Virtual Assistant Attention Detection

Ethan Callanan, Jae Makitalo, Kavin Zhu, Ella Duffy

QMIND - Queen's AI Hub

Queen's University, Kingston, Ontario K7L 3N6, Canada.

Virtual assistants have seen an incredible rise in popularity over the past few years; Last year there were approximately 4.2 billion virtual assistant devices in use, and by 2024 experts predict that number will rise to over 8.4 billion. Despite the number of devices, they are almost all voice and audio based for activation, requests, and responses. This project utilizes a multitask cascaded convolutional network (MTCNN) to detect faces, then a convolutional neural network (CNN) classifier to visually recognize when a user is speaking to the device. This provides an alternative way to interact with virtual assistants, offering more choice to the user and increasing the accessibility of these products. The facial detection model performed with 93.7% accuracy, and the CNN classifier with 97.2% testing accuracy and 0.079 mean square error (MSE) testing loss on our dataset. We aim to improve our model by expanding the dataset to include more people and more variation in the camera angles. This will improve the generalizability of the classifier to less familiar contexts.
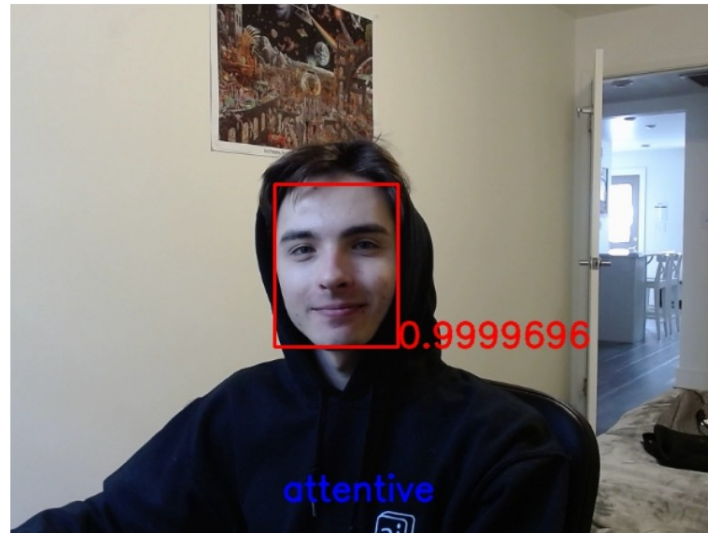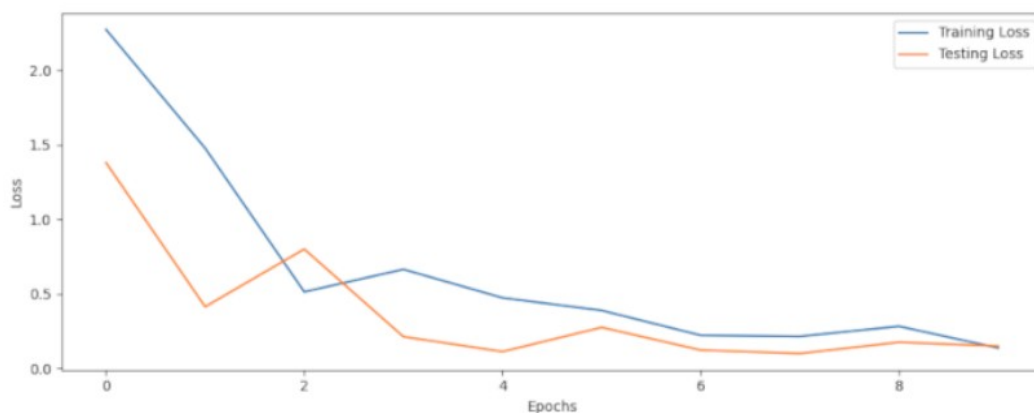

Figure 1: Live face detection and attention classification by the model.

Figure 2: Learning curve of classifier.