Báo cáo đồ án

I. Tính cấp thiết của đề tài:

Thế giới vừa mới trải qua những tổn thất nghiêm trọng do chủng virus corona mang đến. Trong giai đoạn dịch bệnh căng thẳng, việc các cá nhân tự bảo vệ bản thân bằng các biện pháp phòng tránh như là một điều kiện bắt buộc và quan trọng nhất trong việc giảm thiểu các hậu quả của virus mang lại cho bản thân cũng như mọi người xung quanh. Với việc lây nhiễm chính bằng đường không khí, việc trang bị cho bản thân những chiếc khẩu trang như một lớp giáp bảo vệ bản thân với sự xuất hiện của virus trong không khí chưa bao giờ là thừa lúc này. Việc tiếp xúc giữa các cá thể với nhau trong những khu vực có mật độ cá thể lớn là điều kiện tuyệt vời đề cho dịch bệnh được lây lan, khi mà một số chủng virus đã được chứng minh là có thể lây lan trong không khí ở một khoảng cách rất xa. Chính vì vậy mà mỗi cá nhân trong thời đại dịch bệnh luôn được khuyến khích, thậm chí là bắt buộc phải trang bị cho cá nhân một chiếc khẩu trang khi tiếp xúc với mọi người trong cộng đồng, đó như là biện pháp để bảo vệ cho an toàn của bản thân cũng như an toàn của mọi người. Tuy vậy, có nhiều cá nhân vẫn chưa nhận thức được sự quan trọng của khẩu trang trong việc phòng chống lây nhiễm virus, và tiếp xúc với mọi người trong cộng đồng với tình trạng không đeo khẩu trang. Việc phát hiện và cảnh báo các cá nhân đó ở những nơi công cộng có thể không xảy ra kịp thời được, từ đó có thể tạo điều kiện cho virus được lây lan nhiều hơn nếu như những cá thể không mang khẩu trang đó dương tính.

Máy tính với sức mạnh tính toán vượt trội có thể hỗ trợ con người trong việc phát hiện và cảnh báo các cá thể trong cộng đồng trong việc mang khẩu trang một cách kịp thời. Nhận thấy được tính thiết thực của vấn đề này và sự hiệu quả của mô hình máy tính vào việc phát hiện và cảnh bảo đeo khẩu trang, nhóm quyết định chọn đề tài “Ứng dụng deep learning vào phát hiện cá thể mang khẩu trang”.

II. Nội dung của đề tài:

1. Phát hiện vật thể

Object detection, hay còn gọi là phát hiện vật thể, là một lĩnh vực trong thị giác máy tính và trí tuệ nhân tạo. Thị giác máy tính là một ngành khoa học đã và đang phát triển rất mạnh, đặc biệt là trong những như sau này, khi có sự ra đời của các cấu trúc mạng neural. Trong nội dung của đồ án, nhóm tìm hiểu một nhánh của thị giác máy tính, object detection với hướng tiếp cận theo mạng neural. Hơn 90% thông tin mà con người xử lý hằng ngày được tiếp nhận thông qua kênh thị giác, bởi vì đặc thù của thị giác nên tốc độ xử lý của con người khi tiếp nhận những thông tin này cũng nhanh hơn nhiều so với thông tin được tiếp nhận từ những kênh khác. Bằng việc áp dựng cơ chế tương tự như vậy vào máy tính, các nhà khoa học đã phát minh ra hướng tiếp cận theo neural, đây là một hướng tiếp cận khác, mạnh mẽ hơn nhiều, và đánh bại các hướng tiếp cận cũ trong việc xử lý bài toán với độ phức tạp cao với thời gian xử lý nhanh. Thông tin được nhận vào từ thị giác được xử lý qua hệ các dây thần kinh neural được nối với nhau một cách phức tạp, từ đó trích xuất ra được các thông tin trừu tượng được xử dụng để xử lý và tính toán ra các thông tin mới theo mục đích của bài toán.

2. Áp dụng vào bài toán nhận diện khẩu trang:

Con người nhìn vào một cá thể liền có thể lập tức nhận ra được là cá thể đó có đang mang khẩu trang hay không mà không cần phải mất thời gian nhiều cho việc tính toán và tiếp nhận. Và nếu như một mô hình học máy, được xây dựng với một cấu trúc hợp lý và được cung cấp cho một nguồn kiến thức đủ phong phú để máy có thể học được cách nhận biết và phân biệt các cá thể có mang khẩu trang hay không, thì cộng với tốc độ xử lý cực nhanh của máy tính, máy tính hoàn toàn có thể hỗ trợ con người trong việc hỗ trợ phát hiện nhận dạng khẩu trang trong tầm nhìn của máy tính.

Một bức ảnh được tiếp nhận vào mắt người, thì ta có thể có biết được khẩu trang nằm ở đâu trên bức ảnh, và khẩu trang có được người trước mặt đeo không thông qua những thông tin trừu tượng hoặc cụ thể mà não ta trích xuất được, ví dụ như: khẩu trang có che được miệng, hay khẩu trang có che được mũi hay không. Khẩu trang có được đeo vào vị trí hợp lý trên khuông mặt hay không,… Những thông tin đó hoàn toàn có thể được biểu diễn trong máy tính dưới dạng các đặc trưng, và máy tính có thể học được và trích xuất được thông tin từ việc sử dụng những đặc trưng đó trong việc tính toán.

Trong hướng tiếp cận theo mạng neural, bức ảnh được biểu diễn dưới dạng một ma trận với các giá trị tương ứng cho từng vị trí trên bức ảnh, giá trị của mỗi vị trí trong ma trận đặc trưng cho mức độ mạnh yếu về sắc độ trong bức ảnh. Việc biểu diễn thông tin như thế này làm cho bức ảnh trở thành một ma trận mang thông tin với giá trị có thể tính toán được.

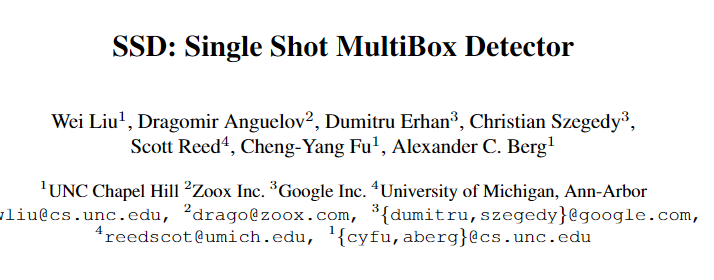
3. Hướng tiếp cận của đồ án:

Trong khi liệt kê ra các yêu cầu chức năng phi chức năng của mô hình hiện thực, nhóm nhận thấy mô hình như là một phần hỗ trợ con người trong việc giám sát và phát hiện các cá thể đeo khẩu trang trong cộng đồng, cho nên tốc độ xử lý là yếu tố được ưu tiên, nhưng vẫn phải đảm bảo được độ chính xác tương đối. Có một số hướng tiếp cận theo mô hình neural mà nhóm tìm hiểu được bao gồm:

* Mô hình multi feature map detection, ví dụ như SSD. Với việc phát hiện và nhận dạng vật thể thông qua một lần forward, nên mô hình SSD mang lại tốc độ xử lý khá cao, nếu được thiết kế hợp lý, có thể áp dụng việc nhận diện theo thời gian thực, bên cạnh đó độ chính xác cũng khá cao, có thể nhận dạng được chính xác các trường hợp dễ.
* Mô hình Region based convolutional neural network, R-CNN. Việc nhận dạng và tính toán có sự hỗ trợ của giải thuật trong việc xác định vùng có tiềm năng chứa vật thể trong bức ảnh. Độ chính xác của mô hình khá cao, và dự đoán khá chính xác vị trí của vật thể trong khung ảnh. Tuy vậy, tốc độ của mô hình trong việc tính toán khá lâu, việc áp dụng vào thực tế ở một bài toán yêu cầu xử lý theo thời gian thực như việc nhận dạng đeo khẩu trang là khá khó khăn và bất khả thi, vì độ trễ có thể làm cho mô hình mất đi nhiều thông tin có thể nhận được trong thời gian tính toán, cộng với khối lượng tính toán lớn như vậy là cản trở trong việc áp dụng ở trong các thiết bị nhúng hay thiết bị di động với cấu hình thấp.
* Mô hình YOLO, thay đổi trong việc chia bức ảnh ra các vùng đặc trưng với các bounding box khác nhau, mô hình bao gồm phần cnn để trích xuất thông tin từ bức ảnh mà chỉ gồm các lớp fully connected, nên tốc độ xử lý của mô hình YOLO rất nhanh, có thể đáp ứng được các bài toàn yêu cầu xử lý theo thời gian thực.

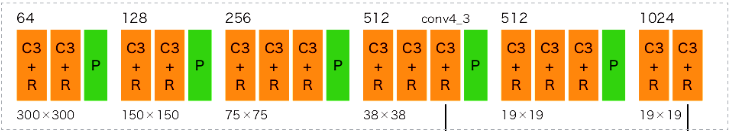
Trong nội dung hiện thực của đồ án, nhóm quyết định hiện thực và sử dụng mô hình theo cấu trúc SSD, cụ thể là mô hình SSD300 để áp dụng vào bài toán nhận dạng khẩu trang.

4. Thiết kế mô hình  
Mô hình SSD300 mà nhóm hiện thực dựa trên thiết kế theo mô hình được publish ở paper của nhóm nghiên cứu:



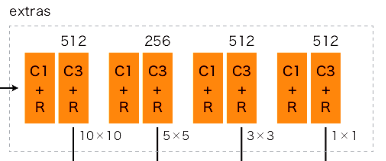
Mô hình có thể chia ra thành những module chính là: module học thông tin từ bức ảnh, module trích xuất đặc trưng từ bức ảnh, module dự đoán vị trí của vật thể trong bức ảnh, module dự đoán loại vật thể ở trong vị trí tương ứng.

Phần module học thông tin từ bức ảnh, nhóm sử dụng kiến trúc vgg16:

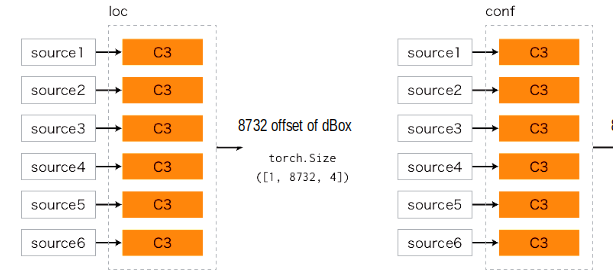


Gồm các lớp tích chập và sẽ giảm kích thước của bản đồ đặc trưng sau mỗi khối, nhằm tăng số kênh thông tin nhận vào ở khối sau, có nghĩa là thông tin nhận được ít cụ thể hơn nhưng ở phạm vi tổng quát hơn. Càng về cuối mô hình thì thông tin học được càng trừu tượng và tổng quát hơn so với những lớp đầu.

Phần module trích xuất đặc trưng từ bức ảnh, nhóm sử dụng kiến trúc như sau:



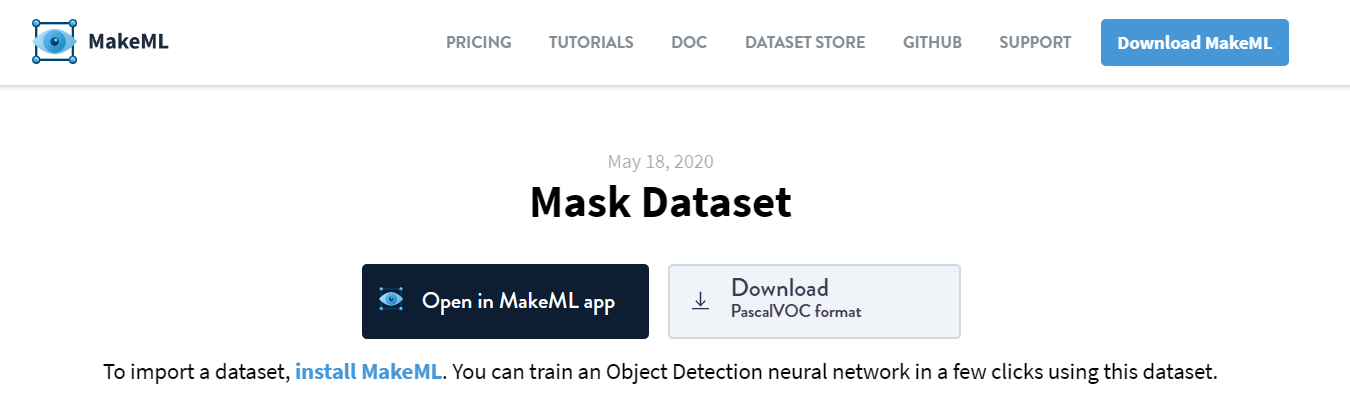
Sau hai lớp trên sẽ là hai module dự đoán vị trí và lớp của đối tượng phát hiện được.



Phần cấu hình chi tiết của mỗi layer nhóm hiện thực giống như paper thiết kế.

5. Tập dữ liệu huấn luyện:

Nguồn dữ liệu:



Dữ liệu gồm 853 bức ảnh được đánh định dạng theo tập dữ liệu VOC. Annotation là một file xml, với các key và field cần thiết như là kích thước của bức ảnh và số kênh màu trong bức ảnh. Các vật thể trong bức ảnh với thông tin vị trí của đối tượng đó, lớp mà vật thể đó thuộc về.

Mô hình của nhóm sẽ trích xuất thông tin từ các file annotation này về vị trí toạ độ của các đối tượng trong bức ảnh của như lớp mà vật thể thuộc về. Vị trí của vật thể được biểu diễn dưới dạng một bounding box với các thông tin về toạ độ của hai điểm ở góc trên bên trái và góc dưới bên phải.

Các vật thể trong tập dữ liệu được đánh nhãn thuộc vào 3 lớp chính là “with\_mask”, “without\_mask”, “mask\_weared\_incorrectly”, tương ứng với nhãn đeo khẩu trang, không đeo khẩu trang, hay đeo khẩu trang sai cách. Tổng số vật thể để học từ trong tập dữ liệu tầm 3-4000 vật thể, tuy vậy dataset gây ra khó khăn cho nhóm khi dữ liệu bị imbalance, cụ thể là số label được đánh nhãn “with\_mask” lớn hơn số label được đánh nhãn “without\_mask” vài lần, và lớn label “mask\_weared\_incorrectly”.

Chính vì sự mất cân đối về số lượng của các lớp trong tập dữ liệu huấn luyện, làm cho mô hình mà nhóm hiện thực có xu hướng đoán chính xác các vật thể thuộc các lớp đa số như là lớp “with\_mask” và “without\_mask” mà bỏ qua các vật thể thuộc lớp còn lại. Điều này xảy ra trong quá trình tối ưu, nhằm nâng cao độ chính xác lúc huấn luyện mà mô hình có xu hướng dự đoán vật thể thuộc vào hai lớp đa số đó mà bỏ qua lớp thiểu số còn lại.

**Chèn ảnh số lượng dữ liệu của từng lớp.**

6. Hiện thực:

III. Tổng kết:

a. Đánh giá:

b. Hạn chế, hướng khắc phục