TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Nghiên cứu và đề xuất mô hình dự đoán …

NGUYỄN VĂN A
nguyenvanabc@sis.hust.edu.vn

Ngành: …

Giảng viên hướng dẫn: TS. Trần Văn B _____

Chữ kí GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 06/2022

LỜI CẢM ƠN

Trong quá trình hoàn thành đồ án tốt nghiệp này, tôi muốn gửi lời cảm ơn sâu sắc đến thầy Ngô Thành Trung, nhờ sự hướng dẫn nhiệt tình, chỉ bảo tận tình và kiến thức sâu rộng mà thầy chia sẻ đã giúp tôi hoàn thiện tốt bài tốt nghiệp này. Bên cạnh đó, tôi cũng xin bày tỏ lòng biết ơn đến các thầy cô Đại học Bách khoa, đặc biệt là các thầy cô ngành Khoa học máy tính , những người đã truyền cảm hứng và sự am hiểu về chuyên môn cho tôi trong suốt thời gian học tập vừa qua.

Một lần nữa, tôi xin chân thành cảm ơn gia đình, bạn bè và thầy cô đã luôn ở bên, hỗ trợ và động viên tôi trong suốt quá trình thực hiện đồ án tốt nghiệp. Sự quan tâm và giúp đỡ của mọi người đã giúp tôi vượt qua nhiều thử thách và đạt được kết quả như hôm nay. Cuối cùng, tôi cũng tự hào vì bản thân đã chăm chỉ và quyết tâm hoàn thành đồ án. Xin cảm ơn tất cả!

LỜI CAM KẾT

Họ và tên sinh viên: Hoàng Ngọc Lâm
Điện thoại liên lạc: 0865543340
Email: lam.hn194089@sis.hust.edu.vn
Lớp: Khoa học máy tính 02-K64
Hệ đào tạo: Khoa học máy tính

Tôi – *Hoàng Ngọc Lâm* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *TS. Ngô Thành Trung*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày tháng năm

Tác giả ĐATN

Họ và tên sinh viên

TÓM TẮT NỘI DUNG ĐỒ ÁN

Trong bối cảnh hiện đại, nhận diện khuôn mặt là một công nghệ quan trọng với nhiều ứng dụng trong an ninh, kiểm soát truy cập. Tuy nhiên, các phương pháp truyền thống dựa trên hình ảnh 2D gặp nhiều hạn chế khi đối diện với các yếu tố như ánh sáng, góc nhìn và biểu cảm khuôn mặt. Các phương pháp hiện đại hơn, bao gồm việc sử dụng hình ảnh 3D và các kỹ thuật học sâu, đã mang lại những cải tiến đáng kể, nhưng vẫn chưa thể giải quyết hoàn toàn các thách thức này. Để khắc phục những hạn chế trên, tôi lựa chọn hướng tiếp cận kết hợp cả hình ảnh 2D và 3D nhằm nâng cao độ chính xác và tính ổn định của hệ thống nhận diện khuôn mặt. Hình ảnh 3D cung cấp thông tin về hình dạng và cấu trúc khuôn mặt, giúp hệ thống phân biệt và nhận diện tốt hơn trong điều kiện ánh sáng và góc nhìn hạn chế.

Giải pháp của tôi bao gồm ba bước chính: tiền xử lý dữ liệu, xây dựng và huấn luyện mô hình, thử nghiệm đánh giá mô hình. Bước đầu tiên, sử dụng bộ dữ liệu khuôn mặt được thu thập và tiền xử lý để thu được bộ dữ liệu 2D và 3D. Sau đó, các mô hình học sâu được huấn luyện trên cả dữ liệu 2D và 3D để khai thác tối đa các đặc trưng của khuôn mặt. Cuối cùng, mô hình được đánh giá thông qua các thử nghiệm so sánh đặc trưng trưng khuôn mặt. Đóng góp chính của đồ án là phát triển một mô hình nhận diện khuôn mặt kết hợp 2D-3D với độ chính xác cao và khả năng hoạt động ổn định trong nhiều điều kiện khác nhau. Kết quả thử nghiệm cho thấy mô hình đạt được độ chính xác cả thiện hơn so với việc chỉ sử dụng các thông tin 2D và 3D riêng lẻ, từ đó mở ra tiềm năng ứng dụng rộng rãi trong thực tiễn.

Sinh viên thực hiện

(Ký và ghi rõ họ tên)

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	Ý nghĩa
2D	Hai chiều (Two-dimensional)
3D	Ba chiều (Three-dimensional)
AUC	Diện tích dưới đường cong ROC (Area Under the ROC Curve)
CNN	Mạng nơ-ron tích chập (Convolutional Neural Network)
DCNN	Mạng nơ-ron tích chập sâu (Deep Convolutional Neural Network)
GPU	Đơn vị xử lý đồ họa (Graphics Processing Unit)
IR	Hồng ngoại (Infrared)
LBP	Mẫu nhị phân cục bộ (Local Binary Pattern)
LDA	Phân tích biệt thức tuyến tính (Linear Discriminant Analysis)
PCA	Phân tích thành phần chính (Principal Component Analysis)
RGB	Mô hình không gian màu đỏ-xanh-lam (Red-Green-Blue)
ROC	đường cong đặc trưng hoạt động của bộ thu nhận (Receiver Operating Characteristic)
SVM	Máy vector hỗ trợ (Support Vector Machine)

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

1.1 Bài toán nhận diện khuôn mặt

Nhận diện khuôn mặt là một bài toán quan trọng trong lĩnh vực thị giác máy tính, cho phép xác định hoặc nhận dạng danh tính của một người từ hình ảnh hoặc video chưa khuôn mặt. Nhận dạng sinh trắc học là quá trình xác định các cá nhân dựa trên các nét độc nhất, có thể phân biệt được bao gồm: dấu vân tay, mống mắt, giọng nói, ... trong đó nhận diện khuôn mặt là một phương pháp nhận diện có tính thuận tiện và tích hợp cao. Công nghệ nhận diện khuôn mặt có nhiều ứng dụng thực tế trong các lĩnh vực như an ninh bảo mật, theo dõi giám sát, mạng xã hội và giải trí, bán hàng và marketing, ... Bên cạnh những thành công trong việc ứng dụng thực tế, nhận diện khuôn mặt cũng gặp phải các khó khăn hạn chế như các biến dạng và thay đổi do góc nhìn, ánh sáng, biểu cảm, khả năng mở rộng với số lượng khuôn mặt lớn, ...

Bài toán nhận diện khuôn mặt là một bài toán có tầm quan trọng lớn đối với cả công nghệ và xã hội. Việc giải quyết được các thách thức của bài toán sẽ mở ra nhiều cơ hội cho các ứng dụng tiên tiến, cải thiện bảo mật an ninh và nâng cao chất lượng cuộc sống xã hội.

1.2 Các giải pháp hiện tại và hạn chế

Nhận diện khuôn mặt là một lĩnh vực nghiên cứu phát triển nhanh chóng với nhiều ứng dụng thực tiễn trong đời sống hàng ngày. Những năm gần đây, nhận diện khuôn mặt có được sự đột phá lớn với sự ra đời của mạng nơ-ron học sâu. Tuy nhiên trước năm 2014, nhận diện khuôn mặt chủ yếu được thực hiện bằng các phương pháp không sử dụng học sâu.

1.2.1 Phương pháp truyền thống

Trước khi các mô hình học sâu trở nên phổ biến, nhiều phương pháp nhận diện khuôn mặt truyền thống đã được phát triển và ứng dụng rộng rãi. Các phương pháp này chủ yếu dựa vào việc trích xuất và phân tích các đặc trưng cục bộ và toàn cục của khuôn mặt. Dưới đây là một số phương pháp tiêu biểu:

- Phương pháp Eigenface PCA được giới thiệu lần đầu tiên vào đầu những năm 1991 bởi Matthew Turk và Alex Pentland tại Học viện Công nghệ Massachusetts (MIT). Eigenfaces sử dụng phân tích thành phần chính (PCA) để giảm số chiều của dữ liệu hình ảnh và tìm ra các đặc trưng chính đại diện cho khuôn mặt.
- Phương pháp Local Binary Patterns (LBP) được giới thiệu lần đầu vào năm 1996 bởi Timo Ojala, Matti Pietikäinen và David Harwood tại Đại học Oulu,

Phần Lan. LBP là một phương pháp trong xử lý ảnh dựa trên việc biến đổi các giá trị pixel thành mẫu nhị phân bằng cách so sánh pixel với các pixel lân cận, giúp mô tả các đặc trưng cục bộ như cạnh, góc, cấu trúc của khuôn mặt.

- Phương pháp Fisherfaces ra đời từ năm 1997, khi Peter N. Belhumeur, Joao P. Hespanha, và David J. Kriegman giới thiệu phương pháp này trong bài báo "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection". Fisherfaces sử dụng phân tích biệt thức tuyến tính (LDA) để tối đa tỷ lệ giữa phương sai của các lớp khác (khuôn mặt khác nhau) và phương sai trong cùng lớp (cùng một khuôn mặt).

Các phương pháp truyền thống không sử dụng mạng học sâu chủ yếu dùng các thuật toán để trích xuất khuôn mặt thành các vector đặc trưng sau đó so sánh vector hoặc dùng bộ phân loại SVM để nhận diện khuôn mặt. Các phương pháp như Eigenfaces, Fisherfaces và LBP đã đóng góp quan trọng vào lĩnh vực nhận diện khuôn mặt nhưng gặp nhiều hạn chế khi có sự thay đổi về ánh sáng, góc chụp và biểu cảm của khuôn mặt, làm giảm hiệu suất nhận diện so với các phương pháp học sâu hiện đại.

1.2.2 Phương pháp hiện đại

Các năm gần đây với sự phát triển của các mô hình học sâu đã mang lại nhiều đột phá lớn trong lĩnh vực nhận diện khuôn mặt. Các mô hình mạng học sâu không chỉ mang lại độ chính xác cao mà còn có khả năng nhận diện trong các điều kiện khó khăn.

Các phương pháp truyền thống sử dụng dữ liệu khuôn mặt chủ yếu là hình ảnh 2D, còn với các phương pháp hiện đại với sự phát triển của các mô hình học sâu có thể dễ dàng trích xuất đặc trưng từ hình ảnh khuôn mặt thì các phương pháp này có thể sử dụng thông tin hình ảnh phong phú như khuôn mặt 2D, khuôn mặt 3D, hình ảnh hồng ngoại. Dưới đây là một số ưu nhược điểm của các dữ liệu:

- Hình ảnh 2D:
  - Ưu điểm: Thiết bị thu thập hình ảnh 2D có giá thành rẻ, có nhiều cơ sở dữ liệu và thuật toán hỗ trợ, mô hình nhận diện khuôn mặt 2D dễ dàng sử dụng và tích hợp.
  - Nhược điểm: Hiệu suất nhận diện khuôn mặt bị giảm trong điều kiện ánh sáng kém, khi thay đổi tư thế khuôn mặt và biểu cảm khuôn mặt. Ngoài ra mô hình 2D khó chống lại việc gian lận trong nhận diện khuôn mặt.
- Hình ảnh 3D:
  - Ưu điểm: Hình ảnh khuôn mặt 3D cung cấp thêm thông tin chiều sâu của

khuôn mặt giúp cải thiện hiệu suất nhận diện, bên cạnh đó cũng khắc phục được một số điểm yếu của khuôn mặt 2D như bị tác động bởi ánh sáng và tăng cường việc phát hiện gian lận giả mạo khuôn mặt.

  - Nhược điểm: Thiết bị thu nhận hình ảnh 3D có giá thành cao, hệ thống nhận diện đòi hỏi phần cứng và phần mềm phức tạp và thời gian xử lý dữ liệu lâu hơn so với hệ thống sử dụng hình ảnh 2D.

- Hình ảnh hồng ngoại (IR):
  - Ưu điểm: Hiệu suất nhận diện tốt trong điều kiện ánh sáng yếu hoặc không có ánh sáng vì sử dụng bức xạ nhiệt của khuôn mặt. Hình ảnh hồng ngoại có thể phát hiện sự khác biệt về nhiệt độ và cấu trúc da, giúp giảm khả năng bị lừa bởi các hình ảnh hoặc mặt nạ giả mạo.
  - Nhược điểm: Thiết bị thu nhận hình ảnh hồng ngoại có chi phí đắt đỏ. Hình ảnh hồng ngoại có độ phân giải thấp và bị ảnh hưởng bởi nhiệt độ của môi trường dẫn đến giảm hiệu suất nhận diện

Trong nhận diện khuôn mặt, các mô hình học sâu như Convolutional Neural Network (CNN) đã chứng tỏ khả năng vượt trội trong việc trích xuất các đặc trưng phức tạp từ hình ảnh khuôn mặt so với các phương pháp truyền thống.

- DeepFace: Được ra đời vào năm 2014 là một đóng góp quan trọng của Facebook AI Research (FAIR). Mô hình này được phát triển bởi các nhà nghiên cứu của FAIR với mục đích áp dụng deep learning vào nhận dạng khuôn mặt. DeepFace sử dụng mô hình khuôn mặt để ánh xạ từng phần khuôn mặt vào một không gian sau đó sử dụng mạng học sâu CNN để trích xuất đặc trưng, từ đó so sánh và nhận diện các khuôn mặt.
- FaceNet: FaceNet là một hệ thống nhận dạng khuôn mặt được Google Research phát triển vào năm 2015. FaceNet ánh xạ hình ảnh khuôn mặt vào một không gian vector, sau đó sử dụng hàm Triplet Loss để tối ưu sao cho khoảng cách Euclid giữa các vector tương ứng với các khuôn mặt giống nhau là nhỏ, và giữa các khuôn mặt khác nhau là lớn.

Sự kết hợp giữa các mạng CNN và các hàm Loss đã giúp trích xuất các đặc trưng khuôn mặt một cách hiệu quả, giúp tăng hiệu suất nhận diện khuôn mặt. Tuy nhiên các mô hình sử dụng mạng học sâu cần đòi hỏi yêu cầu tài nguyên tính toán lớn và lượng dữ liệu huấn luyện phong phú, đa dạng.

1.3 Kết hợp nhận diện khuôn mặt 2D và 3D

Các mô hình nhận diện khuôn mặt 2D đạt được nhiều thành công nhờ vào sự phát triển của mạng học sâu nhưng vẫn còn các hạn chế về sự thay đổi góc nhìn, ánh sáng, và biểu cảm khuôn mặt. Trong khi đó, khuôn mặt 3D có thể cung cấp thêm các thông tin về hình dáng và cấu trúc của khuôn mặt giúp nhận diện chính xác trong các điều kiện hạn chế về ánh sáng và góc nhìn.

Các phương pháp kết hợp dữ liệu 2D và 3D trong nhận diện khuôn mặt có các chiến lược khác nhau để tối ưu hóa hiệu suất nhận diện:

1. Kết hợp cấp độ ban đầu: Sử dụng dữ liệu 2D và 3D từ giai đoạn ban đầu của quá trình nhận diện. Papatheodorou và Rueckert (2004) đã kết hợp tọa độ 3D với giá trị màu xám 2D tương ứng để tạo ra dữ liệu khuôn mặt 4D trong nhận diện khuôn mặt
2. Kết hợp cấp độ đặc trưng: Trích xuất các đặc trưng riêng biệt từ dữ liệu 2D và 3D, sau đó hợp nhất chúng thành vector đặc trưng. Werghi et al. (2015a) giới thiệu phương pháp Mesh-LBP để trích xuất đặc trưng hình dạng và kết cấu khuôn mặt, sau đó kết hợp để nhận diện dựa trên khoảng cách của các đặc trưng.
3. Kết hợp cấp độ cuối: Sử dụng hai hoặc nhiều bộ phân lớp khuôn mặt để tính toán điểm tương đồng của dữ liệu 2D và 3D, sau đó áp dụng các chiến lược kết hợp khác nhau để tổng hợp điểm tương đồng. Chang et al. (2005c) đã sử dụng các phương pháp dựa trên PCA cho hình ảnh độ sâu và hình ảnh màu, và hợp nhất các điểm tương đồng với một lược đồ trọng số.

Việc nghiên cứu mô hình nhận diện khuôn mặt sử dụng kết hợp khuôn mặt 2D và 3D sẽ tăng được độ chính xác của mô hình do trích xuất được nhiều các thông tin từ cả hai dữ liệu 2D và 3D. Bên cạnh đó còn giúp mô hình có tính bảo mật cao hơn do hệ thống chỉ sử dụng thông tin 2D dễ bị tấn công bởi ảnh giả mạo còn dữ liệu 3D với khả năng ghi lại chiều sâu và cấu trúc bề mặt, khó bị sao chép và giả mạo hơn.

1.4 Đóng góp của đồ án

Đồ án đự đoán việc sử dụng thông tin khuôn mặt 3D vào nhận diện khuôn mặt để cải thiện hiệu suất của mô hình nhận diện khuôn mặt. Đồ án này có 3 đóng góp chính như sau:

1. Tiền xử lý dữ liệu để trích xuất thông tin 3D từ các ảnh chụp 2D theo phương pháp Photometric Stereo.
2. Xây dựng mô hình khuôn mặt 2D, 3D, kết hợp 2D và 3D và huấn luyện các

mô hình theo hướng mạng phân loại và mạng Triplet.

3. Đánh giá hiệu suất của các mô hình.

1.5 Bố cục đồ án

Phần còn lại của báo cáo đồ án tốt nghiệp này được tổ chức như sau.

Trong chương 2, đồ án trình bày về ngữ cảnh của bài toán nhận diện khuôn mặt cũng như nền tảng lý thuyết của các phương pháp nhận diện khuôn mặt. Các phương pháp truyền thống như PCA, LDA, LBP dễ triển khai, tính toán nhanh tuy nhiên còn gặp khó khăn trong các điều kiện hạn chế về ánh sáng, tư thế. Cùng với sự phát triển của các mạng học sâu, các phương pháp nhận diện khuôn mặt hiện đại đã sử dụng mạng học sâu có được kết quả cải thiện rất nhiều so với phương pháp truyền thống. Việc sử dụng linh hoạt các mạng CNN cùng các hàm loss đã huấn luyện ra các mô hình nhận diện khuôn mặt có hiệu suất cao dù trong điều kiện hạn chế.

Tiếp theo đó trong chương 3 sẽ trình bày về chi tiết phương pháp đề xuất của đồ án. Đầu tiên trình bày về bộ dữ liệu Photoface Database được phát hành vào tháng 6 năm 2011, là một tập hợp dữ liệu hình ảnh khuôn mặt bao gồm 3174 phiên từ 453 người khác nhau. Mỗi phiên chứa 4 hình ảnh BMP của người đó dưới các điều kiện chiếu sáng khác nhau cùng với thông tin chi tiết về vị trí góc chiếu sáng cũng như tọa độ của 11 đặc điểm khuôn mặt và thông tin bổ sung về giới tính, kính đeo, tư thế, che khuất, cảm xúc và các điều kiện khác. Tiếp theo trình bày về phương pháp Photometric Stereo sử dụng trong việc tiền xử lý dữ liệu để tạo ra thông tin về ảnh 3D từ các ảnh 2D trong mỗi phiên ảnh. Đây là một kỹ thuật trong thị giác máy tính để ước tính bề mặt của vật thể bằng cách quan sát vật thể đó trong các điều kiện ánh sáng khác nhau. Kế tiếp chương này trình bày về xây dựng 3 mô hình nhận diện khuôn mặt 2D, 3D và kết hợp 2D với 3D với backbone chính là InceptionResNetV1 cùng với phương pháp huấn luyện như theo 2 hướng là phân loại(Classification) với hàm mất mát Cross Entropy Loss và mạng Triplet với hàm mất mát Triplet Loss.

Sau khi đã xây dựng và huấn luyện mô hình nhận diện khuôn mặt, chương 4 sẽ trình bày về các chỉ số đánh giá mô hình và kết quả đánh giá của các mô hình với bộ dữ liệu kiểm thử. Sau khi đưa dữ liệu khuôn mặt vào mô hình sẽ trả về kết quả là một vector, sau đó dựa trên khoảng cách euclid hoặc độ tương đồng cosine của vector để đánh giá hai khuôn mặt có cùng là một người hay không. Đường cong ROC biểu diễn mối quan hệ giữa tỷ lệ dương tính giả (False Positive Rate - FPR) và tỷ lệ dương tính thật (True Positive Rate - TPR) khi ngưỡng phân loại thay đổi. Chỉ số AUC là diện tích dưới đường cong ROC, phản ánh khả năng phân biệt giữa

các khuôn mặt cùng hay khác người (một dạng của bài toán phân loại nhị phân). Chỉ số này được chọn để đánh giá vì nó sẽ đánh giá qua tất cả các ngưỡng và cung cấp cái nhìn tổng thể về khả năng phân loại của mô hình bên cạnh đó còn không bị ảnh hưởng bởi sự mất cân bằng, phản ánh chính xác hiệu suất mô hình.

Chương 5 đưa ra các kết luận thu được từ đồ án và tổng kết các vấn đề mà đồ án đã giải quyết cũng như các vấn đề còn tồn đọng. Sau khi có kết quả thực nghiệm thì đã chứng minh được việc sử dụng kết hợp thông tin khuôn mặt 2D và 3D giúp cải thiện tốt hơn cho hiệu suất nhận diện khuôn mặt. Tiếp theo đó đưa ra các định hướng phát triển của đồ án trong tương lai.

CHƯƠNG 2. NỀN TẢNG LÝ THUYẾT

Trong chương này sẽ đi sâu vào ngữ cảnh và nền tảng lý thuyết của bài toán nhận diện khuôn mặt. Đầu tiên, đồ án sẽ giới thiệu về ngữ cảnh của bài toán nhận diện khuôn mặt. Tiếp theo trình các phương pháp truyền thống đã được sử dụng trước sự xuất hiện của mạng học sâu, bao gồm PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis) và LBP (Local Binary Patterns). Sau đó sẽ trình chuyển sang các phương pháp hiện đại, đặc biệt là những đột phá nhờ sự phát triển của mạng học sâu (Deep Neural Networks) như mô hình DeepFace của Facebook và FaceNet của Google. Thông qua việc tìm hiểu các phương pháp và kỹ thuật này, chúng ta sẽ có cái nhìn toàn diện về những thách thức và tiến bộ trong lĩnh vực nhận diện khuôn mặt, từ đó làm nền tảng cho các nghiên cứu và phát triển tiếp theo trong đồ án.

2.1 Ngữ cảnh của bài toán

Nhận diện khuôn mặt là một trong những ứng dụng quan trọng và phổ biến trong lĩnh vực thị giác máy tính (Computer Vision). Công nghệ này giúp chúng ta xác định hoặc nhận dạng danh tính của một người dựa trên đặc điểm sinh trắc học về khuôn mặt từ hình ảnh hoặc video. Phương pháp nhận diện sinh trắc học bằng khuôn mặt là một phương pháp có tính thuận tiện và khả năng tích hợp, ứng dụng cao trong cuộc sống.

2.1.1 Tính quan trọng của nhận diện khuôn mặt

Nhận diện khuôn mặt không chỉ là một công nghệ tiên tiến mà còn mang lại nhiều giá trị thiết thực trong cuộc sống hàng ngày. Dưới đây là một số ứng dụng mà công nghệ nhận diện khuôn mặt đem lại:

- Bảo mật: Nhận diện khuôn mặt được sử dụng rộng rãi trong các hệ thống bảo mật, truy cập giúp xác định các cá nhân được phép truy cập sử dụng hệ thống hoặc cảnh báo khi phát hiện những người không được phép truy cập.
- An ninh: Hệ thống camera giám sát tích hợp nhận diện khuôn mặt có thể tự động nhận dạng và theo dõi các cá nhân trong các khu vực công cộng giúp kiểm soát tốt các tình huống xảy ra.
- Giải trí: Trên các nền tảng mạng xã hội như Facebook, nhận diện khuôn mặt giúp tự động gắn thẻ người dùng trong ảnh và video, cải thiện trải nghiệm người dùng và quản lý nội dung hiệu quả hơn.
- Marketing: Các hệ thống nhận diện khuôn mặt giúp phân tích hành vi khách hàng, quảng cáo và cung cấp trải nghiệm mua sắm tốt hơn thông qua việc theo

dõi khách hàng.

2.1.2 Những thách thức trong nhận diện khuôn mặt

Mặc dù nhận diện khuôn mặt đạt được nhiều tiến bộ và ứng dụng thành công trong nhiều lĩnh vực, nhưng công nghệ này vẫn đối mặt với nhiều thách thức cần được giải quyết để cải thiện độ chính xác và khả năng ứng dụng trong thực tế. Dưới đây là một số thách thức chính trong nhận diện khuôn mặt:

- Biến đổi về góc nhìn: Khuôn mặt của một người có thể trông rất khác nhau khi được chụp từ các góc độ khác nhau. Sự biến đổi về góc nhìn có thể làm giảm độ chính xác của hệ thống nhận diện khuôn mặt, đặc biệt là khi góc nhìn lệch quá nhiều so với hình ảnh gốc trong cơ sở dữ liệu ban đầu.
- Sự thay đổi về ánh sáng: Điều kiện ánh sáng khi nhận diện có thể thay đổi đáng kể đặc trưng khuôn mặt, gây ra bóng đổ và làm mờ các đặc điểm khuôn mặt. Nhận diện khuôn mặt trong các nơi có điều kiện ánh sáng kém hoặc ánh sáng thay đổi liên tục là một thách thức lớn.
- Biểu cảm khuôn mặt: Biểu cảm khuôn mặt có thể làm thay đổi hình dạng của các vùng đặc trưng của khuôn mặt (như mắt, miệng, ...), làm cho việc nhận diện trở nên khó khăn hơn. Các hệ thống nhận diện khuôn mặt cần phải linh hoạt để nhận diện chính xác khuôn mặc dù người đó đang cười, nhăn mặt hay có bất kỳ biểu cảm nào khác.
- Sự che khuất: Việc đeo kính, mũ, khẩu trang hoặc các vật dụng khác có thể che khuất một phần khuôn mặt, làm giảm độ chính xác của hệ thống. Điều này đặc biệt quan trọng trong bối cảnh hiện nay khi việc đeo khẩu trang trở nên phổ biến từ đại dịch COVID-19.
- Gian lận và giả mạo: Các hệ thống nhận diện khuôn mặt cần phải chống lại các hình thức gian lận và giả mạo như sử dụng ảnh in, video hoặc mặt nạ để đánh lừa hệ thống. Điều này đòi hỏi các biện pháp xác thực bổ sung như phát hiện chống giả mạo.

2.1.3 Lịch sử nghiên cứu nhận diện khuôn mặt

Các nghiên cứu về nhận diện khuôn mặt có thể chia làm hai phần chính là các phương pháp nhận diện khuôn mặt truyền thống (sử dụng các thuật toán chưa áp dụng mô hình học sâu) và các phương pháp nhận diện khuôn mặt hiện đại (sử dụng các mô hình học sâu)

- Phương pháp Eigenface PCA được giới thiệu lần đầu tiên vào đầu những năm 1991 bởi Matthew Turk và Alex Pentland tại Học viện Công nghệ Massachusetts (MIT). Công trình của họ đã công bố một phương pháp mới trong nhận dạng

khuôn mặt dựa trên kỹ thuật phân tích thành phần chính (PCA). Bài báo của họ, "Eigenfaces for Recognition" đã trở thành một cột mốc quan trọng trong lĩnh vực nhận dạng khuôn mặt, cung cấp nền tảng cho nhiều nghiên cứu và ứng dụng sau này.

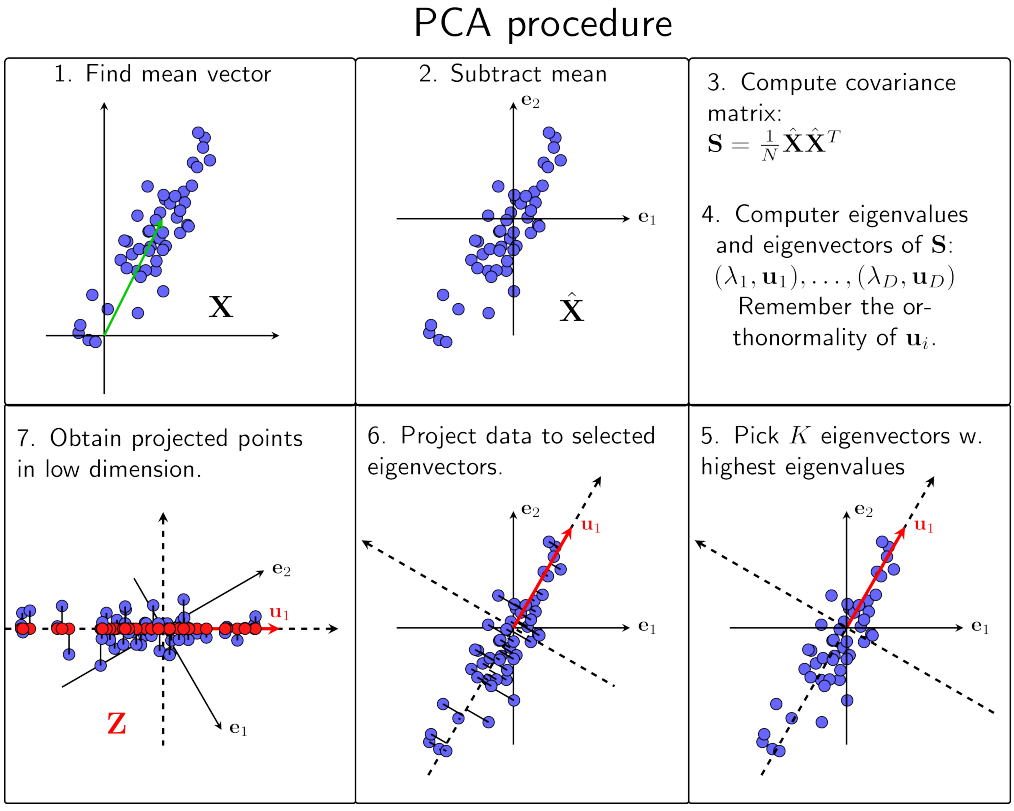
- Phương pháp Local Binary Patterns (LBP) trong nhận diện khuôn mặt được giới thiệu lần đầu vào năm 1996 bởi Timo Ojala, Matti Pietikäinen và David Harwood tại Đại học Oulu, Phần Lan. Công trình nghiên cứu của họ đã đưa ra ý tưởng cơ bản và các ứng dụng ban đầu của LBP trong việc mô tả và phân tích các đặc trưng cục bộ của hình ảnh khuôn mặt. Kể từ đó, LBP đã trở thành một trong những kỹ thuật quan trọng và được sử dụng rộng rãi trong lĩnh vực thị giác máy tính, đặc biệt là trong các ứng dụng nhận dạng và phân tích hình ảnh.
- Phương pháp Fisherfaces ra đời từ năm 1997, khi Peter N. Belhumeur, Joao P. Hespanha, và David J. Kriegman giới thiệu phương pháp này trong bài báo "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection". Bài báo này đưa ra Fisherfaces như một cải tiến so với Eigenfaces bằng cách sử dụng Linear Discriminant Analysis (LDA) để cải thiện khả năng nhận biết các khuôn mặt giống và khác nhau. Phương pháp Fisherfaces đã trở thành một trong những kỹ thuật quan trọng trong lĩnh vực thị giác máy tính và nhận dạng khuôn mặt.
- DeepFace được ra đời vào năm 2014 là một đóng góp quan trọng của Facebook AI Research (FAIR). Mô hình này được phát triển bởi các nhà nghiên cứu của FAIR với mục đích áp dụng deep learning vào nhận dạng khuôn mặt. Mô hình đạt được thành công với khả năng nhận diện khuôn mặt vượt trội và độ chính xác cao. Chính nó cũng đã đặt nền móng cho việc áp dụng deep learning trong lĩnh vực nhận dạng khuôn mặt và có ảnh hưởng lớn đối với các nghiên cứu và ứng dụng sau này.
- FaceNet là một hệ thống nhận dạng khuôn mặt được Google Research phát triển vào năm 2015. Mô hình này sử dụng mạng neural network để học và biểu diễn các đặc trưng khuôn mặt trong không gian đặc trưng, kết hợp cùng với hàm mất mát Triplet Loss giúp biểu diễn của các khuôn mặt cùng một người sẽ gần nhau hơn so với khuôn mặt khác người. FaceNet có khả năng nhận diện khuôn mặt với độ chính xác cao, ngay cả trong các điều kiện khác nhau về ánh sáng, góc chụp và biểu cảm khuôn mặt.

2.2 Các phương pháp nhận diện khuôn mặt truyền thống

Các phương pháp nhận diện khuôn mặt truyền thống đã từng được áp dụng rộng rãi trước khi công nghệ hiện đại và deep learning trở nên phổ biến. Các phương pháp truyền thống đơn giản và dễ dàng triển khai cũng như không đòi hỏi quá cao về bộ dữ liệu. Một số phương pháp đã đạt được nhiều thành công và đạt nền tảng phát triển cho nhận diện khuôn mặt như Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) và Local Binary Patterns (LBP). Dưới đây là trình bày về nền tảng lý thuyết chính của các phương pháp trên.

2.2.1 Principal Component Analysis

Principal Component Analysis (PCA) là một phương pháp thống kê được sử dụng để giảm chiều dữ liệu bằng cách biến đổi các dữ liệu gốc sang một không gian mới thu được các thành phần chính là tổ hợp tuyến tính của dữ liệu gốc. Các thành phần chính này được chọn sao cho chúng giữ lại được nhiều nhất phương sai của dữ liệu gốc. Nguyên lý hoạt động cơ bản của PCA (hình 2.1) bao gồm các bước sau:



Hình 2.1: Phương pháp Principal Component Analysis (PCA)

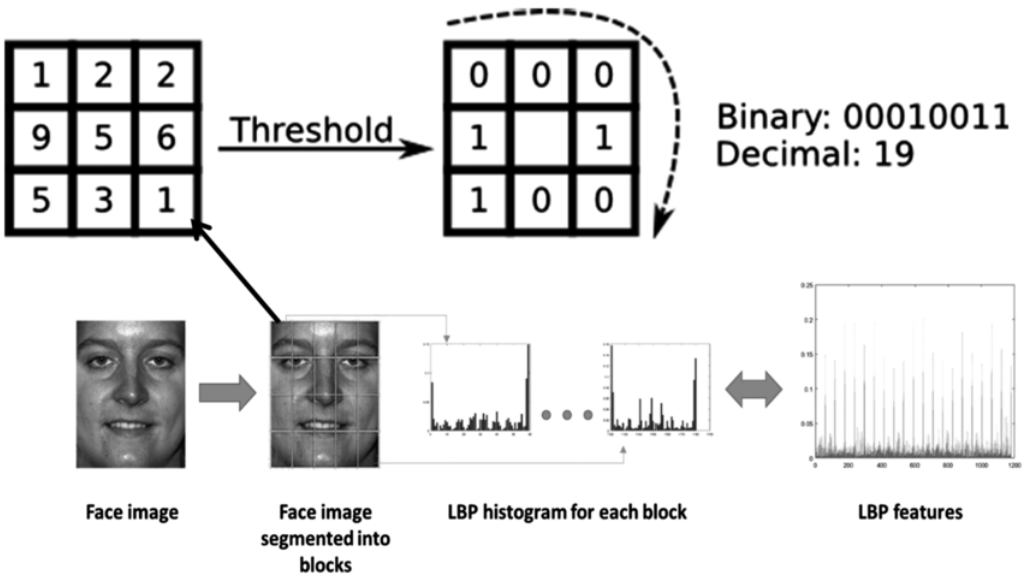
1. Chuẩn hóa dữ liệu: Tính hình ảnh khuôn mặt trung bình từ tập dữ liệu. Sau đó mỗi hình ảnh trong tập dữ liệu sẽ được trừ đi hình ảnh trung bình này để thu các vector của từng khuôn mặt.

2. Tính ma trận hiệp phương sai: Tính ma trận hiệp phương sai từ các vector thu được ở bước trên.
3. Tính Eigenfaces: Tính toán các vector riêng và giá trị riêng của ma trận hiệp phương sai. Các vector riêng này đại diện cho các trục chính của không gian dữ liệu mới, trong khi các trị riêng biểu diễn độ lớn của phương sai theo các trục đó.
4. Chọn trục chính: Chọn ra các trục chính có trị riêng lớn nhất, vì chúng biểu diễn phần lớn nhất của phương sai trong dữ liệu.
5. Chiếu dữ liệu: Cuối cùng, chiếu dữ liệu gốc lên không gian các trục chính đã chọn. Điều này giúp giảm chiều dữ liệu, trong khi vẫn giữ lại các thông tin đặc trưng quan trọng của khuôn mặt.

Principal Component Analysis (PCA) là một phương pháp giảm chiều dữ liệu hiệu quả, giúp giữ lại các đặc trưng quan trọng nhất và loại bỏ nhiễu. PCA dễ hiểu và triển khai, nhưng không phù hợp với dữ liệu phi tuyến và có thể dẫn đến mất mát một số thông tin quan trọng.

2.2.2 Local Binary Patterns

Local Binary Patterns (LBP) là một kỹ thuật mô tả kết cấu của hình ảnh bằng cách xem xét các mối quan hệ cục bộ giữa các điểm ảnh. Ý tưởng chính của LBP là sử dụng các mẫu nhị phân để biểu diễn thông tin về kết cấu xung quanh một điểm ảnh cụ thể trong hình ảnh. Phương pháp LBP hoạt động (hình 2.2) theo các bước:



Hình 2.2: Minh họa phương pháp Local Binary Patterns (LBP)

1. Chọn một điểm ảnh trung tâm: Đối với mỗi điểm ảnh trong hình ảnh (trừ các điểm ảnh ở biên), chọn nó làm điểm ảnh trung tâm.

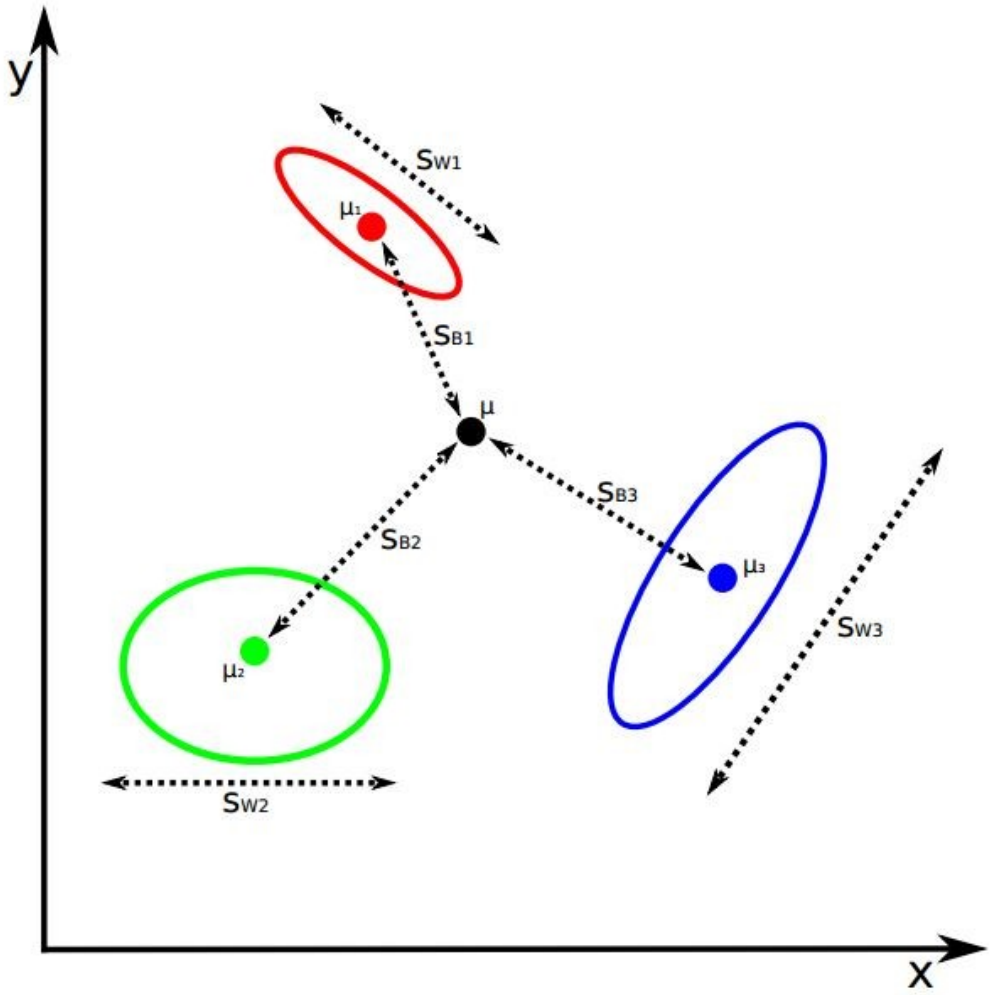
2. So sánh với các điểm ảnh lân cận: So sánh điểm ảnh trung tâm với các điểm ảnh lân cận theo một mẫu cố định (thường là 3x3). Nếu giá trị của điểm ảnh lân cận lớn hơn hoặc bằng giá trị của điểm ảnh trung tâm, gán giá trị 1, ngược lại gán giá trị 0.
3. Tạo mẫu nhị phân: Tạo một số nhị phân từ các giá trị 0 và 1 thu được từ bước 2, đọc theo chiều kim đồng hồ hoặc ngược chiều kim đồng hồ. Ví dụ, đối với một mẫu 3x3, ta có thể tạo ra một số nhị phân 8 bit.
4. Chuyển đổi số nhị phân thành số thập phân: Chuyển đổi số nhị phân vừa tạo thành một số thập phân. Số này là giá trị LBP của điểm ảnh trung tâm.
5. Xây dựng histogram LBP: Lặp lại các bước 1-4 cho tất cả các điểm ảnh trong hình ảnh và xây dựng một histogram từ các giá trị LBP thu được. Histogram này biểu diễn tần suất xuất hiện của mỗi mẫu nhị phân trong toàn bộ hình ảnh và được sử dụng như là một đặc trưng của hình ảnh.

Local Binary Patterns (LBP) là phương pháp đơn giản nhưng hiệu quả trong việc trích xuất các đặc trưng cục bộ từ hình ảnh, có khả năng kháng nhiễu tốt và chi phí tính toán thấp. Tuy nhiên, LBP không nắm bắt được thông tin toàn cục, nhạy cảm với các biến đổi nhỏ và nhiễu cục bộ.

2.2.3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) là một kỹ thuật giảm chiều dữ liệu tương tự như PCA, nhưng thay vì tối đa hóa phương sai toàn cục của dữ liệu, LDA tối đa hóa tỷ số giữa phương sai giữa các khuôn mặt khác nhau và phương sai của khuôn mặt cùng người (hình 2.3). Nguyên lý hoạt động cơ bản của LDA bao gồm các bước sau:

1. Chuẩn hóa dữ liệu: Tính vector trung bình của mỗi lớp trong tập dữ liệu. Tính vector trung bình của toàn bộ tập dữ liệu.
2. Tính ma trận within-class scatter (S_W): Ma trận này biểu diễn sự phân tán của các dữ liệu trong một lớp so với vector trung bình của lớp này.
3. Tính ma trận between-class scatter (S_B): Ma trận này biểu diễn sự phân tán của các các vector trung bình của các lớp so với vector trung bình tổng thể của tập dữ liệu.
4. Tìm vector chiếu: Tìm vector chiếu sau sao tỷ lệ giữa S_B và S_W được tối ưu hóa. Vector này chính là hướng chiếu mà tối ưu hóa sự phân biệt giữa các lớp.
5. Chiếu dữ liệu: Sử dụng vector chiếu để chiếu các điểm dữ liệu từ không gian gốc xuống không gian mới.



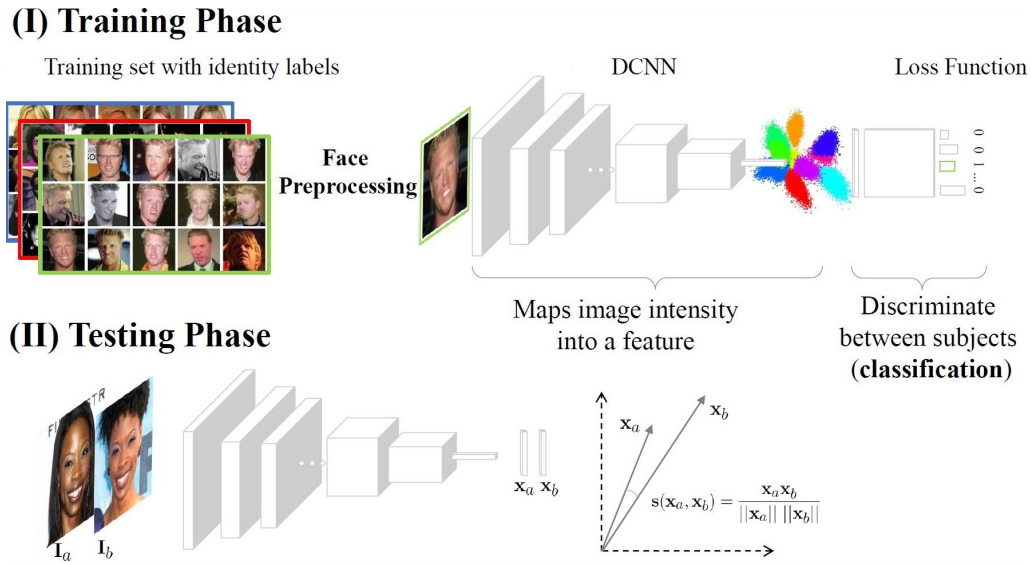
Hình 2.3: Nguyên lý phương pháp Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) tối đa hóa khoảng cách giữa các lớp và giảm thiểu sự phân tán trong cùng một lớp giúp tăng cường khả năng phân biệt giữa các khuôn mặt khác nhau. LDA giả định rằng các lớp dữ liệu có phân phối Gaussian và có cùng ma trận hiệp phương sai. Điều này thường không phù hợp với các tập dữ liệu khuôn mặt thực tế vì chúng không tuân theo phân phối này.

2.3 Các phương pháp nhận diện khuôn mặt hiện đại

Các phương pháp nhận diện khuôn mặt truyền thống thường dựa vào việc rút trích các đặc trưng như cạnh, góc mắt, hoặc tỉ lệ các phần tử trên khuôn mặt để nhận diện và chịu ảnh hưởng thường lớn từ các điều kiện ánh sáng và góc độ khác nhau. Trong khi đó, các phương pháp sử dụng học sâu như mạng tích chập (CNN) đã mang đến bước tiến lớn trong nhận diện khuôn mặt. CNN có khả năng học và rút trích các đặc trưng từ dữ liệu mà không cần đến sự can thiệp của con người. Nhờ vào khả năng học từ dữ liệu lớn, các mô hình học sâu như CNN có thể tự động hóa quá trình nhận diện khuôn mặt một cách hiệu quả hơn, giảm thiểu sự phụ thuộc vào các đặc trưng được chọn và tăng cường độ chính xác đáng kể trong các điều

kiện biến đổi và đa dạng. Một số mô hình như DeepFace của Facebook và FaceNet của Google đã đạt đạt được nhiều thành công khi ra đời nhờ khả năng nhận diện chính xác và hiệu quả dù trong các điều kiện hạn chế.



Hình 2.4: Luồng hoạt động của các phương pháp sử dụng mạng học sâu

Luồng hoạt động chung của các mô hình mạng học sâu trong nhận diện khuôn mặt (hình 2.4):

1. Quá trình huấn luyện: Sử dụng bộ dữ liệu được gán nhãn để học các trọng số trong mạng DCNN để tối ưu hàm mất mát trong bài toán phân loại.
2. Quá trình kiểm thử: Sử dụng mạng DCNN đã học để trích xuất vector đặc trưng khuôn mặt rồi so sánh chúng (có thể sử dụng khoảng cách euclid hoặc độ tương đồng cosine) để trả về kết quả nhận diện khuôn mặt.

2.3.1 DeepFace

DeepFace là một dự án nghiên cứu của Facebook được công bố vào năm 2014, nhằm mục đích nghiên cứu và phát triển các giải pháp nhận diện khuôn mặt sử dụng công nghệ deep learning. Đây là một trong những bước đầu tiên và đáng chú ý trong việc áp dụng deep learning vào lĩnh vực nhận diện khuôn mặt, và đã góp phần đáng kể vào sự phát triển trong lĩnh vực này. Mô hình mạng học sâu mà DeepFace sử dụng:

1. Sử dụng mạng neural bao gồm các lớp sau: convolutional layer - max pooling - convolutional layer - 3 locally connected layers - fully connected layer.
2. Đầu vào của mô hình là hình ảnh RGB của khuôn mặt, được scale về độ phân giải 152x152.
3. Trong quá trình huấn bài toán phân loại như nhận diện khuôn mặt, mục tiêu

chính là tối đa xác suất dự đoán đúng lớp (face ID) cho mỗi ảnh đầu vào. Điều này được thực hiện thông qua việc giảm thiểu hàm mất mát Cross Entropy Loss cho mỗi mẫu huấn luyện.

4. Mô hình học ra một vector thực có kích thước 4096, là vector đặc trưng của ảnh khuôn mặt. Vector đặc trưng này được xử lý tiếp để nhận diện khuôn mặt bằng cách so sánh với danh sách các vector đặc trưng của khuôn mặt đã biết.

2.3.2 FaceNet

FaceNet là một hệ thống nhận dạng khuôn mặt sử dụng deep learning, được phát triển bởi Google Research vào năm 2015. FaceNet sử dụng một mạng neural tích chập (CNN) và hàm mất mát Triplet Loss để học cách biểu diễn của các khuôn mặt. Mạng học sâu này nhận đầu vào là hình ảnh khuôn mặt ở kích thước chuẩn và đưa ra một vector đặc trưng (feature vector) có số chiều cố định. Mô hình mạng học sâu mà FaceNet sử dụng:

1. Facenet sử dụng một backbone được gọi là Inception-ResNet v1. Đây là một kiến trúc mạng nơ-ron sâu kết hợp giữa Inception và ResNet, được thiết kế đặc biệt để trích xuất đặc trưng từ ảnh khuôn mặt.
2. FaceNet đã có cải tiến quan trọng đó chính là sử dụng Triplet Network (cải tiến của mạng Siamese Network). Kiến trúc này có ba nhánh đầu vào chia sẻ cùng một mạng CNN. Đầu ra của CNN là các vector biểu diễn đặc trưng của từng ảnh khuôn mặt.
3. Để đảm bảo học tốt vector biểu diễn của các khuôn mặt FaceNet sử dụng hàm Triplet Loss. Mục tiêu của hàm mất mát này tối ưu khoảng cách giữa các vector biểu diễn của các ảnh khuôn mặt cùng một người là nhỏ, đồng thời khoảng cách giữa các vector biểu diễn của các ảnh khuôn mặt của các người khác nhau là lớn.

2.4 Mô hình Inception-ResNet và các hàm mất mát

Với sự phát triển của các mô hình học sâu cùng các hàm mất mát đã giúp trích xuất các đặc trưng khuôn mặt một cách hiệu quả, tăng hiệu suất nhận diện khuôn mặt. Dưới đây là mô hình kiến trúc InceptionResNetv1 và hai hàm mất mát Cross Entropy Loss và Triplet Loss sẽ đưuọc sử dụng trong đồ án.

2.4.1 Kiến trúc của InceptionResNetv1

InceptionResNetv1 là một trong những mô hình mạng học sâu tiên tiến kết hợp hai kiến trúc mạng nổi tiếng: Inception và ResNet. Mô hình này được thiết kế để tận dụng ưu điểm của cả hai kiến trúc nhằm cải thiện hiệu suất và độ chính xác trong các tác vụ nhận diện và phân loại ảnh.

1. Inception Modules
    - Inception Modules trong mạng Inception sử dụng các bộ lọc kích thước khác nhau (1x1, 3x3, 5x5) trong cùng một lớp để rút trích nhiều loại đặc trưng từ ảnh đầu vào. Điều này giúp mô hình có khả năng học được nhiều đặc trưng đa dạng hơn.
    - Reduction Modules là các module đặc biệt được thiết kế để giảm kích thước của đặc trưng mà không làm mất nhiều thông tin quan trọng, giúp giảm yêu cầu về tính toán và bộ nhớ.
2. Residual Connections
    - Residual Connections trong ResNet giúp giải quyết vấn đề gradient vanishing (mất mát gradient) khi mạng trở nên quá sâu. Các kết nối dư này giúp truyền gradient trực tiếp qua nhiều lớp, giúp mô hình học hiệu quả hơn.
    - Các shortcut connections này cho phép các lớp phía trước truyền trực tiếp thông tin tới các lớp phía sau, giúp mạng có thể học được các hàm ánh xạ phức tạp hơn.
3. Kết hợp Inception và ResNet
    - InceptionResNetv1 kết hợp các Inception Modules với Residual Connections, tạo ra một mạng học sâu có khả năng học được các đặc trưng phong phú và phức tạp, đồng thời giảm thiểu vấn đề gradient vanishing.
    - Các module Inception được sửa đổi để có thể thêm các kết nối dư, làm cho việc huấn luyện mạng trở nên ổn định và nhanh hơn.

InceptionResNetv1 được thiết kế để tối ưu hóa hiệu suất trên các tác vụ nhận diện ảnh, cũng như tối ưu hóa yêu cầu về tính toán giúp nó trở thành lựa chọn lý tưởng cho các ứng dụng thực tế yêu cầu tốc độ và hiệu suất cao.

2.4.2 Hàm Cross Entropy Loss

Cross Entropy Loss là một trong những hàm mất mát phổ biến nhất được sử dụng trong các bài toán phân loại, đặc biệt là trong các mạng neural sâu. Hàm này đo lường sự khác biệt giữa phân phối dự đoán của mô hình và phân phối mục tiêu thực tế.

Cho một bài toán phân loại với n lớp, Cross Entropy Loss được định nghĩa như sau:

$$\mathcal{L} = -\sum_{i=1}^{n} y_i \log(p_i)$$

Trong đó:

- y_i là nhãn thực tế cho lớp i.
- p_i là xác suất dự đoán cho lớp i.

Trong khi các hàm mất mát khác phạt dựa trên các giá trị sai thì Cross Entropy Loss phạt mô hình dựa trên cả độ chính xác và mức độ chắc chắn của dự đoán. Ví dụ, nếu mô hình dự đoán xác suất 0.6 cho nhãn đúng thay vì 0.9, nó sẽ bị phạt vì không tự tin vào dự đoán của mình.

2.4.3 Hàm Triplet Loss

Triplet Loss là một hàm mất mát đặc biệt được sử dụng trong Triplet Network để học các vector biểu diễn (embedding vectors) sao cho các điểm dữ liệu tương tự nhau nằm gần nhau trong không gian vector và các điểm khác nhau thì nằm xa nhau.

Hình 2.5: Hàm mất mát Triplet Loss

Trong hình 2.5 một bộ triplet bao gồm ba mẫu: một $anchor(A)$, một $positive(P)$ và một $negative(N)$ thì Triplet Loss được tính theo:

$$\mathcal{L}(A, P, N) = \max\left(\|f(A) - f(P)\|_2^2 - \|f(A) - f(N)\|_2^2 + \alpha, 0\right)$$

Trong đó:

- $f(x)$ là hàm biểu diễn (embedding function) của mẫu x.
- $\|\cdot\|_2$ là khoảng cách Euclidean (L2 norm).
- α là biên độ ($margin$), một giá trị không âm để đảm bảo rằng khoảng cách giữa $anchor$ và $negative$ lớn hơn một giá trị nhất định so với khoảng cách giữa $anchor$ và $positive$.
- $\max(\cdot, 0)$ là hàm ReLU để đảm bảo rằng Triplet Loss luôn không âm.

Khi Triplet Loss được tối ưu hóa tốt, nó sẽ tạo ra một không gian embedding trong đó các ảnh của cùng một người được nhóm lại với nhau và cách xa các nhóm

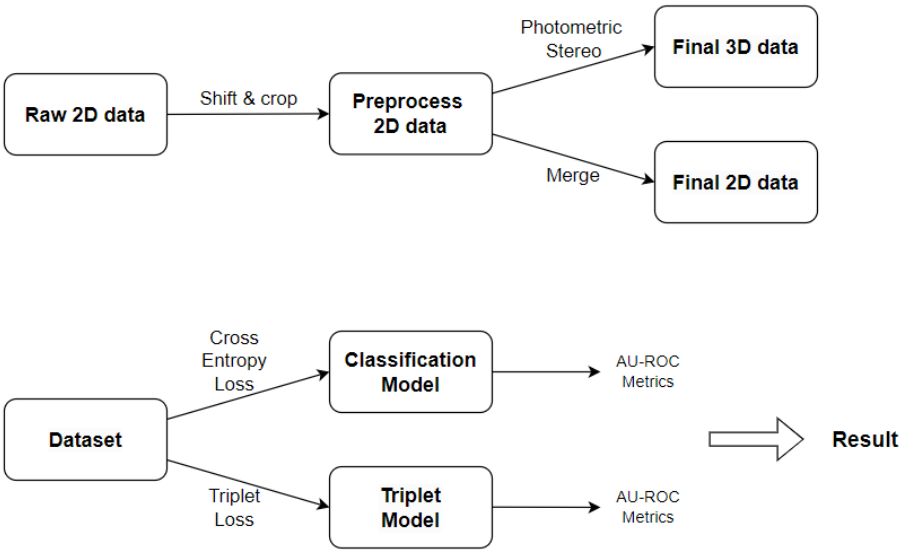
của những người khác, giúp cải thiện đáng kể hiệu suất của các hệ thống nhận diện và xác minh khuôn mặt.

CHƯƠNG 3. PHƯƠNG PHÁP ĐỀ XUẤT

Chương 2 đã trình bày ngữ cảnh của bài toán nhận diện khuôn mặt, các phương pháp đã được phát triển trong lịch sử lĩnh vực này cũng như các cơ sở lý thuyết về một số kiến thức nền tảng được sử dụng trong đồ án. Chương này sẽ trình bày tổng quan về luồng hoạt động của phương án đề xuất và từng phần chi tiết bên trong luồng hoạt động.

3.1 Tổng quan giải pháp

Mô hình đề xuất của đồ án sử dụng bộ dữ liệu Photoface Database được phát hành vào tháng 6 năm 2011. Bộ dữ liệu sẽ được tiền xử lý để thu được bộ dữ liệu khuôn mặt 2D và bộ dữ liệu khuôn mặt 3D. Sau đó bộ dữ liệu được sử dụng để huấn luyện các mô hình nhận diện khuôn mặt, cuối cùng sẽ được đánh giá để kiểm nghiệm hiệu suất của các mô hình.



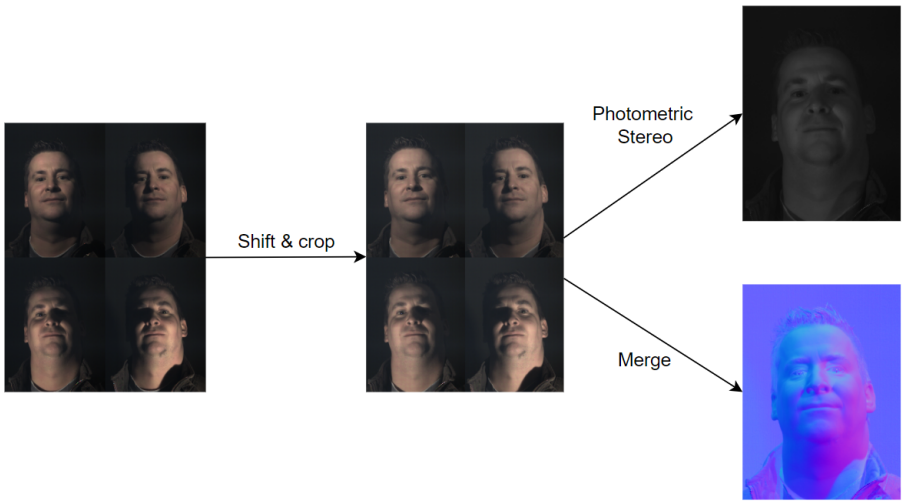
Hình 3.1: Luồng hoạt động của phương pháp đề xuất

Hình 3.1 mô tả các đóng góp chính của giải pháp trong đồ án nhận diện khuôn mặt. Luồng hoạt động có thể chia thành 2 phần chính đó là bước tiền xử lý dữ liệu và huấn luyện mô hình.

3.2 Tiền xử lý dữ liệu

Phần này trình bày về bộ dữ liệu được sử dụng trong đồ án, các bước tiền xử lý để thu được dữ liệu phục vụ cho việc huấn luyện mô hình (hình 3.2). Đầu tiên mỗi phiên ảnh trong bộ dữ liệu được xử lý để dịch chuyển về vị trí phù hợp do khi chụp ảnh các khuôn mặt tại thời điểm khác nhau nên vị trí của khuôn mặt có khả năng bị lệch một vài pixel. Chúng ta cần phải xử lý dịch chúng lại cho đúng vị trí để có thể

áp dụng hiệu quả phương pháp Photometric Stereo. Do vị trí của các nguồn sáng khác nhau nên bốn ảnh 2D trong một phiên chụp có điều kiện ánh sáng khác nhau nên cũng cần xử lý để đưa ra một ảnh 2D tốt nhất. Sau đó từ bộ dữ liệu 2D đã qua xử lý bước đầu chúng ta sẽ tạo bộ dữ liệu huấn luyện 2D và 3D. Dưới đây là trình bài chi tiết về bộ dữ liệu và các bước tiền xử lý.



Hình 3.2: Tiền xử lý dữ liệu

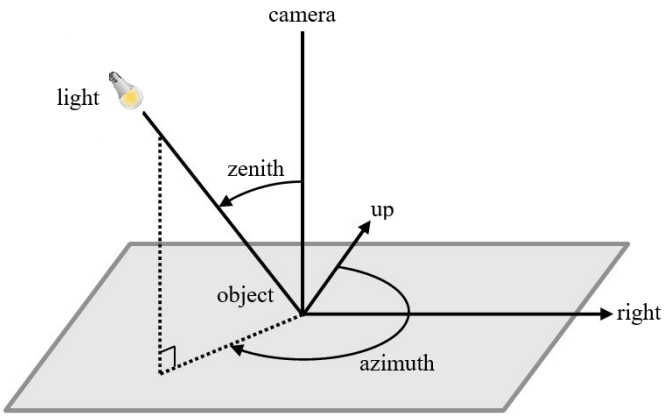
3.2.1 Bộ dữ liệu

Bộ dữ liệu Photoface Database bao gồm 3174 phiên của 453 người khác nhau. Bộ dữ liệu chia thành các thư mục là id của từng người, bên trong mỗi thư mục id là các phiên chụp tại các thời điểm khác nhau. Trong mỗi thư mục phiên có:

- 4 hình ảnh bmp thô của khuôn mặt được chụp từ hệ thống thu thập khuôn mặt ví dụ như trong hình 3.3. Những hình ảnh này chụp dưới ánh sáng khác nhau, các thông tin về góc chiếu sáng được ghi lại trong tệp LightSource.m.

Hình 3.3: Ví dụ 4 ảnh gốc ban đầu trong một phiên chụp ảnh

- LightSource.m cung cấp các góc zenith và azimuth của bốn nguồn sáng cho phiên cụ thể như trong hình 3.4.



Hình 3.4: Thông tin về nguồn sáng

- metadata.txt chứa tọa độ x và y của 11 điểm đặc trưng trên khuôn mặt bao gồm: khóe mắt ngoài bên trái, khóe mắt trong bên trái, điểm giữa trán, khóe mắt trong bên phải, khóe mắt ngoài bên phải, bên trái mũi, đầu mũi, bên phải mũi, góc miệng trái, góc miệng phải và điểm giữa cằm.
- metadataII.txt chứa thêm thông tin metadata: giới tính, kính đeo, râu, tư thế, chất lượng, che khuất, cảm xúc, thông tin khác.

3.2.2 Dịch ảnh khuôn mặt về vị trí phù hợp

Ban đầu 4 hình ảnh khuôn mặt mới ở vị chính tương đối với nhau do các thời điểm chụp khác nhau nên khuôn mặt có bị xê dịch dịch một ít. Nhiệm vụ của bước này là đưa 4 hình ảnh này về đúng với vị trí chính xác.

Đồ án đề xuất phương pháp xử lý tự động dịch hình ảnh khuôn mặt về đúng vị trí. Dưới đây là các bước hoạt động cơ bản của phương án đề xuất:

1. Sử dụng thuật toán Canny để xác định các điểm là cạnh trong ảnh thu được một ảnh nhị phân. Thuật toán Canny được sử dụng để phát hiện các cạnh trong hình ảnh bằng cách áp dụng bộ lọc Gaussian để làm mờ và giảm nhiễu, tính toán gradient sử dụng đạo hàm Sobel để xác định độ lớn và hướng của gradient tại mỗi điểm ảnh. Sau đó, thuật toán loại bỏ các điểm không phải cực đại cục bộ và phân loại các cạnh dựa trên hai ngưỡng để xác định các cạnh mạnh và yếu, cuối cùng là liên kết các cạnh để loại bỏ các cạnh yếu không liên quan.
2. Tiếp theo chọn một ảnh làm ảnh gốc sau đó xét 3 hình ảnh còn lại: từng ảnh một sẽ dịch đi một khoảng trong phạm vi 20 pixel, mỗi một lần dịch chuyển xong sẽ thu được một ảnh mới và sẽ so sánh với ảnh gốc ban đầu và đặt score của cặp ảnh này chính là số pixel cạnh chung của cả hai ảnh, cuối cùng ảnh

dịch chuyển có score cao nhất chính là ảnh cần tìm.

3. Cuối cùng chúng ta thu được một ảnh gốc và ba ảnh dịch chuyển với các điểm score, một phiên ảnh mà điểm score của ba ảnh dịch đều lớn hơn 100 thì chính là một phiên hợp lệ được xử lý bằng phương pháp tự động này.

Bộ dữ liệu sau khi sử dụng phương pháp trên thì còn lại 1350 phiên ảnh hợp lệ của 255 người khác nhau và kết quả thu được có độ chính xác cao.

3.2.3 Tạo bộ dữ liệu huấn luyện

Sau khi thu được các phiên chứa 4 ảnh đã qua xử lý bước tiếp theo sẽ bỏ đi những ảnh kém chất lượng hoặc bị che khuất dựa trên file metadataII.txt. Bước tiếp theo sẽ tạo bộ dữ liệu 2D và 3D từ 4 ảnh 2D đã qua xử lý

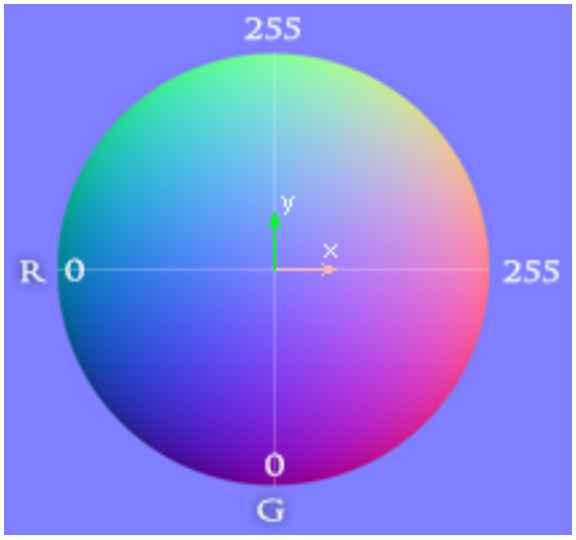
a, Bộ dữ liệu 2D

Chúng ta thu được 4 ảnh 2D sau khi đã qua bước xử lý ban đầu nhưng mỗi ảnh đều có phần bị quá sáng hoặc quá tối do vị trí đặt vị trí chiếu sáng ở 4 góc khác nhau. Để thu được một hình ảnh 2D có chất lượng đủ tốt và có đầy đủ các thông tin thì em đã tiến hành gộp lấy hình ảnh trung bình của cả 4 ảnh. Mỗi giá trị pixel trên hình ảnh cuối cùng sẽ được tính bằng giá trị trung bình của 4 hình ảnh 2D. Đây là một cách đơn giản và cũng hiệu quả để chọn lấy hình ảnh từ 4 khuôn mặt có điều kiện chiếu sáng khác nhau.

b, Bộ dữ liệu 3D

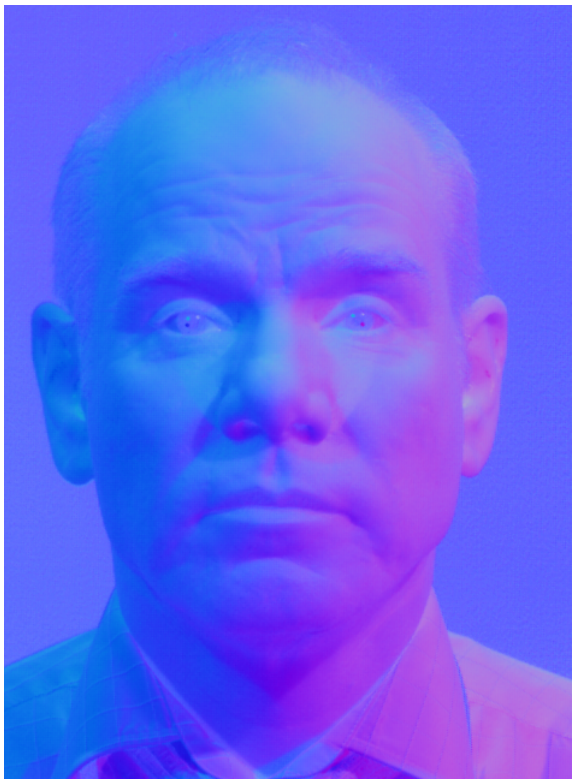
Khác với tạo hình ảnh 2D ở phần trước hình ảnh 3D cần phải sử dụng một phương pháp xử lý ảnh để có thể tạo được thông tin 3D từ các hình ảnh 2D. Phương pháp mà em đề xuất sử dụng trong đồ án là phương pháp Photometric Stereo.

Đây là một phương pháp trong lĩnh vực thị giác máy tính (computer vision) dùng để tái tạo bề mặt 3D của một đối tượng từ nhiều hình ảnh 2D chụp dưới các điều kiện ánh sáng khác nhau. Photometric Stereo hoạt động dựa theo nguyên lý ánh sáng phản xạ từ bề mặt của một vật thể phụ thuộc vào hướng ánh sáng chiếu tới và tính chất phản xạ của bề mặt đó. Bằng cách chụp nhiều ảnh của vật thể dưới các góc chiếu sáng khác nhau, ta có thể thu thập thông tin về sự thay đổi độ sáng và từ đó suy ra thông tin về độ nghiêng và hướng của bề mặt tại mỗi điểm ảnh (không gian ánh xạ của độ nghiêng, hướng trong hình 3.5).



Hình 3.5: Không gian ánh xạ trong normal map

Phần code Photometric Stereo được theo khảo trên Github. Đầu vào của phương pháp này là các ảnh 2D với các điều kiện ánh sáng khác nhau và thông tin về các góc $slant$ (chính bằng góc $zenith$) và $tilt$ (bằng $90°$ - $azimuth$). Kết quả đầu ra của phương pháp là một ảnh normal map (ví dụ hình 3.6) biểu diễn các vector pháp tuyến bề mặt của khuôn mặt. Mỗi pixel trong normal map chứa thông tin về hướng của pháp tuyến tại điểm đó.



Hình 3.6: Hình ảnh khuôn mặt 3D

c, Phát hiện khuôn mặt

Bước cuối cùng của tiền xử lý dữ liệu, các bộ dữ liệu cần đi qua một mô hình Face Detection để phát hiện khuôn mặt và cắt lấy phần khuôn mặt để sử dụng cho huấn luyện mô hình. Đồ án đề xuất sử dụng mô hình Face Detection của MediaPipe. MediaPipe là một framework mã nguồn mở được phát triển bởi Google, cung cấp các giải pháp về thị giác máy tính và xử lý hình ảnh thời gian thực và phát hiện khuôn mặt (Face Detection) là một trong những ứng dụng phổ biến nhất. Mô hình sử dụng mạng học sâu (deep learning) để xác định các đặc điểm của khuôn mặt trong hình ảnh, khi một khuôn mặt được phát hiện, MediaPipe sẽ trả về một hộp giới hạn xung quanh khuôn mặt đó, xác định vị trí và kích thước của khuôn mặt trong hình ảnh. MediaPipe Face Detection không chỉ dễ dàng triển khai sử dụng mà còn cung cấp kết quả phát hiện khuôn mặt với độ chính xác cao, ngay cả trong điều kiện ánh sáng yếu hoặc góc nhìn khó.

Cuối cùng sau khi qua bước tiền xử lý dữ liệu ta thu được 2 bộ dữ liệu khuôn mặt 2D và 3D với 473 phiên ảnh của 128 người khác nhau, với mỗi người có ít nhất 2 phiên ảnh và không nhiều hơn 5 phiên ảnh nhằm tránh mất cân bằng dữ liệu.

3.3 Huấn luyện mô hình

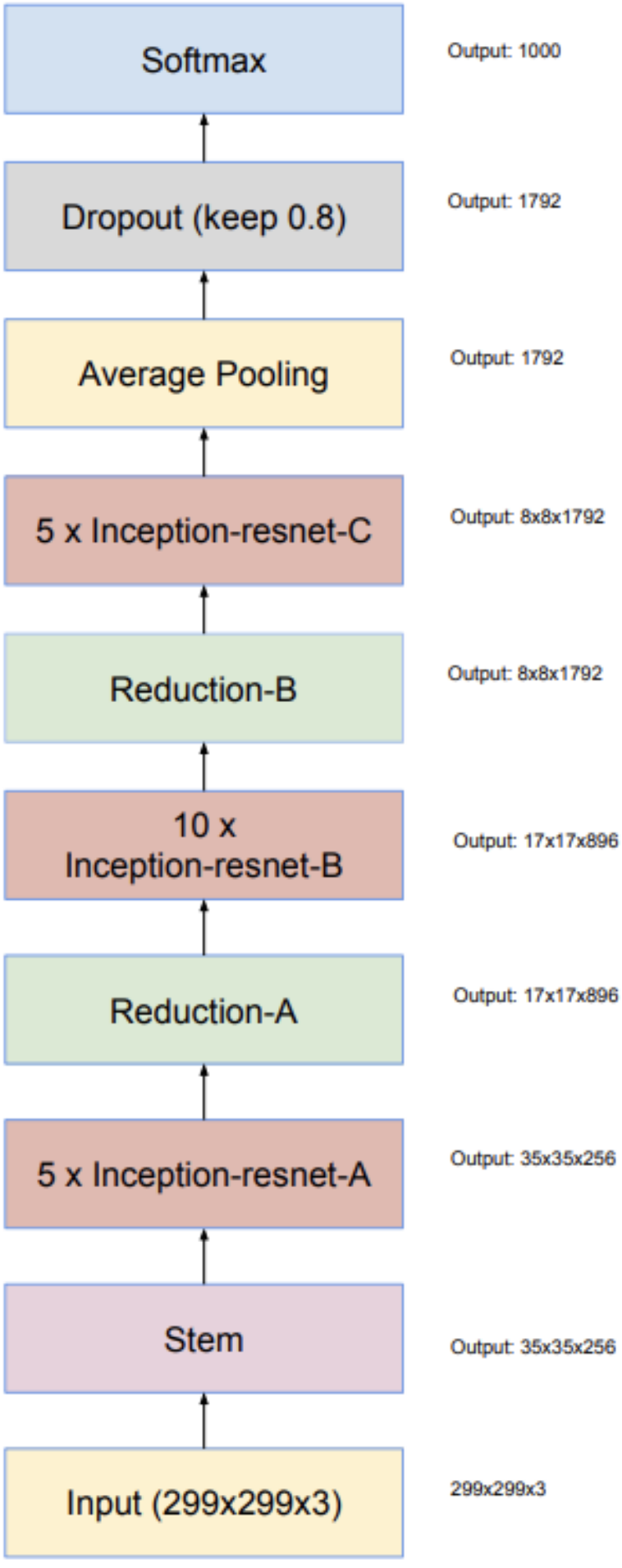
Phần này trình bày về các vấn đề như xây dựng các mô hình nhận diện khuôn mặt, các tham số và huấn luyện mô hình.

3.3.1 Xây dựng mô hình

Đồ án đề xuất hai mô hình để đánh giá khả năng nhận diện khuôn mặt Classification Model và Triplet Model. Cả hai mô hình này đều sử dụng backbone là mạng tích chập InceptionResNetv1 để học ra vector đặc trưng cho khuôn mặt với số chiều là 512 (với dữ liệu 2D hoặc 3D) hoặc 1024 (với dữ liệu kết hợp 2D và 3D).

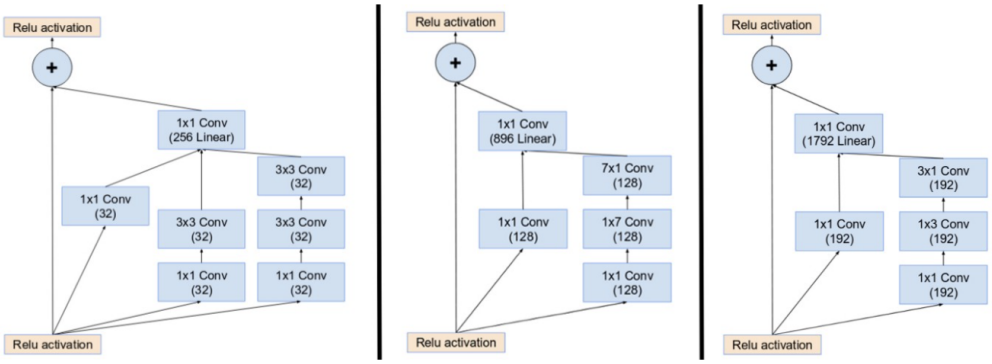
InceptionResNetv1 là một trong những mô hình mạng học sâu tiên tiến kết hợp hai kiến trúc mạng nổi tiếng: Inception và ResNet. Mô hình này được thiết kế để tận dụng ưu điểm của cả hai kiến trúc nhằm cải thiện hiệu suất và độ chính xác trong các tác vụ nhận diện và phân loại ảnh.

Hình 3.7 là tổng quan kiến trúc của mạng InceptionResNetv1.



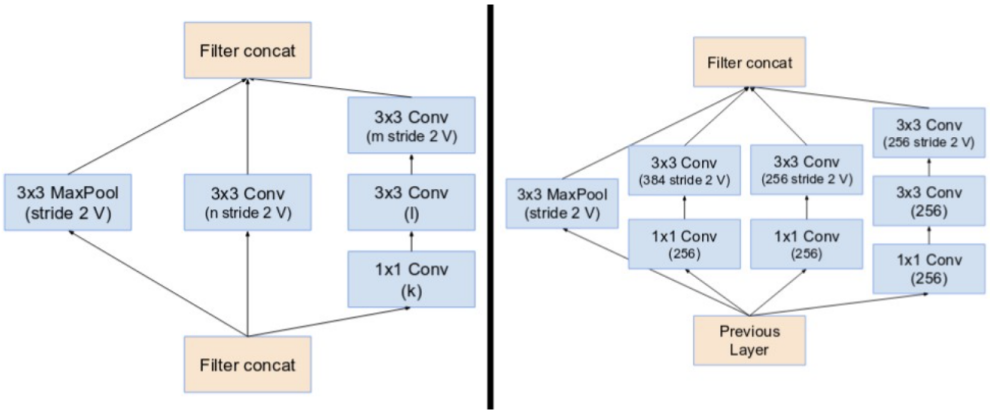
Hình 3.7: Kiến trúc InceptionResnetv1

Hình 3.8 là chi tiết về 3 khối Incpetion-ResNet: A, B, C lần lượt từ trái qua phải.



Hình 3.8: Kiến trúc khối Incpetion-ResNet

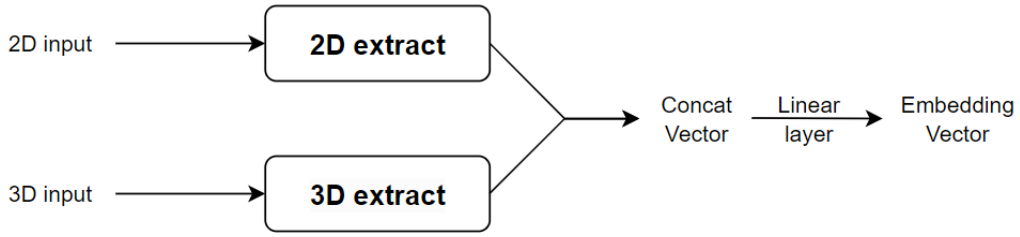
Hình 3.9 là chi tiết lần lượt về 2 khối Reduction A và Reduction B.



Hình 3.9: Kiến trúc khối Reduction

Mô hình classification sẽ nhận một đầu vào và sẽ sử dụng hàm Cross Entropy Loss để học ra được một vector ở cuối mô hình sau đó vector này được đưa qua tầng logit cuối cùng và được phân lớp dựa vào hàm softmax. Và vector cuối cùng này chính là vector đặc trưng đại diện cho khuôn mặt đầu vào.

Mô hình triplet sẽ nhận ba đầu vào lần lượt là anchor (dữ liệu gốc), positive (dữ liệu cùng lớp với anchor), negative (dữ liệu khác lớp với anchor). Phần chính của mô hình này vẫn là mạng backbone InceptionResNetv1 nhưng sẽ không sử dụng tầng logit cuối cùng mà sẽ học bộ ba vector. Quá trình học của mô hình này sẽ đi tối ưu hàm Triplet Loss được tính từ bộ ba vertor sao cho khoảng cách giữa $anchor$ và $positive$ nhỏ hơn khoảng cách giữa $anchor$ và $negative$, với một biên độ $margin$.



Hình 3.10: Mô hình sử dụng kếp hợp thông tin khuôn mặt 2D và 3D

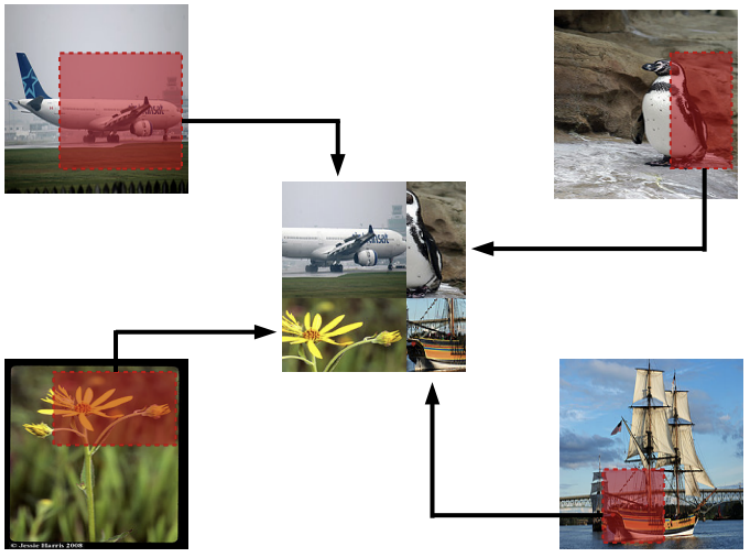
Với mô hình chỉ sử dụng thông tin khuôn mặt 2D hoặc 3D thì chính ta sẽ dử dụng bộ dữ liệu 2D và 3D đã được tiền xử lý để huấn luyện luôn embedding vector. Còn với mô hình sử dụng kết hợp thông tin 2D và 3D (hình 3.10) thì sẽ sử dụng cả hai hình ảnh để làm đầu vào của mô hình, chúng sẽ đi qua hai mạng InceptionResNetv1 để trích xuất ra hai vector sau đó được concat lại và đi qua tầng linear cuối cùng để thu được vector đặc trưng cuối cùng.

3.3.2 Huấn luyện mô hình

Phần huấn luyện mô hình sẽ trình bày các kĩ thuật và tham số để huấn luyện mô hình.

a, Tạo data loader

Bộ dữ liệu được chia thành hai tập train và valid với tỉ lệ là 80:20. Tập train và valid này không được chia theo phiên ảnh mà chia theo từng người riêng biệt. Khi tạo data loader sẽ được thêm các phép tăng cường dữ liệu random crop (hình 3.11) và gauss noise (hình 3.12) để giảm hiện tượng overfitting khi huấn luyện mô hình:



Hình 3.11: Ví dụ về data augmentation: Random crop

Hình 3.12: Ví dụ về data augmentation: Gaussian noise

- Random crop: Trong bước phát hiện khuôn mặt và cắt khuôn mặt đã được scale lên 1.2 lần bounding box để thu được khuôn mặt rộng hơn để sử dụng cho phép tăng cường này. Khuôn mặt sẽ được cắt ngẫu nhiên chiều dài và ngẫu nhiên chiều rộng từ 80% đến 100% kích thước ban đầu sau đó được resize về kích thước 256x256.
- Gaussian noise: Phép tăng cường dữ liệu này sẽ chọn ngẫu nhiên một vùng ngẫu nhiên chiều dài và ngẫu nhiên chiều rộng từ 10% đến 50% kích thước ban đầu để thêm nhiễu. Nhiễu được thêm vào với giá trị trung bình là 0 và độ lệch chuẩn được chọn ngẫu nhiên từ 0 tới 0.1.

b, Các tham số huấn luyện mô hình

Mô hình sử dụng thuật toán tối ưu Adam kết hợp điều chỉnh tham số learning rate theo epoch với bộ lập lịch Cosine Annealing Warm Restarts để tạo ra chu kì tăng giảm learning rate giúp mô hình có thể vượt qua các điểm tối ưu cục bộ tốt hơn.

Mô hình được huấn luyện với:

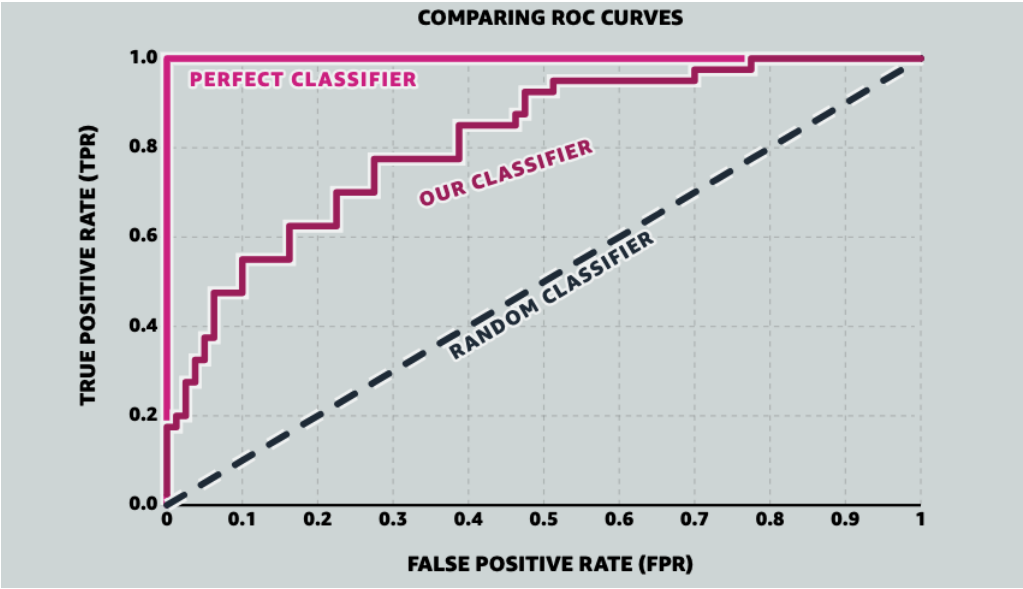
- Số lượng epoch là 1000.
- Learning rate khởi tạo ban đầu là $1e$-4 và giảm xuống $1e$-6 và tăng lại như ban đầu sau 50 epoch.
- Tài nguyên GPU P100 trên Kaggle Notebook.

CHƯƠNG 4. ĐÁNH GIÁ THỰC NGHIỆM

Chương 3 đã trình bày về cách xử lý bộ dữ liệu cũng như phương pháp xây dựng và huấn luyện mô hình. Sau khi thu được mô hình chúng ta cần phải đánh giá mô hình nào có hiệu suất tốt hơn và hiệu suất nhận dạng khuôn mặt có cải thiện không. Chương 4 sẽ trình bày về phương pháp đánh giá và kết quả đánh giá của mô hình.

4.1 Phương pháp và tham số đánh giá

Đường cong ROC (Receiver Operating Characteristic) biểu diễn mối quan hệ giữa tỷ lệ dương tính giả (False Positive Rate - FPR) và tỷ lệ dương tính thật (True Positive Rate - TPR) khi ngưỡng phân loại thay đổi. Chỉ số AUC là diện tích dưới đường cong ROC sẽ đánh giá được hiệu suất nhận biết của mô hình. Đường cong ROC được tạo ra bằng cách vẽ TPR (True Positive Rate) trên trục y và FPR (False Positive Rate) trên trục x. Hình 4.1 minh họa về đường ROC và tương quan về chỉ số AUC với hiệu suất của mô hình.



Hình 4.1: Minh họa đường ROC và độ đo AUC

Một số khái niệm:

- True Positive (TP): Số lượng mẫu dương tính đúng.
- False Positive (FP): Số lượng mẫu âm tính sai.
- True Negative (TN): Số lượng mẫu âm tính đúng.
- False Negative (FN): Số lượng mẫu dương tính sai.

True Positive Rate (TPR):

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

False Positive Rate (FPR):

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Đường cong ROC được vẽ bằng cách thay đổi ngưỡng quyết định của mô hình từ 0 đến 1 và tính toán các giá trị TPR và FPR tương ứng. Một mô hình phân loại tốt sẽ có đường ROC gần với góc trên bên trái của đồ thị.

AUC (Area Under the Curve) là diện tích dưới đường cong ROC. AUC cung cấp một giá trị số duy nhất để đánh giá hiệu suất tổng thể của mô hình phân loại.

- **AUC = 1**: Mô hình hoàn hảo, phân biệt chính xác tất cả các mẫu dương tính và âm tính.
- **AUC = 0.5**: Mô hình phân loại ngẫu nhiên, không phân biệt được mẫu dương tính và âm tính.
- **AUC < 0.5**: Mô hình tệ hơn phân loại ngẫu nhiên, tức là có thể hoán đổi dự đoán của mô hình để có kết quả tốt hơn.

Thông thường một mô hình sẽ có chỉ số AUC nằm trong khoảng từ 0.5 tới 1, các mô hình có chỉ số AUC càng gần 1 thì càng tốt.

Sau khi thu được các vector đặc trưng khi đưa khuôn mặt qua mô hình thì các vector này sẽ được so sánh độ tương động cosine hoặc khoảng cách euclid với các vector đặc trưng khác và dựa vào ngưỡng để xác định đây là khuôn mặt của cùng một người hay là hai người khác nhau. Nhãn dương được quy ước là khuôn mặt cùng một người còn nhãn âm là khuôn mặt không cùng một người.

Mô hình sẽ được đánh giá thông qua hàm loss và chỉ số AUC dựa trên cả tập train và tập valid trong quá trình học.

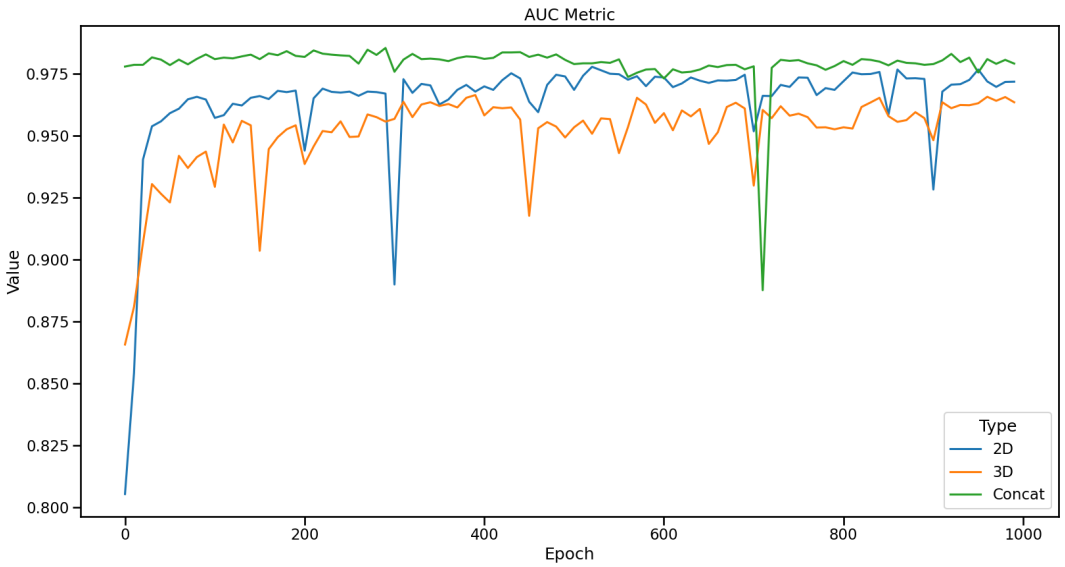
4.2 Kết quả mô hình classification

Bảng 4.1 là kết quả các giá trị tối ưu của loss và AUC mà mô hình phân loại học được. Các giá trị trong bảng là giá trị tối ưu mà các chỉ số đánh giá đạt được tại các epoch khác nhau. Với việc chênh lệnh lớn của loss train và loss valid chúng ta có thể nhận thấy rằng mô hình đã bị overfitting. Việc sử dụng mô hình kết hợp thông tin 2D và 3D đã làm giảm việc overfitting của mô hình và cải thiện tốt được chỉ số AUC của mô hình.

Model	**2D**	**3D**	**Concat**
AUC train	0.9993	0.9984	1.0000
AUC valid	0.9798	0.9789	0.9866

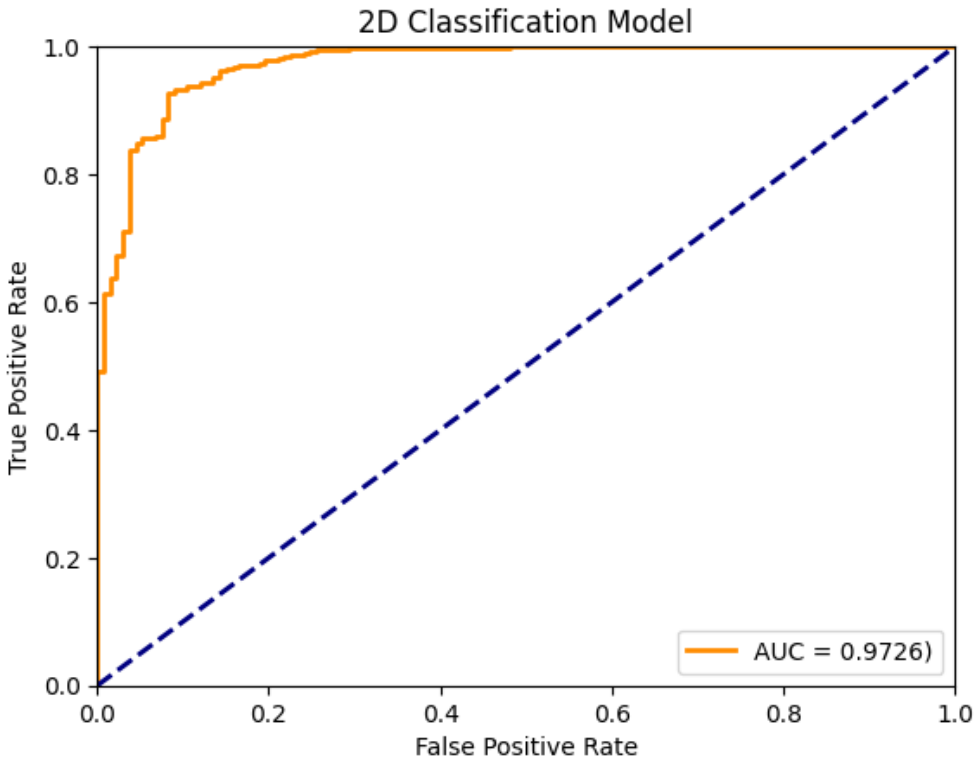
Bảng 4.1: Kết quả mô hình phân loại

Hình 4.2 cho thấy kết quả đánh giá AUC của các mô hình với bộ dữ liệu valid trong quá trình train.

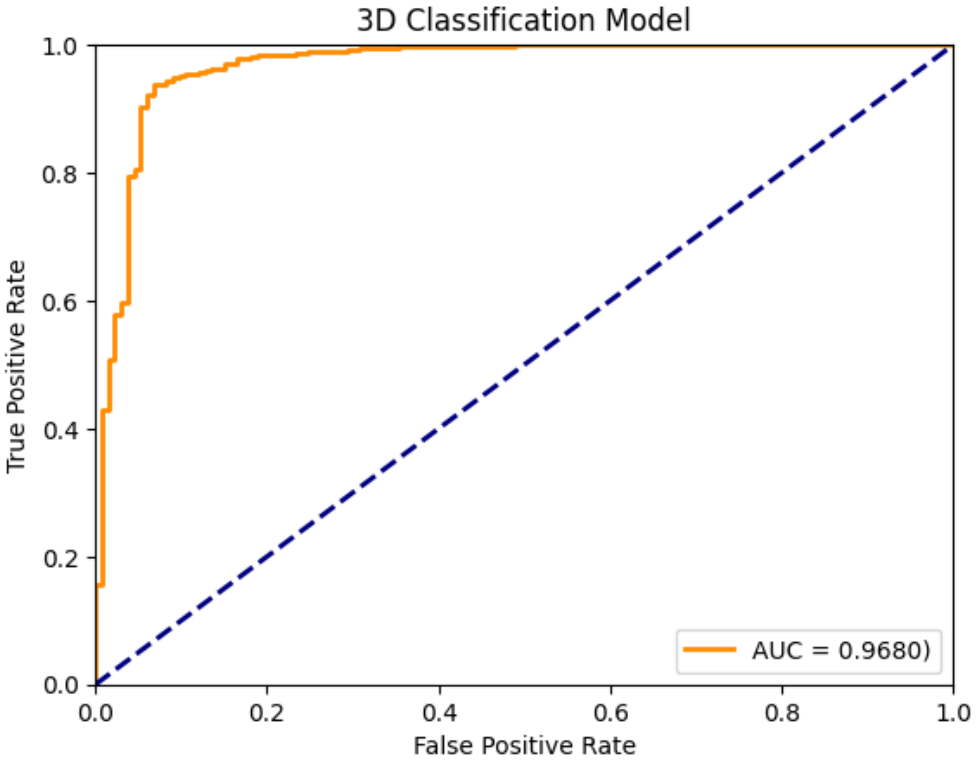


Hình 4.2: Kết quả đánh giá mô hình phân loại trên tập dữ liệu valid

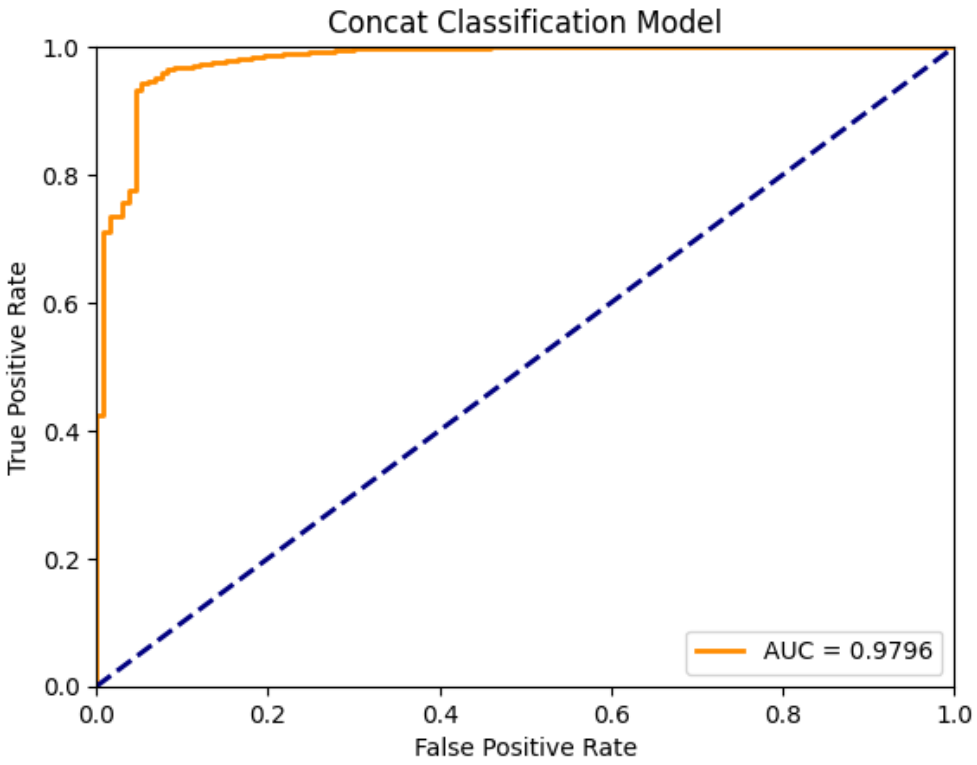
Hình 4.7, 4.8, 4.9 cho thấy kết quả đánh giá AUC của các mô hình có giá trị hàm loss thấp nhất.



Hình 4.3: Kết quả AUC của mô hình phân loại 2D



Hình 4.4: Kết quả AUC của mô hình phân loại 3D



Hình 4.5: Kết quả AUC của mô hình phân loại kết hợp 2D và 3D

Từ kết quả đánh giá của mô hình phân loại chúng ta nhận thấy rằng:

- Bộ dữ liệu 3D có hiệu suất kém hơn so với bộ dữ liệu 2D.
- Việc kết hợp thông tin 2D và 3D giúp cải thiện mô hình như giảm hiện tượng overfitting, tăng hiệu suất nhận diện (cải thiện chỉ số AUC).

$\Longrightarrow$ Kết hợp thông tin 2D và 3D giúp mô hình phân loại trích xuất đặc trưng khuôn mặt hiệu quả hơn dẫn đến cải thiện hiệu suất nhận diện khuôn mặt.

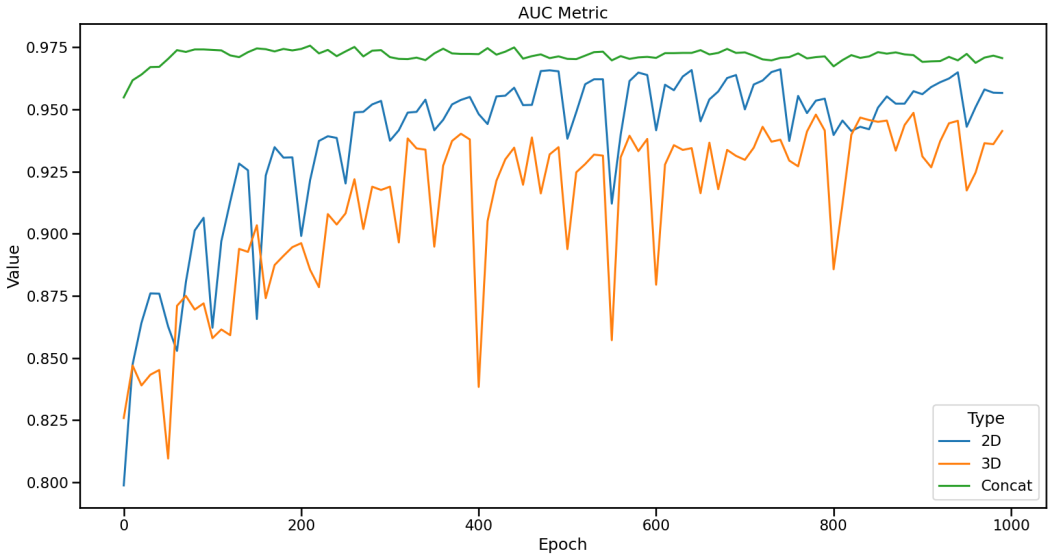
4.3 Kết quả mô hình triplet

Bảng 4.2 là kết quả các giá trị tối ưu của loss và AUC mà mô hình triplet học được. Các giá trị trong bảng là giá trị tối ưu mà các chỉ số đánh giá đạt được tại các epoch khác nhau. So với mô hình phân loại thì mô hình triplet có kết quả tốt hơn. Khi huấn luyện giá trị của hàm loss khá sát nhau chứng tỏ quá trình huấn luyện đã có thể học được các đặc trưng của khuôn mặt. Kết quả của chỉ số AUC cũng đã cho thấy mô hình không bị overfitting nhiều. Cũng tương tự như mô hình phân loại, khi kết hợp thông tin 2D và 3D đã cải thiện được chỉ số AUC của mô hình cũng như khả năng trích xuất đặc trưng của khuôn mặt.

Model	**2D**	**3D**	**Concat**
AUC train	0.9962	0.9980	0.9889
AUC valid	0.9735	0.9627	0.9756

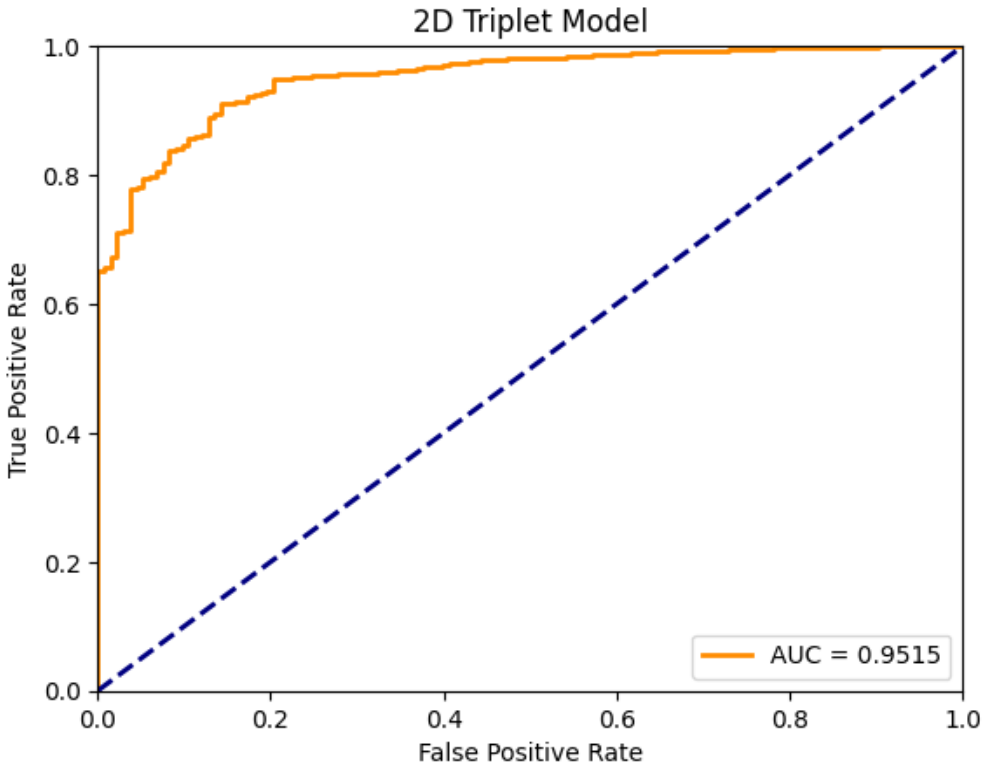
Bảng 4.2: Kết quả mô hình triplet

Hình 4.6 cho thấy kết quả đánh giá AUC của các mô hình với bộ dữ liệu valid trong quá trình train.

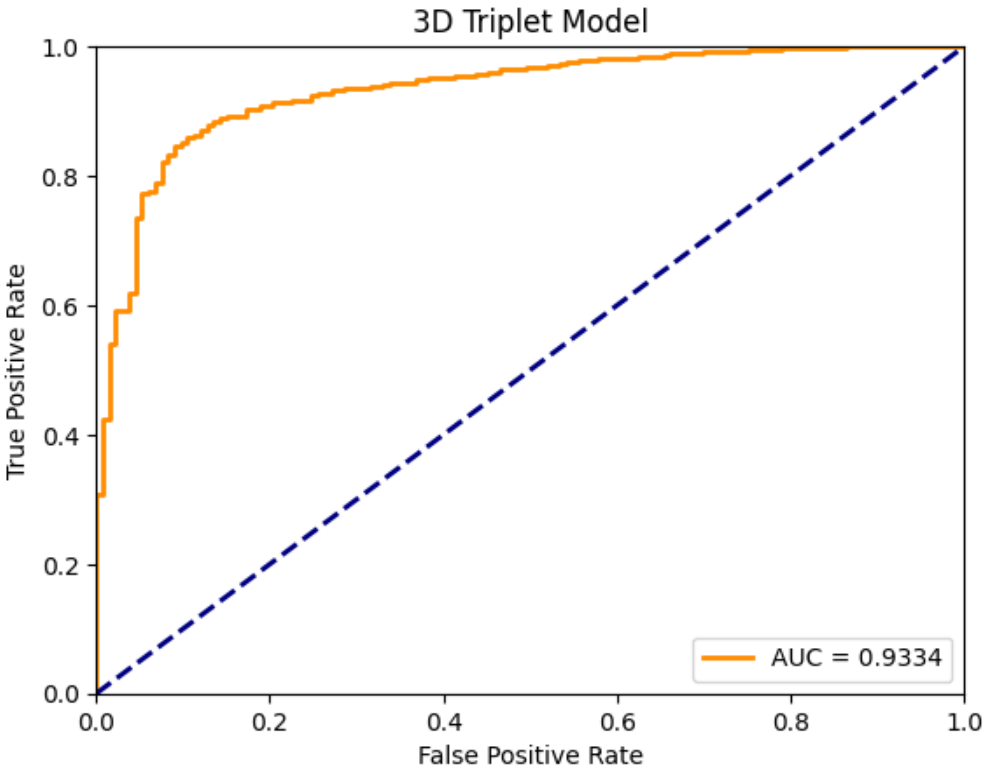


Hình 4.6: Kết quả đánh giá mô hình triplet trên tập dữ liệu valid

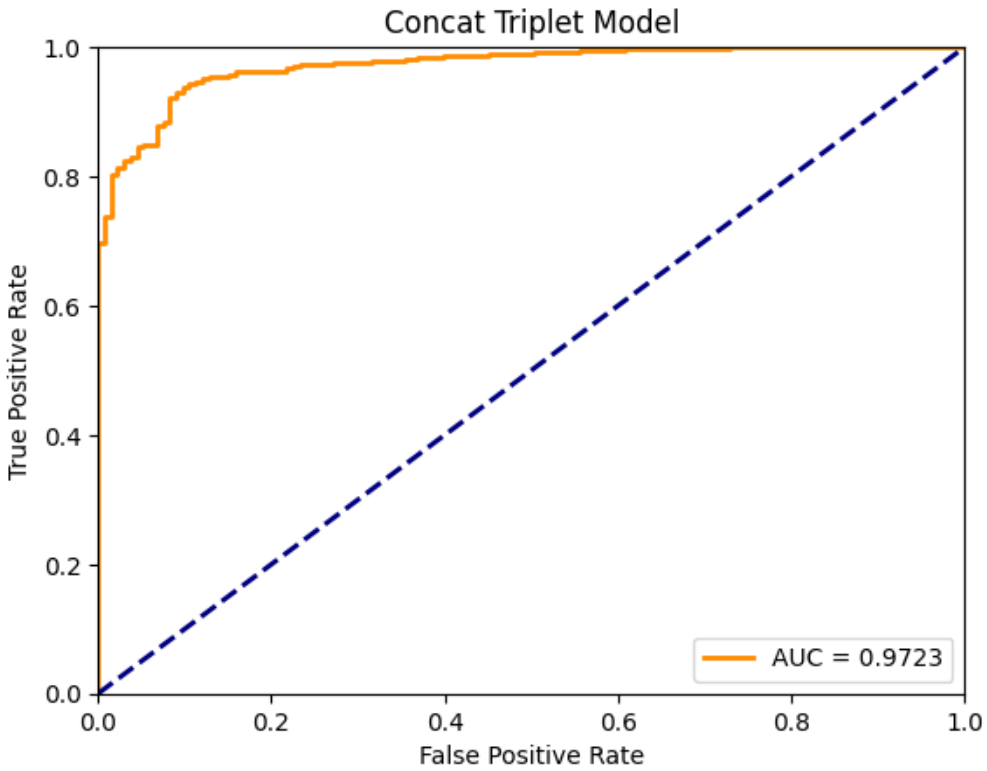
Hình 4.7, 4.8, 4.9 cho thấy kết quả đánh giá AUC của các mô hình có giá trị hàm loss thấp nhất.



Hình 4.7: Kết quả AUC của mô hình triplet 2D



Hình 4.8: Kết quả AUC của mô hình triplet 3D



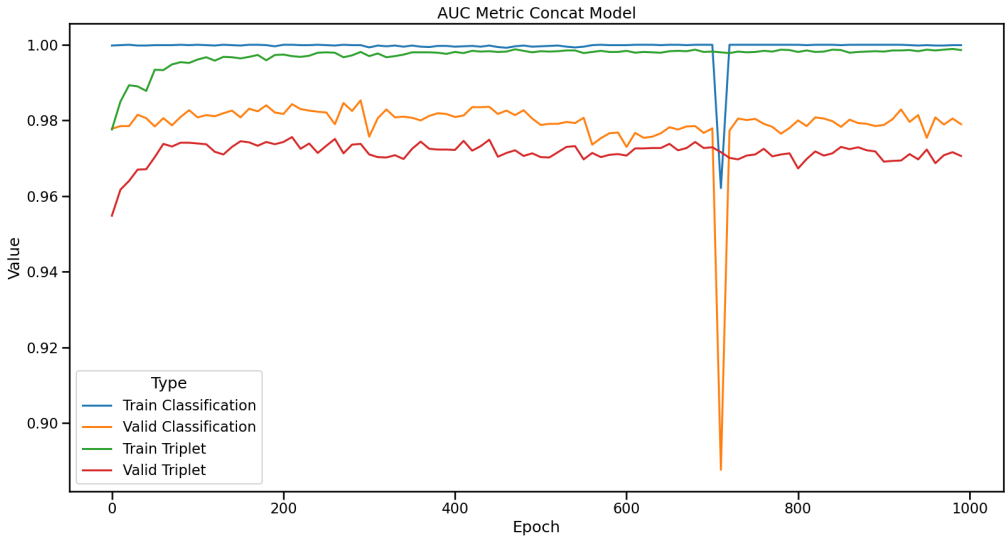
Hình 4.9: Kết quả AUC của mô hình triplet kết hợp 2D và 3D

Từ kết quả đánh giá của mô hình triplet chúng ta nhận thấy rằng:

- Bộ dữ liệu 3D có hiệu suất kém hơn so với bộ dữ liệu 2D.
- Việc kết hợp thông tin 2D và 3D giúp cải thiện mô hình như giảm hiện tượng overfitting, tăng hiệu suất nhận diện (cải thiện chỉ số AUC).

$\Longrightarrow$ Kết hợp thông tin 2D và 3D giúp mô hình triplet trích xuất đặc trưng khuôn mặt hiệu quả hơn dẫn đến cải thiện hiệu suất nhận diện khuôn mặt.

Đánh giá chung:



Hình 4.10: Kết quả AUC của hai mô hình kết hợp 2D và 3D

- Dựa trên tổng quan kết quả đánh giá thực nghiệm (hình 4.10) có thể thấy được mô hình triplet có khả năng học tốt hơn và đỡ bị overfitting so với mô hình phân loại.
- Trong ba bộ dữ liệu thì bộ dữ liệu 2D kết hợp với 3D có kết quả đánh giá tốt hơn so với 2 bộ dữ liệu riêng lẻ, từ đây có thể cho thấy việc kết hợp thông tin 2D và 3D đã bổ sung thông tin cho nhau và cải thiện được chất lượng mô hình.

CHƯƠNG 5. KẾT LUẬN

5.1 Kết luận

Đồ án đã trình bày về bài toán nhận diện khuôn mặt và đề xuất phương pháp sử dụng kết hợp thông tin 3D để cái thiện hiệu suất nhận diện mô hình. Bộ dữ liệu khuôn mặt được xử lý để trích xuất thông tin 3D, đây là một bộ dữ liệu khó do có điều kiện ánh sáng không đồng đều và cần tự tạo hình ảnh thông tin 3D. Mô hình nhận diện được xây dựng dựa trên backbone InceptionResnetv1 theo hai hướng là mô hình phân loại và mô hình triplet. Sau đó kết hợp bộ dữ liệu cùng các phép tăng cường dữ liệu để huấn luyện hai mô hình để trích xuất đặc trưng. Cuối cùng tiến hành thực nghiệm và đánh giá mô hình nhận diện khuôn mặt đã xây dựng.

Trong quá trình xây dựng, đồ án còn gặp một số khó khăn như chưa thể giải quyết được toàn bộ dữ liệu ban đầu để thu được nhiều dữ liệu gốc. Từ đó dẫn đến mô hình chưa có nhiều dữ liệu để học dẫn đến mô hình vẫn còn bị overfitting đặc biệt là mô hình phân loại.

Kết quả cho thấy mô hình triplet cho kết quả học khả quan hơn mô hình phân loại. Khi sử dụng kết hợp thông tin 2D và 3D để huấn luyện mô hình đã giúp cho hiệu suất mô hình tăng lên. Điều này đã chứng minh rằng khuôn mặt 3D đã đóng góp thông tin đặc trưng giúp mô hình nhận đặt được chỉ số AUC từ 0.97 tới 0.98, đây là một hiệu suất nhận diện tốt cho một mô hình nhận diện.

5.2 Hướng phát triển trong tương lai

Nhằm khắc phục những hạn chế hiện tại và tiếp tục nâng cao hiệu suất của mô hình nhận diện khuôn mặt sử dụng thông tin 3D, mô hình có thể phát triển theo một số hướng sau đây:

1. Tăng cường và mở rộng bộ dữ liệu:
    - Thu thập thêm dữ liệu: Mở rộng bộ dữ liệu bằng cách thu thập thêm hình ảnh khuôn mặt từ nhiều nguồn khác nhau, bao gồm các điều kiện ánh sáng và góc chụp đa dạng hơn.
    - Tạo dữ liệu 3D: Sử dụng các kỹ thuật và công cụ tiên tiến khác để tạo dữ liệu 3D từ các hình ảnh 2D hiện có, giúp tăng số lượng dữ liệu huấn luyện và cải thiện tính đa dạng của dữ liệu.
2. Thử nghiệm kiến trúc mô hình:
    - Thay đổi backbone: Thử nghiệm các backbone khác để tìm một backbone phù hợp với bài toán nhận diện khuôn mặt và phụ hợp với bộ dữ liệu.

- Thay đổi kiến trúc mô hình: Đồ án lấy đầu vào là 2 ảnh 2D và 3D đi vào mô hình song song, chúng ta có thể thử nghiệm các kỹ thuật kết hợp khác để sử dụng tốt thông tin dữ liệu 2D và 3D cho mô hình nhận diện khuôn mặt.

TÀI LIỆU THAM KHẢO

[1] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings of Computer Vision and Pattern Recognition*, IEEE, pp. 586–591, 1991.

[2] T. Ojala, M. Pietikäinen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Featured Distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1701–1708, 2014.

[5] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

[6] Y. Jing, X. Lu, and S. Gao, "3D Face Recognition: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2415–2435, 2017.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4278–4284, 2017.

[8] Stefanos Zafeiriou, Mark Hansen, Gary Atkinson, Vasileios Argyriou, Maria Petrou, Melvyn Smith, and Lyndon Smith, "The photoface database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 132–139, 2011.

[9] S. Kautkar, G. Atkinson, and M. Smith, "Face recognition in 2D and 2.5D using ridgelets and photometric stereo," *Pattern Recognition*, vol. 45, pp. 3317–3327. 2012.

[10] X. Wang, K. Wang, and S. Lian, "A Survey on Face Data Augmentation," 2019.