 ISSN: 2454-4248
05 – 07

Video Based Emotion Recognition Using CNN and BRNN

Dhivya Devi K
PG Scholar
II Year ME-Embedded System Technologies
Sri Shakthi Institute of Engineering and Technology
Coimbatore 641062, Tamilnadu, India
E-mail: dhivyadevik2019@srishakthi.ac.in

Dr. Nirmala M
Associate Professor-ECE Dept
Sri Shakthi Institute of Engineering and Technology
Coimbatore 641062, Tamilnadu, India
E-mail: drnirmalamadian@siet.ac.in

***Abstract*—** Video-based Emotion recognition is rather challenging than vision task. It needs to model spatial information of each image frame as well as the temporal contextual correlations among sequential frames. For this purpose, we propose hierarchical deep network architecture to extract high-level spatial temporal features. Two classic neural networks, Convolutional neural network (CNN) and Bi-directional recurrent neural network (BRNN) are employed to capture facial textural characteristics in spatial domain and dynamic emotion changes in temporal domain. We endeavor to coordinate the two networks by optimizing each of them to boost the performance of the emotion recognition as well as to achieve greater accuracy as compared with baselines.

Keywords- *Video-based Emotion recognition, spatial temporal features, Convolutional neural network, dynamic emotion changes.*

*****

I. INTRODUCTION

Fuzzy clustering techniques are best suited to segment the pressure ulcer images because the uncertainty of pressure ulcer image is widely presented in data. The most and powerful segmentation is the Fuzzy C Means (FCM) clustering algorithm because, more information is preserved. The focus is this work is to improve the FCM approach and applies it to pressure ulcer image segmentation for detecting Soft white tissue present in pressure ulcer image. The method used to improve FCM are Total Variation (TV) Regularization where noise from the image is eliminated but results in stair casing effect, which is further improved by Anisotropic Diffusion (AD) is to eliminate the stair casing effect (Zhu et al 2008).

II. OVERVIEW OF EXISTING SYSTEM

Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks:

Part-based Hierarchical Bidirectional Recurrent Neural Network (PHRNN) models facial morphological variations and dynamical evolution of expressions, which is effective to extract “temporal features” based on facial landmarks (geometry information) from consecutive frames. A Multi-Signal Convolutional Neural Network (MSCNN) is proposed to extract a Multi-Signal Convolutional Neural Network (MSCNN) is proposed to extract “spatial features” from still frames. We “spatial features” from still frames.

III. PROPOSED SYSTEM ARCHITECTURAL DESIGN

Convolutional Neural Networks (CNN):

Convolutional Neural Networks (CNN) show remarkable performance in image processing due to their strong ability of automatically extracting discriminative representations from single image in multiple tasks such as image classification , object detection , emotion recognition and face recognition , where the spatial dependencies within each image are well modeled. For the temporal dependencies, RNN provides a very elegant way of dealing with sequential data that embodies correlations between data points that are close in the sequences.

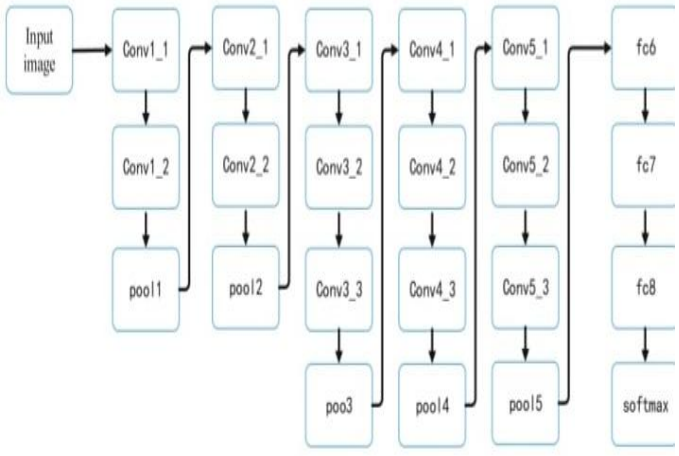

Figure 1: Overall Process of the System

 ISSN: 2454-4248
05 – 07

Fine-Tuning VGG_FACE16 Model:

The VGG_Net [13, 22] is a deep network comprising five stacks of ConvNet, three fully-connected layers and one softmax layer. In other words, it consists of thirteen convolutional layers and three fully-connected layers. All convolution layers are followed by a rectification layer (ReLU) and a max pooling layer. The resulting vector from the last fully connected (FC) layer is regarded as an input of the softmax layer to compute the class probabilities. The model is regularized using weight decay and dropout, which is applied after the first two FC layers to avoid over fitting.

IV. BRNN

BRNN is employed to learn temporal dependencies between the past and future frames. A fully connected layer is used to gather the outputs and learn a sequence representation followed by an 8-class softmax layer for classification. Original RNN is a network with memory and deep in time which is developed for modeling the dependencies in time sequences. Therefore, RNN model is greatly suitable for our sequence classification task with the advantage of encoding contextual information for sequences [30, 31]. Here we employ a bi-directional recurrent neutral network to simultaneously capture forward and backward dynamic transforms of sequence, i.e., two RNN are respectively used to traverse the temporal sequence in a forward or backward behavior. The BRNN can principally be optimized by back propagation through time (BPTT) [24, 25, 31–33] as used in RNN, but the process of forward and backward pass are more complicated. The BPTT is a transformation of back propagation for time sequences. During BPTT, the forward and backward pass are done over the unfolded the bi-directional recurrent nets. There are three procedures in BPTT. The first one is to pass forward running all the input data for one time slice through the BRNN and determine all predicted outputs. The next step is to calculate the part of the objective function derivative for the time slice used in the forward pass. The last step is to update weights.

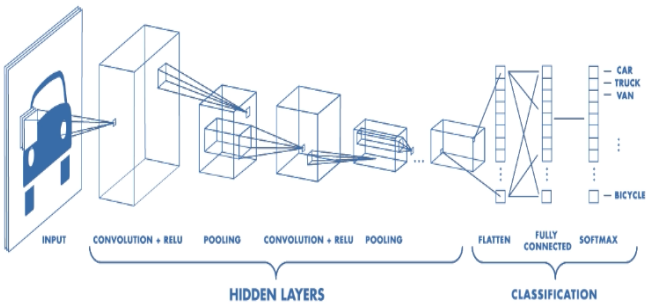

Figure 2: Convolution Neural Network structure

V. EXPERIMENTAL RESULTS

The proposed method is implemented in stages. In the first stage, we used a simple segmentation approach to classify the foreground and background of the images. The foreground consists of single bacilli, touching bacillus and other artifacts. The segmented foreground objects are then given to a trained convolutional neural network (CNN) and the CNN will classify the objects into bacilli and non-bacilli.

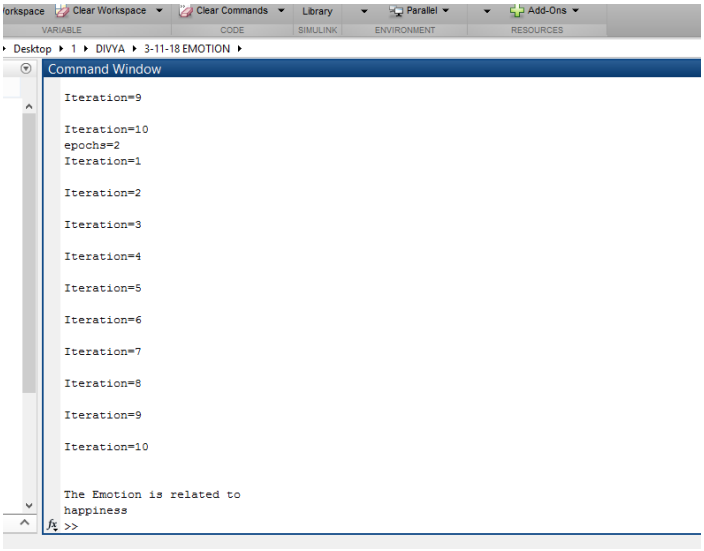

Figure 3: Preprocessing of image and obtained the detection image

Pre-processing (Pixel Normalization): In order to increase robustness, the noisy medical image is pre-processed using Bright Pixel Normalization.

BRNN: A fully connected layer is used to gather the outputs and learn a sequence representation followed by an 8-class softmax layer for classification. Original RNN is a network with memory and deep in time which is developed for modeling the dependencies in time sequences. Therefore, RNN model is greatly suitable for our sequence classification task with the advantage of encoding contextual information for sequences [30, 31]. Here we employ a bi-directional recurrent neutral network to simultaneously capture forward and backward dynamic transforms of sequence, i.e., two RNN are respectively used to traverse the temporal sequence in a forward or backward behavior. The BRNN can principally be optimized by back propagation through time (BPTT) [24, 25, 31–33] as used in RNN, but the process of forward and backward pass are more complicated. The BPTT is a transformation of back propagation for time sequences. During BPTT, the forward and backward pass are done over the unfolded the bi-directional recurrent nets. There are three procedures in BPTT. The first one is to pass forward running

 ISSN: 2454-4248
05 – 07

all the input data for one time slice through the BRNN and determine all predicted outputs. The next step is to calculate the part of the objective function derivative for the time slice used in the forward pass. The last step is to update weights.

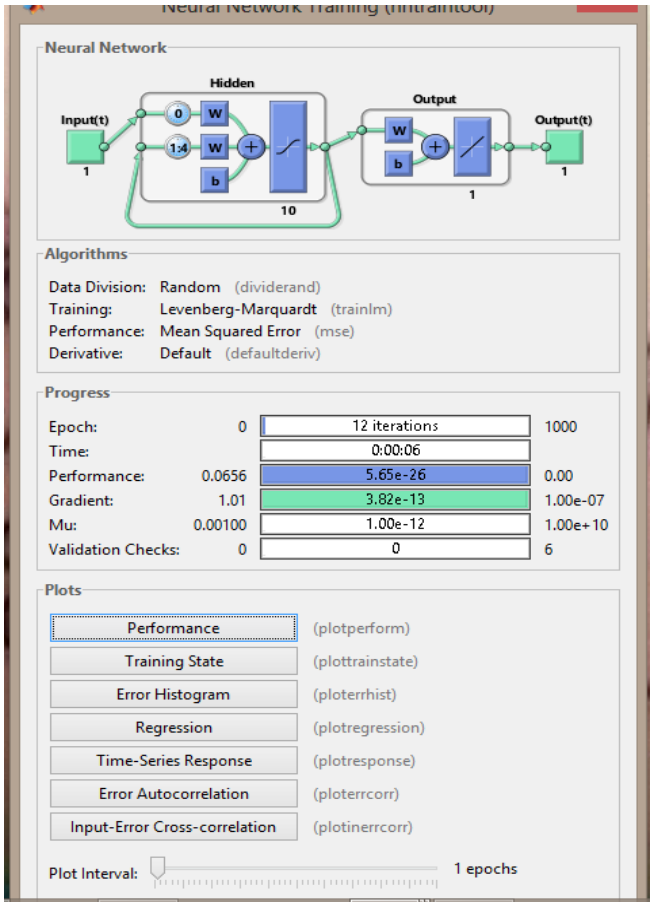

Figure 4: Comparison of simulation graph results

VI. CONCLUSIONS

In this work, we present a facial expression recognition framework consisting of CNN and BRNN that collaborate with each other for emotion recognition based on sequences in CHEAVD. CNN performs image feature extraction of facial expressions and BRNN models contextual dependencies for sequence representations of different emotions. By combing the CNN and BRNN, we learn more powerful emotion feature representations in spatial-temporal domain which have been confirmed to be effective for improving the accuracy of emotion classification. Experimental results over the challenge dataset demonstrate better performance of our framework compared with the baseline.

REFERENCES

[1]. Zheng, W., Zhou, X., Xin, M.: Color facial expression recognition based on color local features. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1528–1532 (2015)

[2]. Zheng, W.: Multi-view facial expression recognition based on group sparse reduced-rank regression. IEEE Trans. Affect. Comput. 5(1), 71–85 (2014)

[3]. Zheng, W., Tang, H., Lin, Z., Huang, T.S.: Emotion recognition from arbitrary view facial images. In: Maragos, P., Paragios, N., Daniilidis, K. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 490–503. Springer, Heidelberg (2010)

[4]. Zheng, W., Tang, H., Lin, Z., et al.: A novel approach to expression recognition from non-frontal face images. In: IEEE 12th International Conference on Computer Vision, pp. 1901–1908 (2009)

[5]. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an appli-cation to facial expressions. IEEE Trans. Pattern Analy. Mach. Intell. 29

[6]. Klaser, M., Marszałek, M., Schmid, C.: A spatio-temporal descriptor based on 3d-gradients. In: BMVC 2008-19th British Machine Vision Conference. British Machine Vision Association, vol. 275, pp. 1–10 (2008)

[7]. Jain, S., Hu, C., Aggarwal, J.: Facial expression recognition with temporal modeling of shapes. In: ICCV Workshops, pp. 1642–1649 (2011)

[8]. Wang, Z., Wang, S., Ji, Q.: Capturing complex spatio-temporal relations among facial muscles for facial expression recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3422–3429 (2013)

[9]. Liu, M., Li, S., Shan, S., Wang, R., Chen, X.: Deeply learning deformable facial action parts model for dynamic expression analysis. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9006, pp. 143–157. Springer, Heidelberg (2015)

[10]. Wöllmer, M., Kaiser, M., Eyben, F., et al.: LSTM-modeling of continuous emotions in an audiovisual affect recognition framework. Image Vis. Comput. 31(2), 153–163 (2013)