TRƯỜNG ĐẠI HỌC CÔNG NGHỆ-ĐHQGHN

KHOA ĐIỆN TỬ VIỄN THÔNG

---ooOoo---

BÁO CÁO GIỮA KỲ

Môn: Thị giác máy (INT3412_1)

Đề tài : Face Detection

Chủ đề: Ứng dụng của Deep Learning để phát hiện giới tính con người sử dụng mô hình SmallerVGGNet của mạng thần kinh chuyển đổi (CNN)

Giáo viên giảng dạy: *Ph.D Lê Thanh Hà*

Sinh viên thực hiện : *Trần Quyết Thắng - 19020622*

Mai Hồng Nhật - 19020590

Trần Ngọc Thắng - 19020624

Hà Nội – Tháng 5 năm 2023

LỜI CẢM ƠN

Chúng em xin trân trọng cảm ơn thầy giáo *PhD. Lê Thanh Hà* đã tận tình giảng dạy và hướng dẫn để chúng em hoàn thành tốt báo cáo này. Phần lớn nội dung trình bày được lấy từ bài giảng, tài liệu tham khảo do thầy cung cấp. Tuy nhiên do thời gian làm bài tập lớn không có nhiều, tài liệu chuyên sâu không đầy đủ, cũng như chưa có nhiều kinh nghiệm về đề tài này. Nên trong báo cáo của chúng em khó có thể tránh khỏi thiếu sót. Rất mong nhận được sự đóng góp của thầy để nội dung đề tài của báo cáo được hoàn chỉnh hơn. *Chúng em xin chân thành cảm ơn !*

LỜI NÓI ĐẦU

Hiên nay, sự quan tâm đến phân loại giới tính tự động đã tăng lên nhanh chóng, đặc biệt là với sự phát triển của mạng xã hội trực tuyến nền tảng, ứng dụng truyền thông xã hội và ứng dụng thương mại. Hầu hết các hình ảnh được chia sẻ trên có các cách thể hiện khác nhau, góc độ khác nhau và độ phân giải thấp. Trong những năm gần đây, công nghệ Deep Learning sử dụng mạng nơ ron tích chập CNN đã trở thành phương pháp mạnh mẽ để phân loại hình ảnh. Nhiều nhà nghiên cứu đã chỉ ra rằng các mạng nơ ron tích chập có thể đạt được hiệu suất tốt hơn bằng cách sửa đổi các lớp mạng khác nhau của kiến trúc mạng. Hơn nữa, việc lựa chọn chức năng kích hoạt thích hợp của các tế bào nơ ron, trình tối ưu hóa và hàm mất mát ảnh hưởng trực tiếp đến hiệu suất của mạng nơ ron tích chập. Trong nghiên cứu này, em xin đề xuất 1 mô hình CNN có tốc độ xử lý nhanh, tốn ít tài nguyên là SmallerVGGNet. CNN được đề xuất này có kiến trúc mạng đơn giản với các tham số phù hợp có thể được sử dụng khi cần đào tạo nhanh với số lượng dữ liệu huấn luyện hạn chế.

Bài báo cáo này trình bày về việc huấn luyện mô hình SmallerVGGNet để phát hiện con người bằng khuôn mặt. Nội dung bài báo cáo chia làm 4 chương:

Chương 1: Cơ sở lý thuyết

Chương 2: Thiết kế mô hình SmallerVGGNet

Chương 3: Thuật toán và kết quả thực nghiệm

Chương 4: Kết luận

Mục Lục

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG

Chương 1: Cơ sở lý thuyết

1. Deep Learning là gì?

Deep learning là một phần của học máy (machine learning) giúp các mạng nơ-ron nhân tạo được đào tạo để nhận dạng các mẫu và đưa ra quyết định. Các thuật toán học sâu sử dụng dữ liệu để học và cải thiện, và nhiều tầng nơ ron nhân tạo xử lý dữ liệu để nhận dạng và phân loại các đặc trưng.

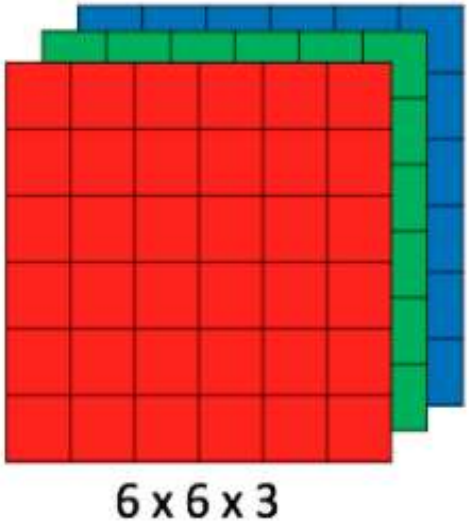
Những kĩ thuật học sâu đang trở thành xu hướng phát triển mạnh mẽ hiện nay và được áp dụng rộng rãi trong nhiều lĩnh vực khác nhau, bao gồm:

- Nhận diện ảnh, video, âm thanh và ngôn ngữ tự nhiên
- Xử lý ngôn ngữ tự nhiên
- Tự động hóa trong sản xuất và kinh doanh
- Thương mại điện tử và dữ liệu khách hàng
- Y tế và chăm sóc sức khỏe
- Xe tự hành

Các ứng dụng của học sâu có thể giúp giảm thiểu thời gian, chi phí và cải thiện hiệu quả trong nhiều ngành công nghiệp và cuộc sống. Tuy nhiên, để áp dụng thành công học sâu trong một ứng dụng, đơn vị trình bày cần có đầy đủ kiến thức về toán nâng cao (advanced mathematics) và lập trình máy tính (computer programming). Ngoài ra, học sâu yêu cầu phải có dữ liệu lớn và đa dạng để đào tạo, thường là dữ liệu khủng (big data).

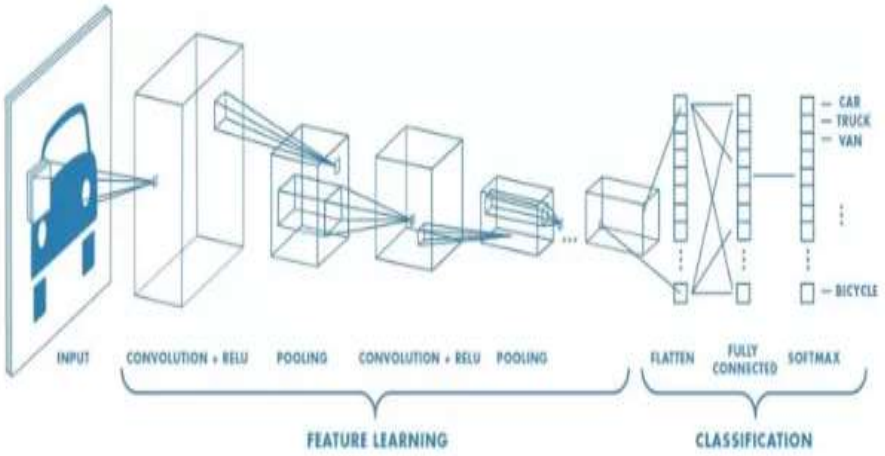
2. CNN (Convolutional Neural Network) là gì?

- Trong mạng neural, mô hình mạng neural tích chập (CNN) là 1 trong những mô hình để nhận dạng và phân loại hình ảnh. Trong đó, xác định đối tượng và nhận dạng khuôn mặt là 1 trong số những lĩnh vực mà CNN được sử dụng rộng rãi.
- CNN phân loại hình ảnh bằng cách lấy 1 hình ảnh đầu vào, xử lý và phân loại nó theo các hạng mục nhất định (Ví dụ: Chó, Mèo, Hổ, ...). Máy tính coi hình ảnh đầu vào là 1 mảng pixel và nó phụ thuộc vào độ phân giải của hình ảnh. Dựa trên độ phân giải hình ảnh, máy tính sẽ thấy H x W x D (H: Chiều cao, W: Chiều rộng, D: Độ dày).



Hình 1: Mảng ma trận RGB 6x6x3 (3 ở đây là giá trị RGB).

- Về kỹ thuật, mô hình CNN để training và kiểm tra, mỗi hình ảnh đầu vào sẽ chuyển nó qua 1 loạt các lớp tích chập với các bộ lọc (Kernals), tổng hợp lại các lớp được kết nối đầy đủ (Full Connected) và áp dụng hàm Softmax để phân loại đối tượng có giá trị xác suất giữa 0 và 1. Hình dưới đây là toàn bộ luồng CNN để xử lý hình ảnh đầu vào và phân loại các đối tượng dựa trên giá trị.



Hình 2: *Cấu trúc cơ bản của mạng nơ ron tích chập*

3. VGGNet là gì?

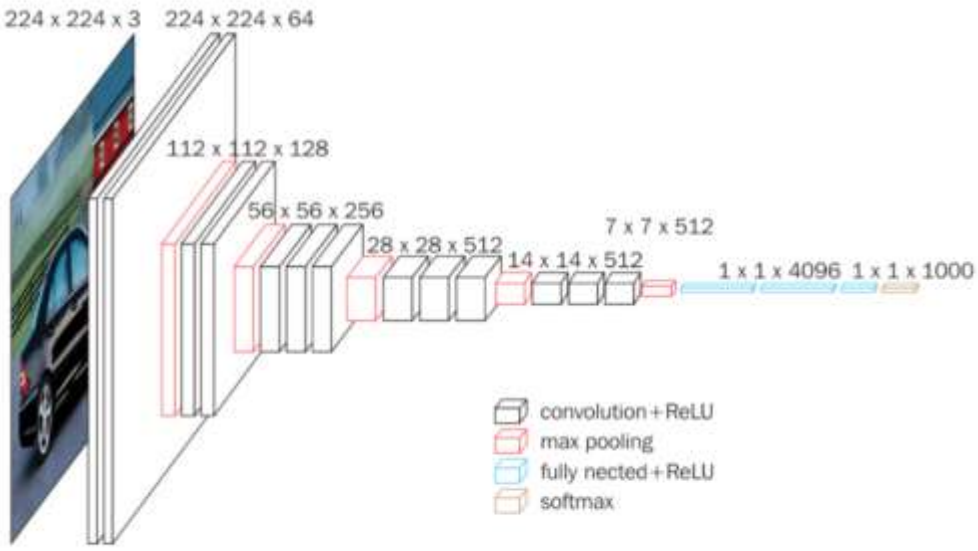
- VGGNet là một kiến trúc mạng thần kinh tích chập (CNN) được phát triển bởi nhóm nghiên cứu Visual Geometry Group (VGG) tại Đại học Oxford, Anh Quốc. Kiến trúc này đã đạt được thành tích rất cao trong cuộc thi ImageNet Large Scale Visual Recognition Challenge (ILSVRC) năm 2014. VGGNet sử dụng một chuỗi các lớp tích chập và lớp max pooling, trước khi tiếp tục với một hoặc một số lớp fully-connected layer để phân loại.

- VGGNet có thể được xây dựng với nhiều độ sâu khác nhau, tùy thuộc vào số lượng lớp và kích thước bộ lọc được sử dụng. Điều này có thể đạt được thông qua các cấu hình khác nhau, được đánh số là VGG-16 và VGG-19. VGG-16 bao gồm 16 lớp, bao gồm 13 lớp tích chập và 3 lớp fully-connected layer cuối cùng để phân loại, trong khi VGG-19 có thêm 3 lớp tích chập nữa.
- VGGNet là một trong những kiến trúc CNN phổ biến nhất và đã được sử dụng trong nhiều bài toán xử lý hình ảnh như nhận dạng đối tượng, phân tích khuôn mặt, phân loại hình ảnh y khoa và các ứng dụng khác. Kiến trúc này tạo ra một mô hình có khả năng phân loại chính xác cao và thường được sử dụng như là một mô hình cơ sở trong các bài toán phân loại hình ảnh.

4. Hoạt động của VGGNet:

VGGNet được sử dụng để giải quyết các bài toán phân loại ảnh, bao gồm nhận diện đối tượng, phân loại chủ đề và nhận dạng khuôn mặt. Đầu vào của VGGNet là một ảnh bitmap có kích thước cố định, chẳng hạn như 224 x 224 pixel.

- Quá trình hoạt động của VGGNet bắt đầu với việc truyền ảnh đầu vào qua chuỗi các lớp tích chập (convolutional layers) và lớp max pooling, tạo ra các feature map có độ sâu tăng dần tại các mức trừu tượng cao hơn. Mỗi lớp tích chập có thể được xem như một bộ phát hiện đặc trưng, nơi các bộ lọc nhỏ được tự động học để phát hiện các đặc trưng nhỏ khác nhau trong ảnh, chẳng hạn như các cạnh, đường nét và các đối tượng cục bộ.
- Sau khi xử lý qua tất cả các lớp tích chập và max pooling, các feature map sẽ được đưa vào một hoặc nhiều lớp fully connected, nơi chúng được kết hợp lại để tạo ra một vector đặc trưng cuối cùng cho ảnh đầu vào. Cuối cùng, vector đặc trưng này được đưa vào một lớp phân loại softmax để xác định xác suất của ảnh đầu vào thuộc về từng lớp phân loại khác nhau.
- Khi huấn luyện, mục tiêu là điều chỉnh trọng số của các lớp tích chập và fully connected để tối thiểu hóa sai số giữa đầu ra dự đoán và nhãn đích thực sự của ảnh. Việc này được thực hiện thông qua quá trình backpropagation, trong đó đạo hàm của sai số được tính trở lại từ lớp phân loại softmax đến các lớp trước đó của mạng, và cập nhật trọng số dựa trên các đạo hàm này thông qua thuật toán gradient descent.Sử dụng hàm kích hoạt để tìm đối số phù hợp và phân loại hình ảnh.



Hình 3: Cấu trúc VGG

Một trong những lợi ích của kiến trúc VGGNet là nó cho phép tái sử dụng các đặc trưng học được, cho phép huấn luyện và phân loại nhanh hơn đối với các bộ dữ liệu lớn. Các đặc trưng học được của VGGNet có thể được lưu vào một file riêng biệt và sử dụng lại cho các mô hình phân loại khác. Tuy nhiên, vì kiến trúc VGGNet có nhiều lớp tích chập và fully connected phức tạp, nó cần nhiều tài nguyên tính toán để huấn luyện và dự đoán. Để giải quyết vấn đề này, một số biến thể của VGGNet đã được giới thiệu phiên bản tối giản hơn như VGG Lite, MiniVGGNet, SmallerVGGNet … tốn ít tài nguyên hơn cho phép tăng độ phức tạp của mô hình để đạt được hiệu suất phân loại tốt hơn.

Chương 2: Thiết kế mô hình SmallerVGGNet

1. Khái niệm:

Smallervggnet là một kiến trúc mạng thần kinh tích chập (CNN) được sử dụng trong các bài toán xử lý hình ảnh. Kiến trúc này được phát triển dựa trên VGGNet, tuy nhiên, nó có kích thước nhỏ hơn thông qua việc giảm số lượng lớp tích chập và kích thước bộ lọc. Smallervggnet vẫn giữ nguyên cấu trúc của VGGNet, với các lớp tích chập xen kẽ với các lớp max pooling, trước khi tiếp tục với các lớp dày đặc (fully-connected layer) để phân loại.

Smallervggnet được sử dụng chủ yếu trong các bài toán nhận diện hình ảnh, bao gồm nhận dạng đối tượng, nhận dạng khuôn mặt, phân tích tín hiệu y học và các ứng dụng khác. Nhờ kích thước nhỏ hơn, smallervggnet tiêu tốn ít tài nguyên hơn, đồng thời đạt được hiệu suất tương đương hoặc tốt hơn so với các kiến trúc mạng CNN lớn hơn.

Smallervggnet hiện nay đã trở thành một trong những kiến trúc mạng CNN phổ biến và được sử dụng rộng rãi cho các bài toán nhận diện hình ảnh và phân loại.

2. Tiềm năng phát triển:

Nghiên cứu này đề cập đến vấn đề phân loại giới tính con người từ những bức ảnh chụp hoặc đang quay bằng camera . Trong xử lý ảnh, phân loại tự động giới tính con người là một chủ đề thú vị, bởi giới tính chứa đựng những thông tin rất quan trọng và phong phú về các hoạt động xã hội của con người. Giới tính là một trong những kiến thức cơ bản và rõ ràng nhất của con người, nhưng kiến thức này mở ra cơ hội thu thập thông tin ước tính trong các ứng dụng thực tế khác nhau. Trong quá trình phân loại giới tính, giới tính của cá nhân được ước tính bằng cách xác định các đặc điểm riêng biệt của nữ tính và nam tính.

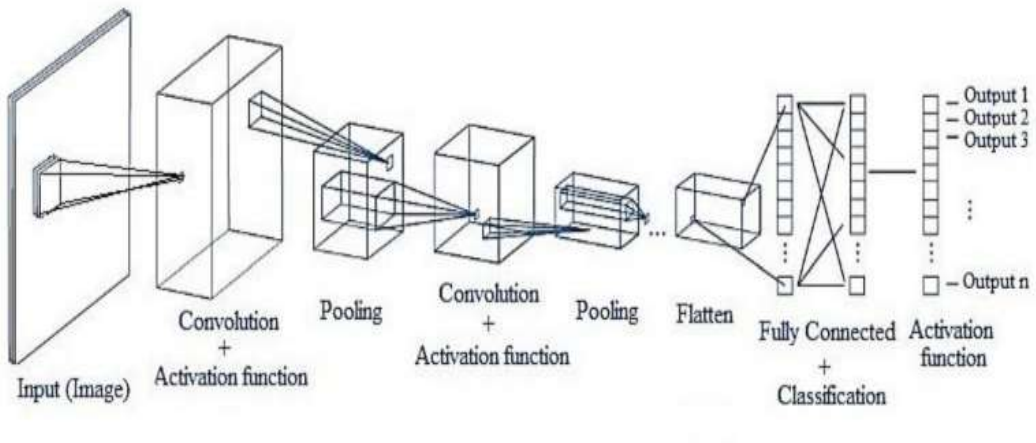
Việc phân loại giới tính có tiềm năng rất lớn, đặc biệt là trong các lĩnh vực sử dụng trí tuệ nhân tạo và xử lý hình ảnh. Ví dụ: hệ thống máy tính có chức năng phân loại giới tính có thể được sử dụng trong các lĩnh vực bao gồm tương tác giữa người và máy tính (HCI) như ứng dụng di động và trò chơi điện tử, khảo sát nhân khẩu học, ngành an ninh và giám sát, và các ứng dụng truyền thông xã hội.

3. Quá trình đào tạo mô hình:

Mô hình máy ảnh được tạo bằng cách đào tạo SmallerVGGNet từ đầu trên khoảng 2200 hình ảnh khuôn mặt (~1100 cho mỗi lớp). Vùng khuôn mặt được cắt bằng cách áp face detectiondụng cvlib trên các hình ảnh được thu thập từ Google Hình ảnh. Nó đạt được độ chính xác đào tạo khoảng 96% và độ chính xác xác thực ~90%. (20% tập dữ liệu được sử dụng để xác thực).

Mạng thần kinh nhân tạo SmallerVGGNet được đào tạo hoạt động tốt như mọi người. Nó đã áp dụng phân tích thành phần chính cho cấu trúc ba chiều và dữ liệu hình ảnh mức xám từ đầu người được quét laze một cách riêng biệt. Họ đã so sánh chất lượng của thông tin trong dữ liệu hình ảnh đầu và mức xám ba chiều. Theo kết quả thử nghiệm, họ nói rằng dữ liệu đầu ba chiều cung cấp khả năng phân loại giới tính tốt hơn so với hình ảnh mức độ xám. Kết hợp sử dụng với các bộ lọc (Gabor, Sobel) để thu nhận và xử lý trước hình ảnh để xử lý hình ảnh như độ sắc nét, độ sáng, độ bão hòa hoặc độ tương phản.

4. Thiết kế:



Hình 4: Cấu trúc CNN của SmallerVGGNet

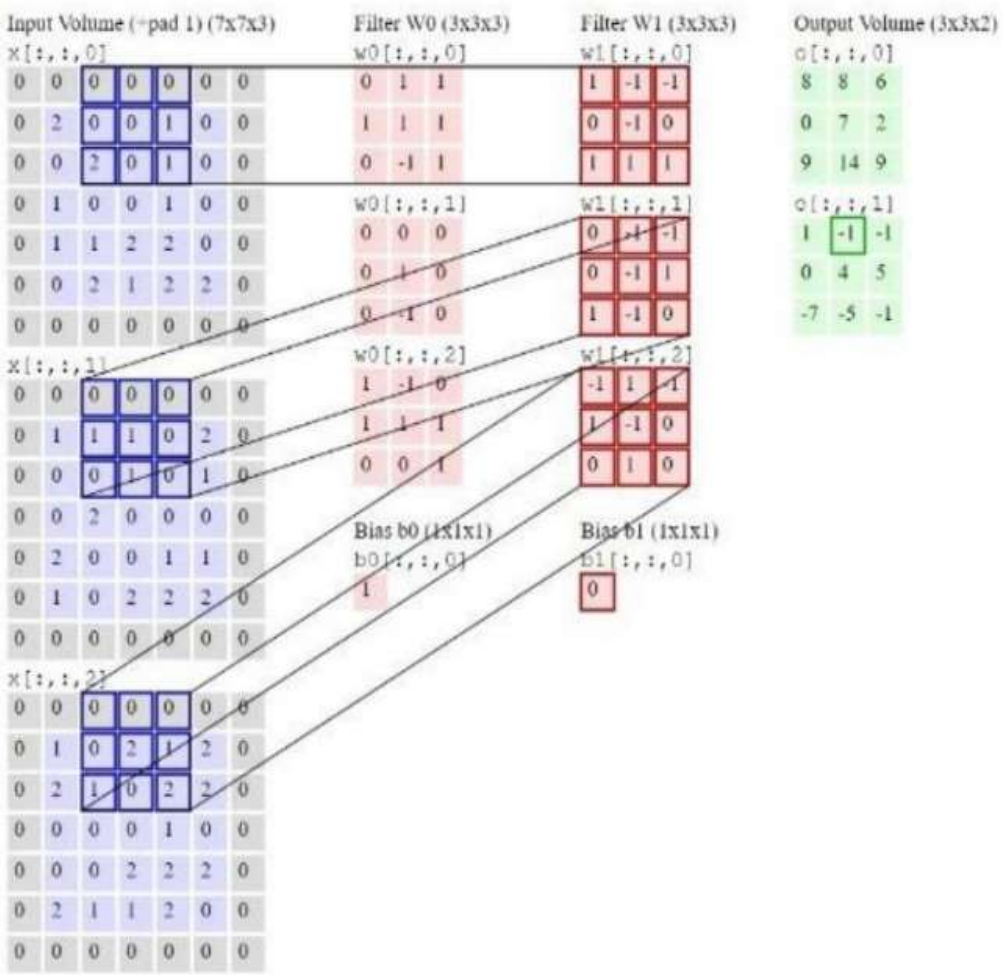
4.1 Lớp đầu vào (Input Layer):

Lớp đầu vào là lớp đầu tiên của CNN. Lớp này được đưa ra ở một kích thước nhất định. Kích thước hình ảnh trong lớp này rất quan trọng đối với sự thành công của các CNN được thiết kế. Tăng kích thước của dữ liệu hình ảnh đến có thể làm tăng

sự thành công của hệ thống cũng như quá trình đào tạo. Khi kích thước của dữ liệu hình ảnh được chọn thấp, quá trình đào tạo sẽ giảm nhưng khả năng thành công của hệ thống có thể giảm. Khi chọn kích thước dữ liệu ảnh vào lớn, sẽ làm tăng quá trình huấn luyện nhưng có thể làm tăng khả năng thành công của hệ thống được thiết kế. Do đó, sẽ rất hữu ích nếu dữ liệu hình ảnh được chọn ở kích thước phù hợp với hệ thống được thiết kế.

4.2 Lớp tích chập (Convolution Layers):

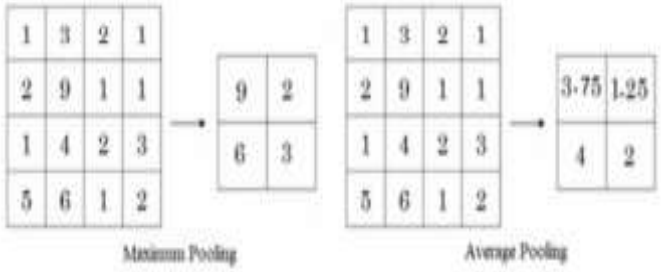
Lớp này được gọi là lớp trích xuất đặc trưng của CNN và quá trình trích xuất được thực hiện bằng cách lọc các hình ảnh ở một kích thước nhất định. Tùy thuộc vào kích thước của hình ảnh, các bộ lọc có thể có các kích thước khác nhau như 2x2, 3x3, 5x5, đầu ra của lớp này tạo ra giá trị đầu ra sau khi các chức năng kích hoạt Sigmoid, Tanh và Đơn vị tuyến tính chỉnh lưu (ReLU) được áp dụng.



Hình 5: Một ví dụ về các lớp tích chập

4.3 Lớp tổng hợp (Pooling Layers):

Trong CNN, lớp tổng hợp thường được đặt sau lớp tích chập và mục đích chính của lớp này là giảm kích thước của ma trận đầu vào cho lớp tích chập. Như trong lớp tích chập, một số bộ lọc nhất định được xác định trong lớp gộp, các bộ lọc này có thể được di chuyển theo một bước nhất định trên ảnh và giá trị của các pixel trong ảnh có thể được tính theo hai cách khác nhau, gộp trung bình và gộp tối đa [27]. Trong tổng hợp tối đa, các giá trị tối đa của các pixel trong hình ảnh được chọn và trong tổng hợp trung bình, các giá trị trung bình của các pixel trong ảnh được lấy. Có hai tham số quan trọng trong lớp tổng hợp, một trong số đó là kích thước bộ lọc và thứ hai là tham số sải chân.



Hình 6: Tổng hợp tối đa và tổng hợp trung bình với bộ lọc 2x2 trong hình ảnh đầu vào 4x4

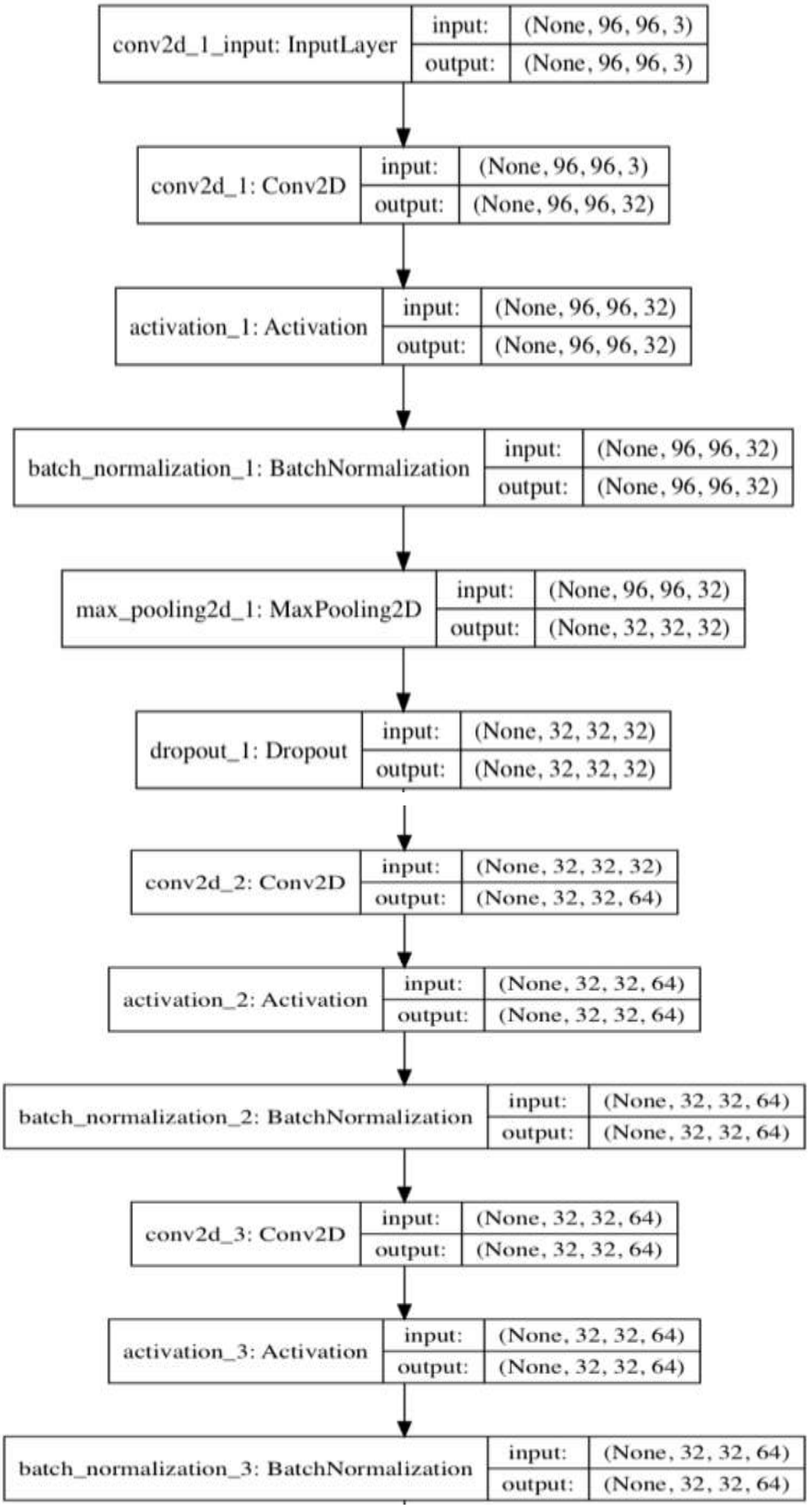
4.4 Lớp kết nối đầy đủ (Fully Connected Layer):

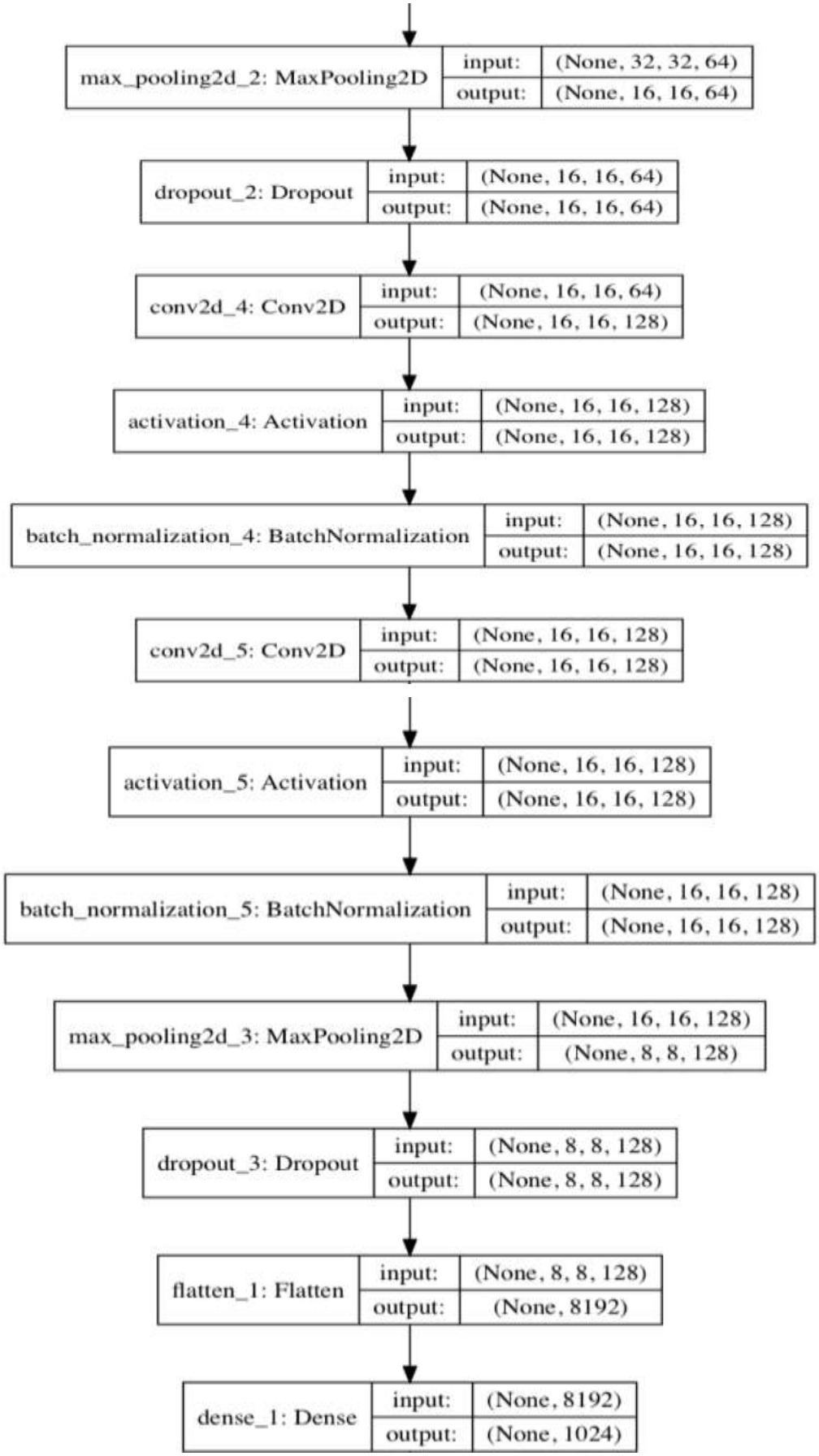
Trong kiến trúc của CNN, lớp được kết nối đầy đủ được giải quyết sau lớp tích chập và lớp tổng hợp. Lớp này được kết nối bởi tất cả các nút trong lớp chập hoặc gộp, số lượng lớp có thể khác nhau tùy theo cấu trúc của CNN. Số lượng trọng số bằng số nút trong các lớp tích chập nhân với số nút trong các lớp được kết nối.

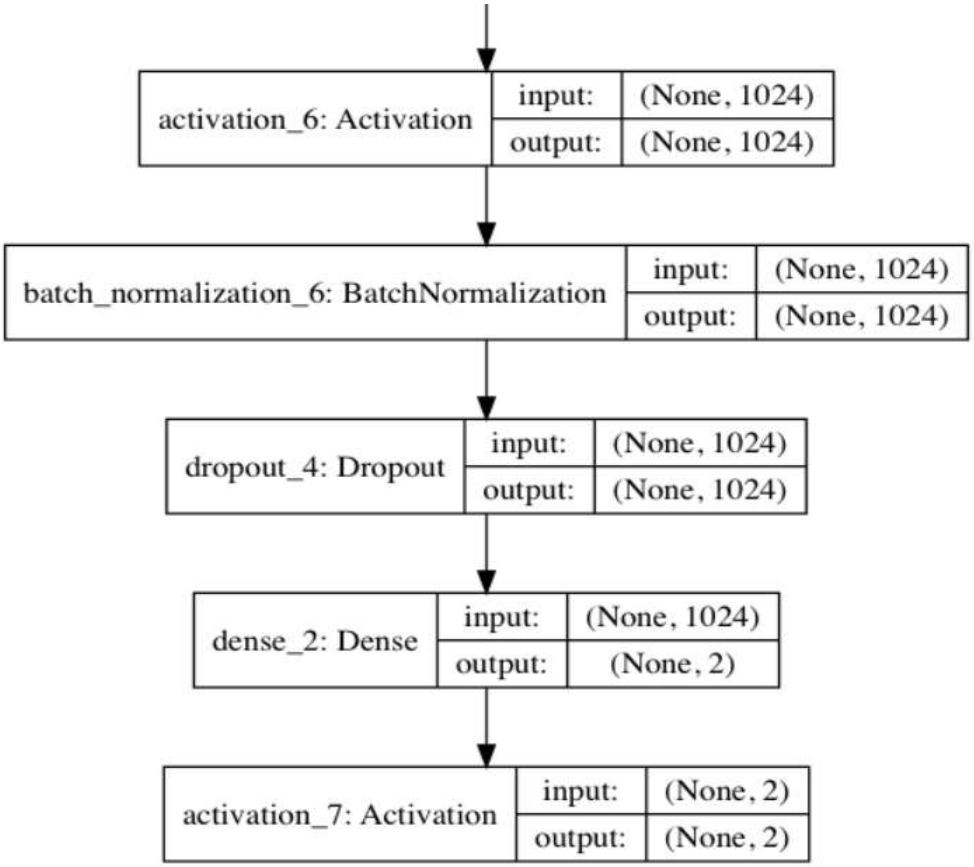
4.5 Lớp phân loại (Classification Layer):

Lớp CNN này xuất hiện sau lớp được kết nối đầy đủ, quá trình học của CNN diễn ra trong lớp này. Lớp này lấy đầu vào từ lớp được kết nối đầy đủ và đầu ra của nó bằng số đối tượng được phân loại.

5. SmallerVGGNet để phân loại giới tính:


conv2d_1_input: InputLayer
input: (None, 96, 96, 3)
output: (None, 96, 96, 3)
conv2d_1: Conv2D
input: (None, 96, 96, 3)
output: (None, 96, 96, 32)
activation_1: Activation
input: (None, 96, 96, 32)
output: (None, 96, 96, 32)
batch_normalization_1: BatchNormalization
input: (None, 96, 96, 32)
output: (None, 96, 96, 32)
max_pooling2d_1: MaxPooling2D
input: (None, 96, 96, 32)
output: (None, 32, 32, 32)
dropout_1: Dropout
input: (None, 32, 32, 32)
output: (None, 32, 32, 32)
conv2d_2: Conv2D
input: (None, 32, 32, 32)
output: (None, 32, 32, 64)
activation_2: Activation
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
batch_normalization_2: BatchNormalization
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
conv2d_3: Conv2D
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
activation_3: Activation
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)
batch_normalization_3: BatchNormalization
input: (None, 32, 32, 64)
output: (None, 32, 32, 64)


max_pooling2d_2: MaxPooling2D
input: (None, 32, 32, 64)
output: (None, 16, 16, 64)
dropout_2: Dropout
input: (None, 16, 16, 64)
output: (None, 16, 16, 64)
conv2d_4: Conv2D
input: (None, 16, 16, 64)
output: (None, 16, 16, 128)
activation_4: Activation
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
batch_normalization_4: BatchNormalization
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
conv2d_5: Conv2D
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
activation_5: Activation
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
batch_normalization_5: BatchNormalization
input: (None, 16, 16, 128)
output: (None, 16, 16, 128)
max_pooling2d_3: MaxPooling2D
input: (None, 16, 16, 128)
output: (None, 8, 8, 128)
dropout_3: Dropout
input: (None, 8, 8, 128)
output: (None, 8, 8, 128)
flatten_1: Flatten
input: (None, 8, 8, 128)
output: (None, 8192)
dense_1: Dense
input: (None, 8192)
output: (None, 1024)



Hình 7: Kiến trúc của SmallerVGGNet để phân loại giới tính

Tất cả các lớp tích chập và hai lớp được kết nối đầy đủ đầu tiên được kích hoạt bởi chức năng kích hoạt có tên ReLU đóng góp rất nhiều vào thành công gần đây của mạng nơ-ron sâu.

- ReLU là một chức năng kích hoạt không có tham số có thể được định nghĩa như sau:

$$f(x) = \begin{cases} x, nếu\ x > 0 \\ 0, nếu\ x \leq 0 \end{cases} \quad (1)$$

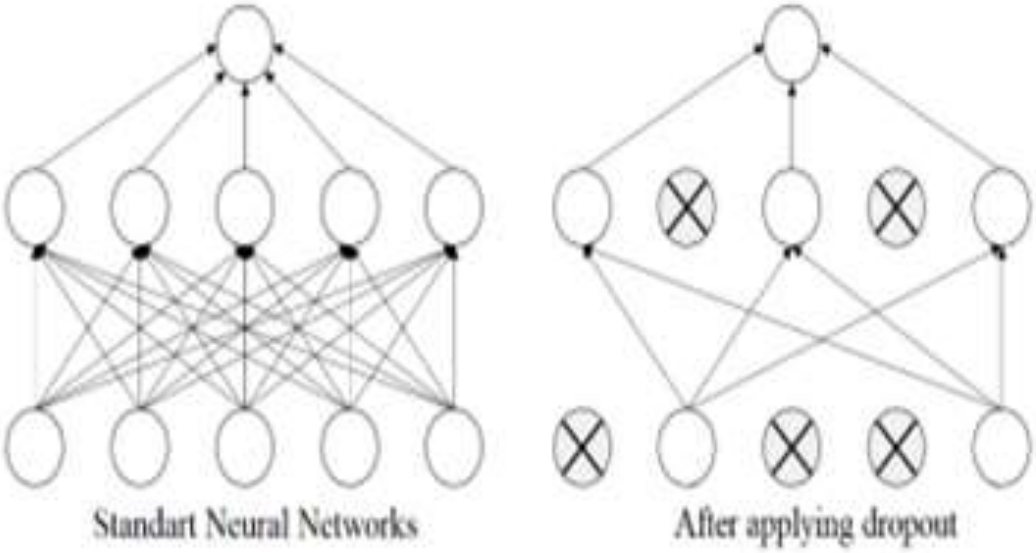
- Chức năng kích hoạt sigmoid được sử dụng cho lớp được kết nối đầy đủ cuối cùng. Hàm kích hoạt sigmoid nhận một giá trị thực và xuất giá trị trong khoảng 0 và 1 có thể được định nghĩa như sau:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

- Crossentropy nhị phân được sử dụng làm hàm mất mát có thể được định nghĩa như sau:

$$crossentropy(t, o) = -(t * \log(o) + (1 - t) * \log(1 - o)) \qquad (3)$$

Bỏ học (Dropout a technique) là một kỹ thuật được sử dụng để cải thiện sự phù hợp quá mức trên các mạng nơ-ron, bỏ học ngắn gọn đề cập đến việc bỏ qua các nơ-ron trong giai đoạn huấn luyện một số nơ-ron nhất định được chọn ngẫu nhiên. Tỷ lệ bỏ học 0,25 được sử dụng sau lớp phân loại dense_2.



Hình 8: Các lớp bị bỏ học

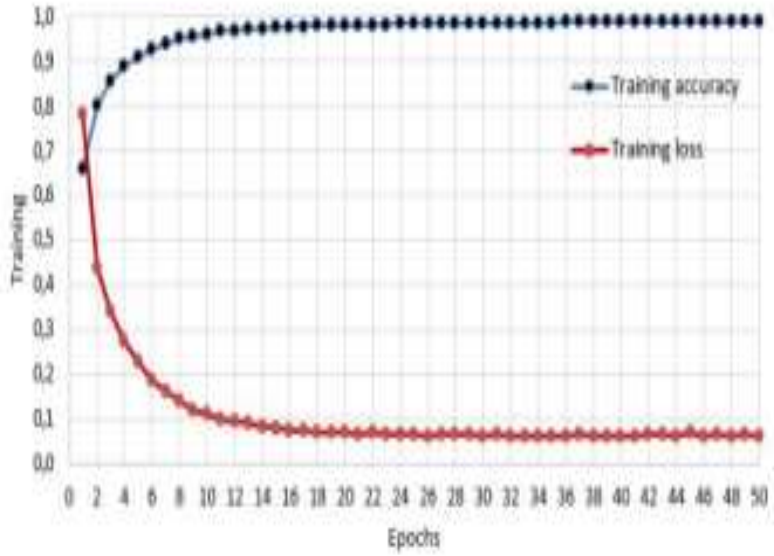
Chương 3: Thuật toán và kết quả thực nghiệm

1. Thuật toán:

Tham khảo code chạy với Python 3.10:

https://github.com/arunponnusamy/gender-detection-keras

2. Kết quả thực nghiệm giới tính:



Hình 9: *Sự hội tụ của độ chính xác phân loại và giá trị tổn thất phân loại trong giai đoạn đào*

Khi nhìn vào hình trên, có vẻ như CNN được đề xuất thường xuyên giảm tổn thất phân loại trong giai đoạn đào tạo. Mất phân loại của các CNN được đề xuất được khởi tạo ở mức 0,78, nó giảm liên tục từ kỷ nguyên đầu tiên và giảm xuống 0,06 trong các kỷ nguyên cuối cùng. Độ chính xác phân loại thay đổi tỷ lệ nghịch với sự mất phân loại trong giai đoạn huấn luyện. Độ chính xác phân loại của các CNN đề xuất ban đầu là 65,6%, nó tăng liên tục từ kỷ nguyên đầu tiên và đạt 98,8% ở các kỷ nguyên cuối cùng. Nhìn chung, con số hội tụ của CNN cho thấy hệ thống được thiết kế để phân loại giới tính được đào tạo tốt mà không có quá phù hợp. Các CNN đã đào tạo được thử nghiệm bằng hình ảnh không có nhãn giới tính trong giai đoạn thử nghiệm. Từ kết quả thử nghiệm thử nghiệm, ta thu được ma trận nhầm lẫn đưa ra:

		Predicted	
		Male	Female
Actual	Male	7033	1087
	Female	814	8558

Bảng 1:Ma trận nhầm lẫn của giai đoạn thử nghiệm

Đối với các vấn đề phân loại, độ chính xác, độ nhạy và độ đặc hiệu là các biện pháp đánh giá quan trọng có thể được tính toán như dưới đây bằng cách sử dụng ma trận nhầm .

Độ đúng = $TP + TN/TP + TN + FP + FN$ (4)

Độ chụm = $TP/TP + FP$ (5)

Độ đặc hiệu = $TN/TN + FP$ (6)

Trong đó:

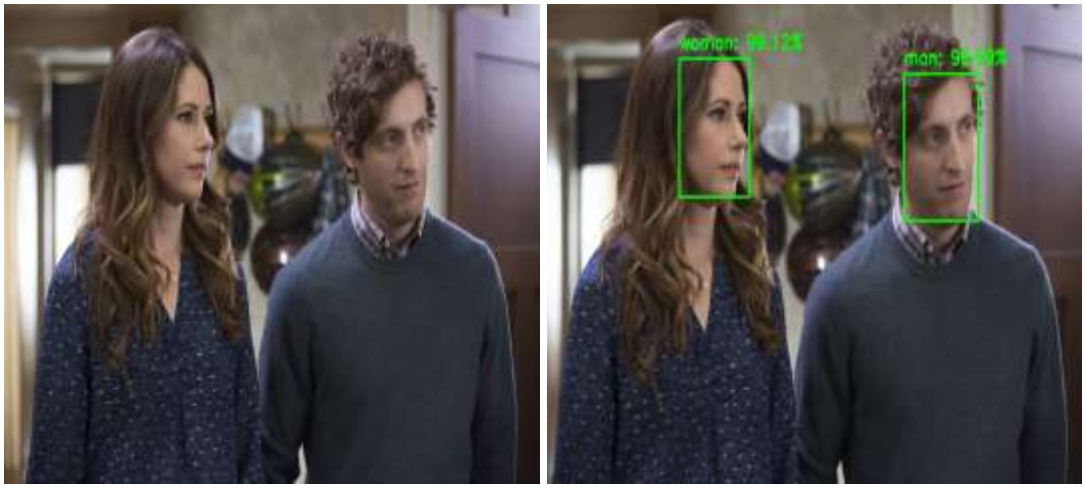
- Dương tính giả (TP) là số nam được phân loại đúng
- Âm tính đúng (TN) là số nữ được phân loại đúng
- Dương tính giả (FP) là số nữ sai phân loại là nam
- Âm tính giả (FN) là số nam bị phân loại sai là nữ

Theo các phương trình 4-6, độ đúng, độ chụm và độ đặc hiệu các biện pháp đánh giá được tính toán 89,13%, 89,14% và 88,8%, tương ứng. Khi thước đo độ chính xác được coi là CNN được đề xuất đã dự đoán chính xác phần lớn giới tính từ hình ảnh. Ngoài ra, các giá trị của độ chính xác và độ đặc hiệu dường như gần với giá trị của độ đúng.

Điều này rất quan trọng đối với sự mạnh mẽ của các CNN được đề xuất bởi vì độ chính xác cho thấy cách các CNN được đề xuất phân loại chính xác nam giới trường hợp và tính đặc hiệu chỉ ra cách các CNN được đề xuất phân loại chính xác các trường hợp nữ. Độ chính xác và độ đặc hiệu kết quả cho thấy các CNN được đề xuất đã dự đoán các trường hợp nam giới tốt hơn một chút so với trường hợp nữ.

3. Chạy mô phỏng:

- Sau khi git clone file code và sửa code trên visual studio code với python 3.10.
- Tải các gói thư viên python:
  - ✓ numpy
  - ✓ opencv-python
  - ✓ tensorflow
  - ✓ keras
  - ✓ requests
  - ✓ progressbar
  - ✓ cvlib
- Với đầu vào là hình ảnh: python detect_gender.py -i <input_image>

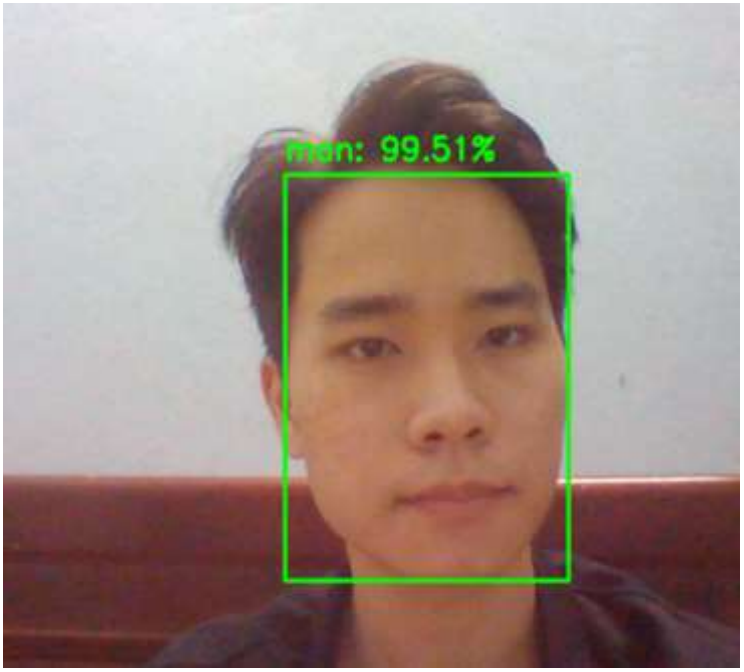


Hình 10: Test lần 1 có 2 bước : bước 1 là 168ms/step, bước 2 là 25ms/step.

Hình 11: Test lần 2 có 2 bước : bước 1 là 179ms/step, bước 2 là 29ms/step.

Hình 12: Test lần 3 có 4 bước : bước 1 là 228ms/step, bước 2 là 39ms/step, bước 3 là 32ms/step, bước 4 là 30ms/step.

- Với webcam: python detect_gender_webcam.py



Hình 13: Test lần 4 với mỗi bước từ 24ms/step đến 45ms/step.

⇨ Nhận xét :

✓ Ưu điểm : Tốc độ sử lý khá nhanh, tốn ít tài nguyên.

✓ Nhược điểm: Độ chính xác càng thấp khi độ tuổi càng trẻ, lọc nhiễu kém và do kỹ thuật bỏ học (Dropout a technique) nên còn lúc không xác định khuôn mặt.

Chương 4: Kết luận

Trong nghiên cứu này, em đã đề xuất một cấu trúc CNN mới để phân loại giới tính từ các hình ảnh khuôn mặt đó là SmallerVGGNet. Tài liệu này là nhiều nghiên cứu về phân loại giới tính đã sử dụng ảnh có độ phân giải cao được chuẩn bị trong môi trường phòng thí nghiệm. Nhưng trên mạng xã hội, hầu hết các hình ảnh còn chụp quá mờ do bị nhiễu và chất lượng webcam còn thấp.

Do đó, để đạt được thành công như mong muốn của các CNN được đề xuất, thì bộ dữ liệu đối tượng bao gồm hình ảnh khuôn mặt được chụp bằng camera với các góc độ và biểu cảm khác nhau thuộc nhiều độ tuổi và giới tính khác nhau đã được sử dụng phải có độ sắc nét. Các CNN đề xuất được huấn luyện trong thời gian ngắn với số lượng ảnh nhỏ và đạt độ chính xác phân loại tốt theo nghiên cứu thực nghiệm.

Đối với công việc trong tương lai, các CNN được đề xuất sẽ được sử dụng trong các ứng dụng di động, máy camera… bao gồm: phân loại giới tính và thống kê giới tính người trên mạng xã hội. Ngoài ra, các CNN mới sẽ được phát triển để dự đoán tuổi hoặc danh tính bằng cách sử dụng kiến thức và kinh nghiệm thu được từ nghiên cứu này.

TÀI LIỆU THAM KHẢO

[1] **Code:** https://github.com/arunponnusamy/gender-detection-keras

[2] **Báo cáo:** https://www.researchgate.net/publication/328405250_Human_Gender_Prediction_on_Facial_Mobil_Images_using_Convolutional_Neural_Networks

[3] https://viblo.asia/p/deep-learning-tim-hieu-ve-mang-tich-chap-cnn-maGK73bOKj2

[4] https://www.kaggle.com/code/blurredmachine/vggnet-16-architecture-a-complete-guide

[5] https://longvan.net/deep-learning-la-gi-ung-dung-cua-deep-learning.html