



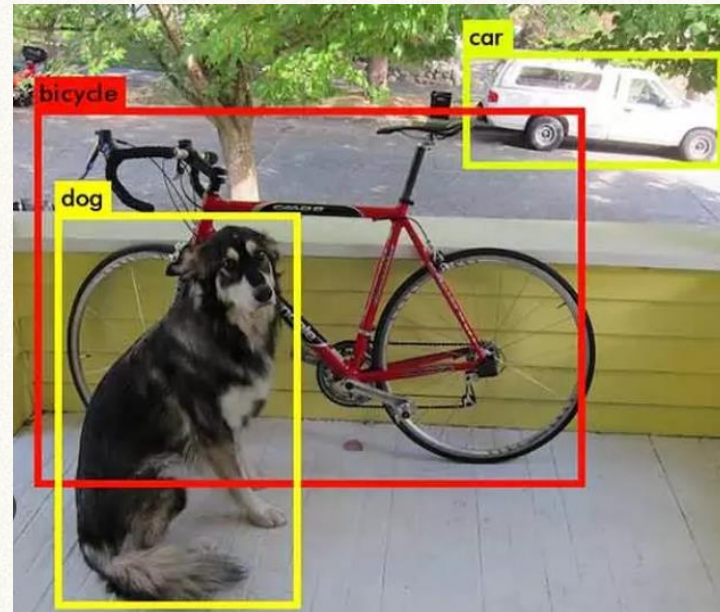
YOLO trong Object Detection

Huỳnh Thị Mỹ Thanh
Mssv: 19127273



YOLO là gì?

YOLO trong object detection có nghĩa là “You only look once”. Tức là chúng ta chỉ cần nhìn 1 lần là có thể phát hiện ra vật thể. YOLO không chỉ là dự báo nhãn cho vật thể như các bài toán classification mà nó còn xác định location của vật thể.

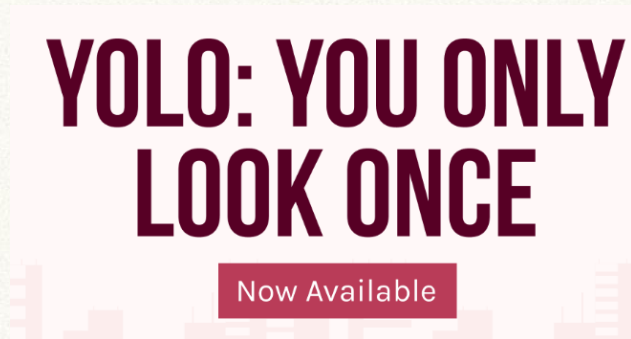




Sự ra đời và phát triển

2016

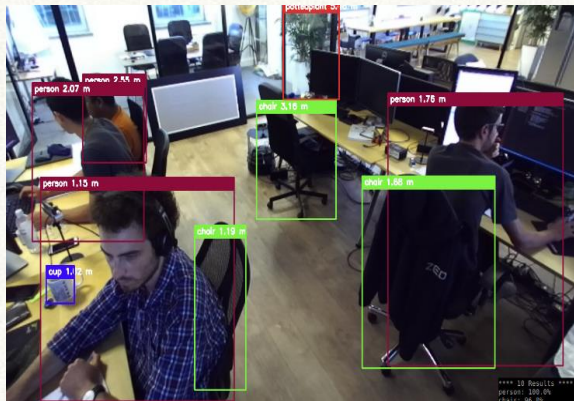
YOLO xuất hiện với tác giả chính là ông Joseph Redmon và đã tạo sự chú ý đáng kể từ cộng đồng nghiên cứu nhờ vào việc dễ sử dụng (end-to-end) và tốc độ chạy nhanh (real-time).



Sau đó YOLO đã phát triển thêm 2 version mạnh hơn là YOLO-v2 vào năm 2017 và YOLO-v3 vào năm 2018. Ngoài ra sau này còn có các nhà nghiên cứu khác cũng đã cải tiến và phát triển theo mục tiêu của các tác giả, và hiện chúng ta có: YOLO-v4, YOLO-v5, YOLO-v7 ...



Ứng dụng



Camera giám sát thông minh



Camera phát hiện sản phẩm cần thu hoạch





Ưu điểm

- Tốc độ xử lý nhanh: Tốc độ xử lý đáp ứng được thời gian thực, phát hiện đối tượng và xác định vị trí của chúng một cách nhanh chóng.
- Độ chính xác khá tốt.
- Dễ dàng triển khai mô hình.





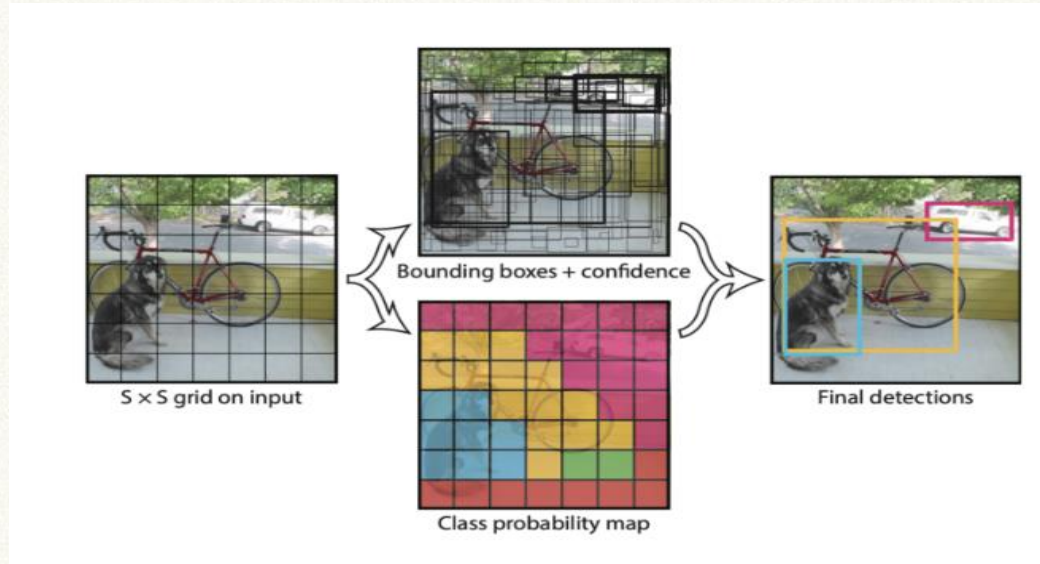
Thách thức

- Tuy YOLO có độ chính xác khá tốt nhưng nó vẫn chưa thể sánh bằng được với các phương pháp như Fast R-CNN.
- YOLO thường bỏ sót những đối tượng nhỏ trong hình ảnh (những đối tượng có kích thước rất nhỏ so với kích thước hình ảnh) và gặp khó khăn bởi việc phát hiện những đối tượng bị che khuất.
- Thời gian training khá dài.





Hệ thống detection của YOLO



Input: YOLO chia ảnh đầu vào thành kích thước $S \times S$ ô lưới.

Output: Là một ma trận 3 chiều có kích thước $S \times S \times (5 \times N + M)$ với số lượng tham số của mỗi ô là $(5 \times N + M)$ với N và M lần lượt là số lượng Box và Class mà mỗi ô cần dự đoán.





Mỗi ô dự đoán sẽ chứa 5 thông tin (x,y,w,h) và confidence, trong đó:

- (x,y) là tọa độ trọng tâm của đối tượng
- (w,h) chiều rộng, chiều cao của bounding box
- Confidence: độ tin cậy. Được tính bằng cách so sánh thông tin predict ra và thông tin được label. Công thức:

$$\Pr(\text{Class}_i | \text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}}$$

xác suất của mỗi class cho ô.

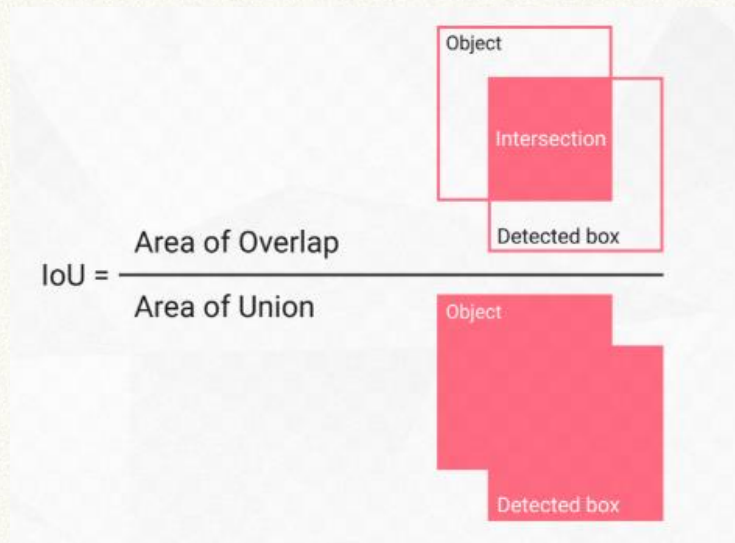
xác suất có vật thể trong ô.

phần giao nhau của
bounding box predict và
bounding box được label.





IoU

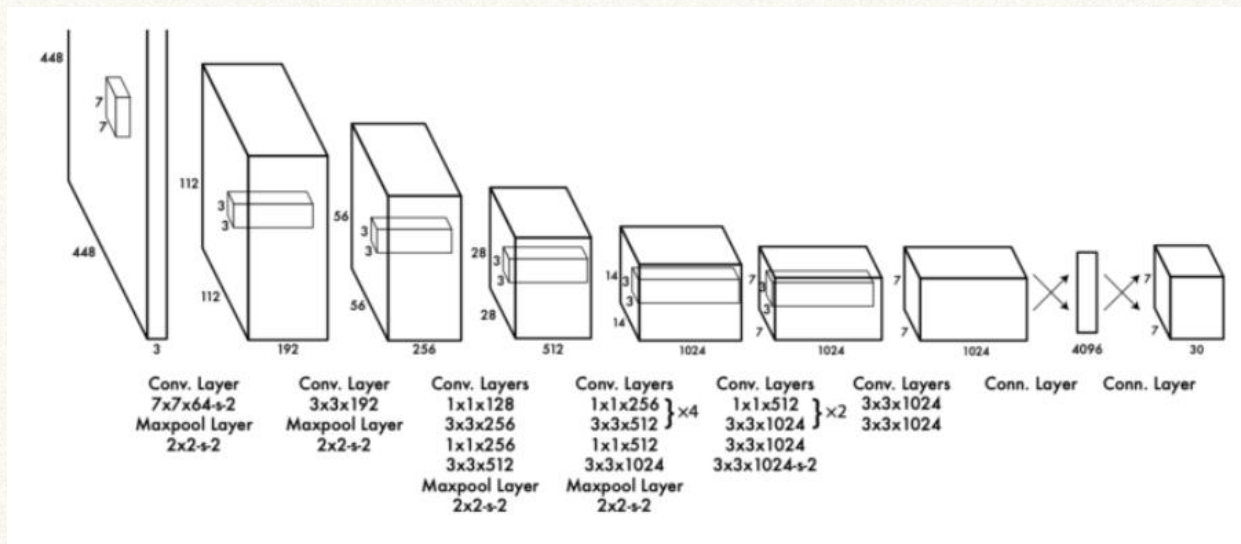


Nếu $\text{IoU} > 0.5$, detected box được đánh giá là tốt





Network architecture



Về cơ bản, YOLO chứa 3 loại layer: Convolutional, Maxpool, và Fully Connected.





LOSS FUNCTION

$$\begin{aligned} & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \quad \leftarrow \text{Loss của tâm (x,y) của box được predict và box} \\ & \quad + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \quad \leftarrow \text{Loss của chiều dài và rộng (h,w) của bbox được} \\ & \quad \quad + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \quad \leftarrow \text{Loss của confidence score khi có object trong ô} \\ & \quad + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \quad \leftarrow \text{Loss của confidence score khi không có} \\ & \quad \quad + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad \leftarrow \text{Loss của xác suất tại ô có object} \end{aligned}$$





Các version của YOLO:

Version	Tốc độ (FPS)	Độ chính xác (Map)	Số lượng lớp
v1	45	63.4	20
v2	78	76.8	20
v3	91	57.9 – 65.7	80
v4	65-75	43.5 – 52.3	80





Họ R-CNN trong bài toán Object Detection

	RCNN	YOLO
Phương pháp tiếp cận	Region proposal	End to end
Tốc độ xử lý	Chậm hơn	Nhanh hơn
Số lượng tham số	Nhiều	Ít
Độ chính xác	Cao hơn	Thấp hơn

RCNN và các biến thể của nó (Fast R-CNN, Faster R-CNN) được coi là phương pháp đạt hiệu suất tốt nhất trong các phương pháp dựa trên mạng neural. RCNN cung cấp một bước phát hiện đối tượng rất chính xác và có khả năng phát hiện các đối tượng nhỏ và phức tạp.





Kết luận

- ❖ YOLO là phương pháp giải quyết có tốc độ xử lý nhanh nhất hiện nay.
- ❖ Tuy nhiên độ chính xác của nó vẫn chưa thể so sánh được với các phương pháp khác như Faster RCNN
- ❖ Càng về sau này, các version của YOLO càng có những cải thiện đáng kể về độ chính xác
- ❖ Trong tương lai, hướng phát triển có thể là việc tăng độ chính xác để nâng cao hiệu quả.



**THANKS
FOR YOUR
WATCHING**

