**TRƯỜNG ĐẠI HỌC LÂM NGHIỆP**

**KHOA CƠ ĐIỆN CÔNG TRÌNH**



**BÁO CÁO ĐỒ ÁN**

**Môn học: Khai phá dữ liệu**

**Đề tài:**

**Xây dựng mô hình và ứng dụng demo để nhận diện đối tượng qua hình ảnh.**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn:** | **ThS. Mai Hà An** |
| **Sinh viên thực hiện:**  **Mã sinh viên:** | **Vũ Huy Hoàng**  **2174801040732** |
| **Lớp:** | **K66B-HTTT** |

**Hà Nội, tháng 4 năm 2025**

**MỤC LỤC**

[**CHƯƠNG 1: TỔNG QUAN VẤN ĐỀ** 3](#_Toc195215196)

[**1.1. Giới thiệu** 3](#_Toc195215197)

[**1.2. Huấn luyện mô hình là gì?** 3](#_Toc195215198)

[**1.3. Quy trình huấn luyện mô hình** 3](#_Toc195215199)

[**1.4. Giới thiệu về YOLO** 4](#_Toc195215200)

[**CHƯƠNG 2: MỤC TIÊU VÀ PHƯƠNG PHÁP NGHIÊN CỨU** 5](#_Toc195215201)

[**2.1. Mục tiêu** 5](#_Toc195215202)

[**2.2. Đối tượng nghiên cứu** 5](#_Toc195215203)

[**2.3. Phương pháp nghiên cứu** 5](#_Toc195215204)

[**2.4. Phạm vi nghiên cứu** 6](#_Toc195215205)

[**2.5. Công cụ và nền tảng sử dụng** 6](#_Toc195215206)

[**CHƯƠNG 3: KẾ HOẠCH THỰC HIỆN** 7](#_Toc195215207)

[**3.1. Dữ liệu** 7](#_Toc195215208)

[**3.2. Gán nhãn** 7](#_Toc195215209)

[**3.3. Chuẩn bị dữ liệu** 7](#_Toc195215210)

[**3.4. Huấn luyện mô hình** 8](#_Toc195215211)

[**3.5. Ứng dụng Web Demo** 8](#_Toc195215212)

[**CHƯƠNG 4. ĐÁNH GIÁ, NHẬN XÉT** 9](#_Toc195215213)

[**4.1. Đánh giá** 9](#_Toc195215214)

[**4.2. Nhận xét** 9](#_Toc195215215)

[**4.3. Hướng cải thiện** 9](#_Toc195215216)

[**CHƯƠNG 5.KẾT LUẬN** 10](#_Toc195215217)

# **CHƯƠNG 1: TỔNG QUAN VẤN ĐỀ**

## **1.1. Giới thiệu**

Trong thời đại công nghệ phát triển mạnh mẽ, trí tuệ nhân tạo (AI), đặc biệt là các mô hình học sâu trong lĩnh vực thị giác máy tính (Computer Vision), ngày càng được ứng dụng rộng rãi trong các lĩnh vực như an ninh, giao thông, đô thị thông minh và đời sống hàng ngày. Một trong những ứng dụng nổi bật của thị giác máy tính là nhận diện đối tượng (object detection), giúp hệ thống tự động xác định, phân loại và theo dõi các đối tượng xuất hiện trong ảnh hoặc video.

Trong đề tài này, ta tập trung vào việc xây dựng một mô hình học sâu có khả năng nhận diện các đối tượng phổ biến trong môi trường thực tế, cụ thể là: con người, ô tô, xe máy, ghế và đèn đường. Việc mô hình có thể nhận diện chính xác những đối tượng trên không chỉ hỗ trợ cho các hệ thống giám sát và phân tích hành vi, mà còn mở rộng tiềm năng ứng dụng vào các lĩnh vực như thành phố thông minh và xe tự hành.

## **1.2. Huấn luyện mô hình là gì?**

Huấn luyện mô hình là quá trình “dạy” cho một thuật toán học máy hiểu và nhận diện các đặc trưng từ dữ liệu. Trong thị giác máy tính, huấn luyện mô hình nghĩa là cung cấp cho mô hình một tập ảnh đã được gán nhãn để nó học cách phân biệt các đối tượng khác nhau trong ảnh.

Quy trình này yêu cầu một lượng lớn dữ liệu đầu vào được chuẩn bị kỹ càng, đồng thời cần chọn đúng kiến trúc mô hình phù hợp với mục tiêu bài toán. Sau khi huấn luyện thành công, mô hình sẽ có khả năng nhận diện chính xác các đối tượng trong những hình ảnh mới chưa từng gặp.

## **1.3. Quy trình huấn luyện mô hình**

Quy trình huấn luyện bao gồm nhiều bước chính:

* **Thu thập dữ liệu**: Thu thập hình ảnh có chứa các đối tượng cần nhận diện (người, xe máy, ô tô, v.v.) từ nhiều bối cảnh, điều kiện ánh sáng và góc chụp khác nhau.
* **Tiền xử lý dữ liệu**: Chuẩn hóa kích thước ảnh, đổi tên, chuyển định dạng, phân chia tập train/test.
* **Gán nhãn dữ liệu**: Dùng công cụ như Roboflow để xác định vị trí và loại đối tượng trong ảnh, lưu theo định dạng YOLO.
* **Huấn luyện mô hình**: Sử dụng mô hình YOLOv8 để học các đặc trưng đối tượng từ dữ liệu.
* **Đánh giá mô hình**: Đánh giá hiệu quả mô hình bằng các chỉ số như mAP, precision, recall.
* **Triển khai mô hình**: Tích hợp mô hình vào một ứng dụng thực tế như hệ thống camera giám sát, web demo nhận diện đối tượng, v.v.

## **1.4. Giới thiệu về YOLO**

YOLO (You Only Look Once) là một trong những mô hình nổi bật nhất trong lĩnh vực nhận diện đối tượng thời gian thực, được đánh giá cao bởi tốc độ xử lý nhanh và độ chính xác cao. YOLO chia ảnh đầu vào thành các lưới nhỏ và thực hiện nhận diện đồng thời toàn bộ ảnh trong một bước duy nhất, thay vì hai bước như các mô hình truyền thống.

Các phiên bản phát triển của YOLO bao gồm:

* **YOLOv1 – v3**: Các phiên bản đầu với tốc độ tốt, dần cải thiện khả năng nhận diện đối tượng nhỏ và đa lớp.
* **YOLOv4**: Phiên bản mạnh mẽ được cộng đồng tối ưu hóa cho cả tốc độ lẫn độ chính xác.
* **YOLOv5**: Phiên bản nổi bật với khả năng huấn luyện nhanh, cài đặt đơn giản.
* **YOLOv6 & v7**: Tập trung vào hiệu suất trên thiết bị di động và nhúng.
* **YOLOv8**: Phiên bản mới nhất (Ultralytics), hỗ trợ đa tác vụ như classification, detection, segmentation, tracking với giao diện dễ sử dụng.

# **CHƯƠNG 2: MỤC TIÊU VÀ PHƯƠNG PHÁP NGHIÊN CỨU**

## **2.1. Mục tiêu**

Mục tiêu của đề tài là xây dựng một hệ thống nhận diện đối tượng có khả năng:

* Phát hiện và phân loại chính xác các đối tượng phổ biến như con người, ô tô, xe máy, ghế và đèn đường trong ảnh.
* Áp dụng mô hình vào thực tế thông qua một ứng dụng Web Demo, giúp người dùng có thể tải ảnh hoặc sử dụng webcam để phát hiện đối tượng theo thời gian thực.

## **2.2. Đối tượng nghiên cứu**

Đối tượng nghiên cứu bao gồm 5 loại thường gặp trong môi trường đô thị và công cộng:

1. Con người
2. Ô tô
3. Xe máy
4. Ghế đá
5. Đèn đường

Việc nhận diện chính xác các đối tượng này giúp phục vụ hiệu quả trong các lĩnh vực như:

* Giám sát an ninh trong thành phố
* Điều phối giao thông thông minh
* Robot tự hành nhận diện chướng ngại vật
* Hệ thống hỗ trợ người khiếm thị trong việc định vị vật thể

## **2.3. Phương pháp nghiên cứu**

Dự án sử dụng phương pháp kết hợp giữa:

* **Nghiên cứu lý thuyết**: Tìm hiểu các kiến trúc YOLO, đánh giá các phiên bản từ v5 đến v8, chọn phiên bản phù hợp với yêu cầu bài toán.
* **Thực nghiệm**: Tiến hành thu thập ảnh, gán nhãn, huấn luyện mô hình, đánh giá và triển khai thử nghiệm mô hình trên nền tảng web.

## **2.4. Phạm vi nghiên cứu**

* Nhận diện 5 đối tượng: con người, ô tô, xe máy, ghế đá và đèn đường.
* Sử dụng mô hình YOLOv8 phiên bản nhẹ (YOLOv8n) để tối ưu tài nguyên huấn luyện trên Google Colab.
* Sử dụng dữ liệu ảnh tĩnh trong quá trình huấn luyện và thử nghiệm.
* Xây dựng ứng dụng demo dạng web phục vụ thử nghiệm mô hình sau huấn luyện.

## **2.5. Công cụ và nền tảng sử dụng**

* **Google Colab**: Môi trường huấn luyện mô hình YOLOv8 với GPU miễn phí (Tesla T4).
* **Roboflow**: Công cụ gán nhãn ảnh với định dạng đầu ra tương thích YOLO.
* **Ultralytics YOLOv8**: Thư viện chính để huấn luyện và chạy mô hình nhận diện đối tượng.
* **Python 3.10**: Ngôn ngữ lập trình chính cho quá trình huấn luyện và demo.
* **Flask / FastAPI**: Framework nhẹ để xây dựng ứng dụng Web demo.

# **CHƯƠNG 3: KẾ HOẠCH THỰC HIỆN**

## **3.1. Dữ liệu**

* **Số lượng ảnh**: 50 ảnh cho mỗi loại đối tượng, tổng cộng 250 ảnh.
* **Kích thước ảnh**: Chuẩn hóa về 640x640 pixel.
* **Định dạng nhãn**: Sử dụng định dạng YOLO (label.txt).
* **Công cụ gán nhãn**: Roboflow.
* **Nguồn ảnh**: Tự thu thập từ thiết bị di động.

## **3.2. Gán nhãn**

Công cụ sử dụng: Roboflow – một nền tảng hỗ trợ xử lý dữ liệu thị giác máy tính, bao gồm cả gán nhãn, chuyển đổi định dạng và chia tách dữ liệu.

Quy trình gán nhãn:

* Tạo một project mới trên Roboflow và đặt tên phù hợp với nội dung nghiên cứu.
* Upload toàn bộ ảnh vào dự án.
* Tiến hành vẽ bounding box và gán nhãn cho từng đối tượng trong ảnh với tên lớp tương ứng: person, car, motorbike, chair, streetlight.
* Xuất dữ liệu sau khi hoàn tất dưới định dạng YOLOv8.
* Roboflow sẽ tự động chia tập dữ liệu thành: train, val, test.

## **3.3. Chuẩn bị dữ liệu**

Sau khi export từ Roboflow, tải xuống bộ dữ liệu đã chia theo cấu trúc:

├── images/

│ ├── train/

│ ├── val/

├── labels/

│ ├── train/

│ ├── val/

Tải dữ liệu này lên **Google Drive**, tạo liên kết đến **Google Colab** để huấn luyện.

## **3.4. Huấn luyện mô hình**

Cài đặt thư viện Ultralytics trên Google Colab.

Tạo file .yaml mô tả dataset (gồm đường dẫn ảnh và danh sách tên lớp).

Sử dụng lệnh yolo detect train để huấn luyện mô hình với các tham số đã cấu hình.

Trong quá trình huấn luyện, theo dõi:

* Loss (objectness, classification, bbox regression)
* mAP@0.5 và mAP@0.5:0.95
* Precision / Recall

Nếu cần, có thể dừng huấn luyện sớm (early stopping) hoặc fine-tune với mô hình tốt nhất (best.pt).

## **3.5. Ứng dụng Web Demo**

Xây dựng một ứng dụng Web demo sử dụng Flask.

Tính năng chính:

* Người dùng tải ảnh lên hoặc chụp ảnh trực tiếp từ webcam.
* Hệ thống chạy mô hình YOLOv8 đã huấn luyện để nhận diện các đối tượng.
* Kết quả trả về là ảnh có vẽ bounding box và tên từng đối tượng.

Giao diện đơn giản, dễ sử dụng:

* Khung tải ảnh
* Kết quả nhận diện
* Tùy chọn dùng ảnh có sẵn

# **CHƯƠNG 4. ĐÁNH GIÁ, NHẬN XÉT**

## **4.1. Đánh giá**

* **Mô hình được sử dụng**: Roboflow 3.0 – Object Detection (Fast).
* **Checkpoint gốc**: COCOn (huấn luyện từ bộ dữ liệu COCO gốc).
* **Mô hình cho kết quả như sau:**
  + **mAP@50**: 40.5% – mức độ chính xác tương đối, cần cải thiện thêm.
  + **Precision**: 1.0% – độ chính xác khi dự đoán đối tượng còn rất thấp, mô hình đang dự đoán quá nhiều sai.
  + **Recall**: 62.0% – khả năng phát hiện đúng đối tượng ở mức khá, nhưng chưa đủ tốt để ứng dụng thực tế.

## **4.2. Nhận xét**

* Dựa vào kết quả huấn luyện:
  + **Recall cao** nhưng **Precision thấp**, mô hình đang phát hiện được hầu hết đối tượng nhưng không phân biệt chính xác từng loại.
  + **mAP@50 = 40.5%** là khởi đầu chấp nhận được nhưng còn cần cải thiện nhiều.

## **4.3. Hướng cải thiện**

* + **Tăng số lượng ảnh** trong tập huấn luyện, đặc biệt là ảnh chất lượng cao.
  + **Cân bằng dữ liệu** giữa các lớp (tránh lớp quá nhiều hoặc quá ít ảnh).
  + **Augmentation**: sử dụng Roboflow để tạo dữ liệu mới bằng cách xoay, lật, tăng sáng tối...

# **CHƯƠNG 5.KẾT LUẬN**

Trong đồ án này, nhóm đã thực hiện nghiên cứu và triển khai một mô hình nhận diện đối tượng sử dụng YOLOv8, kết hợp với nền tảng Roboflow để xử lý và gán nhãn dữ liệu. Các đối tượng được lựa chọn bao gồm con người, ô tô, xe máy, ghế và đèn đường – đều là những thành phần phổ biến trong môi trường đô thị.

Quá trình thực hiện bao gồm các bước từ thu thập và gán nhãn dữ liệu, huấn luyện mô hình, đánh giá hiệu quả đến triển khai mô hình trong một ứng dụng web demo. Kết quả huấn luyện trên Roboflow cho thấy mô hình đạt được:

mAP@50: 40.5% – thể hiện mức độ chính xác trung bình trong việc phát hiện đúng đối tượng.

Recall: 62% – khả năng nhận diện được đối tượng đúng tương đối ổn.

Precision: 1% – độ chính xác còn thấp, phản ánh việc mô hình nhận diện nhầm khá nhiều.

Kết quả này cho thấy mô hình đã bước đầu học được các đặc trưng của các đối tượng nhưng vẫn cần được cải thiện thêm về hiệu suất, đặc biệt là độ chính xác trong nhận diện.

Việc ứng dụng Roboflow đã giúp đơn giản hóa quy trình xử lý dữ liệu, gán nhãn và huấn luyện mô hình – đặc biệt phù hợp với các dự án nhỏ hoặc sinh viên nghiên cứu. Đồng thời, nhóm cũng đã xây dựng một nền tảng ứng dụng web để kiểm thử mô hình trong thực tế, giúp chứng minh tính ứng dụng cao của mô hình trong các bài toán giám sát, theo dõi đối tượng, hoặc hỗ trợ giao thông thông minh.