Grenze International Journal of Engineering and Technology, Jan Issue

A Comparative Study on Facial Emotion Recognition using Deep Neural Networks

Asha S[1] and Dr. R Sundarrajan[2]
[1]Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education Srivilliputtur, Tamilnadu,India
Email: ashaaravind2016@gmail.com
[2]Department of Information Technology, Kalasalingam Academy of Research and Education Srivilliputtur, Tamilnadu, India
Email: r.sundarrajan@klu.ac.in

***Abstract*— Emotions are strongly associated with individuals' mood and personality. In the field of Human Computer Interaction, human face plays a very vital role. According to studies made by researchers' majority of the information conveyed through facial expressions than verbal communication. In day-to-day life, human expresses different types of feelings such as Happiness, Anger, Sadness, Fear, Disgust and Surprise which is considered as "Universal Emotions". It has always been difficult for computers to recognize human emotions. Thus, a substantial effort was made by the researchers to build the Facial Emotion Recognition system and which was considered as the best tool for recognizing emotions through facial expressions.**

In this paper, a detailed study on different methods that can be used in facial emotion recognition is done. For this study, the literature is collected from various reputable research published. This survey paper is based on the current approaches to face detection and feature extraction techniques for FER and also presented the real-time applications

***Index Terms*— deep networks, facial emotion recognition, convolutional neural network, recurrent network, deep belief network, machine learning.**

I. INTRODUCTION

Our daily lives are greatly affected by our emotions. The related facial expressions convey these feelings in a consistent manner. Following the years of research, there is now widespread agreement that there are seven basic facial expressions which describes the emotions of human. Happiness, Anger, Sadness, Fear, Disgust and Surprise are among the basic feelings [1][2]. The emotional feelings of a human can be inferred from verbal as well as nonverbal data collected by the various sensors, such as changes in face expression [3], voice tonality [4], and physiological signals [5]. The psychoanalyst Mehrabian [6] proposed a study on communication of messages between the humans. In that study, it reveals 55% of messages are conveyed by the expression of face, 38% by the supporting language like voice, speech and so on, and only 7% by oral languages. From unlocking our mobile phones to accessing our personal documents on our computers, facial recognition technology plays a vital role and parcel in our daily activities. Using different technologies for emotion recognition through facial expressions has recently generated a lot of interest. Combined with Machine Learning (ML) and Deep Learning (DL), Artificial Intelligence is developing FER softwares that will transform different industries in terms of safety, convenience,

Grenze ID: 01.GIJET.10.1.431

and smarter AI-backed outputs. AI backed facial emotion recognition has applications in different industries. Currently, the Facial Emotion Recognition system plays a main part of artificial intelligence which helps in recognition of the facial expressions. Many of the research works in emotion recognition system follows the framework of pattern recognition [7]. It consists of three phases: face detection, facial feature extraction, facial expression classification. Instead of performing the above-mentioned phases separately and reducing the time of recognition different algorithms for emotion recognition are introduced by the researchers which include feature extraction and classification based on physiological signals, facial expressions, body movements. In this paper, a comparative study of different approaches for real-time emotion recognition of basic emotions are presented.

II. General Idea For Facial Emotion Recognition

In the conventional approach, the facial image needs to be pre-processed for extracting the features from different feature locations on the face before the facial emotion is recognized. After the appropriate features have been collected, they are normalized and utilized to train and test the classifier for recognizing facial expressions. The flow diagram of FER is shown in fig 1. Here, colored or grayscale photographs from image databases were used to create the images. Pre-processing is carried after following image collection. To improve system performance, it is typically done prior to feature extraction. It entails scaling, smoothing, and cropping of acquired photos to produce processed facial images that convey a particular emotion.

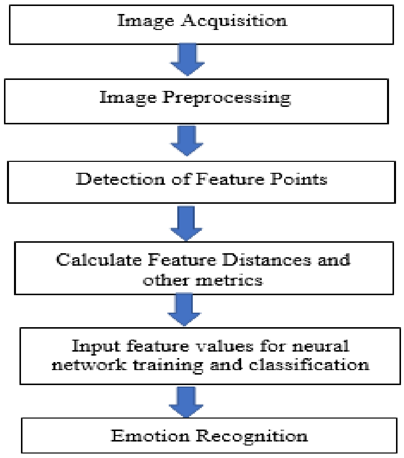

Fig 1: Flow diagram of conventional facial emotion recognition system.

After the pre-processing feature extraction is done. The feature points are defined in eyes, eyebrows, lip region etc of the detected faces. Based on the feature points extracted from the face region, feature distances are computed using distance equations like Manhattan distance. People have varying face sizes and forms, hence the identified faces and the distances between their features vary from person to person. Therefore, it is necessary to standardize these feature distances. The extracted features are then input to the system for training and classification, finally emotion recognition is done using various machine learning algorithms like SVM [8], PCA [9], Fisherfaces [10] etc.

III. Face Emotion Recognition Using Deep Learning

Recently deep learning techniques gained a wide popularity in the area of research and is used in a variety of applications [11]. Deep learning follows a layered designs with several transformations which are not linear and representations to try and it capture high-level abstractions. In Deep Neural Networks, the feature extraction, feature distance calculations etc can be done by the system itself compared to the conventional system. In the coming sections, several DL techniques that can be applied for Facial Emotion Recognition, were introduced.

A. Convolutional Neural Networks (CNN)

Convolution neural networks are recommended for recognition of facial expressions [12][13][14]. A convolution neural network can forecast the localized facial feature object identification and can also perform different deep neural network task and several forms of facial emotion recognition. In order to efficiently evaluate video images and lessen the effect of individual character variation on facial emotion recognition, the combined form of optical flow and deep neural network stacked sparse autoencoder, a novel technique was developed. K-nearest neighbors (KNN) and regions of interest (ROI) were coupled by some researchers to create the ROI-KNN approach, a quick and easy method for recognising facial expressions. Prior to ROI-KNN, deep neural networks methods reduced measurement error rates and dealt with data shortages [15].

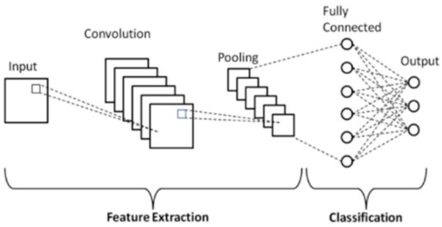

Fig 2. Model of Convolutional Neural Network

B. Deep belief network (DBN)

A generative graphical model called Deep Belief Network is a machine learning technology. Some researchers contend that deep neural networks may have also been involved. In the realm of image identification, the Deep Belief Neural Network is a popular feed-forward type Deep Neural Network. It contains several hidden layers. All of the layers in the DBN architecture were interconnected, but not the units inside each layer. A deep belief network develops replacement skills in its input data when trained on a set of examples without supervision. After that, the layers serve as trait detectors. Following this learning process, a deep belief network was built under close supervision to achieve data classification [16]. Facial expression recognition achieved notable results using DBN.

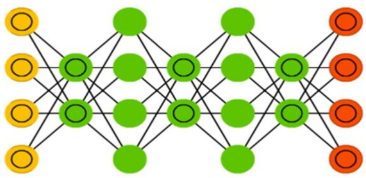

Fig 3. Deep belief network

C. Generative Adversarial Network (GAN)

(GAN) which was the powerful neural network class was first developed by Goodfellow et al. [17] in 2014. This model consists of mainly two models which automatically identify the patterns and learn the patterns input data. Generator and Discriminator are the models used in GAN. These two models compete each other, thereby increasing the result accuracy. Generator is a neural network which produces fake data and given to discriminator. The examples generated by the generator was the negative training examples for the discriminator. The main aim of the Generator is to enable the discriminator to differentiate its output as real. The next model Discriminator is also a type of neural network which produces a real output from the given fake data by the generator. During the discriminator's training process, the discriminator connects to two different loss functions and also the discriminator ignores the loss function of generator and just uses its loss function.

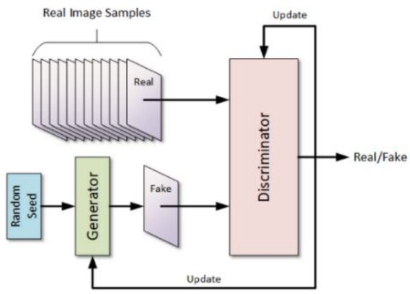

Fig 4. Generative Adversarial Network

D. Recurrent neural network (RNN)

Recurrent Neural Network is a connectionist model and is well suited for sequential data prediction with random lengths since it can collect temporal information. RNNs may train a deep neural network using a single feed-forward approach and have recurrent edges that can span adjacent time steps and share the same parameters throughout the step.

The RNN is trained using the traditional back propagation through time technique [18]. In order to solve the gradient vanishing and explosion difficulties that are frequently encountered when training RNNs, Hochreiter & Schmidhuber [19] developed the new model called long-short term memory (LSTM).

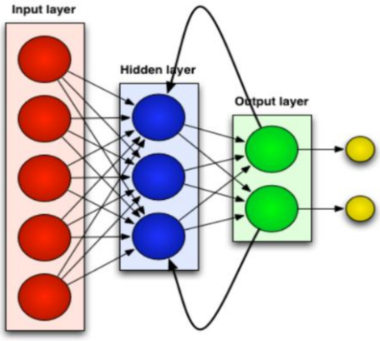

Fig 5. Model of Recurrent neural network

E. Deep autoencoder (DAE)

In order to discover effective codings for dimensionality reduction, DAE was first introduced in [20]. An autoencoder is a type of deep learning algorithm that made to take in information and change it into a different representation. Auto encoders play a vital part in image construction. An Autoencoder consists of three layers:

• Encoder
• Code
• Decoder

The input picture is transformed into a latent space representation by the encoder layer. It creates a compressed representation of the supplied image in a smaller dimension. In the compressed state, the original picture has been twisted. The decoder layer's compressed input is represented by the code layer. The decoder layer restores the image's original dimensions after decoding it. It is possible to reconstruct the decoded image from the depiction of the latent space, and this lossy reconstruction of the original image is done using the latent space representation.

One of the various DAE variations, the denoising autoencoder [21], restores the original and undisturbed input from partially damaged input data.

A directed graphical model with particular kinds of latent variables is used by the variational auto- encoder [22] and the convolutional autoencoder [23] to create sophisticated generative models, respectively.

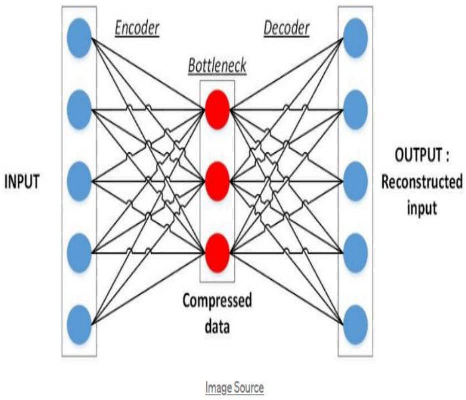

Fig 6. Deep autoencoder

F. Vision Transformer

Vision transformer is a pure transformer architectural model which is now widely used for emotion recognition. In this model the images are split into fixed-size patches. Then the images are correctly embedded each of the patches using linear embeddings as well as positional embeddings. And finally, this is given as the input to the transformer encoder for classification [24]. The transformer encoder contains a multi-head self-attention layer to concatenates all the attention results to correct dimensions. The multi-layer perceptron layer of the transformer contains the gaussian error linear unit. This layer contains the MLP head for image classification. The layer normalization is applied prior to each block for improving the overall training time and performance.

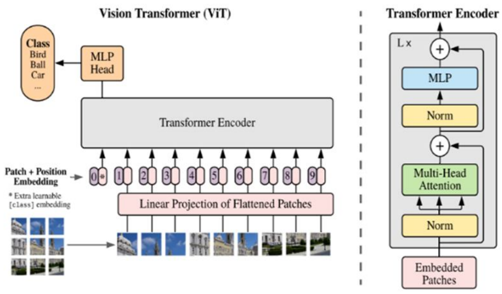

Fig 7. Vision transformer model

IV. Comparison Of Various Deep Learning Technique For Fer

TABLE I. Comparison Table Of Deep Learning Techniques

TECHNIQUES	ADVANTAGES	DISADVANTAGES
Convolutional Neural Networks (CNN)	1. Exploit the spatial hierarchy of the features present in the image. 2. Capture the patterns regardless of their position or orientation in an image.	1. Difficulty in handling temporal dynamics 2. sensitive to variation in data distribution
Deep belief network (DBN)	1. Allows the network to learn complex and abstract features from raw input data 2. used for transfer learning	1. expensive to train because it has complex data models 2. limited scalability
Generative Adversarial Network (GAN)	1. Can learn to generate realistic emotional instances by capturing the underlying data distributions. 2. Can be used for domain adaptations.	1. lack of interpretability 2. computationally expensive and requires significant computational resources to train and deploy
Recurrent neural network (RNN)	1.Capture temporal dependencies and can effectively modal sequential data. 2.Can be enhanced with attention mechanism.	1. Challenge of capturing long term dependencies. 2. Faces issues like Exploding or Gradient Vanishing.
Deep Auto Encoder (DAE)	1. Are capable of learning complex nonlinear representations of input data. 2. Can be trained in an unsupervised manner	1. may eliminate the important information in the input data. 2.Lack of interpretability
Vision Transformer (ViT)	1. Highly scalable and can be applied to images of varying size. 2. Also consider relationship between all image regions.	1. requires more computational resources and training data 2. limited ability to capture spatial information

V. Conclusion

In the field of digital image processing and human computer interaction, identifying facial expressions for emotion recognition has been difficult. Over the past 20 years, this area has been the subject of numerous research projects.

Facial expression detection has received a lot of interest recently because to the multiple uses and implementations it has seen in a range of industries. In conventional method of emotion recognition, pre-processing, feature extraction, normalization etc has to be done before training and testing the network. It is necessary to increase the accuracy of feature extraction algorithms as well as take use of facial actions in order to measure the intensity of emotions in conventional method. However, using deep learning techniques, feature extraction, feature distance calculations etc are done by the deep neural network. This paper presented different deep neural network study and their comparisons which will be beneficial for upcoming research and development.

REFERENCES

[1] Matsumoto, D. et al. Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience. Asian Journal of Social Psychology (4), pp. 113-131 (2001)

[2] Ekman, P. & Keltner, D. (1997). Universal facial expressions of emotion: An old controversy and new findings. In Segerstråle, U. C. & Molnár, P. (Eds.), Nonverbal communication: Where nature meets culture (pp. 27-46). Mahwah, NJ: Lawrence Erlbaum Associates.

[3] E. Sariyanidi et al., "Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition ", IEEE Trans. Pattern Anal. Mach. Intell., oct. 2014, doi: 10.1109/TPAMI.2014.2366127.

[4] C.-N. Anagnostopoulos et al., "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011", Artif. Intell. Rev., vol. 43, no 2, p. 155-177, févr. 2015, doi: 10.1007/s10462-012-9368-5.

[5] L. Shu et al., "A Review of Emotion Recognition Using Physiological Signals", Sensors, vol. 18, no 7, p. 2074, juill. 2018, doi: 10.3390/s18072074.

[6] Mehrabian, A.: 'Communication without words', Psychol. Today., 1968, 2, (4), pp. 53–56.

[7] Shih, P., & Liu, C. (2006). Face detection using discriminating feature analysis and support vector machine.

[8] Tian, Y.L., Kanade, T., Cohn, J.F.: 'Facial expression analysis', in 'Handbook of face recognition' (Springer, New York, NY, 2005), pp. 247–275.

[9] Ajit P. Gosavi, S. R. Khot,' Facial Expression Recognition Using Principal Component Analysis', International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013

[10] Hyung-Ji Lee et.al., Face recognition using Fisherface algorithm and elastic graph matching, Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205), 07 August 2002.

[11] L. Deng, D. Yu et al., "Deep learning: methods and applications," Foundations and Trends R in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014.

[12] A. Fathallah et al., "Facial Expression Recognition via Deep Learning - IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) ", 2017.

[13] L. Nwosu et al., "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts - IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech) ", 2017.

[14] H. Jung et al., "Development of deep learning-based facial expression recognition system - 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) ", 2015.

[15] Xiao, S et al., " Facial expression recognition using ROI-KNN deep convolutional neural networks". Acta Automatica Sinica 42(6), 883–891 (2016)

[16] Hassan, Mohammad & Alam, Md Golam Rabiul & Uddin, Md. Zia & Huda, Shamsul & Almogren, Ahmad & Fortino, Giancarlo. (2018). Human Emotion Recognition Using Deep Belief Network Architecture. Information Fusion. 51. 10.1016/j.inffus.2018.10.009.

[17] Goodfellow et al., "Generative adversarial nets," in Advances in neural information processing systems, 2014, pp. 2672–2680.

[18] P. J. Werbos, "Backpropagation through time: what it does and how to do it," Proceedings of the IEEE, vol. 78, no. 10, pp. 1550–1560, 1990.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[20] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," science, vol. 313, no. 5786, pp. 504–507, 2006.

[21] P. Vincent et al., "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," Journal of Machine Learning Research, vol. 11, no. Dec, pp. 3371–3408, 2010.

[22] Q. V. Le, "Building high-level features using large scale unsupervised learning," in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013, pp. 8595–8598.

[23] S. Rifai et al., "Contractive auto-encoders: Explicit invariance during feature extraction," in Proceedings of the 28th International Conference on International Conference on Machine Learning. Omnipress, 2011, pp. 833–840.

[24] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Representations, 2021