**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** **04: Xây dựng hệ thống nhận diện**

**giới tính và tuổi sử dụng OpenCV**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Mã sinh viên** | **Sinh viên thực hiện** | **Lớp hành chính** |
| **1** | 20210094 | Đặng Đình Tấn | DCCNTT12.10.1 |

**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 04: Xây dựng hệ thống nhận diện**

**giới tính và tuổi sử dụng OpenCV**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Mã sinh viên** | **Sinh viên thực hiện** | **Lớp hành chính** |
| **1** | 20210094 | Đặng Đình Tấn | DCCNTT12.10.1 |

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

**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: *04***  **Tên học phần: XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH**  **Lớp Tín chỉ: XATGMT.03.K12.01.LH.C04.1\_LT** | |
| **Cán bộ chấm thi 1**  *(Ký và ghi rõ họ tên)*  Lương Thị Hồng Lan | **Cán bộ chấm thi 2**  *(Ký và ghi rõ họ tên)* |

| **TT** | **TIÊU CHÍ** | **THANG ĐIỂM** | **Đặng Đình Tấn** | **TÊN SV 2** | **TÊN SV 3** | **TÊN SV 4** | **TÊN SV 5** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 20210094 | MÃ SV 2 | MÃ SV 3 | MÃ SV 4 | MÃ SV 5 |
| **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**

[**MỤC LỤC** 6](#_Toc184634087)

[**DANH MỤC HÌNH ẢNH** 8](#_Toc184634088)

[**LỜI NÓI ĐẦU** 9](#_Toc184634089)

[**CHƯƠNG 1: CƠ SỞ LÝ THUYẾT** 10](#_Toc184634090)

[**1.1. Xử lý ảnh và thị giác máy tính** 10](#_Toc184634091)

[**1.1.1. Khái niệm** 10](#_Toc184634092)

[**1.1.2. Tầm quan trọng** 11](#_Toc184634093)

[**1.1.3. Ứng dụng** 11](#_Toc184634094)

[**1.2. Học máy** 12](#_Toc184634095)

[**1.2.1. Học máy (Machine learning)** 12](#_Toc184634096)

[**1.2.2. Các phương pháp học máy** 13](#_Toc184634097)

[**1.3. Công cụ sử dụng** 14](#_Toc184634098)

[**1.3.1. Ngôn ngữ Python** 14](#_Toc184634099)

[**1.3.2. Visual studio code** 15](#_Toc184634100)

[**1.3.3. Các thư viện sử dụng** 16](#_Toc184634101)

[**CHƯƠNG 2: XÂY DỰNG HỆ THỐNG** 18](#_Toc184634102)

[**2.1. Bài toán nhận diện giới tính độ tuổi** 18](#_Toc184634103)

[**2.1.1. Bài toán nhận diện khuôn mặt** 18](#_Toc184634104)

[**2.1.2. Bài toán nhận diện giới tính và độ tuổi** 19](#_Toc184634105)

[**2.2. Mô hình CNN (Convolutional Neural Networks))** 20](#_Toc184634106)

[**2.2.1. Khái niệm về mạng nơ-ron tích chập CNN** 20](#_Toc184634107)

[**2.2.2. Mô hình** 21](#_Toc184634108)

[**2.3. Thư viện OpenCV** 26](#_Toc184634109)

[**2.3.1. OpenCV là gì** 26](#_Toc184634110)

[**2.3.2. Tính năng và ứng dụng OpenCV** 26](#_Toc184634111)

[**CHƯƠNG 3: KẾT QUẢ THỰC NGHIỆM** 29](#_Toc184634112)

[**3.1. Dữ liệu và xử lý dữ liệu** 29](#_Toc184634113)

[**3.1.1. Dữ liệu** 29](#_Toc184634114)

[**3.1.2. Xử lý dữ liệu** 29](#_Toc184634115)

[**3.2. Đào tạo mô hình** 30](#_Toc184634116)

[**3.3. Kết quả thực nghiệm** 34](#_Toc184634117)

[**KẾT LUẬN** 37](#_Toc184634118)

[**TÀI LIỆU THAM KHẢO** 39](#_Toc184634119)

# **DANH MỤC HÌNH ẢNH**

|  |  |  |
| --- | --- | --- |
| Hình | Tên hình ảnh | Trang |
| 1.1 | Thị giác máy tính | 19 |
| 1.2 | Nhận diện khuôn mặt | 21 |
| 2.1 | Kiến trúc hệ thống | 24 |
| 2.2 | Phân phối bộ dữ liệu (ACROSS Gender) | 25 |
| 2.3 | Phân phối bộ dữ liệu (Qua lứa tuổi) | 25 |
| 2.4 | Mô hình lớp đầu vào | 28 |
| 2.5 | Mô hình tuổi | 28 |
| 2.6 | Mô hình giới tính | 29 |
| 2.7 | Dự đoán độ tuổi và giới tính | 30 |
| 2.8 | Hình ảnh được dự đoán | 31 |
| 2.9 | Các file mã nguồn | 31 |
| 3.1 | Giao diện chính | 33 |
| 3.2 | Giao diện up ảnh | 33 |
| 3.3 | Giao diện loading | 34 |
| 3.4 | Giao diện hoàn tất | 34 |

# **LỜI NÓI ĐẦU**

Trong thời đại công nghệ phát triển như hiện nay, việc ứng dụng trí tuệ nhân tạo (AI) và các công nghệ xử lý hình ảnh vào các lĩnh vực khác nhau của đời sống đã và đang mang lại những bước tiến đột phá. Một trong những ứng dụng quan trọng và mang tính thực tiễn cao là hệ thống nhận diện khuôn mặt và dự đoán độ tuổi, giới tính giúp tăng cường tính năng bảo mật, cá nhân hóa dịch vụ, và hỗ trợ các ứng dụng thông minh khác trong nhiều ngành nghề.

Công nghệ nhận diện khuôn mặt và dự đoán độ tuổi và giới tính ngày càng được chú trọng và phát triển, đặc biệt trong bối cảnh nhu cầu về bảo mật và trải nghiệm người dùng ngày càng tăng cao. Các hệ thống này không chỉ giúp nhận diện nhanh chóng và chính xác danh tính của người dùng mà còn mở ra cơ hội cung cấp các dịch vụ tùy chỉnh theo độ tuổi, từ đó nâng cao sự hài lòng của khách hàng. Việc ứng dụng nhận diện khuôn mặt và độ tuổi có thể được tìm thấy trong các lĩnh vực như giáo dục, y tế, thương mại, giải trí, và thậm chí là quản lý hành chính.

Với sự tiến bộ của công nghệ, đặc biệt là trong lĩnh vực xử lý hình ảnh và trí tuệ nhân tạo, việc xây dựng một ứng dụng nhận diện khuôn mặt và dự đoán độ tuổi, giới tính đã trở nên khả thi và hiệu quả hơn bao giờ hết. Sự kết hợp giữa OpenCV, một thư viện mạnh mẽ trong xử lý hình ảnh, và Keras, một thư viện học sâu (deep learning) phổ biến, cho phép chúng ta phân tích hình ảnh khuôn mặt và dự đoán độ tuổi một cách chính xác. Ứng dụng có thể hoạt động trong thời gian thực nhờ khả năng tích hợp với các camera hoặc thiết bị ghi hình, đồng thời tận dụng sức mạnh của các thuật toán học sâu để cải thiện hiệu suất nhận diện và dự đoán.

Đề tài này tập trung vào việc xây dựng một ứng dụng nhận diện khuôn mặt và dự đoán độ tuổi bằng OpenCV và Keras, nhằm khám phá tiềm năng ứng dụng của AI trong thực tiễn. Hệ thống được thiết kế để hoạt động hiệu quả trong các điều kiện thực tế và có thể dễ dàng tích hợp vào các giải pháp hiện đại. Đây không chỉ là một công cụ kỹ thuật mà còn là một giải pháp hữu ích trong việc cải thiện chất lượng dịch vụ, nâng cao trải nghiệm người dùng và thúc đẩy sự phát triển của các ứng dụng thông minh trong đời sống.

# **CHƯƠNG 1: CƠ SỞ LÝ THUYẾT**

## **1.1. Xử lý ảnh và thị giác máy tính**

### **1.1.1. Khái niệm**

* **Xử lý ảnh:**

Xử lý ảnh (Tiếng Anh: digital image processing) hay xử lý ảnh kỹ thuật số là sự sử dụng máy tính số để xử lý các ảnh kỹ thuật số thông qua một thuật toán. Là một phân ngành tổng thể của xử lý tín hiệu số, xử lý ảnh kỹ thuật số có nhiều lợi thế hơn so với xử lý ảnh analog. Nó cho phép áp dụng nhiều thuật toán hơn cho dữ liệu đầu vào và có thể tránh được các vấn đề như sự tích tụ nhiễu và méo hình trong quá trình xử lý. Vì ảnh được định nghĩa trên hai chiều (hoặc nhiều hơn), xử lý ảnh kỹ thuật số có thể được mô hình hóa dưới dạng hệ thống đa chiều. Sự phát triển của xử lý ảnh kỹ thuật số chủ yếu bị ảnh hưởng bởi ba yếu tố: đầu tiên là sự phát triển của máy tính; thứ hai là sự phát triển của toán học (đặc biệt là sáng tạo và cải tiến lý thuyết toán rời rạc); thứ ba là nhu cầu ứng dụng rộng rãi trong môi trường, nông nghiệp, quân sự, công nghiệp và khoa học y tế ngày càng tăng.

A colorful sphere with a black and white grid

Description automatically generated with medium confidence

* **Thị giác máy tính:**

Thị giác máy tính là một công nghệ mà máy sử dụng để tự động nhận biết và mô tả hình ảnh một cách chính xác và hiệu quả. Ngày nay, các hệ thống máy tính có quyền truy cập vào khối lượng lớn hình ảnh và dữ liệu video bắt nguồn từ hoặc được tạo bằng điện thoại thông minh, camera giao thông, hệ thống bảo mật và các thiết bị khác. Ứng dụng thị giác máy tính sử dụng trí tuệ nhân tạo và máy học (AI/ML) để xử lý dữ liệu này một cách chuẩn xác cho việc xác định đối tượng và nhận diện khuôn mặt, cũng như phân loại, đề xuất, giám sát và phát hiện.



### **1.1.2. Tầm quan trọng**

Tuy rằng công nghệ xử lý thông tin hình ảnh đã xuất hiện từ lâu nhưng phần lớn quy trình vẫn đòi hỏi sự can thiệp của con người, tốn nhiều thời giờ và dễ bị lỗi. Ví dụ: việc triển khai hệ thống nhận diện khuôn mặt trước đây yêu cầu nhà phát triển phải gắn thẻ thủ công hàng ngàn hình ảnh bằng các điểm dữ liệu chính, chẳng hạn như chiều rộng sống mũi và khoảng cách giữa hai mắt. Tự động hóa các tác vụ này đòi hỏi sức mạnh điện toán rộng lớn vì dữ liệu hình ảnh không có cấu trúc và phức tạp để máy tính có thể sắp xếp. Do đó, ứng dụng thị giác tốn kém và hầu hết các tổ chức không thể tiếp cận.

Ngày nay, tiến bộ trong lĩnh vực này kết hợp với sự tăng cường đáng kể của sức mạnh điện toán đã cải thiện cả quy mô và độ chính xác của quy trình xử lý dữ liệu hình ảnh. Các hệ thống thị giác máy tính được hỗ trợ bởi tài nguyên điện toán đám mây hiện giờ trở nên dễ tiếp cận với tất cả mọi người. Bất kỳ tổ chức nào cũng có thể sử dụng công nghệ này để xác minh danh tính, kiểm duyệt nội dung, phân tích video phát trực tuyến, phát hiện lỗi và nhiều tính năng khác.

### **1.1.3. Ứng dụng**

**Bảo mật và an toàn:**

Chính phủ và doanh nghiệp sử dụng thị giác máy tính để tăng cường bảo mật tài sản, cơ sở và trang thiết bị. Ví dụ: camera và cảm biến giám sát các không gian công cộng, cơ sở công nghiệp và môi trường bảo mật cao. Các thiết bị này gửi cảnh báo tự động nếu có điều gì bất thường xảy ra, chẳng hạn như một người đột nhập trái phép vào khu vực bị hạn chế.

**Chăm sóc sức khỏe:**

Y tế là một trong những ngành đi đầu trong áp dụng công nghệ thị giác máy tính. Đáng chú ý, phân tích hình ảnh y khoa tạo hiển thị trực quan mô và các cơ quan giúp chuyên gia y tế chẩn đoán nhanh và chính xác, từ đó cải thiện kết quả điều trị và kéo dài tuổi thọ. Ví dụ:

* Phát hiện khối u thông qua phân tích nốt ruồi và tổn thương trên da
* Phân tích tia X tự động
* Phát hiện triệu chứng từ phép chụp MRI

**Phương tiện tự hành:**

Công nghệ xe tự hành sử dụng thị giác máy tính để nhận biết hình ảnh theo thời gian thực và xây dựng bản đồ 3D từ nhiều camera được trang bị cho vận chuyển tự động. Phương tiện tự hành có khả năng phân tích hình ảnh và xác định những người đi đường khác, biển báo trên đường, người đi bộ hoặc chướng ngại vật.

Ở các phương tiện bán tự hành, thị giác máy tính sử dụng máy học (ML) để giám sát hành vi của người lái. Ví dụ: nó tìm kiếm các dấu hiệu phân tâm, mệt mỏi và buồn ngủ dựa theo vị trí đầu của người lái xe, theo dõi mắt và chuyển động của phần trên cơ thể. Nếu công nghệ nhận được một số dấu hiệu cảnh báo nhất định, nó sẽ cảnh báo cho người lái xe và giảm nguy cơ xảy ra sự cố khi lái xe.

**Nông nghiệp:**

Ứng dụng thị giác máy tính tăng cường hoạt động tổng thể của ngành nông nghiệp, từ tăng năng suất đến giảm chi phí với công nghệ tự động hóa thông minh. Hình ảnh vệ tinh cũng như cảnh quay UAV giúp phân tích các vùng đất rộng lớn và cải thiện các phương pháp canh tác. Ứng dụng thị giác máy tính tự động hóa các tác vụ như giám sát tình trạng ruộng đồng, xác định bệnh ở cây trồng, kiểm tra độ ẩm của đất và dự đoán thời tiết cũng như năng suất cây trồng. Giám sát động vật bằng thị giác máy tính là một chiến lược canh tác thông minh quan trọng khác

## **1.2. Học máy**

### **1.2.1. Học máy (Machine learning)**

Machine learning là một nhánh của trí tuệ nhân tạo (AI) và khoa học máy tính, tập trung vào việc sử dụng dữ liệu và thuật toán để bắt chước hành động của con người, dần dần cải thiện độ chính xác của nó.

Machine learning còn là một thành phần quan trọng của lĩnh vực khoa học dữ liệu đang phát triển. Thông qua việc sử dụng các phương pháp thống kê, các thuật toán được đào tạo để đưa ra các phân loại hoặc dự đoán và khám phá những thông tin chi tiết từ chính các dự án khai thác dữ liệu.

Thông qua các thông tin chi tiết có được để thúc đẩy việc đưa ra quyết định đối

với các ứng dụng và doanh nghiệp, tác động mạnh đến các chỉ số tăng trưởng. Khi dữ

liệu lớn tiếp tục nhu cầu mở rộng và phát triển đòi hỏi nhu cầu tuyển dụng các nhà khoa học dữ liệu sẽ tăng lên. Họ sẽ được yêu cầu giúp xác định các câu hỏi kinh doanh có liên quan nhất và dữ liệu để trả lời chúng.

Bài toán của machine learning thường được chia làm hai loại là dự đoán (prediction) và phân loại (classification). Các bài toán dự đoán thường là giá nhà, giá xe, v.v, còn các bài toán phân loại thường là nhận diện chữ viết tay, đồ vật, v.v.

Các thuật toán phổ biến của machine learning:

* Neural networks: Mô phỏng cách thức hoạt động của bộ não con người, với một số lượng khổng lồ các nút xử lý được liên kết. Neural networks là thuật toán được dùng trong việc nhận dạng các mẫu và đóng một vai trò quan trọng trong các ứng dụng bao gồm dịch ngôn ngữ tự nhiên, nhận dạng hình ảnh, nhận dạng giọng nói và tạo hình ảnh.
* Linear regression: Thuật toán này được sử dụng để dự đoán các giá trị số, dựa trên mối quan hệ tuyến tính giữa các giá trị khác nhau.
* Logistic regression: Thuật toán giúp đưa ra dự đoán cho các biến phản hồi phân loại, chẳng hạn như câu trả lời “có/không” cho các câu hỏi. Nó có thể được sử dụng cho các ứng dụng như phân loại thư rác và kiểm soát chất lượng trên dây chuyền sản xuất.
* Clustering: Các thuật toán phân cụm có thể xác định các mẫu trong dữ liệu để nó có thể được nhóm lại. Máy tính có thể giúp các nhà khoa học dữ liệu bằng cách xác định sự khác biệt giữa các mục dữ liệu mà con người đã bỏ qua.
* Decision trees: Là thuật toán được sử dụng để dự đoán giá trị số (hồi quy) và phân loại dữ liệu. Decision trees sử dụng một chuỗi phân nhánh của các quyết định được liên kết có thể được biểu diễn bằng sơ đồ cây. Một trong những ưu điểm của decision trees là chúng dễ xác thực và kiểm tra, không giống thuật toán Neural networks.
* Random forests: Trong một khu rừng ngẫu nhiên, thuật toán máy học dự đoán một giá trị hoặc danh mục bằng cách kết hợp các kết quả từ một số cây quyết định.

### **1.2.2. Các phương pháp học máy**

* Học có giám sát

Trong loại học máy này, các nhà khoa học dữ liệu cung cấp các thuật toán với dữ liệu huấn luyện được gắn nhãn và xác định các biến mà họ muốn thuật toán đánh giá về các mối tương quan. Cả đầu vào và đầu ra của thuật toán đều được chỉ định.

Để giải quyết một vấn đề nhất định về học có giám sát, người ta phải thực hiện

các bước sau:

Bước 1: Xác định loại ví dụ đào tạo. Trước khi làm bất cứ điều gì khác, người dùng nên quyết định loại dữ liệu nào sẽ được sử dụng làm tập huấn luyện. Ví dụ, trong trường hợp phân tích chữ viết tay, đây có thể là một ký tự viết tay đơn lẻ, toàn bộ từ viết tay, toàn bộ câu chữ viết tay hoặc có thể là một đoạn văn viết tay đầy đủ.

Bước 2: Tập hợp một tập hợp đào tạo. Tập huấn luyện cần phải đại diện cho việc sử dụng hàm trong thế giới thực. Do đó, một tập hợp các đối tượng đầu vào được tập hợp và các đầu ra tương ứng cũng được thu thập, từ các chuyên gia con người hoặc từ các phép đo.

Bước 3: Xác định biểu diễn đặc điểm đầu vào của hàm đã học. Độ chính xác của hàm đã học phụ thuộc nhiều vào cách biểu diễn đối tượng đầu vào. Thông thường, đối tượng đầu vào được chuyển đổi thành một vectơ đặc trưng, chứa một số đặc điểm mô tả đối tượng. Số lượng các đối tượng địa lý không được quá lớn, vì điều này có thể xảy ra; nhưng phải chứa đủ thông tin để dự đoán chính xác kết quả đầu ra.

Bước 4: Xác định cấu trúc của hàm đã học và thuật toán học tương ứng. Ví dụ, kỹ sư có thể chọn sử dụng máy vectơ hỗ trợ hoặc cây quyết định.

Bước 5: Hoàn thiện thiết kế. Chạy thuật toán học tập trên tập huấn luyện đã tập hợp. Một số thuật toán học có giám sát yêu cầu người dùng xác định các thông số điều khiển nhất định. Các tham số này có thể được điều chỉnh bằng cách tối ưu hóa hiệu suất trên một tập hợp con (được gọi là tập xác nhận) của tập huấn luyện hoặc

thông qua xác nhận chéo.

Bước 6: Đánh giá độ chính xác của hàm đã học. Sau khi điều chỉnh tham số và học hỏi, hiệu suất của chức năng kết quả phải được đo trên một bộ thử nghiệm tách biệt với bộ huấn luyện.

* Học không giám sát

Loại học máy này liên quan đến các thuật toán đào tạo trên dữ liệu không được gắn nhãn. Thuật toán quét qua các tập dữ liệu để tìm kiếm bất kỳ kết nối có ý nghĩa nào. Dữ liệu mà các thuật toán đào tạo cũng như các dự đoán hoặc khuyến nghị mà chúng xuất ra được xác định trước.

* Học bán giám sát

Cách tiếp cận này đối với học máy liên quan đến sự kết hợp của hai loại trước đó. Các nhà khoa học dữ liệu có thể cung cấp một thuật toán chủ yếu là dữ liệu đào tạo được gắn nhãn, nhưng mô hình có thể tự do khám phá dữ liệu và phát triển sự hiểu biết của riêng mình về tập dữ liệu.

* Học tăng cường

Các nhà khoa học dữ liệu thường sử dụng học tăng cường để dạy máy hoàn thành một quy trình gồm nhiều bước trong đó có các quy tắc được xác định rõ ràng. Các nhà khoa học dữ liệu lập trình một thuật toán để hoàn thành một nhiệm vụ và cung cấp cho nó các tín hiệu tích cực hoặc tiêu cực khi nó tìm ra cách hoàn thành một nhiệm vụ. Nhưng phần lớn, thuật toán tự quyết định những bước cần thực hiện trong quá trình thực hiện.

## **1.3. Công cụ sử dụng**

### **1.3.1. Ngôn ngữ Python**

Python là một ngôn ngữ lập trình cấp cao, thường được ứng dụng để phát triển website và phần mềm, tự động hóa công việc cũng như phân tích dữ liệu. Đây là một ngôn ngữ đa năng, cho phép xây dựng nhiều loại chương trình khác nhau mà không bị giới hạn bởi một lĩnh vực cụ thể nào, ví dụ như hình ảnh trong đồ họa có thể thấy một chương trình đang hoạt động.

A computer screen with a logo

Description automatically generated

Tính linh hoạt của Python, kèm theo tính thân thiện với người dùng mới, đã góp phần làm cho nó trở thành một trong những ngôn ngữ lập trình phổ biến nhất hiện nay. Theo khảo sát từ công ty phân tích RedMonk, Python đứng thứ hai trong danh sách ngôn ngữ lập trình được yêu thích nhất trong năm 2021.

Về mặt ứng dụng, Python chủ yếu được dùng trong phát triển website và phần mềm, tự động hóa tác vụ, phân tích và trực quan hóa dữ liệu. Nhờ vào dễ học, nhiều người không chuyên về lập trình, như kế toán viên hay nhà khoa học, đã áp dụng Python trong các công việc hàng ngày liên quan đến quản lý tài chính.

Trong lĩnh vực phân tích dữ liệu và học máy, Python đóng vai trò quan trọng khi cho phép các nhà phân tích thực hiện các phép toán thống kê phức tạp, tạo ra các hình ảnh dữ liệu, xây dựng thuật toán machine learning, xử lý và phân tích dữ liệu, cũng như hoàn thành các nhiệm vụ liên quan khác. Python có khả năng tạo ra nhiều loại biểu đồ khác nhau như biểu đồ đường, thanh, hình tròn và 3D. Đặc biệt, nó còn cung cấp nhiều thư viện hỗ trợ lập trình viên trong việc phân tích dữ liệu và học máy, chẳng hạn như TensorFlow và Keras.

Khi bạn phải thực hiện những công việc lặp đi lặp lại, tự động hóa bằng Python sẽ giúp bạn làm việc hiệu quả hơn. Quá trình viết mã để tạo ra các quy trình tự động này được xem là scripting. Trong lĩnh vực lập trình, tự động hóa có thể hỗ trợ kiểm tra lỗi trên nhiều tập tin, chuyển đổi định dạng tệp, thao tác đơn giản và loại bỏ bản sao trong dữ liệu.

### **1.3.2. Visual studio code**

Visual Studio Code chính là ứng dụng cho phép biên tập, soạn thảo các đoạn code để hỗ trợ trong quá trình thực hiện xây dựng, thiết kế website một cách nhanh chóng. Visual Studio Code hay còn được viết tắt là VS Code. Trình soạn thảo này vận hành mượt mà trên các nền tảng như Windows, macOS, Linux. Hơn thế nữa, VS Code còn cho khả năng tương thích với những thiết bị máy tính có cấu hình tầm trung vẫn có thể sử dụng dễ dàng.

A screenshot of a computer

Description automatically generated

Visual Studio Code hỗ trợ đa dạng các chức năng Debug, đi kèm với Git, có Syntax Highlighting. Đặc biệt là tự hoàn thành mã thông minh, Snippets, và khả năng cải tiến mã nguồn. Nhờ tính năng tùy chỉnh, Visual Studio Code cũng cho phép các lập trình viên thay đổi Theme, phím tắt, và đa dạng các tùy chọn khác. Mặc dù trình soạn thảo Code này tương đối nhẹ, nhưng lại bao gồm các tính năng mạnh mẽ.

Dù mới được phát hành nhưng VSCode là một trong những Code Editor mạnh mẽ và phổ biến nhất dành cho lập trình viên. Nhờ hỗ trợ nhiều ngôn ngữ lập trình phổ biến, tích hợp đầy đủ các tính năng và khả năng mở rộng, nên VSCode trở nên cực kì thân thuộc với bất kì lập trình viên nào.

### **1.3.3. Các thư viện sử dụng**

* **OpenCV**

OpenCV (viết tắt của Open Source Computer Vision Library) là một thư viện mã nguồn mở chuyên dùng trong xử lý ảnh và thị giác máy tính. Công nghệ cung cấp các công cụ và thư viện để phân tích và xử lý ảnh, video từ việc xác định các đối tượng trong ảnh đến việc nhận diện khuôn mặt hoặc theo dõi chuyển động khác**.**

* **Keras**

Keras được ra mắt vào năm 2015 và tác giả là François Chollet, nhà nghiên cứu và kỹ sư nổi tiếng trong lĩnh vực trí tuệ nhân tạo. Keras đã nhanh chóng trở thành một trong những công cụ phổ biến nhất trong cộng đồng Deep Learning nhờ vào giao diện thân thiện với người dùng và khả năng tích hợp với các framework mạnh mẽ như TensorFlow. Các đặc tính nổi bật của Keras:

* API cấp cao dễ sử dụng: Keras cung cấp các API cấp cao cho phép người dùng định nghĩa mô hình học sâu một cách trực quan và dễ hiểu. Thay vì phải viết code chi tiết cho từng lớp và kết nối trong mạng neural, Keras cho phép bạn xây dựng mô hình chỉ với vài dòng mã.
* Tập trung vào phần quan trọng: Thay vì phải lo về các chi tiết kỹ thuật phức tạp, Keras giúp người dùng tập trung vào việc thiết kế mô hình, chọn hàm mất mát và các thuật toán tối ưu, làm cho quá trình xây dựng mô hình trở nên nhẹ nhàng hơn.
* **TensorFlow**

TensorFlow là thư viện học sâu được Google phát triển vào năm 2015 và nhanh chóng trở thành một trong những công cụ phổ biến nhất trong lĩnh vực Deep Learning với các ưu điểm nổi trội như:

* Hiệu suất tính toán cao: TensorFlow được thiết kế để xử lý các phép toán trên tensor (ma trận nhiều chiều) một cách tối ưu trên nhiều loại phần cứng như CPU, GPU và cả TPU (phần cứng đặc biệt của Google). Điều này cho phép mô hình có thể xử lý lượng dữ liệu lớn và giảm đáng kể thời gian huấn luyện các mô hình phức tạp.
* Tự động tính toán gradient: Một trong những tính năng nổi bật của TensorFlow là khả năng tự động tính toán gradient - một khái niệm toán học quan trọng trong học sâu. Gradient cho biết hướng thay đổi của các tham số mô hình để giảm thiểu sai số dự đoán. TensorFlow giúp chúng ta không phải tự tính toán gradient một cách thủ công, tiết kiệm rất nhiều thời gian và công sức.
* Cung cấp API cấp thấp: TensorFlow cung cấp cho người dùng khả năng kiểm soát chi tiết mọi khía cạnh của mô hình, từ cấu trúc mạng đến thuật toán tối ưu hóa. Điều này rất hữu ích cho các nhà nghiên cứu muốn thử nghiệm những mô hình phức tạp.
* Khả năng triển khai đa nền tảng: TensorFlow hỗ trợ triển khai mô hình trên nhiều nền tảng như web, di động, máy chủ và các thiết bị nhúng, giúp mô hình dễ dàng được ứng dụng vào thực tiễn hơn.

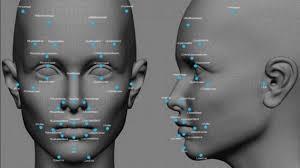
# **CHƯƠNG 2: XÂY DỰNG HỆ THỐNG**

## **2.1. Bài toán nhận diện giới tính độ tuổi**

### **2.1.1. Bài toán nhận diện khuôn mặt**

* **Giới thiệu bài toán.**

Nhận dạng đối tượng trong ảnh là một lĩnh vực quan trọng của trí tuệ nhân tạo (AI) và thị giác máy tính (Computer Vision). Mục tiêu của bài toán là xác định và phân loại các đối tượng xuất hiện trong ảnh hoặc video. Ví dụ, hệ thống có thể nhận diện một khuôn mặt, chiếc xe, hay đồ vật cụ thể trong hình ảnh đầu vào.

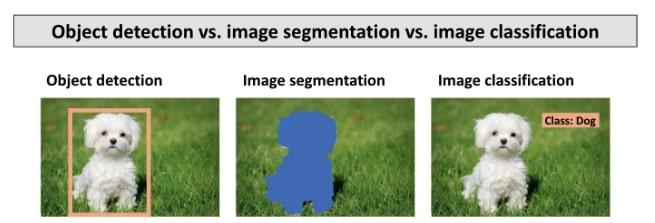


*Hình 1.1: Thị giác máy tính*

* Nhận dạng đối tượng có rất nhiều ứng dụng thực tế:
* Phân loại sản phẩm trong thương mại điện tử.
* Phân tích hình ảnh y tế, ví dụ như phát hiện các khối u trong ảnh chụp X-quang.
* Hỗ trợ điều khiển tự động, ví dụ trong xe tự lái, nhận dạng biển báo giao thông và người đi bộ.
* **Các vấn đề đặt ra với bài toán nhận diện.**

Trong bài toán nhận dạng, các công việc chính cần thực hiện bao gồm:

1. Phát hiện đối tượng (Object Detection): Xác định vị trí của đối tượng trong ảnh. Công việc này thường trả về các hộp giới hạn (bounding boxes) chứa đối tượng.
2. Trích xuất đặc trưng (Feature Extraction): Trích xuất các đặc trưng quan trọng của đối tượng, như hình dáng, màu sắc, hoặc kết cấu, để phân loại.
3. Phân loại đối tượng (Object Classification): Xác định loại của đối tượng dựa trên các đặc trưng đã trích xuất.
4. Phân đoạn đối tượng (Object Segmentation): Phân chia từng pixel của ảnh để xác định chính xác các vùng thuộc về đối tượng.

******

*Hình 1.2: Nhận diện khuôn mặt*

### **2.1.2. Bài toán nhận diện giới tính và độ tuổi**

* **Giới thiệu bài toán**

Nhận diện giới tính và tuổi từ hình ảnh khuôn mặt là một trong những ứng dụng phổ biến của thị giác máy tính và trí tuệ nhân tạo. Bài toán này liên quan đến việc phát hiện khuôn mặt từ hình ảnh hoặc video, sau đó phân tích các đặc điểm khuôn mặt để dự đoán giới tính (nam, nữ) và ước lượng tuổi của người đó.

Hệ thống nhận diện khuôn mặt, giới tính và độ tuổi có nhiều ứng dụng thực tiễn, bao gồm:

* Thương mại: Cá nhân hóa trải nghiệm khách hàng, phân tích hành vi người tiêu dùng dựa trên giới tính và độ tuổi.
* Giáo dục: Quản lý học sinh, sinh viên; tối ưu hóa các phương pháp giáo dục dựa trên nhóm tuổi.
* Y tế: Hỗ trợ bác sĩ trong việc thu thập thông tin bệnh nhân theo độ tuổi và giới tính.
* An ninh: Kiểm soát truy cập, nhận diện cá nhân trong các khu vực nhạy cảm.
* Quảng cáo: Định hướng nội dung phù hợp với giới tính và độ tuổi của khách hàng mục tiêu.
* **Ứng dụng thực tiễn**

Quảng cáo thông minh: Phân tích đối tượng khách hàng theo giới tính và độ tuổi để cung cấp các gợi ý quảng cáo phù hợp.

Xác thực danh tính: Hỗ trợ trong hệ thống an ninh để xác minh thông tin cá nhân.

Trải nghiệm khách hàng: Tích hợp vào các hệ thống bán lẻ, như cửa hàng tự động, để cải thiện dịch vụ.

Y tế và chăm sóc sức khỏe: Hỗ trợ theo dõi sự phát triển của trẻ em hoặc chẩn đoán các vấn đề lão hóa từ khuôn mặt.

* Hoạt động của hệ thống

*Đầu vào:*

Hình ảnh hoặc video từ camera.

Dữ liệu khuôn mặt được phát hiện và cắt từ khung hình.

Xử lý trung tâm:

Phát hiện khuôn mặt: Sử dụng OpenCV với các mô hình phát hiện khuôn mặt (Haar Cascade, DNN).

Dự đoán giới tính: Sử dụng mô hình CNN để phân loại nam hoặc nữ.

Dự đoán độ tuổi: Sử dụng mô hình học sâu để ước tính tuổi từ các đặc trưng khuôn mặt.

*Đầu ra:*

Hiển thị giới tính và độ tuổi ước tính trên giao diện ứng dụng.

Tùy chọn xuất dữ liệu để phân tích hoặc lưu trữ.

Đặc điểm kỹ thuật của hệ thống

## **2.2. Mô hình CNN (Convolutional Neural Networks))**

### **2.2.1. Khái niệm về mạng nơ-ron tích chập CNN**

Mạng nơ ron tích chập là một trong những mạng truyền thẳng đặc biệt. Mạng nơ ron tích chập là một mô hình học sâu phổ biến và tiên tiến nhất hiện nay. Hầu hết các hệ thống nhận diện và xử lý ảnh hiện nay đều sử dụng mạng nơ ron tích chập vì tốc độ xử lý nhanh và độ chính xác cao. Trong mạng nơ ron truyền thống, các tầng được coi là một chiều, thì trong mạng nơ ron tích chập, các tầng được coi là 3 chiều, gồm: chiều cao, chiều rộng và chiều sâu . Mạng nơ ron tích chập có hai khái niệm quan trọng: kết nối cục bộ và chia sẻ tham số. Những khái niệm này góp phần giảm số lượng trọng số cần được huấn luyện, do đó tăng nhanh được tốc độ tính toán. [6]

A blue cube with a black arrow pointing at it

Description automatically generated with medium confidence

Hình 1. 2: Các tầng (layer) trong CNN là 3 chiều

Convolutional Neural Networks (CNN) là một trong những mô hình deep learning phổ biến nhất và có ảnh hưởng nhiều nhất trong cộng đồng thị giác máy tính (Computer Vision). CNN được dùng trong trong nhiều bài toán như nhân dạng ảnh, phân tích video, ảnh MRI, hoặc cho bài các bài của lĩnh vự xử lý ngôn ngữ tự nhiên,và hầu hết đều giải quyết tốt các bài toán này.

### **2.2.2. Mô hình**

Một kiến trúc CNN bao gồm các lớp: convolution layer, pooling layer và fully connected layer. Ở giữa các lớp convolution và pooling thường có các hàm kích hoạt phi tuyến. Ảnh khi đưa vào mạng sẽ được lan truyền qua tầng convolution layer, giá trị tính được từ các tầng convolution sẽ đi qua một hàm kích hoạt, sau đó giá trị này sẽ được lan truyền qua pooling layer. Cuối cùng ảnh sẽ được lan truyền đến tầng fully connected layer và đi qua hàm kích hoạt Softmax, thường thì cuối cùng sẽ thu được một vector chứa xác suất phần trăm thuộc về các lớp đối với các bài toán phân loại.

Ví dụ minh họa về một kiến trúc mạng nơ ron tích chập đầy đủ:

A diagram of a graph

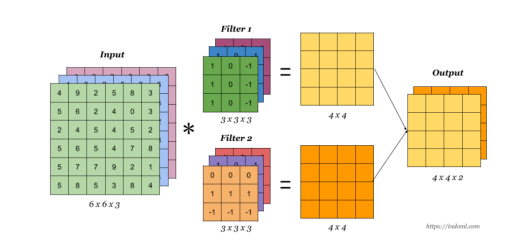
Description automatically generated

Hình 1. 3: Ví dụ minh họa về cấu trúc CNNs – LeNet – 5[7]

* **Convolution layer**

Convolution layer là lớp quan trọng nhất và cũng là lớp đầu tiên của của mô hình CNN. Lớp này có chức năng chính là phát hiện các đặc trưng có tính không gian hiệu quả. Trong tầng này có 4 đối tượng chính là: ma trận đầu vào, bộ filters, và receptive field, feature map. Conv layer nhận đầu vào là một ma trận 3 chiều và một bộ filters cần phải học. Bộ filters này sẽ trượt qua từng vị trí trên bức ảnh để tính tích chập (convolution) giữa bộ filter và phần tương ứng trên bức ảnh. Phần tương ứng này trên bức ảnh gọi là receptive field, tức là vùng mà một neuron có thể nhìn thấy để đưa ra quyết định, và mà trận cho ra bởi quá trình này được gọi là feature map. Khi đưa ảnh vào mạng, bộ filter sẽ quét qua toàn bộ ảnh cho nên các đặc trưng cơ bản của ảnh như là góc, cạnh, màu sắc và texture sẽ được mạng phát hiện ra bất kể nó nằm ở vị trí nào trong ảnh. Do đó tầng convolution được xem như là một bộ trích chọn đặc trưng (feature detector) vì nó có chức năng chính là phát hiện đặc trưng cụ thể của bức ảnh đầu vào. [8]

Khi áp dụng phép tính tích chập cho xử lý hình ảnh, người ta nhận thấy rằng kỹ thuật tích chập này sẽ giúp biến đổi các thông tin đầu vào thành các yếu tố đặc trưng (nó tương tự như bộ phát hiện nhằm phát hiện ra các đặc trưng như cạnh, hướng, ...). Hình 1.4 minh họa cho việc áp dụng phép tính tích chập trên ảnh và cho ra kết quả là một bản đồ đặc trưng - feature map. Cụ thể hơn, tích chập sẽ trích xuất đặc trưng của ảnh đầu vào qua các vùng ảnh nhỏ. Các vùng này được gọi là Local Receptive Field (LRF). Tích chập sẽ tính toán trên các LRF chồng lấp lên nhau. Độ chồng lắp này phụ thuộc vào hệ số trượt S (stride) của từng kiến trúc mạng cụ thể. Nếu sử dụng với hệ số trượt S = α, thì tương ứng LRF (bằng kích thước với kernel) sẽ dịch chuyển α đơn vị pixel sau mỗi lần tích chập.



Hình 1. 4: Minh họa cách thức tính chập của một ảnh RGB và ma trận kernel

Ảnh đầu vào sau khi thực hiện quá trình tích chập sẽ thu được bản đồ đặc trưng, số LRF ở ảnh đầu vào sẽ tương ứng với số neural ở feature map và kernel sẽ là trọng số liên kết mỗi LRF với một neural ở bản đồ đặc trưng. Lớp tích chập có thể chứa một hoặc nhiều feature map. Nếu lớp tích chập có K feature map, thì ta nói lớp conv này có độ sâu là k. Để hình dung rõ hơn về quá trình này, sau đây sẽ minh họa quá trình trích xuất đặc trưng từ ảnh đầu vào cụ thể như sau: thực hiện xử lý tính giá trị đầu ra của một ảnh có kích thước W1×H1×D1 (W1 và H1 lần lượt là chiều rộng và chiều cao của ảnh và D1 là chiều sâu hay thực chất là giá trị tại 3 kênh màu tương ứng của ảnh RGB). khi đó, một Conv như một cửa sổ trượt (sliding window, còn được gọi là kernel, filter hay feature detector) với kích thước F ×F - giả sử trong trường ta sử dụng K filter. Trong quá trình xử lý, mỗi filter sẽ được tính toán với tất cả các LRF trong hình và S = α. Trong một số trường hợp để cân bằng giữa số bước di chuyển và kích thước của ảnh, người ta đã chèn thêm P pixel với một giá trị màu

được gán (thông thường là 0) xung quanh viền của ảnh. sau cùng ta thu được ma trận đầu ra (feature map) với kích thước W2 ×H2 ×D2. [9] [10] [11]

A diagram of a grid with numbers and a red line

Description automatically generated

Hình 1. 5: Mô phỏng quá trình tích chập trong CNN

* **Các tham số của lớp tích chập – Convolutional Layer:**

Các tham số cơ bản của tầng convolution chính là kích thước filter, stride và padding. Trong đó quan trọng nhất chính là kích thước bộ filter, vì nó tỉ lệ thuận với số tham số cần học tại mỗi tầng convolution và là tham số quyết định receptive field của tầng này. Kích thước filter phổ biến thường dùng là 3x3. Thông thường chúng ta nên chọn kích thước filter nhỏ, vì các lý do sau:

* Rút trích được các đặc trưng có tính cục bộ cao.
* Phát hiện được các đặc trưng nhỏ.
* Rút trích đa dạng đặc trưng, hữu ích cho các tầng sau.
* Kích thước ảnh giảm chậm, cho phép xây dựng một kiến trúc mạng sâu, học
* được nhiều hơn.
* Chia sẻ trọng số tốt.

A diagram of a number

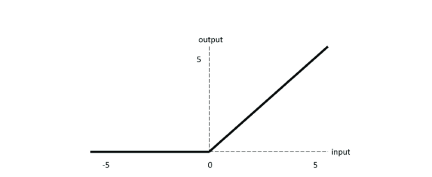
Description automatically generated with medium confidence

Hình 1. 6: Minh họa về bộ lọc filter

Ngoài ra, tham số stride cũng cần lưu ý bởi vì nó thể hiện số pixel cần phải dịch chuyển mỗi khi trượt bộ filter qua bức ảnh. Tham số padding cũng rất quan trọng bởi vì nó sẽ giúp giữ nguyên kích thước ma trận đầu ra của mỗi tầng convolution, do đó ta có thể xây dựng được một kiến trúc mạng với số tầng tùy ý.

* **Hàm kích hoạt**

Hàm kích hoạt là một hàm số nhận vào một giá trị đầu vào và kết quả là một giá trị có miền giá trị nằm trên một khoảng (hay nửa khoảng) nào đó. Một số các hàm kích hoạt phổ biến có thể kể đến đó là Sigmod, Tanh, Relu. Hàm kích hoạt rất quan trọng bởi vì nó sẽ tăng khả năng dự đoán của mạng neural và giúp mô hình học được các quan hệ phi tuyến phức tạp tiềm ẩn trong dữ liệu. Thông thường hàm kích hoạt sử dụng ở giữa các tầng convolution và pooling là hàm Relu. [9] [10] [11]



Hình 1. 7: Đồ thị hàm kích hoạt Relu

Hàm Relu có công thức toán học là 𝒇(𝒙) = 𝐦𝐚𝐱 (𝟎, 𝒙). Hàm Relu được ưa chuộng vì tính toán đơn giản, giúp hạn chế tình trạng vanishing gradient, và cũng cho kết quả tốt hơn. Relu cũng như những hàm kích hoạt khác, được đặt ngay sau tầng convolution, Relu sẽ gán những giá trị âm bằng 0 và giữ nguyên giá trị của đầu vào khi lớn hơn 0.

* **Lớp Pooling**

Lớp Pooling được sử dụng sau lớp Relu theo như mẫu thiết kết các lớp theo như trình bày của đại học Standford. Pooling giúp cho mạng giảm số lượng tham số, từ đó giúp đơn giản hóa quá trình tính toán của CNN và qua đó góp phần giải quyết vấn đề overfiting khi huấn luyện mạng.

Có nhiều toán tử pooling như Sum-pooling, Max-pooling, L2-pooling nhưng Max-pooling được sử dụng phổ biến nhất trong kiến trúc mạng CNN vì nó cho kết quả hơn so với những toán tử còn lại. Ngoài ra, Max-pooling còn giúp tạo ra tính bất biến dịch chuyển (translation invariance) cho đặc trưng. Cụ thể, dù đối tượng trong hình ảnh đầu vào có sự dịch chuyển nhỏ thì mạng vẫn có khả năng phân lớp chính xác được đối tượng. Đó là bởi vì max-pooling chọn ra neural có giá trị đầu ra tại mỗi vùng neural của lớp trước và tổng hợp thành lớp sau. Việc chọn neural có giá trị tín hiệu lớn nhất được xem như chọn ra đặc trưng tốt nhất để xử lý. Chính vì thế, khả năng phân lớp chính xác đối tượng dựa trên đặc trưng này vẫn không thay đổi và giúp cho CNN đạt được tính ổn định khi đối tượng di chuyển.

Trong lớp tích chập, có thể có nhiều feature map, tương ứng với mỗi feature map sẽ có một lớp max-pooling. Hình 1.8 là một ví dụ minh họa, với lớp đầu vào kích thước [28 × 28] giả sử ta thu được ba feature map kích thước [24 × 24] và ba lớp max-pooling kích thước [12×12], ta sử dụng kernel kích thước [5×5] và max-pooling lấy giá trị lớn nhất tại mỗi vùng [2×2] neural. [9]

A diagram of a number and a slice

Description automatically generated with medium confidence

Hình 1. 8: Minh họa kỹ thuật Pooling trong mô hình CNN

* **Lớp Normalization**

Lớp Normalization (Norm) là lớp giúp chuẩn hóa dữ liệu đầu ra cho các lớp trong CNN trước khi được truyền đi tiếp. Trong những kiến trúc CNN lớn và phức tạp, lớp Norm sẽ chuẩn hóa các giá trị neural trước khi chúng được truyền đến hàm Relu. Hàm Relu tuy giúp rút ngắn thời gian huấn luyện, nhưng nếu không điều chỉnh trọng số phù hợp, hàm Relu sẽ rất dễ gặp phải vấn đề "dying Relu" khiến cho mạng trở nên chậm hơn khi huấn luyện. Lớp Norm lúc này sẽ chuẩn hóa và tạo ra các giá trị tích chập phù hợp để tránh cho Relu rơi vào giá trị 0. Tránh việc gradient xấp xỉ bằng 0 khiến cho tốc độ học của mạng trở nên rất chậm.

* **Lớp đầy đủ - Fully connected layer**

Tầng cuối cùng của mô hình CNN trong bài toán phân loại ảnh là tầng fully connected layer. Tầng này có chức năng chuyển ma trận đặc trưng ở tầng trước thành vector chứa xác suất của các đối tượng cần được dự đoán. Ví dụ trong một bài toán phân lớp có 10 lớp, tầng fully connected layer sẽ chuyển ma trận đặc trưng của tầng trước thành vector có 10 chiều thể hiện xác suất của 10 lớp tương ứng. [8]

A diagram of a network

Description automatically generated

Hình 1. 9: Minh họa Fully connected layer

## **2.3. Thư viện OpenCV**

### **2.3.1. OpenCV là gì**

OpenCV (viết tắt của Open Source Computer Vision Library) là một thư viện mã nguồn mở chuyên dùng trong xử lý ảnh và thị giác máy tính. Công nghệ cung cấp các công cụ và thư viện để phân tích và xử lý ảnh, video từ việc xác định các đối tượng trong ảnh đến việc nhận diện khuôn mặt hoặc theo dõi chuyển động khác.

A logo with a colorful logo

Description automatically generated with medium confidence

Hệ thống OpenCV cung cấp các thuật toán và công cụ mạnh mẽ được dùng để phát hiện, phân tích hình ảnh. Ngoài ra, công cụ còn được sử dụng rộng rãi trong nhiều lĩnh vực liên quan đến thị giác máy tính, robot và xe tự động.

### **2.3.2. Tính năng và ứng dụng OpenCV**

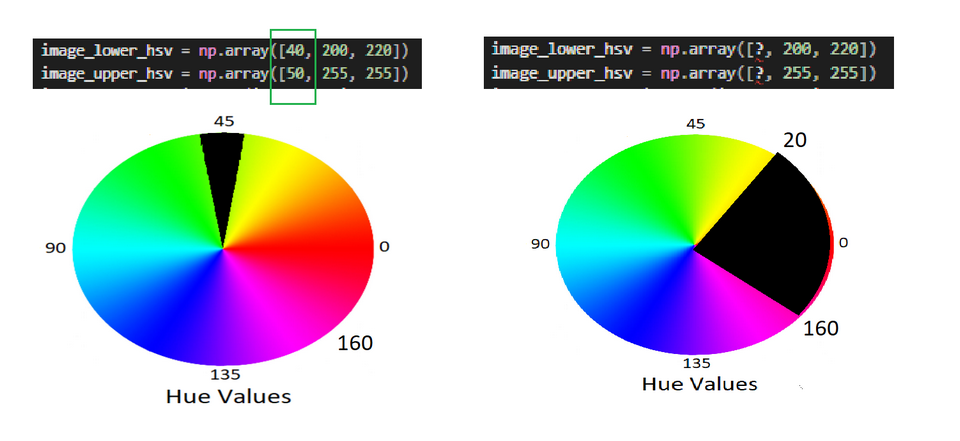
* **Tính năng**

OpenCV (Open Source Computer Vision Library) là một thư viện mã nguồn mở và rất phổ biến trong lĩnh vực thị giác máy tính và xử lý ảnh.



* Xử lý ảnh cơ bản: OpenCV cung cấp các công cụ để đọc, ghi, xử lý và xử lý ảnh cơ bản. Bao gồm các hoạt động như xử lý histogram, ánh sáng, màu sắc và cắt ghép ảnh.
* Nhận diện và theo dõi đối tượng: OpenCV cung cấp các thuật toán để nhận diện và theo dõi đối tượng trong ảnh và video, bao gồm cả nhận diện khuôn mặt, phát hiện đối tượng.
* Thực hiện xử lý ảnh nâng cao: OpenCV hỗ trợ các tính năng xử lý ảnh nâng cao như biến đổi hình thái, lọc và biểu diễn hình ảnh ở không gian màu khác nhau.
* Xử lý video: OpenCV cung cấp các công cụ để xử lý video, phân tích chuyển động, trích xuất vật thể và phân tích hành vi.
* Máy học và nhận biết: OpenCV tích hợp sẵn các thuật toán máy học và nhận biết như SVM, k-nearest neighbors, decision trees và deep learning.
* Xử lý các biểu đồ và ảnh y khoa: OpenCV hỗ trợ việc xử lý và phân tích các loại ảnh đặc biệt như ảnh y khoa và ảnh hàng không.
* Tích hợp với các ngôn ngữ lập trình: OpenCV được tích hợp sâu với các ngôn ngữ lập trình như C++, Python, Java và MATLAB.
* **Phân loại modun**

Các mô-đun phổ biến trong thư viện OpenCV cung cấp các chức năng khác nhau để xử lý ảnh và thị giác máy tính. Dưới đây là một số mô-đun quan trọng và định nghĩa của chúng:



* **Core**

Mô-đun cơ bản chứa các cấu trúc dữ liệu cơ bản và các hàm xử lý cơ bản như đọc và ghi ảnh, xử lý ma trận, tính toán điểm ảnh,… Chứa các cấu trúc dữ liệu cơ bản như Mat (là lớp chứa ma trận ảnh và dữ liệu pixel), Point, Size, Rect cũng như các hàm xử lý ma trận và điểm ảnh.

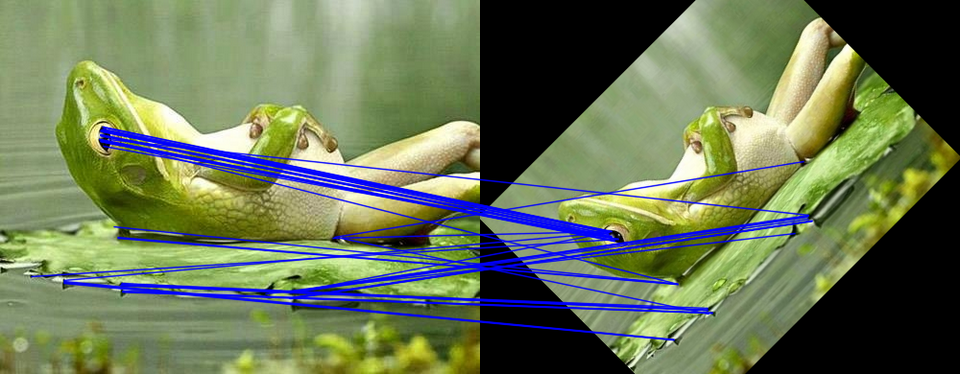
* **Imgproc**

Mô-đun này chứa các hàm xử lý ảnh căn bản như biến đổi hình thái học, lọc ảnh, chuyển đổi không gian màu. Bao gồm các hàm xử lý ảnh như làm mờ, làm sắc nét, biến đổi hình thái học, chuyển đổi không gian màu, cân bằng histogram, cắt ảnh, so sánh hình ảnh,…

Video

Cung cấp các công cụ và thuật toán để làm việc với dữ liệu video như theo dõi đối tượng, khám phá chuyển động. Chẳng hạn như các hàm và công cụ cho xử lý video như theo dõi đối tượng, phát hiện chuyển động, tạo video từ ảnh, xử lý dữ liệu video,…

* **Features2d**



Được sử dụng để tìm kiếm các keypoint trên ảnh và tính toán các descriptors để sử dụng trong ứng dụng nhận dạng. Điển hình như các thuật toán như SIFT, SURF, ORB và các công cụ để tìm keypoint, tính toán descriptors cho việc nhận dạng hình ảnh.

* **Objdetect**

Cung cấp các công cụ và thuật toán để phát hiện và theo dõi đối tượng trong ảnh và video. Cụ thể như các công cụ và thuật toán để phát hiện khuôn mặt, đối tượng, và theo dõi chuyển động trong ảnh và video.

* **Ứng dụng**

**Nhận diện khuôn mặt**

OpenCV cung cấp các thuật toán mạnh mẽ để nhận diện và phân tích khuôn mặt trong ảnh, video. Công nghệ sẽ được sử dụng trong các ứng dụng nhận diện khuôn mặt tự động, nhận dạng người dùng và các hệ thống an ninh.

**Xác định đối tượng**

OpenCV cung cấp các công cụ để xác định và phân tích các đối tượng trong ảnh, video. Công cụ đã trở thành một phần quan trọng của các ứng dụng như theo dõi vật thể, nhận diện biển số xe và quản lý hàng hóa.

**Xử lý ảnh y tế**

Công nghệ OpenCV còn được sử dụng để phân tích và xử lý hình ảnh y tế. Một số hoạt động tiêu biểu bao gồm việc phát hiện và phân tích các dấu hiệu bất thường trong hình ảnh y khoa, tạo ảnh MRI. Nền tảng được tích hợp với nhiều ứng dụng khác trong lĩnh vực y tế.

# **CHƯƠNG 3: KẾT QUẢ THỰC NGHIỆM**

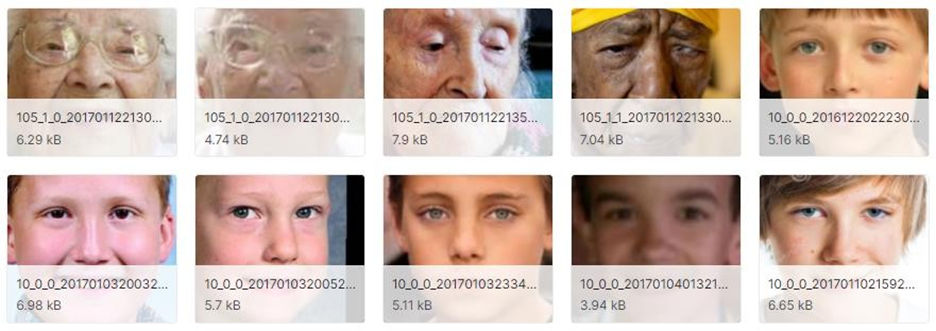
## **3.1. Dữ liệu và xử lý dữ liệu**

### **3.1.1. Dữ liệu**

Bởi vì hệ thống có cấu trúc như trên nên mục tiêu lấy dữ liệu ra để xử lý hay học máy thì em đã tìm kiểu được tệp tài liệu dưới đây.

Mục đích chính là phát hiện tuổi và giới tính thông qua bộ dữ liệu nhất định. Chúng ta sẽ sử dụng các phương pháp python và Keras đơn giản để phát hiện tuổi và giới tính. Tập dữ liệu có thể được tải xuống từ [UTKFace](https://www.kaggle.com/jangedoo/utkface-new).

Bộ dữ liệu UTKFace là bộ dữ liệu khuôn mặt quy mô lớn với độ tuổi dài (từ 0 đến 116 tuổi). Bộ dữ liệu bao gồm hơn 20.000 hình ảnh khuôn mặt có chú thích về độ tuổi, giới tính và dân tộc. Các hình ảnh bao gồm sự thay đổi lớn về tư thế, biểu cảm khuôn mặt, độ sáng, độ che khuất, độ phân giải, v.v. Bộ dữ liệu này có thể được sử dụng cho nhiều tác vụ khác nhau, ví dụ như phát hiện khuôn mặt, ước tính độ tuổi, tiến trình/hồi quy tuổi, định vị mốc, v.v.



Nó chứa 23k+ dữ liệu khuôn mặt từ nhiều lứa tuổi, chủng tộc và giới tính khác nhau.

### **3.1.2. Xử lý dữ liệu**

* **Tiền xử lý ảnh**

Nhận diện khuôn mặt: Giai đoạn đầu tiên thường là xác định vị trí khuôn mặt trong hình ảnh. Điều này có thể được thực hiện bằng các thuật toán phát hiện khuôn mặt được đào tạo trước.

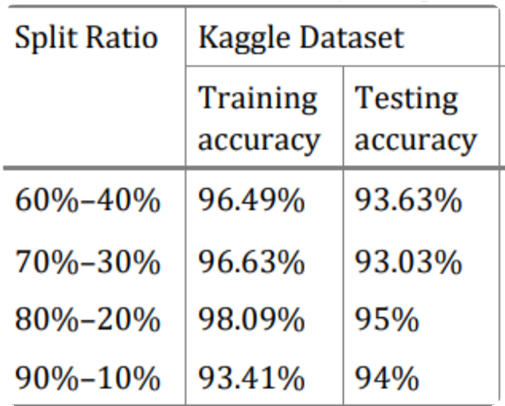
Tiền xử lý: Sau khi phát hiện khuôn mặt, hình ảnh sẽ được tiền xử lý để chuẩn hóa kích thước, độ tương phản và các đặc trưng khác, giúp cho mô hình học máy dễ dàng xử lý hơn.

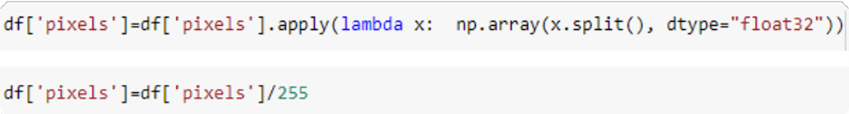
* **Xử lý dữ liệu**

Chúng ta sẽ chuyển đổi tất cả các cột thành một mảng bằng cách sử dụng np.array và thành dtype float. Sau đó, chúng ta sẽ chia tập dữ liệu thành xTrain, yTrain, yTest và xtest. Cuối cùng, chúng tôi sẽ áp dụng mô hình tuần tự và kiểm tra các dự đoán.

Cụ thể, trước tiên, chúng ta đọc tệp CSV chứa năm cột tuổi, dân tộc, giới tính, img\_name và pixel bằng cách sử dụng gấu trúc read\_csv hàm. Năm hàng đầu tiên có được bằng cách sử dụng phương thức DataFrame.head(). Chúng tôi đã chuyển đổi các pixel có tên cột thành một mảng bằng cách sử dụng thư viện NumPy và Định hình lại chúng thành kích thước 64, 64 bằng cách sử dụng openCv. Chúng tôi cũng đã chuyển đổi các giá trị trong float.

Chúng tôi chia thêm các giá trị cho 255.





Sau đó tôi nhóm các lứa tuổi thành 4 lớp:



## **3.2. Đào tạo mô hình**

* **Mô hình lớp đầu vào**

Mô hình đầu vào bao gồm một lớp đầu vào, 3 lớp tích chập, 2 lớp MaxPooling, một lớp bỏ học và lớp phẳng.

Các hình ảnh có kích thước 64 x 64. Bạn chuyển đổi ma trận hình ảnh thành một mảng, thay đổi tỷ lệ nó trong khoảng từ 0 đến 1, định hình lại nó để nó có kích thước 64 x 64 x 1 và cung cấp nó làm đầu vào cho mạng.

Chúng tôi đang sử dụng hai lớp 2D tích chập:

* Lớp đầu tiên sẽ có bộ lọc 32-3 x 3,
* Lớp thứ hai sẽ có bộ lọc 64-3 x 3 và
* Ngoài ra, có hai lớp gộp tối đa mỗi lớp có kích thước 2 x 2.
* Ngoài ra, mô hình còn chứa các lớp Flatten và Dropout trong mô-đun tuần tự để mô hình được đào tạo tốt nhất cho những gì được yêu cầu.

A screen shot of a computer program

Description automatically generated

Hình 2.4: Mô hình lớp đầu vào

* **Mô hình tuổi**

Mô hình tuổi có một chuỗi các lớp bỏ học và dày đặc và giảm các lớp thần kinh trong mỗi bước.

Chúng tôi sử dụng kích hoạt relu vì age\_detection liên quan đến phân loại đa lớp

A screenshot of a computer program

Description automatically generated

Hình 2.5: Mô hình tuổi

* **Mô hình giới tính**

Mô hình giới tính có một chuỗi các lớp bỏ học và dày đặc và giảm các lớp thần kinh trong mỗi bước.

Chúng tôi sử dụng kích hoạt sigmoid vì gender\_detection chỉ liên quan đến hai phân loại lớp duy nhất (M & F)

A screenshot of a computer program

Description automatically generated

Hình : Mô hình giới tính

Cuối cùng, cả ba mô hình đều được tạo ra để hoạt động đồng bộ dưới một mô hình duy nhất:



* **Dự đoán tuổi và giới tính**

Sau khi xử lý trước dữ liệu thô (hình ảnh), dữ liệu đã xử lý tức là. Vectơ tính năng phù hợp với mô hình đã đào tạo của chúng ta. Vì vậy, độ tuổi và giới tính được ước tính cho phù hợp.

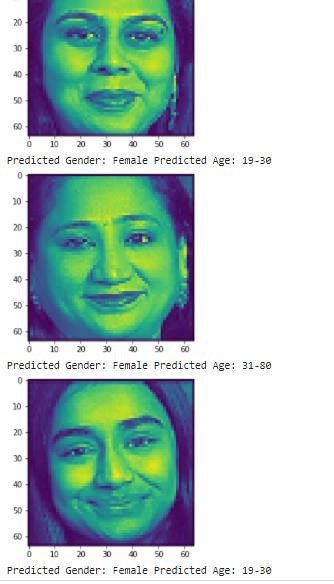
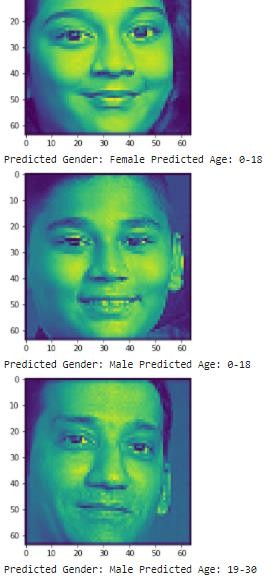
Chúng tôi phân loại tuổi dựa trên vallue biến dạng được cung cấp bởi mô hình được đào tạo.

*A screenshot of a computer program

Description automatically generated*

*Hình 2.7: Dự đoán độ tuổi và giới tính*

Dự đoán cuối cùng về tuổi và giới tính của các khuôn mặt được phát hiện từ hình ảnh nguồn đã cho.



*Hình 2.8: Hình ảnh được dự đoán*

## **3.3. Kết quả thực nghiệm**

Với trình độ của em thì em đã xây dựng được trương trình bề nổi là wepapp. Mô hình được đào tạo được tích hợp với giao diện người dùng bằng cách sử dụng khung Flask.

* **Giao diện chính**

A screenshot of a computer

Description automatically generated

*Hình 3.1: Giao diện chính*

* **Giao diện sau khi tải ảnh lên**

**A screenshot of a computer

Description automatically generated**

*Hình 3.2: Giao diện up ảnh*

* **Giao diện loading**

A screenshot of a computer

Description automatically generated

*Hình 3.3: Giao diện loading*

* **Giao diện khi xuất ra kết quả**

**A screenshot of a computer

Description automatically generated**

*Hình 3.4: Giao diện hoàn tất phân tích*

**Đánh giá kết quả:**

Chúng tôi đã giải quyết việc phân loại nhóm tuổi và giới tính của hình ảnh khuôn mặt trong thế giới thực chưa được lọc. Chúng tôi đặt ra nhiệm vụ như một vấn đề phân loại đa lớp và do đó, đào tạo mô hình với hàm tổn thất dựa trên phân loại làm mục tiêu đào tạo.

Thuật toán tiền xử lý hình ảnh, xử lý một số biến thể được quan sát thấy ở các khuôn mặt trong thế giới thực không được lọc điển hình và điều này xác nhận khả năng áp dụng mô hình cho nhóm tuổi và phân loại giới tính trong tự nhiên.

Cuối cùng, chúng tôi điều tra độ chính xác phân loại theo độ tuổi và giới tính; Phương pháp được đề xuất của chúng tôi đạt được hiệu suất hiện đại, ở cả nhóm tuổi và phân loại giới tính. Đối với các công trình trong tương lai, chúng tôi sẽ xem xét kiến trúc CNN sâu hơn và thuật toán xử lý hình ảnh mạnh mẽ hơn để ước tính tuổi chính xác. Ngoài ra, ước tính tuổi rõ ràng của khuôn mặt con người sẽ là nghiên cứu thú vị để điều tra trong tương lai.

# **KẾT LUẬN**

* **Kết quả đạt được**

Hệ thống đã thành công trong việc nhận diện giới tính và độ tuổi qua việc phân tích các đặc điểm khuôn mặt. Các thử nghiệm cho thấy hệ thống có thể chính xác trong việc xác định giới tính và độ tuổi của người dùng dựa trên các dấu hiệu khuôn mặt và các đặc điểm sinh lý. Hệ thống đưa ra kết quả nhanh chóng và đáng tin cậy, đồng thời cung cấp cảnh báo âm thanh hoặc thông báo về độ tuổi và giới tính khi cần thiết.

Hệ thống hoạt động ổn định trong điều kiện ánh sáng khác nhau và với các góc nhìn khác nhau của khuôn mặt**.**

* **Kết quả chưa đạt được**

Khả năng xử lý trong điều kiện ánh sáng cực đoan: Mặc dù hệ thống hoạt động tốt trong điều kiện ánh sáng bình thường, nhưng nó vẫn gặp khó khăn trong các điều kiện ánh sáng quá yếu hoặc quá sáng. Đặc biệt, việc nhận diện giới tính và độ tuổi có thể bị ảnh hưởng khi ánh sáng quá yếu hoặc quá sáng, khiến việc nhận diện khuôn mặt gặp khó khăn.

Hiệu suất với các vật cản và góc nhìn lệch: Hệ thống có thể gặp khó khăn trong việc nhận diện giới tính và độ tuổi khi có vật cản hoặc khi người dùng quay đầu quá nhiều, gây ảnh hưởng đến khả năng phân tích khuôn mặt.

Đáp ứng với các tình huống đặc biệt: Trong một số tình huống như người đeo kính râm hoặc có đặc điểm khuôn mặt không điển hình, khả năng nhận diện có thể bị ảnh hưởng, đặc biệt là trong việc xác định độ tuổi và giới tính chính xác.

* **Dự kiến phát triển và cải tiến thuật toán**

Áp dụng thuật toán vào các lĩnh vực khác nhau: Mở rộng ứng dụng của thuật toán nhận diện giới tính và độ tuổi vào các lĩnh vực như giáo dục (nhận diện độ tuổi học sinh), chăm sóc khách hàng, hoặc các ứng dụng về an ninh.

Cải thiện khả năng nhận diện trong điều kiện ánh sáng cực đoan: Nghiên cứu và áp dụng các kỹ thuật xử lý ảnh tiên tiến như cân bằng ánh sáng tự động và cải thiện thuật toán nhận diện khuôn mặt để hoạt động tốt hơn trong các điều kiện ánh sáng khác nhau, từ đó nâng cao độ chính xác trong việc nhận diện giới tính và độ tuổi.

Tăng cường độ chính xác với vật cản và góc nhìn lệch: Phát triển và áp dụng các mô hình học máy mạnh mẽ hơn, như mạng nơ-ron sâu (Deep Learning), để nhận diện giới tính và độ tuổi chính xác hơn trong các điều kiện không lý tưởng, bao gồm khi có vật cản hoặc góc nhìn lệch.

Tích hợp cảm biến phụ trợ: Kết hợp các cảm biến khác như cảm biến chuyển động hoặc cảm biến nhịp tim để nâng cao khả năng xác định các yếu tố ảnh hưởng đến giới tính và độ tuổi (ví dụ: giúp phân biệt người lớn và trẻ em trong môi trường đông đúc).

Nâng cao khả năng tùy chỉnh: Cung cấp các tùy chọn điều chỉnh nâng cao cho người dùng, bao gồm khả năng cá nhân hóa các tham số nhận diện (như các yếu tố liên quan đến độ tuổi và giới tính) dựa trên các đặc điểm khuôn mặt và môi trường.

Tăng cường hiệu suất và tính tương thích: Tối ưu hóa thuật toán để giảm thiểu mức tiêu thụ tài nguyên và nâng cao hiệu suất xử lý thời gian thực. Đảm bảo hệ thống tương thích với các loại camera và phần cứng khác nhau.

Thực hiện thử nghiệm và đánh giá mở rộng: Tiến hành thử nghiệm trong các điều kiện và môi trường khác nhau để xác định điểm yếu và cải thiện tính chính xác của hệ thống trong việc nhận diện giới tính và độ tuổi.

# **TÀI LIỆU THAM KHẢO**

[1]. G. Levi và T. Hassner, "Phân loại tuổi và giới tính bằng cách sử dụng mạng nơ-ron tích chập," Hội thảo IEEE về Phân tích và Mô hình hóa khuôn mặt và cử chỉ (AMFG), IEEE Conf. về Nhận dạng mẫu và Thị giác Máy tính (CVPR), Boston, 2015.

[2]. Eugenio Culurciello (19/4/2017) “The History of Neural Networks”.

[3]. Michael A. Nielsen (2015) “Neural Network and Deep Learning”.

[4]. XLA\_C2\_3.pdf, Trường Đại Học Công Nghệ Đông Á

[5]. Phân lớp DL.pdf, Trường Đại Học Công Nghệ Đông Á

[6]. Hou, Qiqi, et al. "Facial landmark detection via cascade multi-channel

convolutional neural network." 2015 IEEE International Conference on Image

Processing (ICIP). IEEE, 2015.

[7]. Xiao, Shengtao, Shuicheng Yan, and Ashraf A. Kassim. "Facial landmark detection via progressive initialization." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.

[8]. Altenberger, Felix, and Claus Lenz. "A non-technical survey on deep convolutional neural network architectures." arXiv preprint arXiv:1803.02129 (2018).

[9]. Cao, Changyu et al. “A convolutional neural network face recognition algorithm based on data augmentation.” (2019).

[10]. Wang Q., Xiong D., Alfalou A., Brosseau C. Optical image authentication scheme using dual polarization decoding configuration. Opt. Lasers Eng. 2019;112:151–161. doi: 10.1016/j.optlaseng.2018.09.008.

[11]. Vinay A., Hebbar D., Shekhar V.S., Murthy K.B., Natarajan S. Two novel detector-descriptor based approaches for face recognition using sift and surf. Procedia Comput. Sci. 2015;70:185–197