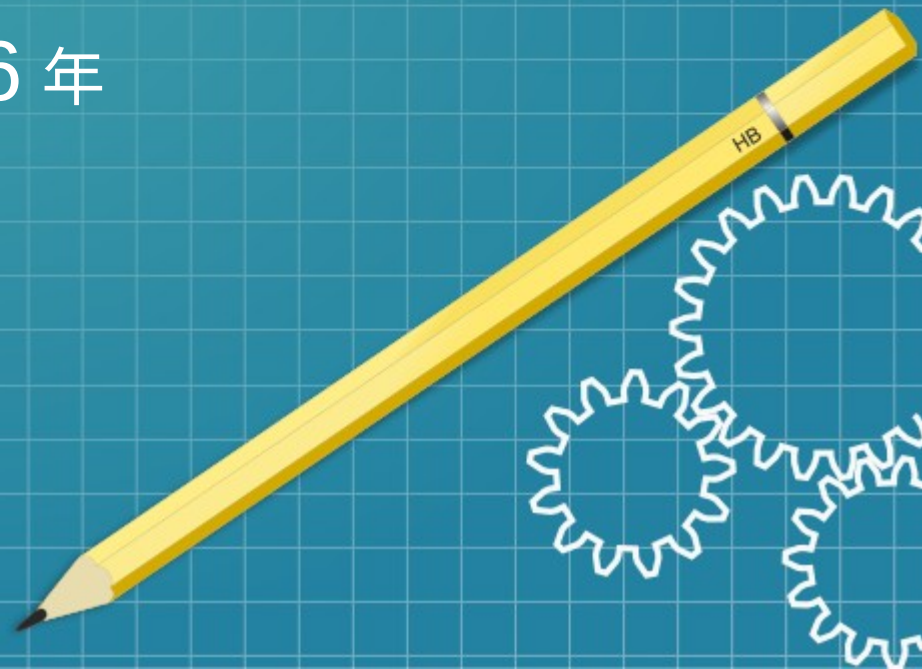




You Only Look Once: Unified, Real-Time Object Detection

Joseph Redmon * , Santosh Divvala *† , Ross Girshick ¶ , Ali Farhadi *†
University of Washington * , Allen Institute for AI † , Facebook AI Research ¶

出典：2016 年

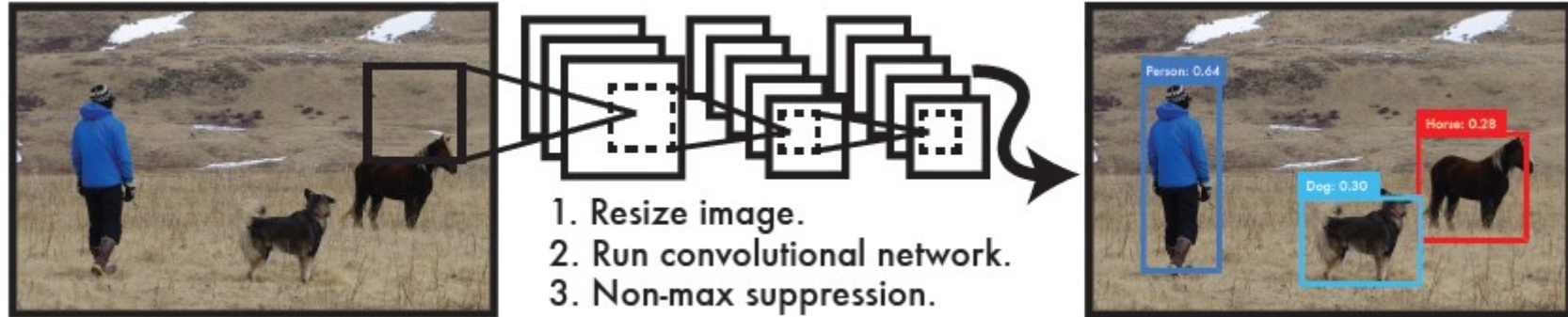


Abstract



- 今回の論文は YOLO(V1)
- 最新は YOLO V3
- 2016 年に発表され、Fast R-CNN と比較している
- 特徴
- 検出ネットワークが単一のため、シンプル
- 非常に高速 通常版 (15 FPS) 高速版 Titan X GPU(150 FPS)

Introduction

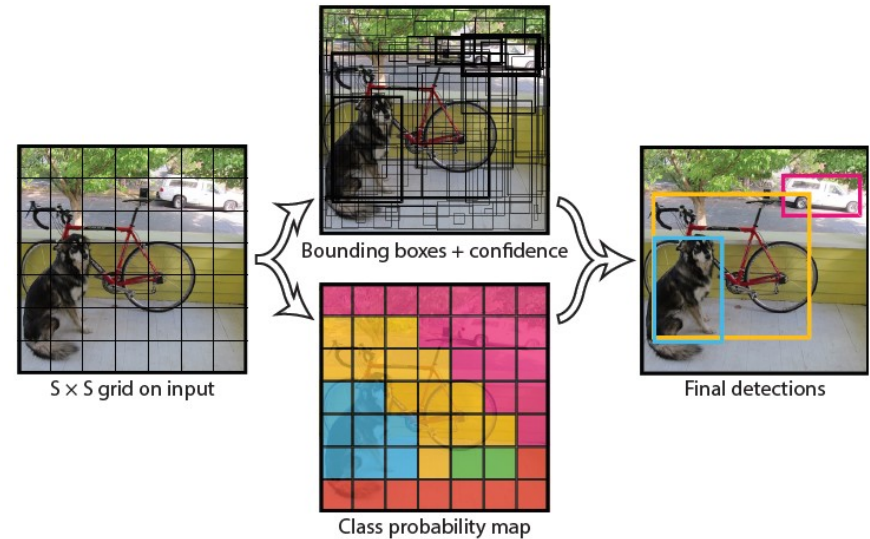


- YOLO は 3 ステップで処理ができる
 - (1) 入力画像を 448×448 にサイズ変更
 - (2) 画像上で単一の畳み込みネットワークを実行
 - (3) モデルの検出スコアによって検出結果を閾値を決める

単一の畳み込みネットワークにより、バウディングボックスおよびクラス確率を同時に予測

Unified Detection

- 回帰問題として検出をモデル化
- 1. 画像を $S \times S$ のグリッドに分割
- 2. B 各グリッドセルに対して、
バウンディングボックス提案
- 3. C ボックスに対する信頼性、クラス確率を予測
- これらの予測は以下のテンソルとして符号化される
 $S \times S \times (B * 5 + C) \text{ tensor}$.
- 評価には $S=7, B=2, C=20$
 $7 \times 7 \times (2 * 5 + 20) = 7 \times 7 \times 30 \text{ tensor}$.





- 非常に高速
- 複雑なパイプラインは必要

optimize

- 損失関数

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

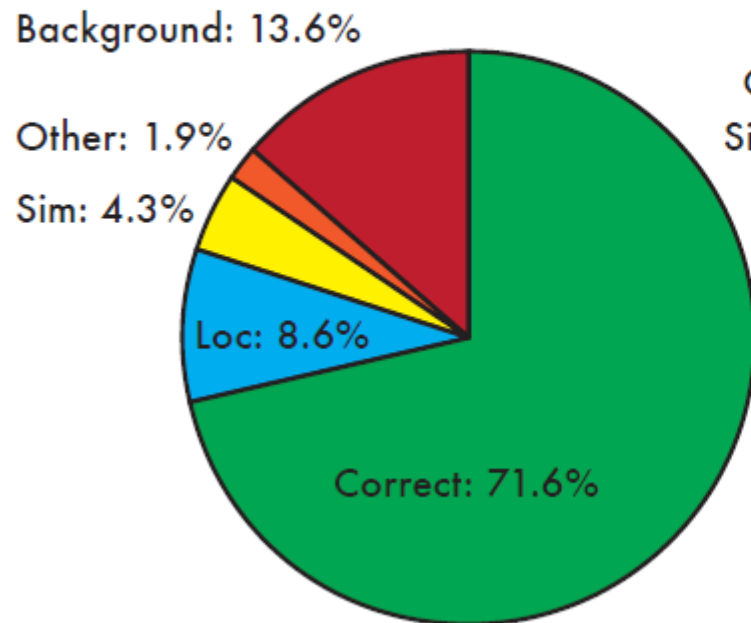
Limitations of YOLO



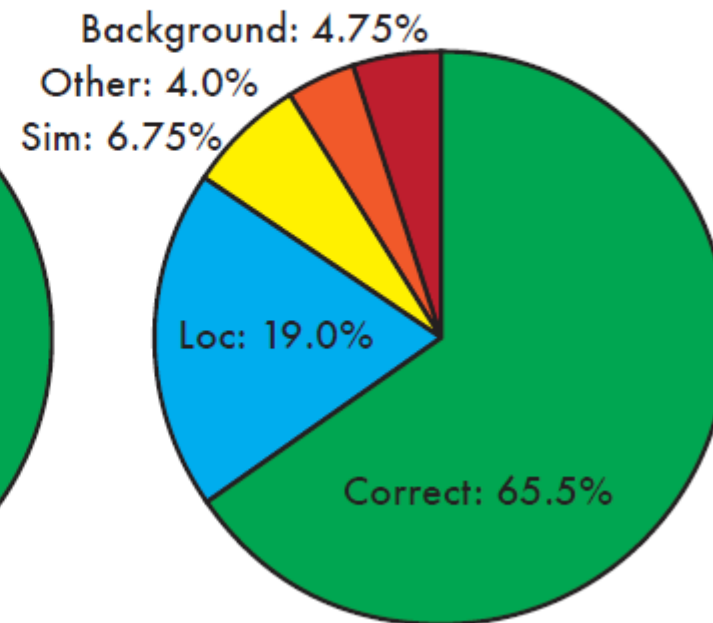
- YOLO の限界
- 各グリッドが二つのボックスのみを予測し、一つのクラスしか持ていない
- そのため、バウンディングボックスの予測に強い空間的制約を課す
- 鳥の群れのようにグループの小さな物体が苦手
- データからバウンディングボックスを予測することを学習している
- そのため、異常なアスペクト比または構成でオブジェクトに一般化するのに苦労

Fast R-CNN YOLO 比較

Fast R-CNN



YOLO



Correct: correct class and $\text{IOU} > 0.5$

Localization: correct class, $0.1 < \text{IOU} < 0.5$

Similar: class is similar, $\text{IOU} > 0.1$

Other: class is wrong, $\text{IOU} > 0.1$

Background: $\text{IOU} < 0.1$ for any object

YOLO の認識結果

