Facial Expression Detection & Emoji Conversion

CSCI 2470 Deep Learning Project

by Elbert Wu (ewu12), Bohao Wang (bwang98), Ying Sun(ysun141)

Introduction

Under the exploding development of computer vision, facial recognition is now becoming increasingly important. In fact, facial recognition in today's era is no longer just about whether some specific facials are from an identical person or not in order to achieve the function of facial recognition unlocking. Instead, facial recognition is now evolving towards the direction of capturing changes in facial expressions and making accurate classifications of people's emotions. At the same time, the rise of the idea, meta, has given us some inspirations. If we can successfully capture people's expressions through facial images with the application of deep learning techniques and convert them into corresponding emoji expressions in the virtual world, this would help us build more realistic connections between the real world and the virtual world. This project is made up of three parts. The first part is a classification problem under the umbrella of supervised learning to do emotional analysis, and the second part is simply converting the resulting classifications into emojis which have corresponding emotions. The third part is the emotion detections by using web-camera.

Data

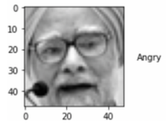
The dataset we plan to use is FER-2013. It contains approximately 30,000 facial grey scale images of 7 different expressions. The emotion expressions include happy, disgust, sad, neutral, angry, surprise, and fear. 35,887 grayscale 48x48-pixel images are contained. The data set we use is from Kaggle. However, this data is imbalanced as it has much more data which corresponds to the emotion of happy and the emotion of fear than other emotions

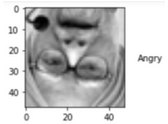
Data Source:
https://www.kaggle.com/datasets/nicolejyt/facialexpressionrecognition

Methodology

1. Preprocess

Data Augmentation is applied in order to create variations of the data and further avoid the problem of overfitting. A matrix transformation is applied to pixels such that the image is flipped up and down.





2. Model

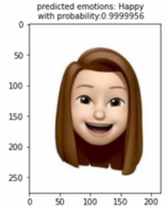
Convolutional Neural Network(CNN) would be the basic structure of this project. Support Vector Machine(SVM) is also applied as it has advantages of doing classifications. However, the accuracy score of using support vector machine classifier is less than 50%, and it has extremely long running time. This issue lies in the fact that this data set is a large data set, especially after data augmentation. Our final model has three CNN layers and three fully connected neural network.

3. Train & Validation & Test

The original data set has three types of sets, the training set, the public testing set, and the private training set, which includes 80%, 10%, and 10% of the total data respectively. Based on divisions, I consider the public testing set as the validation set and the private testing set as the testing set.

4. Convert to Emoji

Ying provides Emoji through Apple Emoji conversion techniques.

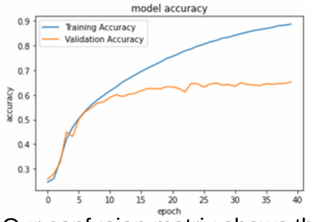


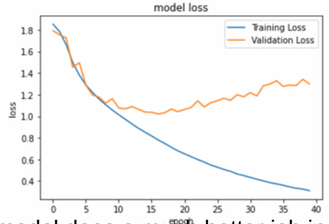
5. Web Camera

We could access the computer's camera to detect live emotions

Result

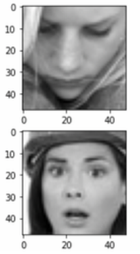
With the batch size of 64 and 40 epochs, our training accuracy is almost 89%, and the validation accuracy is almost 66%. The testing accuracy is 64.29%. From the graphs below, we see there exists some degrees of overfitting problems ad the gap between the training curve and the validation curve becomes significantly large.





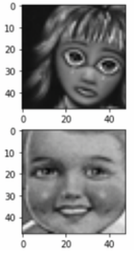
Our confusion matrix shows that this model does a much better job in detecting emotional expression of happy, which is number 3, and fear, which is number 5. This is due to the imbalanced data structure as discussed in the section of Data.

Some examples of output is given below



predicted emotions: Sad
with probability:0.9916331

correct emotion: Sad



predicted emotions: Sad
with probability:0.95543855

correct emotion: Sad

predicted emotions: Surprise
with probability:0.9983491

correct emotion: Fear

predicted emotions: Happy
with probability:0.99902093

correct emotion: Happy

Discussion

In general, this model does a moderate job in terms of its accuracyt score and its ability of converting to emojis and detecting live images. However, we might need to further add more data augmentations such as rotations or so in order to avoid overfitting problems. We might need to consider combining SVM and CNN together or use modified CNNs such as Visual Geometric Group Network(VGG Net) as well.