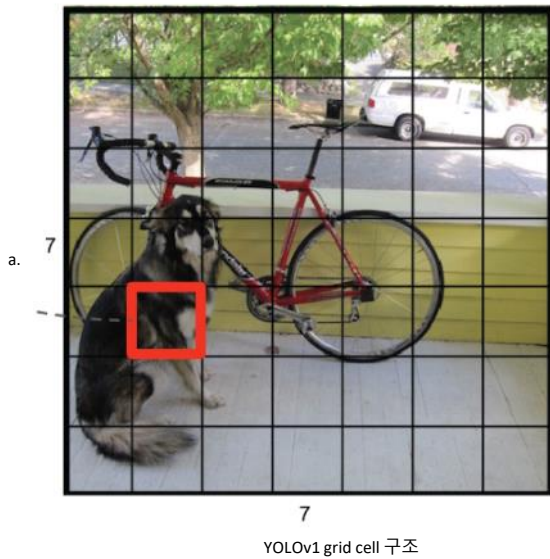


YOLOv1 Loss 의문점

2024년 3월 24일 일요일 오전 2:42



YOLOv1 prediction 구조를 보면 하나의 grid cell 당 1개의 클래스 확률, 2개의 bbox 좌표, 2개의 bbox에 대한 confidence를 추론한다. Loss function 계산 시 loss는 3가지로 구분되는데 classification loss는 1개인 반면 regression loss와 confidence loss는 2개이다.

Each grid cell predicts B bounding boxes and confidence scores for those boxes. These confidence scores reflect how confident the model is that the box contains an object and also how accurate it thinks the box is that it predicts. Formally we define confidence as $\Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}}$. If no object exists in that cell, the confidence scores should be zero. Otherwise we want the confidence score to equal the intersection over union (IOU) between the predicted box and the ground truth.

위의 내용은 YOLOv1 논문 내용을 발췌한 것으로 confidence score는 box 좌표의 정확성 그리고 해당 box 내부에 사물이 있을지에 대한 확신의 정도를 나타내며 $\Pr(\text{Object}) * \text{IOU}$ 로 정의한다.

Detection 모델들의 loss를 보면 여러 bbox에 대해 iou 함수를 사용하여 높은 iou 값을 가지는 bbox를 선택 후 해당 box에 대해 loss를 계산하는데(이 부분이 첫번째로 헷갈림, 다음에 이어질 내용들의 전제, 모든 detection 모델이 iou 적용 값에 대해 loss를 계산하는지 모르겠음) 아래 내용들에 대한 의문이 있음.

1. 논문 정의 상 confidence loss는 사물의 존재 여부와 iou 두가지에 영향을 받으며 iou는 결국 bbox 좌표를 얼마나 정확히 예측했는지를 나타낸다. 하지만 bbox에 대한 regression loss가 따로 있기 때문에 iou 연산이 필요한지 자체에 대해 의문이 있음.
2. 참고 코드에서는 regression loss 및 confidence loss에 하나의 박스만 선택 후 loss 연산에 참여시키는데 그러면 실제 학습은 박스 1개에 대해서만 이루어지기 때문에 학습 속도 저하가 예상됨.
3. 추론 과정에서 box를 2개 추론하는 실질적 의미가 없음. 실제로 box 여러 개를 추론해도 실질적으로 작동하는 box는 1개임. YOLOv1 모델 자체의 문제라고 생각함. 아래 그림들 참조

1. x - coordinate of bbox center inside cell ($[0; 1]$ wrt grid cell size)
2. y - coordinate of bbox center inside cell ($[0; 1]$ wrt grid cell size)
3. w - bbox width ($[0; 1]$ wrt image)
4. h - bbox height ($[0; 1]$ wrt image)
5. c - bbox confidence $\sim P(\text{obj in bbox2})$

YOLOv1 prediction 구조

Regression loss

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

Confidence loss

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2$$

Classification loss

$$+ \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

YOLOv1 prediction loss function

2개의 박스 중 하나를 선택하기 위한 인덱스
iou 수치가 아니며 regression에 수치 반영되지 않음

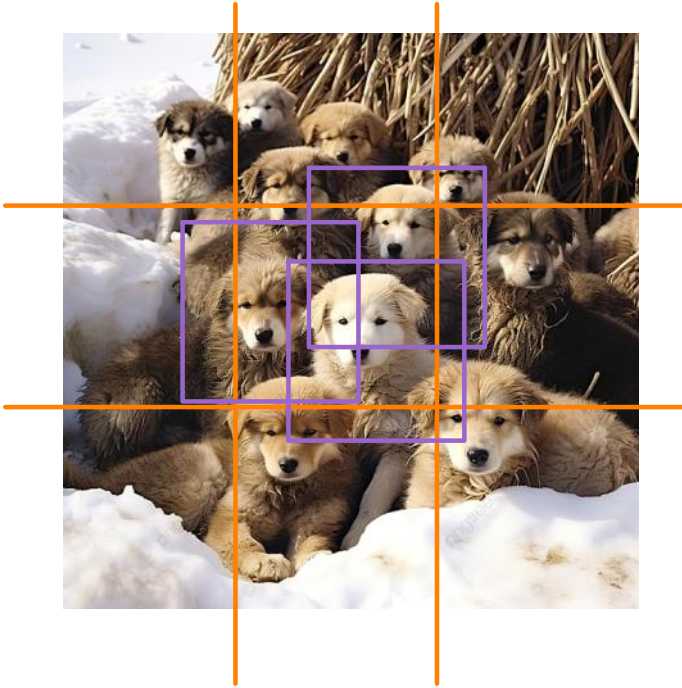
```
# confidence loss -----
chosen_confidence = (
    bestbox * predictions[..., 25:26] + (1 - bestbox) * predictions[..., 20:21]
)
target_confidence = targets[..., 20:21]

object_loss = self.mse(
    torch.flatten(exists_box * chosen_confidence),
    torch.flatten(exists_box * target_confidence),
)

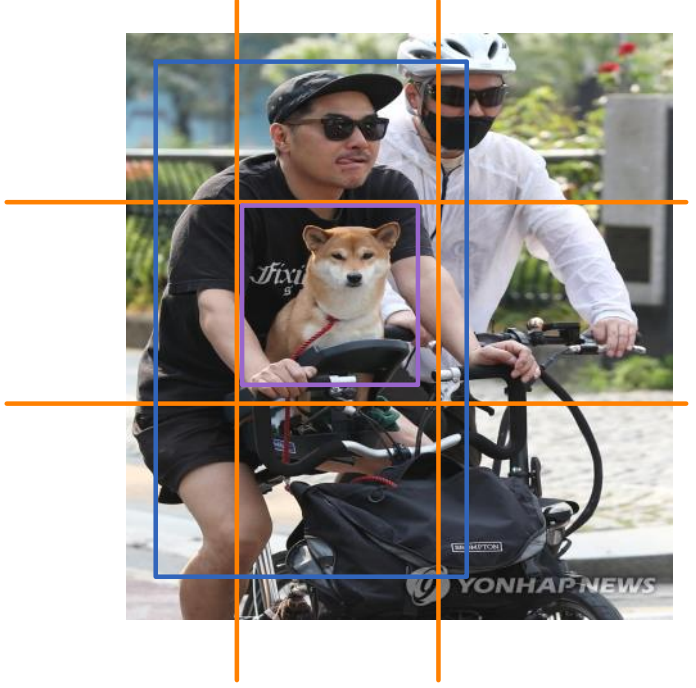
no_object_loss = self.lambda_noobj * self.mse(
    torch.flatten((1 - exists_box) * chosen_confidence),
    torch.flatten((1 - exists_box) * target_confidence),
)

confidence_loss = object_loss + no_object_loss
```

참고 코드



하나의 grid cell 안에 같은 class를 가지는 객체가 다수 존재
하지만 객체간 IOU가 높기 때문에 NMS 과정에서 여러 객체를 하나라고 인식할 가능성이 높음



하나의 grid cell 안에 다른 class를 가지는 객체가 존재
Grid cell 당 하나의 클래스만 추론 가능하므로 box loss와 confidence loss가 좋더라도 박스 2개중 1개는 오답



하나의 grid cell 안에 같은 두개의 가구 클래스가 존재함
두 객체 간 IOU 값이 크게 차이 나기 때문에 추론 결과에서 2가지 값이 전부 살아날 확률이 높음

수식 오류

Regression
loss

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

Confidence
loss

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2$$

Classification
loss

$$+ \sum_{i=0}^{S^2} \mathbb{I}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

1. Regression loss에서 xi, yi, wi, hi -> xij, yij, wij, hij
2. lobj는 두 bbox 중 어느 bbox가 ground truth와 iou가 높은지를 나타내는 마스크인데 confidence loss에 있는 lnoobj는 lobj의 반전이 아님