Pakistan Institute of Engineering and Applied Sciences

Department of Computer & Information Sciences

Term Project Report

Subject:

Artificial Intelligence

Submitted By:

Group 11 Abdur Rafay
Ijtihad Hussain
Mohsin Raza
Shoaib Akhtar

Submitted to:

Dr. Umer Farooq

EMOTION RECOGNITION USING FACIAL EXPRESSIONS

Introduction

- Emotion recognition is the process of identifying human emotion. People vary widely in their accuracy at recognizing the emotions of others. By using Facial Emotion Recognition, businesses can process images, and videos in real-time for monitoring video feeds or automating video analytics, thus saving costs and making life better for their users.
- A technology which Uses biometric markers to detect emotions in human faces.
- This technology is a sentiment analysis tool and is able to automatically detect the expressions.
- So basically, there's already work done on emotion recognition and we can easily find it online but we have added some more by ourselves.
- AI can detect emotions by learning what each facial expression means and applying that knowledge to the new information presented to it. Emotional artificial intelligence, or emotion AI, is a technology that is capable of reading, imitating, interpreting, and responding to human facial expressions and emotions.
- Companies have also been taking advantage of emotion recognition to drive business outcomes. For the upcoming release of Toy Story 5, Disney plans to use facial recognition to judge the emotional responses of the audience. Apple even released a new feature on the iPhone X called Animoji, where you can get a computer simulated emoji to mimic your facial expressions. It’s not so far off to assume they’ll use those capabilities in other applications soon.

Background

Facial expressions are important in facilitating human communication and interactions. Also, they are used as an important tool in behavioral studies and in medical rehabilitation. Facial image based mood detection techniques may provide a fast and practical approach for non-invasive mood detection. The purpose of the present study was to develop an intelligent system for facial image based expression classification using committee neural networks.

With the development of artificial intelligence and deep learning, numerous FER algorithms have been proposed to deal with the expression information in facial representations, which has improved the accuracy of recognition gradually and achieved better performance than traditional methods. The tasks of FER can be mainly divided into two categories: static images (represented by photographs) and dynamic sequence (represented by videos) that take the dynamic relationship between the continuously changing images into account and therefore pose additional challenges than the former. In addition to the vision-based methods, other biometric techniques can also be adopted to assist the recognition of expression.

Limited by the hardware and insufficient processing capability, the majority of the traditional methods for FER employed hand-craft features or shallow learning, such as local binary patterns (LBP) and nonnegative matrix factorization (NMF). With the development of processing capabilities and computer simulation, all kinds of machine learning algorithms, such as Artificial Neural Networks (ANNs), Support Vector Machines (SVM), and Bayesian classifiers, were applied to FER, and the high accuracy has been verified in controlled environments so that the faces can be detected effectively. However, these methods were weak in generalization ability while this is the key to evaluate the practicality of a model. Deep learning algorithms can solve this problem, and it is also robust in the uncontrolled environments. Recent works have shown that convolutional neural networks (CNNs), because of their effectiveness in feature extraction and classification tasks, performed well in addressing the computer vision problems especially in FER, and numerous models based on CNN structure are proposed constantly and have achieved better results than previous methods. Simonyan and Zisserman adopted an architecture of very small (3×3) convolution filters to conduct a comprehensive evaluation of networks with increasing depth and the two best-performing ConvNet models were available publicly to facilitate the further research in this field. By increasing the depth and width of the

network while keeping the computational budget constant, Szegedy et al. introduced a deep convolutional neural network architecture named “Inception” in which the utilization of the computing resources can be improved significantly, and Jahandad et al. worked on 2 convolutional neural network architectures (Inception-v1 and Inception-v3) based on “Inception” and proved that these 2 models performed better than others, and Inception-v1 with 22-layer-deep network performed better than 42-layer-deep Inception-v3 network when facing low-resolution input images and 2D images of signatures; however, Inception-v3 outperformed in ImageNet challenge. The general trend of neural networks is to increase the depth of the network and the width of layer. In theory, the deeper the neural network models, the stronger the learning capabilities, but the more difficult to train. He et al. proposed a residual learning framework to reduce training difficulty of deeper networks and proved thoroughly that these residual networks are easier to optimize while increasing accuracy from the considerably increased depth. In addition, a part of researchers proposed that the accuracy of recognition can be further improved by combining CNNs with recurrent neural networks (RNNs) in which the CNNs are adopted as the inputs to RNNs.

During the past decades, online education has developed rapidly whether at universities or training institutions, which offers potential application opportunities for FER. Significantly different from the traditional face-to-face courses, online courses are often considered of less constraining force and effective communication, which will inevitably lead to faculty’s suspicions towards this novel educational method while there are several studies that argue that the students’ learning outcomes achieved by online education may be comparable to traditional face-to-face courses, except for the skills that require optimum precision and a greater degree of haptic awareness. It is undeniable that the rapid growth of online education can effectively provide the convenience and flexibility for more students, so it also has broad development space in the future; therefore, how to ensure that students keep the same level of concentration and learning efficiency as the traditional courses during online education is critical to promote the further development of online education.

In brief, the main contribution of this paper is as follows. By combining the existing online education platforms with facial expression recognition model based on the architecture of convolutional neural network, this work proposed a framework that enables real-time monitoring of students’ emotions in online courses and ensures that the feedback expressed by facial expression can be provided to teachers timely, so that they can flexibly adjust the teaching programs and ultimately improve the quality and efficiency of online education.

Applications & Importance:

- Market research
- Video game testing
- It is the most ideal way to assess the effectiveness of any business content
- It can provide valuable information about the sentiment of a target audience towards a marketing message, product or brand.
- It can really tell about a business startup whether it will succeed or not.

Data Set:

We used the following data set in our project https://www.kaggle.com/msambare/fer2013

Data Preprocessing:

The dataset contained only 7 emotions, so we manually added two more emotions which are Flustered and Confused.

Libraries:

- Numpy
- Keras
- tensorflow

```
In [13]: from tensorflow.keras.preprocessing.image import ImageDataGenerator
         from tensorflow.keras.utils import to_categorical, plot_model
         from tensorflow.keras import models, layers, regularizers
         from keras.preprocessing.image import ImageDataGenerator , img_to_array, load_img
         import numpy as np
```

Sample Data:

Setting Hyperparameters:

```
In [16]: batch_size =64
         target_size = (200,200)

         train_datagen = ImageDataGenerator(rescale=1./255)
         val_datagen   = ImageDataGenerator(rescale=1./255)
```

```
In [19]: # Compile Model
         model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

In [22]: num_epochs = 100
         STEP_SIZE_TRAIN = train_generator.n//train_generator.batch_size
         STEP_SIZE_VAL   = val_generator.n//val_generator.batch_size
```

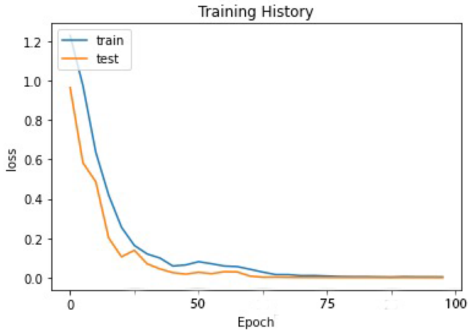
Model Summary:

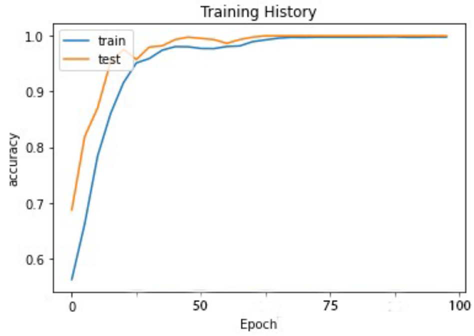
```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 48, 48, 32)        320

 conv2d_1 (Conv2D)           (None, 48, 48, 32)        9248

 max_pooling2d (MaxPooling2D  (None, 24, 24, 32)       0
 )

 conv2d_2 (Conv2D)           (None, 24, 24, 64)        18496

 conv2d_3 (Conv2D)           (None, 24, 24, 64)        36928

 max_pooling2d_1 (MaxPooling  (None, 12, 12, 64)       0
 2D)

 conv2d_4 (Conv2D)           (None, 12, 12, 128)       73856

 conv2d_5 (Conv2D)           (None, 12, 12, 128)       147584

 max_pooling2d_2 (MaxPooling  (None, 6, 6, 128)        0
 2D)

 flatten (Flatten)           (None, 4608)              0

 dense (Dense)               (None, 512)               2359808

 dense_1 (Dense)             (None, 64)                32832

 dense_2 (Dense)             (None, 9)                 585

=================================================================
Total params: 2,679,657
Trainable params: 2,679,657
Non-trainable params: 0
_________________________________________________________________
```

Model Training:

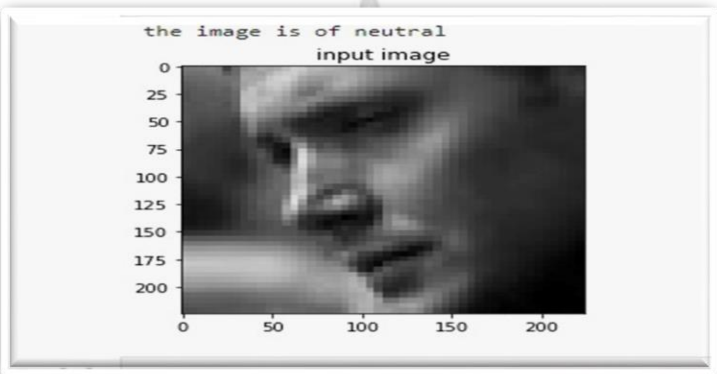
```
In [88]: # Train Model
         history = model.fit(train_generator, steps_per_epoch=STEP_SIZE_TRAIN, epochs=num_epochs, verbose=1, validation_data=val_generato

         Epoch 1/100
         62/62 [==============================] - 36s 562ms/step - loss: 2.1505 - accuracy: 0.1802 - val_loss: 2.3002 - val_accuracy: 0.
         1094
         Epoch 2/100
         62/62 [==============================] - 37s 599ms/step - loss: 2.1158 - accuracy: 0.1809 - val_loss: 2.2702 - val_accuracy: 0.
         1339
         Epoch 3/100
         62/62 [==============================] - 37s 601ms/step - loss: 2.1076 - accuracy: 0.1855 - val_loss: 2.2648 - val_accuracy: 0.
         1272
         Epoch 4/100
         62/62 [==============================] - 37s 593ms/step - loss: 2.0974 - accuracy: 0.1883 - val_loss: 2.1739 - val_accuracy: 0.
         1585
         Epoch 5/100
         62/62 [==============================] - 36s 582ms/step - loss: 2.0843 - accuracy: 0.1837 - val_loss: 2.2300 - val_accuracy: 0.
         1562
```

Graphs:





Testing:



Conclusion:

In this study, by combining the online courses platforms and a compact deep learning model based on the architecture of CNN, we construct a framework to analyze students' emotions according to their facial expressions from the perspective of computer simulation. The overall result can be presented in a histogram intuitively, and teachers can adjust their teaching strategies accordingly to improve the efficiency of online teaching.

With reference to the studies of Ekman et al. and FER2013, the emotions were classified into anger, disgust, fear, happiness, sadness, surprise, flustered, confused and neutral in the proposed framework. To verify the applicability of this framework in a real environment, we captured an image including the facial images of all participants at one time in a real online meeting; there were 12 participants in this meeting, and the captured time was determined at the end of meeting. A total of 12 faces were captured, of which 11 were effectively recognizable faces that contain enough feature points. By inputting this image into the applied CNN model, we obtained the emotional tags for each valid face and got the overall emotion at that time. It has been proved that the framework has good applicability in practical activities and plays a positive role in solving the problems, such as the lack of binding force on students, and teachers cannot get timely feedback. Ultimately, it will contribute to improving the quality of online education.

Despite the above benefits, there is still much room for improvement in this framework and its applications. From the perspective of technologies, with the development of computer simulation, algorithms with better performance and shorter operation time, including preprocessing and deep learning models, will be continuously developed over time. For instance, the image preprocessing contains face detection, alignment, rotation, and resize, but when facing problems, such as backlight, shadows, and facial incompleteness, caused by complex environments, these current methods are always powerless, and these shortcomings may be solved in the future. What is more, although the CNN model in the proposed framework currently performs well, it will be replaced by models with higher learning capabilities and higher classification accuracy in the future. In order to ensure the competitiveness of the framework in a longer period, it needs to be adjusted and maintained regularly, and more advanced algorithms and technologies should be adopted to update it.

In addition, with a large number of participants in the online courses, we have no way to ensure that everyone keeps the high level of concentration, and then students' expressions may not fully represent their emotions due to these subjective factors. Taking measures like setting thresholds can filter out some invalid information and highlight the main emotions in the image. Finally, the teaching efficiency can be improved as a result.