EmoAI Smart Classroom: A Student Emotional and Behavioral Engagement Recognition System

Folasewa Abdulsalam Adekunle Afolabi Soufiane Hourri
Obafemi Awolowo University Cadi Ayyad

Abstract

Due to various distractions and a lack of contact between students and professors, the issue of student disengagement has become crucial in the contemporary educational system. Large offline classes make this problem worse because it is emotionally tasking for professors to keep an eye on student involvement and maintain the proper amount of connection at the same time. It is possible to better grasp and enhance the dynamics of a lecture by having a better understanding of the basic academic emotional states exhibited by students in the classroom. Hence, an emotional and behavioral engagement system that detects and analyzes academic affective states and behaviors of students in the classroom while sending feedback to the lecturer was proposed in this project. The Dataset used is DAiSEE (Dataset for Affective States in E-Environment), the architecture used is YOLO V8, Unified Modelling Language (UML) was used to design the system, Python was used to implement using RoboFlow Framework, the system was also evaluated through testing of pre-recorded classroom videos.

Review of Related Works

In a study by [11], they developed a vision-based automated system for student group engagement in a large offline classroom environment by analyzing students' academic affective states through facial expressions. The facial expression dataset used was created from classroom lecture video frames of more than forty students. A Convolutional Neural Network-based architecture was used as the framework for the recognition system. Multi-modal analytics went a step higher in a research conducted by [12] where they monitored students' mental states (boredom, frustration, concentration, and confusion) during classroom discussions. They developed an intelligent multi-sensor system using modalities such as facial, heart rate, and acoustic indicators. They trained a set of machine learning algorithms such as support vector machine, random forest, and multilayer perceptron using features from facial, heart rate, and auditory modalities. However, the implementation of this study is expensive because it needs physical devices such as Apple Watch, and Air pods to measure

Methodology

Data Collection

DAiSEE (Dataset for Affective States in E-Environment) is the first multi-label video classification dataset comprising 9,068 video snippets captured from 112 users for recognizing user affective states of boredom, confusion, engagement, and frustration in the wild. the dataset has four levels of labels namely-very low, low, high, and very high for each of the affective states The DAiSEE dataset was collected from a link provided in an article published on Paper with Code.

Fig 1 Screenshot of DAiSEE

Frame Extraction

In order to capture the temporal changes in affect, a tool known as FFMPEG was used to extract frames at 30 frames per second (30fps). The frames extracted from the video (9,068) totaled 2,723,882.

Data Annotation

Semi-Manual annotation using Label Assist on Roboflow (see fig ii). 707 images from the frames were loaded into the Annotation section of RoboFlow, and each frame was mapped to a class. The classes were: Confusion, Boredom, Engaged, and Frustration. The labeled data then formed the data for model building

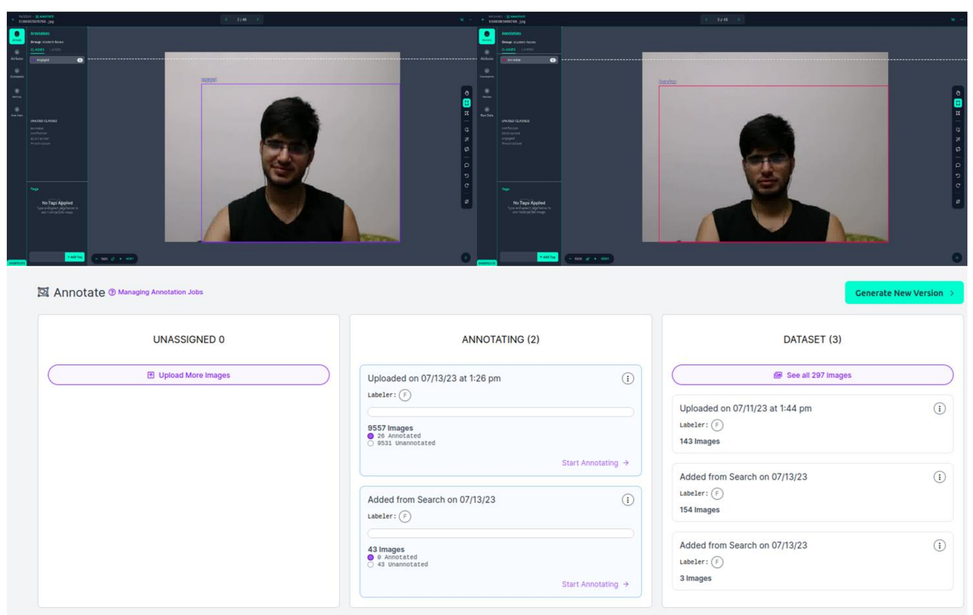

Fig 2 Data Annotation

Model Building

This talks about the data pipeline from splitting to making inferences on Roboflow. The stages involved are explained below:

Data Split

To prepare the dataset for benchmarking, we create a data split into train, validation, and test sets.80- 10-10 is the ratio split used on the dataset. 80 for Training, 10 for testing and validation. (see fig iii)

Data Preprocessing

Grayscale, Resizing, Auto Orient, bounding box brightness were applied.

Training the Model

Yolo V8, the newest and fastest state-of-the-art model for computer vision tasks. The pre-processed data was trained on the pre-loaded model

Results

The result of the model training gave 74.5% precision with the following loss results (see fig iv).

Inference

The system was tested by uploading pre-recorded classroom videos. The result is shown in fig v

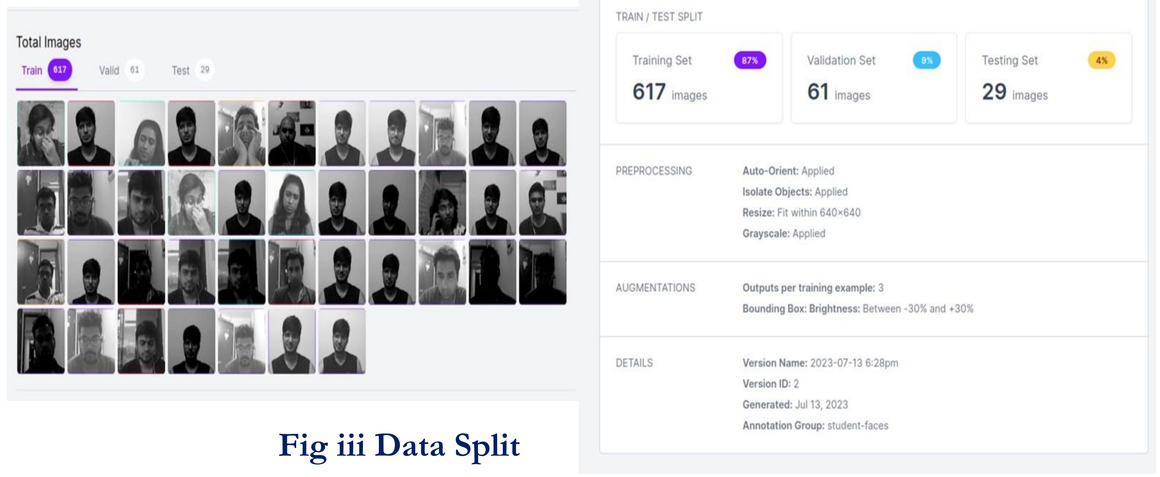

Fig iii Data Split

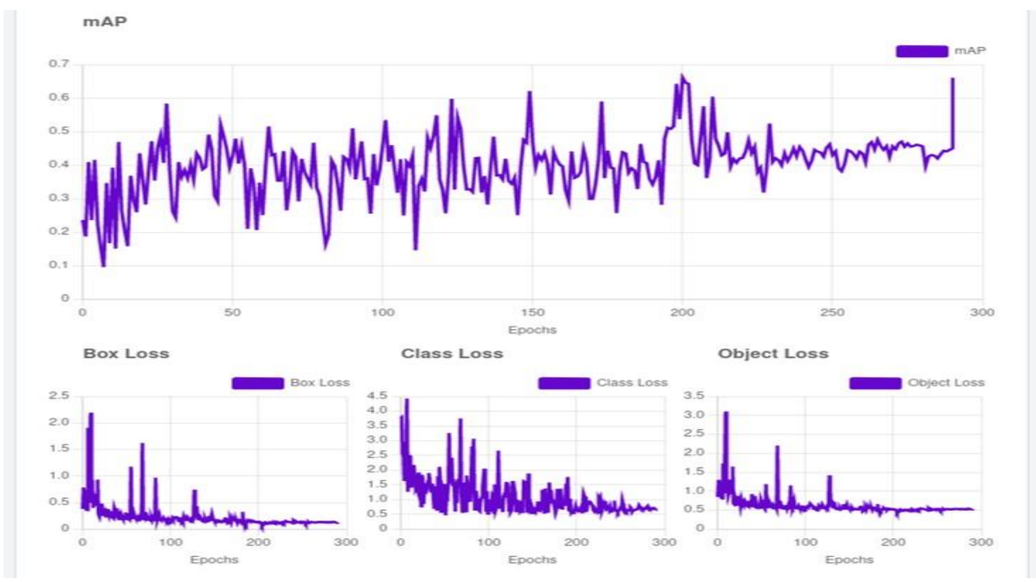

Fig iv Class Loss Results

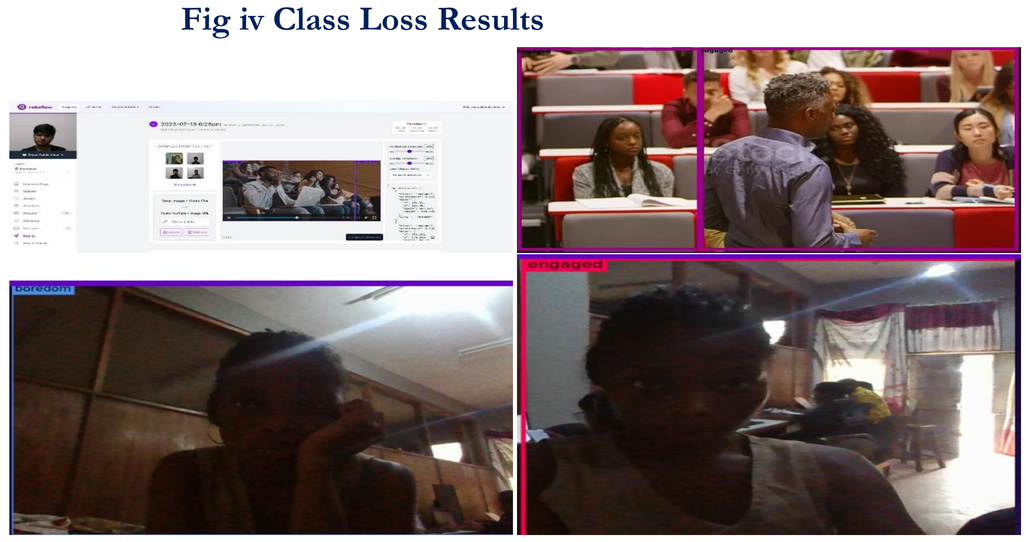

Fig v Model Inference on Classroom Video

Conclusion and Future Work

In this work, we presented the DAiSEE dataset to develop an engagement detection system. The unique proposition is the fact that it is intended for offline classrooms. We hope that this system will be the third eye for tutors in the classroom while lectures are being delivered without being invasive. This work, based its working architecture on the work done by [11]. However, in place of the Convolutional Neural Network used, YOLO V8, which is the current state of the art was used. Although over two million frames were collected, while building the system, seven hundred data were used as an initial proof of concept. The environment provided for evaluation is one in which the tutor uploads a pre-recorded classroom. video after the lecture. One of the limitations of the system is that the system cannot recognize faces wearing glasses. Going forward, the system can be upgraded to work in real-time to give alerts when it detects the engagement level of the students is low or below the threshold. Tutors can be more aware of the atmosphere in the classroom which in turn can help improve the dynamics of lectures delivered.