

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**UIT**  
**TRƯỜNG ĐẠI HỌC**  
**CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO ĐỒ ÁN**

**MÔN CS231.O12.KHCL – NHẬP MÔN THỊ GIÁC MÁY TÍNH**

**Đề tài: BÀI TOÁN NHẬN DIỆN TUỔI VÀ GIỚI TÍNH CỦA MỘT NGƯỜI**  
**TỪ MỘT BỨC HÌNH CHO TRƯỚC**

**Giảng viên: Thầy Mai Tiến Dũng**

**Sinh viên thực hiện: Trần Thanh Hà**

**MSSV: 21521749**

TP. HỒ CHÍ MINH, Tháng 1 Năm 2024

NHẬN XÉT CỦA GIẢNG VIÊN

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

TPHCM, Ngày ... Tháng ... Năm ...

Người nhận xét

(Ký tên)

## Mục lục

<b>Phần 1: Mở đầu</b>	3
1. Lý do chọn đề tài	3
2. Đối tượng nghiên cứu	4
3. Cơ cấu đồ án	4
<b>Phần 2: Bài toán và nguồn tham khảo</b>	4
1. Input và Output	5
2. Nguồn tham khảo	5
<b>Phần 3: Mô hình và bộ dữ liệu</b>	6
1. Phương pháp	7
2. Mô hình VGG-16	8
3. Bộ dữ liệu – UTK Face	9
<b>Phần 4: Kết quả thực nghiệm và đánh giá</b>	9
1. Thực nghiệm	9
2. Kết quả huấn luyện	10
<b>Kết luận</b>	12

## Phần 1: Mở đầu

### 1. Lý do chọn đề tài

Nhu cầu mua sắm đặc biệt là các dịch vụ mua sắm trên mạng đang phát triển mạnh mẽ trong thời gian gần đây nhờ sự lên ngôi của công nghệ cùng với đó là sự thuận tiện cũng như nhanh chóng mà các dịch vụ đó đem lại. Nhu cầu đưa ra các loại sản phẩm phù hợp với thị hiếu khách hàng càng được đặt lên hàng đầu, việc chỉ dựa đơn giản vào những thông tin cơ bản như tuổi và giới tính từ những bức hình được khách hàng chụp lên đang là một phương thức tốt để giải quyết vấn đề này.

## 2. Đối tượng nghiên cứu

Đề án nghiên cứu việc áp dụng các kỹ thuật thông qua các ứng dụng của thị giác máy tính thông qua đó nhằm trích xuất đặc trưng và dự đoán ra được những kết quả mong muốn.

## 3. Cơ cấu đề án

Nội dung của bản báo cáo bao gồm:

Phần 1: Khái quát chung đề tài.

Phần 2: Giới thiệu về bài toán và nguồn tham khảo.

Phần 3: Giới thiệu về mô hình sử dụng và bộ dữ liệu.

Phần 4: Kết quả thực nghiệm và đánh giá.

## Phần 2: Bài toán và nguồn tham khảo

## 1. Input và Output

**Input:** Một tấm ảnh gồm một khuôn mặt được chụp chính diện.



Hình 2.1: Một ví dụ về Input được lấy trực tiếp trong bộ dữ liệu.

**Output:** Giới tính và tuổi dự đoán từ bức ảnh.

Age: 34 Gender: Male

Hình 2.2: Một ví dụ về Output được lấy từ trong bộ dữ liệu.

## 2. Nguồn tham khảo

Nguồn tham khảo chính của bài toán đến từ bài báo *Gender Recognition Through Face Using Deep Learning*.

Được đăng tại hội nghị ICCIDS 2018.

**Cite:**

*Amit Dhomne, Ranjit Kumar, Vijay Bhan,*

*Gender Recognition Through Face Using Deep Learning,*

*Procedia Computer Science,*

*Volume 132,*

*2018,*

*Pages 2-10,*

*ISSN 1877-0509,*

<https://doi.org/10.1016/j.procs.2018.05.053>.

### **Tóm tắt sơ bộ bài báo:**

Nhận dạng giới tính tự động hiện đã liên quan đến việc mở rộng việc sử dụng trong nhiều phần mềm và phần cứng khác nhau, đặc biệt là do sự phát triển của các trang web mạng xã hội trực tuyến và phương tiện truyền thông xã hội. Tuy nhiên, hiệu suất của hệ thống hiện có với các bức ảnh khuôn mặt thể giới vật lý, hình ảnh có phần không xuất sắc, đặc biệt là khi so sánh với kết quả của nhiệm vụ liên quan đến nhận dạng khuôn mặt. Trong bài báo này, chúng tôi đã khám phá rằng bằng cách thực hiện phương pháp học và phân loại và với việc sử dụng kỹ thuật Mạng nơ-ron tích chập sâu (D-CNN), có thể đạt được mức tăng trưởng hiệu suất đáng kể đối với các nhiệm vụ phân loại giới tính như vậy, đó là lý do tại sao chúng tôi quyết định đề xuất một kiến trúc mạng nơ-ron tích chập hiệu quả VGG-16 có thể được sử dụng trong trường hợp cực đoan khi lượng dữ liệu đào tạo được sử dụng để học D-CNN dựa trên kiến trúc VGG-16 bị hạn chế. Chúng tôi kiểm tra công trình liên quan của mình trên hình ảnh khuôn mặt chưa được lọc hiện tại để nhận dạng giới tính và hiển thị nó để thể hiện các phương pháp cập nhật nâng cao hiện tại.

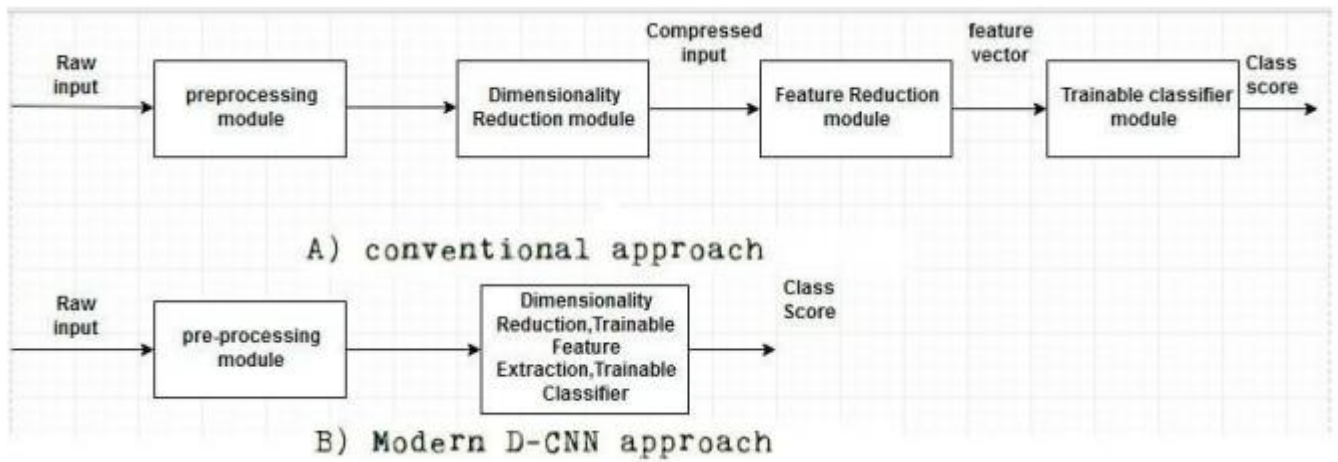
## **Phần 3: Mô hình và bộ dữ liệu**

## 1. Phương pháp

- Trong quá trình huấn luyện mạng nơ-ron, đặc biệt là mạng VGG-16, việc chọn lọc đặc trưng được sử dụng để cho phép mạng học các trọng số chịu trách nhiệm trích xuất đặc trưng.
- Mạng D-CNN có khả năng trích xuất nhiều đặc điểm khác nhau từ một hình ảnh đầu vào chưa qua xử lý, đồng thời yêu cầu rất ít hoặc không cần tiền xử lý.
- Lợi ích chính của việc sử dụng D-CNN là chúng giúp dễ dàng hơn cho các lớp mạng (đầu vào, ẩn và đầu ra) trong việc học các tham số, trọng số và giảm thiểu tổn thất thông qua thuật toán Backpropagation.
- Thêm vào đó mạng D-CNN cho thấy những kết quả xuất sắc trong nhiều ứng dụng như: human tracking system (HTS), surveillance system, traffic signal recognition (TSR) ...

Các hướng tiếp cận mạng CNN:

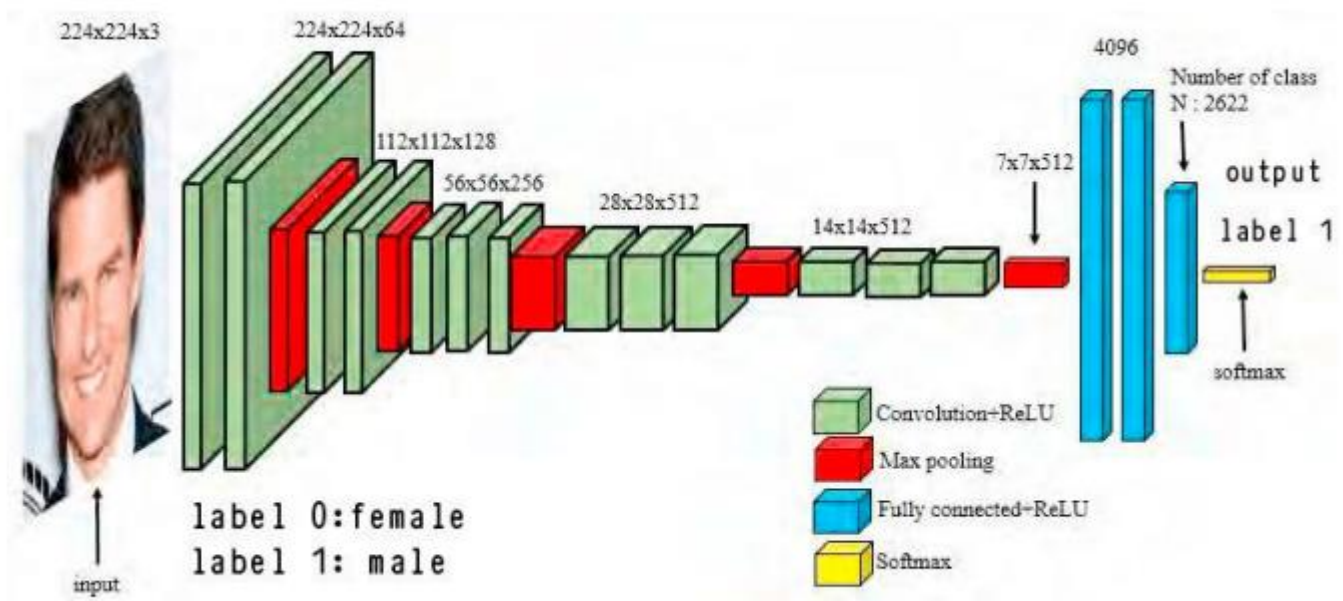
- Bắt đầu từ input thô, đưa ra các khối tiền xử lý, thực hiện giảm chiều dữ liệu sau đó mới đưa qua các khối trích xuất đặc trưng, rồi đưa qua các mô đun huấn luyện để xuất ra điểm của các lớp.
- Tuy nhiên đối với mạng CNN hiện đại, từ input thô, chúng ta chỉ cần đưa qua một khối tiền xử lý duy nhất và đưa vào mạng CNN hiện đại là khối thực hiện chung 3 bước: giảm chiều dữ liệu, trích xuất đặc trưng, huấn luyện rồi đưa ra kết quả.



Hình 3.1: Các cách tiếp cận mạng CNN, trên là các tiếp cận phổ thông, dưới là cách tiếp cận hiện đại.

Trong bài này sử dụng một mạng CNN là VGG-16.

## 2. Mô hình VGG-16



Hình 3.2: Mô phỏng một mô hình VGG-16 kèm với các chú thích trên hình.



VGG16 là mạng convolutional neural network được đề xuất bởi K. Simonyan and A. Zisserman, University of Oxford. Model sau khi train bởi mạng VGG16 đạt độ chính xác 92.7% top-5 test trong dữ liệu ImageNet gồm 14 triệu hình ảnh thuộc 1000 lớp khác nhau. Ở đây mô phỏng lại kiến trúc của VGG-16 bằng các thư viện.

- Input: VGG-16 nhận đầu vào là ảnh 224x224.
- Convolution layers: VGG-16 dùng filter 3x3. Đồng thời cũng dùng 1x1 cho bước biến đổi linear.
- Re-LU activation: Là một hàm linear cung cấp một output khớp với positive input và đưa ra 0 nếu đó là negative input.
- Hidden-layers: Sử dụng ReLU giảm thời gian chạy và tốn ít bộ nhớ.
- Pooling layers: Giảm số chiều và feature map sau các lớp convolution.
- Fully connected layers: VGG-16 dùng 3 lớp fully connected. Hai lớp đầu có 4096 kênh và lớp cuối có 1000 kênh.

### 3. Bộ dữ liệu – UTK Face

UTK Face là một bộ dữ liệu từ kaggle. Bao gồm 20,000 tấm ảnh kèm chú thích cho ba đặc trưng tuổi, giới tính, chủng tộc (Ở đây bỏ qua chủng tộc).

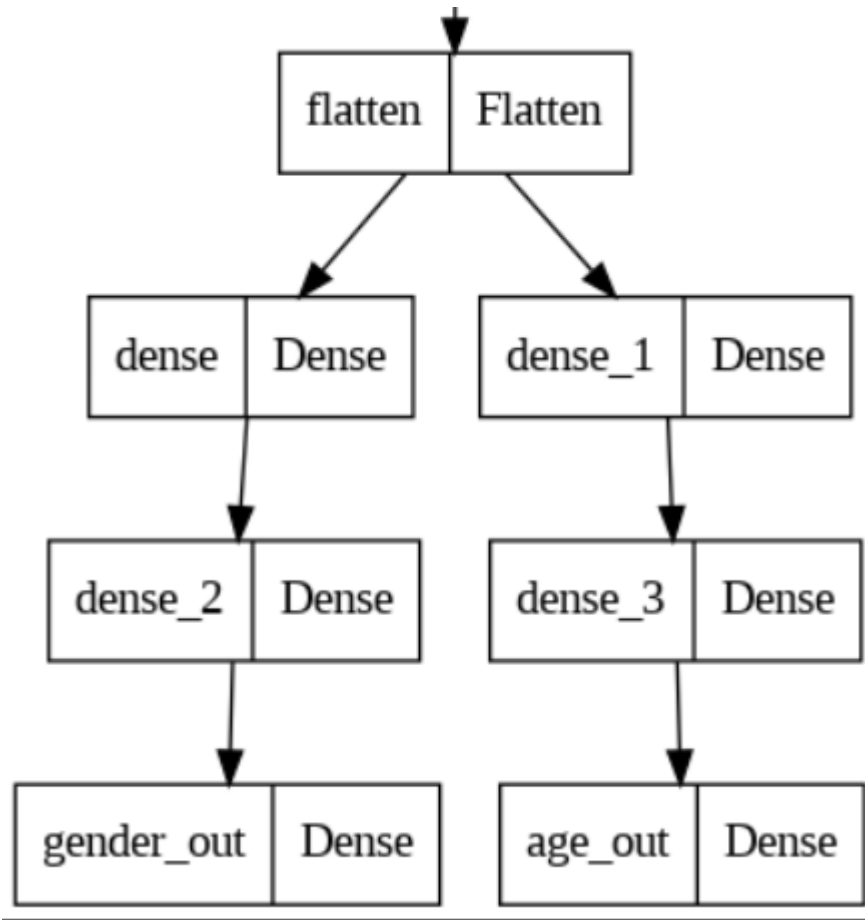
Các tấm ảnh được cắt gọn cẩn thận và chỉ bao gồm một khuôn mặt duy nhất.

Thực nghiệm thực hiện chia bộ dữ liệu thành train – dev – test theo tỉ lệ 7-2-1.

## Phần 4: Kết quả thực nghiệm và đánh giá

### 1. Thực nghiệm

**Điểm khác biệt so với bài báo:** Thực hiện tinh chỉnh mô hình để thực hiện dự đoán thêm đặc trưng tuổi, điều này khác so với bài báo chỉ dự đoán giới tính.



Hình 4.1: Mô phỏng lớp fully connected sau khi đã tinh chỉnh.

Để dự đoán giới tính, sử dụng lớp activation là sigmoid.

Để dự đoán tuổi, sử dụng lớp activation là ReLU.

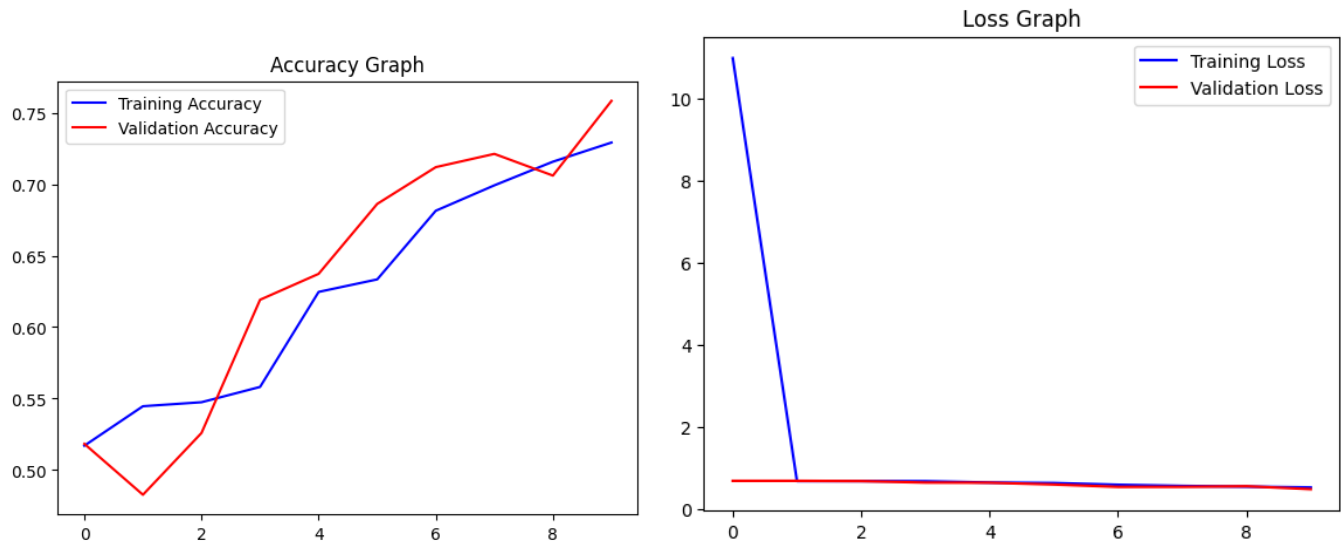
Cuối cùng sử dụng độ đo là accuracy với hàm optimizer là adam.

### Thông số mô hình:

Batch\_size: 32

Epochs = 2

## 2. Kết quả huấn luyện



Hình 4.2 và 4.3: là biểu đồ accuracy và loss của mô hình sau khi chạy.

## Một số hình ảnh dự đoán:



Nhận xét: Dự đoán giới tính khá tốt nhưng tuổi dự đoán còn sai nhiều, điều này có thể do mô hình được tinh chỉnh có độ chính xác chưa tốt, dù vậy nếu sử dụng khoảng tuổi thì có thể chính xác hơn.

## Kết luận

Thông qua bài báo cáo, chúng ta có thể thấy được tầm quan trọng của việc dự đoán giới tính và tuổi cho các doanh nghiệp bán lẻ.

Hiện nay khi các mô hình máy học có rất nhiều và việc dự đoán đã không còn khó khăn nữa, đồng thời với các mô hình hiện đại đạt được độ chính xác rất cao.

Tuy nhiên, trong bài thực nghiệm độ chính xác cho đặc trưng tuổi còn chưa cao dù đã sử dụng mô hình rất tối ưu cho dự đoán là VGG-16. Điều này có thể là do việc tinh chỉnh cá nhân là sai sót.

Nhưng bỏ qua điều đó, việc dự đoán giới tính lại vô cùng chính xác và độ lệch tuổi không quá cao, ở mức chấp nhận được.

## Tài liệu tham khảo:

<https://doi.org/10.1016/j.procs.2018.05.053>.

<https://www.kaggle.com/datasets/jangedoo/utkface-new>.