Facial Emotion Recognition Using Deep Cnn Based Features

Jyostna Devi Bodapati, N. Veeranjaneyulu

Abstract*: The objective of emotion recognition is identifying emotions of a human. The emotion can be captured either from face or from verbal communication. In this work we focus on identifying human emotion from facial expressions. Facial emotion recognition is one of the useful task and can be used as a base for many real-time applications. It can be used as a part of many interesting and useful applications like Monitoring security, treating patients in medical field, marketing research, E-learning etc;. We humans can easily identify the emotion of other humans without any effort. Automatic detection of emotion of a human face is important due to its use in real-time applications. The recent advance in GPU has taken many applications like face recognition, hand written digit recognition and object recognition to the next level. Especially the pretrained CNN based features better represent the images. Pretrained CNN features represent the most discriminative features and hence allows for better performance. Feature representation plays a major role on the performance of any machine learning algorithm. After observing unbelievable performance with deep learning models, we propose to use the deep convolutional features to better represent the given image instead of using the traditional handcrafted features. The downside of the deep learning models is that they require large datasets to obtain better performance. To leverage the use of deep learning models without the requirement of large datasets is to use pre-trained models. For feature extraction pre-trained Convolutional Neural Networks model (VGG16) is used and the concept of Deep Neural Networks model is used for classification. To show the performance of the proposed model, Extended Cohn-Kanade (CK+) benchmark dataset is used for the experimental studies. Based on the experimental results we claim that these unsupervised features better represent the images compared the handcrafted features.

Index Terms: Facial Emotion Recognition, CNN, RBF kernels, Extended Cohn-Kanade, Multi-class SVM.

I. INTRODUCTION

The objective of face emotion recognition (FER) is identifying emotions of a human. The emotion can be captured either from face or from verbal communication. Psychological characteristics such as heartbeat and blood Pressure, speech, hand gestures, body movements, Facial expressions identify emotions of a person.

Facial emotion recognition is one of the useful tasks and can be used as a base for many real-time applications. It can be used as a part of many interesting and useful applications like Monitoring security, treating patients in medical field, marketing research, E-learning etc;. We humans can easily identify the emotion of other humans without any effort. Automatic detection of emotion of a human face is important due to its use in real-time applications. The recent advance in GPU has taken many applications like face recognition, hand written digit recognition and object recognition to the next level. Especially the pretrained CNN based features better represent the images. Pretrained CNN features represent the most discriminative features and hence allows for better performance. Feature representation plays a major role on the performance of any machine learning algorithm. In the last few years there has been tremendous progress in the area of facial image classification. In facial image classification, the objective is to identify the person to whom the given face belongs to. A major break through in this face classification task is possible with the introduction of deep neural network (DNN) based models. The reason for the success of deep learning models is the effective feature representation [15]. Representation of the images is a key factor for the success of face classification, object classification, and many other computer vision tasks. The neuron output of fully-connected layers of pre-trained CNN models are excellent representation of images and can better represent the most discriminative features that are useful for tasks like [18] image classification, scene classification etc;. The neuron outputs of convolution layers of pre-trained CNN models can better represent the regional features of the image and are useful for detecting objects in images and to draw bounding boxes over the object regions [19]. In this paper a novel simple FER model is proposed that makes use of deep convolution features. In this work we try to stress the importance of deep convolution features and how they can help to improve the performance of the FER. Especially how the deep neural features can affect the task is studied in detail. Features from convolution layer of the pretrained model are obtained and then fed as input to the classifier. We propose to use the deep convolutional features to better represent the given image instead of using the traditional hand crafted features. The downside of the deep learning models is that they require large datasets to obtain better performance. To leverage the use of deep learning models without the requirement of large datasets is to use pre-trained models. For feature extraction pre-trained Convolutional Neural Networks model (VGG16) is used and the concept of Deep Neural Networks model is used for classification.

Revised Manuscript Received on May 10 ,2019

Jyostna Devi Bodapati, Department of CSE, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, AP, India.

N. Veeranjaneyulu, Department of IT, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, AP, India.

Multi-class SVM based classifier is used for classification as it is proved as the robust classifier [4]. Based on the experimental results we observed that the proposed method is simple and gives better performance compared to the other methods. To show the performance of the proposed model, Extended Cohn-Kanade (CK+) benchmark dataset is used for the experimental studies. Based on the experimental results we claim that these unsupervised features better represent the images compared to the hand crafted features.

II. BACKGROUND WORK

The objective of FER is to identify the emotion of a human face. That is given a face of a human the system has to automatically identify the type of emotion of the face as happy, anger, disgust, fear, happiness, sadness, and surprise. Several studies have been conducted in this area as there are large number of applications which directly or indirectly uses FER. FER can be used as a part of many interesting and useful applications like Monitoring security, treating patients in medical field, marketing research, E-learning etc;.

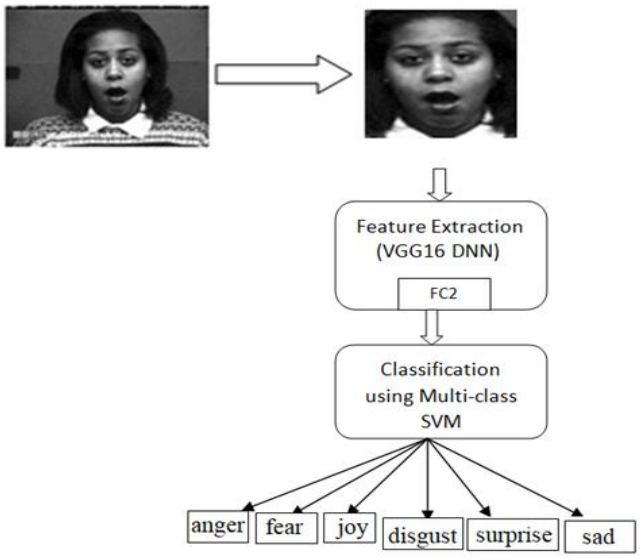
A fuzzy based emotion recognition method based on eyes and mouth features has been proposed [1]. In [5] a multimodal face recognition algorithm has been proposed. In [6] the effect of different kernels on the performance of FER has been investigated. For the study different kernels like linear, polynomial, sigmoid and RBF kernels are considered. In [7] the proposed method is used to automatically find the type of emotion from the video clips. The video clips are taken from the real world and different type of classifiers like SVM, logistic regression, and partial least squares methods are used. A method is proposed in [8] to identify the facial expressions of unknown people. In [9] in addition to the type of emotion recognition, whether the emotion is positive or negative is also identified. A method which is feasible for real time applications is proposed in [10] that uses line-based caricatures. In [11] a decision tree based facial expression classification is proposed which is computationally inexpensive. Another novel method, which uses Gabor filter based features and LVQ based classification for FER [12].

In [2], a novel face region detection method is proposed that uses YCbCr color model in combination with Maximum Morphological Gradient Combination (MMGC). Once the facial region is detected, the searching of emotion is limited to that region alone. In [3] the famous face region detection algorithm, Viola-Jones, is used for facial region detection and then KNN is used for classification of the type of emotion from the facial region.

III. PROPSED WORK

Initially the face region is extracted from the given face images using viola jones face detection algorithm. Then from the cropped face regions, deep features are extracted from the last fully connected layer of VGG16. Multi-class SVM based classifier is used for classification. Based on the experimental results we observed that the proposed method is simple and gives better performance compared to the other methods.

VGG-16, a pretrained convolution model, [15] is known for its state of the art performance in various applications [15] like image classification, object detection etc;. It is trained on ImageNet dataset. VGG16 consists of a total of 13 layers with 5 convolution layers interleaved with pooling layers. Features are extracted from the FC2 layer of VGG16, which gives 2048 dimensional feature vector for each image. Static feature representation is used where all the images are represented with equal size.



IV. EXPERIMENTAL RESULTS

The objective of this work is to identify the emotion of the human given an image of that human. Due to its applications in the real time world, this face emotion recognition has become so popular in the recent past. In this paper we have proposed an algorithm for facial emotion recognition that makes use of deep neural based features. To prove the efficiency of our proposed method we use CK+ dataset, the bench mark dataset for facial expression detection, for experimental studies.

Summary of the Dataset: CK+ [14]: To investigate the performance of the proposed method, benchmark dataset is used for experimental studies. The Extended CK+ dataset is used for our experimental studies as it is the standard dataset used extensively for the FER experiments. This dataset contains a total of 593 video clips of 123 different subjects. The video sequences are varying in nature and varies in expression from neutral face to the final expression. A total of 8 expressions are being considered for the experiments including contempt.

Result analysis: To understand the performance of the proposed model, CK+ dataset is used for our experimental studies. All the images of the dataset are of gray scale type. The input image is given as input to the viola jones algorithm to get the face region alone. Viola jones returns the face region of size 30x30. Then these cropped images are fed as input to the pre-trained VGG16 model which gives a feature vector of size 2048.

These feature vectors are passed to the multi-class SVM to get the type of emotion of the face. Accuracy is used as the performance measure to understand the efficiency of the proposed method.

Dataset	Pretrained model	Face detection	Accuracy
CK+	Resnet50	Yes	86.04%
CK+	Resnet50	No	81.36%

Table1: Comparison of the proposed model with and without face detection stage.

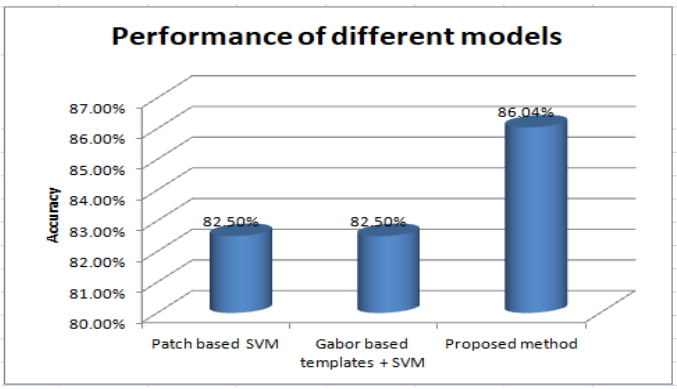
Table1 shows the performance of the proposed model with the face detection algorithm and without using the face detection algorithm. When face detection algorithm is used the accuracy is 86.04% and without using the face detection algorithm the proposed model gives an accuracy of 81.36%. Based on this we can understand that the face detection stage helps us to improve the performance of the algorithm. The reason behind this is by detecting the facial region, it crops out the unnecessary parts of the image and considers only the required parts of the image.

In this subsection the proposed method is compared with the existing methods. To understand the performance of the proposed deep feature based algorithm we compare the model performance with the existing methods.

Model	Accuracy
Patch based SVM [16]	82.5%
Gabor based templates + SVM [17]	82.5%
Proposed method	**86.04%**

Table2. Comparison of the proposed method with the hand crafted feature based models.

The above table shows that the proposed method is better compared to the existing methods. Based on the results we observe that the deep features can better represent the discriminative features of the data when compared to the other methods.



Graph1: comparison of the proposed with existing models

Graph1 shows that the proposed method is better compared to the existing methods. Based on the results we claim that our proposed method that uses deep features is far better than the other methods.

V. CONCLUSION AND FUTURE WORK

The proposed FER method gives better performance than the existing methods. The deep neural based features can represent the emotion of a face better than the handcrafted features. Especially the performance is far better when used with the face detection algorithm as it can crop the facial part alone. The deep features can represent the face features better than the handcrafted features. In future we can extend this and apply the deep neural network based classification to further improve the performance of the algorithm

REFERENCES

1. Farahani, Fatemeh Shahrabi, Mansour Sheikhan, and Ali Farrokhi. "A fuzzy approach for facial emotion recognition." 2013 13th Iranian Conference on Fuzzy Systems (IFSC). IEEE, 2013
2. Oh, Byung-Hun, and Kwang-Seok Hong. "A study on facial components detection method for face-based emotion recognition." 2014 International Conference on Audio, Language and Image Processing. IEEE, 2014
3. Reney, Dolly, and Neeta Tripathi. "An Efficient Method to Face and Emotion Detection." 2015 Fifth International Conference on Communication Systems and Network Technologies. IEEE, 2015.
4. N. Cristiana, T. Shawe, An Introduction to Support Vector Machine, Cambridge University Press, 2000.
5. Pantic, Maja, and Leon JM Rothkrantz. "Toward an affect-sensitive multimodal human-computer interaction" Proceedings of the IEEE 91.9 (2003): 1370-1390.
6. Adeyanju, Ibrahim A., Elijah O. Omidiora, and Omobolaji F. Oyedokun. "Performance evaluation of different support vector machine kernels for face emotion recognition." 2015 SAI Intelligent Systems Conference (IntelliSys). IEEE, 2015.
7. Liu, Mengyi, et al. "Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild." Proceedings of the 16th International Conference on multimodal interaction. ACM, 2014.
8. Salunke, Vibha V., and C. G. Patil. "A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks." 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). IEEE, 2017.
9. Lee, Hyeon-Jung, and Kwang-Seok Hong. "A study on emotion recognition method and its application using face image." 2017 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2017.
10. Gao, Yongsheng, et al. "Facial expression recognition from line-based caricatures." IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 33.3 (2003): 407-412.
11. Noh, Sungkyu, et al. "Feature-adaptive motion energy analysis for facial expression recognition." International Symposium on Visual Computing. Springer, Berlin, Heidelberg, 2007.
12. Bashyal, Shishir, and Ganesh K. Venayagamoorthy. "Recognition of facial expressions using Gabor wavelets and learning vector quantization." Engineering Applications of Artificial Intelligence 21.7 (2008): 1056-1064.
13. Siddiqi, Muhammad Hameed, et al. "Depth camera-based facial expression recognition system using multilayer scheme." IETE Technical Review 31.4 (2014): 277-286.
14. P. Lucey, J. F. 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," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010, pp. 94–101.
15. Johnson, Justin, Andrej Karpathy, and Li Fei-Fei. "Densecap: Fully convolutional localization networks for dense captioning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
16. Zhao, Guoying, and Matti Pietikäinen. "Boosted multi-resolution spatiotemporal descriptors for facial expression recognition." Pattern recognition letters 30.12 (2009): 1117-1127.

17. Zhang, Ligang, Dian Tjondronegoro, and Vinod Chandran. "Random Gabor based templates for facial expression recognition in images with facial occlusion." Neurocomputing145 (2014): 451-464.
18. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
19. Johnson, Justin, Andrej Karpathy, and Li Fei-Fei. "Densecap: Fully convolutional localization networks for dense captioning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

AUTHORS PROFILE

Jyostna Devi Bodapati working as Assistant Professor in the Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research (Deemed to be University), Vadlamudi, India since 2010. She has published good number of research articles in reputed International journals and Conferences. Her areas of interests are Deep learning, Pattern Recognition and Kernel methods for pattern analysis

Dr. N. Veeranjaneyulu is a Professor at Vignan's Foundation for Science, Technology and Research (Deemed to be University), Vadlamudi, India. He has 21 years of teaching experience in the field of Computer Science. He has around 50 publications in International journals and Conference proceedings. His areas of interests are cloud computing and Big-data Analytics

.