

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

CÔNG NGHỆ THÔNG TIN



NHẬN DẠNG - CSC14006

BÁO CÁO CUỐI KỲ

Đề tài: FACE DETECTION BASED ON SKIN COLOR

THẦY (CÔ) HƯỚNG DẪN

PGS.TS. LÊ HOÀNG THÁI

Thầy. ĐƯƠNG THÁI BẢO

Thầy. TRƯỜNG TÂN KHOA

SINH VIÊN THỰC HIỆN

MSSV	Họ và Tên	Email
22127147	Đỗ Minh Huy	dmhuy22@clc.fitus.edu.vn
22127249	Trần Thanh Long	ttlóng22@clc.fitus.edu.vn
22127322	Lê Phước Phát	lpphat22@clc.fitus.edu.vn
22127330	Nguyễn Đức Phúc	ndphuc22@clc.fitus.edu.vn

THÀNH PHỐ HỒ CHÍ MINH, THÁNG 05, 2025

LỜI NÓI ĐẦU

Thị giác máy tính, điển hình là nhận dạng mẫu, là một lĩnh vực quan trọng mang tính khoa học và công nghệ. Ứng dụng của nó, điển hình là phát hiện và nhận dạng gương mặt dựa trên nhiều yếu tố, mang tính cấp thiết trong các lĩnh vực khác nhau như: y học, vật lý, toán học, tìm kiếm, bảo mật, và rất nhiều lĩnh vực khoa học khác nhau, ...

Phát hiện khuôn mặt là một phần trong lĩnh vực xử lý ảnh, là một vấn đề cơ bản trong ứng dụng thị giác máy tính. Đây được xem là một trong những giai đoạn của hệ thống nhận dạng mặt người ban đầu của hệ thống nhận dạng gương mặt.

Tuy nhiên, việc phát hiện gương mặt bị ảnh hưởng bởi nhiều yếu tố khác nhau theo khía cách chủ quan lẫn khách quan. Về mặt chủ quan, hiệu suất của bài toán thường bị ảnh hưởng bởi chất lượng hình ảnh, cơ chế cũng như kiến trúc của ảnh. Về mặt khách quan, hiệu suất cũng bị ảnh hưởng với màu sắc, ánh sáng ngoại lai, nhiễu, ...

Trong đồ án này, nhóm sẽ trình bày phương pháp loại bỏ đi một số yếu tố ảnh hưởng đến độ chính xác trong phát hiện mặt người. Phương pháp này xem như là một bước tiền xử lý hình ảnh trước khi đưa qua các mô hình thuật toán nhận diện gương mặt. Về phương pháp, kỹ thuật này sẽ dựa vào vùng da mặt đã được phân đoạn để phát hiện gương mặt, loại bỏ các vùng không gian dư thừa trong ảnh.

Mặc dù các thành viên trong nhóm đã rất cố gắng để hoàn thành thật tốt bài tập nghiên cứu cá nhân nhưng chắc chắn không thể tránh khỏi những hạn chế và thiếu sót không mong muốn. Nhóm chúng em mong nhận được sự thông cảm và ý kiến đóng góp của thầy cô trong lớp học phần **NHẬN DẠNG MẪU - CSC14006** để em có thể rút kinh nghiệm và hoàn thiện những bài tập sau tốt hơn.

Nhóm chúng em xin được gửi lời cảm ơn chân thành tới thầy **PGS.TS Lê Hoàng Thái**, thầy **Dương Thái Bảo** và thầy **Trương Tân Khoa** vì đã luôn tận tình chỉ bảo, hướng dẫn, và giải đáp mọi thắc mắc của em trong suốt quá trình thực hiện bài tập cá nhân này.

Tóm lại, thông qua bài tập này, chúng ta có thể nhìn nhận được tập quan trọng của phát hiện và nhận diện gương mặt dựa trên yếu tố vùng da nói riêng và ứng dụng nhận dạng nói chung trong đời sống khoa học - xã hội hiện đại ngày nay.

TP.HCM, mùa xuân 2025.

LỜI CAM ĐOAN

Nhóm thực hiện đồ án **SKIN COLOR IN FACE ANALYSIS** gồm các thành viên trên đều là sinh viên khoa Công nghệ Thông tin Chất lượng cao, thuộc trường Đại học Khoa học Tự nhiên, ĐHQG-HCM. Nhóm cam đoan rằng bài tập nghiên cứu này là do các thành viên trong nhóm tìm hiểu, nghiên cứu và thực hiện dưới sự giám sát và hướng dẫn của các thầy **PGS.TS. Lê Hoàng Thái**, thầy **Dương Thái Bảo**, và thầy **Trương Tấn Khoa**. Các dữ liệu được nêu trong đồ án là hoàn toàn trung thực, phản ánh đúng kết quả mô phỏng thực tế. Tất cả các tài liệu được sử dụng trong nghiên cứu này được các thành viên trong nhóm thu thập bằng cách tự thân và từ các nguồn khác nhau, và những tài liệu này được liệt kê đầy đủ trong phần tài liệu tham khảo. Tất cả đều được trích dẫn đúng đắn. Trong trường hợp có vi phạm bản quyền, các thành viên trong nhóm sẽ chịu trách nhiệm cho hành động đó. Do đó, trường **Đại học Khoa học Tự nhiên, ĐHQG-HCM** không chịu trách nhiệm về bất kỳ vi phạm bản quyền nào được thực hiện trong bài tập nghiên cứu này.

TP.HCM, ngày 06 tháng 04 năm 2025.

Người cam đoan

Nhóm trưởng

LÊ PHƯỚC PHÁT

MỤC LỤC

DANH MỤC HÌNH VẼ	i
CHƯƠNG 01. GIỚI THIỆU	1
CHƯƠNG 02. CƠ SỞ LÝ THUYẾT	5
2.1 Cải thiện chất lượng ảnh (Image enhancement)	5
2.2 Phân đoạn vùng da (skin image segmentation) dựa trên mô hình lai RGB–YCbCr–HSV	6
2.3 Khử nhiễu bằng lọc trung vị	6
2.4 Tính số Euler để loại bỏ vùng không phải khuôn mặt	7
2.5 Xác định khuôn mặt dựa trên Bounding Box và độ lệch tâm	7
CHƯƠNG 03. PHƯƠNG PHÁP NGHIÊN CỨU	10
3.1 Phương án nghiên cứu ban đầu	10
3.1.1 Tiền xử lý	11
3.1.2 Thuật toán xác định màu da	11
3.1.3 Thuật toán xác định khuôn mặt	12
3.1.4 Thảo luận	12
3.2 Phương án nghiên cứu cải tiến	13
3.2.1 Thu thập dữ liệu	14
3.2.2 Tiền xử lý dữ liệu	17
3.2.3 Fine-tuned mô hình phân đoạn UNet	17
3.2.4 Xác định và hiển thị vùng ROIs da	19
3.2.5 Áp dụng các mô hình pretrained phát hiện gương mặt	20
3.3 Kết luận	20
CHƯƠNG 04. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ	21
KẾT LUẬN	22

Kết luận chung	22
Hướng phát triển	22
Kiến nghị và đề xuất	23
TÀI LIỆU THAM KHẢO	24

DANH MỤC HÌNH VẼ

Hình 1.1 Quy trình phát hiện khuôn mặt dựa trên phân đoạn da và đặc trưng hình học	5
Hình 3.1 Ví dụ thử nghiệm với phương pháp chuyển không gian màu	13
Hình 3.2 Pipeline xử lý nghiên cứu Face Detection based on Skin Segmentation	13
Hình 3.3 Ảnh minh họa bộ dữ liệu CelebAMask-HQ	14
Hình 3.4 Ảnh minh họa bộ dữ liệu Pratheepan	16
Hình 3.5 Kết quả dự đoán của mô hình	18
Hình 3.6 Ảnh sau khi được xử lý hiển thị vùng ROIs và contours	19
Hình 3.7 Kết quả cuối cùng của pipeline	20
Hình 4.1 Ứng dụng phát hiện gương mặt	21

CHƯƠNG 01. GIỚI THIỆU

Trong thị giác máy tính, màu sắc luôn là một đặc trưng thường xuyên được sử dụng bởi tính đơn giản, bởi lẽ các phép toán với màu sắc có thể được cài đặt một cách nhanh chóng và hiệu quả, đồng thời trong môi trường ổn định với độ sáng đồng đều, màu sắc thường không bị tác động bởi sự thay đổi của hình học. Do đó, trong một số trường hợp, chỉ cần màu sắc là đủ để nhận dạng mọi sự vật.

Tuy nhiên, sự khó khăn chính hiện nay trong sử dụng màu sắc cho các ứng dụng thị giác máy tính là các camera thường không thể phân biệt được sự thay đổi của màu sắc do quang phổ của các nguồn ánh sáng khác nhau chiếu vào. Do đó, màu sắc thường nhạy cảm với sự thay đổi của nguồn sáng, điều này xảy ra rất thường xuyên trong môi trường không được kiểm soát. Các sự thay đổi này có thể được gây ra bởi sự thay đổi độ sáng (như bóng của vật thể), hoặc sự thay đổi cường độ ánh sáng (như ánh sáng mặt trời hoặc các nguồn sáng huyền quang từ bóng đèn, ...), hoặc cả hai sự thay đổi trên. Ngoài ra, các camera khác nhau và các mức điều chỉnh khác nhau có thể tạo ra nhiều bức ảnh khác nhau trong mắt con người.

Vào thời điểm hiện tại, vấn đề trên đã có nhiều hướng giải quyết được đề ra nhằm giảm thiểu sự nhạy cảm với sự thay đổi của nguồn sáng một cách không mong muốn. Trong đó bao gồm hai hướng chính sau đây:

- Thông tin màu sắc sẽ được chia làm hai thành phần chính bao gồm cường độ màu sắc và sắc độ. Khi đó, ta sử dụng sắc độ để giảm bớt tác động của sự thay đổi độ sáng, và áp dụng các thuật toán cố định màu sắc nhằm loại bỏ ảnh hưởng của sự thay đổi cường độ ánh sáng. Tuy nhiên, hiệu quả của phương pháp này vẫn còn hạn chế [1].
- Ngoài ra, ta có thể để các mô hình tự thích nghi với sự thay đổi của nguồn sáng. Phương pháp này có thể mang lại những kết quả rất đáng mong đợi được trình bày trong báo cáo này.

Trong thị giác máy tính, việc giảm thiểu sự phụ thuộc vào cường độ ánh sáng là điều thường được ưu tiên. Mục tiêu của chúng ta là phải loại bỏ hoàn toàn ảnh hưởng của màu sắc của nguồn sáng bằng cách xác định một biểu diễn màu chỉ phụ thuộc vào hệ số phản xạ bề mặt. Tuy nhiên, cho đến nay, điều này vẫn chưa đạt được trong lĩnh vực thị giác máy. Hệ thống thị giác của con người vượt trội hơn trong khía cạnh này, vì nhận thức màu sắc của con người phụ thuộc đáng kể vào hệ số phản xạ bề mặt, mặc dù ánh sáng đến mắt là kết quả của sự kết hợp giữa hệ số phản xạ bề mặt, màu sắc của nguồn sáng và cường độ ánh sáng.

Đối với những thành tựu đạt được trong lĩnh vực nhận dạng khuôn mặt, màu sắc thường được sử dụng như bước tiền xử lý đầu tiên để thực hiện các bước nghiên cứu sâu hơn và đòi hỏi tính toán nhiều hơn. Ví dụ như trong việc phát hiện khuôn mặt, thay vì phải xét hết từng vị trí và độ lớn khác nhau trong ảnh để phát hiện, thì với màu sắc, ta có thể tiền xử lý bức ảnh đó, và chỉ lấy những phần có màu sắc giống với màu da.

Trong chương 09 của cuốn sách **Handbook of Face Recognition** [2] của hai tác giả Stan Z. Li và Anil K. Jain hay trong bài báo **SKIN COLOR IN FACE ANALYSIS** [3] của ba tác giả J. Birgitta Martinkauppi, Abdenour Hadid, và Matti Pietikäinen đã trình bày chi tiết về vai trò của màu sắc trong việc phân tích hình ảnh khuôn mặt, đặc biệt là phát hiện và nhận diện màu da. Các nội dung ấy sẽ được nhóm chúng em trình bày theo cách hiểu bản thân một cách chi tiết và đầy đủ nhất trong **chương 02. Tổng quan tài liệu và cơ sở lý thuyết**. Nội dung ấy sẽ trình bày các vấn đề như sau:

- Giới thiệu về việc sử dụng màu sắc trong lĩnh vực phân tích ảnh khuôn mặt.
- Giới thiệu về thành phần của màu sắc và ảnh hưởng của từng nguồn sáng lên màu sắc.
- Phân tách các nguồn dữ liệu da.
- Giới thiệu việc mô hình hóa màu da.
- Đánh giá việc sử dụng màu sắc trong phát hiện khuôn mặt.
- Lợi ích của màu sắc trong nhận diện khuôn mặt.
- Mô hình phân đoạn hình ảnh dựa theo ngữ nghĩa
- Mô hình phát hiện khuôn mặt hiện đại kết hợp học sau

Sau khi chúng ta tìm hiểu các nội dung trên, nhóm chúng em sẽ nghiên cứu các phương án tối ưu nhất dựa trên các nghiên cứu hiện đại nhất lúc bấy giờ và đưa ra đề xuất cho đồ án này để có thể hoạt động một cách hiệu quả, được trình bày trong **chương 03. Phương pháp nghiên cứu**.

Sau đó, chúng em sẽ nói về hướng cài đặt mã nguồn cũng như trình bày việc báo cáo đồ án cuối kỳ như thế nào trong phần **chương 04. Thực nghiệm và Kết quả**. Tuy nhiên, do chúng em chỉ mới đề xuất hướng đi nên chúng em chưa có kết quả đầu ra để phân tích cụ thể, mong các thầy có thể bỏ qua thiếu sót này trong bài báo cáo này.

Cuối cùng, nhóm chúng em sẽ đưa ra kết luận cuối cùng về phương pháp nghiên cứu cũng như hướng đi sắp tới của nhóm em trong đồ án **Face Detection based on Skin Color** trong phần **Kết luận và Đề nghị**.

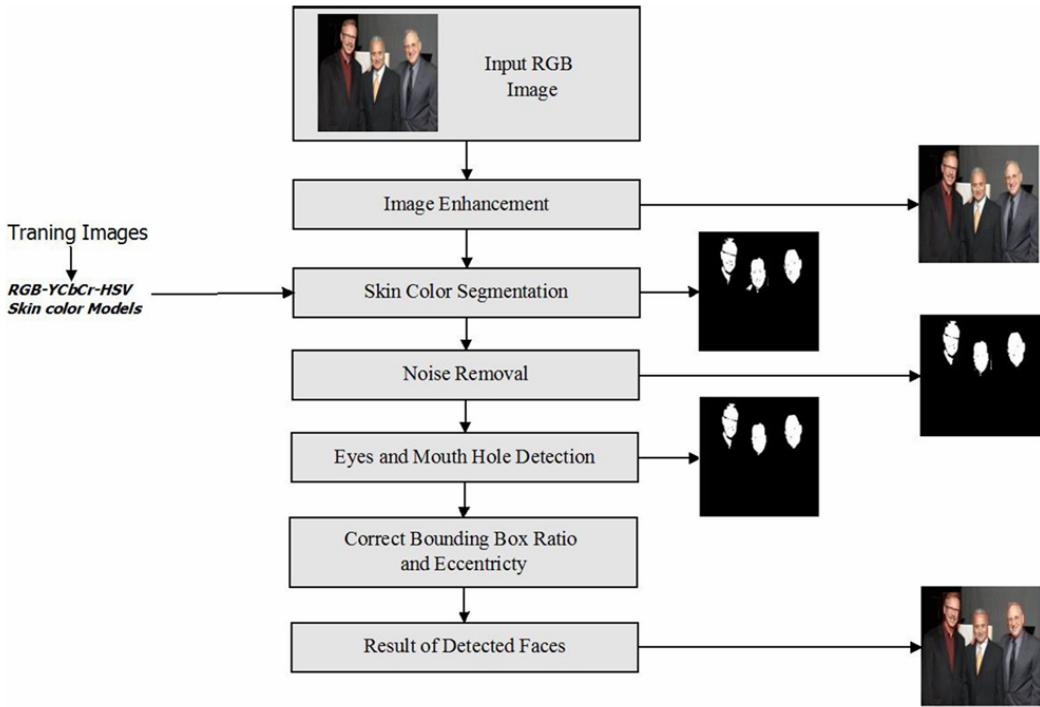
Mong các thầy có thể xem xét và đưa ra những lời khuyên bổ ích cho chủ đề và nội dung đồ án này để nhóm chúng em có thể kịp thời tinh chỉnh cũng như tìm ra hướng nghiên cứu và thực nghiệm hợp lý.

CHƯƠNG 02. TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ LÝ THUYẾT

Các phương pháp phát hiện khuôn mặt được phân thành bốn nhóm chính:

- **Phương pháp so khớp mẫu (Template Matching):** Phương pháp này lưu trữ nhiều mẫu chuẩn để mô tả khuôn mặt, có thể là toàn bộ khuôn mặt hoặc các đặc trưng riêng lẻ như mắt, mũi... Vị trí khuôn mặt được tìm dựa trên giá trị tương quan với mẫu chuẩn. Tuy nhiên, nó gặp khó khăn khi phải đối mặt với các thay đổi về góc nhìn, biểu cảm, điều kiện chiếu sáng,...
- **Phương pháp bất biến đặc trưng (Feature Invariant Methods):** Dựa vào các đặc trưng khuôn mặt như mắt, mũi, tai, môi và màu da. Những phương pháp này có thể xử lý các điều kiện ánh sáng thay đổi, hoặc các tư thế quay ngoài/phẳng.
- **Phương pháp dựa trên tri thức (Knowledge-Based Methods):** Dựa vào các quy tắc mô tả mối quan hệ giữa các đặc trưng khuôn mặt. Tuy nhiên, độ chính xác của phương pháp phụ thuộc mạnh vào cách thiết kế quy tắc. Quy tắc nghiêm ngặt có thể bỏ sót khuôn mặt, còn quy tắc đơn giản có thể gây nhầm lẫn.
- **Phương pháp dựa trên biểu diễn (Appearance-Based Methods):** Những phương pháp này yêu cầu huấn luyện mô hình trên một tập ảnh. Sau khi huấn luyện, mô hình có thể được áp dụng cho ảnh mới để phát hiện khuôn mặt. Phát hiện khuôn mặt bằng mạng nơ-ron nhân tạo cũng thuộc nhóm này.

Phương pháp nhóm tụi em sử dụng cách tiếp cận lai tạo từ 2 phương pháp **bất biến đặc trưng và dựa trên tri thức**



Hình 1.1 Quy trình phát hiện khuôn mặt dựa trên phân đoạn da và đặc trưng hình học

2.1 Cải thiện chất lượng ảnh (Image enhancement)

Ánh sáng ảnh hưởng lớn đến phân đoạn da. Vì vậy, ảnh đầu vào được chuyển sang không gian YCbCr, tính giá trị trung bình Y (độ sáng) và áp dụng hệ số hiệu chỉnh T:

Các bước sau đây tóm tắt quá trình cải thiện chất lượng ảnh:

1. Chuyển ảnh từ không gian màu *RGB* sang *YCbCr* và chuẩn hóa thành phần độ sáng *Y*.
2. Tính giá trị trung bình của thành phần *Y*.

$$Y_{avg} = \frac{\sum Y(x, y)}{W \times H} \quad (2.1)$$

Trong đó, *W* và *H* lần lượt là chiều rộng và chiều cao của ảnh. Ngưỡng *T* được thiết lập dựa trên giá trị trung bình *Y_{avg}*.

$$T = \begin{cases} 1.4, & \text{nếu } Y_{avg} < 64 \\ 0.6, & \text{nếu } Y_{avg} > 192 \\ 1, & \text{còn lại} \end{cases}$$

Sau đó, thành phần đỏ (R) và xanh lá (G) được hiệu chỉnh bởi T và ghép lại với B tạo thành ảnh mới đã cải thiện chất lượng.

2.2 Phân đoạn vùng da (skin image segmentation) dựa trên mô hình lai RGB-YCbCr-HSV

Hình ảnh được biểu diễn bằng không gian màu RGB kém hiệu quả hơn trong việc xây dựng các mô hình màu da vì vùng màu da không được phân biệt rõ ràng trong tất cả ba cường độ kênh RGB.

Để phát hiện khuôn mặt, bước đầu tiên là phân đoạn vùng da nhằm giảm độ phức tạp tính toán. Mô hình lai sử dụng đồng thời ba không gian màu:

- **RGB:** $R > 95, G > 40, B > 20, \max - \min > 15, |R - G| \geq 15, R > G, R > B$
- **YCbCr:** $85 \leq Cb \leq 135, 10 \leq Cr \leq 45, Y \geq 80$
- **HSV:** $H \in (0^\circ - 35^\circ) \text{ hoặc } (325^\circ - 360^\circ), S \in (0.2 - 0.6), V \geq 20$

Vùng nào thoả mãn đồng thời ba điều kiện trên sẽ được gán là vùng da.

2.3 Khử nhiễu bằng lọc trung vị

Nhiễu là kết quả của lỗi trong khi xử lý hình ảnh, dẫn đến các giá trị pixel không phản ánh cường độ thực của ảnh. Nhiễu có thể xuất hiện trong hình ảnh do một số lý do như:

- Trong khi quét hình ảnh từ ảnh chụp trên phim, hạt phim là nguồn gây nhiễu. Nếu phim bị hỏng, thì sẽ gây nhiễu.
- Nếu hình ảnh được lưu trực tiếp dưới dạng kỹ thuật số, cơ chế thu thập dữ liệu có thể gây nhiễu

Có nhiều loại bộ lọc nhiễu khác nhau như bộ lọc thông thấp, FFT (biến đổi Fourier nhanh) và bộ lọc trung vị. Vấn đề khi sử dụng bộ lọc thông thấp hoặc FFT là đôi khi chúng loại bỏ thông tin quan trọng. Do đó, nhóm sử dụng bộ lọc làm mịn trung vị trên hình ảnh để làm mịn nhiễu của nó.

Ảnh được khử nhiễu bằng phương pháp lọc trung vị vì trung vị ít nhạy cảm hơn nhiều so với giá trị trung bình đối với các giá trị cực trị (giá trị ngoại lai)

Lọc trung vị được sử dụng để loại bỏ nhiễu mà không làm mất chi tiết ảnh. Mỗi pixel được thay thế bằng giá trị trung vị trong vùng lân cận, giúp loại bỏ điểm bất thường mà không làm mờ viền ảnh.

2.4 Tính số Euler để loại bỏ vùng không phải khuôn mặt

Do mắt và miệng tạo thành lỗ trên mặt người, những vùng da không có lỗ sẽ bị loại bỏ. Số Euler được tính theo công thức:

$$E = C - H$$

Trong đó:

- E là số Euler,
- C là số thành phần liên thông,
- H là số lỗ trong vùng đó.

Nếu $E \geq 0$, vùng bị loại bỏ.

2.5 Xác định khuôn mặt dựa trên Bounding Box và độ lệch tâm

Sau khi lọc bằng Euler, hai đặc trưng hình học tiếp theo được áp dụng:

- **Tỷ lệ Bounding Box:** $R = \frac{W}{H}$, chấp nhận nếu $R > 0.35$
- **Độ lệch tâm (Eccentricity):** $0.1 < E < 0.89$ thì được coi là hình dạng khuôn mặt hợp lệ

Vùng nào thoả mãn cả hai điều kiện sẽ được xác định là vùng khuôn mặt.

Các bước đầy đủ được minh họa trong thuật toán:

Thuật toán phát hiện khuôn mặt

Bắt đầu

1. Đọc một ảnh đầu vào.
2. Chuyển đổi ảnh RGB sang ảnh không gian màu YCbCr.
3. Để bù sáng, thực hiện các bước sau:
 - a. Tìm giá trị nhỏ nhất $\min(Y)$ và lớn nhất $\max(Y)$ của thành phần Y và chuẩn hóa thành phần Y .
 - b. Tính giá trị trung bình của Y .

$$Y_{avg} = \frac{\sum Y(x,y)}{W \cdot H}$$

- c. Đặt giá trị ngưỡng với các điều kiện sau:

$$T = \begin{cases} 1.4 & \text{nếu } Y_{avg} < 64 \\ 0.6 & \text{nếu } Y_{avg} > 192 \end{cases}$$

- d. Sửa đổi thành phần R(Red) và G(Green) của ảnh bằng cách sử dụng ngưỡng T :

$$R_{new} = \begin{cases} R^T & \text{nếu } T \neq 1 \\ R & \text{trong trường hợp khác} \end{cases}, \quad G_{new} = \begin{cases} G^T & \text{nếu } T \neq 1 \\ G & \text{trong trường hợp khác} \end{cases}$$

- e. Kết hợp R_{new} , G_{new} và B(Blue) để có được ảnh màu cuối cùng.

4. Thực hiện phân đoạn màu da bằng cách sử dụng giá trị ngưỡng sau:

Nếu $V = [R, G, B]$ thỏa mãn các điều kiện:

$$(R > 95 \cap G > 40 \cap B > 20 \cap (\max(V) - \min(V) > 15) \cap |R - G| \geq 15 \cap R > G \cap R > B)$$

VÀ

$$(Y > 80 \cap 10 \leq Cr \leq 45 \cap 85 \leq Cb \leq 135)$$

VÀ

$$((0 < H < 35) \cup (325 < H < 360)) \cap (0.2 < S < 0.6) \cap (V \geq 20)$$

Vùng da

Ngược lại

Vùng không phải da

5. Thực hiện loại bỏ nhiễu bằng bộ lọc trung vị (median filter).
6. Tính toán số lượng thành phần liên thông trong ảnh đã phân đoạn da.
7. Thực hiện tính toán Euler E .

Nếu $E \geq 0$

Không có lỗ, loại bỏ vùng này.

Ngược lại

Chấp nhận.

8. Đặt tỷ lệ hộp giới hạn (bounding box) B_i và tỷ lệ độ lệch tâm (eccentricity ratio) Ec_i lên ảnh. (Giả sử B_i biểu thị các thuộc tính/tỷ lệ của hộp giới hạn).
9. **Nếu** điều kiện dựa trên B_i và Ec_i được thỏa mãn (ví dụ: $B_i \cap Ec_i = \text{true}$ được diễn giải là các điều kiện này đúng)

Vùng khuôn mặt

Ngược lại

Vùng không phải khuôn mặt.

Kết thúc

CHƯƠNG 03. PHƯƠNG PHÁP NGHIÊN CỨU

Sau khi chúng ta đã nghiên cứu về phần cơ sở lý thuyết, chúng ta biết được màu sắc là một đặc trưng hữu dụng trong xử lý ảnh khuôn mặt, thể hiện rõ nhất trong việc phân vùng da và phát hiện khuôn mặt, mặc dù độ hữu ích của màu sắc trong nhận dạng khuôn mặt vẫn chưa có câu trả lời tối ưu nhất định.

Vấn đề ở đây đầu tiên là việc lựa chọn không gian màu và mô hình màu da, bởi chưa có một giải pháp nào tối ưu cho vấn đề này, mà việc lựa chọn còn bị phụ thuộc vào yêu cầu của phần mềm và môi trường xung quanh. Một khi ta chọn được mô hình màu da, màu sắc có vai trò quan trọng trong phát hiện khuôn mặt, thể hiện trong quá trình tiền xử lý và lựa chọn những vùng trên ảnh có màu da. Nhờ đó, những bước lọc tiếp theo có thể được thực hiện nhằm tìm ra những khuôn mặt trong những vùng có màu da đó. Nhờ đó, các thuật toán phát hiện khuôn mặt sử dụng màu sắc sẽ hoạt động nhanh hơn rất nhiều so với các thuật toán sử dụng ảnh đen trắng, đặc biệt là với các ảnh có kích thước lớn.

Trong lĩnh vực nhận dạng khuôn mặt, việc thông tin về màu sắc có đem lại lợi ích cho việc nhận dạng hay không còn mang đến nhiều tranh cãi. Các kết quả thu được cho thấy rằng màu sắc chưa cho thấy hết toàn bộ tiềm năng và cần được nghiên cứu sâu hơn. Do đó, vào thời điểm hiện tại, có lẽ các hệ thống nhận dạng khuôn mặt chưa thể sử dụng màu sắc.

Đó cũng chính là lý do và động lực giúp nhóm chúng em thử nghiệm đề tài này. Trong bài báo cáo này, nhóm chúng em sẽ thử qua cả hai phương án nghiên cứu ban đầu lẫn cải tiến để xem xét hiệu quả sau đó, nhóm chúng em sẽ đi đến kết luận chọn phương án nghiên cứu tối ưu nhất. Phần này sẽ giúp giải đáp câu hỏi "Which state-of-the-art method will you deliver?".

3.1 Phương án nghiên cứu ban đầu

Phương án chuyển đổi không gian màu, sử dụng các không gian màu HSV và YCbCr để phát hiện màu da, và tạo một mask cho vùng da. Sau đó ta map tọa độ của mask này áp dụng lên ảnh gốc, các pixel nào nằm ngoài mask sẽ được filter về màu đen (tức giá trị pixel bằng 0), chỉ giữ lại các pixel ở ảnh gốc mà khớp với vùng mask. Sau đó, ta sẽ sử dụng một mô hình phát hiện khuôn mặt hiện đại để có thể phát hiện được khuôn mặt dựa trên ảnh mà ta đã áp dụng mask vào.

Việc sử dụng các phương pháp để phát hiện vùng da trước khi đưa vào các mô hình hiện đại để phát hiện khuôn mặt sẽ giúp máy sẽ loại bỏ được các vùng pixel không là ứng cử viên có thể là khuôn mặt (bởi các vùng này không phải là các vùng da). Các mô

hình phát hiện khuôn mặt của ta sẽ chỉ tập trung vào những vùng mà chúng ta đã giới hạn giúp nó, việc này sẽ giúp cho việc xác định và phát hiện khuôn mặt trở nên chính xác hơn và ít tốn chi phí hơn vì nó không tập trung quá nhiều vào những phần chi tiết không đáng có.

3.1.1 *Tiền xử lý*

Ban đầu, một bước tiền xử lý được thực hiện nhằm điều chỉnh độ sáng của bức ảnh. Cách tiền xử lý như sau:

- Chuyển bức ảnh RGB sang kênh màu YCbCr, và lọc ra kênh màu Y.
- Tính giá trị trung bình của kênh màu Y trên toàn bộ bức ảnh, gọi là Y_{avg} .
- Đặt giá trị ngưỡng T dựa theo giá trị của Y_{avg} : $T = \begin{cases} 1.4, & Y_{avg} < 64 \\ 0.6, & Y_{avg} > 192 \\ 1, & Khi khác \end{cases}$
- Thay đổi giá trị của hai kênh màu R và G: $R_{new} = R^T$, $G_{new} = G^T$.

3.1.2 *Thuật toán xác định màu da*

Để xác định màu da, ta sử dụng thuật toán đặt ngưỡng đơn giản, với các ngưỡng được cho như sau:

- $R > 95$
- $G > 40$
- $B > 20$
- $R - G \geq 15$
- $R > B$
- $85 \leq Cb \leq 135$
- $10 \leq Cr \leq 45$
- $Y \geq 80$
- $0 < H < 35$ hoặc $325 < H < 360$
- $0.2 < S < 0.6$
- $V \geq 20$

Sau các phép so sánh, ta được một binary mask cho vùng da. Để làm giảm nhiễu cho mask này, ta sử dụng phương pháp median filter đơn giản.

3.1.3 Thuật toán xác định khuôn mặt

Ta thực hiện thuật toán tìm thành phần liên thông để xác định các ứng cử viên cho khuôn mặt. Từ các ứng cử viên này, ta sẽ kiểm tra một số điều kiện để xác định đó có phải khuôn mặt hay không:

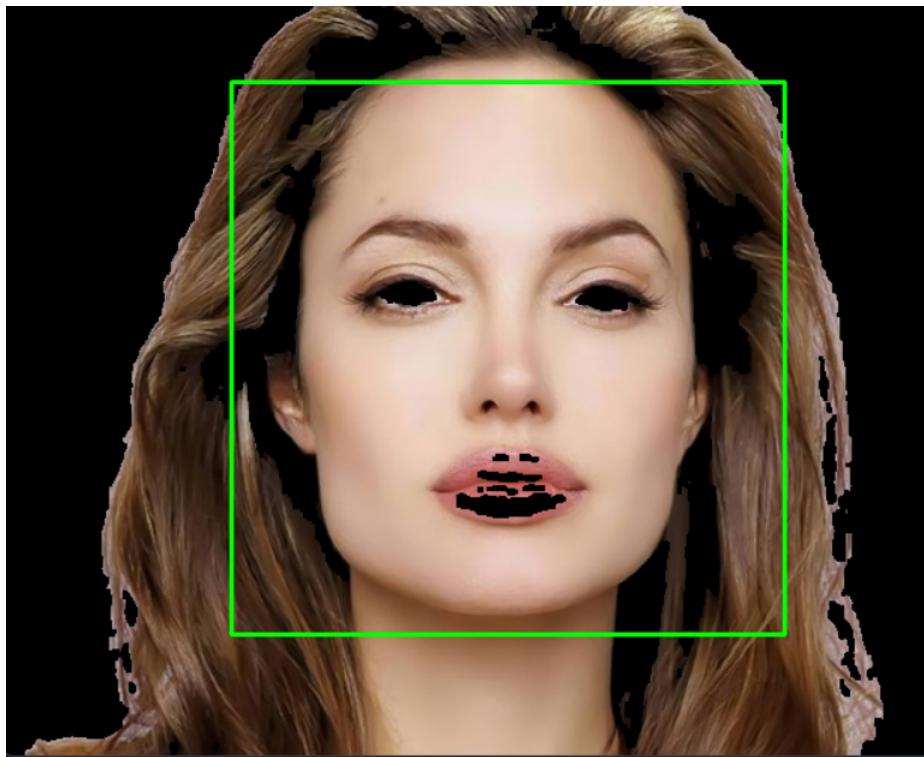
- Với mỗi ứng cử viên, ta tính hệ số Euler $E = C - H$, với C là số thành phần liên thông và H là số lỗ trống. Đây là một heuristics, trong việc khuôn mặt phải có đầy đủ hai mắt và miệng. Ở đây, điều đó được thể hiện qua phép so sánh $E \geq 0$. Nếu ứng cử viên không thoả điều kiện này, ta sẽ bỏ qua nó.
- Sau đó, ta tìm bounding box và bounding ellipse của ứng cử viên. Ta đặt R là tỉ lệ giữa chiều rộng và chiều cao của bounding box, và E là tỉ lệ giữa đường chéo phụ và đường chéo chính của bounding ellipse. Khi đó, thành phần liên thông đó được tính là khuôn mặt khi nó thoả mãn 2 điều kiện: $R > 0.35$, và $0.1 \leq E \leq 0.89$.

Sau các bước này, ta đã có xác định khuôn mặt trong ảnh.

3.1.4 Thảo luận

Tuy nhiên thì phương pháp này tồn tại cả những ưu điểm cũng như khuyết điểm, sau đây sẽ là các ưu điểm và khuyết điểm của phương pháp này mà chúng em nhận thấy được sau quá trình thử nghiệm:

- **Ưu điểm:** Ta có thể thấy được ngay rằng phương pháp này rất ít tốn chi phí về cả bộ nhớ lẫn thời gian để chúng ta có thể xác định và giới hạn lại phạm vi tìm kiếm. Bởi việc chỉ dùng các ngưỡng của các không gian màu như HSV và YCbCr để có thể lọc các pixel ít có khả năng là da người thì nó chỉ mất chi phí tính toán của chúng là 0(1) cho từng pixel với việc chỉ cần so sánh giá trị pixel với các ngưỡng đã đề sẵn.
- **Nhược điểm:** Nhưng chính việc đơn giản là chỉ sử dụng các ngưỡng để lọc các pixel không phải là vùng da thì nó sẽ dễ cho ra các sai sót và không tối ưu được việc giới hạn phạm vi tìm kiếm. Đây là một vấn đề mà chúng em gặp phải khi thử nghiệm phương pháp này.

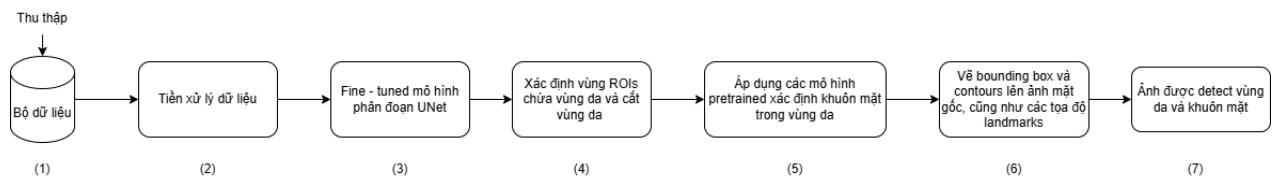


Hình 3.1 Ví dụ thử nghiệm với phương pháp chuyển không gian màu

- Ở đây ta có thể thấy là các background của ảnh, mắt, môi đều đã được thuật toán lọc vùng da của ta filter đi bớt và chỉ còn lại màu đen thôi, nhưng mà do ngưỡng màu của da và tóc nó cùng nằm trong một ngưỡng nên là ở đây tóc của chúng ta cũng không được lọc đi bởi thuật toán chuyển đổi không gian màu này. Chính vì vậy mà ta không thể tối ưu được việc giới hạn phạm vi phát hiện khuôn mặt. Do vậy mà chúng tôi đã nghiên cứu một phương pháp cải tiến mới đó là dùng Segmentation để có thể phân đoạn vùng da người một cách tối ưu hơn mà không dính những vật thể khác vào, để có thể tối ưu được vùng tìm kiếm khuôn mặt.

3.2 Phương án nghiên cứu cải tiến

Trong phương án cải tiến này, nhóm chúng em sẽ trình bày các giai đoạn thực hiện nghiên cứu như hình 3.2.



Hình 3.2 Pipeline xử lý nghiên cứu Face Detection based on Skin Segmentation

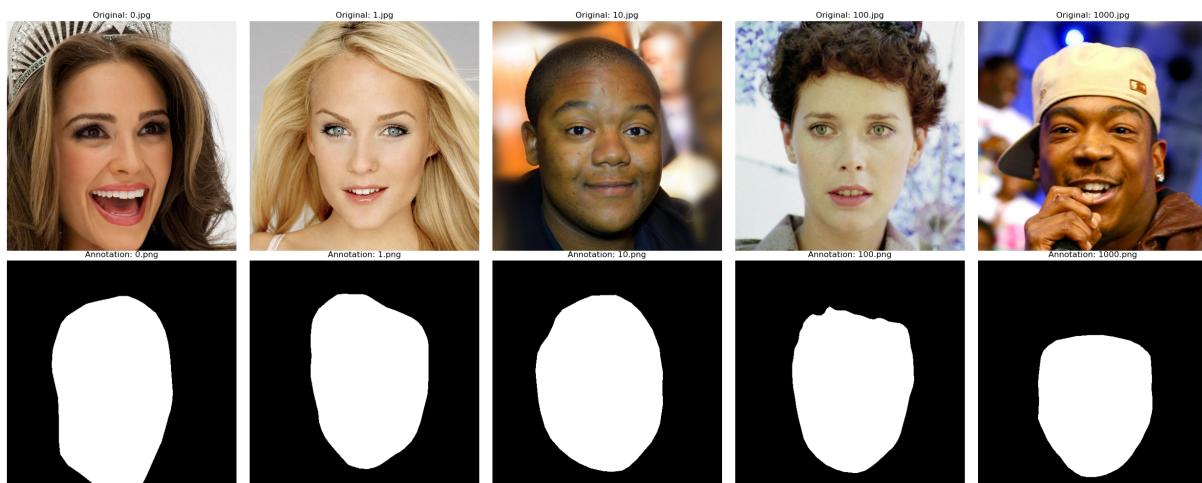
3.2.1 Thu thập dữ liệu

Về phần thu thập dữ liệu, nhóm chúng em sẽ đi khảo sát các dữ liệu được gán nhãn sẵn vùng da mặt người, trong đó nhóm đã tìm hiểu được những bộ dataset sau đây có thể khả thi cho quá trình thực hiện đồ án.

3.2.1.1 CelebAMask-HQ Dataset

CelebAMask-HQ [4] được giới thiệu lần đầu trong bài báo **MaskGAN: Towards Diverse and Interactive Facial Image Manipulation** [5] bởi các tác giả Cheng-Han Lee and Ziwei Liu và Lingyun Wu and Ping Luo vào năm 2020, là một bộ dữ liệu về mặt người theo diện rộng bao gồm 30000 ảnh mặt người với độ phân giải cao được chọn từ bộ dữ liệu CelebA theo CelebA-HQ. Mỗi ảnh có một bộ mask phân đoạn các đặc trưng khuôn mặt dựa theo CelebA. Những bộ masks của bộ dữ liệu CelebAMask-HQ đều được gán nhãn thủ công với kích thước ảnh mask là 512 x 512 và 19 lớp nhãn bao gồm tất cả các thành phần sinh học và phụ kiện có trên khuôn mặt từng người như da mặt, mũi, mắt trái, mắt phải, lông mày, tai, miệng, môi, tóc, nón, mắt kính, bông tai, dây chuyền, cổ và quần áo.

Bộ dữ liệu này được sử dụng rộng rãi trong nhiều nghiên cứu về phân đoạn khuôn mặt (face segmentation) và phân tích khuôn mặt (face parsing), được dùng để huấn luyện và đánh giá các mô hình nhận diện gương mặt và GANs cho việc tạo sinh và chỉnh sửa gương mặt.



Hình 3.3 Ảnh minh họa bộ dữ liệu CelebAMask-HQ

Tuy nhiên, do đề tài bài toán về phát hiện khuôn mặt dựa trên vùng da được phân đoạn nên nhóm chúng em chỉ sử dụng các ảnh gốc và ảnh annotations về vùng da mặt được gán nhãn là skins từ bộ dữ liệu dataset như được biểu diễn trong hình 3.4. Trong hình trên, hàng đầu tiên hiển thị toàn bộ ảnh gốc RGB có trong bộ dữ liệu CelebAMask-

HQ, đồng thời từng annotation tương ứng về vùng da mặt được hiển thị ở dòng dưới.

Bộ dữ liệu này cũng tồn tại những ưu điểm và khuyết điểm nhất định như sau:

- **Ưu điểm**

- Độ phân giải và chất lượng hình ảnh của bộ dữ liệu này rất tốt, cho phép các mô hình segmentation học được rất nhiều chi tiết đặc trưng.
- Dữ liệu chứa sự đa dạng về ánh sáng, góc chụp, biểu cảm và các yếu tố khác, hỗ trợ các mô hình học được các đặc trưng tổng quát và có khả năng ứng dụng trong nhiều tình huống thực tế.
- Bộ dữ liệu được công bố rộng rãi và sử dụng làm benchmark trong nhiều nghiên cứu, tạo điều kiện so sánh hiệu năng của các mô hình khác nhau.
- **Nhược điểm:** Bộ dữ liệu này chỉ tập trung nhận diện vùng da mặt của một cá nhân nhất định chứ chưa có thể nhận diện nhiều vùng da mặt khác nhau có trong khung ảnh.

3.2.1.2 Pratheepan Dataset

Bộ dữ liệu Pratheepan được xây dựng với mục tiêu nghiên cứu việc phát hiện da người từ các hình ảnh gương mặt, được công bố lần đầu tiên trong bài báo **A Fusion Approach for Efficient Human Skin Detection** [6] vào năm 2012.

Các ảnh trong bộ dữ liệu này được tải ngẫu nhiên từ Google cho nghiên cứu phát hiện da người, bao gồm các ảnh chụp bằng nhiều loại máy ảnh khác nhau với các kỹ thuật cải thiện màu sắc và dưới các điều kiện ánh sáng khác nhau.

Bộ dữ liệu được tổ chức thành 4 thư mục chính:

- **FacePhoto:** chứa ảnh khuôn mặt của một cá nhân duy nhất, gồm 32 ảnh mẫu.
- **FamilyPhoto:** chứa ảnh nhiều đối tượng (gia đình), cấu trúc nền ảnh phức tạp hơn, bao gồm 46 ảnh mẫu.
- **GroundT_FacePhoto:** chứa ảnh groundtruth tương ứng cho thư mục FacePhoto.
- **GroundT_FamilyPhoto:** chứa ảnh groundtruth tương ứng cho thư mục FamilyPhoto.



Hình 3.4 Ảnh minh họa bộ dữ liệu Pratheepon

Bộ dữ liệu này cũng có những ưu điểm và khuyết điểm như sau:

- **Ưu điểm:** Bộ dữ liệu này không chỉ cung cấp vùng da cho từng cá nhân nhưng cung cấp vùng da cho nhiều người trong cùng một khung ảnh.
- **Nhược điểm:** Bộ dữ liệu này có số lượng ảnh hạn chế, với chỉ khoảng 78 ảnh, dẫn đến bộ dữ liệu này có thể không đủ lớn để huấn luyện các mô hình deep learning phức tạp, đặc biệt khi so sánh với các bộ dữ liệu quy mô lớn khác. Đồng thời, các annotation này được thực hiện trên vùng da toàn bộ các bộ phận cơ thể chứ không chỉ cho da mặt, nên sẽ có rất nhiều nhiễu và ảnh mask sẽ không được xử lý sạch sẽ.

3.2.1.3 Các bộ dữ liệu khác

Ngoài ra, nhóm cũng xem xét các bộ dữ liệu khác trên Kaggle như bộ Skin Tone Classification hoặc Monk Skin Tone Examples (MST-E) Dataset.

MST-E là một tập dữ liệu gồm các ví dụ của 19 người trải dài trên thang điểm MST 10 điểm. Nó chứa 1515 hình ảnh và 31 video. Mỗi người được chụp ảnh ở nhiều tư thế và điều kiện ánh sáng khác nhau và có/không có phụ kiện như mặt nạ và kính. Sau đó, Tiến sĩ Monk đã chú thích hình ảnh của những người này, cung cấp cho chúng tôi tông màu da MST thực tế. Tuy nhiên các bộ dữ liệu này chỉ tập trung về màu da trên gương mặt chứ chưa tập trung sâu sắc về vùng da người.

3.2.1.4 Kết luận

Nhóm sẽ tiếp tục khảo sát nhiều bộ dữ liệu nữa và kết hợp chúng lại để đạt được hiệu quả cao nhất đáp ứng yêu cầu bài toán trên.

3.2.2 Tiết xử lý dữ liệu

Đầu tiên, ta sẽ đọc ảnh gốc dưới định dạng RGB và ảnh mask dưới dạng grayscale với threshold. Nếu kích thước ảnh gốc không bằng với kích thước ảnh mask, thì ta sẽ resize ảnh mask theo kích thước ảnh gốc.

Sau đó, ta sẽ sử dụng albumentations để resize ảnh và mask theo kích thước tiêu chuẩn 256 x 256 cũng như áp dụng augmentation (horizontal flip) để tăng độ chính xác trong quá trình huấn luyện.

Sau đó ta sẽ trộn nhiều bộ dataset lại với nhau với kết hợp nhiều cặp (ảnh, mask) trên các folder tương ứng và nạp vào DataLoader để sử dụng trong quá trình huấn luyện và validation.

Trước khi huấn luyện, ta sẽ chia bộ dữ liệu gốc thành 3 tập khác nhau như tập huấn luyện với tỷ lệ 70%, tập validation với tỷ lệ 15% và tập testing với tỷ lệ 15%.

3.2.3 Fine-tuned mô hình phân đoạn UNet

Việc sử dụng mô hình UNet nhằm để phân đoạn vùng da trên mặt người. Mục tiêu phân đoạn nhằm để xác định vùng da mặt người một cách chính xác nhất, loại bỏ các không gian nền nhiễu hay các ánh sáng ngoại lai chiếu vào khuôn mặt làm giảm hiệu phát hiện khuôn mặt. Sau khi xác định được vùng da thì ta chỉ cần áp dụng các mô hình phát hiện gương mặt huấn luyện sẵn để xác định khuôn mặt nhằm làm tăng hiệu suất phát hiện khuôn mặt nhất có thể.

3.2.3.1 Kiến trúc mô hình

Trong kiến trúc mạng UNet này, nhóm em sẽ giữ nguyên kiến trúc UNet, cụ thể như sau:

- **Encoder Layers:** gồm 4 block, mỗi block gồm 2 lớp convolution kernel 3x3, được theo sau bởi các Batch Normalization, ReLU và một lớp Dropout 2d nhằm giảm overfitting. Sau mỗi block trừ block cuối, ta sẽ sử dụng một Max Pooling 2x2 để giảm kích thước không gian.
- **Bottleneck Layer:** là tầng nằm giữa encoder và decoder, có số lượng kênh gấp đôi so với encoder cuối cùng, nhằm giúp học các đặc trưng trừu tượng nhất.
- **Decoder Layers:** cũng gồm 4 block. Mỗi block bắt đầu bằng một lớp ConvTranspose2d (up-convolution) để tăng kích thước không gian gấp đôi, sau đó kết hợp (concatenate) kết quả với đầu ra của block encoder tương ứng (skip connection) nhằm tái sử dụng thông tin không gian chi tiết. Sau đó, áp dụng 2 lớp convolution (cùng cấu trúc như các block encoder: Conv2d, BatchNorm, ReLU, và Dropout2d

sau lớp đầu tiên).

- **Output Layer:** Một lớp convolution 1x1 được sử dụng để chuyển đổi số kênh từ số features của block cuối cùng của decoder thành số kênh mong muốn của output (trong trường hợp bài toán nhị phân segmentation, out_channels=1). Sau đó, áp dụng hàm sigmoid để chuyển các giá trị output về dạng xác suất cho mỗi pixel.

3.2.3.2 Huấn luyện mô hình

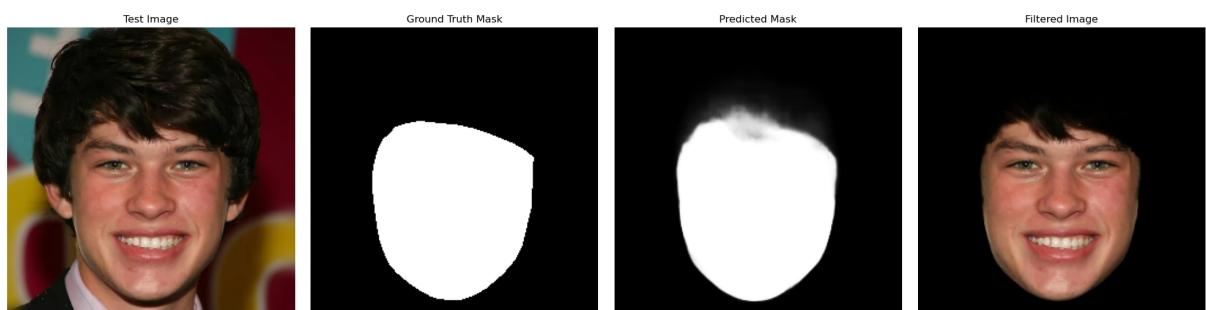
Trong quá trình huấn luyện cũng như đánh giá mô hình, chúng ta sẽ sử dụng các hàm mất mát như binary cross entropy để đo lường sự khác biệt pixel-wise cũng như Dice Loss để đo độ trùng khớp về mặt hình học của mask so với ảnh gốc. Đồng thời, ta cũng áp dụng các scheduler như ReduceLROnPlateau để giúp tự động giảm learning rate hay early stopping nhằm tránh overfitting.

3.2.3.3 Đánh giá mô hình

Sau khi huấn luyện, mô hình được đánh giá trên tập test để kiểm tra khả năng tổng quát hóa. Các metrics nhóm em dùng như sau:

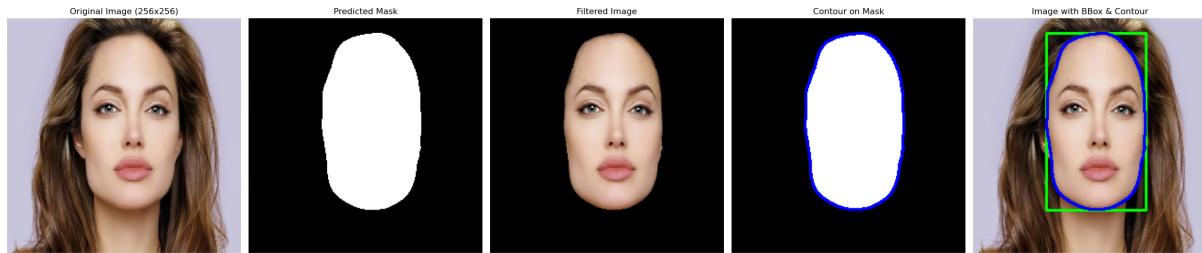
- **IoU (Intersection over Union):** xác định tỷ lệ giữa vùng giao nhau và vùng hợp nhất của dự đoán và nhãn thực tế.
- **Precision và Recall:** nhằm đánh giá độ chính xác và khả năng phát hiện đúng các vùng quan tâm.
- Các độ đo metrics khác cũng được sử dụng.

Hình 3.5 là ảnh kết quả dự đoán của mô hình sau khi được huấn luyện.



Hình 3.5 Kết quả dự đoán của mô hình

3.2.4 Xác định và hiển thị vùng ROIs da

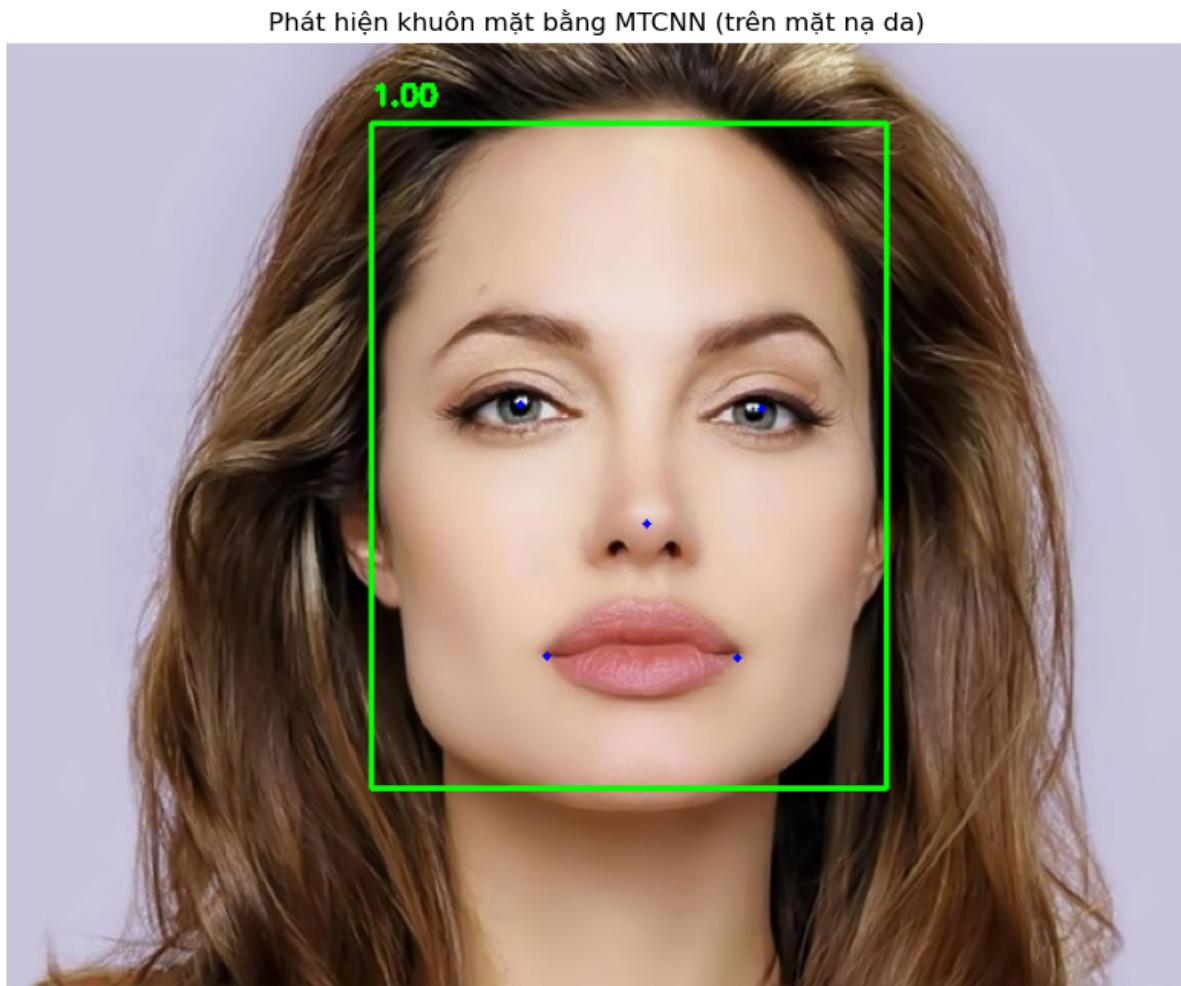


Hình 3.6 Ảnh sau khi được xử lý hiển thị vùng ROIs và contours

Sau khi ta đã phân đoạn được vùng da mặt, ta sẽ xác định vùng tọa độ ROIs vùng da trên ảnh mask và áp dụng lên ảnh gốc thông qua vùng liên thông, sau đó ta sẽ hiển thị lên ảnh gốc. Mục đích của bước này nhằm xác định hết tất cả các vùng da mặt trên ảnh nhằm loại bỏ đi các ảnh nền thừa và nhiễu bởi các chi tiết giống khuôn mặt.

Hình 3.6 là ảnh kết quả của bước này. Trong đó, ảnh ban đầu là ảnh gốc, ảnh tiếp theo là ảnh mask được dự đoán bởi mô hình, ảnh thứ ba là ảnh sau khi cắt bỏ vùng da mặt, ảnh thứ 4 là ảnh sau khi vẽ contours, và ảnh cuối cùng là ảnh kết quả của quá trình trên.

3.2.5 Áp dụng các mô hình pretrained phát hiện gương mặt



Hình 3.7 Kết quả cuối cùng của pipeline

Sau đó, cuối cùng ta sẽ dùng các mô hình phát hiện gương mặt trên vùng ROIs đã được xác định bước trên. Ảnh kết quả cuối cùng sẽ hiển thị confidence của mô hình sau khi detect gương mặt, các điểm landmarks lên ảnh gốc.

Việc áp dụng các mô hình pretrained hiện nay vẫn được nhóm em nghiên cứu và đưa ra hướng giải quyết, trong đó có hai mô hình phát hiện gương mặt được nhóm em quan tâm là RetinaFace và MTCNN, cả hai kiến trúc này đều đã được trình bày trong các phần trước đó.

3.3 Kết luận

Sau khi bàn bạc và thảo luận, nhóm chúng em quyết định dùng phương án cải tiến với mong phương án này có thể đạt được hiệu quả cao trong quá trình nghiên cứu.

CHƯƠNG 04. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Trong chương này, nhóm chúng em sẽ trình bày việc trình bày chương trình chạy kiểm thử như nào trong quá trình vấn đáp, cũng như phần này sẽ giải đáp câu hỏi "How will you conduct the demonstration?"

Nhóm chúng em sẽ xây dựng một ứng dụng đơn giản mang tên **Face Detection based on Skin Segmentation**.

Đầu tiên chương trình sẽ cho phép người dùng bấm nút quay video bản thân và chương trình sẽ hiện ra cửa sổ camera để quay hình ảnh nhận được.

Sau đó, chương trình sẽ qua các giai đoạn xử lý và trả ra kết quả cho người dùng tương ứng với đoạn video được segmentation gương mặt và được detect gương mặt với bounding box và contours.

Đồng thời, nhóm chúng em cũng sẽ hiện các bản đánh giá kết quả quá trình mô hình cũng như các số liệu thống kê chi tiết trong quá trình huấn luyện cũng như kiểm thử.



Hình 4.1 Ứng dụng phát hiện gương mặt

KẾT LUẬN VÀ ĐỀ NGHỊ

Kết luận chung

Sau một thời gian tìm hiểu và nghiên cứu đề tài này, nhóm em đạt được một số kết quả sau đây:

- Tìm hiểu được các hướng tiếp cận dùng màu sắc để phát hiện và nhận dạng gương mặt.
- Tìm hiểu về các phương pháp chuyển đổi kênh màu trong bài toán phát hiện gương mặt.
- Áp dụng một số kỹ thuật xử lý ảnh để xây dựng một chương trình thử nghiệm phát hiện khuôn mặt dựa vào màu da.
- Tìm hiểu các mô hình phân đoạn hình ảnh theo ngữ nghĩa hiện đại ngày nay và các mô hình phát hiện gương mặt kết hợp học sâu.

Tuy nhiên, vẫn còn tồn tại một số bất cập sau đây:

- Chương trình vẫn chưa xử lý được phần đoạn nhiều khuôn mặt trong cùng một khung ảnh.
- Dữ liệu vẫn còn thiếu và chưa thể gán nhãn label cho từng ảnh cụ thể.
- Chương trình chạy chậm do phải dùng nhiều model xử lý.
- Vẫn còn phải sử dụng nhiều model hiện đại nên dung lượng để lưu trữ dữ liệu cũng như GPUs cho quá trình huấn luyện còn hạn chế.

Hướng phát triển

Nhóm chúng em sẽ cố gắng thực hiện nhiều thực nghiệm trên nhiều bộ dataset khác nhau để chọn ra bộ dataset tối ưu nhất.

Thực nghiệm trên nhiều mô hình tốt hơn để đưa ra sự lựa chọn xác đáng nhất.

Nhóm cũng đã nghĩ đến hướng sau khi phát hiện gương mặt thì nhóm có thể làm thêm tác vụ để phát hiện màu da dựa trên khuôn mặt đã phát hiện. Bởi lẽ, sau khi xác định gương mặt thì việc phát hiện màu da cũng rất quan trọng trong các ứng dụng phân tích gương mặt.

Kiến nghị và đề xuất

Nhóm em cũng chưa chắc chắn hướng đi của nhóm sẽ là chính xác tuyệt đối cho bài toán phát hiện gương mặt dựa trên vùng da. Mong các thầy có thể đưa ra những lời khuyên bổ ích cho nhóm em.

TÀI LIỆU THAM KHẢO

- [1] B. Funt, K. Barnard, and L. Martin, “Is machine colour constancy good enough?” in *Computer Vision — ECCV’98*, H. Burkhardt and B. Neumann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 445–459.
- [2] A. K. Jain and S. Z. Li, *Handbook of Face Recognition*. Berlin, Heidelberg: Springer-Verlag, 2005.
- [3] J. B. Martinkauppi, A. Hadid, and M. Pietikäinen, *Skin Color in Face Analysis*. London: Springer London, 2011, pp. 223–249. [Online]. Available: https://doi.org/10.1007/978-0-85729-932-1_9
- [4] C. M. Laboratory. (2019) Celebamask-hq: A benchmark for high-quality face segmentation. Accessed: April 12, 2025. [Online]. Available: https://mmlab.ie.cuhk.edu.hk/projects/CelebA/CelebAMask_HQ.html
- [5] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, “Maskgan: Towards diverse and interactive facial image manipulation,” 2020. [Online]. Available: <https://arxiv.org/abs/1907.11922>
- [6] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, “A fusion approach for efficient human skin detection,” *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 138–147, 2012.