International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 1923–1931

Facial Expression Recognition With Convolutional Neural Networks

Neelansh Gulati
Department of Electronics and Communication
Bharati Vidyapeeth's College of Engineering
New Delhi, India neelanshgulati100@gmail.com

Arun Kumar Dubey
Department of Information Technology Bharati Vidyapeeth's College of Engineering
New Delhi, India arudubey@gmail.com

Abstract

In this project,development of convolutional neural networks (CNN) for a facial expression recognition has been performed.The main objective involves segregation of each facial image into one of the seven facial emotion categories contemplated in this study from the fer2013 dataset.The process consists of preparing the dataset into a form which will work on a generalized algorithm,followed by face detection stage,in which the face is detected from the image that we wish to classify and then implementing the CNN algorithm to segregate input image into one of seven classes.We experimented with different models such as VGG16,VGG19 and ResNet50,achieving different accuracy for different models.

Keywords— Facial expression recognition,convolutional neural networks,resnet,deep learning, vgg.

I. INTRODUCTION

In order to accentuate various aspects of speech and to demonstrate response,individuals interconnect with one another via oration.One of the necessary approaches through which human beings express thoughts is via facial expressions which constitute very significant section of communication.

However nothing is said verbally,there is a lot that needs to be comprehended regarding the information that is exchanged via nonverbal message.Facial messages deliver telepathic sign,and have a vital function in social connections. Although people understand facial expressions practically without any attempt or gap,authentic expression recognition is nevertheless a question.

Due to significance of facial expression in communicating emotions, recognition of human emotions has become a topic of substantial research interest.There are few particular facial expressions that corresponds to comprehensive meaning. In 1971 paper titled "Constants Across Cultures in the Face and Emotion", Ekman et al. determined six facial expressions with comprehensive meaning.These expressions include sadness, fear, anger, disgust, happiness, and surprise[1]. Researchers aim to identify these emotions through computer vision.

Here,we offer a technique established upon CNN for recognition of facial expression.The input fetched into our system consists of an image,then prediction of facial expression label is performed through CNN.Expression label should fall under one of following class: disgust , happiness, sadness, fear,anger, and neutral.

One of the common problems encountered in visual classification is networks with stack of convolutional
layers.Association linking depth of the network[8][9] and classification performance have been analysed by few researchers because of the openness of public database and efficacious system.Networks are capable of capturing advance data known as high-level characteristic with the help of increasing number of convolutional layers.Networks incorporate mid-level and low-level aspects obtained from convolutional layer in order to obtain such information.Usually high-level feature is considered by classifier with hidden layers.VGG[10] is an example of such network as it comprises of certain blocks of convolutional layers.

ISSN: 2233-7857 IJFGCN

Keras has the following CNN which have been pre-trained on the ImageNet records:

- VGG16
- VGG19
- ResNet

A. VGG16 and VGG19

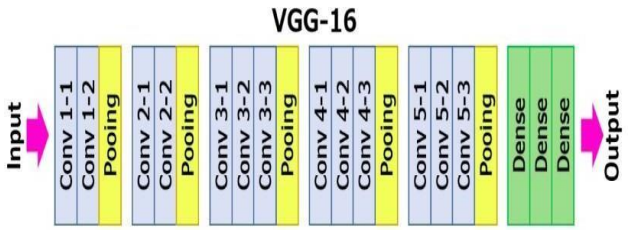


Figure. 1. VGG16 convolutional neural network model

Number 16 in the name VGG−16 refers to the fact that it has 16 layers that have some weights. This is a pretty large network, and has a total of about 138 million parameters,which is pretty large even by modern standards.However, the simplicity of the VGG16 architecture made it quite appealing and we can tell that this architecture is really quite uniform.There are a few convolutional layers followed by means of a pooling layer which reduces the height and width of a volume.In this architecture,the number of filter we use is roughly doubling on every step or doubling through every stack of convolutional layers and that is another simple principle used to design the architecture of this network.The main downside was that it was a pretty large network in terms of the number of parameters to be trained.VGG19 neural network is bigger than VGG16, but because VGG16 does almost as well as the VGG19,a lot of people use VGG16. Primary disadvantage with VGGNet are as follows:

- It is extremely sluggish to train.
- Large network architecture.

B. ResNet

ResNet is not like traditional sequential network architecture such VGG and AlexNet, instead it is a kind of “exotic architecture” based upon micro-architecture modules.This expression indicates the collection of smaller chunk required to establish the mesh.A group of these chunks and other standard layers causes the macro-architecture.

Proposed by He et al,the ResNet architecture is significant in displaying that highly deep networks are usually trained with the assistance of standard SGD via residual modules:

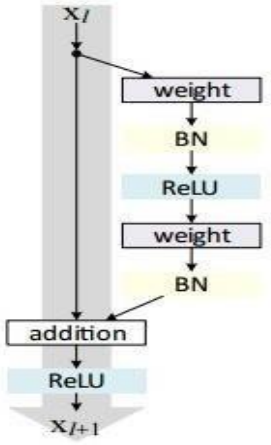
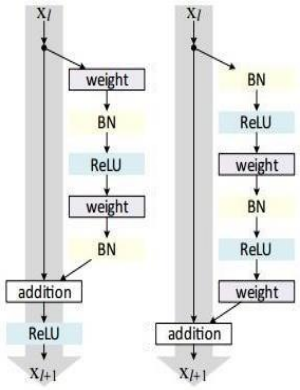


Figure. 2. ResNet original residual module proposed by He et al

Updating residual module to use identity mapping helps in achieving more accuracy.



(a) (b)

Figure. 3. (a) ResNet original residual module (b) ResNet updated residual moduleAlthough ResNet is profoundly deep compared to VGG16 andVGG19, model measurement is extensively minor due to the deployment of global average pooling instead of fully-connected layers.

II. RELATED WORK

In the year 1971,Ekman and Friesen presented several elemental emotions,namely: fear, happy,disgust, angry, surprise, sad, and neutral.Emotions can be detected by examining facial component expression.

Facial Action Coding System(FACS)[2] is the basic approach through which classification of facial expression can take place.Optical information of facial muscles are contained in FACS,also called action units.It is pretty direct to observe that majority of picked facial parts are present near the nose,eyes or mouth.Typically,muscles surrounding these areas are inclined to energize in order to demonstrate one's emotions.Consequently,experts are particularly stressing on determining and examining these areas for recognition of emotion.CK,CK+, and MMI are amongst the few public datasets supplying facial images with equivalent AU's.Facial expression and AUs are well stimulated and easily identifiable as these images were captured in the lab environment.The remarkable design process on CK and CK+ databases is CNN with four Inception modules[3].Initially, this consists of distinctive dimensions of kernel was launched in order to define huge picture classification issue.Rather than considering semantic net,Shan et al proposed an approach which covered SVM and strength handbuilt face function LBP in order to attain an aggressive efficiency on MMI database.To deal with restrictions mentioned above,different databases had been presented with unconstrained circumstance like DISFA[11] or AM- FED[4].

As information contained in these databases is minute, normal studying strategies can obtain related outcomes[4][5].

Standard computer inspired techniques with handbuilt characteristic is capable of attaining fierce outcome on information obtained in stimulated conditions,i.e., nose, mouth, and eyes which are determined via easy estimations.Experts have made an attempt to resolve the emotion recognition in the wild with the help of technique known as transfer leaning and their approach did great in emotion classification,but overall performance of this approach was still an issue.

In FER2013 venture of ICML 2013[6],Tang brought a CNN mutually realized with linear SVM for facial expression recognition[12].Using these in region of softmax classifier,their approach surpassed others and won the first vicinity.Encouraged by means of the triumph of GoogLeNet, Mollahosseini et al. has put forward a structure which consist of four inception modules.In 2016, Zhou et al. put forward multi-scale CNNs[7].This mannequin had three different networks with variable input load.Additionally, late fusion was incorporated in order to acquire ultimate grouping outcome.By merging

more than one CNNs and editing loss function, Yu et al. attained greater performance in comparison of prior approaches. Multi-scale Convolutional Neural Network was put forward by Wang et al,in which the they utilised absolute feature maps in the network for classification.Nonetheless,the usage of all generated facets barring determination may decrease the complete efficiency owing to insignificant data in shallow layers.An enhancement on facial expression recognition was attained with the aid of incorporating CNN and SIFT attributes.The efforts of Al Shabi et al. is the contemporary today's approach on FER2013 dataset.

III. METHODOLOGYalgorithm and generate efficient results. In the face detection stage, the face is detected from the image that we wish to classify. The emotion classification step consists of implementing the CNN algorithm to segregate input image into one of the seven classes.

- Normalization - Normalization of an image is done to remove illumination variations and obtain improved face image.
- Grayscaling - Grayscaling is the process of transforming a colored image input into an image whose pixel value relies upon the intensity of light on the image. Grayscaling is done as colored images are difficult to process by an algorithm.

A. Dataset

The dataset deployed for implementing was the FER2013 dataset from the Kaggle challenge on FER.The dataset contains 35,887 gray-scale images out of which 28,709 are for training purposes, 3589 for testing, and the rest for testing.Images in the FER2013 dataset comes under one of the seven categories,namely:neutral,happy,fear ,surprise,disgust, angry,and sad.

Emotion labels in the dataset:

- 0:-4593 images-Angry
- 1:-547 images-Disgust
- 2:-5121 images-Fear
- 3:-8989 images-Happy
- 4:-6077 images-Sad
- 5:-4002 images-Surprise
- 6:-6198 images-Neutral

Figure. 4. Sample images from the FER2013 dataset

B. Process of Facial Expression Recognition

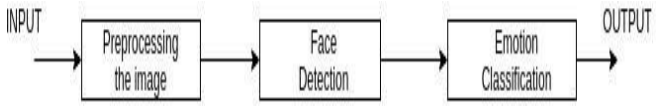


Figure. 5. Process Flow of Facial Expression Recognition

Facial Expression Recognition has three stages. The first stage which is the preprocessing stage consists of preparing the dataset into a form which will work on a generalized

Figure. 6. Grayscaling of images in FER2013

- Resizing - The image is resized to remove the unnecessary parts of the image. This reduces the memory required and increases computation speed.

C. *Face Detection*

Input of convolutional neural network consists of an image and the output is returned as set of probabilities related to seven distinct emotion categories.Generally the size of input image fetched to a CNN that are trained on ImageNet are of size 224×224,227×227,256×256, and 299×299; however,other dimensions can be considered as well.Standard preprocess input from Keras was assigned as our preprocess function.

D. *Emotion Classification*

In this step, the system classifies the image into one of the seven universal emotions – Neutral,Disgust ,Surprise, Anger, Happiness, Fear, and Sadness as labelled in the FER2013 dataset.The dataset was first divided into training and test datasets, and then it was trained on the training set.The approach followed was to experiment with different architectures on the CNN.We used pre-trained network architecture weights from disk and illustrate the model and the Convolutional Neural Network is then instantiated using the pre-trained ImageNet weights.After that we fetched the input image from the disk using OpenCV and display the image to our screen.

IV. EXPERIMENT AND RESULTS

In this project,we focused on facial expression recognition and focused to segregate facial images into one of distinct emotion class with the help of CNN (VGG19,VGG16,ResNet) pre-trained on the ImageNet database using Python and Keras deep learning library employing FER2013 database.

A. *Loss and accuracy over time*

The graphs below presents the loss and accuracy with each epoch and it can be observed that as the loss decreases,and the accuracy increases with each epoch.The training versus

testing curve for loss remains ideal over the first forty epochs, after which it begins to deviate from the ideal values.

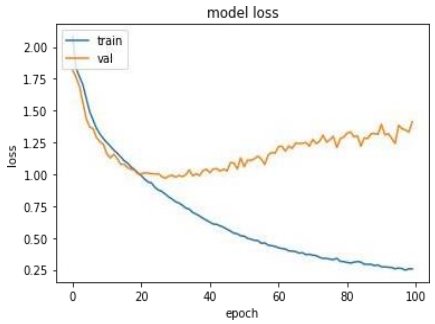


Figure. 7. Graph of training and validation loss per epoch

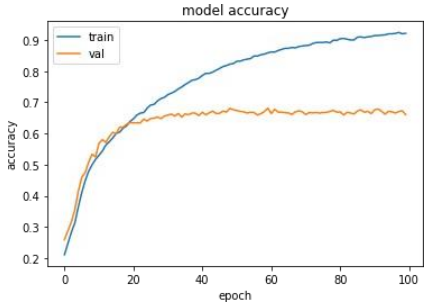


Figure. 8. Graph of training and validation accuracy per epoch

B. *Confusion matrix*

The confusion matrix generated over the test data is shown in figure 9.The blocks along the diagonal show that the test data has been classified well and it can be observed that the number of correct

classifications is low for disgust.The numbers on either side of the diagonal represent the number of wrongly classified images.As these numbers are lower compared to the numbers on the diagonal, it can be concluded that the algorithm has worked correctly and accomplished advanced level results.

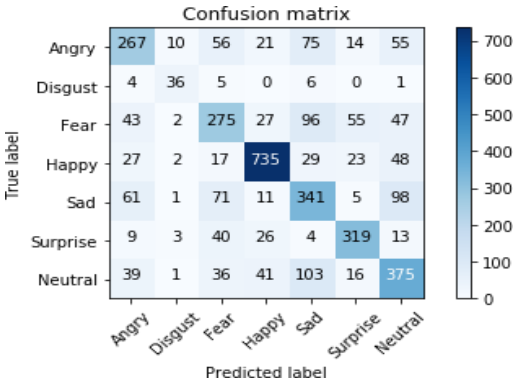


Figure. 9. Confusion matrix represented as a heatmap

C. Contrast with the advanced methods

TABLE 1. Various method and their respective accuracy on FER2013 dataset

Method	**Accuracy in percentage**
Shen et al.[13]	61.86
Ergen et al.[14]	57.10
VGG16	63.07
VGG19	61.16
ResNet	57.07

To assess the performance of the proposed models with advanced models,table 1 lists down the accuracy obtained by distinct methods on the FER2013 dataset.VGG16 model achieved accuracy of 63.07%,which tops the table, and VGG19 model ranks #3 in the list.While ResNet model has almost similar accuracy compared to the model ranked #4 in the list.

V. CONCLUSION/FUTURE WORK

Our main focus was recognize facial expression with CNN and we focused to segregate facial images into one of distinct emotion class.We experimented with CNN models on the FER2013 dataset for facial expression recognition and further estimated theirefficiency.The results shows that VGG16 has a better performance than the other proposed models,while results show that VGG19 model could also achieve decent accuracy in facial expression recongnition.In future work,we will try to implement hybrid features to improve the model accuracy along with depth analysis about the face classification problem.

REFERENCES

1. P. Ekman and W. V. Friesen. Emotional facial action coding system. Unpublished manuscript, University of California at San Francisco, 1983.
2. K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in Int. Conf. Comput. Vis. Pattern Recognition (CVPR) (Las Vegas, USA, 2016), pp. 770–778.
3. A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, Int. Conf. Neural Information Processing Systems (NIPS) (Lake Tahoe, Nevada, USA, 2012), pp. 1097–1105.
4. K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image

recognition, Int. Conf. Learning Representations (ICLR) (San Diego, California, USA, 2015), pp. 1–14.
5. P. Ekman and W. V. Friesen, Facs - facial action coding system (1977), https://www.cs. cmu.edu/ face/facs.htm.
6. A. Mollahosseini, D. Chan and M. H. Mahoor, Going deeper in facial expression recognition using deep neural networks, Winter Conf. Applications of Computer Vision (WACV) (Lake Placid, NY, USA, 2016), pp. 1–10.
7. [S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh and J. F. Cohn, Disfa: A spontaneous facial action intensity database, IEEE Trans. Affective Comput. 4(2) (2013) 151– 160.
8. D. McDuff, R. el Kaliouby, T. Senechal, M. Amr, J. F. Cohn and R. Picard, Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected \ in-the-wild", Int. Conf. Computer Vision and` Pattern Recognition (CVPR) - Workshops (Portland, Oregon, USA, 2013), pp. 881–888.
9. X. Zhang, M. H. Mahoor and S. M. Mavadati, Facial expression recognition using lp-norm mkl multiclass-svm, Mach. Vis. Appl. 26 (2015) 467–483.

10. P. Carrier and A. Courville, Challenges in representation learning: Facial expression recognition challenge (2013), https://goo.gl/kVzT48.
11. Y. Tang, Deep learning using linear support vector machines, Workshop on Challenges in Representation Learning, International Conference on Machine Learning (ICML) (Atlanta, USA, 2013).
12. S. Zhou, Y. Liang, J. Wan and S. Z. Li, Facial expression recognition based on multi-scale cnns, in Chin. Conf. Biometric Recognition (Chengdu, China, 2016), pp. 503–510. K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in Int. Conf. Comput. Vis. Pattern Recognition (CVPR) (Las Vegas, USA, 2016), pp. 770–778.
13. G. Zeng, J. Zhou, X. Jia, W. Xie, L. Shen, Hand-crafted feature guided deep learning for facial expression recognition, in: Proceedings of the 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), 2018, pp. 423–430, doi: 10.1109/FG.2018.0 0 068 .
14. Tumen , O.F. Soylemez , B. Ergen , Facial emotion recognition on a dataset using convolutional neural network, in: Proceedings of the Artificial Intelligence and Data Processing Symposium, 2017, pp. 1–5

ISSN: 2233-7857 IJFGCN