Counting people in the crowd using social media images for crowd management in city events

Gong, X.; Daamen, W.; Bozzon, Alessandro; Hoogendoorn, S.P.

DOI
10.1007/s11116-020-10159-z

Publication date
2021

Document Version
Final published version

Published in
Transportation

Transportation (2021) 48:3085–3119
https://doi.org/10.1007/s11116-020-10159-z

Counting people in the crowd using social media images for crowd management in city events

V. X. Gong[1,2] · W. Daamen[1] · A. Bozzon[1] · S. P. Hoogendoorn[1]

Published online: 12 January 2021

Abstract

City events are getting popular and are attracting a large number of people. This increase needs for methods and tools to provide stakeholders with crowd size information for crowd management purposes. Previous works proposed a large number of methods to count the crowd using different data in various contexts, but no methods proposed using social media images in city events and no datasets exist to evaluate the effectiveness of these methods. In this study we investigate how social media images can be used to estimate the crowd size in city events. We construct a social media dataset, compare the effectiveness of face recognition, object recognition, and cascaded methods for crowd size estimation, and investigate the impact of image characteristics on the performance of selected methods. Results show that object recognition based methods, reach the highest accuracy in estimating the crowd size using social media images in city events. We also found that face recognition and object recognition methods are more suitable to estimate the crowd size for social media images which are taken in parallel view, with selfies covering people in full face and in which the persons in the background have the same distance to the camera. However, cascaded methods are more suitable for images taken from top view with gatherings distributed in gradient. The created social media dataset is essential for selecting image characteristics and evaluating the accuracy of people counting methods in an urban event context.

Keywords Crowd size estimation · Social media analysis · Image processing · Face recognition · Input for crowd management

✉ V. X. Gong
x.gong-1@tudelft.nl

W. Daamen
w.daamen@tudelft.nl

A. Bozzon
a.bozzon@tudelft.nl

S. P. Hoogendoorn
s.p.hoogendoorn@tudelft.nl

[1] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands

[2] Mathematics and Computer Science, Faculty Electrical Engineering, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

Introduction

City events, such as sports matches, thematic carnivals and national annual festivals, are carried out in urban areas, and may attract a large number of people during a short time period. The scale and intensity of these events demand systematic approaches supporting stakeholders (e.g., event organisers, public and safety authorities) to manage the crowd. Stakeholders aim to reduce risks of incidents caused by internal and external threats, and maintain an acceptable level-of-service (Fruin 1971; Marana et al. 1998) of the event area. The level-of-service describes the ranges of average area occupancy for a single pedestrian (Polus et al. 1983). A higher level-of-service of the event area indicates lower density of people in that area, which is safer than the lower level-of-service that refers to high density of people. Using such information together with a set of other qualitative and quantitative interpretation of the crowd, such as sentiment (Gong et al. 2019) and composition (Gong et al. 2018a), stakeholders apply predefined measures to manage the crowd. The level-of-service information can be inferred from density of people in that area, which can be further calculated using the number of people in the area and the area of the event place. Taking the national holiday, i.e. the King's Day, in the Netherlands as an example. To estimate the level-of-service in a popular attraction such as the Dam Square in the Amsterdam during the King's Day for crowd management, we can calculate the density of crowd in that area. According to Duives et al. (2015), the density of pedestrians in an area during a certain period can be defined as the number of attendees per unit area. Once the level-of-service of an area is estimated, crowd managers may apply certain measures to avoid incidents such as overcrowding in that area. Therefore, the number of people in the crowd is a valuable input for estimating the level-of-service in event area, and further for crowd management.

The crowd size can be estimated by stewards, or using crowd observation and monitoring algorithms based on data from surveys (Fang et al. 2008), cameras (Davies et al. 1995), counting systems (Daamen et al. 2016), mobile phones (Yuan 2014; Earl et al. 2004) and public transportation systems (Luo et al. 2018; Wang et al. 2018). However, conventional methods have various disadvantages. Crowd sizes estimated by stewards and surveys contain human errors. Collecting data from cameras or counting systems can be expensive, in particular for large city events when many cameras are needed. Similar to public transportation data, sensors cannot be employed globally and may involve privacy issues. In the meantime, with the advance of technology, social media is widely used by people in city events. People on social media share their expressions by sending text and images together with timestamps and locations. Despite the drawbacks of social media data such as sparsity and dependence on individuals, images on social media may be a promising data source to estimate the size of a crowd. Figure 1 illustrates the relationship between the crowd size in reality and the crowd size estimates from social media images sent by the crowd during city events.

Counting people in an image is extensively studied (Chen et al. 2013; Idrees et al. 2013; Lempitsky and Zisserman 2010; Zhang et al. 2015, 2016b). More than 50 methods are reviewed in surveys (Sindagi and Patel 2018; Saleh et al. 2015; Ryan et al. 2015) about estimating amount of people in the crowd from a single image. However, there is no work investigated crowd size estimation from social media image concerning diverse characteristics in the context of city events, and no work proposed corresponding dataset for exploring such problem. There is a lack of an in-depth understanding of which methods are

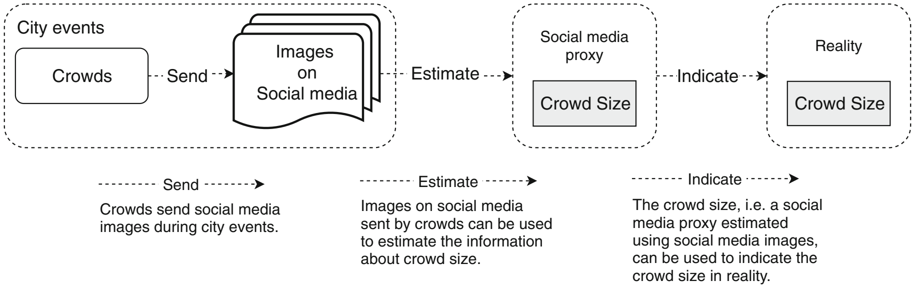

City events
Crowds
Send
Images on Social media
Estimate
Social media proxy
Crowd Size
Indicate
Reality
Crowd Size
Send
Crowds send social media images during city events.
Estimate
Images on social media sent by crowds can be used to estimate the information about crowd size.
Indicate
The crowd size, i.e. a social media proxy estimated using social media images, can be used to indicate the crowd size in reality.

Fig. 1 Illustration of relationship between crowd size in reality and the crowd size estimated using social media images sent by crowds in city events

most effective in this context, and whether their performance will be affected by the diverse image characteristics.

These research gaps lead to two research questions:

- RQ1. Which methods are suitable to estimate crowd size using social media images in the context of city events?
- RQ2. What is the effect of image characteristics on the accuracy of the crowd size estimation methods described in RQ1?

To answer these research questions, we first set the scope of the crowd size to be measured in this research. Then, we selected a set of methods from diverse categories introduced by Saleh et al. (2015) to estimate the crowd size from images. Based on the knowledge we gain from investing the existing methods, we propose new methods to estimate crowd size. We leave it for future work. In order to test the effectiveness of each method, we constructed a dataset with the number of people in the crowd annotated as well as diverse image characteristics. Next, we applied each selected method on the constructed dataset to estimate the crowd size of each image. We then analyzed the impact of image characteristics on the performance of each method. Finally, we selected the most promising method and identified under which image characteristics it is most effective.

This paper is organised as follows. We present the related works in the next section. Then, we set the scope of the crowd size to be estimated, followed by an introduction of the research methodology to examine the estimation accuracy of different methods. Further, the image characteristics from social media data are described, as well as the potential crowd counting methods. The next section introduces the social media dataset, followed by the experimental design to examine the effectiveness of the selected methods for the constructed dataset. The resulting experiment findings and corresponding analysis are then shown, followed by discussion. The paper ends with conclusions.

Related works

In this section, we review related works about counting people in the crowd. As such works propose and investigate crowd counting methods in a certain context and use datasets to assess their effectiveness, we review such works in terms of their methods, context and dataset.

Sindagi and Patel (2018) compared 27 crowd counting methods in their survey. They classify these methods into two types, i.e. traditional approaches and convolutional neural network (CNN) based approaches. The traditional approaches detect people through handcrafted features, such as facial features, head and interest parts, extracted from an image (Liu et al. 2019; Zhou et al. 2015; Chen et al. 2012). Traditional approaches can be further classified to detection-, regression- and density estimation-based approaches. While the CNN-based approaches count the number of people in an image using the advancements driven primarily by CNN network (Wang et al. 2020; Sindagi et al. 2019; Ma et al. 2019; Sindagi and Patel 2019b; Jiang et al. 2019; Shi et al. 2019; Sindagi and Patel 2019a). The CNN based methods can be further classified based on network property and training approach. Experiments (Sindagi and Patel 2018) show that CNN-based approaches reach better performance in high dense crowds with variations in object scales and scene perspective.

Saleh et al. (2015) reviewed a set of approaches related to crowd counting and density estimation. They categorise the reviewed approaches into direct approaches and indirect approaches. The direct approach (i.e. object-based target detection) count people by identifying individual segments in the crowd and then accumulate them as the result. While, the indirect approaches (e.g. pixel-based, texture-based, and corner points-based analysis) count crowd with machine learning algorithms or statistical analysis, which are considered to be more robust compared with direct methods.

Ryan et al. (2015) evaluated 22 crowd counting methods in their survey where they categorise those works as holistic, intermediate and local methods. The holistic approaches describe each image using global image features and then map these features with a crowd size estimate through a regression or classification model. In contrast, the local approaches use local features in an image to identify individuals and then accumulate them as the result. The intermediate ones Accumulates information about local objects into histogram bins, and this information is represented at a holistic level.

Though the three classifications have different names, they are similar in terms of the crowd counting approaches (i.e. mechanism for counting people) and whether neural networks are used. For instance, the category of traditional methods (Sindagi and Patel 2018) is similar to the category of direct methods (Saleh et al. 2015) and local methods (Ryan et al. 2015). The category of CNN-based methods are similar to the category of indirect methods and holistic methods. In this regard, we follow the classification proposed by Saleh et al. (2015) in this research, i.e. direct methods and indirect methods, to select methods for counting the crowd size.

Crowd counting analyses are also carried out in various context, e.g. counting people for crowd management through video surveillance on pedestrian walkway (Zhang et al. 2015), in shopping mall (Idrees et al. 2013; Chen et al. 2013), in city area (Li et al. 2018; Xiong et al. 2017; Sam et al. 2017), for violent behaviour detection in diverse environment (Marsden et al. 2017). Moreover, Crowd counting approaches show different effectiveness in diverse context according to previous studies. The holistic crowd counting approaches, such as M-CNN (Zhang et al. 2016b) and CNN-boosting (Walach and Wolf 2016), are more adept than local approaches in low dense pedestrian walkway (Chan et al. 2008). Whereas CNN-based methods, e.g. Cascaded-MTL (Sindagi and Patel 2017) and Switching-CNN (Sam et al. 2017), outperform others in the context of city events (Zhang et al. 2016a, b), as in such context crowds are in high dense levels with variation in scene perspective. However, there are no works that examine crowd counting in the context of city-event using social media images.

In the meantime, a set of datasets are proposed and employed in crowd counting analysis (Wang et al. 2020; Chan et al. 2008; Chen et al. 2012; Idrees et al. 2013; Zhang et al. 2015, 2016a, b). These datasets are diverse concerning the dense level and scene variation across images. UCSD dataset (Chan et al. 2008), as among the first dataset created for such research, contains 2,000 images with a total of 49,885 annotated pedestrians, captured by video surveillance in a pedestrian walkway. Though a large number of images it contains, it has low-density crowd with an average of around 15 people in an image which is insufficient for crowd counting analyse in a highly dense environment. With effort in improving dense level diversity, Chen et al. (2012) proposed Mall dataset which contains 2,000 images with 62,325 people annotated with an average of 33 people in an image captured by a shopping mall video surveillance. However, both of these two datasets are less variable in terms of scene perspective as they are all captured by fixed video surveillance. Idrees et al. (2013) increased scene variation in their UCF_CC_50 dataset by collecting 50 images from the Internet with a set of keywords e.g. concerts, protests, stadiums and marathons. Though it has various density level the too few images it has made it insufficient for training and testing machine learning-based methods. Zhang et al. (2016b) introduced a dataset containing 1198 images with 330,165 annotated heads, among which 482 images are randomly chosen from the Internet and the rest of images are captured from video surveillance in the street of metropolitan areas in the city. It is widely used in crowd counting and density estimation researches as it contains diverse density levels with rich variation in scene perspective across images. However, there is no dedicated dataset proposed with images captured actively by the individuals themselves, such as from social media, in the context of city events.

Definition of crowd size

In this section, we define the crowd size that will be estimated, consisting of crowd size levels and specific numbers of people for less populated environments.

Crowd size levels

As indicated in the previous section, the crowd size information is essential for estimating the level-of-service in an area (Marana et al. 1998) for crowd management. The crowd size during city events is diverse, e.g. it can be small in the beginning, and become large during the peak of event activities. When the number of people is large in an image, it is difficult to get the ground truth data as manual counting becomes error-prone. Thus, in this research, we estimate the crowd size in different levels (categories) where each level corresponds a range of the number of people in an image.

Jiang et al. (2014) categorise crowd size into five levels, i.e. 0–10, 10–30, 30–60, 60–100, >100 persons. However, the number of people captured in the camera monitoring of a fixed area is far less diverse than the number of people in social media images, e.g. social media selfies may only contain a few people, while some panorama pictures may contain a large number of people in a square or on the street in city events. Therefore, we adjusted the categories of crowd size used by Jiang et al. (2014) into a larger scope.

We performed a prior investigation on the pilot social media dataset, introduced later. We found that around 30% of the images do not contain people, and around 50% of the images contain less than 20 people, most of which are selfies and group pictures. Therefore, we set the first crowd size level to contain no people in images, denoted as 0, and the second level containing a number of people between 1 and 20, denoted as 1. Furthermore, we define 3 levels with a number of people above 20. Though the scoping of levels 2, 3, and 4 is different from the categories used in Jiang et al. (2014), this difference of scoping has less impact on this research as the proportion of images containing more than 20 people is less than 20% in social media images.

We summarise the scoping of crowd size level as follows:

- Level 0 denotes the number of people is 0, i.e., no people.
- Level 1 denotes the number of people between 1 and 20.
- Level 2 denotes the number of people between 20 and 100.
- Level 3 denotes the number of people between 100 and 250.
- Level 4 denotes more than 250 persons.

Number of people in the crowd in less populated images

In a less populated environment, where the number of people is less than 20 (level 1), it is possible to count the specific number of people in an image and derive the ground truth data. Thus, exploring the accuracy of the methods for level 1 can be more precise. Therefore, in this research, we also estimate the specific number of people in less populated images, i.e. images with crowd size level 1.

Research approach

In this chapter, we introduce the research approach and assessment metrics for answering the research questions. We illustrate each step of the research approach shown in Fig. 2.

To investigate the first research question, i.e. the effectiveness of methods in crowd size estimation using social media images in city events, we select a set of crowd counting methods with diverse techniques reviewed in surveys (Sindagi and Patel 2018; Saleh et al. 2015). To estimate the crowd size using selected methods, we construct a social media dataset of images that are collected during city events considering various aspects, such as event topics, editions, length of duration, cities and area in the city. The selected methods are applied on the annotated dataset, yielding the estimated crowd size for each image. The accuracy of crowd size estimation is calculated by comparing the estimation result to the

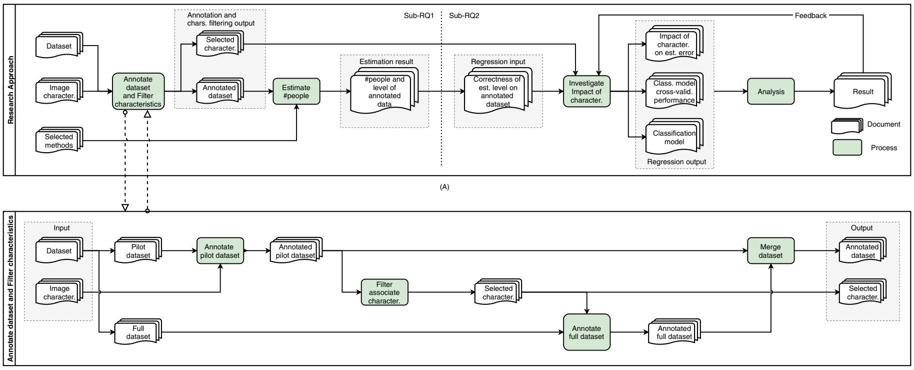


Fig. 2 The process of exploring the effectiveness of the selected methods on crowd size estimation using social media images in city events, and investigating the impact of image characteristics on the estimation results. The symbols in green denote process steps. The symbols in grey denote input and output. (Color figure online)

ground truth, which indicates whether the crowd size level of an image is the same as the ground truth.

To answer the second research question, i.e. investigate the impact of image characteristics on the crowd size estimation of different methods, we generated a set of image characteristics from both the crowd management perspective and the social media image perspective. Image characteristics generated from crowd management include conditions such as indoor or outdoor, and the urban environment where pictures are captured such as square, street, canal, and park. Image characteristics generated from social media image perspective consist of characteristics may affect crowd counting effectiveness, such as people_present, view, and selfie_face.

We zoom in to the dataset annotation and image characteristics filtering in Fig. 2b. As the image characteristics generated in the previous step may contain highly correlated characteristics, we perform a characteristic selection procedure to filter out high correlated image characteristics. To do so, we randomly selected a set of images from the total dataset as a pilot dataset and annotated these with values of images characteristics and the crowd size as ground truth. After checking the correlation on all characteristics annotated in the pilot dataset, the least correlated image characteristics are screened out. The full dataset is then annotated with the selected image characteristics and the crowd size. The output of this sub-process is the annotated dataset merging the Pilot dataset and the Full dataset, named Total dataset.

To investigate the impact of image characteristics on the accuracy of the selected methods, we train a classifier using a logistic regression algorithm for machine learning (Dreiseitl and Ohno-Machado 2002) for each method on the value of image characteristics with the correctness of the crowd size level estimation. This step outputs a set of impacts (coefficient) for each image characteristic. It also produces a classification model as a by-product with the average performance of the model calculated from the cross-validation. We analyse these outputs and provide it as a feedback to improve the investigation of the impact of image characteristics on effectiveness of methods in crowd size estimation.

Comparison metrics

We use a set of comparison metrics to analyse the accuracy of methods in crowd size estimation and the impact of images characteristics on the estimation result.

The estimation accuracy of methods

The estimation performance is assessed using the estimation accuracy, which is calculated for each method i. The estimation error A_i is calculated by the amount of correctly identified images M_i^{true} divided by the total sample size M_i, see Eq. 1. We are aware of the drawbacks of this measure with respect to the (hard) boundaries of the crowd size levels, e.g. assuming an image contains 99 people, i.e. ground truth crowd size level 2, while the estimated number of people is 101, i.e. crowd size level 3. Though the difference between ground truth value and the estimated number of people in the image is small, the estimated crowd size level is incorrect, which seems to be an overreaction. To compensate this, we also check the ground truth and estimation close distance in adjacent levels.

To explore the insights of the estimation performance of different methods, we further show the distribution of estimated crowd size levels compared with the ground truth in Table 7, and the distribution of the estimated number of people in crowd size level 1 compared with the ground truth in Fig. 6.

$$A_i = \frac{M_i^{true}}{M_i} \tag{1}$$

Classification performance

The investigation of the impact of image characteristics on the accuracy of the methods in crowd size estimation outputs the impact of each image characteristic, a classification model, and the cross-validation performance of this classification model. The performance of the classification model is indicated by the cross-validation performance. This cross-validation performance of the classification model is measured with metrics for binary classification, i.e. *Precision*, *Recall* and *F1_Score* (Powers 2011). *Precision* refers to the percentage of classified results are correct among all classified results, while the *Recall* refers to percentage of correct items have been classified among all correct items. The *F1_Score* is simply the harmonic mean of the precision and the recall.

Social media image characteristics

In this section, we identify a set of scene characteristics of the images (in the following referred to as Image Characteristics) to investigate their impact on the accuracy of crowd counting methods. The image characteristics consist of requirements from crowd management such as indoor/outdoor and urban environment shown in each image (as the purpose of this research is to provide information about crowd size for crowd management), and characteristics of images posted from social media in terms of, e.g. image type (selfie or group picture) and distribution of crowds, that may affect the performance of crowd counting methods. The image characteristics are further categorised to three types, i.e. global, frontend and backend image characteristics, which will be introduced in the following sections. The detailed definitions of image characteristics are given in Table 1, and corresponding examples are shown in Fig. 3.

Crowd management perspective characteristics

As in this research, we investigate the effective of methods counting people for crowd management in city events carried out in an open space (denoted as outdoor), such as in the street, sports stadium, and conference centre, rather than in a closed space (denoted as indoor) such as in a room or shop. Thus we identify images if they are taken in indoor or outdoor. In the meantime, crowd managers typically apply predefined measures to manage the crowd. The confinedness of an area where the crowd is located is relevant to select measures; do people have an area to go? Is there water or a railway track or another inaccessible area around? Therefore, we need to identify the urban environment shown in the images, such as square, street, canal, park, which may affect the effective of methods counting people in the crowd. We categorise these two image characteristics as global characteristics because they exist in all images.

Table 1 Identified image characteristics

Perspective	Category of image characteristic	Image characteristic	Definition	Values and example image(s)
Crowd management	Global	Condition	Whether this photo is taken indoor or outdoor?	Indoor: the picture is captured indoor, e.g., Fig. 3a
				Outdoor: the picture is captured outdoor, e.g., Fig. 3b
	Global	Urban environment	The place where the picture is captured, such as square, street, canal, park and others	Square: e.g., Fig. 3d, e
				Street: e.g., Fig. 3c
				Canal: e.g., Fig. 3h
				Park: e.g., Fig. 3b
				Others: e.g., Fig. 3a, f, g, i
Social media	Global	People present	Whether this image contain people	Yes: the picture contains people, e.g., Fig. 3a–c
				No: the picture contain no people, e.g., Fig. 3c
	Global	View	The view of the camera to people, i.e. top, parallel, or between top and parallel	Top: the people in the picture is captured from top, e.g., Fig. 3a, 3d
				Parallel: the people in the picture is captured in the same level with the camera, e.g., Fig. 3b, f
				Between: the people in the picture is captured between top and parallel, e.g., Fig. 3e
	Front	Has selfie	Whether this image contain selfie	Yes: there are selfie people in the frontend of the picture, e.g., Fig. 3b, f
				No: there is no selfie people in the frontend of the picture, e.g., Fig. 3a, e
	Front	Selfie face	The different types of faces captured in this image, such as full, partial, back, mixed, or none face	Full: full face is shown in the picture, e.g., Fig. 3f
				Partial: only part of the face is shown, e.g., Fig. 3g
				Back: the back face or hack head is shown, e.g., Fig. 3h
				Mixed: mixed faces, e.g., Fig. 3b
				None: such as only part of the body is shown, e.g., Fig. 3i

Table 1 continued

Perspective	Category of image characteristic	Image characteristic	Definition	Values and example image(s)
	Back	Has gatherings	Whether there are people in the backend of the picture	Yes: there are people in the backend of the picture, e.g., Fig. 3e, j No: there is no people in the backend of the picture, e.g., Figure 3b, f, h, i
	Back	Gatherings distribution	The distribution of the gatherings in the backend of the picture, i.e. fixed, gradient	Fixed: gatherings in the backend of the picture are in the same distance from the camera, e.g., Figure 3j, m Gradient: gatherings in the backend of the picture are gradually far away from the camera, e.g., Fig. 3d, e, l
	Back	Gatherings clarity level	The clarity level of people who is the most clearest one in the gatherings in respect to be identified	A: Very clear. The face of each person is shown clearly. The detail of features on the face can be identified, e.g., Fig. 3f B: Clear, the face of each person is clear. The features on the face can be identified, but the detail of features can not be identified, e.g., Fig. 3a, b, g, h C: Clear. The face of each person is still observable but the features on the face is not distinct. The shape of each person is clear, e.g., Fig. 3e D: Less clear. Each person is only shown as a shape. The face is not observable, e.g., Fig. 3j E: Very unclear. Each person is shown as a dot. Their face and shape is totally unobservable, e.g., Fig. 3d

**Frontend and backend* there are two layers in the picture where people are present, that is, in the Frontend layer and in the Backend layer, illustrated in Fig. 4b

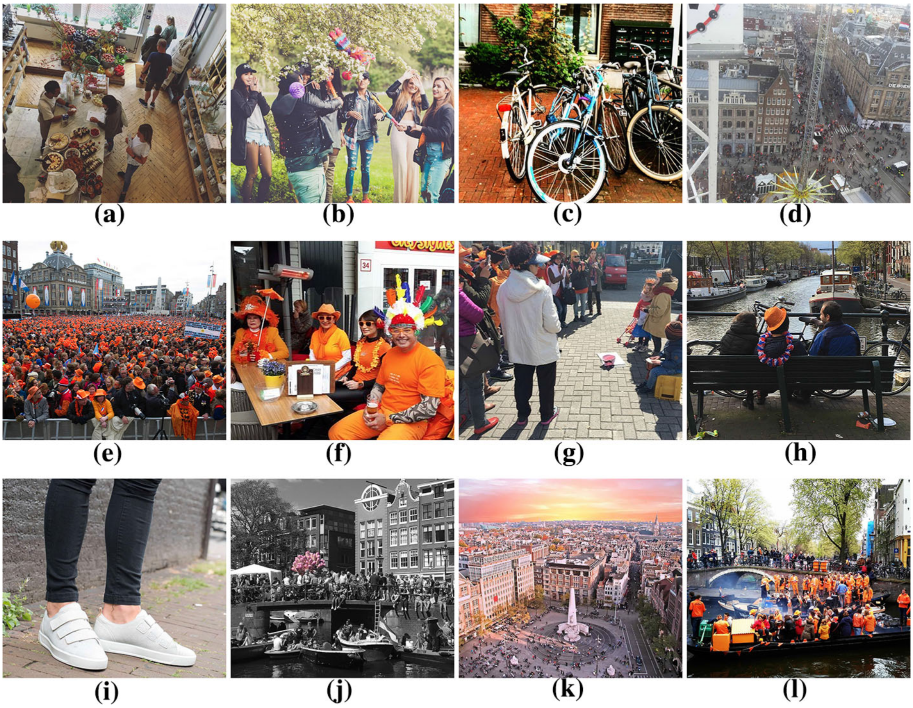
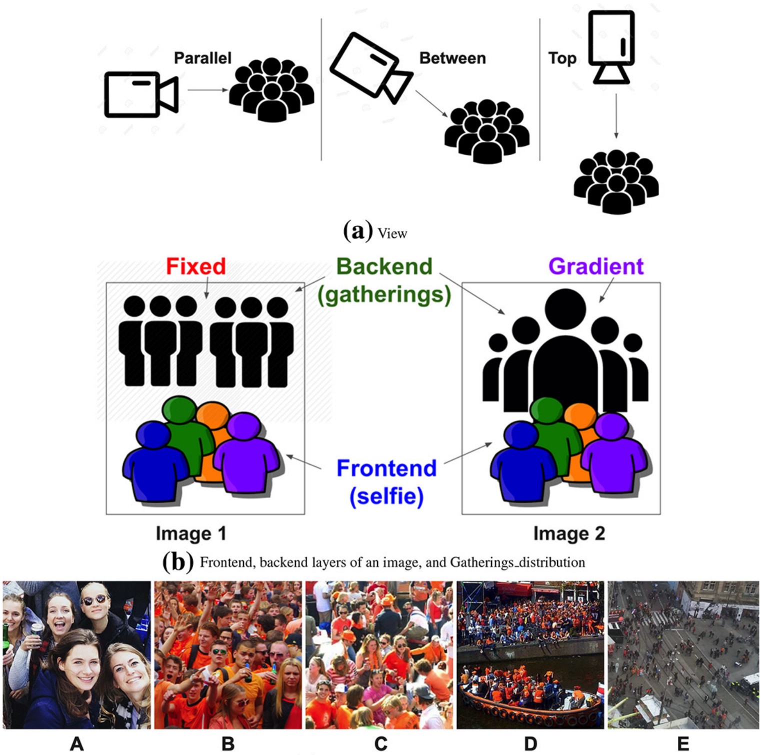


Fig. 3 Sample of images with diverse characteristics, as listed in Table 1. These images are collected from social media in city events of King's Day 2016–2018, and Europride 2017, both in Amsterdam, the Netherlands

Social media perspective characteristics

To identify characteristics on social media images which may affect the effective of counting people methods, We reviewed social media images from the pilot dataset (which will be introduced in the section Data Collection). As all selected methods (which will be introduced in the section Selection of Crowd Size Estimation Methods) counting people based on the content of images, therefore we focus on characteristics of images content, rather than other information of images, such as meta-data. Firstly, we found that most of reviewed images contain people. Among all these images, people are captured by camera from different viewpoints, such as Top, Parallel, and between Top and Parallel, illustrated in Fig. 4a, which may affect effectiveness of methods which counting people through identifying shapes and faces. These two image characteristics are also categorised as global image characteristics as they exist in all images. In addition, we found that people are present in the two layers in the picture, i.e. the frontend layer which close to the camera lens, and the backend layer which far away from the camera lens, illustrated in Fig. 4b. Images with people in the frontend are normally selfies. People in the backend of an image are so-called gatherings (denoted as gatherings). The different size and shapes of people in the frontend and backend may affect effectiveness of people counting methods. Based on the viewpoints and layers where the people are captured, images containing people can be further categorized into three types, i.e. only selfie, selfie with gatherings, and only



(a) View

(b) Frontend, backend layers of an image, and Gatherings_distribution

(c) Gatherings_clarity_level

Fig. 4 The schematic interpretation of image characteristics in terms of views, has_gatherings, gatherings_distribution, and gatherings_clarity_level, listed in Table 1. Social media images are collected from social media in city events of King's Day 2016–2018, and Europride 2017, in Amsterdam, the Netherlands

gatherings. For selfies containing people in the frontend, the face of people may be diverse, e.g. Full face, Partial face, Blocked face, Back face (or Hack head), or No face (i.e. only show body rather than face). This may affect faces-based methods in counting people. For images containing gatherings, the distribution of gatherings can be divided into two types, i.e. Fixed and Gradient, illustrated in Fig. 4b. Fixed gatherings indicate that the people have a similar distance to the camera. Gradient gatherings, on the contrary, have different distances to the camera, with smaller people having a longer distance. The different distributions of gatherings lead to different size and clarity of faces and shapes, which may affect the effectiveness of people counting methods. Among all images with gatherings in the backend, the clarity of people is different: some images are quite blurry, while others are very clear. The different clarity of gatherings in the backend may affect the effectiveness of methods which count people based on faces and textures (e.g. cascaded methods). We then categorize images with people in gatherings into different levels in terms of clarity, illustrated in Fig. 4c, ranging from A to E, where A indicates the highest clarity and E indicates the lowest clarity. In clarity level A, the faces and detail feature on faces are clear and can be identified. In level B, the face is clear. The features on the faces are observable but can not be identified. In level C, only the faces are observable while the features on the face are not distinct. The shape of people is clear. In level D, each people is only shown as a shape. In the level E, each person is shown as a dot. The detail rules and examples for distinguishing the levels of clarity are listed in the row of Gathering clarity level in Table 1. Further, We categorise image characteristics of Has selfie and Selfie face types as frontend image characteristics as they are exist in the frontend layer of images. While, the image characteristics of Has gatherings, Gatherings distribution, and Gatherings clarity level are categorised as backend image characteristics, because they exist in the backend layer of images.

Selection of crowd size estimation methods

In this section we select the methods to perform the crowd size estimation on social media images. As indicated in the introduction, there is no existing literature comparing the performance of crowd size estimation methods using social media data in the context of city events. However, counting the number of people in an image is not a novel problem. Many works discussed this topic and proposed methods to solve this problem, see (Chen et al. 2013; Idrees et al. 2013; Lempitsky and Zisserman 2010; Zhang et al. 2015, 2016b). More than 60 methods are reviewed in surveys about counting people from images (Sindagi and Patel 2018; Saleh et al. 2015; Ryan et al. 2015). As introduced in Introduction Section, these methods can be categorized with respect to different approaches, i.e. direct approaches and indirect approaches. The methods using direct approaches identify persons using handcrafted features in an image and accumulate them as the amount of people in an image. The handcrafted features refer to properties derived beforehand by human experts using the information present in the image itself, such as the face, head, shoulder and legs of people (Nanni et al. 2017). The methods with indirect approaches count people using non-handcrafted features applied with learning algorithms or statistical analyses. The non-handcrafted features are also called learned features, which is learned by machine learning algorithms using data rather than handcrafted features.

In this research we select several methods to count people from images. The selection criteria are as follows: the selected methods should be 1) diverse in mechanism, 2) diverse in specific features used for identifying and counting people, and 3) should have a high performance in comparison to related methods. To meet the first criterion, we select both direct and indirect methods. For direct methods, we further consider methods with different features for identifying and counting people, such as faces and objects. We select Faceplusplus (Zhou et al. 2015) and Darknet Yolo (Redmon and Farhadi 2017), which identify people through face recognition and object recognition, respectively, and reach high performance (Zhou et al. 2015; Redmon and Farhadi 2017). The selected indirect methods are convolutional neural network based Cascaded methods (Sindagi and Patel 2017) with version A and B, which reach significantly better results in comparison with related methods (Sindagi and Patel 2017). The details of the selected methods are listed in Table 2 and described as follows.

Face recognition: Faceplusplus (Face++)

Faceplusplus (Face++) is a method widely used for identifying people by their faces (Zhou et al. 2015). The face recognition model is established based on a deep convolutional neural network which is trained with 5 million labelled faces with about 20,000

Table 2 Selected methods for counting people in the crowd

Approach	Approach category	Name	Features	Performance
Face recognition	Direct methods	Faceplusplus	Face of people	99.50% accuracy in Wild (LFW) test
Object recognition		Darknet Yolo	Shape of people	Outperformed related methods
Convolutional neural network machine learning	Indirect methods	Cascaded A	Learned features	Lowest MAE in random pictures about city events
		Cascaded B	Learned features	Lowest MAE in pictures of busy streets in city events ∼

**MAE* mean absolute error

individuals. It reaches 99.50% accuracy in the test of recognizing faces in the database Labelled Faces in the Wild (LFW) (Zhou et al. 2015), a database of faces designed for studying the problem of unconstrained face recognition. This method detects faces in each image and provides the amount of faces in each image, based on which we can calculate the crowd size of each image.

Object recognition: Darknet Yolo (you only look once)

Darknet Yolo (You Only Look Once) is a state-of-the-art, neural network based machine learning method for real time object detection (Redmon et al. 2016; Redmon and Farhadi 2017). It is constructed based on the Darknet, an open source neural network framework (https://pjreddie.com/darknet/). It can recognise a large number of objects including persons and reaches a mean average precision of 78.6 on the PASCAL Visual Object Classes Challenge 2007 (PASCAL VOC 2007), which outperformed other algorithms widely used in this fields such as Fast R-CNN, SSD300 and SSD500 (Redmon and Farhadi 2017). This method detects people and exports the number of people in each image.

CNN-based: cascaded methods

The cascaded method is a state-of-the-art, convolutional neural network based (CNN-based) machine learning method for estimating the number of people in a high density context in an image (Sindagi and Patel 2017). It consists of two trained models, i.e. Cascaded A and Cascaded B. Both models are trained with different parts of the Shanghai Tech dataset (Zhang et al. 2016b) which contains 1,198 annotated images with a total of 330,165 people. Cascaded A is trained with part of images in the Shanghai Tech dataset which are randomly crawled from Internet about the Shanghai Tech event and most of them have a large number of people. While, the Cascaded B is trained with part of images in the Shanghai Tech dataset which are taken from busy streets of metropolitan areas in Shanghai during the Shanghai Tech event (Sindagi and Patel 2017; Zhang et al. 2016b). According to the comparison (Sindagi and Patel 2017), the performance of both cascaded models outperformed popular methods used in this fields such as MCNN (Zhang et al. 2016b), Idrees (Idrees et al. 2013; Walach and Wolf 2016; Zhang et al. 2015). The Cascaded methods estimate number of people in each image. Then, for each image the crowd size level corresponding to this number of people is assigned.

Data collection and annotation

In this section, we construct a dataset containing annotated social media images in city events, for deriving ground truth to investigate the effectiveness of different methods in crowd size estimation, as well as impact of image characteristics on this effectiveness. To collect these social media images, we first select a set of events and activities during these events. Then, social media images taken during these events and activities are collected from Instagram, the most popular image based social network (Yang et al. 2016; Gong et al. 2018a). After collecting the data, we use these social media images to derive the ground truth (annotated dataset). As we also want to investigate the impact of image characteristics, these images are also used for identifying image characteristics. We introduce each step in detail in the following.

Table 3 City events and their characteristics for constructing social media dataset

Name	Year	City	Area	Date	Term	Topic
King's Day	2016	Amsterdam	City center	27-04-2016	1 day	King's birthday celebration
	2017	Amsterdam	City center	27-04-2017	1 day	King's birthday celebration
	2018	Amsterdam	City center	27-04-2018	1 day	King's birthday celebration
Europride	2017	Amsterdam	City center	29-07-2017 to 06-08-2017	9 days	LGBT* festival
Sail	2015	Amsterdam	IJ (bay) area	19-08-2015 to 23-08-2015	5 days	Nautical event
Feyenoord	2017	Rotterdam	Around stadium	07-05-2017 after 14:30	Less than 1 day	Football fan riots

*$LGBT$ lesbian, gay, bisexual, and transgender

Event selection

Social media data are collected from various city events and the activities during these events. To avoid bias in selecting city events as well as activities, We identify the requirements regarding the events and activities considering diversity in terms of cities, event characteristics, and their major activities. The selected events are listed in Table 3. It contains 4 different city events with different topics, i.e. King's Day is a celebration of the King's birthday in the Netherlands, Europride is a LGBT festival, Sail is a nautical event, and Feyenoord represents the Feyenoord football fan riots in 2017. We have collected data during three editions of King's Day event, in 2016, 2017 and 2018. The selected events are diverse in duration, ranging from less than 1 to 9 days. They are also diverse in cities and areas in the city, e.g. while the Feyenoord event is in Rotterdam, all other events take place in the city of Amsterdam. Except for the Sail event and the Feyenoord event, which took place in the IJ area (bay area) and around the football stadium, respectively, the events took place in the city centre.

Social media data collection

Instagram, the image based social media network, is widely used by people to share pictures (Yang et al. 2016; Gong et al. 2018a). Therefore, we collect images from Instagram to construct the social media dataset. Instagram images are collected through the API of Instagram platforms using SocialGlass (http://social-glass.tudelft.nl/), an integrated system for collecting and processing social media data (Bocconi et al. 2015; Psyllidis et al. 2015). In the end, we collected 2,028 Instagram images sent during selected events.

Data annotation

The total dataset is split randomly into two sub-datasets, i.e. pilot dataset and full dataset, each containing around 50% of Instagram images collected during selected events. The pilot dataset is used for identifying and selecting image characteristics, while the full dataset will be later annotated with the image characteristics selected from the pilot dataset, and further merged with the pilot dataset to derive ground truth and investigate the impact of image characteristics on crowd size estimation. Tables 4 and 5 lists composition of two datasets in terms of crowd levels and images characteristics.

Pilot data annotation

The pilot dataset is manually annotated with regard to the crowd size as the ground truth, and values of identified image characteristics in section of Social media image characteristics.

Characteristics selection

As the image characteristics are identified from two different perspectives, i.e. the crowd management and social media. There might be overlaps, i.e. strong associated characteristics, exist between these two perspectives. To analyse the impact of image characteristics on the crowd size estimation accuracy, it is necessary to identify the associated characteristics and select the representative one.

Table 4 Descriptive statistics of the annotated social media image dataset in terms of crowd level and categories of image characteristics, part 1

Dataset		Crowd level						Global image characteristic									
Name	# images ~	% of each dataset						People present, % of each dataset			Condition, % of people present			View, % of people present			
		0	1	2	3	4		Yes	No		Indoor	Outdoor		Top	Parallel	Between	
Pilot	956	30.45	51.93	13.44	2.65	1.53	100.00	69.55	30.45	100.00	16.69	83.31	100.00	1.02	94.58	4.39	100.00
Full	1072	32.70	48.32	11.81	4.72	2.45	100.00	67.30	32.70	100.00	21.59	64.91	100.00	3.37	84.08	12.55	100.00
Total	2028	31.64	**50.02**	12.58	3.74	2.02	100.00	68.36	31.64	100.00	19.24	**73.74**	100.00	2.25	**89.12**	8.64	100.00

The cell in **bold** denotes it takes the largest proportion in the Total dataset in terms of the crowd level or categories of image characteristics

The **0, 1, 2, 3, 4** denote the crowd level

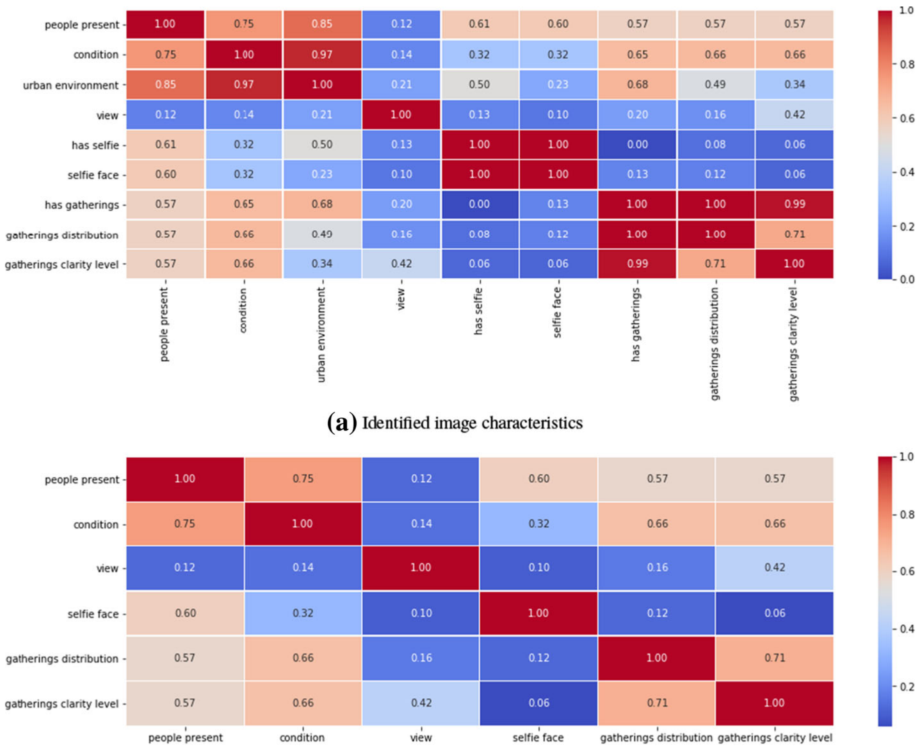
The cell valued **73.74%** in column **Outdoor** in **“Condition, % of images people present”** denotes the 73.74% of people present images, i.e. images containing people, were taken outdoors

Table 5 Descriptive statistics of the annotated social media image dataset in terms of crowd level and categories of image characteristics, part 2

Dataset		Frontend image characteristic							Backend image characteristics									
Name	% of total dataset	Selfie, % of people P.	Selfie face, % of selfie images						Gathering., % of people P.	Gatherings Distribution, % of images contain Gathering			Gatherings clarity level, % of images contain Gatherings					
			Full	Part	Back	Mixed	None	Fixed		Fixed	Gradient		A	B	C	D	E	
Pilot	47.14	66.03	75.83	9.53	4.66	9.09	0.89	100.00	62.37	21.36	78.64	100.00	23.24	39.67	26.76	9.86	0.47	100.00
Full	52.86	60.19	81.39	9.64	2.47	2.91	3.59	100.00	59.51	25.62	74.38	100.00	10.43	32.65	28.12	23.58	5.22	100.00
Total	100.00	62.99	**78.60**	9.59	3.57	6.02	2.23	100.00	60.88	23.53	**76.47**	100.00	16.72	**36.10**	27.45	16.84	2.88	100.00

The **bold** cell denotes it takes the largest proportion in the Total dataset in terms of categories of image characteristics.

The **A, B, C, D, E** denote the gatherings clarity level



(a) Identified image characteristics

(b) Selected image characteristics

Fig. 5 The association heatmap of identified image characteristics, calculated by Cramer's V based on annotated pilot dataset

We employ Cramer's V (Cramér 1999) to calculate the effect size of an association between each pair of characteristics. Cramer's V varies between 0 and 1. A value close to 0 shows little association between two characteristics. However, a value close to 1 indicates a strong association. Heatmaps in Fig. 5a shows the strength of association between each pair of characteristics. Redder cells denote strong associations while bluer ones denote weaker associations.

According to Fig. 5a, strong associations are observed between condition and 'urban environment', 'has selfie' and 'selfie face', 'has gatherings' and 'gatherings distribution'. For each above pair, one image characteristic will be removed, if:

- The association with other characteristics is higher than another characteristic.
- It is not required by crowd management, listed in Table 1.

For the pair 'condition' and 'urban environment', the 'condition' characteristic indicating an indoor or outdoor environment is key information for crowd management (Martella et al. 2017). It is more important than a specific location as is indicated by the 'urban environment' characteristic. Thus, the 'condition' characteristic will be kept. For the other two pairs, i.e. 'has selfie' with 'selfie face', and 'has gatherings' with 'gatherings distribution', the latter characteristics in each pair, i.e. 'selfie face' and 'gatherings distribution', contain richer information than the former ones, and has less association in average with other image characteristics. Consequently, the latter ones are kept.

Thus, the following image characteristics will be removed: 'urban environment', 'has selfie', and 'has gatherings'. The association heatmap of the selected image characteristics is shown in Fig. 5b.

Full dataset annotation

After selecting the image characteristics, we annotated the selected image characteristics and the crowd size (ground truth) for the rest of the dataset (Full dataset). For this annotation we have used crowd-sourcing (Schenk and Guittard 2009): the selected image characteristics and crowd size of each image are determined by multiple people and the majority judgement is taken as the ground truth. We performed the crowd-sourcing operation using figure Eight (https://www.figure-eight.com/), a popular crowd-sourcing platform.

After the annotation of the full dataset, it is merged with the annotated pilot dataset to come up with the annotated total dataset.

Dataset descriptive statistics

The descriptive statistics of the dataset in terms of selected image characteristics are listed in Tables 4 and 5, with highlights for the largest proportions in the Total dataset in terms of crowd level or categories of image characteristics. The total dataset contains 2,028 Instagram images, of which the pilot dataset contained around 47.14% and rest are part of the full dataset. The crowd level distributions among all social media images show similar pattern across datasets, i.e. almost one third of them contain no people, and half of them contain number of people less than 20. While, Around 12% of them contain number of people between 20 to 100. Images containing more than 100 people are rare. Though the dataset is imbalanced in terms of crowd size levels, it reflects the reality of the crowd size levels in social media images, which is imbalanced. The effectiveness of algorithms tested using this dataset collected from social media, therefore, fits the purpose of this research, i.e. to compare the effectiveness of methods on crowd size estimation in city events using social media images.

With regard to the image characteristics in both datasets, among more than two third of images in which people are present, 74% of them in average are taken in the outdoor, and around 89% of them are taken in the parallel view. With regard to the frontend image characteristics, about 63% of images which contain people are selfie photos, and 79% of such selfies captured the full face of people. With regard to the backend image characteristics, among all images containing gatherings, around 76% of them the gatherings are shown in gradient distribution, i.e. the gatherings are gradually far away from the camera. In around 36% and 27% of such cases, clarity levels are B and C, implying that we can see the face without the detailed features of the faces of those people in the gatherings, or we can only see their shapes, respectively.

Experimental setup

We set up experiments to perform crowd counting analysis using social media images in city events to answer two research questions of this study. To answer the first research question we set up an experiment to study the crowd size estimation accuracy of different methods. To answer the second research question we set up an experiment to investigate the impact of selected image characteristics on the correctness of crowd size estimation by each method. In the following subsections, we introduce each experiment in terms of their variables, expected results and process.

Experiment 1: crowd size estimation accuracy

We set up the first experiment to assess the estimation accuracy for each selected method to estimate the crowd size. The independent variable in this experiment is the accuracy of the selected methods and the annotated social media images in the dataset. The dependent variable is the crowd size of each image estimated by the each method. In addition to the estimated crowd size, this experiment also outputs the measures A_i defined in Eq. 1 for comparing the effectiveness of different methods. The experiment process is listed as follows:

- For each method, we perform crowd size estimation on social media images in the dataset, yielding a set of crowd size estimated for each image.
- Calculate the measures A_i defined in Eq. 1 on the set of estimated crowd size for each method.

Experiment 2: impact of image characteristics on crowd size estimation

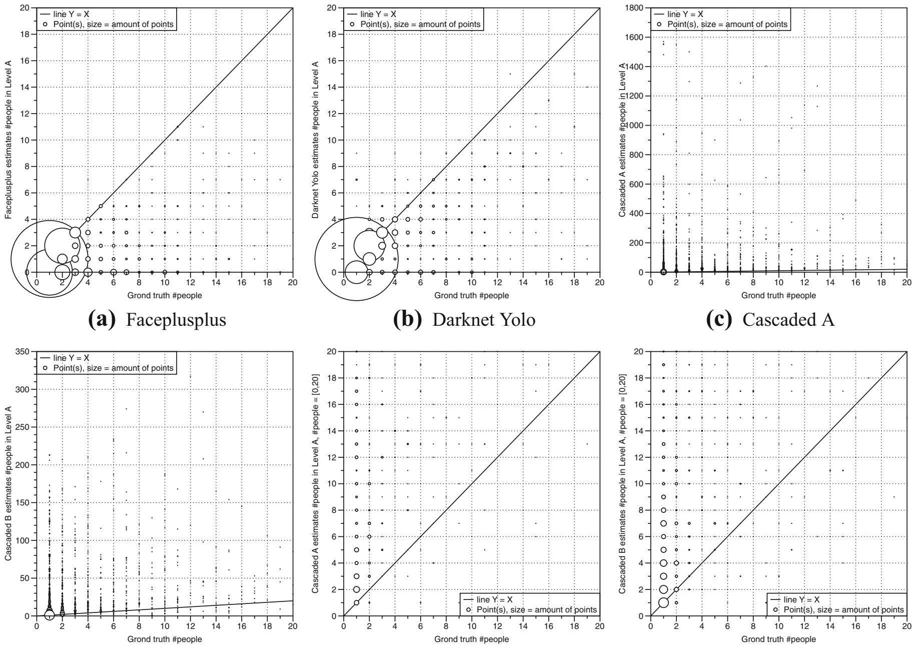
The second experiment is set up to investigate the impact of image characteristics on the accuracy of the estimation of crowd size level estimation by the different methods. The independent variables consist of the selected image characteristics and the correctness of crowd size level estimated by each method. The dependent variable is the impact (coefficient) of each image characteristic on the crowd size level correctness for each method. After performing the experiment process, in addition to the image characteristics impact, the experiment also outputs a classification model with cross-validation performance as a side product for each method. The classification model classify input images into bi-categories, i.e. whether the crowd size in an image can be correctly estimated, while taking into account the image characteristics. For instance, the crowd size in a selfie image containing people with full faces, captured in parallel view without mass gatherings in the backend may be correctly estimated by Faceplusplus or Darknet Yolo methods rather than Cascaded methods. The process of this experiment is listed as follows:

- Calculate whether the estimated crowd size level is the same with the ground truth for each image, and set the result as the dependent bi-categorical variable.
- Train a binary classifier using logistic regression algorithm for machine learning (Dreiseitl and Ohno-Machado 2002) for each method with image characteristics and the estimation correctness by this method. To assess how the classifier will generalize to an independent dataset (Kohavi 1995), we apply fivefold cross-validation (Guyon 1997) in the training process.

Table 6 Crowd size estimation by different methods

	Crowd size level estimation				# People in crowd size Level 1 estimation			
	# Underestimate	# Accurate est.	# Overestimate	Est. accuracy (%)	# Underestimate	# Accurate est.	# Overestimate	Est. accuracy (%)
Faceplusplus	679	1318	31	65.00	635	381	8	37.21
Darknet Yolo	531	1460	37	**72.01**	515	390	119	**38.09**
Cascaded A	30	501	1496	24.72	371	21	632	2.05
Cascaded B	65	719	1244	35.43	305	50	669	4.88

The **bold** cell denotes model with the highest estimation accuracy



(a) Faceplusplus **(b)** Darknet Yolo **(c)** Cascaded A

(d) Cascaded B **(e)** Cascaded A, #people ¡= 20 **(f)** Cascaded B, #people ¡= 20

Fig. 6 The distribution of specific number of people estimated in crowd size Level 1 for each method

- Record the impact (coefficient) of each selected image characteristic from the trained model and measure the classification performance from cross-validation by *Precision*, *Recall* and *F1_Score* (Powers 2011) introduced in the previous chapter.

Crowd counting analysis: findings of the experiment on crowd size estimation and image characteristics impacts

In this section, we analyse the accuracy of the estimated crowd size for each method, as well as the impact of image characteristics on the crowd size level estimation for each method listed in Table 2.

Crowd size estimation from social media images in city events

Table 6 lists the results of the crowd size estimation of different methods using social media data in city events. Here, we distinguish different levels of estimation. When the estimated level is 1 (so less than 20 persons), we also specify the exact estimated number of people.

Crowd size level estimation

According to Table 6, Faceplusplus (65.00%) and Darknet Yolo (72.01%) reach 2–3 times higher accuracy than Cascaded A (24.72%) and Cascaded B (35.43%). Faceplusplus and Darknet Yolo underestimate the crowd size in a large number of images, while Cascaded A and B predict too high values. As Faceplusplus and Darknet Yolo count people by identifying their faces or shapes, the crowd size in dense images is underestimated, as the faces and shapes might not be available in this type of images.

To compare the estimated levels with the ground truth, we show the distribution of estimated levels for each method in Table 7. The diagonal of the table shows the percentage of images that are correctly estimated by each method. According to the table, Faceplusplus and Darknet Yolo produce higher percentage of correct estimation in less dense levels 0 and 1. Instead, the Cascaded A and B produce more correct estimations for higher levels. This may also be caused by the distinct features used by different methods in detecting people, i.e. Faceplusplus and Darknet Yolo detect people by faecs or shapes while Cascaded methods use learned features. We can thus conclude that Faceplusplus and Darknet Yolo are more feasible in low-density environments, while Cascaded A and B are fit for high-density environments. While comparing the Faceplusplus and Darknet Yolo, the latter method reaches better accuracy than the former one; which may indicate that in social media images, even in low dense environment, shapes are more available or valuable than faces to be detected for counting people. In the meantime, as the constructed dataset contains more low dense images collected from social media, the estimation accuracy for Cascaded methods is obviously lower than Faceplusplus and Darknet Yolo.

Specific number of people in crowd size level 1 estimation

According to Table 6, Darknet Yolo reaches the highest estimation accuracy (38.09%) in the estimation of the specific number of people in crowd size level 1, closely followed by Faceplusplus (37.21%). The Cascaded A and B methods reach very low accuracy (2.05%, 4.88%). Similar to the observations in crowd size level estimation, the tendency of under- and overestimation of different methods may caused by the different features they used for detecting people, as we described in the previous section.

To explore the relationship among the ground truth, the estimation value and amount of such estimation for each method, we plotted in Fig. 6 with a ground truth value on the X axis, an estimated value on the Y axis and the number of corresponding estimation points in size. Points on the diagonal ($Y = X$) denote correct estimation. According to Fig. 6a, b, the Faceplusplus and Darknet Yolo methods reach the highest accuracy in the range of 0 to 4. Instead, the accurate estimation for the Cascaded methods, according to Fig. 6c–f, are distributed more equally than Faceplusplus and Darknet Yolo. This is consistent with the mechanism of different methods, i.e. Faceplusplus and Darknet Yolo are more feasible in low dense environment while Cascaded A and B are more feasible in high dense environment.

When comparing the two Cascaded methods, the accurate estimation in Cascaded B are more equally distributed than the Cascaded A. This is nature that most of social media image sent during city events are captured in outdoor event area, which are more feasible for the method Cascaded B, which are trained with busy street area in city events, than Cascaded A, which are trained with random pictures of city events.

Table 7 Crowd size level estimation by different methods

Method		Ground Truth					
	Levels	0	1	2	3	4	Total
Face ++	0	95.00%	30.90%	45.60%	55.10%	59.50%	54.00%
	1	5.00%	69.00%	53.60%	43.60%	40.50%	45.80%
	2	0.00%	0.10%	0.80%	1.30%	0.00%	0.20%
	3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Darknet Yolo	0	94.40%	16.40%	24.50%	19.20%	16.70%	41.50%
	1	5.60%	83.50%	74.30%	78.20%	81.00%	58.20%
	2	0.00%	0.10%	1.10%	2.60%	2.40%	0.30%
	3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Cascaded A	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	1	31.70%	32.40%	4.20%	5.10%	0.00%	26.80%
	2	35.90%	39.80%	40.20%	7.70%	9.50%	36.80%
	3	18.20%	15.00%	29.90%	47.40%	14.30%	19.10%
	4	14.30%	12.70%	25.70%	39.70%	76.20%	17.20%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Cascaded B	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	1	51.00%	50.90%	6.90%	3.80%	4.80%	42.50%
	2	35.90%	37.70%	52.50%	25.60%	16.70%	38.10%
	3	11.60%	11.10%	36.40%	65.40%	40.50%	17.20%
	4	1.60%	0.30%	4.20%	5.10%	38.10%	2.20%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Face ++ denotes the Faceplusplus method.
The **Percentage** value in each cell denotes the percentage of correctly estimated crowd size levels.
Darker cells denote higher accuracy than lighter ones.

Impact of image characteristics on crowd level estimation

We perform the second experiment to investigate the impact of image characteristics on crowd size level estimation for separate methods. The result is listed in Tables 8 and 9, where each dark cell indicates an image characteristic (e.g. ‘Condition’) with current value (e.g. ‘Outdoor’) in the column has maximum positive impact for the corresponding method in the row. Namely, images with characteristics in such value have larger possibility to the correctly estimated the crowd size level by the corresponding method. For instance, the cell

Table 8 The impact of image characteristics on crowd level estimation for each method, part 1

Method	Global image characteristics									Frontend image characteristics					
	People present		Condition			View				Selfie face					
	Yes	No	Outdoor	Indoor	Unknown	Parallel	Between	Top	Unknown	No selfie	Full face	Part face	Back face	Mixed face	Only body
Face ++	− 0.719	**0.719**	0.526	**0.689**	− 1.215	**0.880**	0.541	− 0.206	− 1.215	− 0.279	**1.197**	0.114	− 0.873	0.630	− 0.787
Darknet Yolo	− 0.366	**0.366**	0.457	**0.758**	− 1.215	**0.874**	0.358	− 0.017	− 1.215	− 0.442	**0.693**	0.329	0.100	0.157	− 0.837
Cascaded A	**1.354**	− 1.354	− 0.139	**0.398**	− 0.259	− 0.098	0.167	**0.190**	− 0.259	− 0.096	0.034	0.083	**0.259**	− 0.262	− 0.018
Cascaded B	**1.429**	−1.429	− 0.217	**0.504**	− 0.287	− 0.028	0.081	**0.234**	− 0.287	− 0.306	0.005	0.060	− 0.142	− 0.025	**0.409**

The **Face ++** denotes the Faceplusplus method.

Each **bold** cell indicates an image characteristic (e.g. ‘Condition’) with current value (e.g. ‘Outdoor’) in the column has maximum positive impact for the corresponding method in the row. Namely, images with characteristics in such value have larger possibility to the correctly estimated the crowd size level by the corresponding method

Table 9 The impact of image characteristics on crowd level estimation for each method, part 2

Method	Backend image characteristics									Classification model performance E {score}
	Gathering distribution			Gathering clarity level						
	No gatherings	Fixed	Gradient	No gatherings	A	B	C	D	E	
Face ++	0.137	**0.239**	− 0.375	0.288	**0.376**	0.080	− 0.210	− 0.111	− 0.424	0.866
Darknet Yolo	0.020	**0.329**	− 0.348	**0.612**	0.156	0.136	− 0.235	− 0.203	− 0.466	**0.869**
Cascaded A	− 0.323	0.005	**0.318**	**0.479**	0.034	0.068	− 0.214	− 0.241	− 0.125	0.765
Cascaded B	−0.276	**0.142**	0.134	**0.390**	0.223	− 0.066	− 0.131	− 0.239	− 0.176	0.740

The **Face ++** denotes the Faceplusplus method

The **E{F1-score}** in the column of "Classification model performance" (the last column) denotes the average F1-score calculated in multi-folds validation in training of classification model with impact scored 0.88 in column ‘Parallel’ indicates that images captured in ‘view’ (i.e. an image characteristic) of ‘parallel’ (i.e. the value of the image characteristic ‘View’) shows most positive impact on crowd size level estimation than any other value in view characteristic using the method Faceplusplus. Simply, it is more likely that Faceplusplus estimates the crowd size level correctly for images with taken in parallel view.

According to Tables 8 and 9, the image characteristic ‘people present’ shows a negative impact for Faceplusplus and Darknet Yolo, but maximum positive impact for the Cascaded methods. This may be caused by the Faceplusplus and Darknet Yolo methods which tend to underestimate crowd size, and thus increased the correct estimation accuracy in processing images containing less people, in particular, no people.

We also found that indoor pictures show positive impact to all methods. It may be caused by indoor images containing less people which reduces the difficulties in crowd size estimation.

We observed that pictures taken in parallel view show higher positive impact for Faceplusplus and Darknet Yolo, while top view pictures are better interpreted by Cascaded methods. This may be caused that Faceplusplus and Darknet Yolo, counting people by faces and shapes, require more detail information about people than Cascaded methods, counting people through learned features as introduced in the previous chapter.

With regard to ‘gathering distribution’, all methods except Cascaded A shows higher estimation accuracy with fixed distribution of gatherings. This is natural that, compared with gradient distribution, the gatherings in fixed distribution contains less people and the people have a similar size in the image, which reduces the difficulties in detecting and counting people.

The findings show that all methods except Faceplusplus tend to correctly estimate the crowd size of images with gatherings. The Faceplusplus instead reaches a higher estimation accuracy with images containing clearest gatherings (so in level A) than no gatherings. It is natural that for other three methods, the gatherings which are in small size in the backend of the images increase the difficulties for crowd size estimation. However, as a face recognition based method, the Faceplusplus still can recognise small but clear faces (so in Level A) in the gatherings.

To assess the effectiveness of the impact of image characteristics on the crowd size estimation accuracy for each method, we tested the cross-validation performance of the by-product classifier constructed with impact of image characteristics. According to Table 9, both Faceplusplus and Darknet Yolo reach *F1_Score* at 0.86, while Cascaded A and B reach 0.76 and 0.74, respectively. It indicates that Faceplusplus and Darknet Yolo reach higher possibility (confidence) to produce correct crowd size level estimation than Cascaded methods when characteristics of images are in most positive impact values.

Discussion

In this section, we discuss the research and findings in terms of the dataset, the effectiveness and feasibility of algorithms, and how the crowd managers use estimated crowd size information to manage the crowd in city events.

The constructed dataset shows that social media data is indeed broadly available in city event area in large size, rather than data from sensors (e.g. Wi-Fi, counting system) which require extra resources in data collection. According to the constructed social media dataset in city events, 70% of images contain people. Thus, these images are valuable for crowd size estimation. Particularly, 20% of total images contain more than 20 people in each picture, which are essential in crowd size estimation for crowd management to reduce risks and incidents. However, the unbalance of number of images in terms of the way people are showing in the images (so as the different perspective the people are captured from camera, the selfie or panoramic pictures), and dense levels may affect the accuracy of methods in crowd size estimation, because methods identify and count people depending on such information.

With regard to social media data and the selected methods listed in Table 2, the findings show that direct methods (Faceplusplus and Darknet Yolo) tend to underestimate the crowd size while indirect methods (Cascaded A and B) tend to overestimate the crowd size. Also, direct methods are more effective with social media images in parallel view with full face and fixed clarity gatherings. Whereas, indirect methods are more effective when images are taken in top view with gatherings in gradient distribution. This may be explained by the mechanism of different types of methods, i.e. the direct methods detect and count people in an image by information of face or shape of people, are more feasible with images captured in parallel view (i.e. a viewpoint at more or less the same height as that of the people in the photo) with full faces of people and clearly observable gatherings around the people. In contrast, the indirect methods, such as Cascaded methods, which detect and count people through non-handcrafted features do not require the same information (e.g. face or shape of people) for crowd size estimation.

Moreover, the feasibility of different types of methods with social media images in different dense, i.e. direct methods fit low dense images while indirect methods fit high dense images, may indicate the crowd size estimation accuracy may be improved by selecting different methods under conditions which methods work best. For instance, to improve the crowd size estimation accuracy, we may classify the low dense and high dense images, or split frontend layer containing selfie people with backend layer containing gatherings in one image, and then assign low dense frontend images (layers) to direct methods and assign high dense backend images (layers) to indirect methods. To reach this, it is required to detect the conditions automatically and assign images with different characteristics to the different methods.

When applying these techniques in crowd management, crowd managers use the estimated crowd size following a two-phase process (Martella et al. 2017), i.e. planning phase and operational phase. In the planning phase, the historical event data, e.g. event programs, stewards report and social media data from previous events (such as Sail 2015, King’s Day 2016) or the same events in previous editions (such as King’s Day in 2018, 2017 and 2016) can be used for simulating the change of the crowd size in the event area. This knowledge can be used for inferring guidelines and making the crowd management plan for the coming events. The crowd management plan is made by answering a set of questions (Li 2019; Still 2000), such as when the crowd size reaches the peak, how the crowd size peak changes temporally and geographically. Different crowd size estimation methods can be used in different dense environments. For instance, using direct methods in the beginning or end of the event which are in low dense crowds, while using the indirect methods in the high dense environment in the peak of the events according to the event programs and stewards reports. In the operational phase, a set of what-if scenarios are prepared and crowd management measures for a certain scenario are predefined. During the event, the real-time social media images collected from this event can be used to estimate the current crowd size in the monitoring area. The current crowd size can be used for calculating properties for a crowded urban area, such as the level-of-service (Fruin 1971; Marana et al. 1998), crowd density (Gong et al. 2018b; Blanke et al. 2014; Botta et al. 2015; Quercia et al. 2015), which can be further used for assessing potential risks (Helbing et al. 2005).

Once the threshold of a property in a certain scenario is met, e.g. the level-of-service reaches lower levels (e.g. D, E, F) that refers to higher density of people (Fruin 1971) which indicates potential risks, the predefined crowd management measures can be applied to manage the crowd and avoid such risks taking place (Li 2019; Still 2000).

Summary and conclusion

Knowing the crowd size is essential for crowd safety when managing the crowd. Conventional solutions to derive such information depend on manual observations, which are expensive, prone to observation biases, and not suitable for global observations.

In this paper, we investigate the accuracy of four methods to estimate crowd size using social media images in city events, and we also investigate the impact of image characteristics on the estimation effectiveness for different methods.

To perform this research, we select four methods for crowd size estimation analysis from two different types. We created a dataset consisting of social media images collected from various events and major activities. Each image is annotated with a set of image characteristics and crowd size as ground truth. This dataset has been used for investigating the crowd size estimation accuracy of selected methods and the impact of image characteristics on the crowd size estimation.

Findings show that direct methods (Faceplusplus and Darknet Yolo) reach better estimation accuracy than indirect methods (Cascaded method A and B). Specifically, Darknet Yolo reaches the highest accuracy in crowd size level estimation (72.01%) and in estimating the specific number of people when less than 20 (38.09%). The findings indicate that, social media images taken in parallel view with selfie people in full face and gatherings in fixed distributed are more feasible to Faceplusplus and Darknet Yolo for crowd size estimation, while images taken from top view with gatherings in gradient distribution are more suitable to Cascaded methods. We recommend to use Darknet Yolo method to estimate and predict the crowd size in city events based on social media images, in terms of levels of crowd size as well as specific number of people if it is a low dense environment.

Results of this research may be influenced by the construction bias of the social media dataset, which are introduced by the diverse characteristics of city events. Though we considered a set of event characteristics in dataset construction, it may not be able to cover all diversities in the reality. Thus, the crowd size estimation performance using social media images may be affected in city events beyond the scope of consideration in terms of diversity of event characteristics. Moreover, the bias in social media usage in terms of users' age and gender may also affect crowd size estimation using social media. For instance, knowing that social media is more popular in younger generation (Yang et al. 2016; Gong et al. 2018a), city events with less younger participants may generate less social media images, which may not sufficient for training and improving the crowd size estimation methods, and also not sufficient for methods to estimate the crowd size during events. Therefore, it may affect crowd size estimation for crowd management.

In future work, we plan to propose and explore more methods for crowd size estimation using social media images, such as hybrid methods that integrate the advances of direct and indirect methods, in such a way to improve the estimation effectiveness for crowd management. Also, we intend to enlarge the social media dataset by adding more diverse events and activities, and by adding more annotated image characteristics. Last but not least, we will investigate the feasible approach to derive image characteristics automatically from social media dataset. In this way, the annotated dataset can be enlarged. It can also be used for proposing new methods for crowd size estimation, such as training a classifier to estimate the crowd size using social media images.

Acknowledgements The research leading to these results has received funding from the European Research Council under the European Union Horizon 2020 Framework Programme for Research and Innovation. It is established by the Scientific Council of the ERC Grant Agreement No. 669792 (Allegro).

Author contributions The authors confirm contribution to the paper as follows: study conception and design: all authors; data collection: VXG; analysis and interpretation of results: VXG, WD; draft manuscript preparation: VXG, WD, AB; study supervision: SPH. All authors reviewed the results and approved the final version of the manuscript.

References

Blanke, U., Tröster, G., Franke, T., Lukowicz, P.: Capturing crowd dynamics at large scale events using participatory gps-localization. In: IEEE 9th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 1–7. IEEE (2014)

Bocconi, S., Bozzon, A., Psyllidis, A., Titos Bolivar, C., Houben, G.J.: Social glass: a platform for urban analytics and decision-making through heterogeneous social data. In: Proceedings of the 24th International Conference on World Wide Web, pp. 175–178. ACM (2015)

Botta, F., Moat, H.S., Preis, T.: Quantifying crowd size with mobile phone and twitter data. R. Soc. Open Sci. 2(5), 150162 (2015)

Chan, A.B., Liang, Z.S.J., Vasconcelos, N.: Privacy preserving crowd monitoring: counting people without people models or tracking. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–7. IEEE (2008)

Chen, K., Loy, C.C., Gong, S., Xiang, T.: Feature mining for localised crowd counting. In: BMVC, vol. 1, p. 3 (2012)

Chen, K., Gong, S., Xiang, T., Change, Loy C.: Cumulative attribute space for age and crowd density estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2467–2474 (2013)

Cramér, H.: Mathematical Methods of Statistics, vol. 9. Princeton University Press, Princeton (1999)

Daamen, W., Yuan, Y., Duives, D., Hoogendoorn, S.: Comparing three types of real-time data collection techniques: counting cameras, wi-fi sensors and gps trackers. In: Proceedings of the Pedestrian and Evacuation Dynamics (2016)

Davies, A.C., Yin, J.H., Velastin, S.A.: Crowd monitoring using image processing. Electron. Commun. Eng. J. 7(1), 37–47 (1995)

Dreiseitl, S., Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review. J. Biomed. Inform. 35(5–6), 352–359 (2002)

Duives, D.C., Daamen, W., Hoogendoorn, S.P.: Quantification of the level of crowdedness for pedestrian movements. Physica A Stat. Mech. Appl. 427, 162–180 (2015)

Earl, C., Parker, E., Tatrai, A., Capra, M., et al.: Influences on crowd behaviour at outdoor music festivals. Environ. Health 4(2), 55 (2004)

Fang, Z., Yuan, J., Wang, Y.C., Lo, S.M.: Survey of pedestrian movement and development of a crowd dynamics model. Fire Saf. J. 43(6), 459–465 (2008)

Fruin, J.J.: Designing for pedestrians: A level-of-service concept. HS-011 999 (1971)

Gong, V.X., Daamen, W., Bozzon, A., Hoogendoorn, S.: Crowd characterization using social media data in city-scale events for crowd management. Technical report (2018a)

Gong, V.X., Yang, J., Daamen, W., Bozzon, A., Hoogendoorn, S., Houben, G.J.: Using social media for attendees density estimation in city-scale events. IEEE Access 6, 36325–36340 (2018b)

Gong, V.X., Daamen, W., Bozzon, A., Hoogendoorn, S.P.: Estimate sentiment of crowds from social media during city events. Transp. Res. Rec. p 0361198119846461 (2019)

Guyon, I.: A scaling law for the validation-set training-set size ratio. AT&T Bell Laboratories, pp. 1–11 (1997)

Helbing, D., Buzna, L., Johansson, A., Werner, T.: Self-organized pedestrian crowd dynamics: experiments, simulations, and design solutions. Transp. Sci. 39(1), 1–24 (2005)

Idrees, H., Saleemi, I., Seibert, C., Shah, M.: Multi-source multi-scale counting in extremely dense crowd images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2547–2554 (2013)

Jiang, M., Huang, J., Wang, X., Tang, J., Wu, C.: An approach for crowd density and crowd size estimation. JSW 9(3), 757–762 (2014)

Jiang, X., Xiao, Z., Zhang, B., Zhen, X., Cao, X., Doermann, D., Shao, L.: Crowd counting and density estimation by trellis encoder–decoder networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6133–6142 (2019)

Kohavi, R., et al.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai, Montreal, Canada, vol. 14, pp. 1137–1145 (1995)

Lempitsky, V., Zisserman, A.: Learning to count objects in images. In: Advances in Neural Information Processing Systems, pp. 1324–1332 (2010)

Li, J.: Crowds inside out: understanding crowds from the perspective of individual crowd members' experiences. Ph.D. thesis, Delft University of Technology (2019)

Li, Y., Zhang, X., Chen, D.: Csrnet: dilated convolutional neural networks for understanding the highly congested scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1091–1100 (2018)

Liu, W., Salzmann, M., Fua, P.: Context-aware crowd counting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5099–5108 (2019)

Luo, D., Bonnetain, L., Cats, O., van Lint, H.: Constructing spatiotemporal load profiles of transit vehicles with multiple data sources. Transp. Res. Rec. p 0361198118781166 (2018)

Ma, Z., Wei, X., Hong, X., Gong, Y.: Bayesian loss for crowd count estimation with point supervision. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 6142–6151 (2019)

Marana, A.N., Velastin, S.A., Costa, L.F., Lotufo, R.: Automatic estimation of crowd density using texture. Saf. Sci. 28(3), 165–175 (1998)

Marsden, M., McGuinness, K., Little, S., O'Connor, N.E.: Resnetcrowd: a residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–7. IEEE (2017)

Martella, C., Li, J., Conrado, C., Vermeeren, A.: On current crowd management practices and the need for increased situation awareness, prediction, and intervention. Saf. Sci. 91, 381–393 (2017)

Nanni, L., Ghidoni, S., Brahnam, S.: Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognit. **71**, 158–172 (2017)

Polus, A., Schofer, J.L., Ushpiz, A.: Pedestrian flow and level of service. J. Transp. Eng. 109(1), 46–56 (1983)

Powers, D.M.: Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation (2011)

Psyllidis, A., Bozzon, A., Bocconi, S., Bolivar, C.T.: A platform for urban analytics and semantic data integration in city planning. In: International Conference on Computer-Aided Architectural Design Futures. Springer, pp. 21–36 (2015)

Quercia, D., Schifanella, R., Aiello, L.M., McLean, K.: Smelly maps: the digital life of urban smellscapes. In: AAAI (2015)

Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)

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

Ryan, D., Denman, S., Sridharan, S., Fookes, C.: An evaluation of crowd counting methods, features and regression models. Comput. Vis. Image Underst. 130, 1–17 (2015)

Saleh, S.A.M., Suandi, S.A., Ibrahim, H.: Recent survey on crowd density estimation and counting for visual surveillance. Eng. Appl. Artif. Intell. 41, 103–114 (2015)

Sam, D.B., Surya, S., Babu, R.V.: Switching convolutional neural network for crowd counting. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4031–4039. IEEE (2017)

Schenk, E., Guittard, C., et al..: Crowdsourcing: what can be outsourced to the crowd, and why. In: Workshop on Open Source Innovation, Strasbourg, France, Citeseer, vol. 72, p. 3 (2009)

Shi, M., Yang, Z., Xu, C., Chen, Q.: Revisiting perspective information for efficient crowd counting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7279–7288 (2019)

Sindagi, V.A., Patel, V.M.: Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE (2017)

Sindagi, V.A., Patel, V.M.: A survey of recent advances in CNN-based single image crowd counting and density estimation. Pattern Recogn. Lett. 107, 3–16 (2018)

Sindagi, V.A., Patel, V.M.: Ha-ccn: hierarchical attention-based crowd counting network. IEEE Trans. Image Process. 29, 323–335 (2019a)

Sindagi, V.A., Patel, V.M.: Multi-level bottom-top and top-bottom feature fusion for crowd counting. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1002–1012 (2019b)

Sindagi, V.A., Yasarla, R., Patel, V.M.: Pushing the frontiers of unconstrained crowd counting: new dataset and benchmark method. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1221–1231 (2019)

Still, G.K.: Crowd dynamics. Ph.D. thesis, University of Warwick (2000)

Walach, E, Wolf, L.: Learning to count with CNN boosting. In: European Conference on Computer Vision, pp. 660–676. Springer (2016)

Wang, Q., Gao, J., Lin, W., Li, X.: Nwpu-crowd: a large-scale benchmark for crowd counting. arXiv preprint arXiv:200103360 (2020)

Wang, Y., de Almeida Correia, G.H., van Arem, B., Timmermans, H.H.: Understanding travellers' preferences for different types of trip destination based on mobile internet usage data. Transp. Res. C Emerg. Technol. 90, 247–259 (2018)

Xiong, F., Shi, X., Yeung, D.Y.: Spatiotemporal modeling for crowd counting in videos. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5151–5159 (2017)

Yang, J., Hauff, C., Houben, G.J., Bolivar, C.T.: Diversity in urban social media analytics. In: International Conference on Web Engineering, pp. 335–353. Springer (2016)

Yuan, Y.: Crowd monitoring using mobile phones. In: 2014 Sixth International Conference on Intelligent Human–Machine Systems and Cybernetics, vol. 1, pp. 261–264. IEEE (2014)

Zhang, C., Li, H., Wang, X., Yang, X.: Cross-scene crowd counting via deep convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 833–841 (2015)

Zhang, C., Kang, K., Li, H., Wang, X., Xie, R., Yang, X.: Data-driven crowd understanding: A baseline for a large-scale crowd dataset. IEEE Trans. Multimed. 18(6), 1048–1061 (2016a)

Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y.: Single-image crowd counting via multi-column convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 589–597 (2016b)

Zhou, E., Cao, Z., Yin, Q.: Naive-deep face recognition: touching the limit of lfw benchmark or not? arXiv preprint arXiv:150104690 (2015)

V. X. Gong is currently a Postdoc researcher in Human-Centered Artificial Intelligence, Delft University of Technology, The Netherlands. He uses social media data, such as Twitter and Instagram, to identify relevant information, develop data models, estimate and analyse crowds' characteristics in city events, including demographic and city-role composition, Spatial-temporal distribution, sentiment estimation, Points of Interest preferences, word use, crowd size and density estimation, which help crowd managers make better decisions. More information please check his website: https://Gong.im.

W. Daamen is currently an Associate professor with the Chair of Traffic Operations and Management, Department of Transport and Planning, Delft University of Technology, The Netherlands. Her research interests include theory, modelling, and simulation of traffic (pedestrians, cyclists, vehicles, and vessels), and innovative methods have been developed to collect microscopic traffic data, which are used to underpin theories and models describing traffic operations.

A. Bozzon is currently a Professor and the Head of the department of Human-Centered Artificial Intelligence, and a part-time professor with the Web Information Systems Group, Delft University of Technology, The Netherlands. His research interests are at the intersection of crowd-sourcing, user modelling, and web information retrieval. He studies and creates novel social data science methods and tools that combine the cognitive and reasoning abilities of individuals and crowds, with the computational powers of machines, and the insights from large amounts of heterogeneous data.

S. P. Hoogendoorn is currently a Professor and the Head of the Chair of Traffic Operations and Management, Department of Transport and Planning, Delft University of Technology, The Netherlands. He is also a Principal Investigator with the AMS Amsterdam Institute for Advanced Metropolitan Solutions, The Netherlands, a Faculty Fellow with the IBM Benelux Center of Advanced Studies, The Netherlands, and a Strategic Advisor with ARANE, The Netherlands. In the past several years, his research has involved theory, modelling and simulation of traffic and transportation networks. He focused on innovative approaches to collect microscopic traffic data and the use of these data to underpin the models and theories that he has developed, using new techniques for model identification.

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