771764

MSc Research Project

Report

Identification and Classification in Urban Canyons in Leeds City Centre

Abstract

The project aims to provide an evidence basis for the introduction of green infrastructure in urban heat island (UHI) areas as vegetation can help to reduce the UHI intensity. Urban Heat Islands (UHI) occur where the temperatures in the dense urban areas of a city are higher than the surrounding countryside (Oke, 1982). This project aims to create an algorithm capable of automatically finding and characterising urban canyons based on LiDAR images, allowing those with potentially high heat intensity to be targeted with interventions. This is because one of the locations that have the highest air pollution and UHI are urban canyons. An urban canyon is a place where the street is flanked by buildings on both sides creating a canyon-like environment. (Nicholson, 1975). The initial study area is Leeds city centre.

Road and LiDAR data was processed in QGIS and Python. 2,237 urban canyons were identified in Leeds city centre. These were further split down into 4 subcategories high on both sides (HH), lower both sides (LL), high on one side (H) and lower on one side (L). The LiDAR image was separated into 900 images which were manually classified into canyon or not canyon. They were then used to train a CNN. Additional images of Manchester city centre were also created to add to the training sample to improve model accuracy. The final accuracy was 81.4%. A CNN to identify canyon types was not successful due to a lack of data.

Remote sensed temperature data and modelled solar energy was used to further split the Leeds canyons to identify hot and sunny canyons as these would most likely benefit from green infrastructure. In total,1,277 canyons were identified as having some hot and sunny aspects. Of the HH canyons, 335 have some hot and sunny aspects.

The research shows the urban canyons can be identified from LiDAR images and subcategorised further by using remotely sensed temperature and solar energy data. The analysis can provide a basis for further research on UHI in Leeds City Centre.

Introduction

The University of Hull and Leeds Beckett University are collaborating on the GIAUrban (Green Infrastructure Assets in Urban Heat Islands - UHIs) research project. The project aims to provide an evidence basis for the introduction of green infrastructure in urban heat island areas as vegetation can help to reduce the UHI intensity and airborne pollutants.

Urban Heat Islands (UHI) occur where the temperatures in the dense urban areas of a city are higher than the surrounding countryside (Oke, 1982). They are issues because increased air temperatures have an impact on people's health (Buchin et al., 2016) and can lead to increased energy consumption (Santamouris et al, 2015). With an increasing proportion of the population living in cities (UN,2014), combined with global warming, there is expected to be an increase in UHIs. The use of vegetation, or Green Infrastructure Assets (GIAs), in cities offers a potential solution to mitigate against higher temperatures as the vegetation provides shade and also as water evaporation not only cools the plants but the air around them(Alexandri et al., 2008).

One of the aims of the project is to monitor the impact of GIAs on Urban Heat Islands at a local level. UHI effects are site-specific, therefore an assessment of potential mitigation strategies must be carried out at a local level (GIAUrban, n.d.). Currently, the focus of the research is on Leeds City Centre. As part of the project, several EarthSense Sensors have been put around Leeds (Parker, 2020). These record an hourly measurement of temperature, humidity and air pollution.

One of the places that have the highest air pollution and UHI are urban canyons. An urban canyon (also known as a street canyon) is a place where the street is flanked by buildings on both sides creating a canyon-like environment. (Nicholson, 1975). Air quality describes how polluted the air is. There are several measures of air quality. The five 5 most damaging pollutants are Particulate matter (PM), Ammonia (NH3), Nitrogen Oxide (NOx), Sulphur dioxide (SO2), Non-methane volatile organic compounds (NMVOCS) (Department for Environment, Food & Rural Affairs, 2018). Health issues that can arise from exposure to these pollutants are cardiovascular and respiratory diseases, which can lead to reduced life expectancy (Public Health England, 2018).

This project aims to create an algorithm capable of automatically finding and characterising urban canyons based on remotely sensed imagery (e.g. LiDAR), allowing those with potentially high air pollution to be targeted with interventions.

Literature review

One of the key measures of an urban canyon is the aspect ratio which is calculated as the height (H) divided by the width (W). A regular canyon has a measurement of about 1 and no large opening along its length. An avenue canyon has an aspect ratio of less than 0.5. A deep canyon has a value of about 2. The length (L) of a canyon is calculated as the distance between two major junctions. Short canyons have a Length/Height ratio of about 3, medium canyons have a ratio of about 5 and long canyons have a ratio of about 7.
A canyon can further be divided into symmetric or asymmetric. Symmetric canyons are where the building are an even size on each side. Asymmetric canyons are where one side is higher than the other (Vardoulakis et al., 2003).

Another measure of urban canyons is sky view factor (SVF). This is a measure between 1 and 0 of the amount of sky that is visible, with 0 being no sky and 1 being open skies (Oke, 1988). There are several techniques used that can be used to calculate SVF, these include geometric methods, fish-eye photographic methods, Global Positioning System method, simulation methods, big data approaches using street view images and from Light detection and ranging (LiDAR) data (Miao et al., 2020). LiDAR data has been used to calculate SVF from the shadow cast by buildings (Hodul, Knudby and Ho, 2016).

Traditionally, canyons have been measured by surveying the building and infrastructure in an area. This is time-consuming and expensive. There is now a wealth of data such as Google Street View (GSV) images (Hu et al., 2020), LiDAR data (Bonczak and Kontokosta, 2019), and ONS road map data which can be used to map an area and calculate the height and width of an urban canyon without physical measurement. Using these data would allow for a more widespread categorisation of urban canyons than ever before.

Methodology

Leeds City Centre Canyons Identification

Data

Two public available data were used to identify canyons. OS Open Roads data (www.ordnancesurvey.co.uk, n.d.) and UK LiDAR data (environment.data.gov.uk, n.d.) last return LiDAR data producing a "bald earth" Digital Elevation Model (DEM) at 1-meter resolution.

Area

The initial study area was Leeds city centre. Leeds is a large city in Northern England. The area contains a mixture of buildings including office blocks, shopping centres, housing and cultural buildings. The area extends from just above the inner ring road to the river just below the railway station (Figure 1)

Figure 1 - LiDAR Leeds City Centre

Method

The road and LiDAR data was imported in QGIS, an open-source GIS software (QGIS.org, 2021). The roads were divided into 10m sections. 20m perpendicular lines were added to the 10m road segments. The perpendicular lines were sampled for height at every 1m. This data and data about 10m road segments was output into .xlsx format and imported into Python for further processing and analysis. (A. Hardy, personal communication, 23/4/2021)

The 1m sample data was transposed to create one row for every canyon cross-section. Roads with no data or mostly missing data were removed. A range of measures was calculated.

$$\text{max\_height010} = \text{maximum}(\text{SAMPLE\_1\_0} - \text{SAMPLE\_1\_10})$$
$$\text{max\_height1120} = \text{maximum}(\text{SAMPLE\_1\_11} - \text{SAMPLE\_1\_20})$$
$$\text{min\_height317} = \text{minimum}(\text{SAMPLE\_1\_3} - \text{SAMPLE\_1\_17})$$
$$\text{max\_height07} = \text{maximum}(\text{SAMPLE\_1\_0} - \text{SAMPLE\_1\_7})$$
$$\text{max\_height1320} = \text{maximum}(\text{SAMPLE\_1\_13} - \text{SAMPLE\_1\_20})$$

$$\text{Height} = (\text{max\_height010} + \text{max\_height1120})/2 - \text{min\_height317}$$

$$\text{Aspect Ratio (AR)} = \text{Height} / 20$$

Cross-sections with an aspect ratio of less than 0.5 were removed. This leads to 2,237 canyons.

The canyons were classified into 4 groups using K means clustering carried out on Height and DiffHeight, with DiffHeight calculated as:

$$\text{DiffHeight} = \text{abs}(\text{max\_height07} - \text{max\_height1320})$$

This gave a split at 15 meters high and 10 meters height difference. This leads to 4 groups HH, LL, H and L, which are:

- HH High both sides - Height greater than equal 15 metres and DiffHeight less than equal 10 metres
- LL Low both sides - Height less than 15 metres and DiffHeight less than equal 10 metres
- H High one side - Height greater than equal 15 and DiffHeight greater than 10 metres
- L Low one side - Height less than 15 metres and DiffHeight greater than 10 metres

Results

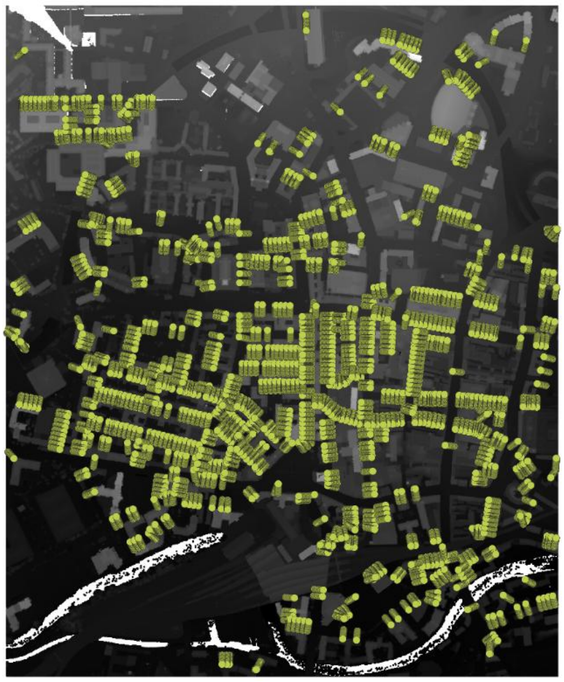

Figure 2a Canyons Leeds City Centre

Figure 2b Canyons Leeds City Centre by Type

Legend
– LL - L
– H - HH

Convolutional Neural Network (CNN) To Predict Urban Canyons from LiDAR images of Leeds

Convolutional Neural networks (CNN) are a type of neural network that are commonly used to recognised features with images. Their structure is based on the relationship between neurons in the human brain. CNN's contain many neurons which have learnable weights and biases. Inputs are received by the neuron a weighted sum is calculated which passed through an activation function to produce an output which is the input to another neuron (Ong, 2020). Figure 3 shows a typical structure of a CNN.

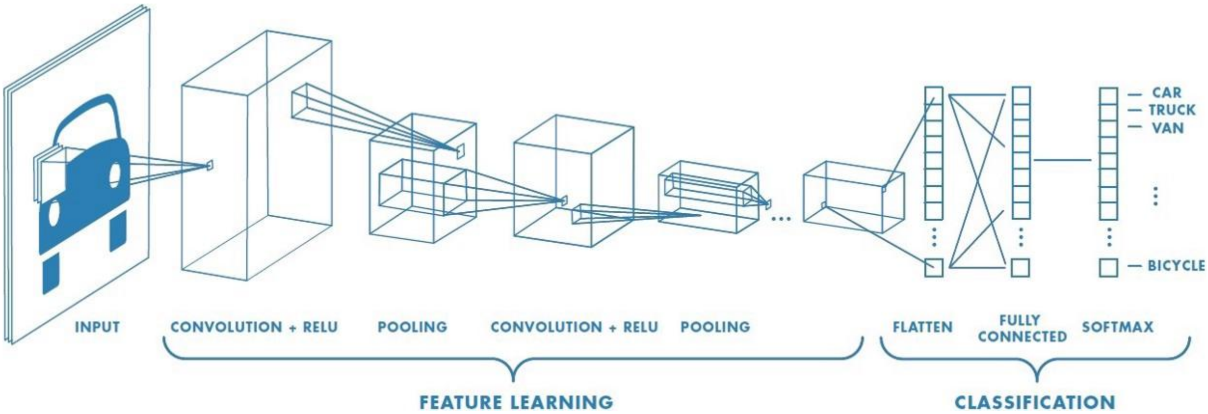

Figure 3 Structure of a CNN (Ong, 2020)

Data

The LiDAR image (Figure 1) and canyons by type image (Figure 2b) were output as PNG files from QGIS. Both images were made to be the same size and cover the same area of Leeds. Both were split into 30 x 30 parts, giving a total of 900 images using an online image processing tool (www.imgonline.com.ua, n.d.). The LiDAR images were manually separated into the canyon or no canyon based on if the corresponding canyons image contains any dots. An 80/20 train-validation split was used.

To increase the volume of images the base images were rotated left, right and 180 degrees so to get better model accuracy (Table 1a).

Model Structure

The model was built using Keras (Chollet, F. & others, 2015) and TensorFlow (Abadi et al., n.d.)

The model consisted of the following layers

Input layer
max_pooling2d layer
Conv2D
max_pooling2d layer
Conv2D
max_pooling2d layer
Conv2D
max_pooling2d layer
Dense layer
Dropout layer
Final output layer

Results

	Train -Base	Test - Base	Train - Rotated	Test – Rotated
Canyon	332	83	1336	334
No Canyon	388	97	1544	386

Table 1a – Canyon Volume

	Accuracy 20 epochs	Loss 20 epochs	Accuracy 40 epochs	Loss 40 epochs
Train	0.8486	494.60	0.9806	95.16
Validation	0.6778	1.12	0.6750	2.77

Table 1b – CNN Leeds Results

The increasing loss on the validation sample indicates that there is an overfitting of the model on the development sample. To reduce this additional images were sourced from Manchester city centre.

Manchester data

The Manchester city centre data was processed in the same way as the Leeds data. Figure 3 shows the canyon by type for Manchester.

Figure 4 Manchester Canyons by type

Legend

- LL - L
- H - HH

To increase the volume of images, the base images were rotated left, right and 180 degrees to try and obtain better model accuracy. Also, additional of non-canyons images from just outside Leeds City Centre were used to try and get a more balanced sample.

	Train Leeds	Train Manchester	Train Outside Leeds CC	**Train Total**	Test Leeds	Test Manchester	Test Outside Leeds CC	**Test Total**
Canyon	2272	1336		**3608**	568	334		**902**
No Canyon	512	1544	2156	**4212**	128	386	513	**1027**

Table 2 – Leeds and Manchester Volumes

Convolutional Neural Network (CNN) To Predict Urban Canyons from LiDAR images of Leeds and Manchester

The same CNN structure as for Leeds city centre CNN was used. The final results are shown below.

	Accuracy 20 epochs	Loss 20 epochs	Accuracy 40 epochs	Loss 40 epochs
Train	0.8510	1334.95	0.9533	481.59
Validation	0.7853	0.5650	0.8136	1.0596

Figure 5a – CNN Leeds and Manchester Results

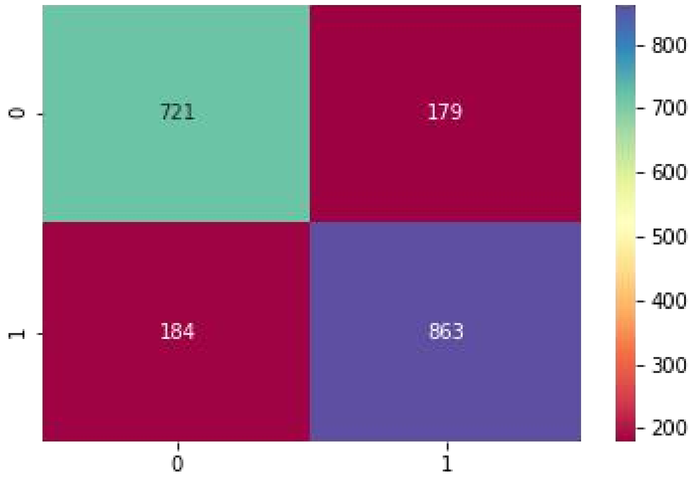

Figure 5b – Confusion Matrix

Sensitivity 0.82 Specificity 0.80

Other models

To increase the accuracy of the model on the validation sample and reduce overfitting, two other modelling techniques were tested. These were regularization and augmentation. Augmentation is the process of increasing the number of images by applying random transformations to alter the image. (Shorten and Khoshgoftaar, 2019). Regularization is the process of making the fitted function smoother by penalising the weights of a function(Team, n.d.).

For the augmented model, the combined Manchester and Leeds images were each augmented 5 times to give 18,272 canyons and 21,078 no canyon images. The model produced a validation accuracy of 61.6% on the same validation sample as the original Leeds and Manchester CNN was validated on. There are limits on how much augmentation can be used to improve the overfitting of a model as there were only 902 canyon and 1053 non-canyon base images. Some of the images only had the canyons in one corner of the image and the augmentation process could have caused this to become too distorted.

For the regularization model, a regularization function was added to two of the layers. The rest of the model remained the same. The accuracy on the validation images was 80.1%. Neither CNN produced a better accuracy than the base Leeds/Manchester CNN on the validation images.

Modelling Canyons types

A further CNN was created to see if was possible to predict canyon type from LiDAR data alone. The same model structure as for Leeds/Manchester CNN was used.

	Train	Test
H	614	154
HH	1366	326
L	531	133
LL	1155	289

Table 3a Canyons Types Volume

	Accuracy 20 epochs	Loss 20 epochs	Accuracy 40 epochs	Loss 40 epochs
Train	0.7909	436.80	0.9609	92.028
Validation	0.4611	2.81	0.4389	6.73

Table 3b CNN Canyon Types Results

Identifying Hot and Sunny Canyons

The final stage was to use remotely sensed temperature and solar data to hot and sunny canyons. These are the canyons that would most likely benefit from green infrastructure.

Data

Direct solar beam for July and August generated using the same LiDAR image as in Figure 1, (in mW m^2 per sr).

ASTER satellite kinetic temperature image in Kelvin. It has been projected from a 20 m^2 pixel resolution to the same 1m spatial extent and spatial resolution as LiDAR and solar data.

Method

The canyons were classified into 4 groups using K means clustering carried out on the Solar and Temperature variables.

This suggested a split at 3024 Kelvin and 2741 mW m^2 per sr. This gives 4 groups Hot and Sunny, Hot and Not Sunny, Cool and Sunny and Cool and Not Sunny, as described below:

- Hot and Sunny - Temperature greater than equal 3024 K and solar energy greater than equal 2741 m^2
- Hot and Not Sunny - Temperature greater than 3024 K and solar energy less than equal 2741 m^2
- Cool and Sunny - Temperature less than 3024 K and solar energy greater than 2741 m^2
- Cool and Not Sunny - Temperature less than 3024 K and solar energy less than 2741 m^2

Figure 6a Canyons Solar/Temperature Leeds City Centre

Figure 6b Canyons Solar/Temperature Leeds City Centre HH

Legend

- Cool and Not Sunny
- Cool and Sunny
- Hot and Sunny
- Hot and Not Sunny

Discussion

The location of the canyons in Leeds city centre (Figure 2b) does fit with local knowledge of the area. The business district of Leeds centred around Park Row does show a concentration of canyons. The visualisation of the canyons on the map of Leeds aids in identifying areas with a high concentration of canyons that would benefit from green intervention.

The base CNN based on Leeds images does achieve high accuracy on the training sample (Table 1b). This shows that it is possible to create a model that does predict urban canyons but the increasing loss on the validation sample and lower accuracy do indicate overfitting. The model could likely be improved by the inclusion of more images from another city to create a large and more robust training sample.

To test if images from another city would improve the robustness of the model, LiDAR and road data from Manchester city centre, another large city in northern England, were processed in QGIS and Python. The inclusion of more images increased the accuracy on the validation sample from 67.5% to 81.3% (Figure 5a). This shows that a larger number of images has helped the CNN learn the feature of canyons more accurately and reduce overfitting. Validation on another city would also help to prove the CNN is robust.

(Hu et al., 2020) achieved an accuracy of 89.3% identifying canyons from Google Street View images. To further increase accuracy, there may be benefits to splitting the data into 3 groups canyons, non-canyons and indeterminate. The indeterminate class is where the canyon is less than a quarter of the image. Also, more images could be created by splitting the images differently. The Leeds LiDAR image could be split 20 x 20 as well as 30 x 30.

A further CNN was developed to try and predict if a canyon was one of the 4 canyon types - HH, H, LL, L. This model achieved an accuracy of 43.9% on the validation sample (Table 3b) To successfully model canyon types, substantially more data would be required. This would involve creating images from many more UK city centres. One issue with model canyon types is that more than one canyon type can be present in an image and it is a judgement as to which is the most dominant type. The fact that images can contain multiple canyon types does make it more difficult for a CNN to learn the features of each canyon type. Also, changes to the structure of the CNN such as additional layers could aid with the identification of canyon types.

The clustering of the canyon using solar beam and temperature data does show clusters of cooler canyons around Park Lane. There are two clusters of hot canyons on either side. However, the low emissivity roofing material that the building is constructed from could have an impact on how the temperatures are sensed. New building material could radiate less heat than older brick and slate buildings. In total,1,277 canyons were identified as having some hot and sunny aspects. Of the HH canyons, 335 have some hot and sunny aspects.

This work has not included the length of the canyon in the calculation, this is because it is difficult to include length in geospatial analysis. The length of the canyon could be included in future works. This data obtained in this analysis could be variables in machine learning models that are used to predict air quality in Leeds city centre (Šimić et al., 2020). Additional analysis and clustering could be carried out if EarthSense Sensors (Parker, 2020) are placed in Leeds city centre. Air pollution data could be used to test if there was a relationship between canyon type and air pollution.

Conclusion

OS Road and remote sensed LiDAR data was used to identify 2,237 urban canyons in Leeds city centre.

It is possible to build a model that can predict urban canyons from LiDAR images. The CNN based on images from Leeds and Manchester produced an accuracy of 81.4% on the validation sample. Augmenting the existing images further did not produce increased accuracy on the validation sample due to the small number of base images leading to over augmentation. Further improvements to the accuracy could be made by expanding the number of base images by adding more data from other UK cities.

A CNN to predict canyon types was created but achieved an accuracy of 43.9% on the validation sample. Substantially more data would be required from many city centres to create a more robust model.

Remote sensed solar and temperature data was used to identify canyons that would get the greatest benefits from green infrastructure. In total,1,277 canyons were identified as having some hot and sunny aspects.

This work could feed into further modelling and analysis of air quality and UHI in Leeds city centre.

References

R. Oke, The energetic basis of the urban heat island, Q. J. R. Meteorol. Soc. 108 (1982) 1–24.

S. E Nicholson (1975), A pollution model for street-level air. Atmospheric Environment, 9 (1975), pp. 19-31

Buchin, M.T. Hoelscher, F. Meier, T. Nehls, F. Ziegler, Evaluation of the health-risk reduction potential of countermeasures to urban heat islands, Energy Build. 114 (2016) 27–37

M. Santamouris, C. Cartalis, A. Synnefa, D. Kolokotsa, On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings - A review, Energy Build. 98 (2015) 119–124.

UN, United Nations, World Urbanization Prospects: The 2014 Revision,(2014),http://dx.doi.org/10.4054/DemRes.2005.12.9.

E. Alexandri, P. Jones. Temperature decreases in an urban canyon due to green walls and green roofs in diverse climates. Build. Environ., 43 (2008), pp. 480-493

GIAUrban. (n.d.). Home. [online] Available at: http://www.giaurban.uk/ [Accessed 26 Mar. 2021].
Parker, J. (2020). The Leeds urban heat island and its implications for energy use and thermal comfort. Energy and Buildings, p.110636.

Department for Environment, Food & Rural Affairs (2018). Air quality: explaining air pollution – at a glance. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/air-quality-explaining-air-pollution/air-quality-explaining-air-pollution-at-a-glance.
Public Health England (2018). Health matters: Air pollution. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/health-matters-air-pollution/health-matters-air-pollution.

Vardoulakis, Sotiris; Bernard E.A. Fisher; Koulis Pericleous; Norbert Gonzalez-Flesca (2003). "Modelling air quality in street canyons: a review" (PDF). Atmospheric Environment. 37 (2): 155–182. Bibcode:2003AtmEn..37..155V. doi:10.1016/s1352-2310(02)00857-9.

Oke, T.R. Street design and urban canopy layer climate. Energy Build. 1988, 11, 103–113.
Miao, C., Yu, S., Hu, Y., Zhang, H., He, X. and Chen, W. (2020). Review of methods used to estimate the sky view factor in urban street canyons. Building and Environment, 168, p.106497.

Hodul, M., Knudby, A. and Ho, H. (2016). Estimation of Continuous Urban Sky View Factor from Landsat Data Using Shadow Detection. Remote Sensing, 8(7), p.568.

Chuan-Bo Hu, Fan Zhang, Fang-Ying Gong, Carlo Ratti, Xin Li, Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning, Building and Environment, Volume 167,2020,

Bonczak, B. and Kontokosta, C.E. (2019). Large-scale parameterization of 3D building morphology in complex urban landscapes using aerial LiDAR and city administrative data. Computers, Environment and Urban Systems, 73, pp.126–142.

www.ordnancesurvey.co.uk. (n.d.). OS Open Roads | High Level View of Roads | Vector Map Data | Free Download. [online] Available at: https://www.ordnancesurvey.co.uk/business-government/products/open-map-roads.

environment.data.gov.uk. (n.d.). ArcGIS Web Application. [online] Available at: https://environment.data.gov.uk/DefraDataDownload/?Mode=survey.
QGIS.org, 2021. QGIS Geographic Information System. QGIS Association. http://www.qgis.org

www.imgonline.com.ua. (n.d.). Cut photo into equal parts online - IMG online. [online] Available at: https://www.imgonline.com.ua/eng/cut-photo-into-pieces.php [Accessed 9 Jul. 2021].

Ong, W. (2020). What is CNN ? A 5 year old guide to Convolutional Neural Network. [online] Medium. Available at: https://medium.com/analytics-vidhya/what-is-cnn-a-5-year-old-guide-to-convolutional-neural-network-c9d0338c3bf7 [Accessed 17 Aug. 2021].

Chollet, F. & others, 2015. Keras. Available at: https://github.com/fchollet/keras.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D., Steiner, B., Tucker, P., Vasudevan, V., Warden, P. and Wicke, M. (n.d.). TensorFlow: A System for Large-Scale Machine Learning This paper is included in the Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). TensorFlow: A system for large-scale machine learning. [online] . Available at: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.

Shorten, C. and Khoshgoftaar, T.M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1).

Team, K. (n.d.). Keras documentation: Layer weight regularizers. [online] keras.io. Available at: https://keras.io/api/layers/regularizers/.

Šimić, I., Lovrić, M., Godec, R., Kröll, M. and Bešlić, I. (2020). Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon. Environmental Pollution, 263, p.114587.