

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN**  
**CHỦ ĐỀ:**  
**NHẬN DIỆN SẢN PHẨM BÁN LẺ**

Giảng viên hướng dẫn: **ThS. Phạm Nguyễn Trường An**

**TS. Lê Đình Duy**

Sinh viên thực hiện: **19522245 – Võ Nhật Thanh**

**19522351 – Trần Trung Tín**

**19522174 – Nguyễn Ngọc Tân**

**TP. HỒ CHÍ MINH – NĂM 2021-2022**

## I. Giới thiệu

### 1.1. Mô tả bài toán

Trong cuộc sống hiện đại, việc xếp 1 hàng dài để chờ thanh toán trong các siêu thị, trung tâm thương mại khiến nhiều người khó chịu và mất rất nhiều thời gian, làm cho các doanh nghiệp tổn thất rất nhiều. Trong đó việc quét mã vạch là một trong những công đoạn tốn nhiều thời gian nhất. Đặc biệt trong thời gian dịch bệnh covid-19 đang hoành hành gần đây thì việc thanh toán chậm làm ùn tắc, người xếp hàng tập trung đông là rất nguy hiểm.



Hình 1. Mọi người đang tập trung đông trong siêu thị

Nhận thấy những nguy hiểm và những rủi ro rất lớn nên nhóm chúng em đã cùng nhau tìm những phương pháp để giải quyết. Chúng em nhận thấy được việc công nghệ nhận diện hình ảnh ngày càng phát triển và các dữ liệu hình ảnh sản phẩm nhiều thì nhóm chúng em xin đề xuất 1 giải pháp tăng tốc quá trình thanh toán như sau:

- + Sử dụng camera hoặc điện thoại được gắn vào phía trên của quầy thanh toán, sau đó nhận diện hình ảnh của các món hàng trong một khu vực được quy định trong thời gian thực hiện liên kết với màn hình của quầy. Sau đó sẽ phân chia ra các món hàng nào nhận diện được món nào không nhận diện được. Đối với các món không nhận diện được vẫn sẽ tính tiền theo kiểu quét mã vạch.

Với giải pháp này của chúng em thì các cửa hàng, siêu thị sẽ giảm được đáng kể thời gian thanh toán, tránh được ùn tắc quầy thu ngân ở những khung giờ cao điểm. Cửa hàng sẽ tận dụng được tối đa các cơ sở vật chất sẵn có để tăng trải nghiệm người đi mua hàng.

### 1.2. Input, Output

- **INPUT:** Là một bức ảnh từ trên chiếu xuống quầy thanh toán với ánh sáng trắng rõ ràng trong đó bao gồm nhiều món hàng không xếp chồng lên nhau trong một khu vực được định sẵn bằng một cái khay hay đường viền màu được chụp từ một camera hoặc điện thoại.
- **OUTPUT:** Là bức ảnh từ input nhưng có các đường viền vuông bao quanh các món hàng có màu xanh, là món hàng nhận diện được và trên các thanh trên cùng được gắn nhãn tên của món đồ.

### 1.3. Mô tả bộ dữ liệu

Nhóm chúng em có tìm hiểu được một số bộ dữ liệu được thu thập sẵn có liên quan đến đề án của chúng em như là:

- + **SOIL-47:** Là tập dữ liệu sản phẩm tập trung vào việc thử nghiệm các thuật toán nhận dạng đối tượng dựa trên màu sắc, nó chứa 47 danh mục sản phẩm với 21 hình ảnh cho mỗi danh mục, được chụp từ 20 chế độ xem khác nhau. Hai bộ những hình ảnh như vậy được chụp trong các điều kiện ánh sáng khác nhau để các thuật toán kiểm tra yêu cầu cường độ chiếu sáng bất biến.



Hình 2. Một số hình ảnh từ bộ dữ liệu Soil-47

- + **Grozi-120:** Là tập dữ liệu được đề xuất cho các cửa hàng tạp hóa nhận biết trong môi trường tự nhiên. Nó chứa 120 tạp hóa danh mục sản phẩm. Đối với mỗi loại sản phẩm, có 2 loại hình ảnh một loại hình ảnh được thu thập từ web, loại hình ảnh khác được thu thập bên trong một cửa hàng tạp hóa. Tổng cộng 11.870 hình ảnh được thu thập với 676 từ web và 11.194 từ cửa hàng.

1		2		3		4		5	
6		7		8		9		10	
11		12		13		14		15	
16		17		18		19		20	
21		22		23		24		25	
26		27		28		29		30	
31		32		33		34		35	

### Hình 3. Một số sản phẩm trong tập dữ liệu Grozi-120

Nhận thấy các tập dữ liệu có sẵn chủ yếu phần lớn là các sản phẩm của nước ngoài, mẫu mã khác biệt nên không thể áp dụng được cho các cửa hàng, siêu thị ở Việt Nam. Vì vậy, nhóm quyết định chỉ tham khảo các bộ dữ liệu và tự đi thu thập một bộ dữ liệu riêng biệt.

Các sản phẩm nhóm dự định thu thập là các sản phẩm bán lẻ ở các cửa hàng tạp hóa lớn, siêu thị, thân quen với mọi người như là: nước ngọt các loại, các loại bánh kẹo, các loại đồ mỹ phẩm, các vật dụng cá nhân,....

**Khó khăn:** Nhưng do tình hình dịch bệnh covid-19 đang diễn biến, tình hình các vùng của các thành viên trong nhóm đang phức tạp và hạn chế ra ngoài nên việc ra ngoài để thu thập data là việc cực kì khó khăn. Nên bộ dữ liệu mà nhóm thu thập được chủ yếu là các sản phẩm có sẵn trong nhà của mỗi thành viên dẫn đến bộ dữ liệu bị hạn chế.

## II. Các nghiên cứu trước

### 2.1. Bài báo “Deep learning for Retail Product Recognition: Challenges and Techniques”

Tác giả: Yuchen Wei, Son Tran, Shuxiang Xu, Byeong Kang and Matthew Springer.

Họ sử dụng nhiều phương pháp trên nhiều bộ dữ liệu khác nhau, các phương pháp được sử dụng:

- + Pháp pháp cổ điển: Nhận dạng sản phẩm được thực hiện bằng cách trích xuất các tính năng trên hình ảnh của bao bì
- + Deep Learning
- + Convolutional Neural Networks
- + Deep Learning for Object Detection + Product
- + Recognition Based on Deep Learning

Các bộ dữ liệu họ đã sử dụng:

- + Grozi-120
- + DS2 dataset
- + RPC dataset
- + Cigarette Dataset
- + Grocery Store Dataset
- + Grozi-3.2k

Kết quả đạt được:

- + Với tập dữ liệu RPC

Clutter mode	Methods	cAcc (↑)	ACD (↓)	mCCD (↓)	mCIoU (↑)	mAP50 (↑)	mmAP (↑)
Easy	Single	0.02%	7.83	1.09	4.36%	3.65%	2.04%
	Syn	18.49%	2.58	0.37	69.33%	81.51%	56.39%
	Render	63.19%	0.72	0.11	90.64%	96.21%	77.65%
	Syn+Render	<b>73.17%</b>	<b>0.49</b>	<b>0.07</b>	<b>93.66%</b>	<b>97.34%</b>	<b>79.01%</b>
Medium	Single	0.00%	19.77	1.67	3.96%	2.06%	1.11%
	Syn	6.54%	4.33	0.37	68.61%	79.72%	51.75%
	Render	43.02%	1.24	0.11	90.64%	95.83%	72.53%
	Syn+Render	<b>54.69%</b>	<b>0.90</b>	<b>0.08</b>	<b>92.95%</b>	<b>96.56%</b>	<b>73.24%</b>
Hard	Single	0.00%	22.61	1.33	2.06%	0.97%	0.55%
	Syn	2.91%	5.94	0.34	70.25%	80.98%	53.11%
	Render	31.01%	1.77	0.10	90.41%	95.18%	71.56%
	Syn+Render	<b>42.48%</b>	<b>1.28</b>	<b>0.07</b>	<b>93.06%</b>	<b>96.45%</b>	<b>72.72%</b>
Averaged	Single	0.01%	12.84	1.06	2.14%	1.83%	1.01%
	Syn	9.27%	4.27	0.35	69.65%	80.66%	53.08%
	Render	45.60%	1.25	0.10	90.58%	95.50%	72.76%
	Syn+Render	<b>56.68%</b>	<b>0.89</b>	<b>0.07</b>	<b>93.19%</b>	<b>96.57%</b>	<b>73.83%</b>

Experimental results of the ACO task on RPC dataset

+ Với tập dữ liệu D2S

Approaches	mAP
Mask R-CNN	78.3
FCIS	68.3
Faster R-CNN	78.0
RetinaNet	80.1

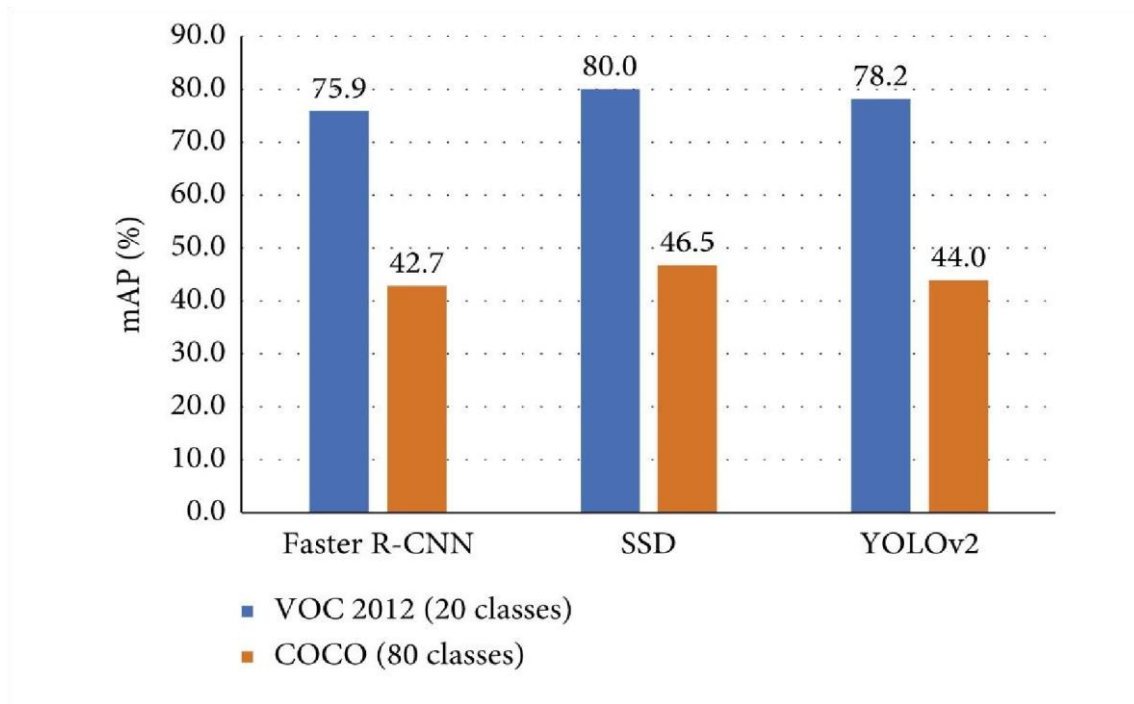
## 2.2. Những thách thức của bài toán

**Phân loại quy mô lớn:** Số lượng các sản phẩm riêng biệt cần được xác định trong siêu thị có thể rất lớn, khoảng vài nghìn, đối với một cửa hàng tạp hóa quy mô vừa vượt xa khả năng thông thường của máy dò đối tượng.

Hiện tại, YOLO, SSD, Faster R-CNN và Mask R-CNN là các phương pháp phát hiện đối tượng hiện đại, đánh giá thuật toán của chúng bằng PASCAL bộ dữ liệu VOC và MS COCO. Tuy nhiên, PASCAL VOC chỉ chứa 20 lớp đối tượng và MS COCO chứa ảnh của 80 loại đối tượng.

Điều này có ý nghĩa là các thiết bị phát hiện đối tượng hiện đại không thích hợp để áp dụng trực tiếp vào nhận dạng sản phẩm bán lẻ do những hạn chế của chúng đối với các danh mục có quy mô lớn. Dưới đây là bảng so sánh kết quả trên bộ kiểm tra VOC 2012 (20 loại đối tượng) và COCO (80 loại đối tượng) với các thuật toán khác nhau, bao gồm Faster R-CNN, SSD và YOLOv2.





**Giới hạn dữ liệu:** Các phương pháp tiếp cận dựa trên học sâu đòi hỏi phải một lượng lớn dữ liệu được chú thích để đào tạo, đặt ra một thách thức đáng kể trong những trường hợp chỉ có một số lượng nhỏ các ví dụ.

Các công cụ gán nhãn hình ảnh, yêu cầu lao động thủ công để gán nhãn mọi đối tượng trong mỗi hình ảnh. Thông thường, có ít nhất hàng chục nghìn hình ảnh đào tạo trong một tập dữ liệu phát hiện đối tượng chung, rõ ràng cho thấy rằng việc tạo một tập dữ liệu với đủ dữ liệu đào tạo cho học sâu là công việc tốn nhiều thời gian.

**Các sản phẩm có tính chất thủy tinh:** Do sự tương đồng trực quan về hình ảnh, màu sắc, văn bản và kích thước hệ mét giữa các sản phẩm nội thủy tinh, các sản phẩm bán lẻ thực sự khó được xác định, việc máy tính phân loại các sản phẩm nội thủy tinh này sẽ rất phức tạp.

**Tính linh hoạt:** Nhìn chung, với số lượng sản phẩm mới ngày nhiều, các cửa hàng tạp hóa cần thường xuyên nhập các mặt hàng mới để thu hút khách hàng. Hơn nữa, sự xuất hiện của các sản phẩm hiện có thường xuyên thay đổi theo thời gian. Do những lý do trên, một hệ thống ghi nhận thực tế nên linh hoạt mà không cần hoặc không phải đào tạo lại bất cứ khi nào một sản phẩm/gói sản phẩm mới được giới thiệu. Tuy nhiên, mạng nơ-ron tích tụ luôn bị “catastrophic forgetting”, chúng không thể nhận ra một số đối tượng đã học trước đó khi thích nghi với một nhiệm vụ mới.



Hình 5. Các hãng liên tục thay về ngoài sản phẩm và ra mắt nhiều loại mới

### III. Xây dựng bộ dữ liệu

#### 3.1. Tại sao cần thu thập dữ liệu thủ công

Do các dữ liệu có sẵn không đáp ứng được ngữ cảnh của bài toán như: background xung quanh sản phẩm, sản phẩm nội địa Việt Nam, ...

Việc tự thu thập dữ liệu giúp kiểm soát được các yếu tố ngoại cảnh như góc quay, sản phẩm, ánh sáng,... tùy vào đó mà đưa ra các tiêu chí để thu thập dữ liệu.

#### 3.2. Các tiêu chí thu thập dữ liệu

Với ngữ cảnh bài toán là chụp một bức ảnh từ trên xuống trong môi trường siêu thị ánh sáng rõ ràng vì vậy yêu cầu đặt ra là cố gắng mô phỏng được khoảng cách từ camera đến sản phẩm, ánh sáng, góc quay cụ thể như sau:

- + Vị trí camera điện thoại cách sản phẩm từ 20-30 cm
- + Các vật phải nằm trong một khu vực giới hạn ( đối với tập test)
- + Background màu trắng
- + Các sản phẩm có nhiều tư thế có thì phải lấy được tất cả các tư thế của vật
- + Ưu tiên lấy các mặt có màu sắc logo rõ ràng
- + Đảm bảo điều kiện ánh sáng trong phòng tốt, mô phỏng càng giống ánh sáng trong siêu thị càng tốt
- + Góc camera 45~80 độ
- + Độ phân giải tối thiểu là 1280x720



Hình 6. Ảnh được chụp từ khoảng cách 30cm có background trắng, ánh sáng rõ ràng, với 2 góc chụp khác nhau và 2 tư thế khác nhau của vật

### 3.3. Cách thức thu thập

Chuẩn bị các thiết bị:

- + Một cái giá treo điện thoại
- + Một chiếc điện thoại chụp ảnh với độ phân giải tối thiểu 720p
- + Một nguồn sáng trong phòng tối
- + Một cục sạc dự phòng cho điện thoại

Cách thức thu thập dữ liệu train:

- + Chuẩn bị ánh sáng, góc chụp, tư thế của vật
- + Chụp từ khoảng cách 20-30 cm
- + Sau đó thay đổi tư thế của vật và góc chụp rồi tiếp tục chụp
- + Chụp mỗi loại khoảng 50 ảnh với các tư thế và góc chụp khác nhau

Cách thức thu thập dữ liệu test:



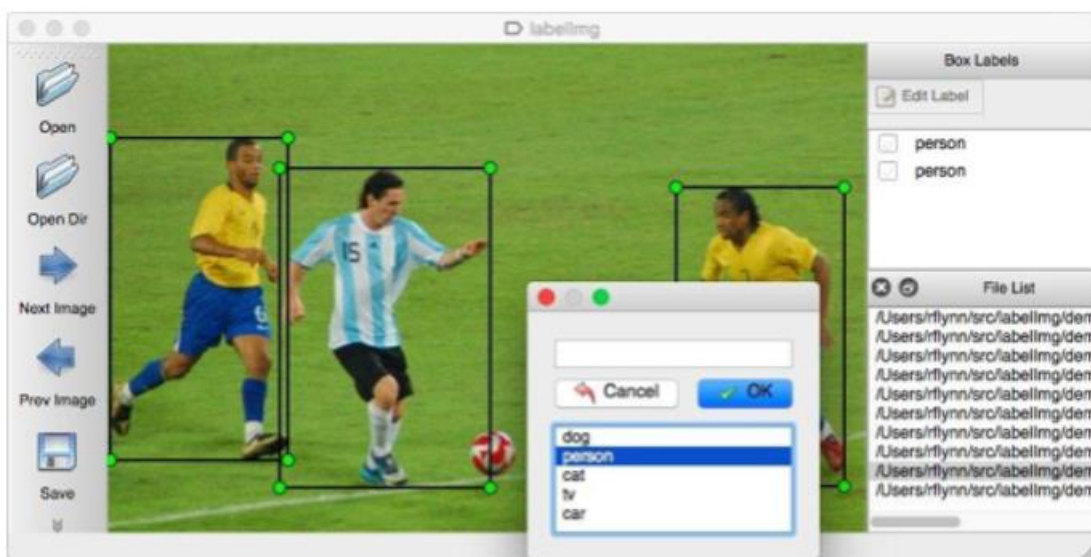
- + Chuẩn bị một background trắng giới hạn 30x30cm, góc chụp từ trên xuống, ánh sáng trong phòng tốt
- + Sau đó bỏ 2-6 vật không chồng lên nhau vào một hình rồi chụp lại



Hình 7. Một số ảnh từ tập test

### 3.4. Gán nhãn dữ liệu

Sau khi tổng hợp được dữ liệu thô thì chúng em bắt đầu gán nhãn, chúng em sử dụng tool Labellmg.



Hình 9: Tool Label

### 3.5. Tổng quan bộ dữ liệu

Bộ dữ liệu gồm 2 tập dữ liệu là tập train, test và valid của 145 class là các sản phẩm tiêu dùng trong siêu thị như: mì, bánh kẹo, nước ngọt, đồ dùng cá nhân, ... Bộ dữ liệu gồm 9709 ảnh.

Bộ dữ liệu dùng để train bao gồm:

- + 8455 ảnh chỉ có một vật được chụp từ 145 sản phẩm
- + trung bình mỗi vật có 58 ảnh
- + Mỗi ảnh chỉ có một vật, vậy có tổng cộng 8455 vật được gán nhãn

Bộ dữ liệu test bao gồm: 630 ảnh tự chụp từ 3-10 vật trong mỗi ảnh.

Bộ dữ liệu vali bao gồm: 630 ảnh tự chụp từ 3-10 vật trong mỗi ảnh.



Hình 10. Bộ dữ liệu đa dạng về hình dạng



Hình 11. Các vật có hình dạng giống nhau



Hình 12. Vật có hình dạng giống nhau nhưng khác màu sắc

**Nhận xét:**

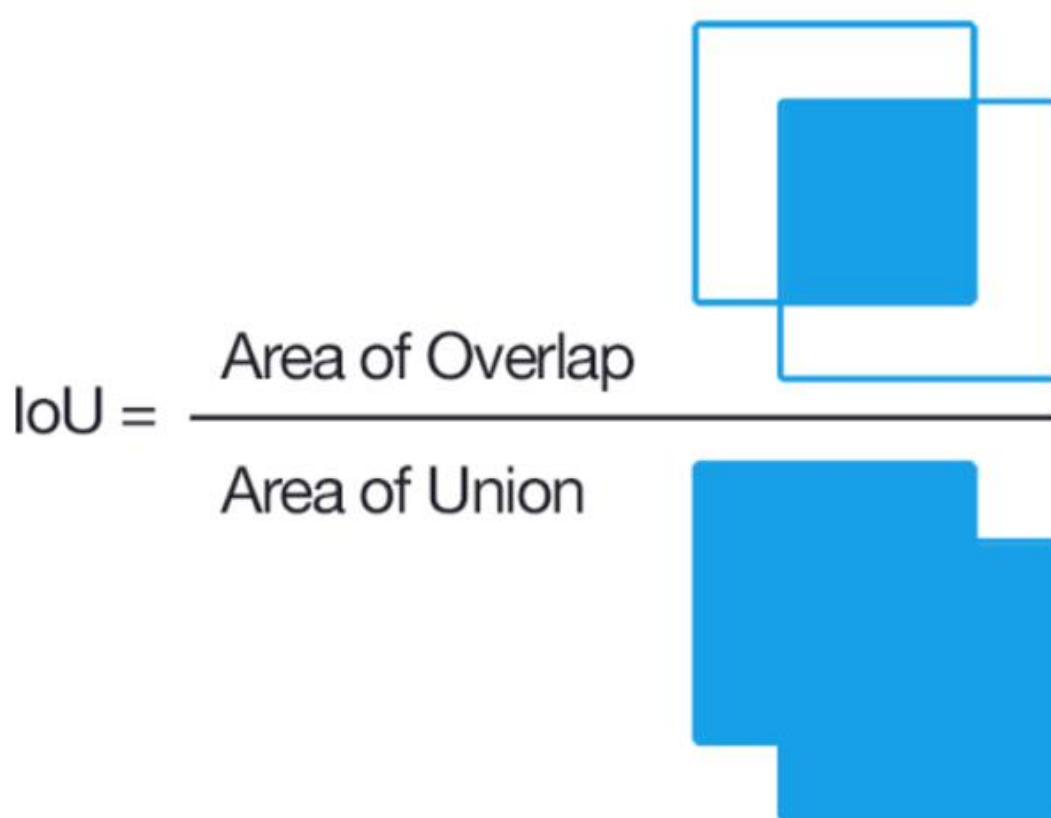
- + Bộ dữ liệu do chúng em thu thập thủ công nên khá sạch, số lượng class trung bình không quá lớn chỉ ngang với các nghiên cứu trước (150 so với 120 của Grozi-120 và 200 của RPC)

#### IV. Training và đánh giá Model

##### 4.1. Phương thức đánh giá Model

Để đánh giá model Object detection người ta sử dụng các thông số như IoU, AP, mAP,....

IoU là độ đo overlap giữa Ground-truth-bounding box là đường bao mà ta gán nhãn với Predicted bounding box là đường bao mà model dự đoán


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Hình 13. Công thức tính IoU

Precision: đo lường mức độ chính xác là dự đoán của mô hình tức là tỷ lệ phần trăm dự đoán của mô hình là chính xác.

Recall: đo lường như thế nào tốt mô hình tìm thấy tất cả các mẫu tích cực.

AP: là chỉ số có quan hệ mật thiết với chỉ số Precision (phần trăm bounding box được dự đoán đúng) và Recall (tỷ lệ phần trăm các bounding box được đoán đều chính xác).

AP: là độ chính xác với IoU = 0.5



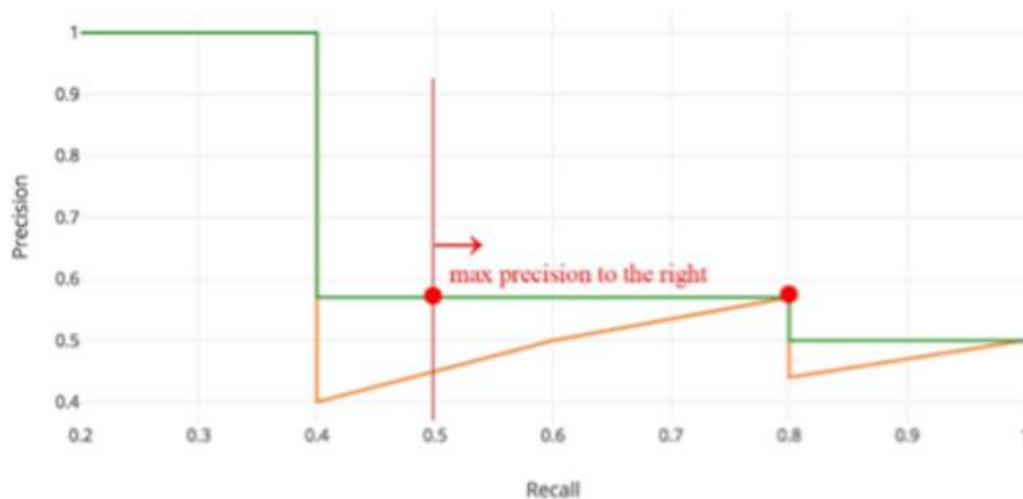
AP75: là độ chính xác với IoU = 0.75

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

Hình 14. Cách tính Precision và Recall.

Trong đó:

- True Positive (TP): đối tượng ở lớp Positive
- True Negative (TN): đối tượng ở lớp Negative
- False Positive (FP): đối tượng ở lớp Positive
- False Negative (FN): đối tượng ở lớp Negative



Hình 15. Các tính AP dựa trên Precision và Recall

Chỉ số mAP là trung bình tổng chỉ số AP của tất cả các class.

#### 4.2. YOLOv5

Shortly after the release of YOLOv4 Glenn Jocher introduced YOLOv5 using the Pytorch framework.

The open source code is available on GitHub

Author: Glenn Jocher

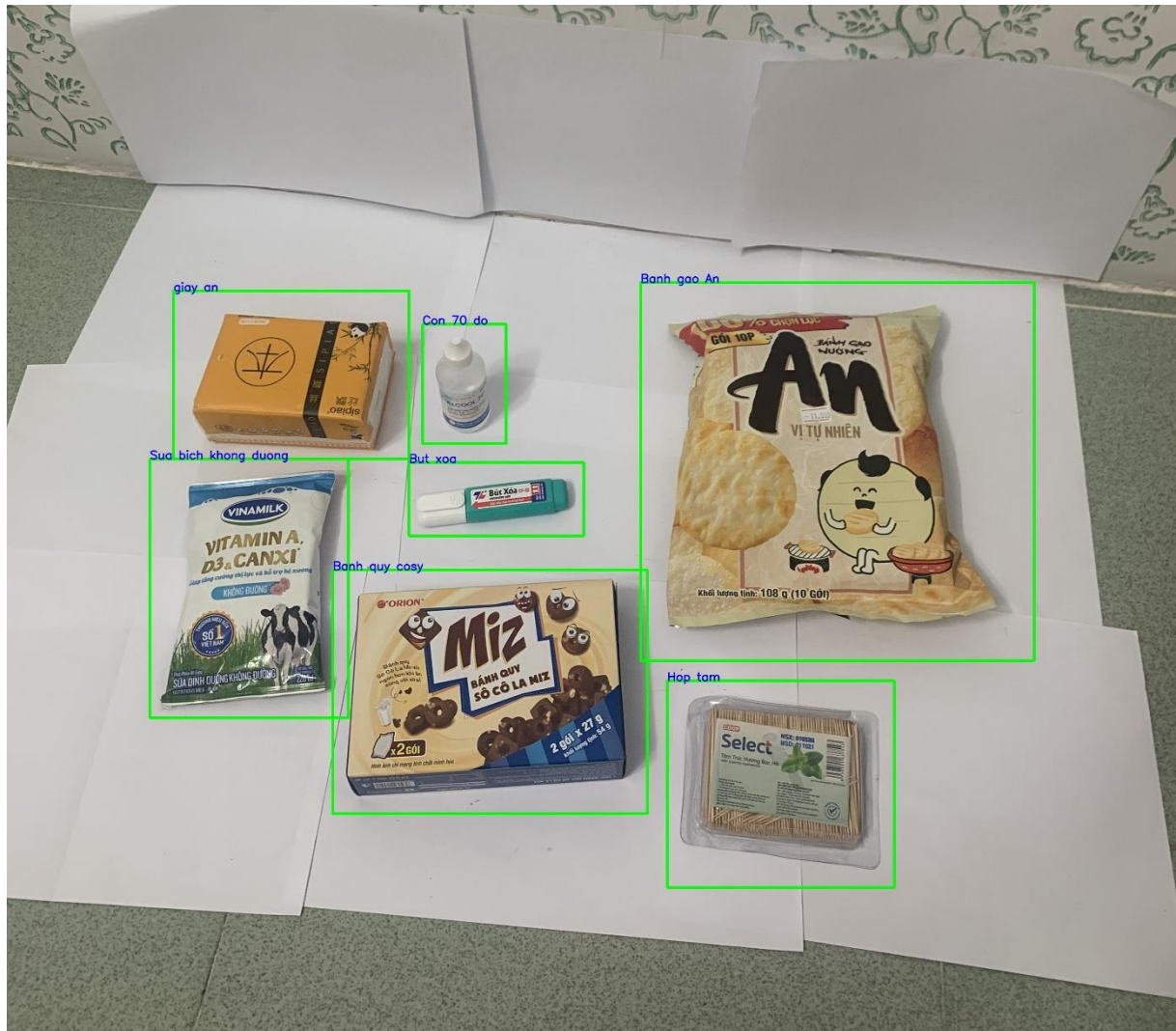
Released: 18 May 2020

Train thử 2000 ảnh 40 classes

- Kết quả từ tập val được lấy từ tập train khi train tới **6000 iteration** thì:
  - + **AP50** đạt khoảng : **90,5%**



+ Kết quả trên tập test : **mAP = 67,15 %**



Hình 16: kết quả

## V. Ứng dụng và hướng phát triển

Hiện nay qua việc khảo sát nhóm nhận thấy việc khách hàng chờ đợi thanh toán trong các cửa hàng tiện lợi và siêu thị vào các khung giờ cao điểm tốn rất nhiều thời gian, bên cạnh đó dịch covid19 đang diễn biến rất phức tạp thì việc giảm thiểu thời gian thanh toán càng được chú trọng.

Đầu tư về thiết bị để thu thập dữ liệu nên tạo một không gian riêng để dành cho chụp hình các sản phẩm để tăng khả năng nhận biết sản phẩm lúc thanh toán một cách tốt nhất.

Liên kết với ngân hàng hoặc ví điện tử để thanh toán và nhận hóa đơn ngay trên điện thoại, hạn chế nhận hóa đơn giấy như hiện nay.

## VI. Tài liệu tham khảo

SOIL-47                      Surrey                      Object                      Image                      Library:  
<https://www.ee.surrey.ac.uk/CVSSP/demos/colour/soil47>

Grozi-120 Database: <http://grozi.calit2.net/grozi.html>

Deep Learning for Retail Product Recognition: Challenges and Techniques:  
<https://www.hindawi.com/journals/cin/2020/8875910>

LabelImg is a graphical image annotation tool and label object bounding boxes in images: <https://github.com/tzutalin/labelImg>

mAP (mean Average Precision):  
<https://dothanblog.wordpress.com/2020/04/24/map-mean-average-precision>

Bounding boxes augmentation for object detection:  
[https://albumentations.ai/docs/getting\\_started/bounding\\_boxes\\_augmentation](https://albumentations.ai/docs/getting_started/bounding_boxes_augmentation)

Làm thế nào để đánh giá một mô hình máy học?  
<http://tutorials.aiclub.cs.uit.edu.vn/index.php/2021/05/18/evaluation/>