A Real-time Application to
Alert Instructors of Their Confused Students

Yu Fang

March 26, 2024

Abstract

In light of the evolving educational landscape, our project seeks to leverage advanced facial recognition and gaze tracking technologies to analyze and enhance student engagement in real-time [1–3]. Building on preliminary findings from the previous semester's investigation into teaching styles and student interactions, this research aims to delve deeper into the students' facial expression. We will first train a machine learning model to recognize confusion or not using openly available datasets [4, 5], then apply it to classroom videos we obtained for Edusense project. Furthermore, we will develop this methodology into a real-time web application to demonstrate it's real-time tracking capabilities. This project will further explore the ever-expansive utility of automated sensing, providing instructors more tools than ever to manager their own teaching for the purposes of fostering a more engaging and effective learning environment.

Introduction

Our previous work revealed a good performance in predicting classroom activity (COPUS [6], 1) using automated sensing output, including, but not limited to, student gaze movement in 3D. This semester, we further explored on how students' gaze reflect their attentiveness throughout a semester, and even throughout an 80-minute class. We decided to identify motifs of students gaze and to explore the possibility of using this as the proxy for student engagement. The data will be provided by Edulyze [7]. Throughout the investigation, we also managed to use DeepFace [2] to improve identity tracking. In light of the capabilities of DeepFace, we propose taking a step further and identify when an instructor has lost their students' attention. One idea would be to predict the state of confusion via students' facial expression. We also believe that providing instantaneous feedback on the average students' comprehension would facilitate and can serve as a guide for the pace of lectures.

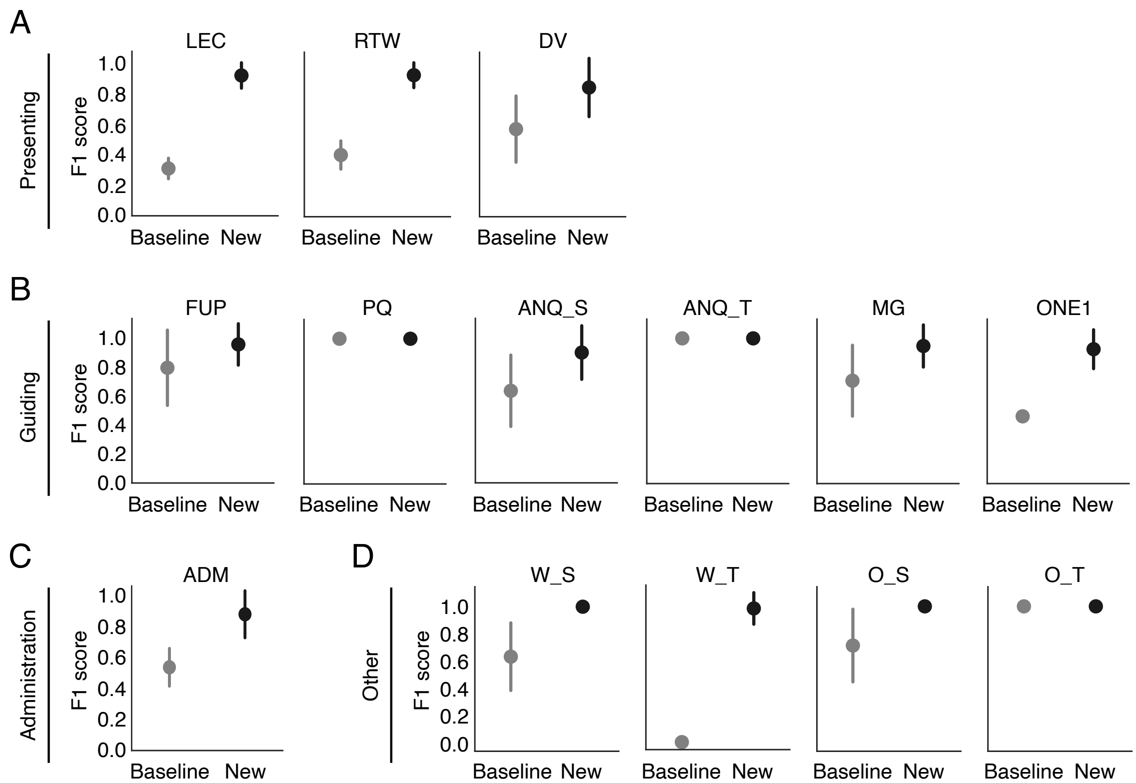

Figure 1: Automated sensing correctly predicting student activity. Mean ± standard deviation of performance on 20 randomly partitioned held-out new classroom data for each of the four main categories (A: Presenting; B: Guiding; C: Administration; D: Other). New is after incorporating 5 sessions from the new classroom - 407, whereas baseline is before. Note that the description of these shorthanded activity can be found in Smith et al. [6]

Methodology

To investigate state of confusion, we will need to first build a facial expression classifier that accurately predicts the confused. We will first inquire facial images of confusion from two sources [4, 5]. Upon obtaining the dataset, we will re-build the emotion recognition module from DeepFace using similar amounts of images of confused versus not-confused (Fig. 2). Once we built this binary classifier and validated the performance, we will test it on classroom videos we have from previous study (Fig. 3). Since we have more comprehensive sensing data from prior investigations, we can summarize the subtle dynamics in posture that accompanies a confused facial expression. Lastly, we will develop a web application provides real-time feedback on how well the students' comprehend the current material (Fig. 4).

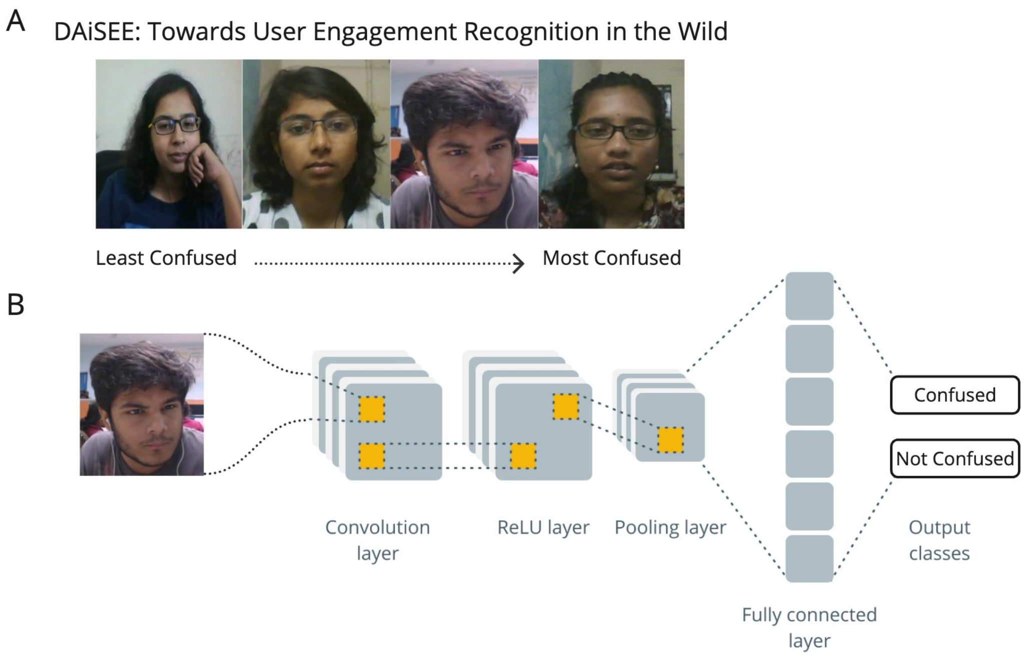

Figure 2: Train a custom Convolution Neural Network specifically at identifying students' confusion. A) Each image has been annotated as confused or not in the DAiSEE dataset [5]. B) To implement a confusion classifier in DeepFace, we will train a convolutional neural network based on the DAiSEE dataset.

Data acquisition and preprocessing

Analytical framework

Previous work has demonstrated utility and insights primarily obtained by obtaining low-level features from videos of students interacting with their instructor and amongst themselves. To provide more real-time insights of how students' comprehend course materials, we will perform facial expression analyses. Specifically, we will obtain facial expression datasets that were annotated with confused (among other expressions). These images and it's respective annotations will be mapped via a Convolutional Neural Network (Fig. 2). Since we do not possess actual ground truth of students' being confused (could be subjective without interviewing the students), we believe mapping our previous work of a composite features that are associated with gaze and posture could help shed some interpretable insights into how well we did (Fig. 3).

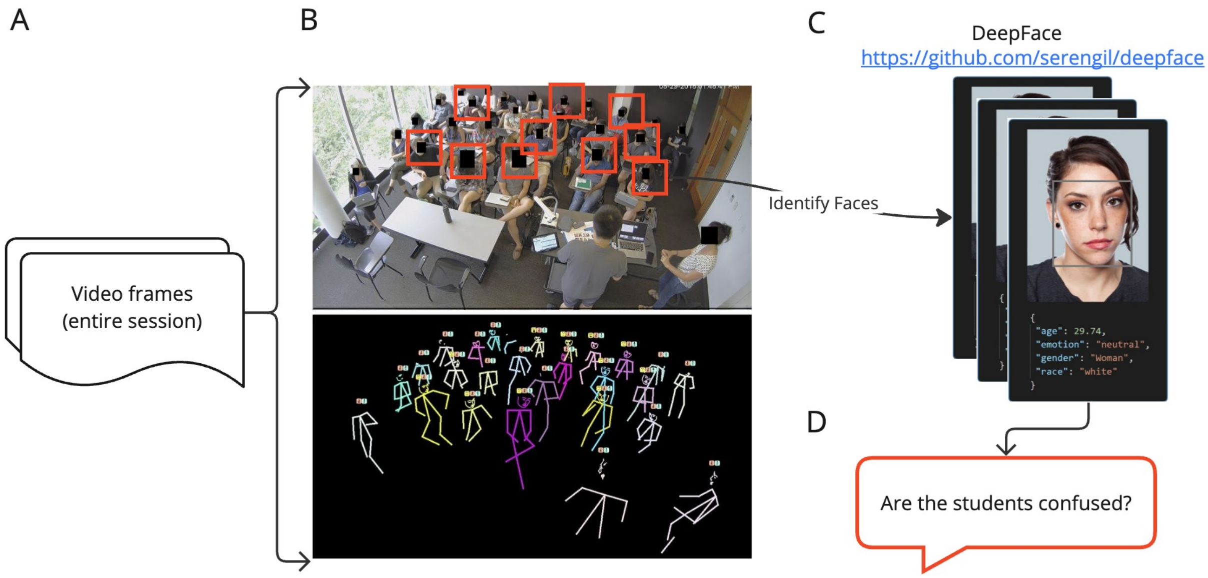

Figure 3: Predict the state of confusion during pre-recorded classes using custom-trained neural network. A-B) In each frame, facial images are segmented out using DeepFace [2]. C-D) These segmented faces are then feed into a custom-trained convolution neural network specifically trained to identify students' state of confusion.

Real-time application

To further expand the utility of identifying students' state of confusion, we will develop a real-time web application for and alert instructors during a class if their students overall become confused of the course material (Fig. 4).

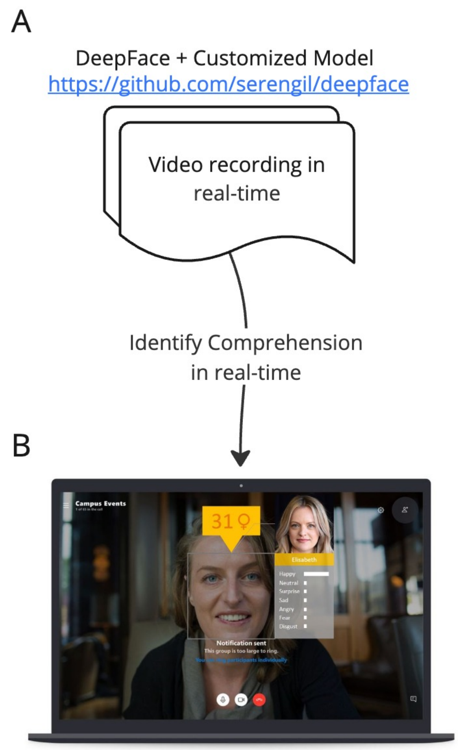

Figure 4: Web application to enable the instantaneous feedback during a class. A) Real-time classroom video is analyzed every N seconds. B) Output each student's confusion prediction, while also provide a population level confusion in real-time.

Expected outcomes

Overall, we are trying to identify students ability to comprehend during a class. We aim to unveil nuanced insights into specific time points at which students' become confused and potentially start to lose attention. By incorporating the pose dynamics already obtained through Edulyze, we expect to find engagement motifs that correlate with state of confusion. If we do, this approach will empower educators to identify student's attentiveness in real-time to potentially enhance their audience's attentiveness.

Discussion

Our research aims to bridge the gap between technological advancements and educational methodologies by harnessing the power of facial recognition, facial expression analysis, and pose dynamics to fully capture student engagement. Beyond post-hoc analyses, we will also realize a real-time web application that could alert the instructors during the class, not after. Overall, we aspire to contribute to the broader goal of enhancing academic success and have instructors dynamically react to real-time reduction in audience's engagement levels.

References

1. Parkhi, O. M., Vedaldi, A. & Zisserman, A. Deep Face Recognition. *British Machine Vision Conference,* 1–41 (Dec. 2015).

2. Serengil, S. I. & Ozpinar, A. LightFace: A Hybrid Deep Face Recognition Framework. *Proceedings - 2020 Innovations in Intelligent Systems and Applications Conference, ASYU 2020* (Oct. 2020).
3. Serengil, S. I. & Ozpinar, A. HyperExtended LightFace: A Facial Attribute Analysis Framework. *7th International Conference on Engineering and Emerging Technologies, ICEET 2021* (2021).
4. Landowska, A., Szwoch, M. & Szwoch, W. Methodology of affective intervention design for intelligent systems. *Interacting with Computers* **28,** 737–759. ISSN: 09535438 (Nov. 2016).
5. Gupta, A., Cunha, A. D. '., Awasthi, K. & Balasubramanian, V. DAiSEE: Towards User Engagement Recognition in the Wild. https://arxiv.org/abs/1609.01885v7 (Sept. 2016).
6. Smith, M. K., Jones, F. H., Gilbert, S. L. & Wieman, C. E. The classroom observation protocol for undergraduate stem (COPUS): A new instrument to characterize university STEM classroom practices. *CBE Life Sciences Education* **12,** 618–627. ISSN: 19317913. https://www.lifescied.org/doi/10.1187/cbe.13-08-0154 (Dec. 2013).
7. Ahuja, K. *et al.* EduSense: Practical Classroom Sensing at Scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol* **3,** 71. https://doi.org/10.1145/3351229 (2019).