

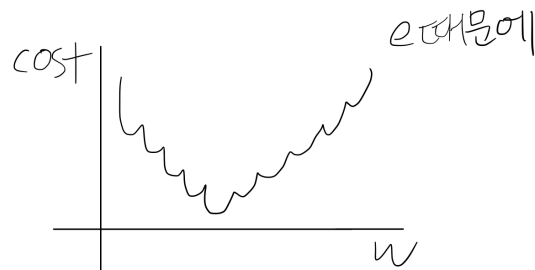
YOLOv3: An Incremental Improvement

M2019170

허성실

1. Class prediction

- Use multi-label classification (ex. woman and person)
-> sum of output may be greater than 1
- ① Softmax function -> independent logistic classifier
- ② Mean square error -> binary cross-entropy loss



2. Bounding box prediction

- $\text{Pr}(\text{Object}) > \text{threshold}(\text{default } 0.5)$ 인 bounding box만 남김.
- 각 Ground truth와의 IOU가 가장 큰 bounding box만 objectness score에 1을 줌.
- 즉, IOU가 가장 큰 bounding box를 제외한 나머지 bounding box에서는 confidence error만 고려. classification error와 localization error를 고려하지 않음.

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

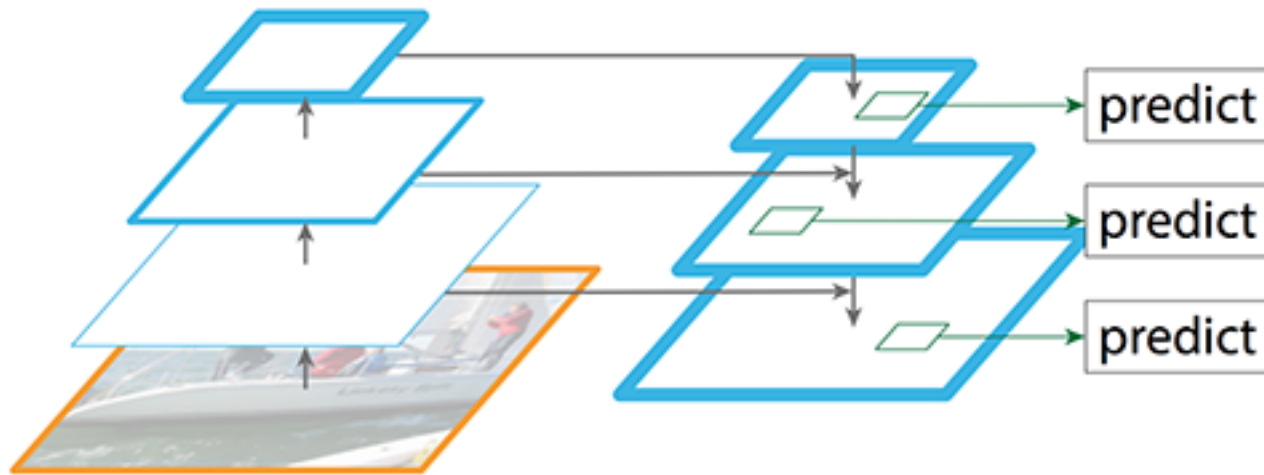
$$b_y = \sigma(t_y) + c_y$$

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

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

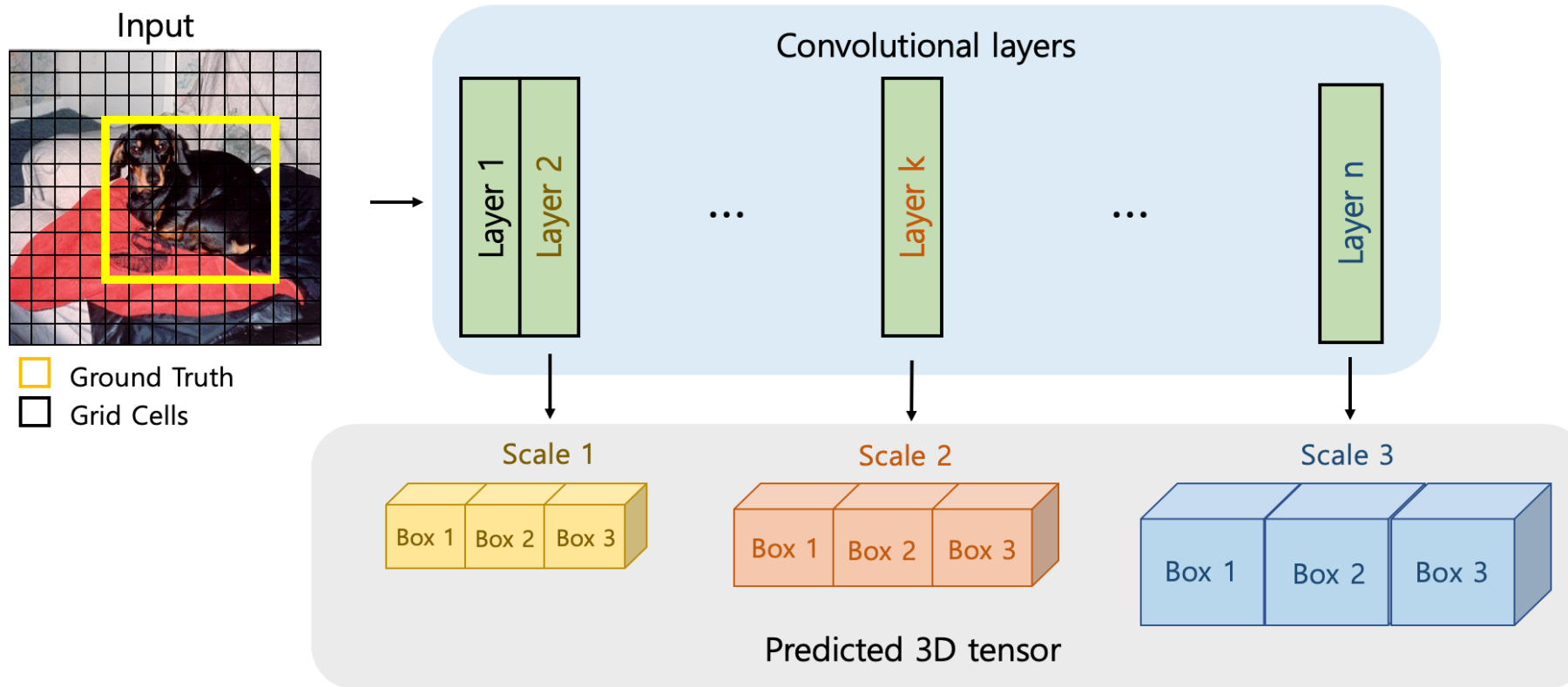
3. Predictions Across Scales

- Similar concept to Feature Pyramid Networks



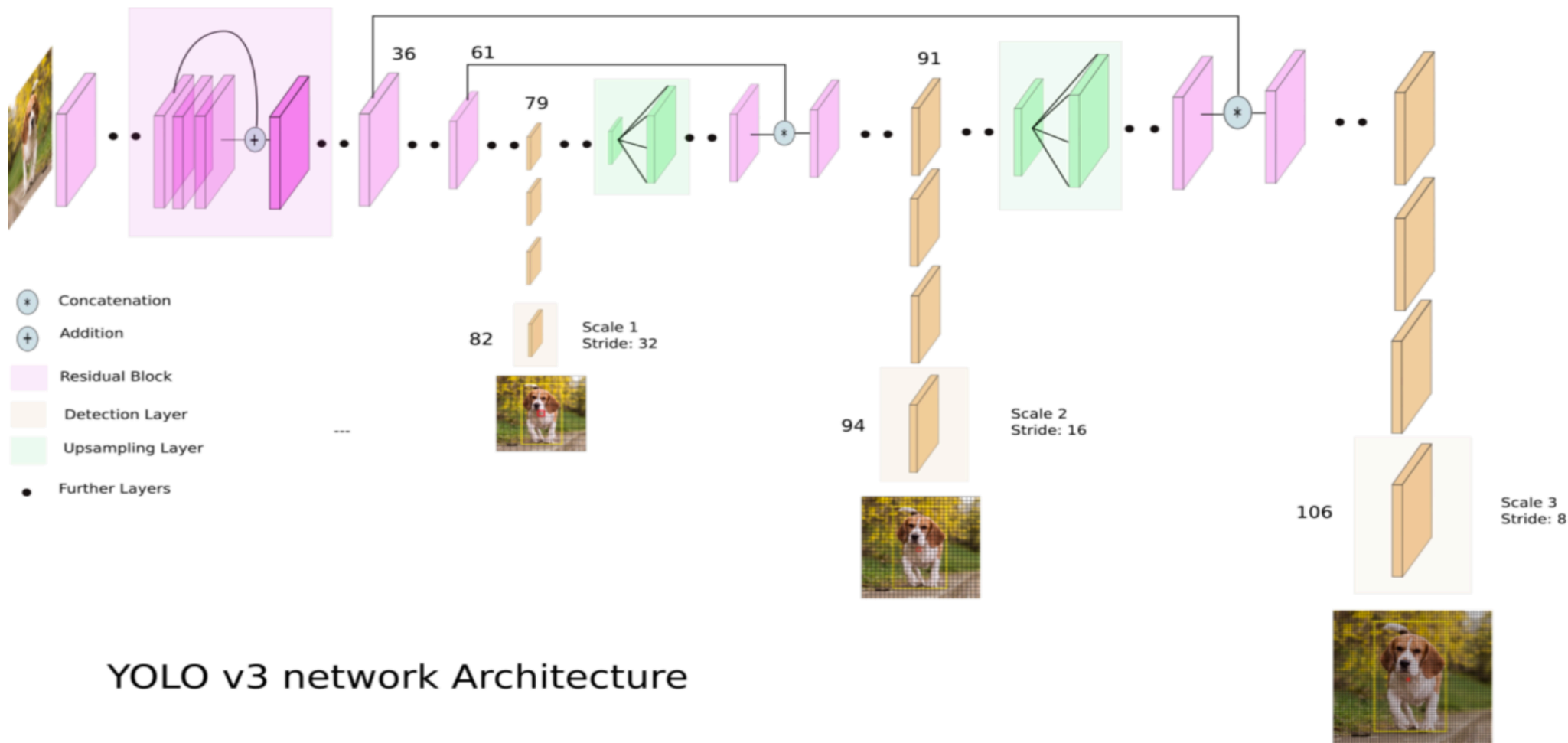
3. Predictions Across Scales

- Predicts 3 boxes at 3 different scales.
- tensor size : $N \times N \times [3 \times 5 + 80]$



3. Predictions Across Scales

- 마지막 layer를 x2 upsampling 한 것과, 그전의 레이어와 합친 Feature를 이용해 prediction.
- > upsample한 feature로부터 semantic한 정보, 이전 feature map에서 finer-grained한 정보



3. Predictions Across Scales

- Still use k-means clustering to determine bounding box priors.
- (10 x 13), (16 x 0 3), (33 x 23), (30 x 61), (62 x 45), (59 x 119), (116 x 90), (156 x 198), (373 x 326),

4. Feature extractor

- Darknet-53 아키텍처 사용
- Darknet-19에 비해 image classification 성능은 Resnet-101만큼 좋고, 속도도 1.5x 빠름.

Result

