International Journal of
Image Processing and
Pattern Recognition

ISSN: 2456-6985
Volume 10, Issue 1, 2024
January—June
DOI (Journal): 10.37628/IJIPPR

https://journalspub.com/journal/ijippr

Review IJIPPR

Facial Expressions Detection Using Machine Learning

Shravani Jadhav[1], Pratik Kakade[2], Amar Raut[3,*], Achala Deshmukh[4]

Abstract

Facial expressions are necessary in human communication and understanding emotions. First and foremost, identifying emotions is a critical factor for many businesses to comprehend how their customers are responding to the things they have introduced. It can also be utilized to find out if the amenities provided to their staff are meeting their needs. Additionally, it has a plethora of other applications, such as utilizing a camera to identify someone's attitude without having to approach them. Additionally, with little adjustments, the same technique may be used to a variety of different sectors, including face detection, attendance tracking, mask detection, and many more. Detecting facial expressions automatically has significant uses in different platforms, like interaction between human and machines, emotion analysis, or mental health diagnostics. This research focuses on developing a robust and efficient system for facial expression detection using deep learning techniques. The experimental findings show how well the suggested method works to reliably identify and categorize expressions. The system achieves performance on benchmark datasets, showcasing its potential for real-world applications. Additionally, we compare our results with those of conventional machine learning approaches, emphasizing the superior performance of deep learning methods in tasks involving facial emotion identification. This work advances the field of computer vision by offering a sophisticated method for automatic face emotion recognition. The accuracy, resilience, and real time capabilities of the built system make it a useful tool for a variety of applications, such as virtual reality, gaming, healthcare, and customer experience analysis. Facial emotion analysis is effectively employed in surveillance films, expression analysis, gesture recognition, computer games, smart homes, depression treatment, patient monitoring, anxiety, lying detection, psychoanalysis, paralinguistic communication, operator tiredness detection, and robotics.

Keywords: Facial emotion detection, convolutional neural network (CNN), image edge computing, real-time expression detection

INTRODUCTION

In a number of fields, including affective computing, emotion analysis, and human–computer interaction, facial expression recognition is an essential task. Conventional techniques for recognizing facial expressions frequently depend on manually created features and basic learning models, which may not be sufficient to accurately capture the intricate spatial and temporal patterns found in facial expressions [1]. There is increasing interest in using deep neural networks for face expression detection applications due to the progress made in deep learning. This work investigates the application of deep learning methods, especially convolutional neural networks (CNNs) for the recognition of facial expressions and provides a performance comparison. There are

***Author for Correspondence**
Amar Raut
E-mail: rautamar328@gmail.com

[1-3]Student, Department of Electronics & Telecommunication, Sinhgad College of Engineering, Pune, Maharashtra, India
[4]Professor, Department of Electronics & Telecommunication, Sinhgad College of Engineering, Pune, Maharashtra, India

Received Date: April 24, 2024
Accepted Date: May 04, 2024
Published Date: May 30, 2024

Citation: Shravani Jadhav, Pratik Kakade, Amar Raut, Achala Deshmukh. Facial Expressions Detection Using Machine Learning. International Journal of Image Processing and Pattern Recognition. 2024; 10(1): 11–16p.

several features that can be used for emotion recognition, including face, speech, and even text. For a variety of reasons, including their visibility, their abundance of traits that aid in emotions detection, and their ease of collection relative to other methods of human recognition, facial expressions rank among these attributes as some of the most, if not the most, popular. Deep learning, and CNN networks in particular, have made it possible to extract and learn numerous aspects for a respectable approach for recognizing facial expressions.

However, it is noteworthy that a few facial features, such as the lips and eyes, provide the majority of the cues when it comes to expression of emotions. The following categories apply to human emotions: fear, scorn, disgust, rage, surprise, sadness, happiness, and neutrality. These feelings are quite nuanced. Because even little variations produce distinct expressions, it can be difficult to discern even the smallest variations in facial muscular contortions. Additionally, because emotions are so context-dependent, different people or even the same people may exhibit the same feeling in different ways [2].

Even while we can concentrate just on the parts of the face that exhibit the greatest range of emotions, such as the mouth and eyes, it is still crucial to figure out how to identify and classify these gestures. For these tasks, machine learning and neural networks have been employed.

RELATED WORK

Acquiring the spoken command or face expression feature vector is CNN's main objective. Next, a set of long short-term memory (LSTM) models with shared weight reflect an input sequence provided by a CNN with an input sub picture or spectrogram corresponding to face expression and spoken instruction, respectively. In short, two sequential recurrent convolution networks (SRCNs) are developed: SRCN-DFER for dynamic face emotion recognition and SRCN-WSCR for wireless voice command recognition. Emotion recognition has been shown to have many positive uses in people's lives as artificial intelligence has grown.

Unfortunately, most of the emotion detection methods that are now in use perform poorly, which impedes the development of these methods in practical applications. To tackle this problem, we proposed a deep automated encoder-based multi-modal emotion recognition method based on expression–electroencephalogram (EEG) interaction. First, decision trees are employed as an objective feature selection method. The test samples' expression category is then determined by examining the solution vector coefficients in relation to the facial expression characteristics that the sparse representation found. The data from the EEG and facial expressions are then combined using a bimodal deep automated encoder [3].

An individual's sentiments and reactions can be greatly influenced by the emotional context of a particular situation. However, existing emotion detection techniques focus mostly on feature analysis of the target subject and do not sufficiently combine these aspects with the scene's contextual information. In order to address this problem, we present a new emotion recognition model that combines three independent and prioritized deep CNN with a feature fusion enhancement method to efficiently combine subject features, body, pose, information, and facial information in the entire picture [4].

DATASET DESCRIPTION

The proposed model is evaluated on “FER-2013, learn facial expressions from an image” datasets as shown in Figure 1. This dataset is available on Kaggle website. The data consists of 48 by 48-pixel grayscale images. Every image has a face that is generally centered and takes up the same amount of space since the faces are automatically registered. Depending on the emotion expressed in the facial expression, each face must be assigned to one of seven groups. Train.csv is composed of the "pixels" and "emotion" columns.

For each emotion shown in the image, a numeric code between 0 and 6 is included in the "emotion" column. For every image, a string enclosed in quotes may be found in the "pixels" column. You have to guess what the mood column will be test.csv just has the "pixels" column [5].

Figure 1. Example of images from Facial Expression Recognition – 2013 (FER-2013) dataset.

METHODOLOGY

To find faces in photos, use face detection methods (such as deep learning-based detectors or Haar cascades). Add more variety to the collection by transforming it using operations like flipping, scaling, and rotation. Enhancement improves the model's ability to generalize. Take pre-processed photos and extract their features. CNN is one type of network that can be used as feature extractors. Select a suitable machine learning model. Divide the dataset into sets for validation and training. Use the parameters like accuracy, precision, recall, F1-score to assess the trained model. Construct a real-time system by integrating the trained model. Depending on the intended use case, deploy the real-time facial expression detection system as a mobile app, online service, or application. The software development cycle comprises several stages. It is easy to follow and works best for this project's implementation. During this stage, business requirements, use case definitions, and corresponding documentation are studied. During this stage, data model designs are defined, and various data preparation and analysis are conducted [6].

Implementation

At this point, the model's actual development will take place. The back-end and front-end components of the agent will be developed using suitable algorithms, mathematical models, and design patterns based on the requirements and data model designs from earlier stages.

Testing

In this step, the developed model built upon the earlier phases will be put to the test. A number of validation tests are run on the trained model. Deployment: The model is prepared for deployment or use in simulated scenarios once its accuracy scores have been verified.

Maintenance

The model will face a variety of inputs and circumstances during the use of the produced solution, which may have an impact on the model's overall correctness. Alternatively, as time goes on, the model might not meet the updated company needs. As a result, frequent maintenance is needed to keep the model operating in the intended manner [7].

ALGORITHM

Artificial intelligence has advanced rapidly in terms of bridging the capability gap between computers and humans. In order to attain excellent outcomes, both enthusiasts and researchers concentrate on different aspects of the area. Among these many fields is computer vision. The field's objective is to enable machines to perceive and comprehend the world in a manner akin to that of humans, and to even employ that capacity for a variety of uses, such as classification of images, recommendation systems, natural language processing (NLP), and image video recognition. Over time, deep learning has been utilized to create and improve computer vision through the usage of CNN as the primary algorithm. CNNs, also known as Conv Nets, are deep learning algorithms that can identify and categorize various objects and features in an input image. They may also be trained with weights and biases. Comparatively speaking, a Conv Net requires far less preprocessing than other categorization methods. Even when filters are manually generated using outdated techniques, Conv Nets can learn these traits and filters given enough training.

Face Recognition Is Divided into Three Steps

Face Alignment and Detection

Finding faces in the input image is the first step. A machine learning algorithm known as a Haar cascade classifier, can be used to accomplish this. The face in a picture or video has to be found by the system. The majority of cameras today come equipped with a face detection feature. Facebook and other social media sites employ face detection in addition to other methods to enable users to apply effects to the pictures and videos that they shoot using their apps. The fact that faces are frequently not turned to face the camera directly presents a problem in face detection. Faces turned away from the main focus appear very different on a computer [8].

Feature Measurement and Extraction

The process of extracting features from faces begins with face alignment and detection. Herein lies the role of the CNN. CNNs can extract advanced features from images that are used to find the faces in database.

Face Recognition

The last stage involves matching the features that were retrieved from the faces in a database. The Euclidean distance metric, which calculates the similarity between two vectors, is typically used for this.

RESULTS

The Intel Core i5 computer used in the research has eight gigabytes of memory. The developer environment is called Spyder. For preprocessing pictures, the OpenCV and SRGAN programs are utilized. The crucial component for the key feature extraction is the Media Pipe 0.8.6 library. The machine learning classifiers are implemented and the evaluation metrics for the suggested model are computed using Scikit-learn 0.24.2. Also utilized are the libraries NumPy, Pandas, Math, OS, and Matplotlib. The suggested framework is assessed using the following five metrics: recall, accuracy, precision, F1-score, and training duration. The performance metrics of the trained model on a particular dataset are usually included in the outcomes of machine learning projects [9, 10].

You may evaluate how successfully the model learns to distinguish between different face emotions, such as joyful, sad, or furious, by looking at the accuracy versus epoch plot as shown in Figure 2. A well-crafted graphic suggests that the model is successfully encapsulating the characteristics that differentiate these expressions. All things considered, the accuracy versus epoch plot is a useful tool for tracking a facial expression detection model's training process and making sure it performs well without overfitting.

X-Axis

Epoch: This indicates how many times the model has been run through the training dataset in order for it to learn. An epoch is a single pass. The model is exposed to the data more and should ideally keep getting better as the number of epochs rises. The fraction of correctly classified facial expressions is shown on the Y-axis (accuracy). A value of 1 denotes perfect accuracy, or the model's 100% recognition of all expressions.

The Plot Often Shows Two Curves

Accuracy of Training

The accuracy of the model using actual training data is shown by this curve. As the model learns to recognize the patterns in the training data, this curve ought to rise over the course of the epochs.

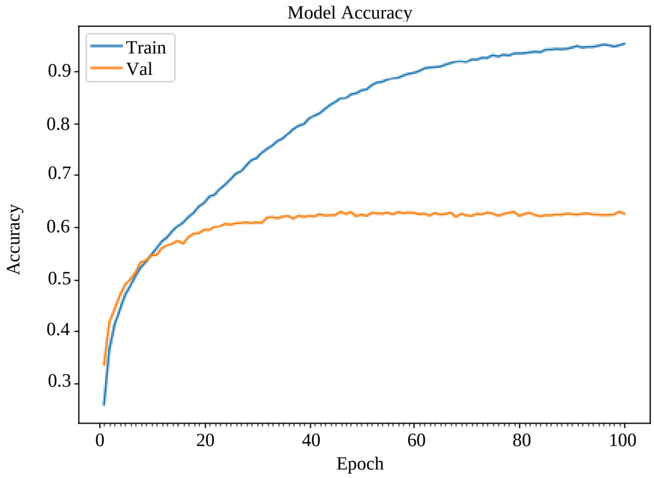

Figure 2. Accuracy plot.

Table 1. Comparison of convolutional neural network (CNN) with some most common algorithms used in facial expressions image processing.

Model	Accuracy	Precision	Recall	F1-Score
Recurrent neural network (RNN)	0.80	0.78	0.82	0.80
K-nearest neighbor (K-NN)	0.79	0.77	0.81	0.79
Decision trees	0.75	073	0.77	0.75
Random forest	0.85	0.84	0.86	0.85
Convolutional neural network (CNN)	0.88	0.87	0.89	0.88

Conversely, overfitting happens when a model learns from the training data too effectively and performs badly on unknown data.

Validation Accuracy

This curve shows the model's performance on an untrained, independent validation dataset. This curve ought to rise as well, albeit more slowly than the training accuracy.

Analysis of Plot

Good Performance: A good plot indicates a steady increase in training and validation accuracy, followed by a plateau in validation accuracy (no discernible improvement). This suggests that the model is picking things up quickly and generalizing well to new data.

If the accuracy curves remain low during the course of the epochs, the model might be underfit. This suggests that either the model is overly simplistic or it has not been trained thoroughly enough to capture the complexity in the data.

Despite not being able to generalize, the model is able to recall the training set as shown in Table 1. This can be addressed with strategies like regularization (adding constraints to the model) or early halting (stopping training before overfitting starts).

DISCUSSION

The findings of the simulation indicate that the suggested method performs well in identifying human emotions. The suggested method performs better than other efforts in this field, as can be seen. Machine learning–based facial expression recognition has enormous potential to improve mental health diagnosis, facilitate better human-computer connection, and advance a number of industries, including education, healthcare, and entertainment. In the upcoming years, this technology's progress will be driven by ongoing study, ethical considerations, and technological advancements.

CONCLUSION

When the algorithm predicts anything incorrectly, in this case, the right label is often the second most likely emotion. The powerful face recognition model provided by this study's facial expression identification system expands on the corpus of knowledge by connecting behavioral variables to physiological biometric traits. The recognition system uses geometrical structures as its base matching template, and these structures are linked to physiological aspects of the human face that are relevant to various facial expressions, such as happy, sad, fear, neutral, etc. The behavioral part of this system connects the idea that underlies different expressions as the basis for property. The property bases of genetic algorithmic genes are separated into a visible and hidden category. The gene training set provides a strong expressional recognition model and evaluates the expressional uniqueness of each face in the context of biometric security. In the modern world, it is imperative to create intelligent computers that can recognize the expressions on the faces of different people and respond accordingly.

It has been proposed that internet of things (IoT) sensors can be integrated with emotion-oriented deep learning techniques. In this instance, it is anticipated that doing so will raise FER's performance at par with that of humans, which will be highly beneficial in the fields of investigation, security, surveillance, and healthcare. A comparative analysis of deep learning methods for face expression identification is presented in this work. The outcomes of the experiment show how well the convolutional recurrent neural network model captures temporal and spatial data to enhance face expression recognition. The results of this study have implications for a number of applications, including affective computing, emotion recognition, and human–computer interaction. They also advance deep learning–based systems for facial expression analysis.

Acknowledgements

We would like to thank our respected mentor and guide Dr. A. M. Deshmukh, for helping and guiding us. Her constant encouragement and suggestions were very useful in the completion of this project work.

REFERENCES

1. Agarwal S, Santra B, Mukherjee DP. Anubhav: recognizing emotions through facial expression. De Visual Computer. 2018; 34 (2): 177–191.
2. Pan X, Ying G, Chen G, Li H, Li W. A deep spatial and temporal aggregation framework for video-based facial expression recognition. IEEE Access. 2019; 7: 48807–48815.
3. Aneja D, Colburn A, Faigin G, Shapiro L, Mones B. Modeling stylized character expressions via deep learning. In: Lai S-H, Lepetit V, Nishino K, Sato Y, editors. Proceedings of the Asian Conference on Computer Vision. Cham, Switzerland: Springer; 2016. pp. 136–153.
4. Zhao S, Cai H, Liu H, Zhang J, Chen S. Feature selection mechanism in CNNs for facial expression recognition. In: Proceedings of the British Machine Vision Conference, Newcastle, UK, September 3–6, 2018.
5. Oloyede MO, Hancke GP. Unimodal and multimodal biometric sensing systems: a review. IEEE Access. 2019; 4: 7532–7555.
6. Phelps EA. Emotion and cognition: insights from studies of the human amygdala. Annu Rev Psychol. 2020; 57: 27–53.
7. Hossain MS, Muhammad G. An audio-visual emotion recognition system using deep learning fusion for a cognitive wireless framework. IEEE Wireless Commun. 2019; 26 (3): 62–68.
8. Huo J, Li WB, Shi YH, Gao Y, Yin HJ. WebCaricature: a bench-mark for caricature recognition. arXiv preprint: 1703.03230, 2020. Available at https://arxiv.org/abs/1703.03230
9. Jiang C-S, Liu Z-T, Wu M, She J, Cao W-H. Efficient facial ex- pression recognition with representation reinforcement network and transfer self-training for human–machine interaction. IEEE Trans Ind Informatics. 2023; 19 (9): 9943–9952. doi: 10.1109/TII.2022.3233650.
10. Rasheed A, San O, Kvamsdal T. Digital twin: values, challenges and enablers from a modeling perspective. IEEE Access. 2020; 8: 21980–22012. doi: 10.1109/ACCESS.2020.2970143.

