

**HO CHI MINH CITY**

**UNIVERSITY OF TECHNOLOGY AND EDUCATION**

**FACULTY OF HIGH QUALITY TRAINING**

**🙡🕮🙣**



**REPORT OF SUBJECT**

**ARTIFICIAL INTELLIGENCE**

|  |  |
| --- | --- |
| **Lecturer: Assoc. Prof. Dr. Nguyen Truong Thinh** | |
| **Name:** | **ID:** |
| **Nguyen Cong Danh** | **20146231** |
| **Class: 222ARIN337629E** | |

**Ho Chi Minh, May 2023**

**TABLE OF CONTENTS**

[CHAPTER 1: INTRODUCTION 3](#_Toc134602898)

[1.1. Machine Vision 3](#_Toc134602899)

[1.2. Artificial Intelligence and Deep Learning 5](#_Toc134602900)

[CHAPTER 2: COLLECT DATA FEATURES 7](#_Toc134602901)

[2.1. Collection 7](#_Toc134602902)

[2.2. Analysis collection data 9](#_Toc134602903)

[CHAPTER 3: TRAINING MODEL 13](#_Toc134602909)

[3.1. Building Pipeline 13](#_Toc134602910)

[3.2. Processing Data 15](#_Toc134602911)

[3.3. Training and Evaluate Model 17](#_Toc134602912)

[CHAPTER 4: APPLYING MODEL FOR MACHINE VISION 22](#_Toc134602916)

[4.1. Object detection 22](#_Toc134602917)

[4.2. Object Counting 2](#_Toc134602918)3

[CHAPTER 5: RESULTS 24](#_Toc134602920)

**CHAPTER 1: INTRODUCTION**

* 1. **Phân tích đám đông với mô hình CNN**

Phân tích đám đông đóng một vai trò quan trọng trong việc hiểu động lực, hành vi và mô hình của đám đông trong các môi trường khác nhau như không gian công cộng, trung tâm giao thông và địa điểm tổ chức sự kiện. Nó liên quan đến việc phân tích các nhóm lớn người để rút ra những hiểu biết có giá trị cho các mục đích lập kế hoạch, an toàn và quản lý đám đông. Một cách tiếp cận hiệu quả để phân tích đám đông là sử dụng Mạng thần kinh chuyển đổi (CNN), một mô hình học sâu mạnh mẽ được biết đến với khả năng học các mẫu và tính năng phức tạp từ hình ảnh.

**1.2 Artificial Intelligence và Deep Learning**

- Trong phân tích đám đông, Artificial Intelligence (AI) và Deep Learning, đặc biệt là mô hình Convolutional Neural Network (CNN), có thể được sử dụng để nhận dạng và phân tích đám đông.

CNN là một loại mạng neural network được thiết kế đặc biệt cho việc xử lý dữ liệu không gian như hình ảnh. Nó sử dụng các lớp convolutional để học các đặc trưng từ dữ liệu và lớp pooling để giảm kích thước của đầu vào. Sau đó, thông qua các lớp fully connected, CNN có khả năng phân loại dữ liệu thành các lớp khác nhau.

Trong bài báo cáo này, tôi chỉ tập trung vào huấn luyện mô hình CNN sau đó áp dụng mô hình vào một ứng dụng thực tế nào đó.Ngoài kia còn có nhiều phương pháp tiên tiến, cũng như các mô hình pretrain được xây dựng rất tốt độ chính xác cao, có thể đưa ra lết quả tốt hơn trong bài báo cáo này.Như đã nói, trong bài báo cáo này bao gồm việc thu tự thập dữ liệu thực tế, nghiêng cứu mô hình, xây dựng một đường data pipeline bằng google colab và cuối cùng là ứng dụng mô hình để giải quyết một vấn đề thực tế nào đó từ đó kết luận được nguyên nhân,lý do, ưu khuyết điểm của mô hình CNN.

**CHAPTER 2: COLLECT DATA FEATURES**

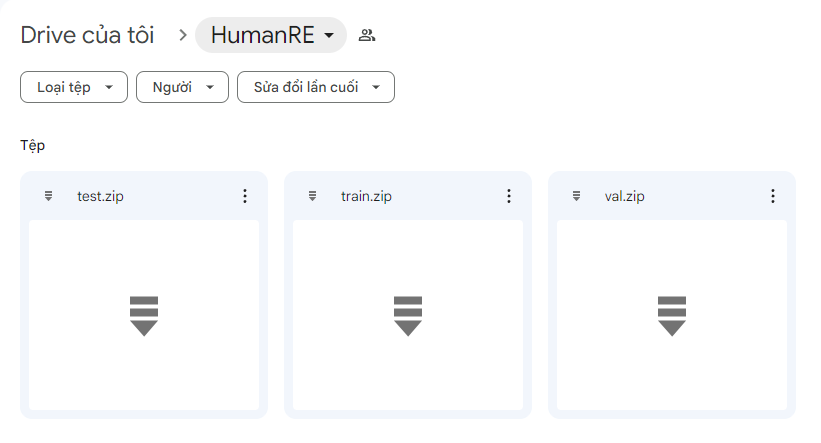
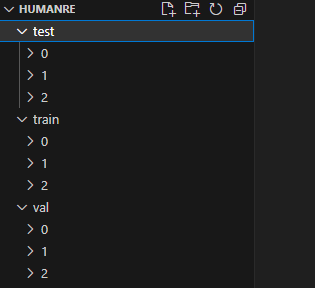
**2.1 Collection**

Công cuộc chuẩn bị dữ liệu là một phần quan trọng. Dữ liệu huấn luyện có vai trò quyết đinh việc nhận diện, quyết định mô hình có hoạt động tốt hay không. Các mô hình AI thông minh lớn trên thế giới đều được huấn luyện từ nguồn data dồi dào.Dữ liệu giúp trích xuất ra những đặc trưng của vật thể vì thế khi mô hình tiếp cận với dữ liệu mới nó sẽ phán đoán chính xác vật thể , yêu cầu của bài toán đó là gì.

Trong bài toán nhận diện vật thể này, dữ liệu thu thập bao gồm ba tệp dữ liệu được nén lại tệp thứ nhất là tệp “train.zip” là dữ liệu thô với các kích thước khác nhau dung để huấn luyện mô hình, bên trong đó thì gồm có hai thư mục thư mục thứ nhất chứa dữ liệu ảnh người có tên “1”, “2” và “0” thư mục còn lại là ảnh không có người là môi trường xung quanh, nó có thể là cái bàn, ghế và những vật thể có màu tối trùng với khu vực ảnh người khi chuyển sang dạng ảnh nhị phân và được bỏ vào folder “0” để loại bỏ vì nó không phải là những đặc trưng của cơ thể con người .Hình ảnh con người khi chuyển về dạng ảnh nhị phân, theo khảo sát thì tôi nhận thấy rằng để nhận diện chính xác con người ta nên tập trung vào phần đầu của họ, tóc của mỗi người sẽ có màu đen thế nên tôi sẽ chọn phần đầu của mỗi người làm dữ liệu cho tệp “train” và đương nhiên trong thực tế không chỉ có tóc mỗi người sẽ mặc trang mục màu tối hoặc sáng, về phần đó tôi sẽ cho nó vào thư mục train thứ hai có tên “2”. Lúc này, data trong tệp huấn luyện đã bao quát hết khu vực thư nghiệm bao gồm đầy đủ các điều kiện thời tiết và chiếu sáng.

Tệp thứ hai lại là tệp “validation.zip” tệp này cũng sẽ được phân loại và gán nhãn như tệp “train.zip” nhưng dữ liệu của nó không thuộc dữ liệu train nghĩa là vẫn sẽ tồn tại 2 thư mục “0” và “1”, ”2” nhưng những dữ liệu này là những dữ liệu mới.Validation set giúp kiểm tra độ chính xác trong mô hình huấn luyện.

Cuối cùng là tệp ”test.zip” tệp này bao gồm các ảnh có ảnh người tron cả 2 trường hợp và ảnh nên để lẫn lộn. tệp này sẽ được sử dụng sau khi mô hình đã được huấn luyện xong. Để biết được mô hình sau khi huấn luyên có được tốt hay không ta sử dụng dữ liệu mới để tiếp cận mô hình training.Trong tệp dữ liệu này bắt buộc phải được chuẩn hóa giống như dữ liệu của mô hình ta sẽ nói ở phần sau.Nó cũng được phân loại thành 3 folder cho các trường hợp cần thiết.

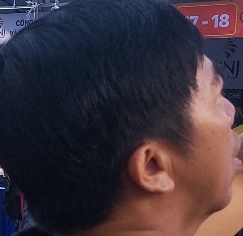
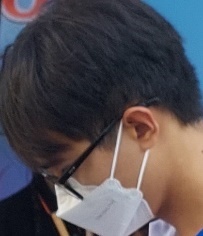
 

**2.2 Analyze collected data**

Dữ liệu huấn luyện là quan trọng,do đó ta hãy phân tích một chút vì khi huấn luyện và sử dụng thì dữ liệu trong tệp test và dữ liệu trong tệp train nó phải có dạng khá giống nhau.  
Như đã nói từ trước ảnh sẽ được phân loại là trường hợp người sẽ chia làm 2 phần như vậy bởi lẽ khi chuyển sang anh dạng nhị phân ta chỉ có thể thấy phần đầu , phân thân hoặc phần chân (tuy vào phần nào mặc trang phục tối) thể nên khi sử dụng thuật toán CNN để nhận diện thì ta phải chú ý đến việc xử lí ảnh cho đúng.  
Sau đây là ảnh từ tệp 1 và 2:

**Trường hợp nhận dạng đúng:**

Tệp 1:



Tệp 2:



Tệp 0:



**Trường hợp nhận dạng sai:**

Trường hợp một sẽ nhận dạng sai khi ảnh crop không gần sát với khuôn mặt mà bị kéo xuống phần vai hoặc chỉ crop phần tóc đen ở đỉnh đầu mà không crop đến hết toàn bộ khuôn mặt.

Trường hợp hai sẽ nhận dạng sai khi khu vực ảnh crop cơ thể có chứa quá nhiều đặc tính của môi trường hoặc là một bức ảnh tổng thể (chưa được crop).

Dữ liệu này được thu thập tự camera cá nhân, các nguồn ảnh từ các nhiếp ảnh.

**CHAPTER 3: TRAINING MODEL**

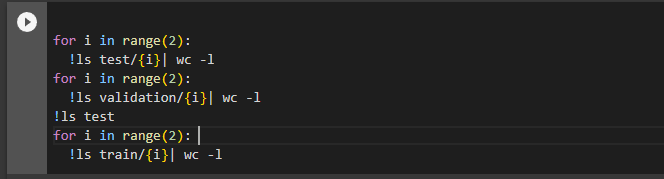


**3.1. Building data Pipeline from GG Colab**

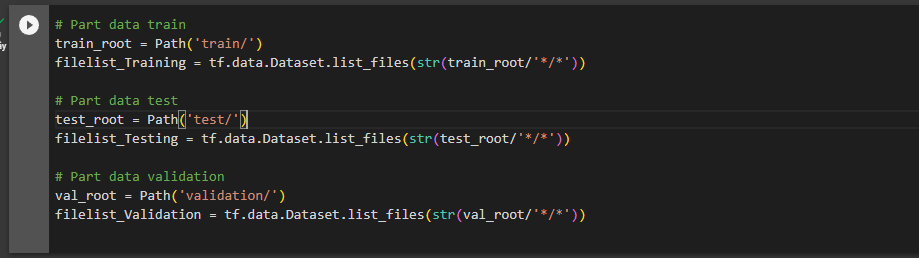
Lý do xây dựng một data pipepline đó chính là chúng ta có thể trực tiếp quản lí dữ liệu trên google drive và có thể trực tiếp lấy dữ liệu xuống để huấn luyện mô hình.Một lý do khác là vì có thể liên kết qua google colab ta có thể dùng nó để training model bằng thời gian chạy GPU, GPU của google sẽ giúp giảm thời gian train đi rất nhiều và tiết kiệm tài nguyên xử lí của máy tính.

Để làm được việc này ta cần phải tải dữ liệu lên google colab và tạo đường dẫn cho cho các thư mục dữ liệu  
Code tham khảo

Kiểm tra dữ liệu:

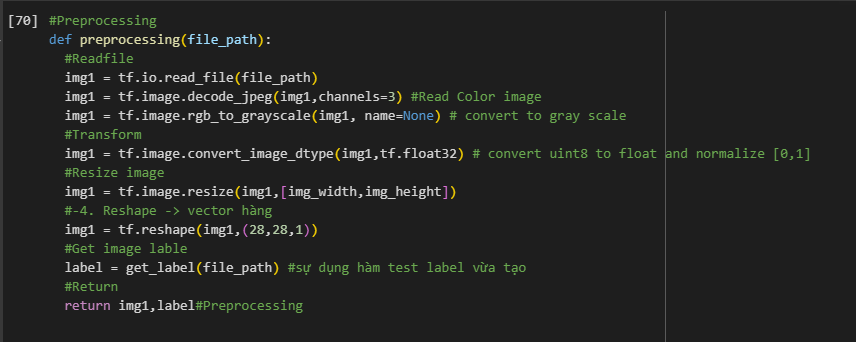


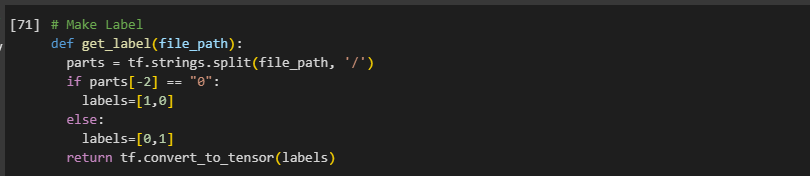
Tạo đường dẫn:



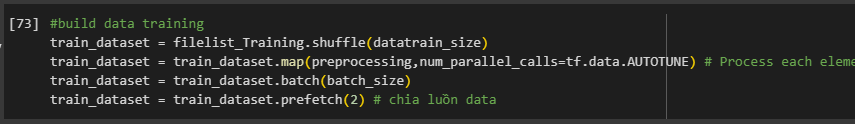
**3.2. Processing Data**

Tiến hành xử lí ảnh đầu vào và đồng thời gán nhãn cho từng lớp.Trong hàm preprocessing, ta sẽ đọc nó dưới dạng ảnh màu (channels=3).sau đó convert sang ảnh xảm .  
Tiếp theo đó là normalize tấm ảnh về giá trị từ 0 đến 1, sau đó tiến hành resize tấm ảnh về kích thước (28,28), cuối cùng ra reshape tấm ảnh thành một vector.

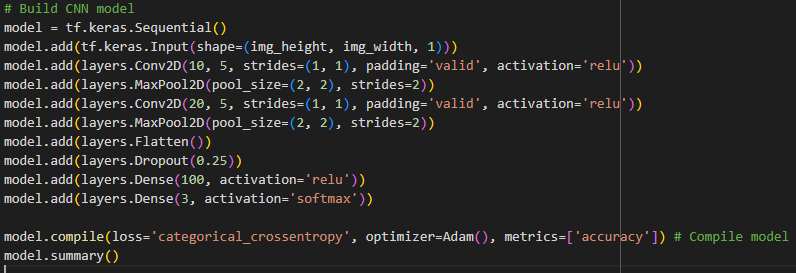


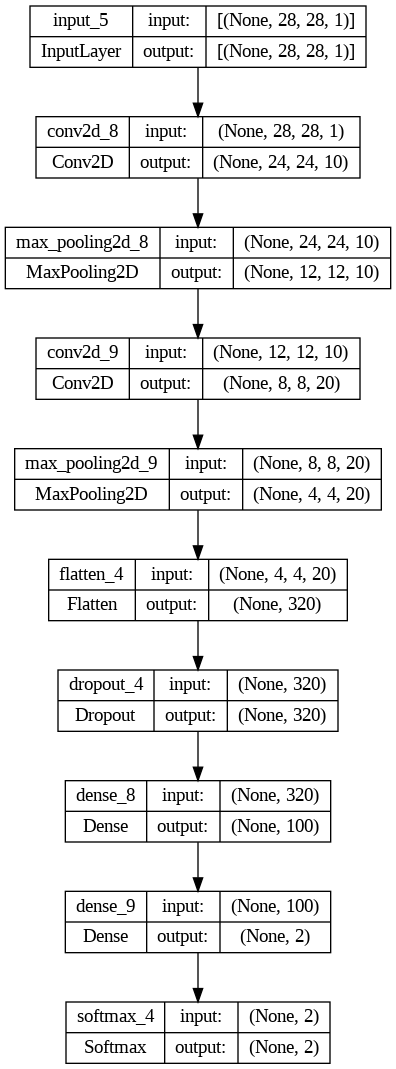


**Bắt đầu xây dụng dữ liệu huấn luyện:**

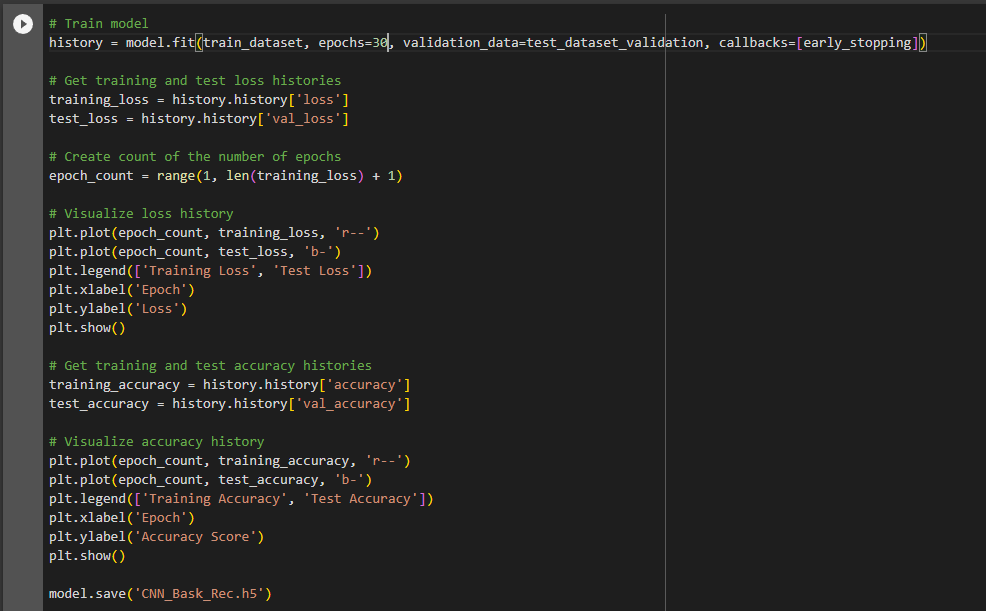


**3.3. Build Model**



****

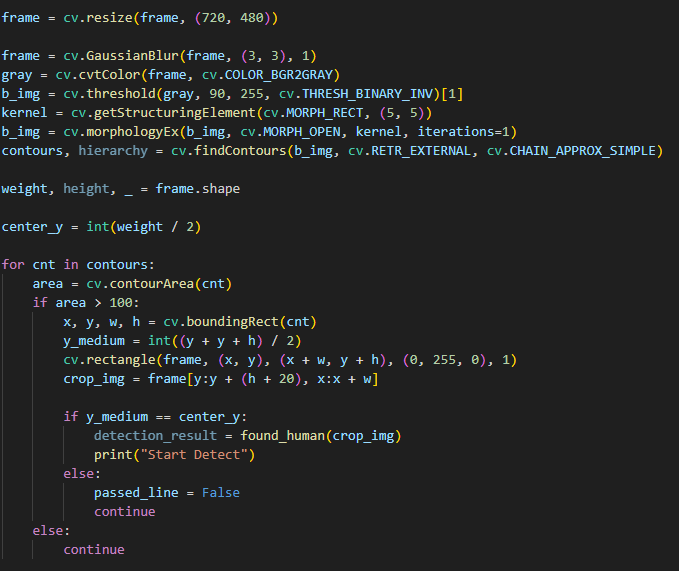
**3.4. Training and Evaluate Model**



**CHAPTER 4: APPLYING MODEL FOR MACHINE VISION**

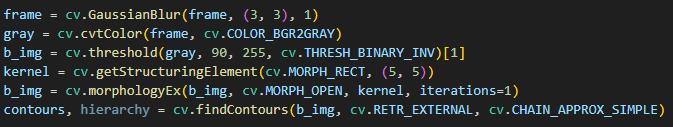
**4.1 Object detection**

-Nhìn vào yêu cầu đề bài là Phát hiện và Couting số người, việc đầu tiên ta cần thực hiện là viết chương trình phát hiện (detect) hình ảnh con người dựa trên mô hình đã huấn luyện. Nếu vật thể phát hiện được là con người thì sẽ tiến hành đếm số người trong khung ảnh.

**Figure 4.1**: Thuật toán phát hiện người

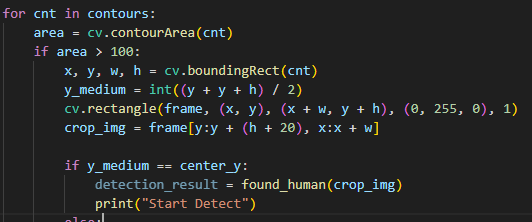
-Đầu tiên, để nói về thuật toán phát hiện con người, phải chắc chắn rằng ta đã có được mô hình huấn luyện con người cần tác động. Tiếp đó sẽ là các bước xử lý như sau, ở đây tôi dùng một video có sẵn, nó có nên trắng và người đi lại nhìn chung cũng tối và rất để detect:

**Bước 1**: Dùng phương pháp xử lí ảnh truyền thống để tìm ra hình ảnh con người dang di chuyển



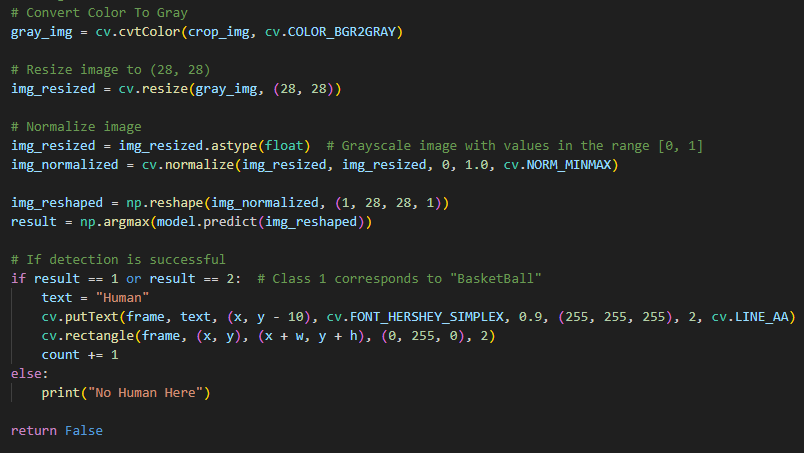
**Figure 4.2**: Xác định ngưỡng threshold của hình dạng người

**Bước 2**: Dùng hàm FindContours để ra khu vực có người xuất hiện và tiến hành cắt phần ảnh đó. Xử lý ảnh qua các bước resize 28x28, chuẩn hóa hình ảnh để phù hợp đầu vào của mô hình huấn luyện.



**Figure 4.3**: Xử lý ảnh cắt để đưa vào mô hình

**Bước 4**: Dự đoán xem phần đã được cắt có phải là con người hay không. Nếu không sẽ tiến hành dự đoán ở những khung hình sau.



**Figure 4.4**: Dự đoán vật thể

Lưu ý một việc rằng ảnh xử dụng trong vùng được cắt phải đưa về định dạng đúng với lúc ảnh được train. Ơ đây ảnh sẽ là ảnh xám resize thành 28x28 và cuối cùng được reshape thành dạng vector.

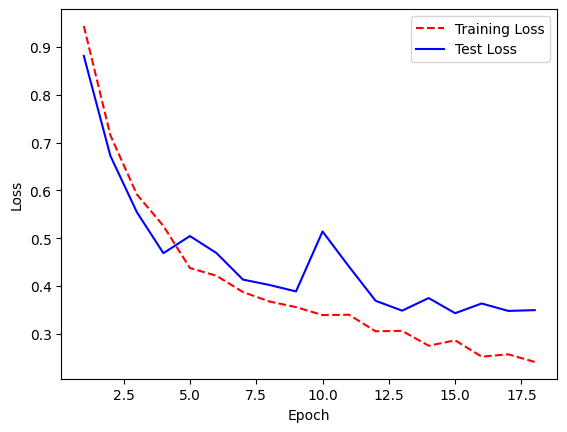
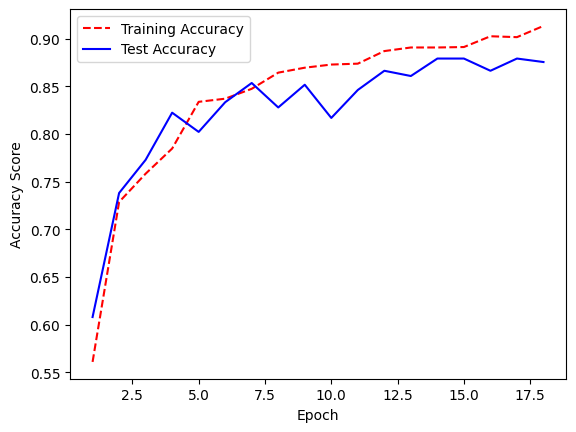
**4.2. Object Couting**

Để có thể đếm được số người trong một khu vực nhất định ta không thể load,crop từng khu vực bouding box trong frame ảnh ra sau đó tiến hành detec được. Vì làm thế hệ số sẽ không thể chạy nổi với 24 khung hình trên giây, đồng thời việc đếm sẽ sai.   
Ta phải khác phục vấn đề sau bằng cách kẻ một line, hoặc phân vùng để đếm nếu tấm của bouding box đó chạm vào khu vực phân vùng nó sẽ bắt đầu crop khung ảnh của bouding box và đồng thời tiến hành predict bằng mô hình CNN, đưa ra kết quả dự đoán nếu đúng là con người nó sẽ cộng lên một.



**CHAPTER 5: RESULTS AND DISCUSSION**

**Evaluate Model**

**Train and loss history:  
 ** **Giá trị Loss:** trong qua trình huấn luyện cho thấy đang giảm dần từ epoch đầu tiên đến epoch cuối cùng.Điều này cho thấy rằng mô hình đang học cách giảm thiểu được sai số.

**Giá Trị Accuracy**: Độ chính xác của mô hình huấn luyện tăng dần qua từng epoch nhìn vào biểu đồ ta có thể nhận biết được rằng với epoch đầu tiên mô hình chỉ có độ chính xác 50% và tăng dần đến 90% ở những epoch cuối cùng. Mô hình đang học cách phân loại chính xác các mẫu dữ liệu.

**Giá trị Validation Loss và Validation Accuracy**: Giá trị validation loss trong quá trình huấn luyện giảm dần từ epoch đầu tiên đến epoch cuối cùng, từ 0.8844 xuống còn 0.348 và giá trị validation accuracy tăng dần từ 0.68lên 0.901. Mô hình có hiệu suất tốt trong việc dự đoán và có khả năng dự đoán chính các giá trị mục tiêu trong tệp kiểm tra.

**Các thông số đánh giá khác:**

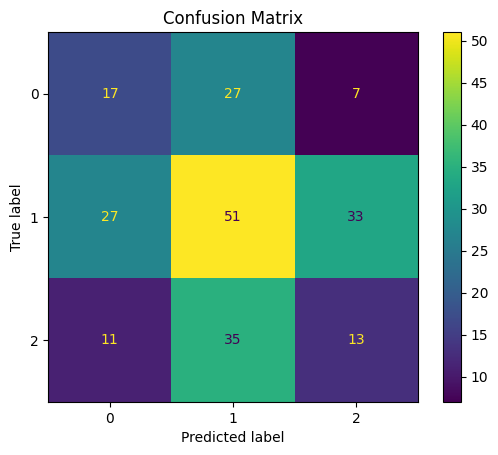
Trên tập dữ liệu thực tế này, tôi đã cố tình ước lượng bên trong đó 50% là các trường hợp đúng và 50% là các trường hợp sai.Dựa vào tệp test đó tôi nhận được kết quả Accuracy của model là 41%,Macro trung bình trên tất cả các lớp là 0.37%,Weighted avg (Trung bình trọng số theo số lượng mẫu) 40%

**Confusion Matrix:**Với ảnh chưa được qua xử lí khi đi vào mô hình mô hình sẽ nhận dạng sai, vì vậy khi bắt đầu tính toán confusion matrix ta có thể thấy được.

Tổng quan, các lớp đều có precision, recall và f1-score khá thấp, cho thấy mô hình gặp khó khăn trong việc phân loại chính xác các lớp dữ liệu. Mô hình có hiệu suất kém và cần được cải thiện để đạt được kết quả tốt hơn trên tập kiểm tra.

Ngoài ra, accuracy trên tập kiểm tra là 0.41, tức là mô hình dự đoán đúng khoảng 41% trên tổng số mẫu. Tuy nhiên, cần lưu ý rằng accuracy không phản ánh đúng hiệu suất của mô hình trong các lớp cụ thể, mà chỉ là tỷ lệ dự đoán chính xác trên toàn bộ tập dữ liệu.

Nhưng trên thực tế khi đã được xủ lí thì ta có được kết quả rất tốt như trên mục 4.2.

****