International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)
ISSN: 2305-0543 (Online), ISSN: 2411-6173 (Print)
Vol. 11(41), Jul. 2021, PP. 5043-5045

Nigeria Consultants' Network (CN)

Transfer Learning: An Enhanced Feature Extraction Techniques for Facial Emotion Classification

[1]Christine Bukola Asaju
Computer Science Department, Federal Polytechnic, Idah, Kogi State
Chrisamaju02@gmail.com

[2]Ugwu Kingsley
Computer Science Department, Federal Polytechnic, Idah, Kogi State
StateKinosoft123@gmail.com

[3]Eze Chinemerem Christian
Computer Science Department, Federal Polytechnic, Idah, Kogi State
ezechinemerem@gmail.com

[4]Ugbedeojo Musa
Computer Science Department, Federal Polytechnic, Idah, Kogi State
Musa.ugbedeojo@yahoo.com

***Abstract*— The process of extracting features from images from the scratch in the field of machine learning is a very challenging and time-consuming task. Previous work focused on a technique for feature extraction in which a CNN is pre-trained on some data sets and then in turn used to learn the pattern from other datasets on which it was not originally trained. This paper proposed transfer learning as the ability of a pre-trained CNN model to learn patterns from data which it was not originally trained on. Several pre-trained CNN architectures have been proposed by researchers to improve the system performance. Some of these include the ResNet, Xception, VGG16, VGG19, InceptionResnetV2, InceptionV3, DenseNet, MobileNet, etc., all trained on ImageNet datasets. However, not all of these networks have the ability to effectively perform feature extraction on facial datasets. This paper attempts to compare the efficiencies of five pre-trained networks namely, VGG19, VGG16, ResNet50, inceptionResNetV3, and InceptionV2 which have been pre-trained on the ImageNet dataset as features extractors on the DISFA plus facial emotion dataset for classification. The extracted features are used for training and testing a simple linear regression classifier. Testing the classifier network with features extracted by these networks produced different accuracies. It was discovered in the process that out of the five networks being tested, the VGG19 architecture performed more accurately than other pre-trained networks on the classifier with an accuracy of 97%. The work, therefore, presumes that the VGG19 network has a better generalization ability on facial data than other networks.**

Keywords—Convolutional Neural Network, Machine Learning, Transfer Learning, Pre-trained network

I. INTRODUCTION

The natural ability to transfer knowledge from one task to another is possessed by every intelligent human being. Acquired knowledge in one task is easily transferred or used in solving problems in related areas [1]. Ideas can easily be cross-fertilized between related tasks. For example, when an individual knows how to ride a bike, the individual might easily learn how to drive a car. The individual needs not to learn everything from the scratch when the new challenge comes, the knowledge of the initial one could be transferred to the new task. [1]. This transfer of knowledge is not so for machines. When models are developed for a particular task, another model has to be designed for a similar task from scratch. Further research in machine learning brought about the concept of transfer learning [2]. Defines transfer learning as a machine learning method, where a model that was developed for a task is being used again as the beginning level for a model on a second task.

II. RELATED WORKS

In the work [3], the researchers opined that transfer learning can improve learning through the transfer of knowledge from a similar domain. Authors in [4] and [5] describes transfer learning as the application of information from a known region to an unknown region. Transfer learning can adopt knowledge from a simulation environment to a real-world domain [6].

The researchers in [7] describes transfer learning models as pre-trained models. Deep neural networks trained on large scale datasets such as ImageNet have demonstrated to be excellent in the tasks of transfer learning. Convolutional neural network (CNN) pre-trained on image-Net as the strength of the most state of the art approaches to transfer learning. Several pre-trained models are as described in [7].

Some of such CNN includes Residual Networks (ResNet-50)[8], VGG16, VGG19 [9], InceptionResnetV2 [10], InceptionV3 [11] etc. Transfer learning is of two types namely; Pre-trained networks being treated as feature extractors, secondly, removing the fully connected layers of an existing network and replacing it with a new FC layer, and fine-tuning these weights to recognize new object classes [1]. The research focuses on different pre-trained CNN networks being treated as feature extractors on DISFAplus facial emotion data set followed by a classifier for emotion classification. The major focus of the work is to determine the most efficient pre-trained network that is suitable for use in facial emotion recognition. Therefore, the work trained several pre-trained networks and tested with the same data set. The work further compares the results of the test experiment to determine the most efficient.

A. Facial Emotion Expression

Facial emotion expressions are the spontaneous movements that occurs when one or more of the facial muscles of the face are engaged [12]. The facial emotion

Article History:
Received Date: Feb. 30, 2021, Accepted Date: Jun. 25, 2021, Available Online: Jul. 01, 2021

expression is the most appropriate source of non-verbal communication and contains so much information. Paul Ekman in his work, further discovered that there are seven basic emotions, namely; happiness, sadness, fear, disgust, anger, contempt, and surprise in [13]. To successfully train a model to classify the various emotions, the features need to be extracted. The work considers transfer learning technique for feature extraction of the facial emotion datasets, using different pre-trained networks namely; VGG19, VGG16, RESNET-50, InceptionResNetV2, and InceptionV3 and for further classification using simple linear regression classification. The various pre-trained network that was used for the work is described briefly in the next section.

B. Pre-trained Networks

ResNet-50

Pre-trained ResNet-50 network is a 50- layer deep convolutional network, where the main idea is to skip blocks of convolutional layers by using short cut connections [15]. ResNet-50 network was trained on the ImageNet 2012 classification data set, which consists of 1000 classes. The models are trained on 1.28 million training images and evaluated on 50 thousand validation images. The final result was obtained on 100 thousand test images, with Adam Optimizer. During training, ResNet got an accuracy of 75% after training on 100 epochs. It was evaluated with both top-1 and top-5 error rates. ResNet use zero padding and no extra parameters.

VGG16 &VGG19

The Visual Geometry group (VGG) at Oxford, came up with another CNN model which is a successor of the AlexNet, called VGG. The VGG has several variants, of which VGG19 and VGG 16 is part. VGG19 has 19 layers (16 convolution layers, 3 fully connected layers, 5 Maxpool layers, and 1SoftMax layer). VGG16 on the other hand is a CNN with 16 layers deep. A pre-trained version of VGG16 was trained on more than a million images from the ImageNet database. The VGG16 can classify images into over 1000 categories. [16]

InceptionResnetV2

InceptionResnetV2 is a CNN network that 164 layers deep, and can classify images into 1000 objects. It was pre-trained on more than a million images from the ImageNet database [17].

InceptionV3

InceptionV3 is a CNN network that was also pre-trained using a data set of 1000 classes from the original imageNet data set according to [17]. Each of the network is used as features extractor and used to train a classifier. Their test result is given in the next section.

III. EXPERIMENT

This work, acquired the video data set of facial emotions the DISFAplus dataset by request. The DISFAplus is also known as extended Denver Intensity of spontaneous Facial Action dataset. It provides a manually labelled frames of five levels of intensity of FACS (Facial Action Coding System).

The dataset is pre-processed and was spilt into training, testing and validation, then using python codes, with Keras library and TensorFlow as the back- end, applied each of the pre-trained networks to extract the features of faces from the video sequence. After each extraction by the pre-trained networks, the outputs were used as input to training, validating and testing a linear classifier. Setting the parameter at 50 epochs, model was tested using 6083 samples that have not been exposed to the network. Each of the network gave different classification accuracies, with VGG19 giving 97% which is the highest. The model diagram shown in Figure 1. While confusion matrix for VGG19 test result is given in figure 2. The test results for all the models are shown in Table 1.

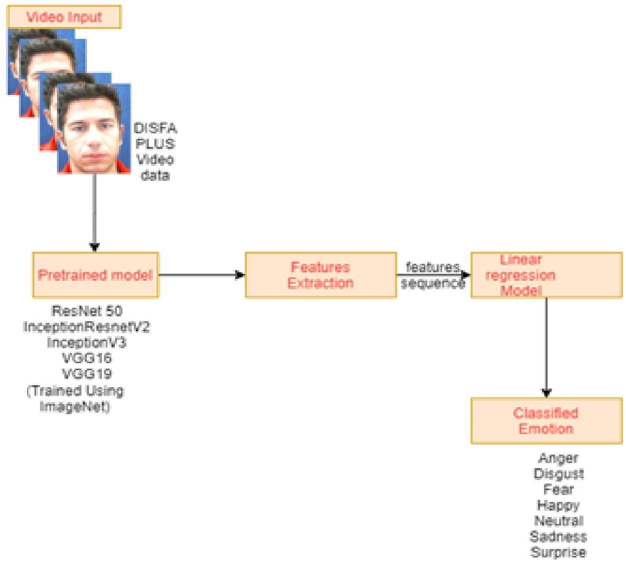

Fig 1. Transfer learning model diagrams

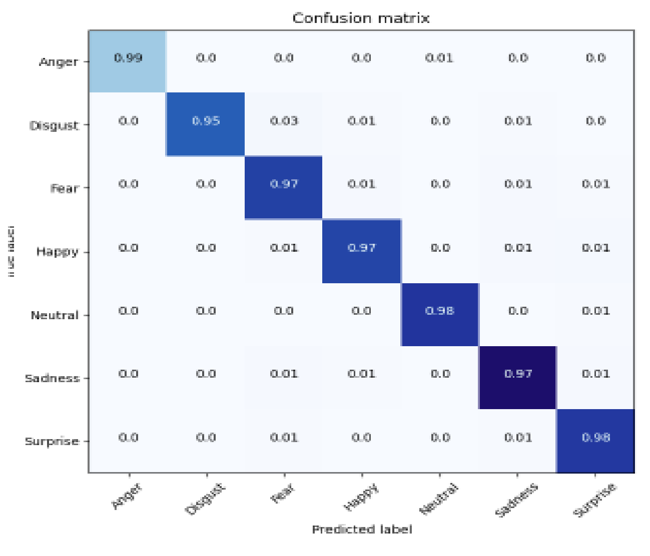

Fig 2: Confusion matrix for VGG19 test result

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

TABLE I. MODEL TEST RESULTS OF A LINEAR CLASSIFIER

CNN pre-trained model	Test accuracy at 50 epochs (%)	Remark
InceptionV3	79	4th
InceptionResnetV2	80	3rd
ResNet50	73	5th
VGG16	94	2nd
VGG19	97	1st

IV. CONCLUSION

This paper discussed an experimental work on the work observed from the experiment that transfer learning is possible, but dependent on pre-trained architectures/data sets. For the task of facial expression recognition, VGG 19 was able to give a higher recognition rate. It was further observed that, not all the pre-trained networks were able to perform efficiently in classification with the data it was not originally trained on, which indicates the ability of VGG19 to have a higher generalization ability than others. The work contributes to knowledge in the following way; It will help researchers to understand and identify the most appropriate pre-trained CNN model to apply to a given feature extraction task. The work is applicable in efficient facial emotion classification task.

REFERENCES

[1] Sarkar D. A Comprehensive Hands-on Guide to Transfer Learning with Real-World Application Deep Learning. Midum publisher, 2018

[2] Brownlee J. A Gentle Introduction to Transfer Learning for deep learning. Machine Learning Mastery. Journal of machine Learning Mastery, 2017 Dec.

[3] Akhauri S, Zheng L, Lin M. Enhanced Transfer Learning for Autonomous Driving with Systematic Accident Simulation.arXiv preprint arXiv:2007.12148. 2020 Jul 23.

[4] Chen J, Liu X. Transfer learning with one-class data. Pattern Recognition Letters. 2014 Feb 1; 37:32-40.

[5] Chen J, Liu X, Tu P, Aragones A. Learning person- specific models for facial expression and action unit recognition. Pattern Recognition Letters. 2013 Nov 1; 34(15):1964-70.

[6] Akhauri S, Zheng L, Lin M. Enhanced Transfer Learning for Autonomous Driving with Systematic Accident Simulation. arXiv preprint arXiv:2007.12148. 2020 Jul 23.

[7] Simon M, Rodner E, Denzler J. ImageNet pre-trained models with batch normalization. arXiv preprint arXiv:1612.01452. 2016 Dec 5.

[8] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.

[10] Wan X, Ren F, Yong D. Using Inception-Resnet V2 for Face-based Age Recognition in Scenic Spots. *In IEEE 6th Int'l Conf on Cloud Computing and Intelligence Systems (CCIS)* 2019 Dec 19(pp. 159-163). IEEE.

[11] Wang C, Chen D, Hao L, Liu X, Zeng Y, Chen J, Zhang G. Pulmonary image classification based on inception-v3 transfer learning model. *IEEE Access.* 2019 Oct 7; 7:146533-41.

[12] Theckedath D, Sedamkar RR. Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. SN Computer Science. 2020 Mar; 1(2):1-7. [13] Ekman P. Basic emotions. Handbook of cognition and emotion. 1999 Feb 25; 98(45-60):16.

[14] Mavadati M, Sanger P, Mahoor MH. Extended disfa dataset: Investigating posed and spontaneous facial Expressions. In proceedings of the IEEE conference on computer vision and pattern recognition workshops 2016 (pp. 1-8).

[15] Rezende E, Ruppert G, Carvalho T, Ramos F, De Geus P. Malicious Software Classification using Transfer Learning of Resnet-50 deep neural network. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) 2017 Dec 18 (pp. 1011-1014). IEEE.

[16] Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. British Machine Vision Association, 2015.

[17] Baust M, Ludwig F, Rupprecht C, Kohl M, Braunewell S. Understanding Regularization to Visualize Convolutional Veural Networks. arXiv preprint arXiv:1805.00071. 2018 Apr 20.

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173