Applied Deep Learning

Emotion Detection

Using the Convolutional Neural Network (CNN) Algorithm

Nagham Dahi[1]
YCH3Y3

[1] Eötvös Loránd University, Budapest, Hungary
ych3y3@inf.elte.hu

1 Introduction:

Emotion recognition focuses on recognizing human emotions based on a variety of methods, such as auditory and visual expressions, body language, physiological signals, etc.

Sentiment models can be broadly categorized into discrete models and dimension models. The first classifies feelings into separate entities, for example, anger, disgust, fear, happiness, sadness and surprise in Eckman's theory. The latter describes feelings using their basic dimensions, for example, valence, arousal, and dominance, which measure feelings from unpleasant to pleasant, passive to active, and controlled, respectively. Challenges persist in recognizing emotions under natural conditions due to high variance within a layer and low variance between layers, e.g.changes in facial position and subtle differences between expressions.

Facial emotion recognition (FER) is as important for human-computer interaction as clinical practice and behavioural description. Accurate and robust FER through computer models remains a challenge due to the heterogeneity of human faces and differences in images such as different facial poses and lighting. Among all the FER techniques, deep learning models, especially Convolutional Neural Networks (CNNs) have shown great potential due to strong automatic feature extraction and computational efficiency.

2 Related Work

Among all the FER techniques, deep learning models, especially Convolutional Neural Networks (CNNs) have shown great potential due to strong automatic feature extraction and computational efficiency. In image classification, convolutional neural networks (CNNs) have shown great potential due to their computational efficiency and feature extraction ability.

A typical CNN includes a convolutional layer, a pooling layer, and a fully connected layer. This makes it effective in processing still images. However, at that time, the application of CNNs was limited by the lack of training data and computing power.

Various techniques have been suggested to improve performance further. For example, the sigmoid activation function has been replaced by the corrected linear unit (ReLU) activation to avoid gradient scattering problems and speed up training. Various aggregation methods such as mean aggregation and max aggregation are used to reduce the input and help with generalization. Dropout, regularization, and data augmentation are used to prevent overfitting. Batch normalization was developed to help prevent gradual fading and bursting.

In this work, we achieve the highest classification accuracy for a single network in FER2013. We adopt the VGGNet architecture, fine-tune its hyperparameters, and experiment with various optimization methods. To our knowledge, our model achieves a single-evolving network accuracy of 73.28% in FER2013 without using additional training data.

3 Experiments:

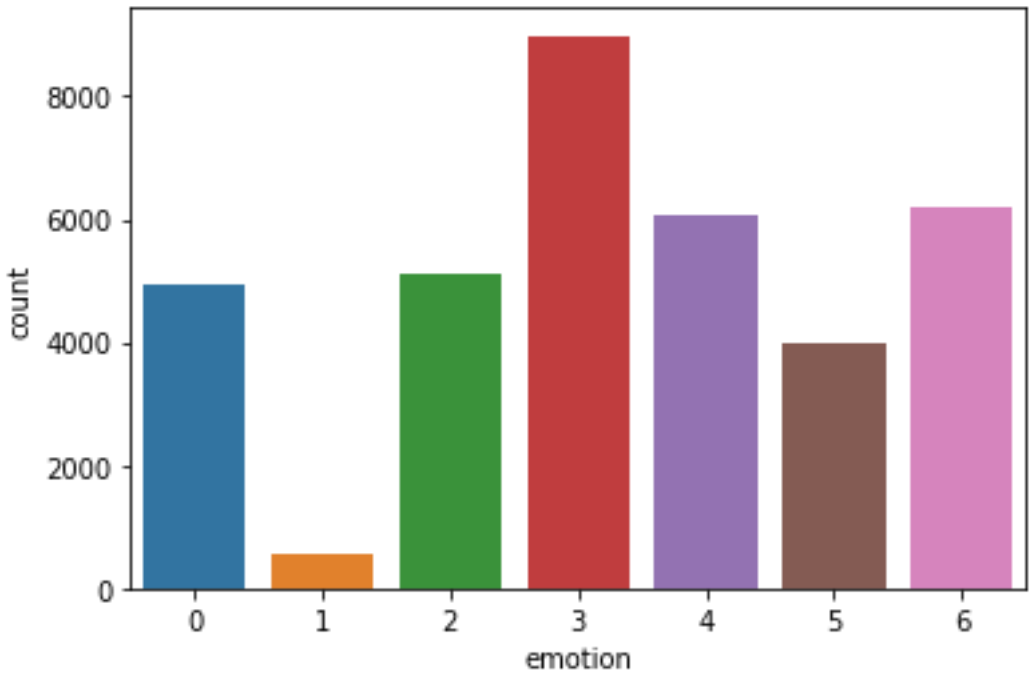
A. Dataset and processing:

Specific dataset for emotion recognition that includes challenging and challenging natural conditions is FER2013. It was presented at the International Conference on Machine Learning (ICML) in 2013 and has become a standard in comparing model performance in emotion recognition. Human performance in this data set is estimated at 65.5%. When comparing different methods and measuring our results, we are highly interested in previous work trained and evaluated in this data set.

This data is not in image format but it's in a format called data-frame. The pixels column of the data-frame contains all the pixel values. The pixel values associated with each image because each image is grey-scaled and of resolution 48x48.

	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

And there are 7 categories of emotions in this data-set (anger', 'disgust', 'fear', 'happiness', 'sadness', 'surprise' 'neutral')



If we visualize the images of each emotion category we see the data contains a wide range of images like, male, female, kids, olds, white, black etc. and It contains some non-human images like cartoons.

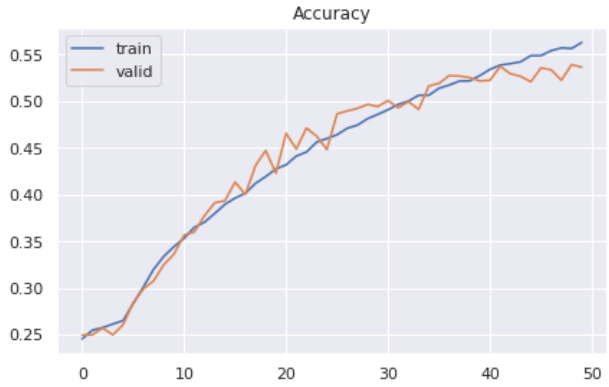
B. Creating the Model:

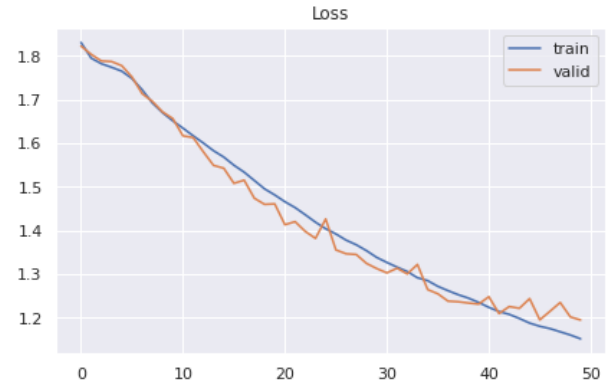
creating an Convolutional Neural Network (CNN) for this task and then feed batches of 48x48 gray-scaled images. But at first we have to pre-process our data, after split our data to train-test sets then we convert them to arrays. And we have to normalize the data to be compatible with our upcoming model.

We start building the model of CNN: we use convolution layer with 64 neural and filter size (3, 3), maxpooling ,and dropout. After all we flatten the data to have fully connected neural network. We use Dense (fully connected layer).

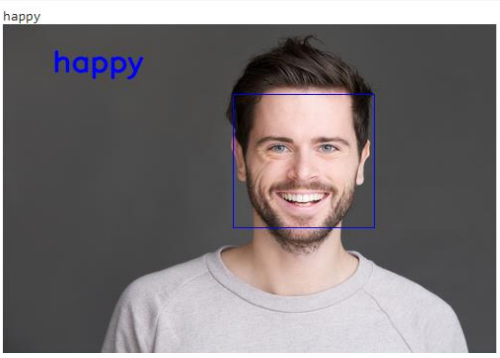
We compile our model and then train and fit the model for 50 epochs, validation loss: 1.1698 and validation accuracy: 0.5553

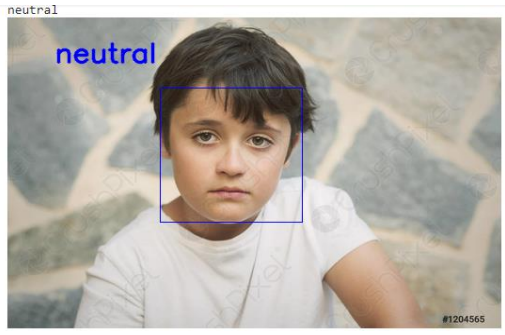
If we plot the training and validation metrics, the accuracy gradually increases and achieved 55% accuracy on both training and validation set.

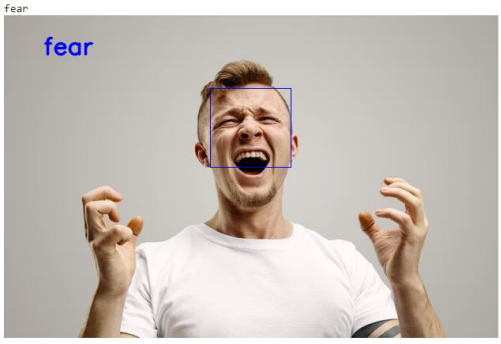




We test our model with some samples :







4 Conclusion

From the results of the conducted experiments, in general, they succeeded in designing a system with a general description of the search object using the convolutional neural network (CNN) method to predict 7 (seven) human facial expressions using the face. Emotion recognition (FER) from the FER-2013 dataset.

The system design in this study uses the process described as follows:

1) The training process uses the 2013 facial expression recognition dataset or the FER-2013 dataset and uses the Convolutional Neural Network (CNN) method as feature extraction and a suitable facial prediction process.

2) Direct recognition of facial expressions (real time), finding of facial bodies using the Haar Cascade method, and a convolutional neural network or CNN method used to classify facial expressions.

However, the test results have weaknesses in the system, so it would be better to use a new dataset or enter more training data so that when testing the data, it will produce higher accuracy, especially for tests of more than 5m area, viewing angles and when rotating images.

References

[1] Y. Khaireddin, and Z. Chen, : Facial Emotion Recognition: State of the Art Performance on FER2013, Dept. of Electrical and Computer Engineering, Boston University, Boston, MA, USA.

[2] L. Zahara, P. Musa, E. Wibowo, I. Karim, and S. Musa : The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi, Indonesia.