COMP 541, Project

8 November 2020 Sunday 18:41

Nov 9

Yolo v1

→ Reason globally on the entire image & all objects.

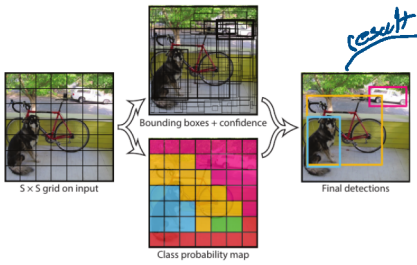
→ Divide img into $S \times S$ grid. Center of rectangles are responsible for the objects. Each cell preds BBOX & confidence.

↳ Confidence = $Pr(Obj) \times IoU^{pred}_{truth}$

↳ BBOX: x, y, w, h, conf
- x, y: center
- w, h: relative to whole img.

↳ Also, predict conditional class probs: $Pr(cls_i | obj)$ ↳ per grid cell

→ $Pr(cls_i | obj) \times Pr(obj) \times IoU^{pred}_{truth} = Pr(cls_i) \times IoU^{pred}_{truth}$



result → $S \times S \times (B \times 5 + C)$

↳ # of bbox it is predefined. why?

→ Reduction layers (1×1) after conv layers. (3×3)

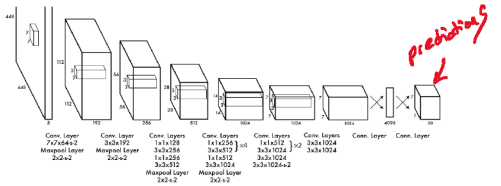


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection.

- Pretrained on ImageNet.

10

- Loss: sum-squared error

↳ penalizes localization error and cls. err equally

↳ to remedy loss diff. between small and large bboxes, predict sqrt of w, h.

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

↙ multipart loss

$\mathbb{1}_i^{obj}$: if there is obj in the cell

$\mathbb{1}_{ij}^{obj}$: j^{th} bbox predictor is responsible

Limitations

↳ struggles w/ small objects

↳ group of obj. close to each other like flocks of birds

↳ main source of errors is incorrect localizations

? R-CNN, region proposals, Selective Search

? Grasp Detection