**TRƯỜNG ĐẠI HỌC AN GIANG**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO**

**TRÍ TUỆ NHÂN TẠO**

**ĐỀ TÀI: TÌM HIỂU PHƯƠNG PHÁP ĐẾM VẬT NUÔI**

**NHÓM 7**

**PHAN ANH KIỆT DTH205836**

**NGUYỄN TRẦN NGỌC NGÂN DTH205876**

**NGUYỄN THỊ XUÂN HUỲNH DTH205802**

**LÊ HOÀI BẢO DTH205717**

**PHẠM LÝ TƯỜNG DTH206052**

**AN GIANG, 12-2023**

**TRƯỜNG ĐẠI HỌC AN GIANG**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO**

**TRÍ TUỆ NHÂN TẠO**

**ĐỀ TÀI: TÌM HIỂU PHƯƠNG PHÁP ĐẾM VẬT NUÔI**

**NHÓM 7**

**PHAN ANH KIỆT DTH205836**

**NGUYỄN TRẦN NGỌC NGÂN DTH205876**

**NGUYỄN THỊ XUÂN HUỲNH DTH205802**

**LÊ HOÀI BẢO DTH205717**

**PHẠM LÝ TƯỜNG DTH206052**

**GIẢNG VIÊN: NGUYỄN VĂN HÒA**

**AN GIANG, 12-2023**

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

**Giảng viên hướng dẫn**

*(ký và ghi rõ họ tên)*

# MỞ ĐẦU

Hiện nay, ngành công nghiệp chăn nuôi trong nước ta đang ngày càng phát triển. Chăn nuôi truyền thống đang dần được thay thế bằng chăn nuôi công nghệ cao giúp giảm chi phí nguồn nhân lực, giúp nông dân tối đa được nguồn chi phí và được nhiều lợi nhuận hơn.

Khi cuộc cách mạng công nghiệp 4.0 mở ra, nhiều công nghệ mới tiếp tục được áp dụng trong các trang trại chăn nuôi như: Internet vạn vật (IoT), trí tuệ nhân tạo (AI), big data… Thông qua áp dụng công nghệ hiện đại đã biến quá trình chăn nuôi truyền thống (dựa vào tay chân là chính) thành chăn nuôi tự động hóa (chủ yếu dựa vào phần mềm, máy máy móc thiết bị).

Có quá nhiều trang trại và số lượng vật nuôi không ngừng tăng lên . Nông dân không thể dể dàng kiểm soát số lượng lớn vật nuôi bằng thủ công được , Để giải quyết vấn đề đó công nghệ nhận dạng và đếm số lượng vật nuôi bằng AI ra đời .Mục tiêu của báo cáo là giúp chúng ta nhận dạng vật nuôi và đếm số lượng của chúng.

**MỤC LỤC**

[MỞ ĐẦU iii](#_Toc2343)

[MỤC LỤC HÌNH ẢNH 5](#_Toc17516)

[PHẦN MỞ ĐẦU 1](#_Toc7420)

[1.Đặt vấn đề 1](#_Toc13021)

[2.Mục đích đề tài 1](#_Toc5754)

[3. Phương pháp nghiên cứu 1](#_Toc12309)

[4.Phạm vi nghiên cứu 1](#_Toc14685)

[5.Nội dung nghiên cứu 1](#_Toc28906)

[6. Cấu trúc báo cáo 2](#_Toc3849)

[NỘI DUNG 3](#_Toc27016)

[CHƯƠNG 1: CƠ SỞ LÝ THUYẾT 3](#_Toc29161)

[1.1. Trí tuệ nhân tạo 3](#_Toc11738)

[a. Khái niệm 3](#_Toc25793)

[b. Hướng nghiên cứu về trí tuệ nhân tạo 3](#_Toc521)

[c. Phân loại 4](#_Toc7587)

[d. Ứng dụng 4](#_Toc1075)

[1.2. Tổng quan OpenCV 4](#_Toc10588)

[1.3. Object detection và object tracking 4](#_Toc24705)

[1.4. Tổng quan về YOLO 5](#_Toc29601)

[a. Tổng quan 5](#_Toc19269)

[b. Dòng thời gian phát triển của YOLO 6](#_Toc8510)

[c. YOLOv8 sử dụng cho đếm vật nuôi 6](#_Toc6353)

[1.5. Tổng quan về DeepSort 10](#_Toc18713)

[CHƯƠNG 2: QUÁ TRÌNH NHẬN DẠNG VÀ ĐẾM VẬT NUÔI 11](#_Toc32157)

[1. Chuẩn bị dữ liệu 11](#_Toc19814)

[2. Gán nhãn cho đối tượng trong ảnh 11](#_Toc30842)

[3. Huấn luyện 11](#_Toc16549)

[4. Nhận dạng đối tượng 11](#_Toc485)

[5. Theo dõi đối tượng 12](#_Toc20238)

[6. Đếm vật nuôi 12](#_Toc23108)

[Công trình được trích dẫn 13](#_Toc12353)

**MỤC LỤC HÌNH ẢNH**

[Hình 1: Hình ảnh của AI 3](#_Toc25709)

[Hình 2: Nhận dạng vật nuôi 5](#_Toc1508)

[Hình 3: YOLO 5](#_Toc6379)

[Hình 4: Progression Timeline 6](#_Toc3967)

[Hình 5: Cấu trúc mạng CNN 7](#_Toc6903)

[Hình 6: So sánh độ chính xác và hiệu suất của các mô hình YOLO 8](#_Toc28775)

[Hình 7: Huấn luyện YOLOv8 9](#_Toc30781)

[Hình 8: Nhiều chiếc xe hơi được gắn nhãn (Nguồn: Open Image Dataset v7) 9](#_Toc14645)

[Hình 9: Nhận dạng đối tượng trong hình 12](#_Toc23057)

# PHẦN MỞ ĐẦU

## 1.Đặt vấn đề

Chăn nuôi là lĩnh vực trong nông nghiệp, về việc nuôi dưỡng quản lý và chăm sóc động vật là quá trình nuôi lớn vật nuôi để sản xuất ra các sản phẩm như: thực phẩm, lông và sức lao động… và vấn đề quản lý vật nuôi luôn là một vấn đề lớn đối với ngành nhất là đối với những trang trại có quy mô lớn và số lượng đàn nhiều, khó có thể kiểm soát được số lượng, và làm thế nào để ta có thể kiểm soát được một số lượng vật nuôi với quy mô lớn.

Hiện nay trên thế giới có rất nhiều trang trại hay doanh nghiệp có trang thiết bị, công nghệ hay máy móc tiên tiến để có thể sắp xếp cũng như quản lý vật nuôi với số lượng đàn lớn, điều này giúp dể dàng quản lý vật nuôi một cách chặt chẻ và khoa học hơn. Từ đó giúp nâng cao chất lượng, hiệu quả trong quản lý, giám sát số lượng đàn một cách chính xác, giảm chi phí, nâng cao chất lượng.

Những năm gần đây, trí tuệ nhân tạo AI ngày càng phát triển, cụ thể hơn thì AI là Machine Learning và Deep Learning đang từ ngày thay thế nhiều hoạt động bằng chứng là cuộc cách mạng công nghiệp 4.0. và vận dụng từ những điều này cùng với việc học môn Trí tuệ nhân tạo cho đề tài tìm hiểu về nhận diện và đếm vật nuôi.

## 2.Mục đích đề tài

- Nghiên cứu và tìm hiểu phương pháp nhân diện áp dụng với việc đếm số lượng đàn vật nuôi.

- Tìm hiểu về chương trình phương pháp giúp nhân diện, đếm số lượng vật nuôi.

## Phương pháp nghiên cứu

- Tìm hiểu các phương pháp, thuật toán, mô hình, có thể sử dụng để đếm động vật qua chuyển động của chúng.

- Tìm hiểu cách mà các phương pháp, mô hình hoạt động, độ chính xác và tốc độ của các mô hình.

## 4.Phạm vi nghiên cứu

- Dữ liệu nghiên cứu được lấy từ các địa điểm chăn nuôi các loại gia xúc, gia cầm như heo, gà, vịt quanh địa phương, yêu cầu có lắp camera.

## 5.Nội dung nghiên cứu

- Tìm hiểu tổng quan về Object Detection and Object Tracking.

- Tìm hiểu về mô hình nhận dạng vật nuôi YOLOv8 và OpenCV.

- Tìm hiểu về phương pháp theo dõi vật nuôi với DeepSORT.

## Cấu trúc báo cáo

Báo cáo “ Tìm hiểu phương pháp đếm vật nuôi ” bao gồm tất cả 2 phần:

**Chương 1:** Cơ sở lý thuyết về *“Tìm hiểu phương pháp đếm vật nuôi”.*

**Chương 2:** Quá trình nhận dạng và đếm vật nuôi.

# NỘI DUNG

## CHƯƠNG 1: CƠ SỞ LÝ THUYẾT

#### Trí tuệ nhân tạo

##### Khái niệm

Trí tuệ nhân tạo (AI) đại diện cho một lĩnh vực nghiên cứu trong khoa học máy tính, tập trung vào giải quyết những thách thức liên quan đến khả năng nhận thức, tương tác và học tập, đặc biệt là những khía cạnh gắn liền với khả năng tư duy của con người.



Hình 1: Hình ảnh của AI

##### Hướng nghiên cứu về trí tuệ nhân tạo

**Các nhánh nghiên cứu của trí tuệ nhân tạo:**

- Khả năng học (nhánh học máy, học sâu)

Học máy, học sâu (Machine Learning, Deep Learning): nghiên cứu và phát triển các kỹ thuật giúp cho máy tính có thể học tri thức từ dữ liệu đầu vào.

- Khả năng biểu diễn tri thức và suy diễn (nhánh phương pháp biểu diễn tri thức và suy diễn)

Biểu diễn tri thức (Knowledge representation): Cung cấp cơ sở để máy tính có thể thực hiện việc suy diễn như con người.

- Khả năng nghe-nhìn (nhánh thị giác máy tính, xử lý giọng nói)

Thị giác máy tính (Computer vision): Nghiên cứu và phát triển các kỹ thuật,khả năng của máy tính để nhận diện, hiểu và xử lý hình ảnh và video

để giúp máy tính có thể nghe và nhìn như con người.

- Khả năng sử dụng ngôn ngữ (nhánh xử lý ngôn ngữ tự nhiên)

Xử lý ngôn ngữ tự nhiên (Natural language processing): Giúp cho máy tính có thể hiểu được ngôn ngữ mà con người đang sử dụng.

- Khả năng thể hiện cử chỉ (nhánh Robotics)

Robotics: Hỗ trợ robot thể hiện các hành động và cử chỉ như con người.

##### Phân loại

* Artificial Narrow Intelligence (ANI - Trí tuệ nhân tạo hẹp)
* Artificial General Intelligence (AGI - Trí tuệ nhân tạo chung)
* Artificial Super Intelligence (ASI - Siêu trí tuệ nhân tạo).

##### Ứng dụng

- Xử lý ngôn ngữ tự nhiên (NLP):

+ Chatbots và Virtual Assistants: AI được sử dụng để xây dựng các

trợ lý ảo và chatbot để tương tác với người dùng và cung cấp

thông tin hoặc giải quyết vấn đề.

 + Phân tích tình cảm: AI có thể phân tích tình cảm trong văn bản,

giúp doanh nghiệp hiểu đánh giá của khách hàng về sản phẩm

hoặc dịch vụ.

- Thị giác máy tính (Computer Vision):

 + Nhận diện khuôn mặt: AI có thể nhận diện và xác định khuôn

mặt trong hình ảnh hoặc video, thường được sử dụng trong hệ

thống đảm bảo an ninh và xác thực.

 + Xử lý hình ảnh y tế: Giúp bác sĩ chẩn đoán bệnh dựa trên hình

ảnh y tế như X-quang, MRI, CT scans.

- Tự động hóa và Robotica:

 + Ô tô tự lái: AI được sử dụng trong xe ô tô tự lái để phát hiện và

tránh các vật thể, biển báo giao thông, và điều khiển xe.

 + Robot công nghiệp: Robot được trang bị trí tuệ nhân tạo để thực

hiện các nhiệm vụ trong môi trường sản xuất, từ lắp ráp đến kiểm

tra chất lượng.

##### Tổng quan OpenCV

OpenCV viết tắt cho Open Source Computer Vision Library. OpenCV là thư viện nguồn mở hàng đầu cho Computer Vision và Machine Learning, và hiện có thêm tính năng tăng tốc GPU cho các hoạt động theo real-time.OpenCV được thiết kế để hỗ trợ hiệu quả về tính toán và chuyên dùng cho các ứng dụng real-time (thời gian thực). ảnh. OpenCV cung cấp một loạt các tính năng và công cụ tiện ích để phân tích, xử lý và trích xuất thông tin từ ảnh và video.

1. Xử lý ảnh và video

2. Nhận dạng khuôn mặt và phát hiện vật thể

3. Xử lý thời gian thực và thị giác máy tính

4. Tích hợp với ngôn ngữ lập trình

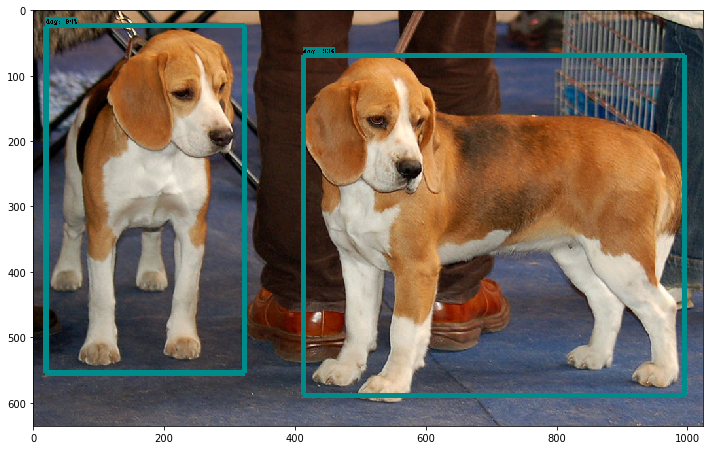
5. Hỗ trợ đa nền tảng

##### Object detection và object tracking

Object detection và tracking là hai khái niệm quan trọng trong lĩnh vực thị giác máy tính. Object detection là bài toán xác định vị trí và đối tượng trong một hình ảnh hoặc video. Trong khi đó, object tracking là bài toán theo dõi một hoặc nhiều đối tượng chuyển động theo thời gian trong một video.

Một số thuật toán phổ biến trong object detection bao gồm: YOLO (You Only Look Once), Faster R-CNN (Region-based Convolutional Neural Network), và SSD (Single Shot Detector) .

Các thuật toán phổ biến trong object tracking bao gồm SORT (Simple Online and Realtime Tracking) và Deep SORT (Deep Learning-based Object Tracking).

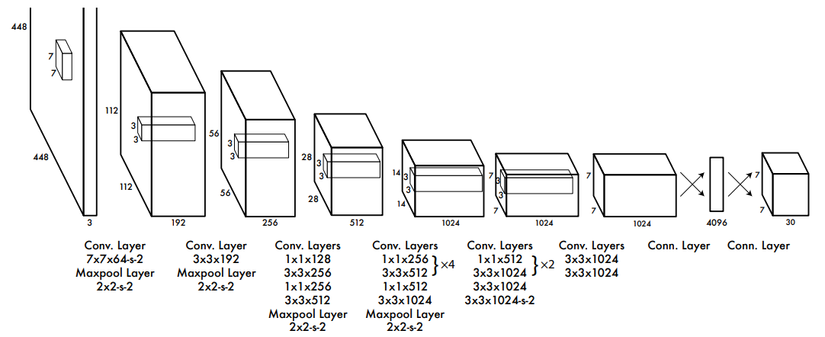


Hình 2: Nhận dạng vật nuôi

##### Tổng quan về YOLO

##### Tổng quan

Họ các mô hình về YOLO (You Only Look Once) dùng để nhận dạng đối tượng được thiết kế để nhận diện các vật thể real-time.Yolo là một mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng. Yolo được tạo ra từ việc kết hợp giữa các convolutional layers và connected layers.Trong đóp các convolutional layers sẽ trích xuất ra các feature của ảnh, còn full-connected layers sẽ dự đoán ra xác suất đó và tọa độ của đối tượng.



Hình 3: YOLO

##### Dòng thời gian phát triển của YOLO



Hình 4: Progression Timeline

##### YOLOv8 sử dụng cho đếm vật nuôi

YOLOv8 là một mô hình nhận dạng đối tượng dựa trên mạng convolutional neural network (CNN) được phát triển bởi Joseph Redmon và nhóm nghiên cứu của ông tại Đại học Washington.

Mạng CNN là một tập hợp các lớp Convolution chồng lên nhau và sử dụng các hàm nonlinear activation như ReLU và tanh để kích hoạt các trọng số trong các node. Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.

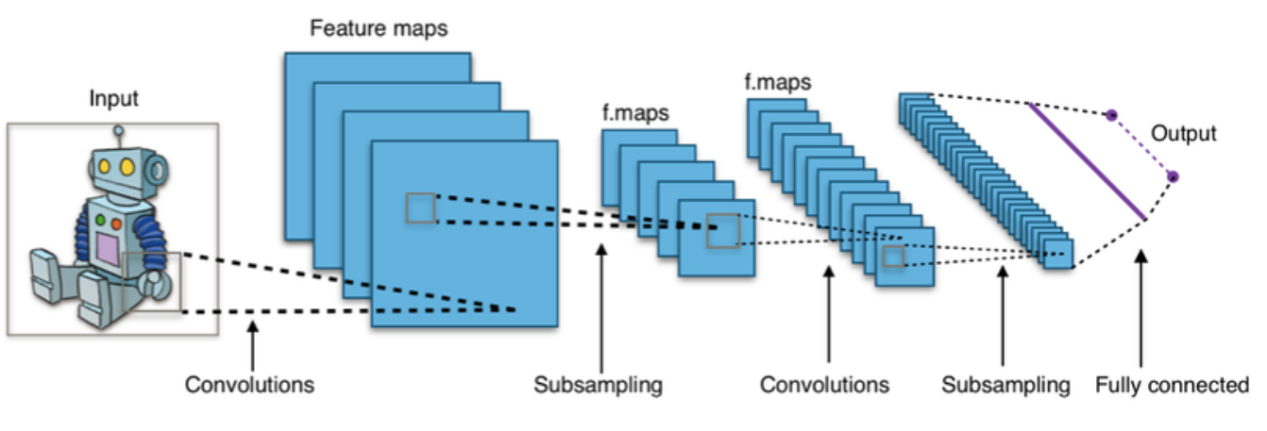
Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo. Trong mô hình mạng truyền ngược (feedforward neural network) thì mỗi neural đầu vào (input node) cho mỗi neural đầu ra trong các lớp tiếp theo.

Mô hình này gọi là mạng kết nối đầy đủ (fully connected layer) hay mạng toàn vẹn (affine layer). Còn trong mô hình CNNs thì ngược lại. Các layer liên kết được với nhau thông qua cơ chế convolution.

Layer tiếp theo là kết quả convolution từ layer trước đó, nhờ vậy mà ta có được các kết nối cục bộ. Như vậy mỗi neuron ở lớp kế tiếp sinh ra từ kết quả của filter áp đặt lên một vùng ảnh cục bộ của neuron trước đó.

Mỗi một lớp được sử dụng các filter khác nhau thông thường có hàng trăm hàng nghìn filter như vậy và kết hợp kết quả của chúng lại. Ngoài ra có một số layer khác như pooling/subsampling layer dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu).

Trong quá trình huấn luyện mạng (traning) CNN tự động học các giá trị qua các lớp filter dựa vào cách thức mà bạn thực hiện. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Layer cuối cùng được dùng để phân lớp ảnh.



Hình 5: Cấu trúc mạng CNN

Trong mô hình CNN có 2 khía cạnh cần quan tâm là **tính bất biến** (Location Invariance) và **tính kết hợp** (Compositionality). Với cùng một đối tượng, nếu đối tượng này được chiếu theo các gốc độ khác nhau (translation, rotation, scaling) thì độ chính xác của thuật toán sẽ bị ảnh hưởng đáng kể.

Pooling layer sẽ cho bạn tính bất biến đối với phép dịch chuyển (translation), phép quay (rotation) và phép co giãn (scaling). Tính kết hợp cục bộ cho ta các cấp độ biểu diễn thông tin từ mức độ thấp đến mức độ cao và trừu tượng hơn thông qua convolution từ các filter.

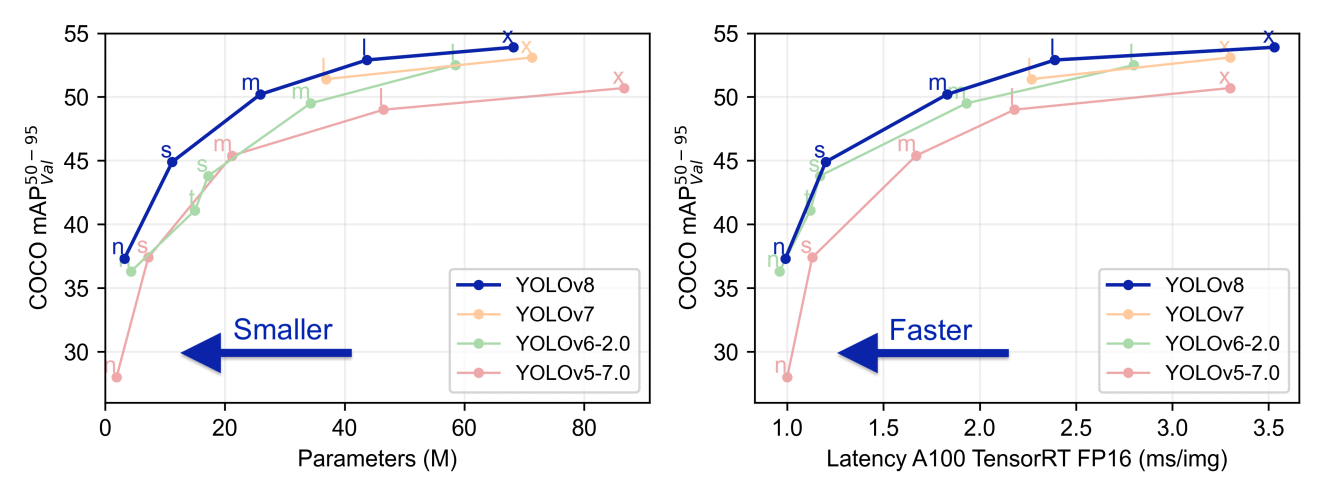
Đó là lý do tại sao CNNs cho ra mô hình với độ chính xác rất cao. Cũng giống như cách con người nhận biết các vật thể trong tự nhiên.

###### Tính năng chính:

* Tăng cường mô hình bằng cách thêm các kênh phân tán để tăng tốc độ tính toán.
* Sử dụng kỹ thuật Attention để cải thiện khả năng nhận dạng đối tượng của mô hình.
* Áp dụng phương pháp đào tạo mới để tăng tốc độ hội tụ.
* Sử dụng kiến trúc mạng nơ-ron mới: Sử dụng kiến trúc YOLOv4 làm cơ sở để tăng hiệu suất và độ chính xác của mô hình.
* Tích hợp cơ chế tự động điều chỉnh tỷ lệ tăng kích thước của hình ảnh đầu vào (AutoScale).
* Hỗ trợ giám sát bằng video (Video Supervision): Mô hình có khả năng phát hiện và giám sát vật thể trong các video và đưa ra dự đoán liên tục trên toàn bộ video.
* Tích hợp công nghệ Ensemble.
* Tính năng điều chỉnh tỷ lệ tự động (AutoAnchor): Cải thiện việc phát hiện đối tượng với nhiều tỷ lệ khác nhau.

###### Ưu điểm

- Tốc độ: YOLOv8 được đánh giá là nhanh chóng và thời gian phản hồi thấp, giúp xử lý các tác vụ nhận diện đối tượng và phân-segment ảnh trong thời gian thực.  
 - Độ chính xác: YOLOv8 được xây dựng trên các tiến bộ về học sâu và thị giác máy tính, đảm bảo độ chính xác cao trong việc nhận diện đối tượng.  
 - Sự linh hoạt: YOLOv8 hỗ trợ việc nhận diện đối tượng và phân-segment trên cả GPU và CPU, tận dụng các công nghệ như TensorRT của Nvidia và OpenVino của Intel.



Hình 6: So sánh độ chính xác và hiệu suất của các mô hình YOLO

###### Hạn chế

- Để sử dụng YOLOv8 hiệu quả cần phải:

+ Có kiến thức chuyên sâu về Machine Learning, Deep Learning và các thuật toán liên quan.

+ Cần phải được huấn luyện trên một tập dữ liệu đủ lớn và đa dạng để đạt được hiệu quả cao nhất.

+ Yêu cầu các tài nguyên tính toán cao để đạt được tốc độ xử lý nhanh và chính xác.

- Thuật toán YOLOv8 không phải là mã nguồn mở và chỉ có sẵn thông qua các thỏa thuận cấp phép với người tạo ra nó, Joseph Redmon.

- YOLOv8 có thể không hoạt động tốt trong tất cả các môi trường và có thể cần thêm điều chỉnh hoặc tối ưu hóa để đạt được hiệu suất tối ưu.

###### Chuẩn Bị Dữ Liệu Huấn Luyện

Để huấn luyện mô hình YOLOv8, ta cần có tập dữ liệu hình ảnh được gắn nhãn và chú thích đối tượng tương ứng của chúng. Các chú thích bao gồm nhãn lớp đối tượng (class) và tọa độ của hộp giới hạn (bounding box) của nó trong ảnh. Dữ liệu này được sử dụng để đào tạo mô hình phát hiện chính xác các đối tượng trong hình ảnh mới.

Việc tạo ra tập dữ liệu tùy chỉnh có thể là một quá trình khó khăn. Có thể mất hàng chục hoặc thậm chí hàng trăm giờ để thu thập hình ảnh, gắn nhãn và chuyển đổi chúng ở định dạng thích hợp. Nhưng ngày nay đã có nhiều tools hỗ trợ việc xử lý dữ liệu như LabelImg, RectLabel hay Roboflow.

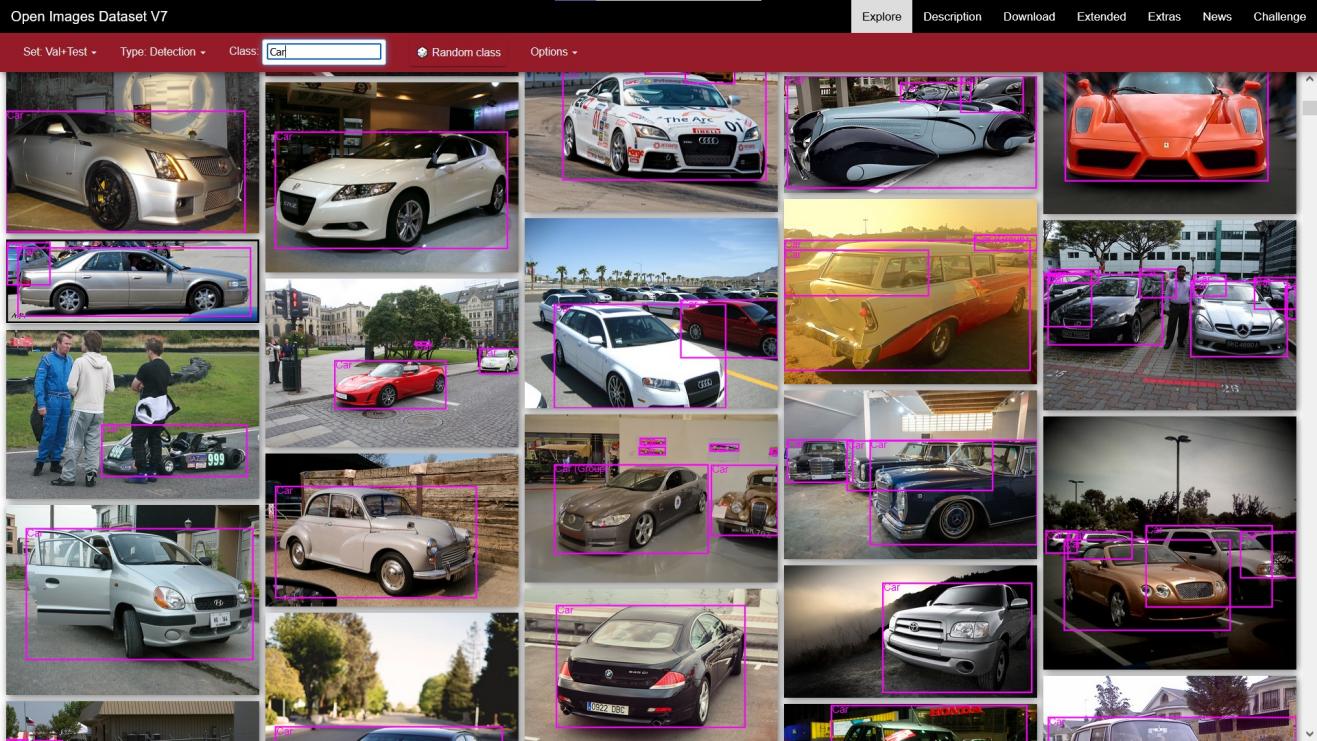
### IMG_256

Hình 7: Huấn luyện YOLOv8

###### Các Bước Tạo Dataset

• Thu thập dữ liệu: Bước đầu tiên trong việc tạo tập dữ liệu tùy chỉnh là xác định vấn đề và loại dữ liệu ta cần. Khi ta đã xác định vấn đề của mình, ta cần thu thập dữ liệu. Điều này có thể liên quan đến việc chụp ảnh, tải xuống hình ảnh từ internet hoặc thu thập dữ liệu từ các nguồn khác.

• Gắn nhãn cho dữ liệu: Sau khi ta đã thu thập dữ liệu, ta cần dán nhãn cho nó. Điều này liên quan đến việc thêm chú thích vào dữ liệu cho biết vị trí và loại của các đối tượng trong ảnh. Ta có thể sử dụng nhiều công cụ khác nhau để gắn nhãn cho dữ liệu của mình.



Hình 8: Nhiều chiếc xe hơi được gắn nhãn (Nguồn: Open Image Dataset v7)

• Chuyển đổi dữ liệu: Tùy thuộc vào khung bạn đang sử dụng, bạn có thể cần chuyển đổi dữ liệu sang một định dạng cụ thể. Ngoài ra, bạn có thể thay đổi kích thước hình ảnh, chuẩn hóa giá trị pixel hoặc áp dụng các kỹ thuật tăng cường dữ liệu để tăng kích thước tập dữ liệu của mình.

###### Quá trình huấn luyện

Việc cài đặt tham số rất quan trọng vì nó quyết định hiệu suất và độ chính xác của mô hình. Tham số là các biến ảnh hưởng đến cách mô hình học và khái quát hóa các mẫu trong dữ liệu.

##### Tổng quan về DeepSort

Deep SORT được Nicolai Wojke và Alex Bewley phát triển ngay SORT nhằm giải quyết các vấn đề thiếu sót liên quan đến số lượng ID switches cao. Hướng giải quyết mà deep SORT đề xuất dựa trên việc sử dụng deep learning để trích xuất các đặc trưng của đối tượng nhằm tăng độ chính xác trong quá trình liên kết dữ liệu. Ngoài ra, một chiến lược liên kết cũng được xây dựng mang tên Matching Cascade giúp việc liên kết các đối tượng sau khi đã biến mất 1 thời gian được hiệu quả hơn.

**Luồng xử lí của Deep SORT được thực hiện tuần tự qua các bước dưới đây :**

*Bước 1:* Sử dụng Faster Region CNN (với backbone là VGG16) để phát hiện các đối tượng trong khung hình hiện tại.

*Bước 2:* Deep SORT sử dụng Kalman Filter để dự đoán các trạng thái track mới dựa trên các track trong quá khứ. Các trạng thái này lúc mới khởi tạo sẽ được gán 1 giá trị mang tính thăm dò (tentative). Giá trị này nếu vẫn đảm bảo duy trì được trong 3 frame tiếp theo, trạng thái sẽ chuyển từ thăm dò sang xác nhận (confirmed), và sẽ cố gắng được duy trì theo dõi trong 30 frame tiếp theo. Ngược lại, nếu mất dấu khi chưa đủ 3 frame, trạng thái sẽ bị xóa khỏi trình theo dõi.

*Bước 3:* Sử dụng những track đã được xác nhận, tiến hành đưa vào chiến lược đối sánh phân tầng (matching cascade) nhằm liên kết với các detection phát hiện được dựa trên độ đo về khoảng cách và đặc trưng.

*Bước 4:* Các track và các detection chưa được liên kết sẽ được đưa đến 1 lớp lọc tiếp theo. Sử dụng giải thuật Hungary giải bài toán phân công với ma trận chi phí IOU để liên kết lần 2

*Bước 5:* Xử lí, phân loại các detection và các track

*Bước 6:* Sử dụng Kalman filter để hiệu chỉnh lại giá trị của track từ những detection đã được liên kết với track và khởi tạo các track mới.

## CHƯƠNG 2: QUÁ TRÌNH NHẬN DẠNG VÀ ĐẾM VẬT NUÔI

#### Chuẩn bị dữ liệu

Hình ảnh của vật nuôi cần nhận dạng càng nhiều và rõ nét thì đạt độ chính xác càng cao. Hình ảnh có thể lấy từ internet, máy ảnh hoặc từ video.

#### Gán nhãn cho đối tượng trong ảnh

Sử dụng công cụ gán nhãn (labelImg,...) để gán nhãn và tạo hộp giới hạn (bounding box) cho đối tượng vật nuôi cần nhận dạng.

#### Huấn luyện

Truyền ảnh và nhãn vào thuật toán để tiến hành huấn luyện. Yolo chủ yếu sử dụng phương pháp học có giám sát. Giải thuật học sử dụng là mạng thần kinh tích chập (CNN). Kiến trúc của YOLO v8 được chia thành hai thành phần chính: xương sống và phần đầu. Xương sống, dựa trên kiến trúc CSPDarknet53 đã được sửa đổi, chịu trách nhiệm trích xuất tính năng từ hình ảnh đầu vào. Nó sử dụng 53 lớp tích chập cùng với các kết nối một phần xuyên giai đoạn để tăng cường luồng thông tin giữa các lớp. Phần đầu của mạng, bao gồm nhiều lớp chập theo sau là các lớp được kết nối đầy đủ, là nơi diễn ra quá trình phát hiện đối tượng thực tế. Phần mạng này dự đoán các hộp giới hạn, điểm đối tượng và xác suất của lớp.

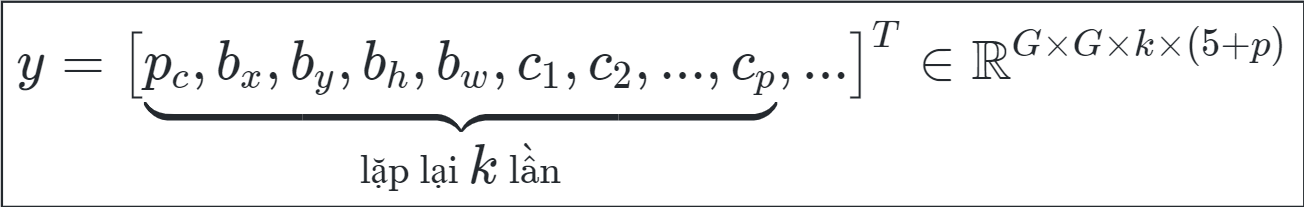
Trong quá trình huấn luyện, YOLOv8 thực hiện nhiều kỹ thuật tăng cường cho ảnh huấn luyện. Một trong những kỹ thuật này là tăng cường dữ liệu mosaic.

Tăng cường dữ liệu mosaic là một kỹ thuật đơn giản, trong đó bốn hình ảnh khác nhau được nối lại với nhau và đưa vào mô hình như đầu vào. Điều này giúp mô hình học được các đối tượng thực sự từ các vị trí khác nhau và trong trạng thái bị che khuất.

#### Nhận dạng đối tượng

Bước 1: Thuật toán sẽ phân chia tấm ảnh đầu vào thành một lưới G x G. Thường thì sẽ là 3×3, 7×7, 9×9... việc chia ô này có ảnh hưởng tới việc mô hình phát hiện đối tượng.

Bước 2: Với mỗi lưới, chạy một mạng CNN dự đoán y có dạng sau:



pc: là có đối tượng nhận dạng trong ô lưới đang xét không. Nếu có thì giá trị là 1, nếu không có thì giá trị là 0.

bx: là tọa độ x của neo.

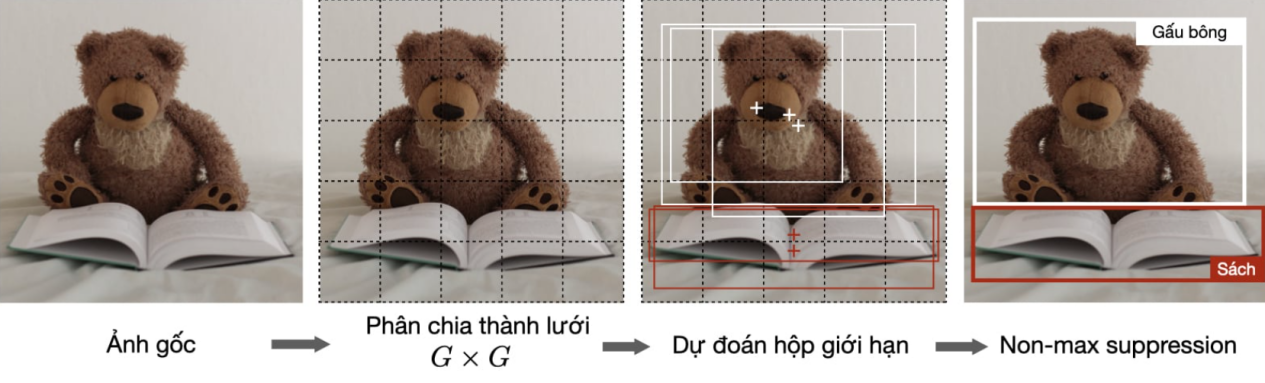
by: là tọa độ y của neo.

bh: là chiều cao của hộp giới hạn.

bw: là chiều ngang của hộp giới hạn.

c1: là đối tượng của class 1 có trong vị trí (bx,by) không. Nếu có là 1 nếu không có thì giá trị là 0.

Bước 3: Chạy thuật toán non-max suppression để loại bỏ bất kỳ hộp giới hạn có khả năng bị trùng lặp.



Hình 9: Nhận dạng đối tượng trong hình

Kỹ thuật non-max suppression hướng tới việc loại bỏ những hộp giới hạn bị trùng chồng lên nhau của cùng một đối tượng bằng cách chọn chiếc hộp có tính đặc trưng nhất. Sau khi loại bỏ tất cả các hộp có xác suất dự đoán nhỏ hơn 0.6, những bước tiếp theo được lặp lại khi vẫn còn tồn tại những hộp khác.

1. **Theo dõi đối tượng**

Xây dựng vùng dùng để theo dõi đối tượng và đếm đối tượng. Khi đối tượng tiến vào vùng theo dõi thì gán nhãn cho đối tượng. Khi đối tượng ra khỏi vùng theo dõi thì xóa nhãn đối tượng.

1. **Đếm vật nuôi**

Đếm khi đối tượng đi qua đường được tạo bằng hàm trong openCV. Cộng số lượng khi con vật tiến tới và qua khỏi đường vạch. Trừ số lượng khi con vật đi ngược lại qua khỏi đường vạch.

# Công trình được trích dẫn

|  |  |
| --- | --- |
| [1] | aws, “Trí tuệ nhân tạo (AI) là gì?,” *aws,* 2020. |
| [2] | PGS.TS Đặng Trần Khánh, “XU HƯỚNG NGHIÊN CỨU VÀ ỨNG DỤNG TRÍ TUỆ,” p. 23, 2018. |
| [3] | VDB, “YOLOv8 có gì nâng cấp so với các phiên bản trước?,” *Vinbigdata,* p. 1, 2023. |
| [4] | phamdinhkhanh, “Các thuật toán Object Detection,” p. 1, 2019. |
| [5] | B. T. Tung, “Một góc nhìn về Object Tracking,” *VIBLO,* p. 1, 2020. |
| [6] | D. N. Đỗ gia Huy, “Mô Hình Nhận Dạng Vật Thể YOLOv8,” *HackMD,* p. 1, 2023. |
| [7] | V. Hoàng, “Tìm hiểu về YOLO trong bài toán real-time object detection,” *VIBLO,* p. 1, 2019. |
| [8] | B. T. Tung, “SORT - Deep SORT : Một góc nhìn về Object Tracking,” *VIBLO,* p. 1, 2021. |
| [9] | Hunonic, “Ứng dụng của trí tuệ nhân tạo trong cuộc sống,” *hunonic,* p. 1, 2022. |
| [10] | D. Đ. Trình, “Các tính năng của OpenCV mà bạn không thể bỏ qua vào năm 2022,” *WEBSITEHCM.com,* p. 1, 2022. |