Facial Emotion Recognition Using Deep Cnn Based Features

Jyostna Devi Bodapati, N. Veeranjaneeyulu

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. 

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. 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.

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

VGG16, 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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pretrained model</th>
<th>Face detection</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CK+</td>
<td>Resnet50</td>
<td>Yes</td>
<td>86.04%</td>
</tr>
<tr>
<td>CK+</td>
<td>Resnet50</td>
<td>No</td>
<td>81.36%</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patch based SVM [16]</td>
<td>82.5%</td>
</tr>
<tr>
<td>Gabor based templates + SVM [17]</td>
<td>82.5%</td>
</tr>
<tr>
<td>Proposed method</td>
<td>86.04%</td>
</tr>
</tbody>
</table>

Table2. Comparison of the proposed method with the handcrafted 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

Facial Emotion Recognition Using Deep Cnn Based Features


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.