IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, October 2024

A Review of Emotion Detection Datasets and Deep Learning Techniques

Shaik Karimunnisa Begum[1] and Dr. Kusum Rajawat[2]
Research Scholar, Department of Computer Science[1]
Research Guide, Department of Computer Science[2]
Sunrise University, Alwar, Rajasthan, India

Abstract: *Facial expression-based emotion identification is an intriguing research subject with applications in safety, health, customer service, and human-machine interactions. This technology's success has led to the development of several deep learning architectures to improve performance. This article reviews current deep learning research in facial emotion recognition (FER). Highlight the contributions, architecture, and databases used, and show progress by comparing the suggested techniques to the outcomes. Deep learning models, especially CNNs, offer great promise among all FER techniques because to their automated feature extraction and computational efficiency.*

Keywords: Facial databases, deep learning, and facial emotion identification

I. INTRODUCTION

A method called facial expression recognition (FER) makes use of photos of human faces to forecast basic facial emotions. FER has drawn a lot of interest because to its prospective applications in computer interfaces, autonomous driving, health management, human aberrant detection, and other related fields. Human-computer interaction has been the focus of an increasing amount of study in recent years due to the growing popularity of artificial intelligence and pattern recognition. Human facial expressions play an important role in social communication. Both verbal and nonverbal communication are often used. Nonverbal communication between people includes things like body language, facial expressions, eye contact, and language. The categorical model, which characterizes emotions in terms of distinct fundamental emotions, is still the most often used viewpoint for FER even though alternative emotion description models like the continuous model employing the affect dimensions are believed to reflect a wider spectrum of emotions.

Artificial neural networks, a kind of machine learning that uses vast quantities of data to learn, are used in deep learning. These networks are analogues of the human brain. In the same way that humans learn from experience, the deep learning system would repeat a task.

Each time, it will improve until the intended outcome is achieved. A model is trained using computer picture data for control, identification, or verification. The patterns and traits present in the photos are analyzed using deep neural networks. FER systems may be divided into two groups based on feature representations: dynamic sequence FER and static picture FER. While dynamic-based techniques evaluate the temporal relationship between successive frames of the given facial expression sequences, static-based approaches use basic spatial information from the current photos to encode the feature representation.

A number of variables, including head deflection, partial blockage of face regions, and variations in illumination, continue to pose hurdles for facial emotion identification in the most current investigations. These interferences may considerably impede face detection ability and decrease FER accuracy. Therefore, deep learning could have been a good way to deal with these issues.

Pattern recognition has already advanced significantly thanks to convolutional neural networks (CNNs), especially in the areas of face recognition and handwritten mathematical expression detection. CNN can automatically decipher and learn the target's abstract signatures when it has a deep network. CNN, or any other deep network, can effectively perform FER under severe circumstances because of its deep layers and intricate design.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, October 2024

The following are common problems with emotion recognition systems: (a) misclassification issues; (b) minor alignment issues impact performance; (c) a fully connected neural network is unable to learn the complex local pixel relationship in image data; (d) in an architecture based on convolutional neural networks, local spatial features such as eyes, noses, etc., are learned well, but global spatial features such as the animal or face as a whole are not learned much; (e) dataset issues impact performance; and (f) contains mislabeled data.

This is how the remainder of the survey is put together. A comprehensive review of relevant work on FER techniques is given in Section II. The performance details and comparison of the suggested model are given in Section III. The conclusion is explained in Section IV.

II. LITERATURE SURVEY

Evolutional spatial-temporal networks were suggested by K. Zhang et al. [1] to extract various characteristics for face emotion recognition. A PHRNN model is proposed to extract dynamic geometry information. It is simpler to explain the development of dynamical expressions when landmarks are separated into several sections according to variations in face morphology. Subsequently, they created an MSCNN model to augment the still appearance data by using both recognition and verification signals as supervision. In order to enhance the variety of distinct expressions and decrease the difference between similar expressions, the two signals correlate to two distinct loss functions. The CK+, Oulu-CASIA, and MMI databases are the models used in this database. Across three facial expression datasets, this strategy reduces the error rates of the previous best approaches by 45.5%, 25.8%, and 24.4%, respectively. In the confusion matrices across the three databases, they discovered that it is challenging to record spatial-temporal information about expressions with little motion. As a consequence, specific methods to measure these emotions, like metriclearning, are required, as are more complex frameworks to explain the actions of these crucial areas.

A weighted mixture deep neural network (WMDNN) was introduced by B. Yang et al. [2] in 2018 to automatically extract the parameters that are necessary for FER tasks. A number of preprocessing methods, including face detection, rotation correction, and data augmentation, are required for input facial data. face grayscale pictures and its corresponding local binary pattern (LBP) images are included in WMDNN's two channels of face images. To automatically extract expression-related variables from face grayscale photos, a partial VGG16 network is built using initial parameters taken from the VGG16 model pretrained on ImageNet. Figure 1 illustrates the extraction of a shallow convolutional neural network (CNN) from LBP face pictures using DeepID features. The CK+, JAFFE, and Oulu-CASIA datasets are used. Accurate facial expression classification is possible using the suggested method. Despite the achieved precision, some recognition is lost. Subsequent study will concentrate on streamlining the network used for speeding, followed by a focus on additional face image channels that may be used to enhance the fusion network. It only performs well on some characteristics.

In 2019, H. Zhang et al. [3] proposed a facial expression detection technique based on a CNN model and image edge detection that efficiently extracts the facial attributes. The edge of each layer of the data is retrieved in the convolution process after the facial expression image is normalized. To maintain the texture image's edge structure information, the retrieved edge information is placed on each feature image. The maximum pooling approach is then used to reduce the dimensionality of the retrieved implicit features. The suggested technique is compared to the traditional neural network FRR-CNN model and the R- CNN algorithm in order to validate the robustness of this method for facial expression recognition under complicated backgrounds, as shown in Fig 2. It was created by combining the Fer-2013 facial expression database with the LFW data set in a scientific way. The suggested method achieves an average recognition rate of 88.56% with fewer iterations, and the training speed on the training set is roughly 1.5times quicker than the contrast approach. The issue of datasets is very challenging, and noisy variation (such as face posture, occlusion, and blurring) of these datasets affects the performance and needs more robust models that satisfy real conditions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, October 2024

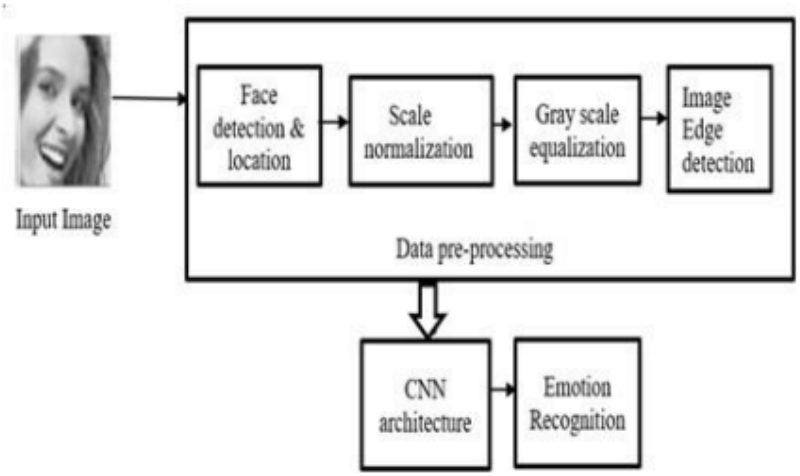

Fig. Block diagram of CNN approach, H. Zhang et al.

Convolution neural networks (CNNs) are frequently used to address computer vision problems like object identification, object tracking, image classification, and image segmentation. In 2020, Garima Verma et al. [4] proposed a CNN-based deep learning model for analyzing facial expressions and predicting emotions. The proposed method includes two component models. In the first, a CNN model identifies a secondary emotion, like love or affection, while in the second, it categorizes the primary emotion in a picture, such pleasant or sad. This model uses the JAFFE and FER2013 datasets. There are preprocessing procedures for creating the model. The core CNN is made up of three convolution layers and three completely connected layers, each with 1024 neurons. The last completely linked layer is a two-neuron layer that applies the classification using the SoftMax function. Each of the five 2D convolution layers of the secondary CNN is then linked to the max-pooling layer. To cut down on overfitting and training time, the network is enhanced with two dropout layers, each with a dropout rate of 0.2. The FER2013 and JAFFE datasets were used to train the model, which produced an accuracy rate. It is still difficult to deal with the mislabeled data. Thus, sophisticated CNN-based models are required.

A lightweight convolutional neural network (CNN) for real-time and bulk face emotion recognition was proposed and built by N. Zhou et al. in 2021 [5]. This approach use multi-task cascaded convolutional networks (MTCNN) to do face recognition and send the gathered face coordinates to the facial emotion classification model they first built. One of the cascade detection functions in multi-task cascaded convolutional networks may be employed separately, reducing memory consumption. The classification model employs global average pooling in lieu of the fully connected layer seen in the typical deep convolution neural network model. By assigning each feature map channel to the appropriate category, the black box features of the fully linked layer are reduced to some extent. The model simultaneously adds the normalization term and merges residual modules with depth-wise separable convolutions, which reduces a significant number of parameters and increases portability. The FER2013 dataset is used here. The model was accurate. Real-life facial expressions may contain a lot of noise, such as images with too bright or too dark illumination, blurred images, the majority of the face being blocked, and other factors that make recognition challenging, even though this model achieves accuracy when compared to other recent works.

Emotion Recognition Technology

Neural Networks: Like neurons in the brain, these computer systems are composed of interconnected nodes. It can cluster and categorize raw data, uncover hidden patterns and connections, and learn and become better over time by using algorithms. A model known as a convolutional neural network is used in the majority of studies. CNN is a feed-forward neural network that evaluates visual pictures by processing input in a grid-like layout. CNN is used to identify and classify objects in a picture. In a convolution neural network, information may be extracted from an image thanks

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, October 2024

to several hidden layers. CNN is composed of four levels. Using a convolutional layer, which has several filters, is a method of extracting significant features from an image. ReLU then sets all negative pixels to zero using an element-by-element process. A down-sampling method called pooling lowers a feature map's dimensionality. The updated feature map is now sent via a pooling layer to produce a pooled feature map. Artificial neural networks known as recurrent neural networks are widely used in natural language processing and voice recognition. In order to address a layer's output, RNNs work on the principle of conserving that output and feeding it back into the input. Neural networks use attention as a tactic to simulate cognitive attention. With the premise that the network should focus more on that small but important component of the data, the impact enhances certain parts of the input data while decreasing others.

Facial Databases used in EmotionRecognition

One of the main performance metrics of deep learning is the training of the neural network using examples, and researchers may now use a number of face emotion recognition datasets to aid in this process. Each one differs from the others in terms of population, lighting, facial attitude, and the quantity and size of photos and videos. Below is a discussion of a few of them:

CK+: These are the most extensively used laboratory-controlled database for FER system assessment is the Extended (CK+) database. The CK+ dataset contains 593 video sequences from 123 different people, ranging in age from 18 to 50, gender, and ethnicity. Each image depicts a face transition from neural to a selected peak emotion, captured at 30 frames per second in 640x640pixel resolution.

JAFFE: The Japanese Female Facial Expression database is a lab-controlled image library with 213 examples of posed emotions from ten Japanese women. Each individual possesses three to four images representing each of the six fundamental facial expressions, as well as one neural image. Because there are few samples per subject or expression, the database is difficult to utilize.

MMI: The MMI database contains 326 sequences from 32 people and is lab-controlled. A total of 213 sequences are labelled with six fundamental expressions, and 205 sequencesare caught in frontal view. Furthermore, MI has more difficult conditions, such as substantial inter-personal variances due to respondents' non- uniform expressions and the fact that many of them wear accessories.

KDEF: The laboratory-controlled Karolinska directed emotional faces (KDEF) database consists of images of 70 actors with five different angles labelled with six basic facial expressions plus neutral. In addition to these commonly used datasets, others that aresuitable for training deep neural networks have emerged in the last two years.

Oulu-CASIA: There are 2,880 records in the Oulu-CASIA database. Only the final three peak frames and the first frame are usually used. From the 480 videos gathered by the VIS system. The 10-fold cross validation studies were then carried out using regular indoor lights.

FER2013. The Google image search API automatically collects FER 2013, a large-scale and unrestricted database, as shown in Fig

After rejecting incorrectly tagged frames and modifying the cropped region, all images were registered and resized to 48*48 pixels. FER2013 comprises 28,709 training images, 3,589 validation images, and 3,589 test images.

Fig. Sample images of different datasets.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, October 2024

III. DISCUSSION AND COMPARISON

Numerous studies on face emotion recognition have been thoroughly examined in terms of their design, datasets, and recognition rates. The thorough analysis of earlier studies in Table I is presented in the comparison table.

Evolutional Spatial-Temporal Networks are suggested in one study employing the CK+, Oulu-CASIA, and MMI datasets. The accuracy rates that were achieved are 98.50%, 86.25%, and 81.18%. The study makes use of the FER method, which is based on the WMDNN and CK+, Oulu-CASIA, and JAFFE databases. The corresponding average recognition accuracies are 0.970, 0.922, and 0.923. The CNN-based approach with edge detection is suggested in the next study. The FER 2013 database and LRF dataset were then carefully combined. This results in an 88.56% recognition rate. In terms of training speed, the training set outperforms the contrast method by a factor of 1.5. The CNN model is suggested in the study. After that, it is trained using the JAFFE and FER2013 datasets. Accuracy results were 97.07% and 94.12%, respectively. In the subsequent study, a CNN model is used, trained on the FER2013 dataset, and achieved an accuracy of 67%. Deep learning models are more robust and provide useful outcomes, as shown by earlier studies. Lastly, it is anticipated that FER 2013 achieves poorer accuracy than the others based on all of the datasets.

IV. CONCLUSION

Recognizing facial expressions is a difficult task. To this accomplish goal of recognition, variety of methodologies and procedures have been developed. This paper presented recent FER research, allowing us to keep up-to-date on the most recent discoveries in this field. We have discussed different architectures recently introduced by various researchers as well as various databases comprising spontaneous image collected in the real world and rest of them are created in laboratories, in order to have and accomplish accurate emotion recognition. The convolutional neural network is employed by the majority of themin all of the above studies since it provides better accuracy. Even though they achieve better accuracy, some of them fail to extract multiple features. Therefore, we need a hybrid model for using attention-based vision transformers with transfer learning.

REFERENCES

[1]. K. Zhang, Y. Huang, Y. Du and L. Wang, "Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks," in IEEE Transactions on Image Processing, vol. 26, pp. 4193-4203, Sept. 2017.

[2]. B. Yang, J. Cao, R. Ni and Y. Zhang, "Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images," in IEEE Access, vol. 6, pp. 4630-4640, 2018.

[3]. H. Zhang, A. Jolfaei and M. Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Networkand Image Edge Computing," in IEEE Access, vol. 7, pp. 159081-159089, 2019.

[4]. Garima Verma, Hemraj Verma, "Hybrid Deep Learning Model for Emotion Recognition Using Facial Expressions", Rev of Socionetwork Strat, vol.14, pp. 171- 180, 2020.

[5]. N. Zhou, R. Liang and W. Shi, "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," in IEEE Access, vol. 9, pp. 5573- 5584, 2021.

[6]. D. H. Kim, W. J. Baddar, J. Jang, et Y. M. Ro, "Multi- Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition ", IEEE Trans. Affect. Comput., vol. 10, pp. 223-236, avr. 2019.

[7]. Y. Li, J. Zeng, S. Shan, et X. Chen," Occlusion Aware Facial Expression Recognition Using CNN With AttentionMechanism ", IEEE Trans. Image Process., vol. 28, no 5, pp. 2439-2450, 2019.

[8]. Agrawal et N. Mittal, "Using CNN for facial expressionrecognition: a study of the effects of kernel size and number of filters on accuracy ", Vis. Comput., 2019.

[9]. D. Liang, H. Liang, Z. Yu, et Y. Zhang,"Deep convolutional BiLSTM fusion network for facial expression recognition ", Vis. Comput., vol.36, pp. 499- 508,2020.

[10]. Z. Yu, G. Liu, Q. Liu, et J. Deng, "Spatio-temporal convolutional features with nested LSTM for facial expression recognition ", Neurocomputing, vol. 317, pp. 50-57, Nov. 2018.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, October 2024

[11]. M. Mohammadpour, H. Khaliliardali, S. M. R. Hashemi, et M. M. AlyanNezhadi, "Facial emotion recognition using deep convolutional networks ", in 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), pp. 0017-0021,2017.

[12]. E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network,"6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 317-320, 2020.

[13]. Lu Lingling liu, "Human Face Expression Recognition Based on Deep Learning-Deep Convolutional Neural Network",International Conference on Smart Grid and Electrical Automation (ICSGEA),2019