ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
UIT-HCM

REAL-TIME FACE MASK DETECTION

[1,2] **Nguyen Nhat Truong** 20522087@gm.uit.edu.vn
[1,2] **Lai Chi Thien** 20520309@gm.uit.edu.vn
[1,2] **Le Thi Phuong Vy** 20520355@gm.uit.edu.vn
[1,2] **Le Truong Ngoc Hai** 20520481@gm.uit.edu.vn
[1,2] **Tran Thi My Quyen** 20520731@gm.uit.edu.vn

[1] **University of Information Technology**
[2] Viet Nam National University Ho Chi Minh City - VNUHCM

I. INTRODUCTION:

In recent years, the raging Covid-19 pandemic has significantly affected the social and financial situation around the world in general and Vietnam in particular. To ensure citizens health as well as manage citizens compliance with regulations in epidemic prevention, it is necessary to build an automatic detector to determine how people are wearing their masks, i.e., to check whether they wear masks correctly or not. Thanks to that, we can easily communicate reminders and punishments to people who do not wear masks or wear masks incorrectly when going out

II. PROBLEM DESCRIPTION:

Goals

- Detects the objects and their masking state in a photo or video.
- The detection of masks helps to detect abnormalities and manage Covid-19 epidemic prevention activities in the community more effectively.

Overview

- Real-time face mask detection problem is a problem belonging to the group of OD problems, a subfield in CV community and has many practical applications.
- Based on previous related scientific works, our project is aimed at detecting masks in videos or photos taken from a camera with a direct view.
- The classes considered in the study are “with_mask”, “without_mask” and “mask_worn_incorrect”, including:
  + with mask: People wearing masks cover up their nose and mouth.
  + without mask: People who don’t wear masks.
  + mask worn incorrectly: People wearing masks that don’t fully cover up their nose or mouth, or both.

a. Input **b. Output**

Fig 1. Input and output illustration of face mask detection

III. APPLYING COMPUTATIONAL THINKING:

Decomposition

The main problem was initially posed to detect masks on videos and images in real-time from webcams or cameras. Since an available video or real-time video from the camera is a sequence of images that are played in succession, the action on the video can be broken down into several tasks to be performed on each image.

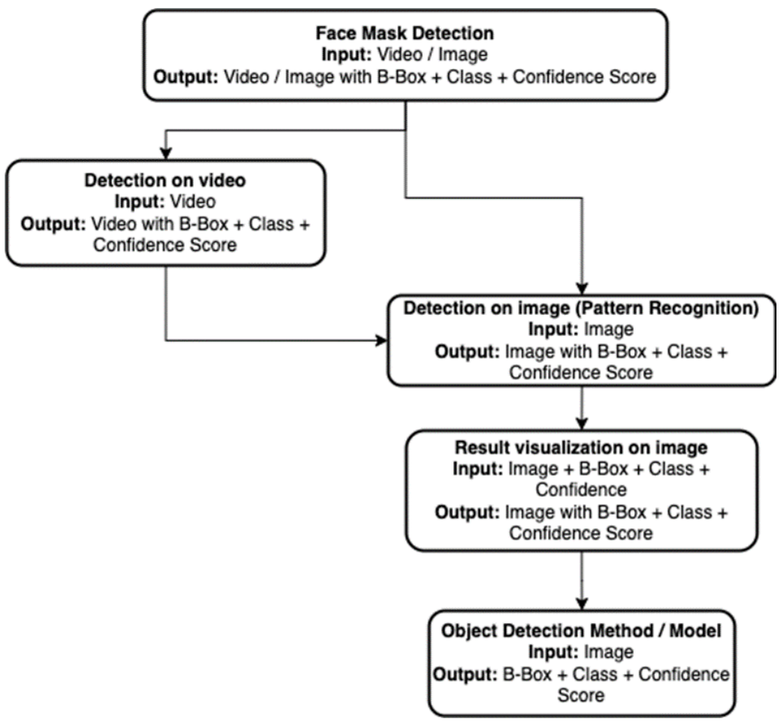

Fig 2: Decomposition Tree

Pattern Recognition

Real-time detection problem: we find that there is a similarity between the detection process for different images, the images will all be processed in the same way to give the same final format of results. Therefore, Pattern Recognition will be applied to process each image in the input video.

Face mask detection problem:

- *Input*: An image taken from a webcam or camera with or without a human face in PNG format.
- *Output*: An image consists of a minimal bounding box containing the possible location of a person's face in the image and a label indicating the mask-wearing state to which the object belongs.

Object detection problem:

- *Input*: 1 image may or may not have an object of interest.
- *Output*: An image with a minimal bounding box containing the location of the object of interest and its label.

Comparing the characteristics of the **mask detection** problem and the **object detection** problem, we can conclude that: The input of the two problems is similar and the output of the mask detection problem is a more specific case of the object detection problem. Therefore, we can apply the algorithms of the Object Detection problem to the Face Mask Detection problem by changing the object we are interested in (class) into a set consisting of with_mask, without_mask, mask_weared_incorrect.

Abstraction / Data Representation

The input of the main problem is Video. A video is made up of two main components, images and sounds. However, since this is a problem that focuses entirely on image processing and feature-based detection, the audio part may not be considered as the input of the problem.

Besides, some constraints for input are also set as follows:

- Specific limits on resolution, ensuring maximum efficiency: the minimum size of the input image is 128x128.
- Requirements of a bounding box for each object: The obtained bounding box will be the smallest bounding box containing all human face objects.
- The input of the problem can be an image, a video, or a source streamed from the camera.
- Less than 25 people in the photo will produce good results.
- Minimum face size for good detection: Area greater than or equal to 1/25 of the total area of the input image.
- The problem can handle Occlusion cases, but the accuracy may be reduced.

IV. ALGORITHMS AND EXPERIMENT:

YOLOv5

- YOLOv5 is an object detection method of the YOLO family and belongs to One-stage Detector, the input image will be fed into the model only once to return the bounding box and class results of the object.
- YOLOv5 was introduced with many different versions, in which the lightest version, giving the fastest detection time is YOLOv5n.

Metrics

The metrics used in this study include:

- Precision
- Recall
- Average Precision (AP)
- Mean Average Precision (mAP)

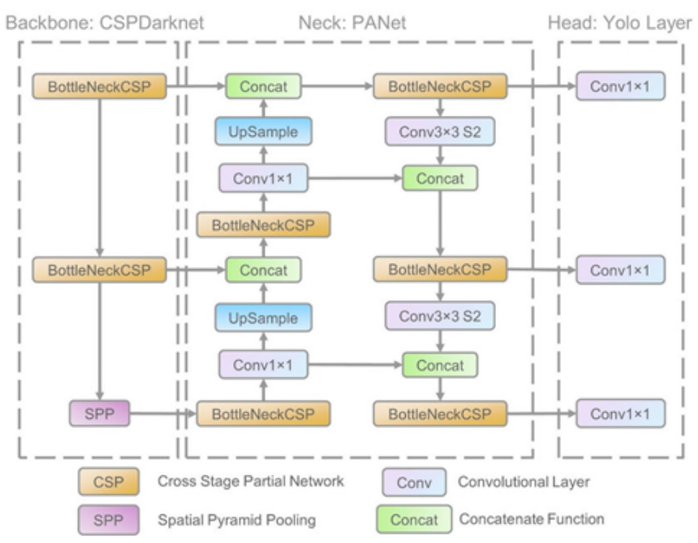

Fig 3: YOLOv5 architecture

Class	Precision	Recall	mAP@.5	mAP@.5:.95:
all	0.89	0.799	0.893	0.571
mask_weared_incorrect	1	0.731	0.906	0.561
with_mask	0.886	0.871	0.919	0.617
without_mask	0.783	0.795	0.855	0.535

Table 1: Experimental results

Conclusion

The mask detection problem in general gave relatively acceptable results, the model can be performed in real time with not too great a delay, the solution also solves the occluded and small objects quite well. However, in some cases of wearing the mask the wrong way, the model is still confused with the rest of the cases and appears a situation where one object locates but two bounding boxes are created. In the future, the team will continue to learn and overcome the existing weaknesses.