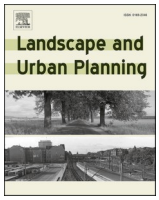
Landscape and Urban Planning 234 (2023) 104715

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan



Research Paper

Smart curbs: Measuring street activities in real-time using computer vision

Arianna Salazar-Miranda [a], Fan Zhang [b,*], Maoran Sun [a], Pietro Leoni [a], Fábio Duarte [a,c], Carlo Ratti [a]

[a] Senseable City Lab, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, United States
[b] Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology, Hong Kong, China
[c] Center for Real Estate, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

HIGHLIGHTS

- A framework to study street activity using a mobile device and computer vision.
- A deep learning-based model to detect various human mobility types.
- Constructing measures of street use to study their spatio-temporal dynamics.

ARTICLE INFO

Keyword:
Street activity
Pedestrians
Transportation modes
Image analysis
Computer vision

ABSTRACT

Streets are conduits of human activity. Despite their importance, studying street activity has been obscured by a lack of data on how people use them, with most approaches limited to studying a single point in time or small geographic areas. This paper proposes a new framework to measure street activity in real-time. Our framework leverages machine learning and computer vision to classify pedestrian activities and transportation modes using images collected from moving vehicles. We apply our methodology to measure street activity in Paris for five weeks. We produce activity maps for this period showing that streets vary dramatically in their capacity to support pedestrian activity and that these differences are highly persistent. Our proposed framework can be used to measure street activities in other contexts and cities, providing urban researchers with an approach to guide planning interventions, identify infrastructural deficiencies, and inform design policies that foster active streets.

1. Introduction

Streets comprise 80 percent of all public space in cities, serving as primary conduits for social and commercial encounters and exchanges.[1] Studying pedestrian activity in streets is a necessary and important step for better understanding urban life, the dynamics of commerce and the exchange of ideas, civic engagement, and the formation of social capital in cities.[2] From a practical perspective, measuring pedestrian activity in streets can also inform the evaluation of policy interventions such as changes in zoning and the construction of bike lanes and green spaces.

Traditional approaches to studying streets and public spaces have relied on personal interviews and direct observation, following the work of scholars in the field of urban studies (Lynch, 1960; Jacobs, 1961; Milgram, 1977; Whyte, 1980; Nasar, 1990; Gehl, 2011). Although these efforts provide a rich in-depth view of how spaces are used, they are labor-intensive and hard to scale, limiting them to single points in time or small geographic areas. More recently, access to cellphone location data has enabled scholars to examine human activity at multiple geographic scales—from streets to entire cities— and over time, circumventing the limitations of traditional approaches (Ratti, Frenchman, Pulselli, & Williams, 2006; González, Hidalgo, & Barabási, 2008; Jiang, Ferreira, & Gonzalez, 2016; Salazar Miranda, Fan, Duarte, & Ratti, 2021; Schläpfer et al., 2021). Despite their expanded coverage, recent studies relying on cellphone location data cannot distinguish between different activity *types* taking place in streets, providing a less detailed description than direct observation.

This paper proposes *a new methodology for measuring street activity* that combines the strength of approaches based on direct observation

* Corresponding author.
E-mail address: cefzhang@ust.hk (F. Zhang).

[1] See for example, City Transportation Officials (2013).
[2] See Jacobs (1961), Lynch (1960), Milgram (1977), Whyte (1980).

https://doi.org/10.1016/j.landurbplan.2023.104715
Received 28 May 2022; Received in revised form 8 February 2023; Accepted 11 February 2023
Available online 27 February 2023

with the scalability of more recent approaches using cellphone location data. Our approach leverages machine learning and computer vision techniques to measure various forms of street activity using images collected in real time from moving vehicles. These images are collected by a camera-based device that attaches to the front shield of moving vehicles and can capture and locally process images of its surroundings. The key advantage of our approach is that it offers real-time and comparable measures of street activity that can be collected for entire cities. Our approach circumvents the scale limitation typical of direct observation methods while also providing an accurate and detailed way to study street activity in real-time. In addition, the rich street imagery gathered in our study can be used to distinguish between different types of pedestrian activity and other curb uses, such as the number of cars and micro-mobility modes. These data allow us to fully describe the spatial and temporal patterns of street activity in a large area of Paris for the duration of our study.

To show the applicability of our framework, we partnered with Regie Autonome des Transports Parisiens (RATP)—a state-owned public transport operator that manages the largest bus fleet in Paris—to install our camera-based device in one of their buses operating along a fixed route. Our device gathered and processed 226,639 real time images of streets over five weeks, which we then used to classify different types of street activities in the areas of Paris covered by the bus route.

Using these data, we construct two key measures of real-time pedestrian activity. First, we compute the number of *passersby per photo*, which captures the transient daily activity on streets: A measure that urbanists have emphasized as critical to achieve vibrant urban environments (Jacobs, 1961). Second, we compute the *number of stayers per photo*, which proxies for the magnetism of streets and their ability to engage pedestrians in optional and social activities (Gehl, 2011). Following the same approach, we also compute measures for the number of cars and micro-mobility transportation modes per photo. We average these measures to a 100x100-meter grid overlapping the bus line for our analysis.

Using these measures, we produce activity maps for the area of Paris covered by the bus route and analyze how street activity varies geographically and throughout different times of the day. Our results show that streets vary dramatically in how they can support pedestrian activity and that these differences persist over time. Streets attracting the most people do so consistently during our study period.

The paper is organized as follows. Section 3 introduces the framework to analyze street activity dynamics using computer vision. This section explains how real-time images can be collected and describes how to classify street activities. Section 4 describes the deployment in Paris and documents the spatial and temporal patterns of street use using our measures. Finally, Section 5 discusses the results and the opportunities they open for future research.

2. Literature Review

This paper contributes to the literature on place-making that emphasizes the importance of understanding street life. Traditional studies have relied on direct observation (Whyte, 1980), films and photography (Appleyard, 1964), or cognitive maps (Lynch et al., 1990). These approaches offer rich insights into how streets are used but are limited to a few locations and require costly and time-consuming hand-coding of thousands of hours of video from specific parts of a city. For example, in William Whyte's study of public spaces in New York City, he studied how the placement of street furniture and its design influenced people's enjoyment of public space using first-hand observations and interviews combined with stills and time-lapse photography. More recently, Gehl and Svarre (2013) highlighted the importance of furniture and the abundance of access points to generate social mixing, which he measured by observing contact between people. We extend this work by introducing a methodology that leverages computer vision techniques and real-time images obtained from a moving vehicle. Our approach provides consistent measures of human activity and can be used to observe and measure the dynamics of street activity at a fine scale for entire cities where the bus infrastructure is available.

Our study also complements a growing literature in urban studies using emerging technologies such as sensors, Wi-Fi sniffing, Bluetooth tracking, and street imagery to study human dynamics in cities (Ben-Joseph, 2011; Duarte & Ratti, 2021). One branch of this literature exploits passively collected cellphone location data from mobile applications (Ratti et al., 2006; Shoval, 2008; González et al., 2008; Jiang, Ferreira, & González, 2012; Louail et al., 2014; Tan, Zegras, Wilhelm, & Arcaya, 2018; Heine et al., 2021; Salazar Miranda et al., 2021; Bongiorno et al., 2021). Other studies leverage data from static sensors to measure the number of pedestrians using urban infrastructure (Williams et al., 2019). Different from these approaches, we use street images combined with computer vision to identify different uses of streets. Thereore, our study is closer to recent papers using computer vision and street imagery to predict travel patterns in cities. For example, Yin, Cheng, Wang, and Shao (2015) use Google Street View images to predict pedestrian volumes in three cities: Buffalo, NY, Washington, D.C., and Boston, MA. Similarly, Goel et al. (2018) use Google Street View images collected in Great Britain to estimate city-level travel patterns for active and motorized modes. A key difference of our study relative to previous ones is that we collect real-time images instead of images collected at a single point in time, such as those obtained from Google Street View. Our approach allows us to get a richer and more comprehensive picture of which activities are happening and assess how they change throughout the day.

3. A Framework for Measuring and Classifying Street Activity

This section provides an overview of the steps involved in our framework for measuring street activity, and then explains each of the steps in detail. Fig. 1 illustrates the steps involved in our framework. In a pre-deployment phase, we train a deep learning model—the *street activity detection model*—that takes an image as input and outputs a classification for pedestrian activities and transportation modes occurring in it. The deployment phase involves the installation of a camera-based device that gathers real-time street images and processes these data locally using the street activity detection model.

3.1. Pre-deployment Phase: The Street Activity Detection Model

Training the street activity detection model: The first step in our framework is to train the street activity detection model. This process requires a training dataset containing labeled images of street activities, which we use to train and validate the model prior to its deployment. The training dataset can be context and application-specific. For the application of our framework to Paris, we construct our training dataset by combining data from two sources: Google Street View (GSV) and MS COCO images—a large-scale object detection, segmentation, and captioning dataset commonly used to train computer vision models (Lin et al., 2014).

GSV images are useful in this context because they are taken by a moving vehicle, which guarantees that the camera's view is similar to the one taken by the moving bus. We downloaded 6400 GSV images covering the entire city of Paris using the GSV API and then down-selected 1000 images representing a diverse set of activities. We then combined the GSV data with 675 additional images from the MS COCO dataset. The images in MS COCO include many cities and countries and are representative of a broader set of street activities that might be missing or under-represented in the GSV data, such as bicycles and scooters. Together, these two datasets include a comprehensive range of street activities found in streets.

To manually label the training data set with the street activities found in each image, we recruited two student participants. Participants drew a contour line around each street activity using the VGG Image

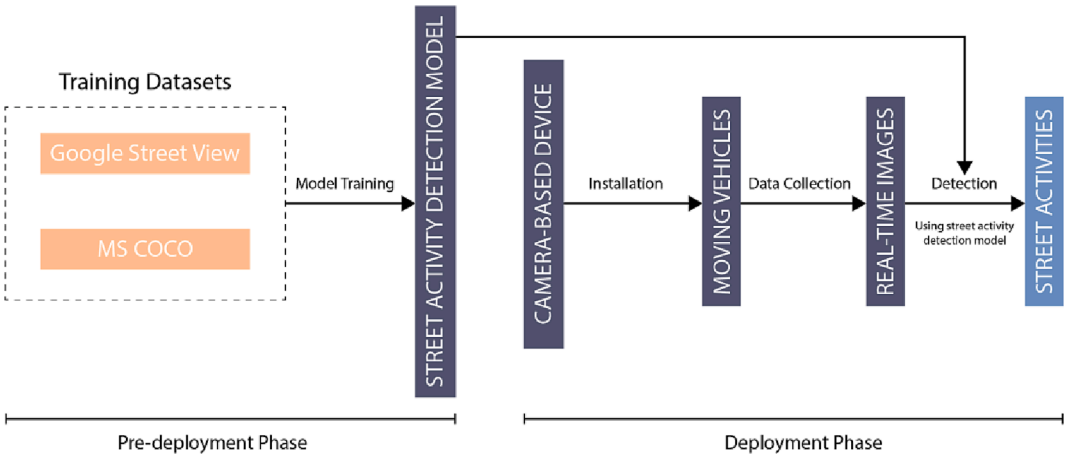


Fig. 1. A framework for measuring street activity. The figure illustrates the two phases of our framework. The pre-deployment phase, which involves training a street activity detection model using images from Google Street and MS COCO. The deployment phase involves the installation of camera-based devices on a fleet of moving vehicles to gather and process data on street activity using the street activity detection model.

Annotator (VIA) (Dutta & Zisserman, 2019). The street activities and transportation modes identified by the participants included: pedestrians, cars, vans, buses, trucks, and micro-mobility modes such as motorcycles, bikes, and scooters. Participants also distinguished between different types of pedestrian activity and annotated "passersby" if people were standing or walking and "stayers" if they were sitting. Fig. A1 in the Supplementary Material plots the total number of activities labeled, showing that most categories contain around 600 labels, except for passersby and cars, which have 8000 and 4500 labels, respectively.

To avoid bias in labeling the training data set, we set two protocols in place: First, we created a two-stage labeling process where each participant first labeled 50 images. Then, every label was reviewed by one author to ensure that both participants used the same annotation criteria. Second, a review was conducted by an author at the end of the entire labeling process to ensure that all the categories labeled were consistent across the image samples.

The final training dataset consists of 1,675 hand-classified images. For purposes of validation, we split this dataset into training and test samples, with 1,424 and 251 images, respectively. Fig. 2 provides examples of the images in our training dataset.

Model and performance: Our street activity detection model is based on the YOLOv4 architecture, which uses a Convolutional Neural Network (CNN) algorithm (See Bochkovskiy et al., 2020). We trained the model using our training sample on an Ubuntu platform using four GeForce 2080 Ti GPUs and the Darknet deep learning framework. The resulting model takes as input an image of 416 by 416 resolution and outputs a prediction class for each human activity or transportation use detected in an image and a confidence score for each prediction. The confidence score ranges from 0 to 1, with one corresponding to higher confidence and zero representing the lowest. We then transferred the weights of the trained model to a YOLOv4-tiny structure, which is explicitly designed for mobile devices and allows us to run the detection model on the camera-based device.

Table 1 reports the performance of the street activity detection model on the test sample. Column 1 indicates the metric used to evaluate the overall performance of the model (average precision or recall). The remaining columns indicate the performance of the model in predicting each individual street activity. The precision is computed as the number of true positives over the sum of true positives and the number of false positives. The recall is computed as the number of true positives (TP) divided by the sum of true positives and the number of false negatives (FN).

The street activity detection model can predict most activities and transportation modes with relatively high accuracy. Column 2 shows that the overall precision and recall of the model are 0.68 and 0.72 (with confidence and IoU thresholds set to 0.25), respectively.[3] Our model can accurately predict the most common transportation modes and street activities, such as cars and passersby, with an average precision above 0.8. This means that when the model is confronted with one of these categories, it predicts it correctly 8 out of 10 times. The highest average precision is for passersby (0.87), while the lowest is for stayers (0.48). The results are similar when we use the recall measure, although the magnitudes tend to be smaller.

To assess the performance of our model, we compare it to one of the most commonly used benchmark models in urban studies: CITYSCAPES. Similar to our model, the CITYSCAPES model is used to classify elements in images—for example, distinguishing between buildings, people, and cars. The CITYSCAPES model is trained using data from over 50 cities and has been used widely for a range of urban research applications (Cordts et al., 2016), including the segmentation of street images to predict ridesharing waiting time, classify vegetation coverage, and predict walking behavior (Wang et al., 2022; Zhou, Tao, Yan, & Sun, 2021; Yang et al., 2019). The bottom panel of Table 1 reports the average precision for the CITYSCAPES model trained in 2021—the same year as our deployment. The table shows that our model outperforms the CITYSCAPES model in average precision across all categories, with the most significant improvements in the micro-mobility and car categories.

3.2. Deployment Phase: The Camera-Based Device

The camera-based device that gathers real-time street images and processes these data is depicted in Fig. 3. The 3D-printed case is 20 cm long by 12.5 wide. Inside the device, we include a wide-angle camera for image detection, a voltage converter, an uninterruptible power source UPS module with a lithium battery to manage power outages, a USB GPS that records the location where each image was taken (latitude and longitude), an IMU module used to record the speed of the bus, and a 4G Wi-Fi modem to transfer the data to a server.

[3] To further assess the accuracy of the street activity detection model we also calculated the F1-score and Mean Average Precision (mAP) to be 0.7 and 0.4, respectively. The mAP was computed by calculating the AP on a different IOU threshold from 0.5 to 0.9 with 0.05 interval increments. The mAP of 0.4 is in line with the performance of Yolo V4 on an MS COCO object detection task, which is 0.41. The F1-score is comparable to the performance of our model reported using the average precision and recall. In addition, Fig. ?? plots the precision-recall curve for the model, which shows that precision and recall are both relatively high between the 0.2 and 0.3 confidence thresholds, further validating our 0.25 confidence threshold choice.

Fig. 2. Examples of images in the training dataset. The figure shows example images from the training dataset obtained from Google Street View (GSV) and MS COCO.

Table 1
Model Performance by Detected Street Activity Type.

	Street Activity Detection Model				
Performance Metric	Overall	Car	Passersby	Stayers	Micro-mobility
Average Precision	0.68	0.82	0.87	0.48	0.56
Recall	0.72	0.73	0.77	0.47	0.58
			Cityscapes model		
Average Precision	0.45	0.62	0.45	0.45	0.32

Notes: Table reports the object detection results for the model. Column 1 indicates the metric used to evaluate the model. Column 2 shows the overall performance, and the remaining columns indicate the performance for each of the activities and transportation modes detected in our experiment. Note that CITYSCAPES does not distinguish by type of pedestrian activity. Therefore, we report the average precision for the "people" category used in the CITYSCAPES model for both the passersby and stayers categories. The CITYSCAPE model also does not have a micro-mobility category. Thus, we report the average precision for the rider, motorcycle, and bicycle activities.

The device can be attached to the front windshield of any moving vehicle. The front windshield location is desirable for two reasons: First, it reduces side occlusions stemming from large vehicles moving at the same speed as the vehicle where the device is installed. Second, in the case of buses, the front position has easy access to its electric panel and guarantees that the sensor is distant from passengers, ensuring no interruption to its power and passenger safety. One might worry that the front placement of the sensor might limit its ability to capture activities on the curbside. To optimize the camera's position, we designed a piece that allows the camera to rotate so it can face the side of the vehicle. Fig. A3 shows the placement of the camera and its field of view. Each photo covers an area of approximately 50 × 50m from the device.

The computing module of the device is a Raspberry Pi 4 connected to all of the sensors. The device was programmed to capture images of the external surroundings of the bus approximately every three seconds. The GPS then records the exact location where the image was taken and saves this information. Importantly, we designed the device so that the street activity detection model is run locally using the YOLOv4-tiny implementation without saving or transferring any of the images. Finally, the device computes a count of the street activities it detects and transfers these counts with the location of the image (latitude and longitude) and a time stamp to a server.

Overall, our approach has two important advantages. First, the operation of the device is completely automatic and does not require any intervention by the driver or bus operator. Second, and most importantly, the device processes the images locally without saving them, allowing us to construct rich metrics on street use while respecting individual privacy.

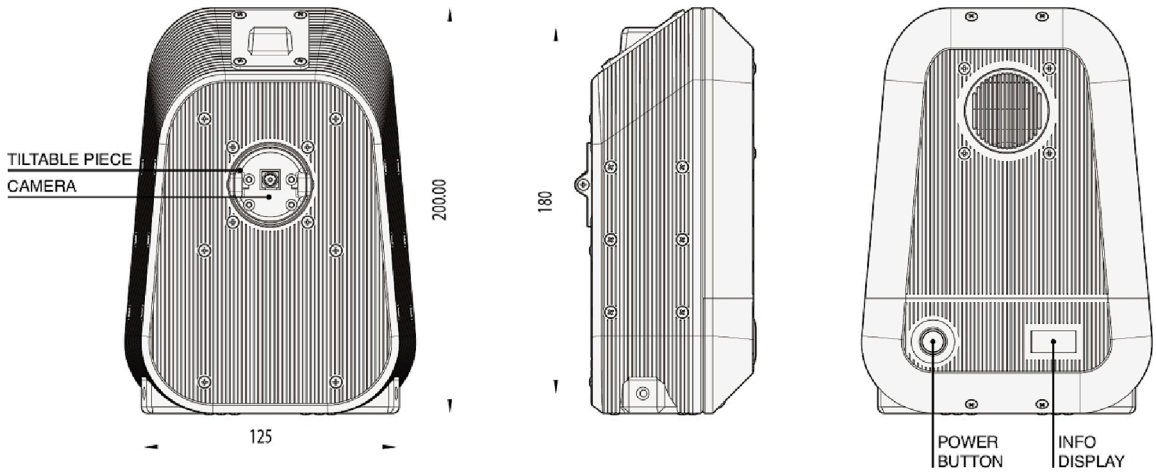


Fig. 3. Camera-based design. The figure shows the schematics of the device from the back (left panel), side (middle panel), and front (right panel). The device measures 20 cm long by 12.5 wide and is attached to the front bus windshield to maximize the camera's field of view.

4. Deployment: Spatial and Temporal Patterns of Street Activity in Paris

We partnered with RATP to install our camera-based device in one of their bus lines to measure street activity. We focused on the 70 bus line because it crosses many of Paris's eastern and western areas, revealing diverse economic and social environments and because it passes through many different street types: from main streets close to the Seine river to commercial and residential areas towards the south of the city. Our device processed data for five weeks starting in May 2021, collecting a total of 226,639 photos between 5:00 AM and 11:00 PM—the bus shift schedule.[4]

With these data, we analyze street activity in the areas of Paris covered by our deployment. We first construct pedestrian activity and transportation mode measures extracted from the collected images. For pedestrian activity, we construct two measures: the *number of passersby per photo*, which captures transient pedestrian activity, and the *number of stayers per photo*, which proxies for the magnetism of streets and their ability to engage people in optional and social activities (Gehl & Svarre, 2013). Distinguishing between these two types of human activity is important for the street offerings available in a city. For instance, a street dominated by passersby is more likely to be used as a transport connector between two areas, while a street attracting mostly stayers is more likely to have cafes, terraces, and benches, among other infrastructure that compels pedestrians to stay in the area.

For transportation modes, we compute measures for the number of cars and micro-mobility modes per photo. We define micro-mobility as an aggregate category containing objects classified as bicycles, motorbikes, and scooters. Finally, we average all our measures to 100x100 meter grids overlapping the bus line.[5] Averaging activity counts across grids guarantees that even if the bus stalls and detects the same activity in different photos (for example, when waiting at a traffic light), these activities will not be counted more than once. To minimize measurement error, we focus our analysis on the grids for which the camera-based device collected at least ten street images. Our final sample consists of 151 grids covering our study area.

The left panel of Fig. 4 shows the total number of activities detected by our model for the study period. The two most common street activities are cars and passersby, with our algorithm identifying close to 200,000 cars and 130,000 passersby. The right panel shows the data broken down by hour. The number of cars and passersby increases daily, peaking around 5:00 PM. Instead, the number of micro-mobility modes and stayers remains relatively constant throughout the day. Note that our data collection period coincides with Covid restrictions in Paris, so we should expect lower levels of street activity overall.

We now explore the temporal changes in street activity using the number of passersby and stayers per photo as our two main measures of pedestrian activity on streets. Fig. 5 plots the time series of pedestrian activity throughout the five weeks of our study. The left panel shows the data for passersby and the right panel shows the data for stayers. In both panels, we show the time series separately for the morning (between 5:00 AM and 2:00 PM), afternoon (between 2:01 PM and 7:00 PM), and night (between 7:01 PM and 11:00 PM). We find that the average grid in our study area has 1.01 passersby per photo and 0.006 stayers. When interpreting these numbers, one should keep in mind that each photo covers a depth and width of roughly 50 meters. This implies that one can also think of these measures as capturing the number of passersby and stayers every 2,500m^2. One could also convert these numbers to people per grid by multiplying them by four since each grid covers an area of 10,000m^2. This means that grids in our sample had 4 passersby and 0.024 stayers on average at each point in time.

The figure also shows that the number of passersby and stayers varies with the time of the day. The left panel shows that passersby are the most common during the afternoon, with grids in our sample having 1.21 passersby per photo. Grids averaged 0.81 passersby per photo in the morning and 0.32 at night, respectively. As shown by the right panel, these patterns differ for stayers, which do not vary as much with the time of the day.

In terms of temporal changes, the figure shows that the number of passersby remained roughly constant during our study period. We do, however, observe a temporary increase in stayers during the last two weeks of May. This temporary increase was observed at all times but was more pronounced in the afternoon, which might be explained by the relaxing of the night curfew that allowed cafes, bars, and restaurants to offer outdoor service starting on May 19th.

We now turn to explore how street activity varies across space. We focus on the average number of stayers and passersby in each grid during

[4] Data was collected consistently across the five weeks, except for 10 days when the bus was not operating due to maintenance or repairs.

[5] We aggregate the data from each image to the 100x100 meter grid using the GPS location of the device at the time the image was taken. By construction, the location of the device is not exactly the same as the centroid of the image. Instead, the centroid of the image is located approximately 35 m away from the device. This follows from the fact that the images, on average, cover an area with a depth and width of 50 m, which implies that their centroid is located at $\sqrt{2} \times 25 \approx 35m$ away from the camera. However, note that the distance between the device and the centroid of the image is constant and does not affect the data aggregation into 100x100m grids. In particular, each grid represents the street activity captured by the camera 35 m away from it.

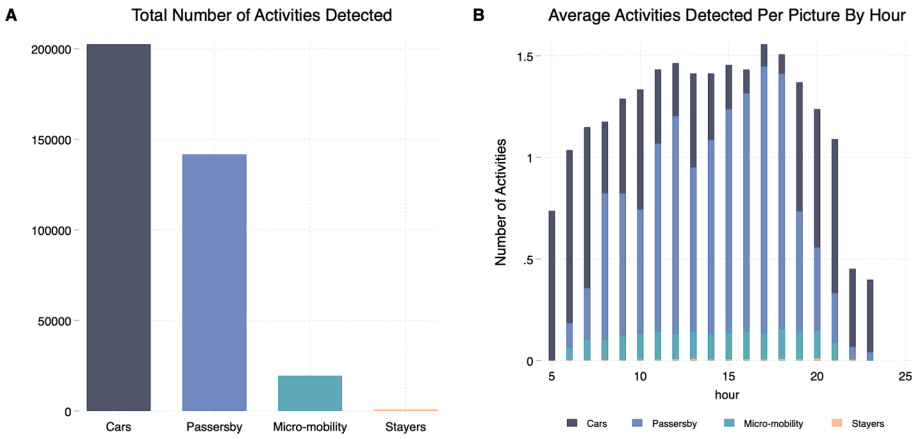


Fig. 4. Street activity by type. The left panel shows the total activities by type detected in the five-week period. The right panel shows the total activities detected by type and hour of the day.

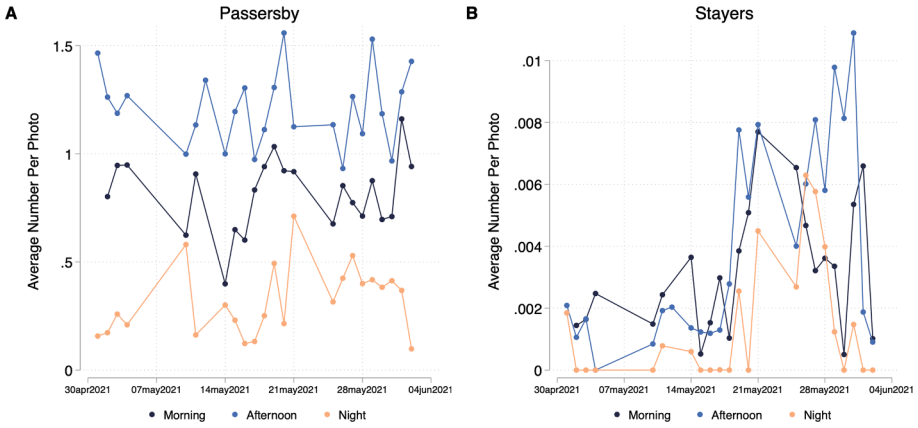


Fig. 5. Street activity by type over time. The figure shows the average number of activities per photo for the morning, afternoon, and night periods across the 5-week period. The left panel plots the time series for passersby, and the right panel for stayers. The markers in the figure indicate the days for which data was collected. Days without markers correspond to those for which we do not have data.

the study period. Fig. 6 shows the distribution of passersby and stayers across grids. Panels A and C show the geographic distribution of grids belonging to different percentile breaks of the street activity distribution, each represented by a different color. In particular, we group grids into those belonging to the 0–10th, 10–25th, 25–50th, 50–75th, and 75–90th percentiles, and those in the top 10 percent of grids with the highest levels of street activity. Grids in the top 10 percent of the passersby distribution are scattered throughout our study area but are more frequent towards the Grenette area. In contrast, grids in the top 10 percent of the stayers distribution are more frequently encountered in the areas of Javel and Auteuil.

Panels B and D plot the histograms indicating the breakpoints for different percentiles of the distribution of passersby (top) and stayers (bottom). The histogram for passersby reveals large variation in activity across grids. Grids in the top 10 percent of the distribution—or hotpots (colored in orange)—have 1.90 passersby per photo on average, while grids in the bottom 10 percent (colored in dark blue) have only 0.33. The magnitudes for stayers shown in the right panel are much smaller, with many grids having zero stayers per picture.

We have documented which areas of the city are frequented by passersby and stayers. We now characterize the spatial distribution of these "high attractor" areas in space. To do this, we compute the distance from every grid to the closest hotspot, defined as grids falling in the top 10 percent of the passersby or stayers distribution for the entire five-week period. Hotspots are defined using the average number of passersby or stayers in each grid. If hotspots were clustered in a single area, one would expect the distribution of distances to be uniform and to have wide support. Fig. 7 shows a narrow distribution of distances with a fat lower tail, indicating that many grids are close to hotspots: 50% of the grids are within 500 meters of a hotspot. For comparison, the figure also plots a simulated distribution of distances in a counterfactual scenario where all hotspots are assumed to be clustered in the middle of our study area. This counterfactual distribution is uniform and wide, ranging from 0 to 3600 meters. In this case, only 8% of the grids would fall within 500 meters of a hotspot. Overall, these results show that street activity hotspots are frequent and dispersed over our study area.

Panels C and B of Figs. 6 and 7 show similar patterns for stayers, with the main difference being that the level of street activity measured by stayers is much lower.

As a final exercise, we explore whether the location of street activity across grids persists over time. For this exercise, we work with a combined measure of street activity that sums the number of passersby and stayers, capturing the total pedestrian activity per photo. We computed a transition matrix week-by-week, which gives the probability that a grid goes from one group of percentiles (for example, the top ten percent of grids with the highest activity) to another group in the next week (for example, grids between the 50th and 75th percentiles of activity). We use the six groups of percentiles introduced in Fig. 6 (from 0–10,10–25,25–50,50–75,75–90,top 10 percent). Fig. 8 provides the resulting 6 × 6 matrix of weekly transition probabilities averaged across all weeks, and with the cells color-coded so that lighter colors indicate higher values. The figure shows that the transition matrix has a strong

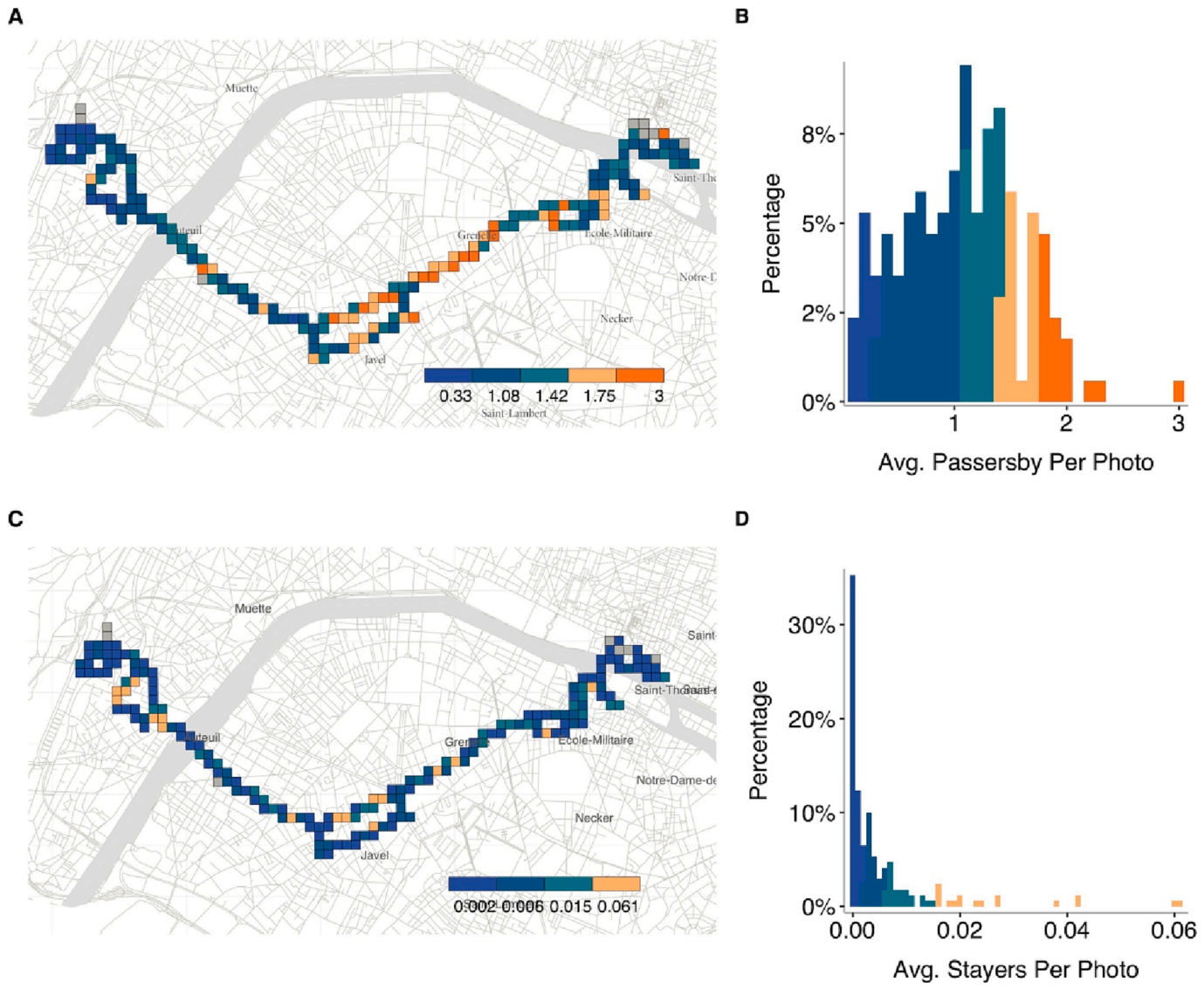


Fig. 6. Number of stayers and passersby across space. The top panel shows the map and histograms with the average number of passersby by grid across the five-week period. The histograms are colored according to the breakpoints for different percentiles of the passersby distribution. The bottom panel shows the analogous for stayers. Grids shown in grey overlap the bus line but either are missing photograph data or have less than ten photos. Grids are constructed based on the location of the GPS location of the camera device at the time each picture was taken.

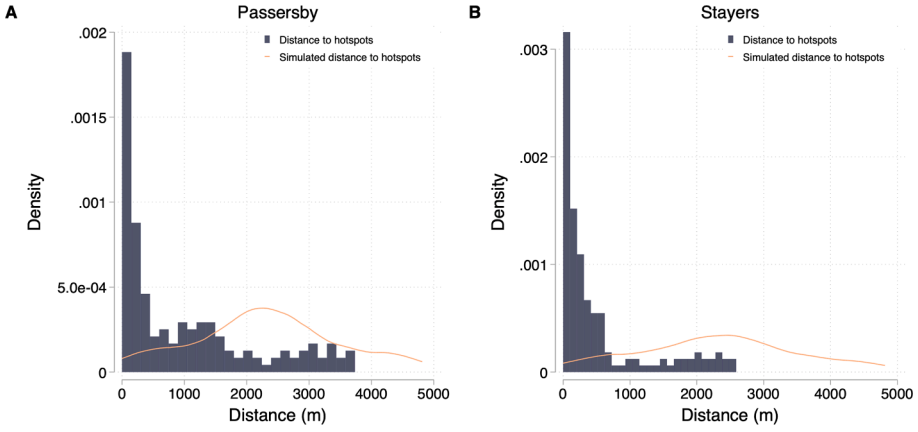


Fig. 7. Distribution of street activity by distance to hotspots. The figures show histograms of distance to the closest hotpot (in blue) and simulated distances to a counterfactual cluster of hotspots in the middle of our study area (in orange). The left panel shows the results for passersby ad the right panel for stayers.

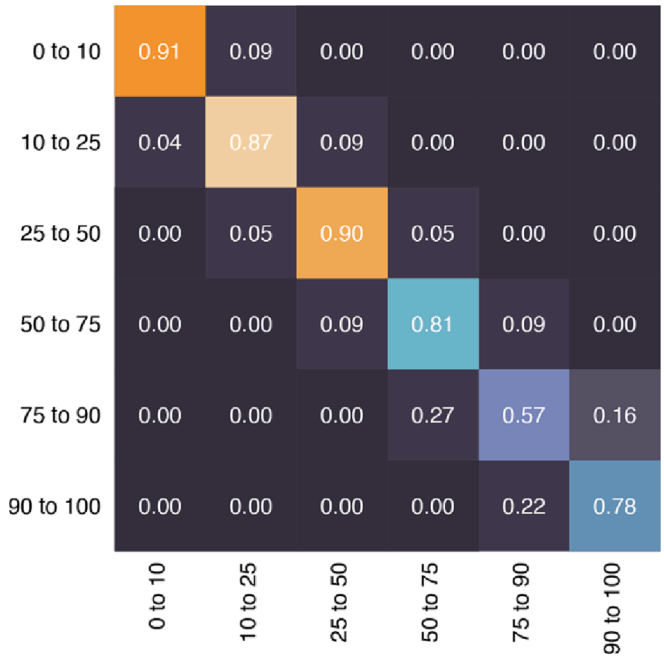


Fig. 8. Persistence of street activity. The figure shows a transition matrix showing the probabilities of moving from a group of percentiles in week i to another group in week $i+1$, averaged across our study period. Each row in the matrix gives the probability of moving from the group indicated in the row header to the group indicated in the column header next week.

diagonal component, with all diagonal entries exceeding 0.57. The closer the diagonal entries are to 1, the higher the persistence of the spatial distribution of street activity. Thus, street activity appears to be highly persistent across space.

5. Conclusion

This paper introduces a novel framework for obtaining detailed measures of street activity using real-time images. Our approach leverages machine learning and computer vision techniques to identify various forms of street activity from images collected in real-time by a moving vehicle. Our methodology can be used to distinguish between different types of pedestrian activity and other curb uses, such as the number of cars and micro mobility modes in a street and provides an accurate and detailed way to study street use.

Besides developing this framework, this paper provides a first application in Paris revealing several patterns of street activity:

1. The number of passersby (1 per picture, on average) is two orders of magnitude larger than the number of stayers (0.006 per picture, on average).
2. Street activity varies during the day and peaks in the afternoon.
3. Our measures of street activity show an increase in the number of stayers (from 0.002 per picture to 0.01 during the afternoon) when Covid restrictions were lifted, around the week of May 19th.
4. Street activity is highly dispersed across space, and this dispersion is highly persistent.
5. Hotspots of street activity are not clustered in a single area but are dispersed across Paris.

Our findings align with theoretical work suggesting that cities feature multiple sub-centers of activity (Christaller & Baskin, 1966) and empirical evidence documenting the stability of highly attractive centers across cities of different sizes (Louail et al., 2014). Extending our framework to measure street activity in other cities would also provide data to understand how theories of the built environment and street networks, such as Space Syntax, shape human activity (Hillier, Perm, Hanson, Grajewski, & Xu, 1993).

Going beyond this specific application, our methodology offers a tool that can be used in other contexts and cities for collecting data on street use. These real-time measures can be used to evaluate the success of street interventions, such as pedestrianizing a street or adding a new bus stop. Finally, our measures, which distinguish between people and different vehicular modes, can also be used to identify bottlenecks of conflicting uses between pedestrians, curbside delivery, and micro-mobility modes.

This paper opens two additional avenues for future research. Future work could extend the methodology for other areas of Paris and cities worldwide, allowing for a better understanding of how street activity patterns vary across contexts. In addition, one can use these measures to understand the determinants of street activity and what explains the success of some streets over others.

Data availability

The authors do not have permission to share data.

Acknowledgement

We thank LandWey, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Volkswagen Group America, FAE Technology, Samoo Architects & Engineers, Shell, GoAigua, ENEL Foundation, the cities of Helsingborg, Amsterdam, and Stockholm, and all other members of MIT Senseable City Laboratory Consortium for supporting this research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.landurbplan.2023.104715.

References

Appleyard, D. (1964): The view from the road. Published for the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University by the M.I.T. Press, Massachusetts Institute of Technology, Cambridge.

Ben-Joseph, E. (2011). *Companion to urban design*. Routledge companions: Routledge, Milton Park, Abingdon, Oxon.

Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao (2020): Yolov4: Optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934.

Bongiorno, C., Zhou, Y., Kryven, M., Theurel, D., Rizzo, A., Santi, P., Tenenbaum, J., & Ratti, C. (2021). Vector-based pedestrian navigation in cities. *Nature Computational Science, 1*(10), 678–685.

Christaller, Walter, & Baskin, C. W. (1966). *Central places in southern Germany*. Englewood Cliffs, N.J.: Prentice-Hall.

Cordts, M., M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele (2016): The Cityscapes Dataset for Semantic Urban Scene Understanding, in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Duarte, F., & Ratti, C. (2021). *What Urban Cameras Reveal About the City: The Work of the Senseable City Lab*. Springer Singapore.

Dutta, A., and A. Zisserman (2019): The VIA Annotation Software for Images, Audio and Video, in Proceedings of the 27th ACM International Conference on Multimedia, MM '19, New York, NY, USA. ACM.

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

Gehl, J., & Svarre, B. (2013). *How to Study Public Life*. Springer.

Goel, R., Garcia, L. M., Goodman, A., Johnson, R., Aldred, R., Murugesan, M., Brage, S., Bhalla, K., & Woodcock, J. (2018). Estimating city-level travel patterns using street imagery: A case study of using Google Street View in Britain. *PLoS ONE, 13*(5), 1–22.

González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.

Heine, C., C. Marquez, P. Santi, M. Sundberg, M. Nordfors, and C. Ratti (2021): Analysis of mobility homophily in Stockholm based on social network data, PLoS ONE, 16(3 March), 1–14.

Hillier, B., Perm, A., Hanson, J., Grajewski, T., & Xu, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B, 20*, 29–66.

Jacobs, J. (1961). *The death and life of great American cities*. New York: Random House Inc.

Jiang, S., Ferreira, J., & González, M. C. (2012). Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery, 25*(3), 478–510.

Jiang, S., Ferreira, J., & Gonzalez, M. C. (2016). *Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore*. IEEE.
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* (pp. 740–755). Springer.
Louail, T., Lenormand, M., Cantu Ros, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2014). From mobile phone data to the spatial structure of cities. *Scientific Reports, 4*, 1–12.
Lynch, Banerjee (1990). *Tridib., Southworth, Michael., K. City sense and city design: writings and projects of Kevin Lynch*. Cambridge, Mass: MIT Press.
Lynch, K. (1960). *The image of the city*. MIT Press.
Milgram, S. (1977). *Psychological Maps of Paris*. Mcgraw-Hill Book Company.
Nasar, J. L. (1990). The Evaluative Image of the City. *Journal of the American Planning Association, 56*(1), 41–53.
Of City Transportation Officials, N.A. (2013): Urban Street Design Guide. Island Press, Washington, OCLC: ocn844728420.
Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design, 33*(5), 727–748.
Salazar Miranda, A., Fan, Z., Duarte, F., & Ratti, C. (2021). *Desirable streets: Using deviations in pedestrian trajectories to measure the value of the built environment* (p. 101563). Computers: Environment and Urban Systems.
Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., & West, G. B. (2021). The universal visitation law of human mobility. *Nature, 593*(7860), 522–527.
Shoval, N. (2008). Tracking technologies and urban analysis. *Cities, 25*(1), 21–28.
Tan, S. B., Zegras, P., Wilhelm, E., & Arcaya, M. C. (2018). Evaluating the effects of active morning commutes on students' overall daily walking activity in Singapore: Do walkers walk more? *Journal of Transport and Health, 8*, 220–243.
Wang, M., Chen, Z., Rong, H. H., Mu, L., Zhu, P., & Shi, Z. (2022). Ridesharing accessibility from the human eye: Spatial modeling of built environment with street-level images, Computers. *Environment and Urban Systems, 97*, Article 101858.
Whyte, W. (1980): The social life of small urban spaces. Project for Public Spaces.
Williams, S., Ahn, C., Gunc, H., Ozgirin, E., Pearce, M., & Xiong, Z. (2019). Evaluating sensors for the measurement of public life: A future in image processing. *Environment and Planning B: Urban Analytics and City Science, 46*(8), 1534–1548.
Yang, Y., He, D., Gou, Z., Wang, R., Liu, Y., & Lu, Y. (2019). Association between street greenery and walking behavior in older adults in Hong Kong. *Sustainable Cities and Society, 51*, Article 101747.
Yin, L., Cheng, Q., Wang, Z., & Shao, Z. (2015). 'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Applied Geography, 63*, 337–345.
Zhou, H., Tao, G., Yan, X., & Sun, J. (2021). Influences of greening and structures on urban thermal environments: A case study in Xuzhou City, China. *Urban Forestry & Urban Greening, 66*, Article 127386.