

# CV-AS4-Jiazheng Li

Jiazheng Li

April 2023

## 1 Task 1

### 1.1 Q1

Several commercial computer vision systems (Microsoft, IBM, Face++) have been criticized due to their asymmetric accuracy across sub-demographics. Studies found that the commercial face gender classification systems all perform better on male and on light faces. Existing public face datasets are strongly biased toward Caucasian faces, and other races (e.g., Latino) are significantly underrepresented.

### 1.2 Q2

The main metric they used is classification accuracy, which is actually used in different settings. Below are all metrics used in the experimental part.

- Figure 1: Simply calculate the ratios of different races in the datasets.
- Table 2&3: Classification accuracy for cross-dataset usage.
- Table 4: Classification accuracy and maximum accuracy disparity.
- Figure 5: Distribution of pairwise distances of faces in 3 datasets measured by L1 distance on face embedding.
- Table 5: Mean and Standard deviation of classification accuracy on different datasets and tasks.
- Table 6&Figure 6: Classification accuracy of different commercial CV APIs.
- Table 7: Face detection rates of commercial CV APIs on FairFace dataset.

### 1.3 Q3

For face datasets, bias is defined as they have obvious different numbers of images for different races, genders and ages, especially White people dominating the datasets. For measurement, it could be calculated as Figure 1 in the original paper. For face recognition tasks, bias is defined as the model exhibits different ability for different races, genders and ages. For measurement, we could use standard deviation and maximum accuracy disparity.

## 2 Task 2

I used deepface to detect and determine the age the faces in the fairface validation set (1.25 padding). It has 10,954 images in total and deepface detects 7,357 faces (detection rate: 67.16%). Here is the accuracy of age prediction for each subset.

Table 1: Accuracy

	Black	East Asian	Indian	Latino	Mid Eastern	SE Asian	White	Mean	STD
Female	.239	.204	.208	.203	.229	.178	.223	.212	.019
Male	.228	.249	.246	.247	.251	.256	.261	.248	.010
Mean	.233	.227	.227	.225	.240	.217	.242	.230	.024

The accuracy and detection rate are relatively low, which is possibly because the images are cropped and the task is difficult (9 classes). From the table we can see that

- As to gender, mean accuracy for female is lower while standard deviation is higher to some extent.
- As to race, mean accuracy is not consistent.

Also, I use the maximum accuracy disparity as the original paper which is 0.167. In summary, according to the mean accuracy, standard deviation and maximum accuracy disparity, I argue the backend (opencv) of deepface is biased.