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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Reference-free video-to-real distance approximation-based urban social distancing analytics amid COVID-19 pandemic

Fan Zuo\textsuperscript{a,}\textsuperscript{*}, Jingqin Gao\textsuperscript{a}, Abdullah Kurkcu\textsuperscript{b}, Hong Yang\textsuperscript{c}, Kaan Ozbay\textsuperscript{d}, Qingyu Ma\textsuperscript{e}

\textsuperscript{a} C2SMART Center, Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY, 11201, USA
\textsuperscript{b} Ulteig, 5575 DTC Parkway, Suite 200, Greenwood Village, CO, 80111, USA
\textsuperscript{c} Department of Computational Modeling and Simulation Engineering, Old Dominion University, 1117 ENGR & COMP SCI BLDG, Norfolk, VA, 23529, USA
\textsuperscript{d} C2SMART Center, Department of Civil and Urban Engineering & Center for Urban Science and Progress (CUSP), Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY, 11201, USA
\textsuperscript{e} Department of Computational Modeling and Simulation Engineering, Old Dominion University, 1117 ENGR & COMP SCI BLDG, Norfolk, VA, 23529, USA

ARTICLE INFO

Keywords:
Social distancing
COVID-19
Close contact
Pedestrian
Deep learning
Computer vision

ABSTRACT

Introduction: The rapidly evolving COVID-19 pandemic has dramatically reshaped urban travel patterns. In this research, we explore the relationship between “social distancing,” a concept that has gained worldwide familiarity, and urban mobility during the pandemic. Understanding social distancing behavior will allow urban planners and engineers to better understand the new norm of urban mobility amid the pandemic, and what patterns might hold for individual mobility post-pandemic or in the event of a future pandemic.

Methods: There are still few efforts to obtain precise information on social distancing patterns of pedestrians in urban environments. This is largely attributed to numerous burdens in safely deploying any effective field data collection approaches during the crisis. This paper aims to fill that gap by developing a data-driven analytical framework that leverages existing public video data sources and advanced computer vision techniques to monitor the evolution of social distancing patterns in urban areas. Specifically, the proposed framework develops a deep-learning approach with a pre-trained convolutional neural network to mine the massive amount of public video data captured in urban areas. Real-time traffic camera data collected in New York City (NYC) was used as a case study to demonstrate the feasibility and validity of using the proposed approach to analyze pedestrian social distancing patterns.

Results: The results show that microscopic pedestrian social distancing patterns can be quantified by using a generalized real-distance approximation method. The estimated distance between individuals can be compared to social distancing guidelines to evaluate policy compliance and effectiveness during a pandemic. Quantifying social distancing adherence will provide decision-makers with a better understanding of prevailing social contact challenges. It also provides insights into the development of response strategies and plans for phased reopening for similar future scenarios.
1. Introduction

According to the World Health Organization (WHO) and the U.S. Centers for Disease Control and Prevention (CDC), social distancing is currently the most effective nonpharmaceutical way to slow the spread of the novel coronavirus disease 2019 (COVID-19) due to the aerosol-transmission method through which the virus is spread from person to person.

Social distancing refers to efforts including avoiding mass gatherings, closing public places, and keeping a sufficient distance (commonly at least 6 feet) between people to reduce disease spread by maximizing physical distance and minimizing frequency of human contacts (Ferguson et al., 2005). For example, on Mar 20, 2020, New York Governor Andrew Cuomo announced the “New York State on PAUSE” executive order requiring all non-essential businesses to close in-office personnel functions, and temporarily banning all social gatherings (NYS, 2020). Similar guidelines issued by other city and state government agencies have urged individuals to maintain a minimum of 6-feet of social distance from others in public settings.

Although social distancing orders are mandated, how people are responding to these policies is not clear. People may ignore these guidelines, or may still go outside for essential activities (e.g., to work, or to purchase groceries). In such a context, investigating crowd density and actual frequency of social contact between people is crucial to measuring the effectiveness of the policy and reducing the chances of community transmission. Numerous studies have been performed to analyze close contact in different types of indoor environments such as in hospitals (Isella et al., 2011a; Hornbeck et al., 2012; Vanhems et al., 2013), schools (Salathé et al., 2010; Hoang et al., 2019; Stehlé et al., 2011b), offices (Zhang et al., 2019, 2020), and conference venues (Isella et al., 2011b; Smieszek et al., 2016; Stehlé et al., 2011a). In contrast, there is a very limited effort to measure the contact behavior between individuals in an outdoor environment. This has been largely attributed to the lack of effective and safe monitoring solutions for continuous data collection in complex environments. Moreover, the dynamics of human interaction in open environments do not evolve to an equilibrium state; instead, these interactions are likely to fluctuate, vary in time, and remain vulnerable to other uncertainties.

When the pandemic begins to ease, the volume of people traveling within a city will start to increase. However, it is likely that this return will be gradual, as more discretionary trips are postponed, social-distancing habits become ingrained, or a general fear of travel persists. It is also possible that this crisis may lead to changes in how people prefer to travel within urban areas, potentially eschewing typically crowded modes such as public transportation (Wang et al., 2020) or shared mobility (Liu et al., 2021), and choosing more solitary modes, such as walking, instead (Bian et al., 2021). Despite the perceived solitude of walking, this shift in preference may lead to higher levels of interactions on dense city sidewalks. These behavioral changes and their impacts on urban environments are unpredictable and have been previously underexamined. Thanks to the recent developments of emerging technologies, such as wireless positioning systems using WiFi or Bluetooth, or computer vision techniques, many new solutions have been developed to facilitate social distancing practices and monitor the dynamics of human interactions in an urban setting (Kurkcu and Ozbay, 2017; Zuo et al., 2020; Zuo et al., 2020).

This study aims to introduce a low-cost, continuous, remote, real-time social-distancing big data acquisition and pedestrian detection framework (Fig. 1). This framework leverages numerous traffic camera feeds along with a deep learning-based video processing method for analyzing time-dependent social distancing patterns in the outdoor environment at the local level. Since many cities have already installed traffic cameras that can be used for object detection, no additional equipment or extra implementation costs are required. This approach is fully remote and free of risk as it does not require the presence of human investigators in the field. This is critical in safely deploying effective field data collection during the crisis as interaction with human subjects is reduced or prohibited.

The quantified heterogeneity in terms of pedestrian density, social distance distribution, and temporal variations can be used to inform residents of the potential risk of exposure in an urban environment and assist in evaluating the effectiveness of relevant public interventions. The proposed framework can also provide authorities with insight into density trends during the reopening phases, to assist in developing effective response strategies or to plan for similar future scenarios.

The rest of this paper is organized as follows:

Fig. 1. Proposed data acquisition and pedestrian detection framework.
1. A summary of related work on social distancing and pedestrian monitoring technologies.
2. A description of our proposed real-time social-distancing big data acquisition and video processing framework, as well as a distance approximation methodology.
3. A discussion of experimental results through case studies.
4. Conclusions and future research perspectives

2. Literature review

Many recent works have shown evidence of the effectiveness of social distancing practices during the COVID-19 pandemic. For example, De Oliveira et al. (De Oliveira et al., 2020) showed that the proportion of individuals staying home had a strong inverse correlation with the time-dependent reproduction number R(t) (rho < −0.7), a measure of disease transmissibility, using aggregated mobile phone data. Engle et al. (Engle et al., 2020) used county-level location-based information extracted from mobile devices and combined it with COVID-19 case data and population characteristics to estimate the effects of disease prevalence and social distancing orders’ impact on mobility. Their results show that an official stay-at-home policy leads to a 7.87% decrease in individual mobility and that a rise in the local infection rate from 0% to 0.003% corresponds to a 2.31% reduction in mobility.

Recent advancements in information and communication technologies have made it possible to collect empirical data relevant to human behavior and social contact in real-world conditions instead of relying on survey data. Solutions such as tracking WiFi and Bluetooth traces (Faggian et al., 2020; Berke et al., 2020), using smartphone applications (Cho et al., 2020; Inn, 2020; Udugama et al., 2020), and active radio frequency identifications devices (RFID) or other wireless sensors (Guo et al., 2020; Zhang et al., 2020; Vanhems et al., 2013, A. Barrat et al., 2014) have been recently employed to collect data on proximity of human-to-human interactions. However, many of these applications are better suited for indoor localization, and they may have reduced accuracy in dynamic environments. In addition, there are significant concerns about the scalability and privacy of such applications (Berke et al., 2020; Cho et al., 2020).

Table 1 summarizes the primary crowd detection technologies that can be used for both outdoor and indoor environments. A few of the technologies have a medium to high installation or equipment cost (e.g., LIDAR or Thermal sensors), and most integrate with existing systems, such as mobile phones and wearable devices (e.g., WiFi, Bluetooth, cellular, UWB, inertial sensors, and visible light). A drawback of such approaches is that position accuracy may drop when a user’s mobile phone is located inside a pocket or bag, and the device must remain in certain operational modes (e.g., Bluetooth mode) (Nguyen et al., 2020). The need for deploying additional hardware in-field or installing software applications on an individual’s smartphone also make it less feasible and less economical in the context of the COVID-19 crisis. Moreover, the temporal resolution and detection time of these technologies is not sufficient for measuring many close contacts within a short duration (Zhang et al., 2019; Cattuto et al., 2010; Vanhems et al., 2013). Thus, very little precise data, especially continuous data combining environment facts and high temporal resolution, exist on COVID-19 related social distancing behavior.

As shown in Table 1, computer vision offers a low-cost, appealing alternative to smartphones or networked wearable sensors for safety assessment (Xie et al., 2016, 2019; Li et al., 2016) or pedestrian detection (Manlises et al., 2015) as surveillance cameras have already been installed in many cities (Zuo et al., 2019). The use of computer vision technology, object detection in particular, can turn surveillance cameras into “smart” sensors capable of detecting crowd density and identifying compliance with social distancing requirements in real-time (Nguyen et al., 2020). More importantly, computer vision provides a risk-free approach to data collection during the crisis. As Lobe et al. (2020) emphasized, social distancing mandates are restricting traditional in-person investigations of all kinds. For example, the Institutional Review Board (IRB) suspended all in-person human subjects research activities in response to the pandemic.

There are two main types of approaches that are commonly used in the computer vision domain. The first one is a region-based approach, such as Fast-RCNN (Girshick, 2015), Faster-RCNN (Ren et al., 2015), Mask R-CNN (He et al., 2017), and RetinaNet (Lin et al., 2017), which detect humans from images in two stages, including region proposal and processing according to regions (Zhao et al., 2017), which detect humans from images in two stages, including region proposal and processing according to regions (Zhao et al., 2017).

Table 1
Crowd detection/positioning technologies for both outdoor and indoor environments.

| Technology             | Range                  | Cost       | Position Accuracy | Privacy   |
|------------------------|------------------------|------------|-------------------|-----------|
| Wi-Fi                  | Outdoor: 100 m         | $          | Medium            | Low       |
|                        | Indoor: 50 m           | $          | (1 m–5m)          | Low       |
| Bluetooth              | Outdoor: 55–78 m       | $          | Medium            | Low       |
|                        | Indoor: 15–35 m        | $          | (1 m–2m)          | Low       |
| Cellular               | Short to Long          | $          | High (Less than 50cm) | Low to High |
| Ultrawideband (UWB)    | Short to Medium        | $          | High (Less than 10cm) | High      |
| Radar/LIDAR            | 100 m                  | $–$$      | High              | High      |
| Seismic sensor         | Short (~15 m)          | $          | High              | High      |
| Computer Vision        | Varies by camera       | $–$$      | Low to High       | Low       |
| Infrared/Thermal sensor| IRP 1–10 m, THC a km   | $–$$      | Medium            | Low to High |
| Inertial sensors       | Not applicable         | $          | Medium (Less than 1 m) | High      |
| Visible light          | Short                  | $          | High (1 cm–20 cm) | High      |

Note: Synthesized based on References (Nguyen et al., 2020) (Bernas et al., 2018) and (FHWA, 2013).
et al., 2019). Although region-based approaches have a high detection accuracy, their applications may be limited due to their high complexity. The second type of approach is called a unified-based approach that includes the You Only Look Once (YOLO) model (Redmon et al., 2016), improved versions of this model (Redmon and Farhadi, 2017, 2018; Bochkovskiy et al., 2020), and Single Shot Multibox Detector (SSD). These approaches map image pixels to bounding boxes and class probabilities to detect humans or objects and is usually faster than region-based methods (Nguyen et al., 2020). Combined with distance approximation methods, these computer vision techniques can identify whether or not a group of people is complying with social distancing requirements, helping to revealing human contact patterns and effectiveness of social distancing policies.

Another consideration is that most recent relevant studies (De Oliveira et al., 2020; Ghader et al., 2020; Engle et al., 2020) used aggregated mobility data with the assumption of homogeneous behavior within a city, state, or county. Such an approach is helpful for macro-level analyses. However, it is possible that different cities or even different areas within the same city may have non-synchronized social distancing patterns in open urban spaces. The computer vision-based approach leveraging existing data from cameras in cities is able to provide detailed spatiotemporal heterogeneity of social distancing patterns at a micro-level as well.

3. Data acquisition

The closed-circuit television (CCTV) system is a valuable source of traffic condition information for many transportation systems. Traffic video data can provide rich information, such as traffic volume, travel speed, and incident information, to facilitate traffic operations and management (NYSDOT, 2019). This paper collected traffic video data from NYC to support the study of social distancing. NYC Department of Transportation (NYCDOT) traffic cameras provide frequently updated still images from 731 locations in the five boroughs (Fig. 2). During the COVID-19 pandemic, these cameras are running normally, though it should be noted that these cameras only provide live feeds and do not record any footage and are sometimes repositioned to view traffic from varying directions.

We developed a real-time data acquisition framework (Fig. 1) to automatically and continuously collect large-scale social distance behavior data of pedestrians from sampled streaming footage of active traffic cameras distributed in the five boroughs of NYC, including in the Bronx, Brooklyn, Manhattan, Queens, and Staten Island, where pedestrian activities are observed (e.g., cameras facing crosswalks or sidewalks). The raw video data are sampled and fed into a deep learning architecture for pedestrian data extraction. This paper uses data collected from 11 sampled locations as a case study.

Our data acquisition framework contains a multi-task deep learning model that embeds the conventional neural network (CNN) for pedestrian detection and characterization. The simultaneous accomplishment of multiple tasks ensure its computational efficiency and inference performance for large-scale data collection practices. Different model architectures are evaluated and compared. Human reviewers help collect ground-truth data for model training and validation by manually labeling pedestrians from the collected video image data. Our final data collection approach is risk-free without requiring the presence of in-field human investigators and offers the opportunity to collect a massive amount of perishable data to objectively and timely profile unique social distance patterns using NYC as a living laboratory. It should be mentioned that the collected data did not capture private information, such as faces or traces, that

Fig. 2. Traffic camera locations in NYC(NYCDOT).
could identify any individuals.

4. Methodology

The proposed approach for quantifying social distancing patterns is centered on the idea that one can obtain pedestrian density and distance between each pedestrian pair by using a pre-trained object detection model, real distance approximation, and several post-processing filters. Fig. 3 presents the overall workflow of the proposed methodology. The program is developed using Python and the primary machine learning modules are supported by open-source libraries, including TensorFlow (Abadi et al., 2016), Keras (Chollet, 2015), and ImageAI (Olafenwa and Olafenwa, 2018).

4.1. Pre-trained object detection model

Pre-trained state-of-the-art object detection models are shown to have good generalization capability allowing efficient deployment to new environments, even with different video resolutions or camera angles (Du, 2018; Zhang et al., 2016). Therefore, a pre-trained pedestrian detection model is selected for use in this study. The object detection model detects objects (e.g., pedestrians) from each frame and finds boxes around the objects. The detector then returns a list of predicted potential classes of each object with probabilities and classifies the object into a relevant class type based on the highest probability. The Non-Max Suppression method (Felzenszwalb et al., 2009) with a fine-tuning threshold is used to control the output balance between false positives and false negatives by setting the minimum acceptable probability of identified class and rejecting all identified classes with probabilities lower than the threshold (see Fig. 1, lower right).

![Fig. 3. Proposed workflow of collecting and mining pedestrian social distance data from publicly available surveillance video data.](image-url)
4.2. Object detection model selection and evaluation

Three well-known state-of-the-art object detection models, YOLOv3, RetinaNet, and Mask RCNN are selected as backbone models. The performances of these models are compared with ground truth data extracted from two selected locations to specify the best model parameters for the subsequent multi-location analysis. The comparison is also used to validate their detection accuracy on the existing video surveillance systems used in this study. Both RetinaNet and Mask RCNN used ResNet-101 as the network architecture. These models are all pre-trained using the COCO dataset (Lin et al., 2014).

Each video contains 4 hours of surveillance data with a 30-s collection frequency. Precision-Recall (PR) curves were generated by tuning the non-max suppression threshold for the evaluation. A high area under the PR curve shows that the detector is returning accurate results (high precision) and a majority of all positive results (high recall). Based on performance and computing costs, the YOLOv3 model was selected for subsequent analysis. Fig. 4 shows the PR curves.

4.3. Post-processing filters

Post-processing filters are applied to improve detection accuracy and eliminate detection errors such as duplicate detections, oversized detections, or detection at impossible locations. Four post-processing filters are developed for this study:

- **Remove overlapped bounding boxes for the same object:** Sometimes, the model generates multiple bounding boxes for the same object, e.g., identifying a truck as both truck and a car. Therefore, an Intersection over Union (IoU) based filter is developed. For areas of two bounding boxes \( B_1 \) and \( B_2 \), the IoU is calculated by:

  \[
  IoU = \frac{B_1 \cap B_2}{B_1 \cup B_2}
  \]

  This study uses an IoU = 0.8, that is, a bounding box with a lower-class probability is removed when the intersection over the union of two overlapped bounding boxes is larger than 80% of the union.

- **Remove incorrect size:** Any large bounding boxes (larger than 75% of the size of the input image) are removed.
- **Customize detection area:** This filter is used to remove any detected objects that appear in irrelevant or inaccessible areas (e.g., sky/building, parked lanes).
- **Significant height difference at near hyper-plane:** A significant height difference of two pedestrians standing close to each other may affect the accuracy of a social distance calculation due to the mechanism that the algorithm is built-on. A position filter is deployed to determine if two pedestrians are close in proximity. Since all pedestrians are perpendicular to the same horizontal plane (the earth), the vertical position of proximate pedestrians should be similar in the image. If the vertical position difference between the bottom lines of detected bounding boxes is lower than a threshold, then the related objects are considered close. The level of vertical position difference is calculated as:

  \[
  Diff_{1,2} = \frac{|y_{1,\text{bottom}} - y_{2,\text{bottom}}|}{\min(h_{1,p}, h_{2,p})}
  \]

  where \( y_{1,\text{bottom}} \) and \( y_{2,\text{bottom}} \) are the vertical locations of the bottom line of bounding boxes and \( h_{1,p} \) and \( h_{2,p} \) are the estimated pixel-heights of each bounding box. If \( Diff \) is lower than 0.25, two detected pedestrians are considered close to one another, and the one with the higher pixel height will be assumed to have a real height equal to the pre-set height (5.74 feet in this study).

Fig. 4. The performances of the three models tested at selected locations.
4.4. Real distance approximation

The major challenge of detecting social distancing patterns from surveillance videos is the accuracy of the measurement of the actual distance between pedestrians. Difficulties usually arise from the perspective effect and a lack of references. A common solution is to compute a homography, a matrix that represents the transformation between two planes, to morph the video frames from perspective view into a top-bottom view, then using preset objects with known measurement or existing objects with available measurement as references to compute the distances in the transformed frames (Szeliski, 2010). Field visits, measurements from Google Maps or civil infrastructure documents, and chessboard-like or point matrix calibration plates are often used to obtain reference distances.

Existing surveillance cameras at different locations usually have different views of perspective and may be repositioned at different times. For these reasons, it may be challenging to get a reliable reference to calculate distances using camera feeds. A generalized method that can be used to measure distance from multiple cameras is needed. Considering these complications, we propose a novel method (Fig. 5) to measure distance without the use of field measurement and homography computing.

First, we slice the image into several hyperplanes, which are perpendicular to the horizontal plane and the vanishing lines. Because of the perspective effect, the number of pixels in the image corresponding to a single real-world length can vary on different hyperplanes. In other words, each hyper-plane will have a specific real-pixel distance rate (RP-rate)—the “further” the hyper-plane, the larger the RP-rate.

Second, it is assumed that each person in the image is perpendicular to the horizontal plane and has the same height \( h_r \). Detection results may be affected by this height assumption, but because of the size and resolution of the videos, the impact of the height assumption is acceptable.

Next, we identify the centroids of all the bounding boxes of all detected pedestrians and estimate the pixel-distance \( l \) between the centroids of bounding boxes for each pair. The pixel-height \( h_i^p \) of the bounding box is used as the pixel-height of the detected person \( i \). For each pedestrian pair, it is always possible to find a “box” with four hyper-planes; two pedestrians belong to two of the hyper-planes, and the other two hyperplanes are perpendicular and intersect with the pedestrians (Fig. 5 upper left). The generalized RP-rate \( r_{i}^{RP} \) for any person \( i \) can be represented as:

\[
r_{i}^{RP} = \frac{h_i}{h_i^p}
\]

We then slice the line between two centroids into small enough segments with a small pixel-distance \( \Delta p \), and use each segment’s RP-rate \( r_{i}^{RP} \) to calculate the real distance of that segment \( \Delta d \):

\[
\Delta d = \Delta p \Delta r_{i}^{RP}
\]

Fig. 5. Proposed real distance approximation method.
Thus, the distance between person $a$ and $b$ can be calculated using the following formula:

$$D_{ab} = \int_a^b \Delta d = \int_a^b r_{iab} dp$$  \hspace{1cm} (5)$$

Because the hyper-planes are perpendicular to vanishing lines, and the vanishing lines are straight lines, the transfer process of pixel-height $h_p$ between two hyper-planes is correlated with pixel distance linearly, and the formula is shown below:

$$h_p = \frac{h_2^p - h_1^p}{l} \cdot p + h_1^p$$  \hspace{1cm} (6)$$

where $h_1^p$ and $h_2^p$ are the estimated pixel-heights of each person, $l$ is the estimated pixel-distance between two persons, and $p$ is the variable that indicates the pixel distance.

Finally, the real distance between two centroids can be calculated by combining equation (3), (5) and (6):

$$D = \int_0^l r_{iab} dp = \int_0^l \frac{h_2^p}{h_p} dp = \int_0^l \frac{h_2 - \frac{h_2^p - h_1^p}{l} \cdot p + h_1^p}{h_p} dp$$

$$\log \left( \frac{\frac{h_2^p - h_1^p}{l} \cdot p + h_1^p}{h_p} \right)$$

$$= h_r \cdot \log \left( \frac{h_2}{h_p} \right) - h_r \cdot \log \left( \frac{h_1}{h_p} \right)$$

$$= h_r \cdot l \cdot \log \left( \frac{h_2}{h_1} \right)$$  \hspace{1cm} (7)$$

Theoretically, it should be noted that the transfer of the RP-rate is not linear. To simplify the computation, we assume that the RP-rate is transferred linear, so the equation (7) can be roughly simplified into:

$$D = \left( \frac{h_r}{h_1^p + h_2^p} \right) \cdot l$$  \hspace{1cm} (8)$$

Then, the K-D Tree algorithm (Bentley, 1975), a space partitioning data structure for organizing points in a K-Dimensional space, is applied to quickly identify all pedestrians in close contact. The program will plot a line between the pair of pedestrians who are not following social distancing guidelines. All the frame-based outputs and the estimated distancing data will be stored in the final results after the whole video is completely processed.
5. Results and discussion

We sampled 11 key locations from the existing NYCDOT traffic surveillance system as a case study to evaluate the proposed method. These locations are spatially dispersed in the five boroughs of NYC and have different land use or sociodemographic characteristics (e.g., close to hospitals, meal distribution centers, or subway entrances). Because the cameras are sometimes repositioned to view traffic from varying directions, representative weekdays with similar camera conditions are selected for these locations.

The program is run on an instance configured with Intel Core i7-7700HQ@2.8GHz CPU, 16GB RAM, and NVIDIA GeForce GTX 1060 GPU, Windows 10 64-bit operating system. The proposed method is implemented in real-time. The average running time is around 0.13 s/frame/location without real-time visualization and 0.38 s/frame with visualization.

5.1. Detection output

The proposed framework outputs all traffic-related objects (i.e., person, car, truck, bicycle, bus) as well as the total number of pedestrians that are in close contact in each frame. It builds a bounding box around the detected objects and assigns a class name and probability. The framework also highlights the pedestrians that are in close contact with blue lines in each frame. Fig. 6 presents an example of the video processing output at one of the selected locations. Fig. 6 also illustrates the number of detected objects in the video frame, including pedestrians, cars, and buses.

The average pedestrian density and overall sociability metrics considering various distances (e.g., 3-feet, 6-feet, and 12-feet) suggested by different agencies and experts (WHO, 2020, US CDC, 2020; Bourouiba, 2016) for the sampled locations are summarized in Fig. 7 and Table 2. Pedestrian density drops to the lowest point from mid-April to mid-May, where one pedestrian or less is captured in recorded frames for more than half the time. The results show a positive skew, with the mean higher than the median, except on May 27, 2020, which displays a negative skew, and on Jun 18, 2020, which is normally distributed. A gradual increase in pedestrian volumes is observed, starting from the end of May, reaching a peak in the middle of June. Accordingly, the ratio of people following social distancing guidelines (the social distancing rate) dropped slightly from April to May for all of the suggested distances. This rate slightly increases at the beginning of June and fluctuates around the new increased value.

Using the data generated by the proposed algorithm, we compared the total number of pedestrians in close contact (distance < 12, 6, and 3 feet) and the number of newly reported positive cases in NYC over time (Fig. 8). It is apparent that when the number of daily cases (annotations in the circle) decreases, more people go outside and get in closer contact with each other, with a lower number of people complying with social distancing guidelines. There may be other underlying factors not identified in this paper that impact the decision for people to go outside in greater numbers and reduce observed distance between each other after April 15, 2020.

5.2. Heatmaps

Spatial patterns of social distancing have also been explored through the analysis of heatmaps for each site. This intends to highlight hotspots at which pedestrians are in close contact with each other. The heatmaps generated in this section use a 12-feet rule as a case study. Each pedestrian pair less than 12 feet apart is identified and clustered to generate heatmaps for selected study locations and times. It is important to note that these heatmaps do not represent pedestrian densities. Instead, they illustrate detection areas that capture high frequencies of close contact events. Constructing these heatmaps will improve the identification of social distancing guideline compliance rates by focusing on areas that show the highest occurrence of proximate pedestrian pairs. In other words, generated heatmaps serve as a visual cue to help understand and identify areas with a high risk of close contact between pedestrians (Fig. 9).

Each column in the figure below shows a different location, while each row represents the results of a different week. It can be seen

![Average Pedestrian Density Comparison](image)

Fig. 7. Box plot of the average pedestrian density of each frame among different dates. The box shows the interquartile range (IQR), which is the 1st to the 3rd quantile of the data (Q1 and Q3). The whiskers extend 1.5 times the IQR from the Q1 and Q3. The green triangles are the means of the data, and the “notches” indicate the 95% confidence interval of the median. We did not include the outliers (data points stand outside of the whiskers) in this plot, but Table 2 shows the maximum pedestrian densities for each date.
that the locations where people tend to get physically closer one another remain relatively consistent among the studied weeks, while the intensity of such occurrences may change. Increased intensity can be explained by more people going outside as the number of daily COVID-19 cases decreases. More people present on a sidewalk will increase the chances of an encounter with other approaching pedestrians.

There are other locations shifting hotspots represent new patterns. These high-risk areas are determined to be located near some critical facilities (e.g., public transit stops), which may be an indicator of the varying service frequency or the changing number of users or visitors to these facilities over time.

### 5.3. 24-Hour temporal density distribution

Based on the object detection outputs from the cameras, temporal density distribution profiles are constructed for one study location where a set of pre-pandemic data is available to investigate potential temporal pattern changes. Fig. 10 shows the 24-h
Fig. 9. Heatmaps of clustered pedestrians (distance < 12 feet). For row two to four, from top to bottom: Apr 2, 2020, May 13, 2020, and Jun 24, 2020. Due to the camera movement, the third column’s camera cannot capture the right sidewalk on May 13 and Jun 24, so the related heatmaps do not contain hot zones on the right-side.
distribution of pedestrian densities at the selected site. This location had a typically high and consistent pedestrian density throughout the day before the outbreak (as shown by the solid black line in Fig. 10). Pedestrian density remains very low in April and May and gradually increases in June after the city’s reopening. Compared to pre-pandemic levels, peak hours are shifted. In addition, afternoons (3:00 p.m.-5:00 p.m.) become the period with the highest pedestrian density for this location in May, before the city’s reopening. These temporal changes in pedestrian behavior may deserve more attention amid the COVID-19 pandemic, and appropriate response measures can be carried out (e.g., examining open street strategies, recommending staggered work hours to nearby companies). When pedestrian demand rebounds, more frequent close contact on streets is expected to occur (as already shown in earlier analysis). The public should be continuously reminded of the potential risk of exposure in crowded environments, including in the open space of urban streets.

6. Conclusions

In this study, a continuous, real-time social-distancing and pedestrian density detection system was presented. The system leverages existing video surveillance cameras along with a deep learning-based computer vision model for analyzing pedestrian density and social distancing patterns at the local level without the need for new equipment or tools. Our work extends the existing literature in the following three ways:

1. It develops a scalable, deep learning-based, real-time pedestrian density and social distancing pattern detection system.
2. It provides a generalized real-pixel distance rate method to approximate real distance between a pair of pedestrians regardless of the different perspectives of the cameras; and
3. The proposed method is used to investigate the temporal and spatial changes in pedestrian density and social distancing behavior during the COVID-19 outbreak and the subsequent recovery process, relying on an existing video surveillance system in NYC to do so.

The obtained results can be used to observe social distancing patterns and pedestrian activity in the city. The results can help government agencies make better-informed decisions (e.g., enforcing specific rules at hotspots of close contacts to mitigate contagion risk) (Bian et al., 2021), and allow the public to assess risk levels at different sites with the use of different channels such as a visualization platform (Zuo et al., 2020) or frequently published social media updates (Ye et al., 2021). In addition, the acquired social distancing data can be used as an input for developing predictive models for understanding dynamic spatiotemporal infection risk of COVID-19 in urban areas.

One of the essential tasks in detecting pedestrians in a video sequence is to localize all subjects that are human. This issue raises the need to build bounding boxes in the video feed that enclose pedestrians. In most surveillance systems using image processing, the objective is to locate pedestrians and track their motion over consecutive frames. In contrast, this study does not aim to track individuals but rather detect pedestrian density at selected intersections. To achieve this objective, static images taken at 0.5-min intervals are used to detect pedestrians at each selected intersection. The performance of image processing tools is comparatively better when it is only used to quantify the presence of pedestrians instead of tracking them.

Since the proposed approach is intended to quantify pedestrian density and distance between pedestrians, accuracy and precision are highly important, especially when assessing compliance with guidelines (e.g., determining whether social distance is less than 6 feet). It should be able to better capture actual distance between individuals than other solutions, such as using smartphone GPS for estimating the proximity of individuals. Minimizing the number of false positives can provide much more accurate results. Other concerns raised in image processing applications involve the privacy and individual rights of pedestrians. These concerns can be addressed with some additional measures, including encryption methods for the identity of individuals, maintaining transparency about the fair use of data, and receiving consent at study locations.

There are some potential exogenous and endogenous limitations associated with the presented approach. The exogenous limitations are caused by the data “pollution,” and can be remedied by increasing data quality. For example, the approach may not perform well at night, under extreme climate conditions, or when glare affects video footage. It is also limited by the quality of the input images; a higher resolution of the input data can improve the identification accuracy.

The endogenous limitations are caused by the mechanism of the approach. One major endogenous limitation is associated with the

![Fig. 10. 24-h temporal distributions of pedestrians at one sampled location.](image-url)
complicated scenarios such as overcrowded pedestrian groups. The pedestrian identification accuracy is another limitation. For example, biased positions and incomplete or oversized bounding boxes around pedestrians will introduce instrumental errors to the distance calculation. The useful information contained in the training data is one of the critical components that can further remedy this type of limitation. Different object detection models are also subject to restrictions of their internal structures and components. Moreover, the accuracy of pedestrian detection may decrease as the number of pedestrians per frame increases. This may also decrease the accuracy of distance detection among pedestrians since their bounding boxes will be placed on top of each other. However, the goal of this study is to understand whether social distancing rules are followed and how it changes over time. The problem of inaccurate distance detection may not arise when pedestrian density is low, which makes the inference of social distancing much easier. Furthermore, this study focuses on computing adherence to social distancing guidelines, which can be more efficiently and accurately reached using readily available image processing algorithms, rather than the exact distance values between individuals. The results in this study quantify when, where, and how people potentially avoid infectious contacts, and it presents a timeline of social distancing trends through the course of the pandemic. Initial results showed that when the number of daily cases decreases, more people go out and reduce social distance between each other. Data from the earlier days of the COVID-19 pandemic confirmed that individuals reduced their contact rate in response to a higher perceived risk due to the increased number of daily COVID-19 cases. Continuously tracking close contacts in the city will help reveal hotspots and provide data-driven decision-making in deploying appropriate countermeasures (e.g., planning one-way pedestrian lanes) to reduce the high frequency of close contacts.

The proposed framework is scalable as the key components (input format, backbone object detection model, post-processing filters) are changeable and extendable. The generalized real distance approximation method allows the implementation of social distance calculations under different environment conditions (e.g. cameras with different angles). The framework is capable of supporting applications under different scenarios, including online frame-based data monitoring and offline aggregated statistical analysis. The proposed framework is tested using video feeds from 11 different locations in this study. The selection of various locations provided information about the variable application of social distancing over time. The results collected from different locations showed that people tend to get closer to each other at certain locations such as bus stops, curbsides, and crosswalks, and the intensity of these occurrences changes with time. As an ongoing effort, this approach will be extended to cover 100 camera locations to continuously evaluate the changes in crowd density and social distancing practice between pedestrians. With cloud computing power, it is possible to translate these findings to action in near real-time, tracking density trends during the reopening phases and using these trends for predictive analysis (e.g., determining an optimal Open Street location or predicting future cycling rates), to assist in developing effective response strategies or to plan for similar future scenarios. Although the current study shows the implementation of the proposed framework in an outdoor environment, the same framework can be easily adopted in an indoor environment with high pedestrian activity, such as in subway or train stations.

The interior of public transport (e.g., buses and subways) is another interesting possible application. The major difference between this environment and either outdoor environments or transportation facility interiors includes the available height of the camera and the space size. Since pedestrian overlapping can significantly bias results, the limited height and space can make the distance calculation inaccurate. However, a similar analysis can be considered with some modifications if granted access to the onboard cameras from buses or subways, with a potential approach involving identifying each passengers’ head to process similar calculations to the ones outlined in this paper. The applications of similar approaches in different environments deserve more investigations in the future.

Financial disclosure

The work in this paper is sponsored by C2SMART, a Tier 1 U.S. Department of Transportation-funded University Transportation Center (UTC) led by New York University, and Ulteig, an engineering consulting firm. The authors did not receive any specific funding for this work.

CRediT authorship contribution statement

Fan Zuo: Conceptualization, Methodology, Software, Validation, Writing – original draft, Visualization. Jingqin Gao: Conceptualization, Resources, Data curation, Visualization, Writing – original draft, Writing – review & editing. Abdullah Kurkcu: Conceptualization, Validation, Formal analysis, Writing – original draft, Writing – review & editing. Hong Yang: Conceptualization, Resources, Data curation, Formal analysis, Writing – review & editing, Writing – original draft. Kaan Ozbay: Supervision, Conceptualization, Writing – review & editing, Writing – original draft. Qingyu Ma: Data curation, Writing – original draft, All authors reviewed the results and approved the final version of the manuscript.

Declaration of competing interest

Each of the authors confirms that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. The named authors have no conflict of interest, financial or otherwise.

Acknowledgments

This work was supported by the Connected Cities for Smart Mobility towards Accessible and Resilient Transportation (C2SMART)
Smieszek, T., Castell, S., Barrat, A., Cattuto, C., White, P.J., Krause, G., 2016. Contact diaries versus wearable proximity sensors in measuring contact patterns at a conference: method comparison and participants’ attitudes. BMC Infect. Dis. 16, 341.

State of New York, 2020. Governor Cuomo signs the ‘New York state on PAUSE’ executive order [online]. Accessed. https://www.governor.ny.gov/news/governor-cuomo-signs-new-york-state-pause-executive-order. (Accessed 6 April 2020).

Stehlé, J., Voisin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Régis, C., Pinton, J.-F., Kanafer, N., Van Den Broeck, W., 2011a. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. BMC Med. 9, 87.

Stehlé, J., Voisin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaggiotto, M., Van Den Broeck, W., Régis, C., Lina, B., 2011b. High-resolution measurements of face-to-face contact patterns in a primary school. PloS One 6.

State of New York, 2020. Governor Cuomo signs the ‘New York state on PAUSE’ executive order [online]. Available: https://www.governor.ny.gov/news/governor-cuomo-signs-new-york-state-pause-executive-order. (Accessed 6 April 2020).

Szeliski, R., 2010. Computer Vision: Algorithms and Applications, Springer Science & Business Media.

Udugama, B., Kadhiresan, P., Kozlowski, H.N., Malekjahani, A., Osborne, M., Li, V.Y., Chen, H., Mubareka, S., Gubbay, J., Chan, W.C., 2020. Diagnosing COVID-19: the disease and tools for detection. ACS Nano 14 (4), 3822–3835.

US, C.D.C., 2020. Social distancing [online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html. (Accessed 6 April 2020).

Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Kanafer, N., Régis, C., Kim, B.-A., Comte, B., Voirin, N., 2013. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PloS One 8.

WHO, 2020. Coronavirus disease (COVID-19) advice for the public [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public.

Xie, K., Li, C., Ozbay, K., Dobler, G., Yang, H., Ciang, A.-T., Ghandehari, M., 2016. Development of a comprehensive framework for video-based safety assessment. In: IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2016. IEEE, pp. 2638–2643.

Xie, K., Ozbay, K., Yang, H., Li, C., 2019. Mining automatically extracted vehicle trajectory data for proactive safety analytics. Transport. Res. C Emerg. Technol. 106, 61–72.

Xie, K., Ozbay, K., Yang, H., Li, C., 2019. Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learning Syst. 30, 3227–3232.

Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O., 2016. Understanding Deep Learning Requires Rethinking Generalization arXiv preprint arXiv:1611.03530.

Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X., 2019. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PloS One 6.

Zuo, F., Ozbay, K., Kurkcu, A., Gao, J., Yang, H. and Iyer, S., 2020. Microscopic simulation based study of pedestrian safety applications at signalized urban crossings in a connected-automated vehicle environment and reinforcement learning based optimization of vehicle decisions. Advances in Transportation Studies, 2(Special issue), pp.113-126.