

L^AT_EX Final Report for Computer Vision Group 1

Murtadha Marzouq

Graduate Student

mmarzouq@charlotte.edu

Param Patel

Graduate Student

ppate211@charlotte.edu

Haochen Ye

Graduate Student

hye5@charlotte.edu

Sam Aldehayyat

Undergraduate Student

saldehay@charlotte.edu

Yuepei Yu

Undergraduate Student

yyu20@uncc.edu

Abstract

In educational settings, assessing classroom occupancy is integral for attendance keeping purposes. This project aims to develop a sophisticated computer vision system aimed at automating the attendance process by accurately detecting and recording the presence of students within images taken in a classroom by an instructor. We will utilize well-known datasets and methods for examining the images which will be introduced further along in the paper. It will contain a detailed explanation of the problem, planning the solution, implementing it and the lessons learnt throughout this process.

Keywords: Computer Vision, Object Detection, Image Processing, Student Counting, Classroom Management.

1. Introduction

Attendance is an indispensable metric for a student's academic journey, where consistent participation correlates with academic success. The process of monitoring and maintaining attendance records presents significant challenges, which mostly originates from constraints in time and resources. By harnessing the benefits of computer vision, this project addresses the problem of automating the attendance-taking process in classrooms using a tool that is at every teacher's disposal, their cell phone camera.

2. Problem Statement

Attendance enforcement and record keeping in the classroom is often overlooked because of added time and resources. Attendance is often used as a certain percentage of a student's grade to entice them into showing up to a classroom, when it should be an innate desire to attend class to further their knowledge. Secondly, classes are often over-saturated with the number of students for one subject so

keeping track of names and their attendance records is hard for a single instructor. Resorting to methods like iClicker remotes and in-person quizzes further encroaches upon valuable class time that should ideally be dedicated to teaching. The COVID-19 pandemic has also made it difficult to enforce attendance in the classroom due an overdependence on 'fall-back' virtual learning platforms like Zoom and Canvas. Keeping in mind all of these facets, we aim to simplify and modernize the experience for both the instructor and students.

3. Dataset

The first step to building an image recognition system that will help identify the number of students, we need to start by having a high-quality set of images. The dataset serves as the foundation for training and testing the image recognition system, and the quantity, quality, and diversity of the images in this dataset are all also factors that affect its performance. Constructing a high-quality dataset primarily includes data collection, data annotation, data preprocessing, and dataset partitioning for evaluation and deployment.

3.1. Data Collection

To ensure we have enough data, we've collected information from various sources. Initially, we started by acquiring a portion of publicly available data online, specifically leveraging the [Microsoft COCO\(COCO Dataset\)](#) [3]. This dataset contains a wide array of images featuring individuals from various backgrounds, orientations, ethnicities, and settings. We used this specific dataset because it is quite well-known and has numerous sections which are applicable to our specific use-case such as the person classifier. Beyond that open-source database, our validation images also incorporate our individual student experiences. For instance, in each class that we are in, we've taken a set of photographs from different angles with the consent of

067 our classmates. These photos hold significant importance
068 for the practical application of our system, as they closely
069 mirror the real-world environment (classrooms) where our
070 recognition system will be put into use. To further enrich
071 our dataset's diversity, we've collaborated with other teams
072 working on similar projects and obtained a portion of shared
073 data.
074

075 **3.2. Data Labeling**

076 To achieve clearer identification of objects within images,
077 especially the target subjects, each image in our dataset
078 needs to be carefully outlined and annotated. These labels
079 effectively specify the position and identity of each object
080 to be recognized within the images. Annotating individual
081 objects can be accomplished through manual annotation,
082 where humans identify and mark the regions of individuals
083 in the pictures, or through other tools designed for
084 object recognition and annotation. The data obtained from
085 the COCO dataset also comes with precise annotations. To
086 enhance the accuracy of image recognition, we require a
087 substantial amount of labeled data to train our image recogni-
088 tion system. The accuracy of labeling directly impacts
089 the reliability of our system, making accurate annotation the
090 foundation of our system.
091

092 **3.3. Data Preprocessing**

093 Thoroughly preparing the data is another part of preparing
094 our dataset. This step is essential to make sure that all the
095 images in our dataset have consistent quality and can be
096 effectively used for both training and testing our system.
097 During this phase, we carry out various tasks on the data,
098 including data we obtained from the COCO dataset and our
099 own captured data. In this stage, we resized images to a uni-
100 form size, removed noise, standardized the brightness and
101 contrast in the images, and aligned facial features to im-
102 prove accuracy. Preprocessing the data not only improved
103 the overall quality of our dataset but also made the train-
104 ing process smoother by providing the model with standard-
105 ized inputs. This standardization is to ensure that our image
106 recognition system can perform well under different light-
107 ing conditions, facial expressions, and variations in image
108 quality.
109

110 **3.4. Data Partitioning**

111 After going through the previous steps, the final critical
112 step in dataset construction involves dividing the dataset
113 into different subsets. These subsets typically include a
114 training set, a validation set, and a test set. The training
115 set is used to train our image recognition system, allowing
116 it to learn the patterns and features of individuals in

117 images. The validation set is used to fine-tune the model's
118 parameters and implement early stopping techniques to
119 prevent overfitting. The test set is used to evaluate the
120 overall performance of the trained model, providing a
121 fair assessment of its accuracy and robustness. Currently,
122 we have around 4,000 training data samples, over 700
123 validation data samples, and approximately 70 test data
124 samples. Since the COCO dataset provides a substantial
125 amount of suitable data, we included a portion of it in our
126 training set. In the validation set, we incorporated data
127 obtained from other collaborating teams to fine-tune the
128 model. Finally, for the test set, we used data from our daily
129 attendance photos to validate the reliability of our system
130 in a real-world attendance environment.
131

4. Methodology & Computer Vision Algorithm

133 **4.1. Methodology**

1. Research datasets to use [2]
2. Research models to use 3.
- Selected COCO Dataset and YOLOv6 [1] model
4. Train that model to recognize humans in an image [4]
5. Test the trained model to recognize the students in the class from images taken by group members

The model will be trained using a portion of the Microsoft COCO dataset. The model will be used to recognize the students in the class, apply bounding boxes around the students recognized and sum up the number of bounding boxes to get a total count of students in that image. As an addition, we noticed that we were not able to capture an entire classroom in 1 image from the front, we have also presented an option to the instructor to take numerous pictures and stitch them together into 1 image and then do a total sum of students found.

150 **4.2. Computer Vision Algorithm**

We chose to use the YOLO (You Only Look Once) [4] model. YOLO uses a single neural network to simultaneously predict bounding boxes and class probabilities, making it ideal for applications demanding instantaneous detection such as our project.

Since it's introduction in 2016, YOLO has undergone significant improvements. YOLOv2 was released in 2017, YOLOv3 in 2018, etc. these numerous iteration included refined accuracy and speed, and incorporated multiple scales for object detection [5]. YOLO showcases the hallmarks of a good model due to it's advancements that continue to propel it to the forefront of real-time object detection technology. This lengthy history also makes it an ideal choice to use for our project since we can easily find numerous resources and blog posts referring this model.[6]

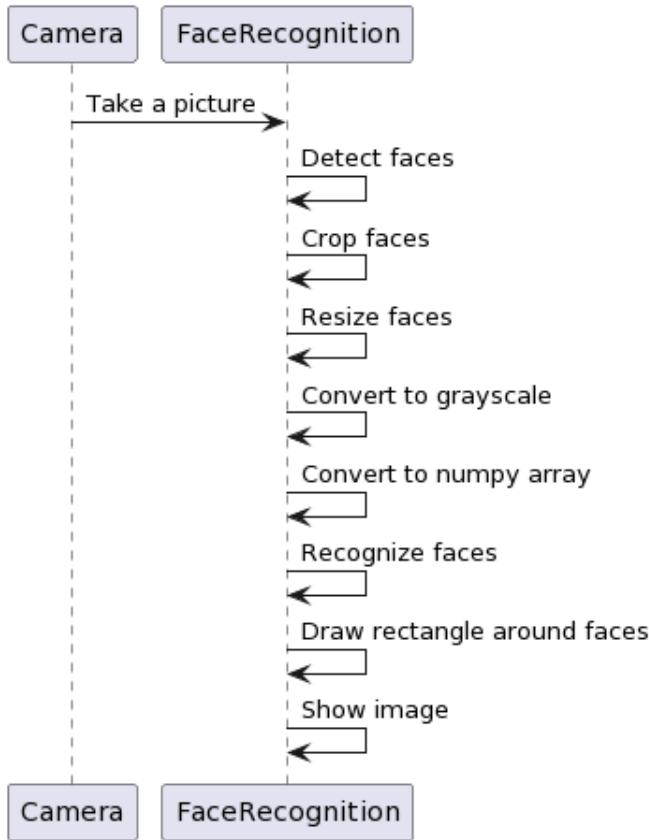


Figure 1. Flowchart of the system

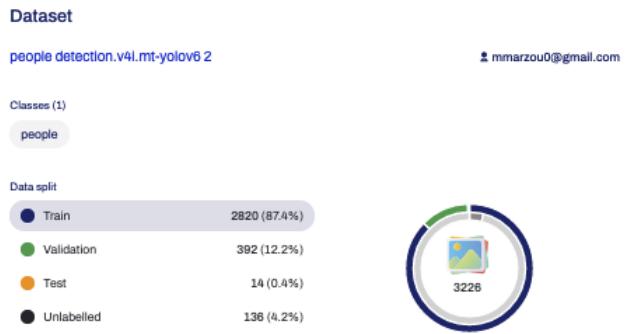


Figure 2. training data

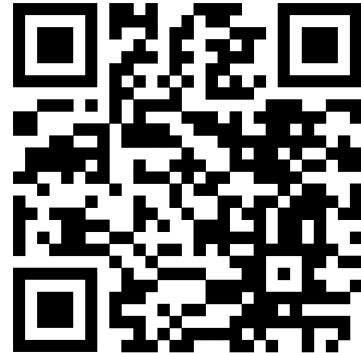


Figure 3. QR code for Demo

167

5. Results

168

The results of our efforts in this project, including the metrics of the model itself and an example of its working state, will be outlined in this section.

171

Our own model was initially trained, requiring a significant amount of time and effort. However, a switch was made to a pre-trained model, which presents the following metrics:

176

Our aim was to make our demo publicly available for everyone to test. Azure was utilized to host our solution and the below QR code was created that could be easily scanned, directing users straight to the tool.

181

Upon arriving at our site, two buttons will be visible: one for uploading a single image and another for uploading three or more images. In this workflow, the complete functionality is demonstrated by selecting more than three images. This selection involves pre-processing to stitch the images together, ensuring uniform size and zoom across all images. The stitching process was performed using the

189

OpenCV Library.



Figure 4. Pre-Processed Image

190

The figure below shows the image before it is fed into the model after performing the pre-processing and stitching.

191

192

193

The image below shows the results of the model. The model was trained for 100 epochs and the loss was 0.0001. The model that was used was pre-trained on yolov6-n vari-

194

195

196



Figure 5. Before

197

ant of the YOLO family.

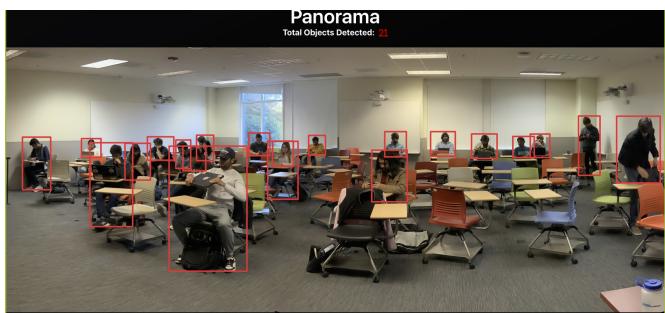


Figure 6. After

198

199 There were 21 total students found in this classroom.

200

6. Conclusion

201 In conclusion, our project embarked on an exploration of
 202 different computer models and an ambitious application
 203 of it in our own environment by applying the scientific
 204 method. Our endeavor wasn't just confined to theoretical
 205 exploration; it entailed hands-on research and data gathering,
 206 which gave us a deeper understanding of these models'
 207 nuances [6]. The meticulous collection and curation of our
 208 own dataset from our own classrooms bolstered our ability
 209 to comprehensively evaluate and fine-tune the model and
 210 offered valuable insights. [7]

211 There are numerous promising avenues for future work on
 212 our project. Integrating individual accounts for teachers
 213 to establish class sections would offer enhanced organization
 214 and tailored experiences. Furthermore, integration with
 215 Canvas to extract student enrollment counts would be a crucial
 216 step toward seamless existence with preexisting educational
 217 resources at our university. A history log could detail
 218 attendance percentages and graphical representations as a
 219 visual tool for assessment and intervention. These future
 220 enhancements hold the potential to revolutionize classroom
 221 attendance taking and facilitate a more engaged learning en-
 222 vironment.

7. Bibliography

223

1. Caya, M. V. C., Caringal, M. E. C., Manuel, K. A. C. (2021). Tongue Biometrics Extraction Based on YOLO Algorithm and CNN Inception. In 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). <https://doi.org/10.1109/hnicem54116.2021.9732037> 224
225
226
227
228
229
230
231
2. A Survey of Computer Vision-Based Human Motion Capture. <https://arxiv.org/pdf/2001.09099.pdf> 232
233
234
3. Karam, L. J., Zhu, T. (2015). Quality labeled faces in the wild (QLFW): a database for studying face recognition in real-world environments. In B. E. Rogowitz, T. N. Pappas, H. de Ridder (Eds.), IST/SPIE Electronic Imaging. SPIE.. <https://doi.org/10.1117/12.2080393> 235
236
237
238
239
240
4. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In Computer Vision – ECCV 2014 (pp. 740–755). Springer International Publishing.. https://doi.org/10.1007/978-3-319-10602-1_48 241
242
243
244
245
246
5. Chuyi Li and Lulu Li and Yifei Geng and Hongliang Jiang and Meng Cheng and Bo Zhang and Zaidan Ke and Xiaoming Xu and Xiangxiang Chu. (2023). YOLOv6 v3.0: A Full-Scale Reloading. In 2023 4th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). IEEE. <https://arxiv.org/abs/2301.05586> 247
248
249
250
251
252
253
6. Yuan, Y., Chen, X., Wang, J. (2020). Object-Contextual Representations for Semantic Segmentation. In Computer Vision – ECCV 2020 (pp. 173–190). Springer International Publishing. https://doi.org/10.1007/978-3-030-58539-6_11 254
255
256
257
258
7. Godil, A., Shackleford, W., Shneier, M., Bostelman, R., Hong, T. (2014b). Performance Metrics for Evaluating Object and Human Detection and Tracking Systems. National Institute of Standards and Technology. <https://doi.org/10.6028/nist.ir.7972> 259
260
261
262
263
8. Misbah Ahmad, Imran Ahmed, and Awais Adnan, “Overhead View Person Detection Using YOLO”, “10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), IEEE”. <https://ieeexplore.ieee.org/document/8992980> 264
265
266
267
268
269

270 **8. Team**

271 **Murtadha Marzouq**



272

273 [Resume Link](#)

274

Param Patel



275

276 [Resume Link](#)

277

Haochen Ye



278

279 [Resume Link](#)

280

Sam Aldehayyat

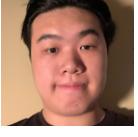


281

282 [Resume Link](#)

283

Yuepei Yu



284

285 [Resume Link](#)

286 This team had been build with the following fellowship in
287 minds:

- 288 1. Liberté, Egalité, Fraternité
- 289 2. Diversity and Inclusion
- 290 3. Teamwork
- 291 4. Respect
- 292 5. Integrity

293