Introduction

New Yorkers love the subway, but they don't love the bus system. While enormous, the NYC bus system is inconsistent, unreliable, and according to Riders Alliance, the slowest bus service in America[1]. However, bus riders should have cause for optimism, as congestion pricing has resulted in a 4% improvement in bus speeds across the MTA, and a corresponding increase in ridership[2]. Now, more than ever, New Yorkers are asking the age-old question: "where is my bus?"

Many of the buses in the NYC fleet have built-in trackers which publish their locations. However, these trackers are frequently broken or disabled, leading to a lack of rider trust. After a particularly frustrating bus experience this weekend, we thought: what if we could use computer vision to identify buses and show riders a camera image of where they are?

Data Sources

Our main data source for this project is the system of traffic cameras run by the NYC Department of Transportation (DOT). Images from each camera are published at 2 second intervals (example image given below). These cameras are located at major intersections, not every intersection, so this will guide which bus routes we are able to effectively monitor. For inference, we plan to scrape data from the camera system live. For training, we may need to build our own persistent scraper or use an archive of surveillance footage like that from NYC Mesh[3].

Another source of data is the bus transponder data provided by the MTA[4]. Though there can be gaps in the data, it can be used to evaluate the performance of our vision-based approach for bus tracking.

---

[1] Riders Alliance. (n.d.). *Better Buses*. Riders Alliance. https://www.ridersalliance.org/better-buses
[2] Duggan, K. (2025, March 20). *Congestion Pricing's Big Winner? Bus Riders*. Streetsblog New York City. https://nyc.streetsblog.org/2025/03/20/congestion-pricings-big-winner-bus-riders
[3] https://www.nycmesh.net/blog/nyc-surveillance-archive/
[4] https://bustime.mta.info/wiki/Developers/Index;jsessionid=3BEEFC1A449E9DDB6B82FDFCDDD256C2

Modelling Tasks

Our task can be broken down into a few subproblems:

1. Can we identify the presence or absence of a city bus in the frame, being careful not to include other kinds of buses or vehicles?
2. Can we infer the bus line from its signage? The images are pretty low-resolution (example image given below) and many bus routes overlap. We will likely need to find distinguishing characteristics based on the contextual information to identify the buses. For example, a bus might consistently turn in one direction at a specific intersection.

Models

Object recognition within images is a popular task and many foundational models exist that we can build off of. Models are generally split into two-stage and one-stage detectors. An example of a one-stage detector model is the single shot detector (SSD). The SSD uses an underlying neural network to recognize key features in the image, bounds the features in anchor boxes, and scores them[5]. A two-stage detector, like R-CNN, will run through the image twice to further refine where the key features are located[6].

[5] GeeksforGeeks. (2024, July 11). *How single-shot detector (SSD) works?*. GeeksforGeeks. https://www.geeksforgeeks.org/how-single-shot-detector-ssd-works/

[6] Kundu, R. (2023, January 27). *YOLO Algorithm for Object Detection Explained [+Examples]*. V7 Labs. https://www.v7labs.com/blog/yolo-object-detection