Dataset for Emotional Intelligence on Online Learning Platforms

Naman Pundir, Dr. David Doermann
University at Buffalo, SUNY
{namanpun, doermann}@buffalo.edu

Abstract

Since Covid-19 pandemic, online learning and teaching has developed and used rapidly with several available platforms and different education models. If we compare traditional teaching and learning with online platforms there are many advantages of online learning as not constrained by location and a wide variety of features using these platforms which can't be achieved in traditional teaching models. In traditional learning models it's difficult to see every face reaction and students' understanding but in online models we can see each face and can detect students' emotions for the sentiment analysis of the online meeting or lecture. This paper proposes and creates a dataset for an online learning platform which can be used to train the model and test the model on live lectures to determine students' understanding of the lecture or meeting which can help to improve the lecturer and learner interaction. There are 6 emotions in humans according to the facial expression namely happy, sad, fear, disgust, surprise, and anger. [1, 2], This paper works on creating a dataset from online learning platform to train and validate model to detect emotions in online learning platforms. Detection of emotions from the texts can be also done using LSTM or transformers like BERT and Hugging face [3].

[GitHub Repository for the paper: https://github.com/namanpundir/Dataset-for-Emotional-Intelligence-on-Online-Learning-Platforms]

1 Introduction

There are 6 emotion categories that are widely used to describe humans' basic emotions, based on facial expression: anger, disgust, fear, happiness, sadness and surprise.

FIGURE 1: Types of emotions in humans. [4]

With the development of artificial intelligence and deep learning, numerous Facial expression and recognition algorithms have been proposed to deal with the expression information in facia representations, which has improved the accuracy of recognition gradually and achieved better performance than traditional machine learning methods.

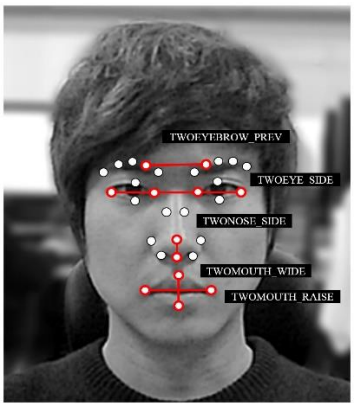

FIGURE 2: Facial analysis points used for face expressions. [5]

1.1 Related Work

A lot of related work has been done in the field of emotion detection and sentiment analysis on images and videos but few on Online learning platforms as the new method of online learning of online video classes is not that old. Some deep learning approaches [4][5] has been done to recognize the emotions in the online lectures videos by training the deep learning model on FER-2013 Dataset [6]. But these models are trained on FER 2013 and applied on Online learning platform to obtain estimation of the sentiment behind the lectures. This paper creates a dataset with ground truth of the given videos which helps to estimate an accuracy of the model after training with FER 2013 and testing it on online lectures recording.

2 Online Learning Platforms

There are many online learning platforms which is used wordwide for education purposed or industrial work. Since pandemic the need of such online platforms increased rapidly. Advances in technological delivery modalities have created a huge number of online education platforms and substantially improved educational flexibility, allowing teachers to use a variety of technical approaches to help teaching without fear of losing control. Conventional classroom-based courses have a maximum number of students, and students from different regions can speak in real time without having to consider traffic as well as other difficulties. For students' reference, the same instructional materials that are used in traditional classes can be published to these platforms.

FIGURE 3: Online video conference and meeting applications. [6]

Teachers can use the video meeting approach and use the camera included in devices to collect and recognize students' facial expressions in real time in platforms that have online teaching features, such as DingTalk, Zoom Meetings, Microsoft Teams, Skype, Google Meet and Rain classroom and much more.

3 Datasets: Online Learning Lectures

To create our dataset and ground truth to train and validate models to detect emotions and sentiments of students in the lectures we have taken publicly available 14 zoom meetings lectures:

https://drive.google.com/drive/folders/1qwQJyK34oEJ0fsC3BMAxUyuTQ-9HCOOL?usp=sharing

These videos comprise of different subject and different professors' lectures. Size of this dataset is 4.5 GB and the length of lectures varies from 23 minutes to 115 minutes.

FIGURE 4: Snapshot of videos used to create this dataset.

4 Frames at different Timestamps.

To process these videos, we took screenshots or frames of these videos at different timestamps to get the overall sentiment behind the whole lecture. We have taken 21 frames at different timestamps using the following calculation:

X = (total minutes * 60) / (number of timestamps)

After each X second, we took the frame from the video. So, from each video we took 21 frames. So, we got 294 frames in total from all 14 videos.

https://drive.google.com/drive/folders/1rhFvavKbacYHUlYnmp_GAho2k4y-7W-_?usp=sharing

These frames comprise of multiple faces of the participants in the lecture including the instructor. To get the overall sentiment behind the lecture these frames will be used to estimate the average response of the students for the lecture.

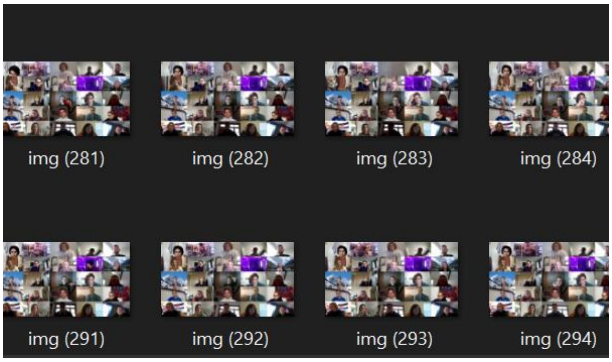

FIGURE 5: Snapshot of frames used to create this dataset.

5 Cropping Faces from Frames

Face recognition involves three basic steps which include face detection, face recognition, and face extraction. Any system needs to encapsulate the image and then manage as well as record the vital features to determine the location of the face. For the recognition of the captured image it keeps records of various features such as skin color, skin tone etc.

Face recognition takes an image from a video or a camera as input and returns the image topic that has been identified. Regions of the face, differences in the face structure, and other facial features are examples of facial features. The features from the camera are grabbed as part of the extraction process. Face Detection entails removing the backdrop and focusing on the target. Apart from the foreground, there are no additional items in the scene. [7]

The most important aspect of face detection is finding faces. Faces can be discovered using a variety of approaches. Even though face detection is the most crucial stage in image processing, the techniques used to accomplish it need to be updated in order to improve its performance and lower the barriers it faces. Face detection methods rely heavily on the accuracy of face detection; as a result, face detection is the most crucial step in the entire process of face detection and tracking.

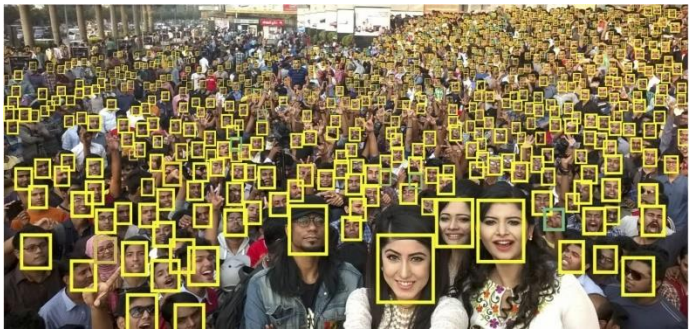
FIGURE 6: Face Detection

To extract emotions from these videos we took the above frames and each frame comprise of faces of participants. To detect the emotions behind these participants we cropped faces from these 294 frames. We used OpenCV to crop faces. OpenCV is a python library mainly used for image processing and computer vision. In this article first, we detect faces after that we crop the face from the image. Face detection is the branch of image processing that uses to detect faces. We will use a pre-trained Haar Cascade model to detect faces from the image. A haar cascade is the object detection method used to detect objects from the image. This algorithm was trained by numerous images. [8]

We use cv2.CascadeClassifier for load haarcascade file in face cascade. X, y is pixel location of faces, w, h are width and height of faces. cv2.rectangle() function used for draw rectangle over the detected object, image is input image, **(x, y),(x+w, y+h)** are locations of rectangle i.e bounding box or bbox of the cropped faces.

We have used “haarcascade_frontalface_alt2.xml”.

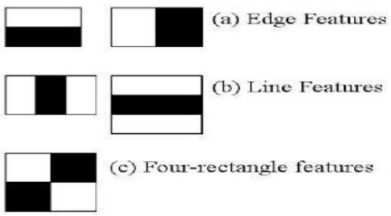

FIGURE 7: Haar Cascade Features

After cropping faces and finding bounding boxes coordinates, we have created a json file “bbox_master.json” which contain all the faces and their bounding boxes details in a json file format.

"img (1)_0.jpg": "(369, 413, 589, 633)"

Name convention tells about the frame number (1) and the face number in the frame "_0".

We got 2784 faces after cropping faces from 294 frames on 14 videos dataset. Our next task was to label emotions on these 2784 faces.

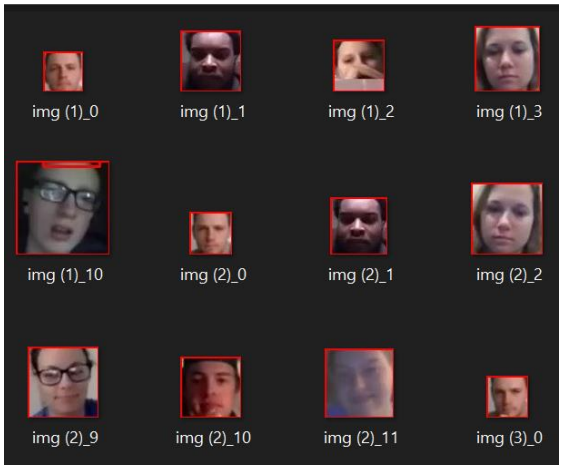

FIGURE 8: Snapshot of cropped faces used to create this dataset.

6 Survey for Emotions on Cropped Faces

To label these 2784 faces we performed a public survey to get the emotions behind these faces and labelling them.

https://docs.google.com/spreadsheets/d/1w7OrNx-WuMGUk0IdWaP79oGEjP4byCn6y1xooM5S5S0/edit?usp=sharing

We have created 139 google forms comprise of 20 faces each to get the public opinion on the emotions to make our dataset more accurate. Google Forms is a survey administration software included as part of the free, web-based Google Docs Editors suite offered by Google. Each form comprise of 20 faces and each face has a checkbox option with six emotions happy, sad, fear, angry, disgust, and surprise. Each form was filled by 4 different participants. It was a big challenge to create 139 forms comprise of 2784 faces and getting public opinion on these forms which took a total of 556 submissions.

These responses were received by different participants and to make the dataset nonbiased random forms were given to different participants. The motive behind different input from several participants was to get a variety of opinion on the emotions and then going with the majority of the participants for a better insight of the emotions behind the faces.

FIGURE 9: Snapshot of survey used to create this dataset.

Timestamp	Name		
21/11/2021 23:55:52	Naman	Neutral	Neutral
12/12/2021 23:32:04	Aman	Sad	Angry
13/12/2021 11:50:46	Harshita Malhotra	Neutral	Neutral
17/12/2021 14:23:03	Guresh Mehta	Neutral	Neutral

FIGURE 10: Snapshot of survey response used to create this dataset.

After getting all 556 responses from these 139 surveys forms, we took most of the emotion for each face to make it more accurate. We downloaded all these 139 responses as csv files.

https://drive.google.com/drive/folders/1hDIpZGLwBNNznuSib-Mg827C7rMfnYDX?usp=sharing

These sheets were sorted and uploaded as data frame using pandas library. After creating these data frames we extracted the major emotion choose by survey participants for each 2784 faces.
We extracted emotion for each cropped face and labelled it against the faces.

7 Final Ground Truth

After getting these emotions for each face we mapped it with the bbox_master.json file to map various bounding boxes of faces of each image to their corresponding emotions that we obtained from the public survey.
We finally created a nested dictionary which contains each image and each face bounding box with the emotions from survey as the value against them.

"image63":
{"(94, 163, 435, 504)": "Neutral",
"(23, 125, 1057, 1159)": "Neutral",
"(46, 116, 797, 867)": "Neutral"

Here, image63 refer to the frame number 63 and the key value here is the bounding box coordinates of the faces detected in this frame and the value belongs to the emotion behind that bounding box from the survey responses.
So, the final ground truth contains a nested dictionary which include all 294 frames and all 2784 faces detected in them with their emotions.
Now each video 21 frames have their own cropped faces bounding boxes detail and their corresponding emotions coming from the survey.

The "Final_groundtruth.json" file contains the final ground truth of the dataset which comprise of above nested dictionary.

8 Conclusion

In this dataset we can now detect the emotion and sentiment analysis behind the videos. All the videos have 21 frames and all 21 frames contain some faces. To conclude the final sentiment of the students in the lecture can be determined by taking most of the emotion in the frames and then taking most of the emotion in all 21 frames which will return the majority emotion and sentiment behind the lecture.
To get a sentiment estimate insight about the long lectures of these 14 lecture videos we can now determine by processing this dataset. This dataset can be used to train the model and to use on new lectures to determine the sentiment behind the lectures. The basic logic behind determining the sentiment is to detect the emotions of the students at different timestamps and to take the average sentiment of the students at these different timestamps to give an estimate insight about the lectures.

9 Scope of Work

Using this dataset along with FER2013 we can train our model and can apply to new lecture videos or business meeting videos to determine the sentiment of the participants in the meeting. As we know online learning platform is in trend after the pandemic and it also solved the problem of distant learning. There is a big field of sentiment analysis on real time videos whether it is online learning lectures or business meetings.
Despite the above benefits, there is still much room for improvement in this dataset and its applications. For instance, taking 100 videos and more frames with less timestamps and working with more faces and other thing is with the survey for more accurate and non-biased responses we can go with a bigger survey of at least 10 participants for one face and then taking most choices.

10 Applications

Emotion detection and sentiment analysis is one of the hottest topics in the tech industry.

Applications like **social media monitoring** which help companies about some of the most truthful points of view about products, services, and business because users offer their opinions unsolicited.

Sentiment analysis **Customer Support** reads regular human language for meaning, emotions, tone, and more, to understand customer requests, just as a human would.

Sentiment analysis can be used to gain insights from the **Customer Feedback** available and save employee hours and can also work with common mistakes like misspelled and misused words.

11 References

[1] C. Darwin and P. Prodger, *The Expression of the Emotions in Man and Animals*, Oxford University Press, Oxford, MA, USA, 1998.

[2] Y.-I. Tian, T. Kanade, and J. F. Cohn, *"Recognizing action units for facial expression analysis,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 97–115, 2001

[3] Thomas Wolf, Lysandre Debut, *Hugging Face's Transformers: State-of-the-art Natural Language Processing,* 14 July 2020

[4] Chao Ma; Chongliang Sun; Donglei Song; Xuan Li; *Hao Xu "A Deep Learning*

Approach for Online Learning Emotion Recognition", 2018

[5] Weiqing Wang, Kunliang Xu, Hongli Niu,*"Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation",* Sept 2020.

[6] Yousif Khaireddin, Zhuofa Chen ,*"The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi"* 2013

[7] I. q J. Goodfellow, D. Erhan, P. L. Carrier et al*., "Challenges in representation learning: a report on three machine learning contests,"* Neural Information Processing, Springer, Berlin, Germany, pp. 117–124, 2013.

[8] Y. Kim, S. Kang, S. Lee et al., *"Efficiently detecting outlying behavior in video-game players,*" PeerJ, vol. 3, p. e1502, 2015.

[9] John Gottfried, *Virtual conferencing and meeting systems: Resources for online connections, Feb 2015*

[10] Maliha Khan; Sudeshna Chakraborty, *"Face Detection and Recognition Using OpenCV"*, Oct 2019

[11] Li Cuimei; Qi Zhiliang *"Human face detection algorithm via Haar cascade classifier combined with three additional classifiers"*,Oct 2017

[12] Weiqing Wang, Kunliang ku, *"Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation", Sep 2020*