



TRƯỜNG ĐẠI HỌC
SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH
HCMC University of Technology and Education

DỰ ĐOÁN TUỔI TÁC, CẢM XÚC, GIỚI TÍNH DỰA TRÊN KHUÔN MẶT REALTIME

Bùi Minh Tú

Đại học Sư phạm kỹ thuật thành phố Hồ Chí Minh, 19146295@student.hcmute.edu.vn

Mô tả: Nhận dạng khuôn mặt người là một công nghệ được ứng dụng rộng rãi trong đời sống hằng ngày của con người như các hệ thống giám sát, quản lý vào ra, tìm kiếm thông tin một người nổi tiếng,... Có rất nhiều phương pháp nhận dạng khuôn mặt để nâng cao hiệu suất tuy nhiên dù ít hay nhiều những phương pháp này đang vấp phải những thử thách về độ sáng, hướng nghiêng, kích thước ảnh, hay ảnh hưởng của tham số môi trường. Bài báo này sẽ đề cập đến việc sử dụng thuật toán CNN (Mạng nơ-ron tích chập) để xây dựng 3 mô hình giới tính, tuổi và cảm xúc để dự đoán dựa trên khuôn mặt real-time. Để thực hiện đề tài này, dữ liệu hình ảnh đã thu thập được từ nhiều nguồn khác nhau với khoảng 47000, 9000 và 28000 hình ảnh tương ứng với mỗi mô hình. Công việc thực hiện bao gồm thay đổi kích thước hình ảnh, giải mã và tiêu chuẩn hóa rồi sau đó sẽ được đào tạo và kiểm tra để xác định độ chính xác của các mô hình.

1. Đặt vấn đề

Với sự phát triển không ngừng của khoa học và công nghệ, đặc biệt là các thiết bị được hỗ trợ công nghệ xử lý ảnh ngày càng hiện đại và được sử dụng phổ biến trong đời sống con người đã làm cho lượng thông tin thu được bằng hình ảnh ngày càng tăng và phổ biến. Theo đó, lĩnh vực xử lý ảnh cũng được chú trọng phát triển, ứng dụng rộng rãi trong đời sống xã hội hiện đại. Không chỉ dừng lại ở việc chỉnh sửa, tăng chất lượng hình

ảnh mà với công nghệ xử lý ảnh hiện nay chúng ta có thể giải quyết các bài toán nhận dạng chữ viết, nhận dạng dấu vân tay, đặc biệt là nhận dạng khuôn mặt...

Công nghệ nhận diện khuôn mặt hiện là một công nghệ đang được sử dụng khá phổ biến tại các quốc gia phát triển. Công nghệ này có khả năng xác định hoặc xác nhận một người từ hình ảnh kỹ thuật số được lấy mẫu trước đó hoặc từ một khung hình trong một nguồn video khác. Đây là một phương pháp xác minh độc đáo khi thiết bị sẽ dựa vào những điểm khác nhau tiêu biểu nhất trên khuôn mặt của một người để tiến hành phân biệt giữa người này với người khác, dự đoán tuổi tác hay cảm xúc của mọi người. Chính vì đặc điểm này thì ngoài được ứng dụng trong việc quản lý nhân sự ra thì nó còn là sự lựa chọn của rất nhiều đơn vị hoạt động trong lĩnh vực an ninh, bảo mật.

Xây dựng một ứng dụng nhận dạng khuôn mặt nhằm giúp việc quản lý, điểm danh, giao dịch hay thống kê theo từng yêu cầu phục vụ cho các mục đích khác nhau là cấp thiết. Ngoài ra, giúp sinh viên tìm hiểu về thư viện OpenCV, tìm hiểu các phương pháp xác định khuôn mặt (Face Detection).

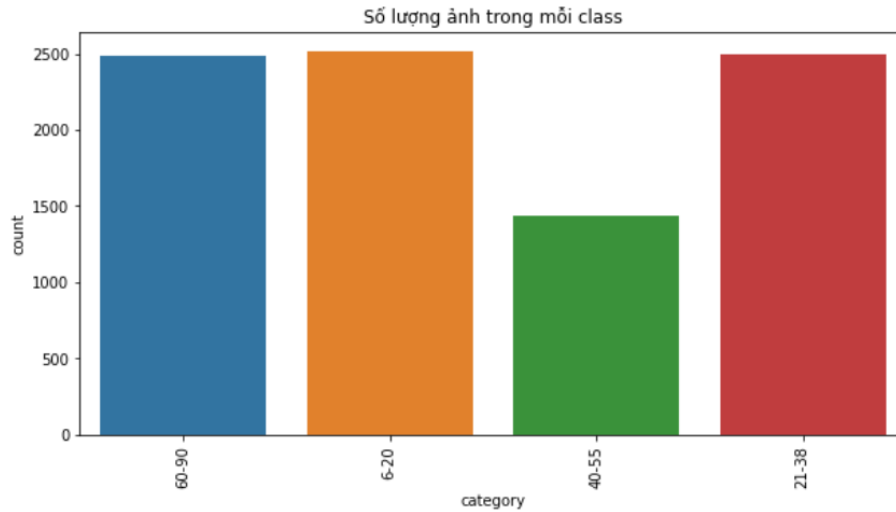
2. Phương pháp nghiên cứu

Đây là phần trình bày phương pháp và tài liệu, bao gồm các chi tiết của tập dữ liệu thử nghiệm, các bước xử lý trước dữ liệu được thực hiện để thiết lập các thử nghiệm và phương pháp được áp dụng để giải quyết mục đích nghiên cứu.

2.1. Tập dữ liệu

2.1.1. “Age recognition dataset ”

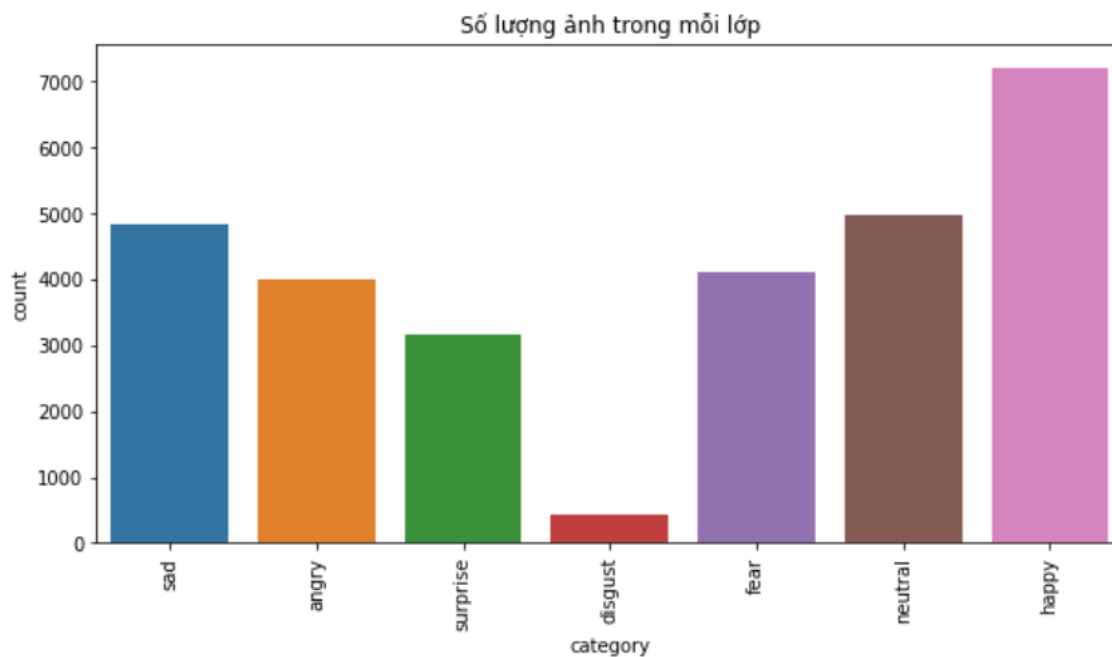
Đây là tập dữ liệu về tuổi tác được đăng lên Kaggle bởi Rashik Rahman. Gồm có 4 classes bao gồm: 6-20, 25-30, 42-48, 60-98. Do các độ tuổi trong các lớp không sát nhau nên em đã lọc, tìm kiếm thêm và phân lại các độ tuổi giúp việc dự đoán dễ hơn. Sau khi phân lại thì tập cũng gồm 4 classes: 6-20, 21-38, 40-55, 60-90.



Hình 1. Biểu đồ thể hiện số lượng ảnh trong mỗi lớp trong tập “Age recognition dataset”

2.1.2. “FER-2013”

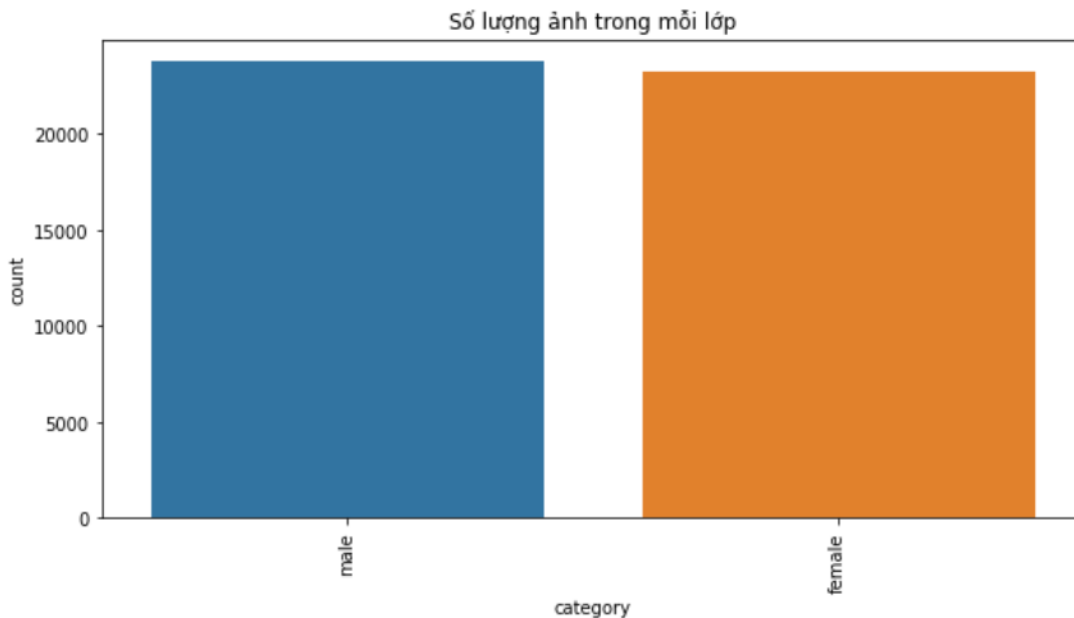
Đây là tập dữ liệu về cảm xúc được đăng lên Kaggle bởi Mansas Sambare. Gồm có 7 classes bao gồm: angry, disgust, fear, happy, neutral, sad, surprise. Theo biểu đồ bên dưới ta có thể thấy dữ liệu trong các lớp không đều và đang gặp tình trạng imbalanced.



Hình 2. Biểu đồ số lượng ảnh trong mỗi lớp của tập “FER-2013”

2.1.3. “Gender Classification Dataset”

Đây là tập dữ liệu về giới tính gồm 2 lớp female và male, được đăng lên Kaggle bởi Ashutosh Chauhan.



Hình 3. Biểu đồ số lượng ảnh trong 2 lớp của “Gender Classification Dataset”

2.2. Xử lý dữ liệu

Do tập dữ liệu có kích thước nhỏ sẽ ảnh hưởng đến hiệu suất của mô hình. Vì vậy cần phải tăng kích thước lên bằng kỹ thuật tăng cường hình ảnh. Tăng cường hình ảnh là một kỹ thuật hữu ích được sử dụng để tăng tính đa dạng của tập dữ liệu huấn luyện sao cho các bản sao thực tế nhưng ngẫu nhiên của hình ảnh gốc có thể được tạo ra thông qua các phép biến đổi đơn giản như thay đổi hình học và không gian màu, cắt hình ảnh, chèn nhiễu và xóa ngẫu nhiên. Sử dụng ImageDataGenerator được cung cấp bởi thư viện Keras để tăng cường dữ liệu hình ảnh.

2.3. Tăng cường dữ liệu

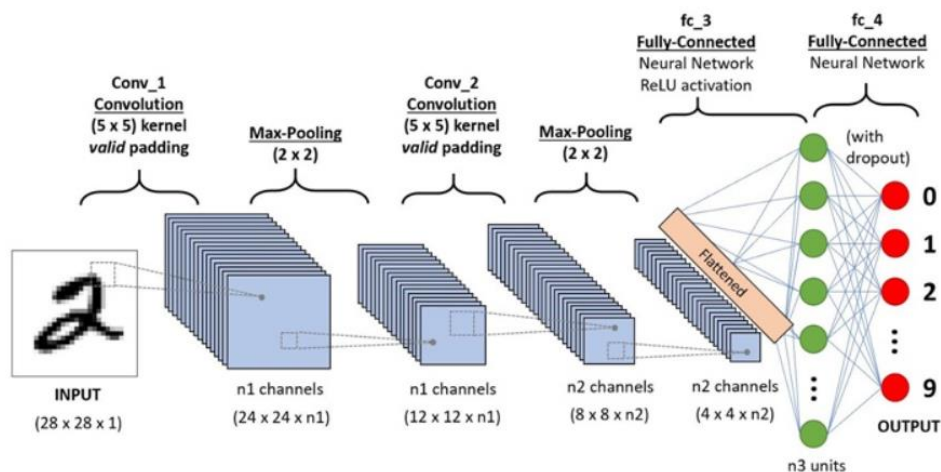
Tăng cường dữ liệu giúp khắc phục vấn đề “không đủ dữ liệu”, khắc phục các điểm overfitting và giúp cho mô hình hoạt động tốt hơn với đa dạng mẫu dữ liệu. Data

Augmentation (Tăng cường dữ liệu) là một kỹ thuật được sử dụng để mở rộng kích thước của tập huấn luyện bằng cách tạo thêm dữ liệu đã được sửa đổi từ dữ liệu ban đầu. Trong project này, các phương thức được sử dụng bao gồm:

horizontal_flip, vertical_flip, rescale=1./255, validation_split=0.1, rotation_range=40, brightness_range= [1.1, 1.5], zoom_range=0.2, width_shift_range = 0.2, height_shift_range = 0.2, shear_range= 0.2, fill_mode='nearest'

2.4. Thuật toán CNN

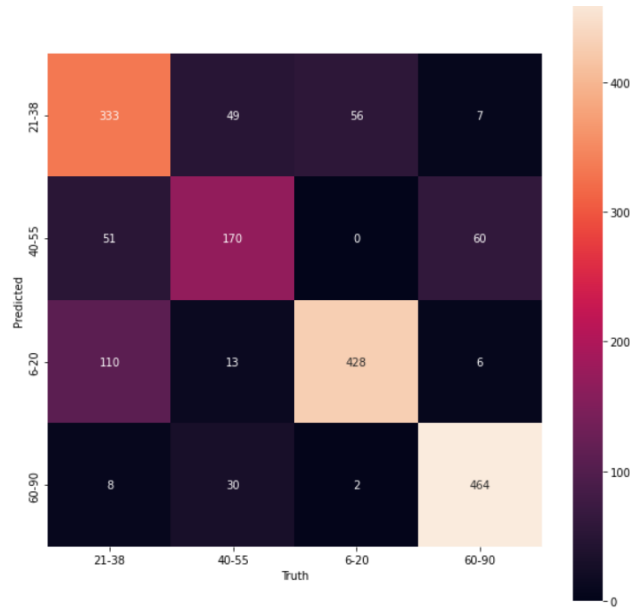
Convolutional Neural Networks (CNN) là một trong những mô hình deep learning phổ biến nhất và có ảnh hưởng nhiều nhất trong cộng đồng Computer Vision. CNN được dùng trong nhiều bài toán như nhận dạng ảnh, phân tích video, ảnh MRI, hoặc cho bài các bài của lĩnh vực xử lý ngôn ngữ tự nhiên, và hầu hết đều giải quyết tốt các bài toán này. Trong đề tài này, mô hình CNN được xây dựng với ảnh input có kích thước 96x96.^[1]



Hình 4. Cấu trúc CNN

2.5. Đánh giá mô hình

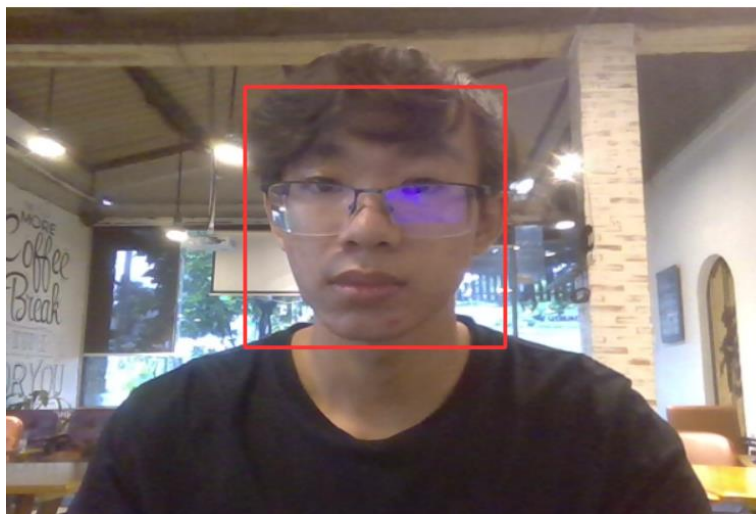
Ma trận nhầm lẫn (confusion matrix), training curve, precision, recall, f1 score, support vector machine được dùng để đánh giá mô hình. Tuy nhiên thì độ chính xác “accuracy” là công cụ đánh giá chuẩn xác và nhanh nhất.



Hình 5. Confusion matrix của model tuổi tác

2.6. Phát hiện khuôn mặt (face detection)

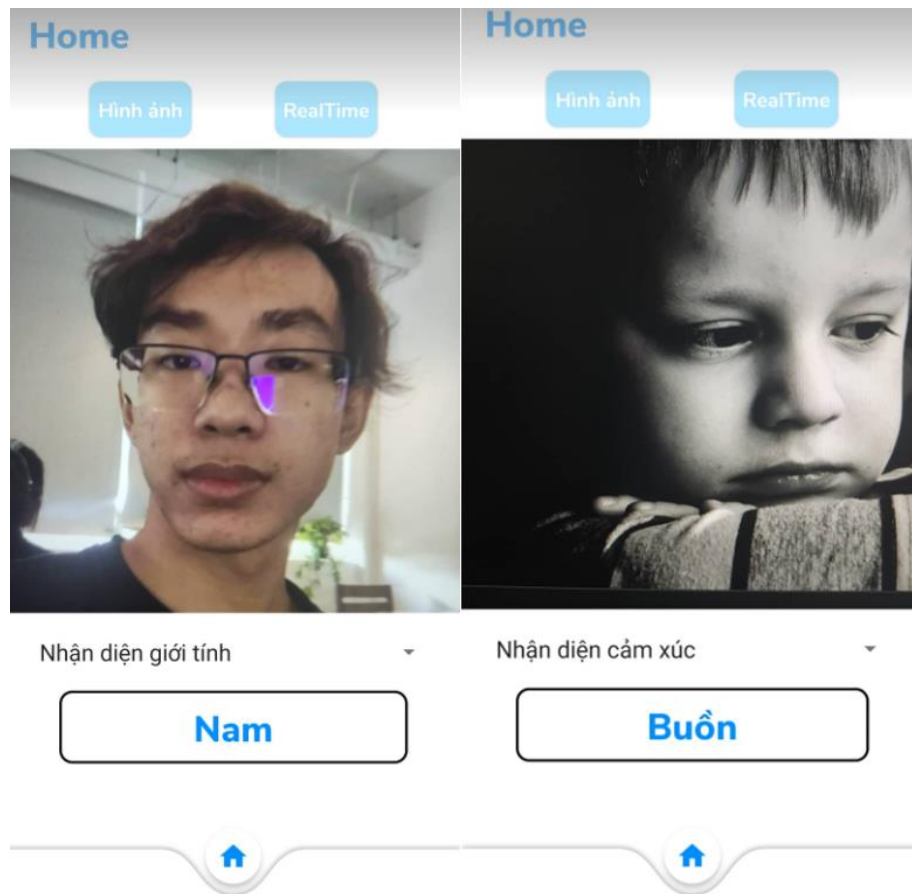
Để trích chọn đặc trưng cho mỗi khuôn mặt, trước tiên ta cần tìm ra vị trí khuôn mặt trong bức hình. Vì bộ dữ liệu sẽ bao gồm nhiều ảnh có điều kiện ánh sáng cũng như các góc độ của khuôn mặt khác nhau, chính vì vậy việc lựa chọn face detector cũng rất quan trọng để đảm bảo hiệu quả cao nhất cho hệ thống. Vì vậy việc sử dụng haarcascade xml để phát hiện khuôn mặt là điều cần thiết.



Hình 6. Phát hiện khuôn mặt với Haar cascade

2.7. Chạy thực nghiệm mô hình trên app android

Thiết kế ứng dụng android bằng Android studio, chuyển đổi model sang file tfLite và triển khai lên phần mềm. Phần mềm sử dụng camera sau của điện thoại để nhận dạng. Có hai chế độ lấy dữ liệu đầu vào là sử dụng camera chạy real time, lấy ảnh bằng cách chụp ảnh hoặc từ thư viện trên điện thoại để dự đoán.



Hình 7. Giao diện trên app

Kết quả

Sau khi huấn luyện, mô hình dự đoán tuổi có độ chính xác là 78%, dự đoán cảm xúc là 50%, dự đoán giới tính là 95%.

Sau khi triển khai, khả năng dự đoán giới tính khá tốt khi đa phần là đúng, khả năng dự đoán tuổi tương đối tốt khi vẫn dự đoán được với tỉ lệ khá cao, còn dự đoán cảm xúc

là chỉ ở mức tạm ổn khi độ chính xác không cao, dự đoán sai tương đối nhiều và cần được cải thiện nhiều trong tương lai.

Kết luận

Sau khi thực hiện đề tài, các kết quả đạt được đã đáp ứng được mục tiêu đề ra. Tuy nhiên, kết quả này vẫn chưa thực sự tốt do còn nhiều điểm thiếu sót khi độ chính xác của mô hình dự đoán cảm xúc chỉ là 50%. Nguyên nhân là do mạng CNN và việc xử lý dữ liệu đầu vào chưa đủ tốt. Điều này dẫn đến app android hoạt động chưa thực sự đáp ứng được yêu cầu khi dự đoán.

Từ những thiếu sót trên sinh viên nhận thấy được đề tài cần được tiếp tục cải thiện nhiều trong tương lai.

Định hướng phát triển cho đề tài: Sàng lọc dataset, thu thập thêm dữ liệu phù hợp, cải thiện mạng CNN, thử các phương pháp xử lý dữ liệu khác, nâng cấp giao diện và chức năng trên ứng dụng để nhiều người có thể tiếp cận.

Tài liệu tham khảo:

1. “Tìm Hiểu Convolutional Neural Networks Cho Phân Loại Ảnh”, Quốc Thịnh,
https://pbcquoc.github.io/cnn/?fbclid=IwAR1zP7AnsFEYCYUfPFiirvV6IcQF4FpLEvmLMHPQzX5pI6X_RD5ICMf78I