International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Learning Effective Video Features for Facial Expression Recognition via Hybrid Deep Learning

A. Rajesh kumar, G. Divya

Abstract*: *Facial Expression Recognition is one of the recent trends to detect human expression in streaming video sequences. To identify emotions of video like sad, happy or angry. In this paper, the proposed method employs two individual deep convolution neural networks (CNNs), including a permanent CNN processing of static facial images and a temporary CN network processing of optical flow images, to separately learn high-level spatial and temporal characteristics on the separated video segments. Such two CNNs are fine tuned from a pre-trained CNN model to target video facial expression datasets. The spatial and temporal characteristics obtained at the segment level are then incorporated into a deep fusion network built with a model of deep belief network (DBN). This deep fusion network is used to learn spatiotemporal discriminative features together

***Keywords* : Machine learning algorithms, Neural networks,**

I. INTRODUCTION

In our day-to-day communications, a primary method of shipping human feelings assumes a vital task. Outward appearance recognition is a mind-boggling and fascinating problem, discovering its applications in driver safety, social insurance, human-PC relation, etc. Outward appearance identification has gained significant attention among scientists in the PC vision field due to its wide range of uses. While different novel theories have been proposed as of late, perceiving external appearance with high accuracy and speed stays testing due to the multifaceted nature and changeability of external appearances

1. Instead of utilizing the entire face locale, three sorts of dynamic districts are applied to characterize outward appearances. A strategy to look through advanced dynamic areas is proposed by similitude of dynamic districts.
2. A choice level combination structure is proposed, which is useful to expand the exactness of outward appearance acknowledgment

Manuscript published on January 30, 2020.
* Correspondence Author
A. Rajeshkumar*, Department of Computer Science and Engineering, , Saveetha school of engineering , Saveetha Institute of Medical and Technical Sciences, Chennai, India.
G. Divya, Assistant Professor, Department of Computer Science and Engineering, Saveetha school of engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.

II. LITERATRE SURVEY

1.Farahani, Fatemeh Shahrabi, Mansour Sheikhan, and Ali Farrokhi. "A fuzzy approach for facial emotion recognition."
Feelings assume a basic job in individuals collaboration. The capacity to comprehend human feelings by PC is alluring in a few applications, particularly by watching outward appearances. This paper displays another fluffy based technique for feeling acknowledgment from eyes and mouth includes in various ages. The technique recognizes eyes and mouth from blend of various shading spaces. Four parameters comprise of educational, mouth opening, enlightening/width proportion and mouth width are chosen for fluffy investigation. Utilizing Mamdani-type suggestion relations, facial ascribes and their mapping to feeling space are encoded. Applying the technique on Ebner's outward appearance database demonstrates 78.8% exactness.

2. Oh, Byung-Hun, and Kwang-Seok Hong. "A study on facial components detection method for face-based emotion recognition."
In this paper, we propose a facial parts discovery strategy for face-based feeling acknowledgment. To begin with, the face area is recognized through a blend of the YCbCr shading model and the Maximum Morphological Gradient Combination (MMGC) picture. The scanning locale for the facial segment recognition is restricted to the identified face area. The facial parts are recognized utilizing the histogram strategy, the mass naming technique, and the MMGC picture. Additionally, a labeling activity for dependable assessment was performed. The location execution was assessed utilizing a labeling coordinate. In the trial result, the facial parts discovery demonstrated 81.4% exactness and a 53.6 ms recognition time. We show the unwavering quality of the proposed assessment strategy through various investigations utilizing the labeling technique.

3.Reney, Dolly, and Neeta Tripathi. "An Efficient Method to Face and Emotion Detection."
.Face discovery and feeling determination is the one of the present point in the security field which gives answer for different difficulties. Next to customary difficulties in caught facial pictures under uncontrolled settings, for example, changing stances, diverse lighting and appearances for face acknowledgment and distinctive sound frequencies for feeling acknowledgment. For the any face and feeling location framework database is the most significant part for the examination of the face highlights and sound Mel recurrence segments.

For database creation highlights of the face are determined and these highlights are store in the database. This database is then use for the assessment of the face and feeling by utilizing various calculations.

4. N. Cristiana, T. Shawe, An Introduction to Support Vector Machine,

This is the main thorough prologue to Support Vector Machines (SVMs), another age learning framework dependent on ongoing advances in factual learning hypothesis. SVMs convey best in class execution in true applications, for example, content categorisation, written by hand character acknowledgment, picture arrangement, biosequences examination, and so on., and are currently settled as one of the standard devices for AI and information mining. Understudies will discover the book both invigorating and open, while experts will be guided easily through the material required for a decent handle of the hypothesis and its applications. The ideas are presented bit by bit in available and independent stages, while the introduction is thorough and careful.

5.Pantic, Maja, and Leon JM Rothkrantz. "Toward an affect-sensitive multimodal human-computer interaction"

The use of instructive information mining (EDM) methods to intelligent learning programming is progressively being utilized to widen the scope of develops ordinarily fused in understudy models, moving from conventional appraisal ofstudent information to the evaluation of commitment, influence, technique, and metacognition. Scientists are additionally expanding the scope of situations inside which these develops are surveyed. In this investigation, we create sans sensor influence identification for EcoMUVE, a vivid multi-client virtual condition that shows center school understudies about loss in environments. In this investigation, models were built for five diverse instructively applicable emotional states (fatigue, disarray, charm, drew in fixation, and disappointment).

III. EXISTING SYSTEM

In Existing System Edge Detection algorithm is used which really only serves the edges of the rice we are going to dissect. They used Edge Detection Technique in the existing framework to detect feeling in the face. The primary drawback of this system is that the boundaries of actions are separated due to this internal area, which demonstrates that the feeling can not be segmented. Equally high over lapping of pictures will build the SNR ratio, thus diminishing the accuracy.

IV. PROPOSED SYSTEM

The objective of emotion recognition is identifying emotions of a human. The emotion can be captured either from face or from verbal communication. In this work we focus on identifying human emotion from facial expressions. Facial emotion recognition is one of the useful task and can be used as a base for many real-time applications. It can be used as a part of many interesting and useful applications like Monitoring security, treating patients in medical field, marketing research, E-learning etc.;. We humans can easily identify the emotion of other humans without any effort. Automatic detection of emotion of a human face is important due to its use in real-time applications. The recent advance in GPU has taken many applications like face recognition, hand written digit recognition and object recognition to the next level. The representation of features plays a major role in the success of any algorithm for machine learning. After finding incredible performance with deep learning models, we recommend using the deep convolutionary features to better represent the given image rather than using traditional handcrafted features. The downside of deep learning models is that to achieve better performance, they require large data sets.

V. METHODOLOGY

The half-and-half-deep learning model contains two individual data sources, i.e. the sound system with a CNN model handling sound sign, and the visual system with a 3D-CNN model that prepares visual information. In a hybrid structure operated with a DBN model, the yields of the fully linked layers of these two structures are combine.

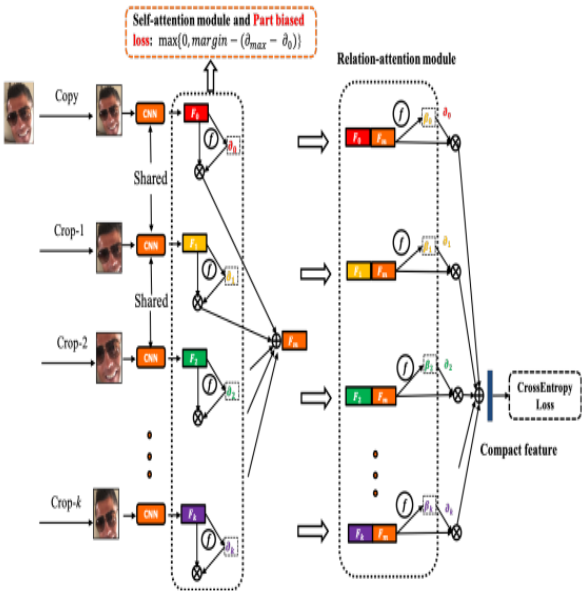

Figure 1: System Architecture

A.Audio Input Generation:

It is known that the 1-D spectrogram, represented by the squared magnitude of the time-varying spectral characteristics of audio signals, contains tremendous low-level acoustic information related to the speaker's emotion expression, such as energy, pitch, formants, and so on [50]. However, CNNs are commonly used to process 2-D or 3-D images in vision tasks [23]. To leverage the available CNN models and make our deep model initialization easier, it is hence intuitive to transform the 1-D spectrogram into a 2-D array as the input of CNN.

B.Visual Input Generation:

After splitting the video sample into segments, we use the video segments as the 3D-CNN input. For each frame in the video segment, we run face detection, estimate the eye distance, and finally crop a RGB face image of size 150 × 110 × 3, as done.

In detail, we employ the robust real-time face detector presented by Viola and Jones to perform automatic face detection on each frame. From the results of automatic face detection, the centers of two eyes can be located in a typical up-right face.

C.Emotion Classification

Once the fusion network has been equipped, a 2048-D joint feature representation on each audio-visual segment can be determined. Since each audio-visual video sample has a different number of segments, average pooling is applied from each video sample to all segment features to form the global fixed-length representation of the video feature. Our experiments compared average pooling and max pooling, and found better performance of average pooling.

VI. RESULT:

Convolutional Neural Network technique is applied in the tensor flow. Tensor flow is an API to detect the emotions in streaming video and identifying a particular sequence with a frame. Figure 2 show different emotions of a video with various timing of an video . CNN yields 87.9 % accuracy of the detection when compared to other previous approaches

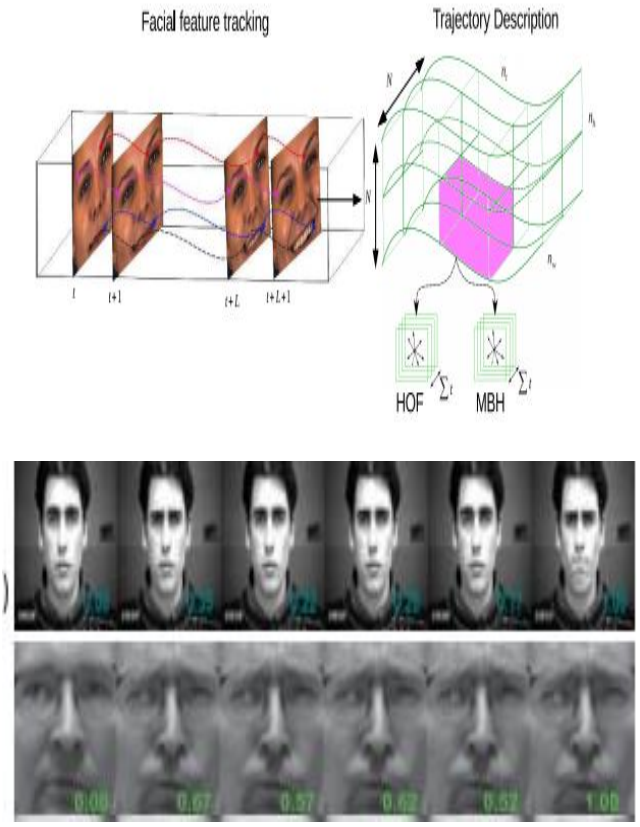

Figure 2: Emotions detection

VII. CONCLUSION:

This project proposes a hybrid deep learning model, which consists of the spatial CNN network, the temporal CNN network, and the DBN fusion network, to apply for FER in video sequences. We implement our proposed method in two stages. (1) We employ the existing VGG16 model pre-tained on ImageNet data to individually fine-tune the spatial CNN network and the temporal CNN network on target video-based facial expression data.

(2) To fuse the learned spatio-temporal CNN characteristics deeply, we train a deep DBN model to learn spatio-temporal discriminative characteristics together. Test findings on three datasets of public video-based facial expression, i.e. BAUM-1s RML and MMI, show the advantages of our proposed method. We will be expanding our research to practical applications in the future.

REFERENCES

1. Farahani, Fatemeh Shahrabi, Mansour Sheikhan, and Ali Farrokhi. "A fuzzy approach for facial emotion recognition." 2013 13th Iranian Conference on Fuzzy Systems (IFSC). IEEE, 2013
2. Oh, Byung-Hun, and Kwang-Seok Hong. "A study on facial components detection method for face-based emotion recognition." 2014 International Conference on Audio, Language and Image Processing. IEEE, 2014
3. Reney, Dolly, and Neeta Tripathi. "An Efficient Method to Face and Emotion Detection." 2015 Fifth International Conference on Communication Systems and Network Technologies. IEEE, 2015.
4. N. Cristiana, T. Shawe, An Introduction to Support Vector Machine, Cambridge University Press, 2000.
5. Pantic, Maja, and Leon JM Rothkrantz. "Toward an affect-sensitive multimodal human-computer interaction" Proceedings of the IEEE 91.9 (2003): 1370-1390.
6. Adeyanju, Ibrahim A., Elijah O. Omidiora, and Omobolaji F. Oyedokun. "Performance evaluation of different support vector machine kernels for face emotion recognition." 2015 SAI Intelligent Systems Conference (IntelliSys). IEEE, 2015.
7. Liu, Mengyi, et al. "Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild." Proceedings of the 16th International Conference on multimodal interaction. ACM, 2014.
8. Salunke, Vibha V., and C. G. Patil. "A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks." 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). IEEE, 2017.
9. Lee, Hyeon-Jung, and Kwang-Seok Hong. "A study on emotion recognition method and its application using face image." 2017 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2017.
10. Gao, Yongsheng, et al. "Facial expression recognition from line-based

AUTHORS PROFILE

A. Rajeshkumar*,Department of Computer Science and Engineering, , Saveetha school of engineering , Saveetha Institute of Medical and Technical Sciences, Chennai, India. Email:rajeshaaa143@gmail.com.
This publication is based on my final year project in B.E

G. Divya, Assistant Professor, Department of Computer Science and Engineering, Saveetha school of engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.
Email: mailtodivya16@gmail.com . My guide for the project.