Facial Emotion Detection Using Deep Learning

Final Project of CSE465: Pattern Recognition & Neural Network

Team Members:

Simon Bin Ahmed 1811 772 042

Ahnaf Ismat Tasin 1731 754 042

Syed Maaher Hossain 1731045642

Abstract. Human emotion recognition from images is a strong and difficult study task in social communication. Deep learning-based emotion recognition outperforms standard image processing approaches. Our project presents the development of an artificial intelligence system capable of detecting facial expressions that identifies emotion. It covers the emotion detection technique in two straightforward steps: face detection and emotion classification. Our project utilizes a deep learning architecture based on convolutional neural networks to identify emotion in consecutive pictures. We were finally able to attain an accuracy of 0.8639 and validation accuracy of 0.6307.

Keywords
Facial recognition; Facial technology; Expression recognition; Image classification; Deep learning; Convolutional neural network; Haar Cascade Algorithm; ReLU

Introduction

Recognizing human expressions and emotions has captivated researchers' interest, as the ability to recognize one's own expressions aids in human-computer interaction, helps target appropriate advertising campaigns, and ultimately results in augmented and enhanced human communication by augmenting humans' emotional intelligence. There are several methods for examining the identification of human expressions, including facial expressions, body posture, and voice tone. We concentrated on face expression recognition in our project. Facial Emotion Recognition is a booming field of study, with several breakthroughs occurring in areas like as automated translation systems and machine-human interaction.

Related Work

A detailed study on facial emotion recognition is discussed in [1] which exposes the properties of the dataset, facial emotion recognition study classifier. Visual features of image are examined and some of the classifier techniques are discussed in [2] which is helpful in the further inspection of the methods of emotion recognition. This paper [3] examined the prediction of the future reactions from images based on the recognition of emotions, using different classes of classifiers. Some of the classification algorithms like K-Nearest Neighbor, Random Forest are applied to classify emotions. Neural networks arise tremendously which attempts to solve problems in data science. Various range of CNN, modelled and trained for facial emotion recognition is evaluated in [4]. Facial emotion Recognition is drawing its own importance in the research field. Facial emotion recognition is inspected and analyzed on all research areas [5]. This paper [6] studies different databases used for facial emotion recognition, features selected from facial expression images, classifiers used to classify different classes of emotions. This paper [7] explains about learning significant features such as Support vector machine training, local invariant feature learning, salient discriminative feature analysis for facial emotion recognition.

Methodology

Our proposed framework for Facial Emotion Detection is divided into two modules where the first module is “Model Training” and the second module is “Emotion Detection”. Model training exploits deep learning techniques, more specifically convolutional neural networks to train a classification model using publicly available facial expression image dataset from Kaggle. For the training procedure we used Keras [8] deep learning framework with Tensorflow [9] in the backend.

The second module accepts input in the form of sequential random pictures. The face was identified using the Haar Cascade algorithm. From images with random backgrounds, we determined the facial boundary. Then we clipped that area and calculated the likelihood of the input being matched with each label in the previously constructed dictionary. Whichever has the greater percentage will be sent to the output.

Dataset

The dataset was collected from Kaggle & this set of data consists of 48x48 pixel grayscale images of faces. We used greyscale images because if we used RGB images, the number of channels would increase by a huge margin and this would be very slow to compute as we do not have a dedicated GPU of our own. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

Model Training

Our model was trained using the Keras deep learning framework in conjunction with Tensorflow on the backend. We utilized convolutional neural networks and max-pooling to train our model.

The input layer consisted of a collection of frontal face expression photographs that were all scaled and displayed in 48x48 uniform proportions with a single channel representing greyscale images. The input layer is then processed using three successive pairs of convolution and max pooling layers, each of which has a 3x3 convolution filter and a 2x2 max pooling filter. We incorporate a Rectified Linear Unit (ReLU) as an activation function between each convolution and max pooling. Following the third max-pool layer, we added a layer to the network to flatten the output of the previous max-pool layer and transfer it to the first densely connected layer. Additionally, another (second) fully connected layer is then in work with activation function as ReLU before getting the output layer with activation function as Softmax. To train the final network, we utilized the Adam optimizer, categorical loss, and Accuracy as our cross entropy metric. The abstraction of our final trained network is depicted in Figure 1.

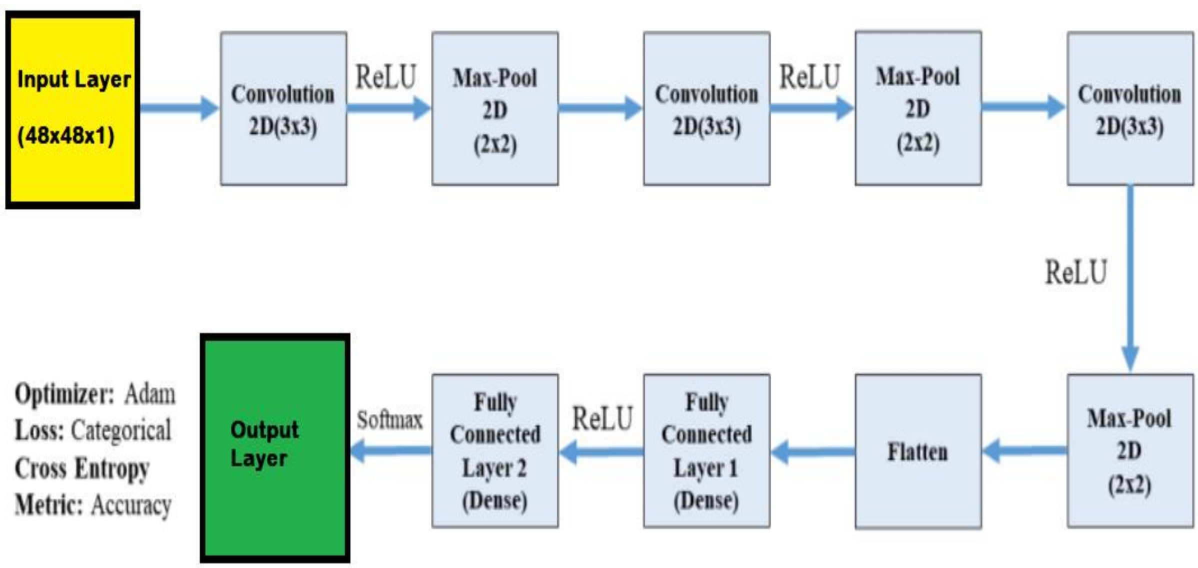


Fig. 1. Illustration of the Trained Network

Activation Function

Here we used the rectified linear activation function or ReLU as activation function over other existing activation functions as ReLU overcomes the vanishing gradient problem which aligns our model to learn faster and perform better [10].

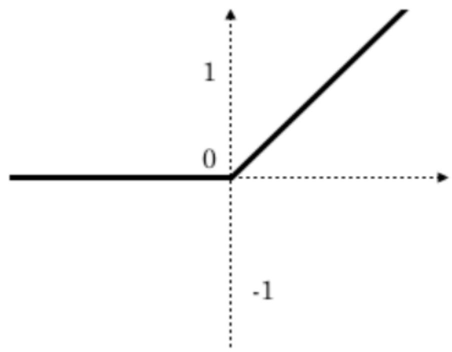


Fig. 2. The rectified linear activation function (ReLU)

Face Detection

It is an Object Detection Algorithm used to identify faces in an image or a real time video. The algorithm uses edge or line detection features proposed by Viola and Jones in their research paper [11].

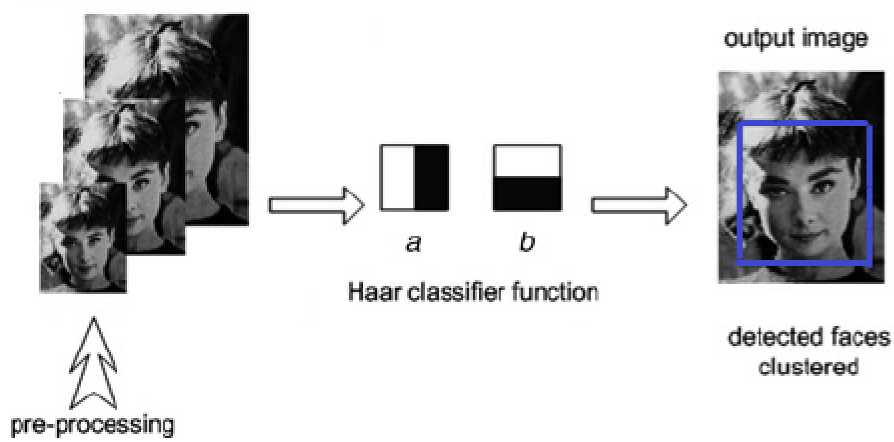
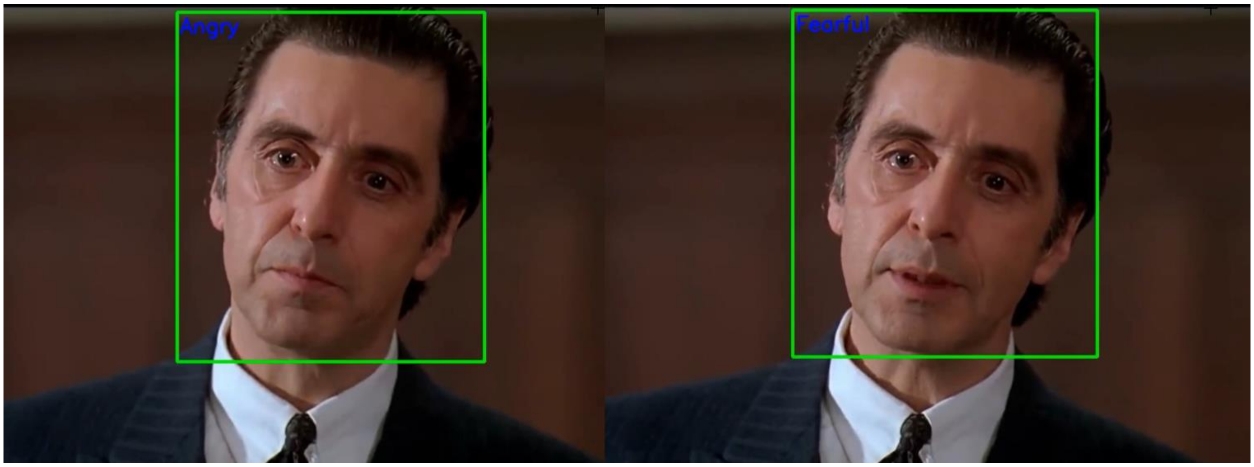
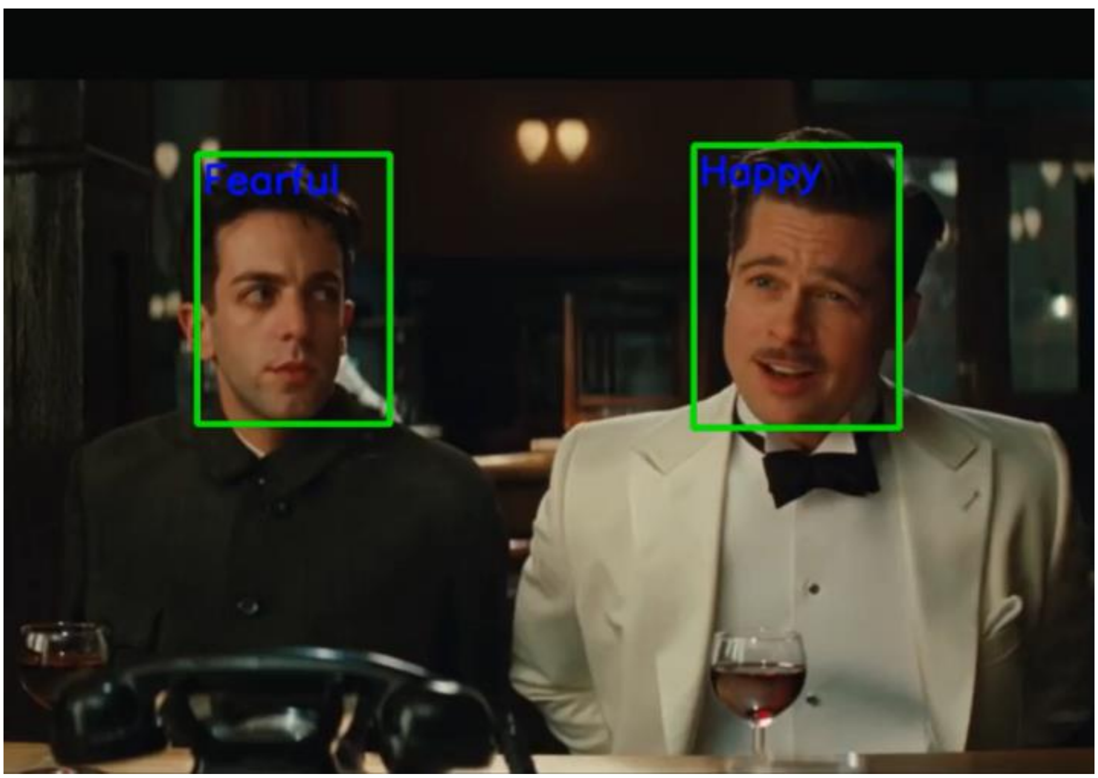


Fig. 3. Frontal Face Configuration (Haar Cascade)

Results



The input that we collected will be compared to the dictionary's label, and that label with the highest probability will be displayed in the output.



Our model is smart enough to identify the expression from images even if multiple faces are present in a single input image.

Additionally, it is able to capture expressions from sequential data. In sequential data, expressions change rapidly but our model is smart enough to identify the emotion and the change in emotion.

Validation

```
448/448 [==============================] - 451s 1s/step - loss: 0.4147 - accuracy: 0.8520 - val_loss: 1.1447 - val_accuracy: 0.6272
Epoch 48/50
448/448 [==============================] - 494s 1s/step - loss: 0.4026 - accuracy: 0.8548 - val_loss: 1.1612 - val_accuracy: 0.6285
Epoch 49/50
448/448 [==============================] - 471s 1s/step - loss: 0.3909 - accuracy: 0.8582 - val_loss: 1.1641 - val_accuracy: 0.6285
Epoch 50/50
448/448 [==============================] - 427s 954ms/step - loss: 0.3807 - accuracy: 0.8639 - val_loss: 1.1960 - val_accuracy: 0.6307
```

To train deep learning models, we utilized the Adam optimizer, which is a replacement optimization technique for gradient descent. Similarly, we employ a loss function called categorical cross entropy. In general, we utilize categorical entropy in multi-class classification problems when an input might belong to only one of a large number of potential categories and the model must select which one to use. Finally, we were able to achieve an accuracy of 0.8639 and a validation accuracy of 0.6307. We hypothesize that this is the cause for the model's low validation accuracy on significant differences between the type of data used to train the model and the testing data provided for assessment.

Discussion & Limitation

Due to our poorly configured hardware, training and evaluating the model took a long time. The primary disadvantage and constraint encountered during the project was the lack of powerful hardware for the training module. The training was conducted on a Dell 7000 Series notebook equipped with an Intel Core i7 processor, 4GB of RAM, and a lack of a dedicated graphics card. The final training session lasted around eight hours. Due of the long training period, experimenting with other neural networks was not feasible within the time constraints. On the other hand, we are convinced that, in addition to utilizing a larger dataset and a more powerful GPU, the training module may be significantly enhanced to obtain greater prediction accuracy.

Conclusion

We examined the fundamentals of deep learning, more precisely convolutional deep neural networks, in this project. Additionally, we became familiar with the most widely used deep learning framework among academics and developers. Due to hardware constraints and the inability to train the model on a big enough dataset, the accuracy did not reach the desired level. However, we are certain that further exploration of the neural network and longer training will result in improved performance.

Source Code Repository

References

1. G. Hintin,Greves and A. Mohemed, "*Emotion Recognition with Deep Recurrent Neural Networks*" in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649.
2. K.Weint and C.-W. Huaang, "*Characterizing Types of Convolution in Deep Convolutional Recurrent Neural Networks for Robust Speech Emotion Recognition*" pp. 1–19, 2017.
3. A. M. Yusuf, M. B. Mustaafa, M. and M. Malekzedeh, "*emotion recognition research: an analysis of research focus*" Int. J. Speech Technol., vol. 0, no. 0, pp. 1–20, 2018.
4. A. M. Badshaah, J. Ahmed and S. W. Baek, "*Emotion Recognition from Spectrograms with Deep Convolutional Neural Network*" 2017 Int. Conf. Platf. Technol. Serv., pp. 1– 5, 2017.
5. A. Routrey, M. Swaen and P Kabisetpathy, "*Database, features and classifiers for emotion recognition: a review*" Int. J. Speech Technol., vol. 0, no. 0, p. 0, 2018.
6. D. Lie and E. M. Provest, "*EMOTION RECOGNITION USING HIDDEN MARKOV MODELS WITH DEEP BELIEF NETWORKS.*"
7. C. Busiso et al., "*IEMOCAP: Interactive emotional dynamic motion capture database*" Lang. Resour. Eval. vol. 42, no. 4, pp. 335–359, 2008.

8. https://keras.io/
9. https://www.tensorflow.org/learn
10. Christian Versloot "*ReLU, Sigmoid and Tanh: today's most used activation functions*", Machinecurve.com, 4 September 2019
11. P. Viola, Michael J. Jones "*Rapid Object Detection using a Boosted Cascade of Simple Features*" Computer Vision and Pattern Recognition, 2001