Urban Big Data Centre

Data and expertise for improving cities

www.ubdc.ac.uk

UBDC Technical Note 2022/2 (September 2022)

Glasgow CCTV Object Detection Counts

Authors: Luís Serra (University of Glasgow)

Maralbek Zeinullin (University of Glasgow)

UBDC Technical Notes and Working Papers Series DOI:

10.5281/zenodo.7054623

Economic and Social Research Council

University of Glasgow

Urban Big Data Centre
University of Glasgow
7 Lilybank Gardens
Glasgow, G12 8RZ

T: +44 (0)141 330 2764
E: ubdc@glasgow.ac.uk
W: www.ubdc.ac.uk

The University of Glasgow is a registered Scottish charity: Registration Number SC004401

Glasgow CCTV Object Detection

Computer Vision

Luís Serra
and
Maralbek Zeinullin

Technical Summary

(version 1.0.0)

September 7, 2022

Preface

This summary was written at the Urban Big Data Centre (UBDC), University of Glasgow. Any questions regarding this summary should be directed to `luis.serra@glasgow.ac.uk` or `maralbek.zeinullin@glasgow.ac.uk`.

1 Introduction

This document aims to provide information about the data available in the Glasgow CCTV Object Detection Platform (henceforth referred to as "the platform"). The information accessible in this document can be considered a type of descriptive metadata. The description of the data is organized in *Frequently Asked Questions* format with the purpose to provide platform users with the necessary information to fully understand the data available from this platform.

2 What data is provided by the platform

The platform provides counts of cars, persons, bicycles, motorcycles, buses and trucks taken every 30 minutes in different locations of Glasgow. These locations are all public spaces, mainly street junctions (check section 6 for more details). A snapshot is taken by a CCTV camera operating in each location and the objects of interest visible on the image are detected and subsequently counted. Overall, 48 snapshots per day are taken for each camera.

At 0:00AM each day, all the half hourly counts from the previous day are added to a *csv* file and made available for download on the platform. The structure of this file is explained on table 1.

3 Attributes of cameras used to collect data

The cameras used to collect data are CCTV cameras operating in Glasgow with the technical specifications shown in appendix A.

Table 1: Description of the data available to download on the platform.

Field	Description	Example
id	row number	45
camera_id	camera unique identification	52
timestamp	date-time the image was captured. **T** stands for Time and **Z** stands for the Zero timezone, offset by 0 from the Coordinated Universal Time (UTC)	2019-11-30T04:30:01.428Z
latitude	latitude location of camera in WGS84 coordinate system (EPSG: 4326)	55.856741
longitude	longitude location of camera in WGS84 coordinate system (EPSG: 4326)	-4.268628
car	count of cars visible on captured image; small vans may also be considered in this category	2
person	count of persons visible on captured image	5
bicycle	count of bicycles visible on captured image	1
motorcycle	count of motorcycles visible on captured image	0
bus	count of buses visible on captured image	0
truck	count of lorries visible on captured image; larger vans may also be considered in this category	0
camera_ok	This field indicates whether an image was captured and corresponding counts displayed correctly. The field can display one of three labels: no label(or empty field), "yes" or "no". This verification step was added on 2022-01-20T12:00:02.591Z - records prior to it are empty. The label **yes** means the image was captured; the label **no** means the image was not captured and the counts displayed are not correct.	yes

4 How many cameras provide data to the platform?

Since the project started collecting data in November 6th 2019, some cameras were added to the system whereas others were discontinued. At time of

writing (August 3rd 2022) 37 cameras are in operation, in sharp contrast with the earliest prototypical project phase when just three cameras were active. A committee from Glasgow City Council discusses potential candidates to be added/removed from the system, depending on the interests of the Council.

5 How are images captured?

At a determined frequency, currently every 30 minutes, each camera automatically reverts to a preset 'PTZ'[1] position, chosen to offer an optimal view of the public space and detectable objects. A video of two seconds duration is captured encoded with H.264 compression[2]. The capture process reaches completion with the camera returning to its previous "operational" position. The two seconds video is then sourced to ffmpeg to extract a single frame in *jpeg* format, which will then be presented for analysis by the object detection algorithm.

The automated positioning and capture of imagery does not occur if the camera is being or has recently been manually operated. This feature is designed to minimise the disruption to core community safety and crime prevention monitoring activities. In this case, the field **camera_ok** in the table 1 will present the value **no**.

6 Location of cameras

The cameras for this project can be found in the Avenues' map[3] and are positioned at an average height of six metres. Note that the map may not display all cameras, particularly those added most recently. Conversely, several cameras that are currently not producing imagery and data for the platform may be included within the map. Figure 1 shows an example of the information available for a camera within the CCTV network.

[1] PTZ stands for **P**an, **T**ilt and **Z**oom. Pan refers to movement to the left and right, Tilt refers to the up and down movement and Zoom allows the camera to get closer or away from the object being filmed.

[2] H.264 is a video coding format for recording and distributing full HD video and audio.

[3] Produced by Glasgow City Council and published here with their permission.

Figure 1: Information blueprint for cameras in the Avenues' map. (1) Exemplar of screenshot captured by camera A82; (2) camera id; (3) description with the location of the camera; (4) location of the camera in the map; (5) compass direction of shot (green line). In this example, it is visible that camera A82 shots roughly to the East South East direction. (6) direction button which allows the retrieval of the camera's coordinates (see figure 2).

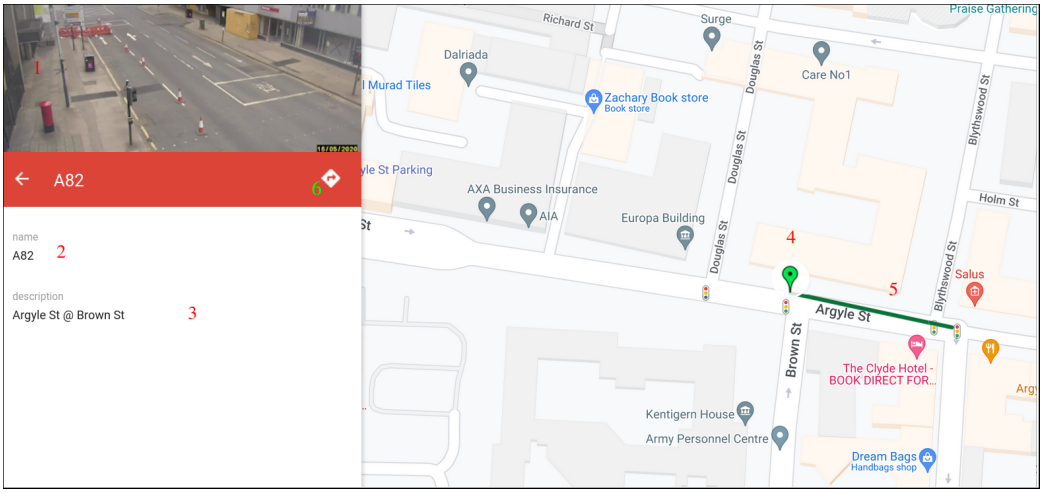


Figure 2: WGS84 latitude and longitude coordinates (red box) of camera A82, available in the address bar.

7 Image attributes

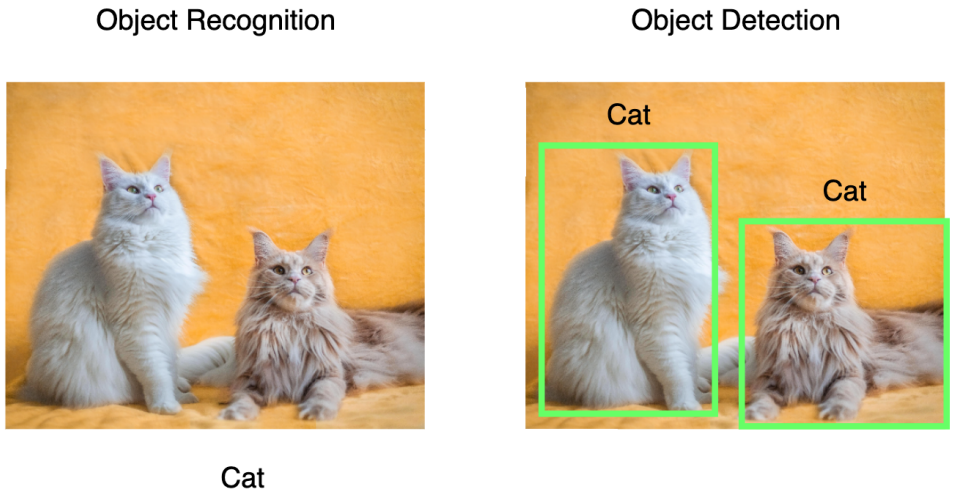
The images captured by the CCTV cameras are encoded in jpeg format, with RGB colour mode and a resolution of 1280 x 720 pixels.

8 How are objects detected on images?

Computer Vision (CV) models are algorithms developed to automate tasks typically associated with the human visual system. While the foundation of the field was laid back in 1960s, modern computing capacity has prompted the development of many advanced CV algorithms. **Object Detection** is one of the techniques which received widespread application and is being broadly used to assist in the detection of various objects present on imagery from public spaces. Object detection is often confused with object recogni-

tion. In fact, the former is a combination of object recognition and object localisation (Figure 3).

Figure 3: Recognition vs Detection



8.1 What is an Object Detection Model?

The task of the object detection model is to take an image as an input and output one or more bounding boxes with a class label. R-CNN model family is one of the most popular model architectures used for object detection. The R-CNN acronym stands for "Regions with Convolutional Neural Networks features". The model architecture consists of three modules. First an input image undergoes through the region proposal process. Up to several thousands of candidate regions may be produced during this step. A convolutional neural network (CNN) is then used to extract features from those regions. Finally, each bounding box is assigned a class during the object classification stage. Figure 4 shows the process of detecting an object in a single image.

8.2 Model used on this project

At time of writing (August 2022) data available from the Glasgow CCTV Object Detection Platform is created using the *faster_rcnn_inception_resnet_v2_atrous_coco*

Figure 4: Object detection process

Object detection process. The region proposal stage can generate up to 2,000 region proposals (simplified here for better visualisation). Note: Figure adapted from https://machinelearningmastery.com/object-recognition-with-deep-learning/.

model, developed in 28/01/2018. Details of the model and download links can be found at its official GitHub page.

Like any programming task, the development of object detection models requires a high-level programming interface to begin with. Deep Learning frameworks provide a convenient way to train and deploy the models. Since our project was initiated in 2019, TensorFlow 1 framework was utilised to facilitate object detection model deployment. Afterwards, Tensorflow 2 was developed and currently, TensorFlow 1 version is no longer supported by the developer. Nevertheless, TF1 is still functioning and the main advantage of this frameworks is its easy deployment on server.

8.2.1 Objects detected by the model

This model was trained using Microsoft's COCO (*Common Objects in Context*) dataset and is capable of detecting 80 various object classes. This includes persons, bicycles, cars and other common objects. The official COCO Dataset webpage presents details of all the object types that can be detected.

8.2.2 Accuracy of the model

According to the official Github page mentioned above, this model architecture demonstrated 37 mAP (Mean Average Precision) during the COCO dataset evaluation. The COCO evaluator is a gold standard for the object

detector. mAP is a metric used to calculate the detection accuracy of the model taking in consideration different IoU[4] thresholds. By using different IoU thresholds, models are rewarded if they not only perform better at classifying objects but also if they perform better at localising objects on the image. mAP is a percentage. The higher the value the more accurate is the model. More information regarding this metric can be found here.

9 Data available from the platform

As mentioned in 8.2.1, the object detection model used in this project is able to recognise and detect 80 different classes of objects. After the object detection model processes the image revealing all possible objects from its catalogue, the backbone script that runs the model then filters persons, cars, motorcycles, bicycles, buses and trucks. The process ends by dumping the counts of the filtered objects into the csv file available from the platform and whose structure was already described in table 1. Although the geometry of bounding boxes are captured, these details are not currently accessible via the platform. One particular notable shortcoming with the current detection is that individuals within large gatherings of people are not easily detected - in such cases the model is likely to report a single person

10 Terms of use

[4]**I**ntersection **O**ver **U**nion. This metric measures the overlap between the predicted bounding box and the ground truth bounding box.

A Technical specifications of CCTV cameras

PRODUCT SPECIFICATION	30x HD ULTRA		40x HD ULTRA
Colour / Mono	Yes		Yes
Imager (CMOS)	1/1.9" Sony Exmor		1/1.9" Sony Exmor
Zoom - Optical + Digial	30X, 12X		40X, 12X
Zoom - Total	360X		480X
Pixels (Effective)	Approx 2.38 Megapixels		Approx 2.38 Megapixels
Horizontal view angle	61.2° (W) -2.32° (T)		60° (W) -1.8° (T)
Lens Size	6-180mm		6-240mm
Min Sensitivity - colour	0.002 Lux		0.002 Lux
Min Sensitivity - mono Int	0.0002 Lux		0.0002 Lux
Signal Noise Ratio	>50dB		>50dB
Picture Flip	Yes		Yes
Picture Freeze	Yes		Yes
PRODUCT FEATURES	ALL PREDATOR units have the following functions		
Focus/Iris	Auto/Manual		
Presets	360		
Tours	4 (max 90 presets per tour)		
Learned Patrols	4 mimic tours - upto 30 minutes duration each		
Privacey Zones	Upto 24		
Variable pan speed/coverage	0.1 - 120°/sec, 360° continuouse rotation, absolute positioning		
varible tile speed / coverage	0.1 - 120°/sec, 180°, absolute positioning		
Auto Homing	Goes to preset, tour or mimic tour after prescribed time		
Col/Mono Changeover	4 levels -3 fixed, 1 custom/user defined		
Operating temperature	-50°C to +60°C with optional heater		
Power	24VAC or 30VDC (AC: 22 - 28 / DC: 24 – 36) Option for 12VDC		
Power Consumption	160 watts maximum load (full operation with opional heater and long range lamps). Normal idle condition: 25 watts		
Low Power Mode	18 watts		
Mounting	101.6mm (4") PCD. Options for Wall, Corner, Side Pole and Soffit		
Certification	CE & FCC approval: Ingress IP68: Impact IK10: Vibration & Shock certified approvals		
SIRIUS ILLUMINATION OPTIONS	IR & WHITE LIGHT (Normal)		STANDARD PAINT OPTIONS / RAL
SIR110	IR 110 metres		Black 9005
SIR160	IR 160 metres		White 9003
SIR140WL	IR 140 metres, WL 140 metres		Cool Grey 7035
250SWL	WL 250 metres		Mid grey 7001
SIR250WL	IR 250 metres, WL 250 metres		WEIGHTS
SIR275	IR 275 metres		No LED's 9.5kg
SIR400	IR 400 metres		Single LED array 10Kg
SIR550S Spot	IR 550 metres		Dual LED array 10.5kg
ANALOGUE MODE			All Ultra Hybrids are fitted with the Hinged Mounting Adapter
Analogue Video	PAL 1V p-p 75Ω		
Telemetry	RS485		
Protocols	360, Forward Vision, Pelco P/D		
DIGITAL MODE - CODEC	25fps	ASPECT	30fps
HD1080P	1920X1080	16:9	1920X1080
HD720P	1280X720	16:9	1280X720
D1	720X576	4:3 (16:9 option)	720X480
VGA	640X480	4:3	640X480
FWQVGA	432X240	16:9	432X240
Distribution	Multi Codec, 2 stream profiles		
Image Compression	H.264, MPEG4, MJPEG		
Protocols	ONVIF(S),TCP/IP,HTTP,HTTPS,RTP, RTSP,360SDK,QoS,802.1X,NTP,FTP		
Compression Rate Control	CBR, VBR, CVBR		

Hinged Mounting Adapter (HMA) is designed to assist in the installation of the Predator allowing access to the cable connections 'hands free'.

Installers can now install the Predator using standard cables, so no need for pre-made cables. Enabling the Predator to be positioned up to 1 00 metres from the power supply (dependent on Predator/cable specification).

360VISION TECHNOLOGY

360 Vision Technology Ltd
Unit 7 Seymour Court Manor Park Runcorn
Cheshire WA7 1 SY UK
Tel: +44(0)1928 570000
info@360visiontechnology.com
www.360visiontechnology.com

DESIGNED AND MANUFACTURED IN THE UK