International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Facial Expression Recognition Using a Temporal Ensemble of Multi-Level Convolutional Neural Networks

Bavusaipeta srinidhi, M.Tech Student,Department of Information Technology,JNTUH University College of Engineering,Jagtial,Telangana-505501,
Email: srinidhi.b26@gmail.com

Dr.S.Suresh Kumar,Assistant Professor,Department of Information Technology,JNTUH University college of Engineering ,Jagtial,Telangana-505501,
Email:sureshsanampudi@jntuh.ac.in

ABSTRACT: Emotion recognition is indispensable in human-machine interaction systems. It comprises locating facial regions of interest in images and classifying them into one of seven classes: angry, disgust, fear, happy, neutral, sad, and surprise. Despite several breakthroughs in image classification, particularly in facial expression recognition, this research area is still challenging, as sampling in the wild is a demanding task. In this study, a two-stage method is proposed for recognizing facial expressions given a sequence of images. At the first stage, all face regions are extracted in each frame, and essential information that would be helpful and related to human emotion is obtained. Then, the extracted features from the previous step are considered temporal data and are assigned to one of the seven basic emotions. In addition, a study of multi-level features is conducted in a convolutional neural network for facial expression recognition. Moreover, various network connections are introduced to improve the classification task. By combining the proposed network connections, superior results are obtained compared to state-of-the-art methods on the FER2013 dataset. Furthermore, the performance of our temporal model is better than that of the single architecture of the 2017 EmotiW challenge winner on the AFEW 7.0 dataset.

Keywords *– EmotiW challenge, ensemble model, facial expression recognition in the wild, FER2013, hierarchical features, multi-level convolutional neural networks.*

1. INTRODUCTION

The current trend is not only to simplify human machine interaction but also to make machines perceive human feelings. Several aspects can be taken into account to recognize human emotion such as facial expression, voice, or body gesture. Generally, facial information is considered the most common characteristic as we do not need to jump through hoops to record sound in a noisy environment or analyze body gesture while dealing with occlusion problem. However, facial expression recognition is still a challenging issue owing to the variety of head poses and background settings. The challenge in facial expression recognition is to effectively locate and understand facial regions of interest. In the past few years, these tasks were performed by traditional computer vision methods, such as landmark detection, and object modeling. However, it was tacitly assumed that the recognition was performed in a controlled environment. That is, the face and the background were not supposed to be overly complicated to identify. Thus, the possibility of facial expression recognition in the wild using deep neural networks has attracted increasing attention. This would enable machines to react as individual human beings by considering various aspects in an unconstrained sampling scenario, e.g., dynamic illumination, occlusion, and head poses.

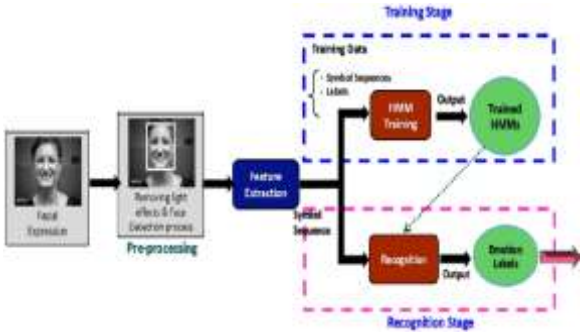

Fig.1: Example figure

Even without saying anything, it can convey a vast range of emotions. Facial expressions convey a person's thoughts, feelings, and actions, and facial expression recognition software can detect these expressions in a photograph of a person's face. During the early 20th century, Americans Ekman and

Friesen [2] created a common set of six globally shared sentiments termed the "basic emotions" (angry; afraid; disgusted; sad; surprised; happy). Facial expression detection has gained a lot of attention recently because of its impact on clinical practice, friendly robots, and education. According to a number of research, emotions have a substantial impact on education. Despite the fact that teachers are already receiving feedback from exams and questionnaires, these methods aren't necessarily the most effective. Teachers can use the facial expressions of their pupils to alter their teaching strategies and resources. A deep learning method widely used in image classification, to identify students' emotions through facial expression analysis. A multistage image processing method is used to extract feature representations. Each of these seven emotions can be recognized in a three-step process that begins with face detection and ends with recognition.

2. LITERATURE REVIEW

Constants across cultures in the face and emotion:

This study addresses the question of whether any facial expressions of emotion are universal. Recent studies showing that members of literate cultures associated the same emotion concepts with the same facial behaviors could not demonstrate that at least some facial expressions of emotion are universal; the cultures compared had all been exposed to some of the same mass media presentations of facial expression, and these may have taught the people in each culture to recognize the unique facial expressions of other cultures. To show that members of a preliterate culture who had minimal exposure to literate cultures would associate the same emotion concepts with the same facial behaviors as do members of Western and Eastern literate cultures, data were gathered in New Guinea by telling subjects a story, showing them a set of three faces, and asking them to select the face which showed the emotion appropriate to the story. The results provide evidence in support of the hypothesis that the association between particular facial muscular patterns and discrete emotions is universal.

Automatic Facial Expression Analysis of Students in Teaching Environments:

Based on students' facial expressions, the teacher in class can know the students' comprehension of the lecture, which has been a standard of teaching effect evaluation. In order to solve the problem of high cost and low efficiency caused by employing human analysts to observe classroom teaching effect, in this paper we present a novel and high-efficiency prototype system, that automatically analyzes students' expressions. The fusion feature called Uniform Local Gabor Binary Pattern Histogram Sequence (ULGBPHS) is employed in the system. Using K-nearest neighbor (KNN) classifier, we obtain an average recognition rate of 79% on students' expressions database with five types of expressions. The experiment shows that the proposed system is feasible, and is able to improve the efficiency of teaching evaluation.

Recognizing student facial expressions: A web application:

The project described in this paper investigates the idea of performing emotion analysis of a student population participating in active face-to-face classroom instruction. Machine learning algorithms are employed on live recordings collected by webcams that are installed in classrooms. The visualization application required to be remotely accessible by the lecturer so the application was engineered as a web application. The output, being a timeline of student emotions monitored throughout and in parallel with the lecture, serves to enable the lecturer and other interested parties to improve the delivery of education.

The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions:

Student engagement is a key concept in contemporary education, where it is valued as a goal in its own right. In this paper we explore approaches for automatic recognition of engagement from students' facial expressions. We studied whether human observers can reliably judge engagement from the face; analyzed the signals observers use to make these judgments; and automated the process using machine learning. We found that human observers reliably agree when discriminating low versus high degrees of engagement (Cohen's $\kappa = 0.96$). When fine discrimination is required (four distinct levels) the reliability decreases, but is still quite high ($\kappa = 0.56$). Furthermore, we found that engagement labels of 10-second video clips can be reliably predicted from the average labels of their constituent frames (Pearson

r=0.85), suggesting that static expressions contain the bulk of the information used by observers. We used machine learning to develop automatic engagement detectors and found that for binary classification (e.g., high engagement versus low engagement), automated engagement detectors perform with comparable accuracy to humans. Finally, we show that both human and automatic engagement judgments correlate with task performance. In our experiment, student post-test performance was predicted with comparable accuracy from engagement labels (r=0.47) as from pre-test scores (r=0.44).

Automatic Detection of Learning-Centered Affective States in the Wild:

Affect detection is a key component in developing intelligent educational interfaces that are capable of responding to the affective needs of students. In this paper, computer vision and machine learning techniques were used to detect students' affect as they used an educational game designed to teach fundamental principles of Newtonian physics. Data were collected in the real-world environment of a school computer lab, which provides unique challenges for detection of affect from facial expressions (primary channel) and gross body movements (secondary channel)-up to thirty students at a time participated in the class, moving around, gesturing, and talking to each other. Results were cross validated at the student level to ensure generalization to new students. Classification was successful at levels above chance for offtask behavior (area under receiver operating characteristic curve or AUC =.816) and each affective state including boredom (AUC =.610), confusion (.649), delight (.867), engagement (.679), and frustration (.631) as well as a fiveway overall classification of affect (.655), despite the noisy nature of the data. Implications and prospects for affectsensitive interfaces for educational software in classroom environments are discussed.

3. METHODOLOGY

Recently, several breakthroughs in image classification have been achieved by using deep convolutional neural networks (CNNs). These architectures consist of two main components: an automatic feature extractor and a classifier. The former produces low-level, mid-level, and high-level features describing simple, moderate, and complex textures, respectively, for the object of interest. Generally, a strong classifier learns the target from a large number of high-level features, and thus a large amount of data should be used to train the network. Owning to the availability of public large-scale image datasets and powerful hardware, deep learning techniques are widely used in various image classification issues, e.g., ImageNet, PASCAL VOC, CIFAR, and Facial Expression Recognition Challenge 2013 (FER2013).

Disadvantages:

1. image classification issues.
2. Despite several breakthroughs in image classification, particularly in facial expression recognition, this research area is still challenging, as sampling in the wild is a demanding task.

In this study, a two-stage method is proposed for recognizing facial expressions given a sequence of images. At the first stage, all face regions are extracted in each frame, and essential information that would be helpful and related to human emotion is obtained. Then, the extracted features from the previous step are considered temporal data and are assigned to one of the seven basic emotions. In addition, a study of multi-level features is conducted in a convolutional neural network for facial expression recognition.

Advantages:

1. Improve the classification task.
2. The performance of our temporal model is better.

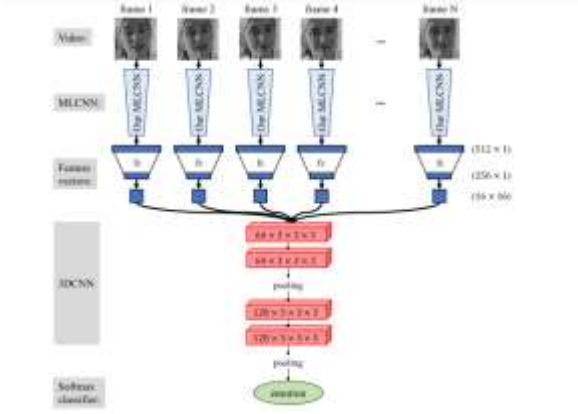

Fig.2: System architecture

MODULES:

- Start camera

- Facial emotion recognition
- Facial expression detection

4. IMPLEMENTATION

CNN:

Convolutional Neural Network is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used. CNN takes an image as input, which is classified and process under a certain category such as dog, cat, lion, tiger, etc. The computer sees an image as an array of pixels and depends on the resolution of the image. Based on image resolution, it will see as h * w * d, where h= height w= width and d= dimension. For example, An RGB image is 6 * 6 * 3 array of the matrix, and the grayscale image is 4 * 4 * 1 array of the matrix. In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, filters (Also known as kernels). After that, we will apply the Soft-max function to classify an object with probabilistic values 0 and 1.

Convolutional Neural Network Architecture

A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

Convolution Layer

The convolution layer is the core building block of the CNN. It carries the main portion of the network's computational load.

This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field. The kernel is spatially smaller than an image but is more in-depth. This means that, if the image is composed of three (RGB) channels, the kernel height and width will be spatially small, but the depth extends up to all three channels.

Pooling Layer

The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually.

There are several pooling functions such as the average of the rectangular neighborhood, L2 norm of the rectangular neighborhood, and a weighted average based on the distance from the central pixel. However, the most popular process is max pooling, which reports the maximum output from the neighborhood.

Fully Connected Layer

Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular FCNN. This is why it can be computed as usual by a matrix multiplication followed by a bias effect.

The FC layer helps to map the representation between the input and the output.

5. EXPERIMENTAL RESULTS

Fig.3: Home screen

Fig.4: User registration

Fig.5: User login

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Fig.6: Main page

Fig.7: User input

Fig.8: Prediction result

6. CONCLUSION

In this study, hierarchical features in CNNs for facial expression recognition were investigated. The experimental results indicated that in addition to high-level features, a number of mid-level features also play a significant role in the classification task. However, this information should be handled carefully. The proposed single MLCNN automatically selected important mid-level and high-level features based on their contribution, and outperformed other single and ensemble models on the FER2013 testing set. To obtain a better representation for facial expression, an ensemble model was proposed by combining MLCNN variants. This model achieved competitive performance and is thus a promising method for facial expression recognition in the wild. In video-based facial expression recognition, the temporal model whose backbone is the proposed ensemble of MLCNNs also achieved comparable performance and outperformed other single video models of the recent winners in the EmotiW challenge.

7. FUTURE SCOPE

Two drawbacks should be considered in future work. As the proposed architecture includes an ensemble model for feature extraction, it requires a number of resources for processing. Moreover, multimodality is taken into account to address the unbalanced data problem in facial expression recognition.

REFERENCES

[1] R. G. Harper, A. N. Wiens, and J. D. Matarazzo, Nonverbal communication: the state of the art. New York: Wiley, 1978.

[2] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” Journal of Personality and Social Psychology, vol. 17, no 2, p. 124-129, 1971.

[3] C. Tang, P. Xu, Z. Luo, G. Zhao, and T. Zou, “Automatic Facial Expression Analysis of Students in Teaching Environments,” in Biometric Recognition, vol. 9428, J. Yang, J. Yang, Z. Sun, S. Shan, W. Zheng, et J. Feng, Éd. Cham: Springer International Publishing, 2015, p. 439-447.

[4] A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, “Recognizing student facial expressions: A web application,” in 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, 2018, p. 1459-1462.

[5] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions,” IEEE Transactions on Affective Computing, vol. 5, no 1, p. 86-98, janv. 2014.

[6] N. Bosch, S. D'Mello, R. Baker, J. Ocumpaugh, V. Shute, M. Ventura, L. Wang and W. Zhao, “Automatic Detection of Learning-Centered Affective States in the Wild,” in Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI ’15, Atlanta, Georgia, USA, 2015, p. 379-388.

[7] Krithika L.B and Lakshmi Priya GG, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric,” Procedia Computer Science, vol. 85, p. 767-776, 2016.

[8] U. Ayvaz, H. Gürüler, and M. O. Devrim, "USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS," Information Technologies and Learning Tools, vol. 60, no 4, p. 95, sept. 2017.
[9] Y. Kim, T. Soyata, and R. F. Behnagh, "Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education," IEEE Access, vol. 6, p. 5308-5331, 2018.
[10] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment," Procedia Computer Science, vol. 125, p. 2-10, 2018.
[11] C.-K. Chiou and J. C. R. Tseng, "An intelligent classroom management system based on wireless sensor networks," in 2015 8th International Conference on Ubi-Media Computing (UMEDIA), Colombo, Sri Lanka, 2015, p. 44-48.
[12] I. J. Goodfellow et al., "Challenges in Representation Learning: A report on three machine learning contests," arXiv:1307.0414 [cs, stat], juill. 2013.
[13] A. Fathallah, L. Abdi, and A. Douik, "Facial Expression Recognition via Deep Learning," in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, 2017, p. 745-750.
[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, vol. 1, p. I-511-I-518.
[15] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no 1, p. 119-139, août 1997.
[16] Opencv. opencv.org.
[17] Keras. keras.io.
[18] Tensorflow. tensorflow.org .
[19] aionlinecourse.com/tutorial/machine-learning/convolution-neural-network. Accessed 20 June 2019
[20] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, p. 1-6.