URBAN ANALYTICS ON GREEN COVERAGE IN MALAYSIA WITH STREET VIEW IMAGES

CHONG XIAN JUN

UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA

DECLARATION OF THESIS / UNDERGRADUATE PROJECT REPORT AND COPYRIGHT

Author's full name : CHONG XIAN JUN

Date of Birth : 14/07/1998

Title : URBAN ANALYTICS ON GREEN COVERAGE IN MALAYSIA WITH STREET VIEW IMAGES

Academic Session : 2022/2023

I declare that this thesis is classified as:

☐	**CONFIDENTIAL**	(Contains confidential information under the Official Secret Act 1972)*
☐	**RESTRICTED**	(Contains restricted information as specified by the organization where research was done)*
✓	**OPEN ACCESS**	I agree that my thesis to be published as online open access (full text)

1. I acknowledged that Universiti Teknologi Malaysia reserves the right as follows:

2. The thesis is the property of Universiti Teknologi Malaysia

3. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.

4. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

SIGNATURE OF STUDENT	SIGNATURE OF SUPERVISOR
MCS211047	TS. DR. CHAN WENG HOWE

Date: 12 JANUARY 2023   Date: 12 JANUARY 2023

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

"I hereby declare that I have read this proposal and in my opinion this proposal is sufficient in term of scope and quality for the award of the degree of Master of Science (Data Science)."

Signature   : _____

Name of Supervisor I   : TS. DR. CHAN WENG HOWE

Date   : 12 JANUARY 2023

BAHAGIAN A - Pengesahan Kerjasama*

Adalah disahkan bahawa projek penyelidikan tesis ini telah dilaksanakan melalui

kerjasama antara _____dengan _____

Disahkan oleh:

Tandatangan :   Tarikh :

Nama :

Jawatan :

(Cop rasmi)

Jika penyediaan tesis atau projek melibatkan kerjasama.

═══

BAHAGIAN B - Untuk Kegunaan Pejabat Sekolah Pengajian Siswazah

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar   **:**

Nama dan Alamat Pemeriksa Dalam   **:**

Nama Penyelia Lain (jika ada)   **:**

Disahkan oleh Timbalan Pendaftar di SPS:

Tandatangan   :   Tarikh:

Nama   :

URBAN ANALYTICS ON GREEN COVERAGE IN MALAYSIA WITH STREET
VIEW IMAGES

CHONG XIAN JUN

A proposal submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Science (Data Science)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2023

DECLARATION

I declare that this proposal entitled *"URBAN ANALYTICS ON GREEN COVERAGE IN MALAYSIA WITH STREET VIEW IMAGES "* is the result of my own research except as cited in the references. The proposal has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature   : ....

Name   : CHONG XIAN JUN

Date   : 12 JANUARY 2023

DEDICATION

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

ABSTRACT

The importance of urban green spaces has been widely recognized as a critical component for sustainable development and the well-being of urban dwellers. However, in Malaysia, there is a lack of research on urban green spaces especially from the perspective of human users. This project addresses this gap by studying the green view index (GVI), a metric that measures the amount of green space visible from a particular location in a representative sample area Malaysia, Johor Bahru city canter. It provides valuable insights into the distribution and quality of green space within urban areas. To compute the GVI, this research employs machine learning models to automatically predict the GVI from street view images collected from Google Street View. The results are plotted on a map and made into a dashboard to allow stakeholders to easily understand the state of urban green coverage in the city and uncover patterns in the distribution of urban green spaces. This research contributes to the understanding of the importance of urban green spaces in Malaysia, thereby supporting urban planning and decision-making.

ABSTRAK

Kepentingan ruang hijau bandar telah diiktiraf secara meluas sebagai komponen kritikal untuk pembangunan mampan dan kesejahteraan penduduk bandar. Walau bagaimanapun, di Malaysia, terdapat kekurangan penyelidikan mengenai ruang hijau bandar terutamanya dari perspektif pengguna manusia. Projek ini menangani jurang ini dengan mengkaji indeks paparan hijau (GVI), metrik yang mengukur jumlah ruang hijau yang boleh dilihat dari lokasi tertentu di kawasan sampel perwakilan Malaysia, canter bandar Johor Bahru. Ia memberikan pandangan berharga tentang pengedaran dan kualiti ruang hijau di dalam kawasan bandar. Untuk mengira GVI, penyelidikan ini menggunakan model pembelajaran mesin untuk meramalkan GVI secara automatik daripada imej paparan jalan yang dikumpulkan daripada Google Street View. Hasilnya diplot pada peta dan dijadikan papan pemuka untuk membolehkan pihak berkepentingan memahami dengan mudah keadaan liputan hijau bandar di bandar dan mendedahkan corak dalam pengagihan ruang hijau bandar. Penyelidikan ini menyumbang kepada pemahaman tentang kepentingan ruang hijau bandar di Malaysia, dengan ini menyokong perancangan bandar dan membuat keputusan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

API	-	Application Programming Interface
CNN	-	Convolutional Neural Network
FCN	-	Full Convolutional Network
GSV	-	Google Street View
GVI	-	Green View Index
HRNet-OCR	-	High-Resolution Net- Object Contextual Representations
IOU	-	Intersection over Union
MAE	-	Mean Absolute Error
ResNet	-	Residual Neural Network
RMSE	-	Root Mean Squared Error
sGVI	-	Standardised Green View Index
TSV	-	Tencent Street View

CHAPTER 1

INTRODUCTION

1.1   Problem Background

Urban greens, including trees, shrubs, and associated vegetation in cities, have long been recognized for their importance in the urban environment. In fact, there have been countless studies in recent years reiterating the instrumental importance of urban forests or trees in lowering urban temperatures and creating more comfortable microclimates (Pauliet, 2003; Schwaab, 2021), providing physical and mental health benefits (Lee, 2010), etc. making the urban areas more livable for its inhabitants.

Unfortunately, there are hardly any regulations guiding the management of urban greens in Malaysia. The Malaysian urban landscape planning practices are mostly based only on open space coverage (Rusli and Rudin, 2010). For instance, under the National Urbanization Policy (NUP), two hectares of open space per 1000 urban population were set as the planning standard and a minimum of 10% or 0.2 hectares of open spaces are dedicated for all development (NUP9.ii). While open spaces are an essential part of urban greens, this single variable on itself is not enough to provide comprehensive data to inform planning that can optimize the experience of the places' inhabitants.

Along with the increase in global awareness on the importance in urban green spaces, urban green space is increasingly seen as an integral part of cities planning (James, 2009), fueling the growth of research in this field. According to the systemic review based on PRISMA framework by Rajoo et. al. (2021), there is a growing trend in research articles concerning urban forests in Malaysia from 2007 onwards. Out of the 43 records reviewed, only 4 of them were focused on the spatial analysis of urban green space (Kanniah, 2017; Masum et. Al., 2017; Kasim et. Al., 2019, Nor et. Al.,

2019). This shows an untapped opportunity where the spatial analysis on urban green coverage in Malaysia can be investigated.

Nevertheless, all of the spatial studies are based on macro-scale aerial views, making them not representative of the human experience on the ground level on the day-to-day micro level. While these studies are essential, they are not able to relate to the direct human user perspectives when they are in the city. In other words, satellite imagery is useful for high level planning of preservation and restoration, but not informative for the environmental design of the everyday human experience.

It is long overdue that we adopt better metrics to measure urban greeneries coverage that can allow us to integrate urban green into our daily experience in the Malaysian urban environment.

1.2   Problem Statement

In this study, we try to tackle the problem by measuring and mapping urban green in a Malaysian city to analyze its coverage from the eyes of the users. In other words, this study measures: "how much greeneries can people in a Malaysian city see?"

The quantified study of the urban users' views is made possible by the establishment of Green View Index (GVI), which makes use of street view images and detects the proportion of greeneries in the images collected. While the study is limited to the street views, the studies and information extracted remain to be highly informative for planners as the streets are the public space with the most human activity.

Nevertheless, there are several challenges to overcome in the study, which lead to our research questions.

1.3   Research Questions

The questions that this research attempts to answer are:

1. How do we collect the large amount of street view data required as the raw data to and study the urban green coverage?

2. Given the large amount of street view images, what kind of learning algorithms can we use to compute GVI efficiently?

3. How do we interpret and analyze the computed GVI and make it informative for the urban planning process?

1.4   Aim and Objectives

The aim of this project is to visualize the state of urban green coverage in Malaysian city from the users' perspective with Green View Index (GVI). Dashboard is developed as the final output that allows us to interpret the findings intuitively. Objectives of this research are as follow:

1. To collect and process street view images data to measure urban green coverage from the perspective of human users.

2. To compute GVI effectively from street view images with unsupervised segmentation and supervised machine learning models.

3. To develop a dashboard to visualize the estimated GVI of selected sites and assess the distribution of green views in Malaysian city.

1.5   Scope of Study

Due to limitations in time and resources, a smaller scope in Johor Bahru city center is chosen as a representative sample site for this study. It is chosen due to it being the central hub of Johor Bahru with a high population density. As there are no official boundaries demarcating the Johor Bahru city center area, local experience is used to help formulate the definition of city center area. In this case, the main road *Jalan Lingkaran Dalam* acts as the edge of the city center.

Meanwhile, the street views images used to compute GVI are collected in December 2022 from the Street View API provided by Google Maps due to its availability and accessibility. As the site has a nearly eclipse outline, the actual boundary of the studied site is approximated by a circle to facilitate the function of Google Maps. With several experimentations, the sample site is set to be a circle with a 1000m radius from the *Komtar JBCC*.
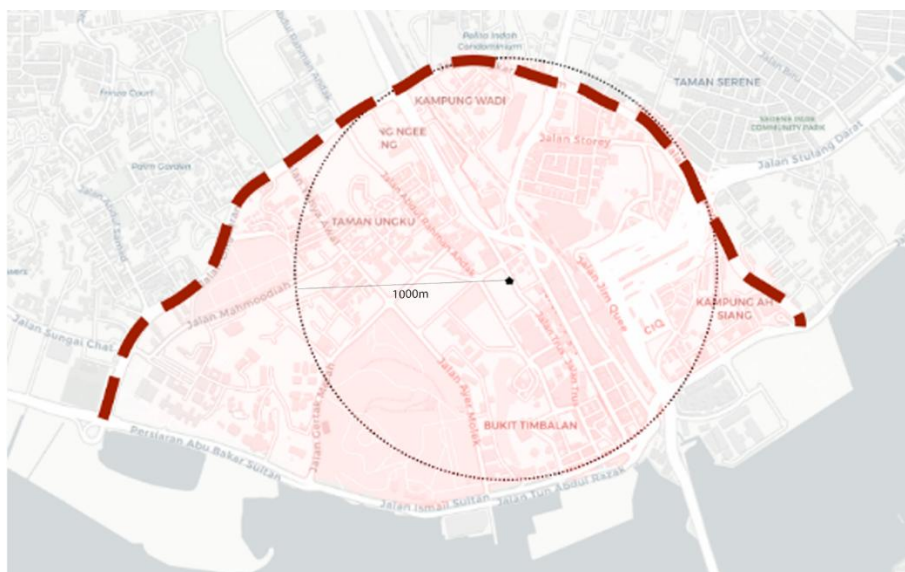
Figure 1.1   Site plan boundary.

4

1.6   Significance of Study

The project is expected to provide comprehensive insights on the urban green coverage from users' perspectives measured by GVI in Johor Bahru city centre. This information will be useful to discover the state of urban green coverage in a Malaysian city and inform the process of urban planning of the said area, while also shining some light on the conditions of the other similar urban areas in Malaysia.

Besides, this study also seeks to compare various existing methods used to compute Green View Index (GVI) to find out the most accurate methods for the task.

1.7   Project Organization

The project is laid out in 5 chapters, in the first chapter we go through the introduction to the problem background and statement, research questions and objectives, along with the scope to establish the framework of study. In Chapter 2, extensive literature review is done to understand the state of urban greeneries research in Malaysia, the use of GVI and the methods used to compute GVI. Several methods are reviewed, and their respective evaluation metrics are explored. In Chapter 3, detailed methodology for the project procedure is established, from data preparation to model development and evaluation, and lastly reporting. Chapter 4 cover initial findings about GVI in the study site. Chapter 5 concludes the project with its achievements, limitations and how it can be improved.

CHAPTER 2

LITERATURE REVIEW

2.1   Overview of Literature Review

By and large, urban greeneries in Malaysia are managed only very indirectly through the management of open space coverage. However, this is hardly an effective measure to optimize the effects of tree coverage in the urban environment. Thus, various studies on urban forests were published to investigate various effects of urban greeneries, e.g. in ecology, biodiversity, accessibility, etc. In this project, the Green View Index (GVI) measured by using street view images along with machine learning algorithms is proposed as another metric to augment the urban planning process in Malaysia.

While previous studies have successfully established the methods to measure and compute GVI in cities, there is still room for improvements in this area, particularly in the context of Malaysian urban areas.

In this literature review, we will study the research questions in several different sections. Firstly, we will establish the background context by reviewing the existing studies of urban greeneries coverage in Malaysia. Later, we dive into the literature discussing the importance and the methods to measure GVI. Lastly, algorithms and models developed in previous studies to detect and compute GVI from the street views are explored to find out the most efficient way to adopt for this project. The extracted insights will provide a foundation for our study, which aims to investigate the GVI measurement in Johor Bahru city center.

2.2   Research on Urban Greeneries in Malaysia

In Malaysia, the management of urban greeneries are based crudely on the measurement of open and green space (land specified as public area, loosely correlated with green area) coverage according to the Department of Town and Country Planning, Peninsular Malaysia (JPBD). Several standards are established, e.g., 2 hectares of open space are reserved for each 1000 urban populations, green areas to be established as buffer zones to limit urban development. While these measurements are essential for sustainable urban planning, they are not sufficient to optimize the urban greeneries.

While there is a lack of urban greeneries management practice, there are various research being done in Malaysia attempting to fill in the gap. In a systematic review done on urban forestry research with PRISMA framework (Rajoo et. al., 2021), there is a consistent growth in the research concerning urban forests in Malaysia from 2007 onwards. Nevertheless, out of the 43 records reviewed, only 4 of them were focused on the spatial analysis of urban green space (Kanniah, 2017; Masum et. Al., 2017; Kasim et. Al., 2019, Nor et. Al., 2019). This shows an untapped opportunity where the spatial analysis on urban green coverage in Malaysia can be investigated.

On the topic of urban green spatial analysis in Malaysia, the research done are typically conducted based on aerial images on macro or micro planning scale. For instance, on a high-level master planning scale, Kasim et. al. (2019) published a study that documented the changes in urban green spaces between 2002, 2012 and 2017 with the use of high-resolution aerial imagery; Kanniah (2017) made use of time-series Landsat satellite imagery to monitor green cover changes in Kuala Lumpur from 2001 to 2016. On the lower-level planning, Ludin and Rusli (2009) monitored the quality and distribution of open spaces in Johor Bahru Tengah Municipal Council with remote sensing data.

However, all of the studies mentioned above are investigated with aerial top-down view for instrumental planning. However, top-down view studies are abstract and not directly informative for the environmental design of the everyday human experience, making it not relatable to the end-user experience. To the best of our

knowledge, there is no objective study done from the perspective of urban environment users, which is a crucial measurement central to the planning of experience of the users in the urban environments. This leaves a research opportunity for such a project to happen.

2.3   Green View Index (GVI) with Street Views

Conventionally, aerial remote data has been used for the task of trees mapping or any other similar land surveying tasks. Even with the advancement of new techniques such as remote sensing methods such as LiDAR, aerial top-down view remains to be the main perspective in which urban greeneries coverage is measured. However, we have all known that the top-down view is hardly how we humans experience our environment as we perceive it in a perspective view, making the studies being hard to relate for the end-user experience of urban dwellers. This is especially relevant in urban environments where man-made facilities are juxtaposed with the mixture of trees, forming urban treescapes with massively different appearances for the same green area.

Many studies have attempted to study the visual impact of urban trees. However, a lot of them were qualitative and subjective in nature (J. Yang et. al., 2009). One of the most popular methods used were ranking, in which the participants of the survey were shown pictures or videos of urban forests and requested to score the pictures. It does not take much to understand the limitations of the studies: qualitative, subjective and unscalable. J. Yang et. al. (2009) established a new metric called Green View Index (GVI) which measures the amount of greenery that people can see on the ground at different locations in a city, laying the foundation for all the future quantified studies of visual effect of urban treescapes.

The index calculation was simple but ingenious. On strategically sampled points on a target site, eye-level photographs were taken at each point and the GVI index of each point was then interpreted by calculating the ratio between areas containing foliage over the areas of whole photos. In short, it can be defined as the

ratio of greenery within the people's field of view with a range between 0 and 1, of which 0 represents no greeneries at all and 1 means that the image is full of greeneries. Due to the streets being the main public space where urban dwellers experience, Green View Index (GVI) is typically measured by using images taken from the street. The formula to calculate GVI is shown in Equation (2.1) below. M refers to the total number of horizontal directions taken on a single point, N refers to the number of vertical view angles for each sample site; $Area_g$ refers to the area of a GSV image covered by greeneries, while $Area_g$ refers to the total area of a single GSV image.

$$GVI = \frac{\sum_{j=1}^{N} \sum_{i=1}^{M} Area_g}{\sum_{j=1}^{N} \sum_{i=1}^{M} Area_t}$$

(2.1)

Equation (2.1) shows the formula to calculate GVI.

After the establishment of GVI, it has since become the foundation for future researchers to conduct quantitative studies on urban treescapes for landscape and urban planning studies. With the development of services like Google Street View, GVI has become even more viable as a technique to map urban green view and become an increasingly popular metric for urban green space research. Table 2.1 displays several examples of the use of GVI in various studies and their applications.

Table 2.1 Examples of other GVI studies and their applications.

Application	Study	Discussion
Investigate green view distribution	View-based greenery: A three-dimensional assessment of city buildings' green visibility using Floor Green View Index (Yu et. al. 2016)	GVI is modified to Floor Green View Index (FGVI) to quantify the area of visible urban vegetation from a certain floor of building.

	How green are the streets? An analysis for central areas of Chinese cities using Tencent Street View (Long and Liu, 2017)	Analyze street greeneries in 245 central area in Chinese Cities and detect patterns of street green view distribution.
	Treepedia 2.0: Applying Deep Learning for Large-scale Quantification of Urban Tree Cover (Cai et. al., 2018)	Develop deep learning method to compute GVI with significantly higher accuracy.
City walkability	Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning (Ki and Lee, 2021)	Use deep learning to compute GVI and find out its relationship with walking time.
Socioeconomic investigation	Who lives in greener neighborhoods? The distribution of street greenery and its association with residents' socioeconomic conditions in Hartford, Connecticut, USA (Li et. al., 2015)	Aggregate GVI at the block group level to compare differences in GVI based on the socioeconomic status of an area.

2.4   GSV configurations for Street Views Collection

One of the greatest limitations of J. Yang et. al.'s study was the excessive resources required for data collection on large scale as it involved extensive manual labor to take street view photos. With the advancement of mapping services, it has

become a lot easier for us to retrieve our needed street view images via Google Street View API without needing to manually take the photos of tree coverage on the street. Other than the ability to collect a larger amount of data easily, the use of Google Street View (GSV) also allows us to be more precise on our camera settings to get unbiased data for urban street treescape mapping. However, the use of GSV requires its users to have a clear understanding on the image taking configurations for the image collection.

One of the good references for this is the street-level urban greenery assessment conducted by Li et. al., 2017. They sampled 258 points of street view data collection randomly across its study site, with at least 100m intervals on average between each point. To ensure each sampling point includes all the green areas that a pedestrian can possibly see, they took 3 panoramic views upwards, straight and downwards. In the Google Street View API parameters, the "heading" was set to 0, 60, 120, 180, 240, and 300 respectively to capture a full 360-degree panorama; the "pitch" was set to −45, 0, and 45 to capture different views and 18 images were taken at each sampling point.

In fact, other than the method mentioned above, there have been many street view image configurations used in different research for different purposes. Dong et. al. (2018) reviewed multiple street view configurations used by previous research. In their review, they used Tencent Static Image (TSV) service to simulate all the GSV configurations coming up in other research.

Table 2.2 Different configurations of GSV

Configuration Name	Description	Discussion
GVI4	GVI computed with 4 horizontal TSV images with heading angle of 90∘. (Long and Liu, 2017)	Greater numbers of headings allow for a lower field of vision (zoomed in), but also

GVI	GVI computed with 6 horizontal TSV images with heading angle of 60◦. (Zhang and Dong, 2018)	takes up more computational resources to compute.
GVI8	GVI computed with 8 horizontal TSV images with heading angle is 45◦ (Dong et. al., 2018)	
GVI18	Similar approach as method used by Li et. al. GVI computed with 6 horizontal TSV images with heading angle of 60◦ and 3 pitches for each heading. 18 images taken in total. (Li et. al., 2015)	Very comprehensive, but computationally expensive.
PGVI	GVI computed with cylindrical panorama stitched together from 6 horizontal TSV pictures. The distortion for PGVI might affect accuracy of GVI. (Cheng et. al., 2017)	Cylindrical panorama cause distortion in views.

While there are many configurations used, it is important that one experiments with different configurations and chooses a configuration that can minimizes distortion and overlap to reduce error in GVI estimation.

2.5   Prediction Models for GVI

Yang et. al. (2009) made GVI calculations by manually selecting areas containing foliage using Adobe Photoshop selection, which is time and resource consuming and requires lots of automation.

2.5.1   Unsupervised Segmentation Model

Li et. al. (2015) assessed the tree coverage in each image by extracting the green pixels with the unsupervised segmentation method, As pixels consisting of vegetation have typically higher reflectance values in the green band than red and blue, pixels consisting of greeneries can be extracted by selecting only the pixels with higher values in green band than red and blue. Nonetheless, it risks inaccuracies by including non-tree green objects and omitting shaded parts of trees that appears to be not green enough.

Figure 2.1 example of green pixels that are not actual greeneries

2.5.2 Deep Learning Models to Predict GVI

With the advancement of deep learning and computer vision models, semantic segmentation can be deployed to detect and mask the greeneries in street view images effectively. One of the most common deep learning algorithms for the task is convolutional neural network (CNN). Convolutional neural networks (CNNs) are a type of specialized neural network for processing data with a grid-like topology (LeCun et al., 1997) that have been widely used for image classification and other computer vision tasks by learning hierarchical representations of the data through the use of convolutional layers (Katole et al., 2015).

The basic building block of CNN is a convolutional layer. It extracts features from the input data by applying a set of learnable filters to the input data. These filters are designed to learn different features at different scales, such as edges, corners, textures, and patterns to produce a set of feature maps. By stacking up the convolutional layers in a neural network, the architecture allowed the use of hierarchical representations to effectively capture the spatial relationships between pixels in an image and learn complex patterns and features by itself as shown in Figure 2.2.
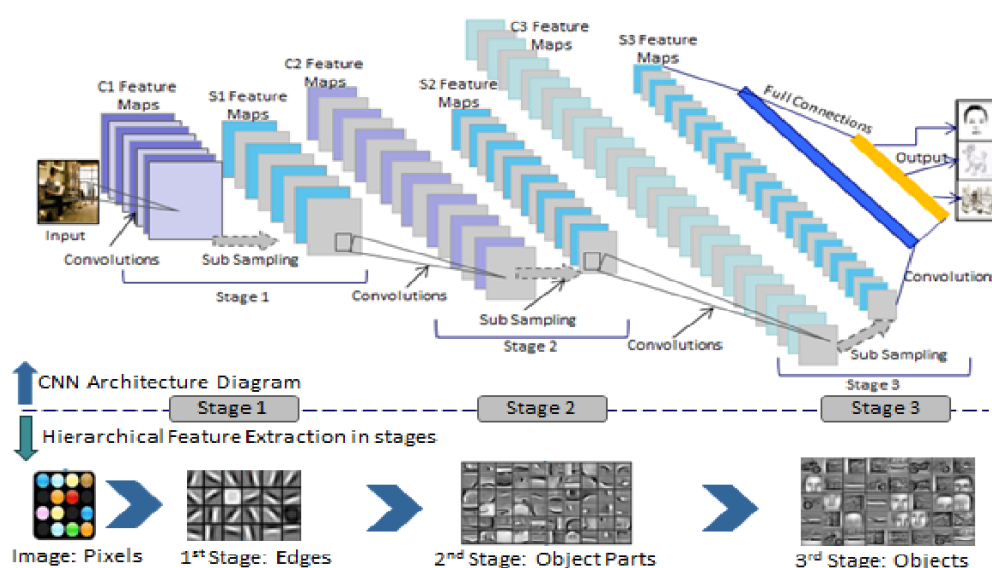

Figure 2.2 Diagram illustrating working principles of CNN.
Source: https://www.semanticscholar.org/paper/Hierarchical-Deep-Learning-Architecture-For-10K-Katole-Yellapragada/f78e280123b1c0c68f84da3cc6c66615f6e7cebd/figure/0

On top of CNN, many deep learning models are developed for image segmentation, e.g. ResNet (He et. al., 2017), FCN (Long et. al., 2015), SegNet (Badrinarayanan et. al., 2017), etc.

While CNN is the most commonly used architecture for image segmentation, there are many different models based on different architecture as well. For instance, recurrent neural networks (RNNs) based models, attention-based models, encoder-decoder based models, and others (Minaee et. al., 2022). As there are too many deep learning models developed for image segmentation, it is beyond the scope of this study to conduct an exhaustive review on every single model. So, this review will include only models that have been used to compute GVI.

Table 2.3 Summary of the deep learning models used to compute GVI.

Model	Method	Training and Calibration	Prediction Output
DCNN semantic segmentation (Cai et. al., 2018)	Pyramid Scene Parsing Network (PSPNet) (Zhao et. al., 2016) with 65,818,363 parameters	Pre-trained on full Cityscapes dataset, then trained on Cityscapes dataset, and finally on 320 GSV images collected by Cai et. al.	Pixel-segmented GSV image
DCNN end-to-end (Cai et. al., 2018)	Deep Residual Network (ResNet) (He et. al., 2017) with 28,138,601 parameters	Pre-trained on ImageNet dataset, then trained on Cityscapes dataset, and finally on 320 GSV images collected by Cai et. al.	Single GVI value between 0 and 1

HRNet-OCR (Zhang and Hu, 2022)	HRNet-OCR model (Yuan et. al., 2019) with 10,500,000 parameters	Trained on Cityscape dataset without calibration	Pixel-segmented GSV image
FCN-8 (Yu et. al., 2021)	Details of models not provided	Trained on ADE20K dataset, fine tuned with GSV images collected by author.	Pixel-segmented GSV image
DeepLabV3+ (Xia et. al., 2021)	Based on DeepLabV3+ model proposed by Chen et. al., 2018)	Trained on Cityscapes dataset, fine tuned with GSV images collected by author.	Pixel-segmented GSV image

As the details of FCN-8 model is not included in study, the accuracy measure is therefore not included in the accuracy comparison of models in Table 2.4.

Table 2.4 Accuracy comparison between both models

Model	Mean IOU (%)	Mean Absolute Error (%) compared with true GVI	Pearson Correlation Coefficient with true GVI	5-95% of GVI Estimation Error
Unsupervised Segmentation (Li et. al.)	44.7	10.1	0.708	-26.6, 18.7
DCNN semantic segmentation (Cai et. al., 2018)	61.3	7.83	0.830	-20.0, 12.37

DCNN end-to-end (Cai et. al., 2018)	NA	4.67	0.939	-10.9, 7.97
HR-NetOCR (Zhang and Hu, 2022)	80.6	NA	NA	NA
DeepLabV3+ (Xia et. al., 2021)	78.37	NA	NA	NA

According to the comparison of accuracies, we can see that HR-NetOCR (Zhang and Hu, 2022) has the highest mean IOU (%) and DCNN end-to-end model (Cai et. al., 2018) has the lowest mean absolute error compared with the true GVI. Thus, these two models will be selected and studied in this project.

2.5.3 DCNN end-to-end model

DCNN end-to-end model proposed by (Cai et. al., 2018) is based on a a 50 layered deep residual network (ResNet) architecture (He et. al., 2015) by adding 3 more layers of dense connections at the end with the final layer consisting of a single sigmoid unit. Instead of the two-step process of pixel-wise segmentation of greeneries and GVI computation, this model is designed to directly estimate GVI as its output. Therefore, a sigmoid function is used for the final layer as the logistic regression function returns value between 0 and 1, which is the same range as the GVI.

To understand the deeper mechanism of this model, it is important to get know about the ResNet architecture. ResNet is a modified version of the plain convolutional neural network to deal with the degradation problem that is common amongst very deep neural network with many layers, i.e., the accuracy of network gets saturated with increasing depth and degrades rapidly beyond the saturation point. This impedes the improvements of model performance by adding more layers

to the model. He et. al. suggests that the issue is caused by the solvers having difficulties in approximating identity mappings by multiple non-linear layers. This is because "if the added layers can be constructed as identity mappings, a deeper model should have training error no greater than its shallower counterpart", which was proven not to be the case in their experiment.

ResNet architecture deals with the challenge by introducing shortcut connections that directly connecting the input data to the output of the stacked layers that acts as identity mapping that plain deep neural network has problem with. Formally, denoting the desired underlying mapping as H(x), we let the stacked nonlinear layers fit the residual mapping of F(x) = H(x)−x instead of the desired mapping of H(x). The original mapping is recast into F(x)+x. While both functions asymptotically approximate the desired function, the residual mapping is easier to optimize as it is easier to just push F(x) to 0 in the case of identity mapping. The building block of residual learning is displayed in Figure 2.3, followed by a diagram displaying how ResNet works in Figure 2.4.
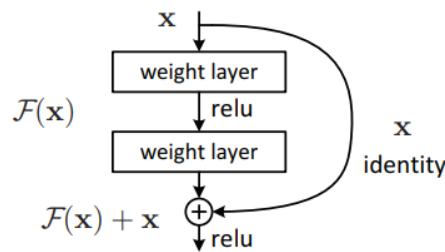

Figure 2.3 Building block of residual network.

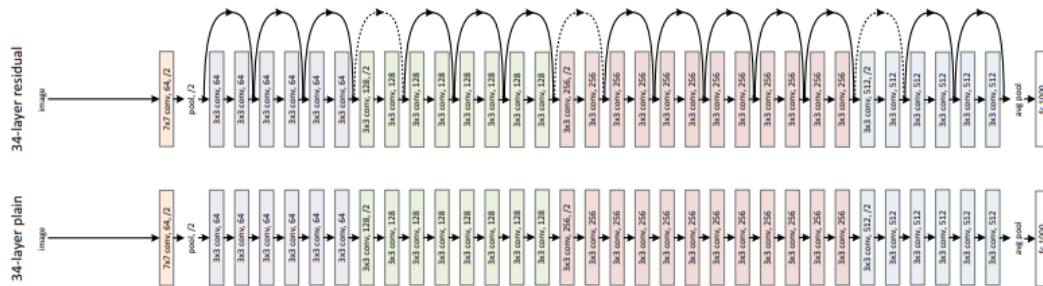source: [1512.03385] Deep Residual Learning for Image Recognition (arxiv.org)

Figure 2.4 Residual network architecture diagram vs plain network.

source: [1512.03385] Deep Residual Learning for Image Recognition (arxiv.org)

Moving back to the discussion on DCNN end-to-end model, it has 28,138,601 parameters trained and fine-tuned with several datasets. The process of model training was first initialized with weights for ResNet that have been pretrained on the ImageNet dataset, then pre-trained with the transformed Cityscapes dataset and associated true GVI labels, and finally trained on small labelled GSV dataset collected by Cai et. al.

2.5.4   HRNet-OCR Model

High-Resolution Network-Object Contextual Representation (HRNet-OCR) is a stacked image segmentation model that combines HRNet and OCRNet used by Zhang et. al. in 2022 for GVI computation. The motivation for choosing such a method for semantic segmentation is that the HRNet can be used to find out meaningful semantic features and the OCRNet explicitly transforms the pixel classification problem into an object region classification problem (Yuan et al., 2020).

HRNet serves as the backbone of the model as the computation of GVI is a position-sensitive vision problem that can be improved significantly with high-resolution representations. Nevertheless, existing deep convolutional neural networks (DCNNs) frameworks are based on low-resolution representation subnetwork that is formed by connecting high-to-low resolution convolutions in series, and recover the high-resolution representation from the encoded low-resolution representation. Compared with DCNNs, HRNet improves the performance in position-sensitive image segmentation task by maintaining high-resolution representations through the whole process, allowing for a semantically richer and spatially more precise representation (Wang et. al., 2020). The improvements are enabled via two key characteristics: (i) Connect the high-to-low resolution convolution streams in parallel maintain the high resolution instead of recovering high resolution from low resolution, and accordingly the learned representation is potentially spatially more precise*; (ii) Repeated fusions of representations from multi-solution streams generate reliable high-resolution representations with strong*

position sensitivity. In short, HRNet achieves both complete semantic information and accurate location information by parallelizing multiple branches of the resolution, coupled with the constant interaction of information between different branches (Sun et al., 2019). *The architecture of HRNet is illustrated in Figure 2.5.*
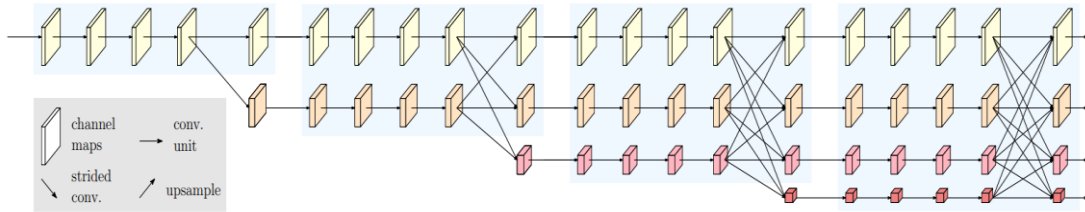

Figure 2.5 HRNet basic architecture diagram.

source: 1908.07919v2.pdf (arxiv.org)

Meanwhile, object contextual representation (OCR) is attached to the HRNet to enhance the performance of the image segmentation model. The main idea of OCR is consistent with the original definition of the semantic segmentation problem, i.e. the class of each pixel is the class of the object to which the pixel belongs (Yuan et. al., 2021). In other words, OCR takes the context of each pixel into account when assigning a class label to it. In contrast with the previous relational context schemes that consider the contextual pixels separately and only exploit the relations between pixels and contextual pixels or predict the relations only from pixels without considering the regions, the proposed approach structures the contextual pixels into object regions and exploits the relations between pixels and object regions (Yuan et. al., 2021). By incorporating object information into the network, the resultant model can better capture the context of each pixel in the image and improve the accuracy of semantic segmentation.

On top of the HRNet backbone, the contextual pixels are divided into a set of soft object regions with each corresponding to a class, i.e., a coarse soft segmentation learned under the supervision of the ground-truth segmentation. The representation of each object region is then estimated by aggregating the representations of the pixels in the corresponding object region. Lastly, the representation of each pixel is augmented with the object-contextual representation (OCR), which is the weighted

21

aggregation of all the object region representations with the weights calculated according to the relations between pixels and object regions.
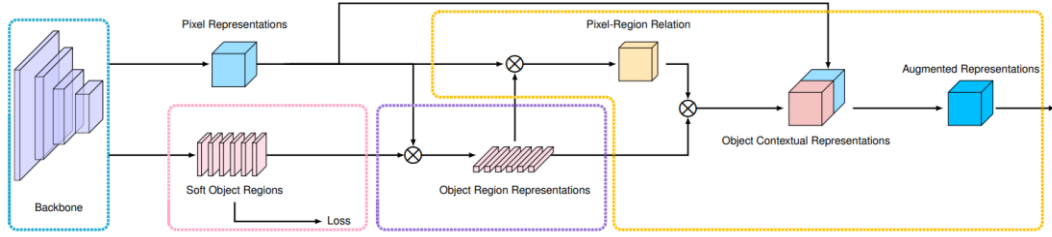

Figure 2.6 OCR basic architecture diagram.

source: https://arxiv.org/pdf/1909.11065.pdf

Zhang et. al. downloaded 5000 images with the correlated fine label in Europe open source labelled dataset Cityscapes to develop their segmentation model used for GVI calculation. The entire dataset is divided into train, validation, and test sets with the ratio of 75:10:15 respectively.

2.6   Performance Measure for GVI Prediction

Cai et. al. (2018) proposed two metrics to evaluate the performance of tree cover estimation, i.e. mean Intersection over Union (IOU) to measure the accuracy of the location of labeled greeneries pixels, and mean absolute error (MAE) for the accuracy of overall GVI.

In computer vision, the mean IOU is commonly used for object detection and segmentation. It is defined as the ratio between the area where predicted objects or pixels overlap with the target object over the area of the union of both (Padilla et. al.). In the context of GVI computation, the predicted and target values are both the pixels consisting of greeneries in the street view images.
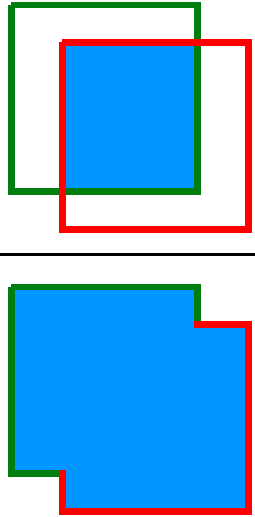
22

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$

Figure 2.7 Illustration of the intersection over union (IOU)
Source:
https://www.researchgate.net/publication/343194514_A_Survey_on_Performance_
Metrics_for_Object-Detection_Algorithms

The IOU of greeneries pixels can be calculated with formula shown in Equation (2.2). n is the number of images in test data, TP_i refers to true positive predicted greeneries label in image i, FP_i refers to the false positive predicted greeneries label in image i, FN_i refers to the false negative predicted greeneries label in image i.

$$IoU = \frac{1}{n} * \sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i + FP_i}$$

(2.2)

Equation (2.1) shows the formula to calculate IOU.

As for the accuracy of GVI value, it can be measured with the usual evaluation metrics used in regression. For instance, root mean square error (RMSE) proposed and mean absolute error (MAE) by (Dong et. al., 2018) and (Cai et. al., 2018) respectively. A model with the lowest MAE or RMSE has the highest accuracy. Their formulas are as shown below as Equation (2.3), Equation (2.4). $\hat{y}_i$ refers to the predicted label of a pixel, y_i refers to the true label of a pixel. M refers to the number of pixels in a GSV image, n refers to the number of images in test set.

23

$$RMSE_j = \sqrt{\sum_{i=1}^{M} \frac{(\hat{y}_i - y_i)^2}{M}}, \qquad \overline{RMSE} = \frac{1}{n}\sum_{j=1}^{n} RMSE_j$$

(2.3)

Equation (2.3) shows the equation to calculate RMSE.

$$MAE_j = \frac{1}{M}\sum_{i=1}^{M} |y_i - \hat{y}_i|, \qquad \overline{MAE} = \frac{1}{n}\sum_{j=1}^{n} |MAE_j|$$

(2.4)

Equation (2.4) shows the equation to calculate MAE.

Other than measuring the accuracy of prediction, Pearson correlation coefficient (r) is used to evaluate whether the predicted GVI can accurately model the underlying patterns in the true GVI. The calculation of Pearson correlation coefficient is demonstrated in Equation (2.5), where r is the correlation coefficient, x_i is the predicted GVI value for image i, $\bar{x}$ is the mean predicted GVI, y_i is the true GVI for image i, $\bar{y}$ is the mean true GVI.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

(2.5)

Equation (2.5) shows the formula to calculate Pearson correlation coefficient (r).

Finally, there is another metric used to evaluate the variance and distribution of the difference between predicted and true GVI, that is 5-95% Estimation Error. For this metric, closer the central value is to 0 and smaller the range of the value is deemed to be better performing.

2.7   Interpretation of GVI

GVI is defined to be the green view that one can see at a single point, typically on a street level. By combining the GVI extracted from many sampling points, one can

connect the dots and come up with a comprehensive plot of the visibility of urban greeneries in the studied site. For instance, the Treepedia project by the MIT Senseable City Lab is mapping GVI across many major cities in the world to explore their green distributions.

Figure 2.8 Map plots of GVI for Amsterdam and Singapore as an illustration.
Source:Treepedia :: MIT Senseable City Lab

Other than directly mapping GVI on streets, the data can also be aggregated at an area-level by mean or median to ease the extraction of meaningful information by stakeholders such as planners and local governments. Nevertheless, there is a big potential to bias if the GVI is simply aggregated by the mean of points per area due to the difference in density of data collection. Kumakoshi et. al. (2020) modified the GVI calculation to propose a Standardized Green View Index (sGVI) as a weighted aggregation of GVI scores in a study area. The formula is shown in Equation (2.6). i represents the point of GVI calculation, while l_i represents the total length of links (streets) that the point i is associated with, and l is the total length of all links in the zone.

$$sGVI = \sum_{i=1}^{n} GVI_i * \frac{l_i}{l}$$

(2.6)

Equation (2.6) shows the formula to calculate sGVI.

On top of the mapping of GVI or sGVI values on map, multiple secondary data such as the width of road (Dong et. al., 2018), ethnic distribution (Li et. al., 2015) can also be used to investigate the effects or reasons behind differences in GVI.

2.8   Issues

Overall, there are multiple issues that can be identified from the literature review. To the best of our knowledge, there are no GVI studies in Malaysia, resulting in a missing potential to measure urban greeneries coverage from the users' perspectives. As the computation of GVI was labor intensive (manual selection of greeneries), it is important that we choose an effective method to automate the extraction of greeneries from the street view images. In the literature review, we have come across multiple methods such as unsupervised segmentation (Li et. al. 2015) and deep learning models (Cai et. al., 2018) based on convolutional neural networks (CNN) such as PSNet and ResNet that were used by previous researchers for the task. It is important that we explore these methods and choose the most accurate and interpretable model to compute our GVI. Finally, it is also essential that we can visualise the computed GVI so that they can be interpreted effectively in meaningful ways to assist spatial planning of the cities.

CHAPTER 3

RESEARCH METHODOLOGY

3.1   Research Framework

The aim of this project is to develop a dashboard that can effectively map and analyse the state of urban greener coverage in Malaysia from the users' perspective with street view images. To achieve the aim, the research project is broken down into several phases that form the research framework.

In Phase 1, problem formulation is carried out to set up the foundation for research. It starts by determining the field of interest and the topic to continue research. It is then followed by literature review where the current state of research is analysed to form background understanding on the research topic and discover research gap. After that, the aim and objectives of the research is formulated and consolidated along with the defined scope of study.

After problem formulation, Phase 2 of data preparation is carried out at the beginning of the research project. In the preparation step, data is first collected, then pre-processed according to the need of the research. In the context of this research, the collected image data is converted into RGB format and masked to generate ground truth data for supervised learning.

Subsequently, the prepared data is used for models' development and evaluation in Phase 3. In this phase, unsupervised segmentation model is developed as the benchmark model to compare with the DCNN model developed to predict GVI in terms of accuracy. Finally, the predicted results are used for reporting and EDA in Phase 4. The generated insights are then presented on a dashboard as the final product.

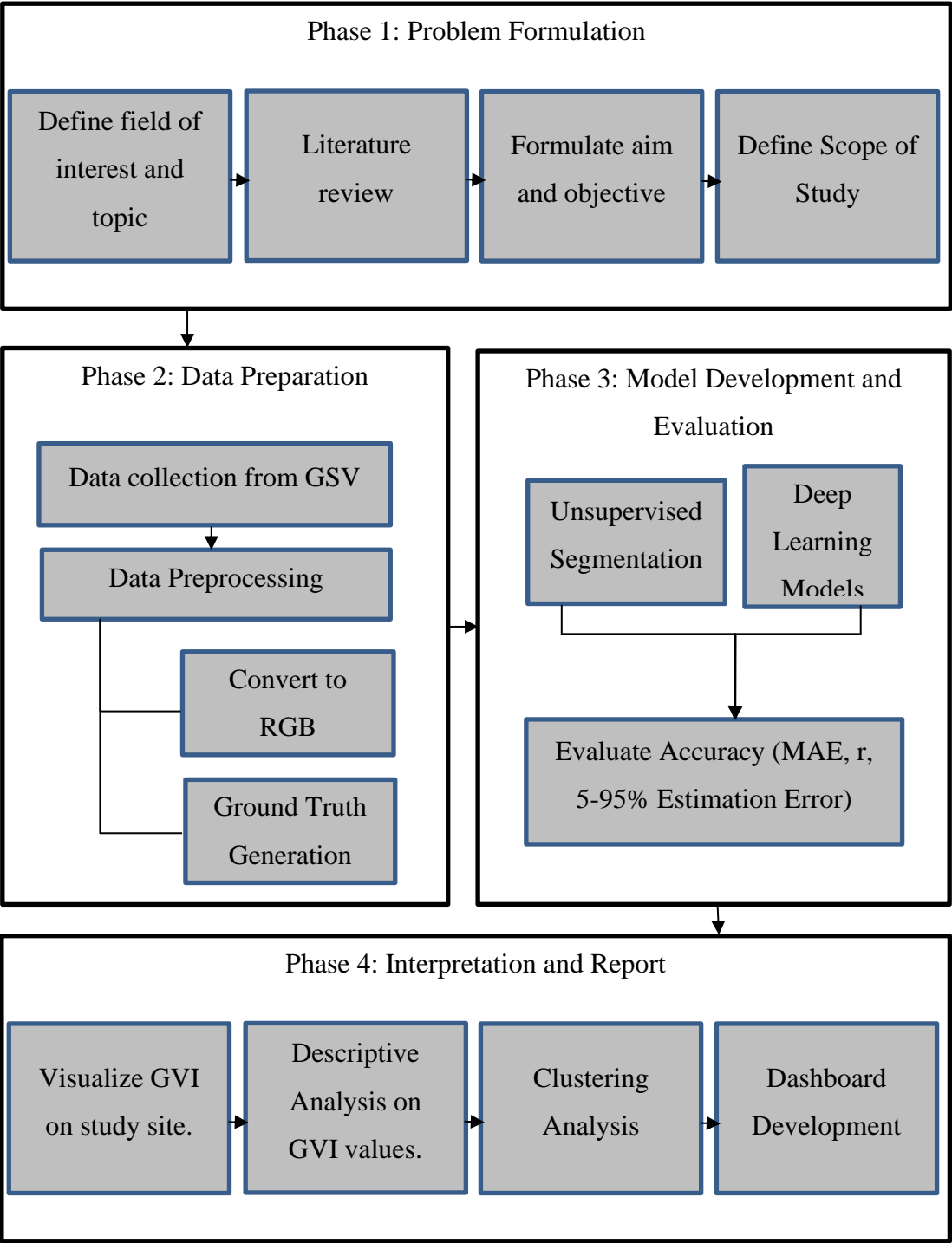The summarised diagram that displays the research framework is displayed in Figure 3.1 below.

Figure 3.1 Project research framework

3.2　Data Preparation

The data preparation for the project can be broken down into 2 parts, i.e. (1) sampling and retrieving street view data and (2) image preprocessing.

3.2.1　Sampling and Retrieving Street View Data

To limit the scope and resources required for the project in a manageable range, the methods used sampling of data collection points are improvised to fit within the Google Street View API framework. The image sampling processes are carried out in several steps:

1. **Decide on the study site and its boundary.**

The city center of Johor Bahru is selected as the study site as it has a high level of activity, movement and population density, making it a good representation of a Malaysian city suitable for our study. Due to the lack of official boundaries demarcating the Johor Bahru city center area, an expedient demarcation is made by using *Jalan Lingkaran Dalam* (dashed line) as an edge.

In order to retrieve street views from Google Street View using coordinates, the study site boundary has been defined as a circle with a radius of 1000m centered on the coordinates of Komtar JBCC. This boundary has been chosen based on several trials to ensure that it covers the busiest part of Johor Bahru city center. The boundary of the study site is illustrated in Figure 3.2 below.
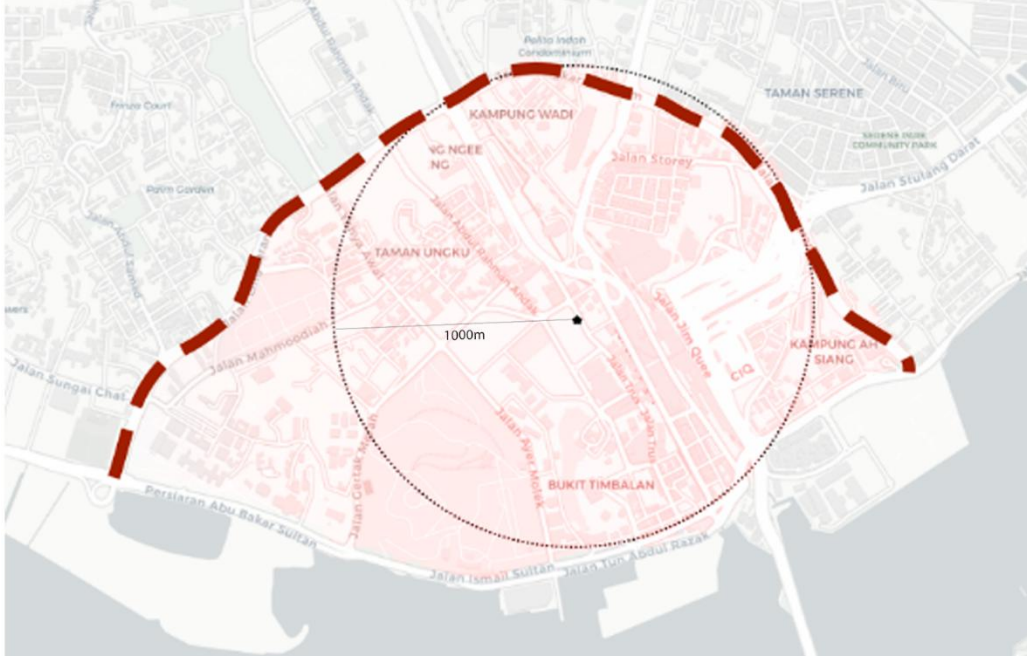
Figure 3.2 Study site plan with boundaries.

2. Sampling random coordinates on streets within the boundary.

The random coordinates are defined by using Pythagoras theorem, following Equation (3.1) to ensure that they are located within the 1000m circular boundary. The center latitude and center longitude represent the latitude and longitude of the center of study site, Komtar JBCC; latitude and longitude represent any random coordinate generated. The radius of the site boundary is converted to degrees of latitude at the equator by dividing it by the approximate distance of 111km per degree of latitude.

$$latitude \leq center\ latitude \ \pm\ \sqrt{radius^2 - (longitude - center\ longitude)^2}$$

(3.1)

Equation (3.1) shows the formula used to limit random sample points within 1000m radius from the canter coordinates *(Komtar JBCC).*

3. Retrieve GSV images from the sampled coordinates.

Various GSV parameters discussed in the literature review, e.g. GVI4, GVI, GVI8, GVI16 are explored from the perspectives of zoom, distortion, overlap of

images and cost to find out the best configuration to use for our task of GVI computation. The value for each parameter is determined in place in the following sequence.

i.   **Field of view**. The field of view of GSV images are set to the minimum of 65 to minimize the proportion of noise included in images collected, e.g. road and sky that are not typically in our field of focused vision unless special attention is given to.

ii.   **Headings**. Based on the field of view used, the headings of visuals decide the images retrieved from a single location excluded parts of the view or overlapped with one another. With trial and error, 6 pictures facing different directions of 0°, 60°, 120°, 180°, 240°, 300° to capture every view that one might see while minimizing overlap as illustrated in Figure 3.3.

iii. **Pitch**. A slight pitch of 10 degree upwards is applied to reduce the proportion of road being captured in the image.

The parameters set is shown as follows, with a for loop used to take multiple images of different headings.

```
params = {
        'key': api_key,
        'size': '1280x1280',
        'fov': 65,
        'location': f'{latitude},{longitude}',
        'heading': 0,
        'pitch': 10
    }
```

Figure 3.3 Sample view collected from a sample point at (1.4574946, 103.7690224)

In total, 6385 images from 1065 sample locations were selected. The street view images are saved as RGB format and their metadata, latitude, longitude and heading are stored as a csv file for possible future use. Table 3.1 displays the first few rows of metadata. The selected sample locations are plotted on a map (Figure 3.4) to visualize the density of sampling.

Table 3.1 The metadata of sampled images.

	lat	lng	heading
0	1.46143	103.7624	300
1	1.461554	103.7623	0
2	1.461554	103.7623	60
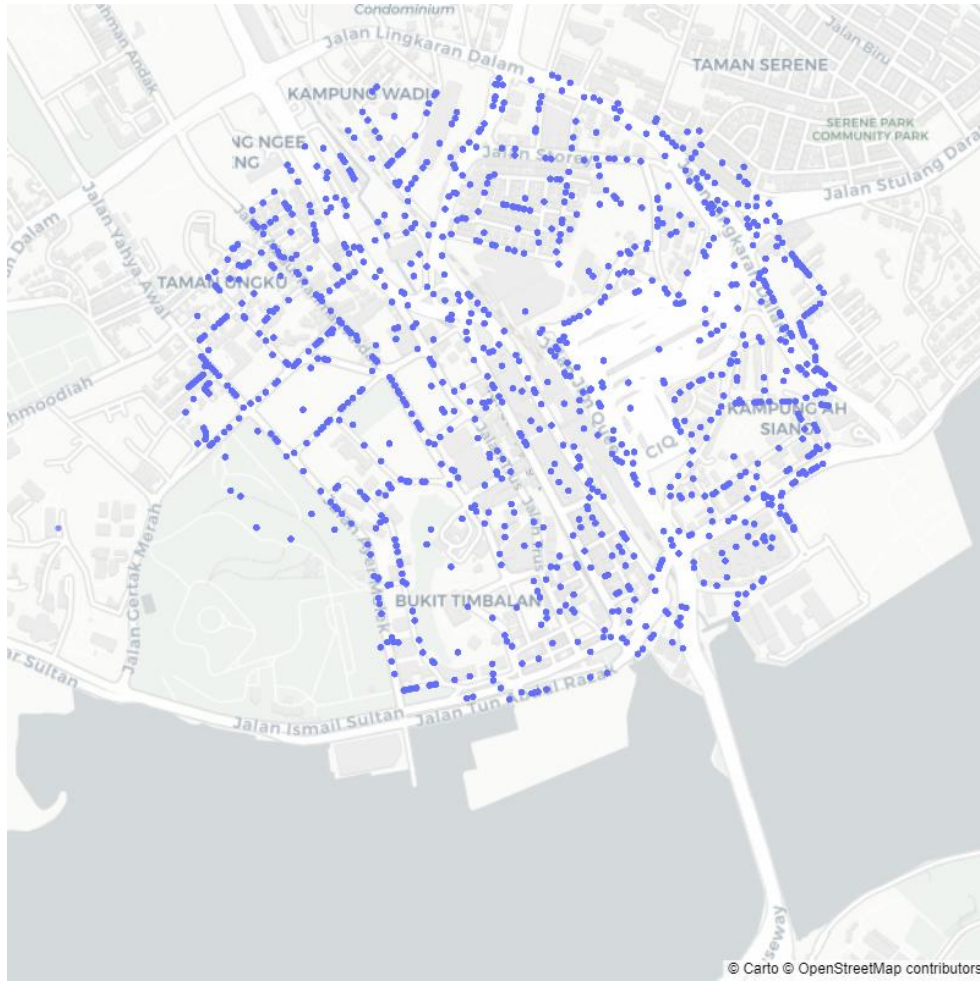
Figure 3.4 Map plot of sampled locations

3.2.2 Splitting Data for Model Development

In the *Treepedia* study published by Cai et. al, the DCNN end-to-end encrypted model deployed performed very effectively for the prediction of GVI by street view image with mean absolute error of only 4.67% and testing time of 40s per 1000 GSV images. Nonetheless, the training of the model took an incredibly long 48 hours on a system equipped with a single NVDIA 1080 Ti GPU.

In their research, 320, 80, and 100 google street view images are used for training, validation and testing respectively. This serves as a benchmark for the number of images used for model training in this project. As the model deployed was published on github: billcai/treepedia_dl_public: Treepedia 2.0: Deep Learning Based Large Scale Quantification of Urban Canopy Cover (github.com), we can use

transfer learning to train the model for the project with relatively fewer data. Thus, in this project, a slightly lower number, 240 images are used for training to reduce the training time, while 60 and 100 images are used for validation and testing respectively.

3.2.3 Data Preprocessing

To develop a deep learning model, labeled data is needed. In this case, the data for model development consists of images and the labels correspond to the presence or absence of vegetation in those images. There are two types of labeled data needed:

1. Extract vegetation pixels from the images. To create the labels, binary masks are applied to the images, where vegetation is represented by white pixels and all other areas are represented by black pixels. This helps to simplify the images and make the training process more efficient.
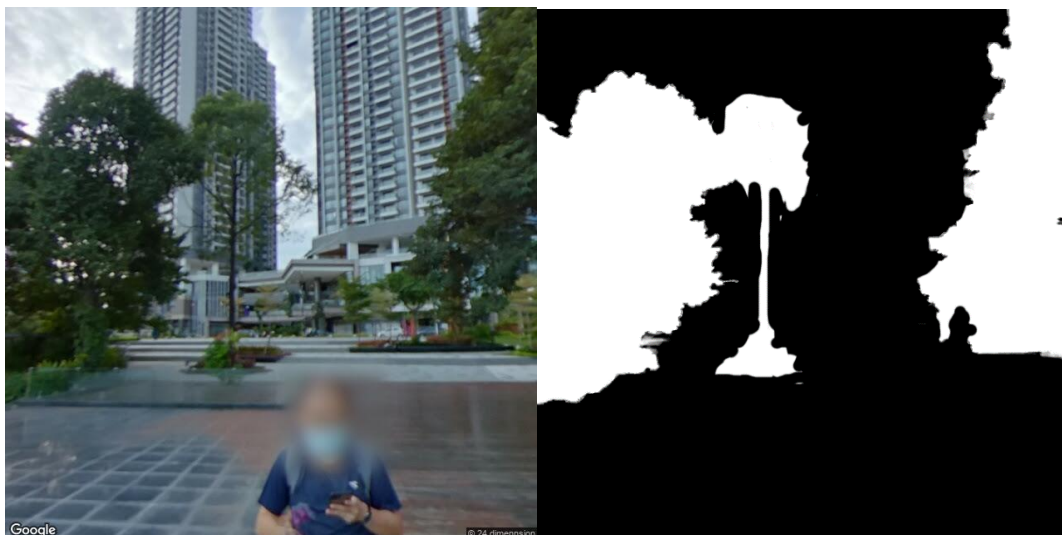
Figure 3.5 A sample of the street view image and its target mask.

2. GVI at every sample point used for model development. To develop models that can accurately predict GVI, GVI calculated from the manually extracted

vegetation pixels are used as the ground-truth. GVI is calculated with
Equation (3.2).

$$GVI = \frac{\sum_{i=1}^{n} \sum_{i=1}^{M} Area_g}{\sum_{i=1}^{n} \sum_{i=1}^{M} Area_t}$$

(3.2)

Equation (3.2) shows the formula to calculate GVI.

At the end of the data preprocessing, input data for model development is prepared: street view images represented in *Numpy* array format (training data), the true GVI of each street view image used for training and testing (target data).

3.3   Model Development for GVI Prediction

In this project, the classical unsupervised segmentation method proposed by Li et. al., 2017 is benchmarked against two supervised deep learning models DCNN end-to-end model and HRNet-OCR model.

The unsupervised segmentation model is used as the benchmark as it is the most widely adopted approach used to compute GVI of street view images. Besides, compared with supervised deep learning models, the unsupervised segmentation approach comes with several major benefits, e.g. greater interpretability, less susceptible to human induced error in the data preprocessing stage, no need of time and computation resources for model training.

Due to the benefits mentioned above, deep learning models need to predict GVI with higher accuracies to show that they worth more than what they cost.

3.3.1 Unsupervised Segmentation Method

Firstly, separate bands of red, green and blue are extracted from the RGB images. As we wish to find pixels that have higher green band values than the other two bands, differences are computed by subtracting the red band and blue band from the green band respectively. If both differences are larger than 0, they are multiplied to form a new Difference image. The generated Difference image is a grayscale mask that represents the number of differences between the green band of a pixel with the other two bands. The lighter the pixel or the higher its value, the 'greener' the pixel appears in the original picture and vice versa.

To eliminate noise, a threshold is assigned such that selected pixels with value larger than threshold are assigned to value of 255 and those below assigned 0. The final outcome classifies the selected pixels as vegetation, while the others as not vegetation. After this, the GVI of an image can be easily computed from the classification outcome with Equation (3.2). Figure 3.6 simplifies and illustrates the whole process.
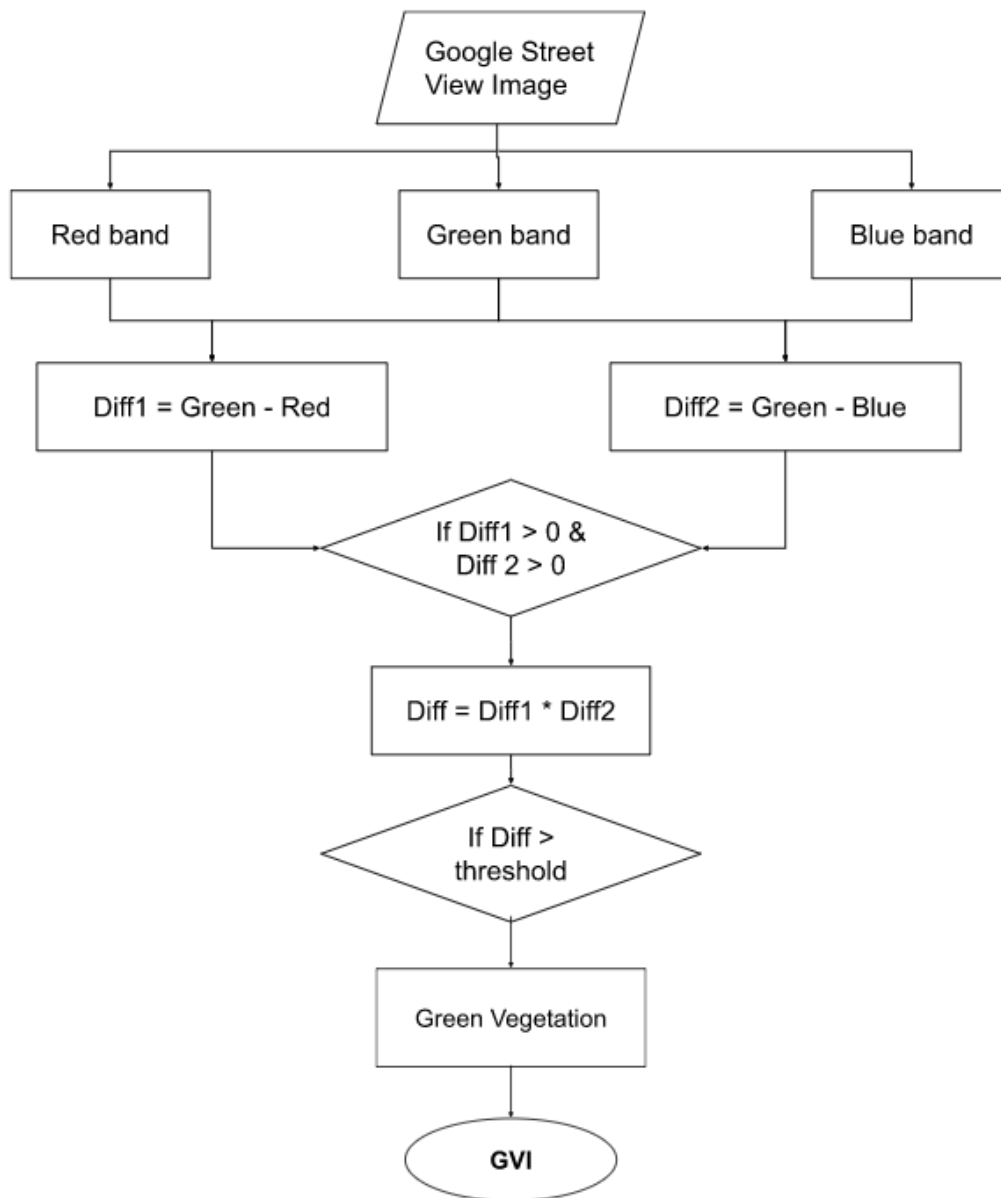
Figure 3.6 The process of unsupervised segmentation for green vegetation extraction and GVI calculation.

3.3.2 Deep Learning Models

The DCNN end-to-end model is downloaded from the Github repository uploaded by Cai: billcai/treepedia_dl_public: Treepedia 2.0: Deep Learning Based Large-Scale Quantification of Urban Canopy Cover (github.com), while HRNet-

OCR is forked from [GitHub - openseg-group/openseg.pytorch: The official Pytorch implementation of OCNet series and SegFix.](#) Nonetheless, as the output of HRNet-OCR model is pixel-segmented image which is different from that of GVI prediction generated from DCNN end-to-end model, the activation function of the last layer of neuron is changed to a sigmoid function to change the output to GVI values between 0 and 1 so that both deep learning models can be compared on the same level. Transfer learning is done on the pre-trained model by fine-tuning it based on local street view images collected. The trained value is then used to directly predict the GVI values of GSV images.
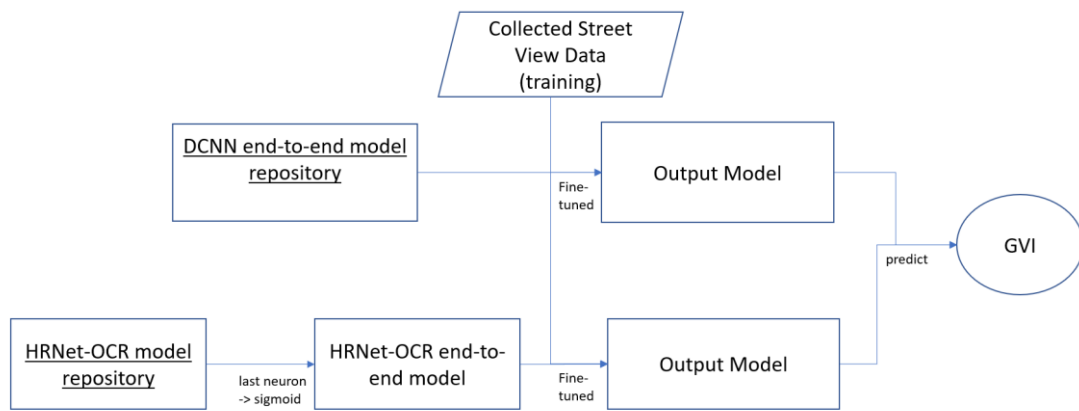
Figure 3.7 The process to develop deep learning models to predict GVI.

3.4   Evaluation of Models

The final outputs of the models are all predicted GVI values between 0 and 1. The accuracy of the prediction models are defined by its error rate: the lower its error rate, the higher its accuracy. With the 100 images used (tentatively 45 as I haven't finished labeling the data), following the metrics used by studies covered in Chapter 2, mean absolute error (MAE) in percentage, 5-95 percentile of MAE, Pearson correlation coefficient with the true GVI values (r), are used as the parameters to assess the error rates of the proposed prediction models. Smaller values for MAE and r value closer to 1 represents higher accuracy. Meanwhile, 5-95 percentile of MAE is used to measure the centrality and spread of the error. For this metric, the desired outcome is for the spread to be small and centered on 0.

Other than accuracy, the models are also evaluated by efficiency. The time taken for model training are recorded, as well as running time for the models to predict all GVI values in the study site.

3.5   Dashboard and Report

Finally, the predictions generated are used to build a dashboard that displays the prediction of GVI values at all sampling points in the study site. This helps to visualize the urban greeneries coverage in the eyes of users, allowing the planners to optimize it.

Based on GVI, we can find out the areas or points where people are less likely to see greeneries and find out the possible reasons. This can be done by conducting clustering analysis on the GVI values to find out common characteristics of clusters.

Besides, the performance and accuracies of different models are also documented to determine the most effective model used for GVI prediction.

CHAPTER 4

INITIAL FINDINGS

4.1   Performance of Models in GVI Computation

With the tentative 45 samples used for testing, the unsupervised segmentation model shows a higher accuracy in prediction compared with the base DCNN end-to-end model with initial weights prior to further training. The comparison in accuracy is displayed in Table 4.1. Due to the lack of time and the long training time required, DCNN model has not been further trained at the moment.

Table 4.1 Accuracy comparison between different models for GVI computation.

Model	Mean Absolute Error (%)	Pearson Correlation Coefficient with true GVI (r)	5-95% of GVI Estimation Error
Unsupervised segmentation model	7.31	0.909	-0.115, 0.183
DCNN end-to-end (initial weights without further training)	14.67	0.893	-0.357, 0.073
DCNN end-to-end (trained)	To be further trained	To be further trained	To be further trained
HRNet-OCR end-to-end model	To be further trained	To be further trained	To be further trained

The unsupervised segmentation model has mean absolute error, correlation coefficient with true GVI, 5-95% of GVI estimation error at 7.31%, 0.909 and (-0.115, 0.183) compared with that of 14.67%, 0.893, (-0.357, 0.093). Not only is unsupervised segmentation more accurate than the untrained DCNN model, it is more stable as it has smaller spread centered closer to 0 for 5-95% GVI estimation error.

4.2   Visualization of GVI Prediction in Study Site

Based on the model with higher accuracy, the predicted GVI values within the study site are visualized to allow further analysis on its urban vegetation coverage. In this image, red scatter plots represent spots with low GVI while green plots represent high GVI values.
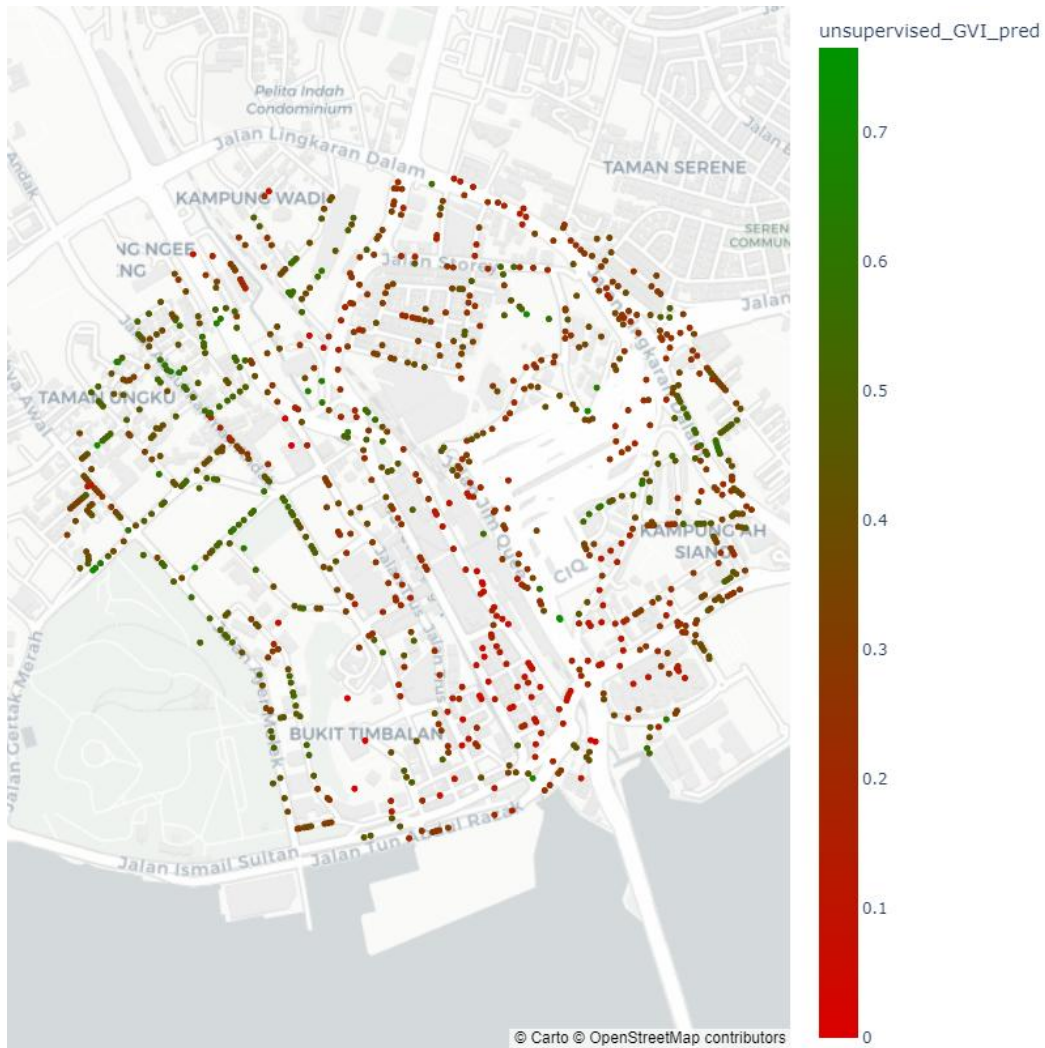
Figure 4.1 GVI predictions based on unsupervised segmentation visualized.

Based on Figure 4.1, there are several vague patterns discovered in the distribution of GVI values. For instance, the East side of the study site appears to have lower GVI values than the West; the South-East part of the site has the highest density of red points with low GVI. These observations suggest that GVI values in the study site vary due to their locations.

In order to investigate the reasons contributing to the differences, clustering analysis is carried out.

4.3   Clustering Analysis of GVI

As the pattern observed in Figure 4.1 has very irregular shape, hierarchical clustering is used to capture the clusters of different sizes and shapes. After several trials, 4 clusters are used to capture the data points with different characteristics in Figure 4.2.
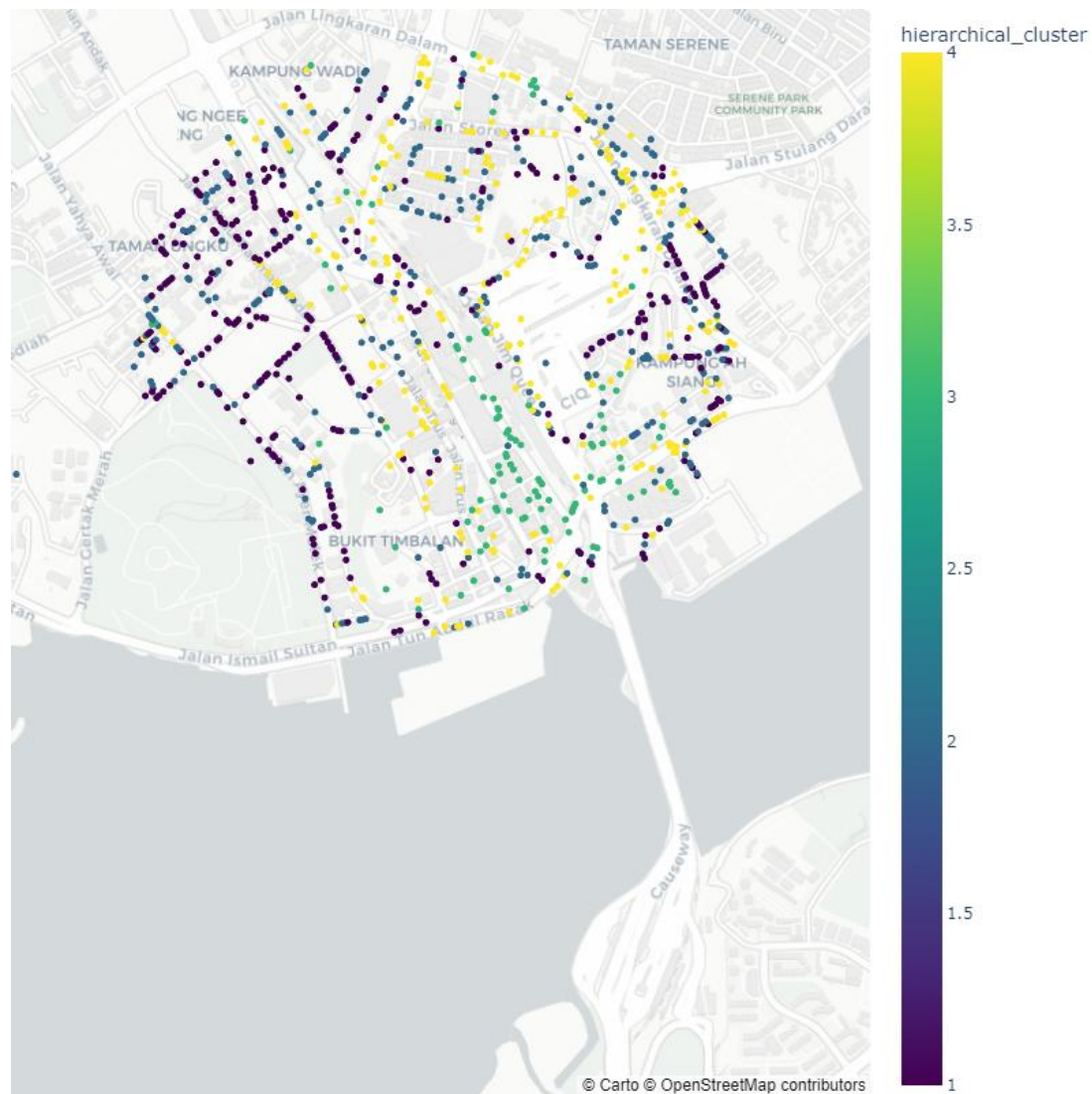
Figure 4.2 Visualization of hierarchical clusters.

The mean predicted GVI values for each cluster differ, with cluster 1 having the highest predicted GVI, followed by 2, 4 and 3 as shown in Table 4.1.

Table 4.2 shows the hierarchical cluster and GVI prediction.

hierarchical_cluster	unsupervised_GVI_pred
1	0.492871
2	0.342162
4	0.241678
3	0.114488

For the two clusters with lower mean predicted GVI value, hierarchical cluster 3 are mainly commercial units and terrace houses with medium density and data in cluster 4 are typically on very wide roads or even highways. (Illustration and further explanation to be added).

Cluster 1: Streets with adjacent to large open space with tall vegetation.

Cluster 2: Huge vehicular streets with green views at a distance.

Cluster 4: Busy streets occupie with buildings with very limited vegetation.

Cluster 3: Highways and main roads without any observable vegetation.

4.4   Chapter Summary

Unsupervised segmentation method yields decent results in predicting GVI when compared with the DCNN model without fine tuning. To continue with the study, the DCNN and HRNet-OCR models need to be fine-tuned and applied to the context of GVI prediction in study site. Based on the tentative predicted results, the distribution of green coverage from the perspective of humans on street is quantified and visualised on our site plan and clustered to uncover underlying patterns that affect GVI to inform spatial analysis of the urban environment.

CHAPTER 5

CONCLUSION

5.1   Progress Summary and Discussion

Based on several research making use of computer vision techniques to compute GVI of urban environments, this project attempts to compute GVI in Johor Bahru city center as a representative sample of Malaysian cities.

Two models were investigated to automate the computation of GVI, i.e. unsupervised segmentation model (Li et. al., 2017) and DCNN end-to-end prediction model (Cai et. al., 2018). Based on the initial findings, while both models perform with relatively high accuracy, the unsupervised segmentation model is more accurate compared with the DCNN end-to-end model provided in the study of Cai et. al. before fine tuning. This can be a result of the variance in data used. As the target data used for development of the DCNN model is extracted by tracing vegetation manually, it causes a large variance between different sets of target data. This makes the DCNN model needing to be fine-tuned and retrained every time it is adopted by different entities, which can be computationally expensive and unattractive for potential users. Meanwhile, the unsupervised segmentation is more robust as it just detects all green pixels within the images, but it has lower accuracy than the well-trained DCNN model. Depending on the needs of users, different models may be selected.

The predicted GVI values are visualized and plotted on map. This is an essential step to make the generated data useful for further analytics. In this case, clustering analysis is done to uncover the different characteristics that might contribute to the difference in GVI values. For instance, from the map plot, it is observed that wide roads have very low GVI compared to the smaller streets. This allows us to extract plausible reasons that might contribute to the increase or decrease of GVI, which are important to inform the urban planning process.

47

5.2   Limitations

In this project, we have visualized the GVI values in Johor Bahru city center that enables future research. Nonetheless, it also comes with several limitations.

Firstly, the scope of the project is too small. As the current project only covers a radius of 1000m, it only includes a very small part in Johor Bahru. The small scope does not allow us to uncover more patterns that can further inform urban planning. Besides, due to the limitations of Google API, the range of the study site is set to be a circle with constant radius. This arbitrary site boundary does not allow for more nuanced analysis on factors that can result in differences of GVI, e.g., local governance, zoning, socioeconomic level of the site, etc. A larger study site in GIS can be used to conduct more comprehensive study on a larger scale. Complementary information such as zoning and socioeconomic levels can be juxtaposed to the GVI of site to extract

Regarding the sampling process, there is an important limitation too. As the GSV images are retrieved by random coordinates within the set site boundary, the sampling does not consider the distance between sampling points. This method is suboptimal as it can either (1) not cover the study site adequately for comprehensive analysis or (2) taking too many samples at similar points which results in bias. This can be optimized by integrating GIS information consisting of street information and setting constraints such as distances between sampling points on the same street to be within a particular range.

For the calculation of GVI, the limitation lies in the variation of target data as they are created by manually tracing vegetation in Photoshop. This can introduce a lot of noise and bias for the model developed for GVI computation, which is especially so when we try to scale up and include more data. This can be improved by establishing a rigorous selection method to follow when manually selecting vegetation. As GVI is point data, they can be aggregated to produce a weighted mean GVI of an area to make them more interpretable for the stakeholders.

In conclusion, there are multiple limitations in terms of the scope of study, sampling methods and model development for GVI calculation that can be improved in further studies.

5.3   Future Improvements

Based on the limitations established, several improvements can be made in the further development of the project. Firstly, a larger site can be used to uncover more patterns and contexts behind distribution of GVI. Second, a GIS model including street information can be used to optimize the sampling method by setting the minimum and maximum distance between sampling locations, thereby reducing bias introduced. Thirdly, a more rigorous and formal process for ground truth generation when masking greeneries manually will be established to reduce the man-made variation. Finally, GVI values can be aggregated to represent weighted mean GVI of an area for greater interpretability.

REFERENCES

Abdul Aziz, N.A., Konijnendijk, C., Maruthaveeran, S., Nilsson, K., 2011.
Greenspace Planning and Management in Klang Valley, Peninsular Malaysia.
Journal of Arboriculture 37, 99–107. https://doi.org/10.48044/jauf.2011.014

Aziz, N. A. A. (2012). Green space use and management in Malaysia. Forest &
Landscape.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep Convolutional
Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on
Pattern Analysis and Machine Intelligence 39, 2481–2495.
https://doi.org/10.1109/TPAMI.2016.2644615

Cai, B.Y., Li, X., Seiferling, I., Ratti, C., 2018. Treepedia 2.0: Applying Deep Learning
for Large-Scale Quantification of Urban Tree Cover, in: 2018 IEEE
International Congress on Big Data (BigData Congress). Presented at the 2018
IEEE International Congress on Big Data (BigData Congress), IEEE, San
Francisco, CA, pp. 49–56.
https://doi.org/10.1109/BigDataCongress.2018.00014

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder
with Atrous Separable Convolution for Semantic Image Segmentation, in:
Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision
– ECCV 2018, Lecture Notes in Computer Science. Springer International
Publishing, Cham, pp. 833–851. https://doi.org/10.1007/978-3-030-01234-
2_49

Cheng, L., Chu, S., Zong, W., Li, S., Wu, J., Li, M., 2017. Use of Tencent Street View
Imagery for Visual Perception of Streets. ISPRS International Journal of Geo-
Information 6, 265. https://doi.org/10.3390/ijgi6090265

Dong, R., Zhang, Y., Zhao, J., 2018. How Green Are the Streets Within the Sixth Ring Road of Beijing? An Analysis Based on Tencent Street View Pictures and the Green View Index. International Journal of Environmental Research and Public Health 15, 1367. https://doi.org/10.3390/ijerph15071367

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition.

Hu, J., Zhang, L., Liang, W. and Wang, Z. (2009) 'Incipient mechanical fault detection based on multifractal and MTS methods', Petroleum Science, 6(2), pp. 208–216. Zhang, J., Hu, A., 2022. Analyzing green view index and green view index best path using Google street view and deep learning. Journal of Computational Design and Engineering 9, 2010–2023. https://doi.org/10.1093/jcde/qwac102

James, P., Tzoulas, K., Adams, M.D., Barber, A., Box, J., Breuste, J., Elmqvist, T., Frith, M., Gordon, C., Greening, K.L., Handley, J., Haworth, S., Kazmierczak, A.E., Johnston, M., Korpela, K., Moretti, M., Niemelä, J., Pauleit, S., Roe, M.H., Sadler, J.P., Ward Thompson, C., 2009. Towards an integrated understanding of green space in the European built environment. Urban Forestry & Urban Greening 8, 65–75. https://doi.org/10.1016/j.ufug.2009.02.001

Kanniah, K.D., 2017. Quantifying green cover change for sustainable urban planning: A case of Kuala Lumpur, Malaysia. Urban Forestry & Urban Greening 27, 287–304. https://doi.org/10.1016/j.ufug.2017.08.016

Kasim, J. A., Yusof, M. J. M., & Shafri, H. Z. M. (2019). Monitoring urban green space (UGS) changes by using high resolution aerial imagery: A case study of Kuala Lumpur, Malaysia. Pertanika J. Sci. Technol, 27, 1971-1990.

Katole, A.L., Yellapragada, K.P., Bedi, A.K., Kalra, S.S., Siva Chaitanya, M., 2015. Hierarchical Deep Learning Architecture for 10K Objects Classification, in: Computer Science & Information Technology (CS & IT). Presented at the Second International Conference on Computer Science & Engineering,

Academy & Industry Research Collaboration Center (AIRCC), pp. 77–93. https://doi.org/10.5121/csit.2015.51408

Kumakoshi, Y., Chan, S., Koizumi, H., Li, X., Yoshimura, Y., 2020. Standardized Green View Index and Quantification of Different Metrics of Urban Green Vegetation. Sustainability 12, 7434. https://doi.org/10.3390/su12187434

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), 1995.

Lee, A.C.K., Maheswaran, R., 2011. The health benefits of urban green spaces: a review of the evidence. Journal of Public Health 33, 212–222. https://doi.org/10.1093/pubmed/fdq068

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening 14, 675–685. https://doi.org/10.1016/j.ufug.2015.06.006

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Long, Y., Liu, L., 2017. How green are the streets? An analysis for central areas of Chinese cities using Tencent Street View. PLOS ONE 12, e0171110. https://doi.org/10.1371/journal.pone.0171110

Luo, J., Zhai, S., Song, G., He, X., Song, H., Chen, J., Liu, H., Feng, Y., 2022. Assessing Inequity in Green Space Exposure toward a "15-Minute City" in Zhengzhou, China: Using Deep Learning and Urban Big Data. International Journal of Environmental Research and Public Health 19, 5798. https://doi.org/10.3390/ijerph19105798

Masum, K.M., Mansor, A., Sah, S.A.M., Lim, H.S., 2017. Effect of differential forest management on land-use change (LUC) in a tropical hill forest of Malaysia.

Journal of Environmental Management 200, 468–474. https://doi.org/10.1016/j.jenvman.2017.06.009

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image Segmentation Using Deep Learning: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968

Nor, A.N.M., Abdullah, S.A., 2019. Developing Urban Green Space Classification System Using Multi-Criteria: The Case of Kuala Lumpur City, Malaysia. Journal of Landscape Ecology 12, 16–36. https://doi.org/10.2478/jlecol-2019-0002

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. In 2020 international conference on systems, signals and image processing (IWSSIP) (pp. 237-242). IEEE.

Pauleit, S., 2003. Urban street tree plantings: identifying the key requirements. Proceedings of the Institution of Civil Engineers - Municipal Engineer 156, 43–50. https://doi.org/10.1680/muen.2003.156.1.43

Rusli, N., Ludin, A.N.M., n.d. (PDF) Evaluation of Open Space and Recreation Area in Johor Bahru Tengah Municipal Council [WWW Document]. URL https://www.researchgate.net/publication/327200128_Evaluation_of_Open_Space_and_Recreation_Area_in_Johor_Bahru_Tengah_Municipal_Council (accessed 12.20.22).

Schwaab, J., Meier, R., Mussetti, G., Seneviratne, S., Bürgi, C., Davin, E.L., 2021. The role of urban trees in reducing land surface temperatures in European cities. Nat Commun 12, 6763. https://doi.org/10.1038/s41467-021-26768-w

Seiferling, I., Naik, N., Ratti, C., Proulx, R., 2017. Green streets − Quantifying and mapping urban trees with street-level imagery and computer vision. Landscape

and Urban Planning 165, 93–101. https://doi.org/10.1016/j.landurbplan.2017.05.010

Sundara Rajoo, K., Karam, D.S., Abdu, A., Rosli, Z., James Gerusu, G., 2021. Urban Forest Research in Malaysia: A Systematic Review. Forests 12, 903. https://doi.org/10.3390/f12070903

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B., 2021. Deep High-Resolution Representation Learning for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 43, 3349–3364. https://doi.org/10.1109/TPAMI.2020.2983686

Xia, Y., Yabuki, N., Fukuda, T., 2021. Development of a system for assessing the quality of urban street-level greenery using street view images and deep learning. Urban Forestry & Urban Greening 59, 126995. https://doi.org/10.1016/j.ufug.2021.126995

Yu, H., Zhou, Y., Wang, R., Qian, Z., Knibbs, L.D., Jalaludin, B., Schootman, M., McMillin, S.E., Howard, S.W., Lin, L.-Z., Zhou, P., Hu, L.-W., Liu, R.-Q., Yang, B.-Y., Chen, G., Zeng, X.-W., Feng, W., Xiang, M., Dong, G.-H., 2021. Associations between trees and grass presence with childhood asthma prevalence using deep learning image segmentation and a novel green view index. Environmental Pollution 286, 117582. https://doi.org/10.1016/j.envpol.2021.117582

Yuan, Y., Chen, X., & Wang, J. (2020, August). Object-contextual representations for semantic segmentation. In European conference on computer vision (pp. 173-190). Springer, Cham.

Zhang, Y., Dong, R., 2018. Impacts of Street-Visible Greenery on Housing Prices: Evidence from a Hedonic Price Model and a Massive Street View Image Dataset in Beijing. ISPRS International Journal of Geo-Information 7, 104. https://doi.org/10.3390/ijgi7030104

Zhang, J., Hu, A., 2022. Analyzing green view index and green view index best path using Google street view and deep learning. Journal of Computational Design and Engineering 9, 2010–2023. https://doi.org/10.1093/jcde/qwac102

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid Scene Parsing Network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, pp. 6230–6239. https://doi.org/10.1109/CVPR.2017.660