JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FACIAL EMOTION DETECTION USING DEEP LEARNING

S Penchala Reddy[1], J V S Sai Varshini[2], A Mounika[3], T Vineela[4], V Sunil[5], O Prameela[6]

[1]Dept of ECE, PBR Visvodaya Institute of Technology and Science, India
[2]Dept of ECE, PBR Visvodaya Institute of Technology and Science, India
[3]Dept of ECE, PBR Visvodaya Institute of Technology and Science, India
[4]Dept of ECE, PBR Visvodaya Institute of Technology and Science, India
[5]Dept of ECE, PBRVisvodaya Institute of Technology and Science, India
[6]Dept of ECE, PBR Visvodaya Institute of Technology and Science, India

Abstract: Human Emotion detection from image is one of the most powerful and challenging research tasks in social communication. Deep learning (DL) based emotion detection gives performance better than traditional methods with image processing. This paper presents the design of an artificial intelligence (AI) system capable of emotion detection through facial expressions. It discusses about the procedure of emotion detection, which includes basically three main steps: face detection, features extraction, and emotion classification. Facial emotion recognition (FER) is a crucial component in the development of human-computer interaction systems and artificial intelligence applications aimed at understanding human emotions. This paper presents a comprehensive review of recent advancements, methodologies, challenges, and future directions in facial emotion recognition, it discusses the significance of FER in various domains such as psychology, human-computer interaction, and affective computing. Next, it provides an overview of the underlying techniques employed in FER, including traditional machine learning approaches and deep learning architectures. Furthermore, we examine the challenges associated with FER, such as variations in facial expressions, occlusions, and dataset biases. Highlight the importance of ethical considerations and privacy concerns in FER applications. This project proposed a convolutional neural network (CNN) based deep learning architecture for emotion detection from images. The performance of the proposed method is evaluated using two datasets Facial emotion recognition challenge (FERC-2013) and Japanese female facial emotion (JAFFE). The accuracies achieved with proposed better percentage

***Index Terms* - Face detection, Features extraction, and Emotion classification.**

I. INTRODUCTION

Emotion is a spontaneous mental state that lasts for a few seconds or minutes. It solely indicates the current state of mind of a person and not their person's current emotional state. The interaction between two parties is impaired when one of the parties is not able to recognize or understand the other's emotions. This applies to human-human interaction but also to human-computer interaction. Affective computing is a field that attempts to enhance the interactions between human and machine by developing artificial systems that are able to recognize human emotions and react according to them. For the interface between humans and machines to be natural, machines should have the capability of recognizing human emotions. This could be applied in sociable robotics where robots are able to support people in simple functions like distributing food or sweeping of the house. These human-machine interactions right now are still negligible and if the robot knew more about the person with whom it was interacting, it would be better. Knowing what people feel when interacting with machines would allow the said machines to adapt and improve their interaction by having a suitable reaction to people's feelings. Taking the example of humanoid robots that offer services to people, the robot-human interaction would significantly improve if these robots were able to regulate their responses to people's recent emotions. Face and speech recognition can even be used to remotely monitor patients. However, recent research shows that body language comprises a substantial quantity of intuitive information. Body language can be conveyed in distinctive manners, from facial expressions to body posture, eye movement, gestures, touch or even personal space. The finest modalities to utilize and how to merge them in achieving the finest recognition rate of human emotions is still an object of research.

II. LITERATURE SURVEY

Pujol et al. developed a fuzzy approach for recognizing skin in color photographs, based on the assumption that each color tone is a fuzzy set. For the creation of our fuzzy design, we used the RGB, HSV, and YCbCr color systems (where Y is the brightness and Cb,Cr are the chroma elements). Therefore, To compute all the parameters required for the fuzzy systems, a fuzzy three-partition entropy methodology is adopted, and also a face identification technique is devised to confirm the segmentation findings. Sun et al. recommended a novel deep learning-centered face identification strategy that produced advanced detection results on widely known FDDB face identification benchmark. Specifically, the authors improved the advanced Faster RCNN model by merging various schemes, comprising concatenation of features, negative hard mining, multi-level training, model pre-training, and critical parameter standardization. Deep learning application which is based on deep convolutional neural networks (DCNN) has observed a significant achievement in the subject of the face recognition in recent times. But, the identification of smallscaled faces is one of the major open challenges. The deepness of the CNN may quickly shrink the probable feature map for little faces, and most scale invariant detection techniques can't process faces smaller than 15X15 pixels. Wu et al. proposed a Faster R-convolutional neural networks based different scales face detector (DSFD) to overcome this problem. The proposed network can enhance the accuracy of face recognition while working as an actual-time Quicker R-CNN.

III. Existing Method

Humans understand that emotions are not easy to quantify or replicate artificially from this complex set of actions. Many researchers use their version of emotion definitions and assumptions. This makes research in human facial emotions troublesome because all the studies that have been done have significant variance in them and do not draw a generalized conclusion. Although all humans have naturally occurring sets of emotions that can be perceived even cross-culturally, this is also mentioned in the Discrete Emotion Theory, which says that such emotions are distinguishable by an individual’s features . Ekman claimed that these emotions are perceived by humans not only culturally but also universally. His proposed model suggested that emotions are categorized into Fear, Happiness, Sad, Surprise, Disgust, and Anger. These categorical emotions are classified using facial and vocal data, which allows them to perform a better human FER efficiently. Alternatively, there is another proposed model by Plutchik , who claims that there are more basic emotions (i.e., joy, fear, anger, sadness, trust, disgust, surprise, and anticipation).

Different datasets use different combinations of emotions for research. For example, very few kids’ datasets have ‘angry’ emotions. Those using it have recorded by posing. Recording the angry emotion from spontaneous expressions is difficult. But for Adult datasets, it is straightforward to pose for an angry emotion. Datasets like ‘RML’ have recorded the emotions in a controlled environment with good lighting conditions. Most datasets cut movie clips or tv shows and use them for classification. The category of emotions differs from dataset to dataset.

IV. Disadvantages of Existing Method

Multiple datasets that include different populations and recording environment variations help the researchers design a more robust deep learning system for emotion detection. In this section, the authors have discussed the datasets available for emotion detection, used by researchers worldwide for FER system evaluation. They have divided this section into two parts - Kids and Adults. This is also illustrated in Figure 12. provides a visual of different kids and adult datasets used for emotion detection using video and audio.

However, when these different categories of a dataset are compared, authors comment that there is a scarcity of kids’ datasets as compared to FER, so they suggest creating a new novel dataset that is balanced and has a high quality of data in the kids’ category and set up a new benchmark accuracy on it.

V. Proposed Method

The suggested facial recognition system's general block design is shown in Figure 3.1. Firstly, the videos are subjected to frame separation then each of those images are subject to face detection process for discovery the suitable face in the photo which aids to emphasis on the relevant portion for identifying the face emotional identification. After identifying each image's face, the following step has been taken to mine the face's features. Since these retrieved characteristics have a greater number of properties, the computing complexity rises to a upper level. To combat this issue, duplicated or useless facts are ignored with the use of feature optimization methods, which also aids in the discovery of the crucial set of features. When you've found the best set of features, give them to the classifier to find the face emotion.

Haar Cascade is a feature-based object detection algorithm to detect objects from images. A cascade function is trained on lots of positive and negative images for detection. Haar cascade uses the cascade function and cascading window. It tries to calculate features for every window and classify positive and negative. If the window could be a part of an object, then positive, else, negative.

The face of image is recognized with the aid of popular Viola and Jones face recognition technique.

Feature extraction is also applied to translate images from image domain to feature domain that assists in removing duplicated data and reduction in computation time. Local feature extraction as local binary pattern and HOG (global feature extraction) are coupled to obtain superior feature extraction.

Post feature extraction and optimization, classification is performed by SVM, which provides an effective method of extracting features as well as a set of criteria for categorization. A distinct hyper-plane represents SVM, which is a discriminative classification algorithm. The SVM classifier is extensively utilized in a diversity of domains, including bioinformatics, signal processing, and computer vision, because of its excellent precision and ability to process data with multiple dimensions. SVM excels in handling two-class problems, which are linked to Vapnik Chervonenkis theories and structure rules.

SIFT is a technique for retrieving and detecting consistent local feature descriptors that are independent of picture contrast, rotation, and scale. Recently, It has been updated to provide better picture retrieval.

A CNN typically contains several non-linear layers for feature extraction and a single MLP classifier that inputs in the attributes and executes the categorization. In most of the CNNs, there are three types of layers: Convolutional, pooling, and fully connected.

This section presents paradigm for facial emotion categorization by deep CN design. The anticipated design is fully CNN based architecture. This design consists of nine convolutional layers which includes ReLU, Batch standardization and average.

We use 3 coalesced CNN models for face detection, which allows us to recognize faces quicker and more accurately. Furthermore, the faces are linked on the basis of nose, mouth and eye features. On this foundation, an image pyramid is built, which is 54 then sent to the three-stage network for learning. The network generates potential areas in the primary step, while the final two stages refine the discovered landmarks. In the third phase, the detection result is achieved. We use a Large-margin SoftMax loss to minimize the learning error, which also assists to limit the over-fitting issue. Furthermore, the Softmax with a large margin enhances intra-class compaction.

RESEARCH METHODOLOGY

In next phase, we concentrate on the facial emotion recognition system, in which we make two advancements to the feature extraction and classification procedure. Mainly, SIFT (Scale-invariant feature transform) features are mined in feature abstraction procedure. Anyway, several crucial points which are gathered during SIFT retrieval seem to provide duplicated data. So, the Whale Optimization Algorithm (WOA) technique is used to optimize the important points. Consequently, the retrieved features from the elected key points are magnified by weight factor that should be enhanced by algorithm. The DBN is provided the best features for face emotion categorization, where WOA and feature weight are used to further maximize the number of hidden neurons. Eventually, with the aid of WideResNet and Caffe prototypes, we utilized CNN and introduced an integrated model for emotion classification, age, and gender estimate.

IV. RESULTS AND DISCUSSION

The Challenges in Representation Learning was held during the ICML 2013 conference, the Facial Expression Recognition 2013 (FER-2013) database was presented. Faces were dynamically recorded in the database, which was built using the Google picture exploration API. All of the other six fundamental expressions, and also the neutral, are assigned. 63 The resultant data includes approximately 36000 photos, the majority of which are taken in natural environments. This dataset contains total 35887 number of images where each category has 4953, 547, 5121, 8989, 6198, 6077, and 4002 for Anger, Disgust, Fear, Happiness, Neutral, Sadness, and Surprised correspondingly

Layer Name	Layer Type	Layer description	Outcome dimension
Data	Input	Random crop	3 × 277 × 277
Conv 1	Conv	96 kernels, ReLU	96 × 55 × 55
Pool	Max Pool	96 kernels with size 3	96 × 27 × 27
Conv 2	Conv	256 kernels with size 5	256 × 27 × 27
Pool	Max Pool	Pooling with size 3	256 × 13 × 13
Conv 3	Conv	384 kernels, ReLU	384 × 13 × 13
Conv 4	Conv	384 kernels, ReLU	384 × 13 × 13
Conv 5	Conv	256 kernels, ReLU	256 × 13 × 13
Pool	Pool	Pooling with size 3	256 × 6 × 6
Fc	Fully Connected	ReLU and dropout	4096 × 1 × 1
Fc	Fully Connected	ReLU and dropout	4096 × 1 × 1
Fc		Classification	2 × 1 × 1

Table 6.1 *Depiction of each layer employed for feature extraction, fine-tuning forgender classification.*

System	*Performance in CK+ dataset*
Anticipated OLPP based detection	*98.79%*
PCA+Gabor -1 [160]	*93.00%*
PCA+LBP - 2 [160]	*96.83%*

Table.6.2 *Comparative study of accuracy*

Methods	PSO	GA	FF	ABC	GWO	WOA
Accuracy	0.922449	0.926531	0.946939	0.942857	0.946939	0.95102
Specificity	0.954762	0.957143	0.969048	0.966667	0.969048	0.971429
Sensitivity	0.728571	0.742857	0.814286	0.8	0.814286	0.828571
FPR	0.045238	0.042857	0.030952	0.033333	0.030952	0.028571
Precision	0.728571	0.742857	0.814286	0.8	0.814286	0.828571
NPV	0.954762	0.957143	0.969048	0.966667	0.969048	0.971429
FNR	0.271429	0.257143	0.185714	0.2	0.185714	0.171429
F1-score	0.728571	0.742857	0.814286	0.8	0.814286	0.828571
FDR	0.271429	0.257143	0.185714	0.2	0.185714	0.171429
MCC	0.683333	0.7	0.783333	0.766667	0.783333	0.8

Table 6.3 *Comparative analysis using whale optimization*

REFERENCES

[1]. Noroozi, F., Marjanovic, M., Njegus, A., Escalera, S., & Anbarjafari, G. (2017). Audio-visual emotion recognition in video clips. IEEE Transactions on Affective Computing, 10(1), 60-75.

[2]. Löffler, D., Schmidt, N., & Tscharn, R. (2018, February). Multimodal expression of artificial emotion in social robots using color, motion and sound. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 334-343).

[3]. Augello, A., Dignum, F., Gentile, M., Infantino, I., Maniscalco, U., Pilato, G., & Vella, F. (2018). A social practice-oriented signs detection for human- humanoid interaction. Biologically inspired cognitive architectures, 25, 8-16.

[4]. Noroozi, F., Kaminska, D., Sapinski, T., & Anbarjafari, G. (2017). Supervised vocal-based emotion recognition using multiclass support vector machine, random forests, and adaboost. Journal of the Audio Engineering Society, 65(7/8), 562-572.

[5]. Zhang, Y., Wang, Z. R., & Du, J. (2019, July). Deep fusion: An attention guided factorized bilinear pooling for audio-video emotion recognition. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1- 8). IEEE.

[6]. Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. Frontiers in psychology, 6, 711.

[7]. Ekman, P. (2009). Darwin's contributions to our understanding of emotionalexpressions. Philosophical Transactions of the Royal Society B: BiologicalSciences, 364(1535), 3449-3451.

[8]. Kortli, Y., Jridi, M., Merzougui, M., Alasiry, A., & Atri, M. (2020, December). Comparative Study of Face Recognition Approaches. In 2020 4th International Conference on Advanced Systems and EmergentTechnologies (IC_ASET) (pp. 300-305). IEEE.

[9]. Pasandi, M. E. M. (2014). *Face, Age and Gender Recognition using Local Descriptors* (Doctoral dissertation, University of Ottawa).

[10]. Ekmekji, A. (2016). Convolutional neural networks for age and gender classification. Stanford University.

[11]. Cao, K., & Jain, A. K. (2018). Automated latent fingerprint recognition. *IEEE transactions on pattern analysis and machine intelligence*, *41*(4), 788-800.

[12]. Giorgi, R., Bettin, N., Ermini, S., Montefoschi, F., & Rizzo, A. (2019, June).An Iris+ Voice Recognition System for a Smart Doorbell. In *2019 8th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1-4). IEEE.

[13]. Hájek, J., & Drahanský, M. (2019). Recognition-based on eye biometrics: Iris and retina. In *Biometric-Based Physical and Cybersecurity Systems* (pp.37-102). Springer, Cham.

[14]. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., & Ortega-Garcia, J. (2018). Exploring recurrent neural networks for on-line handwritten signature biometrics. *Ieee Access*, *6*, 5128-5138.

[15]. Ishii, L. E., Nellis, J. C., Boahene, K. D., Byrne, P., & Ishii, M. (2018). Theimportance and psychology of facial expression. *Otolaryngol Clin North Am*, *51*(6), 1011-1017.

[16]. Wójcik, W., Gromaszek, K., & Junisbekov, M. (2016). Face recognition: Issues, methods and alternative applications. Face Recognition-Semisupervised Classification, Subspace Projection and Evaluation Methods, 7-28.

[17]. Skalski, P. D., Neuendorf, K. A., & Cajigas, J. A. (2017). Content analysis in the interactive media age. The content analysis guidebook, 2, 201-42.

[18]. Ranjan, R., Patel, V. M., & Chellappa, R. (2017). Hyperface: A deep multi- task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE transactions on pattern analysis and machine intelligence*, *41*(1), 121-135.

[19]. Wang, L., Yu, X., Bourlai, T., & Metaxas, D. N. (2019). A coupled encoder decoder network for joint face detection and landmark localization.Image and Vision Computing, 87, 37-46.

[20]. Lu, X., Duan, X., Mao, X., Li, Y., & Zhang, X. (2017). Feature extraction and fusion using deep convolutional neural networks for face detection. Mathematical Problems in Engineering, 2017.