**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A blue circle with text

Description automatically generated**

**BÁO CÁO ĐỒ ÁN**

**MÔN : ĐỒ ÁN CHUYÊN NGÀNH**

**ĐỀ TÀI : EMOTIONAL CLASSIFICATION (CNN)**

**MÔ HÌNH : VGG**

**Giáo viên hướng dẫn : Thầy Nguyễn Quốc Huy**

**Thành viên nhóm**

Nguyễn Hoài Lâm – 3120410274

Nguyễn Phan Huy Lượng – 3120410312

**TP. HỒ CHÍ MINH, Ngày 24 tháng 03 Năm 2024**

[**ABSTRACT 1**](#_heading=h.8seq58ckciyf)

[**I . INTRODUCTION 1**](#_heading=h.nte8hzpyj6nz)

[**II . RELATED WORK 1**](#_heading=h.45xpw8dtx84f)

[**2.1. Tổng quan 1**](#_heading=h.n64j8kwrg7s8)

[**2.1.1. Giai đoạn ban đầu (1960-1990) 2**](#_heading=h.kkqt26q539g)

[**2.1.2. Giai đoạn dựa trên học máy (1990-nay) 2**](#_heading=h.1upsvf8hmhc4)

[**2.1.3. Giai đoạn hiện đại (2010-nay) 2**](#_heading=h.qv64p33bkq8j)

[**2.2. Dataset 4**](#_heading=h.ncmvg8x8lnjw)

[**2.3. Công nghệ sử dụng 4**](#_heading=h.xj90q5yvna5y)

[**III . EMOTIONAL CLASSIFICATION USING VGG19 5**](#_heading=h.6hdtb91wpdrz)

[**IV . EXPERIMENT 7**](#_heading=h.se75qdxa63tr)

[**4.1. Truyền dữ liệu và in ra màn hình. 7**](#_heading=h.7tm13yoo5188)

[**4.2. Biến đổi dữ liệu 7**](#_heading=h.obdy1tgea652)

[**4.3. Tách dữ liệu 8**](#_heading=h.wkkis5b53idn)

[**4.4. Khởi tạo mô hình VGG19 9**](#_heading=h.77flpzp2mfjx)

[**4.5. EarlyStopping và ReduceLROnPlateau 12**](#_heading=h.fp711rlp3ijv)

[**4.6. Thiết lập thông số cho mô hình 13**](#_heading=h.cghmxprpfs4)

[**4.7. Huấn luyện và đánh giá mô hình 13**](#_heading=h.240ybcgl6ok2)

[**4.8. Kiểm tra mức độ hoạt động của mô hình VGG19 15**](#_heading=h.vh4u4p65vteh)

[**V . CONCLUSION 16**](#_heading=h.g7sxpvyd5s74)

# **ABSTRACT**

Nhận diện biểu cảm khuôn mặt (FER) có ý nghĩa rất lớn và là một nhiệm vụ quan trọng trong thị giác máy tính với các ứng dụng từ tương tác con người - máy tính đến nghiên cứu tâm lý học. Bài báo này trình bày một bản đánh giá về kỹ thuật FER, tập trung vào cách mà mô hình sẽ hoạt động và đưa ra những đánh giá về mô hình phân tích, chỉ ra được những khuyết điểm tồn tại trong mô hình và so sánh nó với các mô hình hiện đại hơn. Trong bài báo cáo này, nhóm chúng em sẽ tập trung vào phương pháp phân tích biểu cảm khuôn mặt dựa trên mô hình VGG.

# **I . INTRODUCTION**

FER (Facial Emotion Recognition) hay Emotional Classification là một lĩnh vực trong trí tuệ nhân tạo và thị giác máy tính, nhằm nhận dạng và phân loại cảm xúc trên khuôn mặt của con người thông qua các phương pháp phân tích hình ảnh và dữ liệu. Mục tiêu của FER là tự động nhận diện và phân loại các biểu hiện cảm xúc như vui vẻ, buồn bã, tức giận, sợ hãi, kinh ngạc, và khác nhau trên khuôn mặt.

FER thường sử dụng các kỹ thuật học máy và thị giác máy tính để xử lý hình ảnh số hoặc video của khuôn mặt. Các phương pháp thông thường bao gồm việc trích xuất đặc trưng từ hình ảnh khuôn mặt, như các điểm mốc (landmarks) trên khuôn mặt, đồng thời áp dụng các mô hình máy học để phân loại cảm xúc dựa trên các đặc trưng này.

Ứng dụng của FER rất đa dạng, từ việc cải thiện giao tiếp giữa con người và máy tính đến ứng dụng trong lĩnh vực giáo dục, y tế, marketing, và giải trí. Ví dụ, FER có thể được sử dụng trong việc phát hiện tâm trạng của người dùng khi sử dụng các ứng dụng trực tuyến, trong việc tăng cường trải nghiệm người dùng, hoặc trong việc phát hiện tình trạng tâm trạng của bệnh nhân để cung cấp phản hồi hoặc hỗ trợ điều trị. Trong bài báo này, nhóm chúng em đã tập trung vào nghiên cứu phân loại biểu cảm khuôn mặt sử dụng mô hình VGG19 trên tập dữ liệu FER2013.

# **II . RELATED WORK**

## **2.1. Tổng quan**

Nhận dạng Biểu cảm Khuôn mặt (FER) là một lĩnh vực nghiên cứu trong trí tuệ nhân tạo (AI) tập trung vào việc phát triển các thuật toán để tự động xác định và phân loại các biểu cảm cảm xúc trên khuôn mặt con người. Lịch sử của FER có thể được chia thành ba giai đoạn chính:

### **2.1.1. Giai đoạn ban đầu (1960-1990)**

Giai đoạn này chứng kiến sự phát triển của các phương pháp FER dựa trên quy tắc, sử dụng các kiến ​​thức thủ công về biểu cảm khuôn mặt để trích xuất các đặc điểm khuôn mặt và phân loại chúng thành các biểu cảm cụ thể. Một số phương pháp nổi tiếng trong giai đoạn này bao gồm:

* Phương pháp dựa trên cơ mặt: Phân tích chuyển động của các cơ mặt để xác định biểu cảm.
* Phương pháp dựa trên hình dạng: Phân tích hình dạng của các bộ phận khuôn mặt, chẳng hạn như lông mày, mắt và miệng, để xác định biểu cảm.
* Phương pháp dựa trên kết cấu: Phân tích kết cấu của da mặt, chẳng hạn như nếp nhăn và nụ cười, để xác định biểu cảm.

### **2.1.2. Giai đoạn dựa trên học máy (1990-nay)**

Sự ra đời của học máy đã mang đến những tiến bộ đáng kể cho FER. Các phương pháp dựa trên học máy có thể tự động học các mối quan hệ phức tạp giữa các đặc điểm khuôn mặt và biểu cảm cảm xúc từ dữ liệu huấn luyện. Một số phương pháp học máy phổ biến được sử dụng trong FER bao gồm:

* Học máy có giám sát: Sử dụng tập dữ liệu được chú thích thủ công để huấn luyện mô hình học máy phân loại các biểu cảm khuôn mặt.
* Học máy không giám sát: Học các mô hình biểu diễn biểu cảm khuôn mặt từ dữ liệu không được chú thích.
* Học sâu: Sử dụng mạng nơ-ron nhân tạo học sâu để trích xuất các đặc điểm khuôn mặt và phân loại chúng thành các biểu cảm.

### **2.1.3. Giai đoạn hiện đại (2010-nay)**

Giai đoạn này chứng kiến sự phát triển nhanh chóng của các phương pháp FER dựa trên học sâu, đạt được độ chính xác cao hơn bao giờ hết. Các phương pháp FER hiện đại có thể xử lý dữ liệu khuôn mặt trong thời gian thực và hoạt động trong nhiều điều kiện ánh sáng và nhiễu khác nhau. Một số ứng dụng của FER bao gồm:

* Giao tiếp giữa người và máy tính: Cho phép máy tính hiểu và phản hồi cảm xúc của con người.
* Bảo mật: Xác định người dùng dựa trên khuôn mặt và biểu cảm của họ.
* Tiếp thị: Phân tích cảm xúc của khách hàng để cung cấp dịch vụ và sản phẩm phù hợp hơn.
* Giáo dục: Đánh giá sự tham gia của học sinh và cung cấp phản hồi kịp thời.

Nhận dạng Biểu cảm Khuôn mặt là một lĩnh vực nghiên cứu đầy hứa hẹn với tiềm năng to lớn để cải thiện cách con người tương tác với máy tính và thế giới xung quanh.

Một số công trình nghiên cứu tiêu biểu:

| **Năm** | **Tên công trình** | **Tác giả** | **Mô tả** | **Thành tựu và Dataset** |
| --- | --- | --- | --- | --- |
| 2012 | **ImageNet Classification with Deep Convolutional Neural Networks** | Krizhevsky et al | Sử dụng kiến trúc LeNet-5 với 5 lớp tích chập và 2 lớp kết nối đầy đủ | Đạt được tỷ lệ lỗi top-5 là 15.3% trên tập dữ liệu ImageNet |
| 2013 | **Very Deep Convolutional Networks for Large-Scale Image Recognition** | Simonyan et al | Sử dụng kiến trúc VGGNet với 16 lớp tích chập và 3 lớp kết nối đầy đủ. | Đạt được tỷ lệ lỗi top-5 là 11.2% trên tập dữ liệu ImageNet |
| 2013 | **Deep Facial Emotion Recognition** | Yu et al | Sử dụng CNN để phân loại cảm xúc từ khuôn mặt | Đạt độ chính xác 84.6% trên tập dữ liệu FER2013 |
| 2014 | **Going Deeper with Convolutions** | Szegedy et al | Sử dụng kiến trúc VGGNet với 16 lớp tích chập và 3 lớp kết nối đầy đủ | đạt được tỷ lệ lỗi top-5 là 7.3% trên tập dữ liệu ImageNet |
| 2015 | **Emotion Recognition in the Wild** | EmoNet | Sử dụng CNN để phân loại cảm xúc từ hình ảnh trong môi trường tự nhiên | Đạt độ chính xác 65.8% trên tập dữ liệu EmoNet |
| 2015 | **Deep Residual Networks for Image Recognition** | He et al | Đề xuất kiến trúc ResNet với các kết nối trực tiếp giữa các lớp giúp học sâu hơn | đạt được tỷ lệ lỗi top-5 là 4.5% trên tập dữ liệu ImageNet |
| 2016 | **Deep Affective Computing** | AVEC | Sử dụng CNN kết hợp RNN để mô hình hóa sự phụ thuộc thời gian giữa các cảm xúc | Đạt độ chính xác 89.0% trên tập dữ liệu AVEC |
| 2018 | **Emotion Recognition with Transformer and Conditional Generative Adversarial Network** | Zhang et al | Sử dụng mô hình Transformer kết hợp CNN để phân loại cảm xúc | Đạt độ chính xác 92.2% trên tập dữ liệu NRC-Canada |
| 2019 | **A Survey on Transfer Learning for Visual Recognition** | Zhuang et al | Tóm tắt các phương pháp Transfer Learning giúp áp dụng kiến thức từ các mô hình được đào tạo trên tập dữ liệu lớn sang các bài toán khác | Cải thiện hiệu suất và khả năng khái quát của mô hình |
| 2020 | **Multi-Task Learning for Visual Recognition: A Survey** | Zhuang et al | Tóm tắt các phương pháp Multi-task Learning giúp mô hình học được nhiều nhiệm vụ gleichzeitig | Cải thiện hiệu suất và khả năng khái quát của mô hình |
| 2021 | **Graph Attention Convolutional Neural Network for Facial Expression Recognition** | Xin Li, et al | một mô hình CNN sử dụng graph attention network để capture mối quan hệ giữa các bộ phận khuôn mặt | Đạt độ chính xác 63.2% trên tập dữ liệu FER2013. |
| 2022 | **Federated Learning for Facial Expression Recognition** | Wei Wang, et al | Sử dụng federated learning để phân loại cảm xúc trong môi trường phân tán. | Đạt độ chính xác 64.5% trên tập dữ liệu FER2013. |

## **2.2. Dataset**

Bộ dữ liệu FER2013 (Facial Expression Recognition 2013) chứa các hình ảnh cùng với các danh mục mô tả cảm xúc của người trong đó. Bộ dữ liệu chứa các hình ảnh mức xám có kích thước 48×48 pixel với 7 cảm xúc khác nhau như Angry (Tức giận), Disgust (Ghê tởm), Fear (Sợ hãi), Happy (Hạnh phúc), Sad (Buồn bã), Surprise (Ngạc nhiên) và Neutral (Trung tính). Bộ dữ liệu bao gồm 28709 ảnh trong tập huấn luyện, 3589 ảnh trong tập kiểm tra công khai và 3589 ảnh trong tập dữ liệu kiểm tra riêng tư.

## **2.3. Công nghệ sử dụng**

Mạng VGG (Visual Geometry Group) là một loạt các mô hình mạng nơ-ron sâu (deep neural network) được phát triển bởi nhóm nghiên cứu Visual Geometry Group tại Đại học Oxford. Có nhiều biến thể của mạng VGG được phát triển dựa trên kiến trúc gốc.

Cụ thể, hai kiến trúc phổ biến nhất của mạng VGG là VGG16 và VGG19, tương ứng với số lượng layers trong mạng. VGG16 bao gồm 16 layers (bao gồm các lớp tích chập và các lớp pooling), trong khi VGG19 bao gồm 19 layers. Cả hai đều có các lớp tích chập theo cấu trúc sâu và nằm giữa các lớp pooling.

Tuy nhiên, người ta có thể tạo ra các biến thể khác của mạng VGG bằng cách thay đổi số lượng và cấu trúc của các layers hoặc thêm các lớp mới để phù hợp với nhiều mục đích khác nhau trong lĩnh vực thị giác máy tính và học sâu. Điều này có thể dẫn đến sự đa dạng về biến thể của mạng VGG.

VGG19 (mạng nơ-ron sâu 19 lớp) là một biến thể của mô hình VGG, tóm lại bao gồm 19 lớp (16 lớp tích chập, 3 lớp kết nối đầy đủ, 5 lớp MaxPool và 1 lớp SoftMax). Khái niệm của mô hình VGG19 (cũng gọi là VGGNet-19) tương tự như VGG16 chỉ khác là mô hình này hỗ trợ 19 lớp. Số "16" và "19" cho biết số lượng lớp trọng số trong mô hình (các lớp convolutional). Điều này có nghĩa là VGG19 có ba lớp convolutional nhiều hơn so với VGG16.

# **III . EMOTIONAL CLASSIFICATION USING VGG19**

Emotional classification dựa trên dữ liệu hình ảnh sử dụng Convolutional Neural Networks (CNNs) là quá trình nhận diện và phân loại cảm xúc từ hình ảnh.

Convolutional Neural Network (CNN) là một kiến trúc học sâu được sử dụng rộng rãi trong phân tích hình ảnh. Một CNN thường bao gồm tích chập, tổng hợp và đầy đủ các lớp liên kết xếp chồng lên nhau. Sự tích chập có thể là được định nghĩa là một thao tác lọc lặp đi lặp lại đối với đầu vào dẫn đến kết quả trong một bản đồ đặc trưng. Để phân loại hình ảnh, chúng ta có thể nói CNN làm nổi bật các phần quan trọng của hình ảnh để phân biệt các lớp học.

Mô hình VGG19 là một mạng nơ-ron nhân tạo(CNN) được phát triển bởi nhóm nghiên cứu Visual Geometry Group (VGG) thuộc Đại học Oxford. Mô hình này được giới thiệu lần đầu tiên trong bài báo "Very Deep Convolutional Networks for Large-Scale Image Recognition" vào năm 2014. VGG19 đã đạt được kết quả ấn tượng trong nhiều cuộc thi nhận dạng hình ảnh, bao gồm ImageNet Large Scale Visual Recognition Challenge (ILSVRC) năm 2014.

VGG19 có cấu trúc gồm 16 lớp, được chia thành 5 khối chính, mỗi khối bao gồm một số lớp tích chập (convolutional layers), lớp gộp cực đại (max pooling layers) và lớp bình thường hóa (normalization layers). Các lớp tích chập trích xuất các đặc điểm từ hình ảnh đầu vào, lớp gộp cực đại giảm kích thước của biểu diễn đặc trưng, và lớp bình thường hóa giúp cải thiện độ ổn định của mô hình trong quá trình huấn luyện.

Để sử dụng VGG19 cho phân loại cảm xúc khuôn mặt, ta có thể thực hiện theo các bước sau:

**1. Chuẩn bị dữ liệu:**

* Thu thập một tập dữ liệu hình ảnh khuôn mặt có chú thích cảm xúc. Tập dữ liệu nên đa dạng về chủng tộc, giới tính, độ tuổi, điều kiện ánh sáng và biểu cảm cảm xúc.
* Chú thích cảm xúc cho mỗi hình ảnh, thường là một trong bảy loại cảm xúc cơ bản: hạnh phúc, buồn bã, tức giận, sợ hãi, ngạc nhiên, khinh miệt và trung lập.
* Chia tập dữ liệu thành ba tập: tập huấn luyện, tập xác nhận và tập kiểm tra. Tập huấn luyện được sử dụng để huấn luyện mô hình, tập xác nhận được sử dụng để theo dõi hiệu suất của mô hình trong quá trình huấn luyện, và tập kiểm tra được sử dụng để đánh giá hiệu suất mô hình sau khi đã được huấn luyện xong.

**2. Tiền xử lý dữ liệu:**

* Xử lý trước hình ảnh để đảm bảo chúng có kích thước và định dạng giống nhau. Điều này có thể bao gồm việc thay đổi kích thước, cắt xén và chuẩn hóa hình ảnh.
* Áp dụng các kỹ thuật tăng cường dữ liệu để tăng kích thước tập dữ liệu và cải thiện khả năng tổng quát của mô hình. Kỹ thuật tăng cường dữ liệu có thể bao gồm việc xoay, lật, thay đổi tỷ lệ và thêm nhiễu vào hình ảnh.

**3. Trích xuất đặc điểm:**

* Sử dụng VGG19 đã được đào tạo trước để trích xuất các đặc điểm hình ảnh từ mỗi hình ảnh trong tập dữ liệu. Mô hình VGG19 được đào tạo trên một tập dữ liệu khổng lồ gồm hình ảnh và chú thích, cho phép nó học các đặc điểm hình ảnh cấp cao có thể được sử dụng cho nhiều tác vụ khác nhau.
* Loại bỏ lớp đầu ra cuối cùng của VGG19, vốn được sử dụng cho nhiệm vụ phân loại 1000 lớp trong ImageNet.
* Thêm một lớp kết nối đầy đủ mới vào đầu ra của VGG19, với số lượng nút bằng số lượng lớp cảm xúc (trong trường hợp này là 7).

**4. Huấn luyện mô hình phân loại:**

* Sử dụng thuật toán học máy, chẳng hạn như SVM hoặc logistic regression, để huấn luyện mô hình phân loại các đặc điểm hình ảnh được trích xuất từ VGG19 thành các lớp cảm xúc.
* Sử dụng hàm mất mát, chẳng hạn như entropy chéo phân loại, để đánh giá độ sai lệch giữa dự đoán của mô hình và nhãn cảm xúc thực tế.
* Sử dụng phương pháp tối ưu hóa, chẳng hạn như Adam hoặc SGD, để cập nhật trọng số của mô hình theo hướng giảm thiểu hàm mất mát.

**5. Đánh giá mô hình:**

* Đánh giá hiệu suất của mô hình phân loại trên tập dữ liệu kiểm tra.
* Sử dụng các chỉ số hiệu suất, chẳng hạn như độ chính xác, độ chính xác trung bình (mean precision) và độ thu hồi trung bình (mean recall) để đánh giá hiệu quả của mô hình.
* Phân tích kết quả để xác định các điểm mạnh và điểm yếu của mô hình.

**\*Một số lưu ý:**

* Hiệu suất của mô hình phân loại cảm xúc phụ thuộc vào nhiều yếu tố, bao gồm chất lượng dữ liệu, cấu trúc mô hình, thuật toán học máy và phương pháp tối ưu hóa.
* Có thể cần điều chỉnh các siêu

# **IV . EXPERIMENT**

Dưới đây là phương pháp dự đoán biểu cảm khuôn mặt sử dụng mô hình VGG19, tập dữ liệu sẽ lấy từ bộ dữ liệu FER2013. Sau quá trình đánh giá cho thấy độ chính xác của mô hình là khoảng 70% - nằm ở mức khá khi so sánh với các mô hình hiện đại sau này.

Sau đây là mô tả chi tiết về cách mà mô hình hoạt động:

## **4.1. Truyền dữ liệu và in ra màn hình.**

Tiến hành import dữ liệu từ bộ dữ liệu FER2013, sau đó in ra màn hình số lượng biểu cảm, lần lượt là 0:'anger', 1:'disgust', 2:'fear', 3:'happiness', 4: 'sadness', 5: 'surprise', 6: 'neutral'

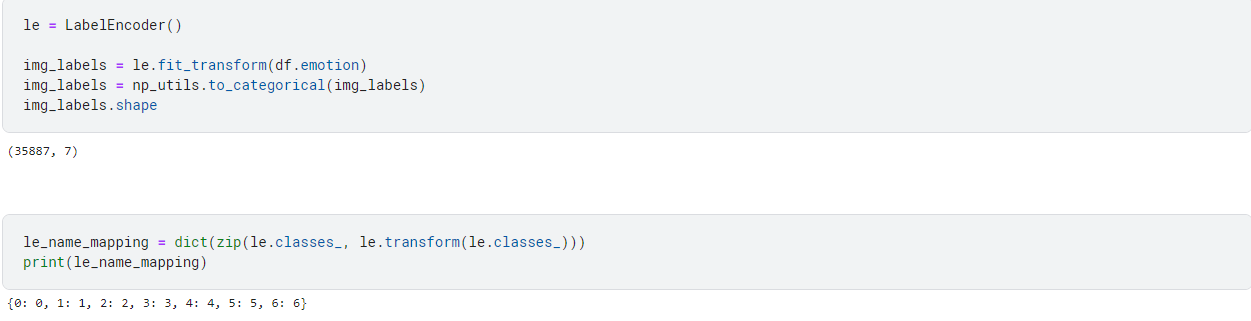


## **4.2. Biến đổi dữ liệu**

Bước này sẽ biến đổi dữ liệu từ nhiều kiểu dữ liệu thành một kiểu thống nhất là Float32 với kích thước là 48x48.

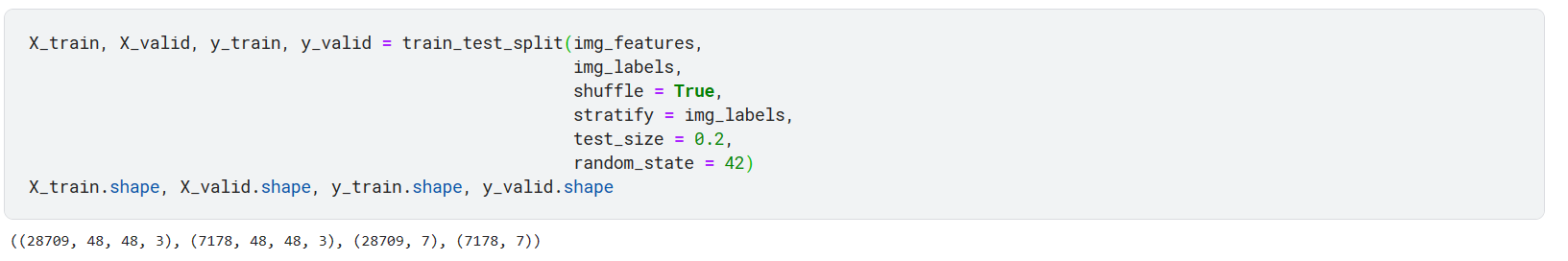


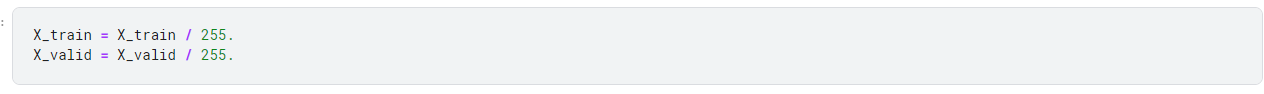
Tùy chỉnh để phù hợp hơn với mạng nơron.

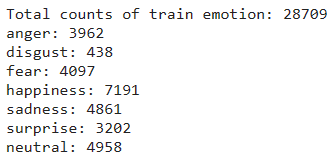


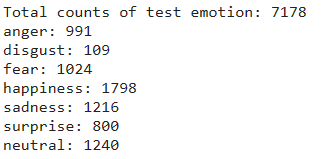
## **4.3. Tách dữ liệu**

Bước này tiến hành phân tách dữ liệu thành hai phần train và test. Dữ liệu từ tập train sẽ dùng để huấn luyện mô hình còn dữ liệu từ tập test sẽ dùng để đánh giá mức độ hiệu quả của mô hình.









## **4.4. Khởi tạo mô hình VGG19**

Mô hình VGG19 và một mô hình có sẵn trong thư viện Tensorflow Keras. Trọng số của mô hình sẽ được khởi tạo từ trọng số đã được đào tạo trước trên tập dữ liệu ImageNet, loại bỏ các layers fully connected.



Model: "vgg19"

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

Layer (type) Output Shape Param #

=============================================================

input\_2 (InputLayer) [(None, 48, 48, 3)] 0

block1\_conv1 (Conv2D) (None, 48, 48, 64) 1792

block1\_conv2 (Conv2D) (None, 48, 48, 64) 36928

block1\_pool (MaxPooling2D) (None, 24, 24, 64) 0

block2\_conv1 (Conv2D) (None, 24, 24, 128) 73856

block2\_conv2 (Conv2D) (None, 24, 24, 128) 147584

block2\_pool (MaxPooling2D) (None, 12, 12, 128) 0

block3\_conv1 (Conv2D) (None, 12, 12, 256) 295168

block3\_conv2 (Conv2D) (None, 12, 12, 256) 590080

block3\_conv3 (Conv2D) (None, 12, 12, 256) 590080

block3\_conv4 (Conv2D) (None, 12, 12, 256) 590080

block3\_pool (MaxPooling2D) (None, 6, 6, 256) 0

block4\_conv1 (Conv2D) (None, 6, 6, 512) 1180160

block4\_conv2 (Conv2D) (None, 6, 6, 512) 2359808

block4\_conv3 (Conv2D) (None, 6, 6, 512) 2359808

block4\_conv4 (Conv2D) (None, 6, 6, 512) 2359808

block4\_pool (MaxPooling2D) (None, 3, 3, 512) 0

block5\_conv1 (Conv2D) (None, 3, 3, 512) 2359808

block5\_conv2 (Conv2D) (None, 3, 3, 512) 2359808

block5\_conv3 (Conv2D) (None, 3, 3, 512) 2359808

block5\_conv4 (Conv2D) (None, 3, 3, 512) 2359808

block5\_pool (MaxPooling2D) (None, 1, 1, 512) 0

=================================================================

Total params: 20,024,384

Trainable params: 20,024,384

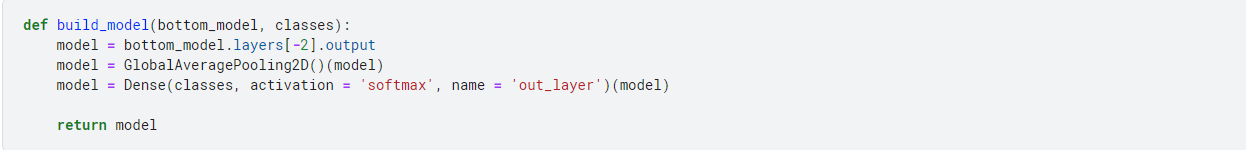
Non-trainable params: 0

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

Khởi tạo hàm **build\_mode**

Hàm build\_model trả về một mô hình mới với kiến trúc đã được xây dựng từ bottom\_model, trong đó các lớp cuối cùng đã được thay đổi để phù hợp với số lượng lớp đầu ra mong muốn.

Trích xuất output của lớp thứ hai từ cuối cùng của mô hình bottom\_model. Điều này thường làm trong các tình huống sử dụng chuyển giao học (transfer learning), muốn loại bỏ các lớp fully connected cuối cùng và sử dụng đầu ra của lớp trước đó làm đầu vào cho mô hình mới.



Model: "model"

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

Layer (type) Output Shape Param #

=============================================================

input\_1 (InputLayer) [(None, 48, 48, 3)] 0

block1\_conv1 (Conv2D) (None, 48, 48, 64) 1792

block1\_conv2 (Conv2D) (None, 48, 48, 64) 36928

block1\_pool (MaxPooling2D) (None, 24, 24, 64) 0

block2\_conv1 (Conv2D) (None, 24, 24, 128) 73856

block2\_conv2 (Conv2D) (None, 24, 24, 128) 147584

block2\_pool (MaxPooling2D) (None, 12, 12, 128) 0

block3\_conv1 (Conv2D) (None, 12, 12, 256) 295168

block3\_conv2 (Conv2D) (None, 12, 12, 256) 590080

block3\_conv3 (Conv2D) (None, 12, 12, 256) 590080

block3\_conv4 (Conv2D) (None, 12, 12, 256) 590080

block3\_pool (MaxPooling2D) (None, 6, 6, 256) 0

block4\_conv1 (Conv2D) (None, 6, 6, 512) 1180160

block4\_conv2 (Conv2D) (None, 6, 6, 512) 2359808

block4\_conv3 (Conv2D) (None, 6, 6, 512) 2359808

block4\_conv4 (Conv2D) (None, 6, 6, 512) 2359808

block4\_pool (MaxPooling2D) (None, 3, 3, 512) 0

block5\_conv1 (Conv2D) (None, 3, 3, 512) 2359808

block5\_conv2 (Conv2D) (None, 3, 3, 512) 2359808

block5\_conv3 (Conv2D) (None, 3, 3, 512) 2359808

block5\_conv4 (Conv2D) (None, 3, 3, 512) 2359808

global\_average\_pooling2d (G (None, 512) 0

lobalAveragePooling2D)

out\_layer (Dense) (None, 7) 3591

=============================================================

Total params: 20,027,975

Trainable params: 20,027,975

Non-trainable params: 0

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

## **4.5. EarlyStopping và ReduceLROnPlateau**

* EarlyStopping:

Early stopping là một kỹ thuật giúp dừng quá trình huấn luyện sớm nếu không có sự tiến triển đủ lớn trong việc cải thiện hiệu suất của mô hình trên tập validation.

Trong trường hợp này, nếu độ chính xác trên tập validation không cải thiện ít nhất 0.00005 trong 11 epochs liên tiếp, quá trình huấn luyện sẽ dừng lại. restore\_best\_weights=True để khôi phục trọng số của mô hình về trạng thái tốt nhất đã được ghi nhận khi Early Stopping được kích hoạt.

* ReduceLROnPlateau:

ReduceLROnPlateau là một kỹ thuật giúp giảm tốc độ học của mô hình nếu không có sự tiến triển đủ lớn trong việc cải thiện hiệu suất của mô hình trên tập validation.

Trong trường hợp này, nếu độ chính xác trên tập validation không cải thiện trong 7 epochs liên tiếp, tốc độ học sẽ giảm đi một nửa (factor=0.5), nhưng không thấp hơn mức 1e-7 (min\_lr=1e-7).

Bằng cách kết hợp cả hai callback này, đảm bảo rằng quá trình huấn luyện sẽ dừng lại khi không có sự tiến triển đủ lớn và tốc độ học sẽ được điều chỉnh tự động để giúp mô hình học hiệu quả hơn.



## **4.6. Thiết lập thông số cho mô hình**

batch\_size = 32

epochs = 50

Hàm tối hư hóa SGD với thông số learning\_rate = 0.001, momentum=0.9, nesterov=True.

Hàm loss được đặt là 'categorical\_crossentropy' (hàm mất mát phù hợp với bài toán phân loại nhiều lớp), và các metric được đánh giá là 'accuracy' (độ chính xác).



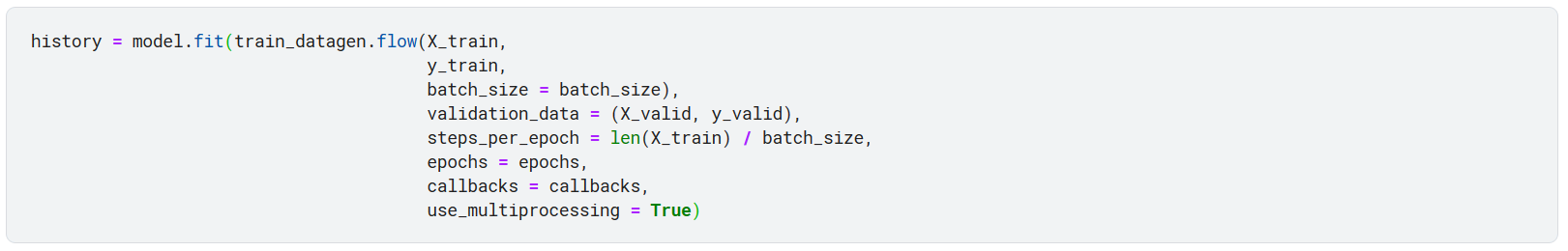
## **4.7. Huấn luyện và đánh giá mô hình**

* Huấn luyện

Sử dụng model.fit để tiến hành huấn luyện mô hình dựa trên những thống số đã thiết lập trước đó.

‘callbacks = callbacks’: Danh sách các callback được sử dụng trong quá trình huấn luyện, chẳng hạn như EarlyStopping, ReduceLROnPlateau.

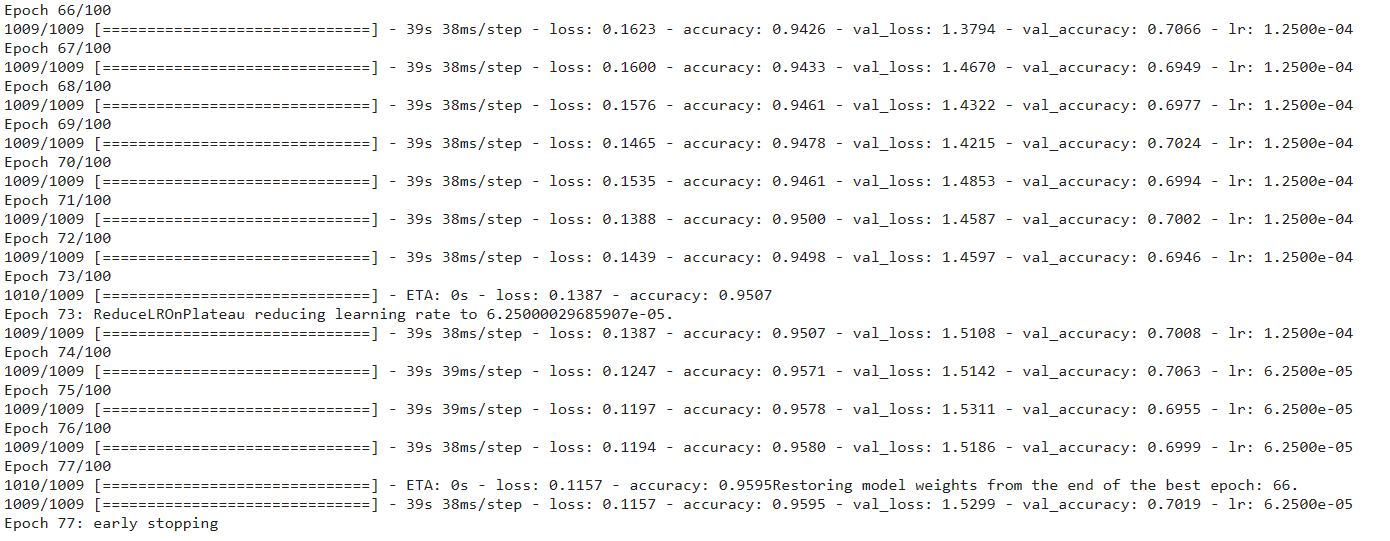
‘use\_multiprocessing = True’: Cài đặt này cho phép sử dụng multiprocessing để tăng tốc độ xử lý của dữ liệu. Điều này có thể hữu ích khi huấn luyện mô hình trên dữ liệu lớn hoặc nếu bạn muốn tận dụng các lõi CPU đa lõi. Tuy nhiên, nó không phải lúc nào cũng phù hợp và có thể không hoạt động tốt trên mọi hệ thống, do đó bạn cần thử nghiệm và điều chỉnh nó tùy thuộc vào tình huống cụ thể của mình.



* Kết quả

Kết quả thực nghiệm sau 100 epoch (vòng lặp):

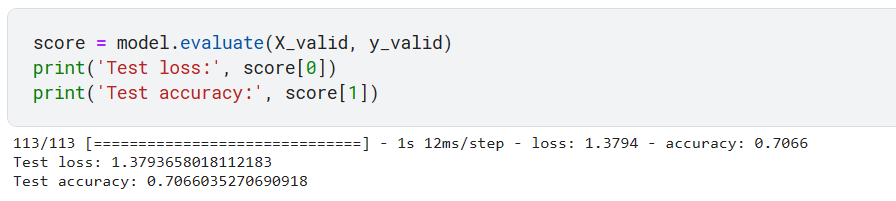
* Train\_loss: 0,1157
* Train\_acc: 0,9595
* Val\_loss: 1,5299
* Val\_Acc: 0,7019



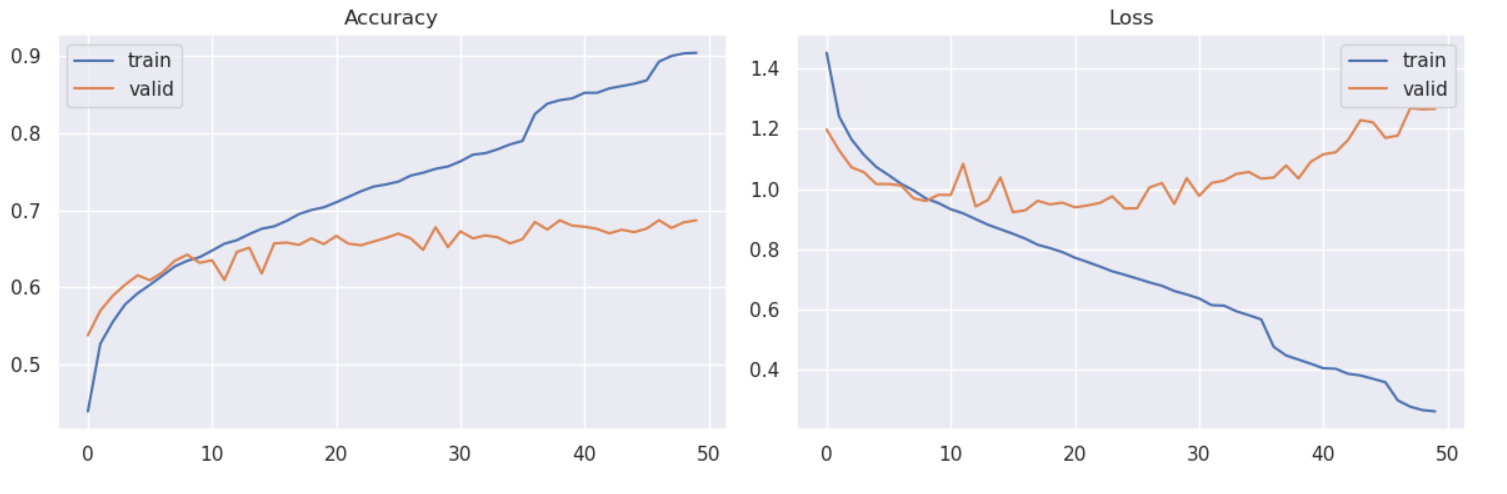
* Đánh giá

Độ chính xác trên tập test: 0,7066 ~ 70,66%.

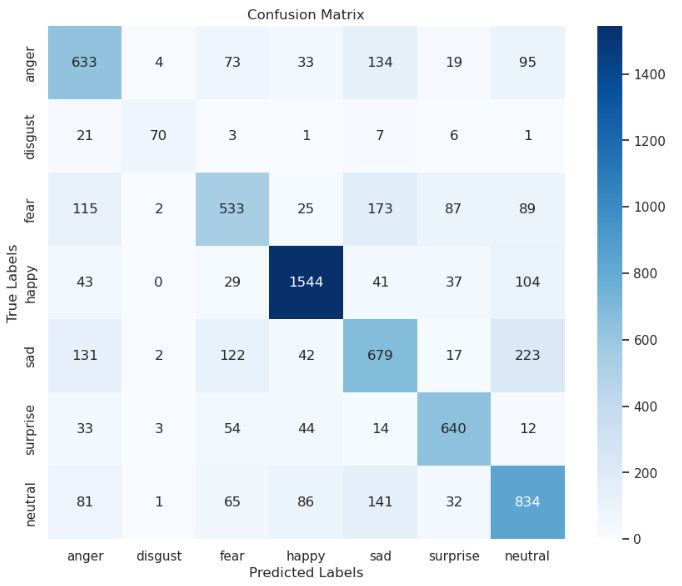
Tỉ lệ mất mát :1,26 ~ 126%.



Mô hình biểu thị mức độ tương quan giữa các thông số loss và accuracy trên hai tập dữ liệu train và test.



Ma trận nhầm lẫn (confusion matrix) biểu thị mức độ chính xác của mô hình giữa trên kết quả dự đoán và kết quả thực tế.



## **4.8. Kiểm tra mức độ hoạt động của mô hình VGG19**

So sánh mô hình trên các tập dữ liệu khác nhau

| **VGG model** | **FER2013** | **RAF DB** |
| --- | --- | --- |
| Valid loss | 1,3793 | 0,8557 |
| Valid Accuracy | 0,7066 | 0,8386 |

So sánh mô hình VGG với những mô hình khác trên cùng tập dữ liệu FER2013

| **Model** | **Year** | **Accuracy** | **Bài báo** |
| --- | --- | --- | --- |
| [Ensemble ResMaskingNet with 6 other CNNs](https://paperswithcode.com/paper/facial-expression-recognition-using-residual) | 2021 | 76,82 | [Facial Expression Recognition using Residual Masking Network](https://paperswithcode.com/paper/facial-expression-recognition-using-residual) |
| [EmoNeXt](https://paperswithcode.com/paper/emonext-an-adapted-convnext-for-facial) | 2023 | 76,12 | [EmoNeXt: an Adapted ConvNeXt for Facial Emotion Recognition](https://paperswithcode.com/paper/emonext-an-adapted-convnext-for-facial) |
| [Segmentation VGG-19](https://paperswithcode.com/paper/a-novel-facial-emotion-recognition-model) | 2023 | 75,92 | [A novel facial emotion recognition model using segmentation VGG-19 architecture](https://paperswithcode.com/paper/a-novel-facial-emotion-recognition-model) |
| [Local Learning Deep+BOW](https://paperswithcode.com/paper/local-learning-with-deep-and-handcrafted) | 2018 | 75,42 | [Local Learning with Deep and Handcrafted Features for Facial Expression Recognition](https://paperswithcode.com/paper/local-learning-with-deep-and-handcrafted) |
| [LHC-Net](https://paperswithcode.com/paper/local-multi-head-channel-self-attention-for) | 2021 | 74,42 | [Local Multi-Head Channel Self-Attention for Facial Expression Recognition](https://paperswithcode.com/paper/local-multi-head-channel-self-attention-for) |
| [Residual Masking Network](https://paperswithcode.com/paper/facial-expression-recognition-using-residual) | 2021 | 74,14 | [Facial Expression Recognition using Residual Masking Network](https://paperswithcode.com/paper/facial-expression-recognition-using-residual) |
| [ResNet18 With Tricks](https://paperswithcode.com/paper/fer2013-recognition-resnet18-with-tricks) | 2021 | 73,70 | [Fer2013 Recognition - ResNet18 With Tricks](https://paperswithcode.com/paper/fer2013-recognition-resnet18-with-tricks) |

# **V . CONCLUSION**

Mô hình VGG19 là một mô hình mạnh mẽ và hiệu quả cho nhận dạng cảm xúc khuôn mặt, đặc biệt là với các tập dữ liệu có kích thước lớn như FER2013. Tuy nhiên, mô hình này cũng có một số hạn chế về tính toán và khả năng phân loại các cảm xúc phức tạp. Do đó, mô hình VGG19 có thể không phù hợp cho các ứng dụng thực tế đòi hỏi thời gian xử lý nhanh hoặc cần phân loại các cảm xúc một cách tinh tế.