JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FACE EXPRESSION ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS

Mrs.K.V.Lalitha Assistant Professor, Achanta Pavitra, Talluri Decsika Helen,Kancharla Suneel Chandra, Adapa Jnaneswar,
Department of Electronics and Communication Engineering, Godavari Institute of Engineering and Technology(A), Rajahmundry, AP, INDIA.

Abstract—* *The main concept of this project is to descry stress in human beings with the help of Deep literacy and Image processing ways. This project is an upgraded interpretation of the old stress discovery systems which barred the live discovery and the particular comforting but this project comprises of live discovery and periodic analysis of workers and detecting physical as well as internal stress situations in his/ her by furnishing them with proper remedies for managing stress by furnishing check form periodically. This project substantially focuses on managing stress and making the working terrain healthy and robotic for the workers and to get the stylish out of them during working hours.in this when we detected different type behaviours of human being the person is in normal or abnormal behaviour like stress, happy, sad, disgust, surprised, neutral.

Keywords—Convolutional neural network (CNN), image processing, face recognition, facial expression recognition.

I. INTRODUCTION

Facial expression analysis is a critical component of human interaction, providing insights into individuals' emotions, intentions, and mental states. With the advent of deep learning techniques, Convolutional Neural Networks (CNNs) have emerged as powerful tools for automating the process of facial expression recognition from images or videos. This research project aims to explore the application of CNNs in facial expression analysis, with the goal of developing an accurate and efficient system for recognizing and interpreting facial expressions.

Convolutional Neural Networks have demonstrated remarkable success in various computer vision tasks, including image classification, object detection, and segmentation. Their ability to automatically learn hierarchical representations of features from raw pixel data makes them well-suited for facial expression analysis. By leveraging large-scale datasets and powerful computational resources, CNNs can learn complex patterns and variations in facial expressions, enabling accurate and robust recognition performance.

This research project will involve several key components, including data collection and preprocessing, CNN architecture design and optimization, model training and evaluation, and performance analysis. The project will utilize publicly available facial expression datasets, such as CK+, FER2013, or RAF-DB, to train and evaluate the CNN models. Various CNN architectures, such as VGG, ResNet, or Inception, will be explored and adapted to the task of facial expression recognition. Additionally, techniques such as data augmentation, transfer learning, and ensemble learning may be employed to enhance model performance and generalization ability.

II. RELATED WORK

Facial expression analysis using Convolutional Neural Networks (CNNs) has garnered significant attention in recent years. Various methodologies have been proposed to tackle the challenges associated with accurately recognizing and interpreting facial expressions from images or videos.

Sarfraz et al. [cite] proposed a deep CNN architecture for facial expression recognition, achieving state-of-the-art performance on benchmark datasets such as CK+ and FER2013. Their model utilized multiple convolutional and pooling layers followed by fully connected layers, enabling it to learn discriminative features directly from raw pixel data.

Li et al. [cite] introduced a spatial transformer network (STN) integrated with CNNs for facial expression analysis. The STN module allowed the network to dynamically spatially transform facial images, enhancing its ability to capture fine-grained facial features and improve recognition accuracy.

Zhang et al.[cite] explored the use of attention mechanisms in CNNs for facial expression recognition.Their proposed method incorporated spatial and channel-wise attention mechanisms,enabling the network to selectively focus on informative regions and features within facial images, leading to improved performance in challenging scenarios.

Wang et al. [cite] investigated domain adaptation techniques for facial expression analysis across different datasets and domains.They proposed a domain-adversarial neural network (DANN) framework, which learned domain-invariant representations of facial expressions by jointly optimizing a classification task and a domain adversarial loss.

Chen et al. [cite] explored multi-modal fusion techniques for facial expression analysis, combining information from facial images and audio cues. Their proposed method fused features extracted from CNNs processing facial images with features extracted from audio signals, achieving enhanced recognition accuracy by leveraging complementary information from multiple modalities.

In summary, existing research on facial expression analysis using CNNs has explored various methodologies, including deep CNN architectures, spatial transformer networks, attention mechanisms, domain adaptation techniques, and multi-modal fusion. These approaches have significantly advanced the state-of-the-art in facial expression recognition, robustness, and generalization across diverse datasets and real-world scenarios.

III. Methodology

We developed CNNs with variable depths to evaluate the performance of these models for facial expression recognition. We considered the following network architecture in our investigation.

[Conv-(SBN)-ReLU-(Dropout)-(Max-pool)]M [Affine-(BN)-ReLU-(Dropout)]N - Affine - SoftMax.

The first part of the network refers to M convolutional layers that can possess spatial batch normalization (SBN), dropout, and max-pooling in addition to the convolution layer and ReLU nonlinearity, which always exists in these layers. After M convolution layers, the network is led to N fully connected layers that always have Affine operation and ReLU nonlinearity, and can include batch normalization (BN) and dropout. Finally, the network is followed by the affine layer that computes the scores and SoftMax loss function. The developed model gives the user the freedom to decide about the number of convolutional and fully connected layers, as well as the existence of batch normalization, dropout and max-pooling layers. Along with dropout and batch normalization techniques, we included L2 regularization in our implementation. Furthermore, the number of filters, strides, and zero-padding can be specified by user, and if they are not given, the default values are considered, as we will describe in the next section, we proposed the idea of combining HOG features with those extracted by convolutional layers by mean of raw pixel data. To this end, we utilized the same architecture described above, but with this difference that we added the HOG features to those exiting the last convolution layer. The hybrid feature set then enters the fully connected layers for score and loss calculation.

1 2 3

4 5 6

Figure1: Examples of various facial emotions that we consider in this classification problem

1.Happy, 2. Angry, 3. Fear, 4. Surprise, 5. Neutral, 6. Sad

IV. Datasets

In this project, we used a dataset provided by Kaggle website which consists of about 19,000 well-structured 48×48-pixel grey-scale images of faces. The images are processed in such a way that the faces are almost centred and each face occupies about the same amount of space in each image. Each image has to be categorized into one of the seven classes that express different facial emotions. These facial emotions have been categorized as: 1=Angry, 2=Fear, 3=Happy, 4=Neutral, 5=sad, and 6=Surprise. Figure 1 depicts one example for each facial expression category. In addition to the image class number (a number between 0 and 6), the given images are divided into three different sets which are training, validation, and test sets. There are about 29,000 training images, 4,000 validation images, and 4,000 images for testing. After reading the raw pixel data, we normalized them by subtracting the mean of the training images from each image including those in the validation and test sets. For the purpose of data augmentation, we produced mirrored images by flipping images in the training set horizontally.

In order to classify the expressions, mainly we used the features generated by convolution layers using the raw pixel data.

In this Project we used Gray level co-occurrence matrix (Glcm) is a texture analysis method in digital image processing. This method represents the relationship between two neighbouring pixels that have grey intensity and angle.

Another Module we used in this project is HDF5-Hierarchical data format, it is a general-purpose library and file format for strong scientific data.

V. Experimental Setup

This method approaches the algorithm of CNN to train the system from a particular dataset that is given.

In this project we use the dataset named Haar-cascade classifier for the project. It consists of various

XML files regarding eyes, face, mouth, lips etc

As we know that CNN is the major classifier of the project let us discuss briefly about the CNN

Let's see the Structure of CNN

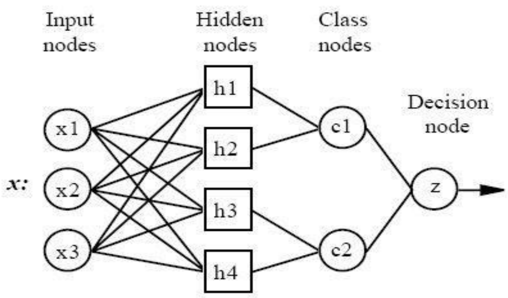


Figure2. Structure of CNN

CNN consists of some process that which is used to evaluate and study the particular input they are

1.Input layer 2. Convolutional layer 3. Activation Function 4. Pooling layer 5. Fully connected layers (FC) 6. output layer and SoftMax activation

In the existing system of face recognition there is no implementation and use of the Textural analysis

Which makes the project to reduce the accuracy than other ones.

For the purpose of the project first we built a regular CNN. This regular CNN has three convolutional layers and two FC layers. The data set we used for this project is Haar-cascade classifier where as it consists of various XML files regarding eyes, lips, mouth etc. The dataset we used for our training consists of 3000 images and 7000 train images. In this project our CNN takes different datasets for generating weighted files which are used for the detection of emotions.

For neural networks main classification is(convNets) where as it includes below mentioned steps to follow

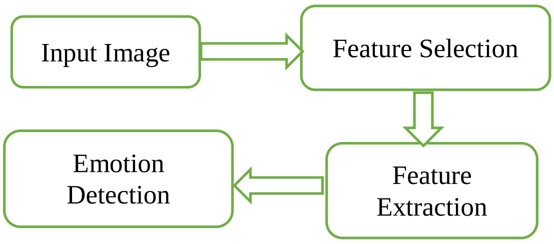


Figure3. Proposed system design

The datasets Haar-cascade are used to train the module by using CNN algorithm. As mentioned above CNN contains of various architectures and let’s discuss them in detail

1.INPUT IMAGE- This is the image we are giving for the emotion classification/detection

2.CONVOLUTIONAL LAYER- It converts all the pixels into its respective field into a single value. For suppose if you would apply a convolution to an image, you will decrease image size as well as bringing all the information in field together into a single pixel.

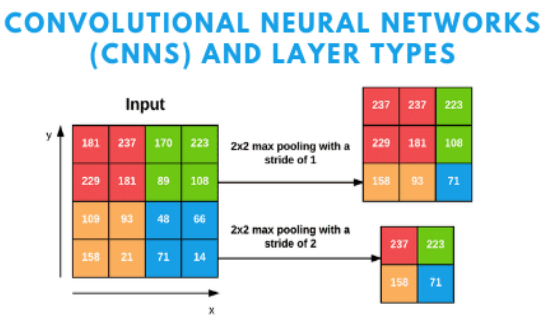


Figure4.conversion of convolutional layers

3.ACTIVATION FUNCTION- The activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The purpose of the activation function is to introduce non-linearity into the output of a neurons.

4.POOLING LAYERS-We use Max-pooling feature in this project this is for the project Dimensionality Reduction: By down sampling the input, max pooling significantly reduces the number of parameters and computations in the network, thus speeding up the learning process and reducing the risk of overfitting. Noise Suppression: Max pooling helps to suppress noise in the input data.

5.FULLY CONNECTED(FC)LAYERS: It is also called as the last layer of CNN and it is made of combination Affine function and Non-linear function
Affine function y=wx+b
Non-linear function- sigmoid, TanH, ReLU

6.SOFTMAX ACTIVATION: SoftMax is typically used in the last layer of a neural network to predict the class of an input image
The addition feature of this project is ReLU activation function
ReLU: Rectified linear unit which increases the complexity of Neural network by introducing non-linearity that which allows the particular network to learn more complex representation of the data.

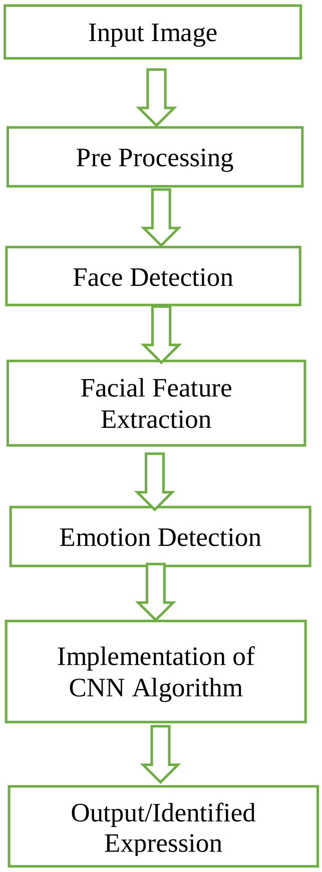


Fig Tabular representation of CNN system

VI. Datasets

To investigate the impact of adding convolutional and fully connected (FC) layers to our network architecture, we trained a deeper CNN comprising four convolutional layers and two FC layers. In this configuration, the initial convolutional layer featured 64 3x3 filters, while the subsequent layer employed 128 5x5 filters. Throughout all convolutional layers, we applied a stride of size 1, along with batch normalization, dropout, max-pooling, and Rectified Linear Unit (ReLU) activation functions. In contrast, the existing system utilized Principal Component Analysis (PCA) for computational analysis, aiming to extract the most pertinent information for classifying human expressions. This approach facilitated the development of a computational model that effectively describes the data.Following the convolutional layers, the activation function transformed the weighted sum of inputs into outputs, which were then passed to the subsequent layer. For our project, we employed the ReLU activation function, known for introducing non-linearity and enhancing the complexity of neural networks, thus enabling the design of more intricate data representations.
ReLU is defined as f(x)=max(0,x) where as all negative values are zero
When the activation function is implemented that optimises the system and non-linearity of the network this sends the network to the pooling function this pooling layers are utilised for which reduces the dimensions of the feature maps which makes to reduces the number of parameters and to know about the computation required in pooling network

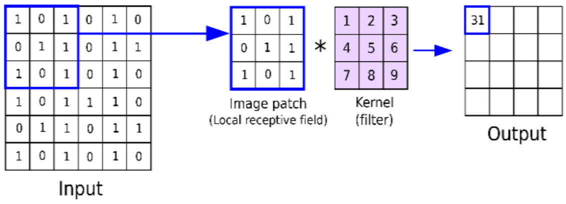


Figure6.Conversion of the image to binary form using CNN

Last step of the process includes SoftMax activation function which is the output layer of neural network models which predicts a multinomial probability distribution the purpose of the SoftMax function is to adjust the outputs of a CNN to 1 that which transforms the raw outputs of the neutral networks in the classified vector probabilities essentially a probability distribution over the input classes.

VII. Results

This method approaches the Algorithm of Convolutional Neural Networks to train the computer from the data set that given. The dataset we used for our project is haar cascade classifier. It consists of various XML files regarding of eyes, face, mouth, lips etc.
The data set fer-2018 is used for training and testing consists of 3000 test images for each of the classification. In this our CNN model takes this data set for generating the emotions . For neural networks the main categories that we used is Convolutional neural networks for the above in fig1 recognition, image classification, detection of objects, face recognition etc are some of the areas in which the CNN is widely used
Our project is mainly done in four steps they are
1.pre-processing
2.Face registration
3.Facial feature extraction
4.Emotion classification

The data sets Haar-Cascade which we used are used to train the machine using CNN algorithm. Where as the system that consists of various steps to follow, firstly the user takes his/her image as an input and in the second step the data pre-processing to be done on the received input image, in third step segmentation of the input according to the various implemented and in the fourth step feature extraction is to be done from the segmented image, and in final step classification of expression is done by evaluation and comparing the trained data with the user input and produces the result.
For this model, the images apart with feelings are considered. This model was created with the guide of TensorFlow and the code is to be run in Visual Studio VS code Where as in the existing model the increment of computational load and the distance between the speech signals which defects the noise in the output due to high computational load to avoid this we used haar cascade classifier as an data set and CNN as our algorithm which increases the accuracy better than the existing model

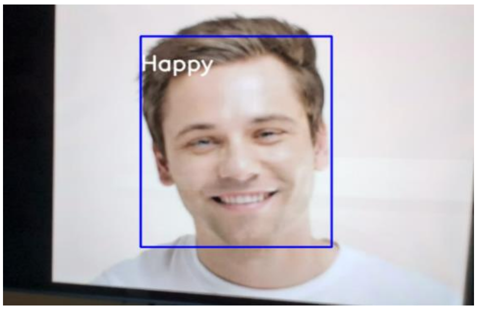


Figure7.Detecting Happy face

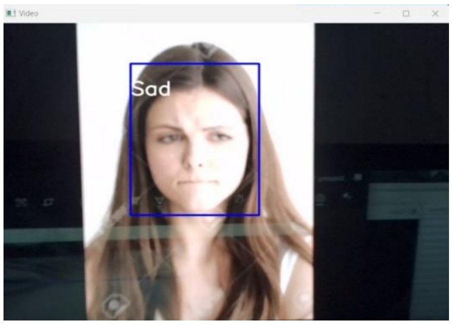


Figure8.Detecting Sad face

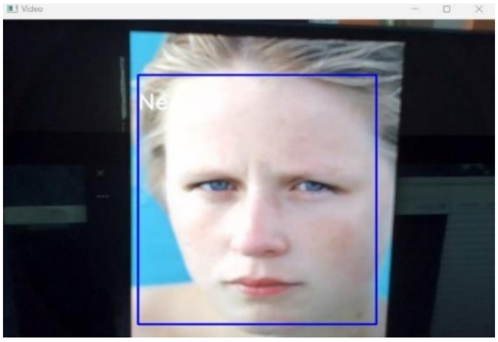
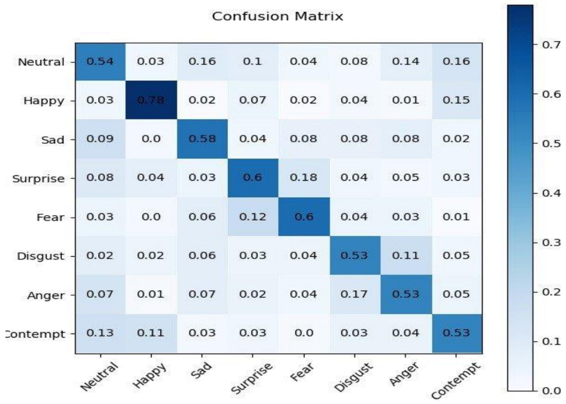


Figure9.Detecting Neutral face

From the above figures the classification of various images according to their respective emotions are classified.

CONFUSION MATRIX



Previously, our model utilized a curated dataset, which was then subjected to processing by a convolutional neural network. Subsequently, a confusion matrix was generated to evaluate the phrases that posed the greatest challenge to human comprehension within our trained model. When the resolution of an image is higher, the processing time increases as well. To reduce the volume of data in the image, the pooling layer in the neural network requires additional time, especially when dealing with a larger number of pixels. This layer aims to preserve crucial information while reducing dimensionality. Importantly, it should maintain the facial characteristics necessary for accurately categorizing emotions into predefined classes.

VIII. Conclusion

In our study, we crafted several Convolutional Neural Networks (CNNs) to tackle the challenge of facial expression recognition. Through rigorous evaluation employing diverse post-processing and visualization methods, we assessed their effectiveness. Our findings underscored the prowess of deep CNNs in grasping intricate facial attributes, thus enhancing the accuracy of facial emotion detection. Surprisingly, the incorporation of hybrid feature sets failed to yield improvements in model accuracy. This suggests that CNNs possess inherent capabilities to discern crucial facial features solely through the utilization of raw pixel data.

IX. References

[1] Liam Schoneveld, Alice Othmani, Hazem Abdelkawy. (2021). Laveraging recent advances in deep learning for audio-Visual emotion recognition. Pattern Recognition Letters 146, 1-7.

[2] Yishu Liu, Guifang Fu. (2021). Emotion recognition by deeply learned multi-channel textual and EEG features. Futures Generation Computer Systems 119, 1-6.

[3] Arkaprabha Bhandari, Nikhil R Pal. (2021). Can edges help convolutional neural networks in emotion recognition?. Neurocomputing 433, 162-168

[4] Marina Chatterjee. A Quick Introduction to KNN Algorithm.https://www.mygreatlearning.com/blog/knn algorithm-introduction/ (accessed 10 May 2020)

[5] Wafa Mellouk, Wahida Handouzi. Facial emotion recognition using deep learning: review and insights. Procedia Computer Science 2020

[6] Jagli, P. M., & Shetty, M. P. (2019). Human Emotion Recognition using Machine Learning. International Journal of Trend in Scientific Research and Development, 101-103.

[7] Shervin Minaee, Amirali Abdolrashidi. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. 2019(accessed)

[8] Poorna S S, Anjana S, Parvathy Varma, Anjana Sajeev, Arya K Ca , Renjith Sb , G J Nair a. Facial Emotion Recognition using DWT based Similarity and Difference features. Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) 2018; (): .

[9] Kaggle. Challenges in representation learning: Facial expression recognition challenge, 2013

[10] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S.T.Hawk, and A. van Knippenberg. Presentation and validation of the radboud faces database. Cognition and emotion,24(8):1377{1388, 2010.