FACIAL EMOTION RECOGNITION SYSTEM

Megha Singh[1], Shiva Kant Sharma[2], Shouhaddo Paul[3], Jithin P Sajeevan[4], Sandarya Paul[5]

[1,2] Student, Computer Science & Engineering, Acharya Institute of Technology/ VTU, Karnataka, India
[3] Student, Computer Science & Engineering, Brindavan College of Engineering/ VTU, Karnataka, India
[4] Research Scientist & Project Coordinator, LRDC, Lithoera Technologies Pvt Ltd, Karnataka, India
[5]Project Supervisor, Department of Computer Science, Lithoera Technologies, Karnataka, India

ABSTRACT

This digital era involves technology for everything, right from day to day communication to doing complicated stents which even humans cannot do. Communication among humans is done by speaking, using sign languages, displaying emotions, or by body gesturing. To do any kind of communication or to express feelings, the most important role is played by expressions. Facial expressions are the most effective way to express one's emotions or feelings. Hence it becomes one of the most important and interesting fields to work with. Though it has always been a very easy task for humans to recognize emotions, it is quite challenging for computers to do so. With the advancement in ML and CV, it is now possible to recognize the emotions from images or live videos. Here, the technique used is called facial emotion recognition using convolutional neural networks (FERC). Using an aggregate of modern deep CNN and FER2013 [1] dataset, the test accuracy of 67.7% was obtained, outperforming many previous works without requiring additional training data or face registration.[2].

Keyword: - *Convolutional neural networks(CNN), Emotion recognition (ER), Facial emotion recognition(FER), Machine Learning(ML), Computer Vision (CV)*

1.Introduction

Facial motion plays a vital role in expressing these emotions. The muscles of the human face can be changed to communicate different feelings. Human beings can easily recognize these signals even if they are slightly displayed, by simultaneously processing information acquired by our other senses. Based on psychological studies, it can be assumed that human emotion perception follows a similar trend. It was possible to correctly highlight the emotion with 67.68% accuracy. The two-level CNN works in series, and here the last layer of perceptron adjusts the exponent values and weights with each iteration. There are seven types of human emotions shown to be universally recognized across different cultures: happiness, sadness, anger, disgust, fear, surprise, contempt; even for the most complicated expressions where a mixture of emotions could be used as descriptors, a cross-cultural agreement is still observed. Such an advancement could bring applications in many fields like security, marketing, medicine, and entertainment. The task of emotion recognition can be challenging for two reasons:
1) There does not exist a large database for training images.
2) depending on whether the input image is static, classifying emotion can be difficult.
The recent issue is particularly difficult for real-time detection where facial expressions vary dynamically. Provide with the computational need and complexity of a CNN, optimizing a network for efficient computation for frame-by-frame classification is necessary [3]. Also, reckoning for variations in subject position lighting in the environment is challenging.

2.The Facial Expression Recognition (FER-2013) Dataset

The dataset contains 35,685 examples of 48 x 48 pixel grayscale images of human faces through which it classifies facial expressions. The dataset contains 7 classes where images are categorized based on their emotions [2]. 48 x 48 pixels of grayscale images are contained in the data [4], the human faces have been automatically inscribed so that the frontal face should be centered and should occupy the same amount of space in each picture.
The objective was to perform a gig and categorize each human face stationed according to their respective emotions shown in the facial expression in one of the 7 categories {0: neutral, 1: Disgust, 2: Happy, 3: Fear, 4: Sad, 5: Surprise, 6: Angry}. The training folder from the dataset contains 7 following subfolders which further contain respective images. The training set contains 28,709 images whereas the public leaderboard test set consists of 3,589 images. The final test set on which the winner of the competition was evaluated, had 3589 more images..

3. METHODOLOGY

First of all data preprocessing was done, then the architecture was selected(here CNN) and then a model was proposed, compiled, fitted. Then the parameters were tuned and the model was tested. The complete code has been archived in the Facial-Emotion-Recognition-System repository as ipynb files for further research and analysis for researchers [5].

3.1Data Pre processing

Digital image processing is a manner in which computer algorithms execute image processing on digital images. It contains a variety of algorithms to be applied for the input data - the digital image processing focuses on improving the image info(features) by extinguishing undesirable distortions and/or improvement of some crucial image features so that our Artificial Intelligence-Computer Vision models can profit from this improved information to work on. The height and width of the image have been shifted to get better convergence. The image has been rescaled to normalize the image pixels between [0,1] as it will lead to faster execution. The resolution of images is fixed to 48 x 48 pixels.

3.2 CNN Architecture

Convolutional neural networks are a class of deep [3] learning, in the deep neural networks which are maximum commonly applied to evaluate visual imagery [6]. They are also acknowledged as transfer invariant or space invariant artificial neural networks, based on their contributed-weights architecture and rendition invariance characteristics.

3.2.1 Proposed model

The input to the network is an image of dimensions (48, 48, 1). The first two layers have 64 channels of 3*3 filter size and same padding with activation "relu", Then the whole batch is normalized and Then after a max pool layer of stride (2, 2), and later 50% of the neurons are dropped as to optimize and regularize the algorithm and help reducing variance. It is then stacked to 3 more layers with properties as before but with numbers of filters following 128, 256, 512. After the stack of convolution, batch normalization,max-pooling layer, and dropouts, a (2, 2, 512) feature map was obtained. The output is flattened to make it a (1, 2048) feature vector. After this there is 4 connected layer with dropouts of 50% after each layer, the first layer takes input from the last feature vector and outputs a (1, 512) vector and drops 50% of the neurons, the second layer also outputs a vector of size (1, 256), the third layer gets us (1,128) but the fourth layer output a 7 channels for 7 classes offer dataset, then after the output of 3rd fully connected layer is passed out to the softmax layer to normalize the classification vector. After the output of classification vectors the top-5 categories for evaluation. All the hidden layers primarily use the Rectified Linear Unit as its activation function [7].

3.3 Compilation and fitting the model

3.3.1 Optimizers

Optimizers well-organized the loss function and model parameters by updating the model in response to the output of the loss function. In simple words, they cast out our model into its most accurate and exact form by futzing with the weights [8]. The loss function is the counselor to the grade, telling the optimizer when it's moving in the right or wrong direction.

3.3.2 Adam

Adaptive moment estimation is abbreviated as adam and is another way of using past gradients to calculate current gradients. Adaptive moment estimation also utilizes the concept of momentum by adding fractions of previous gradients to the current one. This optimizer has become quite widespread and is practically accepted for use in training neural nets. Adam's Algorithm

1. Momentum is combined with RMSprop [9] to get Adam optimizer.
2. It Cares an exponentially weighted average of post gradients and stores it in variable V.
3. It calculates an exponentially weighted array of the sequence of the post gradients and stores it in variable
4. It updates parameters in a direction on counting information from '1' to '28'[10].

3.4 Hyper parameters tuned in Adam

$Vdw \leftarrow \beta1.Vdw + (1-\beta1)dw$, $Vdb \leftarrow \beta1.Vdb + (1-\beta1)db$ #momentum β1
$Sdw \leftarrow \beta2.Sdw + (1-\beta2)dw\hat{}2$, $Sdb \leftarrow \beta2.Sdb + (1-\beta1)db$ #Rmsprop β2

Adam works by first stabilizing the stochastic estimate of the gradient by taking a moving average, and then rescaling the step by a moving average of the standard deviations of each dimension. You can view the latter as the inverse of an adaptive multiplier on the learning rate for that dimension: a big standard deviation estimate means a low learning rate. A very small standard deviation estimate means a very large learning rate.
The following are the parameters tuned for our model:
Learning_rate=0.0005,beta_1=0.9, beta_2=0.999,epsilon=1e-06
Epsilon for numerical stability ,beta_1 =dw , beta_2 =dw^2

5. RESULT AND PREDICTION

After150 epochs the accuracy obtained on the training data was 66.4 % and in test data was 67.6 % using the CNN architecture.Fig 4.1 shows the relationship between loss and value loss and Fig 4.2 shows the accuracy and value accuracy [11]. Fig 4.3 shows the accuracy obtained by our model. Fig 4.4 shows the output generated by our model.

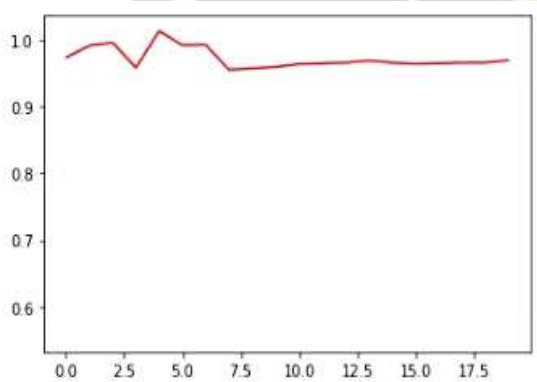

Fig -4.1 : val_loss(red)

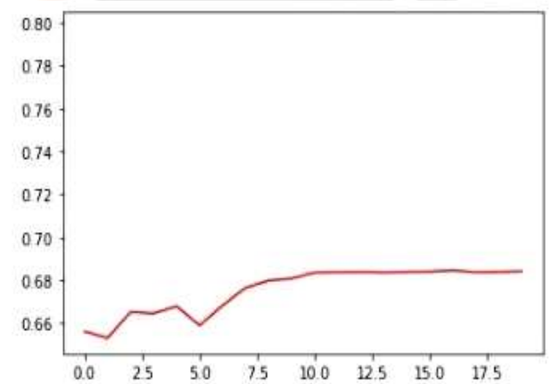

Fig -4.2 : val_accuracy(red)

```
448/448 [==============================] - 41s 91ms/step - loss: 0.9188
- accuracy: 0.6648 - val_loss: 0.8826 - val_accuracy: 0.6768
```

Fig-4.3 : Accuracy

Fig-4.4 : Model prediction on the live feed

5. CONCLUSIONS

This task aimed to implement real-time facial emotion recognition. First, it is suggested to hyper tune the architecture of the CNN used for the model to fit perfectly with the problem at hand. Some areas of this hyper tuning include finding and removing redundant parameters, adding new parameters in more useful places in the CNN's structure, adjusting the learning rate decay schedule, adapting the probability and location of dropout, and experimenting to find ideal stride sizes. Vgg-16 has been quite an old model now, VGG was great for the results it attained back in 2014. It has so many weight parameters, the models are very heavy, 550 MB + of weight size. But deeper networks can have higher test error and generalize lesser if done simply like what VGG does. Other latest state of art computer vision models like efficient-net may be used to get better results. This can still be optimized and accuracy can be increased.

6. REFERENCES

[1]. Lxyuan0420. (2019, July 21). Facial Expression Recognition using CNN. Retrieved December 19, 2020, from https://www.kaggle.com/lxyuan0420/facial-expression-recognition-using-cnn

[2] Busso, Carlos & Yildirim, Serdar & Bulut, Murtaza & Lee, Chul & Kazemzadeh, Abe & Lee, Sungbok & Neumann, Ulrich & Narayanan, Shrikanth. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. ICMI'04 - Sixth International Conference on Multimodal Interfaces. 205-211. 10.1145/1027933.1027968.

[3]. Aneja D., Colburn A., Faigin G., Shapiro L., Mones B. (2017) Modeling Stylized Character Expressions via Deep Learning. In: Lai SH., Lepetit V., Nishino K., Sato Y. (eds) Computer Vision – ACCV 2016. ACCV 2016. Lecture Notes in Computer Science, vol 10112. Springer, Cham. https://doi.org/10.1007/978-3-319-54184-6_9

[4]. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2020.2981446.

[5]. Github code archieved at: "https://github.com/Megha-Singh-10/Facial-Emotion-Recognition-System"

[6]. Minaee, Shervin, and Amirali Abdolrashidi. "Deep-emotion: Facial expression recognition using attentional convolutional network." arXiv preprint arXiv:1902.01019 (2019).

[7]. Edwards, Jane, Henry J. Jackson, and Philippa E. Pattison. "Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review." Clinical psychology review 22.6: 789-832, 2002.

[8]. He, Kaiming, 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, pp. 770-778, 2016.

[9]. Kingma, Diederik P. "Adam: A Method for Stochastic Optimization/Diederik P." Kingma, Jimmy Ba, URL: https://arxiv. org/abs/1412.6980.

[10]. ML Cheatsheet retrived from https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html.

[11]. Quinn M., Sivesind G., and Reis G., "Real-time Emotion Recognition From Facial Expressions", 2017