



HCMUTE



FME

KHOA CƠ KHÍ CHẾ TẠO MÁY

FACULTY OF MECHANICAL ENGINEERING

PHÁT HIỆN ĐỘ TUỔI VÀ GIỚI TÍNH THÔNG QUA ẢNH VÀ CAMERA REAL-TIME

Tác giả : Nguyễn Phúc Dũng MSSV : 20146486

Thông tin

*Trường đại học Sư phạm Kỹ
thuật thành phố Hồ Chí Minh,
Khoa Cơ khí Chế tạo máy.*

21/05/2022

Tóm tắt

Phát hiện độ tuổi và giới tính là việc xác định tuổi và giới tính của những người trong một hình ảnh hoặc camera. Quá trình này cho phép thu thập và sử dụng dữ liệu cần thiết rất hiệu quả trong các quy trình phát triển kinh doanh. Do đó, mục đích của bài báo này là tự động phát hiện độ tuổi và giới tính của con người bằng cách sử dụng hình ảnh khuôn mặt của con người kết hợp với phương pháp sử dụng mạng nơ-ron tích chập CNN. Hơn 10000 hình ảnh chụp khuôn mặt đã được sử dụng và thông qua các phương pháp xử lý công nghệ, các mô hình đã được đào tạo cho nhiệm vụ phân loại và phát hiện dựa trên cơ sở học chuyển đổi. Trong mô hình công nghệ này, độ chính xác của mô hình đào tạo lên đến 94%, cho ra kết quả khá chính xác so với khuôn mặt thực tế. Hãy cùng xem các công nghệ này hoạt động như thế nào.

Từ khóa

Age and Gender

Face image

Camera Real-time

Webapp

CNN

Trainning Model

Python

Artificial Intelligence

1. Giới thiệu chung

Các công nghệ phân tích khuôn mặt dựa trên trí tuệ nhân tạo đã đạt được tầm quan trọng rất lớn. Yếu tố quan trọng nhất đằng sau sự phổ biến của những công nghệ này là chúng có thể được sử dụng rất hiệu quả trong đời sống kinh doanh. Nhờ những công nghệ này, chẳng hạn như phân khúc khách hàng, phát triển sản phẩm và nhật ký kinh doanh đều trở nên dễ dàng. Các công nghệ trí tuệ nhân tạo được sử dụng rộng rãi nhất dựa trên phân tích khuôn mặt là công nghệ phát hiện tuổi và nhận diện giới tính khuôn mặt. Những công nghệ này cung cấp các giải pháp cửa hàng trực tuyến và cửa hàng thực, phương pháp tiếp thị, cải tiến dịch vụ và phát triển sản phẩm hiệu quả cao.

Đây là những thuật toán rất phức tạp và cực kỳ khó sử dụng. Để có thể thực hiện các tác vụ phức tạp như vậy và nhận dạng chính xác, thuật toán cần được đào tạo với các bộ dữ liệu rất lớn. Ngày nay, các thuật toán Deep Learning được sử dụng rộng rãi cho việc này, nhưng các công ty như Microsoft cũng phát triển phần mềm cho các quy trình này. Mặt khác, các giải pháp trí tuệ nhân tạo dựa trên web như Camera Real-time luôn ở bên bạn để đáp ứng nhu cầu kinh doanh của bạn.

Phát hiện tuổi và phát hiện giới tính AI có nhiều cách sử dụng khác nhau. Ví dụ: tính năng phát hiện độ tuổi và giới tính trong cửa hàng và trước cửa hàng rất hiệu quả ngay cả trong các doanh nghiệp bán lẻ. Bằng cách sử dụng các công nghệ này, doanh nghiệp có thể xác định phân khúc khách hàng, thay đổi hoặc phát triển các chiến lược tiếp thị theo phân khúc này và thay đổi phạm vi sản phẩm của họ.

Phát hiện tuổi và giới tính là điều cần thiết để xác thực, tương tác giữa người và máy tính, phân tích hành vi, đề xuất sản phẩm dựa trên sở thích của người dùng và nhiều lĩnh vực khác. Nhiều công ty cần thu thập dữ liệu về độ tuổi và giới tính, nhưng có rất ít giải pháp. Những phát triển quan trọng đã được thực hiện trong vài năm qua để đáp ứng nhu cầu này. Trong thập kỷ qua, các hệ thống phân loại trí tuệ nhân tạo đã được sử dụng thay cho các hệ thống phân loại thủ công để phát hiện tuổi và giới tính. Với sự ra đời của trí tuệ nhân tạo, tỷ lệ thành công của việc giải quyết vấn đề đã tăng lên.

2. Phương pháp thực hiện và dữ liệu

Trong nghiên cứu này, mục đích là xây dựng được một hệ thống phát hiện giới tính và độ tuổi của con người từ hình ảnh người dùng đưa vào hoặc qua webcam trong thời gian thực sử dụng thuật toán mạng nơ-ron tích chập CNN (Convolutional Neural Network).

2.1. Bộ dữ liệu

Hệ thống cơ sở dữ liệu ở trong bài nghiên cứu này được tạo ra từ các nguồn mở trên internet như google, kaggle, github,... Bộ dữ liệu cho mô hình phát hiện giới tính có khoảng 5645 hình ảnh chụp khuôn mặt của người trên mọi quốc gia trên thế giới. Và bộ dữ liệu cho mô hình phát hiện độ tuổi có khoảng 5080 hình ảnh chụp khuôn mặt của mỗi người trong mỗi khoảng độ tuổi khác nhau. Tất cả đều phục vụ cho mục đích huấn luyện mô hình để mô hình có thể đạt được độ chính xác cao. Ngoài ra, tiến hành tạo thêm một bộ dữ liệu thử nghiệm để trong quá trình huấn luyện mô hình sẽ đưa ra được độ chính xác thực tế so với tập dữ liệu thử nghiệm này. Tiến hành dán nhãn cho từng tập dữ liệu để máy có thể học được giới tính cũng như độ tuổi của nhiều người khác nhau. Bảng 1 cho thấy sự phân bố hình ảnh ở trong từng tệp dữ liệu có nhãn là nam, nữ và các khoảng độ tuổi khác nhau. Ngoài ra, một số hình ảnh trong tệp dữ liệu có nhãn là nam hoặc nữ trong các độ tuổi khác nhau được đưa ra làm ví dụ trong Hình 1. Chất lượng hình ảnh trong tập dữ liệu này thay đổi do các yếu tố khác nhau như chất lượng thiết bị, vị trí, độ sáng,...

Bảng 1. Dữ liệu Giới tính và độ tuổi

	Loại hình	Số lượng ảnh khuôn mặt	Tổng
train	Nam	3508	10726
	Nữ	2137	
	Độ tuổi	5081	
test	Nam	176	447
	Nữ	170	
	Độ tuổi	101	



Hình 1. Hình ảnh đã gắn nhãn từ dữ liệu các khuôn mặt

2.2. Tăng cường dữ liệu (data augmentation)

Hiện nay trong deep learning thì vấn đề dữ liệu có vai trò rất quan trọng. Chính vì vậy có những lĩnh vực có ít dữ liệu để cho việc train model thì rất khó để tạo ra được kết quả tốt trong việc dự đoán. Do đó người ta cần đến một kỹ thuật gọi là tăng cường dữ liệu (data augmentation) để phục vụ cho việc nếu bạn có ít dữ liệu, thì bạn vẫn có thể tạo ra được nhiều dữ liệu hơn dựa trên những dữ liệu bạn đã có.

Để cải thiện khả năng tổng quát hóa của các mô hình CNN, cần có dữ liệu lớn trong giai đoạn đào tạo. Tuy nhiên, thường không có đủ dữ liệu có thể thu được đối với một số vấn đề về thị giác máy tính. Đặc biệt, việc thu thập và ghi nhãn dữ liệu cho các vấn đề về khuôn mặt người dùng để phân tích là một công việc tốn nhiều thời gian và công sức. May mắn thay, các phương pháp khác nhau đã được phát triển để khắc phục vấn đề này. Tăng cường dữ liệu, là một trong những phương pháp nên được áp dụng, cải thiện khả năng tổng quát hóa của mô hình, ngăn chặn ghi đè và tăng độ chính xác của mô hình. Trong nghiên cứu này, phương pháp tăng cường dữ liệu đã được sử dụng trước và trong giai đoạn đào tạo. Thông thường khi nhìn vào bộ dữ liệu trước khi huấn luyện có thể thấy rằng số lượng

hình ảnh dán nhãn giới tính nữ ít hơn khoảng 1371 tấm ảnh so với giới tính nam. Chính vì thế để giảm khoảng cách này hoặc tăng độ chính xác cho cả bộ mô hình, các phương pháp xử lý hình ảnh đã được áp dụng ngẫu nhiên cho cả giới tính nam và nữ. Bằng cách sử dụng các phương thức Flip (lật theo chiều dọc hoặc ngang), random crop, rotation (phạm vi góc + 10, - 10), phóng to (phạm vi, 0,8– 1,2), thu nhỏ, ... thì 1371 tấm ảnh được dán nhãn giới tính nữ ấy đã được tăng cường cùng với tăng cường cả giới tính nam thì độ chính xác của mô hình sẽ được tăng lên rất cao. Hơn nữa, trong khi đào tạo các mô hình đã chọn, tăng dữ liệu thời gian thực (dịch chuyển, xoay, phóng to và thu nhỏ) đã được sử dụng để tránh ghi đè. Hình ảnh sau khi tăng cường dữ liệu được thể hiện ở hình 2:



Hình 2. Dữ liệu giới tính sau khi tăng cường dữ liệu

2.3. Thuật toán mạng tích chập CNN (Convolutional Neural Network)

Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến. Nó giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay.

CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh. Để tìm hiểu tại sao thuật toán này được sử dụng rộng rãi cho việc nhận dạng (detection).

Mạng CNN là một tập hợp các lớp Convolution chồng lên nhau và sử dụng các hàm nonlinear activation như ReLU và tanh để kích hoạt các trọng số trong các node. Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.

Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo. Trong mô hình mạng truyền ngược (feedforward neural network) thì mỗi neural đầu vào (input node) cho mỗi neural đầu ra trong các lớp tiếp theo.

Mô hình này gọi là mạng kết nối đầy đủ (fully connected layer) hay mạng toàn vẹn (affine layer). Còn trong mô hình CNNs thì ngược lại. Các layer liên kết được với nhau thông qua cơ chế convolution.

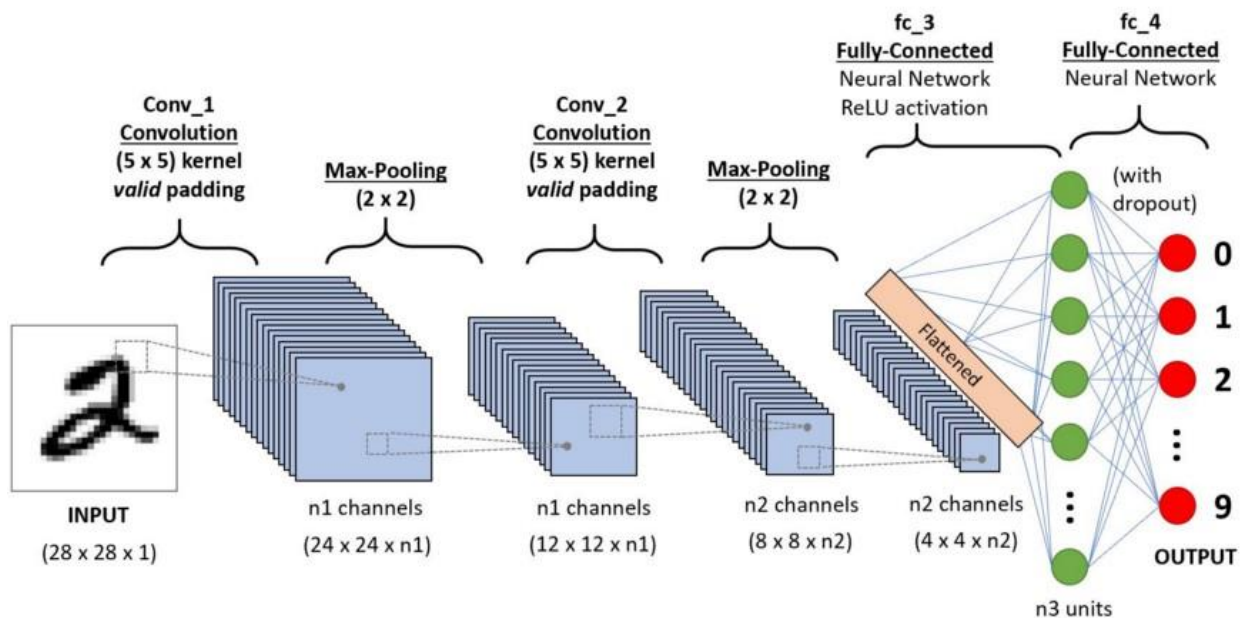
Layer tiếp theo là kết quả convolution từ layer trước đó, nhờ vậy mà ta có được các kết nối cục bộ. Như vậy mỗi neuron ở lớp kế tiếp sinh ra từ kết quả của filter áp đặt lên một vùng ảnh cục bộ của neuron trước đó.

Mỗi một lớp được sử dụng các filter khác nhau thông thường có hàng trăm hàng nghìn filter như vậy và kết hợp kết quả của chúng lại. Ngoài ra có một số layer khác như pooling/subsampling layer dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu).

Trong quá trình huấn luyện mạng (training) CNN tự động học các giá trị qua các lớp filter dựa vào cách thức mà bạn thực hiện. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Layer cuối cùng được dùng để phân lớp ảnh.

Trong mô hình CNN có 2 khía cạnh cần quan tâm là tính bất biến (Location Invariance) và tính kết hợp (Compositionality). Với cùng một đối tượng, nếu đối tượng này được chiếu theo các góc độ khác nhau (translation, rotation, scaling) thì độ chính xác của thuật toán sẽ bị ảnh hưởng đáng kể.

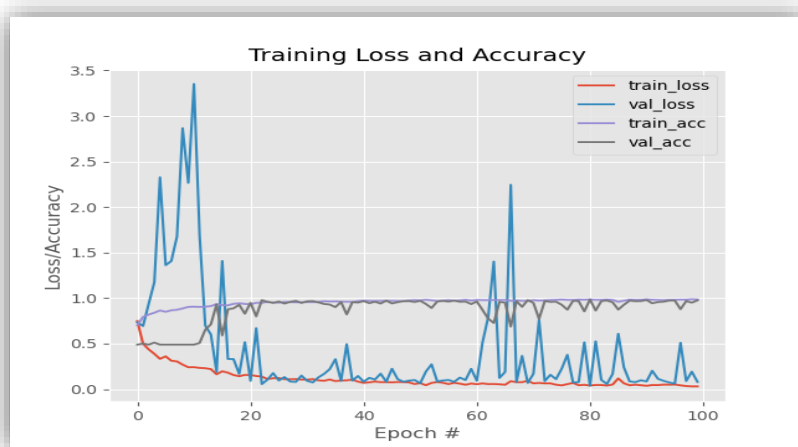
Pooling layer sẽ cho bạn tính bất biến đối với phép dịch chuyển (translation), phép quay (rotation) và phép co giãn (scaling). Tính kết hợp cục bộ cho ta các cấp độ biểu diễn thông tin từ mức độ thấp đến mức độ cao và trừu tượng hơn thông qua convolution từ các filter. Đó là lý do tại sao CNNs cho ra mô hình với độ chính xác rất cao. Cũng giống như cách con người nhận biết các vật thể trong tự nhiên.



Hình 3. Cấu trúc mạng CNN

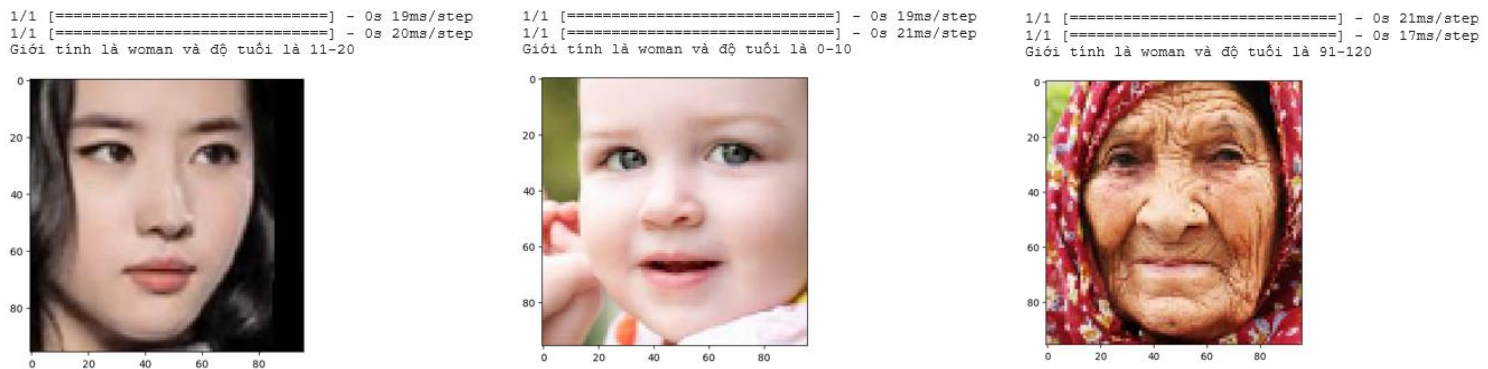
3. Kết quả

Các mô hình áp dụng trong các bài nhận diện hình ảnh sử dụng thuật toán CNN đạt được độ chính xác khá cao (hầu như là trên 94%). Mô hình đã có thể nhận diện được gần như chính xác các bức ảnh đầu vào và sử dụng camera theo thời gian thực phát hiện chính xác. Độ chính xác của mô hình phụ thuộc vào số lượng và chất lượng của dữ liệu đầu vào dùng để train mô hình. Dữ liệu ảnh càng đa dạng (các góc quay khác nhau, độ nghiêng, màu sắc, số lượng ảnh,..) thì độ chính xác của mô hình càng cao. Hình 4 dưới đây biểu thị độ chính xác của mô hình theo các lần học.



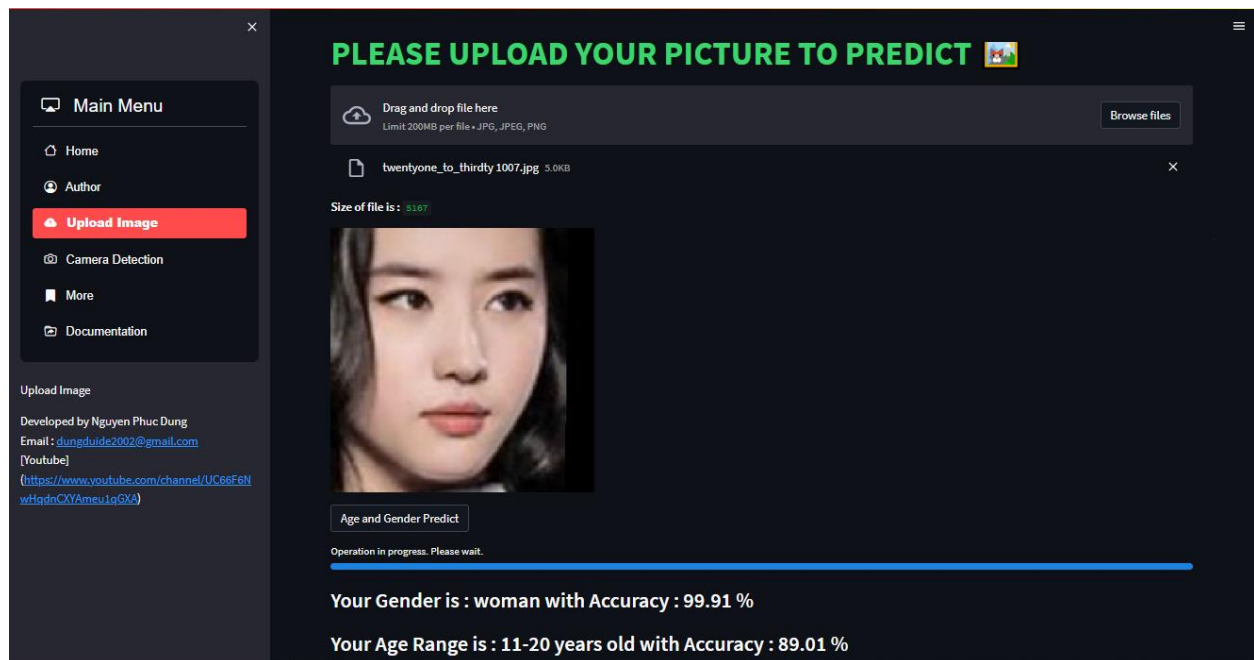
Hình 4. Đồ thị độ chính xác của mô hình và giá trị mất mát theo các lần học

Tiến hành đưa các hình ảnh lên Google Drive, sau khi xử lí bộ dữ liệu sẽ được đưa vào dự đoán trong Google Colab, ta thu được kết quả như hình 5.

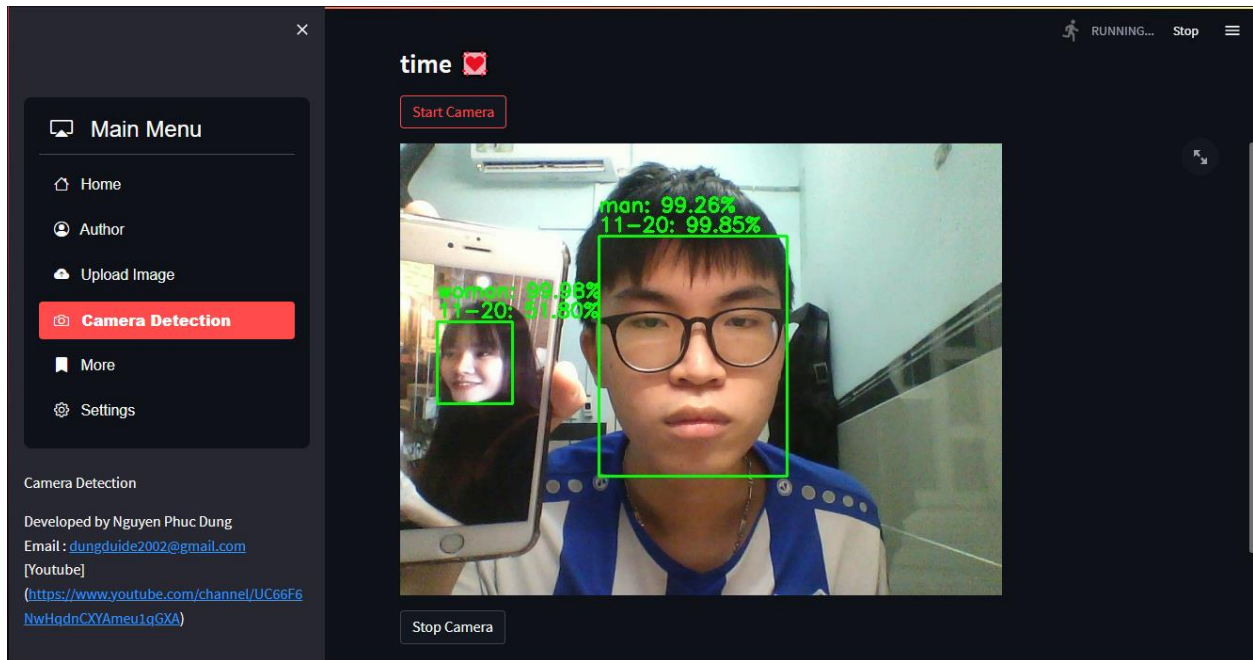


Hình 5. Kết quả dự đoán được trên Google Colab

Để mô hình được phát triển hơn, tiếp cận gần hơn được với nhiều khách hàng người dùng thông qua không gian mạng internet, sử dụng thư viện streamlit để xây dựng webapp với giao diện cho phép tải ảnh lên để phát hiện hoặc cũng có thể sử dụng webcam để phát hiện độ tuổi và giới tính trong thời gian thực như hình 6 và hình 7.



Hình 6. Hình ảnh được tải lên để phát hiện



Hình 7. Quá trình xử lý Camera Real Time trên webapp

Tuy nhiên, dù mô hình có độ chính xác khá cao nhưng không có nghĩa là mô hình sẽ nhận diện đúng được 100% các bức ảnh đầu vào, sẽ vẫn còn những trường hợp nhận diện sai. Lí do là vì chất lượng dữ liệu dùng để train mô hình chất lượng kém, số lượng bị hạn chế chính vì thế nên mô hình vẫn còn những trường hợp sai, không thể hoàn hảo 100% được. Nhưng nhìn chung, kết quả vẫn chấp nhận được.

4. Kết luận

CNN là một phương thức rất hữu hiệu trong việc xử lý các dữ liệu ảnh. Thông thường sử dụng CNN giúp cho việc nhận dạng các đối tượng có trong ảnh có độ chính xác khá cao (khoảng trên 80%, có thể lên tới gần 100%) tùy thuộc vào các dữ liệu đầu vào và cách chúng ta cho máy học theo mô hình CNN như thế nào.

Thế nhưng sử dụng mô hình CNN không đồng nghĩa với việc chắc chắn sẽ nhận dạng đúng. Việc sai sót vẫn là không thể tránh khỏi. Việc này phụ thuộc khá nhiều vào số lượng, chất lượng dữ liệu và kinh nghiệm chọn phương pháp học của người dạy.

Như bạn có thể thấy, các công nghệ phát hiện tuổi và giới tính có thể rất quan trọng đối với doanh nghiệp của bạn. Việc sử dụng những công nghệ này giúp bạn có thể đầu tư và phát triển doanh nghiệp của mình ở nhiều cấp độ khác nhau. Nếu bạn muốn tận dụng khả năng của công nghệ để đầu tư phù hợp vào doanh nghiệp của mình, bạn có thể bắt đầu sử dụng Camera Real-time. Để dễ dàng sử dụng thì mô hình huấn luyện đã được tích hợp sẵn vào webapp để người dùng có thể dễ dàng sử dụng.

Việc sử dụng mô hình CNN sẽ giúp cho việc xử lý trên dữ liệu ảnh tốt hơn chính xác hơn, điều này giúp cho các dự án hoặc các ứng dụng có liên quan đến lĩnh vực này sẽ dễ dàng phát triển hơn.

TÀI LIỆU THAM KHẢO

- [1] <https://www.cameralyze.co/blog/age-gender-detection-top-use-cases>
- [2] <https://topdev.vn/blog/thuat-toan-cnn-convolutional-neural-network/>
- [3] <https://aptech.fpt.edu.vn/cnn-la-gi.html>
- [4] <https://viblo.asia/p/deep-learning-tim-hieu-ve-mang-tich-chap-cnn-maGK73bOKj2>
- [5] <https://www.youtube.com/watch?v=sTNMLLWnG1U>
- [6] <https://www.youtube.com/watch?v=JmvmUWIP2v8&t=638s>
- [7] <https://www.youtube.com/watch?v=Ebb4gUI2IpQ>