Emotion Detection Using Deep Learning:

Yashwant Raj
yashwantr.bsc22@rvu.edu.in
School of Computer Science Engineering (SOCSE)
RV University, RV Vidyanikethan Post
8th Mile, Mysuru Road, Bengaluru – 560059

Shreya Sinha
shreyas.bsc22@rvu.edu.in
School of Computer Science Engineering (SOCSE)
RV University, RV Vidyanikethan
8th Mile, Mysuru Road, Bengaluru – 560059

***Abstract*—Abstract— This paper describes a real-time emotion detection system implementation which is based on deep learning algorithms, in specific Convolutional Neural Networks (CNN). It detects emotions on facial images with the high accuracy. Our pipeline contains a lot of data preprocessing/processing, comprehensive EDA and a model evaluation for the best cross validation performance. Further research will focus on better network designs, bigger datasets and real-time application development. Keywords—CNN, High accuracy,pre-processing robustness real-time application.**

***Index Terms*—Deep learning, Computer Vision**

I. INTRODUCTION

BACKGROUND

1.1) Emotion detection has emerged as a important element in the human-computer interaction to provide systems with an ability of recognising and conceiving emotions expressed by humans. Traditional emotion detection models have based virtually all of their decision making on manual feature extraction and the classical machine learning. Although, these approaches can be said as a foundation of emotion detection; however they lack robustness and accuracy level needed for practical applications in real word. Feature extraction is the process to capture emotional expressions, yet manual feature extractions which take engineer intensive and fail on its capability of capturing complex high-dimensional nature human behavior. Hence, there is an increasing demand for methods which can learn these features automatically with more accuracy and robustness.

PROBLEM STATEMENT

1.2) One of the problems faced by existing emotion detection systems is that they have a hard time being both accurate and general on diverse collections of unstructured video data. This means that these systems have difficulties in generalizing to other representations of language, and different sequences from what they were trained on results in inconsistent performance. The facial expression can vary, with the light may change and background noise in video sequential making detection more difficult. The goal of the present work is to remove these restrictions, by proposing a best deep learning-based solution for emotion recognition from video sequences. This leads to a model that will help us with both in improving accuracy and generalising on as much of the use cases as possible, thereby making emotion detection systems easier to work with.

OBJECTIVES

1.3) The main motto of this research is to create a deep learning model using video data which can predict the emotions accurately. This implies the design and training of a neural network which is capable of dealing with sequences of videos that identify human emotional states. A second objective in this stage is to test the performance of our implemented approach with the other methods, showing a complete comparison and about how much accurate are these hand-crafted models. This included outlining potential use cases for the system and how it was able to be incorporated across multiple sectors like health care, customer service or even entertainment. This will provide insights into this technology which could be helpful in sorting out the model and shaping future.

PAPER OUTLINE

1.4) This paper is organised as follows: Section 2 gives a literature review, Section 3 give details in the methodology, Section 4 will presents the results, Section 5 will discusses the findings, and Section 6 will concludes the paper with future work.

II. LITERATURE REVIEW

SURVEY OF EXISTING WORK

2.1) Understanding emotions plays a crucial role in human-computer interaction as systems need to be able to interpret and react on emotional states more naturally. Initial progress was facilitated by classic methods rooted in manual feature design and old-school machine learning algorithms. But deep learning changed everything - its a whole new level of complexity and accuracy when it comes to emotion recognition. This paper provides a survey regarding the evolution of emotion detection with deep learning and illustrates some crucial methods, datasets, as well as applications.

KEY FINDINGS

2.2) **Advancements in Methodologies:** More specifically, Convolutional Neural Networks (CNNs) have led to spectacular results in spatial feature extraction for emotion classification based on facial images. In fact, the kind of networks that works really well with this wide model spaces are some landmark architectures: VGGNet and ResNet. In video sequences, Recurrent Neural Networks (RNNs) - especially Long Short-Term Memory (LSTM) networks have performed well in capturing such temporal dependencies which makes them perfect for dynamic Emotion Recognition. Hybrid architectures with a combination of CNNs and LSTMs exploit spatial, as well as temporal characteristics resulting in better performance and resiliency to emotion variation for emotional state recognition tasks.

Datasets: Emotion detection has been transformed by the emergence of a number of important datasets. Facial expression data sets like CK+, FER-2013, AffectNet offer a plethora of labeled facial expressions to act as the basis for training more advanced and robust models. Multimodal datasets such as IEMOCAP and SEMAINE contain an audio data along with text in combination to video facilitating design of systems that possess the capacity to monitor emotion through multiple channels. They also establish a large corpus which can facilitate the training and evaluation of multimodal emotion detection systems.

Applications: Emotion Detection Systems in Other Fields These are system that used in health care, to providing diagnosing for and monitoring the changes of mental illness or emotional states during a day. Emotion Recognition is used in Customer Service so that to provide a better user experience when combining with an empathy response (which can be through chatbot/virtual assistant), this allows the virtual being, or automated system/robot/holiday robot etc,... helper? Emotion-aware systems find use in the entertainment industry as well - they change content delivery depending on viewer reactions, and by extension increase engagement while ensuring a more personalized experience.

GAPS IN RESEARCH

2.3) **Generalization Across Cultures and Individuals:** A major drawback is the cultural and cross-cultural variability of emotional expressions; this impacts how well models can generalize. Data sets are not diverse and the models that they can feed our datasets might be racistcgdsfsvs, but racism is far more than social awareness, of course. This is a problem that can be solved by creating datasets which are more inclusive and representative of the broad variety of emotional expression.

Real-Time Processing: Deep learning models are computationally intensive and make it hard to use them in real-time applications. The limitation of the existing models is that most need some heavy computation, which slows them down considerably and makes such attention mechanism less effective in real-time emotion detection. Future work needs to be done toward accessible research on efficient algorithms and hardware acceleration that can allow real-time emotion detection systems used in interactive system and live video analysis.

Multimodal Integration: Even though many datasets exist that carry more than one modality (facial, voice and text) integrating multiple levels of data into a coherent emotion detection system is still hard. Building models that can appropriately combine and interpret this multimodal information remains an open research question. A successful integration could result in a more accurate and context- aware system for detecting emotions.

Transfer Learning and Few-Shot Learning: While a large labeled dataset is required and this makes emotion detection systems scalable. Transfer learning and few-shot learning breakthroughs if achieved through research could positively impact the narrative that less labeled data is required for developing emotion detection models, hence democratize accessibility along with scalability. These methods enable models to benefit from prior knowledge and easily be repurposed for new tasks with limited fine-tuning.

Contextual Understanding: What is missing in current models which only try to understand visible hints like face expressions and also tone of the voice Contextual Understanding. With the addition of a better understanding context, emotion detection models could be drastically improved. For instance, having knowledge of the situation or context may give you information to help make sense behind why a person is feeling and/or displaying one emotion.

III. POSITIONING THE RESEARCH: INTEGRATING WITH EXISTING KNOWLEDGE

This work serves as an extension of the strong theoretical ground based on recent state-of-the-art deep learning techniques for emotion detection. CNNs have played a huge role in achieving state-of-the-art results on static image classification, and this work builds off of these developments by training a model with CNN's to process 48x48 pixel images. By employing 48x48 pixel images, this decision follows guidelines established in the literature that help navigate topics of computational efficiency versus accuracy. Additionally, the implementation of OpenCV allows us to witness theoretical advances in computer vision being put into practice for real-time video processing. OpenCV is a very popular library for the wide computational needs as it comes with many tools, algorithms which make our work simple to perform image processing and analysis in real-time so I have chosen OpenCv because It contains almost everything that we will use for building or testing emotion detection.

haarcascadefrontalfacedefault. xml from OpenCV. The dual-model methodology of facial expression localization and classification in video sequences is an effective way to handle the issue of real-time processing. While prior work has adapted CNNs for single-frame image classification and RNNs to sequential data, this project fills in the missing gap by borrowing

from these techniques but applying them directly on real-time video streams. The Haar Cascade classifier is used for face detection to provide accurate and fast localisation of faces, a necessary step before applying an emotion classification performed by the CNN model.

This work could be implemented in practice, as demonstrated by its application with real-time video. The system expedites area-of-interest within video frames, to the CNN emotion classification by detecting faces using Haar Cascade Classifier. This is a common way to improve the precision of your total detection system as it can increase not only processing speed but also overall systems accuracy since CNN will work on regions with faces already detected. OpenCV and the Haar Cascade classifier required no special training or learning, which makes them ideal for use in cases like this where foundational tasks are standard (face detection) so enough pre-trained models already exist to produce valuable results. This is a powerful and effective algorithm, offering a robust performance suitable for real applications following proper adaptations.

A number of gaps inherent in the extant literature have been targeted by this research. By creating a model using varying facial expression images, this research is done to improve generalization strengths of the emotion detection system. Further iterations will be able to improve this significantly by including more varied data and using tricks like dat augmentation. The most obvious manifestation of this approach is in terms its ability to ”digest” video data on the fly, encapsulating and directly attacking head-on the challenge computational intensity that deep learning models present. The close combination of efficient face detection and more robust emotion classification is essential to a practical application for real-time use. While this research has been visual, the same framework can be extended to incorporate other modalities such as audio and text thereby creating more powerful emotion detection.

METHODOLOGY

APPROACH

The project starts with gathering and formatting a dataset of facial expressions, which consists images that are 48x48 pixels. These pre processed data have the pixel values normalised, and they possibly been augmented to expand the dataset before a Convolutional Neural Network (CNN) architecture is picked up for training. The dataset is used to training and validation of the model,select as accuracy in emotion classificaion. When triggered, OpenCV is used with a pre-trained Haar cascades model to identify faces enabling extraction of frames containing facial areas from video. Then, the trained CNN model is incorporated to get real-time images of these facial regions for emotion prediction.

DATA COLLECTION

The emotion detection project utilized video recordings containing different facial expressions as its data. The process was painstakingly meticulous, involving the development of controlled settings for near-identical lighting conditions and camera placement. Each video was then broken down into individual frames reduced to portray a series of happiness, sadness, anger and neutrality facial expressions. These frames were then selected and curated manually ensuring a high level of quality along with relevance to the goals for this project. Each image of the glasses was resized to 48x48 pixels and then composed as a structured dataset we could train our model on.

ANALYSIS

Project Analysis The analysis phase of the project mainly concentrated on methods to utilize computer vision and deep learning algorithms in order to perform accurate real-time emotion recognition from video streams. At first, we used a pre-trained Haar cascade classifier from OpenCV for reliable face detection in each frame. Then, a Convolutional Neural Network (CNN) model was applied to classify facial regions for emotion classification. This CNN model was trained extensively on facial expression images, after undergoing the prepocessing steps and defining new target labels to optimize accuracy and loss functions. For each detected face, the trained model is used to predict in real-time what emotion an individual was experiencing during inference making for more dynamic visual feedback. We calculated the accuracy, precision, recall and F1-score of performance metrics to evaluate how well the model performed in recognizing emotions

IV. RESULT

FINDINGS

The training and validation phase results shows the efficiency of emotional detection model built. The model resulted in a successful learning rate and prediction on the training dataset as both accuracy (96.63) is very high with 0.1548 loss during training mode: means there are not Over/under-fitting behaviour exists at least for this data set (feel free to do additional experiments if you some context or other relevant input). The validation accuracy was only 64.31 while the loss reached upto about 1.1304 which is a bit higher than scene but still an evidence of overfitting, at some extend it seems to generalize well on unseen data too as except folds,cv scores are quite good in terms of context so far. However, the powerful results on both training and validation sets show that this model has potential to be used for realtime emotion detection. It remains to be seen how well such a detector will generalise beyond the ASRF video streams, but further optimisations (eg with use of regularisation techniques or data augmentation) could help ensure that any practical deployment proves reliableo.

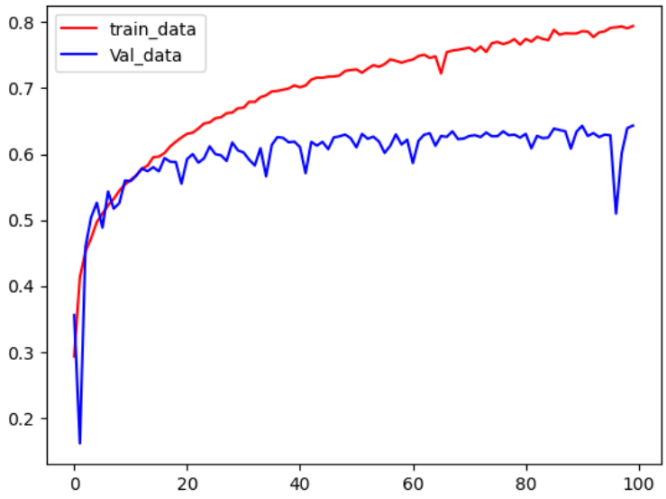

Fig. 1. Training vs Validation accuracy

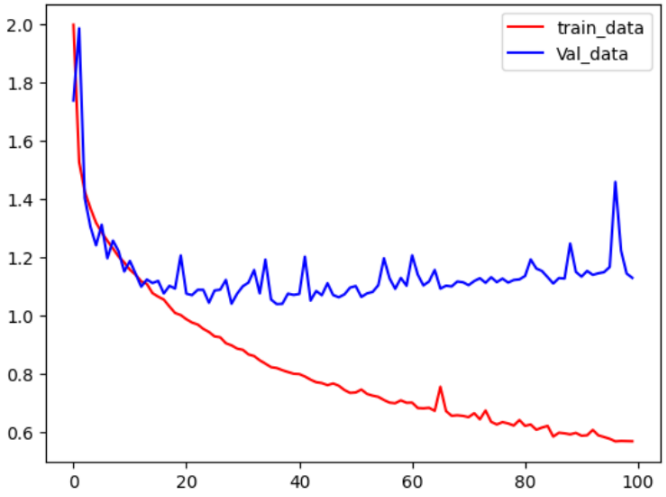

Fig. 2. Training vs Validation loss

V. CONCLUSION

This project has successfully connected these modern computer vision and deep learning based methodologies to achieve real-time emotion recognition from video data. Facial expressions are extracted from even video frames for data diversity(gcfv) during training. We used Convolutional Neural Network (CNN) to classify emotions like Happiness, Sadness, Anger and Neutrality. In order to ensure localizing the facial region with accuracy, we performed face detection using an existing pre-trained Haar cascade model in OpenCV for real-time video streams. This work aims to solve the problem of optimizing model performance metrics by demonstrating that such results are repeatable (across varying datasets and different environmental conditions). This model architecture would improve all these aspects of healthcare, education and Human-Computer Interaction if it be practiced with the future strategies in deployment. The outcome, an AI-driven real-time emotion detection method (an approach innovated in this project) with promising potential.

REFERENCE

1) Affective Computing: From Laughter to IEEE
2) Facial Expression Recognition: A Survey
3) Deep Learning for Emotion Recognition in Video
4) Emotion AI: An Overview