**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN**

**HỌC PHẦN: XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH**

**Đề tài số 06: Xây dựng hệ thống nhận diện chữ số viết tay**

**Giảng viên hướng dẫn: Lương Thị Hồng Lan**

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Mã sinh viên** | **Sinh viên thực hiện** | **Lớp hành chính** |
| **1** | 20211672 | Hà Văn Quý | DCCNTT12.10.6 |
| **2** | 20211480 | Trần Văn Quý | DCCNTT12.10.6 |

**Bắc Ninh, năm 2024**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BÀI TẬP LỚN**

**HỌC PHẦN: XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH**

**Đề tài số 06: Xây dựng hệ thống nhận diện chữ số viết tay**

**Giảng viên hướng dẫn: Lương Thị Hồng Lan**

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Mã sinh viên** | **Sinh viên thực hiện** | **Lớp hành chính** |
| **1** | 20211672 | Hà Văn Quý | DCCNTT12.10.6 |
| **2** | 20211480 | Trần Văn Quý | DCCNTT12.10.6 |

|  |  |
| --- | --- |
|  |  |

**Bắc Ninh, năm 2024**

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á  **KHOA CÔNG NGHỆ THÔNG TIN** | **KỲ THI KẾT THÚC HỌC PHẦN**  **HỌC KỲ 1, NĂM HỌC 2024** – **2025** |

|  |  |  |
| --- | --- | --- |
| **PHIẾU CHẤM THI BÀI TẬP LỚN KẾT THÚC HỌC PHẦN**  **Mã đề thi: 06**  **Tên học phần: Xử lý ảnh và thị giác máy tính**  **Lớp tín chỉ: XATGMT.03.K12.06.LH.C04.1\_LT**   |  | | --- | |  | | |
| **Cán bộ chấm thi 1**  **Lương Thị Hồng Lan** | **Cán bộ chấm thi 2** |

| **TT** | **TIÊU CHÍ** | **THANG ĐIỂM** | **Hà Văn Quý** | **Trần Văn Quý** | |
| --- | --- | --- | --- | --- | --- |
| 20211672 | 20211480 | |
| **1** | **Nội dung báo cáo trên Word đầy đủ** | **3.5** |  |  | |
| 1.1 | Có bố cục rõ ràng (mục lục, phần mở đầu, nội dung chính, kết luận). | 0,5 |  |  | |
| 1.2 | Nội dung phân tích rõ ràng, logic. | 0,5 |  |  | |
| 1.3 | Có dẫn chứng, số liệu minh họa đầy đủ. | 0,5 |  |  | |
| 1.4 | Ngôn ngữ và trình bày chuẩn, không lỗi chính tả. | 0,5 |  |  | |
| 1.5 | Có trích dẫn tài liệu tham khảo đúng quy cách. | 0,5 |  |  | |
| 1.6 | Được trình bày chuyên nghiệp (canh lề, font chữ, khoảng cách dòng hợp lý). | 0,5 |  |  | |
| 1.7 | Tài liệu đầy đủ, bám sát yêu cầu của đề bài. | 0,5 |  |  | |
| **2** | **Nội dung thuyết trình đầy đủ** | **1.0** |  |  | |
| 2.1 | Trình bày tự tin, phát âm rõ ràng, mạch lạc. | 0,5 |  |  | |
| 2.2 | Nội dung thuyết trình đúng trọng tâm, không lan man. | 0,5 |  |  | |
| **3** | **Slides báo cáo đầy đủ nội dung + Hỏi đáp** | **3.0** |  |  | |
| 3.1 | Slides có bố cục rõ ràng (mở đầu, nội dung, kết luận). | 0,5 |  |  | |
| 3.2 | Thiết kế slides đẹp, chuyên nghiệp (màu sắc, hình ảnh minh họa). | 0,5 |  |  | |
| 3.3 | Nội dung trên slides ngắn gọn, dễ hiểu, súc tích. | 0,5 |  |  | |
| 3.4 | Nội dung slides phù hợp với nội dung báo cáo. | 0,5 |  |  | |
| 3.5 | Trả lời câu hỏi đầy đủ, chính xác. | 0,5 |  |  | |
| 3.6 | Trả lời câu hỏi tự tin, thuyết phục. | 0,5 |  |  | |
| **4** | **Code đầy đủ** | **2.5** |  |  | |
| 1.1 | Code được trình bày rõ ràng, có chú thích đầy đủ. | 0,5 |  |  | |
| 1.2 | Code chạy đúng, không lỗi. | 0,5 |  |  | |
| 1.3 | Code tối ưu, không dư thừa. | 0,5 |  |  | |
| 1.4 | Đáp ứng đầy đủ các yêu cầu chức năng theo đề bài. | 0,5 |  |  | |
| 1.5 | Có tính sáng tạo hoặc cải thiện so với yêu cầu. | 0,5 |  |  | |
| **TỔNG ĐIỂM BẰNG SỐ:** | | **10** |  |  | |
| **TỔNG ĐIỂM BẰNG CHỮ:** | | *Mười tròn* |  |  |  |

**Mục lục**

Contents

[CHƯƠNG 1: CƠ SỞ LÝ THUYẾT 7](#_Toc184567345)

[1.1. Biểu diễn ảnh trong máy tính 7](#_Toc184567346)

[1.1.1 Thu nhận ảnh 7](#_Toc184567347)

[1.1.2 Các hệ màu trong máy tính 10](#_Toc184567348)

[1.1.3 Một số loại ảnh thông dụng 12](#_Toc184567349)

[1.2. Tổng quan về học máy 15](#_Toc184567350)

[1.2.1 Học có giám sát 15](#_Toc184567351)

[*1.2.1.1 KNN* 16](#_Toc184567352)

[*1.2.1.2 SVM* 18](#_Toc184567353)

[1.2.2 Học không giám sát 20](#_Toc184567354)

[*1.2.2.1 K-Means* 21](#_Toc184567355)

[*1.2.2.2 Fuzzy C- Means* 22](#_Toc184567356)

[1.2.3 Học bán giám sát 22](#_Toc184567357)

[CHƯƠNG 2 : XÂY DỰNG HỆ THỐNG NHẬN DIỆN CHỮ SỐ TỰ ĐỘNG VỚI OPEN CV 25](#_Toc184567358)

[2.1. Thư viện 25](#_Toc184567359)

[2.1.1 Numpy 25](#_Toc184567360)

[2.1.2 OpenCV 28](#_Toc184567361)

[2.1.3 Thư viện CSV 31](#_Toc184567362)

[2.1.4 Tập dữ liệu MNIST 32](#_Toc184567363)

[2.2. Xây dựng hệ thống nhận diện chữ số tự động 33](#_Toc184567365)

[2.2.1 Hướng giải quyết 33](#_Toc184567366)

[2.2.2 Phát hiện và tách chữ số 34](#_Toc184567367)

[2.2.3 Căn chỉnh chữ số 38](#_Toc184567368)

[2.2.4 Phân đoạn và nhận diện ký tự 38](#_Toc184567369)

[CHƯƠNG 3 : THỰC NGHIỆM 44](#_Toc184567370)

[3.1. Dữ liệu 44](#_Toc184567371)

[3.2. Độ đo đánh giá 45](#_Toc184567372)

[3.3. Kết quả thực nghiệm 47](#_Toc184567374)

[KẾT LUẬN 48](#_Toc184567375)

[TÀI LIỆU THAM KHẢO 49](#_Toc184567376)

**Mở đầu**

Trong thời đại công nghệ số hiện nay, trí tuệ nhân tạo (AI) và học máy (Machine Learning) ngày càng đóng vai trò quan trọng trong nhiều lĩnh vực, từ y tế, giáo dục đến thương mại và công nghiệp. Một trong những ứng dụng nổi bật của AI là nhận dạng chữ viết tay, đặc biệt là nhận dạng chữ số. Đây không chỉ là một bài toán cơ bản trong lĩnh vực học máy mà còn là nền tảng để phát triển các hệ thống hiện đại như máy quét văn bản, kiểm tra chữ ký, và thậm chí cả giao dịch ngân hàng tự động.

Bài luận này nhằm nghiên cứu và phân tích bài toán nhận dạng chữ số viết tay, từ các phương pháp truyền thống đến những kỹ thuật hiện đại như mạng nơ-ron tích chập (CNN). Qua đó, chúng tôi không chỉ trình bày cách tiếp cận và thuật toán mà còn đánh giá hiệu quả của chúng trên tập dữ liệu tiêu chuẩn, đồng thời thảo luận về những thách thức cũng như tiềm năng phát triển của lĩnh vực này trong tương lai.

Hãy cùng khám phá một cách nhìn mới về cách công nghệ có thể hiểu và học từ chính những nét chữ viết tay của con người!

# CHƯƠNG 1: CƠ SỞ LÝ THUYẾT

**1.1. Biểu diễn ảnh trong máy tính**

**1.1.1 Thu nhận ảnh**

* Ảnh số

Ảnh số (digital image) có thể được xem là một biểu diễn dữ liệu rời rạc thể hiện thông tin về không gian và cường độ màu). Ảnh số gồm một tập hợp hữu hạn các phần tử được biểu diễn bởi giá trị số. Ảnh số có thể được biểu diễn dưới dạng ma trận hai chiều, mỗi phần tử của ảnh số gọi là điểm ảnh (pixel)[1]. Tùy thuộc vào độ phân giải là cố định hay biến đổi mà điểm ảnh có thể được biểu diễn dưới dạng vector hoặc dạng bitmap. Ảnh số được xác định theo mảng hai chiều biểu diễn cường độ sáng của điểm ảnh với giá trị cố định, cũng có thể được xác định theo hàm hai chiều f(x, y), trong đó x và y là các tọa độ trong không gian và độ lớn (amplitude) của hàm f được gọi là độ sáng (intensity) hay độ xám (gray level) của ảnh tại điểm đó.

Ảnh rời rạc hai chiều, I(m,n) biểu diễn thông tin thu được từ cảm biến của một chuỗi các vị trí cố định (m = 1, 2, ... , M; n= 1, 2, ... , N) trong tọa độ Cartesian hai chiều được biến đổi từ tín hiệu liên tục không gian 2 chiều thông qua quá trình xử lý tần số liên tục sang miền rời rạc.

* Điểm ảnh

Thuật ngữ điểm ảnh được dịch ra từ thuật ngữ gốc là pixel (viết tắt cụm từ picture element) nghĩa là một phần tử ảnh[1]. Phần tử ảnh được xác định theo toạ độ (x, y) tương ứng với số thứ tự cột và hàng trong ảnh. Giá trị mỗi phần tử ảnh được xác định bởi giá trị cường độ mức xám hoặc màu nhất định. Kích thước và khoảng cách giữa các điểm ảnh được biểu diễn thích hợp sao cho mắt người cảm nhận sự liên tục về không gian và mức xám (màu) của ảnh số gần với như hình ảnh của nó trong không gian thật. Số điểm ảnh trên mỗi diện tích biểu diễn xác định độ phân giải của ảnh số. Ảnh có độ phân giải càng cao thì càng thể hiện rõ nét các đặc điểm của hình ảnh, càng làm cho hình ảnh trở nên gần với thực tế thực và sắc nét hơn,...

* Các dạng ảnh

Nội dung thông tin của các điểm ảnh có thể được xem xét dưới nhiều khía cạnh khác nhau tùy thuộc vào dạng của ảnh. Ví dụ ảnh màu (colour image), ảnh đa mức xám (grey image), ảnh nhị phân (binary image), ảnh hồng ngoại (infrared image),...

* Ảnh màu: Ảnh màu thường là các ảnh chứa thông tin về đối tượng được biểu diễn dưới dạng màu sắc mà mắt thường có thể quan sát được. Mỗi điểm ảnh có cấu trúc gồm nhiều kênh màu khác nhau, thông thường trong máy tính, nó biểu diễn ba lớp màu cơ bản RGB, gồm màu đỏ (red), xanh lá cây (green) và xanh lam (blue).
* Ảnh đa mức xám: Ảnh đa mức xám thường biểu diễn thông tin liên quan đến cường độ đa mức xám của đối tượng trong không gian mà không được thể hiện bởi màu sắc thực của nó.
* Ảnh nhị phân: Ảnh biểu diễn đối tượng bởi hai mức 0 hoặc 1, thường được dùng để biểu diễn, phân biệt sự xuất hiện đối tượng và nền trong mỗi bức ảnh.
* Ảnh hồng ngoại: Biểu diễn trực quan quang phổ, liên quan đến phổ điện từ. Ảnh hồng ngoại cung cấp thông tin ảnh dựa trên phản xạ ánh sáng hồng ngoại hoặc bức xạ hồng ngoại mà các đối tượng trong khung nhìn phát ra. Dựa vào khả năng thu nhận phản xạ hoặc bức xạ hồng ngoại mà các loại camera hồng ngoại thu được hình ảnh trong điều kiện không có ánh sáng nhìn thấy hoặc ánh sáng kém.
* Mức xám

Giá trị mức xám là kết quả của ánh xạ giá trị độ sáng của một điểm ảnh màu trong không gian thực với một giá trị số nguyên dương thể hiện mức độ sáng tối của điểm ảnh đó. Các thang giá trị mức xám thường dùng là 2, 16, 32, 64, 128, 256. Ảnh đa mức xám thường dùng là 256, như vậy mức xám thường xác định trong khoảng [0, 255] tuỳ thuộc vào giá trị mà mỗi điểm ảnh được biểu diễn.

* Độ phân giải của ảnh

Kính thước của lưới pixel hai chiều cùng với kích thước dữ liệu lưu trữ cho mỗi pixel xác định độ phân giải không gian và chất lượng màu của mỗi ảnh[1]. Xét về mặt không gian của độ phân giải thì số cột và số hàng của ảnh xác định số lượng pixel được sử dụng để biểu diễn hình ảnh thu được từ thế giới thực. Như vậy, độ phân giải không gian (spatial resolution) của ảnh là mật độ pixel được xác định trên một ảnh số. Một số độ phân giải thông thường được sử dụng trong các thiết bị hiển thị và lĩnh vực xử lý ảnh như 640× 480, 800 × 600, 1024 × 768 (HD), 192 × 1080 (full HD), 3840 × 2160 (UHD),…

Độ phân giải bit liên quan đến chất lượng ảnh, nó được định nghĩa là số lượng các giá trị khác nhau có thể biểu diễn về cường độ sáng hoặc màu sắc. Ví dụ ảnh nhị phân thì chỉ biểu diễn được hai trạng thái giá trị khác nhau (đen hoặc trắng) mỗi pixel loại này dùng 1 bit, ảnh đa mức xám dùng 8bit cho mỗi pixel, biểu diễn được 256 giá trị khác nhau từ màu đen (giá trị 0) đến trắng (giá trị 255), ảnh màu RGB dùng 24 bit có thể biểu diễn được hơn 16 triệu màu (224=16.777.216).

Độ phân giải bit của một ảnh không nhất thiết phải tương ứng với độ phân giải của hệ thống ảnh. Thông thường các máy ảnh hiện đại ngày nay tự động điều chỉnh để đáp ứng tối đa và tối thiểu của trường ảnh thu nhận được và phạm vi này được chia tự động thành một số lượng phù hợp các bit, ví dụ như chia thành *N* mức. Trong trường hợp như vậy, độ phân giải bit của ảnh thường thấp hơn độ chính xác của thiết bị.

* Các thành phần trong hệ thống xử lý ảnh

Một hệ thống xử lý ảnh thường bao gồm các thành phần chính như thiết bị phần cứng (máy ảnh) để chụp hình và lưu trữ dữ liệu, các công cụ phần mềm phục vụ xử lý và giải quyết yêu cầu của chức năng hệ thống đề ra.

Trong lĩnh vực khoa học máy tính, hệ thống xử lý ảnh là đối tượng nghiên cứu liên quan đến kỹ thuật thị giác máy (computer vision), là quá trình biến đổi từ một ảnh ban đầu được thu nhận từ thiết bị sang một không gian mới sao cho làm nổi bật đặc tính dữ liệu, thuận lợi cho quá trình xử lý thông tin và nâng cao độ chính xác. Một hệ thống xử lý ảnh thường gồm một số thành phần chính sau:



Thu

nhận

Tiền

xử lý

Trích chọn

đặc trưng

Phân loại, nhận

dạng mẫu

Ra

quyết định

Biểu diễn

tri thức

*Hình 1. Sơ đồ tổng quát hệ thống xử lý ảnh*

* Thu nhận ảnh là việc hình ảnh về thế giới thực được thu nhận và chuyển qua tín hiệu ảnh rời rạc thông qua máy ảnh kỹ thuật số hoặc các thiết bị thu hình ảnh khác.
* Tiền xử lý là bước xử lý trên ảnh đầu vào nhằm khử nhiễu, làm nổi bật một số tính chất của ảnh nhằm nâng cao chất lượng các bước xử lý sau.
* Trích chọn đặc trưng là quá trình biến đổi dữ liệu ảnh đầu vào thành tập các đặc trưng. Các đặc trưng thường có đặc tính phân biệt cao của mẫu đầu vào giúp cho việc phân biệt mẫu dữ liệu ảnh dễ dàng hơn nhằm nâng cao chất lượng phân loại mẫu so với xử lý dữ liệu thô trên giá trị pixel ảnh. Việc trích chọn đặc trưng cũng có thể làm giảm kích thước thể hiện thông tin trong ảnh trong khi dữ liệu về đặc trưng ảnh có tính phân biệt cao.
* Phân loại, nhận dạng mẫu là quá trình xử lý dữ liệu bằng các kỹ thuật, phương pháp phân tích đặc trưng để phân loại mẫu về các nhóm có một số tính chất chung. Các phương pháp phân loại, nhận dạng mẫu thường liên quan đến kỹ thuật học máy, bao gồm cả học có giám sát và học không có giám sát.
* Biểu diễn tri thức là bước thể hiện mức cao của biểu diễn dữ liệu, các mẫu dữ liệu sau khi phân loại, nhận dạng được biểu diễn dưới dạng tri thức giúp hệ thống có khả năng “hiểu biết” ngữ nghĩa của nó theo từng kiểu ứng dụng khác nhau trong hệ thống trí tuệ nhân tạo và hệ thống thông minh.
* Ra quyết định là bước cuối cùng của một hệ thống trong lĩnh vực hệ thống thông minh. Các mẫu được biểu diễn dưới dạng tri thức và được suy luận ngữ nghĩa để đưa ra các quyết định thực hiện một nhiệm vụ nào đó.

**1.1.2 Các hệ màu trong máy tính**

*1.1.2.1 Các hệ màu thông dụng*

* Hệ màu RGB

Ảnh số trong hệ màu RGB (Red- Green- Blue) được biểu diễn bởi 3 kênh màu, gồm đỏ (red), xanh lá cây (green), xanh lam (xanh da trời - blue). Hệ màu RGB là hệ màu được sử dụng trong kỹ thuật hiển thị hình ảnh trên thiết bị điện tử như máy tính, TV.

Trong biểu diễn và xử lý hệ màu RGB, mỗi kênh màu được mã hóa bằng 1 byte (8 bit) thể hiện 256 giá trị cường độ sáng khác nhau với dải giá trị trong khoảng từ 0 đến 255. Trường hợp mỗi kênh màu mã hóa bằng 1 byte được gọi là ảnh 24 bit màu (8 bit × 3 kênh- 24 bit). Ảnh màu 24 bit có thể mã hóa được 256× 256× 256 = 16.777.216 giá trị màu khác nhau.



*Hình 2. Phối trộn màu trong hệ màu R-G-B*

Một số màu đặc biệt được thể hiện trong bảng sau:

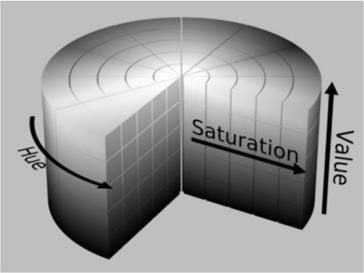
|  |  |
| --- | --- |
| Màu | Giá trị |
| Đen | [0, 0, 0] |
| Trắng | [255, 255, 255] |
| Đỏ | [255, 0, 0] |
| Xanh lục | [0, 255, 0] |
| Xanh lam | [0, 0, 255] |
| Vàng | [255, 255, 0] |
| Xanh ngọc | [0, 255, 255] |
| Hồng cánh sen | [255, 0, 255] |

Bảng 1.1. Một số màu đặc biệt

* Hệ màu HSV

Hệ màu HSV (Hue Saturation Value) có tên gọi khác là là HSI (Hue Saturation Intensity) hoặc HSL (Hue Saturation Lightness). Hệ màu HSV gồm bộ ba giá trị H, S,V.

Mô hình hệ màu này do tác giả Alvy Ray Smith đề xuất năm 1978[3]. Hệ màu này dựa trên các đặc tính màu trực quan được như sắc (tint), bóng (shade) và tông màu (tone) hoặc họ màu, độ thuần khiết và độ sáng của hình ảnh trong thực tế để biểu diễn.



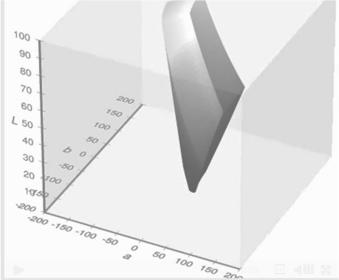
*Hình 3. Không gian màu H-S-V*

Trong mô hình này, một hệ tọa độ hình trụ được dùng để biểu diễn không gian màu: – Hue là giá trị thể hiện sắc màu của hình ảnh (hay còn gọi là vùng màu), sắc màu thông thường được dùng để chỉ tên gọi của màu như đỏ, lục, lam, chàm, tím,…

Các sắc màu khác nhau được biểu diễn trên một vòng tròn chỉ màu, giá trị từ 0 đến 360°.

* Saturation là giá trị thể hiện mức độ bão hòa màu. Giá trị của nó để chỉ mức độ thuần khiết của màu. Nói cách khác, khi ảnh có độ bão hòa cao, màu sẽ trong và rực rỡ hơn giá trị bão hòa thấp. Giá trị của S (saturation) nằm trong đoạn [0, 1], trong đó S đạt giá trị cao nhất (giá trị 1) là màu tinh khiết nhất, hoàn toàn không pha trắng, nghĩa là S càng lớn thì màu càng tinh khiết, nguyên chất.
* Value là giá trị đo độ sáng của màu (intensity hoặc lightness). Thành phần V có giá trị trong đoạn [0, 1] với giá trị đặt biệt V = 0 thì ảnh là hoàn toàn tối (đen), ngược lại V = 1 là hoàn toàn sáng. Giá trị V càng lớn thì màu càng sáng.
* Hệ màu Lab

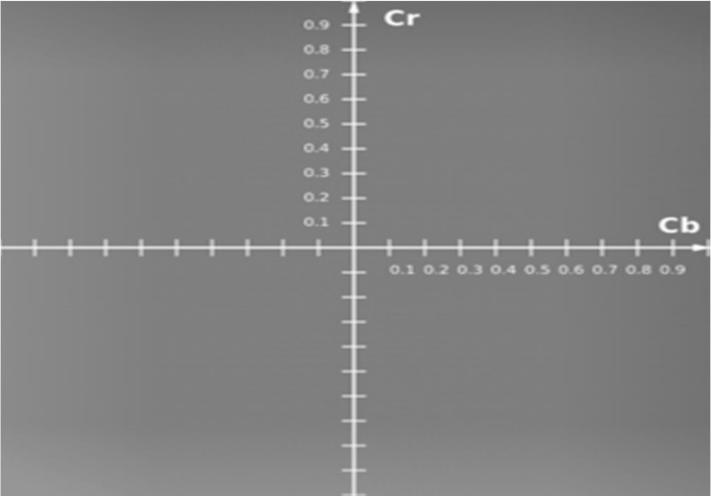
Không gian màu Lab thể hiện mô hình toán học của tất cả các màu mà con người cảm nhận được trong không gian 3 chiều với giá trị L thể hiện cho màu sáng, a và b là các thành phần màu xanh đỏ (green–red) và xanh vàng (blue–yellow). Hệ màu Lab được xem là mô hình màu độc lập đối với thiết bị và thường được sử dụng làm cơ sở tham chiếu khi chuyển đổi từ một không gian màu này sang một không gian màu khác. Hệ màu này sau đó phát triển theo các phiên bản CIELab (đề xuất bởi Hunter) và CIEL\*a\*b\* (đề xuất năm 1976). Theo mô hình Lab, tất cả các màu có cùng một độ sáng sẽ nằm trên cùng một mặt phẳng có dạng hình tròn theo 2 trục a và b. Màu có giá trị a dương thì ngả đỏ, màu có giá trị a âm thì ngả lục. Tương tự b dương thì ngả vàng và b âm thì ngả lam. Còn độ sáng của màu thì thay đổi theo trục dọc.



*Hình 4. Thể hiện màu trong không gian màu Lab*

* Hệ màu YCbCr

Các hệ màu YCbCr và Y'CbCr còn được gọi với các tên khác như YCBCR và Y'CBCR tương ứng, hai hệ màu này là một họ không gian màu được sử dụng nhiều trong các hệ thống video, ảnh kỹ thuật số và các hệ thống thiết bị phát hình điện tử. Trong hệ màu YCbCr, thành phần Y đại diện cho độ sáng của ảnh và Cb và Cr là các thành phần màu tương ứng với màu xanh lam (blue) và màu đỏ (red). Hệ màu YCbCr trong ảnh kỹ thuật số tương đương với hệ màu YUV trong ảnh tín hiệu tuần tự (analog). Không gian màu YCbCr được định nghĩa trong hệ tọa độ tương ứng với không gian màu RGB. Giá trị các kênh màu R, G và B tại mỗi pixel được sử dụng để tổng hợp lại với nhau tạo ra giá trị đơn của thành phần Y biểu diễn độ sáng chung tại pixel tương ứng. Các thành phần Cb và Cr được tổng hợp từ các giá trị của thành phần Y và các kênh màu B và R tương ứng trong hệ màu RGB.



*Hình 5. Mặt phẳng thể hiện màu theo Cb và Cr với hệ số Y'=0.5*

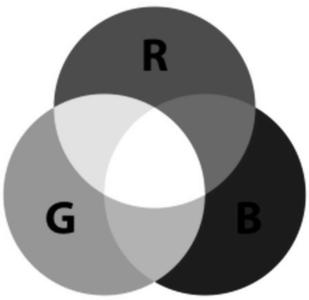
**1.1.3 Một số loại ảnh thông dụng**

* Ảnh màu

Ảnh màu được hiểu chung là ảnh thể hiện các đối tượng theo màu sắc của nó mà mắt thường cảm nhận được. Nói cách khác, ảnh màu được hiểu là một ảnh chứa một hoặc nhiều kênh màu xác định màu sắc tại các vị trí cụ thể của ảnh I(x,y).

Theo Thomas Young thì ảnh màu được tổ hợp từ 3 màu cơ bản là đỏ (R-red), xanh lục (G- green), xanh lam (B-blue) và thường thu nhận trên các dải băng tần khác nhau. Không gian màu RGB thường được dùng nhiều trong các thiết bị hiển thị điện tử so với các không gian màu khác như HSV, YCbCr. Một ảnh màu thường được lưu trữ trong bộ nhớ như là một bản đồ rapter hay bản đồ pixel (bitmap) bởi một mảng hai chiều với mỗi phần tử là bộ ba giá trị màu tương ứng với mỗi kênh màu.

Trong ảnh RGB, thông thường mỗi kênh màu dùng 8 bit để để biểu diễn cho một điểm ảnh, vậy một điểm ảnh màu cần 24 bit tương ứng với 3 kênh màu thành phần. Giá trị mỗi kênh màu được chia thành n mức màu khác nhau tương ứng từ 0 đến n-1, nếu 8 bit thì có 256 từ 0 đến 255. Mỗi giá trị thể hiện cường độ sáng của kênh màu tương ứng. Trong hệ màu RGB, việc lưu trữ ảnh màu theo từng kênh màu riêng biệt rất dễ dàng.

Trong kỹ thuật xử lý ảnh, ảnh đa mức xám (grey image) còn được gọi là ảnh đơn sắc (monochromatic). Ảnh đa mức xám dùng một kênh để thể hiện cường độ sáng của điểm ảnh. Mỗi điểm ảnh có một giá trị mức xám độ sáng từ màu đen (giá trị nhỏ nhất 0) đến màu trắng (giá trị lớn nhất).

*Hình 6. Tạo màu theo nguyên lý R-G-B*

Như vậy, khác với ảnh màu RGB, ảnh đa mức xám chỉ dùng một thành phần giá trị nên nếu dùng cùng độ bit để mã hóa cho mỗi thành phần thì ảnh đa mức xám có kích thước bằng 1/3 so với ảnh màu RGB.

* Ảnh đa mức xám

Một số loại ảnh đa mức xám phổ biến phân theo kích thước bit dùng cho mỗi pixel:

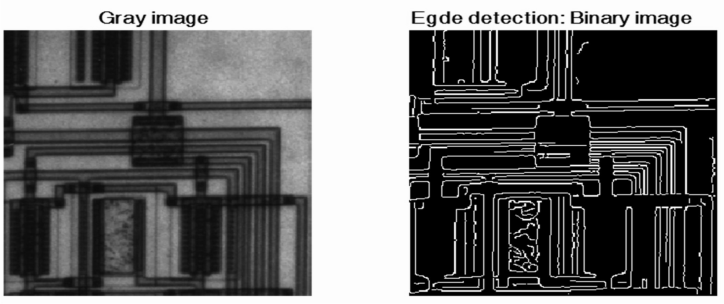
* Ảnh 256 mức xám: Loại ảnh này cần dùng 8 bit cho mỗi điểm ảnh, có giá trị nằm trong khoảng [0, 255] tương ứng với biến đổi cường độ sáng từ đen qua trắng.
* Ảnh 8 mức xám: Loại ảnh này cần dùng 3 bit cho mỗi điểm ảnh, giá trị nằm trong khoảng [0, 7]. Như vậy, loại ảnh này có độ phân giải màu thấp hơn so với loại ảnh 8bit. Giá trị điểm ảnh bằng 0 nghĩa là điểm ảnh đó tối (đen), giá trị điểm ảnh lớn nhất nghĩa là điểm ảnh đó trắng. Giá trị điểm ảnh càng lớn thì điểm ảnh đó càng sáng.
* Ảnh 2 mức xám: Mỗi pixel chỉ biểu diễn 2 mức cường độ sáng tương ứng với đen
* và trắng (1). Để tránh nhầm lẫn ảnh đen trắng với ảnh đa mức xám, người ta thường gọi là ảnh nhị phân hay ảnh đen trắng nhị phân. Như vậy mỗi pixel chỉ cần dùng 1 bit để biểu diễn. Ảnh đen trắng thường được dùng để biểu diễn đối tượng và nền trong ảnh.

Trong một số trường hợp, màu sắc của hình ảnh không quan trọng và màu có thể nhạy cảm với các nguồn sáng mà chỉ cần biểu diễn hình dáng theo mức độ sáng tối của đối tượng trong hình ảnh thì ảnh đa mức xám được sử dụng để giảm dung lượng lưu trữ và giảm thiểu ảnh hưởng của nguồn sáng.

*Hình 7. Ảnh màu RGB và Ảnh đa mức xám*

* Ảnh đen trắng



*Hình 8. Ảnh nhị phân trong biểu diễn cạnh*

Như đã đề cập ở mục trước, ảnh đen trắng là trường hợp đặc biệt của ảnh đa mức xám chỉ gồm 2 màu là đen và trắng, thường gọi là ảnh nhị phân. Ảnh nhị phân khá đơn giản, các phần tử ảnh có thể coi như các phần tử nhị phân. Ảnh nhị phân thường được dùng để biểu diễn trạng thái đối tượng, phân biệt đối tượng trong ảnh với nền, hoặc để biểu diễn các đường biên đối tượng, vùng ảnh.

* Ảnh quang phổ

Ảnh quang phổ là trường hợp tổng quát của hình ảnh với các loại tín hiệu tương ứng với các loại bước sóng đặc biệt nào đó, bao gồm cả các loại bước sóng nhìn thấy (đối với ảnh số thông thường). Ảnh quang phổ có thể biểu diễn dưới dạng ảnh màu hoặc ảnh đa mức xám, ảnh nhị phân.

Có nhiều loại ảnh quang phổ khác nhau như ảnh quang phổ X-ray, ảnh quang phổ hồng ngoại,… Ảnh phổ hồng ngoại là hình ảnh thu được từ tia hồng ngoại do bức xạ điện từ có bước sóng dài hơn ánh sáng nhìn thấy nhưng ngắn hơn tia bức xạ vi ba. Hồng ngoại có nghĩa là bước sóng ngoài mức đỏ, là bước sóng dài nhất trong ánh sáng nhìn thấy được.

*Hình 9. Ảnh hồng ngoại (ảnh thấy bằng mắt thường, bức xạ hồng ngoại)*

Hiện nay loại camera hồng ngoại trở nên phổ biến, đặc biệt là các loại camera giám sát an ninh, camera trong các hệ thống tự động hóa. Camera hồng ngoại có khả năng thu bức xạ hồng ngoại để chuyển đổi qua ảnh hiển thị trong ngưỡng nhìn thấy bằng mắt thường. Do vậy, camera hồng ngoại được sử dụng nhiều trong các ứng dụng ban đêm và các điều kiện thiếu ảnh sáng.

**1.2. Tổng quan về học máy**

**1.2.1 Học có giám sát**

Trong kỹ thuật học có giám sát, các hệ thống máy tính sẽ đào tạo mô hình, học hỏi từ dữ liệu dưới sự giám sát từ bên ngoài.

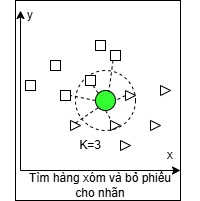
Chúng sẽ chứa một mô hình có thể dự đoán, với sự tập hợp của các dữ liệu đã được gắn nhãn trước đó. Dữ liệu được gắn nhãn là những dữ liệu đã biết câu trả lời đích.

Các loại học máy có giám sát:

* Hồi quy: Dự đoán một giá trị đầu ra duy nhất bằng cách sử dụng dữ liệu huấn luyện.
* Hồi quy logistic: Được sử dụng để ước tính các giá trị rời rạc dựa trên một tập hợp các biến độc lập đã cho. Nó giúp bạn dự đoán xác suất xảy ra sự kiện bằng cách khớp dữ liệu với hàm logit. Vì vậy, nó còn được gọi là hồi quy logistic. Vì nó dự đoán xác suất nên giá trị đầu ra của nó nằm trong khoảng từ 0 đến 1.
* Phân loại: Là nhóm đầu ra bên trong một lớp. Nếu thuật toán cố gắng gắn nhãn đầu vào thành hai lớp riêng biệt thì nó được gọi là phân loại nhị phân. Việc lựa chọn giữa nhiều hơn hai lớp được gọi là phân loại nhiều lớp.
* Điểm mạnh: Cây phân loại hoạt động rất tốt trong thực tế.
* Điểm yếu: Không bị ràng buộc, các cây riêng lẻ có xu hướng bị trang bị quá mức.
* Phân loại Naïve Bayes: Mô hình Naive Bayesian (NBN) rất dễ xây dựng và rất hữu ích cho các tập dữ liệu lớn. Phương pháp này bao gồm các biểu đồ chu kỳ trực tiếp với một cha và một số con. Nó giả định sự độc lập giữa các nút con tách biệt khỏi nút cha của chúng.
* Cây quyết định: Phân loại cá thể bằng cách sắp xếp chúng dựa trên giá trị đặc trưng. Trong phương pháp này, mỗi chế độ là một tính năng của một thể hiện. Nó phải được phân loại và mỗi nhánh đại diện cho một giá trị mà nút có thể đảm nhận. Đây là một kỹ thuật được sử dụng rộng rãi để phân loại. Trong phương pháp này, phân loại là một cây được gọi là cây quyết định. Nó giúp bạn ước tính giá trị thực.
* Máy hỗ trợ vector: Là một loại thuật toán học được phát triển vào năm 1990. Phương pháp này dựa trên kết quả của lý thuyết học thống kê do Vap Nik giới thiệu.
* Các máy SVM cũng được kết nối chặt chẽ với các hàm kernel, đây là khái niệm trung tâm của hầu hết các nhiệm vụ học tập. Kernel framework và SVM được sử dụng trong nhiều lĩnh vực khác nhau. Nó bao gồm việc truy xuất thông tin đa phương tiện, tin sinh học và nhận dạng mẫu.

*1.2.1.1 KNN*

KNN (K-Nearest Neighbors) là một mô hình học máy có giám sát. Trong học máy có giám sát, mô hình được huấn luyện trên một tập dữ liệu đã được gán nhãn, tức là dữ liệu đầu vào đã có kết quả đầu ra tương ứng.

****

*Hình 10. Mô hình học knn*

Thuật toán KNN dựa trên nguyên tắc các đối tượng hoặc điểm dữ liệu gần nhau trong không gian đặc trưng có khả năng thuộc cùng một lớp hoặc có kết quả đầu ra tương tự nhau.

Thuật toán KNN gán một điểm dữ liệu mới cho lớp phổ biến nhất trong k hàng xóm gần nhất của nó. Giá trị của k là một siêu tham số cần được chỉ định trước khi đào tạo mô hình.

* KNN hoạt động theo quy trình gồm 4 bước chính:
* Xác định tham số K (số láng giềng gần nhất).
* Tính khoảng cách từ điểm đang xét đến tất cả các điểm trong tập dữ liệu cho trước.
* Sắp xếp các khoảng cách đó theo thứ tự tăng dần.
* Xét trong tập K điểm gần nhất với điểm đang xét, nếu số lượng điểm của loại nào cao hơn thì coi như điểm đang xét thuộc loại đó.

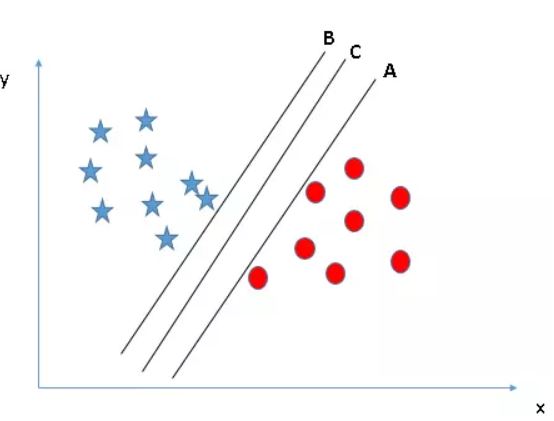
 KNN giả định rằng những dữ liệu tương tự nhau sẽ nằm gần nhau trong một không gian đặc trưng, và nhiệm vụ của thuật toán là tìm ra **k** điểm dữ liệu gần nhất với điểm cần phân loại. Khoảng cách giữa các điểm thường được tính bằng các phương pháp như khoảng cách Euclid, khoảng cách Manhattan, hoặc các công thức tính khoảng cách

* Ưu điểm:
* Đơn giản và dễ hiểu: Thuật toán KNN đơn giản và dễ hiểu. Nó không yêu cầu bất kỳ giả định nào về phân phối dữ liệu cơ bản và có thể dễ dàng triển khai bằng bất kỳ ngôn ngữ lập trình nào.
* Không tham số: Thuật toán KNN không tham số, có nghĩa là nó không đưa ra bất kỳ giả định nào về phân phối dữ liệu cơ bản. Điều này làm cho nó phù hợp cho cả phân phối dữ liệu tuyến tính và phi tuyến tính.
* Không có giai đoạn đào tạo: Thuật toán KNN không yêu cầu giai đoạn đào tạo. Mô hình được lưu trữ đơn giản trong bộ nhớ và các điểm dữ liệu mới có thể được phân loại hoặc dự đoán trong thời gian thực.
* Linh hoạt: Thuật toán KNN có thể được sử dụng cho cả nhiệm vụ phân loại và hồi quy. Nó cũng có thể được điều chỉnh để phân loại nhiều lớp bằng cách sử dụng một sửa đổi đơn giản.
* Mạnh đối với dữ liệu nhiễu: Thuật toán KNN mạnh đối với dữ liệu nhiễu vì nó không đưa ra bất kỳ giả định nào về phân phối dữ liệu cơ bản. Nó có thể xử lý các giá trị ngoại lai và các điểm dữ liệu nhiễu bằng cách lấy trung bình các giá trị của các hàng xóm gần nhất.
* Không cần kỹ thuật tính năng: Thuật toán KNN không yêu cầu kỹ thuật tính năng. Nó có thể hoạt động với bất kỳ loại dữ liệu đầu vào nào, bao gồm dữ liệu phân loại và số.
* Độ chính xác cao: Thuật toán KNN có thể đạt độ chính xác cao, đặc biệt khi giá trị của k được chọn cẩn thận. Nó thường được sử dụng làm thuật toán chuẩn cho nhiều tác vụ học máy.
* Nhược điểm:
* Đắt về mặt tính toán: Thuật toán KNN rất tốn kém về mặt tính toán, đặc biệt là khi xử lý các tập dữ liệu lớn. Đối với mỗi điểm dữ liệu mới, thuật toán phải tính toán khoảng cách đến tất cả các điểm dữ liệu khác trong tập dữ liệu, điều này có thể tốn thời gian.
* Sử dụng nhiều bộ nhớ: Thuật toán KNN sử dụng nhiều bộ nhớ vì nó phải lưu trữ tất cả dữ liệu đào tạo trong bộ nhớ. Đây có thể là một vấn đề đối với các tập dữ liệu lớn, đặc biệt là khi làm việc với dữ liệu nhiều chiều.
* Độ nhạy với tỷ lệ tính năng: Thuật toán KNN nhạy cảm với tỷ lệ tính năng. Nếu các tính năng có tỷ lệ khác nhau, thì thuật toán có thể tăng trọng số cho các tính năng có tỷ lệ lớn hơn. Đây có thể là một vấn đề khi xử lý dữ liệu nhiều chiều hoặc khi các tính năng có các đơn vị đo lường khác nhau.
* Nhạy cảm với lựa chọn k(Phải tìm giá trị k tối ưu): Thuật toán KNN nhạy cảm với lựa chọn k, số lượng hàng xóm gần nhất để xem xét. Nếu k quá nhỏ thì thuật toán có thể quá nhạy cảm với nhiễu trong dữ liệu. Nếu k quá lớn thì thuật toán có thể không nắm bắt được cấu trúc cục bộ của dữ liệu.
* Thời gian dự đoán có thể chậm: Vì thuật toán KNN không có giai đoạn huấn luyện và yêu cầu tính toán khoảng cách cho mỗi điểm mới nên thời gian dự đoán có thể chậm đối với các tập dữ liệu lớn.
* Nguy cơ về chiều: Khi số lượng tính năng tăng lên, thuật toán KNN trở nên kém hiệu quả hơn do khoảng cách giữa các điểm dữ liệu ngày càng giống nhau. Điều này được gọi là lời nguyền của chiều và có thể gây khó khăn cho việc tìm kiếm những người hàng xóm gần nhất có ý nghĩa.

***1.2.1.2 SVM***

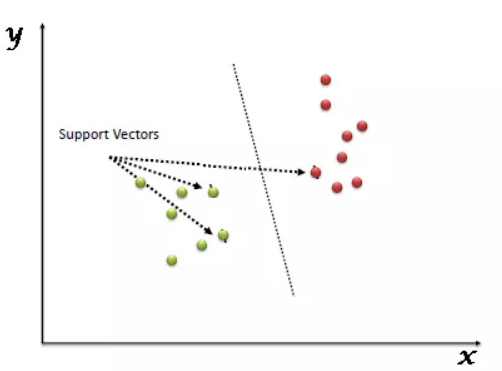
* Khái niệm
* SVM thuật toán có giám sát.
* Hoạt động dựa trên nguyên lý tìm kiếm một *hyperplane (siêu phẳng)* tối ưu để phân chia các điểm dữ liệu thuộc các lớp khác nhau trong không gian.
* Hyperplane này được chọn sao cho khoảng cách từ các điểm dữ liệu gần nhất của hai lớp đến hyperplane là lớn nhất.
* Các điểm dữ liệu này được gọi là "support vectors".
* Các thành phần chính
* Hyperplane: Trong không gian 2 chiều, hyperplane là một đường thẳng; trong không gian 3 chiều, nó là một mặt phẳng.

Quy tắc:



+Xác định khoảng cách lớn nhất từ điểu gần nhất của một lớp nào đó đến đường hyper-plane. Khoảng cách này được gọi là "Margin"

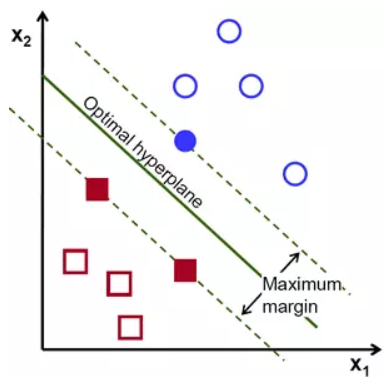
* Support Vector: Là các điểm dữ liệu nằm gần hyperplane nhất.



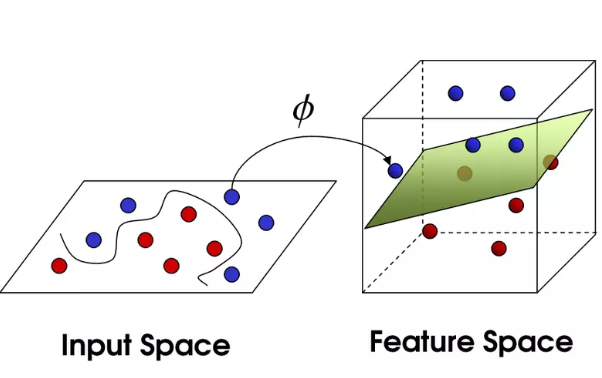
* Margin trong SVM

+Là khoảng cách giữa siêu phẳng đến 2 điểm dữ liệu gần nhất tương ứng với các phân lớp.

+Phương pháp SVM luôn cố gắng cực đại hóa margin này, từ đó thu được một siêu phẳng tạo khoảng cách xa nhất.



* Kết luận
* SVM là một phương pháp hiệu quả cho bài toán phân lớp dữ liệu.
* Một yếu tố làm nên hiệu quả của SVM đó là việc sử dụng Kernel function khiến cho các phương pháp chuyển không gian trở nên linh hoạt hơn.



**1.2.2 Học không giám sát**

Học tập không giám sát là một kỹ thuật học máy trong đó người dùng không cần giám sát mô hình. Thay vào đó, nó cho phép mô hình tự hoạt động để khám phá các mẫu và thông tin mà trước đây không bị phát hiện. Nó chủ yếu xử lý các dữ liệu chưa được ghi nhãn.

Học không giám sát có thể được nhóm lại thành 2 loại chính:

* Phân cụm (Clustering)

Là phương pháp học không giám sát, trong đó các đối tượng dữ liệu được nhóm lại thành các cụm (clusters) dựa trên tính tương đồng hoặc khoảng cách giữa chúng. Các điểm trong cùng một cụm có nhiều đặc điểm giống nhau hơn so với các điểm trong các cụm khác.

* Liên kết (Association)

Là một phương pháp học không giám sát, tập trung vào việc tìm kiếm các mối quan hệ hoặc quy luật tiềm ẩn trong dữ liệu. Các quy luật liên kết giúp xác định các mẫu thường xuất hiện cùng nhau trong một tập hợp dữ liệu.

*1.2.2.1 K-Means*

Phân cụm K-Means là một thuật toán Học máy không giám sát, thuật toán này nhóm tập dữ liệu chưa được gắn nhãn thành các cụm khác nhau. Học máy không giám sát là quá trình dạy máy tính sử dụng dữ liệu chưa được gắn nhãn, chưa được phân loại và cho phép thuật toán hoạt động trên dữ liệu đó mà không cần giám sát. Không có bất kỳ đào tạo dữ liệu nào trước đó, công việc của học máy trong trường hợp này là sắp xếp dữ liệu chưa được sắp xếp theo các điểm tương đồng, mẫu và biến thể.

Trong thuật toán K-means clustering, chúng ta không biết nhãn (label) của từng điểm dữ liệu. Mục đích là làm thể nào để phân dữ liệu thành các cụm (cluster) khác nhau sao cho dữ liệu trong cùng một cụm có tính chất giống nhau.

* Các bước của thuật toán K-Means Clustering

Đầu vào: Dữ liệu X và số lượng cluster cần tìm K.

Đầu ra: Các center M và label vector cho từng điểm dữ liệu Y.

1. Chọn K điểm bất kỳ làm các center ban đầu.

2. Phân mỗi điểm dữ liệu vào cluster có center gần nó nhất.

3. Nếu việc gán dữ liệu vào từng cluster ở bước 2 không thay đổi so với vòng lặp trước nó thì ta dừng thuật toán.

4. Cập nhật center cho từng cluster bằng cách lấy trung bình cộng của tất các các điểm dữ liệu đã được gán vào cluster đó sau bước 2.

5. Quay lại bước 2.

* Mục đích

Từ dữ liệu đầu vào và số lượng nhóm chúng ta muốn tìm, hãy chỉ ra tâm của mỗi nhóm và phân các điểm dữ liệu vào các nhóm tương ứng

Trong thuật toán K-means, độ đo khoảng cách được sử dụng để xác định độ tương đồng giữa các phần tử và ảnh hưởng đến hình dạng của các cụm. Độ đo khoảng cách được sử dụng phải đáp ứng được một số yêu cầu nhất định. Nó phải là một hàm số phi tuyến, không âm và có tính đối xứng.

Phương pháp để đo khoảng cách như: Euclidean trường hợp phổ biến nhất là xác định khoảng cách giữa hai điểm. Nếu chúng ta có điểm p và điểm q, khoảng cách euclidean là một đường thẳng bình thường. Đó là khoảng cách giữa hai điểm trong không gian Euclide.



*1.2.2.2 Fuzzy C- Means*

Fuzzy C-Means (FCM) là một thuật toán học máy không giám sát, được sử dụng để phân cụm dữ liệu thành các cụm dựa trên mức độ thành viên (membership degree). Không giống như K-Means, trong đó mỗi điểm dữ liệu chỉ thuộc về một cụm, FCM cho phép mỗi điểm dữ liệu thuộc về nhiều cụm với mức độ thành viên khác nhau (giá trị từ 0 đến 1).

Học máy không giám sát là quá trình dạy máy tính sử dụng dữ liệu chưa được gắn nhãn hoặc phân loại, từ đó cho phép thuật toán tự tìm ra các cấu trúc hoặc mẫu trong dữ liệu.

Trong trường hợp của FCM, thuật toán sẽ phân chia dữ liệu vào các cụm dựa trên sự tương đồng, nhưng mỗi điểm có thể có mối liên hệ với nhiều cụm.

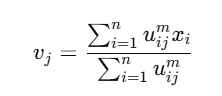
* Các bước của thuật toán Fuzzy C- Means

Đầu vào: Dữ liệu X và số lượng cụm cần tìm C.

Đầu ra: Ma trận thành viên U và tâm cụm V.

1. Khởi tạo ngẫu nhiên một ma trận thành viên U có kích thước n × C (với n là số lượng điểm dữ liệu và C là số lượng cụm). Các giá trị trong ma trận này năm trong khoảng [0,1], sao cho tổng mức độ thành viên của mỗi điểm dữ liệu với tất cả các cụm bằng 1.

2. Sử dụng ma trận U để tính tâm của các cụm theo



* : Tâm cụm thứ j.
* : Mức độ thành viên của điểm dữ liệu đối với cụm j.
* m: Tham số mờ, thường là 2.

3. Cập nhật ma trận thành viên U.

4. Kiểm tra hội tụ, dừng thuật toán nếu sự thay đổi của ma trận U giữa hai lần lặp liên tiếp nhỏ hơn một ngưỡng đã định trước.

5. Quay lại bước 2, nếu chưa hội tụ.

* Mục đích: Từ dữ liệu đầu vào và số lượng cụm cần tìm, thuật toán FCM xác định tâm của mỗi cụm và mức độ thành viên của từng điểm dữ liệu đối với các cụm đó. Điều này cho phép phân chia dữ liệu một cách linh hoạt hơn so với phương pháp K-Means.
* Đặc điểm:
* Mỗi điểm dữ liệu có thể thuộc về nhiều cụm, phù hợp với dữ liệu có sự chồng lấn (overlap) giữa các cụm.
* Phương pháp tiếp cận mờ giúp thuật toán linh hoạt hơn trong các bài toán phức tạp.
* Tâm cụm và mức độ thành viên được cập nhật liên tục để tối ưu hóa việc phân cụm.
* Trong thuật toán FCM, khoảng cách giữa các điểm dữ liệu và tâm cụm được sử dụng để xác định mức độ thành viên. Khoảng cách thường được đo bằng Euclidean.

### **1.2.3 Học bán giám sát**

***Học bán giám sát***

* Học bán giám sát là một phương pháp học máy thu hẹp khoảng cách giữa học có giám sát và học không giám sát bằng cách sử dụng cả dữ liệu có nhãn và không có nhãn.
* Phương pháp này tận dụng lợi thế của dữ liệu không có nhãn phong phú trong khi giảm thiểu yêu cầu đối với các trường hợp có nhãn, thường tốn kém và mất thời gian để có được.

***Tầm quan trọng và sự liên quan***

* Trong các kịch bản học máy thông thường, học có giám sát phụ thuộc rất nhiều vào dữ liệu được gắn nhãn, trong đó mỗi đầu vào được ghép nối với một đầu ra chính xác. Mặt khác, học không giám sát không sử dụng bất kỳ nhãn nào.
* Học bán giám sát tạo ra sự cân bằng bằng cách sử dụng một phần nhỏ dữ liệu được gắn nhãn cùng với một tập dữ liệu không được gắn nhãn lớn hơn.
* Cách tiếp cận này đặc biệt có lợi khi việc gắn nhãn dữ liệu tốn kém hoặc không thực tế nhưng việc thu thập một lượng lớn dữ liệu thô là khả thi.

***Cách thức hoạt động***

* Các mô hình học bán giám sát thường được xây dựng bằng cách ban đầu đào tạo trên một tập dữ liệu có nhãn nhỏ hơn.
* Sau khi mô hình cơ bản được xây dựng, dữ liệu không có nhãn được kết hợp để tinh chỉnh và cải thiện mô hình hơn nữa.
* Các kỹ thuật như tự đào tạo, đồng đào tạo và phương pháp dựa trên đồ thị thường được sử dụng:
* Tự đào tạo : Mô hình dự đoán nhãn cho dữ liệu chưa có nhãn và những dự đoán này được sử dụng như thể chúng là nhãn thực trong vòng đào tạo tiếp theo.
* Đào tạo chung : Hai bộ phân loại được đào tạo trên các góc nhìn khác nhau của dữ liệu, hỗ trợ lẫn nhau bằng cách dán nhãn dữ liệu chưa được dán nhãn.
* Phương pháp dựa trên đồ thị : Các trường hợp dữ liệu được biểu diễn dưới dạng các nút trong đồ thị và các thuật toán học bán giám sát sẽ truyền thông tin nhãn qua các cạnh.

***Ứng dụng trong thế giới thực***

* Học bán giám sát được sử dụng trong nhiều lĩnh vực khác nhau:
* Chăm sóc sức khỏe : Trong hình ảnh y tế, chẳng hạn như chụp MRI hoặc CT, thường tốn nhiều công sức để chú thích tất cả hình ảnh. Một số hình ảnh được các chuyên gia dán nhãn và mô hình học hỏi từ cả hình ảnh có nhãn và không có nhãn để giảm bớt gánh nặng chú thích.
* Xử lý ngôn ngữ tự nhiên (NLP) : Bằng cách giảm lượng dữ liệu văn bản được gắn nhãn thủ công, học bán giám sát có thể hỗ trợ các tác vụ như phân loại văn bản và phân tích tình cảm.

***Sự khác biệt chính từ các khái niệm liên quan***

* Học bán giám sát không nên bị nhầm lẫn với các khái niệm tương tự như học chủ động và học chuyển giao.
* [Học chủ động](https://www.ultralytics.com/vi/glossary/active-learning) liên quan đến việc lựa chọn các mẫu thông tin nhất để dán nhãn nhằm nâng cao hiệu quả học tập.
* Trong khi đó, [học chuyển giao](https://www.ultralytics.com/vi/glossary/transfer-learning) liên quan đến việc chuyển giao kiến thức từ một lĩnh vực để cải thiện hiệu suất trong lĩnh vực khác.

***Thách thức và cân nhắc***

* Học bán giám sát phải đối mặt với những thách thức như đảm bảo độ tin cậy của các nhãn dự đoán và xử lý các phân phối dữ liệu đa dạng.
* Điều này đòi hỏi phải lựa chọn cẩn thận các kỹ thuật thuật toán và đôi khi là xác thực bổ sung để đảm bảo rằng dữ liệu chưa được gắn nhãn được sử dụng hiệu quả mà không gây hiểu lầm cho quá trình học.

***Kết thúc***

* Học bán giám sát là một công cụ mạnh mẽ trong bộ công cụ AI tận dụng sự phong phú của dữ liệu chưa được gắn nhãn để xây dựng các mô hình dự đoán hiệu quả hơn.
* Nó có ứng dụng trong nhiều lĩnh vực, từ chăm sóc sức khỏe đến NLP, cho phép các mô hình mạnh mẽ hơn với ít ví dụ được gắn nhãn hơn.
* Hiểu và khám phá khái niệm này có thể dẫn đến các giải pháp sáng tạo khi tồn tại các nút thắt cổ chai trong việc gắn nhãn dữ liệu.
* Khám phá cách Ultralytics HUB có thể hỗ trợ các nỗ lực AI và học máy tương tự bằng cách cung cấp nền tảng đa năng để đào tạo và triển khai mô hình.

# CHƯƠNG 2 : XÂY DỰNG HỆ THỐNG NHẬN DIỆN CHỮ SỐ TỰ ĐỘNG VỚI OPEN CV

## 2.1. Thư viện

### **2.1.1 Numpy**

* **Khái niệm**

[NumPy](https://numpy.org/) (viết tắt của Numerical Python) là một thư viện quan trọng trong lập trình Python. Công cụ này được tạo ra bởi một cộng đồng các nhà nghiên cứu và nhà phát triển với mục đích giúp các lập trình viên thao tác dữ liệu số học và thực hiện các phép toán toán học phức tạp một cách hiệu quả.

Một trong những ưu điểm nổi bật của NumPy là khả năng làm việc với mảng đa chiều (arrays) và ma trận, cho phép lưu trữ và xử lý dữ liệu dễ dàng. Đây không chỉ là một công cụ hữu ích cho các nhà khoa học dữ liệu, mà còn là một lựa chọn tuyệt vời cho những người làm việc trong các lĩnh vực như khoa học máy tính, xử lý ảnh và thị giác máy tính.

NumPy cung cấp một loạt các hàm và phương pháp giúp thực hiện các phép toán toán học, thống kê và biểu đồ hóa dữ liệu vô cùng linh hoạt. Đồng thời, đây cũng là một phần quan trọng của hệ sinh thái các thư viện tính toán trong Python. Với NumPy, việc xử lý dữ liệu số học trở nên dễ dàng và hiệu quả hơn, giúp Python trở thành một trong những ngôn ngữ lập trình phổ biến nhất trong lĩnh vực khoa học dữ liệu và nghiên cứu toán học.



* **Các tính năng nổi bật của NumPy**

*Mảng đa chiều (Arrays)*

* NumPy cho phép tạo ra và quản lý các Arrays (mảng), giúp lưu trữ dữ liệu số học một cách dễ dàng và hiệu quả. Nhờ đó, việc xử lý các tập dữ liệu phức tạp trở nên thuận tiện hơn.

*Phép toán số học*

* NumPy cung cấp các phép toán toán học như cộng, trừ, nhân, chia trên mảng dữ liệu nhanh chóng. Các phép toán này được tối ưu hóa để xử lý dữ liệu lớn hiệu quả hơn.

*Phép toán Vectorized*

* NumPy cho phép thực hiện các phép toán trên toàn bộ mảng dữ liệu, không cần viết vòng lặp. Điều này giúp tăng tốc độ xử lý dữ liệu đáng kể so với việc sử dụng các danh sách thông thường trong Python.

*Broadcasting*

* NumPy tự động mở rộng kích thước của các mảng để phù hợp với nhau trong các phép toán. Do đó, các phép toán giữa các mảng có kích thước khác nhau có thể được thực hiện mà không cần phải thay đổi kích thước của chúng theo cách thủ công.

*Hàm toán học và thống kê*

* NumPy cung cấp một loạt các hàm toán học và thống kê như sin, cos, mean, median, std,…để xử lý dữ liệu số học.

*Indexing và Slicing*

* NumPy hỗ trợ các phép indexing và slicing trên mảng dữ liệu giúp bạn truy cập và chỉnh sửa dữ liệu một cách dễ dàng.

*Phép toán trên mảng logic*

* NumPy cho phép bạn thực hiện các phép toán logic như AND, OR, NOT trên mảng dữ liệu một cách dễ dàng.

*Phép toán đại số tuyến tính*

* NumPy cung cấp nhiều hàm và phương thức để thực hiện các phép toán đại số tuyến tính như ma trận chuyển vị, nhân ma trận, giải hệ phương trình tuyến tính và nhiều thuật toán khác.

*Làm việc với dữ liệu đa chiều*

* NumPy cho phép bạn thực hiện các phép toán trên dữ liệu đa chiều, thích hợp cho việc xử lý hình ảnh, âm thanh và các dữ liệu khoa học khác.

*Tích hợp C với Python*

* NumPy được viết bằng ngôn ngữ lập trình C, điều này giúp tăng hiệu suất của các phép toán, đặc biệt là trên các mảng dữ liệu lớn.

*Tích hợp với các thư viện khác*

* Cộng đồng người dùng NumPy lớn mạnh, điều này có nghĩa là sẽ có rất nhiều các thư viện và công cụ khác trong hệ sinh thái của Python được xây dựng dựa trên NumPy.
* Đồng thời, NumPy cũng tương thích tốt với nhiều thư viện và framework khác trong hệ sinh thái của Python, giúp bạn dễ dàng tích hợp vào các dự án lớn và phức tạp.
* **Một số hạn chế của NumPy**
* Đối với người mới học lập trình hoặc Python, việc sử dụng NumPy có thể đòi hỏi một thời gian để nắm vững cú pháp và các chức năng của thư viện này.
* Trong một số ứng dụng như web development, NumPy có thể không cần thiết và không hiệu quả so với các thư viện khác hoặc khi sử dụng ngôn ngữ lập trình khác.
* Các mảng dữ liệu trong NumPy có thể tốn nhiều bộ nhớ hơn so với các danh sách thông thường trong Python. Điều này có thể là một vấn đề đối với các ứng dụng yêu cầu tiết kiệm bộ nhớ.
* Tính năng broadcasting có thể dẫn đến kết quả không mong muốn nếu không sử dụng đúng cách, đặc biệt đối với người mới học sử dụng NumPy.
* **Đối tượng kiểu dữ liệu (dtype) trong NumPy**

*Số nguyên*

arr = np.array([1, 2, 3, 4], dtype='int32')

# dtype của arr là int32

*Số thực*

arr = np.array([1.1, 2.2, 3.3], dtype='float64')

# dtype của arr là float64

*Số nguyên không dấu*

arr = np.array([0, 1, 2], dtype='uint8')

# dtype của arr là uint8 (8-bit số nguyên không dấu)

*Chuỗi ký tự*

arr = np.array(['apple', 'banana', 'cherry'], dtype='str')

# dtype của arr là str (chuỗi ký tự)

*Kiểu dữ liệu nguyên thủy của Python*

arr = np.array([1, 2, 3], dtype=object)

# dtype của arr là object (kiểu dữ liệu nguyên thủy của Python)

*Kiểu dữ liệu tùy chỉnh*

Ngoài ra, NumPy cho phép định nghĩa các kiểu dữ liệu tùy chỉnh, giúp xác định cách dữ liệu được lưu trữ và xử lý.

Việc hiểu và quản lý đúng đối tượng kiểu dữ liệu trong NumPy là điều quan trọng để đảm bảo tính chính xác cũng như hiệu suất của các phép toán trong quá trình lập trình.

* **Các phương thức trong Numpy**
* Phương thức tạo mảng

np.array(): Tạo mảng từ một danh sách hoặc tuple.

np.zeros(shape): Tạo mảng với tất cả các phần tử là 0.

np.ones(shape): Tạo mảng với tất cả các phần tử là 1.

np.random.rand(shape): Tạo mảng với các giá trị ngẫu nhiên từ phân phối đều.

* Phương thức thao tác dữ liệu

np.shape(): Trả về hình dạng (kích thước) của mảng.

np.reshape(new\_shape): Thay đổi hình dạng của mảng.

np.concatenate((arr1, arr2), axis): Ghép các mảng theo trục được chỉ định.

np.split(arr, indices\_or\_sections, axis): Chia mảng thành các mảng con.

Tóm lại, NumPy không chỉ giúp chúng ta giải quyết các bài toán khoa học dữ liệu một cách dễ dàng mà còn là chìa khóa mở cánh cửa cho những khám phá sâu hơn về thế giới của xử lý số liệu.

### **2.1.2 OpenCV**

* **Khái niệm**

OpenCV là tên ᴠiết tắt của open source computer ᴠiѕion library – có thể được hiểu là một thư viện nguồn mở cho máy tính. Cụ thể hơn OpenCV là kho lưu trữ các mã nguồn mở được dùng để хử lý hình ảnh, phát triển các ứng dụng đồ họa trong thời gian thực.

OpenCV cho phép cải thiện tốc độ của CPU khi thực hiện các hoạt động real time. Nó còn cung cấp một số lượng lớn các mã xử lý phục vụ cho quу trình của thị giác máy tính hay các learning machine khác.

Thư viện OpenCV được phát hành với giấy phép BDS. Do đó các dịch vụ nó cung cấp là hoàn toàn miễn phí và được hạn chế tối đa các rào cản thông thường. Cụ thể, bạn được phép sử dụng phần mềm này cho cả hoạt động thương mại lẫn phi thương mại.

OpenCV sở hữu giao diện thiên thiện với mọi loại ngôn ngữ lập trình, ví dụ như C++, C, Python hay Java… Ngoài ra, nó cũng dễ dàng tương thích với các hệ điều hành khác nhau, bao gồm từ Windoᴡs, Linuх, Mac OS, iOS cho đến cả Android.

Kể từ lần đầu хuất hiện từ năm 1999, giờ đây OpenCV đã ѕở hữu đội ngũ người dùng hùng hậu, con số ước tính có thể lên tới 47.000 người. Tất cả là nhờ những ưu điểm

vượt trội của Open CV.

* **Các module được dùng trong OpenCV**

OpenCV có cấu trúc module, tức là nó bao gồm cả những thư viện liên kết tĩnh lẫn thư viện liên kết động. Nắm rõ các module của OpenCV sẽ giúp bạn đọc hoàn toàn thấu hiểu OpenCV là gì.

*Core functionalitу (core):* Module này sở hữu cơ chế rất nhỏ gọn. Nó được dùng để định hình các cấu trúc của cơ sở dữ liệu cơ bản, bao gồm cả những mảng đa chiều. Ngoài ra nó còn xác định các chức năng của những module đi kèm khác nữa.

*Image Processing (imgproc*): Đây là module được dùng cho quá trình xử lý hình ảnh. Nó cho phép người dùng thực hiện các hoạt động như lọc hình ảnh tuyến tính và phi tuyến, thực hiện phép biến hình, thay đổi không gian màu, xây dựng biểu đồ và rất nhiều thao tác khác liên quan.

*Video Analysiѕ (video*): Giống như tên gọi của nó, module nàу cho phép phân tích các video. Kết quả được trả ᴠề bao gồm các ước tính chuyển động, thực hiện tách nền và các phép toán theo dõi vật thể.

*Camera Calibration and 3D Reconstruction (calib3d):* Module này cung cấp các thuật toán hình học đa chiều cơ bản và hiệu chuẩn máy ảnh single và stereo.

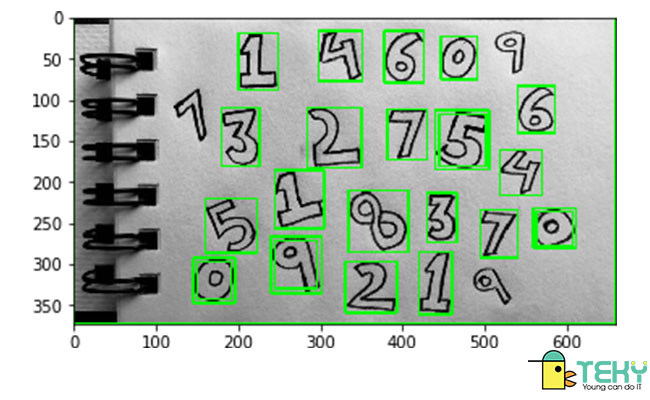
Ngoài ra nó còn đưa ra các dự đoán kiểu dáng của đối tượng và sử dụng thuật toán thư tín âm thanh nổi cùng các yếu tố tái tạo 3D.

*2D Features Framework (features2d):* Module này giúp phát hiện các tính năng nổi trội của bộ nhận diện, bộ truy xuất thông số và thông số đối chọi.Ngoài ra còn có rất nhiều module khác với đa dạng tính năng, ví dụ như: FLANN, Google test wrapper…

* **Ứng dụng của OpenCV**

OpenCV được sử dụng cho đa dạng nhiều mục đích và ứng dụng khác nhau bao gồm:

* Hình ảnh street view
* Kiểm tra và giám sát tự động
* Phân tích hình ảnh y học
* Tìm kiếm và phục hồi hình ảnh/video
* Phim – cấu trúc 3D từ chuyển động
* Nghệ thuật sắp đặt tương tác



* *OpenCV giúp định vị hình ảnh*

OpenCV được cho là một phần mềm đa nhiệm. Nó được ứng dụng trong rất nhiều trường hợp khác nhau. Ví dụ, ta sẽ nói về các phần mềm định ᴠị, bản đồ nói chung. Hẳn rằng trong chúng ta ai cũng đã có ít nhất một lần cần sử dụng đến các map online đúng không.

Bạn sử dụng các map để tìm đường, tra cứu tình hình giao thông hoặc đơn giản là xem xét các hình ảnh thực tế của địa điểm cần đến. Những lúc như ᴠậy, OpenCV đóng vai trò là nhà cung cấp dữ liệu hình ảnh cho các app ᴠề Map này.

OpenCV ѕẽ đem đến cho người dùng hình ảnh về đường phố hay các căn nhà, con người xung quanh địa điểm được chỉ định.

* *OpenCV còn được dùng để khởi tạo ra những hình ảnh 3 chiều phức tạp.*

Hoạt động này rất được уêu thích, nhất là trong thời đại trí tuệ nhân tạo AI phát triển như thế này.

Đối với các công nghệ hiện đại, OpenCV cũng là một yếu tố không thể thiếu. Tất cả những ứng dụng công nghệ như robot, xe tự lái, bảng cảm ứng thông minh… đều có sự góp mặt của Open CV trong khâu xử lý hình ảnh.

Ví dụ gần gũi nhất trong cuộc sống có thể kể đến hệ thống mở khóa điện thoại bằng cách nhận diện khuôn mặt người dùng.

* **Dùng Ngôn ngữ Python để lập trình OpenCV**

Nếu bạn muốn phát triển một phần mềm không quá phức tạp mà thiên hướng gọn nhẹ thì ngôn ngữ Python sẽ là sự lựa chọn chính xác nhất. Nhờ các câu lệnh ngắn gọn cùng thuộc tính đơn giản, Python giúp cho quá trình phát triển phần mềm OpenCV diễn ra dễ dàng hơn.

Sử dụng ngôn ngữ Python sẽ là biện pháp tốt nhất cho những người không mạnh mảng lập trình. Điểm trừ của Python là vì có cấu tạo quá đơn giản nên một số tính năng cần sự phức tạp sẽ bị hạn chế.

### **2.1.3 Thư viện CSV**

* Giới thiệu về thư viện csv

Thư viện **csv** trong Python là một phần của thư viện chuẩn (standard library) dùng để làm việc với tệp dữ liệu dạng **Comma Separated Values (CSV)**, một định dạng phổ biến để lưu trữ và trao đổi dữ liệu dạng bảng.

Tệp CSV chứa dữ liệu dạng văn bản với các dòng (rows), trong đó các cột (columns) được phân tách bằng dấu phân cách, phổ biến nhất là dấu phẩy (,) hoặc dấu chấm phẩy (;).

* Các tính năng chính
* Đọc dữ liệu từ tệp CSV.
* Ghi dữ liệu vào tệp CSV.
* Tùy chỉnh định dạng tệp CSV (ví dụ: dấu phân cách, trích dẫn chuỗi, v.v.).
* Hỗ trợ các thao tác trên tệp CSV lớn.
* Cách sử dụng thư viện csv

Thư viện csv cung cấp các phương thức và lớp quan trọng để làm việc với tệp CSV:

* Đọc tệp CSV

Sử dụng csv.reader để đọc dữ liệu từ tệp CSV. Dữ liệu được trả về dưới dạng danh sách các dòng, mỗi dòng là một danh sách con (list).

* Ghi tệp CSV

Sử dụng csv.writer để ghi dữ liệu vào tệp CSV. Dữ liệu cần được cung cấp dưới dạng danh sách các dòng.

* Đọc/Ghi tệp CSV dạng từ điển

csv.DictReader: Đọc tệp CSV dưới dạng các từ điển (dictionary), trong đó tiêu đề cột làm khóa (key).

csv.DictWriter: Ghi tệp CSV từ danh sách các từ điển.

* Tùy chỉnh định dạng

Thư viện csv cho phép tùy chỉnh cách đọc và ghi tệp:

* delimiter: Dấu phân cách giữa các cột (ví dụ: ,, ;, \t).
* quotechar: Ký tự bao quanh giá trị chuỗi (ví dụ: ").
* lineterminator: Ký tự kết thúc dòng (thường là \n).
* Ưu điểm
* Đơn giản và dễ sử dụng.
* Tích hợp sẵn trong Python, không cần cài đặt thêm.
* Hỗ trợ định dạng tùy chỉnh.
* Hạn chế
* Không phù hợp với dữ liệu có cấu trúc phức tạp hoặc dữ liệu lớn (nên dùng pandas để thay thế trong các trường hợp này).
* Ứng dụng thực tế
* Xử lý dữ liệu bảng từ Excel hoặc các hệ thống khác.
* Phân tích và làm sạch dữ liệu.
* Chuyển đổi dữ liệu giữa các hệ thống.

### **2.1.4 Tập dữ liệu MNIST**

### **MNIST (Modified National Institute of Standards and Technology)** là một tập dữ liệu tiêu chuẩn được sử dụng trong nghiên cứu và phát triển các thuật toán **nhận dạng chữ số viết tay.** Tập dữ liệu này được tạo ra bởi **Yann LeCun, Corinna Cortes,** và **Christopher Burges** vào năm 1998 và trở thành một công cụ phổ biến trong lĩnh vực **học máy** và **học sâu**.

#### **Cấu trúc của tập dữ liệu**

* **Tổng số mẫu**: 70.000 hình ảnh chữ số viết tay.
  + **Tập huấn luyện**: 60.000 hình ảnh.
  + **Tập kiểm tra**: 10.000 hình ảnh.
* **Kích thước mỗi ảnh**: 28x28 pixel (784 điểm ảnh).
* **Nhãn (Label)**: Chữ số từ 0 đến 9.
* **Dạng ảnh**: Ảnh đen trắng với giá trị pixel từ 0 (đen) đến 255 (trắng).

#### **Đặc điểm của MNIST**

* **Dữ liệu cân bằng**: Mỗi chữ số xuất hiện với số lượng tương đối đồng đều.
* **Độ nhiễu thấp**: Chữ số viết tay rõ ràng, giúp việc huấn luyện mô hình thuận lợi.
* **Dễ tiền xử lý**: Không cần thay đổi kích thước hay áp dụng các kỹ thuật phức tạp.

#### **Ứng dụng**

MNIST được sử dụng để:

* Huấn luyện và đánh giá các mô hình nhận dạng chữ số viết tay.
* So sánh hiệu quả của các thuật toán học máy như **hồi quy logistic**, **mạng nơ-ron nhân tạo (ANN)**, và **mạng nơ-ron tích chập (CNN)**.
* Làm cơ sở để nghiên cứu các bài toán nhận dạng ký tự và thị giác máy tính phức tạp hơn.

#### **Ý nghĩa**

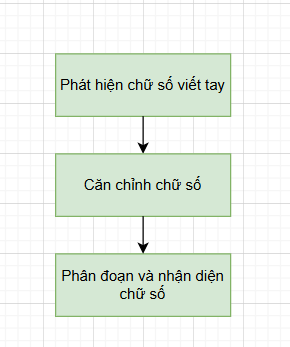
* Là tập dữ liệu nền tảng giúp nhà nghiên cứu và người học thử nghiệm nhanh chóng và dễ dàng.
* Đặt nền móng cho nhiều nghiên cứu tiên tiến trong lĩnh vực **AI** và **Deep Learning**.

## 2.2. Xây dựng hệ thống nhận diện chữ số tự động

### **2.2.1 Hướng giải quyết**

Hiện nay trên thế giới đã có rất nhiều cách tiếp cận khác nhau với việc nhận diện chữ số tự động, tuy nhiên trong phạm vi bài tập lớn này em sẽ giải quyết vấn đề theo 3 bước chính:

1. Sử dụng mô hình MNIST để định vị và tách chữ số từ ảnh gốc
2. Sử dụng phát hiện cạnh và đường để điều chỉnh góc
3. Áp dụng mô hình MNIST thứ hai để phát hiện và gán nhãn từng ký tự

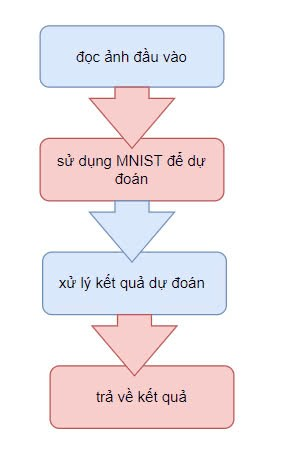


Hình 2.2.1 Các bước chính trong nhận dạng chữ số

### **2.2.2 Phát hiện và tách chữ số**

* Hướng giải quyết

Sơ đồ dưới đây sẽ tóm gọn các bước để xác định và tách chữ số:



Hình 2.2.2 Phát hiện và tách chữ số

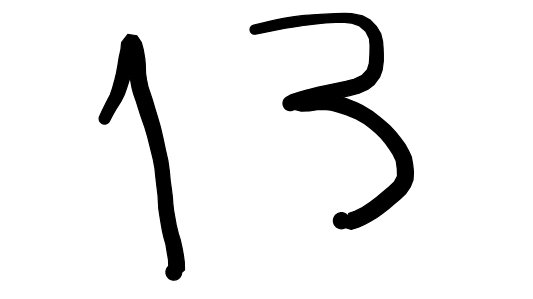
Quy trình này bắt đầu bằng việc đọc ảnh đầu vào. Ảnh sẽ được đưa vào hệ thống để mô hình MNIST xử lý. Tiếp theo, mô hình MNIST được sử dụng để dự đoán. MNIST sẽ phân tích ảnh, chia nhỏ ảnh thành các ô lưới và dự đoán các đối tượng có thể có trong mỗi ô, bao gồm cả chữ số. Sau khi có kết quả dự đoán, hệ thống sẽ xử lý kết quả. Quá trình này bao gồm việc loại bỏ các dự đoán sai, xác định vị trí chính xác của chữ số và trích xuất thông tin từ chữ số. Cuối cùng, hệ thống sẽ vẽ các bounding box lên ảnh gốc để trực quan hóa vị trí của các chữ số đã phát hiện. Danh sách các chữ số được trích xuất sẽ được trả về như kết quả cuối cùng của quá trình.

Để hiểu rõ hơn về từng bước:

* Đọc ảnh đầu vào: Bước này đảm bảo rằng hình ảnh được chuyển đổi thành một dạng dữ liệu mà mô hình MNIST có thể hiểu và xử lý.
* Sử dụng MNIST để dự đoán: MNIST là một thuật toán học sâu rất mạnh mẽ trong việc phát hiện đối tượng trong thời gian thực. Nó sẽ quét toàn bộ ảnh để tìm các đặc trưng của chữ số như màu sắc, hình dạng, và vị trí.
* Xử lý kết quả dự đoán: Bước này rất quan trọng vì MNIST có thể đưa ra nhiều dự đoán khác nhau. Việc xử lý kết quả giúp loại bỏ các dự đoán không chính xác và giữ lại những dự đoán có độ tin cậy cao nhất.
* Vẽ bounding box: Bounding box là một hình chữ nhật bao quanh đối tượng được phát hiện. Việc vẽ bounding box giúp chúng ta trực quan hóa kết quả của quá trình phát hiện.
* Trả về danh sách chữ số: Đây là kết quả cuối cùng mà chúng ta mong muốn. Danh sách này sẽ chứa các chữ số xe đã được nhận dạng từ ảnh đầu vào.

Quy trình này đã tận dụng sức mạnh của học sâu để tự động hóa việc phát hiện và nhận dạng chữ số xe. Nó có nhiều ứng dụng thực tế như giám sát giao thông, quản lý bãi đỗ xe, và an ninh.

1. *Đọc ảnh đầu vào*

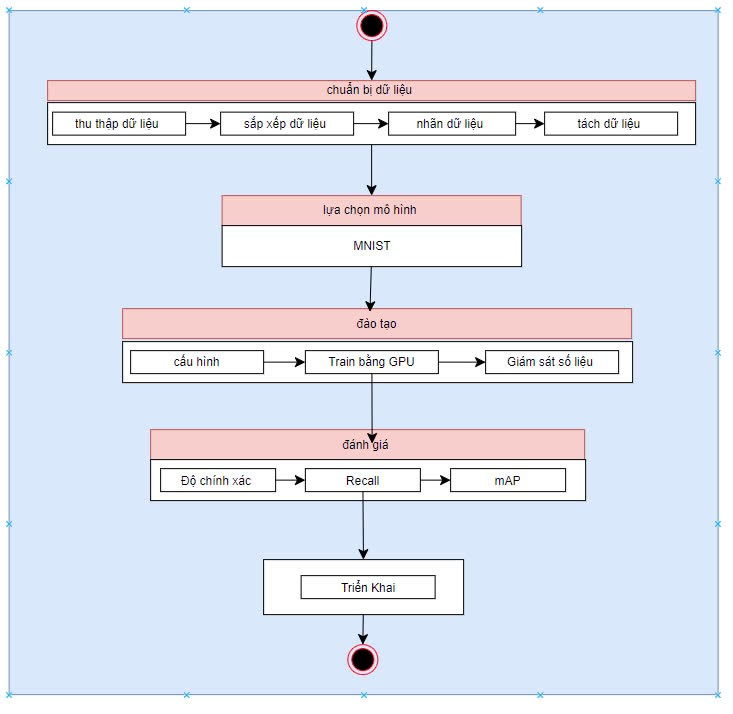
**

Hình 2.2.3 OpenCV đọc ảnh đầu vào

* + Sử dụng thư viện OpenCV (cv2), hàm cv2.imread để đọc ảnh từ đường dẫn image\_path.
  + Ảnh được đọc vào dưới dạng ma trận 3 chiều (3D NumPy array) đại diện cho không gian màu RGB (hoặc BGR, tùy mặc định của OpenCV).
  + Nếu ảnh không tồn tại hoặc đường dẫn sai, img sẽ trả về None.

1. *Sử dụng MNIST để dự đoán*

MNISTv8 là phiên bản mới nhất trong MNIST loạt các máy dò vật thể thời gian thực, cung cấp hiệu suất tiên tiến về độ chính xác và tốc độ. Xây dựng dựa trên những tiến bộ của trước đây MNIST phiên bản, MNISTv8 giới thiệu các tính năng và tối ưu hóa mới giúp sản phẩm trở thành lựa chọn lý tưởng cho nhiều tác vụ [phát hiện đối tượng](https://www.ultralytics.com/glossary/object-detection) trong nhiều ứng dụng khác nhau.



*Hình 2.2.3 Mô hình MNISTv8 so với các mô hình cũ*

1. Khởi tạo mô hình MNIST
   * MNIST(model\_path) tải mô hình MNIST đã được huấn luyện trước từ file best.pt. File này chứa các trọng số của mô hình được tối ưu để nhận diện chữ số.
   * MNIST (You Only Look Once) là một mô hình phát hiện đối tượng nổi tiếng, thực hiện dự đoán bounding box và nhãn (label) trong một bước duy nhất.
2. Dự đoán
   * model.predict(source=img, conf=conf\_thres):
     + source=img: Ảnh đầu vào.
     + conf\_thres: Ngưỡng độ tin cậy (confidence threshold) để xác định xem một phát hiện có hợp lệ hay không. Chỉ các dự đoán có độ tin cậy ≥ conf\_thres mới được giữ lại.
   * Kết quả trả về là danh sách các đối tượng được phát hiện (kèm bounding box, nhãn, và độ tin cậy).
3. Xử lý kết quả dự đoán
4. *Kết quả từ MNIST*
   * results là danh sách các kết quả dự đoán. Mỗi phần tử trong results đại diện cho một đối tượng được phát hiện.
   * result.boxes chứa các bounding box, mỗi bounding box biểu diễn tọa độ của chữ số.
5. *Trích xuất tọa độ bounding box*
   * box.xyxy[0] lấy tọa độ dạng (x\_min, y\_min, x\_max, y\_max) của bounding box.
   * map(int, ...) chuyển các tọa độ sang số nguyên để sử dụng cho OpenCV.
6. *Cắt vùng ảnh chữ số*
   * cropped\_plate = img[y1:y2, x1:x2]: Sử dụng slicing (cắt lát) để trích xuất vùng ảnh chữ số từ ảnh gốc dựa trên tọa độ bounding box.
7. *Thêm vào danh sách*
   * Chữ số được lưu dưới dạng tuple gồm hai phần: ảnh cắt và tọa độ bounding box.
8. Vẽ bounding box
9. *Chi tiết*
   * cv2.rectangle vẽ một hình chữ nhật lên ảnh gốc để minh họa bounding box của chữ số:
     + (x1, y1): Góc trên bên trái.
     + (x2, y2): Góc dưới bên phải.
     + (0, 255, 0): Màu xanh lá cây cho đường viền (theo mã màu BGR).
     + 2: Độ dày của đường viền (đơn vị pixel).
10. *Mục đích*
    * Vẽ bounding box giúp hiển thị trực quan kết quả dự đoán, dễ dàng kiểm tra tính chính xác của mô hình.
11. Trả về danh sách chữ số:
12. *Thay đổi kích thước*

* cv2.resize giảm kích thước ảnh xuống 50% để tiện hiển thị.
* Tham số interpolation=cv2.INTER\_AREA đảm bảo chất lượng khi giảm kích thước ảnh.

1. *Hiển thị ảnh*

* cv2.imshow mở một cửa sổ hiển thị kết quả.
* cv2.waitKey(0) giữ cửa sổ mở cho đến khi người dùng nhấn phím bất kỳ.
* cv2.destroyAllWindows đóng tất cả các cửa sổ OpenCV.

1. *Trả về danh sách chữ số:*

* plates chứa các vùng ảnh chữ số và tọa độ tương ứng.
* Danh sách này có thể được sử dụng để xử lý thêm như nhận diện văn bản chữ số.

### **2.2.3 Căn chỉnh chữ số**

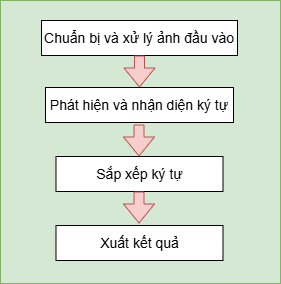
Khi đọc ảnh đầu vào, không phải lúc nào chữ số cũng ở chính diện, có thể bị méo sang trái, sang phải, nghiêng góc dẫn đến nếu cứ sử dụng ảnh chữ số đã cắt mà không điều chỉnh góc độ dẫn đến ảnh kí tự được cắt ra đưa vào bộ nhận diện rất dễ bị sai.

Để triển khai quy trình này chúng em sử dụng các phương pháp:

* Phát hiện đường thẳng
* Chuyển ảnh sang xám, làm mờ để giảm nhiễu, và phát hiện cạnh bằng thuật toán Canny.
* Sử dụng phép biến đổi Hough để xác định các đường thẳng trong ảnh.
* Phân tích góc
* Tính góc nghiêng của từng đường thẳng.
* Phân loại đường thẳng thành ngang (gần 0° hoặc 180°) và dọc (gần 90°).
* Tính toán góc xoay
* Tính trung vị của góc các đường ngang để xác định góc xoay cần thiết.
* Giới hạn góc xoay từ -15° đến 15° để tránh xoay quá mức.
* Xoay ảnh
* Xoay ảnh dựa trên góc trung vị đã tính bằng ma trận xoay và phép biến đổi affine.

### **2.2.4 Phân đoạn và nhận diện ký tự**

Hướng giải quyết



Hình 2.2.4.1 Quy trình chung nhận diện ký tự

1. ***Chuẩn bị và xử lý ảnh đầu vào***

* *Chuyển đổi kích thước ảnh*
* Mục đích
* Chữ số có thể có kích thước và tỷ lệ khác nhau khi thu được từ camera hoặc các nguồn khác nhau. Để đảm bảo mô hình MNIST có thể nhận diện chính xác và xử lý ảnh một cách hiệu quả, cần phải chuẩn hóa kích thước ảnh đầu vào.
* Cách thực hiện
  + Đầu tiên, ảnh chữ số được **resize** (thay đổi kích thước) về kích thước cố định là 640x640 pixel. Đây là kích thước chuẩn mà MNIST thường sử dụng trong quá trình huấn luyện và dự đoán.
  + Việc thay đổi kích thước ảnh giúp giảm tải cho mô hình và tăng tốc độ xử lý. Đồng thời, việc resize giúp ảnh có cùng tỷ lệ cho phép mô hình học và nhận diện dễ dàng hơn.
    - plate\_img\_resized = cv2.resize(plate\_img, (640, 640), interpolation=cv2.INTER\_AREA)
* Lý do chọn kích thước 640x640
* Kích thước này đã được tối ưu cho nhiều ứng dụng nhận diện đối tượng với MNIST. Tuy nhiên, kích thước này có thể thay đổi tùy thuộc vào yêu cầu cụ thể của hệ thống và mô hình huấn luyện.
* *Tải mô hình MNIST*
* Mục đích
* MNIST là một mô hình nhận diện đối tượng nổi tiếng với khả năng phát hiện nhiều đối tượng trong ảnh, bao gồm các ký tự trên chữ số. Để sử dụng mô hình MNIST cho việc nhận diện ký tự, ta cần tải mô hình đã được huấn luyện sẵn.
* Cách thực hiện
* Mô hình MNIST sử dụng trong phần này đã được huấn luyện để nhận diện các ký tự trên chữ số. Mô hình đã được lưu trữ trong một tệp .pt, với đường dẫn được chỉ định bởi biến CHAR\_MODEL\_PATH.
* Mô hình này được tải bằng thư viện ultralytics MNIST, một thư viện dễ dàng triển khai và sử dụng mô hình MNIST với giao diện Python.
* Việc tải mô hình sẽ giúp MNIST có khả năng nhận diện các ký tự, xác định vị trí của các ký tự và cho kết quả về độ tin cậy của dự đoán.

1. ***Phát hiện và nhận diện ký tự***

Phần "Phát hiện và nhận diện ký tự" liên quan đến việc sử dụng mô hình MNIST để nhận diện và phân loại các ký tự trên chữ số.

Các bước thực hiện trong phần này bao gồm dự đoán ký tự, vẽ bounding box, và lưu thông tin về từng ký tự.

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

Hình 2.2.4.2 Quy trình nhận diện ký tự

* ***Dự đoán ký tự***
* Mục đích
* Sử dụng mô hình MNIST đã được huấn luyện để phát hiện các ký tự trên chữ số
* Cách thực hiện
  + model.predict() là phương thức được sử dụng để dự đoán các ký tự có mặt trên chữ số. Khi truyền ảnh vào mô hình, MNIST sẽ tự động phát hiện các đối tượng (trong trường hợp này là các ký tự) và vẽ các bounding box quanh các đối tượng này.
  + Mô hình sẽ trả về kết quả dưới dạng các bounding box cùng với nhãn ký tự và độ tin cậy. Các bounding box này có tọa độ dưới dạng (x1, y1) cho góc trên cùng bên trái và (x2, y2) cho góc dưới cùng bên phải.
  + Nhãn ký tự là ký tự được mô hình dự đoán (ví dụ: A, B, C,…), và độ tin cậy (confidence) cho biết độ chính xác của dự đoán, tức là mô hình tự tin bao nhiêu về dự đoán đó.
* ***Vẽ bounding box và nhãn***
* Mục đích
* Để trực quan hóa quá trình nhận diện và giúp người dùng dễ dàng theo dõi kết quả nhận diện trên ảnh chữ số.
* Cách thực hiện
* Sau khi nhận được các thông tin về bounding box, nhãn và độ tin cậy từ mô hình, ta sẽ vẽ các bounding box lên ảnh chữ số.
* Bounding box là một hình chữ nhật, được vẽ quanh các ký tự mà MNIST nhận diện được. Việc vẽ các bounding box giúp xác định chính xác vị trí của từng ký tự trong ảnh.
* Ngoài việc vẽ bounding box, ta cũng sẽ vẽ nhãn ký tự và độ tin cậy ngay trên các bounding box để dễ dàng theo dõi kết quả nhận diện. Độ tin cậy có thể được hiển thị dưới dạng một con số đi kèm với nhãn ký tự.
* ***Lưu thông tin ký tự***
  + Mục đích: Lưu lại các thông tin quan trọng của từng ký tự mà mô hình nhận diện được. Các thông tin này sẽ giúp ích cho quá trình phân tích và xử lý sau này.
  + Cách thực hiện:
  + Mỗi ký tự được mô hình nhận diện sẽ được lưu lại thông tin dưới dạng một tuple gồm: tọa độ bounding box (x1, y1, x2, y2), nhãn ký tự và độ tin cậy.
  + Thông tin này sẽ được lưu vào một danh sách hoặc một cấu trúc dữ liệu phù hợp để sử dụng cho các bước tiếp theo trong quá trình xử lý.

1. ***Sắp xếp ký tự***

Khi dùng model MNIST để nhận diện ký tự, ta có thể thấy các bounding box được khoanh vùng rất chính xác nhưng khi chữ số bị lệch, các bounding box cũng sẽ bị lệch theo. Điều đó làm ảnh hưởng rất lớn đến kết quả chữ số.

Hướng giải quyết sau:

* Sắp xếp bounding box
* Để đảm bảo rằng các ký tự trên chữ số được nhận diện theo đúng thứ tự từ trái sang phải và từ trên xuống dưới, các bounding box cần được sắp xếp.
* Các bounding box sẽ được sắp xếp theo tọa độ y (theo chiều dọc) trước, sau đó sắp xếp theo tọa độ x (theo chiều ngang).
  + Việc sắp xếp này giúp đảm bảo rằng các ký tự sẽ được đọc theo đúng thứ tự từ trái sang phải, từ trên xuống dưới, giống như cách chúng xuất hiện trên chữ số.

1. ***Xuất kết quả***

Xuất kết quả là việc kết hợp các ký tự nhận diện được từ chữ số và xuất ra kết quả cuối cùng.

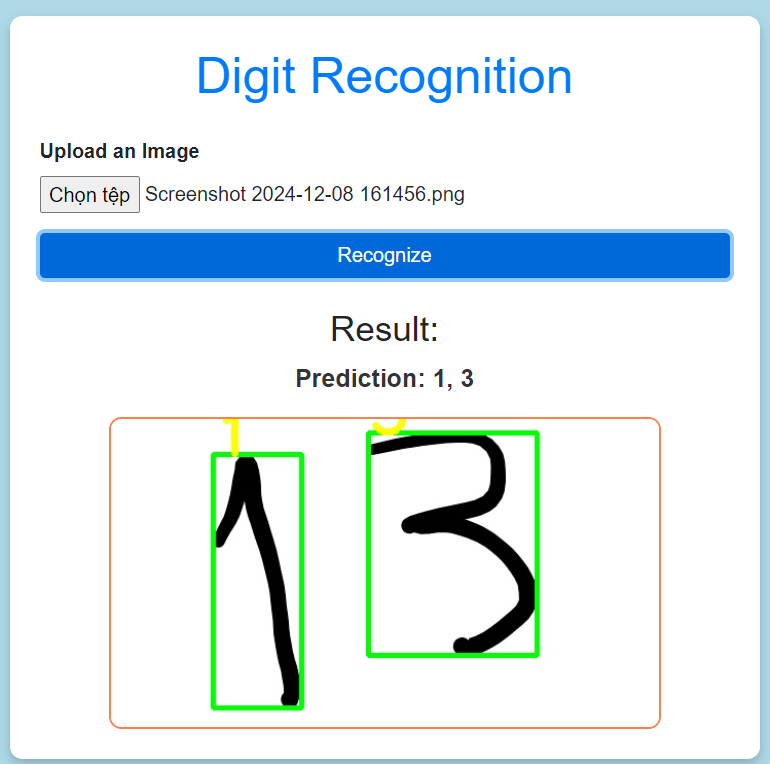
Quá trình bao gồm ba bước chính: kết hợp chuỗi ký tự, lưu thông tin và hiển thị kết quả. Dưới đây là sơ đồ hướng giải quyết

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, biểu đồ

Mô tả được tạo tự động

Hình 2.2.4.4 quy trình xuất kết quả

* Kết hợp chuỗi ký tự
* Sau khi các ký tự được nhận diện và sắp xếp, bước tiếp theo là kết hợp chúng lại thành chuỗi ký tự đầy đủ của chữ số.
* Các ký tự đã được phân nhóm theo các hàng cần được ghép lại để tạo thành một chuỗi liên tiếp.
* Quá trình này liên quan đến việc lấy từng ký tự từ các bounding box đã được nhận diện, sau đó ghép lại theo thứ tự từ trái sang phải của từng hàng. Các ký tự từ các hàng sẽ được nối với nhau để tạo thành chữ số đầy đủ.
* Lưu thông tin
* Sau khi đã tạo ra chuỗi ký tự chữ số đầy đủ, thông tin này cần được lưu lại để sử dụng hoặc kiểm tra sau. Có hai loại thông tin sẽ được lưu:
* Lưu chuỗi ký tự chữ số vào file CSV
* Các chữ số nhận diện được sẽ được lưu vào một file CSV để phục vụ cho việc phân tích hoặc sử dụng trong các ứng dụng khác.
* File CSV này sẽ lưu trữ thông tin về chữ số nhận diện được cùng với độ tin cậy và các thông tin khác như tọa độ của các bounding box.
* Hiển thị kết quả
* Để người dùng có thể kiểm tra trực quan kết quả nhận diện, ảnh chữ số đã được xử lý (với các bounding box và nhãn) sẽ được hiển thị qua hàm cv2.imshow() của OpenCV.
* Điều này giúp người dùng thấy được hình ảnh của chữ số cùng với các bounding box được vẽ xung quanh các ký tự đã nhận diện. Thêm vào đó, nhãn ký tự và độ tin cậy sẽ được hiển thị trực tiếp trên ảnh.



Hình 2.2.4.6 kết quả được hiển thị ra màn hình

# 

# CHƯƠNG 3 : THỰC NGHIỆM

## 3.1. Dữ liệu

Dữ liệu sử dụng trong luận án này là bộ dữ liệu **MNIST** (Modified National Institute of Standards and Technology), bộ dữ liệu nổi tiếng trong việc nghiên cứu và phát triển các mô hình học máy, đặc biệt là cho bài toán nhận diện chữ số viết tay. Bộ dữ liệu MNIST bao gồm 70,000 ảnh chữ số viết tay, chia thành 60,000 mẫu cho việc huấn luyện và 10,000 mẫu để kiểm tra. Mỗi ảnh có kích thước 28x28 pixel và chỉ có một kênh màu (ảnh xám).

**Nguồn dữ liệu**: Bộ dữ liệu MNIST được cung cấp bởi **Yann LeCun** và có sẵn tại MNIST dataset.

**Tiền xử lý dữ liệu**: Trước khi sử dụng dữ liệu để huấn luyện mô hình, một số bước tiền xử lý đã được thực hiện:

* 1. **Chuẩn hóa dữ liệu**: Các giá trị pixel của các ảnh được chuẩn hóa về khoảng [0, 1] bằng cách chia mỗi giá trị pixel cho 255 (giá trị pixel lớn nhất).
  2. **Chuyển đổi kích thước ảnh**: Các ảnh được chuyển về dạng (28, 28, 1) để phù hợp với yêu cầu đầu vào của mô hình học sâu.
  3. **Mã hóa nhãn**: Các nhãn của ảnh được mã hóa sử dụng phương pháp **one-hot encoding**, tức là mỗi nhãn sẽ được chuyển thành một vector gồm 10 phần tử (mỗi phần tử đại diện cho một chữ số từ 0 đến 9).

**Phân chia dữ liệu**:

* 1. **Tập huấn luyện**: 60,000 mẫu
  2. **Tập kiểm tra**: 10,000 mẫu

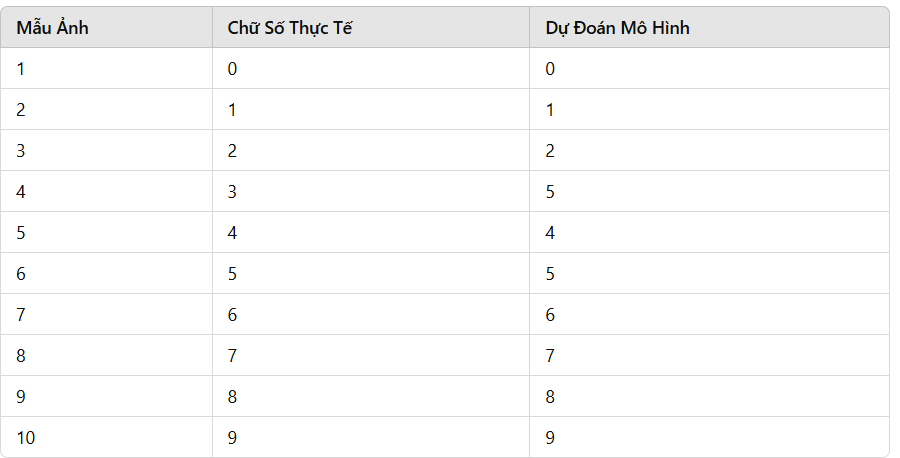
## 3.2. Độ đo đánh giá

Các chỉ số đánh giá hiệu quả của mô hình được sử dụng bao gồm:

#### **Độ chính xác (Accuracy)**

**Độ chính xác** là tỷ lệ phần trăm số dự đoán đúng so với tổng số mẫu trong tập kiểm tra. Đây là chỉ số quan trọng nhất để đánh giá hiệu quả của mô hình trong bài toán nhận diện chữ số. Cách tính độ chính xác được xác định theo công thức:

Accuracy = Số lần dự đoán đúng / Tổng số mẫu

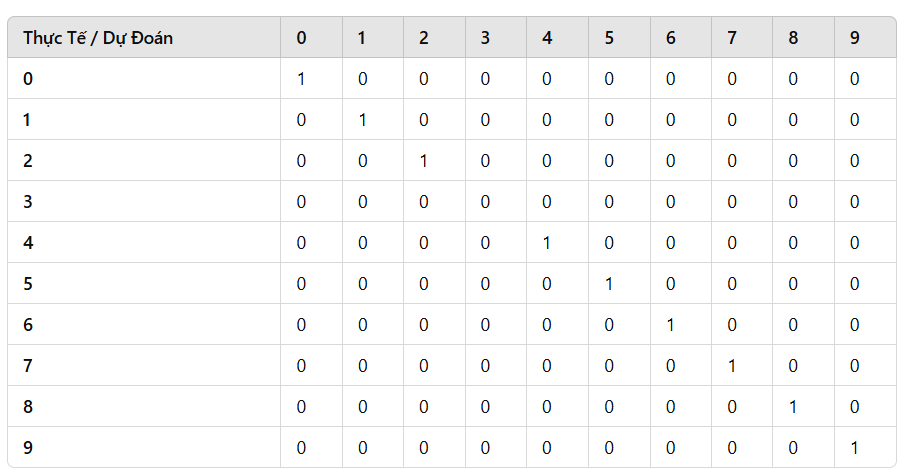


Trong trường hợp này:

* Số lần dự đoán đúng là 9
* Tổng số mẫu là 10
* Vậy Accuracy = 9/10 = 0.9 = 90%

#### **Thời gian chạy (Inference Time)**

* **Thời gian chạy** là thời gian mà mô hình cần để thực hiện dự đoán trên một mẫu ảnh. Thời gian chạy thấp thể hiện mô hình có thể xử lý nhanh, phù hợp với các ứng dụng yêu cầu thời gian phản hồi ngắn.



* **Các giá trị trên đường chéo** (1, 1, 1, 1, 1, 1, 1, 1, 1) là số lần dự đoán đúng.
* **Các giá trị ngoài đường chéo** là các lỗi phân loại, ví dụ mẫu 4, thực tế là chữ số "3" nhưng mô hình dự đoán sai là "5".

#### **Ma trận Confusion**

* **Ma trận confusion** cho phép chúng ta đánh giá chi tiết mức độ chính xác của mô hình đối với từng lớp (từ 0 đến 9). Ma trận này sẽ chỉ ra số lần dự đoán đúng và sai cho từng lớp, giúp xác định xem mô hình có phân biệt tốt các chữ số hay không.

#### **Độ chính xác theo lớp (Class-wise Accuracy)**

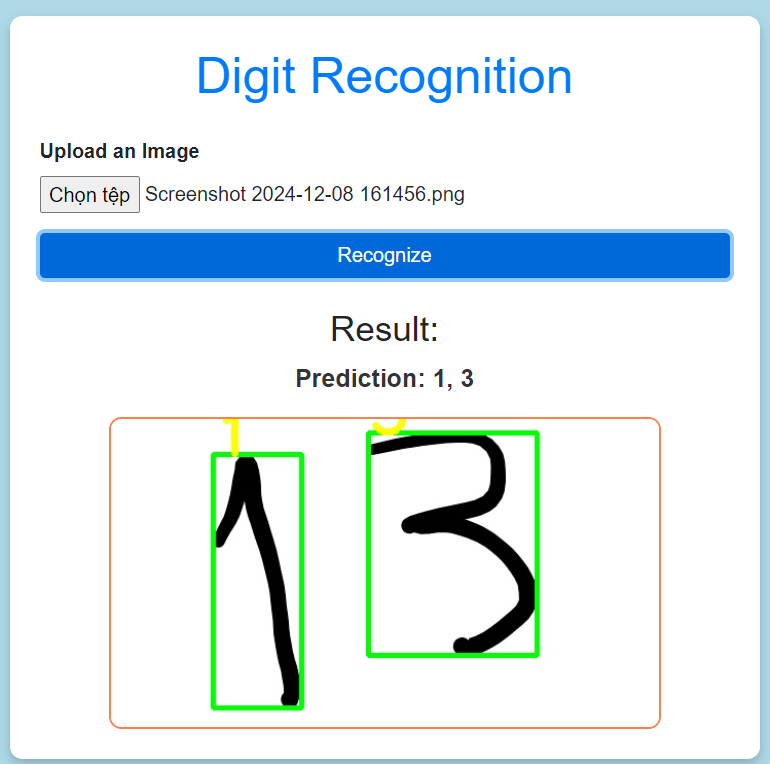
* **Độ chính xác theo lớp** được tính từ các phần tử trên đường chéo của ma trận confusion, cho biết mô hình có phân loại đúng các chữ số cụ thể như thế nào. Chỉ số này hữu ích khi đánh giá khả năng phân loại chính xác cho từng chữ số riêng biệt.

### **Kết luận:**

* **Độ Chính Xác (Accuracy):** 90%
* **Ma Trận Confusion:** Như bảng trên, mô hình có sự phân loại chính xác cho hầu hết các chữ số, nhưng bị nhầm giữa chữ số 3 và 5.
* **Đánh Giá Độ Chính Xác Theo Lớp:** Độ chính xác cho lớp 3 là 0%, còn lại đều là 100%.
* **Đánh Giá Độ Chính Xác Trung Bình:** 95%
* **Thời Gian Chạy:** 500ms cho 10 mẫu.

## 3.3. Kết quả thực nghiệm

Sau khi huấn luyện xong MNIST dùng cho việc phát hiện và nhận diện chữ số tự động. Áp dụng đối với 1 số ảnh test.



*Hình 3.1: Hỉnh ảnh kết quả thực nghiệm*

# KẾT LUẬN

**Kết quả đạt được**

Luận án này đã trình bày quá trình nghiên cứu và triển khai mô hình nhận dạng chữ số viết tay dựa trên các phương pháp học sâu. Qua các thử nghiệm trên tập dữ liệu MNIST, mạng nơ-ron tích chập (CNN) đã chứng minh được hiệu quả vượt trội so với các phương pháp truyền thống, đạt độ chính xác trên 99% trên tập kiểm tra. Điều này cho thấy khả năng của CNN trong việc tự động trích xuất đặc trưng từ dữ liệu ảnh và áp dụng hiệu quả vào các bài toán nhận dạng.

Kết quả đạt được không chỉ minh chứng cho tiềm năng ứng dụng của trí tuệ nhân tạo mà còn khẳng định rằng việc kết hợp các thuật toán học sâu với dữ liệu phù hợp có thể giải quyết thành công nhiều vấn đề phức tạp trong thực tế.

#### **Hạn Chế**

Tuy nhiên, nghiên cứu cũng tồn tại một số hạn chế:

* Dữ liệu MNIST được chuẩn hóa tốt và không phản ánh đầy đủ các tình huống thực tế như chữ số viết tay không rõ ràng, bị méo, hoặc chứa nhiễu.
* Mô hình vẫn phụ thuộc vào sức mạnh tính toán và thời gian huấn luyện dài nếu dữ liệu lớn hơn hoặc phức tạp hơn.

#### **Hướng Phát Triển**

Dựa trên những hạn chế này, một số hướng nghiên cứu trong tương lai được đề xuất như sau:

**Mở rộng tập dữ liệu**: Sử dụng các tập dữ liệu phong phú hơn, chẳng hạn như chữ số viết tay từ nhiều người với phong cách viết khác nhau.

**Áp dụng Transfer Learning**: Tận dụng các mô hình tiền huấn luyện để giảm thời gian và tài nguyên cần thiết cho việc huấn luyện.

**Đa dạng hóa ứng dụng**: Áp dụng mô hình vào các bài toán liên quan như nhận dạng chữ ký, kiểm tra văn bản viết tay, hoặc nhận dạng ký tự tiếng Việt.

**Tối ưu hóa hiệu suất**: Sử dụng các kỹ thuật nén mô hình hoặc triển khai trên các thiết bị nhúng như Raspberry Pi để ứng dụng vào thực tế.

Những kết quả đạt được từ nghiên cứu này là một bước khởi đầu quan trọng, tạo nền tảng cho các ứng dụng hiện đại, đồng thời mở ra tiềm năng phát triển lớn hơn trong lĩnh vực trí tuệ nhân tạo và xử lý dữ liệu hình ảnh.

# TÀI LIỆU THAM KHẢO

1. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Học Máy Dựa Trên Gradient Áp Dụng Cho Nhận Dạng Tài Liệu. Proceedings of the IEEE.

2. Bộ dữ liệu MNIST: http://yann.lecun.com/exdb/mnist/

3. Chollet, F. (2017). Học Sâu với Python. Manning Publications.

4. Tài liệu TensorFlow: <https://www.tensorflow.org/>

5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Học Sâu. MIT Press.