ISSN 2395-1621

Automatic Recognition of Student Engagement using Deep Learning and Facial Expression

Prof. M. P. Nerkar, Apurva Sawant, Sayali Jawade, Rutuja Shinde,
Rushikesh Thakur

[1]AISSMS IOIT, Pune, nerkar.minal@gmail.com
[2] AISSMS IOIT, Pune, apurvasawant45@gmail.com
[3]AISSMS IOIT, Pune, sayalisjawade55@gmail.com
[4]AISSMS IOIT, Pune, shinderj12@gmail.com
[5]AISSMS IOIT, Pune, thakurrushikesh068@gmail.com

ABSTRACT

Learning is a vital part of education for students of all ages. If the relation between facial expressions and student learning can be examined and captured by an automatic facial expression recognition system, at that point it could alter the education field by giving guides a chance to establish the tone and content of teaching as per their student's interest levels. This paper describes the model for recognition of student engagement using deep learning and facial expressions. Student's videos are captured and are trained on available dataset. Dataset of around 8571 videos is used for recognition of student engagement with the use of CNN. In the first step, the face detection algorithm is trained to detect faces of students from the video. In the next step the deep learning-based model CNN works in three layers to give the engagement result.

Keywords— Deep learning, Facial Expression Recognition, Student Engagement

ARTICLE INFO

Article History

Received: 8th March 2020

Received in revised form :

8th March 2020

Accepted: 10th March 2020

Published online :

11th March 2020

I. INTRODUCTION

Learning is essential part of the education for students. Student Engagement allude to the level of understanding, interest, premium, good faith, and energy when they are learning. One major question is by which manner students are engaged on classroom, this can affect the students learning. the student engagement is significant in many areas such as classroom, MOOCS (massively open online courses), training, tutoring systems. Even though the opportunity for learning is expanded after some time, there is a great dropout rate in every one of these settings. One reason behind this is lack of student engagement. The learning process will be interesting if the instructor or the coach can monitor the engagement level of the students. Student engagement can be calculated by traditional way, taking exams of the students or by observing student's behaviour and performance by teacher. Instead Deep learning and facial expression recognition-based techniques can be used to determine the student engagement. It is very inconvenient to analyse the student engagement in the classroom while teaching. Automated recognition of student's engagement helps the teacher to recognize the student's engagement and they can adapt various teaching strategies to engage the students. In this project, the student engagement is evaluated based on the facial expression recognition. Facial expressions pass on emotions and give verification about student's character and points. Facial recognition is a cutting-edge innovation that aides in observing and distinguishing human expressions from a picture or video. This framework detects facial emotions based on the frames extracted from the video. Dataset of around 7142 videos were used for the training and testing purposes and 1429 videos were used for the validation of the system. Facial expression recognition technology is a sentimental tool which can automatically detect the human face expressions. It has the wide range of applications in Medicine, Psychology and Business.

II. RELATED WORK

1. *FACIAL EMOTION RECOGNITION*

Facial expressions convey attitude and intentions of people. These facial expressions where studied from more a century ago. Much progress has been made in the facial expression recognition, but more work is still necessary to get a satisfactory framework. Most of the work uses a framework of six universal emotions: fear, sadness, anger, surprise, happiness and disgust, with a further neutral category.

Kahouet al. [5] won the 2013 Emotion Recognition in the Wild (EmotiW) Challenge for recognising facial expressions using Convolution Neural Networks CNNs. Kahouet al. applied CNNs for extracting visual features accompanied by audio features in a multi-modal data representation.

Tang et al. [10] won the 2013 Facial Expression Recognition (FER) challenge for building another CNN model followed by a linear support vector machine which was trained to recognize facial expressions.

Yu et al. [11] applied a face detection method to detect faces and remove noise in their target data samples. They employed a CNN model that was pre trained on the FER-2013 dataset. Goodfellow [4] and fine-tuned on the Static Facial Expression in the Wild (SFEW) dataset . Zhang et al. [12] applied CNNs which captured spatial information from video frames. The spatial information was then combined with temporal data to recognize facial expressions.

Pramerdorferet al. [6] achieved the state-of-the-art result on the FER-2013 dataset by combining the modern deep architectures such as VGGnet .Mollahosseiniet al. [2] applied face registration processes, aligning and extracting faces, to achieve better performances. They trained CNN models across different well-known datasets like FER to enhance the generalizability of recognizing facial expressions.

2. *Engagement Recognition*

Engagement is an important part of human-innovation connections and is characterized diversely for a variety of uses, for example, web search tools, web based gaming stages, and versatile health applications Tang [10]. According to Monkaresiet al. [1], most of the definitions describe engagement as attentional and emotional involvement in a task. Engagement has been distinguished in three distinctive time scales: the whole video of a learning session, 10-second video clasps and pictures. applied linear support vector machines (SVMs) and Gabor features, to classify four engagement levels: very engaged in the task, engaged, nominally engaged and not engaged at all. In this work, the dataset includes 10-second videos annotated into the above four levels of engagement by observers, who are analysing the videos. Bosch et al. [7] detected engagement using Bayesian classifiers and AU's. Monkaresiet al. [8]used a face tracking engine to extract facial features and a classification toolbox(WEKA)to classify the features into two classes: engaged or not engaged. They annotated their dataset, including 10-second video cuts, using self-reported data collected from students during and after their tasks. In any case, the engagement levels of students can change during 10-second video cuts, so allotting a single name to each clasp is troublesome and once in a while off base.

III. Methodology

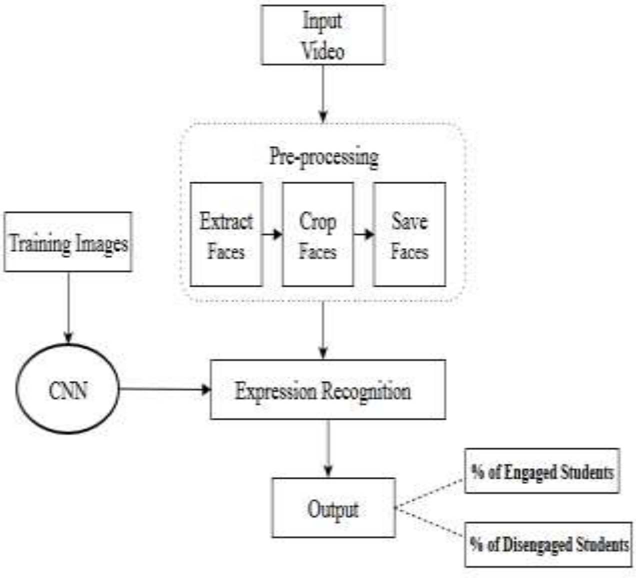

Fig 1.System Architecture

1. *Extract Frames:*

First step is to take video as an input and extract frames after specific time interval. For this we have used “ffmpeg” library.

2. *Pre-process the data:*

This is a pre-processing step. Facial Emotion Recognition has a particular area of focus in the entire image. This region is also called as Region of Interest (ROI). Therefore, faces (ROI) in each frame are cropped out using Python's OpenCV Library. OpenCV is a cross-platform library which mainly focuses on image processing, Face detection and Object detection. These cropped images are then fed as an input to the next layer.

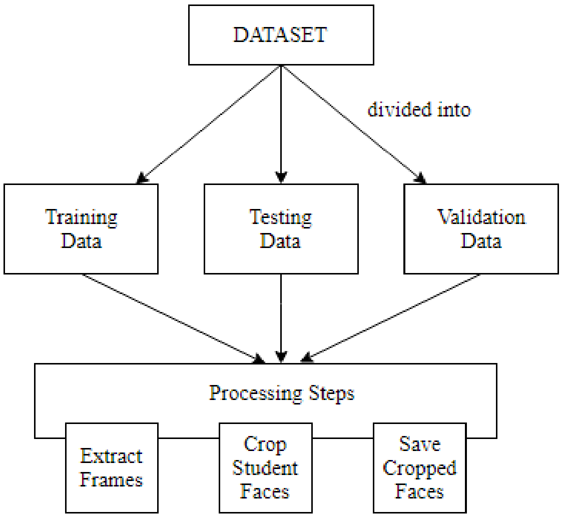

Fig.2 Preprocessing

3. *Building the Convolution Neural Network (CNN):*

The cropped images are then resized to 200*200 pixels. The resized images are then converted to gray scale and passed as an input to the neural network. The CNN has following Layers:

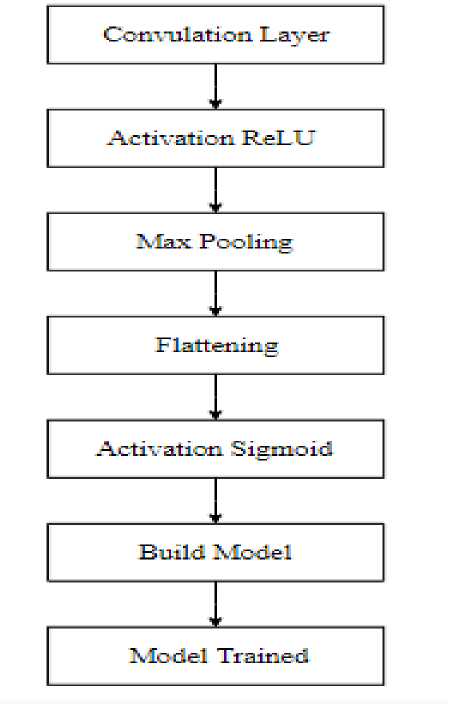

Fig.3 Training Phase

3.1. Convolution 2D Layer:

Convolutional Neural Networks apply a filter to the input to make a feature map that outlines the presence of identified features in the input. This layer consists of 32 filters having 3X3 kernels and produces a feature map.

3.2. Activation ReLU Function:

ReLU (Rectified Linear Unit) Activation function is used in this operation. It is a non-linear operation which replaces all negative pixel values in the feature map with zero.

3.3. Max Pooling:

In this layer, 2X2 Max Pooling operation is carried out. The Pooling Layer reduces the image dimensionality without losing important patterns or features of the image. We took a 2X2 window to perform pooling operation. Here, we take the largest element in the window and replace it with the window.

3.4. Drop-Out:

Dropout of 25% is taken which intends to decrease the complexity of the model with the objective to forestall overfitting.

3.5. Flatten

Flattening is converting over the information into a 1-dimensional array for inputting it to the following layer. We flatten the output of the convolutional layers to make a single long element vector. This vector is given as an input to the dense layers. This layer gives probabilities to each class of emotion to the output layer. As the dataset is multi-labeled, we used sigmoid() to get the probabilities.

4. *TRAINING THE MODEL*

The above model was trained using around 10,000 images of the training dataset. Around 1800 images were used for validation of the model. The validation data helps us with validating how well the model will perform on unseen input. RMSprop optimizer was used for training the model. RMSprop optimizers were effective in increasing the learning rate of the model. The trained model is saved.

5. *LOAD MODEL*

The HDF5 file can be loaded to predict the emotion of the input image.

6. *CLASS PREDICTION*

Finally, the trained model is used to get predictions on new images. Engagement, Boredom, Confusion and Frustration are 4 classes for images. The model successfully predicts which class does the input image belongs to from the 4 classes.

IV. EXPECTED RESULT

The model will be able to predict overall percent of students engaged and disengaged during the entire lecture.

Dataset consists of 8571 videos divided into training, testing and validation sets. Frames are extracted from all the videos after specific time interval. In the first step, faces are extracted and passed to the Convolution Neural Network. In the next step, the network distinguishes whether the students are engaged or disengaged.

The results are displayed using Pie Chart as below.

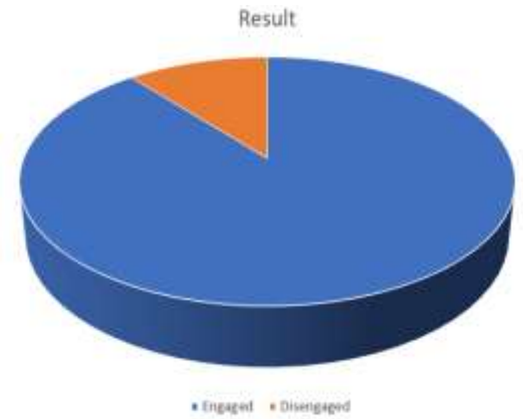

Fig.4 Result

This network is expected to give an accuracy of about 95%.

V. CONCLUSION

Facial Emotion Recognition system analysed various facial emotions such as Boredom, Engagement, Frustration and Confusion. Video can be uploaded to this system and the system will predict the percentage of students Engaged and Disengaged during the lecture by extracting frames at regular intervals. Proposed system reveals overall percentage of engaged and disengaged students throughout the lecture. This result will help teachers analyse response of students to the lecture and accordingly they can adopt different strategies which will keep students more engaged and concentrated. The performance of this system is greatly affected by quality of the video being uploaded and arrangement of the students during lecture. Future research will be directed towards adding more features like audio and eye glaze for predicting the engagement levels which is ought to increase the accuracy of the model.

REFERENCES

[1] Monkaresi, H., Bosch, N., Calvo, R.A.,D’Mello, S.K.: Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Transactions on Affective Computing 8(1), 15–28 (2017)

[2] Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: WACV. pp. 1–10. IEEE (2016)

[3] Khorrami, P.R., How deep learning can help emotion recognition. 2017, University of Illinois at Urbana-Champaign.

[4] Goodfellow, I.J., et al., Challenges in representation learning: A report on three machine learning contests. Neural Networks, 2015. 64: p. 59-63.

[5] Kahou, S.E., Bouthillier, X., Lamblin, P., Gulcehre, C., Michalski, V., Konda, K., Jean, S., Froumenty, P., Dauphin, Y., Boulanger-Lewandowski, N., et al.: Emonets: Multimodal deep learning approaches for emotion recognition in video. Journal on Multimodal User Interfaces 10(2), 99–111 (2016).

[6] Pramerdorfer, C., Kampel, M.: Facial expression recognition using convolutional neural networks: State of the art. arXiv preprint arXiv:1612.02903 (2016).

[7] Bosch, N., D’Mello, S., Baker, R.,Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., Zhao, W.: Automatic detection of learning-centered affective states in the wild. In: IUI. pp. 379–388. ACM (2015).

[8] Joseph F. Grafsgaard, Joseph B. Wiggins,Kristy Elizabeth Boyer, Eric N. Wiebe, and James C. Lester, “Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis”, 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction.

[9] Joseph Psotka, Sharon A. Mutter, “Intelligent Tutoring Systems: Lessons Learned”, Lawrence Erlbaum Associates, ISBN 0-8058-01928, 1988.

[10] Tang, Y.: Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239 (2013).

[11] Yu, Z., Zhang, C.: Image based static facial expression recognition with multiple deep network learning. In: ICMI. pp. 435–442. ACM (2015).

[12] Zhang, K., Huang, Y., Du, Y., Wang, L.:Facial expression recognition based on deep evolutional spatial-temporal networks. IEEE Transactions on Image Processing 26(9), 4193–4203 (2017).

© 2020, IERJ All Rights Reserved