Introducing WhichFace: A mobile application designed to improve face recognition abilities

Marco A. Sama, Kang Huang, & Kaishan Zhang

ECE1778

Professor Jonathan Rose

Word count (total): 1,497

Specialist Content: 379

All others: 1,118

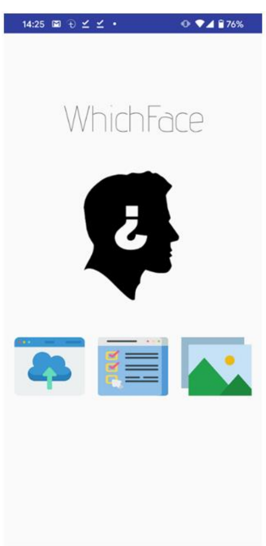
Introduction

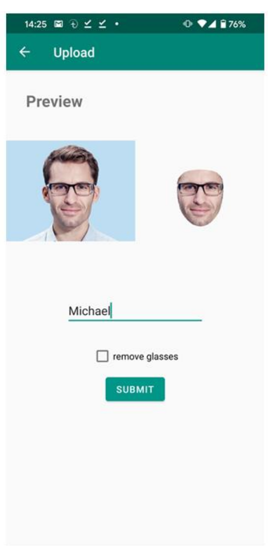
The purpose of *Which*Face is to help users recognize human faces. To that end, *Which*Face extracts faces from an image and isolates it against background noise and other non-face identity information (e.g., hair, colour of the face). The application also possesses the ability to remove glasses, which can be used to recognize someone as, like hairstyle, glasses come in many varieties and may be unique to certain individuals.

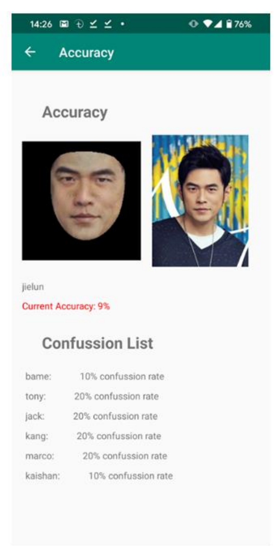
This application is important as face blindness, referred to as prosopagnosia, affects roughly 1% of the population. The inability to recognize important people can lead to difficulty in social situations and in the workplace. *Which*Face allows users to upload familiar faces and train their recognition ability on them. Even if face-learning effects do not generalize to novel faces, they will at least gain some recognition ability for the target familiar faces used in training.

Statement of Functionality

The *Which*Face application contains three main functions that users can select from the main screen (leftmost screenshot). First, users can upload photos from the device storage (middle-left screenshot). Photos can be either downloads or taken via the device camera. The user enters the name and can select an option to remove glasses. A preview of the isolated face is shown next to the uploaded photo. When the image is submitted, the *Which*Face application stores the isolated face, and will remove glasses if the user has that option selected.

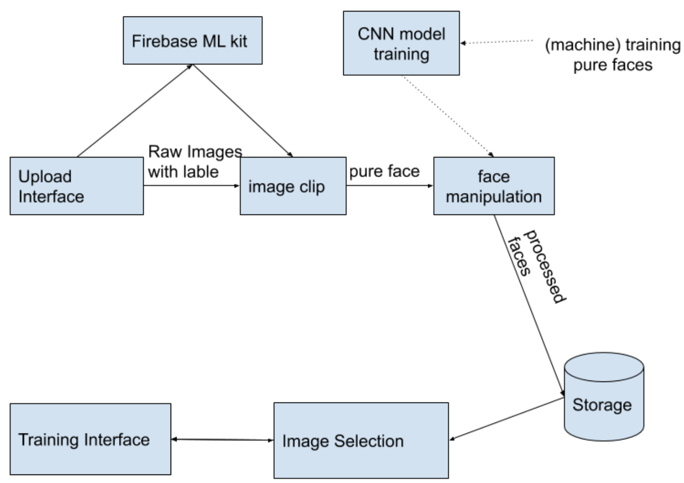






The second, and main function of the application, is the test feature (middle right screenshot). Users select between 1 and 10 images for testing. They see each image on screen, one at a time, and enter the name using the device keyboard. When selecting faces from the photo library to train on, the application will always select those with the lowest accuracy (note, new faces submitted are given a default accuracy of 0%). After testing, the application reviews accuracy for each labelled face. Finally, the third feature is the photo library (rightmost screenshot). This shows all the uploaded images, along with the label accuracy rate and confusability. Confusability tells users how often they mistake a face label with another face (i.e., how often do they mistake "Joe" for "Jim").

Overall Design



*Which*Face starts with the upload interface where users input the image into the application, and label the image. This image is fed through the Firebase ML kit which segments and provides a face mask for the image. The image clip goes through the face manipulation component, which is put through CNN model trained on face recognition. The model processes and removes the face from the rest of the image. Processed faces are then placed in the photo library storage. Through image selection, users choose which images from storage they want to use to train their recognition abilities. Then, users are led through the training interface, viewing faces and assigning them a label.

Reflection

Marco

This was my first team project where each member had diverse education backgrounds. This experience taught me how to partition goals appropriate for each teammate's expertise background. When I began the project, I had many ideas and many features I would have liked to implement. If I had the option to do this again, I would have narrowed my focus to one idea with one or two specific features. This would save time and increase productivity, as the team would not have to sift through the feasibility of each idea.

Kang

I learned how to cooperate in a team environment when developing an application, specifically how to combine progress from two developers using GitHub, where we discussed the detail behind features before implementing them. If I were to start again, I would not try to program the entire application from scratch, but look up open source starter code. This would allow me to focus my abilities on the core features of the application (e.g., the face manipulation) rather than the skeletal environment.

Kaishan

I learned to focus on developing the easier parts of a project first, and then try to complete more complex ideas. This makes it less stressful if some part of the project need to be redone during development. I also learned how to develop a project focused on someone else’s ideas, making it come to fruition. This required me to learn new techniques I would not have thought of otherwise. If I had to do this again, I would focus on developing the easier features first, before trying to figure out the more complex parts.

Contribution by Each Group Member

Marco

I was the specialist. I came up with the idea, did research behind face-processing deficits, and designed the UI. I would demo the application during each stage, testing its ergonomics and ease of use. Outside of that, my expertise came from knowing what non-identity information from a face can contribute to recognition (e.g., hair, glasses) so that I can instruct the programmers to have those features removed during application development.

Kang

I was one of the programmers. I designed the UI for the application based on the recommendations of the specialist. I also built the image uploader, including the part where it detects and cuts faces from the raw image, then uploads both the raw and cut image to the cloud server.

Kaishan

I was also a programmer. Before Spiral 2, I focused on implementing the basic functionality of image selection, the photo library, and the user-training interface. After Spiral 2, I implemented the DYI training model with the existing LFW dataset and a pretrained VGG19 model to successfully achieve our goal of removing glasses from the face. I also built the confusability table near the end of development.

Specialist Context

*Which*Face itself does not have any direct influence to my research field, as I study how face perception works, with references to the clinical side as sometimes how something breaks can inform us how something works. However, the use of online and mobile testing has started to gain traction in this area of research. Knowing that an application can be developed to test and train users allows me to explore the possibility of large-scale behavioural testing using mobile devices. Thus, a novel application designed around psychophysical testing can be developed for mass mobile participant recruitment for the specific types of research questions we ask. It also creates possibilities for long-term studies that investigate changes over time. Specifically, when we test participants in the lab, we measure our variables one time from that individual. Our results are a static snapshot of that individual. Deploying our studies through an application would let us investigate multiple temporal variables that change dynamically, and provide an opportunity to test how these variables play out in the real world. For example, we could track how statistical processing may change across time of day. One difficult variable that researchers have chased with little success is exploring how these statistical visual processes change during phases of the female menstrual cycle, as it is difficult to recruit participants at particular phases. With a mobile application, women could test themselves at specific parts of each phase with good precision.

The difficulty with developing an application to conduct research is the introduction if extraneous variables that cannot be controlled for outside of a lab setting. External lighting and environmental noise is held constant when testing within a lab environment. With 20-30 participants in a lab, this noise can significantly influence our data. However, with large-scale studies where hundreds of participants can participate, it is assumed that any noise introduced by these variables will be averaged out. Other variables can be controlled for by the application. Using the camera, we can measure the distance the user is from the application, controlling viewing angle.

Thus, while *Which*Face itself is not applicable directly to my area of research, it opens the possibility of deploying research studies through phone applications to explore a wide variety of variables that cannot be measured in the lab.

Future Work

Apart from the UI, there are a few possible areas for improvement. The most pertinent area to improve would be the testing phase. Currently, users see a face and enter the name. If there is a typo or the label is not case-matched, the application records this as an incorrect response. To rectify this, labels would have to be case invariant, and could even match to situations where spelling is close (e.g., “Jon” and “John”). We would also recommend adding in a possible drop-down menu to select names, instead of typing them in, or a speech-recognition option. Additionally, we could implement an easy mode where, instead of labelling faces, users see two faces and one name which they have to match to one of the faces.

Outside of the testing phase, we think an autorecognizer would be an important addition when uploading faces. When uploading a face, the autorecognizer would scan the internet and try to find a preexisting match. This would work well for famous faces.

Statement of Public Post

All three authors of the application agree that the video of the final presentation, the report, and the source code may be committed to the Course Website.