



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

Xử lý ảnh và ứng dụng

Bài toán: Helmet Detection

Giảng viên hướng dẫn : Cáp Phạm Đình Thăng

Nhóm CS406 Primary:

Trương Huỳnh Thúy An	22520033
Trương Hồng Anh	22520084
Hoàng Đức Chung	22520161
Nguyễn Hải Đăng	22520189

NỘI DUNG

I. Giới thiệu đề tài

II. Phát biểu bài toán

III. Phương pháp thực hiện

IV. Phương pháp triển khai

V. Thực nghiệm

VI. Kết luận

VII. Demo



I. Giới thiệu đề tài



I. Lý do chọn đề tài

Với sự phát triển nhanh chóng của xe máy tại Việt Nam nhờ vào tính linh hoạt và chi phí thấp, phương tiện này cũng đã trở thành yếu tố chính góp phần vào tỷ lệ tai nạn giao thông, chiếm khoảng **60% tổng số vụ tai nạn**, theo Bộ Giao thông Vận tải. Trong 10 năm qua, gần **7 triệu trường hợp vi phạm không đội mũ bảo hiểm** đã bị xử lý (Cục Cảnh sát Giao thông), cho thấy việc tuân thủ quy định này còn hạn chế. Điều này làm đặt ra nhu cầu cấp thiết trong việc nghiên cứu các giải pháp hiệu quả nhằm giảm thiểu thương vong và bảo vệ an toàn cho người tham gia giao thông.



II. Phát biểu bài toán



II. Phát biểu bài toán

Input: Hình ảnh được trích xuất từ camera giám sát giao thông ghi lại người điều khiển xe máy.

Output: Người tham gia giao thông trong hình ảnh được phát hiện và phân loại:

- Người đội mũ bảo hiểm: được khoanh vùng bằng hộp màu xanh
- Người không đội mũ bảo hiểm: được khoanh vùng bằng hộp màu đỏ



Input



Output

III. Phương pháp thực hiện

1. YOLOv5m

2. YOLOv8m

3. YOLO11m

4. Deep Ensemble of Recurrent Transformers (DERT)

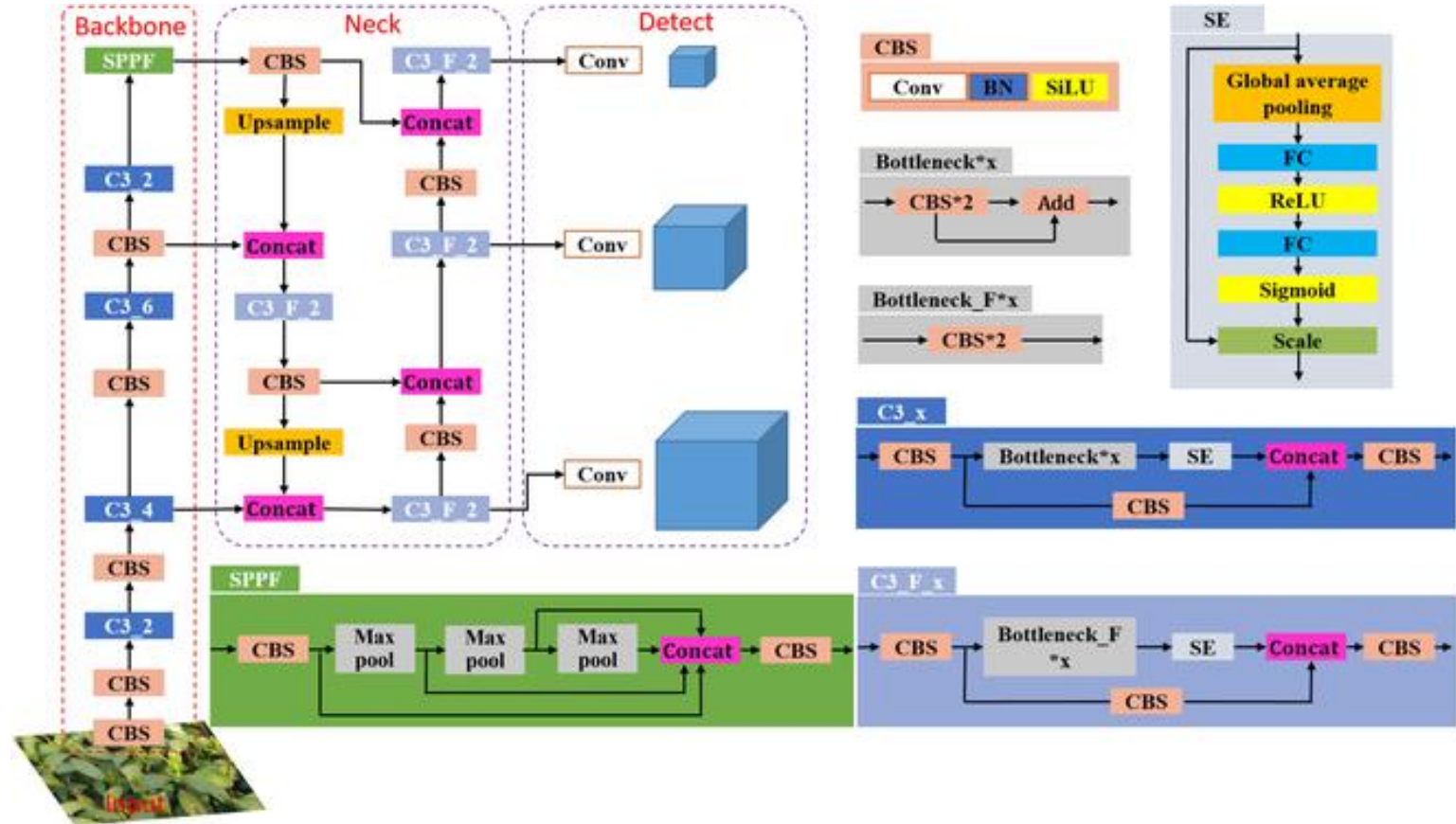


III.1 YOLOv5m

YOLOv5 được giới thiệu bởi **Ultralytics** vào năm 2020, Kế thừa kiến trúc chung của YOLO và đã trở thành một trong những phiên bản YOLO hiệu quả nhất với **tốc độ cao và độ chính xác tốt**. Đây là một mô hình mạnh mẽ cho các tác vụ phát hiện vật thể thời gian thực.

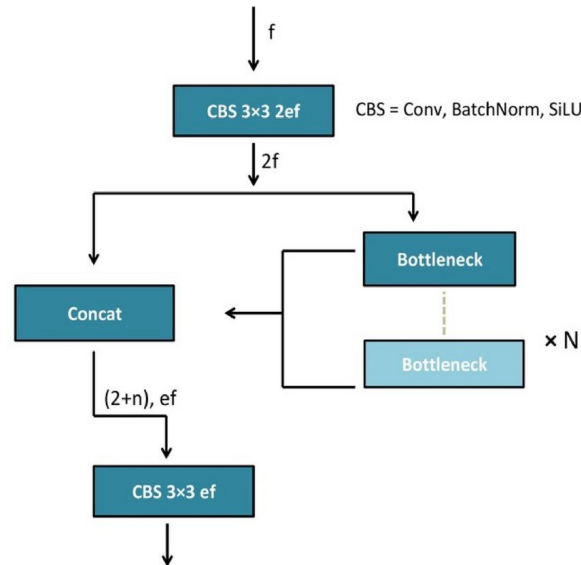


III.1 YOLOv5m



III.2 YOLOv8m

YOLOv8 được giới thiệu vào năm 2023 và được duy trì bởi nhóm Ultralytics. Không đơn giản chỉ là một bản cập nhật của **YOLOv5** mà còn mang lại nhiều cải tiến mạnh mẽ trong cả kiến trúc lẫn hiệu suất. Các **cải tiến về khả năng nhận diện đối tượng nhỏ, tối ưu hóa tốc độ và hiệu quả** cùng với sự hỗ trợ của nhiều tác vụ đa dạng khiến **YOLOv8** trở thành một công cụ ứng dụng mạnh mẽ trong thị giác máy tính.

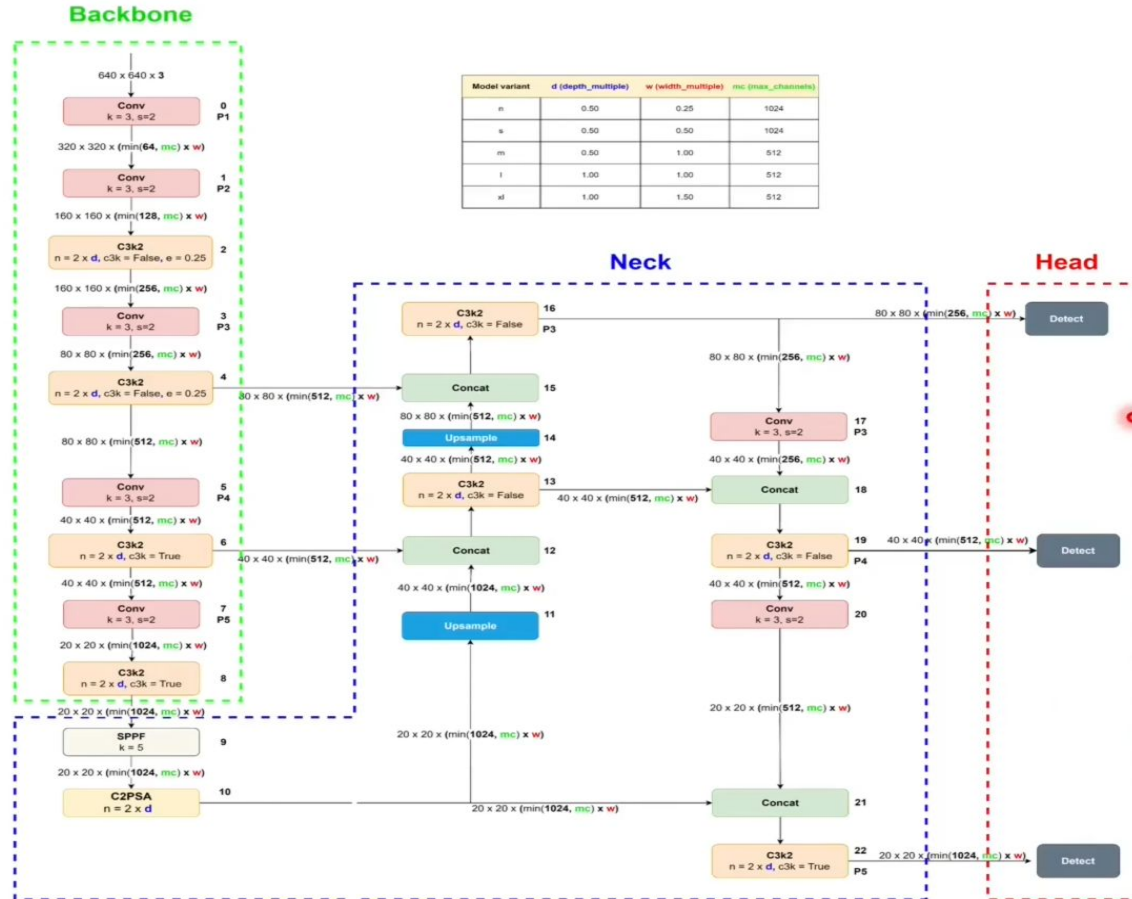


III.3 YOLO11m

YOLO11 được ra mắt vào ngày 30 tháng 9 năm 2024 là **phiên bản mới nhất** của Ultralytics YOLO, **nâng cấp độ chính xác, tốc độ và hiệu quả cao hơn**. Xây dựng dựa trên các phiên bản YOLO trước, **YOLO11** mang đến những cải tiến đáng kể về **kiến trúc và phương pháp đào tạo**, khiến nó trở thành lựa chọn linh hoạt cho nhiều tác vụ thị giác máy tính .

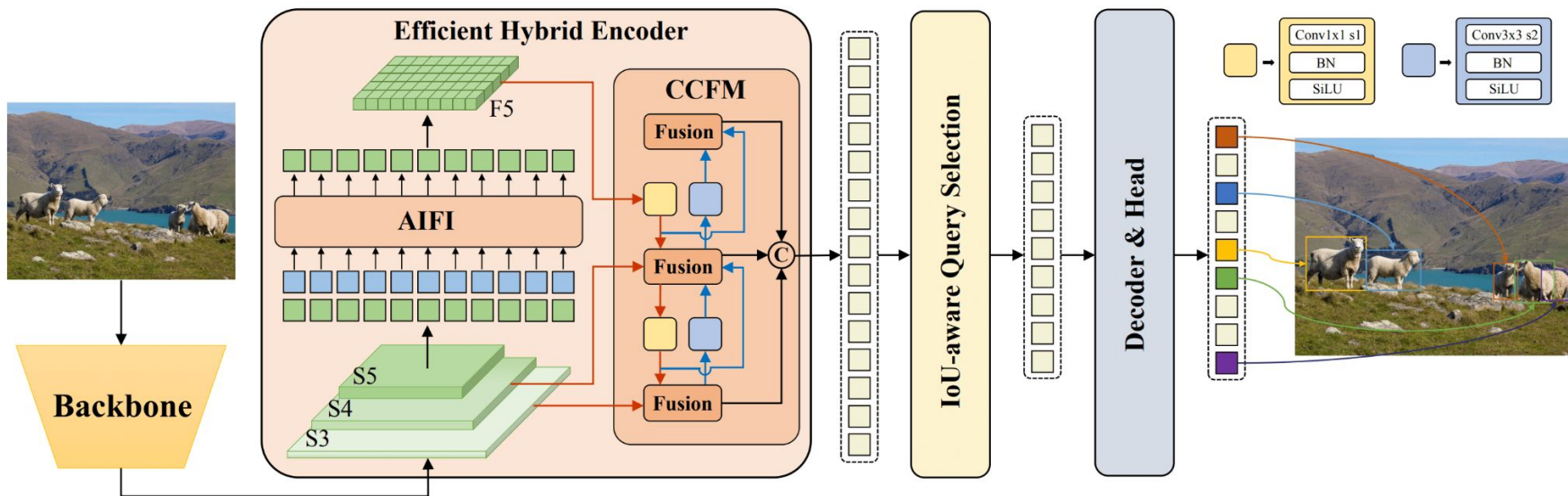


III.3 YOLO11m



III.4 RT-DETR

RT-DETR là một mô hình hiện đại, kết hợp các ưu điểm của Transformer và các kiến trúc truyền thống như YOLO để đạt được cả **tốc độ cao** và **độ chính xác tốt**. Nó được thiết kế để chạy trên phần cứng hạn chế, ví dụ như GPU tiêu chuẩn, mà không làm giảm hiệu suất.



IV. Phương pháp triển khai

1. Slicing Aided Hyper Inference (SAHI)

2. Weighted Box Fusion (WBF)

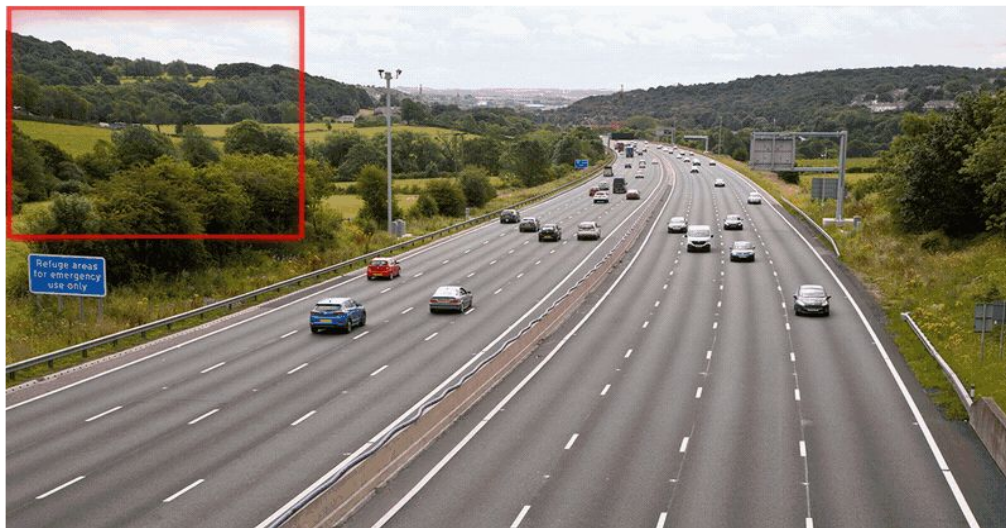


IV.1 SAHI

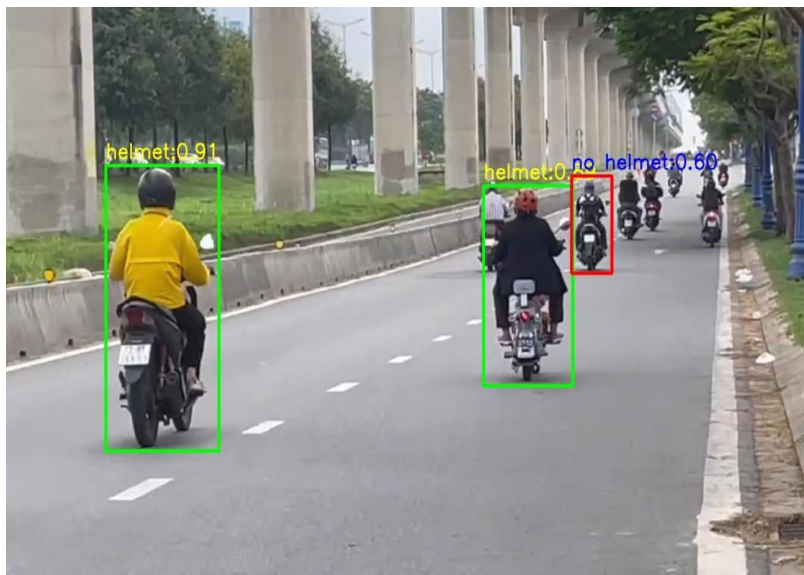
Slicing Aided Hyper Inference là một kỹ thuật và thư viện mã nguồn mở được thiết kế để cải thiện hiệu suất của các mô hình phát hiện vật thể (Object Detection), đặc biệt là trên các đối tượng nhỏ hoặc khi xử lý các hình ảnh có độ phân giải ảnh cao.

SAHI sử dụng cách tiếp cận “chia nhỏ và tổng hợp” để tối ưu hóa quá trình suy luận:

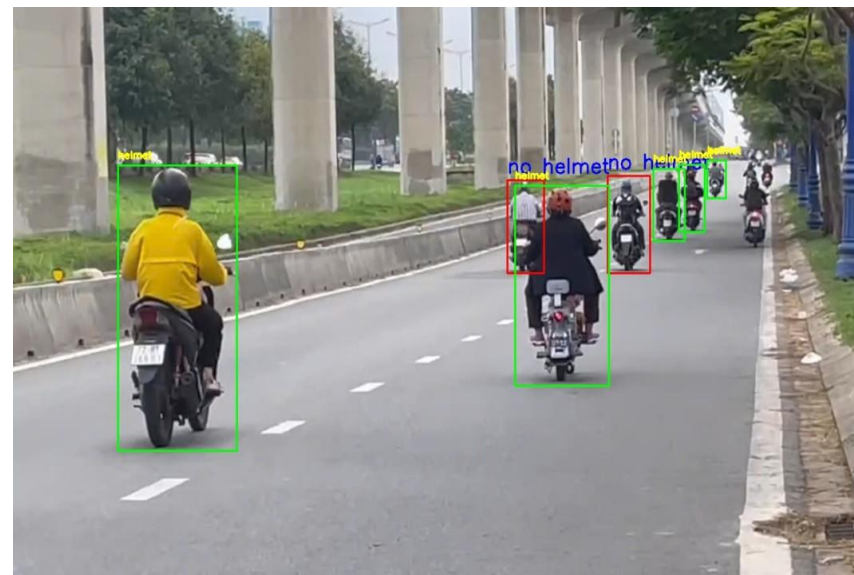
- Chia hình ảnh lớn thành các lát nhỏ (slicing)
- Chạy suy luận trên từng lát ảnh
- Tổng hợp các kết quả lại (prediction merging) và trả về kết quả.



IV.1 SAHI



YOLOv8m

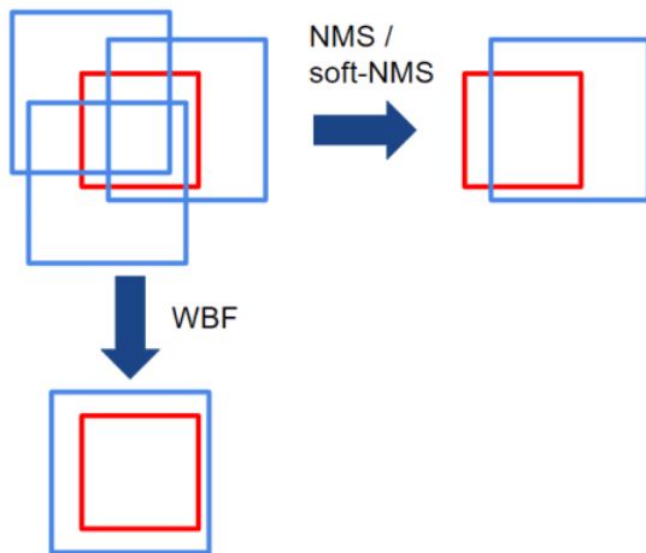


YOLOv8m with SAHI



IV.2 WBF

Weight Boxes Fusion (WBF) là phương pháp hậu xử lý thay thế **Non-Maximum Suppression (NMS)** để kết hợp các hộp dự đoán từ nhiều mô hình phát hiện vật thể. WBF sử dụng trung bình trọng số để tạo ra dự đoán chính xác hơn, giúp cải thiện hiệu suất tổng thể so với NMS. Phương pháp này được giới thiệu trong bài báo *“Weight Boxes Fusion: Ensembling boxes from different object detection models”*.



V. Thực nghiệm

1. Dataset

2. Độ đo

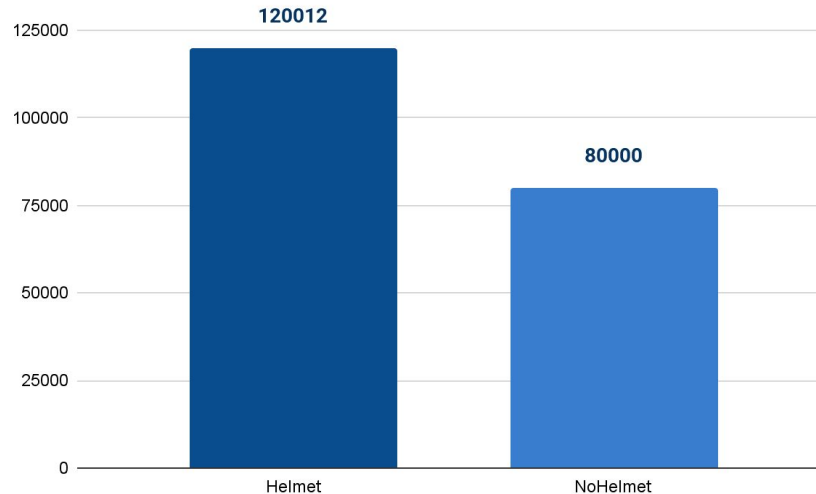
3. Kết quả



V.1 Dataset

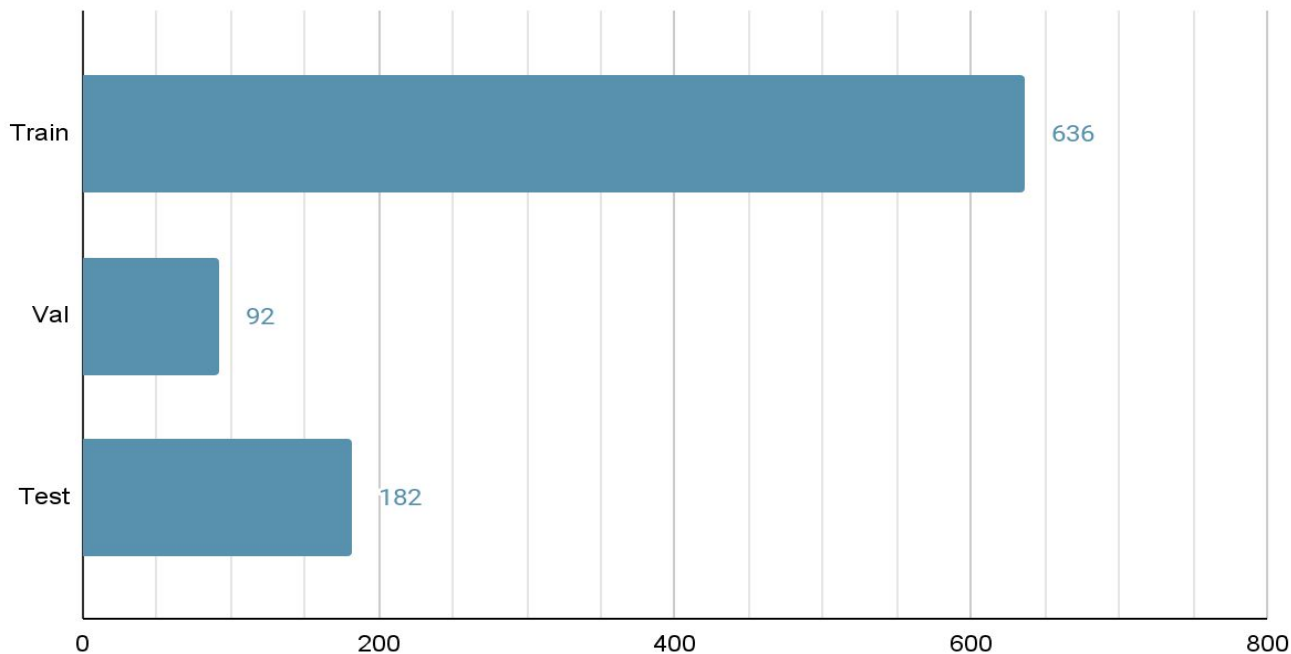
Bộ dữ liệu HELMET được thu thập từ OSF gồm **91.000 khung hình** được gán nhãn, được trích xuất từ **910 video clip**, **mỗi clip** bao gồm **100 khung hình** với độ phân giải **1920x1080** pixel, thu thập tại 7 thành phố ở Myanmar.

Bộ dữ liệu này cung cấp các thành phần chính: thư mục hình ảnh, tệp phân chia dữ liệu (data_split.csv), và tệp chú thích (annotation.zip) chứa thông tin chi tiết về vị trí, số lượng trên xe máy, và trạng thái sử dụng mũ bảo hiểm, hỗ trợ hiệu quả cho việc phát triển các mô hình nhận diện và phân tích an toàn giao thông.



V.1 Dataset

Bộ dữ liệu HELMET, có tổng cộng 910 video được chia thành 3 tập train (70%), validation (10%) và test (20%)



V.2 Độ đo

F1 Score là một chỉ số đánh giá hiệu suất quan trọng, được tính dựa trên trung bình điều hòa giữa **Precision** và **Recall**. Nhằm đánh giá hiệu suất mô hình một cách cân bằng, đặc biệt trong các trường hợp có dữ liệu có sự chênh lệch lớn giữa các lớp.

$$\text{F1 score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}}$$

- Precision: Tỷ lệ bounding boxes đúng trên tổng số dự đoán
- Recall: Tỷ lệ bounding boxes đúng trên tổng số ground truth
- Đúng khi IoU ≥ 0.5

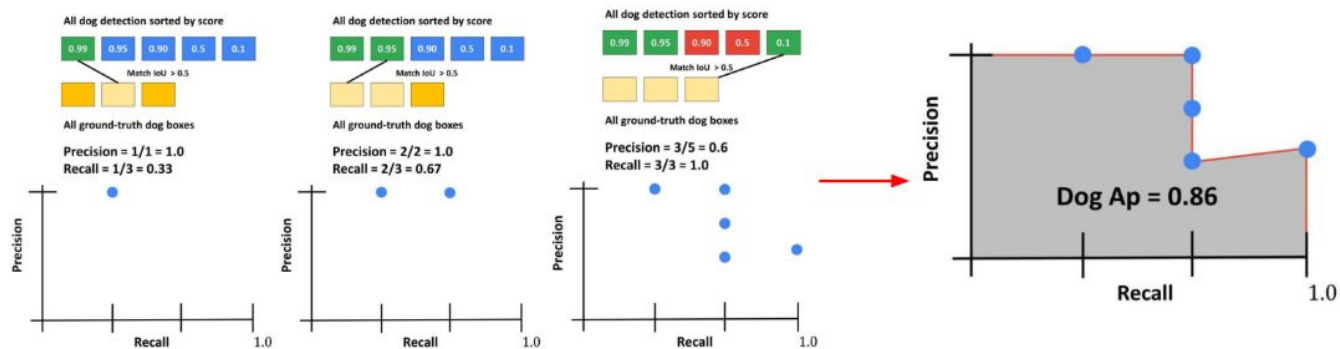


V.2 Độ đo

Mean Average Precision (mAP) là chỉ số đánh giá hiệu suất tổng thể của mô hình trong bài toán object detection. Được tính bằng trung bình của **Average Precision** (AP) trên tất cả các lớp.

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

AP đo lường hiệu suất của mô hình bằng cách tính diện tích đường cong Precision-Recall.



V.3 Kết quả

Mô hình	F1 score	Mean Average Precision (mAP@50)	Mean Average Precision (mAP@50-95)	Time (ms/image)	Num parameters
YOLOv5m	0.8172	0.8768	0.7690	11.6567	25,066,294
YOLOv8m	0.8293	0.8867	0.7836	13.9800	25,857,478
YOLO11m	0.8371	0.8863	0.7886	16.2077	20,054,550
RT-DERT	0.8210	0.8165	0.7050	39.2820	32,810,186



VI. Kết luận



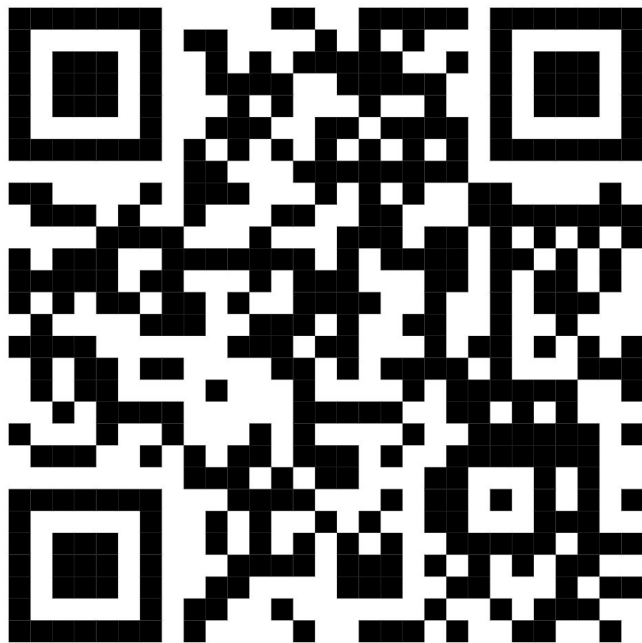
VI. KẾT LUẬN

YOLO11m thể hiện sự **cải tiến vượt trội** so với các phiên bản YOLO trước đó. Cụ thể, mô hình đạt **hiệu suất tốt hơn** với F1 score và mAP@0.5:0.95 cao hơn, cho thấy khả năng phát hiện **chính xác và ổn định** hơn. Đồng thời, YOLO11m sử dụng ít tham số hơn so với YOLOv5m và YOLOv8m, giúp tối **ưu hóa tài nguyên**.

Tuy nhiên, thời gian suy luận của YOLO11m vẫn cao hơn, cho thấy cần có sự đánh đổi giữa độ chính xác và tốc độ khi lựa chọn mô hình cho các ứng dụng thời gian thực.



VII. DEMO



<https://s.net.vn/6TxJ>



Thank You

