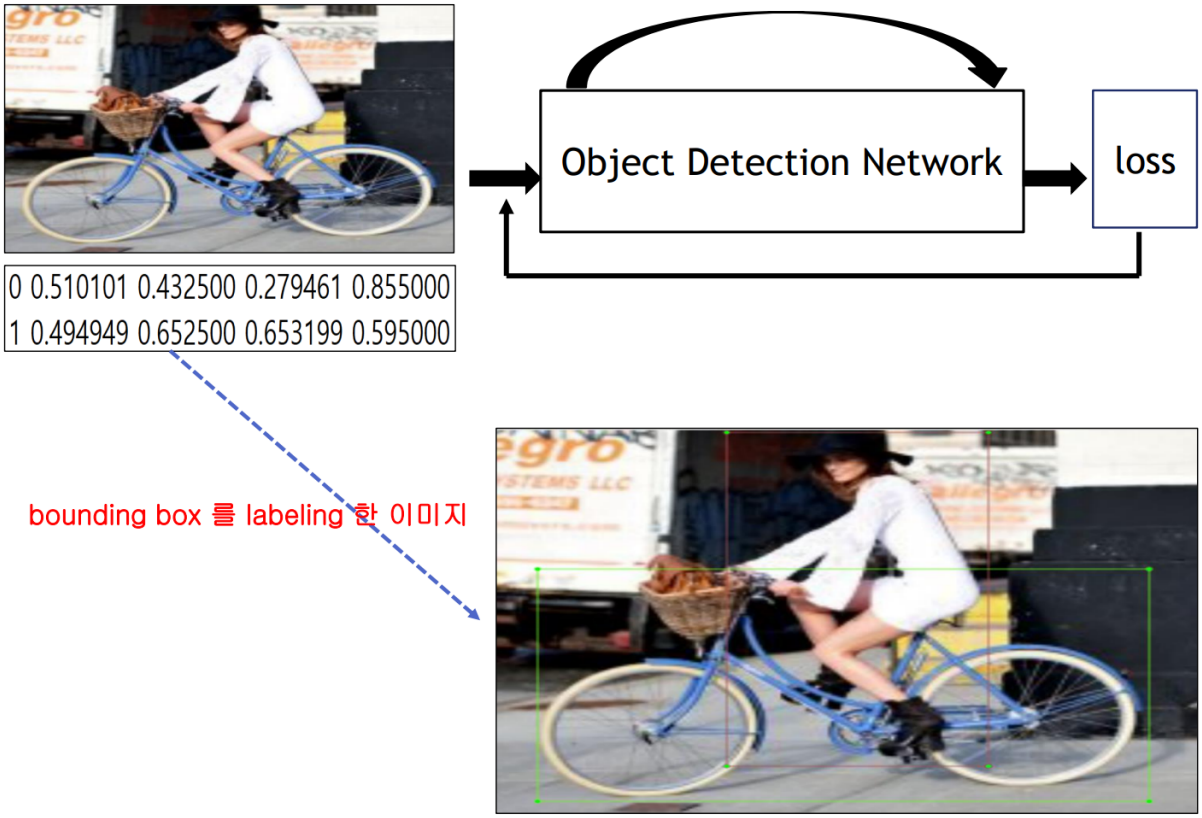


Object Classification using Softmax



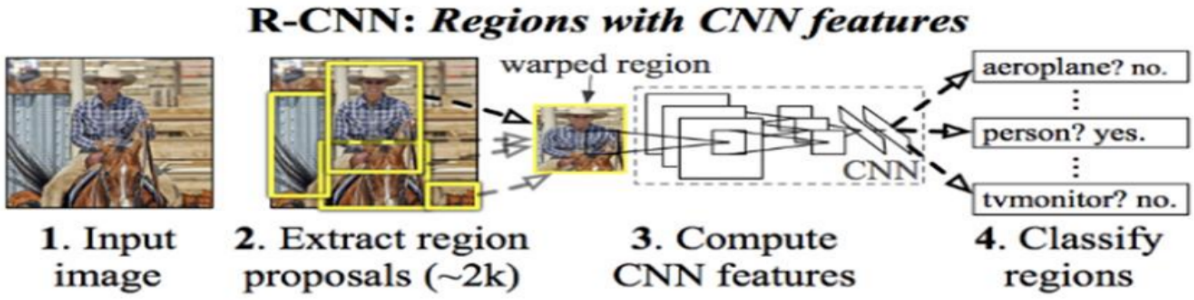
Object Localization using Bounding Box



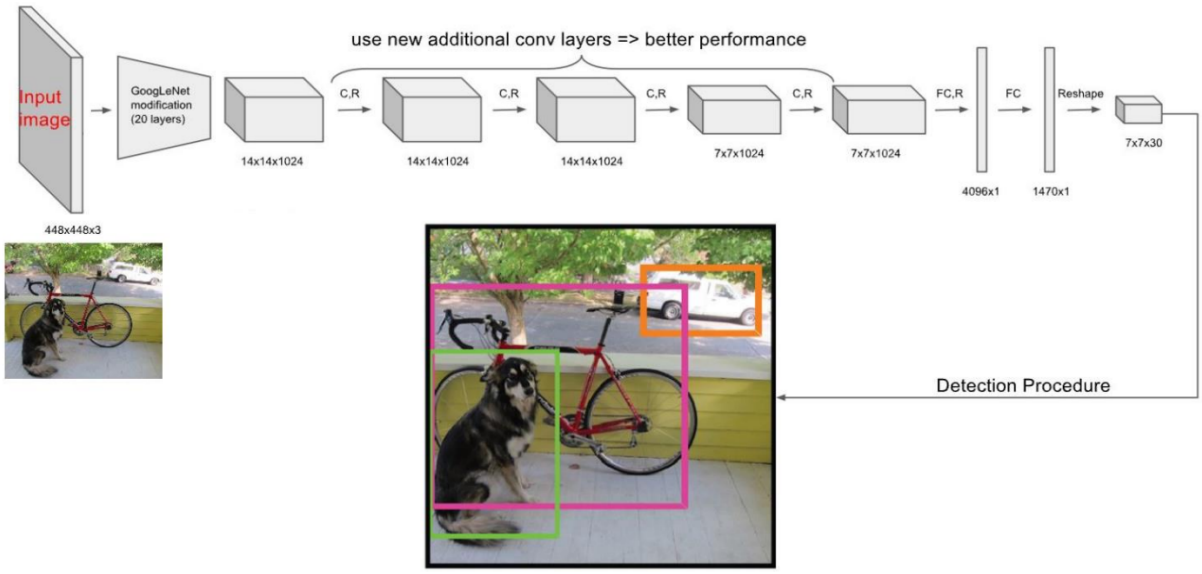


2. YOLO (You Only Look Once)

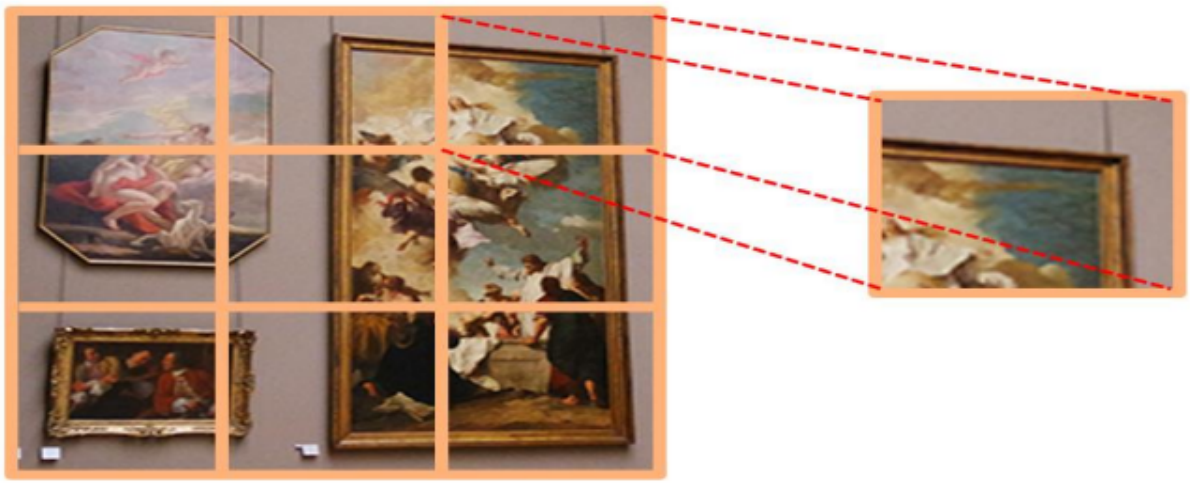
(1) two-shot-detection 개요

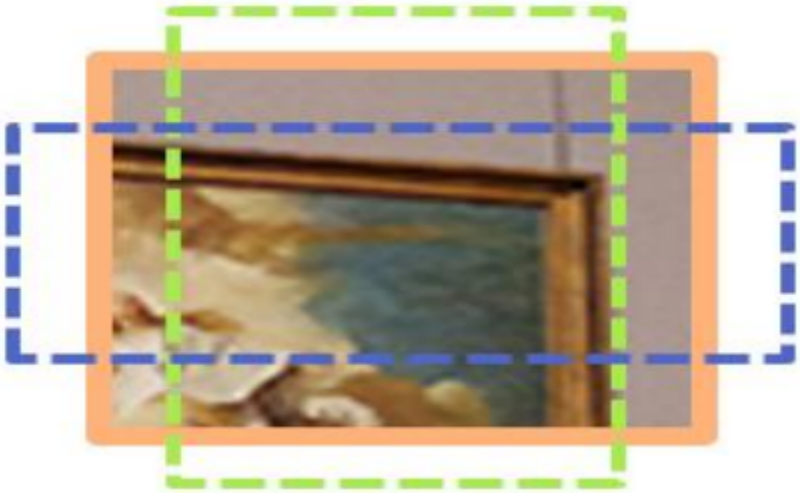


(2) one-shot-detection 개요(YOLO)



(3) grid cell, bounding box

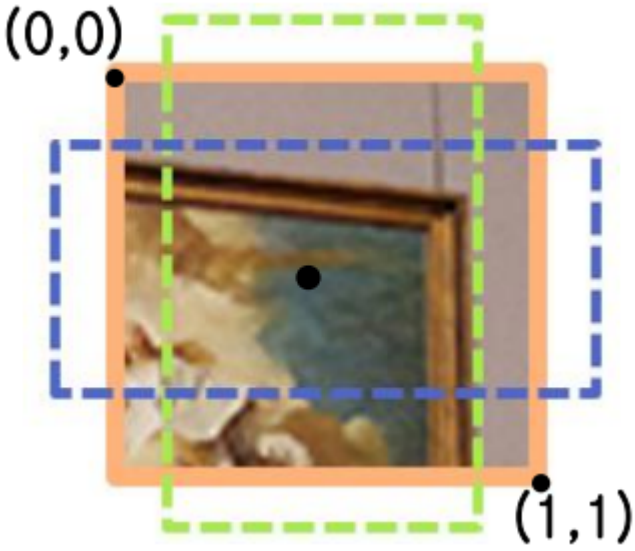




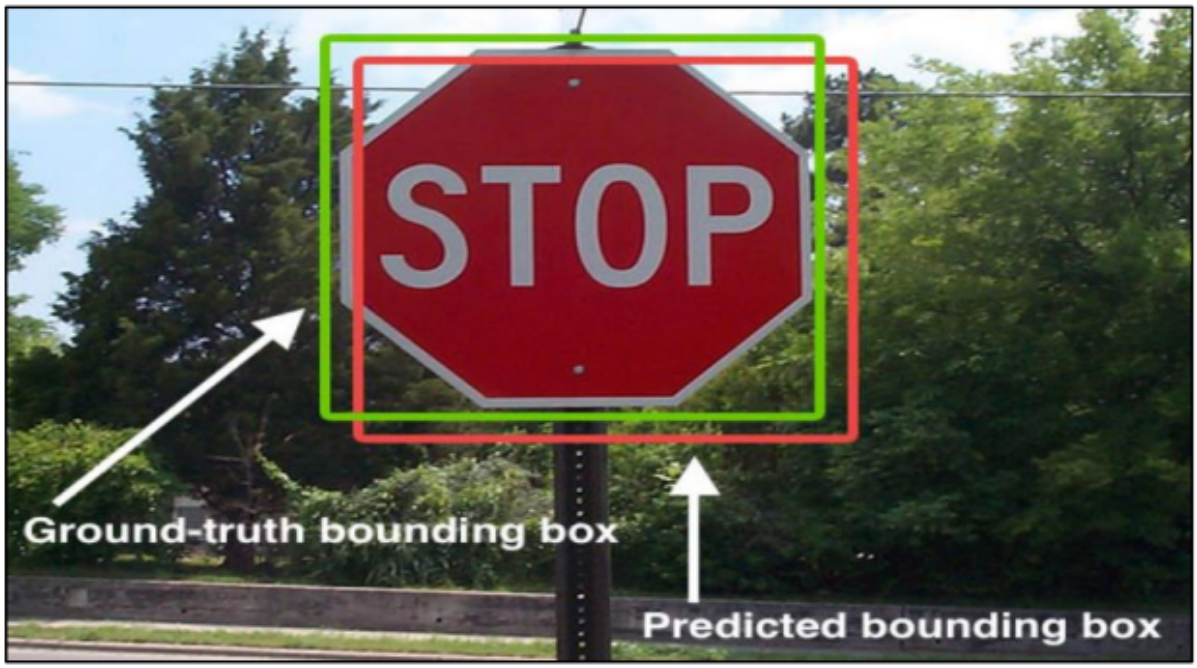
해당 그리드 셀에 물체가 존재할 확률

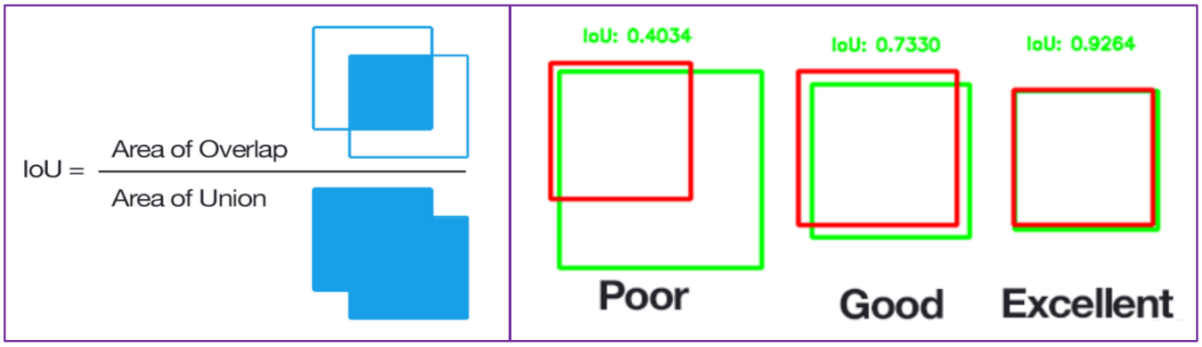
$$\text{confidence score} = \Pr(\text{Object}) * \text{IoU}_{\text{pred}}^{\text{truth}}$$

$$\text{conditional class probabilities} = \Pr(\text{Class}_i|\text{Object})$$



(4) ground-truth, IoU



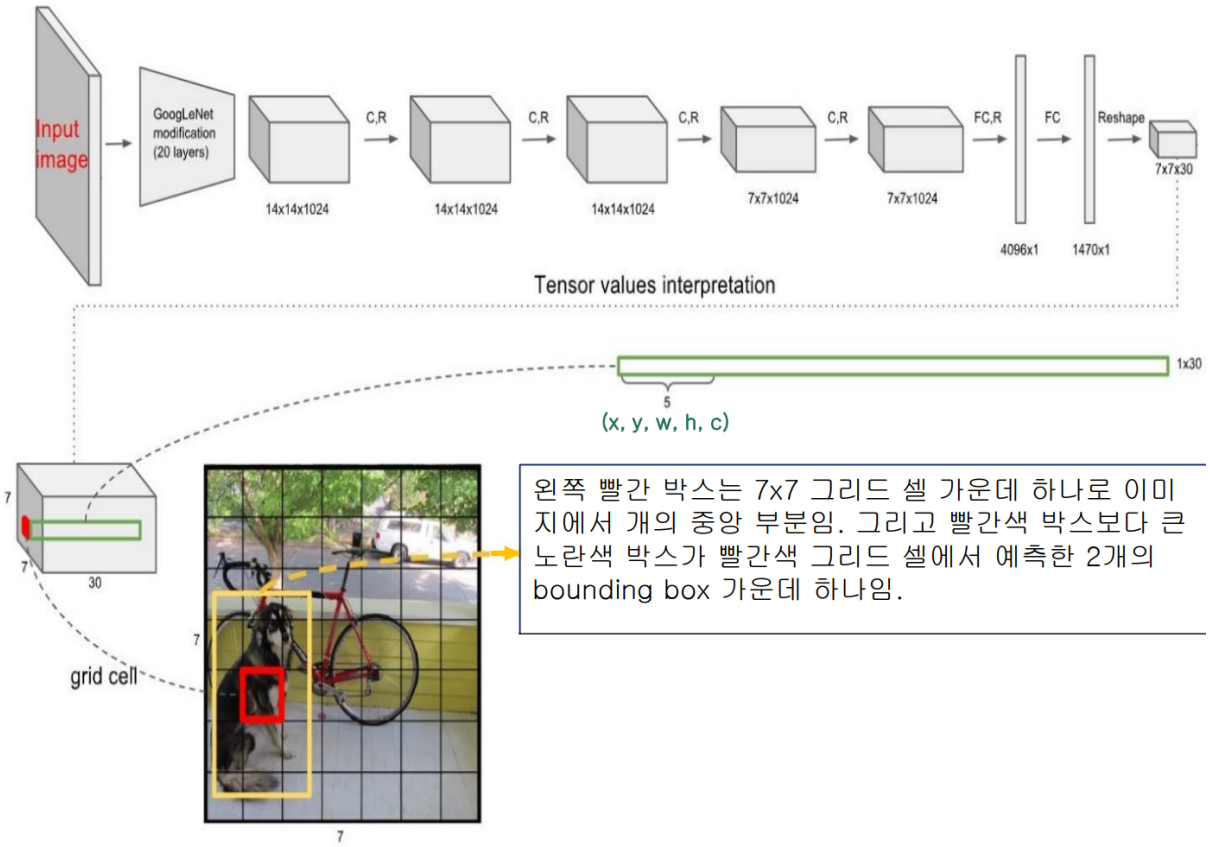


(5) YOLO Inference Process

Input Image

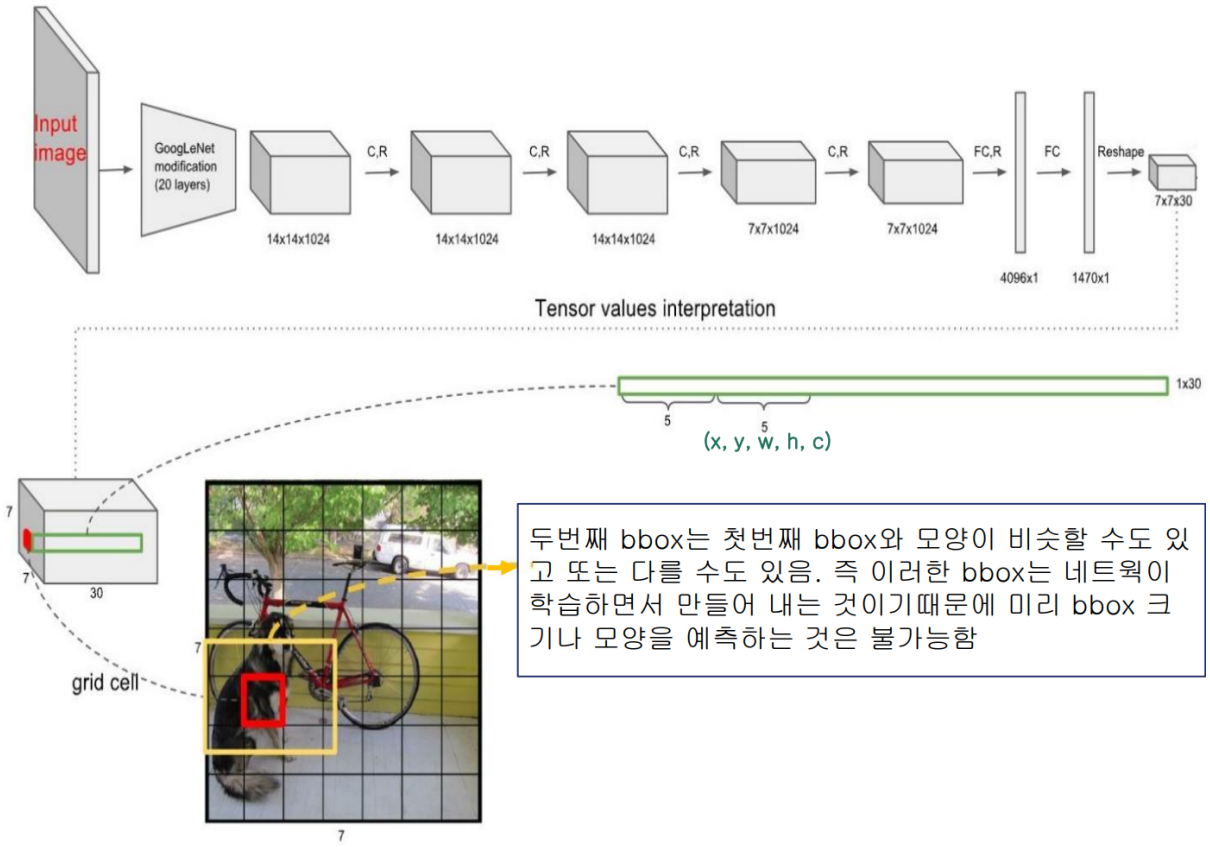
Inference result
(Prediction result)

1) 1st bounding box of a grid cell



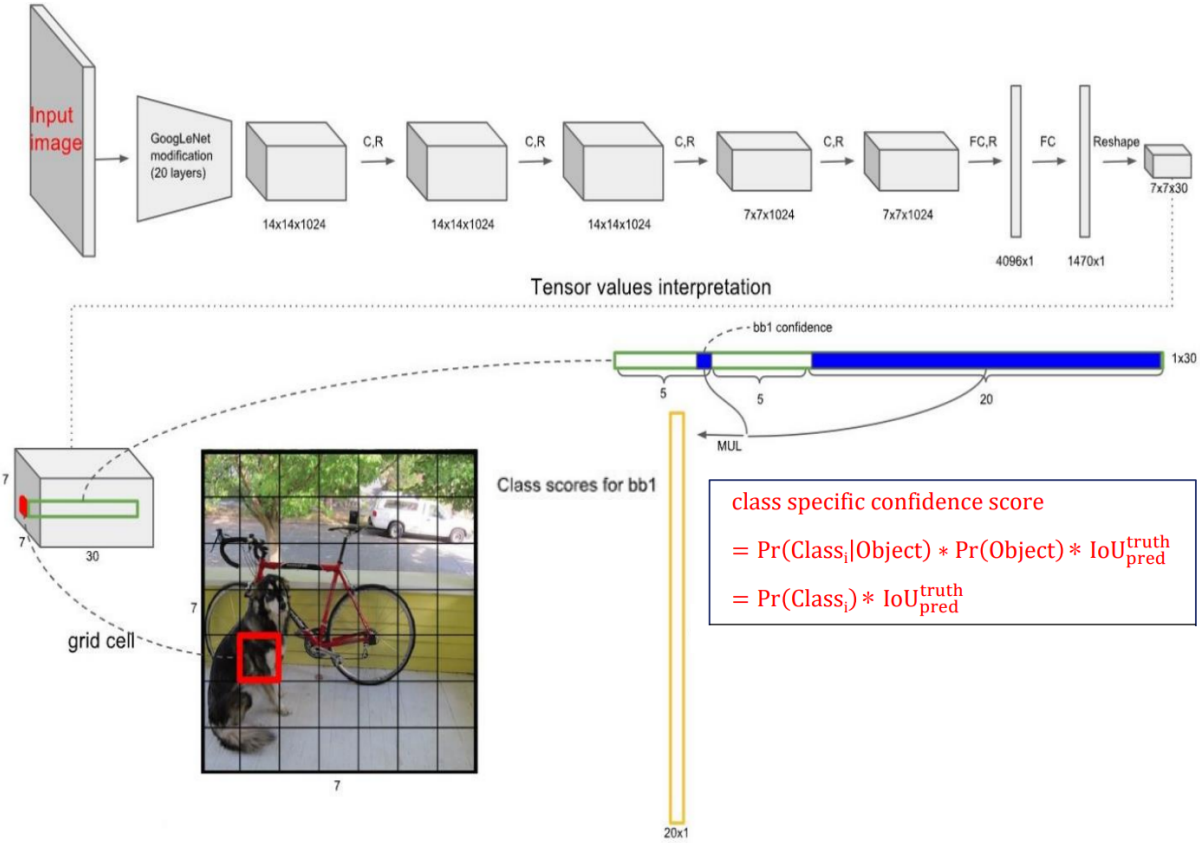
왼쪽 빨간 박스는 7x7 그리드 셀 가운데 하나로 이미지에서 개의 중앙 부분임. 그리고 빨간색 박스보다 큰 노란색 박스가 빨간색 그리드 셀에서 예측한 2개의 bounding box 가운데 하나임.

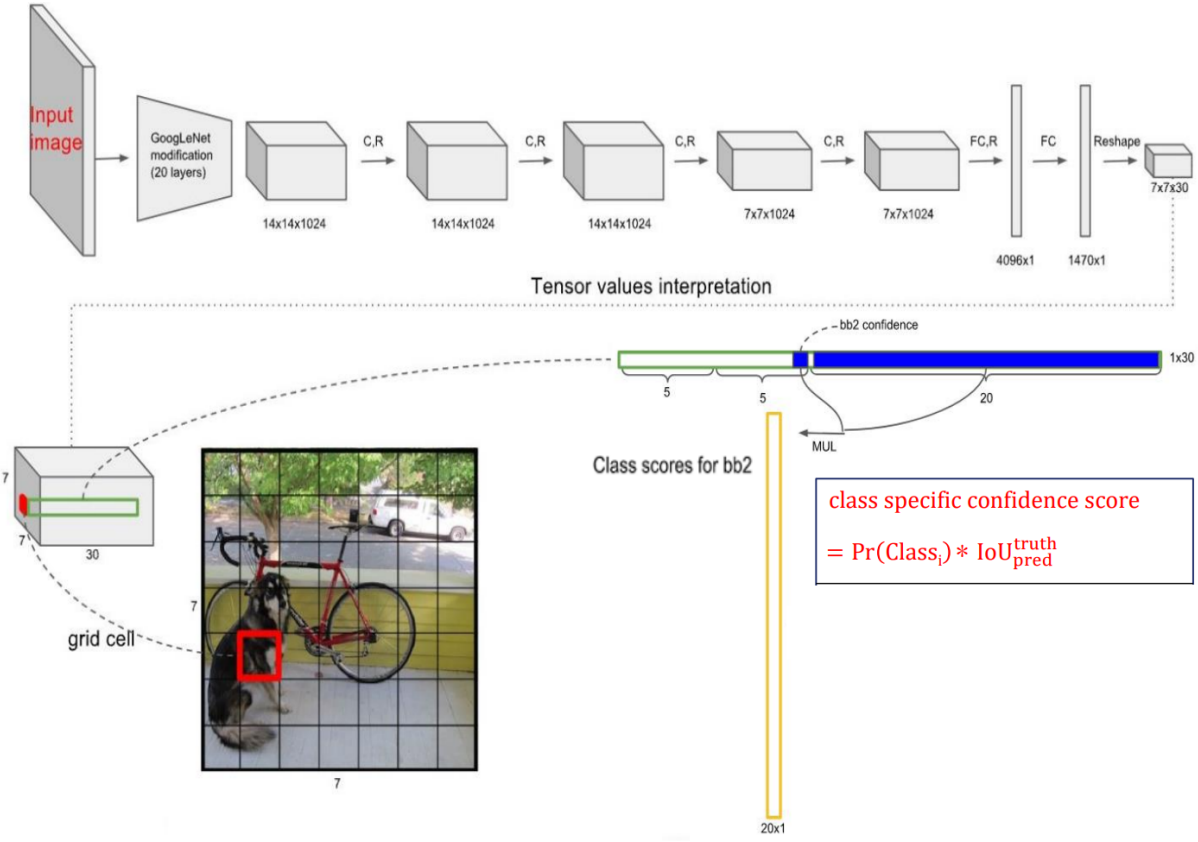
2) 2nd bounding box of a grid cell

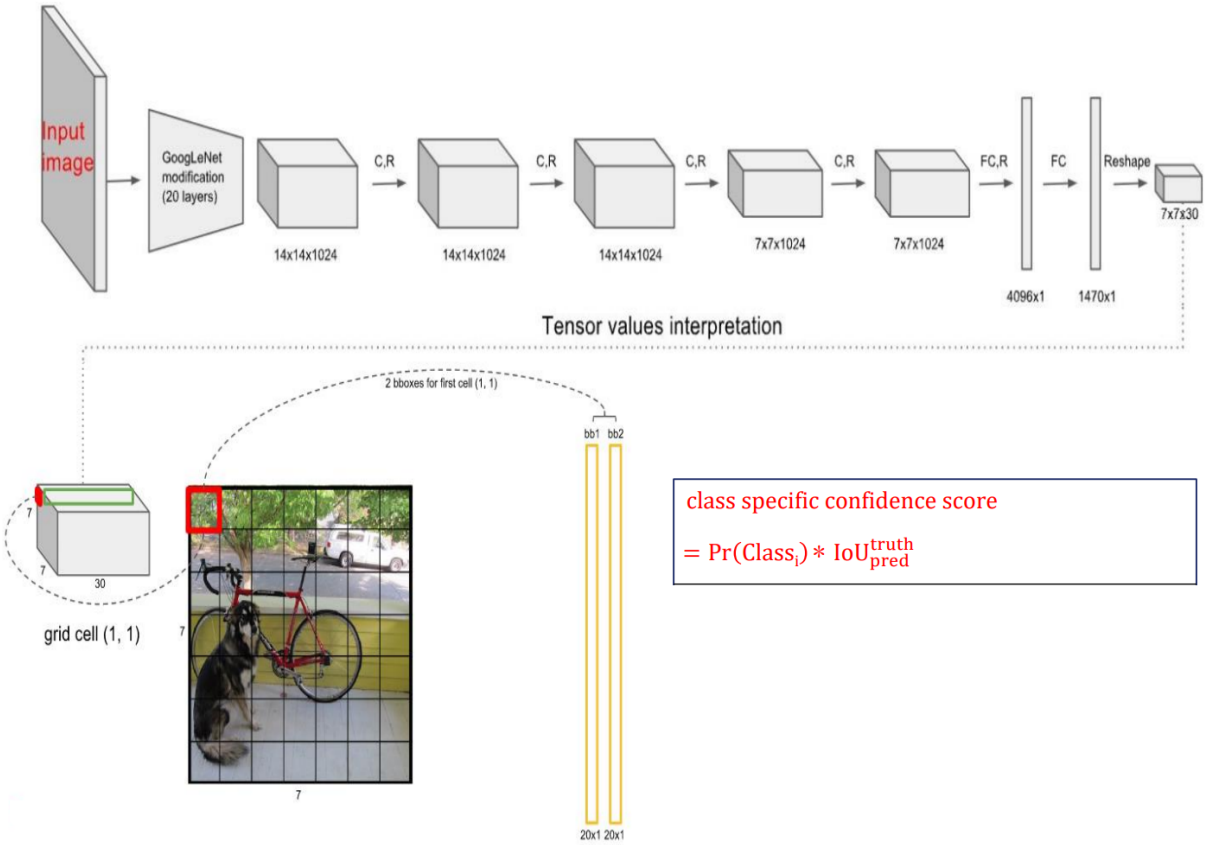


두번째 bbox는 첫번째 bbox와 모양이 비슷할 수도 있고 또는 다를 수도 있음. 즉 이러한 bbox는 네트웍이 학습하면서 만들어 내는 것이기때문에 미리 bbox 크기나 모양을 예측하는 것은 불가능함

3) class specific confidence score

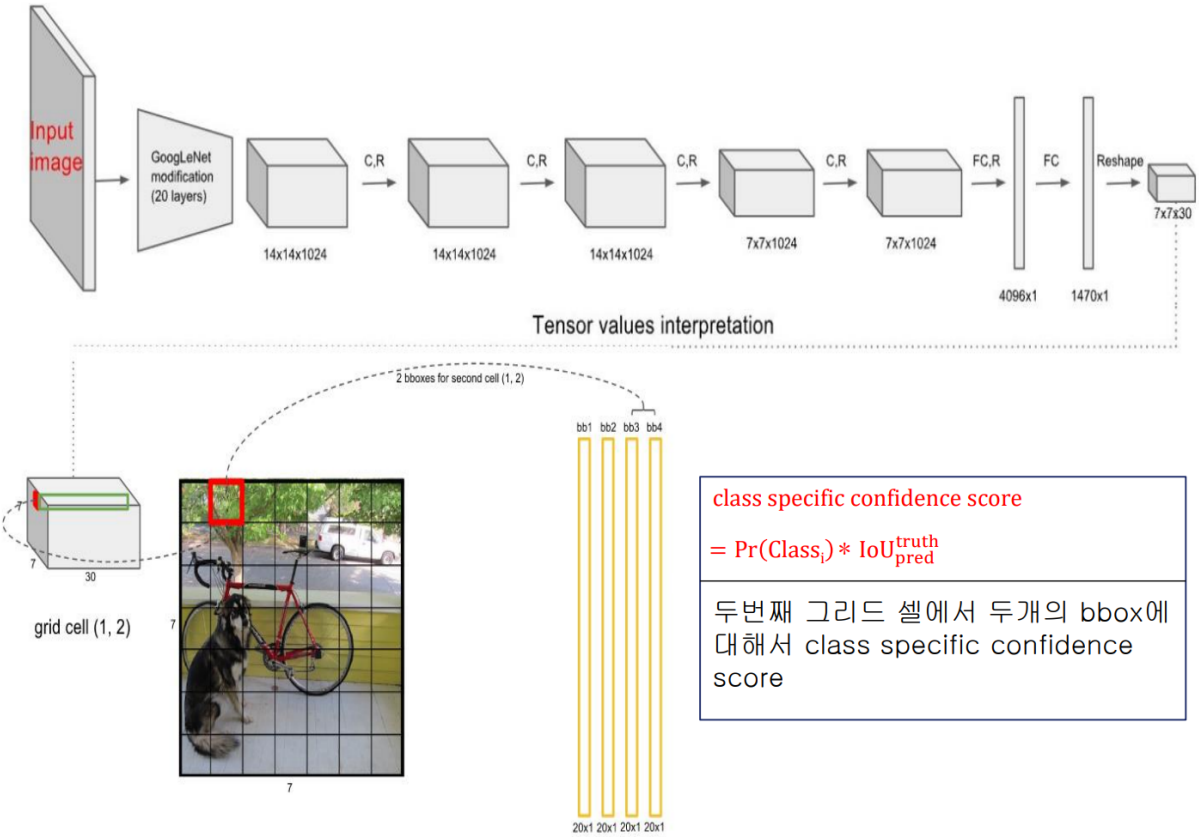

Input image
GoogLeNet modification (20 layers)
14x14x1024
C,R
14x14x1024
C,R
14x14x1024
C,R
7x7x1024
C,R
7x7x1024
FC,R
4096x1
FC
1470x1
Reshape
7x7x30
Tensor values interpretation
bb1 confidence
1x30
5
5
20
MUL
7
7
30
grid cell
7
7
Class scores for bb1
20x1
class specific confidence score
$= \Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IoU}^{\text{truth}}_{\text{pred}}$
$= \Pr(\text{Class}_i) * \text{IoU}^{\text{truth}}_{\text{pred}}$


Input image
GoogLeNet modification (20 layers)
14x14x1024
C,R
14x14x1024
C,R
14x14x1024
C,R
7x7x1024
C,R
7x7x1024
FC,R
4096x1
FC
1470x1
Reshape
7x7x30
Tensor values interpretation
bb2 confidence
1x30
5
5
20
MUL
7
7
30
grid cell
7
7
Class scores for bb2
20x1
class specific confidence score
$= \Pr(\text{Class}_i) * \text{IoU}^{\text{truth}}_{\text{pred}}$



class specific confidence score

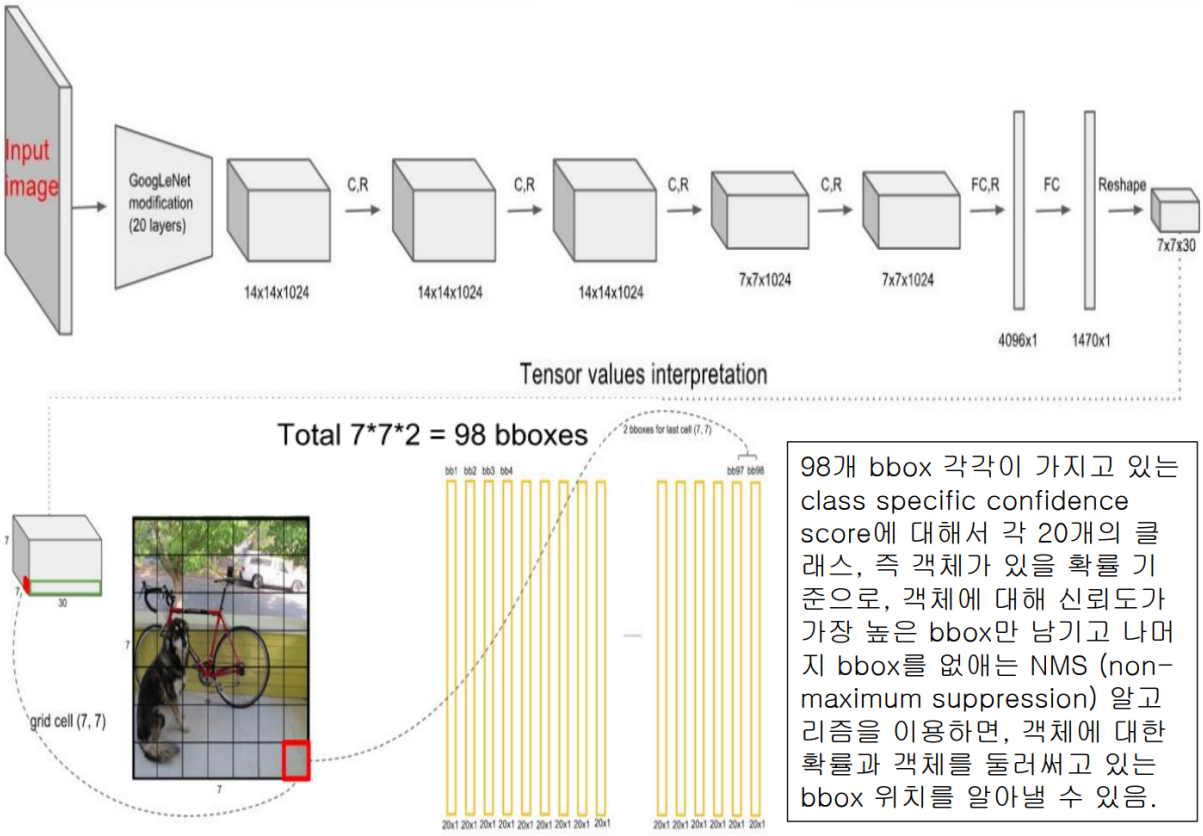
$= \Pr(\text{Class}_i) * \text{IoU}_{\text{pred}}^{\text{truth}}$



class specific confidence score

$= \Pr(\text{Class}_i) * \text{IoU}_{\text{pred}}^{\text{truth}}$

두번째 그리드 셀에서 두개의 bbox에 대해서 class specific confidence score



98개 bbox 각각이 가지고 있는 class specific confidence score에 대해서 각 20개의 클래스, 즉 객체가 있을 확률 기준으로, 객체에 대해 신뢰도가 가장 높은 bbox만 남기고 나머지 bbox를 없애는 NMS (non-maximum suppression) 알고리즘을 이용하면, 객체에 대한 확률과 객체를 둘러써고 있는 bbox 위치를 알아낼 수 있음.

4) NMS(Non-Maximum Suppression)