
FACIAL EMOTION RECOGNITION SYSTEM USING CONVOLUTION NEURAL NETWORK

P. Bhargavi¹, CH.Devayani¹, Y.Vijay¹, P.Manikanta¹, N.Aravind Babu, D.Useen²

¹*Student, ECE, Chalapathi Institute of Engineering and Technology, Lam, Guntur.*

²*Assistant Professor, ECE, Chalapathi Institute of Engineering and Technology, Lam, Guntur.*

ABSTRACT: - Automatic facial expression recognition is an actively emerging research in Emotion Recognition. This project extends the deep Convolutional Neural Network (CNN) approach to facial expression recognition task. This task is done by detecting the occurrence of facial Action Units (AUs) as a sub part of Facial Action Coding System (FACS) which represents human emotion. In the CNN fully-connected layers. This research uses the extended Cohn Kanade (CK+) dataset which is collected for facial expression recognition experiment.

KEY WORDS: - Emotion recognition(Human emotion), Convolution neural network, Data sets(Cohn kanade).

1. INTRODUCTION

Automated facial expression recognition is a task in computer vision – and robotics. This problem is an emerging topic of research, especially in social signal processing and affective computing. The challenge in automated facial expression recognition is to recognize each different facial expression and classify into its respective emotion classes [1]. This topic has a wide implementation area, such as in entertainment, education, ecommerce, health, and security [2]. Two approaches for facial expression recognition are the detection of action unit and detection of facial point [3]–[9]. The first approach is implemented by using a framework called FACS (Facial Action Coding System). The framework quantifies facial expression of human by observing the changes in facial muscle when an emotion is triggered [10]. FACS characterizes facial muscle's movement around 44 areas on face; or so-called action units (AUs). Hence, facial expression can be recognized through the existence and intensity of several AUs. Facial expression has two main steps; AU detection and AU recognition.

To do such task, we employ Deep Convolutional Neural Network which has an architecture that consists of filter layers and a classification layer. A filter stage involves a convolutional layer, followed by a temporal pooling layer and a soft max unit. Deep learning methods have been proposed to solve the facial semantic feature recognition tasks [3] and to detect facial point based on Restricted Boltzman Machine [7]. We use database of facial expression which has a ground truth called CK datasets.dataset has been annotated and validated by the expert of AUs. Through the dataset with ground truth we can measure the performance of the proposed method.

2. EXISTING METHOD :- The complexity of facial expression recognition comes from the variability of human facial expression, and it cannot easily model by using prototypic template of facial expression [1], [4]. The first research on this topic initiated by Tian et.al. who proposed facial expression recognition by utilizing FACS [1]. After then, many researches were proposed to detect AU occurrence and AU intensity [3]–[6], [11], [12]. Different approach is by detecting facial points and translate its expression meaning [2], [7], [13]. Two main tasks of Facial Expression Recognition and Analysis (FERA) are feature extraction and expression

classification [5], [12], [14], [15]. Ming et. al. defined three phases of facial expression recognition: facial image registration, feature extraction, and facial expression classification [16]. Most of the existing FERA methods used various pattern recognition techniques to classify different facial expressions based on facial features, which include geometric-feature based approaches, appearance-feature based approaches, texture-based approaches and hybrid features (fusion of them) based approaches. Geometric-feature based approaches use the location of facial feature points (e.g. eye corners, lip corners etc.) or the shape of facial components (e.g. eyes, brows, mouth etc.). Appearancefeature based approaches use the texture feature of the facial image which is robust to the misalignment and the variation of the illumination [12]. One of the most used is Gabor wavelet features. The LBP features were originally proposed for texture analysis, while due to their tolerance to illumination changes and the computational simplicity, they have become very popular for face analysis recently [16]. Feature extraction is an initial step and very important task because it determines the result. The facial expression recognition has two methods: classification and regression methods. Deep learning has been actively used nowadays, including in facial expression recognition [3], [6], [7], [18]. In this paper we contribute a Deep Learning approaches with dropout mechanism to reduce overfitting. 2.1.

Convolutional Neural Network

Convolutional Neural Networks (CNN) are neural network architecture which has multilayers [19]. CNN input and output are array vectors called as feature map. The array dimension depends on the type of input. As an example, audio input has one dimensional array as well as text input; image has 2D array. The output feature map describes the feature extracted from the input. CNN consists of three main layers: convolutional filter layer, pooling/subsampling layer, and classification layer.

Facial expression recognition's methods It describes the summary of facial expression recognition methods and database. The major disadvantages of feature based approaches are big effort should be put on to design and employ various feature extraction methods which are human crafted features. To overcome this drawback, we propose a new approach based on deep learning, a machine crafted features that automatically extract the facial features. **Novelty and Contribution** This research offers two novelty and contribution to the field of Facial Expression Recognition. First, we found that in many research facial features extraction is quite complicated to be designed manually by human, since it is a crucial part against all phases. Here we design automatic feature extraction using deep learning convolutional neural network to detect the occurrence of Action Units. Second, as a contribution we employ CK+ dataset and this makes it different with previous research which used SEMAINE and BP4D. The legendary CK Database is the first database which has comprehensive data and ground truth about facial expression and action units. It has already been validated by the expert. Table 1. Comparison of facial expression recognition methods

PROPOSED METHODOLOGY

We use the definition of basic emotion by Ekman and Friesen who separated emotion into six classes, namely happy, sad, surprise, fear, disgust, angry [10]. Furthermore, we extend two more classes: contempt and neutral; as exist in the original CK dataset. This research focus on recognizing eight different classes of emotion through facial expression analysis using CNN. Fig. 1 depicts those eight classes of emotion.



Fig. 1. Eight classes of basic emotion 4.2.

The dataset used in this research is the Extended Cohn Kanade database (CK+ database). CK+ consists of 10,708 images from 123 different subjects. It has eight classes: neutral, anger, contempt, disgust, fear, happy, sadness, and surprise. Dataset is being preprocessed before training phase. Images are reshaped into 100x100 pixels and then passing into the CNN system.

CNN for facial expression recognition

The architecture of proposed CNN is depicted on Fig.

2. It has two convolutional layers, and two subsampling layers. The first convolutional layer used six masks, or so called c1 layer. The next layer is subsampling layer which has two layers (s1). The second convolutional layer (or c2) has 12 masks. The last subsampling neural network has two layers. The last is fully connected layer which resulting in the class classification.

Experiments

We conducted experiments using different numbers of training and testing data. CK+ database has 10,708 images and we used it all in our experiment. As we can see on table 2, we use varied number of training data as well as testing data. From the experiments we can see that there is a significant decreasing in mean square error as the number of training data raises. While the testing data is the remaining numbers of images data which are not used as a training data. We can observe that the number of testing data is linear to the mean square error. The smaller the size, the smaller the MSE as well. --

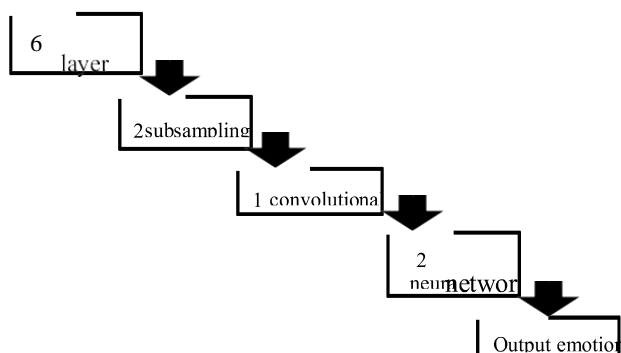


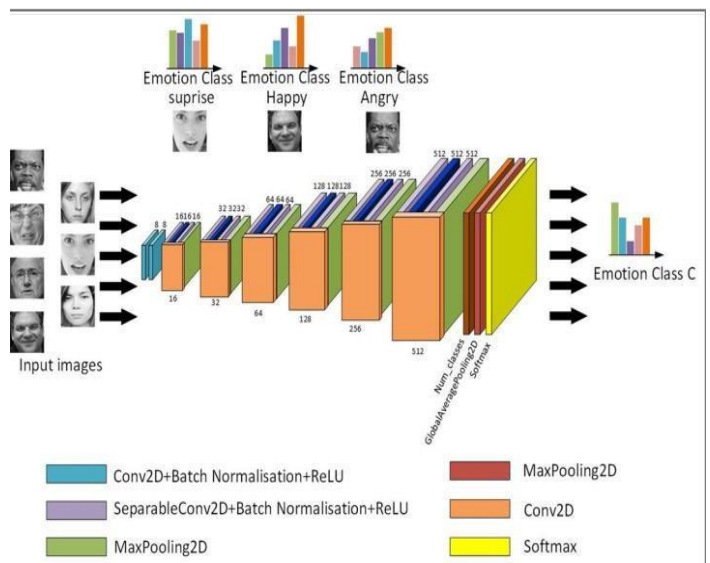
Fig. 2. The Architecture of CNN for Facial Expression Recognition Table By measuring the element of data in using data measurement tools for each class we got the result's summary. The accuracy rate for angry class is 87.73%; contempt 90.95%; disgust 93.46%;

fear 91.75%; happy 96.38%; sad 91.15%; surprise 98.09%; and neutral 92.96%. The average accuracy rate for the entire testing is 92.81%

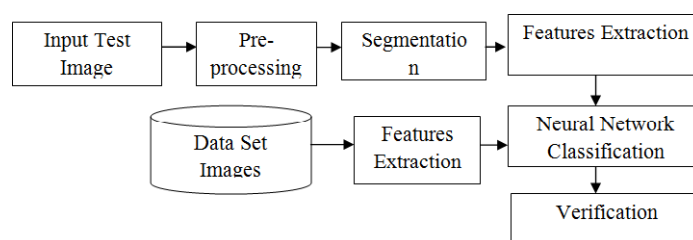
Results and Discussions

The scenario of experiment gives us prominent result about the system performance with the average accuracy rate of 92.81%. The lowest accuracy rate is anger class, 87.73% and the highest one is surprise class, 98.09%. Each class has a misclassification results which indicates that the system needs further improvement. For the next research we should consider changing the whole architecture to give a better result.

2. Experiment results			
No	Num of training data	Num of testing data	MSE
1	8000	2708	0.6381
2	9000	1708	0.4614
3	10000	708	0.3729



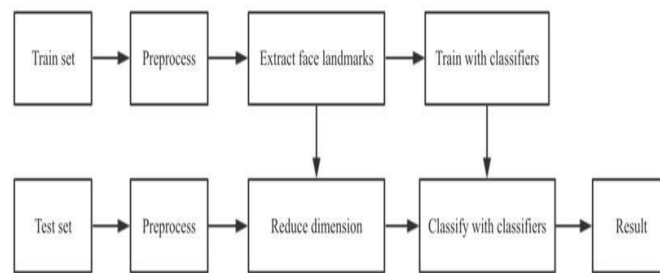
Working



DCNNs are powerful neural networks, but it is important to consider some elements during their design:

Input layer The size of the input volume should be divisible by two several times. These size values range from 32 to 512 [16].

Convolutional layer It is preferable to stack several small filters to use one equivalent large filter because the small filters express more powerful features of the input by preserving the nonlinearities; and require fewer parameters [16].



Conclusion:

We proposed Convolutional Neural Network architecture for facial expression recognition. There are 8 classes of facial expression we tried to recognize. Using the CK+ database we trained using different training data size and the result is the mean square error declines as the number of training data increases. From the experiment we can conclude that the mean square error declines as the training data grows. Furthermore, the performance of the system reaches 92.81% of the accuracy rate. For the next effort, we will put concern on the design of CNN architect to gain the better result.

References:

- [1] Y.-L. Tian, T. Kanade, and J. F. Cohn, “Recognizing action units for facial expression analysis,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR 2000 Cat NoPR00662*, vol. 1, no. 2, pp. 1–19, 2001.
- [2] F. De la Torre and J. F. Cohn, “Facial expression analysis,” *Vis. Anal. Hum.*, pp. 377–410, 2011.
- [3] A. Gudi, H. E. Tasli, T. M. Den Uyl, and A. Maroulis, “Deep Learning based FACS Action Unit Occurrence and Intensity Estimation,” vol. 2013, 2015.
- [4] Z. Ming *et al.*, “Facial Action Units Intensity Estimation by the Fusion of Features with Multi-kernel Support Vector Machine To cite this version : Facial Action Units Intensity Estimation by the Fusion of Features with Multi-kernel Support Vector Machine,” 2015.