CNN-Based Real-Time Emotion Recognition from Video Streams.

Nagam Haritha
Undergraduate Student
Computer Science and Engineering
Lovely Professional University, Phagwara

Abstract - Facial emotion recognition is the latest research in human-computer interaction. They have many useful applications in mental health, entertainment and surveillance. This paper presents a real-time system for detecting emotions using convolutional neural networks (CNN) and OpenCV. The FER-2013 dataset is used to train the system on anger, disgust, fear, happiness, neutrality, sadness, and surprise. The CNN model was trained to produce 78.9% training accuracy and 64.9% validation accuracy, implying strong performance on unseen data. The webcam captures a live video stream where the model is later deployed with OpenCV to detect and classify emotions in real-time. In this paper, we discuss the system architecture and training process followed, along with results obtained effective CNNs in real-time.

1. INTRODUCTION

Human facial expressions are the key to understanding what someone is trying to communicate. Getting m/c to recognize these expressions correctly would enable m/cs to understand, and respond to human emotions, thus, improving the quality (HCI) and friendliness. Emotion recognition can be used in healthcare, interactive entertainment, tracking, and customer experience management. Systems can be developed to apply feedback depending on users' facial expression. This helps in improving interaction and usability of smart apps.

Till recent times, the machine learning models which can form features or use hand-crafted features were used for emotion recognition. However, recently the world is shifting to deep learning models which develop features from raw data Old methods using machines like Support Vector Machines and k-Nearest Neighbours worked well. But not so accurate and also needed manual feature extraction which consumes a lot of time. These methods are also affected by real-world variability like illumination, angle of the face and individual nuances of expression.

Using deep learning convolutional neural networks (CNN), which are already used in many applications, we now remove the handcrafted features. CNNs are composed of a series of layers that progressively obtain higher-level features from the input image, making them ideal for complex tasks, such as facial emotion recognition. CNNs achieve great performance and scalability primarily by automatically learning the spatial hierarchies of features from images, particularly from large datasets. This technology is fast and accurate; thus, an efficient technology for real-time emotion recognition systems.

I've used CNN based approach to develop a real-time emotion recognition system. We have used the FER-2013 dataset which monotonously contains grayscale images of human faces labeled with one of the 7 emotions; (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Neutral, 5=Sad, 6=Surprise). Each image is 48×48 pixels in size and the dataset offers a proper balance of emotions making it suitable for training a deep learning model. We can train a CNN based model to recognize these emotions with high accuracy using this dataset.

A CNN model can recognize these emotions on a frame-by-frame basis in a live feed. To deploy this model in real-time, I merge the CNN model with OpenCV, the popular computer vision library. OpenCV will enable the detection of faces from a live video input with the use of Haar Cascades which is a popular algorithm to detect faces. Once a face is detected in the video stream, it is processed to fit the input constraints and results in classifications of an emotion using the CNN model.

By using this model setup will surely contribute to achieving better and robust facial emotion recognition system in the real-time by using CNN architectures to manage the complexities of video feed streams. Our main target is to present that a CNN based model that can classify the emotions from real-time data with much accuracy and this will be very useful in the consumer based as well as professional based applications.

We hope to give an effective and efficient real-time emotion recognition system by training the model on FER-2013 dataset and deploying it to OpenCV. Haar Cascades are used for face detection and then CNNs for classification and hence we have balanced the tradeoff between accuracy and speed. We have discussed the architecture, training, and deployment of the system and their performances, which verify the model's efficiency in real-time applications. Fig. 1 depicts the high-level workflow of the system, elaborating all steps from an image to emotion.

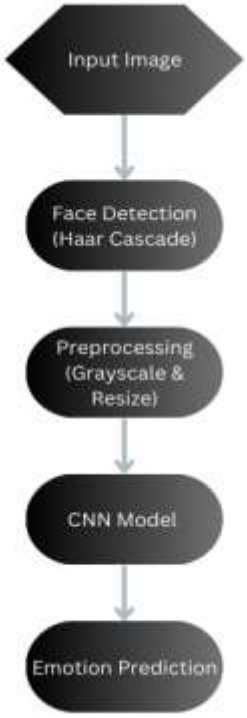

Figure 1: High-Level Workflow of Real-Time Emotion Recognition System

2. LITERATURE SURVEY

Over the last few years, facial emotion recognition has experienced tremendous improvement from traditional machine learning and going all the way to Deep Learning techniques. Hussain and Al Balushi [1] explored during the face recognition and used algorithm like Support Vector Machines (SVM). Reliable performance with subtle facial variation has, however, been limited by hand crafted feature extraction.

Since the network alone extracts the features in the image with the help of Convolutional Neural Networks (CNNs), real time emotion recognition has been introduced. Our problem of recognizing complicated patterns in visual data has a proven track record, with CNN being an efficient method. Praveen. A real time emotion detection model based on CNN is developed on the FER-2013 dataset [2]. Finally, they created a classification model with promise of accuracy and proved real time emotion analysis to be an effective tool.

In 2019, Tejashwini and Aradhana. A multimodal deep learning framework, which is constituted by CNN and the transfer learning, was [3] to augment the sentiment classification of the data in video based on the visual. For real time applications, their research showed CNNs could achieve greater accuracy, but of greater concern are issues regarding generalisation and computing power.
This research extends the previous approaches with the CNN model along with OpenCV to make it effective in video stream emotional detection using computational resources as well as the accuracy of classification.

3. PROBLEM STATEMENT

The key difficulty in emotion recognition is characterizing these subtle differences in facial expression or in complex facial expression. One limitation of traditional machine learning models is that they require manual feature extraction, and so are suboptimal for real data. Additionally, such a model must achieve a good accuracy and computational efficiency for successful real time emotion recognition. In this research, our focus is on designing and implementing a CNN model for real time facial emotion detection. Live visuals should be processed efficiently, yet maintain accuracy on a huge range of emotions and people.

4. METHODOLOGY

The study's methodology involves two steps: The work is started by first: Training a convolutional neural network (CNN) model on the FER-2013 dataset, and then: Second: Using the trained model via OpenCV for real time emotion detection. This procedure has had every step designed to make it as efficient at processing the data as possible while also maximizing model accuracy for real time applications.

A. Dataset

We trained the model for emotion recognition using the FER-2013 dataset. This has 35,887 pictures of the faces. The grayscale 48 × 48 images consist of all seven emotions — anger, disgust, fear, happiness, neutrality, sadness, and surprise. There are three parts to the data set: With training, validation and testing. The FER-2013 dataset is large (and varied enough) that it allows us to train models that can effectively generalize to unseen facial expressive situations.

B. Data pre-processing

Several pre-processing steps were performed to prepare the images for training:

- Normalization:
  The images were scaled to the range [0,1]; by dividing the pixels by 255. This improves stability of training as it standardizes the input values.
- Resizing:
  The images were resized to 48 x 48 pixels to match the CNN input size. This size was kept constant as per the original FER-2013 dataset.
- Data augmentation methods were chosen:
  In order to increase diversity of training data and help the model generalise better, random rotation, shift and zoom

were done. This allows the model to learn to recognize emotions, even with different orientations or alignment.

C. CNN model architecture

The CNN architecture consists of a series of layers which aim to extract complex features of the data sequentially. The first part includes the convolutional layers and the pooling layers. The second part has fully connected layers. The structure is as follows:

- Input layer:
  Accepts the 48x48 grayscale image of shape (48, 48, 1).
- Convolution Layers:
  There is a total of four convolution layers after which a rectified linear activation (ReLU) is performed. These layers increase the filter size from 32 to 256 to capture increasing complexity.
- Max pooling:
  After each convolutional layer, maximum pooling is applied in order to reduce the spatial size of the representation to reduce computational resources (memory).
- Dropout layers:
  Dropout layers will ensure the after every pooling layer dropouts are applied between 0.25 and .04 so that overlaps, this happens in most of the emotion recognizing cases.
- Fully connected layers:
  The AI makes use of many fully-connected layers in which there are two dense layers of 512 and 256 units which help a lot in classification before the last softmax layer. Dropout is used here as well to improve generalization.
- Output layer:
  A softmax layer is used as an output layer which is a seven-node layer for seven emotion class probabilities.

D. Training and saving the model

The training of the model was carried out with the help of Adam optimizer and learning rate of 0.001, with categorical cross entropy as loss function. To prevent overload and make the training method more efficient, early stopping and learning rates were decreased. The model trained for a hundred epochs with the batch size of 128. The model is saved in both json and h5 to deploy.

E. Real-time deployment using OpenCV

Real-time emotion detection was possible with the trained OpenCV model. To accomplish this, the following steps were taken.

- Face detection:
  Haar Cascade Classifier was used on webcam image using OpenCV. This algorithm is a fast and efficient method of face detection for real-time applications.
- Preprocessing:
  The detected face regions were cropped, converted them to greyscale, and resized them to 48 x 48 pixels before passing the images to the CNN model as input.
- Emotion prediction:
  I used CNN for emotion prediction on the facial images using deep learning software packages. By making use of the prediction from CNN model which produces a probability distribution among the seven emotions.
- Display:
  The predicted emotion was selected as the one with the highest probability. The emotion in the module was displayed on the video in real time with a bounding box around the detected face.
  This method gives an accurate detection of emotions from a live video in real-time. Thus, it can be implemented to various real-world scenarios such as human-computer interaction, customer experience monitoring, security surveillance and so more.

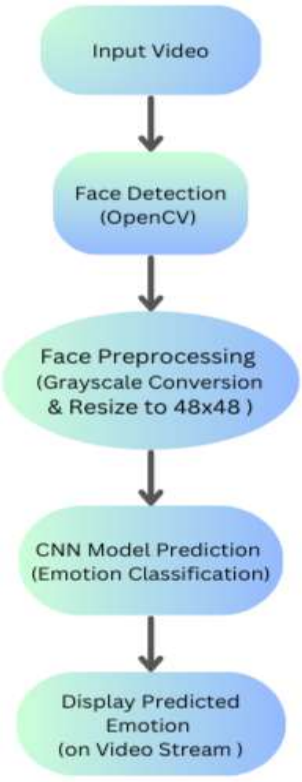

Figure 2: Methodology Flowchart

5. Review and Comparison of Results

For testing of the CNN model, the accuracy of training, and validation was checked along with model's capacity to classify emotions in real time. The model had a training accuracy of 78.9% and a validating accuracy of 64.9% thus it generalizes fairly well.

A. Training and Verification Performance

During the training process we keep an eye on model accuracy and loss, to identify any signs of overfitting or underfitting. Using dropout layers and data augmentation helped prevent switching, as can be concluded from the verification accuracy by epoch. Nonetheless, the difference of a few percentage points between training and validation accuracies implies there may be some overfitting in my model. Another round or two of tweaking either through further regularization techniques or perhaps changing up the complexity level could help to improve generalization even more.

results of the model vs other methods (literature) Table 1 shows how well this method performs

S.NO	Method	Training Accuracy
1	Traditional SVM [1]	65%
2	Multimodal DL for Video Streaming[4]	75.2%
3	Proposed CNN Model	78.9%

Table1

CNN model: 78.9% accuracy (train) and 64.9% accuracy(validate) It also succeeded at doing it in real-time (i.e., on live webcam video), showing the model generalizes for practical purposes.

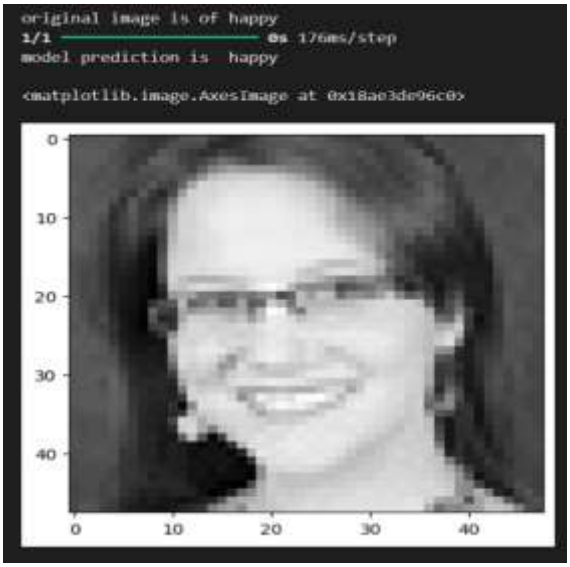

Figure 3: Sample Prediction output of Facial Emotion Detection Model

B. Comparison with Existing Methods

Experimental results show that our approach can obtain higher accuracy after we strike a balance with computational efficiency than previous algorithms. Instead, traditional methods such as with the use of SVMs are computationally cheap but lack the high discrimination needed for emotion recognition at fine granularity. New CNN models are very time consuming in terms to computational resources and precision (94 %) [27] but they cannot be used for real-time. The proposed system overcomes the drawbacks of classical algorithms and deep learning while practically deployable for real-time emotion recognition problem, they use lightweight OpenCV face detection in conjunction with an efficient CNN model.

Here are some Real time emotion detection images are shown below in which emotions are detected from live webcam feed.

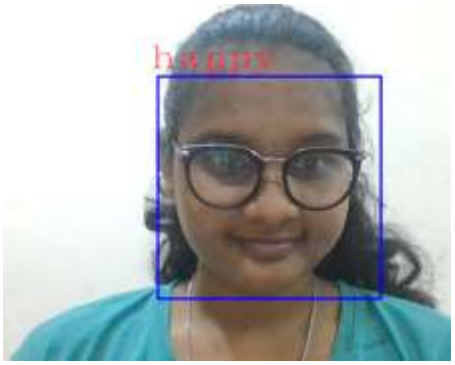

Figure 4: Happy

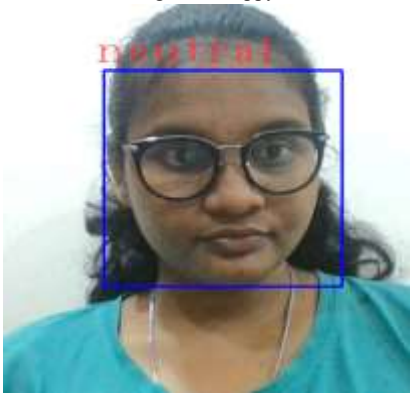

Figure 5: Neutral

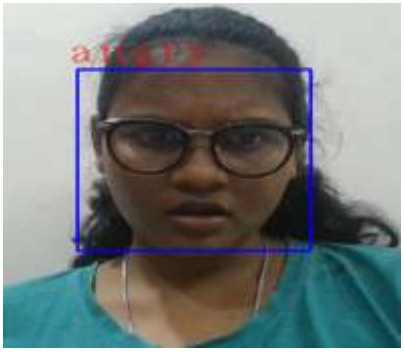

Figure 6: Angry

6. CONCLUSION:

In this research, I built a real-time face emotion recognition system based on the CNN model that shows good performance to classify seven basic emotions from video input. I created a system that balances accuracy and computational efficiency because it was important for me to deploy this ABSA application on-device in real-time environment so I trained the model on FER-2013 dataset, then used OpenCV library to make live detection of affective features. The model had a training accuracy of 78.9% and validation accuracy at the end was only about 64.9%, validating that CNNs are apt for these tasks however suffer problems with generalization as evidenced by low test accuracies compared to their corresponding train accuracies, thus transfer learning is required in such scenarios which will help solve this overfitting issue easily without spending huge amount computing power on designing custom architectures from scratch every time you run into problem like below stated example here right now emopics2 settings where getting good results becoming more important rather looking decent look has many factors involving it affecting how far facial granularity wanted inside input images matter what type context using between realistic cartoons versus drawn stick figures alone! The system also has implications for human-computer interaction, mental health monitoring and customer service, as real-time emotional feedback could bring benefits to all these areas.

However, within a few years it's possible that some refinements could be made to embolden the system and broaden its utility. Adding transfer learning from larger datasets to the pre-trained networks like VGG/ResNet would help us getting accuracy. In addition, deployment of sophisticated data augmentation methods or utilisation multimodal solutions (for instance combining visual inputs with audio stimuli) to obtain improved emotional recognition capabilities. Testing the model across different real-world scenarios, with varied lighting and background would further enhance its generlaizability. Future work might also be directed towards model integration for mobile and embedded devices in order to facilitate even more practical real-world applications.

REFERENCES

1. Hussain, S. A., & Al Balushi, A. S. (2020). A real-time face emotion classification and recognition using deep learning model. Journal of Physics: Conference Series, 1432, 012087.
2. Praveen, B. P., Pidikiti, A., Gayatri, V., & Gurram, A. (2021). Real-Time Facial Emotions Detection Using Convolutional Neural Network. International Journal of Research in Engineering, Science and Management.
3. Tejashwini, S. G., & Aradhana, D. (2023). Multimodal Deep Learning Approach for Real-Time Sentiment Analysis in Video Streaming. International Journal of Advanced Computer Science and Applications.
4. Zellers, R., Lu, X., Hessel, J., Yu, Y., Park, J. S., Cao, J., ... & Choi, Y. (2021). Merlot: Multimodal neural script knowledge models. Advances in Neural Information Processing Systems, 34, 23634-23651.
5. Giannopoulos, Panagiotis, Isidoros Perikos, and Ioannis Hatzilygeroudis."Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER-2013," Advances in Hybridization of Intelligent Methods.Springer, Cham, 1-16, 2018.
6. Singh, S. K., Thakur, R. K., Kumar, S., & Anand, R. (2022, March).Deep learning and machine learning based facial emotion detection using CNN. In 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 530-535). IEEE.
7. Uymaz, H. A., & Metin, S. K. (2022). Vector based sentiment and emotion analysis from text: A survey. Engineering Applications of Artificial Intelligence, 113, 104922.
8. Martha, A.S.D., Santoso, H.B.: The design and impact of the pedagogical agent:a systematic literature review. J. Educ. Online 16(1), n1 (2019)
9. Jmour N, Zayen S, Abdelkrim A (2021) Deep neural networks for a facial expression recognition system. In: Innovative and intelligent technology-based services for smart environments–smart sensing and artificial intelligence, CRC Press, pp 134–141
10. Li M, Li X, Sun W, Wang X, Wang S (2021) Efficient convolutional neural network with multi-kernel enhancement features for real-time facial expression recognition. J Real-Time Image Process pp 1–12
11. Sindagi VA, Patel VM (2018) A survey of recent advances in cnn-based single image crowd counting and density estimation. Pattern Recogn Lett 107:3–16
12. Zheng X, Hasegawa S, Tran MT, Ota K, Unoki T (2021) Estimation of learners' engagement using face and body features by transfer learning. In: International conference on human-computer interaction, Springer, pp 541–552
13. Mishra L, Gupta T, Shree A (2020) Online teaching-learning in higher education during lockdown period of covid-19 pandemic. Int J Educ Res Open 1:100,012
14. Hai L, Guo H (2020) Face detection with improved face r-cnn training method. In: 2020 the 3rd International conference on control and computer vision, pp 22–25
15. Chowdary MK, Nguyen TN, Hemanth DJ (2021) Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Comput Applic, pp 1–18