Biosc.Biotech.Res.Comm. Special Issue Vol 13 No 14 (2020) Pp-185-188

BBRC
Bioscience Biotechnology Research Communications

Facial Expression Recognition Using Transfer Learning on Deep Convolutional Network

Ramchand Hablani
[1]*Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India*

ABSTRACT

A novel method is proposed for facial expression recognition. We have implemented two techniques for automatic facial expression recognition. First, we applied transfer learning to AlexNet, and VGG19 for classification. Second, we used AlexNet and Vgg19 for feature extraction and cascaded it with an SVM for classification. We achieved 86.11% accuracy with AlexNet and 94.44% with AlexNet-SVM cascade. We also achieved 94.44% accuracy with VGG19 and 86.11 with VGG19-SVM cascade. We used JAFFE Data Set to train our four models. Our system achieves an improvement in accuracy for JAFFE Data Set.

KEY WORDS: FACIAL EXPRESSION RECOGNITION, ALEXNET, VGG19, TRANSFER LEARNING, FEATURE EXTRACTION.

INTRODUCTION

We humans are very good at recognising various facial expressions. At an early age a child learns to recognise different facial expressions. Automatic facial expression recognition by computer in varying external conditions has not achieved that much success [Ramchand Hablani,2013] [Sarika Jain, 2014]. Deep learning especially convolutional neural network (CNN) has successfully been applied for object recognition in the computer vision domain [S. Li, 2018]. First few layer of CNN extract abstract features of the image, in the later part of CNN this abstract features combined to form a meaningful features. For different classification problems, researchers have proposed different architectures of CNN, some of the architectures are AlexNet, GoogleNet VGG19 and FaceNet [Wang J, 2018].

ARTICLE INFORMATION

*Corresponding Author: hablanir@rknec.edu
Received 15th Oct 2020 Accepted after revision 29th Dec 2020
Print ISSN: 0974-6455 Online ISSN: 2321-4007 CODEN: BBRCBA

Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal

Clarivate Analytics

crossref

NAAS Journal Score 2020 (4.31)

Online Contents Available at: http//www.bbrc.in/
Doi: http://dx.doi.org/10.21786/bbrc/13.14/44

There are basically three approaches for using convolutional neural network for image classification; the first approach is using a trained network (trained on few millions of images) for differencing [S. Li, 2018]. The drawback of this approach is that if we are presenting images of objects which do not belong to any class of trained network, then test accuracy is not up to the mark [Minaee S, 2019]. Suppose network is not trained on various facial expression images, then the test accuracy of this trained network on facial expression images is very low. The second approach for using CNN for image classification is to train the complete network from scratch. There are two problems in this approach. The first one is that few millions of images are required to train network from the scratch. For the task of facial expression recognition, that much large data set is not available. Second problem is that a very powerful GPU system is required to train CNN from scratch.

There is a third approach between these two extreme approaches, called transfer learning. In transfer learning, the fully connected layers are changed according to our classification problem. Weights of convolutional base are freeze, and weights of fully connected layers are learned by presenting the images and corresponding target outputs to the network. There is another approach in transfer learning, flattened values of the convolutional base are

considered as extracted features of network, and then these features are used as input to some robust classifier like SVM. This classifier is trained by using features and corresponding target values. We have implemented both the methods of transfer learning on AlexNet and VGG19. We have used SVM as a classifier to classify the given image into one of the six possible emotions. We train, and fine tune [Minaee S, 2019], the AlexNet using 227x227x3 input images, and train and fine tune VGG19 using 224x224x3 input images. Then, we classify each image as one of six possible emotions.

2. Literature Review

2.1 Convolutional Neural Network: There are four operations in the CNN

1. Convolution operation: Convolutional neural network is based on convolution operation. The convolution operation in one dimension is define as

$$y(n)=x(n)*h(n) \quad (1)$$

$$y(n) = \Sigma_k x(k)h(n-k) \quad (2)$$

Where x(n) is the input, h(n) is called the kernel (filter) and y(n) in is the output. The convolution operation in two dimensions is define as

$$y(m,n) = x(m,n)*h(m,n) \quad (3)$$

$$y(m,n) = \Sigma_j\Sigma_i x(i,j)h(m-i,n-j) \quad (4)$$

The parameters of the kernel h are not fixed but are learnt from the data.

2. Activation function (ReLU): After the convolution operation which is linear, 2nd operation is nonlinear that is an activation function. ReLU is applied as an activation function.

3. Max Pooling or Sub Sampling: Pooling layer reduces the size of outputs of convolution filters. Pooling layer doesn't have any learnable parameters. There are two types of pooling operations, max pooling and average pooling. Max pooling is generally used in convolutional neural network. The convolution, activation functions and pooling is called a convolutional layer. In CNN architecture, there are many convolutional layers. All these layers of convolutional layers is called conversional base of the network.

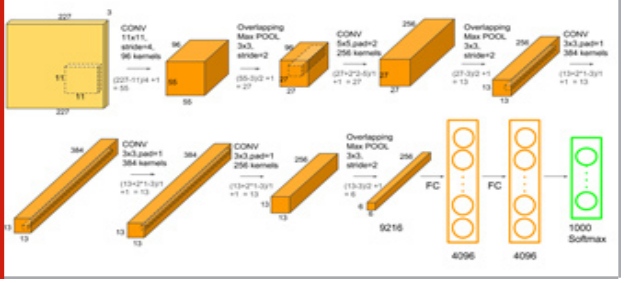
4. Fully Connected Layer: The output of the conventional base is flattened to give input to the fully connected layers of the network. The last layer of the fully connected layers is the output layer. The number of neurones in the output layer is equal to the number of classes in the training data.

2.2 Feature Extraction from pre-trained Convolutional Neural Network: As mentioned above, the output of the convolutional base is converted into one dimensional vector. This one dimensional vector can be used as a feature vector. There are two approaches for using this feature vector [A. Mollahosseini, 2015]. The first approach is to input this feature vector to fully connected layers of convolutional neural network. The only change to be made is to change the size of the output layer. The size of the output layer must be equal to number of classes in our classification problem [Jyh-Yeong Chang, 2001]. In our case there are 6 classes one for each facial expression. Softmax is used as an activation function for the output layer. The second approach is to use this feature vector as an input to some robust classifier like SVM, and train that classifier with this feature vectors as a input and the corresponding class as a target output. We have used SVM as a classifier for facial expression recognition [R. R. Selvaraju, 2017].

2.2.1 AlexNet Architecture: The size of the input layer of AlexNet is 227x227x3.The RGB images of size 227x 227 are presented as input to the network. There are a total 8 layers in a network; out of which 5 are convolutional layers and remaining 3 are fully connected layers. On the input images, 96 filters of size 11x11 with stride equal to 4 are applied and get a tensor of size 55x55x96. Max pooling of size 3x3 is applied on this tensor and gets a tensor of size 27x27 x 96. After that 256 filters of size 5x5 with same convolution are applied and get a tensor of size 27x27x256.Then max pooling is applied on this tensor and gets a tensor of size 13x13x256. After that 384, 384 and 256 filters of size 3x3 with same convolution are applied one after another without applying any pooling operation. So we get a tensor of size 13x13x256.After that max-pooling is applied on this tensor and get a tensor of size 6x6x256. This tensor is now converted into one dimensional vector 0f size 9216. Finally 3 layers of size 4096, 4096 and 1000 neutrons are added as fully connected layers.

2.2.2 VGG19 Architecture: The size of the input layer of VGG19 is 224x224x3.The RGB images of size 224x 224 are presented as input to the network. There are a total 19 layers in a network; out of which 12 are convolutional layers and 5 are max pooling layers and remaining 2 are fully connected layers.

Figure 1: AlexNet Architecture



3. Facial Expression Database: The Japanese female facial expression (JAFFE) is the data set for facial expression recognition [Sunny Bagga, 2013]. There are a total 213 images in which 10 female subjects are expressing their

faces in 6 universal emotions like anger, disgust, fear, happiness, sadness and surprise. The seventh expression is neutral [Ankita Vyas, 2014]. These are 8 bit gray scale images with resolution of 256 X 256.

Figure 2: VGG19 Architecture

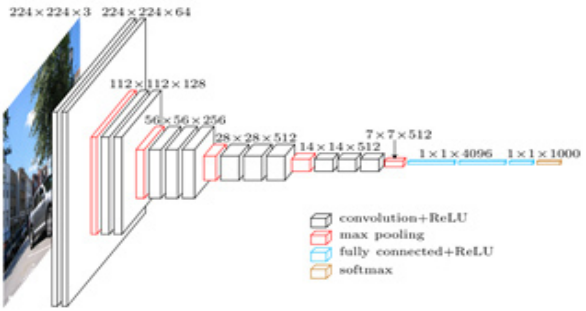


Figure 3: JAFFE dataset

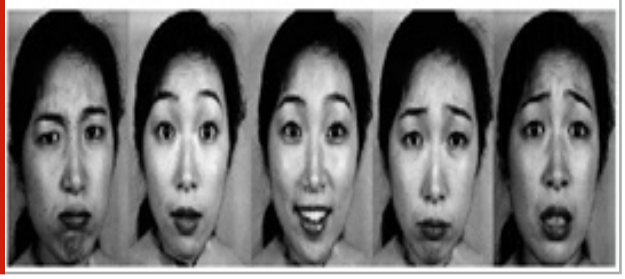


Figure 4: Transfer Learning

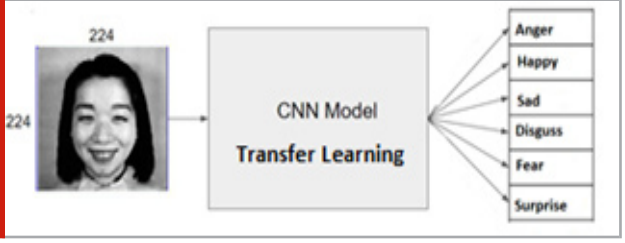
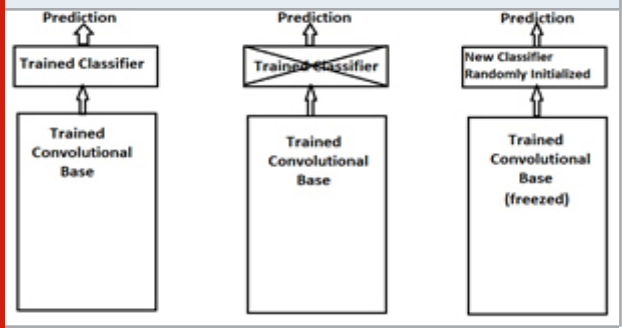


Figure 5: Feature extraction with new classifier



4. Proposed Method: In the proposed system, transfer learning and feature extractions are used to classify one out of 6 possible classes. Two approaches transfer learning and feature extraction with SVM classifier, have been applied onto pre-trained networks AlexNet and VGG19. In the first approach, the size of output layer in fully connected layers is changed to 6, one for each facial expression. The convolutional base is freeze, and weights of fully connected layers are trained by using JAFFE dataset. In the second approach, the output of conventional base is converted into one dimensional feature vector, this feature vector and target values are used to train SVM classifier.

5. Experiments and Results: The last three layers of pre-trained network of AlexNet and VGG19 are configured for 1000 classes however; we performed fine-tuning of these classes since we wanted to detect 6 classes for facial expressions. For our classification, output layer of fully connected network was changed to 6 and we also retrained the subsequent two softmax and classification output layers. In our second approach, features were extracted from the fully connected layer FC7 of AlexNet and VGG19. Based on this, the feature vector with training labels were formed and fed them to the SVM Classifier. Since AlexNet only accepts the image with input size of 227x227x3,(VGG19 accept images with input size of 224x224x3), so we replicated the channel three times to converted our single channel gray scale images of JAFFE to three channels. Also, the input image size was different so, we resized the image.

Table 1. Confusion Matrix for Transfer Learning with AlexNet

Expressions	Fear	Surprise	Sad	Angry	Disgust	Happy
Fear	6	0	0	0	1	0
Surprise	0	6	1	0	1	0
Sad	0	0	5	0	0	0
Angry	0	0	0	6	2	0
Disgust	0	0	0	0	2	0
Happy	0	0	0	0	0	6

Table 2. Confusion Matrix for AlexNet-SVM cascade

Expressions	Fear	Surprise	Sad	Angry	Disgust	Happy
Fear	6	1	0	0	0	0
Surprise	0	5	0	0	0	0
Sad	0	0	5	0	0	0
Angry	0	0	0	6	0	0
Disgust	0	0	0	0	6	0
Happy	0	0	1	0	0	6

Random splitting with 80-20 ratios was done for the train and test images, which lead to different accuracy for each split. We achieved 86.11% and 94.44% accuracy for transfer learning and feature extraction-SVM approach respectively for AlexNet. The mini Batch size of 2 and 20 epoch gives optimal results for transfer learning. We achieved 94.44% and 86.11% accuracy for

transfer learning and feature extraction-SVM approach respectively for VGG19.

Table 3. Confusion Matrix for Transfer Learning with VGG19

Expressions	Fear	Surprise	Sad	Angry	Disgust	Happy
Fear	5	1	0	0	0	0
Surprise	0	5	0	0	0	0
Sad	1	0	6	0	0	0
Angry	0	0	0	0	6	0
Disgust	0	0	0	0	6	0
Happy	0	0	0	0	0	6

CONCLUSION

In proposed system, we have implemented transfer learning and classification using SVM on extracted features from convolution base of pre-trained models. We have used AlexNet and VGG19 as pre-trained CNN models. We have applied proposed methods on JAFFE dataset. Proposed system has achieved the accuracy of 94.44% which is comparable to best existing system.

The main conclusions of our proposed work are:

1. For small dataset transfer learning Deep CNN is a better option.
2. SVM is a powerful classifier, it can be trained using features extracted from pre-trained Deep CNN.

We will extend our work for Cohen Kanade dataset for facial Expressions, and real time images.

Table 4. Confusion Matrix for VGG19-SVM cascade

Expressions	Fear	Surprise	Sad	Angry	Disgust	Happy
Fear	5	1	0	0	0	0
Surprise	0	3	0	0	0	0
Sad	0	0	5	0	0	0
Angry	0	0	0	6	0	0
Disgust	1	2	0	0	6	0
Happy	0	0	1	0	0	6

REFERENCES

A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going Deeper in Facial Expression Recognition using Deep Neural Networks," CoRR, vol. 1511, 2015

Ankita Vyas, Ramchand Hablani. "Effect of Different Occlusion on Facial Expressions Recognition", Int. Journal of Engineering Research and Applications, Vol.4- Issue 10,October 2014. (ISSN: 2248-9622) pp.40-44

Jyh-Yeong Chang* and Jia-Lin Chen "Automated Facial Expression Recognition System Using Neural Networks" Journal of the Chinese Institute of Engineers, Vol. 24, No. 3, pp. 345-356 (2001)

Minaee S., Abdolrashidi A., "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network", 2019

Ramchand Hablani, Narendra Chaudhari and Sanjay Tanwani, "Recognition of Facial Expressions using Local Binary Patterns of Important Facial Parts", International Journal of Image Processing(IJIP) IJIP-738. ISSN(online) 1985-2304 , vol.7, Issue 2,2013.

R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D.Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in 2017 IEEE International Conference on Computer Vision (ICCV), Oct 2017.

Sarika Jain, Sunny Bagga, Ramchand Hablani, Narendra Chaudhari and Sanjay Tanwani, "Significance of Facial Features in Performance of Automatic Facial Expression Recognition", The CSI Conference on Big Data 2014 (CSI-BIG -2014), Indore, India.

S. Li and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348, 2018

Sunny Bagga, Sarika Jain, Ramchand Hablani, Narendra Chaudhari, and Sanjay Tanwani, "Automatic Facial Expression Recognition Using LBP of Essential Facial Parts and Feed Forward Neural Network", International Conference on Emerging Trends and Applications in Computer Science (IC ETACS– 2013), Shilong , India.

Wang J., and Mbuthia M., "FaceNet: Facial Expression Recognition Based on Deep Convolutional Neural Network", 2018