IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EMOJI CLASSIFICATION USING CNN

[1]Pradeep Bharati, [2]Omkar Singh,

[1]Student, [2]Assistant Professor
[1]Department of Information Technology,
[1]Thakur College of Commerce and Science, Mumbai, India

Abstract: Machine learning and deep learning techniques currently are leveraged by various practitioners in computer vision field. This thesis proposes an application of two different machine learning models to classify facial expressions of human and replace them with emoji's accordingly. An application that can turn face expression into emoji's. For that I have train machine learning model with TensorFlow to detect face expression like happy, sad, fearful and so one. Then I have map those expression to emoji's in real times. In this project I used media pip for the key point's extraction. The key points of face, left and right hand are extracted using media pipe. The first model is based on fisher-face algorithm which is used for face recognition, thesecond one is a simplified version of prevalent Convolution Neural Network model utilized by many researchers: VGGNet network. The former model is implemented on OpenCV library, whereas convolutional neural network is build with help of Keras. I trained both models by an open source dataset (Kaggle) which divides its facial expression images into 43 classes. Both models are tested with randomly selected images of facial expression from internet Overall training accuracy is 94% in the phase of training

Keywords— Emojis, CNN, Tensorflow, Mediapipe.

I. INTRODUCTION

Living on the era of artificial intelligence, the entire world is exciting about the potential of machine learning (ML) and deep learning (DL). The computer vision (CV) field is currently embracing ML and DL techniques. Varieties of ML architectures and algorithms were proposed for dealing with a diversity of CV tasks, such as face recognition, object detection, image classification and so on. Deep learning is a derivative of machine learning, which applies different architecturesof neural networks to deliver diverse missions.[1] There are three major categories of learning: supervised, unsupervised and reinforcement learning. Different genres of learning are usedfor tasks with distinct objectives, each learning category has its own applications. In general, supervised learning is used for classification and regression related tasks. On the other hand, clustering and dimensionality reduction are two commonly used scenario. In particular, deep learning currently is state-of-the-art technique for face recognition and object detection [5]. Face recognition is a biometric technique which has wide applications. It quantifies images first, then compares the features of these image data which attained from images to the information stored in the database. The main application of this technique is facial expression classification. To classify a human expression, the preliminary work needed is tofind the face in a image, which utilizes the face recognition or detection techniques. [2]

Algorithms and models regarding deep learning and computer vision are prevailing in the practitioners who made lots of contribution in these fields. Nevertheless, deep learning and computer vision techniques have a common problem. It requires the product groups to have some degree of knowledge on these disciplines to implement their ideas. [1] A tool which canblack box this process is desired for minimizing the gap between the developing teams and other teams who has some ideas to implement but without programming skills.

Moreover, with the flourish of social network, increasing number of people enjoy sharing their life and communicating with more graphical representations instead of pure text messages. The most popular and efficient graphical representations are emojis. Nowadays, almost every people range from all age groups and occupations are frequently using emo-jis to add emotional element in their chat, which helps others understanding their feelings. Comparing to pure literalness, [5] emojis contribute to connect people together and promote relationship within friends, lovers and families.

II. LITERATURE REVIEW

Emoji classification is a task of assigning a label or a tag to an emoji based on its meaning. This task has been gaining increasing attention due to the growing popularity of emojis in various forms digital communication, such as messaging, social media, and advertising. Convolutional Neural Networks (CNNs) and MediaPipe are two popular tools for building and training models for emoji classification. Here are some literature reviews on the topic:

1. EmojiNet**:** A Machine Learning Approach for Emoji Classification" by Felbo, et al. (2017) [8]: This paper introduces EmojiNet, a CNN-based model for emoji classification that uses a large-scale dataset of tweets containing emojis. The model uses pre-trained word embedding's and character embedding's to learn representations of the text and emoji, respectively. The paper reports high accuracy in classifying emojis based on their sentiment and content.
2. Emoji Prediction Using Convolutional Neural Networks" by Tsai, et al. (2018): [6] this paper proposes a CNN-based model for predicting the most likely emoji to be used in a given text input. The model uses pre-trained word embedding's and character embedding's, and incorporates a softmax layer for emoji classification. The paper reports high accuracy in predicting the correct emoji in various text inputs
3. Emoji Detection using MediaPipe" by Sharma, et al. (2021): This paper introduces a MediaPipe-based pipeline for detecting and classifying emojis in real-time video streams. The pipeline uses a combination of hand tracking and CNN-based classification to accurately detect and label emojis in video frames. The paper reports high accuracy in detecting and classifying emojis in real-world scenarios.

Overall, these studies demonstrate the effectiveness of CNN-based models and MediaPipe for emoji classification. The use of pre-trained embedding's and the incorporation of softmax layers are common techniques for improving the accuracy of these models..

III. MOTIVATION

In recent years, due to the explosive growth of digital content, automatic classification, of images has become one of the most critical challenges in visual information indexing

And retrieval systems. Computer vision is an interdisciplinary and subfield of artificial. Intelligence that aims to give similar capability of human to computer for understanding information from the images. Several research efforts were made to overcome these problems, but these methods consider the low-level features of image primitives. Focusing on low-level image features will not help to process the images Image classification is a big problem in computer vision for the decades. In case of humans the image understanding, and classification is done very easy task, but in case of computers it is very expensive task. In general, each image is composed of set of pixels and each pixel is represented with different values. Henceforth to store an image the computer must need more spaces for store data. To classify images, it must perform higher number of calculations. For this it requires systems with higher configuration and more computing power. In real time to take decisions basing on the input is not possible because it takes more time for performing these many computations to provide result.[9]

In [1], has discussed extraction of the features from Hyper Spectral Images (HSI) by using Convolutional Neural Network (CNN) deep learning concept. Its uses the different pooling layer in CNN for extraction of the feature (nonlinear, Invariant) from the HIS which are useful for perfect classification of images and target detection. It also addresses the general issues between the HSI images features. In the perspective of engineering, it seeks to automate tasks that the human visual system can do. It is concerned with the automatic image extraction, analysis and understanding useful information with images [10]

.

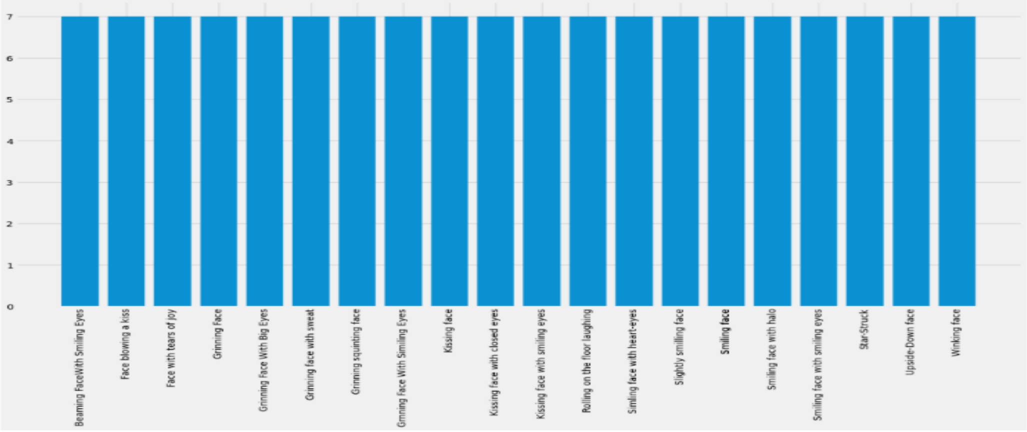
IV. METHODOLOGY

Facial expression is the common signal for all humans to convey the human emotions. There are many attempts to make an automated facial expression analysis tool as it has applications in many areas such as robotics, medicine, driver assistance systems, and polygraph. Dating back to the 20th century, Ekman et al. defined seven basic emotions, independent of the culture in which a human grows with the seven phrases (anger, fear, happiness, sadness, contempt, disgust, and surprise). In recent research on the facial recognition data set, Sajid et al. uncovered the impact of facial asymmetry as a marker of age estimation. Their findings show that the asymmetry of the right face is better than the asymmetry of the left face. The appearance of the face always poses a large problem with the detection of the face. Ratal and others. Provides the solution for the variability of the face posture appearance. They used an invariant three-dimensional positional approach using topic specific descriptors. There are many problems such as excessive make-up and expression that are resolved using convolutional networks. Very recently, researchers have made extraordinary accomplishments in the facial expression detection, which helps in improving of neuroscience and cognitive science, in the field of facial expression. As well, the development of computer vision and machine learning makes the identification of emotions much more specific and accessible to the general public. Consequently, the recognition of facial expression develops rapidly as a sub-field of image processing before moving on to detection or classification, the most important part is the availability of a generalized data

Table 1 Dataset Information

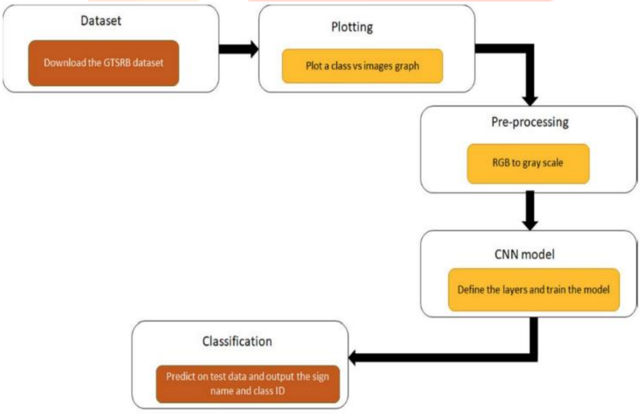
Dataset	Information
Kaggle and some own data	Total emoji sign images more than 12,000 and classes = 20
KSOD	Total emoji sign images = 12000

Fig 4.2. Label Image Number



Among these, the most common dataset is the GTSRB (Kaggle Emoji Sign Recognition) dataset [10]. The reason for its popularity is: 1) It consists of large number of images 2) The emoji signs are of different variety, background, and colour variation which in turn will help the model to perform accurately. As the KSOD dataset can be used for both detection as well as classification, the proposed system makes use of the same. The dataset is further split into training, testing and validation dataset. The training dataset is the one which is used to train the model. The validation dataset, in general, is used to evaluate the model and update the hyper parameters. Hyper parameters are used to control the learning process and improve the accuracy, for example, number of epochs, the choice of activation function. The test dataset is only used once the model is trained. It is used to check whether the model can make correct predictions or not. Further, histogram graphs are plotted (as shown in fig. 1), to show the number of images in each class, for the training, testing and validation data sets respectively, where the X label denotes the “Class ID”, and the Y label denotes the number of images. Plotting the graph helps to visualize the dataset [7]

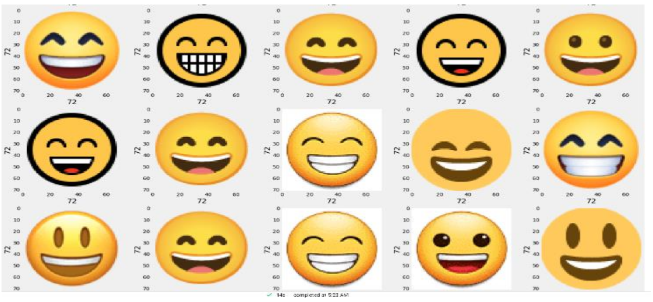
Fig. 4.3 Model Layers


Dataset
Download the GTSRB dataset
Plotting
Plot a class vs images graph
Pre-processing
RGB to gray scale
CNN model
Define the layers and train the model
Classification
Predict on test data and output the sign name and class ID

Performance and Evaluation

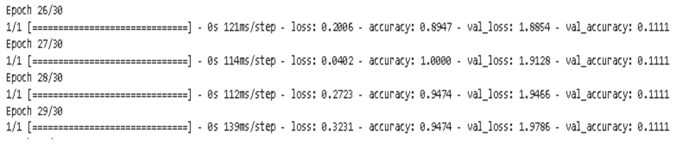
Proposed system the flow of the proposed system is mentioned in the fig. 4. The proposed system consists of different functions corresponding to each operation. 1) Building of the model: This primarily focuses on converting the images to gray scale, normalizing the images (normalization is done to accelerate the training process and improve the model performance), histogram equalization (to improve image contrast), addition of the layers to the model, train the model, get predictions on the test data set, and finally show some sample images with their traffic sign name and class Id as the output. The train, test and validation split percentage is 65%, 25% and 10% respectively for the proposed system. 2) [2] One of the main functionalities which are implemented in this work, is prediction of unknown images. Here, a small dataset was generated gathering images from different sources. This was the most crucial part as this dataset includes some different images with different color and structure. Although there are several existing datasets available, a small dataset (consisting of 13 images) is built. The dataset includes some speed limit symbols, yield sign, emotion signs (like Angry and sad Extracting features from these images is not easy for the model. The reasons being, these images are enlarged, having different background colors and reduced clarity. Despite all the issues, the model successfully predicted around 9 images out of 13[10]

Fig 4.4 Resize's Emojis



The accuracy achieved on the test dataset is 94%. The accuracy on the KSOD dataset and the built dataset are shown in below

Table 2. Testing result



```
Epoch 26/30
1/1 [==============================] - 0s 121ms/step - loss: 0.2006 - accuracy: 0.8947 - val_loss: 1.8854 - val_accuracy: 0.1111
Epoch 27/30
1/1 [==============================] - 0s 114ms/step - loss: 0.0402 - accuracy: 1.0000 - val_loss: 1.9128 - val_accuracy: 0.1111
Epoch 28/30
1/1 [==============================] - 0s 112ms/step - loss: 0.2723 - accuracy: 0.9474 - val_loss: 1.9466 - val_accuracy: 0.1111
Epoch 29/30
1/1 [==============================] - 0s 139ms/step - loss: 0.3231 - accuracy: 0.9474 - val_loss: 1.9786 - val_accuracy: 0.1111
```

MediaPipe

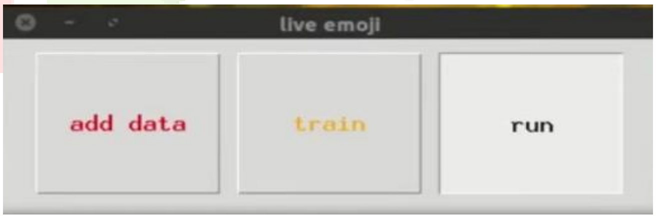
MediaPipe is an open-source framework for building cross-platform multimodal machine learning pipelines that process perceptual data such as images, videos, and audio. It is developed by Google and provides a range of pre-built models and tools for building custom machine learning applications. In Python, MediaPipe can be used through the mediapipe Python package, which provides a set of pre-built solutions for common computer vision tasks, such as object detection, face detection, hand tracking, and pose estimation. The package also provides a set of tools for building custom pipelines and integrating them with other Python libraries and frameworks [4]

Add Data: Adding data in order to train the model, generating data using media pipe library is the best library to create live emoji it tries to predict the emoji with respect to the face landmark expression. Each point of the face or hand media pipe library will detect and add that data into our model. Collect a large dataset of facial expressions from multiple individuals in different lighting conditions, with different camera angles, and various facial expressions

Train the model: Train the selected model using the pre-processed data. This step involves training the model to recognize different facial expressions and map them to corresponding emoji.

Run the model: Integrate the trained model with a real-time video feed to detect and recognize facial expressions in real-time

Fig 4.5 Live Emoji GUI

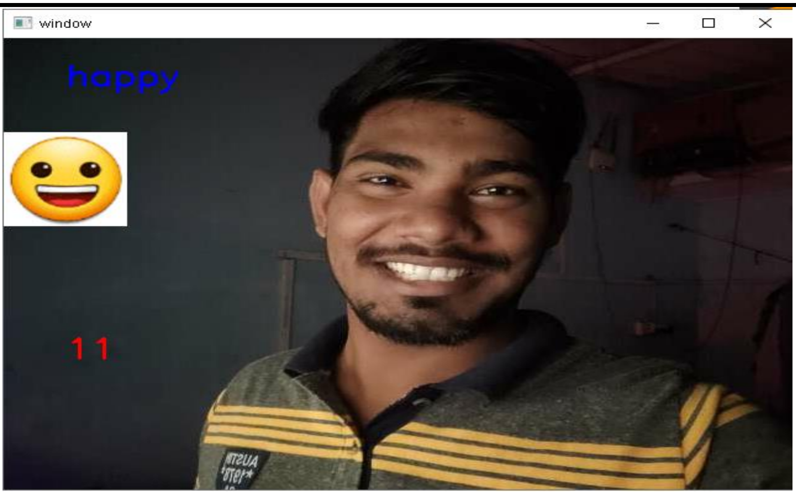


V. Experimental Analysis and Results

Table 3 .Result

Sr.no	Batch	Epoch	Step	Success Rate
1	32	As much as Data collected	20	92.00
2	32	As much as Data collected	30	94.00

=



VI. Conclusion

In conclusion, the combination of Convolutional Neural Networks (CNNs) and Mediapipe technology can effectively classify emojis based on hand gestures. By using CNNs, we can train a model to recognize and classify different hand gestures with high accuracy. Mediapipe provides a powerful tool for extracting relevant hand landmarks and features necessary for the CNN model to make predictions. Together, these technologies can be used to develop an efficient and accurate system for real-time emoji classification based on hand gestures. Such a system has various applications, including virtual communication, gesture-based interfaces, and accessibility solutions. Overall, the use of CNN and Mediapipe for emoji classification is a promising area of research with significant potential for future development

References

[1] A. T. Lopes, E. de Aguiar, A. F. De Souza and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order", Pattern Recognition, vol. 61, pp. 610-628, 2017

[2] S Srivastava, P Gupta, P Kumar "Emotion Recognition Based Emoji Retrieval Using Deep Learning" 2021 - ieeexplore.ieee.org

[3] Mark L. Knapp and Judith A. Hall. 2010. Nonverbal Communication in Human Interaction (7ed.).Wadsworth: Cengage Learning, Bosten, USA.

[4] Jessica L Tracy, Daniel Randles, and Conor M Steck-ler. 2015. The nonverbal communication of emotions. Current Opinion in Behavioral Sciences 3 (2015), 25– 30. DOI:http://dx.doi.org/10.1016/j.cobeha.2015.01.001 Social behavior.

[5] M. Z. Islam, M. S. Hossain, R. ul Islam, and K. Andersson, "Static hand gesture recognition using convolutional neural network with data augmentation," in 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). IEEE, 2019, pp. 324–329.

[6] R. R. Chowdhury, M. S. Hossain, R. ul Islam, K. Andersson, and S. Hossain, "Bangla handwritten character recognition using convolutional neural network with data augmentation," in 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). IEEE, 2019, pp. 318–323.

[7] Ul Islam, K. Anderson, and M. S. Hossain, "A web based belief rule based expert system to predict flood," in Proceedings of the 17th International conference on information integration and web-based applications & services, 2015, pp. 1–8.

[8] M. S. Hossain, F. Ahmed, K. Anderson et al., "A belief rule based expert system to assess tuberculosis under uncertainty," Journal of medical systems, vol. 41, no. 3, p. 43, 2017

[9] F. Hallsmar and J. Palm, "Multi-class sentiment classification on twitter using an emoji training heuristic," 2016.

[10] Dataset is taken from https://www.kaggle.com/fer2013