Emotion Detection and Classification

Salehin Rahman Khan, Ashikul Haque Khan, Maisha Maliha, Faiyaz Khaled, Ehsanul Amin Khan
CSE, BRAC University
Mohakhali 1212, BRAC University
Dhaka, Bangladesh
{Salehin9029, ashik.mw, maisha.ruhe, mahin.faiyaz, eh41995san}@gmail.com

Abstract- A key advancement that has been made in the refinement of Artificial Intelligence is the capacity to order the feeling of the human administrator. In this paper we introduce the plan of an artificially clever system equipped for predicting the emotions through facial articulations. Three promising neural network structures are altered, prepared, furthermore, subjected to different classification assignments, after which the best performing system is additionally streamlined. The relevance of the display is depicted in a live video application that can promptly restore the emotion of the client with an accuracy of seventy one percent.

Keywords— emotion; detection; classification;

1. INTRODUCTION

Ever since computers were developed, scientists and engineers thought of artificial intelligent systems that will be mentally and/or physically equivalent to humans. Within the past few decades, the rise of generally available computational power provided a a helping hand for developing quick learning machines whereas the web provided a vast quantity of data for training. These two developments boosted the research on smart self-learning systems, with neural networks among the foremost promising techniques.

1.1 BACKGROUND

One of the present best uses of artificial intelligence utilizing neural systems is the acknowledgment of appearances in photographs and recordings. Generally procedures process visual information and look for general examples displayed in human appearances. Face acknowledgment can be utilized for reconnaissance purposes by law masters as well as in swarm administration. Other present-day applications include programmed obscuring of countenances on Google Streetview film and programmed acknowledgment of Facebook companions in photographs. A significantly further developed improvement in this field is feeling acknowledgment. Notwithstanding just distinguishing faces, the PC utilizes the game plan and state of eyebrows and lips to decide the outward appearance and thus the feeling for every child. One conceivable application for this lies in the territory of reconnaissance and behavioral investigation by law enforcement. Besides, such procedures are utilized as a part of computerized cameras to consequently take pictures when the client grins. In any case, the most encouraging applications include the refinement of artificial intelligent frameworks. On the off chance that PCs can follow along of the psychological condition of the client, robots can respond upon this and carry on properly. Feeling recognition in this way assumes a key-part in enhancing human-machine association.

1.2 RESEARCH OBJECTIVE

In this research we tend to primarily target neural network based on artificially intelligent systems capable of deriving the feeling of an individual through the footage of his or her face. Completely different approaches from existing literatures are experimented with and so as the results of the various choices in design. The main research question, therefore, becomes: **How can an artificial neural network be used for interpreting the facial expression of a human?**

The rest of this article depicts the few stages taken to answer the primary research question i.e. the sub-questions. In segment 2, a writing study will elucidate what the part of outward appearances is in feeling acknowledgment and what sorts of systems are reasonable for computerized picture arrangement. The third segment clarifies how the neural systems under thought are organized and how the systems are prepared. Segment 4 depicts how the final model performs after which a conclusion and a few proposals follow in the last area. It might be noticed that the point of our work isn't to outline a feeling recognizer without any preparation yet rather to audit plan decisions and upgrade existing procedures with some new thoughts.

2. LITERATURE REVIEW

For the advancement of a framework that can recognize feelings through outward facial appearances, previous research on the way humans reveal emotions as well as the theory of automatic image categorization is reviewed. In the rest part of this segment, the part of translating outward appearances in emotions will be talked about. The last part discusses about previous techniques on programmed picture classification.

2.1 HUMAN EMOTIONS

A key element in human communication is the all inclusiveness of outward appearances and non-verbal communication. As of now in the nineteenth century, Charles Darwin distributed published upon globally shared outward appearances that played critical part in non-verbal correspondence. In 1971, Ekman and Friesen pronounced that facial practices are universally related with specific feelings. Clearly Humans, but also animals, create comparative solid developments with a certain mental state, regardless of their place of birth, race, training etc. Consequently, if legitimately demonstrated, this all inclusiveness can be an exceptionally advantageous element in human-machine communication: an all around prepared framework can comprehend feelings, autonomous of whoever the subject is.

One should remember that outward appearances are neither really straightforwardly translatable into feelings nor the other way around. Outward appearance is furthermore a capacity of e.g. mental state, while feelings are moreover communicated by means of non-verbal communication and voice. More intricate feeling acknowledgment frameworks ought to subsequently likewise incorporate these last two commitments. Be that as it may, this is out of the extent of this examination and will remain a proposal for future work. Perusers keen on investigating the classification of feelings through discourse acknowledgment are alluded to Nicholson et al.

As a last purpose of consideration, feelings ought not be mistaken for state of mind, since mind-set is thought to be a long haul mental state. As needs be, state of mind acknowledgment frequently includes longstanding investigation of somebody's conduct and articulations, and thus be precluded in this work.

2.2 DISCUSSION

(i) A breakthrough publication on automatic image classification in general is given by Krizhevsky and Hinton. This work shows a deep neural network that resembles the functionality of the human visual cortex. Using a self-developed labeled collection of 60000 images over 10 classes called the CIFAR-10 dataset, a model to categorize objects from pictures is obtained. Another important outcome of the research is the visualization of the filters in the network, such that it can be assessed how the model breaks down the pictures.

(ii) In another work which adopts the CIFAR-10 dataset, very wide and deep network architecture is developed, combined with GPU support to decrease training time. On popular datasets, such as the MNIST handwritten digits, Chinese characters, and the CIFAR-10 images, near-human performance is achieved. The extremely low error rates beat prior state-of-the-art results significantly. However it has to be mentioned that the network used for the CIFAR-10 dataset consists of 4 convolutional layers with 300 maps each, 3 max pooling layers, and 3 fully connected output layers. As a result, although a GPU was used, the training time was several days.

3. DATASET ANALYSIS

For emotion recognition, many datasets are available for research, varying from a few hundred high resolution photos to tens of thousands of smaller images. Neural networks and deep networks In particular, are known for their need for large amount of training data. The performance of the model highly depends on the choice of images used for training. So, it signifies that the dataset should be both highly qualitative and quantitative as well.

The datasets mainly differ on quantity, quality and cleanness of the images.We use the FER-2013 Faces Database, where there is a set of 28,709 pictures of people displaying 7 emotional expressions. The expressions are angry, disgusted, fearful, happy, sad, surprised and neutral.

The FER-2013 set has large size of dataset which is very beneficial for the robustness of a model. Another reason for choosing this dataset is that once trained upon the set, images from ‘clean’ datasets can easily be classified. We have divided the dataset into two parts, one of them for training and another for testing. Figure 1 shows the number of images per emotion present in the training set. Figure 2 shows Number of images per emotion in the training set.

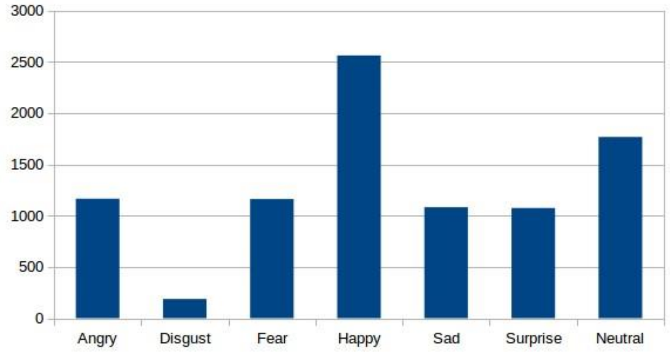

Figure 1: The Number of images per emotion present in the training set

Figure 2: Number of images per emotion in training set

There are two columns in the dataset. The input column is pixels which contain a string surrounded in quotes for each image and the output column which is emotion contains numeric code ranging from 0 to 6 where the numbers represents each individual emotions. With use of the Haar Feature-Based Cascaded Classifier inside the OpenCv framework, all data are preprocessed. For every image, only the square part containing the face is taken, rescaled, and converted to an array with 48x48 grey-scale values.

4. METHODS

A. Basically, we are training the dataset by neural network. The first network is based on the research and algorithm by Krizhevsky and Hinton. This is the smallest network out of the three that we have used, which means that it has the lowest computational demands among all. The network consists of three convolutional layers and two fully connected layers, combined with maxpooling layers for reducing the image size and a dropout layer to reduce the chance of over fitting. The hyper parameters are chosen such that the number of calculations in each convolutional layer remains roughly the same. This ensures that information is preserved throughout the network. In this research we only distinguishing seven emotions and for the limited computing resources the size of the network is considered to be large.

B. The AlexNet convolutional network was developed for classifying images in more than 1000 different classes, using 1.2 million sample pictures from the ImageNet dataset. Due to the fact that in this research the model only has to distinguish seven emotions, and due to our limited computing resources, the size of the original network is considered to be too large. Hence, instead of 5 convolutional layers we applied 3, and in the subsequent 3 fully connected layers the number of nodes of each fully connected was reduced from 4096 to 1024. While the original network was divided for parallel training, it was observed that was not necessary for the smaller version. The network also makes use of local normalization to speed up the training and dropout layers in order to reduce the over fitting.

C. The last experiments are performed on a network based by the work of Gudi. It take images of testing set and finally the usage of open CV to take live images from video and convert them to 48 x 48. And thus it gives us our desired real time feedback which is the main goal of our project.

5. TEST RESULTS

The state of the art networks obtained from test sets, which is about 67% . Our resources are limited and the result are pretty much good. The accuracy on the RaFD test set have completely different images than the trained data. The accuracy rates of the final model are given in table below

Table 2: Accuracy of the networks

Network	FERC-2013 Validation	FERC-2013 Test	RaFD
A	63%		50%
B	53%		46%
C	63%		60%
Final	66%	63%	71%

For every emotion, a table is generated. Very high accuracy rates are basically obtained on happy (90%), neutral (80%) and surprised (77%). These are the main facts from which the emotions will be differentiated.

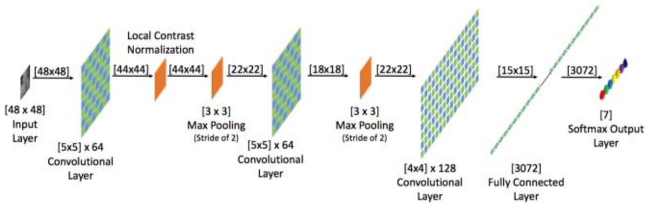

Figure 3: Overview of the Network Architecture of the Final Model.

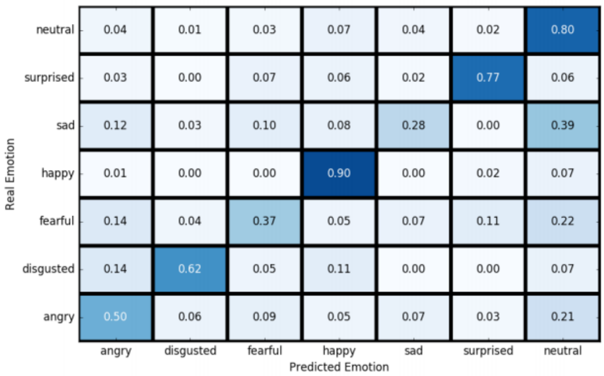

Figure 4: Performance Matrix of the Final Model

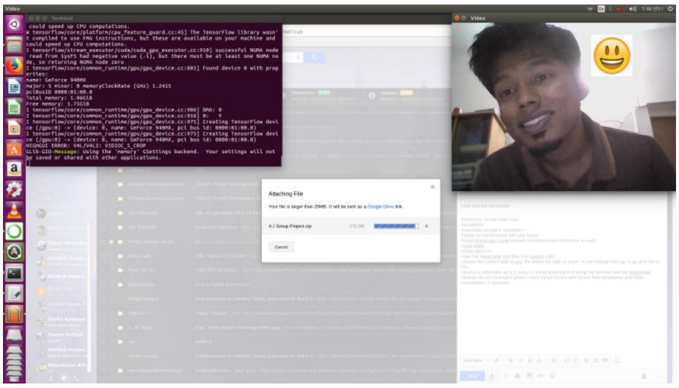

Figure 5: Screenshot of the Output

6.CONCLUSION

Our Project hit Staggering 90% accuracy when it comes to happy emotion detection and classification with an overall accuracy of 71% which is fair. We believe that as a prototype it's a great result which can be enhanced in the future to be implemented in industries. The influence of human emotions upon effectiveness of our work and the quality of the results is unquestionable. The scenarios proposed in this paper should help to evaluate the scale of this phenomenon in artificial intelligence, which has not been investigated before. The proposed approach to emotion recognition should lead to more reliable results when compared with approaches based on a single information channel or questionnaires.

7.References

[1] T. Ahsan, T. Jabid, and U.-P. Chong. Facial expression recognition using local transitional pattern on gabor _ltered facial images. *IETE Technical Review*, 30(1):47{52, 2013.

[2] D. Ciresan, U. Meier, and J. Schmidhuber. Multicolumn deep neural networks for image classification In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642{3649. IEEE, 2012.

[3] C. R. Darwin. *The expression of the emotions in man and animals*. John Murray, London, 1872.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition,2009. CVPR 2009. IEEE Conference on*, pages 248{255. IEEE, 2009.

[5] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[6] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern recognition*, 36(1):259{275, 2003.

[7] A. Gudi. Recognizing semantic features in faces using deep learning. *arXiv preprintarXiv:1512.00743*, 2015.

[8] Kaggle. Challenges in representation learning: Facial expression recognition challenge, 2013.

[9] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097{1105, 2012.

[11] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. van Knippenberg. Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377{1388, 2010.

[12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94{101. IEEE, 2010.

[13] Y. Lv, Z. Feng, and C. Xu. Facial expression recognition via deep learning. In *Smart Computing (SMARTCOMP), 2014 International Conference on*, pages 303{308. IEEE, 2014.

[14] J. Nicholson, K. Takahashi, and R. Nakatsu. Emotion recognition in speech using neural networks. *Neural computing & applications*, 9(4): 290{296, 2000.

[15] OpenSourceComputerVision. Face detection using haar cascades. URL http://docs.opencv.org/master/d7/d8b/tutorial_py_face_detection.html.

[16] TFlearn. Tearn: Deep learning library featuring a higher-level api for tensorow. URL: http://tflearn.org/.