NHẬN DIỆN KHUÔN MẶT ĐEO KHẨU TRANG SỬ DỤNG MẠNG NƠ-RON TÍCH CHẬP TRONG BỐI CẢNH ĐẠI DỊCH COVID-19

Tác Giả : Đặng Lê Gia Vũ

Khoa Công nghệ Thông tin, Trường Đại học Ngoại ngữ - Tin học Thành phố Hồ Chí Minh
** Liên hệ tác giả : Email: 18dh110138@st.huflit.edu.vn | Điện thoại: 096.778.1273*

Tóm tắt

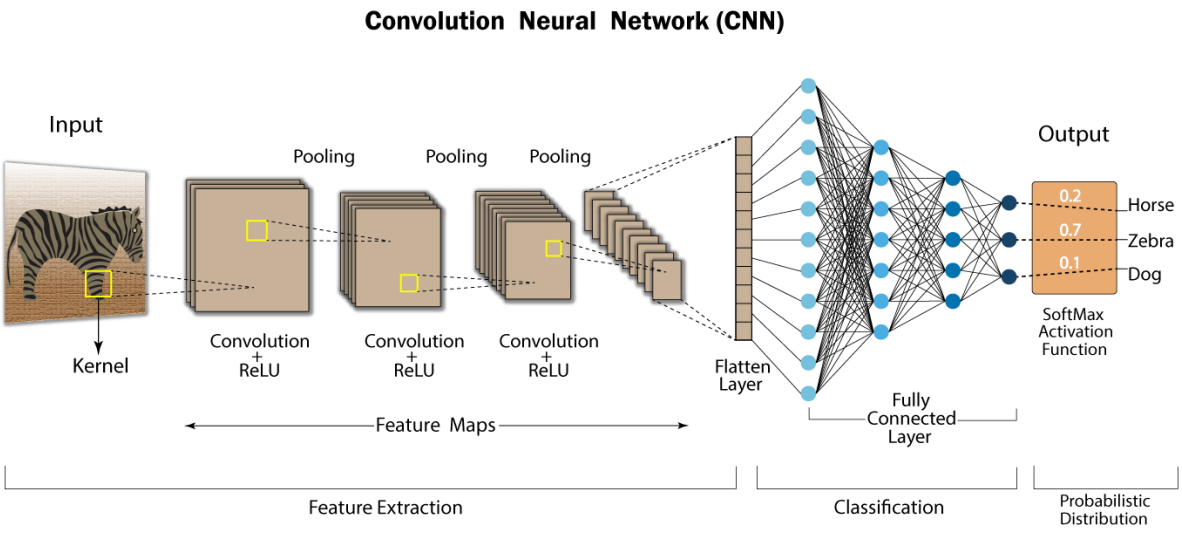
Trong bài nghiên cứu này, tác giả đã sử dụng mạng nơ-ron tích chập và ứng dụng kiến trúc MobileNetV2, một trong những phương pháp học sâu (Deep Learning) để thực hiện phân đoạn theo ngữ nghĩa hình ảnh khuôn mặt có trong video theo thời gian thực với OpenCV, từ đó phát hiện một người có đeo khẩu trang hay không. Hệ thống sử dụng một tập dữ liệu bao gồm 4095 hình ảnh khác nhau gồm: các hình ảnh đeo khẩu trang và không đeo khẩu trang để thực hiện việc huấn luyện. Bài nghiên cứu này có thể tích hợp với các hệ thống sẵn có tại sân bay, sân ga, nơi làm việc hay trường học để đảm bảo an toàn cho cộng đồng. Mô hình đạt độ chính xác 96-99% phụ thuộc vào hình ảnh đầu vào từ camera, cho thấy đây là một trong những phương pháp góp phần hạn chế lây nhiễm dịch bệnh, nhằm bảo vệ sức khỏe cộng đồng.

Từ khoá: Mạng nơ-ron tích chập; Nhận diện khuôn mặt đeo khẩu trang; MobileNetV2; Xử lý ảnh; Thị giác máy tính; OpenCV.

1. GIỚI THIỆU

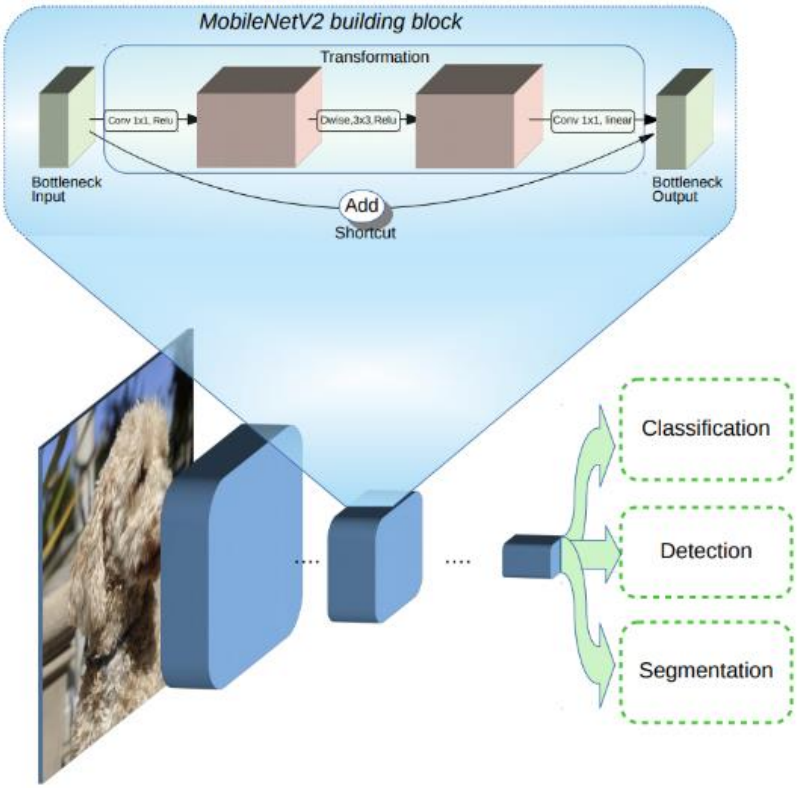
Với bối cảnh đại dịch COVID-19 đang hoành hành, việc đeo khẩu trang, duy trì khoảng cách và khử khuẩn là đặc biệt quan trọng. Virus lây lan từ người sang người thông qua đường tiếp xúc trực tiếp, gián tiếp (qua các vật dụng hoặc bề mặt bị nhiễm mầm bệnh), hoặc tiếp xúc gần với người nhiễm bệnh qua dịch tiết từ miệng và mũi. Dịch tiết này bao gồm nước bọt, dịch tiết hô hấp hoặc các giọt bắn. Đeo khẩu trang trong thời điểm hiện tại là điều bắt buộc với tất cả những người có nguy cơ bị nhiễm bệnh. Khi đó, chúng ta sẽ làm giảm khả năng lây truyền của loại virus chết người này. Trong bối cảnh dịch bệnh như vậy, ngày càng khó để theo dõi con người có đeo khẩu trang một cách thường xuyên hay không. Sức người là có hạn, vì thế cần phải phát minh ra một phần mềm để xử lý vấn đề này.

Mô hình trong bài nghiên cứu này được huấn luyện thông qua mạng nơ-ron tích chập. Được lấy cảm hứng từ vỏ não thị giác, cho phép máy tính có khả năng “nhìn nhận” và “phân tích”. Sử dụng để nhận dạng hình ảnh bằng cách đưa các dữ liệu hình ảnh vào nó qua mạng nơ-ron với nhiều lớp, mỗi lớp sẽ có một bộ lọc các tích chập. Sau khi đi qua các lớp này, chúng ta có được nét đặc trưng và dùng nó để tiến hành nhận dạng đối tượng. Hình 1 là kiến trúc của một mạng nơ-ron tích chập (CNN).

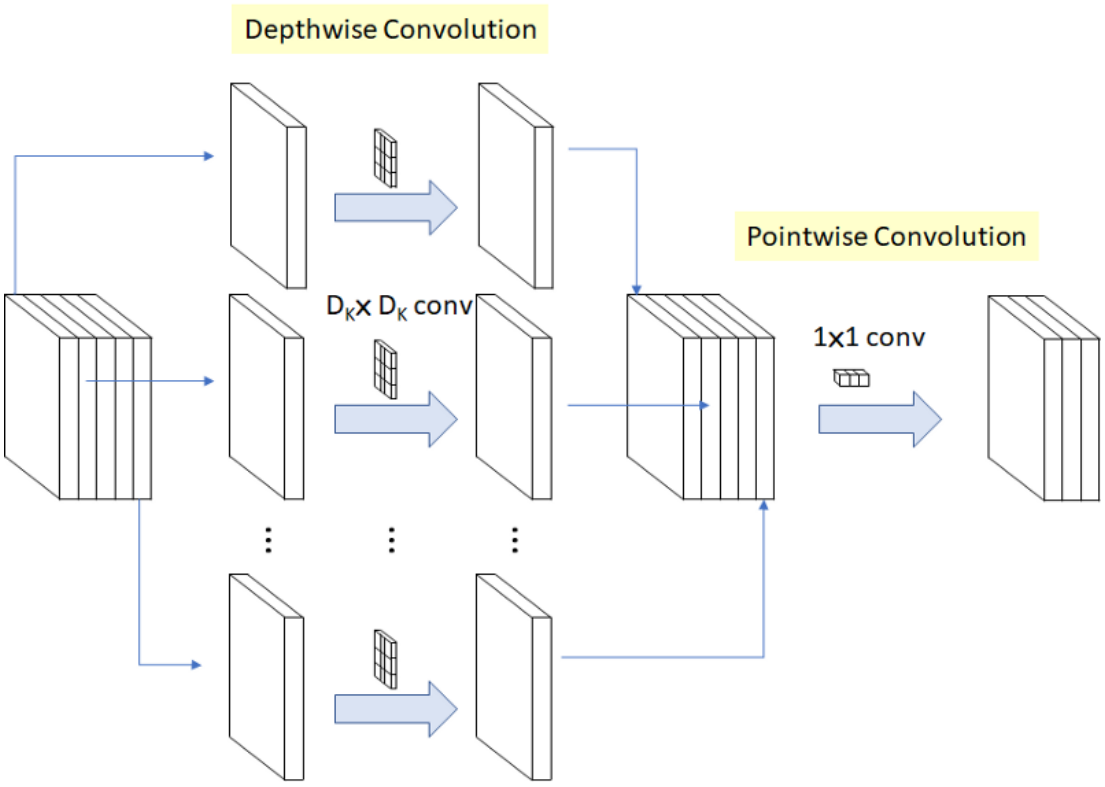


Hình 1. Mô hình mạng nơ-ron tích chập (CNN Model)

Mô hình mạng nơ-ron tích chập được sử dụng ở đây là kiến trúc MobileNetV2. Mô hình MobileNet là một mô hình mạng sử dụng các phép tích chập phân tách theo chiều sâu (Depthwise Separable Convolution). Hình 2 – 3 lần lượt là kiến trúc của mạng MobileNetV2 và cấu trúc của một Depthwise Separable Convolution.



Hình 2. Kiến trúc của mạng MobileNetV2



Hình 3. Cấu trúc của một Depthwise Separable Convolution

Áp dụng framework MobileNet2 với bộ dữ liệu được xáo trộn một cách ngẫu nhiên (nhưng vẫn đảm bảo các nhãn của tập dữ liệu không thay đổi). Siêu tham số được sử dụng lần lượt là tốc độ học (Learning Rate), được dùng để điều chỉnh các mô hình tối ưu hóa, xác định kích thước bất kỳ và giảm hàm mất mát. Đây là một tham số cực kỳ quan trọng để dẫn đến sự hội tụ hoặc vượt quá mô hình (kiểm soát tốc độ mô hình thay đổi). Các siêu tham số khác cũng được sử dụng như : kích thước lô (Batch Size) và huấn luyện theo chu kỳ (Epochs). Ngoài ra mô hình còn sử dụng OpenCV để thực hiện ghi lại các khung hình có trong luồng video theo thời gian thực

2. BÀI NGHIÊN CỨU LIÊN QUAN

I. B. Venkateswarlu và các cộng sự [1] đã đề xuất ra một bài nghiên cứu "Face mask detection using MobileNet and Global Pooling Block" sử dụng kiến trúc MobileNet kết hợp với khối tổng hợp toàn cầu (Global Pooling Block) để phát hiện khuôn mặt đeo khẩu trang. Mô hình này đạt độ chính xác 99% và 100% trên hai tập dữ liệu là DS1 và DS2. Mặc dù mô hình được đề xuất trong bài hơn các mô hình hiện có về chất lượng cũng như thời gian chuẩn bị nhưng mô hình này không thể phát hiện nhiều khuôn mặt đeo khẩu trang cùng một lúc.

Marielet Guillermo và các cộng sự [2] đã đề xuất ra một bài nghiên cứu "COVID-19 Risk Assessment through Face Mask Detection using MobileNetV2 DNN". Nghiên cứu này được tạo ra nhằm thúc đẩy tầm quan trọng của việc kiểm soát dịch bệnh và các biện pháp phòng ngừa như sử dụng khẩu trang nơi công cộng. Việc thực hiện nghiên cứu này thông qua 3 giai đoạn chính : thu thập dữ liệu đeo và không đeo khẩu trang, đào tạo huyến luyện mô hình và thử nghiệm mô hình. Họ đạt được kết quả chính xác là 92%.

Md.Sanzidul Islam và các cộng sự [3] đã đề xuất ra một bài nghiên cứu "Adeep Learning Based assestive system to classify COVID-19 Face Mask for Human Safety with YOLOv3" sử dụng kiến trúc YOLOv3. Bài nghiên cứu của họ tập trung vào xây dựng phát hiện đối

tượng để nhận diện đối tượng tùy ý và phát hiện khuôn mặt đeo khẩu trang từ video. Họ đạt được kết quả chính xác là 96%.

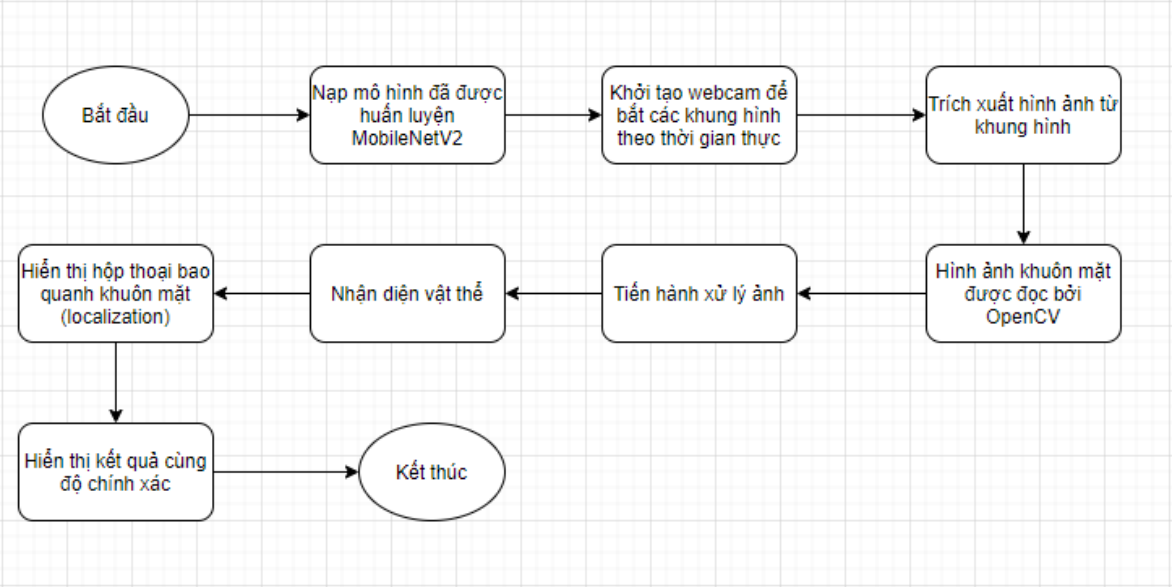
3. PHƯƠNG PHÁP NGHIÊN CỨU

3.1.Tổng quan

Bài nghiên cứu sử dụng mạng nơ-ron tích chập (CNN) để phân loại hình ảnh có đeo khẩu trang hay không. Mạng nơ-ron tích chập là một loại hình học sâu (Deep Learning) nhận một hình ảnh đầu vào. Sau đó ánh xạ các đặc trưng khác nhau qua từng lớp để thu về kết quả. Mô hình đề xuất ở đây được thiết kế và mô hình hóa bằng cách sử dụng các thư viện và framework của ngôn ngữ lập trình Python như : Tensorflow, Keras và OpenCV. Mô hình sử dụng là mô hình MobileNetV2 mạng nơ ron tích chập.

Phương pháp sử dụng MobileNetV2 được gọi là sử dụng học tập chuyển giao (Transfer Learning). Học tập chuyển giao là việc ứng dụng các mô hình được đào tạo trước đó làm bàn đạp dự đoán, để đào tạo mô hình hiện tại giúp cho việc đào tạo tiết kiệm thời gian hơn.. Các siêu tham số đưa vào điều chỉnh là : tốc độ học (Learning Rate), số lượng chu kỳ (Number Of Epochs) và kích thước lô (Batch Size). Tập dữ liệu bao gồm : 2165 tấm ảnh có đeo khẩu trang và 1930 tấm ảnh không đeo khẩu trang. Trong bài nghiên cứu này, tác giả đã thử nghiệm mô hình với các siêu tham số khác nhau và kết quả sẽ được đề cập phần tiếp theo của bài nghiên cứu này.

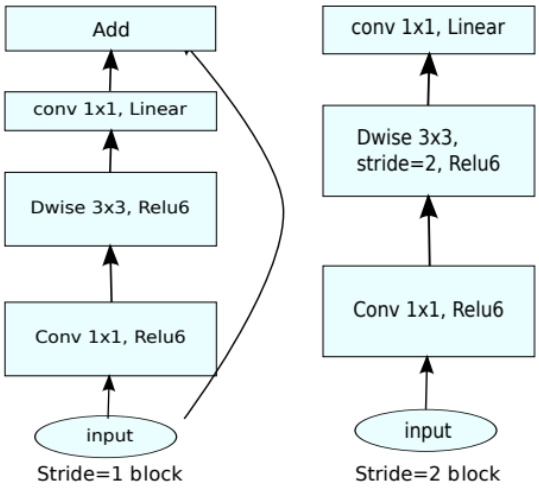
Đầu tiên, tác giả sẽ tiến hành cung cấp tập dữ liệu cho mô hình, sau đó khởi chạy quá trình huấn luyện đào tạo mô hình trên tập dữ liệu đã cung cấp. Sau khi mô hình được huấn luyện xong, tiến hành bật luồng video để lấy các khung hình liên tục có trong video bằng camera. Quá trình này sẽ được chuyển qua các lớp có trong mô hình MobileNetV2 để phân loại hình ảnh có hoặc không đeo khẩu trang. Nếu người đó đang đeo khẩu trang thì hộp thoại hiển thị đối tượng sẽ hiển thị màu xanh ngược lại sẽ là màu đỏ kèm theo tỷ lệ chính xác. Mô hình nhận diện này sử dụng công nghệ trí tuệ nhân tạo để phát hiện có đeo khẩu trang hay không. Nó có thể dễ dàng kết nối, tích hợp với bất kỳ hệ thống có sẵn nào. Mô hình đạt độ chính xác từ 96 - 99% tùy thuộc vào hình ảnh đầu vào cung cấp bởi camera. Hình 4 cho biết quy trình thực hiện của mô hình nhận diện khuôn mặt đeo khẩu trang.



Hình 4. Sơ đồ mô tả quy trình nhận dạng khuôn mặt đeo khẩu trang

3.2.**Kiến trúc của mạng MobileNetV2**

Như đã giới thiệu, MobileNetV2 là một mô hình mạng sử dụng các phép tích chập phân tách theo chiều sâu (Depthwise Separable Convolution) do Google phát triển. Trong mạng MobileNetV2 có 53 lớp convolution và 1 AvgPool với gần 350 GFLOP. Nó gồm có 2 thành phần chính : Inverted Residual Block và Bottleneck Residual Block. Hơn nữa, nó còn có 2 loại lớp chập chính là : 1x1 Convolution và 3x3 Depthwise Convolution. Mỗi thành phần có 3 lớp khác nhau gồm : 1x1 Convolution với phi tuyến Relu6, Depthwise Convolution và 1x1 Convolution không có phi tuyến. Hình 5 – 6 lần lượt là 2 thành phần khác nhau trong mạng MobileNetV2 và 3 loại lớp khác nhau có trong mỗi thành phần.



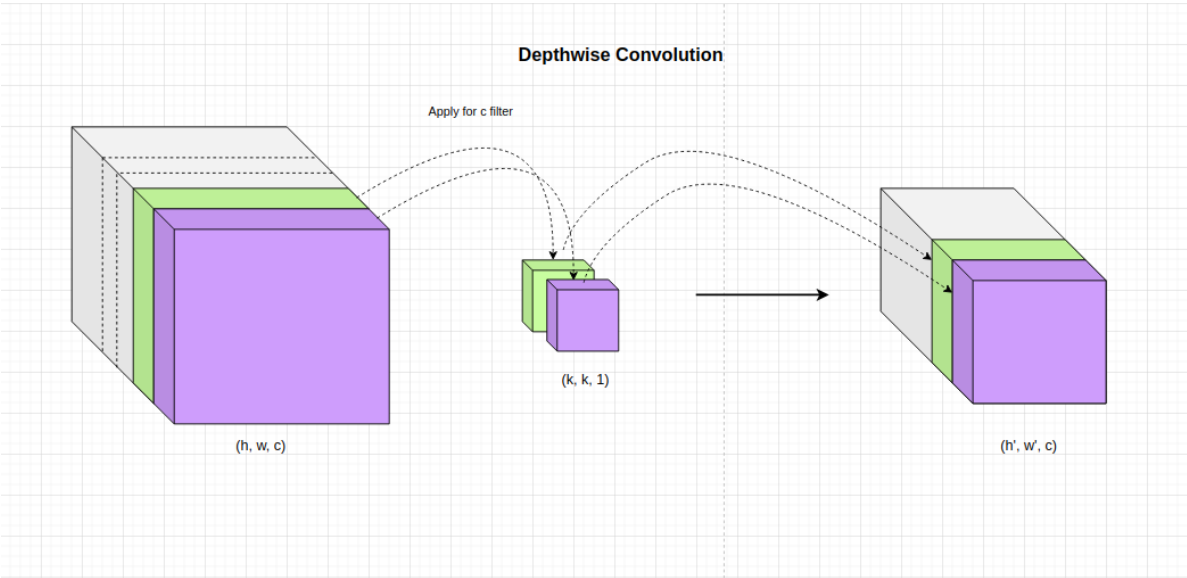
(d) Mobilenet V2

Hình 5. Hai thành phần khác nhau có trong mạng MobileNetV2

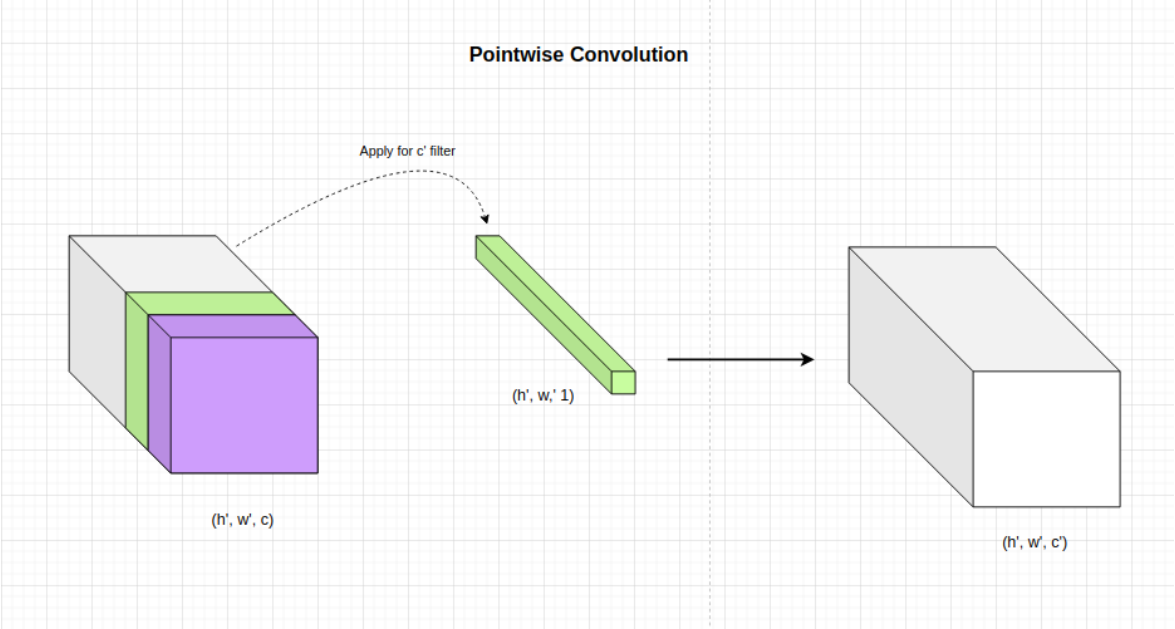
Input	Operator	Output
$h \times w \times k$	1x1 conv2d , ReLU6	$h \times w \times (tk)$
$h \times w \times tk$	3x3 dwise s=s, ReLU6	$\frac{h}{s} \times \frac{w}{s} \times (tk)$
$\frac{h}{s} \times \frac{w}{s} \times tk$	linear 1x1 conv2d	$\frac{h}{s} \times \frac{w}{s} \times k'$

Hình 6. Ba loại lớp khác nhau có trong mỗi thành phần của mạng MobileNetV2

Chúng ta nhận định rằng độ sâu là một trong những nguyên nhân chính dẫn tới sự gia tăng số lượng tham số của mô hình. Tích chập tách biệt chiều sâu sẽ tìm cách loại bỏ sự phụ thuộc vào độ sâu khi chập mà vẫn tạo ra được kết quả (Output Shape) có kích thước tương đương so với khi chập thông thường. Nó gồm có hai lớp : Tích chập theo chiều sâu (Depthwise Convolution) và Tích chập điểm (Pointwise Convolution). Nhìn chung MobileNetV2 có một số điểm cải tiến giúp cho nó có độ chính xác cao hơn, số lượng tham số và số lượng các phép tích ít hơn. Hình 7 và 8 lần lượt là kiến trúc của một mạng MobileNetV2 và cấu trúc của một Depthwise Separable Convolution.



Hình 7. Tích chập theo chiều sâu : các khối input (tensor3D) được chia thành những lát cắt ma trận theo chiều sâu, thực hiện tích chập trên từng lát cắt.



Hình 8. Tích chập theo điểm : có tác dụng thay đổi độ sâu của output bước trên từ c sang c'.Chúng ta sẽ áp dụng c' bộ lọc kích thước 1x1xc. Như vậy witdth, height không đổi mà chỉ có depth thay đổi

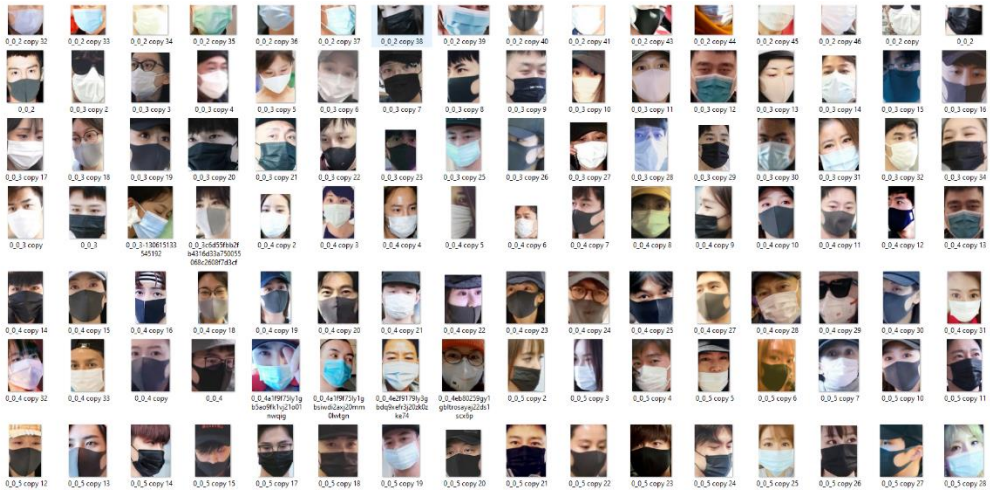
Để cùng tạo ra một output shape có kích thước $h' \times w' \times c'$ thì số lượng tham số đưa vào ở một mạng tích chập thông thường sẽ là : $c' \times k \times k \times c$. Trong khi đó ở tích chập tách biệt theo chiều sâu ta chỉ cần : $k \times k \times c + c' \times c$. Vì độ sâu sẽ tăng dần, nên $c' > c$. Vì vậy ta có tỷ lệ : $\frac{c' \times k \times k \times c}{k \times k \times c + c' \times c} = \frac{c' \times k \times k}{k \times k + c'}$

Đây là một mức giảm khá lớn về kích thước mô hình. Vì thế mà MobileNet có kích thước gọn nhẹ, thích hợp dùng để phát triển được những ứng dụng AI trên các thiết bị như di động.

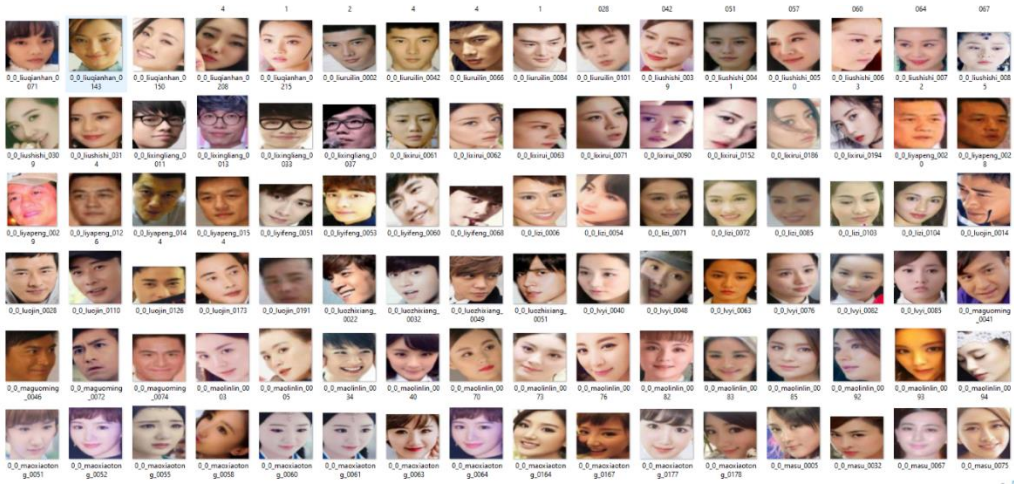
4. XÂY DỰNG MÔ HÌNH

4.1.1. Thu thập dữ liệu

Quá trình huấn luyện của bài nghiên cứu bắt đầu bằng việc thu thập dữ liệu. Bộ dữ liệu dùng huấn luyện bao gồm hình ảnh về những người đeo khẩu trang và không đeo khẩu trang. Với tập dữ liệu (lấy từ nhiều nguồn như Kaggle, BingSearchAPI, RMFD) bao gồm 4095 tấm ảnh. Trong đó có 2165 tấm ảnh đeo khẩu trang và 1930 tấm ảnh không đeo khẩu trang. Tại mỗi bước, hình ảnh sẽ được cắt cho đến khi nhận diện được khuôn mặt. Sau đó sẽ tiến hành gán nhãn dữ liệu, dữ liệu được thu thập được gán nhãn 2 nhóm: có đeo khẩu trang và không đeo khẩu trang. Hình 10 – 11 lần lượt là hình ảnh đeo khẩu trang và không đeo khẩu trang có trong tập dữ liệu.



Hình 9. Tập dữ liệu chứa các hình ảnh khuôn mặt đeo khẩu trang



Hình 10. Tập dữ liệu chứa các hình ảnh khuôn mặt không đeo khẩu trang

4.1.2. Tiền xử lý

Tiền xử lý là giai đoạn trước khi thực hiện việc đào tạo và kiểm tra dữ liệu. Có 4 bước trong quá trình tiền xử lý bao gồm : điều chỉnh kích thước hình ảnh, chuyển đổi hình ảnh sang mảng, tiền xử lý đầu vào với MobileNetV2 và cuối cùng là biến đổi các nhãn cho dữ liệu để gán vào dữ liệu. Hình 9 mô tả quy trình xây dựng mô hình.

Điều chỉnh kích thước hình ảnh là một bước tiền xử lý cực kỳ quan trọng trong lĩnh vực thị giác máy tính, nó sẽ làm tăng hiệu quả huấn luyện mô hình. Kích thước hình ảnh càng nhỏ thì càng tốt cho mô mình đem đi huấn luyện. Trong bài nghiên cứu này, hình ảnh đầu vào sẽ được điều chỉnh về kích thước 224 x 224 pixels.

Bước tiếp theo trong quy trình tiền xử lý chính là biến đổi các hình ảnh trong tập dữ liệu thành một mảng. Hình ảnh được ánh xạ thành mạng để có thể gọi các phần tử đó bên trong vòng lặp. Sau đó đem vào mạng MobileNetV2 để xử lý.

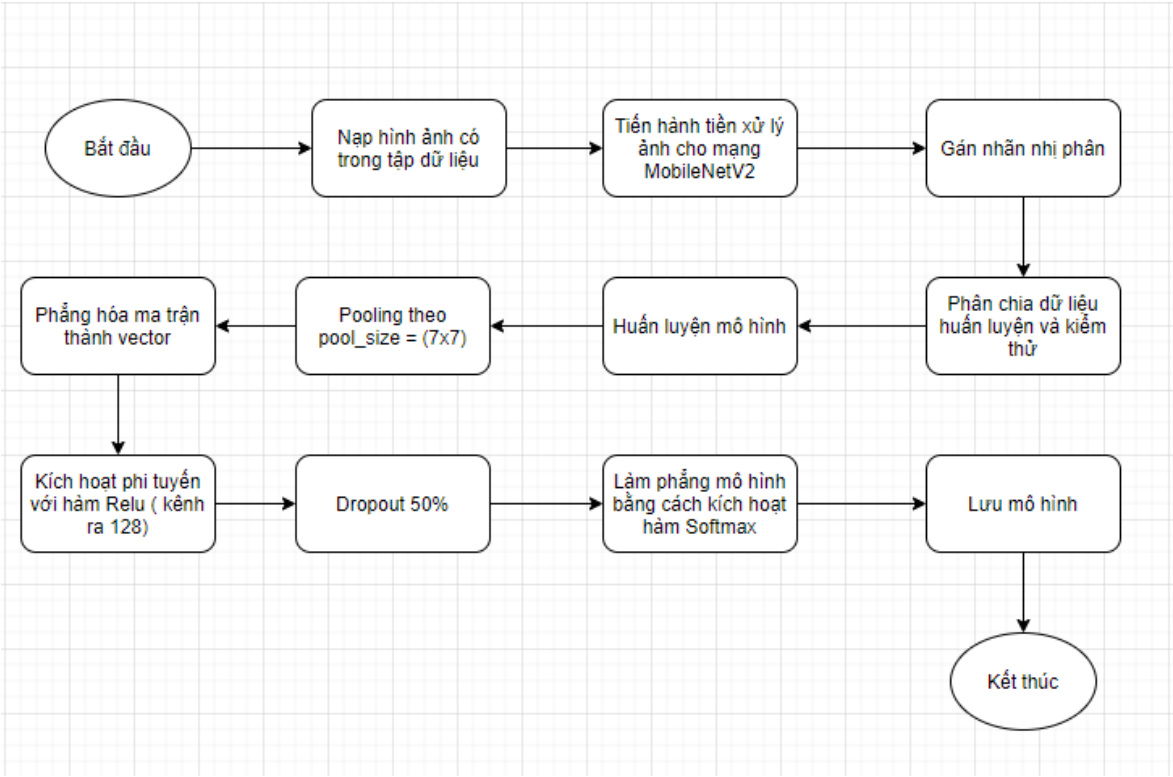
Bước cuối cùng trong giai đoạn tiền xử lý là thực hiện mã hóa các nhãn dữ liệu. Bởi vì một số thuật toán không thể hoạt động được trên các nhãn dữ liệu thông thường, vì vậy cần ánh xạ các nhãn dữ liệu thành một nhãn các số để thuật toán có thể hiểu và xử lý.

4.1.3. Phân chia dữ liệu

Sau giai đoạn tiền xử lý, tập dữ liệu sẽ được chia thành 2 lô (Batch) bao gồm : 75% được sử dụng để huấn luyện và 25% để kiểm tra. Mỗi lô chứa dữ liệu hình ảnh có đeo và không đeo khẩu trang.

4.1.4. Các siêu tham số

Các siêu tham số được đưa vào để tinh chỉnh trong quá trình thiết kế mô hình lần lượt là : tốc độ học (Learning Rate), số chu kỳ (Epoch) và kích thước lô (Batch Size).

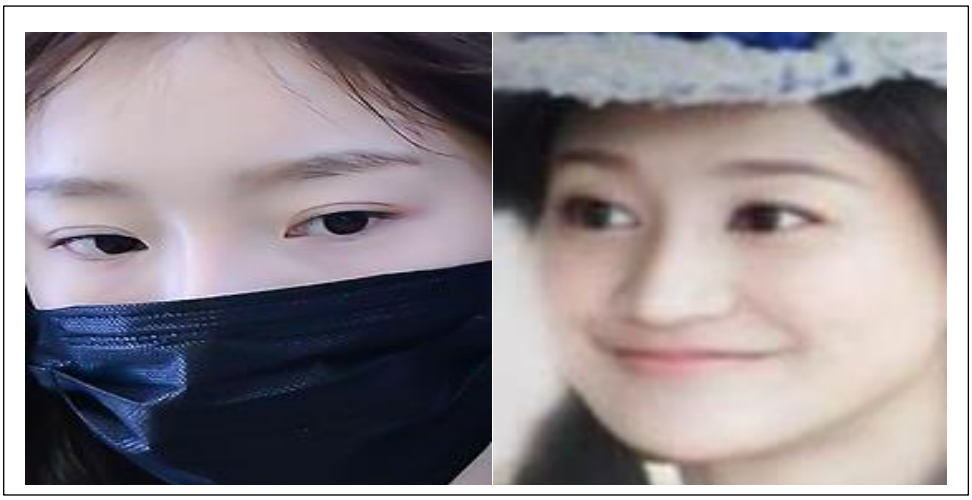


Hình 11. Sơ đồ mô tả các bước thiết kế mô hình

5. QUÁ TRÌNH HUẤN LUYỆN

5.1.1. Dữ liệu đầu vào

Dữ liệu đầu vào là các hình ảnh có trong tập dataset bao gồm hình ảnh đeo khẩu trang và không đeo khẩu trang. Sau đó được biến đổi về kích thước 224x224 pixel để tiến hành huấn luyện. Hình 9 là ví dụ về hai loại hình ảnh có trong tập dữ liệu.



Hình 12. Tập dữ liệu gồm 2 loại hình ảnh : đeo và không đeo khẩu trang

5.1.2. Dữ liệu đầu ra

Dữ liệu đầu ra là một tensort2D chứa nhãn ở dạng one-hot-encoded.

5.1.3. Các tầng có trong mô hình

Mô hình gồm có các tầng như sau: tầng kiến trúc của MobiletNetV2, AveragePooling2D với pool_size là 7x7, tầng Flatten, tầng Dropout và tầng Fully Connected (Dense). Sử dụng hàm Softmax để trích xuất tạo vector thể hiện xác suất của mỗi lớp.

5.1.4. Hàm mất mát và hàm tối ưu

Để đánh giá tập trọng số cần xác định lỗi cho cả quá trình huấn luyện (loss) và kiểm tra (val_loss) ta sử dụng hàm Cross Entropy là hàm nhị phân chéo binary_crossentropy. Cụ thể tính toán loss của mỗi trường hợp bằng cách tính giá trị trung bình như sau :

$$loss_{(y,\hat{y})} = -\left(\frac{1}{n}\sum_{i=1}^{n} y_i \, log(\hat{y}_i) + (1-y_i)\, log(1-\hat{y}_i)\right) \qquad (1)$$

Với n là số lượng giá trị vô hướng trong đầu ra của mô hình, hàm loss trả về một số thực không âm thể hiện sự chênh lệch giữa hai đại lượng đang xét : $\boldsymbol{y}$ là xác suất của nhãn đúng và $\hat{\boldsymbol{y}}$ là xác suất nhãn đang được dự đoán.

Như chúng ta đã biết, thuật toán tối ưu (optimizer) là cơ sở để xây dựng mô hình mạng nơ-ron với mục đích "học" các đặc trưng của dữ liệu đưa vào, từ đó tìm được cặp trọng số weight và bias hợp lý để tối ưu hóa mô hình. Để tiến hành tối ưu và hội tụ nhanh, trong bài nghiên cứu này tác giả sử dụng thuật toán Gradient Descent "Adam" - Adam (Adaptive Moment Estimation). Adam optimizer là một thuật toán được kết hợp với kỹ thuật của RMSprop và momentum. Thuật toán sử dụng hai internal states moementum (m) và squared momentum (v) của gradient cho các tham số. Sau mỗi lô (batch) huấn luyện, giá trị của m và v được cập nhật lại bằng cách sử dụng exponential weighted averaging. Trong khi momentum có thể xem như một quả bóng đang chạy xuống dốc, Adam lại giống như một quả bóng nặng với ma sát. Adam sử dụng bình phương gradient để chia tỷ lệ learning rate như RMSProp và tận dụng "đà" giống như momentum. Cụ thể công thức tối ưu của nó là :

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad (2)$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) {g_t}^2 \quad (3)$$

$$\hat{m}_t = \frac{m_t}{1-{\beta_1}^t} \quad (4)$$

$$\hat{v}_t = \frac{v_t}{1-{\beta_2}^t} \quad (5)$$

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \quad (6)$$

Để tốt hơn về chi phí tính toán, (4) – (5) – (6) có thể viết lại thành (7) – (8) như sau :

$$\eta_t = \eta \frac{\sqrt{1-{\beta_2}^t}}{1-{\beta_1}^t} \quad (7)$$

$$\theta_t = \theta_{t-1} - \eta_t \frac{m^t}{\sqrt{v_t} + \varepsilon} \quad (8)$$

6. ĐÁNH GIÁ THỰC NGHIỆM

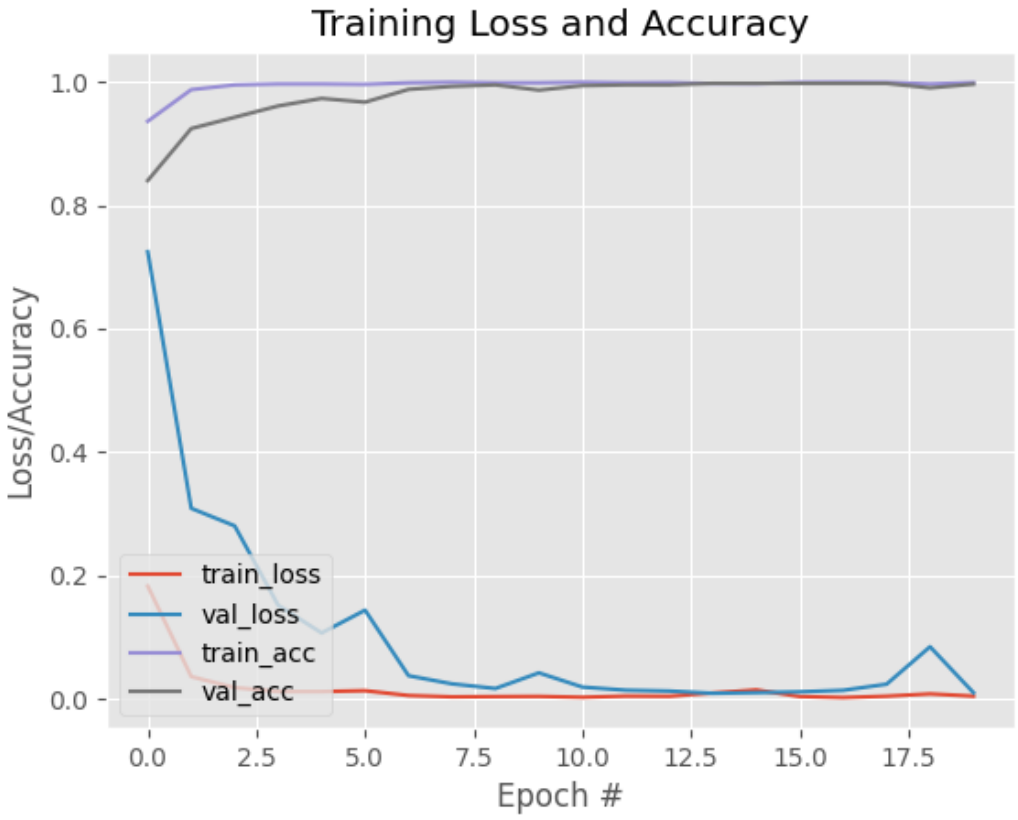
Tác giả đã thử nghiệm mô hình cho từng trường hợp khác nhau, bên dưới là bảng kết quả của những tình huống đã thử nghiệm với tốc độ học (Learning Rate) là 0.0001. Số lượng chu kỳ (Number Of Epochs) là 20 và kích thước lô (batch size) là 32. Từ đó ghi nhận kết quả chính xác lên đến 99%. Trong bài nghiên cứu [3], mô hình sử dụng kiến trúc YOLOv3 đạt độ chính xác 96% từ đó cho thấy mô hình sử dụng MobileNetV2 có độ chính xác cao hơn.

Bảng 1. Bảng biểu diễn quá trình lặp để kiểm tra sự mất mát và độ chính xác

Epoch	Loss	Accuracy	Val loss	Val acc
1/20	0.3796	0.8531	0.1356	0.9805
2/20	0.1389	0.9648	0.0793	0.9792
3/20	0.1030	0.9654	0.0621	0.9829
4/20	0.0793	0.9679	0.0552	0.9829
5/20	0.0725	0.9806	0.0508	0.9853
6/20	0.0565	0.9824	0.0465	0.9853
7/20	0.0541	0.9840	0.0457	0.9853
8/20	0.0494	0.9855	0.0460	0.9853
9/20	0.0445	0.9836	0.0460	0.9866
10/20	0.0439	0.9846	0.0435	0.9878
11/20	0.0405	0.9883	0.0402	0.9878

12/20	0.0372	0.9877	0.0393	0.9878
13/20	0.0348	0.9880	0.0381	0.9853
14/20	0.0328	0.9898	0.0397	0.9866
15/20	0.0353	0.9904	0.0366	0.9853
16/20	0.0321	0.9898	0.0355	0.9866
17/20	0.0359	0.9883	0.0418	0.9890
18/20	0.0307	0.9911	0.0361	0.9890
19/20	0.0328	0.9898	0.0391	0.9890
20/20	0.0325	0.9901	0.0376	0.9890

Từ bảng 1, có thể thấy độ chính xác càng càng tăng dần bắt đầu vào chu kỳ thứ 2 (epoch 2) và độ mất mát giảm dần sau đó. Mô hình đạt độ chính xác khoảng 99%, đây là một giá trị tương đối cao. Bảng sau có thể được biểu diễn ở dạng đồ thị như trong hình 13.



Hình 13. Đồ thị biểu diễn mất mát và độ chính xác trong quá trình huấn luyện

Khi đường biểu diễn tỉ lệ chính xác đang ổn định, đồng nghĩa rằng không cần lặp lại nhiều lần hơn để tăng độ chính xác của mô hình. Vì vậy, bước tiếp theo là thực hiện đánh giá mô hình như trong bảng 3.

Để kiểm tra tính hiệu quả của mô hình, cần tính toán tỷ lệ chính xác trung bình trên tất cả các dự đoán, sử dụng thang đo ma trận nhầm lẫn (Confusion Matrix) như sau :

Bảng 2. Bảng biểu diễn Confusion Matrix

	Mô hình dự đoán đeo khẩu trang **(Predicted as Positive)**	**Mô hình dự đoán không đeo khẩu trang** **(Predicted as Negative)**
Ảnh đeo khẩu trang **(Actual Positive)**	TP (True Positive)	FN (False Negative)
Ảnh không đeo khẩu trang **(Actual Negative)**	FP (False Positive)	TN(True Negative)

Trong đó : Các hàng của ma trận là nhãn lớp thực tế, các cột của ma trận là nhãn lớp dự đoán

- *TN là số lượng khuôn mặt không đeo khẩu trang được phân loại chính xác.*
- *FN là số lượng khuôn mặt đeo khẩu trang bị phân loại nhầm là khuôn mặt không đeo khẩu trang.*
- *TP là số lượng khuôn mặt đeo khẩu trang được phân loại chính xác.*
- *FP là số lượng khuôn mặt không đeo khẩu trang bị phân loại nhầm là khuôn mặt không đeo khẩu trang.*

Từ đó, độ chính xác của mô hình được tính theo công sức sau :

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

Đây là tỉ lệ của tất cả các trường hợp hợp phân loại đúng trên toàn bộ trường hợp trong mẫu kiểm tra.

Một thang đo khá là phổ biến thường được dùng để đánh giá mô hình phân lớp, đó là F-Measure hay F-Score. Bảng 3 biểu diễn đánh giá mô hình trên thang F-Score. Được tính dựa trên 2 độ đo khác nhau là Precision và Recall. Cụ thể như sau :

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

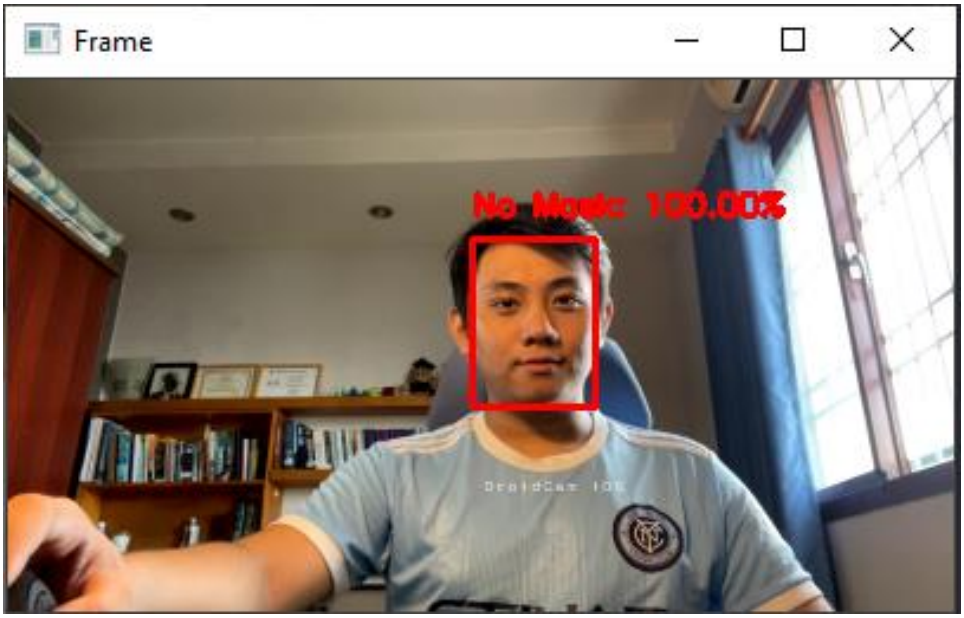
$$F_{score} = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} = \frac{TP}{TP+\frac{FN+FP}{2}} \quad (12)$$

Bảng 3. Bảng đánh giá mô hình

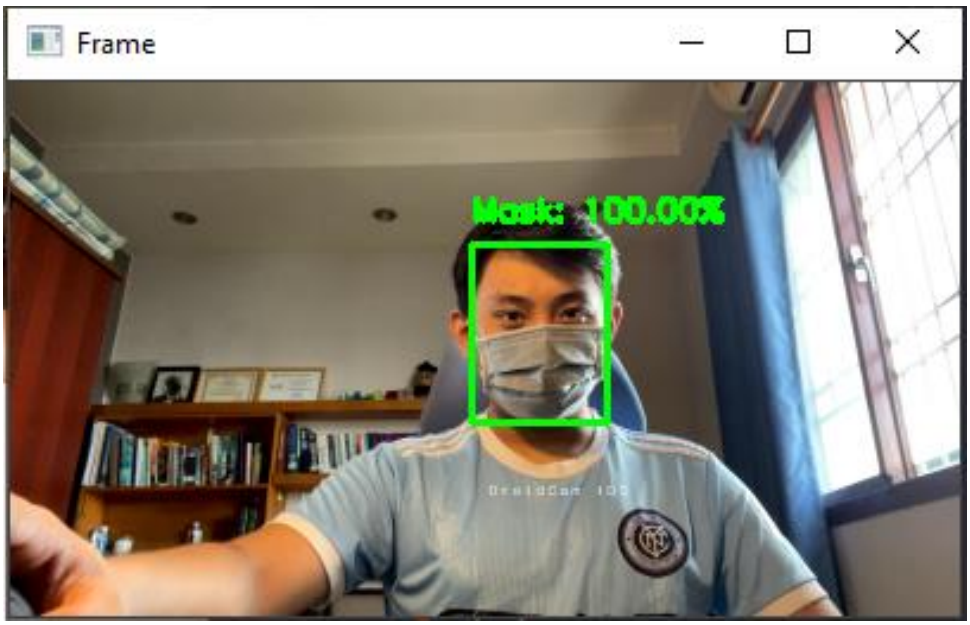
	Precision	Recall	F1-Score	Support
With mask	0.99	0.99	0.99	433
Without mask	0.99	0.98	0.99	386
Accuracy				
Macro avg	0.99	0.99	0.99	819
Weighted avg	0.99	0.99	0.99	819

7. TRIỂN KHAI MÔ HÌNH

Mô hình thực hiện thông qua việc thu thập các khung hình có trong video để thực hiện thuật toán nhận diện khuôn mặt. Khi quét được khuôn mặt nào, nó sẽ tìm các khung hình tiếp theo có chứa khuôn mặt và quá trình này được xử lý lặp đi lặp lại. Từ các khung hình đã xác định được khuôn mặt,ta lần lượt thay đổi kích thước hình ảnh, chuyển đổi sang mảng, tiền xử lý bằng MobileNetV2. Bước tiếp theo là tiến hành dự đoán dữ liệu đầu vào từ dữ liệu đã được lưu vào mô hình trước đó.Bên cạnh đó, khung hình có trong camera sẽ được hiển thị là có đeo khẩu trang hay không và phần trăm dự đoán. Hình 14 và 15 là một ví dụ về việc triển khai mô hình.



Hình 14. Kết quả dự đoán khi không đeo khẩu trang



Hình 15. Kết quả dự đoán khi có đeo khẩu trang

8. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Để kiểm soát sự lây lan của đại dịch COVID-19, tác giả đã phát triển ra một giải pháp để phát hiện xem một người có đeo khẩu trang hay không bằng cách sử dụng mạng tích chập (CNN) và học tập chuyển giao (Transfer Learning). Mô hình sử dụng tập dữ liệu bao gồm: 2165 tấm ảnh có đeo khẩu trang và 1930 tấm ảnh không đeo khẩu trang để tiến hành huấn luyện. Những hình ảnh này được lấy từ các nguồn khác nhau như tập dữ liệu Kaggle, BingSearchAPI và RMFD. Bài nghiên cứu sử dụng MobileNetV2, OpenCV, Tensorflow, Keras và mạng tích chập (CNN). Trong tương lai, tác giả sẽ mở rộng bài nghiên cứu này sang việc nhận diện nhiều khuôn mặt cùng một lúc. Bài nghiên cứu này có thể sử dụng để tích hợp với các hệ thống sẵn có tại nơi công cộng một cách dễ dàng nhằm hạn chế sự lây lan của virus, để bảo vệ sự sống trọn vẹn cho toàn nhân loại. Độ chính xác lên đến 99% khiến bài nghiên cứu có thể đem vào ứng dụng trong thực tiễn, từ đó đem lại một công cụ hữu ích trong công cuộc bảo vệ sức khỏe toàn dân.

TÀI LIỆU THAM KHẢO

[1] I. B. Venkateswarlu, J. Kakarla and S. Prakash, "Face mask detection using MobileNet and Global Pooling Block" 4 2020 IEEE 4th Conference on Information & Communication Technology (CICT), 2020.

[2] Athena Rosz Ann Pascua,Marielet Guillermo, "COVID-19 Risk Assessment through Face Mask Detection using MobileNetV2 DNN", International Symposium on Computational Intelligence and Industrial Applications, 2020.

[3] Rafiuzzaman Bhuiyam, Sharun akter Khsushbbu, Md.Sanzidul Islam, "Adeep Learning Based assestive system to classify COVID-19 Face Mask for Human Safety with YOLOv3",IEEE,2019.

[4] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. "Mobilenetv2: The next generation of on-device computer vision networks. URL Https://Ai. Googleblog. Com/2018/04/Mobilenetv2- next-Generation-of-on. 2020.

[5] Cheng, Kar Keung; Lam, Tai Hing; Leung, Chi Chiu, "Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity". The Lancet, 2020.

[6] Hussain, S. A., & Al Balushi, A. S. A. "A real time face emotion classification and recognition using deep learning model". Journal of Physics: Conference Series, 1432(1), 12087, 2020.

[7] Ejaz, M. S., Islam, M. R., Sifatullah, M., & Sarker, A., "Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition". 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 1–5.

[8] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic". Measurement: Journal of the International Measurement Confederation, 2020, doi: 10.1016/j.measurement.2020.108288

[9] Murray, O. M., Bisset, J. M., Gilligan, P. J., Hannan, M. M., & Murray, J. G., "Respirators and surgical facemasks for COVID-19: implications for MRI. Clinical Radiology, 75(6), 405–407. https://doi.org/10.1016/j.crad.2020.03.029, 2020.

[10] Feng, S., Shen, C., Xia, N., Song, W., Fan, M., & Cowling, B. J., "Rational use of face masks in the COVID-19 pandemic". The Lancet Respiratory Medicine, 8(5), 434–436, 2020.