

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

-----**-----



BÁO CÁO ĐỒ ÁN CUỐI KỲ

CS331.N21.KHCL

PHÁT HIỆN VÀ LÀM MỜ BIỂN SỐ XE VIỆT NAM

Sinh viên thực hiện:

20521938 Trần Phương Thảo

Hồ Chí Minh, 06 - 2023

[illegible]

MỤC LỤC

NHẬN XÉT CỦA GIẢNG VIÊN	2
-------------------------------	---

A. TỔNG QUAN BÀI TOÁN:	4
------------------------------	---

1. GIỚI THIỆU:	4
----------------------	---

2. MÔ TẢ:	4
-----------------	---

B. PHƯƠNG PHÁP:	4
-----------------------	---

1. TIÊU CHÍ CHỌN MÔ HÌNH:	4
---------------------------------	---

C. THỰC NGHIỆM:	6
-----------------------	---

1. DỮ LIỆU:.....	6
------------------	---

1.1. MÔ TẢ:	6
-------------------	---

1.2. XỬ LÝ:	8
-------------------	---

1.3. THỐNG KÊ:.....	8
---------------------	---

2. ĐÁNH GIÁ MÔ HÌNH:	10
----------------------------	----

3. KẾT QUẢ THỰC NGHIỆM:	17
-------------------------------	----

D. CẢI THIỆN VÀ HƯỚNG PHÁT TRIỂN:	18
---	----

1. CẢI THIỆN:	18
---------------------	----

1.1. DỮ LIỆU:.....	18
--------------------	----

1.2. MÔ HÌNH:	18
---------------------	----

2. HƯỚNG PHÁT TRIỂN:	18
----------------------------	----

TÀI LIỆU THAM KHẢO.....	19
-------------------------	----

A.TỔNG QUAN BÀI TOÁN:

1. GIỚI THIỆU:

Cách mạng 4.0 phát triển dẫn đến việc phát triển của nhiều ngành nghề, một trong đó là các nghề như youtuber, vlogger... là những nghề thường quay những video, clip sau đó đăng lên mạng xã hội. Tuy nhiên khi quay cảnh đường phố thì các biển số xe thường không được làm mờ. Điều này có thể để lộ ra thông tin người trong ảnh đến đó và thời điểm nào dẫn đến sự không hài lòng của chủ sở hữu.

Sản phẩm của đề án này sẽ mang đến những lợi ích như đẩy nhanh quá trình tự động che biển số xe cho video; người dùng không tốn quá nhiều thời gian và kỹ năng để xử lý; ...

2. MÔ TẢ:

- Input:

Một video có độ dài bất kỳ với định dạng mp4 và độ phân giải tối thiểu 480p.

- Output:

Một video với thời lượng và nội dung giống như đầu vào với biển số xe (được cấp ở Việt Nam) đã được làm mờ ([link demo](#)).

B.PHƯƠNG PHÁP:

1. TIÊU CHÍ CHỌN MÔ HÌNH:

- Tốc độ: tốc độ xử lý từng video phải nhanh hơn so với phương pháp làm mờ thủ công.
- Khả năng phát hiện vật thể tốt: mô hình phải có khả năng phát biển số tốt tránh bỏ sót hay nhầm lẫn giữa biển số và nền.

- Hỗ trợ đầu vào là video.

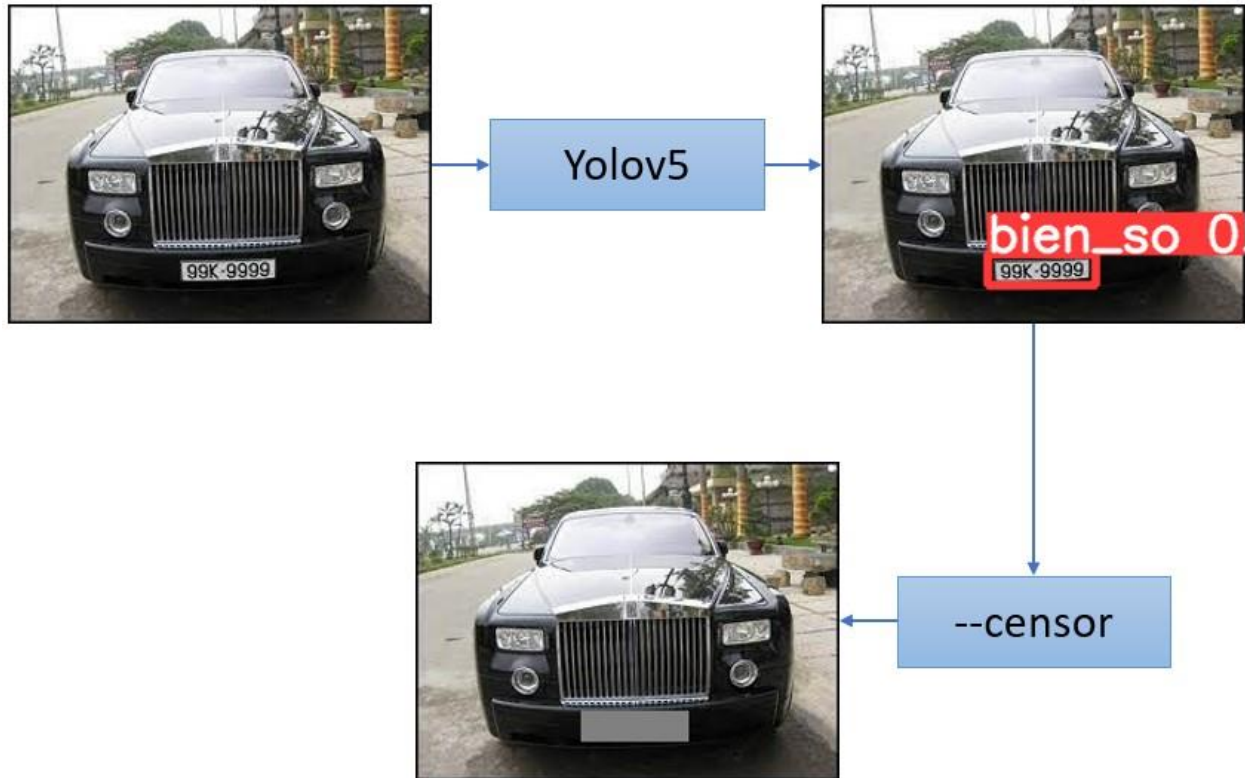
2. CÁCH THỰC HIỆN:

Từ những tiêu chí trên, có thể thấy được mô hình YOLOv5 là 1 trong những state-of-the-art objects detector đáp ứng được.

YOLOv5 được giới thiệu vào năm 2020 bởi Glenn Jocher. Trong bài toán này, ta tiến hành training như sau:

- Image size: 640. Là kích thước ảnh đầu vào mà mô hình sẽ resize lại trước khi học.
- Epochs: 1000. Số lần mô hình sẽ học.
- Batch size: auto. Việc đưa cả dataset vào GPU để train là điều không thể nên sẽ chia dataset thành các batch (với kích thước phù hợp với GPU) để train ở mỗi epoch.
- Weight: YOLOv5l. Vì có tốc độ và độ chính xác chỉ sau YOLOv5x nhưng số lượng parameters chỉ bằng một nửa của YOLOv5x.

Pipeline trong bài toán này:



Hình 1: Các bước thực hiện

C.THỰC NGHIỆM:

1. DỮ LIỆU:

1.1. MÔ TẢ:

- Dữ liệu được thu thập từ các kênh Youtube và thư viện mìn AI.
 - Nguồn: [HaNoi Travel](#), [Vietnam Bus Vlog](#), [New Sài Gòn](#), [Mìn AI](#),...
 - Chất lượng video: tối thiểu 720p.
- Đồng thời dữ liệu được thu thập bằng cách chụp hay quay bởi điện thoại thông minh.

- Địa điểm: trên các con đường ở các quận của thành phố Hồ Chí Minh, bãi giữ xe trường Đại học Công nghệ Thông tin và Học viện Hành chính Quốc gia TPHCM, ...
- Cách thu thập:
 - o Sử dụng điện thoại thông minh.
 - o Quay vào nhiều thời điểm trong ngày và các ngày khác nhau
 - o Quay chủ yếu vào biển số xe của các xe đang lưu thông, đang đỗ... trên đường bộ.



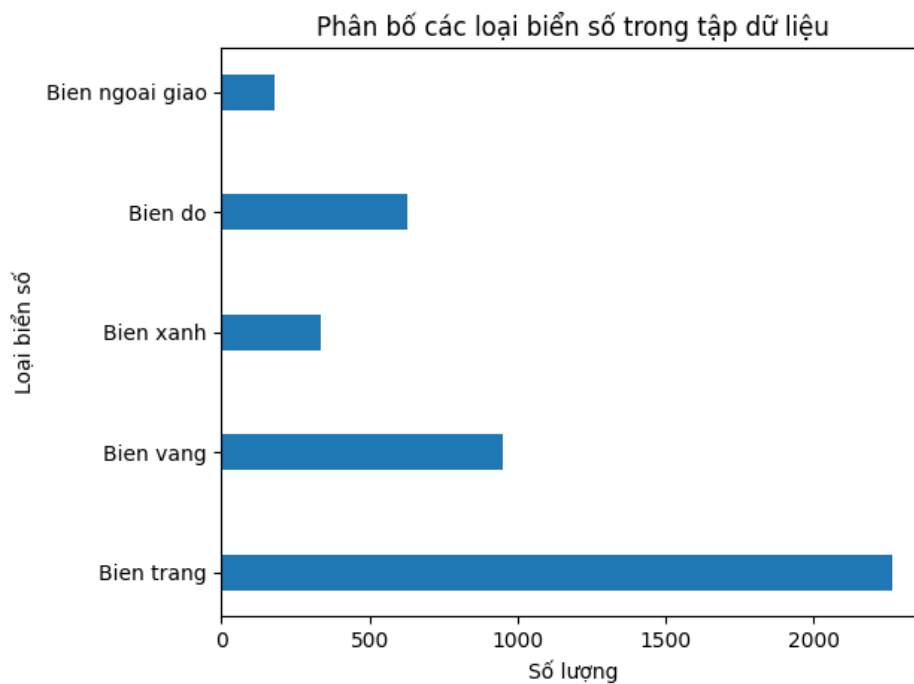
Hình 2: Một số ảnh trong dữ liệu.

1.2. XỬ LÝ:

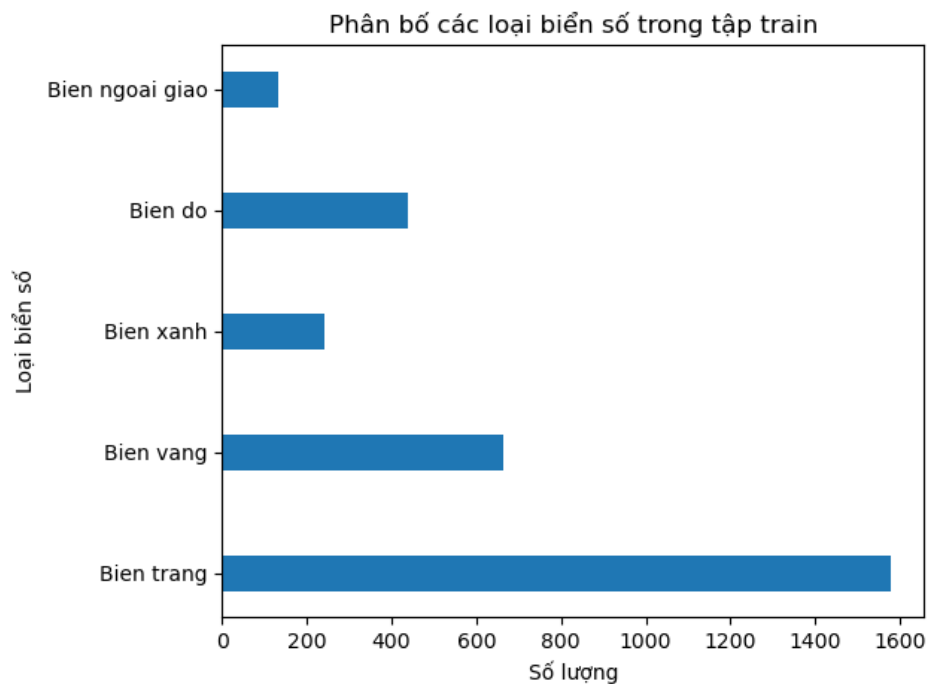
Đối với các video, tiến hành cắt ra thành các frame với 2 frame liên tiếp trong 1 video cách nhau 0.5 giây và lưu thành ảnh. Sau đó thực hiện quá trình lọc và loại bỏ các ảnh bị nhòe, mờ và không chứa biển số.

Tiêu chí gán nhãn: các biển số có thể nhìn rõ và đầy đủ trong ảnh (nếu phóng to mà vẫn không thấy và các biển số chỉ thấy được 1 phần thì sẽ không dán nhãn), và các bounding box phải bao sát biển số.

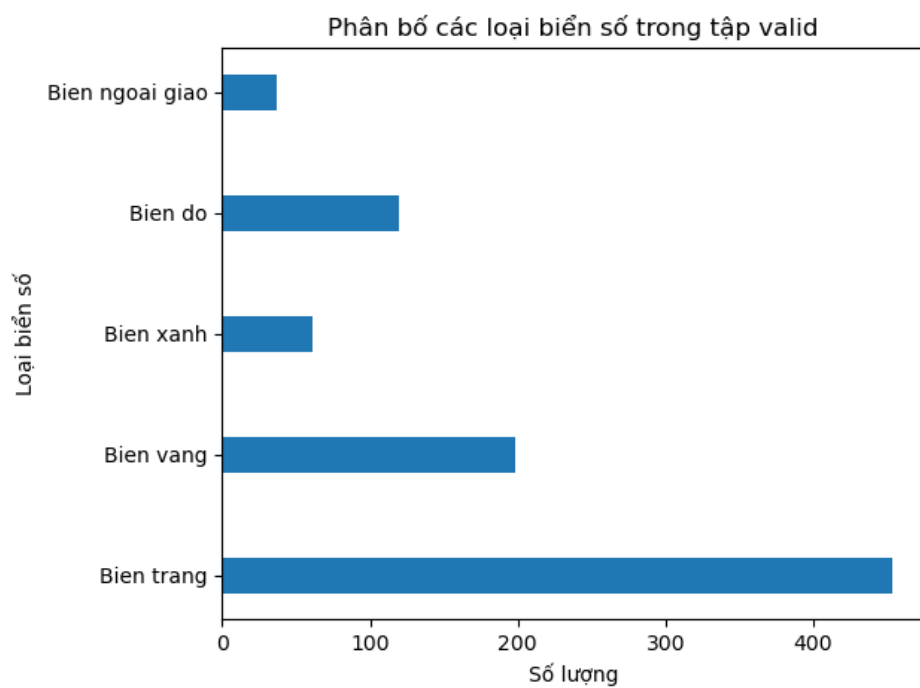
1.3. THỐNG KÊ:



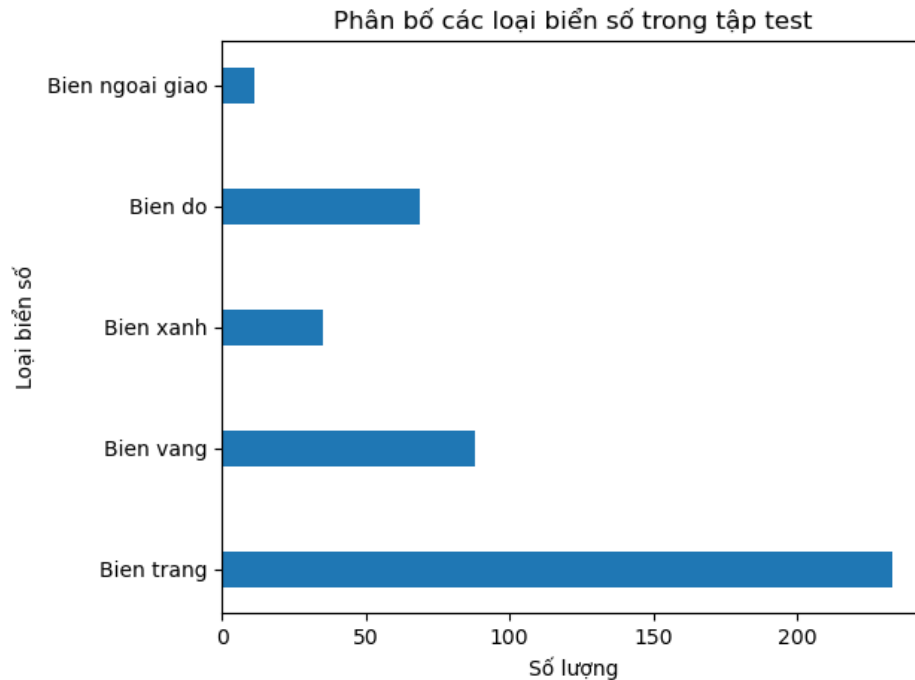
Hình 3: Phân bố các loại biển số trong tập dữ liệu



Hình 4: Phân bố các loại biến số trong tập train



Hình 5: Phân bố các loại biến số trong tập valid



Hình 6: Phân bố các loại biển số trong tập test

2. ĐÁNH GIÁ MÔ HÌNH:

Việc đánh giá được thực hiện đối với 2 giá trị mô hình trả về:

- Tọa độ bounding box: tọa độ hình chữ nhật mà mô hình cho rằng vùng ảnh đó có biển số xe.
- Confidence score: phản ánh khả năng bounding box đó có biển số xe và mức độ tự tin của mô hình về khả năng đó. Nếu bounding box không nằm trong vùng ảnh có biển số xe thì confidence score lý tưởng là 0. Và confidence score có xu hướng cao hơn đối với các bounding box bao sát vùng có biển số xe hơn.

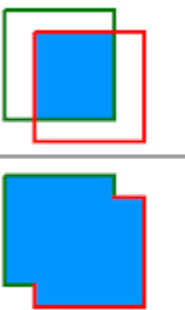
2.1. CẤU HÌNH THỰC HIỆN ĐÁNH GIÁ:

Thực hiện đánh giá trên Google Colab với cấu hình GPU Tesla T4.

2.2. THANG ĐO:

a. IoU (Intersection over Union):

Chỉ số IoU chỉ ra độ trùng khớp giữa bounding box được mô hình dự đoán và ground truth do con người gán sẵn. với mỗi ground truth box ta sẽ tiến hành tính IoI với tất cả các bounding box mà mô hình dự đoán.

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{area of overlap}}{\text{area of union}}$$


Khi có được IoU, ta chọn threshold để phân loại bounding box thành 3 loại: TP, FP và FN.

- TP (true positive): bounding box có IoU với 1 ground truth box cao nhất trong tất cả các giá trị IoU của ground truth box đó và cao hơn threshold.
- FP (false positive): bounding box không phải TP.
- FN (false negative): ground truth box không có bounding box nào được tính là TP.

Số lượng của 3 loại bounding box trên được dùng để tính toán các thang đo bên dưới. Với các threshold khác nhau thì số lượng 3 loại bounding box cũng khác nhau.

Giả sử ta đánh giá trên ảnh sau:



Hình 7: Ảnh ground truth (bên trái) và predicted bounding box (bên phải)

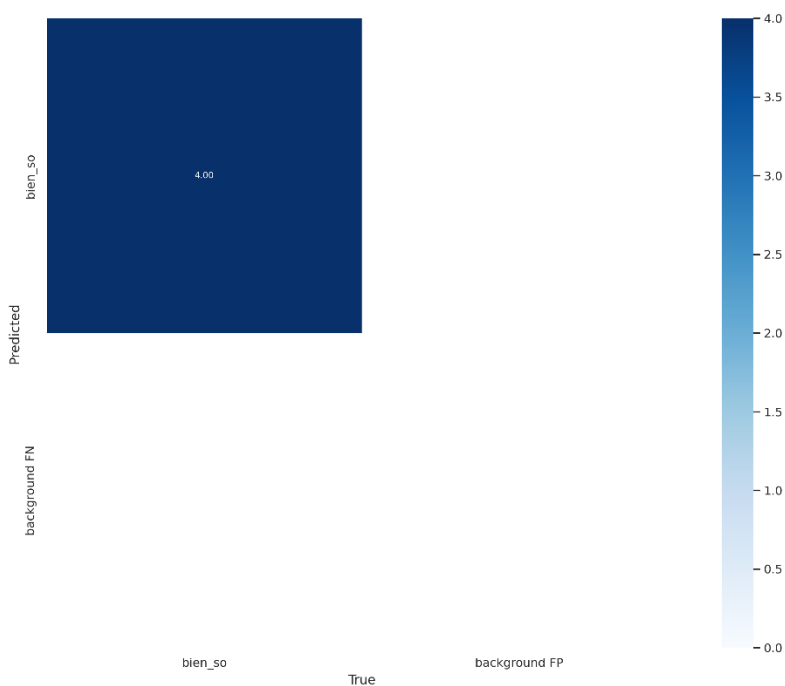
Trong đó:

- Hình chữ nhật màu tím bên ảnh trái là các ground truth box (4 ground truth box).
- Hình chữ nhật màu đỏ bên ảnh phải là các bounding box (4 bounding box).
- Các bounding box được mô hình dự đoán đều có ngưỡng confidence score > 0.25 .

Trong ảnh trên, sau khi tiến hành tính toán, ta thu được bảng sau:

<i>Bounding box</i>	0	1	2	3
<i>Ground truth</i>				
0	0.88462			
1		0.78704		
2			0.83117	
3				0.79257

Bảng 1: IoU của các bounding box trong hình 7.



Hình 8: Confusion matrix với threshold 0.45.

b. Precision:

Precision là tỷ lệ trường hợp bounding box có $\text{IoU} \geq \text{threshold}$ trong các bounding box được dự đoán.

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

Precision có thể không giảm cùng với Recall. Định nghĩa về Precision cho thấy việc giảm threshold có thể làm tăng mẫu số, bằng cách tăng số lượng kết quả trả về. Nếu threshold đang chọn được đặt quá cao, giảm threshold có thể khiến tất cả các kết quả mới là TP, tăng Precision và ngược lại sẽ khiến FP tăng, giảm Precision.

c. Recall:

Recall đo lường tỷ lệ bounding box được dự đoán có $IoU \geq$ threshold trên tổng số ground truth box.

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{all\ ground\ truths}$$

Trong đó, $TP + FN$ không phụ thuộc vào threshold. Điều này có nghĩa việc giảm threshold có thể tăng Recall bằng cách tăng số lượng kết quả TP. Ngoài ra, việc hạ thấp threshold có thể không thay đổi Recall, trong khi Precision lại dao động.

d. PR Curve (Precision – Recall Curve) và AP (Average Precision):

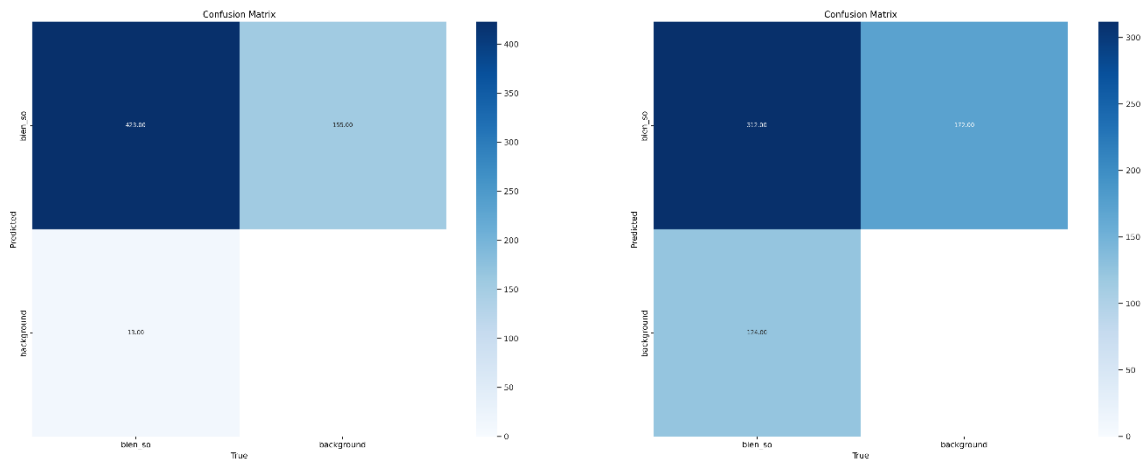
PR Curve cho biết sự cân bằng giữa Precision và Recall đối với các giá trị confidence score khác nhau. Diện tích dưới đường cong (AP) cao nghĩa là Precision và Recall đều cao. Trong bài toán này, Precision cao đồng nghĩa FP thấp và Recall cao đồng nghĩa FN thấp. Điểm cả hai cao cho thấy bounding box mô hình dự đoán chứa biên số (Precision cao) và cũng như mô hình dự đoán được phần lớn các ground truth (Recall cao).

Một mô hình có Recall cao nhưng Precision thấp nghĩa là mô hình trả về nhiều bounding box không trùng ground truth box. Một mô hình có Precision cao nhưng Recall thấp thì ngược lại,

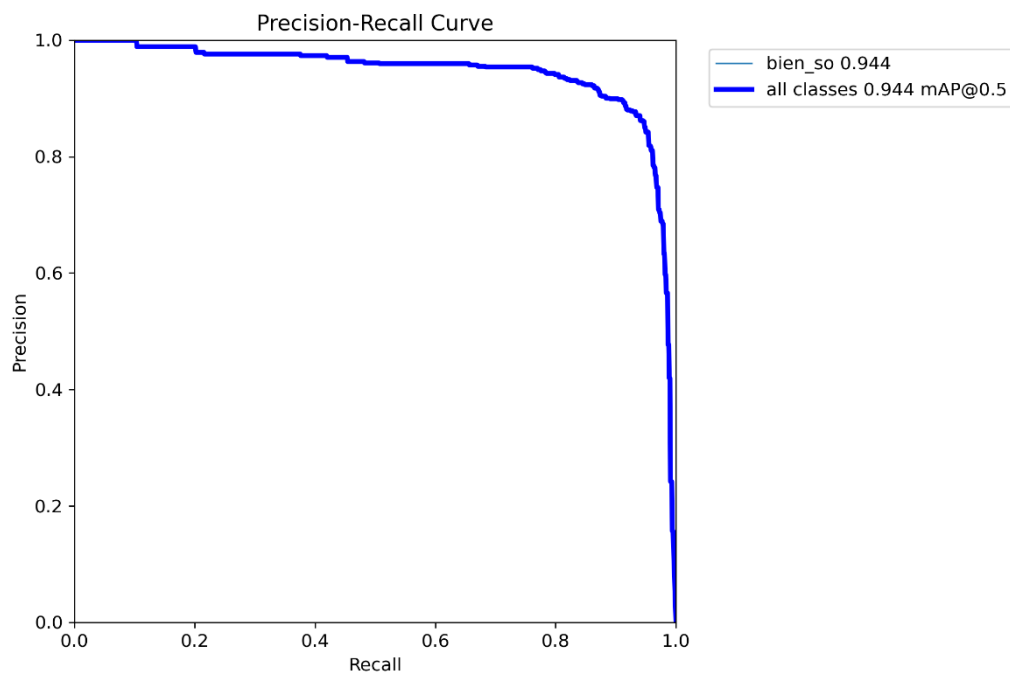
dự đoán rất ít bounding box nhưng đều chứa biển số. Một mô hình lý tưởng với Precision và Recall đều cao sẽ dự đoán nhiều bounding box và đều chứa biển số.

Khi so sánh các mô hình khác nhau, ta sẽ so sánh PR Curve của từng mô hình. Tuy nhiên, PR Curve thường là đường ZigZag lên xuống nên việc so sánh không dễ dàng. Chính vì vậy AP – diện tích dưới PR Curve là 1 con số giúp chúng ta đánh giá mô hình cũng như so sánh các mô hình khác nhau nhanh chóng và dễ dàng hơn. Do Precision và Recall dao động từ 0 đến 1 nên AP cũng dao động từ 0 đến 1.

2.3. KẾT QUẢ:



Hình 9: Confusion matrix trên tập test với threshold 0.5 (bên trái) và 0.75 (bên phải)



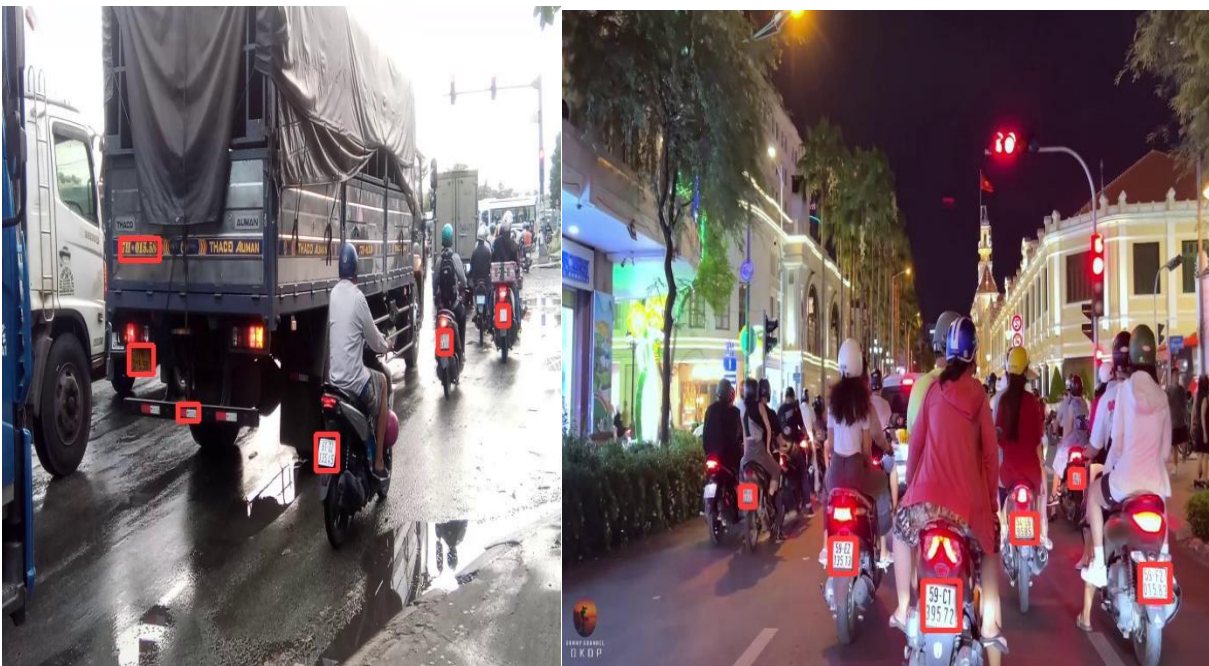
Hình 10: Precision – Recall Curve với IoU threshold 0.5

AP@.5	AP@.75	AP@.5:95	FPS
0.944	0.605	0.548	32,57

Bảng 2: Kết quả thu được trên tập test với IoU threshold 0.5

3. KẾT QUẢ THỰC NGHIỆM:

Các kết quả mô hình dự đoán:



Hình 11: Kết quả mô hình dự đoán không tốt (bên trái) và tốt (bên phải)

Nhận xét:

- Phân hình chữ nhật che biển số không bị lệch khỏi vị trí của biển số, che khá tốt các biển số có thể nhìn rõ trong ảnh hay video, thích hợp để ứng dụng che biển số.
- Có những phần che không chưa biển số nhưng không quá nhiều và không ảnh hưởng đến nội dung ảnh hay video.
- Với các video khi quay bị rung lắc thì mô hình vẫn che tốt.

D.CẢI THIỆN VÀ HƯỚNG PHÁT TRIỂN:

1. CẢI THIỆN:

1.1. DỮ LIỆU:

- Thu thập thêm các biên màu nhiều hơn, đặc biệt là biên đỏ và xanh.
- Bổ sung ảnh nền để giảm sự lẫn lộn của mô hình về biên số và nền.
- Tăng số lượng các biên số trong các điều kiện khó khăn: biên số bị ố màu, bị ảnh hưởng bởi ánh sáng, bị bám bụi, ...

1.2. MÔ HÌNH:

Tiếp tục train mô hình trên các kích thước ảnh khác nhau để tìm ra kích thước phù hợp cho ra kết quả cao và tốc độ xử lý nhanh.

2. HƯỚNG PHÁT TRIỂN:

- Phát triển lên 1 trang web để mọi người có thể sử dụng.
- Tìm ra mô hình có tốc độ xử lý nhanh và kết quả tốt, áp dụng vào xử lý real time, phù hợp cho các trường hợp livestream.

TÀI LIỆU THAM KHẢO

- [1] YOLOv5 - <https://github.com/ultralytics/yolov5>
- [2] The practical guide for Object Detection with YOLOv5 algorithm - <https://towardsdatascience.com/the-practical-guide-for-object-detection-with-yolov5-algorithm>
- [3] mAP (mean Average Precision) for Object Detection - <https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection>