

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Nguyễn Công Hòa

**NGHIÊN CỨU HỌC SÂU TRONG NHẬN DẠNG KHUÔN MẶT
ỨNG DỤNG CHO BÀI TOÁN ĐIỂM DANH TỰ ĐỘNG HỌC SINH**

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 8.48.01.01

TÓM TẮT LUẬN VĂN THẠC SĨ

HÀ NỘI - 2020

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. PHAN THỊ HÀ

Phản biện 1:

Phản biện 2:

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ
tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: giờ ngày tháng năm

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

MỞ ĐẦU

Đảng, Nhà Nước ta xác định “Giáo dục và đào tạo là quốc sách hàng đầu” và để thực hiện tốt mục tiêu, nhiệm vụ chiến lược này, việc ưu tiên ứng dụng công nghệ thông tin trong quản lý và trong việc hỗ trợ các hoạt động dạy - học là một vấn đề rất cấp bách, đã được thể hiện qua rất nhiều văn bản chỉ đạo của Đảng, Nhà Nước và của Bộ giáo dục và đào tạo.

Việc quản lý học sinh trong các nhà trường phổ thông hiện nay hoàn toàn dựa theo hình thức thủ công, hiện tượng học sinh bỏ cả buổi học, bỏ tiết, ngồi học không đúng lớp mình học, học sinh không phải của nhà trường ... gây ra cho công tác kiểm diện và quản lý học sinh gặp nhiều khó khăn. Hơn nữa, phụ huynh cũng muốn giám sát xem con mình có mặt ở trường, ở lớp hay không? hiện đang là một nhu cầu rất lớn.

Học sâu đã và đang rất phát triển, được ứng dụng rộng rãi trong các bài toán nhận dạng như: nhận dạng hình ảnh, nhận dạng giọng nói, xử lý ngôn ngữ tự nhiên ... và thu được những thành tựu to lớn với độ chính xác ngày càng cao. Trong đó nhận dạng khuôn mặt để xác định danh tính, giao dịch, kiểm soát an ninh ... ngày càng trở nên phổ biến.

Xuất phát từ thực tế trên, đề tài “*nguyên cứu Học sâu trong nhận dạng khuôn mặt ứng dụng cho bài toán điểm danh tự động học sinh*” với hy vọng có thể ứng dụng thành công mô hình Học sâu hiện đại trong việc xây dựng hệ thống điểm danh tự động dựa vào nhận dạng khuôn mặt, đặc biệt là ứng dụng cụ thể vào điểm danh tự động học sinh trong mỗi lớp học của trường THPT Thanh Oai B, huyện Thanh Oai, Hà Nội.

Mục tiêu của luận văn là nghiên cứu học sâu trong nhận dạng khuôn mặt ứng dụng cho bài toán điểm danh tự động đối với quá trình quản lý học sinh trong các nhà trường phổ thông hiện nay nhằm nâng cao chất lượng quản lý học sinh trong công tác giáo dục đào tạo của nhà trường.

Đối tượng nghiên cứu của luận văn: Mô hình mạng nơron tích chập và bài toán điểm danh tự động.

Phạm vi nghiên cứu của luận văn: Xây dựng hệ thống điểm danh tự động đối với học sinh trong lớp học tại trường THPT Thanh Oai B, huyện Thanh Oai, Hà Nội.

Nội dung của luận văn được trình bày trong ba chương với nội dung chính như sau:

Chương 1: Tổng quan về bài toán điểm danh tự động

Nội dung chính của chương 1 là tìm hiểu khái quát về xử lý ảnh và bài toán nhận dạng khuôn mặt.

Chương 2: Tìm hiểu về học sâu và mô hình mạng nơron tích chập

Nội dung chính của chương 2 là tìm hiểu tổng quan về học máy, Học sâu, mô hình mạng nơron tích chập (CNN) cũng như cách hoạt động, cấu trúc và việc huấn luyện của mô hình mạng nơron tích chập.

Chương 3: Nhận dạng khuôn mặt ứng dụng cho bài toán điểm danh tự động

Nội dung chính của chương 3 là trình bày chi tiết các bước xây dựng hệ thống điểm danh tự động dựa trên nhận dạng khuôn mặt.

Chương 1. TỔNG QUAN VỀ BÀI TOÁN ĐIỂM DANH TỰ ĐỘNG

Chương này tập trung vào những khái niệm về xử lý hình ảnh, giới thiệu về bài toán nhận dạng khuôn mặt bao gồm các ứng dụng, tầm quan trọng và những khó khăn trở ngại hiện nay khi áp dụng bài toán về nhận dạng khuôn mặt, và cuối cùng những phương pháp nhận dạng khuôn mặt hiện nay. Chương này cũng chỉ ra vai trò và tầm quan trọng của bài toán điểm danh tự động học sinh tại trường THPT Thanh Oai B, Huyện Thanh Oai, Hà Nội.

1.1. Tổng quan về xử lý ảnh

1.1.1. Một số khái niệm.

1.1.2. Các vấn đề của xử lý ảnh

- a) Điều chỉnh mức xám của ảnh
- b, Trích chọn đặc điểm
- c, Nhận dạng:
- d, Nén ảnh

1.1.3. Ứng dụng của xử lý ảnh trong thực tế.

- ❖ Xử lý và phục hồi hình ảnh
- ❖ Lĩnh vực y tế
- ❖ UV imaging
- ❖ Truyền và mã hóa
- ❖ Thị giác máy tính và robot
- ❖ Phát hiện vật cản
- ❖ Công nghệ nhận dạng

1.2. Bài toán nhận dạng khuôn mặt

1.2.1. Khái niệm.

1.2.2. Một số trở ngại của công nghệ nhận dạng khuôn mặt

- ❖ Góc chụp khuôn mặt: Chụp thẳng, chụp nghiêng, chụp hất lên ...
- ❖ Một số thành phần xuất hiện thêm hoặc không xuất hiện trên khuôn mặt như: đeo kính, đeo khẩu trang, trang điểm, mọc râu ... làm cho việc nhận dạng khuôn mặt thiếu chính xác.
- ❖ Khi con người thể hiện sự biểu cảm như: cười, khóc, nhăn mặt ... cũng ảnh hưởng đến kết quả nhận dạng.

❖ Ngoài ra một số tác nhân khác cũng gây ảnh hưởng đến kết quả như: Ảnh quá sáng, quá mờ, chất lượng ảnh ...

❖ Nền ảnh phức tạp



Hình 1.1: Nền ảnh phức tạp

❖ Màu sắc của da mặt

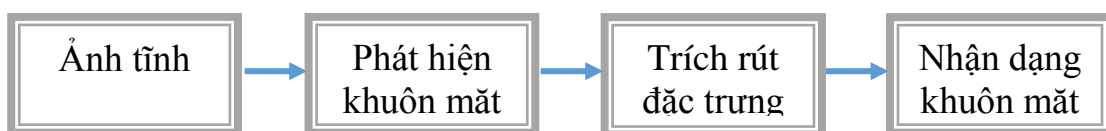
1.2.3. Tầm quan trọng của bài toán nhận dạng khuôn mặt

1.2.4. Các ứng dụng đặc trưng của bài toán nhận dạng khuôn mặt

- Ứng dụng trong giám sát an ninh
- Ứng dụng trong các ngành bán lẻ, dịch vụ
- Ứng dụng trong doanh nghiệp, công sở
- Ứng dụng trong chính phủ
- Ứng dụng trong trường học
- Ứng dụng trong lĩnh vực Y tế - sức khỏe
- Ứng dụng trong các thiết bị IOT
- Phân tích cảm xúc

1.2.5. Xây dựng hệ thống nhận dạng khuôn mặt.

Một hệ thống nhận dạng khuôn mặt có thể khái quát chung gồm có 3 bước cơ bản sau:



Hình 1.2: Hệ thống nhận dạng khuôn mặt

1.2.6. Một số phương pháp nhận dạng khuôn mặt

Dựa vào các tiêu chí mà người ta chia ra thành nhiều phương pháp nhận dạng khuôn mặt nhưng phổ biến hiện nay là các loại sau:

- + Phương pháp tiếp cận toàn cục.
- + Phương pháp tiếp cận dựa trên các đặc điểm cục bộ.

+ Phương pháp lai.

1.3. Vai trò và tầm quan trọng của bài toán điểm danh tự động học sinh tại trường THPT Thanh Oai B, Huyện Thanh Oai, Hà Nội.

Điểm danh là công việc được tiến hành hàng ngày và thường xuyên trong các buổi học tại các nhà trường phổ thông hiện nay trong đó có trường THPT Thanh Oai B, huyện Thanh Oai, Hà Nội.

❖ Thực trạng:

- + Học sinh thường đến lớp muộn, nghỉ học hoặc bỏ học cả buổi hoặc bỏ tiết học.
- + Hình thức điểm danh thủ công: Giáo viên đầu tiết học đếm sĩ số lớp ghi tên những học sinh vắng tiết học đó vào Sổ đầu bài.
- + Cuối tuần, cuối tháng, cuối học kỳ, cuối năm học giáo viên chủ nhiệm phải tổng hợp ngày nghỉ của học sinh để xếp thi đua và xét lên lớp cho học sinh. Công việc này rất mất thời gian, thiếu khách quan, thiếu chính xác.

❖ Giải pháp:

Cần có một hệ thống điểm danh tự động, mỗi phòng lắp một thiết bị điểm danh tại mỗi cửa phòng học. Giáo vụ, Ban giám hiệu có thể theo dõi kết quả điểm danh học sinh vào bất kể thời gian nào trong buổi học. Cuối tháng, cuối kỳ tổng hợp và gửi danh sách cho giáo viên chủ nhiệm.

1.4. Kết luận chương

Trong chương 1, luận văn đã trình bày khái quát về xử lý ảnh và đặc biệt đã trình bày tương đối chi tiết về nhận dạng khuôn mặt ứng dụng cho bài toán của luận văn “*Điểm danh tự động học sinh*”. Trong chương tiếp theo, luận văn sẽ trình bày sơ lược về Học máy và hướng người đọc đến phần quan trọng là Học sâu.

Chương 2. TÌM HIỂU VỀ HỌC SÂU VÀ MÔ HÌNH MẠNG NƠON TÍCH CHẬP

Kỹ thuật Học sâu là một phạm trù nhỏ của lĩnh vực Học máy, Học sâu tập trung giải quyết các vấn đề liên quan đến mạng thần kinh nhân tạo (Artificial Neural Network - ANN) nhằm nâng cấp các công nghệ như nhận dạng giọng nói, thị giác máy tính và xử lý ngôn ngữ tự nhiên. Bởi vậy trong chương này, luận văn sẽ trình bày khái quát về Học máy, đi sâu vào kỹ thuật Học sâu cùng một số thuật toán và ứng dụng của nó trong thực tế, đồng thời cũng trình bày chi tiết về mô hình mạng nơon tích chập (CNN) cũng như cách hoạt động, cấu trúc và việc huấn luyện mô hình CNN.

2.1. Tổng quan về Học máy (Machine learning)

Học máy là một công nghệ phát triển từ lĩnh vực trí tuệ nhân tạo. Các thuật toán Học máy là các chương trình máy tính có khả năng học hỏi về cách hoàn thành các nhiệm vụ và cách cải thiện hiệu suất theo thời gian.

2.2. Các thuật toán Học máy

2.2.1. Học có giám sát (supervised learning)

2.2.2. Học không giám sát (unsupervised learning)

2.2.3. Học bán giám sát (Semi-Supervised Learning)

2.2.4. Học củng cố (Reinforcement learning)

2.3. Tìm hiểu về Học sâu (Deep learning)

2.3.1. Học sâu là gì?

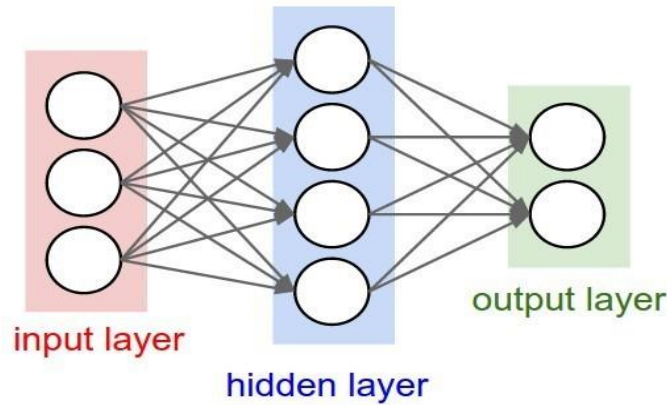
2.3.2. Lịch sử Học sâu

2.3.3. Tổng quan về mạng nơon nhân tạo

a) Tìm hiểu về mạng nơon nhân tạo.

b) Kiến trúc ANN

Kiến trúc chung của ANN được mô tả trong hình 2.4 gồm 3 lớp: Lớp đầu vào (Input Layer), lớp ẩn (Hidden Layer) và lớp đầu ra (Output Layer).



Hình 2.1: Kiến trúc 3 phần của ANN

c) Huấn luyện ANN

- Lan truyền ngược (back-propagation) là phương pháp huấn luyện mạng ANN với mục tiêu xác định trọng số tối ưu cho mạng thông qua việc lặp đi lặp lại 2 quá trình: lan truyền tiến (tính giá trị đầu ra của mạng từ đó tính sai số giữa giá trị này với giá trị mong muốn). Tiếp theo là quá trình lan truyền ngược sai số (dựa vào sai số sẽ cập nhật lại các trọng số).

+) Lan truyền tiến: Dữ liệu từ tập huấn luyện thông qua tầng nhập sẽ được chuyển vào tầng tiếp theo. Tại mỗi neural của mỗi tầng sẽ tiến hành thực hiện việc tính toán thông qua các hàm kết hợp, giá trị này sau khi truyền qua hàm kích hoạt là đầu ra mỗi neural.

+) Lan truyền ngược sai số: Dựa trên sai số được tính từ quá trình lan truyền tiến, mạng sẽ cập nhật lại các trọng số theo nguyên tắc lan truyền ngược sai số. Trong đó kỹ thuật cơ bản được áp dụng trong quá trình cập nhật trọng số đó là gradient descent.

Như vậy, để huấn luyện hay để một mạng học từ dữ liệu thì mạng thực hiện 2 bước lan truyền tiến và lan truyền ngược sai số. Quá trình này thực hiện cho tới khi sai số đạt được một ngưỡng nào đó hoặc thực hiện qua số bước lặp được người huấn luyện mạng đặt ra.

d) Thuật toán lan truyền ngược (Back – Propagation)

e) Giảm lỗi cho mạng

❖ **Underfitting**

❖ **Overfitting** (quá khớp)

2.3.4. Ứng dụng của Học sâu

a) Nhận dạng giọng nói

b) Dịch thuật

c) Ô tô tự vận hành hay máy bay không người lái.

- d) Tô màu, phục chế hình ảnh
- e) Nhận dạng khuôn mặt
- f) Y học và dược phẩm
- g) Dịch vụ giải trí và mua sắm

2.4. Tìm hiểu về CNN [2]

2.5. Cấu trúc của CNN

CNN là một kiểu mạng ANN truyền thẳng, trong đó kiến trúc chính gồm nhiều thành phần được ghép nối với nhau theo cấu trúc nhiều lớp đó là: Convolution, Pooling, ReLU và Fully connected.

2.5.1. Lớp tích chập (Convolution)

2.5.2. Lớp phi tuyến Relu

2.5.3. Lớp Pooling

2.5.4. Lớp Fully-connected (FC)

2.6. Huấn luyện mô hình CNN

2.7. Tìm hiểu về Multi-task Cascaded Convolutional Networks

2.7.1. Multi-task Cascaded Convolutional Networks là gì?

Multi-task Cascaded Convolutional Networks được viết tắt là MTCNN. Nó bao gồm 3 mạng CNN xếp chồng và đồng thời hoạt động khi detect khuôn mặt. Mỗi mạng có cấu trúc khác nhau và đảm nhiệm vai trò khác nhau trong task. Đầu ra của MTCNN là vị trí khuôn mặt và các điểm trên mặt như: mắt, mũi, miệng...

2.7.2. MTCNN Workflow

- a. Mạng P-Net
- b. Mạng R-Net
- c. Mạng O-Net

2.7.3. Lý do lựa chọn MTCNN để detect khuôn mặt

2.8. Tìm hiểu về mô hình ResNet

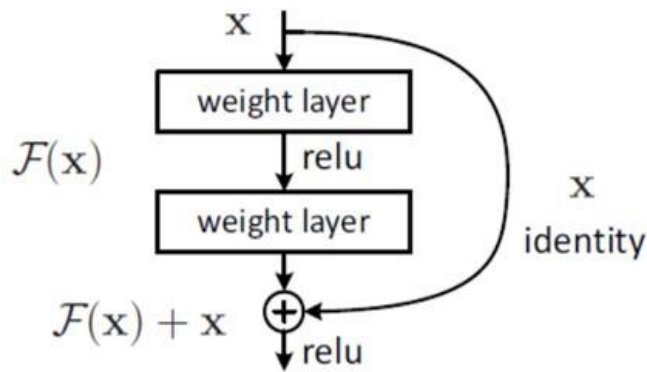
2.8.1. Giới thiệu về mô hình ResNet

2.8.2. Điểm nổi bật của mô hình ResNet

2.8.3. Kiến trúc ResNet

ResNet có kiến trúc gồm nhiều residual block, ý tưởng chính là skip layer bằng cách thêm kết nối với layer trước. Ý tưởng của residual block là feed forward $x(\text{input})$ qua một

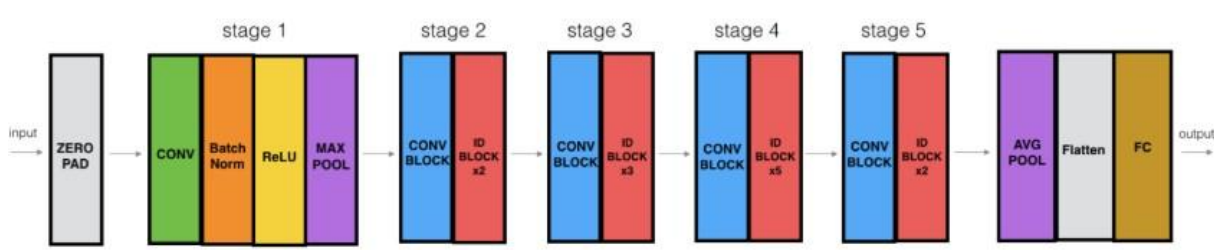
số layer conv-max-conv, ta thu được $F(x)$ sau đó thêm x vào $H(x) = F(x) + x$. Model sẽ dễ học hơn khi chúng ta thêm feature từ layer trước vào.



Hình 2.2: Một khối xây dựng của ResNet

2.8.4. Mô hình ResNet

Mô hình chi tiết ResNet-101.



Hình 2.3: Mô hình ResNet-101

2.9. Kết luận chương

Trong chương 2, luận văn đã trình bày 2 nội dung chính sau:

- + Khái quát về học máy, đặc biệt là Học sâu. Qua đó luận văn cũng đã trình bày sơ lược về một số mô hình mạng Học sâu cũng như ứng dụng của chúng.

- + Trình bày tương đối chi tiết về mô hình, hoạt động của mạng nơron tích chập CNN cũng như cách xây dựng nó. Bên cạnh đó, luận văn cũng đã trình bày về hai mô hình CNN chính sẽ sử dụng để xây dựng hệ thống nhận dạng khuôn mặt là MTCNN và ResNet.

Sau khi đã tìm hiểu về học máy cũng như cách xây dựng và đánh giá mô hình học máy và những nội dung kiến thức liên quan đến xử lý hình ảnh bằng mạng CNN, trong chương tiếp theo luận văn sẽ trình bày chi tiết quá trình xây dựng hệ thống nhận dạng khuôn mặt ứng dụng bài toán điểm danh tự động học sinh và xây dựng ứng dụng quản lý việc điểm danh bằng phần mềm.

Chương 3. NHẬN DẠNG KHUÔN MẶT ỨNG DỤNG CHO BÀI TOÁN ĐIỂM DANH TỰ ĐỘNG

Nội dung chương này tập trung vào xây dựng mô hình nhận dạng khuôn mặt, phương pháp huấn luyện và đánh giá mô hình. Bên cạnh đó chương cũng sẽ giới thiệu về các công nghệ được sử dụng, phương pháp xây dựng bộ dữ liệu huấn luyện. Cuối cùng là trình bày về nghiên cứu và thiết kế phần cứng cho thiết bị điểm danh và cách tối ưu luồng xử lý.

3.1. Xây dựng hệ thống nhận dạng khuôn mặt

3.1.1. Công nghệ sử dụng

- MXNet
- SQLite
- Thuật toán k-NN (K-Nearest Neighbors)

Đầu ra của mô hình nhận dạng khuôn mặt sử dụng mạng Resnet sau khi bỏ tầng Sigmoid activation là một vector 128 chiều, nghĩa là mỗi khuôn mặt sẽ được trích xuất các đặc điểm và tổng hợp trong vector này. Đối với cơ sở dữ liệu khuôn mặt của học sinh, mỗi học sinh sẽ có 3-4 ảnh chụp khuôn mặt của mình trong hệ thống. Vấn đề đặt ra là sử dụng thuật toán nào để so khớp khi đưa một ảnh chụp khuôn mặt bất kỳ vào phần mềm và xác định được đó là khuôn mặt của ai. Để giải quyết vấn đề này, tôi nhận thấy thuật toán k-NN khá hiệu quả và đảm bảo được yêu cầu đề ra.

K-NN thực chất là thuật toán để phân lớp đối tượng dựa vào khoảng cách gần nhất giữa các đối tượng cần xếp lớp. Tập dữ liệu huấn luyện là các điểm dữ liệu đã được gán nhãn, khi đưa vào một điểm dữ liệu cần phân lớp, thuật toán sẽ tính độ giống giữa điểm dữ liệu này với các điểm dữ liệu trong bộ dữ liệu huấn luyện, sau đó lấy trung bình độ giống với lớp dữ liệu tương ứng. Giá trị trung bình này lớn nghĩa là điểm dữ liệu này gần với các điểm dữ liệu trong lớp đó, khả năng thuộc vào lớp đó sẽ cao. Và giá trị trung bình lớn nhất nghĩa là điểm dữ liệu đó có khả năng thuộc vào lớp đó nhất. Tuy nhiên ta cũng cần phải xác định một ngưỡng tối thiểu (threshold) để đảm bảo một điểm dữ liệu có thể không thuộc vào lớp nào nếu nằm dưới ngưỡng tối thiểu đó.

Áp dụng k-NN vào trong bài toán này, mỗi học sinh trong cơ sở dữ liệu sẽ có 4 ảnh và mỗi học sinh sẽ được coi là một lớp dữ liệu, mỗi ảnh trong lớp dữ liệu sẽ được coi là một điểm dữ liệu. Như vậy mỗi lớp dữ liệu gồm tối đa 4 điểm dữ liệu. Lặp qua từng học sinh, ta sẽ tính độ giống vector đặc điểm của ảnh học sinh được đưa vào với vector đặc điểm của

từng khuôn mặt của học sinh trong cơ sở dữ liệu, sau đó lấy giá trị trung bình [3]. Kết thúc lặp ta sẽ có giá trị các độ giống trung bình của ảnh đưa vào với khuôn mặt từng học sinh. Dựa vào giá trị này ta sẽ xác định được khuôn mặt đó là của học sinh nào.

$$d_k = \frac{1}{n} \sum_{i=1}^n \text{similar}(q, y_i) \quad (3-1)$$

Trong đó:

d_k : Là trung bình độ giống của vector q với các khuôn mặt trong lớp k (trường hợp này $n = 4$)

q : vector đặc điểm của khuôn mặt đưa vào hệ thống

y_i : Vector khuôn mặt thứ i của lớp k .

Để đo độ giống (similar) giữa 2 vector ở đây tôi sử dụng khoảng cách Euclid:

$$\text{similar}(q, y) = \sqrt{\sum_{j=1}^m (q_j - y_j)^2} \quad (3-2)$$

Viết gọn lại:

$$d_k = \frac{1}{n} \sum_{i=1}^n \left(\sqrt{\sum_{j=1}^m (q_j - y_j)^2} \right) \quad (3-3)$$

Để xác định được khuôn mặt đưa vào là của học sinh nào, ta tiến hành lấy giá trị d_k lớn nhất theo công thức sau:

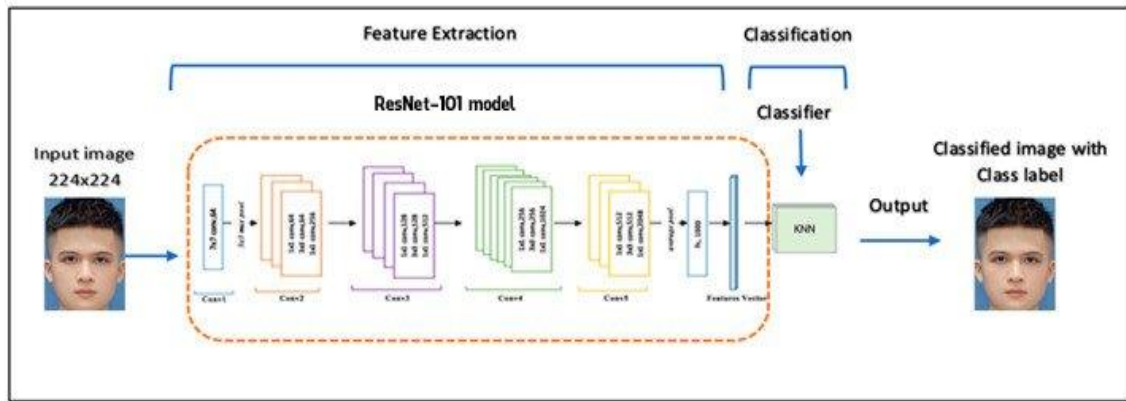
$$c = \max_i(d_i) \mid i = (1, C), d_i \leq \text{threshold} \quad (3-4)$$

Để tránh trường hợp một người không có trong cơ sở dữ liệu nhưng sử dụng hệ thống điểm danh, ta không thể lấy giá trị trung bình độ giống lớn nhất để kết luận đó là ai vì sẽ dẫn đến việc nhận nhầm mà còn phải so sánh với giá trị ngưỡng (threshold) để đảm bảo giá trị về độ giống luôn phải nhỏ hơn ngưỡng đó.

- **Thiết kế giao diện thiết bị bằng QtDesigner :**

3.1.2. Xây dựng hệ thống nhận dạng khuôn mặt

Luận văn sử dụng mô hình ResNet (đã tìm hiểu ở chương 2) để trích xuất đặc trưng khuôn mặt. Sau đây tôi sẽ trình bày về cách áp dụng mô hình Resnet- 101 để thực hiện việc nhận dạng khuôn mặt:



Hình 3.1: Các bước thực hiện nhận diện khuôn mặt sử dụng Resnet-101

Camera sẽ chụp ảnh học sinh sau đó ảnh sẽ được đưa qua mạng MTCNN để mạng này phát hiện vị trí và trích xuất khuôn mặt. Một ảnh đầu vào có thể có nhiều khuôn mặt. Ảnh khuôn mặt sẽ được chỉnh về kích thước chuẩn là 224 x 224px để giảm số chiều vector. Ảnh lúc này trở thành 1 ma trận 224 x 224 x 3 [2].

Sau đó tiếp tục đưa ma trận này vào mạng Resnet để trích xuất đặc điểm khuôn mặt. Sau khi qua mô hình ta sẽ nhận được một features vector 128 chiều. Với vector 128 chiều này ta sẽ sử dụng để làm căn cứ để xác định độ tương đồng giữa 2 khuôn mặt và đưa vào thuật toán phân lớp KNN để gán nhãn cho khuôn mặt [7].

3.1.3. Xây dựng dữ liệu huấn luyện

Dữ liệu huấn luyện được chia ra làm 2 loại: Dữ liệu thu thập từ internet và dữ liệu tự xây dựng.

Trong phạm vi luận văn này tôi sử dụng các bộ dữ liệu sau đây:

- **ORL**
- **Cơ sở dữ liệu khuôn mặt GTAV**
- **Cơ sở dữ liệu khuôn mặt Georgia Tech**

Face FEI

- **Các khuôn mặt được gán nhãn trong tự nhiên (LFW)**
- **Các khuôn mặt được dán nhãn được chuẩn hóa chính diện trong tự nhiên (F_LFW)**
- **Dữ liệu tự xây dựng (hình 3.2):** Dữ liệu tự xây dựng gồm 300 bức ảnh chụp học sinh khối 11 trường THPT Thanh Oai B với các góc chụp khác nhau như thẳng, nghiêng ($\pm 30^\circ$, $\pm 45^\circ$, $\pm 60^\circ$)

3.1.4. Huấn luyện mô hình nhận dạng khuôn mặt

a) Hàm mất mát

Để phân tích sự giống nhau giữa hai hình ảnh, chúng ta cần biến đổi hình ảnh đầu vào của mình thành một hình biểu diễn nhỏ hơn, chẳng hạn như một vector duy nhất. Biểu diễn này thường được gọi là embedding. Chúng ta cần xây dựng các embedding vector để chúng có các thuộc tính sau:

- Hai hình ảnh giống nhau tạo ra hai embedding vector và khoảng cách toán học giữa chúng là nhỏ.
- Hai hình ảnh rất khác nhau tạo ra hai embedding vector và khoảng cách toán học giữa chúng lớn.







Để làm được điều đó, chúng ta cần huấn luyện một mạng nơ-ron để tạo ra các embedding vector tốt chứa các thuộc tính này. Để so sánh khuôn mặt, 'hai hình ảnh giống nhau' mà tôi tham chiếu ở trên có thể là cùng một khuôn mặt trong hai ảnh khác nhau (tạo ra các embedding vector có khoảng cách gần) và 'hai hình ảnh rất khác nhau' có thể là hai khuôn mặt khác nhau, tạo ra các embedding vector khoảng cách lớn.

Như vậy số chiều của vector nhúng là bao nhiêu thì đủ? Ý tưởng chính của việc encode hình ảnh khuôn mặt là vector đầu ra phải có khả năng biểu diễn các thuộc tính đủ để phân biệt khuôn mặt này với khuôn mặt khác. Để giải quyết vấn đề này, chiều dài của vector được coi là một tham số cần tìm. Với mạng Resnet-50, tôi lựa chọn 128 số để biểu diễn các thuộc tính của khuôn mặt tương ứng với vector đầu ra là 128 chiều.

Có hai cách chính để tìm ra các tham số cho mạng CNN:

Thứ nhất, chúng ta có thể đơn giản coi phần bên phải của hệ thống như một bộ phân loại nhị phân với $Y = 1$ nếu các ảnh đầu vào thuộc cùng một lớp và $Y = 0$ nếu ảnh đầu vào không thuộc cùng một lớp.

X		Y
		1

		0
		1
		1

Hình 3.2: Mô tả phương pháp tính độ lỗi

Cách thứ 2, vì chúng ta muốn so sánh 2 hình ảnh và có khoảng cách giữa 2 embedding vector là nhỏ nếu 2 bức ảnh giống nhau và khoảng cách là lớn nếu 2 bức ảnh khác nhau [4].

Như vậy:

- Một bức ảnh đầu tiên, được gọi là Anchor (neo)
- Một hình ảnh từ cùng lớp với Anchor, được gọi là Positive
- Một hình ảnh từ một lớp khác với Anchor, được gọi là Negative



Hình 3.3: Mô tả phương pháp tính độ lỗi dựa trên điểm neo

Với bộ ba của ba hình ảnh này, (gọi các vector embedding của các ảnh này là A, P và N), như vậy mô hình cần phải đưa ra :

$$\text{Khoảng cách}(A, P) < \text{khoảng cách}(A, N)$$

Hay viết lại theo cách khác:

$$\text{Khoảng cách (A, P)} - \text{khoảng cách (A, N)} < 0$$

Để tránh trường hợp khoảng cách = 0 ta thêm tham số margin vào như sau:

$$\text{distance(A,P)} - \text{distance(A,N)} + \text{margin} < 0$$

Như vậy hàm mất mát sẽ có dạng [5]:

$$L = \max(d(A,P) - d(A,N) + \text{margin}, 0)$$

b) Phân chia tập dữ liệu huấn luyện

Để huấn luyện mô hình nhận dạng khuôn mặt, tôi chia tập dữ liệu huấn luyện thành 2 phần: 70% tập dữ liệu dùng để huấn luyện và 30% tập dữ liệu để kiểm thử mô hình.

c) Kết quả huấn luyện

3.2. Lập trình nhúng cho thiết bị điểm danh

Để triển khai việc điểm danh bằng nhận dạng khuôn mặt cần phải tạo ra thiết bị đặt ở mỗi lớp học để thực hiện việc điểm danh. Thiết bị này có gắn camera để thực hiện nhận dạng học sinh và kết nối internet để gửi thông tin người điểm danh về hệ thống theo dõi để thống kê và báo cáo. Vì vậy yêu cầu thiết bị phải có tốc độ xử lý nhanh, chính xác.

Trong quá trình nghiên cứu tôi đã đưa ra một số giải pháp cùng với các ưu/nhược điểm của các giải pháp như sau:

- **Sử dụng máy tính PC kèm webcam**
- **Sử dụng máy tính nhúng và camera tích hợp dạng module**

Như vậy với ưu điểm vượt trội của máy tính nhúng, tôi lựa chọn giải pháp này để thực hiện triển khai mô hình điểm danh học sinh.



Hình 3.4: Máy tính nhúng Raspberry pi cùng màn hình

3.2.1. Máy tính nhúng raspberry Pi 4:

3.2.2. Cài đặt hệ điều hành

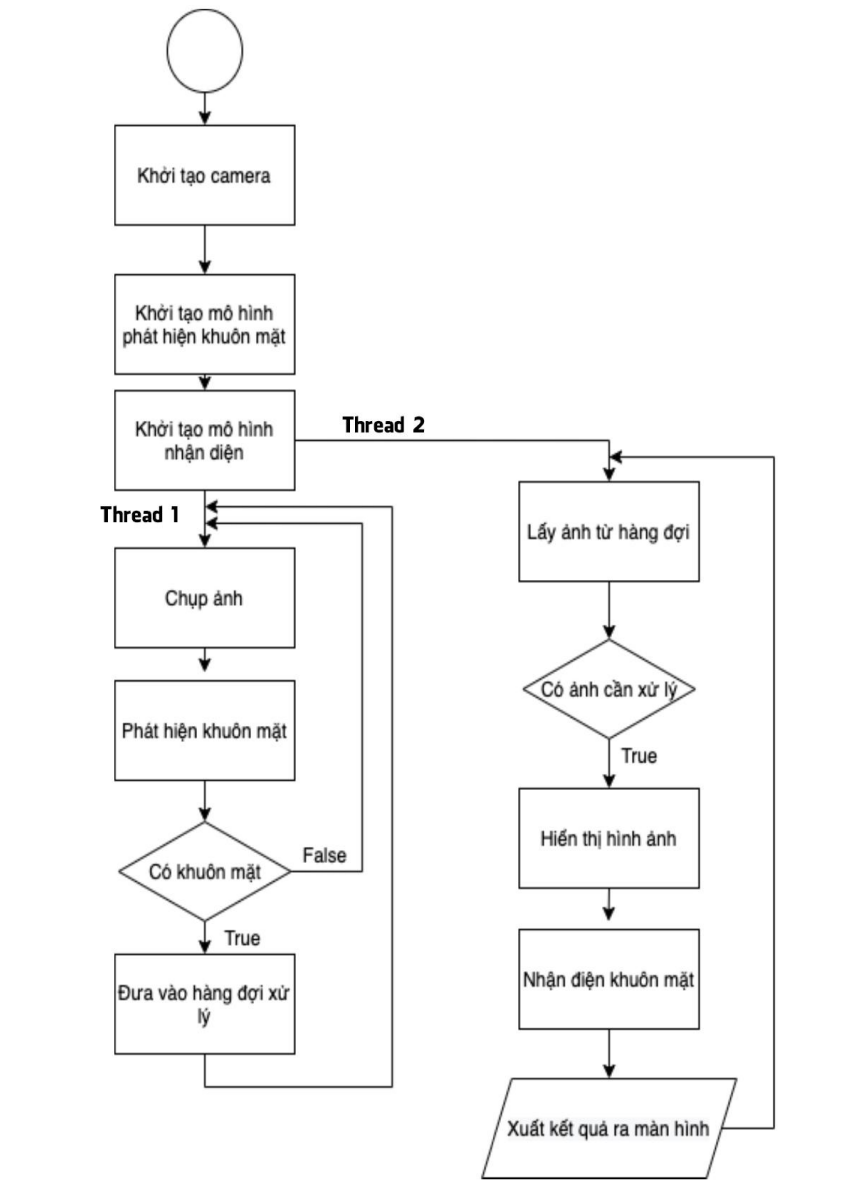
3.2.3. Xây dựng giao diện cho thiết bị

3.2.4. Xử lý nâng cao

Trong quá trình thực hiện sản phẩm, ban đầu tôi sử dụng thuật toán tuần tự như sau để thực hiện việc xử lý hình ảnh. Vấn đề đặt ra trong quá xử lý hình ảnh đó là thời gian xử lý của mô hình trên thiết bị chậm dẫn đến tốc độ khung hình của camera thể hiện trên màn hình không mượt và giật, đôi lúc dẫn đến treo thiết bị.

Ngoài ra có một vấn đề nữa nếu sử dụng thuật toán này, đó là việc khi mô hình phát hiện khuôn mặt tìm ra nhiều hơn 1 hình ảnh khuôn mặt (có thể do nhiều người điếm danh cùng lúc) thì máy chấm công sẽ dẫn đến bị treo, xảy ra hiện tượng thất cổ chai về mặt xử lý do mô hình nhận dạng khuôn mặt không xử lý kịp.

Để khắc phục điều này, tôi đã sử dụng hàng đợi để xử lý. Hàng đợi sẽ được xây dựng dựa trên thuật toán sau:

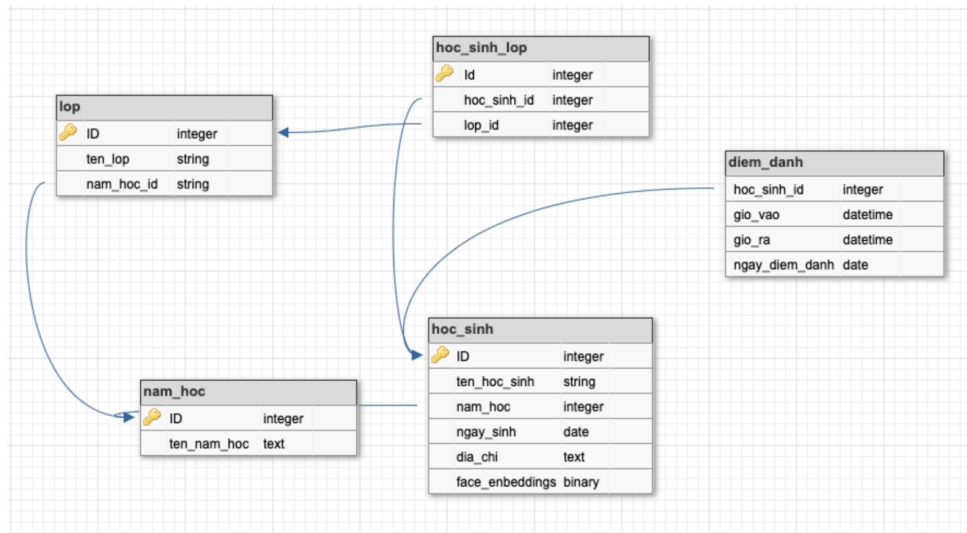


Hình 3.5: Thuật toán xử lý ảnh sau khi cải tiến.

Để thực hiện xử lý hàng đợi, cần tạo ra thêm 1 luồng xử lý song song với xử lý chính để giúp việc xử lý được nhanh hơn. Ngoài ra tôi cũng thay đổi thời gian nhận diện từ liên tục thành chụp ảnh sao mỗi 500ms, điều này ảnh hưởng ít đến tốc độ mà giúp tăng trải nghiệm của người dùng khi sử dụng máy điểm danh.

Thiết lập tham số `queue.size() < 2` để giới hạn số lượng cần xử lý trong hàng đợi, nếu hàng đợi đang có dữ liệu cần xử lý thì ta sẽ không đẩy thêm vào mà chờ cho hàng đợi xử lý xong.

3.3. Xây dựng cơ sở dữ liệu



Hình 3.6: Database hệ thống điểm danh

3.4. Demo và đánh giá kết quả

Khi mỗi khuôn mặt được đưa vào để nhận dạng thì mô hình Resnet sẽ phân lớp và lấy ra vector đặc trưng, sau đó hệ thống sẽ nhận dạng với thuật toán k-NN. Để đưa các thông tin nhận dạng được ra màn hình ở đây em sử dụng thư viện OpenCV.

Ban đầu, người dùng sẽ phải khai báo thông tin về khuôn mặt bằng cách chụp ảnh trên máy điểm danh. Sau đó giáo viên chủ nhiệm sẽ tiến hành khớp thông tin của học sinh với dữ liệu hình ảnh từ máy điểm danh gửi về phần mềm, mỗi học sinh có thể có nhiều hình ảnh khuôn mặt với nhiều góc độ chụp khác nhau để tăng độ chính xác.

Sau đó, người dùng với vai trò là học sinh sẽ thực hiện điểm danh, học sinh sẽ phải đứng trước máy điểm danh, sau đó đợi hệ thống nhận diện để xác định danh tính. Sau khi xác định xong danh tính, hệ thống sẽ gửi dữ liệu lên phần mềm và lưu trữ dữ liệu trên máy chủ cloud để cung cấp số liệu xuất ra báo cáo về kết quả điểm danh như sĩ số, tỉ lệ chuyên cần của từng lớp và gửi đến giáo viên chủ nhiệm .vv...

❖ Đánh giá độ chính xác của hệ thống

Sau khi triển khai hệ thống tại trường THPT Thanh Oai B, hệ thống hoạt động ổn định và đạt độ chính xác cao, tỉ lệ nhận dạng sai khuôn mặt rất nhỏ. Để khắc phục những trường hợp nhận dạng không chính xác, khuyến cáo đến người dùng là thử chụp lại ảnh, thay đổi các góc chụp chính xác và sau đó cập nhật lại cho hệ thống. Hầu hết các trường hợp bị nhận dạng sai đều được khắc phục. Cụ thể như sau:

- Trong quá trình triển khai thực tế, thiết bị vẫn xảy ra hiện tượng nhận nhầm học sinh. Trên thiết bị có báo cáo trường hợp thiết bị nhận sai danh tính, tỉ lệ này theo báo cáo tại các lớp học thì chiếm khoảng 5,4% trong tất cả các học sinh điểm danh.

- Các trường hợp bị nhận nhầm khuôn mặt được khắc phục bằng cách chụp lại các ảnh có thể gây nhầm lẫn, cần chụp lại ảnh của cả 2 người bị nhầm lẫn nhau.

3.5. Kết luận chương

Kết thúc chương cuối cùng này, luận văn đã hoàn thành hệ thống nhận dạng khuôn mặt. Kết quả thu được cũng tương đối khả quan so với mong muốn. Bên cạnh đó, trong tương lai khi có tập dữ liệu tốt hơn, mô hình phức tạp hơn nhờ có khả năng xử lí của máy móc tốt hơn thì mô hình CNN sẽ có thể mang lại một kết quả hài lòng hơn nữa.

KẾT LUẬN

Kết quả đạt được của luận văn

Qua nghiên cứu và thực nghiệm, luận văn đã đạt được những kết quả chính sau:

- Nghiên cứu tổng quan về xử lý ảnh.
- Nghiên cứu chi tiết mô hình học sâu tiêu biểu đó là mô hình CNN.
- Áp dụng mô hình CNN cho bài toán nhận dạng khuôn mặt.
- Ứng dụng các thuật toán đã tìm hiểu để giải quyết bài toán phân lớp thông qua mô hình huấn luyện bằng dữ liệu.
- Xây dựng tập dữ liệu huấn luyện và kiểm tra: Bước đầu xây dựng được quy trình thu thập dữ liệu từ thực tế, xử lý dữ liệu thô để đưa vào huấn luyện cho mô hình.
- Xây dựng hệ thống nhận dạng khuôn mặt.
- Triển khai được mô hình trên thiết bị phần cứng thực tế và nền tảng lưu trữ trên máy chủ cloud.

Bước đầu cho thấy hiệu quả của CNN trong việc nhận dạng ảnh đạt được kết quả tương đối khả quan.

Hướng phát triển của luận văn

Luận văn này đạt được một số kết quả nêu trên, nhưng luận văn còn nhiều hạn chế trong việc xây dựng mô hình và xử lý ảnh. Vì vậy, hướng nghiên cứu tiếp theo của luận văn sẽ là:

- Nghiên cứu thêm về mô hình CNN để có thể tăng độ chính xác cho việc nhận dạng trên thực tế.
- Nghiên cứu về mô hình phân lớp để phân loại được không chỉ là khuôn mặt người mà còn có thể dự đoán được tuổi tác và giới tính.
- Có thể phát triển ứng dụng trên bộ dữ liệu đầy đủ và chi tiết hơn.