

NHẬN DẠNG ĐEO KHẨU TRANG ĐÚNG CÁCH BẰNG RESNET, YOLO

Giảng viên: Lê Đình Duy

Giảng viên: Phạm Nguyễn Trường An

19522243-Nguyễn Văn Thành (nhóm trưởng)

19521642-Vũ Quốc Huy

19522244-Võ Huy Thành

Nội Dung

- Giới thiệu bài toán
- Giới thiệu dataset
- Giới thiệu 2 model
- Đánh giá kết quả
- Kết luận



GIỚI THIỆU BÀI TOÁN

BÀI TOÁN NHẬN DIỆN ĐEO KHẨU TRANG

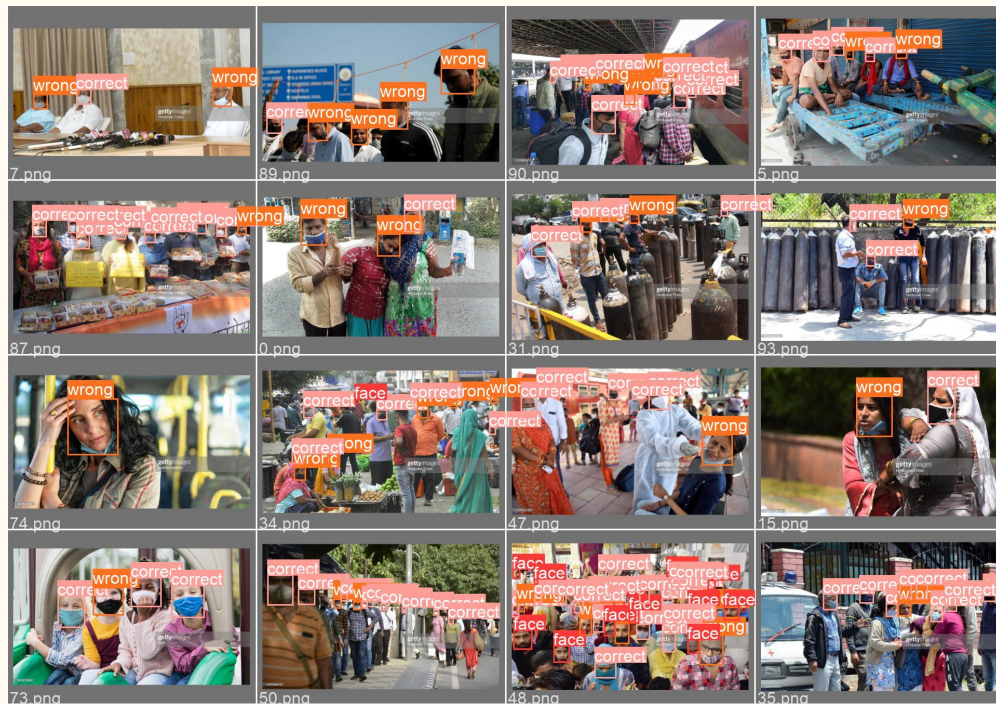
Input: Một ảnh, một video mà có chứa khuôn mặt người.

Output: Bounding Box phần khuôn mặt. Nhãn kèm theo.

Ứng dụng thực tế: Có thể kiểm soát việc người dân có đeo khẩu trang ở nơi công cộng, đeo khẩu trang có đúng cách không để đưa ra biện pháp tuyên truyền, hướng dẫn, hoặc xử phạt nếu cố tình.

Link video demo:

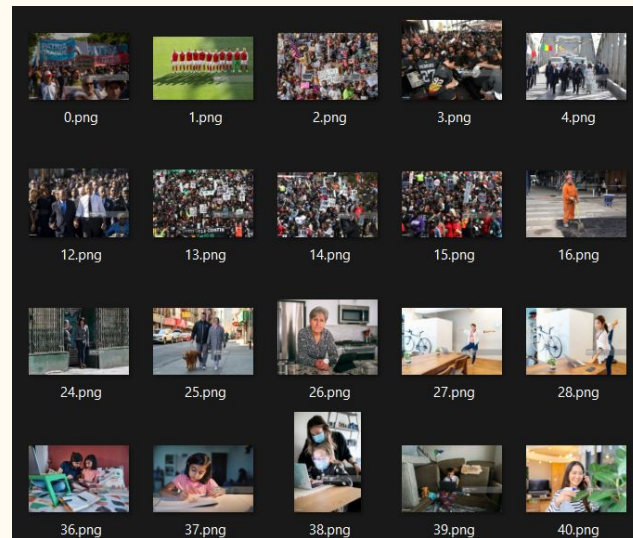
<https://drive.google.com/file/d/16SgtA6DX-Hul5GPZwI-3FRyWFLfdsyqW/view?usp=sharing>



GIỚI THIỆU DATASET

GIỚI THIỆU DATASET

1. Các ảnh được thu thập chủ yếu ở trang greetyimage.
2. Các ảnh được lưu trữ dưới dạng png. Có thể có nhiều đối tượng tập trung trong một ảnh, hoặc 1 đối tượng tập trung trên 1 ảnh.
3. Kích thước ảnh là 716x477.



GIỚI THIỆU DATASET

Nhãn cho dữ liệu:

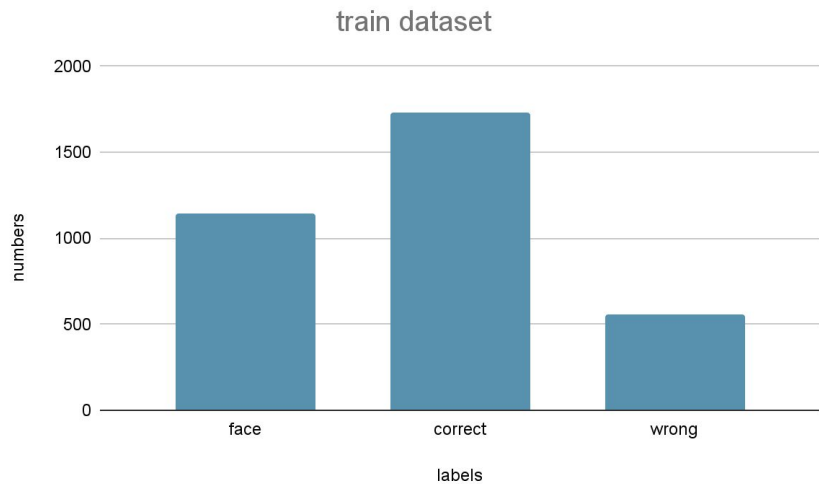
- **face:** trên khuôn mặt không có khẩu trang.
- **Correct:** trên khuôn mặt có khẩu trang và được đeo đúng cách.
- **wrong:** trên khuôn mặt có khẩu trang nhưng đeo sai cách.

Dùng labelImage để gán nhãn cho hình.



GIỚI THIỆU DATASET

1. Bộ dataset gồm có 815 hình.
2. Trong tập train có 705 hình. Gồm có 3431 nhãn, trong lớp face có 1144 nhãn, trong lớp correct có 1727 nhãn, trong lớp wrong có 560 nhãn.
3. Trong tập test có 108 hình. Gồm có 535 nhãn, trong lớp face có 47 nhãn, lớp correct có 298 nhãn, lớp wrong có 190 nhãn.



GIỚI THIỆU DATASET

```
2 0.886872 0.359833 0.050279 0.087866  
2 0.752095 0.379707 0.032123 0.064854  
1 0.552374 0.441423 0.037709 0.058577
```

Định dạng nhãn cho model yolo là một file .txt tương ứng với một hình. Ví dụ hình 0.png thì nhãn của nó sẽ là file 0.txt.

Số đầu tiên là class mà đối tượng trong bounding-box thuộc. Như trên hình là '2' là class: wrong.

4 số tiếp theo là vị trí của 4 góc của bounding-box.

```
<object>  
  <name>incorrect</name>  
  <pose>Unspecified</pose>  
  <truncated>0</truncated>  
  <difficult>0</difficult>  
  <bndbox>  
    <xmin>118</xmin>  
    <ymin>64</ymin>  
    <xmax>191</xmax>  
    <ymax>146</ymax>  
  </bndbox>  
</object>
```

Định dạng nhãn cho model resnet là một file .xml tương ứng với một hình.

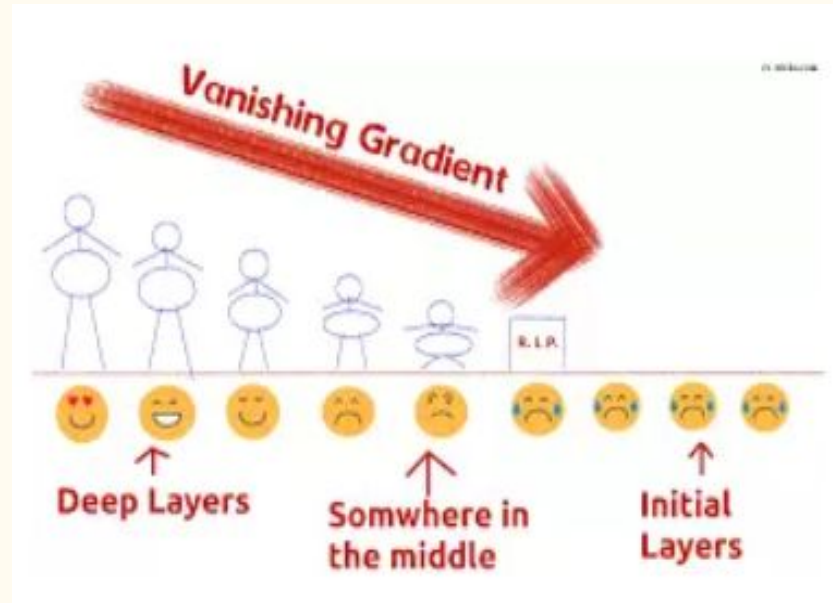
name: tên của class

bndbox: tạo độ của bounding-box.

GIỚI THIỆU RESNET

GIỚI THIỆU RESNET

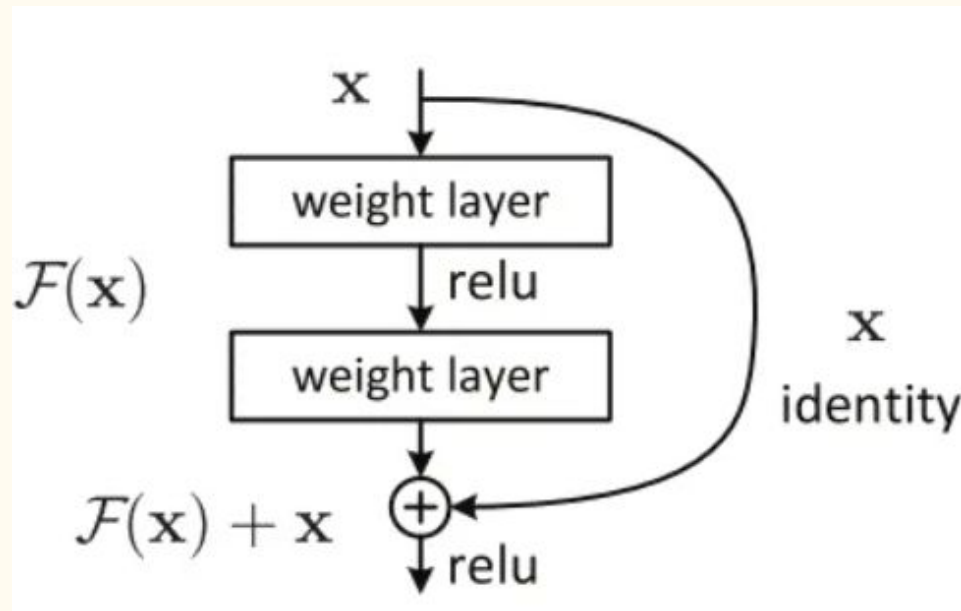
- Resnet (viết tắt của residual network) là một mạng học sâu (CNN), được thiết kế để làm việc với hàng trăm, nghìn lớp chập.
- Khi mà xây dựng mạng CNN với nhiều lớp chập sẽ xảy ra hiện tượng Vanishing Gradient.
- Vanishing Gradient: Nếu mạng CNN có quá nhiều lớp, khi thực hiện Backpropagation thì khi càng về lớp thấp hơn thì gradient nó sẽ nhỏ hơn. Vì vậy khi thực hiện gradient descent để cập nhật các trọng số dẫn tới chúng không hội tụ và làm cho mạng trở nên kém đi



GIỚI THIỆU RESNET

Để cải thiện vấn đề trên. Resnet xuất hiện với ý tưởng là sử dụng kết nối tắt đồng nhất để xuyên qua một hay nhiều lớp.

Nhờ khả năng biểu diễn mạnh mẽ của Resnet, nên nó được ứng dụng nhiều trong thị giác máy tính.



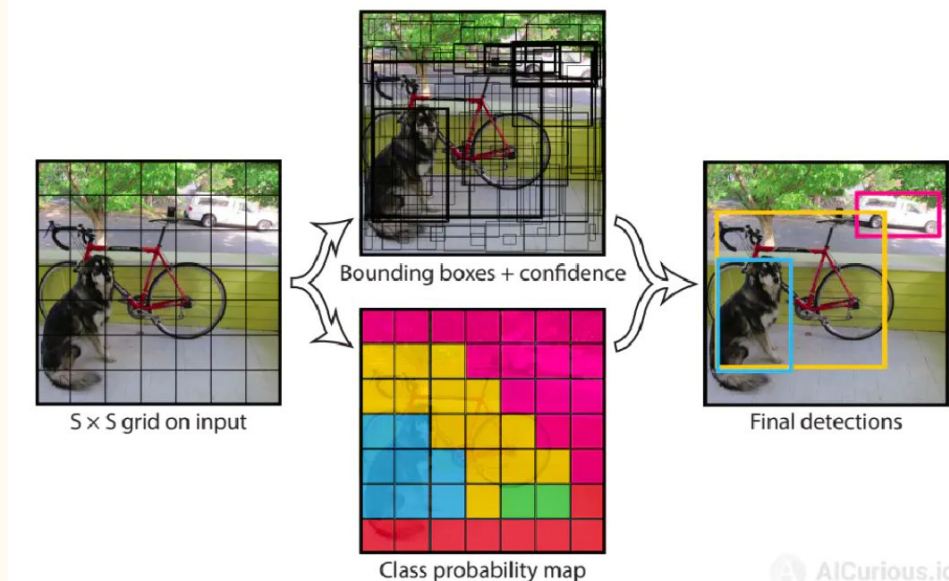
GIỚI THIỆU YOLO

GIỚI THIỆU YOLO

Yolo (you only look once) được ra mắt bởi Joseph Redmon vào tháng 5 năm 2016, nó là một mô hình CNN được dùng để phát hiện đối tượng.

Ý tưởng của Yolo:

- Chia ảnh thành một lưới các ô.
- Với mỗi ô, mô hình sẽ đưa ra dự đoán cho B bounding box.
- Ứng với mỗi box trong B bounding box này sẽ là 5 tham số x, y, w, h , confidence với (x, y) là tọa độ tâm, w, h là chiều cao chiều rộng và confidence là độ tự tin của dự đoán.
- Với mỗi ô thì mô hình cũng dự đoán xác suất rơi vào mỗi class.



ĐÁNH GIÁ KẾT QUẢ

ĐÁNH GIÁ KẾT QUẢ

Để so sánh hai model với nhau. Chúng ta sẽ kiểm tra từng model sau khi được huấn luyện trên cùng một tập train.

Chúng ta sẽ so sánh dựa trên các giá trị $mAP@0.5$, và $mAP@0.5:0.95$ để xem xét model nào tốt hơn.

- Nếu mAP lớn, thì có nghĩa là các Precision và Recall có giá trị đều khá cao. Từ đó suy ra model tốt.
- Nếu mAP nhỏ, thì có nghĩa là các Precision và Recall có giá trị khá thấp. Từ đó suy ra model không tốt.

AP (mAP) là gì?

ĐÁNH GIÁ KẾT QUẢ


IoU là gì?

\Rightarrow IoU là tỉ lệ giữa phần giao và phần hợp của của vùng dự đoán và vùng đối tượng thật.

Kết quả của IoU nằm trong khoảng $(0,1)$. Để xác định dữ liệu đó là phát hiện sai hay đúng, thì chúng ta dựa vào một ngưỡng (threshold) cho trước. Nếu IoU lớn hơn hoặc bằng threshold thì là phát hiện đúng, còn nhỏ hơn là phát hiện sai.

True Positive (TP): IoU lớn hơn hoặc bằng threshold

False Positive (FP): IoU nhỏ hơn threshold


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

ĐÁNH GIÁ KẾT QUẢ

- Precision: Đánh giá độ tin cậy của kết luận đưa ra (bao nhiêu % kết luận của model đưa ra là chính xác)
- Recall: Đánh giá khả năng tìm kiếm toàn bộ các ground truth của mô hình (mô hình có khả năng tìm kiếm bao nhiêu % positive samples).

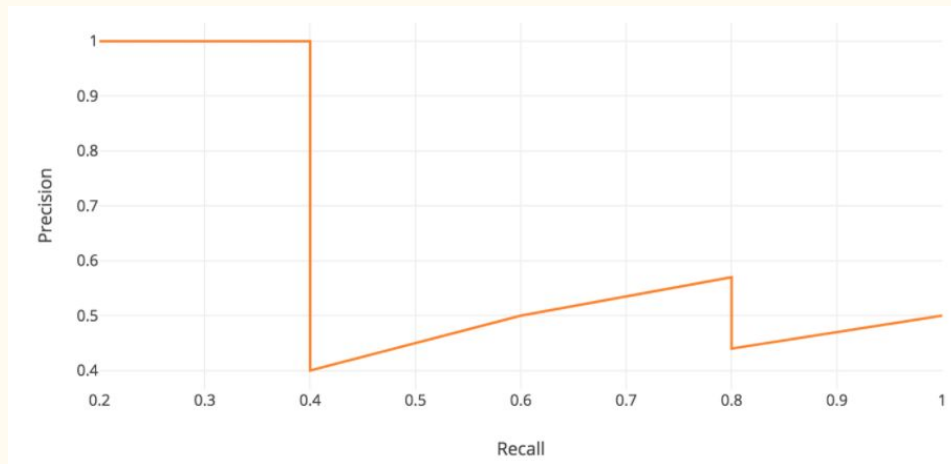
Rank	Correct?	Precision	Recall
1	True	1.0	0.2
2	True	1.0	0.4
3	False	0.67	0.4
4	False	0.5	0.4
5	False	0.4	0.4
6	True	0.5	0.6
7	True	0.57	0.8
8	False	0.5	0.8
9	False	0.44	0.8
10	True	0.5	1.0

ĐÁNH GIÁ KẾT QUẢ

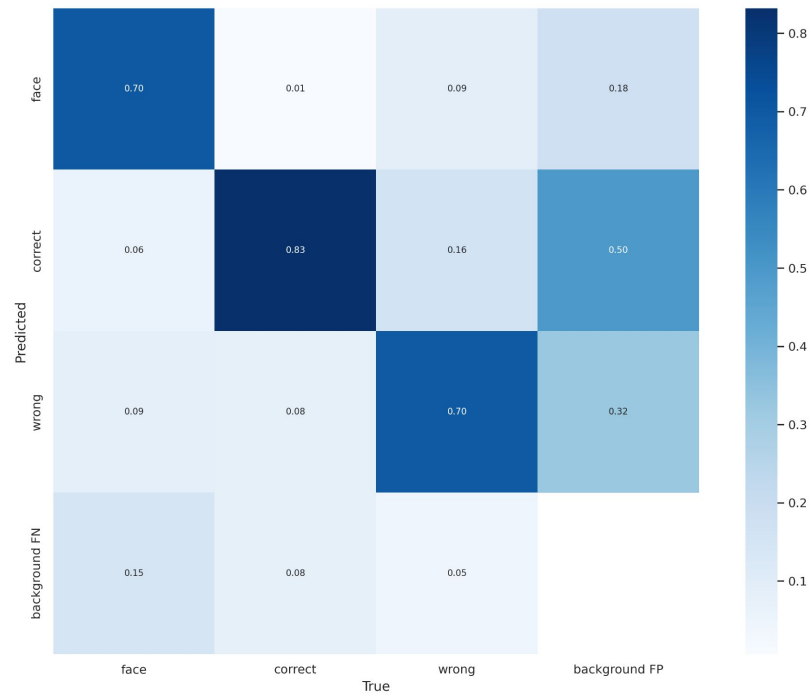
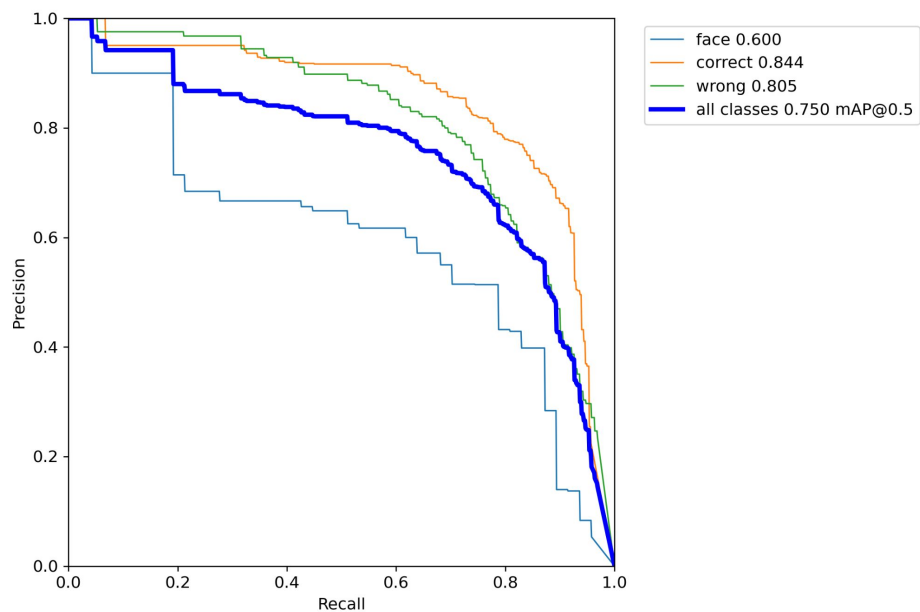
AP là giá trị phía dưới đường biểu diễn mối quan hệ precision-recall.

mAP là trung bình AP cho tất cả các lớp.

Trong một vài trường hợp thì AP chính là mAP.



KẾT QUẢ TRÊN YOLO



SO SÁNH RESNET VỚI YOLO

	RESNET	YOLOv5
mAP@0.5	0.802	0.750
mAP@0.5:0.95	0.550	0.358

Kết quả:

- Ta thấy về cả hai giá trị mAP@0.5 và mAP@0.5:0.95 thì mạng Resnet đều lớn hơn Yolo. Vậy về độ chính xác thì mô hình Resnet tốt hơn Yolo trên cùng bộ dataset được huấn luyện.
- Về thời gian huấn luyện thì thời gian huấn luyện của Yolo nhanh hơn so với thời gian huấn luyện của Resnet.