LaTeX Proposal for Computer Vision

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 oversaturated 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 it's 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-knnown and has numerous sections which are applicable to our specific use-case such as the person classier. 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 our classmates. These photos hold significant importance for the practical application of our system, as they closely mirror the real-world environment (classrooms) where our recognition system will be put into use. To further enrich our dataset's diversity, we've collaborated with other teams working on similar projects and obtained a portion of shared data.

3.2. Data Labeling

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

3.3. Data Preprocessing

Thoroughly preparing the data is another part of preparing our dataset. This step is essential to make sure that all the images in our dataset have consistent quality and can be effectively used for both training and testing our system. During this phase, we carry out various tasks on the data, including data we obtained from the COCO dataset and our own captured data. In this stage, we resized images to a uniform size, removed noise, standardized the brightness and contrast in the images, and aligned facial features to improve accuracy. Preprocessing the data not only improved the overall quality of our dataset but also made the training process smoother by providing the model with standardized inputs. This standardization is to ensure that our image recognition system can perform well under different lighting conditions, facial expressions, and variations in image quality.

3.4. Data Partitioning

After going through the previous steps, the final critical step in dataset construction involves dividing the dataset into different subsets. These subsets typically include a training set, a validation set, and a test set. The training set is used to train our image recognition system, allowing it to learn the patterns and features of individuals in images. The validation set is used to fine-tune the model's parameters and implement early stopping techniques to prevent overfitting. The test set is used to evaluate the overall performance of the trained model, providing a fair assessment of its accuracy and robustness. Currently, we have around 4,000 training data samples, over 700 validation data samples, and approximately 70 test data samples. Since the COCO dataset provides a substantial amount of suitable data, we included a portion of it in our training set. In the validation set, we incorporated data obtained from other collaborating teams to fine-tune the model. Finally, for the test set, we used data from our daily attendance photos to validate the reliability of our system in a real-world attendance environment.

4. Methodology & Computer Vision Algorithm

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.

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]

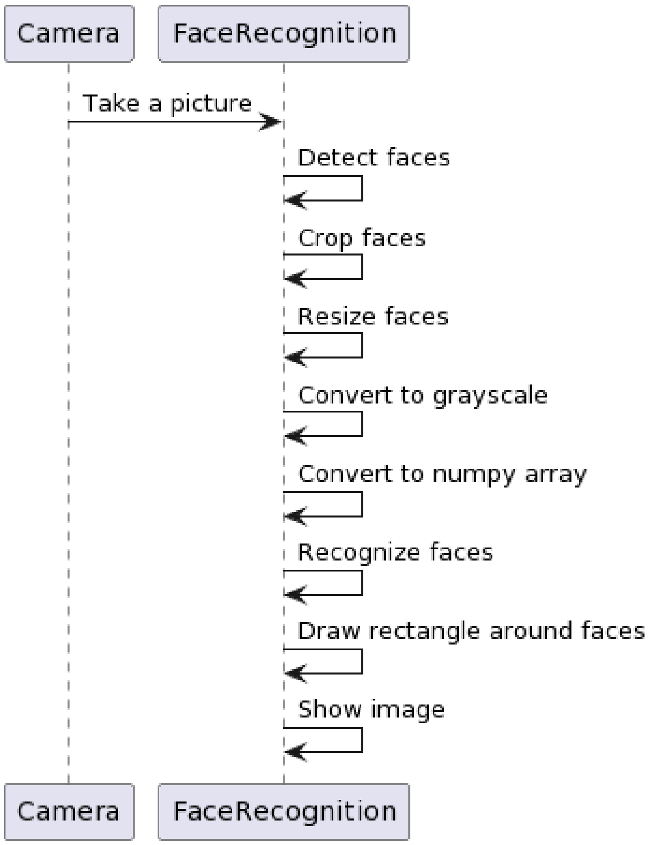


Figure 1. Flowchart of the system

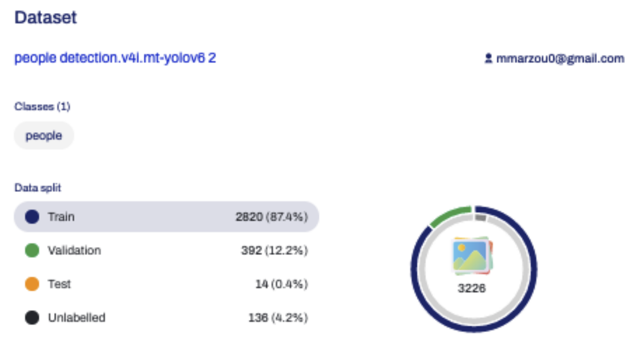


Figure 2. training data

Figure 3. QR code for Demo

5. Results

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.

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:

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.

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 OpenCV Library.

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

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 variant of the YOLO family.

There were 21 total students found in this classroom.

Figure 4. Pre-Processed Image

Figure 5. Before

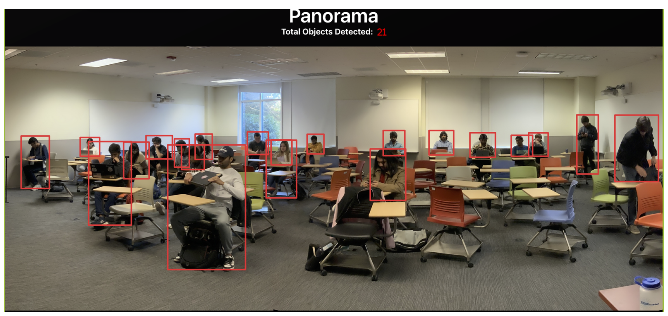


Figure 6. After

6. Conclusion

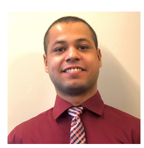
In conclusion, our project embarked on an exploration of different computer models and an ambitious application of it in our own environment by applying the scientific method. Our endeavor wasn't just confined to theoretical exploration; it entailed hands-on research and data gathering, which gave us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

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

7. Bibliography

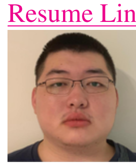
1. *A Survey of Computer Vision-Based Human Motion Capture*. https://arxiv.org/pdf/2001.09099.pdf
2. *A Survey of Computer Vision-Based Human Motion Capture*. https://arxiv.org/pdf/2001.09099.pdf
3. *A Survey of Computer Vision-Based Human Motion Capture*. https://arxiv.org/pdf/2001.09099.pdf
4. *A Survey of Computer Vision-Based Human Motion Capture*. https://arxiv.org/pdf/2001.09099.pdf
5. *A Survey of Computer Vision-Based Human Motion Capture*. https://arxiv.org/pdf/2001.09099.pdf
6. *A Survey of Computer Vision-Based Human Motion Capture*. https://arxiv.org/pdf/2001.09099.pdf
7. *A Survey of Computer Vision-Based Human Motion Capture*. https://arxiv.org/pdf/2001.09099.pdf

8. Team



Murtadha Marzouq
Resume Link

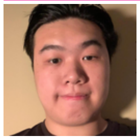
Param Patel
Resume Link



Haochen Ye
Resume Link

Sam Aldehayyat

Resume Link



Yuepei Yu

Resume Link

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

1. Liberté, Egalité, Fraternité
2. Diversity and Inclusion
3. Teamwork
4. Respect
5. Integrity

References