Music Playlist Generator Based on Facial Expression

By Tysean Canada & Elizabeth Fatade

Abstract

Emotion detection from facial expressions is a topic of increasing interest in the field of machine learning. In this paper, we showcase how we were able to train a neural network and build a model for detecting emotions from facial images using a combination of Convolutional Neural Networks (CNNs) and relevant python libraries to contribute to our project. Our approach achieves an average enough accuracy in emotion detection. Our project uses emotions to play music that will complement an individual's current emotion. The emotions we are recognizing are happy, angry, sad, neutral, surprise, fear, and disgust. We've done research to curate a playlist based on each emotion to play relevant music.

In order to evaluate our method, we conduct experiments on a dataset of facial images and compare the results to those of other state-of-the-art emotion detection techniques. The results show that our method is able to detect emotions such as happiness, sadness, anger, and etc… Furthermore, our project is able to run in real-time, making it suitable for applications in various fields including psychology, media, and human-computer interaction.

Introduction

The ability to automatically detect emotions from a person's face has numerous potential applications, such as improving human-computer interaction, aiding in the diagnosis and treatment of mental health disorders, and helping individuals with social communication impairments. We plan to use this technique in our project to help alleviate some negative moods and reduce individuals' stress levels. If our true mission is to lower stress levels in individuals through musical selections, we need to be cognizant of the effects that music can have on minds for negative reasons. In the article it states "Depression has also been linked to certain music listening choices and motivations. It has been found that depressed people are more likely to use music to reflect mood and express emotion"(Thomson). With that being said, it is imperative that our playlist and musical selections work to reduce stress and not add to depressive thoughts.

Traditionally, emotion detection has been approached using methods such as facial action coding, which involves manually identifying and labeling specific facial movements and gestures. However, this approach is time-consuming and subject to human error. With the advent of machine learning, it is now possible to automatically detect emotions from facial images using advanced algorithms.

Our project will be a step in advancement to the way we listen to music. Instead of having to pick a playlist, the mood detector would be able to play music that compliments your current feelings. This is important for a numerous of reasons. One reason is because our project can be a new technology to be used in the therapy industries to provide people with comforting and soothing music to their problems. Another reason is that this project is just the beginning of what mood detection methods could be used for. They could be used for many other therapeutic remedies other than the World's largest, which is music.

The rest of this paper is organized as follows. In Section 1: Method, we describe our proposed method in detail. In Section 2: Execution , we go over how we were able to put the project together piece by piece. In Section 3: Our Contributions, we elaborate on the ways we were able to make this project unique including what we were able to add and work on from other sources. Section 4: Results and Discussion present the results of our experiments as well as the . Finally, in Section 5, we conclude the paper and discuss possible future work.

Overall, our approach offers a promising solution for detecting emotions from facial images, with potential applications in various fields including psychology, neuroscience, and human-computer interaction. In the article it states ""I control my emotions by not expressing them." Items are answered on a seven-point scale ranging from "Strongly disagree" to "Strongly agree". Reappraisal was expected to be positively correlated with MMR, since music is shown to advance mental restructuring of thoughts."(Saarikallio). We chose to do music as the application of the mood detection method for this specific reason. Music allows people to get in tune with their emotions.

Section 1: Method

First, we will need to gather a dataset of images of faces that are labeled with their corresponding emotions. There are several publicly available datasets that can be used for this purpose, such as the CK+ dataset and the MMI Facial Expression Database. These datasets typically contain images of people displaying a range of emotions, including happiness, sadness, anger, fear, surprise, and disgust. The dataset that was used for this project comes from Keggle, an online community of data scientists and machine learning practitioners. The link to this dataset can be found as a part of our sources.

Once we have collected a dataset, we will need to preprocess the images to make them suitable for training a machine learning model. This may involve resizing the images to a consistent size, converting them to grayscale, and possibly applying other image processing techniques to improve their quality, in addition to data cleaning.

Next, we will need to train a machine learning model to detect emotions from faces. One approach is to use a convolutional neural network (CNN), which is a type of neural network that is well-suited to image classification tasks. To train the CNN, we will feed it our preprocessed images along with their corresponding labels (i.e., the emotions they depict). As the CNN processes the images, it will learn to recognize patterns in the pixel data that are associated with different emotions. The trained CNN will be saved as a model in an h5 file.

Once our CNN has been trained, we can use it to predict the emotions of new, unseen images of faces. This can be done by providing the CNN with an input image and then examining the output of the network to determine which emotion it has predicted for that image.

In order to improve the accuracy of our emotion detection model, we may want to consider using techniques such as transfer learning and data augmentation. Transfer learning involves using a pre-trained model.

Section 2: Execution

As per usual, we began by looking through many different sources trying to find the perfect way to start and what we would need to implement. We decided to train our own model rather than using an already trained model as a way to sort of challenge ourselves and add an extra component to this project. This did come with its own challenges, however.

We obtained our dataset from Kaggle as it is a very popular site containing many datasets. The dataset is titled XYZ. The set of images were divided into two directories: train and validation, with each containing images of Anger, Happy, Disgust, Fear, Neutral, Sad and Surprise. After going through the directories we did notice that there was significantly more "happy" images in both the train and validation directories. We did not do any data cleaning initially before training the model to get an idea of what type of errors were in the data we were dealing with.

Examples of Images received from the Kaggle dataset

We were able to find useful pieces of code from online sources to train our neural network, utilizing deep libraries such as Keras and Tensorflow.

For each picture used to train our model we set them to have a size of 48 and used a batch size of 128. These numbers came from observations of others who used these numbers. Using the 7 emotion classes that we described above, we built a CNN with 5 layers and a learning rate of 0.0001. After trial and error with different batch sizes, learning rates and layers, we concluded that these three numbers worked the best together. We trained our data over 48 epochs and plotted the models accuracy. We discuss more of the accuracy results in the results section.

After obtaining our model we used that in combination with the OpenCV python library to create a video frame that captured live video and displayed the dominant emotion the model received from the users face in a box around the face detected.

In addition to OpenCV, we passed the dominant emotion on to a playlist maker generated using Pygame mixer that had a list of saved songs under a certain folder that corresponds to the emotion. The music player allows for pause, play, next and end functionalities with autonext features (immediately plays the next song after the first one finishes).

Section 3: Our Contributions

One major component to our project is the User Interface design allowing the user to have an interactive experience with our program. The user interface for this project is something that we created from scratch without the use of any source code. Our plan for this aspect of the code was to make the designs recognizably innovative with artwork from our presentations earlier in the class.

Another contribution to this project was working on the dataset to increase the accuracy of the random images. The way we did that was added photos from another random data set that catered to all emotions other than happiness. With the balanced out dataset our program runs more smoothly and can better detect moods because the datasets are no longer biased to the happiness training dataset. This was essential in the development of our program providing the accuracies we desired.

Section 4: Results & Discussion

Emotion detection from faces, also known as facial expression recognition, is a rapidly growing field of study with numerous potential applications. By using advanced computer vision algorithms, it is possible to analyze a person's facial expressions and identify emotions such as happiness, sadness, anger, fear, and surprise. This can be done in real-time using video feeds from cameras, or by processing images of faces.

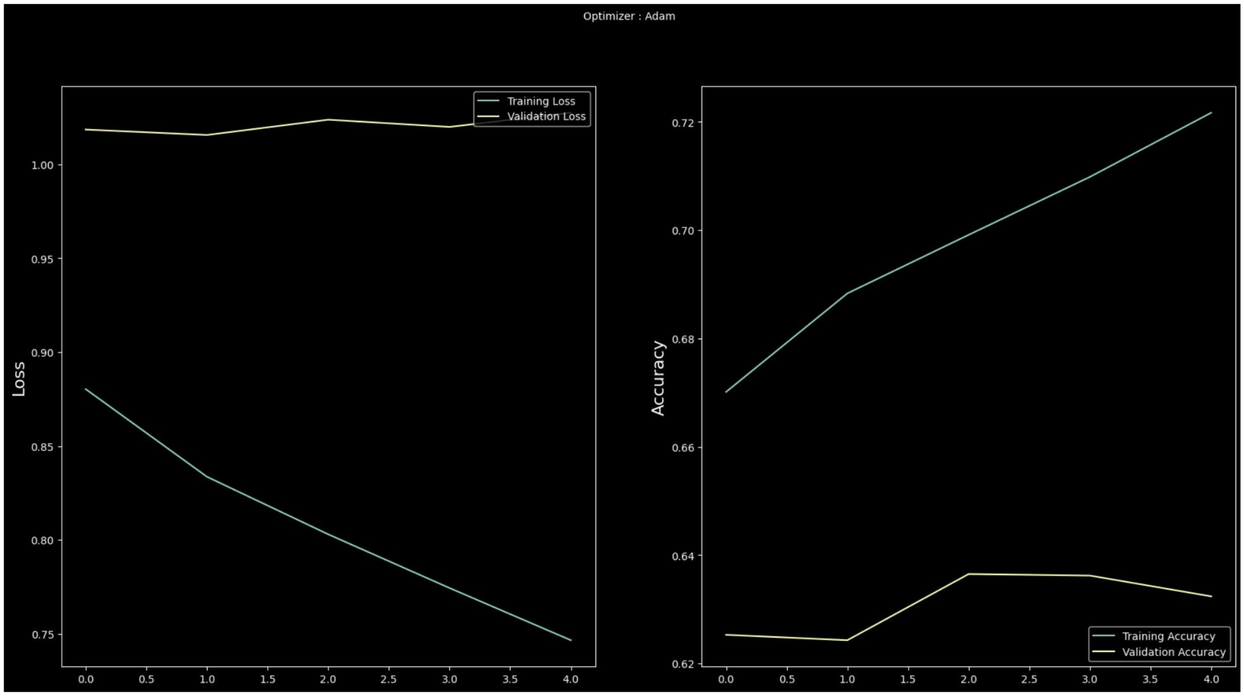
One of the key challenges in emotion detection from faces is that people can express the same emotion in many different ways. For example, a person might smile with their mouth closed to show happiness, or furrow their brow to show anger. Additionally, people from different cultures may express emotions differently, making it difficult for a facial expression recognition system to accurately identify emotions in all cases.

Another issue is dealing with images or live videos of people who have glasses or facial hair. This changes the scope of our project forcing us to increase a number of variables that attribute to many human's qualities. In an article states "Such a system should also be invariant to different distraction like glasses, different hair styles, mustache, facial hairs and different lightening conditions." (Mehendale)With these added variables to human faces, this increases the margin for error and lessens the accuracy of our facial detection methods.

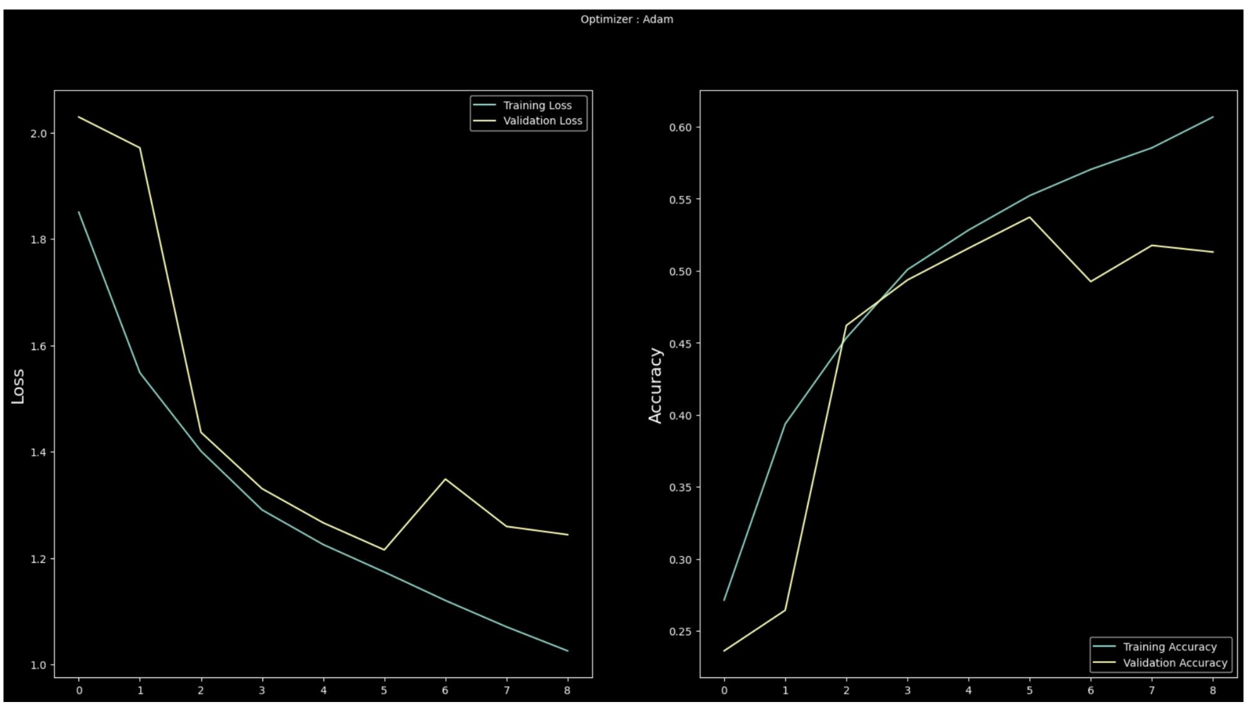
When testing our model without cleaning the data set (as we stated earlier), we were able to plot the accuracy results on a graph, Figure (a) and notice in great detail that our model was definitely overfitting. In addition to that, it appeared that the dominant emotion it was recognizing was "happy", as expected since the happy folder contained significantly more images than the others. Our dataset was unbalanced and needed to be cleaned up a bit to provide more accurate results. We removed a number of happy images and went through most images taking out the ones that seeed unclear or that had unrecognizable faces.

We plotted the training loss vs validation loss and training accuracy vs validation accuracy. Before cleaning our dataset, we discovered that the the training loss goes down over time, achieving low error values whilst the validation loss was very high meaning the model is producing erroneous output. Our plot of training vs validation accuracy was quite the opposite as it had a low validation accuracy and higher raining accuracy, which was wrong.

Figure(b) shows the results of our accuracy after cleaning our dataset and retraining our model. As shown there is less difference between our valiadationa and training losses and accuracy.



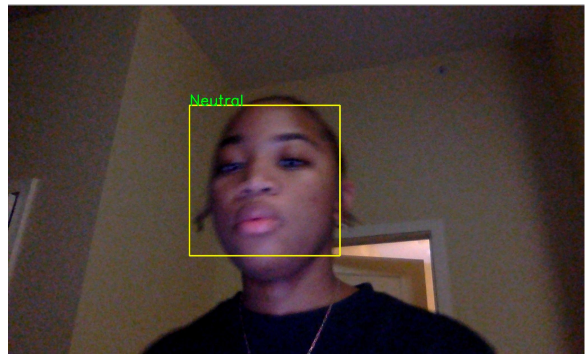
Figure(a): Accuracy Results before we cleaned our dataset



Figure(b): Accuracy Results after we cleaned our dataset

Overall, emotion detection from faces is a promising area of research that has the potential to improve a wide range of applications. However, it is important to carefully consider the potential ethical implications of using this technology, and to continue to improve the accuracy and reliability of emotion detection algorithms. In an article it states “Although the overall performance of the feature-level and decision-level bimodal classifiers was similar, an analysis of the confusion matrices of both classifiers reveals that the recognition rate for each emotion type was totally different.” (Busso) This quote refers to the different approaches we could have taken to complete this project and detect moods. One approach used in that research paper was a bimodal method that checked for each feature on an individual’s face and classified it as one of the specific emotions in their given scope. As this may be a reliable way to determine someone’s mood through facial expressions, one may note that it is necessary to look at the face as a holistic view because people's faces are not perfect with symmetry. Another con to that approach would be that skin colors can easily change the accuracy of the program.

At the end our model had an accuracy of about 72% which is not the greatest but a good start.



Figure(c): Image from trial run of program

```
In [5]: playlist = make_playlist(label)
        play_music(playlist)

        ['You Rock My World.wav', 'Tems - Higher(Lyrics Video).wav', 'novocane by frank ocean clean.wav', '[CLEAN] Tyler, The Creator - WUSYANAME (feat. YoungBoy Never Broke Again & Ty Dolla $ign).wav']
        Press 'p' to pause, 'r' to resume
        Press 'n' to play the next song
        Press 'e' to exit the program
          p
        Press 'p' to pause, 'r' to resume
        Press 'n' to play the next song
        Press 'e' to exit the program
          r
        Press 'p' to pause, 'r' to resume
        Press 'n' to play the next song
        Press 'e' to exit the program
          n
        4
        Press 'p' to pause, 'r' to resume
        Press 'n' to play the next song
        Press 'e' to exit the program
          e
```

Figure(d): Image of the interaction with the mixer library after classifying a neutral face

Section 5: Future Plans

Moving forward, we can work towards having a spotify api so that we would not need to save all the music required for the playlists to our device and for a more seamless connection. In addition to that, we can try using a pre trained model or improving our model for better accuracy.

References

Akmadan. "Emotion_Detection_CNN"

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., & Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. *Proceedings of the 6th International Conference on Multimodal Interfaces - ICMI '04*. https://doi.org/10.1145/1027933.1027968

Mehendale, Ninad. "Facial Emotion Recognition Using Convolutional Neural Networks (FERC)." *SN Applied Sciences*, vol. 2, no. 3, 2020, https://doi.org/10.1007/s42452-020-2234-1.

Oheix, Jonathan. "Face Expression Recognition Dataset" https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset

Proverbio, Alice Mado, et al. "Multimodal Recognition of Emotions in Music and Facial Expressions." *Frontiers in Human Neuroscience*, vol. 14, 2020, https://doi.org/10.3389/fnhum.2020.00032.

Rani, Jyoti, and Kanwal Garg. "Emotion Detection Using Facial Expressions -A Review ." Apr. 2014.

Saarikallio, Suvi. "Development and Validation of the Brief Music in Mood Regulation Scale (B-MMR)." *Music Perception*, vol. 30, no. 1, 2012, pp. 97–105., https://doi.org/10.1525/mp.2012.30.1.97.

Saarikallio, Suvi H. "Music in Mood Regulation: Initial Scale Development." *Musicae Scientiae*, vol. 12, no. 2, 22 Aug. 2008, pp. 291–309., https://doi.org/10.1177/102986490801200206.

Thomson, Cassandra Jane, et al. "The Relationship between Music-Related Mood Regulation and Psychopathology in Young People." *Musicae Scientiae*, vol. 18, no. 2, 2014, pp. 150–165., https://doi.org/10.1177/1029864914521422.