**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN GIỮA KỲ**

**THỊ GIÁC MÁY TÍNH**

*Người thực hiện*: **ĐẶNG CHÍ VIỄN - 20105361**

**LÊ MINH TOÀN - 20099181**

**VIÊN VĨ NGHIỆP - 20050361**

*Người hướng dẫn*: **TS TRẦN TẤN THÀNH**

*TP. Hồ Chí Minh, tháng 4 năm 2025*

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

A logo with a black background

AI-generated content may be incorrect.

**ĐỒ ÁN GIỮA KỲ**

**THỊ GIÁC MÁY TÍNH**

*Người thực hiện*: **ĐẶNG CHÍ VIỄN - 20105361**

**LÊ MINH TOÀN - 20099181**

**VIÊN VĨ NGHIỆP - 20050361**

*Người hướng dẫn*: **TS TRẦN TẤN THÀNH**

*TP. Hồ Chí Minh, tháng 4 năm 2025*

# LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em xin chân thành cảm ơn đến quý thầy TS.Trần Tấn Thành. Cảm ơn thầy đã tận tình giảng dạy, truyền đạt kiến thức và đồng hành cùng chúng em trong học kì này. Chúc thầy nhiều sức khỏe và thành công trong sự nghiệp giảng dạy.

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi / chúng tôi và được sự hướng dẫn của TS. Trần Tấn Thành. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Công nghiệp TP Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

# PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**TÓM TẮT**

Bài toán "Xác định tuổi và giới tính" là một bài toán trong lĩnh vực thị giác máy tính, nhằm mục đích dự đoán tuổi và giới tính của một người dựa trên ảnh khuôn mặt. Đây là một bài toán có nhiều ứng dụng thực tiễn, bao gồm:

* Hệ thống nhận diện khuôn mặt: Hỗ trợ trong bảo mật, kiểm soát truy cập.
* Phân tích thị trường: Giúp các công ty hiểu rõ hơn về khách hàng của mình.
* Y tế: Hỗ trợ đánh giá tuổi sinh học, theo dõi sức khỏe.

Mục tiêu của báo cáo này là nghiên cứu, thử nghiệm và đánh giá mô hình CNN (Convolutional Neural Network) để giải quyết bài toán này, đồng thời so sánh với mô hình ResNet để tìm ra phương pháp phù hợp nhất.

Tóm tắt về mô hình CNN và ResNet

CNN (Convolutional Neural Network)

* CNN là một mô hình học sâu được thiết kế đặc biệt để xử lý dữ liệu hình ảnh.
* Hoạt động dựa trên các lớp tích chập giúp tự động trích xuất đặc trưng từ ảnh.
* Phù hợp với các bài toán nhận diện khuôn mặt, phân loại ảnh nhờ khả năng học đặc trưng hiệu quả.
* Hạn chế: Khi mạng quá sâu có thể gặp vấn đề mất mát thông tin hoặc vanishing gradient.

ResNet (Residual Network)

* ResNet là một mô hình học sâu với kiến trúc mạng dư (residual network) giúp giải quyết vấn đề suy giảm gradient khi tăng số lớp.
* Sử dụng kết nối tắt (skip connections) giúp duy trì thông tin giữa các lớp, cho phép huấn luyện mạng rất sâu.
* Phù hợp với các bài toán phức tạp đòi hỏi độ chính xác cao.
* Hạn chế: Đòi hỏi tài nguyên tính toán lớn hơn so với CNN truyền thống.

**SỬ DỤNG XỬ LÝ ẢNH ĐỂ XÁC ĐỊNH ĐỘ TUỔI VÀ GIỚI TÍNH**

# **CHƯƠNG 1: GIỚI THIỆU**

## **Tổng quan**

Công nghệ nhận diện khuôn mặt đã trở thành một lĩnh vực trọng tâm trong ngành thị giác máy tính và trí tuệ nhân tạo. Việc trích xuất thông tin như giới tính, độ tuổi và cảm xúc từ hình ảnh khuôn mặt đang ngày càng được ứng dụng rộng rãi trong các hệ thống an ninh, thương mại và tương tác người-máy.

### **1.1.1 Bối cảnh**

Trong lĩnh vực nhận dạng hình ảnh và thị giác máy tính, các mô hình học sâu (Deep Learning) ngày càng được ứng dụng rộng rãi nhằm giải quyết các bài toán phức tạp, trong đó có nhận diện và phân loại đối tượng. Một trong những bài toán quan trọng là xác định tuổi và giới tính dựa trên hình ảnh, được ứng dụng trong nhiều lĩnh vực như giám sát an ninh, chăm sóc khách hàng, y tế và nghiên cứu nhân khẩu học.

Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) và các biến thể hiện đại như ResNet (Residual Network) đã chứng minh hiệu quả vượt trội trong các tác vụ nhận dạng hình ảnh. Với khả năng học các đặc trưng phức tạp từ dữ liệu hình ảnh, ResNet và CNN có thể hỗ trợ việc xác định tuổi và giới tính một cách chính xác hơn so với các phương pháp truyền thống.

Với sự phát triển mạnh mẽ của các mạng nơ-ron tích chập (CNN) và khả năng xử lý thời gian thực của các hệ thống phần mềm, việc xây dựng một hệ thống nhận diện giới tính, độ tuổi và cảm xúc không còn là bài toán quá phức tạp. Dự án này sử dụng mô hình học sâu và OpenCV để phát hiện và phân loại khuôn mặt trong ảnh hoặc video trực tiếp.

### **1.1.2 Lý do chọn đề tài**

Nhận dạng tuổi, giới tính và cảm xúc là một bài toán có nhiều ứng dụng thực tiễn nhưng vẫn còn nhiều thách thức do sự đa dạng về khuôn mặt, ánh sáng, góc chụp và biểu cảm. Việc sử dụng các mô hình học sâu, đặc biệt là ResNet và CNN, sẽ giúp cải thiện độ chính xác và khả năng tổng quát hóa của hệ thống. Hơn nữa, với sự phát triển của công nghệ tính toán và dữ liệu lớn, nghiên cứu này có tiềm năng mang lại các giải pháp hữu ích cho nhiều lĩnh vực khác nhau.

## **1.2 Mục tiêu nghiên cứu**

* Xây dựng mô hình nhận dạng tuổi, giới tính và cảm xúc sử dụng ResNet và CNN.
* Đánh giá độ chính xác của mô hình trên các tập dữ liệu khác nhau.
* So sánh hiệu suất của ResNet với các mô hình CNN thông thường.
* Đề xuất phương pháp tối ưu hóa mô hình để cải thiện độ chính xác.

## **1.3 Phạm vi nghiên cứu**

* Thu thập và xử lý tập dữ liệu hình ảnh khuôn mặt từ các nguồn mở.
* Áp dụng và huấn luyện các mô hình CNN và ResNet để xác định tuổi, giới tính và cảm xúc.
* Đánh giá kết quả thông qua các chỉ số hiệu suất như độ chính xác (accuracy), độ nhạy (recall) và độ đặc hiệu (specificity).
* Thử nghiệm trên các tập dữ liệu khác nhau để kiểm tra tính tổng quát của mô hình.

## **1.4 Ý nghĩa khoa học thực tiễn**

* **Ý nghĩa khoa học:** Góp phần nghiên cứu và phát triển các phương pháp nhận diện hình ảnh tiên tiến dựa trên học sâu. Cung cấp cái nhìn sâu hơn về cách các mô hình ResNet và CNN hoạt động trong bài toán xác định tuổi và giới tính.
* **Ý nghĩa thực tiễn:** Hỗ trợ cải tiến các hệ thống nhận diện khuôn mặt trong giám sát an ninh, chăm sóc khách hàng thông minh, nhận diện nhân khẩu học trong nghiên cứu xã hội và các ứng dụng y tế như đánh giá độ lão hóa khuôn mặt.

# **CHƯƠNG 2: CƠ SỞ LÝ THUYẾT**

## **2.1. Mạng nơ-ron tích chập (CNN)**

### **2.1.1. Khái niệm**

Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) là một mô hình học sâu (Deep Learning) chuyên xử lý dữ liệu có cấu trúc lưới, đặc biệt là hình ảnh. CNN hoạt động bằng cách sử dụng các bộ lọc (filters/kernels) để tự động trích xuất đặc trưng từ hình ảnh, giúp giảm thiểu việc cần thiết kế đặc trưng thủ công như trong các mô hình truyền thống.

Mạng CNN bao gồm nhiều lớp chính như:

* Lớp tích chập (Convolutional Layer): Trích xuất đặc trưng từ hình ảnh.
* Lớp phi tuyến (Activation Function, thường là ReLU): Tăng tính phi tuyến tính của mô hình.
* Lớp giảm kích thước (Pooling Layer): Giảm số lượng tham số và tính toán.
* Lớp kết nối toàn phần (Fully Connected Layer - FC): Thực hiện phân loại hoặc dự đoán.

### **Ứng dụng của CNN**

CNN được ứng dụng rộng rãi trong nhiều lĩnh vực, đặc biệt là thị giác máy tính (Computer Vision). Dưới đây là một số ứng dụng quan trọng:

- Nhận diện và phân loại hình ảnh:

* CNN giúp nhận diện các đối tượng trong ảnh, chẳng hạn như phân loại mèo và chó.
* Ứng dụng trong **Google Photos**, **Facebook**, **Instagram** để gợi ý gắn thẻ bạn bè.

- **Nhận diện khuôn mặt**

* CNN được sử dụng trong **FaceID của Apple**, **hệ thống giám sát**, và **bảo mật sinh trắc học**.

- **Xe tự hành**

* CNN giúp xe tự hành nhận diện biển báo, người đi bộ, vật cản trên đường.
* Các công ty như **Tesla, Waymo** sử dụng CNN trong hệ thống lái tự động.

- **Y tế – Chẩn đoán bệnh**

* Phân tích ảnh y khoa như **X-ray, MRI, CT scan** để phát hiện ung thư, bệnh tim,...
* Ví dụ: **AI của Google Health** có thể phát hiện bệnh võng mạc tiểu đường.

- **Nhận diện chữ viết tay & OCR**

* Dùng trong **Google Lens, CamScanner** để chuyển ảnh thành văn bản số.
* Ứng dụng trong tự động đọc số serial trên hóa đơn, chứng từ.

### **Ưu điểm và nhược điểm**

- Ưu điểm:

* Tự động trích xuất đặc trưng:
* Không cần thiết kế đặc trưng thủ công như trong các phương pháp truyền thống (SIFT, HOG,...) mà CNN có thể học đặc trưng trực tiếp từ dữ liệu.
* Điều này giúp tiết kiệm thời gian và tăng hiệu quả trong nhận diện hình ảnh.
* Tối ưu hóa tính toán với Pooling Layer:
* Lớp Pooling giúp giảm kích thước dữ liệu đầu vào mà vẫn giữ được thông tin quan trọng.
* Giảm số lượng tham số, tăng tốc độ xử lý và giảm nguy cơ overfitting.
* Khả năng tái sử dụng đặc trưng (Feature Reusability):
* Các bộ lọc (filters) học được có thể tái sử dụng ở nhiều bài toán khác nhau (ví dụ: các cạnh, góc, đường nét).
* Dùng mô hình đã huấn luyện trước (Pre-trained models như VGG, ResNet) giúp giảm công sức huấn luyện từ đầu.
* Hiệu suất cao trong xử lý hình ảnh:
* CNN vượt trội so với mạng nơ-ron truyền thống (MLP) trong xử lý dữ liệu hình ảnh.
* Đạt độ chính xác cao trong nhận diện hình ảnh, phân loại, phát hiện vật thể,...
* Giảm số lượng tham số so với MLP:
* Mạng MLP yêu cầu số lượng kết nối rất lớn nếu xử lý ảnh (mỗi pixel là một neuron).
* CNN sử dụng tích chập (convolution) giúp giảm số lượng tham số đáng kể mà vẫn duy trì hiệu suất cao.

- Nhược điểm:

* Cần lượng dữ liệu lớn để huấn luyện
* CNN yêu cầu hàng ngàn, thậm chí hàng triệu hình ảnh để hoạt động hiệu quả.
* Nếu dữ liệu quá ít, mô hình dễ bị **overfitting**.
* Yêu cầu tài nguyên tính toán cao
* Huấn luyện CNN cần **GPU mạnh** hoặc **TPU** để xử lý nhanh chóng.
* Với các mô hình lớn như **ResNet-152, EfficientNet**, cần nhiều bộ nhớ và thời gian huấn luyện lâu.
* Không hiệu quả với dữ liệu không gian thấp (Low-dimensional data)
* CNN rất tốt với ảnh/video nhưng không phù hợp với dữ liệu dạng bảng (tabular data) như trong tài chính, y tế (có thể dùng MLP hoặc XGBoost).
* Nhạy cảm với nhiễu và dữ liệu lệch
* Nếu ảnh đầu vào bị **nhiễu**, **mờ**, hoặc có quá nhiều thay đổi, CNN có thể hoạt động kém hơn.
* Nếu dữ liệu huấn luyện không cân bằng (ví dụ: quá nhiều ảnh mèo, ít ảnh chó), mô hình có thể thiên vị.
* Không giải thích được quyết định
* CNN là một "hộp đen" (**black box**), khó giải thích vì sao mô hình đưa ra kết quả.
* Điều này gây khó khăn khi áp dụng CNN vào lĩnh vực cần giải thích rõ ràng như y tế.

### **Các lỗi thường gặp khi sử dụng CNN**

* Overfitting
* Khi mô hình học quá kỹ từ dữ liệu huấn luyện, nó có thể hoạt động kém trên dữ liệu mới.
* Cách khắc phục:
  + Thêm Dropout (ví dụ: Dropout 0.5).
  + Sử dụng Data Augmentation (xoay ảnh, làm mờ,...).
  + Thu thập thêm dữ liệu hoặc dùng mô hình pre-trained.
* Vanishing Gradient (Độ dốc biến mất)
* Khi sử dụng mạng quá sâu, gradient có thể trở nên quá nhỏ, làm chậm quá trình học.
* Cách khắc phục:
  + Dùng Batch Normalization.
  + Sử dụng ReLU thay vì sigmoid/tanh.
  + Áp dụng Residual Network (ResNet) để giúp gradient lan truyền tốt hơn.
* Dữ liệu không cân bằng (Imbalanced Data)
* Nếu tập huấn luyện có quá nhiều hình ảnh của một lớp và quá ít hình ảnh của lớp khác, CNN có thể đưa ra dự đoán thiên vị.
* Cách khắc phục:
  + Dùng Weighted Loss Function để cân bằng dữ liệu.
  + Data Augmentation để tạo thêm dữ liệu từ lớp ít dữ liệu hơn.
* Thời gian huấn luyện lâu
* Nếu không tối ưu mô hình, việc huấn luyện có thể mất hàng giờ đến hàng ngày.
* Cách khắc phục:
  + Dùng Transfer Learning với mô hình có sẵn.
  + Giảm kích thước ảnh đầu vào.
  + Sử dụng mô hình nhẹ hơn như MobileNet, EfficientNet thay vì ResNet/VGG.
* Kiểm tra kém trên dữ liệu thực tế
* Nếu mô hình hoạt động tốt trên tập huấn luyện nhưng kém trên dữ liệu thực tế, có thể do leakage (rò rỉ dữ liệu) hoặc dữ liệu huấn luyện không đại diện tốt.
* Cách khắc phục:
  + Kiểm tra lại dữ liệu huấn luyện.
  + Dùng cross-validation để đảm bảo tính tổng quát.

## **2.2.** **Mạng ResNet (Residual Network)**

## **2.2.1 Khái niệm**

ResNet (Residual Network) là một kiến trúc mạng nơ-ron tích chập (CNN) sâu, được giới thiệu bởi Microsoft Research vào năm 2015 trong bài báo "Deep Residual Learning for Image Recognition".

* Mục đích chính của ResNet: Giải quyết vấn đề vanishing gradient khi huấn luyện các mạng nơ-ron rất sâu.
* Cách hoạt động: Sử dụng Residual Blocks với Skip Connections (kết nối tắt) giúp thông tin có thể truyền qua nhiều lớp mà không bị mất mát hoặc suy giảm.
* Kết quả: ResNet có thể đạt đến hàng trăm lớp mà vẫn huấn luyện hiệu quả, trong khi các mô hình trước đây chỉ giới hạn khoảng 20–30 lớp.

### **2.2.2 Ứng dụng của ResNet**

ResNet được sử dụng rộng rãi trong các tác vụ liên quan đến thị giác máy tính, đặc biệt là trong nhận diện hình ảnh và học sâu.

* Phân loại hình ảnh (Image Classification)
* ResNet giúp tăng độ chính xác trong các bài toán phân loại hình ảnh.
* Đã đạt thành tích xuất sắc tại cuộc thi ImageNet 2015, vượt qua các mô hình như VGG.
* Phát hiện vật thể (Object Detection)
* ResNet được sử dụng làm backbone trong các mô hình như Faster R-CNN, YOLO, Mask R-CNN để phát hiện vật thể trong ảnh.
* Nhận diện khuôn mặt
* Ứng dụng trong bảo mật, FaceID của Apple, hệ thống giám sát,...
* Thị giác máy tính trong y tế
* Chẩn đoán bệnh qua ảnh X-ray, CT, MRI,...
* Ví dụ: Phát hiện khối u ung thư bằng Deep Learning.
* Ứng dụng trong xe tự hành
* Dùng trong phân tích hình ảnh camera để nhận diện biển báo, vạch đường, người đi bộ,...
* Xử lý ảnh vệ tinh
* Ứng dụng trong nông nghiệp, quân sự, quan sát trái đất, dự báo thiên tai.

### **2.2.3 Ưu điểm và nhược điểm**

- Ưu điểm:

Giải quyết vấn đề vanishing gradient

* Nhờ Skip Connections, ResNet giúp truyền thông tin qua nhiều lớp mà không bị mất mát gradient, giúp mạng học tốt hơn.

Huấn luyện mạng cực sâu dễ dàng

* Các mô hình như ResNet-50, ResNet-101, ResNet-152 có hàng chục đến hàng trăm lớp nhưng vẫn có thể huấn luyện hiệu quả.

Cải thiện độ chính xác của mô hình

* Nhờ khả năng học tốt các đặc trưng phức tạp, ResNet đạt kết quả vượt trội so với các mô hình như VGG trong phân loại hình ảnh.

Hiệu quả tính toán tốt hơn so với VGG

* ResNet có số lượng tham số ít hơn so với VGG nhưng đạt độ chính xác cao hơn.

Dễ dàng tích hợp với các mô hình khác

* ResNet được sử dụng làm backbone cho nhiều mô hình deep learning khác.

- Nhược điểm:

Đòi hỏi tài nguyên tính toán cao

* Mạng ResNet sâu có thể cần GPU mạnh (NVIDIA RTX, A100) để huấn luyện nhanh hơn.

Skip Connection không phải lúc nào cũng hiệu quả

* Trong một số trường hợp, nếu số lớp quá lớn, Skip Connection có thể không giúp cải thiện độ chính xác.

Độ trễ tính toán cao hơn so với các mô hình nhẹ hơn

* Các mô hình như MobileNet, EfficientNet có thể chạy nhanh hơn trên thiết bị di động hoặc nhúng.

Dễ bị quá tải bộ nhớ (Memory Bottleneck)

* Khi huấn luyện ResNet với hình ảnh lớn và batch size cao, có thể gặp lỗi thiếu bộ nhớ GPU.

### **2.2.4 Các lỗi thường gặp**

Vanishing Gradient:

* Nếu thiết kế ResNet không hợp lý hoặc batch size quá nhỏ, mô hình vẫn có thể gặp vanishing gradient.
* Cách khắc phục:
  + Dùng Batch Normalization để ổn định quá trình huấn luyện.
  + Kiểm tra learning rate để tránh bị mắc kẹt trong local minima.

Overfitting:

* Khi ResNet quá sâu và dữ liệu không đủ lớn, mô hình có thể học quá mức từ tập huấn luyện và hoạt động kém trên dữ liệu thực tế.
* Cách khắc phục:
  + Sử dụng Dropout.
  + Dùng Data Augmentation để tạo thêm dữ liệu.
  + Dùng Early Stopping để tránh huấn luyện quá lâu.

Quá tải bộ nhớ GPU (Out of Memory - OOM Error):

* Khi batch size quá lớn hoặc mô hình ResNet quá sâu, có thể gặp lỗi tràn bộ nhớ GPU.
* Cách khắc phục:
  + Giảm batch size (ví dụ từ 64 xuống 32).
  + Sử dụng Gradient Accumulation để giảm tải bộ nhớ GPU.
  + Dùng các phiên bản nhẹ hơn như ResNet-18, ResNet-34 thay vì ResNet-152.

Mô hình hội tụ chậm (Slow Convergence):

* Khi dùng ResNet rất sâu, quá trình hội tụ có thể bị chậm, làm tăng thời gian huấn luyện.
* Cách khắc phục:
  + Dùng learning rate scheduler (ReduceLROnPlateau, CosineAnnealing).
  + Dùng Optimizer mạnh như AdamW hoặc LARS để tăng tốc độ học.

Không cải thiện kết quả dù tăng số lớp:

* Trong một số trường hợp, tăng số lớp không cải thiện độ chính xác mà còn gây giảm hiệu suất do overfitting hoặc vanishing gradient.
* Cách khắc phục:
  + Kiểm tra lại số lớp hợp lý (ví dụ: ResNet-50 có thể tốt hơn ResNet-152 cho một số bài toán).
  + Sử dụng Transfer Learning thay vì huấn luyện từ đầu.

## **2.3 Các chỉ số đánh giá mô hình**

* Độ chính xác (Accuracy): Tỷ lệ dự đoán đúng trên tổng số mẫu thử.
* Độ nhạy (Recall): Khả năng mô hình nhận diện đúng các mẫu thuộc một lớp nhất định.
* Độ đặc hiệu (Specificity): Khả năng mô hình phân biệt chính xác giữa các lớp.
* F1-score: Trung bình điều hòa giữa độ chính xác và độ nhạy, giúp đánh giá mô hình toàn diện hơn.

## **2.4 So sánh giữa mô hình ResNet và CNN**

|  |  |  |
| --- | --- | --- |
| **Tiêu chí** | **CNN** | **ResNet** |
| Độ sâu mạng | Tương đối nông, thường gồm một số lớp tích chập và pooling | Rất sâu, sử dụng hàng chục hoặc hàng trăm lớp |
| **Khả năng xử lý đặc trưng** | Học các đặc trưng cơ bản nhưng có thể mất thông tin nếu quá nhiều lớp | Giữ lại thông tin tốt hơn nhờ kết nối tắt |
| **Tốc độ huấn luyện** | Nhanh hơn, yêu cầu tài nguyên tính toán ít hơn | Chậm hơn do kiến trúc phức tạp |
| **Yêu cầu tài nguyên** | Ít hơn, phù hợp với phần cứng hạn chế | Cao hơn, cần GPU mạnh để huấn luyện |
| **Ứng dụng phù hợp** | Các bài toán đơn giản hoặc dữ liệu không quá lớn | Các bài toán phức tạp, dữ liệu lớn |

# **CHƯƠNG 3: THỰC NGHIỆM**

## **3.1 Tập dữ liệu:**

Tập dữ liệu UTKFace trên Kaggle được lựa chọn do những ưu điểm sau:

* Số lượng mẫu lớn: Gồm hơn 23.000 ảnh khuôn mặt với nhãn tuổi, giới tính, và cảm xúc.
* Dữ liệu đa dạng: Phạm vi tuổi từ 0 đến 116 tuổi, bao gồm nhiều dân tộc và giới tính.
* Dữ liệu sạch: Hình ảnh đã được tiền xử lý, căn chỉnh khuôn mặt, giúp giảm bớt công đoạn xử lý thủ công.
* Dễ tiếp cận: Dữ liệu sẵn có trên Kaggle, dễ dàng tải về và sử dụng.

Link của data: (https://www.kaggle.com/code/uzzivirus/age-and-gender-detection/input)

## **3.2. Quá trình huần luyện:**

- Tiền xử lý dũ liệu:

Quá trình tiền xử lý dữ liệu bao gồm các bước sau:

1. Chuẩn hóa kích thước ảnh: Chuyển tất cả ảnh về kích thước 128x128 pixels để đảm bảo đồng nhất.
2. Chuyển đổi dữ liệu thành tensor: Mỗi ảnh được chuyển thành tensor có dạng (128,128,1).
3. Chuẩn hóa giá trị pixel: Chia tất cả giá trị pixel cho 255 để đưa về khoảng [0,1], giúp tăng tốc độ hội tụ khi huấn luyện.
4. Xem xét phân bố dữ liệu: Kiểm tra xem tập dữ liệu có cân bằng giữa các độ tuổi, giới tính và cảm xúc hay không.

Phân bố tập dữ liệu

* Giới tính: Tập dữ liệu có số lượng nam và nữ khá cân bằng.
* Độ tuổi: Số lượng mẫu ảnh giảm dần ở các nhóm tuổi cao.

- Chia nhỏ dữ liệu để train.

## **3.3. Kết quả:**

- Chuẩn bị:

Bước 1: Sử dụng dataframe của pandas để lưu trữ path của ảnh để train

A screenshot of a computer

AI-generated content may be incorrect.

Bước 2 : Kiểm tra ảnh đã được load trước khi sử dụng

A computer screen with white text

AI-generated content may be incorrect.

Bước 3 : Vì tập dữ liệu quá lớn nên chia nhở để train tránh dẫn đến overfitting đồng thời / thời gian train model quá lâu , đồng thời xáo trộn ngẫu nhiên tập ảnh đầu vào tránh bị trùng lập làm cho model không chính xác

A screenshot of a computer

AI-generated content may be incorrect.

- Xây dựng mô hình:

Khởi tạo mô hình CNN để train model với tập dữ liệu đầu vào đã được xử lý:



Kết quả mô hình xây dựng thu được và train model:

A screenshot of a computer program

AI-generated content may be incorrect.

A black and white screen

AI-generated content may be incorrect.

Thống kê giá trị mất mát (loss) và MAE (Mean Absolute Error):

A computer screen shot of a program code

AI-generated content may be incorrect.

- Thử nghiệm mô hình:

A screenshot of a computer

AI-generated content may be incorrect.

## **3.4. Đánh giá:**

- Đánh giá về mô hình

* Mô hình có kiến trúc CNN với ba lớp tích chập (Conv2D) kết hợp Batch Normalization và MaxPooling2D để trích xuất đặc trưng từ ảnh khuôn mặt.
* Sau khi Flatten dữ liệu, mô hình sử dụng các lớp Dense với kích thước giảm dần để phân loại giới tính và dự đoán tuổi.
* Có hai đầu ra:
  + gender\_output (giới tính): Một node đầu ra với sigmoid activation, phù hợp cho bài toán phân loại nhị phân.
  + age\_output (tuổi): Một node đầu ra, có thể sử dụng linear activation cho bài toán hồi quy.
* Tổng số tham số: 11 triệu, mức trung bình đối với mô hình nhận diện ảnh.

- Đánh giá kết quả huấn luyện

* Độ chính xác phân loại giới tính:
* Bắt đầu từ khoảng 62% (Epoch 1) và tăng lên ~95% (Epoch 20).
* Mô hình phân loại giới tính khá tốt.
* Lỗi dự đoán tuổi (MAE - Mean Absolute Error):
* Bắt đầu từ 17.7 (Epoch 1) và giảm xuống ~5.3 (Epoch 20).
* Lỗi trung bình 5.3 tuổi là khá ổn, nhưng có thể cải thiện thêm.
* Mất mát (Loss):
* Giảm dần theo thời gian, cho thấy mô hình hội tụ tốt.

**3.5 Giao Diện**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a person

AI-generated content may be incorrect.**

## **3.6. Hướng phát triển trong tương lai:**

Một trong những mục tiêu chính trong tương lai là nâng cao độ chính xác của mô hình trong việc xác định tuổi và giới tính. Điều này có thể đạt được thông qua việc tăng cường tập dữ liệu, sử dụng các phương pháp học sâu mới như mạng nơ-ron đối kháng (GANs) để tạo ra dữ liệu tổng hợp, hoặc cải tiến các kiến trúc mạng như Transformer và EfficientNet.

Mặc dù mô hình hiện tại đã đạt được kết quả khả quan, nhưng việc áp dụng vào các tình huống thực tế như nhận diện trong điều kiện ánh sáng yếu, góc nhìn khác nhau hoặc khuôn mặt bị che khuất vẫn còn nhiều thách thức. Cần phát triển các phương pháp xử lý dữ liệu tiên tiến hơn để tăng cường khả năng nhận diện trong các tình huống này.

Ngoài các ứng dụng truyền thống như giám sát an ninh và chăm sóc khách hàng, mô hình có thể được áp dụng trong các lĩnh vực khác như phân tích cảm xúc, dự đoán sức khỏe tâm thần từ hình ảnh khuôn mặt, hoặc sử dụng trong các ứng dụng y tế để chẩn đoán bệnh lý thông qua đặc điểm khuôn mặt.