Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Estimation of left behind subway passengers through archived data and video image processing

Charalampos Sipetas⁎, Andronikos Keklikoglou, Eric J. Gonzales

Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, MA 01003, United States

ARTICLE INFO

Keywords:
Public transit
Crowding analysis
Left-behind passengers
Object detection tools
Logistic regression models

ABSTRACT

Crowding is one of the most common problems for public transportation systems worldwide, and extreme crowding can lead to passengers being left behind when they are unable to board the first arriving bus or train. This paper combines existing data sources with an emerging technology for object detection to estimate the number of passengers that are left behind on subway platforms. The methodology proposed in this study has been developed and applied to the subway in Boston, Massachusetts. Trains are not currently equipped with automated passenger counters, and farecard data is only collected on entry to the system. An analysis of crowding from inferred origin–destination data was used to identify stations with high likelihood of passengers being left behind during peak hours. Results from North Station during afternoon peak hours are presented here. Image processing and object detection software was used to count the number of passengers that were left behind on station platforms from surveillance video feeds. Automatically counted passengers and train operations data were used to develop logistic regression models that were calibrated to manual counts of left behind passengers on a typical weekday with normal operating conditions. The models were validated against manual counts of left behind passengers on a separate day with normal operations. The results show that by fusing passenger counts from video with train operations data, the number of passengers left behind during a day’s rush period can be estimated within 10% of their actual number.

1. Introduction

Public transportation serves an important role in moving large numbers of commuters, especially in large cities. Transit performance is an important determinant of ridership, and transit services that offer short and reliable waiting times for commuters offer a competitive alternative to driving, which contributes to reduced congestion and improved quality of life. Crowding is a major challenge for public transit systems all over the world, because it increases waiting times and travel times and decreases operating speeds, reliability, and passenger comfort (Tirachini et al., 2013). Studies show that crowding in public transit increases anxiety, stress, and feelings of invasion of privacy for passengers (Lundberg, 1976). The COVID-19 pandemic has also highlighted the public health risks associated with passenger crowding in transit vehicles. Although transit ridership dropped precipitously during the pandemic in cities around the world, concerns about crowding on transit continue as economies re-open, commuters return to work, and agencies plan for the future.

When overcrowded, commuters may not be able to board on the first train or bus that arrives. These commuters are left behind the vehicle that wished to board, and their number is directly related to various basic performance measures of public transportation.
systems related to the user's experience, including service quality and reliability.

The Transit Capacity and Quality of Service Manual (TCQSM) establishes some guidelines for measuring the quality of service to track passenger related metrics (Kittelson et al., 2013). It states that crowding affects several aspects of transit availability and all the elements of comfort and convenience related to the quality of service. The indicators of availability as presented on the TCQSM are frequency, service span, and access. Crowding causes vehicles to operate more slowly or passengers to be left behind at stations and stops, resulting in an effective reduction in service frequency for users. This limits the availability of transit service for users compared to the same uncrowded system. Access is also affected by crowding if users are unable to enter a station or board a vehicle due to the conditions (Kittelson et al., 2013). The effects of crowding on public transport using data from a stated preference survey are investigated by Batarce et al. (2016).

The indicators of comfort and convenience as defined by TCQSM are passenger load, reliability, and travel time. More crowded vehicles increase the likelihood of passengers having to stand or squeeze with other passengers, which deteriorates the quality of the passenger experience. Evidence has shown that the users perceive waiting times and travel times to be longer in crowded conditions than in uncrowded conditions due to this discomfort (Fan et al., 2016). Likewise, crowding contributes to diminished reliability regarding on-time performance and consistent headways, because boarding and alighting are delayed when there are many people in vehicles and at stations (Carrion and Levinson, 2013). Another consequence for reliability is that left-behind passengers wait on the platform for at least one extra headway if they are unable to board the first vehicle that arrives, which causes travel delays. Finally, actual travel times are also increased by crowding due to longer dwell times and inconsistent headways.

Most of the literature on crowding has focused on passenger discomfort. Several demand models include crowding as a parameter, and it has been shown that waiting and in-vehicle travel time savings are inversely proportional to the number of people in the platform or vehicles (Douglas and Karpouzis, 2005). This is the basis for estimating the crowding externality or crowding cost (Kraus, 1991). Many studies have investigated the value of crowding from the perspective of the user, in terms of value of time and willingness to pay an extra fee to avoid crowding (Li and Hensher, 2011; Haywood and Koning, 2015; Haywood et al., 2017; Hörcher, 1991). Many studies have investigated the value of crowding from the perspective of the user, in terms of value of time and willingness to pay an extra fee to avoid crowding (Li and Hensher, 2011; Haywood and Koning, 2015; Haywood et al., 2017; Hörcher, 1991). Furthermore, various studies have aimed to determine the effect that crowding has on passengers' travel decisions and path choice (Raveau et al., 2014). For instance, research that has been completed in Seoul, South Korea, suggests that crowding affects the path choice in networks that are large and connected enough to offer multiple path choices to users between origin-destination pairs (Kim et al., 2015).

To date, most studies that infer travel patterns of transit system users are developed using farecard data from transit systems that require passengers to tap in and tap out for zone-based fare collection (Ma et al., 2013). Such systems give exact information about n-destination pairs (Kim et al., 2015). The actual travel times are also increased by crowding due to longer dwell times and inconsistent headways.

Most of the literature on crowding has focused on passenger discomfort. Several demand models include crowding as a parameter, and it has been shown that waiting and in-vehicle travel time savings are inversely proportional to the number of people in the platform or vehicles (Douglas and Karpouzis, 2005). This is the basis for estimating the crowding externality or crowding cost (Kraus, 1991). Many studies have investigated the value of crowding from the perspective of the user, in terms of value of time and willingness to pay an extra fee to avoid crowding (Li and Hensher, 2011; Haywood and Koning, 2015; Haywood et al., 2017; Hörcher, 1991). Moreover, various studies have aimed to determine the effect that crowding has on passengers' travel decisions and path choice (Raveau et al., 2014). For instance, research that has been completed in Seoul, South Korea, suggests that crowding affects the path choice in networks that are large and connected enough to offer multiple path choices to users between origin-destination pairs (Kim et al., 2015).

To date, most studies that infer travel patterns of transit system users are developed using farecard data from transit systems that require passengers to tap in and tap out for zone-based fare collection (Ma et al., 2013). Such systems give exact information about arrivals, departures, and travel times of passengers in the system. Pelletier et al. (2011) presents the various uses of smartcard technology. Most existing studies to detect left-behind passengers have been conducted with data from the London Underground, which offers both tap-in and tap-out information (Zhu et al., 2018). A recent study on estimating left-behind passengers uses Automatic Fare Collection (AFC) and Automatic Vehicle Location (AVL) systems in order to define a measure of cumulative transit vehicle capacity shortage (Miller et al., 2018). This measure is proven to be correlated with the number of left behind passengers. Manual counts from video feeds were used to calibrate the model. The results indicate that the model performs better in very crowded conditions. This study can only be implemented with tap-in only systems. A detailed table with comparison between existing methods for estimating left-behind passengers is included in Ma et al. (2019).

Given the importance of crowding in public transportation systems on passengers and agencies, there is a need for methods to measure the numbers of people in crowded transit systems. This study focuses on the specific problem of measuring the number of passengers that are left behind on subway platforms when trains are too crowded to board. Currently, systems like the Massachusetts Bay Transportation Authority (MBTA) in Boston, Massachusetts, rely on manual counts to measure left-behind passengers, but this is a costly process that is not undertaken on a regular basis. An automated method for estimating the number of left-behind passengers will allow public transportation agencies to monitor overcrowding in a systematic way. This will not only allow agencies to identify the locations, times, and magnitude of the problem but also to plan appropriate service improvements where passenger experience is most impacted.

The methodology presented here addresses the challenging case of a system with flat fares in which passengers' farecards are only used upon station entry (e.g., Boston, New York, Chicago), thus their flows inside the system can only be inferred through modeling based on assumptions. Focused data collection at MBTA stations was performed to investigate the use of emerging technologies for estimating the number of left-behind passengers by processing video feeds from surveillance cameras in transit stations. Models that account for both transit operation and demand are developed. These models are evaluated and used to estimate the distribution of passengers' actual waiting times in Boston's transit system. This is important for tracking the percent of users waiting less than a scheduled headway to board a train, which is the relevant service reliability metric used by the MBTA.

The paper is organized as follows. First, a review of digital image processing technologies for counting people in videos is presented. Second, the general framework of the proposed methodology is explained. Third, a method for modeling crowding at transit stations by making use of available data sources is described in order to identify suitable locations to test the proposed passenger counting method. Fourth, the types of data that need to be manually collected to train and validate the proposed model are presented. Fifth, the application of the digital image processing techniques to detect left-behind passengers are described. Sixth, the proposed data collection and modeling method is developed and validated to estimate the number of passengers that are left behind when trains are too crowded to board. Finally, a discussion of results and conclusions reveals the potential of this technique to be used to automatically quantify the numbers of left-behind passengers in transit systems where farecard data is only collected upon station entry.
2. Digital image processing for object detection

There are a number of technologies that can be used to observe, count, and track pedestrians and pedestrian movements in an area. Digital image processing for object detection is an appealing approach for transit systems because surveillance videos are already being recorded in transit stations for safety and security purposes. The video feed records passenger positions and movements in the same way that a person would observe them, as opposed to infrared or wireless signal detectors that merely detect the movement of a person past a point or their proximity to a detector. The detection of objects in surveillance videos is an invaluable tool for passenger counting and has numerous applications. For example, object detection can be used for passenger counting or tracking, recognizing crowding, and hazardous object recognition. In a relevant application, Velastin et al. (2006) uses image processing techniques to detect potentially dangerous situations in railway systems. Computer vision is the duplicate of human vision aiming to electronically perceive, understand and store information extracted from one or more images (Sonka et al., 2014).

There are various techniques to use computers to process an image for object detection by extracting useful information. Recent methods use feature-based techniques rather than segmentation of a moving foreground from a static background, which was used in the past. Then, the detected features are extracted and classified, typically using either boosted classifiers or Support Vector Machine (SVM) methods (Viola, 1993; Cheng et al., 2015). SVM is one of the most popular methods used in object detection algorithms and especially for passenger counting, because it offers a method to estimate a hyperplane that splits feature vectors extracted from pedestrians and other samples (Cheng et al., 2015), differentiating pedestrians from other unwanted features. Boosting uses a sequence of algorithms to weight weak classifiers and combine them to form a strong hypothesis when training the algorithm to attain accurate detection (Zhou, 2012). Current methods for object detection take a classifier for an object and evaluate it at several locations and scales in a test image, which is time-consuming and creates numerous computational instabilities at large scales (Deng et al., 2010).

The most recent methods, such as Region Based Convolutional Neural Network (R-CNN), use another method to decrease the region over which the classifier runs and includes the SVM. First, category-independent regions are proposed to generate potential bounding boxes. Second, the classifier runs and extracts a fixed-length feature vector for each of the proposed regions. Finally, the bounding boxes are refined by the elimination of duplicate detections and resoring the boxes based on other objects on the scene using SVMs (Girshick et al., 2014). The bounding box is a rectangular box located around the objects in order to represent their detection (Coniglio et al., 2017; Lézoray and Grady, 2012). The resulting object detection datasets are images with tags used to classify different categories (Deng et al., 2009; Everingham et al., 2010).

An open-source software tool called You Only Look Once (YOLO) uses a different method than the above-mentioned techniques for object detection. It generates a single regression problem to estimate bounding box coordinates and class probabilities simultaneously by using a single convolutional network that predicts multiple bounding boxes and class probabilities for these boxes (Redmon, 2016; Redmon et al., 2016). Another advantage of YOLO is that, unlike other techniques such as SVMs, it sees the entire image globally instead of sections of the image. This feature enables YOLO to implicitly transform contextual information to the code about classes and their appearance and at the same time makes YOLO more accurate, making fewer than half the number of errors compared to Fast R-CNN (Redmon et al., 2016). YOLO uses parameters for object detection that are acquired from a training dataset. YOLO can learn and detect generalizable representations of objects, outperforming other detection methods, including R-CNN. The ability to train YOLO on images has the potential to directly optimize the detection performance and increase the bounding box probabilities (Redmon et al., 2016).

The calibration of parameters for object detection using an algorithm like YOLO requires training datasets with a large number of tagged images. Although a custom training set that is specific to the context of application (e.g., MBTA transit stations) would be desirable for achieving the most accurate object detection outcomes, it is very costly to create a large tagged training set from scratch. The Common Objects in Context (COCO) dataset is a large-scale object detection, segmentation, and captioning dataset that is freely available to provide default parameter values for YOLO. The COCO dataset is not specific to passengers or transit stations, but it is a general dataset that includes 328,000 images, 2.5 million tagged objects and 91 object types, including “person” (Lin et al., 2014). Nevertheless, the tool is effective for identifying individual people in camera feeds, and the use of general training data allows the same tool to be applied in other contexts without requiring additional training data.

3. Methodology

The proposed methodology aims to estimate the number of left behind passengers at a transit station when trains are too crowded to board. Fig. 1 presents a flowchart of the data and methods used in this study in order to provide a roadmap for the analysis described in this paper. The methods rely heavily on two data sources that are automatically collected and recorded (shown in blue): train tracking records that indicate train locations over time, and surveillance video feeds. Additional archived data on inferred travel patterns from farecard records is used only to identify the most crowded parts of the system (shown in purple), and manual counts are used to estimate and validate models (shown in red). For model implementation, the proposed models require only the automatically collected input data.

3.1. Identification of study locations and times

The first step of the analysis presented in this paper is to identify the stations and times of day when crowding is most likely to cause passengers to be left behind on the platform. This analysis is used only for determining where to collect data to demonstrate the implementation of the proposed model. This step could be skipped for cases in which the locations for implementation are already...
The identification of study sites involves a crowding analysis that makes use of two data sources: train tracking records, which denote the locations of trains over time; and Origin–Destination-Transfer (ODX) passenger flows, which are inferred from passenger farecard data. Peaks in train occupancy and numbers of boarding passengers show where and when passengers are most likely to be left behind, as described in Section 4.1. Then, Section 4.2 describes an analysis of surveillance camera views to determine which stations have unobstructed platform views and station geometry that allows the automated video analysis techniques to be used to count passengers.

3.2. Automated dwell time estimation

Train tracking data, which includes the time each train enters a track circuit, is automatically recorded into the MBTA Research Database. By comparing this data against manual observations of the times that train doors open and close in the station, a linear regression model is estimated to predict dwell time from the train tracking records, as described in Section 5.1. This model is used to obtain automated dwell time estimates as inputs to the model of left behind passengers.

3.3. Automated passenger counts from video

Automated counts of the number of passengers on each station platform are obtained using YOLO, an automated image detection algorithm. The parameters of the algorithm are associated with the freely-available COCO training dataset, as described in Section 2. The threshold for object identification is calibrated, as described in Section 6.1, by applying the algorithm to the surveillance video feed and comparing with manual counts of the passengers remaining on the platform after the doors have closed (Section 5.2) and the passengers entering and exiting the platform (Section 5.3). With the parameter values and calibrated threshold, YOLO produces estimates of the number of passengers on the platform as a time series. The number of passengers that remain on the platform after the doors close is a raw automated passenger count, as shown in Section 6.2. These raw counts are not very accurate as a direct measure (Section 6.3), but they provide a useful input for modeling the number of left behind passengers.

3.4. Model estimation and validation for left behind passengers

A logistic regression is used to predict the probability that a passenger is left behind on the station platform based on automated dwell time estimates and/or automated passenger counts from video. The model parameters are estimated using the manually
observed counts of passengers left behind on the station platforms as the observed outcome. In this study, data collected on November 15, 2017, were used for model estimation. The diagnostics, parameters, and fit statistics are presented for three models in Section 7.1.

The quality of the proposed models is evaluated through validation against manually collected counts on a different day. In this study, the estimated models are used to predict the number of left behind passengers using automated dwell time estimates and automated passenger counts on January 31, 2018. The accuracy of the model predictions is then calculated relative to manually observed passenger counts on the same day, as shown in Section 7.2.

3.5. Model implementation

Implementation of the model to make ongoing estimates of the numbers of passengers left behind each departing train requires only train tracking data and surveillance video feeds as model inputs. The manual observations of door opening/closing times and the number of passengers on the platforms are used only for estimating model parameters. The models then produce predictions of the number of passengers left behind each departing train based only on data that is automatically collected. Therefore, the numbers of left behind passengers and the associated impact on the distribution of wait times experienced by passengers can be tracked as a performance measure over time. If data feeds were processed as they are recorded, it would also be possible to implement the models to make real-time predictions of the left behind passengers.

4. Identification of study sites

To test the implementation of object detection with video in transit stations, a first step is to identify locations and times to collect video feeds as well as direct manual observations of left-behind passengers. For this study, stations were selected based on a crowding analysis and evaluation of station geometry and camera view characteristics. The goal was to identify stations with the greatest likelihood of passengers being left behind during a typical morning or afternoon rush and where object detection techniques would be most successful. The analysis focused on the Orange Line, which is 11 miles long with 20 stations. Oak Grove and Forest Hills are the northern and southern end stations, respectively. There are two main reasons for choosing this specific line. First and most important, it has no branch lines, so all travelers can reach their destination by boarding the next available train. This simplifies the identification of left-behind passengers. Second, it passes through several transfer stations in the center of Boston, which highlights its significance for passengers’ daily commuting.

4.1. Crowding analysis

A crowding analysis is a necessary step to identify the times and stations where crowding is observed and left behinds have the highest probability of occurring. The data used in this part of the analysis have been extracted from the Rail Flow database in the MBTA Research and Analytics Platform. The Rail Flow dataset includes aggregated boarding and alighting counts by time of day with 15-min temporal resolution averaged across all days in a calendar quarter. An example is given in Fig. 2 for 5:15–5:30 pm in Winter 2017. These data are derived from the Origin–Destination-Transfer (ODX) model, which makes use of AFC and AVL systems to infer the flow of passengers within the subway (Sánchez-Martínez, 2017).

The ODX model identifies records from AFC that can be linked in order to infer transfers or return trip patterns. For example, a passenger using a Charlie Card (MBTA’s farecard) to enter a rail station and later board a bus near a different rail station can be assumed to have used the rail system and then transferred to the bus. Another passenger who enters one rail station in the morning and enters a different rail station in the afternoon may be completing a round-trip commute, so the destination of the morning and afternoon trips can be inferred by linking the two trips. Some trip origins and/or destinations cannot be inferred, for example if the fare is paid with cash or the trip has only one farecard transaction. For more details about the ODX model, the reader is referred to Sánchez-Martínez (2017), where the model’s application inferred the origins of 98% and the destinations of 73% of the total number of fare transactions.

For the crowding analysis in this paper, cumulative counts of passengers boarding and alighting at each station have been created along the direction of train travel using the aggregated rainflow data. For a 15-min time period, $B(n, t)$ is the cumulative count of all passengers that board trains in the direction of interest at stations preceding and including station $n$ during time interval $t$. Similarly, $A(n, t)$, is the cumulative count of passengers that are assumed to have exited trains traveling in the direction of interest at stations preceding and including station $n$ during time interval $t$. It should always be true that $A(n, t) \leq B(n, t)$, because passengers can only alight a train after boarding it.

The difference between the cumulative boardings, $B(n, t)$, and alightings, $A(n, t)$, is the estimated passenger flow, $Q(n, t)$, between station $n$ and $n + 1$ during each 15-min time period.

$$Q(n, t) = B(n, t) - A(n, t)$$

This calculation is approximate, because cumulative counts are calculated for a single 15-min time period, and real trains take more than 15 min to traverse the length of a line.

To calculate the number of passengers per train, the passenger flow per time period must be converted to passenger occupancy, $O(n, t)$ (passengers/train), which is calculated by multiplying the passenger flow by the scheduled headway of trains, $h(t)$ (minutes), at time $t$. 

5
The headway is divided by 15 min to account for the fact that the passenger flow is per 15-min time period. This measure is an approximation of the number of passengers onboard each train that is based on the assumptions that headways are uniform and passengers are always able to board the next arriving train. In reality, variations in headways may lead to increased crowding after longer headways, increasing the likelihood that some passengers will be left behind.

The 2017 MBTA Service Delivery Policy (SDP) (MBTA, 2017) provides guidelines for reliability and vehicle loads. In the 2010 MBTA SDP (MBTA, 2010), the maximum vehicle load was explicitly defined as 225% of seating capacity in the peak hours (start of service to 9:00 am; 1:30 pm – 6:30 pm) and 140% of the seating capacity in other hours. The 2017 SDP notes that accurately monitoring the passenger occupancy of heavy rail transit is not yet feasible on the MBTA system. Nevertheless, the guidelines from Table B2 in the 2017 SDP are used to identify general crowding levels, recognizing that each Orange Line train is six cars long and has a total of 348 seats.

A visualization of average train occupancy for the Winter 2017 Rail Flow data is shown in the color plot in Fig. 3a. The color for each station and 15-min time interval corresponds to the value of $O(n, t)$. Since the trains have 348 seats, red parts of the plot indicate large numbers of standing passengers, with dark red indicating crowding near vehicle capacity. This figure shows that in the northbound direction, the most severe crowding occurs between Downtown Crossing and North Station shortly before 6:00 pm. Note that the crowding appears to decrease before rebounding again at 6:30 pm. This is due to the change in scheduled headway at 6:30 pm from 6 min to 10 min, which increases occupancy, as calculated in Eq. (2).

$$O(n, t) = Q(n, t) \times \frac{h(t)}{15}$$

The headway is divided by 15 min to account for the fact that the passenger flow is per 15-min time period. This measure is an approximation of the number of passengers onboard each train that is based on the assumptions that headways are uniform and passengers are always able to board the next arriving train. In reality, variations in headways may lead to increased crowding after longer headways, increasing the likelihood that some passengers will be left behind.

The 2017 MBTA Service Delivery Policy (SDP) (MBTA, 2017) provides guidelines for reliability and vehicle loads. In the 2010 MBTA SDP (MBTA, 2010), the maximum vehicle load was explicitly defined as 225% of seating capacity in the peak hours (start of service to 9:00 am; 1:30 pm – 6:30 pm) and 140% of the seating capacity in other hours. The 2017 SDP notes that accurately monitoring the passenger occupancy of heavy rail transit is not yet feasible on the MBTA system. Nevertheless, the guidelines from Table B2 in the 2017 SDP are used to identify general crowding levels, recognizing that each Orange Line train is six cars long and has a total of 348 seats.

A visualization of average train occupancy for the Winter 2017 Rail Flow data is shown in the color plot in Fig. 3a. The color for each station and 15-min time interval corresponds to the value of $O(n, t)$. Since the trains have 348 seats, red parts of the plot indicate large numbers of standing passengers, with dark red indicating crowding near vehicle capacity. This figure shows that in the northbound direction, the most severe crowding occurs between Downtown Crossing and North Station shortly before 6:00 pm. Note that the crowding appears to decrease before rebounding again at 6:30 pm. This is due to the change in scheduled headway at 6:30 pm from 6 min to 10 min, which increases occupancy, as calculated in Eq. (2).
Amoredetailed visualization combines transit vehicle location records and inferred origin–destination trip flows from a specific date. As mentioned already, the ODX trip flows are constructed with simplifying assumptions about passenger movements; for example, all passengers entering a station are assumed to board the first arriving train. Despite such assumptions, however, the model is valuable for many applications. The trajectories in Fig. 3b are associated with the recorded arrival and departure times of train at each station. The colors are associated with the estimated train occupancy based on the inferred boardings and alightings, assuming that no passengers are left behind. The trajectory plot shows that the headways between trains can vary substantially, especially for

**Fig. 3.** Inferred passenger crowding for northbound Orange Line trains (from Forest Hills to Oak Grove). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
the stations north of Downtown Crossing. Longer headways are followed by more crowded trains, because more passengers have arrived to board since the previous train. The occurrence of left-behind passengers would make actual train occupancies slightly lower for the trains following long headways. Those left-behind passengers would then be waiting to board the next train, thereby increasing the occupancy on one or more subsequent trains.

Tracking the average number of passengers onboard trains provides an indicator for the likelihood of passengers being left behind, because full trains leave little room for additional passengers to board. During the most crowded times of the day, it is also useful to look at the numbers of passengers boarding and alighting trains at each station. Passengers are most likely to be left behind at stations where trains arrive with high occupancy, few passengers alight, and many more passengers wait to board. By this measure, North Station in the afternoon peak appears to be an ideal candidate for observing left behind passengers. Using the same method for the southbound direction, Sullivan Square station was identified as an ideal candidate location for data collection in the morning peak. Other candidate stations include Back Bay, Chinatown and Wellington stations.

4.2. Station geometry and camera views

In addition to identifying stations with the greatest likelihood of passengers getting left behind crowded trains, the stations that are selected for detailed analysis should also have characteristics that are amenable to successful testing of video surveillance counting methods. There are a variety of station layouts and architectures that contribute complicating factors to the analysis of left behind passengers, and the goal of this study is to identify the potential for the adopted detection method under the best possible conditions. Ideal conditions for the proposed analysis are:

- Dedicated Platform for Line and Direction of Interest – In this case, all passengers on a platform are waiting for the same train, so any passenger that does not board can be counted as being left behind. In the case of an island platform, observed passengers may be waiting for trains arriving on either track. In the MBTA system, more than half of the station platforms for heavy rail rapid transit in the city center (the most crowded part of the system) meet this criterion.\(^1\)
- High Quality Camera Views – Surveillance cameras vary in age, quality, and placement throughout the MBTA system. Newer cameras have higher definition video feeds. The quality of the view is also affected by lighting conditions, especially at above-ground station where sunlight and shadows can affect the clarity of the images.
- Platform Coverage of Camera Views – The surveillance systems are designed to provide views of the entire platform area for security purposes. In some stations, the locations of columns obfuscate the views, requiring more cameras to provide this coverage.

Surveillance camera views were considered from five stations on the Orange Line (Back Bay, Chinatown, North Station, Sullivan Square, and Wellington) that were identified through crowding analysis as candidate stations. Ultimately, North Station was selected as the study site for the northbound direction afternoon peak period because the station exhibits consistent crowding and the geometry provided good camera views. Samples of the camera views from this station are shown in Fig. 4.

5. Manual data collection

Manual observations on the platform needed to be collected to establish a ground truth against which to compare alternative methods for measuring and estimating the number of passengers left behind crowded trains. Detailed data collection at North Station was conducted during afternoon peak hours (3:30–6:30 pm) on midweek days during non-holiday weeks (Wednesday, November 15, 2017, and Wednesday, January 31, 2018). Three observers worked simultaneously on the station platform to record observations.

5.1. Train door opening and closing times

Although Train-Tracking Records (TTR) report the times that each train enters the track circuit associated with a station, there is no automated record of the precise times that doors open and close. Since passengers can only board and alight trains while the doors are open, recording these times manually is important for identifying when passengers board trains, when they are left behind, the precise dwell time in the station, and the precise headway between trains. Each of the three observers recorded the times of doors opening and closing. The average of these observations is considered the true value.

A simple linear regression model shows that observed dwell times (time from doors opening to doors closing) can be accurately estimated from automatic records of TTR arrival and departure times associated with each station. Fig. 5 shows the data and regression results combining manual counts for November 15, 2017 and January 31, 2018. There is no systematic difference between records from different days, and the \(R^2\) is greater than 0.9, indicating a good fit.

\(^1\) All stations from Tufts Medical Center through Haymarket and the northbound platform at North Station on the Orange Line (11 platforms), three out of four Blue Line stations in downtown Boston (5 platforms), and all northbound platforms for the Red Line from South Station to Porter (8 platforms) meet this criterion.
5.2. Number of passengers left behind

Each observer counted the number of passengers left behind on the station platforms after the train doors closed. In order to avoid double-counting, each observer was responsible for observing passengers in a two-car segment of the six-car train (front, middle, and back). Some judgement was necessary in determining which passengers to count, because some passengers linger on the platform after alighting the train and some choose to wait for a later train even when there is clearly space available to board. The goal of the left-behind passenger count is to measure the number of passengers that are left behind due to crowding within $\pm 2$ passengers of the true number.

5.3. Number of passengers waiting on platform

In addition to counting the number of passengers left behind by crowded trains, it is important for model calibration to get an accurate count of the number of passengers waiting to board each arriving train. Given the large number of commuters using the heavy rail system during commuting hours, it is not possible to accurately count this total number of passengers in person. Surveillance video feeds of escalators, stairs, and elevators used to access the platform of interest were used to manually count the number of passengers entering and exiting the platform offline. Specifically, an open-source software tool was used to track passenger movements by logging keystrokes to the video timestamp during playback (Campbell, 2012). Counts were conducted by watching the surveillance video playback of each entry and exit point from the platform and logging the entry and exit of each individual passenger. The resulting data log records the time (to the nearest second) that each passenger entered and exited the platform. Since the platforms of interest serve only one train line in one direction, all entering passengers are assumed to wait to board the next train, and all exiting passengers are assumed to have alighted the previous train. Combining these counts with the direct observations of the number of passengers left behind each time the doors close provides an accurate estimate of the number of passengers that were successfully able to board each train. Fig. 6 illustrates the cumulative numbers of passengers entering the platform (blue curve) and boarding the trains (orange curve). The steps in the orange curve correspond to the times that the train doors close. If passengers are assumed to arrive onto the platform and board trains in first-in-first-out (FIFO) order, the red arrow represents the waiting time that is experienced by the respective passenger, which is estimated as the difference between the arrival and the boarding time.

A timeseries of the actual number of passengers waiting on the platform is constructed by counting the cumulative arrivals of passengers to the platform over time and assuming that all passengers board departing trains except those that are observed to be left behind. This ground truth for data collected on November 15, 2017, is shown in blue in Fig. 7. The sawtooth pattern shows the growing number of passengers on the platform as time elapses from the previous train. The drops correspond to the times when doors close. At these times, the platform count usually drops to zero. When passengers are left behind, the timeseries drops to the number of left behind passengers. One such case is illustrated with the red arrow just before 17:30 in Fig. 7.
6. Automated detection of passengers on platforms in video feeds

6.1. Calibration of parameters

The YOLO algorithm uses pattern recognition to identify objects in an image. The COCO training dataset was used to define the object detection parameters in YOLO, as described in Section 2. A threshold for certainty can also be calibrated to adjust the number of identified objects in a specific frame. If the threshold is set too high, the algorithm will fail to recognize some objects that do not adequately match the training dataset. If the threshold is set too low, the algorithm will falsely identify objects that are not really present. In order to identify the optimal threshold, frames from 14 camera views were analyzed. Each frame was analyzed separately for threshold values ranging from 6% to 25% to determine the optimal threshold value in relation to a manual count of passengers visible in the frame. The optimal threshold across all camera views is 7%, which minimizes the mean squared error between YOLO and manual counts as shown in Table 1. Fig. 8 shows the identified objects at each threshold level for the same frame from a camera installed in North Station.

6.2. Raw video counts

The input for YOLO is a set of frames, each of which are analyzed independently to detect objects. The algorithm runs quickly enough to analyze each frame in less than one second, so the surveillance video feeds are sampled at one frame per second to allow YOLO to run faster than real time. Although the analysis for this paper was conducted off line, it would be possible to implement the algorithm in real time.

The output from YOLO is a text file that lists the objects detected for each frame and the bounding box for the object within the image. A time series count of passengers on the platform is simply the number of “person” objects identified in the corresponding
frames from each sample video feed. Fig. 9a shows the raw passenger counts on the platform at North Station for the time period from 5:00 pm – 6:30 pm on November 15, 2017. Although there are noisy fluctuations, there is a clear pattern of increasing passenger counts until door opening times (green). A surge of passenger counts while doors are open (between green and red) represents the passengers alighting the train and exiting the platform. Passenger counts drop off dramatically following the door closing time (red), except in cases that passengers are left behind. For example, the third train in Fig. 9a arrives after a long headway and shows roughly nine passengers left behind.

To facilitate analysis of the automatic passenger counts from the surveillance videos, it is useful to work with a smoothed time series of passenger counts. Using a smoothing window of ±10 seconds, the smoothed series is shown in Fig. 9b. This smoothed time series is more suitable for a local search to identify the minimum passenger count following each door closing time. This represents the count of left-behind passengers identified through the automated object detection process.

6.3. Accuracy of detected left behind passengers

The smoothed video counts from the three surveillance camera feeds used to monitor the northbound Orange Line platform at North Station are shown as the green curve in Fig. 10. The automated passenger counting algorithm clearly undercounts the total number of passengers on the platform. The reason for this large discrepancy is that the algorithm can only identify people in the foreground of the images, where each person is large. Therefore, the available camera views do not actually provide complete coverage of the platform for automated counting purposes. Furthermore, when conditions get very crowded, it becomes more difficult to identify separate bodies within the large mass of people.

The problem of undercounting aside, it is clear that the automated counts generate a pattern that is representative of the total number of passengers on the platform. Using regression, the smoothed timeseries can be linearly transformed into a scaled timeseries (the orange curve in Fig. 10), which minimizes the squared error compared with the manually counted timeseries. Using this scaling method, the data from November 15, 2017, were used to compare estimated counts of left-behind passengers in the peak periods with the directly observed values. This provides a measure of the accuracy of automated video counts. The total number of left-behind passengers estimated by this method is presented in Table 2, where the Root Mean Squared Error (RMSE) is calculated by comparing the number of passengers left-behind each time the train doors close.

The scaling process, which makes the blue and orange curves in Fig. 10 match as closely as possible, results in substantially overcounted left beinds, because the scaling factor tends to over-inflate the counts when there are few passengers on the platform. As a direct measurement method, automated video counting is not satisfactory, at least as implemented with YOLO. However, Fig. 10

Table 1: Passenger count errors for YOLO thresholds.

| Threshold | Mean Error | Mean Squared Error |
|-----------|------------|--------------------|
| 25%       | 6.82       | 62.0               |
| 20%       | 5.82       | 43.7               |
| 15%       | 4.59       | 28.5               |
| 10%       | 3.06       | 14.1               |
| 9%        | 2.23       | 8.0                |
| 8%        | 1.35       | 4.6                |
| 7%        | 0.12       | 1.2                |
| 6%        | −1.53      | 6.8                |

Fig. 8. Effect of threshold values on object detection in a sample video frame, North Station.
shows a clear relationship between the video counts and passengers being left behind on station platforms, so there is potential to use the video feed as an explanatory variable in a model to estimate the likelihood of passengers being unable to board a train.

### 7. Modeling left-behind passengers

In order to improve the accuracy of estimates of the number of passengers left behind on subway platforms, a logistic regression model is formulated to estimate the probability that each passenger is left behind based on explanatory variables that can be collected automatically. A logistic regression is used to estimate the number of passengers left behind by way of estimating the probability that each waiting passenger is left behind, because the logistic function has properties that are more amenable to this application. Since passengers are only left behind when platforms and trains are very crowded, a linear regression has a tendency to provide many negative estimates of left behind passengers, which are physically impossible. The binary logit model, by contrast is intended for estimating the probability that one of two possible outcomes is realized (e.g., a passenger is either left behind or not left behind). The estimated probability from a logit model is always between 0 and 1, so the resulting estimate of the number of left-behind passengers is always non-negative and cannot exceed the total number of waiting passengers.

For estimation of the logistic regression, each passenger is represented as a separate observation, and all passengers waiting for the same departing train are associated with the same set of explanatory variables. Over the course of a 3-h rush period, there are typically about 30 trains serving North Station, serving 1,500 to 3,000 passengers per period, and leaving behind well over 100
Logistic regression models are generally expected to give stable estimates when the dataset for fitting includes at least 10 observations for each outcome, so there is sufficient data to estimate parameters for a model that is structured this way.

The logistic function defines the probability that a passenger is left behind by

\[ P(x) = \frac{1}{1 + e^{-\beta_0 + \beta x}} \]  

(3)

where \( x \) is a vector of explanatory variables, \( \beta \) is a vector of estimated coefficients for the explanatory variables, and \( \beta_0 \) is an estimated alternative-specific constant. The estimation of the model can be thought of as identifying the values of \( \beta_0 \) and \( \beta \) that best fit the observed outcomes

\[ y = \begin{cases} 
1, & \beta_0 + \beta x + \epsilon > 0 \\
0, & \text{else}
\end{cases} \]  

(4)

where \( y = 1 \) corresponds to a passenger being left behind, and \( y = 0 \) corresponds to a passenger successfully boarding.

The underlying assumption in this formulation is that the likelihood of being left behind can be expressed in terms of a linear combination of explanatory variables and a random error term, \( \epsilon \), which is logistically distributed. The explanatory variables that are considered in this study are as follows:

1. Dwell time (time from door opening to door closing) or difference of TTR arrival and departure times
2. Video count of passengers on platform following doors closing

These explanatory variables can all be monitored automatically, without manual observations. Video counts of passengers on the platform following doors closing are obtained from the object detection process described above. Although dwell time is an appropriate explanatory variable because doors stay open longer when trains are crowded, the dwell time is not directly reported in archived databases. As demonstrated in Fig. 5, observed dwell times can be accurately estimated from automatic records of TTR arrival and departure times. This leads to using TTR reported values of difference between train arrival and departure instead of dwell times for the model development. Since these are essentially the same explanatory variable, we call this difference “dwell time” for the remainder of the paper.

7.1. Model estimation

Initially, three models were estimated, making use of only TTR data (Model 1), only video counts (Model 2), and then fused TTR
and video counts (Model 3). The data from November 15, 2017, were used to develop these models. The number of passengers waiting on the platform (as described in Section 5.3) are used to determine the number of observations for estimating the parameters of the logit model. In total, 2167 passengers boarded arriving trains at North Station during the rush period and 198 of them were left behind. This leads to a sample size of 2365 passengers for the logistic models.

Models 1 and 2 are simple logistic regressions, each with only one independent variable. Neither model has influential values (i.e., values that, if removed, would improve the fit of the model). Model 3 uses both TTR data and video counts, so it is important to diagnose the model’s fit, especially with respect to the assumptions of the logistics regression. First, multicollinearity of explanatory variables should be low. The correlation between dwell time and video count is -0.643 and the variance inflation factor is 1.7, both indicating that the magnitude of multicollinearity is not too high. Second, no influential values were identified. Third, the logistic regression is based on the assumption that there is a linear relationship between each explanatory variable and the logit of the response, \( \log\left(\frac{p}{1-p}\right) \), where \( p \) represents the probabilities of the response. Fig. 11 shows that dwell time is approximately linear with the logit response, while there is somewhat more variability with respect to the video counts. Neither plot suggests that there is a systematic mis-specification of the model.

A summary of the estimated model coefficients and fit statistics is presented in Table 3. The log likelihood is a measure of how well the estimated probability of a passenger being left behind matches the observations. The null log likelihood is associated with no model at all (every passenger is assigned a 50% chance of being left behind), and values closer to zero indicate a better fit. The \( \chi^2 \) value is a related measure of model fit, with values closer to 1 indicating a better model.

For all three models, the estimated coefficients have the expected signs and magnitudes. The positive coefficients for dwell time and video counts indicate a positive relationship with the probability of having left-behind passengers, which is intuitive. In order to compare models, the likelihood ratio statistic is used to determine whether the improvement of one model is statistically significant compared to another. The likelihood ratio test statistic is calculated by comparing the log likelihood of the restricted model (with fewer explanatory variables) to the unrestricted model (with more explanatory variables):

\[
D = 2(LL_{\text{unrestricted}} - LL_{\text{restricted}}) 
\]  
(5)

Comparing Model 1 (restricted) to Model 3 (unrestricted), one additional variable in Model 3, indicates one degree of freedom, which

### Table 3
Logistic regression model parameters, North Station, November 2017.

| Parameter                  | Model 1 |      | Model 2 |      | Model 3 |      |
|----------------------------|---------|------|---------|------|---------|------|
| Value | p-stat | Value | p-stat | Value | p-stat |      |
| Constant for Left-Behind  | -10     | 0.00 | -4.18   | 0.00 | -7.57   | 0.00 |
| Dwell Time (sec)           | 0.0903  | 0.00 | 0.370   | 0.00 | 0.0487  | 0.00 |
| Video Count                |         |      |         |      | 0.232   | 0.00 |
| Null Log Likelihood, \( LL_0 \) | -1639.29 |      | -1639.29 |      | -1639.29 |      |
| Model Log Likelihood, \( LL \) | -551.94  |      | -533.28 |      | -514.43 |      |
| \( \hat{p}^2 \)           | 0.663   |      | 0.675   |      | 0.686   |      |
| AIC                        | 1107.9  |      | 1070.6  |      | 1034.9  |      |
Table 4
Validation of probability and count of left behind passengers, North Station, January 31, 2018.

| Train | Manual Count | Unscaled Video | Scaled Video | Model 1 | Model 3 |
|-------|--------------|----------------|--------------|---------|---------|
|       | Prob. | Count | Count | Count | Prob. | Count | Prob. | Count |
| 1     | 6.5%  | 2     | 1    | 7    | 3.2%  | 2     | 2.2%  | 2     |
| 2     | 0.0%  | 0     | 2    | 15   | 1.0%  | 0     | 1.5%  | 0     |
| 3     | 0.0%  | 0     | 2    | 15   | 0.4%  | 0     | 0.9%  | 0     |
| 4     | 0.0%  | 0     | 1    | 7    | 0.8%  | 0     | 1.1%  | 1     |
| 5     | 0.0%  | 0     | 0    | 3    | 0.3%  | 0     | 0.5%  | 0     |
| 6     | 0.0%  | 0     | 0    | 0    | 0.7%  | 0     | 0.8%  | 0     |
| 7     | 0.0%  | 0     | 0    | 0    | 0.6%  | 0     | 0.7%  | 0     |
| 8     | 0.0%  | 0     | 2    | 15   | 1.2%  | 1     | 1.7%  | 1     |
| 9     | 0.0%  | 0     | 2    | 15   | 2.1%  | 1     | 2.2%  | 1     |
| 10    | 0.0%  | 0     | 1    | 7    | 0.6%  | 0     | 0.9%  | 0     |
| 11    | 15.2% | 23    | 2    | 15   | 18.2% | 27    | 7.4%  | 11    |
| 12    | 0.0%  | 0     | 1    | 7    | 1.2%  | 1     | 1.3%  | 1     |
| 13    | 0.0%  | 0     | 1    | 7    | 2.1%  | 1     | 1.8%  | 1     |
| 14    | 0.0%  | 0     | 2    | 15   | 0.8%  | 0     | 1.3%  | 0     |
| 15    | 14.4% | 24    | 3    | 22   | 15.6% | 19    | 8.4%  | 10    |
| 16    | 5.8%  | 5     | 4    | 30   | 7.0%  | 6     | 6.6%  | 6     |
| 17    | 14.6% | 19    | 3    | 22   | 19.5% | 24    | 9.6%  | 12    |
| 18    | 10.6% | 14    | 7    | 77   | 18.2% | 20    | 33.9% | 37    |
| 19    | 8.7%  | 9     | 3    | 22   | 6.4%  | 3     | 5.1%  | 2     |
| 20    | 2.4%  | 1     | 4    | 30   | 0.9%  | 0     | 2.3%  | 1     |
| 21    | 3.5%  | 4     | 3    | 22   | 3.8%  | 3     | 3.8%  | 3     |
| 22    | 3.0%  | 3     | 2    | 15   | 1.7%  | 1     | 2.0%  | 2     |
| 23    | 0.0%  | 0     | 1    | 7    | 0.8%  | 0     | 1.0%  | 0     |
| 24    | 0.0%  | 0     | 3    | 22   | 0.8%  | 0     | 1.6%  | 1     |
| 25    | 0.0%  | 0     | 3    | 22   | 1.1%  | 1     | 2.0%  | 2     |
| 26    | 2.7%  | 2     | 2    | 15   | 2.7%  | 1     | 2.5%  | 1     |
| 27    | 6.7%  | 7     | 2    | 15   | 3.2%  | 3     | 2.8%  | 2     |
| 28    | 3.6%  | 2     | 3    | 22   | 0.5%  | 0     | 1.3%  | 1     |
| 29    | 3.6%  | 2     | 1    | 7    | 0.8%  | 0     | 1.0%  | 0     |
| 30    | 2.7%  | 3     | 2    | 15   | 2.5%  | 2     | 2.4%  | 2     |
| Total |        |       |      |      |       |       |       |       |
|       | 120   | 66    | 490  | 116  | 100   |
| MAE   |       |       |       | 3.5  | 12.9  |
| RMSE  |       |       |       | 6.5  | 17.6  |

requires \( D > 3.84 \) to reject the null hypothesis at the 0.05 significance level. Comparison between Models 1 and 3 gives \( D = 75.02 \), indicating that Model 3 provides a significant improvement over Model 1 by adding video counts. Comparison between Models 1 and 2 gives \( D = 37.7 \), which is also a significant improvement. The Akaike Information Criterion (AIC) is an additional model fit statistic that weighs the log likelihood against the complexity of the model. Although Model 3 has more parameters, the AIC is greater than for Model 1 or Model 2, indicating that the improved log likelihood justifies the inclusion of both TTR and video count data.

7.2. Model validation

7.2.1. Number of passengers left behind

The logistic regression provides an estimate of the probability that passengers are left behind each time the train doors close. In order to translate this probability into a passenger count, the estimated number of passengers waiting on the platform from the scaled video count is used as an estimate of the number of passengers waiting to board. Table 4 shows the validation results when the models were applied to data collected on January 31, 2018, for North Station. The scaling factor used for the number of passengers waiting on the platform is estimated from November 15, 2017 data. Considering the estimated number of left behind passengers for each train separately, it is observed that these models achieve higher accuracy when there are a few passengers left behind. Overall, Model 1 exhibits error of only 3.3% since it estimates that 116 passengers are left behind in total when 120 passengers were observed to be left behind. Model 3 gives a lower estimate of 100 passengers being left behind, which leads to an error of approximately 17%.

As shown in Table 2 and Table 4, direct video counts (unscaled and scaled) do not provide accurate estimates of the total numbers of passengers left behind without some additional modeling. The unscaled video counts underestimate the total, while the scaled video counts overestimate the total. The logistic regression provides much better results. Although there are some discrepancies for specific train departures, the estimated numbers of passengers left behind are not significantly biased and the total number of passengers left behind during the three-hour rush period is similar to the manually counted total.

The logistic regressions estimate the probability of a passenger being left behind using only the explanatory variables listed in Table 3. However, the estimated number of left behind passengers is calculated by multiplying the probability by the scaled video count of passengers on the platform at the time the doors opened, as estimated from the TTR data. Therefore, the estimated number of
passengers left behind with Model 1 and Model 3 rely only on TTR data that is currently being logged and supplemented by automated counts of passengers in existing surveillance video feeds. The models therefore utilize explanatory variables that are monitored automatically, and they can be deployed for continuous tracking of left behind passengers without needing additional manual counts.

The logistic models could actually perform even better if there were a way to obtain a more accurate count of the number of passengers waiting for a train. During the morning peak period, the count of farecards entering outlying stations can provide a good estimate for the number of passengers waiting to board each inbound train. This is more challenging at a transfer station, like North Station, in which many passengers are transferring from other lines. In some cases, strategically placed passenger counters could provide useful data. Nevertheless, Table 5 presents the performance of the developed logistic regression models if their estimated probabilities are multiplied by the actual number of passengers on the platform instead of the estimated number as in Table 4. This reveals the value of more accurate data, because Model 3 decreases its error compared to Table 4. Model 3 in Table 5 estimates 122 passengers being left behind in the afternoon rush on the observed date when the previous estimate was 100, which is a reduction of error from 17% to 2% for this model compared to the 120 observed left behind passengers.

7.2.2. Occurrence of a train leaving behind passengers

Another way to evaluate the performance of the developed models is to consider whether or not trains that leave behind passengers can be distinguished from trains that allow all passengers to board. Through the course of data collection and analysis, the number of passengers being left behind because of overcrowding can only be reliably observed within approximately ±2 passengers. The reason for this is that sometimes people choose not to board a train for reasons other than crowding, and one or two passengers left on the platform did not appear to be consistent with problematic crowding conditions.

If a train is defined to be leaving behind passengers when more than 2 passengers are left behind, the results presented in Table 4 can be reinterpreted to evaluate each method by four measures:

1. **Number of Trains Leaving Behind Passengers**: The number of trains in a time period that leave behind passengers due to overcrowding.
2. **Correct Identification Rate**: The percent of trains that are correctly classified as leaving behind passengers or not leaving behind passengers, as compared to the manual count. This value should be as close to 1 as possible.
3. **Detection Rate**: The percent of departing trains that were manually observed to leave behind passengers that are also flagged as such by the estimation method. This value should be as close to 1 as possible.
4. **False Detection Rate**: The percent of departing trains that are estimated to leave behind passengers but have not, according to manual observations. This value should be as close to 0 as possible.

There is an important distinction to make here, because there are two ways that the model to identify trains leaving behind passengers can be used:

(a) to estimate the number of trains that leave behind passengers, in which case we only care about measure 1; or
(b) to identify which specific trains are leaving behind passengers, in which case measures 2 through 4 are important.

Depending on how the data will be used, application (a) or (b) may be more relevant. For example, application (a) provides an aggregate measure of the number of trains leaving behind passengers. Application (b), on the other hand, is what would be needed to get toward a real-time system for identifying (even predicting) left-behind passengers.

A comparison of the four measures is presented in Table 6 for the 30 trains that departed North Station between 3:30 pm and 6:30 pm on January 31, 2018. Unscaled video counts provide a good estimate of the number of trains that leave behind passengers.
(measure 1), but suffer from a low detection rate and high false detection rate. Scaled video counts are poor estimators for the occurrence of left-behind passengers because they are high enough to trigger too many false detections. The modeled estimates both perform well in approaching the actual number of trains leaving behind passengers. Model 3 has the best performance for measures 2 through 4. It never falsely identifies a train as leaving behind passengers, and it correctly detects most occurrences of passengers being left behind. Like the count estimates above, both Model 1 and Model 3 rely on the scaled video counts to estimate the number of passengers waiting on the platform when the train doors open, so a fusion of TTR records and automated video counts provide the most reliable measures.

7.2.3. **Distribution of experienced waiting times**

Another application of the model is to consider the distribution of waiting times implied by the estimated probabilities that passengers are left behind each departing train. From the direct manual counts, a cumulative count of passengers arriving onto the platform and of passengers boarding trains provides a time series count of the number of passengers on the platform. If passengers are assumed to board trains in the same order that they enter the platform, the system follows a first-in-first-out (FIFO) queue discipline. Although it is certainly not true that passengers follow FIFO order in all cases, this assumption allows the cumulative count curves to be converted into estimated waiting times for each individual passenger. The FIFO assumption yields the minimum possible waiting time that each passenger could experience, and the waiting time for each passenger can be represented graphically by the horizontal distance between the cumulative number of passengers entering the platform and boarding trains (see Fig. 6 for data from November 15, 2017). The yellow curve in Fig. 12a represents the cumulative distribution of waiting times that are implied by the observed numbers of passengers entering the platform if all passengers on the platform are assumed to be able to board the next departing train. We call this the expected waiting time. The blue curve in Fig. 12a is the cumulative distribution of waiting times if the number of left-behind passengers are accounted for when trains are too crowded to board. We call this the observed waiting time, because it reflects direct observation of passengers waiting on the platform using manual counts. The distribution indicates the percentage of passengers that wait less than the published headway for a train departure, which is the reliability metric used by the MBTA. For the Orange Line during peak hours, the published headway is 6 min (360 s). Currently, the MBTA is only able to track the expected wait time as a performance metric. The difference between the yellow and blue curves indicates that failing to account for left-behind passengers leads to overestimation of the reliability of the system.

The models developed in this study provide the estimated probability that a passenger is left behind each time the train doors close. In the absence of additional passenger count data, a constant arrival rate is assumed over the course of the rush period, the door closing times from TTR and the probability of passengers being left behind from Model 3 can be used to estimate the cumulative passenger boardings onto trains over time. Under the same FIFO assumptions described above, the distribution of experienced waiting times can be estimated based on train-tracking and video counts. By this process a cumulative distribution of waiting times is estimated using probabilities from Model 3 is shown as a red curve in Fig. 12b, which we call the uniform arrivals modeled wait time. Table 7 includes the values of experienced waiting times for the observed, the expected, and the modeled distributions. This table also shows how the accuracy of estimating waiting times can be improved if we consider the actual arrival rate under the same assumptions used to develop the uniform arrivals modeled wait time. We call this distribution the actual arrivals modeled wait time. The Earth Mover’s Distance (EMD) is used to measure the difference between the observed distribution and the expected, uniform arrivals and actual arrivals modeled distributions (Rubner et al., 2000). As shown in Table 7, the EMD for the expected case is much higher than the EMD for the modeled cases, which indicates that the proposed model reduces errors.

The modeled distributions of waiting times closely approximate the observed distribution. This suggests that the estimated probabilities of passengers being left behind each departing train are consistent with the overall passenger experience. The percentage of passengers experiencing waiting times lower or equal to the 6 min published headway is 79% for both the observed and uniform arrivals model curve, and 77% for the actual arrivals model curve. The automated count of left behind passengers provides a close approximation of the actual service reliability when applied to the independent data collected on January 31, 2018. The expected distribution, which does not account for left-behind passengers produces an estimate of 81% of passengers waiting less than 6 min. The expected distribution overestimates the reliability of the system by failing to account for the waiting time that left-behind passengers experience.

8. Conclusions

This paper presents a method for measuring passengers that are left behind overcrowded trains in transit stations without records of exiting passengers. A study performed by Miller et al. (2018) also addresses this challenging case using manual video counts to calibrate the developed models. The methodology proposed in this paper uses archived data with automatic video counts as inputs to estimate the total number of left behind passengers during peak demand periods. The automatic video counts are obtained through the implementation of image processing tools. This paper presents an investigation of the effects of accounting for left behind passengers on the estimation of the current reliability metric used by the MBTA, the experienced waiting times.

Following a preliminary study of crowding conditions on the MBTA’s Orange Line, data collection and analysis focused specifically on northbound trains at North Station during the afternoon peak hours. Data was collected on two typical weekdays and confirmed that overcrowding is a common problem, even on days without disruptions to service. This is an indication that the system is operating very near capacity, and even small fluctuations in headways lead to overcrowded trains that result in left-behind passengers.

This study specifically investigated the potential for measuring the number of left-behind passengers using existing data sources
and automated passenger counts derived from existing surveillance video feeds. The analysis of automated passenger counts was based on the implementation of a fast, open-source algorithm called You Only Look Once (YOLO) using existing training sets that identify people as well as other objects. The performance is fast enough that frames from surveillance video feeds could potentially be analyzed in real time.

Although video counts were not accurate in isolation, the development of models to use automated video counts with automated train-tracking records (Model 3) demonstrated good results for different applications. In predicting the number of trains leaving

![Distribution of passenger wait times, North Station January 31, 2018.](image)

(a) Comparison between expected (without left-behind passengers) and observed (with actual left-behind passengers) distributions of waiting times

![Uniform Arrival Model](image)

(b) Comparison between observed (with actual left-behind passengers) and estimated (with estimated left-behind passengers) distributions of waiting times

Fig. 12. Distribution of passenger wait times, North Station January 31, 2018.
behind passengers, the developed models can correctly identify whether or not passengers were left behind for 93% of the trains. The number of passengers that are left behind during the afternoon rush period can be estimated within 17% of their actual number using only automated video counts and automatically collected train tracking records. With actual counts of the numbers of passengers on the station platform at each train arrival the model can predict the number of left behind passengers with 2% of the actual number. Furthermore, the modeled distribution of experienced waiting times reduced the total EMD error by more than 50% compared to the error of the operator’s expected distribution, where left-behind passengers are not considered. This highlights the need of accounting for left-behind passengers when tracking the system’s reliability metrics.

There are a number of ways that this study could be extended. One approach would be to implement and evaluate the developed models over more days. In terms of passenger flow data, the ODX model has some known drawbacks given existing limitations, such as lack of tap-out farecard data or passenger counters on trains. In systems without these limitations, the developed models could achieve higher accuracy. The methodology presented here could also be combined with the previous study by Miller et al. (2018) in order to improve the overall process for estimating left behind passengers in subway systems without tap-out. Comparing the two studies, Miller et al. (2018) achieves higher accuracy for very crowded conditions, whereas our method performs better when there are few passengers left behind. The automated object detection presented in our study could also be combined with the model proposed by Miller et al. (2018) as part of its real-time implementation in case of special events where real-time AFC is not available. In the area of image processing, a number of steps could be taken to improve the accuracy of video counts and extend the feasibility to more challenging station environments. Suggested approaches include comparing the algorithm with other fast and accurate video detection algorithms and training the algorithm to detect heads rather than whole bodies. Although there are limitations to any single data source, the potential for improving performance metrics through data fusion and modeling continues to grow.

CRediT authorship contribution statement

Charalampos Sipetas: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. Andronikos Keklikoglou: Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. Eric J. Gonzales: Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgement

This study was undertaken as part of the Massachusetts Department of Transportation Research Program. This program is funded with Federal Highway Administration (FHWA) and State Planning and Research (SPR) funds. Through this program, applied research is conducted on topics of importance to the Commonwealth of Massachusetts transportation agencies.

References

Batarce, M., Muñoz, J.C., de Dios Ortúzar, J., 2016. Valuing crowding in public transport: implications for cost-benefit analysis. Transp. Res. Part A: Policy Pract. 91, 358–378.
Campbell, R., 2012. Simple Player. Computer Software, Berkeley, California.
Carrion, C., Levinson, D.M., 2013. Uncovering the influence of commuters’ perception on the reliability ratio. In: Transportation Research Board 92nd Annual Meeting. Cheng, D., Wang, J., Wei, X., Gong, Y., 2015. Training mixture of weighted SVM for object detection using EM algorithm. Neurocomputing 149, 473–482.
Coniglio, C., Meurie, C., Lézoray, O., Berbineau, M., 2017. People silhouette extraction from people detection bounding boxes in images. Pattern Recogn. Lett. 93, 182–191.
Deng, J., Berg, A.C., Li, K., Fei-Fei, L., 2010. What does classifying more than 10,000 image categories tell us? In: European Conference on Computer Vision. Springer,
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.

Douglas, N., Karpuzis, G., 2005. Estimating the cost to passengers of station crowding. In: 28th Australasian Transport Research Forum (ATRF), Sydney.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The Pascal visual object classes (VOC) challenge. Int. J. Comput. Vision 88 (2), 303–338.

Fan, Y., Guthrie, A., Levinson, D., 2016. Waiting time perceptions at transit stops and stations: effects of basic amenities, gender, and security. Transp. Res. Part A: Policy Pract. 90, 251–264.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587.

Haywood, L., Koning, M., 2015. The distribution of crowding costs in public transport: new evidence from Paris. Transp. Res. Part A: Policy Pract. 77, 182–201.

Haywood, L., Koning, M., Monchambert, G., 2017. Crowding in public transport: who cares and why? Transp. Res. Part A: Policy Pract. 100, 215–227.

Hörcher, D., Graham, D.J., Anderson, R.J., 2017. Crowding cost estimation with large scale smart card and vehicle location data. Transp. Res. Part B: Methodol. 95, 105–125.

Kim, K.M., Hong, S.-P., Ko, S.-J., Kim, D., 2015. Does crowding affect the path choice of metro passengers? Transp. Res. Part A: Policy Pract. 77, 292–304.

Kittelson and Associates Inc., Brinckerhoff, P., Inc, K.G., Institute, T.A.T., Arup, 2013. Transit service and quality of service manual. Tech. rep., Transportation Research Board of the National Academies.

Kraus, M., 1991. Discomfort externalities and marginal cost transit fares. J. Urban Econ. 29 (2), 249–259.

Lézoray, O., Grady, L., 2012. Image Processing and Analysis with Graphs: Theory and Practice. CRC Press.

Li, Z., Hensher, D.A., 2011. Crowding and public transport: a review of willingness to pay evidence and its relevance in project appraisal. Transp. Policy 18 (6), 880–887.

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: European Conference on Computer Vision. Springer, pp. 740–755.

Lundberg, U., 1976. Urban commuting: crowdedness and catecholamine excretion. J. Hum. Stress 2 (3), 26–32.

Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J., 2013. Mining smart card data for transit riders’ travel patterns. Transp. Res. Part C: Emerg. Technol. 36, 1–12.

Ma, Z., Koutsopoulos, H.N., Chen, Y., Wilson, N.H., 2019. Estimation of denied boarding in urban rail systems: alternative formulations and comparative analysis. Transp. Res. Rec. 0361198119857034.

MBTA, 2010. Service Delivery Policy. Massachusetts Bay Transportation Authority, Boston, MA.

MBTA, 2017. Service Delivery Policy. Massachusetts Bay Transportation Authority, Boston, MA.

Miller, E., Sánchez-Martínez, G.E., Nassir, N., 2018. Estimation of passengers left behind by trains in high-frequency transit service operating near capacity. Transp. Res. Rec. 2672 (8), 497–504.

Pelleiter, M.-P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: a literature review. Transp. Res. Part C: Emerg. Technol. 19 (4), 557–568.

Raveau, S., Guo, Z., Muñoz, J.C., Wilson, N.H., 2014. A behavioural comparison of route choice on metro networks: time, transfers, crowding, topology and socio-demographics. Transp. Res. Part A: Policy Pract. 66, 185–195.

Redmon, J., 2016. Darknet: Open source neural networks in C. pjreddie. com.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788.

Rubner, Y., Tomasi, C., Guibas, L.J., 2000. The earth mover’s distance as a metric for image retrieval. Int. J. Comput. Vis. 40 (2), 99–121.

Sánchez-Martínez, G.E., 2017. Inference of public transportation trip destinations by using fare transaction and vehicle location data: dynamic programming approach. Transp. Res. Rec. 2652 (1), 1–7.

Sonka, M., Harvac, V., Boyle, R., 2014. Image Processing, Analysis, and Machine Vision. Cengage Learning.

Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp. Res. Part A: Policy Pract. 53, 36–52.

Velastin, S.A., Boghossian, B.A., Vicencio-Silva, M.A., 2006. A motion-based image processing system for detecting potentially dangerous situations in underground railway stations. Transp. Res. Rec. Emerg. Technol. 14 (2), 96–113.

Viola, P., 1993. Feature-based recognition of objects. In: Working Notes of the AAAI Fall Symposium on Machine Learning in Computer Vision. pp. 60–64.

Zhou, Z.-H., 2012. Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC.