CO3519 Assignment 2 – Facial Emotion Recognition using Advanced AI

Author : James Birkenhead | G20983016

## ***Introduction***

Facial Emotion Recognition (FER) encompasses computer vision tasks which are aimed at identifying various human emotions depicted via a face and then categorizing them into various emotional categories. The process of detecting emotions normally revolves around analysing facial landmark positions such as nose, eyebrows and mouth. Changes to these positions can be analysed and then categorised into emotions.

FER is deemed to be an important and advantageous application of Artificial Intelligence technologies, since much communication is non-verbal. The exact amount is contested, however some studies suggest up to 60-80% of communication is non-verbal, providing a significant area for potential use and deployment. FER also has numerous applications across various sectors from education, to autopilot systems, to neuroscience.

The previous paper on this topic discussed the implementation of an FER algorithm using Machine Learning via Histogram of Oriented Gradients feature extraction and Random Forest Emotion Classification. This was somewhat powerful; however, it also had some limitations.

* Performance was mixed across various emotions. Whereas the algorithm performed well on some emotions such as happiness and surprise, the algorithm struggled with classifying anger and sadness.
* Due to the simplicity of the model. underfitting was a problem. This meant that certain patterns in the data couldn’t be captured, and performance on the testing data for some emotions was comparatively poor.
* While some of the observed discrepancies could’ve potentially been due to biases in the training datasets, there was nonetheless deemed to be areas for improvement which could be built on by using a more complex algorithm.

To try and improve the performance of FER, this paper will discuss the implementation of an Advanced AI algorithm based on Convolutional Neural Networks (CNNs). These networks are specialized deep learning algorithms designed for tasks that involve object recognition. CNNs contain several key components…

* **Convolutional Layers** apply a sliding window function to the matrix of pixels. The sliding function is called a filter, and several filters can be applied.
* **Rectified Linear Unit Activation Functions** are applied after each convolution operation, aiding the network to learn non-linear relationships between image features. This makes the network stronger at identifying different patterns.
* **Pooling layers** pull the most significant features from a convoluted matrix by applying aggregation operations which reduce the dimension of the matrix. This in turn reduces the amount of memory when training the network
* **Fully connected layers** are the last layer of the CNN and their inputs correspond to the flattened one dimensional matrix generated by the last pooling layer.

An example architecture of a CNN can be found below…

A diagram of a diagram of a block diagram

AI-generated content may be incorrect.

The potential CNN architectures that will be discussed include VGG and ResNet.

## ***Literature Review***

Visual Geometry Group (VGG) is a standard deep CNN. VGG-16 is an example of such network and consists of 16 layers. This model has been found to achieve a test data accuracy of 92.77% with the ImageNet dataset. Images of size 224x224x3 are passed through various layers. The convolution layer has a 3x3 filter size while the max pooling layer has a filter size of 2x2. VGG-19 consists of three more convolutional layers than VGG-16 which should allow it to capture more complex image features.

ResNet50 is another CNN consisting of 48 convolution layers, making it a deeper network than VGG-16 and VGG-19. **Add more background!!!**

Mascarenhas and Agarwal (2021) carried out feature extraction on a dataset of 6000 images which could be classified into 5 different categories; shoes, beauty, jewellery, watches and bags. Once the features were extracted, they were fed into the various models to classify the images. All of the images needed to be compressed since they were of too great a size to be inputted into the convolutional layers. Classification was tested using each algorithm with different numbers of epochs.

A table of data with numbers

AI-generated content may be incorrect.

The study found that ResNet50 exhibited the highest accuracy followed by VGG-19 and VGG-16. They also found that epoch values above 20 induced overfitting and values below 15 induced underfitting, indicating that an optimal range of epoch values is between 15 and 20.

<https://ieeexplore.ieee.org/abstract/document/9687944>

https://www.ibm.com/think/topics/convolutional-neural-networks

## ***References***

European Data Protection Supervisor (2021). *Facial Emotion Recognition*. [online] Available at: https://www.edps.europa.eu/system/files/2021-05/21-05-26\_techdispatch-facial-emotion-recognition\_ref\_en.pdf.

Mascarenhas, S. and Agarwal, M. (2021). *A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification*. [online] IEEE Xplore. doi:https://doi.org/10.1109/CENTCON52345.2021.9687944.