CS5242 Group 10 Project Plan: Facial Mask Detection

1 Motivation

COVID-19 has greatly affected people's lives. Masks have proved to be an effective method to stop the spread of the virus since 2020. Although Singapore has partially loosened its coronavirus controls, people are still required to wear masks in public places such as buses and subways. Therefore, intelligent identification of whether passengers are wearing masks is a research topic of great significance.

2 Description

We aim to locate human faces in the image, and detect whether the person is wearing a mask or not. The classification results are divided into three categories: face with a mask, face with no masks, and no human faces. The method is elaborated in the proposed solution part.

We expect our model to achieve considerable accuracy without loss of generalizability and interpretability. Through visualization and cross-model comparisons, we are able to provide both qualitative and quantitative performance analysis for our experimental results, which inspire future discussions into facial mask detection problem.

3 Data Collection

In order to ensure the randomness and representatives of the dataset, we decided to obtain pictures of wearing masks from different websites using scrapper, including but not limited to Flickr, Google, and Video Snapshot. Use the keyword 'people', 'people wearing mask' and 'bus' to collect three types of picture, so that there is no need to manually make labels. To make the follow-up task more accurate, we used YOLO to capture faces. Then to ensure the accuracy of the dataset, the images will be manually checked to guarantee if the face was captured correctly, or if the 'mask' label was correct.

4 Problem Formulation

Facial Mask Detection is a classification task, as shown in equation (1), where the input is an image, i, and the target is an integer indicator, c, of the predicted label l.

$$l = \begin{cases} \text{wearing a mask} & c = 0 \\ \text{not wearing a mask} & c = 1 \\ \text{not a face} & c = 2 \end{cases} \tag{1}$$

Formally, given a database $\mathcal{D}$, where each sample is an image-label pair $\langle i, c \rangle$, we aim to learn a predictor, p, that maximizes the prediction accuracy, as shown below in equation (2):

$$Acc = \mathop{\mathbb{E}}_{\langle i,c\rangle \in D} \mathbb{1}_{argmax(p(i))=c} \tag{2}$$

5 Proposed Solution

The process can be divided into the following parts.

- Develop our own scrapper to fetch image data from the Internet.
- Data cleansing, and exploratory data analysis(EDA).
- Implement MLPs, CNNs, RNNs as the baseline, Transformer and ResNet as improvement, and optimize hyperparameters.
- Performance analysis, visualization, results interpretation, and conclusion.

The procedure above serves as the basis of our facial mask detection model. For the part of EDA, we may utilize statistical distribution, feature importance analysis, clustering, and other visualization techniques. For model performance analysis, we may include accuracy and loss, confusion matrix, speed, bias and variance, and so on. We will also look into hyper-parameter settings and different data resources to see their impact on model performances.

In order to further enhance model affects, we apply extra techniques including data augmentation, cross-domain data alignment, and pre-trained models. These methods provide us an opportunity to perform more fine-grained classification, such as whether the mask covers a person's nose and mouth so that he is wearing his mask in a correct and standard way.

The initiative points are listed below:

- New models exploration: YOLO, VIT
- New data augmentation methods: cropping, zooming, darkening brightening/color modification, grayscaling, changing contrast adding noise, random erasing
- Multiple data resources: Google, Flickr, Unsplash, Pexels, Pixabay, Rawpixel

The project milestone and task allocation are shown below in Fig.1 and Table.1, respectively.

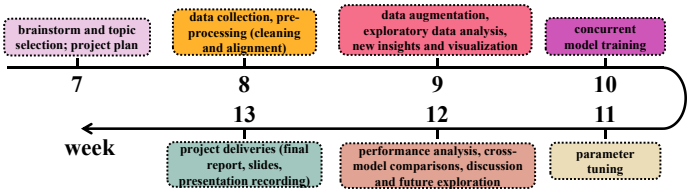

Figure 1: Weekly MileStone.

Huang Ziyu	Data collection Model training: ANN(Transformer)
Li Zhaofeng	Data visualization Model training: RNN Results visualization
Wang Yuchen	Data cleansing: detect and locate the face in an image Model training: MLP, CNN

Table 1: Task Allocation.