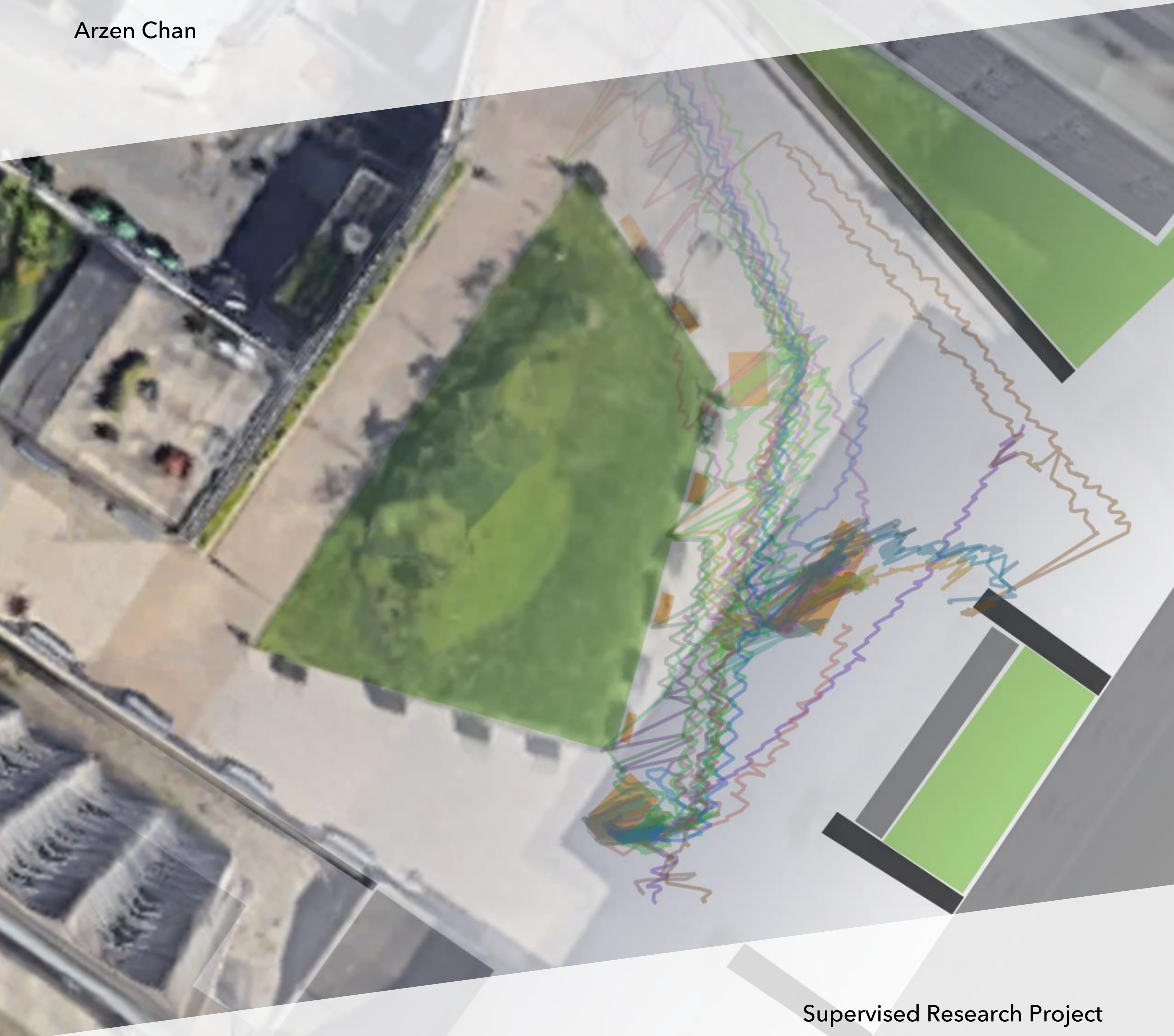
Using Computer Vision in Public Space Design

Applying computer vision tracking methods to public space analysis and design

Arzen Chan



Supervised Research Project
Submitted to Prof. Nik Luka
School of Urban Planning, McGill University
Completed in partial fulfilment of the requirements for the degree Master of Urban Planning
August 2023

Table of Contents

Acknowledgements

I would like to thank Professor Nik Luka for being my supervisor for this research and his endless support and flexibility with the progress of this work. In addition, I would like to thank Tim Elrick, director of the GIC at McGill, for allowing me to use the Centre as a vantage point for data collection and Professors Ahmed El-Geneidy and Raja Sengupta for early guidance in formulating this research. Lastly, I'd like to thank my colleagues in the School of Urban Planning, whose support, suggestions, and camaraderie carried me through this project.

1 Introduction

Public spaces have long been places of meeting, socialization, and gathering for people. The importance of effective public space means the design of such spaces can result in a public space being conducive or detrimental to its conduciveness towards fostering these activities (Gehl, 2011). As such, having detailed information on the behaviour of people within public spaces can aid greatly in the effective design and redesign of public spaces such as squares, plazas, parks, streets, events spaces, and many more (Gehl & Svarre, 2013).

Research on public spaces have had wide and varied applications and methods over the years. Of particular attention are the works of William Whyte (1980) and Jan Gehl (2011), whose empirical approached to studying public spaces lead them to conduct detailed surveys using time-consuming direct observation in order to record the counts of people, lines of movement (termed traces), and activities within public spaces to draw conclusions on how people actually interact with them. These methods typically involved time-lapse videography or direct observation by a researcher for the duration of the process and were manually recorded onto maps. As a result, the methods used by this type of research are labour intensive and may not be justifiable for every application within the urban design sphere (Gehl & Svarre, 2013).

Meanwhile in the field of computer science, much progress has been recently made on the technology of computer vision, allowing computers to identify and track movements of objects and people recorded in a video of a public space with relatively little overhead and comparative skill required from a researcher (Aharon et al., 2022; Lakmali et al., 2020; Sun et al., 2021). As a result, there is a major overlap between these two fields of study, with the new computer vision technology potentially providing a new technical method of gathering the same empirical data that was classically manually gathered by human researchers in the field of urban studies.

This research is intended to be a test of using new technologies on old techniques from the standpoint of an urban designer or planner, not a computer scientist. As a result, the goal is to determine if the current state of computer vision technology is capable of being a significant asset to the field of urban design and planning. Currently, there exists extensive literature for the study of public spaces based on manual methods in the field of urban studies and another body of literature testing the technical nuances of computer vision in the field of computer science (Brunetti et al., 2018). What is missing is a confluence of these areas of literature to both inform computer science research to the particular needs of public space research and to apply the research into a real-world application. The labour-intensive methods from manual tracing methods can be reinforced and replicated by automated methods without the need for specialty infrastructure or active participants. A result is that the costs of these methods can be greatly reduced, meaning having such information from a wide timeframe is much more feasible for projects redesigning public spaces around the world. The emergence of computing power and technology has reduced transactional costs throughout the economy (Schniederjans & Hales, 2016) and the study of public spaces, both practically and academic, can take advantage of these new developments.

I suggest that such technology should be implemented as an early step in studying any public space that is slated for redesign in order to provide an additional and vital layer of information—how people end up doing and experiencing what Gehl called "life between buildings"—at relatively little cost. I demonstrate that, despite some limitations involving the lack of demographic data and some inaccuracies from the computer vision model, computer vision is a technology that can be applied on public space analysis and is effective at providing large-scale data. It should therefore be integrated into future background research of public spaces and information gathered via this technique should inform future design interventions.

2 Current Practice in Studying Public Space & Computer Vision

This chapter offers a general overview of current debates on three topics: works on people's behaviour in public spaces, works on how planners and policy makers approach designing public space, and works around computer vision on public environments. These fields inform this project from the standpoint of theory, practice, and technical expertise.

2.1 Studies of Public Spaces

Due to their importance in the life of people in cities, public spaces have gained the attention of the academic community. The theory behind the study and application of public space can be roughly divided into two camps: deductive and inductive. An example of how the former was codified is the work by Christopher Alexander and his colleagues (1977) where the theory, assumptions, and knowledge of public space design is tested and applied against real-world situations. The latter is the hallmark of William Whyte and Jan Gehl, whose investigations into how people behave in public spaces as they currently exist inform theory and practice today and are still considered enduring classics in scholarship. It is in the footsteps of their inductive approaches that this research was undertaken, coupled with an attempt to make the work easier via automation thanks to new computer advancements. Foremost among relevant older research are Whyte's 1980 work *The Social Life of Small Urban Spaces* and Gehl's 1971 *Life Between Buildings*, for which the 2011 edition was consulted. Both these pieces of research on public spaces particularly took note on how people experience public space and the ways they move through it. The play of interactions with the spaces for seating, people within the space, visibility, and movement all mesh to form the way a public space is used. Small elements such as changes in elevation, seating, noise, shadows, and edges were found to have distinct impacts on the ways people moved and stayed in the space.

The nature of a public space can be divided in several ways by studying its use. One way highlighted by Gehl (2011) is the differentiation between a space for staying and a space for walking (p.133). The former involves people sitting, standing, and generally lingering in the space while the latter makes up the dominant form of movement through or around a space. In plazas, it is common for many people to be lingerers while the activity of the movers acts as much of the entertainment for those lingering and people watching. Therefore, having the space for transit also ends up benefiting those who linger. The levels in which these people can interact can be modified with the placement and orientation of benches and other features in a space.

Methodologically, the manual tabulation of tracing research techniques is what is addressed in this research. Gehl & Svarre helpfully published How to Study Public Life in 2013, summarizing the techniques developed under his research during and since Life Between Buildings. Eight key methods are outlined: counting, mapping, tracing, tracking (shadowing), looking for traces, photographing, keeping a diary, and test walks. For this research, **counting, mapping, tracing, and tracking** are the most important, with tracing being especially relevant. Each involves observing individuals and how they interact and move through a space and the highlighted tracing method is where movement is registered by the researchers and a map is created of these patterns (Figure 1). Insights from this method include walking sequence, choice of direction, flow, and which entrances are used most and least.

Figure 1 - Tracing as demonstrated by Gehl in How to Study Public Life *(2013, p.28).*

Gehl & Svarre also highlighted problems associated with "Automated Registration Methods", noting that automation is especially helpful in processing larger amounts of data but that the technology is not very common and potentially expensive. It should be noted, however, that technological advancements have been occurring very rapidly—something discussed in the next section—and the book is currently a decade old, being published in 2013. On the effectiveness of automated registration methods, he also says it "must often be supplemented by a careful evaluation of the data collected, which can end up being as time-consuming as direct observation" (p.23). As such, the labour needed to check and validate the findings of the computer vision systems used in this research must be weighed against the labour savings potentially generated via the automation.

There are currently many techniques for tracking individuals through a space, each with their own drawbacks, which were helpfully listed by Hanzl & Ledwon (2017). The foremost is the classic direct and time lapse observation as described by Gehl with the corresponding high amount of labour required for analysis. Also described are video recording and analysis, of which the processing is the topic of this paper. These techniques are limited due to requiring post processing, being disturbed by occlusion, and the necessity for a top-down placement of the camera. Using GPS devices, GSM/UMTS phone positioning, Bluetooth, and WLAN network location has also been tested and is useful due to the prevalence of mobile devices. However, the latter three techniques cannot provide accurate data and GPS data, which can potentially provide metre accuracy, becomes less accurate in urban areas due to signal reflection off tall buildings. Wearable devices, Radio Frequency Identification (RFID), laser scanning, floor sensors, and 3D motion sensing have also been experimented with but are less applicable to larger outdoor installations.

Beyond these quantitative methods, qualitative methods such as field notes, direct semi-structured interviews, and focus groups are common methods for engaging the public in place-based research for planning and design (Silverman & Patterson, 2022) in addition to other ethnographic techniques (Taplin et al., 2005; Tate, 2023). Such methods provide detailed and participant centred information, allowing diverse publics who make use of a given environment to have voice, rather than just the researcher. They can thus provide empowerment and sensitivity that researcher-generated information can otherwise lack. In addition, such methods provide valuable narrative, context, and meaning to any quantitative information.

All these techniques can be used in order to gain useful information on public spaces and how people used them. By gaining this information, a professional who is tasked with redesigning a public space can encourage and nourish the interactions people seek—or at least experience—in such contexts. This can be anything from understanding the number of people who enter a square to much more in-depth information about the actual activities and personal experience people have within a space. Having this rich-picture data can enhance project implementation and outcomes by informing the need, type, and placement of interventions in order to better reach the goals of public space design (Calderon & Chelleri, 2013; Carmona, 2019; North, 2012).

2.2 Planning for Public Spaces in Practice

The importance of public spaces is well known in the planning profession and efforts to ensure new spaces are good and effective have been made in practice and theory. The aforementioned studies on public spaces outline some key principles on how planners should approach these spaces in order to make them conducive to effective community building and placemaking. Tools at the designer's arsenal include the placement of trees, benches, tables, greenery, and more. As a result, the guiding principles of urban space planning inform how such interventions should be used to react to public spaces as they are in the real world.

Resources such as *Squares: A Public Place Design Guide for Urbanists* (Childs, 2004), *People Places: Design Guidelines for Urban Open Space* (Marcus & Francis, 1998), and the toolkits developed by the 501(c)(3) tax-exempt organization Project for Public Spaces (https://www.pps.org/) offer guidance on how to approach the design of these spaces for a practicing urban designer or planner. These references, along with recommendations made by Gehl and colleagues in aforementioned works, provide parallel directions into the way that public space ought to be designed from a prescriptive perspective. The result is a general consensus on what constitute good planning and process for the design of public spaces meaning suggesting such processes ought to be followed. These pieces highlight the benefits of community centred design in creating spaces that are centres of the community to attract a variety of people performing different tasks and activities throughout.

The principles and working methods expressed by such contributions have made their way into official city documents to act as policy surrounding the creation and recreation of public spaces within the locales. In the Canadian context, all major cities (Toronto, Montréal, Vancouver, Calgary, Edmonton, and Ottawa) have design documents moving towards best practices for public space for the downtown core. Of these, the City of Toronto's *Downtown Parks and Public Realm Plan* (2018), prepared by Gehl Studio, has the most detailed analysis, incorporating a detailed survey of the movements and types of people moving through a space. They feature five dimensions of study: (1) how many people use a space, (2) who uses the space, (3) where people move and stay, (4) what do people do, and (5) how long people stay. With a detailed survey of 16 locations, they were able to determine in depth information of people's activities in each public space along with demographic information on age and gender. This type of study provides valuable insight to tailor designs of public spaces to the people who are actively using the space, allowing a designer to bolster the existing environment based on empirical data.

Other cities have more gestural efforts towards their public space plans lacking the detail of Toronto's study and providing little direct guidance into the process which the governments would like planners and designers to follow. Vancouver's *Downtown Public Space Strategy* (City of Vancouver, 2020) highlights important principles for the city to follow when designing these spaces. Less empirical than Toronto's, they codify much of the same language that was previously mentioned in this paper. As guiding principles, the strategy has six: (1) For All, By All, (2) The "Right" Supply, (3) Design for People, (4) Day-long, Year-long, (5) Nature and Resilience, and (6) Connecting Places. To demonstrate the type of specificity featured in this strategy, The "Design for People" principle prescribes having ample "space to pause" by featuring benches every 100m and encouraging building setbacks along with other interventions (City of Vancouver, 2020, p. 60). Such concrete targets are made for each of the Strategy's six directions and applied to the entirety of the downtown area, though the Strategy does highlight areas of particular opportunity. Edmonton features a similar document in their Downtown Public Spaces Plan (City of Edmonton, 2020). However, the Edmonton plan is much more particular with where the interventions are to be placed and strategic directions being concrete goals such as acquiring 2 hectares of public land within the downtown area. Also including area-wide recommendations, the document outlines a specific set of goals to move forward with but has little in the way of recommendations towards specific design goals of the spaces themselves besides the broad goals of what the spaces should act like.

Other municipalities have less focused documents. Ottawa has the *Downtown Ottawa Urban Design Strategy* (City of Ottawa, 2004) and Calgary has the *Centre City Design Guidelines* (City of Calgary, 2015). Both these documents serve a similar purpose in outlining broad urban design goals for the downtown area, with sections highlighting similar goals to the other cities mentioned for public space. In Montréal, the topic of public space design is limited to a subsection of the p*lan d'urbanisme* in the section "High-Quality Architecture and Urban Landscapes" (Ville de Montréal, 2004). In addition to the familiar goals found in other cities, the plan calls for the "establishment of urban design guidelines". Should such a document exist, it was not publicly made available, but it can be presumed such a document would be similar to others on this exploratory list.

The principles as outlined in theory appear to be well integrated into the planning design principles of Canadian cities. However, aside from Toronto's plan, there is little attention brought into the process of data collection itself, relying on broader notions of principles. While this has the benefit of leaving plans for new and reworked public spaces to have their own process, it also leaves the process of designing new public spaces up to the designer's own interpretations and practice. Toronto's much more detailed plan gives a more concrete example of how to approach studying public space. However, the methods employed are labour intensive and not always reasonable for all plans and a highly prescriptive approach may limit the placemaking abilities of a design (Carmona, 2019). The openness of method allows a more varied approach, but no suggestion is made to attempt to explore the utilization of new technologies to better inform the process. This is but one example of how there appears to be opportunities for new technologies to support the study and design of public spaces in meaningful ways that have not yet been fully explored in planning practice, if only for the difficulties of scarce time and resources that practitioners often struggle with.

2.3 Computer Vision and Public Environments

Computer vision is the broad field of study that focuses on the using of computers to identify meaning from input images. For instance, identifying objects in an image or video, what they are, and tracking their movement through the frame. While such an activity is trivial for a human, computers only see the image or video as a grid of pixels without a relational understanding. Creating meaningful relations between these pixels in order to extract high-level information is the overall goal of computer vision.

The field is not necessarily new. Early computer vision started in the late 1960s with tests in using computers to distinguish between male and female faces (Arkad′ev & Braverman, 1967; Dobson, 2023). In recent years, the rapid acceleration of easily accessible computing power and new computational techniques has greatly transformed the field and allowed the technology to move into the mainstream. Recent advancements in the field incorporate deep learning (DL) featuring convolutional neural networks (CNNs) in order to classify images. This contrasts against traditional computer vision techniques that relied on the researcher to design the entirely of the system (O'Mahony et al., 2020). The traditional framework for analysis can be broken down into three steps: Image Acquisition, Feature Extraction, and Classification (Brunetti et al., 2018). Here a researcher builds the model for which the computer used to classify elements of an image. The recent advancement of CNNs in the field mean the researcher need not go through the effort of manual classification in the feature extraction step to create the image. Instead, a large selection of training data is collected consisting of thousands of classified images—for instance images of people labelled as such to the computer—and the CNN changes its pattern recognition system in order to be able to find the same patterns in a new image and recognize a particular labelled feature.

When it comes to public space, such methods incorporate techniques of multiple object tracking (MOT) on the space to track movement patterns. Such methods can be used for pedestrians, cars, bicycles, and others but most relevant to public squares is the tracking of pedestrians, technically referred to as "multiple pedestrian tracking" (MPT) (Sun et al., 2021). The field has expanded rapidly in recent years, with less than 100 papers being published in the year 2000 to over 600 published during 2016 tagged with Human Detection and Tracking (Brunetti et al., 2018). The tracking of pedestrians has been highlighted as particularly complex due to people being "non-rigid" objects making them good candidates for the study of MOT (Luo et al., 2021). As a result, there is a large amount of literature has been done in the computer science sphere on pedestrian tracking. Specific subsets of the technology such as pose estimation, action recognition, and behaviour analysis can be used on numerous different applications such as movement prediction (Yang et al., 2022), surveillance (Wilkowski et al., 2020), and human behaviour analysis (Hanzl & Ledwon, 2017).

Much of the literature is highly technical, attempting to use models and techniques to improve accuracy and consistency. For instance, researchers building the BoT-SORT MOT method particularly for pedestrian tracking built upon the pre-existing ByteTrack tracker by modifying the common use of discrete Kalman filers with custom equations that better captured the extremities of a walking person's feet and incorporating camera motion compensation (Aharon et al., 2022). Meanwhile, other researchers have investigated the use of clustering techniques to account for crowds (Stadler & Beyerer, 2021). This is only scraping the surface of technical computer science literature, even when limiting ourselves to works on pedestrian tracking. A vast field of literature exists for MOT that was not examined in this paper but provides much of the background technical framework that the sampling of works in this section is built upon which made the application of this technology in this setting possible.

As these studies are most interested in the success of the technology in tracking, the works often have a more transportation-based slant to it due to their interest in the motion of individuals. For instance, Yang et al. (2022) took the technology of computer vision in order to aid non-motorized users of six road intersections in Seattle. The researchers were successful in using the You Only Look Once library to identify users and OpenPose to identify the movement of the individual parts of the users' body to predict movement. The output of the analysis is part of a proposed "Vision Enhanced Non-Motorized Users Services (VENUS) Smart Node" which can directly be input into the traffic signal system or relayed to other municipal services such as emergency services and maintenance departments. Lakmali et al. (2020) specifically look into using movement tracking to help plan for the smooth event planning and crowd management.

Nonetheless, the technology has been used on topics more related to urban design such as performing landscape analysis (Tang et al., 2020) and human behaviour analysis (Hanzl & Ledwon, 2017; Lakmali et al., 2020). The last topic is the method that is to be investigated in the following analysis section to determine its viability in the planning practice.

There is a history of learning about public spaces via a deductive and empirical approach. However, when it comes to applying such an approach to real-world implementations, there is much less in the way of direct guidance and prescription towards actually using these techniques on a regular basis. In a different ballpark, the work of computer scientists has created technology that is able to replicate many of the techniques outlined in previous studies, allowing for the low-cost implementation of these techniques. Combining these fields of literature can allow the urban design practice to take advantage of this new technology to aid in informing practitioners who are studying a space.

3 Methodology

The primary aim of this study is to investigate the effectiveness of using computer vision technology on replicating manual methods. As such, much of the research was devoted into applying this automatic method to a particular test case. This section will describe the nuances of the square that this technology was tested on in this study as well as how it was implemented and compared against an implementation of the manual tracing method on the same square.

3.1 Site Selection and Data Collection

For the sake of this test, a small square in front of McGill's Otto Maass Chemistry Building was selected as the primary test location. The square was selected due to is small scale, frequent visitation, variety of exit points, and good vantage point provided from Burnside Hall. The square is largely enclosed by the Maass building to the north and east, Burnside to the west, and a walled-off loading dock to the northwest. The result is a square with clearly defined hard edges and egress points. There are three entrance passages to the square, one at each corner except at the east. The north passage is the most easterly entrance to the main campus area and is the nearest access point to the Maass and Burnside buildings from the Metro system. Access via this

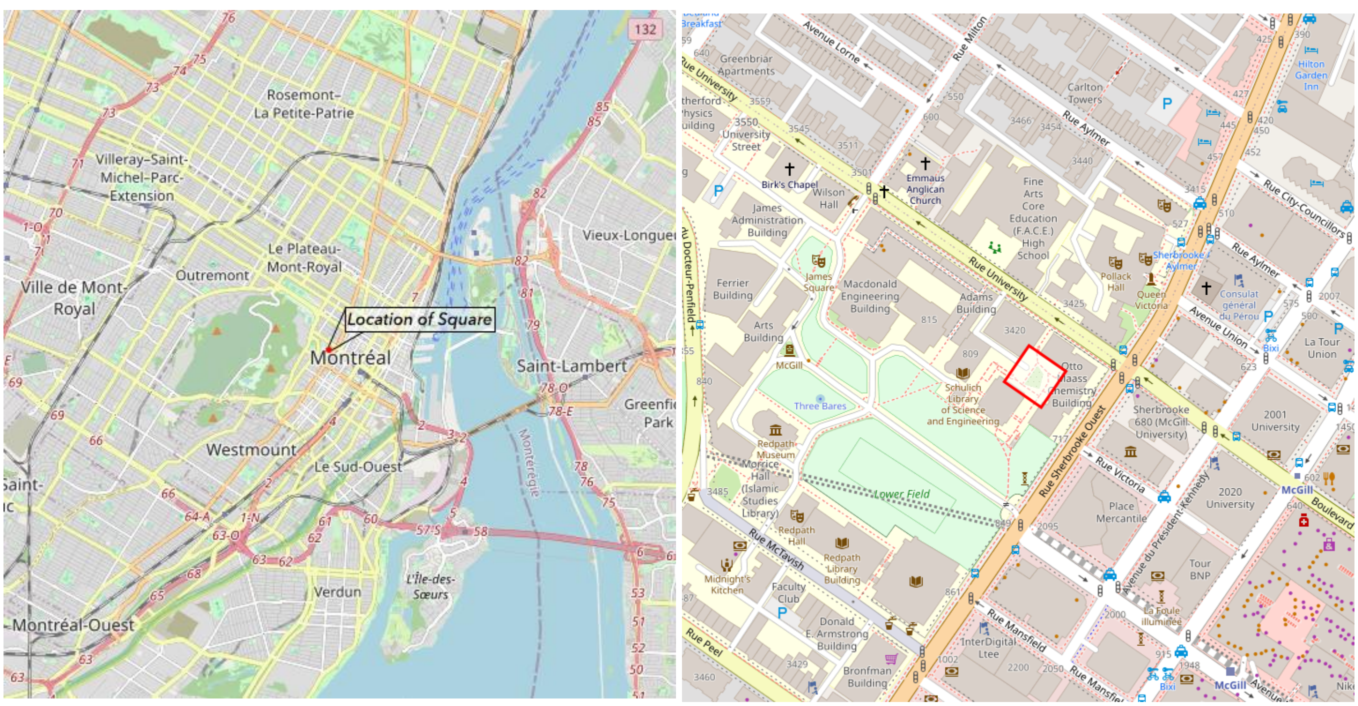


Figure 2 - Maps showing the location of the square.
Left: Map showing location of the square in the Montréal.
Right: Zoom in on the immediate area around the square.
Source: OpenStreetMap

entrance is provided via a set of stairs while the other two egress points have ramps providing step-free access. Towards the eastern corner of the square is the main entrance to the Maass building. The square itself is about half paved, with the remainder open grass and is about 925m². The square is largely isolated and closed off from its surrounding environments by the large university buildings that surround it. As a result, it is relatively invisible from the street and other parts of the campus (Figure 7).

The area around the square has an array of institutional, office, and commercial uses with most of the nearest buildings being related to McGill University itself, a university of nearly 40,000 students (McGill University, 2022). However, data collection was done in the summer

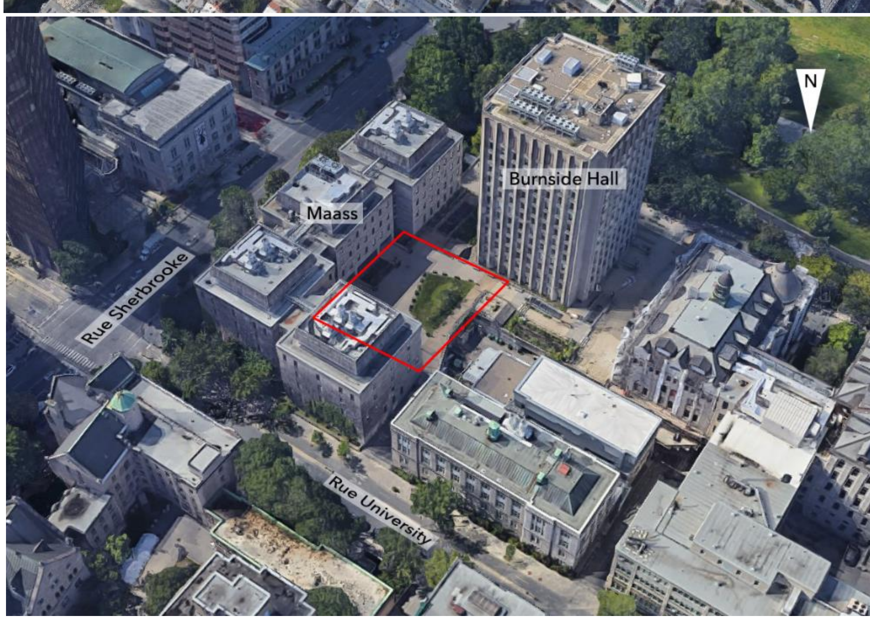


Figure 4 - Satellite imagery of the area around the square with nearby Metro entrances marked. Source: Google Maps

Figure 4 – Satellite imagery of the area at a lower angle, looking south.

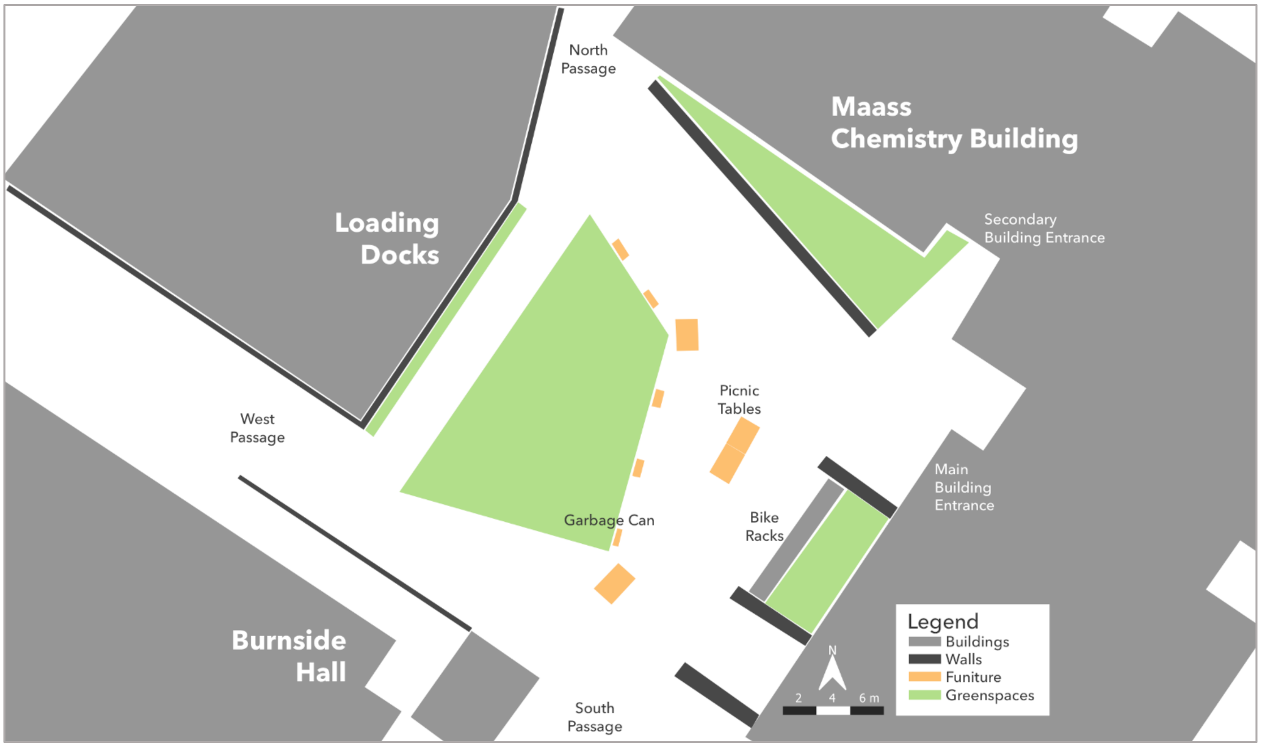


Figure 6 - Diagram of the square with labels for important features.

when classes were not in order and as such much of the student body was not on campus. McGill is a frequent tourist destination for both prospective students and visitors to the city, but this square itself is unlikely to draw much activity compared to other sites throughout campus as the entrances are modest compared to the main gates. However, the occasional tour group did walk through the square which allowed for the test of the system of large groups of people (Figure 25). More broadly, the site is situated within the downtown of Montréal, with numerous tall office buildings. The main McGill campus is one of the largest greenspaces in the immediate area, though Mont Royal Park is within walking distance. Nonetheless, the main campus's greenspaces

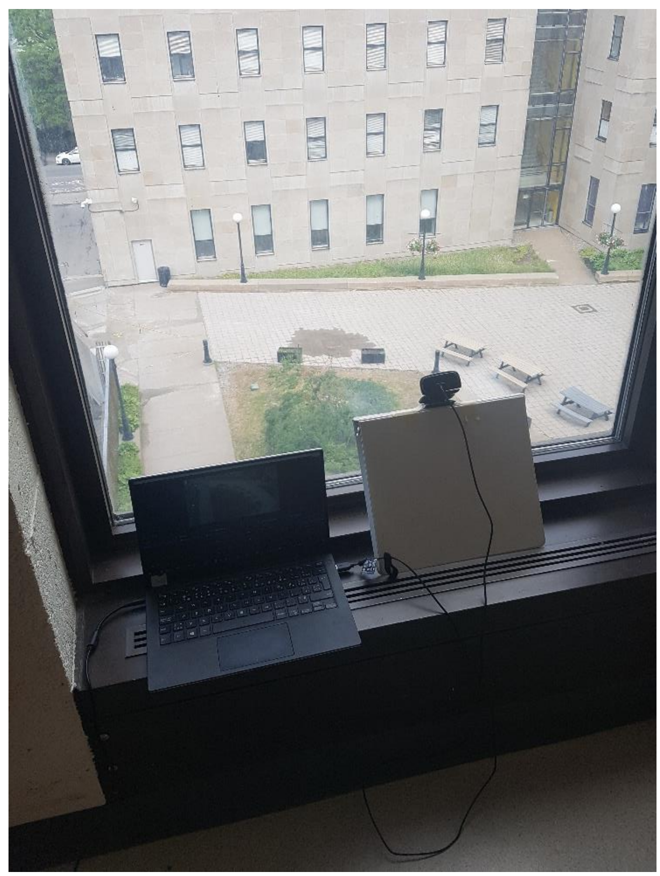


Figure 5 - Camera setup overlooking the square.

host many events during the school year. Though none were occurring at the time of data collection, the campus remains relatively lively, with many people using the green space for socialisation and casual park games. This main green area is on the other side of Burnside Hall, however, and the Hall acts as a distinct visual and physical barrier between the square under examination and the main campus.

A camera was set up on the fifth floor of Burnside Hall overlooking the square and set to record video over the course of several hours on different days in order to get a variety of conditions (Figure 6). Partial tests were run on a variety of shorter recordings and time lapse footage in order to refine the process of analysis. The full analysis was done on a recording on August 7th, 2023, a Monday, from 11:30am to 4:30pm, a period of five hours. In this test, the camera was a Logitech B525 webcam recording at 720p at 30 frames per second.

During the recording period, conditions were mild and pleasant for being outside. Temperatures were recorded at the nearby McTavish weather station to be at around 21 degrees, just cresting at 22 degrees at 1pm, with partly cloudy coverage and a relative humidity of 68% increasing to 80% through the recording (Environment and Climate Change Canada, 2023).

Figure 7 – Views of the square:
Top Left: Looking east at north passage stairs from University Ave. Top Right: Top of stairs at the north passage.
Bottom Left: From Otto Maass Building entrance. Bottom Right: Southern passage.

Mild wind speeds of 7 to 8 km/h were also recorded with no rain. These are relatively cooler conditions than what has been recorded in the past five years, with the previous five years having an average high of 28.1 degrees on the same day. As a result, the conditions that were recorded were more pleasant to be outside than is typical of the day and provide an ideal case for people to be using the outdoors as a place of leisure. The date was also representative of typical summer conditions on the campus itself. No special events were occurring at the time and start-off-semester activities had yet to return. Compared to previous data collection sessions, this data provided the ideal combination of activity, clear footage, and long recording time.

3.2 Computer Vision Registration & Processing

The You Only Look Once (YOLO) library was used to identify figures crossing and lingering in the square. No custom training was undertaken, and the included version 8 model was used. The model proved successful at identifying figures, however a low confidence threshold of 0.05 was required to be set in order to get continuous readings. Otherwise, people would not be consistently detected through the square, with numerous dropouts. As the square is strictly pedestrian only, the process was limited to identifying only people. Partly, this was because the model had no need to detect other classes, but it also rectified the constant misidentification of other objects, such as identifying picnic benches as cars. This was likely exacerbated by the low confidence threshold set and there is no system in YOLO to set different thresholds for different object classes (person, car, bicycle etc.) on the same run. As there was no need to track other objects during this test, different runs tracking different objects was not undertaken.

The basic object detection output was saved in CSV format featuring the x and y coordinates of the tracked person in the image with a low 0.05 confidence threshold used as a successful detection of a person as the confidence of the model was high variable. This was done in a time lapse mode of 2 fps fast forward in order to save on frames analysed. This form of data provides a momentary snapshot at each frame—in this case each half second—but does not provide any data linking points together, which is the purview of the tracker method.

Tracking was done via the BOT-SORT tracker, which has been proven to be successful in previous research (Aharon et al., 2022). Once again, the low and highly variable confidence of the object detection became an issue and low tracking thresholds of 0.005 and a high match threshold of 0.99 was used to gain some consistency in the tracks. Relatedly, the video had to be processed at the full frame rate, and time-lapse video proved unsuccessful in being used for tracking. During time-lapse video, there is not enough overlap or consistency between frames for the tracker to be able to identify related objects to create a trace. As such, the process had to be done essentially in real time. However, in order to save processing speed in later transformations, every detected point in only every third frame was saved into the output file. Care was taken to ensure the bottom of the identified box was saved as this was the location of the person's feet. It should also be noted that YOLO follows a XY coordinate system with a top-left origin. This is standard in computer images but is not typical to other fields. However, the unusual coordinates have little bearing as they are replaced with georeferenced coordinates at a later stage.

The points and traces were imported into QGIS with the latter converted to paths using the inbuilt point-to-path tool using the saved IDs of each detected object to connect the points. Both were georeferenced with the inbuilt georeferencer tool using the projective transformation. Due to the limitations of the georeferencer only being able to view on file at a time, it was helpful to create vector line boxes around key georeferenced points such as corners or lampposts, so they are easily spotted in the georeferencer. These were later removed to maintain the original data.

3.3 Manual Tracing

Following methods outlined by Gehl & Svarre (2013), manual tracing methods were undertaken for a half hour as a control and comparison against the computer vision technology. Video was able to be done via time-lapse and the same two frames per second time lapse speed was used for point data as the motion of individuals is still obvious to the human observer. While tracing methods were classically done on paper, this manual tracing was done digitally by directly drawing lines in QGIS. As a result, each individual's path was traced as their own unique line that approximated the path of movement through the square. Lines were drawn directly onto a geolocated square and therefore no transformations were necessary to geolocate the lines.

4 Results & Analysis

The points and paths gained via this processing method proved able to gather in depth information with relatively little human input. Aside from the camera setup, loading of files into the process, and geolocation, the process runs on its own. Over a 3.5-hour period, the tracker was able to trace 450 people through the square. This provided detailed and relatively accurate information on individuals that moved through the square but more importantly showed general behaviour trends. About 42,000 points were detected over 5 hours, which was spread out over about 35,000 frames at 2 fps.

The time discrepancy between the path and point data was due to the unreliability of the computer used for processing, which would unpredictably crash. Due to the long length of the processing time needed for the high number of processed frames required for the tracker to be successful, the computer would frequently crash while processing. It should be stressed, however, that this is not related to the software being tested here, but issues related to the machine's hardware; these issues would not occur on a more reliable machine.

4.1 Urban Design Findings

As a demonstration of the efficacy of this system, I was able to use the data that was gained from the computer vision process to identify some key points about how people treat this small square. It should be noted that these findings are limited to solely using the data gained from the computer vision process and that utilizing a mixture of other methods should be done when doing a complete analysis of a public space and to confirm findings highlighted by this automatic process.

The foremost finding is that few people linger in the space. This is not characteristic of the campus itself, as there is a very lively field on the other side of Burnside, but rather the nature of this square. The square is mostly used as a space of transit between the campus entrance and other destinations further into the campus, with roughly 12.5% of all transit beginnings and ends originating or heading to the east, 33.6% to the south, 39% to the north, and 14.8% into the Otto Maass building itself. Average linger time was 124.5 seconds, about 2 minutes, while median linger time was detected at 42 seconds. The longest linger time detected was 77 minutes. This information would indicate that there are a large amount of people spending little time in the square while certain outliers spend considerable time lingering. It should be noted, however, that should the tracking and detection model only begin tracking an individual near the end of their time in the square or lose track of an individual, the linger time would be artificially short. As the dataset has numerous traces that end and start at unexpected locations in the middle of the square, the average and especially the median may be artificially low. Among the traces that were recorded to linger more than 10 minutes, almost all seemed to stay at picnic benches, with only one person spending time at the short wall on the southern side (Figure 9).

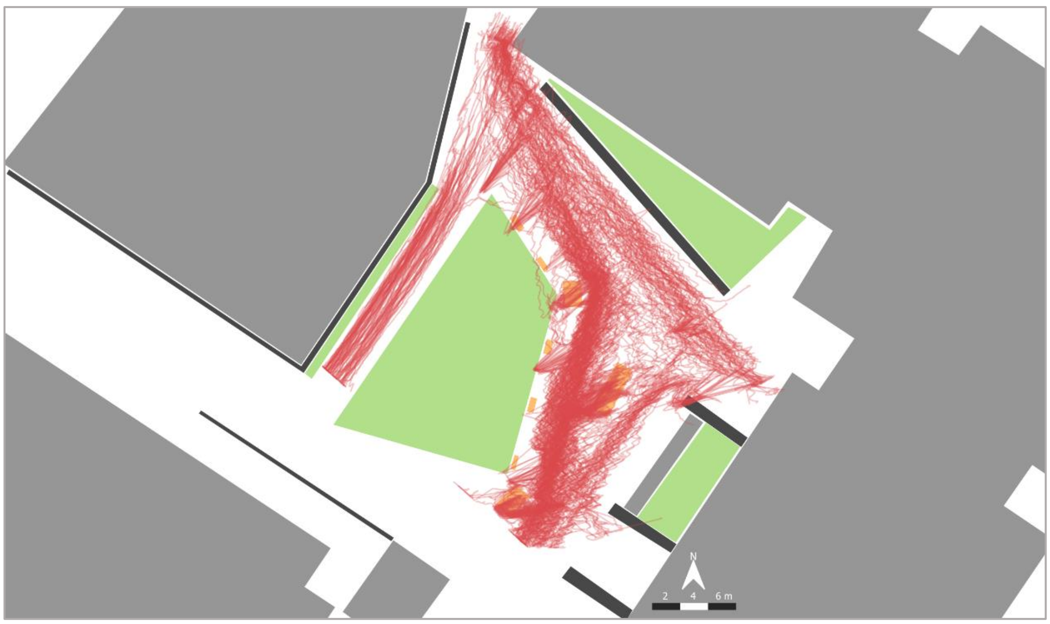


Figure 8 - Map of all detected traces during the test period.

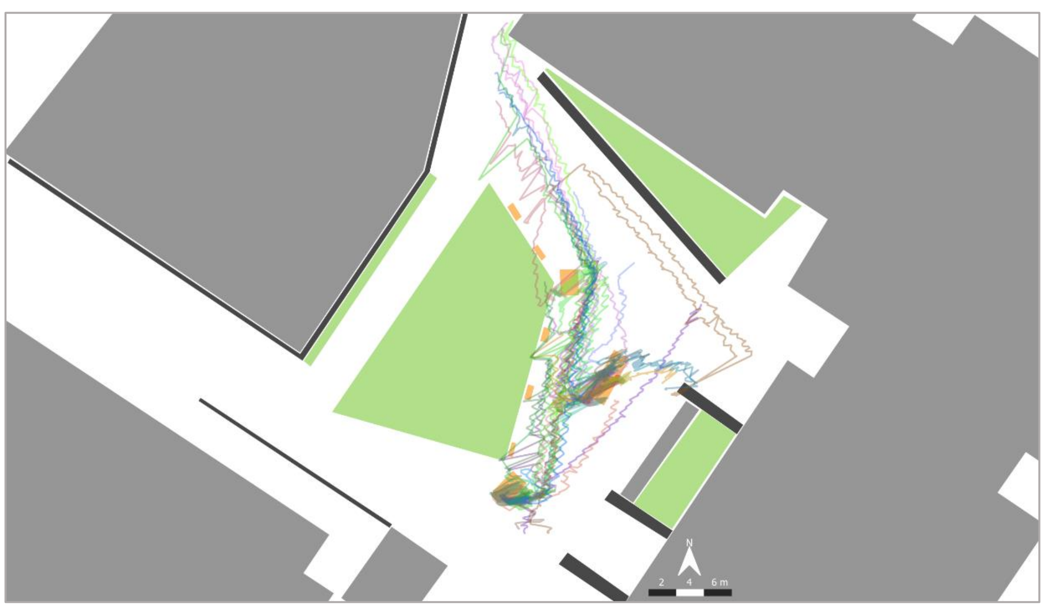


Figure 9 - All traces measured where individuals lingered more than 10 minutes. Note the clustering at picnic benches.

The system is able to easily identify when people and where people linger most. The paths with 10-minute or longer linger times informs us that people gathered for lunch at the picnic tables. This gathering also occurred most at specific times. During the recording period, by far the most popular hour for people to be in the square was from 1:30pm to 2:30pm, with over 20,000 detections. By 3:30pm to 4:30pm, only around 4,200 detections were found. It should be noted that the magnitude of these numbers is less important as they are the raw numbers of detections which include false positives and are multiplied by the frame rate. Rather, the large difference in detections clearly demonstrate the lunch rush. A similar spike was also reflected by the traces, with 155 traces between 12:30 and 1:30pm compared against 129 from 11:30-12:30 and 120 from 1:30 to 2:30. These numbers' absolute value has more bearing in this case, as the detections attempt to demonstrate the actual unique number of people who are present. As such, the computer vision model suggests that about 134 people per hour use the square during the mid-day, with about 15% more people using it from 12:30 to 1:30.

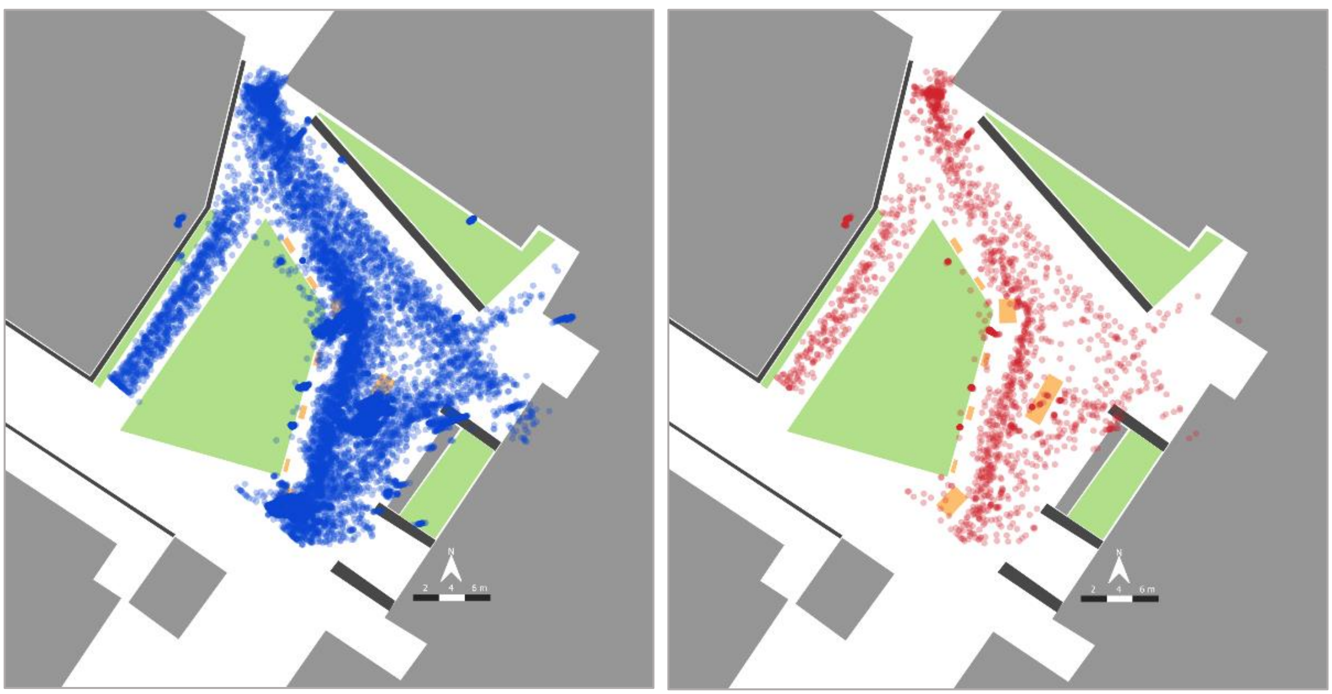


Figure 10 – Left: points from 1:30pm to 2:30pm. Right: points from 3:30pm to 4:30pm.

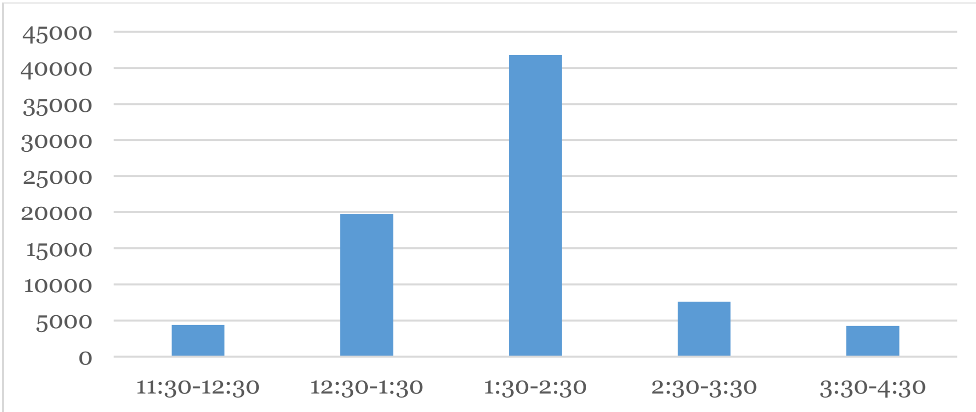


Figure 11 - Chart showing the counts of points per hour.
This gives a comparative record of how many people were in the square at each hour, but a less meaningful absolute value.

In this square, people quite ardently remain on the pavement. Only 4 instances of minor corner cutting were detected by the model, visible in the map of all traces (Figure 8), and no instances of deeper movement within the greenspace was detected. Nonetheless, people generally walked via the most direct path towards their destination. Although the automatic traces have a tendency to zigzag—an issue that will be discussed in next section—the general trend of straight paths remains visible. However, the most direct path will be deviated from to avoid particularly small gaps. The northern most picnic table leaves enough space for someone to walk between the grass and table. However, very few people did that, preferring to take the much more spacious route around the table.

This data informs us of the ways in which the square is used in its current for. Should the design be reconfigured, this current use ought to be taken into account. The narrow time frame of high traffic in conjunction with almost all lingering occurring at the picnic tables indicates that using the space for seating is the most common use of the space. As a result, should a design want to encourage the current use of the square, enhancing the seating experience would be beneficial, particularly from 12 to 3pm. However, the square should be kept open for the sake of travellers, as the vast majority of people use the square to move through, not to linger. For that, since most of the traffic starts or ends at either the north or south, ample space should be kept maintaining a clear path for people moving between those two entrances. Little occurs within the greenspaces, suggesting that activities related to the greenspace are not typically related to macro-scale observable activities. This does not discount less direct interaction, such as observation from afar, and experiential qualities the greenery may bring.

4.2 Comparison Against Manual Methods

By raw numbers, manual registration of the half hour between 11:30am and 12:00pm counted 54 people compared to the computer counting 51. In this time period, there did not seem to be cases of double counting, where the tracker loses track of an individual and creates a new one, which would result in a detection rate of 94%. This sounds highly successful, but there are large differences that are harder to measure. Many traces are shorter than the manually registered traces as the tracker was only able to start detecting the person part way through the square or lost track of them as they moved through. As a result, many traces are incomplete. In addition, there are two clear points near the northern point of the greenspace where many paths deviate unexpectedly. This was due to a bollard partly obscuring individuals. While this is not the bollard that created many false detections, the partial obscuring did create a much longer bounding box, dropping the saved point down in the frame significantly. A similar issue occurred for the other deviation, where a bushy planter would merge with partly obscured individuals. Upon transformation for geolocation, these errors were multiplied. In addition, the tracker may lose track of an individual moving through the square and re-register them later one, creating two counts when there was only one person. In a large data set, this would be difficult to differentiate between two separate people who were only tracked part the way through their path.

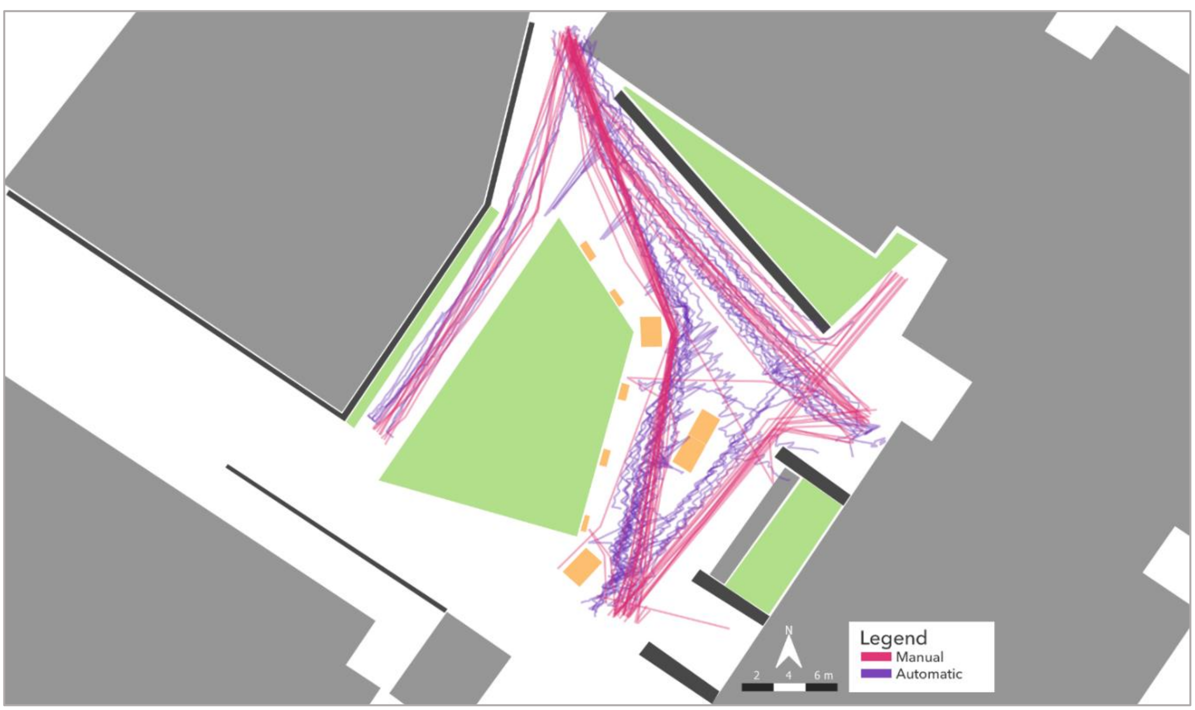


Figure 12 - Comparing automatic to manual tracing.
Note that despite manual tracing being neater, many of the same trends are observed.

Nonetheless, paths were relatively similar to the manually registered paths and created the same trends that are observed in the manual method.

There is an imprecision in the manual registration as well as it depends on the amount of detail the researcher desires to put down. The method involved estimating the location a person was in the square which has some relatively open areas, making placement difficult. The computer vision process, on the other hand, seems to promise a more precise transcription of an individual's movements as the frame-by-frame movement can be tracked. However, the paths are unstable, and tend to zigzag as the computer adjusts its bounding boxes. As a result, the two methods result in relatively similar information, being most useful for providing gestural indications of movement through the space.

4.3 Speed, Errors, and Methodological Findings

The average speed the analysis process ran at was 50ms per frame on an Intel i5-12400 CPU, the equivalent of 20 fps. The original test footage was recorded 30fps video meaning 33 1/3 ms per frame meaning the process was not able to run in real time on that configuration. Rather, 1 hour of footage at 30 fps took approximately 1.5 hours to process. Different frame rates were not tested for their efficacy in the tracing method. This speed is less of an issue for the point data as time lapse video is effective for that processing. For this test, point data was run on time-lapse video at 2 fps and a 50ms per frame speed was also observed meaning 1 hour of footage took only about 6 minutes to process. As such, depending on the amount of time resolution the researcher would like, the frame rate and therefore processing time, can be adjusted as they see fit. This allows for flexibility between the available resources a user of this technology has. A higher time resolution would provide more detail as to exactly when people were at particular times at the expense of longer processing times.

Figure 13 - This bollard was frequently detected as a person. Meanwhile, a person in a white t-shirt, likely blending into the light-coloured background, is not detected. The number is the confidence of the detection model, here 14%.

There are obvious outliers in the raw point data. Much of these came from false detections. Particularly, a bollard near the middle of the square was very frequently detected to be a person by the model (Figure 13). Lampposts created unusual detections as well, resulting in random clusters of points. Removing these outlying points is a matter of data cleaning and is necessary in order to create effective heatmaps of activity. While the removals of the datapoints far from the main areas of activity is trivial, the central bollard creates an issue as it is impossible to determine whether the detection is a false positive, a false positive and a detection, or a detection merged with the bollard. However, the stationary bollard creates a very dense cluster of points, and removing the cluster leaves the surrounding points still able to provide heat map information. These long-term false positives did not create erroneous traces, however, as the tracker was able to isolate these false detections and did not save them as tracked individuals.

Geolocating was not strictly necessary for determining how people use the space and findings can be found simply by analysis the lines and points over the original image. It does have the potential to be useful for standardization for comparing different camera angles, slight shifts in the same camera between different recording sessions, or comparing against pre-existing datasets.

5 Discussion

This section outlines the general findings of the research, from the information that can be gather via this method to the practicality and issues related to the technology at this time. Generally, this endeavour proved successful, being able to replicate many of the datapoints that can be gathered via manual methods. However, there are limitations such as the lack of data gathered on demographics and activities in additional to technical limits related to false and missed identifications. Limiting its widespread implementation, the setup process is rather technical in its current implementation, relying on some skill in computer coding a some understanding of how the computer vision system works. Lastly, limitations on camera location may serve to limit how much this method can be used for all sites.

5.1 Effectiveness & Information Gained Using Computer Vision Technology

Time information is a major asset of this technique. While technically possible using manual tracing methods, detailed information on *when* an individual moves across the space is gathered as a by-product of tracing. If such data is to be gathered manually, it increases the complexity of the registration process as timestamps would need to be kept by the registrant. In this automated process, such data are trivial to extract from the frame information as a conversion to a time stamp from the frame is a matter of simple unit conversion. In addition, the frame information is necessary to link points together into a path to create the traces. Aside from this additional information, the general findings are relatively similar between manual and automatic techniques. Even if the test was not exhaustive, a high level of accuracy was observed, and the same general trends could be deduced. Therefore, the automatic process was able to replicate much of the work done in the manual registration process at a lower amount of labour than otherwise would have been needed.

5.2 Limitations

There are several limitations that could be major blind spots to future implementations of this technology which cannot be solved. Primarily, this involves the nature of recording video from afar and limited data that can be collected observationally. By recording and observing, only elements of a person that is visible from a macro perspective is made available and the internal thoughts and identities of a person remains invisible. However, some issues may be the result of this technology only recently becoming available to the mainstream and still maturing along with its implementation with affordable, consumer-grade equipment in this test.

5.2.1 Demographics and Activities

The most glaring limitation of using pedestrian tracking technology for tracing techniques is the complete lack of demographic information. Such information is highly important for catering public spaces particularly towards the needs of the people who use it. For instance, in the detailed surveys done in Toronto's *Parks and Public Realm Plan*, the distribution of age and gender was logged. From this, informed if not necessarily definitive comments on the gender balance and age-friendliness of spaces could be made. Other demographic information such as income, immigrant status, language skills, sexual orientation, and others are impossible to gain via observation, and even the previous demographic information is left up to the judgement of the individual observer. As such, there is a broader limitation of observation on public space that is not limited to utilizing computer vision technology.

With a sufficiently high-quality camera, refinement of the computer vision model, and large amounts of training data, it may be possible to identify age and gender. Determining the gender of an individual as was demonstrated by even the oldest computer vision techniques (Arkad′ev & Braverman, 1967). However, having such a detailed record brings privacy concerns to the individuals. In addition, the technology will do little to address biases that are pre-existing within systems (Tommasi et al., 2015). It should be noted, however, that such bias also occurs in the human observer and that their assumptions on individuals that they observe are the ones that are being recorded. Much like how biases can be ameliorated in the human observer via training, the same can be done on a computer observer. The effectiveness of such a model remains to be seen.

In addition to personal details, the nature of an individual's visit is also absent from this type of analysis. For instance, people gathering to study and people gathering to eat lunch look the same in the data. This loses a major dimension of how a public space functions. As two examples related to points highlighted in the urban design findings of this study, why do people not cross the greenspace and what do people do while at the picnic table? For the former, this may be due to the large bush in the middle of the central greenspace making it impractical to cut across, which is visible in frame, but this is a speculation on myself, the researcher's, part and is not grounded in empirical data. For the second point, while recording I was able to observe people eating lunch, doing work, and waiting for people at the picnic tables. In addition, as mentioned previously, McGill is a tourist draw. Are the people moving through the square to visit or work? This information is lost in the analysis process, erased to simply be a timestamp noting that a person was indeed there. It would be comparatively easy for a person to go back through the footage to compare against, but such intervention means a relatively high amount of human labour needing to introduced into the method, tending towards the points made by Gehl & Svarre (2013) that automatic registration methods may require just as much human input to check the data. To test this, a more complete analysis featuring this method should be done and the usefulness of the data versus all other data should be compared against the amount of effort needed to set up and run the system.

5.2.2 Error in Tracking and Detection

Much of the literature around computer vision has discussed the error in the computer vision process and it remains an issue in this research. Most obvious are when the model loses track of an individual while they are moving. This is most common due to the inability of the frame-by-frame detection model to continue detecting the individual, and thus the tracker is unable to find any more connections.

While momentary breaks in traces are of little concern due to connecting points still typically making a good path, losing track of an individual may result in a path ending in an unexpected location and the creation of a new path that should be connected to a previous one. The result is a multiplication of the number of individuals that actually entered a space. It should be noted however, that this is partially cancelled out by individuals that were missed by the tracker. On the other hand, this is also a problem for manual tracking as the researcher will only be able to reasonably keep track of a limited number of people and should the space become very crowded it will be difficult to maintain consistency. The computer has none of these scaling issues and has the same ability to track any number of people, provided the view is clear.

Occlusion—the obscuring of part of the frame—is another obvious concern and partial occlusion can exacerbate the issue mentioned above. This test scenario is a largely ideal case, with few trees and little in the way of other built element obscuring the frame, but many public spaces are highly foliaged and could cause major occlusion issues for the camera. Indeed, all the plans and guidelines for public spaces consulted in the beginning of this paper highlight the need to increase urban greenery. As a result, having a significantly foliaged public space will only make the implementation of this type of study more complicated. Progress has been done in the computer science field to address partial occlusion (Brunetti et al., 2018), but little can be done if the occluding object completely covers the subjects of study.

Related to occlusion, tracked individuals may be dropped or missed altogether if they happen to blend into the background. In the test case, it was common for people wearing dark clothing blending into the dark asphalt or lamp posts, which lead to the detector failing to identify them against the background.

Even if the tracker is able to maintain a good track on a given individual through their entire path, there is a tendency for the tracked object's shape to shift as the detector adjusts the bounding box to best suit what it detects. As a result, the paths tend to "zigzag" through the frame. This zigzagging is exacerbated when transformed into the final map as the necessary skewing and stretching of the vertical axis to address the oblique angle of the camera amplifies any errors in the path. This may have additionally exacerbated in this implementation from the tracking of people from the bottom of the bounding box in order to get ground coordinates. Such zigzagging may be eliminated with smoothing functions or better tracking and detection models that provide more stability to the bounding boxes.

Lastly, there are issues around misidentifying objects. As mentioned in the previous chapter, a bollard in the middle of this test case provided a high amount of false positive detection and lampposts also would occasionally create false positives as well. In this test case, the model was limited to detecting people. However, as more objects are added to the model, more false positives could be detected. For instance, in early trial runs, with more classes of objects being included in the model, picnic benches were frequently misidentified as cars. One could also foresee particular features of a public square, such as statues, being difficult to isolate. In the case of the bollard, the bollard itself was not a gathering spot, so the high number of false detections is obvious. However, should a frequently false detection act as a gathering point, this separation of the false versus true detections could prove impossible. This would especially be a concern if it is an intermittent gathering point as the false detection would be covered by positive detections in a consistent crowd.

Many of these concerns can be ameliorated with improved algorithms and better training with the latter having the best opportunity for improvement in this context. This test case used the pretrained YOLOv8 model, which is a general detection model and is not tailored for this specific test case of identifying people from afar and from above. Therefore, training a new model particularly on images that fit into this type of image has the potential to vastly improve the accuracy of the results. Improved algorithms ma also be in the works, the BoT-SORT algorithm built into the YOLO library for tracking and used in this test was created only last year (Aharon et al., 2022) and incremental improvements continue to be produced.

5.2.3 Setup, Tuning & Running

Setup for the software and getting useful output, while not particularly advanced, requires some moderate knowledge of programming and computing. The models still remain largely in the domain of computer science, particularly in the open-source implementation that was used in this test. In addition, the tuning requires some trial and error in order to get the desired results. It should be noted that the tracking implementation used in this research was only packaged with YOLO in March of 2023 (this year). As such, the implementation is functional and works well, but refinement will occur in the future (Luo et al., 2021). These barriers limit the ability for the ordinary practicing planner or urban designer to use this tool. With basic GIS and data science coding skills becoming increasingly a common in the field (Wikle & Fagin, 2015), it would not be difficult for a practitioner to be able to use the models, but a cleaner, easier to use, packaged program would allow for the widespread usage of such methods.

The tracking process can be relatively slow. While the point information can be done on time lapse, and therefore higher speed, the tracking must be done at relatively close to live frame rate to have ample overlap for the tracker to be able to make the connection between multiple object detections. Low frame rates that are found in time lapse video were proven to be ineffective. Depending on computer speed, this can mean a processing speed well in excess of the original capture’s length. Such speed can be significantly improved on more powerful machines, however, and outsourcing the processing to high-speed processor services may be able to mitigate much of the speed concerns and allow for the processing to happen at faster than live speed. Another method of improving speed would be to add a layer of processing to remove frames that contain no movement, removing the need to run the heavy detection algorithm on those frames. This would only be effective on relatively quiet public spaces, however, and the speed improvements would depend on the speed of identifying still frames and the number of still frames present. In addition, there would be the potential for losing important frames where lingering individuals remain stationary enough for motion detection to fail to notice them. However, despite potential slowness, it should be noted that while the processing is occurring, no human input is required. As a result, the processing can occur overnight or while other more important work is being done. Therefore, the amount of human labour time is significantly lower that otherwise would be needed for manual methods. In addition, improvements to algorithm efficiency are likely occur in the future (Bhatt et al., 2021).

5.2.4 Physical Constraints

As highlighted by Hanzl and Ledwon (2017), this form of data collection relies on a top-down view of the space being studied. This does not need to be perfectly top down, this test was done with only a moderate top-down view, as the data can be corrected via geolocation transformations. However, the more vertical the viewing angle, the less errors would be amplified in the transformation process. The availability of high vantage points in urban areas is relatively high, and in the case of this urban test scenario the nearby Burnside Hall provided a good vantage point, but not all public spaces will have a point with enough elevation nearby or such points are not available to researchers. As a result, the physical barriers holding back a researcher from setting up a camera in a good vantage point may prove to limit the applicability of this system. Drone footage may act as an alternative, but as this type of study relies on long-term video, battery constraints and long-term disturbance by the drone means this may not be a useful alternative. More expensive would be the erection of a type of pole to mount and elevate the camera. Alternately, the presence of CCTV, should such footage be made available, may be another simple avenue for gaining input video.

The cameras used may also provide some challenges to the maximum scope of this type of study. For larger squares, the camera's field of view becomes a limitation to how much coverage it can have. This is partly related to the issue of gaining enough height to overlook the square but is also due to the amount of detail the camera is able to collect. In this test case a 720p was recorded and higher resolutions may be taken to see more detail. However, higher resolutions require more processing by the computer, so a compromise must be taken between the speed of the analysis and the amount of detail that needs to be gained for a useful output. This scope issue may be solved using multiple cameras and stitching the data together, though connecting the datasets may prove complicated as IDs between sets would be unlikely to match up. Alternately, a wide-angle lens may be used, but how effective the geolocation process would be at correcting the distorted output from a wide-angle lens would need to be tested. High contrast images also cause issues depending on camera. In this case, harsh shadows would occasionally make some parts of the image much too bright for anything to be visible in and the detector and tracker would fail to see anything in the over exposed sections. Higher quality cameras with better dynamic range may be effective at solving this issue or the data could be limited certain weather conditions or times of day, a less desirable solution. All the proposed solutions to these issues tend to inflate the equipment cost. This research was somewhat of a minimal implementation test, and affordable and basic equipment was used. More expensive equipment such as higher quality cameras could easily substitute the video capture process. These potential camera related physical constraints should also be tested to investigate the amount of hinderance they actually occur and whether the proposed solutions are necessary.

5.3 Refining and Applying the Method in the Future

Moving forward with computer vision technology on research, implementation into using this technology on real projects requires addressing some of the aforementioned issues both in the form of technical and applied barriers. Technical issues can be addressed via the improvement of the algorithms themselves or the tailoring of the technology specifically to the urban design use case. The former is largely in the domain of computer science research and, while not complete, has been continuing to make massive strides. The latter has more potential to be created by an invested individual or company and would with the creation of a new training dataset specifically made on public square information. The effectiveness of the system should also be tested against a much wider array of scenarios and in real projects where the method can be tested against the actual needs of a design team. Lastly, for the widespread adoption of the system, the setup should be simplified, wrapped up in an app or simple computer program to allow designers who are less technically minded to use the system easily and effectively.

5.4 Conclusion

The study of public space will allow for the creation of better-informed interventions and designs when such public spaces are slated for redesign and improvement (Gehl, 2011; Whyte, 1980). Having more concrete data would be beneficial but traditionally the resources were too limited to apply a big-data approach to this study. However, this paper has demonstrated that, albeit with some caveats, new advancements in computer vision technology can effectively eliminate some of these resource limits and a roughly equivalent study can be done on relatively little cost.

There are major limits with the most glaring being the lack of demographic and activity information that a computer would have great difficulty extracting. In addition, other constrains such as occlusion, errors, and physical limitations on where a camera can be places further limits on how this technology can be applied. Technical error should be left in the domain of computer science, where improve algorithms for pattern detection can be developed. On the urban design front, there needs to be a design to implement this technology. Somewhat between the two lies a space for the simplification and tailoring the model towards an urban design specific form. Packaging the system into an easier to use program and training a bespoke CNN model on public space data are two steps forward that are higher level than the much more technical development of improved models. In this way, some of the limits can be addressed and accounted for and the relative ease of the setup and process means high returns on relatively little investment.

What is missing currently is any widespread initiative to adopt such strategies in documents and guides that are meant to provide suggestions on how to approach this type of study. Most cities leave the method vague and up to the individual design firms while general guides do not suggest the use of newer technologies. There are valid concerns should the technology be widespread. Efforts must be made to ensure the privacy of individuals and their right to enjoy a public space. During the recording of this study, the low-quality camera was an asset in this case, with it being difficult to identify individuals at the resolution made available. In addition, it should be stressed that this data should absolutely not be taken as any sort of end point. Public spaces have deeper meanings than the simple movements through them. They have stories, meanings, and attachments that are invisible to the outside observer. As such, a proper community engagement where the actual users of the space's voices are the ones being heard and centred should be done.

Nonetheless, this type of analysis can act as an additional layer of data to inform decisions. This test demonstrates that advancements in computer vision technology has made Gehl & Svarre's (2013) claim that automatic methods for tracing are prohibitively expensive is false. Free and open-source libraries such as YOLO are now effective at replicating the manual registration techniques with a high amount of accuracy on low-cost equipment. As a result, this test proves the ability for computer vision to be a valuable asset to the study of public spaces in the future and that such a technology should be put into practice regularly. Much in the same way that cheap cameras allowed William Whyte and Jan Gehl to conduct their studies on public spaces, new technologies can allow another proliferation of this type of study on the public realm.

Works Cited

Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). *BoT-SORT: Robust Associations Multi-Pedestrian Tracking* (arXiv:2206.14651). arXiv. https://doi.org/10.48550/arXiv.2206.14651

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., Shlomo, A., & Center for Environmental Structure. (1977). *A pattern language: Towns, buildings, construction*. Oxford University Press; WorldCat.org. http://catdir.loc.gov/catdir/enhancements/fy1311/74022874-t.html

Arkad′ev, A. G., & Braverman, Ė. M. (1967). *Computers and pattern recognition* ([U.S. ed.].). Thompson Book Co.

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. *Electronics*, *10*(20). https://www.mdpi.com/2079-9292/10/20/2470

Brunetti, A., Buongiorno, D., Trotta, G. F., & Bevilacqua, V. (2018). Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. *Neurocomputing*, *300*, 17–33. https://doi.org/10.1016/j.neucom.2018.01.092

Calderon, C., & Chelleri, L. (2013). Social Processes in the Production of Public Spaces: Structuring Forces and Actors in the Renewal of a Deprived Neighbourhood in Barcelona. *Journal of Urban Design*, *18*(3), 409–428. https://doi.org/10.1080/13574809.2013.800449

Carmona, M. (2019). Principles for public space design, planning to do better. *URBAN DESIGN International*, *24*(1), 47–59. https://doi.org/10.1057/s41289-018-0070-3

Childs, M. C. (2004). *Squares: A public place design guide for urbanists* (1st ed.). University of New Mexico Press; WorldCat.org. http://catdir.loc.gov/catdir/toc/ecip0419/2004015733.html

City of Calgary. (2015). *Centre City Urban Design Guidelines*.

City of Edmonton. (2020). *Downtown Public Spaces Plan*.

City of Ottawa. (2004). *Downtown Ottawa Urban Design Strategy*.

City of Toronto. (2018). *Downtown Parks and Public Realm Plan: Public Space Public Life Study*.

City of Vancouver. (2020). *Downtown Public Space Strategy*.

Dobson, J. E. (2023). *The Birth of Computer Vision*. University of Minnesota Press. https://muse.jhu.edu/pub/23/monograph/book/103328

Environment and Climate Change Canada. (2023, August 7). *Hourly Data Report for August 07, 2023—Climate*. https://climate.weather.gc.ca/climate_data/hourly_data_e.html?hlyRange=1994-02-01%7C2023-08-22&dlyRange=1994-07-26%7C2023-08-22&mlyRange=%7C&StationID=10761&Prov=QC&urlExtension=_e.html&searchType=stnName&optLimit=yearRange&StartYear=1840&EndYear=2023&selRowPerPage=25&Line=0&searchMethod=contains&txtStationName=McTavish&timeframe=1&time=LST&time=LST&Year=2023&Month=8&Day=7#

Gehl, J. (2011). *Life between buildings: Using public space*. Island Press.

Gehl, J., & Svarre, B. (2013). *How to study public life* (1–1 online resource (xiii, 179 pages)). Island Press; WorldCat.org. https://doi.org/10.5822/978-1-61091-525-0

Hanzl, M., & Ledwon, S. (2017, October 10). *Analyses of human behaviour in public spaces*.

Lakmali, R. G., Genovese, P. V., & Abewardhana, D. (2020, December 30). *Pedestrian Movement Tracking and Tracing in Public Space*.

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T.-K. (2021). Multiple Object Tracking: A Literature Review. *Artificial Intelligence*, *293*, 103448. https://doi.org/10.1016/j.artint.2020.103448

Marcus, C. C., & Francis, C. (1998). *People places: Design guidelines for urban open space* (2nd ed.). Van Nostrand Reinhold; WorldCat.org.

McGill University. (2022). *2022 Quick Facts*. About McGill. https://www.mcgill.ca/about/quickfacts

North, A. (2012). *Operative Landscapes: Building Communities Through Public Space*. https://doi.org/10.1515/9783034610858

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. (2020). Deep Learning vs. Traditional Computer Vision. In K. Arai & S. Kapoor (Eds.), *Advances in Computer Vision* (Vol. 943, pp. 128–144). Springer International Publishing. https://doi.org/10.1007/978-3-030-17795-9_10

Schniederjans, D. G., & Hales, D. N. (2016). Cloud computing and its impact on economic and environmental performance: A transaction cost economics perspective. *Decision Support Systems*, *86*, 73–82. https://doi.org/10.1016/j.dss.2016.03.009

Silverman, R. M., & Patterson, K. L. (2022). *Qualitative research methods for community development* (Second edition., 1–1 online resource (xii, 145 pages) : illustrations). Routledge; WorldCat.org. https://www.taylorfrancis.com/books/9781003172925

Stadler, D., & Beyerer, J. (2021). Multi-Pedestrian Tracking with Clusters. *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–10. https://doi.org/10.1109/AVSS52988.2021.9663829

Sun, Z., Chen, J., Chao, L., Ruan, W., & Mukherjee, M. (2021). A Survey of Multiple Pedestrian Tracking Based on Tracking-by-Detection Framework. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(5), 1819–1833. https://doi.org/10.1109/TCSVT.2020.3009717

Tang, X., Zhang, L., Chen, Z., Wan, J., & Li, L. (2020). Urban Street Landscape Analysis Based on Street View Image Recognition. *2020 International Conference on Urban Engineering and Management Science (ICUEMS)*, 145–150. https://doi.org/10.1109/ICUEMS50872.2020.00040

Taplin, D. H., Scheld, S., & Low, S. M. (2005). Rapid Ethnographic Assessment in Urban Parks: A Case Study of Independence National Historical Park. *Human Organization*, *61*(1), 80–93. https://doi.org/10.17730/humo.61.1.6ayvl8t0aekf8vmy

Tate, L. E. (2023). Using Rapid Ethnography to Unpack Performances of Community Authenticity: An Art Festival Case from Victoria, British Columbia. *Journal of Planning Education and Research*, *43*(3), 653–665. https://doi.org/10.1177/0739456X20920922

Tommasi, T., Patricia, N., Caputo, B., & Tuytelaars, T. (2015). A Deeper Look at Dataset Bias. In J. Gall, P. Gehler, & B. Leibe (Eds.), *Pattern Recognition* (pp. 504–516). Springer International Publishing. https://doi.org/10.1007/978-3-319-24947-6_42

Ville de Montréal. (2004). *Plan d'urbanisme*.

Whyte, W. H. Jr., 1917-1999. (1980). *The social life of small urban spaces*. Conservation Foundation; WorldCat.org.

Wikle, T. A., & Fagin, T. D. (2015). Hard and Soft Skills in Preparing GIS Professionals: Comparing Perceptions of Employers and Educators. *Transactions in GIS*, *19*(5), 641–652. https://doi.org/10.1111/tgis.12126

Wilkowski, A., Stefańczyk, M., & Kasprzak, W. (2020). Training Data Extraction and Object Detection in Surveillance Scenario. *Sensors*, *20*(9), Article 9. https://doi.org/10.3390/s20092689

Yang, H., Ling, Y., Kopca, C., Ricord, S., & Wang, Y. (2022). Cooperative traffic signal assistance system for non-motorized users and disabilities empowered by computer vision and edge artificial intelligence. *Transportation Research Part C: Emerging Technologies*, *145*, 103896. https://doi.org/10.1016/j.trc.2022.103896

Appendix A – Additional Maps

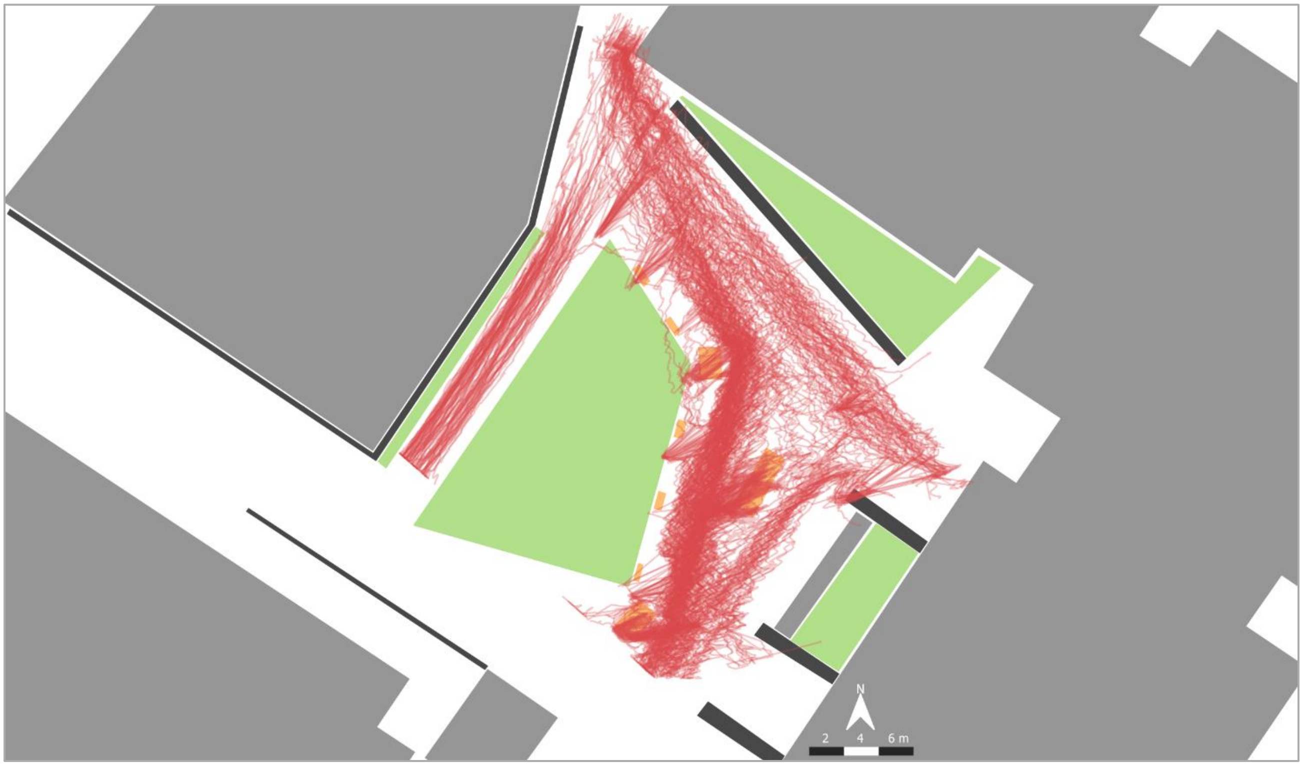


Figure 14 - All detected paths.

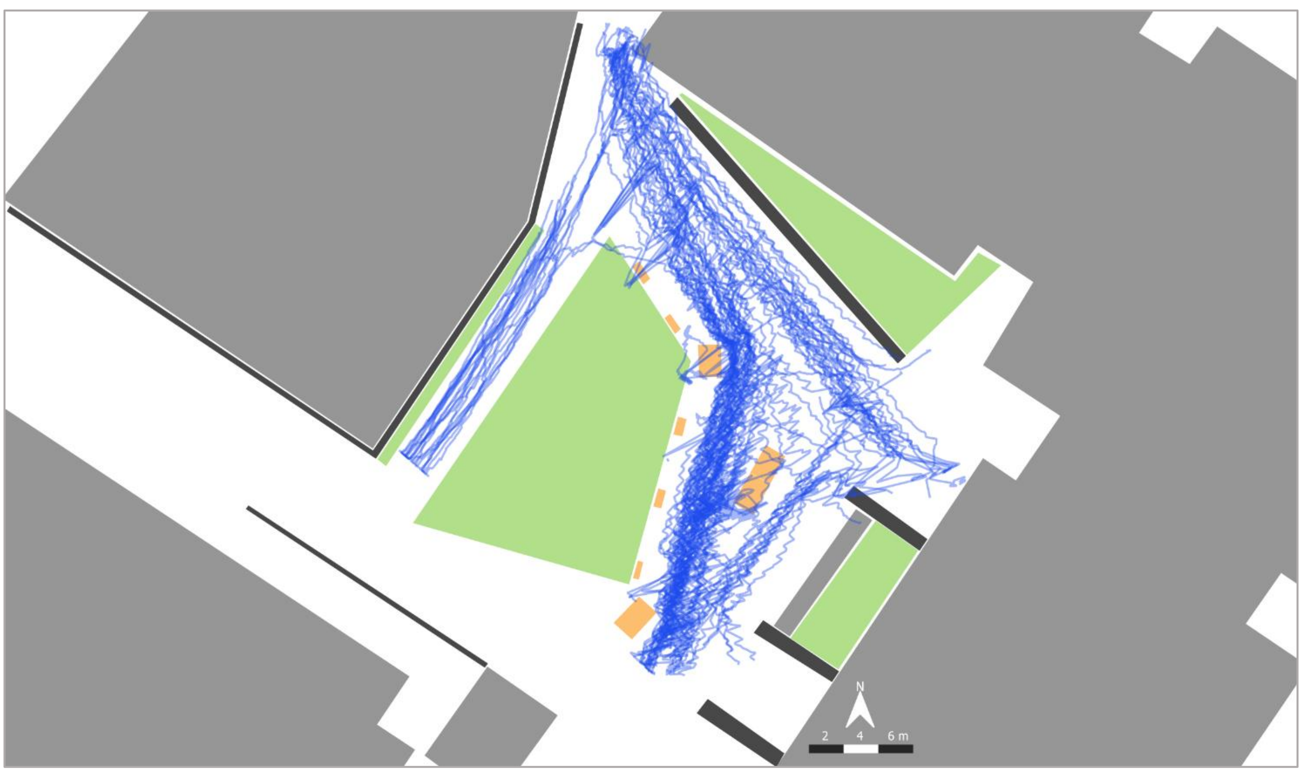


Figure 15 - Paths from 11:30 to 12:30pm

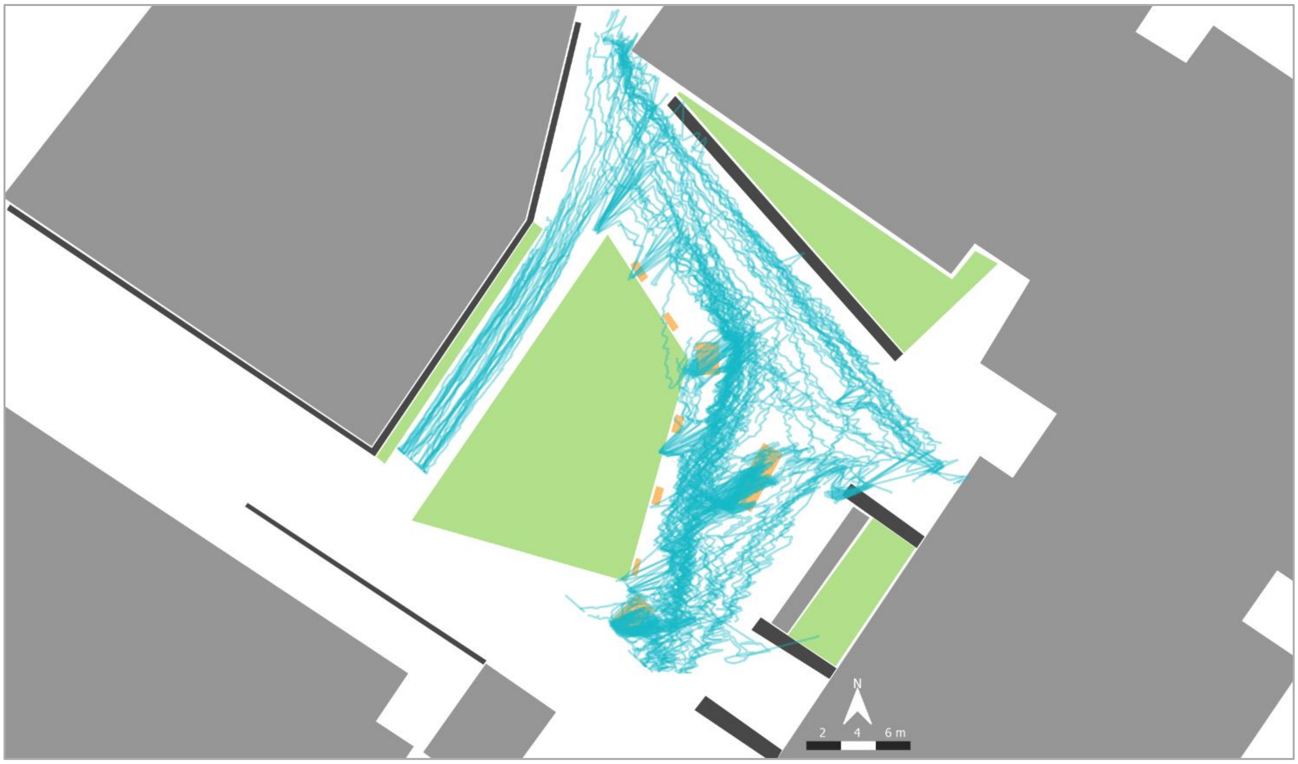


Figure 16 - Paths from 12:30 to 1:30pm

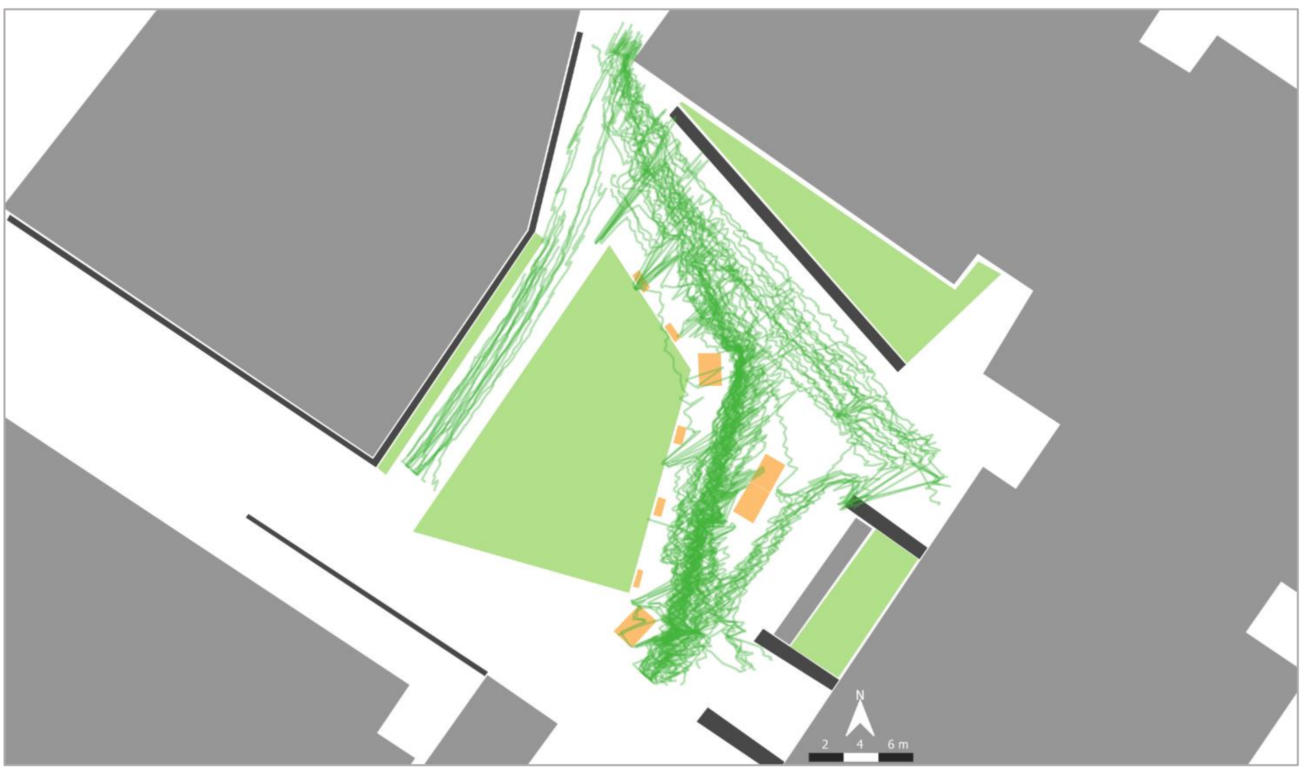


Figure 17 - Paths from 1:30 to 2:30pm

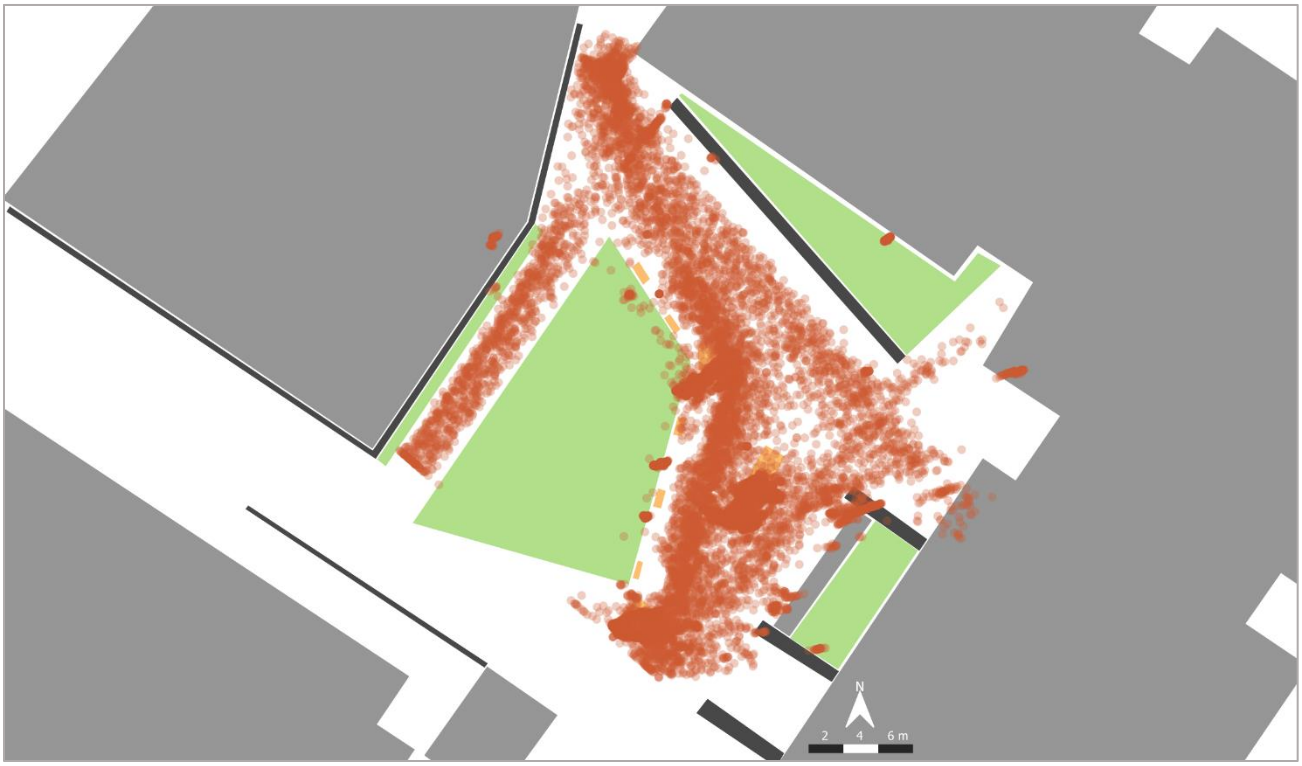


Figure 18 - All detected points.

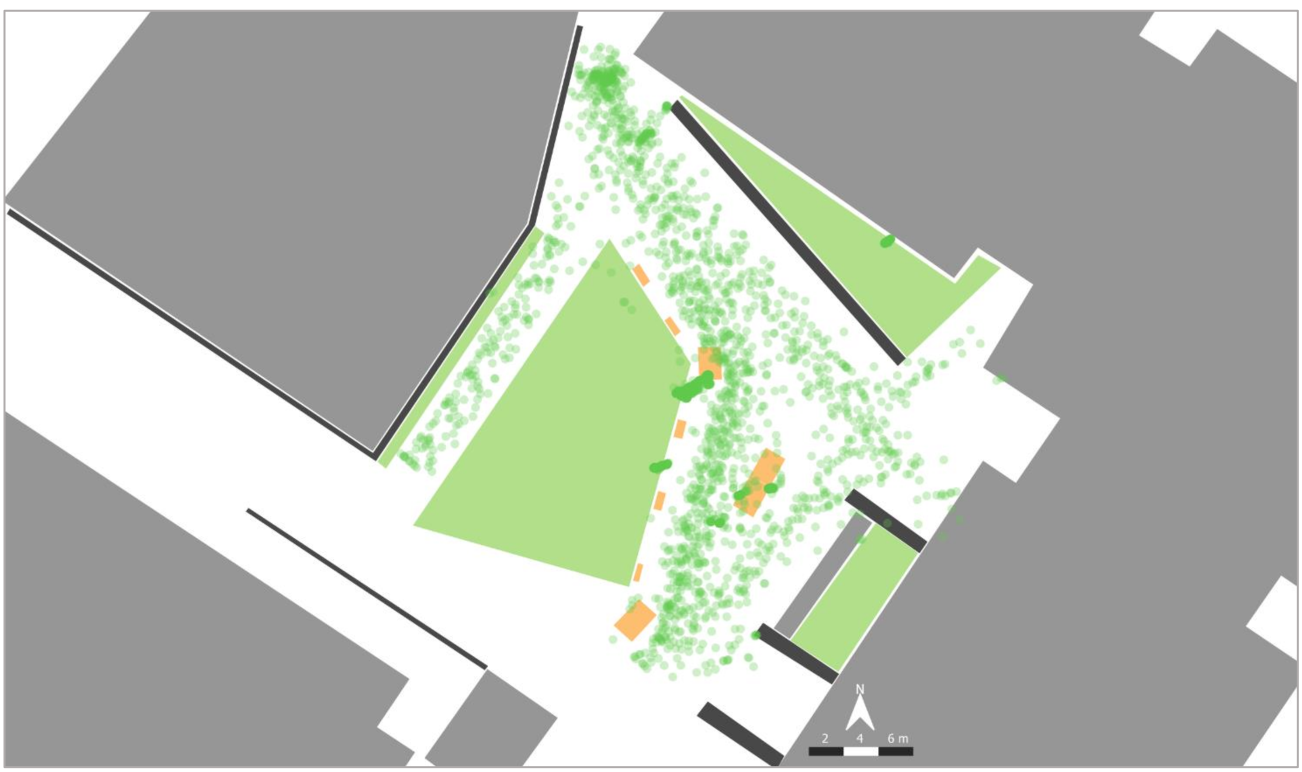


Figure 19 - Points from 11:30 to 12:30pm

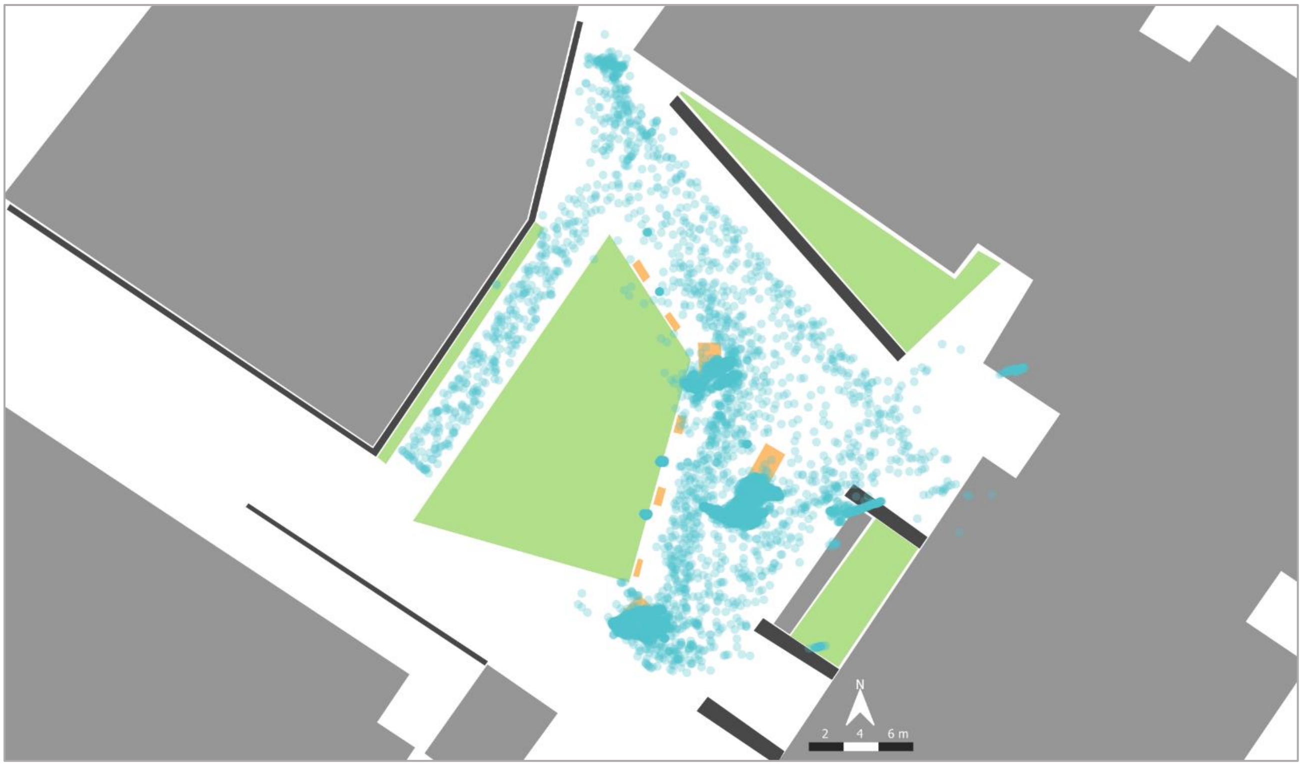


Figure 20 - Points from 12:30 to 1:30 pm

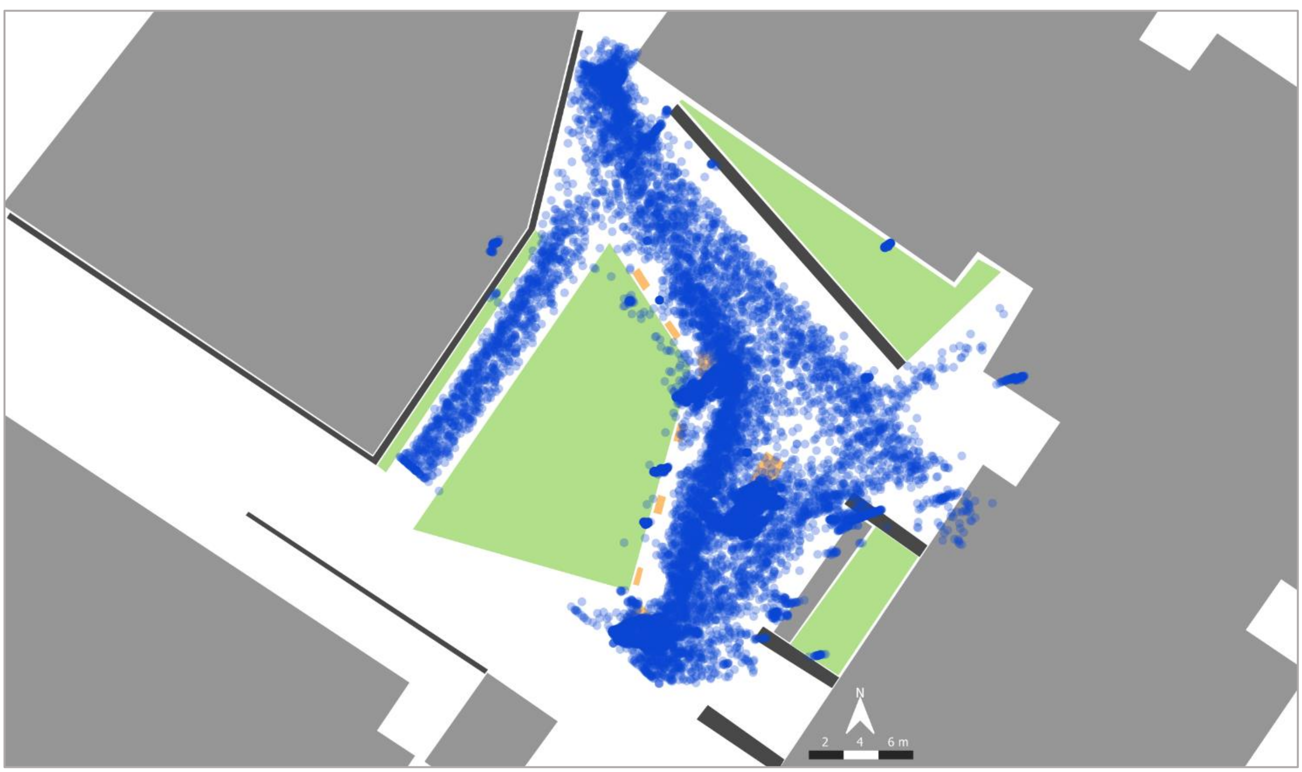


Figure 21 - Points from 1:30 to 2:30 pm

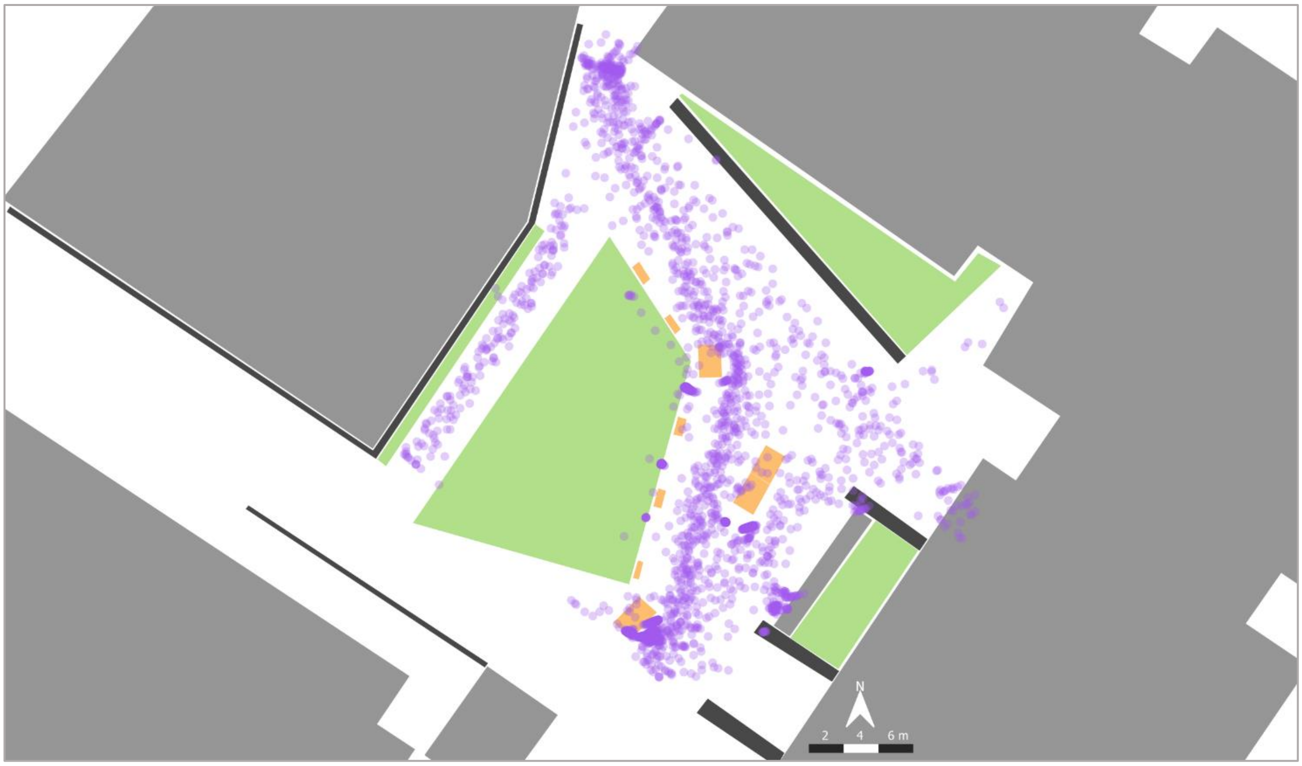


Figure 22 - Points from 2:30 to 3:30 pm

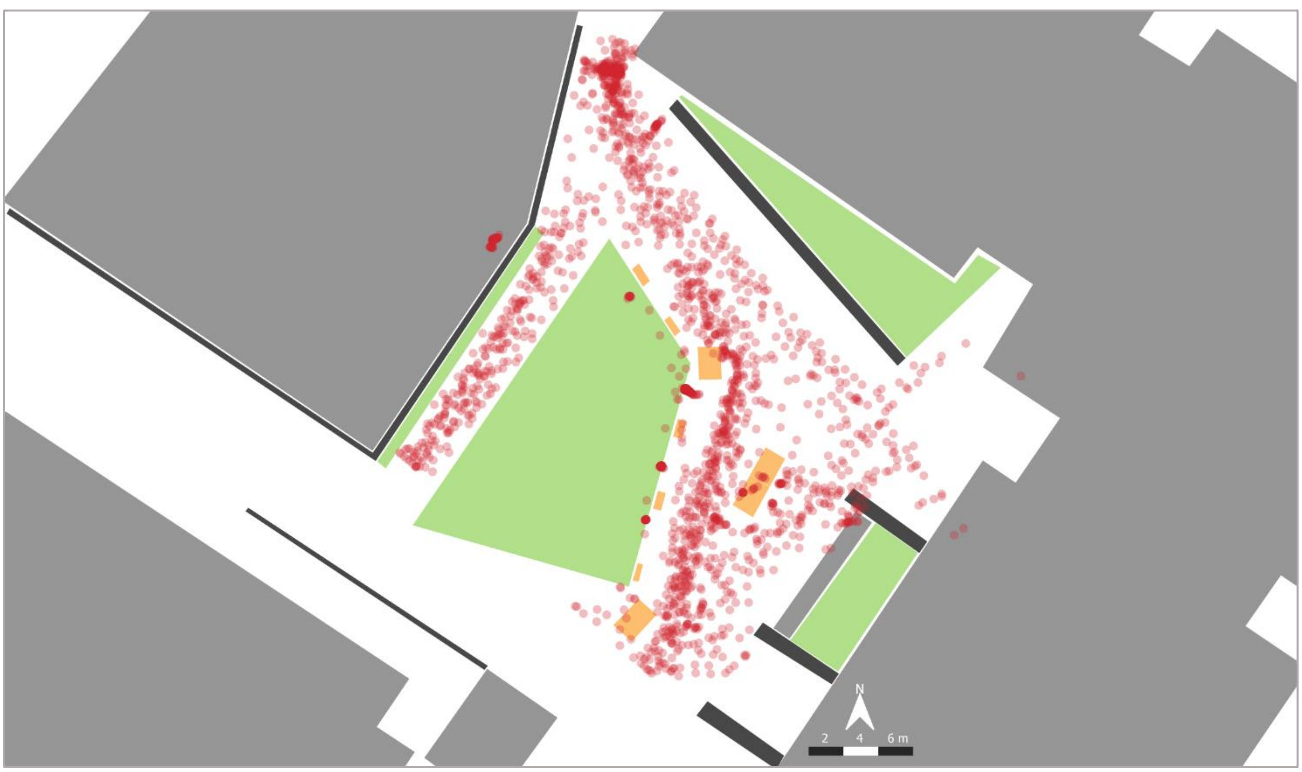


Figure 23 - Points from 3:30 to 4:30 pm

Appendix B – Additional Images

Figure 24 - Raw, non-georeferenced traces. Many findings can be deduced from this version, but georeferencing is useful for standardization and comparison with other datasets.

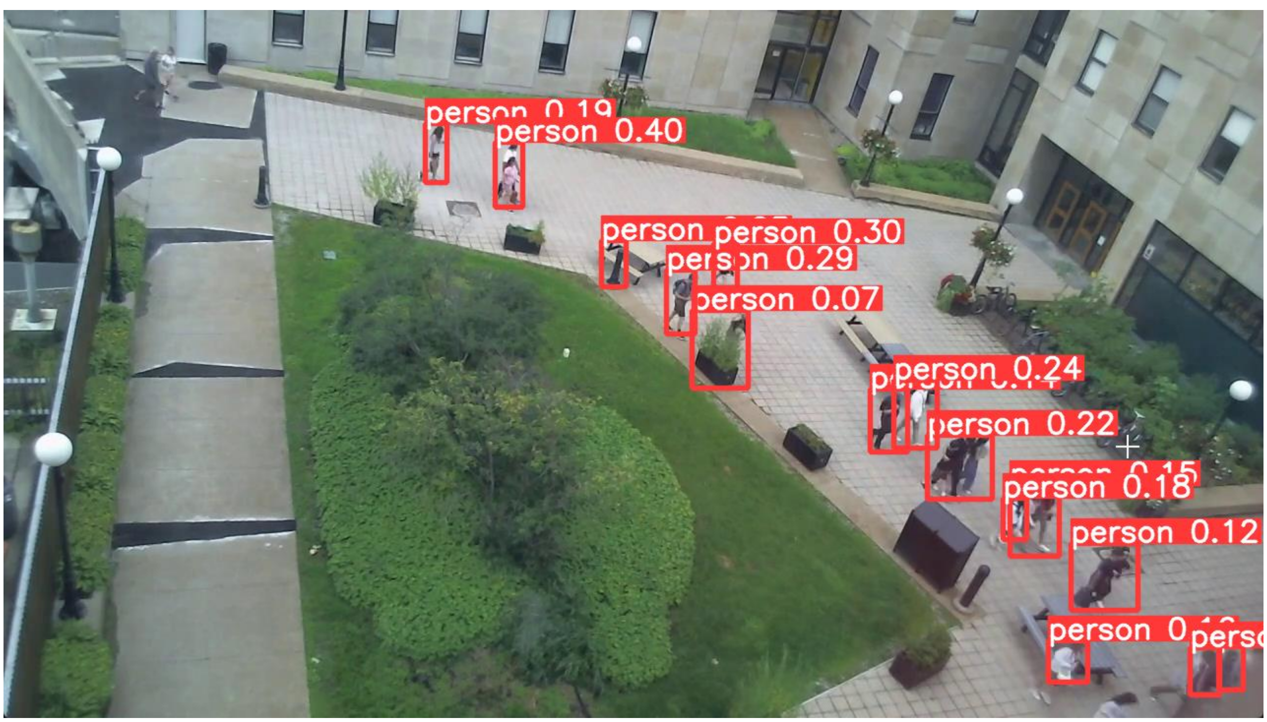


Figure 25 - The model is successful with large amounts of people as well. However, note the expanded bounding box around the planter. This is not a false detection but is inaccurate to where the person is. The second to last person was also detected as one when there are actually two.

Figure 26 - The model fails to detect and track two people, one with a board.

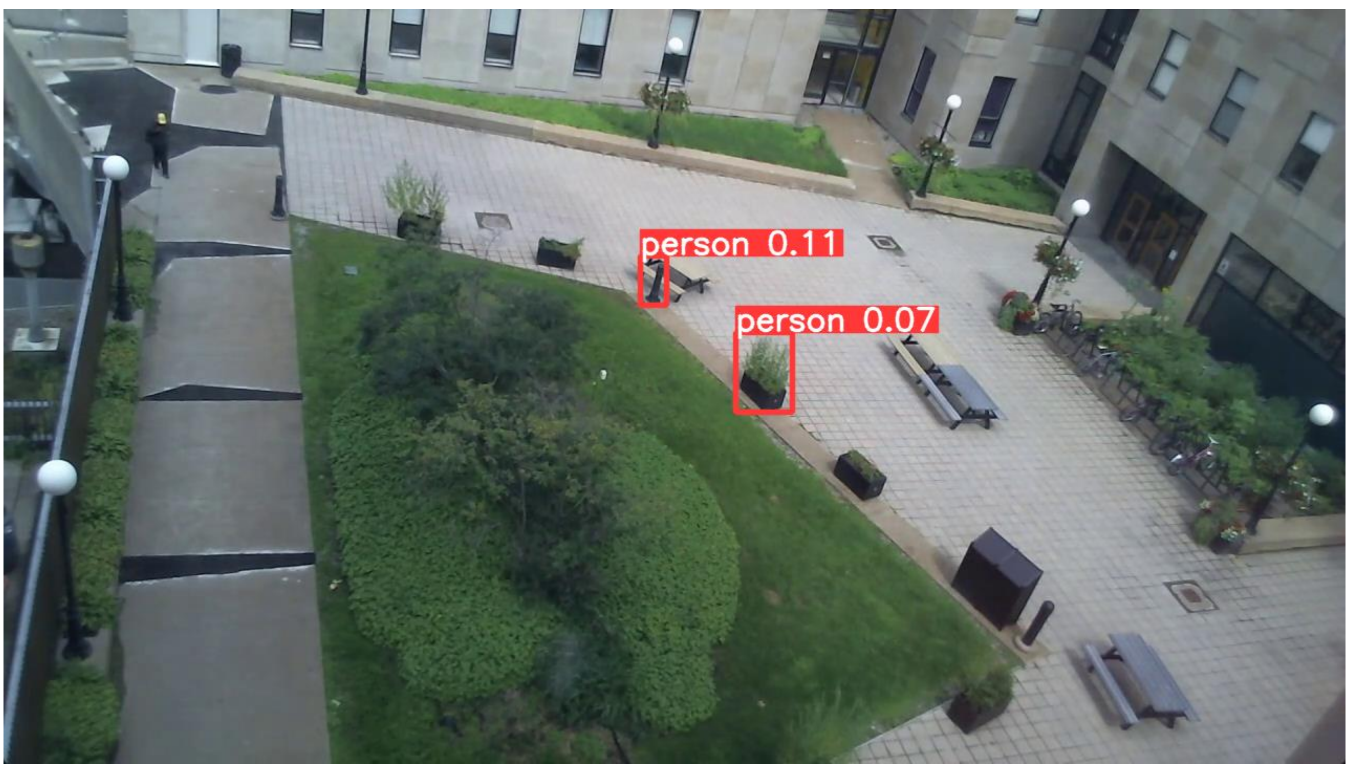


Figure 27 - The model thinks a bollard and plant are people but missed the person dressed in black blending into the asphalt.