

NHẬN BIẾT CẢM XÚC CON NGƯỜI SỬ DỤNG ĐẶC TRƯNG HAAR VÀ MẠNG NƠON TÍCH CHẬP

Nguyễn Đăng Bình¹, Nguyễn Hữu Hải^{*2}

¹Khoa CNTT, Trường Đại học Khoa học, Đại học Huế

²Trường TH&THCS Gio An, Gio Linh, Quảng Trị

*Email: haihuu@gmail.com

Ngày nhận bài: 24/01/2024; ngày hoàn thành phản biện: 01/02/2024; ngày duyệt đăng: 5/3/2024

TÓM TẮT

Tự động nhận dạng cảm xúc khuôn mặt (FER) là một phần của hệ thống tương tác người - máy, là lĩnh vực được nghiên cứu rộng rãi trong và ngoài nước. Các phương pháp FER hiện tại có thể được phân loại thành hai nhóm chính: các phương pháp tiếp cận dựa trên học máy thuần túy không ứng dụng mạng nơ-ron và các phương pháp tiếp cận dựa trên học sâu có ứng dụng mạng nơ-ron tích chập. Trong nghiên cứu này, chúng tôi tập trung nghiên cứu cách nhận diện khuôn mặt bằng phương pháp Haar của Viola-Jones, đồng thời cài đặt hệ thống mạng nơ-ron tích chập để phân loại cảm xúc khuôn mặt cũng như việc kết hợp phương pháp học máy truyền thống và mạng nơ-ron tích chập hiện đại. Chúng tôi tiến hành đối sánh kết quả với các nghiên cứu khác trong khoảng 5 năm trở lại đây trên cùng bộ dữ liệu được sử dụng rộng rãi FER2013.

Từ khóa: Haar, mạng CNN, nhận dạng cảm xúc, mạng nơ-ron tích chập.

1. MỞ ĐẦU

Nhận biết cảm xúc con người (FER) nằm trong số những ứng dụng quan trọng của lĩnh vực thị giác máy tính, đặc biệt trong bối cảnh ngày nay, khi nhu cầu giao tiếp của con người ngày càng phát triển. Cách thức giao tiếp ngày càng đa dạng và kèm theo đó là các thông tin cảm xúc. Phát hiện cảm xúc nhằm điều chỉnh hệ thống để phục vụ con người một cách tốt hơn. Việc nhận biết cảm xúc giúp cải thiện sự hiểu biết và tương tác giữa con người với con người, con người với máy tính. Xây dựng mô hình để nhận diện cảm xúc hiệu quả cũng đang là bài toán khó và hấp dẫn với nhiều nhà khoa học cũng như các nhà phát triển ứng dụng từ trước đến nay.

Vậy nên, chúng tôi chọn đề tài: "*Nhận biết cảm xúc con người sử dụng đặc trưng Haar và mạng nơ-ron tích chập*" để nghiên cứu. Mục tiêu và phạm vi nghiên cứu là tìm

hiểu cách FER bằng cách sử dụng Convolutional Neural Networks kết hợp bộ đặc trưng Haar, ứng dụng vào nhận biết cảm xúc con người. Cách phân bố các lớp mạng, điều chỉnh tham số để đạt được hiệu quả tốt nhất.

2. PHƯƠNG PHÁP NGHIÊN CỨU

Để nhận dạng cảm xúc, chúng ta thường sử dụng công nghệ về lĩnh vực CV với mục đích tự động nhận biết và mô tả hình ảnh một cách chính xác và hiệu quả. Ứng dụng CV sử dụng trí tuệ nhân tạo và học máy để xử lý những dữ liệu này một cách chính xác nhằm xác định đối tượng và nhận diện khuôn mặt, cũng như phân loại, đề xuất, giám sát và phát hiện.

2.1. Tổng quan các nghiên cứu gần đây về nhận dạng cảm xúc

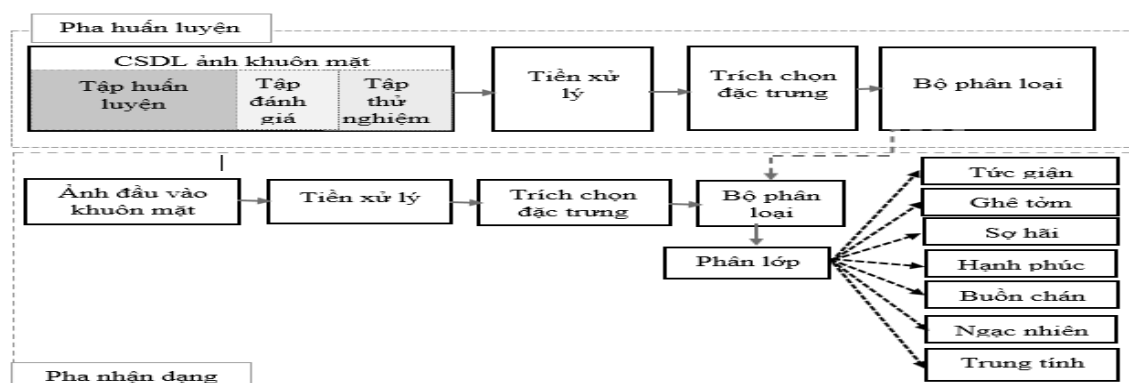
Đa số các nghiên cứu gần đây đều sử dụng Deep Learning (DL). [2] đã trích xuất một tập hợp các bản đồ đặc trưng sâu bằng mô hình CNN được huấn luyện trước từ các hình ảnh đầu vào, nơi các đặc điểm sâu cục bộ được thu thập dày đặc, [2] không chỉ cho thấy hiệu suất vượt trội trên dữ liệu quy mô lớn mà còn có khả năng học biểu diễn tính năng hiệu quả từ một tập dữ liệu tương đối nhỏ. Một số thì tạo ra chiến lược hợp nhất để tăng hiệu suất phân loại [1].

2.2. Một số phương pháp nhận dạng đối tượng

FER được chia thành nhiều hướng nghiên cứu theo các tiêu chí khác nhau, tuy nhiên ta có thể chia thành hai phương pháp chính là: có ứng dụng mạng nơ-ron nhân tạo và không sử dụng mạng nơ-ron nhân tạo.

2.2.1. Phương pháp không ứng dụng mạng nơ-ron nhân tạo

Hệ thống FER với phương pháp này thường qua các giai đoạn: tiền xử lý hình ảnh khuôn mặt, trích xuất đặc trưng và phân loại.



Hình 1. Mô tả hệ thống FER không sử dụng mạng nơ-ron nhân tạo

Tiền xử lý hình ảnh mục đích để cải thiện hiệu suất của hệ thống và được thực hiện các quy trình: căn chỉnh độ rõ, lọc nhiễu, chia tỷ lệ hình ảnh, điều chỉnh độ tương phản v.v...

Trích xuất đặc trưng là một giai đoạn quan trọng, nó phát hiện ra việc chuyển từ mô tả đồ họa sang mô tả dữ liệu ẩn, trích chọn những đặc trưng riêng nhất của hình ảnh, sau đó những mô tả dữ liệu này có thể được sử dụng làm đầu vào cho bài toán phân loại.

Phân loại là giai đoạn cuối cùng, để phân loại ra các loại cảm xúc trên khuôn mặt: hạnh phúc, buồn bã, bất ngờ, tức giận, sợ hãi, ghê tởm và bình thường. Sử dụng các phương pháp phân loại như: Cây quyết định, Support Vector Machine, Hidden Markov Model,...

2.2.2. Phương pháp có ứng dụng mạng nơ-ron nhân tạo

Phương pháp này thường có các bước chính thực hiện qua các giai đoạn: tiền xử lý hình ảnh, phân lớp sử dụng học sâu.

-*Tiền xử lý ảnh*: xử lý 1 số vấn đề của ảnh đầu vào hệ thống, xử lý trước quá trình training thông qua các bước thực hiện: Căn chỉnh khuôn mặt để phát hiện khuôn mặt, tăng dữ liệu hình ảnh đảm bảo đủ dữ liệu training, chuẩn hóa dữ liệu khuôn mặt. Sử dụng các phương pháp CNN, DBN, DAE, RNN, GAN...

-*Phân loại*: Trong phương pháp truyền thống bước trích xuất đặc trưng và bước phân loại tính năng là độc lập với nhau, trong Deep learning có thể thực hiện FER theo cách từ đầu đến cuối. Một lớp softmax được thêm vào cuối mạng để điều chỉnh lỗi lan truyền ngược, sau đó xác suất dự đoán của từng mẫu có thể được mạng trực tiếp xuất ra.

DL sử dụng các kỹ thuật học sâu hiệu quả nhờ cấu trúc sâu và khả năng tự học các tính năng quan trọng từ bộ định dạng và tạo ra kết quả đầu ra, giải quyết tốt bài toán này. DL chủ yếu được phát triển dựa trên nguyên lý kỹ thuật mạng nơ-ron nhân tạo.

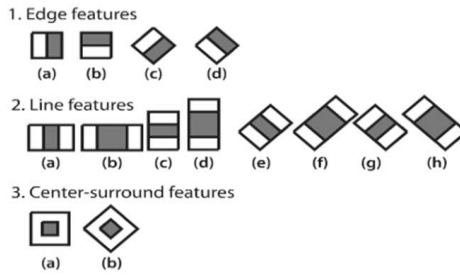
2.3. Mô hình kết hợp đặc trưng Haar và mạng nơ-ron tích chập

2.3.1 Đặc trưng Haar-like

Trích chọn đặc trưng là quá trình chuyển đổi dữ liệu đầu vào thành một tập hợp các đặc trưng có ý nghĩa cao hơn, nhằm giảm chiều dữ liệu, loại bỏ dữ liệu không quan trọng và tạo ra một biểu diễn tốt nhất của dữ liệu. Mục tiêu của trích chọn đặc trưng là giúp cho mô hình học máy hoặc thuật toán có khả năng học và tổng hợp thông tin hiệu quả để thực hiện các tác vụ như phân loại, nhận diện, phát hiện và dự đoán. Đầu vào là vec tơ biểu diễn, đầu ra là ma trận có kích thước $m \times n$. Khái niệm đặc trưng được sử dụng để biểu thị một phần thông tin liên quan để giải quyết các nhiệm vụ

nhất định. Các đặc trưng mô tả thuộc tính của đối tượng bao gồm: điểm ảnh, vùng ảnh,..

Đặc trưng Haar-like là một trong những phương pháp hiệu quả nhất để phát hiện đối tượng, được đề xuất bởi Paul Viola và Michael Jones [5]. Có 4 đặc trưng Haar-Like cơ bản được sử dụng và được chia làm 3 tập đặc trưng như sau: Đặc trưng cạnh (Edge features), đặc trưng đường thẳng (Line features), đặc trưng xung quanh tâm (Center-surround features).



Hình 2. Chi tiết các bộ lọc Haar

Về cơ bản đặc trưng Haar hoạt động bằng cách sử dụng nhiều bộ lọc Haar như Hình 2 qua nhiều lượt (cascade) để tạo thành hệ thống nhận diện hoàn chỉnh.

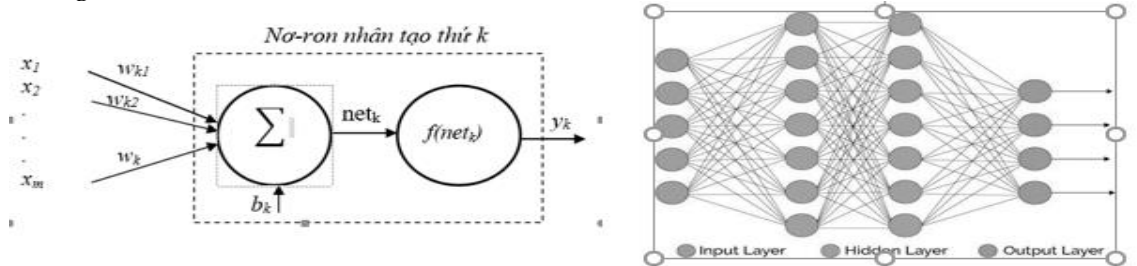


Hình 3. Mô tả bộ lọc Haar-like hoạt động

Để tăng tính hiệu quả và tốc độ nhận diện khuôn mặt Viola và Jones dùng thuật toán AdaBoost kết hợp các bộ phân loại yếu tạo thành bộ phân loại mạnh theo mô hình phân tầng.

2.3.2 Mô hình mạng nơ-ron nhân tạo

Mạng nơ-ron (Neural Networks - NNs) bao gồm các nơ-ron kết nối với nhau như được tìm thấy trong các bộ não sinh học. Thành phần là các nơ-ron và cách mà chúng kết nối với nhau.



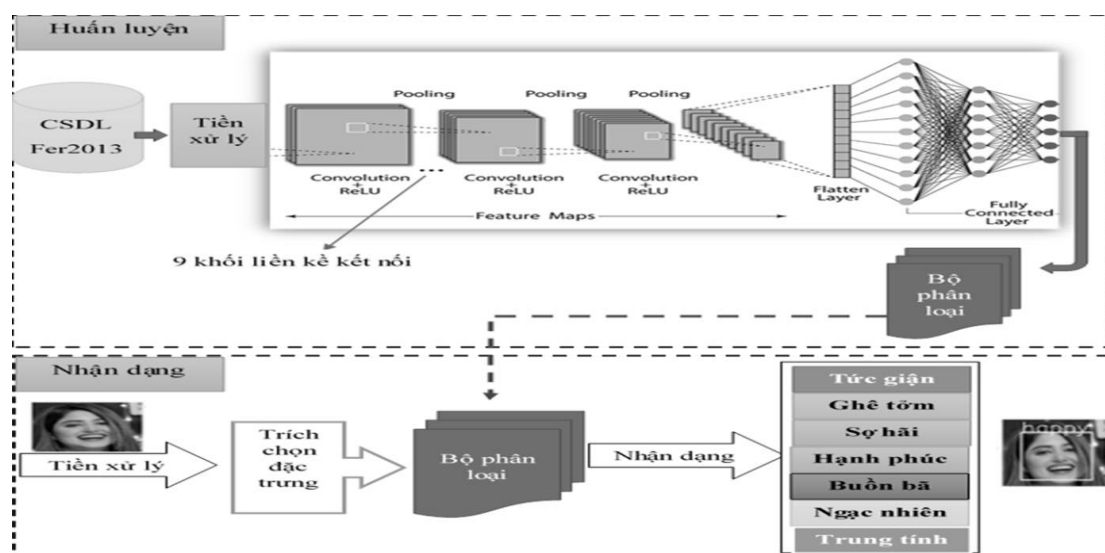
Hình 4. Mô hình nơ-ron và mạng nơ-ron nhân tạo

Mạng nơ-ron nhân tạo có khả năng tự động học cách biểu diễn đặc trưng từ dữ liệu đầu vào, quản lý độ phức tạp của mô hình, học từ dữ liệu lớn cũng như học tính toán phi tuyến khi dữ liệu không phải lúc nào cũng tuân theo các mối quan hệ tuyến tính đơn giản.

2.3.3 Giải pháp CNN kết hợp đặc trưng Haar

Hệ thống được chúng tôi triển khai phát hiện khuôn mặt người sử dụng bộ phân loại tăng Haar từ hình ảnh đầu vào đã cho và sau đó cắt khuôn mặt và chuẩn hóa hình ảnh thành 48x48 pixel. Đầu vào được truyền qua CNN và đầu ra cho kết quả là một lớp (0-6) tức là 0: 'tức giận', 1: 'ghê tởm', 2: 'sợ hãi', 3: 'hạnh phúc', 4: 'buồn', 5: 'ngạc nhiên', 6: 'trung tính'.

Hệ thống chúng tôi xây dựng gồm 9 khối liên kết nối, mỗi khối bao gồm một lớp tích chập, hàm kích hoạt ReLU và lớp chuẩn hoá Batch Normalization (Hình 5).



Hình 5. Mô tả giải pháp kết hợp của chúng tôi.

- Đối với lớp tích chập, được sử dụng để trích xuất các tính năng quan trọng, chỉ giữ lại những đặc trưng chính. Trong từng lớp tích chập chúng tôi cài đặt các thông số như: số kênh đầu vào, số kênh đầu ra, kích thước cửa sổ trượt, lề padding và thông số bias như Bảng 1.

- Hàm ReLU đảm bảo tính phi tuyến của mô hình huấn luyện sau khi đã thực hiện một loạt các phép tính toán tuyến tính qua các lớp tích chập.

- Lớp chuẩn hoá Batch Normalization là làm tăng độ chính xác của mô hình, cho kết quả test tốt hơn, chạy nhanh hơn, hội tụ sớm hơn.

Bảng 1. Tham số các lớp tích chập chúng tôi sử dụng

TT	Tên lớp	Số kênh đầu vào	Số kênh đầu ra	Kích thước bộ lọc	Padding	Chuẩn hoá batchnormal
1	convblock1	3	32	(3,3)	1	32
2	Convblock2	32	64	(3,3)	1	64
3	Convblock3	64	128	(3,3)	1	128
4	Convblock4	128	265	(3,3)	1	265
5	Convblock5	256	512	(1,1)	1	512
6	Convblock6	512	1024	(3,3)	1	1024
7	Convblock7	1024	1024	(3,3)	1	1024
8	Convblock8	1024	512	(3,3)	1	512
9	Convblock9	512	256	(3,3)	0	256

Chúng tôi cũng đưa lớp pooling vào sau các lớp tích chập 2, 5 và 7 để làm giảm kích ảnh đầu ra trong khi vẫn giữ được các phần quan trọng của ảnh đầu vào. Từ đó giảm được số lượng tham số cũng như tăng hiệu quả tính toán.

Sau 9 khối lớp kết nối, chúng tôi chèn vào hệ thống các lớp **Dropout**, **Avgpooling**, lớp làm phẳng **Flatten** và 2 lớp kết nối đầy đủ **Full Connected** với mục đích như sau:

- Lớp Dropout: việc tối ưu các tham số để làm cho giảm hàm mất mát, giảm tình trạng quá khớp overfitting (hiện tượng mạng nơron quá bám sát vào tập dữ liệu huấn luyện và không đáp ứng được với các tập dữ liệu mới). Lớp Avgpooling - gộp trung bình toàn cầu: thực hiện một kiểu lấy mẫu xuống cực đoan, trong đó bản đồ đặc trưng có kích thước chiều cao \times chiều rộng được ghép xuống thành mảng 1×1 bằng cách lấy trung bình của tất cả các phần tử trong mỗi bản đồ đặc trưng, trong khi độ sâu của bản đồ đặc trưng được giữ lại. Thao tác này thường chỉ được áp dụng một lần trước khi các lớp được kết nối đầy đủ. Ưu điểm là: giảm số lượng tham số có thể học được và cho phép CNN chấp nhận đầu vào có kích thước thay đổi. Lớp làm phẳng **Flatten**: được sử dụng để chuyển ma trận 2 chiều thành một vectơ tuyến tính (biến đổi dữ liệu từ không gian đa chiều thành không gian một chiều). Vectơ tuyến tính được đưa đến một lớp được kết nối đầy đủ để phân loại. Lớp kết nối đầy đủ (**Full connected**): là đầu ra từ lớp tích chập trích xuất các tính năng quan trọng và lớp tổng hợp làm giảm kích thước của hình ảnh vẫn giữ các tính năng quan trọng. Sau khi trải qua giai đoạn của lớp được kết nối đầy đủ, lớp cuối cùng là hàm kích hoạt **softmax** được sử dụng để lấy xác suất của đối tượng dựa trên đầu vào thuộc về các lớp cụ thể (phân loại).

3. KẾT QUẢ VÀ THẢO LUẬN

3.1 Mô tả bài toán và bộ dữ liệu

Bài toán được đặt ra là phải làm sao để khi đưa hình ảnh khuôn mặt một người bất kỳ vào, hệ thống phải tự động phân loại ra loại cảm xúc mà người này đang thể hiện là tương ứng với loại cảm xúc nào trong cơ sở dữ liệu.

Đầu vào là hình ảnh, video từ camera, đầu ra là trạng thái cảm tương ứng trên hình.

Tập dữ liệu chúng tôi sử dụng là FER2013, bao gồm 35.887 ảnh xám: hình ảnh khuôn mặt kích thước 48x48 pixel từ nhiều góc độ khác nhau. Gồm 8.989 ảnh 'Happy', 6.077 ảnh 'Sad', 6.198 ảnh 'Neutral', 4002 ảnh 'Suprised', 5121 ảnh 'Scared', 547 ảnh 'Disgust' và 4593 ảnh 'Angry'. Trong đó, hình ảnh đào tạo 28709, tập ảnh thử nghiệm 3589. Một số cảm xúc từ bộ dữ liệu FER-2013 được hiển thị trong Hình 6.



Hình 6. Tập dữ liệu mẫu FER 2013

3.2. Thực nghiệm và đánh giá kết quả

Phương pháp được đề xuất bao gồm chín lớp chập, ba lớp tổng hợp tối đa, một lớp tổng hợp trung bình, một lớp làm phẳng và hai lớp kết nối đầy đủ. Lớp đầu ra là softmax được sử dụng để tính toán kết quả từ các mạng nơ-ron tích chập. Thông qua kết quả thí nghiệm, phương pháp đề xuất của chúng tôi đạt được độ chính xác xác nhận là 68,88%.

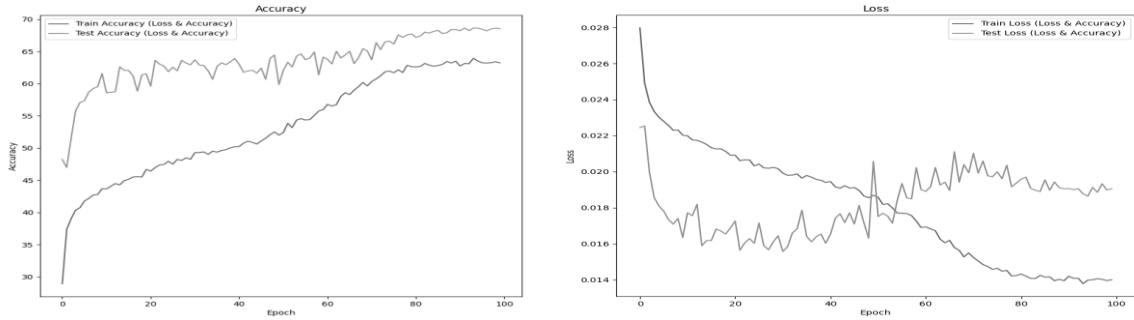
Chúng tôi sử dụng máy tính có cấu hình: Windows 10 64 bit, i5 3200 4 lõi, RAM 8Gb, ngôn ngữ lập trình Python 3.7 trên môi trường Anaconda,.

```
Test set: Average loss: 0.0171, Accuracy: 2467/3589 (68.74%)  
  
EPOCH: 38 LR: 0.00019191894422458902  
Loss=0.8434934616088867 Batch_id=448 le=4.813817529909954e-05 Accuracy=63.07: 100% |██████████| 449/449 [01:12<00:00, 6.16it/s]  
  
Test set: Average loss: 0.0177, Accuracy: 2468/3589 (68.77%)  
  
EPOCH: 39 LR: 4.813817529909954e-05  
Loss=0.8717096447944641 Batch_id=448 le=2.0023904120337228e-07 Accuracy=63.03: 100% |██████████| 449/449 [01:13<00:00, 6.10it/s]  
  
Test set: Average loss: 0.0173, Accuracy: 2472/3589 (68.88%)
```

Hình 7 Kết quả quá trình huấn luyện mô hình

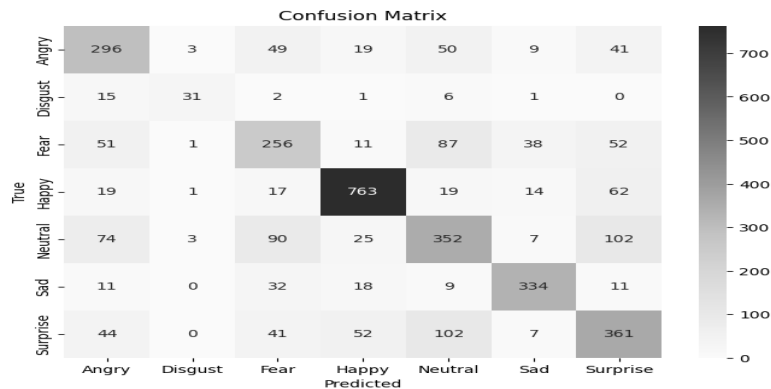
Thực hiện trong môi trường có ánh sáng tốt, không bị ngược sáng, góc quan sát người của camera là chính diện, không đứng quá xa so với camera và khuôn mặt không bị che khuất.

Chúng tôi đã tiến hành các thử nghiệm sau giá trị bỏ học. Nó đã được tìm thấy rằng không có overfitting trong mô hình như thể hiện trong Hình 8 và Hình 9. Thời gian tính toán của quá trình đào tạo mất 87s mỗi epoch với kích thước lô 64. Trong quá trình đào tạo, chúng tôi nhận thấy rằng tổn thất đào tạo giảm trong khi độ chính xác của đào tạo tăng lên khi nhiều epoch được đào tạo.



Hình 8. Đồ thị xác nhận độ chính xác mô hình huấn luyện, Hình 9. Đồ thị xác nhận độ mất mát mô hình huấn luyện

Chúng tôi đã tính toán các số liệu đánh giá như ma trận nhầm lẫn, độ chính xác như thể hiện trong Hình 9.



Hình 9. Ma trận nhầm lẫn của mô hình

Mô hình đã nhận diện thành công bộ sưu tập hình ảnh khuôn mặt đầu vào. Các thí nghiệm với hình ảnh đầu vào đã được thực hiện, hệ thống đã nhận diện thành công cảm xúc khuôn mặt người. Khi so sánh hiệu suất giữa các phương pháp trước đây và phương pháp được chúng tôi đề xuất nhận thấy tính khả quan cao, hiệu suất đạt được khá tốt.

Bảng 2. So sánh các phương pháp trước đây và phương pháp chúng tôi tìm hiểu xây dựng.

Phương pháp	Cơ sở dữ liệu áp dụng	Số trạng thái biểu cảm	Độ chính xác xác thực
Tu`men et al [6]	FER2013	7	57.1%

Knyazev et al [7]	FER2013	7	60.0%
Liu et al [8]	FER2013	7	62.44%
Giannopoulos et al [9]	FER2013	7	64.20%
Haque et al [10]	FER2013	7	63.11%
Agrawal et Mittal [4]	FER2013	7	65,77%
Phương pháp của chúng tôi	FER2013	7	68.88%

Quá trình thực nghiệm, chúng tôi thu lại một số hình ảnh như Hình 10, Hình 11 sau:



Hình 10. Mô tả một số cảm xúc qua thực nghiệm



Hình 11. Nhận diện cảm xúc đông người

4. KẾT LUẬN

Từ những nội dung trên, chúng tôi kết luận rằng nghiên cứu đã đạt được mục tiêu của mình. Tuy nhiên, còn rất nhiều cơ hội để cải thiện và mở rộng nghiên cứu này. Đặc biệt là việc nghiên cứu các biến thể khác của mô hình học sâu hoặc áp dụng nó cho các lĩnh vực khác nhau cũng như việc cung cấp tập dữ liệu lớn và đa dạng hơn, chẳng

hạn như tập dữ liệu khuôn mặt của người Châu Á...để đảm bảo tính khả thi, phổ quát và độ chính xác của mô hình.

Đối với thuật toán Haar-like để phát hiện khuôn mặt, chúng tôi nhận thấy đáp ứng rất tốt với việc nhận dạng khuôn mặt trong tư thế trực diện, cũng như đáp ứng về tính tức thời của khung hình qua camera và webcam máy tính vì tính đơn giản gọn nhẹ của thuật toán. Tuy nhiên vẫn còn một số hạn chế khi triển khai hệ thống trên thực tế như tính ngược sáng, góc nghiêng, phần che khuất của khuôn mặt cũng như khó kiểm chứng qua hệ thống camera ngoài thực địa nên mô hình có thể còn tối ưu được nhằm đạt hiệu quả cao nhất.

Trong tương lai, có thể có nhiều cơ hội để mở rộng và phát triển nghiên cứu về chủ đề nhận dạng cảm xúc. Chúng tôi hy vọng rằng những kết quả và hiểu biết từ luận văn này có thể đóng góp một phần nhỏ vào lĩnh vực thị giác máy tính, lý thuyết nhận dạng và giúp làm sáng tỏ một phần nào đó về học máy và trí tuệ nhân tạo. Chúng tôi cũng mong rằng nghiên cứu này có thể tiếp tục đóng góp vào sự hiểu biết và phát triển trong lĩnh vực ứng dụng mạng nơron nhân tạo và giúp giải quyết các thách thức và vấn đề quan trọng liên quan đến chủ đề nhận dạng cảm xúc này.

TÀI LIỆU THAM KHẢO

- [1]. Akriti Jaiswal, A. Krishnama Raju, Suman Deb,(2020). "Facial Emotion Detection Using Deep Learning". *International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.*
- [2]. Shi, J., Zheng, X., Li, Y., Zhang, Q., & Ying, S. (2018). "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease." *IEEE Journal of Biomedical and Health Informatics*, 22(1), 173–183. doi:10.1109/jbhi.2017.2655720
- [3]. Baoguang Shi, Bai X, Yao C (2016). "Script identification in the wild via discriminative convolutional neural network". *Pattern Recognition* 52:448–458
- [4]. A. Agrawal et N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy", *Vis. Comput.*, janv. 2019, doi: 10.1007/s00371-019-01630-9.
- [5]. Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features" 2001
- [6]. Tu'men, V., O.F. So'ylemez, and B. Ergen. Facial emotion recognition on a" dataset using convolutional neural network. in 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). 2017. *IEEE*.
- [7]. Knyazev, B., et al., Convolutional neural networks pretrained on large face recognition datasets for emotion classification from video. arXiv 2017.

- [8]. Liu, K., M. Zhang, and Z. Pan. Facial expression recognition with CNN ensemble. In 2016 international conference on cyberworlds (CW). 2016. IEEE.
- [9]. Giannopoulos, P., I. Perikos, and I. Hatzilygeroudis, Deep learning approaches for facial emotion recognition: A case study on FER-2013, in Advances in hybridization of intelligent methods. 2018, Springer. p. 1-16.
- [10]. Haque, M.I.U. and D. Valles. A facial expression recognition approach using DCNN for autistic children to identify emotions. in 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). 2018. IEEE.

HUMAN EMOTION DETECTION USING HAAR FEATURES AND CONVOLUTION NEURAL NETWORKS

Nguyen Dang Binh¹, Nguyen Huu Hai^{2*}

¹ Faculty of Information Technology, University of Sciences, Hue University

² Gio Linh district, Quang Tri province

*Email: haihuu@gmail.com

ABSTRACT

Automatic Facial Expression Recognition (FER) is a part of human-machine interaction systems, which is widely researched both domestically and internationally. Current FER methods can be classified into two main groups: pure machine learning-based approaches without neural network application, and deep learning-based approaches employing convolutional neural networks (CNNs). In this study, we focus on investigating face detection using the Viola-Jones method, while implementing a convolutional neural network system to classify facial expressions, as well as combining traditional machine learning methods with modern CNNs. We conduct a comparison of results with other studies conducted over the past 5 years using the widely-used FER2013 dataset.

Keywords: Haar, CNN network, emotion recognition, convolutional neural network,



Nguyễn Hữu Hải sinh ngày 18/01/1981 tại Quảng Trị. Ông tốt nghiệp cử nhân ngành Sư phạm Tin năm 2011 tại ĐHSP- ĐH Huế và thạc sĩ chuyên ngành Khoa học máy tính tại Đại học Khoa học – ĐH Huế vào năm 2023. Từ năm 2005 đến nay, ông là giáo viên công tác tại Trường TH&THCS Gio An, Gio Linh, Quảng Trị.

Lĩnh vực nghiên cứu: Mạng nơron nhân tạo.



Nguyễn Đăng Bình sinh ngày 08/11/1974 tại Thừa Thiên Huế. Năm 1996, ông tốt nghiệp Đại học ngành Toán - Tin tại trường Đại học Sư phạm Huế. Ông nhận bằng thạc sĩ Công nghệ thông tin tại Trường Đại học Bách Khoa Hà Nội năm 2002; bằng Tiến sĩ ngành Công nghệ thông tin tại Viện Công nghệ Kyushu, Nhật Bản năm 2006 và hoàn thành nghiên cứu Sau tiến sĩ tại Viện Thị giác và Đồ họa máy tính tại Đại học Công nghệ Graz, Cộng hòa Áo, năm 2008. Hiện ông công tác tại khoa Công nghệ Thông tin, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Học máy, Thị giác máy tính, Nhận dạng và Xử lý ảnh.