ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC MÁY TÍNH

Nguyễn Tấn Tài - 23521376

Phương pháp học chuyển giao để phân loại nhóm tuổi từ ảnh chân dung

A Transfer Learning Approach for Age Group Classification from Portrait Images

CỬ NHÂN NGÀNH KHOA HỌC MÁY TÍNH

TP. HỒ CHÍ MINH, 2025

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC MÁY TÍNH

Nguyễn Tấn Tài - 23521376

Phương pháp học chuyển giao để phân loại nhóm tuổi từ ảnh chân dung

A Transfer Learning Approach for Age Group Classification from Portrait Images

CỬ NHÂN NGÀNH KHOA HỌC MÁY TÍNH

GIẢNG VIÊN HƯỚNG DẪN
TS. MAI TIẾN DŨNG

TP. HỒ CHÍ MINH, 2025

Lời cảm ơn

Để hoàn thành khóa luận này, nhóm sinh viên chúng em may mắn nhận được sự hỗ trợ, động viên và hướng dẫn tận tình từ quý thầi.

TP. Hồ Chí Minh, ngày 01 tháng 01 năm 2026.

Nguyễn Tấn Tài

Khoa Khoa Học Máy Tính, Lớp CS420.Q12

MỤC LỤC

Danh sách hình vẽ

Danh sách bảng

Thông tin các thành viên và đánh giá mức độ hoàn thành

Họ và tên: Nguyễn Tấn Tài
Mã số sinh viên: 23521376
Nhiệm vụ đảm nhận:

- Thu thập và tiền xử lý dữ liệu ảnh khuôn mặt
- Xây dựng và huấn luyện các mô hình học sâu (CNN, Vision Transformer)
- Thực hiện thí nghiệm, đánh giá và phân tích kết quả
- Chuẩn bị bài thuyết trình trên lớp
- Viết và hoàn thiện báo cáo khóa luận

Mức độ hoàn thành: 100%

Chương 1

TỔNG QUAN

Trong chương này, tác giả sẽ giới thiệu về đề tài, trình bày lý do chọn đề tài, mục tiêu nghiên cứu, đối tượng và phạm vi nghiên cứu.

1.1 Đặt vấn đề

1.1.1 Giới thiệu đề tài

Trong kỷ nguyên số hiện nay, việc phân tích và hiểu các thuộc tính của con người từ dữ liệu hình ảnh đang trở thành một lĩnh vực nghiên cứu sôi động với nhiều ứng dụng thực tiễn. Trong số đó, nhận dạng tuổi từ ảnh khuôn mặt nổi lên như một công nghệ nền tảng quan trọng. Khóa luận này tập trung vào bài toán **"Nhận dạng nhóm tuổi từ ảnh chân dung"** – một bài toán con của nhận dạng tuổi, với mục tiêu phân loại một khuôn mặt vào một nhóm tuổi được định nghĩa trước (ví dụ: Trẻ em, Thanh niên, Trung niên, Người cao tuổi). Sự phát triển mạnh mẽ của các phương pháp Học sâu (Deep Learning) trong những năm gần đây đã mở ra những hướng tiếp cận mới, mang lại độ chính xác cao và khả năng ứng dụng rộng rãi cho bài toán này.

1.1.2 Mô Tả Bài Toán

Bài toán nhận dạng nhóm tuổi từ ảnh chân dung được mô tả một cách hình thức như sau:

- **Đầu vào:**

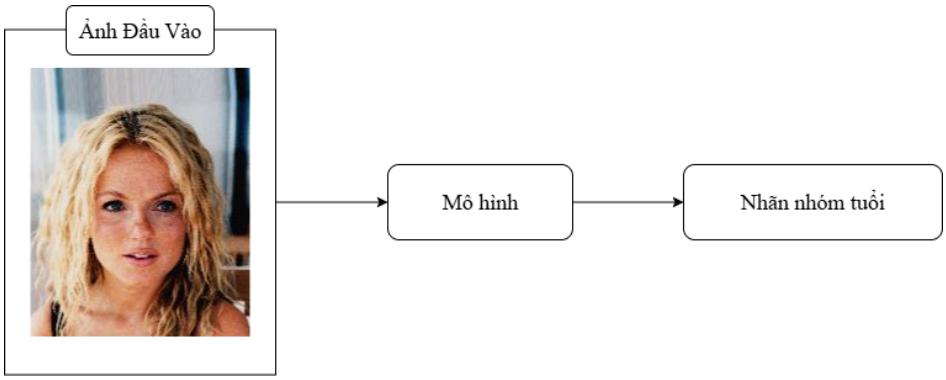
  - **Giai đoạn huấn luyện:** Một tập dữ liệu huấn luyện
  - $D_{train} = \{(I_1, y_1), (I_2, y_2), \ldots, (I_N, y_N)\}$, trong đó mỗi cặp (I_i, y_i) bao gồm:
    * Ảnh đầu vào I_i: ảnh màu hoặc xám có kích thước $W \times H \times C$ chứa khuôn mặt người đã được phát hiện và căn chỉnh.
    * Nhãn gốc y_i: nhãn nhóm tuổi tương ứng, với $y_i \in \{G_1, G_2, \ldots, G_C\}$.
  - **Giai đoạn dự đoán:** Một ảnh đầu vào I chưa biết nhãn.

- **Đầu ra:** Một nhãn y dự đoán, trong đó $y \in \{G_1, G_2, \ldots, G_C\}$. Mỗi G_i đại diện cho một nhóm tuổi cụ thể được định nghĩa trước.

- **Mục tiêu:** Cần xây dựng một hàm ánh xạ F (mô hình học sâu) sao cho:

$$F(I) \to y \tag{1.1}$$

với y là nhãn nhóm tuổi được dự đoán cho ảnh đầu vào I.
Minh họa: Quy trình của bài toán có thể được minh họa qua sơ đồ sau:



Hình 1.1: Sơ đồ minh họa bài toán nhận dạng nhóm tuổi từ ảnh chân dung

Ví dụ về phân loại nhóm tuổi:

- **Nhóm 1 (G_1):** Trẻ em (1 - 12 tuổi)
- **Nhóm 2 (G_2):** Thiếu niên (13 - 19 tuổi)
- **Nhóm 3 (G_3):** Thanh niên (20 - 39 tuổi)
- **Nhóm 4 (G_4):** Trung niên (40 - 59 tuổi)

- **Nhóm 5 (G_5):** Người cao tuổi (Trên 60 tuổi)

Về bản chất, đây là một bài toán **Phân loại (Classification)** trong Học máy, với số lớp là C (số nhóm tuổi).

1.1.3 Lý do chọn đề tài

Đề tài **"Nhận dạng nhóm tuổi từ ảnh chân dung"** được lựa chọn nghiên cứu dựa trên các lý do cốt lõi sau:

- **Tính thời sự và ứng dụng thực tiễn cao:**
  - Trong bối cảnh chuyển đổi số, nhu cầu về các hệ thống có khả năng hiểu và tương tác thông minh với con người ngày càng trở nên cấp thiết.
  - Nhận dạng nhóm tuổi là nền tảng cho nhiều ứng dụng quan trọng: hệ thống quảng cáo thông minh có thể điều chỉnh nội dung theo độ tuổi người xem; các nền tảng truyền thông xã hội có thể tự động kiểm soát nội dung phù hợp với lứa tuổi; hay trong lĩnh vực an ninh, hỗ trợ việc giám sát và nhận dạng đối tượng tự động.
  - Sự phổ biến của camera và thiết bị di động thông minh tạo ra nguồn dữ liệu hình ảnh phong phú, thúc đẩy nhu cầu phát triển các giải pháp phân tích tự động.
- **Tính thách thức về mặt khoa học và kỹ thuật:**
  - Bài toán chứa đựng nhiều khó khăn thú vị: sự khác biệt về tốc độ lão hóa giữa các cá nhân, chủng tộc và giới tính; ảnh hưởng của điều kiện chụp ảnh (ánh sáng, góc máy, độ phân giải); tác động của trang điểm, biểu cảm khuôn mặt; và sự không rõ ràng ở ranh giới giữa các nhóm tuổi.
  - Đây là bài toán đa ngành, kết hợp giữa thị giác máy tính, học máy và hiểu biết về nhân chủng học, tạo ra môi trường nghiên cứu đa dạng và hấp dẫn.
- **Tiềm năng phát triển và cải tiến:**
  - Sự phát triển không ngừng của học sâu (deep learning), đặc biệt là các mô hình mạng nơ-ron tích chập (CNN) và mô hình Visual Transformers,

tiếp tục mở ra những hướng tiếp cận mới để nâng cao độ chính xác và độ ổn định của hệ thống.

- Việc kết hợp và tối ưu hóa các kiến trúc mạng hiện đại để giải quyết các thách thức đặc thù của bài toán vẫn là một lĩnh vực mở, hứa hẹn nhiều đóng góp có giá trị.

1.2 Mục tiêu nghiên cứu

Khóa luận này được thực hiện nhằm hướng đến các mục tiêu nghiên cứu cụ thể sau:

- **Mục tiêu tổng quát:**
  - Nghiên cứu, thiết kế và xây dựng một hệ thống tự động có khả năng nhận dạng nhóm tuổi từ ảnh chân dung sử dụng phương pháp học sâu. Hệ thống hướng đến việc đạt được độ chính xác cao, có khả năng tổng quát hóa tốt trên dữ liệu thực tế và có thể ứng dụng trong nhiều tình huống cụ thể.
- **Mục tiêu cụ thể:**
  1. **Nghiên cứu lý thuyết nền tảng:**
     - Tìm hiểu sâu về các kiến trúc mạng nơ-ron tích chập (CNN) hiện đại được sử dụng trong thị giác máy tính.
     - Nghiên cứu các kỹ thuật Tiền xử lý ảnh và Tăng cường dữ liệu (Data Augmentation) nhằm nâng cao chất lượng dữ liệu đầu vào.
     - Phân tích bài toán Nhận dạng thuộc tính khuôn mặt, đặc biệt tập trung vào bài toán ước lượng tuổi và phân loại nhóm tuổi.
     - Nghiên cứu cơ sở lý thuyết về học chuyển giao và các kiến trúc mạng học sâu phổ biến.
     - Khảo sát và lựa chọn các mô hình pre-trained phù hợp (ResNet, EfficientNet, ViT, . . .).
     - Thiết kế và tinh chỉnh mô hình cho bài toán phân loại nhóm tuổi.
     - Đánh giá hiệu quả của học chuyển giao so với các mô hình baseline.

2. **Chuẩn bị dữ liệu:**
   - Thu thập và lựa chọn các tập dữ liệu ảnh khuôn mặt công khai phù hợp (ví dụ: UTKFace, Adience).
   - Tiến hành tiền xử lý dữ liệu: phát hiện và căn chỉnh khuôn mặt, chuẩn hóa kích thước, cân bằng ánh sáng, và làm sạch dữ liệu.
   - Phân chia dữ liệu một cách khoa học thành các tập Huấn luyện (Training), Kiểm định (Validation) và Kiểm thử (Test) để đánh giá khách quan.
3. **Xây dựng mô hình:**
   - Đề xuất một kiến trúc mạng nơ-ron dựa trên phương pháp **học chuyển giao**, trong đó sử dụng các mô hình đã được huấn luyện trước (pre-trained) trên tập dữ liệu lớn làm bộ trích xuất đặc trưng cho bài toán phân loại nhóm tuổi.
   - Nghiên cứu và áp dụng các chiến lược tinh chỉnh (fine-tuning) phù hợp đối với các lớp của mô hình nhằm nâng cao hiệu quả phân loại và khả năng tổng quát hóa.
4. **Thực nghiệm và đánh giá:**
   - Tiến hành huấn luyện mô hình đề xuất trên tập dữ liệu đã chuẩn bị, sử dụng các hàm mất mát và bộ tối ưu phù hợp.
   - So sánh hiệu quả của mô hình đề xuất với các mô hình cơ sở (baseline) hoặc các phương pháp đã được công bố khác.
   - Sử dụng các chỉ số đánh giá phổ biến trong phân loại như Độ chính xác (Accuracy), Precision, Recall, và F1-Score để đo lường hiệu suất một cách toàn diện.
5. **Phân tích và kết luận:**
   - Phân tích định tính các kết quả, nhận diện các trường hợp mà mô hình dự đoán đúng/sai và tìm hiểu nguyên nhân.
   - Rút ra các kết luận về ưu điểm, hạn chế của phương pháp đề xuất.
   - Đề xuất các hướng nghiên cứu tiếp theo để có thể cải tiến hoặc mở rộng nghiên cứu trong tương lai.

1.3 Thách Thức của Đề Tài và Hướng Giải Quyết

1.3.1 Các Thách Thức Chính

Việc xây dựng một hệ thống nhận dạng nhóm tuổi chính xác và mạnh mẽ phải đối mặt với nhiều thách thức sau:

- **Sự đa dạng về nhân khẩu học và di truyền:**
  - **Thách thức:** Tốc độ và dấu hiệu lão hóa khác nhau rất lớn giữa các cá nhân, giới tính và chủng tộc. Yếu tố gen và môi trường sống tạo nên sự khác biệt khó lường, khiến cho các mô hình tổng quát khó đưa ra dự đoán chính xác cho mọi đối tượng.
- **Ảnh hưởng của điều kiện thu thập dữ liệu:**
  - **Thách thức:** Chất lượng ảnh đầu vào không đồng nhất do sự khác biệt về:
    - **Ánh sáng:** Ảnh thiếu sáng hoặc dư sáng làm mất chi tiết các đặc trưng quan trọng như nếp nhăn, độ đàn hồi da.
    - **Góc chụp (Pose) và biểu cảm:** Khuôn mặt không thẳng, biểu cảm cười hay cau mày có thể làm biến dạng các đặc điểm cấu trúc, che khuất vùng da chứa thông tin về tuổi.
    - **Trang điểm và phụ kiện:** Trang điểm có thể che giấu hoặc làm giảm các dấu hiệu tuổi tác. Kính mát, khẩu trang che khuất các vùng quan trọng như mắt và má.
- **Tính mơ hồ ở ranh giới các nhóm tuổi:**
  - **Thách thức:** Ranh giới giữa các nhóm tuổi (ví dụ giữa 39 và 41 tuổi) là do con người quy ước và không có sự thay đổi đột ngột về mặt sinh học. Điều này dẫn đến sự chồng lấn đáng kể về đặc trưng giữa các nhóm liền kề, gây khó khăn cho việc phân loại.
- **Chất lượng và sự cân bằng của dữ liệu:**

  - **Thách thức:** Các tập dữ liệu công khai thường thiếu sự cân bằng về số lượng mẫu giữa các nhóm tuổi (ví dụ: ít ảnh người cao tuổi và trẻ em hơn ảnh người trung niên). Điều này dễ dẫn đến hiện tượng mô hình bị thiên lệch về nhóm tuổi có nhiều dữ liệu.

1.3.2 Hướng Giải Quyết Đề Xuất

Để giải quyết các thách thức nêu trên, khóa luận đề xuất một framework dựa trên **phương pháp học chuyển giao (Transfer Learning)** với các hướng tiếp cận chính sau:

- **Ứng dụng Học chuyển giao (Transfer Learning):**
  - **Giải pháp:** Sử dụng các mô hình học sâu đã được huấn luyện trước (pre-trained) trên các tập dữ liệu quy mô lớn làm bộ trích xuất đặc trưng, sau đó tiến hành tinh chỉnh (fine-tuning) cho nhiệm vụ phân loại nhóm tuổi từ ảnh chân dung.
  - **Cơ sở:** Các mô hình pre-trained đã học được những đặc trưng thị giác tổng quát và giàu thông tin, giúp mô hình đạt hiệu quả cao ngay cả khi dữ liệu huấn luyện cho bài toán cụ thể còn hạn chế, đồng thời giảm nguy cơ overfitting.
- **Quy trình tiền xử lý dữ liệu mạnh mẽ:**
  - **Giải pháp:** Áp dụng một quy trình tiền xử lý dữ liệu nghiêm ngặt, bao gồm:
    * **Phát hiện và căn chỉnh khuôn mặt** (sử dụng MTCNN hoặc RetinaFace) nhằm chuẩn hóa góc nhìn và vị trí khuôn mặt, giảm ảnh hưởng của pose.
    * **Tăng cường dữ liệu (Data Augmentation):** Áp dụng các kỹ thuật như điều chỉnh độ sáng, độ tương phản, xoay và lật ảnh ngẫu nhiên để tăng tính đa dạng của dữ liệu và cải thiện khả năng tổng quát hóa của mô hình.
    * **Cân bằng dữ liệu:** Sử dụng các kỹ thuật tăng cường dữ liệu có định hướng cho các nhóm tuổi có ít mẫu nhằm giảm thiểu hiện tượng mất cân bằng dữ liệu.

- **Lựa chọn và tinh chỉnh kiến trúc mạng sâu:**
  - **Giải pháp:** Khảo sát và lựa chọn các kiến trúc mạng nơ-ron tích chập hiện đại đã được tiền huấn luyện như ResNet, EfficientNet hoặc Vision Transformer để làm nền tảng cho mô hình phân loại nhóm tuổi.
  - **Cơ sở:** Việc tinh chỉnh có chọn lọc các lớp của mô hình giúp kế thừa tri thức đã học, đồng thời thích nghi hiệu quả với đặc thù của bài toán phân loại nhóm tuổi.

1.4 Phạm Vi Và Đối Tượng Nghiên Cứu

1.4.1 Phạm Vi Nghiên Cứu

Để đảm bảo tính khả thi và tập trung, khóa luận được giới hạn trong các phạm vi sau:

- **Về bài toán:** Nghiên cứu chỉ tập trung vào bài toán **phân loại nhóm tuổi** (ví dụ: chia thành 5 nhóm: 1-12, 13-17, 18-30, 31-45, >45). Bài toán ước lượng tuổi chính xác (regression) sẽ không nằm trong phạm vi của khóa luận này.
- **Về dữ liệu:** Dữ liệu nghiên cứu chủ yếu là các **ảnh chân dung tĩnh** chứa một khuôn mặt người. Các video hoặc ảnh động sẽ không được xem xét. Nghiên cứu sử dụng các bộ dữ liệu công khai, có sẵn nhãn tuổi.
- **Về phương pháp:** Nghiên cứu giới hạn trong việc ứng dụng các mô hình **Học sâu (Deep Learning)**, đặc biệt là các mạng CNN và Transformers. Các phương pháp truyền thống dựa trên đặc trưng thủ công (hand-crafted features) sẽ không được khảo sát sâu.
- **Về đánh giá:** Các thử nghiệm và đánh giá chính sẽ được thực hiện trên một số tập dữ liệu công khai phổ biến (như UTKFace). Việc triển khai hệ thống trên phần cứng thực tế với các ràng buộc về tốc độ xử lý thời gian thực nằm ngoài phạm vi của khóa luận.

1.4.2 Đối Tượng Nghiên Cứu

Các đối tượng chính được tập trung nghiên cứu trong khóa luận bao gồm:

- **Bài toán nhận dạng nhóm tuổi từ ảnh khuôn mặt:** Bao gồm các đặc thù, thách thức, các hướng tiếp cận và các chỉ số đánh giá liên quan.
- **Các mô hình học sâu cho thị giác máy tính:** Đặc biệt là các kiến trúc CNN (ResNet, VGG, EfficientNet,...) và các kiến trúc Vision Transformer (ViT16,...) để cải thiện hiệu suất và độ ổn định của mô hình.
- **Các tập dữ liệu ảnh khuôn mặt:** Các bộ dữ liệu công khai có gán nhãn tuổi, các phương pháp tiền xử lý và tăng cường dữ liệu để nâng cao chất lượng huấn luyện mô hình.
- **Quy trình huấn luyện và đánh giá mô hình:** Các phương pháp đánh giá, các hàm mất mát, kỹ thuật tối ưu hóa được sử dụng để xây dựng và thẩm định một mô hình học máy hiệu quả.

1.5 Đóng Góp Của Khóa Luận

Khóa luận kỳ vọng sẽ đạt được những đóng góp sau:

- **Về Mặt Học Thuật:**
  - Cung cấp một cái nhìn tổng quan và hệ thống về bài toán nhận dạng nhóm tuổi, các thách thức và hướng giải quyết dựa trên học sâu.
  - Đề xuất một kiến trúc mô hình phân loại, góp phần minh chứng cho hiệu quả của việc học các thuộc tính liên quan đồng thời trong việc cải thiện độ chính xác của nhiệm vụ chính.
- **Về Mặt Ứng Dụng:**
  - Xây dựng một mô hình prototype (nguyên mẫu) có khả năng nhận dạng nhóm tuổi, có thể được tích hợp vào các hệ thống ứng dụng thực tế.
  - Cung cấp bộ dữ liệu đã được tiền xử lý và mã nguồn thực nghiệm, có thể trở thành tài liệu tham khảo cho các nghiên cứu tiếp theo.
- **Về Mặt Kinh Nghiệm:** Quá trình thực hiện khóa luận giúp tác giả nâng cao kỹ năng nghiên cứu, phân tích vấn đề, và triển khai các mô hình học máy một cách bài bản.

1.6 Cấu trúc khóa luận

Chương 1: Trình bày tổng quan về đề tài phân loại nhóm tuổi từ ảnh chân dung, bao gồm bối cảnh nghiên cứu, lý do lựa chọn đề tài, mục tiêu, phạm vi và đối tượng nghiên cứu, cũng như các thách thức và hướng tiếp cận đề xuất.

Chương 2: Trình bày cơ sở lý thuyết liên quan đến thị giác máy tính và học sâu, tập trung vào các kiến trúc mạng nơ-ron đã được huấn luyện trước và phương pháp học chuyển giao. Đồng thời, chương này tổng hợp và phân tích các công trình nghiên cứu liên quan đến bài toán nhận dạng tuổi và phân loại nhóm tuổi từ ảnh khuôn mặt.

Chương 3: Trình bày phương pháp đề xuất cho bài toán phân loại nhóm tuổi từ ảnh chân dung dựa trên học chuyển giao, bao gồm quy trình tiền xử lý dữ liệu, lựa chọn mô hình pre-trained, chiến lược tinh chỉnh (fine-tuning) và thiết kế mô hình phân loại.

Chương 4: Trình bày quá trình cài đặt thực nghiệm, mô tả các tập dữ liệu sử dụng, các độ đo đánh giá, và phân tích kết quả thực nghiệm. Chương này cũng so sánh hiệu quả của các mô hình học chuyển giao khác nhau cho bài toán phân loại nhóm tuổi.

Chương 5: Tổng kết các kết quả đạt được của khóa luận, phân tích những hạn chế và thách thức còn tồn tại, đồng thời đề xuất các hướng phát triển và nghiên cứu tiếp theo trong tương lai.

Chương 2

Cơ sở lý thuyết và các công trình nghiên cứu liên quan

2.1 Tổng quan về thị giác máy tính và học sâu

Thị giác máy tính (Computer Vision) là một lĩnh vực quan trọng của trí tuệ nhân tạo, tập trung vào việc giúp máy tính có khả năng thu nhận, phân tích và hiểu thông tin từ hình ảnh và video. Trong những năm gần đây, sự phát triển mạnh mẽ của học sâu (Deep Learning), đặc biệt là các mạng nơ-ron tích chập (Convolutional Neural Networks – CNN), đã tạo ra những bước tiến vượt bậc trong các bài toán như phân loại ảnh, nhận dạng khuôn mặt và ước lượng tuổi [3].

Trong bài toán phân loại nhóm tuổi từ ảnh khuôn mặt, mô hình cần học được các đặc trưng thị giác phức tạp liên quan đến cấu trúc khuôn mặt, nếp nhăn và các đặc điểm hình thái. Những đặc trưng này rất khó được mô hình hóa bằng các luật thủ công, do đó các phương pháp học sâu và học chuyển giao ngày càng trở nên phổ biến và hiệu quả.

2.2 Phương pháp phân loại ảnh truyền thống

Trước khi học sâu trở thành xu hướng chủ đạo, các hệ thống phân loại ảnh chủ yếu dựa trên việc trích xuất đặc trưng thủ công kết hợp với các thuật toán học máy truyền thống. Quy trình chung bao gồm ba bước chính: tiền xử lý ảnh, trích xuất đặc

trưng và huấn luyện mô hình phân loại.

Các đặc trưng phổ biến như Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) và histogram cường độ xám thường được sử dụng để mô tả thông tin biên, kết cấu và phân bố mức xám của ảnh. Sau khi trích xuất, các vector đặc trưng được đưa vào các mô hình phân loại như Support Vector Machine (SVM), K-nearest Neighbors (KNN) hoặc Decision Tree.

Theo hướng tiếp cận được trình bày trong tài liệu của DigitalOcean [1], phương pháp truyền thống có ưu điểm là đơn giản, dễ triển khai và yêu cầu tài nguyên tính toán thấp. Tuy nhiên, hiệu quả của các phương pháp này phụ thuộc mạnh vào chất lượng đặc trưng thủ công và thường gặp hạn chế khi xử lý các bài toán có độ biến thiên lớn như phân loại nhóm tuổi từ ảnh khuôn mặt.

2.3 Mạng nơ-ron tích chập và học chuyển giao

Mạng nơ-ron tích chập (CNN) là nền tảng của hầu hết các hệ thống phân loại ảnh hiện đại. CNN sử dụng các phép tích chập để tự động học các đặc trưng từ dữ liệu ảnh, từ các đặc trưng mức thấp (cạnh, góc) đến các đặc trưng mức cao (hình dạng khuôn mặt) [3].

Học chuyển giao (Transfer Learning) là kỹ thuật sử dụng các mô hình đã được huấn luyện trước trên tập dữ liệu lớn như ImageNet để áp dụng cho các bài toán mới. Thay vì huấn luyện mô hình từ đầu, các trọng số đã học được được giữ nguyên hoặc tinh chỉnh một phần. Phương pháp này giúp giảm thời gian huấn luyện và cải thiện độ chính xác, đặc biệt trong trường hợp dữ liệu huấn luyện có kích thước hạn chế.

2.4 Các kiến trúc mạng nơ-ron sâu tiêu biểu

2.4.1 ResNet

ResNet được đề xuất trong công trình *Deep Residual Learning for Image Recognition* [3]. Điểm nổi bật của ResNet là việc sử dụng các kết nối tắt (skip connections), cho phép mô hình học phần dư thay vì ánh xạ trực tiếp. Nhờ đó, ResNet có thể mở rộng lên hàng chục hoặc hàng trăm lớp mà vẫn duy trì khả năng huấn luyện ổn định.

Các biến thể như ResNet18 và ResNet34 thường được sử dụng trong các bài toán phân loại ảnh do sự cân bằng giữa độ chính xác và chi phí tính toán.

2.4.2 EfficientNet

EfficientNet được giới thiệu trong công trình *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* [5]. Thay vì mở rộng mạng theo một chiều duy nhất, EfficientNet đề xuất phương pháp mở rộng đồng thời độ sâu, độ rộng và độ phân giải ảnh đầu vào thông qua một hệ số tỷ lệ thống nhất.

Cách tiếp cận này giúp EfficientNet đạt hiệu năng cao hơn so với các mô hình CNN truyền thống trong khi vẫn duy trì số lượng tham số ở mức hợp lý. Trong nghiên cứu này, các phiên bản EfficientNet-B0 và EfficientNet-B3 được sử dụng cho bài toán phân loại nhóm tuổi.

2.4.3 Vision Transformer (ViT)

Vision Transformer (ViT) được đề xuất trong công trình *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* [2]. ViT chia ảnh đầu vào thành các khối nhỏ (patches) và xử lý chúng như các chuỗi token tương tự trong xử lý ngôn ngữ tự nhiên.

Thông qua cơ chế self-attention, ViT có khả năng học các mối quan hệ toàn cục giữa các vùng khác nhau trong ảnh. Khi kết hợp với học chuyển giao từ các mô hình đã được huấn luyện trước, ViT cho thấy hiệu quả tốt trong nhiều bài toán thị giác máy tính.

2.5 Nhận xét và định hướng nghiên cứu

Từ các công trình nghiên cứu liên quan, có thể thấy rằng phương pháp phân loại ảnh truyền thống và các mô hình học sâu đều có những ưu điểm và hạn chế riêng. Phương pháp truyền thống phù hợp làm mô hình đường cơ sở do tính đơn giản và khả năng triển khai nhanh [1], trong khi các mô hình học chuyển giao như ResNet, EfficientNet và Vision Transformer cho hiệu năng vượt trội trong các bài toán phức tạp [3, 5, 2].

Do đó, nghiên cứu này triển khai song song cả hai hướng tiếp cận nhằm đánh giá toàn diện hiệu quả của từng phương pháp trong bài toán phân loại nhóm tuổi từ ảnh khuôn mặt.

Chương 3

PHƯƠNG PHÁP ĐỀ XUẤT

3.1 Tổng quan chương

Chương này trình bày chi tiết các phương pháp được sử dụng để giải quyết bài toán phân loại nhóm tuổi từ ảnh khuôn mặt. Hai hướng tiếp cận chính được triển khai và so sánh gồm: (i) phương pháp phân loại ảnh truyền thống dựa trên trích xuất đặc trưng thủ công kết hợp với các bộ phân loại học máy cổ điển; và (ii) phương pháp học sâu dựa trên học chuyển giao (Transfer Learning) với các mô hình CNN và Transformer đã được huấn luyện trước trên các tập dữ liệu quy mô lớn.

3.2 Phương pháp phân loại ảnh truyền thống dựa trên trích xuất đặc trưng

3.2.1 Tiền xử lý dữ liệu ảnh

Ảnh khuôn mặt đầu vào được chuẩn hóa thông qua các bước tiền xử lý cơ bản nhằm giảm nhiễu và đảm bảo tính đồng nhất của dữ liệu, bao gồm:

- Chuyển ảnh màu sang ảnh xám nhằm giảm số chiều dữ liệu và tập trung vào cấu trúc hình học của khuôn mặt.
- Chuẩn hóa kích thước ảnh về cùng độ phân giải 48×48 pixel.
- Chuẩn hóa giá trị điểm ảnh về khoảng $[0, 1]$.

Quá trình tiền xử lý này giúp dữ liệu phù hợp với các kỹ thuật trích xuất đặc trưng truyền thống, đồng thời giảm chi phí tính toán trong quá trình huấn luyện mô hình.

3.2.2 Trích xuất đặc trưng thủ công

Từ các ảnh đã được tiền xử lý, các đặc trưng cấp thấp được trích xuất nhằm mô tả hình dạng và kết cấu khuôn mặt. Cụ thể, khóa luận sử dụng các phương pháp sau:

- **Histogram of Oriented Gradients (HOG):** Mô tả phân bố hướng gradient trong ảnh, phản ánh cấu trúc biên và hình dáng tổng thể của khuôn mặt.
- **Local Binary Pattern (LBP):** Mô tả kết cấu cục bộ của ảnh, đặc biệt hiệu quả trong các bài toán nhận dạng khuôn mặt.
- **Histogram cường độ xám:** Biểu diễn phân bố mức xám toàn cục của ảnh.

Các đặc trưng này được trích xuất độc lập và sử dụng làm đầu vào cho các mô hình phân loại học máy.

3.2.3 Chuẩn hóa và giảm chiều đặc trưng

Do không gian đặc trưng thu được thường có số chiều lớn, dữ liệu được tiếp tục xử lý thông qua:

- Chuẩn hóa bằng phương pháp **StandardScaler** nhằm đảm bảo các đặc trưng có cùng thang đo.
- Giảm chiều bằng phương pháp **Principal Component Analysis (PCA)**, trong đó giữ lại 95% phương sai nhằm giảm nhiễu và hạn chế hiện tượng quá khớp (overfitting).

3.2.4 Mô hình phân loại

Trên không gian đặc trưng sau khi giảm chiều, các bộ phân loại học máy cổ điển được áp dụng, bao gồm:

- Máy véc-tơ hỗ trợ tuyến tính (*Linear Support Vector Machine*);
- Thuật toán K láng giềng gần nhất (*K-Nearest Neighbors*);

- Cây quyết định (*Decision Tree*).

Hiệu năng của các mô hình được đánh giá thông qua các chỉ số *Accuracy*, *Precision*, *Recall* và *F1-score*.

3.3 Phương pháp học chuyển giao (Transfer Learning)

3.3.1 Động cơ lựa chọn học chuyển giao

Trong bối cảnh dữ liệu huấn luyện hạn chế, phương pháp học chuyển giao cho phép tận dụng tri thức đã học từ các tập dữ liệu lớn (chẳng hạn như ImageNet), từ đó mang lại các lợi ích sau:

- Giúp mô hình hội tụ nhanh hơn;
- Cải thiện độ chính xác phân loại;
- Giảm nguy cơ xảy ra hiện tượng quá khớp.

3.3.2 Tiền xử lý và tăng cường dữ liệu

Ảnh đầu vào được xử lý theo chuẩn của các mô hình đã được huấn luyện trước, bao gồm:

- Chuẩn hóa kích thước ảnh về 224×224 pixel;
- Áp dụng các kỹ thuật tăng cường dữ liệu (*Data Augmentation*) như xoay nhẹ, lật ngang, và điều chỉnh độ sáng/độ tương phản;
- Chuẩn hóa ảnh theo thống kê của tập dữ liệu ImageNet.

Bên cạnh đó, các ảnh có chất lượng thấp (ví dụ: ảnh mờ hoặc thiếu thông tin màu sắc) được tự động phát hiện và loại bỏ nhằm nâng cao chất lượng tập dữ liệu huấn luyện.

3.3.3 Xử lý mất cân bằng dữ liệu

Do sự phân bố không đồng đều về số lượng mẫu giữa các nhóm tuổi, phương pháp **Weighted Random Sampling** được sử dụng để tăng tần suất xuất hiện của các lớp thiểu số trong quá trình huấn luyện, từ đó cải thiện khả năng học của mô hình đối với các lớp này.

3.3.4 Kiến trúc mạng đề xuất

Các mô hình học sâu hiện đại được sử dụng như bộ trích xuất đặc trưng, bao gồm:

- **EfficientNet (B0, B3):** Kiến trúc CNN được thiết kế nhằm tối ưu sự cân bằng giữa độ chính xác và chi phí tính toán.
- **ResNet:** Mạng nơ-ron residual giúp huấn luyện các mạng sâu một cách hiệu quả.
- **Vision Transformer (ViT-B/16):** Mô hình Transformer áp dụng cho thị giác máy tính, khai thác mối quan hệ toàn cục giữa các vùng ảnh.

Trong các mô hình này, phần backbone được giữ nguyên trọng số ban đầu, trong khi phần phân loại (*classifier*) được thiết kế lại để phù hợp với số lượng lớp nhóm tuổi.

3.3.5 Chiến lược fine-tuning

Quá trình tinh chỉnh mô hình được thực hiện theo các bước:

- Đóng băng (freeze) toàn bộ backbone trong giai đoạn huấn luyện ban đầu;
- Mở khóa (unfreeze) một số block cuối của mạng để tinh chỉnh các đặc trưng bậc cao;
- Sử dụng tốc độ học khác nhau cho backbone và classifier nhằm đảm bảo quá trình huấn luyện ổn định.

3.3.6 Huấn luyện và tối ưu

Quá trình huấn luyện mô hình được cấu hình như sau:

- Hàm mất mát: *Cross-Entropy Loss*;
- Bộ tối ưu: *AdamW*;
- Điều chỉnh tốc độ học động bằng *ReduceLROnPlateau*;
- Áp dụng kỹ thuật *Early Stopping* nhằm tránh hiện tượng quá khớp.

3.3.7 Đánh giá mô hình

Hiệu năng của mô hình được đánh giá trên tập validation thông qua:

- Độ chính xác tổng thể (*Accuracy*);
- Báo cáo phân loại chi tiết gồm *Precision*, *Recall* và *F1-score* cho từng nhóm tuổi;
- Ma trận nhầm lẫn (*Confusion Matrix*) nhằm phân tích chi tiết các lỗi phân loại.

3.4 So sánh hai hướng tiếp cận

Phần này trình bày sự so sánh giữa phương pháp phân loại truyền thống và phương pháp học chuyển giao dựa trên các tiêu chí:

- Độ chính xác phân loại;
- Khả năng tổng quát hóa;
- Chi phí tính toán và thời gian huấn luyện.

3.5 Kết luận chương

Chương này đã trình bày chi tiết các phương pháp được đề xuất nhằm giải quyết bài toán phân loại nhóm tuổi từ ảnh khuôn mặt. Kết quả phân tích cho thấy phương pháp học chuyển giao có nhiều ưu điểm vượt trội so với phương pháp trích xuất đặc

trưng thủ công. Đây là cơ sở quan trọng cho việc trình bày và phân tích kết quả thực nghiệm trong chương tiếp theo.

Chương 4

THỰC NGHIỆM

4.1 Tập Dữ Liệu và Thiết Lập Thực Nghiệm

4.1.1 Mô Tả Tập Dữ Liệu

Tên tập dữ liệu: UTKFace Dataset

Nguồn gốc tập dữ liệu: Tập dữ liệu UTKFace được xây dựng và công bố bởi các nhà nghiên cứu tại Đại học Tennessee, Knoxville, và được giới thiệu chính thức tại trang web của nhóm tác giả [6].

Nguồn sử dụng trong thực nghiệm: Trong nghiên cứu này, dữ liệu UTKFace được tải và sử dụng thông qua phiên bản công khai trên nền tảng Kaggle [4], nhằm thuận tiện cho quá trình thu thập và tái lập thí nghiệm.

Mô tả chung:

> UTKFace là một tập dữ liệu khuôn mặt quy mô lớn, bao gồm các ảnh khuôn mặt được gán nhãn về tuổi, giới tính và chủng tộc. Tập dữ liệu này được sử dụng rộng rãi trong các nghiên cứu về ước lượng tuổi và phân loại thuộc tính khuôn mặt từ ảnh.

Thông tin chi tiết:

- **Tổng số ảnh ban đầu:** 23,625 ảnh
- **Độ phân giải ảnh:** 200×200 pixels
- **Khoảng tuổi:** Từ 1 đến 116 tuổi

- **Các thuộc tính được gán nhãn:**
  - Tuổi (Age)
  - Giới tính (Gender): 0 – Nam, 1 – Nữ
  - Chủng tộc (Race): 0 – White, 1 – Black, 2 – Asian, 3 – Indian, 4 – Others
- **Định dạng ảnh:** JPEG

Tập dữ liệu sau đó được tiền xử lý, sàng lọc và phân chia thành các tập huấn luyện, kiểm định và kiểm thử phục vụ cho các thí nghiệm trong nghiên cứu này.

4.1.1.1 Tiền Xử Lý Dữ Liệu

Các bước tiền xử lý được áp dụng trên tập dữ liệu:

- **Phân chia nhóm tuổi:** Chuyển đổi tuổi chính xác thành 5 nhóm tuổi:
  - **Nhóm 1 (1-12):** Trẻ em
  - **Nhóm 2 (13-19):** Thiếu niên
  - **Nhóm 3 (20-39):** Thanh niên
  - **Nhóm 4 (40-59):** Trung niên
  - **Nhóm 5 (60+):** Người cao tuổi

4.1.1.2 Phân Chia Dữ Liệu

Tập dữ liệu được phân chia theo tỷ lệ 80-10-10:

- **Tập huấn luyện (Training set):** 16,514 ảnh (80%)
- **Tập kiểm định (Validation set):** 3,552 ảnh (10%)
- **Tập kiểm thử (Test set):** 3,559 ảnh (10%)

Bảng 4.1: Phân bố dữ liệu theo nhóm tuổi sau khi chia

Nhóm tuổi	**Tập huấn luyện**	**Tập kiểm định**	**Tập kiểm thử**	**Tổng**
Trẻ em (1–12)	2,382	510	512	3,404
Thiếu niên (13–17)	825	176	178	1,179
Thanh niên (18–30)	8,316	1,782	1,782	11,880
Trung niên (31–45)	3,180	681	683	4,544
Cao tuổi (46+)	1,811	403	404	2,617
Tổng	**16,514**	**3,552**	**3,559**	**23,625**

4.1.2 Độ Đo Đánh Giá Mô Hình

Để đánh giá hiệu suất của mô hình trong bài toán phân loại nhóm tuổi, các độ đo phổ biến sau được sử dụng:

- **Độ chính xác (Accuracy):**

  Độ chính xác biểu thị tỷ lệ mẫu được dự đoán đúng trên tổng số mẫu trong tập kiểm thử. Trong bài toán phân loại nhiều lớp, độ chính xác được tính trực tiếp như sau:

$$Accuracy = \frac{\text{Số lượng mẫu dự đoán đúng}}{\text{Tổng số mẫu}} \tag{4.1}$$

- **Precision, Recall và F1-score theo từng lớp:**

  Đối với từng nhóm tuổi, các độ đo *Precision*, *Recall* và *F1-score* được sử dụng để đánh giá chi tiết hiệu năng mô hình:

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.4}$$

- **Macro-F1 và Macro Recall (Balanced Accuracy):**

  Do tập dữ liệu có sự chênh lệch đáng kể về số lượng mẫu giữa các nhóm tuổi, các độ đo trung bình không trọng số được sử dụng để đánh giá công bằng hơn hiệu năng mô hình trên từng lớp.

– **Macro Recall (Balanced Accuracy):**

$$Balanced\ Accuracy = \frac{1}{C}\sum_{i=1}^{C} Recall_i \tag{4.5}$$

với $C = 5$ là số nhóm tuổi.

– **Macro-F1:**

$$Macro\text{-}F1 = \frac{1}{C}\sum_{i=1}^{C} F1_i \tag{4.6}$$

Các chỉ số Macro phản ánh rõ ràng sự suy giảm hiệu năng trên các lớp thiểu số, đặc biệt là nhóm *Thiếu niên*, dù độ chính xác tổng thể của mô hình vẫn ở mức cao.

- **Weighted Average:**

  Ngoài ra, các độ đo trung bình có trọng số theo số lượng mẫu của từng lớp (*weighted average*) cũng được báo cáo nhằm phản ánh hiệu năng tổng thể trong điều kiện phân bố dữ liệu không đồng đều.

- **Ma trận nhầm lẫn (Confusion Matrix):** Biểu diễn trực quan số lượng mẫu của từng lớp được dự đoán đúng hoặc sai. Mỗi phần tử M_{ij} thể hiện số lượng mẫu thuộc lớp thật i nhưng được dự đoán là lớp j.

4.2 Phương pháp phân loại ảnh truyền thống

4.2.1 Tổng quan phương pháp

Phương pháp phân loại ảnh truyền thống trong nghiên cứu này được triển khai theo đúng quy trình đã trình bày ở Chương 3, bao gồm ba giai đoạn chính: (i) tiền xử lý và chuẩn hóa dữ liệu ảnh, (ii) trích xuất đặc trưng thủ công, và (iii) huấn luyện các mô hình phân loại học máy dựa trên các đặc trưng đã trích xuất.

Trong Chương 4, nội dung tập trung mô tả chi tiết cách thức triển khai thực nghiệm, các tham số cấu hình cụ thể cũng như thiết lập huấn luyện của từng thành phần trong quy trình. Phương pháp truyền thống được sử dụng như một mô hình đường cơ sở (baseline) nhằm đánh giá hiệu quả của các đặc trưng thị giác cổ điển trong bài toán phân loại nhóm tuổi và làm cơ sở so sánh với các phương pháp học sâu và học chuyển giao ở các mục tiếp theo.

4.2.2 Tiền xử lý và chuẩn hóa ảnh

Toàn bộ ảnh khuôn mặt trong tập dữ liệu được chuyển sang ảnh xám bằng phép biến đổi *Grayscale* với một kênh đầu ra duy nhất. Việc sử dụng ảnh xám giúp giảm số chiều dữ liệu đầu vào, đồng thời loại bỏ thông tin màu sắc không cần thiết đối với các đặc trưng thủ công được sử dụng trong nghiên cứu này.

Sau đó, tất cả ảnh được chuẩn hóa về cùng kích thước 48×48 pixels nhằm đảm bảo tính đồng nhất của dữ liệu đầu vào cho quá trình trích xuất đặc trưng. Giá trị điểm ảnh được chuyển đổi sang kiểu số thực và chuẩn hóa về khoảng $[0, 1]$ bằng cách chia cho 255, giúp cải thiện tính ổn định của các thuật toán học máy ở các bước tiếp theo.

Quá trình tiền xử lý được áp dụng nhất quán cho cả tập huấn luyện và tập kiểm định.

4.2.3 Trích xuất đặc trưng

Ba loại đặc trưng thủ công phổ biến được sử dụng trong nghiên cứu này, tương ứng với các lựa chọn trong quá trình thực nghiệm:

- **Histogram of Oriented Gradients (HOG):** Đặc trưng HOG được trích xuất với số hướng gradient bằng 8, kích thước mỗi ô (cell) là 4×4 pixels và block gồm 1×1 cell. Đặc trưng này tập trung mô tả thông tin biên và cấu trúc hình dạng tổng thể của khuôn mặt.

- **Local Binary Patterns (LBP):** Đặc trưng LBP được trích xuất với các tham số $P = 8$, $R = 1$ và phương pháp *uniform*. Đặc trưng này phản ánh các mẫu kết cấu cục bộ trên bề mặt khuôn mặt.

- **Histogram cường độ xám:** Phân bố mức xám toàn cục của ảnh được biểu diễn thông qua histogram với 32 bins trong khoảng giá trị $[0, 1]$.

Đối với mỗi loại đặc trưng, quá trình trích xuất chỉ được thực hiện một lần và các vector đặc trưng thu được được lưu lại nhằm tăng hiệu quả tính toán trong các lần huấn luyện tiếp theo.

Sau khi trích xuất, các vector đặc trưng được chuẩn hóa bằng phương pháp *StandardScaler* để đưa các đặc trưng về cùng thang đo. Tiếp theo, phương pháp *Principal Component Analysis (PCA)* được áp dụng nhằm giảm chiều dữ liệu, với số thành phần

chính được lựa chọn sao cho giữ lại 95% phương sai của dữ liệu ban đầu. Việc giảm chiều giúp hạn chế nhiễu và giảm nguy cơ quá khớp trong quá trình huấn luyện mô hình.

4.2.4 Mô hình phân loại và cấu hình tham số

Ba mô hình học máy truyền thống được sử dụng trong thực nghiệm bao gồm:

- **Máy vector hỗ trợ tuyến tính (Linear SVM):** Mô hình Linear SVM được cấu hình với hệ số điều chuẩn $C = 1.0$, hàm mất mát *squared hinge*, số vòng lặp tối đa là 2000 và ngưỡng hội tụ $tol = 10^{-3}$. Mô hình sử dụng bộ sinh số ngẫu nhiên với *random state* bằng 42 nhằm đảm bảo khả năng tái lập kết quả.
- **K-nearest Neighbors (KNN):** Mô hình KNN được thiết lập với số láng giềng $k = 5$ và sử dụng trọng số theo khoảng cách, giúp giảm ảnh hưởng của các mẫu huấn luyện nằm xa trong không gian đặc trưng.
- **Cây quyết định (Decision Tree):** Mô hình cây quyết định được giới hạn độ sâu tối đa ở mức 10 và sử dụng *random state* bằng 42 nhằm hạn chế hiện tượng quá khớp và đảm bảo tính ổn định của mô hình.

Các mô hình được huấn luyện trên tập huấn luyện sau khi đã áp dụng chuẩn hóa và giảm chiều dữ liệu, và được đánh giá trên tập kiểm thử bằng các độ đo hiệu năng được trình bày ở Mục 4.4.

4.3 Phương pháp học chuyển giao (Transfer Learning)

4.3.1 Tiền xử lý và lọc dữ liệu

Trước khi tiến hành huấn luyện, tập dữ liệu UTKFace được tiền xử lý nhằm loại bỏ các mẫu không đảm bảo chất lượng. Hai tiêu chí chính được áp dụng như sau:

- **Độ mờ ảnh (Blur):** Các ảnh có độ nét nhỏ hơn ngưỡng

$$\texttt{BLUR\_THRESHOLD} = 15$$

bị loại bỏ nhằm hạn chế ảnh hưởng tiêu cực đến khả năng học đặc trưng.

- **Độ phong phú màu sắc:** Các ảnh có số lượng màu hiệu quả nhỏ hơn

$$\texttt{COLOR\_THRESHOLD} = 500$$

được xem là thiếu thông tin và không được sử dụng.

Sau quá trình sàng lọc, tập dữ liệu cuối cùng gồm **19,745 ảnh khuôn mặt hợp lệ**, được sử dụng cho toàn bộ các thí nghiệm.

Tập dữ liệu sau tiền xử lý được chia theo phương pháp *stratified split* nhằm đảm bảo phân bố các nhóm tuổi được duy trì đồng đều giữa các tập dữ liệu. Tỷ lệ phân chia được sử dụng là 80% cho tập huấn luyện, 10% cho tập kiểm định và 10% cho tập kiểm thử, tương ứng với:

- **Tập huấn luyện (Training set):** 15,796 ảnh
- **Tập kiểm định (Validation set):** 1,974 ảnh
- **Tập kiểm thử (Test set):** 1,975 ảnh

Bảng 4.2: Phân bố dữ liệu theo nhóm tuổi sau khi tiền xử lý và phân chia

Nhóm tuổi	**Train**	**Val**	**Test**	**Tổng**
Trẻ em (1–12)	2,312	289	290	2,891
Thiếu niên (13–19)	851	106	107	1,064
Thanh niên (20–39)	7,649	956	957	9,562
Trung niên (40–59)	3,062	383	383	3,828
Người lớn tuổi (60+)	1,920	240	240	2,400
Tổng	**15,794**	**1,974**	**1,977**	**19,745**

—

4.3.2 Biến đổi dữ liệu

Nhằm tăng khả năng tổng quát hóa của mô hình, các phép biến đổi dữ liệu khác nhau được áp dụng cho tập huấn luyện và tập kiểm định.

4.3.2.1 Tập huấn luyện

Các phép biến đổi được sử dụng bao gồm:

- Thay đổi kích thước ảnh về 224×224 pixels;
- Lật ngang ngẫu nhiên;
- Xoay ngẫu nhiên trong khoảng $\pm 10^{\circ}$;
- Điều chỉnh nhẹ độ sáng và độ tương phản;
- Chuẩn hóa theo thống kê ImageNet.

4.3.2.2 Tập kiểm định và kiểm thử

Chỉ áp dụng thay đổi kích thước và chuẩn hóa ImageNet nhằm đảm bảo tính nhất quán trong quá trình đánh giá.

—

4.3.3 Cấu hình huấn luyện

Các tham số huấn luyện chung được áp dụng cho tất cả các mô hình như sau:

- **Batch size:** 64
- **Số epoch tối đa:** 50
- **Hàm mất mát:** Cross-Entropy Loss
- **Early stopping:** patience = 8, $\Delta_{min} = 0.001$
- **Scheduler:** ReduceLROnPlateau (factor = 0.5, patience = 5)

—

4.3.4 Phương pháp học chuyển giao

Nghiên cứu này tập trung đánh giá hiệu quả của phương pháp học chuyển giao (Transfer Learning) trong bài toán phân loại nhóm tuổi. Các mô hình được khởi tạo bằng trọng số huấn luyện trước trên ImageNet và được tinh chỉnh để phù hợp với bài toán phân loại 5 nhóm tuổi.

—

4.3.5 Các mô hình và chiến lược huấn luyện

4.3.5.1 EfficientNet-B0

- **Backbone:** EfficientNet-B0 (ImageNet)
- **Chiến lược huấn luyện:** Đóng băng toàn bộ backbone và mở khóa 4 block cuối.
- **Kiến trúc classifier:**

$$\text{Dropout}(0.4) \rightarrow \text{Linear}(1280 \rightarrow 512) \rightarrow \text{ReLU} \rightarrow \text{Dropout}(0.2) \rightarrow \text{Linear}(512 \rightarrow 5)$$

- **Learning rate:**
  - Backbone: 1×10^{-4}
  - Classifier: 1×10^{-3}

—

4.3.5.2 EfficientNet-B3

- **Backbone:** EfficientNet-B3
- **Chiến lược huấn luyện:** Fine-tuning 4 block cuối.

$$\text{Dropout}(0.4) \rightarrow \text{Linear}(1536 \rightarrow 768) \rightarrow \text{ReLU} \rightarrow \text{Dropout}(0.2) \rightarrow \text{Linear}(768 \rightarrow 5)$$

- **Learning rate:**
  - Backbone: 5×10^{-5}
  - Classifier: 1×10^{-3}

—

4.3.5.3 Vision Transformer (ViT-B/16)

- **Backbone:** ViT-B/16
- **Chiến lược huấn luyện:**
  - Đóng băng patch embedding và encoder;
  - Mở khóa 3 transformer block cuối.

$$\text{Linear}(768 \to 5)$$

- **Learning rate:**
  - Encoder: 1×10^{-4}
  - Head: 1×10^{-3}

—

4.3.5.4 ResNet-18

$$\text{Dropout}(0.3) \to \text{Linear}(512 \to 256) \to \text{ReLU} \to \text{Dropout}(0.2) \to \text{Linear}(256 \to 5)$$

Learning rate được phân tầng từ 10^{-5} đến 10^{-3} tương ứng với từng residual block.

—

4.3.5.5 ResNet-34

Kiến trúc classifier và chiến lược huấn luyện tương tự ResNet-18, với backbone sâu hơn nhằm khai thác đặc trưng phức tạp hơn.

—

4.3.6 Tổng hợp tham số mô hình

Bảng 4.3: Tổng hợp kiến trúc classifier và số lượng tham số

Mô hình	Classifier	Tổng tham số	Tham số huấn luyện	Fine-tuning
EfficientNet-B0	1280→512→5	∼5.3M	∼3.7M	4 block cuối
EfficientNet-B3	1536→768→5	∼12.0M	∼9.9M	4 block cuối
ResNet-18	512→256→5	∼11.3M	∼11.3M	layer1–4
ResNet-34	512→256→5	∼21.8M	∼21.8M	layer1–4
ViT-B/16	768→5	∼86.0M	∼21.2M	3 encoder blocks

4.4 Đánh giá kết quả đạt được

Phần này trình bày đánh giá kết quả phân loại nhóm tuổi từ ảnh khuôn mặt dựa trên hai phương pháp: phân loại ảnh truyền thống và học chuyển giao. Việc đánh giá được thực hiện thông qua các độ đo Accuracy, Precision, Recall và F1-score, đồng thời phân tích ma trận nhầm lẫn nhằm làm rõ hành vi dự đoán của mô hình.

4.4.1 Đánh giá theo các độ đo định lượng

4.4.1.1 Kết quả của phương pháp phân loại ảnh truyền thống

Bảng 4.4 trình bày kết quả đánh giá của các mô hình truyền thống sử dụng đặc trưng thủ công (HOG/LBP/Histogram) kết hợp với các bộ phân loại cổ điển.

Bảng 4.4: Kết quả đánh giá của các mô hình phân loại ảnh truyền thống

Mô hình	Accuracy	Macro Precision	Macro Recall	Macro F1
Linear SVM	0.66	0.57	0.51	0.51
KNN	0.62	0.54	0.45	0.48
Decision Tree	0.54	0.42	0.36	0.38

Có thể nhận thấy rằng Linear SVM đạt kết quả tốt nhất trong nhóm phương pháp truyền thống, tuy nhiên các độ đo Macro F1-score vẫn ở mức thấp. Điều này cho thấy

các mô hình truyền thống gặp khó khăn trong việc phân loại đồng đều các nhóm tuổi, đặc biệt là các lớp có số lượng mẫu nhỏ như *Thiếu niên*.

4.4.1.2 Kết quả của phương pháp học chuyển giao

Bảng 4.5 tổng hợp kết quả đánh giá của các mô hình học chuyển giao dựa trên mạng nơ-ron sâu và Transformer.

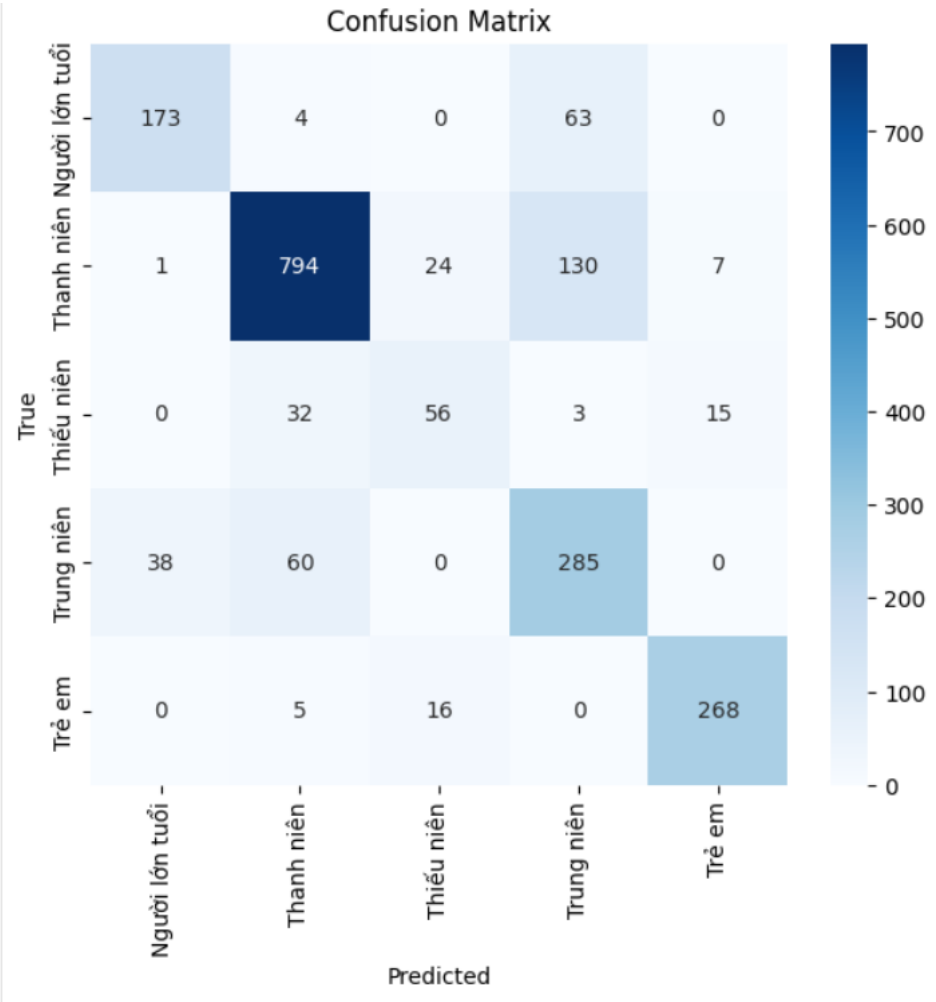
Bảng 4.5: Kết quả đánh giá của các mô hình học chuyển giao

Mô hình	**Accuracy**	**Macro Precision**	**Macro Recall**	**Macro F1**
EfficientNet-B0	0.80	0.75	0.73	0.74
EfficientNet-B3	0.80	0.76	0.75	0.75
ResNet18	0.78	0.74	0.72	0.72
ResNet34	0.78	0.73	0.73	0.73
ViT-B/16	0.76	0.70	0.68	0.69

Kết quả cho thấy các mô hình học chuyển giao đều vượt trội hơn rõ rệt so với phương pháp truyền thống. Trong đó, EfficientNet-B3 đạt kết quả cao nhất với Accuracy xấp xỉ 80% và Macro F1-score đạt 0.75. Các mô hình CNN nhìn chung cho hiệu quả cao hơn Transformer trong bối cảnh dữ liệu không quá lớn.

4.4.2 Phân tích ma trận nhầm lẫn

Hình **??** minh họa ma trận nhầm lẫn của mô hình học chuyển giao tiêu biểu. Các giá trị trên đường chéo chính thể hiện số lượng mẫu được phân loại đúng cho từng nhóm tuổi.



Hình 4.1: Ma trận nhầm lẫn của mô hình học chuyển giao

Có thể nhận thấy rằng các nhóm *Thanh niên* và *Trẻ em* được phân loại với độ chính xác cao, thể hiện qua số lượng lớn các mẫu nằm trên đường chéo chính. Ngược lại, sự nhầm lẫn chủ yếu xảy ra giữa các nhóm tuổi liền kề như *Người lớn tuổi* và *Trung niên*, cũng như giữa *Thanh niên* và *Trung niên*. Điều này phản ánh đặc điểm sinh học tự nhiên khi các đặc trưng khuôn mặt thay đổi dần theo độ tuổi và có sự chồng lấn giữa các nhóm.

4.4.3 Nhận xét chung

Tổng hợp các kết quả thực nghiệm cho thấy phương pháp học chuyển giao mang lại hiệu quả vượt trội cả về độ chính xác tổng thể lẫn khả năng phân loại cân bằng giữa các lớp. Trong khi đó, phương pháp truyền thống chỉ phù hợp như một mô hình

đường cơ sở (baseline) nhằm minh họa giới hạn của các đặc trưng thủ công trong bài toán phân loại nhóm tuổi từ ảnh khuôn mặt.

Chương 5

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 Tổng kết kết quả đạt được

Khóa luận này đã nghiên cứu và triển khai hai hướng tiếp cận chính cho bài toán phân loại nhóm tuổi từ ảnh khuôn mặt, bao gồm: (i) phương pháp phân loại ảnh truyền thống dựa trên trích xuất đặc trưng thủ công kết hợp với các mô hình học máy cổ điển; và (ii) phương pháp học sâu dựa trên học chuyển giao (Transfer Learning) sử dụng các kiến trúc mạng nơ-ron hiện đại đã được huấn luyện trước.

Đối với phương pháp truyền thống, các đặc trưng thị giác phổ biến như HOG, LBP và histogram cường độ xám được trích xuất từ ảnh khuôn mặt sau tiền xử lý. Các đặc trưng này sau đó được chuẩn hóa, giảm chiều bằng PCA và đưa vào các bộ phân loại Linear SVM, KNN và Decision Tree. Kết quả thực nghiệm cho thấy Linear SVM đạt hiệu năng tốt nhất trong nhóm này với độ chính xác đạt khoảng 66%. Tuy nhiên, các chỉ số Macro-F1 và Macro Recall đều ở mức thấp, phản ánh hạn chế rõ rệt của các đặc trưng thủ công trong việc phân biệt đồng đều các nhóm tuổi, đặc biệt là các lớp thiểu số.

Đối với phương pháp học chuyển giao, các mô hình CNN và Transformer hiện đại như EfficientNet-B0, EfficientNet-B3, ResNet-18, ResNet-34 và Vision Transformer (ViT-B/16) được sử dụng làm bộ trích xuất đặc trưng. Kết quả cho thấy tất cả các mô hình học chuyển giao đều vượt trội hơn đáng kể so với phương pháp truyền thống.

Trong đó, EfficientNet-B3 đạt hiệu năng cao nhất với độ chính xác xấp xỉ 80% và Macro F1-score đạt 0.75. Các mô hình CNN nhìn chung cho kết quả ổn định và hiệu quả hơn so với Vision Transformer trong bối cảnh tập dữ liệu có quy mô trung bình.

Phân tích ma trận nhầm lẫn cho thấy các nhóm tuổi có số lượng mẫu lớn như *Thanh niên* và *Trẻ em* được phân loại với độ chính xác cao, trong khi các nhầm lẫn chủ yếu xảy ra giữa các nhóm tuổi liền kề, phản ánh đặc điểm sinh học tự nhiên của sự thay đổi khuôn mặt theo độ tuổi.

Nhìn chung, các kết quả đạt được đã chứng minh tính hiệu quả của phương pháp học chuyển giao trong bài toán phân loại nhóm tuổi từ ảnh khuôn mặt, đồng thời khẳng định vai trò của các mô hình học sâu hiện đại trong việc thay thế các phương pháp trích xuất đặc trưng thủ công truyền thống.

5.2 Hạn chế và thách thức

Mặc dù đạt được những kết quả khả quan, khóa luận vẫn tồn tại một số hạn chế và thách thức nhất định.

Thứ nhất, tập dữ liệu sử dụng có sự mất cân bằng rõ rệt giữa các nhóm tuổi, đặc biệt là nhóm *Thiếu niên* và *Người lớn tuổi*. Mặc dù đã áp dụng các kỹ thuật xử lý mất cân bằng dữ liệu, hiệu năng trên các lớp này vẫn còn hạn chế.

Thứ hai, việc phân chia tuổi thành các nhóm rời rạc có thể làm mất đi một phần thông tin liên tục vốn có của thuộc tính tuổi. Điều này dẫn đến các trường hợp nhầm lẫn giữa các nhóm tuổi liền kề, đặc biệt trong những khoảng tuổi chuyển tiếp.

Thứ ba, các mô hình học sâu, đặc biệt là Vision Transformer, đòi hỏi chi phí tính toán và tài nguyên phần cứng lớn. Trong điều kiện dữ liệu và tài nguyên huấn luyện còn hạn chế, tiềm năng của các mô hình này chưa được khai thác tối đa.

Cuối cùng, nghiên cứu hiện tại chỉ tập trung vào ảnh khuôn mặt tĩnh và chưa xem xét ảnh hưởng của các yếu tố ngoại cảnh như biểu cảm, ánh sáng, góc chụp hay chất lượng ảnh đến hiệu năng của mô hình.

5.3 Hướng phát triển và nghiên cứu trong tương lai

Dựa trên các kết quả và hạn chế đã phân tích, một số hướng phát triển và nghiên cứu tiếp theo có thể được đề xuất như sau.

Trước hết, có thể mở rộng và cân bằng tập dữ liệu bằng cách thu thập thêm dữ liệu từ các nguồn khác hoặc áp dụng các kỹ thuật sinh dữ liệu tổng hợp nhằm cải thiện hiệu năng trên các lớp thiểu số.

Thứ hai, thay vì phân loại nhóm tuổi rời rạc, các hướng tiếp cận kết hợp giữa hồi quy tuổi và phân loại nhóm tuổi, hoặc các mô hình ordinal classification, có thể được nghiên cứu nhằm phản ánh tốt hơn mối quan hệ thứ tự giữa các nhóm tuổi.

Thứ ba, việc khai thác các kiến trúc học sâu tiên tiến hơn như Vision Transformer quy mô lớn, các mô hình hybrid CNN–Transformer hoặc các cơ chế attention chuyên biệt cho khuôn mặt có thể giúp cải thiện hiệu năng phân loại.

Cuối cùng, nghiên cứu trong tương lai có thể mở rộng sang các bài toán đa nhiệm, kết hợp đồng thời ước lượng tuổi, giới tính và chủng tộc, hoặc ứng dụng mô hình vào các hệ thống thực tế như giám sát thông minh, phân tích hành vi người dùng và tương tác người–máy.

5.4 Kết luận chung

Khóa luận đã hoàn thành mục tiêu đề ra là xây dựng, triển khai và đánh giá các phương pháp phân loại nhóm tuổi từ ảnh khuôn mặt. Thông qua các thí nghiệm và phân tích chi tiết, nghiên cứu đã chứng minh ưu thế rõ rệt của phương pháp học chuyển giao so với các phương pháp truyền thống. Các kết quả đạt được không chỉ có ý nghĩa học thuật mà còn tạo tiền đề cho các nghiên cứu và ứng dụng thực tiễn trong lĩnh vực thị giác máy tính và trí tuệ nhân tạo.

TÀI LIỆU THAM KHẢO

[1] DigitalOcean. *Image Classification Without Neural Networks*. https://www.digitalocean.com/community/tutorials/image-classification-without-neural-networks. 2020.

[2] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *International Conference on Learning Representations (ICLR)* (2021). URL: https://arxiv.org/abs/2010.11929.

[3] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016). URL: https://arxiv.org/abs/1512.03385.

[4] Jangedoo. *UTKFace Dataset (Kaggle)*. https://www.kaggle.com/datasets/jangedoo/utkface-new. Accessed: 2025. 2018.

[5] Mingxing Tan and Quoc Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning* (2019). URL: https://arxiv.org/abs/1905.11946.

[6] Zhifei Zhang, Yang Song, and Hairong Qi. *UTKFace: Large-Scale Face Dataset with Age, Gender and Ethnicity Labels*. https://susanqq.github.io/UTKFace/. University of Tennessee, Knoxville. 2017.