Emotion analysis from facial expressions using deep learning

Meghana Joshi

Department of Computer Science and Engineering
BMS College Of Engineering
Bengaluru, India

***Abstract*—Automated facial expression recognition has been proven to greatly improve the human–machine interface. Automatic facial expression recognition is an actively emerging research in Emotion Recognition. A large number of deep learning approaches have been applied in recent years due to their outstanding recognition accuracy which can be attained after training the model with large amounts of data. This research paper throws light on how to enhance Convolutional Neural Network method to recognize 6 basic emotions and compare some pre-processing methods to show its influences on the performance. The preprocessing methods are : resizing, mean, normalization, standard deviation, scaling and edge detection**

Index Terms -Emotion recognition, Facial Expression Recognition, Convolutional Neural Network

I. Introduction

As the exposure of humans with machines has been increasing, some sort of machine perception could be developed (machines are able to appreciate and understand their surroundings)[1]. Humans use their senses to gain insights about their environment. Therefore, machine perception aims to mimic human senses in order to interact with their environment[2][3]. Nowadays, machines have several ways to capture their environment state through cameras and sensors. Hence, using this information with suitable algorithms allows us to generate machine perception. In the last years, the use of Deep Learning algorithms has been proven to be very successful in this regard[4].

Emotion recognition can simply be expressed as the process of identifying human emotion. Since people vary widely in their ability to accurately recognize the emotions of others, the use of technology to help people with emotion recognition has started to develop, however, it still is a nascent field. Generally, the technology works best if it uses multiple modalities in context (facial expressions from video, spoken expressions from audio or written expressions from text). The process of expression recognition can be categorized into three parts: image acquisition and pre-processing, image feature extraction and selection, and finally classification. It is quite well known that basic emotions consist of anger, happiness, sadness, disgust, fear and surprise. The technology of automatic facial expression analysis finds its use in a wide variety of fields such as - clinical psychology, neurology, lie detection, pain assessment and multimodal human computer interface (HCI). In order to employ facial expression analysis, there exist two approaches for feature extraction and action unit detection from facial action coding system (FACS). These techniques are (i) geometric feature-based and (ii) appearance-based feature extraction. Geometric feature based technique represents facial points which form feature vectors. These vectors describe the face from a geometrical approach. Appearance based feature extraction is applied in the extraction of feature vectors, either in specific or holistic face areas of the facial image, such as LBP (Local Binary Patterns) application.

Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning that can further be utilized as emotion recognition and facial expression analysis. However, its performance directly depends on the data size ie. the greater the data set, the better the performance is. The size of facial expressions datasets are still insufficient for deep learning to be implemented thoroughly. Therefore, some researchers tend to apply augmentation techniques in the pre-processing step such as resizing, scaling, mean, standard deviation, normalization, rotation, edge detection. This research study aims to compare the accuracy of each pre-processing method and the combinations between them, then analyse it in order to see the variability of accuracies.

II. Related Work

A. Research Related Studies

According to the authors of Facial expression recognition with Convolutional Neural Networks[5], there is a single solution

for facial expression recognition which uses a combination of Convolutional Neural Networks and specific image pre-processing steps. Convolutional Neural Networks (CNN) can achieve higher accuracy with big data. This experiment employed three largely used public databases (Ck+, JAFFE and BU-3DFE) to train and allow for real time facial expression recognition with the help of standard computing power.

According to the author of Convolutional Neural Networks Based Method for Improving Facial Expression Recognition[6], the research used different imbalanced data sets of facial expression. First, the data is pre-processed after which it is balanced using the following classifier models - Decision Tree (DT), Multilayer perceptron (MLP), Convolutional Neural Network (CNN). CNN is determined to produce the best recognition accuracy.

B. Research Methodology

This proposal aims to explore the opportunities provided by facial expression recognition by performing feature extraction and classification for each of the emotions - anger, happiness, sadness, disgust, fear and surprise. Technically, this paper's goal consists of applying some pre-processing steps like resizing, normalization, scaling, mean, standard deviation, rotation, and edge detection to train a deep neural network with labeled images of static facial emotions. Later, this network could be used as part of a software to detect emotions in real time. This capability can be used by machines to improve their interaction with humans by providing more adequate and germane responses.

1) Data Sets

The experiments can be performed using a posed dataset JAFFE, which means every subject was given instructions by experts to demonstrate the six abovementioned emotions. This dataset can be intensively used for emotion recognition and facial expression analysis. Furthermore, the male and female facial expression (JAFFE) has 100 images that represent some basic emotions from male and females . The data consists of gray scale images of faces in 48 x 48 pixel. The faces are more or less face centered and occupy about the same amount of space in each image. Our ultimate goal is to categorize each face based on the emotion shown in the facial expression into one of six categories. We categorise each image based on two columns- "emotion" and "pixels". The "emotion" column contains a numeric code ranging from 0 to 6, for the emotion that is present in the image. The "pixels" column contains a string surrounded in quotes for each image. The resulting string can be looked at as an array of 2304 pixel values arranged in a particular order. Our task now is to predict the "emotion" column, which varies from 0-6 (Neutral emotion is also added for comparison purposes).

Table 1. The JAFFE datasets

	JAFFE
Label	7 emotions
The Number	100
Participants	100 (male female)
Resolution	48/48 pixel array
Format	.tiff

2) Pre-processing

Data Preprocessing is a technique that is applied in order to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible from the analysis point of view. Pre-processinggenerally encompasses steps such as resize, normalization, scaling, mean, standard deviation, rotation, edge detection. Image pre-processing is the study of algorithms which take an image as input and give a vector(features) or image(matrix) as output. We do this kind of preprocessing to extract the hidden information and not to lose any information or to remove noise or to normalise the pixel data.

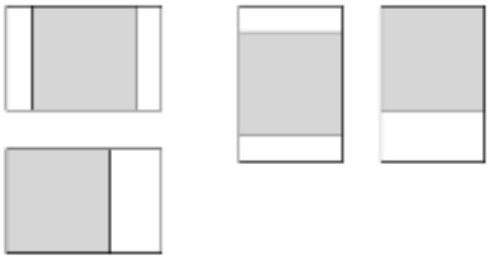
Resizing:

An image will be adjusted to determine width, height and converted to gray scale image during the process of resizing.

Example:

Before resizing (1920, 1080)

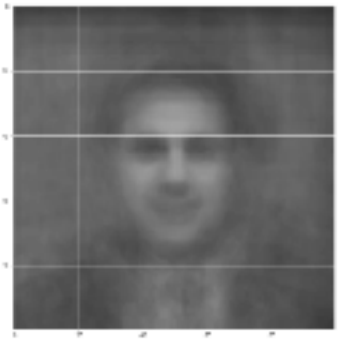
After resizing (888, 500)



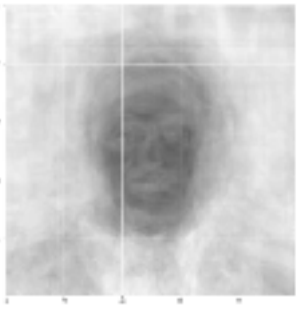
Scaling:
While scaling an image, a new image with a higher or lower number of pixels must be generated.

Example image: Square 250 x 250

Mean:
This method is used to reduce the amount of intensity variation between one pixel and the other. It is often used to reduce noise in images.



Standard Deviation:
When we process an image with standard deviation , it will show the variance which is nearer or farther to the mean value of the given image.



Normalization:
It is a process that changes the range of pixel intensity values in images with poor contrast due to glare. It is also called contrast stretching.

Rotation:
Images can be rotated to any degree clockwise or anti clockwise. We just need to define rotation point, degree of rotation and the scaling factor.

Edge Detection:
The process of image detection involves detecting sharp edges in the image. This edge detection is essential in object detection.

3) Convolutional Neural Networks

The study of the visual cortex is closely related to the development of the convolutional neural networks. Back in 1968, Hubel and Wiesel presented a study focused on the receptive fields of the monkey's visual cortex [7]. This study was relevant because of the striate cortex (primary visual cortex) architecture description and the way that cells are arranged on it.In deep learning a convolution neural network (CNN) is a class of deep neural methods, most commonly applied to analyzing visual imagery. CNN uses a variation of multilayer perceptrons designed to require minimal preprocessing; they are also known to shift invariant or space invariant artificial neural networks (SIANN) based on their shared weight architecture and translation invariance characteristics. CNN uses relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that the traditional algorithms were hand-engineered. A convolution neural

network considers an input and an output layer as well as multiple hidden layers. Each convolution neuron processes data only for its respective field although fully connected feedforward neural networks can be used to learn features as well as classify data. Convolution layer brings a solution to many large input sizes associated with images, where each pixel is a selected variable. For this CNN brings a solution to this problem as it reduces the number of face parameters, allowing the network to be developed with few parameters by back propagation.\It is in principle the same as a traditional multilayer perceptron neural network goes through a fully connected layer to classify the image. The input area of a neuron is called its respective field. Learning in a neural network program by making adjustment the biases and weight. The vector of weight and the bias are called a filter and represents some feature of the input.

III. Conclusion

Two main contributions are presented in this paper. One contribution is that the proposed pre-processing method can assist the CNN model to gain the higher accuracy rate in the applications of facial image processing. It is evident that the capacity of the model satisfies the complexity task for facial expression recognition on those resolutions. We can boost the performance of CNN using data augmentation like combining data from step (b) cropping and (f) adding noises. The future work involves exploring image synthesis techniques that may be considered as a solution of augmentation data in deep learning. It aims to prevent data starvation and overfitting for small amounts of data.

Acknowledgment

The work reported in this paper is supported by the college through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-III] of the MHRD, Government of India.

References

[1] Matthew Turk. Perceptive media: machine perception and human computer interaction.
CHINESE JOURNAL OF COMPUTERS-CHINESE EDITION-23(12):1235–1244, 2000.

[2] G. Mather. Essentials of Sensation and Perception. Foundations of Psychology. Taylor & Francis, 2014.

[3] Michael Cooney. Ibm: In the next 5 years computers will learn, mimic the human senses.
http://www.networkworld.com/article/2162228/data-center/ibm--in-the-next-5-years-computers-will-learn--mimic-the-human-senses.html

[4] Geoffrey Hinton et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE, 29(6):82–97, 2012.

[5] https://doi.org/10.1016/j.patcog.2016.07.026 by André TeixeiraLopes,Edilsonde Aguiar,Alberto F.De Souza,Thiago Oliveira-Santos

[6] Rashid, Tarik. (2016). Convolutional Neural Networks Based Method for Improving Facial Expression Recognition.

[7] David H. Hubel and Torsten N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology (London), 195, 1968.