**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KHOA HỌC MÁY TÍNH**



**BÀI BÁO CÁO ĐỒ ÁN**

**Môn: CÁC VẤN ĐỀ CHỌN LỌC TRONG THỊ GIÁC MÁY TÍNH**

**Đề tài:**

**DISTANCE ESTIMATION AND OBJECT DETECTION**

**GVHD: Đỗ Văn Tiến**

**Sinh viên thực hiện:**

**Hoàng Kim Ngọc Anh – 21520560**

Tp. Hồ Chí Minh, tháng 12 năm 2024

**MỤC LỤC**

[**Chương 1: MỞ ĐẦU 3**](#_Toc137219695)

[**1.1. Đặt vấn đề 3**](#_Toc137219696)

[**1.2. Mục tiêu 4**](#_Toc137219697)

[**1.3. Đối tượng và phạm vi nghiên cứu 5**](#_Toc137219698)

[**1.4. Nội dung và phương pháp thực hiện 5**](#_Toc137219699)

[**1.4.1. Xác định câu hỏi nghiên cứu 5**](#_Toc137219700)

[**1.4.2. Tạo bảng dữ liệu từ bộ dữ liệu thô 5**](#_Toc137219701)

[**1.4.3. Phân tích dữ liệu 6**](#_Toc137219702)

[**1.4.4. Xây dựng mô hình để dự đoán khoảng cách từ ảnh đến vật thể 6**](#_Toc137219703)

[**1.4.5. Chuẩn hóa dữ liệu và xây dựng các mô hình phân loại để dự đoán lớp của vật thể. 8**](#_Toc137219704)

[**1.4.6. Giải thích kết quả của các mô hình phân loại 8**](#_Toc137219705)

[**1.4.7. Kết quả dự kiến 8**](#_Toc137219706)

[**1.4.8. Kế hoạch thực hiện: 10 tuần 9**](#_Toc137219707)

[**Chương 2: TỔNG QUAN 10**](#_Toc137219708)

[**2.1. Tình hình nghiên cứu trên thế giới 10**](#_Toc137219709)

[**2.2. Tình hình nghiên cứu trong nước 11**](#_Toc137219710)

[**Chương 3: ƯỚC TÍNH KHOẢNG CÁC VÀ NHÃN VẬT THỂ ĐƯỢC CHỤP TRÊN ĐƯỜNG PHỐ 11**](#_Toc137219711)

[**3.1 . Tìm hiểu dữ liệu 11**](#_Toc137219712)

[**3.2. Phương pháp giải quyết bài toán 21**](#_Toc137219713)

[**3.3. Phân loại nhóm vật thể 22**](#_Toc137219714)

[**3.4. Xử lý dữ liệu thô 24**](#_Toc137219715)

[**3.5. Tiền xử lý dữ liệu 27**](#_Toc137219716)

[**3.5.1. Mô tả chi tiết các bước xử lý trong tiền xử lý dữ liệu 30**](#_Toc137219717)

[**3.6. Xây dựng mô hình 30**](#_Toc137219718)

[**3.6.1 . Chia dữ liệu huấn luyện, dữ liệu kiểm tra 30**](#_Toc137219719)

[**3.6.2. Chuẩn hóa dữ liệu 31**](#_Toc137219720)

[**3.6.3. Chuẩn hóa dữ liệu 31**](#_Toc137219721)

[**3.6.4. Huấn luyện các mô hình dự đoán dựa trên các pretrained model 32**](#_Toc137219722)

[**3.6.4.1. YOLOv8n 32**](#_Toc137219723)

[**3.6.4.2. Fast RCNN 33**](#_Toc137219724)

[**3.6.2. Kiểm định mô hình dự đoán 36**](#_Toc137219727)

[**3.6.3. Các bước còn lại 36**](#_Toc137219728)

[**3.7. Dự đoán thực tế 36**](#_Toc137219729)

[**3.8. Kết quả thực nghiệm 37**](#_Toc137219730)

# **Chương 1: MỞ ĐẦU**

## **1.1. Đặt vấn đề**

Trong thời đại công nghệ số hóa, việc phát triển các hệ thống thông minh có khả năng nhận diện và phân tích hình ảnh ngày càng trở nên quan trọng. Một trong những ứng dụng tiêu biểu là việc ước tính khoảng cách từ vật thể đến camera, vốn đóng vai trò cốt lõi trong các lĩnh vực như lái xe tự hành, robot thông minh, thực tế ảo (AR/VR), và giám sát an ninh.

Ước tính khoảng cách chính xác không chỉ giúp tăng cường khả năng nhận thức không gian của các hệ thống mà còn cải thiện độ an toàn và hiệu quả vận hành. Ví dụ, trong hệ thống lái xe tự hành, việc xác định khoảng cách từ xe đến các vật thể như người đi bộ, xe cộ hoặc chướng ngại vật là yếu tố then chốt để ra quyết định điều khiển và tránh va chạm. Tương tự, trong lĩnh vực robot, thông tin khoảng cách cho phép robot lập kế hoạch đường đi và thực hiện các tác vụ với độ chính xác cao.

Mặc dù các phương pháp truyền thống như sử dụng cảm biến LIDAR hoặc SONAR có thể cung cấp thông tin khoảng cách chính xác, chi phí cao và yêu cầu phần cứng phức tạp khiến chúng trở thành lựa chọn không khả thi trong một số ứng dụng thực tiễn. Điều này đã thúc đẩy sự phát triển của các giải pháp dựa trên hình ảnh và các mô hình máy học, tận dụng thông tin từ ảnh RGB để ước tính khoảng cách một cách nhanh chóng và hiệu quả.

Tuy nhiên, việc dự đoán khoảng cách từ hình ảnh gặp phải nhiều thách thức, bao gồm sự đa dạng về môi trường, điều kiện ánh sáng, kích thước và đặc điểm của vật thể. Do đó, bài toán này đòi hỏi các mô hình không chỉ có khả năng xử lý dữ liệu hình ảnh phức tạp mà còn cần tối ưu hóa để đạt được độ chính xác cao trên các tập dữ liệu thực tế.

Trong bối cảnh đó, báo cáo này sẽ tập trung vào việc phát triển một mô hình dự đoán khoảng cách từ vật thể đến camera, sử dụng các kỹ thuật máy học hiện đại. Mục tiêu là đề xuất một giải pháp hiệu quả, có tính tổng quát cao và ứng dụng thực tế. Báo cáo cũng sẽ phân tích các phương pháp tiếp cận, thách thức kỹ thuật, và các kết quả thực nghiệm để làm rõ tính khả thi của giải pháp đề xuất.

## **1.2. Mục tiêu**

* Mục tiêu đề tài: Dựa vào dữ liệu KITTI Benchmark Vision (2015) đề dự đoán khoảng cách từ máy ảnh đến vật thể trên đường phố, đồng thời dự đoán nhãn của vật thể này
* Mục đích nghiên cứu: Phát triển một mô hình máy học dựa trên dữ liệu hình ảnh để dự đoán khoảng cách từ vật thể đến camera một cách chính xác và hiệu quả. Nghiên cứu nhằm tối ưu hóa hiệu suất dự đoán trong các môi trường thực tế, đồng thời cung cấp giải pháp thay thế chi phí thấp cho các công nghệ đo khoảng cách truyền thống như LIDAR hoặc SONAR.

## **1.3. Đối tượng và phạm vi nghiên cứu**

* Dữ liệu đề tài nghiên cứu giới hạn trong tập dữ liệu KITTI Benchmark Vision bao gồm:
  + Tập dữ liệu hình ảnh/ video: dữ liệu của đường phố được chụp từ máy ảnh đặt trên nóc ô tô: Hình ảnh, video có độ phân giải 1242x375 và video 15fps
  + Tập dữ liệu annotation: Dữ liệu có thể được đánh dấu thủ công hoặc lấy từ bộ dữ liệu bao gốm các chỉ số: Loại đối tượng (car, van, truck, pedestrian, tram, etc), độ bị cắt (từ 0: không đến 1: rời khỏi khung hình), mức độ che khuất (0: hoàn toàn thấy đến 3: không rõ), góc quan sát đối tượng (trong khoảng [-π, π], kích thước 3d của đối tượng (cao, rộng, dài), tọa độ bounding box của đối tượng (góc trái trên phải dưới), tọa độ 3d của đối tượng trong hệ trục camera (x,y,z), góc quay quanh trục y của hệ trục camera ([-π, π]), độ tin cậy của dự đoán (giá trị càng cao là tốt hơn).

## **1.4. Nội dung và phương pháp thực hiện**

### 1.4.1. Xác định câu hỏi nghiên cứu

Câu hỏi nghiên cứu là phân tích và so sánh các mô hình máy học nào có khả năng trích xuất đặc trưng hoặc dự đoán chính xác khoảng cách từ vật thể đến camrera hoặc xác định các đặc trưng nào trong dữ liệu ảnh và đánh dấu ảnh hưởng nhiều nhất đến độ chính xác của dự đoán. Các phương pháp để đối phó với điều kiện môi trường ví dụ như ánh sáng, môi trường và vật thể cũng được cân nhắc. Cuối cùng không thể nhắc tới về việc so sánh các phương pháp để tìm ra phương án hiệu quả nhất về chi phí, độ chính xác và khả năng ứng dụng.

### 1.4.2. Tạo bảng dữ liệu từ bộ dữ liệu thô

Chọn ra các thuộc tính cần thiết cho đề tài từ các bảng dữ liệu trong bộ dữ liệu KITTI Benchmark Vision và kết hợp chúng thành một bảng dữ liệu dùng cho đề tài.

Phương pháp thực hiện:

* Pandas dataframe.groupby() được sử dụng để nhóm dữ liệu theo danh mục và áp dụng chức năng cho danh mục nhằm tổng hợp dữ liệu một cách hiệu quả.
* Pandas dataframe.merge() được sử dụng để hợp nhất hai đối tượng DataFrame bằng thao tác nối kiểu cơ sở dữ liệu. Việc nối được thực hiện trên các cột hoặc chỉ mục.

### 1.4.3. Làm sạch dữ liệu

Làm sạch dữ liệu để chuẩn bị cho phân tích. Nội dung này thường liên quan đến việc:

* Xóa dữ liệu trùng lặp và bất thường.
  + Sử dụng pandas dataframe.duplicated().sum() để kiểm tra và thống kê các giá trị trùng lặp trong bảng dữ liệu.
  + Sử dụng pandas dataframe.drop\_duplicates() để xóa các trùng lặp
  + Sử dụng pandas dataframe.isna().sum() để kiểm tra và thống kê các giá trị thiếu trong bảng dữ liệu.
  + Sử dụng pandas dataframe.dropna(how="all") để xóa các hàng dữ liệu thiếu tất cả các cột.
* Điều chỉnh sự không nhất quán:
  + Sử dụng pandas dataframe.replace() để thay thế giá trị được chỉ định bằng một giá trị được chỉ định khác. Phương thức này sẽ tìm kiếm trên toàn bộ Data Frame và thay thế mọi trường hợp của giá trị đã chỉ định.
* Chuẩn hóa cấu trúc và định dạng dữ liệu
  + Sử dụng Label Encoding để chuyển đổi các cột dữ liệu kiểu object thành các cột số để phục vụ cho việc trực quan hoá dữ liệu và các mô hình học máy chỉ lấy dữ liệu số.
* Xử lý khoảng trắng và các lỗi cú pháp khác.

### 1.4.4. Phân tích dữ liệu

Sau khi trực quan hóa dữ liệu ta có phân tích sau:

* Các giá trị trong bảng đều có tỉ lệ với nhau, từ màu nhạt dần biểu thị tỉ lệ thuận đến màu đen biểu thị tỉ lệ nghịch
* Đối với class phân loại là z location (khoảng cách đến camera) thì có thuộc tính tương quan mạnh yếu xuất hiện chẳng hạn như:
* Thuộc tính tọa độ bounding box và góc quan sát đối tượng ảnh hưởng nhiều nhất đến kết quả của bài toán dự đoán.

### 1.4.5. Chuẩn hóa dữ liệu và xây dựng các mô hình phân loại để dự đoán lớp vật thể và khoảng cách.

* Sử dụng sklearn sklearn preprocessing.StandardScaler() để chuẩn hóa dữ liệu.
* Sử dụng các mô hình pretrained và mô hình phân loại như:
* Decision Tree
* Fast RCNN
* YOLOv8n

### 1.4.6. Giải thích kết quả của các mô hình phân loại

Sử dụng các độ đo đánh giá để đánh giá kết quả của các mô hình dự đoán:

* Precision cho biết tỉ lệ các đối tượng dự đoán đúng so với tất cả các đối tượng mà mô hình dự đoán có mặt.
* Recall đo lường khả năng của mô hình trong việc phát hiện tất cả các đối tượng có mặt trong ảnh.
* mAP50 đánh giá độ chính xác trung bình của mô hình tại mức IoU = 0.5.
* mAP50-95 đánh giá độ chính xác trung bình ở nhiều mức IoU từ 0.5 đến 0.95, giúp đánh giá toàn diện hiệu suất của mô hình.

### 1.4.7. Kết quả dự kiến

* Phát triển một mô hình máy học có khả năng dự đoán khoảng cách từ vật thể đến camera với độ chính xác cao, đáp ứng yêu cầu thực tế.
* Xác định và đánh giá tầm quan trọng của các đặc trưng đầu vào như bounding box, vị trí 3D, góc quay, v.v., đối với kết quả dự đoán.
* Mô hình hoạt động tốt trên nhiều loại vật thể và trong các điều kiện khác nhau như thay đổi ánh sáng hoặc môi trường.
* Chứng minh mô hình đề xuất là giải pháp thay thế chi phí thấp, dễ triển khai, và đủ chính xác so với các công nghệ đo khoảng cách truyền thống.
* Tích hợp mô hình vào các ứng dụng thực tế như lái xe tự hành hoặc robot để kiểm tra tính khả thi.

### 1.4.8. Kế hoạch thực hiện: 10 tuần

|  |  |
| --- | --- |
| Thời gian | Công việc |
| 18/10/2024 | Nhận thông tin về đồ án môn học |
| 25/10/2024 | Chọn đề tài, tìm hiểu dữ liệu và phân tích vấn đề |
| 01/11/2024 | Tìm hiểu dữ liệu và phân tích vấn đề |
| 08/11/2024 | Tìm hiểu về thuật toán sử dụng, trường dữ liệu và sử dụng dữ liệu mẫu |
| 17/11/2024 | Thay đổi mục tiêu đề tài, tối ưu hóa và rút gọn dữ liệu |
| 25/11/2024 | Finetune mô hình, lựa chọn mô hình cuối cùng cho từng module |
| 03/12/2024 | Xây dựng app và kiểm thử mô hình |
| 15/05/2024 | Lên kế hoạch cuối cho báo cáo đồ án |
| 20/12/2024 | Chuẩn bị nội dung file word về báo cáo đồ án |

# **Chương 2: TỔNG QUAN**

## **2.1. Tình hình nghiên cứu trên thế giới**

Trong những năm gần đây, dự đoán khoảng cách từ vật thể đến camera đã trở thành một chủ đề nghiên cứu quan trọng trong lĩnh vực thị giác máy tính và học sâu. Dưới đây là một số xu hướng và tiến bộ chính:

- Các công nghệ truyền thống như LIDAR, RADAR và SONAR đã được sử dụng rộng rãi để đo khoảng cách với độ chính xác cao. Tuy nhiên, chúng thường đắt đỏ và yêu cầu phần cứng phức tạp, không phù hợp cho các ứng dụng chi phí thấp.

- Các nghiên cứu gần đây đã chuyển hướng sang sử dụng ảnh RGB và các thuật toán học máy hoặc học sâu để ước tính khoảng cách. Ví dụ: Các mô hình học sâu như Monocular Depth Estimation sử dụng ảnh đơn để dự đoán bản đồ độ sâu (depth map) hoặc là kết hợp các mô hình phát hiện đối tượng (YOLO, Faster R-CNN) với thuật toán dự đoán khoảng cách.

- Sử dụng mạng neural sâu (Deep Neural Networks) như Convolutional Neural Networks (CNNs) và Transformer đã cải thiện đáng kể độ chính xác dự đoán khoảng cách từ hình ảnh.

- Nhiều nghiên cứu đã tập trung vào việc tăng cường khả năng của các mô hình trong các môi trường phức tạp, bao gồm điều kiện ánh sáng yếu, nhiều đối tượng chồng chéo, hoặc góc quay camera thay đổi.

## **2.2. Tình hình nghiên cứu trong nước**

Tại Việt Nam, lĩnh vực dự đoán khoảng cách từ vật thể đến camera sử dụng dữ liệu hình ảnh vẫn đang trong giai đoạn phát triển, nhưng đã ghi nhận một số tiến bộ đáng chú ý nhờ sự quan tâm ngày càng tăng đối với các công nghệ liên quan đến thị giác máy tính và học sâu.

# **Chương 3: ƯỚC TÍNH KHOẢNG CÁC VÀ NHÃN VẬT THỂ ĐƯỢC CHỤP TRÊN ĐƯỜNG PHỐ**

## **3.1 . Tìm hiểu dữ liệu**

Dữ liệu sử dụng cho bài toán dự đoán KITTI Benchmark Vision

Dữ liệu dùng cho bài toán này bao gồm 1 video dài 8 phút 32 giây với 15 fps có độ phân giải 1242x374. Các dữ liệu này bao gồm cả file annotation đánh dấu cho các vật thể có trong bộ dữ liệu.

Chi tiết bảng annotation:

|  |  |  |  |
| --- | --- | --- | --- |
| Data Field | Description | Data Type | Ghi chú |
| type | Phân biệt các loại vật thể có trong dữ liệu | object | Có các giá trị: Car, Van, Truck. Pedestrian, Cyclist, Misc,… |
| truncated | Độ bị cắt khỏi khung hình | float | Có giá trị trong khoảng:  0 – không  1 – rời khỏi khung hình |
| occludded | Mức độ bị che khuất khỏi tầm nhìn | integer | Có giá trị trong khoảng:  0 – hoàn toàn thấy  3 – không rõ |
| alpha | Góc quan sát của đối tượng | float | Có giá trị trong khoảng: [-π, π]. |
| bbox | Tọa độ bounding box bao quanh vật thể | float | Được chia ra làm 4 đặc trưng: xmin, ymin, xmax, ymax |
| dimension | Kích thước 3D của đối tượng | float | Được chia ra làm 3 đặc trưng: cao, rộng, dài |
| location | Tọa độ 3D của đối tượng trong hệ trục camera | float | Được chia ra làm 3 đặc trưng: x,y,z |
| rotation\_y | Góc quay quanh trục Y trong hệ trục camera | Float | Có giá trị trong khoảng: [-π, π]. |
| score | Độ tin cậy của dự đoán | float | Càng cao càng tốt |

## **3.2. Phương pháp giải quyết bài toán**

Nhóm thực hiện giải quyết bài toán “Dự đoán khoảng cách và nhãn của vật thể từ ảnh chụp của camera” theo hướng data mining và computer vision, cụ thể ở đây là áp dụng kỹ thuật feature extraction để lấy được đặc trưng các vật thể trong từng khung hình. Và các cột thuộc tính trong dữ liệu sử dụng cho bài toán dự đoán chính là các cột thông tin bounding box và góc độ chụp của các vật thể như xe ô tô, xe tải, người đi bộ,…. Em sẽ tiến hành lập trình các giải thuật xử lý dữ liệu trong lĩnh vực khai phá dữ liệu để biến các bảng dữ liệu thô nêu trên thành một bảng dữ liệu (data) theo ý muốn. Trong bảng (data) này sẽ có một cột dữ liệu là cột nhãn (‘zloc’) để dự đoán, hay có thể gọi là cột dự đoán khoảng cách trong bài toán. Các cột còn lại sẽ được chọn từ các bảng dữ liệu thô để làm các thuộc tính huấn luyện cho mô hình dự đoán.

## **3.3. Phân loại nhóm vật thể**

Tính toán số lượng vật thể có dữ liệu trên tổng số vật thể của toàn bộ video. Chi tiết số lượng loại vật thể như sau:

|  |  |
| --- | --- |
| **Loại vật thể** | **Số lượng** |
| Car | 5526 |
| Pedestrian | 1045 |
| Van | 568 |
| Cyclist | 329 |
| Truck | 193 |
| Misc (Khác) | 180 |
| Tổng số | 7481 |

Giá trị ‘Khác’ biểu thị hai ý nghĩa: Những vật thể không đáng nhắc tới hoặc không phù hợp huấn luyện hoặc bị nhầm nhãn

Để đơn giản hoá đề tài và tăng độ chính xác cho các mô hình nhóm thực hiện bài toán dự đoán khả năng tốt nghiệp với hai module: dự đoán khoảng cách và dự đoán nhãn vật thể.

## **3.4. Xử lý dữ liệu thô**

Trong phần này, nhóm sẽ trình bày chi tiết phần xử lý dữ liệu thô. Mục tiêu của xử lý dữ liệu thô là tạo ra bảng dữ liệu chính cho đề tài bao gồm các cột thuộc tính và một cột nhãn.

Để áp dụng phân lớp máy học vào bảng dữ liệu này thì mỗi dòng trong bảng chính là thông tin của 1 vật thể trong 1 khung hình và mỗi cột chính là các thuộc tính liên quan đến vật thể đó như: nhãn, khoảng cách, bounding box,... Trong đó cột cuối cùng của bảng chính là cột zloc – giá trị khoảng cách từ camera đến vật thể. Danh sách các bảng dữ liệu dùng cho đề tài đã được nhóm giới thiệu ở những phần trước. Bây giờ việc cần làm là xác định các cột thuộc tính nào sẽ có mặt trong bảng dữ liệu.

### 3.5.1. Mô tả các bước xử lý trong tiền xử lý dữ liệu

* Khám phá dữ liệu:
  + Hiểu cấu trúc dữ liệu: số lượng cột, dòng, loại dữ liệu (số, chuỗi, danh mục).
  + Kiểm tra giá trị thiếu (missing values) và giá trị ngoại lệ (outliers).
  + Xác định các cột không cần thiết hoặc trùng lặp.
  + Kiểm tra phân phối dữ liệu và mối tương quan giữa các biến.
* Xử lý các giá trị bị thiếu:
  + Loại bỏ các hàng hoặc cột chứa quá nhiều giá trị thiếu.
  + Điền giá trị thay thế (imputation)
* Xử lý các giá trị ngoại lệ:
  + Phát hiện giá trị ngoại lệ bằng các phương pháp như Z-score, IQR (Interquartile Range).
  + Loại bỏ hoặc điều chỉnh giá trị ngoại lệ.

## **3.6. Xây dựng mô hình**

### 3.6.1 . Chia dữ liệu huấn luyện, dữ liệu kiểm tra

Trong bảng dữ liệu (Data) ta có 7841 dòng dữ liệu tương ứng 7841 vật thể. Nhóm sẽ dự đoán khoảng cách và nhãn của vật dựa trên đặc trưng trích xuất ra từ ảnh và các dữ liệu bounding box + góc độ máy ảnh trong bộ dữ liệu.

Dữ liệu kiểm tra sẽ là một tập dữ liệu khác tương tự được chia ra bao gồm cùng số lượng loại nhãn vật thể và các khoảng cách khác nhau

### 3.6.2. Chuẩn hóa dữ liệu

Đối với dữ liệu huấn luyện thì:

* Các cột thuộc tính chứa giá trị rời rạc được mã hóa theo phương pháp Label Encoder
* Các cột thuộc tính chứa giá trị số thì được giữ nguyên
* Các nhãn sẽ được đặt tên dễ hiểu, sắp xếp đúng thứ tự và đồng nhất.
* Các dữ liệu thông qua kĩ feature extraction sẽ được lưu giữ vả sử dụng tách biệt

Đối với dữ liệu kiểm tra thì:

* Các cột thuộc tính chứa giá trị rời rạc được mã hóa theo phương pháp Label Encoder
* Các cột thuộc tính chứa giá trị số thì được giữ nguyên
* Các nhãn sẽ được đặt tên dễ hiểu, sắp xếp đúng thứ tự và đồng nhất.

### 3.6.3. Chuẩn hóa dữ liệu

Với cả dữ liệu huấn luyện và dữ liệu kiểm tra thì:

* Các cột thuộc tính về bounding box được nhóm chuẩn hóa theo phương pháp Standard Scaler:
  + Phương thức chuẩn hóa bằng cách loại bỏ giá trị trung bình và cân bằng tỉ lệ theo tỉ lệ phương sai.
  + Điểm chuẩn hóa của một mẫu x được tính bằng công thức:

Z = (x-u)/s

Trong đó:

* u là trung bình của các mẫu huấn luyện hoặc bằng 0 nếu with\_mean = False
* s là độ lệch chuẩn của các mẫu huấn luyện hoặc bằng 1 nếu with\_std=False

Cân bằng và tỉ lệ xảy ra riêng biệt với mỗi phương thức bằng cách tính toán số liệu thống kê có liên quan trong các mẫu của bảng huấn luyện. Giá trị trung bình và độ lệch chuẩn được lưu lại và tái sử dụng khi hàm transform được gọi lên.

Chuẩn hóa dữ liệu cho một bảng dữ liệu là điều cần thiết cho nhiều mô hình máy học: Những mô hình này vận hành không được tốt nếu những phương thức riêng lẻ này không được chú trọng đến để điều chỉnh những dữ liệu quan trọng.

Ví dụ với những yếu tố sử dụng trong hàm chủ động trong thuật toán (ví dụ như là RBF kernel của Support Vector Machine hay là bộ điều chỉnh L1 và L2 của mô hình tiệm cận), những thuật toán này yêu cầu tất cả các tính năng cân bằng xung quanh giá trị 0 và sắp xếp các phương sai trong cùng một trình tự. Nếu một tính năng có một phương sao mà thứ tự của nó có kích cỡ lớn hơn các phương sai khác, điều này dẫn tới sự áp đảo trong hàm chủ động và làm bộ tính toán không thể học từ các tính năng khác một cách chính xác.

Bộ điều chỉnh tỉ lệ này có thể áp dụng để giảm bớt ma trận CSR hoặc CSC thông qua gọi lên with\_mean = False để tránh làm hỏng cấu trúc thưa thớt trong dữ liệu.

### 3.6.4. Huấn luyện các mô hình dự đoán dựa trên các pretrained model

#### **3.6.4.1. YOLOv8n**

Mô hình Yolov8n là phiên bản nhẹ nhất trong dòng YOLOv8, một mô hình học sâu tiên tiến dành cho các nhiệm vụ phát hiện đối tượng (Object Detection), phân đoạn ảnh (Image Segmentation), và ước tính điểm chính (Keypoint Estimation). Được phát triển bởi Ultralytics, YOLOv8 được thiết kế để kế thừa những cải tiến từ các phiên bản YOLO trước đây (YOLOv4, YOLOv5, YOLOv7) và tối ưu hóa cả về hiệu năng lẫn tốc độ xử lý.

Quy trình xây dựng mô hình Yolov8n bao gồm các bước sau:

1. Chuẩn bị dữ liệu: Dữ liệu huấn luyện được sắp xếp thành các mẫu với nhãn đã biết. Mỗi mẫu gồm các đặc trưng đầu vào.
2. Định dạng chú thích (YOLO format): Mỗi tệp chú thích chứa các dòng theo cấu trúc <class\_id> <x\_center> <y\_center> <width> <height>
3. Chia tập dữ liệu: Chia thành 3 tập: huấn luyện (train), kiểm tra (test), và xác thực (validation): 70%-20%-10%.
4. Tiền xử lý dữ liệu: Xử lý hình ảnh: thay đổi kích thước. Đảm bảo dữ liệu không chứa lỗi (ảnh hỏng, chú thích sai).
5. Huấn luyện mô hình: Tinh chỉnh tham số, huấn luyện và theo dõi. Lưu lại model có kết quả tốt nhất để dùng làm dự đoán

Mô hình Yolov8n có các ưu điểm như kích thước nhỏ gọn: YOLOv8n là mô hình "Nano", tức là phiên bản nhẹ nhất và nhanh nhất, được thiết kế để hoạt động tốt trên các thiết bị có tài nguyên hạn chế như CPU, GPU nhỏ hoặc thiết bị nhúng; Nhờ kiến trúc đơn giản và số lượng tham số ít, YOLOv8n đạt tốc độ xử lý cao, lý tưởng cho các ứng dụng thời gian thực; Mặc dù kích thước nhỏ, YOLOv8n vẫn đảm bảo hiệu năng tốt trong các bài toán phát hiện đối tượng cơ bản.

#### **3.6.4.2. Fast RCNN**

Fast R-CNN là một mô hình học sâu được phát triển để cải thiện hiệu suất và tốc độ của các phương pháp phát hiện đối tượng trước đó, chẳng hạn như R-CNN (Region-based Convolutional Neural Networks). Fast R-CNN được đề xuất bởi Ross B. Girshick vào năm 2015 và là một trong những bước đột phá quan trọng trong lĩnh vực phát hiện đối tượng (object detection).

Mô hình Fast R-CNN có các đặc điểm quan trọng sau:

1. Sử dụng CNN toàn cục (Global CNN): Trong Fast R-CNN, ảnh đầu vào được đưa qua một mạng CNN duy nhất (chẳng hạn như AlexNet) để trích xuất các đặc trưng đặc biệt cho toàn bộ ảnh (tạo ra một bản đồ đặc trưng). Điều này giúp tiết kiệm thời gian so với việc tính toán lại đặc trưng cho từng vùng đề xuất như trong R-CNN.
2. Region of Interest (RoI) Pooling: Fast R-CNN sử dụng kỹ thuật RoI Pooling (Region of Interest Pooling) để trích xuất đặc trưng cho từng vùng đề xuất từ bản đồ đặc trưng toàn cục. Điều này giúp đưa các vùng đề xuất có kích thước khác nhau về cùng một kích thước cố định, phù hợp với các lớp Fully Connected trong mạng.
3. Học đồng thời (End-to-end training): Toàn bộ mô hình Fast R-CNN có thể được huấn luyện đồng thời bằng cách sử dụng một hàm mất mát chung, thay vì phải huấn luyện từng phần riêng biệt như R-CNN.
4. Ba nhiệm vụ trong một mô hình: Fast R-CNN không chỉ dự đoán lớp của đối tượng mà còn dự đoán chính xác các tọa độ của bounding box (hình chữ nhật bao quanh đối tượng), giúp tăng độ chính xác trong việc phát hiện và phân loại.

**3.6.4.3. VGG16**

VGG16 (Visual Geometry Group 16) là một kiến trúc mạng nơ-ron sâu (deep neural network) được đề xuất bởi nhóm nghiên cứu tại Đại học Oxford, trong bài báo "Very Deep Convolutional Networks for Large-Scale Image Recognition" (2014). VGG16 là một trong những mô hình thành công nhất trong việc phân loại hình ảnh và đã đạt được hiệu suất cao trong các cuộc thi như ImageNet.

- Đặc điểm của VGG16:

- Đơn giản nhưng hiệu quả: Một trong những điểm mạnh của VGG16 là kiến trúc tương đối đơn giản, với chỉ các lớp convolution 3x3 và lớp pooling 2x2, nhưng vẫn rất mạnh mẽ trong việc học các đặc trưng của hình ảnh.

- Ứng dụng rộng rãi: VGG16 đã được sử dụng trong nhiều ứng dụng khác nhau ngoài phân loại hình ảnh, bao gồm nhận diện đối tượng, nhận dạng khuôn mặt, và học chuyển giao (transfer learning).

- Transfer Learning: VGG16 đã được huấn luyện trên một tập dữ liệu lớn như ImageNet, do đó, mô hình có thể học được nhiều đặc trưng hữu ích cho các tác vụ khác nhau. Việc sử dụng VGG16 đã được áp dụng rộng rãi trong transfer learning, nơi các trọng số của mô hình được sử dụng lại và chỉ cần huấn luyện lại các lớp cuối cùng để áp dụng vào các bài toán mới.

### 3.6.2. Kiểm định mô hình dự đoán

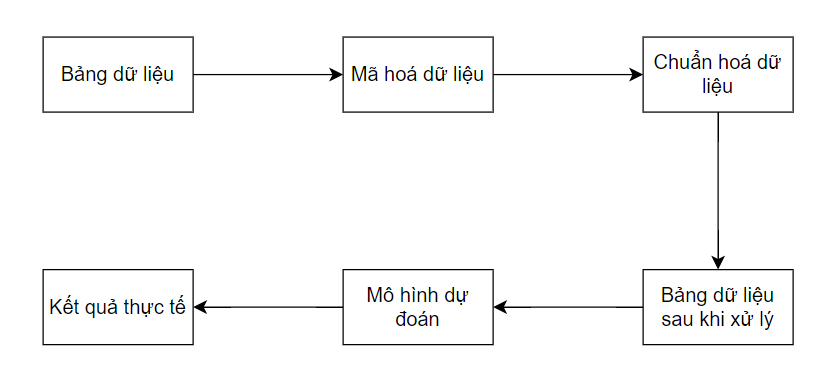
Tại đây, các độ đo: precision, recall, mAP50 và mAP50-95 được nhóm sử dụng để kiểm định chất lượng của mô hình dự đoán trong các trường hợp thử nghiệm.

### 3.6.3. Các bước còn lại

Nhóm sẽ thực hiện chọn mô hình có kết quả thực nghiệm tốt nhất cho từng test và tiến hành so sánh giữa các test với nhau để chọn ra những thuộc tính quan trọng nhất cho bài toán dự đoán khoảng cách giữa máy ảnh và vật thể trên đường phố.

## **3.7. Dự đoán thực tế**

Hiện tại chúng ta sẽ có bảng dữ liệu chính với các cột thuộc tính và một cột nhãn. Quy trình cần thực hiện là:



Hình 1. Quy trình để dự đoán thực tế

## **3.8. Kết quả thực nghiệm**

Trong phần này nhóm sẽ trình bày kết quả kiểm định mô hình dự đoán. Để có được kết quả kiểm định mô hình, nhóm đã lập trình bằng ngôn ngữ Python với các framework như: pytorch, opencv, pandas, numpy, sklearn, …

Bộ dữ liệu chính đang có thông tin đầy đủ của các vật thể trong bộ dữ liệu KITTI Vision Benchmark. Bài toán dự đoán khoảng cách của vật thể đến máy ảnh sẽ có ý nghĩa cao hơn khi ta dự đoán thông qua 1 khung hình, cụ thể là một khung hình được cắt ra từ một video được lưu từ nguồn hoặc stream.

Vì vậy nhóm xin thử nghiệm mô hình với trường hợp thử nghiệm như sau:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Yolov8 | | | | Faster R-CNN | | | |
| Precision | Recall | mAP50 | mAP50-95 | Precision | Recall | mAP50 | mAP50-95 |
| Car | 0.923 | 0.902 | 0.958 | 0.774 | 0.940 | 0.902 | 0.942 | 0.823 |
| Pedestrian | 0.883 | 0.641 | 0.787 | 0.451 | 0.873 | 0.732 | 0.871 | 0.726 |
| Van | 0.929 | 0.852 | 0.926 | 0.723 | 0.926 | 0.868 | 0.933 | 0.841 |
| Cyclist | 0.861 | 0.710 | 0.801 | 0.518 | 0.834 | 0.774 | 0.842 | 0.759 |
| Truck | 0.948 | 0.931 | 0.975 | 0.813 | 0.943 | 0.926 | 0.921 | 0.782 |
| Misc | 0.880 | 0.792 | 0.877 | 0.612 | 0.861 | 0.832 | 0.784 | 0.721 |

Ta có thể nhận thấy, với một số chỉ số như Precision, Recall thì 2 mô hình được đánh giá xấp xỉ nhau, tuy nhiên Faster R-CNN vẫn nhỉnh hơn trên 2 độ đo mAP50 và mAP50-95.

Tuy nhiên với yêu cầu ứng dụng trong thời gian thực để xử lý ảnh trực tiếp, nhóm sẽ ưu tiên chọn mô hình Yolov8n để áp dụng vào thực nghiệm.

* Đánh giá chung:
* Ưu điểm: Mô hình YOLOv8 cho kết quả rất tốt với các lớp "Car", "Truck" và "Van", với Precision, Recall, mAP50 và mAP50-95 đều ở mức cao. Mô hình có thể phát hiện chính xác và ít gặp lỗi sai đối với các lớp này.
* Nhược điểm: Mô hình gặp khó khăn trong việc phát hiện Pedestrian và Cyclist, với Recall, mAP50 và mAP50-95 thấp hơn so với các lớp khác. Điều này có thể yêu cầu cải thiện mô hình để tăng khả năng phát hiện các đối tượng này, đặc biệt trong môi trường có nhiều người đi bộ hoặc xe đạp.

Trong tổng thể, mô hình hoạt động rất tốt trong các tình huống phổ biến như phát hiện xe (Car, Van, Truck) nhưng cần cải thiện hiệu suất với các đối tượng khó phát hiện hơn như Pedestrian và Cyclist.

* Kết luận: Mô hình YOLOv8 đã cho kết quả ấn tượng trong việc phát hiện đối tượng, đặc biệt là với các lớp "Car", "Van", và "Truck", với độ chính xác cao và khả năng phát hiện tốt (Precision, Recall, mAP50 và mAP50-95 đều đạt giá trị cao). Điều này cho thấy mô hình này có thể hoạt động hiệu quả trong các ứng dụng yêu cầu phát hiện các phương tiện giao thông hoặc các đối tượng có kích thước lớn. Tuy nhiên, đối với các lớp "Pedestrian" và "Cyclist", mô hình gặp phải một số vấn đề khi độ chính xác (Recall, mAP50, mAP50-95) còn thấp. Điều này có thể gây ảnh hưởng đến hiệu quả của mô hình trong các ứng dụng đòi hỏi khả năng phát hiện các đối tượng nhỏ hơn hoặc những đối tượng có sự chuyển động nhanh như người đi bộ và xe đạp.
* Phương hướng phát triển:
  + Tăng cường và bổ sung dữ liệu
  + Fine tuning
  + Cải thiện các lớp phụ trợ (Auxiliary Networks)
  + Sử dụng các mô hình kết hợp (Ensemble Models)