1.1 Thực nghiệm

1.1.1 Dữ liệu

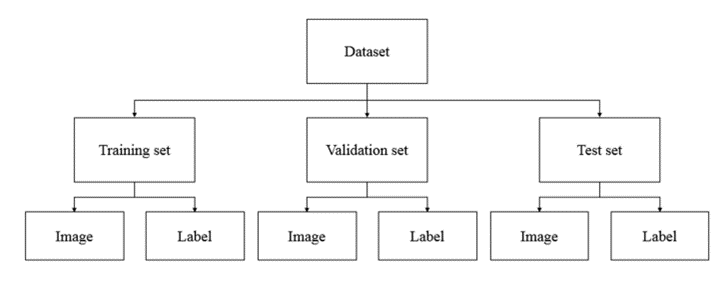
Bài tập lớn này sử dụng bộ dữ liệu fer-2013. Bộ dữ liệu bao gồm 35887 dòng dữ liệu, với mỗi dòng dữ liệu bao gồm 3 cột dữ liệu: một cột chứa nhãn cảm xúc, cột thứ hai chứa hình ảnh lưu dưới dạng pixel tỷ lệ 48 \* 48 và cột cuối cùng là phương thức sử dụng.

Tập dữ liệu này chứa 7 nhãn cảm xúc cơ bản của con người:

* 0: Angry
* 1: Disgust
* 2: Fear
* 3: Happy
* 4: Sad
* 5: Surprise
* 6: Neutral

Bộ dữ liệu FER được chia thành 3 nhóm chính:

* Training dataset là dữ liệu được dùng để huấn luyện cho mạng.
* Validation dataset là nhóm dữ liệu dùng để kiểm tra độ chính xác của hệ thống trong quá trình huấn luyện
* Test set: là nhóm dữ liệu dùng để đánh giá độ chính xác của hệ thống, sau khi đã huấn luyện xong.



Hình **1.**2 Cấu trúc dữ liệu [[1]](#footnote-1)

FER2013 là một tập dữ liệu cảm xúc khuôn mặt lớn sử dụng cho việc huấn

luyện mạng nơ-ron, tuy nhiên việc phân bố dữ liệu không đồng đều giữa các trạng

thái cảm xúc với nhau. Việc này có ảnh hưởng đến kết quả huấn luyện.

Bên cạnh đó, FER2013 cũng cho thấy đây là một tập dữ liệu mạng tính thách thức vì bên cạnh những dữ liệu mang đúng nhãn cảm xúc thì có những ảnh mang nhãn không đúng hoặc không thể nhận diện được khuôn mặt trong ảnh.

Trong bộ dữ liệu thì có ba loại cảm xúc là vui, buồn và tự nhiên là phổ biến còn các loại cảm xúc khác như là sợ hãi, ngạc nhiên và kinh tởm là những loại cảm xúc có ít dữ liệu để training nên độ chính xác sẽ thấp hơn so với ba loại cảm xúc trên.

1.1.2 Xử lý dữ liệu

Đầu vào dữ liệu là một file csv mà trong đó chứa dữ liệu ảnh dưới dạng pixel nên phải xử lý dữ liệu.

Đầu tiên, tạo mảng chứa dữ liệu:

* X\_train, X\_test, X\_check: chứa chuỗi pixel
* y\_test, y\_train, y\_check: chứa nhãn cảm xúc

Tiếp theo, chuyển dữ liệu từ dạng chữ thành dạng số.

Đối với, y\_test, y\_train, y\_check mã hóa thành dạng số nguyên 1D.

Sau đó, chuyển X\_train, X\_test, X\_check về dạng dữ liệu tensor 4D để dùng cho mô hình train.

1.1.3 Công nghệ sử dụng

Ngôn ngữ lập trình sử dụng: Python.

Các thư viện sử dụng để hiện thực bài toán: Keras, TensorFlow.

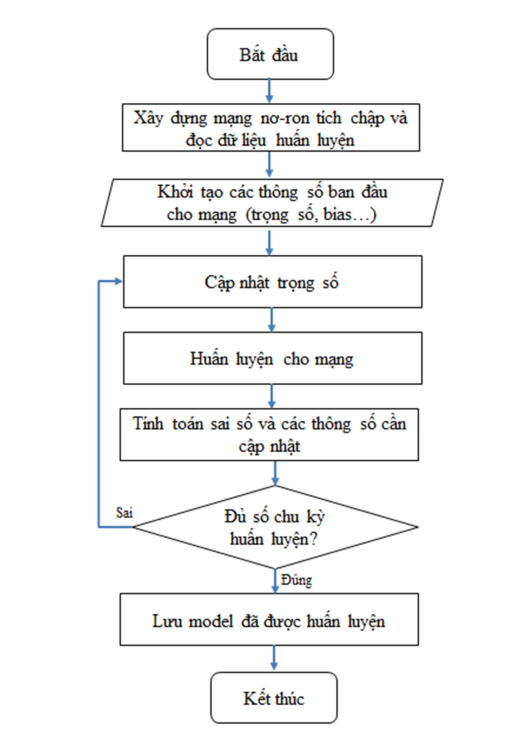
Các công cụ sử dụng giải quyết bài toán: Google Colab, Google Drive.

1.1.4 Cách đánh giá

Trong bài tập lớn này, sử dụng Accuracy và categorical cross entropy làm độ đo của mô hình. Accuracy là một trong những thước đo phổ biến để sử dụng để đánh giá các mô hình.

Accuracy được tính theo công thức: Accuracy = Số lần đoán đúng/ Số lần dự đoán.

1.2 Kết quả đạt được



Hình **1.**3 Lưu đồ giải thuật

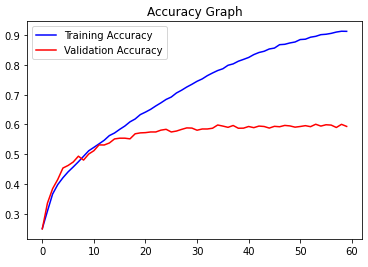
Đề tài này sử dụng mô hình CNN với 3 layer cho lớp tích chập, sau mỗi lớp tích chập sử dụng một lớp MaxPooling, mỗi lớp tích chập sẽ sử dụng hàm kích hoạt ReLU và ở lớp cuối cùng là lớp Fully Connected sử dụng hàm Softmax. Hàm Loss sử dụng hàm categorical cross entropy, thuật toán tối ưu sẽ sử dụng thuật toán Adam.

Với mô hình CNN trên, mô hình được huấn luyện với bộ dữ liệu fer – 2013. Quá trình huấn luyện được trải qua 60 chu kỳ dữ liệu ( 60 epoch), chọn epoch có chỉ số tốt nhất để đánh giá kết quả.

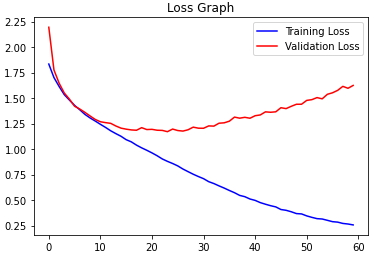
Trong quá trình huấn luyện mô hình, sau mỗi chu kỳ các thông số độ chính xác của mỗi chu kỳ và độ chính xác của tập dữ liệu validation. Sau mỗi chu kỳ thì độ chính xác và độ chính xác của tập dữ liệu cũng dần tăng lên, độ chính xác sau khi kết thúc mô hình vào khoảng 90.99% và độ chính xác của tập validation khoảng 58.76%.

Sau mỗi chu kỳ thì độ chính xác của quá trình huấn luyện tăng dần, còn với độ chính xác của tập dữ liệu validation thay đổi ít đối với những chu kỳ huấn luyện cuối.

Kết thúc huấn luyện mô hình, tiến hành kiểm tra độ chính xác của mô hình thông qua tập dữ liệu private test, độ chính xác đo được trên tập dữ liệu này là khoảng 60.29%.



Hình **1.4** Độ chính xác trong quá trình huấn luyện



Hình **1.5** Loss trong quá trình huấn luyện

|  |  |  |
| --- | --- | --- |
| Features | 7 emotions | 5 emotions |
| HoG features | 29.0% | 34.4% |
| Face landmarks | 39.2% | 46.9% |
| Face landmarks + HOG | 48.2% | 55.0% |
| Face landmarks + HOG on slinding window | 50.5% | 59.4% |
| CNN | 60.29% | N/A |

Bảng **1.**1 Bảng so sánh giữa các mô hình[[2]](#footnote-2)

1.3 Kết luận

1.3.1 Kết quả đạt được

Về lý thuyết, đề tài này đã trình bày được những phương pháp giải quyết bài toán, sử dụng phương pháp Support Vector Machine và phương pháp học sâu Convolutional Neural Network. Trong bài, đã sử dụng phương pháp học sâu Convolutional Neural Network, CNN là một trong những phương pháp phổ biến trong các bài toán về nhận diện cảm xúc.

Về thực nghiệm, trong bài tập lớn này sử dụng CNN để tiến hành triển khai mô hình với bộ dữ liệu fer-2013. Kết quả của mô hình cho thấy mô hình có thể nhận diện được cảm xúc thông qua các ảnh cụ thể.

1.3.2 Hạn chế

Mô hình CNN có thể nhận diện được cảm xúc nhưng mô hình vẫn còn một số loại cảm xúc khó nhận diện nên dẫn đến mô hình nhận diện kết quả sai.

Cảm xúc của con người rất phức tạp, có một số loại cảm xúc gần như là giống nhau nên khả năng nhận diện cảm xúc của mô hình càng khó.

1.3.3 Hướng phát triển

Với đề tài này có một số hạn chế nên có một số hướng phát triển để đạt được hiệu quả cao hơn so với mô hình trên:

* Sử dụng các thuật toán khác nhau để tìm ra thuật toán có độ chính xác cao nhất để giải quyết bài toán.
* Sử dụng kết hợp các thuật toán khác nhau để nâng cao hiệu quả của mô hình nhận diện cảm xúc.
* Nghiên cứu, tìm hiểu, thu thập bộ dữ liệu đặc trưng để huấn luyện mô hình.
* Tiến hành tiền sử lý dữ liệu để nâng cao hiệu quả và độ chính xác của mô hình.

1. Phung, Van Hiep, and Eun Joo Rhee. "A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets." *Applied Sciences* 9.21 (2019): 4500. [↑](#footnote-ref-1)
2. https://github.com/amineHorseman/facial-expression-recognition-svm [↑](#footnote-ref-2)