Emotions Recognition

1st KETUL MADHVANI
Institute of Technology
Nirma University
Ahmedabad, Gujarat, India
20BCE125@nirmauni.ac.in

2nd KINAR SHETH
Institute of Technology
Nirma University
Ahmedabad, Gujarat, India
20BCE129@nirmauni.ac.in

3rd LAV PANCHAL
Institute of Technology
Nirma University
Ahmedabad, Gujarat, India
20BCE140@nirmauni.ac.in

4th MALHAR SHINDE
Institute of Technology
Nirma University
Ahmedabad, Gujarat, India
20BCE140@nirmauni.ac.in

***Abstract*—FER (Facial emotion recognition) is one of the very prominent and well-known areas in the domains of deep learning and artificial intelligence. In this paper, we have presented our deep learning model (which makes use of transfer learning), which will classify a given input human facial expression image into one of these seven categories of facial expressions:- 1) Angry, 2) Disgusted, 3) Scared, 4) Happy, 5) Sad, 6) Surprised, and 7) Neutral. For training this deep learning model, we have made use of TensorFlow, Keras, MobileNetV2, HaarCascade models/algorithms/libraries.**

***Index Terms*—Facial Recognition, TensorFlow, Machine Learning, Deep Learning, Keras, MobileNetV2, HaarCascade**

I. INTRODUCTION

Facial Expressions provide important signifiers of human emotions. Recognizing facial expressions that communicate fundamental emotions like fear, happiness, disgust, etc. is known as facial emotion recognition. It is useful in human-computer interactions and can be used in customer satisfaction surveys, online gaming, digital advertising, and healthcare [1], [3]. FER is a member of the group of technologies known as "affective computing," a multidisciplinary area of study on the capacity of computers to recognize and understand affective states and human emotions that frequently relies on Artificial Intelligence technology. In the upcoming generations of computer vision systems, effective interaction-based computer systems may be significantly reliant on them. And the issue of inaccurate identification of images has been resolved because of developments in computer vision, which enable high emotion identification accuracy in photographs taken in predictable surroundings and under controlled circumstances [4].

Still, the challenges persist in emotion recognition due to changes in facial poses and the small differences in various expressions. These could be broadly classified into high intra-class variation and low inter-class variation. Improvements in the computer vision field aim to improve the classification accuracy of such problems. There are many methods of implantation, using different ways to classify the data. We'll discuss it in the next section.

II. FACIAL EXPRESSION RECOGNITION MODELS

The different algorithms used for facial expression recognition are discussed below [5]:

- **K-Nearest Neighbor**: Each of us uses a certain set of facial expressions to convey our feelings. For instance, when we smile, the zygomaticus major muscle raises our cheeks, causing our lips to u-shape. So, by utilizing KNN to identify the grin pattern and contrasting it with a test image, we can identify the happy feeling. The same is true for other expressions.
- **Support Vector Machines (SVM)**: SVM is another famous FER technique. In order to find the plane with the greatest margin separation, it seeks to find a hyperplane in an N-dimensional space that divides the different types of data points. The method of feature extraction determines how effectively recognition works.
- **Deep Belief Network**: This algorithm uses the greedy technique in several layers and requires faces to be ideally aligned in an image.
- **Multi-Layered Perceptron (MLP)**: MLP analyses the hidden pattern in photos for each emotion using their pixel values and employs a backpropagation technique to test different hyperparameters.
- **Convolutional Neural Network**: While working with images, the convolution technique enables us to reduce computation without sacrificing the system's accuracy. Similar to MLP, it develops a facial expression recognition system using pixel values.

III. DATASET USED FER

We chose to go with the FER 2013 dataset available on Kaggle. FER2013 is one particular emotion recognition dataset that takes into account the challenging naturalistic circumstances and difficulties. It was presented at the International Conference on Machine Learning (ICML) in 2013 and quickly established itself as a standard for evaluating the effectiveness of various emotion identification models. On this dataset, human performance is thought to be 65.5% [6]. FER2013 has a collection of over 30,000 48x48 facial RGB photos with various emotions, and its primary classifications fall into one of seven categories:0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. Each class has approximately 5,000 samples except Disgust, which has around 600 image samples.

IV. Potential Problems in the Dataset

Since nothing is a 100% perfect, this dataset has come along with its own set of problems. A few of them are:

- **Imbalance**: Facial expression images of one or more particular expressions are much more in number than those of other expressions, leading to an imbalance amongst the data samples of various classes, which can make the deep learning model biased towards that particular facial expression whose images are more.
- **Intra-class variation**: In the dataset, all of the images might not be actual human facial images, some of them might be cartoons, drawings, sketches, paintings, etc. which don't constitute ideal data samples, accounting for variations in classes.
- **Occlusion**: The covering of some portions of the face by any other body part or external thing so that the facial expression indicators (eyes, nose, mouth, cheeks, etc.) are obstructed/occluded. Due to this, the deep learning model can't work properly on such images.
- **Contrast variation**: The light contrast in the image background or in the face itself (certain portions of the image being too light or too dark) leads to this problem.
- **Eyeglasses**: Eyes are one of the most fundamental and major (and accurate) indicators of human expressions. If the eyes of the images in the dataset are covered with eyeglasses, it becomes quite difficult for the deep learning model to classify the images accurately.
- **Outliers**: Ambiguous images which are not images of actual human facial expressions at all, but rather those of some random thing or cropped things, etc.

V. Our Approach

We have used the Deep Learning (DL) approach to train the dataset. In training this DL model, a transfer learning technique has been used. The advantage provided by this technique is, in case there arises the need of adding one more human facial expression to the exisiting classes, i.e., broadening the model to classify the dataset images into one more new class (a human expression), instead of coding and training the model from scratch and spending hours again to train the model, only the particular weights associated majorly with the newly introduced class (a human facial expression) can be modified and the model can be trained and made ready in a much shorter time compared to the former scenario. Hence, using transfer learning techniques, the deep learning model also becomes quite robust and suitable to changing requirements. Note that this approach can only be used for a new classification or regression problem which is, fundamentally, quite similar in nature to the earlier problem.

Using MobileNetV2, which classifies 224x224 images into 1000 classes as the base architecture, through transfer learning, a model has been made which classifies the dataset images into 7 classes, i.e, these 7 human facial expressions:-

- 0: Anger
- 1: Disgust
- 2: Fear
- 3: Happiness
- 4: Sadness
- 5: Surprise
- 6: Neutrality

VI. Implementation of the model

We have used the Tensorflow library to create this deep learning model, CV2 library to scan the dataset images, Matplotlib and Numpy libraries for data visualization and handling storage objects, respectively. We used the 'for loop' to assign the path to the folder on the device where images have been stored, all the FER 2013 dataset images have been scanned and made ready for use. Then, all the images have been resized from 48x48 format to 224x224 format. This is done because standard image classifiers work in this 224x224 image format And the FER2013 dataset has images in the 48x48 format. Hence, to use transfer learning, these FER 2013 dataset images have to be resized. Then we created two arrays for storing images and their labels (classification. i.e., the label to state which of the 7 facial expression categories that particular image falls under). The image array data samples have been normalized by simply dividing the array's numerical value by 255. We imported the Keras library to train our deep learning model. A pre-trained deep learning model, MobileNetV2 has been used to obtain weights for transfer learning. 'Relu' function, and 'softmax' (in the last, fully connected layer/ the classification layer) function have been used as activation functions. Since in the last layer of MobileNetV2 model classifies images into 1000 classes, we removed that layer to add some new layers and classify the images into 7 classes. We then used the 'Adam' optimization algorithm for stochastic gradient descent, to train this deep learning model. For testing this deep learning model on any random facial expression we downloaded the images from the internet. But we will have to first crop the image to include only the relevant portions of the image which shows the facial expression, and removes the excessive blank space from the image background. This is necessary to provide the model a sufficiently suitable image, which has more facial expression portions than other elements. For doing this (cropping the image to only majorly include the facial expression), we have made use of the 'haarcascade frontal face', a pre-trained deep learning model, which is an algorithm for face detection. But the 'haarcascade frontal face' algorithm only works on gray images, and not on RGB (Red Green Blue) images. Hence, we will have to convert the image, which is to be cropped for including only the facial expression, into a gray format image. Then, finally, the model has been trained for 5 epochs, which provides a fairly sufficient classification accuracy for the input images. In this way, the deep learning model, using transfer learning from various pre-trained models and algorithms, has been trained and is now ready for use. Providing input images to the model, it will classify them into one of the seven classes of human facial expressions.

Conclusion

The deep learning model has been made and trained following the methodology described above. For approximately five epochs, the model took around 5-6 hours to get trained, giving an accuracy of about 65%. Naturally, the more the number of epochs, the greater the model's accuracy of correctly classifying the input images into the corresponding facial expression. By a rough estimate, an accuracy of about 75% can be achieved after ten epochs, 90% after 20 epochs, and 95% after 25 epochs, which is an excellent accuracy value for a FER (facial emotion recognition) model.

Acknowledgment

We are expressing our sincere gratitude to Ms. Rupal Kapdi who has been a powerful source of inspiration and motivation for all of us to do this project and explore Machine Learning.

References

[1] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2003, vol. 5, doi: 10.1109/CVPRW.2003.10057.

[2] F. Abdat, C. Maaoui, and A. Pruski, "Human-computer interaction using emotion recognition from facial expression," in Proceedings - UKSim 5th European Modelling Symposium on Computer Modelling and Simulation, EMS 2011, 2011, doi: 10.1109/EMS.2011.20.

[3] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," Pattern Recognition, vol. 36, no. 1. 2003, doi: 10.1016/S0031-3203(02)00052-3.

[4] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6. 2015, doi: 10.1109/TPAMI.2014.2366127.

[5] "Facial Emotion Recognition Project using CNN with Source Code," ProjectPro. [Online]. Available: https://www.projectpro.io/article/facial-emotion-recognition-project-using-cnn-with-source-code/570. [Accessed: Nov. 26, 2022]

[6] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, 2015, doi: 10.1016/j.neunet.2014.09.005.