

NHẬN DIỆN DANH TÍNH MẶT NGƯỜI KHI ĐEO VÀ KHÔNG ĐEO KHẨU TRANG (REAL TIME BẰNG CAMERA)

Trần Quang Huy

Đại học Sư Phạm Kỹ Thuật TP.HCM

Môn học: Trí tuệ nhân tạo

GVGD: PSG.TS Nguyễn Trường Thịnh

Mã học phần: ARIN337629

Tóm tắt

Nghiên cứu này đồng thời là bài báo cáo cuối kì môn Trí tuệ nhân tạo nên mục tiêu chính của đề tài này là nhận diện danh tính người khi đeo và không đeo khẩu trang, và số người nhận diện cụ thể là 5 người. Các ảnh mẫu được tạo, với độ phân giải 150x150, kết quả thử nghiệm gồm 3688 mẫu cho khả năng phát hiện chính xác 91.61% Bằng cách xây dựng mô hình CNN (Convolutional Neural Network) – Mạng nơ-ron tích chập để có thể nhận dạng người khi đeo và không đeo khẩu trang thông qua các đặc tính của mặt người khi đeo và không đeo khẩu trang, và mô hình sẽ được áp dụng và chạy thời gian thực.

ĐẶT VẤN ĐỀ

Hiện tại, dịch bệnh COVID-19 vẫn có diễn biến phức tạp ở Việt Nam nói riêng và trên toàn thế giới nói chung, đại dịch này qua từng ngày lại cướp đi hàng triệu sinh mạng trên toàn thế giới do tốc độ lây lan của virus rất nhanh và chúng có khả năng biến đổi thành nhiều loại biến thể giúp thích nghi với nhiều môi trường sống

khác nhau, nó không chỉ ảnh hưởng đến kinh tế, thương mại, dịch vụ mà còn ảnh hưởng không nhỏ đến tâm lý xã hội của người dân và chưa có vacxin điều trị cho nên cần kiểm soát tốt nguồn lây nhiễm. Chính vì thế có những nghiên cứu chỉ ra rằng giữ khoảng cách xã hội, rửa tay và đeo khẩu trang là các biện pháp có hiệu quả trong việc giảm tỷ lệ mắc COVID-19, trong đó đeo khẩu trang mang lại hiệu quả cao nhất. Theo trung tâm kiểm soát bệnh tật CDC (Centers for Disease Control and Prevention), khẩu trang được khuyến nghị là một rào chắn đơn giản để giúp ngăn chặn các giọt bắn từ đường hô hấp bay vào không khí và lên người khác. Khuyến nghị này dựa trên những nghiên cứu về vai trò của những giọt nước bọt bắn lên các bề mặt kim loại, con người từ đó đi vào đường hô hấp, kết hợp với bằng chứng mới xuất hiện từ các nghiên cứu lâm sàng và trong phòng thí nghiệm cho thấy khẩu trang làm giảm việc phun các giọt bắn khi đeo khẩu trang qua miệng và mũi. Vì vậy việc sử dụng khẩu trang đặc biệt quan trọng ở thời điểm hiện tại nhằm hạn chế việc lây lan bệnh [5].

Như vậy, phát hiện người đeo khẩu trang đề cập đến việc phát hiện xem một người đeo khẩu trang hay không và vị trí của khuôn mặt đó trở thành một nhiệm vụ quan trọng nhưng nghiên cứu liên quan đến vấn đề này chưa phổ biến. Trên thế giới, đã có một số nghiên cứu phát hiện người đeo hay không đeo khẩu trang. Tại Việt Nam, Robot do Đại học Công nghệ (ĐH Quốc gia) chế tạo để phát hiện người không đeo khẩu trang và nhắc nhở sử dụng những thiết bị, linh kiện có sẵn. Camera tích hợp cảm biến laser để đo khoảng cách, phát hiện người ra vào cửa. Cùng lúc, camera sẽ thu lại hình ảnh và truyền tới máy tính. Các chi tiết chính trên khuôn mặt như mắt, mũi, miệng sẽ được số hóa và xử lý. Nếu máy tính phát hiện mũi và miệng bị che, có nghĩa là người đó đang đeo khẩu trang. Còn trong trường hợp miệng bị che nhưng mũi hở, hoặc cả mũi và miệng đều hở, nghĩa là người đó đeo chưa đúng cách hoặc

không đeo. Như vậy, phát hiện khẩu trang là một nhiệm vụ cần thiết trong tình hình hiện nay, mục tiêu ngoài nhắc nhở những cá nhân không đeo khẩu trang, mà còn tạo ra dữ liệu thống kê giúp chính quyền dự đoán sự bùng phát của COVID-19 trong tương lai.

NỘI DUNG NGHIÊN CỨU

2.1. Bài toán phát hiện khẩu trang

Để phát hiện đối tượng trong ảnh hay video thì bài toán cần phải xử lý được người trong ảnh hoặc trong video có đeo khẩu trang hay không. Bài toán phân loại này gồm hai bước là xây dựng mô hình và vận hành mô hình. Xây dựng mô hình sẽ dựa trên một tập dữ liệu với dữ liệu là khuôn mặt của 5 người đeo và không đeo khẩu trang được gán nhãn sẵn và đầu ra của mô hình sẽ là 10 với 5 khuôn mặt không đeo khẩu trang và 5 khuôn mặt đó khi đeo khẩu trang. Sau khi xây dựng mô hình phân lớp từ dữ liệu mẫu, sau đó mô hình này sẽ dự đoán những khuôn mặt của 5 người khi chưa biết nhãn.

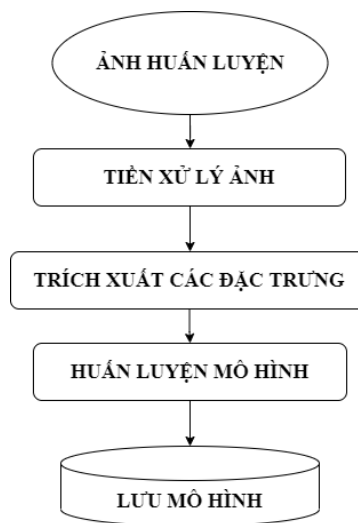
2.2. Tiền xử lý dữ liệu

Sau khi thu thập bộ dữ liệu của 5 khuôn mặt khi đeo và không đeo khẩu, sau đó tiến hành tải bộ dữ liệu lên Kaggle để chia sẻ dữ liệu. Bộ dữ liệu gồm 10 thư mục, 5 thư mục được gán nhãn của 5 người khi không đeo khẩu trang, 5 thư mục được gán nhãn của khuôn mặt 5 người khi đeo khẩu trang [1]. Tập dữ liệu đầu tiên là khuôn mặt người đi đeo khẩu trang. Tiến hành chia tập dữ liệu thành 2 phần. Phần thứ nhất là `train_dataset`, dùng được huấn luyện cho mô hình. Phần thứ hai là `validation_dataset`, được dùng để kiểm tra độ chính xác của mô hình. Ở đây, nhằm tăng cường dữ liệu cho bộ dữ liệu sẽ sử dụng hàm `ImageDataGenerator` để biến đổi đào tạo, giúp cho mô hình đã học có thể mạnh mẽ và chính xác hơn. Các thuộc tính

zoom_range và shear_range lần lượt thực hiện zoom ngẫu nhiên, làm méo ảnh. Sử dụng phương thức phân tách dữ liệu validation_split, và chọn giá trị giá trị là 0.2 tương ứng với sẽ lấy 80% của tập dữ liệu để huấn luyện và 20% còn lại sẽ được dùng để kiểm tra. Và chuyển đổi tỷ lệ ảnh từ số nguyên 0-255 và dạng float 0-1. Kích thước của những tấm ảnh đầu vào là 150 x 150.

2.3. Huấn luyện mô hình

2.3.1. Thuật toán huấn luyện mô hình



Hình 2.1. Thuật toán huấn luyện mô hình

Mô hình huấn luyện được xây dựng dựa trên mô hình mạng nơ-ron tích chập CNN với thuật toán như hình 2.1.

2.3.2. Xây dựng các lớp cho mô hình

Các lớp được xây dựng dựa trên mô hình mạng nơ-ron tích chập CNN. Mô hình được xây dựng với 5 lớp, với lớp đầu tiên ảnh đầu vào có kích thước 150x150x3 (tương ứng với 3 màu đỏ, xanh lục, xanh lam trong hệ màu RGB thông thường), sau mỗi lớp tích chập là những lớp chuyển tiếp để giảm kích thước của nơ-ron đi và

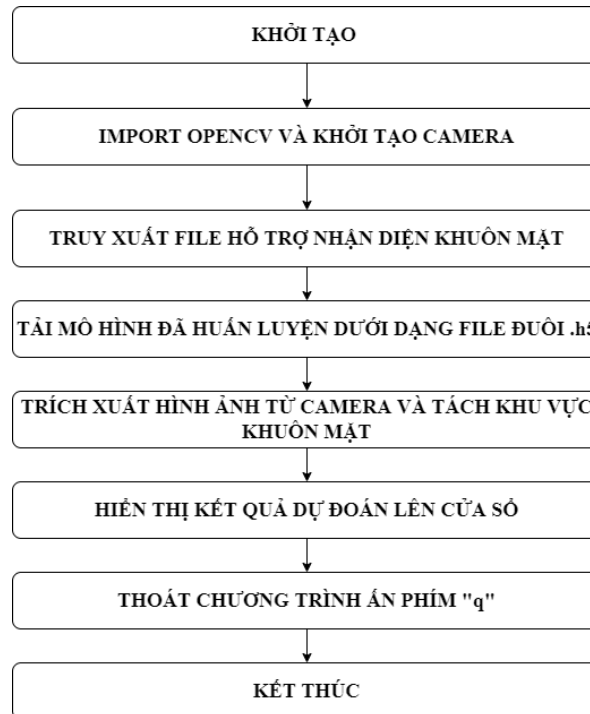
tập trung vào các phần tử quan trọng nhất. Trong phần này ta sử dụng lớp pooling có kích thước (2x2) và sử dụng kiểu max pooling hay lấy giá trị lớn nhất, chiều sâu được giữ nguyên. Một lớp được kết nối đầy đủ làm phẳng các đối tượng được xác định trong các lớp trước đó. Trong lớp Dense cuối cùng, sử dụng hàm softmax để xuất ra một véc – tơ thể hiện xác suất của mỗi lớp. Để giúp cho việc huấn luyện mô hình trở nên tốt hơn trong hàm compile sử dụng để training model em tiến hành sử dụng thuật toán tối ưu optimizers cụ thể là Stochastic Gradient Descent (SGD) , tuy SGD sẽ làm giảm tốc độ của 1 epoch, nhưng sau vài lần epoch SGD sẽ hội tụ rất nhanh, cùng với đó là momentum giúp cho giải quyết được bài toán Gradient Descent không tiến được tới điểm global minimum mà chỉ dừng lại ở local mimum. Ở hàm loss , sử dụng categorical_crossentropy để có thể dùng classifier nhiều class. Tập dữ liệu sau khi được chia thành 1 tập huấn luyện và 1 tập để kiểm tra, ta tiến hành huấn luyện mô hình với batch_size là 32 , epochs là 200 và verbose là 1 để có thể hiển thị tiến trình hoạt động của việc huấn luyện. Sau quá trình huấn luyện, mô hình đạt độ chính xác 91.61%

```
Model: "sequential_1"
-----
Layer (type)                 Output Shape              Param #
-----
conv2d_4 (Conv2D)            (None, 150, 150, 128)    3584
-----
max_pooling2d_4 (MaxPooling2 (None, 75, 75, 128)      0
-----
conv2d_5 (Conv2D)            (None, 75, 75, 128)    147584
-----
max_pooling2d_5 (MaxPooling2 (None, 37, 37, 128)      0
-----
conv2d_6 (Conv2D)            (None, 37, 37, 256)    295168
-----
max_pooling2d_6 (MaxPooling2 (None, 18, 18, 256)      0
-----
conv2d_7 (Conv2D)            (None, 18, 18, 512)    1188160
-----
max_pooling2d_7 (MaxPooling2 (None, 9, 9, 512)      0
-----
flatten_1 (Flatten)          (None, 41472)           0
-----
dropout_1 (Dropout)          (None, 41472)           0
-----
dense_2 (Dense)              (None, 128)             5308544
-----
dense_3 (Dense)              (None, 512)             66048
-----
wearmask (Dense)             (None, 10)              5130
-----
Total params: 7,006,218
Trainable params: 7,006,218
Non-trainable params: 0
```

Hình 2.2. Tổng các tham số của mô hình

2.4. Kiểm nghiệm mô hình trên thời gian thực

2.4.1. Thuật toán phát hiện người đeo khẩu trang thời gian thực

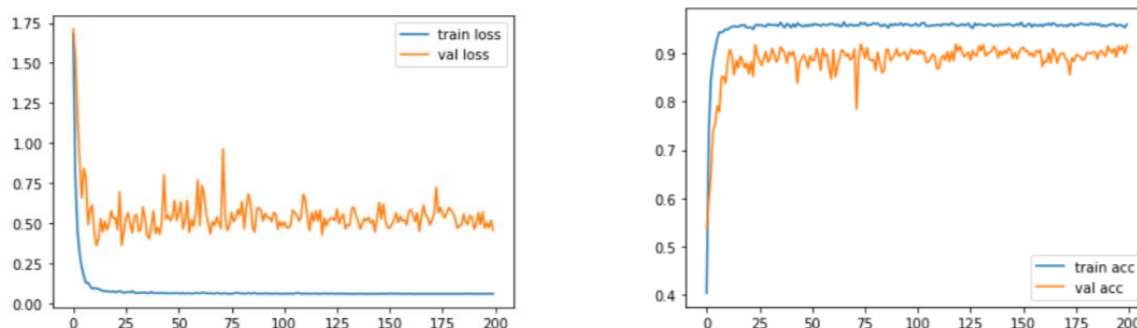


Hình 2.2. Thuật toán phát hiện người đeo khẩu trang trong thời gian thực

Để phát hiện khuôn mặt, thuật toán được sử dụng là thuật toán phát hiện khuôn mặt Viola-Jones [4]. Sau khi ảnh được cắt từ video, nó sẽ được định dạng lại kích thước giống với kích thước của dữ liệu trong mô hình huấn luyện. Sau đó, ảnh này sẽ được chuyển đổi thành những tham số có định dạng giống mô hình mẫu. Khi đó các tham số sinh ra được từ tập dữ liệu huấn luyện sẽ được sử dụng để thẩm định lại tính thích hợp của mô hình trên tập dữ liệu của hình ảnh vừa được trích xuất. Dựa vào kết quả thu được ta tiến hành hiển thị lên màn hình kết quả của 1 trong 5 người có đeo khẩu trang hay không.

KẾT QUẢ

Sau 1 lần training mô hình với độ chính xác (accuracy) 91.61%, độ mất mát giảm còn 45.35% . Hình cho thấy dạng đồ thị của độ mất mát và độ chính xác của mô hình



Hình 2.3. Đồ thị bên trái là độ mất mát, bên phải là độ chính xác của mô hình

Chương trình đã thực hiện kiểm tra trong nhiều điều kiện khác nhau, cho ra kết quả khá tốt như hình 2.4 và 2.5 hình dưới đây



Hình 2.4. Kết quả khi phát hiện danh tính người đeo khẩu trang trong thời gian thực



Hình 2.5. Kết quả khi phát hiện danh tính người đeo khẩu trang trong thời gian thực

KẾT LUẬN

Bài báo cáo đã xây dựng thành công mô hình nhằm xác định danh tính của 5 người khi đeo và không đeo khẩu trang trong thời gian thực. Tuy nhiên điều kiện ánh sáng kèm việc phát hiện khẩu trang vẫn chưa thực sự hiệu quả. Chương trình khi phân biệt người đeo khẩu trang và không đeo khẩu trang đôi lúc vẫn chưa phân biệt được đúng. Chương trình được viết bằng ngôn ngữ Python và sử dụng một số thư viện mã nguồn mở như Keras, Python, OpenCV,... Dựa vào kết quả thu được, để có thể giúp cho mô hình tốt hơn, tập dữ liệu phải được tăng cường mạnh mẽ hơn nữa. Trong tương lai chương trình có thể kết hợp với các thiết bị phần cứng như Raspberrypi [2], Arduino, STM32 ,... để xây dựng các hệ thống giám sát và phát hiện danh tính người, nhắc nhở người dân đeo khẩu trang ở những nơi đông đúc như trường học, trung tâm thương mại,...

TÀI LIỆU THAM KHẢO

- [1] <https://www.kaggle.com/datasets/infloop1/testdataset-cl3b-ver2..>
- [2] W. Gay, Book Raspberrypi Hardware Reference, Apress 2014.
- [3] https://en.wikipedia.org/wiki/Convolutional_neural_network.. [Online].
- [4] P. V. a. M. J. Jones, Robus real-time face detection, Int. J. Comput.Vision, Vol.57, No.2,pp.137-154, May 2004.
- [5] <https://123docz.net/document/7269131-phat-hien-khau-trang-su-dung-mo-hinh-hoc-sau-mobilenetv2.htm..> [Online].