International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Facial Expression Recognition using Attentional Convolutional Network

Bala Sai Mani Kanta[1], Gaganpal Singh[2], Rounak Mitra[3], Dr. Arun Nehru[4]

[1,2,3]Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India
[4]Assistant Professor, Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

---

Abstract - Recognition of facial expression has been an important area of study over the past few decades and it is still difficult due to the high variability in intra-class Standard solutions to this problem are based on side-crafted features such as SIFT, HOG and LBP accompanied by a classifier trained in an image or video repository that performs relatively well on image datasets collected in a controlled environment, but fail to perform as well on more demanding datasets with greater variability of image and partial faces. Several studies in recent years have suggested an end-to-end system for recognition of facial expression, using deep learning models. Given these works performing better, there still seems to be a great scope for change. In this thesis, we propose a deep learning approach focused on attentive convolutionary network, capable of focusing on important sections of the face and achieving substantial improvement on multiple datasets, like FER-2013, CK+, FERG, and JAFFE than previous models. We also use a visualization technique which, based on the performance of the classifier, can find important face regions for detecting different emotions. We demonstrate through experimental studies that different emotions appear responsive to different parts of the face.

***Key Words*: Deep learning, Face Expression, MATLAB.**

1. INTRODUCTION

The most strong and natural non verbal form of emotional communication is facial expression. Recognizing facial expressions isn't an simple problem. The manner in which people show their emotions can differ greatly. Also images of the same person with the same facial expression that vary with brightness, context and posture, and these differences are highlighted if different subjects are considered (because of form differences, ethnicity among others). Accordingly, perception of facial expression is still a problem in computer vision. To suggest a solution for the detection of facial expression that uses a combination of Convolutional Neural Network and pre-processing measures for different images. It identified the groundbreaking solution that delivers successful face expression and deep learning with convolutionary neural networks (CNNs) has achieved great success in classifying various face emotions as happy, angry, sad and neutral. A number of neuron-wise and layer-wise method of visualization were implemented using a CNN, trained from a given image dataset with a publicly accessible. Therefore, on diagnosis, it is observed that neural networks can capture the colors and textures of lesions that are unique to the respective emotions, which imitate human decision making.

1.1 RELATED WORK

The six key (besides neutral) emotions were defined in one of Paul Ekman's most influential works of emotional acknowledgement, joy, sorrow, anger, surprise, fear, and abhorrence. Ekman later created FACS using this concept, and has set the standard for emotion recognition research ever since. Neutral was later also included, resulting in seven important emotions in most human perception datasets. Earlier work on emotion detection, based on the traditional two-step machine learning approach, where certain features are extracted in the first step from the images, and in the second step, a classifier (such as SVM, neural network, or random forest) is used to classify emotion. Based gradient (HOG) histogram of local binary patterns (LBP), Gabor wavelets, and hair features are some of the typical hand-crafted features used to identify facial expression. Instead, a classifier would allocate the best emotion to the picture. Both methods tended to work well on simpler datasets, but they started to expose their vulnerability with the introduction of more complex data sets (which have more intra-class variability).

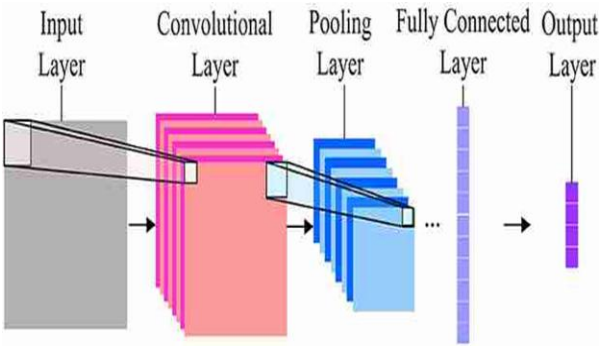
1.2 EXISTING SYSTEM

Now a days facial recognition is used in a variety of real-world applications. Despite its importance, the face-recognition process still entails many challenges, such as facial expression changes, posture variations, occlusions, and lighting. Face is fundamentally a crucial feature of the human body and is used visually for identification purposes. A facial expression is a nonverbal form of communication that plays a vital role in the sharing of experiences. Its significance increases exponentially for people with hearing impairments where the only way to communicate is facial expression. Face area is further broken down into local and national characteristics. Due to various cultures, facial expressions vary from person to person, inborn capacity that adapts from their ancestors, or more specifically due to the unique facial muscle activity of individuals. It hasn't focused on classifying CNN. It hasn't focused on increasing the recognition rate and classification accuracy of the extent of face expression

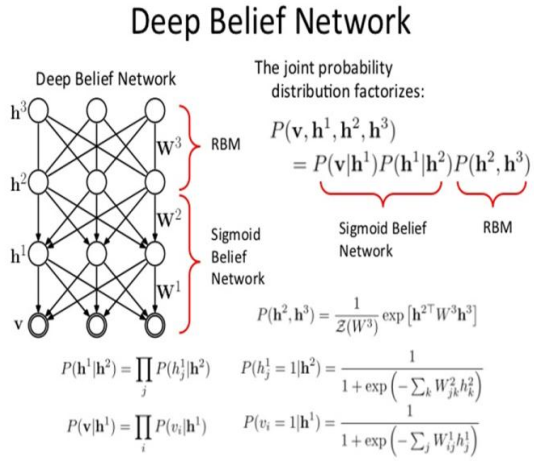
2. DEEP NEURAL NETWORKS

Convolutional neural network (CNN)

CNN is a type of artificial neural network which is used extensively for image based task. It's meaning can be divided into two part: Convolution:- convolution is a mathematical operation of two functions to produce a third function that express how the shape of one can be modified by other, Neural Network:- an artificial neural network is a set of software or/and hardware building blocks modeled on biological nervous system to solve tasks without being programmed explicitly. There are basically five layers to Convolution neural network: 1) Input layer: Input layer is where you give your input to train the model. The model interpreters images as spatial arrangement of pixels; 2) Convolution layer: A convolution layer is used to convert the image(Read a large matrix of pixels) into an another matrix format (read smaller) by employing a special matrix called 'filter'. There are two important concepts regarding the convolution layer: a) Padding: padding is adding extra zeroes at borders so as to avoid information loss and shrinkage. If you notice, while doing convolution the dimensions of image shrinks, to avoid loss in information or maintaining the first size of image padding is applied; b) Strides: Strides essentially tells what moves we are taking when applying convolution in both horizontal and vertical directions.. For example strides of 1 means we are taking one steps in both directions; 3) Pooling layer: pooling layers are wont to reduce the dimensions of image, hence speed up the computation process. There are concepts of max pooling, min pooling, average pooling etc.; 4) Fully connected layer: After extracting the features of the image using convolution, we flatten the output and send through fully connected layers; 5) Output layer: The output from the completely linked layers is fed to a softmax layer that provides probabilistic scores of an image that belongs to a particular class mark.



The spatial size of the output volume can be determined based on the size of the input volume (W), the receptive field size of the Conv Layer neurons (F), the phase in which they are added (S), and the amount of zero padding used (P) at the boundary. You can be persuaded that the right formula for calculating how many "appropriate" neurons are given by (W-F 2P)/S 1.Deep belief network is a network composed of several intermediate layers of Restricted Boltzmann (RBM) structure and the final layer as a classifier. A deep belief network may also be a generative graphic model, or perhaps a type of deep neural network, made up of several layers of ptential variables, with different inter-connections between layers but not between units within each layer. When educated on a disorganized group of examples without guidance, a network of deep beliefs will learn to recreate their inputs in a probabilistic way. The layers then serve as detectors of the elements. A deep belief network is also more equipped with supervision after this learning phase in order to conduct classification. Deep belief networks are often known as combinations of simple, unsupervised networks such as restricted Boltzmann machines (RBMs), where each sub-network's secret layer becomes the next visible layer. A Restricted Boltzmann framework is an undirected generation-energy model with a clear input layer and secret layer and interlayer links but not between layers. This composition contributes to a quick, layer-by-layer unsupervised method of training, where contrastive divergence is applied successively to each sub-network beginning from the lowest pair of layers (the lowest visible layer can be a training set). The discovery that deep conviction networks are often grippingly educated, one layer at a time, contributing to at least one of the first deep learning algorithms to be effective. Generally, there are numerous tempting applications and uses of deep conviction networks in real-life structures and scenarios.



3. CONCLUSIONS

This paper uses an attentive convolutionary network to suggest a new paradigm for facial expression recognition. We believe that attention is an essential piece of facial expression detection that can allow neural networks with less than ten layers to compete (and even outperform) with much deeper emotion recognition networks. We also received comprehensive experimental analysis of our research on four common databases of recognition of facial expression, and showed promising results. We have used a

visualization tool to highlight the unfinished regions of facial images that are the most important parts of them in the identification of various facial expressions

REFERENCES

1] Cowie, Roddy, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G. Taylor. "Emotion recognition in human-computer interaction." IEEE Signal processing magazine 18, no. 1: 32-80, 2001.

[2] Aneja, Deepali, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones. "Modeling stylized character expressions via deep learning." In Asian Conference on Computer Vision, pp. 136-153. Springer, Cham, 2016.

[3] Edwards, Jane, Henry J. Jackson, and Philippa E. Pattison. "Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review." Clinical psychology review 22.6: 789-832, 2002.

[4] Chu, Hui-Chuan, William Wei-Jen Tsai, Min-Ju Liao, and YuhMin Chen. "Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning." Soft Computing: 1-27, 2017.

[5] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[6] He, K., Gkioxari, G., Dollr, P., & Girshick, R. "Mask r-cnn" International Conference on Computer Vision (pp. 2980-2988). IEEE, 2017.

[7] Clavel, Chlo, Ioana Vasilescu, Laurence Devillers, Gal Richard, and Thibaut Ehrette. "Fear-type emotion recognition for future audio-based surveillance systems." Speech Communication 50, no. 6: 487-503, 2008.

[8] Minaee, S., Abdolrashidiy, A., Wang, Y. "An experimental study of deep convolutional features for iris recognition", In Signal Processing in Medicine and Biology Symposium, IEEE, 2016.

[9] Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in neural information processing systems, 2015.

[10] Minaee, Shervin, Imed Bouazizi, Prakash Kolan, and Hossein Najafzadeh. "Ad-Net: Audio-Vi

[11] Tzirakis, Panagiotis, George Trigeorgis, Mihalis A. Nicolaou, Bjrn W. Schuller, and Stefanos Zafeiriou. "End-to-end multimodal emotion recognition using deep neural networks." IEEE Journal of Selected Topics in Signal Processing 11, no. 8: 13011309, 2017

[12] Aneja, Deepali, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones. "Modeling stylized character expressions via deep learning." In Asian Conference on Computer Vision, pp. 136-153. Springer, Cham, 2016.

[13] Edwards, Jane, Henry J. Jackson, and Philippa E. Pattison. "Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review." Clinical psychology review 22.6: 789-832, 2002.