

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

Phạm Quang Huy¹

¹K68B, Khoa công nghệ thông tin

GVHD: TS. Đặng Thành Trung - Khoa Công Nghệ Thông Tin, Trường Đại học Sư Phạm Hà Nội

Tóm Tắt: 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.

Từ khóa: Giáo dục trực tuyến; nhận diện cảm xúc; mạng nơ ron tích chập.

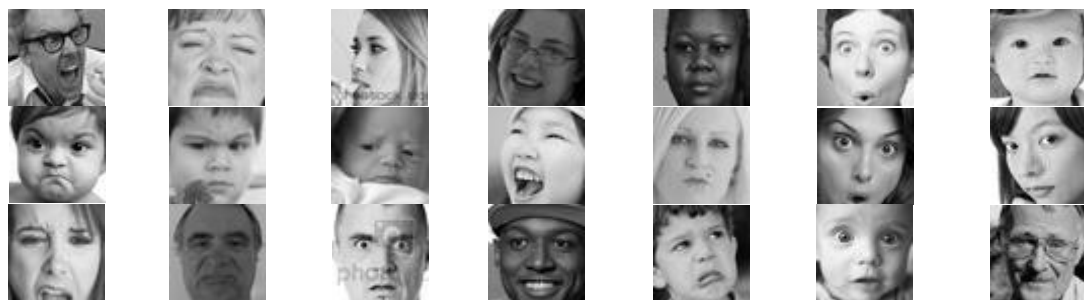
1. Mở đầu

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 [23], chúng tôi đã đề xuất một phương pháp cho phép theo dõi cảm xúc của sinh viên theo thời gian thực trong các lớp học trực tuyến và đảm bảo rằng các nhãn cảm xúc được phản hồi thông qua nét mặt có thể cung cấp cho giáo viên một công cụ đánh giá khách quan kịp thời, nâng cao chất lượng và hiệu quả của giáo dục trực tuyến [24].

Biểu cảm trên khuôn mặt là một trong những dấu hiệu phổ biến và tự nhiên nhất để con người truyền tải trạng thái cảm xúc [5] và ý nghĩ của họ [1, 2]. Có rất nhiều ứng dụng liên quan trực tiếp đến vấn đề này như: đánh giá sức khỏe [3], hỗ trợ lái xe, giao tiếp, ... [4]. Đã có nhiều nghiên cứu chỉ ra tầm quan trọng của việc ứng dụng nhận diện cảm xúc trong các lĩnh vực [22], tuy nhiên nhận diện cảm xúc sinh viên dựa trên biểu cảm khuôn mặt trong giáo dục trực tuyến là một đề tài còn mới mẻ trong nước và được đề ý trong khoảng một năm trở lại đây do ảnh hưởng của dịch bệnh nghiêm trọng.

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. 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

tuyên. 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.



Giận dữ Ghê tởm Sợ hãi Vui vẻ Buồn bã Ngạc nhiên Bình thường

Hình 1. Một số hình ảnh được gán nhãn cảm xúc trong CSDL FER2013

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 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. 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. Phương pháp nhận diện cảm xúc tự động được giới thiệu dựa trên một mô hình mạng tích chập là phương pháp đánh giá một cách khách quan, mô hình cho phép nhận diện bảy loại cảm xúc khác nhau của con người qua đó nhằm đánh giá chất lượng lớp học thông qua biểu đồ thống kê chỉ số cảm xúc của sinh viên trong lớp học theo thời gian thực.

Dựa trên hoàn cảnh dịch bệnh nghiêm trọng và mối liên kết chặt chẽ giữa nét mặt và cảm xúc của một người [1, 2], nhận diện cảm xúc sinh viên dựa trên biểu cảm khuôn mặt là một đề tài còn non trẻ và đầy thách thức. Tuy nhiên, để làm nên sự khác biệt đó là việc phát hiện khuôn mặt dựa trên phương pháp Haar-Cascade [20], kết hợp mô hình nhận diện cảm xúc dựa trên mạng nơ-ron tích chập [23] và ứng dụng thành công và lớp học trực tuyến tại Khoa Công Nghệ Thông Tin, Trường Đại học Sư Phạm Hà Nội là một nghiên cứu hoàn toàn mới.

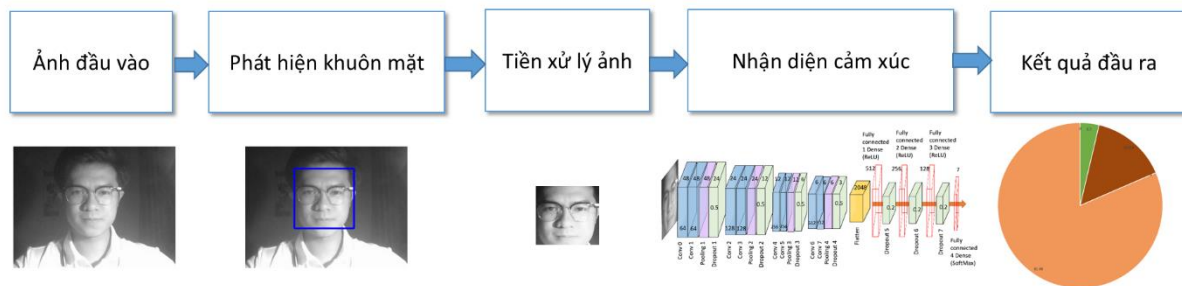
2. Nội dung nghiên cứu

2.1 Tổng quan về ứng dụng học sâu trong nhận diện cảm xúc sinh viên trong giáo dục trực tuyến

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 [24].

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 2 minh họa một cách trực quan các bước của lược đồ.



Hình 2. Lược đồ phương pháp đề xuất

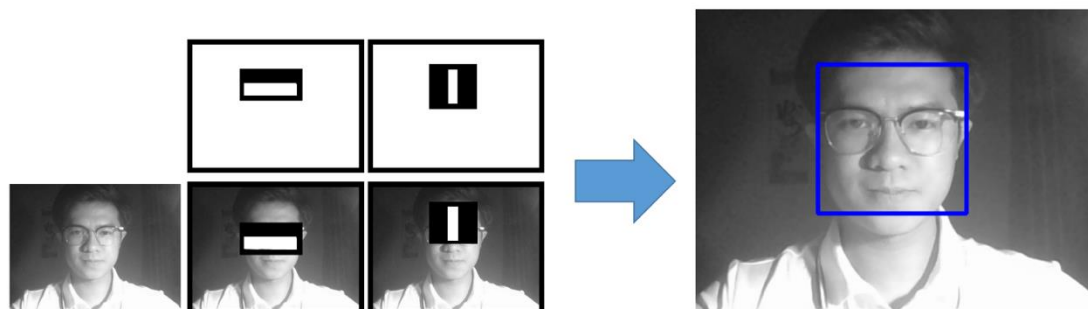
2.3 Hình ảnh đầu vào lớp học trực tuyến

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 lý 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.

2.4 Phát hiện khuôn mặt

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. 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. 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. Hình 3 minh họa một ví dụ về một khuôn mặt

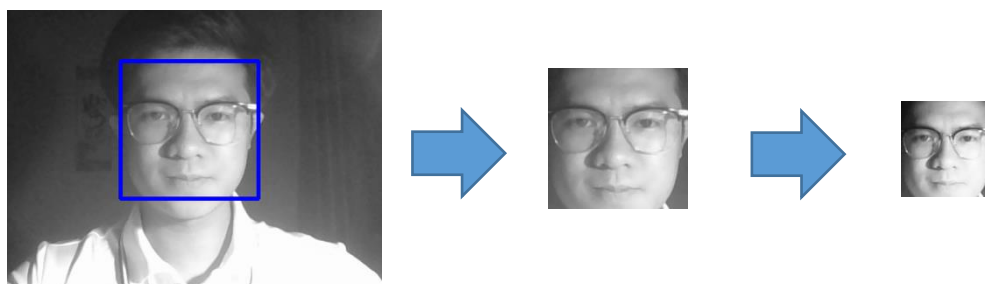
đã được phát hiện dựa trên phương pháp Haar-Cascade và được tô viền xung quanh khuôn mặt.



Hình 3. Phát hiện khuôn mặt bằng phương pháp Haar-Cascade

2.5 Tiền xử lý hình 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 4. 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 hoá 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), giảm nhiễu với bộ lọc Gaussian, 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 hoá về kích thước 48x48), ... Hình 4 mô phỏng hình ảnh khuôn mặt sau khi được tiền xử lý.

2.6 Xây dựng mô hình nhận diện cảm xúc

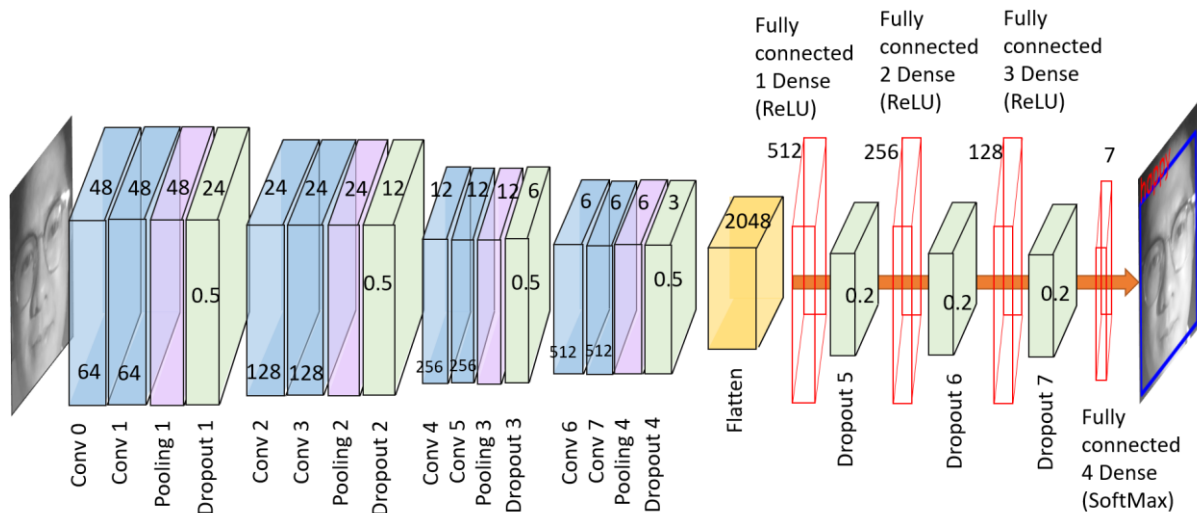
Sau khi hình ảnh khuôn mặt đã được tiền xử lý và chuẩn hoá, giai đoạn tiếp theo trong lược đồ đề xuất sẽ là việc nhận diện cảm xúc từ thông tin hình ảnh đầu vào. Trong nghiên cứu này, chúng tôi đề xuất một mô hình học sâu mạng tích chập CNN dựa trên mô hình

gốc của Kuo [22] do sự vượt trội về hiệu suất và độ chính xác của nó so với các cách tiếp cận khác. Hình 5 minh họa một cách chi tiết các lớp của mô hình nhận diện.

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 1.

Bảng 1. 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



Hình 5. Kiến trúc mạng tích chập cho nhận diện cảm xúc

3. Kết quả

3.1 Bộ dữ liệu tập huấn

Bộ dữ liệu FER2013 được sử dụng để đào tạo mô hình nhận diện cảm xúc, bộ dữ liệu bao gồm các ảnh đa mức xám có kích thước 48x48. Trong CSDL này, hình ảnh khuôn mặt đã được cắt bỏ phần ảnh nền dư thừa xung quanh và khuôn mặt được căn giữa hình ảnh. Các hình ảnh được gán nhãn với bảy loại cảm xúc khác nhau: giận dữ, ghê tởm, sợ hãi, vui vẻ, buồn, ngạc nhiên, bình thường. Thông tin chi tiết về CSDL được mô tả trong Bảng 2.

Bảng 2. 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

3.2 Kết quả thử nghiệm và đánh giá

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. Bảng 3 mô tả cấu hình phần cứng Google Colab được sử dụng trong nghiên cứu này.

Bảng 3. 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

Để đánh giá mô hình đề xuất, chúng tôi sử dụng bộ ảnh kiểm thử từ bộ dữ liệu FER2013 như đã trình bày ở trên, các kết quả thực nghiệm thu được được mô tả trong Bảng 4. Kết quả đầu ra cho thấy có đến 3443 trên tổng số 3589 ảnh có kết quả dự đoán đúng, tỷ lệ chính xác là 95,9% đối với bộ dữ liệu ảnh kiểm thử FER2013, Bảng 5 minh họa một số ảnh cụ thể trong quá trình kiểm thử đối với bộ dữ liệu trên.

Bảng 4. Kết quả thí nghiệm kiểm tra mô hình với bộ dữ liệu kiểm thử

CSDL	Số lượng ảnh tập huấn	Số lượng ảnh kiểm thử	Số lượng kết quả đúng	Tỷ lệ chính xác	Thời gian trung bình (ms)
FER2013	28709	3589	3443	95,9%	56,76




Bảng 5. Một số kết quả thử nghiệm

Ảnh	Nhãn CSDL	Nhãn kết quả
	Vui vẻ	Vui vẻ
	Sợ hãi	Sợ hãi
	Tức giận	Tức giận
	Buồn	Buồn

3.3 Ứng dụng thực tế

Để kiểm tra hiệu quả của phương pháp được đề xuất trong các ứng dụng thực tế, chúng tôi đã ứng dụng thực tiễn với khuôn mặt của tác giả làm hình ảnh đầu vào trong thời gian thực và đưa mô hình mạng nơ-ron tích chập vào nhận dạng cảm xúc trong ảnh. Kết quả thực nghiệm cho thấy mô hình nhận dạng cảm xúc trong thời gian thực đạt hiệu quả tốt. (bổ sung thêm các khuôn mặt khác với nhiều người).

Bảng 6. Kết quả thực nghiệm với khuôn mặt tác giả trong thời gian thực

Ảnh			Thời gian nhận diện(s)	Nhãn kết quả
			0.15	Bình thường
			0.14	Ngạc nhiên
			0.12	Sợ hãi

Tiếp theo, chúng tôi đã sử dụng hình ảnh học trực từ một số lớp học trên ứng dụng Zoom và đưa mô hình mạng nơ-ron tích chập vào nhận dạng cảm xúc trong ảnh, đây là hình ảnh được chụp trước khi kết thúc lớp học người giáo viên đã có vài phát biểu trước khi kết thúc lớp học trong một bầu không khí vui vẻ. Chúng tôi đã tiến hành thực nghiệm thu thập thông tin hình ảnh trong một số môn của Khoa Công nghệ thông tin, trường ĐH Sư phạm Hà Nội. Các môn học được thực nghiệm bao gồm cả ngành Sư phạm Tin và Công nghệ thông tin. Các lớp học bao gồm chủ yếu là các bạn sinh viên năm thứ 2 và năm thứ 3. Trong một nghiên cứu của Toguc và Ozkara [25] có chỉ ra rằng, mức độ cảm xúc vui vẻ của sinh viên sẽ được cải thiện đáng kể trong vòng vài phút trước khi kết thúc bài giảng, do đó, các thực nghiệm của chúng tôi được thực hiện tại một thời điểm ngẫu nhiên giữa tiết học (từ phút 30 – 40, với tiết học có thời lượng 50 phút).

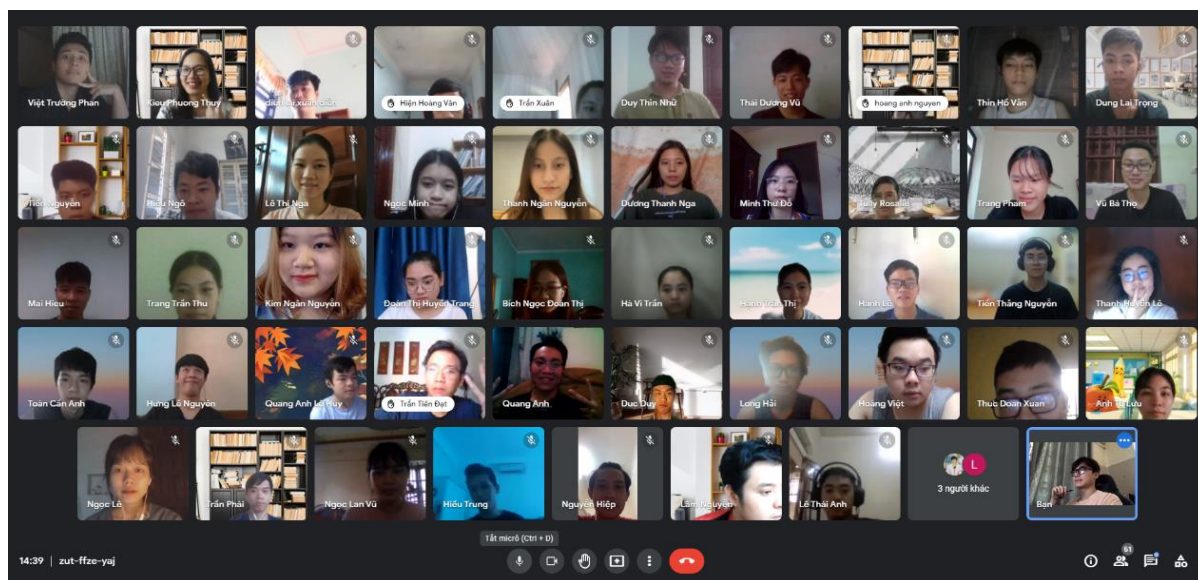
Bảng 7. Kết quả thử nghiệm tại lớp học Khoa Công nghệ thông tin, trường ĐH Sư phạm Hà Nội

Tên môn	Số lượng sinh viên	Số khuôn mặt phát hiện được	Số khuôn mặt được gán nhãn	Tỷ lệ nhận diện	Thời gian trung bình (ms)
Một số vấn đề xã hội của CNTT	48	27	27	56,2%	1817.491
Phần mềm nhúng và di động	47	15	15	32%	1413.18

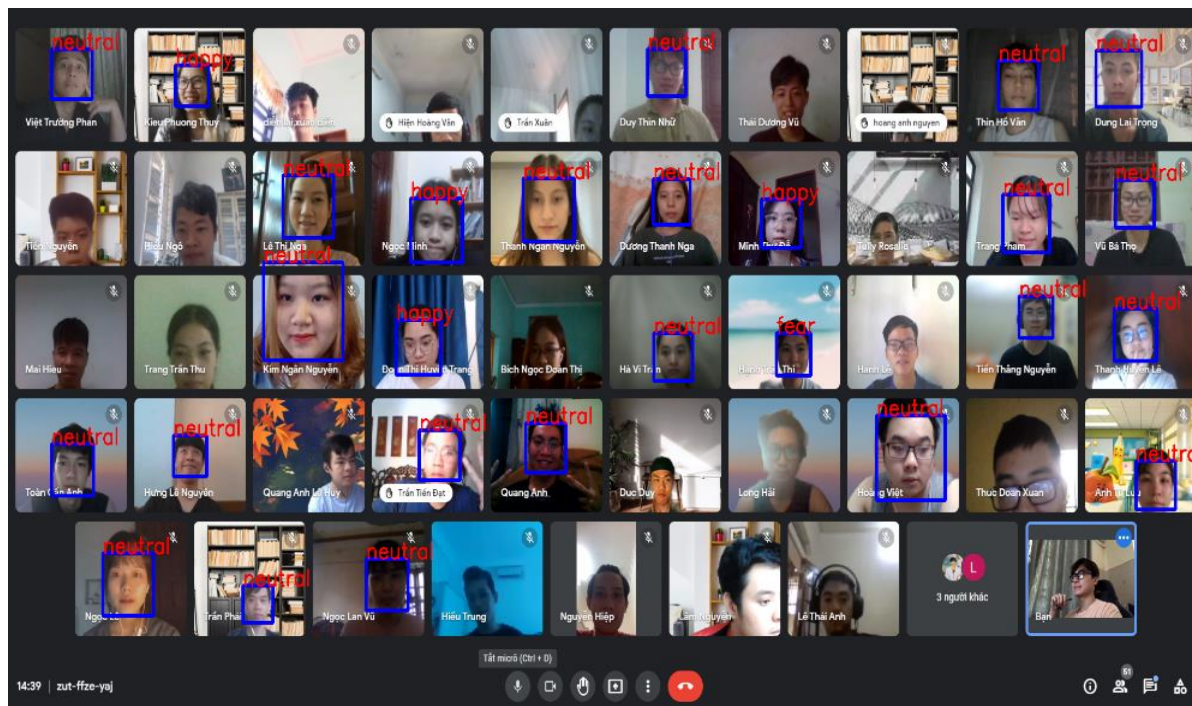
Phát triển phần mềm cho thiết bị di động K69	28	17	17	60,7%	1332.91
--	----	----	----	-------	---------

Hình 6 và Hình 7 minh họa một ví dụ về việc đánh giá cảm xúc của lớp học. Hầu hết các khuôn mặt đã được phát hiện và đánh dấu bằng các đường viền hình chữ nhật; các biểu cảm của các khuôn mặt được tiền xử lý một cách rõ nét và đã được nhận diện với các nhãn tương ứng. Trong tổng số 48 khuôn mặt, có 4 khuôn mặt được gắn nhãn “vui vẻ”, 22 khuôn mặt được gắn nhãn “bình thường” và 1 khuôn mặt được gắn nhãn “sợ hãi”. Các khuôn mặt chưa được tô viền và đánh nhãn, nguyên nhân là do các hình ảnh khuôn mặt này thiếu đi các chi tiết nét đặc trưng của khuôn mặt cơ bản hoặc do ánh sáng chưa đủ từ các thiết bị ghi hình của sinh viên.

Hình 8 minh họa thống kê về số lượng cảm xúc và tỷ lệ % cảm xúc nhận diện được tại một lớp học, từ đó chúng ta có thể quan sát tổng thể các cảm xúc một cách trực quan và phán đoán trạng thái cảm xúc của lớp cho phù hợp. Tuy nhiên, cần lưu ý rằng cảm xúc tổng thể của khuôn mặt có thể được đánh giá bằng nhiều phương pháp khác nhau, trong bài nghiên cứu này chúng tôi sử dụng phương pháp tìm ra giá trị lớn nhất của cảm xúc có trong kết quả dự đoán. Ở một số khuôn mặt được đánh dấu là “bình thường” có xác suất cao hơn nhiều so với “vui vẻ”, trong khi ở một số khuôn mặt được đánh nhãn là “vui vẻ” thì xác suất cảm xúc “bình thường” có thể chỉ thấp hơn một chút so với cảm xúc “vui vẻ”. Nhìn chung, kết quả của thí nghiệm này có thể hỗ trợ thuật lợi cho hoạt động của mô hình khi áp dụng vào môi trường thực tế.



Hình 6. Hình ảnh lớp học trực tuyến môn học một số vấn đề xã hội của CNTT

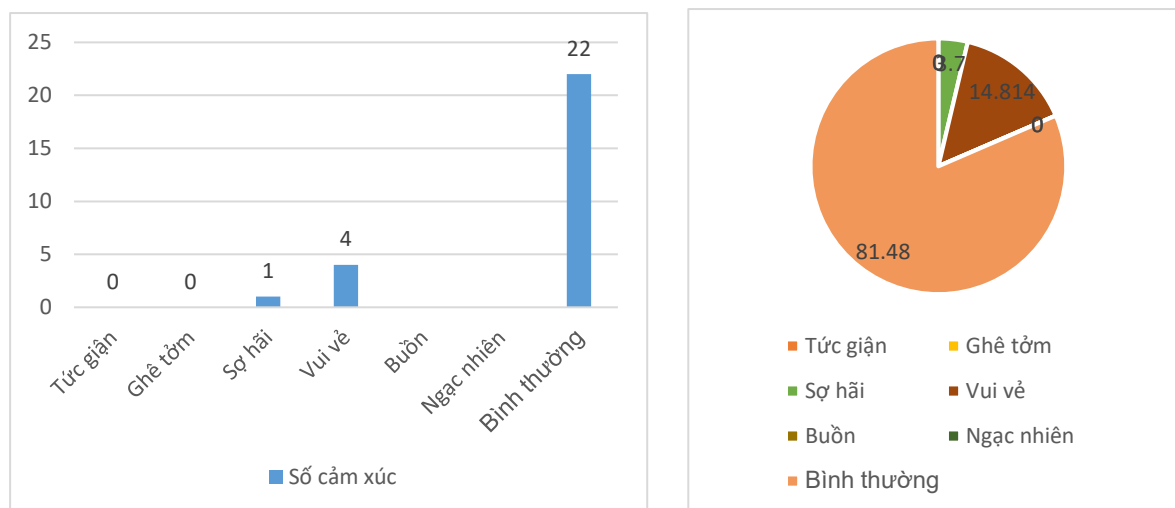


Hình 7. Nhận diện cảm xúc khuôn mặt trong lớp học trực tuyến môn học một số vấn đề xã hội của CNTT

3.5 Kết luận và hướng phát triển trong tương lai

Các kết quả thu được được trình bày dưới dạng biểu đồ một cách trực quan giúp giảng viên, người quản lý giáo dục có thể điều chỉnh phương pháp giảng dạy, kế hoạch giảng dạy sao cho phù hợp và nâng cao hiệu quả của việc giảng dạy trực tuyến. Từ ví dụ lớp học trực tuyến môn học “Một số vấn đề xã hội của CNTT” kết quả đầu ra được tổng hợp tại Hình 4.4 cho thấy có tổng cộng 27 nhãn cảm xúc, trong đó có 22 nhãn cảm xúc là bình thường, 4 nhãn cảm xúc là vui vẻ và 1 nhãn cảm xúc là sợ hãi, từ đó có thể nhận thấy rằng chỉ số cảm xúc bình thường và vui vẻ chiếm lần lượt là 81,48% và 14,81%, hai cảm xúc này thuộc nhóm cảm xúc có tác động tích cực đến quá trình học như đã trình bày ở chương trước đó. Kết luận rằng lớp học có chỉ số cảm xúc ảnh hưởng tích cực đến quá trình học tại thời điểm đánh giá cảm xúc.

Để đánh giá mô hình đề xuất, chúng tôi đã sử dụng bộ cơ sở dữ liệu hình ảnh chuẩn FER 2013 để thực nghiệm. Các kết quả thực nghiệm cho thấy, mức độ nhận diện cảm xúc với độ chính xác 95,9% đối với bộ CSDL FER2013. Các kết quả thu được cho thấy mức độ tin cậy của mô hình đề xuất là chấp nhận được và hoàn toàn có thể đáp ứng được các ứng dụng thực tế.



Hình 8. Biểu đồ đánh giá cảm xúc lớp học trực tuyến môn học “Một số vấn đề xã hội của CNTT”

Dựa trên các kết quả thực nghiệm, chúng tôi cũng đã tiến hành áp dụng mô hình vào môi trường thực tế. Một số môn học của Khoa Công nghệ thông tin, trường ĐH Sư phạm Hà nội được sử dụng làm môi trường thu thập và đánh giá. Các hình ảnh được thu thập từ 3 môn của 3 lớp. Tổng số 123 sinh viên tham gia 3 lớp học được thu thập trong đó 59 khuôn mặt chứa đầy đủ các đặc điểm đặc trưng của khuôn mặt nên có thể phát hiện một cách hiệu quả. Một số kết quả thực nghiệm cũng đã thu được và đã thể hiện được trên các lược đồ tương ứng. Các kết quả thực nghiệm cho thấy một kết quả tiềm năng và thú vị.

Hiện tại, kết quả nhận diện vẫn còn hạn chế và còn nhiều nhược điểm trong quá trình đào tạo mô hình cũng như thuật toán nhận diện khuôn mặt, do chất lượng hình ảnh chụp còn chưa đủ tốt, việc phát hiện hình ảnh khuôn mặt có tỷ lệ chưa cao và thuật toán nhận diện khuôn mặt cũng như nhận diện cảm xúc và đánh nhãn cho khuôn mặt còn có thể cải thiện hơn. Do đó, trong thời gian tới, việc nâng cấp khả năng phát hiện khuôn mặt trong điều kiện hạn chế: độ phân giải thấp, môi trường chụp hạn chế, ảnh bị mờ chữ lên mặt, ... và đào tạo mô hình cũng như cải thiện thuật toán sẽ được chúng tôi tiếp tục nghiên cứu nhằm tăng khả năng ứng dụng thực tế của mô hình đề xuất.

Ngoài ra, với số lượng lớn người tham gia các lớp học trực tuyến lớn, nhưng màn hình học trực tuyến tại mỗi thời điểm là hạn chế, do đó, không thể đảm bảo việc đánh giá được toàn bộ người đang học cùng một lúc. Một số giải pháp cũng đã được đề xuất và sẽ được giới thiệu trong các nghiên cứu tiếp theo.

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 on Pattern 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 Recognition Workshops, 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] R. Mustafina, M. Ilina, I. Shcherbakova. *Emotions and their Effect on Learning*. Revista internacional de filosofía y teoría social cesa-fces-universidad del zulía. maracaibo-venezuela issn 1316-5216 / issn-e: 2477-9555. 2020.

ABSTRACT

APPLICATION OF AI ON REAL-TIME EMOTION DETECTION STUDENT IN E-LEARNING

Summary: Online education is gradually becoming a new trend full of potential and challenges. Especially in the current serious situation of the COVID-19 epidemic, most schools are closed, online education is considered one of the most optimal solutions today. There are many previous studies that have shown that there is a strong relationship between a person's facial expressions and emotions. Therefore, to objectively assess the quality of online classes, we propose an automatic emotion recognition method based on a convolutional network (CNN). The model allows to identify seven different types of human emotions. The proposed method is experimentally based on two databases on emotion recognition, FER2013 and CK Plus. Experiment on three online classes including three classes of students of IT faculty, Hanoi National University of Education. The results show that the proposed model is not only effective with standard data sets, but also works well in different experimental environments.

Keywords: Online education; emotion recognition; Convolution Neural Network.

