Urban Object Detection Kit: A System for Collection and Analysis of Street-Level Imagery

Maarten Sukel
University of Amsterdam
Amsterdam, The Netherlands
m.m.sukel@uva.nl

Stevan Rudinac
University of Amsterdam
Amsterdam, The Netherlands
s.rudinac@uva.nl

Marcel Worring
University of Amsterdam
Amsterdam, The Netherlands
m.worring@uva.nl

ABSTRACT

In this paper, we propose Urban Object Detection Kit, a system for the real-time collection and analysis of street-level imagery. The system is affordable and portable and allows local government agencies to receive actionable intelligence about the objects on the streets. This system can be attached to service vehicles, such as garbage trucks, parking scanners and maintenance cars, thus allowing for large-scale deployment. This will, in turn, result in street-level imagery captured at a high collection frequency, while covering a large geographical region. Unlike more traditional panoramic street-level imagery, the data collected by this system has a higher frequency, making it suitable for the highly dynamic nature of city streets. For example, the proposed system allows for real-time detection of urban objects and potential issues that require the attention of city services. It paves the way for easy deployment and testing of multimedia information retrieval algorithms in a dynamic real-world setting. We showcase the usefulness of object detection for identifying issues in public spaces that occur within a limited time span. Finally, we make the kit, as well as the data collected using it, openly available for the research community.

CCS CONCEPTS

• **Computing methodologies** → Visual inspection; *Mobile agents*; **Distributed computing methodologies**; • **Information systems** → **Multimedia and multimodal retrieval**.

KEYWORDS

urban object detection, urban computing, urban multimedia data collection, real-time street-level imagery

ACM Reference Format:
Maarten Sukel, Stevan Rudinac, and Marcel Worring. 2020. Urban Object Detection Kit: A System for Collection and Analysis of Street-Level Imagery. In *Proceedings of the 2020 International Conference on Multimedia Retrieval (ICMR '20), June 8–11, 2020, Dublin, Ireland*. Dublin, Ireland, 8 pages. https://doi.org/10.1145/3372278.3390708

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICMR '20, June 8–11, 2020, Dublin, Ireland
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7087-5/20/06…$15.00
https://doi.org/10.1145/3372278.3390708



Figure 1: Example of how several types of objects are detected on the city streets using our proposed Urban Object Detection Kit.

1 INTRODUCTION

Cities around the world have the task of making sure that public spaces are safe, clean and well maintained. Since cities are highly dynamic environments, finding issues that require the attention of relevant services is an extremely challenging task. In practice, citizens often have to report an issue first before the action is taken. Automatic detection of urban issues would allow for a proactive response. In this paper, we propose a system that is capable of real-time collecting and analyzing street-level imagery.

In recent years, a significant body of research has addressed the problem of capturing street-level imagery, which has resulted in successful applications, such as Google Street View, Mapillary and OpenStreetCam. However, panoramic images are typically captured infrequently, which makes them impractical for real-time monitoring of streets and public spaces in general. The real-time monitoring of city streets is essential because the issues in cities can occur within a short time span. In this paper we particularly focus on urban objects that are related to these issues and pose significant challenges to the local government agencies.

Examples of urban objects we aim at detecting are graffiti on the walls, trash on the street (cf. Figure 1), broken traffic poles or bicycle wrecks that need to be removed. However, the system could also be used for estimating parking spaces availability, measuring pedestrian density on sidewalks or monitoring the growth of vegetation. For the detection of these issues, we collect and analyse street-level imagery in real-time which is combined with the geographical data.

A more traditional approach to finding urban issues is using citizen reports, which are routinely collected by the cities around

the world using standardized protocols such as Open311 [21]. A disadvantage of citizen reports is that they are highly participatory in nature, requiring active input of the citizens, and therefore not all problems are generally reported. Another problem is that neighborhoods report issues to varying extents, resulting in biased data collection. For instance, a neighborhood with a lot of citizens that know how to report issues will have more detection, but the actual number of issues occurring is heavily skewed by the effects like this. This could be caused by a difference in income, education, immigration background or participation in general. Sections of the city with underprivileged population could potentially be underrepresented in the citizen reports, as it is shown in [33] that the volume of citizen reports is negatively correlated to voter turnout and census return rates, but positively correlated to the activity of campaign donations. Similarly, social multimedia use, technological literacy, and social participation levels, in general, may vary significantly throughout the city even in case of the "welfare states", which could introduce unwanted biases and negatively impact decision making [16, 17]. These biases can be further propagated to policies created using such collected data. In this paper we aim at gathering and processing participatory multimedia data in a more systematic and "objective" fashion, which facilitates an alternative approach to gaining insights about the locations in a city, allowing for better maintenance of public spaces and richer insights for creating policy. The data produced by the system can be combined by data collected through citizen reports to create new insights. It also opens up the possibility to detect objects in city streets that require attention before they become a nuisance.

The focus on detecting urban issues has been chosen due to the needs of the local governments and a large potential impact on neighborhood livability should the problems be addressed in a timely manner. To illustrate the scale and geographical distribution of these issues, Figure 2 presents a map summarizing a single month of citizens reporting issues with garbage on the street. Namely, during December of 2019 alone, the citizens of Amsterdam generated more than 12.500 reports about the issues related to garbage in the public space. In Section 4 we provide an example of automatically detecting such urban issues in real-time street-level imagery captured by the moving service vehicles.

Urban issues occur constantly and can't be tracked if street-level images are only recorded at long time intervals. This makes traditional sources such as Google Street View inadequate for the task. The use of social multimedia data or citizen reports is helpful, but the data collected with our system is of a different nature. Namely, the data collected by Urban Object Detection Kit will have a high collection frequency and high geographical density because the use of the system by the government is rewarded with actionable intelligence of what is occurring or needs to occur on the city streets. In Section 4 several examples are provided to illustrate the benefits of using the system. The high frequency with which government vehicles, such as garbage trucks, parking scanners, and local police vehicles, collect images will likely result in a high geographical resolution as well. Because of the real-time nature of the data that will be collected, this system can be considered as a living lab that can be easily deployed in various environments and use cases.

The Urban Object Detection Kit proposed in this paper provides multimedia research community access to real-time street-level imagery, allowing for a whole new range of research initiatives and applications of multimedia information retrieval. Also we conjecture that the proposed system and the data collected using it will open new possibilities for research in the fields of e.g. urban studies, criminology, and fairness and accountability, and serve as a general archival source of data about the state of the streets over the period in which it is collected.



Figure 2: Citizen reports about garbage in a single month of 2019 in Amsterdam. More than 12.500 reports about garbage are made, geographically spread all over the city.

We further conjecture that the amount of available street-level imagery will increase with technological advancements. For example, higher bandwidth networks will allow streaming an increasing amount of real-time footage. Also, the ubiquity of vehicles equipped with high-resolution cameras will be increasing when autonomous driving becomes more common. The system we propose, and the data it collects, will allow the field to advance even before such urban multimedia data is widely available.

The contributions of this paper can be summarized as follows:

- We develop an open-source framework for collecting, distributing and analyzing real-time street-level imagery as well as the actionable intelligence about urban issues.
- Our proposed framework empowers multimedia community with the technology and data for conducting experiments on real-time street-level imagery captured in a complex urban environment.
- We showcase a method for extracting valuable information about urban issues from the collected data and explore several examples of possible use cases.

2 RELATED WORK

Over the last decades, the multimedia community has contributed a large number of excellent approaches to gathering, indexing, analyzing and making sense of large amounts of heterogeneous data. This section provides a brief overview of related work on using urban multimedia data from various sources for analyzing processes in the city.

2.1 Social Multimedia

Content sharing and social networking platforms have been a cornerstone of collecting and analyzing urban multimedia [10, 28, 31,

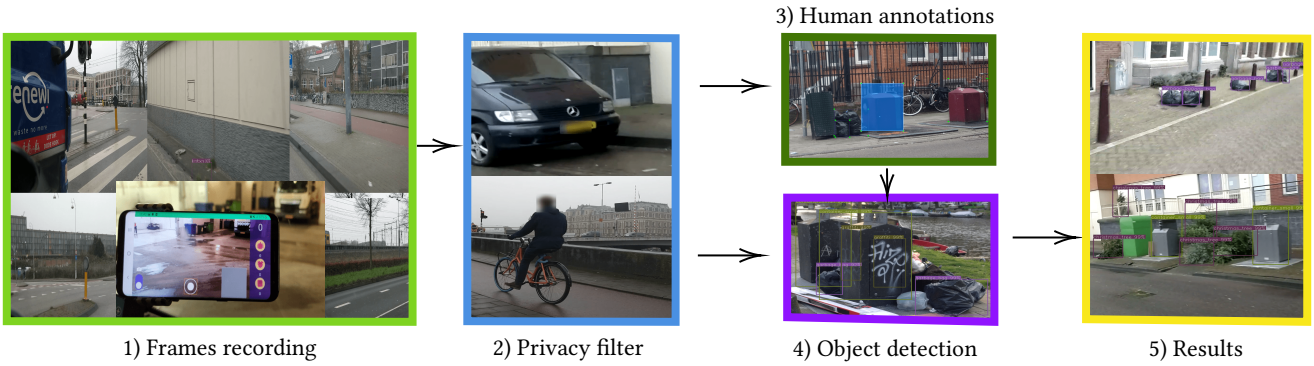

Figure 3: To collect and analyse real-time street level imagery, 1) *N* vehicles drive around the city with smartphones collecting data through a web app; 2) Faces and license plates are blurred; 3) Crowdsourcing is used for annotating the collected images; 4) Object detection models are trained and 5) used for detection of urban objects.

36]. The data from these platforms was successfully used for i.a. detecting social events such as graduations or a concert, or news events [1, 6, 8, 24, 38]. However, we need to detect urban issues that are of a different nature than the social or news events and are more difficult to find on social networking websites. Despite the significant effort invested in detecting multimedia events and even the introduction of successful international benchmarks serving as the epicenter of research on the topic [5, 13, 23], none of the existing collections or sources contain visual content with sufficient spatial and especially temporal granularity.

2.2 Street-level Imagery

The use of street-level imagery for multimedia retrieval tasks is a well-established field of research. Commercial products like Google Street View that use these techniques are widely available and commercially successful [4]. Previous work has focused on the annotation of such data [18] and its usage for tasks like monitoring urban assets [11] and crowd-mapping static assets [25]. Our system augments this field of work by providing a structured method to process real-time data. This can be leveraged for detecting urban issues that occur for brief observation intervals.

In [15] a dataset comparable to what is recorded by the Urban Object Detection Kit is presented that offers coordinates, direction, and spatial keywords, collected by volunteers. The system we introduce will be used by both volunteers and vehicles that are active in the cities around the world on a very regular basis, such as garbage trucks and local police vehicles. Also, the data will be annotated increasingly fine-grained with the growth of the system, by adding both textual description and the location of the described objects on the image.

In [9] an overview is given of the approaches using deep learning and computer vision techniques for understanding cities, but the data considered in the study is not captured or analyzed in real-time. Finally, in [29] Sugimoto et al. demonstrate that the videos recorded by people walking through a neighborhood can be used to create omni-directional video of a geographic area. This leads us to believe that our proposed framework could be used by the citizens as well and perhaps even in the fashion similar to life-logging [7].

2.3 Citizen as a Sensor Paradigm

As mentioned earlier, another paradigm to detecting urban issues relies on the utilization of citizen reports, routinely collected by a local government using e.g. Open311 [21]. For example, recent related work has shown that the urban issues can be efficiently detected based on multimedia information contained in citizen reports [30]. However, as described earlier, these reports are not as fine-grained with regard to categories of the issues, the location or the time frame. In this paper, we offer a viable alternative to citizen requests, by using the coverage of e.g. the vehicles from the parking services and garbage collection trucks that drive around the city continuously. This will provide a completely new source of information about the city.

2.4 Smartphones and Wearable Cameras

Using smartphones for data collection has been proven a cost-effective alternative for acquiring data of relatively high quality. Alavi et al. provide an overview of smartphone technology used for monitoring of construction sites and evaluating the road and pavement quality [2]. The conclusion is that using smartphones for monitoring civil engineering projects is promising due to a low cost and ubiquity. But to advance research it is required that academia, the public sector, and the private sector collaborate. The SDK we propose in this system has the goal to enhance this collaboration on a large scale. The registering of a person's environment using a wide range of sensors, popularly known as lifelogging, is another example of a field relying on the data collected using portable cameras worn by the individuals. Gurrin et al. provide a comprehensive summary of the multimedia information retrieval research using such data [7]. In this work, we aim to record visual and geographical data, not with the purpose of logging the life events of a single individual but registering urban dynamics.

2.5 Autonomous Driving Applications

With the advent of autonomous driving applications, the research on urban object detection intensified considerably, often focusing on the detection of traffic participants or anything related to traffic such as cars, roads, and cyclists. In such applications, radar is often

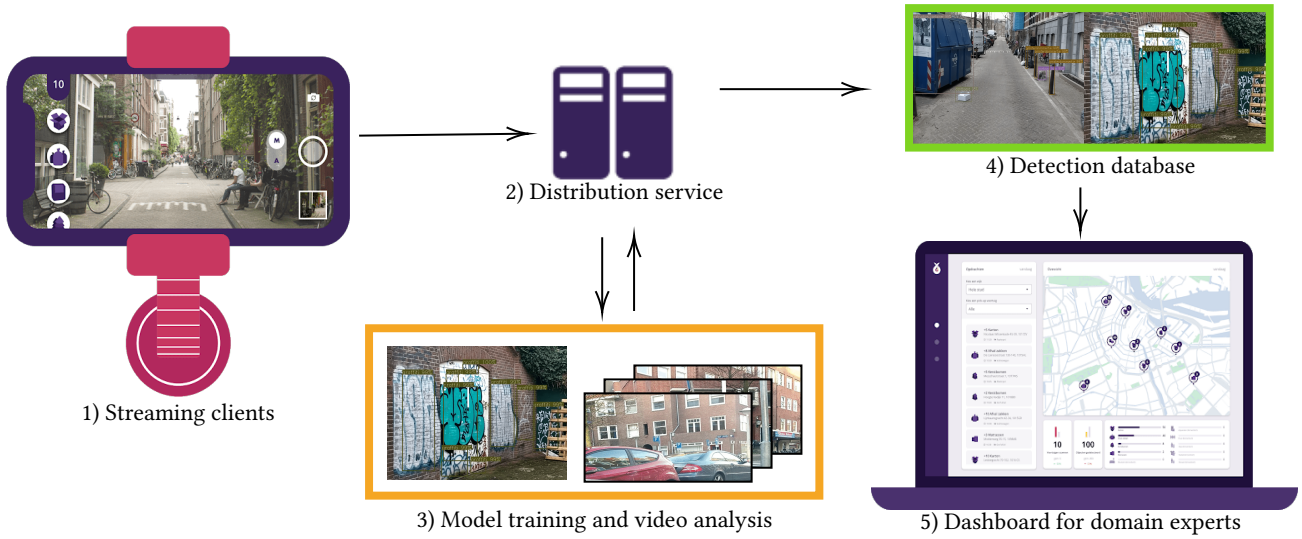

Figure 4: System pipeline - 1) Frame recording using the scanning clients. 2) Frames are sent to the distribution server 3) Frames are analysed by the machine learning workers. 4) Output data is stored, distributed through APIs and 5) displayed on a dashboard.

used instead of or in addition to visual data [22]. Similarly, point clouds gathered using LIDAR are frequently used as in [37] and [3]. In this work, we focus on data generated by smartphones as explored in Section 4.

3 APPROACH

In this section, we describe the system for collecting, distributing and analysing real-time street-level imagery. More information about the system can be found on [19] and the source code of the described components can be found in [20]. A schematic overview of the system is shown in Figure 4.

The system works by using smartphones attached to the vehicles of local government, while they drive around the city performing their already appointed tasks. Examples are garbage trucks, local police vehicles or vehicles that are automatically scanning for parking violations. The system can also be expanded to the vehicles of volunteers, or even integrated into cars that are already equipped with cameras, for example with the purpose of autonomous driving. Because of the low cost and ubiquity of smartphones, using them allows for cheap collection of data whenever setting up the system in a new environment. The decision for such a low-cost solution has been made to make the threshold of becoming a streaming client very low and therewith increase the number of data gatherers. It also makes the system scalable and possibly autonomous.

Whenever the system is recording, activated manually or in automatic mode by detecting movement of the vehicle, it will send recorded frames, their GPS location and the time of capturing to a central server. In this server object detection tasks can be performed, and frames can be distributed to human workers for annotations or stored for research purposes. In Section 4 we describe experiments with the initial use cases of the system, displaying information about urban issues to the civil servants. The system interface allows them to efficiently plan routes for maintaining, cleaning and patrolling neighbourhoods of the city.

3.1 Streaming Clients

The scanning client consists out of several elements – a smartphone, a phone holder, a car charger, a web app, and a sticker. It works by installing the phone on the dashboard of the car, where multiple clients can be placed on a single-vehicle. The phone is plugged into the car socket to prevent a power failure. Whenever the scanning commences the license plate is registered and the streaming is started, sending 3 frames each second to the server combined with the geographical location and the timestamp. Our progressive app can be accessed by almost any mobile device while behaving like a native app. In Figure 5 the interface of the scanning client is shown. Because the scanning of the environment can be perceived as controversial by some, a sticker is placed on the vehicle informing citizens what the scanner is doing and where they can find more information. The scanning client gives visual feedback about the detected objects. The client can also be put into automatic mode, starting the recording when the vehicles start moving.

Experiments have been performed using a Motorola Moto G8 plus smartphone, which has a 48MP camera with a 26mm (wide) lens. Testing on several other user devices did not show a significant difference in performance of the scanning client and the object detection. In our system the frames are reduced to 608x608 pixels for object detection. Thus it is likely that a lower-resolution camera will be as effective.

To prevent any invasion of privacy for the recording drivers, volunteers or civil servants, data recorded with the scanning client and made publicly available will not have information about the operator. To prevent an increase in workload for the operators, the scanning clients have an auto mode which does not require any



Figure 5: Example of a smartphone running the scanning client to the windshield of a car.

interaction. In the future, the technology can be integrated into vehicles, which will lead to a further decrease in the workload for the operators.

3.2 Distributing Service

The core of the whole system is the distributing service, which allows for a connection between a scalable number of streaming clients and a scalable number of frame processors. This will also allow for a larger scale of data collection, and as long as the processors can process the number of incoming frames the system can operate. The distribution service is a FastAPI [32] server which connects with the scanning clients using websockets to receive frames and metadata. Subsequently, it sends these to the frame analyzer using RabbitMQ [26].

3.3 Frame Analyzer

The system is currently configured to process all images using frame analyzers. These are machines that take frames from the distribution service and return the outcome of the analysis. The distribution service supports a varying number of frame analyzers, allowing the increase of capacity when required.

In Section 4 an object detection model is described that is used for the initial experiments. Because the system has to run in real-time and still be affordable, the video stream is analyzed frame by frame. A high inference speed is crucial for analysing large number of frames in real-time. For this reason no visual information is used to de-duplicate detected objects, but instead, metadata is used to filter the detections. This is achieved by only recording and streaming frames to the frame analyzer when the vehicle is moving, and by filtering out duplicate detections using their geocoordinates.

3.4 Privacy Filter

Since the collection of data in city streets can be a controversial subject due to privacy concerns, an important element of the proposed system is that all images are stored only after filtering out privacy-sensitive information. To allow for filtering privacy-sensitive information without causing any decrease in inference speed, license plates (cf. Figure 6) and faces (cf. Figure 7) are detected using the same model as for the object detection. For license plate detection, we devise a method similar to [12], but no recognition is used in this work since the detection is only used to remove the license plate from the images. For face detection we use a method similar to [34] and [14], which show that YOLOv3 [27] yields robust detections and high inference speed, while being capable of detecting even small faces. Another possibility of providing an even higher level of anonymity would be filtering out persons completely, which is possible with similar techniques. An even higher level of anonymity could be achieved by also filtering out complete vehicles and house numbers, to which the model could be adapted. Finally, the modular nature of our proposed framework would allow the integration of privacy-by-design solutions too, such as [35].

3.5 Detected Object Overview

The output data of the frame analysers can for example be used to create a dashboard for civil servants that gives them actionable intelligence about issues in the city. The detected objects can be displayed in a geographical interface, that shows where they where found. This can be further used to e.g. monitor how clean the streets are, but also to create the most efficient routes or detect areas of the city that need special attention.

Richly-annotated imagery will be released frequently, accompanied with the additional data coming from e.g. various GIS databases and neighbourhood-level statistics, to allow for diverse use cases and algorithm development and benchmarking on increasingly large scale.

Furthermore, data can be accessed in real-time using an API and used in e.g. frame-level object detection algorithms. This will allow for the real-time street imagery recorded from a moving vehicle to be used for deployment and testing of multimedia information retrieval algorithms in a dynamic, realistic and challenging setting.



Figure 6: Example of license plate filter

4 SHOWCASE

Since the system and the data it collects can be used for many different applications, this section will show and discuss some of them, starting with using an object detection model to find urban issues, then elaborating on the possibility for the citizens to submit service requests with minimal effort and allowing the domain experts and citizens alike to measure crowdedness of the city streets in a privacy friendly manner.



Figure 7: Example of face filter

The Urban Object Detection Kit is continuously being tested and developed for several use cases, and altered by using feedback from domain experts. At the time of writing this paper, several cities in the Netherlands are using the system to collect data for real-time street-level image analysis and output of these tests are continuously used to improve the system.

4.1 Street-level Real-time Object Detection

To show an example of a possible application, an object detector was trained on a small collection of images with several annotated classes that should be identified in real-time. The same model is used for both making the data anonymous and detecting relevant objects. The detected objects are displayed in a dashboard which is used by civil servants to effectively deploy their garbage trucks for cleaning, repair vehicles for maintenance of the public spaces and local police for enforcing. This will make the system valuable for the intended users, which will, in turn, promote its use. Another reason it is beneficial for local governments to use the system is that they have an affordable method for detecting urban issues, which complements "more traditional" approaches relying on citizen service requests. Because of the benefits the system gives to local governments, they will be motivated to deploy it frequently. This will result in a constant stream of real-time street-level imagery and continuous increase in the size of the collection, allowing for more interesting research possibilities. Finally, the areas that are not accessible by cars can also be accessed by boat, bike or on foot in a fashion similar to lifelogging.

The object detector has been trained on a small dataset of 1400 annotated recordings of street-level imagery. The classes that have been annotated can all be linked to a specific service of the local government – *garbage bags have to be removed* (cf. Figure 10) or *placing of portable toilet or construction containers needs to be verified by civil servant.* (cf. Figure 1). The detection can also be used to monitor the state of assets in city streets, e.g. the collecting of visual information about poles as shown in Figure 9.

The classes have been selected with domain experts from several different city maintenance departments: Garbage collection, asset maintenance and local police. While the most essential objects have been detected in the pilot study presented in this paper, in subsequent iterations the system will be expanded based on the demand of local governments.

For performing the task of object detection, a YOLOv3 model [27] was trained using 1288 annotated images, while the evaluation has been done on a test set of 112 images. An overview of the classes is given in Table 1. Object detection already yields useful results on the small dataset and we expect that the detection accuracy will keep improving with the increased training set size. Unsurprisingly, the classes with more training data and homogeneous visual appearance score the highest, as shown in Table 2. When annotating additional data we will seek to create a more balanced dataset that will allow for a much better performance on a highly heterogeneous set of objects.



Figure 8: Example of container and Christmas tree detection

Table 1: Number of annotated objects for each class in the training set used in our showcase.

Class	Number of objects
All	7986
Container small	1251
Garbage bag	1154
License plate filter	958
Face filter	894
Cardboard	888
Pole	751
Graffiti	706
Christmas tree	180
Mattress	50

The detected objects and their geographical location are displayed in a dashboard for civil servants, and used for route optimization, creating policy or researching the city (cf. Figure 4).

Both the performance and the number of classes will increase with a more intensive use of the system. Once more data is collected, richly annotated by local governments for the classes of their interest and made publicly available, various types of detection models will be created and evaluated.

Table 2: The result of the object detection experiment.

Class	N objects	P	R	mAP	F1
All	514	0.21	0.74	0.61	0.31
Container small	120	0.38	0.95	0.94	0.54
Garbage bag	92	0.13	0.87	0.70	0.23
License plate filter	86	0.25	0.78	0.60	0.38
Face filter	62	0.22	0.79	0.62	0.34
Cardboard	59	0.18	0.59	0.43	0.28
Graffiti	43	0.12	0.86	0.73	0.21
Pole	39	0.47	0.95	0.87	0.63
Christmas tree	12	0.11	0.83	0.59	0.19
Mattress	1	0	0	0	0



Figure 9: Example of garbage bag, traffic bollard (pole) and cardboard detection

4.2 Other Possible Applications

The availability of real-time street-level imagery allows for new types of research, several of which will be discussed in this section.

4.2.1 Visual citizen reports. When citizens make a report about issues in the city streets, they often have to fill out a form or make a call to a designated phone number. By giving them access to the web app, they could simply report an issue and it will be automatically routed to the right department. We further conjecture that making the reporting process easier and less laborious would positively reflect on citizen participation and ultimately lead to more relevant reports with a clearer information about the task at hand. Because textual descriptions are not always completely clear, this can cause a delay when the service requests are routed to the wrong department. Since a visual registration of the object can be confirmed, this will likely work more effectively for issues with clear visual properties. In Figure 1 we illustrate how this would work for garbage bags, poles and portable toilets and in Figure 10 how it would help reporting garbage next to containers.

4.2.2 Crowdedness monitor. Cities around the world are getting increasingly busy and crowded, which makes accurately monitoring the size of the crowd useful for tracking city dynamics, creating policy and ensuring safety. The system can be used for monitoring crowdedness on street level and in real time by e.g. counting the different types of vehicles, bikes or people (while anonymous) present in the streets.



Figure 10: If a citizen makes a recording of garbage, it can be detected that garbage bags are placed next to containers. This shows that garbage bags need to be picked up and it is a strong indication that the container is full or broken.

5 CONCLUSION

We proposed a system for collection, distribution and analysis of real-time street-level imagery. In addition to providing local governments with timely insights about the issues in the public space that need their attention, the proposed system facilitates the development and testing of multimedia information retrieval algorithms in a complex urban setting. Our system is released as open source, together with a dataset already collected during a pilot intended to test viability of the system. We further showcased the potential of detecting objects related to the issues that occur in urban environments and discussed several other potential use cases. In the future work, we will extend our openly available and richly annotated collection of street-level urban imagery collected in several cities and test the framework on a much larger and diverse set of urban objects.

5.1 Acknowledgements

This project has received funding from the City of Amsterdam.

The authors express their gratitude to the members of the development team: Claudia Pinhão, Mark van der Net, Stan Guldemond, Sven Brilleman and Youssef Kassem.

REFERENCES

[1] Kashif Ahmad, Nicola Conci, Giulia Boato, and Francesco GB De Natale. 2016. USED: a large-scale social event detection dataset. In *Proceedings of the 7th International Conference on Multimedia Systems*. 1–6.

[2] Amir H Alavi and William G Buttlar. 2019. An overview of smartphone technology for citizen-centered, real-time and scalable civil infrastructure monitoring. *Future Generation Computer Systems* 93 (2019), 651–672.

[3] Yassin Alkhalili, Manisha Luthra, Amr Rizk, and Boris Koldehofe. 2019. 3-D Urban Objects Detection and Classification From Point Clouds. In *Proceedings of the 13th ACM International Conference on Distributed and Event-based Systems*. 209–213.

[4] Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. 2010. Google street view: Capturing the world at street level. *Computer* 43, 6 (2010), 32–38.

[5] George Awad, Asad A. Butt, Jonathan Fiscus, David Joy, Andrew Delgado, Martial Michel, Alan F Smeaton, Yvette Graham, Wessel Kraaij, Georges Quénot, Roeland J.F. Ordelman, Gareth J.F. Jones, and Benoit Huet. 2017. Trecvid 2017: Evaluating ad-hoc and instance video search, events detection, video captioning and hyperlinking. In *Proceedings of TRECVID*, Vol. 2017.

[6] Longbiao Chen, Jérémie Jakubowicz, Dingqi Yang, Daqing Zhang, and Gang Pan. 2017. Fine-Grained Urban Event Detection and Characterization Based on Tensor Cofactorization. *IEEE Trans. Human-Machine Systems* 47, 3 (2017), 380–391.

[7] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Found. Trends Inf. Retr.* 8, 1 (June 2014), 1–125. https://doi.org/10.1561/1500000033

[8] Yuqian Huang, Yue Li, and Jie Shan. 2018. Spatial-Temporal Event Detection from Geo-Tagged Tweets. *ISPRS International Journal of Geo-Information* 7, 4 (2018), 150.

[9] Mohamed R Ibrahim, James Haworth, and Tao Cheng. 2020. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities* 96 (2020), 102481.

[10] Bogdan Ionescu, Adrian Popescu, Mihai Lupu, Alexandru Lucian Gundefinednscundefined, Bogdan Boteanu, and Henning Müller. 2015. Div150Cred: A Social Image Retrieval Result Diversification with User Tagging Credibility Dataset. In *Proceedings of the 6th ACM Multimedia Systems Conference* (Portland, Oregon) *(MMSys '15)*. Association for Computing Machinery, New York, NY, USA, 207–212. https://doi.org/10.1145/2713168.2713192

[11] Vladimir A Krylov, Eamonn Kenny, and Rozenn Dahyot. 2018. Automatic discovery and geotagging of objects from street view imagery. *Remote Sensing* 10, 5 (2018), 661.

[12] Rayson Laroca, Luiz A Zanlorensi, Gabriel R Gonçalves, Eduardo Todt, William Robson Schwartz, and David Menotti. 2019. An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. *arXiv preprint arXiv:1909.01754* (2019).

[13] Martha Larson, Mohammad Soleymani, Pavel Serdyukov, Stevan Rudinac, Christian Wartena, Vanessa Murdock, Gerald Friedland, Roeland Ordelman, and Gareth J. F. Jones. 2011. Automatic Tagging and Geotagging in Video Collections and Communities. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval* (Trento, Italy) *(ICMR '11)*. Association for Computing Machinery, New York, NY, USA, Article Article 51, 8 pages. https://doi.org/10.1145/1991996.1992047

[14] Chong Li, Rong Wang, Jinze Li, and Linyu Fei. 2020. Face Detection Based on YOLOv3. In *Recent Trends in Intelligent Computing, Communication and Devices*. Springer, 277–284.

[15] Ying Lu, Hien To, Abdullah Alfarrarjeh, Seon Ho Kim, Yifang Yin, Roger Zimmermann, and Cyrus Shahabi. 2016. GeoUGV: User-generated mobile video dataset with fine granularity spatial metadata. In *Proceedings of the 7th International Conference on Multimedia Systems*. 1–6.

[16] Sako Musterd, Szymon Marcińczak, Maarten van Ham, and Tiit Tammaru. 2017. Socioeconomic segregation in European capital cities. Increasing separation between poor and rich. *Urban Geography* 38, 7 (2017), 1062–1083. https://doi.org/10.1080/02723638.2016.1228371 arXiv:https://doi.org/10.1080/02723638.2016.1228371

[17] Sako Musterd and Wim Ostendorf. 2013. *Urban segregation and the welfare state: Inequality and exclusion in western cities*. Routledge.

[18] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE International Conference on Computer Vision*. 4990–4999.

[19] City of Amsterdam. 2020. Object Detection Kit. http://www.odk.ai.

[20] City of Amsterdam. 2020. Object Detection Kit Repository. https://gitlab.com/odk/odk-stack.

[21] Dietmar Offenhuber. 2015. Infrastructure legibility—a comparative analysis of open311-based citizen feedback systems. *Cambridge Journal of Regions, Economy and Society* 8, 1 (2015), 93–112.

[22] Rodrigo Pérez, Falk Schubert, Ralph Rasshofer, and Erwin Biebl. 2019. Deep Learning Radar Object Detection and Classification for Urban Automotive Scenarios. In *2019 Kleinheubach Conference*. IEEE, 1–4.

[23] Georgios Petkos, Symeon Papadopoulos, Vasileios Mezaris, Raphael Troncy, Philipp Cimiano, Timo Reuter, and Yiannis Kompatsiaris. 2014. Social event detection at MediaEval: a three-year retrospect of tasks and results. In *Proc. ACM ICMR 2014 Workshop on Social Events in Web Multimedia (SEWM)*.

[24] Shengsheng Qian, Tianzhu Zhang, and Changsheng Xu. 2016. Multi-modal multi-view topic-opinion mining for social event analysis. In *Proceedings of the 24th ACM international conference on Multimedia*. ACM, 2–11.

[25] Sihang Qiu, Achilleas Psyllidis, Alessandro Bozzon, and Geert-Jan Houben. 2019. Crowd-Mapping Urban Objects from Street-Level Imagery. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1521–1531. https://doi.org/10.1145/3308558.3313651

[26] RabbitMQ. [n.d.]. RabbitMQ. https://github.com/rabbitmq.

[27] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[28] S. Rudinac, A. Hanjalic, and M. Larson. 2013. Generating Visual Summaries of Geographic Areas Using Community-Contributed Images. *IEEE Transactions on Multimedia* 15, 4 (June 2013), 921–932. https://doi.org/10.1109/TMM.2013.2237896

[29] Naoki Sugimoto, Yuko Iinuma, and Kiyoharu Aizawa. 2019. Walker's Movie Map: Route Vies Synthesis Using Omni-Directional Videos. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 1050–1052. https://doi.org/10.1145/3343031.3350587

[30] Maarten Sukel, Stevan Rudinac, and Marcel Worring. 2019. Multimodal Classification of Urban Micro-Events. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 1455–1463. https://doi.org/10.1145/3343031.3350967

[31] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: The New Data in Multimedia Research. *Commun. ACM* 59, 2 (Jan. 2016), 64–73. https://doi.org/10.1145/2812802

[32] Tiangolo. [n.d.]. FastAPI. https://github.com/tiangolo/fastapi.

[33] Ariel White and Kris-Stella Trump. 2018. The promises and pitfalls of 311 data. *Urban affairs review* 54, 4 (2018), 794–823.

[34] Wang Yang and Zheng Jiachun. 2018. Real-time face detection based on YOLO. In *2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII)*. IEEE, 221–224.

[35] Tengqi Ye, Brian Moynagh, Rami Albatal, and Cathal Gurrin. 2014. Negative FaceBlurring: A Privacy-by-Design Approach to Visual Lifelogging with Google Glass. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management* (Shanghai, China) *(CIKM '14)*. Association for Computing Machinery, New York, NY, USA, 2036–2038. https://doi.org/10.1145/2661829.2661841

[36] J. Zahálka, S. Rudinac, and M. Worring. 2015. Interactive Multimodal Learning for Venue Recommendation. *IEEE Transactions on Multimedia* 17, 12 (Dec 2015), 2235–2244. https://doi.org/10.1109/TMM.2015.2480007

[37] Younes Zegaoui, Marc Chaumont, Gérard Subsol, Philippe Borianne, and Mustapha Derras. 2019. Urban object classification with 3D Deep-Learning.

[38] Sicheng Zhao, Yue Gao, Guiguang Ding, and Tat-Seng Chua. 2017. Real-time multimedia social event detection in microblog. *IEEE transactions on cybernetics* 48, 11 (2017), 3218–3231.