
Introduction to Object Detection

Lesson of Content

1. What is Object Detection?
2. Key Concepts in Object Detection
3. Popular Object Detection Models
4. Evaluation Metrics for Object Detection
5. Challenges in Object Detection

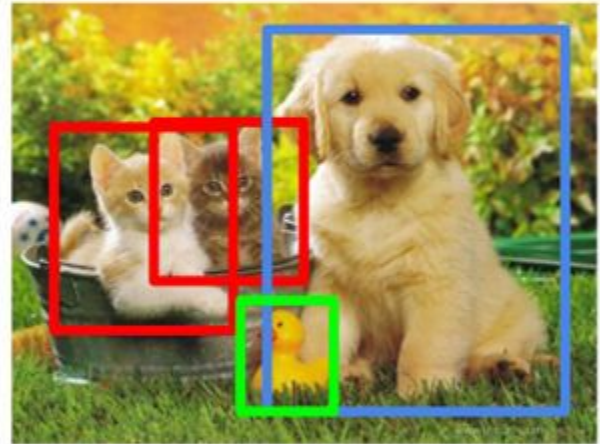
1. What is Object detection ?

Classification



CAT

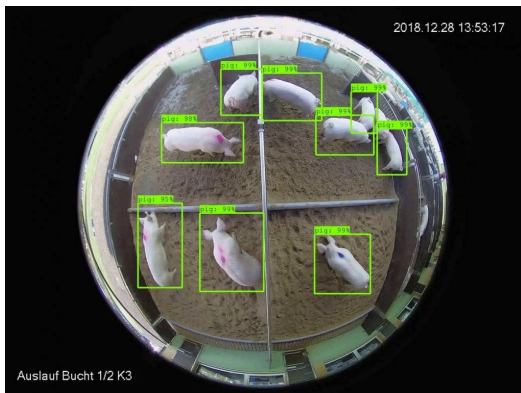
Object Detection



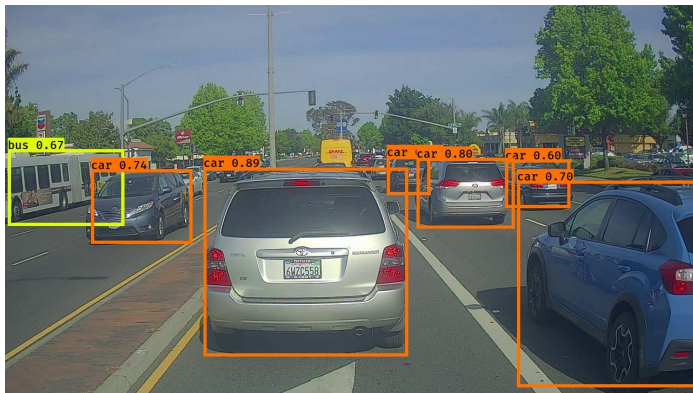
CAT, DOG, DUCK

1. What is Object detection ?

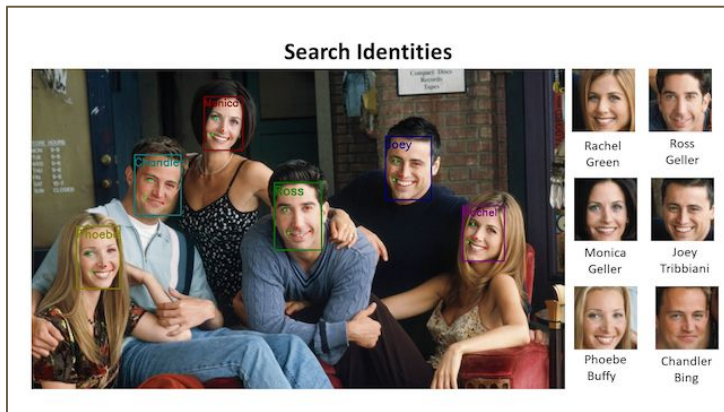
Some Application



Animal monitoring

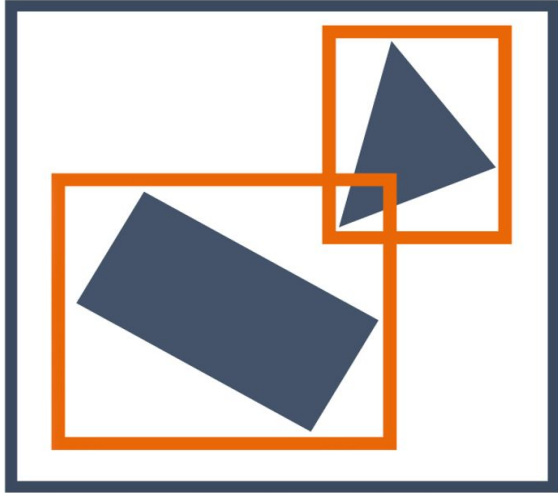


Semi-automatic driving



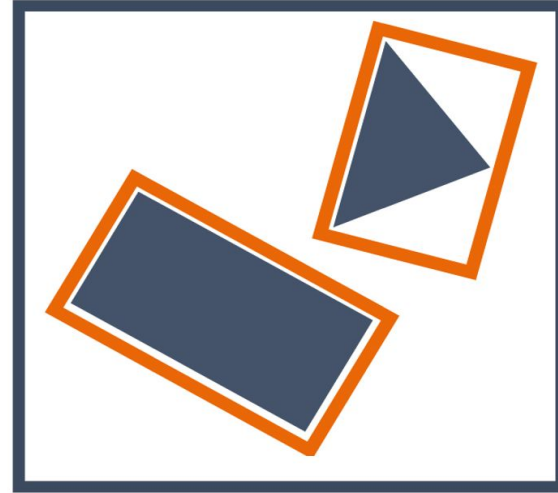
Face Recognition

Bounding boxes



Normal bounding boxes

[x1, y1, x2, y2]

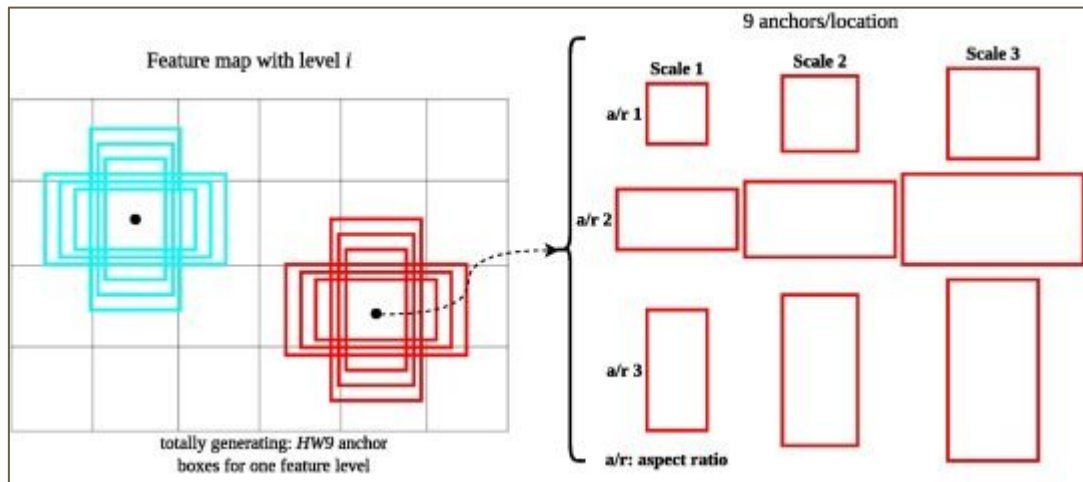


Oriented bounding boxes

[x1, y1, x2, y2, x3, y3, x4, y4]

Anchor Boxes

Anchor boxes là những boundingbox có **tỉ lệ xác định**. Trong quá trình huấn luyện **tỉ lệ này sẽ được canh chỉnh** sao cho phù hợp với các đối tượng thực tế



Việc khởi tạo các anchor boxes cho biết tỉ lệ, nhưng **không có kích thước chính xác**.

Thường sẽ dùng thêm **K-Means** để khởi tạo các kích thước này dựa trên dữ liệu huấn luyện. (Nó sẽ là các kích thước phổ biến).

=> Giúp có sự ước lượng ban đầu tốt.

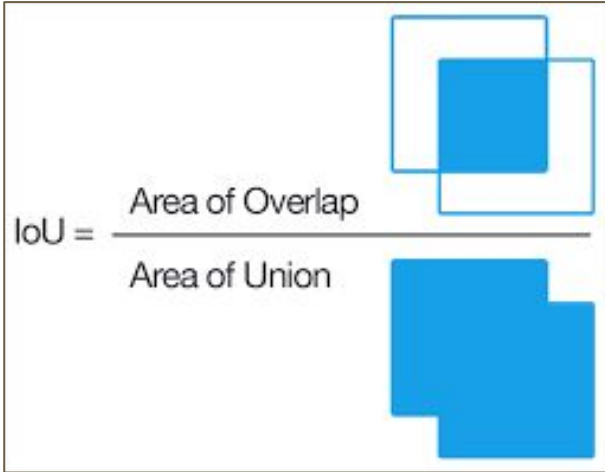
=> Giảm mất mát trong hàm loss, giúp tăng tốc quá trình huấn luyện.

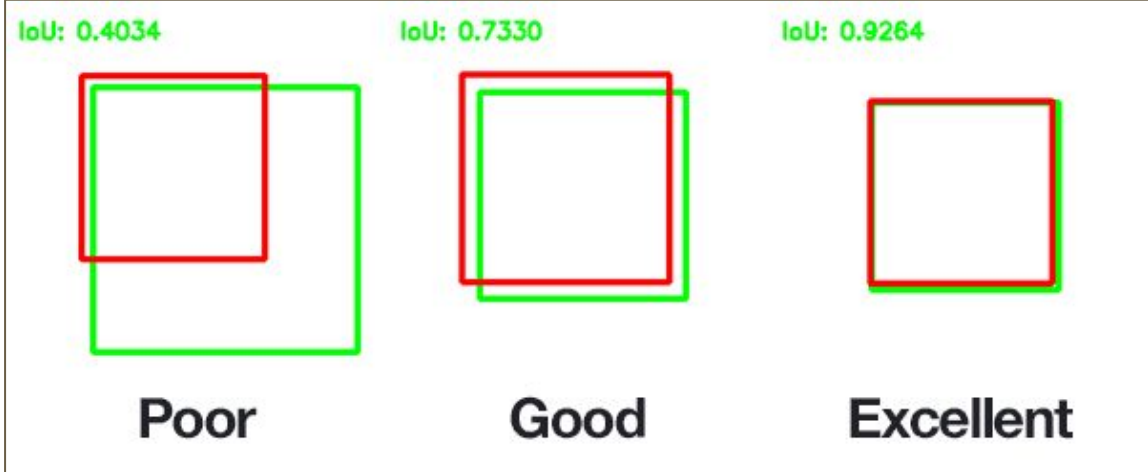
Example: Yolov3 có 3 ngõ ra lần lượt 13x13, 26x26, 52x52 grid cells.

=> Số lượng anchor boxes là: $(13 \times 13 + 26 \times 26 + 52 \times 52) \times 3 = 10.647$ anchors

Intersection over Union (IoU)

IOU là phép đo để xem xét độ chồng lấn của hai bboxes.


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



Dùng trong tính **mAP** (**mean Average Precision**) và **NMS** (**Non-Max Suppression**).

Confidence Score

Confidence score thể hiện mức độ chắc chắn dự đoán của mô hình

$$\text{Confidence Score} = P(\text{Object}) \times \text{IoU}(\text{predicted_box}, \text{true_box})$$

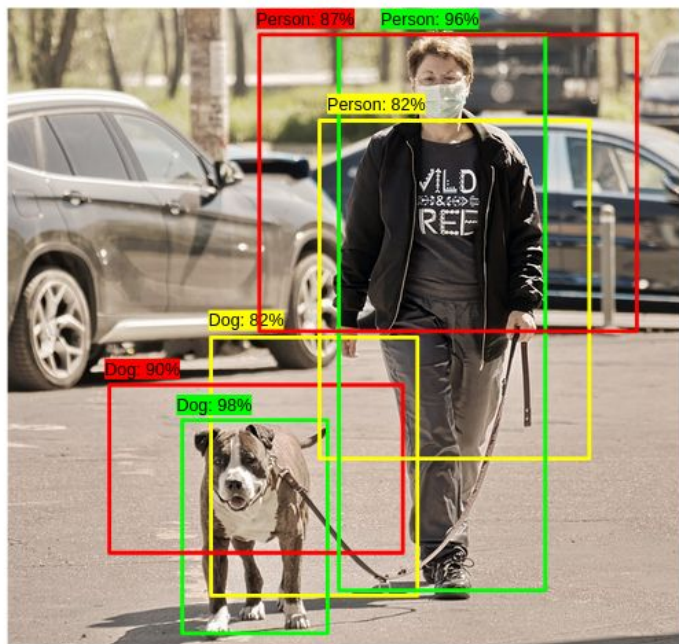
- **Pr(Object)**: Xác suất có một đối tượng trong boundingbox
- **IoU (predicted_box, ground_truth)**: Độ chồng lấn của bbox dự đoán và bbox thực tế (ground truth)

Lưu ý:

- Chỉ áp dụng công thức cho huấn luyện mô hình
- Khi inference : Confidence Score = P(object)

Non-Max Suppression (NMS)

NMS như một bước hậu xử lý sau dự đoán của mô hình. Giúp loại bỏ các boundingbox của cùng một đối tượng nhưng có độ tin cậy không cao.



B1: Sắp xếp Bounding Boxes: Sắp xếp các bboxes dựa theo điểm tin cậy (confidence score) từ cao đến thấp.

B2: Lấy Bounding boxes có điểm cao nhất: Lấy bounding box có score cao nhất là dự đoán cuối cùng.

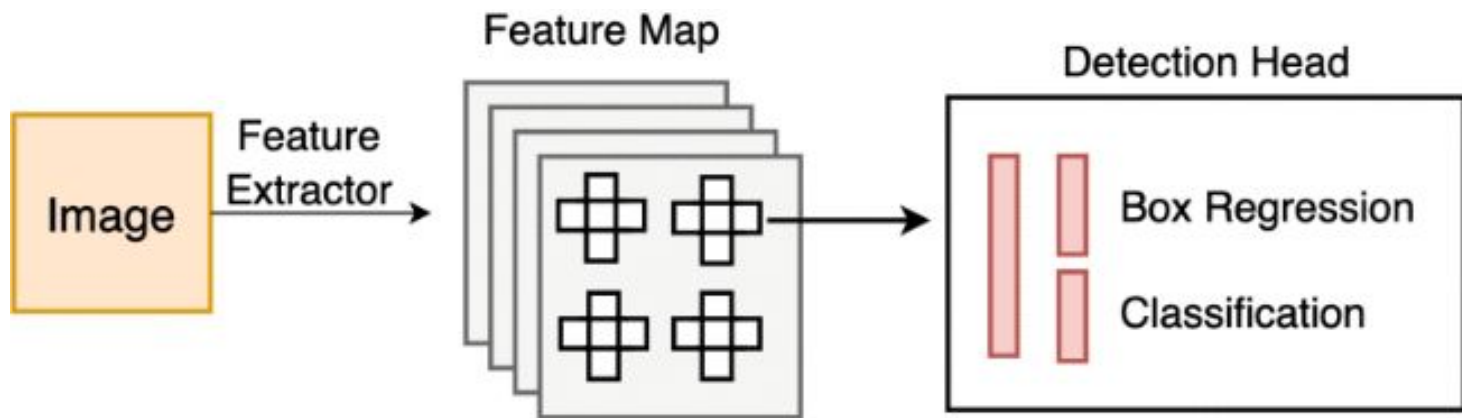
B3: So sánh IoU: Tính IoU giữa bboxes vừa chọn với các boundingbox còn lại trong danh sách.

B4: Loại bỏ các Bounding boxes trùng lặp: Loại bỏ các bboxes có IoU cao hơn một ngưỡng thường là 0.5 với box có confidence cao nhất.

B5: Lặp lại: Quay lại với bước 2.

One Stage Object Detection

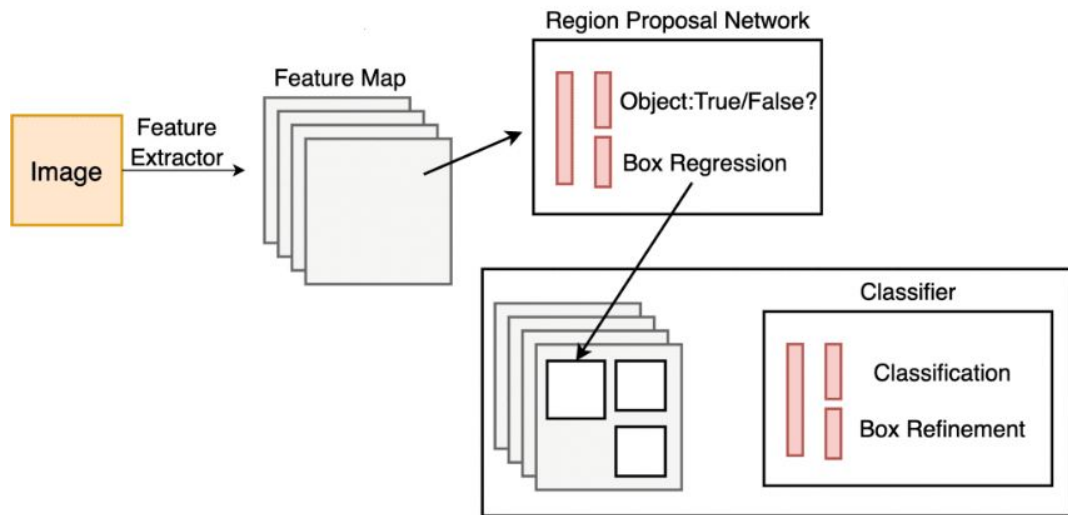
Là mô hình một giai đoạn thực hiện việc phát hiện và phân loại đối tượng. Thường áp dụng trong các ứng dụng cần xử lý thời gian thực.



Một số mô hình : Yolo, SSD.

Two Stage Object Detection

Mô hình gồm hai giai đoạn chủ yếu tập trung vào việc cải thiện được độ chính xác. Bao gồm: Tạo ra các khu vực đề xuất (region proposals) mà có thể chứa đối tượng, sau đó phân loại và hiệu chỉnh các đề xuất này.



Một số mô hình : R-CNN, Fast R-CNN, Faster R-CNN

4. Evaluation Metrics for Object Detection

	backbone	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
<i>Two-stage methods</i>							
Faster R-CNN+++ [3]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [6]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [4]	Inception-ResNet-v2 [19]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [18]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
<i>One-stage methods</i>							
YOLOv2 [13]	DarkNet-19 [13]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [9, 2]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [2]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [7]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet [7]	ResNeXt-101-FPN	40.8	61.1	44.1	24.1	44.2	51.2
YOLOv3 608 × 608	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9

AP^{small}

% AP for small objects: area < 32²

AP^{medium}

% AP for medium objects: 32² < area < 96²

AP^{large}

% AP for large objects: area > 96²

AP₅₀: Là chỉ số của AP khi dùng ngưỡng IOU = 0.5. Có nghĩa những đối tượng như đoán đúng class và có IOU >= 0.5 thì xem là một dự đoán chính xác, và ngược lại.

AP₇₅: Tương tự

AP S/M/L (Small, medium, Large): Chỉ xét những đối tượng theo điều kiện nhỏ, vừa, lớn

AP (Average Precision)

Chỉ số để đánh giá hiệu suất của mô hình phát hiện đối tượng. AP tính trung bình của precision ở tất cả các mức recall, từ 0 đến 1.

$$AP = \sum_{k=0}^{k=n-1} [Recalls(k) - Recalls(k + 1)] * Precisions(k)$$

$Recalls(n) = 0, Precisions(n) = 1$
 $n = \text{Number of thresholds.}$

AP (Average Precision)

Thứ tự	Tỷ tin	Đúng/Sai

1	0.98	Đúng
2	0.95	Đúng
3	0.90	Sai
4	0.87	Đúng
5	0.85	Sai
6	0.80	Đúng
7	0.75	Đúng
8	0.70	Sai
9	0.60	Đúng
10	0.55	Sai



Thứ tự	Recall	Precision

1	0.1	1.0
2	0.2	1.0
3	0.2	0.67
4	0.3	0.75
5	0.3	0.6
6	0.4	0.67
7	0.5	0.71
8	0.5	0.625
9	0.6	0.67
10	0.6	0.6

Recall: Số lượng dự đoán đúng tích lũy

Precision: Số lượng dự đoán đúng trên tổng số dự đoán.

Ví dụ:

idx = 5

recall = $3/10 = 0.3$

precision = $\frac{3}{5} = 0.6$

AP (Average Precision)

Thứ tự	Recall	Precision
1	0.1	1.0
2	0.2	1.0
3	0.2	0.67
4	0.3	0.75
5	0.3	0.6
6	0.4	0.67
7	0.5	0.71
8	0.5	0.625
9	0.6	0.67
10	0.6	0.6

$$AP = \sum_i (R_i - R_{i-1}) P_i$$

$$\begin{aligned}
 AP &= (0.1 - 0) * 1.0 + \\
 &\quad (0.2 - 0.1) * 1.0 + \\
 &\quad (0.3 - 0.2) * 0.75 + \\
 &\quad (0.4 - 0.3) * 0.67 + \\
 &\quad (0.5 - 0.4) * 0.71 + \\
 &\quad (0.6 - 0.5) * 0.67 + \\
 &\quad (0.6 - 0.6) * 0.6 = \\
 &0.1 + 0.1 + 0.075 + 0.067 + 0.071 + 0.067 = 0.48
 \end{aligned}$$

mAP (mean Average Precision)

Là trung bình AP trên tất cả các lớp object. Đây là thước đo tổng quan nhất cho bài toán object detection cho nhiều lớp.

Mean Average Precision Formula

$$\text{Mean Average Precision} = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

n = the number of classes

AP_k = the average precision of class k

Cáo: $AP = 0.87$

Mèo: $AP = 0.91$

Hổ: $AP = 0.85$

Sư tử: $AP = 0.89$

Chó: $AP = 0.88$

$$mAP = \frac{AP_{cáo} + AP_{mèo} + AP_{hổ} + AP_{sư tử} + AP_{chó}}{5}$$

$$mAP = \frac{0.87 + 0.91 + 0.85 + 0.89 + 0.88}{5} = 0.88$$

mAP càng cao càng tốt, thường một mô hình gọi là ổn khi $mAP \geq 0.75$.

5. Challenges in Object Detection

- **Variability in size and Scale:** Các đối tượng trong một hình ảnh có thể xuất hiện ở nhiều kích thước, tỷ lệ với góc độ khác nhau, làm cho việc phát hiện trở nên phức tạp.
- **Overlap and Occlusion:** Khi một hoặc nhiều đối tượng bị che khuất hoặc chồng chéo lên nhau, việc phát hiện và phân loại trở nên khó khăn.
- **Detection of Low-Resolution Objects:** Đối với các đối tượng ở xa hoặc ở độ phân giải thấp trong ảnh, việc phát hiện và phân loại cũng khó khăn.
- **Real-time and Computational Efficiency:** Các bài toán cho xe tự lái hoặc giám sát thì việc phát hiện đối tượng phải là real-time. Đòi hỏi cân bằng giữa độ chính xác và tốc độ xử lý.
- **Data Imbalance:** Trong một số tập dữ liệu, một số lớp có thể có nhiều mẫu, nhưng một số lớp khác chỉ có một lượng mẫu nhỏ. Điều này là cho mô hình thiên vị về những lớp có nhiều mẫu hơn.