UvA-DARE (Digital Academic Repository)

Detecting Urban Issues With the Object Detection Kit

Sukel, M.; Rudinac, S.; Worring, M.

DOI

[10.1145/3394171.3414427](https://doi.org/10.1145/3394171.3414427)

Publication date

2020

Document Version

Final published version

Published in

MM '20

License

Article 25fa Dutch Copyright Act

[Link to publication](#)

Citation for published version (APA):

Sukel, M., Rudinac, S., & Worring, M. (2020). Detecting Urban Issues With the Object Detection Kit. In *MM '20: proceedings of the 28th ACM International Conference on Multimedia : October 12-16, 2020, Virtual Event, USA* (pp. 4518–4520). Association for Computing Machinery. <https://doi.org/10.1145/3394171.3414427>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Detecting Urban Issues With the Object Detection Kit

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 Worrying
University of Amsterdam
Amsterdam, The Netherlands
m.worrying@uva.nl

ABSTRACT

This technical demo will present the Object Detection Kit, a system capable of collecting, analyzing and distributing street level imagery in real-time. It provides civil servants with the actionable intelligence about issues on city streets and, at the same time, equips the multimedia research community with a framework and data facilitating easy deployment and testing of algorithms in a challenging urban setting. The system is available as open source. In the Object Detection Kit demo we will demonstrate how the framework can be used to detect urban issues and showcase the capabilities of the system.

CCS CONCEPTS

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

KEYWORDS

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

ACM Reference Format:

Maarten Sukel, Stevan Rudinac, and Marcel Worrying. 2020. Detecting Urban Issues With the Object Detection Kit. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20)*, October 12–16, 2020, Seattle, WA, USA. Seattle, United States, 3 pages. <https://doi.org/10.1145/3394171.3414427>

1 INTRODUCTION AND RELATED WORK

Cities around the world are increasing their sustainability efforts, but at the same time the number of people living in cities is increasing, making it more difficult to keep the public spaces clean, safe and well-maintained. To help local governments in this challenging task, novel solutions are needed facilitating more efficient functioning of city services. In response to this call to action, the City of Amsterdam has already deployed a multimedia analysis system that automatically categorizes and yearly routes hundreds of thousands of citizen reports about the issues in the public environment [5]. However, the action is often taken only when citizens report issues, which results in an imbalance in the maintenance of the public environment. A neighborhood with plenty of citizens that actively report issues will be maintained better and more frequently than neighborhoods where citizens are less inclined to notify the local

Figure 1: The streaming app, that gives the user feedback about what urban objects are detected

government of any issues. To solve this imbalance in city maintenance, and support sustainability efforts new methods are required to automatically and quickly identify issues in the public space, even before they become nuisance for the citizens.

A potential solution would be to use computer vision for monitoring issues on city streets, which would allow for more efficient maintenance, planning and decision making. However, to fully capitalize on this potential, an inexpensive and widely deployable framework for real-time acquisition, distribution and analysis of street-level imagery is needed. This framework should use cameras with network capabilities that are low in cost and widely available, such as smartphones. In addition, the system should provide an incentive for the local governments to ensure sustainable deployment in a real-world setting.

Recent advances in multimedia and computer vision technologies make it possible to accurately discriminate between a variety of urban object categories [7, 8], and data sets such as Cityscapes [2] and Mapillary [3] supply the data required to train such models. However, the object categories commonly featured in existing work do not explicitly focus on urban objects that require attention from the local government. Due to the close cooperation with domain experts and integration with local government, the framework we will demonstrate allows for the creation of a large collection of real world annotated street-level imagery. This real-time street level imagery is also different from panoramic images that are captured at a lower frequency, such as with Google Street View [1]. Highly dynamic nature of urban environments requires a much higher acquisition frequency for extracting useful and actionable intelligence.

In this demo we demonstrate the Object Detection Kit [4, 6], a system for the real-time collection and analysis of street-level imagery. Besides operating in real-time, the system is designed to be affordable and portable, to ensure deployment and sustained use in a dynamic urban environment. To realize a large geographical

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MM '20, October 12–16, 2020, Seattle, WA, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7988-5/20/10.

<https://doi.org/10.1145/3394171.3414427>

Figure 2: (1) Scanners are attached to vehicles, and frame streaming to the distribution server is started. A2) Privacy sensitive information is filtered out from the images that will be used for annotation. A3) Together with domain experts the data is annotated for relevant urban objects. B2) The data is analysed and B3) actionable intelligence is shown to civil servants.

coverage, the system offers local governments an incentive in the form of actionable intelligence about issues in the city. We have worked together closely with domain experts and civil servants to create a system that informs the local government of essential actionable intelligence about issues on the streets of their city. In addition, the system collects large quantities of real-time street level imagery. This collection allows for the deployment and testing of multimedia analysis algorithms in a highly challenging urban setting and at large scale. Effectively, we are transforming entire cities into urban living labs where a wide range of algorithms and use cases can be tested.

2 THE SYSTEM

Figure 2 shows an system overview. It consists of a progressive web app that can run on smartphones. Due to the simple deployment of the scanners, the system is well suited for the dynamic nature of urban environments. The smartphones can be operated by hand, or attached to a vehicle such as a bike, a garbage truck, or a car. When the scanners start streaming frames they are sent to a server that distributes the frames over several machines that will analyse the images. The frames are analyzed by automatically detecting urban objects, while at the same time blurring out any privacy sensitive information such as license plates and faces.

The output data of the analyzed frames, combined with the geographical location from the smart phones GPS is used to create a detailed map of issues in the city and presented to civil servants in a dashboard. This data can for example be used for collecting resources more efficiently and sustainable. For example, if it is known where mattresses are located they can be collected separately from other bulky waste and brought too a specialized facility where they

can be turned into resources. In addition, the objective information about urban issues can be used for identifying and mitigating possible bias and/or unfairness in maintaining different neighbourhoods. For improved efficiency, the same model is used to detect urban objects and remove any privacy sensitive data from the images. Afterwards, the anonymous visual data is annotated in close co-operation with domain experts responsible for city maintenance. Such created dataset will provide research community with large collection of richly annotated street-level imagery. Availability of high-quality research data, together with a modular system design, will allow for easy integration of successful community-contributed algorithms and their use in improving city livability.

3 OBJECT DETECTION KIT DEMO

In the Object Detection Kit demo we will demonstrate how the framework can be used to detect urban issues. We will focus on how different components of the system work together, and showcase the capabilities of the system. We will demonstrate the web app, as shown in Figure 1 that streams frames while receiving feedback about detections in real time. In addition, we will show how the framework distributes the frames over several machines with GPUs to analyse larger quantities of frames. Every machine with a GPU, for example a cloud instance or even a laptop, can be turned into a worker that processes images. The distribution server will distribute the images between all the active workers.

Finally, we will demonstrate how the system is being used by local government in the streets of Amsterdam. When conditions for the demo allow it, the area near the conference venue will be scanned to give an overview of registered issues in the neighborhood.

REFERENCES

- [1] 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.
- [2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3213–3223.
- [3] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Buló, and Peter Kotschieder. 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE International Conference on Computer Vision*. 4990–4999.
- [4] City of Amsterdam. 2020. Object Detection Kit. <http://www.odk.ai>.
- [5] 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>
- [6] 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*. 509–516.
- [7] Andrew Tao, Karan Sapra, and Bryan Catanzaro. 2020. Hierarchical Multi-Scale Attention for Semantic Segmentation. *arXiv preprint arXiv:2005.10821* (2020).
- [8] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. 2020. Resnest: Split-attention networks. *arXiv preprint arXiv:2004.08955* (2020).