



An integrated framework of mobile crowd estimation for the 2019, July 1st rally in Hong Kong

T. Edwin Chow¹ · Paul S. F. Yip² · Kwan-Po Wong³

Received: 9 July 2021 / Revised: 24 May 2022 / Accepted: 18 April 2023 /
Published online: 2 May 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Traditional approach of mobile crowd estimation involves counting a group of individuals at a specific place, manually, in real-time. It is a laborious exercise that can be physically and mentally demanding. In Hong Kong, a large rally can last more than six hours, making the manual count method susceptible to human errors. While crowd counting using object detection and tracking has been well-established in computer vision, such application has remained relatively small scale within a controlled indoor setting (e.g. counting people at fixed gateways in a mall). No attempt to date has applied the automatic crowd counting method to count hundreds of thousands of people along an open stretch of rally route within the complex urban outdoor landscape. This research proposed an integrated approach that combines the capture-recapture method in statistics and a Convolutional Neural Network (CNN) method in computer vision to count the mobile crowd. The research teams implemented the integrative approach and counted 276,970 people with a 95% confidence interval of 263,663 to 290,276 in the 2019, July 1st Rally in Hong Kong. This work counted the attendance of a large-scale rally as a proof of concept to fill in a gap in the empirical studies. The intellectual merits and research findings shed useful insights to improve mobile population estimation and leverage alternative data sources to support related scientific applications.

Keywords Convolutional neural network · Crowd counting · Capture-recapture · Mobile crowd estimation · Big rally

✉ T. Edwin Chow
chow@txstate.edu

¹ Department of Geography and Environmental Studies, Texas State University, San Marcos, TX 78666, USA

² Department of Social Work and Social Administration, University of Hong Kong, Hong Kong SAR, China

³ C&R Wise AI Limited, Cambridge, UK

1 Introduction

Crowd counting of demonstration and protest are receiving more attention due to recent outbreaks of rallies across the US, Europe and Asia. As rally participation is often regarded as a manifestation of opposing social or political forces, the exercise of crowd estimation often faces cynical scepticism about under- or over-estimating the count based on the estimators' perceived political agenda. Similar to the dispute over the actual head-count in the U.S. Million Man March of 1995, the organizers of June 9th, 2019 rally in Hong Kong also claimed to be well-attended with a million people, whereas the police count only contended with a peak attendance of 240,000. A week later, the organizer 'doubled down' on their count of the June 16th rally to be 2 million plus one person, in remembrance of a protester who died the day before, the police count differed with a modest estimated peak attendance of 338,000. The controversy over crowd size resulted in further distrust against "conservative estimates by the media and police" [17]. In the absence of an accurate count, the public knowledge would be compromised with guesswork and 'alternative facts' [2]. While political disputes over crowd size may never go away completely, it is a legitimate responsibility of the estimators (e.g. the rally organizers, establishment forces, researchers, etc.) to lay down the foundation of "credible estimates" by documenting the crowd counting methodology in a transparent and objective manner [17]. In an attempt to provide reliable and accurate event data, there have been some recent efforts to catalogue the low and high estimates of public protests reported by the media into a database [9].

In crowd counting, Jacobs [13] pioneered the areal-density method by simply multiplying the occupying area with the average density of the 'peak' crowd. The exact time when 'peak' crowd occurred, however, is somewhat arbitrarily defined (e.g. a certain time after the rally has formally started). Nevertheless, the capture of the 'peak' crowd is only a snapshot in time and space. It is a common technique to count a 'static' crowd which does not move, such as a mass gathering or a sit-in. Hence, this method is less appropriate for a dynamic situation whereas a moving crowd that marches over a long haul where the protesters can join in and move out at any given time and location [29].

To count a moving crowd, existing approaches can be classified into statistical, computer-vision and simulation approaches. Assuming everyone starts from the origin and leaves at the destination, the total crowd is simply the overall count in theory. A statistical approach is to count the number of protesters passing through a certain check point(s) during the rally and account for crowd dynamics of a moving crowd (e.g. early departure and late arrival) using a survey [12, 29, 30]. As the head-count is often conducted manually, counting a moving herd in practice is indeed a laborious and skilful exercise. In the real-world application, it is common to conduct a temporal sampling by counting the crowd at a fixed time interval, such as counting one every five minutes. These head-counts at a fixed time interval will then be multiplied linearly with the duration to compute the total head-count. The accuracy of this approach depends on the quality of manual count, temporal variation of crowd count between counting, and sampling bias of field survey in the adjustment of early departure and late arrival [29, 30].

To overcome possible inconsistency and erroneous counting due to human fatigue, crowd counting in computer vision deploys a [semi]-automatic approach to estimate the crowd from a still image or video. In this case, a computer algorithm relies on visual cues from a scene to identify and count human objects, so that the total count at that location represents everyone who were captured by the field of view (FOV). While this can count a moving crowd within that FOV, counting a large crowd requires careful planning

to estimate the crowd size captured by a network of non-overlapping FOV. The performance of computer vision algorithm is very sensitive to environment factors (e.g. brightness, shadow, etc.), and hence it is important to carefully setup and calibrate the camera parameters.

Besides statistical and computer vision approach, an emerging branch of crowd counting that has been underexploited is computational social science. By simulating each protester as a rule-driven agent in a streetscape environment, Torrens and McDaniel [28] first pioneered the use of agent-based model (ABM) to simulate various crowd behaviors and manifest the resulting geographic pattern in a riot setting. Using crowd simulation, ABM is an unexploited asset to generate deductive explanation to the number of protesters attended in a finite and confined landscape [18].

The aims of this research are two-fold. In addressing the inefficiency for each approach aforementioned, this study proposes an integrated approach that combines the statistical capture-recapture method and a computer-vision method using a Convolutional Neural Network (CNN) to count the mobile crowd. The overall project is also augmented by a geosimulation approach to model the complex crowd behaviours based on crowdsourced data to verify the rally count [6]. However, the geosimulation approach is beyond the scope of this manuscript and will be detailed in a separate manuscript. The advantage of an integrated approach presents an opportunity to internally verify and calibrate the count, and hence contributing to a reproducible and replicable framework that is suitable to address the controversial nature of rally count. The research teams implemented the integrative approach to count a large-scale rally as a proof of concept. As protests with hundred thousand of attendees are relative scarce events, this study presents a rare empirical study to fill in a gap in the literature. The research question to be answered by this study is – how many people were there in the 2019, July 1st Rally in Hong Kong?

This paper first introduced the science of crowd counting and common approaches, to be followed by an in-depth review in the latest development from the literature. The methodology section outlines the proposed integrative framework and the results section presents the empirical findings from counting the crowd in the 2019, July 1st rally in HK. Finally, the findings are discussed with an outlook to future work.

2 Literature review

Based on the Jacob's method of areal-density calculation, the area occupied by the crowd can be measured by field surveying or aerial photograph [20]. As the crowd is not evenly distributed over space and time, crowd density often varies a lot. Hence, a critical component in the areal-density method is to obtain a representative crowd density for calculation. An improved practice is to divide the occupying area into smaller counting grids, account for spatially-varying crowd density to calculate the total count. Environmental factors like shading may also hamper the ability to count accurately from aerial photography. It is also noted that the areal-density method can only takes a snapshot of the crowd, ideally at the time when 'peak' crowd occurs. Due to uncertain 'peak' time and crowd movement, the estimated 'peak' crowd is thought to be less than the total crowd. To count a moving crowd along a route, when and where to capture the aerialphotos would affect the calculated crowd size.

In order to estimate the protesters who left before or joined after that check point, one can deploy a post-rally telephone poll to adjust the overall count accordingly using

statistical measures [12]. Basic demographic variables are often collected in such a questionnaire to statistically adjust the weight and reflect the underlying demographic distribution of the overall population. While the *posteriori* polling method can be versatile to collect additional data about the protest (e.g. sociological and political aspects of respondents), the precise crowd size can only be estimated on a later date [3]. To get a precise count in real-time, an improvement of this method is to have two check points and an in-situ questionnaire during the rally using a capture-recapture approach [30].

The location of proposed check point(s) is critical to the success of these statistical head-count. Those who by-pass the check point(s) would be left uncounted despite statistical adjustment [30]. In recent years, the intensified conflicts between the protesters and police forces often led to rally detour or new routes to be adopted in an ad hoc manner. For example, some protesters in Hong Kong and the Black Live Matters movement in the US had adopted guerrilla tactics at times, which made it challenging to be counted accurately. Moreover, linear multiplication of sampled count over the time period assumes a steady and uniform flow of protesters, which can be presumptuous as crowd behaviours (e.g. walk velocity) can be highly fluctuating from minute to minute. Considering the complicated crowd dynamics and its interaction with the police's crowd management tactics, temporal sampling can induce some uncertainties into crowd estimation. Moreover, tallying a large rally crowd that lasts for several hours is physically and mentally demanding to human observers. Hence, the overall count is subjected to sampling error, operational error associated with human fatigue and miscount, as well as random error due to the chaotic situation.

Crowd counting by computer vision can be classified into local detection, motion tracking or holistic [15, 16, 24]. Local detection usually involves the identification for each person by the detection of a full-body monolithic resemblance or specific body parts (e.g. face detection). Motion tracking relies on not the detection of a specific object but the clustering of feature trajectories, which can be derived from histogram or clustering methods. The holistic approach “guesstimate” the crowd size with a synoptic scan of the whole region of interest without segregating individual detection or tracking.

Among the three approaches, local detection that uses shape matching of all image feature, except texture, were reported to be superior [24]. In the context of a rally, part-based detection can more useful for crowd counting as the protesters are walking slowly with various signs and outfits (e.g. hat, umbrella) in a tight space, resulting in a lot of overlapping and blocked views of a full human body in the scene. Motion tracking is a viable alternative, but tracking the trajectories of feature-points might be challenging in counting a high-density crowd [10]. Using a regression of low-level holistic features of motion flow, it is argued that this approach is also effective while protecting individual privacy [5]. In general, linear classifiers are simpler but less effective, whereas non-linear classifiers can handle a complex scene but it is relatively slow, making it not suitable for real-time detection. Recent studies have demonstrated successful training and applications of deep-learning classifiers based on CNN to estimate crowd density and count. Miao et al. [19] utilized a 3D CNN that considers both the detection of spatial features from image segmentation as well as the temporal dependence across video frames for tracking.

While it is possible to train a classifier from an inventory of benchmark dataset for crowd counting, such application mostly focused on surveillance or traffic control purposes, which are indoor and count only up to dozens of people per frame. Recent breakthroughs in CNN extended the counting of dense crowds up to hundreds of people per scene. In dense crowd counting, most studies adopted a supervised training to allow the machine-learning algorithm to reinforce the learning progress using multi-scale density maps [27] or transformer [8]. Opposite to a supervised approach, Sam et al. [25] developed

a CNN using Grid-Winner-Takes-All approach to update only the weight of the winning neurons in an iterative almost-unsupervised learning. While their CNN regressor were able to self-learn 99.9% of all parameters in the first four convolution layers, a supervised training session is still needed in the last two layers to select the appropriate features for crowd density map [25]. Moreover, the two benchmark datasets tested in dense crowd counting are limited to several thousand people presented in still images. Reddy et al. [21] also utilized a scene-adaptive guiding network to count unlabelled crowd scenes, but their testing was limited to a smaller crowd (i.e. dozens) in still images. Real-time crowd counting in the unique context of a rally where hundreds of thousands of people marching in a heterogeneous landscape remain untested. Moreover, as the environmental factors (e.g. illumination and shadowing) changes throughout the day, the model must be adaptive to the dynamic changes of an urban landscape where the rally took place. Hence, the computer-vision approach of crowd counting in a rally requires sophisticated in-situ training and real-time calibration to solve the practical problem of fine-tuning the model parameters in the field. Developing an adaptive classifier to yield reliable counting and profiling of a dense crowd remains challenging [8, 19, 21].

Despite crowd counting and tracking has been a common research topic in computer vision [16], such application has remained relatively small scale and held within a controlled indoor setting (e.g. counting people at fixed gateways in a mall, simulating a crowd evacuating from a structural building). No attempt to date has applied the automatic crowd counting method to count hundreds of thousands of people along an open stretch of rally route within the complex urban outdoor landscape. External factors, such as camera angle, illumination changes throughout the day, inclusion of various protest objects in front of the human body, as well as complex crowd behaviors in a rally contributes to many practical challenges using the computer vision approach. While automatic crowd counting can be deployed continuously throughout the entire rally without temporal sampling, its count is only reported at specific check points and hence it would not be able to account for those who left before or arrived behind the check points. The appropriate setup of monitoring locations also requires stage planning and local calibration a priori, making it less flexible to cope with complex crowd dynamics (such as a rally re-route). Therefore, theoretical and practical issues dealing with the hardware and software needed to process the computationally-intensive crowd counting for a large protest in real-time is absent in the empirical literature.

An emerging inquiry in crowd management is to simulate interactive crowd behaviors in the context of a large-scale social conflict scenario by assigning various agents, such as policemen, journalists, and different types of protesters (e.g. aggressive, “quiet”) [14]. While human behaviors are complex, chaotic, random, and stochastic, it is possible to express basic principles of wayfinding from a starting place to the destination and explore the emerging patterns of human dynamics in a rally [6]. Nevertheless, a major problem in ABM to model human behaviors was to design large-scale simulations [7]. Only limited research has attempted observing human dynamics from a large-scale simulation contextualized in an actual empirical event. As each rally is unique in a historical context, the simulation approach is the only viable option to study these real-world events in an experimental setting. For the purpose of crowd counting, however, ABM requires tremendous empirical data, including both the physical environment (e.g. streets, rally route(s), origin(s), entrances and exit gateways, destination(s), etc.) and the crowd (e.g. trajectories, walking velocity, early departure rate, late arrival rate, density, etc.) to calibrate the model parameters and warrant a proper simulation [6]. While it is possible to model any number of protesters in an ABM, simulating a larger crowd size involves exponential computational

intensity and run time. The deductive approach also made it challenging to prove or disprove a particular estimated crowd size as the “ground truth” if any. So far, no research to date has used ABM for crowd counting of a real-world rally.

Previous efforts of crowd counting often rely on one single approach, may it be the areal-density, statistical, computer vision or simulation approach [6, 19, 20, 30]. While there were a couple studies examining different crowd size numbers, they remain limited to discuss different approaches and explore possible integration [17, 24]. There has been a lack of empirical study that cross-examine their strengths and weaknesses, and if their results are comparable. It is not known how these approaches can be executed in counting a large-scale protest and the associated challenges in practice. Hence, the integrative framework proposed in this study expand the literature by exploring their supplementary information in the provision of a head-count.

3 Methodology

3.1 Study area

Hong Kong (HK) was once a British colony that transformed into a Special Administrative Region (SAR) of China under “one country two system” starting in 1997. Since the outbreak of Severe Acute Respiratory Syndrome (SARS) in 2003, July 1st marks not only the anniversary date in remembrance of the historical transition of HK sovereignty, it is also a day when many Hongkongers demonstrated their discontent about various socioeconomic and political issues onto the streets. On this day, the rally organizers typically host the crowd at Victoria Park in HK Island in the afternoon, and start a 4-km march towards either the government headquarter or the Chater Garden in Central as the final destination.

Since 2003, there have been a lot of disputes about the crowd size over the years, in which the organizers would give an optimistic account of total rally participation whereas the police would report a figure that represents the ‘peak’ crowd at an unknown time. An alarming trend, however, is that the ratio of estimated count between the organizer and police is often several folds (Fig. 1), reflecting the divergence of political landscape and

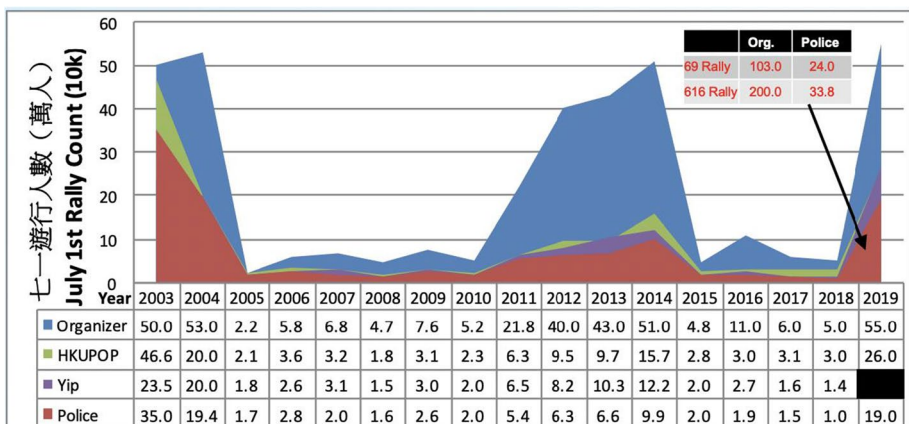


Fig. 1 The estimated crowd size of July 1st Rally in Hong Kong during 2003 – 2019

polarized society in HK. The two independent research teams, namely the HKUPOP and Yip’s team from the University of Hong Kong, have been estimating a medium count between the two political forces. In response to the controversial extradition bill, it was noted that there were two big protests took place weeks before the anniversary rally on June 9th and June 16th, 2019 in which the organizers claimed an attendance of 1 and 2 million people respectively. The police reported a much lower estimation in both instances (Fig. 1). In light of the unmet political demand, the discrepancy of estimated crowd size exacerbated the distrust against the government, and it was anticipated that the 2019 July 1st rally would draw a large crowd to fill the narrow streets in HK.

3.2 Methods

This research adopts an integrated approach that combines statistics, computer vision and geosimulation to estimate the moving crowd (Fig. 2). The capture-recapture method in statistics provides a baseline model to estimate the total crowd size whereas the deep-learning algorithm in computer vision provides an alternative source of crowd counting that is less subjected to human fatigue. The real-time manual head-count and field survey also provided field data to calibrate the AI parameters and overall crowd size as well. The geosimulation approach detailing an agent-based model of the July 1st rally is already documented elsewhere [6]. Using the observed rally time as a constrain to verify and evaluate the credibility of a specific crowd size, it involves iterative simulations of many associated parameters (e.g. crowd speed, density, patch capacity, early departure

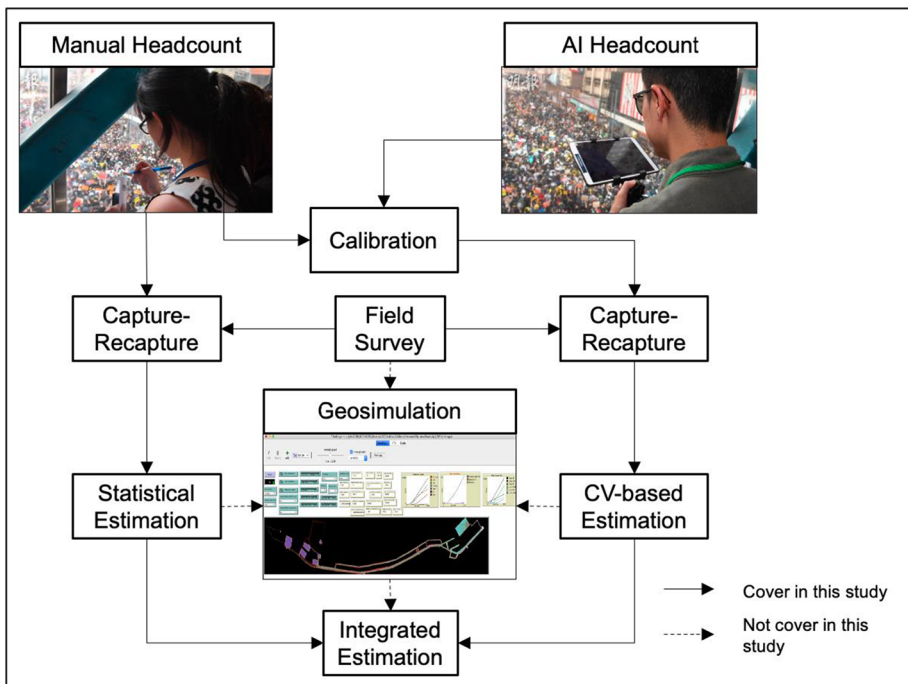


Fig. 2 The data flow diagram of the proposed integrative framework for crowd counting

rate, late arrival rate, way-finding mechanism, etc.). In this study, however, only the head-count from the statistical and computer vision approaches were compared and cross-examined to derive probable crowd size as an integrated estimation accordingly.

3.2.1 Capture-recapture method in statistics

To count the moving crowd, the research team set up two check points at the bridges on Percival Street (Point A) and Arsenal Street (Point B) along the main rally route on Hennessy Road (Fig. 3). Considering many protesters did not start the march from Victoria Park but joint in along Causeway Bay and Wan Chai, those two locations are strategically located near the start and the end of rally route where most people will be counted at either location. Ignoring those who entered and left the rally between the check points, the crowd size (N) can be estimated as:

$$N = N_A + (1 - \emptyset) \times N_B \tag{1}$$

where N_A and N_B represent the estimated count at checkpoints A and B respectively, and \emptyset denotes the percentage of individuals who passes both checkpoints. The parameter \emptyset is empirically derived from a field survey at check point B where the protesters were asked whether they passed check point A or not. By considering the proportion of protesters who were “double counted” at both checkpoints, the protesters who joint after check point A but counted at checkpoint B (i.e. late entry) is derived as a function of N_B as: $(1 - \emptyset) \times N_B$. As N_A counted those who passed by checkpoint A, the latter term in Eq. 1 is for those who joined the rally after Point A and passed through Point B (i.e. late arrival). This method effectively estimates the total crowd size (N) by double counting and spot checking for quick statistical adjustment [30]. For those who left before Point A, those who joint the rally only in between Point A and Point B, and those who joint after Point B would not be

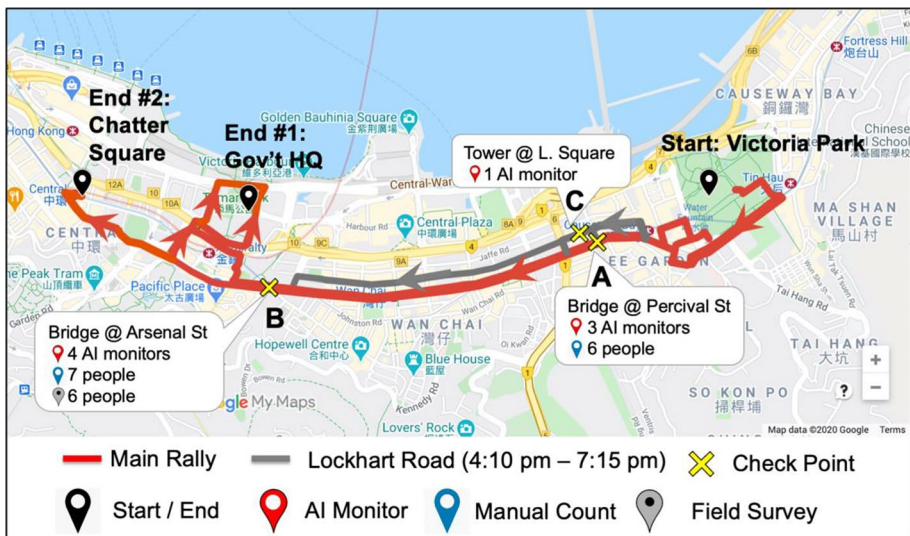


Fig. 3 The study area of July 1.st Rally in 2019

counted. The proportion of these protesters was relatively small and had limited impact on the estimate based on fieldwork observation.

The accuracy of coefficient \varnothing is subjected to sample size (m), and its standard error (SE) can be estimated as:

$$SE_{\varnothing} = \sqrt{\frac{\varnothing(1-\varnothing)}{m}} \quad (2)$$

while the standard error of N can be estimated using the delta method as:

$$SE(N) = \sqrt{\text{var}(N_A) + (1-\varnothing)^2 \text{var}(N_B) + N_B^2 \frac{\varnothing(1-\varnothing)}{m}} \quad (3)$$

At each check point, there were six to seven volunteers who manually counted the protesters for one minute every five minutes. The incremental count at a given checkpoint was a product of average estimated count among multiple volunteers and the time interval of five minutes. Besides manual count, this research also tested a [semi]-automatic and independent source of crowd counting that is less subjected to human fatigue. This statistical approach served as the baseline model for comparison purposes.

3.2.2 Deep-learning algorithm in computer vision

Based on image detection and object tracking in computer vision research, this research uses a Convolutional Neural Network (CNN) to count the mobile crowd at the check points. Extending the framework of Bochkovskiy, Wang and Liao [4], the proposed deep-learning (i.e. Artificial Intelligence) algorithm of this research analyses the video frames captured within the Instantaneous Field of View (IFOV) of a mobile device equipped with a camera. The AI algorithm detects each human object, tracks their movement, and counts each unique individual as they enter and leave a trigger zone. By tallying up every protester passing the trigger zone, the computer vision approach estimates the total crowd observed at the check points.

The counting algorithm consists of 3 components—object detection, object tracking, and object counting. For object detection, we adopted the well-known YOLOv3, which can detect human objects in an image. The detection returns a set of bounding boxes for the target object. Object tracking algorithm tracks the movement of target objects in the subsequent image frames. Tracking is performed by calculating the proximity of bounding boxes in between the image frames. Each unique object tracked through multiple image frames is labelled with a unique identity number.

To count an object, the counting algorithm checks if the object passes a red virtual tripwire from one direction to another (Fig. 4). If the object stays in the counting area which is bounded by the blue line as shown after passing the tripwire, the object will only be counted once. If the object leaves the counting area and enters again, the object will be counted again if it passes the red line in the desired direction. Therefore, objects lingering in the counting area will not be double counted. While the count of human objects passing the red tripwire was updated instantaneously, the tallied count was sent to the cloud database every 30 s.

The network was based on the unmodified version of the darknet from YOLOv3 [22]. The pre-trained model was obtained from the source (<https://pjreddie.com/darknet/yolo/>) and it was converted to the CoreML-based model for iOS-based mobile devices. An iOS

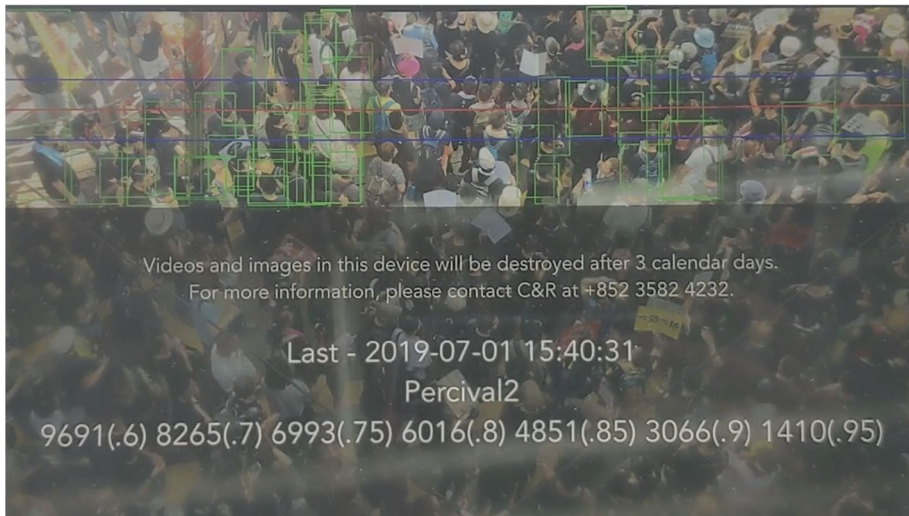


Fig. 4 The mobile app to detect, track and count human objects with seven pre-defined CNN models at the bottom of screen

mobile app was developed to run the model, including the tracking and counting algorithms. The app also used 4G connectivity to transmit counting data with the GPS location every 30 s to the central server from eight locations. In the deployment, an iPad mini with A12 Bionic chip and an 10 k mAh battery was used at each monitoring station.

The YOLOv3 object classes were originally trained based on the Microsoft COCO (Common Objects in Context) dataset which contains images of many human and other real-world objects. The model was a CNN with 53 convolutional layers. Using the COCO and ImageNet datasets, YOLOv3 was superior or on par in terms of accuracy comparing to SSD, FPN-FRCN, and RetinaNet variants (Redmon & Farhadi, 2018). Moreover, YOLOv3 was generally 2–4 times faster and could process more frames per second than these methods. Hence, it was more suitable to support real-time application in this study. In general, the accuracy of object detection depends on the image quality and clarity, including lighting condition, field of view, and distance to the target objects. The protest in Hong Kong took place within an urban canyon surrounded by high-rise buildings, which present a unique challenge for each monitor to detect, track and count human objects in a changing environment over time (e.g. shadow, brightness, etc.). As the protest event could not be rehearsed, repeated or controlled, and therefore any a priori and post priori training of the CNN model was impossible. This study adopted a heuristic approach of parameter calibration for each device using seven different pre-set CNN models with varying confidence levels (Fig. 4).

For every 30-min interval, a one-minute video was recorded and sent to the control centre to be compared with manual counts. Several analysts conducted head-counting to derive the average count for each minute-long video for Quality Control/Quality Assurance (QA/QC) purposes. These manual counts were compared with the counts estimated by the seven pre-set CNN models with different confidence levels. The confidence level of the model with the closest count was used to determine the count in the last 30-min interval. By considering the overlapping portion of the rally route covered by each monitor station, the final crowd size is the weighted sum of all counts from each group of the devices.

Due to privacy concern and the storage space required to record the passing crowd over five hours, the research team decided not to video-taped the entire duration of protest. Instead, the videos were processed in real-time to count the moving crowd. This research utilized a cloud-computing infrastructure in which the count and position data were only transmitted to the cloud-based Google Sheets in real-time for tallying, but not the images and videos.

To cover the entire road surface reserved for protesters, three and four devices were setup at check points A and B respectively. The cameras were aiming at the back of protesters to protect personal privacy from face detection. As the front of a human body might be covered by objects (e.g. signs, mask, fans, umbrella, etc.), a view at the back also enhances the chance of object detection of human features. Besides the two check points at the fixed locations, the research team also prepared an additional field team equipped with a floating device to respond to any re-routing which often happened during the two big rallies in June 2019.

In fact, many protesters broke the police barricade at East Point Road after waiting for hours to get into Victoria Park without vail on July 1st, 2019. Parallel to the main rally route on Hennessy Road, some protesters started a new marching route along Lockhart Road around 4 pm that by-passed check point A at Percival Street. Learning from similar detours of the rally crowd in the two protests in June, the research team planned ahead and were able to deploy an additional checkpoint C on Lockhart Street and counted the protesters along that “tributary” from 4:10 pm to 7:15 pm (Fig. 3).

Moreover, to better understand the distribution and magnitude of early departure, the research team also stationed dozens of volunteers at various entrances of Mass Transit Railway (MTR) as well as bus stops in Causeway Bay and Wan Chai. The volunteers would visually distinguish the unique dress code of protesters and counted those who entered the gate in between the passing of head and tail crowds in the rally. Similarly, volunteers were counting passengers leaving by the buses during the rally as well. Many bus routes were rerouted and this study only included the ones near Canal Street, the most popular transportation hubs along the rally. As many roads in the area were often blocked for crowd management purposes, the MTR and bus were the major operational transportation system in that area besides foot traffic. This was the first time to collect empirical data about early departure and its spatio-temporal distribution. The data provided useful insights to verify the adjusted counts from capture-recapture method and calibrated the geosimulation (to be reported in a separate manuscript).

4 Results

According to the public media, the rally started at 2:45 pm and finished around 9:15 pm, lasting for about 6.5 h (i.e. 390 min). At the bridge on Percival Street (check point A), marching protesters were counted between approximately 3:15 pm to 7 pm in the evening, lasting for about 225 min. It was noticed that many protesters did not queue into Victoria Park to start marching but gathered along the rally route ahead of the head crowd in Causeway Bay. The head and tail crowds passed the Arsenal Street (check point B) between 3:50 pm to 8:45 pm for about 295 min. As mentioned, the Lockhart section was occupied with protesters between approximately 4 pm to 7:15 pm. The Lockhart crowd first started pushing the barricade in spite of the police pushback, and the crowd was later released to allow free flow around 4:50 pm. The team set up an AI monitor and started counting the

crowd at barricade 4:51 pm for about 2 h. Despite the planned destination at Chater Garden in Central, many protesters detoured and marched towards the Government Headquarter instead (Fig. 3). Nevertheless, this had little effect to the count as many protesters were already counted at either check point earlier.

During the rally, the count tallied by the computer vision approach (hereafter AI count) was compared with several rounds of real-time manual head-count to identify the best CNN model and associated parameters. The data sources of these manual head-count include several 1-min video clips recorded by the field crew of AI team to the rally count as well as the count tallied by the HKU team who conducted the statistical approach (hereafter manual count). For example, it was found that the CNN model with 95% confidence level for device 3 reported 635 people passing was check point A at 6:16 pm, and it was the most accurate among the seven models by comparing with the 600 people reported by HKU team. Hence, the data series from that model was adopted to derive the total AI count at that check point during that period. This calibration process was conducted at both check points throughout the rally period.

After calibrating the CNN parameters, it was estimated that there were 198,698 protesters (N_A) counted at check point A, while there were 223,395 protesters (N_B) counted at check point B (Table 1). Due to unfavourable conditions after sunset, it was noted that the AI count after 7:19 pm was not reliable despite the best effort in calibration, and hence the AI count was not included thereafter. In other words, the AI count at N_B only reflected the head-count up until then. Although there were much less protesters after sunset, however, protesters were still marching on the rally route until the tail crowd passed around 8:45 pm. To compensate for the protesters that did not get counted after sunset at N_B , a second-order polynomial was used to fit the trend of time-series AI data with a R^2 of 0.84 (Fig. 5). Hence, the adjusted AI count at N_B was 239,648. On the other hand, the crowd passing check point C was estimated to be 22,772. It was observed from the field that most protesters marching on Lockhart Road merged into the main rally on Hennessey Road and were counted at check point B, so no adjustment to the final count was needed.

The manual counts by HKU have 202,825 and 135,525 at those check points respectively. Unfortunately, the manual head-count at N_B was stopped after the passing of tail crowd at N_A (i.e. around 7:30 pm) due to miscommunication. Similar to the trend-fitting analysis of AI count, a second-order polynomial fitted the trend of time-series data (with a R^2 of 0.90) and projected the total count to 190,387. Based on a field sample of 500 protesters (m), it was found that 68.2% of protesters (q) were counted at both check points, an equivalent of 129,844 protesters. This figure implies approximately 31.8% of the crowd (i.e. 60,543 protesters) joint the rally after check point A but were counted at check point B. Using the statistical approach and manual count as the baseline model, it was estimated that 263,368 people attended the rally.

Table 1 Raw counts using the statistical approach (i.e. manual) and computer vision approach (i.e. AI), their trend fitting and final adjustment

	Raw count		Trend fitting to fill data @ N_B after 19:19		
	Manual	AI	Manual	AI	Adjustment
N_A	202,825	198,698	202,825	198,698	200,762
N_B	135,525	223,395	190,387	288,909	239,648
Theta	68.2%	-	68.2%	-	68.2%
N	245,922	269,738	263,368	290,571	276,970

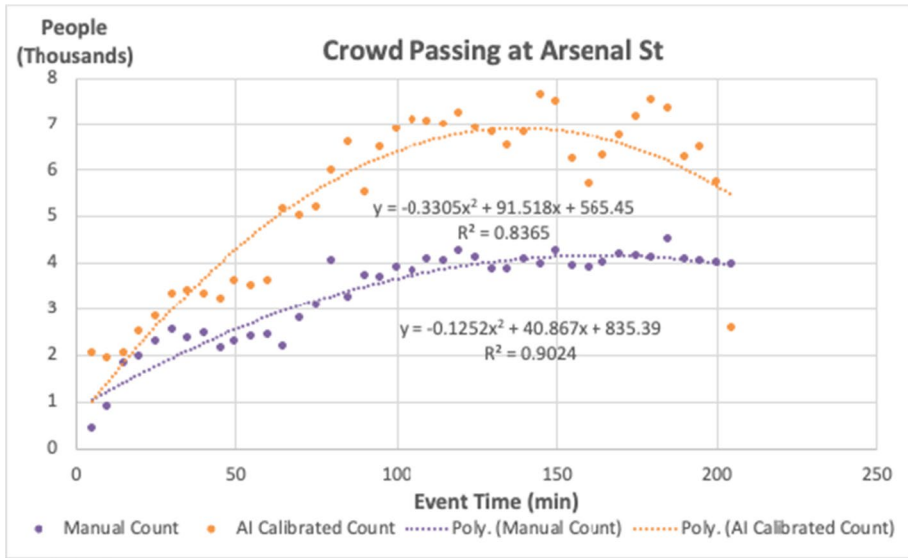


Fig. 5 Trend-fitting of the manual and AI counts at Arsenal Street

This study used the average of both manual and AI approaches for final estimation. Based on Eq. 1, the estimated total crowd size was 276,970 (Table 1). Assuming a Poisson distribution for N_A and N_B , the standard error was calculated to be 6,788.9 (Eq. 3). Hence, the 95% confidence interval (CI) for the estimate would be $276,970 \pm (1.96 * 6788.9) = (263,663, 290,276)$.

For the early departure count, the volunteers started counting protester-like passengers in the public transportation (i.e. leaving the rally area) shortly after the arrival of head crowd at their locations, which was around 3:15 pm and 3:40 pm near Causeway Bay and Wan Chai respectively. It was observed that few protesters left at the beginning of rally and increased as the rally proceeded. More protesters left from the Causeway Bay MTR stations, the location closer to Victoria Park, initially during the first two hours and later superseded by those who left from Wan Chai MTR stations around 6:40 pm (Fig. 6). As the tail crowd passed those locations as late as 8:40 pm, there were 7,562 and 13,567 protesters left from Causeway Bay and Wan Chai MTR stations respectively. Much fewer protesters, totalling up to 457, left by buses at Canal Street throughout the event. The protesters who left between the passing of head and tail crowds were estimated to be 21,586. It was noted from the field survey that the estimated crowd who left early (i.e. counted at A but not B) were 72,981 (202,825 – 129,844). Considering many protesters might have stayed around longer (e.g. rest, dinner) or left the rally area by foot or public transportation at other locations, the estimated counts based on field survey were in line with the early departure data reported by the volunteers.

5 Discussion

This study reported a total rally attendance of 276,970 with a 95% confidence interval between 263,663 and 290,276. Given that the police only estimated the ‘peak’ crowd at an unknown time and the tendency for the organizers to be optimistically overestimated, our

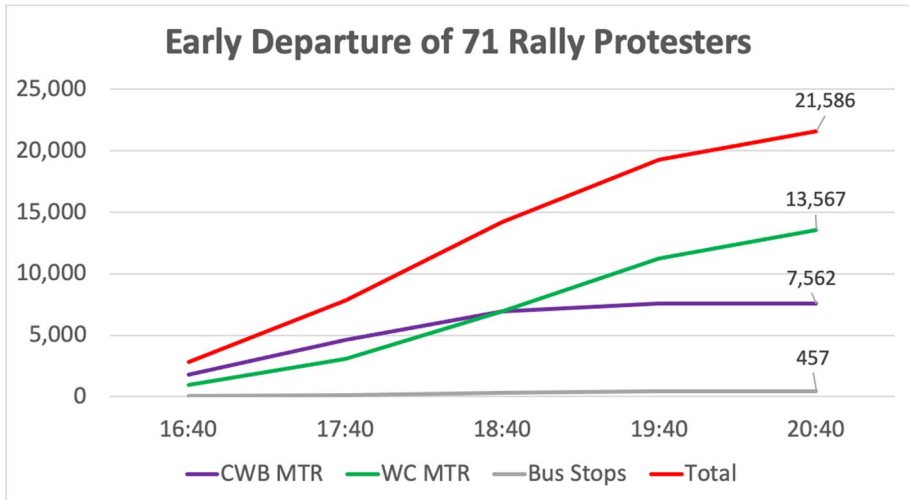


Fig. 6 The estimated protesters who left early in the 2019, July 1st Rally in Hong Kong

study estimated a number in between the ‘official’ figures of 190,000 and 550,000 provided by the police and organizers respectively. This 276,970-figure is comparable to the 265,000 estimated by the Hong Kong Public Opinion Research Institute (HKPORI), which adopted a laborious approach of manual head-count at two locations similar to this study. However, their initial count did not account for early departure or late arrival, which was later adjusted to multiply by a factor of 1.3 to 1.6 based on telephone survey from previous years [11]. The Reuters team estimated the crowd size to be about 220,000 [26] by counting the crowd based on dozens of 30-s videos irregularly throughout the rally near Arsenal Street. However, their approach used a single check point late in the rally course and excluded those who left before it. Considering the estimated early departure of 72,981 based on the field sample, the Reuters count is also similar to the overall count reported in this study. While it may be impossible to know the true attendance, the crowd size estimated from this study lies within a comparable range to these independent sources.

This research was the first empirical study to deploy a deep-learning algorithm to conduct head-count in the setting of a rally. The findings from this paper were of multiple folds – unlike most computer vision studies that counted dozens (or hundreds at most) of pedestrians, we counted hundreds of thousands of protesters. In this study, the rally took place in the busy commercial district where there were plenty bridges that provide low altitude platform to set up the cameras on tripod. Relative to the head size of passengers, iPad Mini (5th generation) has 1920×1080-pixel back camera that is more than sufficient to detect human objects. Nevertheless, the varying crowd densities associated with this rally crowd size stretched the object detection and tracking algorithms to extreme limits. As human objects could be partially or wholly blocked by another object, including but not limited to another human body part(s), umbrella, posters, etc., within the IFOV and in between the frames, the CNN approach was versatile and adaptive to detect, track and count the unique human objects in the scene. The cameras were set at a low oblique angle (e.g. within 20° from nadir) which also helped object detection/tracking and minimized any reflection from nearby objects among other ‘noises’. The empirical research also identified a limitation of the CNN approach

of crowd counting during the evening hours. It was likely that the low-light and low-contrast conditions made it hard to identify, track and count individual human objects. While this research used only the default optical lens equipped in standard iPad Mini to sense the objects, it is possible to leverage thermal infrared imageries to enhance the extraction of human objects as well. It is not known how the performance of thermal infrared imageries in object tracking for counting purposes.

Given the advancement in computer vision, digital image analysis can extract demographic attributes (e.g. gender, age) from each target object [1]. These demographic attributes may provide additional insights to understand differential crowd behaviors and weight the survey response accordingly to improve early departure, late arrival and ultimately the total head-count. It must be noted, however, that the use of AI in crowd monitoring should be exercised with cautious and sensitivity towards personal privacy and human rights. The misuse of facial recognition for mass surveillance can violate privacy rights, suppress freedom of expression and intimidate peaceful demonstration of social discourse. In fact, this was a major concern in our research design – that we decided to turn the device cameras at the back of the crowd instead of facing the protesters to protect personal privacy. The results of this study proved that this approach was sufficient for crowd counting purposes without compromising the concerns for privacy.

Moreover, this empirical study was different from previous ones as the counting took place in a complex urban canyon over an extended period of time. Unlike the controlled environment for indoor surveillance, the environmental factors of the outdoor landscape, such as the scene brightness and shadows, were changing dynamically not only over space but time as well. The ability to detect, differentiate and track each unique human object differed from bright afternoon to dark evening throughout the rally duration. This research calibrated and fine-tuned the model parameters at the beginning of rally, but did not adjust the model parameters to the change of environmental factors over time. Despite the versatility of proposed AI approach to count protesters in this rally, future research may consider machine training in near real-time and continuous calibration of model parameters throughout the rally period. Given the chaotic nature of rally counting, there were practical challenges to calibrate model parameters in real-time. For example, the telecommunication signal was unstable at times during the rally, which made coordination and real-time communication between research team members difficult.

The count for people leaving Causeway Bay and Wan Chai areas by MTR totalled up to 21,586 during the rally time (Fig. 6). This count represents about 29.5% of the field-surveyed early departure (72,981) who left early before Arsenal St from the rally. Previous studies focused on accounting for late arrivals [29, 30], this is the first study that reports empirical early departure in a large-scale rally. In addition, the data detailed the spatio-temporal distribution of where and when the protesters tended to leave. This information can be useful for crowd management and transportation planning for special events. It was found that most protesters relied on the MTR as the primary transportation mode (Fig. 6). This data would have omitted those who left the rally on foot (e.g. nearby residents), other transportation (e.g. taxi, small vans, private vehicles) or those who chose to stayed longer or went further beyond the observed extent (e.g. Tin Hau). Considering these groups might remain relatively small, there might be some degree of undercount in the final estimation of the rally crowd. The count for early departure also relied on the observers' experience to separate protesters from the general public based on apparels and appearance, which can be subjected to field verification in future studies.

6 Conclusion

The July 1st rally was one of the many protests in HK during 2019, and it was followed with more smaller protests afterwards. Since the passing and enforcement of national security law in 2020, there were no more protests in HK at the time of press. Hence, the crowd count for this rally would provide insights to this historical event.

This study demonstrated an innovative approach to count a large crowd of protesters by integrating statistical and computer vision techniques. The research team estimated a crowd size of 276,970 protesters who took on to the streets in Hong Kong on July 1, 2019 with a 95% confidence interval between 263,663 and 290,276. This scientific approach provides a transparent and objective framework to count a protest, which is intended to get a step closer to the “ground truth” for record keeping purposes.

Unlike manual count by human, whose accuracy and quality of count can be subjected to experience and fatigue over time, the AI method can be arguably to be more ‘objective’. Moreover, the AI approach can tally the count continuously whereas manual approach often relies on a temporal sampling of one over n-minutes interval (e.g. count one minute every five minutes). It is often that the manual approach would multiple the one-minute count over that period by assume the same static flow, and hence it loses temporal variation. Nevertheless, many model parameters in the AI approach still require human analyst to initialize, fine-tune and decide based on dynamic environmental factors at the site. Moreover, this research showed that the AI approach was limited during low-light and low-contrast conditions in the evening hours. Nevertheless, these two approaches were in fact supplementary and necessary in this integration. While the proposed method holds tremendous potential to count a mobile crowd, it remains an intellectual and practical pursuit to evaluate the accuracy of such count. Despite this study calibrated the AI count with manual count, both methods were subjected to some uncertainties and could be benefitted from future studies that examine the AI approach in particular within a controlled environment and a known number of human objects. More field testing (i.e. actual rallies) are also needed to better train the CNN model.

The AI approach of computer vision to detect, track and count human objects in near real-time opens up new research agenda for crowd management. Similar approach can be used to count a crowd in open settings, such as park visitation, outdoor concerts, emergency evacuation and other special events. Moreover, it is also possible to venture this approach into new applications, such as traffic monitoring, urban studies, transportation planning, etc. Despite its potential, the researchers also noted the cautious use of AI technologies (e.g. facial recognition) in crowd management with regards to privacy concerns.

Acknowledgements The authors express their appreciation for the volunteers and students who participated in the counting activities. Providing accurate, independent and verifiable estimates for crowd sizes has always been a challenging task.

Authors’ contribution Chow oversaw and coordinated the crowdsourcing of protester trajectory and field count of exit crowd for geosimulation. Yip led the statistical team in the capture-recapture approach of field counting. Wong’s team developed and tested the DNN approach of object detection and tracking for automatic crowd counting.

Funding The field work is supported by the CABLE TV and Mingpao Inc.

Data availability The data that support the findings of this study are available from the corresponding author (TE Chow). Due to privacy concerns and restrictions apply to the availability of these data, which were used

under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of C&R Wise AI Limited.

Code availability The code of the Deep Neural Network (DNN) was adopted [4, 23] and modified. The code of geosimulation is available at GitHub: <https://github.com/chowte/hkrallies>.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent to publication The authors consent to the publication of content drafted in this manuscript.

Conflicts of interest The authors have no conflicts of interest in this work.

References

1. Abirami B, Subashini TS, Mahavaishnavi V (2020) Gender and age prediction from real time facial images using CNN. *Mater Today Proc* 33(7):4708–4712
2. Barrera O, Guriev S, Henry E, Zhuravskayaa E (2020) Facts, alternative facts, and fact checking in times of post-truth politics. *J Public Econ* 182:104123
3. Beyerlein K, Barwis P, Crubaugh B, Carnesecca C (2018) A new picture of protest: The national study of protest events. *Sociol Methods Res* 47(3):384–429
4. Bochkovskiy A, Wang C-Y, Liao H-YM (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. Available: <https://www.semanticscholar.org/paper/YOLOv4%3A-Optimal-Speed-and-Accuracy-of-Object-Bochkovskiy-Wang/2a6f7f0d659c5f7dcd665064b71e7b751592c80e>. Accessed 15 Oct 2020
5. Chan AB, Morrow M, Vasconcelos N (2009) Analysis of crowded scenes using holistic features. In: *Workshop on Performance Evaluation of Tracking and Surveillance*, Miami
6. Chow T-KE (2019) A crowdsourcing-geocomputational framework of mobile crowd simulation and estimation. *Cartogr Geogr Inf Sc* 46(1):2–20
7. Conte R, Paolucci M (2014) On agent-based modeling and computational social science. *Front Psychol* 5:668
8. Do PT (2021) Attention in Crowd Counting Using the Transformer and Density Map to Improve Counting Result. In: *8th NAFOSTED Conference on Information and Computer Science*
9. Fisher DR, Andrews KT, Caren N, Chenoweth E, Heaney MT, Leung T, Perkins LN, Pressman J (2019) The science of contemporary street protest: New efforts in the United States. *Sci Adv* 5(10):eaaw5461
10. Hashemzadeh M, Pan G, Yao M (2014) Counting moving people in crowds using motion statistics of feature-points. *Multimedia Tools Appl* 72:453–487
11. HKPORI (2019) Translated title: HKPORI will provide a platform to count July 1 Rally 香港民研為民間社會提供七一遊行人數點算平台. Available: <https://www.pori.hk/press-release/2019-07-01.html>. Accessed 26 May 2021
12. Hong Kong University Public Opinion Programme (2018) July 1st Rally Crowd Estimation Project. Available: <https://www.hkupop.hku.hk/english/features/july1/headcount/2018/index.html>. Accessed 5 Oct 2020
13. Jacobs H (1967) To count a crowd. *Columbia J Rev* 6(1):37
14. Lemos C, Coelho H, Lopes RJ (2014) Agent-based modeling of protests and violent confrontation: a micro-situational, multi-player, contextual rule-based approach. In: *Proceedings of the 5th World Congress on Social Simulation*, São Paulo, Brazil
15. Liu Y, Wang Z, Shi M, Satoh S, Zhao Q, Yang H (2020) Towards Unsupervised Crowd Counting via Regression-Detection Bi-knowledge Transfer. In: *Association for Computing Machinery*, Seattle, WA
16. Loy CC, Chen K, Gong S, Xiang T (2013) Crowd Counting and Profiling: Methodology and Evaluation. In: *Modeling, Simulation and Visual Analysis of Crowds*, New York: Springer, pp. 347–382

17. Mcphail C, Mccarthy J (2004) Who counts and how: estimating the size of protests. *Context*, pp. 12–18
18. Mekni M (2014) A novel spatial behavioral approach for agent-based crowd simulation. *Int J Math Comput Simul* 8:46–59
19. Miao Y, Han J, Gao Y, Zhang B (2019) ST-CNN: Spatial-Temporal Convolutional Neural Network for crowd counting in videos. *Pattern Recogn Lett* 125:113–118
20. Raybould M, Mules T, Fredline E, Tomljenovic R (2000) Counting the herd. Using aerial photography to estimate attendance at open events. *Event Manag* 6(1):25–32
21. Reddy MKK, Rochan M, Lu Y, Wang Y (2022) AdaCrowd: unlabeled scene adaptation for crowd counting. *IEEE Trans Multimedia* 24:1008–1019
22. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
23. Redmon J, Farhadi A (2018) Available: <https://arxiv.org/abs/1804.02767v1>. Accessed 25 May 2021
24. Ryan D, Denman S, Sridharan S, Fookes C (2015) An evaluation of crowd counting methods, features and regression models. *Comp Vision Image Underst* 130:1–17
25. Sam DB, Sajjan NN, Maurya H, Babu RV (2019) Almost Unsupervised Learning for Dense Crowd Counting. In: *Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, Hawaii*
26. Scarr S, Sharma M, Hernandez M (2019) How many protesters took to the streets on July 1?. Available: <https://graphics.reuters.com/HONGKONG-EXTRADITION-CROWDSIZE/0100B05W0BE/index.html>. Accessed 26 May 2021
27. Sun Z (2022) Coarse-to-Fine Network for Crowd Counting. In: *International Conference on Electrical Engineering, Big Data and Algorithms*
28. Torrens PM, McDaniel AW (2013) Modeling geographic behavior in riotous crowds. *Ann Am Assoc Geogr* 103(1):20–46
29. Watson R, Yip P (2011) How many were there when it mattered? Estimating the sizes of crowds. *Significance* 8(3):104–107
30. Yip PSF, Watson R, Chan K-S, Lau EHY, Chen F, Xu Y, Xi L, Cheung YTD, Ip BYT, Liu D (2010) Estimation of the number of people in a demonstration. *Aust N Z J Stat* 52(1):17–26

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.