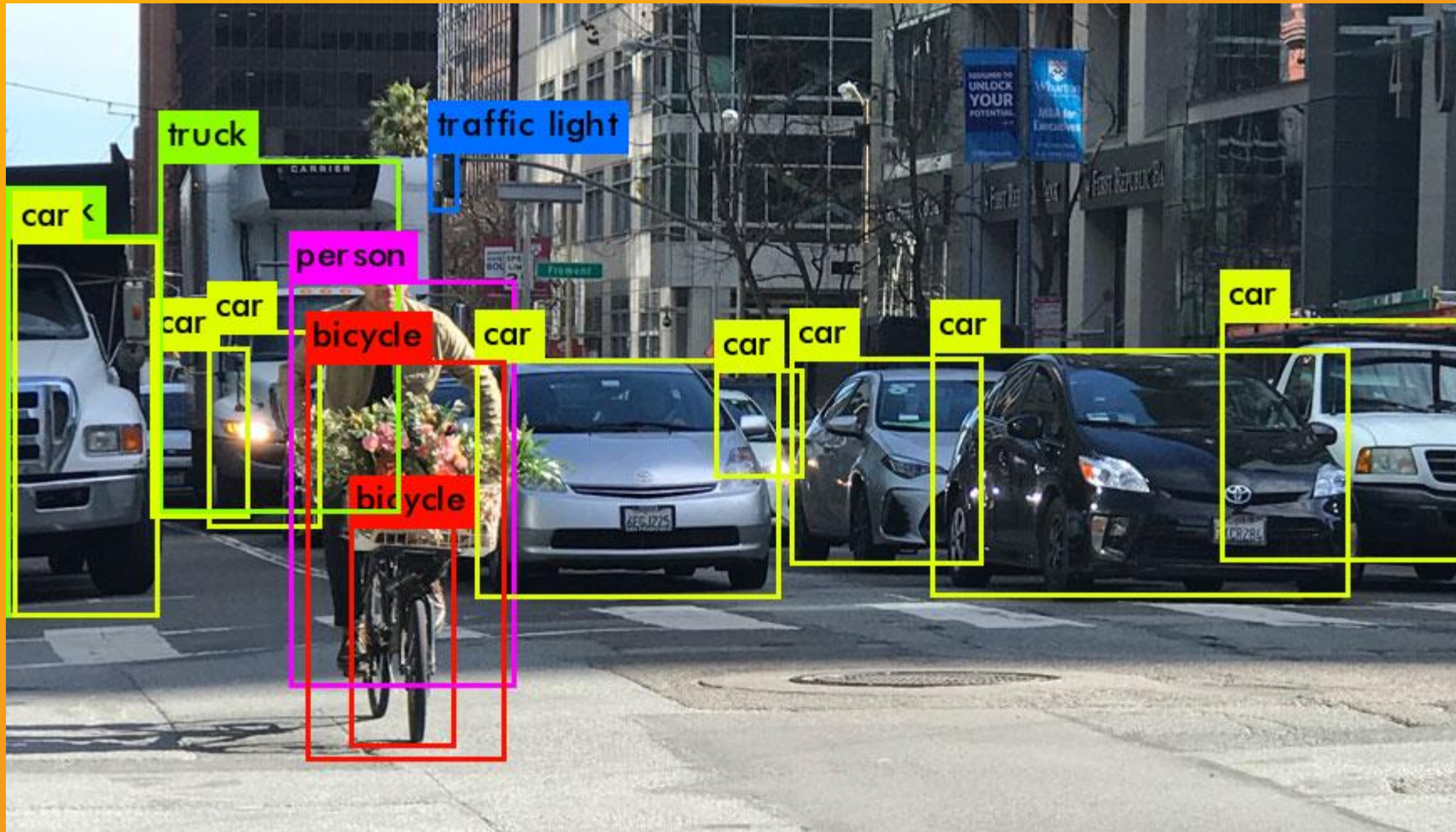
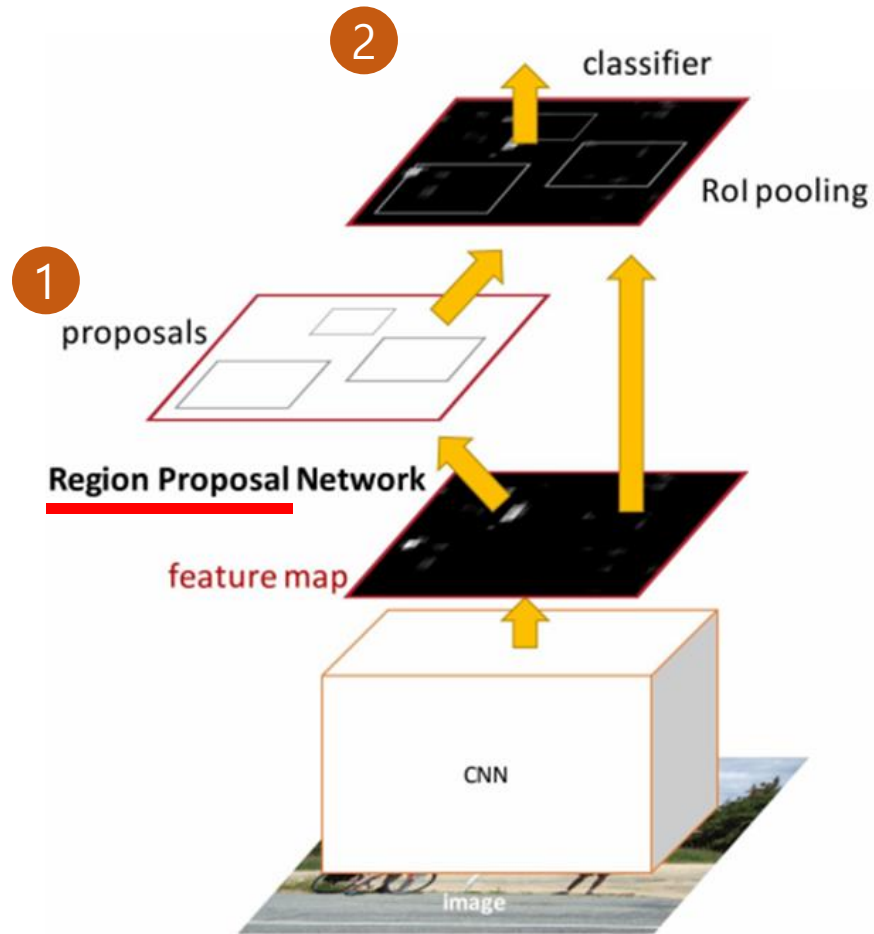


YOLO (You Only Look Once) – 실시간 객체 검출 알고리즘

KIM JEONG HYUN



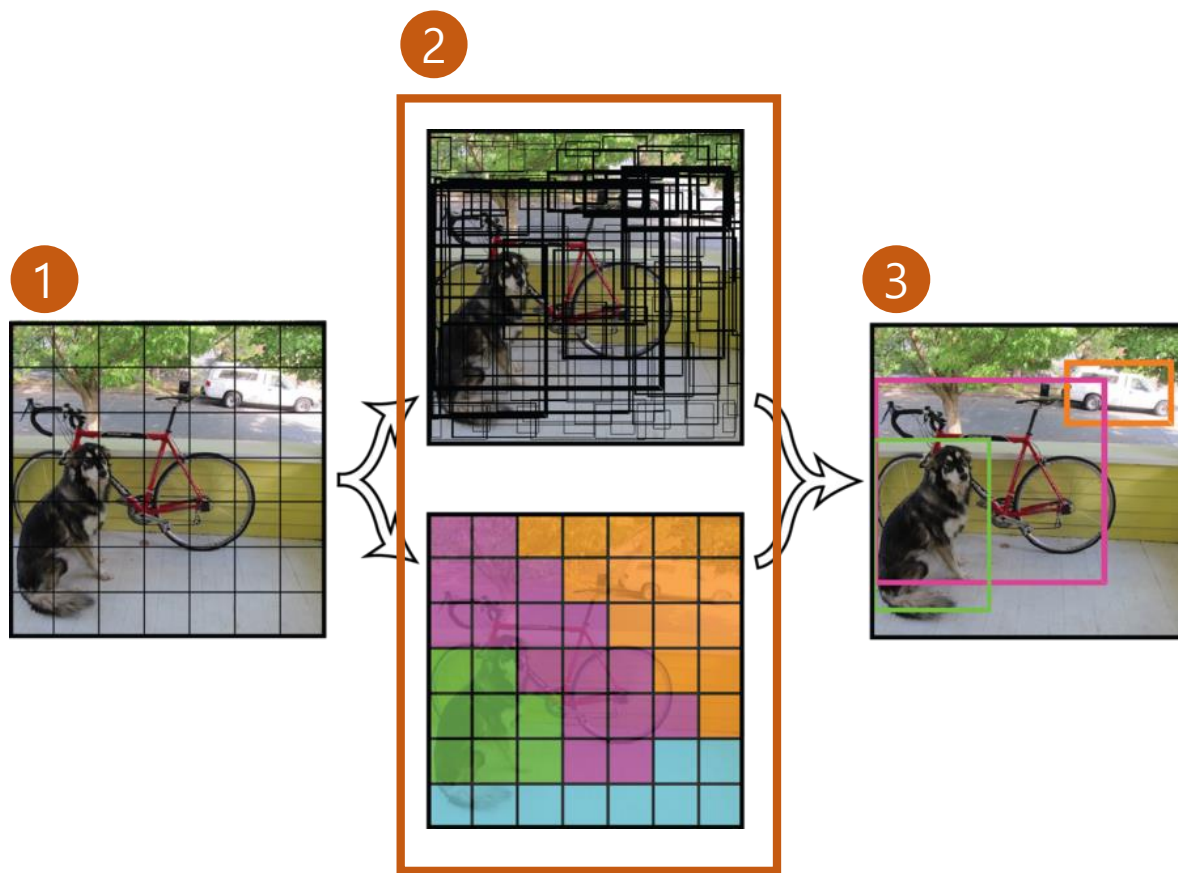
<https://www.youtube.com/watch?v=kxX09i4fds>



Fast R-CNN

오브젝트가 있을 것 같은 영역을 뽑아서 제안
객체 검출의 정확성은 높지만 신속하지 못함

1 2 = Two-stage Methods

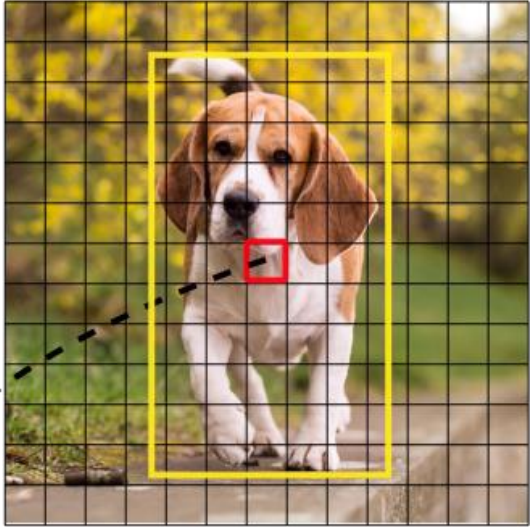


YOLO (You Only Look Once)

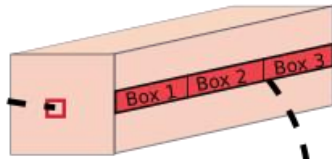
- 1** 프레임을 7*7 그리드로 분할
- 2** 각 그리드를 중심으로 Bound box 생성
 - 1) Bound box를 그리드 셀(7*7)의 2배만큼 생성
→ 프레임 상의 오브젝트 위치 예측
 - 2) 제안된 Bound box의 오브젝트 class 구분
- 3** 객체별 class와 영역 검출

 = **Single-stage Methods**

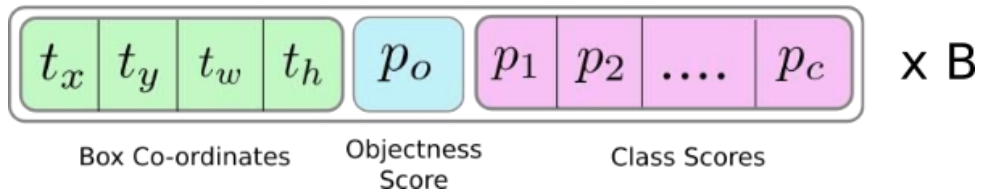
Image Grid. The Red Grid is responsible for detecting the dog



Prediction Feature Map



Attributes of a bounding box



Interpreting the output(=featuremap)

Input Image = 416 * 416 size

Stride = 32

→ cell = 13*13

임의의 cell 하나 당 3개의 bounding box(=B)

B 하나 당 5 + C개의 attr.

(t_x, t_y = B 의 x, y좌표

t_w, t_h = B 의 크기

P_o = 정확도(objectness score)

$C = P_x = \text{cat, dog, car, ...}$)

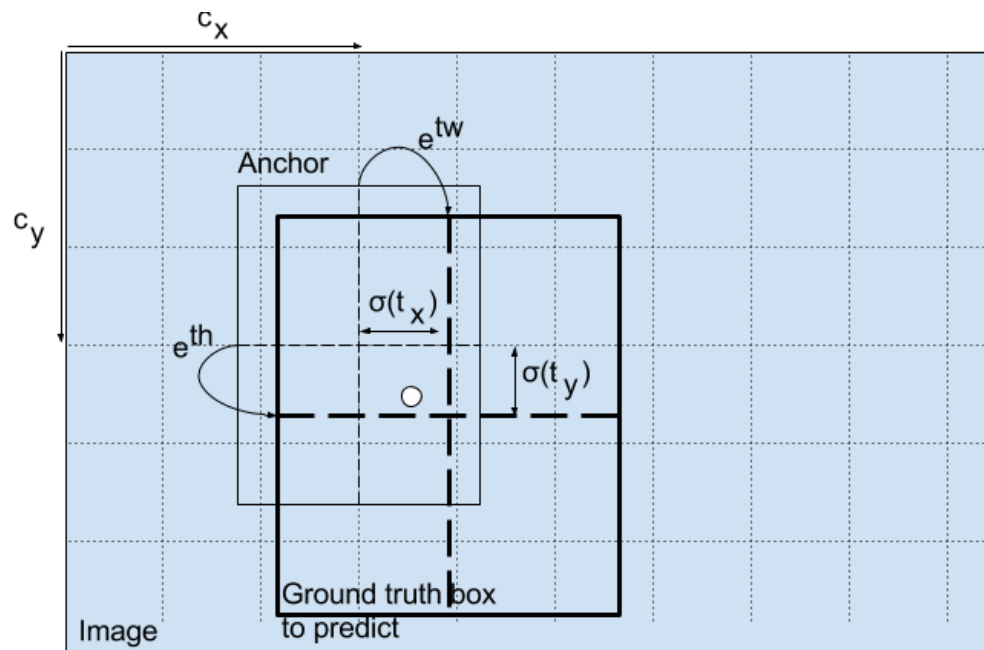
→ 총 $B * (5 * C)$ 개의 feature map

$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$



Anchor map

Anchor?: simply offsets to pre-defined default bounding boxes

x, y 좌표를 전체로 확대


: cx, cy = cell의 x, y좌표

→ $cx(cy) + tx(ty)$ = 전체 image에서 B의 좌표

이 때 전체 image의 크기를 벗어나지 않기 위해
tx, ty에 Sigmoid 함수를 사용 → 0~1



Multiple Grids may detect the same object
NMS is used to remove multiple detections

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


Output Processing

유사한 위치/크기의 feature map이 대량 발생

NMS를 활용

- IoU를 이용해 겹치는 비율로 유사도 측정
- 겹치는 Box들의 합집합으로 feature map 확정