

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA: KHOA HỌC MÁY TÍNH



BÁO CÁO ĐỒ ÁN

Bộ môn: CS331 - THỊ GIÁC MÁY TÍNH NÂNG CAO

ĐỀ TÀI:

**NHẬN DIỆN SẢN PHẨM BÁN LẺ
Ở QUÀY THANH TOÁN CỦA CỬA HÀNG
VỚI MÔ HÌNH NHẬN DẠNG YOLOv5**

Giảng viên hướng dẫn:

TS. Mai Tiến Dũng - dungmt@uit.edu.vn

Sinh viên thực hiện:

Hoàng Ngọc Bá Thi - 19522255@gm.uit.edu.vn

Trần Văn Truyền - 19522448@gm.uit.edu.vn

Phạm Minh Trí - 19522390@gm.uit.edu.vn

MỤC LỤC

I.	GIỚI THIỆU ĐỒ ÁN.....	2
II.	BỘ DỮ LIỆU.....	2
1.	Xây dựng bộ dữ liệu.....	3
2.	Tăng cường bộ dữ liệu.....	6
III.	HUẤN LUYỆN MÔ HÌNH NHẬN DIỆN YOLOv5.....	9
1.	Kiến trúc của mô hình YOLOv5.....	9
2.	Huấn luyện mô hình YOLOv5-s.....	13
IV.	THỰC NGHIỆM VÀ ĐÁNH GIÁ MÔ HÌNH.....	13
1.	Các thông số đánh giá mô hình.....	13
2.	Kết quả thực nghiệm.....	14
V.	TỔNG KẾT.....	18
VI.	TÀI LIỆU THAM KHẢO.....	18

I. GIỚI THIỆU ĐỒ ÁN

Ngày nay, phương thức quét mã BarCode trên hàng hóa để kiểm kê, thanh toán đã trở thành phương thức mặc định cho các cửa hàng hiện đại. So với quy trình kiểm thủ công thì phương thức quét mã BarCode cho tốc độ nhanh hơn, và chính xác hơn rất nhiều. Tuy nhiên, chúng ta cũng thường xuyên gặp tình trạng máy quét không đọc được mã BarCode do mã vạch in trên bao bì sản phẩm bị méo mó, biến dạng, hoặc khó nhìn thấy do hình dạng của sản phẩm, điều này gây mất thời gian cho người bán hàng và người mua khi phải nhập chính xác 1 dãy mã vạch dài vào máy thanh toán.

Từ thực tiễn trên, nhóm chúng em quyết định nghiên cứu một phương pháp áp dụng Máy học để quét sản phẩm khác với mong muốn khắc phục được những điểm bất tiện của phương thức quét BarCode kể trên.

Đồ án này hướng đến phương thức nhận diện hàng hóa như sau: chỉ cần đặt các sản phẩm vào khu vực thanh toán, hệ thống sẽ thông qua ảnh chụp từ camera để xuất ra thông tin sản phẩm có trong khu vực thanh toán, từ thông tin này, người bán chỉ cần xuất hóa đơn và thanh toán. Ở đồ án này, nhóm sử dụng **30** sản phẩm khác nhau và một mặt phẳng màu trắng để mô phỏng khu vực thanh toán.

Bài toán đặt ra: Hệ thống cần nhận diện được sản phẩm trong khu vực thanh toán. Đây là bài toán nhận diện vật thể (Object Detection).

INPUT: Ảnh chụp khu vực thanh toán có chứa các mặt hàng của cửa hàng, siêu thị. Ảnh có góc chụp thẳng đứng từ trên xuống, vuông góc với bề mặt đặt sản phẩm

OUTPUT: Thông tin của các sản phẩm (có thể là: mã sản phẩm, tên sản phẩm,...) có trong ảnh chụp khu vực thanh toán.

II. BỘ DỮ LIỆU

Bộ dữ liệu do nhóm tự thu thập và xây dựng được chia thành 3 phần như sau:

- **Tập Train:** gồm 406 bức ảnh, mỗi ảnh chỉ chứa một sản phẩm.
- **Tập Validation:** gồm 30 bức ảnh, mỗi ảnh chứa nhiều sản phẩm khác nhau.
- **Tập Test:** gồm 29 bức ảnh, mỗi ảnh chứa nhiều sản phẩm khác nhau.

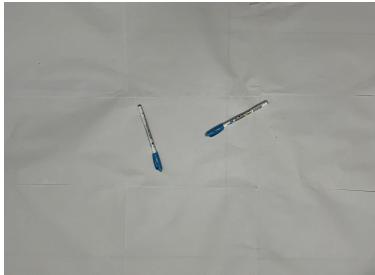
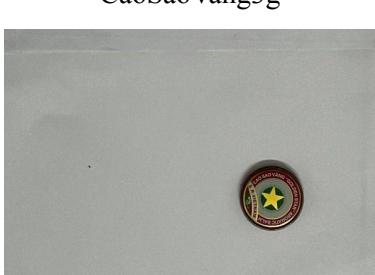
Tất cả các ảnh đều đã được tiền xử lý điều chỉnh kích cỡ về 416x416 pixels để phù hợp với đầu vào của mô hình YOLO.

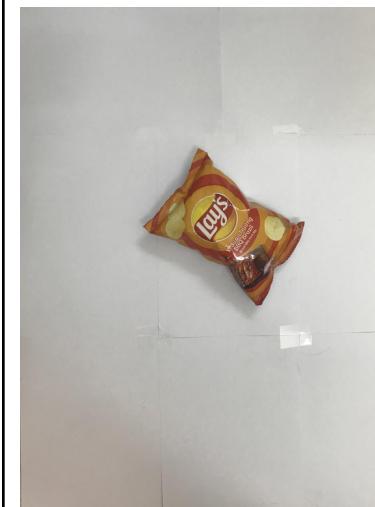
Kèm theo mỗi ảnh là 1 file Annotation (định dạng .txt) lưu thông tin của bounding box có trong ảnh.

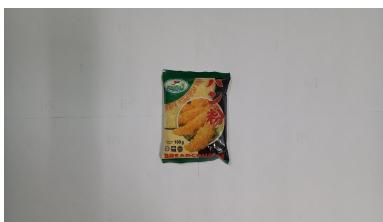
1. Xây dựng bộ dữ liệu

1.1 Thu thập dữ liệu

Nhóm tiến hành thu thập dữ liệu của 30 sản phẩm khác nhau:

UsbApacer8GB 	ButGelXoaDuocMazzicXANH 	ButGelXoaDuocMazzicTIM 
MiGoiHaoHao 	BongDenKingLed9W 	CaoSaoVang3g 
KhauTrangKF94TienHung10c 	WaxVuotTocHengKou100g 	DaugoidauRomano180g 

GoiPhoGaCungDinh	MyCayYEUL	MyLyModern
		
MienAnLienPhuHuong55g	LuongKhoHuuNghiCACAO	BanhLaysBQQ30g
		
BanhLaysBoNuong30g	BanhOreo66g	HatNemChayKnorr170g
		

HatNemAzingon170g	BotChienXuPanko100g	BotChienGionTaiKy150g
		
TuongCaCholimex270g	NuocTuongHuongViet500ml	DauHaoMaggi530g
		
LonSting320ml	FantaCamLon320ml	SaxiChuongDuongLon330ml
		
PepsiChai390ml	7UpChai390ml	SuaBichCGHL220ml
		

Tiêu chuẩn trong việc thu thập dữ liệu:

- Sản phẩm đặt trên bề mặt phẳng màu trắng có kích thước (DxR): 80cm × 60cm.
- Góc chụp từ trên xuống, vuông góc với mặt phẳng .
- Chất lượng ảnh rõ hình, không bị mờ, nhòe .
- Ánh sáng vừa đủ, không quá tối, không quá sáng, không bị chói lóa sản phẩm .
- Phải có đủ ảnh sản phẩm ở cả mặt trước và sau.
- Với sản phẩm chỉ có 1 mặt in nhãn mác thì chỉ chụp mặt có nhãn mác.
- Với sản phẩm như Lon, Chai, Lọ thì chụp rõ phần nhãn mác chính.

1.2 Gắn nhãn dữ liệu

Nhóm sử dụng công cụ có tên “LabelImg”, đây là công cụ mã nguồn mở giúp gắn nhãn hình ảnh, từ dữ liệu về nhãn đã gắn, chúng ta có thể tạo ra bộ dữ liệu dùng cho việc huấn luyện và kiểm thử các mô hình máy học. LabelImg hỗ trợ 2 định dạng dữ liệu gồm: file XML cho Pascal VOC và file TXT cho YOLO.

Nguồn tải công cụ và hướng dẫn cài đặt, hướng dẫn sử dụng có tại trang web:

<https://github.com/tzutalin/labelImg>

Nhận xét:

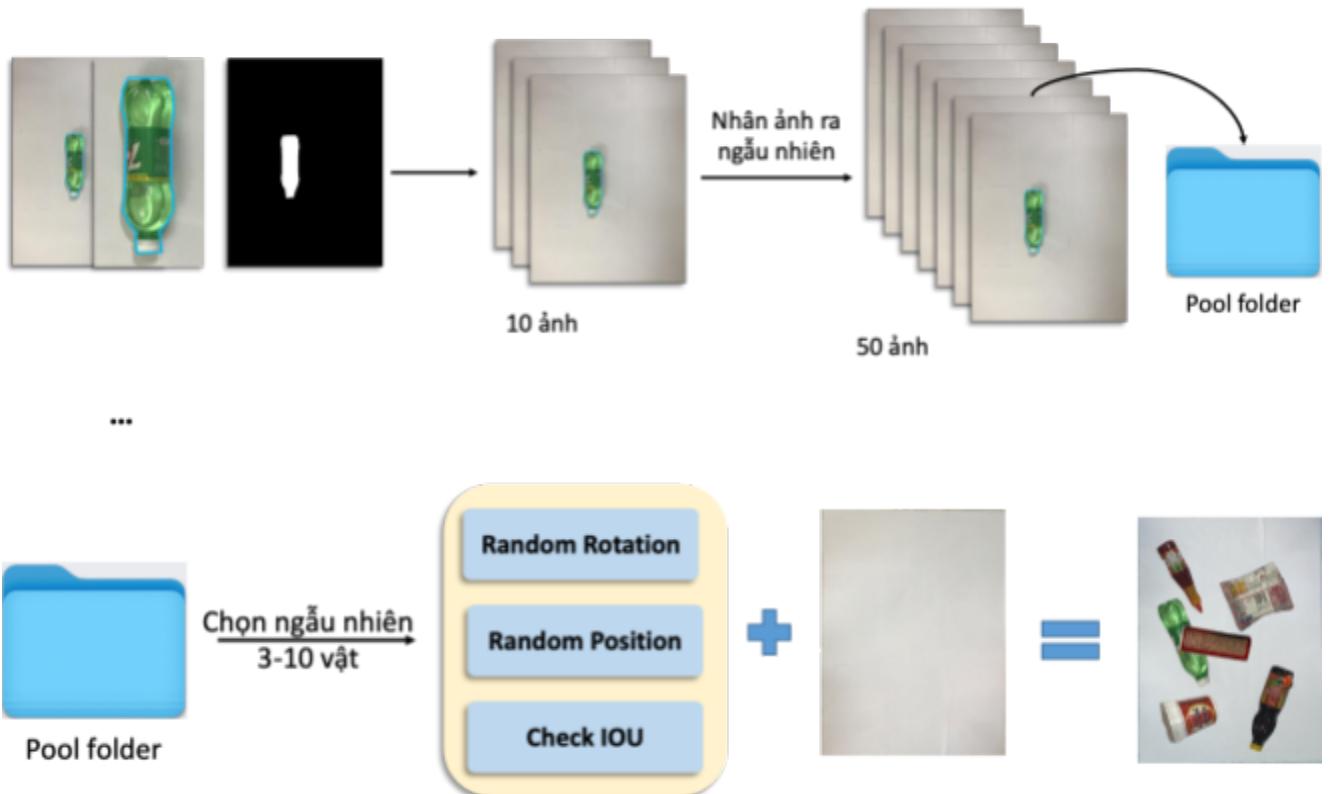
- Các ảnh trong tập Train của chúng ta chỉ chứa 1 vật trong khi tập Test lại là ảnh chứa nhiều vật (từ 5-10 vật) làm giảm khả năng dự đoán của mô hình.
- Việc chụp vật ở nhiều vị trí và xoay vật ở nhiều góc độ sẽ mất rất nhiều thời gian và tài nguyên.
- Trong quá trình thu thập dữ liệu thì một số vật có số lượng ảnh nhỏ gây nên tình trạng mất cân bằng dữ liệu.
=> ***Do đó nhóm quyết định thực hiện việc tăng cường dữ liệu để tăng độ hiệu quả cho quá trình huấn luyện mô hình.***

2. Tăng cường bộ dữ liệu

Để khắc phục các tình trạng trên chúng em sử dụng phương pháp tăng cường dữ liệu bằng cách ghép vật vào chung 1 ảnh nền như sau:

- Tạo một cái polygon bounding box để cắt vật ra khỏi background phía sau tạo ra một cái mask cho toàn bộ ảnh tập train.
- Nhận ảnh ngẫu nhiên từ trong tập dữ liệu gốc cho mỗi class là 50 ảnh rồi cho vào Pool folder.

- Chọn 3-10 vật ngẫu nhiên từ pool folder đó.
- Random Rotation:** xoay vật 360° ngẫu nhiên đồng thời điều chỉnh bounding box phù hợp. Bước này thực hiện trong 100 lần nếu Check IoU phù hợp của bước sau thì quay lại chọn lại vật ngẫu nhiên.
- Random Position:** chọn vị trí ngẫu nhiên trên ảnh background để ghép ảnh vào. Bước này thực hiện trong 100 lần nếu Check IoU phù hợp của bước sau thì quay lại bước Random Rotation
- Check IoU:** Kiểm tra IoU của vật mới thêm vào so với tất cả các vật đã ghép trong ảnh không được vượt quá ngưỡng IoU đã cho, Nếu không thỏa sẽ quay lại bước Random Position.

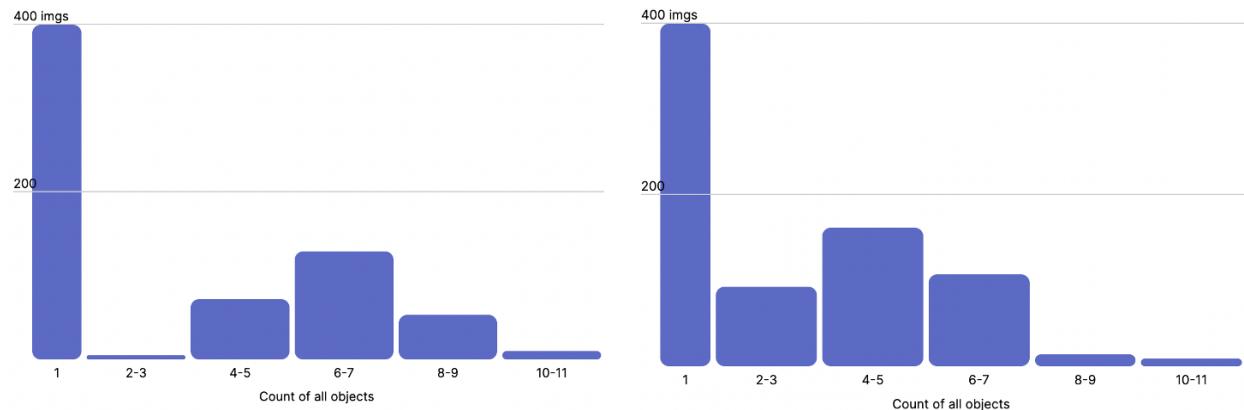


Nhận xét:

- Việc ghép nhiều vật trong một ảnh giúp cho bộ dữ liệu tăng về số lượng vật có trong ảnh, đồng thời bổ sung thêm ảnh phù hợp cho tập test và cân bằng được lượng dữ liệu cho các vật.
- Điểm yếu của phương pháp này là khi tính IoU giữa các vật thì có khả năng vật vật thể bị che khuất rất nhiều thậm chí có thể bị chắn hoàn toàn bởi vật khác làm mất thông tin của vật. Còn nếu chúng ta để IoU quá thấp thì phân bố về số lượng vật được ghép trong ảnh sẽ giảm và đồng thời các vật cũng có xu hướng cách xa nhau khiến cho bộ dữ liệu khó có thể phát hiện những vật đứng sát nhau.



Số lượng vật trong dataset sau khi tăng cường dữ liệu



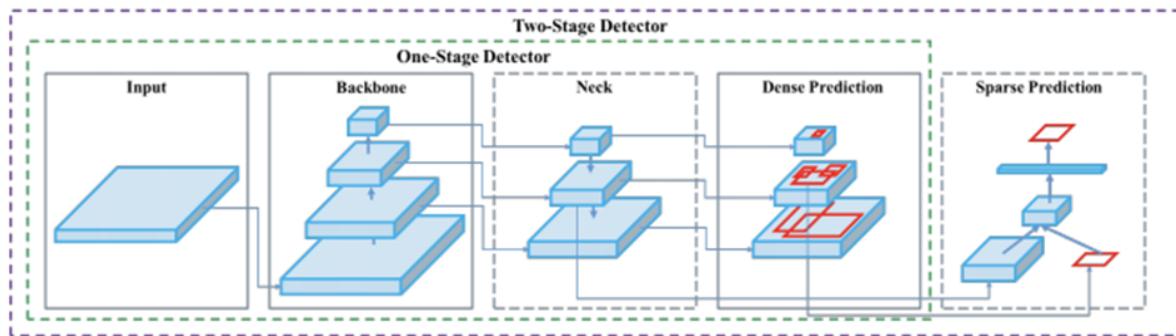
Phân bố về số lượng vật trong một ảnh khi dùng IoU 0.2 và 0.1 lần lượt như trên

III. HUÂN LUYỆN MÔ HÌNH NHẬN DIỆN YOLOv5

1. Kiến trúc của mô hình YOLOv5

1.1. Tổng quan

YOLO Là mô hình nhận dạng vật thể trích xuất đặc trưng ảnh đầu vào và đưa qua mô hình dự đoán để tạo khung đánh dấu vật thể trong ảnh và gán nhãn cho vật thể đó.



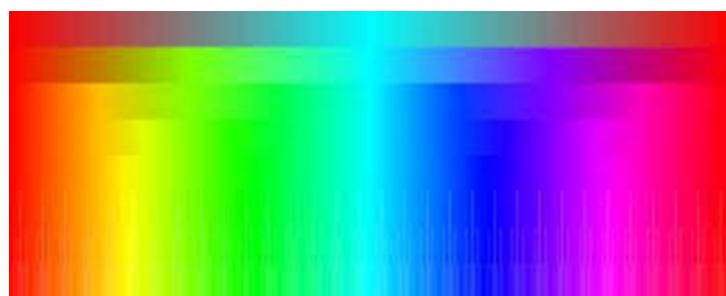
Các thành phần chính trong kiến trúc của YOLO

Mạng YOLO gồm 3 phần chính:

- **Backbone:** Chứa một mạng neural tích chập để tổng hợp và trích xuất các đặc trưng ảnh ở từng mức độ chi tiết khác nhau.
- **Neck:** Gồm chuỗi các lớp để pha trộn và kết hợp đặc trưng ảnh để chuyển tiếp đến bước dự đoán.
- **Head:** Xử lý đặc trưng và thực hiện bo khung, phân lớp vật thể.

1.2. CSP Backbone

YOLOv5 sử dụng CSP để xây dựng đặc trưng ảnh. Mô hình CSP được chọn vì nó có khả năng giải quyết vấn đề trùng lắp và thất thoát dải màu trong các mạng ConvNet Backbone khác dẫn đến việc thiếu hụt tham số và giảm hiệu suất máy tính trong quá trình train model.



Mô hình CSP dựa trên DenseNet, thiết kế để kết nối các lớp trong mạng neural tích chập nhằm giảm thiểu trùng lắp dải màu, thúc đẩy mạng lưới tái sử dụng các đặc trưng ảnh và giảm số lượng tham số trong mạng lưới. Có 3 backbone được xem xét sử dụng:

- CSPResNext50
- CSPDarknet53
- EfficientNet-B3

Backbone model	Input network resolution	Receptive field size	Parameters	Average size of layer output (WxHxC)	BFLOPs (512x512 network resolution)	FPS (GPU RTX 2070)
CSPResNext50	512x512	425x425	20.6 M	1058 K	31 (15.5 FMA)	62
CSPDarknet53	512x512	725x725	27.6 M	950 K	52 (26.0 FMA)	66
EfficientNet-B3 (ours)	512x512	1311x1311	12.0 M	668 K	11 (5.5 FMA)	26

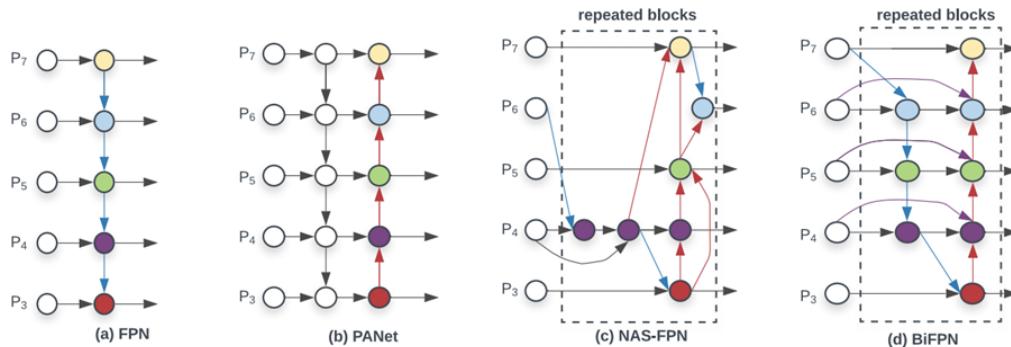
Có thể thấy EfficientNet vượt trội hơn các mạng khác có cùng kích thước về phân loại hình ảnh. Tuy nhiên, tác giả của YOLOv5 đã chọn mạng CSPDarknet53 làm backbone mà không đưa ra lý do cụ thể.

1.3. PA-Net Neck

Neck có nhiệm vụ trộn và kết hợp các features map đã học được thông qua quá trình trích xuất đặc trưng backbone và quá trình nhận dạng.

YOLO cho phép tùy biến sử dụng các cấu trúc cho phần Neck như:

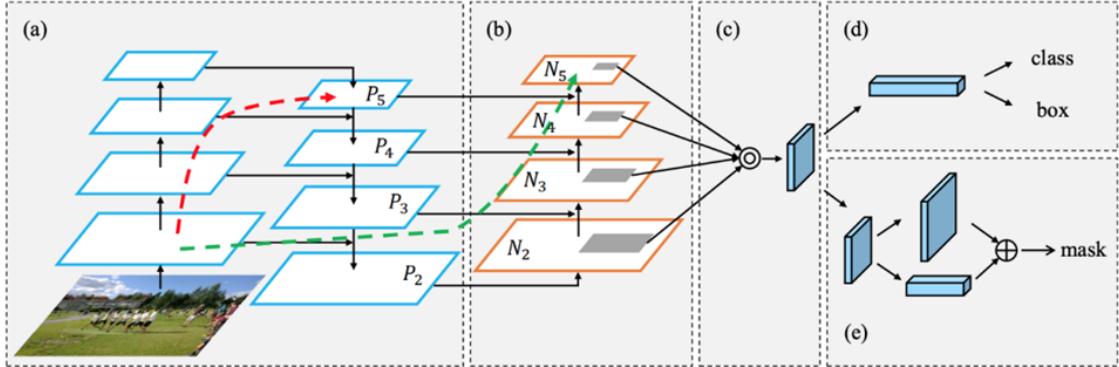
- FPN
- PA-Net
- NAS-FPN
- BiFPN



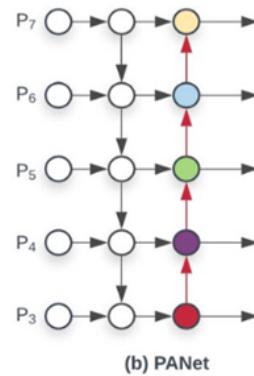
Mỗi một phần tử P_i nêu trên đại diện cho một lớp đặc trưng trong CSP Backbone

YOLOv5 dùng PA-Net để tổng hợp các đặc trưng.

PA-Net là một cải tiến của FPN nhằm cải thiện thông tin cục bộ trên các top layers.

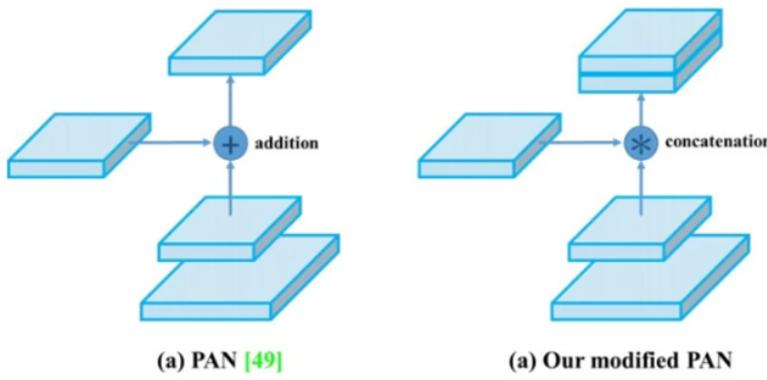


PA-Net có thêm 1 bottom-up pathway, trong đó, mỗi layer lấy đầu vào là feature maps của stage trước đó, đi qua một ma trận convolution 3×3 . Đầu ra được add với feature map của top-down pathway với stage tương ứng. Thiết kế của Neck có thể được mô tả như hình dưới:



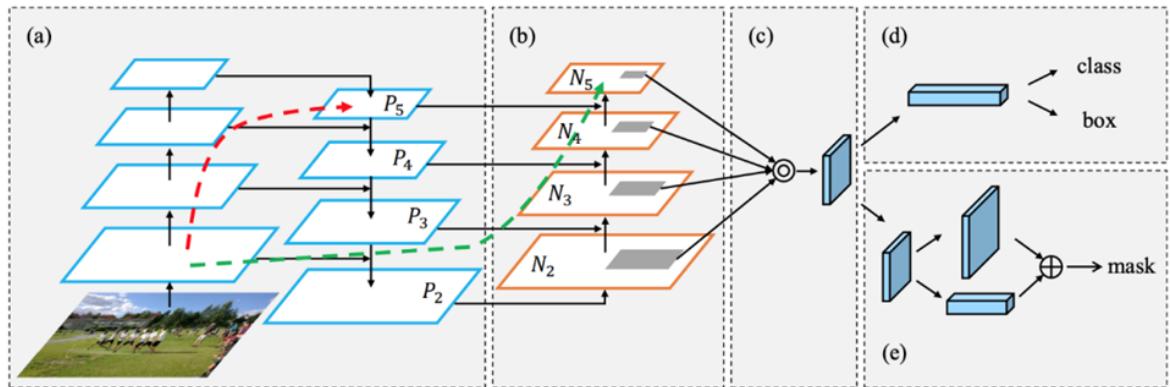
(b) PANet

Trong YOLOv5, hàm add được chỉnh sửa thành hàm concat:



Tại các stage của augmented bottom-up pathway, object được detect một cách độc lập với các kích thước khác nhau. Điều này có thể dẫn tới sự dư thừa về dữ liệu hoặc tại mỗi stage sẽ không sử dụng thông tin từ các stage khác. Do đó, tại mỗi stage, các feature maps sẽ được đẩy qua một mạng SPP (=ROIAlign), rồi sau đó đưa qua lớp fully

connected layers, kết quả thu được sau các lớp fully connected layer này sẽ được element-wise max operation để thu được prediction.



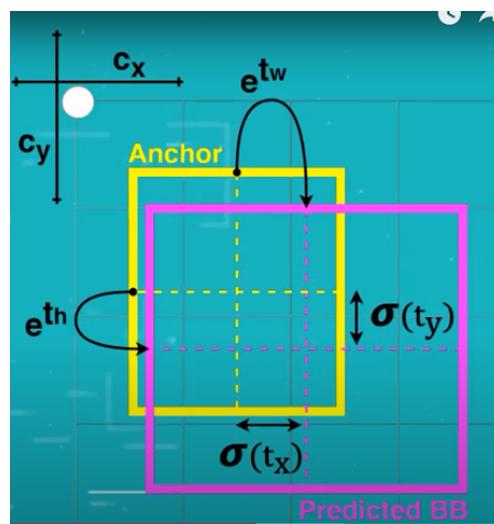
1.4. Head – Bước nhận dạng

YOLOv5 sử dụng Head với các anchor box để nhận dạng vật thể có kích thước khác nhau trong ảnh.

Anchor box thực ra là các bounding box nhưng được tạo sẵn (chứ không phải kết quả của quá trình nhận dạng – prediction).

Với mỗi ô lưới sẽ có 9 anchor box khác nhau với kích thước:

- Ô lưới 13×13 : $(116 \times 90), (156 \times 198), (373 \times 326)$.
- Ô lưới 26×26 : $(30 \times 61), (62 \times 45), (59 \times 119)$.
- Ô lưới 52×52 : $(10 \times 13), (16 \times 30), (33 \times 23)$.



Trong quá trình đào tạo sẽ kết hợp với thuật toán K-Means cluster và Genetic Learning. Cùng với ground truth để tính ra sai sót giữa ground truth và anchor box bằng cách điều chỉnh các giá trị x,y,w,h, từ đó học được các đặc điểm của vật thể. Phân phối các kích thước và vị trí bounding box có thể rất khác so với bounding box đã được điều chỉnh sẵn trong bộ dữ liệu COCO.

Sự khác biệt giữa bounding box tự học và có sẵn trong bộ dữ liệu COCO thể hiện rõ khi phát hiện vật thể có kích thước chiều dài lớn nhưng chiều rộng nhỏ.

Tất cả bounding box trong YOLOv5 đều tự học để bo khung với mọi bộ dữ liệu bất kì.

2. Huấn luyện mô hình YOLOv5-s

YOLOv5 đề xuất 4 phiên bản với kích thước khác nhau:

- YOLOv5-s: Small - Cỡ nhỏ
- YOLOv5-m: Medium - Cỡ vừa
- YOLOv5-l: Large - Cỡ lớn
- YOLOv5-x: Extra-large - Cực lớn

Việc lựa chọn phiên bản YOLOv5 tùy thuộc vào độ phức tạp của bài toán, phức tạp hơn thì dùng phiên bản lớn hơn và ngược lại.

Trong đồ án này, sau khi khảo sát và thảo luận, chúng em quyết định sử dụng phiên bản YOLOv5-s để huấn luyện, lý do là vì dữ liệu không quá phức tạp, gồm ảnh vật thể đặt trên nền trắng, nên có khả năng đạt độ chính xác cao. Ngoài ra, khi áp dụng thực tế, thì cần một hệ thống không những chính xác mà còn gọn nhẹ và nhanh, cho nên phiên bản YOLOv5-s là sự lựa chọn hợp lý trong trường hợp này.

Quá trình huấn luyện, chúng em sử dụng template training YOLOv5-s của **Roboflow**, huấn luyện model với **batch_size** = 64, **image_size** = 416x416, **epoch** = 1100. Thời gian huấn luyện là **1 giờ 7 phút 32 giây**.

III. THỰC NGHIỆM VÀ ĐÁNH GIÁ MÔ HÌNH

1. Các thông số đánh giá

1.1. IoU – Intersection over Union

Là phần giao trên phần hợp giữa Bounding Box mà mô hình dự đoán được và Bounding Box thực tế (Ground-Truth).

Chỉ số IoU dùng làm ngưỡng để xác định True-Positive, False-Positive và False-Negative.

Thường thì ngưỡng IoU là 0.5:

- Nếu $\text{IoU} > 0.5$: xác định đối tượng đã được nhận diện đúng $\rightarrow \text{TP}$
- Nếu $\text{IoU} < 0.5$: xác định đối tượng đã nhận diện sai $\rightarrow \text{FP}$
- Nếu $\text{IoU} = 0.5$: tùy vào xác định của người thiết kế mô hình mà đưa vào **TP** hay **FP**
- **FN** cho trường hợp không nhận diện được dữ liệu lại cần nhận diện.

1.2. Precision (P) và Recall (R)

$$\begin{aligned}Precision &= \frac{TP}{TP + FP} = \frac{\text{Số dự đoán đúng}}{\text{Tổng số dự đoán}} \\Recall &= \frac{TP}{TP + FN} = \frac{\text{Số lần dự đoán đúng}}{\text{Số lần nhận dạng có thể có}}\end{aligned}$$

Precision: đo độ chính xác của dự đoán, tức tỉ số số lần dự đoán đúng trên tổng lần dự đoán.

➤ Precision đại diện cho độ tin cậy của mô hình nhận diện.

Recall: đánh giá khả năng tìm ra được Ground-truth bounding box, tính bằng cách lấy tỉ số số lần dự đoán chính xác trên số lần nhận dạng đúng có thể có.

➤ Recall đại diện cho độ nhạy của model.

1.3. Average Precision (AP)

Ta có thể đánh giá mô hình dựa trên việc thay đổi một ngưỡng và quan sát giá trị Precision và Recall.

Giả sử có N ngưỡng tính Precision và Recall, mỗi ngưỡng cho cặp giá trị P_k và R_k ($k = 1, 2, \dots, N$), AP được xác định bằng:

$$AP = \sum_{k=0}^{k=N-1} (R_k - R_{k+1})P_k$$

1.4 Mean Average Precision (mAP)

mAP là trung bình AP tính cho tất cả các lớp. Mỗi quan hệ giữa Precision và Recall giúp mAP đánh giá được độ chính xác của mô hình.

Precision – Recall thay đổi khi ngưỡng IoU thay đổi cho nên với các giá trị IOU khác nhau, ta có thể tính được các độ chính xác khác nhau của mô hình.

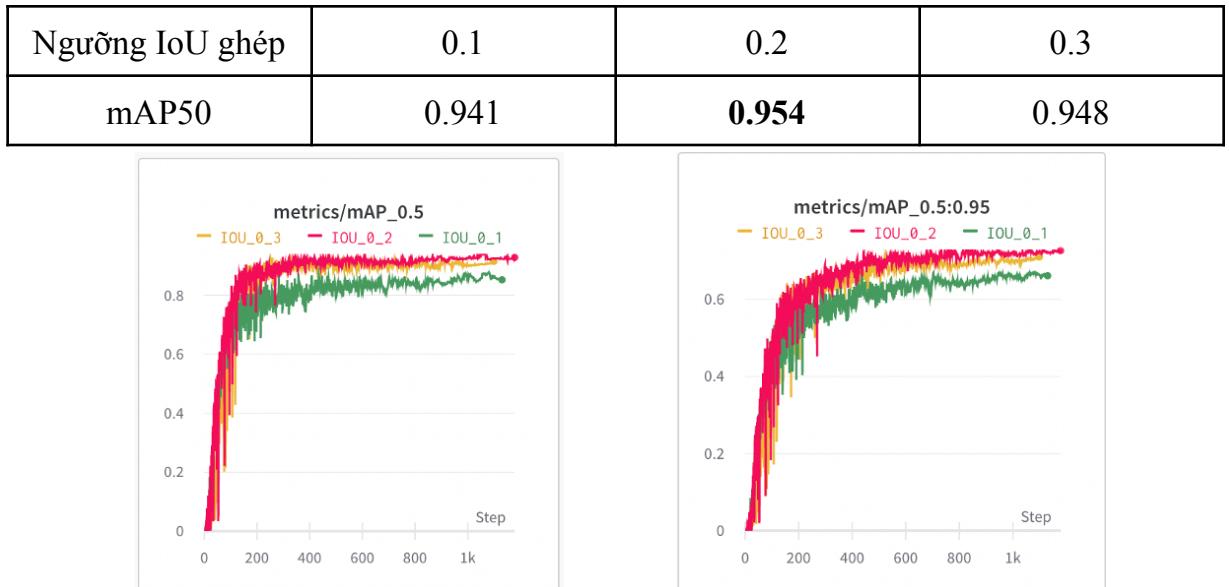
2. Kết quả thực nghiệm

- Mô hình YOLOv5-s được train với 1100 epochs, batch_size là 64, image_size là 416 được train trong khoảng 1h30p.
- Ta được kết quả như sau:

	Không ghép	Ghép	Ghép + Không ghép
mAP50	0.44	0.809	0.954

Nhận xét:

- Từ bảng trên ta có thể thấy việc ghép nhiều vật trong một ảnh giúp cải thiện khả năng dự đoán của mô hình rất đáng kể từ 0.44 lên 0.809.
- Ở đây ta cũng có thể thấy rằng nếu ta vừa dùng ảnh Ghép lẫn ảnh Không ghép để train thì sẽ được kết quả tốt nhất có thể là do việc chỉ sử dụng ảnh ghép thì các bounding box của một số vật này bị dính với vật khác khiến cho mô hình không học được đúng thông tin về vật vì thế nếu chúng ta sử dụng thêm ảnh không ghép sẽ bổ sung thông tin cho vật giúp mô hình dự đoán tốt hơn.



Biểu đồ kết quả mAP của tập val trong quá trình train

Nhận xét:

- Từ bảng và biểu đồ trên ta có thể thấy được nếu ta để ngưỡng IoU khi ghép ảnh quá thấp thì các vật sẽ không đứng sát nhau nên sẽ khó dự đoán ra các ảnh test có những vật sát nhau hay chồng chất lên nhau.
- Còn nếu để ngưỡng IoU quá cao thì ảnh sau khi ghép sẽ mất nhiều thông tin của vật gây khó khăn trong quá trình dự đoán.

Ngoài ra nhóm em có so sánh kết quả giữa YOLOv4-Tiny (Phiên bản rút gọn của YOLOv4) vs YoLOv5-s trong cùng bộ dữ liệu không ghép kết hợp có ghép (ngưỡng IoU = 0.2) như sau:

	Yolov4 tiny	Yolov5s
mAP50	0.847	0.954

Có thể thấy, cùng là phiên bản nhỏ gọn nhưng YOLOv5-s cho kết quả tốt hơn so với YOLOv4-Tiny. Cộng thêm ưu điểm của mình là thời gian huấn luyện nhanh, YOLOv5-s sẽ là một mô hình thích hợp cho bài toán nhận diện sản phẩm bán lẻ.

3. Những trường hợp Detect sai của Yolov5



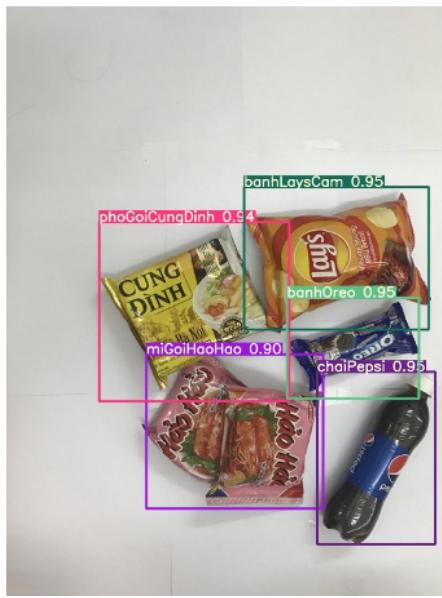
(Nhận diện 2 lon Sting)



(Nhận diện lon Xá xị Chưong Dương thành lon Sting)

Nhận xét:

- Việc nhận diện 2 lon Sting gần nhau rất có thể là do khi ta ghép các vật lại với nhau thì khi 2 vật sát nhau nó sẽ học luôn thông tin của vật cạnh đó đồng thời 1 phần của vật cũng bị che khuất nên gây ra tình trạng trên. Ta có thể tăng ngưỡng confident để loại bỏ bounding box tránh lặp dữ liệu.
- Còn việc nhận diện sai lon Xá xị thành lon Sting khả năng cao là bộ dataset gốc không chụp đủ góc độ của lon xá xị và lon Sting đồng thời màu sắc lẫn hình dạng 2 vật này rất giống nhau từ 2 điều đó dẫn đến mô hình dự đoán sai.

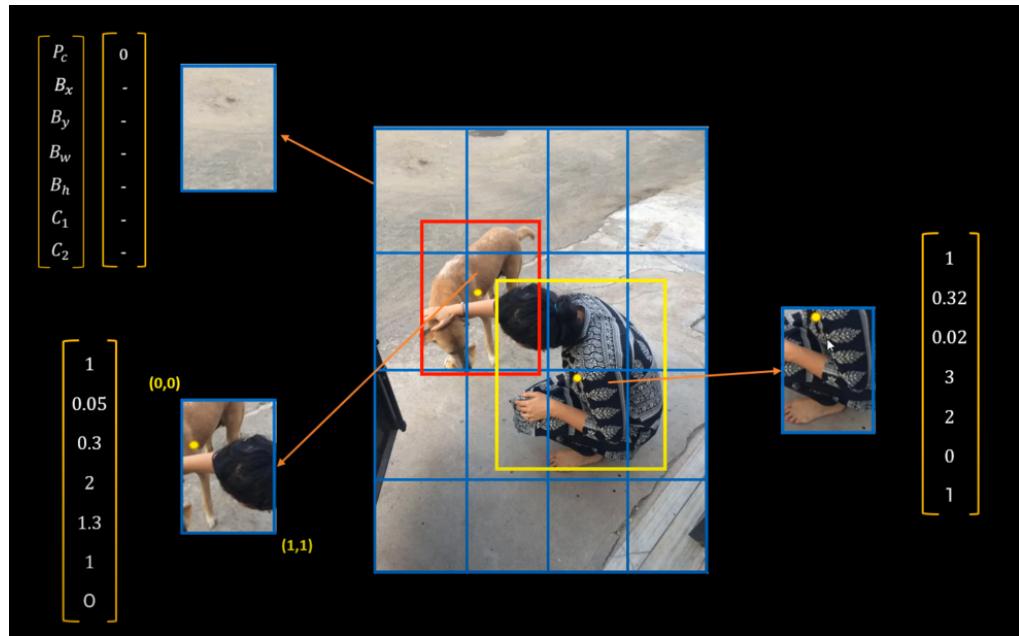


Nhận diện 2 vật sát nhau thành 1 vật (2 gói mì hảo hảo và 2 cây bút xanh)



Nhận xét:

- Việc bị dính 2 vật sát nhau khả năng là do đặc tính của mô hình với các tính năng liên quan đến anchor box mà hai vật giống nhau đứng kế nhau thường bị nhầm lẫn là 1 vật



- Ta có thể thấy rằng mỗi bounding box hay anchor box đều được mô hình hóa bằng một vector đặc trưng cho các chi tiết về tọa độ tâm cũng như kích thước của bounding box.

- Trong quá trình huấn luyện model bounding box sẽ thực hiện điều chỉnh các kích thước của khung box và tọa độ bounding box dựa theo tọa độ tâm và khung của anchor box để dự đoán mô hình. Với 2 vật giống nhau về đặc điểm nhưng đặt sát nhau, mô hình sau quá trình huấn luyện và nhận biết sẽ điều chỉnh tọa độ tâm của 2 vật về gần nhau hơn và cuối cùng là trùng nhau.

- Điều này cũng có thể lý giải bằng một tình huống thực tiễn khi con người quan sát nhiều vật đặt xa nhưng gần sát nhau làm ta liên tưởng rằng tập hợp vật đó là 1 vật đồng nhất. Tuy nhiên ta vẫn nhận biết được đó là nhiều vật vì ta đã quan sát và học tập từ quan sát đó nhiều lần trong khi mô hình YOLO lại áp dụng một phương pháp chỉ quan sát vật một lần duy nhất dẫn đến tình trạng nhầm lẫn với những vật giống nhau đặt gần nhau.

IV. TÔNG KẾT

Từ quá trình thực nghiệm, chúng em nhận thấy YOLOv5-s vẫn tồn tại nhược điểm đến từ chính kiến trúc của nó gây ảnh hưởng đến việc phát hiện vật thể, do đó chúng em đã đưa ra giải pháp ban đầu trong việc sử dụng mô hình (áp dụng đối với người sử dụng mô hình) đó là: *Trước khi tiến hành quét sản phẩm cần sắp xếp các mặt hàng sao cho không bị dính liền nhau hoặc che lấp nhau, điều này cũng giống như việc kiểm hàng trước khi thanh toán ở các phương thức khác.*

Trong phạm vi đồ án này, chúng em đã nghiên cứu áp dụng mô hình phát hiện vật thể YOLOv5 cho phương pháp quét sản phẩm sử dụng Camera và Máy học. Với kết quả của mô hình có độ chính xác cao, thời gian dự đoán nhanh và thời gian huấn luyện mô hình nhanh, chúng em kỳ vọng việc ứng dụng phương pháp quét sản phẩm này sẽ khả thi trong tương lai.

Phương hướng phát triển trong tương lai của chúng em đó là nghiên cứu thuật toán phát hiện vật thể dưới dạng đa giác Polygon và Bounding Box có định hướng (Oriented Bounding Box) nhằm hướng đến mục tiêu tăng độ chính xác cũng như giảm sự sai lệch đến từ mô hình YOLOv5 nguyên bản.

V. TÀI LIỆU THAM KHẢO

1. Glenn Jocher, May 2020| YOLOv5 |URL: <https://github.com/ultralytics/yolov5>
2. Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, April 2020 | “YOLOv4: Optimal Speed and Accuracy of Object Detection” | URL: <https://arxiv.org/abs/2004.10934v1>
3. tzutalin | LabelImg |URL: <https://github.com/tzutalin/labelImg>
4. Roboflow | Models Library | URL: <https://models.roboflow.com/>

—HẾT—