**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

TRƯỜNG ĐẠI HỌC SƯ PHẠM HÀ NỘI

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



**BÁO CÁO TỔNG KẾT**

**TÊN ĐỀ TÀI:**

**ỨNG DỤNG AI TRONG NHẬN DIỆN CẢM XÚC SINH VIÊN THỜI GIAN THỰC TRONG LỚP HỌC TRỰC TUYẾN**

Lĩnh vực khoa học và công nghệ: Học sâu (Deep learning)

Chuyên ngành thuộc lĩnh vực khoa học và công nghệ: Công nghệ thông tin

HÀ NỘI, 2021

HÀ NỘI, 2021

**MỤC LỤC**

[LỜI CẢM ƠN 3](#_Toc101211265)

[NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN 4](#_Toc101211266)

[NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN 5](#_Toc101211267)

[LỜI CAM ĐOAN 6](#_Toc101211268)

[MỤC LỤC 8](#_Toc101211269)

[DANH MỤC HÌNH ẢNH 11](#_Toc101211270)

[DANH MỤC BẢNG BIỂU 12](#_Toc101211271)

[CHƯƠNG 1: GIỚI THIỆU CHUNG 13](#_Toc101211272)

[1.1 Lý do chọn đề tài 13](#_Toc101211273)

[1.2 Giới thiệu 13](#_Toc101211274)

[1.1.1 Bài toán nhận diện cảm xúc dựa trên biểu cảm khuôn mặt 14](#_Toc101211275)

[1.1.2 Cảm xúc có ảnh hưởng trực tiếp đến quá trình học tập 15](#_Toc101211276)

[1.1.3 Nhận diện cảm xúc dựa trên mạng nơ-ron tích chập 15](#_Toc101211277)

[1.3 Tiềm năng ứng dụng của bài toán nhận diện cảm xúc dựa trên biểu cảm khuôn mặt trong giáo dục trực tuyến 16](#_Toc101211278)

[1.4 Mục tiêu của luận văn 17](#_Toc101211279)

[1.5 Cấu trúc luận văn 17](#_Toc101211280)

[CHƯƠNG 2. TỔNG QUAN VỀ MẠNG NƠ-RON VÀ GIỚI THIỆU VỀ MẠNG NƠ-RON TÍCH CHẬP 19](#_Toc101211281)

[2.1 Học sâu và mạng nơ-ron 19](#_Toc101211282)

[2.1.1 Trí tuệ nhân tạo, học máy và học sâu 19](#_Toc101211283)

[2.1.2 Sơ lược lịch sử mạng nơ-ron trong học sâu 19](#_Toc101211284)

[2.1.3 Cấu tạo và quá trình xử lý của một nơ-ron sinh học 20](#_Toc101211285)

[2.2 Mạng nơ-ron trong lĩnh vực học sâu 21](#_Toc101211286)

[2.2.1 Cấu tạo và quá trình xử lý của một nơ-ron trong học sâu 21](#_Toc101211287)

[2.2.2 Các mô hình hàm kích hoạt phổ biến của mạng nơ-ron 22](#_Toc101211288)

[2.3.1 Mô hình mạng nơ-ron tích chập 25](#_Toc101211289)

[2.3.1.1 Tầng tích chập (convolutional) 26](#_Toc101211290)

[2.3.1.2 Tầng tổng hợp (pooling layer) 27](#_Toc101211291)

[2.3.1.3 Tầng kết nối đầy đủ (fully-connected) 27](#_Toc101211292)

[2.3.2 Mô hình quá khớp (over-fitting) và mô hình chưa khớp (under-fitting) 28](#_Toc101211293)

[2.3.3 Chuẩn hóa dữ liệu đầu ra (phương pháp Batch Normalization) 29](#_Toc101211294)

[2.3.4 Thuật toán tối ưu (optimizers) trong huấn luyện mạng nơ-ron 30](#_Toc101211295)

[2.3.4.1 Gradient Descent 30](#_Toc101211296)

[2.3.4.2 SGD với động lượng (SGD with momentum) 30](#_Toc101211297)

[2.3.4.3 RMSProp (Root Mean Square Propogation) 31](#_Toc101211298)

[2.3.4.4 Adagrad 31](#_Toc101211299)

[2.3.4.5 Adadelta 31](#_Toc101211300)

[2.3.4.6 Adam 32](#_Toc101211301)

[2.3.5 Một số mô hình mạng huấn luyện nổi tiếng 32](#_Toc101211302)

[2.3.5.1 Mạng huấn luyện AlexNet 32](#_Toc101211303)

[2.3.3.2 Mạng huấn luyện VGG16 33](#_Toc101211304)

[2.3.6 Huấn luyện mô hình 34](#_Toc101211305)

[2.3 Kết luận 34](#_Toc101211306)

[CHƯƠNG 3. PHƯƠNG PHÁP ĐỀ XUẤT 36](#_Toc101211307)

[3.1 Nhận diện cảm xúc dựa trên mạng CNN 36](#_Toc101211308)

[3.2 Lược đồ đề xuất 36](#_Toc101211309)

[3.3 Hình ảnh đầu vào 36](#_Toc101211310)

[3.4 Phát hiện khuôn mặt 37](#_Toc101211311)

[3.4.1 Tổng quan về Haar Cascade 37](#_Toc101211312)

[3.4.2 Cách hoạt động của phương pháp Haar Cascade 37](#_Toc101211313)

[3.5 Tiền xử lý ảnh 39](#_Toc101211314)

[3.6 Nhận diện cảm xúc 40](#_Toc101211315)

[3.6.1 Bộ dữ liệu đào tạo 40](#_Toc101211316)

[3.6.2 Xây dựng mô hình nhận diện cảm xúc 41](#_Toc101211317)

[3.6.3 Môi trường đào tạo 43](#_Toc101211318)

[3.6.4 Đào tạo mô hình nhận diện cảm xúc 43](#_Toc101211319)

[3.7 Kết luận 45](#_Toc101211320)

[CHƯƠNG 4. KẾT QUẢ THỰC NGHIỆM 46](#_Toc101211321)

[4.1 Cài đặt chương trình nhận diện cảm xúc 46](#_Toc101211322)

[4.2 So sánh với mô hình mạng nơ-ron VGG16 47](#_Toc101211323)

[4.2 Kết quả thực nghiệm với bộ dữ liệu FER 2013 48](#_Toc101211324)

[4.3 Ứng dụng thực tế trên khuôn mặt 49](#_Toc101211325)

[4.4 Thử nghiệm thực tế tại lớp học trường Đại học sư phạm Hà Nội 52](#_Toc101211326)

[4.5 Kết quả 55](#_Toc101211327)

[TÀI LIỆU THAM KHẢO 59](#_Toc101211328)

# **DANH MỤC HÌNH ẢNH**

[Hình 1.1 Một số hình ảnh được gán nhãn cảm xúc trong CSDL FER2013 14](#_Toc101215777)

[Hình 2.1 Hình ảnh một nơ-ron sinh học [25] 20](#_Toc101215778)

[Hình 2.2 Minh họa mạng nơ-ron nhân tạo 21](#_Toc101215779)

[Hình 2.3 Quá trình tính toán của một nơ-ron 22](#_Toc101215780)

[Hình 2.4 Đồ thị hàm Sigmoid 23](#_Toc101215781)

[Hình 2.5 Đồ thị hàm TanH 24](#_Toc101215782)

[Hình 2.6 Đồ thị hàm ReLU 25](#_Toc101215783)

[Hình 2.7 Mô hình lớp (layer) trong mạng CNN 25](#_Toc101215784)

[Hình 2.8 Mô hình lớp (layer) trong mạng CNN 26](#_Toc101215785)

[Hình 2.9 Phương thức tổng hợp Max Pooling và Avarage Pooling 27](#_Toc101215786)

[Hình 2.10 Minh họa mô hình kiến trúc mạng CNN 28](#_Toc101215787)

[Hình 2.11 Các trạng thái over-fitting, under-fitting và cân bằng 29](#_Toc101215788)

[Hình 3.1 Lược đồ phương pháp đề xuất 36](#_Toc101215796)

[Hình 3.2 Đặc trưng hình chữ nhật trong phương pháp Haar-Cascade 38](#_Toc101215797)

[Hình 3.3 Các đặc trưng hình chữ nhật khác trong phương pháp Haar-Cascade 38](#_Toc101215798)

[Hình 3.4 Phát hiện khuôn mặt bằng phương pháp Haar-Cascade 39](#_Toc101215799)

[Hình 3.5 Tiền xử lý hình ảnh đầu vào 39](#_Toc101215800)

[Hình 3.6 Kiến trúc mạng tích chập cho nhận diện cảm xúc 42](#_Toc101215801)

[Hình 3.7 Sơ đồ khối đào tạo mô hình nhận diện cảm xúc 44](#_Toc101215802)

[Hình 4.1 Sơ đồ khối nhận diện cảm xúc 47](#_Toc101215807)

[Hình 4.2 Lớp học trực tuyến môn học “Một số vấn đề xã hội của CNTT” 54](#_Toc101215808)

[Hình 4.3 Nhận diện cảm xúc khuôn mặt trong lớp học trực tuyến môn học “Một số vấn đề xã hội của CNTT” 54](#_Toc101215809)

[Hình 4.4 Biểu đồ đánh giá cảm xúc lớp học trực tuyến môn học “Một số vấn đề xã hội của CNTT” 55](#_Toc101215810)

**DANH MỤC BẢNG BIỂU**

[Bảng 1.1 Mô tả đầu vào và đầu ra của bài toán nhận diện cảm xúc dựa trên biểu cảm khuôn mặt 14](#_Toc101217486)

[Bảng 3.1 Thông tin chi tiết số lượng ảnh và cảm xúc trong bộ CSDL FER2013 40](#_Toc101217487)

[Bảng 3.2 Thông tin số lượng dữ liệu đào tạo theo từng loại cảm xúc trong bộ CSDL FER2013 41](#_Toc101217488)

[Bảng 3.3 Thông tin số lượng dữ liệu kiểm thử theo từng loại cảm xúc trong bộ CSDL FER2013 41](#_Toc101217489)

[Bảng 3.4 Các tham số chi tiết cho mô hình đề xuất 42](#_Toc101217490)

[Bảng 3.5 Cấu hình phần cứng GoogleColab 43](#_Toc101217491)

[Bảng 4.1 So sánh cấu trúc và thời gian đào tạo mô hình đề xuất và VGG16 48](#_Toc101217492)

[Bảng 4.2 Kết quả thí nghiệm kiểm tra mô hình với bộ dữ liệu kiểm thử 48](#_Toc101217493)

[Bảng 4.3 Một số kết quả thử nghiệm 49](#_Toc101217494)

[Bảng 4.4 Kết quả thực nghiệm với khuôn mặt tác giả trong thời gian thực với mô hình đề xuất 50](#_Toc101217495)

[Bảng 4.5 Kết quả thực nghiệm với khuôn mặt tác giả trong thời gian thực với mô hình VGG16 51](#_Toc101217496)

[Bảng 4.6 Kết quả thử nghiệm tại lớp học Khoa Công nghệ thông tin, trường ĐH Sư phạm Hà Nội 53](#_Toc101217497)

# **MỞ ĐẦU**

Với sự phát triển mạnh mẽ của công nghệ thông tin, giáo dục trực tuyến dần trở thành một xu hướng mới đầy tiềm năng và thách thức. Đặc biệt trong hoàn cảnh nghiêm trọng của dịch bệnh COVID-19 như hiện nay, hầu hết các trường học đều đang đóng cửa, giáo dục trực tuyến được xem là một trong những giải pháp tối ưu nhất hiện nay. Có nhiều nghiên cứu trước đây đã chỉ ra rằng, có một mối quan hệ chặt chẽ và ổn định giữa biểu cảm khuôn mặt và cảm xúc của một người nào đó. Do đó, để đánh giá khách quan chất lượng của các lớp học trực tuyến, một phương pháp nhận diện cảm xúc tự động được giới thiệu dựa trên một mô hình mạng tích chập CNN (Convolutional Neural Network). Mô hình cho phép nhận diện bảy loại cảm xúc khác nhau của con người. Phương pháp đề xuất được thực nghiệm dựa trên bộ CSDL về nhận diện cảm xúc là FER2013. Ngoài ra, ba lớp học trực tuyến gồm ba lớp sinh viên khoa CNTT, trường ĐHSPHN cũng được sử dụng để đánh giá. Các kết quả thu được cho thấy mô hình đề xuất không chỉ hiệu quả với các bộ dữ liệu chuẩn mà còn hoạt động mạnh mẽ trong các môi trường thực nghiệm khác nhau.

1. **Tổng quan tình hình nghiên cứu trong và ngoài nước**

Với hầu hết mọi người thì biểu cảm trên khuôn mặt là một trong những tín hiệu mạnh mẽ, tự nhiên và phổ biến nhất để con người truyền tải trạng thái cảm xúc và ý nghĩ của họ­ [1], [2], có rất nhiều ứng dụng liên quan đến vấn đề này như: quản lý sức khỏe [3], hỗ trợ lái xe, giao tiếp, … [4].

Ekman và Friesen [5] đã chỉ ra rằng con người nhận thức được một số cảm xúc cơ bản theo cùng một cách bất kể nền tảng văn hóa hay quốc gia nào và họ đã xác định có sáu loại cảm xúc cơ bản bao gồm: giận dữ, ghê tởm, sợ hãi, vui vẻ, buồn bã và ngạc nhiên. Trong một nghiên cứu mở rộng khác, Ekman và Heider [21] đã bổ sung thêm một loại cảm xúc nữa là khinh bỉ.

Ngoài ra, FER 2013, một bộ cơ sở dữ liệu quy mô lớn được giới thiệu trong IMCL 2013, cũng giới thiệu và phân loại các khuôn mặt với bảy loại trạng thái cảm xúc khác nhau bao gồm: giận dữ, ghê tởm, sợ hãi, vui vẻ, buồn bã, ngạc nhiên và bình thường. Trong các nghiên cứu khác, các nhà khoa học cũng đã giới thiệu nhiều loại mô hình khác nhau để cung cấp nhiều loại cảm xúc hơn do sự phức tạp của nét mặt tuy nhiên, các cảm xúc mở rộng này chiếm một phần khá nhỏ trong các biểu hiện cảm xúc hàng ngày nên chưa được đưa vào trong nghiên cứu này [7]. Hình 1.1 minh hoạ một số biểu cảm khuôn mặt cơ bản kèm theo các nhãn cảm xúc tương ứng trong bộ cơ sở dữ liệu FER2013, sẽ được sử dụng để thử nghiệm trong nghiên cứu này.

Sợ hãi

Ghê tởm

Giận dữ

Bình thường

Vui vẻ

Buồn bã



Ngạc nhiên

Giận dữ

Buồn bã

Bình thường

Vui vẻ

Sợ hãi

Ghê tởm

Hình 1.1 Một số hình ảnh được gán nhãn cảm xúc trong CSDL FER2013

# **Lý do chọn đề tài**

Với sự phát triển mạnh mẽ của công nghệ thông tin, giáo dục trực tuyến dần trở thành một xu hướng mới đầy tiềm năng và thách thức. Đặc biệt trong hoàn cảnh nghiêm trọng của dịch bệnh COVID-19 như hiện nay. Theo số liệu từ WHO tính đến ngày 30/03/2022 có đến 6.132.461 trường hợp tử vong, trước tình hình nguy hiểm của dịch bệnh, điều này buộc các trường học tại Việt Nam cũng như trên toàn thế giới buộc người dân hạn chế ra khỏi nhà khi không cần thiết. Do đó, hầu hết tất cả các trường học đều đang đóng cửa, giao tiếp trực tuyến qua trong thời gian thực không chỉ đem lại hiệu quả trong hoàn cảnh dịch bệnh này mà phương pháp này còn đem lại nhiều lợi ích khác như: giao tiếp từ xa không quan trọng khoảng cách địa lý, tính tiện lợi và linh hoạt,… Nhờ vào những ưu điểm trên, giáo dục trực tuyến được xem là một trong những giải pháp tối ưu nhất hiện nay. Trong suốt những thập kỷ qua, giáo dục trực tuyến đã phát triển nhanh chóng dù là tại các trường đại học hay cơ sở đào tạo [18], điều này mang lại cơ hội ứng dụng tiềm năng cho các hệ thống nhận diện cảm xúc. Vấn đề khó khăn lớn giữa lớp học trực tuyến học trực tiếp truyền thống đó là các lớp học trực tuyến thường được coi là ít ràng buộc hơn và giao tiếp kém hiệu quả, chắc chắn sẽ dẫn đến sự nghi ngờ của giảng viên cũng như sinh viên, sinh viên đối với phương pháp giáo dục mới lạ này trong khi có một số nghiên cứu cho rằng kết quả học tập của sinh viên đạt được bằng giáo dục trực tuyến có thể tương đương với các lớp học truyền thống, ngoại trừ các kỹ năng đòi hỏi độ chính xác tối ưu và mức độ nhận thức xúc giác cao hơn [19]. Không thể phủ nhận rằng tốc độ phát triển nhanh chóng của giáo dục trực tuyến có thể mang lại sự thuận tiện và linh hoạt cho nhiều sinh viên hơn, vì vậy nó cũng có không gian phát triển rộng rãi trong tương lai. Do đó, làm thế nào để đảm bảo rằng sinh viên giữ được mức độ tập trung và hiệu quả học tập như các lớp học truyền thống trong quá trình giáo dục trực tuyến là rất quan trọng để thúc đẩy sự phát triển hơn nữa của giáo dục trực tuyến. Để giải quyết vấn đề này, cần phải có những công cụ đánh giá chủ quan và khách quan làm cơ sở cho những sự thay đổi, cải tiến nhằm nâng cao chất lượng đào tạo.

Tuy nhiên, hiệu quả của các lớp học trực tuyến từ lâu đã bị đặt nhiều dấu hỏi. So với các lớp học trực tiếp truyền thống, các lớp học trực tuyến thiếu sự giao tiếp và phản hồi trực tiếp, kịp thời và hiệu quả giữa giáo viên và học viên.

Vì vậy, nhằm nâng cao chất lượng giáo dục trực tuyến và khắc phục nhược điểm lớn nhất của phương pháp giáo dục này, thì việc tìm ra một giải pháp giúp cải thiện tính tương tác giữa người dạy và người học là vấn đề cấp thiết.

Bằng cách kết hợp các nền tảng giáo dục trực tuyến hiện có với mô hình nhận diện nét mặt dựa trên kiến ​​trúc của mạng nơ-ron tích chập, chúng tôi đã đề xuất một phương pháp cho phép theo dõi thời gian thực cảm xúc của sinh viên trong các khóa học trực tuyến và đảm bảo rằng phản hồi được thể hiện bằng nét mặt có thể cung cấp cho giáo viên một công cụ đánh giá khách quan kịp thời. Giúp các nhà quản lý, giảng viên có thêm một công cụ để họ có thể linh hoạt điều chỉnh chương trình dạy học một cách phù hợp hơn và cuối cùng là nâng cao chất lượng và hiệu quả của giáo dục trực tuyến.

# **4. Mục tiêu của nghiên cứu**

Luận văn đề xuất phương pháp giải quyết bài toán nhận diện cảm xúc dựa trên biểu cảm khuôn mặt trong giáo dục trực tuyến giúp cải thiện hình thức học trực tuyến trong hoàn cảnh dịch bệnh nghiêm trọng. Mô hình mạng nơ-ron tích chập cho phép trích xuất các đặc trưng của biểu cảm khuôn mặt với tốc độ xử lý nhanh và cho ra độ chính xác cao.

# **5. Nội dung nghiên cứu**

### **5.1 Bài toán nhận diện cảm xúc dựa trên biểu cảm khuôn mặt**

Với sự phát triển của công nghệ thông tin, đặc biệt trong lĩnh vực trí tuệ nhân tạo và học sâu, nhiều thuật toán nhận diện cảm xúc được đề xuất để nhận diện các biểu cảm được thể hiện trên khuôn mặt. Bài toán nhận diện cảm xúc dựa trên biểu cảm khuôn mặt chính là việc xác định loại cảm xúc của một người thông qua đặc trưng của nét mặt.

Đầu vào: là hình ảnh từ các lớp học dựa trên các nền tảng trực tuyến, dựa trên hình ảnh của sinh viên được trích xuất từ thiết bị học trực tuyến hình ảnh đầu và là một bức ảnh chứa các khuôn mặt của sinh viên trong lớp học trực tuyến.

Đầu ra: là kết quả dự đoán cảm xúc của sinh viên, kết quả ở đây có thể là kết quả tổng hợp được biểu diễn theo dạng biểu đồ thông số các cảm xúc hoặc tỉ lệ phần trăm giữa các cảm xúc được đánh giá trong một lớp học trực tuyến tại một thời điểm.

Mô tả cụ thể về bài toán thông qua ví dụ sau:

Bảng 1.1 Mô tả đầu vào và đầu ra của bài toán nhận diện cảm xúc dựa trên biểu cảm khuôn mặt

|  |  |  |
| --- | --- | --- |
| Ảnh đầu vào | Ảnh đầu ra | Nhãn kết quả (cảm xúc) |
| D:\Khóa luận\khoa-luan-tot-nghiep\Thư mục ảnh đi kèm\Huy-bình thường-trước.png | D:\Khóa luận\khoa-luan-tot-nghiep\Thư mục ảnh đi kèm\Huy-bình thường-sau.png | Bình thường |

### **5.2 Cảm xúc có ảnh hưởng trực tiếp đến quá trình học tập**

Cảm xúc đóng một vai trò quan trọng trong quá trình học tập. Một số nghiên cứu đặc biệt nhấn mạnh ý nghĩa của cảm xúc trong học tập đặc biệt là cảm xúc trong giáo dục. Những cảm xúc tích cực như vui vẻ, bất ngờ, hào hứng và bình thường cho là sẽ đóng góp vào động lực cả bên trong và bên ngoài, thúc đẩy việc sử dụng các chiến lược học tập linh hoạt, những cảm xúc được cho là tác động tiêu cực có thể cản trở quá trình học tập bất cứ lúc nào bao gồm: sợ hãi , tức giận, buồn bã, chán nản. Do đó, cảm xúc có khả năng tác động đến điều kiện kết quả học tập [34].

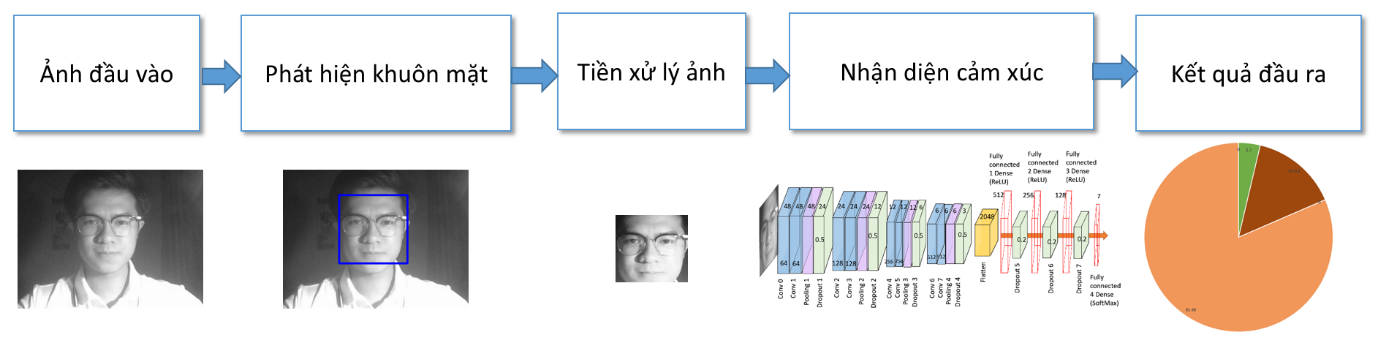
### **5.3 Nhận diện cảm xúc dựa trên mạng CNN**

Từ những ưu điểm của mạng CNN trong những bài toán xử lý ảnh nói chung như đã trình bày ở phần trước, thì việc ứng dụng mạng CNN để xây dựng một mô hình nhận diện cảm xúc và cụ thể là mô hình dựa trên ý tưởng của VGG16 để có thể giải quyết bài toán nhận diện cảm xúc của sinh viên dựa trên biểu cảm khuôn mặt là hoàn toàn khả thi, tuy nhiên việc xây dựng một mô hình mạng dựa trên ý tưởng của mô hình VGG16 để cho ra kết quả nhận diện tốt thì cần có những bước tiền xử lý ảnh đầu vào sao cho hiệu quả, nhằm đảm bảo độ chính xác cũng như hiệu năng của chương trình. Thêm vào đó, dữ liệu đầu ra là các nhãn cảm xúc cần được thống kê một cách cụ thể và khoa học nhằm giúp cho người dạy học có thể nắm bắt số liệu một cách nhanh chóng để điều chỉnh chất lượng dạy học.

Do đó, một lược đồ đề xuất bao gồm năm bước chính là: thu thập ảnh đầu vào, phát hiện khuôn mặt, tiền xử lý ảnh, nhận diện cảm xúc và cuối cùng là thống kê lại kết quả sẽ được đề xuất trong phần này.

### **5.4 Lược đồ đề xuất**

Trong phần này, một lược đồ nhận diện cảm xúc dựa trên các nền tảng học trực tuyến được giới thiệu. Hiện tại, có hai nền tảng học trực tuyến được sử dụng phổ biến tại trường ĐH Sư phạm Hà Nội là Zoom và Google meet. Do đó, các ảnh đầu vào sẽ được thu thập chủ yếu dựa trên hai nền tảng này. Lược đồ nhận diện đề xuất bao gồm năm bước chính bao gồm: thu thập ảnh đầu vào, phát hiện khuôn mặt, tiền xử lý ảnh đầu vào, nhận diện cảm xúc và hiển thị kết quả. Hình 3.1 minh hoạ một cách trực quan các bước của lược đồ. Một biểu đồ thống kê tổng số các cảm xúc hiện có trong lớp sẽ được tổng hợp và cung cấp cho các giảng viên. Dựa trên biểu đồ thống kê này, giảng viên và các nhà quản lý đào tạo sẽ có thêm một kênh đánh giá khách quan để có thể điều chỉnh kế hoạch giảng dạy nhằm nâng cao chất lượng đào tạo.



**Hình 3.1 Lược đồ phương pháp đề xuất**

### **5.5 Hình ảnh đầu vào**

Những tiến bộ trong công nghệ đã tạo ra một số lượng lớn các nền tảng giáo dục trực tuyến và tăng tính linh hoạt trong đào tạo. Những nền tảng công nghệ này cho phép giáo viên áp dụng các phương tiện công nghệ cao và đa dạng để hỗ trợ giảng dạy mà không phải lo lắng về giới hạn số lượng sinh viên trong lớp như các lớp học truyền thống và sinh viên ở các vị trí địa lí khác nhau hoàn toàn có thể giao tiếp trong thời gian thực mà không cần phải đến lớp. Các tài liệu giảng dạy tương tự như các lớp học truyền thống có thể được tải lên các nền tảng này để sinh viên tham khảo thêm. Hiện tại, hầu hết các nền tảng này đều tích hợp chức năng dạy trực tuyến như Zoom, Google meet, MS Team... Khi đó, giảng viên có thể dễ dàng tương tác với sinh viên thời gian thực và cũng dễ dàng thu được hình ảnh khuôn mặt của sinh viên dựa trên các camera tích hợp. Các hình ảnh khuôn mặt này có thể được sử dụng như là tập các dữ liệu đầu vào cho hệ thống đề xuất để có thể đánh giá và nhận diện cảm xúc của người học theo thời gian thực.

### **5.6 Phát hiện khuôn mặt**

#### **5.6.1 Tổng quan về Haar Cascade**

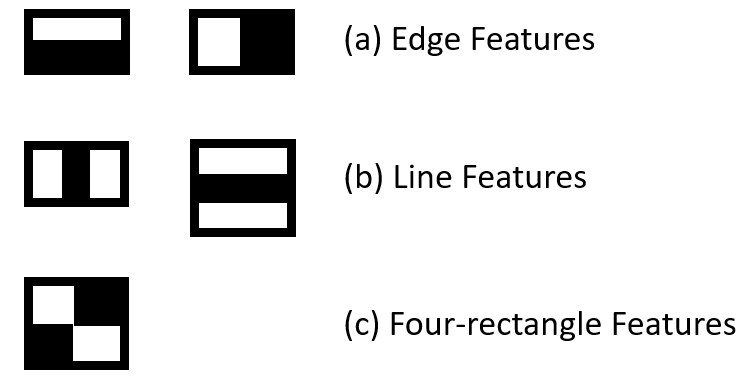
Các hình ảnh khuôn mặt đầu vào có thể chứa nhiều thông tin khác nhau ngoài hình ảnh khuôn mặt cần nhận diện (nhiều chi tiết khác trên ảnh nền, …) do đó, cần phải xác định chính xác vị trí khuôn mặt trong ảnh trước khi tiến hành nhận diện. Trong nhiều trường hợp, người học có thể sử dụng các loại background khác nhau, sẽ khiến cho việc phát hiện khuôn mặt khó khăn hơn. Trong nghiên cứu này, để có thể phát hiện và cắt được chính xác vị trí khuôn mặt trong ảnh, phương pháp Haar-Cascade[20] được ứng dụng dựa trên các đặc trưng Haar.

Haar Cascade là một thuật toán được tạo ra để phát hiện đối tượng (có thể là khuôn mặt, mắt, tay, đồ vật,…) được đề xuất vào năm 2001 bởi Paul Viola và Michael Jones trong bài báo của họ với khẳng định “Phát hiện đối tượng một cách nhanh chóng bằng cách sử dụng tầng (Cascade) tăng cường các tính năng đơn giản”.

Triển khai ban đầu được sử dụng để phát hiện khuôn mặt chính diện và các đặc điểm như Mắt, Mũi và Miệng. Tuy nhiên, có nhiều đặc trưng Haar được đào tạo trước cho các đối tượng khác cũng như cho toàn bộ cơ thể, thân trên, thân dưới, nụ cười và nhiều đồ vật khác.

#### **5.6.2 Cách hoạt động của phương pháp Haar Cascade**

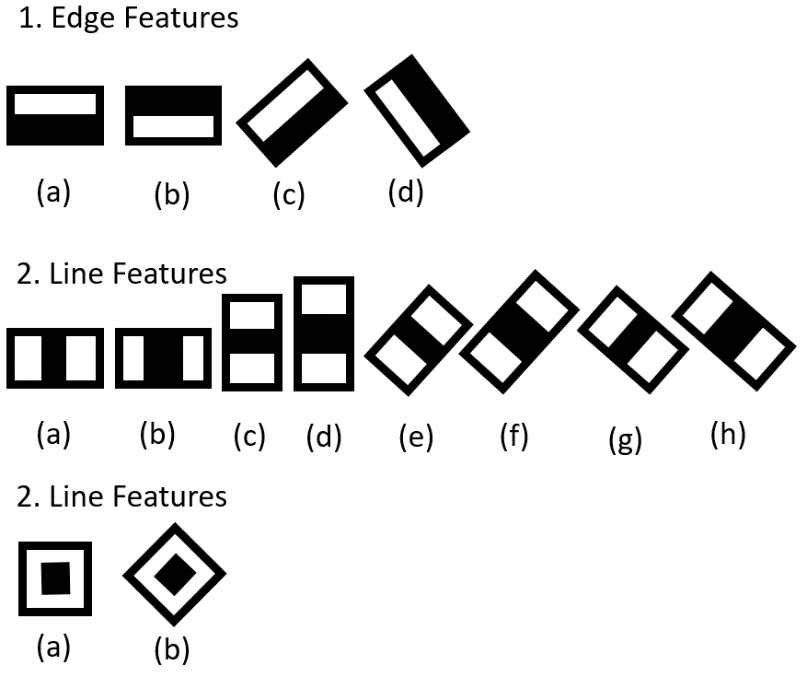
Bằng cách sử dụng các điểm hình chữ nhật như một bộ lọc để phát hiện các đặc điểm khác nhau của khuôn mặt như mắt và các nốt như trong hình hình 3.2. Các cửa sổ trượt hình chữ nhật được chạy lần lượt trên hình ảnh và tổng số pixel nằm trong phần màu trắng được trừ cho tổng số pixel nằm trong phần màu đen.



**Hình 3.2 Đặc trưng hình chữ nhật trong phương pháp Haar-Cascade**

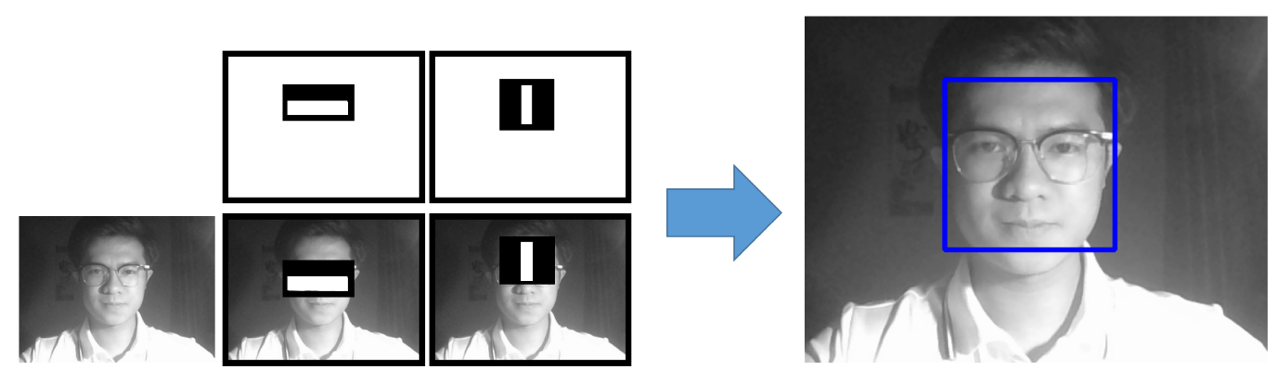
Trong đó:

1. Là các bộ lọc bắt các cạnh trong ảnh
2. Bắt các đường thẳng trong ảnh
3. Về đặc trưng 4 hình vuông

Hoặc các đặc trưng được nằm gọn trong trung tâm một vùng như trong hình 3.3

**Hình 3.3 Các đặc trưng hình chữ nhật khác trong phương pháp Haar-Cascade**

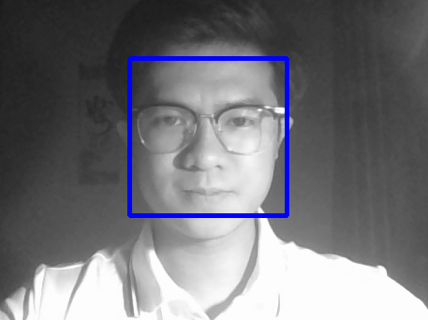
Các đặc trưng Haar cho phép phát hiện các khuôn mặt trong ảnh một cách nhanh chóng, thời gian thực và không phụ thuộc vào vị trí hoặc tỉ lệ ảnh. Haar-cascade cũng có thể được sử dụng để phát hiện nhiều khuôn mặt trong ảnh cùng một lúc. Các đặc điểm chính của từng khuôn mặt bao gồm lông mày, mắt, đầu mũi và miệng có thể được nhận ra một cách hiệu quả, và biểu hiện trên khuôn mặt có thể được phát hiện bằng các đường viền hình chữ nhật cho phù hợp, những đường viền này được xây dựng bởi các điểm đặc trưng ở cạnh của mọi mặt, bao gồm cả mặt trên và mặt dưới, xác định chiều rộng dọc, ngoài cùng bên phải và ngoài cùng bên trái, xác định chiều ngang của hình ảnh khuôn mặt. Để tránh bỏ sót thông tin trên khuôn mặt đồng thời giảm nhiễu nền, các đường viền của hình chữ nhật định vị khuôn mặt sẽ được đặt là 3px.

****Tuy nhiên, cách áp dụng các bộ lọc này khác một chút so với các cửa sổ bộ lọc trong mạng nơ-ron tích chập, trong mạng nơ-ron tích chập bộ lọc chiếm toàn bộ cửa sổ trượt, trong khi ở đặc trưng Haar, bộ lọc chỉ chiếm một phần trong cửa sổ trượt thôi. Hình 3.3 minh hoạ một ví dụ về một khuôn mặt đã được phát hiện dựa trên phương pháp Haar-Cascade với các cửa sổ trượt và được tô viền xung quanh khuôn mặt sau khi phát hiện khuôn mặt.

**Hình 3.4 Phát hiện khuôn mặt bằng phương pháp Haar-Cascade**

### **5.7 Tiền xử lý ảnh**

Sau phát hiện khuôn mặt trong ảnh đầu vào dựa trên phương pháp Haar-Cascade thì việc thực hiện nhận diện cảm xúc là hoàn toàn khả thi. Một ảnh mới (chỉ có khuôn mặt) sẽ được cắt ra để làm hình ảnh đầu vào cho bước nhận diện tiếp theo. Việc cắt hình ảnh khuôn mặt sẽ làm giảm bớt các chi tiết dư thừa trong ảnh, nâng cao hiệu suất nhận diện.







**Hình 3.5 Tiền xử lý hình ảnh đầu vào**

Tuy nhiên, trong quá trình thực nghiệm, các kết quả cho thấy việc nhận diện cảm xúc vẫn chưa thực sự hiệu quả một phần là do chất lượng ảnh đầu vào chưa tốt (quá tối, hoặc nhiễu, …), một phần là do kích thước hình ảnh đầu vào khác nhau, nên kích thước ảnh khuôn mặt sau khi được phát hiện cũng sẽ khác nhau. Do đó, cần phải tiến hành thêm bước tiền xử lý để chuẩn hóa các ảnh khuôn mặt đầu vào trước khi tiến hành nhận diện. Một số thao tác tiền xử lý được thực hiện trong lược đồ đề xuất bao gồm:

* Nâng cấp hình ảnh (dựa trên việc cân bằng histogram): Nhằm giảm sự ảnh hưởng do chiếu sáng (chói), thiếu ánh sáng (ảnh tối), …, các giải thuật xử lý ảnh thường nhạy cảm với ánh sáng, cùng nội dung ảnh nhưng với các điều kiện ánh sáng khác nhau có thể làm sai lệch kết quả xử lý. Do đó, cân bằng sáng ở bước tiền xử lý là một trong những cách giúp làm giảm các ảnh hưởng này.
* Giảm nhiễu với bộ lọc Gaussian: nhiễu gauss có được do bản chất rời rạc của bức xạ (hệ thống ghi ảnh bằng cách đếm các photon-lượng tử ánh sáng). Mỗi pixel trong ảnh nhiễu là tổng giá trị pixel đúng và pixel ngẫu nhiên . Lọc gaussian được thực hiện bằng cách nhân chập ảnh đầu vào với một ma trận lọc Gauss sau đó cộng chúng lại để tạo thành ảnh đầu ra.
* Xoay ảnh dựa trên việc xác định mũi là trung tâm khuôn mặt, thay đổi kích thước ảnh cho phù hợp với kích thước đầu vào của bộ nhận diện (ảnh được chuẩn hóa về kích thước 48x48), … Hình 3.4 mô phỏng hình ảnh khuôn mặt sau khi được tiền xử lý.

### **5.8 Nhận diện cảm xúc**

#### **5.8.1 Bộ dữ liệu đào tạo**

Bộ cơ sở dữ liệu ảnh là một trong các thành phần quan trọng hàng đầu trong các phương pháp học máy nói chung, được sử dụng để phục vụ cho quá trình tính toán tham số và huấn luyện, tinh chỉnh các mô hình. Thông thường, bộ dữ liệu càng lớn và càng được chọn lọc tỉ mỉ cẩn thận thì độ chính xác của mô hình càng được cải thiện.

Bộ dữ liệu đào tạo FER 2013 (Facial Emotion Recognition) Tập dữ liệu nguồn mở được tạo ra cho một dự án bởi PierreLuc Carrier và Aaron Courville, được chia sẻ công khai trong cuộc thi Kaggle (2013). Bộ dữ liệu FER 2013 bao gồm 35.887 ảnh xám: hình ảnh khuôn mặt kích thước 48x48 pixel từ nhiều góc độ khác nhau. Hình ảnh được phân loại thành một trong bảy lớp thể hiện cảm xúc khuôn mặt khác nhau, tất cả được gán nhãn từ 0 – 7 (0 = Giận dữ, 1 = Ghê tởm, 2 = Sợ hãi, 3 = Vui vẻ, 4 = Buồn, 5 = Ngạc nhiên, 6 = Bình thường) [28]. Bảng 2.1 mô tả thông tin bộ CSDL FER 2013.

Bảng 3.1 Thông tin chi tiết số lượng ảnh và cảm xúc trong bộ CSDL FER2013

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CSDL** | **Tổng số ảnh** | **Bộ ảnh huấn luyện** | **Bộ ảnh kiểm thử** | **Kích thước** | **Số lượng trạng thái** |
| FER2013 | 32298 | 28709 | 3589 | 48x48 | 7 |

Bảng 3.2 Thông tin số lượng dữ liệu đào tạo theo từng loại cảm xúc trong bộ CSDL FER2013

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Loại cảm xúc** | Giận dữ | Ghê tởm | Sợ hãi | Vui vẻ | Buồn | Ngạc nhiên | Bình thường |
| **Tổng số ảnh đào tạo** | 4462 | 492 | 4593 | 8110 | 5483 | 3586 | 5572 |

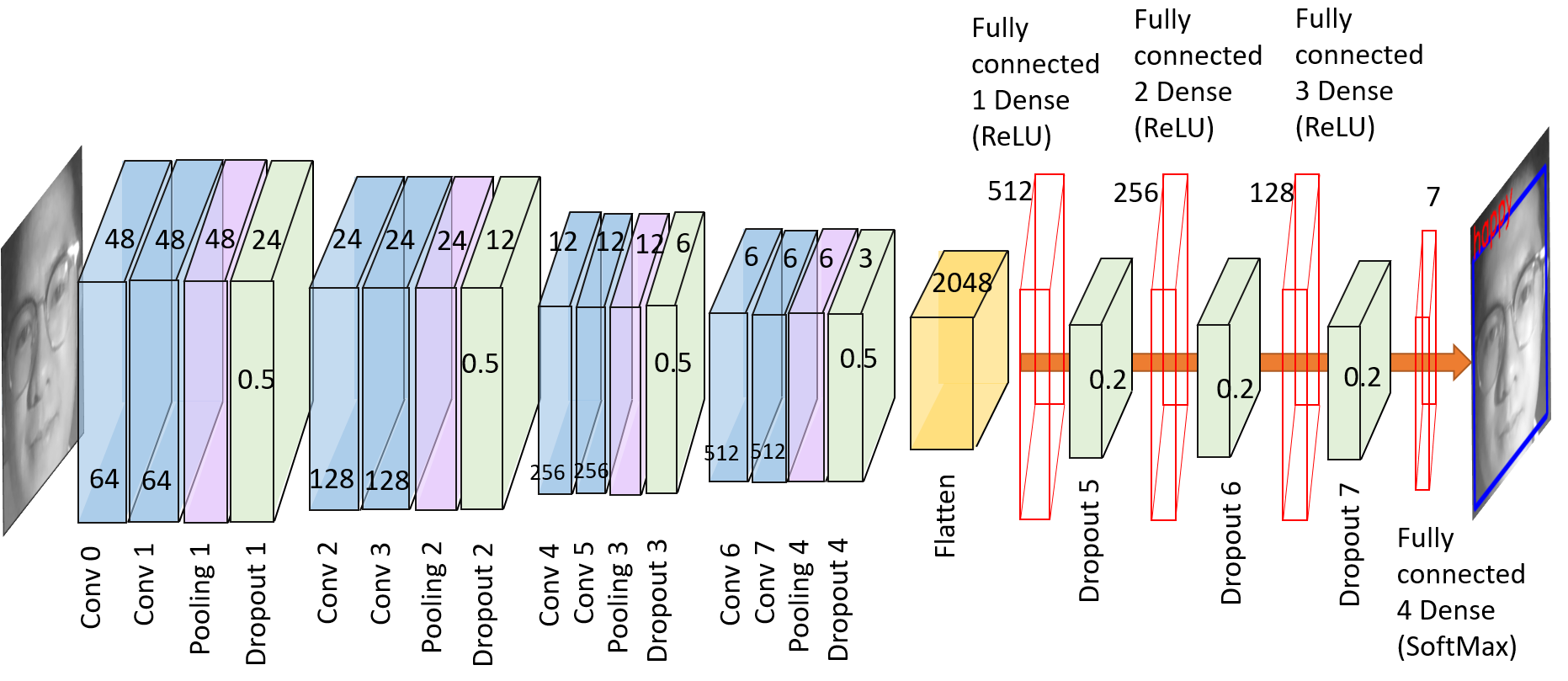
Bảng 3.3 Thông tin số lượng dữ liệu kiểm thử theo từng loại cảm xúc trong bộ CSDL FER2013

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Loại cảm xúc** | Giận dữ | Ghê tởm | Sợ hãi | Vui vẻ | Buồn | Ngạc nhiên | Bình thường |
| **Tổng số ảnh kiểm thử** | 491 | 55 | 528 | 879 | 594 | 416 | 626 |

#### **5.8.2 Xây dựng mô hình nhận diện cảm xúc**

Sau khi hình ảnh khuôn mặt đã được tiền xử lý và chuẩn hoá, giai đoạn tiếp theo trong lược đồ đề xuất sẽ là việc nhận diện cảm xúc từ thông tin hình ảnh đầu vào. Trong nghiên cứu này, chúng tôi đề xuất một mô hình học sâu mạng tích chập CNN dựa trên mô hình gốc của Kuo [22] do sự vượt trội về hiệu suất và độ chính xác của nó so với các cách tiếp cận khác Mô hình nhận diện đề xuất dưới đây được xây dựng dựa trên ý tưởng của mạng VGG16 như đã trình bày ở phần trước. Hình 3.5 minh hoạ một cách chi tiết các lớp của mô hình nhận diện, bao gồm ba khối chính như sau:

* Khối thứ nhất chứa 2 lớp tích chập mỗi lớp gồm 64 bộ lọc (channel); mỗi bộ lọc có kích thước cỡ 3 x 3 và kích thước ảnh đầu vào của bộ lọc có kích thước 48x48x1. Theo sau đó là hai lớp tổng hợp (pooling) có kích cỡ 2x2, bước nhảy là 2x2 và lớp dropout có tỷ lệ là 0.5 nhằm loại bỏ một vài trường hợp trong quá trình huấn luyện mạng. Việc bỏ các điểm đầu vào được thực hiện bằng cách lấy ngẫu nhiên nhưng đảm bảo một ngưỡng xác suất nào đó. Việc bổ sung thêm lớp dropout nhằm tránh trường hợp overfiting trong quá trình huấn luyện.
* Khối thứ hai có cấu trúc tương tự như khối thứ nhất bao gồm 2 lớp tích chập gồm 128 bộ lọc cỡ 3x3, một lớp tổng hợp pooling cỡ 2x2 với bước nhảy 2x2 và cuối cùng là một lớp dropout với tỷ lệ 0.5. Tuy nhiên, khác với khối thứ nhất, kích thước ảnh đầu vào bộ lọc khối thứ 2 sẽ giảm một nửa còn 24x24 để giảm độ phức tạp của thuật toán và tăng độ chính xác về việc trích chọn đặc trưng của ảnh.
* Khối thứ ba có cấu trúc tương tự hai khối trước với kích thước đầu vào được giảm còn 12x12. Trong đó, hai lớp tích chập trong khối được tăng lên 256 bộ lọc nhằm tăng cường độ phức tạp cho mô hình cho phù hợp với số lượng dữ liệu đầu vào điều này giúp hạn chế tình trạng mô hình chưa khớp (under-fitting) cho mô hình.
* Khối thứ tư về cơ bản cũng có cấu trúc tương tự như ba khối trước. Kích thước ảnh đầu vào cũng được tiếp tục giảm đi một nửa còn 12x12. Ngoài ra, hai lớp tích chập trong khối này được tăng cường số lượng kênh lên là 512 đồng thời bổ sung thêm lớp flatten nhằm làm phẳng dữ liệu và kết hợp các đặc trưng của ảnh đề có được đầu ra cho mô hình.
* Khối cuối cùng bao gồm các lớp kết nối đầy đủ (fully connected layer) gồm 4 lớp. Lớp đầu tiên có 512 nơ-ron, trong đó sử dụng hàm kích hoạt ReLUs, các lớp kết nối đầy đủ phía sau lần lượt là 256 và 128 nơ-ron. Lớp kết nối đầu đủ sau cùng gồm 7 nơ-ron và sử dụng hàm softmax làm hàm kích hoạt để phân loại các biểu cảm bao gồm: Tức giận, ghê tởm, sợ hãi, vui vẻ, buồn, ngạc nhiên, bình thường.

Thông tin chi tiết về các lớp trong các khối của mô hình mạng nơ-ron tích chập đề xuất được mô tả trong Bảng 3.4.

**Hình 3.6 Kiến trúc mạng tích chập cho nhận diện cảm xúc**

Bảng 3.4 Các tham số chi tiết cho mô hình đề xuất

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lớp | Số kernel | Kích thước mỗi kernel | Bước nhảy | Kích thước ảnh |
| Input | 0 | 0 | None | 48 x 48 x 1 |
| Conv2D-0 | 64 | 3 x 3 | 1 | 48 x 48 x 64 |
| Conv2D-1 | 64 | 3 x 3 | 1 | 48 x 48 x 64 |
| Pooling 0 | 0 | 2 x 2 | 2 | 24 x 24 x 64 |
| Dropout 0 |  | Dropout=0.5 |  | 24 x 24 x 64 |
| Conv2D-2 | 128 | 3 x 3 | 1 | 24 x 24 x 128 |
| Conv2D-3 | 128 | 3 x 3 | 1 | 24 x 24 x 128 |
| Pooling 1 | 0 | 2 x 2 | 2 | 12 x 12 x 128 |
| Dropout 1 |  | Dropout=0.5 |  | 12 x 12 x 128 |
| Conv2D-4 | 256 | 3 x 3 | 1 | 12 x 12 x 256 |
| Conv2D-5 | 256 | 3 x 3 | 1 | 12 x 12 x 256 |
| Pooling 2 | 0 | 2 x 2 | 2 | 6 x 6 x 256 |
| Dropout 2 |  | Dropout=0.5 |  | 6 x 6 x 256 |
| Conv2D-6 | 512 | 3 x 3 | 1 | 6 x 6 x 512 |
| Conv2D-7 | 512 | 3 x 3 | 1 | 6 x 6 x 512 |
| Pooling 3 | 0 | 2 x 2 | 2 | 3 x 3 x 512 |
| Dropout 3 |  | Dropout=0.5 |  | 3 x 3 x 512 |
| Flatten |  |  |  | 1 x 1 x 2048 |
| Dense-0 | 512 | activation='relu' |  | 1 x 1 x 512 |
| Dropout 4 |  | Dropout=0.2 |  | 1 x 1 x 512 |
| Dense-1 | 256 | activation='relu' |  | 1 x 1 x 256 |
| Dropout 5 |  | Dropout=0.2 |  | 1 x 1 x 256 |
| Dense-2 | 128 | activation='relu' |  | 1 x 1 x 128 |
| Dropout 6 |  | Dropout=0.2 |  | 1 x 1 x 128 |
| Dense-3 | 7 | activation='softmax' |  | 1 x 1 x 7 |
| Output | 0 | 0 | None | 1. x 1 x 7 |

#### **5.8.3 Môi trường đào tạo**

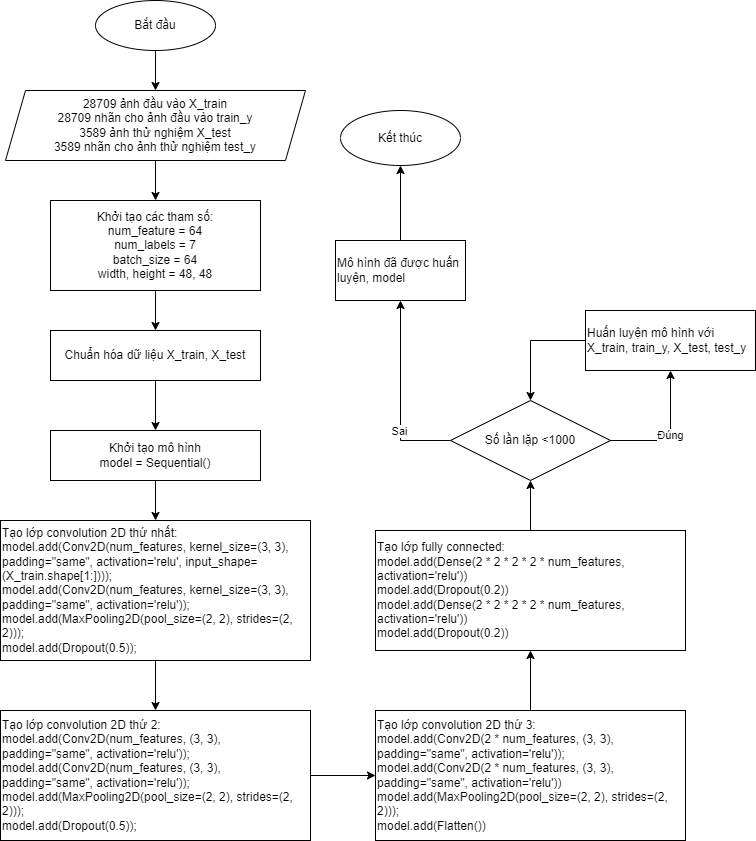
Mô hình đề xuất được huấn luyện với 28709 ảnh trong bộ CSDL FER 2013. Trong quá trình thực nghiệm, mô hình đã được triển khai với ngôn ngữ lập trình Python, quá trình huấn luyện được thực hiện trên Colaboratory hay còn gọi là Google Colab, một dịch vụ máy chủ điện toán đám mây của Google dành cho mục đích nghiên cứu. Dịch vụ này cho phép chạy các dòng code python thông qua trình duyệt, đặc biệt phù hợp với các lĩnh vực nghiên cứu: phân tích dữ liệu, học máy, trí tuệ nhân tạo... Colab cung cấp nhiều loại GPU, thường là Nvidia K80s, T4s, P4s và P100s, tuy nhiên người dùng không thể chọn loại GPU trong Colab, GPU trong Colab thay đổi theo thời gian. Vì là dịch vụ miễn phí, nên Colab sẽ có những thứ tự ưu tiên trong việc sử dụng tài nguyên hệ thống, cũng như giới hạn thời gian sử dụng, thời gian sử dụng tối đa lên tới 12 giờ, Bảng 3.5 mô tả cấu hình phần cứng Google Colab được sử dụng trong nghiên cứu này.

Bảng 3.5 Cấu hình phần cứng GoogleColab

|  |  |  |
| --- | --- | --- |
| **CPU** | **GPU** | **TPU** |
| Intel(R) Xeon(R) CPU @ 2.30 GHz và 13GB RAM | Tesla K80 12GB, GDDR5 VRAM,Intel(R) Xeon(R) CPU @ 2.20 GHz và 13GB RAM | TPU Cloud, Intel(R) Xeon(R) CPU @ 2.30 GHz và 13GB RAM |

#### **5.8.4 Đào tạo mô hình nhận diện cảm xúc**

* Bước 1: Đọc dữ liệu đầu vào từ bộ dữ liệu FER 2013 dưới dạng file .csv và khởi tạo các tham số cần thiết
* Bước 2: Duyệt lần lượt dữ liệu đầu vào và tiến hành chuẩn hóa dữ liệu
* Bước 4: Khởi tạo mô hình đào tạo
* Bước 5: Tiến hành đào tạo mô hình và lưu lại kết quả



Hình 3.7 Sơ đồ khối đào tạo mô hình nhận diện cảm xúc

# **TÀI LIỆU THAM KHẢO**

1. C. Darwin and P. Prodger. The Expression of the Emotions in Man and Animals. John Murray, 1998.
2. Y. Tian, T. Kanade, and J. F. Cohn. *Recognizing action units for facial expression analysis*. IEEE Transactions onPattern Analysis and Machine Intelligence, vol. 23, no. 2, 2001.
3. M. Bani, S. Russo, S. Ardenghi, G. Rampoldi, V. Wickline, S. Nowicki Jr, M. G. Strepparava *Behind the Mask: Emotion Recognition in Healthcare Students*. Med.Sci.Educ. 2021.
4. M. Jeong, B. C. Ko. *Driver’s Facial Expression Recognition in Real-Time for Safe Driving*. Department of Computer Engineering, Keimyung University, Daegu 42601, Korea, 4 December 2018.
5. P. Ekman and W. V. Friesen. *Constants across cultures in the face and emotion*. Journal of Personality and Social Psychology*, vol. 17, no. 2, 124–129, 1971.*
6. Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. *A survey of affect recognition methods: audio, visual, and spontaneous expressions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39–58, 2009.
7. S. Li and W. *Deng. Deep facial expression recognition: a survey*. IEEE Transactions on Affective Computing, In press.
8. C. Shan, S. Gong, and P. W. McOwan. *Facial expression recognition based on local binary patterns: a comprehensive study*. Image and Vision Computing, vol. 27, no. 6, pp. 803–816, 2009.
9. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. *The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression*. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern RecognitionWorkshops, pp. 94–101, San Francisco, CA, USA, July 2010
10. D. Matsumoto. *More evidence for the universality of a contempt expression*. Motivation and Emotion, vol. 16, no. 4, pp. 363–368, 1992.
11. R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn. *Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 1, pp. 38–52, 2011.
12. A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon. *Emotion recognition in the wild challenge 2014: baseline, data and protocol*. In Proceedings of the 16th International Conference on Multimodal Interaction, pp. 461–466, ACM, Istanbul Turkey, November 2014.
13. J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju. *Attention mechanism-based CNN for facial expression recognition*. Neurocomputing, vol. 411, pp. 340–350, 2020.
14. K. Simonyan and A. Zisserman. *Very deep convolutional networks for large-scale image recognition*. 2014, https:// arxiv.org/abs/1409.1556.
15. C. Szegedy, W. Liu, Y. Jia et al. *Going deeper with convolutions*. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, Boston, MA, USA, June 2015.
16. A. Jahandad, S. M. Sam, K. Kamardin, N. N. Amir Sjarif, and N. Mohamed. *Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3*. Procedia Computer Science, vol. 161, pp. 475–483, 2019.
17. K. He, X. Zhang, S. Ren, and J. Sun. *Deep residual learning for image recognition*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, Las Vegas, NV, USA, June 2016.
18. I. Allen and J. Seaman. *Digital compass learning: distance education enrollment report 2017*. Babson Survey Research Group, Babson Park, MA, USA, 2017.
19. E. Dolan, E. Hancock, and A. Wareing. *An evaluation of online learning to teach practical competencies in undergraduate health science students*. The Internet and Higher Education, vol. 24, pp. 21–25, 2015.
20. A.B.Shetty , Bhoomika , Deeksha , J.Rebeiro , Ramyashree. *Facial Recognition using Haar Cascade and LBP Classifiers*. Journal Pre-proof, 28 July 2021.
21. P. Ekman and W. V. Friesen. A new pan cultural facial expression of emotion . Motivation and Emotion, vol. 10, no. 2, pp. 159–168, 1986.
22. C. M. Kuo, S. H. Lai, and M. Sarkis. *A compact deep learning model for robust facial expression recognition*. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
23. Dumoulin, V., & Visin, F. (2016). *A guide to convolution arithmetic for deep learning.*
24. Samer, C. H., Rishi, K., & Rowen. *Image Recognition Using Convolutional Neural Networks. Cadence Whitepaper*, 1–12, 2015.
25. Reinhard Klette. *Concise Computer Vision*. Springer, 2014.
26. T. Szandała *Review and Comparison of Commonly Used Activation. ​* Wroclaw University of Science and Technology Wroclaw, Poland.
27. H.K. Jabbar, R.Z. Khan. *Method To Avoid Over-Fitting And Under-Fitting In Supervised Machine Learning (Comparative Study).* Computer Science, Communication & Instrumentation Devices, 2015.
28. S. Turabzadeh, H. Meng, R. Swash, M. Pleva, and J. Juhar, *Facial Expression Emotion Detection for Real-Time Embedded Systems*, Technologies, vol. 6, no. 1, p. 17, Jan. 2018.
29. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Fei-Fei, L. (2015). *ImageNet Large Scale Visual Recognition Challenge*. International Journal of Computer Vision, 115(3), 211–252.
30. Q. Guan, Y. Wang , B. Ping , D. Li , J. Du, Y. Qin, H. Lu, X. Wan, J. Xiang . *Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study.* Ivyspring International Publisher, 2019.
31. J. Bjorck, C. Gomes, B. Selman, K. Q. Weinberger. *Understanding Batch Normalization.* Cornell University, 2018.
32. D. Anderson, G. McNeill. *Artificial Neural NetWorks Technology.* Kaman Sciences Corporation. DACS State-of-the-Art Report, 1992.
33. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner. *Gradient-based learning applied to document recognition*, 1998.
34. R. Mustafina, M. Ilina, I. Shcherbakova. *Emotions and their Effect on Learning.* Revista internacional de filosofía y teoría social cesa-fces-universidad del zulia. maracaibo-venezuela issn 1316-5216 / issn-e: 2477-9555*.* 2020.