BỘ GIÁO DỤC VÀ ĐÀO TẠO

# TRƯỜNG ĐẠI HỌC THĂNG LONG

****

**BÁO CÁO BÀI TẬP LỚN MÔN: CÔNG CỤ LẬP TRÌNH AI**

Học máy trong xe tự lái

**GIẢNG VIÊN HƯỚNG DẪN: ThS. NGÔ MẠNH CƯỜNG LỚP: 243\_AI220\_01**

**NHÓM: 6**

**SINH VIÊN : A447704 – VŨ HOÀNG HUY A46129 – NGUYỄN HOÀI NAM A45670 – ĐINH VŨ HẠ**

**HÀ NỘI – 2025**

**MỤC LỤC**

[DANH MỤC HÌNH ẢNH 1](#_bookmark0)

[DANH MỤC BẢNG 2](#_bookmark1)

[CHƯƠNG 1. TỔNG QUAN VỀ ĐỀ TÀI 3](#_bookmark2)

* 1. [Lý do lựa chọn đề tài 3](#_bookmark3)
     1. [Nhu cầu thực tiễn 3](#_bookmark4)
     2. [Áp dụng kiến thức môn học 3](#_bookmark5)
  2. [Lịch sử nghiên cứu 4](#_bookmark6)
     1. [Cách nhận diện giao thông trước đây 4](#_bookmark7)
     2. [Sự phát triển của Machine Learning trong nhận diện giao thông 5](#_bookmark8)
        1. [Nhận diện vật thể (Object Detection) 5](#_bookmark9)
  3. [Mục đích, đối tượng và phạm vi nghiên cứu 6](#_bookmark10)
     1. [Mục đích nghiên cứu 6](#_bookmark11)
     2. [Đối tượng nghiên cứu 6](#_bookmark12)
     3. [Phạm vi nghiên cứu 6](#_bookmark13)
  4. [Phương pháp nghiên cứu 7](#_bookmark14)
     1. [Thu thập dữ liệu 7](#_bookmark15)
     2. [Tiền xử lý dữ liệu 7](#_bookmark16)
     3. [Xây dựng mô hình học máy 7](#_bookmark17)
     4. [Đánh giá mô hình 7](#_bookmark18)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 9](#_bookmark19)

* 1. [Học sâu 9](#_bookmark20)
  2. [Bài toán nhận diện vật thể (Object Detection) 9](#_bookmark21)
  3. [Đánh giá mô hình nhận diện vật thể 9](#_bookmark22)
  4. [YOLO (You Only Look Once) 10](#_bookmark23)
  5. [Nguyên lý hoạt động của YOLO 10](#_bookmark24)
  6. [Thuật toán xây dựng YOLO 10](#_bookmark25)
  7. [Ưu điểm và nhược điểm của YOLO 11](#_bookmark26)
  8. [Các chỉ số đánh giá mô hình 11](#_bookmark27)
     1. [Mean Average Precision (mAP) 11](#_bookmark28)

[CHƯƠNG 3. THỰC NGHIỆM 12](#_bookmark29)

* 1. [Tìm hiểu về dữ liệu 12](#_bookmark30)
     1. [COCO 2017 12](#_bookmark31)
     2. [Mục tiêu thực nghiệm 13](#_bookmark32)
  2. [Tiền xử lý dữ liệu 13](#_bookmark33)
     1. [COCO 2017 13](#_bookmark34)
     2. [Xây dựng mô hình 16](#_bookmark35)
     3. [Kết Luận 18](#_bookmark36)
     4. [Định hướng phát triển 18](#_bookmark37)

[TÀI LIỆU THAM KHẢO 19](#_bookmark38)

# DANH MỤC HÌNH ẢNH

[*Hình 1. Thống kê dữ liệu mỗi lớp trong tập huấn luyện 14*](#_Toc195482615)

[*Hình 2.Thống kê dữ liệu mỗi lớp trong tập kiểm thử 15*](#_Toc195482616)

[*Hình 3.Tệp yaml 16*](#_Toc195482617)

[*Hình 4. Đánh giá trên tập train 17*](#_Toc195482618)

[*Hình 5. Đánh giá trên tập validation 17*](#_Toc195482619)

[*Hình 6. Đánh giá trên tập train 18*](#_Toc195482620)

[*Hình 7. Đánh giá trên tập test 18*](#_Toc195482621)

# CHƯƠNG 1. TỔNG QUAN VỀ ĐỀ TÀI

## Lý do lựa chọn đề tài

* + 1. *Nhu cầu thực tiễn*

## An toàn giao thông và nhu cầu tự động hóa phương tiện

* + Với sự phát triển của công nghệ, xe tự lái đang trở thành xu hướng quan trọng trong ngành công nghiệp ô tô. Một trong những yêu cầu cốt lõi của xe tự lái là khả năng nhận diện vật thể, biển báo giao thông và vạch đường, giúp phương tiện vận hành một cách an toàn và chính xác trong môi trường thực tế.

## Giảm thiểu tai nạn giao thông

* + Tai nạn giao thông do con người gây ra chiếm tỷ lệ lớn do các yếu tố như mất tập trung, phản xạ chậm hoặc vi phạm luật giao thông. Việc ứng dụng các hệ thống phát hiện vật thể (Object Detection), nhận diện biển báo (Traffic Sign Detection) xe có thể tự động nhận biết và phản ứng nhanh hơn con người, từ đó giảm thiểu tai nạn và nâng cao mức độ an toàn.

## Hỗ trợ và tối ưu hóa giao thông đô thị

* + Xe tự lái có khả năng dự đoán, thích ứng với điều kiện giao thông và vận hành tối ưu, giúp giảm ùn tắc tại các khu vực đô thị đông đúc. Việc sử dụng các thuật toán học sâu để nhận diện tín hiệu đèn giao thông, biển báo, phương tiện khác và vạch kẻ đường giúp xe có thể di chuyển linh hoạt và tuân thủ luật giao thông mà không cần sự can thiệp của con người.

## Ứng dụng của Machine Learning trong giao thông

* + Các thuật toán tiên tiến như YOLO (You Only Look Once), Faster R-CNN, SSD (Single Shot MultiBox Detector) đã được chứng minh hiệu quả trong bài toán nhận diện đối tượng, giúp xe tự lái xác định các vật thể như ô tô, người đi bộ, xe đạp.
  + Mạng thần kinh tích chập (CNN) và Deep Learning cũng được sử dụng rộng rãi để nhận diện biển báo giao thông, giúp xe hiểu và tuân thủ quy tắc đường bộ.
    1. *Áp dụng kiến thức môn học*
* **Áp dụng Machine Learning vào bài toán thực tế**: Nhận diện vật thể, biển báo và vạch đường là những bài toán quan trọng trong Computer Vision, giúp sinh viên hiểu rõ cách áp dụng Deep Learning và AI vào thực tế. Việc triển khai mô hình

trong bối cảnh giao thông thực tế giúp rèn luyện kỹ năng phân tích dữ liệu và phát triển mô hình học sâu.

* **Làm quen với các công cụ lập trình trí tuệ nhân tạo**: Dự án giúp sinh viên thực hành với các framework phổ biến trong AI, bao gồm:
  + **OpenCV**: Hỗ trợ xử lý hình ảnh, phát hiện đường biên, nhận diện cạnh.
  + **TensorFlow & PyTorch**: Các thư viện mạnh mẽ để xây dựng và huấn luyện mô hình học sâu.
  + **Scikit-learn**: Hỗ trợ phân tích dữ liệu, đánh giá mô hình và tối ưu thuật toán.

## Lịch sử nghiên cứu

* + 1. *Cách nhận diện giao thông trước đây*
       1. *Nhận diện vật thể (Object Detection - xe cộ, người đi bộ, đèn tín hiệu)*

Hệ thống xe tự lái cần phát hiện các vật thể xung quanh để đưa ra quyết định lái xe phù hợp. Trước đây, các phương pháp phổ biến bao gồm:

* + - * + Histogram of Oriented Gradients (HOG) + Support Vector Machine (SVM)

HOG được sử dụng để trích xuất đặc trưng từ hình ảnh, sau đó SVM được huấn luyện để phân loại đối tượng (xe hơi, người đi bộ, v.v.).

Hạn chế: Chỉ hoạt động tốt với hình dạng đối tượng cố định, không phù hợp với góc nhìn và điều kiện thực tế thay đổi.

* + - * + Viola-Jones Object Detection

Dựa trên đặc trưng Haar để nhận diện người đi bộ.

Không hiệu quả khi có quá nhiều loại vật thể khác nhau.

* + - 1. *Nhận diện biển báo giao thông (Traffic Sign Detection)*

Trước đây, biển báo giao thông được phát hiện dựa trên:

* + - * + Phân đoạn màu sắc (Color Segmentation)

Biển báo giao thông thường có màu sắc đặc trưng (đỏ, xanh, vàng), nên các thuật toán như Thresholding, HSV color space được sử dụng để phát hiện biển báo dựa trên màu sắc.

Hạn chế: Không hoạt động tốt khi biển báo bị phai màu, ánh sáng thay đổi hoặc bị che khuất một phần.

* + - * + So khớp mẫu (Template Matching)

Dùng kỹ thuật Correlation Matching hoặc SIFT (Scale-Invariant Feature Transform) để so sánh biển báo trong ảnh với một tập mẫu có sẵn.

Hạn chế: Kém hiệu quả khi biển báo bị méo, bị xoay hoặc bị che khuất một phần.

* + 1. *Sự phát triển của Machine Learning trong nhận diện giao thông*

Với sự phát triển của Machine Learning, nhiều nghiên cứu đã áp dụng thuật toán học máy vào bài toán dự đoán khả năng vỡ nợ.

* + - 1. *Nhận diện vật thể (Object Detection)*

Các thuật toán nhận diện vật thể giúp xe tự lái phát hiện xe cộ, người đi bộ, đèn tín hiệu với độ chính xác cao:

* + - * + **YOLO (You Only Look Once)** – Redmon et al. (2016, 2018, 2023)

Phát hiện vật thể trong thời gian thực với tốc độ rất cao.

* + - * + **Faster R-CNN** – Ren et al. (2015)

Mô hình chính xác cao nhưng tốc độ xử lý chậm hơn YOLO.

Thích hợp với các bài toán cần độ chính xác cao hơn tốc độ.

* + - 1. *Nhận diện biển báo giao thông (Traffic Sign Detection & Classification)*

Deep Learning giúp cải thiện khả năng nhận diện biển báo bằng các mô hình:

## CNN (Convolutional Neural Networks)

Được sử dụng để phân loại biển báo với độ chính xác cao.

Một số mô hình phổ biến: ResNet, MobileNet, EfficientNet.

## GTSRB Dataset & Các nghiên cứu ứng dụng CNN

Bộ dữ liệu GTSRB (German Traffic Sign Recognition Benchmark) thường được sử dụng để huấn luyện mô hình nhận diện biển báo.

CNN đạt độ chính xác > 95% khi nhận diện biển báo giao thông trong điều kiện thực tế.

.

## Mục đích, đối tượng và phạm vi nghiên cứu

* + 1. *Mục đích nghiên cứu*

## Phát triển hệ thống nhận diện thông minh hỗ trợ xe tự lái:

* + Ứng dụng Machine Learning và Computer Vision để xây dựng hệ thống nhận diện vật thể (Object Detection), nhận diện biển báo giao thông (Traffic Sign Detection).
  + Góp phần nâng cao độ an toàn của xe tự lái bằng cách giúp phương tiện hiểu và tương tác chính xác với môi trường giao thông.

## Tăng cường khả năng nhận diện biển báo giao thông (Traffic Sign Detection):

* + Phát triển hệ thống phát hiện và phân loại biển báo giao thông giúp xe tự lái tuân thủ luật lệ đường bộ.
  + Sử dụng Deep Learning (CNN, ResNet, MobileNet) để nâng cao độ chính xác ngay cả khi biển báo bị che khuất, xoay hoặc trong điều kiện ánh sáng kém.

## Cải tiến hệ thống nhận diện vật thể (Object Detection) trong giao thông:

* + Tích hợp hệ thống phát hiện người đi bộ, phương tiện, đèn tín hiệu giao thông để giúp xe đưa ra quyết định lái xe an toàn hơn..
    1. *Đối tượng nghiên cứu*
* **Dữ liệu nghiên cứu**: một phần của bộ Coco Dataset, bộ dữ liệu GTSRB

## Các yếu tố nghiên cứu:

* + **Phương pháp nhận diện vật thể (Object Detection):** Xe, người đi bộ, xe đạp, tín hiệu đèn giao thông.
  + **Phương pháp nhận diện biển báo giao thông (Traffic Sign Detection):** Biển báo giới hạn tốc độ, biển báo cấm, biển báo chỉ dẫn,…
* Thuật toán và mô hình Deep Learning
  + 1. *Phạm vi nghiên cứu*
* Phạm vi dữ liệu
  + **Nguồn dữ liệu**: Tập Coco 2017 dataset
  + **Nguồn dữ liệu:** German Traffic Sign Recognition Benchmark
* Phạm vi thời gian
  + Mô hình huấn luyện trên dữ liệu quá khứ nhưng có thể áp dụng trong điều kiện giao thông hiện tại và tương lai.
* Phạm vi kỹ thuật
  + Thuật toán Machine Learning: YOLOv8 và YOLOv11
  + Đánh giá mô hình:
    - Sử dụng mAP50 và FPS cho nhận diện vật thể và phân loại biển báo.
  + Công cụ và thư viện sử dụng:
    - Python, TensorFlow, OpenCV, Scikit-learn, ultralytics,...

## Phạm vi ứng dụng:

* + Xe tự lái và hệ thống hỗ trợ lái xe thông minh (ADAS).
  + Phân tích giao thông thông minh cho thành phố hiện đại.
  + Cảnh báo nguy hiểm cho người lái xe trong hệ thống hỗ trợ lái xe nâng cao (ADAS).

## Phương pháp nghiên cứu

* + 1. *Thu thập dữ liệu*
* Sử dụng một phần bộ dữ liệu COCO 2017.
* Sử dụng bộ dữ liệu GTSRB
  + 1. *Tiền xử lý dữ liệu*
* Do kích thước của 2 tập dữ liệu quá lớn nên
  + Object Detection: chỉ sử dụng 7 lớp liên quan tới các phương tiện đường bộ và giữ lại 1500 ảnh cho mỗi lớp
  + Traffic Sign Detection: giữ lại tối đa 5000 ảnh ngẫu nhiên cũng như các nhãn tương ứng để huấn luyện mô hình
    1. *Xây dựng mô hình học máy*

 Sử dụng mô hình YOLO cho bài toán Object Detection và Traffic Sign Detection

* + 1. *Đánh giá mô hình.*
* Đánh giá hiệu suất dự đoán bằng các chỉ số:
  + mAP50 (mean Average Precision at IoU 0.5): đánh giá độ chính xác khi chỉ chấp nhận các hộp dự đoán có IoU >= 0.5 so với hộp thực tế
  + FPS (Frame per second): Số frame xử lý trong 1 giây

# CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

## Học sâu

* Học sâu là một nhánh của học máy, trong đó các mô hình học sử dụng mạng nơ- ron nhân tạo với nhiều lớp (layers) để học từ dữ liệu.
* Mô hình học từ dữ liệu quá khứ và tự động trích xuất các đặc trưng phức tạp từ dữ liệu đầu vào.
* Trong bài toán nhận diện vật thể trong xe tự lái:
  + Đầu vào: Hình ảnh từ camera hoặc video (các khung hình chứa các đối tượng cần nhận diện).
  + Đầu ra: Các nhãn tương ứng với các đối tượng nhận diện, ví dụ: xe, người đi bộ, biển báo giao thông, v.v.
  + Mô hình học từ các tập dữ liệu hình ảnh để nhận diện và phân loại các đối tượng trong cảnh quan giao thông.

## Bài toán nhận diện vật thể (Object Detection)

* Nhận diện vật thể là bài toán trong học máy và học sâu, trong đó mục tiêu là phát hiện và phân loại các đối tượng trong một hình ảnh hoặc video.
* Bài toán nhận diện vật thể bao gồm hai phần chính:
  + **Phát hiện (Detection)**: Xác định vị trí của các đối tượng trong hình ảnh (thông qua các hộp giới hạn - bounding boxes).
  + **Phân loại (Classification)**: Gán nhãn cho các đối tượng đã được phát hiện, ví dụ: nhận diện xe, người, biển báo giao thông, v.v.
* Trong xe tự lái, bài toán nhận diện vật thể là rất quan trọng vì nó giúp hệ thống hiểu được môi trường xung quanh, nhận diện các mối nguy hiểm như người đi bộ, xe cộ, biển báo, v.v., và hỗ trợ trong việc điều khiển xe an toàn.

## Đánh giá mô hình nhận diện vật thể

* mAP50 (mean Average Precision): Chỉ số chính được sử dụng trong nhận diện vật thể, tính toán độ chính xác trung bình của mô hình ở tất cả các lớp đối tượng và ở các mức độ khác nhau của độ chính xác.
* FPS (Frame Per Second): Số lượng khung hình xử lý được trong một giây.

## YOLO (You Only Look Once)

* YOLO là một mô hình học sâu trong nhận diện vật thể, được sử dụng để phát hiện và phân loại các đối tượng trong một hình ảnh hoặc video.
* Mô hình YOLO nhận diện tất cả các đối tượng trong một lần duy nhất (do đó có tên "You Only Look Once"), thay vì phân tách nhiệm vụ nhận diện thành nhiều bước như các phương pháp truyền thống.
* Mục tiêu: Dự đoán vị trí và loại của các đối tượng trong một hình ảnh hoặc video, đồng thời đánh dấu chúng bằng các hộp giới hạn (bounding boxes).
* Ứng dụng trong bài toán xe tự lái: YOLO giúp nhận diện các đối tượng trong môi trường giao thông, chẳng hạn như xe cộ, người đi bộ, biển báo giao thông, v.v., hỗ trợ cho hệ thống điều khiển của xe tự lái.

## Nguyên lý hoạt động của YOLO

* YOLO chia mỗi hình ảnh thành một lưới và dự đoán các hộp giới hạn và nhãn của các đối tượng trong từng ô lưới.
  + **Lưới phân mảnh:** Mô hình chia hình ảnh thành các ô lưới nhỏ (ví dụ 13x13 hoặc 19x19), và mỗi ô sẽ chịu trách nhiệm dự đoán một số đối tượng.
  + **Dự đoán trong mỗi ô:** Mỗi ô lưới sẽ dự đoán**:**
    - Các hộp giới hạn (bounding boxes) với vị trí, chiều rộng và chiều cao.
    - Xác suất của mỗi đối tượng (nhãn) trong ô lưới đó.
* Các dự đoán của mô hình YOLO được tổng hợp lại để tạo thành kết quả cuối cùng, bao gồm các đối tượng được phát hiện cùng với vị trí và nhãn của chúng.

## Thuật toán xây dựng YOLO

YOLO xây dựng mô hình nhận diện vật thể bằng cách tối ưu hóa một hàm mất mát tổng hợp các yếu tố:

* Mất mát về vị trí (Localization Loss): Mất mát về độ chính xác trong việc dự đoán vị trí của các hộp giới hạn.
* Mất mát về nhãn (Classification Loss): Mất mát về độ chính xác trong việc phân loại các đối tượng.
* Mất mát về sự hiện diện của đối tượng (Confidence Loss): Mất mát về sự tự tin của mô hình đối với dự đoán của mình.

Các giá trị này được tính cho từng hộp giới hạn và các đối tượng phát hiện, từ đó điều chỉnh các tham số của mô hình qua quá trình huấn luyện.

## Ưu điểm và nhược điểm của YOLO

* Ưu điểm:
  + **Nhanh chóng**: YOLO nhận diện tất cả các đối tượng trong một lần duy nhất, giúp tăng tốc độ xử lý và phù hợp với các ứng dụng thời gian thực, như xe tự lái.
  + **Chính xác**: YOLO có thể nhận diện nhiều đối tượng trong một hình ảnh và đưa ra các dự đoán về cả vị trí và nhãn đối tượng cùng một lúc.
* Nhược điểm:
  + Giới hạn trong việc nhận diện các đối tượng nhỏ: YOLO gặp khó khăn khi nhận diện các đối tượng nhỏ trong một hình ảnh, vì mô hình chia hình ảnh thành các ô lưới lớn, làm giảm độ chính xác của các đối tượng nhỏ.
  + Khả năng phân loại thấp trong một số trường hợp: Mặc dù YOLO rất nhanh, nhưng đôi khi độ chính xác của các phân loại có thể không cao, đặc biệt là khi có nhiều lớp đối tượng trong một hình ảnh.
* Tuy nhiên những nhược điểm này đã được cải thiện qua từng phiên bản YOLO với các phương pháp tối ưu hóa để cải thiện độ chính xác, đặc biệt trong việc phát hiện đối tượng nhỏ và phân loại đối tượng phức tạp.

## Các chỉ số đánh giá mô hình

### Mean Average Precision (mAP)

* + - * Định nghĩa: mAP là chỉ số chính trong Object Detection để đánh giá chất lượng của mô hình. Nó tính toán độ chính xác (Precision) trung bình trên tất cả các lớp đối tượng, sau đó lấy giá trị trung bình của các độ chính xác đó.
      * Ý nghĩa: mAP giúp đánh giá tổng thể về khả năng phát hiện vật thể của mô hình trong suốt toàn bộ bộ dữ liệu.
      * Ứng dụng: Đây là chỉ số quan trọng nhất khi đánh giá các mô hình như YOLO (You Only Look Once) hoặc Faster R-CNN.

# CHƯƠNG 3. THỰC NGHIỆM

## Tìm hiểu về dữ liệu

* + 1. *COCO 2017*
       1. *Tổng quan*

COCO 2017 được phát triển bởi nhóm nghiên cứu tại Microsoft và là một trong những tập dữ liệu phổ biến nhất trong lĩnh vực thị giác máy tính. Tập dữ liệu này chứa các hình ảnh với các đối tượng được chú thích chi tiết trong bối cảnh thực tế, giúp các mô hình học máy phát triển khả năng nhận diện đối tượng và phân tích ngữ nghĩa từ ảnh.

* + - 1. *Cấu trúc của COCO 2017*

Số lượng hình ảnh và chú thích

* Số lượng ảnh huấn luyện: 118,000 ảnh
* Số lượng ảnh kiểm tra: 5,000 ảnh
* Số lượng ảnh thử nghiệm: 41,000 ảnh
* Số lượng lớp (categories): 80 lớp đối tượng
* Số lượng chú thích: Khoảng 2,5 triệu chú thích cho các đối tượng trong ảnh Nội dung chú thích
* COCO 2017 bao gồm nhiều dạng chú thích, giúp mô hình hiểu và nhận diện đối tượng trong ảnh một cách chính xác hơn:
  + Bounding boxes: Chú thích hộp giới hạn cho các đối tượng.
  + Segmentation masks: Mặt nạ phân đoạn cho các đối tượng (instance segmentation).
  + Keypoints: Các điểm quan trọng cho đối tượng (dành cho nhận diện cơ thể người).
  + Image captions: Mô tả ngắn gọn cho mỗi hình ảnh, giúp mô hình học cách tạo chú thích văn bản cho ảnh.
    - 1. *Cấu trúc của dữ liệu COCO 2017*

Tập dữ liệu COCO 2017 được lưu trữ dưới dạng các tệp JSON, trong đó chứa thông tin chi tiết về các ảnh, chú thích và danh mục đối tượng.

Các phần chính trong file JSON:

* images: Mỗi ảnh trong tập dữ liệu sẽ có thông tin bao gồm id, file\_name, width, height.
* annotations: Chứa các chú thích cho từng đối tượng trong ảnh, bao gồm thông tin về category\_id (danh mục đối tượng), bbox (bounding box), segmentation (segmentation mask), iscrowd (đánh dấu nếu đối tượng bị che khuất).
* categories: Danh sách các đối tượng trong ảnh với thông tin như id, name (tên lớp), supercategory (nhóm lớp).
  + 1. *Mục tiêu thực nghiệm*

 Mô hình có thể phát hiện và phân loại vật thể.

## Tiền xử lý dữ liệu

* + 1. *COCO 2017*
       1. *FiftyOne*

FiftyOne là một thư viện Python mạnh mẽ được thiết kế để giúp bạn dễ dàng kiểm tra, phân tích và đánh giá các mô hình Machine Learning (ML), đặc biệt là trong các bài toán Computer Vision như Object Detection. Nó cung cấp các công cụ trực quan hóa và phân tích dữ liệu, giúp bạn hiểu rõ hơn về hiệu suất của mô hình cũng như các vấn đề trong dữ liệu của mình.

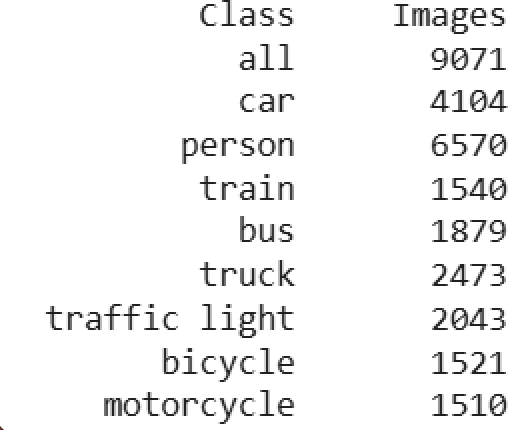
Trong dự án này FiftyOne được sử dụng để:

* **Tải dữ liệu**: FiftyOne giúp tải bộ dữ liệu COCO 2017 cho các tập huấn luyện và kiểm tra từ kho dữ liệu sẵn có.
* **Lọc dữ liệu**: Lọc các đối tượng trong bộ dữ liệu chỉ giữ lại những lớp quan trọng (như "person", "car", "bus", v.v.).
* **Nhóm dữ liệu theo lớp**: Dữ liệu được nhóm theo từng lớp (class), và chỉ giữ lại những mẫu dữ liệu có đối tượng thuộc các lớp này.
* **Chọn ngẫu nhiên dữ liệu**: Mẫu dữ liệu được chọn ngẫu nhiên với giới hạn tối đa 1500 ảnh cho mỗi lớp để đảm bảo sự phân bố hợp lý.
* **Xuất dữ liệu**: Dữ liệu được xuất ra dưới dạng chuẩn YOLOv5, sẵn sàng cho việc huấn luyện mô hình nhận diện vật thể.
* **Thống kê dữ liệu**: Đếm và hiển thị số lượng ảnh trong mỗi lớp sau khi chọn dữ liệu.
  + - 1. *Lọc dữ* liệu *COCO 2017*

Như đã thấy với những thông số trên về tập dữ liệu COCO 2017, đây là một tập dữ liệu rất lớn, nếu sử dụng máy cá nhân thì việc dùng toàn bộ dữ liệu cho việc huấn luyện mô hình là rất khó. Vì thế, nhóm em sẽ chỉ sử dụng một phần rất nhỏ của tập dữ liệu. Nhóm sẽ sử dụng các bước sau để lọc sử liệu:

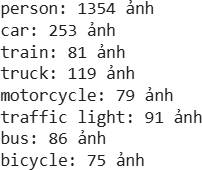
* Lọc số lượng lớp: bộ dữ liệu có đến 80 lớp đối tượng, nhóm sẽ chỉ sử dụng 8 lớp mà nhóm nhận thấy là quan trọng nhất trong bài toán nhận diện vật thể cho xe tự lái với phạm vi là đường bộ:
  + Person: người đi đường
  + Train: tàu
  + Car: ô tô
  + Motorcycle: xe máy
  + Bus: xe buýt
  + Bicycle: xe đạp
  + Truck: xe tải
  + Traffic light: đèn tín hiệu
* Lọc số lượng ảnh: lấy ngẫu nhiên 1500 ảnh từ mỗi lớp.
  + - 1. *Kết quả sau khi lọc*

Sau khi trải qua 2 bước lọc dữ liệu, lượng dữ liệu cuối sẽ gồm tổng cộng khoảng 9000 dữ liệu tất cả với 8 lớp. Lượng dữ liệu mỗi lớp là:



Hình 1. Thống kê dữ liệu mỗi lớp trong tập huấn luyện

Còn lọc dữ liệu đối với tập kiểm thử (validation), sau khi lọc với 8 lớp, dữ liệu từ 5000 giảm còn khoảng 3000, sau đó nhóm lấy ngẫu nhiên 1 nửa số ảnh. Vì vậy, sau khi lọc dữ liệu tập kiểm thử sẽ còn khoảng 1500 ảnh với phân bố từng lớp như sau:



Hình 2.Thống kê dữ liệu mỗi lớp trong tập kiểm thử

* + - 1. *Xuất dữ liệu sang định dạng chuẩn của YOLO*

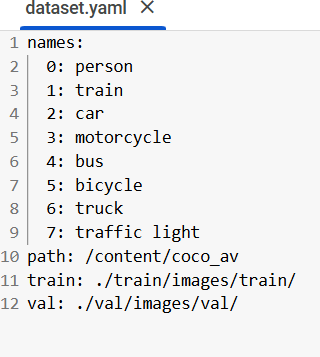
COCO 2017 sử định dạng JSON để lưu trữ thông tin về ảnh và nhãn trong khi đó YOLO sử dụng một định dạng tệp khác, mỗi ảnh trong YOLO có một tệp nhãn riêng, với một dòng riêng biệt cho mỗi đối tượng. Vì vậy, cần phải chuyển định dạng sang định dạng chuẩn của YOLO để sử dụng mô hình YOLO cho việc huấn luyện

Trong thư viện FiftyOne đã có sẵn hàm để xử lý vấn đề này.

* + - 1. *Tệp YAML*

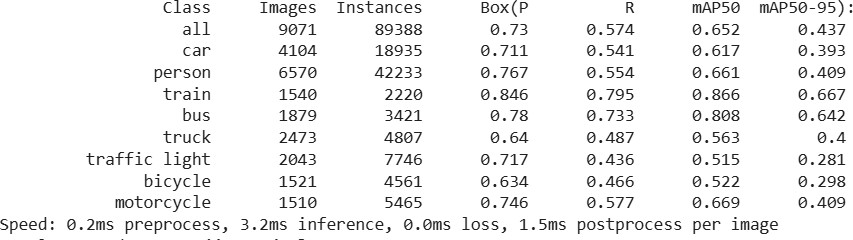
Tệp YAML (YAML Ain't Markup Language) là một định dạng dữ liệu dễ đọc, dễ hiểu và dễ chỉnh sửa cho con người. Nó được sử dụng để mô tả các cấu hình, dữ liệu và cấu trúc trong nhiều ứng dụng khác nhau, đặc biệt trong các hệ thống phần mềm và lập trình.

Trong YOLO (You Only Look Once), tệp YAML được sử dụng để cấu hình thông tin về các tham số của mô hình, tập dữ liệu, số lớp (classes), và các tham số khác khi huấn luyện mô hình. Cấu trúc YAML cho YOLO rất quan trọng vì nó giúp YOLO hiểu về dữ liệu và cách xử lý dữ liệu trong quá trình huấn luyện.

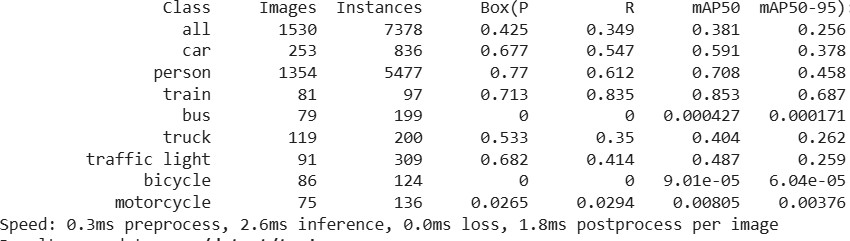


Hình 3.Tệp yaml

* + 1. *Xây dựng mô hình*
       1. *Object Detection*
* Sử dụng mô hình YOLO11 bản nano để huấn luyện với các tham số sau:
  + Epochs = 20
  + Imgsz = 640
  + Batch = 8
  + Workers = 2
    - * 1. *Đánh giá mô hình*

**

Hình 4. Đánh giá trên tập train

**

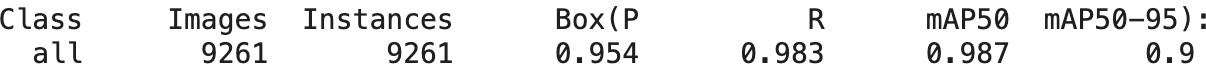
Hình 5. Đánh giá trên tập validation

* + - * 1. *Đánh giá hiệu xuất mô hình*

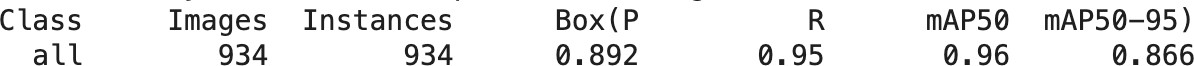
Mô hình mất khoảng 10 – 20 ms để xử lý một khung hình.

FPS trên video đầu ra dao động từ 30 – 40 FPS.

* + - 1. *Traffic Sign Detection*
* Sử dụng mô hình YOLO11 bản nano để huấn luyện với các tham số sau:
  + Epochs = 15
  + Imgsz = 640
  + Batch = 8
  + Workers = 2
    - * 1. *Đánh giá mô hình*

**

Hình 6. Đánh giá trên tập train

**

Hình 7. Đánh giá trên tập test

* + - * 1. *Đánh giá hiệu suất mô hình*

- Tốc độ xử lý 1 frame dao động từ 15 – 20ms/frame ảnh.

* + 1. *Kết Luận*

Cả 2 mô hình đều có tốc độ xử lý nhanh, tuy nhiên mô hình object detection có kết quả chưa thực sự tốt, cần cải thiện thêm. Với mô hình Traffic Sign có kết quả khá tốt.

* + 1. *Định hướng phát triển*

Cải thiện hiệu suất và độ chính xác của mô hình Object Detection. Mở rộng tập dữ liệu biển báo (nhiều quốc gia) cho mô hình Traffic Sign Detection. Tích hợp 2 mô hình và mở rộng thêm các mô hình khác (Lane Detection,…)

# TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | Pandas, “Pandas Documentation,” [Online]. Available: https://pandas.pydata.org. |
| [2] | Matplotlib, “Matplotlib Documentation,” [Online]. Available: https://matplotlib.org/. |
| [3] | Seaborn, “Seaborn Documentation,” [Online]. Available: https://seaborn.pydata.org/. |
| [4] | S. learn, “Scikit learn document,” [Online]. Available: https://scikit- learn.org/stable/user\_guide.html. |
| [5] | XGBoost, “XGBoost document,” [Online]. Available: https://xgboost.readthedocs.io/en/release\_3.0.0/. |
| [6] | machinelearningcoban, “Machinelearningcoban về PCA,” [Online]. Available: https://machinelearningcoban.com/2017/06/15/pca/. |
| [7] | machinelearningcoban, “Machinelearningcoban về random\_forest,” [Online]. Available: https://machinelearningcoban.com/tabml\_book/ch\_model/random\_forest.html. |
| [8] | T. R. S. K. J. M. H. J. H. Connor Carreras, Principles of Data Wrangling: Practical Techniques for Data Preparation. |