

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG THƯƠNG TP.HCM
KHOA CÔNG NGHỆ THÔNG TIN
-----o0o-----



ĐỒ ÁN MÔN TRÍ TUỆ NHÂN TẠO
ĐỀ TÀI: DỰ ĐOÁN ĐỘ TUỔI NGƯỜI DỰA TRÊN ẢNH
KHUÔN MẶT SỬ DỤNG THUẬT TOÁN CNN

NHÓM: 11
LỚP: 12DHTH_TD

TP.Hồ Chí Minh tháng 12 năm 2023

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THỰC PHẨM TP.HCM
KHOA CÔNG NGHỆ THÔNG TIN
-----o0o-----



ĐỒ ÁN MÔN TRÍ TUỆ NHÂN TẠO
ĐỀ TÀI: DỰ ĐOÁN ĐỘ TUỔI NGƯỜI DỰA TRÊN ẢNH
KHUÔN MẶT SỬ DỤNG THUẬT TOÁN CNN

GVHD: Nguyễn Thanh Bình

THỰC HIỆN:

Nguyễn Trọng Đăng Khoa	2001215879
Lê Trí Cường	2033210183
Trần Minh Hòa	2001215784
Nguyễn Thị Mai	2001215951

TP.Hồ Chí Minh tháng 12 năm 2023

MỤC LỤC

MỤC LỤC	3
DANH MỤC CÁC HÌNH ẢNH.....	V
BẢNG TÓM TẮT	1
PHÂN CÔNG CÔNG VIỆC	2
LỜI CẢM ƠN.....	3
LỜI MỞ ĐẦU	4
LÝ DO CHỌN ĐỀ TÀI.....	5
CHƯƠNG 1: GIỚI THIỆU	6
1.1 Giới thiệu đề tài.....	6
1.2 Nội dung nghiên cứu.....	6
1.3 Giới hạn đề tài.....	7
1.4 Bố cục báo cáo	7
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT.....	8
2.1 Các khái niệm cơ bản.....	8
2.1.1 Thuật toán Convolutional Neural Networks (CNN)	8
2.1.2 Một số ứng dụng của thuật toán CNN:	10
2.1.3 Mô tả về ba lớp của mô hình CNN	10
2.2 Các bài toán về CNN	14
2.3 Cascade Classifier	15
2.4 CategoricalCrossentropy và Adam	16
2.5 Các nghiên cứu liên quan.....	16
2.5.1 Các bài báo về CNN cần biết	16

CHƯƠNG 3: CÁC PHƯƠNG PHÁP ĐỀ XUẤT	19
3.1 Yêu cầu bài toán.....	19
3.1.1 Phương pháp đề xuất thực thi.....	19
3.2 Hiện thực	20
CHƯƠNG 4: KẾT QUẢ HIỆN THỰC.....	23
4.1 Yêu cầu phần cứng, phần mềm.....	23
4.1.1 Yêu cầu phần cứng.....	23
4.1.2 Yêu cầu phần mềm.....	23
4.2 Kết quả hiện thực	24
4.2.1 Huấn luyện	24
4.2.2 Ưu điểm.....	26
4.2.3 Nhược điểm.....	26
CHƯƠNG 5: KẾT LUẬN.....	27
5.1 Kết quả đạt được	27
5.2 Phương pháp đề xuất.....	27
5.3 Bước mở rộng trong tương lai	27
TÀI LIỆU THAM KHẢO	30

DANH MỤC CÁC HÌNH ẢNH

Hình 1. Sơ đồ thuật toán CNN.[1].....	9
Hình 2. Lớp tích chập convolution.[2]	12
Hình 3. Lớp Pooling.[2].....	13
Hình 4. Lớp Fully Connected Layer.[1]	13
Hình 5. Biểu đồ phân bố lượng hình ảnh trên mỗi độ tuổi thu thập được.[4]	19
Hình 6. Mô hình cấu trúc CNN. [3].....	20
Hình 7: Dự đoán độ tuổi ở hệ điều hành Android	22
Hình 8. Biểu đồ biểu diễn độ chính xác của mô hình trên tập dữ liệu huấn luyện và tập dữ liệu đánh giá thay đổi qua từng epoch.....	25
Hình 9. Biểu đồ biểu diễn sự hội tụ của mô hình trên tập dữ liệu huấn luyện và tập dữ liệu đánh giá thay đổi qua từng epoch.....	25

BẢNG TÓM TẮT

Từ viết tắt	Nghĩa
CNNs	Convolutional Neural Networks
RAM	Random Access Memory
CPU	Central Processing Unit
GPU	Graphics Processing Unit

PHÂN CÔNG CÔNG VIỆC

MSSV	Họ và Tên	Lớp	Công việc
2033210183	Lê Trí Cường	12DHTH_TD	Word, Powerpoint, Tích Chập, Lập Trình Android (Kết nối API)
2001215951	Nguyễn Thị Mai	12DHTH_TD	Powerpoint, Padding, Pooling, Tạo Server Python (Vận Dụng Model, Tạo API)
2001215784	Trần Minh Hòa	12DHTH_TD	Powerpoint, Mạng Neural, Lập Trình Android(Camera)
2001215879	Nguyễn Trọng Đăng Khoa	12DHTH_TD	Demo, Mạng Neural, Xây Dựng/ Quản Lý Huấn Luyện Model

LỜI CẢM ƠN

Lời đầu tiên, nhóm thực hiện đề tài đồ án cuối kỳ xin gửi lời cảm ơn chân thành đến Giảng viên giảng dạy Thầy Nguyễn Thanh Bình. Trong suốt quá trình nhóm thực hiện đề tài, Thầy đã chỉ dạy các kiến thức để nhóm em thực hiện đề tài.

Bên cạnh đó, em xin chân thành cảm ơn thầy Ngô Dương Hà đã truyền tải các kiến thức sâu hơn về CNNs và Ứng dụng nhận diện để hoàn thành đề tài nay.

Mặc dù nhóm đã cố gắng hoàn thành các công việc, nhiệm vụ mà đề tài đặt ra, tuy nhiên vì chưa có nhiều kinh nghiệm và kiến thức của nhóm còn nhiều hạn chế nên sẽ không thể tránh khỏi những thiếu sót trong quá trình thực hiện, trình bày đề tài. Nhóm mong Thầy, Cô thông cảm và rất mong nhận được những ý kiến đóng góp từ Thầy, Cô.

Nhóm xin chân thành cảm ơn!

LỜI MỞ ĐẦU

Ngày nay, thế giới của chúng ta đang chuyển đổi với tốc độ kỷ lục, và công nghệ thông tin chính là nguyên nhân động lực đằng sau sự thay đổi này. Công nghệ đã xâm nhập vào mọi khía cạnh của cuộc sống của chúng ta, từ cách chúng ta làm việc, giải trí, truyền thông, đến cách chúng ta tương tác với nhau và với thế giới xung quanh. Cách mà chúng ta thu thập, trao đổi và sử dụng thông tin đã trở nên không thể tách rời với cuộc sống hàng ngày.

Cùng với sự thâm nhập của công nghệ thông tin vào đời sống, việc xác định và đánh giá độ tuổi của con người từ hình ảnh khuôn mặt đã trở thành một thách thức quan trọng và thú vị. Độ tuổi không chỉ là một khía cạnh cá nhân quan trọng mà còn đóng góp vào nhiều khía cạnh khác của cuộc sống và xã hội. Thế nên hiện nay, việc áp dụng công nghệ để giải quyết bài toán nhận diện độ tuổi trở nên cần thiết hơn bao giờ hết.

LÝ DO CHỌN ĐỀ TÀI

Trước khi ngành công nghệ thông tin phát triển vượt bậc và trở thành một phần không thể thiếu trong cuộc sống, việc xác định độ tuổi của một cá nhân thực sự là một nhiệm vụ khó khăn và nhúc nhối. Thông tin về độ tuổi thường được đánh giá dựa trên quan sát và kinh nghiệm của người nhìn hoặc thông qua việc kiểm tra giấy tờ tùy thân của người đó. Tuy nhiên, với sự xuất hiện của Convolutional Neural Networks (CNN), mọi thứ đã trở nên khác biệt.

Đã đến lúc chúng ta không còn phải dựa vào đánh giá chủ quan của con người hoặc giấy tờ tùy thân để biết độ tuổi của một người. Thay vào đó, CNN cho phép chúng ta tận dụng các yếu tố quan trọng như hình dạng khuôn mặt, đường chảy, cấu trúc xương, màu da, và nhiều yếu tố khác để đưa ra kết quả xác định độ tuổi một cách chính xác và khoa học.

Từ những yếu tố trên nhóm em quyết định chọn thuật toán CNN để xây dựng lên một hệ thống nhận diện độ tuổi và đặt hệ thống cho nhiều lĩnh vực tiềm năng sau này như: an ninh, y tế, chống giả mạo,... Mà còn còn giúp cho con người tiết kiệm thời gian và tài nguyên cũng như công sức.

CHƯƠNG 1: GIỚI THIỆU

1.1 Giới thiệu đề tài

Đề tài "Nhận diện độ tuổi bằng Convolutional Neural Networks (CNN)" là một lĩnh vực nghiên cứu quan trọng trong lĩnh vực trí tuệ nhân tạo và công nghệ hình ảnh.

Trong thuật toán CNN, mạng nơ-ron đóng một vai trò quan trọng trong việc xác định độ tuổi của con người thông qua xử lý hình ảnh khuôn mặt. Đầu tiên, chúng tự động trích xuất các đặc trưng quan trọng từ hình ảnh, bao gồm hình dạng khuôn mặt, đường chảy, cấu trúc xương, và màu da. Sau đó, mạng nơ-ron này học cách phân loại các đặc điểm này và xác định mối liên quan giữa những đặc điểm này và độ tuổi của người dùng.

Từ đó xác định độ tuổi của con người qua hình ảnh khuôn mặt giúp cho tốc độ xử lý nhanh chóng và độ chính xác cao hơn so với việc xác định bằng phương pháp thủ công như: quan sát và kiểm tra giấy tờ tùy thân.

1.2 Nội dung nghiên cứu

- Tìm hiểu cơ sở lý thuyết liên quan đến mạng CNN.
- Tìm hiểu về các thuật toán liên quan đến CNN nhằm hỗ trợ quá trình nhận diện.
- Tạo một tập dữ liệu lấy từ các trang của Google. Sắp xếp, phân chia tập dữ liệu để số lượng ảnh trong mỗi độ tuổi được cân bằng.
- Điều chỉnh các thông số và số lớp trong kiến trúc mạng. Áp dụng phương pháp tối ưu hóa để nâng cao độ chính xác cho cả phân loại độ tuổi, giới tính.
- Thực thi mô hình, dựa trên kết quả về độ chính xác khi nhận dạng để điều chỉnh thông số, đưa ra nhận xét và đánh giá.
- Xử lý các dữ liệu hình ảnh và video từ đầu vào và xuất thông tin từ đầu ra gồm độ tuổi và giới tính.

1.3 Giới hạn đề tài

Đề tài chỉ tập trung vào mạng CNN cơ bản không đi sâu hơn về CNN, các tập dữ liệu, lập trình ứng dụng trên mọi nền tảng. Không tập trung vào cách chuẩn hóa dữ liệu, không tập trung vào các khía cạnh về chức năng nghiệp vụ của lập trình ứng dụng.

1.4 Bố cục báo cáo

- ❖ Lời cảm ơn: Nhóm làm đồ án chúng em vô cùng cảm ơn thầy Nguyễn Thanh Bình và thầy Ngô Dương Hà đã truyền tải kiến thức giúp chúng em có thể hoàn thành đồ án môn học này ạ.
- ❖ Lời mở đầu: Nhóm chúng em giới thiệu tổng quan về công nghệ áp dụng vào đời sống hiện nay.
- ❖ Lý do chọn đề tài: Chúng em muốn chọn đề tài nhằm giúp cho mọi người có góc nhìn khác về thuật toán CNNs có thể áp dụng cho rất nhiều lĩnh vực khác nhau.
- ❖ Chương 1: Giới Thiệu: Nhóm chúng em cho mọi người có cái nhìn sơ bộ về đề tài và những gì chúng em tìm được trong quá trình bọn em tìm hiểu về CNNs.
- ❖ Chương 2: Cơ sở lý thuyết: Ở đây nhóm đồ án chúng em nêu ra các thuật toán, phương pháp chi tiết để chúng em hoàn thành đồ án này.
- ❖ Chương 3: Phương pháp đề xuất: Nhóm chúng em đề xuất ra các đề để giúp cho ứng dụng có thể hoàn thành tốt hơn.
- ❖ Chương 4: Kết quả - Hiện thực: Ở chương này, nhóm chúng em nêu ra các kết quả và hiện thực mà bọn em thực hiện được và chỉ ra các điểm sai sót cũng như yếu kém của chúng em về đồ án này.
- ❖ Chương 5: Kết luận: Nhóm chúng em nêu ra các kết quả chung nhất về ứng dụng của chúng em. Bên cạnh đó, nhóm em đưa ra các đề xuất về hướng phát triển trong tương lai để giúp mọi người có cái nhìn khác về công nghệ trong tương lai.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

2.1 Các khái niệm cơ bản

2.1.1 Thuật toán Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) là một loại mạng nơ-ron sử dụng rộng rãi trong xử lý hình ảnh và đã chứng minh được hiệu quả đáng kể trong việc nhận diện độ tuổi dựa trên hình ảnh khuôn mặt.

CNN đã có được ứng dụng mạnh mẽ trong nhiều lĩnh vực, đặc biệt là trong thị giác máy tính (Computer Vision), nhận dạng hình ảnh, và xử lý ngôn ngữ tự nhiên.

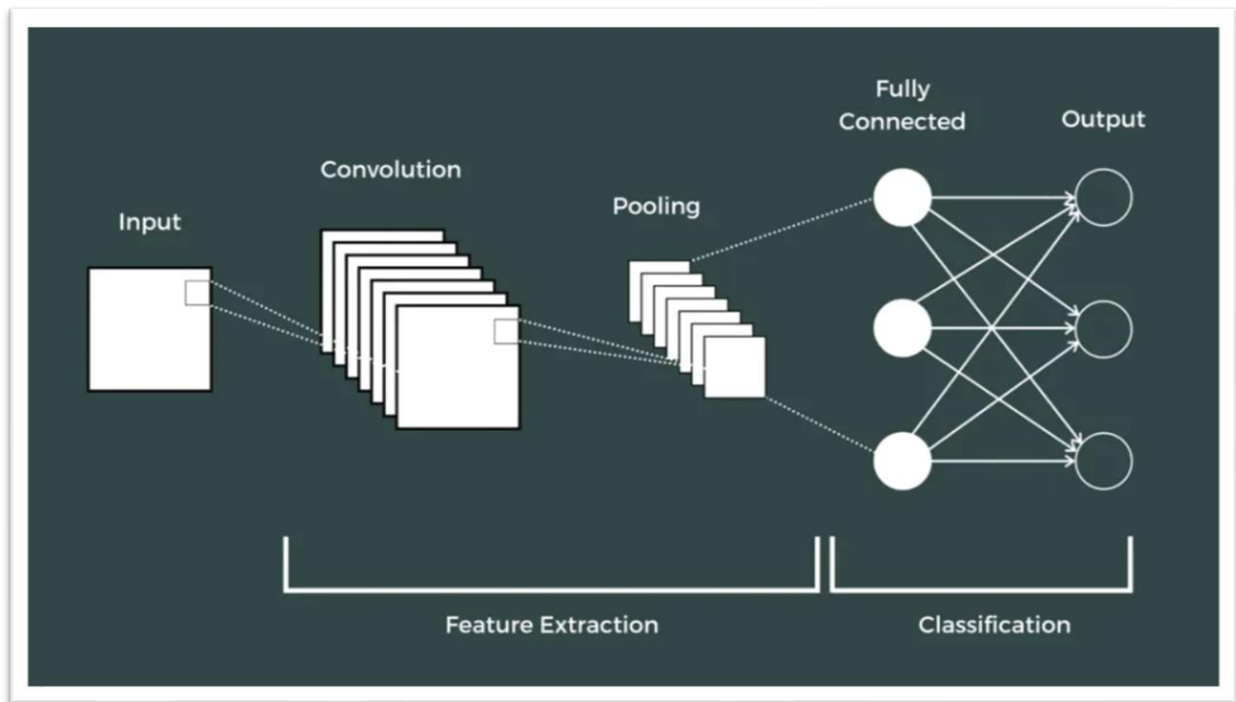
CNN hoạt động bằng cách sử dụng một loạt các tầng (layers) đặc biệt để trích xuất và học các đặc trưng quan trọng từ hình ảnh. Các tầng này bao gồm:

- **Tầng tích chập (Convolutional Layer):** Tầng này sử dụng các bộ lọc (kernels) để trượt qua hình ảnh và thực hiện phép tích chập để trích xuất các đặc trưng cụ thể như cạnh, góc, và các hình dạng cơ bản trên khuôn mặt.
- **Tầng gộp (Pooling Layer):** Tầng này giúp giảm kích thước của dữ liệu bằng cách lấy giá trị trung bình hoặc giá trị tối đa từ các vùng nhỏ trên hình ảnh. Điều này giúp giảm chi phí tính toán và tạo ra tính bất biến với việc thay đổi vị trí của đặc trưng trên hình ảnh.
- **Tầng kết nối đầy đủ (Fully Connected Layer):** Tầng này làm việc như một bộ phân loại cuối cùng. Nó kết nối tất cả đặc trưng đã được trích xuất từ các tầng trước đó và học cách phân loại độ tuổi dựa trên các đặc trưng này.

Liên quan đến quá trình nhận diện độ tuổi của thuật toán CNN:

Trong quá trình huấn luyện, CNN được cung cấp một tập dữ liệu lớn chứa hình ảnh của nhiều người ở các độ tuổi khác nhau. Chúng học cách xác định mối liên quan giữa các đặc trưng trên khuôn mặt và độ tuổi của người dùng, sau đó sử dụng kiến thức này để dự đoán độ tuổi của một người dựa trên hình ảnh khuôn mặt mà chúng nhận được.

Thuật toán CNN được mô tả như sau:



Hình 1. Sơ đồ thuật toán CNN.[1]

Cách thuật toán hoạt động qua mô tả bằng chữ:

Bước 1. **Chuẩn bị hình ảnh đầu vào:** Chuẩn bị hình ảnh cần đọc. Hình ảnh cần phù cần phải có định dạng và kích thước phù hợp với mô hình CNN.

Bước 2. **Tiền xử lý hình ảnh:** Trước khi đưa vào mô hình cần xử lý tiền hình ảnh giúp cho dữ liệu đưa vào phù hợp với mô hình (màu sắc, kích thước, ...).

Bước 3. **Chuyển hình ảnh thành định dạng ma trận:** Mô hình CNN hoạt động bằng dữ liệu số hóa vì vậy cần chuyển sang dạng ma trận. Với mỗi Pixel trong hình tương ứng với một giá trị số.

Bước 4. **Đưa hình ảnh vào mô hình:** Mô hình sẽ thực hiện một loạt phép tích chập và tầng kết nối đầy đủ để trích xuất đặc trưng và thực hiện phân loại hoặc dự đoán dựa trên dữ liệu hình ảnh.

Bước 5. **Nhận kết quả:** Sau khi mô hình hoàn thành quá trình xử lý sẽ đưa ra kết quả. Kết quả là một dự đoán về đối tượng trong hình ảnh, một lớp phân loại, hoặc giá trị số liên quan đến nội dung hình ảnh.

2.1.2 Một số ứng dụng của thuật toán CNN:

Các ứng dụng chung của CNN:

- Nhận dạng khuôn mặt và biometrics (vân tay, móng mắt, ...).
- Nhận diện đối tượng và đối tượng trong thời gian thực (phát hiện và nhận diện đối tượng, động vật, xe hơi, ... và các hệ thống tự lái, an ninh).
- Phân loại hình ảnh (cây, động thực vật, sản phẩm thương mại) để phục vụ cho tìm kiếm hình ảnh trực tuyến, quảng cáo, ...
- Xử lý hình ảnh y tế (tổn thương từ tai nạn, diễn biến trong y tế, ...).
- Tự lái xe và xe tự hành (cảm biến, định hướng, ...).
- Tìm kiếm và gợi ý hình ảnh (các nhu cầu và sở thích cá nhân về hình ảnh).

Các ứng dụng của CNN dành cho nhận diện độ tuổi:

- An ninh và quản lý thời gian làm việc
- Kiểm tra giấy tờ tùy thân tự động
- Y tế và nghiên cứu lão hóa
- Phân loại nội dung trực tuyến
- Y tế thẩm mỹ và làm đẹp
- Giám sát hành vi khách hàng

2.1.3 Mô tả về ba lớp của mô hình CNN

Mô hình CNN (Convolutional Neural Networks) bao gồm ba lớp chính: Lớp tích chập (Convolutional Layer), Lớp gộp (Pooling Layer), và Lớp kết nối đầy đủ (Fully Connected Layer).

2.1.3.1 Mô tả chi tiết các lớp:

Lớp Tích Chập (Convolutional Layer):

- Lớp tích chập là lớp đầu tiên trong mô hình CNN.
- Nhiệm vụ chính của lớp này là trích xuất các đặc trưng cụ thể từ hình ảnh đầu vào bằng cách thực hiện các phép tích chập trên nó.
- Lớp tích chập sử dụng một tập hạt nhân (kernel) có kích thước nhỏ di chuyển qua hình ảnh để tạo ra các đặc trưng (feature maps).
- Các đặc trưng này tương ứng với các đặc điểm như cạnh, đường nét, và các thông tin cụ thể khác trong hình ảnh.
- Số lượng bộ lọc (kernels) và kích thước của chúng có thể thay đổi tùy thuộc vào cấu trúc của mô hình.

Lớp Gộp (Pooling Layer):

- Lớp gộp thường nằm sau lớp tích chập.
- Nhiệm vụ chính của lớp gộp là giảm kích thước của đặc trưng được trích xuất từ lớp tích chập.
- Điều này giúp giảm chi phí tính toán và giúp mô hình trở nên cơ động hơn.
- Phép gộp thường là phép lấy giá trị tối đa (max pooling) hoặc phép lấy giá trị trung bình (average pooling) trong một vùng cụ thể của đặc trưng.
- Kích thước vùng và bước nhảy (stride) có thể thay đổi tùy thuộc vào mục tiêu của mô hình.

Lớp Kết Nối Đầy Đủ (Fully Connected Layer):

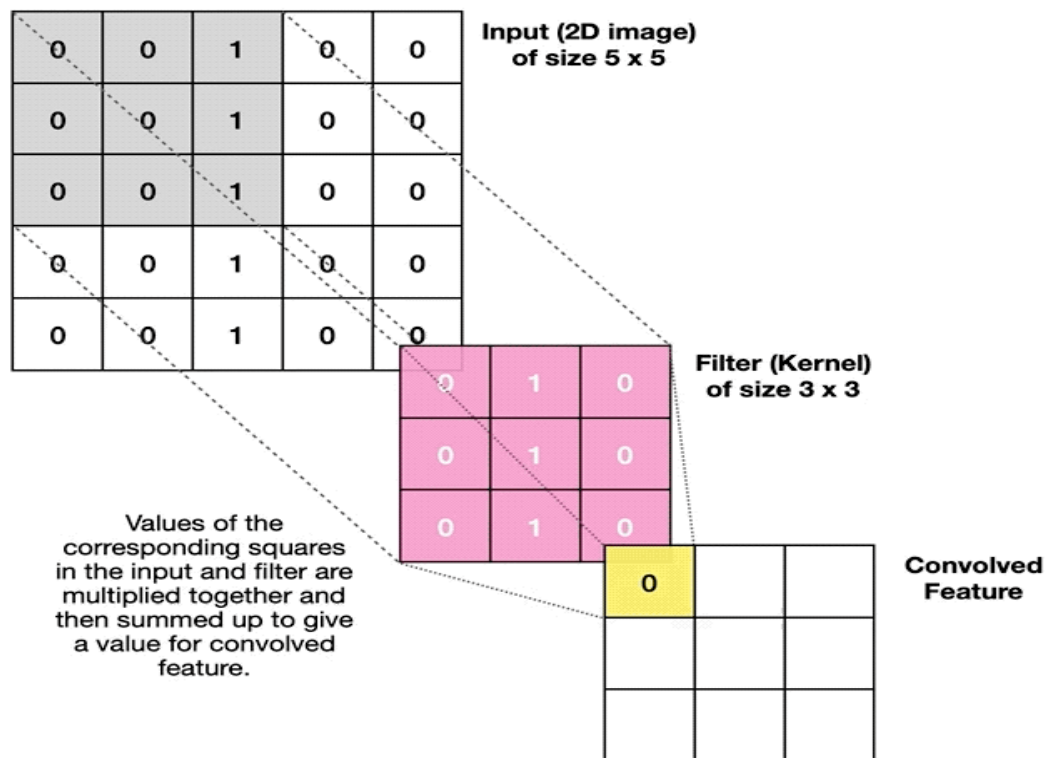
- Lớp kết nối đầy đủ là lớp cuối cùng trong mô hình CNN.
- Lớp này thường sử dụng để kết nối tất cả các đặc trưng đã được trích xuất từ các lớp trước đó để thực hiện phân loại hoặc dự đoán.
- Các đặc trưng từ lớp gộp được duỗi thành một vector và sau đó được đưa qua một hoặc nhiều tầng kết nối đầy đủ để tạo ra đầu ra dự đoán.
- Số lượng nơ-ron trong lớp kết nối đầy đủ thường phụ thuộc vào số lớp đầu ra mong muốn và mục tiêu cụ thể của ứng dụng.

2.1.3.2 Nguyên tắc hoạt động của ba lớp:

a) Nguyên tắc hoạt động của các lớp:

✚ Lớp Tích Chập (Convolutional Layer):

- Lớp tích chập trích xuất các đặc trưng cụ thể từ hình ảnh đầu vào bằng cách sử dụng các bộ lọc (kernels) di chuyển qua hình ảnh.
- Mỗi bộ lọc tạo ra một đặc trưng(feature map) bằng cách thực hiện phép tích chập.
- Các đặc trưng này biểu diễn các thông tin như cạnh, đường nét, và các đặc điểm quan trọng trong hình ảnh.
- Kết quả từ lớp này là một tập hợp các feature maps, mỗi feature map tương ứng với một đặc trưng cụ thể.

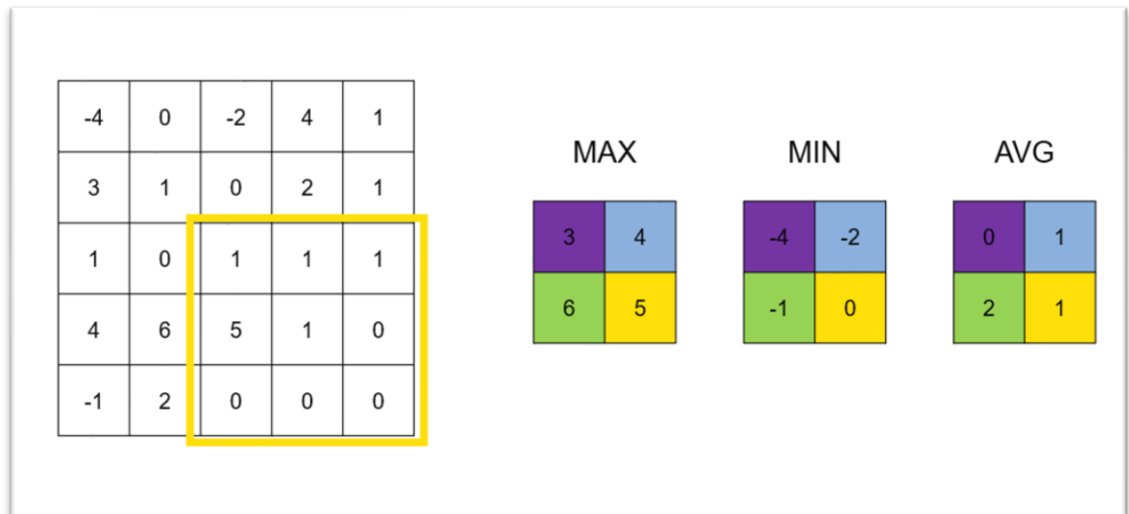


Hình 2. Lớp tích chập convolution.[2]

✚ Lớp Gộp (Pooling Layer):

- Lớp gộp nằm sau lớp tích chập và thường được sử dụng để giảm kích thước của đặc trưng.
- Phép gộp (pooling) thường là phép lấy giá trị tối đa (max pooling) hoặc phép lấy giá trị trung bình (average pooling) trong các vùng cụ thể của đặc trưng.
- Mục tiêu là giảm chi phí tính toán và tạo ra biểu diễn đặc trưng cơ bản.

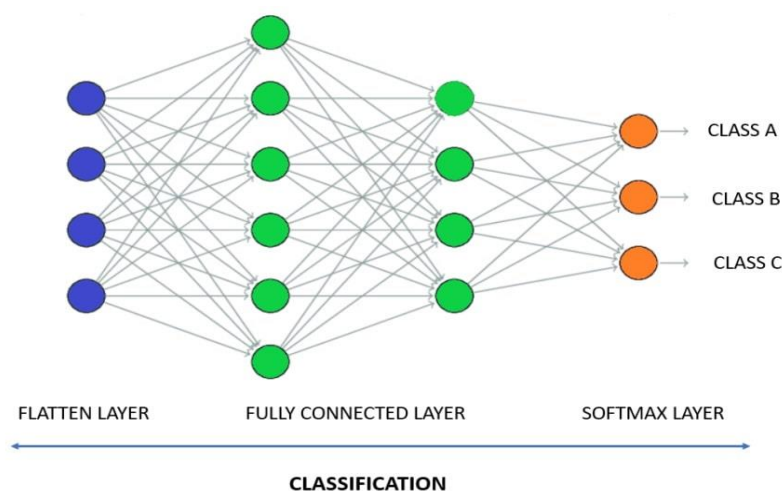
- Kết quả từ lớp này là tập hợp các đặc trưng đã gộp lại với kích thước nhỏ hơn.



Hình 3. Lớp Pooling.[2]

✚ Lớp Kết Nối Đầy Đủ (Fully Connected Layer):

- Lớp kết nối đầy đủ thường là lớp cuối cùng của mô hình và được sử dụng để thực hiện phân loại hoặc dự đoán.
- Các đặc trưng từ lớp gộp được duỗi thành một vector và sau đó được đưa qua một hoặc nhiều tầng kết nối đầy đủ.
- Tầng này học cách kết hợp các đặc trưng để tạo ra đầu ra dự đoán.
- Kết quả cuối cùng từ lớp này là đầu ra dự đoán, chẳng hạn như lớp của đối tượng hoặc giá trị dự đoán.



Hình 4. Lớp Fully Connected Layer.[1]

b) Liên kết giữa các lớp

- Dữ liệu và thông tin truyền qua các lớp theo một dòng luồng dạng chuỗi.
- Các trọng số của mô hình được học thông qua quá trình huấn luyện sẽ ảnh hưởng đến cách thông tin lan truyền từ lớp này đến lớp khác.
- Quá trình lan truyền ngược (backpropagation) được sử dụng để điều chỉnh và cải thiện các trọng số này dựa trên lỗi trong quá trình huấn luyện.
- Tầng tích chập và gộp thường xen kẽ nhau, và đầu ra của lớp gộp được truyền vào lớp kết nối đầy đủ để thực hiện phân loại hoặc dự đoán.

2.2 Các bài toán về CNN**❖ Regression Task:**

Mô Tả: Bài toán đặt ra là xây dựng một mô hình có khả năng dự đoán giá trị số của độ tuổi dựa trên khuôn mặt.

Thách Thức: Yêu cầu một tập dữ liệu lớn, đa dạng và có nhãn chính xác với giá trị độ tuổi để huấn luyện mô hình regression.

❖ Classification Task:

Mô Tả: Đánh giá độ tuổi thành các khoảng nhất định (ví dụ: dưới 18, 18-25, 26-35, và như vậy) thay vì dự đoán giá trị cụ thể.

Thách Thức: Yêu cầu một tập dữ liệu phân loại với nhãn độ tuổi được gán vào các nhóm, thường được đặt theo khoảng độ tuổi.

❖ Data Imbalance:

Mô Tả: Dữ liệu có thể không cân bằng giữa các nhóm độ tuổi, ví dụ, có thể có nhiều hình ảnh của nhóm tuổi trẻ hơn so với nhóm tuổi cao.

Thách Thức: Điều này có thể dẫn đến mô hình có thể chệch hướng, hiểu biểu diễn của đối tượng một cách không chính xác nếu có sự thiếu sót về dữ liệu.

❖ Age Progression/Regression:

Mô Tả: Bài toán về cách mô hình có thể dự đoán sự thay đổi độ tuổi của một cá nhân qua thời gian (progression) hoặc ngược lại (regression).

Thách Thức: Đòi hỏi sự theo dõi và gán nhãn độ tuổi qua các hình ảnh của cùng một cá nhân, có thể là một thách thức khi tìm dữ liệu.

❖ Cross-Age Face Recognition:

Mô Tả: Đây là bài toán nhận diện khuôn mặt của một cá nhân qua các độ tuổi khác nhau.

Thách Thức: Yêu cầu mô hình có khả năng hiểu biểu diễn của một cá nhân không chỉ qua thời gian mà còn qua các biến đổi về độ tuổi.

Facial Age Estimation in the Wild:

Mô Tả: Đối mặt với việc ước lượng độ tuổi từ hình ảnh có điều kiện ánh sáng và nền không kiểm soát.

Thách Thức: Yêu cầu mô hình có khả năng tổng quát hóa từ dữ liệu thu thập trong điều kiện thực tế và không kiểm soát.

❖ *Ethnicity and Gender Bias:*

Mô Tả: Mô hình có thể hiển thị thiên lệch đối với độ tuổi dựa trên sắc tộc hoặc giới tính.

Thách Thức: Yêu cầu dữ liệu đa dạng và chú ý để giảm thiểu sự chệch hướng và đảm bảo tính công bằng.

2.3 Cascade Classifier

a) Cascade Classifier là gì?

Cascade Classifier là một phương pháp phổ biến trong xử lý ảnh và thị giác máy tính để nhận diện đối tượng. Được giới thiệu đầu tiên bởi Viola và Jones vào năm 2001, Cascade Classifier được sử dụng để nhận diện các đối tượng như khuôn mặt trong ảnh hoặc video.

b) Các bước chính của Cascade Classifier

Haar Feature Selection and Integral Image:

- Sử dụng các Haar-like features để biểu diễn các đặc trưng của hình ảnh.
- Tích hợp hình ảnh (integral image) để tăng tốc quá trình tính toán.

Adaboost Training:

- Sử dụng thuật toán Adaboost để lựa chọn các đặc trưng quan trọng.
- Xây dựng một chuỗi các weak classifiers (thường là các máy học đơn giản).

Cascade Structure:

- Chia cascade thành nhiều tầng.
- Mỗi tầng sẽ có một số weak classifiers.
- Các tầng sau sẽ chấp nhận các ảnh dương giảm số lượng ảnh được xem xét.

Khi áp dụng Cascade Classifier để nhận diện đối tượng, quá trình kiểm tra được thực hiện qua từng tầng. Nếu ảnh không vượt qua một tầng nào đó, nó sẽ bị loại bỏ ngay lập tức, giúp tăng tốc quá trình nhận diện.

- Sử dụng Cascade Classifier để xác định vùng khuôn mặt: Cascade Classifier có thể giúp xác định vị trí của khuôn mặt trong hình ảnh.
- Sử dụng CNN để nhận diện độ tuổi: Sau khi có được vùng khuôn mặt, bạn có thể sử dụng mạng CNN để phân loại độ tuổi. Mạng CNN có thể được huấn luyện trước trên bộ dữ liệu độ tuổi lớn để có khả năng nhận diện và phân loại độ tuổi một cách chính xác.

Kết hợp giữa Cascade Classifier và CNN giúp tận dụng ưu điểm của cả hai phương pháp: tăng tốc quá trình vùng quan trọng và khả năng học sâu và chính xác của mạng neural.

2.4 Categorical Crossentropy và Adam

Categorical Crossentropy (Hàm Mất Mát):

- + Loại: Categorical Crossentropy là một loại hàm mất mát được sử dụng cho các vấn đề phân loại, nơi nhiệm vụ là phân loại dữ liệu đầu vào vào nhiều danh mục.
- + Sử Dụng: Được thiết kế đặc biệt cho các vấn đề phân loại đa lớp, nơi mỗi trường hợp chỉ có thể thuộc một lớp. Nếu vấn đề phân loại của bạn có nhiều hơn hai lớp, Categorical Crossentropy là sự lựa chọn phổ biến.

Adam (Bộ Tối Ưu Hóa):

- + Loại: Adam là một thuật toán tối ưu hóa được sử dụng để huấn luyện mô hình học sâu.
- + Đặc Điểm: Adam kết hợp ý tưởng từ hai thuật toán tối ưu hóa phổ biến khác: RMSprop và Momentum. Nó duy trì trung bình động của gradient của tham số và gradient bình phương của chúng, điều chỉnh tỷ lệ học của mỗi tham số cá nhân.
- + Ưu Điểm: Adam nổi tiếng với sự hiệu quả và hiệu suất trong việc huấn luyện mô hình nơ-ron sâu. Thường thì nó hội tụ nhanh hơn so với các biến thể gradient descent truyền thống.

2.5 Các nghiên cứu liên quan

2.5.1 Các bài báo về CNN cần biết

a) Các bài báo và sách cơ bản về CNN

Để hiểu về thuật toán *Convolutional Neural Networks* chúng ta có một số bài báo cần đọc để hiểu về chúng, nhóm đã tìm được một số bài báo về mô tả cho quy trình cũng như các điểm đặc biệt của thuật toán CNN:

- **"Deep Learning"** - Yann LeCun, Yoshua Bengio, và Geoffrey Hinton (2015)
 - Bài báo này tổng quan về lĩnh vực học sâu và nêu rõ những tiến bộ quan trọng của CNN và RNNs.
- **"Deep Learning: A Review"** - Yann LeCun, Yoshua Bengio, và Geoffrey Hinton (2015)
 - Bài báo này cũng của nhóm tác giả trên và tổng quan về lĩnh vực học sâu, bao gồm cả CNN và RNNs.
- **"Deep Learning for Computer Vision"** - Rajalingappaa Shanmugamani (2019)
 - Cuốn sách này tập trung vào việc sử dụng học sâu và CNN trong xử lý hình ảnh và thị giác máy tính.
- **"Understanding Machine Learning: From Theory to Algorithms"** - Shai Shalev-Shwartz và Shai Ben-David (2014)
 - Cuốn sách này bao gồm chương về mạng nơ-ron và học sâu, giúp bạn hiểu về cơ bản của CNN và lý thuyết đằng sau chúng.

Đây là các bài báo tập trung mô tả về các thành phần cơ bản của CNN và các mô hình liên quan đến Deep Learning (học sâu) để người đọc tìm hiểu.

b) Các bài báo và sách về mô hình học sâu và bước tiến của CNN

Ở đây nhóm đã tìm được một số sách và bài báo về mô hình học sâu của CNN. Bên cạnh đó là các bước tiến của mô hình Convolutional Neural Networks:

- **"Deep Residual Learning for Image Recognition"** - Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2015)
 - Bài báo này giới thiệu mô hình ResNet, một mô hình CNN rất sâu và mô tả cách mà nó sử dụng các khối Residual Blocks để cải thiện hiệu suất đáng kể.
- **"An overview of gradient descent optimization algorithms"** - Sebastian Ruder (2016)
 - Bài báo này giới thiệu các thuật toán tối ưu hóa gradient descent, mà là một phần quan trọng trong việc đào tạo mô hình CNN.
- **"Regularization Techniques for Deep Learning"** - Sebastian Ruder (2016)
 - Bài báo này giải thích về kỹ thuật regularization trong học sâu và làm thế nào nó có thể giúp cải thiện hiệu suất của mô hình CNN.

-"Visualizing and Understanding Convolutional Networks" - Matthew D. Zeiler và Rob Fergus (2013)

- Bài báo này giới thiệu các phương pháp để trực quan hóa và hiểu mô hình CNN, giúp bạn nắm vững cách mà các lớp và bộ lọc hoạt động.

-"A Gentle Introduction to Convolutional Neural Networks" - David Stutz (2018)

- Bài viết này giới thiệu các khái niệm cơ bản về CNN một cách dễ hiểu và đi sâu vào các yếu tố quan trọng của chúng.

Các bài báo trên giúp chúng ta hiểu rõ và sâu về thuật toán CNN cũng như các bước tiến mà chúng mang lại. Từ đó cho chúng ta có cái nhìn cụ thể và rõ ràng về các bước, mô hình và kiến trúc cụ thể của một mạng nơ-ron. Cho thấy được tầm quan trọng mà CNN có thể mang lại ở mọi lĩnh vực của khoa học – kỹ thuật.

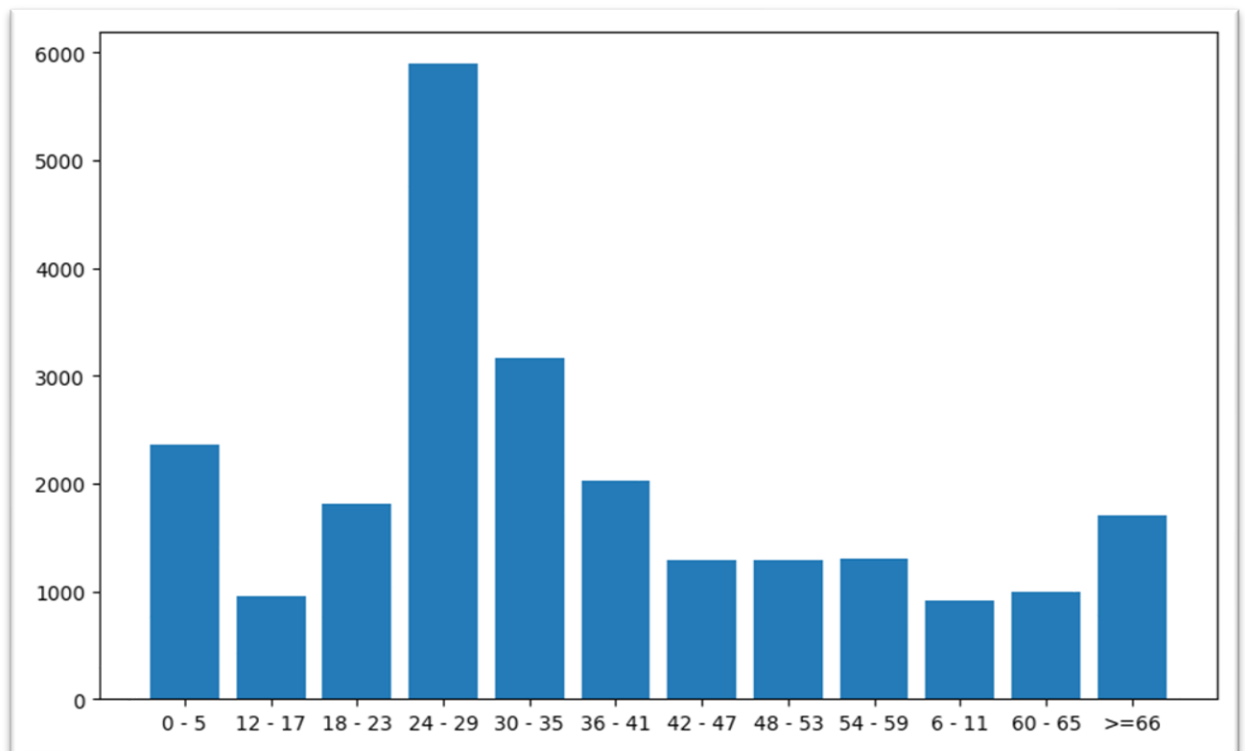
CHƯƠNG 3: CÁC PHƯƠNG PHÁP ĐỀ XUẤT

3.1 Yêu cầu bài toán

Bài toán dự đoán độ tuổi yêu cầu đầu vào là một hình ảnh có một hoặc nhiều khuôn mặt chính diện khuôn mặt, không bị xoay ảnh, hệ ảnh RGB. Hình ảnh sẽ được cắt ra từng khuôn mặt thông qua *Cascade Classifier* để nhận diện khuôn mặt sau đó các hình ảnh đã được cắt đưa vào mô hình CNN để dự đoán độ tuổi. Đầu ra sẽ là độ tuổi của ảnh được đưa vào mô hình.

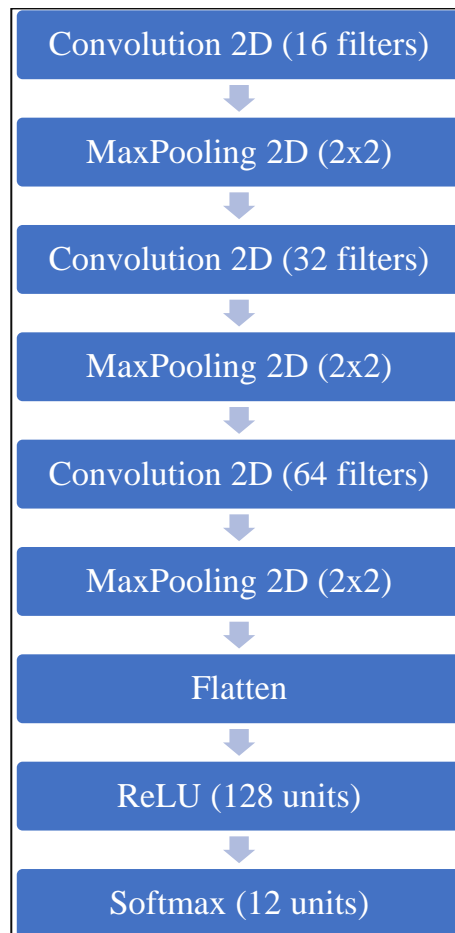
3.1.1 Phương pháp đề xuất thực thi

Dữ liệu chuẩn bị gồm tập dữ liệu *UTKFace[12]* (23,708 hình ảnh) và *natural image* (5,913 hình ảnh) tổng 29,621 hình ảnh, sử dụng tập dữ liệu *UTKFace[12]* phân ra các độ tuổi trong khoảng “0 đến 5”, “6 đến 12”, “12 đến 17”, “18 đến 23”, “24 đến 29”, “30 đến 35”, “36 đến 41”, “42 đến 47”, “48 đến 53”, “54 đến 59”, “60 đến 65”, “trên hoặc bằng 66”. Như vậy ta được số lượng dữ liệu sau:



Hình 5. Biểu đồ phân bố lượng hình ảnh trên mỗi độ tuổi thu thập được.[4]

Dữ liệu sẽ được tách ra 1/3 cho tập dữ liệu kiểm định (val/test), phần còn lại dành cho tập dữ liệu huấn luyện. Tập dữ liệu huấn luyện sẽ được đưa vào mô hình CNN có cấu trúc như sau:



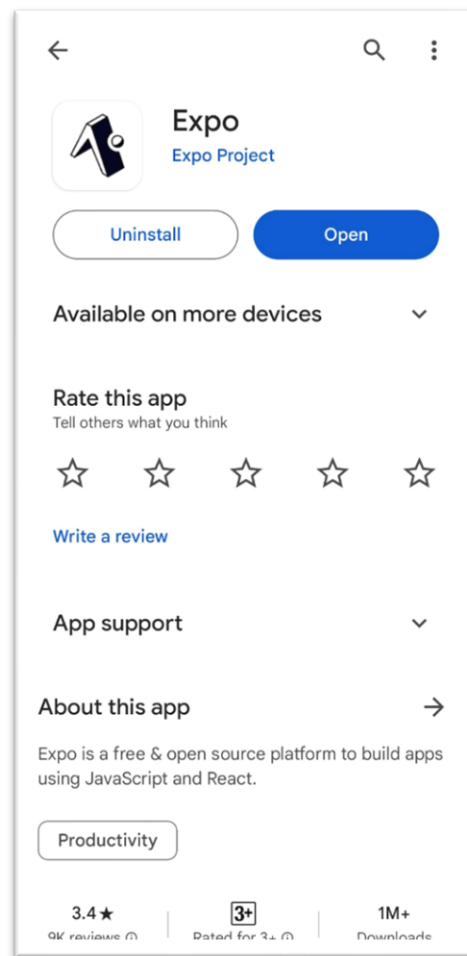
Hình 6. Mô hình cấu trúc CNN. [3]

Các lớp tích chập có kernel size đều là (3, 3), kích thước pool của max pooling là 2x2. Tổng mô hình có 320,172 tham số tham gia quá trình huấn luyện. Model sử dụng CategoricalCrossentropy cho hàm loss, optimizer dùng Adam, các thông số được đề mặc định bởi tensorflow.

Sử dụng react native, expo tạo ra ứng dụng hoạt động trên android thông qua camera trên thiết bị di động để chụp ảnh khuôn mặt. Dùng python để tạo ra server cho model CNN hoạt động.

3.2 Hiện thực

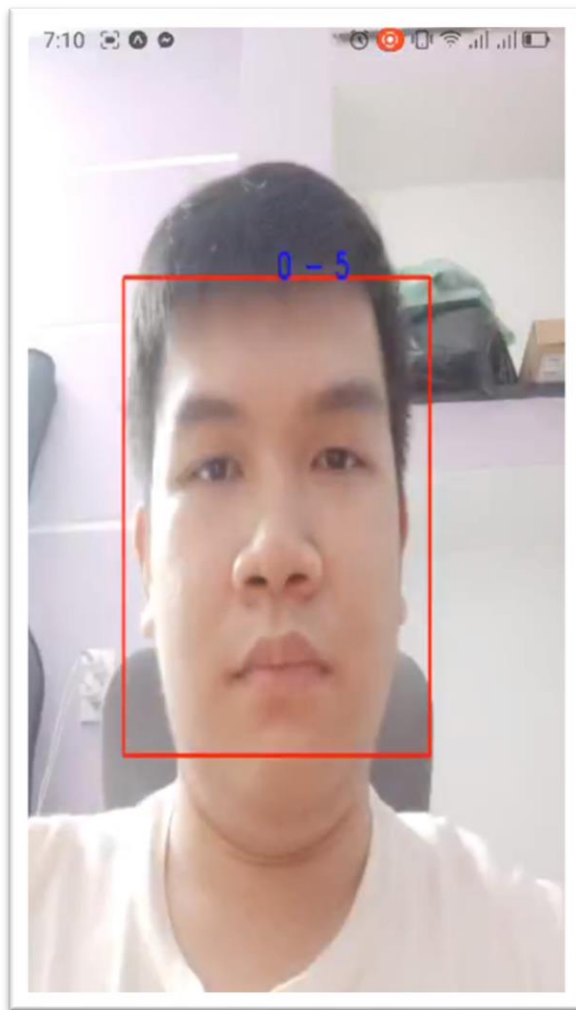
Bước 1: Tải ứng dụng Expo Go từ Google Play Store (Android):



Hình 7. Tải ứng dụng Expo Go

Bước 2: Mở ứng dụng Expo Go trên điện thoại và chọn tùy chọn "Scan QR Code" hoặc "Quét Mã QR" sau đó Expo Go sẽ tự động mở ứng dụng và tải ứng dụng lên.

Bước 3: Khi ta chụp một ảnh, ứng dụng sẽ xử lý và đưa ra kết quả dự đoán về độ tuổi của dữ liệu được đưa vào:



Hình 8. Dự đoán độ tuổi ở hệ điều hành Android

CHƯƠNG 4: KẾT QUẢ HIỆN THỰC

4.1 Yêu cầu phần cứng, phần mềm

4.1.1 Yêu cầu phần cứng

Khi sử dụng phần cứng bọn em ưu tiên sử dụng các phần miễn phí và có sẵn:

- Sử dụng Google colab phiên bản free để thực hiện huấn luyện model CNN
- GPU: TPA
- RAM: 12 GB

Đối với lập trình ứng dụng android, nhóm sử dụng cấu hình này trong quá trình phát triển ứng dụng:

- CPU: Intel Core i3 10100F @3.60 Hz
- Mainboard: Gigabyte H510M S2H
- RAM: 16GB
- GPU: NVIDIA GeForce GTX 1050Ti 4GB GDDR5

Để chạy ứng dụng android, có thể dùng bất cứ máy có hệ điều hành windows 10 trở lên.

4.1.2 Yêu cầu phần mềm

Nhóm em đã tìm hiểu và đưa ra được các phần mềm và công cụ có thể giúp hoàn thành đồ án một cách tốt nhất:

Windows 11 Pro version 10.0.22621:

- Giao diện người dùng: Windows 11 cung cấp một giao diện người dùng hiện đại và thân thiện với người dùng, có thể cải thiện trải nghiệm làm việc và tương tác với hệ thống.
- Tích hợp sâu: Windows hỗ trợ nhiều ứng dụng và công nghệ phát triển, là một môi trường thuận tiện để triển khai ứng dụng.

Android Studio

- Phát triển ứng dụng Android: Là một IDE (Integrated Development Environment) chính cho việc phát triển ứng dụng di động Android.
- Emulator tích hợp: Cung cấp trình giả lập Android tích hợp để kiểm thử ứng dụng trên nhiều thiết bị và phiên bản Android khác nhau.

Visual Studio Code:

- Nhẹ và linh hoạt: Là một trình soạn thảo mã nguồn nhẹ, dễ mở rộng và hỗ trợ nhiều ngôn ngữ lập trình.
- Hỗ trợ Git tích hợp: Cung cấp tích hợp tốt với Git để quản lý mã nguồn và dự án.

GitHub:

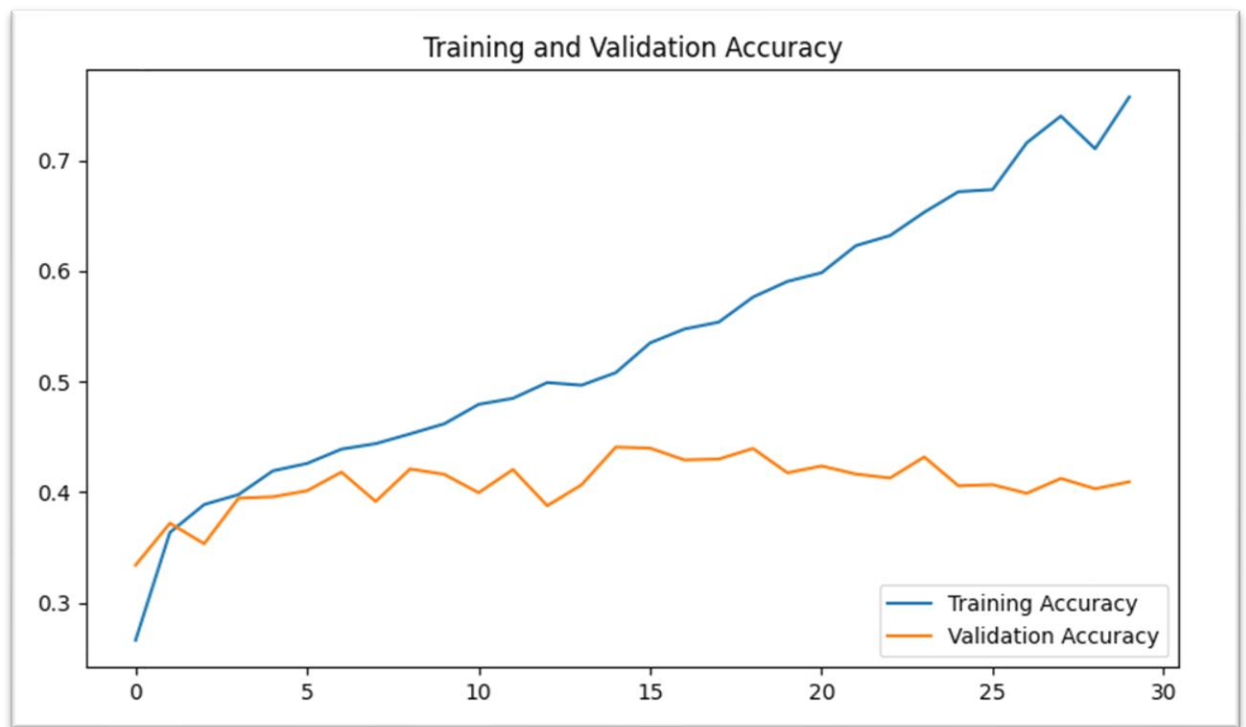
- Quản lý mã nguồn: GitHub là một dịch vụ lưu trữ mã nguồn phổ biến, giúp đội ngũ phát triển làm việc cộng tác và quản lý phiên bản mã nguồn.
- Cộng đồng lớn: GitHub có một cộng đồng lớn, giúp chia sẻ mã nguồn, hợp tác, và sử dụng các thư viện và dự án mở nguồn.

Anaconda:

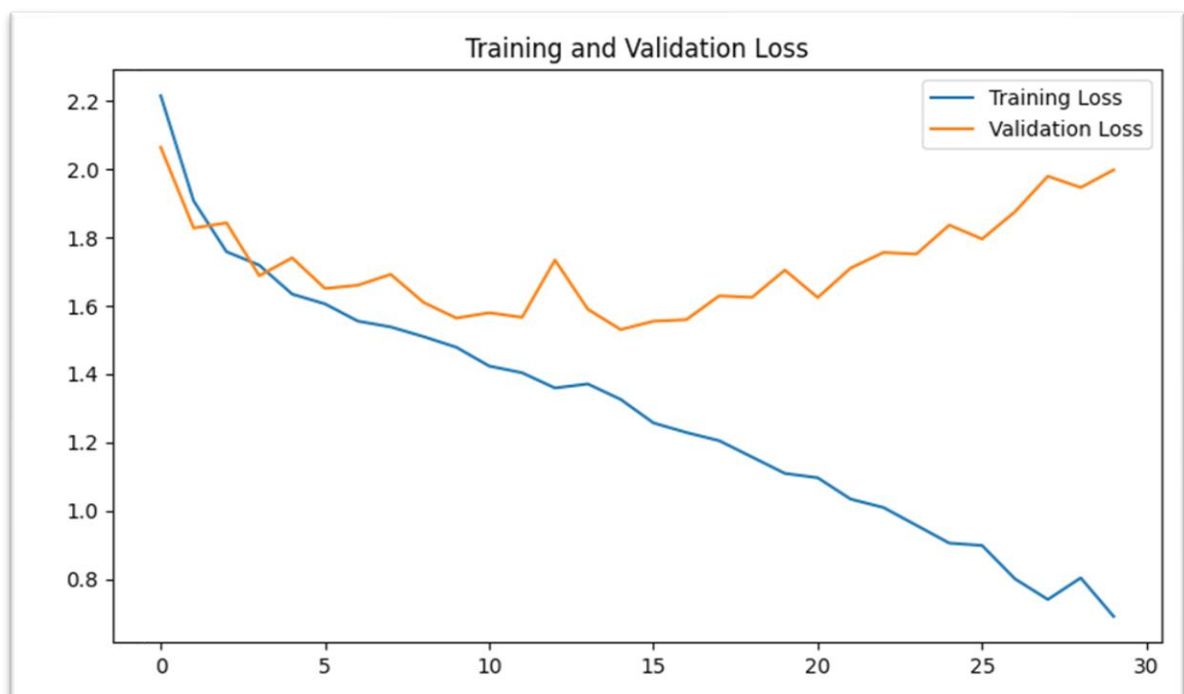
- Quản lý môi trường và gói: Anaconda giúp quản lý môi trường phát triển và các gói Python, giúp giải quyết vấn đề phức tạp liên quan đến phụ thuộc và tương thích gói.
- Jupyter Notebooks tích hợp: Anaconda đi kèm với Jupyter Notebooks, một công cụ mạnh mẽ cho việc thực hiện và chia sẻ công việc phân tích dữ liệu và lập trình.

4.2 Kết quả hiện thực**4.2.1 Huấn luyện**

Quá trình huấn luyện gồm 30 epochs, batch size 512.



Hình 9. Biểu đồ biểu diễn độ chính xác của mô hình trên tập dữ liệu huấn luyện và tập dữ liệu đánh giá thay đổi qua từng epoch



Hình 10. Biểu đồ biểu diễn sự hội tụ của mô hình trên tập dữ liệu huấn luyện và tập dữ liệu đánh giá thay đổi qua từng epoch

4.2.2 Ưu điểm

Nhóm chúng em nhận thấy ứng dụng “Dự Đoán Độ Tuổi Người Dựa Trên Ảnh Khuôn Mặt Sử Dụng Thuật Toán CNNs” có nhiều ưu điểm nổi trội và khá thiết thực:

- Độ chính xác trên tập dữ liệu huấn luyện khá cao đạt: 75.72%.
- Tập dữ liệu khá nhiều và chi tiết giúp hệ thống có thể đọc tốt hơn.
- Được training nhiều lần giúp các lỗi được khắc phục khá nhiều.

4.2.3 Nhược điểm

Bên cạnh những ưu điểm phía trên thì nhược điểm của đồ án chúng em khá nhiều và còn nhiều hạn chế, nhóm chúng em mong muốn tối ưu hơn cho hệ thống của chúng em nên chúng em đã nêu ra nhằm đặt mình vào mục tiêu phải hoàn thành:

- Hội tụ loss chưa nhanh.
- Độ chính xác trên tập val chưa cao, chỉ 40.30%.
- Cần có bộ nhớ khá lớn.

CHƯƠNG 5: KẾT LUẬN

5.1 Kết quả đạt được

Độ chính xác mô hình CNN 71.03% dành cho tập dữ liệu huấn luyện và 40.30% cho tập dữ liệu đánh giá. Model hoạt động tốt trên môi trường python dùng Flask để tạo API, ứng dụng trên android có thể chụp ảnh bằng camera trước và sau. Dùng ảnh đã chụp gửi đến server để dự đoán độ tuổi bằng mô hình CNN.

5.2 Phương pháp đề xuất

❖ Thu thập và chuẩn bị dữ liệu:

- Dùng tập dữ liệu kết hợp UTKFace[12].

❖ Kiến trúc CNN

- Sử dụng kiến trúc như Hình 6.

❖ Đào tạo mô hình

- Dữ liệu được huấn luyện chiếm 1/7, còn lại để kiểm tra đánh giá (val/test) của UTKFace[12]. Huấn luyện với 30 epochs và batch size là 512.

❖ Kiểm định và tinh chỉnh mô hình

- Dữ liệu đánh giá chiếm 1/3 của UTKFace[12].

❖ Đánh giá và Kiểm thử

- Đánh giá trên tập huấn luyện: độ chính xác 75.72% và hàm mất mát đạt giá trị 0.6904 trên tập huấn luyện. Tốc độ hội tụ chậm.
- Kiểm thử trên tập đánh giá(val/test): độ chính xác 40.93% và hàm mất mát đạt giá trị 1.9977 trên tập huấn luyện. Hội tụ bị móc ngược nên mô hình chưa tốt cần tinh chỉnh siêu tham số thêm.

5.3 Bước mở rộng trong tương lai

❖ Tăng cường tập dữ liệu

- **Thu thập Dữ liệu Đa dạng:** Mở rộng tập dữ liệu để bao gồm nhiều đặc điểm người dùng và điều kiện chụp hình khác nhau, bao gồm cả độ tuổi, giới tính, và biến thể ngoại hình.

❖ Kiến trúc mô hình nâng cao

- **Sử Dụng Kiến Trúc Mô Hình Hiện Đại:** Xem xét sử dụng các kiến trúc mô hình CNN hiện đại như EfficientNet, ViT (Vision Transformer), để cải thiện độ chính xác và tốc độ huấn luyện.
- ❖ **Học tập đa nhiệm**
 - **Nhận Diện Nhiều Đặc Trưng:** Mở rộng mô hình để có khả năng nhận diện nhiều đặc trưng khác nhau, chẳng hạn như giới tính, nhóm ethnic, để tạo ra một ứng dụng nhận diện độ tuổi đa nhiệm và đa nhiệm.
- ❖ **Tối ưu hóa và Phân loại sự rõ ràng**
 - **Tối Ưu Hóa Tham Số:** Tiếp tục tối ưu hóa các tham số của mô hình để giảm kích thước mô hình và tăng tốc độ dự đoán.
 - **Phân Loại Rõ Ràng:** Mở rộng phương pháp phân loại để không chỉ dự đoán độ tuổi mà còn cung cấp thông tin chi tiết hơn về người dùng, như nhóm độ tuổi.
- ❖ **Tích hợp thêm công nghệ học máy mới**
 - **Tích Hợp Attention Mechanisms:** Sử dụng cơ chế chú ý (attention mechanisms) để tăng khả năng tập trung vào các đặc trưng quan trọng.
 - **Sử Dụng Transfer Learning:** Tích hợp transfer learning từ các mô hình đã được huấn luyện trước để tận dụng kiến thức đã học.
- ❖ **Phát triển ứng dụng thực tế**
 - **Tích Hợp vào Hệ Thống An Ninh và Quản lý Sự Kiện:** Phát triển tích hợp với hệ thống an ninh và quản lý sự kiện để giúp theo dõi và đáp ứng nhanh chóng khi cần thiết.
 - **Ứng Dụng trong Công Nghiệp và Dịch Vụ:** Mở rộng áp dụng của hệ thống vào các lĩnh vực như bán lẻ, giải trí, quảng cáo, và các dịch vụ trực tuyến khác.
- ❖ **Quản lý dữ liệu và Quyền riêng tư**
 - **Tăng Cường Quản lý Dữ liệu:** Cải thiện quy trình quản lý dữ liệu, đảm bảo sự an toàn và tuân thủ với các quy định về quyền riêng tư.
 - **Cung Cấp Tùy Chọn Người Dùng:** Tích hợp tính năng cho phép người dùng kiểm soát thông tin cá nhân và quyết định việc sử dụng dữ liệu.
- ❖ **Hệ thống giám sát và Bảo trì liên tục**

- ***Theo dõi Hiệu Suất Liên Tục:*** Thiết lập hệ thống giám sát để theo dõi hiệu suất của hệ thống liên tục và cảnh báo về bất kỳ vấn đề nào.
- ***Cập Nhật Định Kỳ:*** Thực hiện cập nhật định kỳ cho mô hình và phần mềm để duy trì sự chính xác và hiệu suất tốt nhất.

TÀI LIỆU THAM KHẢO

- [1].<https://viblo.asia/p/tim-hieu-ve-convolutional-neural-networks-cnn-naQZRkr0lvx> (28 – 09 – 2023)
- [2].<https://vi.wikipedia.org/wiki/CNN> (28 – 09 – 2023)
- [3].<https://topdev.vn/blog/thuat-toan-cnn-convolutional-neural-network/>
- [4].<https://vietnix.vn/cnn-la-gi/> (29 – 09 – 2023)
- [5].<https://techie.vn/nhan-dien-khuon-mat-cong-nghe-voi-nhieu-tiem-nang-phat-trien/> (28 – 09 – 2023)
- [6].https://docs.opencv.org/4.x/db/d28/tutorial_cascade_classifier.html (01 – 10 – 2023)
- [7].<https://vi.wikipedia.org/wiki/Android> (03 – 10 – 2023)
- [8].<https://niithanoi.edu.vn/cach-lam-viec-voi-api-trong-python.html> (03 – 09 – 2023)
- [9].<https://www.tensorflow.org/> (02 – 09 – 2023)
- [10].<https://reactnative.dev/> (02 – 09 – 2023)
- [11].<https://expo.dev/> (01 – 09 – 2023)
- [12]. <https://susanqq.github.io/UTKFace/> (02-11-2023)