

### 3. Kết quả

#### 3.1. Phát hiện khuôn mặt

Trong phần huấn luyện mô hình phát hiện khuôn mặt, nhóm sử dụng tập dữ liệu Wider Face có sẵn. Thông tin về tập dữ liệu:

Nguồn: <http://shuoyang1213.me/WIDERFACE/>

Mô tả: Tập dữ liệu Wider Face là tập dữ liệu chuẩn cho phát hiện khuôn mặt, trong đó hình ảnh được chọn từ tập dữ liệu Wider công khai. Tập này có 32.203 hình

ảnh và gán nhãn 393.703 khuôn mặt có mức độ thay đổi cao về tỷ lệ, tư thế. Bộ dữ liệu

chứa các hình ảnh về 61 sự kiện khác nhau. Trong đó, tập dữ liệu đã được chia sẵn thành:

Train (12880 hình ảnh), Val (3226 hình ảnh), Test (16.097 hình ảnh).

•Độ phân giải: Đa độ phân giải

•Kích thước: Đa kích thước(chủ yếu 640x640)

Phân chia dữ liệu: Train: 2.210 (68,84%), Val: 500(15,58%), Test: 500 (15.58%).

#### 3.1.2. Huấn luyện

Sau khi có được dữ liệu và phân chia dữ liệu theo tỉ lệ phù hợp.Nhóm tiến hành chia dữ liệu thành images và labels để dễ dàng tiến hành train.Nhóm sử dụng mô hình YOLOv7, trong đó chứa các pre-trained model và các kiến trúc mạng. Từ đây, nhóm chỉnh sửa kiến trúc mạng sử dụng cho YOLOv7, dùng yolov7.pt và khai báo một file yami để yolov7 biết

Đường dẫn đến thư mục train, test, số lượng class qua biến nc ,tên của các các class để đi huấn luyện trên Google Colab. Sau khi huấn luyện đã cho ra mô hình phát hiện khuôn mặt. Sau khi huấn luyện nhóm sử dụng AP để tiến hành đánh giá hiệu suất mô hình.

#### 3.1.3. Kết quả giải pháp phát hiện khuôn mặt

Kết quả định tính

- Khi camera có độ phân giải thấp, rất khó phát hiện ở khoảng cách lớn.
- Vẫn còn phát hiện nhầm từ không có khuôn mặt thành có khuôn mặt.
- Trong phạm vi lớp học có thể phát hiện tất cả khuôn mặt nhưng trong phạm vi rộng lớn như sân vận động,... chỉ phát hiện được 50% tổng số khuôn mặt.

Kết quả định lượng

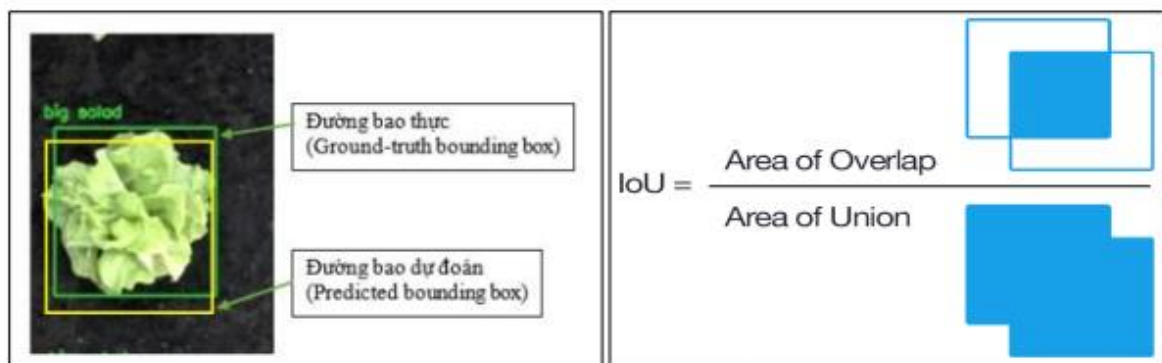
Để tiến hành đánh giá hiệu suất mô hình, nhóm sử dụng mAP (mean Average Precision). mAP là độ đo được sử dụng rất phổ biến cho các bài toán Object Detection.

Tuy nhiên, trong mAP còn có nhiều khái niệm khác về Precision, Recall, IOU. IOU (Intersection over union)

IOU là chỉ số đánh giá được sử dụng để đo độ chính xác của phát hiện đối tượng dựa trên tập dữ liệu cụ thể. Để áp dụng được IOU để đánh giá, cần một đường bao thực

là đường bao mà chúng ta gán cho vật thể và một đường bao dự đoán là đường bao dung model sau khi huấn luyện để nhận dạng. Dựa trên hai đường bao này, ta đo được mức độ giao nhau giữa hai đường bao để xác định khung hình có bị đè chồng lên nhau không.

Tỉ lệ này được tính dựa trên phần diện tích giao nhau giữa hai đường bao với phần diện tích không giao nhau giữa chúng.



Hình 22: Đường bao thực, dự đoán

Hình 23: Công thức tính IOU

Độ đo

Precision x Recall curve

Precision x Recall curve là cách tốt nhất đánh giá hiệu suất của một bộ phát hiện khuôn mặt với độ tự tin được thay đổi theo đồ thị đường cong cho mỗi lớp. Một bộ phát hiện khuôn mặt được coi là tốt nếu precision của nó vẫn cao khi recall cao, hoặc là khi thay đổi ngưỡng thì precision và recall vẫn cao.

Precision: Là độ đo đánh giá độ tin cậy của kết luận đưa ra (bao nhiêu % lời kết luận của model là chính xác).

Recall: Là độ đo đánh giá khả năng tìm kiếm toàn bộ các ground truth của mô hình (bao nhiêu % positive sample mà model nhận diện được).

True Positive (TP): Một kết quả phát hiện đúng ( $IOU \geq \text{ngưỡng}$ ).

False Positive (FP): Một kết quả phát hiện sai (IOU < ngưỡng).

False Negative (FN): Một khuôn mặt nhưng không phát hiện ra.

True Negative (TN): Một kết quả không phải là khuôn mặt nhưng phát hiện là khuôn mặt.

$Precision = \frac{TP}{TP + FP}$	$TP = \text{True positive}$
	$TN = \text{True negative}$
$Recall = \frac{TP}{TP + FN}$	$FP = \text{False positive}$
	$FN = \text{False negative}$

Hình 24: Công thức tính Precision và Recall

### Average Precision (AP)

Cách khác để so sánh hiệu suất của các bộ phát hiện khuôn mặt là tính diện tích của đồ thị đường cong (UAC) của Precision x Recall curve. Đường cong AP thường là đường zizac lên rồi xuống. Trong thực tế, AP là độ chính xác được tính trung bình trên tất cả các giá trị recall giữa 0 và 1.

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

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

Hình	Số khuôn mặt	Số khuôn mặt phát hiện	Tốc độ phát hiện tất cả khuôn mặt (s)
1	1	1	0.8
2	5	5	0.8
3	13	13	0.92
4	33	33	0.83
5	101	49	0.95