**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

TRƯỜNG ĐẠI HỌC BÁCH KHOA

KHOA ĐIỆN – ĐIỆN TỬ

**BỘ MÔN ĐIỆN TỬ**

---------------o0o---------------



**ĐỒ ÁN MÔN HỌC**

**Lập trình ứng dụng nhận diện khuôn mặt để điểm danh học sinh qua websever và CSDL**

**GVHD: Nguyễn Khánh Lợi**

**SVTH: Đặng Lâm Tùng**

**MSSV: 1713856**

**TP. HỒ CHÍ MINH, THÁNG NĂM 20**

***LỜI CẢM ƠN***

*Tp. Hồ Chí Minh, ngày tháng năm .*

**Sinh viên**

**TÓM TẮT ĐỒ ÁN**

Đồ án này trình bày về việc thực hiện ứng dụng điểm danh học sinh ứng dụng nhận diện gương mặt (Face Recognition) bằng phương pháp Facenet trên Tensorflow, đồng thời xây dựng hệ thống server trên nền tảng Flask Python và cơ sở dữ liệu MySQL để điểm danh và quản lý việc điểm danh học sinh theo ngày.

Mục tiêu của đồ án là xây dựng được ứng dụng điểm danh học sinh trên PC có nhiệm vụ thực hiện thu ảnh và hiện thị kết quả nhận diện, với một giao diện người dùng (GUI) thân thiện, dễ sử dụng và một server thực hiện các bước xử lý nhận diện gương mặt, thêm dữ liệu gương mặt mới, điểm danh học sinh.

**MỤC LỤC**

[1. GIỚI THIỆU 1](#_Toc483413745)

[1.1 Tổng quan 1](#_Toc483413746)

[1.2 Nhiệm vụ đề tài 1](#_Toc483413747)

[1.3 Phân chia công việc trong nhóm 1](#_Toc483413748)

[2. LÝ THUYẾT 1](#_Toc483413749)

3. [THIẾT KẾ VÀ THỰC HIỆN PHẦN MỀM (NẾU CÓ) 2](#_Toc483413751)

4[. KẾT QUẢ THỰC HIỆN 2](#_Toc483413752)

5[. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 4](#_Toc483413753)

5[.1 Kết luận 4](#_Toc483413754)

5[.2 Hướng phát triển 4](#_Toc483413755)

[7. TÀI LIỆU THAM KHẢO 4](#_Toc483413756)

[8. PHỤ LỤC 4](#_Toc483413757)

DANH SÁCH HÌNH MINH HỌA

[Hình 5‑1 Kết quả thi công 3](#_Toc310380287)

[Hình 5‑2 Kết quả mô phỏng 3](#_Toc310380288)

**DANH SÁCH BẢNG SỐ LIỆU**

[Bảng 1 Thông số hệ thống 3](#_Toc310380293)

# GIỚI THIỆU

## Tổng quan

## Bài toán face recognition

Bài toán Face Recogntion là một bài toán kinh điển của lĩnh vực Computer vision, trong đó từ ảnh gương mặt người mà ta biết được ai là ai. Trước khi deep learning phát triển thì face recognition không đủ chính xác để sử dụng trong thực tế.

## Nhiệm vụ đề tài

Nội dung 1: Tìm hiểu lý thuyết về nhận diện gương mặt

Nghiên cứu lý thuyết về việc nhận diện gương mặt (face recognition)

Thực hiện chạy demo code face recognition trên máy tính

Nội dung 2: Tìm hiểu thư viện Flask Python, cách tạo web service, các giao thức truyền nhận ảnh.

Nghiên cứu lý thuyết về thư viện Flask Python để tạo web service

Nghiên cứu lý thuyết về giao thức HTTP

Thực hiện tạo web service face recognition bằng Flask, đã thử nghiệm chạy trên local host và AWS.

Nội dung 3: Nghiên cứu hệ quản trị CSDL MySQL để quản lý việc điểm danh học sinh và xây dựng ứng dụng giao diện pyQT để thực hiện nhận diện gương mặt và điểm danh trên PC

Nghiên cứu hệ cơ sở dữ liệu MySQL

Kết nối MySQL với server xử lý face recognition.

Nghiên cứu xây dựng GUI bằng PyQT5 để thực hiện nhận diện gương mặt trên PC.

# LÝ THUYẾT

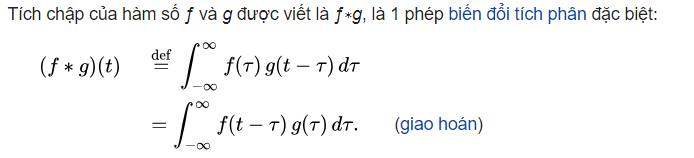
## Chương 1: Lý thuyết về nhận diện gương mặt (face recognition):

1.1 Mạng Neuron tích chập (CNN):

1.1.1 Tích chập (Convolution)

1.Bản chất của phép Convolution

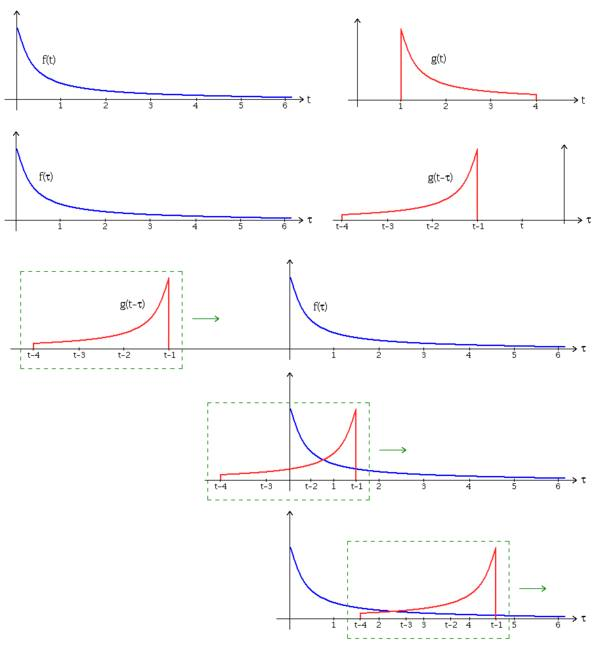
Trong xử lý số tín hiệu , tích chập là một khái niệm hết sức quen thuộc, nó là một phép toán cho 2 hàm số để ra một hàm số khác phép tích chập có thể thể hiện nhiều kiểu biến đổi (có thể là lọc tín hiệu, xét xem tín hiệu có theo mẫu nào không,....)



Tham khảo wiki

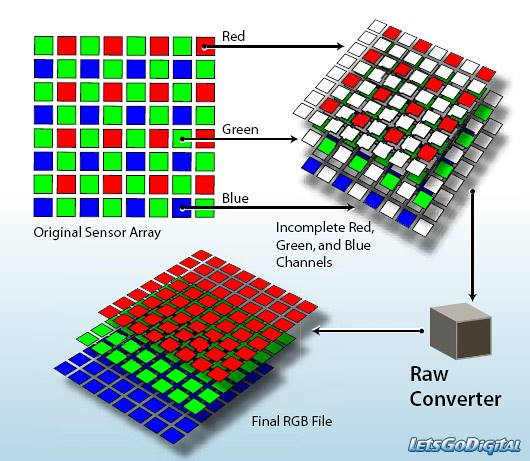
Công thức của tích chập thể hiện tính chất của tích chập đó là khi thực hiện, ta phải trừ ngược lại đáp ứng xung của hệ thống để tích chập thực hiện trước khi tín hiệu xuất hiện (thể hiện việc tín hiệu có thể xuất hiện ở trước điểm 0)

Dưới đây là hình minh họa của Wikipedia về tích chập:



Trong đó ta thực hiện trừ ngược và lật hàm g lại sau đó tích phân trên cả miền của tín hiệu

Một ảnh là một tập các ma trận các kênh màu:

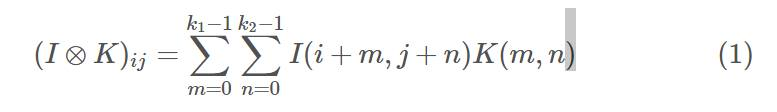


Ta hoàn toàn có thể coi đây như là tín hiệu 2 chiều vậy, chính vì vậy để lọc các tính chất của ảnh ta có thể áp dụng được các công cụ như tích chập. Chính vì thế dẫn đến các bộ lọc được dùng rất nhiều trong computer vision. Tuy nhiên ở đây tín hiệu có 2 chiều, và hơn nữa lại không có trường hợp tín hiệu xuất hiện trước t = 0 nên tích chập ở đây không cần phải lật kernel như tín hiệu 1 chiều. Chính vì vậy nó giống một khái niệm khác là tương quan chéo, mặc dù bản chất không giống nhau, phép này vẫn là tích chập.

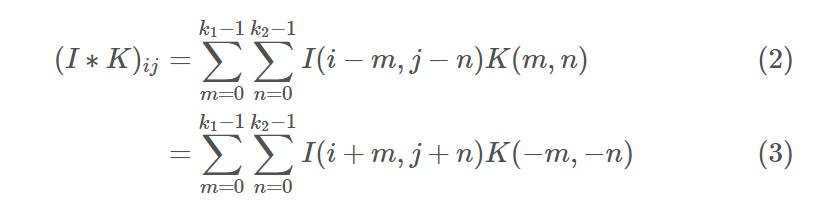


Bộ lọc Gaussian là một tích chập

Công thức tương quan chéo 2 chiều:

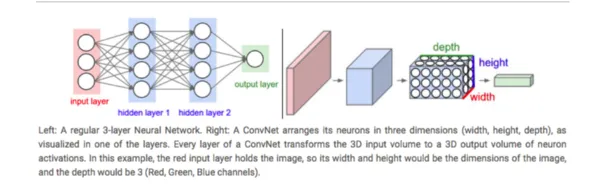


Công thức của tích chập 2 chiều:



Ở đây có thể thấy phép convolution được thực hiện giống với 1 chiều, và phép tích phân được thay thế bằng phép tổng, phù hợp với tín hiệu rời rạc

1.1.2 Mạng neuron tích chập

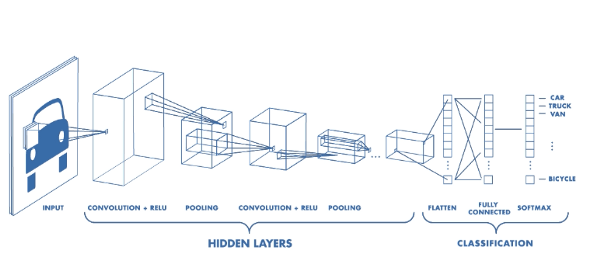


Mạng Nơ-ron Tích Chập có kiến trúc khác với Mạng Nơ-ron thông thường. Mạng Nơ-ron bình thường chuyển đổi đầu vào thông qua hàng loạt các tầng ẩn. Mỗi tầng là một tập các nơ-ron và các tầng được liên kết đầy đủ với các nơ-ron ở tầng trước đó. Và ở tầng cuối cùng sẽ là tầng kết quả đại diện cho dự đoán của mạng.

Đầu tiên, mạng Nơ-ron Tích Chập được chia thành 3 chiều: rộng, cao, và sâu. Kế đên, các nơ ron trong mạng không liên kết hoàn toàn với toàn bộ nơ-ron kế đến nhưng chỉ liên kết tới một vùng nhỏ. Cuối cùng, một tầng đầu ra được tối giản thành véc-tơ của giá trị xác suất.

CNNs gồm hai thành phần:

* Phần tầng ẩn hay phần rút trích đặc trưng: trong phần này, mạng sẽ tiến hành tính toán hàng loạt phép tích chập và phép hợp nhất (pooling) để phát hiện các đặc trưng. Ví dụ: nếu ta có hình ảnh con ngựa vằn, thì trong phần này mạng sẽ nhận diện các sọc vằn, hai tai, và bốn chân của nó.
* Phần phân lớp: tại phần này, một lớp với các liên kết đầy đủ sẽ đóng vai trò như một bộ phân lớp các đặc trưng đã rút trích được trước đó. Tầng này sẽ đưa ra xác suất của một đối tượng trong hình.

 [www.mathworks.com](https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html)

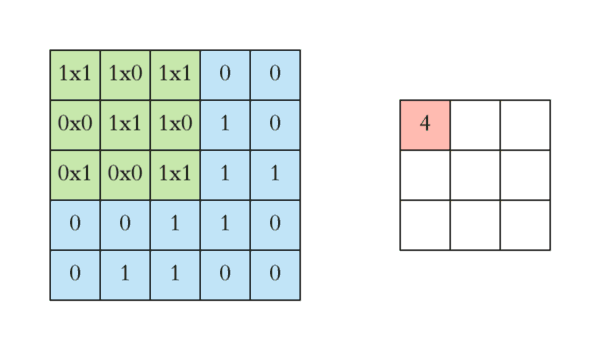
##### Rút trích đặc trưng

Tích chập là một khối quan trọng trong CNN. Thuật ngữ tích chập được dựa trên một phép hợp nhất toán học của hai hàm tạo thành hàm thứ ba. Phép toán này kết hợp hai tập thông tin khác nhau.

Trong trường hợp CNN, tích chập được thực hiện trên giá trị đầu vào của dữ liệu và kernel/filter (thuật ngữ này được sử dụng khác nhau tùy tình huống) để tạo ra một bản đồ đặc trưng (feature map).

Ta thực hiện phép tích chập bằng cách trượt kernel/filter theo dữ liệu đầu vào. Tại mỗi vị trí, ta tiến hành phép nhân ma trận và tính tổng các giá trị để đưa vào bản đồ đặc trưng.

Trong hình dưới đây, thành phần kernel/filter (màu xanh lá) trượt trên đầu vào (màu xanh dương) và kết quả được trả về bản đồ đặc trưng (màu đỏ). Kernel/filter có kích thước là 3×3 trong ví dụ này.

 [towardsdatascience.com](https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2)

Trong thực tế, tích chập được thực hiện hiện trên không gian 3 chiều. Vì mỗi hình ảnh được biểu diễn dưới dạng 3 chiều: rộng, cao, và sâu. Chiều sâu ở đây chính là giá trị màu sắc của hình (RGB).

Ta thực hiện phép tích chập trên đầu vào nhiều lần khác nhau. Mỗi lần sử dụng một kernel/filter khác nhau. Kết quả ta sẽ thu được những bản đồ đặc trưng khác nhau. Cuối cùng, ta kết hợp toàn bộ bản đồ đặc trưng này thành kết quả cuối cùng của tầng tích chập.

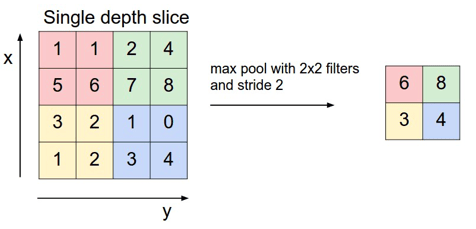
Tương tự như mạng nơ-ron thông thường, ta sử dụng một hàm kích hoạt (activate function) để có đầu ra dưới dạng phi tuyến. Trong trường hợp CNN, đầu ra của phép tích chập sẽ đi qua hàm kích hoạt nào đó ví dụ như hàm ReLU (rectified linear units).

Trong quá trình trượt kernel/filter trên dữ liệu đầu vào, ta sẽ quy định một bước nhảy (stride) với mỗi lần di chuyển. Thông thường ta lựa chọn thường chọn bước nhảy là 1. Nếu kích thước bước nhảy tăng, kernel/filter sẽ có ít ô trùng lắp.

Bởi vì kích thước đầu ra luôn nhỏ hơn đầu vào nên ta cần một phép xử lí đầu vào để đầu ra không bị co giãn. Đơn giản ta chỉ cần thêm một lề nhỏ vào đầu vào. Một lề với giá trị 0 sẽ được thêm vào xung quanh đầu vào trước khi thực hiện phép tích chập.

Thông thường, sau mỗi tầng tích chập, ta sẽ cho kết quả đi qua một tầng hợp nhất (pooling layer). Mục đích của tầng này là để nhanh chóng giảm số chiều. Việc này giúp giảm thời gian học và hạn chế việc overfitting.

Một phép hợp nhất đơn giản thường được dùng đó là max pooling, phép này lấy giá trị lớn nhất của một vùng để đại diện cho vùng đó. Kích thước của vùng sẽ được xác định trước để giảm kích thước của bản đồ đặc trưng nhanh chóng nhưng vẫn giữ được thông tin cần thiết.



[cs231n.github.io/convolutional-networks/](http://cs231n.github.io/convolutional-networks/)

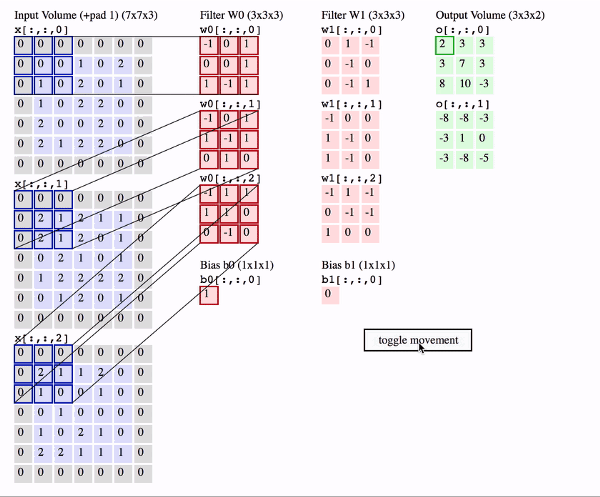
Tổng kết lại khi sử dụng CNN, ta cần chú ý đến 4 tham số quan trọng:

* Kích thước kernel/filter
* Số lượng kernel/filter
* Kích thước bước nhảy (stride)
* Kích thước lề (padding)

##### Phân lớp

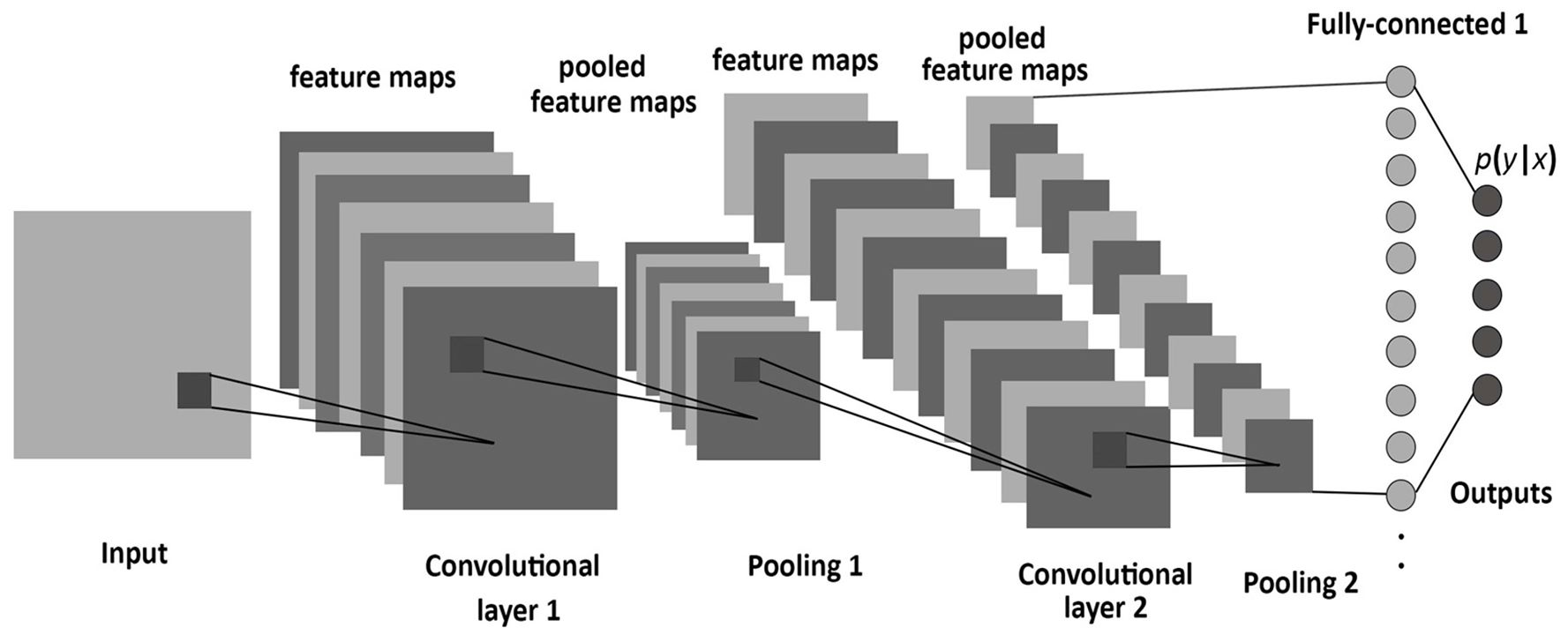
Trong phần phân lớp, ta sử dụng một vài tầng với kết nối đầy đủ để xử lí kết quả của phần tích chập. Vì đầu vào của mạng liên kết đầy đủ là 1 chiều, ta cần làm phẳng đầu vào trước khi phân lớp. Tầng cuối cùng trong mạng CNN là một tầng liên kết đầy đủ, phần này hoạt động tương tự như mạng nơ-ron thông thường.

Kết quả thu được cuối cùng cũng sẽ là một véc-tơ với các giá trị xác suất cho việc dự đoán như mạng nơ-ron thông thường.



1.1.3 Face Recognition với mạng CNN

Để hiểu cho đơn giản CNN hay Mạng neuron tích chập gồm các lớp tích chập sẽ thực hiện các thao tác tách feature của một hình ảnh ra và sau đó sử dụng một mô hình máy học khác như kNN hoặc SVM để phân biệt người này với người khác. Cách thực hiện này là one-shot learning, CNN chỉ đóng vai trò extract feature, việc huấn luyện CNN ở đây là khiến nó tách đặc trưng tốt hơn. Ta cũng có thể thực hiện huấn luyện một CNN với số đầu ra là số người, nhưng điều đó rất khó khăn, đòi hỏi thay đổi cấu trúc mạng và việc thu thập dataset cũng mất thời gian hơn, còn có thể mất cân bằng bộ dataset.



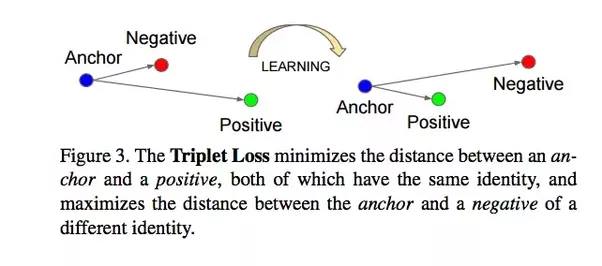
Facenet một mạng CNN có nhiệm vụ tách các đặc trưng của một ảnh mặt. Điểm đặc biệt tạo nên sự khác biệt của Facenet là nó sử dụng hàm lỗi Triplet để tối thiểu hóa khoảng cách giữa các gương mặt tương đồng và tối đa hóa khoảng cách đến những gương mặt không tương đồng, vì vậy facenet có thể phân biệt rất chính xác người với người. Quá trình train CNN được thực hiện trên các bộ dataset lớn (ví dụ như vggface hoặc MS 1 million), kết quả ta muốn là tập embedding vector ra phải tách nhau ra thành từng cụm để thực hiện phân loại bằng kNN hoặc SVM.

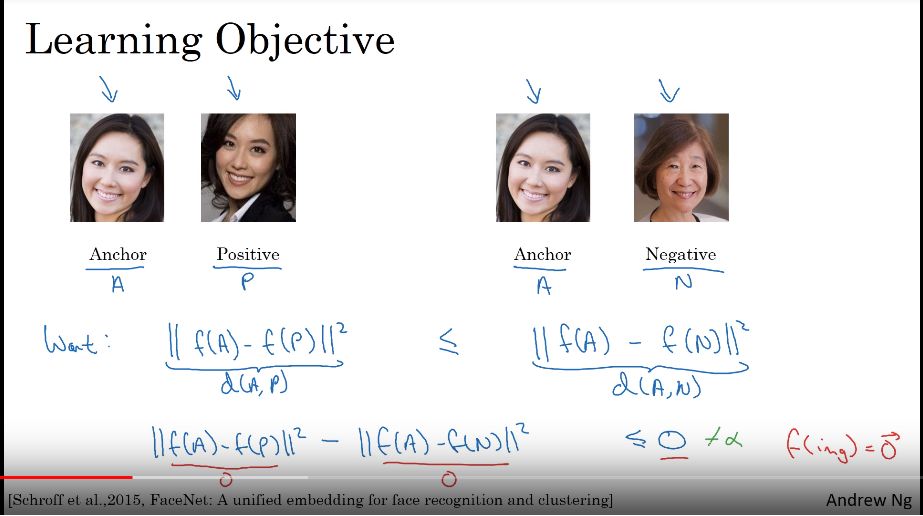
B1: Sử dung một bộ face detector để tách phần ảnh có gương mặt người ra. Các phương pháp hiện đại sử dụng MTCNN (Multi-task Cascaded Convolutional Networks), ưu điểm là mô hình này detect được gương mặt ở nhiều góc độ, và có thể nhận diện cả face landmark với thời gian xử lý khá tốt. Sau đó ta resize hình lại một kích cỡ xác định (tùy chọn), facenet thì là 160x160.

B2: Thực hiện các bước “lọc” bộ dataset. Thực tế là các bộ dataset lớn như vggface hay ms 1m đều có nhiều “nhiễu”, tức các hình không chứa gương mặt hoặc gương mặt khác với người cần xét. Vì vậy có thể dùng một mô hình nhận diện gương mặt khác để tìm và loại bỏ những hình “nhiễu” ấy. Hoặc đơn giản là ta có thể lấy clean list (danh sách các hình chuẩn) đã được lọc sẵn.

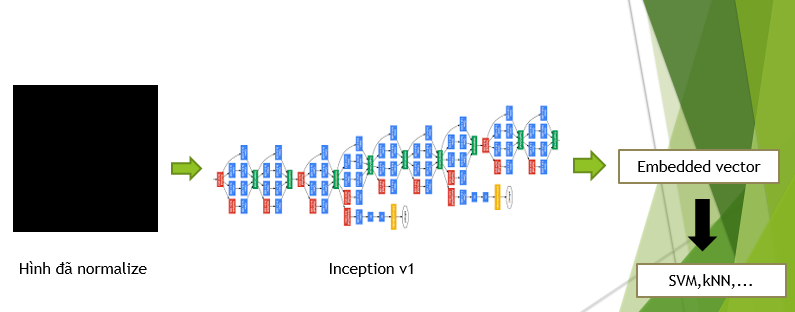
B3: Thực hiện bước normalize cho tập dataset đã chuẩn bị. Cách normalize cũng giống như cách normalize cho CNN bình thường. Các bước train: Đầu vào là tập dataset đã chuẩn bị, CNN có thể là các CNN như inception, mobilenet,… với đầu ra là embbeding vector như đã nói, embedding vector này có thể có 128 chiều hoặc 512 chiều,… tùy ý. Tóm lại ta chỉ cần lấy cấu trúc inception hoặc mobilenet, bỏ lớp cuối và thêm vào lớp embedding vector (không có activation)

B4: Forward pass trên tập data, lưu các embedding vector đã tính trên minibatch lại.  
  
B5: Chia minibatch thành 3 phần là positive, negative và anchor, trong đó anchor là phần “neo” giống như “tâm” của cụm embedding vector, sau đó chọn phần positive là những embedding vector là cùng người với anchor, negative là khác người  
  
B6: Tính Triplet loss là khoảng cách giữa các embbeding vector anchor và positive trừ cho khoảng cách giữa các embbeding vector anchor và negative (tức ta tìm cách làm giảm khoẳng cách từ các vector positive tới anchor và ngược lại đẩy các vector negative đi xa khỏi anchor)









## Chương 2: Giao thức HTTP và thư viện Flask Python

### Giao thức HTTP:

Http (HyperText Transfer Protocol) là giao thức truyền tải siêu văn bản được sử dụng trong www dùng để truyền tải dữ liệu giữa Web server đến các trình duyệt Web và ngược lại. (Wikipedia)

HTTP sử dụng các request như GET, POST, HEAD, PUT,.. . để thực hiện truyền thông tin

### Phương thức POST

Phương thức POST truyền thông tin thông qua HTTP header, thông tin này được mã hóa như phương thức GET. Dữ liệu đươc gửi bởi phương thức POST rất bảo mật vì dữ liệu được gửi ngầm, không đưa lên URL, bằng việc sử dụng Secure HTTP, bạn có thể chắc chắn rằng thông tin của mình là an toàn. Parameters được truyền trong request body nên có thể truyền dữ liệu lớn, hạn chế tùy thuộc vào cấu hình của Server. Không cache và bookmark được cũng như không được lưu lại trong browser history. POST không có bất kì hạn chế nào về kích thước dữ liệu sẽ gửi, có thể gửi dữ liệu nhị phân, hình ảnh. (Trích Viblo)

Để thực hiện truyền ảnh lên server để thực hiện nhận diện gương mặt em sử dụng giao thức POST do POST có thể send file ảnh lên server và để xử lý POST request lên server em sử dụng thư viện Flask của Python.

### Phương thức Get

Phương thức GET gửi thông tin người dùng đã được mã hóa thêm vào trên yêu cầu trang:

http://www.example.com/index.htm?name=value1&name1=value1

Chúng ta thấy rằng GET lộ thông tin trên đường dẫn URL. Băng thông của nó chỉ khoảng 1024 kí tự vì vây GET hạn chế về số kí tự được gửi đi. GET không thể gửi dữ liệu nhị phân , hình ảnh ... Có thể cached và được bookmark (đánh dấu trên trình duyệt). Lưu trong browser history.

### Thư viện Flask

* Flask là một web frameworks, nó thuộc loại micro-framework được xây dựng bằng ngôn ngữ lập trình Python.



* Đặc điểm của thư viện Flask là đơn giản, rất dễ hiểu và dễ sử dụng.
* Thời gian phát triển ứng dụng Flask rất nhanh
* Ví dụ về một chương trình Flask đơn giản:
* from flask import Flask # Gọi thư viện Flask
* app = Flask(\_\_name\_\_) # Tạo đối tượng Flask để tạo web service

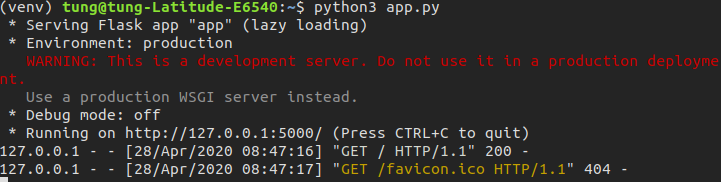
@app.route('/') # Quy định đường html là localhost/ sẽ do hàm hello\_world quản lý

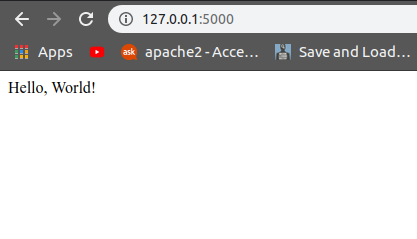
def hello\_world(): # Ta có thể chọn đường dẫn cũng như giao thức (POST hay GET)

return 'Hello, World!' # Trả lời lại request của user

if \_\_name\_\_ == '\_\_main\_\_': # Hàm main để thực chạy app

app.run()





## Chương 3: Hệ quản trị dữ liệu MySQL và thư viện PyQT:

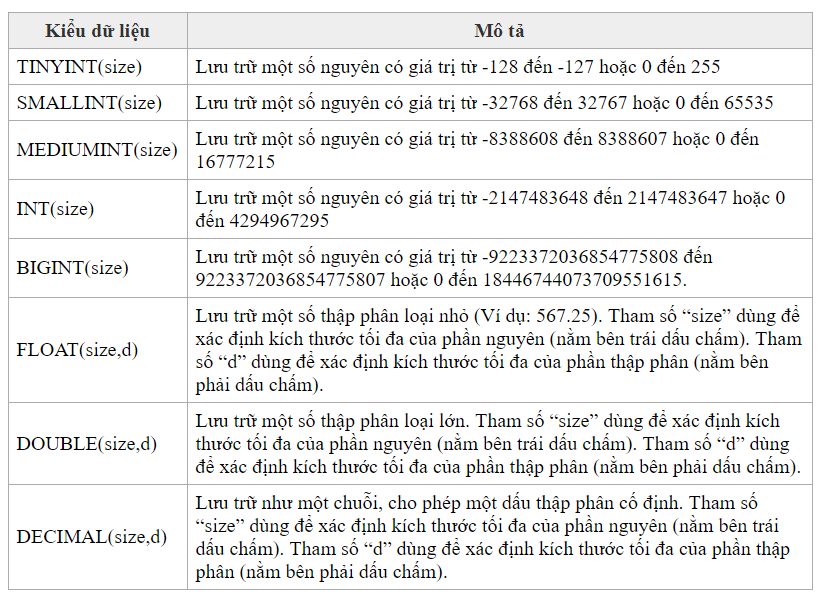
### Hệ quản trị dữ liệu MySQL và ngôn ngữ SQL

* SQL là viết tắt của Structured Query Language là ngôn ngữ truy vấn có cấu trúc. Nó là một ngôn ngữ, là tập hợp các lệnh để tương tác với cơ sở dữ liệu. Dùng để lưu trữ, thao tác và truy xuất dữ liệu được lưu trữ trong một cơ sở dữ liệu quan hệ.
* SQL giúp giúp quản lý hiệu quả và truy vấn thông tin nhanh hơn, giúp bảo trì, bảo mật thông tin dễ dàng hơn.

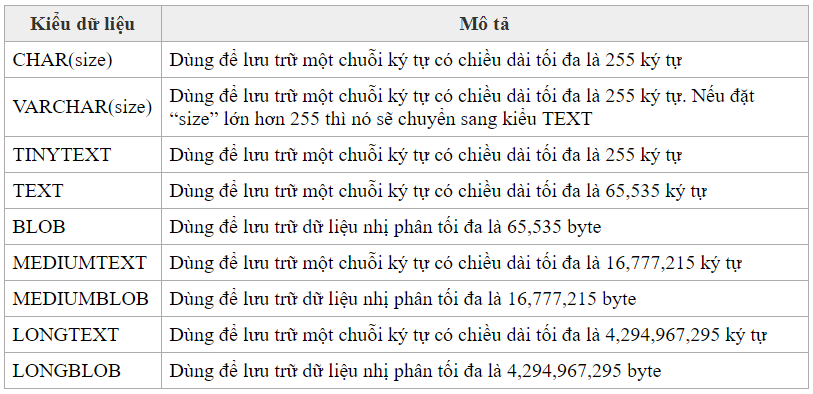
# **2. Các kiểu dữ liệu trong MySQL**

Trong MySQL, kiểu dữ liệu được chia làm ba loại chính: kiểu số, kiểu văn bản, kiểu ngày tháng

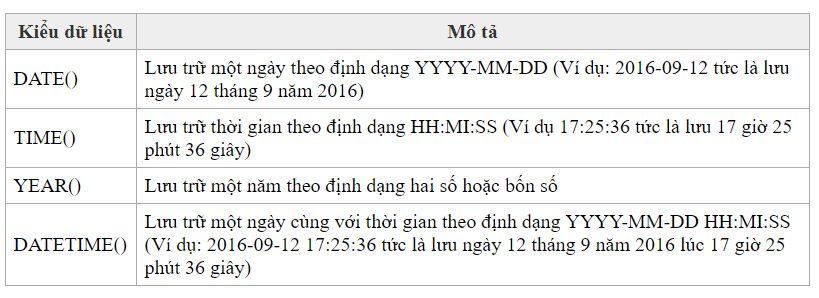
## **2.1 Kiểu Số**



## **2.2 Kiểu văn bản**



## **2.3 Kiểu ngày tháng**



* SQL quản lý data thành từng bảng (table)
* Data trong table có thể liên kết với nhau.
* Ví dụ tạo Table SQL:
* CREATE TABLE if not exists Student (id INT AUTO\_INCREMENT PRIMARY KEY, name VARCHAR(255), address VARCHAR(255), time VARCHAR(255))
* Để chọn một dữ liệu ta dùng câu lệnh SELECT:
* SELECT \* FROM Checkin JOIN Student USING(name)
* Để thêm data vào TABLE:
* INSERT INTO Student(name, address) VALUES (%s,%s)
* MySQL là một hệ thống quản trị cơ sở dữ liệu mã nguồn mở (Relational Database Management System, viết tắt là RDBMS)
* RDBMS là một phần mềm hay dịch vụ dùng để tạo và quản lý các cơ sở dữ liệu (Database) theo hình thức quản lý các mối liên hệ giữa chúng.
* Điểm hiệu qủa của MySQL là nó có chức năng phân quyền và cách kết nối dễ dàng.
* Mỗi chương liên quan đến một vấn đề
* Nếu có sử dụng tài liệu tham khảo, sinh viên cần trích dẫn rõ ràng.
* Với phần lý thuyết không quan trọng, sinh viên có thể đưa vào mục tài liệu tham khảo.

# THIẾT KẾ VÀ THỰC HIỆN PHẦN MỀM

* Yêu cầu đặt ra cho phần mềm

Các chức năng cần thực hiện:

+ Thực hiện được việc nhận diện gương mặt với độ chính xác cao, ổn định (~ 80%) các gương mặt được nhận diện chính xác.

+ Ứng dụng server chạy ổn định, thời gian xử lý nhanh, ổn định.

+ Ứng dụng điểm danh trên PC có giao diện đơn giản, dễ sử dụng. Ứng dụng có thể lưu lại thông tin về các lần điểm danh, thời gian điểm danh theo ngày.

* **Phân tích**
  + Mô hình cần thực hiện ở quy mô nhỏ (một lớp học) nên database không cần quá lớn
  + Để dễ dàng triển khai lên server và Cloud thì frameworks Flask được lựa chọn vì tính đơn giản và dễ triển khai của nó.
  + Với phần xử lý bằng dưới PC, một giao diện GUI viết bằng PyQT hoặc giao diện web sẽ có lợi hơn, ở đây em chọn PyQT làm giao diện chính.
  + Để đạt hiệu quả nhận diện chính xác nhất, em sử dụng mô hình nhận diện gương mặt facenet để lấy feature và SVM để phân loại từng người với nhau. Tuy nhiên cách làm này lại yêu cầu GPU để mang lại hiệu quả cao nhất, nhưng với điều kiện có hạn em chỉ có thể sử dụng CPU.
  + Về cơ sở dữ liệu, MySQL được chọn vì tính bảo mật và tốc độ xử lý khá tốt của nó.

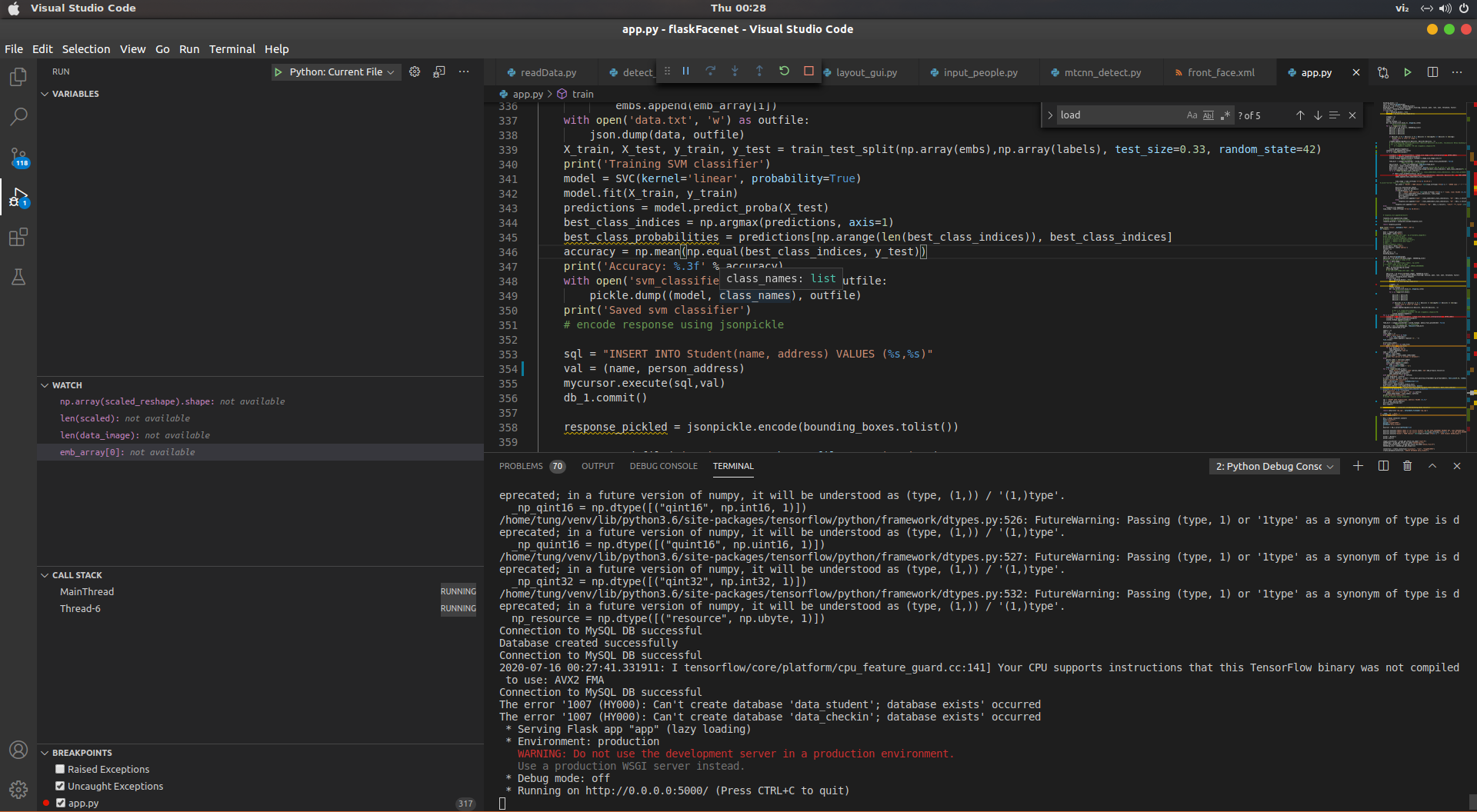
# KẾT QUẢ THỰC HIỆN

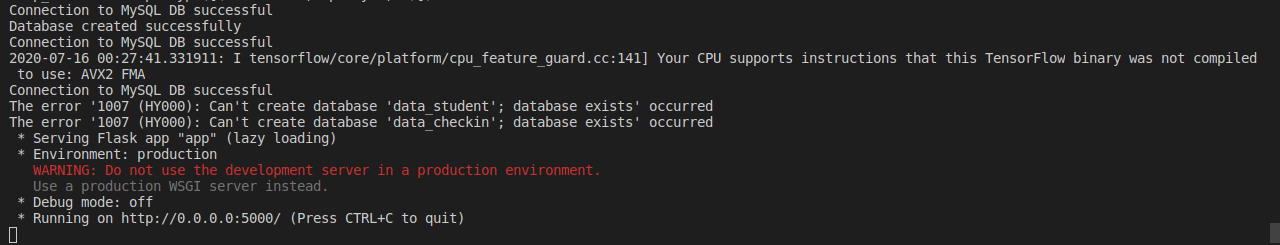
Trong phần này, sinh viên mô tả:

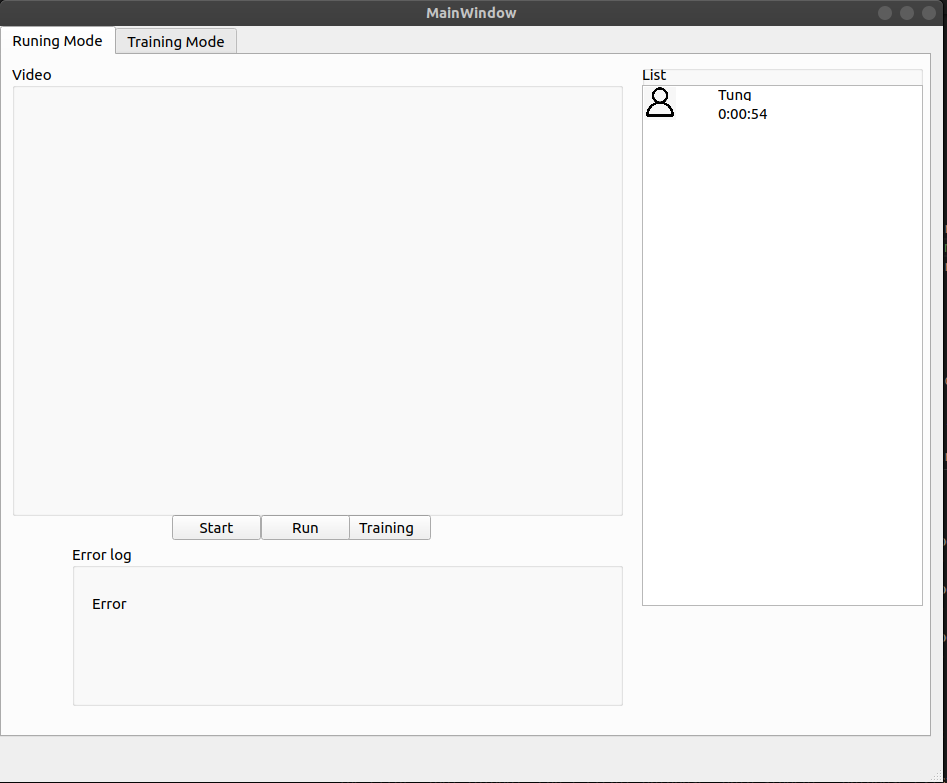
* Trình bày **cách thức đo đạc, thử nghiệm** 
  + Thiết bị sử dụng trong đồ án: Laptop Dell E6540, CPU Intel i7 4800 MQ
  + Các thư viện Python sử dụng: Tensorflow, OpenCV, Flask,….
  + Để thực hiện các lệnh gọi HTTP, phần mềm Postman được sử dụng

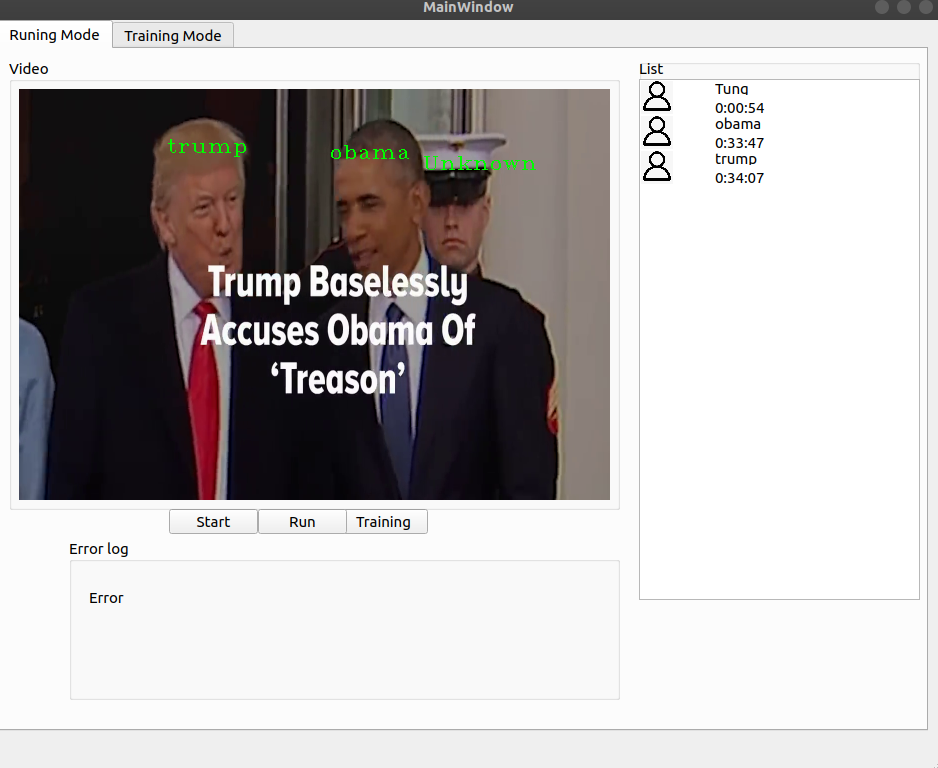
Các bước tiến hành thử nghiệm:

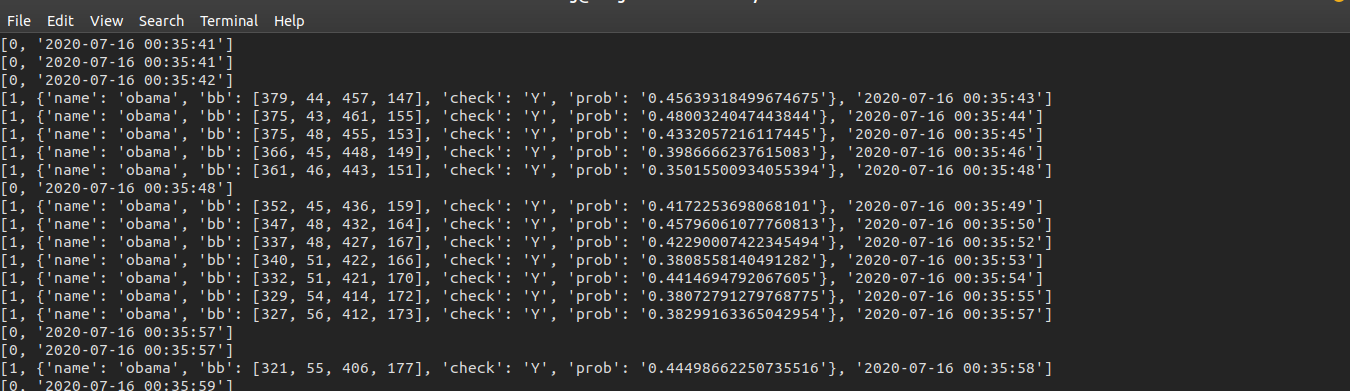
Server được bật lên:

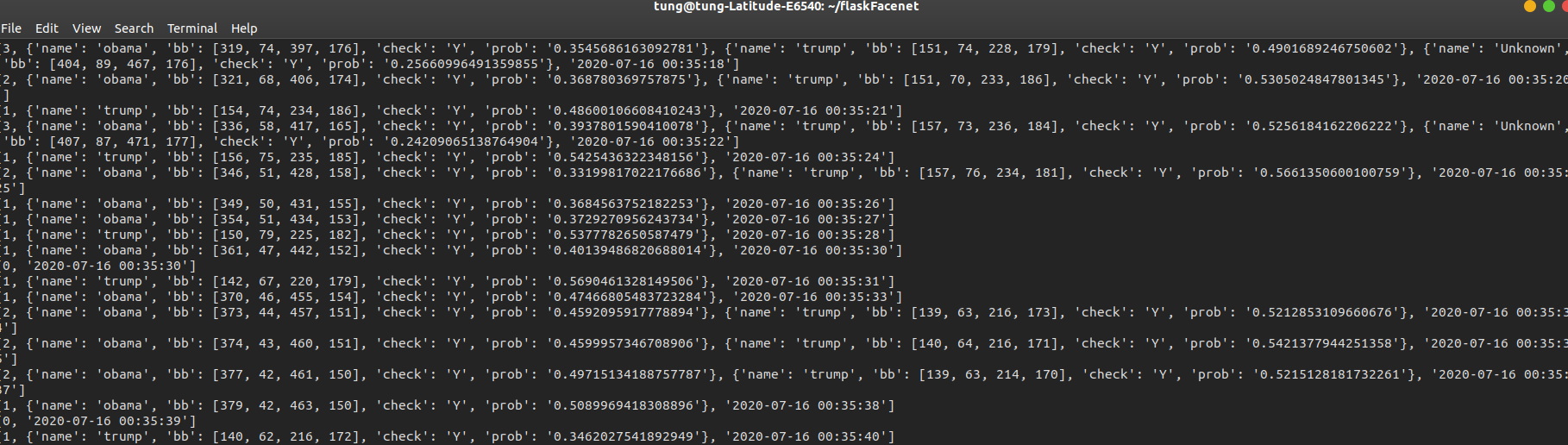


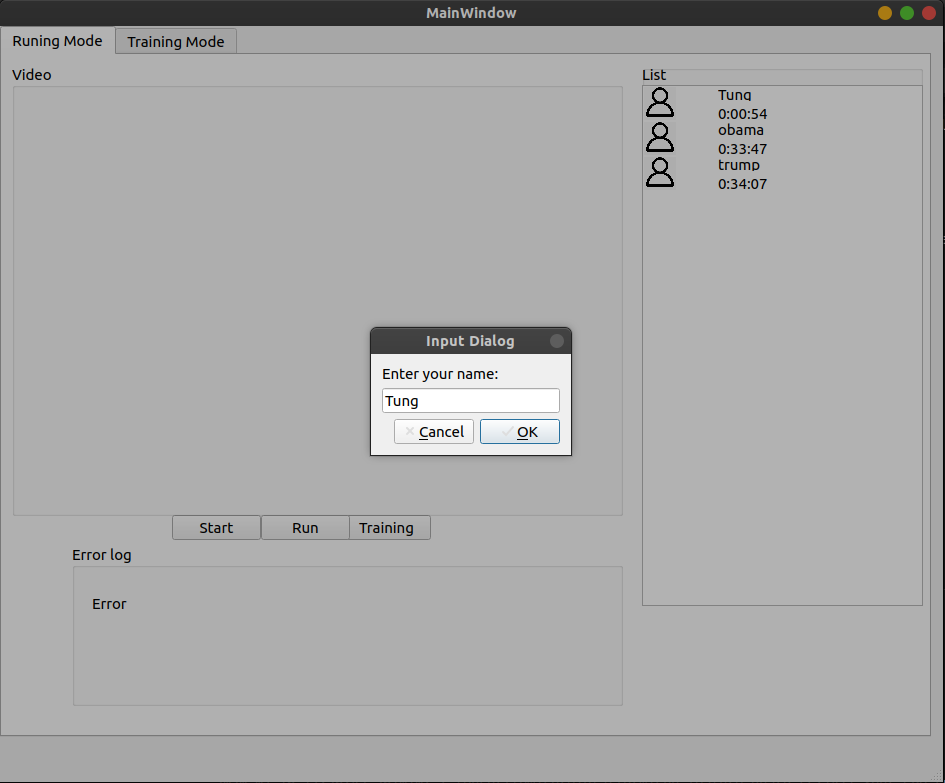


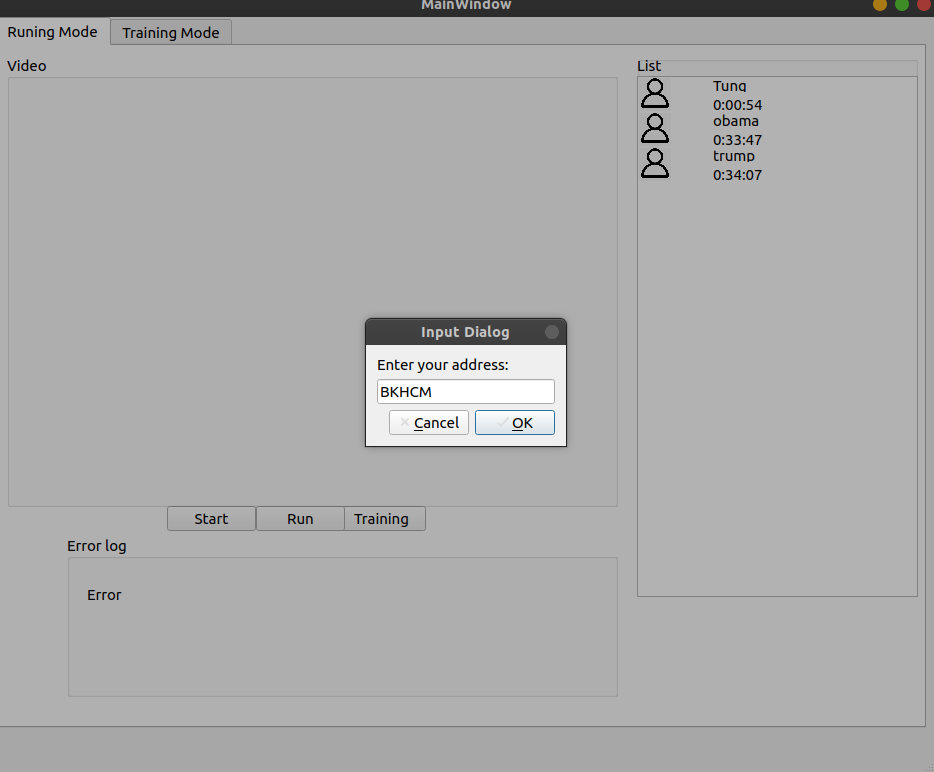
GUI được viết bằng PyQT:

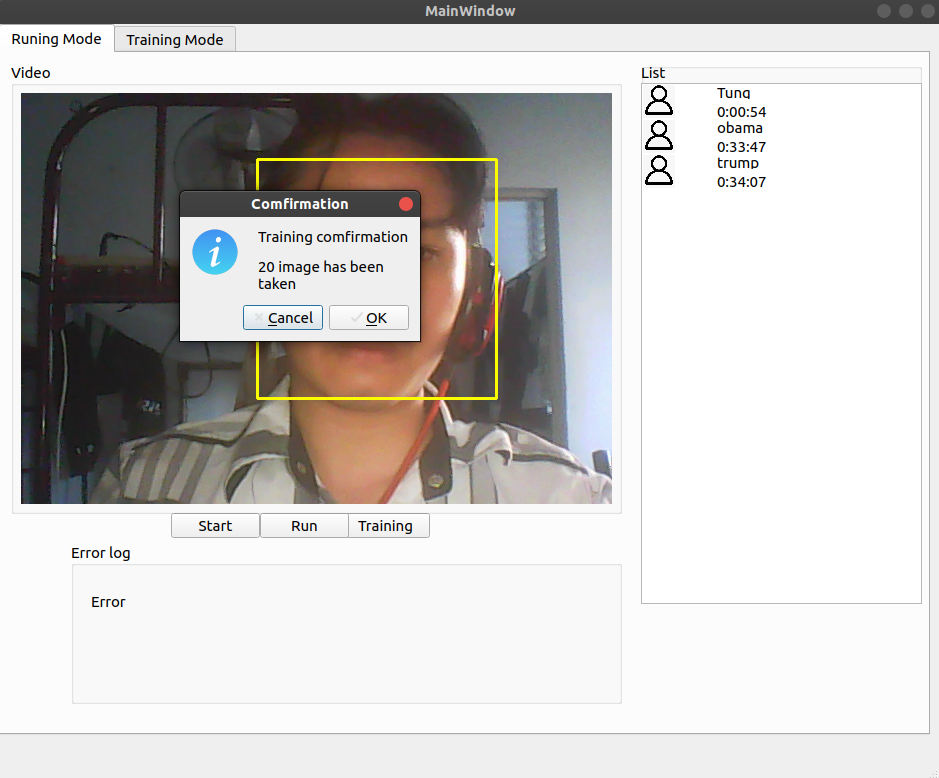




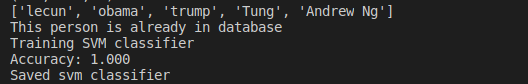








-Nếu gương mặt đã có trong database:



Trả về tên người được chụp, Bounding Box và ngày giờ người đó điểm danh:

Về tốc độ thực hiện chương trình:

Thuật toán Face Recognition Facenet như đã trình bày gồm 3 bước:

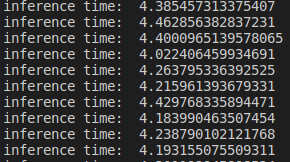
1) Face detection

2) Feature extraction

3) Face recognition

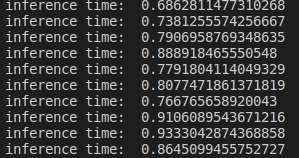
Do đó thời gian xử lý của thuật toán phụ thuộc vào số lượng gương mặt trong hình.

Trong trường hợp không detect được gương mặt thì thuật toán tương đương với thuật toán face detection với MTCNN, số fps trong 10 frame liên tiếp như sau:



=> Thời gian chạy thụât toán MTCNN trung bình là khoảng 1/(4.2) = 0.238s.

Trong trường hợp có 1 gương mặt, thời gian xử lý của server như sau:



=> Thời gian chạy này bao gồm thời gian detect face, thời gian chạy CNN và thời gian chạy SVM để phân loại từng gương mặt. Số fps xấp xỉ 0.9

=> Thời gian xử lý phần này là 1/0.9 = 1.11s

Như vậy thời gian xử lý CNN và SVM là 1.11 – 0.238 = 0.872 s

Tuy nhiên có thể nhận xét là thời gian chạy CNN đóng vai trò chủ yếu trong việc thực thi chương trình.

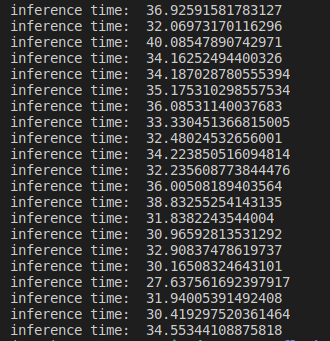
Thực hiện thí nghiệm so sánh thời gian truyền ảnh qua 2 giao thức HTTP và thư viện ZMQ

Thư viện ZMQ là thư viện chuyên cho việc xử lý truyền ảnh qua mạng, dựa trên nền tảng HTTP, thư viện ZMQ có thể được sử dụng với framework Flask của python.

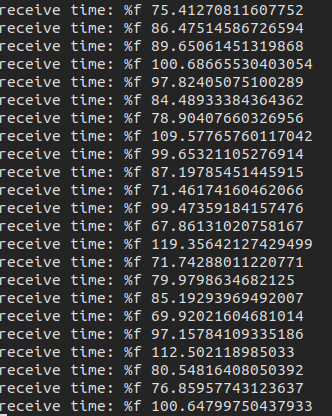
Để chứng minh nghẽn cổ chai trong hệ thống là từ phần xử lý nhận diện, em setup thí nghiệm đối chứng 2 giao thức HTTP được sử dụng trong đồ án và việc truyền nhận bằng thư viện ZMQ trên local host:

Thí nghiệm được đặt ra là truyền 2 video có kích thước 640\*480 qua local host bằng 2 giao thức đang so sánh:

Số FPS được truyền qua giao thức HTTP POST:

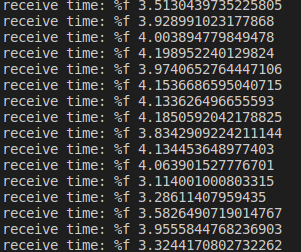


Số FPS của một video được truyền qua ZMQ:



Có thể thấy tốc độ truyền ảnh qua ZMQ nhanh hơn HTTP từ 2 – 3 lần.

Tuy nhiên khi sử dụng ZMQ cho ứng dụng face detection với MTCNN thì số fps tương đương so với HTTP



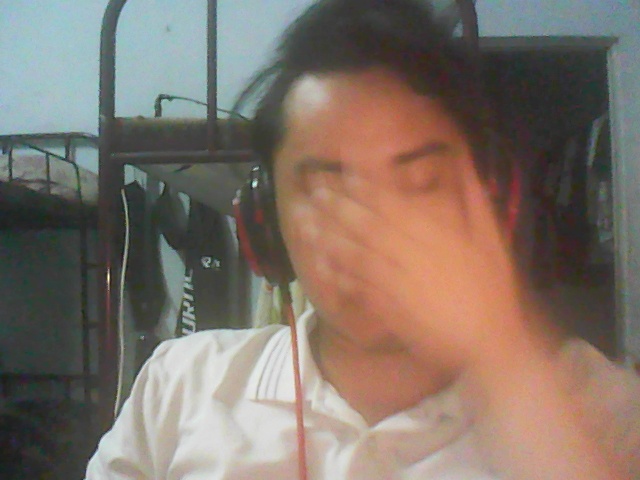
Điều này chứng tỏ rằng trong điều kiện thử nghiệm trên localhost, phương pháp sử dụng giao thức HTTP và ZMQ là có hiệu quả tương đương, nhưng xét về tốc độ truyền nhận data trên server, ZMQ chắc chắn là lựa chọn tốt hơn.

Về độ chính xác:

Do bài toán nhận diện gương mặt bao gồm khá nhiều bước phức tạp, em chỉ có thể thực hiện đo đạc độ chính xác trên 100 hình nhận diện của bản thân như sau:

|  |  |  |
| --- | --- | --- |
| Lần thử | Số frame phát hiện được guơng mặt | Số frame nhận diện được guơng mặt |
| 1 | 77 | 77 |
| 2 | 90 | 89 |
| 3 | 98 | 97 |

Một số trường hợp không detect được gương mặt:





Ta có thể thấy độ chính xác của mô hình facenet khá cao, lỗi xuất hiện chủ yếu ở phần MTCNN, đồng thời MTCNN gặp lỗi khá nhiêu khi detect guơng mặt quay góc

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## Kết luận

Ưu điểm:

Phần mềm thực hiện được các chức năng yêu cầu (nhận diện gương mặt, điểm danh, lưu vào CSDL, có GUI để chạy trên PC)

Nhược điểm:

Do giới hạn của phần cứng, kết quả thực hiện nhận diện gương mặt còn khá chậm

Phương pháp lấy mẫu data để train còn thiếu sót các trường hợp khó. các góc của gương mặt

Giao diện đồ họa (GUI) còn khá sơ sài, thiếu các chức năng tra cứu CSDL người dùng.

## Hướng phát triển

Các hướng phát triển cho đề tài:

+ Sử dụng phần cứng tốt hơn (GPU) để tăng hiệu quả nhận diện gương mặt/

+ Chỉnh sửa giao diện người dùng đẹp hơn, chuyên nghiệp hơn.

+ Thực hiện các bước lấy data training tốt hơn để đạt độ chính xác cao hơn.

# TÀI LIỆU THAM KHẢO

Trong mục này, sinh viên liệt kê những tài liệu đã tham khảo khi thực hiện đề tài luận văn. Những nội dung trình bày ở mục trên có tham khảo tài liệu thì sinh viên cần ghi chú bằng chỉ số (ví dụ [1], [2]). Chỉ số này cần tương ứng danh mục tài liệu tham khảo. Sinh viên xem thêm hướng dẫn cách viết trích dẫn kiểu IEEE.

Ví dụ:

1. Tống Văn On, “Thiết kế mạch số với VHDL & Verilog”, Nhà xuất bản Lao động Xã Hội, 2007.
2. Altera Corp., “SDRAM Controller for Altera’s DE2/ DE1 boards”, [www.altera.com](http://www.altera.com/)

# PHỤ LỤC

Trong phần này, sinh viên có thể trình bày:

* Những kết quả nghiên cứu bổ sung mà trong phần Kết quả luận văn chưa trình bày hết.
* Phần mã nguồn chương trình, sinh viên cũng có thể trình bày trong mục này. Để ngắn gọn, sinh viên chỉ đưa những mã nguồn chính vào phần Phụ lục.
* Sơ đồ toàn mạch chi tiết