Identification of Facial Expressions using Deep Neural Networks

Preetika Soni , Harsh Jain , Parv Bharti, A. Kumar Dubey *

Information Technology Bharati Vidyapeeth's College of Engg,New Delhi, India
Emails: sonipreetika20@gmail.com; harshjain2525@gmail.com; parv.bharti@gmail.com; arudubey@gmail.com
***** Correspondence: arudubey@gmail.com

Abstract

Detecting and analyzing emotions from human facial movements is a problem defined and developed over many years for the benefits it brings. During playback, when developing data sets, data sets with methods become more and more complex, and accuracy and difficulty increase gradually. In the given paper, we will use a deep structured learned network using the two mechanisms - Vgg and Resnet50 with deep layers to classify emotions based on input images in complex environments. Besides that, we also use learning methods combining many modern models to increase accuracy. Experimental results show that the two proposed methods have better results than some modern methods in emotional recognition problems for complex input images and some results reported in scientific studies. Particularly combined learning method gives good accuracy - 66.15% on the dataset FER2013

Keywords: Facial expression; Deep Neural Network; VGG, Resnet;

1. Introduction

Understanding human sentiments[1] is an important area of analysis and experimentation because being able to analyze a person's feelings can grant a person ingress to many more chances to discover unexplored areas of application in daily life ranging from personal laptops to desktop computers, chosen marketing advertisement, and enhanced access to social interaction, by improving individual sentiment ("EQ"). ours. There are many possibilities in which one can inquire into the identification of a person's feelings, from the facial look, body language, and voice pitch. In the given paper, we will pivot on only one domain in this field - the visual identification of emotions. One of the prime motives we decided to focus on facial expressions is because certain facial expressions have a universal meaning, and these feelings have been noted for decades and even centuries. "One who lives in all cultures of faces and emotions" [2]. Seven emotions were identified: sadness, anger, fear, disgust, happiness, surprise, and neutral. Currently, researchers are using these facial expressions in computer vision, such as Kaggle's Facial Expression Recognition Challenge. Therefore, our work (the VGG network and ResNet50 renaming network) is based on identifying these seven main human emotions about deep learning [3]. For us, this issue is of great importance due to its wide range of working in various fields, such as formal recruitment, while also being able to Integrate with various technologies (e.g., Virtual Reality, smart glasses, etc.). Emotions and facial expressions can serve as a new feature of user experience (e.g., Imagine Facebook or Google analyzing your mood and response inOrder to learn more about a person and offer better recommendations and ads). To achieve our goal, we will be using some of the current deep learning architecture models - VGG and ResNet50, while making other changes that involve the use of various deep and machine learning techniques and integration and transfer learning [5]. We have chosen to go with

DOI : https://doi.org/10.54216/FPA.020101

Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

VGG and ResNet50 because they have overcome the previous ImageNet challenge, showed superior results in terms of predictability, and followed the standard CNN architecture. The database we used was FER2013, which contains 35,887 grayscale images. We found these data to be representative because of their size, shape (depending on the face, ethnicity, age, and gender), and similar data distribution across seven key demographic variables. Training efficiency, validation, and various test and training sets were used to test the performance of our models. The procedures will then include some general statistics such as precision and recall to provide a number of optimization models. Our best model was expected to reach at least 50% accuracy.

Emoticons, also known as facial expressions, play a very important role on social networks. The discussion that took place has both linguistic and non-linguistic features. Non-verbal items include visual, body language, facial expressions, body language, etc. The bud smiles indicate joy. The sad expression indicates loss, and angry talk shows unhappiness with the unexpected talk that shows that something unexpected happened. According to C. Darwin and P. Prodger [7], facial expression is one of the natural, universal, and most powerful human traits that express their intentions and emotional conditions. In the field of computer vision and machine learning, there are many scientists who have researched automated facial analysis systems for their use.

2. Related Work

It is most important to extract facial expression features from the face accurately, as its accuracy impacts the results of the final classification of the expression, which is the extraction of facial expression's next step. In spite of the fact that there is a great deal of examination into the history too the advancement of facial acknowledgment, we find that there are a few sorts of strategies that can be utilized for facial recognition. For example, facial recognition depends on the extraction of the component purposes of the component geometry and eigenface.

A geometric-based technique is normally used to separate the area of facial organs as the attributes of grouping. Some referenced that the thought is to acquire relative position and different parameters of unmistakable highlights, for example, eyes, mouth, nose, and jawline [8]. This capacity is old and dull, yet there are as yet two principle shortcomings in geometrical highlights-based capacity: the first is that in vitality work, the weighting coefficients, which is hard to outline, must be controlled by understanding. Another weakness is the procedure of enhancement of vitality work is tedious. As an afterthought, the recognition innovation of the highlight point can't be precisive, and the calculation is costly too.

Eigenface, which was presented first by Matthew Turk and Alex Pentland in 1991, has gotten one of the most famous calculations during these years. It is known as straightforward and successful. A straightforward way to deal with separating the data contained in a picture of a face is, in some way or another, to catch the variety in an assortment of face pictures, autonomous of any judgment of highlights, and utilize this data to encode and look at singular face pictures [5].

As such, the principle thought of eigenface is to change a face picture from pixel space to another space and, at that point, do a figuring of similitude in another space. The eigenface utilizes the PCA function to get the creation of human faces. Alternatively, training set, we wish to obtain the eigenvalue decomposition matrix of human faces' images and corresponding eigenvectors. All the eigenvectors which we accomplish through calculation are called featured faces.

DNN is a significant learning technique that is produced for picture grouping and acknowledgment. Likewise, it has been generally utilized in the field of FER these days. DNN, as a profound learning engineering, can reduce the model's complexity and accurately extract image features [7].

DOI : https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

3. Methodology

3.1 VGG

A progression of VGG has the equivalent structure consisting of three fully connected layers and outcoming through a softmax layer with ReLu function. The general structure contains five combinations of convolutional layers, which are trailed by different depth Maxpool layers. VGG has advanced depending on the past designs. VGG employs 3x3 channels all through the convolutional layers of the system. Each convolutional layer was trailed by a ReLU activation function. The VGG is a much deeper network that contains 16 weighted layers that perform better in ImageNet Large Scale Visual Recognition Challenge.

3.2 VGG With Two Dense Layers

In the above VGG model, we added two dense layers. These layers helped to improve the model by achieving better accuracy than the previous one. The dense layer is a deeply architectured and connected neural network layer. It is the most familiar and generally used layer. The dense layer performs the following operation on the given input and returns the wanted output.

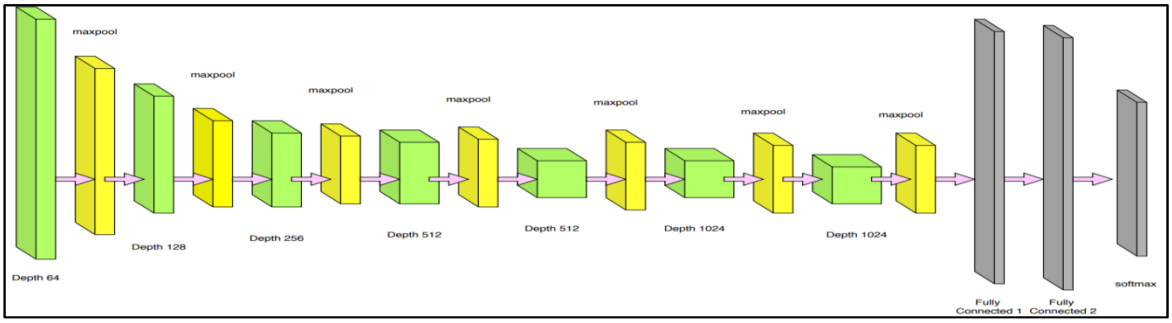

Figure 1: Architecture of modified VGG

As shown in the above figure in this model, the VGG model is modified by adding 2 dense layers in the end, followed up by fully connected layers and the output of which is passed through a softmax layer. This helps to expand the number of layers by expanding the model. The output is obtained after this layer.

3.3 ResNet

ResNet was a very efficient model in comparison to the ImageNet model. It was developed in 2015. ResNet contains a hundred or more layers and has become the best picture acknowledgment model in the PC vision network [4]. The architecture of the ResNet model is very complicated and contains many layers. The fundamental methodology of this model is that the information will pass a layer with a little yield of 1x1 at first. At that point, it will move to a layer of 3x3, and at that point, utilize a layer of 1x1 to deal with a progressively noteworthy number of highlights. Additionally, contrasted and the conventional CNN, for example, VGG, ResNet decreases the necessary parameters. In addition, ResNet can be advanced from top to bottom without any problem of angle scattering. ResNet was a huge advancement at that time when it was proposed because of its efficiency.

DOI : https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

3.4 Resnet With Two Dense Layers

We have added two dense layers to the above ResNet model. These layers helped to improve the model by achieving better accuracy than the previous one. The dense layer adds the nonlinearity property. Thus, they can model any mathematical function.

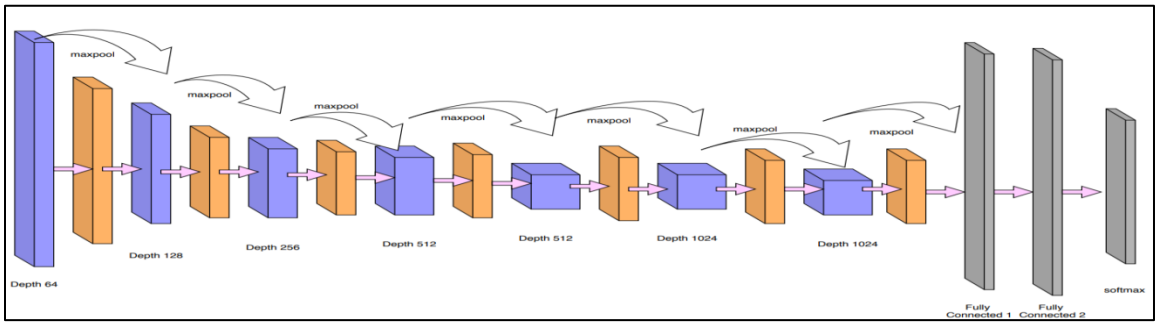

Figure 2: Architecture of modified ResNet

As shown in the figure, in this model, we have connected our input layer with 2 dense layers of different depths to change the depths of the model. Then the layer is further connected with two fully connected dense layers and the output of which is passed through a softmax layer. The output is obtained after this layer.

3.5 Merged Model

This model is a combination of both of the above models, VGG and ResNet. We have merged both these models into a single model to obtain better accuracy on our dataset FER2013. We used the Sequential model inception technique to connect the models, and finally, we used the Sequential merge model to merge both VGG and ResNet.

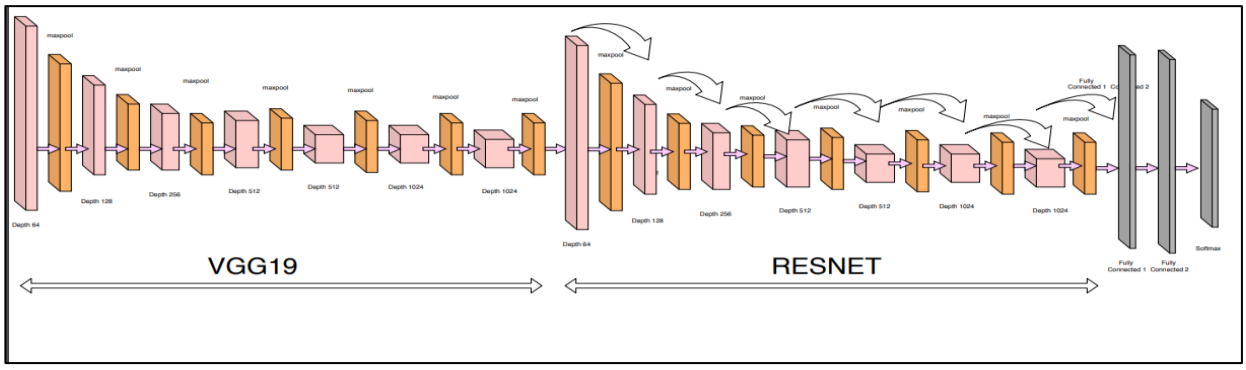

Figure 3: Architecture of Merged Model

Both the models are concatenated to each other, and the output obtained at the end of the ResNet layer is further passed through two fully connected layers. This is further connected to a softmax layer which provides the output.

DOI : https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

4. Experimental Result

4.1 Dataset

This dataset contains 35,887 gray photos of size 48 × 48: Of which 28,708 images are utilized for the purpose of training, 3,858 images are used for training (Public Test), and the remaining 3,588 images were used to perform the test (Private Test). Each photo contains a face of one of seven classes (0: Angry, 1: Disgust, 2: Fear, 3: Happy, 4: Sad, 5: Surprise, 6: Neutral). This dataset was prepared by Pierre-Luc Carrier and Aaron Courville, who was used as a part of their research work.

4.2 Results

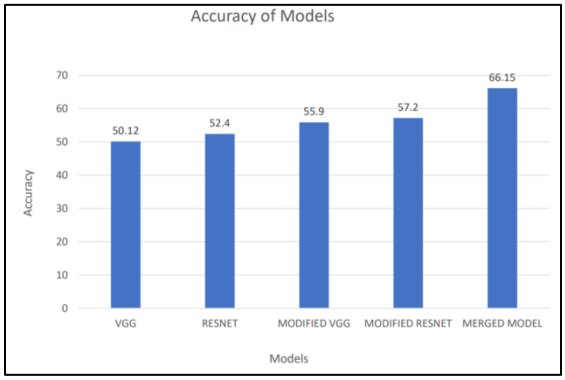

Figure 4: Bar Graph showing the accuracy of different Models

As shown in the above figure with different numbers of epochs, the above result is obtained. The best accuracy obtained is of the merged model of modified VGG with modified RESNET, which is 66.15, which is comparatively higher than the other models which were trained on the same dataset, i.e., FER2013 containing face images with labeled emotions in the dataset.

Confusion matrix or matrix error is a measure arranged in a table that allows the class to visualize the performance of tissue Figure. Each row of the matrix shows the results predicted by the model, while each column represents the actual value (Groundtruth) of data points. Confusion Matrix's name comes from the fact that it makes it easier to see whether the system is confusing two layers or not (that is often mislabeled as other classes). The result of this approach is a square matrix with size in each direction by the number of data layers.

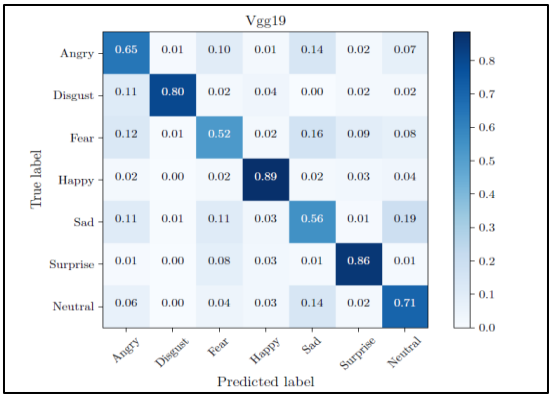

Figure 5: confusion matrix of VGG

DOI : https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

The VGG model shows the highest accuracy for Happy emotion with 89%, followed by Surprise with 86%, Disgust with 80%, Neutral with 71%, Angry with 65%, Fear with 23%, and the lowest accuracy for sad emotion with 21.4%.

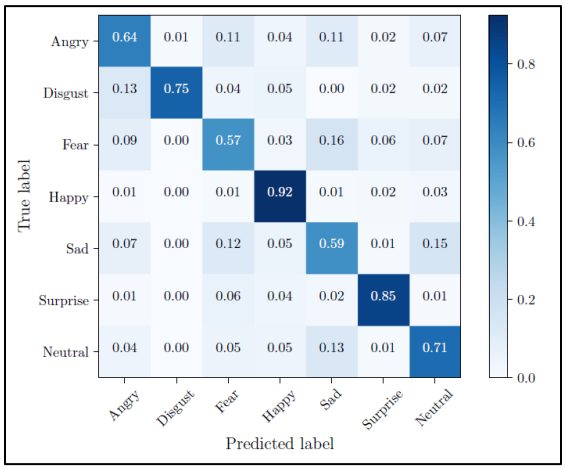

Figure 6: Confusion matrix of ResNet

The ResNet model shows the highest accuracy for Happy emotion with 92%, followed by Surprise with 85%, Disgust with 75%, Neutral with 71%, Angry with 64%, Fear with 21.05%, and the lowest accuracy for sad emotion with 20.33%.

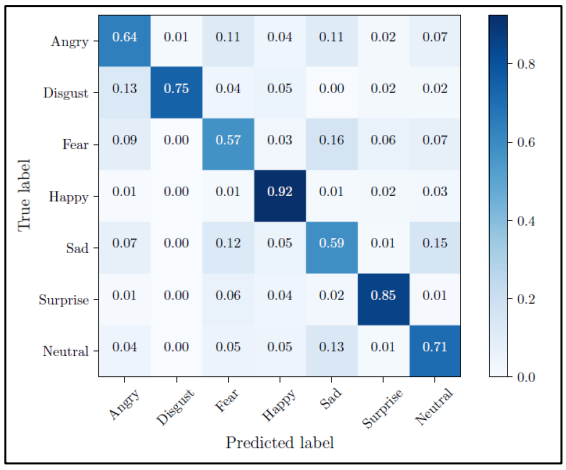

Figure 7: Confusion matrix of Merged Model

The Merged model shows the highest accuracy for Happy emotion with 90%, followed by Surprise with 86%, Disgust with 76%, Neutral with 73%, Angry with 65%, Sad emotion with 62%, and the lowest accuracy for Fear with 21.81%.

From our best model, i.e., the merged model, the following results were obtained :


Supposed to be: Surprised

DOI: https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

Found: 86% Surprised - 12% Neutral

Supposed to be: Angry

Found: 21% Sad - 65% Angry

Supposed to be: Happy

Found: 90% Happy - 4% Neutral

Supposed to be: Neutral

Found: 76% Neutral - 7% Angry

Supposed to be: Sad

Found: 62% Sad - 25% Fear

Supposed to be: Fear

Found: 21% Fear - 4% Neutral

Supposed to be: Disgust

Found: 9% Neutral - 76% Disgust

The comparative study between the different models on the FER2013 dataset is shown in Table 1.:

DOI : https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

Table 1: Comparison table of models

Model Name	Year	Salient Features	Accuracy	Total Layers	Epochs	Employed approach
VGG	2015	Fixed-size kernels	50.12	16	100	A.T. Lopes [1]
RESNET	2016	Shortcut Connections	52.40	50	75	S. Xie [11]
Modified VGG	2017	Wider-size kernels	55.90	18	70	Y. Zhang [9]
Modified RESNET	2017	Residual Connections	57.20	52	75	H. Li [4]
Merge Model (Our approach)	2020	Deeper Connections	66.15	67	150	Combination of VGG and Resnet

5. Conclusion

In this paper, we applied Deep Neural Network with five distinct models for facial expression recognition. At first, we examined the fundamental structures of DNN models, VGG and ResNet. Likewise, we determined and looked at the accuracy of each model. The five DNN models are ResNet, VGGNet, VGGNet with two dense layers, ResNet with two dense layers, and merge model (ResNet and VGG combined), and they are inspected on a similar database which is FER2013. FER2013 is one of the famous datasets. Since the dataset is enormous, it contains some disturbances. For the most part, the restricted outcomes we got are additionally reasonable for other general circumstances. Finally, we found that our merge model accomplished the best overall accuracy, which is 0.66. The consequences of our methodology show that DNN can create some valuable outcomes on FER. As FER is an incredible method to help individuals in regular day-to-day existence in numerous fields, later on, we will work to improve the accuracy of each DNN model. On the other hand, we will examine different factors that may affect accuracy, for example, the potential effect of FER2013. We look forward to finding more effective ways to solve the problems and issues.

Funding: "This research received no external funding."

Conflicts of Interest: “The authors declare no conflict of interest.”

DOI : https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020

References

[1] A. T. Lopes, E. D. Aguiar, and T. Oliveirasantos. A facial expression recognition system using CNN. In Graphics, Patterns and Images, pages 273–280, 2015.
[2] B. E. Bejnordi, J. Lin, B. Glass, M. Mullooly, G. L. Gierach, M. E. Sherman, N. Karssemeijer, J. V. D. Laak, and A. H. Beck. Deep learning-based assessment of tumor associated stroma for diagnosing breast cancer in histopathology images. In IEEE International Symposium on Biomedical Imaging, pages 929–932, 2017.
[3] C. F. Bobis, R. C. Gonza´lez, J. Cancelas, I. A´ lvarez, and J. Enguita. Face recognition using binary thresholding for features extraction. In International Conference on Image Analysis and Processing, page 1077, 1999.
[4] H. Li, H. Li, H. Li, H. Li, and H. Li. Does resnet learn good general-purpose features? In International Conference on Artificial Intelligence, Automation and Control Technologies, page 19, 2017.
[5] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, and D. H. Lee. Challenges in representation learning: A report on three machine learning contests. Neural Netw, 64:59–63, 2015.
[6] M. A. Imran, M. S. U. Miah, and H. Rahman. Face recognition using eigenfaces. Proc Cvpr, 118(5):586–591, 2002.
[7] Shen, Dinggang, Guorong Wu, and Heung-Il Suk. "Deep Learning in Medical Image Analysis." Annual review of biomedical engineering 19 (2017): 221–248. PMC. Web. 25 June 2018.
[8] Y. Tu, S. Li, and M. Wang. Intelligent facial expression recognition system r&c-fer. In Intelligent Control and Automation, 2008. Wcica 2008. World Congress on, pages 2501–2506, 2008.
[9] Y. Zhang, F. Chang, L. I. Nanjun, H. Liu, and Z. Gai. Modified alexnet for dense crowd counting. (cii), 2017.
[10] Yijun Gan. Facial Expression Recognition Using Convolutional Neural Network, 2018
[11] S. Xie, R. Girshick, P. Dollar, Z. Tu and K. He. Aggregated Residual Transformations for Deep Neural Networks. arXiv preprint arXiv:1611.05431v1,2016.
[12] Zhou, Yitao & Ren, Fuji & Nishide, Shun & Kang, Xin. (2019). Facial Sentiment Classification Based on Resnet-18 Model. 463-466. 10.1109/EEI48997.2019.00106.
[13] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.
[14] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Identity Mappings in Deep Residual Networks. 9908. 630-645. 10.1007/978-3-319-46493-0_38.
[15] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.
[16] Cheng, Shuo & Zhou, Guohui. (2019). Facial Expression Recognition Method Based on Improved VGG Convolutional Neural Network. International Journal of Pattern Recognition and Artificial Intelligence.

DOI : https://doi.org/10.54216/FPA.020101
Received: March 13, 2020 Revised: April 18, 2020 Accepted: May 10, 2020