

BÁO CÁO TÌM HIỂU ĐỀ TÀI

NHẬN DIỆN KHUÔN MẶT

Face Detection & Face Recognition



Thực hiện: Phạm Đình Thục

Link Github: <https://github.com/phamdinhtduc020100/Project-FaceRecognizer.git>

Tháng 11 năm 2021

Mục lục

A. LEVEL 01	3
I. Đặt vấn đề	3
II. Mục đích tìm hiểu	3
III. Quá trình tìm hiểu	3
IV. Nội dung tìm hiểu	3
B. LEVEL 02	5
I. Đặt vấn đề	5
II. Mục đích tìm hiểu	6
III. Quá trình tìm hiểu	6
IV. Nội dung tìm hiểu	6
C. LEVEL 03	10
I. Đặt vấn đề	10
II. Mục đích tìm hiểu	10
III. Quá trình tìm hiểu	10
IV. Nội dung tìm hiểu	10
V. Tài liệu tham khảo	13

A. LEVEL 01

I. Đặt vấn đề

Vấn đề nhận dạng khuôn mặt hiện nay đang rất được quan tâm, nó được sử dụng cho nhiều mục đích như ngăn ngừa tội phạm, bảo vệ dữ liệu và thông tin cá nhân thông qua sinh trắc học

Hiện nay, thuật toán nhận dạng đã được các nền tảng lớn sử dụng thường xuyên hơn, như Facebook đã dùng thuật toán nhận dạng khuôn mặt để gắn thẻ bạn bè, người thân trên các tấm ảnh người dùng upload với độ chính xác 98% [1]

II. Mục đích tìm hiểu

- Nắm được các lý thuyết cơ bản của nhận diện khuôn mặt
- Tìm hiểu mức cơ bản quá trình nhận diện khuôn mặt bằng webcam
- Demo mức cơ bản

III. Quá trình tìm hiểu

- Đọc, hiểu tài liệu liên quan đến Nhận diện khuôn mặt
- Áp dụng thư viện OpenCV để chạy mô hình nhận diện khuôn mặt
- Demo cơ bản bằng webcam trên laptop

IV. Nội dung tìm hiểu

Hệ thống nhận dạng khuôn mặt là một ứng dụng máy tính tự động xác định hoặc nhận dạng một người nào đó từ một bức hình ảnh kỹ thuật số hoặc một khung hình video từ một nguồn video. Một trong những cách để thực hiện điều này là so sánh các đặc điểm khuôn mặt chọn trước từ hình ảnh và một cơ sở dữ liệu về khuôn mặt

Hiện nay có nhiều thuật toán, công cụ hỗ trợ cho huấn luyện, thực thi các mô hình nhận diện khuôn mặt. Có thể kể đến như: OpenCV, MTCNN, FaceNet, ...

Nhận diện khuôn mặt gồm có 2 công việc chính:

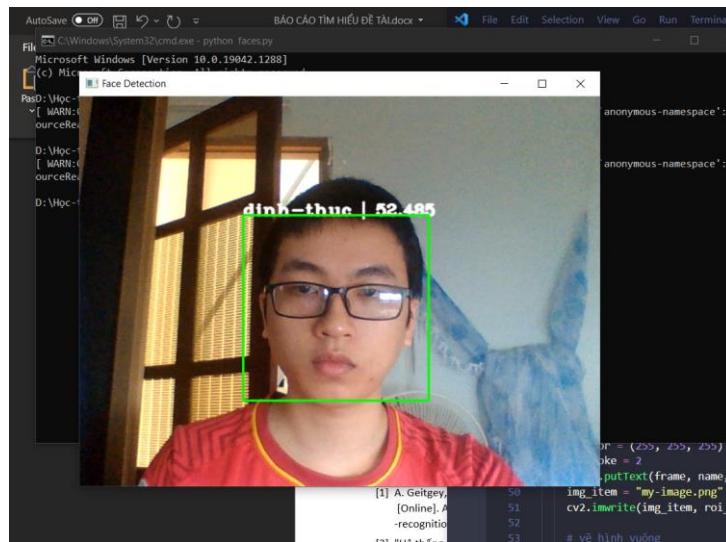
- Phát hiện khuôn mặt
- Nhận diện khuôn mặt đó

Các bước chính của nhận diện khuôn mặt

- Bước 1. Chuẩn bị một cơ sở dữ liệu để lưu trữ thông tin của mỗi người như: tên, tuổi, chức danh...
- Bước 2. Thực hiện chụp ảnh hoặc tải về ảnh có sẵn
- Bước 3. Train model để nhận diện
- Bước 4. Tiến hành nhận dạng khuôn mặt

Ảnh Demo

- Ta demo trên tập dữ liệu nhỏ (5 người) cho ra kết quả

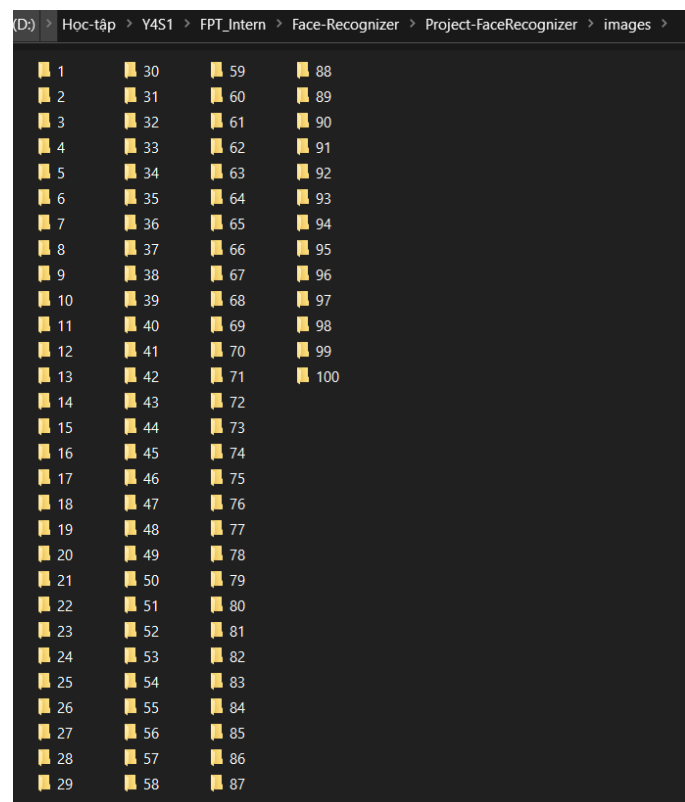


- Tiếp theo ta test trên bộ dữ liệu lớn hơn (khoảng 100 người)
 - Dataset được chọn là: VN-celeb dataset với khoảng 23k khuôn mặt của hơn 1000 người Việt [2]. Ở đây ta dùng 99 tập của 99 người và 1 tập của cá nhân. Tổng cộng 100 tập

Cấu trúc file

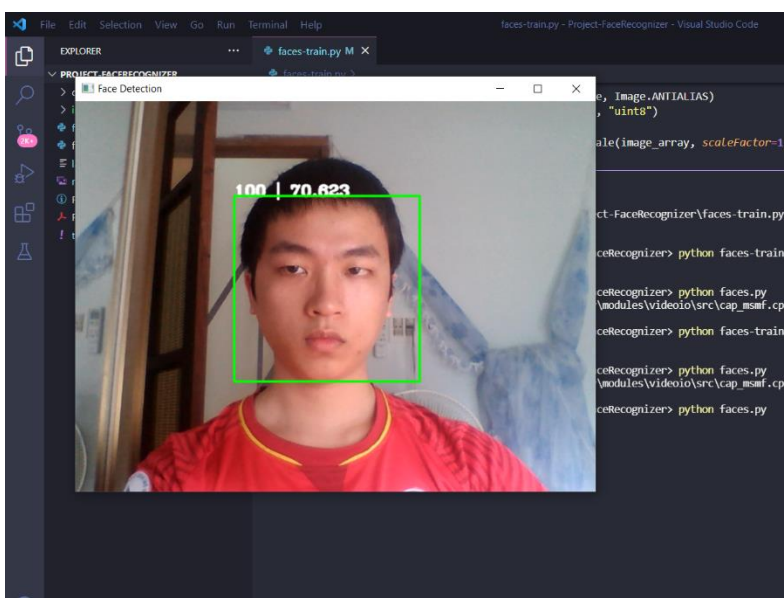
Folder 1 – 99: Người Việt nổi tiếng

Folder 100: đình thực

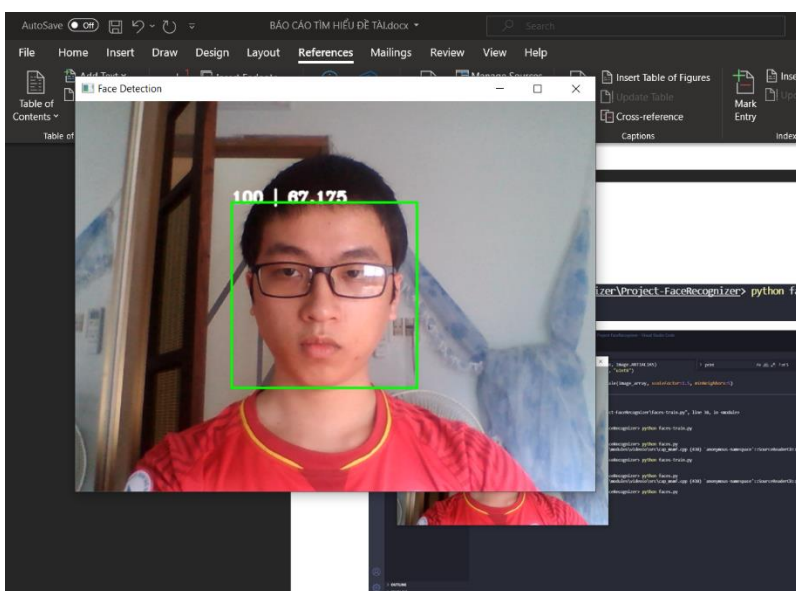


Kết quả test

```
PS D:\Hoc-tap\Y4S1\FPT_Intern\Face-Recognizer\Project-FaceRecognizer> python faces-train.py
Loading ...
finished
```



(Không đeo kính)



(Có đeo kính)

B. LEVEL 02

I. Đặt vấn đề

Với tình hình dịch bệnh COVID-19 trên toàn cầu hiện nay, cùng với đó việc áp dụng các thuật toán nhận diện khuôn mặt truyền thống đang gặp các khó khăn như: người dân đeo khẩu trang, đeo kính và tuân thủ khoảng cách an toàn đang là tiêu chuẩn mới nhằm ngăn ngừa dịch bệnh. Điều này khiến khuôn mặt của họ bị che đi, không còn nhiều các đặc điểm đặc trưng để có thể nhận dạng nhanh chóng. NIST đã tiến hành các bài kiểm tra độ chính xác của nhận dạng khuôn mặt gần đây. Báo cáo của họ tiết lộ rằng các thuật toán FR chính thống đã tăng tỷ lệ lỗi từ 5% đến 50% trên khuôn mặt có đeo khẩu trang. [3]

Vì thế cần có một phương pháp mới để có thể nhận diện một người nào đó dựa vào một số ít các đặc trưng tiêu biểu

II. Mục đích tìm hiểu

- Hiểu được các lý thuyết cơ bản về nhận dạng khuôn mặt cải tiến (khi đeo khẩu trang)
- Hiểu được cách các mô hình học máy, học sâu được áp dụng

III. Quá trình tìm hiểu

- Đọc hiểu đề tài
- Tìm kiếm thông tin đề tài
- Remake source code (nếu có) để demo

IV. Nội dung tìm hiểu

Facial Recognition for People Wearing Masks

- Các bước cho quá trình nhận dạng khuôn mặt cải tiến (khi đeo khẩu trang)
 - **Thu thập dữ liệu**
Đầu tiên, ta cần thu thập hình ảnh liên quan đến khuôn mặt người khi đeo và không đeo khẩu trang
Có thể thu thập bằng nhiều cách: tải các bộ dataset có sẵn, hoặc crawl trên web bằng các thư viện như Selenium hoặc BeautifulSoup
 - **Tiền xử lý**
Ở bước này, ta tiến hành các thao tác xử lý ảnh (rotate, crop, ...) để đưa ảnh về kích thước phù hợp cho quá trình huấn luyện
Để làm giàu dữ liệu ta có thể sử dụng các phương pháp tăng số lượng và chất lượng ảnh như: Mask the face, ...

Phương pháp Mask the face [4]:

- Nó sử dụng một máy dò các cạnh khuôn mặt dựa trên dlib để xác định độ nghiêng của khuôn mặt và các đặc điểm chính của khuôn mặt cần thiết để đeo khẩu trang vào (khẩu trang ảo). Khẩu trang làm mẫu sẽ được biến đổi theo các hướng của khuôn mặt sao cho khớp nhất có thể. Tham khảo mô hình MaskTheFace

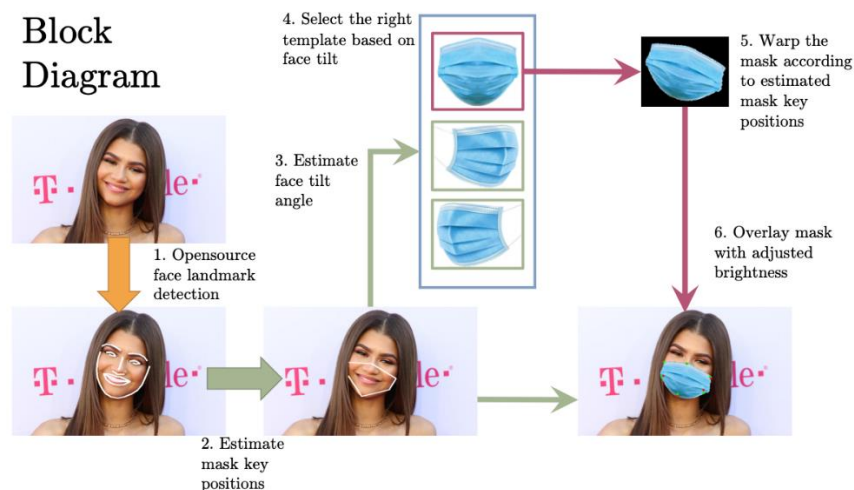




Figure 1: Dataset Generation: Original => Masked => Cropped

- Trong khi sử dụng dlib để tạo mặt nạ, chúng ta ghi lại vị trí của mặt nạ và tạo một binary segmentation maps ứng với từng ảnh để loại bỏ khẩu trang

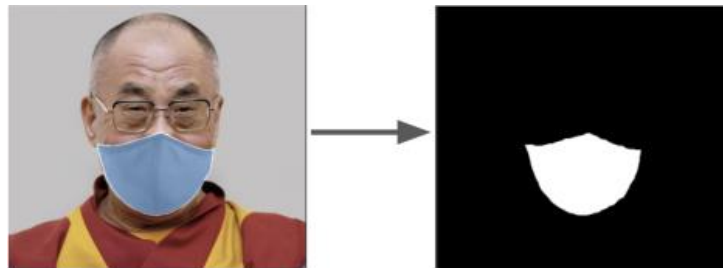


Figure 2: Mask Segmentation

- Ở đây, tác giả dùng nhiều mẫu khẩu trang để làm đa dạng bộ dữ liệu, và thuật toán được điều chỉnh để có thể gắn khẩu trang cho nhiều người trên một khung hình

Ngoài ra ta có thể sử dụng bộ dữ liệu có sẵn: Masked faces in real world for face recognition (MFR2) với 53 danh tính của những người nổi tiếng và chính trị gia với tổng số 269 hình ảnh được thu thập từ internet, mỗi người có trung bình 5 ảnh.



○ Xây dựng mô hình

- Tác giả thử nghiệm trên 2 mô hình: **VGG** and **ResNet**. Kết quả cho ra VGG không thể hội tụ trên tập dữ liệu với các siêu tham số mà nhóm tác giả đã thử nghiệm. ResNet đã hoạt động tốt hơn. Vì thế, **ResNet50** đã được lựa chọn làm phương thức chính cho mô hình này
- Cụ thể, sẽ mở rộng ResNet50 với 1 lớp Dropout và dùng 1 lớp tuyến tính mới để tạo các bản nhúng 512 bit
- Lớp FC cuối, được cấu hình như bản sau

ID	Layers	In	Out	Kernel Size
0	Dropout			
1	AdaptiveAvgPool2d	2048	2048	1×1
2	Flatten			
3	Linear	2048	512	
4	BatchNorm1d			

Table 1: Extension layers

○ Loss function

- Các hệ thống FR truyền thống như FaceNet sử dụng chức năng Triplet Loss. Ở đây tác giả đề xuất một Loss function mới được gọi là **Quintuplet Loss** để xử lý các hình ảnh có đeo và không đeo khẩu trang.
- Quintuplet Loss gồm 5 ảnh: Anchor (A), Positive (P), Negative (N), Masked Positive (Pm) and Masked Negative(Nm).
 - **Anchor** là hình ảnh được nhận dạng
 - **Positive** là hình ảnh không đeo khẩu trang khác của cùng một người
 - **Negative** là hình ảnh không đeo khẩu trang của một người khác
 - **Masked Positive** là hình ảnh đeo khẩu trang của cùng một người
 - **Masked Negative** là hình ảnh đeo khẩu trang của một người khác
 - α là margin

The Loss Function is:

$$\mathcal{L} = ||f(A) - f(P)||^2 + ||f(A) - f(P_m)||^2 - ||f(A) - f(N)||^2 - ||f(A) - f(N_m)||^2 + \alpha$$

Để tránh mô hình bị model collapsing, ta cần chọn Quintuplets phù hợp (không quá khó, không quá dễ). Nếu quá dễ dàng, chúng sẽ không giúp train neural network. Nếu quá khó, chúng có thể dẫn đến cực tiểu cục bộ hoặc model collapse. Để tối ưu, ta cần chọn Quintuplets với điều kiện:

$$0 < ||f(A) - f(N)||^2 - ||f(A) - f(P)||^2 < \alpha$$

$$0 < ||f(A) - f(N_m)||^2 - ||f(A) - f(P_m)||^2 < \alpha$$

○ Xóa khẩu trang và tái tạo hình ảnh (MRII)

- Thực hiện xóa khẩu trang và tái tạo hình ảnh trước khi đưa hình ảnh vào mô hình FR.
- Một pretrained generative model được tạo từ Edgeconnect để tháo khẩu trang của hình ảnh.

- Binary mask segmentation map sẽ được tạo cùng với cùng với hình ảnh trong quá trình mask the face. Ngoài ra ta cũng có thể dùng các thuật toán object detection và localization algorithm.

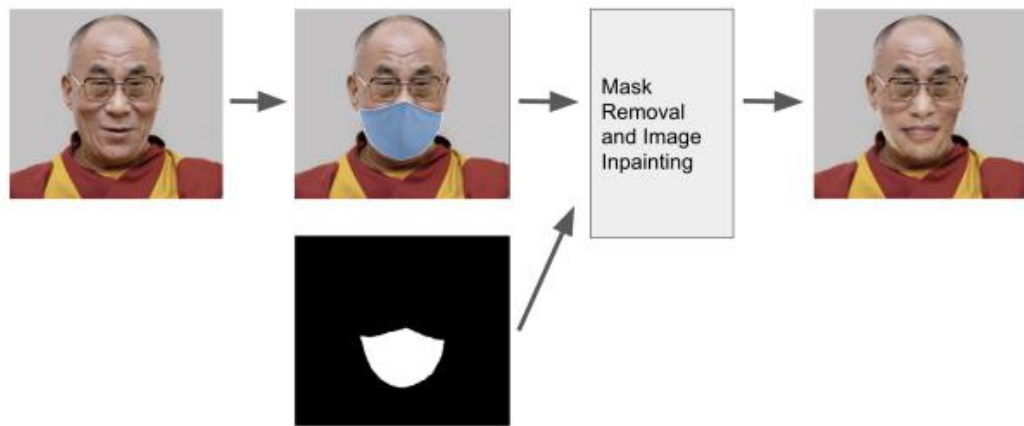


Figure 3: Mask removal and image inpainting

- Quá trình được diễn ra theo các bước
 - o Từ ảnh không gốc đeo (hoặc không) khẩu trang
 - o Thông qua quá trình mask the face (đối với ảnh không đeo khẩu trang) để tạo Binary mask segmentation map
 - o Đối với các ảnh đã đeo khẩu trang, dùng các thuật toán object detection và localization algorithm để tạo Binary mask segmentation map
 - o Tổng hợp hình ảnh đeo khẩu trang cùng với Binary mask segmentation map để gỡ bỏ khẩu trang và tái tạo lại khuôn mặt
 - o Cuối cùng, ta đưa kết quả đến FaceNet để dự đoán
 - o **Kết quả**

Model	FaceNet	MaskFaceNet	MRII
Accuracy	67.28%	75.06%	82.39%

Table 2: Evaluation

- Mô hình **FaceNet** hoạt động không tốt lắm vì nó chưa được đào tạo với hình ảnh khuôn mặt bị che.
- Mô hình **MaskFaceNet** hoạt động tốt hơn FaceNet, cho thấy quintuplet loss function giúp mô hình nhận ra hình ảnh khuôn mặt bị che. Độ chính xác tổng thể của mô hình không cao lắm vì phải sử dụng batch size nhỏ, bị giới hạn bởi bộ nhớ GPU và thời gian đào tạo cũng ít hơn nhiều so với các mô hình như FaceNet. Với nhiều tài nguyên máy tính hơn, MaskFaceNet sẽ có hiệu suất tốt hơn
- Mô hình **Mask removal and image inpainting (MRII)** cải thiện đáng kể hiệu suất của FaceNet trên tập dữ liệu khuôn mặt bị che. Nó cho thấy rằng một số đặc điểm trên khuôn mặt được phục hồi bằng kỹ thuật inpainting có thể giúp hệ thống FR nhận ra hình ảnh khuôn mặt bị che

C. LEVEL 03

I. Đặt vấn đề

Các mô hình nhận dạng khuôn mặt hiện tại có tính tổng quát và mạnh mẽ, nhưng chúng vẫn dễ dàng xác định nhầm các khuôn mặt bị lệch và có tư thế lạ.

Trong tình hình dịch COVID-19, mọi người luôn đeo khẩu trang để bảo vệ mình khỏi bị lây nhiễm, điều này rất khó để các mô hình hiện có xác định chính xác. Mặt nạ có thể dẫn đến các vấn đề sau:

1. Không nhìn thấy được đặc điểm kết cấu của vùng mặt được che.
2. Các khu vực được che cho những người khác nhau hoặc thậm chí giống nhau không phải lúc nào cũng nhất quán

Các công trình trước đây có thể được tóm tắt thành ba loại:

- **Phương pháp matching-based:** Các phương pháp matching-based cố gắng so sánh mức độ giống nhau giữa các hình ảnh bằng quy trình matching, nhưng hiệu suất quá nhạy cảm với các chiến lược lấy mẫu khác nhau
- **Phương pháp restoration-based:** Các phương pháp dựa trên khôi phục nhằm mục đích khôi phục các vùng bị che của hình ảnh thăm dò theo các hình ảnh trong bộ sưu tập, nó làm giảm phần lớn khó khăn trong việc nhận dạng
- **Phương pháp occlusion removal-based:** Để giảm ảnh hưởng của occluded regions, các phương pháp occlusion removal-based trước tiên sẽ phát hiện các occluded regions và sau đó loại bỏ nó trực tiếp. Tuy nhiên, việc loại bỏ các vùng bị che khuất có thể phá vỡ đặc điểm hình dạng của khuôn mặt

II. Mục đích tìm hiểu

Tìm kiếm các phương pháp khác với những phương pháp đã tìm hiểu, nhằm hiểu sâu, rõ hơn về các thuật toán nhận dạng khuôn mặt trong các điều kiện đặc biệt

III. Quá trình tìm hiểu

- Đọc hiểu yêu cầu đề tài
- Tìm kiếm thông tin đề tài
- Mô tả lại các thông tin hiểu được từ quá trình tìm hiểu đề tài
- Trình bày lại các ý đã tiếp thu được

IV. Nội dung tìm hiểu

Mask Aware Network for Masked Face Recognition in the Wild

Ta đề xuất **Mask Aware Network (MAN** - giải pháp này đạt được 5 vị trí đầu tiên và 1 vị trí thứ ba trong ICCV Masked Face Recognition WebFace), bao gồm các mô-đun tìm kiếm loss function và tạo khẩu trang. Cụ thể, với một hình ảnh, trước tiên, ta phát hiện các điểm mốc trên khuôn mặt và sau đó thêm khẩu trang vào một khuôn mặt nhất định với tỷ lệ ngẫu nhiên (khẩu trang một nửa và toàn bộ)

Các điểm mốc trên khuôn mặt được phát hiện có thể được coi là điểm neo cho việc đeo khẩu trang. Như trong Hình 1, khác với các phương pháp đeo khẩu trang khác, cách đeo này linh hoạt và thay đổi hơn. Sau đó, ta cung cấp những hình ảnh này vào **Convolutional Neural Network (CNN)** để trích xuất đặc điểm khuôn mặt. Để tăng cường khả năng phân biệt đối tượng, ta sử dụng LFS để tìm kiếm loss function phù hợp nhất cho tác vụ này.



Figure 1. The top row is the official mask augmentation, the bottom row is our mask augmentation. It is obviously that our augmentation can produce more variable masked faces than official method.

Các bước thực hiện chính:

1. Đề xuất Mask Aware Network để nhận dạng khuôn mặt bị che, bao gồm hai mô-đun được thiết kế tốt có tên là **Mask Generation** và **Loss Function Searching Modules**.
2. Mô-đun Mask Generation sử dụng các điểm mốc trên khuôn mặt làm điểm neo để cải thiện sự đa dạng của các khuôn mặt bị che, đây là yếu tố quan trọng đối với hiệu suất mô hình của chúng tôi.
3. Mô-đun Loss Function Searching đảm bảo rằng chúng ta có thể có được loss function phù hợp nhất cho vấn đề nhận dạng khuôn mặt bị che.

• Tổng quan về MAN

Trước tiên ta xác định training and validation sets. Đối với mỗi hình ảnh trong training set, mô-đun tạo mặt nạ sử dụng các điểm mốc trên khuôn mặt được phát hiện để tạo một khuôn mặt được che một cách ngẫu nhiên, có thể cung cấp nhiều khuôn mặt được che hơn.

Sau đó, ta đưa khuôn mặt đeo (hoặc không đeo) khẩu trang thành nhiều several backbones để trích xuất các đặc trưng. Lưu ý rằng tất cả các several backbones đều có cùng cấu trúc nhưng khác nhau về các tham số.

Để khám phá **loss** phù hợp nhất, ta xác định loss function pool trong đó các loss functions khác nhau được khởi tạo với cùng một α . α_i đại diện cho xác suất sử dụng loss function thứ i trong corresponding backbone, được xác định bởi hiệu suất của corresponding backbone tương ứng trên validation set.

Cuối cùng, ta đồng bộ hóa các thông số backbone của hiệu suất tốt nhất trên validation set với các tham số backbone khác.

- **Mask Generation Module**

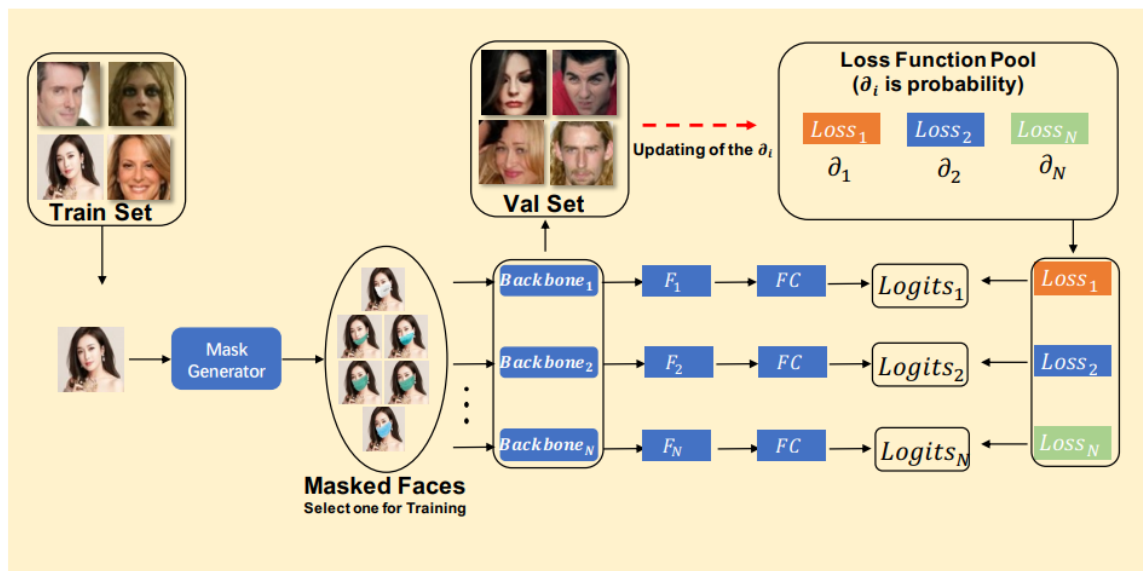


Figure 2. The pipeline of our proposed mask aware network. Given an image from train set, the online mask generator first add mask with 20% probability. Then, the image (or masked image) is fed into several backbones to extract the face recognition features. Note that, the backbones are with the same structure and supervised by different loss functions. The loss function pool contains several loss functions. The using probability of each function is initialized with the same α . The updating of α is determined by the results on the validation set. If the loss function achieves more better results on validation set, the using probability is updated with a larger value.

Để tạo ra sự thay đổi của khuôn mặt, khác với các phương pháp đeo khẩu trang, chúng tôi sử dụng các điểm mốc trên khuôn mặt để tạo ra khuôn mặt được che. Quá trình này có thể được xây dựng như sau,

$$M = G(I, L; \theta)$$

Trong đó, G hàm thêm khẩu trang, I là hình ảnh đầu vào, θ là các thông số của operation

Ta sử dụng máy dò tìm điểm mốc khuôn mặt để phát hiện 96 điểm mốc trên khuôn mặt và sau đó chọn 10 điểm mốc trên khuôn mặt (5 điểm bên trái và 5 điểm bên phải). Dựa vào 10 mốc trên khuôn mặt, chúng ta có thể điều chỉnh vùng đeo khẩu trang một cách ngẫu nhiên

- **Loss Function Searching Module**

- a. **Kết luận**

Trong đề tài này, ta đã nghiên cứu các cách tiếp cận để cải thiện hiệu suất của hệ thống FR trên hình ảnh khuôn mặt của những người đeo khẩu trang. Kết quả đánh giá cho thấy các cách tiếp cận đều có thể cải thiện độ chính xác của hệ thống FR trên hình ảnh khuôn mặt được che mặt. Ta có thể đạt được sự cải thiện hơn nữa nếu có nhiều tài nguyên máy tính hơn hoặc đào tạo lại một số mô hình.

V. Tài liệu tham khảo

- [1] A. Geitgey, "Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning," [Online]. Available: <https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78>.
- [2] "VN-celeb: Dữ liệu khuôn mặt người nổi tiếng Việt Nam và bài toán Face recognition," [Online]. Available: <https://viblo.asia/p/vn-celeb-du-lieu-khuon-mat-nguoi-noi-tieng-viet-nam-va-bai-toan-face-recognition>.
- [3] "Facial Recognition for People Wearing Masks," [Online]. Available: http://cs230.stanford.edu/projects_fall_2020/reports/55790723.pdf.
- [4] "MaskTheFace," [Online]. Available: <https://github.com/aqeelanwar/MaskTheFace>.
- [5] "Hệ thống nhận dạng khuôn mặt," [Online]. Available: https://vi.wikipedia.org/wiki/H%E1%BB%87_th%E1%BB%91ng_nh%E1%BA%ADn_d%E1%BA%A1nh_khu%E1%BA%B7t.
- [6] "[Face Recognize] Thử làm hệ thống chấm công bằng nhận dạng khuôn mặt," [Online]. Available: <https://www.miai.vn/2019/08/13/face-recognize-thu-lam-he-thong-cham-cong-bang-nhan-dang-khuon-mat>.
- [7] "OpenCV Python TUTORIAL #4 for Face Recognition and Identification," [Online]. Available: https://www.youtube.com/watch?v=PmZ29Vta7Vc&ab_channel=CodingEntrepreneurs.