Model Development Document (MDD)

Submitted BY: VIVEK GUPTA

Emotion Detection Using a CNN Model for Prediction

1. Introduction

This document outlines the development of a Convolutional Neural Network (CNN) model for real-time emotion detection based on facial expressions. Emotion recognition has various applications, including human-computer interaction, mental health assistance, education, and surveillance. This project aims to leverage deep learning techniques to accurately classify facial expressions into predefined emotional categories, enhancing the understanding and response to human emotions in various scenarios.

This study focuses on emotion prediction based on facial expressions, utilizing an image-based dataset that captures a range of facial expressions. A CNN model is proposed to identify emotional states based on these expressions. By leveraging the dataset, the model can learn to recognize patterns and accurately predict emotions.

2. Problem Definition

The objective is to build a robust and accurate CNN model capable of classifying human facial expressions into a set of discrete emotions. This model will take images of faces as input and predict the corresponding emotion from a pre-defined set, such as anger, disgust, fear, happiness, sadness, surprise, and neutral. The challenge lies in developing a system that can accurately interpret facial cues under varying conditions, such as lighting, pose, and individual variations in expression, according to ScienceDirect.com.

3. Data collection and preprocessing

The success of a deep learning model relies heavily on the quality and quantity of the training data.

- Dataset: The FER-2013 (Facial Expression Recognition) dataset will be utilized. It contains 35,887 grayscale images of human faces, each representing one of seven emotions: angry, disgust, fear, happy, neutral, sad, and surprise.
- Preprocessing: The data will undergo the following preprocessing steps:
    - Grayscale Conversion: If necessary, images will be converted to grayscale to reduce dimensionality and ensure consistency.
    - Face Detection: Haar cascade classifiers from OpenCV will be used to detect faces within the images.

- Cropping and Resizing: Detected faces will be cropped and resized to a consistent resolution (e.g., 48x48 pixels) suitable for the CNN model.
- Normalization: Pixel values will be normalized (scaled between 0 and 1) to improve training stability.
- One-Hot Encoding: Emotion labels will be converted to a one-hot encoded format, a requirement for multi-class classification.
- Data Augmentation: Techniques like random rotation, shifts, and flips will be applied to the training data to increase its diversity and improve model generalization, preventing overfitting.

- Data Splitting: The dataset will be split into training, validation, and test sets (e.g., 80% training, 10% validation, 10% test) to evaluate the model's performance during development and assess its generalization ability.

4. CNN model architecture

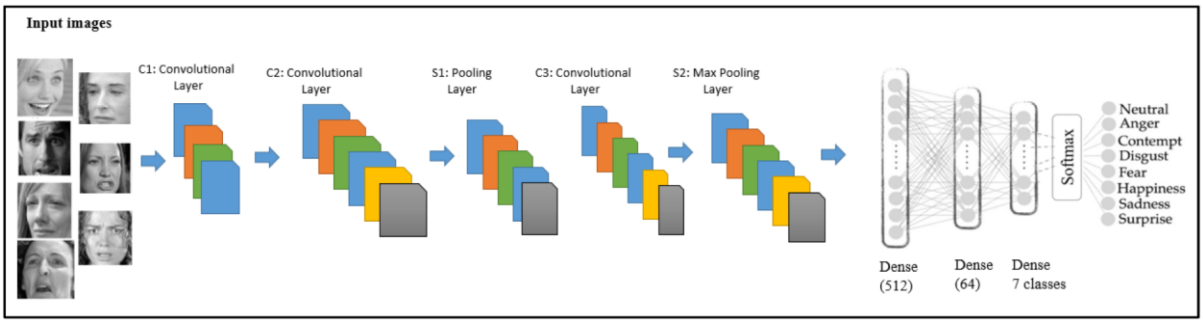

FIG 1: CNN MODEL FOR ED

The CNN architecture will be designed to effectively extract features from facial images and classify them into the seven emotion categories. A typical CNN architecture for this task would include:

- Convolutional Layers: Multiple convolutional layers will be used to learn hierarchical features from the input images, employing different filter sizes (e.g., 3x3, 5x5) and activation functions (e.g., ReLU).
- Max Pooling Layers: Max-pooling layers will downsample the feature maps, reducing the spatial dimensions and computational complexity.
- Batch Normalization: Batch normalization layers will be incorporated to stabilize the training process and improve network performance.
- Dropout Layers: Dropout layers will be used to prevent overfitting by randomly dropping out neurons during training.
- Flatten Layer: The output from the convolutional layers will be flattened into a single-dimensional vector before being fed into fully connected layers.

- Dense Layers: Fully connected layers will be used for classification, with the final layer having seven output nodes (representing the emotion classes) and a Softmax activation function.
- Optimizer and Loss Function: The model will be compiled with an optimizer like Adam and use categorical cross-entropy as the loss function, appropriate for multi-class classification.
- Callbacks: Early stopping and model checkpointing will be used during training to monitor performance and save the best model weights based on validation accuracy.

5. Model training and evaluation

The model will be trained using the preprocessed data, and its performance will be evaluated using various metrics:

- Training: The model will be trained using the training dataset and validated on the validation set for a predefined number of epochs, with a specified batch size.
- Evaluation Metrics:
    - Accuracy: Overall accuracy of the model on the test dataset.
    - Precision, Recall, and F1-score: These metrics will be calculated for each emotion class to assess the model's ability to correctly classify each emotion and handle class imbalances.
    - Confusion Matrix: A confusion matrix will be used to visualize the model's performance in terms of true positives, true negatives, false positives, and false negatives for each emotion class.

An expression transfer method from humans to many stylised characters is proposed as Deep Expr (DEEP). Two Convolutional Neural Networks (CNN) are trained initially to recognise the expressions of both people and cartoon characters. The mapping from humans to characters is then learned using a transfer learning approach, resulting in a shared embedding feature space [7]. Using this embedding, it is possible to get images based on facial expressions and on the expressions of fictional characters. Character expressions based on people can be found using our perceptual model. We put our approach to the test using a variety of retrieval tasks and a stylized character dataset [4]. A facial expression expert and a series of Mechanical Turk trials have also shown that the suggested characteristics' projected ranking order is strongly associated with the actual ranking order.

In this study, a typical neural network with data augmentation is used to recognise face expressions. This method can categorise images into Anger, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Due to their huge number of filters, CNNs are superior for image identification tasks.

SYSTEM DIAGRAM:

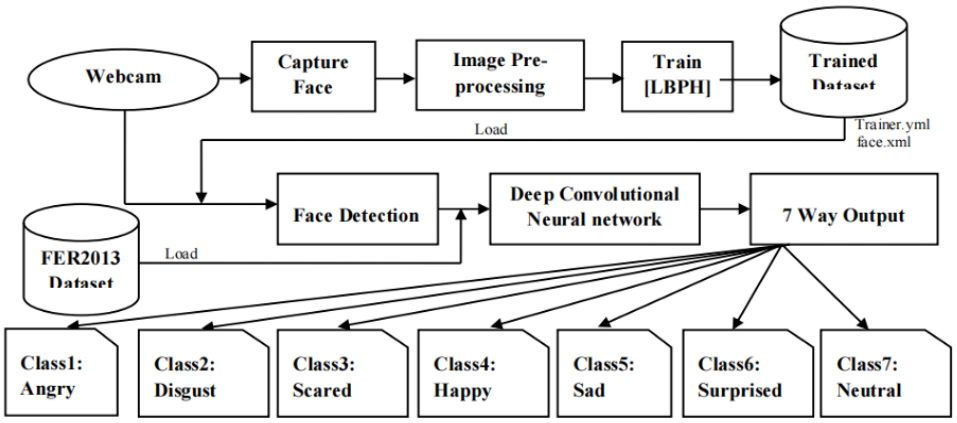

FIG 2. System diagram of facial emotion detection

A webcam is used to capture, identify, and recognise the facial expressions of a person, which is done through the use of software. In the camera, a rectangular frame on the face area is obtained; this identification of the face region from a non-facial region is accomplished by the employment of the Viola Jones method, the LBPH Face Recognizer algorithm, and the Haarcascade frontal face dataset, among other techniques. Captured person faces are preprocessed before being saved in a folder labelled with the subject's ID and name. These photos are trained using the LBPH method, and the resulting trained dataset is saved as Trainer.yml in the Trainer folder. During the Face Detection process: A trained dataset is used to match the face in a video camera with the face in the dataset. If a person's face matches that in the trained dataset, his or her ID and name will be displayed on the screen. In order to classify the obtained face, convolutional neural networks are used in conjunction with the FER2013 database to do the classification [11]. The facial expression represents the chance of acquiring the maximum level of expression based on the characteristics of the individual. One of seven possible facial expressions is presented in conjunction with the recognised picture of the subject.

6. Deployment considerations

- Real-time Inference: The trained model will be capable of real-time emotion detection, accepting webcam feed or image inputs and predicting the emotion instantly.
- OpenCV Integration: OpenCV will be used to capture video frames, detect faces, and draw bounding boxes and emotion labels on the video feed.
- Saving the Model: The model architecture will be saved in JSON format and the trained weights in HDF5 (.h5) format for deployment and future use.
- Ethical Considerations: The ethical implications of emotion recognition, particularly regarding privacy, bias, transparency, and potential misuse, will be acknowledged and addressed throughout the development and deployment process.
    - Informed Consent: Obtaining informed consent for data collection and usage, especially in scenarios involving sensitive emotional data, is crucial.
    - Bias Mitigation: Efforts will be made to address potential biases in the dataset and model to ensure fair and accurate emotion recognition across diverse demographic groups.
    - Transparency and Explainability: The model's decision-making process will be made as transparent as possible, potentially using explainable AI (XAI) techniques, to foster trust and accountability.

System flowchart:

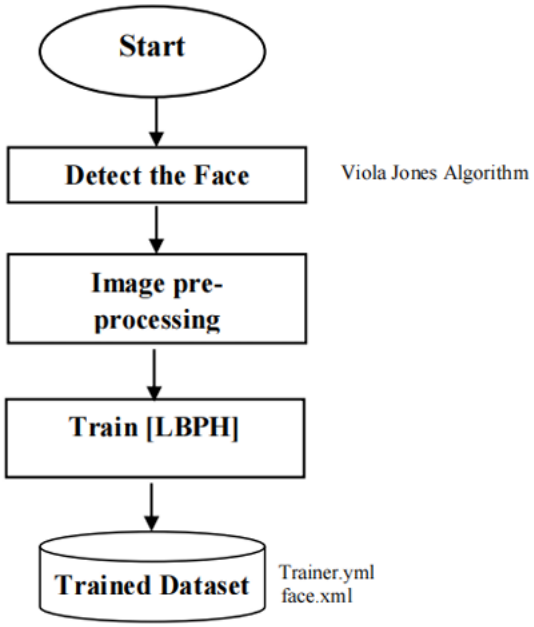

Fig 3.flowchart of training

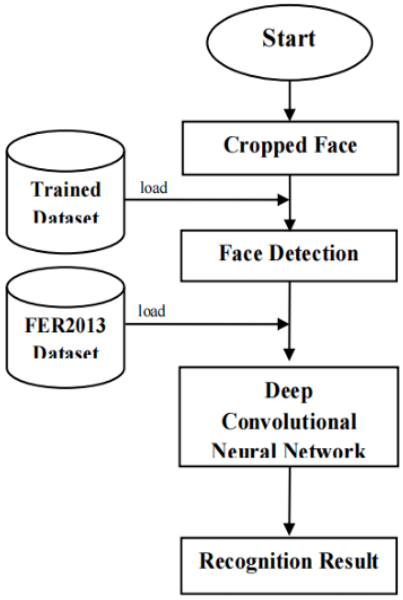

Fig 4. Flowchart of testing

During training Phase, the system received a training data comprising grayscale images of faces with their respective expression label and learns a set of weights for the network. The training step took as input an image with a face. Thereafter, an intensity normalization is applied to the image.

The normalized images are used to train the Convolutional Network. To ensure that the training performance is not affected by the order of presentation of the examples, validation dataset is used to choose the final best set of weights out of a set of trainings performed with samples presented in different orders.

The output of the training step is a set of weights that achieve the best result with the training data. During test, the system received a grayscale image of a face from test dataset, and output the predicted expression by using the final network weights learned during training. Its output is a single number that represents one of the seven basic expression.

DATASETS USED: Several public databases were used in order to assess face expression recognition algorithms: Frontal face dataset from Haarcascade: HaarCascade Classifier is used to recognise faces in pictures utilising characteristics.

The frontal face is detected using the haarcascadefrontalface default.xml [17]. It was created by Viola and Jones in response to a proposal made in 1998[2] by Papa Georgiou et al. To verify that the retrieved faces are all in the same location, we utilised an additional classifier named 'haarcascade eye.xml' from the same OpenCV library[18]. This identifies the region around the eyes and then adjusts the left and right borders of the face window to maintain an equal distance between the eyes and the sides of the face.

Thus, superfluous information (such as hair, ears, and background) is removed, and the retrieved faces have their locations adjusted.FER2013 dataset: FER2013[15] is an open-source dataset generated by Pierre-Luc Carrier and Aaron Courville for an ongoing project and later given publicly for a Kaggle competition[15].

The FER2013 database was launched during the 2013 International Conference on Machine Learning's Challenges in Representation Learning. FER2013 is a massive and unrestricted database that was automatically compiled using the Google image search API. After rejecting incorrectly labelled frames and modifying the cropped region, all photos have been registered and resized to 48*48 pixels. This dataset contains 35,887 grayscale, 48x48-pixel pictures of faces displaying a range of emotions -7 emotions, all labeled-. Emotion labels in the dataset: 0: -4593 images- Angry 1: -547 images- Disgust 2: -5121 images- Fear 3: -8989 images- Happy 4: -6077 images- Sad 5: -4002 images- Surprise 6: -6198 images- Neutral The FER-2013 dataset was created by gathering the results of a Google image search of each emotion and synonyms of the emotions. The images in FER-2013[15]consist of both posed and un-posed headshots.

Fig 5: Example Images of FER2013 dataset

Fig 5: Example Images of FER2013 dataset Figureillustrating variability in illumination, age, pose, expression intensity, and occlusions that occur under realistic conditions. Images in the same column depict identical expressions, namely anger, disgust, fear, happiness, sadness, surprise, as well asneutral. The data file contains 3 columns — Class, Image data, and Usage. a) Emotion class: is a digit between 0 to 6 and represents the emotion depicted in the corresponding picture. Each emotion is mapped to an integer as shown below. 0- 'Angry' 1-'Disgust' 2-'Fear' 3-'Happy' 4-'Sad' 5-'Surprise' 6-'Neutral' b) Image data: is a string of 2,304 numbers and these are the pixel

intensity values of our image, we will cover this in detail in a while. c) Usage: It denotes whether the corresponding data should be used to train the network or test it. 4.

7. Results:

The training and testing datasets are from a Kaggle Facial Expression Recognition Challenge (FER2013). It comprises of precropped grayscale photos of faces classified as pleased, sad, disgusted, angry, surprised, fearful, or neutral. The webcam image will be used as the input for processing the output. The output labels human facial expressions as pleased, sad, disgusted, angry, surprised, fearful, or neutral. Neural Evolutionary Network Convolutional Operation is the first step. Our strategy of attack begins with a convolution operation. This stage will discuss feature detectors, which act as filters for the neural network. Additionally, we will explore feature maps, including how to learn their parameters, how patterns are recognised, the layers of detection, and how the findings are shown.

8. Future work

Further enhancements could include exploring more advanced CNN architectures, incorporating other modalities like audio or physiological signals for multimodal emotion recognition, developing personalized models to account for individual differences in expression, and rigorously addressing the ethical and societal implications of emotion detection technology.

References:

[1]Chu, William Wei-Jen Tsai, Hui-Chuan, YuhMin Chen and Min-Ju Liao. Facial expression recognition with transition detection for students with high-functioning autism in adaptive e-learning.” Soft Computing: 1-27, 2017.

[2] Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Panagiotis, Bjrn W. Schuller, and Stefanos Zafeiriou. ”End-to-end multi-modal emotion recousing neural networks.” IEEE Journal of Topics in Signal Processing 11, no. 8: 13011309, 2017

[3] Gampala, V., Kumar, M.S., Sushama, C. and Raj, E.F.I., 2020. Deep learning based image processing approaches for image deblurring. Materials Today: Proceedings.

[4] Aneja, Gary Faigin, Deepali, Alex Colburn, Barbara Mones, and Linda Shapiro. ”Modeling stylized character expressions via deep learning.” In Asian Conference on Computer Vision, pp. 136-153. Springer, 2016.

[5] Natarajan, V.A., Kumar, M.S., Patan, R., Kallam, S. and Mohamed, M.Y.N., 2020, September. Segmentation of Nuclei in Histopathology images using Fully Convolutional Deep Neural Architecture. In 2020 International Conference on Computing and Information Technology (ICCIT 1441) (pp. 1-7). IEEE.

[6] Mohammed, M. A., Abdulkareem, K. H., Mostafa, S. A., Khanapi Abd Ghani, M., Maashi, M. S., Garcia Zapirain, B., ... & Al-Dhief, F. T. (2020). Voice pathology detection and classification using convolutional neural network model. Applied Sciences, 10(11), 3723.

[7] Korla, S.; Chilukuri, S. T-Move: A Light-Weight Protocol for Improved QoS in Content-Centric Networks with Producer Mobility. Future Internet 2019, 11, 28. https://doi.org/10.3390/fi11020028.

[8] Kumar, P. M., Gandhi, U., Varatharajan, R., Manogaran, G., Jidhesh, R., & Vadivel, T. (2019). Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things. Cluster Computing, 22(4), 7733-7744.

[9] Liang, L., Lang, C., Li, Y., Feng, S., & Zhao, J. (2020). Fine-grained facial expression recognition in the wild. IEEE Transactions on Information Forensics and Security, 16, 482-494.

[10] IEEE Style Citation: Korla Swaroopa, Sireesha Rodda, Shanti Chilukuri, "Differentiated Caching for Improved QoS in Vehicular Content-centric Networks," International Journal of Computer Sciences and Engineering, Vol.6, Issue.10, pp.317-322, 2018. [11] Shao, J., & Qian, Y. (2019). Three convolutional neural network models for facial expression recognition in the wild. Neurocomputing, 355, 82-92.