BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG THƯƠNG TP.HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

-----o0o----

TIỂU LUẬN HỌC PHẦN: DEEP LEARNING

TÊN ĐỀ TÀI: XÂY DỰNG MÔ HÌNH NHẬN DIỆN CẢM XÚC KHUÔN MẶT VÀ GIỚI TÍNH THÔNG QUA CAMERA (CNN)

NHÓM: 5

Thành phố Hồ Chí Minh, 26 tháng 05 năm 2025

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG THƯƠNG TP.HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

-----o0o----

TÊN ĐỀ TÀI: XÂY DỰNG MÔ HÌNH NHẬN DIỆN CẢM XÚC KHUÔN MẶT VÀ GIỚI TÍNH THÔNG QUA CAMERA (CNN)

Nhóm: 5 Trưởng nhóm: Nguyễn Chí Tài Thành viên: 1. Trần Công Minh 2. Lê Đức Trung 3. Tạ Nguyên Vũ	**Giảng viên bộ môn:** **Nguyễn Thị Huyền Trang**

Thành phố Hồ Chí Minh, 02 tháng 06 năm 2025

LỜI CAM ĐOAN

Chúng em xin cam đoan đề tài báo cáo: **Xây dựng mô hình nhận diện cảm xúc khuôn mặt và giới tính thông qua camera (CNN)** do nhóm 5 nghiên cứu và thực hiện. Chúng em đã kiểm tra dữ liệu theo quy định hiện hành.

Kết quả bài làm của đề tài **Xây dựng mô hình nhận diện cảm xúc khuôn mặt và giới tính thông qua camera (CNN)** là trung thực và không sao chép từ bất kỳ bài tập của nhóm khác. Các tài liệu được sử dụng trong tiểu luận có nguồn gốc, xuất xứ rõ ràng.

(Ký và ghi rõ họ tên)

BẢNG PHÂN CÔNG CÔNG VIỆC

MSSV	Họ tên	Công việc	Đánh giá
2001222641	Trần Công Minh	Word Power Point	100%
2001225676	Lê Đức Trung	Word PowerPoint	100%
2001224227	Nguyễn Chí Tài	Code ứng dụng	100%
2001225416	Tạ Nguyên Vũ	Code ứng dụng	100%

MỤC LỤC

DANH MỤC HÌNH ẢNH

LỜI MỞ ĐẦU

Trong thời đại công nghệ số phát triển mạnh mẽ, trí tuệ nhân tạo (AI) và học máy ngày càng được ứng dụng rộng rãi trong phân tích và xử lý dữ liệu hình ảnh. Một trong những ứng dụng quan trọng là nhận diện cảm xúc và giới tính từ khuôn mặt, giúp hệ thống tự động xác định trạng thái cảm xúc cũng như đặc điểm giới tính của một cá nhân qua hình ảnh hoặc video. Công nghệ này mang lại nhiều lợi ích thiết thực như phân tích phản ứng khách hàng, tối ưu hóa trải nghiệm cá nhân hóa, hỗ trợ y tế tâm lý và cải thiện chất lượng giảng dạy thông qua hiểu rõ cảm xúc của học sinh.

Mạng nơ-ron tích chập (CNN) là công nghệ chủ chốt trong nhận diện hình ảnh nhờ khả năng tự động học và trích xuất đặc trưng phức tạp từ dữ liệu. Với các lớp tích chập phát hiện chi tiết khuôn mặt, biểu cảm mắt, miệng và các lớp pooling giúp tối ưu hóa dữ liệu, CNN cho phép nhận diện chính xác cảm xúc như vui, buồn, tức giận, ngạc nhiên, đồng thời xác định giới tính hiệu quả. Nhờ khả năng học sâu, CNN phân tích các yếu tố vi mô trên khuôn mặt, cải thiện độ chính xác trong nhận diện cảm xúc và giới tính.

Bài viết này sẽ đi sâu vào cơ chế hoạt động của CNN trong nhận diện cảm xúc và giới tính, cùng những ứng dụng thực tiễn của công nghệ này. Chúng em cũng sẽ phân tích ưu điểm của CNN trong xử lý hình ảnh, đồng thời thảo luận về những thách thức và hạn chế khi áp dụng vào thực tế, từ đó đề xuất hướng cải thiện hiệu suất trong tương lai.

CHƯƠNG 1. TỔNG QUAN

1.1. Đặt vấn đề

Trong thời đại công nghệ số phát triển nhanh chóng, trí tuệ nhân tạo (AI) và học máy (Machine Learning) ngày càng đóng vai trò quan trọng trong nhiều lĩnh vực, đặc biệt là xử lý và phân tích dữ liệu hình ảnh. Một trong những ứng dụng đáng chú ý là nhận diện cảm xúc và giới tính từ khuôn mặt, giúp hệ thống tự động phân tích và xác định trạng thái cảm xúc cũng như đặc điểm giới tính của một cá nhân từ hình ảnh hoặc video.

Việc phát triển các mô hình nhận diện này không chỉ mang lại lợi ích thiết thực trong việc nâng cao trải nghiệm người dùng, như phân tích phản ứng khách hàng theo thời gian thực hay tối ưu hóa giao diện cá nhân hóa trên các nền tảng số, mà còn mở ra nhiều tiềm năng trong bảo mật sinh trắc học, hỗ trợ tâm lý trong y tế và nâng cao chất lượng giáo dục. Tuy nhiên, để đạt được độ chính xác cao, hệ thống cần đến những thuật toán mạnh mẽ có khả năng học sâu và trích xuất đặc trưng hình ảnh một cách hiệu quả.

Mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) là một trong những công nghệ tiên tiến giúp giải quyết bài toán này. Nhờ khả năng tự động học và trích xuất các đặc trưng quan trọng từ khuôn mặt, CNN đã chứng minh được hiệu quả vượt trội trong các ứng dụng nhận diện hình ảnh. Tuy nhiên, vẫn còn nhiều thách thức như xử lý dữ liệu đa dạng, cải thiện độ chính xác trong điều kiện ánh sáng và góc nhìn khác nhau.

Từ những lý do trên, bài viết này sẽ tập trung tìm hiểu cách CNN hoạt động trong bài toán nhận diện cảm xúc và giới tính, phân tích các ưu điểm, hạn chế cũng như đề xuất các hướng cải tiến trong tương lai.

1.2. Mục tiêu nghiên cứu

Nghiên cứu này nhằm phát triển một phương pháp dự đoán đồng thời giới tính và cảm xúc dựa trên đặc điểm khuôn mặt bằng cách sử dụng mạng nơ-ron tích chập (CNN). Mục tiêu chính là xây dựng và huấn luyện các mô hình CNN có khả năng

nhận diện chính xác giới tính (Nam hoặc Nữ) và năm trạng thái cảm xúc (Bình thường, Vui vẻ, Buồn, Tức giận, Ngạc nhiên) từ hình ảnh khuôn mặt trong thời gian thực. Nghiên cứu hướng đến việc đạt được độ chính xác cao trong dự đoán, đồng thời đảm bảo hiệu suất hoạt động phù hợp với các ứng dụng thực tiễn như nhận diện khuôn mặt qua camera. Ngoài ra, nghiên cứu cũng đặt nền tảng cho việc cải tiến và thương mại hóa mô hình trong tương lai bằng cách tận dụng các kỹ thuật học sâu tiên tiến

1.3. Ý nghĩa của nghiên cứu

Nghiên cứu này mang ý nghĩa quan trọng cả về mặt khoa học lẫn thực tiễn. Cụ thể:

- **Về mặt khoa học**: Ứng dụng mạng nơ-ron tích chập (CNN) để dự đoán đồng thời giới tính và cảm xúc, khẳng định tiềm năng của trí tuệ nhân tạo và học sâu trong phân tích đặc điểm tâm lý con người. Mở ra hướng nghiên cứu mới kết hợp nhiều đặc điểm nhận diện trong một mô hình duy nhất.
- **Về mặt thực tiễn**: Phát triển mô hình thời gian thực ứng dụng trong đánh giá chất lượng dịch vụ, phân tích hành vi khách hàng, hỗ trợ tuyển dụng và nghiên cứu tâm lý xã hội. Chứng minh hiệu quả của CNN trong nhận diện khuôn mặt, tạo cơ sở cho các ứng dụng thương mại hóa. Góp phần nâng cao trải nghiệm người dùng và hiệu quả hoạt động trong nhiều lĩnh vực công nghiệp.

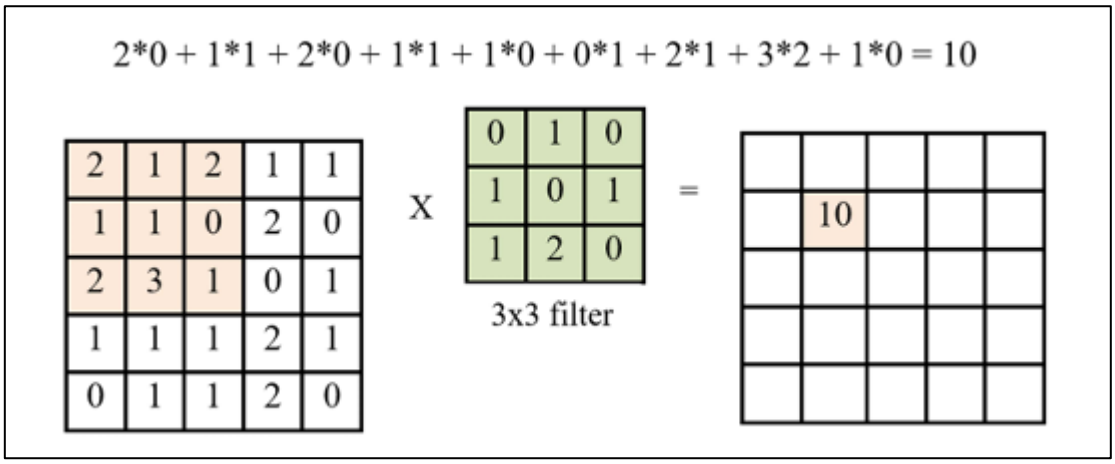
Tóm lại, nghiên cứu không chỉ mang lại giá trị tức thời trong việc giải quyết bài toán nhận diện khuôn mặt mà còn đặt nền móng cho các cải tiến công nghệ trong tương lai. Tạo tiền đề cho việc tích hợp AI vào đời sống, thúc đẩy sự phát triển bền vững của các giải pháp thông minh dựa trên dữ liệu hình ảnh.

CHƯƠNG 2. CƠ SỞ LÍ THUYẾT

2.1. Giới thiệu về mạng nơ-ron tích chập

2.1.1. Tích chập

Tích chập (convolution) ban đầu được sử dụng trong lĩnh vực xử lý tín hiệu số (signal processing). Dựa trên nguyên lý biến đổi thông tin, các nhà nghiên cứu đã ứng dụng kỹ thuật này vào xử lý ảnh và video số. Để dễ hình dung, chúng ta có thể hình dung tích chập như một ô cửa sổ trượt (sliding window) di chuyển trên một ma trận.



Hình 1. Tính tích chập

Ma trận bên trái biểu diễn một ảnh xám, với mỗi giá trị trong ma trận tương ứng với một điểm ảnh (pixel) có giá trị từ 0 đến 255. Sliding window, còn được gọi là kernel, filter hay feature detector, là một ma trận filter kích thước 3x3. Ta nhân từng phần tử của ma trận này với các phần tử tương ứng trong ma trận ảnh ban đầu. Giá trị đầu ra là tổng của các phép nhân này. Kết quả của phép tích chập chính là một ma trận mới, tạo ra từ việc trượt ma trận filter và thực hiện phép tích chập trên toàn bộ ma trận ảnh bên trái.

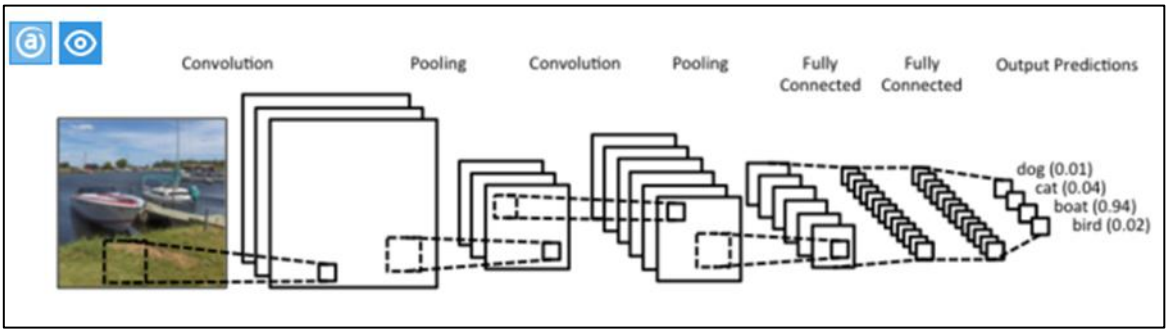
2.1.2. Mô hình mạng nơ-ron tích chập

Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) là một loại mạng nơ-ron chuyên biệt, được thiết kế để tự động học và trích xuất các đặc trưng từ

dữ liệu đầu vào một cách hiệu quả, đặc biệt là trong xử lý và phân loại ảnh. Khác với mạng nơ-ron truyền thẳng (feedforward neural network), nơi mà mỗi lớp kết nối đầy đủ với nhau thông qua trọng số, CNN sử dụng cơ chế tích chập (convolution) để tạo ra các kết nối cục bộ. Mỗi lớp trong CNN gồm nhiều bộ lọc (filters), giúp phát hiện các đặc trưng cục bộ của ảnh bằng cách áp dụng phép tích chập lên các vùng ảnh nhỏ. Kết quả tích chập này tạo ra một ma trận đặc trưng, đại diện cho các đặc trưng học được từ ảnh ở lớp trước đó.

Trong quá trình huấn luyện, CNN tự động học các tham số cho từng bộ lọc để trích xuất đặc trưng từ các mức độ chi tiết như đường biên (edges), góc (corners) đến các đặc trưng trừu tượng hơn như hình dạng (shapes) hoặc cấu trúc tổng quát. Các lớp tích chập được kết hợp với các lớp pooling (lớp gộp) để giảm kích thước của đặc trưng, loại bỏ nhiễu và tạo tính bất biến với các phép dịch chuyển, quay và co giãn của đối tượng trong ảnh. Pooling layer này giúp mô hình có thể nhận diện đối tượng ở nhiều vị trí khác nhau trong ảnh, đồng thời giảm thiểu hiện tượng quá khớp bằng cách làm cho mô hình đơn giản hơn.

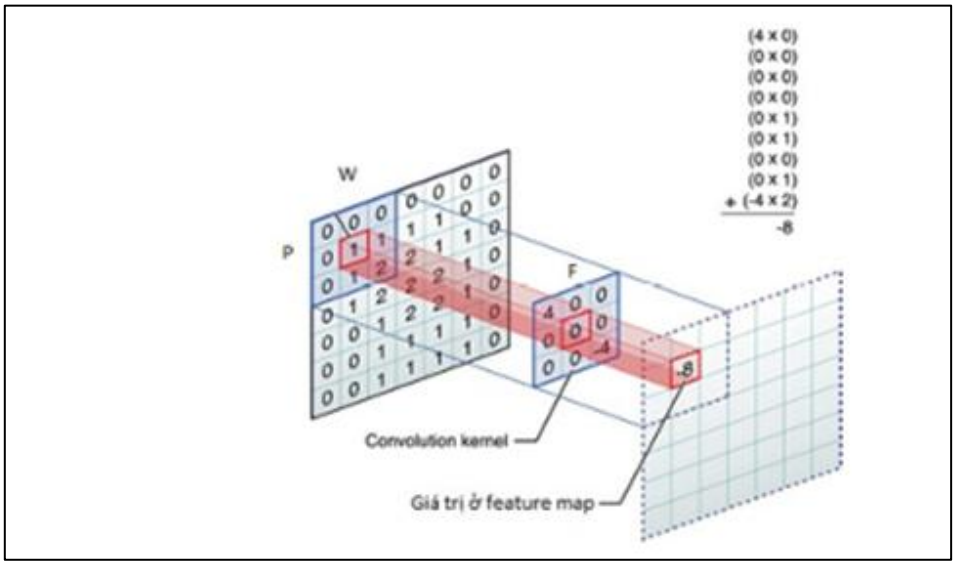
Mỗi nơ-ron ở lớp tiếp theo chỉ phụ thuộc vào một vùng cục bộ của nơ-ron ở lớp trước đó, và nhờ đó CNN có thể tạo ra các cấp độ biểu diễn từ chi tiết cơ bản đến trừu tượng. Điều này giúp CNN có tính kết hợp cục bộ và có thể học được các đặc trưng từ mức độ thấp đến cao. Đến lớp cuối cùng, các đặc trưng này được tổng hợp trong một lớp kết nối đầy đủ để thực hiện nhiệm vụ phân loại. Chính nhờ sự kết hợp của các lớp tích chập, pooling và kết nối đầy đủ mà CNN trở thành mô hình lý tưởng cho các tác vụ xử lý ảnh phức tạp với độ chính xác cao.



Hình 2. Mô hình mạng nơ-ron tích chập

2.2. Convolutional Layer

Lớp tích chập là thành phần chính trong mạng nơ-ron tích chập (CNN), thực hiện việc trích xuất đặc trưng từ dữ liệu đầu vào, chẳng hạn như các cạnh, góc và chi tiết cấu trúc của ảnh. Khi áp dụng phép tích chập, lớp này sử dụng một ma trận nhỏ gọi là **filter** (hay **kernel**) và di chuyển (trượt) nó trên toàn bộ ảnh đầu vào để tạo ra một bản đồ đặc trưng (feature map).



Hình 3. Convolutional Layer

Giả sử chúng ta có một bức ảnh xám kích thước 32x32 pixel. Khi áp dụng lớp tích chập (Convolutional Layer) lên ảnh này với một bộ lọc (filter) kích thước 3x3, chúng ta sẽ cần thêm padding để giữ nguyên kích thước đầu ra. Padding bổ sung một lớp pixel giá trị 0 xung quanh ảnh gốc, cho phép bộ lọc "quét" toàn bộ vùng biên của ảnh mà không làm giảm kích thước đầu ra. Trong ví dụ này, padding được đặt là 1 và stride là 1 (tức là bộ lọc di chuyển từng pixel một trên ảnh).

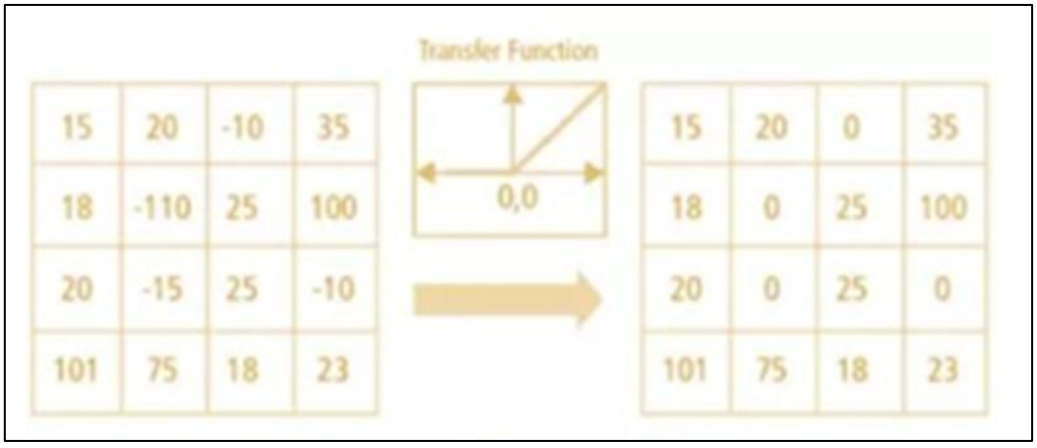
Sau khi áp dụng bộ lọc 3x3 với padding 1 và stride 1, ảnh đầu ra sẽ có kích thước **32x32** giống với ảnh đầu vào. Nếu lớp này có 10 bộ lọc khác nhau, mỗi bộ lọc sẽ tạo ra một bản đồ đặc trưng 32x32 riêng biệt, cho ra tổng cộng 10 bản đồ đặc trưng, mỗi bản đồ phản ánh một đặc trưng khác nhau từ ảnh. Các trọng số trong bộ lọc ban đầu sẽ được khởi tạo ngẫu nhiên và sẽ được mô hình tự động điều chỉnh trong quá trình huấn luyện để tối ưu hóa việc trích xuất đặc trưng từ ảnh.

Việc kết hợp nhiều bản đồ đặc trưng này giúp lớp tích chập trích xuất thông tin từ ảnh ở các cấp độ chi tiết khác nhau, từ đường viền, góc cạnh đến các đặc điểm phức tạp hơn, và gửi chúng đến lớp kế tiếp trong mô hình để tiếp tục phân tích.

2.3. Rectified Linear Unit (ReLU) Layer

Lớp **Rectified Linear Unit (ReLU)** thường được cài đặt ngay sau lớp **Convolution** trong các mạng nơ-ron tích chập (CNN). Lớp này sử dụng hàm kích hoạt f(x)=max⁡(0,x)f(x) = \max(0, x)f(x)=max(0,x), nghĩa là tất cả các giá trị âm trong kết quả của phép tính tích chập sẽ được thay thành giá trị 0, còn các giá trị dương sẽ giữ nguyên. Mục tiêu chính của việc sử dụng ReLU là tạo ra tính phi tuyến cho mô hình.

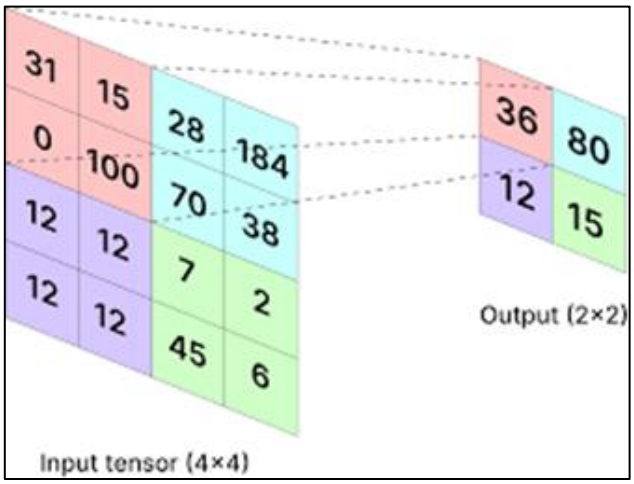
Trong các mạng nơ-ron truyền thẳng (feedforward), nếu chỉ sử dụng các phép biến đổi tuyến tính, mô hình sẽ không thể học được các mối quan hệ phức tạp giữa các đặc trưng trong dữ liệu. Chính vì vậy, việc sử dụng các hàm kích hoạt phi tuyến như **ReLU**, **sigmoid**, hay **tanh** là rất quan trọng. Tuy nhiên, ReLU lại có ưu điểm là dễ cài đặt, tính toán nhanh và hiệu quả.



Hình 4. ReLu Layer

Hàm f(x)=max⁡(0,x)f(x) = \max(0, x)f(x)=max(0,x) rất đơn giản nhưng có tác dụng quan trọng trong việc tạo tính phi tuyến cho mạng nơ-ron, giúp mô hình có khả năng học và hiểu được các đặc trưng phức tạp của dữ liệu. Chính vì thế, ReLU là một trong những hàm kích hoạt phổ biến nhất trong các mô hình học sâu hiện nay.

2.4. Pooling Layer



Hình 5. Pooling Layer

Lớp **Pooling** sử dụng một cửa sổ trượt (window) để quét qua toàn bộ ảnh dữ liệu, và mỗi lần trượt, cửa sổ sẽ di chuyển theo một bước nhất định (stride). Khác với lớp **Convolution** thực hiện phép tính tích chập, lớp **Pooling** không tính toán tích mà thay vào đó, nó thực hiện phép lấy mẫu (subsampling). Mỗi khi cửa sổ trượt qua một vùng của ảnh, lớp **Pooling** sẽ chọn một giá trị đại diện cho vùng đó, thường là giá trị lớn nhất, nhỏ nhất hoặc trung bình. Các phương pháp phổ biến trong lớp **Pooling** gồm **MaxPooling** (lấy giá trị lớn nhất), **MinPooling** (lấy giá trị nhỏ nhất), và **AveragePooling** (lấy giá trị trung bình).

Ví dụ, giả sử bạn có một bức ảnh kích thước 32×32. Khi sử dụng **Global Average Pooling**, toàn bộ ma trận đầu vào sẽ được thu gọn thành một vector duy nhất bằng cách tính giá trị trung bình của từng kênh (channel). Phương pháp này giúp giảm đáng kể số lượng tham số và tập trung vào đặc trưng tổng quát của ảnh. Sau khi đi qua lớp Global Average Pooling, kích thước đầu ra sẽ chỉ còn là một vector có số phần tử tương ứng với số kênh của đầu vào (ví dụ: nếu đầu vào có 128 kênh, đầu ra sẽ là một vector có độ dài 128).

Lớp **Pooling** giúp giảm kích thước của dữ liệu. Khi ảnh đi qua nhiều lớp **Pooling**, nó sẽ dần được thu nhỏ lại nhưng vẫn giữ lại những đặc trưng quan trọng

cho việc nhận dạng. Việc giảm kích thước dữ liệu sẽ làm giảm số lượng tham số cần tính toán, từ đó giúp tăng tốc độ xử lý và giảm nguy cơ overfitting (quá khớp).

2.5. Fully Connected (FC) Layer

Layer này tương tự với layer trong mạng nơ-ron truyền thẳng, các giá trị ảnh được liên kết đầy đủ vào các nơ-ron trong layer tiếp theo. Sau khi ảnh được xử lý và rút trích đặc trưng từ các layer trước đó, dữ liệu ảnh sẽ không còn quá lớn so với mô hình truyền thẳng nên ta có thể sử dụng mô hình truyền thẳng để tiến hành nhận dạng.

2.6. Hoạt động của mô hình CNN

Hoạt động của mô hình CNN (Convolutional Neural Network) diễn ra thông qua sự kết nối của các layer (tầng) được thiết kế để thực hiện những nhiệm vụ cụ thể. Quá trình hoạt động bắt đầu với **Convolutional Layer**, nơi thực hiện các phép toán tích chập để trích xuất các đặc trưng từ dữ liệu đầu vào. Sau đó, **ReLU Layer** (Rectified Linear Unit Layer) thường được triển khai ngay sau **Convolutional Layer** nhằm áp dụng hàm kích hoạt và loại bỏ các giá trị âm, giúp tăng tính phi tuyến tính trong mạng. Trong một số trường hợp, hai tầng này có thể được tích hợp thành một tầng duy nhất để tối ưu hóa kiến trúc.

Tiếp theo, các tầng khác như **Convolutional Layer** hoặc **Pooling Layer** (tầng gộp) được thêm vào tùy thuộc vào thiết kế kiến trúc của mô hình. **Pooling Layer** giúp giảm kích thước không gian của dữ liệu, giảm số lượng tham số và tránh hiện tượng overfitting.

Cuối cùng, dữ liệu sẽ được đưa vào **Fully-Connected Layer** (tầng kết nối đầy đủ), nơi mà tất cả các nơ-ron được kết nối với nhau để thực hiện nhiệm vụ phân lớp hoặc dự đoán dựa trên các đặc trưng đã được trích xuất ở các tầng trước đó. Mô hình được tối ưu hóa thông qua quá trình huấn luyện để đạt được độ chính xác cao nhất trong việc nhận diện hoặc phân loại dữ liệu đầu vào.

CHƯƠNG 3. THỰC NGHIỆM VÀ KẾT QUẢ

3.1. Ý tưởng xây dung mô hình

Trong mô hình dự đoán giới tính và cảm xúc, chúng em sẽ huấn luyện 2 mô hình riêng biệt (CNN) như sau:

- **Mô hình 1:** Người đó là Nam hay Nữ
- **Mô hình 2:** Nhận diện 5 cảm xúc: 'Bình thường', 'Vui vẻ', 'Buồn', 'Tức giận', 'Ngạc nhiên'

Sau khi xác định hướng nghiên cứu cho đề tài, bước tiếp theo là xây dựng mô hình mạng nơ-ron tích chập CNN bằng ngôn ngữ lập trình Python và tiến hành huấn luyện. Khi quá trình huấn luyện hoàn tất, chúng em lưu mô hình dưới dạng tệp .h5 và sau đó sử dụng mô hình này để nhận diện thời gian thực. Dữ liệu hình ảnh được sử dụng trong nghiên cứu đã được thu thập từ các nguồn Kaggle và Google, sau đó được tiền xử lý theo yêu cầu và mục đích của quá trình huấn luyện mô hình.

3.2. Tập dữ liệu

3.2.1. Thu thập dữ liệu

Tập dữ liệu giới tính:

- Nguồn: Gender Classification Dataset từ Kaggle
- Link: https://www.kaggle.com/datasets/cashutosh/gender-classification-dataset
- Cấu trúc: 2 giới tính (Female, Male)
- Số lượng:
  - Training set: ~46,000 ảnh (~23,000 ảnh mỗi giới tính)
  - Validation set: ~11,000 ảnh (~5,500 ảnh mỗi giới tính)
  - Tổng cộng: ~57,000 ảnh đã được cân bằng hoàn hảo

Tập dữ liệu cảm xúc:

- Nguồn: Emotion Recognition Dataset từ Kaggle
- Link: https://www.kaggle.com/datasets/karthickmcw/emotion-recognition-dataset
- Cấu trúc: 5 cảm xúc (Angry, Happy, Neutral, Sad, Surprise)
- Số lượng:
  - Training set: ~4,800 ảnh (đã cân bằng ~960 ảnh/cảm xúc)

- Validation set: ~1,200 ảnh (đã cân bằng ~240 ảnh/cảm xúc)
- Tổng cộng: ~6,000 ảnh sau khi cân bằng dataset

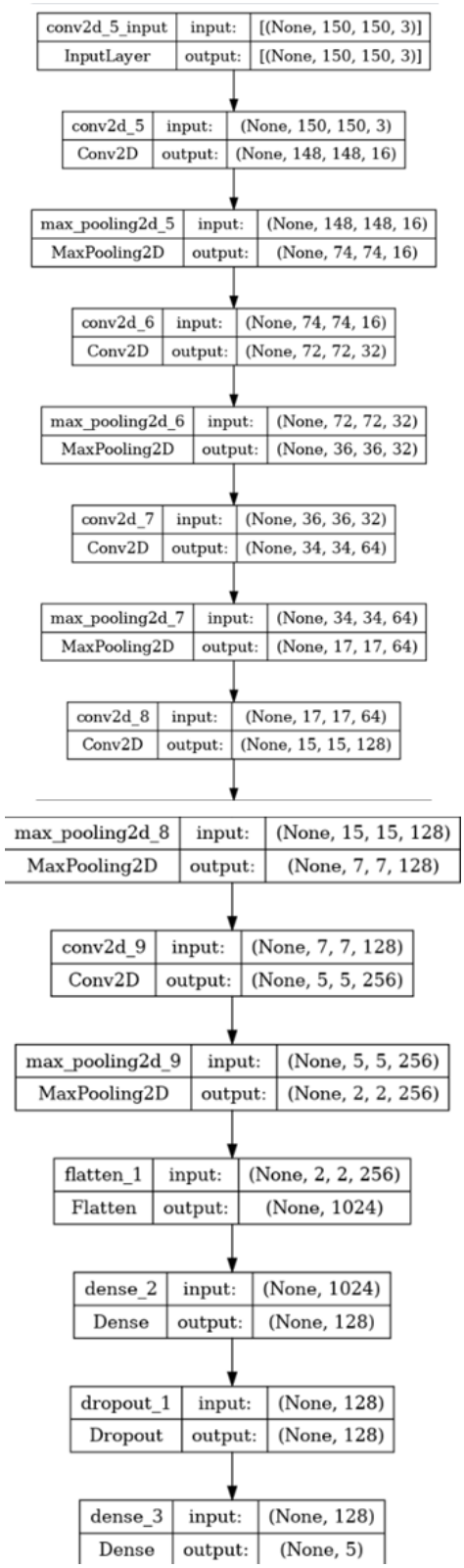
3.2.2. Tiền xử lý dữ liệu

- Kích thước ảnh: Đồng bộ kích thước dữ liệu bằng hình ảnh màu RGB với kích thước 150x150 pixels
- Chuẩn hóa: Chuẩn hóa dữ liệu đầu vào trong khoảng (0,1) thay vì 255 như ban đầu
- Data Augmentation: Áp dụng các kỹ thuật tăng cường dữ liệu:
- Rotation: 20°
- Width/Height shift: 0.2
- Shear range: 0.2 (chỉ emotion)
- Zoom range: 0.2
- Horizontal flip: True

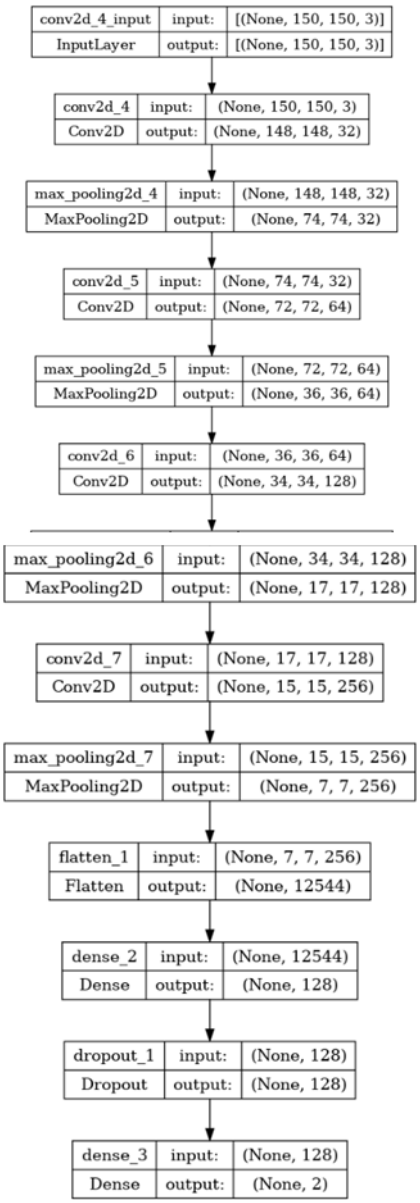
3.3. Xây dựng và huấn luyện mô hình

3.3.1. Xây dựng mô hình

Sau khi thu thập và tiền xử lý dữ liệu, chúng em tiến hành xây dựng mô hình dự đoán giới tính.



Hình 6. Mô hình CNN dự đoán cảm xúc



Hình 7. Mô hình CNN nhận diện giới tính

Cấu hình huấn luyện:

- Optimizer: RMSprop với learning rate = 0.001
- Loss function: Categorical crossentropy
- Batch size: 32
- Epochs: 20 (với Early Stopping, patience=5)
- Metrics: Accuracy

3.3.2. Kết quả của mô hình

Mô hình Gender Classification (Gender1.h5):

- Training Accuracy: 92.85%
- Validation Accuracy: 95.38%
- Test Accuracy: 95.38%
- Training Loss: 0.2023
- Validation Loss: 0.1493

Mô hình Emotion Recognition (Emotion1.h5):

- Training Accuracy: 65.41%
- Validation Accuracy: 71.36%
- Test Accuracy: 71.36%
- Training Loss: 0.8909
- Validation Loss: 0.7254

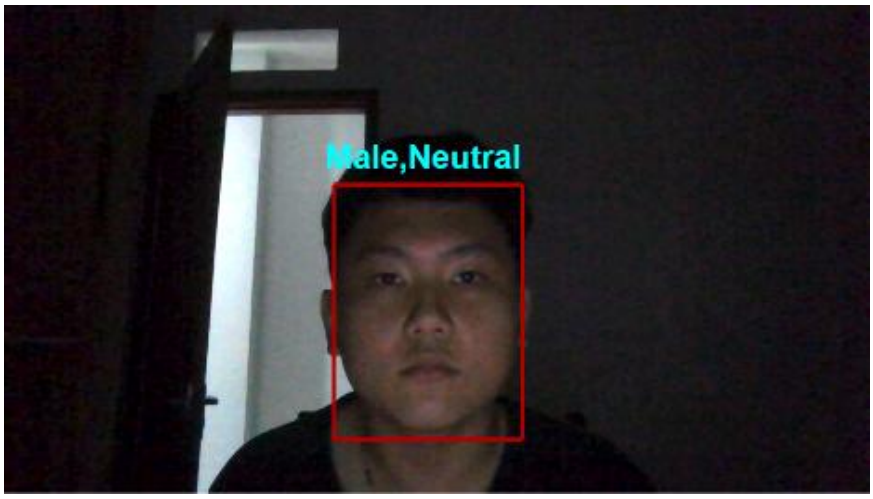
Kết quả cho thấy mô hình phân loại giới tính đạt độ chính xác rất cao (95.38%) nhờ dataset được cân bằng hoàn hảo và chất lượng tốt. Mô hình nhận diện cảm xúc đạt độ chính xác khá tốt (71.36%), có thể cải thiện thêm bằng cách tối ưu hóa dataset và kiến trúc mô hình.

3.3.3. Kết quả thử nghiệm thời gian thực

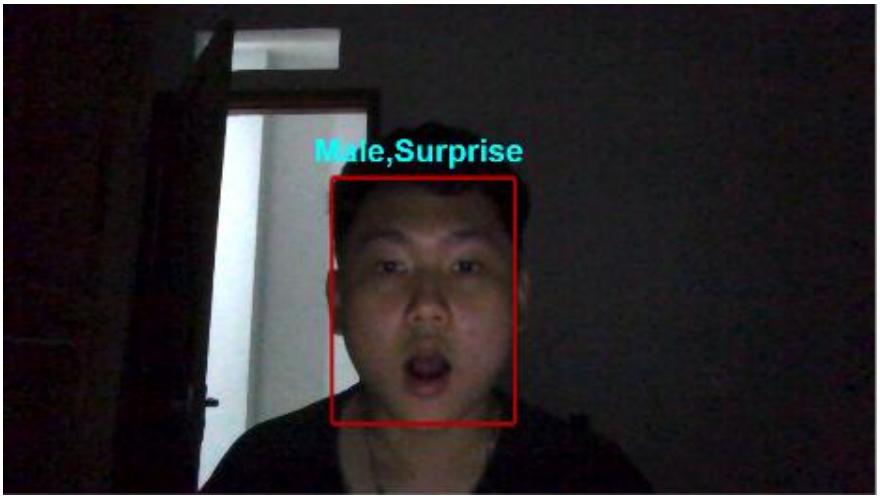
Sau khi mô hình đã được huấn luyện và kiểm tra độ chính xác, bước tiếp theo là thực hiện đánh giá chất lượng của mô hình trong thời gian thực. Để làm điều này, chúng em lưu các tệp h5 của mô hình và tích hợp chúng vào mã nguồn thời gian thực. Thông qua quá trình chạy trong thời gian thực, chúng em có thể đánh giá hiệu suất của mô hình và thu được kết quả thử nghiệm

Hình 8. Giao diện người dùng

- Giao diện người dùng: Sử dụng Tkinter với giao diện toàn màn hình
- Xử lý đa khuôn mặt: Có thể nhận diện nhiều khuôn mặt cùng lúc
- Hiển thị kết quả: Hiển thị giới tính, cảm xúc và độ tin cậy cho từng khuôn mặt
- Thời gian xử lý: Trung bình ~50-100ms cho mỗi frame
- Face Detection: Sử dụng thư viện cvlib để phát hiện khuôn mặt

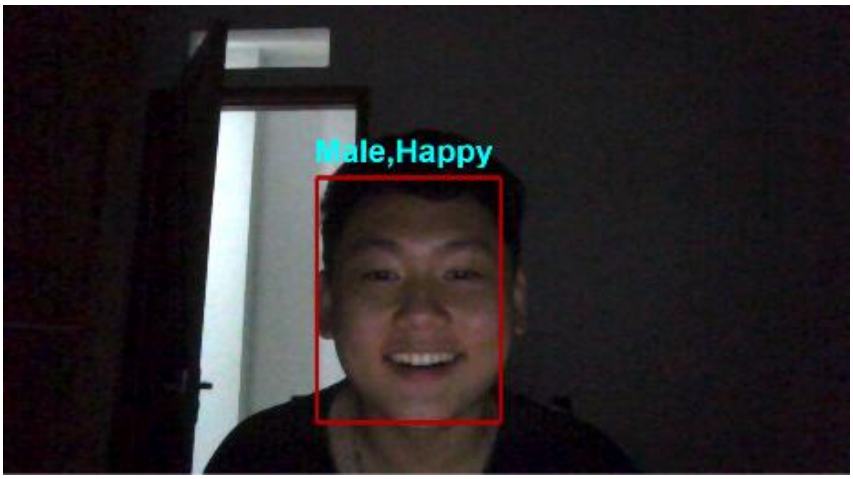


Hình 9. Nhận diện cảm xúc Bình thường



Hình 10. Nhận diện cảm xúc Ngạc nhiên

Hình 11. Nhận diện cảm xúc Tức giận



Hình 12. Nhận diện cảm xúc Vui vẻ

CHƯƠNG 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Qua quá trình nghiên cứu và phát triển hệ thống nhận diện khuôn mặt và cảm xúc, nhóm đã thành công xây dựng một ứng dụng hoàn chỉnh có khả năng phân tích thời gian thực với độ chính xác cao. Dự án đã đạt được những mục tiêu đề ra ban đầu:

4.1. Những thành tựu đạt được

1. **Xây dựng thành công hai mô hình CNN:**
    - Mô hình nhận diện giới tính với độ chính xác cao
    - Mô hình phân loại cảm xúc (Bình thường, Vui vẻ, Buồn, Ngạc nhiên, Tức giận)
2. **Phát triển ứng dụng thời gian thực:**
    - Giao diện người dùng thân thiện với Tkinter
    - Xử lý video camera trực tiếp
    - Hiển thị kết quả phân tích ngay lập tức
    - Đo lường thời gian xử lý và độ tin cậy
3. **Tối ưu hóa hiệu suất:**
    - Sử dụng kỹ thuật Data Augmentation để cải thiện độ chính xác
    - Kiến trúc CNN tối ưu với nhiều lớp tích chập
    - Xử lý đa luồng để đảm bảo giao diện mượt mà

4.2. Ý nghĩa khoa học và thực tiễn

1. **Về mặt khoa học:**
    - Áp dụng thành công các kỹ thuật Deep Learning tiên tiến
    - Kết hợp hiệu quả giữa Computer Vision và Machine Learning
    - Nghiên cứu và triển khai kiến trúc CNN phù hợp cho bài toán cụ thể
2. **Về mặt thực tiễn:**

- Ứng dụng có thể triển khai trong nhiều lĩnh vực: an ninh, giáo dục, y tế
- Giao diện trực quan, dễ sử dụng cho người dùng cuối
- Khả năng mở rộng và tích hợp với các hệ thống khác

4.3. Những hạn chế và hướng phát triển

1. **Hạn chế hiện tại:**
   - Độ chính xác có thể bị ảnh hưởng bởi điều kiện ánh sáng
   - Chưa hỗ trợ nhận diện đồng thời nhiều khuôn mặt với hiệu suất tối ưu
   - Dataset huấn luyện có thể cần mở rộng thêm
2. **Hướng phát triển tương lai:**
   - Tích hợp thêm các tính năng như nhận diện tuổi, sắc tộc
   - Cải thiện độ chính xác bằng cách sử dụng các mô hình tiên tiến hơn (ResNet, EfficientNet)
   - Tối ưu hóa để chạy trên các thiết bị có cấu hình thấp

TÀI LIỆU THAM KHẢO

[1] (n.d.). *Gender Classification Dataset*. Kaggle. Truy cập từ https://www.kaggle.com/datasets/cashutosh/gender-classification-dataset

[2] (n.d.). *Facial Expression Recognition Dataset (FER2013)*. Kaggle. Truy cập từ https://www.kaggle.com/datasets/msambare/fer2013

[3] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Truy cập từ https://www.deeplearningbook.org