

NHẬN DIỆN CẢM XÚC KHUÔN MẶT DÙNG MẠNG NƠ – RON TÍCH CHẬP CNN TRÊN PHẦN CỨNG JETSON TX2 IMPLEMENTATION OF FACIAL EMOTION RECOGNITION USING CNN ON JETSON TX2

Phạm Minh Quyền, Phùng Thanh Huy, Đỗ Duy Tân,
Huỳnh Hoàng Hà, Trương Quang Phúc
Trường Đại học Sư phạm Kỹ thuật TP.HCM, Việt Nam

Ngày toà soạn nhận bài 19/8/2020, ngày phản biện đánh giá 3/9/2020, ngày chấp nhận đăng 29/9/2020

TÓM TẮT

Trong bài báo này, tác giả sử dụng mạng nơ-ron tích chập, một trong những kiến trúc phổ biến của học sâu để nhận diện và phân loại cảm xúc khuôn mặt. Một cách tổng quát, các mạng nơ-ron được chọn có cấu trúc phức tạp và có tham số lớn, tác giả tập trung vào việc xây dựng một mạng nơ-ron đơn giản hơn và phù hợp với bộ dữ liệu thông qua phương pháp so sánh và đánh giá. Ngoài ra, tác giả cũng tập trung vào việc thu thập một tập dữ liệu đủ lớn để đạt kết quả cao. Cụ thể, tác giả lựa chọn sử dụng nền tảng phần cứng nhúng Jetson TX2 của NVIDIA để tận dụng khả năng tính toán của GPU nhằm tối ưu thời gian tính toán và huấn luyện dữ liệu. Trong đó, dữ liệu được sử dụng là FER2013 và RAF để huấn luyện và kiểm tra. Phương pháp sử dụng đã đạt được độ chính xác 72% trên tập dữ liệu kiểm tra.

Từ khóa: Nhận diện; cảm xúc khuôn mặt; mạng nơ-ron; CNN; Jetson TX2.

ABSTRACT

In this paper, a convolutional neural network (CNN), one of the most popular deep learning architectures used for facial extraction research, has been implemented on NVIDIA Jetson TX2 hardware. Different from many existing approaches investigating CNN with complex structure and large parameters, we have focused on building a robust neural network through extensive performance comparison and evaluation. In addition, we have collected a dataset using a built-in camera on a laptop computer. Specifically, we have applied our model on Jetson TX2 hardware to take advantage of the computational power of the embedded GPU to optimize computation time and data training. In particular, both FER2013 and RAF datasets with seven basic emotions have been used for training and testing purposes. Finally, the evaluation results show that the proposed method achieves an accuracy of up to 72% on the testing dataset.

Keywords: recognition; facial emotion; neural network; CNN; Jetson TX2.

1. GIỚI THIỆU

Trong các hệ thống thông minh hỗ trợ tương tác người dùng, quá trình thu thập dữ liệu và đánh giá hành vi của khách hàng là rất cần thiết, việc này đòi hỏi hệ thống có thể ghi nhận được trạng thái cảm xúc của khách hàng thông qua nhận diện cảm xúc khuôn mặt là hết sức cần thiết. Thực vậy, các phương pháp và hệ thống nhận diện cảm xúc khuôn mặt được sự quan tâm của nhiều

nhà nghiên cứu trong và ngoài nước. Trong [1], nhóm nghiên cứu thuộc trường đại học Nanchang, Trung Quốc đã áp dụng kiến trúc mạng nơ-ron tích chập, mạng VGG, và đạt được độ chính xác 73.06%. Mạng VGG được xem là có hiệu quả cao mặc dù có cấu trúc đơn giản hơn so với nhiều kiến trúc mạng CNN khác nhưng lại có số lượng tham số nhiều hơn và tăng chiều sâu hơn ở mỗi lớp. Mạng VGG có kích thước khá lớn và nhiều tham số (136 triệu) nên sẽ tốn rất

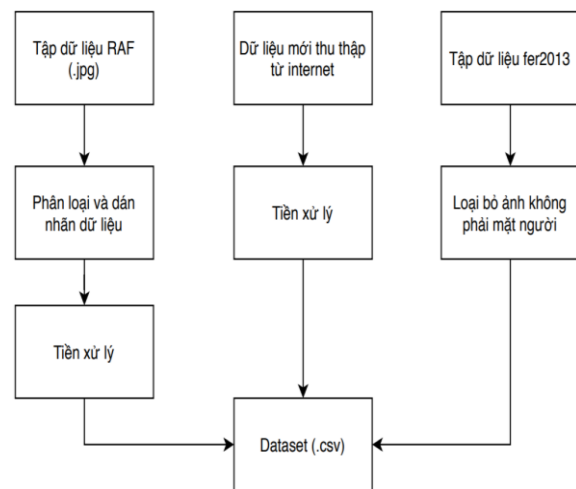
nhiều thời gian huấn luyện dữ liệu. Bên cạnh đó, trong [2], các tác giả đã xây dựng ứng dụng nhận diện cảm xúc khuôn mặt của sinh viên sử dụng CNN với độ chính xác đạt được là 70%. Ở trường đại học Hà Nội, Việt Nam, một nhóm nghiên cứu đã triển khai một dạng kiến trúc khác của CNN đó là BKStart với khoảng 7.17 triệu tham số nhưng vẫn đạt được độ chính xác khá cao (70,4%) trên FER2013 dataset [3]. BKStart khác với VGG bằng việc sử dụng các kernel có kích thước lớn hơn tuy nhiên lại sử dụng stride bằng 1 để tăng hiệu quả trích xuất đặc trưng. Tác giả cũng đã áp dụng mạng Nơ-ron tích chập CNN, một trong những kiến trúc phổ biến trong học sâu, để nhận diện cảm xúc con người.

Bên cạnh sự phát triển của các thuật toán, một yếu tố quan trọng khác đó chính yếu tố phần cứng. Một nhóm nghiên cứu ở Luân đôn, Vương Quốc Anh cũng đã tiến hành thực hiện thiết kế kết hợp với phần cứng Xilinx Spartan-6 LX45 FPGA cho hệ thống nhận diện cảm xúc khuôn mặt [4]. Với phần cứng FPGA, tốc độ xử lý đã được cải thiện đáng kể, có thể đạt được được 30FPS. Tuy có thể cải thiện được khả năng tính toán nhưng lại không thể áp dụng được các thuật toán phức tạp nên dẫn đến độ chính xác còn khá thấp 51.28%. Phần cứng Jetson TX2 được lựa chọn để thực hiện mô hình nhận diện cảm xúc này. Jetson TX2 đã được tích hợp vào đó các lõi CUDA cùng với hệ thống GPU mang lại tốc độ xử lý cao [5]. Ngoài ra Jetson TX2 đáp ứng được hầu hết các yêu cầu khắt khe về thời gian thực nhưng lại mang một kích thước nhỏ gọn, dễ dàng lắp đặt [6].

Tác giả thực hiện mô hình nhận diện cảm xúc khuôn mặt người sử dụng mạng Nơ-ron tích chập CNN trên phần cứng nhúng Jetson TX2 với các mục tiêu, phát triển mô hình nhận có khả năng nhận diện 7 loại cảm xúc: vui, buồn, kinh tởm, ngạc nhiên, bình thường, sợ hãi và giận dữ của con người. Mô hình được xây dựng trên một hệ thống nhúng riêng biệt (Jetson TX2) và có thể đạt được độ chính xác cao (trên 70% cho cả quá trình phát hiện khuôn mặt người và nhận diện cảm xúc).

2. TẬP DỮ LIỆU

Cảm xúc của con người là muôn hình vạn trạng. Mỗi người đều có mỗi cách bày tỏ cảm xúc của mình khác nhau. Chính vì thế, dữ liệu cần để huấn luyện cho mô hình nhận diện cảm xúc khuôn mặt người là rất lớn. Trong đề tài này, tập dữ liệu FER2013 được lựa chọn sử dụng, một phần tập dữ liệu RAF và các dữ liệu được thu thập từ internet. Hình 1 là tổng quan toàn bộ tập dữ liệu và cách xử lý từng loại dữ liệu để đưa bài toán nhận diện cảm xúc:



Hình 1. Tập dữ liệu huấn luyện

2.1 Tập dữ liệu FER2013

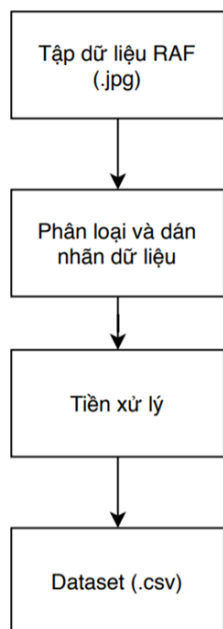
Tập dữ liệu nguồn mở FER2013.csv, được tạo ra cho một dự án bởi PierreLuc Carrier và Aaron Courville, được chia sẻ công khai trong cuộc thi Kaggle (2013).

Dữ liệu trong tập FER2013.csv Bộ dữ liệu này 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). 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’ [4].

2.2 Tập dữ liệu RAF

RAF-DB (Real-world Affective Faces Database) là một tập dữ liệu lớn gồm 15000

hình ảnh khuôn mặt với nhiều loại biểu cảm khác nhau được thu thập từ internet. Hình ảnh trong tập dữ liệu có sự thay đổi lớn về tuổi tác, dân tộc, tư thế đầu và điều kiện ánh sáng và đa dạng về phong cách (mang kính, đội nón) [5].



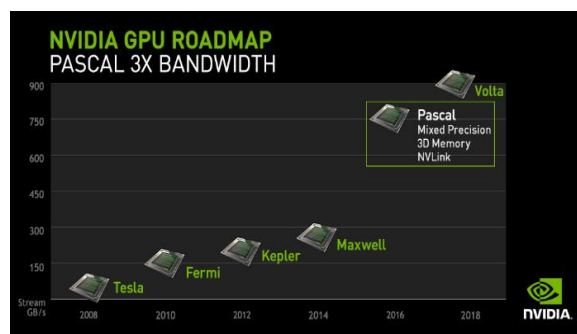
Hình 2. Quá trình xử lý tập dữ liệu RAF

Hình 2 cho thấy quá trình xử lý tập dữ liệu RAF mà nhóm thực hiện để đưa tập dữ liệu này vào tập dữ liệu chung. Ban đầu, tập dữ liệu là tập hợp những hình ảnh cảm xúc khuôn mặt người chưa được phân loại và một tệp .txt chứa dữ liệu phân loại (tên ảnh, nhãn cảm xúc). Từ đó nhóm phân loại ra từng thư mục cảm xúc khác nhau. Tiền xử lý hình ảnh (chuyển ảnh sang ảnh xám và giảm kích thước ảnh về 48x48), chuyển ảnh về dạng pixel và cập nhật vào file Dataset (.csv).

3. PHẦN CỨNG JETSON TX2

Jetson TX2 là một máy tính nhúng có hiệu quả năng lượng cao của NVIDIA được sử dụng rộng rãi cho thị giác máy tính, trí tuệ nhân tạo, máy học. Jetson TX2 là một máy tính nhỏ nhưng khá mạnh mẽ. TX2 có 2 CPU: ARM Cortex-A57 (4 lõi) 2GHz và NVIDIA Denver2 (lõi kép) @ 2GHz [6]. Ngoài ra, ở một máy tính của NVIDIA, GPU là 1 thành phần không thể thiếu với Pascal 256 lõi 1300 MHz. Bộ nhớ 8Gb được chia sẻ giữa CPU và GPU. Tất cả những tác vụ tính toán có thể

tiêu thụ chỉ 7,5W năng lượng. Đây là phần cứng tốt để thực hiện các giải thuật học sâu trên nền tảng hệ thống nhúng [7].



Hình 3. Quá trình phát triển GPU của NVIDIA. [8]

Hình 3 cho thấy quá trình phát triển lõi GPU của NVIDIA đặc biệt là sự đột phá của NVIDIA Pascal ra mắt vào năm 2016 và sau đó đã được đem vào trong kiến trúc của phần cứng Jetson TX2.

4. MẠNG NƠ-RON TÍCH CHẬP

Mạng CNN được xây dựng trên ngôn ngữ Python và sử dụng thư viện Keras. Mạng CNN được xây dựng gồm sáu lớp tích chập, ba lớp Max-pooling và cuối cùng là hai lớp Full-connected. Số lượng filter tương ứng lần lượt là 32, 64 và 128. Kích thước của filter 3x3 và lớp Max-pooling có kích thước stride là 2x2. Lớp Max-pooling được sử dụng sau mỗi hai lớp tích chập. Max pooling được dùng để giảm kích thước ma trận nhưng vẫn làm nổi bật lên được đặc trưng có trong ma trận đầu vào và giảm thiểu công việc tính toán cho phần cứng [9].

Sau mỗi lớp Max Pooling sẽ có một lớp Drop Out nhằm giảm hiện tượng overfitting bằng cách loại bỏ ngẫu nhiên một số unit.

Để giữ được kích thước không gian của ngõ ra, các zero-padding được thêm vào. Sau khi qua các lớp tích chập, ngõ ra sẽ qua Flatten trở thành vector 1 chiều để đưa vào lớp dense (hay lớp full connected).

Hàm kích hoạt được sử dụng trong mạng CNN là Leaky Rectifier Linear Unit (Leaky ReLU) theo công thức sau [10]:

$$f(x) = \max(x, \frac{x}{20}) \quad (1)$$

Giá trị 20 được sử dụng theo tập validation của FER-2013 [11]. Từ công thức (1), giá trị trả về thay vì bằng 0 khi $x < 0$ như RELU, leaky RELU sẽ tạo thành đường dốc với giá trị giảm về nhỏ hơn 0. Ở mạng nơ-ron, kết quả đầu ra cần được chuẩn hóa sang dạng phần trăm để dự đoán giữa các lớp. Khi đó lớp cuối cùng của mạng CNN là lớp Softmax với 7 ngõ ra.

$$a_k = \frac{e^{z_k}}{\sum_{i=1}^n e^{z_i}} \quad (2)$$

Với a_k là tỉ lệ dự đoán lớp thứ k, n là tổng số lớp được dự đoán (7 cảm xúc) và z_k là hệ số của các node trước cho lớp thứ k.

Thuật toán sử dụng để cập nhật trọng số của mạng CNN là Adam. Thuật toán Adam nổi trội hơn một số thuật toán khác bởi khả năng vượt qua vùng local minimum và đạt tới điểm tối ưu nhất (flat minimum). Một đặc điểm ở thuật toán Adam đó chính là khả năng tận dụng “động lực” để hội tụ nhanh hơn. “Động lực” chính là khả năng tận dụng một số phân trong lần cập nhật trọng số trước để cập nhật cho trọng số hiện tại [11].

Kết quả cuối cùng của model cần được đánh giá để biết hiệu quả của quá trình training dữ liệu. Ở đây hàm mất mát chính là công cụ để thực hiện điều đó. Hàm mất mát sử dụng trong bài báo này là hàm “categorical_crossentropy”.

$$L = - \sum_{i=1}^n (y_i * \text{Log}(\hat{y}_i)), \quad (3)$$

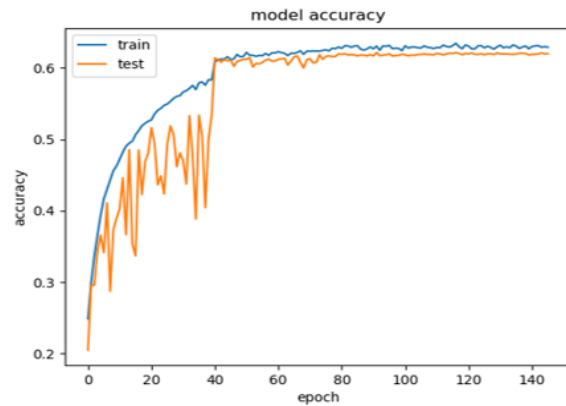
Với L là ký hiệu của hàm mất mát, y_i là kết quả thực và \hat{y}_i là kết quả dự đoán. Biểu thức (3) chính là hàm tính sai số được áp dụng phổ biến và được sử dụng để vẽ đồ thị đánh giá kết quả cuối cùng.

Ngoài hàm mất mát, việc đánh giá model cũng có thể dựa vào độ chính xác:

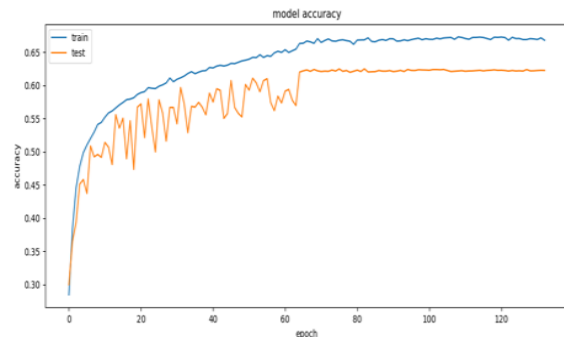
$$\text{Độ chính xác} = \frac{\text{Số lượng dự đoán chính xác}}{\text{Số lượng được dự đoán}} \quad (4)$$

Biểu thức (4) là biểu thức tổng quát để tính độ chính xác, được sử dụng để tính độ chính xác cho quá trình đánh giá kết quả của hệ thống.

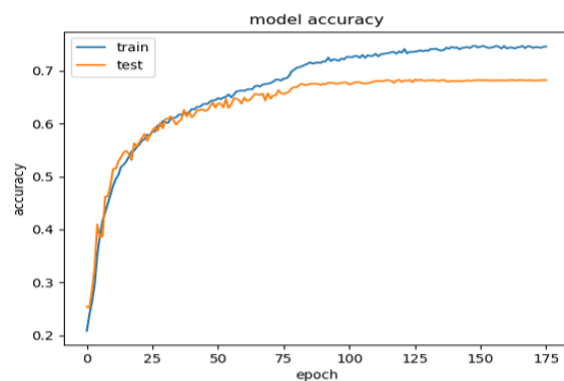
Sgd:



Adagrad:

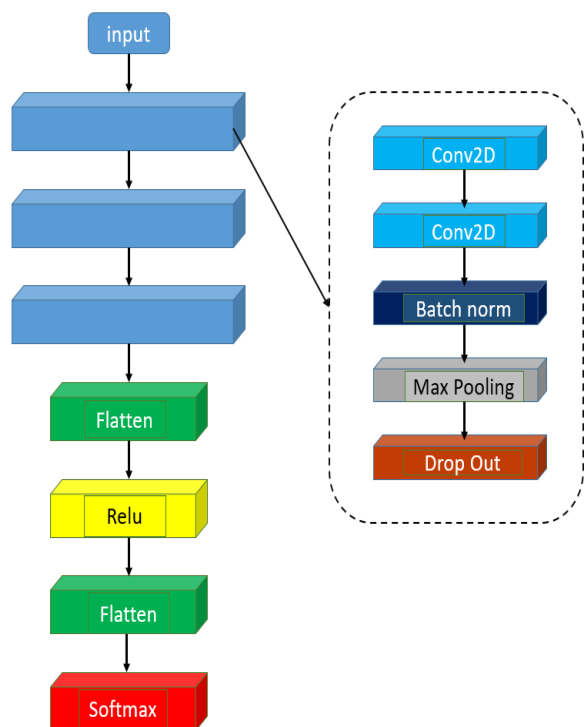


Adam:



Hình 4. So sánh giữa một số thuật toán tối ưu.

Hình 5 cho thấy sự khác nhau trong quá trình huấn luyện dữ liệu giữa các bộ optimizer khác nhau bao gồm Adam, Adagrad và Sgd. Rõ ràng, từ hình 4 cho thấy thuật toán Adam cho kết quả với độ chính xác khá cao so với Sgd và Adagrad. Trong khi đó, quá trình huấn luyện dữ liệu của Sgd và Adagrad lại cho thấy sự giao động khá lớn ở những epoch đầu tiên. Hình 5 trình bày tổng quát toàn bộ kiến trúc mạng CNN.



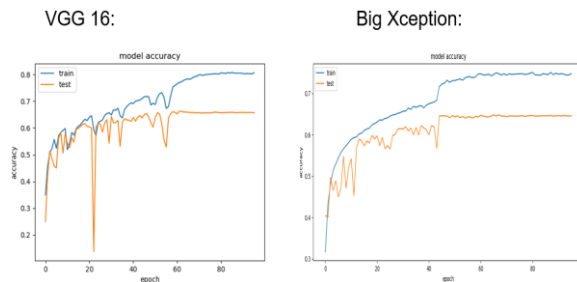
Hình 5. Cấu trúc mạng CNN

Trong bảng 1, chúng tôi so sánh kết quả nhận diện đạt được giữa kiến trúc CNN ConvNet V2 với một số mô hình CNN khác.

Bảng 1. So sánh độ chính xác giữa các kiến trúc CNN

| Model | Accuracy | Param |
|---------------|----------|-------|
| Simple CNN | 0.6291 | 0.64 |
| Simpler CNN | 0.6356 | 0.6 |
| Tiny Xception | 0.6236 | 0.02 |
| Mini Xception | 0.6586 | 0.06 |
| Big Xception | 0.6606 | 0.21 |
| VGG 16 | 0.6656 | 20.86 |
| ConvNet V1 | 0.6803 | 5.01 |
| ConvNet V2 | 0.72 | 9,74 |

Đối với các model có lượng Param thấp như Mini Xception nhưng lại đạt được độ chính xác tốt. Tuy nhiên vẫn chưa đạt được mục tiêu đề ra. Đối với model VGG 16 có lượng tham số cao nhưng đối với dataset có kích thước khá nhỏ khiến model này dễ dẫn đến hiện tượng overfitting. Hình 6 trình bày kết quả nhận diện của 2 kiến trúc mạng CNN phổ biến, VGG16 và Big Xception, với độ chính xác lên tới 66%.



Hình 6. Kết quả nhận diện của 2 kiến trúc mạng CNN phổ biến, VGG16 và Big Xception.

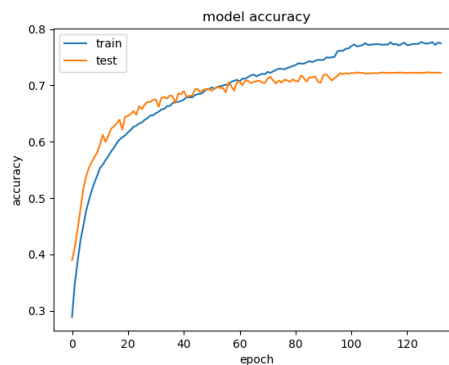
Bảng 2 trình bày kết quả so sánh giữa các thông số ảnh hưởng kết quả huấn luyện dữ liệu. Đối với mỗi thông số khác nhau sẽ phù hợp với một bộ dataset và model khác nhau. Do đó, việc lựa chọn thông số là cần thiết để có thể đạt được hiệu suất tốt nhất trong quá trình huấn luyện dữ liệu.

Bảng 2. Bảng so sánh độ chính xác khi thay đổi các thông số

| Phần cứng | Jetson (GPU) | Jetson (GPU) | Jetson (GPU) | Jetson (GPU) | Jetson (GPU) | Laptop (CPU) |
|-----------------|--------------|--------------|--------------|--------------|--------------|--------------|
| optimizer | Adam | adam | adam | adam | adam | Adam |
| Learning rate | 0,01 | 0.001 | 0.0001 | 0.001 | 0.001 | 0.001 |
| Batch size | 64 | 64 | 64 | 32 | 128 | 64 |
| Thời gian train | 8 giờ | 8 giờ | 12 giờ | 10 giờ | 7 giờ | 16 giờ |
| Acc | 0.25 | 0.69 | 0.69 | 0.68 | 0.68 | 0.68 |
| Loss | 1.8 | 0.91 | 0.91 | 0.93 | 0.92 | 0.92 |

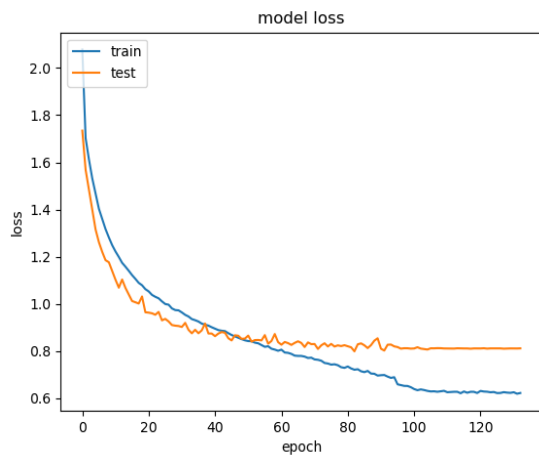
5. KẾT QUẢ ĐÁNH GIÁ

Việc đánh giá model được sử dụng sẽ dựa trên độ chính xác của tập Validation và tập test. Hình 7 cho thấy độ chính xác của quá trình huấn luyện dữ liệu đã hội tụ sau khoảng 85 epoch và đạt được độ chính xác trên tập test là 72%.



Hình 7. Độ chính xác trên tập train và tập test.

Với sự hội tụ của độ chính xác, hình 8 trình bày sai số của model trên tập test validation và tập train cũng đã hội tụ sau khoảng 90 epoch.



Hình 8. Sai số trên tập train và tập test

Confusion matrix là một công cụ hữu ích trong các bài toán phân loại với việc xem xét cả những chỉ số về độ chính xác và độ bao quát của các dự đoán cho từng lớp. Confusion matrix trong bảng 3 cho thấy kết quả của quá trình đánh giá thực tế và kết quả dự đoán của mô hình.

Bảng 3. Confusion matrix

| Actual emotion | Predict emotion | | | | | | |
|----------------|-----------------|---------|------|-------|-----|---------|---------|
| | angry | disgust | fear | happy | sad | suprise | neutral |
| angry | 510 | 35 | 57 | 72 | 157 | 104 | 13 |
| disgust | 9 | 81 | 2 | 3 | 2 | 11 | 2 |
| fear | 60 | 16 | 504 | 74 | 132 | 173 | 109 |
| happy | 4 | 0 | 9 | 1658 | 82 | 19 | 23 |
| sad | 22 | 2 | 53 | 83 | 978 | 98 | 36 |
| suprise | 62 | 6 | 99 | 70 | 308 | 614 | 17 |
| neutral | 5 | 1 | 70 | 46 | 21 | 11 | 655 |

Dựa trên kết quả thu được từ tập test, lớp Angry đạt được độ chính xác xấp xỉ 53%, Disgust đạt 73%, lớp Fear đạt 54%, lớp happy đạt 92%, lớp Sad đạt 77%, lớp Surprise đạt 52% và lớp Neutral đạt 80%. Ba lớp Angry, Fear và Surprise đạt hiệu suất thấp nhất và lớp Happy cùng với lớp Neutral đạt hiệu suất cao nhất.

Để đánh giá model, kết quả thực nghiệm được xem xét trong những điều kiện khác nhau như được liệt kê trong bảng 4.

Bảng 4. Kết quả nhận diện trên nhiều điều kiện khác nhau

| Cảm xúc | Khoảng cách dưới 1m | Khoảng cách trên 1m | Ánh sáng bình thường | Thiếu sáng | Ánh sáng chói. |
|-------------|---------------------|---------------------|----------------------|------------|----------------|
| Bình thường | 57/60 | 60/60 | 57/60 | 33/60 | 19/60 |
| Vui | 53/60 | 31/60 | 53/60 | 34/60 | 14/60 |
| Buồn | 47/60 | 7/60 | 47/60 | 22/60 | 13/60 |
| Giận | 23/60 | 6/60 | 23/60 | 16/60 | 12/60 |
| Khó Chịu | 37/60 | 3/60 | 37/60 | 13/60 | 12/60 |
| Lo sợ | 31/60 | 2/60 | 31/60 | 12/60 | 11/60 |
| Ngạc nhiên | 50/60 | 4/60 | 50/60 | 10/60 | 13/60 |

Dựa vào bảng 4 có thể thấy kết quả nhận diện dễ bị ảnh hưởng bởi các điều kiện khác nhau. Trong đó, khoảng cách càng xa hoặc độ sáng không đủ dễ làm giảm độ chính xác của mô hình nhận diện. Tương tự, trong điều kiện thiếu sáng hoặc quá sáng mô hình cho thấy kết quả nhận diện không tốt. Hơn nữa, đối với khoảng cách xa, kết quả phát hiện khuôn mặt sẽ bị ảnh hưởng đáng kể, phần lớn do độ phân giải của camera khá thấp khiến những bức ảnh xa có số lượng đặc trưng thấp ảnh hưởng tới kết quả nhận diện.

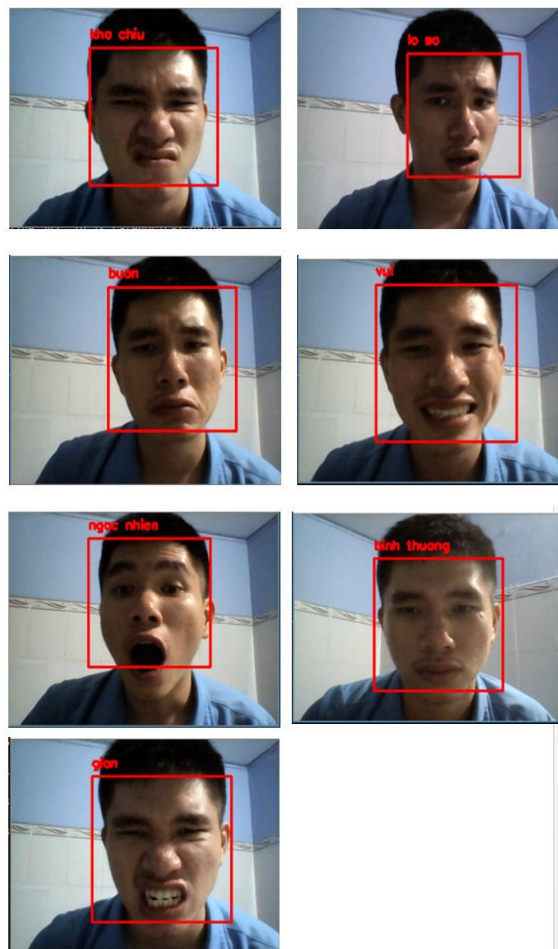
Hình 9 và hình 10 tương ứng cho thấy kết quả nhận diện đúng và kết quả nhận diện sai trên tập dữ liệu test của FER2013 với khoảng hơn 7000 chiếm khoảng 20% tập dữ liệu của FER.



Hình 9. Kết quả nhận diện đúng trên tập dataset



Hình 10. Kết quả nhận diện sai trên tập dataset



Hình 11. Một số hình ảnh nhận diện thực tế.

Hình 11 minh họa một số kết quả nhận diện 7 loại cảm xúc tương ứng bằng cách sử dụng trực tiếp camera được tích hợp sẵn trên phần cứng Jetson TX2.

6. KẾT LUẬN

Trong đề tài này, tác giả đã triển khai một mô hình mạng Nơ-ron tích chập CNN trên nền tảng phần cứng nhúng Jetson TX2 để nhận diện 7 loại cảm xúc cơ bản của con người, với độ chính xác của mô hình đạt được 72%. Mô hình nhận diện được cảm xúc của khuôn mặt trong điều kiện vùng khuôn mặt sáng, khuôn mặt không bị che khuất, hướng khuôn mặt nhìn thẳng và nhìn nghiêng

45°. Tỷ lệ nhận dạng chính xác các cảm xúc vui vẻ, ngạc nhiên và bình thường khá cao, cảm xúc buồn và giận dữ có tỷ lệ dự đoán thấp, riêng với cảm xúc ghê tởm tỷ lệ dự đoán là thấp nhất. Việc sử dụng phần cứng nhúng Jetson TX2 (GPU) cho thấy tốc độ xử lý nhanh hơn đáng kể so với laptop (CPU). Tuy nhiên, số lượng hình ảnh trong tập dữ liệu giữa các lớp có sự chênh lệch quá lớn, lớp khó chịu có ít dữ liệu nhất (1424 ảnh) trong khi đó lớp vui có nhiều dữ liệu nhất (14946 ảnh) trong tổng số 51226 ảnh, điều này dẫn đến sự mất cân bằng dữ liệu giữa các lớp. Từ đó có sự ảnh hưởng rất lớn đến độ chính xác của hệ thống.

TÀI LIỆU THAM KHẢO

- [1] HeJun, Li Shuai, Shen Jinming, Liu Yue, Wang Jingwei, Jin Peng, "Facial Expression Recognition Based on VGGNet Convolutional Neural Network," 2018 Chinese Automation Congress (CAC), Xi'an, China, pp. 4146-4151, 2018.

- [2] Imane Lasri , Anouar Riad Solh , Mourad El Belkacemi, "Facial Emotion Recognition of Students using Convolutional Neural Network," 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, pp. 1-6, 2019.
- [3] D. V. Sang, N. Van Dat and D. P. Thuan, "Facial expression recognition using deep convolutional neural networks," 2017 9th International Conference on Knowledge and Systems Engineering (KSE), Hue, 2017, pp. 130-135.
- [4] 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.
- [5] "Nvidia. Autonomous Machines," NVIDIA, [Online]. Available: <https://developer.nvidia.com/embedded/develop/hardware?fbclid=IwAR2SmS-iYoKCWGrc6fysbbbd7t07Ly8dw-g9IfldJT-ilbP9aT28vSrFs8I>.
- [6] Artiom Basulto-Lantsova , Jose A. Padilla-Medina , Francisco J. Perez-Pinal , Alejandro I. Barranco-Gutierrez, "Performance comparative of OpenCV Template Matching method on Jetson TX2 and Jetson Nano developer kits," 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, pp. 0812-0816, 2020.
- [7] "Technical Specifications," NVIDIA, [Online]. Available: <https://developer.nvidia.com/embedded/develop/hardware>.
- [8] "Nvidia. Autonomous Machines," NVIDIA, [Online]. Available: <https://developer.nvidia.com/embedded/develop/hardware?fbclid=IwAR2SmS-iYoKCWGrc6fysbbbd7t07Ly8dw-g9IfldJT-ilbP9aT28vSrFs8I>.
- [9] J. Flores, "Training a TensorFlow model to recognize emotions," 24 May 2018. [Online]. Available: <https://medium.com/@jsflo.dev/training-a-tensorflow-model-to-recognize-emotions-a20c3bcd6468>.
- [10] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 2018, pp. 1-2.
- [11] Tee Connie , Mundher Al-Shabi , Wooi Ping Cheah, Michael Goh, " Facial Expression Recognition Using a Hybrid CNN–SIFT Aggregator," *Multi-disciplinary Trends in Artificial Intelligence (MIWAI), Lecture Notes in Computer Science*, vol 10607, 2017.

Tác giả chịu trách nhiệm bài viết:

Trương Quang Phúc

Trường Đại học Sư phạm Kỹ thuật Tp. Hồ Chí Minh

Email: phuctq@hcmute.edu.vn