Development of Facial Expression Classifier using Neural Networks

Vandana Jha
Dept. of CSE
KSSEM, VTU
Bangalore, India
vjvandanajha@gmail.com

P. Deepa Shenoy
Dept. of CSE
UVCE, Bangalore University
Bangalore, India
shenoypd1@gmail.com

Venugopal K R
Bangalore University
Bangalore, India
venugopalkr@gmail.com

Abstract—A person's emotional and mental well being, together with the age, sex, race, can be easily depicted by one's face. A crucial role is played by facial expressions in day-to-day social interactions. An individual's emotional level as well as behavioral manners can be interpreted by these expressions. Facial expression classifier is a evolving, demanding and curious problem in computer vision. It has its potential applications in the field of robotics, behavioral science, human computer interaction, video games etc.. It assists in building more intelligent systems which have better ability to interpret human emotions. In this paper, a facial expression classifier is proposed based on Convolution Neural Networks (CNN). CNNs are biologically-inspired variants of multi-layer preceptor (MLP) networks. They use an architecture which is particularly well suitable to classify images. Detection of facial expression can be enhanced by utilizing the ability of learning facial features. The results showed that deep CNNs are apt for this task. We have used CNN to interpret and classify the facial expression label into seven different classes: anger, happiness, fear, sadness, disgust, surprise and neutral.

Index Terms—Convolutional Neural Networks, Facial Expression Detection, OpenCV, Python, TensorFlow.

I. INTRODUCTION

"The face is the mirror of the mind, and eyes without speaking confess the secrets of the heart."

— St. Jerome, Latin Catholic priest

A recent study on nonverbal communication discloses that facial expressions are the means to convey 55% of a person's emotional or intentional information so facial expressions are essential players for human to show emotions in communications. In recent times, researches on emotional analysis have increased significantly. This became possible due to the development of neuroscience and cognitive science, easing the path of emotional analysis. Also, the technical advancement in the areas of computer vision and machine learning made a contribution to the growth of emotional analysis and made it reach to the public. Researches on identifying facial expressions are developing quickly as it is an important subfield of emotional analysis. Identification of facial expressions has many applications in the areas such as human-computer interaction system, multimedia, surveillance and driver safety. Work of Ekman et al. [1] is one of the initial works in this direction of research.

A. Motivation

The motivation behind our work of facial expression classifier is to build an automatic system for reading and understanding facial expressions, which further helps in knowing the mood and the mental status of the person. Generally, facial expressions can be classified into seven basic classes, which encompass anger, happiness, fear, sadness, disgust, surprise and neutral. Therefore, classifying the expressions into these basic classes, is the primary task of our proposed system. Automatic identification and classification of facial expressions is an essential component of natural human-machine interfaces. It has its applications in medical domain as well as in behavioral science, specially, in case of treating coma patients or child psychology. Even though, facial expressions identification is fast and without any effort by human mind, it is still a challenging task to be performed by learned machine. The contribution made by our work, is increased accuracy with less processing time, and this is achieved with the use of Deep Convolutional Neural Network (DCNN) features.

B. Organization

The paper continues with the following organizational structure: A brief analysis of the related work is presented in section II. Section III describes proposed system with tensorflow platform and process involved in recognising the facial expression. Simulation runs on kaggle dataset and all the related results are discussed in section IV. Section V concludes the paper.

II. RELATED WORK

Convolutional Neural Network algorithms are used for facial recognition from quite a time now and based on the approaches or modifications in the existing algorithm, the related work can be divided into two parts:

- Convolutional Neural Network
- Hybrid Convolutional Neural Network.

Convolutional Neural Network - Meng et al. [2] developed an Identity-Aware Convolution Neural Network (IACNN) which made use of both expression-related and identity-related information for recognition of facial expression. They introduced a new auxiliary layer with an identity sensitive

1

contrastive loss, which helps in learning identity-related representations to lessen high inter-subject variations. A joint loss function have been proposed, which examines both expression and identity similarities as well as classification errors for recognising expressions. This loss function is used to adjust both the features simultaneously (expression-related and identity-related features). After concatenating, both the features, it is used to recognise identity-invariant facial expression. Experiments have been conducted on two different facial expression poses datasets. The results have shown that the proposed IACNN model gave better results than the baseline CNN methods as well as most of the existing methods that make use of the dynamic information extracted from image sequences.

Zhang et al. [3] proposed a model in which, given the training facial images and their expression labels, a number of face part patches were generated which contain both the local and global personal identity information. The generated image patches were fed into CNN for deep model training. The same network was shared by all patches for computational convenience. The second to the last layer outputs were concatenated as features and a strong classifier, support vector machine (SVM), was trained upon these features for predicting the expression classes. One of the main limitations of this approach, is the conditions of illumination, the pose of human faces, the difference of races, have significant influences on the facial expression identification. Another limitation in the facial expression identification is how to extract key facial features that can better differentiate different emotions but tuning the best parameters for CNN is very time-consuming.

Rashid et al. [4] presented a technique for identifying facial expressions using different datasets of expressions, which were imbalanced in nature. The data preprocessing was performed and later it was made balanced. Significant features of face was extracted using a technique and these extracted features were given as inputs to a classifier model. It has been determined that the best recognition accuracy is produced by Convolutional Neural Network. A person conveys strong and communicative feelings via various manners, for example; attitude, physique, feeling, movement, natural signs etc. Convolutional Neural Network has been used by this research to implement an approach that can recognize facial expressions in static images. The facial behavior can be classified into wide-ranging sets; the suggested technique were evaluated by using these datasets.

A method has been proposed by Mayya et al. [5] for automatically identifying facial expressions using Deep Convolutional Neural Network (DCNN) features. The prime focus of this model was on facial expressions identification of a person from a single image. The feature extraction time can be significantly reduced by the usage of general purpose graphic processing unit (GPGPU). Experiments have been performed on two facial expression datasets which are publicly available. From the results of these experiments, it has been observed that, the existing recognition rate can be achieved using DCNN features, in classifying different facial expressions into different classes.

Hybrid Convolutional Neural Network - A hybrid Convolutional Neural Network for facial expression recognition has been proposed by [6] with Dense Scale Invariant Feature Transform aggregator. By using Scale Invariant Feature Transform (SIFT) features, the performance on small data can be increased because of the advantage that SIFT does not require extensive training data to generate useful features. Both SIFT and Dense SIFT are studied in detail and observed that Dense SIFT runs faster than regular SIFT. All these advantages and disadvantages were compared and decided to be merged with CNN features. The accuracy in facial expression recognition obtained by the combination of CNN with Dense SIFT is better than the individual model, CNN only and the combination, CNN with SIFT. The increase in accuracy obtained by Dense SIFT model was about 1%. The experiments have been performed on both FERC-2013 [7] and CK+ datasets [8]. The obtained results showed that the combined model using Dense SIFT and CNN have achieved outstanding results.

Dachapally and Raj [9] proposed two independent methods for facial emotion detection. In the first method, autoencoders were used to construct a unique representation of each emotion, while in the second method, an 8-layer convolutional neural network (CNN) was used. The evaluations of this model showed that the CNN model can give better accuracy and performance than the existing methods for emotion recognition, with more depth and fine-tuning of the filters.

Facial expression recognition also helps in understanding the sentiments of the masses in case of any product advertisement or election campaigning and can be helpful in the works carried out in this direction [10], [11].

III. PROPOSED SYSTEM

The proposed system has a fixed input image size as 42x42x1. It comprises of convolution layers, pooling layers and fully connected layers. The ReLu activation function [12] is used, after each convolutional layer and fully connected layer. The output layer contains 7 neurons to display 7 facial expressions: anger, happiness, fear, sadness, disgust, surprise and neutral. Applying a softmax layer is optional in case of multi-class SVM loss function. In the training phase, we can put a softmax layer after the output layer to make use of the cross-entropy loss function.

A. Model Specifications

The model is inspired by the simple design architecure of VGGNet [13] and contains four blocks. The model specification is shown in table I and can be explained as follows: The first block consists of two 3x3 convolutional layers and the number of filters are 32. The second block consists of two 3x3 convolutional layers and the number of filters are 64. The third block consists of two 3x3 convolutional layers and the number of filters are 64. The fourth block consists of three 3x3 convolutional layers and the number of filters are 64.

TABLE I: Model Specifications

Input image size = 42x42x1
1st Block with two 3x3 conv. layers, filters = 32, Stride = 1
max pooling layer of filter size 2x2, Stride = 2
2nd Block with two 3x3 conv. layers, filters = 64, Stride = 1
max pooling layer of filter size 2x2, Stride = 2
3rd Block with two 3x3 conv. layers, filters = 64, Stride = 1
max pooling layer of filter size 2x2, Stride = 2
4th Block with three 3x3 conv. layers, filters = 64, Stride = 1
fully connected layer with 256 neurons
fully connected layer with 256 neurons
Output layer = fully connected layer with 7 neurons for 7 expressions

For all the four blocks, stride is 1. We have employed a max pooling layer with filter size 2x2 and the stride is 2, after each block except the last one. Two fully connected layers with 256 neurons per layer, after the four blocks have been employed. At the end, the output layer is a fully connected layer with 7 neurons representing 7 facial expressions. As discussed earlier, applying a softmax layer is optional. We have added it and applied the cross-entropy loss function in the training phase, after the output layer. This softmax layer can be skipped, if during the training phase, one uses the multi-class SVM loss.

B. Implementation

Facial expression recognition process involves following steps:

- Grayscale Conversion - The input is taken through the webcam as live input. As the system can't understand the color images, they have to be converted to grayscale. We then obtain the frames of our camera one by one and convert them to gray. That is later converted to binary values for evaluation.
- Feature Extraction - Procedure of eye extraction:
  - (a) Basic rectangular blocks to enclose the eyes.
  - (b) Conclusive landmark points for eye contours.
  - (c) Eye contours are extracted.

  Procedure of mouth extraction:
  - (a) Basic rectangular block to enclose the mouth.
  - (b) Conclusive landmark points for mouth contours.
  - (c) Mouth contours are extracted.

We have employed TensorFlow architecture for the facial expression detection task. This is termed as TensorFlow because tensors, in the form of a multi-dimensional array, is employed as input to the system. We can construct a stateful flowchart of operations (called a Graph) that we want to execute on that input. The input goes in at one end, and then it flows through this system of multiple operations and comes out the other end as output. It works in three parts:

- **Preprocessing the data** - In this step, data normalization is performed at two levels: image level and pixel level. This step is performed on training data as well as on the test data.

- **Training** - The complete dataset is divided into training and testing dataset. In the training phase, two different loss functions can be applied: cross-entropy and multi-class SVM. Minimizing the loss function is our task. For which, we have used gradient descent with momentum and the back-propagation algorithm [14]. The training dataset is used to train the model and its accuracy in detecting facial expression is computed. The training is stopped at the highest accuracy and the model is ready for the testing phase.
- **Testing** - In this step, we apply the trained model on test data and find out its accuracy in identifying the facial expression. The size of the given dataset image is 48x48 but for CNN, the input image size is 42x42, so image cropping is a required step. We have employed 10 crops of image size 42x42 where 5 are original crops and another 5 are the mirror image of initial 5 crops. For each cropped image, models trained on both the methods, cross-entropy and multi-class SVM, are applied and its score for the class labels are computed. The class label with the highest score, gives the final class label and associated facial expression category for the image.

IV. RESULTS AND EVALUATIONS

A. Dataset

The experiment has been performed on FERC-2013 [7] dataset. The dataset consists of grayscale, centered, face image of 48x48 pixel. The size of the training dataset is 28,709 images. The size of public test dataset is 3,589 images and the private test dataset consists of another 3,589 images, making total dataset size of 35,887 images. Pierre-Luc Carrier and Aaron Courville have prepared this dataset. Each image is associated with a numeric code (0 = Angry, 1 = Disgust, 2 = Fear, 3 = Happy, 4 = Sad, 5 = Surprise, 6 = Neutral) in training phase. The task is to predict this numeric code (and associated emotion) in the test phase. A sample of FERC-2013 dataset is given in Fig. 1.

B. Evaluation Results

We have constructed the CNN model with the model specifications given in section III. This model is applied for two different loss functions, cross-entropy loss function and multi-class SVM loss function. We have used a softmax layer after the output with cross-entropy loss function in the training phase whereas this is skipped in case of SVM loss function. The ReLu activation function [12] is used after each convolutional layer and fully connected layer. The trained model is evaluated for the public test dataset and private test dataset and the results obtained are shown in table II. The results shows that CNN Model with multi-class SVM loss function performs slightly better than cross-entropy loss function. The other observation is, identifying happy and sad classes of facial expression gives better accuracy than other classes of facial expression as the similar facial features are responsible for the other expressions.

3

Fig. 1: Sample FERC-2013 Dataset

TABLE II: Accuracy of the Model

Architecture	Accuracy on public test data	Accuracy on private test data
CNN Model with cross-entropy loss function	69.4	70.2
CNN Model with multi-class SVM loss function	69.9	70.8

V. CONCLUSIONS

Here, we have proposed an automatic model for facial expression identification. The system is trained to classify human facial expressions into seven classes i.e., anger, happiness, fear, sadness, disgust, surprise and neutral. The system ensures favourable results for facial expression identification with fast processing time and simpler design architecture as compared to the winning team of FERC-2013 competition. Our model is inspired by the simple design architecure of VGGNet [13]. The model is fully automatic and useful to work with video feeds as well as images. It provides approximately 70% accurate results for facial expression identification. In the future, face expressions from videos can be recognized more accurately by incorporating temporal information. Future work can be extended in combining facial expression classifier with the area of sentiment analysis, mainly for product advertising field. It will provide a correct means of getting rid of problem of writting fake online reviews.

REFERENCES

[1] R. Ekman, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[2] Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong, "Identity-aware convolutional neural network for facial expression recognition," in *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 2017, pp. 558–565.

[3] C. Zhang, P. Wang, K. Chen, and J.-K. Kämäräinen, "Identity-aware convolutional neural networks for facial expression recognition," *Journal of Systems Engineering and Electronics*, vol. 28, no. 4, pp. 784–792, 2017.

[4] T. A. Rashid, "Convolutional neural networks based method for improving facial expression recognition," in *The International Symposium on Intelligent Systems Technologies and Applications*. Springer, 2016, pp. 73–84.

[5] V. Mayya, R. M. Pai, and M. M. Pai, "Automatic facial expression recognition using dcnn," *Procedia Computer Science*, vol. 93, pp. 453–461, 2016.

[6] T. Connie, M. Al-Shabi, W. P. Cheah, and M. Goh, "Facial expression recognition using a hybrid cnn–sift aggregator," in *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*. Springer, 2017, pp. 139–149.

[7] Y. Tang, "Challenges in representation learning: Facial expression recognition challenge implementation https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge," *University of Toronto*, 2013.

[8] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," 2010, pp. 94 – 101.

[9] P. R. Dachapally, "Facial emotion detection using convolutional neural networks and representational autoencoder units," *arXiv preprint arXiv:1706.01509*, 2017.

[10] V. Jha, R. Savitha, P. D. Shenoy, and K. Venugopal, "Reputation System: Evaluating Reputation Among All Good Sellers," in *Proceedings of NAACL-HLT*, 2016, pp. 115–121.

[11] V. Jha, R. Savitha, P. D. Shenoy, K. Venugopal, and A. K. Sangaiah, "A Novel Sentiment Aware Dictionary for Multi-Domain Sentiment Classification," *Computers & Electrical Engineering*, vol. 69, pp. 585–597, 2017.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[14] L. Ma and K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 3, pp. 1588–1595, 2004.

4