Applying Computer Vision to Track Library Occupancy

A Technical Report
presented to the faculty of the
School of Engineering and Applied Science
University of Virginia

by

Gabriel Silliman

May 10, 2023

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Gabriel Silliman

Technical advisor: Briana Morrison, Department of Computer Science

Applying Computer Vision to Track Library Occupancy

CS4991 Capstone Report, 2023

Gabriel Silliman
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
gjs4wu@virginia.edu

ABSTRACT

During peak hours at UVA, finding a quiet study space can be difficult without walking between the different libraries and searching each floor. By providing students with an app that shows the crowdedness of each library, students can avoid wandering until they find a quiet place to study. I propose using computer vision on live camera footage from each library floor to identify the number of people currently in a study space. This data can be collected over peak and lull periods on each library floor to generate a qualitative measurement shown to users through a mobile app. This project could prove very useful to students, especially during finals when the libraries are extremely busy. Moving forward, more study spaces beyond the libraries could be added to the app to provide students with more options.

1. INTRODUCTION

At the end of each semester, as deadlines and exams loom, students pack the libraries around the campus of the University of Virginia. Even in the late hours of the night, rows of students fill the booths and tables of all floors of the libraries. I have experienced this firsthand every time finals season rolls around, and I find myself visiting each floor of Clemons Library, one by one, hoping to snag a table from a student leaving their spot. If the library is completely packed, I am forced to walk to another library, hoping it will be a little less full. I find myself texting friends asking if anyone managed to find a table I can join.

Certain apps, such as Google Maps, provide users with live trackers, showing how crowded businesses are at any moment and predicting future busy levels. This is possible due to the mass data collection practices of Google, which has access to millions of users' precise location data. Regardless of the privacy concerns that come with this level of mass location monitoring, it is not practicable or feasible to monitor the number of students in a study space using this method. Instead, computer vision techniques can be used to identify how busy a specific area is by tracking the number of people in the camera's view.

2. RELATED WORKS

Extensive research has been done in object detection using neural networks. Convolutional neural networks are effective at identifying the existence of any objects, including humans, in still images. Scale-driven convolutional neural network (SD-CNN) models have been used to count the number of people in crowded areas (Basalamah, 2019). The model assumes that heads are the dominant, visible feature in each static frame of analysis. The SD-CNN scales image features by their 2-dimensional location in the frame to account for the perspective of the camera's view on the size of each head. Based on this scale map, object proposals are generated and classified into head or background classes, which generate a response map of each head in the image. This data can be used to count a locate each person and analyze the human density in the image. This method is especially effective for analyzing crowds or crowded spaces where multiple people overlap within the frame. Traditional methods that generate edge maps or bounding boxes have difficulty differentiating between people in these layered scenes.

Many other techniques exist for detecting the occupancy of a space without video. CO_2 levels, the number of devices connected to a Wi-Fi network, or motion detection can be used to estimate the number of people in an area (Zhao, et. al, 2022). These methods are all less computationally expensive than object detection on live video and do not collect as much identifying information. However, these methods are estimates and are not as accurate. CO_2 levels depend on the HVAC system's performance in a room. In a library setting, devices in different rooms and floors connect to the same Wi-Fi router. Motion detection is useful for detecting the presence or absence of people, but not for estimating the number of people in a space. Therefore, to accurately calculate the occupancy of an individual room, object detection using neural networks on a video feed is the most effective method.

3. PROJECT DESIGN

The project needs to consider how occupancy data can be collected, transmitted, stored, and accessed by users.

3.1 System Overview

The proposed library occupancy detection system uses the YOLOv7 detection algorithm (Wang, et. al., 2022) to detect the number of people in study spaces in the library. The system uses a video feed from a camera to capture the study spaces, which is then processed by a video processing component that runs the detection algorithm on frames of the video. The occupancy data is then sent to a cloud backend environment for storage and analysis and is displayed to users through a mobile app, which communicates with that cloud backend.

3.2 Requirements

The system must be able to detect the number of people in study spaces accurately and with minimal delay. However, the system does not need to provide occupancy data to the exact second. Therefore, the system can save resources by only running the detection algorithm once a minute. It still must be able to process high-resolution video streams and provide reliable occupancy data to the cloud backend environment. The system must also be scalable, allowing for multiple cameras to be added in new study spaces as needed, and must be easy to use and operate.

3.3 System Components

The following are the three main components of the library occupancy system.

3.3.1 Video Processing Unit

Each video processing unit (VPU) is responsible for capturing the video feed from the camera and processing it using the YOLOv7 algorithm. The VPU includes a camera, video capture device, and small-scale computer such as a Raspberry Pi. Once a minute, the VPU captures video for five seconds and then runs the detection analysis on ten frames per second. The number of people in the room is calculated as the average of all analyzed frames in the period. This lessens the computational load on the Raspberry Pi while still getting accurate results. Once the occupancy data is calculated, the Raspberry Pi sends it to the cloud backend database. After the data is sent, the captured video is deleted. The network bandwidth, cloud storage, and local storage requirements are much lower by only transmitting the occupancy data and not storing the recorded video. Immediately deleting the footage after it is analyzed also lowers the security risk that recordings could be accessed by a malicious or unauthorized individual.

3.3.2 Cloud Backend Environment

The cloud backend environment is responsible for storing and analyzing the occupancy data collected by the VPUs. It includes a cloud-based database for storing occupancy data, as well as a cloud server that provides an API that allows the mobile app to query occupancy data through a REST API call when the user opens the app. I propose that the cloud environment be run on Amazon Web Services, using a DynamoDB NoSQL database to store occupancy data and AWS Lambda functions to process the occupancy data to provide to the mobile app.

3.3.3 Mobile App

The mobile app provides users with real-time occupancy data for the study spaces in the library. As shown in Figure 1, it includes a user-friendly interface that displays a qualitative occupancy measurement for each study space. The app also allows users to view predicted future data calculated based on historical occupancy trends for each study space, as seen in Figure 2.

Rather than showing users an exact count of the number of people in a study space, the app displays one of five qualitative measurements: Empty, Not Busy, Busy, Very Busy, or Packed. These values are calculated

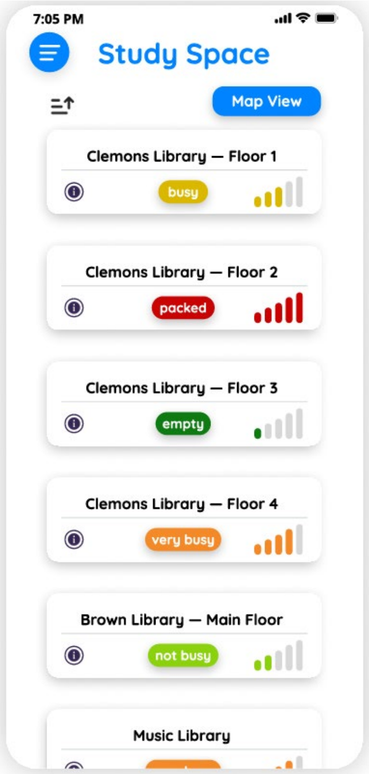

Figure 1: App interface main page

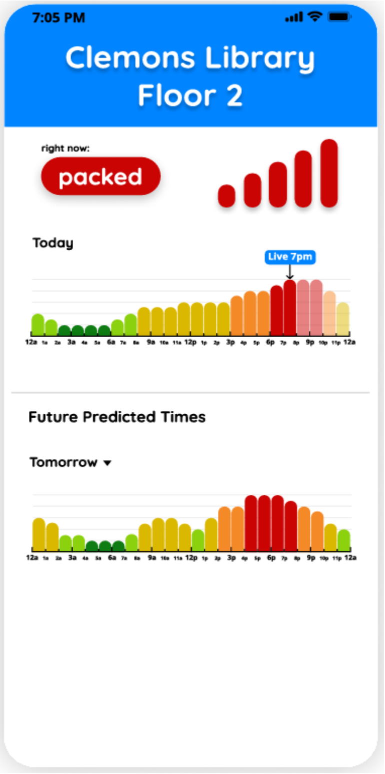

Figure 2: Current and future hourly occupancy levels

Figure 3: App map view

based on the number of people in the space and the total capacity. The mobile app also allows users to view the study spaces as either a list or a map of campus, as shown in Figure 3.

The mobile app is implemented using React Native, which allows the app to be deployed on both iOS and Android platforms. The app consumes the occupancy data from the cloud backend component through a secured REST API call.

3.4 Challenges

One of the main challenges of this system is achieving accurate and reliable object detection in real-time. This requires careful tuning and optimizing the YOLOv7 algorithm to ensure that it can run efficiently on a small-scale CPU while maintaining high accuracy. Another challenge is ensuring the scalability and reliability of the system. As the number of VPUs increases to cover more areas and the number of users of the app increase, the backend cloud environment will need to be scaled appropriately to handle higher traffic. Finally, the system must be designed to be secure and prevent unauthorized access to user data.

3.5 Solution

The YOLOv7 algorithm was chosen for this project due to its efficiency and adaptability, allowing it to be run on small-scale devices. Additionally, due to the lack of a need for second-level precision, the algorithm only runs once a minute to reduce the load on the VPUs. AWS DynamoDB offers seamless scalability, ensuring that with increased users and VPUs, the cloud backend will not be overwhelmed. Finally, from a security standpoint, the choice to process video footage locally and immediately delete it after analysis guarantees that malicious individuals cannot access footage from study spaces. Additionally, queries from the app for occupancy data do not return precise counts of people, but rather qualitative measurements of the occupancy, ensuring more anonymity in the system.

4. RESULTS

While this system is not currently implemented or in use, we can predict its outcome. From personal experience and from speaking to other UVA students, this app would likely be a helpful resource to the UVA community. Study spaces on campus fill up quickly during the week and finding a quiet place to work can be challenging. This app would be especially helpful during finals when the libraries are packed 24/7. The newly designed system provides an easy way to find a place to study without wasting precious time and has built-in security and scalability to ensure its reliability.

5. CONCLUSION

This project introduced a solution to allow students to view live occupancy levels in study spaces on campus through a mobile app. The project provides valuable information for students searching for a quiet place to study by using object-detection algorithms, cloud databases, and a cross-platform mobile application. The system also focuses on prioritizing security and privacy, to protect the personal data of individuals in the study spaces being monitored.

6. FUTURE WORK

In the future, the library occupancy system could be expanded to cover more study spaces on campus. Additional features could be added to the app, such as the ability to reserve private study rooms. There is also an opportunity to collect feedback on the performance of the system through the app by asking users that are currently in a study space to assess the occupancy level. This can be used to evaluate whether the system is performing well or if tweaks to the calculation process should be made.

REFERENCES

[1] Basalamah, S., Khan, S. D., & Ullah, H. (2019). Scale driven convolutional neural network model for people counting and localization in crowd scenes. *IEEE Access, 7*, 71576–71584. https://doi.org/10.1109/ACCESS.2019.2918650

[2] Wang, C.-Y., Alexey Bochkovskiy, & Hong-Yuan Mark Liao. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *ArXiv*, *abs/2207.02696*. https://doi.org/10.48550/arXiv.2207.02696

[3] Zhao, L., Li, Y., Liang, R., & Wang, P. (2022). A State of Art Review on Methodologies of Occupancy Estimating in Buildings from 2011 to 2021. *Electronics*, *11*(19), 3173. https://doi.org/10.3390/electronics11193173