

Bài toán

Hướng giải
quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát
triển

Face Mask Detection

MobileNetV2

Nguyễn Duy Nhất, Huỳnh Yến Nhi

Khóa học Python & Machine Learning 2025
HCMUT EE Machine Learning & IoT Lab

06/09/2025

Nội dung

Bài toán

Hướng giải
quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát
triển

Bài toán

Hướng giải quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát triển

- Trong thời kỳ dịch bệnh (ví dụ COVID-19), việc đeo khẩu trang là cần thiết để đảm bảo an toàn.
- Việc giám sát thủ công (bảo vệ, camera thường) vừa tốn công, vừa thiếu hiệu quả.
- Cần một giải pháp tự động giúp phát hiện nhanh, chính xác tình trạng đeo khẩu trang.

Xây dựng mô hình có khả năng phân loại hình ảnh khuôn mặt thành 2 lớp: Có khẩu trang và Không có khẩu trang.

- Ứng dụng Deep Learning để tự động nhận diện tình trạng đeo khẩu trang.
- Sử dụng mô hình CNN pretrained (MobileNetV2) để trích xuất đặc trưng khuôn mặt.
- Huấn luyện trên tập dữ liệu ảnh đã được tăng cường (augmentation) để đảm bảo độ chính xác cao.

Dữ liệu

Bài toán

Hướng giải quyết

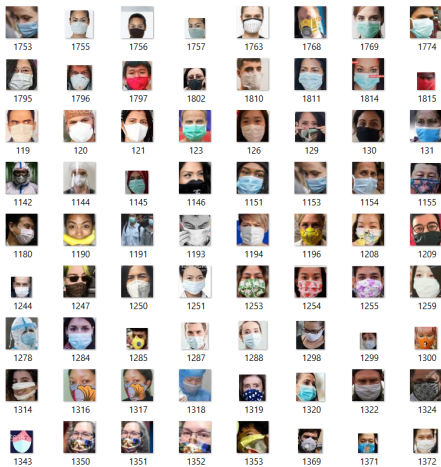
Dữ liệu

Mô hình

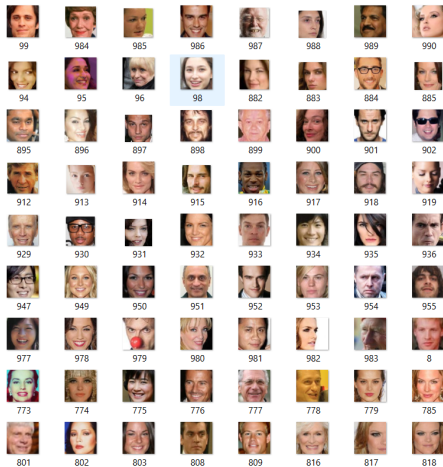
Kết quả

Demo

Hướng phát triển



With Mask



Without Mask

Dataset có khoảng 12.000 ảnh (10.000 train, 1000 validation, 1.000 test) từ Kaggle – Face Mask 12k Images Dataset, gồm 2 lớp: With Mask / Without Mask

- Resize ảnh về kích thước (224, 224).
- Chuẩn hóa giá trị pixel [0,1].
- Data Augmentation: xoay, phóng to/thu nhỏ, dịch chuyển, lật ngang, thay đổi độ sáng... → tăng tính đa dạng dữ liệu.

Lí do chọn mô hình MobileNetV2:

- Nhẹ, nhanh, kiến trúc tối ưu cho thiết bị di động, chạy real-time trên CPU/GPU yếu.
- Độ chính xác đủ cao với bài toán 2 lớp (có/không khẩu trang)
- Với dataset đơn giản như trên, MobileNetV2 chỉ cần Stage 1 (train head) là đã đạt $>99\%$ accuracy.
- Cân bằng tốt giữa tốc độ, độ chính xác và tài nguyên so với ResNet, VGG, Inception.

Mô hình

Bài toán

Hướng giải quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát triển

Tiêu chí	MobileNetV2	EfficientNetB0
Kích thước	Nhỏ, rất nhẹ	Lớn hơn, trung bình đến nặng
Tốc độ suy luận	Rất nhanh, chạy tốt trên CPU/GPU yếu	Nhanh vừa, cần phần cứng mạnh hơn
Độ chính xác	Đủ cao cho bài toán đơn giản	Cao hơn, đặc biệt trên dataset phức tạp
Triển khai	Tốt trên thiết bị di động, nhúng	Tốt nhưng yêu cầu tài nguyên nhiều hơn
Ứng dụng	Real-time, IoT, camera giám sát	Các bài toán cần độ chính xác tối đa

Bảng 1: So sánh MobileNetV2 và EfficientNetB0

- Đã huấn luyện trên 2 mô hình: MobileNetV2 và EfficientNetB0.
- Kết quả:
 - MobileNetV2: Stage 1 đạt 99.29%, fine-tuning Stage 2 đạt 99.90%.
 - EfficientNetB0: Fine-tuning đạt 99.90%.
- Nhận xét: Accuracy khi fine-tuning tương đương nhau.

Kết luận:

Chọn MobileNetV2 vì model nhẹ, tốc độ huấn luyện và suy luận nhanh, dễ dàng đạt độ chính xác cao chỉ với 1 stage, phù hợp triển khai thực tế.

Mô hình

Bài toán

Hướng giải quyết

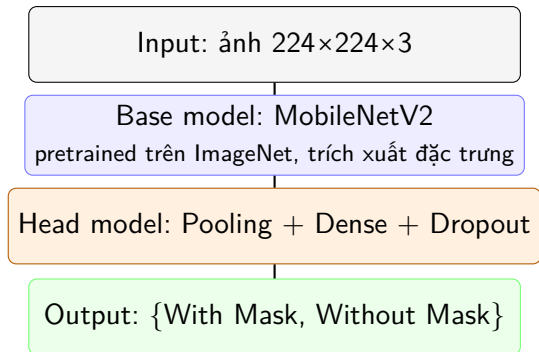
Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát triển



Cấu trúc model

- Sử dụng MobileNetV2 làm base model, bỏ tầng fully connected, giữ trọng số ImageNet
- Thêm head model:
 - Global Average Pooling để giảm số chiều
 - Dense 128 neuron (ReLU) + Dropout 0.5
 - Dense 2 neuron (Softmax) cho phân loại có / không khẩu trang
- Gộp phần trích xuất đặc trưng (MobileNetV2) + phân loại (head model) thành 1 mạng hoàn chỉnh.

Stage 1: Freeze MobileNetV2

- Đóng băng toàn bộ backbone (không train)
- Chỉ huấn luyện head classifier
- Optimizer Adam, learning rate = $1e-4$, early stopping

Stage 2: Fine-tune 50 layer cuối

- Mở khóa 50 layer cuối của MobileNetV2
- Huấn luyện lại cùng với head
- Optimizer Adam, learning rate = $1e-5$, early stopping

Các giai đoạn huấn luyện

Kết quả

Bài toán

Hướng giải quyết

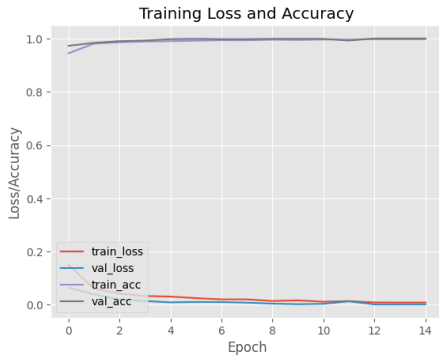
Dữ liệu

Mô hình

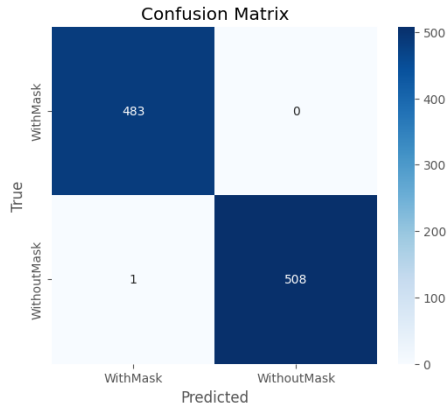
Kết quả

Demo

Hướng phát triển



Training Curve



Confusion Matrix

Kết quả

Bài toán

Hướng giải quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát triển

- Độ chính xác (Accuracy): 99.90%
- Precision & Recall cho cả 2 lớp đều đạt 1.00
- Mô hình phân loại cân bằng, không thiên lệch giữa Mask và No Mask

Class	Precision	Recall	F1-score	Support
WithMask	1.00	1.00	1.00	483
WithoutMask	1.00	1.00	1.00	509
Accuracy			1.00	992
Macro avg	1.00	1.00	1.00	992
Weighted avg	1.00	1.00	1.00	992

Demo

Bài toán

Hướng giải
quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát
triển



Dự đoán: WithoutMask (99.99%)



Dự đoán: WithMask (100.00%)

Demo

Bài toán

Hướng giải
quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát
triển



Dự đoán: WithMask (100.00%)



Dự đoán: WithoutMask (100.00%)

Web streamlit gồm 2 chức năng chính:

- Nhận diện qua ảnh: Upload ảnh từ máy tính hoặc Chụp ảnh trực tiếp từ webcam.
- Nhận diện realtime bằng webcam.

The screenshot shows a web application interface with a sidebar menu on the left and a main content area. The sidebar menu is titled "Menu" and contains two options: "Ảnh" (selected with a red radio button) and "Webcam" (unselected with a white radio button). The main content area has a title "Nhận diện khẩu trang - Test ảnh" (Face Mask Detection - Test image). Below the title, there are two radio buttons for "Chọn chế độ:" (Select mode): "Upload ảnh" (selected with a red radio button) and "Chụp ảnh webcam" (unselected with a white radio button). Below these, there is a section for "Chọn ảnh" (Select image) with a drag-and-drop area that says "Drag and drop file here" and "Limit 200MB per file • JPG, JPEG, PNG". To the right of the drag-and-drop area is a button labeled "Browse files". In the top right corner of the main content area, there is a "Deploy" button and a menu icon.

Kết hợp Caffe Face Detector để nhận diện nhiều khuôn mặt

- Kiến trúc ResNet-10 SSD (Single Shot Multibox Detector).
- Huấn luyện trên tập WIDER FACE.
- Được OpenCV phát hành, tải từ `opencv_face_detector` trong model zoo.
- File cấu hình: `deploy.prototxt`.
- Trọng số: `res10_300x300_ssd_iter_140000.caffemodel`.

Cách kết hợp:

- Caffee SSD phát hiện tất cả khuôn mặt trong ảnh hoặc khung hình webcam.
- Crop và resize từng khuôn mặt về (224×224) .
- Đưa vào Model đã train để phân loại Mask / No Mask.
- Vẽ bounding box và nhãn cho nhiều người cùng lúc.

Ưu điểm: phát hiện nhanh, hỗ trợ nhiều khuôn mặt đồng thời, trực quan trên ảnh/video.

Hướng phát triển²

Bài toán

Hướng giải
quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát
triển

Trường hợp còn sai

Dự đoán: WithMask (100.00%)



Lấy áo che mặt

Dự đoán: WithMask (99.99%)



Đeo sai cách

Hướng phát triển

Bài toán

Hướng giải
quyết

Dữ liệu

Mô hình

Kết quả

Demo

Hướng phát
triển

- Mở rộng nhãn để nhận diện cả trường hợp đeo sai cách.
- Bổ sung dữ liệu có hành vi che mặt bằng tay hoặc áo để tránh nhầm lẫn.
- Triển khai trên thiết bị nhúng hoặc hệ thống camera giám sát thực tế.