



Chuyende - Báo cáo

Phương pháp nghiên cứu (Trường Đại học Công nghệ và Quản lý Hữu Nghị)



Scan to open on Studocu

ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG VIỆT - HÀN
KHOA KHOA HỌC MÁY TÍNH



BÁO CÁO ĐỒ ÁN MÔN HỌC
ĐỀ TÀI:
XÂY DỰNG CHƯƠNG TRÌNH
NHẬN DẠNG GIỚI TÍNH, ĐỘ
TUỔI, CẢM XÚC

Sinh viên thực hiện : **Nguyễn Đại Nam**
Nguyễn Thị Hoa Hồng

Giảng viên : **TS. Phạm Nguyễn Minh Nhựt**
Lớp : **19IT4**

Đà Nẵng, tháng 6 năm 2023

LỜI MỞ ĐẦU

Lời đầu tiên em xin chân thành cảm ơn đến thầy Phạm Nguyễn Minh Nhựt đã giúp em rất nhiều trong quá trình thực hiện đồ án này. Trong quá trình thực hiện đồ án, được sự giúp đỡ tận tình của thầy em đã thu được nhiều kiến thức quý báu giúp em rất nhiều trong quá trình học và làm việc trong tương lai. Trong quá trình thực hiện đồ án không tránh khỏi một số sai sót. Mong nhận được sự góp ý của các thầy để hoàn thiện hơn. Một lần nữa em xin chân thành cảm ơn sự giúp đỡ của thầy trong quá trình thực hiện đồ án để em hoàn thành đồ án này

[illegible]

2

MỤC LỤC

Chương 1 GIỚI THIỆU.....	5
1.1 Tổng Quan.....	5
1.2 Phương pháp – kết quả.....	6
Chương 2 CƠ SỞ LÝ THUYẾT.....	7
2.1 Tổng quan về Deep Learning và bài toán nhận diện khuôn mặt trong thị giác máy tính.....	7
2.2 Giới thiệu Convolutional Neural Network.....	16
2.3 Tổng quan về Machine Learning.....	17
2.4 Thư viện Keras.....	19
2.5 Thư viện FaceNet.....	19
2.6 Thư viện Sklearn.....	21
Chương 3 XÂY DỰNG ỨNG DỤNG.....	23
3.1 Yêu cầu hệ thống.....	23
3.2 Các bước thực hiện.....	23
KẾT LUẬN.....	28
4.1 Kết quả.....	28
4.2 Ưu điểm - nhược điểm.....	28
4.3 Hướng phát triển.....	29
TÀI LIỆU THAM KHẢO.....	30

DANH MỤC HÌNH ẢNH

<i>Hình 1. Deep Learning là gì ?</i>	9
<i>Hình 2. Cách thức hoạt động của Deep Learning</i>	10
<i>Hình 3. Các thuật ngữ Deep Learning</i>	11
<i>Hình 4. Ứng dụng của Deep Learning</i>	14
<i>Hình 5. Hệ thống nhận diện khuôn mặt</i>	15
<i>Hình 6. Kiến trúc CNN trong nhận diện hành động con người</i>	17
<i>Hình 7. FaceNet lấy hình ảnh khuôn mặt làm đầu vào và xuất ra vector embedding</i>	20
<i>Hình 8. Kết quả chương trình</i>	28

Chương 1 GIỚI THIỆU

1.1 Tổng Quan

1.1.1 Mở đầu

Hiện nay, cùng với sự phát triển của xã hội, vấn đề an ninh bảo mật đang được yêu cầu khắt khe tại mọi quốc gia trên thế giới. Các hệ thống nhận dạng con người được ra đời với độ tin cậy ngày càng cao. Một trong các bài toán nhận dạng con người rất được quan tâm hiện nay là nhận dạng khuôn mặt. Vì nhận dạng khuôn mặt là cách mà con người sử dụng để phân biệt nhau. Bên cạnh đó, ngày nay việc thu thập, xử lý thông tin qua ảnh để nhận biết đối tượng đang được quan tâm và ứng dụng rộng rãi. Với phương pháp này, chúng ta có thể thu nhận được nhiều thông tin từ đối tượng mà không cần tác động nhiều đến đối tượng nghiên cứu. Sự phát triển của khoa học máy tính tạo môi trường thuận lợi cho bài toán nhận dạng mặt người từ ảnh số. Các hệ thống nhận dạng offline đã ra đời và có độ tin cậy cao, tuy nhiên các hệ thống nhận dạng online lại chưa đáp ứng được nhiều.

Trong khuôn khổ đề án này, chúng em sẽ tiếp tục giải quyết bài toán nhận dạng offline. Trong đó đối tượng được thu thập thành các file dữ liệu và được chuyển về trung tâm. Tại đó, các số liệu sẽ được phân tích xử lý. Trong phần đầu, chúng em sẽ giải quyết bài toán nhận dạng thông thường, phần hai sẽ là phần nhận dạng độ tuổi, giới tính, cảm xúc.

1.1.2 Mục tiêu của đề án

Sử dụng thư viện Multi Task Convolutional Neural Network để trích xuất khuôn mặt từ một bức ảnh đầu vào cho trước sau đó sử dụng pre-trained model FaceNet để giảm chiều dữ liệu các vector biểu diễn khuôn mặt trở thành vector 128 chiều và áp dụng các thuật toán Machine Learning để huấn luyện model sau đó lựa chọn ra model tốt nhất.

- Input: Ảnh chụp mặt người.
- Output: Giới tính, độ tuổi, cảm xúc của người đó.

1.1.3 Đối tượng nghiên cứu

- Ngôn ngữ Python
- Tìm hiểu về Machine Learning và Deep Learning

- Các thuật toán nhận diện
- Nghiên cứu thư viện Multi Taks Convolutional Neural Network và FaceNet

1.1.4 Phạm vi nghiên cứu

- Xây dựng chương trình nhận dạng giới tính con người sử dụng Machine Learning.
- Ứng dụng đề tài phục vụ cho việc nghiên cứu về nhiều lĩnh vực.

1.2 Phương pháp – kết quả

1.2.1 Phương pháp

**Phương pháp chủ đạo*

- Phương pháp nghiên cứu lý thuyết:
 - + Tìm hiểu tổng quan về Machine Learning và bài toán nhận diện khuôn mặt trong thị giác máy tính.
- + Tìm hiểu thư viện Multi Taks Convolutional Neural Network và FaceNet.
- + Tìm hiểu thư viện Keras và một số thư viện khác.
- Phương pháp nghiên cứu thực nghiệm:
 - + Tiến hành phân tích và cài đặt trên Python
 - + Tiến hành training data cho máy học.

1.2.2 Kết quả

- Tạo ra chương trình nhận dạng giới tính thông qua ảnh có sẵn thực trên hệ điều hành Window.

Chương 2 CƠ SỞ LÝ THUYẾT

2.1 Tổng quan về Deep Learning và bài toán nhận diện khuôn mặt trong thị giác máy tính

2.1.1 Tổng quan về Deep Learning

2.1.1.1 Deep Learning là gì?

Deep Learning là một tập hợp con của Trí tuệ nhân tạo – một kỹ thuật học máy dạy cho máy tính và các thiết bị hoạt động một cách logic. Tại sao lại đặt tên là Deep Learning? Bởi thực tế là nó liên quan đến việc đi sâu vào một số lớp của mạng, bao gồm cả một lớp ẩn. Bạn càng học sâu, bạn càng trích xuất ra những thông tin phức tạp.

Phương pháp Deep learning dựa vào các chương trình phức tạp khác nhau để bắt chước trí thông minh của con người. Phương pháp đặc biệt này dạy cho máy móc nhận biết các họa tiết để có thể phân loại chúng thành các loại khác nhau. Nhận dạng mẫu là một phần thiết yếu của Deep learning và nhờ Machine learning, máy tính thậm chí không cần phụ thuộc vào lập trình mở rộng. Thông qua Deep learning, máy móc có thể sử dụng các tệp hình ảnh, văn bản hoặc âm thanh để xác định và thực hiện bất kỳ tác vụ nào theo cách giống như con người.



Hình 1. Deep Learning là gì ?

2.1.1.2 Tầm quan trọng của Deep Learning

Có thể thấy rằng Deep Learning đang ngày càng phổ biến thời gian gần đây. Nó đóng góp rất nhiều vào việc làm cho cuộc sống hàng ngày của chúng ta thuận tiện hơn, và xu hướng này sẽ phát triển trong tương lai. Cho dù đó chỉ là công nghệ hỗ trợ đỗ xe tự động hoặc nhận diện khuôn mặt tại sân bay, Deep learning đang thúc đẩy rất nhiều cho tự động hóa trong thế giới ngày nay.

Tuy nhiên, sự liên quan của Deep learning có thể được liên kết hầu hết với thực tế là thế giới của chúng ta đang tạo ra lượng dữ liệu theo cấp số nhân ngày nay, điều đó đòi hỏi nhu cầu cấu trúc trên quy mô lớn. Deep learning sử dụng khối lượng dữ liệu ngày càng tăng và tính sẵn có của dữ liệu là hợp lý nhất. Tất cả các thông tin được thu thập từ các dữ liệu này được sử dụng để đạt được kết quả chính xác thông qua các mô hình học tập lặp.

Việc phân tích lặp đi lặp lại các bộ dữ liệu lớn sẽ xóa bỏ các lỗi và sự khác biệt trong quá trình tìm kiếm mà cuối cùng dẫn đến một kết luận đáng tin cậy. Deep learning sẽ tiếp tục tạo ra tác động trong cả kinh doanh và cá nhân và tạo ra rất nhiều cơ hội việc làm trong thời gian tới.

2.1.1.3 Deep Learning hoạt động như thế nào ?

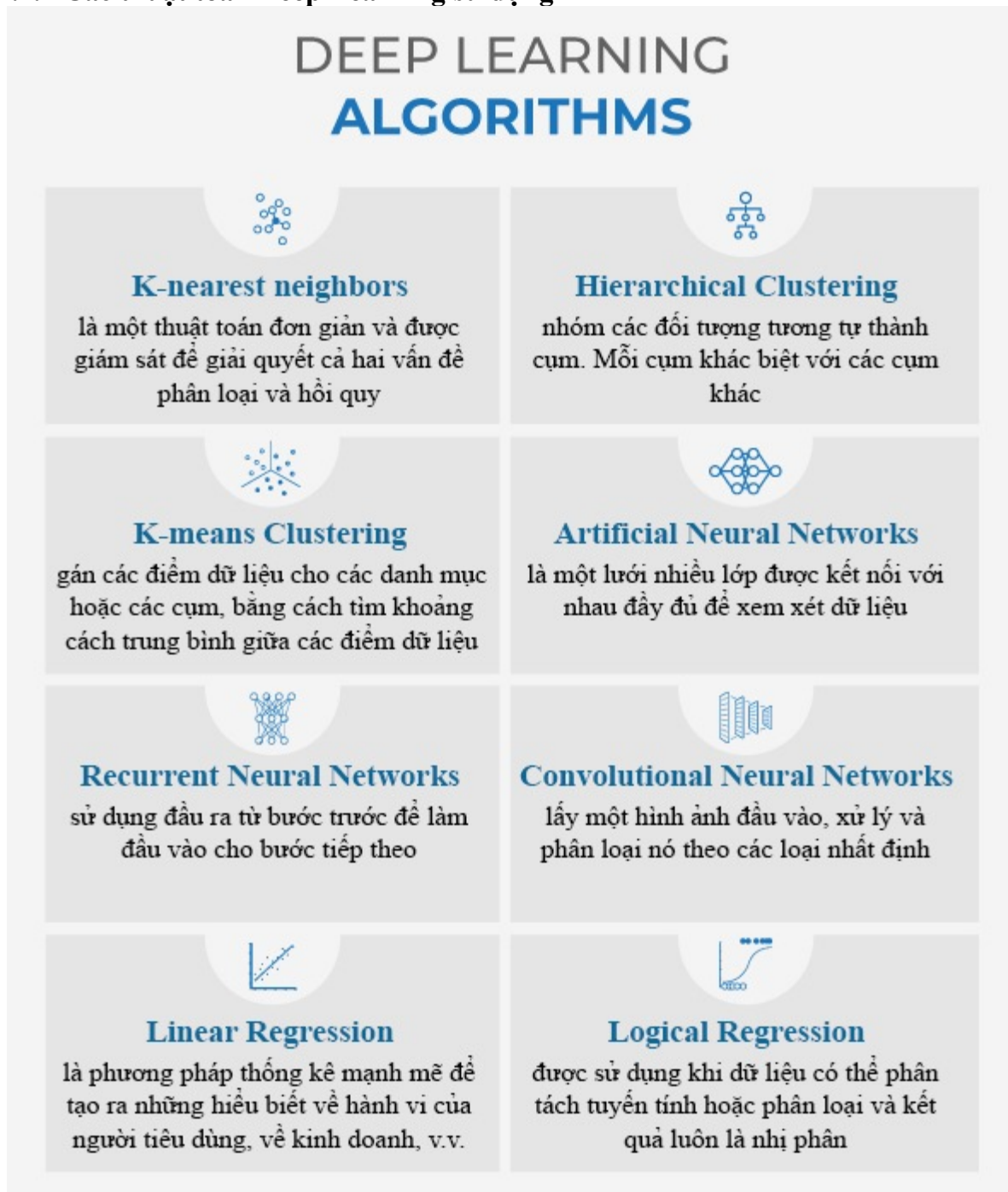
Về bản chất, Deep learning dựa vào các phương pháp lặp để dạy máy móc bắt chước trí thông minh của con người. Một mạng lưới thần kinh nhân tạo thực hiện phương pháp lặp này thông qua một số cấp độ phân cấp. Các cấp độ ban đầu giúp các máy học thông tin đơn giản và khi mức độ tăng lên, thông tin sẽ tiếp tục phát triển. Với mỗi máy cấp độ mới, hãy lấy thêm thông tin và kết hợp nó với những gì nó đã học được ở cấp độ trước. Khi kết thúc quá trình, hệ thống sẽ thu thập một thông tin cuối cùng là đầu vào ghép. Thông tin này đi qua một số hệ thống phân cấp và có ý nghĩa đối với tư duy logic phức tạp.

DEEP LEARNING APPLICATIONS

-  Hệ thống đề xuất trong Netflix, Amazon và Spotify bao gồm các thuật toán Deep learning để nâng cao trải nghiệm của người dùng
-  Phương tiện giao thông tự điều khiển hiểu được thực tế của con đường, biển báo dừng hoặc phương tiện khác thông qua các thuật toán Deep learning
-  Trợ lý ảo như Siri, Alexa và Cortana dịch bài phát biểu của bạn, đặt lịch hẹn và ghi chú với trình xử lý ngôn ngữ tự nhiên Siri, Alexa và Cortana dịch bài phát biểu của
-  Robot cố vấn và Nền tảng đầu tư kỹ thuật số sử dụng thuật toán để tự động thiết lập danh mục đầu tư và tìm kiếm cơ hội đầu tư tốt hơn cho khách hàng
-  Thuật toán Deep learning kiểm tra gen di truyền để dự đoán nguy cơ bệnh tật trong tương lai và các giai đoạn sức khỏe tiêu cực

Hình 2.Cách thức hoạt động của Deep Learning

2.1.1.4 Các thuật toán Deep Learning sử dụng



Hình 3. Các thuật ngữ Deep Learning

2.1.1.5 Một số ứng dụng của Deep Learning

Trợ lý ảo: Amazon Echo, Google Assistant, Alexa và Siri đều đang khai thác các khả năng học sâu để xây dựng trải nghiệm người dùng tùy chỉnh cho bạn. Họ tìm hiểu tiếng nói để nhận ra giọng nói và ngữ điệu của bạn và mang đến cho bạn trải nghiệm thứ cấp của con người thông qua một cỗ máy bằng cách sử dụng các mạng lưới thần kinh sâu sắc bắt chước không chỉ lời nói mà cả giọng điệu của con người. Trợ lý ảo giúp bạn mua sắm, điều hướng, ghi chú và dịch chúng thành văn bản, và thậm chí đặt lịch hẹn cho thẩm mỹ viện.

Nhận dạng khuôn mặt: Nhận dạng khuôn mặt của iPhone sử dụng Deep learning để xác định các điểm dữ liệu từ khuôn mặt của bạn để mở khóa điện thoại hoặc phát hiện ra bạn trong hình ảnh. Deep Learning giúp họ bảo vệ điện thoại khỏi những lần mở khóa không mong muốn và khiến trải nghiệm của bạn không gặp rắc rối ngay cả khi bạn đã thay đổi kiểu tóc, giảm cân hoặc trong điều kiện ánh sáng kém. Mỗi khi bạn mở khóa điện thoại, deep learning sử dụng hàng ngàn điểm dữ liệu để tạo bản đồ độ sâu khuôn mặt của bạn và thuật toán sẵn có sử dụng những điểm đó để xác định xem đó có thực sự là bạn hay không.

Cá nhân hóa: Những ông lớn trong lĩnh vực Thương mại và Giải trí như Amazon và Netflix, v.v. đang cố gắng xây dựng năng lực Deep learning hơn nữa để cung cấp cho người dùng một hệ thống mua sắm hoặc giải trí được cá nhân hóa. Các mục sê-ri phim được đề xuất dựa trên “mẫu” của bạn đều dựa trên học tập sâu. Các doanh nghiệp của họ phát triển mạnh trong việc đưa ra các tùy chọn trong tiềm thức của bạn dựa trên sở thích của bạn, các mục được truy cập gần đây, mối quan hệ với các thương hiệu, diễn viên, nghệ sĩ và lịch sử duyệt web tổng thể trên nền tảng của họ.

Xử lý ngôn ngữ tự nhiên: Một trong những công nghệ quan trọng nhất, Xử lý ngôn ngữ tự nhiên đang đưa AI từ tốt đến tuyệt vời về mặt sử dụng, sự trưởng thành và tinh tế. Các tổ chức đang sử dụng Deep learning rộng rãi để tăng cường tính chất phức tạp này trong các ứng dụng NLP. Tóm tắt tài liệu, trả lời câu hỏi, mô hình hóa ngôn ngữ, phân loại văn bản, phân tích tình cảm là một số ứng dụng phổ biến đã bắt kịp đà phát triển. Một

số công việc trên toàn thế giới phụ thuộc vào sự can thiệp của con người về chuyên môn ngôn ngữ và ngôn ngữ viết sẽ sớm trở lên dư thừa.

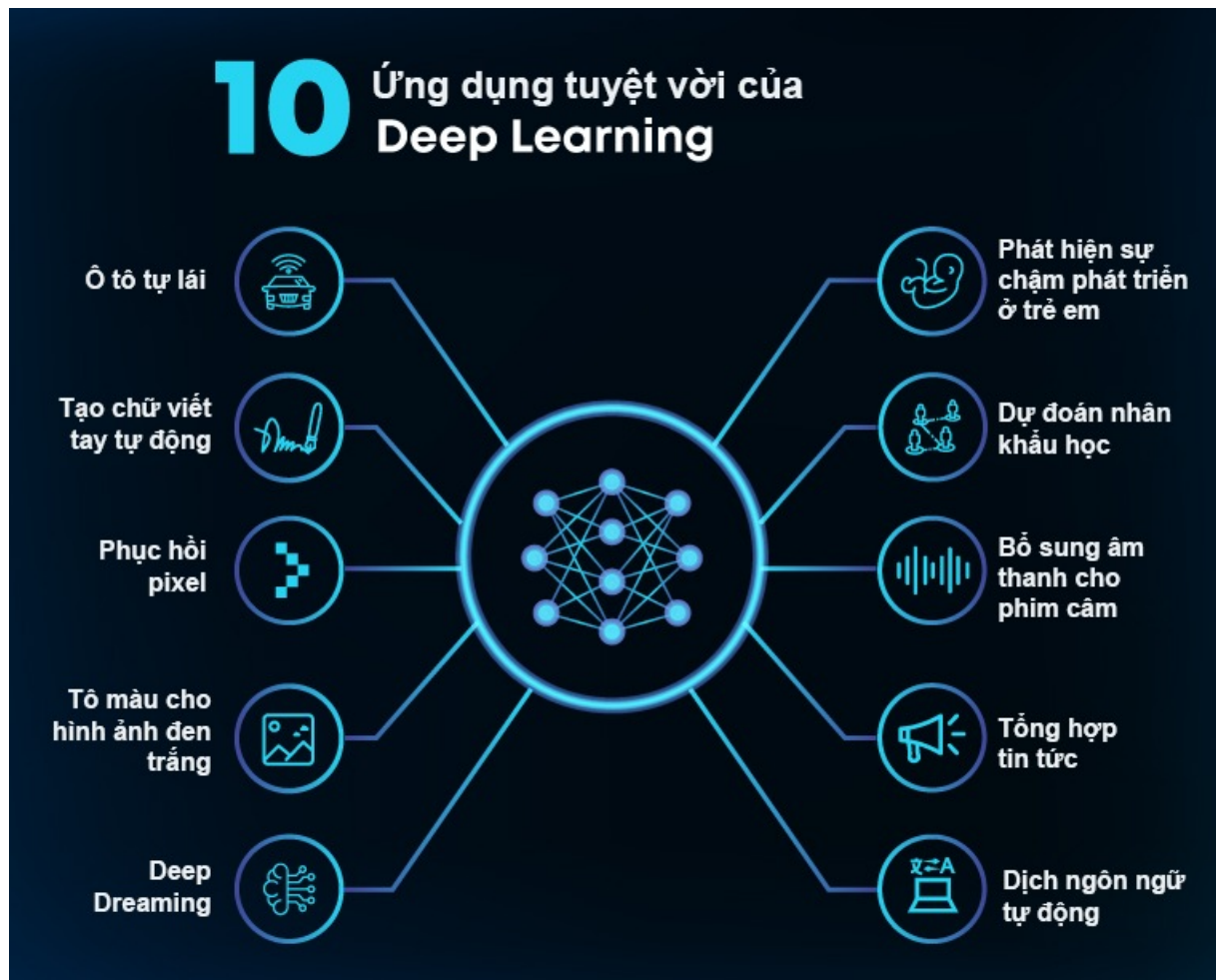
Chăm sóc sức khỏe: Một lĩnh vực khác đã chứng kiến sự tăng trưởng và thay đổi to lớn là lĩnh vực chăm sóc sức khỏe. Từ trợ lý ảo cá nhân đến các bài tập thể dục phục hồi sức khỏe, máy tính đang ghi lại rất nhiều dữ liệu về tình trạng sinh lý và tinh thần mỗi giây của bệnh nhân. Phát hiện sớm các bệnh và tình trạng, hình ảnh định lượng, phẫu thuật robot và có sẵn các công cụ hỗ trợ quyết định cho các chuyên gia hóa ra lại là những người thay đổi cuộc chơi trong lĩnh vực khoa học đời sống, chăm sóc sức khỏe và y học.

Ô tô tự lái: Uber AI Labs ở Pittsburg đang tham gia vào một số công việc to lớn để biến những chiếc xe tự lái thành hiện thực cho thế giới. Deep Learning, tất nhiên, là nguyên tắc chỉ đạo đằng sau sáng kiến này cho tất cả các đại gia ô tô. Các thử nghiệm được thực hiện với một số chiếc xe tự lái đang học tốt hơn với ngày càng nhiều hơn. Deep learning cho phép một chiếc xe không người lái điều hướng bằng cách đưa cho nó hàng triệu tình huống, giúp nó học và xử lý, mang đến một chuyến đi an toàn và thoải mái cho người dùng. Dữ liệu từ các cảm biến, GPS, lập bản đồ địa lý được kết hợp với nhau trong quá trình học sâu để tạo ra các mô hình chuyên xác định đường đi, biển báo đường phố, các yếu tố động như giao thông, tắc nghẽn và người đi bộ.

Tạo văn bản: Chẳng bao lâu nữa, học sâu sẽ tạo ra văn bản gốc (thậm chí là thơ), vì các công nghệ để tạo văn bản đang phát triển rất nhanh. Tất cả mọi thứ từ bộ dữ liệu lớn bao gồm văn bản từ internet đến Shakespeare đều được đưa vào các mô hình học sâu để học và mô phỏng sự sáng tạo của con người với chính tả, dấu câu, ngữ pháp, văn phong và giọng điệu hoàn hảo. Nó đã tạo chú thích / tiêu đề trên rất nhiều nền tảng là minh chứng cho những gì chúng ta hướng đến trong tương lai.

Nhận dạng hình ảnh: Mạng thần kinh chuyển đổi cho phép xử lý hình ảnh kỹ thuật số có thể được phân tách thành nhận dạng khuôn mặt, nhận dạng đối tượng, phân tích chữ viết tay, v.v. Hiện nay máy tính có thể nhận dạng hình ảnh bằng Deep learning. Công nghệ nhận dạng hình ảnh dựa trên công nghệ xử lý hình ảnh kỹ thuật số và sử dụng công nghệ trí tuệ nhân tạo, đặc biệt là phương pháp học máy, để làm cho máy tính nhận ra nội dung trong hình ảnh. Các ứng dụng khác bao gồm tô màu hình ảnh đen trắng và thêm

âm thanh vào các bộ phim câm đã là một kỳ công rất tham vọng đối với các nhà khoa học dữ liệu và chuyên gia trong lĩnh vực này.



Hình 4. Ứng dụng của Deep Learning

2.1.2 Bài toán nhận dạng khuôn mặt trong thị giác máy tính

Nhận dạng khuôn mặt (Face Recognition) là một phương pháp sinh trắc để xác định hoặc xác minh một cá nhân nào đó bằng cách so sánh dữ liệu hình ảnh chụp trực tiếp hoặc hình ảnh kỹ thuật số với bản ghi được lưu trữ cho người đó. Nó được xem là một lĩnh vực nghiên cứu của ngành Biometrics (tương tự như nhận dạng vân tay – Fingerprint Recognition, hay nhận dạng mống mắt – Iris Recognition).

Xét về nguyên tắc chung, nhận dạng khuôn mặt có sự tương đồng rất lớn với nhận dạng vân tay và nhận dạng mống mắt, tuy nhiên sự khác biệt nằm ở bước trích chọn đặc trưng (feature extraction) của mỗi lĩnh vực. Trong khi nhận dạng vân tay và mống mắt đã đạt tới độ chín, tức là có thể áp dụng trên thực tế một cách rộng rãi thì nhận dạng khuôn mặt người vẫn còn nhiều thách thức và vẫn là một lĩnh vực nghiên cứu thú vị với nhiều người. So với nhận dạng vân tay và mống mắt, nhận dạng khuôn mặt có nguồn dữ liệu phong phú hơn (chúng ta có thể nhìn thấy mặt người ở bất cứ tấm ảnh, video clip nào liên quan tới con người trên mạng) và ít đòi hỏi sự tương tác có kiểm soát hơn (để thực hiện nhận dạng vân tay hay mống mắt, dữ liệu input lấy từ con người đòi hỏi có sự hợp tác trong môi trường có kiểm soát).

Các hệ thống nhận dạng khuôn mặt thường được sử dụng cho các mục đích an ninh như kiểm soát an ninh tại tòa nhà, sân bay, máy ATM, tra cứu thông tin của tội phạm, phát hiện tội phạm ở nơi công cộng, ... và ngày càng được ứng dụng rộng rãi trong cuộc sống.

Bên cạnh những thành công đã được ghi nhận thì nhận dạng khuôn mặt cũng còn gặp nhiều khó khăn như về độ sáng, hướng nghiêng, kích thước hình ảnh, diện mạo, biểu hiện cảm xúc của khuôn mặt hay ảnh hưởng của tham số môi trường.

Để xây dựng một hệ thống nhận dạng khuôn mặt có đầu vào của hệ thống là một hình ảnh kỹ thuật số hay một khung hình video từ một nguồn video. Đầu ra là xác định hoặc xác minh người ở trong bức hình hoặc trong video đó là ai. Hướng tới mục tiêu này chúng ta thường chia thủ tục nhận dạng khuôn mặt gồm ba bước: Phát hiện khuôn mặt, trích rút đặc trưng và nhận dạng khuôn mặt.



Hình 5. Hệ thống nhận diện khuôn mặt

Phát hiện khuôn mặt (Face Detection): Chức năng chính của bước này là phát hiện ra khuôn mặt xem nó có xuất hiện ở trong một bức hình hay một đoạn video hay không? Tỷ lệ phát hiện ra khuôn mặt phụ thuộc nhiều vào điều kiện về độ sáng, hướng khuôn mặt, biểu hiện cảm xúc trên khuôn mặt hay các yếu tố môi trường khác. Để hệ thống nhận dạng hoạt động đạt hiệu quả cao thì hình ảnh khuôn mặt sau khi được phát hiện cần chuẩn hóa về kích thước, ánh sáng.

Trích rút đặc trưng (Feature Extraction): Sau khi phát hiện ra khuôn mặt trong bức ảnh, chúng ta tiến hành trích rút những đặc trưng của khuôn mặt. Bước này trích xuất ra một vector đặc trưng đại diện cho một khuôn mặt. Nó phải đảm bảo được tính duy nhất của một khuôn mặt.

Nhận dạng khuôn mặt (Face Recognition): Với hình ảnh đầu vào sau khi phát hiện ra khuôn mặt, trích rút các đặc trưng của khuôn mặt và đem so sánh các đặc trưng này với cơ sở dữ liệu khuôn mặt.

Bài toán nhận dạng khuôn mặt được ứng dụng nhiều trong các lĩnh vực đời sống đặc biệt ở những lĩnh vực công nghệ cao, yêu cầu về an ninh, bảo mật. Do đó để hệ thống nhận dạng khuôn mặt hoạt động mạnh mẽ với tốc độ và độ tin cậy thì có rất nhiều các phương pháp về nhận dạng khuôn mặt được đưa ra. Các phương pháp có thể được phân loại theo các tiêu chí khác nhau như nhận dạng với dữ liệu ảnh đầu vào là ảnh tĩnh 2D (Elastic Bunch Graph, Active Appearance Model). Phương pháp này là phổ biến nhất và tương lai sẽ là 3D (3D Morphable Model). Tuy nhiên trên thực tế người ta hay chia phương pháp nhận dạng khuôn mặt ra thành 2 loại:

- Nhận dạng dựa trên các đặc trưng của các phần tử trên khuôn mặt (Feature Base Face Recognition)
- Nhận dạng dựa trên xét tổng thể toàn khuôn mặt (Appearance Based Face Recognition).

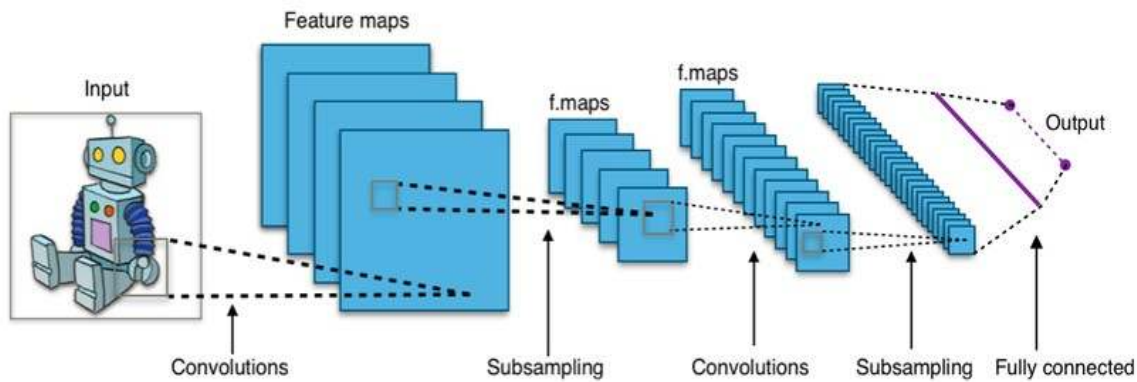
2.2 Giới thiệu Convolutional Neural Network

2.2.1 Tổng quan

Convolutional Neural Network (CNN) bao gồm một hoặc nhiều lớp chập (thường với một bước lấy mẫu con) và sau đó theo sau bởi một hoặc nhiều hơn các lớp kết nối như trong một mạng nơ-ron đa lớp chuẩn. Kiến trúc của một CNN được thiết kế để tận dụng lợi thế của cấu trúc 2 chiều của một hình ảnh đầu vào (hoặc đầu vào 2 chiều khác như một tín hiệu tiếng nói). Điều này đạt được với các kết nối cục bộ và trọng số ràng buộc theo một số hình thức tổng hợp mà kết quả là các đặc trưng không thay đổi. Một lợi ích khác của CNN là dễ dàng huấn luyện hơn và có ít thông số so với các mạng kết nối đầy đủ với cùng một số đơn vị ẩn.

2.2.2 Kiến trúc CNN

CNN gồm một số lớp chập và lớp lấy mẫu con tùy chọn, tiếp theo là các lớp | kết nối. Đầu vào cho một lớp chập là một hình ảnh $m \times m$, với m là chiều cao và chiều rộng của hình ảnh và r là số kênh (ví dụ một ảnh RGB có $r=3$). Lớp chập sẽ có k bộ lọc có kích thước $n \times n$, trong đó n nhỏ hơn kích thước hình ảnh và q có thể bằng hoặc nhỏ hơn số kênh r . Kích thước của các bộ lọc tăng đến cấu trúc kết nối cục bộ, trong đó mỗi bộ lọc xoắn với hình ảnh để tạo ra k bản đồ đặc trưng có kích thước $m - n + 1$. Mỗi bản đồ được lấy mẫu con đặc trưng với việc hợp nhất trung bình hoặc lớn nhất trên 2×2 khu vực lân cận, trong đó phạm vi của p là giữa 2 (cho các hình ảnh nhỏ) và 5 (chỉ các đầu vào lớn hơn). Trước hoặc sau khi lấy mẫu con, một bias bổ sung và tích ma phi tuyến được áp dụng cho mỗi bản đồ đặc trưng.



Hình 6. Kiến trúc CNN trong nhận diện hành động con người

Chập hình ảnh đầu vào với các bộ lọc huấn luyện khác nhau và các bias bổ sung, nhiều bản đồ đặc trưng được tạo trong lớp C1. Mỗi bản đồ đặc trưng trong S2 thu được bởi thao tác tổng hợp các bản đồ đặc trưng tương ứng trong lớp C1. Chập và tổng hợp cực đại trong lớp C3 và S4 thì giống trong lớp C1 và S2. Trong bước nhận diện cuối cùng, các đặc trưng thu được sau khi tổng hợp cực đại trong lớp S4 thì được mã hóa thành một vector 1 chiều.

2.3 Tổng quan về Machine Learning

2.3.1 Machine Learning là gì ?

Machine Learning là một thuật ngữ rộng để chỉ hành động bạn dạy máy tính cải thiện một nhiệm vụ mà nó đang thực hiện. Cụ thể hơn, machine learning đề cập tới bất kỳ hệ thống mà hiệu suất của máy tính khi thực hiện một nhiệm vụ sẽ trở nên tốt hơn sau khi hoàn thành nhiệm vụ đó nhiều lần. Hay nói cách khác, khả năng cơ bản nhất của machine learning là sử dụng thuật toán để phân tích những thông tin có sẵn, học hỏi từ nó rồi đưa ra quyết định hoặc dự đoán về một thứ gì đó có liên quan. Thay vì tạo ra một phần mềm với những hành động, hướng dẫn chi tiết để thực hiện một nhiệm vụ cụ thể, máy tính được “huấn luyện” bằng cách sử dụng lượng dữ liệu và các thuật toán để học cách thực hiện nhiệm vụ.

Có 2 loại Máy học chính bao gồm học có giám sát (supervised learning) và học không giám sát (unsupervised learning).

- Học có giám sát - trong đó, thuật toán tạo ra một hàm ánh xạ dữ liệu vào tới kết quả mong muốn. Một phát biểu chuẩn về một việc học có giám sát là bài toán phân loại:

chương trình cần học (cách xấp xỉ biểu hiện của) một hàm ánh xạ một vector tới một vài lớp bằng cách xem xét một số mẫu dữ liệu – kết quả của hàm đó.

- Học không giám sát - mô hình hóa một tập dữ liệu, không có sẵn các ví dụ đã được gắn nhãn.

2.3.2 Một vài ứng dụng của Machine Learning

- **Cảnh báo giao thông:** Giờ đây, Google Maps có lẽ là ứng dụng được sử dụng với tần suất nhiều nhất mỗi khi bạn tham gia giao thông. Đặc biệt khi các ứng dụng khác về di chuyển như Grab, Be được áp dụng rộng rãi, đồng nghĩa Google Maps được sử dụng liên tục để chỉ đường cho nhà cung cấp dịch vụ hay người sử dụng dịch vụ. Những thông tin về quãng đường tối ưu, thời gian di chuyển nhanh nhất cũng được phân tích cùng lúc trên Google Maps. Dữ liệu lịch sử của tuyến đường đó đã được thu thập theo thời gian và một số dữ liệu có từ các nguồn khác. Mọi người sử dụng bản đồ đều cung cấp vị trí, tốc độ trung bình, tuyến đường. Những thông tin này Google thu thập và tổng hợp thành Dữ liệu lớn về lưu lượng truy cập, thông qua các thuật toán phân tích phức tạp trên Machine Learning, những thông tin này trở nên có nghĩa, chúng giúp Google dự đoán lưu lượng sắp tới và điều chỉnh tuyến đường của bạn theo cách tối ưu nhất.
- **Mạng xã hội Facebook:** Một trong những ứng dụng phổ biến nhất của Machine Learning là Đề xuất gắn thẻ bạn bè tự động trên Facebook hoặc bất kỳ nền tảng truyền thông xã hội nào khác. Facebook sử dụng tính năng nhận diện khuôn mặt và nhận dạng hình ảnh để tự động tìm thấy khuôn mặt của người phù hợp với Cơ sở dữ liệu của họ và do đó đề nghị người dùng gắn thẻ người đó dựa trên DeepFace.
- **Trợ lý cá nhân ảo:** Trợ lý cá nhân ảo hỗ trợ tìm kiếm thông tin hữu ích, khi được yêu cầu qua văn bản hoặc giọng nói. Một số ứng dụng chính của Machine Learning ở đây là:
 - Nhận dạng giọng nói
 - Chuyển đổi lời nói thành văn bản
 - Xử lý ngôn ngữ tự nhiên
 - Chuyển đổi văn bản thành giọng nói

Tất cả những gì bạn cần làm là hỏi một câu hỏi đơn giản như Lịch trình của tôi vào ngày mai là gì? hoặc các chuyến bay có sẵn sắp tới cho chuyến công tác của tôi. Để trả lời, trợ lý cá nhân của bạn tìm kiếm thông tin hoặc nhớ lại các truy vấn liên quan của bạn để thu thập thông tin.

- **Phát hiện gian lận:** Phát hiện gian lận là một trong những tính năng quan trọng nhất của Machine Learning. Bất cứ khi nào khách hàng thực hiện giao dịch – mô hình máy học sẽ chụp kỹ lưỡng hồ sơ của họ để tìm kiếm các mẫu đáng ngờ. Trong Machine Learning, các tính năng như phát hiện gian lận thường được coi là một loại phân loại hoạt động.

2.4 Thư viện Keras

Keras là một thư viện nơ-ron mã nguồn mở được viết bằng ngôn ngữ lập trình Python. Nó có khả năng chạy trên đầu trang của Tensorflow, Microsoft Cognitive Toolkit, R, Theano hoặc StripeML. Được thiết kế để cho phép thử nghiệm nhanh với các mạng thần kinh sâu, nó tập trung vào việc thân thiện với người dùng, mô-đun và mở rộng.

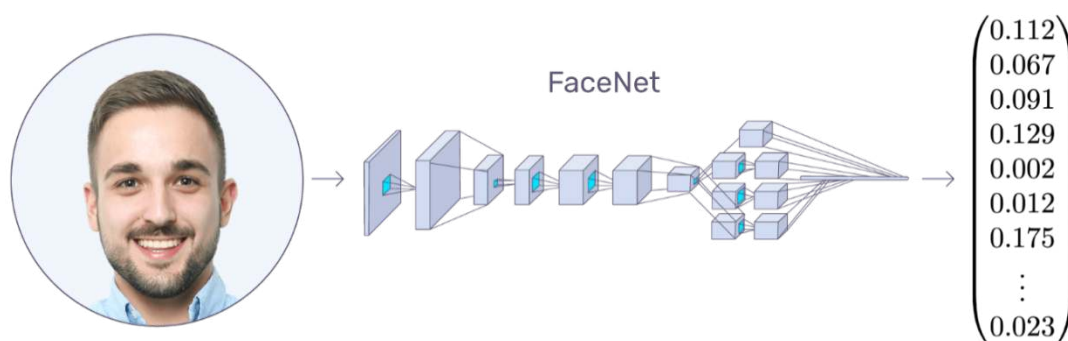
Keras được coi là một thư viện ‘high-level’ với phần ‘low-level’ (còn được gọi là backend) có thể là TensorFlow, CNTK, hoặc Theano (sắp tới Theano sẽ không được duy trì nâng cấp nữa). Keras có cú pháp đơn giản hơn TensorFlow rất nhiều. Với mục đích giới thiệu về các mô hình nhiều hơn là các sử dụng các thư viện deep learning, tôi sẽ chọn Keras với TensorFlow là ‘backend’.

Một số tính năng của Keras:

- Keras ưu tiên trải nghiệm của người lập trình
- Keras đã được sử dụng rộng rãi trong doanh nghiệp và cộng đồng nghiên cứu
- Keras giúp dễ dàng biến các thiết kế thành sản phẩm
- Keras hỗ trợ huấn luyện trên nhiều GPU phân tán
- Keras hỗ trợ đa backend engines và không giới hạn bạn vào một hệ sinh thái.

2.5 Thư viện FaceNet

FaceNet là một mạng lưới thần kinh sâu được sử dụng để trích xuất các tính năng từ hình ảnh của một người mặt. Nó được xuất bản vào năm 2015 bởi các nhà nghiên cứu của Google.



Hình 7.FaceNet lấy hình ảnh khuôn mặt làm đầu vào và xuất ra vector embedding

FaceNet lấy hình ảnh của mặt người làm đầu vào và xuất ra một vector 128 chiều, đại diện cho các tính năng quan trọng nhất của khuôn mặt. Trong học máy, vector này được gọi là nhúng (embeddings). Tại sao phải nhúng? Bởi vì tất cả các thông tin quan trọng từ một hình ảnh được nhúng vào vector này. Về cơ bản, FaceNet lấy một mặt người và nén nó thành một vector gồm 128 số. Khuôn mặt cần định danh cũng có nhúng tương tự.

Facenet chính là một dạng siam network có tác dụng biểu diễn các bức ảnh trong một không gian euclidean n chiều (thường là 128) sao cho khoảng cách giữa các vector embedding càng nhỏ, mức độ tương đồng giữa chúng càng lớn.

Hầu hết các thuật toán nhận diện khuôn mặt trước facenet đều tìm cách biểu diễn khuôn mặt bằng một vector nhúng (embedding) thông qua một layer bottle neck có tác dụng giảm chiều dữ liệu:

- Tuy nhiên hạn chế của các thuật toán này đó là số lượng chiều embedding tương đối lớn (thường ≥ 1000) và ảnh hưởng tới tốc độ của thuật toán. Thường chúng ta phải áp dụng thêm thuật toán PCA để giảm chiều dữ liệu để giảm tốc độ tính toán.

- Hàm loss function chỉ đo lường khoảng cách giữa 2 bức ảnh. Như vậy trong một đầu vào huấn luyện chỉ học được một trong hai khả năng là sự giống nhau nếu chúng cùng 1 class hoặc sự khác nhau nếu chúng khác class mà không học được cùng lúc sự giống nhau và khác nhau trên cùng một lượt huấn luyện.

Facenet đã giải quyết cả 2 vấn đề trên bằng các hiệu chỉnh nhỏ nhưng mang lại hiệu quả lớn:

- Base network áp dụng một mạng convolutional neural network và giảm chiều dữ liệu xuống chỉ còn 128 chiều. Do đó quá trình suy diễn và dự báo nhanh hơn và đồng thời độ chính xác vẫn được đảm bảo.

- Sử dụng loss function là hàm triplet loss có khả năng học được đồng thời sự giống nhau giữa 2 bức ảnh cùng nhóm và phân biệt các bức ảnh không cùng nhóm. Do đó hiệu quả hơn rất nhiều so với các phương pháp trước đây.

2.6 Thư viện Sklearn

Scikit-learn (Sklearn) là thư viện mạnh mẽ nhất dành cho các thuật toán học máy được viết trên ngôn ngữ Python. Thư viện cung cấp một tập các công cụ xử lý các bài toán machine learning và statistical modeling gồm: classification, regression, clustering, và dimensionality reduction.

Thư viện này tích hợp rất nhiều thuật toán hiện đại và cổ điển giúp bạn vừa học vừa tiến hành đưa ra các giải pháp hữu ích cho bài toán của bạn một cách đơn giản.

Thư viện được cấp phép bản quyền chuẩn FreeBSD và chạy được trên nhiều nền tảng Linux. Scikit-learn được sử dụng như một tài liệu để học tập.

Để cài đặt scikit-learn trước tiên phải cài thư viện SciPy (Scientific Python). Những thành phần gồm:

- Numpy: Gói thư viện xử lý dãy số và ma trận nhiều chiều
- SciPy: Gói các hàm tính toán logic khoa học

- Matplotlib: Biểu diễn dữ liệu dưới dạng đồ thị 2 chiều, 3 chiều
- IPython: Notebook dùng để tương tác trực quan với Python
- SymPy: Gói thư viện các kí tự toán học
- Pandas: Xử lý, phân tích dữ liệu dưới dạng bảng

Thư viện tập trung vào việc mô hình hóa dữ liệu. Nó không tập trung vào việc truyền tải dữ liệu, biến đổi hay tổng hợp dữ liệu.

Sau đây là một số nhóm thuật toán được xây dựng bởi thư viện scikit-learn:

- **Clustering**: Nhóm thuật toán Phân cụm dữ liệu không gán nhãn. Ví dụ thuật toán KMeans
- **Cross Validation**: Kiểm thử chéo, đánh giá độ hiệu quả của thuật toán học giám sát sử dụng dữ liệu kiểm thử (validation data) trong quá trình huấn luyện mô hình.
- **Datasets**: Gồm nhóm các Bộ dữ liệu được tích hợp sẵn trong thư viện. Hầu như các bộ dữ liệu đều đã được chuẩn hóa và mang lại hiệu suất cao trong quá trình huấn luyện như iris, digit, ...
- **Dimensionality Reduction**: Mục đích của thuật toán này là để Giảm số lượng thuộc tính quan trọng của dữ liệu bằng các phương pháp như tổng hợp, biểu diễn dữ liệu và lựa chọn đặc trưng. Ví dụ thuật toán PCA (Principal component analysis).
- **Ensemble methods**: Các Phương pháp tập hợp sử dụng nhiều thuật toán học tập để có được hiệu suất dự đoán tốt hơn so với bất kỳ thuật toán học cấu thành nào.
- **Feature extraction**: Trích xuất đặc trưng. Mục đích là để định nghĩa các thuộc tính với dữ liệu hình ảnh và dữ liệu ngôn ngữ.
- **Feature selection**: Trích chọn đặc trưng. Lựa chọn các đặc trưng có ý nghĩa trong việc huấn luyện mô hình học giám sát.
- **Parameter Tuning**: Tinh chỉnh tham số. Các thuật toán phục vụ việc lựa chọn tham số phù hợp để tối ưu hóa mô hình.
- **Manifold Learning**: Các thuật toán học tổng hợp và Phân tích dữ liệu đa chiều phức tạp.

- **Supervised Models:** Học giám sát. Mảng lớn các thuật toán học máy hiện nay. Ví dụ như linear models, discriminate analysis, naive bayes, lazy methods, neural networks, support vector machines và decision trees.

Chương 3 XÂY DỰNG ỨNG DỤNG

3.1 Yêu cầu hệ thống

3.1.1 Phần cứng

- RAM 4 GB
- CPU Intel Core i5
- Ổ cứng SSD

3.1.2 Môi trường thực hiện

- Ngôn ngữ Python 3
- Thư viện Multi Task Convolutional Neural Network và FaceNet
- Hệ điều hành: Window 10

3.2 Các bước thực hiện

3.2.1 Chuẩn bị môi trường

Đầu tiên chúng ta cần cài đặt Python, nên chọn các phiên bản phù hợp với từng khoảng thời gian và dự án hiện tại. Ở đây chúng tôi đã cài đặt phiên bản 3.11

Python version	Maintenance status	First released	End of support	Release schedule
3.12	prerelease	2023-10-02 (planned)	2028-10	PEP 693
3.11	bugfix	2022-10-24	2027-10	PEP 664
3.10	security	2021-10-04	2026-10	PEP 619
3.9	security	2020-10-05	2025-10	PEP 596
3.8	security	2019-10-14	2024-10	PEP 569
3.7	security	2018-06-27	2023-06-27	PEP 537

Sau khi cài đặt môi trường chúng ta bước đến xây dựng thư viện cần sử dụng như:

- Tensorflow
- Keras
- Pandas
- OpenCv
- Numpy

3.2.2 Chuẩn bị dữ liệu

Dữ liệu gồm 2 dataset: 1 dataset gồm 7178 bức ảnh để mô tả 1 cảm xúc khác nhau bao gồm: bình thường, hạnh phúc, tức giận, sợ hãi, ghê sợ, buồn và ngạc nhiên, 1 dataset gồm dữ liệu bao gồm các bức ảnh từ độ tuổi 0 đến 116 gồm hơn 20.000 tấm ảnh khác nhau.

Vì bộ dữ liệu chủ yếu là khuôn mặt của người nên model đưa ra tuy khá tốt nhưng chưa có độ chính xác cao về cảm xúc vì cần phải xét dựa trên bối cảnh của bức ảnh.

3.2.3 Chạy chương trình

Đầu tiên chúng ta sẽ training model giới tính và độ tuổi. Khởi tạo với số lần chạy là 60 epochs cho mỗi model. Sử dụng 75% ảnh để train và 25% để test

```
training.py NhanDienChuyenDe X
training.py
1 import pandas as pd
2 import numpy as np
3 import os
4 import matplotlib.pyplot as plt
5 import cv2
6 from keras.models import Sequential,load_model,Model
7 from keras.layers import Conv2D,MaxPool2D,Dense,Dropout,BatchNormalization,Flatten,Input
8 from sklearn.model_selection import train_test_split
9
10 path = "Datasets"
11
12
13
14
15 images = []
16 age = []
17 gender = []
18
19
20 print('dang nap anh.....')
21
22 for img in os.listdir(path):
23     ages = img.split("_")[0]
24     genders = img.split("_")[1]
25     img = cv2.imread(str(path)+"/"+str(img))
26     img = cv2.cvtColor(img,cv2.COLOR_BGR2RGB)
27     images.append(np.array(img))
28     age.append(np.array(ages))
29     gender.append(np.array(genders))
30
31 age = np.array(age,dtype=np.int64)
32 images = np.array(images) #Forgot to scale image for my training. Please divide by 255 to scale.
33 gender = np.array(gender,np.uint64)
34
35
36 # 75% TRAI VÀ 25% TEST6
37 x_train_age, x_test_age, y_train_age, y_test_age = train_test_split(images, age, random_state=42)
38
39 x_train_gender, x_test_gender, y_train_gender, y_test_gender = train_test_split(images, gender, random_state=42)
40
```

```

training.py NhanDienChuyenDe X
training.py
41 #####
42 #Define age model and train.
43 #####
44
45 age_model = Sequential()
46 age_model.add(Conv2D(128, kernel_size=3, activation='relu', input_shape=(200,200,3)))
47 #age_model.add(Conv2D(128, kernel_size=3, activation='relu'))
48 age_model.add(MaxPool2D(pool_size=3, strides=2))
49
50 age_model.add(Conv2D(128, kernel_size=3, activation='relu'))
51 #age_model.add(Conv2D(128, kernel_size=3, activation='relu'))
52 age_model.add(MaxPool2D(pool_size=3, strides=2))
53
54 age_model.add(Conv2D(256, kernel_size=3, activation='relu'))
55 #age_model.add(Conv2D(256, kernel_size=3, activation='relu'))
56 age_model.add(MaxPool2D(pool_size=3, strides=2))
57
58 age_model.add(Conv2D(512, kernel_size=3, activation='relu'))
59 #age_model.add(Conv2D(512, kernel_size=3, activation='relu'))
60 age_model.add(MaxPool2D(pool_size=3, strides=2))
61
62 age_model.add(Flatten())
63 age_model.add(Dropout(0.2))
64 age_model.add(Dense(512, activation='relu'))
65
66 age_model.add(Dense(1, activation='linear', name='age'))
67
68 age_model.compile(optimizer='adam', loss='mse', metrics=['mae'])
69 print(age_model.summary())
70
71 history_age = age_model.fit(x_train_age, y_train_age,
72                             validation_data=(x_test_age, y_test_age), epochs=60)
73
74 age_model.save('age_model_50epochs.h5')
75
76 #####
77 #Define gender model and train

```

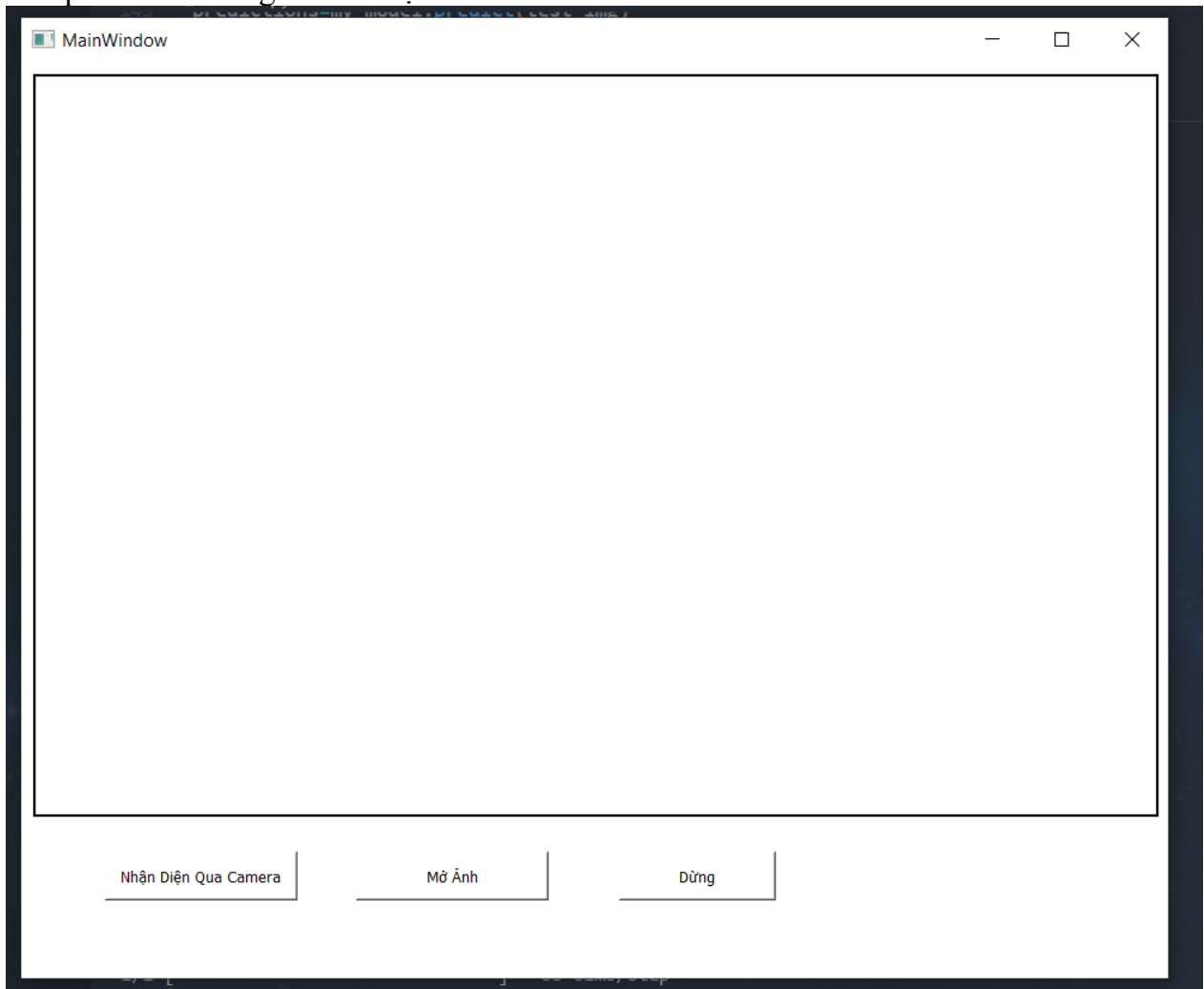
Tương tự làm với training model về cảm xúc. Ở đây sử dụng 50 epochs và tỉ lệ train và test là 70% và 30%

```

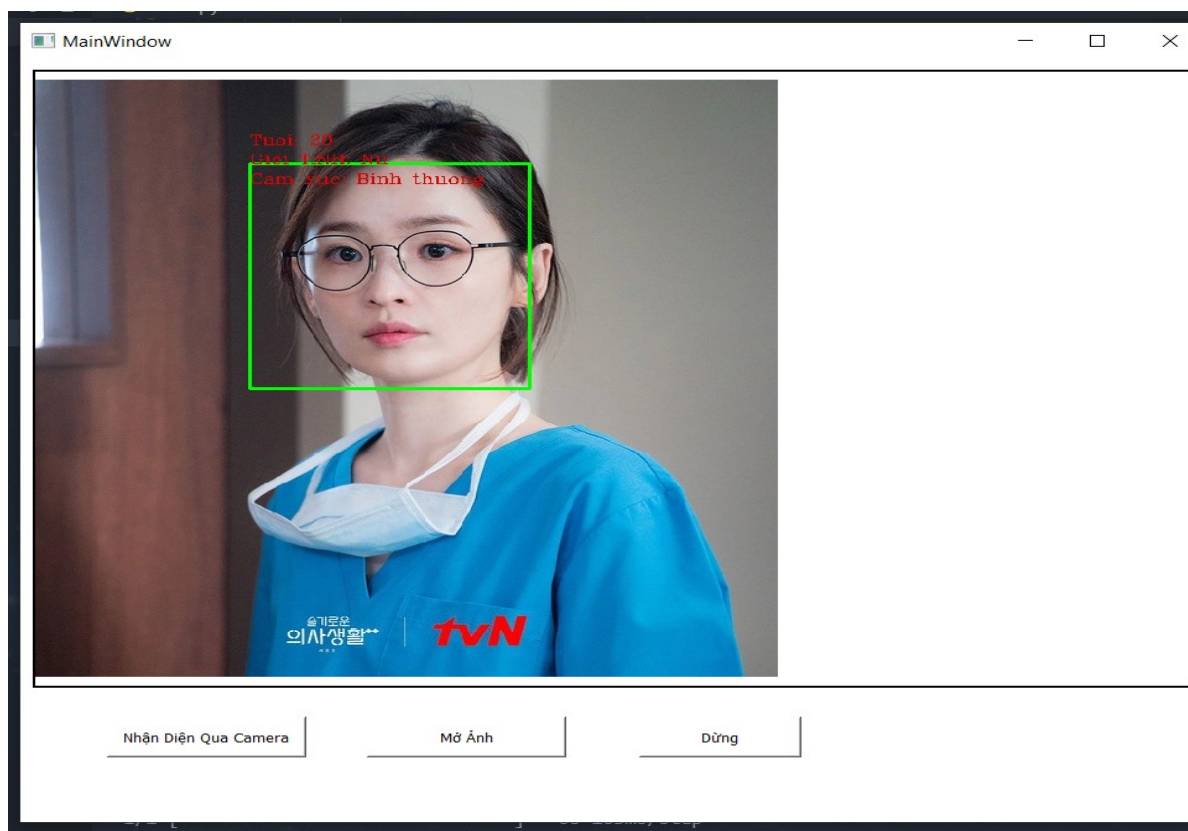
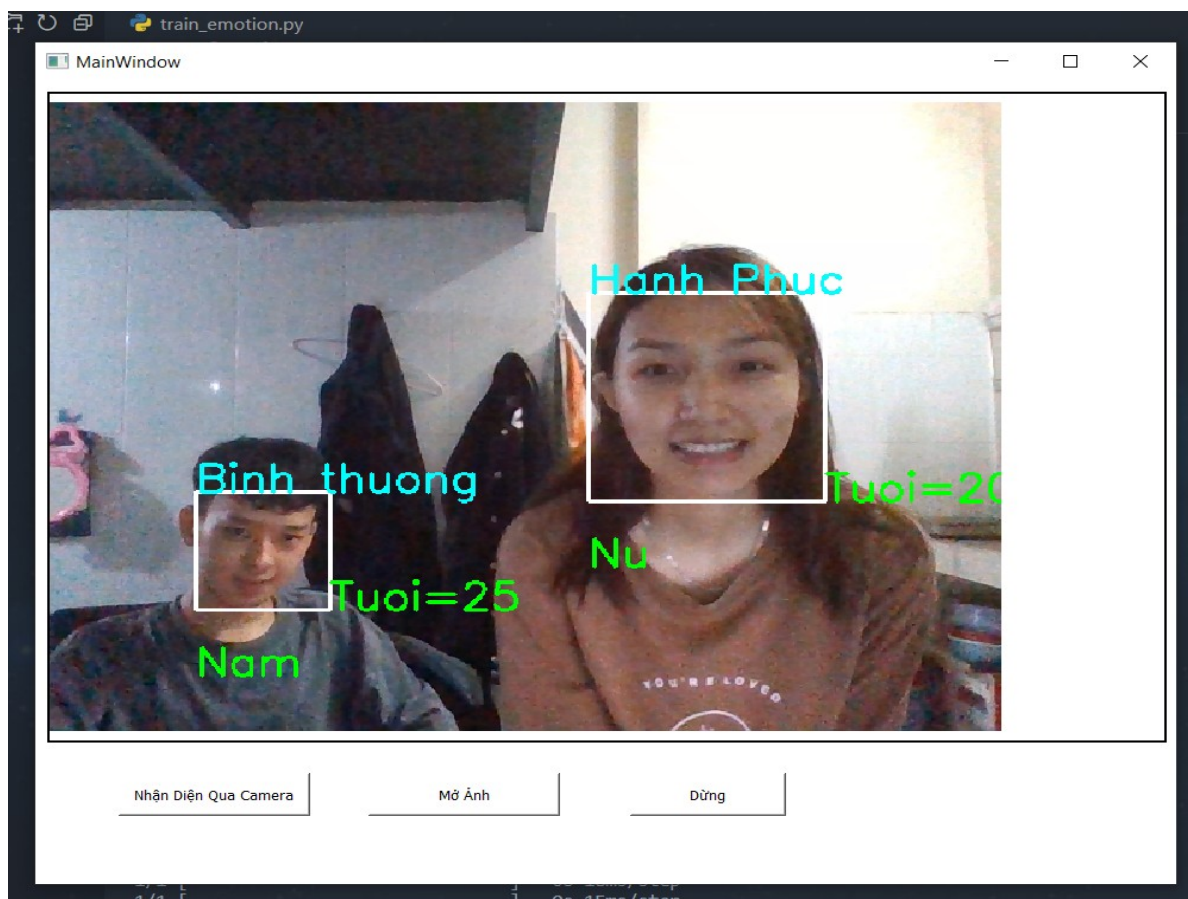
training.py NhanDienChuyenDe • train_emotion.py NhanDienChuyenDe X
train_emotion.py
81 model.add(Dense(7, activation='softmax'))
82
83 model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
84 print(model.summary())
85
86
87 train_path = "dataEmotion/train/"
88 test_path = "dataEmotion/test"
89
90 num_train_imgs = 0
91 for root, dirs, files in os.walk(train_path):
92     num_train_imgs += len(files)
93
94 num_test_imgs = 0
95 for root, dirs, files in os.walk(test_path):
96     num_test_imgs += len(files)
97
98
99 epochs=50
100
101 history=model.fit(train_generator,
102                  steps_per_epoch=num_train_imgs//batch_size,
103                  epochs=epochs,
104                  validation_data=validation_generator,
105                  validation_steps=num_test_imgs//batch_size)
106
107 model.save('emotion_detection_model_100epochs.h5')
108
109 #plot the training and validation accuracy and loss at each epoch
110 loss = history.history['loss']
111 val_loss = history.history['val_loss']
112 epochs = range(1, len(loss) + 1)
113 plt.plot(epochs, loss, 'y', label='Training loss')
114 plt.plot(epochs, val_loss, 'r', label='Validation loss')
115 plt.title('Training and validation loss')
116 plt.xlabel('Epochs')
117 plt.ylabel('Loss')
118 plt.legend()
119 plt.show()

```

Sản phẩm cuối cùng ta thu được như sau:



Chương trình gồm 2 chức năng chính là nhận diện thông qua Webcam, nhận diện qua ảnh



KẾT LUẬN

4.1 Kết quả.

Với kết quả thu được là chương trình nhận giới tính là nam hay nữ khi được cung cấp ảnh, có thể được sử dụng để ứng dụng vào nhiều vấn đề thực tế như nhận diện khuôn mặt ở các cửa hàng tiện lợi, mở khóa điện thoại, nhận diện học sinh, quản lý nhân sự, xu hướng khách hàng...

4.2 Ưu điểm - nhược điểm

Ưu điểm:

- Chương trình nhận diện khá chính xác khi được cung cấp ảnh chất lượng tốt như rõ nét, chân dung,...
- Chương trình có thể nhận diện được nam giả nữ hoặc ngược lại khi chất lượng ảnh tốt.
- Nhận diện được đa số mọi lứa tuổi.
- Nhanh chóng, thuận tiện, chỉ cần ảnh có sẵn.

Nhược điểm:

- Vẫn còn có trường hợp dự đoán giới tính sai như khi chất lượng ảnh thấp ảnh bị mờ nhiều chỗ khuất.
- Chương trình có khả năng dự đoán sai ở 1 vài trường hợp như: độ tuổi của người được cung cấp quá nhỏ để phân biệt, người trong ảnh trang điểm hoặc ăn mặc quá đậm quá khác dẫn đến kết quả dự đoán sẽ bị sai.
- Tuy nhiên sự sai sót này có thể bỏ qua vì bài toán nhận dạng giới tính thường dùng để áp dụng cho các vấn đề về quản lý nhân sự, xu hướng khách hàng,.... và độ tuổi trẻ em thì thực sự chưa có tác động nhiều trên những vấn đề này.

- Xử lý hình ảnh và sử dụng Deep Learning để huấn luyện sâu, cần có thiết bị có phần cứng có khả năng xử lý tính toán cao để có thể áp dụng cho một lượng dữ liệu khổng lồ như hiện nay.
- Dữ liệu nhận diện còn ít, thiếu sự đa dạng.

4.3 Hướng phát triển.

Có nhiều hướng phát triển cho chương trình này, có thể phát triển cả về mặt ứng dụng và mặt thuật toán (để cải thiện hiệu quả phát hiện giới tính). Có thể xây dựng một ứng dụng hoàn chỉnh để dễ sử dụng hơn.

Có thể loại bỏ những hình ảnh trẻ em trong bộ dữ liệu để thay đổi cách học của model vì trong bài toán này khuôn mặt trẻ em có thể xem như là dữ liệu nhiễu.

Bổ sung thêm dữ liệu để giải quyết tỉ lệ chương trình dự đoán sai giới tính.

Ngoài ra có thể phát triển chương trình theo hướng nhận dạng khuôn mặt, xây dựng một hệ thống để học các đặc trưng của những người cần nhận dạng. Khi thực hiện, đầu tiên ta đưa qua bức ảnh qua chương trình phát hiện mặt người để phát hiện nhanh các khuôn mặt có trong ảnh, sau đấy so sánh các khuôn mặt đó với các khuôn mặt mà chương trình đã được “học” từ trước, so sánh các đặc trưng của hai khuôn mặt, nếu trùng thì đưa ra thông tin về khuôn mặt được nhận dạng.

TÀI LIỆU THAM KHẢO

[1]. Mô hình facenet trong facerecognition,

Nguồn: <https://phamdinhhkhanh.github.io/2020/03/12/faceNetAlgorithm.html>

[2]. How to Develop a Face Recognition System Using FaceNet in Keras

Nguồn: <https://machinelearningmastery.com/how-to-develop-a-face-recognition-system-using-facenet-in-keras-and-an-svmclassifier/>

fbclid=IwAR1FR0AygXgXeHSJnQPy81oPxFH-
mvhkams0OPRNNIqFHrmPvGg45r3TS4ZY

[3]. Tài liệu về thư viện MTCNN.

Nguồn: <https://github.com/ipazc/mtcnn>

[4]. Xây dựng hệ thống nhận dạng giới tính.

Nguồn: http://fit.vimaru.edu.vn/sites/default/files/filedinhkem/towards_building_an_automatic_gender_classification_system_using_lpq.pdf

[5]. Tìm hiểu về MTCNN và áp dụng để xác định vị trí khuôn mặt.

Nguồn: <https://viblo.asia/p/tim-hieu-mtcnn-va-ap-dung-de-xac-dinh-vi-tri-cac-khuon-mat-3Q75wkO75Wb>