Data Article

A dataset of images of public streetlights with operational monitoring using computer vision techniques

Ioannis Mavromatis*, Aleksandar Stanoev, Pietro Carnelli*, Yichao Jin, Mahesh Sooriyabandara, Aftab Khan*

Bristol Research and Innovation Laboratory, Toshiba Europe Ltd., 32 Queen Square, Bristol BS1 4ND, United Kingdom

ARTICLE INFO

Article history:
Received 16 September 2022
Accepted 3 October 2022
Available online 7 October 2022

Dataset link: Images of Public Streetlights with Operational Monitoring using Computer Vision Techniques (Original data)

Keywords:
Streetlight
Street furniture
Light and emergency light maintenance
Lighting columns
Computer vision
Machine learning
Smart cities

ABSTRACT

A dataset of street light images is presented. Our dataset consists of ~350 k images, taken from 140 UMBRELLA nodes installed in the South Gloucestershire region in the UK. Each UMBRELLA node is installed on the pole of a lamppost and is equipped with a Raspberry Pi Camera Module v1 facing upwards towards the sky and lamppost light bulb. Each node collects an image at hourly intervals for 24 h every day. The data collection spans for a period of six months. Each image taken is logged as a single entry in the dataset along with the Global Positioning System (GPS) coordinates of the lamppost. All entries in the dataset have been post-processed and labelled based on the operation of the lamppost, i.e., whether the lamppost is switched ON or OFF. The dataset can be used to train deep neural networks and generate pre-trained models providing feature representations for smart city CCTV applications, smart weather detection algorithms, or street infrastructure monitoring. The dataset can be found at 10.5281/zenodo.6046758.

© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
### Specifications Table

| Subject | Computer Science |
|---------|------------------|
| Specific subject area | Computer Vision and Pattern Recognition |
| Type of data | The image data for each street light column are provided as JPEG files [1]. Each zipped directory contains all the images associated with a single light column/lamppost. In addition, a CSV file containing the information about all the street lights, their metadata and the status decision outcome from our post-processing algorithm, i.e., if it is switched ON or OFF. We also include another processed CSV file with the number of occurrences per street light column for easier post-processing of the data. |
| How the data were acquired | Acquired from the UMBRELLA testbed [2], consisting of 140 UMBRELLA nodes, each equipped with a Raspberry Pi Camera Module V1 [3] and an OmniVision OV5647 sensor [4]. The nodes are installed on top of lampposts around the South Gloucestershire region in the UK, across a ~7.2km stretch of road and the University of West of England Campus. Note that all cameras are facing upwards, towards the light bulbs of the lampposts and the sky. |
| Data format | Raw data format, as recorded by each UMBRELLA node, pertaining to the UMBRELLA testbed, in the form of JPEG and CSV files. The raw data are analysed to identify the street lights' functionality, and the results are listed in a CSV file format. |
| Description of data collection | The data has been recorded on hourly intervals at a best-effort basis (for all available nodes on each hourly timeslot), for seven days per week and for a period of six months. A random time delay of 0s to 600s was introduced before taking each photo to spread the analysis resource utilisation on the server side. |
| Data source location | Roughly ~80% of the nodes are installed on a ~7.2 km stretch of public road in the Gloucestershire region in the UK, comprising the following roads: Coldharbour Lane, Filton Road, and A4174 Ring Road. In addition, ~20% of the nodes are installed around the University of the West of England Frenchay Campus in Bristol, UK. |
| Data accessibility | Data available online at the link below: 10.5281/zenodo.6046758 [5] |
| Related research article | |

### Value of the Data

- Our dataset of ~350 k JPEG street light images provides unique camera placements, photographic angles, and distances between the different street lights. Furthermore, partial obstructions by vegetation, street lights outside the camera’s Field of View (FoV) and images altered by the weather conditions provide a unique dataset for Machine Learning (ML) use-cases for inspection, monitoring and. In [6], the feasibility of using centralised, personalised, and federated learning is presented in a streetcare IoT application context.
- Building upon the provided images, Computer Vision, Internet of Things (IoT), and Smart Cities experts can design models and heuristic tools to assess the status of the street and emergency lights in real-time (e.g., as in [7]). This can facilitate further research around Smart City services that automatically detect whether a street light is ON or OFF, alert a maintenance team, and reduce the person-hours required for on-site monitoring.
- Deep Neural Networks (DNNs) and the models’ regularisation can benefit from pre-trained models [8]. Our dataset provides an extensive collection of images taken at different times (day/night), under different weather conditions and exposure settings. ML models pre-trained on this dataset can provide excellent feature representations in DNN-based transfer learning applications, e.g., outdoor smart city CCTV deployments.
- Since the cameras face the sky, this dataset can be used to train real-time weather warnings and traffic management systems. For instance, by building a simple rain detector and estimating the direction of the rain, weather warnings can be generated, pin-pointed very precisely and disseminated to drivers approaching specific streets.
• Finally, the diverse set of streetlamp images provided can be combined with datasets from other street furniture (e.g., traffic lights, street name signs, traffic signs, etc.) or sensors (e.g., LiDARs). Such a dataset can be later used for object recognition algorithms (with sensor fusion) for autonomous vehicles and drone navigation systems on public road infrastructures, e.g., as in [9].

1. Objective

The objective for generating this dataset was to provide a unique and diverse dataset of images that can be used for inspection, monitoring and maintenance within IoT ecosystems and object recognition use-cases (if combined with other datasets). We designed the dataset in such a way so the images are diverse and unique. This comes from the fact that the camera’s FoV, obstructions, and weather conditions affect the image quality providing unique edge use-cases for testing and experimentation. Hence, such a dataset can introduce various challenges when designing detection and classification algorithms, training tools and machine learning models and can benefit researchers in the areas of Computer Vision, Smart Cities, and Machine Learning.

2. Data Description

The raw data files associated with each lamppost (namely the JPEG [1] images and the CSV file) are organised as follows. The dataset is uploaded as a single Zip file, which, when unzipped, unfolds to a number of zipped and two CSV files. As our dataset is fairly large, we provide two Zip files, one containing the complete dataset and another one containing an example dataset with a smaller number of JPEG images and UMBRELLA nodes. The naming convention for both is “streetcare-dataset-complete.zip” and “streetcare-dataset-example.zip”, respectively. The CSV files within the root directory follow a similar naming convention, i.e., “occurrences-complete.csv”, “streetlights-complete.csv” for the complete dataset and “occurrences-example.csv”, “streetlights-example.csv” for the example one.

One can find the Zip files containing our JPEG images within the root zipped directory. These zipped files are named after the “serial ID” of each UMBRELLA node, i.e., a friendly name given to each node. This serial ID is in the form of “RS[ES]-[-A-Z0-9]”. For example, valid serials are “RSE-A-11-C”, “RSE-A-12-C”, “RSE-A-47-C”, etc. All serial IDs represent the type of the node and the sensors equipped, but this is outside this dataset’s scope. With regards to this dataset, all nodes are equipped with the same Raspberry Pi Camera Module (i.e., Camera Module V1 [3]). When unzipped, all the “RSE-” or “RSS-” subdirectories contain the JPEG images taken of this particular UMBRELLA node. The file name of each JPEG file is a random string of characters and numbers with a size of 32 digits. This string is assigned randomly when the file is created and is always unique. An example of the directory tree of our dataset can be seen below:

```
streetcare-dataset-complete.zip
  occurrences-complete.csv
  streetlights-complete.csv
  RSE-6-C.zip
    9397c82e6daf0d6435e13552e2e7d2ca.jpg
    9196c26e24a1ed938d8a170939833c79.jpg
    ...
  RSE-A-12-C.zip
    d69bb15b397f33b2f706f89926f904a9.jpg
    f1b60e40d3d295a9eee2971ce5e426bc.jpg
    ...
streetcare-dataset-example.zip
  occurrences-example.csv
  streetlights-example.csv
  RSE-L-1-C.zip
    cf0f229123a6b1ac659a9f51366c469f.jpg
    b8fb64ece07c34a21791afcf7fad72f.jpg
    ...
  RSS-12-C.zip
    e27bd4046e861690f95da80ed8b628b.jpg
    8aeedbbdf8a033bc808be3feaeaa34.jpg
    ...
```
Table 1
Definition of the labels of each column of the CSV data files in the dataset.

| Field Name | Definition |
|------------|------------|
| id         | A unique ID given to each node. It increments by one for each image captured. |
| serial     | The UMBRELLA nodes “serial ID” (friendly name). |
| date       | Date and timestamp at the moment an image was captured. |
| hostname   | The UMBRELLA nodes hostname, as seen in the host OS. |
| lat        | Latitude of the position of the current node. |
| lon        | Longitude of the position of the current node. |
| image_name | The JPEG image file name (32-digit alphanumeric string with “.jpg” extension). |
| fault_detected | The post-processed result of the streetlights status (operational or not). |
| confidence | The probability of the “fault detected” result is correct. |
| daylight   | The astrological day or night time cycle. |
| red        | The pixel intensity as perceived under the “red” RGB channel. |
| green      | The pixel intensity as perceived under the “green” RGB channel. |
| blue       | The pixel intensity as perceived under the “blue” RGB channel. |

With regards to the CSV files, the “streetlights-XXX.csv” contains in a tabular format all the information about the street lights, their metadata and the status decision outcome from our post-processing algorithm, i.e., ON or OFF. The first line of the file shows the label of each column. The following rows are considered the “entries”; each represents a JPEG image taken from an UMBRELLA node at a specific time. The “streetlights-XXX.csv” files are structured in the following format:

Column_Label_0;Column_Label_1;... Column_Label_n
Value_0_Entry_0;Value_1_Entry_0;... Value_n_Entry_0
Value_0_Entry_1;Value_1_Entry_1;... Value_n_Entry_1
...
Value_0_Entry_m;Value_1_Entry_m;... Value_n_Entry_m

A detailed explanation of each column found in the “streetlights-XXX.csv” file can be found in Table 1. The “occurrences-XXX.csv” contains a list of serial IDs from all nodes in the dataset, and the number of images taken from this specific node. This file is ordered in descending order based on the number of images. As before, the file is in a tabular format with the first line being the labels of the columns, i.e., “serial” and “occurrences”, and later each line is an entry, i.e., a node serial ID and the number of occurrences in the dataset.

3. Experimental Design, Materials and Methods

All UMBRELLA nodes are installed across multiple locations in the South Gloucestershire region of the UK, these being a ~7.2 km stretch of public road (about ~80% of the nodes) and the University of the West of England (UWE) Frenchay Campus (about ~20% of the nodes). The nodes’ locations and regions can be seen in Fig. 1. All UMBRELLA nodes are connected to our backend servers via fibre or a wireless (WiFi) interface. Our servers are used to collect all the images taken and store the information in our database.

Fig. 2 shows an Umbrella node attached to a lamppost. Each rhombus segment contains custom PCBs, designed and manufactured by Toshiba, accommodating ten sensors (e.g., Bosch BME680, accelerometers, microphones, etc.) and seven network interfaces (Bluetooth, fibre, WiFi, LoRa, etc.). At the top of the node, one can find a Raspberry Pi Camera Module v1 [3], used for taking the street light images. The camera module is equipped with an OmniVision OV5647 sensor [4]. All PCBs are connected to a central processing unit (Raspberry Pi 3b+ Compute Module [10]), responsible for controlling the sensors and the camera module, processing the data generated and sending them to the backend servers.

All available nodes periodically take a photo of the lamppost and send it back to our backend for processing and storing. The photos are taken at hourly intervals and collected for 24h time
All nodes are installed on public lampposts across a road of \( \sim 7.2 \) km. The colours represent the nodes connectivity, i.e., green is fibre connected, blue is WiFi.

Fig. 2. UMBRELLA Node on a lamppost, its exploded view, and the camera placement at the top of the node (seen in the red circles).

Table 2
Camera module and sensor settings.

| Settings         | Day       | Night | Night - High Exposure |
|------------------|-----------|-------|----------------------|
| Exposure Mode    | Automatic | Manual|                      |
| ISO              | 100       | 400   | 800                  |
| Shutter Speed    | Automatic | 250000| 600000               |
| LED State        | Off       | Off   |                      |
| Flash            | Off       |       |                      |
| Resolution       | 1024 \times 768 | | |
| Contrast         | 0         |       |                      |
| Brightness       | 50%       |       |                      |

All lampposts operate (are turned ON or OFF) based on the astronomical nighttime, i.e., they are turned ON 15 min before the astronomical dusk and are turned OFF 15 min after the astronomical dawn. Based on that day/night cycle, we change the configuration of our camera accordingly. The different configurations applied to the camera sensor are listed in Table 2. During the "day" time, i.e., when a lamppost is expected to be OFF, a camera is left in automatic exposure and shutter speed settings to compensate for the available sunlight under different
weather conditions. During the “night” time, i.e., when a lamppost is expected to be ON, the exposure, ISO, and shutter speed are increased to compensate for the low-light conditions. During the night time particularly, two images are collected with different configurations. The first one (column "Night" in Table 2) is the photo saved as part of our dataset. The second configuration (column "Night - High exposure" in Table 2) is used for our post-processing and the labelling of the dataset. Two examples of the same photo taken with and without the increased exposure can be seen in Fig. 3. The astrological day and night cycle for each individual photo taken is reported in the column day/night in our CSV file and calculated using the Astral Python library [11].

Each lamppost image is post-processed on the server-side to identify whether the lamppost is ON or OFF based on its expected behaviour and day/night cycle. More specifically, all photos are converted to a 3-channel Red-Green-Blue (RGB) format with a range of values between 0 to 255. The values are reported in our CSV file under the red, green and blue columns. During the nighttime, the results reported under the RGB columns come from the highly exposed photos. When the three channels are reported, the streetlight status is calculated, i.e., operational or not (fault_detected column in Table 1) and a confidence interval (confidence column in Table 1). As operational, it is considered a streetlight when turned OFF during the day or turned ON during the night cycles. On the other hand, a fault is reported when a streetlight is ON during the day or OFF during the night. An example of how the data are reported and visualised in our front end is shown in Fig. 4.

With regards to the fault detection mechanism, during the day, all images are validated against an ML-trained model. Our model is based on a pre-trained VGG-16 model (using the Imagenet dataset). We initially optimised the final layers of the model for the lamppost ON/OFF classification, followed by some fine-tuning for the entire model. The model’s output is a Softmax layer working as a binary classifier (ON/OFF). Our ML model operates with an accuracy of circa ≥ 90%. The neural network architecture can be seen in Fig. 5. More information about the
initial VGG-16 model can be found in [12]. During the night, the node’s RGB values are used to identify its status. For the night use-case, we use the highly exposed images and the (raw) RGB values. More specifically, when the median RGB value is $\geq 200$, the lamppost is considered ON, and the confidence is nearly 1. On the other hand, when $\text{RGB}_{\text{median}} \leq 100$, the lamppost is considered as OFF, and the confidence is 1 again. For the values $100 < \text{RGB}_{\text{median}} < 200$, the green channel is used as a second step for validating the status of the lamppost. It was shown that the images produced by nodes with increased vegetation around them (as in Fig. 3-right) are more prevalent on the green channel; thus, when the green channel is $\geq 200$, the streetlight is considered ON. We list these results with a confidence of 0.5. An example of various lampposts and images labelled during the day or the night time can be seen in Fig. 6.

**Ethics Statements**

Hereby, the authors consciously assure that for the manuscript “A Dataset of Images of Public Streetlights with Operational Monitoring using Computer Vision Techniques”, the following is fulfilled:
(1) This material is the authors’ own original work, which has not been previously published elsewhere.
(2) The paper is not currently being considered for publication elsewhere.
(3) The paper reflects the authors’ own research and analysis in a truthful and complete manner.
(4) The paper properly credits the meaningful contributions of co-authors and co-researchers.
(5) The results are appropriately placed in the context of prior and existing research.
(6) All sources used are properly disclosed (correct citation). Copying of text must be indicated by using quotation marks and proper references.
(7) All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

As our dataset did not involve any human subjects, animal experiments, or social media platform data, approval from an any IRB/local ethics committees was not required. As our camera images are facing the sky, no human subjects are present in the photos. Finally, as our dataset is based on street light images, no survey studies were conducted, and no work was conducted involving chemicals, procedures, or equipment that have any usual hazards inherent in their use, against aminal or human subjects.

I agree with the above statements and declare that this submission follows the policies of Solid State Ionics as outlined in the Guide for Authors and in the Ethical Statement.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Data Availability

Images of Public Streetlights with Operational Monitoring using Computer Vision Techniques (Original data) (Zenodo).

CRediT Author Statement

Ioannis Mavromatis: Conceptualization, Methodology, Software, Data curation, Writing – original draft; Aleksandar Stanoev: Conceptualization, Methodology, Software; Pietro Carnelli: Software, Writing – review & editing; Yichao Jin: Supervision; Mahesh Sooriyabandara: Funding acquisition; Aftab Khan: Software, Writing – review & editing.

Acknowledgments

This work is funded in part by Toshiba Europe Ltd. UMBRELLA project is funded in conjunction with South Gloucestershire Council by the West of England Local Enterprise Partnership through the Local Growth Fund, administered by the West of England Combined Authority.

References
[1] G.K. Wallace, The JPEG still picture compression standard, Commun. ACM 34 (4) (1991) 30–44, doi:10.1145/103085.103089.
[2] BRIL Toshiba Europe Ltd., “UMBRELLA platform and testbed,” 2022 https://www.umbrellaiot.com. Accessed January 31, 2022.
[3] RaspberryPi, “Raspberry Pi Camera Module ver.1 Datasheet,” 2015 https://www.arducam.com/downloads/modules/RaspberryPicamera/RaspberryPiCameraModuleDSV3.0.pdf. Accessed January 31, 2022.

[4] OmniVision, “Camera sensor OV5647 datasheet,” 2009 https://cdn.sparkfun.com/datasheets/Dev/RaspberryPi/ov5647full.pdf. Accessed January 31, 2022.

[5] I. Mavromatis, A. Stanoev, P. Carnelli, Y. Jin, M. Sooriyabandara, and A. Khan, “Images of public streetlights with operational monitoring using computer vision techniques,” Apr. 2022. [Online]. Available: doi:10.5281/zenodo.6046759.

[6] D. Anand, I. Mavromatis, P.E. Carnelli, A. Khan, A federated learning-enabled smart street light monitoring application: benefits and future challenges, in: Proceedings of the ACM MobiCom - Workshop on AI Empowered Mobile and Wireless Sensing (MORSE), Sydney, Australia, 2022.

[7] S.Y. Kadirova, D.I. Kajtsanov, A real time street lighting control system, in: Proceedings of the International Conference on Electrical Machines, Drives and Power Systems (ELMA), 2017, pp. 174–178, doi:10.1109/ELMA.2017.7955426.

[8] Y. Yao, B. Yu, C. Gong, T. Liu, Understanding how pretraining regularizes deep learning algorithms, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–13, doi:10.1109/TNNLS.2021.3131377.

[9] Z. Li, Q. Zeng, Y. Liu, J. Liu, L. Li, An improved traffic lights recognition algorithm for autonomous driving in complex scenarios, Int. J. Distrib. Sens. Netw. 17 (5) (2021).

[10] R. Pi, “Raspberry Pi compute module 3b+,” 2021, https://www.raspberrypi.org/products/compute-module-3-plus/. Accessed January 31, 2022.

[11] S. Kennedy, “Astral library - python calculations for the position of the sun and moon.” 2022 https://github.com/sffjunkie/astral. Accessed January 31, 2022.

[12] A.Z. Karen Simonyan, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the International Conference on Learning Representations 2015, 2015 May [Online]. Available: https://arxiv.org/abs/1409.1556.