Sentiment and Gender classification using countenance

Pankaj Chandra[1*], Santosh Soni[2] Mukul Gupta[3]

[123] Department of Information Technology, School of Studies (Engineering and Technology), Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh

Abstract- This paper posit an implementation for deploying a real time general convolutional neural network (CNN) framework. This model is capable of face discernment, gender and sentiment categorization using real-time countenance, all using one start-of-the-art CNN. We also propose a hybrid Haar Cascade classifier for feature extraction. The model produces an accuracy of 95% on IMFDB dataset for gender and an accuracy of 76% for FER2013 and KDEF dataset for emotions. Along with this, we have deployed visualization technique via guided back propagation.

Keywords—Deep Neural Networks, Hybrid Haar Cascade, Image Processing

INTRODUCTION

Outward appearances assume a significant job in acknowledgment of feelings and are utilized during the non-verbal communication, just as to distinguish individuals. They are significant in day by day enthusiastic correspondence, only alongside the manner of speaking. They are likewise a pointer of emotions, permitting a person to express an emotional state. Individuals, can quickly perceive the emotional condition of an individual. As an outcome, data on the outward appearances are regularly utilized in programmed frameworks of feeling acknowledgment. The point of the examination, exhibited in this paper, is to categorize six fundamental sentiment states: happy, sad, neutral, fear, surprise and angry based on countenance.

Person's face, as the most uncovered piece of the body, permits the utilization of computers' vision frameworks (normally cameras) to examine the picture of the face for perceiving feelings. Light conditions and changes of head position are the primary factors that influence the nature of feeling acknowledgment frameworks utilizing camera.

Strategies based on spatial and temporal dependencies are undeniably additionally encouraging.

The accomplishment of smooth robot to client interaction and cooperation is only possible by correctly recognizing the emotions. In these components utilizing AI (ML) systems has demonstrated to be convoluted due the high changeability of the examples inside each errand. Manual classification of emotions in FER2013 dataset is very tiresome and using ML methods is inefficient because of variability of samples and thousands of parameters.

Due to the hardware and computational restraints in robotics and real time systems, implementation of such complex models is not viable.

Therefore, we propose to implement a start-of-the-art CNN for sentiment and gender classification, using least parameters and achieving human level efficiency.

RELATED WORK

Ordinarily utilized CNNs for feature extraction incorporate a set of completely connected layers toward the end. Completely connected layers will in general contain huge numbers of parameters in a CNN. In particular, VGG16 contains roughly 90% of all its parameters in its last connected layers. Later designs, for example, Inception V3, diminished the sum of parameters in their last layers by including a Global Normal Pooling activity. This decreases each element map into a scalar one by taking the normal over all components in the element map. The normal task powers the system to extricate global features from the picture. Current CNN models, for example, Xception influence from the mix of two of the best test suppositions in CNNs: the utilization of residual modules and depth wise distinguishable convolutions.

METHODOLOGY

We propose to build a general, sequential CNN architecture in which the fully connected layer in the end will be removed and a depth wise distinguishable layers and residual modules will be added. The proposed architecture is shown in Fig. 1. This will lead to significant reductions in number of parameters and alleviate from slow performance. The fully connected layers are replaced by Global Average Pooling layer. This was achieved by having in the last convolutional layer the same number of feature maps as number of classes, and applying a softmax activation function to each reduced feature map. Adam as an optimizer was utilized and the layers are activated using ReLU (Rectified Linear Unit) activation function. To avoid overfitting, l2 (ridge regularizer) was deployed.

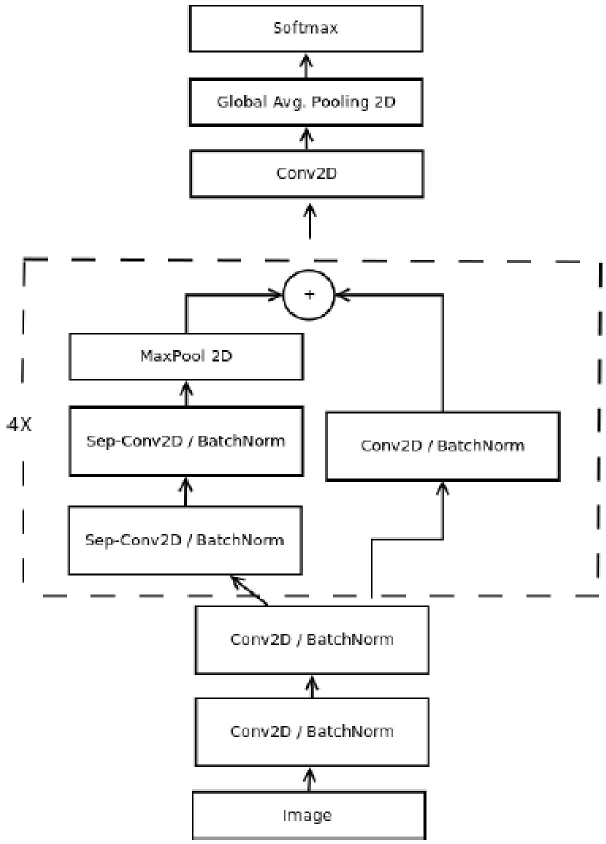

Fig 1: Proposed CNN Architecture

The proposed architecture is a standard fully-convolutional neural network comprising 9 convolution layers, ReLUs, Batch Normalization and Global Average Pooling. This model contains approximately 600,000 parameters. It was trained on the IMFDB gender dataset, which contains 34,512 RGB images where each image belongs to the class "woman" or "man", and it achieved an accuracy of 95% in this dataset. This model was validated for FER2013 dataset. This dataset contains 35,887 grayscale images where each image belongs to one of the following classes {"angry", "fear", "happy", "sad", "surprise", "neutral"}. We achieved an accuracy of 76% in this dataset.

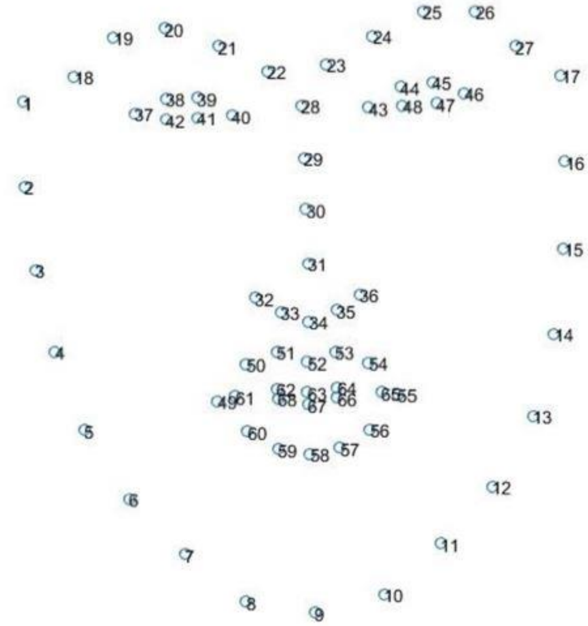

Fig 2: Facial features extraction points

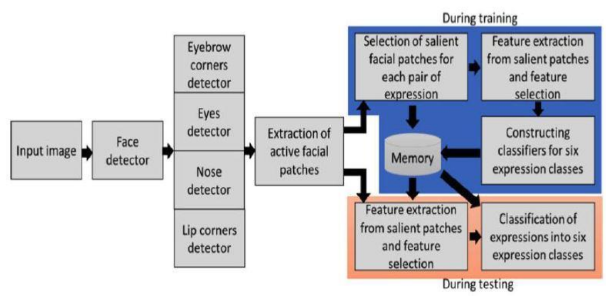

Fig 3: Working Model

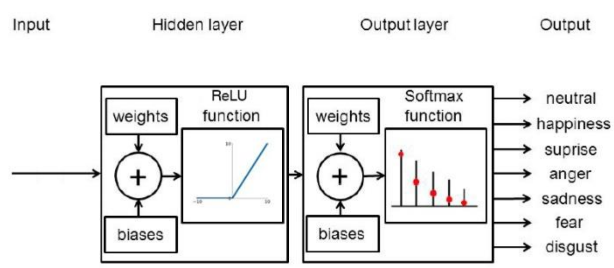

Fig 4: Inside Architecture

Also, as a classifier and for face detection, a hybrid model of Haar Cascade classifier was designed as shown in Fig. 5.

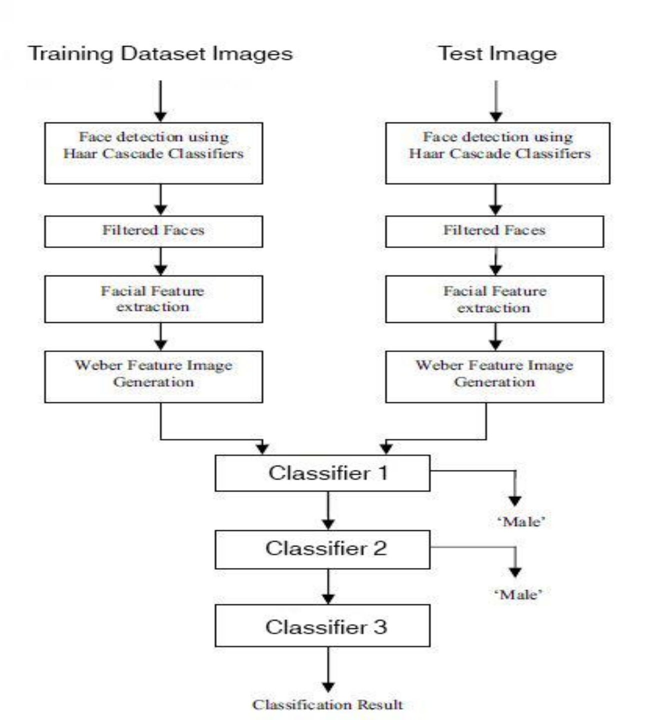

Fig 5: Hybrid Haar Cascade classifier

RESULTS

The real-time results are shown below. The confusion matrix is also shown below. The model gives false positives results for gender classification and misclassification between fear and sad. The selected neuron was always selected in accordance to the highest activation. We can observe that the CNN learned to get activated by considering features such as the frown, the teeth, the eyebrows and the widening of one's eyes, and that each feature remains constant within the same class. These results reassure that the CNN learned to interpret understandable human-like features, that provide generalizable elements. These interpretable results have helped us understand several common misclassifications such as persons with glasses being classified as "angry". This happens since the label "angry" is highly activated when it believes a person is frowning and frowning features get confused with darker glass frames.

Fig 6: Result

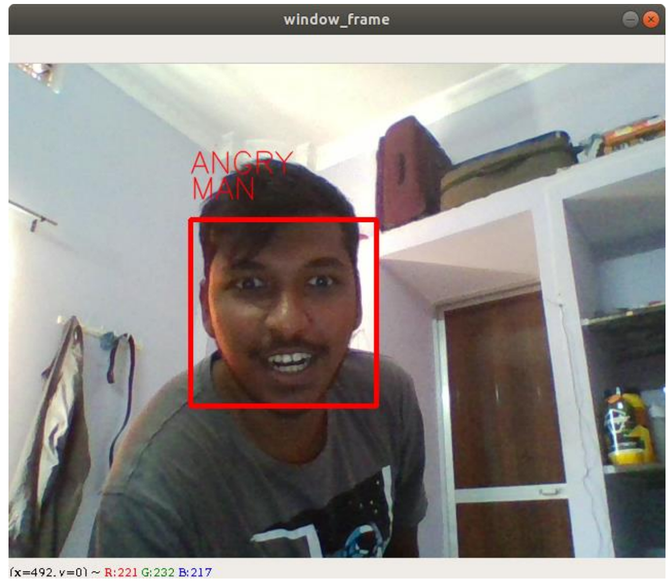

Fig 7: Result

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	425	2	1	3	10	0	1
joy	9	421	0	2	6	0	5
surprise	1	1	429	0	0	11	0
anger	7	1	0	428	1	0	6
sadness	20	4	0	2	416	1	0
fear	5	0	19	0	6	412	1
disgust	2	2	1	9	2	0	427

Fig 8: Confusion Matrix

FUTURE SCOPE

AI models are biased based on their training dataset. Our model too is biased towards Indian faces. Also the use of glasses and improper illumination disrupt the detection and the model is sometime unable to classify gender and sentiments correctly.

CONCLUSION

Our proposed models have been methodically worked so as to lessen the measure of parameters. We started by wiping out totally the completely connected layers and by diminishing the measure of parameters in the remaining convolutional layers by means of profundity astute distinct convolutions. We have demonstrated that our proposed models can be stacked for multi-class orders while keeping up ongoing derivations. In particular, we have built up a dream framework that performs face recognition, gender categorization and sentiment classification in a solitary coordinated module. We have accomplished human-level execution in our orders undertakings utilizing a solitary CNN that use present day design builds. Our design decreases the measure of parameters while getting great outcomes.

References

[1] Franc¸ois Chollet. Xception: Deep learning with depthwise separable convolutions. CoRR, abs/1610.02357, 2016.

[2] Andrew G. Howard et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR, abs/1704.04861, 2017.

[3] Dario Amodei et al. Deep speech 2: End-to-end speech recognition in english and mandarin. CoRR, abs/1512.02595, 2015.

[4] Ian Goodfellow et al. Challenges in Representation Learning: A report on three machine learning contests, 2013.

[5] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 315–323, 2011.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, pages 448–456, 2015.

[8] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[9] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision (IJCV), July 2016.

[10] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[11] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806, 2014.

[12] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2818–2826, 2016.

[13] Yichuan Tang. Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239, 2013.

[14] M. Turk and A. Pentland, "Eigenfaces for Recognition," J. Cognitive Neuroscience, pp. 71–96, 1991.

[15] L. Bui, D. Tran, X. Huang and Girija Chetty, "Classification of Gender and Face based on Gradient faces," Visual Information Processing (EUVIP), pp. 269-272, 2011.

[16] S. Halder, D. Bhattacharjee, M. Nasipuri, D. Basu and M. Kundu, "Face Synthesis (FASY) System for Determining the Characteristics of a Face Image," CoRR, 2010.

[17] Y. Freund and R. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," J Comput Syst Sci 55(1), pp. 119-139, 1997.

[18] Open Computer Vision Library http://sourceforge.net/projects/opencvlibrary/

[19] A. Sohail and P. Bhattacharya, "Detection of Facial Feature Points Using Anthropometric Face Model," IEEE International Conference on Signal-Image Technology and Internet-Based Systems, pp. 656-665, 2006.

[20] B. Wang, W. Li, W. Yang and Q. Liao, "Illumination Normalization Based on Weber's Law With Application to Face Recognition," IEEE Signal Processing Letters, pp. 462-465, 2011.

[21] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[22] B. Moghaddam and M. Yang, "Gender Classification with Support Vector Machines," Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 306–311, 2000.