

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
PHÒNG TN NC&PT XE TỰ HÀNH



BÁO CÁO THỰC TẬP DOANH NGHIỆP
ĐỀ TÀI:

NGHIÊN CỨU QUY TRÌNH XÂY DỰNG CƠ SỞ
DỮ LIỆU VÀ MÔ HÌNH MÁY HỌC NHẬN BIẾT
CÁC LOẠI VẬT NUÔI

GVHD: TS. ĐOÀN DUY

Nhóm thực hiện:

- 1. SV1: DƯƠNG LÂM KHA - MSSV: 16520546**
- 2. SV2: BÙI TẤN DUY - MSSV: 16520277**

TP. Hồ Chí Minh , ngày 30 tháng 12 năm 2021

LỜI CẢM ƠN

Thực tập doanh nghiệp là một môn học mà sinh viên khoa Kỹ Thuật Máy Tính bắt buộc phải trải qua trong chương trình đào tạo. Đây là một học giúp sinh viên có thể tiếp cận với quy trình cũng như cách làm việc của một công ty một cách chân thực nhất. Tuy nhiên, dưới sự ảnh hưởng của đại dịch covid 19 thì việc có thể tìm kiếm một công ty thực tập bên ngoài là rất khó khăn. Nên nhóm chúng em rất cảm ơn Khoa Kỹ Thuật Máy Tính đã tạo điều kiện để chúng em có thể thực tập thuận lợi hơn trong môi trường Phòng Tn Nc&Pt Xe Tự Hành.

Trong quá trình thực tập, nhóm chúng em cũng đã gặp rất nhiều khó khăn. Tuy nhiên, cũng nhờ sự trợ giúp tận tình của các thầy/cô trong khoa mà chúng em đã có thể giải quyết từng khó khăn và hoàn thành khóa thực tập doanh nghiệp. Chúng em xin được chân thành cảm ơn các thầy cô rất nhiều.

Tiếp theo là nhóm xin được cảm ơn sự hỗ trợ trực tiếp của thầy hướng dẫn TS. Đoàn Duy. Thầy đã cho nhóm hướng đi cũng như cách giải quyết các bài toán trong khóa thực tập. Nhóm chúng em xin được cảm ơn thầy rất nhiều.

Bên cạnh đó, nhóm chúng em xin được cảm ơn các bạn sinh viên khác cùng thực tập trong Phòng Tn Nc&Pt Xe Tự Hành đã có những giúp đỡ cho nhóm có thể giải quyết và khắc phục một số vấn đề gặp phải.

Cuối cùng, nhóm xin được chúc quý thầy cô thật dồi dào sức khỏe, niềm tin để tiếp tục hoàn thành sứ mệnh cao đẹp là truyền đạt kiến thức cho thế hệ mai sau, luôn gặt hái được thành công trong sự nghiệp trồng người của mình.

Tp. Hồ Chí Minh, tháng 12 năm 2021

Sinh viên thực hiện

Dương Lâm Kha

Bùi Tấn Duy

Mục lục

LỜI CẢM ƠN.....	1
Danh mục bảng biểu.....	4
Danh mục hình ảnh.....	5
LỜI MỞ ĐẦU	6
CHƯƠNG 1: GIỚI THIỆU	7
1.1. GIỚI THIỆU ĐỀ TÀI.....	7
1.1.1. Lý do thực hiện đề tài.....	7
1.1.2. Mục tiêu tổng quát.....	7
1.1.3. Mục tiêu cụ thể	7
1.1.4. Nội dung dự kiến thực hiện	7
1.2. PHƯƠNG PHÁP THỰC HIỆN	8
1.3. Ý NGHĨA CỦA ĐỀ TÀI.....	9
1.4. GIỚI HẠN VÀ KẾT QUẢ MONG MUỐN	9
CHƯƠNG 2. NỘI DUNG CHÍNH	10
2.1 TÌM HIỂU VỀ MẠNG NƠ RON TÍCH CHẬP - CONVOLUTIONAL NEURAL NETWORKS.....	10
2.1.1 Mạng Nơ Ron – Neural Networks.....	10
2.1.2. Mạng nơ ron tích chập – Convolution neural networks.....	10
2.1.3. Kết quả thực hiện được	11
2.2. OBJECT DETECTION VÀ THUẬT TOÁN YOLO.....	11
2.2.1. Một vài chỉ số đánh giá model object detection.....	12
2.2.2. Thuật toán YOLO	14
2.2.3. Kết quả thực hiện được	15
2.3. PHÂN TÍCH HỆ THỐNG.....	15
2.3.1. Hệ thống nhận diện vật nuôi	15

2.3.2. Kết quả thực hiện.....	16
2.4. XÂY DỰNG TẬP DỮ LIỆU	17
2.4.1. Tổng quan tập dữ liệu – dataset.....	17
2.4.2. Những loại dữ liệu trong dataset	17
2.4.3. Gán nhãn dữ liệu	19
2.4.4. Kết quả thực hiện được	20
2.5. HUẤN LUYỆN YOLOV4.....	20
2.5.1. Quá trình huấn luyện	20
2.5.2. Đánh giá model trên tập validation	21
2.5.3. Đánh giá model trên tập test.....	23
CHƯƠNG 3. KẾT QUẢ ĐẠT ĐƯỢC VÀ CÁC KHÓ KHĂN	25
3.1. ƯU ĐIỂM.....	25
3.2. HẠN CHẾ	25
3.3. KẾT QUẢ ĐẠT ĐƯỢC.....	25
3.4. KHÓ KHĂN.....	25
GÓP Ý VỀ CHƯƠNG TRÌNH ĐÀO TẠO CỦA KHOA	26
TÀI LIỆU THAM KHẢO	27

Danh mục bảng biểu

Bảng 1. Confusion matrix	13
Bảng 2. Thông số về dataset	17
Bảng 3. Average Precision cho 5 lớp	22
Bảng 4. Chi tiết các thông số khi validate	22
Bảng 5. Độ chính xác trung bình trên tập Test.....	23
Bảng 6. Chi tiết các thông số khi testing	24

Danh mục hình ảnh

Hình 1. Minh họa mạng nơ ron network	10
Hình 2. Cấu trúc mạng nơ ron tích chập.....	11
Hình 3. Minh họa cách tính IoU	12
Hình 4. Mô tả cách tính precision và recall.....	13
Hình 5. Kết quả so sánh FPS và AP của YOLOv4 với các kiến trúc khác	15
Hình 6. Hệ thống nhận diện vật nuôi.....	16
Hình 7. Hình ảnh trong dataset chụp ngang vật nuôi.....	17
Hình 8. Hình ảnh trong dataset chụp chính diện vật nuôi.....	18
Hình 9. Hình ảnh trong dataset chụp các góc nghiêng của vật nuôi.....	18
Hình 10. Hình ảnh trong dataset chụp vật nuôi bị che một phần cơ thể.....	18
Hình 11. Hình ảnh trong dataset chụp phía sau vật nuôi	19
Hình 12. Hình ảnh trong dataset có thể có nhiều object	19
Hình 13. Minh họa gán nhãn dữ liệu bằng LabelIMG.....	19
Hình 14. Training loss trong quá trình training.....	20
Hình 15. Giá trị của mAP của model trên tập validate	21
Hình 16. Minh họa nhận diện mèo	24
Hình 17. Minh họa nhận diện chó	24
Hình 18. Minh họa nhận diện gà.....	24
Hình 19. Minh họa nhận diện ngựa	24
Hình 20. Minh họa nhận diện sóc	24

LỜI MỞ ĐẦU

Với đề tài: “NGHIÊN CỨU QUY TRÌNH XÂY DỰNG CƠ SỞ DỮ LIỆU VÀ MÔ HÌNH MÁY HỌC NHẬN BIẾT CÁC LOẠI VẬT NUÔI” thì nhóm sẽ chia ra các công việc bao gồm xây dựng cơ sở dữ liệu và xây dựng một mô hình máy học có thể phát hiện và nhận diện các loài vật nuôi. Với tính chất của bài toán thì chúng em quyết định sẽ xây dựng một mô hình object detection sử dụng thuật toán YOLOv4.

Để có thể hiểu và hiện thực một model YOLOv4 thì chúng em sẽ bắt đầu tìm hiểu các kiến thức cơ bản bao gồm mạng nơ ron tích chập, lý thuyết về thuật toán YOLO cũng như các phiên bản của YOLO. Sau đó, chúng em sẽ tiến hành huấn luyện model trên Google Colab, thực hiện tính toán điều chỉnh các thông số để có thể thu được một model tốt nhất trên tập dữ liệu đã chuẩn bị.

Cuối cùng, nhóm sẽ sử dụng model YOLOv4 đã được và tiến hành testing ở thời gian thực. Đây là bước kiểm chứng cuối cùng để đánh giá chất lượng, tốc độ cũng như độ chính xác của model.

Dựa vào những nội dung trên và những kết quả đã thực hiện được trong quá trình thực tập thì nội dung của báo cáo thực tập sẽ gồm những nội dung sau:

CHƯƠNG 1: GIỚI THIỆU

CHƯƠNG 2: NỘI DUNG CHÍNH

CHƯƠNG 3: KẾT QUẢ ĐẠT ĐƯỢC VÀ CÁC KHÓ KHĂN

CHƯƠNG 1: GIỚI THIỆU

1.1. GIỚI THIỆU ĐỀ TÀI

1.1.1. Lý do thực hiện đề tài

Hiện nay, Deep Learning đang là một trong những công nghệ phổ biến với rất nhiều ứng dụng hữu ích trong cuộc sống. Deep learning đã đem đến những giải pháp hiệu quả cho các doanh nghiệp. Việc sử dụng các mạng thần kinh nhân tạo vào các mô hình AI tiên tiến đã được nghiên cứu và sử dụng rất nhiều trên toàn thế giới. Với mong muốn học và tìm hiểu chi tiết các bước để xây dựng một mô hình máy học cụ thể. Nhóm đã chọn và đăng kí đề tài: “NGHIÊN CỨU QUY TRÌNH XÂY DỰNG CƠ SỞ DỮ LIỆU VÀ MÔ HÌNH MÁY HỌC NHẬN BIẾT CÁC LOẠI VẬT NUÔI”.

1.1.2. Mục tiêu tổng quát

- Xây dựng được một tập dữ liệu gồm có 5 loại vật nuôi.
- Huấn luyện mô hình trên tập dữ liệu đã chuẩn bị để có thể nhận biết được các loại vật nuôi với độ chính xác lớn hơn hoặc bằng 90%.

1.1.3. Mục tiêu cụ thể

- Xây dựng được cơ sở dữ liệu cho mô hình.
- Huấn luyện model để model có thể phát hiện và nhận diện được các loại vật nuôi từ một bức ảnh, một video hoặc từ một camera.
- Đưa ra được quy trình xây dựng cơ sở dữ liệu và mô hình học máy để nhận biết vật nuôi.

1.1.4. Nội dung dự kiến thực hiện

- Nghiên cứu, tìm hiểu và lựa chọn dữ liệu cho mô hình. Dữ liệu bao gồm các ảnh chụp có vật nuôi.
- Nghiên cứu, tìm hiểu về mạng CNN và thuật toán Object detection YOLOv4.
- Huấn luyện mô hình.
- Kiểm tra và đánh giá mô hình.

1.2. PHƯƠNG PHÁP THỰC HIỆN

Nội dung 1: Nghiên cứu, tìm hiểu và lựa chọn dữ liệu cho mô hình. Dữ liệu bao gồm các ảnh chụp có vật nuôi:

- Nghiên cứu tiêu chuẩn chọn ảnh.
- Chọn 5 loại vật nuôi. Tìm ảnh của vật nuôi để xây dựng cơ sở dữ liệu phục vụ cho việc huấn luyện model.
- Các nguồn lấy ảnh: github, kaggle và các nguồn khác trên internet.
- Đưa vào labimg để gán nhãn.

Nội dung 2: Nghiên cứu, tìm hiểu về mạng CNN và thuật toán YOLOv4:

- Sử dụng các nguồn thông tin internet, youtube để tìm hiểu, nghiên cứu về các thành phần của CNN và cách CNN hoạt động.
- Nghiên cứu, tìm hiểu về thuật toán YOLOv4 để phục vụ cho quá trình huấn luyện mô hình.

Nội dung 3: Huấn luyện mô hình:

- Tiến hành huấn luyện mô hình trên Google Colab.
- Theo dõi các chỉ số loss, accuracy trong quá trình training để đánh giá và lựa chọn model tốt nhất.
- Validate trên tập validation để đánh giá độ chính xác trên tập dữ liệu validation.
- Testing trên tập test để kiểm chứng lại độ chính xác.
- Nếu độ chính xác còn thấp hoặc mô hình không ổn định thì xem xét các khả năng làm ảnh hưởng đến chất lượng của model như: thêm data, kiểm tra xem nhãn có bị gán sai hay không, v.v.
- Lặp lại các bước cho đến khi thu được mô hình có độ chính xác lớn hơn hoặc bằng 90%.

Nội dung 4: Kiểm tra và đánh giá model:

- Kiểm tra và đánh giá model trên tập dữ liệu validation và tập test.
- Demo trên một ảnh, một video hoặc từ camera.

1.3. Ý NGHĨA CỦA ĐỀ TÀI

Việc xây dựng được một mô hình máy học có thể phát hiện và nhận diện các loài vật nuôi sẽ là tiền đề để phát triển những ứng dụng cho lĩnh vực này như tìm kiếm vật nuôi bị lạc hay ngăn ngừa những vụ tai nạn giao thông do những loài thú gây ra.

1.4. GIỚI HẠN VÀ KẾT QUẢ MONG MUỐN

Vì đề tài được hiện thực trong quá trình thực tập và môi trường học tập nên nhóm muốn giới hạn lại đề tài để phục vụ cho quá trình học tập và tìm hiểu.

Sau khi tham khảo ý kiến của thầy hướng dẫn là TS. Đoàn Duy thì nhóm xin giới hạn lại phạm vi của đề tài như sau:

- Model được xây dựng để phát hiện và nhận diện 5 lớp tương ứng với 5 loại vật nuôi bao gồm: mèo, chó, gà, ngựa, sóc.
- Tập dữ liệu của mỗi lớp khoảng 100 ảnh (80% training, 20% để validation).
- Độ chính xác dự kiến của mô hình là lớn hơn hoặc bằng 90%.

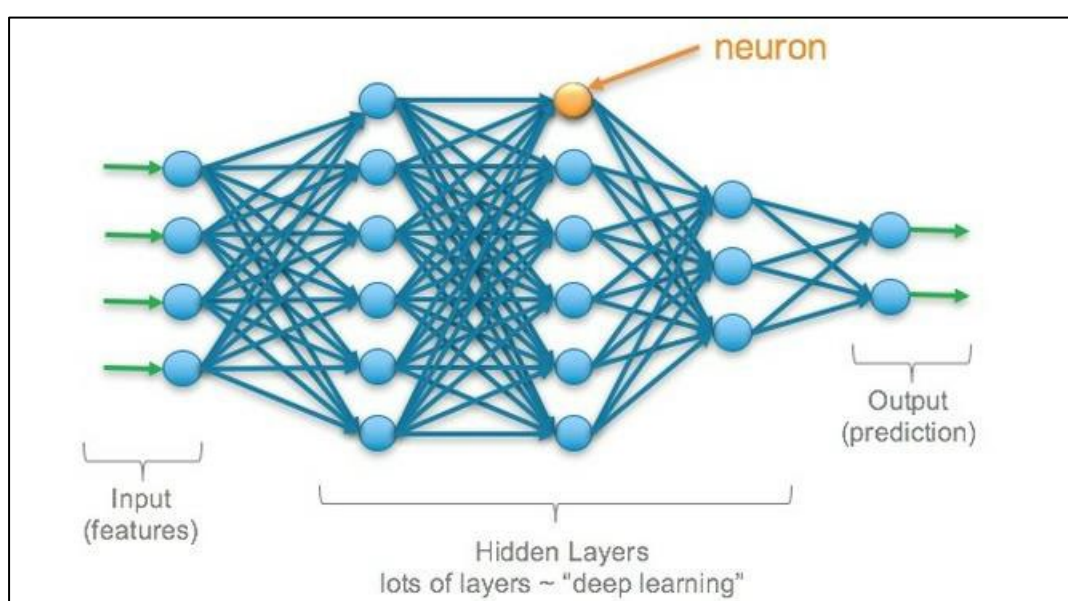
Mặt khác, sau khi đã làm quen với việc xây dựng data và đưa data và huấn luyện model thì nhóm quyết định tăng số lượng ảnh của tập dữ liệu. Lượng dữ liệu quá ít sẽ làm cho model có nhiều sai sót và kém hiệu quả. Do đó, nhóm đã liên tục cập nhật và bổ sung một lượng data nhất định sau mỗi tuần làm việc. Thống kê chi tiết về tập dữ liệu sẽ được cập nhật ở bên dưới.

CHƯƠNG 2. NỘI DUNG CHÍNH

2.1 TÌM HIỂU VỀ MẠNG NƠ RON TÍCH CHẬP - CONVOLUTIONAL NEURAL NETWORKS

2.1.1 Mạng Nơ Ron – Neural Networks

Neural networks (mạng nơ ron) mô phỏng lại mạng nơ ron sinh học, là một cấu trúc khối gồm các đơn vị tính toán đơn giản được liên kết chặt chẽ với nhau, trong đó các liên kết giữa các neuron quyết định chức năng của mạng, được minh họa trên hình 1.2. Mô hình này có khả năng học được các khuôn mẫu phức tạp của dữ liệu qua các lớp (layer) nơ ron có nhiệm vụ biến đổi dữ liệu theo các công thức toán học.



Hình 1. Minh họa mạng nơ ron network

Ở mạng lưới này, lớp đầu tiên sẽ dùng để đưa các đặc tính của vật cần dự đoán vào, được gọi là input layer. Và layer cuối cùng sẽ mang kết quả dự đoán, được gọi là output layer. Các layer ở giữa 2 lớp này được gọi là hidden layer (lớp ẩn). Một mô hình mạng neural có thể có hoặc không có các hidden layer.

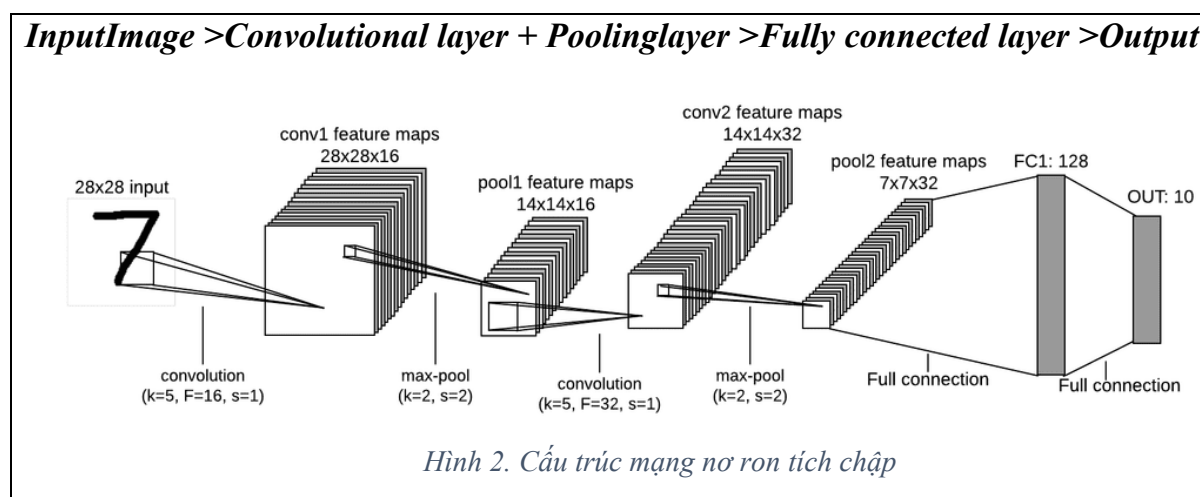
2.1.2. Mạng nơ ron tích chập – Convolution neural networks

Convolution neural networks (CNN) là một trong những kiến trúc thành công nhất và được sử dụng rộng rãi trong cộng đồng deep learning, đặc biệt là cho các tác vụ thị giác máy tính. CNN giúp chúng ta xây dựng được các hệ thống thông minh với độ

chính xác rất cao. Hiện nay, CNN được ứng dụng rất nhiều trong những bài toán nhận dạng object trong ảnh.

Kiến trúc cơ bản của một mạng CNN thông thường gồm 4 lớp: lớp tích chập (convolutional layer), lớp kích hoạt phi tuyến (nonlinear layer), lớp lọc (pooling layer), lớp kết nối đầy đủ (fully connected layer). Các lớp tích chập kết hợp với các lớp phi tuyến sử dụng hàm phi tuyến như ReLU hay hoặc một số hàm khác để tạo ra thông tin trừu tượng hơn cho các lớp tiếp theo. Tùy theo thiết kế và cài đặt cũng như mục đích sử dụng mà mỗi mô hình được người thiết kế thêm hoặc bớt các lớp trên để đạt được mô hình với độ chính xác cao và chi phí tính toán thấp.

Dưới đây là trình bày chi tiết về 4 lớp cơ bản của một mạng CNN.



2.1.3. Kết quả thực hiện được

- Hiểu được kiến trúc của mạng CNN.

2.2. OBJECT DETECTION VÀ THUẬT TOÁN YOLO

Object Detection là một trong những khía cạnh quan trọng trong lĩnh vực thị giác máy tính vì tính ứng dụng của nó trong thực tế. Object detection đề cập đến khả năng của hệ thống máy tính và phần mềm để định vị các đối tượng trong một hình ảnh và xác định từng đối tượng. Object Detection đã được sử dụng rộng rãi để phát hiện khuôn mặt, phát hiện xe, đếm số người đi bộ, hệ thống bảo mật và xe không người lái.

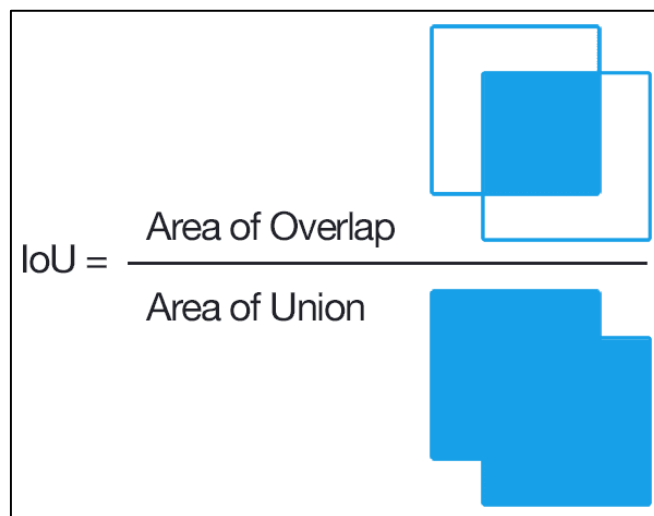
2.2.1. Một vài chỉ số đánh giá model object detection

IOU (Intersection over union):

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. 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).
- Những *predicted bounding box*, đây là bounding box được model sinh ra.

Tỷ lệ giữa đo lường mức độ giao nhau giữa hai đường bao, đường bao dự đoán và đường bao thực gọi là IOU. Tỷ lệ này để nhằm xác định hai khung hình có bị đè chồng lên nhau không và được tính dựa trên phần diện tích giao nhau giữa 2 đường bao với phần tổng diện tích giao nhau và không giao nhau giữa chúng.



Hình 3. Minh họa cách tính IoU

Các tiêu chí IOU được dùng để đánh giá như sau:

- Đối tượng được nhận diện đúng khi $IOU \geq 0.5$
- Đối tượng được nhận diện sai khi $IOU < 0.5$

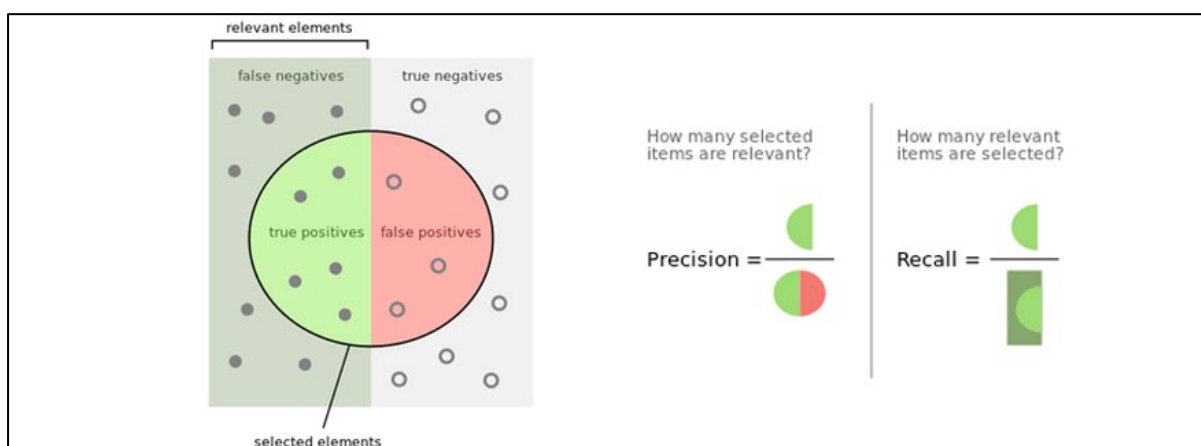
Precision và Recall

Khái niệm về confusion matrix

		Model dự đoán	
		Positive	Negative
Thực tế	Positive	True positive (Dự đoán đúng là Positive)	False negative (Dự sai đúng là Negative)
	Negative	False positive (Dự sai đúng là Positive)	True negative (Dự đoán đúng là negative)

Bảng 1. Confusion matrix

Như bảng trên, ta có thể thấy các trường hợp Positive và Negative giữa thực tế và model dự đoán.



Hình 4. Mô tả cách tính precision và recall

Precision là một chỉ số đo lường mức độ chính xác là dự đoán của mô hình và được tính bằng tỷ lệ giữa số mẫu được tính là True Positive (TP) với tổng số mẫu được phân loại là Positive (bằng chính TP + FP).

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Precision càng cao, tức là số điểm mô hình dự đoán là positive đều là positive càng nhiều. Precision = 1, tức là tất cả số điểm mô hình dự đoán là Positive đều đúng, hay không có điểm nào có nhãn là Negative mà mô hình dự đoán nhầm là Positive.

Recall

Recall là tỷ lệ giữa các điểm positive thực được nhận đúng trên tổng điểm positive thực. Recall được tính bằng công thức:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Recall càng cao có nghĩa tỉ lệ bỏ sót các mẫu positive thực càng thấp.

F1 Scores

Chỉ số F1 càng cao thì model càng tốt. Đây là một trong những tiêu chí giúp chúng ta có thể lựa chọn model một cách tốt nhất. F1 scores được tính như sau:

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

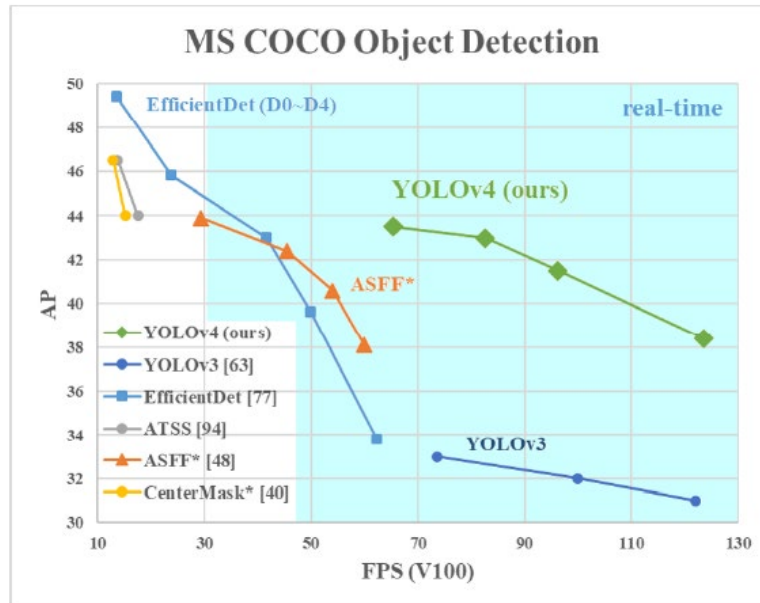
AP và mAP

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). Và mAP là giá trị trung bình của AP trên tất cả các lớp. Chỉ số mAP càng cao thì model càng tốt.

2.2.2. Thuật toán YOLO

YOLO - You Only Look Once là một trong những mô hình phát hiện vật tốt nhất ở thời điểm hiện tại. Dù đều được gọi là YOLO, các phiên bản của mô hình này đều có những cải tiến rất đáng kể sau mỗi phiên bản. Hiện tại, YOLO đã đi qua được 4 phiên bản YOLOv1, YOLOv2, YOLOv3, YOLOv4 và còn một phiên bản YOLOv5 đang được phát triển. Trong phạm vi thực tập, nhóm sẽ chọn phiên bản YOLOv4 để xây dựng model cho bài toán nhận diện vật nuôi.

YOLOv4 có thể coi là một phiên bản cải tiến về tốc độ của YOLOv3. YOLOv4 đã đưa bài toán object detection dễ tiếp cận hơn với những người không có tài nguyên tính toán mạnh. Chúng ta hoàn toàn có thể huấn luyện một mạng phát hiện vật với độ chính xác rất cao bằng YOLOv4 chỉ với GPU 1080ti hoặc 2080ti.



Hình 5. Kết quả so sánh FPS và AP của YOLOv4 với các kiến trúc khác

Hình 5 là kết quả so sánh của YOLOv4 với các thuật toán state-of-the-art tại thời điểm hiện tại. Nhìn vào biểu đồ, ta dễ dàng thấy được sự hiệu quả của YOLOv4 so với các mạng tốt nhất hiện nay. Cụ thể hơn YOLOv4 đạt 43.5% AP trên tập dữ liệu MS COCO ở tốc độ 65 FPS, trên GPU Tesla V100.

Với sự nâng cấp đáng kể về tốc độ cũng như độ chính xác so với YOLOv3, nhóm sẽ chọn YOLOv4 làm thuật toán chính để xây dựng model cho bài toán nhận diện các loài vật nuôi.

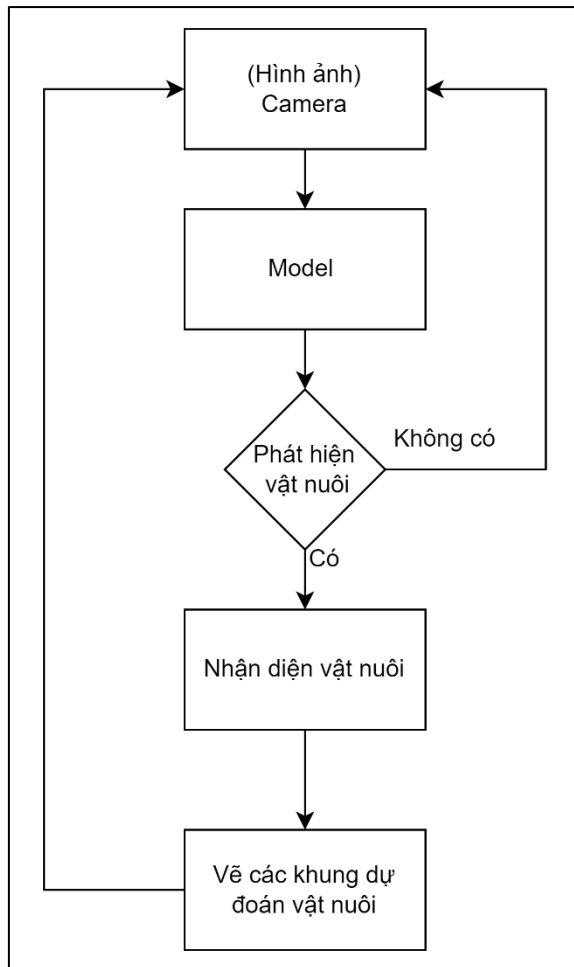
2.2.3. Kết quả thực hiện được

- Hiểu được các thông số đánh giá một model AI để có thể đánh giá model sau này.
- Hiểu được một phần cấu trúc của YOLOv4

2.3. PHÂN TÍCH HỆ THỐNG

2.3.1. Hệ thống nhận diện vật nuôi

Với mong muốn xây dựng một model máy học có thể phát hiện và nhận diện các loài vật nuôi sử dụng thuật toán YOLOv4. Dưới đây là hệ thống phát hiện và nhận diện vật nuôi:



Hình 6. Hệ thống nhận diện vật nuôi

Hình ảnh từ camera hoặc video sẽ được đưa vào model đã được training từ trước để phát hiện vật nuôi. Nếu phát hiện được thì sẽ tiến hành nhận diện đó là vật nuôi nào và tiến hành vẽ các khung dự đoán ra màn hình. Ngược lại, nếu không phát hiện vật nuôi nào thì sẽ quay lại nhận các frame khung hình từ camera hoặc từ ảnh để tiếp tục nhận diện.

Để có thể huấn luyện một model có khả năng nhận biết các loài vật nuôi thì điều đầu tiên cần có đó là cơ sở dữ liệu cho các loài vật nuôi. Chương 4 sẽ trình chi tiết về cơ sở dữ liệu cho bài toán.

2.3.2. Kết quả thực hiện

- Xây dựng được hệ thống nhận diện vật nuôi cho bài toán.

2.4. XÂY DỰNG TẬP DỮ LIỆU

2.4.1. Tổng quan tập dữ liệu – dataset

Với mục tiêu ban đầu xây dựng model có thể nhận biết 5 loại vật nuôi thì nhóm đã chọn ra được 5 loại vật nuôi để chuẩn bị dữ liệu cho quá trình training. Trong đó có 3 loại vật nuôi gần gũi với các gia đình ở nông thôn như chó, mèo, gà. Do data về vật nuôi khá hạn chế nên nhóm chọn 2 loại vật nuôi còn lại là sóc và ngựa.

Dưới đây là số liệu về số lượng ảnh của tập dữ liệu mà nhóm đã chuẩn bị cho quá trình training model.

	Training và validation	Testing
Mèo	585	200
Gà	561	207
Chó	642	200
Ngựa	625	201
Sóc	602	200
Tổng cộng	3015	1008

Bảng 2. Thông số về dataset

2.4.2. Những loại dữ liệu trong dataset

Dữ liệu trong dataset bao gồm các hình ảnh ở nhiều góc độ khác độ. Dưới đây là một số góc độ và tư thế của vật nuôi có trong dataset.

Hình ảnh chụp ngang vật nuôi

Các hình ảnh của vật nuôi nhìn thấy một bên của vật nuôi. Vật nuôi có thể đang đứng, đang ngồi hoặc đang nằm với dạng ảnh chụp ngang.



Hình 7. Hình ảnh trong dataset chụp ngang vật nuôi

Hình ảnh chụp chính diện vật nuôi

Vật nuôi được chụp chính diện, góc độ có thể từ dưới lên hoặc từ trên xuống. Ảnh có thể bị cắt một chút nhưng vẫn đảm bảo có thể nhìn rõ vật nuôi.



Hình 8. Hình ảnh trong dataset chụp chính diện vật nuôi

Hình ảnh chụp các góc nghiêng của vật nuôi

Ảnh chụp vật nuôi với đa dạng các góc nghiêng, đảm bảo vật nuôi không bị thiếu các đặc trưng ở các góc nghiêng.



Hình 9. Hình ảnh trong dataset chụp các góc nghiêng của vật nuôi

Hình ảnh chụp vật nuôi bị che một phần cơ thể

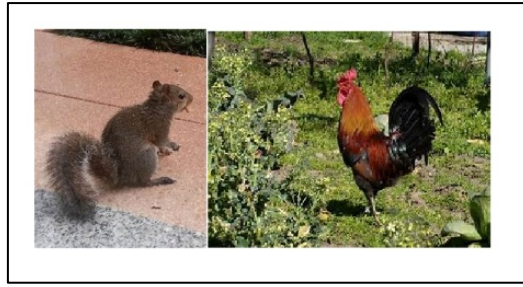
Ảnh chụp vật nuôi bị cắt một phần nhỏ hoặc lớn, một số ảnh bị che mất bộ phận như thân, chân. Một số ảnh có thể chỉ bao gồm phần cổ hoặc mặt của vật nuôi.



Hình 10. Hình ảnh trong dataset chụp vật nuôi bị che một phần cơ thể

Một số dạng khác

Ảnh chụp từ phía sau lên của vật nuôi, một vài ảnh có thể bị cắt một số phần vật thể, số lượng ảnh chụp phía sau tương đối ít so với mặt bằng chung của tập dữ liệu.



Hình 11. Hình ảnh trong dataset chụp phía sau vật nuôi

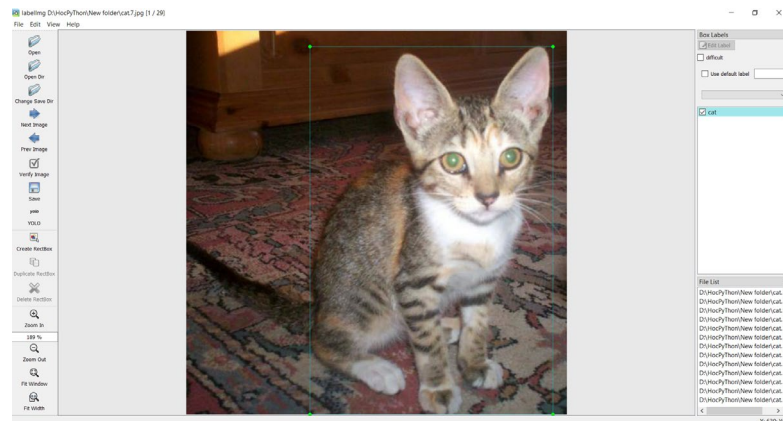
Ngoài ra, một số ảnh có thể có chứa nhiều object, điều này giúp tiết kiệm số lượng ảnh đầu vào nhưng vẫn đảm bảo đầy đủ về mặt dữ liệu của các object.



Hình 12. Hình ảnh trong dataset có thể có nhiều object

2.4.3. Gán nhãn dữ liệu

Sau khi đã thu thập dữ liệu thì nhóm tiến hành gán nhãn dữ liệu cho data. Công cụ dùng để gán nhãn là LabelImg.



Hình 13. Minh họa gán nhãn dữ liệu bằng LabelIMG

Trong quá trình gán nhãn, đối với một số ảnh vật nuôi có râu như mèo và sóc thì quá trình gán nhãn sẽ bỏ qua các đặc điểm đó vì nó chiếm tỷ lệ nhỏ và ảnh hưởng không tốt đến độ chính xác của model. Tập data này sẽ là tập training và validation với 80% training và 20% validation (dữ liệu được chia ngẫu nhiên để đảm bảo tính phân phối dữ liệu).

2.4.4. Kết quả thực hiện được

- Xây dựng được tập dữ liệu cho 5 loại vật nuôi.
- Hoàn tất gắn nhãn cho tập dữ liệu để chuẩn bị huấn luyện.

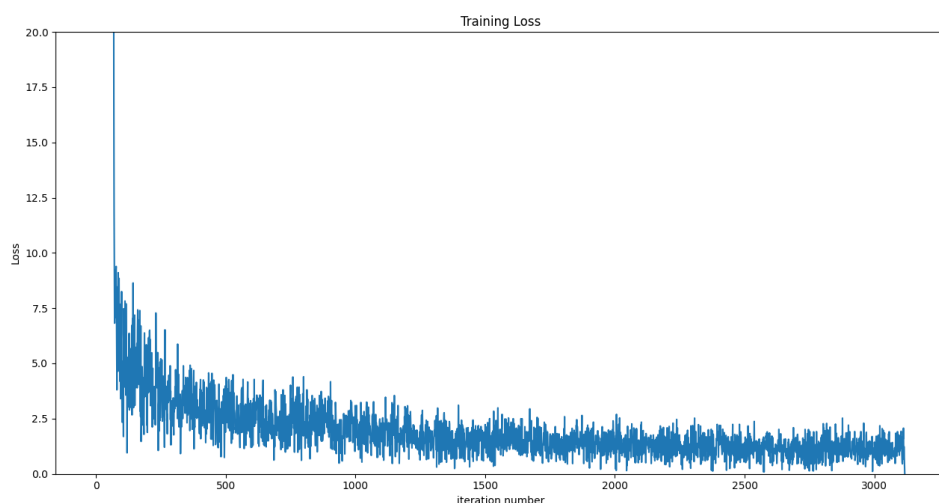
2.5. HUẤN LUYỆN YOLOV4

2.5.1. Quá trình huấn luyện

Sau khi chuẩn bị đầy đủ các file cần thiết để phục vụ quá trình huấn luyện thì quá trình huấn luyện bắt đầu. Quá trình huấn luyện được tiến hành trên Google Colab với GPU hỗ trợ để tăng tốc quá trình huấn luyện.

Để có thể kiểm soát quá trình huấn luyện thì ta theo dõi các chỉ số hàm mất mát (loss function) để đánh giá xem độ mất mát dữ liệu. Cùng với loss, chúng ta có thể theo dõi thêm 1 chỉ số để đánh giá xem có thể dừng training được hay chưa, đó là chỉ số mAP. Nhằm tiết kiệm thời gian training thì sau một số epochs nhất định, mAP sẽ được tính một lần và tùy thuộc vào kích thước của tập train. YOLO sẽ tính như sau: số epochs = $\max(x, y)$ với $x=100$ (x có thể được thiết lập), $y = 4 \times \text{số lượng ảnh train/batch size}$ (batch size được lấy trong file config được thiết lập trước). Cụ thể trong dataset hiện tại với 2474 ảnh dùng để training thì sau 154 vòng sẽ validate mAP một lần.

Nếu sau 100 epochs mà loss có giá trị bão hòa, đồng thời mAP khi được tính sau 154 vòng đạt mức mong muốn thì có thể dừng training.



Hình 14. Training loss trong quá trình training

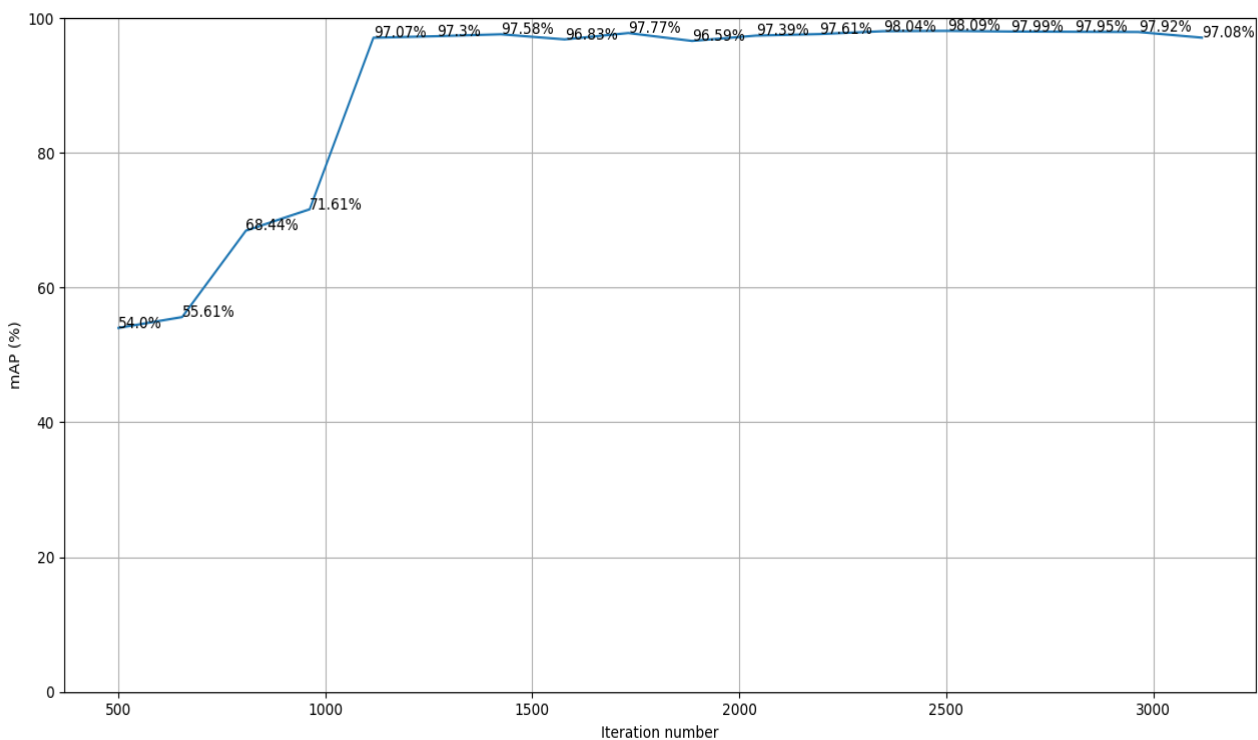
Kết quả thực hiện được:

- Training loss có dấu hiệu bắt đầu bão hòa sau khi train được 3118 vòng.
- Thu được biểu đồ loss training để đánh giá khi nào nên dừng train.

2.5.2. Đánh giá model trên tập validation

Để đánh giá và kiểm tra model, ta sẽ dựa vào các thông số như precision, recall, F1 scores, AP và mAP.

Đầu tiên là kết quả mAP (Mean average precision) khi train 5 class. Sau 154 vòng, YOLOv4 sẽ tính mAP một lần và mAP có kết quả cao nhất sẽ là model có độ chính xác cao nhất.



Hình 15. Giá trị của mAP của model trên tập validate

Quan sát biểu đồ mAP và kết quả của quá trình training thì tại vòng thứ 2502 cho chỉ số mAP có độ chính xác cao nhất là 98.09%.

AP của model trên từng class như sau:

Lớp	Average Precision (AP)
Mèo	98.93%
Gà	93.84%
Chó	99.2%
Ngựa	98.14%
Sóc	99.33%
mAP	98.09%

Bảng 3. Average Precision cho 5 lớp

Mặt khác, nhóm đã cho validate lại cụ thể hơn bằng việc lấy ra các dữ liệu cụ thể hơn để phân tích chi tiết các thông số. Qua đó có thể cải thiện độ chính xác của mô hình dựa vào những dữ liệu thu được.

	Tổng số ảnh	Tổng số object phát hiện được	Số ảnh không nhận diện được (IOU<0.5)	Số ảnh nhận diện nhầm	Chi tiết ảnh nhận nhầm	Số object có IOU>0.9
Mèo	119	130	3	1	Sóc (60.41%)	97
Gà	104	165	0	2	Sóc (87.57%,55.30%)	131
Chó	113	131	2	0	X	115
Ngựa	100	110	1	0	X	103
Sóc	105	108	1	2	Chó (90.95%,61.60%)	81

Bảng 4. Chi tiết các thông số khi validate

Ngoài chỉ số mAP, ta còn thu được các chỉ số sau:

- Precision = 0.94
- Recall = 0.97
- F1 Score = 0.96

Trong trường hợp này, giá trị của F1 score tương đối cao, Precision và Recall cũng khá cân bằng. Điều này cho thấy, model đã khá ổn định trên tập dữ liệu hiện tại.

2.5.3. Đánh giá model trên tập test

Để có thể đánh giá model một cách khách quan hơn, nhóm đã chuẩn bị thêm một tập dữ liệu để test model. Dữ liệu test gồm khoảng 1000 ảnh chia đều cho 5 class.

Kết quả thu được:

Lớp	Độ chính xác
Mèo	95.57%
Gà	93.26%
Chó	93.26%
Ngựa	97.93%
Sóc	90.93%
Trung bình	93.77%

Bảng 5. Độ chính xác trung bình trên tập Test

Thông số của model sau khi testing có sự sụt giảm về độ chính xác. Đó là do tập test được xây dựng từ các ảnh ngẫu nhiên, một số ảnh trong tập test không có trong tập train. Mặt khác, lượng data cho mỗi class cũng còn khá ít nên khó tránh khỏi việc thiếu dữ liệu, từ đó làm giảm chất lượng model.

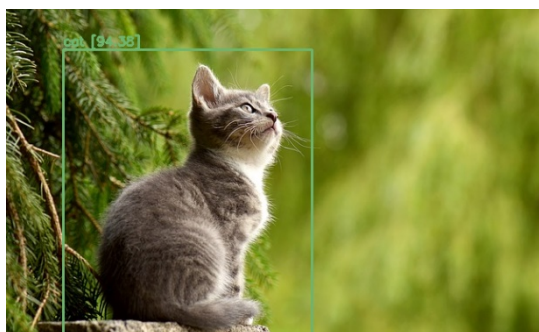
Dưới đây là một số thông số được ghi lại từ quá trình test. Một số class có tình trạng nhận nhầm, cũng như có một số ảnh mà detector không thể dò tìm ra vật nuôi.

	Ảnh đã test	Số object phát hiện được	Số object nhận nhầm
Mèo	200	195	6
Gà	207	228	0
Chó	200	184	14

Ngựa	201	204	1
Sóc	200	186	2

Bảng 6. Chi tiết các thông số khi testing

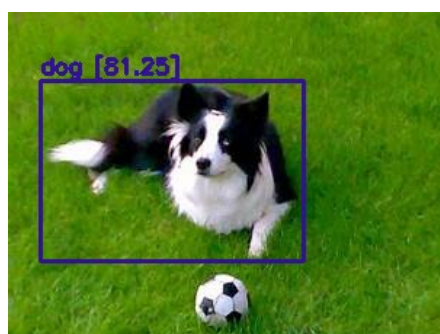
Nhìn chung, model đạt độ chính xác 93.37 % trên tập test, một mức khá ổn. Dưới đây là một vài hình ảnh minh họa cho việc sử dụng model object detection YOLOv4 để nhận diện và phát hiện vật nuôi trên ảnh đầu vào.



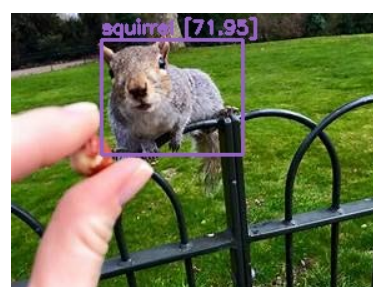
Hình 16. Minh họa nhận diện mèo



Hình 19. Minh họa nhận diện ngựa



Hình 17. Minh họa nhận diện chó



Hình 20. Minh họa nhận diện sóc



Hình 18. Minh họa nhận diện gà

CHƯƠNG 3. KẾT QUẢ ĐẠT ĐƯỢC VÀ CÁC KHÓ KHĂN

3.1. ƯU ĐIỂM

- Làm việc và đạt kết quả đúng như trong tiến độ đã đề ra.
- Kỹ năng làm việc nhóm hiệu quả.
- Kỹ năng đọc tài liệu tiếng anh tương đối tốt.

3.2. HẠN CHẾ

- Chưa tìm hiểu sâu về lý thuyết của thuật toán YOLOv4, nhiều cấu trúc mạng chưa tìm hiểu được.
- Không ứng dụng được lên các thiết bị nhúng mà chỉ testing được trên windows 10 hoặc Google Colab.

3.3. KẾT QUẢ ĐẠT ĐƯỢC

- Hiểu được mạng kiến trúc của mạng nơ ron tích chập và các thành phần của nó.
- Hiểu được một số thông số đánh giá model object detection.
- Tìm hiểu và nắm bắt được các phiên bản của YOLO và cách YOLO hoạt động
- Xây dựng được cơ sở dữ liệu cho bài toán.
- Huấn luyện thành công model có thể phát hiện và nhận diện vật nuôi với độ chính xác trên tập test lớn hơn 90%.
- Ngoài ra, chúng em còn được học và trau dồi thêm các kỹ năng như kỹ năng làm việc nhóm, kỹ năng viết báo cáo, làm slide.

3.4. KHÓ KHĂN

- Đôi khi gặp sự cố về phần cứng (máy tính có vấn đề).
- Do làm việc tại nhà nên kết nối mạng đôi khi không ổn định.
- Data của bài toán còn hạn chế nên model chưa được ổn định.

GÓP Ý VỀ CHƯƠNG TRÌNH ĐÀO TẠO CỦA KHOA

Về chương trình đào tạo của khoa thì chúng em nhận thấy, chương trình đào tạo của khoa đang có một lộ trình rất hợp lí nên nhóm chúng em sẽ không có góp ý nhiều về chương trình đào tạo của khoa. Do đó, chúng em sẽ chỉ nêu ra một số góp ý nhỏ theo quan điểm cá nhân để khoa có thể xem xét và cập nhật trong chương trình đào tạo nếu nó khả quan. Cụ thể, nhóm chúng em có hai góp ý nhỏ sau đây:

Thứ nhất là những môn hoặc những nội dung về hệ thống máy tính. Trong những môn học thay thế khóa luận có môn “Kỹ Thuật hệ thống máy Tính”. Đây là một môn rất hay khi chúng em được học về hệ thống, những thành phần của hệ thống, cách xây dựng cũng như cách đánh giá, phân tích một hệ thống máy tính. Do đây là môn thay thế khóa luận tốt nghiệp nên sinh viên sẽ được học rất trễ (trừ những bạn làm khóa luận). Mặt khác, trong chương trình học có nhiều môn ở những năm 2, năm 3 đã bắt đầu làm về những hệ thống với những quy mô khác nhau. Chính vì thế, nếu có thể được học môn này sớm hoặc các thầy cô có thể hướng dẫn các bạn sinh viên các hệ thống sớm hơn thì các bạn sinh viên có thể tiếp cận vấn đề này một cách sớm nhất và giúp các bạn thuận tiện hơn trong việc học tập và nghiên cứu.

Và thứ hai đó là về các nội dung về trí tuệ nhân tạo, xây dựng các model máy học và triển khai xuống các thiết bị nhúng để có thể xử lí các tác vụ ở thời gian thực. Mặc dù các môn về trí tuệ nhân tạo không phải là những môn chính thức của khoa Kỹ Thuật Máy Tính, nhưng nếu có thể đưa vào chương trình dưới dạng môn tự chọn và đưa vào trong sổ tay. Việc được học và nắm bắt được các khái niệm cơ bản sẽ giúp sinh viên dễ dàng có hướng đi hơn bởi vì đây là một lĩnh vực rất rộng và có rất nhiều kiến thức.

TÀI LIỆU THAM KHẢO

- [1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi ,” You Only Look Once: Unified, Real-Time Object Detection”, University of Washington, Allen Institute for AI, Facebook AI Research.
- [2] Joseph Redmon, Ali Farhadi , “YOLO9000: Better, Faster, Stronger”, University of Washington, Allen Institute for AI.
- [3] Joseph Redmon, Ali Farhadi, “ YOLOv3: An Incremental Improvement”, University of Washington.
- [4] Alexey Bochkovskiy, Chien-Yao Wang , Hong-Yuan Mark Liao , “YOLOv4: Optimal Speed and Accuracy of Object Detection”, Institute of Information Science Academia Sinica, Taiwan.
- [5] Cường Hoàng, 21/01/2021, Series YOLO: #4 Tìm hiểu cấu trúc YOLOv1,v2,v3 và v4-Phần1, <<https://devai.info/2021/01/21/series-yolo-4-tim-hieu-cau-truc-yolov1v2v3-va-v4/>>, truy cập ngày 26/12/2021.
- [6] Cường Hoàng, 24/02/2021, Series YOLO: #4 Tìm hiểu cấu trúc YOLOv1,v2,v3 và v4-Phần1, <<https://devai.info/2021/02/24/series-yolo-4-tim-hieu-cau-truc-yolov1v2v3-va-v4-phan-2/>>, truy cập ngày 26/12/2021.