

11. CẢI TIẾN MÔ HÌNH PHẦN 2

Trong phiên bản 2 này, ta sẽ có các bước làm nhằm cải thiện hiệu suất của mô hình xác định một người có đeo khẩu trang hay không. Các bước chính của phiên bản này bao gồm:

1. Thêm data từ dataset của Kaggle

- Bộ dataset ban đầu chứa: 52 000 ảnh khuôn mặt với tỉ lệ 2 tập là 50:50
- Để có thể cải thiện hiệu suất mô hình ta sẽ thêm các hình ảnh khuôn mặt được thu thập từ kaggle:
 - Dataset in Kaggle (Face Mask Detection): <https://www.kaggle.com/andrewmvd/face-mask-detection>
 - Dataset chứa 853 images và corresponding annotation files được chia thành 3 label:
 - Mask correctly
 - Incorrectly
 - Not wearing mask



Sample images from the face mask dataset (image by author)

(ảnh minh họa)

- Ta cần xử lý để đưa bài toán về dạng binaryClassification
- Thay đổi format file **.xml annotation** thành **YOLO darknet** format
 - YOLO darknet format: `<object-class> <x> <y> <width> <height>`
 - Mỗi dòng sẽ đại diện cho annotation từng đối tượng trong image.
 - `<x> <y>` là tọa độ tâm của bounding box
 - `<width> <height>` tương ứng width and height

< maksssksksss0.xml (1.25 kB)

```
<annotation>
  <folder>images</folder>
  <filename>maksssksksss0.png</filename>
  <size>
    <width>512</width>
    <height>366</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
```

(.xml annotation)

```
1 0.427234 0.123172 0.191749 0.171239
0 0.183523 0.431238 0.241231 0.174121
1 0.542341 0.321253 0.191289 0.219217
```





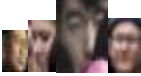
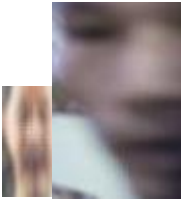

(YOLO darknet format)

- Cách thay đổi: Dùng ứng dụng **Roboflow** (<https://roboflow.com/>)
 - Upload the images và annotations
 - Chọn tỉ lệ tập train và validation (test nếu cần)
 - Thêm các augmentation như: blur, brightness, rotation
 - Generate the new images và thu được YOLO Darknet format
- Thu được tập ảnh với YOLO Darknet format. Tuy nhiên, các dataset được gán nhãn theo 3 label:

```
mask_weared_incorrect # label 0
with_mask # label 1
without_mask # label 2
```

- Ta cần chuyển về dạng bài Binary Classification với 2 label:

```
without_mask # label 0
with_mask # label 1
```
- Sau đi đã gom lại còn 2 label, Ta sẽ rút trích các khuôn mặt ra thành từng ảnh và gán label cho chúng. Kết quả sau cùng của quá trình là:
 - with_mask: 5229 ảnh
 - without_mask: 1353 ảnh

With_mask	Without_mask
 : ảnh kích thước nhỏ, bị mất mát nhiều chi tiết, góc nghiêng.  : ảnh chính diện, rõ ràng	 : under nose  : dưới cằm  : ảnh kích thước nhỏ, bị mất mát dữ liệu  : ảnh mờ  : có phụ kiện (kính), bị che khuất

Sau khi đã tách và gán nhãn có các hình ảnh khuôn mặt, ta sẽ thêm chúng vào chung với dataset ban đầu để train lại mô hình:

Kết quả:

- with_mask: khoảng 31k ảnh
- without_mask: khoảng 27k ảnh

2. Áp dụng các phương pháp Image augmentation

- Image augmentation là các phương pháp nhằm đa dạng tập dữ liệu huấn luyện bằng cách áp dụng các phương pháp xử lý ảnh như:

- Lật ảnh
- Xoay ảnh
- Cắt ảnh
- Đổi màu ảnh
-

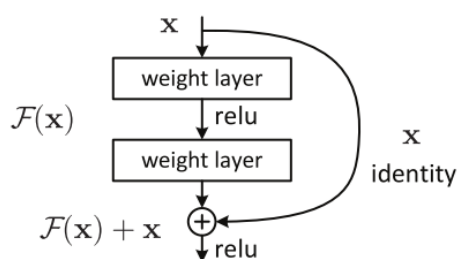
Ví dụ:



Bằng việc sử dụng các phép xử lý ảnh, ta sẽ làm đa dạng bộ dữ liệu huấn luyện. Qua đó, giúp cho mô hình sẽ học tập được các trường hợp khác nhau và làm tăng độ chính xác khi dự đoán dữ liệu thực tế.

3. Tìm hiểu về kiến trúc mô hình Resnet-50

- ResNet (Residual Network) được giới thiệu năm 2015. Hiện tại thì có rất nhiều biến thể của kiến trúc ResNet với số lớp khác nhau như ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152,...
- Mạng ResNet là một mạng CNN được thiết kế để làm việc với hàng trăm hoặc hàng nghìn lớp chập. Một vấn đề xảy ra khi xây dựng mạng CNN với nhiều lớp chập sẽ xảy ra hiện tượng Vanishing Gradient (xảy ra ở Backpropagation – Lan truyền ngược) dẫn tới quá trình học tập không tốt. Mạng ResNet ra đời giải quyết vấn đề đó.

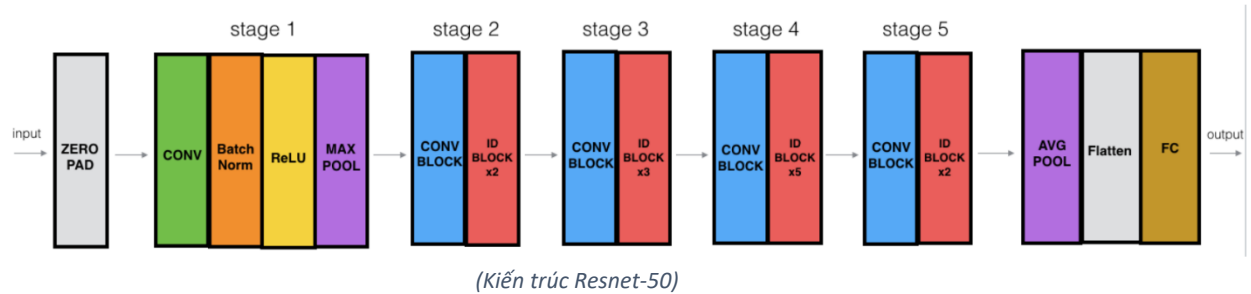


Resnet sẽ đưa ra các “kết nối tắt” để giúp xuyên qua 1 hay nhiều lớp. Các khối có chức năng như vậy được gọi là Residual Block.

Mũi tên trong ảnh xuất phát từ đầu và kết thúc tại cuối một khối dư. Nó sẽ bổ sung Input X vào đầu ra của layer. Tác dụng của việc bổ sung này sẽ giúp chống lại việc đạo hàm bằng 0.

Để giá trị dự đoán có được giá trị gần với giá trị thật nhất bằng cách: $F(x) + X \rightarrow \text{ReLU}$. Trong đó, $X \rightarrow \text{weight1} \rightarrow \text{ReLU} \rightarrow \text{weight2}$.

Kiến trúc mạng Resnet-50:

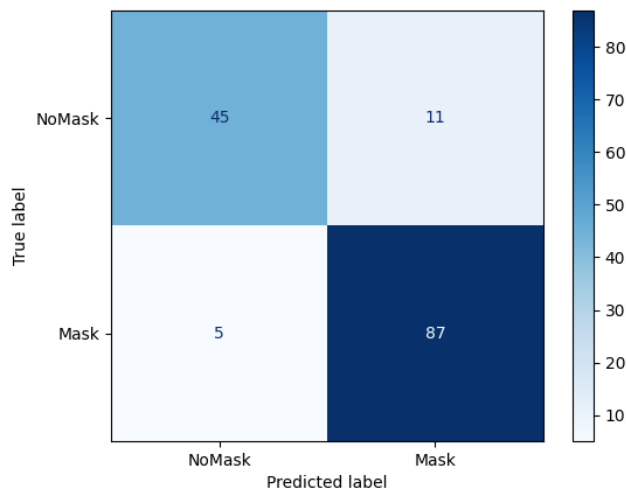


"ID BLOCK" trong hình trên là viết tắt của từ Identity block và ID BLOCK x3 nghĩa là có 3 khối Identity block chồng lên nhau. Nội dung hình trên như sau :

- Zero-padding : Input với (3,3)
- Stage 1 : Tích chập (Conv1) với 64 filters với shape(7,7), sử dụng stride (2,2). BatchNorm, MaxPooling (3,3).
- Stage 2 : Convolutional block sử dụng 3 filter với size 64x64x256, f=3, s=1. Có 2 Identity blocks với filter size 64x64x256, f=3.
- Stage 3 : Convolutional sử dụng 3 filter size 128x128x512, f=3,s=2. Có 3 Identity blocks với filter size 128x128x512, f=3.
- Stage 4 : Convolutional sử dụng 3 filter size 256x256x1024, f=3,s=2. Có 5 Identity blocks với filter size 256x256x1024, f=3.
- Stage 5 :Convolutional sử dụng 3 filter size 512x512x2048, f=3,s=2. Có 2 Identity blocks với filter size 512x512x2048, f=3.
- The 2D Average Pooling : sử dụng với kích thước (2,2).
- The Flatten.
- Fully Connected (Dense) : sử dụng softmax activation

4. Kết quả

Sau khi đã áp dụng các phương pháp augmentation, mô hình Resnet-50 và các thuật toán earlyStopping. Ta sẽ huấn luyện lại mô hình và đánh giá kết quả sau cùng của quá trình:






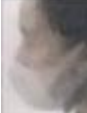

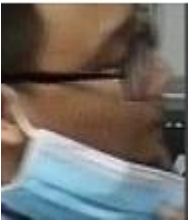

Các thông số thu được từ kết quả:

- $accuracy = \frac{45+87}{45+87+11+5} = 0.8919$
- Tỷ lệ đeo khẩu trang đoán đúng

$$= \frac{87}{87 + 11} = 0.8878$$
- Tỷ lệ không đeo khẩu trang đoán đúng

$$= \frac{45}{45 + 5} = 0.9$$

Các trường hợp bị nhận diện sai

Mask (Nhận diện thành NoMask)	NoMask (Nhận diện thành Mask)
 : ảnh bị che phần mắt  ảnh nghiêng  ảnh nhỏ, góc nghiêng, mất mát dữ liệu  ảnh góc nghiêng, mờ, khẩu trang trùng màu với background  ảnh góc cuối mặt, có phụ kiện (mũ)	 ảnh góc nghiêng, có phụ kiện (mắt kính)  ảnh cuối mặt  ảnh cuối mặt

Nhận xét: Mô hình cho kết quả tốt hơn so với MobileNetV2 (kết quả trên tập test 89% > 50%). Tuy nhiên thời gian huấn luyện còn khá chậm.

Đề xuất giải pháp:

- Sử dụng mô hình Resnet-34 để giảm thời gian huấn luyện mô hình, chấp nhận độ chính xác bị giảm đi để mô hình được huấn luyện nhanh hơn
- Tăng dữ liệu cho tập test để đánh giá chính xác hơn.