

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH
KHOA ĐÀO TẠO CHẤT LƯỢNG CAO
BỘ MÔN CƠ ĐIỆN TỬ



HCMUTE

BÁO CÁO MÔN TRÍ TUỆ NHÂN TẠO
Lớp: 05CLC

Đề tài: NHẬN DẠNG GÓC XOAY KHUÔN
MẶT

GVHD: PGS. TS NGUYỄN TRƯỜNG THỊNH
SVTH: BÙI VŨ LONG
MSSV: 19146205

Tp. Hồ Chí Minh, ngày 11 tháng 06 năm 2022

MỤC LỤC

TỔNG QUAN	4
1. Tổng quan về môn học.	4
2. Tổng quan về đề tài.	6
a. Giới thiệu về đề tài.	6
b. Tầm quan trọng của đề tài.....	6
c. Mục tiêu đề tài.....	7
NỘI DUNG	8
I. GIỚI THIỆU VỀ NHẬN DẠNG KHUÔN MẶT.	8
1. Nhận dạng khuôn mặt.	8
2. Nhận dạng góc xoay trong nhận dạng khuôn mặt.	9
3. Các phương pháp tiếp cận.....	10
II. GIỚI THIỆU VỀ CONVOLUTIONAL NEURAL NETWORK.	11
1. Convolutional Neural Network là gì?.....	11
2. Feature là gì?.....	12
3. Convolutional là gì?.....	12
4. Một số lớp cơ bản của Convolutional Neural Network.	13
5. Cấu trúc của Convolutional Neural Network	14
6. Tham số cho Convolutional Neural Network.....	15
III. XÂY DỰNG KIẾN TRÚC MẠNG CNN NHẬN DẠNG GÓC XOAY KHUÔN MẶT.	16
1. Xây dựng mạng CNNs.....	16
2. Huấn luyện mô hình.....	17
3. Xây dựng mô hình dự đoán:	20
IV. HƯỚNG PHÁT TRIỂN.	21
KẾT LUẬN.	22
TÀI LIỆU THAM KHẢO.	23

MỤC LỤC HÌNH ẢNH

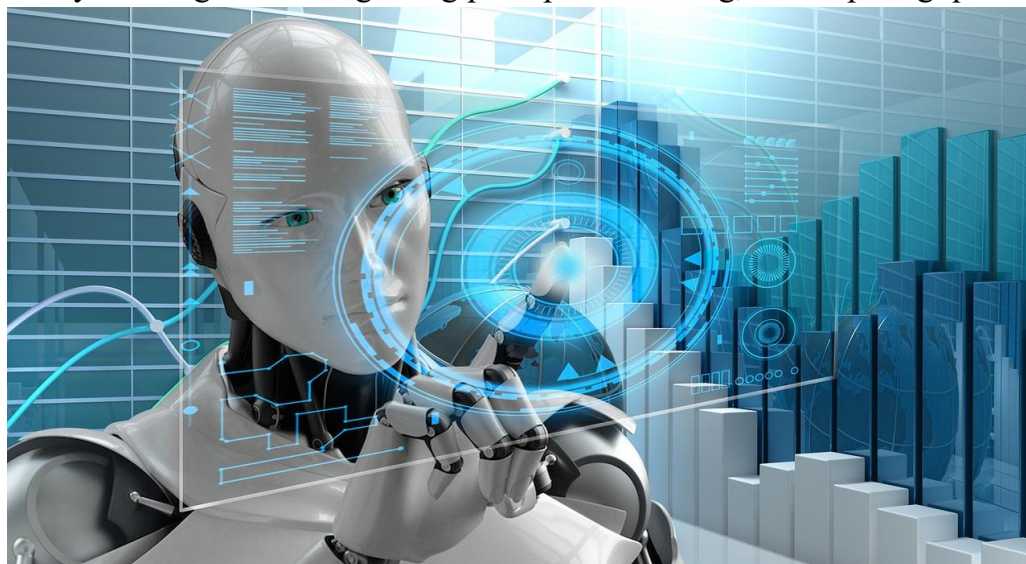
Hình 1. Trí tuệ nhân tạo	4
Hình 2. Robot trí tuệ nhân tạo đầu tiên của Việt Nam.	5
Hình 3. Nhận dạng khuôn mặt	8
Hình 4. Nhận dạng góc xoay khuôn mặt	10
Hình 5. Convolutional Neural Network	11
Hình 6. Mạng lưới Convolutional	12
Hình 7. Lớp Pooling.	13
Hình 8. Cấu trúc mạng CNN.	15
Hình 9. CNN mang lại mô hình vô cùng tốt để xây dựng hệ thống thông minh.	16
Hình 10. Thứ tự các bộ lọc trong CNNs.	17
Hình 11. Mô hình huấn luyện.	18
Hình 12. Tập mẫu huấn luyện.	18
Hình 13. Đọc dữ liệu train và validation.	19
Hình 14. Xây dựng các lớp huấn luyện.	19
Hình 15. Kết quả huấn luyện.	20
Hình 16. Mô hình dự đoán.	20
Hình 17. Kết quả dự đoán.	21

TỔNG QUAN

1. Tổng quan về môn học.

Trí tuệ nhân tạo hay AI (artificial intelligence), đôi khi được gọi là trí thông minh nhân tạo, là trí thông minh được thể hiện bằng máy móc, trái ngược với trí thông minh tự nhiên của con người. Thông thường, thuật ngữ "trí tuệ nhân tạo" thường được sử dụng để mô tả các máy móc (hoặc máy tính) có khả năng bắt chước các chức năng "nhận thức" mà con người thường phải liên kết với tâm trí, như "học tập" và "giải quyết vấn đề".

Khi máy móc ngày càng tăng khả năng, các nhiệm vụ được coi là cần "trí thông minh" thường bị loại bỏ khỏi định nghĩa về AI, một hiện tượng được gọi là hiệu ứng AI. Một câu châm ngôn trong Định lý của Tesler nói rằng "AI là bất cứ điều gì chưa được thực hiện." Ví dụ, nhận dạng ký tự quang học thường bị loại trừ khỏi những thứ được coi là AI, đã trở thành một công nghệ thông thường. Khả năng máy hiện đại thường được phân loại như AI bao gồm thành công hiểu lời nói của con người, cạnh tranh ở mức cao nhất trong trò chơi chiến lược (chẳng hạn như cờ vua và Go), xe hoạt động độc lập, định tuyến thông minh trong mạng phân phối nội dung, và mô phỏng quân sự.



Hình 1. Trí tuệ nhân tạo

Trí tuệ nhân tạo có thể được phân thành ba loại hệ thống khác nhau: trí tuệ nhân tạo phân tích, lấy cảm hứng từ con người và nhân tạo.[7] AI phân tích chỉ có các đặc điểm phù hợp với trí tuệ nhận thức; tạo ra một đại diện nhận thức về thế giới và sử dụng học tập dựa trên kinh nghiệm trong quá khứ để thông báo các quyết định trong tương lai. AI lấy cảm hứng từ con người có các yếu tố từ trí tuệ nhận thức và cảm xúc; hiểu cảm xúc của con người, ngoài các yếu tố nhận thức và xem xét chúng trong việc ra quyết định. AI nhân cách hóa cho thấy các đặc điểm của tất cả các loại năng lực (nghĩa là trí tuệ nhận thức, cảm xúc và xã hội), có khả năng tự ý thức và tự nhận thức được trong các tương tác.

Lợi ích và mặt trái của công nghệ AI:

Lợi ích:

- Phát hiện và ngăn chặn các rủi ro: AI giúp con người dự báo trước các rủi ro và mối nguy hại tiềm ẩn và hạn chế các thiệt hại đem lại. Các rủi ro được AI nhận biết như: Thảm họa thiên nhiên, động đất, sóng thần, núi lửa, dịch bệnh hay có mối nguy hại trong sản xuất kinh doanh.
- Hạn chế sử dụng sức lao động của con người: Nhờ quá trình học máy và tạo ra được các robot trong công nghiệp và đời sống. Con người sẽ không phải tốn nhiều sức lao động trong sản xuất, vận hành. Giờ đây, các máy móc robot sẽ thay con người làm việc đó.
- Xóa bỏ khoảng cách ngôn ngữ: Công nghệ AI sẽ giúp con người trên mọi Quốc gia có thể nói chuyện và hiểu nhau, thoải mái tiếp xúc. Có thêm nhiều cơ hội để học tập và làm việc trên khắp thế giới.
- Cá nhân hóa: Công nghệ AI sẽ đánh giá và thích ứng cũng như học hỏi đối tượng mà nó phục vụ. Từ đó, đưa ra phản ứng phù hợp nhất cho từng đối tượng riêng biệt.



Hình 2. Robot trí tuệ nhân tạo đầu tiên của Việt Nam.

Mặt trái:

- Hẳn các bạn đã từng nghe nói đến Sophia, robot trí tuệ nhân tạo đầu tiên được cấp quyền công dân trên thế giới, được thiết kế và phát triển bởi 1 công ty công nghệ của Mỹ và được kích hoạt lần đầu tiên năm 2015. Sophia được thiết kế để suy nghĩ và cử động sao cho giống với con người nhất đồng thời được trang bị trí tuệ thông minh nhân tạo. Mục đích chế tạo Sophia - theo nhà sản xuất - là nhằm phát minh ra một robot có ý thức, có sự sáng tạo và có khả năng hoạt động như bất kỳ con người nào để giúp đỡ chính con người trong các vấn đề cuộc sống

thường ngày như: phục vụ chăm sóc sức khỏe, hỗ trợ điều trị y tế, giáo dục cùng nhiều các dịch vụ khác

- Chúng ta hãy tự hỏi rằng, nếu như một ngày kia khi chúng ta đối mặt với một AI đã hoàn thiện hơn Sophia, chúng ta sẽ đứng ở vị thế nào? Một AI như thế có quyền công dân và hoàn toàn bình đẳng như chúng ta, nhưng sẽ thông minh hơn chúng ta nhiều lần, hiểu biết gấp hàng ngàn lần chúng ta, tính toán xử lý nhanh gấp hàng triệu lần chúng ta, không bao giờ có sai sót, không bao giờ biết mệt mỏi, không bao giờ đau ốm hay bệnh tật, không bao giờ đòi hỏi nghỉ ngơi, đòi hỏi người khác quan tâm hay giúp đỡ. Và đặc biệt là AI đó vẫn luôn tự học hỏi và nâng cấp không ngừng.
- Chúng ta sẽ suy nghĩ gì khi gặp một AI chỉ mới ra đời từ ngày hôm qua và không cần phải trải qua nhiều chục năm học hành với rất nhiều nỗ lực như chúng ta, một AI với mức chi phí sản xuất rẻ hơn hàng trăm lần so với chi phí để nuôi dạy ta khôn lớn thành người, một AI mang lại hiệu quả công việc cao hơn bản thân ta gấp hàng chục lần mà không đòi hỏi lương hoặc với mức lương không đáng kể. Chúng ta sẽ suy nghĩ gì khi một AI đối xử với chúng ta đúng như một người bề trên hướng dẫn, chăm sóc, bảo vệ kẻ bề dưới? Khi đó, liệu chúng ta có cảm thấy sự tồn tại của mình là xấu xí, thừa thãi, vô dụng?

Trong những năm gần đây, việc ứng dụng AI cho phép các hệ thống vũ khí tự động lựa chọn và tấn công mục tiêu đã gây nhiều tranh cãi về mặt đạo đức. Những ý kiến chỉ trích cho rằng hệ thống này sẽ đe dọa đến an ninh quốc tế và báo trước một cuộc chiến tranh thế giới lần thứ ba sau súng đạn và bom nguyên tử. Những quy tắc và ranh giới cho AI quan trọng hơn so với bất kỳ lĩnh vực nào khác xuất hiện trước đó. Thế nhưng, với tốc độ phát triển chóng mặt của trí tuệ nhân tạo tại một vài tập đoàn lớn (và hầu như đều nghiên cứu bí mật), các tổ chức chính phủ chịu trách nhiệm về các quy định dường như không thể theo kịp.

2. Tổng quan về đề tài.

a. Giới thiệu về đề tài.

Hiện nay, cái bài toán, mô hình về nhận diện khuôn mặt đã không còn xa lạ trong cuộc sống. Tuy nhiên trong bài viết này, chúng ta sẽ đi vào tìm hiểu về nhận diện khuôn mặt khi người nhận diện quay một góc bất kỳ, có nghĩa là hệ thống sẽ phát hiện và thông báo họ đã quay mặt một góc khoảng bao nhiêu độ. Từ đó, thông báo quay mặt lại theo hướng cần thiết.

b. Tầm quan trọng của đề tài.

Đề tài phát triển sẽ được ứng dụng trong rất nhiều lĩnh vực như:

Sử dụng trong camera chấm công, dùng để cảnh báo khi khuôn mặt đặt sai hướng. từ đó, thông báo để người điểm danh thực hiện quay mặt lại sao cho đúng hướng.

Sử dụng trong camera, quay phim chụp hình và giám sát. Giúp theo dõi theo hướng di chuyển của khuôn mặt.

Phát hiện tài xế ngủ gật trong oto.

c. Mục tiêu đề tài.

Hoàn thành được mô hình đề ra.

Chạy thử nghiệm với những góc đơn giản.

Sử dụng mô hình tự tạo để train.

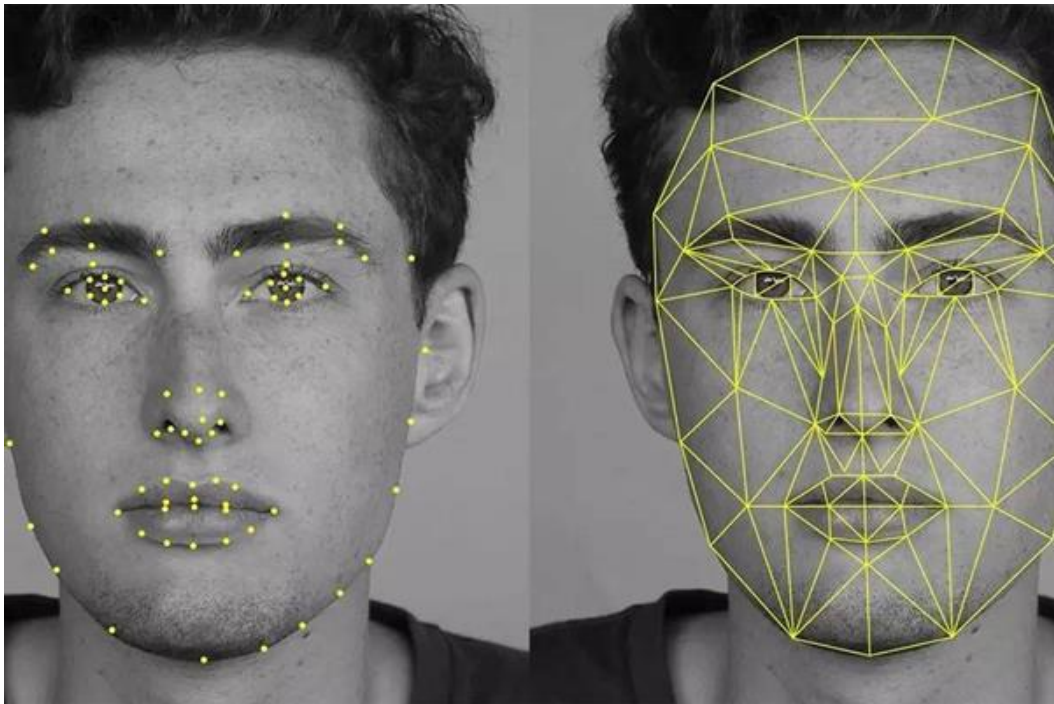
NỘI DUNG

I. GIỚI THIỆU VỀ NHẬN DẠNG KHUÔN MẶT.

1. Nhận dạng khuôn mặt.

Hệ thống nhận dạng khuôn mặt là một ứng dụng máy tính tự động xác định hoặc nhận dạng một người nào đó từ một bức hình ảnh kỹ thuật số hoặc một khung hình video từ một nguồn video. Một trong những cách để thực hiện điều này là so sánh các đặc điểm khuôn mặt chọn trước từ hình ảnh và một cơ sở dữ liệu về khuôn mặt.

Hệ thống này thường được sử dụng trong các hệ thống an ninh và có thể được so sánh với các dạng sinh trắc học khác như các hệ thống nhận dạng vân tay hay tròng mắt.



Hình 3. Nhận dạng khuôn mặt

Cách thức hoạt động:

Quá trình mà sinh trắc học nhận diện khuôn mặt hoạt động bao gồm:

- **Lấy mẫu:** Trước hết, để có thể phân tích khuôn mặt và nhận diện, cần phải tách khuôn mặt ra khỏi khung cảnh còn lại trước đã
- **Phân tích:** Chúng ta nhận diện gương mặt mọi người qua các đặc điểm, hệ thống nhận diện cũng vậy, nhưng nó thực hiện đánh giá các đặc điểm ở 1 level cao hơn
- **So sánh:** Mỗi khuôn mặt đều có nhiều điểm mốc, những phần lồi lõm tạo nên các đặc điểm của khuôn mặt. Các hệ thống nhận diện gương mặt định nghĩa những điểm này là những điểm nút:
- **Kết quả:** Hệ thống sau đó sẽ quyết định kết quả so sánh có phù hợp hay không.

Đặc điểm nổi bật

- Công nghệ nhận diện khuôn mặt có nhiều ưu điểm mà các phương pháp xác thực khác không có

- Một trong những khía cạnh tích cực của nhận diện khuôn mặt là hệ thống không cần sự tương tác của con người dùng,
- Không đòi hỏi người dùng phải chờ đợi trong thời gian dài hoặc làm bất cứ điều gì hơn là nhìn vào camera.
- Nâng cao mức độ an ninh trong hệ thống an ninh qua camera.
- Tạo ra kho dữ liệu với danh sách "Trắng - Tin cậy" và danh sách "Đen - Blacklist"
- Sẵn sàng kết nối với hệ thống kiểm soát ra vào - Access Control.

Ưu điểm và nhược điểm của nhận dạng khuôn mặt

Ưu điểm: Trong số các kỹ thuật sinh trắc học, nhận dạng khuôn mặt có thể không đáng tin cậy và hiệu quả nhất. Tuy nhiên, một trong những lợi thế quan trọng là nó không đòi hỏi sự hợp tác của các đối tượng thử nghiệm. Các hệ thống thiết kế được lắp đặt tại các sân bay, khu chung cư, và những nơi công cộng khác có thể xác định các cá nhân giữa đám đông, mà không bỏ sót một ai. Sinh trắc học khác như dấu vân tay, quét mống mắt, và nhận dạng giọng nói không thể thực hiện được như vậy. Tuy nhiên, câu hỏi đã được đặt ra về hiệu quả của phần mềm nhận dạng khuôn mặt trong trường hợp của an ninh đường sắt và sân bay (theo wikipedia)

Nhược điểm:

- Mặc dù nhận dạng khuôn mặt đã thực hiện được khá tốt ở phía mặt trước và phía chênh lệch 20 độ, nhưng ngay sau khi bạn đi về phía góc khuất, thì nó có vấn đề.
- Một số điều kiện có thể ảnh hưởng tới tính chính xác của phương pháp như:
- Thiếu ánh sáng, đeo kính mát, tóc dài, hoặc một phần khuôn mặt bị che
- Hình ảnh độ phân giải thấp
- Hệ thống sẽ kém hiệu quả nếu biểu hiện khuôn mặt khác nhau, ví dụ như khi cười lớn cũng có ảnh hưởng. Vì vậy nên chụp ảnh hộ chiếu, chứng minh thư, hay khi đi qua cửa hộ chiếu chụp ảnh hộ không cho mình cười là vì thế đấy

2. Nhận dạng góc xoay trong nhận dạng khuôn mặt.

Trong việc điểm danh, chấm công: Nhờ nhận dạng được khuôn mặt người đang không để đúng hướng, khiến cho việc quét khuôn mặt khó khăn, hệ thống sẽ đưa ra cảnh báo và thông báo điều chỉnh khuôn mặt sao cho hợp lý.

Trong máy camera, quay phim chụp hình: Camera sẽ di chuyển theo hướng khuôn mặt người di chuyển, từ đó bắt trọn được mọi khoảnh khắc.

Ngoài ra còn một số ứng dụng khác như: một số hãng xe trang bị để thông báo tài xế ngủ gật; hệ thống bảo mật công nghệ;...



Hình 4. Nhận dạng góc xoay khuôn mặt

3. Các phương pháp tiếp cận.

Để xác định được góc xoay khuôn mặt, ta có một số hướng tiếp cận với vấn đề như sau:

Hướng tiếp cận dựa trên tri thức: Trong hướng tiếp cận này, các luật sẽ phụ thuộc rất lớn vào tri thức của những tác giả nghiên cứu về bài toán xác định khuôn mặt người. Đây là hướng tiếp cận dạng top-down. Dễ dàng xây dựng các luật cơ bản để mô tả các đặc trưng của khuôn mặt và các quan hệ tương ứng. Ví dụ, một khuôn mặt thường có hai mắt đối xứng nhau qua trục thẳng đứng ở giữa khuôn mặt và có một mũi, một miệng. Các quan hệ của các đặc trưng có thể được mô tả như quan hệ về khoảng cách và vị trí. Thông thường các tác giả sẽ trích đặc trưng của khuôn mặt trước tiên để có được các ứng viên, sau đó các ứng viên này sẽ được xác định thông qua các luật để biết ứng viên nào là khuôn mặt và ứng viên nào không phải khuôn mặt. Thường áp dụng quá trình xác định để giảm số lượng xác định sai.

Hướng tiếp cận đặc trưng không thay đổi: Đây là hướng tiếp cận theo kiểu bottom-up. Các tác giả cố gắng tìm các đặc trưng không thay đổi của khuôn mặt người để xác định khuôn mặt người. Dựa trên nhận xét thực tế, con người dễ dàng nhận biết các khuôn mặt và các đối tượng trong các tư thế khác nhau và điều kiện ánh sáng khác nhau, thì phải tồn tại các thuộc tính hay đặc trưng không thay đổi. Có nhiều nghiên cứu đầu tiên xác định các đặc trưng khuôn mặt rồi chỉ ra có khuôn mặt trong ảnh hay không. Các đặc trưng như: lông mày, mắt, mũi, miệng, và đường viền của tóc được trích bằng phương pháp xác định cạnh. Trên cơ sở các đặc trưng này, xây dựng một mô hình thống kê để mô tả quan hệ của các đặc trưng này và xác định sự tồn tại của khuôn mặt trong ảnh. Một vấn đề của các thuật toán theo hướng tiếp cận đặc trưng cần phải điều chỉnh cho phù hợp điều kiện ánh sáng, nhiễu, và bị che khuất. Đôi khi bóng của khuôn mặt sẽ tạo thêm cạnh mới, mà cạnh này lại rõ hơn cạnh thật sự của khuôn mặt, vì thế nếu dùng cạnh để xác định sẽ gặp khó khăn.

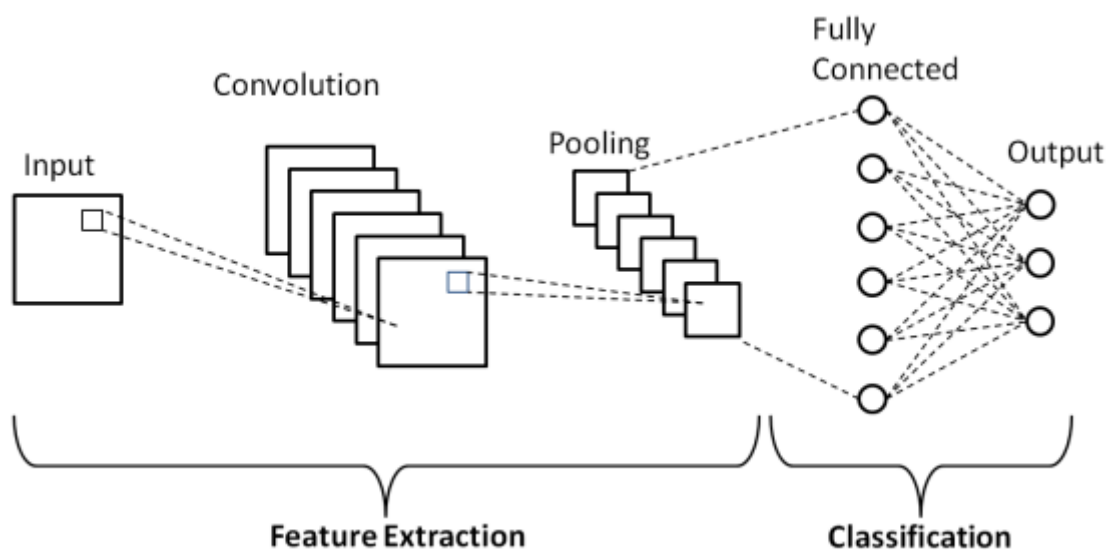
Hướng tiếp cận so khớp mẫu: Trong so khớp mẫu, các mẫu chuẩn của khuôn mặt (thường là khuôn mặt được chụp thẳng) sẽ được xác định trước hoặc xác định các tham số thông qua một hàm. Từ một ảnh đưa vào, tính các giá trị tương quan so với các mẫu chuẩn về đường viền khuôn mặt, mắt, mũi và miệng. Thông qua các giá trị tương quan này mà các tác giả quyết định có hay không có tồn tại khuôn mặt trong ảnh. Hướng tiếp cận này có lợi thế là rất dễ cài đặt, nhưng không hiệu quả khi tỷ lệ, tư thế, và hình dáng thay đổi (đã được chứng minh). Nhiều độ phân giải, đa tỷ lệ, các mẫu con, và các mẫu biến dạng được xem xét thành bất biến về tỷ lệ và hình dáng.

Hướng tiếp cận dựa trên diện mạo: Trái ngược với các phương pháp so khớp mẫu với các mẫu đã được định nghĩa trước bởi những chuyên gia, các mẫu trong hướng tiếp cận này được học từ các ảnh mẫu. Một cách tổng quát, các phương pháp theo hướng tiếp cận này áp dụng các kỹ thuật theo hướng xác suất thống kê và máy học để tìm những đặc tính liên quan của khuôn mặt và không phải là khuôn mặt. Các đặc tính đã được học ở trong hình thái các mô hình phân bố hay các hàm biệt số nên dùng có thể dùng các đặc tính này để xác định khuôn mặt người. Đồng thời, bài toán giảm số chiều thường được quan tâm để tăng hiệu quả tính toán cũng như hiệu quả xác định.

II. GIỚI THIỆU VỀ CONVOLUTIONAL NEURAL NETWORK.

1. Convolutional Neural Network là gì?

Convolutional Neural Network (CNN hoặc ConvNet) được tạm dịch là: Mạng nơ-ron tích tụ. Đây được xem là một trong những mô hình của Deep Learning – tập hợp các thuật toán để có mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý cấu trúc phức tạp. Hiểu đơn giản, CNN là một lớp của mạng nơ-ron sâu, được áp dụng phổ biến nhất để phân tích hình ảnh trực quan.



Hình 5. Convolutional Neural Network.

Hiện tại, chúng ta chưa có định nghĩa một cách chính xác nhất về thuật toán CNN. Mạng CNN được thiết kế với mục đích xử lý dữ liệu thông qua nhiều lớp mạng. Ngoài

ra, CNN có thể giúp bạn tạo ra được hệ thống thông minh, phản ứng với độ chính xác khá cao.

Ví dụ: Bạn có thể sử dụng loại mạng thần kinh này trong các ứng dụng như: nhận dạng hình ảnh, nhận dạng khuôn mặt và hiện tại đã được đưa vào các nền tảng Facebook, Google,...

So với những mạng nơ ron thông thường, mạng CNN nhận đầu vào là một mảng hai chiều và hoạt động trực tiếp trên hình ảnh thay vì tập trung trích xuất tính năng mà bạn thường thấy ở các mạng nơ ron khác.

2. Feature là gì?

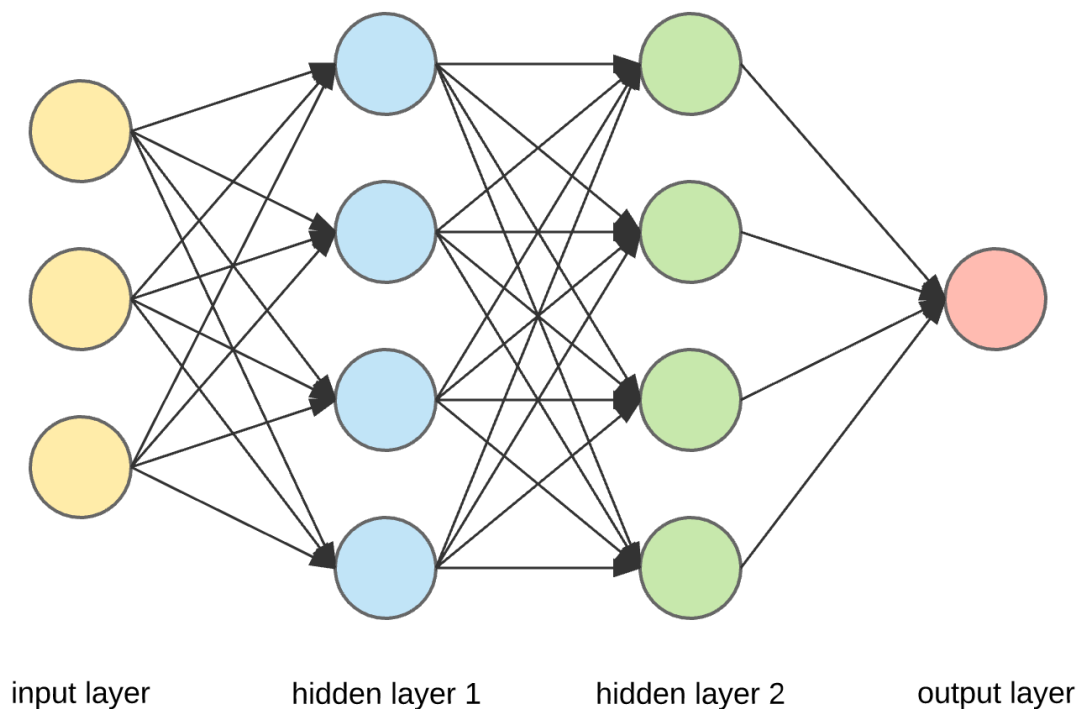
Feature được dịch theo nghĩa tiếng Việt là đặc điểm. Khi sử dụng thuật toán CNN so sánh hình ảnh theo từng mảnh, mỗi mảnh đó được gọi là Feature.

Mỗi Feature được xem như một hình ảnh mini hay gọi là những mảng hai chiều nhỏ. Các Feature được khớp với những khía cạnh chung trong bức ảnh đó. Nghĩa là Feature sẽ tương ứng với khía cạnh nào đó của hình ảnh và chúng sẽ khớp lại với nhau.

3. Convolutional là gì?

Convolutional được hiểu là tích chập. Xét về cơ bản, khi xem một hình ảnh mới, thuật toán CNN sẽ không nhận biết được nó ở vị trí nào, các Feature sẽ khớp với nhau ở đâu?

Chính vì vậy, Convolutional sẽ thử chúng với tất cả các vị trí khác nhau và tạo thành một bộ lọc gọi là Filter. Quá trình này được thực hiện thông qua phần toán nơ ron tích chập.



Hình 6. Mạng lưới Convolutional.

4. Một số lớp cơ bản của Convolutional Neural Network.

#1. Convolutional Layer

Trong ba lớp của Convolutional Neural Network, Convolutional Layer được xem là lớp có vai trò quan trọng nhất. Bởi vì Convolutional Layer sẽ đại diện CNN thực hiện mọi phép toán.

Khi nhắc đến lớp Convolutional Layer, chúng ta cần làm rõ một số khái niệm đó là: Filter Map, Stride, Padding, Feature Map.

- Filter Map

Nếu như ANN kết nối với từng Pixel của hình ảnh đầu vào thì CNN được sử dụng những Filter để áp vào các vùng của hình ảnh. Những Filter Map này có thể xem là một ma trận 3 chiều, bao gồm những con số và các con số chính là Parameter.

- Stride

Trong Convolutional Neural Network, Stride được hiểu là khi chúng ta dịch chuyển Filter Map theo Pixel và dựa vào giá trị từ trái sang phải. Stride đơn giản là biểu thị sự dịch chuyển này.

- Padding

Padding chính là những giá trị 0 được thêm vào lớp Input.

- Feature Map

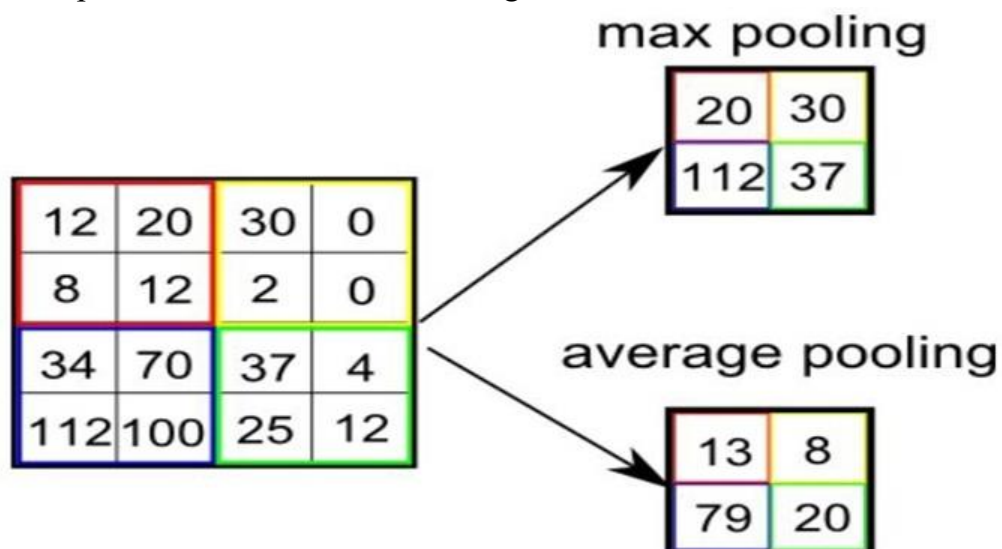
Đây là kết quả hiển thị sau mỗi lần Filter Map quét qua Input. Cứ mỗi lần quét như vậy, bạn sẽ thấy sự xuất hiện của quá trình tính toán được xảy ra.

#2. Pooling Layer

Khi đầu vào quá lớn, các lớp Pooling Layer sẽ được dịch chuyển vào giữa những lớp Convolutional Layer nhằm giảm các Parameter.

Pooling Layer được biết đến với hai loại phổ biến là: Max Pooling và Average Pooling.

Tại Pooling Layer, khi bạn sử dụng lớp Max Pooling thì số lượng Parameter có thể sẽ giảm đi. Vì vậy, Convolutional Neural Network sẽ xuất hiện nhiều lớp Filter Map, mỗi Filter Map đó sẽ cho ra một Max Pooling khác nhau.



Hình 7. Lớp Pooling.

#3. Relu Layer

Đây chính là một hàm kích hoạt trong Neural Network. Chúng ta có thể biết đến hàm kích hoạt này với một tên gọi khác là Activation Function. Nhiệm vụ chính của hàm kích hoạt là mô phỏng lại các Neuron có tỷ lệ truyền xung qua Axon. Trong đó, hàm kích hoạt sẽ bao gồm các hàm cơ bản như: Sigmoid, Tanh, Relu, Leaky Relu, Maxout.

Hiện nay, hàm Relu đang được sử dụng khá phổ biến và thông dụng. Đặc biệt, Relu sở hữu những ưu điểm nổi bật như: hỗ trợ tính toán nhanh nên rất được ưa chuộng sử dụng trong việc huấn luyện các mạng Neuron.

Khi sử dụng Relu, bạn cần lưu ý đến việc tùy chỉnh các Learning Rate và theo dõi Dead Unit. Lớp Relu Layer này được sử dụng sau khi Filter Map được tính toán ra và áp dụng hàm Relu lên tất cả các giá trị trên Filter Map.

#4. Fully Connected Layer

Fully Connected Layer thường sử dụng để đưa ra các kết quả.

Ví dụ: Sau khi các lớp Convolutional Layer và Pooling Layer nhận được các ảnh đã truyền qua chúng, bạn sẽ thu được kết quả là Model đã đọc được khá nhiều thông tin về ảnh. Do đó, để có thể liên kết các đặc điểm này lại và cho ra Output, bạn cần dùng đến Fully Connected Layer.

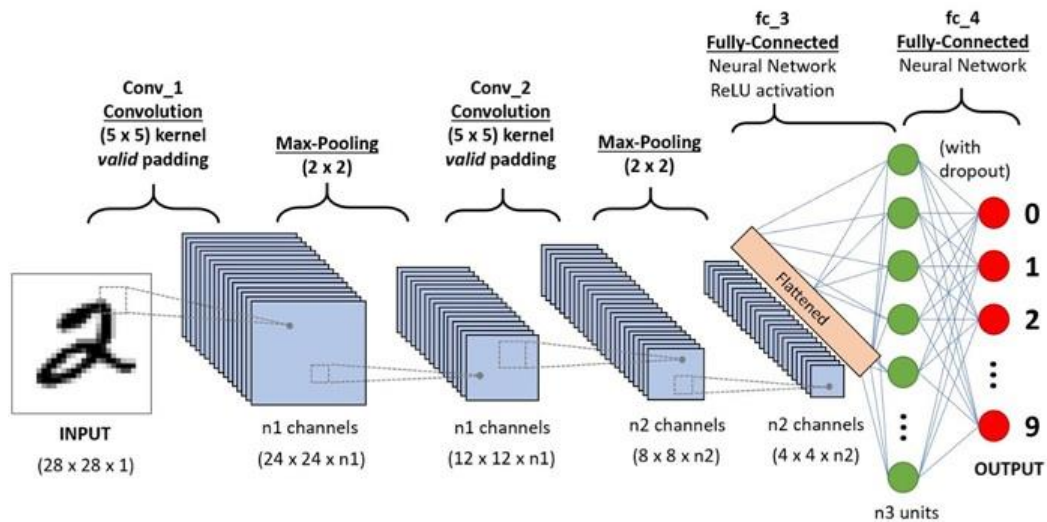
Bên cạnh đó, khi có được các dữ liệu về hình ảnh, Fully Connected Layer sẽ chuyển đổi chúng thành những mục có phân chia chất lượng. Tương tự như kiểu chia chúng thành các phiếu bầu và đánh giá để chọn ra hình ảnh đạt chất lượng tốt nhất. Dù vậy, quá trình này không được coi là quá trình dân chủ nên rất ít sử dụng.

5. Cấu trúc của Convolutional Neural Network

Mạng Convolutional Neural Network là tập hợp nhiều lớp Convolutional chồng lên nhau, sử dụng các hàm Nonlinear Activation và tanh để kích hoạt các trọng số trong các node. Ở mỗi lớp CNN, sau khi được các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho những lớp tiếp theo. Mỗi Layer kết tiếp sẽ là kết quả Convolution từ Layer trước đó nên chúng ta có được các kết nối cục bộ.

Thông qua quá trình huấn luyện mạng, các lớp Layer CNN tự động học các giá trị được thể hiện qua các lớp Filter.

Ví dụ: Trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm kiếm những thông số tối ưu cho các Filter tương ứng theo một thứ tự: Raw Pixel => Edges => Shapes => Facial => High – level Features. Layer cuối cùng được dùng để phân lớp ảnh.



Hình 8. Cấu trúc mạng CNN.

Những phép dịch chuyển, quay hoặc co giãn sẽ được sử dụng Pooling Layer làm bất biến các tính chất kia. Do vậy, CNN đưa ra kết quả có độ chính xác khá cao trong các mô hình.

CNN có cấu trúc cơ bản gồm ba phần chính là: Local Receptive Field, Shared Weights And Bias và Pooling.

- **Local Receptive Field**

Local Receptive Field, tạm dịch: trường tiếp nhận cục bộ. Đây được xem là lớp giúp bạn có thể tách lọc các dữ liệu, thông tin của ảnh và chọn được những vùng ảnh có giá trị sử dụng nhất.

- **Shared Weights And Bias**

Shared Weights, tạm dịch: trọng số chia sẻ. Chức năng chính của lớp này là hỗ trợ bạn làm giảm tối đa số lượng tham số trong mạng CNN. Vì trong mỗi Convolution sẽ bao gồm các Feature Map khác nhau, mỗi Feature Map lại giúp Detect một vài Feature trong ảnh.

- **Pooling Layer**

Pooling Layer, tạm dịch: lớp tổng hợp. Đây được xem gần như là lớp cuối cùng trước khi đưa ra kết quả trong CNN. Chính vì thế, để có được kết quả dễ hiểu và dễ sử dụng nhất thì Pooling Layer có nhiệm vụ làm đơn giản hóa các thông tin đầu ra. Nghĩa là, sau khi hoàn thành quá trình tính toán và quét các lớp thì sẽ đi đến Pooling Layer nhằm lược bớt các thông tin không cần thiết và cho ra kết quả mà chúng ta đang cần.

6. Tham số cho Convolutional Neural Network

Convolutional Neural Network là thuật toán mang lại cho bạn chất lượng mô hình vô cùng tốt để xây dựng hệ thống một cách thông minh nhất. Để chọn được các tham số cho Convolutional Neural Network chuẩn nhất, bạn cần lưu ý đến các số lượng ở 4 yếu tố sau: số Convolution Layer, Filter Size, Pooling Size và cách Train Test.



Hình 9. CNN mang lại mô hình vô cùng tốt để xây dựng hệ thống thông minh.

- **Số Convolution Layer**

Lớp số Convolution Layer càng nhiều thì chương trình của bạn chạy càng được cải thiện. Khi sử dụng các Layer với một số lượng lớn, các tác động có thể được giảm một cách đáng kể. Có thể sau 3 đến 4 Layer thì bạn đã có được kết quả mong muốn.

- **Filter Size**

Phần lớn các Filter Size thường có kích thước là 3×3 hoặc 5×5

- **Pooling Size**

Đối với các loại ảnh thông thường sẽ có kích thước là 2×2 , tuy nhiên nếu đầu vào hình ảnh lớn thì có thể sử dụng Pooling Size 4×4 để đảm bảo chất lượng cho ảnh.

- **Train Test**

Thực hiện Train Test nhiều lần để so sánh các kết quả với nhau. Điều này giúp có được các Parameter tốt nhất.

III. XÂY DỰNG KIẾN TRÚC MẠNG CNN NHẬN DẠNG GÓC XOAY KHUÔN MẶT.

1. Xây dựng mạng CNNs.

Xây dựng tầng Convolutional

Xây dựng tầng Pooling: Sử dụng 64 bộ lọc filter

Xây dựng hàm biến đổi tuyến tính: Về cơ bản, convolution là một phép biến đổi tuyến tính. Nếu tất cả các neuron được tổng hợp bởi các phép biến đổi tuyến tính thì một mạng neuron đều có thể đưa về dưới dạng một hàm tuyến tính.

Kết nối Fully-connected: Cách kết nối các neural ở hai tầng với nhau trong đó tầng sau kết nối đầy đủ với các neural ở tầng trước nó. Đây cũng là dạng kết nối thường thấy ở ANN, trong CNN tầng này thường được sử dụng ở các tầng phí cuối của kiến trúc mạng.

Cụ thể được mô tả như sau: $[[\text{CONV} \rightarrow \text{RELU}]^*N \rightarrow \text{POOL?}]^*M \rightarrow [\text{FC} \rightarrow \text{RELU}]^*K \rightarrow \text{FC}$

Trong đó:

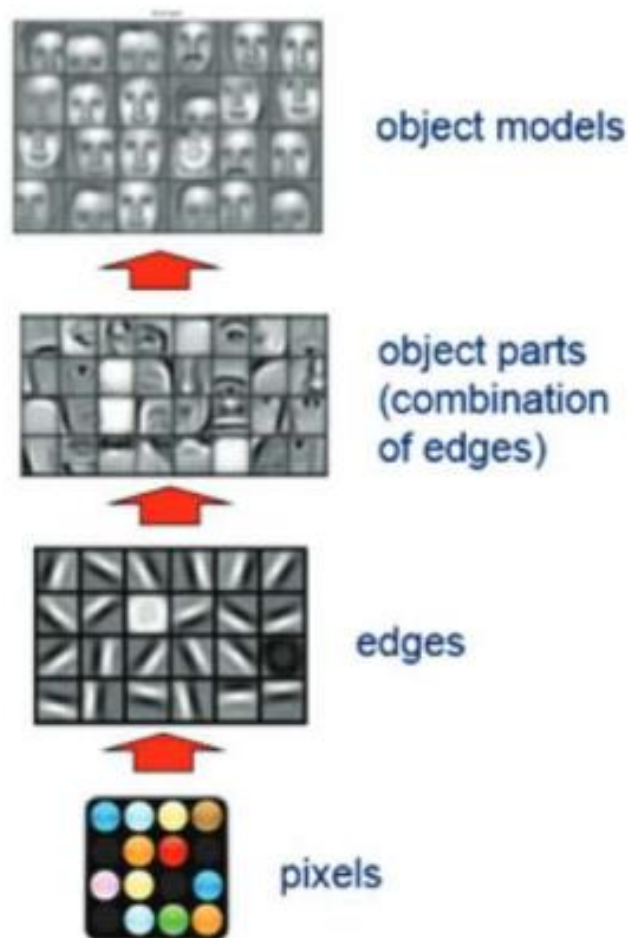
$[\text{CONV} \rightarrow \text{RELU}]^*N$ tức là trong kiến trúc này sau tầng CONV là tầng RELU, trong CNN kiến trúc 2 tầng này có thể lặp N lần

POOL là tầng Pooling cho người thiết kế quyết định có thể có hoặc không.

$[[\text{CONV} \rightarrow \text{RELU}]^*N \rightarrow \text{POOL?}]^*M$ trong kiến trúc CNN có thể lặp lại M lần kiểu sau tầng CONV và RELU và kế tới là tầng Pooling.

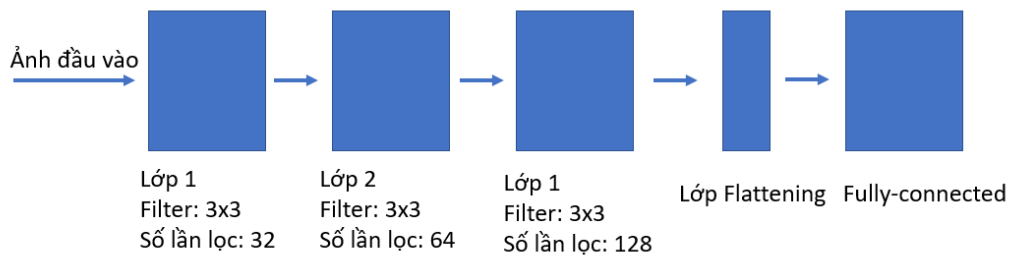
$[\text{FC} \rightarrow \text{RELU}]^*K$ trong CNN có thể lặp K lần cấu trúc kiểu sau tầng FC là tầng RELU nhưng trước nó phải có tầng $[\text{CONV} \rightarrow \text{RELU}]$

Như vậy CNNs sẽ cố gắng tìm ra các thông số tối ưu cho các filter tương ứng theo thứ tự pixel > edges > shapes > facial > high-level features.



Hình 10. Thứ tự các bộ lọc trong CNNs.

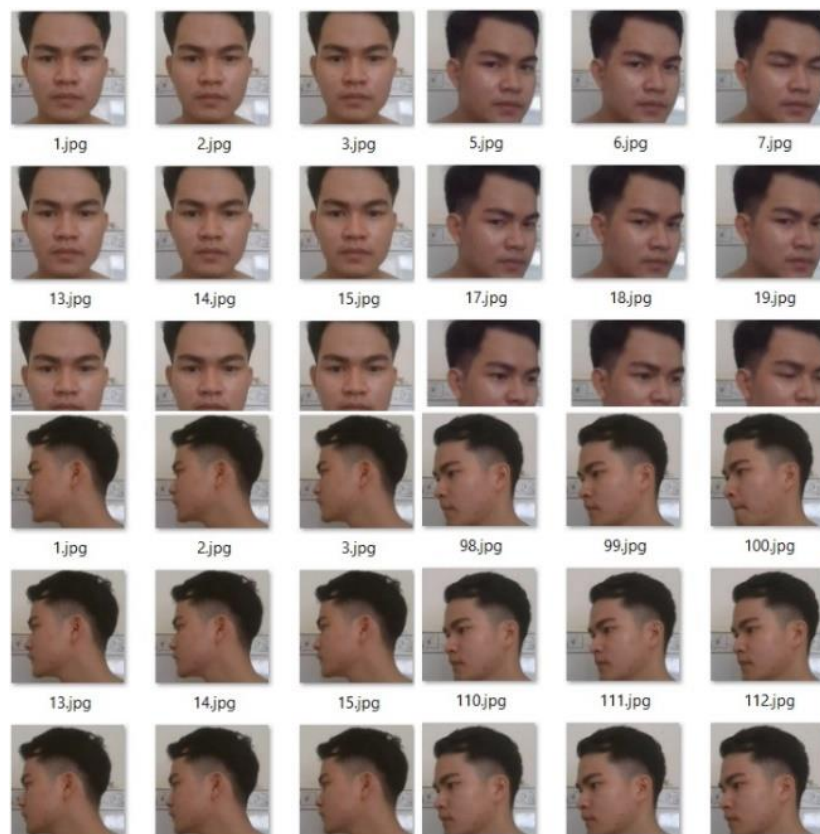
2. Huấn luyện mô hình.



Hình 11. Mô hình huấn luyện.

Bài toán chỉ áp dụng trong mô hình nhỏ chỉ nhận 9 góc độ khác nhau của khuôn mặt. Sự dụng tập ảnh các góc khuôn mặt do chính em tự tạo. Mô hình dùng 80% ảnh để huấn luyện và 20% ảnh để kiểm chứng.

Bước 1: Xây dựng tập dữ liệu là ảnh góc nghiêng các khuôn mặt, phân loại các thư mục và lưu vào thư mục Train.



Hình 12. Tập mẫu huấn luyện.

Bước 2: Thay đổi kích thước huấn luyện thành cỡ 200x200.

Bước 3: Đọc dữ liệu train và validation.

```

train_generator = train_data_gen.flow_from_directory(
    train_image_files_path,
    target_size=(200, 200),
    class_mode='categorical')

validation_generator = validation_data_gen.flow_from_directory(
    valid_image_files_path,
    target_size=(200, 200),
    class_mode='categorical')

```

Found 1555 images belonging to 9 classes.
 Found 330 images belonging to 9 classes.

Hình 13. Đọc dữ liệu train và validation.

Bước 4: Xây dựng mô hình:

```

model=tf.keras.models.Sequential()
# lớp CNN1
model.add(Conv2D(32,(3,3), activation='relu',input_shape=(200,200,3)))
model.add(MaxPooling2D(2,2))
# lớp CNN2
model.add(Conv2D(64,(3,3), activation='relu'))
model.add(MaxPooling2D(2,2))
# lớp CNN3
model.add(Conv2D(128,(3,3), activation='relu'))
model.add(MaxPooling2D(2,2))
# chuyển dữ liệu đầu ra của mạng CNN từ mảng 2 chiều về mảng 1 chiều
model.add(Flatten())
# lớp ẩn
model.add(Dense(128, activation=tf.nn.relu))
# lớp output
model.add(Dropout(0.2))
model.add(Dense(9, activation=tf.nn.softmax))
model.summary()

```

Hình 14. Xây dựng các lớp huấn luyện.

Bước 5: Thiết lập các tham số và thực hiện huấn luyện.

```

from tensorflow.keras.optimizers import Adam
model.compile(optimizer=Adam(learning_rate=0.001),
              loss='categorical_crossentropy',
              metrics=['acc'])

EPOCHS=50
history=model.fit(
    train_generator,
    steps_per_epoch=1,
    epochs=EPOCHS,
    verbose=1,
    validation_data = validation_generator,
    validation_steps=2)

```

```

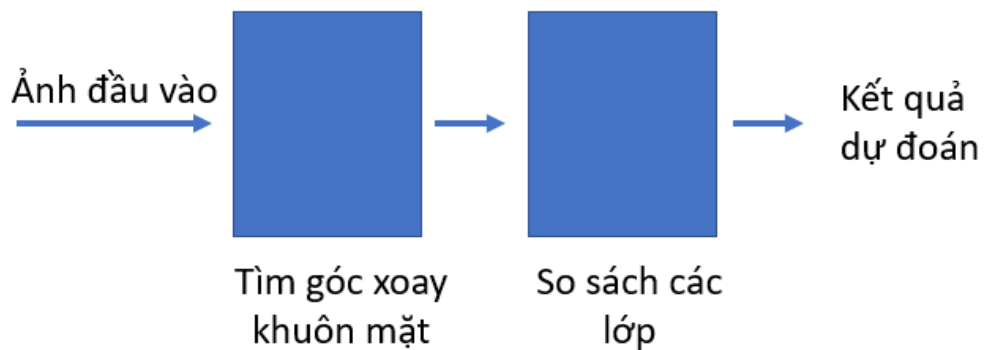
- 1s 664ms/step - loss: 0.8077 - acc: 0.8438 - val_loss: 0.5969 - val_acc: 0.8906
- 1s 629ms/step - loss: 0.7989 - acc: 0.7188 - val_loss: 0.5474 - val_acc: 0.9219
- 1s 627ms/step - loss: 0.5861 - acc: 0.8750 - val_loss: 0.3302 - val_acc: 0.9688
- 1s 620ms/step - loss: 0.6028 - acc: 0.7500 - val_loss: 0.4189 - val_acc: 0.8906

```

Hình 15. Kết quả huấn luyện.

Bước 6: Lưu kết quả huấn luyện.

3. Xây dựng mô hình dự đoán:



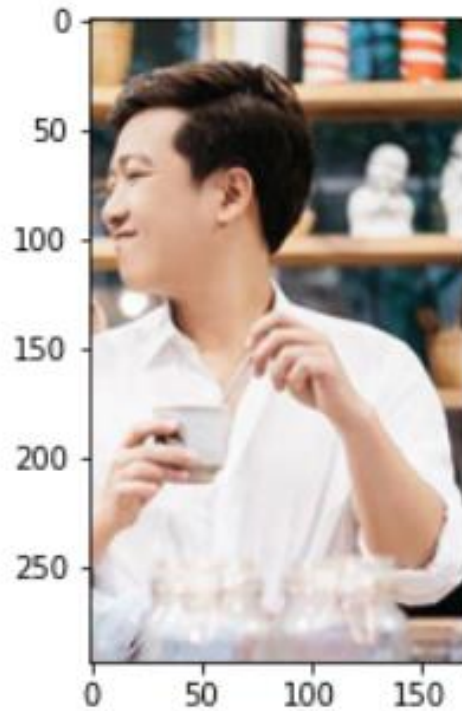
Hình 16. Mô hình dự đoán.

Bước 1: Load file huấn luyện.

Bước 2: Đọc ảnh và xử lý.

Bước 3: So sánh và đưa ra kết quả dự đoán

Góc xoay khuôn mặt dự đoán được là: 90



Hình 17. Kết quả dự đoán.

IV. HƯỚNG PHÁT TRIỂN.

Sự dụng real-time cho thuật toán.

Tỉ lệ nhận dạng còn phụ thuộc vào điều kiện, môi trường sáng xung quanh.

Chưa nhận diện được toàn bộ hướng khuôn mặt.

Sử dụng kết hợp học tăng cường cùng thuật toán để tăng độ chính xác cho mô hình.

KẾT LUẬN.

Hoàn thành cơ bản các yêu cầu đề ra.

Mô hình tuy đã hoàn thành nhưng vẫn còn nhiều thiếu sót.

Chưa áp dụng được realtime vào bài toán và độ nhận diện chính xác chưa cao mặc dù training ra chỉ số xấp xỉ 1.

Thông qua môn học, em đã được tiếp cận với các công nghệ hiện đại. Các thuật toán về AI. Hy vọng trong tương lai sẽ đem áp dụng vào công việc, phục vụ cho sự phát triển khoa học công nghệ trong tương lai.

TÀI LIỆU THAM KHẢO

1. [“Facial Recognition Applications”](#). Animetrics.
2. [Face Recognition Technology Overview](#)
3. Bonsor, K. [“How Facial Recognition Systems Work”](#).
4. Smith, Kelly. [“Face Recognition”](#) (PDF).
5. R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", IEEE Trans. on PAMI, 1993, (15)10:1042-1052
6. <https://wiki.tino.org/convolutional-neural-network-la-gi/>
7. "Sự phát triển toàn diện của trí tuệ nhân tạo có thể hủy diệt nhân loại", Stephen Hawking -