

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



UIT

TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

MÔN HỌC: MÁY HỌC - CS114

BÁO CÁO ĐỒ ÁN

Helmets Of Motorcyclists And Bicycle Riders Detection

Giảng viên hướng dẫn: Lê Đình Duy - Phạm Nguyễn Trường An

Nhóm:

Nguyễn Hải Đăng - 21521920

Trương Hữu Thọ - 21521479

TP Hồ Chí Minh, Ngày 23 tháng 1 năm 2024

TÓM TẮT

Hiện nay, giao thông ngày càng tăng cường, an toàn giao thông trở thành một mối quan tâm hàng đầu. Tuy nhiên vẫn có nhiều người khi tham gia giao thông bằng các phương tiện như xe máy, xe đạp,... lại không đội mũ bảo hiểm, dẫn đến những chấn thương nghiêm trọng khi tham gia giao thông. Để góp phần phát hiện cũng như khắc phục những trường hợp đó, nhóm em đã sử dụng mô hình YOLOv8 và train dựa vào dữ liệu mà nhóm thu thập được. Từ đó có thể phát hiện được người không đội mũ bảo hiểm trong bối cảnh thực tiễn.

Github <https://github.com/NHDang273/CS114.O11.KHCL-21521920>

Mục lục

1	CHƯƠNG 0: UPDATE SAU KHI VẤN ĐÁP	1
1.1	Input và Output của bài toán:	1
1.2	INTERSECTION OVER UNION (IOU) là gì?	2
2	CHƯƠNG 1: Giới thiệu	3
2.1	Bối cảnh	3
2.2	Mục tiêu	4
3	CHƯƠNG 2: Tập dữ liệu	6
4	CHƯƠNG 3: Phương pháp:	7
4.1	What	7
4.2	Why	7
4.3	When	7
4.4	Where	8
4.5	Who	8
4.6	How	8
4.7	Phương pháp đánh giá	9
4.8	Sử dụng phương pháp YOLO	10
5	CHƯƠNG 4: Kết quả	13
6	CHƯƠNG 4: Kết luận	13

1 CHƯƠNG 0: UPDATE SAU KHI VẤN ĐÁP

1.1 INPUT VÀ OUTPUT CỦA BÀI TOÁN:

- Input: Một ảnh có màu có một hoặc nhiều người tham gia phương tiện giao thông (xe đạp, xe máy, ô tô,...) có đội hoặc không đội mũ bảo hiểm.
- Output: Một danh sách các véc tơ bao gồm những thông tin của bounding boxes như sau:
 - Chỉ mục lớp: Chỉ số của lớp đối tượng được phát hiện trong danh sách lớp. (cụ thể trong đây là: 0-rider, 1-helmet, 2-no-helmet.
 - Độ tin cậy: Một số xác suất từ 0 đến 1 đo lường độ chắc chắn của mô hình rằng đối tượng được nhận diện chính xác. Nếu xác suất này thấp, có thể là một dự đoán không chắc chắn.
 - x,y,w,h: Tọa độ của bounding box trong hình ảnh, trong đó: x,y là tọa độ tâm; w,h: là kích thước rộng, dài của bounding box.
- Các thông tin này được biểu diễn dưới dạng một vector có 5 phần tử, theo thứ tự sau:

[class_index, confidence, x, y, w, h]

- Ví dụ, mô hình phát hiện một người đi xe trong hình ảnh, đầu ra có thể là như sau: [0, 0.08832, 240.5, 530.5, 97, 235]
 - 0: class_index là 0, tương ứng với lớp người đi xe (rider).

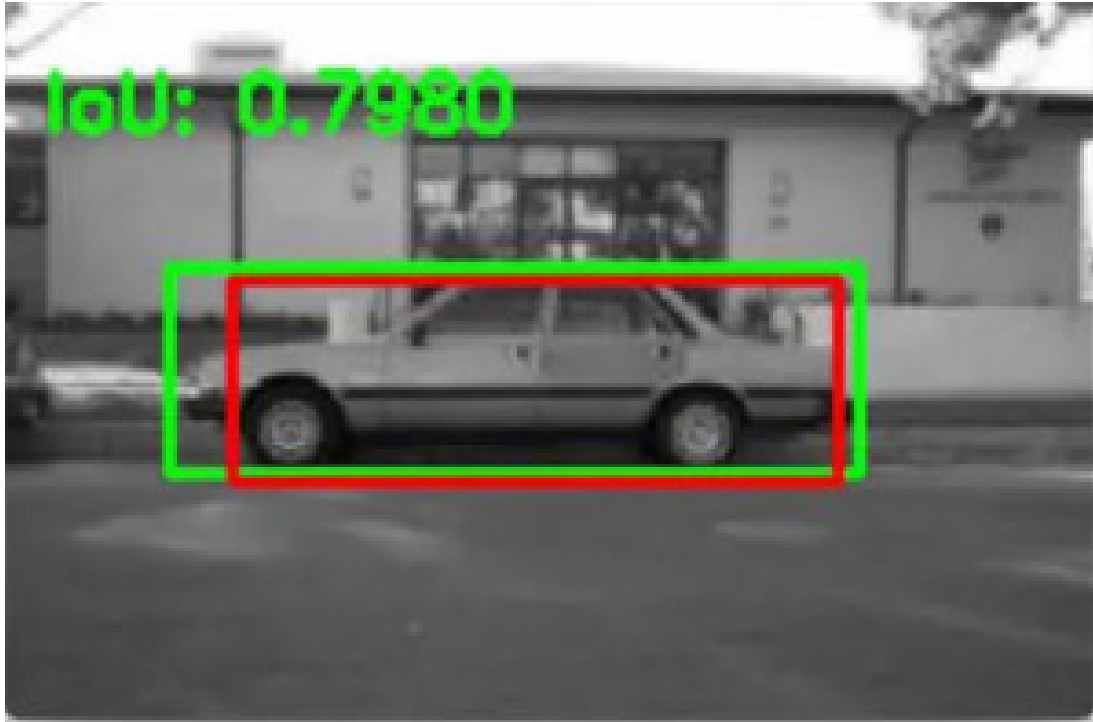
- 0.8832: Confidence là 0.8832, cho thấy rằng mô hình khá chắc chắn về việc phát hiện đối tượng này.
- 240.6,530.5: tức là tọa độ tâm của đối tượng này.
- 97, 255: bounding box có chiều rộng 97 và chiều cao là 255 (đơn vị là pixel).

1.2 INTERSECTION OVER UNION (IOU) LÀ GÌ?

- **Intersection over Union** là chỉ số đánh giá được sử dụng để đo độ chính xác của Object detector trên tập dữ liệu cụ thể. Chúng ta hay gặp chỉ số này trong các Object Detection Challenge, dạng như PASCAL VOC challenge.
- Để áp dụng được IoU để đánh giá một object detector bất kì ta cần:
 - Những ground-truth bounding box (bounding box đúng của đối tượng, ví dụ như bounding box của đối tượng được khoanh vùng và đánh nhãn bằng tay sử dụng trong tập test).
 - Những predicted bounding box được model sinh ra.
 - Đánh giá xem bounding box dự đoán đối tượng khớp với ground truth thật của đối tượng hay không.
 - Chỉ số IoU trong khoảng $[0,1]$ -> IoU càng gần 1 thì bounding box dự đoán càng gần ground truth.
 - Cách tính:

$$\text{IOU} = \frac{\text{area_of_overlap}}{\text{area_of_union}}$$

- * Nếu có điểm số $\text{IoU} > 0.5$, prediction sẽ được coi là một "good" prediction
- * Nếu có điểm số $\text{IoU} < 0.5$, prediction sẽ được coi là một "poor" prediction.



Hình 1: Hình minh họa cách tính IOU

2 CHƯƠNG 1: GIỚI THIỆU

2.1 BỐI CẢNH

- Nón bảo hiểm là một vật dụng nhằm mục đích bảo vệ phần đầu của người đội khi có va đập lúc đi xe đạp, xe máy, giúp đảm bảo an và toàn và ngăn chặn sự ảnh hưởng đến não bộ và hệ thần kinh khi va chạm.
- Tai nạn liên quan đến người điều khiển xe máy và xe đạp, đặc biệt là

khi họ không đội mũ bảo hiểm, đang là nguyên nhân phổ biến khiến họ bị tổn thương nhiều hơn, thậm chí có thể gây tử vong.

- Đồ án này được làm ra với mục tiêu giải quyết vấn đề này bằng cách sử dụng kỹ thuật Machine Learning để tự động nhận dạng việc đội mũ bảo hiểm của người điều khiển xe máy, xe đạp khi tham gia giao thông.

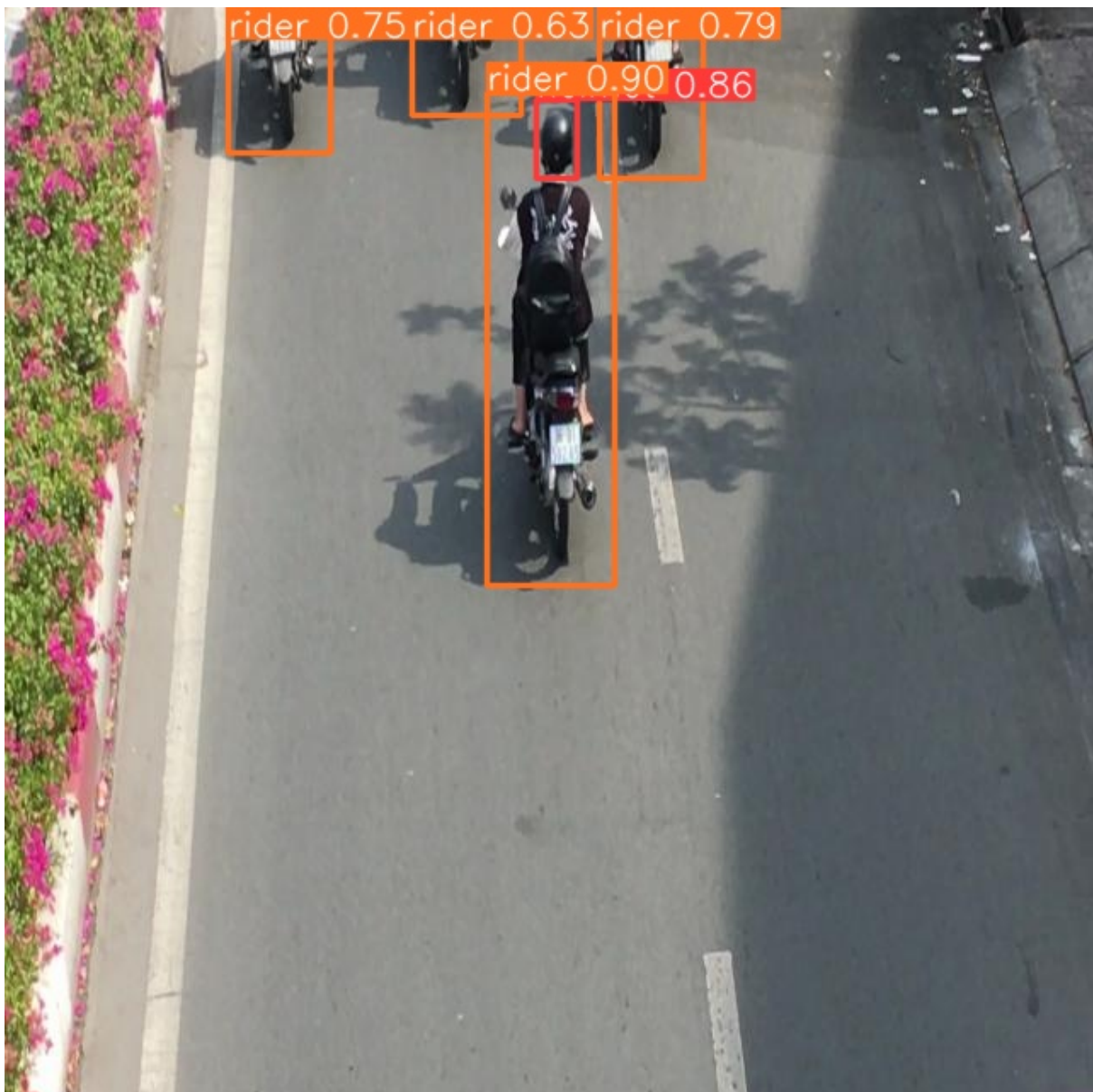
2.2 MỤC TIÊU

- Mục tiêu chính của đồ án là sử dụng một mô hình nhận diện đội mũ bảo hiểm Machine Learning từ động và hiệu quả để nâng cao an toàn giao thông và giảm chấn thương khi gặp tai nạn cho người điều khiển xe máy và xe đạp.

Trong báo cáo này, nhóm em tiếp cận và sử dụng mô hình YOLOv8 để thực hiện tác vụ nhận diện thực thể trên tập dữ liệu mà nhóm thu thập được. Môi trường thực nghiệm được xây dựng và thực hiện trên Google Colab.

- Input: Một ảnh có màu có một hoặc nhiều người tham gia phương tiện giao thông (xe đạp, xe máy, ô tô,..) có đội hoặc không đội mũ bảo hiểm.
- Output: Một danh sách các véc tơ bao gồm những thông tin của bounding boxes như sau:
 - Chỉ mục lớp: Chỉ số của lớp đối tượng được phát hiện trong danh sách lớp. (cụ thể trong đây là: 0-rider, 1-helmet, 2-no-helmet.

- Độ tin cậy: Một số xác suất từ 0 đến 1 đo lường độ chắc chắn của mô hình rằng đối tượng được nhận diện chính xác. Nếu xác suất này thấp, có thể là một dự đoán không chắc chắn.
- x,y,w,h: Tọa độ của bounding box trong hình ảnh, trong đó: x,y là tọa độ tâm; w,h: là kích thước rộng, dài của bounding box.



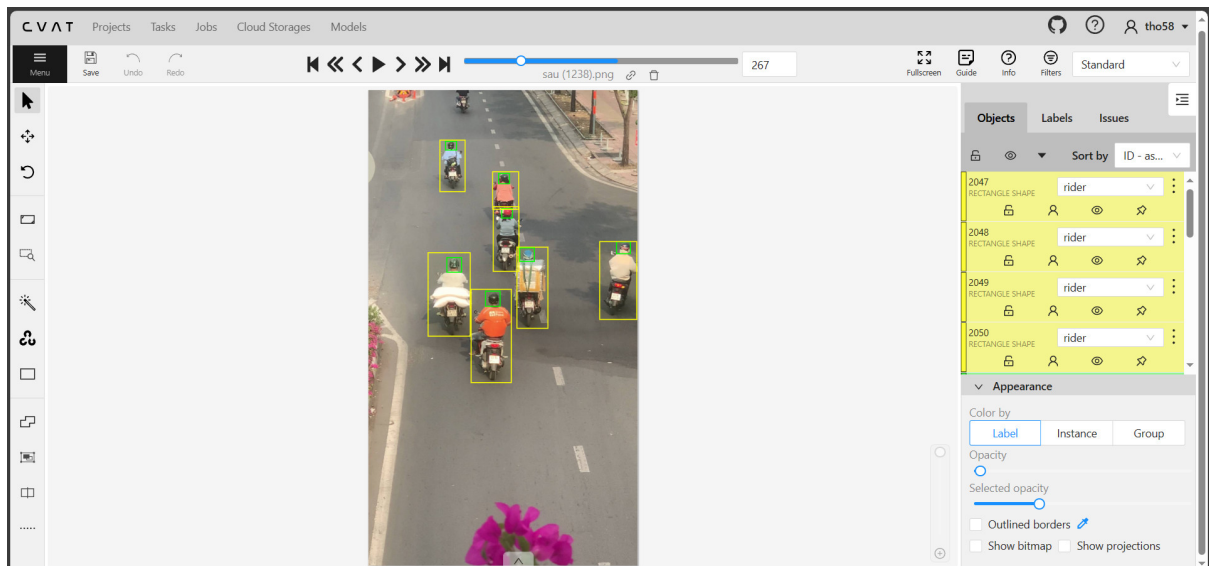
Hình 2: Minh họa về đề tài nhận diện đối tượng nón bảo hiểm

- Ràng buộc:

- Các tấm hình, video phải có độ phân giải 1920x1088, 60fps để mô hình có thể triển khai một cách có hiệu quả và chính xác nhất.
- Thời gian triển khai phải là ban ngày và không mưa.
- Hướng người tham gia phương tiện phải ở chính diện bên dưới hoặc phía sau bên dưới camera.
- Mật độ xe không quá dày đặc (10 chiếc xe đổ xuống/ảnh).

3 CHƯƠNG 2: TẬP DỮ LIỆU

- Bộ dữ liệu bao gồm các ảnh chứa một hoặc nhiều người tham gia phương tiện giao thông (xe đạp, xe máy, ô tô) có đội hoặc không đội mũ bảo hiểm.
- Bộ dữ liệu được nhóm em thu thập ở trên cầu vượt GigaMall Thủ Đức.
- Công cụ thu thập: Iphone 6 plus.
- Tổng ảnh: 2060 ảnh.
- Công cụ gán nhãn: <https://app.cvat.ai>
- Ví dụ:



Hình 3: Hình minh họa quá trình tạo label cho dataset

4 CHƯƠNG 3: PHƯƠNG PHÁP:

4.1 WHAT

Các bức ảnh, video chứa một hoặc nhiều người tham gia phương tiện giao thông (xe đạp, xe máy, ô tô) có đội hoặc không đội mũ bảo hiểm.

4.2 WHY

Việc nhận diện việc đeo mũ bảo hiểm giúp nâng cao an toàn giao thông và giảm nguy cơ tai nạn. Sự tự động hóa thông qua Machine Learning giảm áp lực cho các phương pháp giám sát truyền thống và cung cấp thông tin chính xác.

4.3 WHEN

Thời gian triển khai phải là ban ngày, không mưa, hạn chế giờ cao điểm.

4.4 WHERE

Hệ thống có thể triển khai tại các điểm giao thông quan trọng, đèn giao thông.

4.5 WHO

Đối tượng hướng đến là những người lái xe máy, xe đạp và các cơ quan chức năng và tổ chức quản lý giao thông.

4.6 HOW

YOLOv8

Nhóm chọn sử dụng mô hình YOLOv8 vì khả năng linh hoạt và hiệu suất cao. YOLOv8 cung cấp khả năng nhận diện thời gian thực.

Giới thiệu tổng quan và định dạng yolo

Định dạng annotation của YOLO (định dạng txt) có dạng như sau:

<object-class> <x> <y> <width> <height>

Chú thích:

- object-class: là tên định dạng của bounding box.
- <x> <y> <width> <height>: là các số thực chỉ vị trí tương ứng của object so với ảnh, nằm trong khoảng từ 0 đến 1.
- Cách tính x center là: $\text{<x>} = \text{<absolute_x>} / \text{<image_width>}$ hoặc $\text{<y>} = \text{<absolute_height>} / \text{<image_height>}$
- <x_center> <y_center> là tâm của bounding box.

- **Ví dụ:** 0 0.459361 0.419599 0.165009 0.078250

4.7 PHƯƠNG PHÁP ĐÁNH GIÁ

Dataset được chia làm 3 phần: training set, validation set và test set.

- Training set gồm có 1658 bức ảnh.
- Validation set gồm có 268 bức ảnh.
- Test set gồm có 134 bức ảnh.

Precision và Recall

- Precision là tỉ lệ số dự đoán true positive trong số tổng dự đoán là positive.
- Recall là tỉ lệ số dự đoán true positive trong số những positive thực sự.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

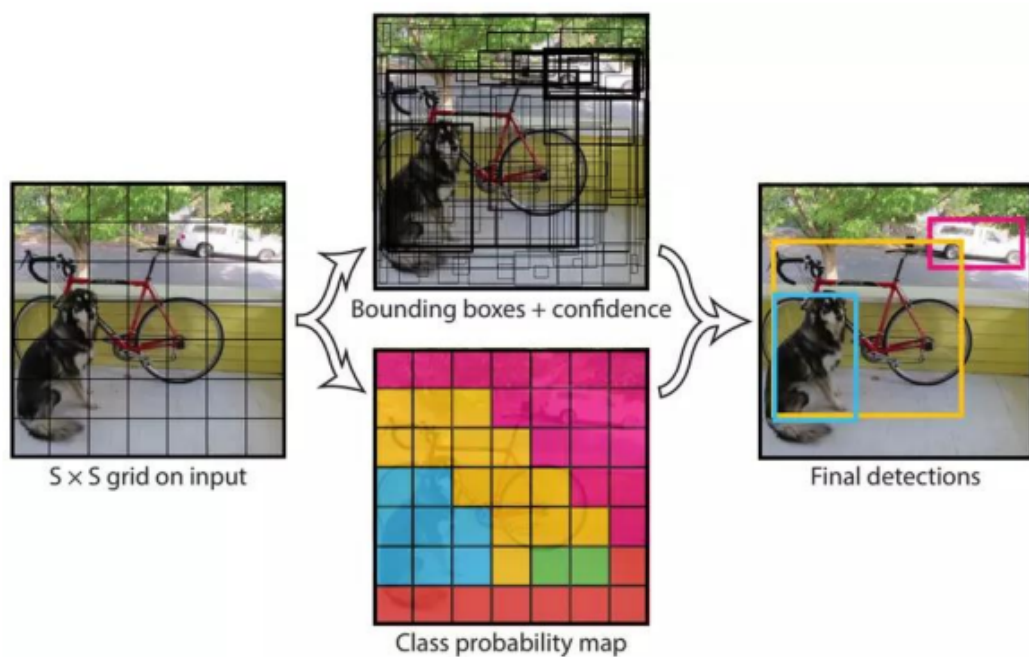
AP và mAP

- AP và mAP: : Từ precision và recall đã được định nghĩa ở trên chúng ta cũng có thể đánh giá mô hình dựa trên việc thay đổi một ngưỡng và quan sát giá trị của Precision và Recall. Khái niệm Area Under the Curve (AUC) cũng được định nghĩa tương tự. Với Precision-Recall Curve, AUC còn có một tên khác là Average precision (AP).
- mAP của mô hình bài toán trên là trung bình của các tất cả các class.

- Average Precision là trung bình của các precision tại các điểm mà mỗi kết quả đúng trả về.
- Mean Average Precision (mAP) là trung bình giá trị AP mỗi class

4.8 SỬ DỤNG PHƯƠNG PHÁP YOLO

Sơ lược về YOLO



Hình 4: Hình minh họa cho cách hoạt động yolo

Đầu vào của mô hình là một tấm ảnh, mô hình sẽ nhận dạng ảnh đó có đối tượng nào hay không, sau đó sẽ xác định tọa độ của đối tượng trong bức ảnh. Ảnh đầu vào được chia thành $S \times S$ ô thường là 3×3 , 7×7 , 9×9 ...

Các bước thực hiện của YOLO:

- **Bước 1:** Với Input là 1 ảnh, đầu ra mô hình là một ma trận 3 chiều có kích thước là $S \times S \times (5 \times N + M)$ với N và M lần lượt là số lượng Box và Class mà mỗi ô cần dự đoán.

- **Bước 2:** Với mỗi ô lưới, chạy một mạng CNN dự đoán các bounding box trong ô đó. Trong tâm của vật thể sẽ được tìm trong các grid và nếu nó nằm trong ô lưới nào, thì ô lưới chứa trọng tâm của đối tượng sẽ chịu trách nhiệm tìm vật thể đó.

Ví dụ:

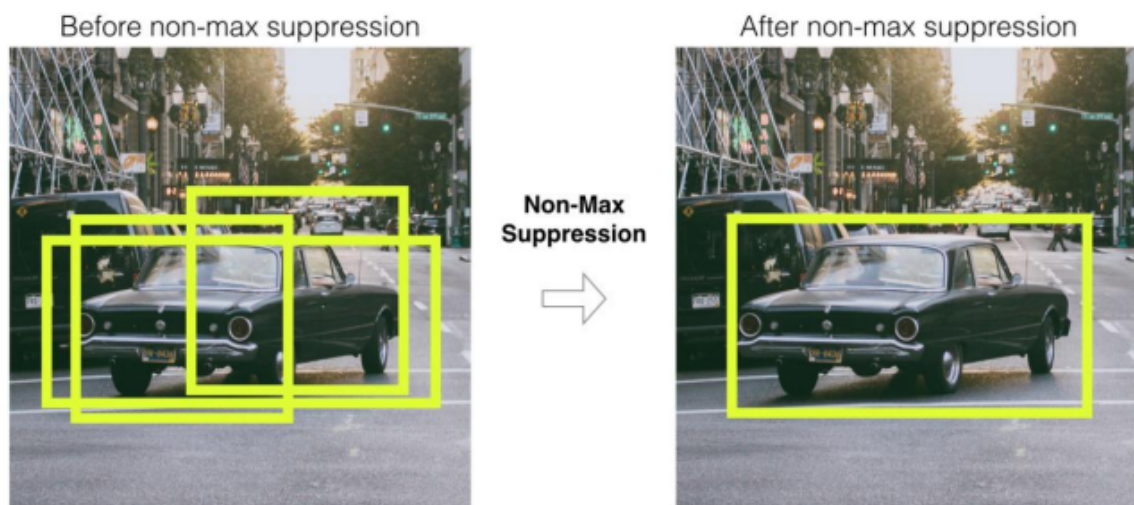
Với hình ảnh trên chỉ thành 7x7 ô lượng Box và Class mà mỗi ô cần dự đoán 2 bounding box và 3 object: con chó, ô tô, xe đạp, thì output sẽ là 7x7x13, mỗi ô sẽ có 13 tham số, kết quả trả về ($7 \times 7 \times 2 = 98$) bounding box. Dự đoán mỗi bounding box gồm có 5 thành phần: (x, y, w, h, prediction) với (x,y) là tọa độ tâm của bounding box, (w, h) lần lượt là chiều rộng và chiều cao của bounding box, prediction được định nghĩa $\text{Pr}(\text{object}) * \text{IOU}(\text{pred}, \text{Truth})$. Với hình ảnh trên như ta tính thì mỗi ô sẽ có 13 tham số, ta có thể hiểu đơn giản như sau: tham số thứ 1 sẽ chỉ ô đó có đối tượng nào hay không $P(\text{Object})$, tham số 2,3,4,5 sẽ trả về x,y,w,h của Box1. Tham số 6,7,8,9,10 tương tự về Box2, tham số 11,12,13 lần lượt là xác suất ô đó có chứa object1 ($P(\text{chó}|\text{object})$), object2 ($P(\text{ô tp6}|\text{object})$), object3 ($P(\text{xe đạp}|\text{object})$). Lưu ý rằng tâm của bounding box nằm ở ô nào thì ô đó sẽ chứa đối tượng, cho dù đối tượng có thể ở các ô khác thì cũng sẽ trả về là 0. Vì vậy việc mà 1 ô chứa 2 hay nhiều tâm của bounding box hoặc đối tượng thì sẽ không detect được, đó là hạn chế của mô hình YOLO1, vậy ta cần tăng số lượng ô chia trong 1 ảnh lên. Đó là lí do vì sao việc chia ô có thể ảnh hưởng tới việc mô hình phát hiện

đối tượng.

- **Bước 3:** Chạy thuật toán non-max suppression (loại bỏ những bounding box bị trùng, chồng lên nhau của cùng một đối tượng bằng cách chọn bounding box có tính đặc trưng nhất).

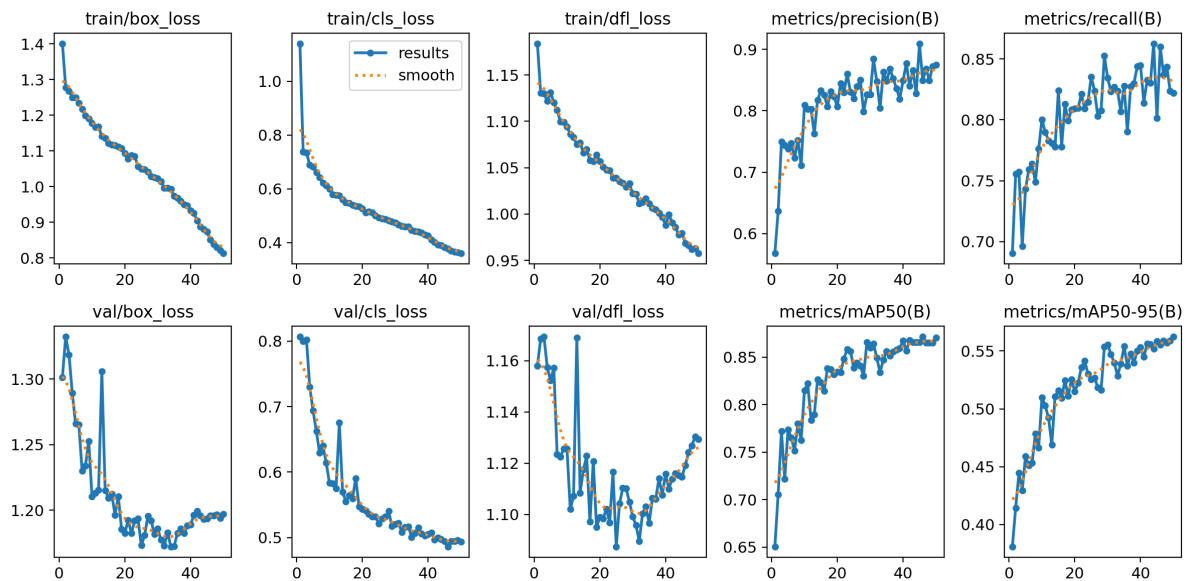
Ở đây thuật toán non-max suppression: Lọc bớt các bounding box dư thừa (trùng nhau và cùng class) bằng non-max suppression. Ý tưởng:

- Các box có confidence_score được xếp hạng từ cao xuống thấp.
[box_0, box1, .. box n]
- Duyệt từ đầu danh sách, với từng box_i , loại đi các box_j có $IOU(box_i, box_j) > threshold$.
- Trong đó $j > i$. threshold làm một giá trị ngưỡng được lựa chọn sẵn. IOU là công thức tính độ over - giao thoa giữa hai bounding box.



Hình 5: Hình minh họa cách thức hoạt động của Non-MaxSuppression

5 CHƯƠNG 4: KẾT QUẢ



Hình 6: Kết quả

Đánh giá

- Hiệu suất dự đoán rider và helmet tương đối tốt.
- Tuy nhiên hiệu suất dự đoán no-helmet không được tốt. Lý do là vì trong bộ dữ liệu có quá ít trường hợp no-helmet, dẫn đến underfitting.

6 CHƯƠNG 4: KẾT LUẬN

- **Hiệu suất mô hình:** Mô hình đã cho thấy hiệu suất tương đối ổn đối với việc nhận dạng người đi xe máy, xe đạp có đội mũ, nhưng hiệu suất khá thấp đối với việc nhận dạng người đi xe máy, xe đạp không đội mũ (do data về class này không nhiều nên đã bị underfitting).
- **Ứng dụng thực tế:** Hệ thống có thể triển khai rộng rãi ở các điểm giao thông nếu được thu thập thêm nhiều data để cải thiện.

- **Tương lai phát triển:** Hệ thống có thể cải tiến thêm bằng cách nhận dạng hành vi vừa nghe điện thoại vừa điều khiển phương tiện, biển số xe để có thể gửi vé phạt khi phát hiện ra hành vi vi phạm an toàn giao thông.