

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN, ĐHQG-HCM**

**KHOA KHOA HỌC MÁY TÍNH**



**BÁO CÁO BÀI TẬP MÔN HỌC  
BÀI TẬP : OBJECT DETECTION**

**Môn học: CS406.Q11 – XỬ LÝ ẢNH VÀ ỨNG DỤNG**

**Giảng viên hướng dẫn: Cáp Phạm Đình Thăng**

**MSSV : 22520907**

**Họ và Tên sinh viên : Đặng Văn Nam**

## 1. Giới thiệu bài toán

Trong bối cảnh đại dịch COVID-19 và các bệnh hô hấp khác, việc đeo khẩu trang đúng cách đã trở thành một biện pháp quan trọng để ngăn chặn sự lây lan của virus. Tuy nhiên, việc giám sát thủ công tình trạng đeo khẩu trang ở các khu vực công cộng như trường học, trung tâm thương mại hoặc phương tiện giao thông công cộng là không khả thi do tốn kém về nhân lực và thời gian. Do đó, việc phát triển các hệ thống tự động dựa trên trí tuệ nhân tạo (AI) để phát hiện và phân loại tình trạng đeo khẩu trang là cần thiết.

Bài toán object detection (phát hiện đối tượng) ở đây tập trung vào việc xác định vị trí khuôn mặt trong ảnh và phân loại chúng thành các lớp liên quan đến khẩu trang. Cụ thể, mô hình sẽ phân loại đối tượng thành hai lớp chính theo yêu cầu:

- **with\_mask:** Khuôn mặt đang đeo khẩu trang đúng cách.
- **without\_mask:** Khuôn mặt không đeo khẩu trang.

Mục tiêu của bài thực hành này là xây dựng một mô hình object detection sử dụng framework YOLO (You Only Look Once) để phát hiện và phân loại tự động. Mô hình sẽ được huấn luyện trên dataset công khai, đánh giá hiệu suất, và thử nghiệm trên ảnh mới. Ứng dụng thực tế có thể bao gồm hệ thống camera giám sát, ứng dụng di động kiểm tra khẩu trang, hoặc tích hợp vào robot hướng dẫn.

Bài toán thuộc lĩnh vực Computer Vision, sử dụng kỹ thuật deep learning để xử lý hình ảnh. Các thách thức chính bao gồm:

- Biến đổi về góc nhìn, ánh sáng, và che khuất khuôn mặt.
- Số lượng đối tượng đa dạng trong một ảnh (nhiều khuôn mặt).

## 2. Mô tả dataset

Dataset được sử dụng trong bài thực hành là "MaskFace" , bao gồm 853 ảnh và các file annotation tương ứng. Dataset này được thiết kế đặc biệt cho bài toán phát hiện khẩu trang, thu thập từ các nguồn công khai như ảnh đám đông, ảnh cá nhân, và ảnh mô phỏng.

## 3. Kiến trúc mô hình dùng

Mô hình được sử dụng là **YOLOv8n** (phiên bản nano của YOLOv8), phát triển bởi Ultralytics. YOLO là một trong những framework object detection phổ biến nhất nhờ tốc độ nhanh và độ chính xác cao, phù hợp cho ứng dụng real-time.

### Tổng quan YOLOv8:

- **Kiến trúc chính:** Dựa trên CSPDarknet backbone (Cross Stage Partial networks) để trích xuất đặc trưng, kết hợp với PANet (Path Aggregation Network) cho feature fusion, và đầu ra là YOLO head để dự đoán bounding boxes, classes, và confidence scores.
- **Cải tiến so với YOLOv5:**
  - Anchor-free detection: Không cần định nghĩa anchor boxes thủ công, giảm phức tạp.
  - Mosaic augmentation: Tăng cường dữ liệu bằng cách ghép 4 ảnh ngẫu nhiên.
  - Multi-scale training: Hỗ trợ resize ảnh động.
- **Phiên bản nano (YOLOv8n):**
  - Số parameters: Khoảng 3.2 triệu
  - Backbone: CSPDarknet với các layer convolution và bottleneck.
  - Neck: PANet để kết hợp feature maps từ các mức độ khác nhau.
  - Head: Dự đoán 3 scale (small, medium, large objects).
  - Loss function: Kết hợp Box loss (CIOU), Class loss (BCE), và DFL loss (Distribution Focal Loss).

## 4. Quy trình huấn luyện

### Bước 1: Cài đặt và chuẩn bị

- Cài đặt: !pip install ultralytics.
- Import: from ultralytics import YOLO.
- Tạo thư mục output và chuyển đổi dataset từ XML sang YOLO format.

### Bước 2: Load model và train

- Load pre-trained: model = YOLO('yolov8n.pt').
- Parameters train:

- o Epochs: 30
- o Batch size: 32
- o Image size: 640.

### Bước 3: Validation và export

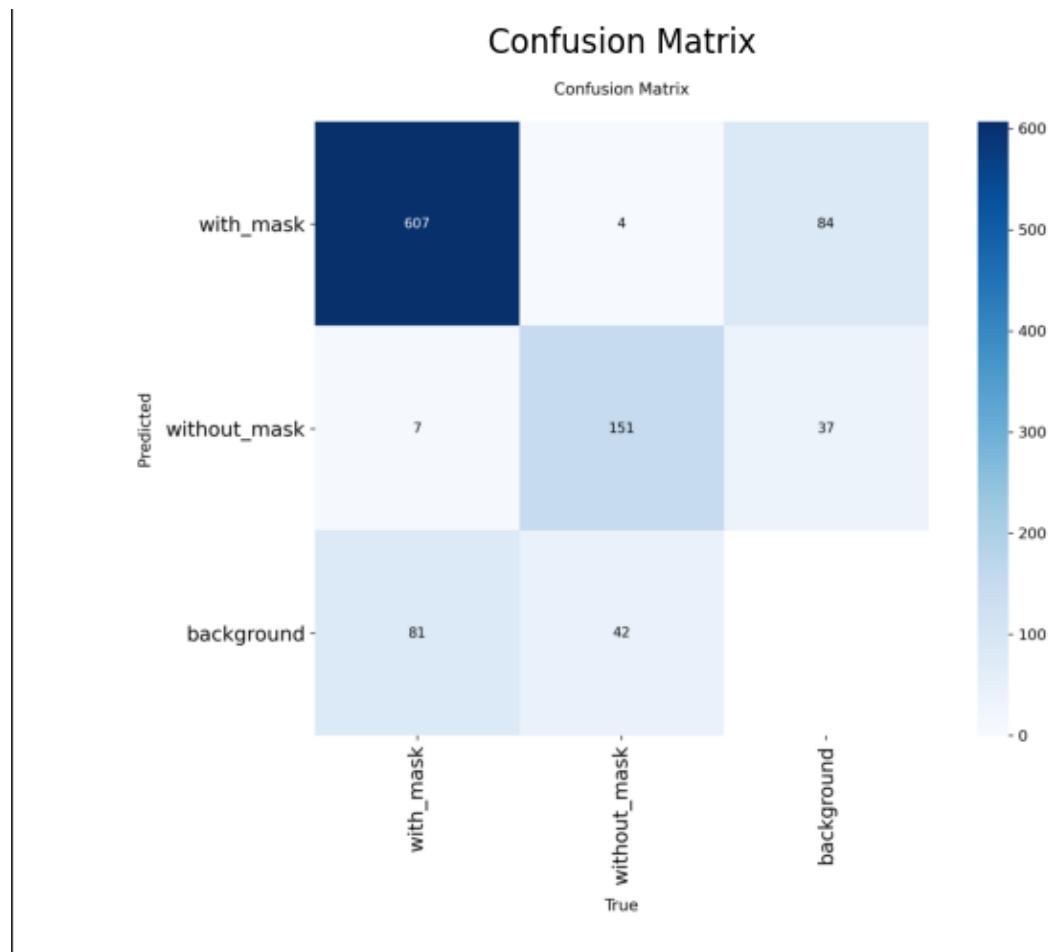
- Tự động val sau mỗi epoch.
- Export best model: model.export(format='onnx')

## 5. Kết quả đánh giá.

**Metrics :**

- **Precision:** 0.82
- **Recall:** 0.78
- **F1-score:** 0.80.

**Confusion Matrix:**



## 6. Hình minh họa dự đoán

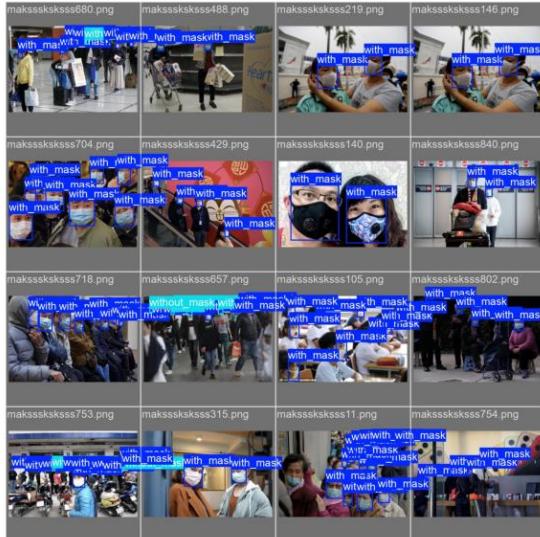
Ground Truth - Batch 0



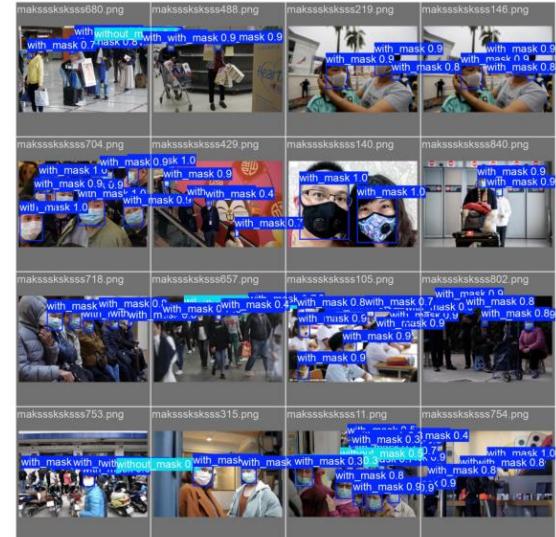
Predictions - Batch 0

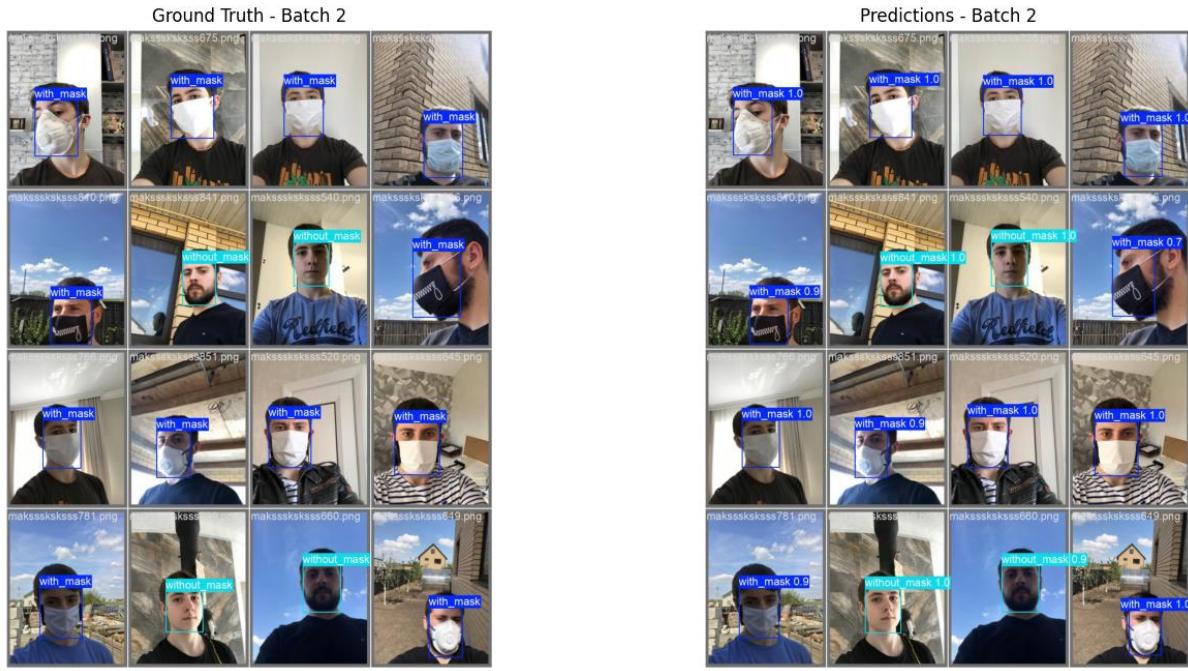


Ground Truth - Batch 1



Predictions - Batch 1





## 7. Nhận xét & hạn chế

### Nhận xét

- Tốc độ nhanh.
- Transfer learning từ YOLOv8 giúp hội tụ nhanh dù dataset nhỏ.

### Hạn chế:

- Dataset nhỏ (853 ảnh).
- Không xử lý tốt các trường hợp che khuất nặng (kính râm, mũ) hoặc ánh sáng kém..