Video-based Inference for Bar line Estimation (VIBE)

Noah Weissman, Darius Zucker, Eden Growney
University of Michigan
500 S State St, Ann Arbor, MI 48109
wnoah@umich.edu, dzucker@umich.edu, egrowney@umich.edu

1. Introduction

The growth of data in the digital age has driven significant advancements in real-time data processing across numerous fields. Real-time data—information processed as it is created—has become a cornerstone for applications ranging from healthcare monitoring to transportation systems. In particular, video data has emerged as a rich source of information, presenting unique challenges. Video streams are dynamic, high-dimensional, and unstructured, requiring advanced techniques to extract insights effectively.

1.1. Background

In computer vision, real-time object detection focuses on identifying and localizing objects in video frames with minimal latency, enhancing public safety, efficiency, and smart urban environments. Traditional methods struggle with dynamic video data, but deep learning, particularly Convolutional Neural Networks (CNNs), has improved accuracy by automating feature extraction. These advancements allow systems to adapt to complex, changing environments.

1.2. Motivation

This project applies real-time data processing and computer vision to estimate wait times at Rick's American Cafe, a popular campus bar. Fluctuating line lengths often cause customer frustration and inefficient service management. A video-based system providing accurate, real-time wait predictions offers a practical solution.

The project benefits both patrons and staff. Patrons gain clearer expectations and reduced uncertainty, while staff can optimize resource allocation during peak hours. Automating this process demonstrates how computer vision can enhance customer satisfaction and streamline operations. This case study highlights the broader potential of real-time video analysis to transform service industry practices and improve everyday experiences.

2. Related Work

2.1. Multi Object Tracking

This CNN-based multi-object tracking (MOT) framework is relevant to our approach as it addresses real-time tracking of individuals across frames, even with movement and occlusion. While originally applied to surveillance and autonomous driving, we adapted it to predict bar queue wait times. By using spatial-temporal attention and ROI maps, our method tracks patrons, even when occluded, providing robust real-time insights into queue movement and wait times.[1].

2.2. Line Estimation in Super Market

A similar wait-time estimation approach is used in supermarkets, factoring in the number of people in line, cart items, and employee efficiency. However, this approach has limitations: it requires costly infrastructure, and accuracy can be affected by occlusions, poor lighting, or sudden customer surges. It also assumes uniform checkout processes, overlooking variability in employee speed or customer behavior. [2].

3. Method

3.1. Computer Vision Model

For our method, we fine-tuned the YOLOv8 architecture with the Roboflow API, and used the Roboflow supervision framework for real-time video analysis to detect and track individuals in a bar queue, enabling accurate wait time predictions. YOLOv8 was chosen for its optimization in real-time object detection, balancing speed and accuracy.

3.2. Data Collection and Annotation

To fine-tune the model, we hand-annotated images of the Rick's line by applying bounding boxes to people, resulting in a robust, far more accurate model. Before fine-tuning, due to the inherent difficulty of the task, the YOLOv8 model predicted an extremely small subset of the people. After fine-tuning, we greatly increased performance, reaching a final precision of 49% and a recall of 32%. Our model's performance in predicting queue sizes may seem poor due to the noisy nature of the task, but it works well overall. By sampling multiple images and applying Gaussian blurring to the time series, we effectively minimized noise and achieved reliable results [3].

3.3. Regions of Interest (ROI) and Tracking

To track crowd dynamics, we divided each frame into two key regions of interest (ROIs): the entry zone (the region near the bar's door) and the waiting zone (the area where people wait in line).

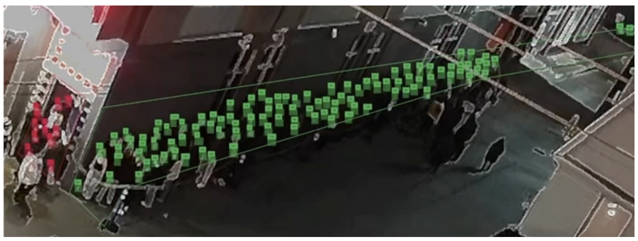

Figure 1. Entry Zone (Red) Waiting Zone (Green)

3.4. Time Series Analysis

We plotted the number of people detected in each ROI over time. Our algorithm aims to find sharp decreases in the waiting zone that are associated with spikes in the entry zone. These "entry events" are marked, and the magnitude of the decrease in the waiting zone is used to estimate the number of people let into the bar.

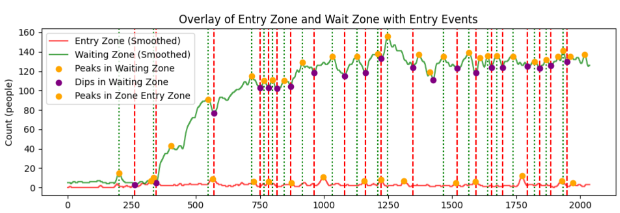

Figure 2. Entry events are defined by the start frames (dashed green lines) and end frames (dashed red lines), overlaid on smoothed counts of people in the Entry Zone (red) and Waiting Zone (green). Peaks and dips in the Waiting Zone and Entry Zone are marked with orange and purple dots, respectively.

3.5. Sliding Window Prediction

A sliding window technique is used to calculate the rate of entry. For our sliding window size of five, up to five entry events are used to calculate the rate of entry at any given moment in time.

$$\text{Rate of Entry} = \frac{\text{Total People Let In}}{\text{Time Elapsed}}$$

This method allows us to predict a running rate of entry by evaluating how frequently people were entering the bar over short intervals. The sliding window approach provided a more accurate and real-time estimation of wait times.

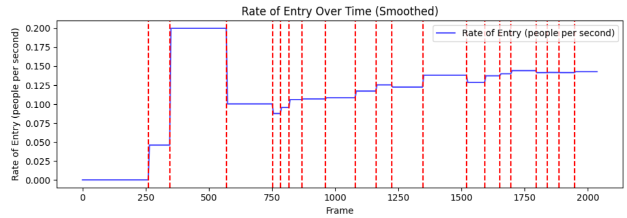

Figure 3. Rate of Entry over time (sliding window size = 5)

3.6. Estimating Wait Time

Finally, at each frame, the running wait time is calculated by the formula below.

$$\text{Wait Time} = \frac{\text{Number of People in Line}}{\text{Rate of Entry}}$$

4. Experiments

Due to the inherent difficulty of accurately watching video recordings and following individuals through the line to calculate wait time (its not just hard for the model, its hard for us too!) we decided to benchmark the model on the percentage of entry events it successfully identified in 1 hour of video. While it successfully identified all 19 entry events, it did predict a false positive at the start of the video when people walked backwards in the waiting zone. This leads us to a precision of 95%, and a recall of 100%. Due to the nature of the problem and the small test, these results should be taken with a grain of salt. More thorough testing including longer footage, different weather conditions, and at-capacity nights must be conducted to prove the robustness of our model and the subsequent time-series analysis.

5. Conclusions

Our YOLOv8-based approach for estimating bar wait times offers an efficient, real-time solution that handles occlusions and tracks individuals across frames with low computational cost. However, accuracy may decrease in certain situations, such as when Rick's gets full, the line extends beyond the camera's field of view, or the camera loses focus. Additionally, the model needs to account for people lining up before Rick's opens.

Future work could focus on improving performance under these challenging conditions, optimizing for crowded settings, and integrating with queue management systems. Incorporating time-related features and external data, such as local events or weather, may further enhance prediction accuracy.

References

[1] Q. Chu, W. Ouyang, H. Li, X. Wang, B. Liu, and N. Yu, "Online Multi-Object Tracking Using CNN-Based Single Object Tracker with Spatial-Temporal Attention Mechanism," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. Available: https://openaccess.thecvf.com/content_ICCV_2017/papers/Chu_Online_Multi-Object_Tracking_ICCV_2017_paper.pdf 1

[2] Q. Wang and M. Zhang, "Queue Time Estimation in Checkout Counters Using Computer Vision and Deep Neural Network," *Journal of Physics: Conference Series*, vol. 1964, no. 6, 2021. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1964/6/062108/pdf 1

[3] Y. Li, W. Guo, X. Liu, Y. Yang, and W. Zuo, "HP-YOLOv8: High-Precision Small Object Detection Algorithm for Remote Sensing Images," *Sensors*, vol. 24, no. 15, p. 4858, 2024. Available: https://www.mdpi.com/1424-8220/24/15/4858 2