A STUDY OF DIFFERENT CONVOLUTION NEURAL NETWORK ARCHITECTURES FOR HUMAN FACIAL EMOTION DETECTION

Logeswaran Sivakumar[1], Jayanth Adithya[2], Dr. N. Rajeswari[3]

[1]*Student, Dept. of Electronics and Communication Enginnering, SVCE, TamilNadu, India*
[2]*Student, Dept. of Electronics and Communication Enginnering, SVCE, TamilNadu, India*
[3]*Associate Professor, Dept. of Computer Science Engineering, SVCE, TamilNadu, India*

Abstract – *Human facial expression detection is a hot research topic in the field of computer vision. In this paper the images of random human faces are used to detect the Facial emotion using different Convolutional neural network models. We have proposed a 17 layered sequential model and have compared it to the different CNN architectures. This comparison is based on the Model accuracy, model losses and the number of layers. This paper also discusses the application of Facial emotion detection in hospitals for monitoring patients.*

Key Words: CNN (Convolutional neural network), ResNet (Residual neural network), VGG(Visual geometry group), MobileNet, DenseNet, Accuracy , Model loss.

1. INTRODUCTION

The main motive of Facial emotion recognition is to provide computers the skill to identify and recognize different emotions in real time. Facial emotion recognition has several applications ranging from E-commerce to enabling safe driving. The facial detection model uses FER 2013 dataset provided by Kaggle that is divided into three sets namely Train, Validation and Testing. The FER 2013 dataset by Pierre- Luc Carrier, contains 32298 examples that have been split into 28709 as training examples in order to train our CNN model and 3589 used as testing examples. Each sample contains 48x48 greyscale image. The train and test sets contain Emotion attribute and pixel attribute. The emotion attribute contains seven numbers where each number signifies a particular emotion. The seven emotions are anger, disgust, fear, happiness, sadness, surprise and neutral. The pixel column contains a string of numbers representing pixel values in row major order. There have been several architectures proposed in the past decade in order to recognize faces namely ResNet, inceptionv3, VGG16, Densenet. In this paper, these architectures are compared based on score, loss, bias and output at each nodes with structured tables and graphs. These models have various real time applications such as vehicle detection in self-driving cars, emotion recognition of patients and real time feedback systems. The key focus of this paper is to design a novel approach for monitoring patients using Facial emotion recognition.

1.1 Significance of Facial Emotion Detection

Facial expressions and other gestures convey non-verbal communication cues that play an important role in interpersonal relations. These cues complement speech by helping the listener to interpret the intended meaning of spoken words. Therefore, facial expression recognition extracts and analyzes information from an image or video feed and delivers unfiltered and unbiased emotional responses as data.

2. Architecture design and Implementation

2.1 ResNet (Residual network)

ResNet created by Microsoft is a classical neural network that is widely used for computer vision. ResNet uses global average pooling rather than fully connected layers and hence it has a smaller modal size of 100mb. ResNet introduced the concept of skip connection where previously the convolutions layers are stacked one after the other but in this skip connection the convolutional layers are stacked one after the other as before but we now add the original input to the layer. ResNet allows offsetting vanishing gradient and skipping the less relevant layers. In this paper we shall analyze the ResNet model that does not include the fully connected layer at the top of the network and weights are randomly initialized for Emotion detection.

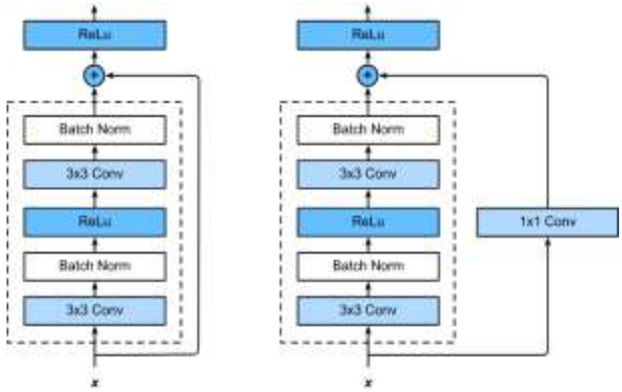

Fig-1: ResNet Model Architecture

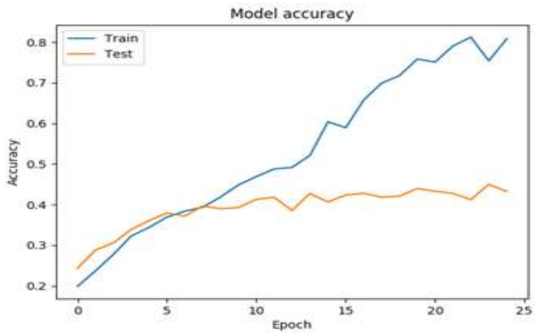

Chart-1: ResNet Model Accuracy

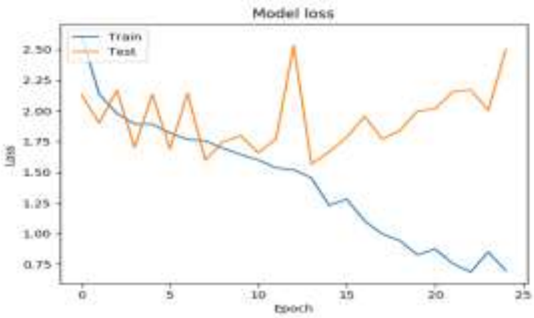

Chart-2: ResNet Model Loss

2.2 VGG19 (Visual geometry group)

Simonyan and Zisserman of University of Oxford created VGG19 having 19 layers, 16 conv and 3 fully connected. VGG19 has 138 Million parameters. VGG19 is trained on more than a million images and can classify into 1000 object categories. In this paper the VGG19 model used does not use the fully connected layers and weights are randomly initialized. Global average pooling is added after the base model of VGG19 and the dense layer is activated using soft max activation.

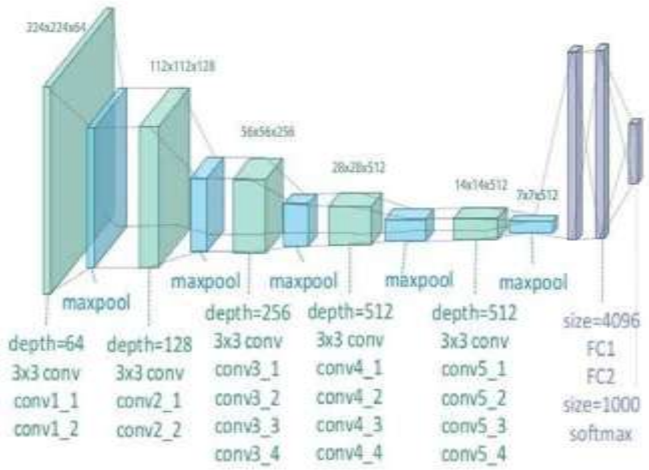

Fig-2: VGG19 Model Architecture

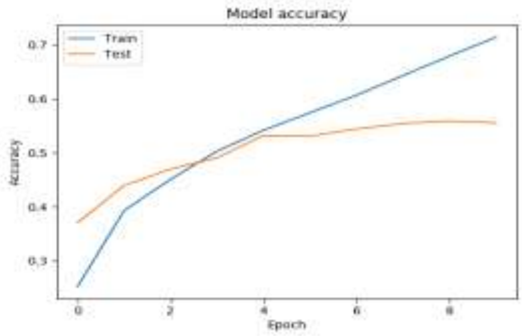

Chart-3: VGG19 Model Accuracy

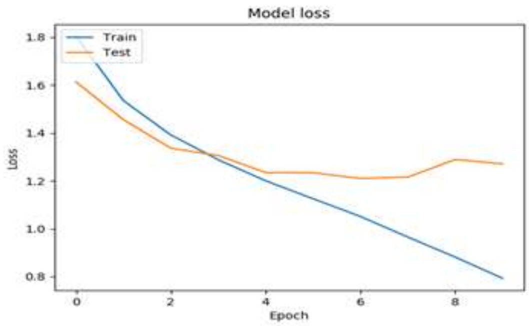

Chart-4: VGG19 Model Loss

2.3 MOBILE NET

Mobile Net proposed by Google is more suitable for mobile and embedded based vision systems. The full Mobile Net network has 30 layers. In the structure first layer is followed by depth wise layer which is followed by point wise layers. The default input size of this model is 224x224. When compared to normal convolutions the number of parameters are significantly less as the architecture uses depth wise separable convolutions, resulting in a light weight deep neural networks. MobileNets tradeoff between latency, size and accuracy. The depth wise convolutional layer convolves each channel separately and later on combines all the channels reducing the computational work.

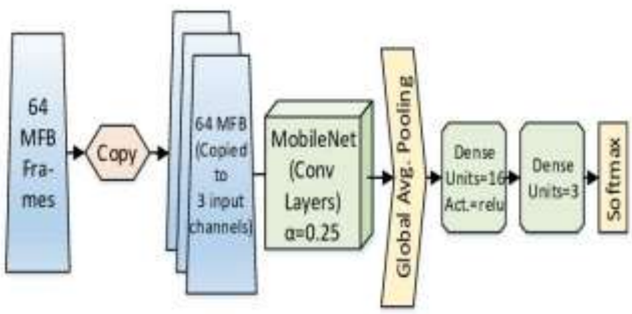

Fig-3: MobileNet Model Architecture

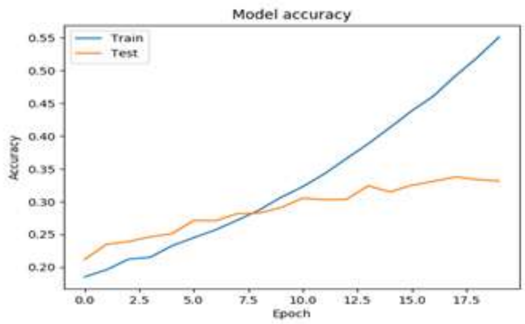

Chart-5: MobileNet Model Accuracy

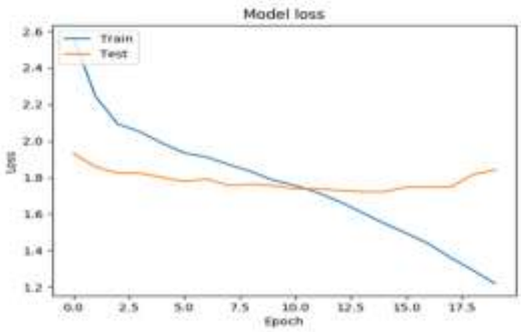

Chart-6: MobileNet Model Loss

2.4 DENSE NET

Dense Net is widely used for visual object recognition. DenseNet is similar to ResNet. DenseNet concatenates output from the previous layers instead of using the summation as used in ResNet. At the last layer the number of features is greater than that of ResNet although having the same width and height. DenseNet architectures uses batch normalization and drop out to bring all the feature values around the same range. In addition to batch normalization and drop out DenseNet uses transition layers in order to reduce the size of the image retaining the number of features. With this approach, Dense Net improves the flow of information and gradients throughout the network, which makes them easy to train.

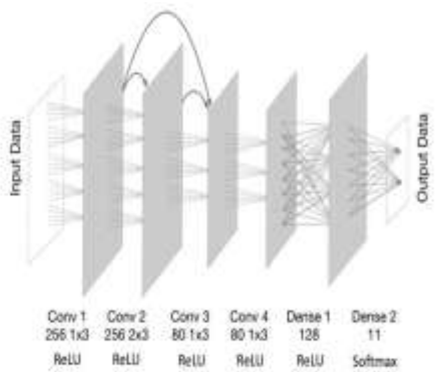

Fig-4: DenseNet Model Architecture

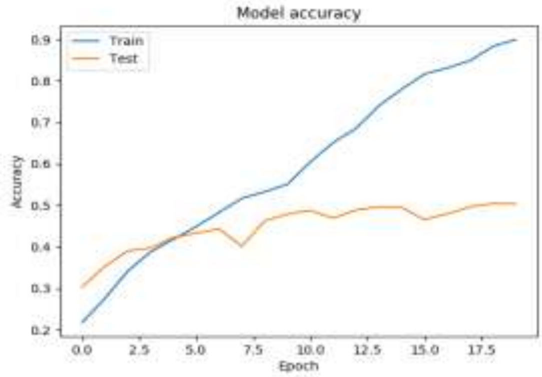

Chart-7: DenseNet Model Accuracy

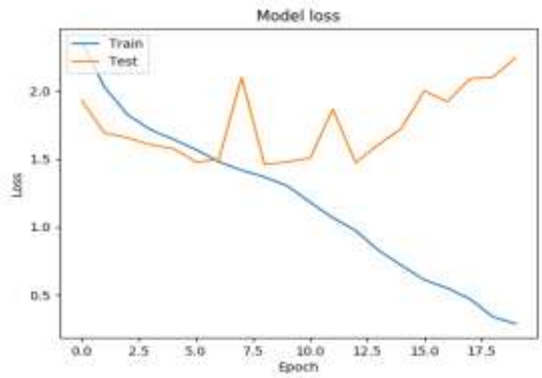

Chart-8: DenseNet Model Loss

3. OUR MODEL

The model proposed by us is a sequential model that contains 17 layers. These layers can be categorized on the basis of the function they perform as Conv2D layer, Dropout layer, max pooling layer and dense layer. The total number of parameters are 1,914,951 which are all trainable parameters. Model accuracy given by the model proposed by us is greater than that of all the previous models. Model loss is also low compared to the other models. Our model predicts fear and neutral emotions accurately.

Fig-5: Our Model Result

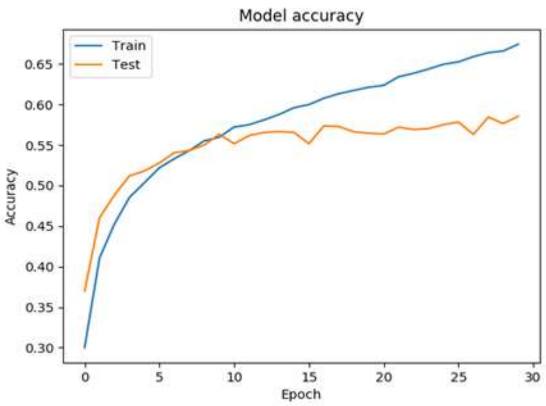

Chart-9: Our Model Accuracy

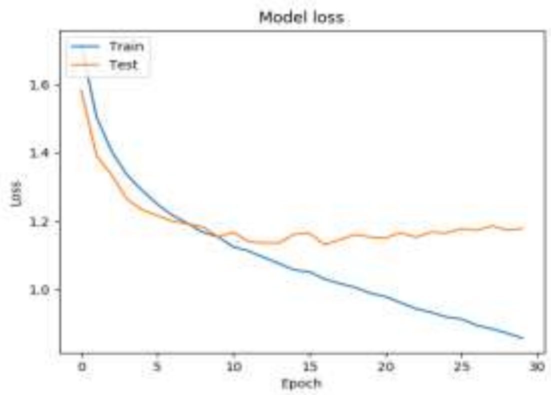

Chart-10: Our Model Loss

4. COMPARISON OF THE MODELS

Table -1: Comparison of different models

Model Name	Accuracy	Loss
ResNet	0.42	1.55
VGG19	0.55	1.3
MobileNet	0.34	1.8
DenseNet	0.49	1.5
Our Model	0.69	1.2

Lower Loss number and higher Accuracy number indicate better performance of the models.

5. APPLICATION

Facial expression detection has several applications like making driving safer and personalized by continuous driver monitoring, facial emotion detection during interviews helps test the candidate's attitude, Video game testing and market research. In this paper we propose a model for patient monitoring. In a densely populated country like India where hospitals are always crowded patients who require support 24/7 are abundant. Constant monitoring of this type of patient is highly not possible. With respect to highly populated countries, the number of patients who need to be under constant support are far greater than the number of people who can provide medical aid. We propose an idea that can reduce the amount of time, effort the hospital staff have to spend monitoring a patient. For instance, a person who needs constant support is recorded and monitored with the help of a camera that is present inside the room. The patient's condition is monitored with the help of a facial emotion detection system. An alert could be given to the medical team when the facial emotion detector detects Fear or shock in the patients face. The process begins with the camera recording the person using a Pi camera that captures images of the patient every 5 secs. The Facial emotion recognition system classifies the type of expression on the patient's face. The classification is performed by feeding into our sequential model. The Facial emotion recognition system notifies the medical team if the emotion that was detected is Fear or Shock. Alerting the medical team can be done using Bluetooth devices at two ends namely at the patient's room and the medical team's room.

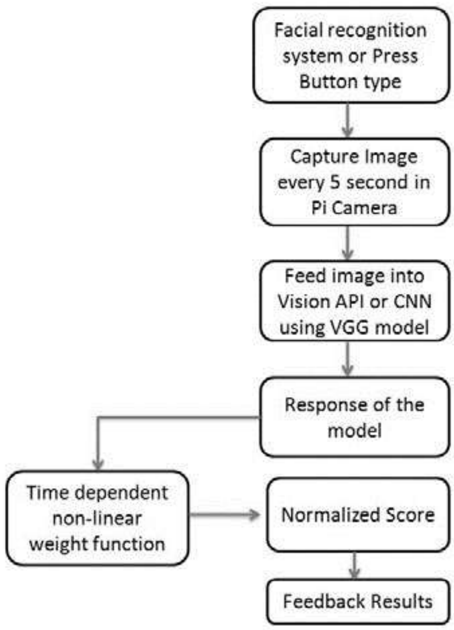

Fig-6: Application flow chart

6. CONCLUSION

In this paper we have focused on designing a Facial emotion recognition system using different CNN architectures. A facial emotion recognition system can be used as a feedback to enhance business and customer service and monitor patients in hospitals. An alternative model architecture having better accuracy was proposed by us and has been compared with different CNN architectures. The comparison was based on model accuracy and model losses and the number of layers used in the model. The paper also discusses a novel use of facial emotion recognition to aid patients, medical team and doctors.

REFERENCES

[1] Hadi Amiri et al. "Learning Text Pair Similarity with Context – sensitive Autoencoders", ACL 2016.
[2] Andre Teixeira Lopes et al. "A Facial Expression Recognition System Using Convolutional Networks", Vol. 00, pg. 273 – 280, 2015.
[3] Arushi Raghuvanshi and Vivek Choksi, "Facial Expression Recognition with Convolutional Neural Networks", CS231n Course Projects, winter 2016.
[4] Dennis Hamester et al., "Face Expression Recognition with a 2-Channel Convolutional Neural Network", International Joint Conference on Neural Networks (IJCNN), 2015.
[5] Siyue Xie and Haifeng Hu, "Facial expression recognition with FRR – CNN", Electronic Letters, 2017, Vol. 53 (4), pg. 235 – 237.
[6] Nima Mousavi et al. "Understanding how deep neural networks learn face expressions", International Joint Conference on Neural Networks, July 2016.
[7] Razavian et al. "CNN Features off-the-shelf: an Astounding Baseline for Recognition", arXiv: 1403.6382v3 [cs.CV], May 2014.
[8] Michael J. Lyons et al. "Coding Facial Expressions with Gabor Wavelets", 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998).
[9] Gary B. Huang et al. "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments." University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.
[10] Francois Chollet, Keras, GitHub, https://github.com/fchollet/keras, 2015.
[11] Alex Krizhevsky et al. "ImageNet Classification with Deep Convolutional Neural Networks", Neural Information Processing Systems (NIPS), 2012.